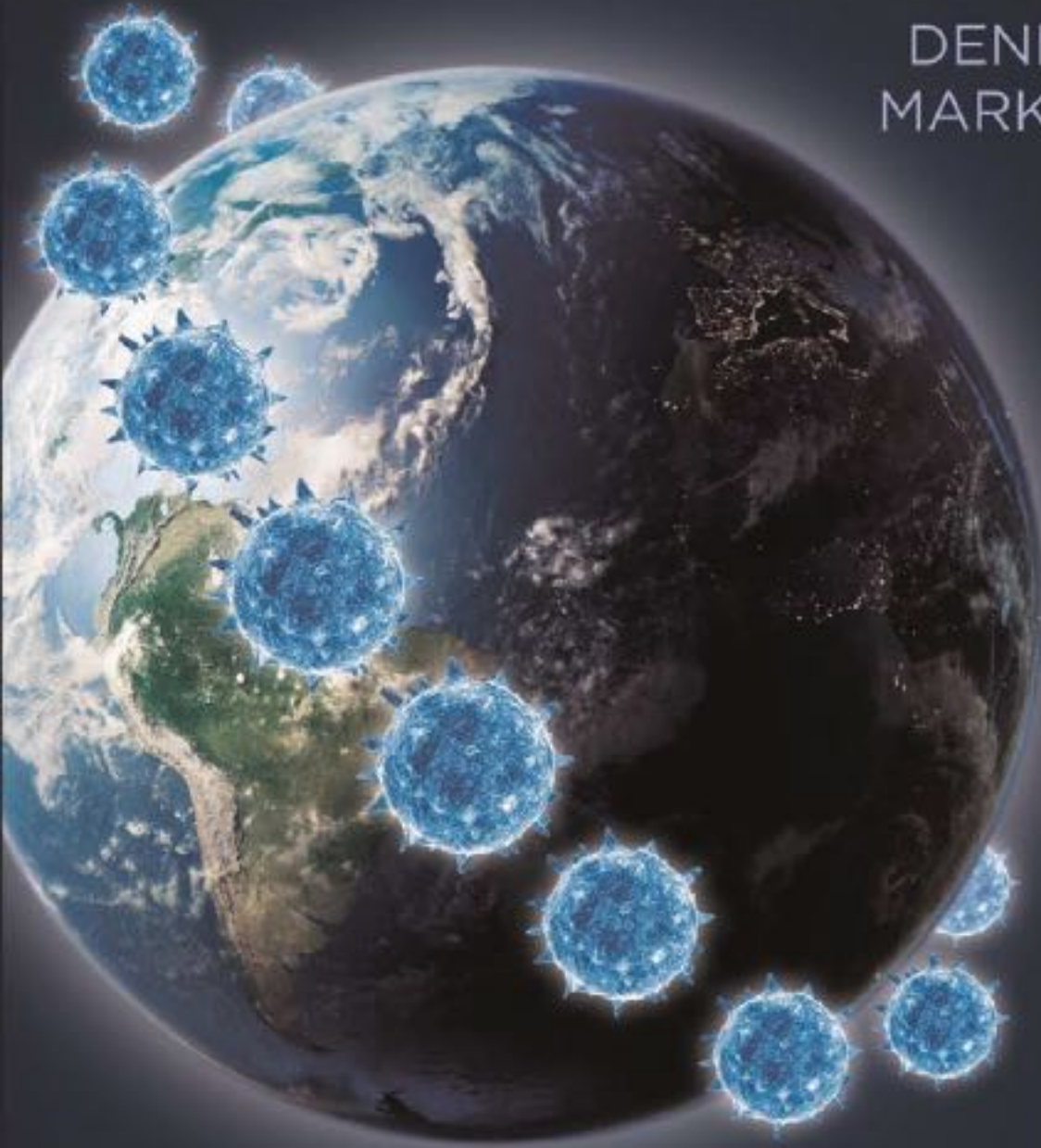


ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Editors in Chief

DENNIS BAMFORD
MARK ZUCKERMAN



ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Volume 1

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

EDITORS IN CHIEF

Dennis H. Bamford

*Molecular and Integrative Biosciences Research Programme
Faculty of Biological and Environmental Sciences
University of Helsinki, Helsinki, Finland*

Mark Zuckerman

*South London Specialist Virology Centre
King's College Hospital NHS Foundation Trust
London, United Kingdom*

and

*Department of Infectious Diseases
School of Immunology and Microbial Sciences, King's College London Medical School
London, United Kingdom*

Volume 1



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2021 Elsevier Ltd. unless otherwise stated. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-12-814515-9

For information on all publications visit our
website at <http://store.elsevier.com>

		Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org		

Publisher: Oliver Walter

Acquisitions Editor: Priscilla Braglia

Content Project Manager: Katarzyna Miklaszewska

Associate Content Project Manager: Gayathri S.

Designer: Matthew Limbert

EDITORS IN CHIEF



Dennis H. Bamford, PhD, is Professor Emeritus of Virology at the Faculty of Biological and Environmental Sciences, University of Helsinki, Finland. He obtained his PhD in 1980 from the Department of Genetics, University of Helsinki. During 1981–1982 he was an EMBO postdoctoral fellow at the Public Health Research Institute of the City of New York, United States, and during 1983–1992 he worked as a Senior Scientist at the Academy of Finland. In 1993 he was appointed Professor of General Microbiology at the University of Helsinki. He was awarded the esteemed Academy Professorship twice, in 2002–2007 and 2012–2016, and he also served twice as the Director of the Finnish Center of Excellence (in Structural Virology, 2000–2005, and in Virus Research, 2006–2011). Prof. Bamford has had continuous external research funding (e.g., from several European Union, Academy of Finland, TEKES and Jusélius Foundation funds, as well as the Human Frontier Science Program). He is an EMBO member and has held several positions of trust in scientific and administrative organizations.

Prof. Bamford has published approx. 400 articles in international peer-reviewed journals in virology, microbiology, biochemistry, and molecular biology (36 of them in high impact journals). About half of the primary articles have been published with international collaborators showing high international integration. He has also been invited to give 56 keynote and plenary presentations in major international meetings. Prof. Bamford has supervised over 35 Master's and

over 40 PhD theses. Seven of his graduate students or post docs have obtained a professorship and a similar number have a principal investigator status. Prof. Bamford has studied virus evolution from a structure-centered perspective, showing that seemingly unrelated viruses, such as bacteriophage PRD1 and human adenovirus have similar virion architecture. When the corona virion architecture was gradually revealed, it was observed that its structural elements were close to those seen in RNA bacteriophage phi6 so that phi6 has been actively used as surrogate for pathogenic viruses - quite a surprise!



Dr. Mark Zuckerman is Head of Virology, Consultant Medical Virologist, and Honorary Senior Lecturer at South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust and King's College London Medical School, Department of Infectious Diseases, School of Immunology and Microbial Sciences in London, United Kingdom. His interests include the clinical interface between developing molecular diagnostic tests relevant to the local population of patients, respiratory virus infections, herpesvirus infections in immunocompromised patients and blood-borne virus transmission incidents in the healthcare setting. He has chaired the UK Clinical Virology Network, Royal College of Pathologists Virology Specialty Advisory Committee and Virology Examiners Panel and is a member of the Specialty Advisory Committee on Transfusion Transmitted Viruses. He is a co-author on four editions of the "Mims' Medical Microbiology" textbook, has written chapters in a number of other textbooks and has over 100 publications in international peer-reviewed journals and is an associate editor for two journals.

EDITORIAL BOARD

Editors in Chief

Dennis H. Bamford

Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Mark Zuckerman

South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust, London, United Kingdom and Department of Infectious Diseases, School of Immunology and Microbial Sciences, King's College London Medical School, London, United Kingdom

Section Editors

Claude M. Fauquet

St Louis, MO, United States

Michael Feiss

Department of Microbiology and Immunology, Carver College of Medicine, University of Iowa, Iowa City, IA, United States

Elizabeth E. Fry

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Said A. Ghabrial[†]

Department of Plant Pathology, University of Kentucky, Lexington, KY, United States

Eric Hunter

Department of Pathology and Laboratory Medicine, Emory University School of Medicine and Emory Vaccine Center, Emory University, Atlanta, GA, United States

Ilkka Julkunen

Institute of Biomedicine, University of Turku, Turku, Finland

Peter J. Krell

Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Mart Krupovic

Archaeal Virology Unit, Institut Pasteur, Paris, France

Maija Lappalainen

HUS Diagnostic Center, HUSLAB, Clinical Microbiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Hubert G.M. Niesters

Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Massimo Palmarini

MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

David Prangishvili

Institut Pasteur, Paris, France and Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

David I. Stuart

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom and Diamond Light Source, Didcot, United Kingdom

Nobuhiro Suzuki

Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

[†]Deceased.

SECTION EDITORS



Claude Fauquet received his PhD in biochemistry from University Louis Pasteur in Strasbourg, France in 1974. Dr. Fauquet joined the Institut de Recherche pour le Développement (IRD) and worked there as a plant virologist for 28 years, and served in Ivory Coast, West Africa for 14 years. In 1991, he founded the International Laboratory for Tropical Agricultural Biotechnology (ILTAB) at The Scripps Research Institute, CA, United States. ILTAB was then hosted by the Donald Danforth Plant Science Center, St. Louis, MO, from 1999 to 2012. In 2003, he co-founded the Global Cassava Partnership for the 21st Century (GCP21), which he directed until 2019 and which goal is to improve the cassava crop worldwide.

Dr. Fauquet is an international leader in plant virology including taxonomy, epidemiology, molecular virology, and in gene-silencing as an antiviral strategy. He was Secretary of the International Committee on Taxonomy of Viruses (ICTV) for 18 years and the editor of several ICTV Reports including the VIIIth ICTV Report in 2005.

He has published more than 300 research papers in reviewed journals and books. He is a fellow of the American Association for the Advancement of Science, of the American Phytopathological Society and a member of the St. Louis Academy of Sciences. In 2007, Dr. Fauquet was knighted “Chevalier de l’Ordre des Palmes Académiques” by the French Minister of High Education and Research.



Dr. Michael Feiss is Professor Emeritus in the Department of Microbiology and Immunology of the Carver College of Medicine at the University of Iowa, IA, United States. Dr. Feiss received his PhD in Genetics at the University of Washington followed by a postdoctoral traineeship in the laboratory of Dr. Allan Campbell at Stanford. Dr. Feiss is a microbial geneticist who studies virus assembly with an emphasis on how a DNA virus, bacteriophage lambda, packages viral DNA into the empty prohead shell. The lab investigates how sites in the viral DNA orchestrate the initiation and termination of the DNA packaging process. This work includes comprehensive examination of the DNA recognition sites. A related interest is study of terminase, the viral DNA packaging enzyme, including the functional domains for protein–DNA and protein–protein interactions. A second focus has been the roles of the bacterial host’s IHF and DnaJ proteins in the lytic life cycle of the virus. More recent work has involved a genetic dissection of the role of terminase’s ATPase center that powers translocation of viral DNA into the prohead. This interest in the ATP hydrolysis-driven packaging motor involves a multidisciplinary collaboration examining the kinetics of DNA packaging during individual packaging events. Finally, recent studies have also looked at how the packaging process has diverged among several lambda-like phages, including phages 21, N15, and Gifsy-1.



Elizabeth E. Fry is a senior postdoctoral scientist in structural biology at the University of Oxford, Oxford, United Kingdom, where she received her DPhil. for studies relating to the structure determination of Foot-and-Mouth Disease Virus. Specializing in structural virology, Dr. Fry has studied many virus/viral protein structures but her primary focus is on picornavirus structure and function, in particular receptor interactions and virus uncoating. She is particularly interested in rationally designing virus-like-particles as next generation vaccines to reduce the inherent risks in handling live viruses.



Said A. Ghabrial[†] received his BSc in 1959 from Cairo University, Cairo, Egypt, and his PhD from Louisiana State University, Baton Rouge, LA, United States, in 1965. Dr. Ghabrial did postdoctoral research at the University of California, Davis, CA, United States, before returning to Cairo, where he served as a plant virologist in the Ministry of Agriculture. He returned to the United States in 1970 to do postdoctoral research at Purdue University, West Lafayette, IN. In 1972, he joined the Plant Pathology Department at the University of Kentucky, Lexington, KY, United States, where he rose to the rank of professor in 1986 and worked until 2013.

Dr. Ghabrial has served as an associate and senior editor of *Phytopathology*. He served on the editorial boards of the *Encyclopedia of Virology*, 3rd edition and *Encyclopedia of Plant Pathology*, and edited a thematic issue of *Advances in Virus Research* on "Mycoviruses". He was a member of the American Phytopathological Society (APS) and the American Society for Virology (ASV); in July 2002 he was elected as a Fellow of the American Phytopathological Society. He also acted as Chair of the ICTV Subcommittee on Fungal Viruses in 1987–1993 and 2011–2014.

His long professional career allowed him to make many scientific achievements in phytopathology and virology. Among them are molecular dissection of a legume-infecting RNA virus, bean pod mottle virus (BPMV), development of BPMV-based vectors, discovery of a transmissible debilitation disease of the phytopathogenic ascomycete, *Helminthosporium victoriae* (*Cochliobolus victoriae*), establishment of a viral etiology of the *H. victoriae* disease, and advancement of structural biology of diverse fungal viruses.



Eric Hunter, PhD, is Professor of Pathology and Laboratory Medicine at Emory University, Atlanta, GA, United States. He serves as Co-Director of the Emory Center for AIDS Research and is a Georgia Research Alliance Eminent Scholar.

Dr. Hunter's research focus has been the molecular virology and pathogenesis of retroviruses, including human immunodeficiency virus. He has made significant contributions to the understanding of the role of retroviral glycoprotein structural features during viral entry and providing unique insights into the assembly and replication of this virus family. In recent years the emphasis of his research has been on HIV transmission and pathogenesis, defining the extreme genetic bottleneck and selection of viruses with unique traits during HIV heterosexual transmission. He has described the selection of fitter viruses at the target mucosa, a gender difference in the extent of selection bias, and a role for genital inflammation in reducing selection. His research has defined the impact of HIV adaptation to the cellular immune response on immune recognition and control of HIV after

transmission, as well as on virus replicative fitness in vitro and in vivo. Recent work highlights the roles that virus replicative fitness and sex of the host play in defining disease progression in a newly infected individual. His bibliography includes over 300 peer-reviewed articles, reviews, and book chapters. He has also been the recipient of four NIH merit awards for his work on retrovirus and HIV molecular biology.

Dr. Hunter served as the Editor in Chief of the journal *AIDS Research and Human Retroviruses* for 10 years. He was Chair of the AIDS Vaccine Research Subcommittee which is charged with providing advice and consultation on AIDS vaccine research to the National Institute of Allergy and Infectious Diseases and continues to serve on editorial boards for several academic journals and on external advisory committees for several government, academic, and commercial institutions.



Ilkka Julkunen graduated as an MD/PhD in 1984 from the Department of Virology, University of Helsinki, Helsinki, Finland. He worked as a postdoctoral research fellow at Memorial Sloan-Kettering Cancer Center in New York, United States, in 1986–1989, followed by positions as a senior scientist, group leader and research professor at Finnish Institute for Health and Welfare in 1989–2013. In 2013 he became a Professor of Virology at the University of Turku, Turku, Finland. The research interests of Dr. Julkunen have concentrated on innate and adaptive humoral immunity in viral and microbial infections. He has studied intracellular signaling and RIG-I and TLR-mediated activation of interferon system in human macrophages and dendritic cells and stable cell lines in response to human and avian influenza, Sendai, Zika and coronavirus infections. In addition, he has analyzed the downregulation of innate immunity by viral regulatory proteins from influenza, HCV, flavi-, filo- and coronaviruses. He has expertise in vaccinology, biotechnology and development of methods to analyze antiviral immunity, he has also been actively involved in research training and collaborations with biotechnological industry.

[†]Deceased.



Peter Krell started his career in virology early as a summer high school student working for the Canadian Forestry Service studying the resistance of nuclear polyhedrosis viruses (now called baculoviruses) to environmental exposure with Dr. Fred T. Bird at the Insect Pathology Research Institute in Sault Ste. Marie, ON, Canada. He received his BSc and MSc in biology from Carleton University studying the iridovirus *Tipula Iridescent Virus* with Dr Peter Lee, in Ottawa, the Canadian capital. For his PhD he headed east to Dalhousie University in Halifax, Nova Scotia on the Atlantic coast. In addition to enjoying the salt sea air, fresh cod, lobster and mussels, he studied the molecular biology of polydnaviruses under the guidance of Dr Don Stoltz. Heading south to Texas A&M University in College Station, TX, United States, as a Postdoctoral Fellow he worked with Dr. Max Summer, of baculovirus fame, and Dr. Brad Vinson continuing to study polydnaviruses, but also became steeped in the early days of molecular baculovirology. He then accepted a faculty position in the Department of Microbiology and Immunology at the University of Guelph in Guelph, ON, Canada. There he switched to baculovirus research, which was more tractable, due in part to available cell cultures and focused on viral DNA replication and functional genomics,

particularly on chitinase, cathepsin and ME53. In collaboration with Dr. Eva Nagy he studied molecular biology of different animal viruses, notably Fowl Avian adenoviruses and their development as vaccine vectors, but also on the birnavirus infectious pancreatic necrosis virus, the coronavirus porcine endemic diarrhea virus, fowlpox virus and the paramyxovirus Newcastle disease virus. He has been involved extensively with virus taxonomy, being active in the International Committee on Taxonomy of Viruses (ICTV) as member of the *Polydnaviridae* and *Baculoviridae* study groups, national representative of Canada on the ICTV, member of the Executive Committee for the ICTV and Chair of the ICTV Invertebrate Virus Subcommittee. In terms of governance, Peter Krell was President of the Canadian Society of Microbiology, Secretary and later President of the Society for Invertebrate Pathology, as well as being on the Editorial Boards of the *Canadian Journal of Microbiology* and the *ASM Journal of Virology*. While at the University of Guelph, he rose through the ranks to Professor and is currently University Professor Emeritus.



Mart Krupovic is the Head of the Archaeal Virology Unit in the Department of Microbiology at the Institut Pasteur of Paris, France. He received his MSc in Biochemistry in 2005 from the Vilnius University, Vilnius, Lithuania and PhD in 2010 in general microbiology from the University of Helsinki, Helsinki, Finland. His current research focuses on the diversity, origin, and evolution of viruses, as well as molecular mechanisms of virus–host interactions in archaea. He has published over 170 journal articles and serves as an editor or on the editorial boards of *Biology Direct*, *Research in Microbiology*, *Scientific Reports*, *Virology*, and *Virus Evolution*. He is also a member of the Executive Committee of the International Committee on Taxonomy of Viruses (ICTV) and chairs the Archaeal Viruses Subcommittee of the ICTV.



Maija Lappalainen, MD, PhD, Associate Professor of Clinical Microbiology, is the Head of Clinical Microbiology in the HUS Diagnostic Center, HUSLAB, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. In her thesis during the years 1987–1992 she studied the incidence and diagnostics of congenital toxoplasmosis. After PhD, her research interest has been in diagnostic clinical virology, viral hepatitis, respiratory infections, viral infections in the immunocompromised patients and viral infections during pregnancy.



Hubert G.M. Niesters (1958) studied biology and chemistry in Nijmegen, the Netherlands. After obtaining his PhD in Utrecht (Prof. dr. M. Horzinek and Prof. dr. B. van der Zeijst, 1987) on the molecular epidemiology of infectious bronchitis virus, he worked as a post-doctoral fellow with Prof. dr. Jim Strauss at the California Institute of Technology (Pasadena, United States) on the replication of Alphaviruses. He received a Niels Stensen fellowship (The Netherlands) and an E.S. Gosney fellowship (Caltech) during this period.

After returning to the Netherlands (1989), he became a research associate in medical microbiology at the Diagnostic Medical Center (Delft) but moved back to clinical virology as a senior research associate in 1991 at the Erasmus University Medical Center Rotterdam (Head Prof. dr. Ab Osterhaus). From 1993 to 2007, he was responsible for the molecular diagnostics unit. During this period, he was involved in the discovery and characterization of several new viruses and variants. In 2007, he became full professor and director of the Laboratory of Clinical Virology within the Department of Medical Microbiology at the University Medical Center Groningen and University of Groningen.

He has been actively involved in the implementation and development of new technologies like real-time amplification and automation within clinical virology. He has been focusing on molecular diagnostics and its use and the clinical value in a transplant setting, as well as in monitoring treatment of hepatitis viruses. Recently, his interest focuses on rapid regional epidemiology, automation including MiddleWare solutions for molecular diagnostics, as well as the cost-benefit of rapid point-of-impact molecular testing. Special interest is focused on raising awareness for the detection of enteroviruses (enterovirus D68) and its relationship with acute flaccid myelitis (AFM).

Since 2017, he is the Chair of the executive board of QCMD (Quality Control of Molecular Diagnostics, Glasgow). He is an auditor and team leader for the Dutch Council of Accreditation and Co-Editor in Chief of the *Journal of Clinical Virology*. He is an (co)-author of more than 250 peer-reviewed papers, chapters and reviews including emerging viruses, such as enterovirus D68 and hepatitis E virus (H-index 80).

For his entire work, he received in 2016 the “Ed Nowakowski Senior Memorial Clinical Virology Award” from the Pan American Society for Clinical Virology.



Massimo Palmarini is the Director of the MRC-University of Glasgow Centre for Virus Research and Chair of Virology at the University of Glasgow, Glasgow, United Kingdom. A veterinarian by training, his research programs focus on the biology, evolution and pathogenesis of arboviruses and the mechanisms of virus cross-species transmission. His work is funded by the MRC and the Wellcome Trust. Massimo Palmarini has been elected Fellow of the Academy of Medical Sciences, of the Royal Society of Edinburgh and of the Royal Society of Biology and he was a Wolfson-Royal Society Research Merit Awardee. He is a Wellcome Trust Investigator.



David Prangishvili, PhD, Honorary Professor at the Institut Pasteur, Paris, France, and Professor at Tbilisi State University, Tbilisi, Georgia, is one of the pioneers in studies on the biology of Archaea and their viruses. His scientific career spans ex-USSR (Institute of Molecular Biology, Moscow; 1970–1976), Georgia (Georgian National Academy of Sciences, Tbilisi; 1976–1991), Germany (Max-Planck Institute for Biochemistry, Munich; University of Regensburg; 1991–2004) and France (Institut Pasteur, Paris, 2004–2020). In the research groups headed by him, several dozens of new species and eight new families of archaeal viruses have been discovered and characterized, which display remarkable diversity of unique morphotypes and exceptional genome contents. The results of his research contribute to the knowledge on viral diversity on our planet and change the field of prokaryotic virology, leading to the notion that viruses of

hyperthermophilic Archaea form a particular group in the viral world, distinctive from viruses of Bacteria and Eukarya, and to the recognition of the virosphere of Archaea as one of the distinct features of this Domain of Life. David Prangishvili is a member of the Academia Europaea, the European Academy of Microbiology, and the Georgian National Academy of Sciences.



David I. Stuart is MRC Professor of Structural Biology in the Nuffield Department of Medicine, Oxford University, Oxford, United Kingdom, Life Science Director at Diamond Light Source and Director of Instruct-ERIC (pan-European organisation providing shared access to infrastructure and methods for structural biology). He has diverse interests in structural virology from picornaviruses, double-stranded RNA viruses and enveloped RNA viruses. His drive to develop structural techniques led to the determination of the structure of Bluetongue virus (1995) and then the first membrane containing virus, PRD1. More recently, he has been at the fore-front of bringing Cryo-EM technology to bear on virus structure determination and its future role in visualizing virus function in cellulo. In addition to basic science he has a strong commitment to structural vaccinology and the development of antiviral drugs.



Dr. Nobuhiro Suzuki, PhD, received his MSc (1985) in phytopathology and PhD (1989) in virology from Tohoku University in Sendai, Japan. Dr. Suzuki currently serves as a full Professor of the Institute of Plant Stress and Resources, formerly Institute of Plant Sciences and Bioresources at Okayama University and as an Editor of *Virus Research*, *Frontiers in Virology*, *Journal of General Plant Pathology*, *Virology Journal*, and *Biology*. He has also been Guest Editor to *PLoS Pathogens*, *PNAS*, and *mBio*, and an Editorial Board member of *Virology* and *Journal of Virology*.

Suzuki Laboratory focuses on characterization of diverse viruses infecting phytopathogenic fungi and exploration of their interplays. Recent achievements include the discovery of a neo-virus lifestyle exhibited by a (+)ssRNA virus and an unrelated dsRNA virus in a plant pathogenic fungus and of multilayer antiviral defense in fungi involving Dicer. Prior to coming to Kurashiki, Okayama Prefecture, he was a visiting fellow of the Center for Agricultural Biotechnology at the

University of Maryland Biotechnology Institute (UMBI), College Park, MA, United States, for 4 years (1997–2001) to study molecular biology of hypoviruses in the laboratory of Professor Donald L. Nuss. Before visiting UMBI, he served as an assistant professor and a lecturer of the Biotechnology Institute at the Akita Prefectural College of Agriculture, Japan, for 11 years (1988–1998) where he was engaged in a project on molecular characterization of rice dwarf phytovereovirus, a member of the family *Reoviridae*. He received awards from the Japanese Phytopathological Society of Japan and Japanese Society for Virology for his outstanding achievements in plant and fungal virology.

FOREWORD

I am delighted to write the foreword to this wonderful Fourth Edition of the *Encyclopedia of Virology*. The Third Edition was published in 2008, how the world has changed in the intervening years. The release of the updated fourth edition could not be more timely or more prescient. It is superb and a huge tribute to the authors, Elsevier the publisher, and to the brilliant editors, Dennis Bamford and Mark Zuckerman.

SARS-CoV-2 has dominated the world since it emerged in 2019 and affected every continent and every aspect of life. A reminder, if it were needed, of the impact of infectious diseases, the importance of virology and the vulnerability and interconnectivity of our world.

There is no doubt that with rapidly changing ecology, urbanization, climate change, increased travel, and fragile public health systems, epidemics and pandemics will become more frequent, more complex and harder to prevent and contain. Most of these epidemics will be caused by viruses, those we know about and maybe able to predict and some we do not know of that will emerge from animals, plants or the environment. Our changing climate will change the epidemiology of viruses, their vectors and the infections they cause, hence the critical importance of this totally revised Fourth Edition of the *Encyclopedia of Virology* which brings together research and an understanding of viruses in animals, plants, bacteria and fungi, the environment, and among humans. Never has a holistic, one-health understanding been more important.

That starts with an understanding of the fundamentals of virology, a field of science that has been transformed in the years since the Third Edition. An understanding transformed by embracing traditional fields of molecular and structural biology, genomics, and influenced by immunology, genetics, pharmacology and increasingly by epidemiology and mathematics. Events of 2020 and 2021 also show why it is so important to integrate within traditional virology an understanding of the animal and human health and behavior, of climate change and its impact on the ecology of viruses, plant sciences and vectors. And why we must understand the viruses we think we know well, and those viruses less extensively studied. Research is critical to this, research that pushes the boundaries of what we know, has the humility to seek answers to things we do not understand and shares that knowledge with the widest possible community. That research will be most exciting at the interface between disciplines, most impactful when dynamic, open, inclusive, global, and collaborative.

This is what the Fourth Edition of the *Encyclopedia of Virology*, the largest reference source of research in virology sets out to achieve. It is a wonderful contribution to a critical field of knowledge. It contains new chapters, every chapter revised and updated by a dedicated global community who have come together to provide what is a brilliant and inspiring reference. It is an honor to contribute in a very small way to the timely release of the Fourth Edition of the *Encyclopedia of Virology*.

Jeremy Farrar

PREFACE

The fourth edition of the Encyclopedia of Virology is encyclopedic, but we wanted to move away from an alphabetical list, apart from where it was more logical, to a vision that encompassed a different structure. Articles describing novel trends as well as original discoveries in specific subfields of virology have been distributed into a set of five volumes, namely Fundamentals of Virology, Human and Animal Viruses, Plant Viruses, Bacterial, Archaeal, Fungal, Algal and Invertebrate Viruses, and Diagnosis, Treatment and Prevention of Virus Infections.

We had hoped that the new edition would 'go viral' but it was ironic that the time to publication 12 years after the previous edition had been made a bit longer due to a virus infection. The world encountered a devastating global pandemic, COVID-19, caused by a new type of a coronavirus, SARS-CoV-2. Scientists in many disciplines all over the world started immediate efforts to discover solutions as to how to mitigate and stop the spread of the pandemic. Virology moved from being a highly specialized subject to one in which everyone became a virologist, proving just how significant the different aspects of virology are in terms of understanding the nature of viral infection. Since the previous edition, the growth in the field of general virology has been enormous, including huge advances in basic science, identification of novel viruses, diagnostic methods, treatment and prevention. Taking this into account, the introduction of the articles within the Encyclopedia are very timely and crucial for providing a wealth of knowledge of the latest findings in the field of virology to a vast range of people, whether school students, undergraduates, postgraduates, teachers, scientists, researchers, journalists and others interested in infections and the conflict between the host and the pathogen.

Pandemic viruses have become a serious public concern in the changing world. We can ask ourselves whether we have reached the point in which nature can no longer cope with the consequences of increased population density and human activities that are harmful to the environment. Although several pandemics have threatened mankind before, this COVID-19 pandemic has highlighted the massive adverse economic consequences towards the wellbeing of society and the importance of research in virology.

We aimed to produce a Major Reference Work that differs in approach to others and binds all the virology disciplines together. Chapters have been included on origin, evolution and emergence of viruses, environmental virology and ecology, epidemiology, techniques for studying viruses, viral life cycles, structure, entry, genome and replication, assembly and packaging and taxonomy and viral–host interactions. Information has been included on all known species of viruses infecting bacteria, fungi, plants, vertebrates and invertebrates. Additional topics include antiviral classification and examples of their use in management of infection, diagnostic assays and vaccines, as well as the economic importance of viral diseases of crops and their control.

This edition used viral classification according to the 9th Report of the International Committee on Taxonomy of Viruses published in 2012. Updating it to the 10th Report in 2020 was affected by the pandemic and can be found online at <http://ictv.global/report/>.

We wish to acknowledge the hard work, interest, flexibility and patience, during such difficult times both socially and professionally, of everybody involved in the process of writing this edition of the Encyclopedia of Virology, especially Katarzyna Miklaszewska, Priscilla Braglia, Sam Crowe and colleagues at Elsevier. We sincerely thank all the authors and section editors for their excellent contributions to this edition.

Book Cover Image: Viruses are obligate parasites and all cells have their own viruses increasing the total number of viruses to the estimated astronomical number of 10^{31} that extends the number of stars in the universe. The viral string illustrates how pandemic viruses surround the globe. The original picture was created by Dr. Nina Atanasova (Finnish Meteorological Institute and University of Helsinki) and amended by Matthew Limbert at Elsevier.

*Dennis H. Bamford
Mark Zuckerman*

HOW TO USE THE ENCYCLOPEDIA

Structure of the Encyclopedia

All articles in the encyclopedia are arranged thematically as a series of entries within subjects/sections, apart from volume 2 where there it was more logical to have articles arranged alphabetically.

There are three features to help you easily find the topic you are interested in: a thematic contents list, a full subject index, and contributors.

1. Thematic contents list: The alphabetical contents list, which appears at the front of each volume, lists the entries in the order that they appear in the encyclopedia.
2. Index: The index appears at the end of volume 5 and includes page numbers for quick reference to the information you are looking for. The index entries differentiate between references to a whole entry, a part of an entry, and a table or figure.
3. Contributors: At the start of each volume there is a list of the authors who contributed to all volumes.

LIST OF CONTRIBUTORS

Stephen T. Abedon

*The Ohio State University, Mansfield, OH,
United States*

Peter Abrahamian

*Agricultural Research Service, US Department of
Agriculture, Beltsville, MD, United States*

Jônatas S. Abrahão

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Florence Abravanel

*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Nicola G.A. Abrescia

*Center for Cooperative Research in Biosciences, Basque
Research and Technology Alliance, Derio, Spain;
Ikerbasque, Basque Foundation for Science, Bilbao,
Spain; and Center for Biomedical Research in the Liver
and Digestive Diseases Network, Carlos III Health
Institute, Madrid, Spain*

Gian Paolo Accotto

*Institute for Sustainable Plant Protection, National
Research Council of Italy, Torino, Italy*

Elisabeth Adderson

*St. Jude Children's Research Hospital, Memphis, TN,
United States and University of Tennessee Health
Sciences Center, Memphis, TN, United States*

Mustafa Adhab

University of Baghdad, Baghdad, Iraq

Alexey A. Agranovsky

Lomonosov Moscow State University, Moscow, Russia

Nasim Ahmed

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Maher Al Rwahnih

University of California, Davis, CA, United States

Olufemi J. Alabi

*Texas A&M AgriLife Research and Extension Center,
Weslaco, TX, United States*

Aurélien A. Albertini

*Institute for Integrative Biology of the Cell (I2BC),
French Alternative Energies and Atomic Energy
Commission, French National Center for Scientific
Research, Paris-Sud University, University of
Paris-Saclay, Gif-sur-Yvette, France*

Aleksandra Alimova

*The City University of New York (CUNY), School of
Medicine, The City College of New York, New York,
NY, United States*

Juan C. Alonso

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Imran Amin

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Stephanie E. Ander

*University of Colorado School of Medicine, Aurora, CO,
United States*

Danielle E. Anderson

Duke-NUS Medical School, Singapore, Singapore

Ida Bagus Andika

Qingdao Agricultural University, Qingdao, China

Ana C.d.S.P. Andrade

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Juana Angel

Pontifical Javeriana University, Bogota, Colombia

Vanesa Anton-Vazquez

King's College Hospital, London, United Kingdom

Guido Antonelli

Sapienza University of Rome, Rome, Italy

Josefa Antón

University of Alicante, Alicante, Spain

Nanako Aoki

*Tokyo University of Agriculture and Technology, Fuchu,
Japan*

Timothy D. Appleby

King's College Hospital, London, United Kingdom

Miguel Arenas

*Department of Biochemistry, Genetics and Immunology,
University of Vigo, Vigo, Spain and CINBIO
(Biomedical Research Center), University of Vigo, Vigo,
Spain*

Basil Arif

*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Vicente Arnau

*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain*

Gaurav Arya

Duke University, Durham, NC, United States

Leyla Asadi

University of Alberta, Edmonton, AB, Canada

Sassan Asgari

The University of Queensland, Brisbane, QLD, Australia

Nina S. Atanasova

*Finnish Meteorological Institute, Helsinki, Finland and
University of Helsinki, Helsinki, Finland*

Houssam Attoui

*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Silvia Ayora

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Walid Azab

Free University of Berlin, Berlin, Germany

Sasha R. Azar

*The University of Texas Medical Branch, Galveston, TX,
United States*

Fengwei Bai

*The University of Southern Mississippi, Hattiesburg, MS,
United States*

Dalan Bailey

The Pirbright Institute, Pirbright, United Kingdom

S.C. Baker

*Loyola University of Chicago, Maywood, IL,
United States*

Fausto Baldanti

*University of Pavia, Pavia, Italy and Scientific Institute
for Research, Hospitalization and Healthcare, San
Matteo Polyclinic Foundation, Pavia, Italy*

Logan Banadyga

*Public Health Agency of Canada, Winnipeg, MB,
Canada*

Ashley C. Banyard

*Animal and Plant Health Agency, Addlestone,
United Kingdom; University of West Sussex, Falmer,
United Kingdom; and St. George's Medical School,
University of London, London, United Kingdom*

Xiaoyong Bao

*The University of Texas Medical Branch, Galveston, TX,
United States*

Yiming Bao

*Beijing Institute of Genomics, Chinese Academy of
Sciences, University of Chinese Academy of Sciences,
Beijing, China*

Alan D.T. Barrett

*The University of Texas Medical Branch, Galveston, TX,
United States*

Diana P. Baquero

*Archaeal Virology Unit, Institut Pasteur, Paris, France
and Sorbonne University, Paris, France*

Moshe Bar-Joseph

*Agricultural Research Organization, Volcani Center, Bet
Dagan, Israel*

Rachael S. Barr

*Bristol Royal Hospital for Children, Bristol,
United Kingdom*

Ralf Bartenschlager

Heidelberg University, Heidelberg, Germany

David L.V. Bauer

Francis Crick Institute, London, United Kingdom

Oliver W. Bayfield

University of York, York, United Kingdom

Sally A. Baylis

Paul-Ehrlich-Institute, Langen, Germany

Philippa M. Beard

*The Pirbright Institute, Pirbright, United Kingdom and The
Roslin Institute, University of Edinburgh, United Kingdom*

Paul Becher

University of Veterinary Medicine, Hannover, Germany

Björn Becker

Saarland University, Saarbrücken, Germany

Karen L. Beemon

Johns Hopkins University, Baltimore, MD, United States

Martin Beer

Friedrich-Loeffler-Institute, Insel Riems, Germany

Jose Miguel Benito

*Health Research Institute of the Jiménez Díaz
Foundation, Autonomous University of Madrid and Rey
Juan Carlos University Hospital, Móstoles, Spain*

Mária Benkő

*Institute for Veterinary Medical Research, Center for
Agricultural Research, Budapest, Hungary*

Max Bergoin
National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada

Sabrina Bertin
Council for Agricultural Research and Economics,
Research Center for Plant Protection and Certification,
Rome, Italy

Shweta Bhatt
University of Copenhagen, Copenhagen, Denmark

Dennis K. Bideshi
California Baptist University, Riverside, CA,
United States and University of California, Riverside,
CA, United States

Yves Bigot
INRAE – French National Research Institute for
Agriculture, Food and Environment, Nouzilly, France

Richard J. Bingham
University of York, York, United Kingdom

Vera Bischoff
Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany

Kate N. Bishop
Francis Crick Institute, London, United Kingdom

Lindsay W. Black
The University of Maryland School of Medicine,
Baltimore, MD, United States

Romain Blanc-Mathieu
Institute for Chemical Research, Kyoto University, Kyoto,
Japan

Soile Blomqvist
National Institute for Health and Welfare, Helsinki,
Finland

Bryony C. Bonning
University of Florida, Gainesville, FL, United States

Lisa M. Bono
Rutgers, The State University of New Jersey,
New Brunswick, NJ, United States

Alexia Bordigoni
Aix-Marseille University, CNRS, IRD, Mediterranean
Institute of Oceanography, Marseille, France and
Aix-Marseille University, IRD257, Assistance-Publique
des Hôpitaux de Marseille, UMR Microbes, Evolution,
Phylogeny and Infections (MEPHI), IHU Méditerranée
Infection, Marseille, France

Mihnea Bostina
University of Otago, Dunedin, New Zealand

Maxime Boutier
University of Liège, Liège, Belgium

P.R. Bowser
Cornell University, Ithaca, NY, United States

Daniel Bradshaw
Public Health England, London, United Kingdom

Claude Bragard
University of Louvain, Louvain-la-Neuve, Belgium

Aaron C. Brault
Centers for Disease Control and Prevention, Fort Collins,
CO, United States

Nicolas Bravo-Vasquez
St. Jude Children's Research Hospital, Memphis, TN,
United States

Rob W. Briddon
University of Agriculture, Faisalabad, Pakistan

Thomas Briebe
Columbia University, New York, NY, United States

Paul Britton
The Pirbright Institute, Pirbright, United Kingdom

Thomas J. Brouwers
Athena Institute, VU Amsterdam, Amsterdam, The
Netherlands

Kevin E. Brown
Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom

Corina P.D. Brussaard
NIOZ Royal Netherlands Institute for Sea Research,
Den Burg, Texel, The Netherlands and Utrecht
University, Utrecht, The Netherlands

Harald Brüssow
Laboratory of Gene Technology, Department of
Biosystems, KU Leuven, Leuven, Belgium

Joachim J. Bugert
Bundeswehr Institute of Microbiology, Munich, Germany

Jozef J. Bujarski
Northern Illinois University, DeKalb, IL, United States
and Polish Academy of Sciences, Poznan, Poland

Laura Burga
University of Otago, Dunedin, New Zealand

Sara H. Burkhard
University Hospital of Zurich, Zurich, Switzerland

Cara C. Burns
Centers for Disease Control and Prevention, Atlanta,
GA, United States

Felicity Burt
University of the Free State, Bloemfontein, South Africa

Kerry S. Burton
Leamington Spa, United Kingdom

Sarah J. Butcher
University of Helsinki, Helsinki, Finland

Mathias Büttner
Leipzig University, Leipzig, Germany

Jesse Cahill
Sandia National Labs, Albuquerque, NM, United States

Marianna Calabretto
Sapienza University of Rome, Rome, Italy

Thierry Candresse
The National Research Institute for Agriculture, Food and the Environment, University of Bordeaux, Villenave d'Ornon, France

Alan J. Cann
University of Leicester, Leicester, United Kingdom

Lorenzo Capucci
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Irene Carlon-Andres
University of Oxford, Oxford, United Kingdom

José M. Casasnovas
National Center for Biotechnology, Spanish National Research Council (CSIC), Madrid, Spain

J.W. Casey
Cornell University, Ithaca, NY, United States

R.N. Casey
Cornell University, Ithaca, NY, United States

Sherwood R. Casjens
University of Utah, Salt Lake City, UT, United States

Antonella Casola
The University of Texas Medical Branch, Galveston, TX, United States

José R. Castón
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Carlos E. Catalano
University of Colorado Anschutz Medical Campus, Skaggs School of Pharmacy and Pharmaceutical Sciences, Aurora, CO, United States

Roberto Cattaneo
Mayo Clinic, Rochester, MN, United States

Patrizia Cavadini
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Supranee Chaiwatpongsakorn
Nationwide Children's Hospital, Columbus, OH, United States

Supriya Chakraborty
Jawaharlal Nehru University, New Delhi, India

Yu-Chan Chao
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Tyler P. Chavers
Centers for Disease Control and Prevention, Atlanta, GA, United States

Keping Chen
Jiangsu University, Zhenjiang, China

Xiaorui Chen
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Yanping Chen
Bee Research Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

Dayna Cheng
National Cheng Kung University, Tainan, Taiwan

Quentin Chesnais
University of Strasbourg, Colmar, France

Sotaro Chiba
Nagoya University, Nagoya, Japan

Wah Chiu
Stanford University, Stanford, CA, United States

David Chmielewski
Stanford University, Stanford, CA, United States

Irma E. Cisneros
The University of Texas Medical Branch, Galveston, TX, United States

Lark L. Coffey
University of California, Davis, CA, United States

Alanna B. Cohen
Rutgers University, New Brunswick, NJ, United States

Jeffrey I. Cohen
Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Seth Coleman
Rice University, Houston, TX, United States

Miquel Coll
*Institute for Research in Biomedicine, Barcelona, Spain
 and Institute for Molecular Biology of Barcelona,
 Barcelona, Spain*

John Collinge
*UCL Institute of Prion Diseases, London,
 United Kingdom*

Carina Conceicao
The Pirbright Institute, Pirbright, United Kingdom

Gabriela N. Condezo
*National Center for Biotechnology, Spanish National
 Research Council, Madrid, Spain*

Michaela J. Conley
*MRC-University of Glasgow Centre for Virus Research,
 Glasgow, United Kingdom*

Charles A Coomer
University of Oxford, Oxford, United Kingdom

Anne K. Cordes
*Hannover Medical School, Institute of Virology, Hannover,
 Germany*

Mauricio Cortes Jr.
*Department of Chemistry, College of Arts and Sciences,
 Fort Wayne, IN, United States*

Robert H.A. Coutts
University of Hertfordshire, Hatfield, United Kingdom

Jeff A. Cowley
CSIRO Livestock Industries, Brisbane, QLD, Australia

Robert W. Cross
*The University of Texas Medical Branch, Galveston, TX,
 United States*

Henryk Czosnek
The Hebrew University of Jerusalem, Rehovot, Israel

Håkon Dahle
*Department of Biological Sciences, University of Bergen,
 Bergen, Norway*

Janet M. Daly
*University of Nottingham, Sutton Bonington,
 United Kingdom*

Subha Das
Okayama University, Kurashiki, Japan

Indranil Dasgupta
University of Delhi, New Delhi, India

Sibnarayan Datta
*Defence Research Laboratory, Defence Research and
 Development Organisation (DRDO), Tezpur, Assam,
 India*

Amy Davis
*St Jude Children's Research Hospital, Memphis, TN,
 United States*

William O. Dawson
*Citrus Research and Education Center, Lake Alfred, FL,
 United States and University of Florida, Lake Alfred, FL,
 United States*

Erik De Clercq
*Rega Institute for Medical Research, KU Leuven,
 Leuven, Belgium*

Raoul J. de Groot
Utrecht University, Utrecht, The Netherlands

Juan C. de la Torre
*The Scripps Research Institute, La Jolla, CA,
 United States*

Marcelo De las Heras
University of Zaragoza, Zaragoza, Spain

Juliana Gabriela Silva de Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Athos S. de Oliveira
University of Brasília, Brasília, Brazil

Nicole T. de Stefano
University of Florida, Gainesville, FL, United States

Greg Deakin
NIAB-EMR, East Malling, United Kingdom

Philippe Delfosse
University of Luxembourg, Esch-sur-Alzette, Luxembourg

Natacha Delrez
University of Liège, Liège, Belgium

Tatiana A. Demina
*Molecular and Integrative Biosciences Research Program,
 Faculty of Biological and Environmental Sciences,
 University of Helsinki, Helsinki, Finland*

Ismail Demir
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

Zihni Demirbağ
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

X. Deng
*Loyola University of Chicago, Maywood, IL,
 United States*

Cécile Desbiez
*Plant Pathology Unit, INRAE – French National
 Research Institute for Agriculture, Food and
 Environment, Montfavet, France*

Christelle Desnues

Aix-Marseille University, CNRS, IRD, Mediterranean Institute of Oceanography, Marseille, France and Aix-Marseille University, IRD 257, Assistance-Publique des Hôpitaux de Marseille, UMR Microbes, Evolution, Phylogeny and Infections (MEPHI), IHU Méditerranée Infection, Marseille, France

Samantha J. DeWerff

University of Illinois at Urbana-Champaign, Urbana, IL, United States

Daniele Di Carlo

Sapienza University of Rome, Rome, Italy

Arturo Diaz

La Sierra University, Riverside, CA, United States

Alfredo Diaz-Lara

University of California, Davis, CA, United States

Ralf G. Dietzgen

The University of Queensland, St. Lucia, QLD, Australia

Michele Digiario

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Michael Dills

Montana State University, Bozeman, MT, United States

Wayne Dimech

National Serology Reference Laboratory, Fitzroy, VIC, Australia

Savithramma P. Dinesh-Kumar

University of California, Davis, CA, United States

Linda K. Dixon

The Pirbright Institute, Pirbright, United Kingdom

Valerian V. Dolja

Oregon State University, Corvallis, OR, United States

Aušra Domanska

University of Helsinki, Helsinki, Finland

Leslie L. Domier

Agricultural Research Service, US Department of Agriculture, Urbana, IL, United States

Pilar Domingo-Calap

Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, Valencia, Spain

Tatiana Domitrovic

Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Sarah M. Doore

Michigan State University, East Lansing, MI, United States

Lucy Dorrell

University of Oxford, Oxford, United Kingdom

Rosemary A. Dorrington

Rhodes University, Grahamstown, South Africa

Andor Doszpoly

Hungarian Academy of Sciences, Budapest, Hungary

Andreas Dotzauer

University of Bremen, Bremen, Germany

Simon B. Drysdale

St George's University Hospitals NHS Foundation Trust, London, United Kingdom and St George's, University of London, London, United Kingdom

Robert L. Duda

University of Pittsburgh, Pittsburgh, PA, United States

Carol Duffy

University of Alabama, Tuscaloosa, AL, United States

Siobain Duffy

Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

David D. Dunigan

University of Nebraska–Lincoln, Lincoln, NE, United States

Stéphane Duquerroy

University of Paris-Saclay, Orsay, France and Institut Pasteur, Paris, France

Bas E. Dutilh

Utrecht University, Utrecht, The Netherlands and Radboud University Medical Center, Nijmegen, The Netherlands

Michael Edelstein

Faculty of Medicine, Bar Ilan University, Ramat Gan, Israel

Herman K. Edskes

National Institutes of Health, Bethesda, MD, United States

Rosina Ehmann

Bundeswehr Institute of Microbiology, Munich, Germany

Toufic Elbeaino

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Joanne B. Emerson

University of California, Davis, CA, United States

Ann Emery

University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Christine E. Engeland
University Hospital Heidelberg and German Cancer Research Center, Heidelberg, Germany and Witten/Herdecke University, Witten, Germany

Luis Enjuanes
National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Katri Eskelin
University of Helsinki, Helsinki, Finland

Rosa Esteban
Institute of Biology and Functional Genomics, CSIC/University of Salamanca, Salamanca, Spain

Mary K. Estes
Baylor College of Medicine, Houston, TX, United States

Cassia F. Estofolete
São José do Rio Preto School of Medicine, São José do Rio Preto, Brazil

Alyssa B. Evans
National Institutes of Health, Hamilton, MT, United States

Øystein Evensen
Norwegian University of Life Sciences, Oslo, Norway

Alex Evilevitch
Department of Experimental Medical Science, Lund University, Lund, Sweden

Montserrat Fàbrega-Ferrer
Institute for Research in Biomedicine, Barcelona, Spain and Institute for Molecular Biology of Barcelona, Barcelona, Spain

Francesco Faggioli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Bentley A. Fane
University of Arizona, Tucson, AZ, United States

Brian A. Federici
University of California, Riverside, CA, United States

F. Fenner
Australian National University, Canberra, ACT, Australia

Isabel Fernández de Castro
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Giovanni Ferrara
University of Alberta, Edmonton, AB, Canada

Elvira Fiallo-Olivé
Institute for Mediterranean and Subtropical Horticulture “La Mayora”–Spanish National Research Council–University of Malaga, Algarrobo-Costa, Málaga, Spain

Andrew E. Firth
University of Cambridge, Cambridge, United Kingdom

Roland A. Fleck
King’s College London, London, United Kingdom

Ricardo Flores
Polytechnic University of Valencia, Higher Council of Scientific Research, Valencia, Spain

Ervin Fodor
University of Oxford, Oxford, United Kingdom

Anthony R. Fooks
Animal and Plant Health Agency, Addlestone, United Kingdom; University of Liverpool, Liverpool, United Kingdom; and St. George’s Medical School, University of London, London, United Kingdom

Patrick Forterre
Archeal Virology Unit, Institut Pasteur, Paris, France and French National Center for Scientific Research, Institute of Integrative Biology of the Cell, University of Paris-Saclay, Gif sur Yvette, France

Rennos Fragkoudis
University of Nottingham, Sutton Bonington, United Kingdom and University of Edinburgh, Edinburgh, United Kingdom

Manuel A. Franco
Pontifical Javeriana University, Bogota, Colombia

Giovanni Franzo
Department of Animal Medicine, Production and Health (MAPS), Padua University, Padua, Italy

Graham L. Freimanis
The Pirbright Institute, Pirbright, United Kingdom

Juliana Freitas-Astúa
Brazilian Agricultural Research Corporation (Embrapa) Cassava and Fruits, Cruz das Almas, Brazil

Elizabeth E. Fry
Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Marc Fuchs
Cornell University, Geneva, NY, United States

Tsutomu Fujimura
Institute of Biology and Functional Genomics, CSIC/University of Salamanca, Salamanca, Spain

- Kuko Fuke
Tokyo University of Agriculture and Technology, Fuchu, Japan
- Toshiyuki Fukuhara
Tokyo University of Agriculture and Technology, Fuchu, Japan
- To S. Fung
South China Agricultural University, Guangzhou, China
- Yahya Z.A. Gaafar
Julius Kuehn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany
- Toni Gabaldon
Barcelona Supercomputing Center-National Center for Supercomputing, Institute of Research in Biomedicine, and Catalan Institution for Research and Advanced Studies, Barcelona, Spain
- Morgan Gaïa
University of Paris-Saclay, Evry, France
- José Gallardo
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain
- Hernan Garcia-Ruiz
University of Nebraska–Lincoln, Lincoln, NE, United States
- Juan A. García
National Center for Biotechnology-Spanish National Research Council, Madrid, Spain
- Matteo P. Garofalo
The University of Texas Medical Branch, Galveston, TX, United States
- Yves Gaudin
Institute for Integrative Biology of the Cell (I2BC), French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, Paris-Sud University, University of Paris-Saclay, Gif-sur-Yvette, France
- Andrew D.W. Geering
The University of Queensland, St. Lucia, QLD, Australia
- Thomas W. Geisbert
The University of Texas Medical Branch, Galveston, TX, United States
- Andrea Gentili
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy
- Volker Gerdts
University of Saskatchewan, Saskatoon, SK, Canada
- Said A. Ghabrial[†]
Department of Plant Pathology, University of Kentucky, Lexington, KY, United States
- Clément Gilbert
Evolution, Genomes, Behavior and Ecology Laboratory, CNRS University of Paris-Sud UMR 9191, IRD UMR 247, Gif-sur-Yvette, France
- Robert L. Gilbertson
University of California, Davis, CA, United States
- Efstathios S. Giotis
Imperial College London, London, United Kingdom and University of Essex, Colchester, United Kingdom
- Laurent Glais
French Federation of Seed Potato Growers/Research, Development, Promotion of Seed Potato, Paris, France and Institute for Genetics, Environment and Plant Protection, Agrocampus West, French National Institute for Agriculture, Food and Environment, University of Rennes 1, Le Rheu, France
- Miroslav Glasa
Biomedical Research Center, Slovak Academy of Sciences, Bratislava, Slovakia
- Ido Golding
University of Illinois at Urbana-Champaign, Urbana, IL, United States
- Esperanza Gomez-Lucia
Complutense University of Madrid, Madrid, Spain
- Zheng Gong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
- Andrea González-González
University of Florida, Gainesville, FL, United States
- Michael M. Goodin
University of Kentucky, Lexington, KY, United States
- Alexander E. Gorbalenya
Leiden University Medical Center, Leiden, The Netherlands
- Paul Gottlieb
The City University of New York (CUNY), School of Medicine, The City College of New York, New York, NY, United States
- M.-A. Grandbastien
INRAE – French National Research Institute for Agriculture, Food and Environment, Versailles, France

[†]Deceased.

Meritxell Granell
*National Center for Biotechnology, Madrid, Spain and
 Institute of Chemical Research of Catalonia (ICIQ),
 Tarragona, Spain*

Patrick L. Green
*The Ohio State University, Columbus, OH,
 United States*

Sandra J. Greive
University of York, York, United Kingdom

Diane E. Griffin
*Johns Hopkins Bloomberg School of Public Health,
 Baltimore, MD, United States*

Jonathan M. Grimes
University of Oxford, Oxford, United Kingdom

Nigel Grimsley
*Integrative Biology of Marine Organisms Laboratory,
 Banuyls-sur-Mer, France and Sorbonne University,
 Banuyls-sur-Mer, France*

Bruno Gronenborn
*Institute for Integrative Biology of the Cell, CNRS,
 University of Paris-Sud, CEA, Gif sur Yvette, France*

Julianne H. Grose
Brigham Young University, Provo, UT, United States

Scott Grytdal
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

Duane J. Gubler
Duke-NUS Medical School, Singapore, Singapore

Peixuan Guo
*College of Pharmacy, The Ohio State University,
 Columbus, OH, United States*

Tongkun Guo
*Beijing Institute of Genomics, Chinese Academy of
 Sciences, Beijing, China*

Anne-Lise Haenni
*Institut Jacques Monod, French National Center for
 Scientific Research, Paris Diderot University, Paris,
 France*

Susan L. Hafenstein
*Pennsylvania State University, Hershey, PA,
 United States*

Ahmed Hafez
*Biotechvana, Valencia, Spain; Pompeu Fabra University,
 Barcelona, Spain; and Minia University, Minya, Egypt*

Marie Hagbom
Linköping University, Linköping, Sweden

Sébastien Halary
*National Museum of Natural History, UMR 7245
 CNRS/MNHN Molécule de Communication et
 Adaptation des Micro-organismes, Paris, France*

Aron J. Hall
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

John Hammond
*Floral and Nursery Plants Research, Agricultural
 Research Service, US Department of Agriculture,
 Beltsville, MD, United States*

Rosemarie W. Hammond
*Agricultural Research Service, US Department of
 Agriculture, Beltsville, MD, United States*

Virginia Hargest
*St Jude Children's Research Hospital, Memphis, TN,
 United States and University of Tennessee Health
 Science Center, Memphis, TN, United States*

Scott J. Harper
Washington State University, Prosser, WA, United States

Balázs Harrach
*Institute for Veterinary Medical Research, Center for
 Agricultural Research, Budapest, Hungary*

Masayoshi Hashimoto
The University of Tokyo, Tokyo, Japan

Muhammad Hassan
University of Agriculture, Faisalabad, Pakistan

Asma Hatoum-Aslan
University of Alabama, Tuscaloosa, AL, United States

Philippa C. Hawes
The Pirbright Institute, Pirbright, United Kingdom

Janelle A. Hayes
*University of Massachusetts Medical School, Worcester,
 MA, United States*

Guijuan He
Virginia Tech, Blacksburg, VA, United States

Klaus Hedman
*University of Helsinki, Helsinki, Finland and Helsinki
 University Hospital, Helsinki, Finland*

Albert Heim
Hannover Medical School, Hanover, Germany

Gary L. Hein
*University of Nebraska–Lincoln, Lincoln, NE,
 United States*

Manfred Heinlein
*IBMP-CNRS, University of Strasbourg, Strasbourg,
 France*

Mercedes Hernando-Pérez
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Carmen Hernández
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Etienne Herrbach
University of Strasbourg, Colmar, France

Stephen Higgs
*Biosecurity Research Institute, Kansas State University,
Manhattan, KS, United States*

Bradley I. Hillman
Rutgers University, New Brunswick, NJ, United States

Deborah M. Hinton
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Judith Hirsch
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Jody Hobson-Peters
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Natalie M. Holste
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Jin S. Hong
Kangwon National University, Chunchon, South Korea

Margaret J. Hosie
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Olivia G. Howell
University of Alabama, Tuscaloosa, AL, United States

Liya Hu
Baylor College of Medicine, Houston, TX, United States

Zhaoyang Hu
Jiangsu University, Zhenjiang, China

Kuan-Ying A. Huang
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Yu Huang
Peking University, Beijing, China

Natalia B. Hubbs
Hanover College, Hanover, IN, United States

Elisabeth Huguet
*Research Institute on Insect Biology, French National
Center for Scientific Research, University of Tours,
Tours, France*

Roger Hull
John Innes Centre, Norwich, United Kingdom

Kiwamu Hyodo
Okayama University, Kurashiki, Japan

Eugénie Hébrard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

Martin Hölzer
University of Jena, Jena, Germany

Tetsuro Ikegami
*The University of Texas Medical Branch at Galveston,
Galveston, TX, United States*

Niina Ikonen
*Finnish Institute for Health and Welfare, Helsinki,
Finland*

Cihan İnan
*Department of Molecular Biology and Genetics,
Karadeniz Technical University, Trabzon, Turkey*

İkbal Agah İnce
*Department of Medical Microbiology, Acibadem
University School of Medicine, Istanbul, Turkey*

Katsuaki Inoue
Diamond Light Source, Didcot, United Kingdom

Toru Iwanami
Tokyo University of Agriculture, Tokyo, Japan

Jacques Izopet
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Fauziah Mohd Jaafar
*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Andrew O. Jackson
China Agricultural University, Beijing, China

Daral J. Jackwood
*The Ohio State University/OARDC, Wooster, OH,
United States*

Jean-Rock Jacques
*Cellular and Molecular Epigenetics (GIGA), Liège,
 Belgium and Molecular Biology (TERRA), Gembloux,
 Belgium*

Tiffany Jenkins
*Nationwide Children's Hospital, Columbus, OH,
 United States and The Ohio State University, Columbus,
 OH, United States*

Jeffrey D. Jensen
Arizona State University, Tempe, AZ, United States

Daohong Jiang
Huazhong Agricultural University, Wuhan, China

Zhihao Jiang
China Agricultural University, Beijing, China

Allison R. Jilbert
The University of Adelaide, Adelaide, SA, Australia

Peng Jing
*Department of Chemistry, College of Arts and Sciences,
 Fort Wayne, IN, United States*

Xixi Jing
Central South University, Changsha, China

Meesbah Jiwaji
Rhodes University, Grahamstown, South Africa

Kyle L. Johnson
*The University of Texas at El Paso, El Paso, TX,
 United States and CQuentia, Memphis, TN, United
 States*

Welkin E. Johnson
Boston College, Chestnut Hill, MA, United States

Ian M. Jones
*University of Reading, Reading, United Kingdom and
 London School of Hygiene and Tropical Medicine,
 London, United Kingdom*

Ramon Jordan
*Agricultural Research Service, US Department of
 Agriculture, Beltsville, MD, United States*

Thomas Joris
*Cellular and Molecular Epigenetics (GIGA), Liège,
 Belgium and Molecular Biology (TERRA), Gembloux,
 Belgium*

Ilkka Julkunen
*Institute of Biomedicine, University of Turku, Turku,
 Finland*

Sandra Junglen
Charité - University Medicine Berlin, Berlin, Germany

Masanori Kaido
Kyoto University, Kyoto, Japan

Laura Kakkola
University of Turku, Turku, Finland

Hannimari Kallio-Kokko
*University of Helsinki and Helsinki University Hospital,
 Helsinki, Finland*

Nassim Kamar
*Toulouse University Hospital, Toulouse, France and
 Toulouse University Paul Sabatier, Toulouse, France*

Phyllis J. Kanki
*Harvard T.H. Chan School of Public Health, Boston,
 MA, United States*

Peter Karayiannis
University of Nicosia, Nicosia, Cyprus

Henry M. Kariithi
*Kenya Agricultural and Livestock Research Organization,
 Nairobi, Kenya*

Brian A. Kelch
*University of Massachusetts Medical School, Worcester,
 MA, United States*

Karen E. Keller
*Horticultural Crops Research Unit, Agricultural Research
 Service, US Department of Agriculture, Corvallis, OR,
 United States*

Japhette E. Kembou-Ringert
University of Tel Aviv, Tel Aviv, Israel

Peter J. Kerr
*University of Sydney, Sydney, NSW, Australia and
 CSIRO Health and Biosecurity, Black Mountain
 Laboratories, Canberra, ACT, Australia*

Tiffany King
*Nationwide Children's Hospital, Columbus, OH,
 United States and The Ohio State University College
 of Medicine, Columbus, OH, United States*

Andrea Kirmaier
Boston College, Chestnut Hill, MA, United States

Thomas Klose
Purdue University, West Lafayette, IN, United States

Barbara G. Klupp
*Friedrich-Loeffler-Institute, Greifswald-Insel Riems,
 Germany*

David M. Knipe
Harvard Medical School, Boston, MA, United States

Nick J. Knowles
The Pirbright Institute, Pirbright, United Kingdom

Guus Koch
*Wageningen Bioveterinary Research, Lelystad,
 The Netherlands*

Renate Koenig
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

Susanne E. Kohalmi
*The University of Western Ontario, London, ON,
Canada*

Hideki Kondo
Okayama University, Kurashiki, Japan

Jennifer L. Konopka-Anstadt
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Eugene V. Koonin
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States and
National Institutes of Health, Bethesda, MD, United States*

Marion P.G. Koopmans
Erasmus Medical Center, Rotterdam, The Netherlands

Richard Kormelink
*Wageningen University and Research, Wageningen,
The Netherlands*

Ioly Kotta-Loizou
Imperial College London, London, United Kingdom

Peter J. Krell
*Department of Molecular and Cellular Biology,
University of Guelph, Guelph, ON, Canada*

Mart Krupovic
Archaeal Virology Unit, Institut Pasteur, Paris, France

Andreas Kuhn
University of Hohenheim, Stuttgart, Germany

Jens H. Kuhn
*National Institutes of Health, Frederick, MD,
United States*

Richard J. Kuhn
Purdue University, West Lafayette, IN, United States

Suvi Kuivanen
University of Helsinki, Helsinki, Finland

Ranjababu Kulasegaram
*Guy's and St Thomas' NHS Foundation Trust, London,
United Kingdom*

Raghavendran Kulasegaran-Shylini
*Department of Pathogen Infection, Faculty of Infectious
and Tropical Diseases, London School of Hygiene and
Tropical Medicine, London, United Kingdom*

Gaurav Kumar
University of Delhi, New Delhi, India

Manish Kumar
Jawaharlal Nehru University, New Delhi, India

Gael Kurath
*US Geological Survey, Western Fisheries Research
Center, Seattle, WA, United States*

Satu Kurkela
*University of Helsinki and Helsinki University Hospital,
Helsinki, Finland*

Wan-Chun Lai
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Kevin Lamkiewicz
University of Jena, Jena, Germany

Rebecca K. Lane
*University of Texas Health Science Center at San
Antonio, San Antonio, TX, United States*

Andrew S. Lang
*Memorial University of Newfoundland, St. John's, NL,
Canada*

Daniel Carlos Ferreira Lanza
Federal University of Rio Grande do Norte, Natal, Brazil

Maija Lappalainen
*HUS Diagnostic Center, HUSLAB, Clinical
Microbiology, University of Helsinki and Helsinki
University Hospital, Helsinki, Finland*

Katherine LaTourrette
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Chris Lauber
*TWINCORE – Center for Experimental and Clinical
Infection Research, Hannover, Germany*

Antonio Lavazza
*The Lombardy and Emilia Romagna Experimental
Zootechnic Institute, Brescia, Italy*

C. Martin Lawrence
Montana State University, Bozeman, MT, United States

Hervé Lecoq
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Young-Min Lee
Utah State University, Logan, UT, United States

Kristen N. LeGault
University of California, Berkeley, CA, United States

James Legg
*International Institute of Tropical Agriculture, Dar es
Salaam, Tanzania*

Anne Legreve
University of Louvain, Louvain-la-Neuve, Belgium

Petr G. Leiman
The University of Texas Medical Branch, Galveston, TX, United States

Stanley M. Lemon
Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Sebastian Leptihn
Zhejiang University-Edinburgh University Institute, Zhejiang University, Haining, China

Dennis J. Lewandowski
University of Florida, Lake Alfred, FL, United States

Sébastien Lhomme
Toulouse University Hospital, Toulouse, France and Toulouse University Paul Sabatier, Toulouse, France

Dawei Li
China Agricultural University, Beijing, China

Guangdi Li
Central South University, Changsha, China

Guoqing Li
Huazhong Agricultural University, Wuhan, China

Yi Li
Peking University, Beijing, China

Zhefeng Li
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Zhenghe Li
Zhejiang University, Hangzhou, China

Jia Q. Liang
South China Agricultural University, Guangzhou, China

Sebastian Liebe
Institute of Sugar Beet Research, Göttingen, Germany

João Paulo Matos Santos Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Bruno Lina
HCL Department of Virology, National Reference Center for Respiratory Viruses, Institute of Infectious Agents, Croix-Rousse Hospital, Lyon, France and Virpath Laboratory, International Center of Research in Infectiology (CIRI), INSERM U1111, CNRS—UMR 5308, École Normale Supérieure de Lyon, University Claude Bernard Lyon, Lyon University, Lyon, France

Walter Ian Lipkin
Columbia University, New York, NY, United States

Jan G. Lisby
Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

Ding X. Liu
South China Agricultural University, Guangzhou, China

Qiang Liu
University of Saskatchewan, Saskatoon, SK, Canada

Sijun Liu
Iowa State University, Ames, IA, United States

Carlos Llorens
Biotechvana, Scientific Park University of Valencia, Valencia, Spain

L. Sue Loesch-Fries
Purdue University, West Lafayette, IN, United States

George P. Lomonosoff
John Innes Centre, Norwich, United Kingdom

L. Letti Lopez
The University of Texas at Austin, Austin, TX, United States

Alan T. Loynachan
University of Kentucky, Lexington, KY, United States

Garry A. Luke
University of St. Andrews, St. Andrews, United Kingdom

M. Luo
University of Alabama at Birmingham, Birmingham, AL, United States

Juan J. López-Moya
Center for Research in Agricultural Genomics and Spanish National Research Council, Barcelona, Spain

Che Ma
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Stuart A. MacFarlane
The James Hutton Institute, Invergowrie, United Kingdom

Saichetana Macherla
J. Craig Venter Institute, La Jolla, CA, United States

Kensaku Maejima
The University of Tokyo, Tokyo, Japan

Fabrizio Maggi
University of Pisa, Pisa, Italy and University of Insubria, Varese, Italy

Melissa S. Maginnis
The University of Maine, Orono, ME, United States

Edgar Maiss
Leibniz University Hannover, Hannover, Germany

Kira S. Makarova
National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, United States

Ariana Manglli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Annette Mankertz
Robert Koch-Institute, Berlin, Germany

Pilar Manrique
The Ohio State University, Wexner Medical Center, Columbus, OH, United States

Shahid Mansoor
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Marco Marklewitz
Institute of Virology, Charité – University Medicine Berlin, Berlin, Germany

Giovanni P. Martelli[†]
University of Bari Aldo Moro, Bari, Italy

Darren P. Martin
University of Cape Town, Cape Town, South Africa

Robert R. Martin
Horticultural Crops Research Unit, Agricultural Research Service, US Department of Agriculture, Corvallis, OR, United States

Manuel Martinez-Garcia
University of Alicante, Alicante, Spain

Francisco Martinez-Hernandez
University of Alicante, Alicante, Spain

Natalia Martín-González
Autonomous University of Madrid, Madrid, Spain

Joaquín Martínez Martínez
Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Manja Marz
University of Jena, Jena, Germany

Andrea Marzi
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, United States

Hema Masarapu
Sri Venkateswara University, Tirupati, India

Chikara Masuta
Hokkaido University, Sapporo, Japan

Carlos P. Mata
University of Leeds, Leeds, United Kingdom

Jelle Matthijnsens
Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

Claire P. Mattison
Centers for Disease Control and Prevention, Atlanta, GA, United States and Cherokee Nation Assurance, Arlington, VA, United States

William McAllister
Rowan University School of Osteopathic Medicine, Stratford, NJ, United States

Alison A. McBride
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Michael McChesney
University of California, Davis, CA, United States

Elaine McCulloch
Quality Control for Molecular Diagnostics (QCMD), Glasgow, United Kingdom

Andrew J. McMichael
University of Oxford, Oxford, United Kingdom

Alexander McPherson
University of California, Irvine, CA, United States

Irene K. Meki
French National Center for Scientific Research, Montpellier, France

Ulrich Melcher
Oklahoma State University, Stillwater, OK, United States

Tomas A. Melgarejo
University of California, Davis, CA, United States

Michael J. Melzer
Department of Plant and Environmental Protection Sciences, University of Hawaii, Honolulu, HI, United States

Luiza Mendonça
University of Oxford, Oxford, United Kingdom

Xiang-Jin Meng
Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Peter P.C. Mertens
University of Nottingham, Sutton Bonington, United Kingdom

[†]Deceased.

Thomas C. Mettenleiter
Friedrich-Loeffler-Institute, Greifswald-Insel Riems, Germany

Philip D. Minor
St Albans, United Kingdom

Ali Mirazimi
National Veterinary Institute, Uppsala, Sweden and Karolinska Hospital University, Huddinge, Sweden

Nischay Mishra
Columbia University, New York, NY, United States

Edward S. Mocarski
Emory University School of Medicine, Atlanta, GA, United States

Florian Mock
University of Jena, Jena, Germany

Volker Moennig
University of Veterinary Medicine, Hannover, Germany

Ian J. Molineux
The University of Texas at Austin, Austin, TX, United States

Aderito L. Monjane
Norwegian Veterinary Institute, Oslo, Finland

Jacen S. Moore
University of Tennessee Health Science Center, Memphis, TN, United States

Marc C. Morais
The University of Texas Medical Branch, Galveston, TX, United States

Cristina Moraru
Institute for Chemistry and Biology of the Marine Environment, Oldenburg, Germany

Hiromitsu Moriyama
Tokyo University of Agriculture and Technology, Tokyo, Japan

Sergey Y. Morozov
Lomonosov Moscow State University, Moscow, Russia

Thomas E. Morrison
University of Colorado School of Medicine, Aurora, CO, United States

Léa Morvan
University of Liège, Liège, Belgium

Benoît Moury
Plant Pathology Unit, INRAE – French National Institute for Agriculture, Food and Environment, Montfavet, France

Muhammad Mubin
University of Agriculture, Faisalabad, Pakistan

Nicolas J. Mueller
University Hospital of Zurich, Zurich, Switzerland

Emmanuelle Muller
The French Agricultural Research Center for International Development, Joint Research Units–Biology and Genetics of Plant-Pathogen Interactions, Montpellier, France and Biology and Genetics of Plant-Pathogen Interactions, University of Montpellier, The French Agricultural Research Center for International Development, French National Institute for Agricultural Research, Montpellier SupAgro, Montpellier, France

John S. Munday
Massey University, Palmerston North, New Zealand

Jacob H. Munson-McGee
Montana State University, Bozeman, MT, United States and Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Hacer Muratoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Kenan C. Murphy
University of Massachusetts Medical School, Worcester, MA, United States

Ugrappa Nagalakshmi
University of California, Davis, CA, United States

Keizo Nagasaki
Kochi University, Nankoku, Japan

Nazia Nahid
GC University, Faisalabad, Pakistan and University of Agriculture, Faisalabad, Pakistan

Venugopal Nair
The Pirbright Institute, Pirbright, United Kingdom

Remziye Nalçacıoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Shigetou Namba
The University of Tokyo, Tokyo, Japan

Rubab Z. Naqvi
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Rachel Nash
The Pirbright Institute, Surrey, United Kingdom

C.K. Navaratnarajah
Purdue University, West Lafayette, IN, United States

Maria A. Navarrete-Muñoz
Biotechvana, Madrid, Spain; Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid; and Rey Juan Carlos University Hospital, Móstoles, Spain

Jesús Navas-Castillo
Institute for Mediterranean and Subtropical Horticulture "La Mayora"—Spanish National Research Council—University of Malaga, Algarrobo-Costa, Málaga, Spain

Muhammad S. Nawaz-ul-Rehman
University of Agriculture, Faisalabad, Pakistan

Christopher L. Netherton
The Pirbright Institute, Pirbright, United Kingdom

Thu V.P. Nguyen
Baylor College of Medicine, Houston, TX, United States

Annette Niehl
Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

Hubert G.M. Niesters
Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Jozef I. Nissimov
University of Waterloo, Waterloo, ON, Canada

Norman Noah
London School of Hygiene and Tropical Medicine, London, United Kingdom

Mauricio L. Nogueira
São José do Rio Preto School of Medicine, São José do Rio Preto, São Paulo, Brazil

Johan Nordgren
Linköping University, Linköping, Sweden

C. Micha Nübling
Paul-Ehrlich-Institute, Langen, Germany

Visa Nurmi
University of Helsinki, Helsinki, Finland

Donald L. Nuss
University of Maryland, Rockville, MD, United States

M. Steven Oberste
Centers for Disease Control and Prevention, Atlanta, GA, United States

Hiroyuki Ogata
Institute for Chemical Research, Kyoto University, Kyoto, Japan

Ane Ogbe
University of Oxford, Oxford, United Kingdom

Hanna M. Oksanen
Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Graziele P. Oliveira
Federal University of Minas Gerais, Belo Horizonte, Brazil

Francesco Origi
University of Bern, Bern, Switzerland

Nikolaus Osterrieder
Free University of Berlin, Berlin, Germany

Robert A. Owens
Beltsville Agricultural Research Center, Beltsville, MD 20705, United States

Emine Özsahin
University of Guelph, Guelph, ON, Canada

Sergi Padilla-Parra
University of Oxford, Oxford, United Kingdom; Department of Infectious Diseases, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom; and Randall Division of Cell and Molecular Biophysics, King's College London, London, United Kingdom

Joshua Pajak
Duke University, Durham, NC, United States

Massimo Palmarini
MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Amanda R. Panfil
The Ohio State University, Columbus, OH, United States

Marcus Panning
Institute of Virology, Freiburg University Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Vitantonio Pantaleo
National Research Council, Research Unit of Bari, Bari, Italy

Anna Papa
Aristotle University of Thessaloniki, Thessaloniki, Greece

Nikolaos Pappas
Utrecht University, Utrecht, The Netherlands

Hanu R. Pappu
Washington State University, Pullman, WA, United States

Kristin N. Parent
*Michigan State University, East Lansing, MI,
United States*

Colin R. Parrish
Cornell University, Ithaca, NY, United States

A. Lorena Passarelli
Kansas State University, Manhattan, KS, United States

Basavaprabhu L. Patil
*ICAR–Indian Institute of Horticultural Research,
Bengaluru, India*

Jade Pattyn
University of Antwerp, Antwerp, Belgium

T.A. Paul
Cornell University, Ithaca, NY, United States

Lillian Pavlik
*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Susan L. Payne
Texas A&M University, College Station, TX, United States

Michael N. Pearson
The University of Auckland, Auckland, New Zealand

Mark E. Peeples
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University College of
Medicine, Columbus, OH, United States*

Ben Peeters
*Wageningen Bioveterinary Research, Lelystad,
The Netherlands*

Joseph S.M. Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Malik Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Judit J. Péntzes
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Miryam Pérez-Cañamás
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Marta Pérez-Illana
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Jaume Pérez-Sánchez
*Institute of Aquaculture Torre de la Sal, Spanish
National Research Council, Castellon, Spain*

Jean-Marie Peron
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Karin E. Peterson
National Institutes of Health, Hamilton, MT, United States

Karel Petrzik
*Biology Center CAS, Institute of Plant Molecular
Biology, České Budějovice, Czech Republic*

Mahtab Peyambari
*Pennsylvania State University, State College, PA,
United States*

Sujal Phadke
J. Craig Venter Institute, La Jolla, CA, United States

Hanh T. Pham
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Mauro Pistello
University of Pisa, Pisa, Italy

Daniel Ponndorf
John Innes Centre, Norwich, United Kingdom

Leo L.M. Poon
The University of Hong Kong, Pok Fu Lam, Hong Kong

Welkin H. Pope
University of Pittsburgh, Pittsburgh, PA, United States

Minna M. Poranen
University of Helsinki, Helsinki, Finland

Claudine Porta
*The Pirbright Institute, Pirbright, United Kingdom and
University of Oxford, Oxford, United Kingdom*

Samuel S. Porter
*National Institute of Allergy and Infectious Diseases,
National Institutes of Health, Bethesda, MD,
United States and University of Maryland, College Park,
MD, United States*

Frank A. Post
*King's College Hospital NHS Foundation Trust, London,
United Kingdom*

Nils Poulicard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

David Prangishvili
*Institut Pasteur, Paris, France and Ivane Javakhishvili
Tbilisi State University, Tbilisi, Georgia*

B. V. Venkataram Prasad
Baylor College of Medicine, Houston, TX, United States

Lalita Priyamvada
Centers for Disease Control and Prevention, Atlanta, GA, United States

Simone Prospero
Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

Elisabeth Puchhammer-Stöckl
Medical University of Vienna, Vienna, Austria

Jianming Qiu
University of Kansas Medical Center, Kansas City, KS, United States

S.L. Quackenbush
Colorado State University, Fort Collins, CO, United States

Killian J. Quinn
King's College Hospital, London, United Kingdom

Diego F. Quito-Avila
Department of Life Sciences, ESPOL Polytechnic University, Guayaquil, Ecuador

Frank Rabenstein
Julius Kühn Institute, Quedlinburg, Germany

Sheli R. Radoshitzky
United States Army Medical Research Institute of Infectious Diseases, Frederick, MD, United States

Saleem U. Rahman
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Mbolarinosy Rakotomalala
FOFIFA, Antananarivo, Madagascar

Norma Rallon
Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid and Rey Juan Carlos University Hospital, Móstoles, Spain

Robert P. Rambo
Diamond Light Source, Didcot, United Kingdom

Bertha Cecilia Ramirez
The Institute for Integrative Biology of the Cell, The French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

María D. Ramos-Barbero
University of Alicante, Alicante, Spain

Chris M. Rands
University of Geneva Medical School and Swiss Institute of Bioinformatics, Geneva, Switzerland

Venigalla B. Rao
The Catholic University of America, Washington, DC, United States

Janne J. Ravantti
University of Helsinki, Helsinki, Finland

Mandy Ravensbergen
Wageningen University and Research, Wageningen, The Netherlands

Georget Y. Reaiche-Miller
The University of Adelaide, Adelaide, SA, Australia

D.V.R. Reddy
International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

Vishwanatha R.A.P. Reddy
The Pirbright Institute, Pirbright, United Kingdom

Juan Reguera
Aix-Marseille University, French National Center for Scientific Research, Marseille, France and French National Institute of Health and Medical Research, Marseille, France

William K. Reisen
University of California, Davis, CA, United States

Jingshan Ren
University of Oxford, Oxford, United Kingdom

Renato O. Resende
University of Brasilia, Brasilia, Brazil

Peter A. Revill
The Peter Doherty Institute of Infection and Immunity, Royal Melbourne Hospital, Melbourne, VIC, Australia

Félix A. Rey
Institut Pasteur, Paris, France

Lara Rheinemann
University of Utah, Salt Lake City, UT, United States

Simone G. Ribeiro
Embrapa Genetic Resources and Biotechnology, Brasília, Brazil

Daniel Rigling
Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Cristina Risco
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Efraín E. Rivera-Serrano

Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Cécile Robin

INRAE – French National Research Institute for Agriculture, Food and Environment, UMR BIOGECO, Cestas, France

Rodrigo A.L. Rodrigues

Federal University of Minas Gerais, Belo Horizonte, Brazil

Elina Roine

University of Helsinki, Helsinki, Finland

Maria R. Rojas

University of California, Davis, CA, United States

Marilyn J. Roossinck

Pennsylvania State University, State College, PA, United States

Vera I.D. Ros

Wageningen University and Research, Wageningen, The Netherlands

Cristina Rosa

Pennsylvania State University, University Park, PA, United States

Hanna Rose

Leibniz University Hannover, Hannover, Germany

David A. Rosenbaum

University of Florida, Gainesville, FL, United States

Shannan L. Rossi

The University of Texas Medical Branch, Galveston, TX, United States

Michael G. Rossmann[†]

Purdue University, West Lafayette, IN, United States

L. Roux

University of Geneva Medical School, Geneva, Switzerland

Simon Roux

US Department of Energy Joint Genome Institute, Walnut Creek, CA, United States

J. Rovnak

Colorado State University, Fort Collins, CO, United States

David J. Rowlands

University of Leeds, Leeds, United Kingdom

Polly Roy

Department of Pathogen Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom and University of Reading, Reading, United Kingdom

Aaron P. Roznowski

The University of Texas at Austin, Austin, TX, United States and University of Arizona, Tucson, AZ, United States

Luisa Rubino

Institute for Sustainable Plant Protection, National Research Council, Bari, Italy

Olli Ruuskanen

Turku University Hospital, Turku, Finland

Eugene V. Ryabov

USDA, Agricultural Research Service, Beltsville, MD, United States

Martin D. Ryan

University of St. Andrews, St. Andrews, United Kingdom

Ki H. Ryu

Seoul Women's University, Seoul, South Korea

Hanns-Joachim Rziha

Eberhard Karls University of Tübingen, Tübingen, Germany

Sead Sabanadzovic

Mississippi State University, Starkville, MS, United States

Roghayeh Safari

Cellular and Molecular Epigenetics (GIGA), Liège, Belgium and Molecular Biology (TERRA), Gembloux, Belgium

Azeez Sait Sahul Hameed

C. Abdul Hakeem College, Melvisharam, India

Nicole Samies

University of Alabama at Birmingham, Birmingham, AL, United States

Carmen San Martín

National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Ruth-Anne Sandaa

Department of Biological Sciences, University of Bergen, Bergen, Norway

Hélène Sanfaçon

Agriculture and Agri-Food Canada, Summerland, BC, Canada

[†]Deceased.

Rafael Sanjuán
*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia-CSIC, Valencia, Spain*

Neeraja Sankaran
Utrecht University, Utrecht, The Netherlands

Fernando Santos
University of Alicante, Alicante, Spain

Cecilia Sarmiento
Tallinn University of Technology, Tallinn, Estonia

Takahide Sasaya
*National Agriculture and Food Research Organization,
Fukuyama, Japan*

Preethi Sathanantham
Virginia Tech, Blacksburg, VA, United States

Panayampalli S. Satheshkumar
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Yukiyo Sato
Okayama University, Kurashiki, Japan

Andreas Sauerbrei
Jena University Hospital, Jena, Germany

Eugene I. Savenkov
*Swedish University of Agricultural Sciences, Uppsala,
Sweden and Linnean Center for Plant Biology, Uppsala,
Sweden*

Carita Savolainen-Kopra
*National Institute for Health and Welfare, Helsinki,
Finland*

Kay Scheets
*Oklahoma State University, Stillwater, OK,
United States*

Uffe V. Schenider
*Copenhagen University Hospital Hvidovre, Hvidovre,
Denmark*

Richard H. Scheuermann
*J. Craig Venter Institute, La Jolla, CA, United States;
University of California, San Diego, CA, United States;
La Jolla Institute for Immunology, La Jolla, CA,
United States; and Global Virus Network, Baltimore,
MD, United States*

Manfred J. Schmitt
Saarland University, Saarbrücken, Germany

James E. Schoelz
University of Missouri, Columbia, MO, United States

Jason R. Schrad
*Michigan State University, East Lansing, MI,
United States*

Declan C. Schroeder
*University of Reading, Reading, United Kingdom and
University of Minnesota, St. Paul, MN, United States*

Stacey Schultz-Cherry
*St. Jude Children's Research Hospital, Memphis, TN,
United States*

Thomas F. Schulz
*Hannover Medical School, Institute of Virology,
Hannover, Germany and German Center for Infection
Research, Hannover-Braunschweig Site, Braunschweig,
Germany*

Catherine A. Scougall
The University of Adelaide, Adelaide, SA, Australia

Kimberley D. Seed
University of California, Berkeley, CA, United States

Joaquim Segalés
*Departament of Animal Health and Anatomy, Faculty of
Veterinary Medicine, Autonomous University of
Barcelona, Barcelona, Spain; Animal Health Research
Center (CReSA) – Institute of Agrifood Research and
Technology (IRTA), Campus UAB, Barcelona, Spain;
and OIE Collaborating Center for the Research and
Control of Emerging and Re-emerging Swine Diseases in
Europe (IRTA-CReSA), Barcelona, Spain*

Mateo Seoane-Blanco
National Center for Biotechnology, Madrid, Spain

Madhumati Sevvana
Purdue University, West Lafayette, IN, United States

Kazım Sezen
*Department of Biology, Karadeniz Technical University,
Trabzon, Turkey*

Arvind Sharma
Institut Pasteur, Paris, France

Sumit Sharma
Linköping University, Linköping, Sweden

James M. Sharp
*University of Zaragoza, Zaragoza, Spain and Edinburgh,
United Kingdom*

Qunxin She
Shandong University, Qingdao, China

Keith E. Shearwin
The University of Adelaide, Adelaide, SA, Australia

Hanako Shimura
Hokkaido University, Sapporo, Japan

Reina S. Sikkema
Erasmus Medical Center, Rotterdam, The Netherlands

Aaron Simkovich
Agriculture and Agri-Food Canada, London, ON,
Canada and The University of Western Ontario,
London, ON, Canada

Peter Simmonds
University of Oxford, Oxford, United Kingdom

Tarja Sironen
University of Helsinki, Helsinki, Finland

Susanna Sissonen
Finnish Institute for Health and Welfare, Helsinki,
Finland

Michael A. Skinner
Imperial College London, London, United Kingdom

Douglas E. Smith
University of California, San Diego, La Jolla, CA,
United States

Melvyn Smith
Viapath Analytics, Specialist Virology Centre, King's
College NHS Foundation Trust, London, United
Kingdom

Thomas J. Smith
The University of Texas Medical Branch, Galveston, TX,
United States

Teemu Smura
Helsinki University Hospital and University of Helsinki,
Helsinki, Finland

Eric J. Snijder
Leiden University Medical Center, Leiden,
The Netherlands

Gisela Soboll Hussey
Michigan State University, East Lansing, MI,
United States

Maria Söderlund-Venemo
University of Helsinki, Helsinki, Finland

Merike Sõmera
Tallinn University of Technology, Tallinn, Estonia

Eun G. Song
Seoul Women's University, Seoul, South Korea

Milan J. Sonneveld
Erasmus University Medical Center, Rotterdam,
The Netherlands

Beatriz Soriano
Biotechvana, Scientific Park University of Valencia and
Institute for Integrative Systems Biology (I2SysBio),
University of Valencia-Spanish National Research
Council, Valencia, Spain

Thomas E. Spencer
University of Missouri, Columbia, MO, United States

Pothur Sreenivasulu
Sri Venkateswara University, Tirupati, India

Ashley L. St. John
Duke-NUS Medical School, Singapore, Singapore

David K. Stammers
University of Oxford, Oxford, United Kingdom

John Stanley
John Innes Centre, Colney, United Kingdom

Glyn Stanway
University of Essex, Colchester, United Kingdom

Thilo Stehle
University of Tuebingen, Tuebingen, Germany and
Vanderbilt University School of Medicine, Nashville, TN,
United States

Gregory W. Stevenson
Iowa State University, Ames, IA, United States

Lucy Rae Stewart
Agricultural Research Service, US Department of
Agriculture, Wooster, OH, United States

C.C.M.M. Stijger
Wageningen University and Research Center, Bleiswijk,
The Netherlands

Peter G. Stockley
University of Leeds, Leeds, United Kingdom

David Stone
Weymouth Laboratory, Centre for Environment, Fisheries
and Aquaculture Science, Weymouth, United Kingdom

Ashley E Strother
The University of Texas Medical Branch, Galveston, TX,
United States

Sundharraman Subramanian
Michigan State University, East Lansing, MI,
United States

William C. Summers
Yale University, New Haven, CT, United States

Liyang Sun
Northwest A&F University, Yangling, China

Wesley I. Sundquist
University of Utah, Salt Lake City, UT, United States

Petri Susi
University of Turku, Turku, Finland

Curtis A. Suttle
University of British Columbia, Vancouver, BC, Canada

Nobuhiro Suzuki
Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

Lennart Svensson
Linköping University, Linköping, Sweden and Karolinska Institute, Stockholm, Sweden

Ronald Swanstrom
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Daniele M. Swetnam
University of California, Davis, CA, United states

Moriah L. Szpara
Pennsylvania State University, University Park, PA, United States

Keisuke Tabata
Heidelberg University, Heidelberg, Germany

Anna Taglienti
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Naoki Takeshita
Tokyo University of Agriculture and Technology, Fuchu, Japan

Kana Takeshita Urayama
Tokyo University of Agriculture and Technology, Fuchu, Japan

Michael E. Taliansky
The James Hutton Institute, Dundee, United Kingdom

Pan Tao
The Catholic University of America, Washington, DC, United States

Jacqueline E. Tate
Centers for Disease Control and Prevention, Atlanta, GA, United States

Satyanarayana Tatineni
Agricultural Research Service, US Department of Agriculture, Lincoln, NE, United States and University of Nebraska–Lincoln, Lincoln, NE, United States

Sisko Tauriainen
University of Turku, Turku, Finland

Norbert Tautz
University of Luebeck, Luebeck, Germany

Paulo Tavares
Institute for Integrative Biology of the Cell, CEA, CNRS, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

Nicholas M.I. Taylor
University of Copenhagen, Copenhagen, Denmark

Xu Tengzhi
University of California, Davis, CA, United States

Raquel Tenorio
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Robert B. Tesh
The University of Texas Medical Branch, Galveston, TX, United States

Vaskar Thapa
Pennsylvania State University, State College, PA, United States

John E. Thomas
The University of Queensland, Brisbane, QLD, Australia

Julie A. Thomas
Rochester Institute of Technology, Rochester, NY, United States

Lynn C. Thomason
Frederick National Laboratory for Cancer Research, Frederick, MD, United States

Elizabeth Ashley Thompson
The University of Southern Mississippi, Hattiesburg, MS, United States

Jeremy R. Thompson
Cornell University, Ithaca, NY, United States

Antonio Tiberini
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Peter Tijssen
National Institute of Scientific Research – Armand-Frappier Health Research Centre, Microbiology and Immunology, Laval, QC, Canada

Yuji Tomaru
Japan Fisheries Research and Education Agency, Kanagawa, Japan

Laura Tomassoli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Ruben Torres
National Biotechnology Center–Spanish National Research Council, Madrid, Spain

Jia Q. Truong
The University of Adelaide, Adelaide, SA, Australia

Erkki Truve[†]
Tallinn University of Technology, Tallinn, Estonia

Chih-Hsuan Tsai
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Roman Tuma
University of Leeds, Leeds, United Kingdom and
University of South Bohemia, České Budějovice, Czech Republic

Topi Turunen
Infectious Disease Unit, Espoo, Finland and Finnish
Institute for Health and Welfare, Helsinki, Finland

Reidun Twarock
University of York, York, United Kingdom

Ioannis E. Tzanetakis
University of Arkansas, Fayetteville, United States

Antti Vaheri
University of Helsinki, Helsinki, Finland

Eeva J. Vainio
Natural Resources Institute Finland (Luke), Helsinki, Finland

Anna M. Vaira
Institute for Sustainable Plant Protection, National
Research Council of Italy, Torino, Italy

Steven M. Valles
Center for Medical, Agricultural and Veterinary
Entomology, Agricultural Research Service, US
Department of Agriculture, Gainesville, FL,
United States

Adrián Valli
National Center for Biotechnology-Spanish National
Research Council, Madrid, Spain

Rodrigo A. Valverde
Louisiana State University Agricultural Center, Baton
Rouge, United States

Pierre Van Damme
University of Antwerp, Antwerp, Belgium

Rene A.A. van der Vlugt
Wageningen University and Research Center,
Wageningen, The Netherlands

Bernard A.M. Van der Zeijst
Leiden University Medical Center, Leiden,
The Netherlands

Koenraad Van Doorslaer
University of Arizona, Tucson, AZ, United States

James L. Van Etten
University of Nebraska–Lincoln, Lincoln, NE,
United States

Suzanne van Meer
University Medical Center Utrecht, Utrecht,
The Netherlands

Monique M. van Oers
Wageningen University and Research, Wageningen,
The Netherlands

Mark J. van Raaij
National Center for Biotechnology, Madrid, Spain

Marc H.V. Van Regenmortel
University of Strasbourg, Strasbourg, France

Piet A. van Rijn
Wageningen Bioveterinary Research, Lelystad, The
Netherlands and North-West University, Potchefstroom,
South Africa

Alain Vanderplasschen
University of Liège, Liège, Belgium

Dana L. Vanlandingham
College of Veterinary Medicine, Kansas State University,
Manhattan, KS, United States

Olli Vapalahti
Helsinki University Hospital and University of Helsinki,
Helsinki, Finland

Mark Varrelmann
Institute of Sugar Beet Research, Göttingen, Germany

Nikos Vasilakis
The University of Texas Medical Branch, Galveston, TX,
United States

Michael Veit
Free University of Berlin, Berlin, Germany

Česlovas Venclovas
Vilnius University, Vilnius, Lithuania

H. Josef Vetten
Julius Kühn Institute, Braunschweig, Germany

Marli Vlok
University of British Columbia, Vancouver, BC, Canada

Anne-Nathalie Volkoff
Diversity, Genomes and Insects-Microorganisms
Interactions, National Institute of Agricultural Research,
University of Montpellier, Montpellier, France

Ian E.H. Voorhees
Cornell University, Ithaca, NY, United States

Alex Vorsters
University of Antwerp, Antwerp, Belgium

[†]Deceased.

Jonathan D.F. Wadsworth
UCL Institute of Prion Diseases, London, United Kingdom

Peter J. Walker
The University of Queensland, St. Lucia, QLD, Australia

Paul Wallace
*Quality Control for Molecular Diagnostics (QCMD),
Glasgow, United Kingdom*

Aiming Wang
*Agriculture and Agri-Food Canada, London, ON,
Canada*

Jen-Ren Wang
National Cheng Kung University, Tainan, Taiwan

Lin-Fa Wang
Duke-NUS Medical School, Singapore, Singapore

Nan Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiangxi Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiaofeng Wang
Virginia Tech, Blacksburg, VA, United States

Katherine N. Ward
University College London, London, United Kingdom

Matti Waris
University of Turku, Turku, Finland

Ranjit Warrier
Purdue University, West Lafayette, IN, United States

Daniel Watterson
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Marta L. Wayne
University of Florida, Gainesville, FL, United States

Friedemann Weber
*FB 10 – Institute for Virology, Justus Liebig University
Giessen, Giessen, Germany*

Sung-Chan Wei
*Institute of Molecular Biology, Academia Sinica, Taipei,
Taiwan*

Robin A. Weiss
University College London, London, United Kingdom

Tao Weitao
*Southwest Baptist University, Bolivar, MO,
United States*

Kerstin Wernike
Friedrich-Loeffler-Institute, Insel Riems, Germany

Rachel J. Whitaker
*University of Illinois at Urbana-Champaign, Urbana, IL,
United States*

K. Andrew White
York University, Toronto, ON, Canada

Anna E. Whitfield
*North Carolina State University, Raleigh, NC,
United States*

Richard Whitley
*University of Alabama at Birmingham, Birmingham,
AL, United States*

Reed B. Wickner
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Luc Willems
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Brian J. Willett
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Alexis Williams
*The University of Texas Medical Branch, Galveston, TX,
United States*

Stephen A. Winchester
*Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom*

Clayton W. Winkler
*National Institutes of Health, Hamilton, MT,
United States*

Stephan Winter
*Leibniz Institute – DSMZ – German Collection of
Microorganisms and Cell Cultures GmbH, Braunschweig,
Germany*

William M. Wintermantel
*Agricultural Research Service, US Department of
Agriculture, Salinas, CA, United States*

Jennifer Wirth
Montana State University, Bozeman, MT, United States

Yuri I. Wolf
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States*

Thorsten Wolff
Robert Koch Institute, Berlin, Germany

Blaide Woodburn
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Michael E. Woodson
The University of Texas Medical Branch, Galveston, TX, United States

Courtney Woolsey
The University of Texas Medical Branch, Galveston, TX, United States

Chien-Fu Wu
Tokyo University of Agriculture and Technology, Fuchu, Japan

Mingde Wu
Huazhong Agricultural University, Wuhan, China

Songsong Wu
Huazhong Agricultural University, Wuhan, China

Yan Xiang
University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

Jiatao Xie
Huazhong Agricultural University, Wuhan, China

Zhuang Xiong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Hajime. Yaegashi
Institute of Fruit Tree and Tea Science, NARO, Morioka, Japan

Mehtap Yakupoğlu
Trabzon University, Trabzon, Turkey

Yasuyuki Yamaji
The University of Tokyo, Tokyo, Japan

Meng Yang
China Agricultural University, Beijing, China

Teng-Chieh Yang
Scarsdale, NY, United States

Qin Yao
Jiangsu University, Zhenjiang, China

Tianyou Yao
Baylor College of Medicine, Houston, TX, United States

Nobuyuki Yoshikawa
Iwate University, Morioka, Japan

George R. Young
Francis Crick Institute, London, United Kingdom

Lawrence S. Young
University of Warwick, Coventry, United Kingdom

Mark J. Young
Montana State University, Bozeman, MT, United States

Ry Young
Texas A&M University, College Station, TX, United States

Isaac T. Younker
University of Alabama, Tuscaloosa, AL, United States

Qian Yu
School of Life Sciences, Jiangsu University, Zhenjiang, China

Sang-Im Yun
Utah State University, Logan, UT, United States

Fauzia Zarreen
University of Delhi, New Delhi, India

Francisco M. Zerbini
Federal University of Viçosa, Viçosa, Brazil

Dong-Xiu Zhang
University of Maryland, Rockville, MD, United States

Jianqiang Zhang
Iowa State University, Ames, IA, United States

Junjie Zhang
Texas A&M University, College Station, TX, United States

Long Zhang
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Pan Zhang
Central South University, Changsha, China

Peijun Zhang
University of Oxford, Oxford, United Kingdom and Electron Bio-Imaging Centre, Diamond Light Source, Didcot, United Kingdom

Tao Zhang
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Yongliang Zhang
China Agricultural University, Beijing, China

Zhenlu Zhang
Shandong Agricultural University, Tai'an, China

Lixia Zhou
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Ling Zhu
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Heiko Ziebell
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

John Ziebuhr
*The Queen's University of Belfast, Belfast,
United Kingdom*

Jeffrey J. Zimmerman
Iowa State University, Ames, IA, United States

Falk Zucker
*Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany*

CONTENT OF ALL VOLUMES

Editors in Chief	v
Editorial Board	vii
Section Editors	ix
Foreword	xv
Preface	xvii
Guide to Use	xix
List of Contributors	xxi

VOLUME 1

The Virus as a Concept – Fundamentals of Virology

A Brief History of Virology <i>David J Rowlands</i>	3
The Origin of Viruses <i>Patrick Forterre and Morgan Gaïa</i>	14
The Virocell Concept <i>Patrick Forterre</i>	23
Virus Taxonomy <i>Jens H Kuhn</i>	28
The Greater Virus World and Its Evolution <i>Eugene V Koonin and Valerian V Dolja</i>	38
The Virus Species Concept <i>Peter Simmonds</i>	47
Genetic Diversity and Evolution of Viral Populations <i>Rafael Sanjuán and Pilar Domingo-Calap</i>	53
Mechanisms of RNA Virus Evolution <i>Lisa M Bono and Siobain Duffy</i>	62
Mechanisms of DNA Virus Evolution <i>Moriah L Szpara and Koenraad Van Doorslaer</i>	71
Paleovirology <i>Clément Gilbert</i>	79
Evolution Steered by Structure <i>Nicola GA Abrescia</i>	87

Pairwise Sequence Comparison in Virology <i>Tao Zhang, Zheng Gong, Tongkun Guo, Zhuang Xiong, and Yiming Bao</i>	100
Computational Analysis of Recombination in Viral Nucleotide Sequences <i>Miguel Arenas</i>	108
Phylogeny of Viruses <i>Alexander E Gorbalenya and Chris Lauber</i>	116
Virus Bioinformatics <i>Nikolaos Pappas, Simon Roux, Martin Hölzer, Kevin Lamkiewicz, Florian Mock, Manja Marz, and Bas E Dutilh</i>	124
Metagenomics in Virology <i>Simon Roux, Jelle Matthijnsens, and Bas E Dutilh</i>	133
Database and Analytical Resources for Viral Research Community <i>Sujal Phadke, Saichetana Macherla, and Richard H Scheuermann</i>	141
Classification of the Viral World Based on Atomic Level Structures <i>Janne J Ravantti and Nicola GA Abrescia</i>	153
Isolating, Culturing, and Purifying Viruses With a Focus on Bacterial and Archaeal Viruses <i>Katri Eskelin and Hanna M Oksanen</i>	162
High Throughput Sequencing and Virology <i>Graham L Freimanis and Nick J Knowles</i>	175
Single-Virus Genomics: Studying Uncultured Viruses, One at a Time <i>Manuel Martinez-Garcia, Francisco Martinez-Hernandez, and Joaquín Martínez Martínez</i>	184
Biophysical Characterizations in the Solution State <i>Robert P Rambo and Katsuaki Inoue</i>	191
Virus Crystallography <i>Jonathan M Grimes</i>	199
Advanced Light and Correlative Microscopy in Virology <i>Sergi Padilla-Parra, Charles A Coomer, and Irene Carlon-Andres</i>	208
Atomic Force Microscopy (AFM) Investigation of Viruses <i>Alexander McPherson</i>	218
Cryo-Electron Microscopy (CEM) Structures of Viruses <i>David Chmielewski and Wah Chiu</i>	233
Analysis of Viruses in the Cellular Context by Electron Tomography <i>Peijun Zhang and Luiza Mendonça</i>	242
Mathematical Modeling of Virus Architecture <i>Reidun Twarock</i>	248
Principles of Virus Structure <i>Madhumati Sevvana, Thomas Klose, and Michael G Rossmann[†]</i>	257
Structures of Small Icosahedral Viruses <i>Elizabeth E Fry, Jingshan Ren, and Claudine Porta</i>	278
Structural Principles of the Flavivirus Particle Organization and of Its Conformational Changes <i>Stéphane Duquerroy, Arvind Sharma, and Félix A Rey</i>	290
Reoviruses (Reoviridae) and Their Structural Relatives <i>Liya Hu, Mary K Estes, and B V Venkataram Prasad</i>	303

[†]Deceased.

Structures of Tailed Phages and Herpesviruses (<i>Herpesviridae</i>) <i>Montserrat Fàbrega-Ferrer and Miquel Coll</i>	318
Adenoviruses (<i>Adenoviridae</i>) and Their Structural Relatives <i>Gabriela N Condezo, Natalia Martín-González, Marta Pérez-Illana, Mercedes Hernando-Pérez, José Gallardo, and Carmen San Martín</i>	329
Negative Single-Stranded RNA Viruses (<i>Mononegavirales</i>): A Structural View <i>Juan Reguera</i>	345
Structure of Retrovirus Particles (<i>Retroviridae</i>) <i>David K Stammers and Jingshan Ren</i>	352
Structure of Helical Viruses <i>C Martin Lawrence</i>	362
Giant Viruses and Their Virophage Parasites <i>Rodrigo AL Rodrigues, Ana CdSP Andrade, Grazielle P Oliveira, and Jônatas S Abrahão</i>	372
Viral Replication Cycle <i>AJ Cann</i>	382
Viral Receptors <i>José M Casasnovas and Thilo Stehle</i>	388
Bacterial and Archeal Virus Entry <i>Minna M Poranen and Aušra Domanska</i>	402
Nonenveloped Eukaryotic Virus Entry <i>Ian M Jones and Polly Roy</i>	409
Enveloped Virus Membrane Fusion <i>Aurélié A Albertini and Yves Gaudin</i>	417
Genome Replication of Bacterial and Archaeal Viruses <i>Česlovas Venclovas</i>	429
Viral Transcription <i>David LV Bauer and Ervin Fodor</i>	439
Translation of Viral Proteins <i>Martin D Ryan and Garry A Luke</i>	444
Recombination <i>Jozef J Bujarski</i>	460
Assembly of Viruses: Enveloped Particles <i>CK Navaratnarajah, R Warriar, and RJ Kuhn</i>	468
Assembly of Viruses: Nonenveloped Particles <i>M Luo</i>	475
Virion Assembly: From Small Picornaviruses (<i>Picornaviridae</i>) to Large Herpesviruses (<i>Herpesviridae</i>) <i>Ling Zhu, Nan Wang, and Xiangxi Wang</i>	480
Genome Packaging <i>Richard J Bingham, Reidun Twarock, Carlos P Mata, and Peter G Stockley</i>	488
Virus Factories <i>Isabel Fernández de Castro, Raquel Tenorio, and Cristina Risco</i>	495
Release of Phages From Prokaryotic Cells <i>Jesse Cahill and Ry Young</i>	501
Virus Budding <i>Lara Rheinemann and Wesley I Sundquist</i>	519

Vesicle-Mediated Transcytosis and Export of Viruses <i>Efraín E Rivera-Serrano and Stanley M Lemon</i>	529
Vector Transmission of Animal Viruses <i>Houssam Attoui, Fauziah Mohd Jaafar, Rennos Fragkoudis, and Peter PC Mertens</i>	542
The Human Virome <i>Alexia Bordigoni, Sébastien Halary, and Christelle Desnues</i>	552
Epidemiology of Human and Animal Viral Diseases <i>Michael Edelstein</i>	559
Zoonosis, Emerging and Re-Emerging Viral Diseases <i>Janet M Daly</i>	569
Antiviral Innate Immunity: Introduction <i>Friedemann Weber</i>	577
Humoral and T Cell-Mediated Immunity to Viruses <i>Ane Ogbe and Lucy Dorrell</i>	584
Antigenicity and Antigenic Variation <i>Kuan-Ying A Huang, Xiaorui Chen, Che Ma, Dayna Cheng, Jen-Ren Wang, and Wan-Chun Lai</i>	597
Antigen Presentation <i>Andrew J McMichael</i>	601
Defense Against Viruses and Other Genetic Parasites in Prokaryotes <i>Kira S Makarova, Yuri I Wolf, and Eugene V Koonin</i>	606
Defective-Interfering Viruses <i>L Roux</i>	617
Ecology and Global Impacts of Viruses <i>Joanne B Emerson</i>	621
The Role of Retroviruses in Cellular Evolution <i>Andrea Kirmaier and Welkin E Johnson</i>	627
The Role of Bacteriophages in Bacterial Evolution <i>Chris M Rands and Harald Brüssow</i>	633
Viruses and Their Potential for Bioterrorism <i>Dana L Vanlandingham and Stephen Higgs</i>	644
The Use of Viral Promoters in Expression Vectors <i>Ian M Jones</i>	652
Oncolytic Viruses <i>Laura Burga and Mihnea Bostina</i>	658
Biotechnology Approaches to Modern Vaccine Design <i>George P Lomonosoff and Daniel Ponndorf</i>	662
Viruses: Impact on Science and Society <i>Neeraja Sankaran and Robin A Weiss</i>	671

VOLUME 2

Viruses as Infectious Agents: Human and Animal Viruses

Adenoviruses (<i>Adenoviridae</i>) <i>Balázs Harrach and Mária Benkő</i>	3
---	---

African Horse Sickness Virus (<i>Reoviridae</i>) <i>Piet A van Rijn</i>	17
African Swine Fever Virus (<i>Asfarviridae</i>) <i>Linda K Dixon, Rachel Nash, Philippa C Hawes, and Christopher L Netherton</i>	22
Akabane Virus and Schmallenberg Virus (<i>Peribunyaviridae</i>) <i>Martin Beer and Kerstin Wernike</i>	34
Alphaviruses Causing Encephalitis (<i>Togaviridae</i>) <i>Diane E Griffin</i>	40
Anelloviruses (<i>Anelloviridae</i>) <i>Fabrizio Maggi and Mauro Pistello</i>	48
Animal Lentiviruses (<i>Retroviridae</i>) <i>Esperanza Gomez-Lucia</i>	56
Animal Morbilliviruses (<i>Paramyxoviridae</i>) <i>Carina Conceicao and Dalan Bailey</i>	68
Animal Papillomaviruses (<i>Papillomaviridae</i>) <i>John S Munday</i>	79
Astroviruses (<i>Astroviridae</i>) <i>Virginia Hargest, Amy Davis, and Stacey Schultz-Cherry</i>	92
Avian Hepadnaviruses (<i>Hepadnaviridae</i>) <i>Allison R Jilbert, Georget Y Reaiche-Miller, and Catherine A Scougall</i>	100
Avian Herpesviruses (<i>Herpesviridae</i>) <i>Vishwanatha RAP Reddy and Venugopal Nair</i>	112
Avian Influenza Viruses (<i>Orthomyxoviridae</i>) <i>Nicolas Bravo-Vasquez and Stacey Schultz-Cherry</i>	117
Avian Leukosis and Sarcoma Viruses (<i>Retroviridae</i>) <i>Karen L Beemon</i>	122
Bluetongue Virus (<i>Reoviridae</i>) <i>Raghavendran Kulasegaran-Shylini and Polly Roy</i>	127
Borna Disease Virus and Related Bornaviruses (<i>Bornaviridae</i>) <i>Susan L Payne</i>	137
Bovine Leukemia Virus (<i>Retroviridae</i>) <i>Thomas Joris, Roghaiyeh Safari, Jean-Rock Jacques, and Luc Willems</i>	144
Bovine Viral Diarrhea, Border Disease, and Classical Swine Fever Viruses (<i>Flaviviridae</i>) <i>Paul Becher, Volker Moennig, and Norbert Tautz</i>	153
Capripoxviruses, Parapoxviruses, and Other Poxviruses of Ruminants (<i>Poxviridae</i>) <i>Philippa M Beard</i>	165
Chikungunya Virus (<i>Togaviridae</i>) <i>Thomas E Morrison and Stephanie E Ander</i>	173
Circoviruses (<i>Circoviridae</i>) <i>Giovanni Franzo and Joaquim Segalés</i>	182
Coronaviruses: General Features (<i>Coronaviridae</i>) <i>Paul Britton</i>	193
Coronaviruses: Molecular Biology (<i>Coronaviridae</i>) <i>X Deng and SC Baker</i>	198

Crimean-Congo Hemorrhagic Fever Virus and Nairoviruses of Medical Importance (<i>Nairoviridae</i>) <i>Ali Mirazimi, Felicity Burt, and Anna Papa</i>	208
Dengue Viruses (<i>Flaviviridae</i>) <i>Ashley L. St. John and Duane J. Gubler</i>	218
Ebola Virus (<i>Filoviridae</i>) <i>Andrea Marzi and Logan Banadyga</i>	232
Enteroviruses (<i>Picornaviridae</i>) <i>Carita Savolainen-Kopra, Soile Blomqvist, and Petri Susi</i>	245
Enveloped, Positive-Strand RNA Viruses (<i>Nidovirales</i>) <i>L. Enjuanes, AE Gorbalenya, RJ de Groot, JA Cowley, J Ziebuhr, and EJ Snijder</i>	256
Epstein-Barr Virus (<i>Herpesviridae</i>) <i>Lawrence S. Young</i>	267
Equine Herpesviruses (<i>Herpesviridae</i>) <i>Gisela Soboll-Hussey, Nikolaus Osterrieder, and Walid Azab</i>	278
Equine, Canine, and Swine Influenza (<i>Orthomyxoviridae</i>) <i>Janet M. Daly and Japhette E. Kembou-Ringert</i>	287
Feline Calicivirus (<i>Caliciviridae</i>) <i>Margaret J. Hosie and Michaela J. Conley</i>	294
Feline Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>Brian J. Willett and Margaret J. Hosie</i>	300
Fish and Amphibian Alloherpesviruses (<i>Herpesviridae</i>) <i>Maxime Boutier, Léa Morvan, Natacha Delrez, Francesco Origi, Andor Doszpoly, and Alain Vanderplasschen</i>	306
Fish Retroviruses (<i>Retroviridae</i>) <i>TA Paul, RN Casey, PR Bowser, JW Casey, J Rovnak, and SL Quackenbush</i>	316
Fish Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Gael Kurath and David Stone</i>	324
Foot-and-Mouth Disease Viruses (<i>Picornaviridae</i>) <i>David J. Rowlands</i>	332
Fowlpox Virus and Other Avipoxviruses (<i>Poxviridae</i>) <i>Efstathios S. Giotis and Michael A. Skinner</i>	343
Hantaviruses (<i>Hantaviridae</i>) <i>Tarja Sironen and Antti Vaheri</i>	349
Henipaviruses (<i>Paramyxoviridae</i>) <i>Lin-Fa Wang and Danielle E. Anderson</i>	355
Hepatitis A Virus (<i>Picornaviridae</i>) <i>Andreas Dotzauer</i>	362
Hepatitis B Virus (<i>Hepadnaviridae</i>) <i>Peter Karayiannis</i>	373
Hepatitis C Virus (<i>Flaviviridae</i>) <i>Ralf Bartenschlager and Keisuke Tabata</i>	386
Hepeviruses (<i>Hepeviridae</i>) <i>Xiang-Jin Meng</i>	397
Herpes Simplex Virus 1 and 2 (<i>Herpesviridae</i>) <i>David M. Knipe and Richard Whitley</i>	404

History of Virology: Vertebrate Viruses <i>F Fenner</i>	414
Human Boca- and Protoparvoviruses (<i>Parvoviridae</i>) <i>Maria Söderlund-Venermo and Jianming Qiu</i>	419
Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (<i>Coronaviridae</i>) <i>Ding X Liu, Jia Q Liang, and To S Fung</i>	428
Human Cytomegalovirus (<i>Herpesviridae</i>) <i>Edward S Mocarski</i>	441
Human Immunodeficiency Virus (<i>Retroviridae</i>) <i>Blaide Woodburn, Ann Emery, and Ronald Swanstrom</i>	460
Human Metapneumovirus (<i>Pneumoviridae</i>) <i>Antonella Casola, Matteo P Garofalo, and Xiaoyong Bao</i>	475
Human Norovirus and Sapovirus (<i>Caliciviridae</i>) <i>Sumit Sharma, Marie Hagbom, Lennart Svensson, and Johan Nordgren</i>	483
Human Papillomaviruses (<i>Papillomaviridae</i>) <i>Alison A McBride and Samuel S Porter</i>	493
Human Parainfluenza Viruses (<i>Paramyxoviridae</i>) <i>Elisabeth Adderson</i>	502
Human Pathogenic Arenaviruses (<i>Arenaviridae</i>) <i>Sheli R Radoshitzky and Juan C de la Torre</i>	507
Human Polyomaviruses (<i>Papillomaviridae</i>) <i>Melissa S Maginnis</i>	518
Human T-Cell Leukemia Virus-1 and -2 (<i>Retroviridae</i>) <i>Amanda R Panfil and Patrick L Green</i>	528
Infectious Bursal Disease Virus (<i>Birnaviridae</i>) <i>Daral J Jackwood</i>	540
Infectious Pancreatic Necrosis Virus (<i>Birnaviridae</i>) <i>Øystein Evensen</i>	544
Influenza A Viruses (<i>Orthomyxoviridae</i>) <i>Laura Kakkola, Niina Ikonen, and Ilkka Julkunen</i>	551
Influenza B, C and D Viruses (<i>Orthomyxoviridae</i>) <i>Thorsten Wolff and Michael Veit</i>	561
Jaagsiekte Sheep Retrovirus (<i>Retroviridae</i>) <i>James M Sharp, Marcelo De las Heras, Massimo Palmarini, and Thomas E Spencer</i>	575
Japanese Encephalitis Virus (<i>Flaviviridae</i>) <i>Sang-Im Yun and Young-Min Lee</i>	583
Kaposi's Sarcoma-Associated Herpesvirus (<i>Herpesviridae</i>) <i>Anne K Cordes and Thomas F Schulz</i>	598
Marburg and Ravn Viruses (<i>Filoviridae</i>) <i>Courtney Woolsey, Thomas W Geisbert, and Robert W Cross</i>	608
Measles Virus (<i>Paramyxoviridae</i>) <i>Roberto Cattaneo and Michael McChesney</i>	619
Molluscum Contagiosum Virus (<i>Poxviridae</i>) <i>Joachim J Bugert and Rosina Ehmann</i>	629

Mumps Virus (<i>Paramyxoviridae</i>) <i>Stephen A Winchester and Kevin E Brown</i>	634
Murine Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>George R Young and Kate N Bishop</i>	643
Newcastle Disease Virus (<i>Paramyxoviridae</i>) <i>Ben Peeters and Guus Koch</i>	648
Orthobunyaviruses (<i>Peribunyaviridae</i>) <i>Alyssa B Evans, Clayton W Winkler, and Karin E Peterson</i>	654
Parapoxviruses (<i>Poxviridae</i>) <i>Hanns-Joachim Rziha and Mathias Büttner</i>	666
Parechoviruses (<i>Picornaviridae</i>) <i>Sisko Tauriainen and Glyn Stanway</i>	675
Parvoviruses of Carnivores, and the Emergence of Canine Parvovirus (<i>Parvoviridae</i>) <i>Colin R Parrish, Ian EH Voorhees, and Susan L Hafenstein</i>	683
Polioviruses (<i>Picornaviridae</i>) <i>Philip D Minor</i>	688
Porcine Reproductive and Respiratory Syndrome Virus and Equine Arteritis Virus (<i>Arteriviridae</i>) <i>Jianqiang Zhang, Alan T Loynachan, Gregory W Stevenson, and Jeffrey J Zimmerman</i>	697
Prions of Vertebrates <i>Jonathan DF Wadsworth and John Collinge</i>	707
Pseudorabies Virus (<i>Herpesviridae</i>) <i>Thomas C Mettenleiter and Barbara G Klupp</i>	714
Rabbit Hemorrhagic Disease Virus and European Brown Hare Syndrome Virus (<i>Caliciviridae</i>) <i>Lorenzo Capucci, Patrizia Cavadini, and Antonio Lavazza</i>	724
Rabbit Myxoma Virus and the Fibroma Viruses (<i>Poxviridae</i>) <i>Peter J Kerr</i>	730
Rabies and Other Lyssaviruses (<i>Rhabdoviridae</i>) <i>Ashley C Banyard and Anthony R Fooks</i>	738
Respiratory Syncytial Virus (<i>Pneumoviridae</i>) <i>Tiffany King, Tiffany Jenkins, Supranee Chaiwatpongsakorn, and Mark E Peebles</i>	747
Rhinoviruses (<i>Picornaviridae</i>) <i>Matti Waris and Olli Ruuskanen</i>	757
Rift Valley Fever Virus and Other Phleboviruses (<i>Phenuiviridae</i>) <i>Tetsuro Ikegami</i>	765
Roseoloviruses: Human Herpesviruses 6A, 6B and 7 (<i>Herpesviridae</i>) <i>Katherine N Ward</i>	778
Rotaviruses (<i>Reoviridae</i>) <i>Juana Angel and Manuel A Franco</i>	789
Rubella Virus (<i>Picornaviridae</i>) <i>Annette Mankertz</i>	797
Saint Louis Encephalitis Virus (<i>Flaviviridae</i>) <i>William K Reisen, Lark L Coffey, Daniele M Swetnam, and Aaron C Brault</i>	805
Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (<i>Coronaviridae</i>) <i>Joseph SM Peiris and Leo LM Poon</i>	814

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (<i>Coronaviridae</i>) <i>Malik Peiris</i>	825
Simian Immunodeficiency Virus (SIV) and HIV-2 (<i>Retroviridae</i>) <i>Phyllis J Kanki</i>	827
Sindbis Virus (<i>Togaviridae</i>) <i>Satu Kurkela</i>	837
Tick-Borne Encephalitis Virus (<i>Flaviviridae</i>) <i>Teemu Smura, Suvi Kuivanen, and Olli Vapalahti</i>	843
Transmissible Gastroenteritis Virus of Pigs and Porcine Epidemic Diarrhea Virus (<i>Coronaviridae</i>) <i>Qiang Liu and Volker Gerds</i>	850
Vaccinia Virus (<i>Poxviridae</i>) <i>Yan Xiang and Rebecca K Lane</i>	854
Varicella-Zoster Virus (<i>Herpesviridae</i>) <i>Jeffrey I Cohen</i>	860
Variola and Monkeypox Viruses (<i>Poxviridae</i>) <i>Lalita Priyamvada and Panayampalli S Satheshkumar</i>	868
Vesicular Stomatitis Virus and Bovine Ephemeral Fever Virus (<i>Rhabdoviridae</i>) <i>Peter J Walker and Robert B Tesh</i>	875
West Nile Virus (<i>Flaviviridae</i>) <i>Fengwei Bai and Elizabeth Ashley Thompson</i>	884
Yellow Fever Virus (<i>Flaviviridae</i>) <i>Ashley E Strother and Alan DT Barrett</i>	891
Zika Virus (<i>Flaviviridae</i>) <i>Nikos Vasilakis, Shannan L Rossi, Sasha R Azar, Irma E Cisneros, Cassia F Estofolete, and Mauricio L Nogueira</i>	899

VOLUME 3

Viruses as Infectious Agents: Plant Viruses

An Introduction to Plant Viruses <i>Roger Hull</i>	3
Emerging and Re-Emerging Plant Viruses <i>Sabrina Bertin, Francesco Faggioli, Andrea Gentili, Ariana Manglli, Anna Taglienti, Antonio Tiberini, and Laura Tomassoli</i>	8
Emerging Geminiviruses (<i>Geminiviridae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, and Muhammad Mubin</i>	21
Movement of Viruses in Plants <i>Manfred Heinlein</i>	32
Plant Antiviral Defense: Gene-Silencing Pathways <i>Vitantonio Pantaleo, Chikara Masuta, and Hanako Shimura</i>	43
Plant Resistance to Viruses: Engineered Resistance <i>Marc Fuchs</i>	52
Plant Resistance to Viruses: Natural Resistance Associated With Dominant Genes <i>Mandy Ravensbergen and Richard Kormelink</i>	60

Plant Resistance to Viruses: Natural Resistance Associated With Recessive Genes <i>Masayoshi Hashimoto, Kensaku Maejima, Yasuyuki Yamaji, and Shigetou Namba</i>	69
Plant Viral Diseases: Economic Implications <i>Basavaprabhu L Patil</i>	81
Retrotransposons of Plants <i>M-A Grandbastien</i>	98
Vector Transmission of Plant Viruses <i>Etienne Herrbach and Quentin Chesnais</i>	106
Viral Suppressors of Gene Silencing <i>Hernan Garcia-Ruiz</i>	116
Virus-Induced Gene Silencing (VIGS) <i>Xu Tengzhi, Ugrappa Nagalakshmi, and Savithramma P Dinesh-Kumar</i>	123
Alfalfa Mosaic Virus (<i>Bromoviridae</i>) <i>L Sue Loesch-Fries</i>	132
Alphaflexiviruses (<i>Alphaflexiviridae</i>) <i>Sergey Y Morozov and Alexey A Agranovsky</i>	140
Alphasatellites (<i>Alphasatellitidae</i>) <i>Rob W Briddon and Muhammad S Nawaz-ul-Rehman</i>	149
Amalgaviruses (<i>Amalgaviridae</i>) <i>Ioannis E Tzanetakis, Sead Sabanadzovic, and Rodrigo A Valverde</i>	154
Badnaviruses (<i>Caulimoviridae</i>) <i>Andrew DW Geering</i>	158
Banana Bunchy Top Virus (<i>Nanoviridae</i>) <i>John E Thomas</i>	169
Barley Yellow Dwarf Viruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	176
Bean Common Mosaic Virus and Bean Common Mosaic Necrosis Virus (<i>Potyviridae</i>) <i>Ramon Jordan and John Hammond</i>	184
Bean Golden Mosaic Virus and Bean Golden Yellow Mosaic Virus (<i>Geminiviridae</i>) <i>Francisco M Zerbini and Simone G Ribeiro</i>	192
Beet Curly Top Virus (<i>Geminiviridae</i>) <i>Robert L Gilbertson, Tomas A Melgarejo, Maria R Rojas, William M Wintermantel, and John Stanley</i>	200
Beet Necrotic Yellow Vein Virus (<i>Benyviridae</i>) <i>Sebastian Liebe, Annette Niehl, Renate Koenig, and Mark Varrelmann</i>	213
Benyviruses (<i>Benyviridae</i>) <i>Annette Niehl, Sebastian Liebe, Mark Varrelmann, and Renate Koenig</i>	219
Betaflexiviruses (<i>Betaflexiviridae</i>) <i>Nobuyuki Yoshikawa and Hajime Yaegashi</i>	229
Betasatellites and Deltasatellites (<i>Toleusatellitidae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, Muhammad Hassan, and Muhammad Mubin</i>	239
Bluner-, Cile-, and Higreviruses (<i>Kitaviridae</i>) <i>Diego F Quito-Avila, Juliana Freitas-Astúa, and Michael J Melzer</i>	247
Brome Mosaic Virus (<i>Bromoviridae</i>) <i>Guijuan He, Zhenlu Zhang, Preethi Sathanantham, Arturo Diaz, and Xiaofeng Wang</i>	252

Bromoviruses (<i>Bromoviridae</i>) <i>Jozef J Bujarski</i>	260
Bymoviruses (<i>Potyviridae</i>) <i>Annette Niehl and Frank Rabenstein</i>	268
Cacao Swollen Shoot Virus (<i>Caulimoviridae</i>) <i>Emmanuelle Muller</i>	274
Carmo-Like Viruses (<i>Tombusviridae</i>) <i>Miryam Pérez-Cañamás and Carmen Hernández</i>	285
Cassava Brown Streak Viruses (<i>Potyviridae</i>) <i>Basavaprabhu L Patil</i>	293
Cassava Mosaic Viruses (<i>Geminiviridae</i>) <i>James Legg and Stephan Winter</i>	301
Caulimoviruses (<i>Caulimoviridae</i>) <i>James E Schoelz and Mustafa Adhab</i>	313
Cheraviruses, Sadwaviruses and Torradoviruses (<i>Secoviridae</i>) <i>Toru Iwanami and René AA van der Vlugt</i>	322
Citrus Tristeza Virus (<i>Closteroviridae</i>) <i>Moshe Bar-Joseph, Scott J Harper, and William O Dawson</i>	327
Closteroviruses (<i>Closteroviridae</i>) <i>Marc Fuchs</i>	336
Comoviruses and Fabaviruses (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	348
Cotton Leaf Curl Disease (<i>Geminiviridae</i>) <i>Nasim Ahmed, Imran Amin, and Shahid Mansoor</i>	355
Cowpea Mosaic Virus (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	364
Cucumber Mosaic Virus (<i>Bromoviridae</i>) <i>Judith Hirsch and Benoît Moury</i>	371
Dianthovirus (<i>Tombusviridae</i>) <i>Kiwamu Hyodo and Masanori Kaido</i>	383
Endornaviruses (<i>Endornaviridae</i>) <i>Toshiyuki Fukuhara</i>	388
Fimoviruses (<i>Fimoviridae</i>) <i>Toufic Elbeaino and Michele Digiaro</i>	396
Furoviruses (<i>Virgaviridae</i>) <i>Annette Niehl and Renate Koenig</i>	405
Geminiviruses (<i>Geminiviridae</i>) <i>Jesús Navas-Castillo and Elvira Fiallo-Olivé</i>	411
Hordeiviruses (<i>Virgaviridae</i>) <i>Zhihao Jiang, Meng Yang, Yongliang Zhang, Andrew O Jackson, and Dawei Li</i>	420
Idaeoviruses (<i>Mayoviridae</i>) <i>Robert R Martin and Karen E Keller</i>	430
Ilarviruses (<i>Bromoviridae</i>) <i>Aaron Simkovich, Susanne E Kohalmi, and Aiming Wang</i>	439

Luteoviruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	447
Machlomovirus and Panicoviruses (<i>Tombusviridae</i>) <i>Kay Scheets</i>	456
Maize Streak Virus (<i>Geminiviridae</i>) <i>Darren P Martin and Aderito L Monjane</i>	461
Nanoviruses (<i>Nanoviridae</i>) <i>Bruno Gronenborn and H Josef Vetten</i>	470
Necro-Like Viruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Giovanni P Martelli[†]</i>	481
Nepoviruses (<i>Secoviridae</i>) <i>Hélène Sanfaçon</i>	486
Ophioviruses (<i>Aspiviridae</i>) <i>Anna M Vaira and John Hammond</i>	495
Orthospoviruses (<i>Tospoviridae</i>) <i>Renato O Resende and Hanu R Pappu</i>	507
Ourmiaviruses (<i>Botourmiaviridae</i>) <i>Gian Paolo Accotto and Cristina Rosa</i>	516
Papaya Ringspot Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	520
Pecluviruses (<i>Virgaviridae</i>) <i>Hema Masarapu, Pothur Sreenivasulu, Philippe Delfosse, Claude Bragard, Anne Legreve, and DVR Reddy</i>	528
Pepino Mosaic Virus (<i>Alphaflexiviridae</i>) <i>Rene AA van der Vlugt and CCMM Stijger</i>	539
Plant Reoviruses (<i>Reoviridae</i>) <i>Yu Huang and Yi Li</i>	545
Plant Resistance to Geminiviruses <i>Basavaprabhu L Patil, Supriya Chakraborty, Henryk Czosnek, Elvira Fiallo-Olivé, Robert L Gilbertson, James Legg, Shahid Mansoor, Jesús Navas-Castillo, Rubab Z Naqvi, Saleem U Rahman, and Francisco M Zerbini</i>	554
Plant Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Ralf G Dietzgen, Michael M Goodin, and Zhenghe Li</i>	567
Plant Satellite Viruses (<i>Albetovirus, Aumaivirus, Papanivirus, Virtovirus</i>) <i>Mart Krupovic</i>	581
Plum Pox Virus (<i>Potyviridae</i>) <i>Miroslav Glasa and Thierry Candresse</i>	586
Poleroviruses (<i>Luteoviridae</i>) <i>Hernan Garcia-Ruiz, Natalie M Holste, and Katherine LaTourrette</i>	594
Pomoviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	603
Potato Virus Y (<i>Potyviridae</i>) <i>Laurent Glais and Benoît Moury</i>	612
Potexviruses (<i>Alphaflexiviridae</i>) <i>Ki H Ryu, Eun G Song, and Jin S Hong</i>	623

[†]Deceased.

Potyviruses (<i>Potyviridae</i>) <i>Adrián Valli, Juan A García, and Juan J López-Moya</i>	631
Quinviruses (<i>Betaflexiviridae</i>) <i>Ki H Ryu and Eun G Song</i>	642
Reverse-Transcribing Viruses (<i>Belpaoviridae, Metaviridae, and Pseudoviridae</i>) <i>Carlos Llorens, Beatriz Soriano, Maria A Navarrete-Muñoz, Ahmed Hafez, Vicente Arnau, Jose Miguel Benito, Toni Gabaldon, Norma Rallon, Jaume Pérez-Sánchez, and Mart Krupovic</i>	653
Rice Tungro Disease (<i>Secoviridae, Caulimoviridae</i>) <i>Gaurav Kumar, Fauzia Zarreen, and Indranil Dasgupta</i>	667
Rice Yellow Mottle Virus (<i>Solemoviridae</i>) <i>Eugénie Hébrard, Nils Poulicard, and Mbolarinosy Rakotomalala</i>	675
Satellite Nucleic Acids and Viruses <i>Olufemi J Alabi, Alfredo Diaz-Lara, and Maher Al Rwahnih</i>	681
Secoviruses (<i>Secoviridae</i>) <i>Jeremy R Thompson</i>	692
Sequiviruses and Waikaviruses (<i>Secoviridae</i>) <i>Lucy Rae Stewart</i>	703
Solemoviruses (<i>Solemoviridae</i>) <i>Cecilia Sarmiento, Merike Sõmera, and Erkki Truve[†]</i>	712
Tenuiviruses (<i>Phenuiviridae</i>) <i>Bertha Cecilia Ramirez and Anne-Lise Haenni</i>	719
Tobacco Mosaic Virus (<i>Virgaviridae</i>) <i>Marc HV Van Regenmortel</i>	727
Tobamoviruses (<i>Virgaviridae</i>) <i>Ulrich Melcher, Dennis J Lewandowski, and William O Dawson</i>	734
Tobraviruses (<i>Virgaviridae</i>) <i>Stuart A MacFarlane</i>	743
Tomato Leaf Curl New Delhi Virus (<i>Geminiviridae</i>) <i>Supriya Chakraborty and Manish Kumar</i>	749
Tomato Spotted Wilt Virus (<i>Tospoviridae</i>) <i>Hanu R Pappu, Anna E Whitfield, and Athos S de Oliveira</i>	761
Tomato Yellow Leaf Curl Viruses (<i>Geminiviridae</i>) <i>Henryk Czosnek</i>	768
Tombusvirus-Like Viruses (<i>Tombusviridae</i>) <i>K Andrew White</i>	778
Tombusviruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Kay Scheets</i>	788
Tritimoviruses and Rymoviruses (<i>Potyviridae</i>) <i>Satyanarayana Tatineni and Gary L Hein</i>	797
Triviruses (<i>Betaflexiviridae</i>) <i>Yahya ZA Gaafar and Heiko Ziebell</i>	805
Tymoviruses (<i>Tymoviridae</i>) <i>Rosemarie W Hammond and Peter Abrahamian</i>	818

[†]Deceased.

Umbraviruses (<i>Tombusviridae</i>) <i>Eugene V Ryabov and Michael E Taliansky</i>	827
Varicosaviruses (<i>Rhabdoviridae</i>) <i>Takahide Sasaya</i>	833
Virgaviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	839
Viroids (<i>Pospiviroidae</i> and <i>Avsunviroidae</i>) <i>Ricardo Flores and Robert A Owens</i>	852
Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	862

VOLUME 4

Viruses as Infectious Agents: Bacterial, Archaeal, Fungal, Algal, and Invertebrate Viruses

Bacterial Viruses

History of Virology: Bacteriophages <i>William C Summers</i>	3
Icosahedral Phages – Single-Stranded DNA (ϕ X174) <i>Bentley A Fane and Aaron P Roznowski</i>	10
Single-Stranded RNA Bacterial Viruses <i>Peter G Stockley and Junjie Zhang</i>	21
Enveloped Icosahedral Phages – Double-Stranded RNA (ϕ 6) <i>Paul Gottlieb and Aleksandra Alimova</i>	26
Membrane-Containing Icosahedral DNA Bacteriophages <i>Roman Tuma, Sarah J Butcher, and Hanna M Oksanen</i>	36
Tailed Double-Stranded DNA Phages <i>Robert L Duda</i>	45
Helical and Filamentous Phages <i>Andreas Kuhn and Sebastian Leptihn</i>	53
Replication of <i>Bacillus</i> Double-Stranded DNA Bacteriophages <i>Silvia Ayora, Paulo Tavares, Ruben Torres, and Juan C Alonso</i>	61
Lytic Transcription <i>William McAllister and Deborah M Hinton</i>	69
Lysogeny <i>Keith E Shearwin and Jia Q Truong</i>	77
Decision Making by Temperate Phages <i>Ido Golding, Seth Coleman, Thu VP Nguyen, and Tianyou Yao</i>	88
Mobilization of Phage Satellites <i>Kristen N LeGault and Kimberley D Seed</i>	98
Portal Vertex <i>Peng Jing and Mauricio Cortes Jr.</i>	105

Prohead, the Head Shell Pre-Cursor <i>Marc C Morais and Michael E Woodson</i>	115
Enzymology of Viral DNA Packaging Machines <i>Carlos E Catalano</i>	124
DNA Packaging: DNA Recognition <i>Sandra J Greive and Oliver W Bayfield</i>	136
DNA Packaging: The Translocation Motor <i>Janelle A Hayes and Brian A Kelch</i>	148
Biophysics of DNA Packaging <i>Joshua Pajak, Gaurav Arya, and Douglas E Smith</i>	160
Energetics of the DNA-Filled Head <i>Alex Evilevitch</i>	167
Bacteriophage Receptor Proteins of Gram-Negative Bacteria <i>Sarah M Doore, Kristin N Parent, Sundharraman Subramanian, Jason R Schrad, and Natalia B Hubbs</i>	175
Tail Structure and Dynamics <i>Shweta Bhatt, Petr G Leiman, and Nicholas MI Taylor</i>	186
Bacteriophage Tail Fibres, Tailspikes, and Bacterial Receptor Interaction <i>Mateo Seoane-Blanco, Mark J van Raaij, and Meritxell Granell</i>	194
Phage Genome and Protein Ejection In Vivo <i>Ian J Molineux, L Letti Lopez, and Aaron P Roznowski</i>	206
Dealing With the Whole Head: Diversity and Function of Capsid Ejection Proteins in Tailed Phages <i>Lindsay W Black and Julie A Thomas</i>	219
Jumbo Phages <i>Isaac T Younker and Carol Duffy</i>	229
CRISPR-Cas Systems and Anti-CRISPR Proteins: Adaptive Defense and Counter-Defense in Prokaryotes and Their Viruses <i>Asma Hatoum-Aslan and Olivia G Howell</i>	242
Bacteriophage: Therapeutics and Diagnostics Development <i>Teng-Chieh Yang</i>	252
Bacteriophage Vaccines <i>Pan Tao and Venigalla B Rao</i>	259
Bacteriophage Diversity <i>Julianne H Grose and Sherwood R Casjens</i>	265
Genetic Mosaicism in the Tailed Double-Stranded DNA Phages <i>Welkin H Pope</i>	276
Bacteriophages of the Human Microbiome <i>Pilar Manrique, Michael Dills, and Mark J Young</i>	283
Bacteriophage: Red Recombination System and the Development of Recombineering Technologies <i>Lynn C Thomason and Kenan C Murphy</i>	291
Nanotechnology Application of Bacteriophage DNA Packaging Nanomotors <i>Tao Weitao, Lixia Zhou, Zhefeng Li, Long Zhang, and Peixuan Guo</i>	302
General Ecology of Bacteriophages <i>Stephen T Abedon</i>	314
Marine Bacteriophages <i>Vera Bischoff, Falk Zucker, and Cristina Moraru</i>	322

Ecology of Phages in Extreme Environments <i>Tatiana A Demina and Nina S Atanasova</i>	342
---	-----

Archaeal Viruses

Diversity of Hyperthermophilic Archaeal Viruses <i>David Prangishvili, Mart Krupovic, and Diana P Baquero</i>	359
Euryarchaeal Viruses <i>Tatiana A Demina and Hanna M Oksanen</i>	368
Vesicle-Like Archaeal Viruses <i>Elina Roine and Nina S Atanasova</i>	380
Virus–Host Interactions in Archaea <i>Diana P Baquero, David Prangishvili, and Mart Krupovic</i>	387
Antiviral Defense Mechanisms in Archaea <i>Qunxin She</i>	400
Discovery of Archaeal Viruses in Hot Spring Environments Using Viral Metagenomics <i>Jennifer Wirth, Jacob H Munson-McGee, and Mark J Young</i>	407
Metagenomes of Archaeal Viruses in Hypersaline Environments <i>Fernando Santos, María D Ramos-Barbero, and Josefa Antón</i>	414
Extreme Environments as a Model System to Study How Virus–Host Interactions Evolve Along the Symbiosis Continuum <i>Samantha J DeWerff and Rachel J Whitaker</i>	419

Fungal Viruses

An Introduction to Fungal Viruses <i>Nobuhiro Suzuki</i>	431
Cross-Kingdom Virus Infection <i>Liyang Sun, Hideki Kondo, and Ida Bagus Andika</i>	443
Diversity of Mycoviruses in Aspergilli <i>Ioly Kotta-Loizou</i>	450
Evolution of Mycoviruses <i>Mahtab Peyambari, Vaskar Thapa, and Marilyn J Roossinck</i>	457
Mixed Infections of Mycoviruses in Phytopathogenic Fungus <i>Sclerotinia sclerotiorum</i> <i>Jiatao Xie and Daohong Jiang</i>	461
Mycovirus-Mediated Biological Control <i>Daniel Rigling, Cécile Robin, and Simone Prospero</i>	468
Mycoviruses With Filamentous Particles <i>Michael N Pearson</i>	478
Prions of Yeast and Fungi <i>Reed B Wickner and Herman K Edskes</i>	487
Single-Stranded DNA Mycoviruses <i>Daohong Jiang</i>	493
Structure of Double-Stranded RNA Mycoviruses <i>José R Castón, Nobuhiro Suzuki, and Said A Ghabrial[†]</i>	504

[†]Deceased.

<i>Ustilago maydis</i> Viruses and Their Killer Toxins <i>Alexis Williams and Thomas J Smith</i>	513
Vegetative Incompatibility in Filamentous Fungi <i>Songsong Wu, Daohong Jiang, and Jiatao Xie</i>	520
Viral Diseases of <i>Agaricus bisporus</i> , the Button Mushroom <i>Kerry S Burton and Greg Deakin</i>	528
Viral Killer Toxins <i>Manfred J Schmitt and Björn Becker</i>	534
Alternaviruses (Unassigned) <i>Hiromitsu Moriyama, Nanako Aoki, Kuko Fuke, Kana Takeshita Urayama, Naoki Takeshita, and Chien-Fu Wu</i>	544
Barnaviruses (Barnaviridae) <i>Peter A Revill</i>	549
Botybirnaviruses (Botybirnavirus) <i>Mingde Wu, Guoqing Li, Daohong Jiang, and Jiatao Xie</i>	552
Chrysoviruses (Chrysoviridae) - General Features and Chrysovirus-Related Viruses <i>Ioly Kotta-Loizou, Robert HA Coutts, José R Castón, Hiromitsu Moriyama, and Said A Ghabrial[†]</i>	557
Fungal Partitiviruses (Partitiviridae) <i>Eeva J Vainio</i>	568
Fusariviruses (Unassigned) <i>Sotaro Chiba</i>	577
Giardiavirus (Totiviridae) <i>Juliana Gabriela Silva de Lima, João Paulo Matos Santos Lima, and Daniel Carlos Ferreira Lanza</i>	582
Hypoviruses (Hypoviridae) <i>Dong-Xiu Zhang and Donald L Nuss</i>	589
Megabirnaviruses (Megabirnaviridae) <i>Yukiyo Sato and Nobuhiro Suzuki</i>	594
Mitoviruses (Mitoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	601
Mycoreoviruses (Reoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	607
Mymonaviruses (Mymonaviridae) <i>Daohong Jiang</i>	615
Narnaviruses (Narnaviridae) <i>Rosa Esteban and Tsutomu Fujimura</i>	621
Phlegiviruses (Unassigned) <i>Karel Petrzik</i>	627
Plant and Protozoal Partitiviruses (Partitiviridae) <i>Hanna Rose and Edgar Maiss</i>	632
Quadriviruses (Quadriviridae) <i>Hideki Kondo, José R Castón, and Nobuhiro Suzuki</i>	642
Totiviruses (Totiviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	648

[†]Deceased.

Yado-kari Virus 1 and Yado-nushi Virus 1 (Unassigned) <i>Subha Das and Nobuhiro Suzuki</i>	658
Yeast L-A Virus (<i>Totiviridae</i>) <i>Reed B Wickner, Tsutomu Fujimura, and Rosa Esteban</i>	664
Algal Viruses	
Algal Marnaviruses (<i>Marnaviridae</i>) <i>Marli Vlok, Curtis A Suttle, and Andrew S Lang</i>	671
Algal Mimiviruses (<i>Mimiviridae</i>) <i>Ruth-Anne Sandaa, Håkon Dahle, Corina PD Brussaard, Hiroyuki Ogata, and Romain Blanc-Mathieu</i>	677
Miscellaneous Algal Viruses (<i>Alvernaviridae, Bacilladnaviridae, Dinodnavirus, Reoviridae</i>) <i>Keizo Nagasaki, Yuji Tomaru, and Corina PD Brussaard</i>	684
Phycodnaviruses (<i>Phycodnaviridae</i>) <i>James L Van Etten, David D Dunigan, Keizo Nagasaki, Declan C Schroeder, Nigel Grimsley, Corina PD Brussaard, and Jozef I Nissimov</i>	687
Invertebrate Viruses	
An Introduction to Viruses of Invertebrates <i>Peter Krell</i>	699
Ascoviruses (<i>Ascoviridae</i>) <i>Sassan Asgari, Dennis K Bideshi, Yves Bigot, and Brian A Federici</i>	724
Baculovirus–Host Interactions: Repurposing Host-Acquired Genes (<i>Baculoviridae</i>) <i>A Lorena Passarelli</i>	732
Baculoviruses: General Features (<i>Baculoviridae</i>) <i>Vera ID Ros</i>	739
Baculoviruses: Molecular Biology and Replication (<i>Baculoviridae</i>) <i>Monique M van Oers</i>	747
Bidensovirus (<i>Bidnaviridae</i>) <i>Qin Yao, Zhaoyang Hu, and Keping Chen</i>	759
Bunyaviruses of Arthropods (<i>Myxoviridae, Nairoviridae, Peribunyaviridae, Phasmaviridae, Phuniviridae, Wupedeviridae</i>) <i>Sandra Junglen</i>	764
Dicistroviruses (<i>Dicistroviridae</i>) <i>Yanping Chen and Steven M Valles</i>	768
Entomobirnaviruses (<i>Birnaviridae</i>) <i>Marco Marklewitz</i>	776
Hytroviruses (<i>Hytroviridae</i>) <i>Henry M Kariithi and Irene K Meki</i>	780
Iflaviruses (<i>Iflaviridae</i>) <i>Bryony C Bonning and Sijun Liu</i>	792
Iridoviruses of Invertebrates (<i>Iridoviridae</i>) <i>İkbal Agah İnce</i>	797
Mesoniviruses (<i>Mesoniviridae</i>) <i>Jody Hobson-Peters and Daniel Watterson</i>	804

Nimaviruses (<i>Nimaviridae</i>) <i>Peter Krell and Emine Ozsahin</i>	808
Nodaviruses of Invertebrates and Fish (<i>Nodaviridae</i>) <i>Kyle L Johnson and Jacen S Moore</i>	819
Nudiviruses (<i>Nudiviridae</i>) <i>Yu-Chan Chao, Chih-Hsuan Tsai, and Sung-Chan Wei</i>	827
Parvoviruses of Invertebrates (<i>Parvoviridae</i>) <i>Judit J Péntzes, Hanh T Pham, Qian Yu, Max Bergoin, and Peter Tijssen</i>	835
Polydnaviruses (<i>Polydnaviridae</i>) <i>Anne-Nathalie Volkoff and Elisabeth Huguet</i>	849
Poxviruses of Insects (<i>Poxviridae</i>) <i>Basil Arif, Lillian Pavlik, Remziye Nalçacıoğlu, Hacer Muratoğlu, Cihan İnan, Mehtap Yakupoğlu, Emine Özşahin, İsmail Demir, Kazım Sezen, and Zihni Demirbağ</i>	858
Reoviruses of Invertebrates (<i>Reoviridae</i>) <i>Peter Krell</i>	867
Rhabdoviruses of Insects (<i>Rhabdoviridae</i>) <i>Andrea González-González, Nicole T de Stefano, David A Rosenbaum, and Marta L Wayne</i>	883
Sarothroviruses (<i>Sarothroviridae</i>) <i>Azeez Sait Sahul Hameed</i>	888
Solinviviruses (<i>Solinviviridae</i>) <i>Steven M Valles and Andrew E Firth</i>	892
Tetraviruses (<i>Alphatetraviridae, Carmotetraviridae, Permutotetraviridae</i>) <i>Rosemary A Dorrington, Tatiana Domitrovic, and Meesbah Jiwaji</i>	897

VOLUME 5

Diagnosis, Treatment and Prevention of Virus Infections

Diagnosis

Introduction to Virus Diagnosis and Treatment <i>Maija Lappalainen and Hubert GM Niesters</i>	3
Electron Microscopy for Viral Diagnosis <i>Roland A Fleck</i>	5
Serological Approaches for Viral Diagnosis <i>Klaus Hedman and Visa Nurmi</i>	15
A Brief History of the Development of Diagnostic Molecular-Based Assays <i>Hubert GM Niesters</i>	22
Sequencing Strategies <i>Sibnarayan Datta</i>	27
Validating Real-Time Polymerase Chain Reaction (PCR) Assays <i>Melvyn Smith</i>	35
Rapid Point-of-Care Assays <i>Jan G Lisby and Uffe V Schenider</i>	45

Standardization of Diagnostic Assays <i>Sally A Baylis, C Micha Nübling, and Wayne Dimech</i>	52
Quality Assurance in the Clinical Virology Laboratory <i>Paul Wallace and Elaine McCulloch</i>	64
Biosafety and Biosecurity in Diagnostic Laboratories <i>Hannimari Kallio-Kokko and Susanna Sissonen</i>	82
Screening for Viral Infections <i>Walter Ian Lipkin, Nischay Mishra, and Thomas Briebe</i>	91
Clinical Diagnostic Virology <i>Marcus Panning</i>	98
Virus Diagnosis in Immunosuppressed Individuals <i>Elisabeth Puchhammer-Stöckl and Fausto Baldanti</i>	105
Diagnosis; Future Prospects on Direct Diagnosis <i>Marianna Calabretto, Daniele Di Carlo, Fabrizio Maggi, and Guido Antonelli</i>	112

Treatment

Antiviral Classification <i>Guangdi Li, Xixi Jing, Pan Zhang, and Erik De Clercq</i>	121
Antiretroviral Therapy – Nucleoside/Nucleotide and Non-Nucleoside Reverse Transcriptase Inhibitors <i>Timothy D Appleby and Killian J Quinn</i>	131
Protease Inhibitors <i>Vanesa Anton-Vazquez and Frank A Post</i>	139
HIV Integrase Inhibitors and Entry Inhibitors <i>Daniel Bradshaw and Ranjababu Kulasegaram</i>	145
Management of Respiratory Syncytial Virus Infections (<i>Pneumoviridae</i>) <i>Rachael S Barr and Simon B Drysdale</i>	155
Management of Influenza Virus Infections (<i>Orthomyxoviridae</i>) <i>Bruno Lina</i>	160
Management of Herpes Simplex Virus Infections (<i>Herpesviridae</i>) <i>Nicole Samies and Richard Whitley</i>	175
Management of Varicella-Zoster Virus Infections (<i>Herpesviridae</i>) <i>Andreas Sauerbrei</i>	181
Treatment and Prevention of Herpesvirus Infections in the Immunocompromised Host <i>Sara H Burkhard and Nicolas J Mueller</i>	190
Management of Adenovirus Infections (<i>Adenoviridae</i>) <i>Albert Heim</i>	197
Management of Hepatitis A and E Virus Infection <i>Sébastien Lhomme, Florence Abravanel, Jean-Marie Peron, Nassim Kamar, and Jacques Izopet</i>	206
Management of Patients With Chronic Hepatitis B (<i>Hepadnaviridae</i>) and Chronic Hepatitis D Infection (<i>Deltavirus</i>) <i>Milan J Sonneveld and Suzanne van Meer</i>	217
Studying Population Genetic Processes in Viruses: From Drug-Resistance Evolution to Patient Infection Dynamics <i>Jeffrey D Jensen</i>	227

Virus-Based Cancer Therapeutics <i>Roberto Cattaneo and Christine E Engeland</i>	233
Prevention	
Surveillance of Infectious Diseases <i>Norman Noah</i>	247
Preparing for Emerging Zoonotic Viruses <i>Reina S Sikkema and Marion PG Koopmans</i>	256
Use of Immunoglobulins in the Prevention of Viral Infections <i>Leyla Asadi and Giovanni Ferrara</i>	267
Vaccine Production, Safety, and Efficacy <i>Thomas J Brouwers and Bernard AM Van der Zeijst</i>	281
Vaccines Against Viral Gastroenteritis <i>Scott Grytdal, Tyler P Chavers, Claire P Mattison, Jacqueline E Tate, and Aron J Hall</i>	289
Human Papillomavirus (HPV) Vaccines and Their Impact <i>Jade Pattyn, Pierre Van Damme, and Alex Vorsters</i>	295
Influenza Vaccination <i>Topi Turunen</i>	300
Polio Eradication <i>M Steven Oberste, Cara C Burns, and Jennifer L Konopka-Anstadt</i>	310
Subject Index	315

FUNDAMENTALS OF VIROLOGY

A Brief History of Virology

David J Rowlands, University of Leeds, Leeds, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Virology is a relatively young scientific discipline, having only come into its own in the late 1900s and the early decades of the 20th century. However, it has had an enormous influence on most aspects of biological science and continues to do so until this day. The importance of virology in the evolution of the life sciences is reflected by the number of Nobel Prizes that have been awarded for work of direct or indirect impact on the field ([Table 1](#)). The study of viruses has been integral to the development of our understanding of the chemical and physical bases of life, the underlying principles of genetics, the immense complexity of host defense mechanisms and the power viruses have to influence even global phenomena such as the climate.

Viruses have profoundly affected humanity for millennia. Changes in the size and structure of human populations have almost certainly influenced our relationships with viruses. At the dawn of humanity small family based groups of hunter/gatherers were probably insufficient for the maintenance of many highly transmissible viruses due to rapid exhaustion of the susceptible host population. Exceptions may be viruses that can initiate persistent infections, e.g., tumor viruses (papillomaviruses, retroviruses, herpesviruses). It is likely that expansion of the human population and development of societal lifestyles provided suitable sized host pools to allow continued transmission of viral pathogens across more integrated populations. Modern sequence and bioinformatics based epidemiological studies have revealed the frequency with which viruses cross between species, allowing them to continuously sample the potential host pool.

Given the limitation of space, this introductory article can provide only a brief overview of the rich history of the origins and evolution of virology. Unfortunately, there is not enough space to acknowledge the many individuals who have contributed to the science of virology. It will necessarily paraphrase and précis many other treatise that cover the subject in greater depth. A list of selected historical highlights during the development of the science of virology is given in [Table 2](#).

Previrology

Diseases of plants, animals and humans had been observed for centuries but it was not until the latter half of the 19th century that progress was made in understanding their underlying causes. Transmissible infections had been observed since times of antiquity, as had the realization that recovery from disease could result in resistance against reinfection. This is well illustrated by the story of smallpox, control of which was initially attempted by the ancient and dangerous practice of variolation and finally by the more controlled process of vaccination developed in the 18th century by Edward Jenner. The germ theory, as proposed by Pasteur, contributed enormously to the concept that diseases were caused by definable agents as opposed to “acts of god”, poisons and miasmas. Demonstration that infectious agents could be grown *in vitro* led Koch to propose the defining characteristics necessary to confirm the causative agent of an infectious disease. These are the famous Koch’s postulates that state:

- (1) The microorganism must be found in abundance in all organisms suffering from the disease, but should not be found in healthy organisms.
- (2) The microorganism must be isolated from a diseased organism and grown in pure culture.
- (3) The cultured microorganism should cause disease when introduced into a healthy organism.
- (4) The microorganism must be re-isolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.

These criteria were useful for defining the etiological agents of diseases caused by cultivatable bacteria and fungi but it became apparent that for many diseases all of the requirements could not be fulfilled, such as the ability to cultivate the causative agents *in vitro*. In addition, for an increasing number of diseases the causative agents were shown not to be removed by filtration through porcelain filters, such as the Chamberland filter candle, and were therefore significantly smaller than bacteria.

Dawn of Virology

The first demonstration that disease could be transmitted by agents other than bacteria was made in 1882 by Mayer, who showed that mechanical transmission of soluble extracts from tobacco plants affected by a mosaic disease to fresh plants resulted in their developing the disease. However, no causative agent for the disease could be cultured *in vitro*. Subsequently, Ivanovsky in 1892 showed that the ability of such extracts to transmit disease was not removed by passing through a filter candle capable of removing known bacteria. This observation was repeated by Beijerinck in 1898 who also made the key finding that the disease causing agent could be recovered following dilution and was, therefore, capable of reproduction and differed

Table 1 Nobel prizes awarded for work of relevance to virology

<i>Year</i>	<i>Laureate</i>	<i>Category</i>	<i>Subject</i>
1905	Robert Koch	Physiology or Medicine	Discoveries relating to Tuberculosis
1951	Max Theiler	Physiology or Medicine	Yellow Fever
1954	John Enders Thomas Weller Frederick Robbins	Physiology or Medicine	Growth of poliovirus in tissues culture
1958	Joshua Lederberg	Physiology or Medicine	DNA transduction in bacteria
1965	Francois Jacob Andre Lwoff Jacques Monod	Physiology or Medicine	Genetic control of enzyme and viral synthesis
1966	Peyton Rous	Physiology or Medicine	Discovery of tumor viruses
1969	Max Delbruck Alfred Hershey Salvador Luria	Physiology or Medicine	Replication mechanisms and genetic structure of viruses
1975	David Baltimore Renato Dulbecco Howard Temin	Physiology or Medicine	Interaction between tumor viruses and cellular genetic material
1976	Baruch Blumberg Carleton Gajdusek	Physiology or Medicine	Discoveries of new mechanisms for origin and dissemination of infectious diseases
1978	Werner Arber Daniel Nathans Hamilton Smith	Physiology or Medicine	Discovery of restriction enzymes and their application to molecular genetics
1982	Aaron Klug	Chemistry	Crystallographic electron microscopy and elucidation of RNA/protein complexes
1984	Niels Jerne Georges Kohler Cesar Milstein	Physiology or Medicine	Discoveries on the control of the immune system and the production of monoclonal antibodies
1986	Ernst Ruska Gerd Binnig Heinrich Rohrer	Physics	Development of electron and scanning tunneling microscopes
1988	James Black Gertrude Elion George Hitchings	Physiology or Medicine	Drug discovery
1989	Michael Bishop Harold Varmus	Physiology or Medicine	Cellular origin of retroviral oncogenes
1993	Richard Roberts Phillip Sharp	Physiology or Medicine	Discovery of split genes
1993	Kary Mullis	Chemistry	Invention of PCR
1996	Peter Doherty Rolf Zinkernagel	Physiology or Medicine	Cell mediated immunity
2008	Harald zur Hausen	Physiology or Medicine	Human papillomavirus and cervical cancer
2008	Francoise Barre-Sinoussi Luc Montagnier	Physiology or Medicine	Discovery of HIV
2017	Jacques Dubochet Joachim Frank Richard Henderson	Chemistry	Cryo-electron microscopy
2018	George Smith Gregory Winter	Chemistry	Phage display of peptides and antibodies

fundamentally from inert poisons. He was also the first to adopt the term virus, from the Latin for slime or poison. This work is often heralded as the birth of virology, a new branch of biology which would expand in the coming century to have major impacts on the study and understanding of virtually all aspects of the life sciences. Other than being very small, the nature of this new infectious agent was not properly understood for many years and was referred to as “*contagium vivum fluidum*”, referring to its apparently non-particulate or fluid nature.

Table 2 Selected historical highlights in virology

Year	Highlight	Investigators
1892	Filterable infectious agent (Tobacco Mosaic Virus - TMV)	Ivanovsky
1898	Concept of virus (TMV)	Beijerinck
1898	First animal virus (FMDV)	Loeffler & Frosch
1901	First human virus (Yellow Fever Virus)	Reed
1911	First solid tumor virus (Rous sarcoma virus)	Rous
1915	First bacterial viruses	Twort
1917	First bacterial virus plaque assay	D'Herelle
1935	Crystallization of TMV	Stanley
1938	Development of electron microscope	Ruska. <i>et al.</i>
1939	First description of single step bacteriophage growth cycle	Ellis, Delbruck
1952	Demonstration that nucleic acid is determinant of infection	Hershey, Chase
1955	<i>In vitro</i> assembly of TMV	Fraenkel-Conrat, Williams
1940s – 50s	Development of tissue culture for in vitro culture and study of viruses	Enders, <i>et al.</i>
1957	Discovery of interferon	Isaacs, Lindenmann
1959	Development of negative stain EM	Brener, Horne
1962	Principles of virus structure	Caspar, Klug
1950–1970	Origins and expansion of molecular virology	numerous
1970–2000	Origins and expansion of cellular immunity	numerous
1970	Discovery of retrovirus reverse transcriptase	Temin, Baltimore
1976	First RNA virus sequence (MS2)	Fiers
1977	First DNA virus sequences (ϕX174, SV40)	Sanger, Fiers, Weissman
1977	Eradication of Smallpox	WHO
1978	First plant virus structure (tomato bushy stunt virus)	Harrison
1983	Discovery of human immunodeficiency virus (HIV)	Montagnier, Barre-Sinoussi, Gallo
1985	First animal virus structures (rhinovirus, poliovirus)	Rossmann, Hogle
1989	First virus discovery by molecular biology techniques (HCV)	Houghton. <i>et al.</i>
1998	Suppressor RNAi in plants	Vance, Baulcombe
2002	SARS coronavirus zoonotic transfer and epidemic	global
2009	Influenza H1N1 pandemic	global
2011	Eradication of rinderpest	OIE
2014	Ebola epidemic – West Africa	
2019	SARSCoV-2 pandemic	global

Thus the first pathogen to be classified as a virus was the etiological agent of a disease of plants; tobacco mosaic virus (TMV). The first animal virus found to pass through bacteria retaining filters was that responsible for foot-and-mouth disease (FMD). This observation was made by Loeffler and Frosch in 1897 in studies to identify the cause of a devastating epidemic disease of cattle that had been recorded as long ago as 1546 by Hieronymus Fracastorius. The origins of bacterial virology dates back to a similar period when an agent present in waters of the river Ganges which destroyed *Vibrio cholera* cultures was shown by Ernest Hankin in 1889 to pass through bacterial filters. In 1901, Walter Reed showed that the causative agent of yellow fever could also pass through a bacterial filter and was the first human pathogen to be classified as a virus by this criterion. Apart from the ability to pass through filters the physical and chemical natures of these new entities were not understood. An important step forward was the demonstration by d'Herelle in 1917 that the bacterial viruses could be quantified by counting the number of plaques (clear areas of cell destruction) that were formed on bacterial lawns. This graphically demonstrated that the infectious agents were particulate and could replicate.

These early observations defined the basic characteristics of viruses and in the first decades of the 20th century the number and variety of infectious agents belonging to this category of pathogens expanded almost exponentially, a phenomenon that continues to this day. As the range of physical and chemical techniques available to interrogate the structure and function of the viruses expanded, it became apparent that, collectively, they included all conceivable ways of storing and expressing genetic information.

The development of virology in the 20th century can be roughly considered as a series of sequential but often overlapping phases, largely driven by characteristics of the host/virus interactions and by technological advances. The original observations that a number of pathological agents infecting plants, animals, and bacteria shared the properties of being smaller than any known bacteria and being non-cultivable *in vitro* begged the question of their nature. Resolving the fundamental nature and function of viruses occurred during the end of the 19th and early decades of the 20th centuries and thereafter expanded rapidly.

The Physico-Chemical Phase – Plant Viruses

Plant virology played an important role in the early stages of unraveling the chemical and physical properties of viruses. In particular, studies of TMV were instrumental in resolving many of the key features of viruses. Growth to high levels in infected plants, ease of

purification and stability of the virus made it an ideal candidate for detailed study. Chemical and physical studies showed that the infectious agent shared many of the properties of proteins, including the ability to induce and be neutralized by antibodies. More detailed chemical analysis suggested that the virus was composed of 95% protein and 5% nucleic acid. The observation that different symptoms could be induced by different viral isolates indicated their genetic nature. Demonstration of the property of birefringence indicated that the infectious agent was large, in molecular terms, and asymmetrical in shape. These properties were confirmed when purified virus particles were first crystallized and then directly observed in the electron microscope in 1939, the first visualization of a virus particle. X-ray diffraction analysis indicated regular hexagonal arrangement and, subsequently, dissociation and reassembly of virus particles in vitro showed them to be composed of multiple copies of components with the inherent ability to self-assemble.

The Biochemical Phase – Bacterial Viruses

The inherent properties of many plant viruses (e.g., TMV) of high level growth and stability accounts for their important roles in unraveling the basic nature and structure of viruses. Bacteria had also been shown to be susceptible to infectious agents with similar properties to viruses. Frederick Twort is accredited with the discovery of bacterial viruses in 1915 when he described the “glassy transformation” of bacterial cultures by a filterable and dilutable agent. Shortly after, d’Horelle made similar observations and introduced the term bacteriophage (bacteria eater). In addition to again demonstrating the filtration resistance of viruses, he developed the first method of quantifying the infectious particles by counting the clear plaques they produced on bacteria culture plates. These studies heralded a number of important advances including the demonstration that bacteriophages attached to their hosts via specific receptor interactions. The phenomenon of lysogeny was also described, in which the bacteriophage genome could remain quiescent within the host and remain associated during bacterial replication but to re-emerge as a lytic infection under appropriate conditions. The dimensions and chemical composition (e.g., DNA, as opposed to RNA in TMV) of bacteriophages were also studied throughout this period.

Two great advantages of bacteriophages for study are the speed with which replication occurs (minutes as opposed to hours or days in other systems) and the ability to infect cultures simultaneously and so study one-step growth kinetics. These properties opened the doors to investigating the very nature of genetic inheritance and expression. The study of bacteriophages in the 1940s and 50s laid the foundations for understanding how viruses replicate which, in turn, established many of the most important aspects of cell function. Many of these ground-breaking discoveries were made by members of the famous “phage group” who first convened at Cold Spring Harbor in 1947. Seminal discoveries made by some of the giants in the field, such as Cohen, Chargaff and Stanley included the demonstration of massive and synchronised biochemical and macromolecular changes in the infected cell during the eclipse phase between initiation of infection and the assembly of new phage particles. Dramatic changes in RNA and DNA synthesis in the infected cells occurred during a period of a few minutes. In addition, it was found that a cellular enzyme was inhibited, thus illustrating the profound changes in cellular function as a consequence of infection. These studies led ultimately to the concept of messenger RNA being the intermediary for gene expression.

A key experiment conducted in 1952 by Hershey and Chase demonstrated that DNA was the molecule responsible for initiating the infectious cycle and is the cornerstone of genetic inheritance. They infected bacteria with phage particles differentially radioactively labeled in protein and DNA and showed that while the protein component remained at the surface and could be removed mechanically after infection had occurred, the DNA was internalized into the bacterial cells. That DNA is the repository of genetic information was further confirmed by Wyatt and Cohen who showed that phage infected bacteria expressed a novel enzyme activity not found in uninfected cells and which therefore must be encoded by the phage genome.

In 1949, Lwoff showed that the genomes of certain DNA phages were incorporated into host cell genomes to induce the lysogenic state, whereby the phage genome is replicated as a component of the host DNA but can be released as an independently replicating lytic phage following environmental stimuli. These phenomena clearly have parallels in the ability of many eukaryotic viruses to integrate into host genomes to initiate oncogenic consequences. The phenomenon of lysogenic integration of phage genome into the host chromosome allowed for genetic mapping of the phage genome. This was achieved by synchronized interruption of bacterial mating to control the proportion of the phage genome transferred to the recipient bacterium during conjugation.

Study of T4 (a lytic phage) and λ (a temperate phage) in particular established the fundamental nature of genetic structure and inheritance. However, an ever expanding number of bacteriophages were discovered which did not conform to the paradigm of double stranded DNA being the sole mechanism for storage and expression of genetic information. Bacteriophages with single stranded DNA and even RNA genomes were discovered, thus exposing the array of alternative lifestyles adopted by viruses. A theme that was further explored by the study of animal viruses.

The Cell Biology Phase – Animal Viruses

The identification and investigation of animal viruses was, and still is, dominated by their involvement in pathogenesis. However, their study has contributed enormously to our basic understanding of the molecular basis of eukaryotic cell function. Although mammalian viral diseases had been recognized for millennia, e.g., smallpox, polio, rabies, foot-and-mouth disease, it was not until the end of the 19th and beginning of the 20th century that the nature of the causative agents was rigorously investigated. Careful use of size specific filters together with investigation of the susceptibility of the infectious agents to exposure to physical and

chemical challenges help to define their basic characteristics. Many important features of the ever increasing array of viruses were demonstrated in these earlier studies. As a few examples, seminal investigations by Walter Reed showed that yellow fever virus (YFV) is transmitted by an insect vector (mosquito), different cell susceptibilities were demonstrated as were the varying outcomes of infection by different viruses, for instance chronic or persistent infection, cell destruction or cell proliferation and oncogenesis.

A major factor that revolutionized the study of animal viruses was the development of in vitro culture of living tissues and cells. Initially primary isolates of cells from susceptible tissues could be cultured for limited periods but a real breakthrough came in 1951 with the selection of the human carcinoma cell line, HeLa, which could be cultured and amplified indefinitely. Subsequently, many other cells lines have been developed and successfully used to facilitate virus research. The advent of cell culture techniques was not only of benefit to fundamental virology research, it also paved the way for vaccine development. The first vaccine against polio was developed by Jonas Salk in 1952 and was made possible by the ability to grow industrial scale quantities of HeLa cells. This massive achievement was made possible by a huge publically funded research project (the March of Dimes) and resulted in the elimination of the annual epidemics of paralytic poliomyelitis that had become the scourge of summer months in the developed world.

The study of viral growth cycles resulted in the discovery of many features of cell biology which underpin our current understanding of cellular functions. For example, understanding the control of gene expression and the role of enhancers evolved through the study of SV40 in the early 1980s. Study of the perturbation of normal cellular replication control mechanisms following integration of viral genes were important for understanding oncogenic transformation. Research on different viruses during the 1980s and 90s provided details of mRNA expression and processing such as mRNA splicing (adenovirus), polyadenylation (pox viruses) and mRNA capping (reoviruses).

Viral Diversity

As more viruses were discovered and characterized the remarkable range of their genome organization became apparent and in 1971 Baltimore proposed a system of classification based on the fundamental nature of viral genomes:

- (1) Double stranded DNA genomes, as in normal cellular organisms.
- (2) Single stranded DNA genomes with +ve (mRNA) sense polarity.
- (3) Double stranded RNA genomes.
- (4) Single stranded RNA genomes with +ve strand polarity.
- (5) Single stranded RNA genomes with -ve strand polarity.
- (6) Single stranded RNA genomes with conversion to DNA during the replication cycle.

In addition to this remarkable array of molecular mechanisms adopted by viruses to store and express their genetic information, genomes were found to be contained within a single molecule or divided into multiple segments, especially in viruses belonging to groups (3), (4) and (5).

As well as this diversity of genetic structure, virus particles were found to display an array of morphologically distinct structural types of widely different complexities. Viral genomes were found to be encapsidated in either of two structural types. The first virus for which structural data was determined (TMV) showed regular helical symmetry, occurring as elongated rods, the dimensions of which were dictated by the length of the genomic RNA. In 1955 it was shown by Fraenkel-Conrat and Williams that separated protein subunits and RNA molecules would reassemble into infectious particles indicating that the structural rules dictating self-assembly were present entirely within the viral components alone. The other major structural theme was encapsidation of the genome within a protein shell of cubic (icosahedral) symmetry. The dimensions of icosahedral nucleocapsids could be varied to accommodate different sized genomes by "simply" varying the number of protein subunits. The principles of icosahedral virus symmetry were determined by Caspar and Klug in 1962. The number of sub-units present in a $T = 1$ icosahedral particle is 60, in which all subunits are in chemically identical environments. For larger icosahedral particles with greater numbers of identical subunits they cannot occupy chemically identical environments, which led to the important concept of quasi-equivalence in which the relationships between different protein subunits within the structure are similar, but not identical. Although these basic "design features" were found to apply generally across the vast spectrum of viruses many often complex modifications and additions to the basic structural motifs were discovered. For example, DNA phages were found to have complex tail structures involved in attachment to the bacterial cell and delivery (injection) of the viral genome. In addition to the proteins directly required for viral assembly and structure it was shown that many viruses obligatorily include enzymes, such as polymerases necessary for transcription of -ve strand RNA genomes into the +ve form necessary for protein translation and the initiation of infection.

Many viruses were found to be wrapped in one (or more) lipid envelopes bearing membrane anchored glycoproteins associated with cell binding and entry through membrane fusion. The presence or absence of an outer membrane envelope was a crucial determinant of the sensitivity of viruses to membrane destroying agents, such as detergents.

The full spectrum of viral diversity has expanded beyond imagination since the identification of TMV and FMDV c.130 years ago. The known virome ranges from viroids, small circular single-stranded replicating RNA pathogens of plants discovered in 1971, to relatively gigantic examples such as the mimiviruses discovered in 2003. Viruses are the most abundant organisms on earth – it is estimated that the world's oceans contain some 10³⁰ viruses, and their influence can be seen globally. They can influence the distribution of phytoplankton and consequently nutrients leading to global changes in atmospheric gasses.

The Structural Phase

As is the case with all scientific disciplines, progress in the study of viruses was dependent on and driven by technological innovation. The dawn of virology as an independent discipline was heralded by the apparently simple procedure of size discrimination by selective filtration, e.g. the initial studies of TMV, FMDV and YFV referred to above.

The invention of the electron microscope in the late 1930s enabled visualization of the TMV particle, for the first time in 1939. EM images confirmed the expectation derived from X-ray analysis of the crystals of the virus, showing that the particles were rod shaped and of defined diameter. Further development and refinement of the electron microscope and improvements in sample preparation and treatment heralded a rich period in the 1970s and 1980s of virus identification and morphological classification. During this period electron microscopy played a major role in the identification and epidemiology of viral infections. In fact, it still has an important role in the identification of viruses which cannot be cultured easily *in vitro*, such as many enteric viruses, and was used to identify coronaviruses as the agents responsible for the recent zoonotic transfer and pandemic spread of SARS-CoV-2.

X-ray crystallography developed steadily during the 20th century and for several decades it was the only technique able to generate structural information on viruses at near atomic resolution. The first icosahedral virus for which this level of structure determination was achieved was tomato bushy stunt virus in 1978 by Harrison and colleagues. The first animal virus structures to be determined at similar resolution were the picornaviruses human rhinovirus and poliovirus in 1985 by Rossmann and Hogle, respectively. The size and complexity of viruses that could be examined in this detail continued to increase as the technology developed. For example, the structure of blue tongue virus, which has $20\times$ greater diameter than picornaviruses, was resolved in the late 1990s by Stuart *et al.* In favorable circumstances, viruses could be co-crystallized in complex with their cell receptors, antibodies or antiviral inhibitors, with important consequences for drug and vaccine design and the understanding of virus attachment to host cells. Despite the tremendous advances in understanding virus structure made possible by X-ray crystallography, the method is restricted by some unavoidable constraints. For example, only structures which can be crystallized are available for study, which eliminates pleomorphic structures typical of many enveloped viruses. Secondly, structures of virus particles could only be analyzed when regularly ordered in the crystalline state, making the study of dynamic and irregular intermediate particles (for example during cell entry) impossible. Electron microscopy had the potential to overcome some of these restrictions but, until recently, did not have the resolving power to study viruses at the atomic level. Enormous advances in cryo-electron microscopy in recent years have overcome these restrictions with the result that the technique has largely superseded X-ray crystallography as the principle method for high resolution structure determination. Tomographic cryo-electron microscopy has the added advantage of facilitating analysis of viruses and virus-induced structures within cells. It is adding greatly to our understanding of viral functions such as the infection process, virus replication, cellular remodeling, and the process of assembly and maturation of new viral particles.

The Sequencing Phase

The 1970s saw the start of the nucleotide sequencing revolution which has come to underpin many areas of biology. The first RNA virus to be sequenced was the bacteriophage MS2 by Walter Fiers group in 1976. This was a major achievement and depended on the isolation and analysis of fragments of radioactively labeled genome generated by specific RNase digestion. The generation of strain specific fragment profiles by RNase T1 has also been used as a sensitive method for discriminating epidemic virus strains. Development of the di-deoxynucleotide chain termination nucleotide sequencing technology by Fred Sanger and colleagues allowed the determination of the first complete genome sequence of a DNA virus, the bacteriophage ϕ X174, in 1977. This method greatly simplified nucleotide sequencing and when coupled with retroviral reverse transcriptase it was readily applicable to RNA viruses. These technical improvements, together with plasmid cloning of DNA fragments produced by restriction enzyme digestion, greatly accelerated and simplified determination of viral nucleic acid sequences. The invention of the polymerase chain reaction (PCR) in 1983 by Mullis added another strong arm to the arsenal of techniques available for understanding and manipulating viral genomes and has become of great importance in viral diagnostics. Finally, development of Next Generation Sequencing technology, in which parallel sequencing of millions of random DNA fragments coupled with the computational methods to combine the reads to obtain the original DNA sequence has increased sequencing power such that complete eukaryotic genomes can be determined in a single day. The sophistication of these technical advances is such that sequence information derived from random genome fragments representing mixtures of multiple organisms can be sorted into individual species. The application of these methodologies has demonstrated the enormous diversity of viruses in environmental samples, such as sea water, without the requirement for individual virus isolation or characterization.

Cloning, *in vitro* manipulation and sequencing of DNA have enabled a number of major developments in virology. The first demonstration that a virus could be recovered from cloned DNA was the bacteriophage Q β in 1978 and in 1981 poliovirus was recovered from HeLa cells transfected with a bacterial plasmid containing a cloned cDNA copy of the viral genome. Subsequently, poliovirus was recovered from a chemically synthesized DNA copy of the genome. In subsequent years *in vitro* manipulation of viral genomes has been used extensively for a variety of reasons. For example, recombinant adenoviruses have been constructed as potential gene therapy agents to replace faulty genes in conditions such as cystic fibrosis. Similarly, recombinant viruses such as insect baculoviruses are widely used to express large quantities of protein products, such as the coat proteins of human papilloma viruses for use as vaccines. DNA manipulation techniques have also been used to develop recombinant viruses for use as live virus vaccines. A notable recent example is a vaccine derived from an attenuated vesicular stomatitis virus which has been manipulated to express the

surface glycoprotein of ebola virus. This novel vaccine was developed and approved for use in record time to join other recombinant ebola vaccines based on adenovirus and vaccinia virus to combat the devastating epidemic of ebola in West Africa in 2014–16.

Pathogenicity and Host Defense Mechanisms

As intracellular parasites, viruses encounter a plethora of host defense mechanisms. The phenomenon of *E. coli* strain specific restriction of phage λ growth was discovered in the 1950s in the laboratories of Luria, Weigle and Bertani; in the 1960s, the endonuclease enzymes responsible for this bacterial immunity were described. Subsequently many thousands of restriction enzymes with different sequence specificities were discovered and became invaluable tools in the explosion of genetic analysis and manipulation that followed. Restriction enzymes are the bacterial equivalent to the innate or non-adaptive responses found in eukaryotes. A prokaryotic adaptive immune response system (CRISPR) was first described in 2005 following two decades of analyses of bacterial genomes including repeat sequence regions. In 2012, modifications of the CRISPR Cas9 system were reported which facilitated the defined manipulation by site specific insertion of DNA sequences into genomes of any organism. This system, derived from bacterial antiviral adaptive immune mechanisms, is now widely used in many areas of biology from the development of crop plants to the manipulation of mosquito genomes.

Equivalent non-adaptive (innate) and adaptive antiviral protective immune systems were also discovered in plants. Receptors able to recognize PAMPs (pathogen associated molecular patterns) were discovered in the 2000s and signaling via these receptor mechanisms induces the expression of a number of genes which affect an antiviral state in the infected cells. Plants were also shown to develop adaptive responses to viral infections by the production of short anti-viral RNA molecules targeted to and able to silence invading viral RNAs.

The ability of humans and other animals to acquire immunity to viruses was well known from the smallpox and rabies vaccination work of Jenner and Pasteur in the 18th and 19th centuries, respectively. Acquired immunity is associated with factors present in serum and can be transferred to naïve animals, as was demonstrated in the late 19th century by Behring and Kitasato. Work in subsequent decades identified antibodies as the serum components responsible for protection, culminating in the resolution of antibody structure in the 1960s and 70s by Edelman, Porter and others. Another milestone was achieved in the 1970s by Milstein *et al.* with the development of techniques to isolate and produce unlimited quantities of monoclonal antibodies. These reagents have proved invaluable in many fields of virology including therapy and diagnostics.

The complex roles of T cells in the evolution of both humoral and cell mediated adaptive antiviral responses were elucidated during the 1970s and 80s. The Nobel Prize winning work by Zinkernagel and Doherty on lymphocytic choriomeningitis virus in mice showed the importance of major histocompatibility (MHC) molecules in the induction and selection of specific anti-viral cytotoxic T cells. Viruses have been key to unraveling the complex interplay between the innate and adaptive arms of the immune system. The discovery of interferon, a soluble factor produced by infected cells and able to induce an antiviral state in naïve cells, by Alick Isaacs in 1957 opened the door to the eventual unraveling of the complicated controls and operations of innate immunity in animal systems.

A number of receptors able to detect foreign (e.g., viral) molecules (PAMPs) were found to induce expression of anti-viral defense proteins via intracellular signaling pathways. The complex interplay between viruses and their mammalian hosts has been a rich field of investigation, especially in the latter decades of the 20th century and into the 21st. Viruses have been found to have evolved numerous mechanisms of avoidance of both innate and adaptive immune responses. It had long been known that escape from humoral, antibody mediated, host defense is acquired largely by antigenic mutation. Influenza provided a prime example with the well-studied phenomena of antigenic drift through sequential acquisition of point mutations and antigenic shift by genome segment exchange. More recently a variety of mechanisms for abrogating and bypassing both adaptive and innate immune defenses have been discovered. For instance, herpes viruses can inhibit the presentation of MHC molecules at the plasma membrane, thus rendering infected cells invisible to cytotoxic T cells. Many viruses, such as HCV, can disrupt the intracellular signaling pathways involved in the induction of IFN; poxviruses produce many “molecular decoys” such as soluble equivalents of host cell-surface linked IFN receptors.

Epidemiology

Diagnostic methods underpin epidemiological studies and progressive improvements in the speed and accuracy of diagnostic techniques have been key to understanding and controlling the spread of viral diseases. In the earlier and mid decades of the 20th century clinical observation was supported by morphological characterization using EM and by immunofluorescence methods and an increased use of primary tissue culture and eventually immortal cell lines.

The development of serological techniques for the detection of viruses, such as complement fixation tests in the 1920s and haemagglutination tests in the 1940s greatly increased the speed and accuracy of viral diagnostic procedures. In the 1970s the enzyme-linked immunosorbent assay (ELISA) largely superseded other serological techniques for rapid laboratory based investigations. The invention of the polymerase chain reaction (PCR) in 1983, together with advances in nucleic acid sequencing techniques again greatly advanced accuracy and sensitivity of diagnostic procedures. Development of deep-sequencing techniques in the 21st century have facilitated epidemiological investigations at unprecedented levels of accuracy, allowing detailed investigations of the spread of viral infections during epidemics.

Table 3 Selected milestones in vaccine development

Year	Virus	Vaccine type	Principle investigators
1798	Smallpox	Attenuated	Jenner
1885	Rabies	Attenuated/inactivated	Pasteur
1938	Yellow fever	Attenuated	Theiler
1945	Influenza	Inactivated	Francis, Salk. <i>et al.</i>
1947	Foot-and-mouth disease	Inactivated (primary tissue)	Fraenkel
1952	Polio	Inactivated	Salk
1957	Polio	Attenuated	Sabin, Koprowski
1962	Foot-and-mouth disease	Inactivated (cell culture)	Mowat, Capstick. <i>et al.</i>
1971	Measles, mumps, rubella	Attenuated	Hilleman
1984	Hepatitis B	Recombinant (VLPs)	Valenzuela
1994	Hepatitis A	Inactivated	Innis
2006	Human papilloma	Recombinant (VLPs)	Cervarix and Gardasil
2019	Ebola	Recombinant (vectored)	Feldmann

Viral Control

Vaccines

Antiviral vaccines have alleviated the burden of many viral diseases of both humans and livestock and have contributed to the total eradication of two important disease; an enormous achievement for mankind. A summary of some of the historical highlights of vaccine development are shown in [Table 3](#).

The origins of vaccination predate the birth of virology as a defined discipline. Variolation was developed in China in the 15th century to protect against the devastating disease of smallpox. It involved infecting the patient by rubbing infectious material from pox lesions onto scarified skin to induce a less severe disease than when the infection was acquired naturally. In 1796 Edward Jenner adapted the same technique but used pustular material from a milkmaid thought to have acquired the infection from a cow. The cowpox was, in effect, a naturally attenuated virus in humans. Although farmers were aware of this method to protect against smallpox, Jenner formally tested the effectiveness of the procedure by challenging a vaccinated boy with virulent smallpox. His insight and dedication led ultimately to the declaration of the eradication of smallpox in 2004, the first disease to be eliminated. The next viral immunization procedure was developed almost a century later in 1885 when Pasteur vaccinated a boy who had been bitten by a rabid dog with material derived from the brain of a rabbit infected with rabies. The material had been “weakened” by drying and was, in effect, a combined attenuated and inactivated antiviral vaccine.

Developments in culture of primary cells *in vitro* in the 1930s and 40s facilitated isolation of a number of important viruses. Yellow fever vaccine was the first to be derived by multiple sequential passage until a non-pathogenic strain had been selected. Some killed vaccines were also produced during these years; for example, influenza was grown in embryonated eggs and chemically inactivated with formaldehyde. FMDV vaccines were produced by growth in epithelial tissue stripped from the tongues of slaughtered cattle and inactivated with formaldehyde. The difficulties and disadvantages of culturing cells from primary tissues were overcome by the selection of immortal cell lines. The first was the HeLa cell line, named after Henrietta Lacks, a cervical cancer patient from whom the cells were obtained in 1951, which enabled the development of the Salk inactivated vaccine against poliovirus. Tissue culture enabled the development and production of virus vaccines in sufficient quantities to conduct mass vaccination campaigns. Many live attenuated or chemically inactivated vaccines were developed using “classical” *in vitro* cell culture in the 1960s–1980s including, for example, measles, mumps, rubella, rabies and hepatitis A. Controlled use of vaccines has resulted in the global eradication of two viral diseases; smallpox in 1980 and the rinderpest in 2011. Rinderpest virus caused periodic devastating plagues in cattle and other ungulates for millennia in Asia, Europe and more recently Africa, and was eliminated using an attenuated vaccine developed by Plowright in the 1960s.

Molecular techniques have been applied in the later decades of the 20th century and early years of the 21st to explore new methods in vaccine development. These include recombinant expression of viral proteins with vaccine potential and in the 1980s the first recombinant viral vaccine was produced by expression of hepatitis B virus-like particles (VLPs) in yeast. More recently VLP vaccines of human papillomaviruses produced in yeast or in insect cells using recombinant baculovirus have been licensed.

Genetic manipulation has also been used to insert viral antigenic proteins into the genomes of non-pathogenic vector viruses such as vaccinia virus, adenoviruses and vesicular stomatitis virus (VSV). Recombinant adenovirus expressing the spike protein of the pandemic SARS-CoV-2 virus is being tested as a potential vaccine against COVID-19 and recombinant VSV expressing the envelope protein of ebola virus has recently been licensed for human use.

Drugs

Antiviral drug development has been slow and difficult and there are still relatively few viral disease for which effective drugs are available. A summary of some of the highlights of antiviral drug development are shown in [Table 4](#). The relative simplicity of viral

Table 4 Selected examples of anti-viral drug development

Year	Drug	Mode of action	Target virus
1957	interferon	Developed as a registered drug in 1980s and 1990s induces antiviral state in cells	HCV, HBV and others
1963	idoxuridine	First approved antiviral drug inhibits viral polymerase	HSV
1968	amantadine	Viral ion channel blocker	influenza
1972	ribavirin	Inhibits viral polymerase	HCV, RSV, etc.
1979	acyclovir	Inhibits viral polymerase	Herpes viruses
1987	zidovudine	Inhibits reverse transcriptase (RT)	HIV
1995	saquinavir	Viral protease inhibitor	HIV
1996	RSV-IGIV	Neutralizing antibodies	RSV
1996	nevirapine	Non-nucleoside RT inhibitor	HIV
1996	HAART	Combined drug formulations	HIV
1999	zanamivir/oseltamivir	Neuraminidase inhibitors	influenza
2011	telaprevir	Viral protease inhibitor	HCV
2014	lepidasvir	NS5A (protease) inhibitor	HCV
2013	sofosbuvir	NS5B (polymerase) inhibitor	HCV
2016	Combination DAA therapy	Triple combinations of direct acting antiviral drugs	HCV

genomes means that there are few virus specific targets for inhibitory compounds. The first drug to be licensed was amantidine in 1964, which inhibits influenza by blocking a viral ion channel. Amantidine suffered from a drawback common to many antiviral drugs in that resistant viruses arose by mutation very rapidly. The anti-herpes drug acyclovir, developed in 1971, was successful as the acquisition of resistance was rare.

The emergence of HIV in the early 1980s presented a new therapeutic challenge as the disease AIDS is progressive and incurable and vaccine development was (and still is) unsuccessful. A repurposed drug, AZT, effectively inhibited the viral reverse transcriptase and was useful until the selection of resistant variants of the virus abrogated its benefits. The problem of drug resistance was eventually addressed by the inclusion of other compounds, inhibitors of the viral protease for example. Acquisition of resistance simultaneously to multiple drugs is statistically rare and usually results in reduced viral fitness.

In 1989 the agent responsible for non-A-non-B hepatitis was identified by Houghton *et al.* This was the first virus to be discovered using molecular biology techniques and was termed hepatitis C virus (HCV). Interferon was first used to treat this chronic infection and was partially effective. Subsequently, intense drug discovery programmes involving many approaches, including structure-guided drug design methods, have culminated in current treatment regimens involving cocktails of drugs targeting different viral targets. This approach, as with HIV, overcomes the rapid selection of variants resistant to individual compounds and, in contrast to HIV, is curative in most cases.

Virus Emergence and Human Activities

The history of virology has been punctuated by eruptions of epidemic and pandemic diseases. These disturbances of virus/host equilibria arise as consequences of three main factors:

(1) Changes in the host conditions and environmental changes.

The evolution of human society and the global dominance of our species has impacted greatly on the interactions with pathogens, including viruses, with ourselves, our domestic animals and food crops. The change from a small group, hunter/gatherer lifestyle to static agricultural economy altered the density distribution of humans, crop plants and livestock alike. Clearly changes in the dynamics of pathogen/host interactions reflected these human driven changes. Expansion of the human population and increasing concentration in urban crowded conditions greatly increase the potential for pathogen spread. These factors almost certainly explain many of the devastating epidemics that have plagued humans since time immemorial. Striking recent examples of the effects of changes in animal husbandry and high density factory farming are outbreaks of “bird flu” that can kill thousands of birds over-night in high density chicken sheds.

Other examples of severe viral epidemics resulting from human activities are the depopulation of the Americas due to the introduction of disease such as smallpox, measles and influenza by European colonizers in the 15th and 16th centuries, decimation of the cattle populations (wild and domesticated) by rinderpest in the 1980s and control of European and Australian rabbit populations following the introduction of the myxomatosis virus from South America in the 1950s.

(2) Mutations of commonly occurring viruses with consequent major effects on virulence and transmission.

Not only do viruses replicate to produce vast numbers of progeny, many use error prone RNA dependent polymerases (e.g., picornaviruses, myxoviruses and retroviruses). As a consequence, virus populations typically comprise a complex mixture of variant mutant genomes referred to as quasi-species. Occasionally mutant viruses capable of exploring a new evolutionary avenue may be selected from the quasi-species pool to initiate a new disease phenotype. Influenza is a classic example of a virus that has caused frequent epidemics due to accumulation of mutations in its surface proteins sufficient to evade host

population immunity through antigenic drift and pandemics following genome segment exchange to produce antigenic shift. This occurred on several occasions throughout the history of virology and has been responsible for some of the most devastating pandemics of the past century; the most notorious being the 1918 “Spanish Flu” pandemic which affected all parts of the globe in a short space of time and is thought to have resulted in 50–100 million deaths.

A further example of a major change in host range associated with “minor” mutation occurred in the 1970s. A new and highly pathogenic disease appeared in domestic dogs and was transmitted across the globe in a short period of time. Sequencing of the “new” virus showed that it differed in just one or two amino acids in the structural protein region from feline parvovirus, it most likely appeared as a new and devastating disease in dogs as a consequence of a mutation which affected the host range of the virus.

(3) Transmission of viruses between species; emerging viruses.

The continual expansion of the human population and encroachment into formally underpopulated habitats has resulted in frequent epidemics and pandemics of “new” diseases via zoonotic inter-species transfer of viruses. Many of the most damaging viral epidemics and pandemics of human disease in the past century have arisen following the introduction of a novel virus originating from animal sources not commonly encountered by humans.

A well-documented example of zoonotic transfer is the so called “four corners” disease which appeared in the western United States in 1993 in a desert region where the states of Utah, Colorado, New Mexico and Arizona meet. Virologists from CDC were called in to investigate a new and highly pathogenic respiratory syndrome that appeared in the local communities. The causative agent was discovered to be a hantavirus (sin nombre virus) commonly found in a species of mouse, *Peromyscus maniculatus*, in which it causes persistent asymptomatic infections. Inhalation of dust contaminated by infected mouse feces and urine during cleaning procedures resulted in severe respiratory disease in humans. This is a typical example of a virus crossing from a wild species to infect humans as a result of unusual exposure, usually due to habitat encroachment.

Many emerging viruses are highly pathogenic, often causing hemorrhagic diseases, but lacking the ability to transmit between humans sufficiently well to initiate widespread epidemics or pandemics. However, there are a number of examples from the past century of zoonotic transfer of animal viruses which are both highly pathogenic and sufficiently transmissible between humans to initiate extensive epidemics and pandemics. Periodic pandemics of highly pathogenic influenza have been recorded and characterized in the past century, and contemporary accounts of devastating plagues in past centuries can be tentatively ascribed to influenza virus. The segmented structure of the influenza genome makes it especially well adapted to explore new host horizons via antigenic shift following dual infection of an intermediate host with viruses originating from different sources (e.g., bird and human). The outcome of such genetic exchange has been the evolution of new viruses including the genes allowing efficient growth in one host (e.g., human) combined with surface glycoproteins derived from another (e.g., bird). Several influenza pandemics have arisen during the past century and have been responsible for extremely high mortalities. The speed of global transmission of the 1918 “Spanish Flu” was facilitated by increased international travel together with disruption due to war.

In 1981 a new disease was observed in America. This was found to be caused by a new retrovirus, the lentivirus termed acquired human immunodeficiency or AIDS virus. From its early identification in the homosexual community of San Francisco, AIDS has gone on to become a major global pandemic, claiming the lives of an estimated 32 million people. Being a retrovirus, integration of the viral genome into the host cell chromosomes ensures that the infection is maintained lifelong. Despite the development of regimens that can control the disease, but require daily consumption of a cocktail of drugs, many millions of individuals are infected, especially in sub Saharan Africa. Sophisticated epidemiological studies have pointed to the origins of the AIDS epidemic being higher primates (chimpanzees and gorillas) in the tropical forests of central Africa. Transmission of AIDS into the human population almost certainly occurred by exposure through acquisition of bush meat.

Interspecies transfer of viruses with zoonotic potential is increasingly associated with human activities. An increase in live animal markets and in bush meat trade can result in increasing juxtaposition of species that would not make contact under natural conditions. These developments create ideal environments for exchange of infectious agents between species. Novel coronaviruses have emerged as significant human pathogens on three occasions since the beginning of the 21st century. In 2003 a new disease, Severe Acute Respiratory Syndrome or SARS, with a high case fatality rate appeared in China and was shown to be caused by a novel coronavirus. In the space of 6 months the virus was transported across the globe, emphasizing the important role of modern rapid international transport in the spread of infectious diseases, and resulted in 770 deaths. Fortunately, effective disease control measures, together with a limited innate ability of the virus to transmit within the population ensured the rapid elimination of the virus. Molecular epidemiological techniques showed that the causative agent of SARS was closely related to a virus isolated from civet cats in Chinese live animal markets and it is thought that these animals were an intermediate species enabling transmission to humans. SARS virus is also closely related to coronaviruses detected in certain bat species, which may be the primary source. It is becoming increasingly clear that bats are the source for many viruses that have the potential to spread into human (and other species) populations and cause serious disease. Bats seem to tolerate persistent infections, with little or no clinical symptoms, by viruses capable of causing serious disease in other species. The reasons for this are not understood and bats are being subjected to intense molecular, immunological and epidemiological studies to attempt to resolve the conundrum.

The second coronavirus to cause serious disease in humans is the MERS or Middle Eastern Respiratory Syndrome virus which was first identified in 2012. Camels appear to be the intermediate reservoir from which infections with this virus have been transmitted to humans although, again, bats are suspected as the original source. MERS is very pathogenic to humans but is not efficiently transmitted between individuals.

The third incursion of a novel coronavirus into the human population occurred at the end of 2019 and again may have originated from a live animal market in China. Sequence analyses suggest the possibility of the virus being acquired from pangolins or bats, both of which were sold in the market. This virus, termed SARS-CoV-2, causes severe, often fatal respiratory disease as was seen with the original SARS virus. However, in contrast to SARS-1, SARS-CoV-2 is transmitted very efficiently and transmission can occur before infected individuals experience symptoms. The combination of these properties has facilitated rapid spread of the virus globally, resulting in the most important pandemic disease since the Spanish Influenza more than a century ago. Within a few months of its appearance in the human population it has been responsible for hundreds of thousands of deaths and has stretched the health provisions of many countries to near breaking point. Attempts to control spread of the infection have resulted in the immobilization of one third of the global population and has caused havoc to economies worldwide. At the time of writing, this pandemic has still not reached its peak globally. Virologist and other infectious disease scientists have warned world governments for many years of the potential dangers from the introduction of emerging pandemic disease and the necessity for adequate preparation for such an event. Unfortunately, these warnings have gone largely unheeded.

Concluding Remarks

As can be seen from this brief historical introduction, the subject of virology has expanded enormously since its birth as a new scientific discipline a little over a century ago. It is now realised that viruses are by far the most numerous organisms on earth with the ability to influence the health and abundance of all cell-based living entities. They utilize the full range of molecular strategies imaginable to store and express their genetic information. They have proved to be invaluable probes to facilitate unraveling the basic molecular processes applicable to all cellular organisms. Investigation of their pathogenic mechanisms has revealed the enormous complexity of the essential anti-viral defense strategies adopted by all living organisms. The size and diversity of the global virome is expanding continuously as can be seen from the detailed articles comprising this Encyclopedia.

Further Reading

- Calil, A.P., Fontes, E.P.B., 2017. Plant immunity against viruses: Antiviral immune receptors in focus. *Annals of Botany* 119, 711–723.
- De Clerque, E., Li, G., 2016. Approved antiviral drugs over the past 50 years. *Clinical Microbiology Reviews* 29, 695–747.
- Enquist, L.W., 2009. Virology in the 21st century. *Journal of Virology* 83, 5296–5308.
- Enquist, L.W., Racaniello, V.R., 2013. Virology: From contagium fluidum to virome. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia: Lippincott, Williams & Wilkins, pp. 1–20.
- Murphy, F.A., 2020. The Foundations of Virology: Discoverers and Discoveries, Inventors and Inventions, Developers and Technologies, second ed. American Society for Virology. https://liveutmb.sharepoint.com/b:/s/collaboration/webfiles/EZ1BzJBerOxDv7yEWY6GpVgBzlj4wfEbo_ugl12V8zACgg?e=IFTTk9.
- Neurath, A.R., 2008. Immune response to viruses: Antibody mediated immunity. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Elsevier, pp. 56–70.
- Oldstone, M.B.A., 2014. History of virology. In: Schmidt, M. (Ed.), *Encyclopedia of Microbiology*, fourth ed. Elsevier, pp. 608–612.
- Plotkin, S., 2014. History of vaccination. *Proceeding of the National Academy of Sciences* 111, 12283–12287.
- Pumpens, P., 2020. Single-Stranded RNA Phages: From Molecular Biology to Nanotechnology. Boca Raton, FL: CRC Press.
- Rasmussen, A.L., 2015. Probing the viromic frontiers. *mBio* 6, 1–3.
- Roingard, P., 2008. Viral detection by electron microscopy: Past, present and future. *Biology of the Cell* 100, 491–501.
- Suttle, C., 2007. Marine viruses – Major players in the global ecosystem. *Nature Reviews: Microbiology* 5, 801–812.
- Zajac, A.J., Harrington, L.E., 2008. Immune response to viruses: Cell-mediated immunity. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Elsevier, pp. 70–77.

Relevant Websites

- <http://vipfdb.scripps.edu/>
Welcome to VIPERdb.
Scripps Research.
- <https://talk.ictvonline.org/taxonomy/>
Taxonomy.

The Origin of Viruses

Patrick Forterre, Archeal Virology Unit, Institut Pasteur, Paris, France and French National Center for Scientific Research, Institute of Integrative Biology of the Cell, University of Paris-Saclay, Gif sur Yvette, France

Morgan Gaïa, University of Paris-Saclay, Evry, France

© 2021 Published by Elsevier Ltd.

Introduction

Discussing the question of viruses origins requires first to specify the defining characteristics of viruses. For instance, if they are defined by their infectious genomes, then all infectious and parasitic mobile genetic elements (MGEs) can be considered to be viruses, even if they do not produce virions (capsidless viruses, *sensu* Koonin and Dolja), and the origin of viruses becomes equivalent to the origin of selfish infectious replicons. With this definition, one can argue that viruses originated at the very onset of life, since mathematical models suggest that genetic parasites probably emerged as soon as the first replicating systems were established. However, if viruses are defined by their specific mode of reproduction, i.e., the production of virions (viral particles) to disseminate their genomes, then the origin of viruses probably took place at a relatively late stage of life evolution, after the emergence of the ribosome, because even the simplest virion contains at least one protein. We will discuss here the problem of virus origins using this more restricted definition, as it allows us to make more reliable inferences based on what we know about modern viruses.

The Evolutionary Connection Between Viruses and Mobile Genetic Elements

Viruses are evolutionarily related to other MGEs, such as plasmids, transposons, and retrotransposons, sharing with them homologous replication proteins and proteins involved in relationships with their hosts (such as toxin-antitoxin systems in plasmids or other systems involved in the arm race between parasites and their hosts). As other MGEs, they are obligate parasites since they lack the ribosome, and are fully dependent of their cellular hosts for protein synthesis. However, they can be precisely distinguished from other MGEs by the fact that their genomes encode proteins involved in the formation of virions. For instance, the smallest virus, with two genes, and the smallest plasmid, with only one, only differ by the presence of a gene encoding a capsid protein in the viral genome.

Importantly, unlike other MGEs, viruses have the ability to profoundly modify the physiology of their hosts, transforming the infected cell into a machinery to produce virions, also called a virocell (as opposed to uninfected cells relying on their own ribosomes for protein synthesis, the ribocells). This process involves the reorganization of the cell metabolism and can sometimes lead to a symbiotic relationship with the host cell (a ribovirocell, with the virus genome integrated into that of its host) but also to the final destruction of the cellular structure, explaining why many viruses are pathogenic. Considering the virocell as a stage in the virus life cycle, along with a free-particles one, justifies considering viruses as *bona fide* organisms since they translate their genome and transcribe their genes into proteins that can be expressed with metabolic functions for instance. They have been proposedly defined as “capsid-encoding organisms” or “virion-producing organisms”, in order to emphasize the production of virions as the hallmark feature of viruses. The virocell concept (i.e., the cyclic nature of viruses, including as virocells) is not compatible with theories suggesting that viruses originated before cells (virus-first theories) since the virus itself is a cellular organism during one stage of its cycle, completely dependent on a pre-existing ribocell (a cell encoding ribosomes) for its formation. The origin of viruses hence becomes the origin of virocells, or more precisely, how some ribocells were first transformed into ribovirocells (i.e., started producing virions able to infect other ribocells) that finally became lethal for the infected cells, producing genuine virocells.

From the above considerations, the question of virus origins can be divided into three separate aspects: (1) the origin of capsidless parasitic MGEs, (2) the origin of capsid proteins and more generally of virions, (3) the mechanisms behind the association of parasitic MGEs and virions to produce *bona fide* viruses. Importantly, it is unclear if the answers to these questions would be the same for RNA and DNA viruses. Indeed, unlike ribosome-encoding organisms (Archaea, Bacteria, and Eukarya), viruses – and other MGEs – exhibit an astonishing diversity of genome chemistry, structure and arrangement. Here, we will separately address the question of the origin of RNA and DNA viruses, as well as the origin of viruses whose life cycle involves both RNA and DNA, although the typical reproduction mechanism of all these viruses suggests some common, recurrent themes to explain why the viral mode of reproduction was probably selected many times in the course of life evolution.

The Origin of RNA Viruses

It is often assumed that RNA viruses originated first since RNA should have antedated DNA as genetic material in early life evolution. Some authors have speculated that both RNA and DNA might have originated and co-existed in a prebiotic world, but it seems more likely that DNA originated from RNA at a cellular stage and that the modern cell metabolism recapitulated the historical pathway that led from RNA to DNA. DNA is indeed a modified form of RNA and the substrates of DNA synthesis, the

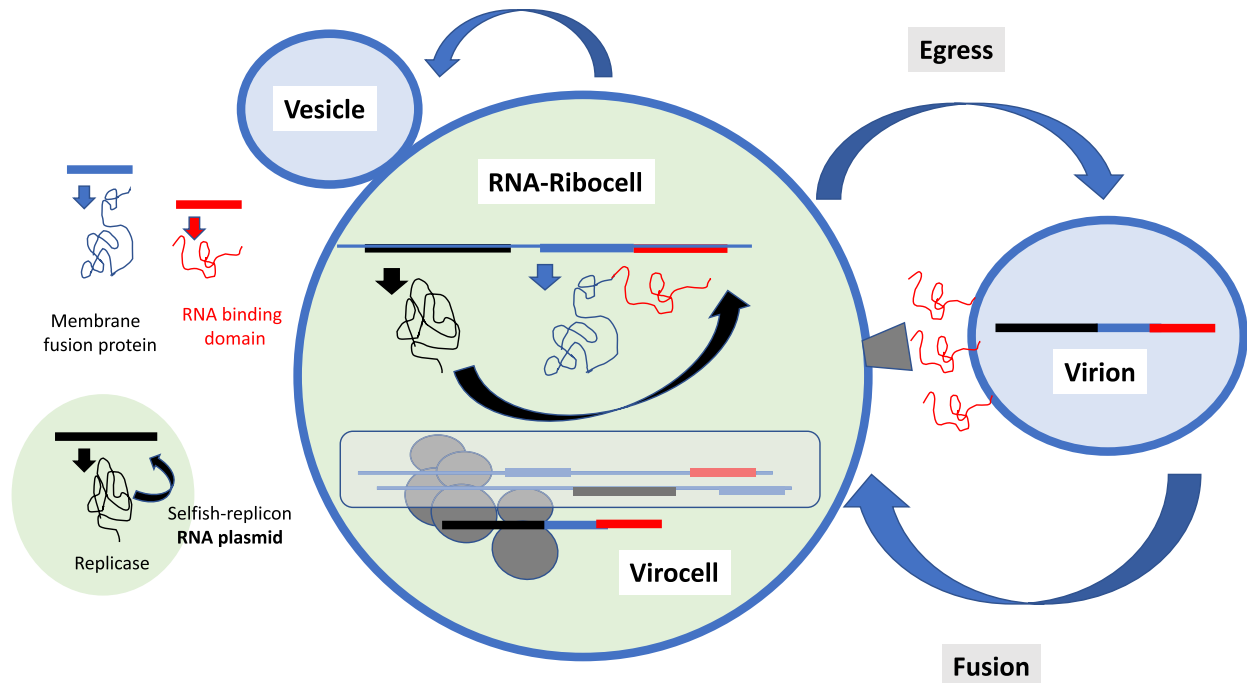


Fig. 1 Schematic hypothesis for the formation of the first RNA viruses in the RNA/protein world. An RNA ribocell (with ribosomes) contains here two linear chromosomes (in the gray area) with genes encoding replicase (black), a membrane fusion protein (blue) and RNA binding protein on different chromosomes. When these three genes are present on the same chromosome (with fusion of the latter two) the chromosome can be transferred by vesiculation from one cell to the other, becoming selfish elements. The vesicle carrying this chromosome (now a viral genome) become a virion whereas the infected cell become a virocell.

deoxynucleotides, are still produced in modern cells from the modification of the RNA substrates, ribonucleotides. It is thus very likely that the historical transition from RNA to DNA took place in *bona fide* cells with RNA genomes. The possible existence of such RNA cells has sometimes been denied because of the intrinsic instability of RNA and the short size of RNA virus genomes. However, RNA repair mechanisms as well as proof-reading endonucleases that can increase RNA replicase fidelity have been discovered, suggesting that ancient RNA cells possibly had RNA genomes longer than those of extant RNA viruses. Moreover, it is unclear if modern RNA viruses are good models to infer the ancient RNA cell physiology because small genomes and unfaithful replication can sometimes be advantageous for viruses. Notably, the RNA cells that gave rise to the first DNA cells probably already had elaborated ribosomes, since the RNA to DNA transition requires a complex cellular metabolism to produce in abundance both ribonucleotides and sophisticated proteins, such as ribonucleotide reductases, thymidylate synthases and reverse transcriptases. The first RNA viruses might have thus originated in such a world of RNA/protein cells from the combination of selfish RNA replicons with gene(s) encoding cellular proteins that facilitated the transfer of these replicons between RNA/protein cells (Fig. 1).

The Origin of Selfish RNA Replicons and Their Relationships With Modern RNA Viruses

In the ancient world of RNA cells, diverse types of selfish and non-selfish RNA replicons (circular or linear RNA plasmids, single or double-stranded) should have coexisted with probably important overlap between both. At some point in the history of life, after the emergence of proto-ribosomes, these RNA replicons started to be replicated by more and more sophisticated replicases (the ancestors of modern RNA-dependent RNA polymerases). It is likely that several different (non-homologous) types of RNA replicases, but also probably of RNA helicases, ligases, and nucleases, originated in the ancient world of RNA/protein cells. Different non-homologous families of these enzymes have indeed been described in the modern biosphere. These families are certainly very ancient since they can be divided into divergent subfamilies widespread in the current living world.

Remarkably, the RNA replicases of all known RNA viruses belong to a single protein superfamily that also includes reverse transcriptases encoded by reverse-transcribing viruses (RTV) and retroelements. RNA viruses and RTV have been therefore recently classified by the ICTV (International Committee on Taxonomy of Viruses) into the same realm, the *Riboviria*. It is tempting to conclude that replicases of *Riboviria* are the descendants of the RNA replicase that was used for replication in ancient RNA cells and that the first RNA viruses originated from a single class of selfish RNA replicons encoding this replicase. However, several other ancient families of RNA- or DNA-dependent RNA polymerases are present in the modern biosphere and could be the descendants of RNA replicases from extinct lineages of RNA cells and/or RNA viruses. This is the case of the monomeric RNA polymerase that is now used for transcription by some bacteriophages (thereafter called bacterioviruses), or else of an intriguing family of RNA/DNA

polymerases distantly related to the typical cellular RNA polymerases that perform transcription in the three domains of life (Archaea, Bacteria, and Eukarya). Some of these atypical and poorly known nucleotide polymerases use RNA as template and are involved in post-transcriptional gene silencing in eukaryotes, whereas others, such as the protein YonO encoded by a *Bacillus subtilis* prophage, function as transcriptase for some bacteriophages widespread in Bacteria.

It is likely that the transition from RNA to DNA-based cells has been a major bottleneck in cellular evolution. It is thus possible that many ancient RNA virus lineages did not survive this bottleneck and that *Riboviria* are the only lineage of ancient RNA viruses that survived the emerging DNA cells (the early *Riboviria* hypothesis). One can even imagine that all ancient RNA viruses disappeared and that *Riboviria* emerged later on during DNA cell evolution (the late *Riboviria* hypothesis).

The early *Riboviria* hypothesis is supported by their extreme diversity and the relative simplicity of most of them. *Riboviria* replicases are extremely divergent between the different subgroups in that realm. *Riboviria* also differ in terms of capsid components, genome structure (single of double-stranded RNA, segmented or not) and transcription/replication mechanisms (positive or negative single-stranded RNA viruses). Moreover, all *Riboviria* have rather small genomes and some of them have very simple life cycle, suggesting that they could correspond to “primitive” viruses, resembling the first RNA viruses that emerged on our planet. In this line of thought, it has been suggested that positive-sense RNA viruses, in which the same molecule functions as both a genome and mRNA, correspond to the most ancient ones.

A Scenario for the Origin of *Riboviria*

Most positive-sense RNA viruses have icosahedral virions with capsid proteins characterized by a simple structural fold called the “single jelly roll” (SJR). The protein domains containing this fold are often linked to another largely unstructured, positively charged N-terminal domain, allowing the packaging of the RNA in the capsid shell. Many cellular proteins widespread in the three domains of life also contain the SJR fold, suggesting that SJR proteins were already diversified at the time of LUCA. These proteins were thus probably already present in the world of RNA/proteins cells. Interestingly, many SJR proteins have carbohydrate recognition properties. As stated by Krupovic and Koonin, “a protein with such a property would be immediately beneficial to the virus because, in addition to providing a protective shell for the genome, it could ensure specific binding of the viral particle to the host cell”. One can thus imagine a rather simple scenario in which the first *Riboviria* originated from the recruitment by a selfish RNA replicon of a gene encoding such SJR protein linked to an RNA binding domain, the carbohydrate recognition properties of the SJR domain allowing the first virions to interact with glycoprotein receptors at the surface of RNA cells (Fig. 1).

The historical validity of this scenario can a priori be tested by the phylogenetic analysis of the RNA replicases combined to an analysis of the SJR capsid type distribution among *Riboviria*. Although sequences of viral RNA replicases are difficult to align, a whole tree of the *Riboviria* realm based on viral RNA replicases (rooted using reverse transcriptases of retroelements) has been published in 2018 by Koonin *et al.* (schematized in Fig. 2(A)). Based on this phylogeny, the authors suggested that *Riboviria* originated indeed from a single ancestor with the SJR capsid protein (thereafter called the SJR ancestral hypothesis). However, several groups of *Riboviria* produce capsid proteins with folds unrelated to the SJR. This is the case in particular of *Leviviridae* that infect Bacteria, negative-sense RNA viruses and for most double-stranded RNA viruses. The authors of the SJR ancestral hypothesis suggested that the SJR capsid protein was replaced several times independently by new types of capsid proteins during the diversification of *Riboviria* lineages.

The rooting proposed by Koonin *et al.* (using reverse transcriptases as an outgroup) does not really support the SJR ancestral hypothesis though because it implies that *Riboviria* with the SJR fold were originally present in only one of the two main branches of the replicase tree (Fig. 2(A), with the SJR of *Coronaviridae* being a secondary acquisition). One can thus imagine that *Leviviridae* (at the origin of one of the two major branches) have emerged independently from other *Riboviria* via the recruitment of different capsid proteins by replicons sharing the same type of replicase. However, if reverse transcriptases originated from RNA replicases, as in the early *Riboviria* scenario, the tree could instead be rooted within the diversity of replicases from positive-sense RNA viruses (Fig. 2(B)). In that case, positive-sense RNA viruses with an SJR capsid are present in the two main branches of the tree, making the ancestral SJR hypothesis a reasonable possibility (Fig. 1).

In any case, it remains to be understood how the first virions acquired the ability to egress from the cell and to infect another cell. Modern *Riboviria* are not good models to infer such events since they have co-evolved for billions of years with their DNA-based cell hosts and became adapted to their cellular fabric. For instance, *Riboviria* infecting eukaryotes have learnt how to manipulate the eukaryotic vesicle trafficking system for delivery and egress (often via the endosomal pathway) and for building viral factories. Similarly, *Riboviria* infecting bacteria are now dependent on complex and rather recent bacterial appendages (F pili in the case of *Leviviridae* and type IV pili in the case of *Cystoviridae*) that were most likely absent in ancient RNA cells. At first sight, eukaryotic *Riboviria* represent a simpler model since they do not have to cross several membranes. However, in the simplest modern *Riboviria* infecting eukaryotes, both egress and exit are completely dependent on the elaborated endosomal pathway of eukaryotic cells and many of them need to manipulate the endoplasmic reticulum or other elaborated eukaryotic structures to build their viral factories. A *minima*, one should imagine that the first cell in which a proto-virion became a virion already had the ability to excrete and to integrate such large macromolecular structure.

It has been suggested that extracellular membrane vesicles (EVs) that are produced by most modern cells could be interesting models to understand the origin of viruses because of the similarities between these vesicles and virions. However, the mechanisms of excretion and fusion of EVs in the three domains of life still remain poorly known. The best documented mechanisms are those occurring in eukaryotes and they are exceedingly complex with again the involvement of the endosomal

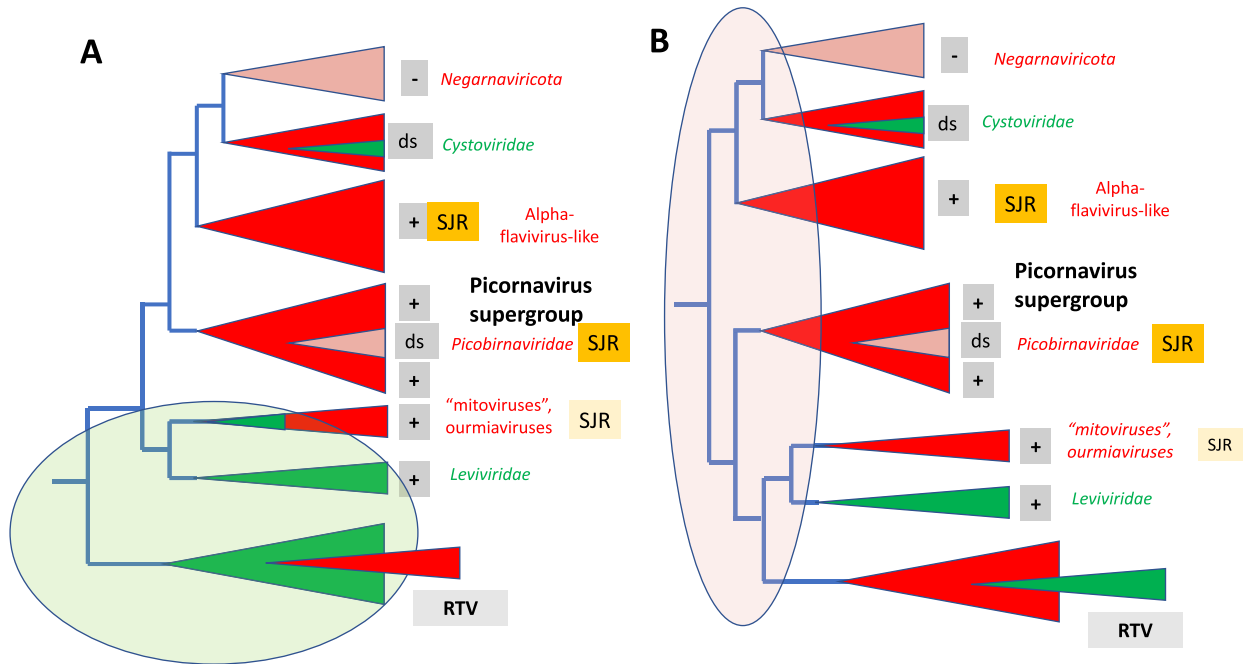


Fig. 2 Schematic history of the *Riboviria* realm based on the phylogenetic analysis of their RNA replicase. RNA viruses or Reverse Transcribing Viruses (RTV) infecting eukaryotes in red and RNA viruses or RTV infecting bacteria in green. The names of viral families infecting Bacteria are in green, the name of viral rank containing only viruses infecting eukaryotes are in red. The phylogeny is based on those published by Wolf *et al.* in 2017. In this publication, the authors proposed to root the tree between replicases of RNA viruses and reverse transcriptase (A). They propose a scenario in which RNA viruses infecting eukaryotes emerged from RNA viruses and/or retroelements infecting prokaryotes (green circle). Alternatively, if reverse transcriptases originated from RNA replicases, the tree can be rooted within replicases of RNA viruses (one possible such rooting, between the major groups of positive sense RNA viruses, is proposed in B). In that case, RNA viruses infecting eukaryotes might derived from ancient RNA viruses that overlapped with the ancient world of protein/RNA cells (pink circle). SJR: members of the respective clade contain viruses with SJR fold in their capsid protein. SJR in faint orange box, secondary capture of the SJR fold (see text). + positive sense RNA viruses, – negative sense RNA viruses, ds: double-stranded RNA viruses.

pathway. In order to imagine evolutionary scenarios for a vesicle-based origin of virions, one should look at simpler types of vesicles. This might be possible, and it is often assumed that ancient cells that predated LUCA already produced vesicles. Interestingly, lipid vesicles are rather easily formed in prebiotic conditions and it has been shown that addition of simple peptides can trigger vesicles fusion. One can thus imagine that some of the first proteins synthesized by proto-ribosomes (long after the origin of the first RNA-based cells) have triggered vesicle production by inducing membrane curvature, whereas other proteins made possible vesicle fusion. These proteins might have been recruited later on in the formation of early virions (Fig. 1). In modern cells, EVs can transfer RNA or DNA from cells to cells (vesiculation), including plasmids (plasmidion) and viral genomes (viral vesicle). This process might have been at the origin of virion dissemination and infectivity.

The Origin of RNA Viruses Infecting Eukaryotes

A major conundrum in the hypothesis that *Riboviria* derived from ancestral RNA viruses from the RNA-cell world is that *Riboviria* are very abundant and diverse in Eukaryotes, but very rare in Bacteria and completely unknown in Archaea (Fig. 3(C)). Bacterial RNA viruses only correspond to two families, *Leviviridae* and *Cystoviridae*, out of the 70 presently described *Riboviria* families, and they are only known to infect a few bacterial lineages. In the *Riboviria* early hypothesis, if *Riboviria* were present at the time of LUCA, one should imagine that they were later on entirely lost in the stem branches of the archaeal domain and that most of them also disappeared in the stem branch of the bacterial domain (Fig. 3(A), *Riboviria* in red). It has been suggested that RNA viruses are absent in Archaea because the last archaeal common ancestor (LACA) was a hyperthermophile. The reasoning is that RNA is much more fragile than DNA at high temperature. By the same token, most RNA viruses might have been lost in Bacteria if the last bacterial common ancestor (LBCA) also lived at high temperature. In support of this hypothesis, reconstructions of ancestral RNA and universal protein sequences by Manolo Gouy *et al.* have suggested that both the LACA and the LBCA were thermophilic or hyperthermophilic organisms. In the case of Bacteria, it is also possible that emergence of the peptidoglycan in the stem branch leading to Bacteria prevented most ancient viral lineages, especially those of RNA viruses, to survive in this lineage. In contrast, it has been suggested that RNA viruses were successful in eukaryotes because they could efficiently replicate in the intracellular membrane-rich cytoplasm of their cells. *Riboviria* indeed often use the endoplasmic reticulum membrane of eukaryotic cells to build their viral factories.

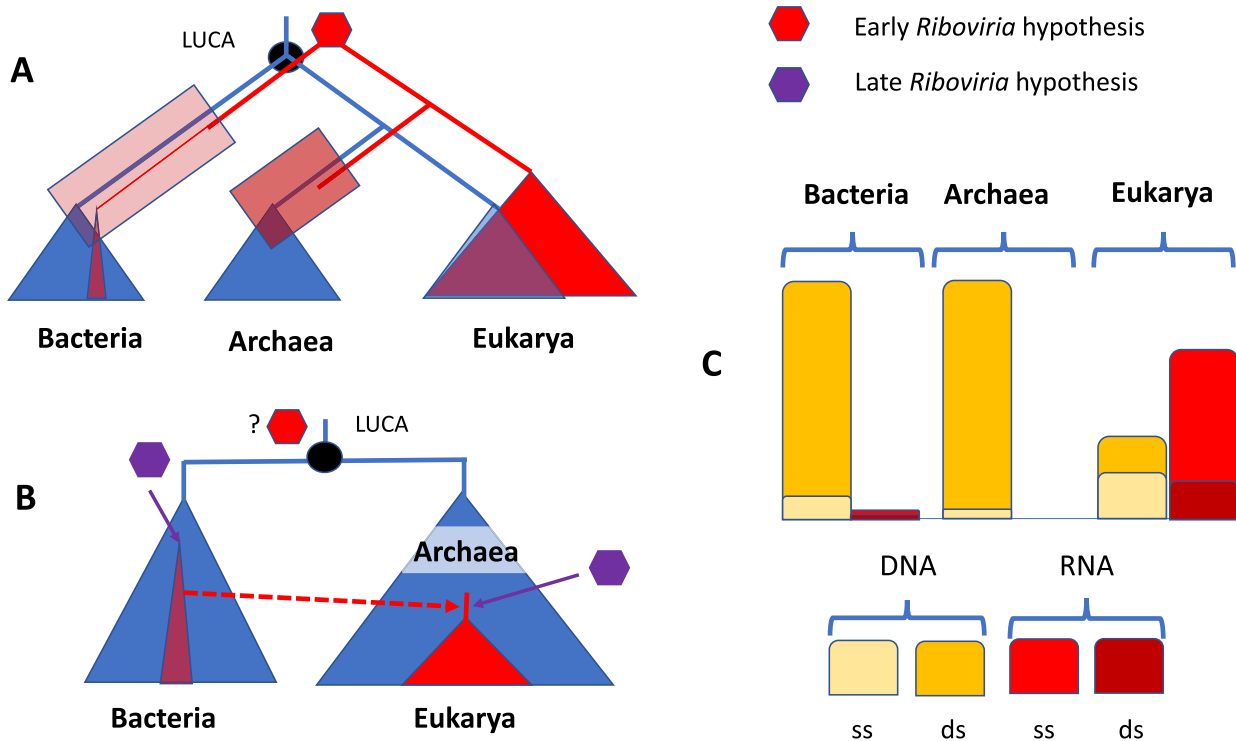


Fig. 3 The mysterious origin of *Riboviria* infecting eukaryotes. (A) Early *Riboviria* hypothesis. In this scenario, based on the Woese's tree of life (the three domains model) *Riboviria* already present at the time of LUCA (black circle). They completely disappeared in the stem branch leading to Archaea during a period of adaptation to very high temperature (pink box). They partly disappeared in the stem branch leading to Bacteria during a period of adaptation to high temperature (light pink box). (B) the late *Riboviria* hypothesis. In this scenario, based on the eocyte tree of life (the two domains model), Eukarya are a subgroup of Archaea. *Riboviria* were present or not at the time of LUCA. If they were present, they later on disappeared in the stem branch leading to Archaea. If they were not present, they appeared (purple capsids) in the stem branch leading to Bacteria and/or eukaryotes. If they appeared in Bacteria (the bacterial ancestor scenario), some subgroups of *Riboviria* were transferred later to the ancestor of eukaryotes via bacterial endosymbionts such as the mitochondrial ancestor. (C) schematic distribution of RNA and DNA viruses in the three domains of life. Adapted from Nasir *et al.* (2014).

The alternative hypothesis to explain the abundance of *Riboviria* in eukaryotes is that modern RNA viruses did not originate from ancient ones that predated LUCA but either from bacterial *Riboviria* (the bacterial ancestor scenario) or *de novo* in the proto-eukaryotic stem lineage (Fig. 3(B), *Riboviria* in red and purple for the early and late scenarios, respectively). These two late eukaryotic *Riboviria* hypotheses are not contradictory and can each be valid for different viral families. In the bacterial ancestor scenario, one should imagine that eukaryotic *Riboviria* originated from *Riboviria* that infected bacteria co-evolving with proto-eukaryotes (dotted line in Fig. 3(B)). For instance, it has been suggested that the ancestors of eukaryotic *Riboviria* might have originated from *Riboviria* that infected the bacterium at the origin of mitochondria. In that scenario, it is however unclear how these RNA viruses “learned” to infect the proto-eukaryote (or the archaeon) that engulfed their bacterial host, especially considering that eukaryotes and bacteria have completely different transcription and translation machineries. Moreover, the RNA replicases of eukaryotic *Riboviria* did not emerge from within those of bacterial RNA viruses in phylogenetic analyzes, except if one assumes that RNA replicases originated from bacterial reverse transcriptases (Fig. 2(A)). On the contrary, the replicases of bacterial *Cystoviridae* and *Leviviridae* (in green in Fig. 2) branch within those of eukaryotic RNA viruses if the tree is rooted within positive-sense viruses with SJR-type capsids (Fig. 2(B)). The second scenario, in which modern eukaryotic RNA viruses originated *de novo* in the proto-eukaryotic stem lineage, should be reconciled with the fact that the replicases of *Riboviria* infecting bacteriophages and eukaryoviruses are homologs. One should imagine in that case that these viruses emerged independently from the recruitment of different capsid proteins by selfish RNA MGEs encoding homologous replicases. It seems however difficult at the moment to imagine that the present diversity of *Riboviria* in eukaryotes could have entirely emerged from the limited number and diversity of *Riboviria* known in Bacteria (even if this diversity is probably largely underestimated).

The Origin of Reverse-Transcribing RNA and DNA Viruses

The genomes of some *Riboviria* is retrotranscribed into DNA in the virocell and new RNA genomes are produced from this DNA template. In a symmetric way, the genome of some DNA viruses (e.g., *Hepadnaviridae* and *Caulimoviridae* families) is retrotranscribed into RNA and new DNA genomes are produced from this RNA template. These fascinating Reverse Transcribed Viruses (RTV) have

been grouped under the *Revtraviricetes* class (*Paramavirae* kingdom), within the *Riboviria* realm with RNA viruses. Most RTV have RNA genomes and belong to the *Ortervirales* order, whose name has been derived from the reverse of retro. *Ortervirales* are divided into several families including the well-known retroviruses that share similar capsid and nucleocapsid proteins/domains and a family. Notably, RTV viruses of the *Hepadnaviridae* family encode a capsid protein which is unrelated to those of viruses of the *Ortervirales*.

The critical step in the life cycle of all these viruses is the reverse-transcription of RNA into DNA by reverse transcriptases. Since a reverse transcriptase activity was necessarily involved in the transition from RNA to DNA genomes, it is tempting to suggest that the origin of DNA coincided with the emergence of the first RTV, in agreement with the idea that DNA emerged in the viral world. The “out of virus” hypothesis for the origin of DNA posits that an RNA virus modified the chemistry of its genome as a protective feature against the defense mechanisms of their hosts that targeted RNA genomes. This might have taken place in two steps with first the emergence of viruses with U-DNA genome (following the emergence or recruitment of a ribonucleotide reductase activity) and finally with T-DNA genomes (following the emergence or recruitment of a thymidylate synthase activity). If this scenario has taken place in an ancient lineage of RNA viruses that became an RTV with an RNA genome, and later on an RTV with a DNA genome, one has to understand what could have been the direct selection pressure favoring the presence of a DNA intermediate in the virocell. This DNA intermediate might have been a storage form allowing the viral genome to persist in an inhospitable environment for foreign RNA. It remains to be understood how RNA cells used to discriminate between an invader RNA and their own RNA genome.

Although both RNA and DNA RTV are evolutionarily related and have been grouped in the same class, *Revtraviricetes*, it is unclear if the RT that participated to the RNA to DNA transition was a direct ancestor of the RT of *Revtraviricetes*. Modern RT are very diverse and can be divided in several phylogenetic clades, those encoded by *Revtraviricetes* belonging to only one of them, all the others corresponding to RT encoded by retroelements widespread in Eukarya and rare in Bacteria. Notably, unlike retroelements, RTV are only present in eukaryotes, their broad distribution within eukaryotes suggesting ancient origin(s), possibly prior to the emergence of eukaryotes. One can again propose either that RTV have been lost in Archaea and Bacteria or that they have originated in the eukaryotic stem lineage. In any case, *Revtraviricetes* should have coevolved with the eukaryotic lineage for a very long time and became adapted to these complex hosts, possibly recruiting new types of capsids.

The RT of *Revtraviricetes* might still be the direct descendants of those that promoted the RNA to DNA transition, but this transition might have been promoted instead by an RT more closely related to the RT of some clades of modern retroelements or from an extinct, or not yet discovered lineage of RTV or retroelement. One can even imagine that the reverse transcriptase that promoted the RNA to DNA transition was not encoded by a *Revtraviricetes* ancestor, not even by another group of *Riboviria*, but by a virus from an extinct lineage.

The Origin of DNA Viruses

Whereas DNA viruses are less abundant than RNA viruses in eukaryotes, they are largely dominant in the bacterial biosphere and every known archaeal virus has a DNA genome. It has been suggested that DNA viruses are less abundant in eukaryotes because the emergence of the nucleus prevented DNA viruses to have an easy access to the DNA replication apparatus of proto-eukaryotes, whereas RNA viruses do not need such access. As a consequence, many ancestral lineages of DNA viruses were eliminated before the time of the last eukaryotic common ancestor (LECA) and only a few DNA viruses finally developed the tricks to replicate within the nucleus. Indeed, many DNA viruses infecting eukaryotes have rather large genomes, encode their own replication apparatus and replicate within cytoplasmic viral factories. However, Bacteria and Archaea are also infected by some DNA viruses with large genomes and autonomous replication machineries, indicating that such viruses are very ancient and predated the origin of the nucleus. Most DNA viruses have double-stranded DNA genomes (dsDNA) (circular or linear) although many DNA viruses infecting eukaryotes and a few of those infecting archaea and bacteria have single-stranded DNA genomes (ssDNA) (circular or linear). Viral dsDNA genomes exhibit a surprising size variability with lengths varying from 1.7 kb (two genes) for circovirus up to 2.5 Mb (more than two thousand genes) for pandoravirus, both infecting eukaryotes. These genomes are replicated by an amazing diversity of DNA replication mechanisms and machineries.

If ancient RNA genomes were progressively transformed into DNA genomes during the RNA to DNA transition, one should imagine that ancient RNA replicators were transformed into DNA replicators and that enzymes previously used to replicate and manipulate RNA were progressively recruited by evolving DNA replicators to manipulate DNA. This is exemplified by the fact that many enzymes involved in RNA or DNA metabolism, such as nucleotide polymerases, nucleases, or helicases are homologous. Some viral enzymes involved in DNA genome replication are homologous to their analogs in cellular organisms but others are specific to both RNA and DNA viruses, as in the case of superfamily III helicases. As a rule, there are usually several families of non-homologous enzymes for each activity (exemplified by the case of RNA and DNA polymerases) suggesting a complex and ancient history of these enzymes that probably overlapped with the RNA to DNA transition. DNA replicators that emerged during the RNA to DNA transition might have later on recruited capsids from RNA viruses or ancient cellular proteins to produce various types of DNA viruses. This evolutionary trajectory has been demonstrated in the case of several families of small ssDNA viruses that infect eukaryotes and have been grouped into the *Cressdnaviricota* phylum. These viruses originated several times independently from different rolling-circle DNA plasmids that recruited their SJR capsid proteins from *Riboviria*. In that case, the diversification of *Riboviria* indeed preceded the emergence of *Cressdnaviricota*. A similar scenario might have taken place earlier for other lineages of

DNA viruses, but it is also possible that DNA viruses originated from the recruitment by DNA plasmids for their capsids of novel types of proteins that were not previously used as capsid proteins by RNA viruses. For instance, members of the order *Ortervirales* are evolutionarily related to viruses of the *Hepadnaviridae* family because they share homologous reverse transcriptase, but they have unrelated capsid proteins.

An open question is the timing of DNA virus emergence: did they originate before or after the LUCA? This question is related to another one, i.e., did LUCA already had a DNA genome or still an RNA genome? Most biologists consider that LUCA already had a DNA genome since all current cellular organisms have DNA genomes. However, in that case, one would have expected the DNA replication proteins in the three domains to be homologous, which is not the case. The RNA LUCA hypothesis is actually supported by the fact that the three major proteins involved in DNA replication in Archaea/Eukarya and in Bacteria (the replicase, the primase, and the helicase) are not homologous. To explain this paradox, it has been suggested by proponents of the DNA LUCA that the replication proteins present in LUCA were the ancestors of the modern archaeal ones and that they were replaced in the bacterial branch by the ancestors of the modern bacterial ones, possibly of viral origin. On the other hand, the proponents of an RNA LUCA argue that LUCA was probably simpler than modern organisms, with smaller ribosomes and a less specific transcription initiation system, suggesting an early stage – possibly still RNA based – in cellular evolution.

The transition from RNA to DNA genomes post-LUCA would also explain why the tempo of protein evolution was drastically reduced during the transition from LUCA to the last common ancestor of the three domains. If LUCA still had an RNA genome, it is possible that DNA viruses originated independently and diversified during this transition in the stem lineages of the three modern domains. In agreement with this hypothesis, many families of DNA viruses are indeed unique to a particular domain. This is especially the case for those infecting cells of the *Archaea* domain and that exhibit a great variety of capsid proteins completely unrelated to those of bacteriophages and eukaryoviruses.

However, there are two exceptions to this observation, corresponding to the so-called cosmopolitan viruses. The latter belong to two realms that include viruses infecting cells from the three cellular domains, *Varidnaviria* (formerly known as the PDR1/Adenovirus lineage) and the *Duplodnaviria* (formerly the HK97 lineage). *Varidnaviria* are characterized by homologous major capsid proteins containing a DJR fold oriented vertically with respect to the capsid surface and a packaging ATPase of the FtsK/HerA superfamily P-loop ATPases family, whereas *Duplodnaviria* (*Duplodna* standing for double-stranded DNA) grouping *Herpesviridae* infecting animals and head-tailed viruses infecting Archaea and Bacteria (*Caudovirales*) are characterized by homologous major capsid proteins containing the so-called HK97 fold and packaging ATPases called Terminases. It has been assumed from their universality that ancestors of these cosmopolitan DNA viruses were already thriving at the time of LUCA. However, in that case, one would have expected their evolutionary history to be congruent with those of cellular organisms in which Archaea and Eukarya are closely related and very divergent from Bacteria in terms of molecular biology and universal protein sequences. This is not the case. On the contrary, *Varidnaviria* and *Duplodnaviria* infecting bacteria are much more similar to those infecting Archaea than to those infecting eukaryotes. For instance, among *Duplodnaviria*, head and tailed viruses infecting Archaea and Bacteria are so similar in terms of virion structure and gene content that they have been classified into the same order, *Caudovirales*. These viruses are extremely divergent from *Herpesviridae*, only sharing with them very distant pATPase, MCP and several other proteins. The same situation is observed in the case of *Varidnaviria*: bacteriophages and archaeoviruses of this realm are mixed in phylogenies based on their capsid protein and packaging ATPases and have been classified into the same class: the *Tectiliviricetes*. They form a group rather distant from the eukaryoviruses of this realm, *Nucleocytoviricota* and their viruses of the *Lavidaviridae* family (virophage) and Polintoviruses. Presently, *Adenoviridae*, which infect eukaryotes, have been classified with *Tectiliviricetes* by the ICTV, even though the structural analysis of their capsid protein suggested that they could be more related to those of other *Varidnaviria* eukaryoviruses.

Phylogenetic analysis of *Varidnaviria* and *Duplodnaviria* based on the concatenation of their major capsid proteins and packaging ATPases have produced trees in which archaeoviruses and bacteriophages are mixed and grouped together apart from eukaryoviruses (Fig. 4 for *Varidnaviria*). Such trees, which are not congruent with the cellular tree of life, suggests that some transfers of viruses have taken place between Archaea and Bacteria, or possibly between proto-archaea and proto-bacteria. An early transfer of *Caudovirales* from Bacteria to Archaea would explain why these viruses are more diverse and widespread in Bacteria than in Archaea.

Several scenarios can be proposed for the evolution of cosmopolitan viruses. They are illustrated in Fig. 4 in the case of *Varidnaviria*. Considering the huge diversity and distribution of *Varidnaviria* in Eukaryotes, one cannot exclude that these viruses originated in proto-eukaryotes and were later on transferred to Archaea and Bacteria (Fig. 4(A)). However, a published hypothesis suggests instead that they originated in Bacteria and were transferred to eukaryotes, possibly via mitochondria, as in the case of one of the two *Riboviria* late hypotheses (Fig. 4(B)). Finally, if the ancestors of *Varidnaviria* were already present at the time of LUCA, one should imagine that they were later on lost in Archaea and recovered from Bacteria (Fig. 4(C)). Similar scenarios can be proposed for *Duplodnaviria*. It is presently unclear if it will be possible in the future to choose between these different scenarios. One can hope that the discovery of new cellular and viral lineages in some isolated corner of our planet through the systematic and exhaustive screening of all environments will bring new insights in that matter, but this is far from a certitude.

An intriguing observation is that whereas viruses belonging to *Varidnaviria* and *Duplodnaviria* produce non-homologous virions and use non homologous packaging ATPase, they sometimes share homologous DNA replication proteins. In particular, some replication proteins used by *Caudovirales* infecting bacteria are related to replication proteins of *Nucleocytoviricota* (formerly known as the NucleoCytoplasmic Large DNA Viruses), the phylum of *Varidnaviria* eukaryovirus that includes giant viruses such as *Mimiviridae*. This suggests that various lineages of DNA viruses recruited homologous replication proteins that emerged in the world of viral and mobile elements during the transition from the RNA to the DNA world.

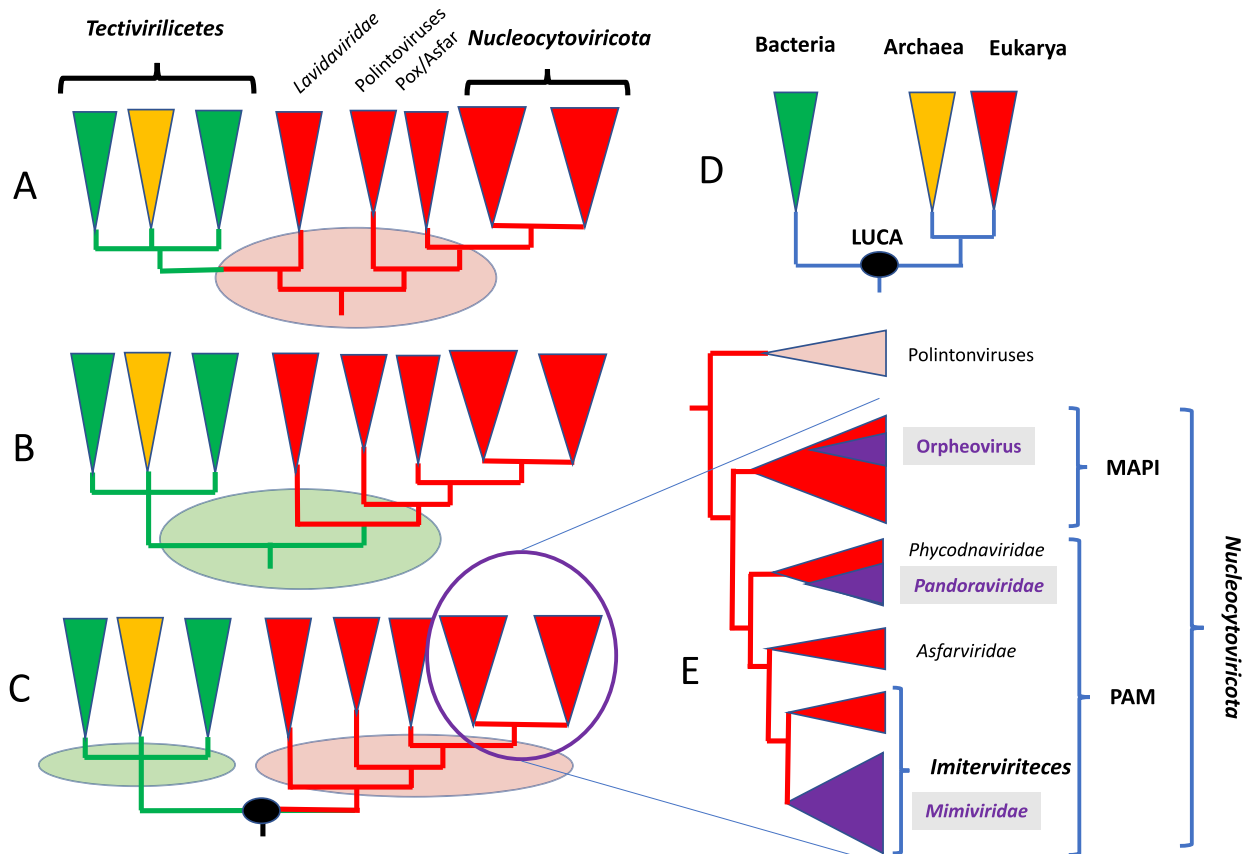


Fig. 4 The history of the realm *Varidnaviria*. A-C: Schematic history of the realm *Varidnaviria* with various rooting. This phylogeny is based on the phylogenetic analysis of their major capsid proteins and packaging ATPases (concatenation) published by Woo and colleagues in 2019: (A) the eukaryotic ancestor scenario, *Varidnaviria* originally infected proto-eukaryotes (pink oval) and were later transferred independently to Archaea and Bacteria. (B) the bacterial ancestor scenario, *Varidnaviria* originally infected proto-bacteria and were later transferred to proto-bacteria (pink oval). (C) the LUCA scenario, *Varidnaviria* were present at the time of LUCA (black circle) and later on lost in early proto-Archaea. There were later introduced in proto-Archaea from proto-bacteria (green oval). (D) the three domains universal tree of life. Whereas Archaea are mixed with Bacteria in the "viral tree" (A, B, C) they are sister group to Eukaryotes in the three domains tree and they are the ancestors of eukaryotes in the eocyte tree (Fig. 3(B)). E, zoom on the phylum *Nucleocytoviricota* phylogeny (rooted with Polintoviruses) based on those published by Guglielmini *et al.* in 2019, giant viruses are in purple.

The Origin of Giant Viruses and Megaphages

Some members of the *Varidnaviria* and *Duplodnaviria* realms have large genomes and virion size. These "giant viruses" (*Mimiviridae*, pandoraviruses, Orpheovirus) or "megaphages" (*Caudovirales*) encode from several hundred up to two thousand genes, and some of them produce virions that can be seen by optical microscopy. The discovery of these viruses has rejuvenated the idea that viruses (at least some of them) possibly derived from ancient cellular lineages of obligate parasites that lost their ribosomes. This hypothesis is not supported by phylogenetic analyzes showing that giant viruses are scattered within their respective phyla and originated several times independently from smaller viruses (see Fig. 4(E) for the distribution of giant viruses – in purple – among *Nucleocytoviricota*). Their genomes most likely increased by gene recruitment and duplication, but also probably via extensive *de novo* gene production.

The "cellular origin" hypothesis has been proposed by some authors to explain the presence in the genome of giant viruses of a majority of genes without cellular homologs, supposedly derived from their ancient cellular ancestors. This was based on the traditional view that considers viruses as inert particles and mostly pick-pockets of cellular genes. However, during the virocell stage, one can anticipate that the well-known processes that trigger the formation of proto-genes and finally new genes from non-coding sequences in cellular genomes similarly take place in viral genomes during their replication. In fact, in all viruses, from the smallest to the largest, the majority of viral genes (called orfans) have no cellular homologs, testifying for the abundance of genes that directly originated in viral genome and were never transferred to their hosts. This is probably due to the high level of genome replication that occurs during the virocell stage of the virus life cycle. The same observation also holds for capsidless mobile genetic elements that also encode mostly "orfan" genes and that are likely another continuous source of new genes in the biosphere.

Some authors have favored the "cellular origin" hypothesis to explain the presence in the overall viral proteome of many proteins that share folds present in proteins from the three domains of life. This observation indeed strongly supports the idea that

viruses were already around at the time of LUCA but do not specifically support the cellular hypothesis scenario. It indicates that ancient viruses of that time would have indeed recruited ancient proteins carrying these folds but, conversely, it is also likely that cellular lineages thriving at the time of LUCA had already integrated in their genomes many genes originally produced in the world of viruses and other mobile genetic elements.

Conclusion

Our ideas about the place of viruses in the history of life have been improved these last decades with our better knowledge of the virus world and the identification of evolutionary relationships between major viral lineages. We have now a rather extensive view of most modern viral lineages. A high number of different types of capsid protein has been determined, confirming that viruses are polyphyletic if viruses' origin is assimilated to the acquisition by selfish replicons of protein(s) suitable for the production of infectious virions. Of the three classical scenarios for the origin of viruses, i.e., the virus first hypothesis, the regression hypothesis (cellular origin), and the escape hypothesis, only the latter has survived but in a rather different setting. Viruses did not originate from some chromosomal portions that escaped from their host genomes after the separation of prokaryotes and eukaryotes, as was still assumed not so long ago, but from selfish replicons that escaped from their hosts by inventing a new way of reproduction, the production of virion. For that purpose, selfish replicons were combined with gene cassettes encoding capsid proteins and other proteins involved in virion production (the chimeric hypothesis). These events took probably place very early in RNA-based cells to produce the first RNA viruses (the ancient escape hypothesis) and more recently to produce DNA viruses through the combination of selfish DNA replicons with new types of virions or virions recruited from RNA viruses. These events have taken place repeatedly and at different times in the history of life.

Many questions remain to be solved. It remains unclear if some modern viruses are the direct descendants of the first virus lineages or if the most ancient viral lineages have all vanished in the mists of time. In particular, we would like to know if reverse transcribing viruses are the direct descendants of those that participated to the historical RNA to DNA transition. The origin of eukaryoviruses with no clear-cut relatives in the two other domains also remains a mystery. Why archaeoviruses and eukaryoviruses are so different whereas the molecular biology of their hosts is rather similar (at least the basic processes)? Why RNA viruses are so abundant in eukaryotes and so scarce in Bacteria? How to reconcile what we know about virus history with the topology of the universal tree of life? Answering these questions will require at least to be clear on the position of eukaryotes in this tree, either as sister group to archaea in the so-called 3D model or deeply anchored within Archaea in the so-called 2D model, a question which is still in debate. A major advance in any case is that viruses are now recognized by more and more evolutionists as both part of the problem and of the answer. They are no longer viewed as byproduct of life's evolution but as major components that have strongly influenced the history of life on our planet.

Further Reading

- Arkhipova, I.R., 2017. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. *Mobile DNA* 6 (8), 19. doi:10.1186/s13100-017-0103-2.
- Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N., Gouy, M., 2008. Parallel adaptations to high temperatures in the Archaeal eon. *Nature* 456, 942–945.
- Forrest, D., James, K., Yuzenkova, Y., Zenkin, N., 2017. Single-peptide DNA-dependent RNA polymerase homologous to multi-subunit RNA polymerase. *Nature Communications* 6 (8), 15774. doi:10.1038/ncomms15774.
- Forterre, P., 2013. The common ancestor of archaea and eukarya was not an archaeon. *Archaea* 2013, 372396. doi:10.1155/2013/372396.
- Forterre, P., 2016. To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life. In: Ankeny, R.A. (Ed.), *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 59. Elsevier, pp. 100–108.
- Gill, S., Catchpole, R., Forterre, P., 2019. Extracellular membrane vesicles in the three domains of life and beyond. *FEMS Microbiology Reviews* 43, 273–303.
- Guglielmini, J., Woo, A.C., Krupovic, M., Forterre, P., Gaia, M., 2019. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 116, 19585–19592.
- Iranzo, J., Puigbò, P., Lobkovsky, A.E., Wolf, Y.I., Koonin, E.V., 2016. Inevitability of genetic parasites. *Genome Biology and Evolution* 8, 2856–2869.
- Kazlauskas, D., Varsani, A., Koonin, E.V., Krupovic, M., 2019. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nature Communications* 10 (1), 3425. doi:10.1038/s41467-019-11433-0.
- Koonin, E.V., Dolja, V.V., 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiology and Molecular Biology Reviews* 78, 278–303.
- Koonin, E.V., Dolja, V.V., Krupovic, M., et al., 2020. Global organization and proposed megataxonomy of the virus world. *Microbiology and Molecular Biology Reviews* 84 (2), e00061.
- Krupovic, M., Dolja, V.V., Koonin, E.V., 2019. Origin of viruses: Primordial replicators recruiting capsids from hosts. *Nature Reviews Microbiology* 17, 449–458.
- Krupovic, M., Koonin, E.V., 2017. Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences of the United States of America* 114 (12), E2401–E2410. doi:10.1073/pnas.1621061114.
- Krupović, M., Bamford, D.H., 2010. Order to the viral universe. *Journal of Virology* 84, 12476–12479.
- Nasir, A., Caetano-Anolles, G., 2017. Identification of capsid/coat related protein folds and their utility for virus classification. *Frontiers in Microbiology* 8, 380. doi:10.3389/fmicb.2017.00380.
- Nasir, A., Forterre, P., Kim, K.M., Caetano-Anollés, G., 2014. The distribution and impact of viral lineages in domains of life. *Frontiers in Microbiology* 5, 194. doi:10.3389/fmicb.2014.00194.
- Raoult, D., Forterre, P., 2008. Redefining viruses: Lessons from mimivirus. *Nature Reviews Microbiology* 6, 315–319.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., et al., 2018. Origins and evolution of the global RNA virome. *mBio* 9 (6), e02329. doi:10.1128/mBio.02329-18.
- Woo, A.C., Gaia, M., Guglielmini, J., Da Cunha, V., Forterre, P., 2019. Evolution of the PRD1-adenovirus lineage: A viral tree of life incongruent with the cellular universal tree of life. *bioRxiv* 741942. doi:10.1101/741942.

The Virocell Concept

Patrick Forterre, Archeology Virology Unit, Institut Pasteur, Paris, France and French National Center for Scientific Research, Institute of Integrative Biology of the Cell, University of Paris-Saclay, Gif sur Yvette Cedex, France

© 2021 Elsevier Ltd. All rights reserved.

The Virion/Virus Paradigm

The virocell concept has been proposed to counterbalance the traditional view which assimilates viruses to their virion (viral particles), the so-called “virus/virion paradigm” (sensu Forterre). From the time of their discovery, viruses were indeed assimilated to the infectious component capable of passing through filters used to retain bacteria. Later on, this infectious component turned out to be the viral particle – the virion – which became assimilated to the virus. The virus/virion paradigm became dominant among biologists because pictures of virions can be easily used to illustrate “what is a virus”, both for scientists and for the general public. This is obvious in the media but also in the scientific literature. The fight against viruses through vaccination also entrenched this way of thinking, since antibodies designed to prevent virus infection usually recognized components of the virion. In recent years, the virus/virion paradigm was also strengthened by the progress in the field of structural virology, when the virion structures could be determined at the molecular and atomic levels. Finally, a universal component of virions, the major capsid proteins, is now used to define viral lineages. The major capsid protein is often considered as the hallmark of the virus, because it is sometimes the only marker than can help discriminating between viruses and plasmids.

In agreement with the virus/virion paradigm, viruses were classified using both the shape of the virion (e.g., order *Caudovirales* groups all “head and tailed viruses”) and the nature of the nucleic acid present in the virion (RNA or DNA, single-stranded or double stranded, positive- or negative-sense RNA). In a few cases, the nucleic acid itself was assimilated to the virion, thus to the virus. Hence, naked infectious RNA molecules derived from *bona fide* RNA viruses were called capsidless viruses by Eugene Koonin and Valerian Dolja and officially recognized as viruses by the ICTV (International Committee for the Taxonomy of Viruses) (e.g., narnaviruses, endornaviruses, hypoviruses, etc.). The virus/virion paradigm is also reflected in most virus definitions. For instance, Jacob and Wolmann defined viruses as “a genetic element enclosed in a protein coat”, whereas Didier Raoult and Patrick Forterre, in a paper proposing a new definition of viruses, still describe a virus as “a capsid encoding organism that is composed of proteins and nucleic acids, self-assembled in a nucleocapsid”. The virus/virion paradigm also explains why André Lwoff includes the presence of a single type of nucleic acid (RNA or DNA) among criteria distinguishing viruses from cells. However, this is disputable since DNA viruses produce viral messenger RNAs that belongs to the virus. Accordingly, the Lwoff criteria is only valid if the virus is assimilated to the virion.

The virus/virion paradigm had several important consequences in the way most scientists consider viruses. This explains why they are often described as giant macromolecular entities – much like the ribosome – instead of living organisms. This is considered as a testament of an absolute gap between the cellular and the viral worlds. Viruses were thus expelled from the living realm by many biologists, including some leading virologists, as inert particles lacking the main attributes of life such as the production of energy and a self-sustaining metabolism.

Critics of the Virion/Virus Paradigm

The virus/virion paradigm was first criticized by Claudiu Bandea who wrote in 1983 that “the living phase of the virus is the vegetative phase when the virus shows the major physiological properties of organisms: metabolism, growth, reproduction”. The intracellular stage of the viral reproductive cycle can be indeed viewed as the most important aspect of this cycle because the virus expresses and replicates its genome during this stage. However, in line with the virus/virion paradigm, this stage is usually called the “eclipse phase” in the literature, being simply defined by the absence of visible virions. A major criticism of the virus/virion paradigm was made by Jean-Michel Claverie, following the discovery in 2003 of the giant Mimivirus and of its huge virus factories located in the cytoplasm of the infected amoeba. Jean-Michel Claverie states in 2006 that the virus/virion paradigm is a case of “when the finger points to the star, the fool looks at the finger” or else that “Interpreting the virion particle as ‘the virus’, is very much like looking at a spermatozoid and calling it a human”. He proposes that: “the virus factory should be considered the actual virus organism”. A weakness of this proposal was that, in contrast to viruses infecting eukaryotes (eukaryoviruses), viruses infecting archaea (archaeoviruses) or bacteria (bacteriophages) do not produce easily discernible virus factories. To extend the idea to the whole viral realm, Patrick Forterre and David Prangishvili combined the definition of Jean-Michel Claverie (the virus is the viral factory) with an old quote of André Lwoff: “the virus transforms the cell into a virus factory” and concluded that the infected cell is the real viral organism. The term virocell, i.e., “a virus expressing itself in a cellular form” was then proposed to illustrate this point (Fig. 1).

The virocells was defined as cells producing infectious virions that have the potential to produce new virocells, as opposed to ribocells, defined as cells containing ribosomes and dividing by binary fission. The notion of virocell is easy to understand when the genome of the infected cell is completely destroyed or inactivated because the infected cell remains obviously a cell, but without “cellular DNA” that could define it as a bacterium, an archaeon or a eukaryote. It became a virocell containing only viral genome.

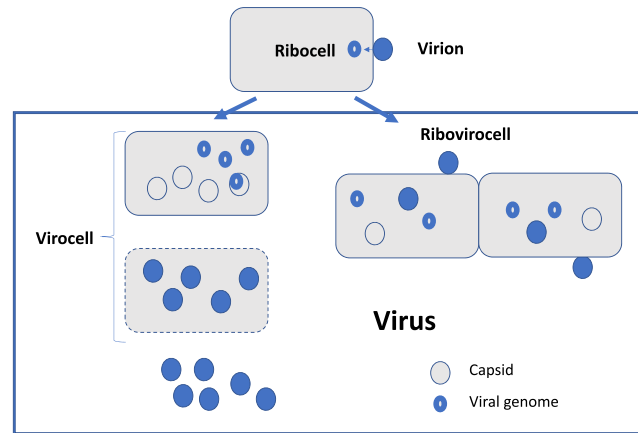


Fig. 1 The various forms of the virus, virion, virocell and ribovirocell.

The virocell concept fits the definition of viruses as “*capsid encoding organisms*” or virion producing organisms, since the viral infection transforms the cell into a novel type of organism, a virocell. In the virocell, the cellular metabolism network is very often entirely re-shaped toward the production of virions. The new metabolism emerging in the virocell can be considered as the metabolism of the virocell itself. Notably, it can lead to the production of new metabolites that can be used as virocell markers (see below). Many examples of virocell metabolism have been reviewed in 2016 by Vardi and colleagues, and techniques such as single cell expression profiling have been specifically designed to analyse virocell metabolism. The rewiring of the cellular metabolism into virocell metabolism involves both virus encoded proteins that modulate the activity of the host enzymes, but also in many case new metabolic enzymes encoded by the virus itself (often called vAMG for virus Associated Metabolic Genes, *sensu* Mya Breitbart). An amazing example of vAMG is provided by bacteriophages infecting cyanobacteria (cyanophages). These viruses often encode components of the photosynthetic pathway that replace labile cellular counterparts, which disappear during the infection process. This allows the virocell to benefit from the energy provided by photosynthesis for the production of virions. The resulting organism is a photosynthetic virus.

In the first publication proposing the virocell concept, it was unclear if the virus itself could be equated to the virocell. Later on, it was made clear that the virus cannot be reduced to the virocell, but that: “*the term virus should describe a biological process and integrate all aspects of the viral reproduction cycle*”, i.e., the virion, the virocell, as well as silent viral genomes (proviruses) free or integrated in cellular chromosomes (Fig. 1). The difference between the virus and virocells can be nicely illustrated in the case of viruses, such as arboviruses, infecting cells from several different species of insects and vertebrates. In that case, the same viral species can produce several types of completely distinct virocells using cells from different hosts. In other cases, with the help of a human scientist, infectious virions can even be produced artificially in the absence of “natural” virocell, using *in vitro* cell lysates mimicking a virocell.

Critics of the Virocell Concept

The virocell concept has been criticized by Purification Lopez-Garcia and David Moreira who labelled it as “*a conceptual trick*” to justify considering viruses as living (see below) or else as “*epistemological cheating*”, considering that “*defining an entity (a virus) in term of itself plus a portion of another entity (a cell) is alien to logic*”. To illustrate their point, these authors suggested that defining the infected cell as a virocell is the same as to consider that a remora fish associated to a shark becomes a remora shark. This critic was refuted by pointing to the difference between a virus and extracellular (the remora) or intracellular (the *Chlamydia*) organisms living together in association (the remora and the shark) or in competition (the *Chlamydia* and the infected bacterium). In both cases, the parasitic organism maintains its organismal or cellular integrity even if its shape is profoundly modified during its reproductive cycle, as in the case of *Chlamydiae*. The latter always maintain a closed intracellular structure surrounded by a membrane, use their own ribosomes for protein synthesis and continue dividing by cell division. In contrast, the virus needs usurping the cellular territory of its victim. The virus robs the membrane of the infected cell to achieve a cellular structure and uses the host ribosomes to synthesize its proteins, transforming the infected cells into a novel cellular organism, the virocell.

The criticism according to which: “*defining an entity (a virus) in term of itself plus a portion of another entity (a cell) is alien to logic*” is valid in the framework of the virus/virion paradigm, since in this context the virus (*an entity*) is the virion (*itself*), whereas the infected cell is assimilated to another entity (*a cell*). If the virus is defined as a process that encompasses all aspects of its reproductive cycle, this criticism is no more valid. One can ask if it is legitimate to include in the virus definition an entity (the virocell) that includes components (e.g., ribosomes, many enzymes, membrane proteins and lipids) that came from the infected cell. The answer is yes, according to the philosopher Thomas Pradeu, who wrote in 2010 that: “*An organism is made of constituents that do not need to have originated in it*”.

Considering the infected cell as an unbreakable entity limits our understanding of the complex phenomenon triggered by the viral infection. This is illustrated by the ambiguity of the term “infected cell” itself. This term can *a priori* define either the bacterium, the archaeon or the eukaryotic cell infected by the virus or the virocell produced by this infection. During the infection process, the cell ‘entity’ is indeed broken since the cell contains two types of cells, the infected ribocell and the emerging virocell. The virocell concept thus requires distinguishing the infected cell from an organism. An example that illustrates this point is an amoeba infected by both intracellular bacteria and a mimivirus, which is itself infected by a virophage. In this case, four organisms, defined by four different evolutionary lineages, share the same cell. The concept of a cell as a vehicle, suggested by Matti Jalasvuori, can be useful to understand this situation. He suggests considering cells as vehicles carrying different types of replicons (cellular chromosome, plasmids and/or viral genomes). This concept can also be applied when considering cells as vehicles carrying different types of organisms.

Notably, the distinction between cells and virocells helps removing some ambiguity that are inherent to the concept of a cell as an unbreakable entity. For instance, Purificacion Lopez-Garcia and David Moreira wrote that “*viruses are evolved by cells*” to deny any active role of the virus in its own evolution. The problem with such reasoning is most obvious when considering new viral genes emerging *de novo* in infected cell, in which the cellular genome has been completely destroyed. In such situation, considering that *viruses are evolved by cells* would mean that viruses can be evolved by dead cells. In the framework of the virocell concept, the sentence *viruses are evolved by cells* remains valid if the term cell means virocell, i.e., “*viruses are evolved by virocells*”.

Diversity of Virocells

Virocells and Ribovirocells

The transformation of the host (victim) ribocell into a virocell is gradual, so that for some time two organisms, the dying host cell and the virocell, co-exist in the same cell, fighting each other (in particular via the CRISPR and anti-CRISPR systems in Archaea and Bacteria). Later on, either one of the two organisms win, or they can manage to coexist (see below). If the ribocell is entirely transformed into a virocell, the system will finally collapse since virocell is a transient organism, which will eventually disappear through self-disruption. This cannot be taken as an argument against the virocell concept since many “regular” cellular organisms also have a short lifespan. In many cases, the infected cell survives and enters into a kind of “pacific coexistence” with the virus, allowing the production of virions, a situation sometime referred to as carrier state or persistence. Cells that produce virions but that can still divide have been called ribovirocells (Fig. 1). In ribovirocell, the virus and the bacterium, the archaeon or the eukaryote have a symbiotic relationship that can be more or less beneficial for the different partners. There are thus different types of virocells, depending of the degree of autonomy and capacity maintained by the original ribocell. Single-cell analyses have shown that a population of infected cells contains in fact various subpopulations of cells expressing different subsets of cellular and viral genes, testifying for different outcomes in the virus-host arms races. Some of these subpopulations probably correspond to virocells or ribovirocells, others correspond to cells becoming resistant to the virus, allowing persistence of the host. The lysogenic state can be viewed as one of the strategies used by ribocells to survive viral infection and for the virus to wait for conditions optimal for virocell production. A lysogenic cell with an integrated or episomal viral genome that is not expressed cannot be called a ribovirocell, since it does not produce virions. As for other definitions proposed by scientists to describe the living world using classification, these definitions somehow arbitrarily divide a reality that is often more complex, virocells, ribovirocells and lysogenic cells forming a continuum, often without clear-cut borders, with overlap and the possibility to shift from one state to the other under mutational or environmental shift.

Virocells With Nucleus

Many eukaryoviruses replicate and transcribe their genomes in viral organelles that mimic the nucleus of eukaryotic cells. These viral factories can be assimilated to virocell nucleus. Beside virus factories, lipid bodies and autophagosomes produced in some infected eukaryotic cells can also serve as new forms of organelles in the virocell. The case of virocells with nuclei (synkaryotic virocells) is especially interesting. At first, it seemed that virocells without nuclei (akaryotic virocells) were only produced during the infection of prokaryotic (akaryotic) cells, archaea and bacteria, whereas synkaryotic virocells were only produced during the infection of eukaryotic (synkaryotic) cells. It came as a surprise when Pogliano and colleagues reported in 2017, that the bacteriophage $\phi 2-1$ produces a viral nucleus during the infection of the bacterium *Pseudomonas chlororaphis* (Fig. 2). The membrane of the viral nucleus is formed by a virus-encoded protein and, remarkably, this nucleus is localized at the centre of the bacterium by another virus-encoded protein, PhuZ, a homologue of eukaryotic tubulin.

Implications of the Virocell Concept

Viruses as Cradles of New Genes

The virocell concept was put forward partly to remind evolutionists that viruses are not only pickpockets of cellular genes (as usually assumed in the virus/virion paradigm) but also a cradle of new genes (thus new functions). This is because new genes

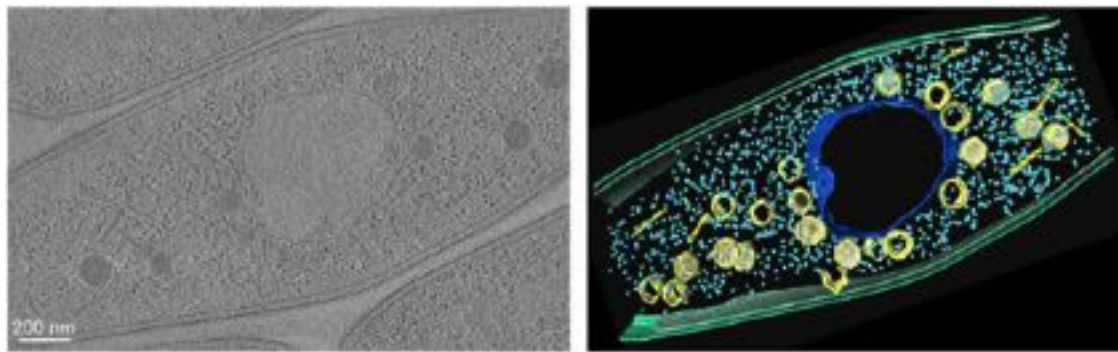


Fig. 2 (Left) Slice through a cryo-electron tomogram of *Pseudomonas chlororaphis* cell infected with phage 201φ2-1. Upon infection, the phage assembles a large protein shell containing phage DNA with capsids docked to this shell for DNA encapsidation and ribosomes excluded from the shell in the host cytoplasm. (Right) Segmentation of the tomogram in left showing host membranes (green), ribosomes (light blue), protein shell (dark blue), capsids (light yellow), tails (dark yellow). Courtesy of Kanika Khanna, Elizabeth Villa, and Joe Pogliano.

continuously arise during the intracellular stage of the virus reproduction cycle (in the virocell), when the viral genome is replicated and expressed. The mechanisms of *de novo* formation of viral genes are most likely the same as those responsible for the formation of new genes in cellular genomes. The mechanism of the emergence of new gene in DNA genomes is now well understood from comparative genomics studies performed in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. Most new genes originate from the selection of short open reading frames (proto-genes) arising randomly in intergenic regions. It is likely that similar mechanisms operate in all three domains of cellular life as well as in DNA viruses. In RNA viruses, new genes can originate by overprinting on the non-coding strand of pre-existing genes, producing overlapping genes that can be unrelated from one viral strain to another. The massive *de novo* creation of viral genes explains why most viral genes have no cellular homologues.

Enumeration of Viruses in the Environment

The estimation of the abundance and importance of viruses in the environment has been traditionally done by counting virus-like particles (VLPs) using epifluorescence or flow cytometry methods. In these studies, the systematic assimilation of the VLP number to the number of viruses once again testifies to the entrenchment of the virus/virion paradigm. This strategy has been questioned on methodological grounds, considering that VLPs can also correspond to extracellular vesicles (EVs). Moreover, on the epistemological grounds, this method was considered to be misleading because, in the framework of the virocell concept: “*estimating the number of viruses by counting their virions is equivalent to counting the number of fish eggs in the ocean to estimate the number of fishes*”. The number of viruses present in the environment is thus grossly overestimated if they are assimilated to the number of VLPs and it would be more accurate to quantify viral activity in the environment by counting virocells. Unfortunately, this is a challenging task. It is much more difficult to estimate the number of virocells in the environment than to count the number of virions. The virocell number should be always lower than the total cell number since the cells detected by epifluorescence or flow cytometry include both virocells and ribocells. However, it could be a large portion of the total cell number. For instance, it has been estimated in several ecological studies that around 40% of bacteria in the ocean are lysed by viruses at any moment. This would mean that at least 40% of total cells enumerated in these environments as bacteria are in fact virocells. Vardi and colleagues proposed that detection of virocell metabolomic signature and of unique virocell-derived molecules could serve as biomarkers of virocells to estimate the impact of viruses on the chemical and microbial composition in the marine environment.

Are Viruses Living?

The debate between scientist regarding the living status of viruses has been on-going for years. The acellular nature of viruses in the virus/virion paradigm was a major argument to consider them as non-living entities because, according to André Lwoff: “*an organism is constituted of cells*”. Moreover, it was assumed that viruses could not be living since they have no metabolism. For instance, Marc Van Regenmortel wrote in 2003 that “*viruses do not possess many of the essential attributes of living organisms, such as the ability to capture and store free energy and they lack the characteristic autonomy arising from the presence of integrated, metabolic activities*”. Proponents of the idea that viruses are not living have criticized the virocell concept as a trick to bypass these traditional arguments. The notion that viruses lack the ability to capture and store free energy because they do not have an integrated metabolism is indeed valid if the virus is assimilated to the virion. In contrast, the concept of metabolising virocells emphasizes that viruses can capture and store free energy, similar to ‘regular cells’.

The philosophers John Dupre and Maureen O’Malley wrote that: “*Life arises when lineage forming entities collaborate in metabolism*”. Viruses fit this definition in the virocell framework, since they are both genetic and metabolic entities. In the course of his reflexion on the nature of viruses and their status as living organisms, the notion of a living organism and the definition of life

itself was questioned. As an example of the difficulties associated to these questions, it is not possible to determine when the transition from living to non-living organisms occurred in the case of the evolution leading from the alpha-proteobacterial ancestor of mitochondria to *bona fide* mitochondria. To maintain the concept of living “alive”, I suggested considering as living not only organisms but all biological entities “as long as they are operational in the process of life”. According to this definition, “entities” corresponds to “individuals”, i.e., objects that are “separable, countable and have acceptable clear-cut spatial boundaries” according to philosopher Stéphane Chauvier. Life can then be defined as “the mode of existence of biological individuals”. With this definition, a living individual could be a protein, a chromosome, a virion or a virocell, considered to be lineages with an evolutionary history designed by Darwinian principles of variation and selection.

Conclusions

In my opinion, the virocell concept clarifies the nature of viruses and opens new philosophical questions about the definition of life. It also has practical potential value in that it calls for reshaping of research programs in virology by putting more spotlight on the intracellular phase of the virus life cycle. Such research has the potential to identify new targets for fighting viral infections by looking for drugs active against the virocell as well as new biomarkers to analyse viral activity in the environment.

Further Readings

- Bandeau, C.I., 1983. A new theory on the origin and the nature of viruses. *Journal of Theoretical Biology* 105, 591–602.
- Breitbart, M., 2012. Marine viruses: Truth or dare. *Annual Review of Marine Science* 4, 425–448.
- Carvunis, A.-R., Rolland, T., Wapinski, I., *et al.*, 2012. Proto-genes and de novo gene birth. *Nature* 487, 370–374.
- Cello, J., Paul, A.V., Wimmer, E., 2002. Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template. *Science* 297, 1016–1018.
- Chaikerasitak, V., Nguyen, K., Khanna, K., *et al.*, 2017. Assembly of a nucleus-like structure during viral replication in bacteria. *Science* 355 (6321), 194–197.
- Chauvier, S., 2008. Particuliers, individus et individuation. In: Ludwig, P., Pradeu, T. (Eds.), *L'Individu. Perspectives Contemporaines*. Paris: Vrin, pp. 11–35.
- Claverie, J.M., 2006. Viruses take center stage in cellular evolution. *Genome Biology* 7 (6), 110.
- Dupre, J., O'Malley, M.A., 2009. Variety of living things: Life at the intersection of lineages and metabolism. *Philosophy, Theory, and Practice in Biology* 1, e003.
- Forterre, P., Gaia, M., 2016. Giant viruses and the origin of modern eukaryotes. *Current Opinion in Microbiology* 31, 44–49.
- Forterre, P., Soler, N., Krupovic, M., Marguet, E., Ackermann, H.W., 2013. Fake virus particles generated by fluorescence microscopy. *Trends in Microbiology* 21 (1), 1–5.
- Forterre, P., 2012. The virocell concept. In: eLS. Chichester: John Wiley & Sons Ltd. doi:10.1002/9780470015902.a0023264.
- Forterre, P., 2016. To be or not to be alive: How recent discoveries challenge the traditional definitions of viruses and life. *Studies in History and Philosophy of Biological and Biomedical Sciences* 59, 100–108.
- Forterre, P., 2013. The virocell concept and environmental microbiology. *The ISME Journal* 7, 233–236.
- Forterre, P., 2011. Manipulation of cellular syntheses and the nature of viruses: The virocell concept. *Comptes Rendus Chimie* 14 (4), 392–399. doi:10.1016/j.crci.2010.06.007.
- Jalasvuori, M., 2012. Vehicles, replicators, and intercellular movement of genetic information: Evolutionary dissection of a bacterial cell. *International Journal of Evolutionary Biology* 2012, 874153. doi:10.1155/2012/874153.
- Koonin, E.V., Dolja, V.V., 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiology and Molecular Biology Reviews* 78, 278–303.
- Krupović, M., Bamford, D.H., 2010. Order to the viral universe. *Journal of Virology* 84, 12476–12479.
- López-García, P., 2012. The place of viruses in biology in light of the metabolism- versus-replication-first debate. *History and Philosophy of the Life Sciences* 34, 391–406.
- Lopez-García, P., Moreira, D., 2012. Viruses in biology. *Evolution: Education & Outreach* 5, 389–398.
- Moliner, C., Fournier, P.E., Raoult, D., 2010. Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution. *FEMS Microbiology Reviews* 34, 281–294.
- Moreira, D., López-García, P., 2009. Ten reasons to exclude viruses from the tree of life. *Nature Reviews Microbiology* 7, 306–311.
- Pradeu, T., Kostyrka, G., Dupré, J., 2016. Understanding viruses: Philosophical investigations. *Studies in History and Philosophy of Biological and Biomedical Sciences* 59, 57–63.
- Pradeu, T., 2016. Mutualistic viruses and the heteronomy of life. *Studies in History and Philosophy of Biological and Biomedical Sciences* 59, 80–88.
- Rancurel, C., Khosravi, M., Dunker, A.K., Romero, P.R., Karlin, D., 2009. Overlapping genes produce proteins with unusual sequence properties and offer insight into *de novo* protein creation. *Journal of Virology* 83, 10719–10736.
- Raoult, D., Forterre, P., 2008. Redefining viruses: Lessons from mimivirus. *Nature Reviews Microbiology* 6 (4), 315–319.
- Rosenwasser, S., Ziv, C., Creveld, S.G.V., Vardi, A., 2016. Virocell metabolism: Metabolic innovations during host-virus interactions in the ocean. *Trends in Microbiology* 24 (10), 821–832.
- Rosenwasser, S., Sheyn, U., Frada, M.J., *et al.*, 2019. Unmasking cellular response of a bloom-forming alga to viral infection by resolving expression profiles at a single-cell level. *PLoS Pathogens* 15 (4), e1007708. doi:10.1371/journal.ppat.1007708.
- Mendoza, S.D., Berry, J.D., Niewegłowska, E.S., *et al.*, 2008. A nucleus-like compartment shields bacteriophage DNA from CRISPR-Cas and restriction nucleases. *bioRxiv* 370791. doi:10.1101/370791.
- Soler, N., Krupovic, M., Marguet, E., Forterre, P., 2015. Membrane vesicles in natural environments: A major challenge in viral ecology. *The ISME Journal* 9 (4), 793–796.
- Thompson, L.R., Zeng, Q., Kelly, L., *et al.*, 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences of the United States of America* 108 (108), 757–764.
- Van Regenmortel, M.H., 2010. Logical puzzles and scientific controversies: The nature of species, viruses and living organisms. *Systematic and Applied Microbiology* 33, 1–6.
- Wimmer, E., 2006. The test-tube synthesis of a chemical called poliovirus. *Embo Reports* 7, 53–59.
- Zhao, L., Saelao, P., Jones, C.D., Begun, D.J., 2014. Origin and spread of de novo genes in *Drosophila melanogaster* populations. *Science* 343, 769–772.

Virus Taxonomy

Jens H Kuhn, National Institutes of Health, Frederick, MD, United States

Published by Elsevier Ltd.

Nomenclature

BC Baltimore Class

DdRp DNA-directed RNA polymerase

HUH His-hydrophobic-His

IAMS International Association of Microbiological Societies

ICNV International Committee on Nomenclature of Viruses

ICTV International Committee on Taxonomy of Viruses

ICVCN International Code of Virus Classification and Nomenclature

IUMS International Union of Microbiological Societies

LHT Lwoff, Horne, and Tournier

MSL Master Species List

RdRp RNA-directed RNA polymerase

Rep replication initiation protein

TaxoProp taxonomic proposal

VDN Archives of Virology's Virology Division News

Glossary

Classification The rational process of assigning viruses to taxa.

Inclusion principle Viruses assigned to a lower taxon based on certain virus characteristics must also have the characteristics needed for classification into a higher-ranked taxon that includes the lower taxon.

Monophyletic Belonging to a group of viruses that consists of all the descendants from a common ancestor.

Nomenclature The (often controversial) assignment of names to taxa.

Polyphyletic Not being connected to a common ancestor.

Polythetic Sharing many, but not all characteristics of a common core set of characteristics.

Species Currently the lowest-ranked taxon in virus taxonomy. Not to be confused with "viruses," which are physical entities assigned to species.

Taxon A man-made category for viruses that share certain characteristics. The now commonly accepted correct plural

of "taxon" is the fake Greek "taxa" ("taxons" would actually be more correct but is not used in English).

Taxonomy Virus classification into taxa and taxon nomenclature. The term "taxonomy" is a misnomer as the term is based on Greek "τάξις [táxis]". "Taxinomy" is a better derivation but is not used in English.

Virions Particles produced by some, but not all viruses, during the viral lifecycle for transfer of the viral genome from one host cell to another. Not to be confused with "virus." Virions differ from virus particles in that the former are assumed to be infectious, whereas the latter may be non-infectious or otherwise defective

Virus A replicative non-cellular physical entity that is obliged to depend on a host cell for replication. Virus is not to be confused with "species," which are man-made categories into which viruses are classified nor be confused with "virion." which are particles that are produced by some, but not all, viruses

History of Virus Taxonomy

Early Taxonomic Developments (1886–1971)

Virus taxonomy is a virology subspecialty that addresses the grouping (classification) of viruses (physical entities) into categories (concepts) called taxa and the development and implementation of a standardized system of naming (nomenclature) for taxa.

Virus taxonomy arguably began with the discovery of "ultrafiltrable" (i.e., non-bacterial) agents today called "viruses." This discovery began in 1886, when Adolf Eduard Mayer (1843–1942) grouped a set of similar infectious diseases of tobacco plants under the name "Mosaikkrankheit des Tabaks" (German for "tobacco mosaic disease"). In 1892, Dmitrij Iosifovič Ivanovskij (1864–1920) reported that the sap of a plant affected by tobacco mosaic disease retained its infectivity even after passage through a Chamberland filter that held back most bacteria. Unaware of Ivanovskij's experiments, Martinus Willem Beijerinck (1851–1931) also observed that filtration of infectious sap did not result in sterility. In 1898, Beijerinck reported that filtered sap retained its infectious nature even after dilution and that the sap of a newly infected plant was similarly potent after passage. Beijerinck coined the term "contagium vivum fluidum" (Latin for "contagious living liquid") for this novel type of pathogen. The term "virus" (Latin for "poison"), previously used for any infectious agent or toxin, became ever more frequently used for this novel type of pathogen, variations of which were then discovered in animals (e.g., foot-and-mouth disease virus), including humans (e.g., yellow fever virus). The (partially) particulate nature of viruses was proven when Félix d'Herelle (1873–1949) reported the invention of the plaque assay for virus particle quantification in 1917, when he discovered viruses infecting bacteria. In 1935, Wendell Meredith Stanley (1904–1971) crystallized tobacco mosaic disease-causing particles and demonstrated that they consisted largely of protein. In 1939, Gustav-Adolf Kausche (1901–1960), Edgar Pfankuch, and Helmut Ruska (1908–1973) published the first-ever electron-microscopic images of virus particles, namely those of tobacco mosaic disease virus.

The first systematic classification of viruses was proposed in 1939 by Bennett on behalf of the newly formed Committee for Virus Nomenclature of the Council of the American Phytopathological Society. Accordingly, viruses were grouped by the type and morphological/cytological manifestation of symptoms they cause in their hosts, their host tropisms, their modes of transmission via disease vectors, and the antigenic, chemical, and physical properties of their particles. In 1941, a classification was suggested by Bawden that was based solely on chemical, antigenic, and morphological properties of virus particles. This suggestion was, in essence, followed by others for many years. In 1948, Francis O. Holmes (1899–1990) suggested a first all-encompassing taxonomic system that was somewhat reminiscent of that used for animals and plants. He proposed a top-rank taxon, the order “*Virales*,” that would include three suborders for viruses that infect bacteria (“*Phaginae*”), plants (“*Phytophaginae*”), or animals (“*Zoophaginae*”). Holmes further created 13 families, 32 genera, and 248 species. However, this system did not gain community support because Holmes separated viruses based on their host tropisms and ignored the considerable morphological similarities of the particles produced by all of the viruses that he separated into the 3 suborders.

From 1953–1957, André Michel Lwoff (1902–1994) created and expanded a set of criteria for the definition of “virus.” This definition informed a hierarchical virus taxonomy brought forth in 1961 by Peter D. Cooper, who divided animal viruses using particle characteristics into those that contain either DNA (“*deoxyviruses*”) or RNA (“*riboviruses*”). Further division of each group was dependent on ether sensitivity, i.e., dependent on whether the particles were wrapped in lipid envelopes (“*lipoviruses*”), or ether resistance (“*clathroviruses*”). From 1963–1967, this system was improved to incorporate virus particle sensitivity to low pH and particle shape and symmetry as classification criteria.

The first taxonomic system that received broad attention is that of Lwoff, Horne, and Tournier (the LHT system), which was first published in 1962 and finalized in 1966. The LHT system grouped viruses into one phylum (“*Vira*”) with two subphyla based on RNA or DNA presence in virions (“*Deoxyvira*” and “*Ribovira*”). Each subphylum included classes based on the symmetry of virion capsids (e.g., the classes “*Deoxyhelica*” for “*helical DNA viruses*” and “*Ribocubica*” for “*cubical RNA viruses*”). These classes were further divided into orders (“*...viraes*”) with suborders (“*...viridales*”), families (“*...viridae*”), subfamilies (“*...virinae*”), genera (“*...virus*”), subgenera (“*...virus*”), and described type species and common names for taxon members. Although none of the higher taxa stood the test of time, several family names used in the system are still used today (e.g., “*Poxviridae*,” “*Adenoviridae*,” “*Paramyxoviridae*” became the current *Poxviridae*, *Adenoviridae*, and *Paramyxoviridae*, respectively), and the suffixes for order, family, subfamily, genus, and subgenus names prevailed, albeit they are now consistently italicized (“*...viraes*,” “*...viridae*,” “*...virinae*,” “*...virus*,” “*...virus*”).

Baltimore Classification (1971)

In 1971, David Baltimore published a working classification of viruses that is still used today in parallel with official virus taxonomy. Rather than assigning viruses to taxa, Baltimore grouped all viruses into six disconnected groups without any subdivisions, now commonly referred to as Baltimore Classes (BCs). BCs were established based on the type of nucleic acid incorporated into virions, and thereby on the type of virus reproduction:

- (1) BC I: viruses with double-stranded DNA genomes that have a replication-expression strategy highly reminiscent of that of cellular organisms (e.g., *Adenoviridae*);
- (2) BC II: viruses with single-stranded DNA genomes (e.g., *Geminiviridae*);
- (3) BC III: viruses with double-stranded RNA genomes (e.g., *Reoviridae*);
- (4) BC IV: viruses with positive-sense RNA genomes (e.g., *Picornaviridae*);
- (5) BC V: viruses with negative-sense RNA genomes (e.g., *Paramyxoviridae*);
- (6) BC VI: viruses with positive-sense RNA genomes that replicate via DNA intermediates produced by reverse transcription of the genome (e.g., *Retroviridae*);
- (7) BC VII (added later to the original system): viruses with double-stranded DNA genomes that package a double-stranded DNA form or an RNA-DNA hybrid into virions and replicate via reverse transcription (e.g., *Hepadnaviridae*).

Until recently, BCs have been widely referred to as informal highest ranks of virus classification and the individual BCs were widely assumed to be monophyletic. However, recent studies (e.g., Wolf *et al.*, Koonin *et al.*) call this assumption into question. Although the partitioning of viruses into BCs will remain highly useful for teaching purposes, BCs are unlikely to be used as top ranks in official virus taxonomy.

Current Virus Taxonomy (1971–Present)

Influenced by the publication of the LHT system, the virology community clearly felt an urgent need for an official virus taxonomy. Consequently, at the 1966 International Congress for Microbiology in Moscow, the Executive Committee of the International Association of Microbiological Societies (IAMS) established the International Committee on Nomenclature of Viruses (ICNV) to develop a globally respected and applicable virus taxonomy for all virus types of all life forms. The ICNV, which in 1974 became today's International Committee on Taxonomy of Viruses (ICTV) under the International Union of Microbiological Societies (IUMS, the successor to IAMS), published its First Report in 1971. From then on until the present, virus taxonomy (today including not only viruses but also viroids and satellites) has been under the jurisdiction of this committee. The vast majority of virology and microbiology specialty journals request (but unfortunately rarely enforces) that manuscript authors follow the official ICTV taxonomy.

Contrary to the LHT system, which approached virus classification from the top taxonomic rank downwards, the ICNV/ICTV established a hierarchical virus taxonomy from the lowest taxonomic rank upwards. This approach was taken because, at the time, hallmark properties common to all viruses (akin to, for instance, 16S rRNA of bacteria or mitochondrial DNA of animals) were unknown. Genome sequence information for most viruses was lacking, and, therefore, phylogenetic analyses were impossible to perform. In addition, high mutation rates and horizontal gene transfer obscured relationships between virus groups.

From the initiation of ICTV's classification efforts, viruses (from hereon, including viroids and satellites) were suspected to be polyphyletic. Thus, the probability of being able to join large groups of viruses under higher-rank taxa using demonstratable evolutionary relationships appeared low. Consequently, viruses were grouped according to the morphological, biophysical, and biochemical properties of their virions and the "phenotype" of infections. After the advent of increasingly efficient genome sequencing methodologies, viruses were grouped according to viral genomic sequences and phylogenetic relationships to one another within relatively closely related virus sets. The result of these activities were numerous individual taxonomic hierarchies for viruses, typically up to the family rank, that were not connected to each other at a top rank.

The historical development of the official ICTV virus taxonomy can be reviewed in published ICTV Reports and interim updates that typically have been published in *Archives of Virology's* Virology Division News (VDN). The latest (10th) ICTV report has been published open-access online and is continuously updated at least annually. Announcements for major updates to the report have been and are being published continuously as Taxonomic Profiles in *The Journal of General Virology*.

The first ICTV report listed only 2 families, 27 genera, 10 subgenera, and 18 "virus groups", but the classification scheme has expanded considerably since then. For instance, the first order (*Mononegavirales*) was established in 1991, and in 2019 the first phylum (*Negarnaviricota*) and realm (*Riboviria*) were made official. By that time, 5560 species had been established (Table 1).

Despite this extension of taxon ranks and associated taxa, many families are currently "free floating" (i.e., families not connected to other families at higher taxonomic ranks). However, the majority of RNA viruses have been assigned to the highest available rank, realm. This assignment was based on the recognition that all these RNA viruses have genomes that share a common viral hallmark gene (VHG), namely one that encodes an RNA-directed RNA polymerase (RdRp) and that these polymerases form a monophyletic group (Fig. 1).

As shown in Fig. 1, most RNA virus families and orders could be grouped into major branches, but these branches have not yet received official status, except for Branch 5. Branch 5, established for negative-sense RNA viruses, is now officially recognized by the ICTV as phylum *Negarnaviricota*. This phylum is currently the only example of a top-down virus taxonomy of all available principle/primary ranks and, thereby, places the viruses of all included taxa into relationship to each other (Fig. 2).

Aside from expanding the number of available ranks and filling them with increasing numbers of taxa and numerous incremental improvements, current virus taxonomy underwent two additional major changes compared to historic efforts:

Table 1 Scope of the official ICTV taxonomy as ratified in spring of 2019. Listed are all permitted principle/primary taxon ranks (black), secondary ranks (grey), and the number of currently established (but not necessarily interconnected) taxa within these ranks. Ranks are ordered from lowest (top, species) to highest (bottom, realm) similar to the classical "Tree of Life" in which the highest ranks are depicted as the tree trunk close to earth and the lowest ranks are depicted as the finest branches up in the canopy. Intra-taxon divergence is high at the highest ranks and low at the lowest ranks

Rank	Number of established taxa
Species	5,560
Subgenus	59
Genus	1,019
Subfamily	79
Family	143
Suborder	7
Order	14
Subclass	0
Class	6
Subphylum	2
Phylum	1
Subkingdom	0
Kingdom	0
Subrealm	0
Realm	1

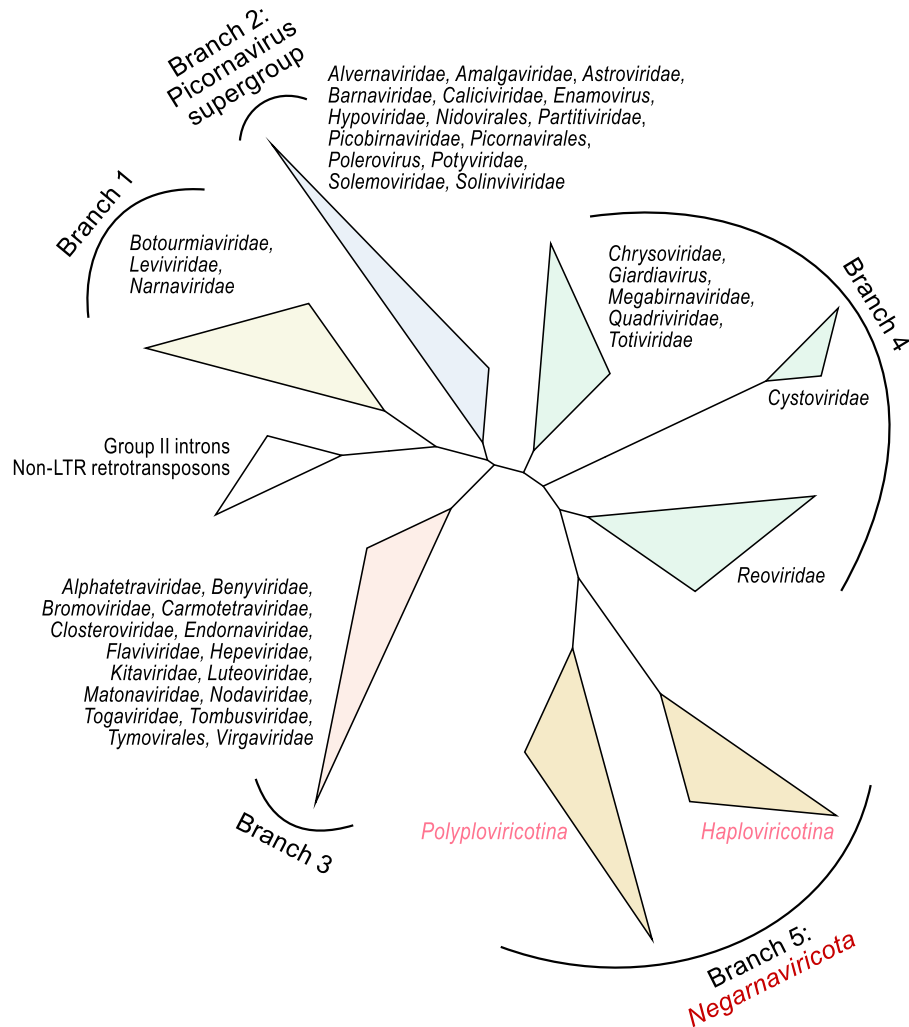


Fig. 1 Phylogenetic analysis of RNA virus RNA-directed RNA polymerases (RdRps) by Wolf *et al.*

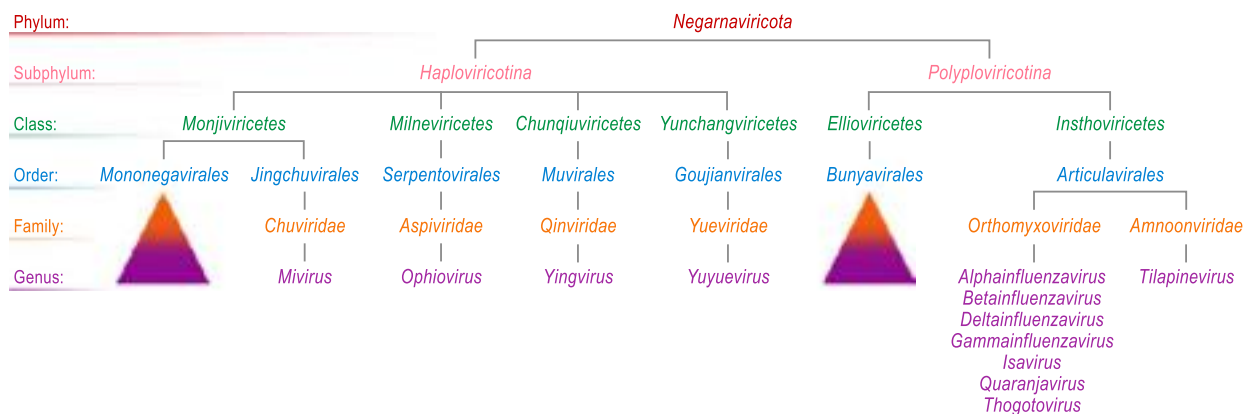


Fig. 2 Current taxonomy of *Negarnaviricota*. Due to space constraints, the current family/genus composition of *Mononegavirales* and *Bunyavirales* are not shown, and species have been omitted.

- (1) In 1991, the ICTV embraced the idea that “virus species” should be conceptually differentiated from “viruses” and that the biological concept of “species” as it has been applied to botanical, mycologic, prokaryotic, and zoologic taxonomy is also universally applicable to virus taxonomy. “Virus species” was first defined as “...a polythetic class of viruses that constitutes a replicating lineage and occupies a particular ecological niche” by van Regenmortel. Currently, “[a] species is the lowest

taxonomic level in the hierarchy approved by the ICTV. A species is a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria". Both species definitions have been highly controversial, and discussions continue on how to improve them. However, scientists broadly agree that the lowest rank of species is needed.

- (2) In 2017, the ICTV acknowledged that most viruses of the global virome are unlikely to ever be cultured or otherwise characterized in a laboratory. Hence, the ICTV accepted virus classification based on information deduced from an at least coding-complete virus genome sequence. This step expanded the classification to include numerous viruses known only from metagenomic datasets and, thereby, opened the door for sophisticated phylogenetic analyses and a vastly improved official description of the virosphere. Importantly, this step by the ICTV only means that the minimum requirement for virus classification is the availability and analysis of a coding-complete virus genome, but it does empathetically not exclude taking other virus characteristics, such as particle morphologies, tissue and host tropism, or infection phenotype, into consideration.

The International Committee on Taxonomy of Viruses (ICTV)

Organization

The ICTV, a non-profit organization, is the only committee of the Virology Division of the IUMS. Contrary to common perception, the ICTV is a relatively large, international organization, and ICTV decisions are achieved in a highly democratic manner based on majority voting at multiple steps involving representatives of the global virology community. The ICTV Executive Committee (EC) consists of 19 elected members with term limits, including a President, a Vice-President, several Secretaries, and 6 Subcommittee Chairs who are experts for large groups of viruses of all cellular organisms. Next to the EC, the ICTV consists of 11 elected Life Members (experts who contributed significantly to the development of virus taxonomy in the past), 42 nominated National Members (national representatives of international microbiological or virologic IUMS member societies), and 101 Study Group Chairs. The latter are ICTV members appointed by EC Subcommittee Chairs, typically for a maximum of two 3-year terms, to assemble an unlimited number of experts, who are not part of the ICTV, but who have specific expertise on viruses of, typically, a viral family or order. All ICTV members perform their duties voluntarily without pay.

Official statutes outline the ICTV remit, actions, and interactions within ICTV and with the community, and a code, the International Code of Virus Classification and Nomenclature (ICVCN), outlines the official rules and recommendations for virus classification into taxa and taxon nomenclature. Importantly, the ICTV is currently not responsible for the nomenclature of viruses (their names and abbreviations) or their subclassifications into isolates, strains, variants, genotypes, serotypes, lineages, or clades. Since no other organization exists that authoritatively administers these community needs, this lack of ICTV responsibility continuously causes confusion and often frustration among virologists who consider these needs part of virus taxonomy.

Taxonomic Process

Contrary to common perception, the ICTV does not impose taxa on the virology community but instead evaluates, votes on, and administers submitted taxonomic proposals (TaxoProps) originating from the virology community. TaxoProps can be submitted electronically to the ICTV at any time by any individual, whether (s)he is a member of the ICTV or not. Populating a TaxoProp is straightforward and guided by standardized Microsoft Word and Excel templates with associated instructions that can be downloaded from the ICTV homepage. Ideally, TaxoProps are submitted to the ICTV by the authors of a scientific research article describing a novel virus or providing novel information on the evolutionary relationships of a known virus. Under these circumstances, the virus discoverers can weigh in on classification and nomenclature discussions. However, virus discoverers do not have to be consulted during the classification process, and, hence, anybody can propose the classification or reclassification of any virus at any time. Typically, the annual deadline for TaxoProp submission is set for a day in June—TaxoProps submitted prior to this date will be considered by the ICTV typically in July; whereas TaxoProps submitted after the submission deadline will be considered in July of the subsequent calendar year.

Submitted TaxoProps undergo several steps of scrutiny:

- (1) After submission to the ICTV, the TaxoProp is forwarded to the most appropriate EC Subcommittee Chair. For instance, a TaxoProp will be forwarded to the EC Bacterial and Archaeal Viruses Subcommittee Chair if the TaxoProp addresses a bacterial virus. The Subcommittee Chair will perform an initial review on whether the proposed taxonomy conforms to current ICVCN rules and whether the TaxoProp is complete with the information necessary for further evaluation. If necessary, the Subcommittee Chair will request additional information or edits from the TaxoProp author(s);
- (2) Once satisfactory to the EC Subcommittee Chair, the TaxoProp is forwarded to the most appropriate Study Group of the EC Subcommittee Chair's subcommittee (either via the Study Group Chair or directly to all Study Group members). For instance, a TaxoProp proposing the classification of a new bacterial virus into *Microviridae* would be forwarded to the ICTV *Microviridae* Study Group. The Study Group then performs an in-depth scientific review of the proposal and then recommends the acceptance, rejection, or need for revision to the Subcommittee Chair. The chair then once more communicates with the TaxoProp authors to improve the TaxoProp. Ideally, this process is repeated until both EC Subcommittee Chair and Study Group (per majority vote) are satisfied with the proposal or have reached a consensus that it needs to be deferred or rejected.

Study Groups therefore serve as mediators between their respective virologist constituency (their virus “subcommunity”) and the ICVCN requirements and concerns of the ICTV. However, an ICTV Study Group does not have veto power. On rare occasion, Study Group recommendations may be overruled by the Subcommittee Chair to prevent gridlock;

- (3) The TaxoProp, if appropriate with comments summarizing EC Subcommittee Chair and/or Study Group concerns or dissent, is then forwarded to the ICTV Proposals Secretary. The Proposals Secretary will once again review the documents for adherence to the ICVCN Rules and TaxoProp preparation guidelines and then post the TaxoProp for public scrutiny online on the ICTV homepage. Public comments on the TaxoProp can be posted directly to the homepage. TaxoProps submitted right before the submission deadline will be visible to the public for at least 4 weeks;
- (4) The TaxoProp will then be distributed to the entire EC for review prior to the annual EC meeting and then be presented in person to the entire EC by the Subcommittee Chair together with his or her recommendation regarding acceptability. The EC then discusses the TaxoProp and, per majority vote, decides whether the TaxoProp is approved, requires further revisions, needs to be deferred for discussion to the next EC meeting in the following calendar year, or rejected. The decision is communicated to the TaxoProp authors (and the public via the ICTV homepage) and the authors are given approximately 8 weeks to revise the TaxoProp if revisions were requested. EC voting is then repeated per email, and the TaxoProp is either accepted, deferred, or rejected;
- (5) EC-accepted TaxoProps are posted once again on the ICTV homepage for public scrutiny. Then, typically around February of each year, the entire ICTV is requested to ratify or reject the EC recommendation by majority email voting. If ratification occurs, the proposed taxonomy becomes official taxonomy and is communicated to the virology community via a VDN taxonomy update article published by the EC, via updates on the ICTV homepage, and by posting a new, so called Master Species List (MSL) on the ICTV homepage.

Virus Classification

Virus classification is the gathering of viruses into progressively more inclusive groups (lower taxa included in higher-ranked taxa) based on common genomic, phylogenetic, and phenotypic properties that, ideally, are proof of evolutionary relationships or descent. As a result, very closely related viruses may be assigned to distinct species that are all included in the same genus, whereas unrelated viruses may be assigned to distinct species that are included in separate realms. The (typically ICTV Study Group-recommended) criteria for virus classification and the demarcation criteria that are used to decide whether a virus belongs to an already established taxon or requires the establishment of a new taxon are modified continuously based on improved understanding of virus micro- and macroevolution. These criteria often differ between taxa because distinct virus groups evolve with different speeds depending on their genome types, absence or presence of replication proof-reading mechanisms, and propensity to reassort genome segments and/or to engage in horizontal gene transfer. By today, no general methodology for virus classification is commonly accepted. Some ICTV Study Groups recommend complete-genome pairwise sequence comparisons to establish percentage similarity cut-offs for taxon demarcation. However, others Study Groups focus on the phylogenies of certain hallmark features (e.g., the core domain of RNA-directed RNA polymerases), concatenated open reading frames, network analyses of structural features of hallmark proteins, or phenotypic features of viruses. Such features may include host tropism, receptor usage, or type of induced host disease. The current classification criteria for viruses of a particular taxon can be found in the most recent ICTV Report.

Differentiating Taxa and Viruses

In virus taxonomy, viruses and taxa need to be strictly differentiated. Within the ICTV, the current majority view is that taxa are man-made categories or “concepts of the mind”, i.e., they are not physical entities. Accordingly, taxa serve as conceptual perfect averages of all physical members that have ever existed, exist, or will exist independent of imperfections of individual members. Therefore, taxa, such as species or families, cannot be discovered or eradicated, and they cannot be studied, infect people, or become infected. Taxa can only be established or abolished. Taxa are represented by physical entities, i.e., viruses, which can be discovered, isolated, studied, used for infections, and eradicated. This view is not necessarily shared by everyone and the discussion about the “nature” of species and whether species can be considered physical entities continues throughout biology.

Nomenclature

Virus Taxa

In biological taxonomies, nomenclature is the process of assigning specific, ideally unique, names to organisms (e.g., “lion”) and to the taxa to which these organisms have been assigned (e.g., lion → “*Panthera leo*”). The primary function of these names is to serve as unique identifiers, i.e., labels, that aid in written and oral communication among scientists about specific groups of organisms. Since people are generally more adept to memorizing (even artificial) names than numbers even in the short term (e.g., while reading a manuscript or listening to a seminar), taxon names typically consist of words rather than numbers. In the case of taxa, specific suffixes may indicate the rank of the named taxon.

Table 2 Rank-specific suffixes for virus taxon names

Rank	Rank-specific suffixes for taxon names	Example
Genus/Subgenus	... <i>virus/virus</i>	<i>Ephemerovirus/Sarbecovirus</i>
Subfamily	... <i>virinae</i>	<i>Rubulavirinae</i>
Family	... <i>viridae</i>	<i>Geminiviridae</i>
Suborder	... <i>virineae</i>	currently none
Order	... <i>virales</i>	<i>Herpesvirales</i>
Subclass	... <i>viricetidae</i>	currently none
Class	... <i>viricetes</i>	<i>Monjiviricetes</i>
Subphylum	... <i>viricotina</i>	<i>Polyploviricotina</i>
Phylum	... <i>viricota</i>	<i>Negarnaviricota</i>
Subkingdom	... <i>virites</i>	currently none
Kingdom	... <i>virae</i>	currently none
Subrealm	... <i>vira</i>	currently none
Realm	... <i>viria</i>	<i>Riboviria</i>

Table 3 Species names in non-virologic taxonomies

Taxonomy type	Example species	Organism assigned to example species
Botanical	<i>Arabidopsis thaliana</i> (L.) Heynh.	Thale cress
Mycological	<i>Pleurotus ostreatus</i> (Jacq.) P. Kumm., 1871	Pearl oyster mushroom
Prokaryotic	<i>Escherichia coli</i>	<i>Escherichia coli</i>
Zoologic	<i>Pan troglodytes</i> Blumenbach, 1775	Common chimpanzee

The ICTV only administers the nomenclature of virus taxa, but not the nomenclature (names or abbreviations) of viruses. Virus taxon nomenclature is regulated by the rules of the ICVCN, which stipulate, for instance, that

- (1) all taxon names are to be capitalized, italicized, and never abbreviated (e.g., *Caudovirales*, *Paramyxoviridae*);
- (2) each taxon name with the exception of names of species may consist only of a single word; and
- (3) all taxon names are to be written in the standard Latin alphabet without diacritical marks.

The rank affiliation of virus taxa is identifiable by rank-specific suffixes appended to taxon names (Table 2). Exceptions are subgenera, which have names with the same suffix as genus names, and species names, which have yet to be standardized.

ICTV nomenclature differs in several aspects from other biological taxonomies. Akin to prokaryotic taxonomy, virus taxonomy requires the italicization of all taxon names, whereas only genus and species names are italicized in botanical, mycological, and zoological taxonomies. In addition, all non-virologic organismal taxonomies, such as botanical, mycological, prokaryotic, and zoological taxonomies, use the so-called Linnaean binomial species format, i.e., a species name consists of two italicized, Latinized words separated by a space. The first (capitalized) word, is the genus name whereas the second (lower-case) word is the so called-species epithet. Depending on the taxonomy, these binomial names are followed by a so-called authority, i.e., typically the last name and year (or an abbreviation thereof) of the person who first established the taxon (Table 3).

The ICTV is currently evaluating whether a similar binomial species naming format should be mandated for virus species names because as of now, virus species naming is chaotic and thereby contributing to the continuous confusion between virus species and viruses (Table 4).

However, each virus species name is required to begin with a capitalized word and all other words are to be written in lower case except if they are proper nouns. Taxon names that have been suggested or official proposed but not yet accepted by the ICTV should be indicated by quotation marks (e.g., "*Autolykiviridae*", "*Megavirales*").

Viruses

Although nomenclature of viruses is not regulated by the ICTV, the ICTV does recommend adherence to certain rules:

- (1) Virus names should not be italicized, and the individual words of a virus name should not be capitalized except if they are proper nouns (Table 4).
- (2) Virus names may be abbreviated (Table 4).
- (3) Virus names and virus name abbreviations ought to be unique.
- (4) The names of groups of viruses belonging to taxa ranked higher than species, such as all members of a family, should be derived from the taxon name (e.g., the members of *Flaviviridae* are called flaviviruses and the members of the order *Nidovirales* are called nidoviruses). Rank-specific suffixes for members of taxa ranked higher than genus have thus far only been suggested

Table 4 Non-exhaustive list of currently used species naming formats in virus taxonomy

Species format	Example	Virus assigned to example species
Identical in spelling to the name of the member virus and only differentiated from that name via italics and, sometimes, capitalization	<i>Cafeteria roenbergensis virus</i>	Cafeteria roenbergensis virus (CroV)
Mimics virus name but is more or less distinct from the name of the member virus	<i>West Nile virus</i>	West Nile virus (WNV)
	<i>Pseudomonas virus D3112</i>	Pseudomonas phage D3112
	<i>Seneca virus A</i>	Seneca Valley virus (SVV)
	<i>Severe acute respiratory syndrome-related coronavirus</i>	severe acute respiratory syndrome coronavirus (SARS-CoV)
Non-Latinized binomial with identical suffixes in both word components	<i>Senegalvirus marseillevirus</i>	Senegalvirus
Non-Latinized genus-species binomial with species epithets being numbers or letters	<i>Aalivirus A</i>	aalivirus A1 (AalV-A1)
Non-Latinized genus-species binomials	<i>Sanfarnavirus 1</i>	
Non-Latinized genus-species multinomial	<i>Alphaarterivirus equid</i>	equine arteritis virus (EAV)
Non-Latinized species-genus binomial	<i>Etaarterivirus ugarco 1</i>	Kibale red colobus virus 2 (KRCV-2)
Non-Latinized species-genus trinomial containing numbers or letters at different positions	<i>Lassa mammarenavirus</i>	Lassa virus (LASV)
	<i>Avian orthoavulavirus 1</i>	avian paramyxovirus 1 (APMV-1)
	<i>Mammalian 1 orthobornavirus</i>	Borna disease virus 1 (BoDV-1)
Non-Latinized species-genus trinomial or multinomials using words	<i>Calla lily chlorotic spot orthotospovirus</i>	calla lily chlorotic spot virus (CCSV)
	<i>Tai Forest ebolavirus</i>	Tai Forest virus (TAFV)
	<i>Lausannevirus</i>	Lausannevirus

for subfamily (“...virins”, e.g., rubulavirins), family (“...virids”, e.g., flavivirids), and order (“...virads”, e.g., nidovirads) but have not yet been widely accepted in the virology community.

The ICTV Virus Metadata Resource (VMR) is a developing resource in which virus names, virus name abbreviations, exemplar isolates, and GenBank accession numbers can be located in association with the taxa to which particular viruses are currently assigned.

Future Developments

Several steps are currently under discussion that, potentially, could bring virus taxonomy in harmony with other biological taxonomies. These include the operational definition of the term “virus” for improved delineation of the ICTV mandate (e.g., if plasmids or satellite RNAs or endogenous virus-like elements were considered viruses, then they ought to be classified); the possible inclusion of virus nomenclature in the ICTV mandate; the development of rank-specific suffixes for vernacular names for the groups of viruses belonging to taxa of all ranks; and the establishment of a mandated Linnaean (Latinized binomial) species naming format.

The acceptance of *Negamaviricota* based on the analysis depicted in Fig. 1 already points toward a path forward to classify all viruses of *Riboviria* into principal/primary ranks, with each depicted branch likely representing a unique phylum. In addition, the viruses of several taxa, such as animal viruses of *Herpesvirales* and the prokaryotic viruses of *Caudovirales*, have long been considered as related and therefore should be joined at higher ranks. Consequently, in 2020, Koonin *et al.* outlined the first systematic top-down virus taxonomy since the LHT system. Accordingly, a total of four realms would be needed to classify most currently known viruses at all available principle/primary ranks. The establishment of the realms was proposed to be justified based on the discovery of specific viral hallmark genes that can be used to evolutionary connect all their constituent viruses: (1) RdRps and RNA-directed DNA polymerases (RdDps) for most RNA viruses (*Riboviria*); (2) replication initiator proteins (Reps) of the His-hydrophobic-His (HUH) superfamily for all single-stranded DNA viruses; (3) vertical jelly-roll capsid proteins for many double-stranded DNA viruses; and (4) HK97-fold capsid proteins for viruses classified in *Caudovirales* and *Herpesvirales*. If accepted by the ICTV, virus classification would finally begin to resemble the classification schemes of cellular organisms. Nevertheless, the establishment of four separated realms would still attest to the polyphyly of viruses, although it suggests that viruses emerged only a few times in evolutionary history (viruses would be “mistletoe on the tree of life”).

Finally, procedural improvements might ensure improved communication of the ICTV with the virology community and thereby increased efficiency of the ICTV. Such improvements may include automatic classification algorithms, fully online TaxoProp submissions and evaluation systems, harmonization of classification criteria across taxa, and faster ratification cycles.

Acknowledgement

The content of this publication does not necessarily reflect the views or policies of the US Department of Health and Human Services or of the institution and company affiliated with the author. The author performed this work as an employee of Battelle Memorial Institute under its prime contract with NIAID, under Contract No. HHSN2722002000161.

Further Reading

- Abudurexiti, A., Adkins, S., Alioto, D., *et al.*, 2019. Taxonomy of the order *Bunyavirales*: Update 2019. *Archives of Virology* 164, 1949–1965.
- Amarasinghe, G.K., Aylón, M.A., Bào, Y., *et al.*, 2019. Taxonomy of the order *Mononegavirales*: Update 2019. *Archives of Virology* 164, 1967–1980.
- Baltimore, D., 1971. Expression of animal virus genomes. *Bacteriological Reviews* 35, 235–241.
- Bawden, F.C., 1941. Nomina ad infinitum. *Chronica Botanica* 6, 17–18.
- Beijerinck, M.W., 1898. Over een Contagium vivum fluidum als oorzaak van de Vlekziekte der Tabaksbladen. *Verhandelingen VII*, 229–235.
- Bennett, C.W., 1939. The nomenclature of plant viruses. *Phytopathology* 29, 422–430.
- Blitvich, B.J., Beaty, B.J., Blair, C.D., *et al.*, 2018. Bunyavirus taxonomy: Limitations and misconceptions associated with the current ICTV criteria used for species demarcation. *The American Journal of Tropical Medicine and Hygiene* 99, 11–16.
- Bradley, D.E., 1965. The morphology and physiology of bacteriophages as revealed by the electron microscope. *Journal of the Royal Microscopical Society* 84, 257–316.
- Brandes, J., Wetter, C., 1959. Classification of elongated plant viruses on the basis of particle morphology. *Virology* 8, 99–115.
- Calisher, C.H., Briesse, T., Brister, J.R., *et al.*, 2019. Strengthening the interaction of the virology community with the international committee on taxonomy of viruses (ICTV) by linking virus names and their abbreviations to virus species. *Systematic Biology* 68, 828–839.
- Calisher, C.H., Mahy, B.W., 2003. Taxonomy: Get it right or leave it alone. *The American Journal of Tropical Medicine and Hygiene* 68, 505–506.
- Cooper, P.D., 1961. A chemical basis for the classification of animal viruses. *Nature* 190, 302–305.
- Dherelle, M.F., 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 165, 373–375.
- Drebot, M.A., Henchal, E., Hjelte, B., *et al.*, 2002. Improved clarity of meaning from the use of both formal species names and common (vernacular) virus names in virological literature. *Archives of Virology* 147, 2465–2472.
- Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), 2005. *Virus taxonomy*. Eighth Report of the International Committee on Taxonomy of Viruses. San Diego, CA: Academic Press.
- Fenner, F., 1976. Classification and nomenclature of viruses. Second report of the International Committee on Taxonomy of Viruses. *Intervirology* 7, 1–115.
- Francki, R.L.B., Fauquet, C.M., Knudson, D.L., Brown, F., 1991. *Archives of virology supplement*. In: *Classification and Nomenclature of Viruses*. Fifth report of the International Committee on Taxonomy of Viruses 2. Springer-Verlag.
- Gorbalenya, A.E., 2018. Increasing the number of available ranks in virus taxonomy from five to ten and adopting the Baltimore classes as taxa at the basal rank. *Archives of Virology* 163, 2933–2936.
- Hamparian, V.V., Hillemann, M.R., Kettler, A., 1963. Contributions to characterization and classification of animal viruses. *Experimental Biology and Medicine* 112, 1040–1050.
- Holmes, F.O., 1948. Order virales – The filterable viruses. In: *Breed, R.S., Murray, E.G.D., Hitchens, A.P. (Eds.), Bergey's Manual of Determinative Bacteriology*, sixth ed. Baltimore, MD: The Williams & Wilkins Company.
- Horne, R.W., Wildy, P., 1961. Symmetry in virus architecture. *Virology* 15, 348–373.
- International Association for Plant Taxonomy. International Code of Nomenclature for algae, fungi, and plants (Melbourne Code). Oberreifenberg: Koeltz Scientific Books, (Regnum Vegetabile; 154).
- International Commission on Zoological Nomenclature. International Code of Zoological Nomenclature, fourth. ed. London: The International Trust for Zoological Nomenclature.
- International Committee on Systematic Bacteriology, 1992. International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision. Washington, DC: ASM Press.
- International Committee on Taxonomy of Viruses Executive Committee, 2020. The new scope of virus taxonomy. Partitioning the virosphere into 15 hierarchical ranks. *Nat Microbiol.* 5 (5), 668–674. doi:10.1038/s41564-020-0709-x.
- Iwanowsky, D., 1892. Über die Mosaikkrankheit der Tabakspflanze. *Bulletin de l'Académie Impériale des Sciences de St.-Petersbourg* 35, 67–70.
- Kausche, G.A., Pfankuch, E., Ruska, H., 1939. Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop. *Naturwissenschaften* 27, 292–299.
- King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), 2011. *Virus Taxonomy – Ninth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier/Academic Press.
- Koonin, E.V., Dolja, V.V., Krupovic, M., *et al.*, 2020. Global organization and proposed megataxonomy of the virus world. *Microbiology and Molecular Biology Reviews* 84 (2), e00061–19. doi:10.1128/MMBR.00061-19.
- Kuhn, J.H., Jahrling, P.B., 2010. Clarification and guidance on the proper usage of virus and virus species names. *Archives of Virology* 155, 445–453.
- Kuhn, J.H., Wolf, Y.I., Krupovic, M., *et al.*, 2019. Classify viruses – The gain is worth the pain. *Nature* 566, 318–320.
- Loeffler, Frosch, 1897. Summarischer Bericht über die Ergebnisse der Untersuchungen der Kommission zur Erforschung der Maul- und Klauenseuche bei dem Institute für Infektionskrankheiten in Berlin. *Centralblatt für Bakteriologie, Parasitenkunde und Infektionskrankheiten*. Erste Abteilung: Medizinisch-hygienische Bakteriologie und tierische Parasitenkunde XXII, 257–259.
- Loeffler, Frosch, 1898. Berichte der Kommission zur Erforschung der Maul- und Klauenseuche bei dem Institut für Infektionskrankheiten in Berlin. *Centralblatt für Bakteriologie, Parasitenkunde und Infektionskrankheiten*. Erste Abteilung: Medizinisch-hygienische Bakteriologie und tierische Parasitenkunde XXIII, 371–391.
- Lwoff, A., 1953. Lysogeny. *Bacteriological Reviews* 17, 269–337.
- Lwoff, A., 1957. The concept of virus. *The Journal of General Microbiology* 17, 239–253.
- Lwoff, A., Horne, R., Tournier, P., 1962a. A system of viruses. *Cold Spring Harbor Symposia on Quantitative Biology* 27, 51–55.
- Lwoff, A., Horne, R.W., Tournier, P., 1962b. Un système des virus. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 254, 4225–4227.
- Lwoff, A., Tournier, P., 1966. The classification of viruses. *Annual Review of Microbiology* 20, 45–74.
- Matthews, R.E.F., 1979. Third report of the international committee on taxonomy of viruses. Classification and nomenclature of viruses. *Intervirology* 12, 129–296.
- Matthews, R.E.F., 1982. Classification and nomenclature of viruses. Fourth report of the International Committee on Taxonomy of Viruses. *Intervirology* 17, 1–199.
- Mayer, A., 1886. Ueber die Mosaikkrankheit des Tabaks. *Die landwirtschaftlichen Versuchs-Stationen* 32, 451–467.
- Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., *et al.*, 1995. *Virus Taxonomy*. Sixth report of the International Committee on Taxonomy of Viruses. Vienna: Springer-Verlag.
- Postler, T.S., Clawson, A.N., Amarasinghe, G.K., *et al.*, 2017. Possibility and challenges of conversion of current virus species names to Linnaean binomials. *Systematic Biology* 66, 463–473.
- Pringle, C.R., 1991a. The 20th meeting of the executive committee of the international committee on virus taxonomy. Virus species, higher taxa, A universal virus database, and other matters. *Archives of Virology* 119, 303–304.
- Pringle, C.R., 1991b. The order *Mononegavirales*. *Archives of Virology* 117, 137–140.
- Reed, W., Carroll, J., Agramonte, A., Lazear, J.W., 1900. The etiology of yellow fever – A preliminary note. *Public Health Papers and Reports* 26, 37–53.
- Van Regenmortel, M.H.V., 2003. Viruses are real, virus species are man-made, taxonomic constructions. *Archives of Virology* 148, 2481–2488.
- Van Regenmortel, M.H.V., 2007. Virus species and virus identification: Past and current controversies. *Infection, Genetics and Evolution* 7, 133–144.
- Van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., *et al.*, 2000. *Virus Taxonomy*. Seventh report of the International Committee on Taxonomy of Viruses. San Diego, California, USA: Academic Press.
- Van Regenmortel, M.H.V., Maniloff, J., Calisher, C.H., 1991. The concept of virus species. *Archives of Virology* 120, 313–317.
- Rivers, T.M., 1927. Filterable viruses. A critical review. *Journal of Bacteriology* 14, 217–258.

- Siddell, S.G., Walker, P.J., Lefkowitz, E.J., *et al.*, 2019. Additional changes to taxonomy ratified in a special vote by the international committee on taxonomy of viruses (October 2018). *Archives of Virology* 164, 943–946.
- Siddell, S.G., Walker, P.J., Lefkowitz, E.J., *et al.*, 2020. Binomial nomenclature for virus species: A consultation. *Archives of Virology*. Available at: <https://link.springer.com/article/10.1007/s00705-019-04477-6>.
- Simmonds, P., Adams, M.J., Benkő, M., *et al.*, 2017. Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology* 15, 161–168.
- Stanley, W.M., 1935. Isolation of a crystalline protein possessing the properties of tobacco-mosaic virus. *Science* 81, 644–645.
- Tauroso, N.M., Shelokov, A., 1967. Arboviruses – A problem in classification. *Archiv für die gesamte Virusforschung* 22, 273–279.
- Vetten, H.J., Haenni, A.-L., 2006. Taxon-specific suffixes for vernacular names. *Archives of Virology* 151, 1249–1250.
- Walker, P.J., Siddell, S.G., Lefkowitz, E.J., *et al.*, 2019. Changes to virus taxonomy and the international code of virus classification and nomenclature ratified by the international committee on taxonomy of viruses (2019). *Archives of Virology* 164, 2417–2429.
- Wildy, P., 1962. Classifying viruses at higher levels: Symmetry and structure of virus particles as criteria. *Symposium of the Society for General Microbiology XII*, 145–163.
- Wildy, P., 1971. *Classification and Nomenclature of Viruses*. First report of the International Committee on Nomenclature of Viruses. Basel: S. Karger.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9, e02329 (18).

Relevant Websites

- <https://talk.ictvonline.org/information/w/ictv-information/383/ictv-code>
ICTV Code (International Code of Virus Classification and Nomenclature).
- <https://talk.ictvonline.org/files/master-species-lists/m/msl/8266>
ICTV Master Species List 2018b.v2.
- <https://www.microbiologyresearch.org/search?value1=ictv+virus+taxonomy+profiles&option1=fulltext>
ICTV Virus Taxonomy Profiles.
- <https://talk.ictvonline.org>
International Committee on Taxonomy of Viruses (ICTV).
- <https://talk.ictvonline.org/files/proposals/>
Pending Proposals.
- <https://talk.ictvonline.org/files/taxonomy-proposal-templates/>
Taxonomy Proposal Templates.
- <https://talk.ictvonline.org/taxonomy/>
Taxonomy
International Committee on Taxonomy of Viruses.
- <https://talk.ictvonline.org/information/w/ictv-information/382/the-statutes-of-the-ictv>
The Statutes of the ICTV.
- <https://talk.ictvonline.org/information/w/virology-division-news>
Virology Division News.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/
Virus Taxonomy: The Classification and Nomenclature of Viruses.
- <https://ictv.global/taxonomy/vmr/>
VMR: Virus Metadata Resource.

The Greater Virus World and Its Evolution

Eugene V Koonin, National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, United States
and National Institutes of Health, Bethesda, MD, United States

Valerian V Dolja, Oregon State University, Corvallis, OR, United States

© 2021 Elsevier Ltd. All rights reserved.

Introduction

All cellular organisms, with the possible exception of some intracellular parasitic bacteria, host diverse obligate intracellular parasites and symbionts including viruses and virus-like mobile genetic elements (MGEs). Viruses are the most abundant biological entities on earth: the total number of virus particles at any given moment reaching the mind-boggling value of 10^{31} , which is about an order of magnitude greater than the total number of cells. Lysis of cells, primarily, bacteria, resulting from virus infection is a major biogeochemical factor. On par with their physical abundance, viruses also represent the largest pool of genetic diversity in the biosphere that constantly supplies genes to cellular life forms. Theoretical arguments suggest that genetic parasites inevitably emerge in even the simplest replicator systems. The ubiquity of viruses in the biosphere and their enormous diversity, along with the theoretical argument on the inevitable emergence of genetic parasites in any replicator system, imply viruses and other MGEs have been major players in the evolution of life for its entire duration.

A defining feature of the virus world is the diversity of the replication and expression strategies among viruses and MGEs (Fig. 1). All possible forms of nucleic acids and their interconversions are employed in the virus world, in a dramatic contrast with the uniform route of information transfer in cellular organisms (Fig. 1). In 1971, David Baltimore published a classification scheme that divided all viruses into 6 classes (a 7th class has been added subsequently). These “Baltimore classes” of viruses differ by the form of nucleic acid that is incorporated into virions (the virus genome):

- (1) dsDNA viruses that mimic the replication-expression strategy of cellular life forms;
- (2) ssDNA viruses that typically replicate via a rolling circle mechanism;
- (3) double-stranded (ds) RNA viruses;
- (4) positive-sense (+)RNA viruses in which the genome is a single-stranded (ss) RNA of the same polarity as the mRNA(s) for virus proteins;
- (5) negative-strand (−)RNA viruses in which the genome is an ssRNA complementary to the mRNA (or, in some groups of viruses, an ambisense RNA);
- (6) RNA-containing reverse-transcribing viruses with a (+)RNA genome that, however, is replicated via a DNA intermediate that is produced by reverse transcription of the genome; and
- (7) DNA-containing reverse-transcribing viruses that also replicate via reverse transcription but package the dsDNA form or an RNA-DNA hybrid into virions.

The Baltimore classes widely differ in their abundance, the diversity of the constituent viruses and other MGEs, and spread among cellular life forms. The two most abundant and diverse of the Baltimore classes are dsDNA viruses that dominate the prokaryotic virome and (+)RNA viruses that are most common in eukaryotes. However, each of the remaining 5 Baltimore classes also is abundant in at least some hosts as exemplified by the prominence of reverse-transcribing viruses in animals.

For several decades, the “Baltimore classes” have been used as an informal highest rank of virus classification. However, the partitioning of viruses into these classes is not, intrinsically, an evolutionary classification. The actual evolutionary relationships among viruses and MGE are studied by the same methods of molecular phylogenetic and phylogenomics that are employed for the study of the relationships among cellular life forms. However, major differences exist between the organizations of the cellular and viral genetic spaces. Even in cellular organisms, the genomes evolve in highly dynamic regimes and display remarkable plasticity, especially in the case of prokaryotes where horizontal gene transfer resulting in extensive gene gain as well as gene loss is a dominant evolutionary process. Nevertheless, all cellular life forms share about 100 universal genes that, mostly, encode protein and RNA components of the protein translation system. The existence of these universal, core genes represents irrefutable evidence of a common ancestry of all cellular life and provides the material for the construction of phylogenetic trees that reflect the central vertical trend in the evolution of cellular organisms and can be used as the framework for evolutionary reconstructions.

In contrast, viruses and MGE lack universal genes and even genes that are conserved in the majority of the genomes. Clearly, a single last common ancestor of viruses never existed, and instead, viruses have evolved on at least several independent occasions. Even when a gene is shared by diverse groups of viruses, it cannot be automatically assigned to their last common ancestor. The uncertainty comes from the fact that viruses continuously capture genes from their cellular hosts, and many cases of independent, convergent acquisition of homologous genes by widely different viruses have been identified. In addition, due to the typical fast evolution of virus genomes, homologous genes in viruses of different groups are often highly diverged such that detection of homologs requires the use of the most powerful methods for protein sequence analysis, and yet, comprehensive identification of homologs remains elusive.

All the problems notwithstanding, large scale evolutionary reconstruction for viruses is both feasible and informative. Although any search for a single common ancestral virus is futile, extensive analyzes of the evolution of large groups of viruses have proved

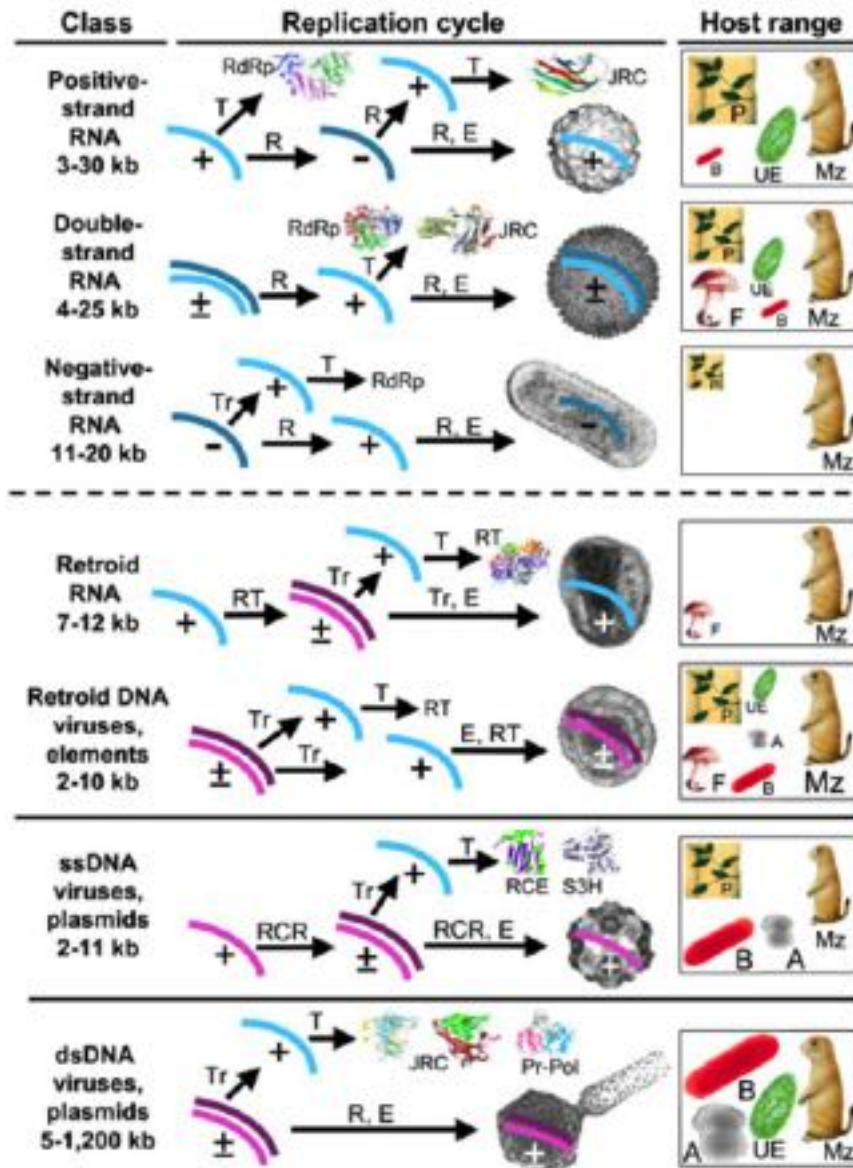


Fig. 1 Information transfer in viruses and the Baltimore classes. The replication-expression-encapsidation scheme, with some of the virus hallmark proteins involved, and the host range are shown for each class. Abbreviations in the schemes: E, encapsidation; JRC, jelly roll capsid (protein); Pr-Pol, protein-primed polymerase; R, replication; RCR, rolling circle replication; RdRp, RNA-dependent RNA polymerase; S3H, superfamily 3 helicase; T, translation; Tr, transcription. Abbreviations in the host range icons: A, archaea; B, bacteria; Mz, metazoa; UE, unicellular eukaryotes; P, plants. From Koonin, E.V., Senkevich, T.G., Dolja, V.V., 2006. The ancient virus world and evolution of cells. *Biology Direct* 1, 29, reproduced under Creative license.

productive. The principal approach employed in such studies is phylogenetic analysis of genes that are conserved across broad ranges of viruses, known as Virus Hallmark Genes (VHGs), and using the VHG phylogenies as a scaffold for the reconstruction of the history of gain and loss of other genes. A complementary approach involves dissection of networks of virus genes sharing into distinct modules. During the last few years, these approaches, perhaps, unexpectedly, have produced fairly detailed and well-substantiated evolutionary reconstructions for all Baltimore classes.

Virus Hallmark Genes

Although there are no universally conserved genes among viruses and MGEs, the distribution of the gene frequencies in viruses is strongly skewed: most of the genes are found only in small groups of viruses but a small set of VHGs includes widespread genes. Not unexpectedly, VHGs encode proteins that are responsible for the key functions in virus replication and virion structure formation. Most of the VHGs are limited to a single Baltimore class but several show a broader spread, spanning two or even more

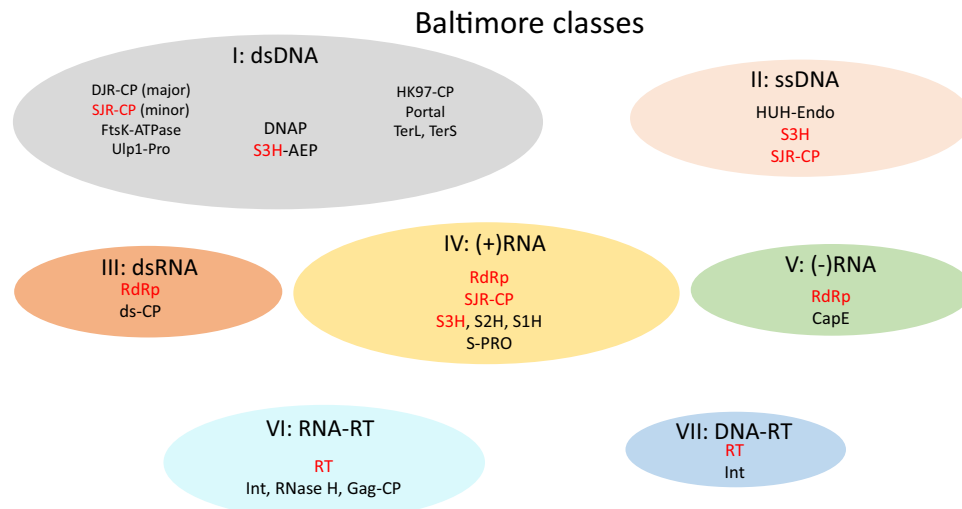


Fig. 2 Virus hallmark genes and the Baltimore classes. For each Baltimore class, the characteristic hallmark genes (not necessarily present in all viruses of the class) are indicated. The “super-hallmark” genes are highlighted in red. Abbreviations: AEP, archaeo-eukaryotic primase; CapE, capping enzyme; DJR-CP, double jelly roll capsid protein; DNAP, DNA polymerase (family B); ds-CP, capsid protein of dsRNA viruses; FtsK-ATPase, DNA packaging ATPases of the FtsK family; Gag-CP, capsid protein of reverse-transcribing viruses, domain of Gag polyprotein; HK97-CP, HK97-like capsid protein; HUH-Endo, rolling circle replication initiation endonuclease (HUH family); Int, integrase; RdRp, RNA-dependent RNA polymerase; RT, reverse transcriptase; SJR-CP, single jelly roll capsid protein; TerL, terminase, large subunit; TerS, terminase, small subunit; Ulp1-Pro, thiol protease homologous to Ulp1 family of deubiquitinating enzymes; S1H, S2H, S3H, helicases of superfamilies 1, 2, and 3.

of the Baltimore classes (Fig. 2). These “super-VHGs” are: (1) RdRps that appear to comprise a monophyletic group of Palm domain polymerases and links the 3 Baltimore classes of RNA viruses (III-V), (2) RTs that unify the two Baltimore classes of reverse-transcribing viruses (and related MGEs), those with RNA and DNA genomes (VI and VII), and is also homologous to the RdRp, (3) Superfamily 3 helicases (S3H), distinct helicases found (almost) exclusively in MGEs including diverse (+)RNA viruses, most of the ssDNA viruses, and many dsDNA viruses, (4) Single jelly-roll capsid proteins (SJR-CPs), the dominant type of capsid proteins in positive-sense ssRNA viruses and ssDNA viruses (Baltimore classes II and IV); (5) Double jelly-roll capsid proteins (DJR-CPs), which are most common among dsDNA viruses and represented also in some ssDNA viruses; and (6) rolling-circle replication initiation endonucleases (RCREs) found in the vast majority of ssDNA viruses (Baltimore class II) but also in some dsDNA viruses (Baltimore class I).

The VHGs appear to widely differ with respect to evolutionary origins. The RdRp, RT, and RCRE, in all likelihood, descend directly from the primordial pool of replicators, given the lack of orthologs of non-viral origin in cellular life forms; S3Hs and SJR-CPs appear to be ancient recruitments of cellular proteins; the DJR-CPs apparently evolved from SJR-CPs in an ancient virus lineage. Regardless of their specific origins, the VHGs undoubtedly entered the virus world and became essential for virus reproduction at early stages of evolution. Therefore, analysis of the evolutionary histories of the VHGs is indispensable for understanding evolution of viruses and the organization of the virus world.

Evolutionary Relationships Within and Between the Baltimore Classes

RNA viruses

The three Baltimore classes that consist of RNA viruses share a single VHG, the RdRp. The sequences of the RdRps are extremely divergent between the three classes. Nevertheless, extensive comparisons of thousands of diverse RdRp and RT sequences as well as the structures of these polymerases from diverse RNA viruses and retroelements have led to an unequivocal conclusion on their monophyly among the Palm domain enzymes. Phylogenetic analyzes of protein families with such a high level of divergence as observed among the RdRps and RTs require extreme caution in interpretation. All the caveats notwithstanding, the tree of the RdRps, in which the most highly conserved RTs, those of group II introns, were included as the outgroup, shows a distinct structure (Fig. 3). The RdRps split into 5 major branches, two of which consist solely of (+)RNA viruses, one combines (+)RNA and dsRNA viruses, the fourth one includes dsRNA viruses, and the fifth one encompasses all (−)RNA viruses (Fig. 3). The monophyly of each of the 5 branches is strongly supported whereas the relationships between the branches are much less clear. The RdRp is the signature of RNA viruses that binds them all together but other genes mark each of the 5 branches and the major sub-branches within those. To confound the evolutionary reconstruction, multiple gene exchanges between RNA viruses from different branches seem to have occurred.

Branch 1 in the RdRp tree consists of the (+)RNA bacteriophages (family *Leviviridae*), one of the 2 groups of RNA viruses that infect prokaryotes, namely, a narrow range of bacteria. In addition to the leviviruses, this branch includes the RdRps of their

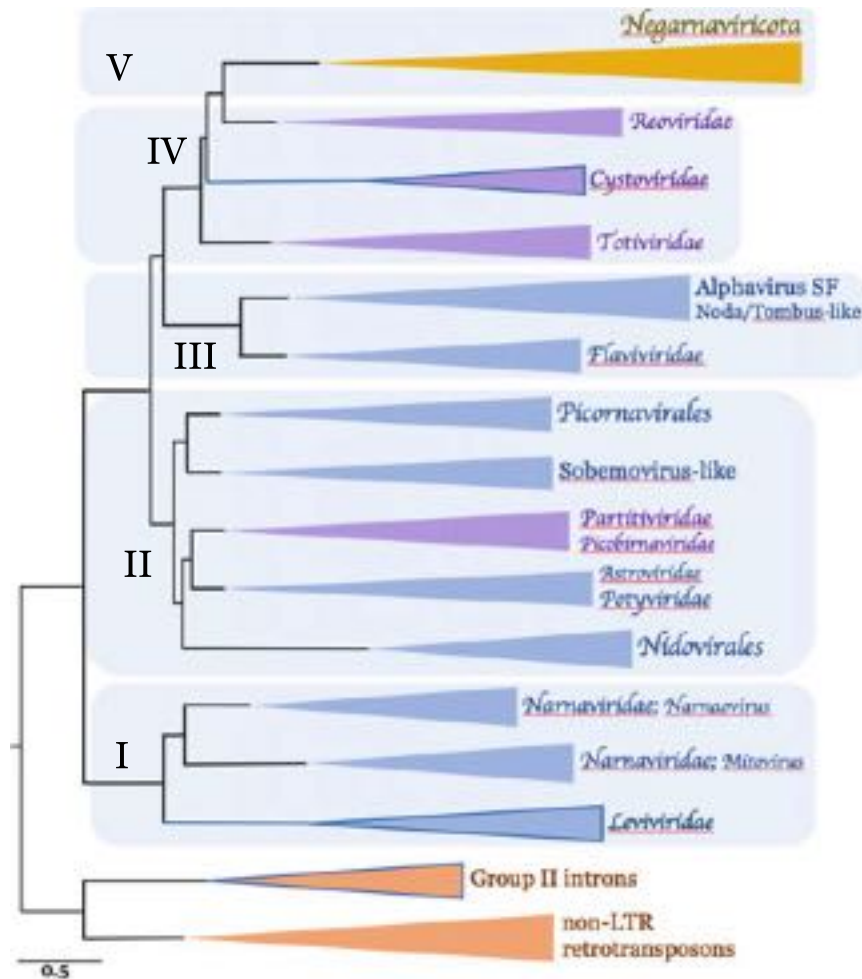


Fig. 3 Schematic phylogenetic tree of the RNA-dependent RNA polymerases of RNA viruses. The tree is rooted with reverse transcriptases of group II introns and non-LTR retrotransposons. The 3 Baltimore classes of RNA viruses are color-coded: blue, (+)RNA; magenta, dsRNA; yellow, (–)RNA. Negarnaviricota, for (–)RNA viruses, is the first formally established virus phylum.

descendants that infect eukaryotes. Apparently, the leviviruses first gave rise to “minimal”, naked replicons that only encode the RdRp and are known as narnaviruses and mitoviruses (both included in the family *Narnaviridae*) although these are not bona fide viruses because they do not form virions. As could be expected of direct descendants of prokaryotic viruses, mitoviruses reproduce in the mitochondria and, most likely, were captured by early eukaryotes from the mitochondrial endosymbiont. Subsequently, these agents were lost from most eukaryotic lineages, surviving only in fungi and some plants. Narnaviruses escaped from mitochondria to the cytosol of eukaryotic cells, and their descendants reentered the world of true viruses by acquiring the gene encoding SJR-CP and, in some lineages, also a movement protein and a helicase. These derivatives of the levivirus-related naked replicons are known as ourmia-like viruses that, although initially thought to be confined to plants, appear to infect invertebrates as well as shown in recent metagenomic studies. Thus, the reconstructed evolutionary scenario for this branch includes a loss of the structural module of the leviviruses (genes for the unique major and minor capsid proteins), yielding naked replicons that subsequently captured an unrelated capsid protein, SJR-CP, from plant viruses.

Branch 2 in the RdRp phylogeny encompasses a vast collection of diverse (+)RNA and dsRNA viruses that are associated with the previously described “picorna-like supergroup”, after picornaviruses, a large, well-characterized family of animal viruses that includes many human pathogens, such as poliovirus. Viruses in this branch have enormously divergent genome lengths, gene compositions, and genome organizations, and produce several types of virions with dissimilar morphologies. This branch includes both viruses with the largest known RNA genomes, nidoviruses, and those with some of the smallest genomes, such as sobemoviruses. Although no genes other than the RdRp are shared by all viruses in this branch, the two hallmark genes encoding SJR-CPs and trypsin-like proteases, appear to be ancestral features. Evolution of the viruses in this branch seems to have involved parallel capture of helicases of three different superfamilies and of a variety of additional genes, most notably, in the nidovirus lineage.

Arguably, the most notable aspect of this RdRp branch is that it combines two Baltimore classes, with a clade of dsRNA viruses that includes picobirnaviruses and partitiviruses embedded among (+)RNA viruses. Both these families of dsRNA viruses have

small, bipartite genomes, one segment encoding the RdRp and the other one encoding the CP that appears to be homologous to the CPs of other dsRNA viruses such as reoviruses. This clade of dsRNA viruses also includes naked RNA replicons that only encode the RdRp, reproduce inside algal mitochondria or chloroplasts, and use the mitochondrial genetic code, thus resembling, in terms of both genome content and life style, mitoviruses of Branch 1. Evolution of this clade in Branch 2 might have followed a similar scenario as the evolution of ourmia-like viruses, namely, combination of an RdRp-encoding capsid-less genetic element inherited from prokaryotes with a capsid protein gene from a eukaryotic virus, in this case, a dsRNA virus.

Branch 3 in the RdRp tree encompasses about as much diversity of gene compositions and genome organizations as Branch 2 but only includes (+)RNA viruses. The branch contains two clades that consist of well-characterized viruses of plants and animals, some of these with large and highly complex genomes (by RNA virus standards), known as the “alpha-like supergroup” and the “flavi-like supergroup”, respectively (Fig. 3). The “alpha-like supergroup” encompasses an enormous range of viruses infecting plants and a smaller variety of animal viruses that share a signature block of VHGs encoding capping enzyme, superfamily 1 helicase, and RdRp. The conservation of this gene array testifies to the monophyly of the “alpha-like supergroup lineage”. The “flavi-like supergroup” is smaller and, until recently, included only the animal viruses of the family *Flaviviridae*. However, metaviromics studies have substantially expanded the group by adding a number of viruses with diverse genome organizations that infect either animals or plants. Similar to the “alpha-like supergroup” viruses, most of the viruses in the “flavi-like supergroup” encode helicases and capping enzymes but these belong to protein families distinct from those in the “alpha-like supergroup”, and clearly, have been acquired independently. Apart from the “alpha-like and flavi-like supergroups”, Branch 3 includes an assortment of lineages that all consist of short-genome viruses that encode only RdRp, CPs and, in some cases, one or two additional proteins. Two distinct, well-characterized clades in this assemblage are noda-like and tombus-like viruses but many more groups identified in metaviromes have joined the group.

Branch 4 in the RdRp tree consists entirely of dsRNA viruses. Similarly to the (+)RNA viruses from Branches 2 and 3, the dsRNA viruses in Branch 4 spread over a wide range of genomic length and complexity, from the minimalist genomes of totiviruses that encode only RdRp and CP to the highly complex reovirus genomes that consist of up to 12 genomic RNA segments containing numerous genes. The monophyly of the dsRNA viruses in this branch is supported by the conservation of the CP structure and capsid organization, with 60 homo- or heterodimers of the CP organized on a unique pseudo $T = 2$ lattice. The primitive toti-like viruses and the complex reo-like viruses form the two major clades within Branch 4. The reo-like clade also includes cystoviruses, the only known group of dsRNA viruses that infect prokaryotes. Thus, the pattern of apparent evolution of viruses of eukaryotes from prokaryotic ancestors that is observed in Branch 1 and, tentatively, in Branch 2 is recapitulated among the dsRNA viruses of Branch 4.

Branch 5 that includes all known (−)RNA viruses is the sister group of the reo-like clade within Branch 4. The monophyly of the (−)RNA virus RdRps is unequivocally supported by phylogenetic analysis, whereas the support for the position of the (−)RNA branch amidst the dsRNA viruses is weaker. However, this placement of (−)RNA viruses is compatible with the structural comparison of the (−)RNA virus and dsRNA virus RdRps. This tree topology perfectly fits the general logic of virus evolution: from the ubiquitous (at least, in eukaryotes) (+)RNA viruses to the less common dsRNA viruses, to the even rarer (−)RNA viruses. The dsRNA viruses and (−)RNA viruses share an essential aspect of the reproduction cycle, namely, packaging of the replication and transcription machineries into virus particles which allows virus reproduction in an infected cell to kick off by transcribing the virion RNA. This compartmentalization of the virus genomes was likely driven by selection for avoidance of the eukaryotic RNAi defense.

Branch 5 splits into two major clades, one of which includes, primarily, viruses with nonsegmented genomes that encode their own capping enzymes, whereas the other clade consists, mostly, of viruses with segmented genomes that use the mRNA “cap-snatching” mechanism (Fig. 3).

Reverse-Transcribing Viruses

Reverse-transcribing viruses are widespread and in many cases, highly abundant in animals, fungi, plants, and protists but, to the best of our current knowledge, are absent in bacteria and archaea. However, bacteria and archaea carry several classes of non-viral, capsid-less retroelements, such as group II introns, retrons and diversity-generating retroelements, whereas eukaryotes harbor several varieties of so-called non-LTR retrotransposons. Similarly to the RNA viruses, retroelements and reverse-transcribing viruses all share a single conserved gene that encodes the key replication enzyme of all these elements, RT. Thus, like the RdRp in the case of the RNA viruses, RT is the only phylogenetic marker for the evolutionary analysis of reverse-transcribing viruses. The phylogenetic tree of the RTs consists of 5 major branches, 4 of which include non-viral retroelements, and 5th, strongly supported branch encompasses all reverse-transcribing viruses. Thus, all reverse-transcribing viruses appear to be monophyletic, evolving from retrotransposons on a single occasion, most likely at the onset of eukaryotic evolution or shortly thereafter.

The virus branch of the RT tree consists of four clades, three of which correspond to the families *Hepadnaviridae*, *Pseudoviridae* and *Belpaoviridae*, whereas the fourth one unites 3 families: *Metaviridae*, *Caulimoviridae* and *Retroviridae*. The viruses in all these families, except for hepadnaviruses, share not only the RT but also two core components of the virions, the capsid and nucleocapsid proteins (the latter apparently lost in members of the *Spumaretrovirinae*, one of the retroviral subfamilies). Belpaoviruses, caulimoviruses, metaviruses, pseudoviruses, and retroviruses employ the same mechanism of replication priming, namely, utilization of host tRNA molecules as primers for reverse genome transcription. In contrast, hepadnaviruses use a protein-priming mechanism that is mediated by the terminal protein domain of the RT. Thus, it appears most likely that the five families of reverse-transcribing viruses are monophyletic, to the exclusion of *Hepadnaviridae*, apparently, the earliest branching family of reverse-transcribing viruses.

Viruses of the families *Hepadnaviridae* and *Caulimoviridae* (often collectively called “pararetroviruses”) that package dsDNA genomes into the virions and accordingly form Baltimore class VII are widely separated in the RT tree. Thus, Baltimore class VII is apparently polyphyletic, which makes class VI a paraphyletic group.

Single-stranded DNA viruses

The great majority of ssDNA viruses have small, circular genomes that replicate via the rolling circle mechanism. The ssDNA viruses infect a broad range of bacteria and some archaea as well as eukaryotes. Although these viruses have been missed in many early studies due to the miniscule size of their virions and genomes, metaviromics has shown that ssDNA viruses are diverse and abundant in various habitats and are a major component of the earth’s virome. The ssDNA viruses include 4 families of bacterial and archaeal viruses, *Inoviridae*, *Microviridae*, *Pleolipoviridae* and *Spiraviridae*, and 9 families of eukaryotic viruses: *Anelloviridae*, *Bacilladnaviridae*, *Bidnaviridae*, *Circoviridae*, *Geminiviridae*, *Genomoviridae*, *Nanoviridae*, *Parvoviridae* and *Smacoviridae*. Unusually, different members of the family *Pleolipoviridae* encapsidate either ssDNA or dsDNA genomes whereas parvoviruses and bidnaviruses have linear genomes. Whereas parvoviruses replicate via a rolling hairpin mechanism, a variant of the classical rolling circle, bidnaviruses encode their own protein-primed family B DNA polymerases.

The ssDNA viruses have “minimal” genomes which typically encode only a single replicative protein and a single structural protein, and in some cases, a few additional small, poorly characterized proteins. With the exception of anelloviruses, bidnaviruses, spiraviruses, and some inoviruses, the replicative protein of ssDNA viruses is the so-called HUH endonuclease that cuts genomic DNA at a specific site and initiates rolling circle (or rolling hairpin) replication. All eukaryotic ssDNA viruses that possess a HUH endonuclease also encode a S3H helicase that most often is fused to the endonuclease to form a two-domain Rep proteins. In contrast, ssDNA viruses of prokaryotes encode solitary HUH endonucleases. Apart from inoviruses, pleolipoviruses, spiraviruses, and the recently discovered FLiP virus that is expected to form a new family, ssDNA viruses have capsids made of SJR-MCPs.

Comparative analysis of the HUH endonucleases, S3Hs and CPs of ssDNA viruses points to multiple, chimeric origins of these viruses. The Rep proteins of eukaryotic ssDNA viruses appear to have evolved from homologous Rep proteins of small, rolling circle-replicating bacterial plasmids that also consist of HUH and S3H domains. Phylogenetic analysis of Rep proteins shows that viral replication machinery evolved from those of plasmids at least 3 times, independently. In parallel, HUH endonucleases of prokaryotic ssDNA viruses seem to have originated from plasmid endonucleases that are not fused to S3H. The CPs of most ssDNA viruses can be traced to a completely different source, namely, the SJR-CPs of different groups of plant and animal (+)RNA viruses. Thus, in a striking example of convergent evolution, ssDNA viruses apparently evolved on multiple, independent occasions via the same route, recombination between a bacterial plasmid and a cDNA copy of a (+)RNA virus.

The ssDNA viruses apparently gave rise to parvoviruses with their linear genomes and the rolling hairpin replication mechanism. Small dsDNA viruses with circular genomes, members of the families *Papillomaviridae* and *Polyomaviridae*, appear to have evolved from ssDNA viruses, most likely, parvoviruses, via a radical transition, inactivation of the HUH domain of the Rep protein that became a DNA-binding domain. Concomitantly, the replication mode of these viruses switched from the rolling circle mechanism to the bidirectional mechanism resembling plasmid and bacterial chromosome replication. The capsid proteins of small dsDNA viruses are highly derived, so it remains unclear whether these evolved from parvovirus CPs or were recruited independently.

Double-stranded DNA viruses

Double-stranded DNA viruses are the most diverse and expansive of Baltimore classes. These viruses dominate the prokaryotic virome and are common in eukaryotes as well, although not to the extent of the RNA viruses. No single gene is shared by all dsDNA viruses, thus, a phylogenetic tree for this Baltimore class cannot be constructed, even in principle. Nevertheless, networks of gene sharing, combined with phylogenies of the genes that are conserved in large virus groups, provide a working platform for classification of dsDNA viruses.

Broadly, the dsDNA virus gene-genome network splits into two large supermodules and three smaller modules. The supermodules are held together by distinct sets of genes that encode structural and morphogenic proteins, and each supermodule appears to be monophyletic, at least, with respect to the core genes.

The DJR-MCP supermodule includes numerous groups of viruses with icosahedral virions that infect bacteria, archaea and eukaryotes. In addition to the signature major capsid proteins, the majority of these viruses also encode SJR minor capsid proteins and genome-packaging ATPases of the FtsK-HerA superfamily. Apart from these 3 morphogenic genes, most of the viruses in this supermodule encode DNA polymerases and often additional components of the replication apparatus, such as S3Hs.

This supermodule includes the tail-less bacterial viruses currently classified in the families *Tectiviridae* and *Corticoviridae* and archaeal viruses in the family *Turriviridae*. Metavirome analyses reveal enormous diversity of unidentified members of this supermodule (one group forming the proposed family “*Autolykiviridae*”). Thus, the DJR-MCP viruses represent a major, perhaps even dominant, but so far poorly characterized component of the prokaryotic virome.

Eukaryotes harbor an even greater diversity of viruses of the DJR-MCP supermodule. Polintons (polintoviruses), virus-like self-synthesizing transposons, as well as diverse polinton-like viruses and virophages, appear to be direct descendants of bacterial tecti-like viruses. Adenoviruses are a derived form that evolved within the same line of descent. In addition to the two capsid proteins and the packaging ATPase, these viruses have acquired another gene for a protein involved in morphogenesis, a protease most likely derived from a deubiquitinating enzyme.

Nucleocytoplasmic large DNA viruses (NCLDV), a vast assemblage of several families of eukaryotic viruses (poxviruses, iridoviruses, phycodnaviruses, mimiviruses and more) with large or “giant” genomes (up to 2.5 Mb, in the case of Pandoraviruses) retain the DJR structural module including the protease, and thus, notwithstanding the striking genomic expansion, appear to have descended from polinton-like viruses. Tecti-like and polinton-like viruses, respectively, also gave rise to two groups of capsid-less elements, mitochondrial and cytoplasmic linear plasmids found in plants, fungi and some protists.

The HK97 supermodule unites the enormous variety of tailed bacterial viruses (bacteriophages), related viruses infecting archaea, and a single, albeit expansive group of animal viruses, the order *Herpesvirales*. The major capsid protein in the icosahedral capsids of these viruses is unrelated to DJR-MCP (the name of this supermodule comes from the name of a phage whose capsid was among the first determined capsid structures of this supermodule). In addition to the HK97-MCPs, viruses in this supermodule share portal proteins and the large and small subunits of the terminases, a distinct packaging ATPases that are not directly related to the packaging ATPases of the DJR-MCP supermodule viruses.

Apart from the structural and morphogenetic proteins, HK97 viruses have extremely diverse genome complexity and gene composition. Some of these viruses encode nearly complete replication machineries, whereas others virtually lack replication genes and rely on host replication apparatus entirely. The taxonomy of prokaryotic HK97 viruses, with the vast order *Caudovirales* classified into five families, *Ackermannviridae*, *Herelleviridae*, *Myoviridae*, *Podoviridae*, and *Siphoviridae*, is currently under reorganization and is likely to expand into many families in the near future.

A distinct, relatively small stand-alone module consists of several families of viruses infecting arthropods, such as baculoviruses and nudiviruses. These viruses have several genes that are distantly related to NCLDV core genes and thus could be peripheral members of the DJR-MCP supermodule.

A large module that is effectively disconnected from the rest of the network consists of viruses of hyperthermophilic Crenarchaeota which adopt diverse, unique virion structures, with capsid or nucleocapsid proteins unrelated to those of other viruses or to each other. Furthermore, most of these archaeal viruses lack genes encoding recognizable homologs for proteins of the replication apparatus encoded by other viruses. However, they are linked by shared genes for transcription regulators, glycosyl-transferases, and uncharacterized ATPases.

Finally, a stand-alone module in the dsDNA virus network includes papillomaviruses and polyomaviruses that evolved from ssDNA viruses, as discussed in the preceding section, and are unrelated to the rest of the dsDNA viruses. Notably, however, papilloma-polyomaviruses are linked to the DJR-MCP supermodule via an unusual group of dsDNA viruses, adomaviruses. These viruses combine a gene for a polyomavirus-like replication protein with genes for structural and morphogenetic proteins homologous to those encoded by adenoviruses. This is another, remarkable case of virus chimerism.

The Evolutionary Status of the Baltimore Classes, Global Organization of the Virus World and Higher Level Virus Taxonomy

Phylogenomic analyzes of all available groups of viruses outlined above place the Baltimore classes into an evolutionary framework (Fig. 4). Only one class, (–)RNA viruses, shows unequivocal evolutionary coherence as a monophyletic taxon. The dsRNA viruses are polyphyletic, with at least two major branches springing up from different parts of the (+)RNA virus tree. Accordingly, (+)RNA viruses, although, technically, monophyletic, represent a paraphyletic taxon with respect to both dsRNA viruses and –RNA viruses.

Among the reverse-transcribing viruses, the RNA-containing ones are paraphyletic with respect to one of the DNA-containing groups (caulimoviruses), whereas the second DNA-containing group (hepadnaviruses) forms a distinct branch. In addition to the splitting and mixing of Baltimore classes, the global organization of the virus world includes a grand unification of all RNA viruses with reverse-transcribing viruses, joining 5 Baltimore classes.

Baltimore Class II, ssDNA viruses, is a distinct case of chimerism and parallel recruitment of key genes from common pools of ancestors, which represents neither monophyly nor polyphyly, in the traditional sense. Furthermore, evolutionary analysis clearly indicates that small dsDNA viruses (polyoma-papillomaviruses) are derivatives of ssDNA viruses, once again demonstrating common origin of two Baltimore classes.

The dsDNA viruses present a clear-cut case of polyphyly of a Baltimore class, with at least 4 independent groups that do not share a common origin: the DJR-MCP and HK97 supermodules, crenarchaeal viruses, and polyoma-papillomaviruses. Thus, the Baltimore classes give a coherent and useful classification of viruses and correlate with the evolutionary relationships that are revealed through phylogenomic analysis. However, this correlation is far from perfect, so that, in general, the Baltimore classes are not monophyletic and, accordingly, cannot be taken as the highest-ranked taxa of viruses.

The Global Network Organization of the Virus World and the Key Events in Virus Evolution

Are there evolutionary connections that reach across the entire virus world? As discussed above, phylogenomic analyzes confidently indicate that some of the major groups of viruses do not share common ancestry, e.g., the 4 large groups of dsDNA viruses. However, this does not rule out the existence of connections. Indeed, strikingly, only 3 simple protein domains present in multiple VHGs, namely, RRM that forms the core of RNA and DNA polymerases, jelly roll, the core domain of SJR-MCP and DJR-MCP, and S3H, link the majority of the viruses in all Baltimore classes into a single network (Fig. 4). Thus, underneath the diverse evolutionary scenarios and the apparent polyphyly of viruses, lies a fundamental structural and evolutionary unity that, in all likelihood, goes back to very early stages of evolution.

- Koonin, E.V., Dolja, V.V., 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiology and Molecular Biology Reviews* 78 (2), 278–303. doi:10.1128/MMBR.00049-13.
- Koonin, E.V., Dolja, V.V., Krupovic, M., 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*. 479–480, 2–25. doi:10.1016/j.virol.2015.02.039.
- Koonin, E.V., Yutin, N., 2019. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Advances in Virus Research* 103, 167–202. doi:10.1016/bs.aivir.2018.09.002.
- Krupovic, M., Koonin, E.V., 2015. Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. *Nature Reviews Microbiology* 13 (2), 105–115. doi:10.1038/nrmicro3389.
- Li, C.X., Shi, M., Tian, J.H., *et al.*, 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* 4. doi:10.7554/eLife.05378.
- Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature Reviews Microbiology* 15 (12), 724–739.
- Shi, M., Lin, X.D., Tian, J.H., *et al.*, 2016. Redefining the invertebrate RNA virosphere. *Nature* 540 (7634), 539–543. doi:10.1038/nature20167.
- Shi, M., Zhang, Y.Z., Holmes, E.C., 2018. Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus Research* 243, 83–90. doi:10.1016/j.virusres.2017.10.016.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio*. 9 (6), pii: e02329-18.

The Virus Species Concept

Peter Simmonds, University of Oxford, Oxford, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

High throughput sequencing (HTS) New technologies capable of large scale nucleotide sequencing sequence assembly and analysis. The method has revolutionized the acquisition of virus sequence information, particularly from environmental samples.

International Committee for the Taxonomy of Viruses (ICTV) An international body with the task of developing

a single, universal taxonomic scheme for all viruses. The ICTV was created as a Committee of the Virology Division of the International Union of Microbiology Societies (IUMS).

Polythetic class A polythetic class is a group of like organism with common properties, although there may be no single property that is shared by all members assigned to that group.

Species in Biology

The primary category of living things is the species. For animals and plants, species correspond to a general division into organisms of single kinds, populations that share appearance and behaviors and importantly, can interbreed to produce offspring similar to their parents. The description of species as “Groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups” is the essence of the widely used biological species concept (BSC) developed by Mayr (see Mayr, 1942 in the Further Reading list). Species are also reproductively isolated from other species and collectively share an exclusive gene pool and a common evolutionary origin. They form a unit of selection that is subject to natural selection that underpins their adaptive fitness.

The application of the BSC is however severely limited by its reliance on inter-breeding ability as its primary inclusion criterion. There are many groups of organisms where this cannot be determined, for example, in the assignment of species to fossil forms and to parthenogenetic animals and plants, to bacteria and in the case of this article, to viruses. In these instances, there have been numerous proposals to define species based on morphology (morphospecies), population cohesion or connectedness, occupation of a specific ecological niche, and more recently by genetic relationships (cladistics or phylogenetic species) as genome sequencing becomes increasingly used for biological characterization. As many as 24–26 different species definitions have been recently reviewed, each with particular applications for particular groups of organisms (see the Mayden, 1997 review in the Further Reading list). The extension of species level classification into microbial and viral taxonomy has required further development of the species concept. What criteria might be used for viruses and the nature of the species thus created has been an immense challenge and area of controversy ever since the species level was first proposed around the time of the publication of the International Committee for the Taxonomy of Viruses (ICTV) 2nd Report in 1976. As described, these are far from being resolved.

Assignment of Viruses to Species

The Master Species List (MSL) maintained by the ICTV is the principal database of virus taxonomy information (see “Relevant Websites section”). Last updated in 2019, it lists a total of 4958 species of viruses, distributed into 846 genera and 143 families. These represent the primary (and obligatory) ranks used for any virus taxonomic assignment. Taxa form a set of hierarchical nested categories, so that one or more species (in the lowest rank classified by the ICTV) are members of a single genus, and one or more genera are assigned to a single virus family.

The Development of the Species Level in Virus Taxonomy

The motivation behind the original classification of viruses was to categorize the diseases that they caused. As infectious agents, viruses are responsible for a substantial proportion of human and veterinary morbidity and mortality, and of often devastating crop failures and other agricultural diseases of plants. The identification of viral causes was an important step in attempts to detect, contain and mitigate effects of infection and vaccinate against them. In the early era of virus discovery and characterization the 1950s and 1960s, the various morphologies, sizes of virus particles and presence of features such as envelopes, tails and spikes visualized by electron microscopy (EM) formed the original basis for family level classification, much of which is still in existence today.

Originally, however there was considerable reluctance to assign viruses in these virus families as species. It was unclear whether a division of viruses after the characteristics of the diseases they caused, as was done in microbiology, would be equivalent to the much more established traditions of species assignment elsewhere in biology. Milne and others furthermore argued that as viruses

were not living entities, and “best classified with other kinds of molecule or molecular complex” (Milne, 1988). They therefore could not form groups that would in any way correspond to animal or plant species. Secondly, without the possibility of applying the BSC to viruses, it was very unclear what level of difference between virus groups would be the equivalent of biologically-defined species elsewhere. Thirdly, it was not obvious whether specific traits used to identify virus groups, such as their disease associations or epidemiological distributions represented the same heritable traits on which species divisions of other organisms were based. A minor mutation within a species might dramatically affect the pathogenicity of a virus, while conversely viruses that caused the same disease, such as pneumonia or meningitis, might indeed belong to many species. Related to this, pre-genomics methods for virus comparisons, such as serological cross-reactivity or DNA hybridization methods were often highly discordant from virus groupings established by comparison of phenotypic properties. As previously noted by members of the ICTV in the late 1970s, many of the different flaviviruses assigned as different species because of their different disease outcomes showed extensive serological cross-reactivity. At the same time, different “types” of influenza A, assigned to a single species based on its disease associations, showed no serological cross-reactivity with each other.

Notwithstanding these identified problems, the first systematic attempts to classify viruses down to the level of species based on their observed properties started in the mid-1970s. These followed the pattern of yellow fever virus, assigned to the species “Yellow fever Flavivirus” (later reformulated as *Yellow fever virus*) in the genus *Flavivirus*, family *Togaviridae* (later re-assigned as *Flaviviridae*). Most early assignments of viruses matched disease, virus and species names together. In many ways, this descriptive approach to classify viruses corresponded to what biologists might do with animals, plants, fungi and bacteria – divide them up into entities that could be distinguished from each other, be it their symptoms, appearance of lesions, disease outcomes and fatality rates in humans or animals. Plant viruses may similarly induce a wide range of characteristic leaf, stem or root lesions, stunting and necrosis in particular host species that similarly provided a battery of descriptive names for the large number of viral crop diseases characterized at that time. Different viruses may additionally be characterized through their seasonality, geographical distributions and transmission by particular arthropod vectors.

Species Assignments and Genetic Relationships

Much of this early spadework in classifying viruses was done from the 1950s to the 1970s in an essentially pre-genomic age. Viruses could be isolated due to advances in sterile cell culture, they could be visualized by EM but could not be genetically characterized until the advent of large scale Sanger sequencing in the late 1970s. Through analyzing the sequences of cloned viral DNA sequences and, after that, of amplicons generated by polymerase chain reaction (PCR)-mediated genome amplification methods in the late 1980s, it became readily apparent that viruses classified in species, genera and families by phenotypic observations corresponded to genetically different lineages, and between different virus families, by major differences in genome organization and gene contents.

Sequence comparisons between viruses provides a much more fine-grained and objective metric of virus relatedness and, as elsewhere in biology, can contribute substantial additional classification information at multiple taxonomic levels. At one extreme it can identify and differentiate between individual virus strains or genotypes hitherto assigned to same species; at the other, it can provide evidence of distant homologies between replication-associated genes such as the RNA-dependent RNA polymerase. Their genetic relationships provide evidence for deeper evolutionary relationship between viruses that had previously been classified into separate families and orders. However, the advent of large scale genetic approaches for virus taxonomy opened up a Pandora’s box of problems for those involved in reconciling the newly derived genetic relationships between viruses with what was originally an largely phenotypically-grounded taxonomy. The relationship between sequence divergence and species assignment is particularly problematic and may never be reconciled.

Criteria Used in Species Definitions

The first formalised definition of a virus species was developed by Marc van Regenmortel and colleagues from the ICTV (Van Regenmortel *et al.*, 1991):

“A virus species is a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche”

This does not define what actual properties of a virus might be used to assign and differentiate between species, but at the time the definition was applied, these were typically based on descriptive features such as host range, antigenicity, epidemiology and distribution (“ecological niche”) rather than genetic relationships. The term polythetic has been used previously in descriptive biological species definitions and indicates that species may share a range of properties, but none of them are individually essential for species assignment. The further requirement that viruses assigned to a species are a single genetic lineage and an implied common ancestor distinct from all other viruses introduces a genomic-based element that was gradually adopted as sequence information became available in that period.

The species definition for viruses was revised by the ICTV in 2012 as follows:

“A virus species is a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria”

This revision remained as unclear as the original definition was about what “multiple” criteria might be used for species delineation. However, the removal of the term “polythetic”, widely believed by the ICTV to be misunderstood, also seemingly removed the need to include phenotypic information in a species description thus enabling viruses to be assigned largely or entirely by metrics of genetic relatedness.

The lack of defined criteria to assign species in both of the ICTV species definitions has undoubtedly contributed to the wide range of quite different descriptive elements and metrics of genetic relatedness used in species definitions in the ICTV taxonomy. This problem is exacerbated by a lack of coordinated oversight into how different viruses are classified and incorporated into the ICTV taxonomy. Most of the work in virus taxonomy is performed by a large number of separate Study Groups of virologists with expertise in particular families. While this division of labor has allowed those with the greatest knowledge of individual viruses and groups to be engaged in virus classification, the relative isolation of Study Groups and independence from each other has also led to the adoption of often quite different criteria for assigning species and genera to viruses they are classifying. A brief scan of the ICTV 9th Report reveals that species definitions for different families and genera vary between purely descriptive elements of virus disease, antigenic properties, haemagglutination or other properties of their isolates through to definitions that may be purely based upon sequence relationships. For the latter, the actual divergence thresholds and genome regions compared used for species divisions vary greatly between virus families. As examples, a <40% amino acid sequence identity of the RdRp divides partitiviruses into separate species while it is 50% in pseudoviruses (capsid gene), 60% in enterovirus species (structural genes) and 90% and 93% in bunyaviruses and hantaviruses (capsid gene). Nucleotide sequence identities might also be used, 30% in endornaviruses (complete genome [CG]), 40% in *Densovirinae* (non-structural [NS] gene), 65% in torquetenoviruses (ORF1), 70% in papillomaviruses (CG), 75% in nanoviruses (CG), 80% in caulimoviruses (*pol* gene), 89% in curtoviruses (CG) and 95% in parvoviruses (NS gene). Very clearly, what constitutes separate species in one virus family might be quite different for species in others.

Species definitions that are based on their clinical or agricultural disease associations furthermore frequently conflict with their underlying genetic relationships. There are currently 54 species assigned to the genus *Flavivirus* (family *Flaviviridae*), that like *Yellow fever virus* described above, bear names that simply derive from their component descriptively named viruses (e.g., dengue fever virus, tick borne encephalitis virus, West Nile virus etc.). Despite the fact that all species of flaviviruses were sequenced many years ago, their genetic relationships remain clearly subservient to their species assignments. There is consequently only a weak relationship between a flavivirus species assignment and its genetic divergence from other species. Two different species may differ by only 1.5% in nucleotide sequence, such as between *Israel turkey meningoencephalomyelitis virus* and *Bagaza virus*, but exceed 40% between others (*San Perlita virus* and *Ilheus virus*). Different types of dengue fever virus assigned to the same species *Dengue fever virus* show greater than 32% nucleotide sequence divergence from each other but these distances that are substantially greater than those between most other flavivirus species. Other anomalies include the assignment of *Louping ill virus* as a separate species from *Tick-borne encephalitis virus* on the basis of its distinct geographical distribution in the UK and different host range, even though phylogenetically, variants assigned to this species lie entirely within the much more diverse viruses assigned to *Tick-borne encephalitis virus*.

In the case of flaviviruses, there seems to be little enthusiasm in the *Flaviviridae* Study Group to reassign species to better match underlying genetic sequence relationships. However, conflicts of this sort elsewhere in the virus taxonomy have led to major reorganizations underlying genetic relationships are revealed. For example, there have been repeated instances of clinically similar viruses having to be re-assigned to different species or even genera or families once their sequences were determined; echovirus 22 and 23 were moved from the genus *Enterovirus* into a new genus, *Parechovirus* in the family *Picornaviridae*. Even more radically, whole families have been split and member species and genera reassigned. As a typical example, the original family *Togaviridae* was divided into *Togaviridae* and *Flaviviridae* when major differences in genome organization and a lack of sequence homology between genes became apparent. The species *Rubella virus* in the genus *Rubivirus* was then subsequently also removed from the *Togaviridae* and assigned to a new family *Matonaviridae* once its lack of genetic relatedness to members of this originally assigned family was acknowledged. Several similar re-organisations of the existing taxonomy are likely in the near future.

Species and Genotypes

While there remains a historical legacy of descriptive species assignments in some families despite their often wildly different genetic relationships, the classification of newly characterized viruses into species is increasingly based on metrics of genetic relatedness. However, this may create the converse problem of poorly matching assignments with medically or biologically relevant properties that virologists and clinicians might want to use for classifying them. For example, all variants of hepatitis E virus (HEV) have, perhaps unhelpfully, been placed into a single species, *Hepevirus A*, along with a range of other HEV-like viruses infecting other mammalian species. (The other species, *Hepevirus B*, *C* and *D*, incorporate even more divergent virus groups with similarly variable host associations.) Because members of *Hepevirus A* were already highly divergent in sequence from each other (30% amino acid sequence divergence in conserved domains of methyltransferase, helicase and polymerase genes), much of the medically relevant classification activity revolves around the assignment of genotypes and subtypes within the species. For example, HEV types 1, 2, 3 and 4 infecting humans possess quite distinct geographical ranges, routes of transmission and risk group associations, properties that might at one time have led to their assignment as different species.

Members of the *Hepacivirus* genus in *Flaviviridae* are a similar case in point; the human hepatitis C virus (HCV) is assigned to the one species *Hepacivirus C* (14) even though the clinically and epidemiologically relevant differentiation is at the level of genotype and subtype within this genus. The sequence divergence between HCV genotypes (>30%) is indeed greater than those between different species in its sister *Pestivirus* genus, such as *Pestivirus A* and *Pestivirus B*. As a further example in another virus family

(*Picornaviridae*), a small proportion of poliovirus infections are associated with a severe neurological outcome, poliomyelitis, that quite distinct from the disease associations of all other enteroviruses. Originally assigned to the species *Poliovirus*, the more recently developed 40% amino acid sequence divergence threshold in the structural gene block now used to delineate enterovirus species places them as just three serotypes among a much larger group of other enteroviruses assigned to the species, *Enterovirus C* that lack these disease associations. As a contrast to this “lumping” effect, species assignments based on lower divergent thresholds may atomize existing virus groups for little apparent purpose. In a much discussed example highlighted by Marc van Regenmortel (2016) (see Further Reading list), the application of an 9% divergence threshold to begomoviruses created over 300 genomically defined species, including the split of the species *Tomato leaf curl virus* into 42 species without any clear or consistent biological relevance.

Overall, these examples and many others demonstrate that species of viruses may not be the lowest element in classification and there may be a considerable degree of variability between virus families and even genera about what constitutes a species and what constitutes a genotype, sub-type or even strain.

Classification of Metagenomically Characterized Viruses

This development of purely genomics-based species assignments has become particularly relevant for the classification of viruses identified and assembled from datasets generated by high throughput sequencing (HTS). The application of such methods to environmental samples and rarely investigated hosts such as insects and other arthropods has revealed a staggeringly diverse sets of new and unfamiliar genetically diverse viruses. These show a degree of genetic and genome organizational diversity that warrants the ongoing and future creation of potentially hundreds of new virus families to accommodate them.

Although some properties of the metagenomically-derived viruses might be inferred from their genome organization and gene complements, as well as their sampling locations and in many cases, inferences about their likely hosts, their classification have to be essentially based on genetic relationships. Many virologists have forcefully and cogently argued against this policy (van Regenmortel *et al.*, 2019). However, the ICTV, supported by the conclusions of an expert group convened to specifically to discuss this issue (Simmonds *et al.*, 2017), has proceeded to assign such “sequence-only” viruses to the official taxonomy equivalent to those already assigned. This policy requires decisions to be made on family-, genus- and species-level assignments in the frequent absence of any phenotypic or other descriptive support. The incorporation of sequence-only viruses had greatly expanded the ICTV taxonomy and will undoubtedly continue to do so as virus characterization in metagenomic datasets proceed.

Emergence of a Genomics-Based Taxonomy of Viruses

The previous section outlines the transitions from an original phenotypically-driven assignment of viruses to species to a new framework that is primarily based upon virus genetic relationships. The current taxonomy is an uncomfortable mix of species defined by these different methods, with large discrepancies in both the nature of the species groups formed, but also in the motivation behind and utility of the species entities that are created (Table 1). The utilitarian assignment of viruses based on phenotype properties creates categories that are immediately of clinical or other biological value (left column), produce informative names and yet often fail to reproduce the underlying genetic relationships of the viruses so classified, nor can they be readily applied for species assignments of sequence-only viruses from metagenomic datasets. Contrastingly, genomics based species assignments (right column) are consistent with virus evolutionary relationships on which a taxonomy ultimately has to be based, and yet for many virologists it fails in its primary purpose in providing relevant categories for applied research. Species may lack descriptive elements that link viruses to their observable properties, their definitions may be based on arbitrary and often misplaced assignment thresholds that may atomize existing species or create varied and inconsistent distinctions between species and genotypes. Thresholds or other metrics of genetic relatedness furthermore produce no obvious basis for species names which are frequently numerically or alphabetically coded.

Table 1 Features of phenotypic and genomics-based species assignments

Phenotypic assignment	Genomics-based assignment
Provides a biologically and clinically relevant division of viruses	Assignments approximate to evolutionary relationships between and within species
Species names are typically descriptive and informative	Species names are often serially numbered without any indication of virus properties
Variable association with genetic divergence between and within species members	Species may contain viruses with different biological or clinical attributes
Assignments may conflict with evolutionary relationships	Sequence divergence threshold used species differentiation are variable between virus groups and often arbitrary
Cannot be applied to viruses without information of biological properties (eg., those derived from metagenomic datasets)	Can be readily applied to viruses with both known and unknown (eg., from metagenomic data) phenotypes

As the taxonomy expansion develops, an increasing proportion of species will be genomically defined, with an increasingly small “rump” of the originally descriptively classified species remaining. Reconciling these legacy assignments with the expanding genomics framework awaits future organizational initiatives by the ICTV.

Species Nomenclature

The principles behind naming virus species has undergone several revisions and the current nomenclature is far from being consistent or logical as a result; it may also change substantially in the future. As described above, naming is influenced by the underlying principles used to delineate species, be it based on descriptive properties or genetic relationships. Historically, the initial reluctance of virologists to assign species extended to and fueled a parallel resistance to adopt the Linnaean binomial nomenclature that is used elsewhere in biology for virus species. After the initial classification of viruses into families and genera was better established in the mid-1970s (5), many of the component viruses within these, ultimately to be defined as species, were informally named using a non-Latinized binomial of the form <virus name> + genus, *e.g.*, West Nile flavivirus (the reverse of the Linnaean genus + species principle). In 1998, the ICTV implemented a formal nomenclature for assigned species as a non-Latinized (usually English) monomial (*i.e.*, species only) name that was based typically on the name of the virus within the species. In contrast to Latinized binomials, there is no specific format for virus species names. Those derived from descriptive species definitions were often formed from two to four word epithets matching the virus name (*e.g.*, *Dengue virus*, *Banana bunchy top virus*). Species names may, however, additionally include numeric or alphabetic identifiers or combinations (*e.g.*, *Enterovirus A*, *Homalodisca coagulata virus-1*, *Torque teno mini virus 12*), that often arise from species defined genomically where a descriptive name may not be obviously relevant.

Although typographically marked by italicization and capitalization, the frequent identity of the names of the virus and the species to which it is assigned (*e.g.*, yellow fever virus is a member of the species *Yellow fever virus*) creates confusion and has engendered frequent misuse of virus and species terms in the literature. The ICTV is currently (in 2019) conducting a community-wide consultation on the advisability of applying a Linnaean nomenclature scheme to virus species; it would more effectively mark species names in text, their Latinized epithets would better differentiate species names from virus names, names would be more informative through inclusion of the genus assignment and finally their use would engender compatibility with the nomenclature used in the rest of biology, including microbiology.

What are Species?

The ICTV maintains a strict division between the identity of a virus and the identity of the taxon to which it is assigned. Under current guidance, a species (or a genus or other taxonomic rank) is strictly defined as a class or category into which viruses, as real world entities are assigned. Thus, poliovirus type 3 is member of the species *Enterovirus C*. Polioviruses possess specific replication and morphological properties and have effects on the cells that they infect and specific disease associations. Contrastingly, *Enterovirus C* as a class or a category does not replicate, has no morphology, no effect on cells and no pathogenicity. This relationship formalizes the requirement for separate names for viruses and taxa, the former representing the “common” name for a virus, which may indeed be different in different languages (rabies virus and virus de la Rage are both members of the species *Rabies lyssavirus*) and the latter representing the name assigned by the ICTV to the taxon.

However, this strict differentiation between organism and the taxon to which it is assigned is widely misunderstood and the names of species and other taxonomic ranks are often conflated in the literature with the viruses they represent. In part, this may originate from a much less clear and often implicit rather than defined differentiation of organisms and taxa in the terminology used in microbiology and elsewhere in biology. For example, an individual may be infected with *Escherichia coli* or by *Streptococcus pyogenes* (implicitly by the bacteria assigned to these species), but he or she cannot be infected by *Enterovirus C*, only by poliovirus, coxsackie virus A21 or one of the many other enteroviruses assigned to that species. It seems unlikely that the ICTV will conform to usage elsewhere in biology in this respect so this potential source of confusion will potentially continue even if a Latinized binomial nomenclature is ultimately adopted.

Summary

While biologists know pragmatically what species are, and have little difficulty in assigning animals, plants, fungi and other organisms to them, formal definitions of species differ considerably in their applicability to different groups of organisms. The plethora of currently used species concepts and definitions underpins much of the ongoing difficulty and uncertainties in how to classify viruses at this taxonomic level. Variants of what might correspond to morphological or phenetic species definitions that classified viruses based on their observed clinical or epidemiological properties have largely given way to a version of the evolutionary species concept, based on virus genetic relationships as the features available for species characterization have changed. The current virus taxonomy is an unresolved composite of different classification methods and further complicated through quite inconsistent metrics of genetic relatedness used to assign species in different virus families and genera. The currently

used formal definition of a virus species contains elements of both phenotypic and genetic characteristics that does not clearly delineate which are necessary or sufficient for species assignments. Species are named by italicized, capitalized monomial names often corresponding the virus name in English, quite different from the Latinized binomial nomenclature used elsewhere in biology. The consequently frequent close resemblance of virus and species names to each other despite a rigidly enforced typological distinction between a species and a member of a species has led to the frequent misuse species terms for viruses. In the future, this may be resolved, at in part, through the adoption of Latinized binomial names that are consistent with their use elsewhere in biology.

References

- Milne, R.G., 1988. Species concept should not be universally applied to virus taxonomy – But what to do instead? *Intervirology* 29, 254–259.
- Simmonds, P., Adams, M.J., Benko, M., *et al.*, 2017. Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology* 15, 161–168.
- Van Regenmortel, M.H.V., 2019. Solving the species problem in viral taxonomy: Recommendations on non-Latinized binomial species names and on abandoning attempts to assign metagenomic viral sequences to species taxa. *Archives of Virology* 164, 2223–2229.
- Van Regenmortel, M.H., Maniloff, J., Calisher, C., 1991. The concept of virus species. *Archives of Virology* 120, 313–314.

Further Reading

- Gibbs, A.J., Gibbs, M.J., 2006. A broader definition of 'the virus species'. *Archives of Virology* 151, 1419–1422.
- Kingsbury, D.W., 1985. Species classification problems in virus taxonomy. *Intervirology* 24, 62–70.
- Mayden, R.L., 1997. In: Claridge, M.F., Dawah, H.A., Wildon, M.R. (Eds.), *Species: The Units of Biodiversity*. London: Chapman and Hall, pp. 381–424.
- Mayr, E., 1942. *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. New York: Columbia University Press.
- Simmonds, P., 2015. Methods for virus classification and the challenge of incorporating metagenomic sequence data. *Journal of General Virology* 96, 1193–1206.
- Simmonds, P., 2018. A clash of ideas – The varying uses of the 'species' term in virology and their utility for classifying viruses in metagenomic datasets. *Journal of General Virology*. doi:10.1099/jgv.0.001010.
- Simmonds, P., Aiewsakun, P., 2018. Virus classification – Where do you draw the line? *Archives of Virology* 163, 2037–2046.
- Van Regenmortel, M.H.V., 1990. Virus species, a much overlooked but essential concept in virus classification. *Intervirology* 31, 241–254.
- Van Regenmortel, M.H.V., 2003. Viruses are real, virus species are man-made taxonomic constructions. *Archives of Virology* 148, 2481–2488.
- Van Regenmortel, M.H.V., 2016. Solving the species problem in viral taxonomy: Recommendations on non-Latinized binomial species names and on abandoning attempts to assign metagenomic viral sequences to species taxa. *Current Topics in Microbiology and Immunology* 13, 59–68.

Relevant Websites

<https://talk.ictvonline.org/taxonomy/>
Taxonomy.
International Committee on Taxonomy of Viruses (ICTV).

Genetic Diversity and Evolution of Viral Populations

Rafael Sanjuán and Pilar Domingo-Calap, Institute for Integrative Systems Biology (I2SysBio), University of Valencia – CSIC, Valencia, Spain

© 2021 Elsevier Ltd. All rights reserved.

This is an update of R. Sanjuán, Quasispecies, in *Encyclopedia of Virology* (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00481-7.

Glossary

Basic reproductive number Average number of new infections generated by an infected host.

Clonal interference Failure of beneficial alleles to get fixed in a population due to the presence of other, more beneficial alleles.

Defective interfering virus Mutant lacking a portion of the viral genome that replicates at the expense of and interferes with functional viruses.

Error catastrophe Process whereby, at high mutation rates, lower-fitness mutants become more abundant than higher-fitness variants.

Fitness Production of viable progeny, usually averaged over individuals of a given genotype and expressed relative to a reference genotype.

Fitness landscape A representation of the genotype-fitness map.

Genetic complementation Compensation of a genetic defect by the presence of other, functional copies of the mutated gene in the same cell.

Genetic hitch-hiking Increase in the population frequency of a deleterious mutation driven by genetic linkage with a beneficial mutation.

Lethal mutagenesis Demographical extinction of a population due to mutation accumulation.

Muller's ratchet Irreversible accumulation of deleterious mutations in a population as a result of random genetic drift.

Mutation rate Probability that a new genetic variant is produced per generation, where a viral generation is often defined as the completion of a cellular infection cycle.

Mutation-selection balance Equilibrium between the influx of new deleterious mutations in a population and the removal of these mutations through selection.

Selection coefficient Difference between the fitness value of a genetic variant and the fitness value of a reference genotype.

Viral quasispecies A population of viruses replicating and competing at high mutation rates.

Introduction

Viruses exhibit highly heterogeneous genome structures and replication strategies. These basic features correlate with viral propensity to mutate and, hence, with viral genetic diversity and evolvability. The mutation rates of viruses vary according to genome composition (RNA or DNA), size, and structure (single- or double-stranded). In general, RNA viruses produce from 10^{-6} to 10^{-4} new base substitutions per nucleotide per cell, whereas these rates range from 10^{-8} to 10^{-6} in DNA viruses (Fig. 1). Furthermore, single-stranded viruses appear to mutate faster than double-stranded viruses, and an inverse correlation between genome size and mutation rate has also been demonstrated. Within the DNA virus group, thus, single-stranded viruses and, presumably, some double-stranded viruses with small genomes, mutate faster than double-stranded viruses with large genomes.

The term quasispecies has been used to refer to highly heterogeneous viral populations in which the frequency of mutations in the population is determined by mutation rate and fitness. The quasispecies concept has thus been applied mainly to RNA viruses, owing to their particularly high mutation rates. High population genetic diversity is believed to play a pivotal role in the ability of RNA viruses to adapt to changing selective pressures and invade new hosts. Indeed, RNA viruses comprise a large number of well-known human emerging or re-emerging pathogens such as HIV-1, hepatitis C virus, influenza virus, Ebola virus, and Zika virus among many others, as well as major plant and animal pathogens. In contrast, DNA viruses are typically less diverse and evolve more slowly, although they also include important pathogens such as herpesviruses and poxviruses. Nevertheless, the divide between RNA and DNA viruses is imperfect, since some RNA viruses evolve slowly and some DNA viruses are fast-evolving, particularly single-stranded and small double-stranded DNA viruses.

This article focuses on the sources of genetic diversity in viruses and on the resulting evolutionary processes at the population level (we do not review long-term viral evolution at the phylogenetic level). The main mechanisms responsible for introducing new mutations are discussed, including replication errors, lack of proofreading, repair avoidance, diversity-generating elements, host-encoded editing of viral genomes, and recombination (Table 1). We review how population diversity is modified by natural selection and genetic drift, and how this leads to different evolutionary processes in viruses. Finally, we discuss some practical implications of viral diversity for pathogenesis, immune escape, and drug resistance.

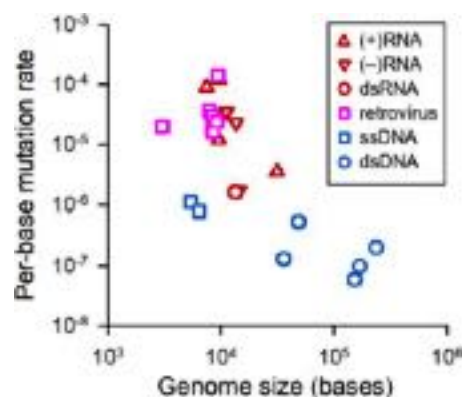


Fig. 1 Mutation rates in different types of viruses. Estimates obtained for viruses belonging to the major indicated groups are shown as a function of genome size. This reveals an effect of genetic material (RNA/DNA) and genome size on viral mutation rates. Data taken from Sanjuán, R., Domingo-Calap, P., 2016. Mechanisms of viral mutation. *Cellular and Molecular Life Sciences* 73, 4433–4448. doi:10.1007/s00018-016-2299-6.

Table 1 Sources of genetic diversity in seven major virus classes^{a,b}

	<i>dsDNA</i>	<i>ssDNA</i>	<i>Para retro</i>	<i>Retro</i>	<i>dsRNA</i>	<i>(+)RNA</i>	<i>(-)RNA</i>
Lack of 3'exonuclease proofreading	–	+/-	+	+	+	+	+
Avoidance of post-replicative repair	–	+/-	+	+	+	+	+
Use of error-prone repair polymerases	+/-	+/-	–	–	–	–	–
Diversity-generating retro-elements	+/-	–	–	–	–	–	–
APOBEC hyper-mutation	+/-	+/-	+	+	–	–	–
ADAR hyper-mutation	–	–	–	–	+/-	+/-	+/-
Reassortment	+/-	+/-	–	–	+/-	+/-	+/-
Template switching	–	–	+	+	+	+	–
Recombinases	+/-	–	–	–	–	–	–

^aAccording to the Baltimore classification of viruses.

^bSource of diversity generally present (+), present in some cases (+/-), or not shown or infrequent (–).

Error-Prone Replication

Intrinsic Selectivity of Viral Polymerases

The fidelity of a polymerase is first determined by its selectivity towards the correct base. Such selectivity is typically on the order of 10^4 – 10^5 , and there appear to be no major differences among RNA-dependent RNA polymerases, reverse transcriptases (RTs), and DNA polymerases. Polymerase variants with altered selectivity and fidelity have been reported in a wide variety of RNA viruses. Serial transfers in the presence of the base analog ribavirin were first shown to select for an increased-fidelity polymerase in polioviruses. Ever since, tens of fidelity mutants have been identified amongst widely different RNA virus families. The evolution of high- and low-fidelity variants in response to selective pressures such as chemically-induced mutagenesis demonstrates that selection can optimize RNA virus replication fidelity. However, most RNA virus fidelity variants show only modest differences in base selectivity compared to the wild type, revealing strong structural or biochemical constraints. Interestingly, the fidelity of viral DNA polymerases appears to be less constrained.

Other Factors Determining Polymerase Selectivity

In addition to the intrinsic fidelity of a polymerase, the rate of base misincorporations depends on factors such as sequence context, divalent cation concentrations, and the relative abundance of dNTPs. For instance, polymerase slippage occurs frequently in homopolymeric templates, producing insertions and deletions. In hepatitis C virus, mutation rates can vary extensively between consecutive bases in the viral genome, suggesting a major role for sequence context in determining replication fidelity. RNA secondary structure can also determine the rate at which replication errors occur.

Lack of Proofreading

Polymerases displaying 3' exonuclease activity can correct most nucleotide misincorporations. However, this activity is generally absent from RNA virus polymerases. Currently, only coronavirus RNA polymerases have been shown to perform

3' exonuclease proofreading. Lack of proofreading is a major factor responsible for the higher rates of spontaneous mutation of RNA viruses compared to DNA viruses. Because RTs also lack 3' exonuclease activity, retroviruses and pararetroviruses mutate at rates similar to those of most RNA viruses. Although DNA virus replicases exhibit 3' exonuclease activity, certain amino acid replacements in their polymerases inactivate proofreading and produce a strong mutator phenotype, as shown for instance in bacteriophage T4. Interestingly, there are also amino acid replacements in the T4 polymerase that increase fidelity up to a hundred-fold compared to the wild type. Finally, polyomaviruses, which use host polymerases for replication, can encode proteins that inactivate the 3' exonuclease proofreading domain of the host polymerase. Therefore, avoidance of proofreading may also be a mechanism whereby some DNA viruses increase their population diversity.

Role of Post-Replicative Repair

Repair Avoidance

Post-replicative repair is a highly efficient system for removing replication errors and repairing DNA damage. As such, lack of repair can increase mutation rates strongly. RNA viruses are not substrates of cellular repair systems, whereas some DNA viruses have evolved specific features to avoid repair. This is the case of bacteriophage Φ X174 and potentially other small DNA phages. In *Escherichia coli*, strand-specific mismatch repair uses the methylation status of GATC sequence motifs to discern between the parental and daughter DNA strands. Interestingly, the Φ X174 genome is devoid of GATC motifs. This prevents methylation and thus, post-replicative repair, allowing the phage to increase its mutation rate relative to that of the host. Eukaryotic viruses might also avoid repair or dysregulate repair pathways to increase their mutation rates, although this has not been investigated in detail.

Error-Prone Repair Polymerases

Eukaryotic viruses are known to interact with DNA damage response pathways. For instance, some small DNA viruses activate these pathways to prolong the S cell-cycle phase and promote their replication. DNA damage response pathways comprise error-prone DNA polymerases for re-synthesis of excised strands, which might contribute to elevating viral mutation rates. On the other hand, some large DNA viruses encode their own DNA repair system. For instance, African swine fever virus encodes an endonuclease, a repair polymerase (pol X) and an ATP-dependent DNA ligase. Pol X is highly error-prone due to a poor base selectivity and lack of 3' exonuclease activity. Use of pol X might contribute to the relatively high genetic diversity shown by African swine fever virus isolates.

Diversity-Generating Retroelements

First described in a *Bordetella* phage, diversity-generating retroelements are genetic cassettes found in different bacteria, plasmids, and DNA bacteriophages. They contain two short repeated sequences, called the template and the variable repeats. The template repeat is transcribed and then reverse-transcribed by an RT encoded by the cassette. During this process, adenines are systematically substituted for random bases. The resulting cDNA is integrated, replacing the former variable repeat with a new, highly mutagenized sequence. This process allows hyper-mutation of specific phage genes involved in host attachment, a trait that is often subject to rapidly changing selective pressures.

Viral Hyper-Mutation Mediated by Host Enzymes

Cellular Cytidine Deaminases

Apolipoprotein B mRNA editing catalytic polypeptide-like enzymes (APOBECs) function as an innate immune response against reverse-transcribing viruses. APOBECs are cytidine deaminases that massively introduce C-to-U base substitutions in the retroviral cDNA, leading to a characteristic G-to-A hyper-mutation pattern in retroviral genomes. It has been estimated that 98% of new HIV-1 mutations are produced by APOBECs, versus only 2% caused by HIV-1 RT errors. Hyper-mutation produces non-infectious genomes, which accumulate as inactive proviruses. In cases of moderate mutation, though, APOBECs might contribute to viral evolution and could even promote immune escape and drug resistance. In addition to retroviruses, APOBECs can edit hepatitis B virus as well as other non-reverse transcribing DNA viruses such as papillomaviruses, herpesviruses, and human polyomaviruses.

Cellular Adenosine Deaminases

Double-stranded RNA-dependent adenosine deaminases (ADARs) constitute another type of host-encoded enzymes capable of editing and hyper-mutating RNA virus genomes. ADARs deaminate adenosines in long double-stranded RNA regions, converting them into inosines, which leads to A-to-G base substitutions. ADAR-induced hyper-mutation was first discovered in measles virus and was then extended to other viruses including human parainfluenza virus, respiratory syncytial virus, lymphocytic choriomeningitis virus, Rift Valley fever virus, influenza virus, noroviruses, and Zika virus, although in many cases evidence for ADAR-mediated mutagenesis is still indirect.

Recombination

Recombination plays an important role in viral evolution. Similar to mutation rates, recombination rates vary extensively among viruses. High rates of recombination may have been favored by natural selection to promote the fixation of new genotypes, to purge deleterious mutations, or as a by-product of other evolutionary processes. For recombination to introduce new genetic variants, it is required that the involved genomes differ by at least two mutations. Furthermore, the recombining molecules have to be present in the same cell. Hence, the emergence of new variants through recombination typically occurs in cells infected with different pre-existing variants. However, this may not happen frequently because super-infection exclusion renders infected cells refractory to re-infection with the same or a related virus. On the other hand, even if rare, recombination between distantly related viruses has the ability to introduce major genetic changes in a single event, such as gene transfer, leading to the emergence of new viral subtypes or even new species. Hence, recombination differs from mutation in the type and abundance of the genetic changes produced.

Reassortment

In segmented and multicomponent viruses, recombination is greatly facilitated by the possibility of re-assorting genome segments. However, barriers to reassortment also exist. For instance, co-packaging of a given genome segment with the segments of another virus requires that their packaging signals are compatible. Reassortment has been extensively studied in influenza virus, for which inter-subtype reassortments between swine, avian and/or human influenza A viruses are known to be at the origin of several pandemics. Reassortment between viruses of the same subtype can also be an important source of diversity, can promote adaptation, and can accelerate immune escape. Reassortment also plays a major role in the evolution of other viruses such as rotaviruses, favoring the emergence of new variants and complicating vaccination efforts.

Template Switching

Even if reassortment is not possible, recombination can still take place through a template-switching mechanism whereby the viral polymerase and the nascent chain dissociate from one template and associate with another. The rate at which this template-switching process occurs has been measured experimentally using genetic markers, and typically ranges from 10^{-5} to 10^{-3} events per nucleotide site per co-infected cell in different retroviruses such as HIV-1, murine leukemia virus, and spleen necrosis virus, as well as in positive-stranded RNA viruses such as mouse hepatitis virus, poliovirus, and tobacco etch virus. Whereas the production of new variants through template switching also requires that the cell is coinfecting with two different variants, in retroviruses such as HIV coinfection with two independent virions is not required because virions are diploid. In contrast to retroviruses and positive-stranded RNA viruses, most negative-stranded RNA viruses show very low rates of template switching. Similar to replication fidelity variants, viral polymerases with different propensities to template switching have been isolated. Such variants have allowed investigating the evolutionary implications of recombination from an experimental evolution approach using poliovirus. This has suggested that recombination increases the ability to purge deleterious mutations from the population and promotes the spread of beneficial mutations. This may accelerate adaptation and increase viral pathogenicity.

Recombination in DNA Viruses

Recombination is a more complex process in DNA viruses than in RNA viruses and involves specific host or viral recombinases. DNA virus homology-dependent recombination can be mediated by cellular double-stranded break repair systems. These systems act on double-stranded DNA viruses but also on single-stranded DNA viruses because the latter produce double-stranded replicative intermediates. Some DNA viruses encode their own recombination systems. For instance, phage λ recombination is active in cells deficient for the *E. coli* recombinase RecA. Also, the herpesviruses recombination machinery plays a major role during viral replication and repair. There is ample evidence that this type of recombination is sequence-dependent. For instance, the inverted terminal repeats of the vaccinia virus genome undergo rapid changes in size because this region contains multiple 10–100 bp repeats that are highly prone to unequal crossovers. Some viruses can use recombinogenic sequence motifs for targeting diversity towards certain genes. In poxviruses, this allows for a recombina-

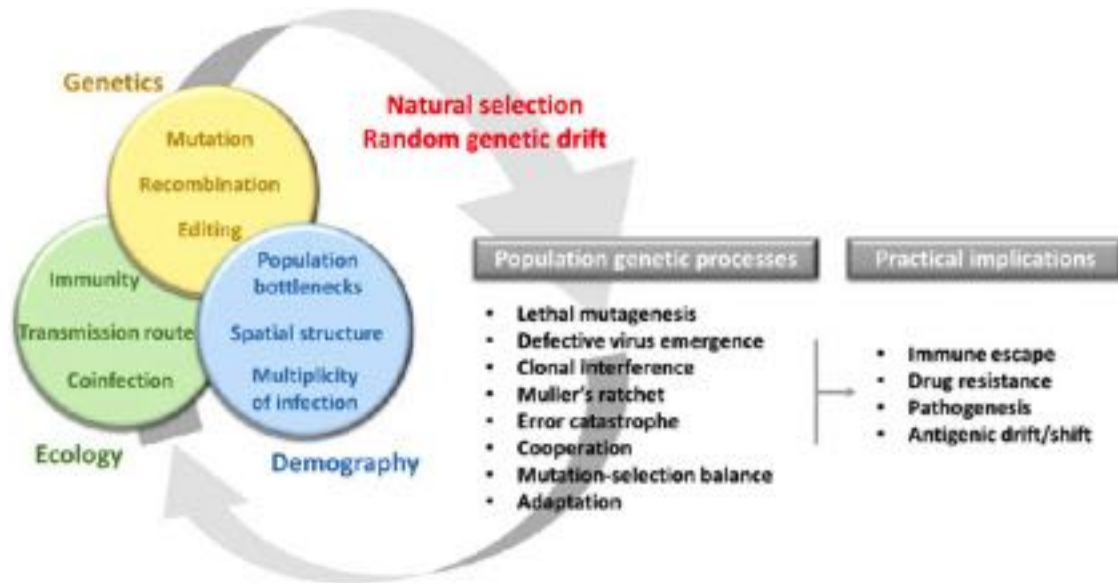


Fig. 2 Genetic, demographical, and ecological processes determining viral population diversity. Mutation is the ultimate source of genetic variation, but recombination also produces new genotypes. Ecological and demographical processes determine how natural selection and random genetic drift acts on these new genetic variants. This leads to a variety of possible population genetic processes, which in turn feedback on viral demography, ecology and genetics. Immune escape, drug resistance, and pathogenesis are a visible consequence of such processes. The list of factors shown is not exhaustive.

tion-mediated gene amplification process that can temporarily increase the dosage of viral genes involved in evading host innate immunity.

Basic Population Genetic Processes Shaping Viral Diversity

The fate of new genetic variants is largely determined by selection. The population frequency of a new mutant will tend to increase, fluctuate randomly, or decrease depending on whether it has a beneficial, neutral, or deleterious effect on viral fitness, respectively. Yet, the fitness of a given allele is context-dependent in several ways, since it varies according to genotype-genotype and genotype-environment interactions. To a large extent, the environment of a virus is dictated by the host, but it can also be modulated by other members of the viral population via competition for cellular resources or cooperative interactions. Allele frequencies can also change as a result of random genetic drift. Drift is determined by the effective size of the viral population, which in turn depends on spatial structure, population size bottlenecks during host-to-host transmission, viral growth dynamics, and so on. Therefore, both selection and drift are conditioned by multiple demographical and ecological factors, which results in a variety of population genetic processes (Fig. 2). Below, we briefly summarize some of these processes.

Mutation-Selection Balance

Since natural populations tend to be adapted to their environments, most new mutations with a fitness effect are deleterious and, hence, selection tends to remove them from the population. When the influx of genetic diversity produced by mutation equals the outflow due to selection, the population reaches an equilibrium called the mutation-selection balance. In haploid populations, the population frequency of a given deleterious allele at the mutation-selection balance simply equals μ/s , where μ is the mutation rate and s the selection coefficient. The selection coefficient ranges from zero for neutral alleles to one for lethal alleles. The time for this equilibrium to be reached depends inversely on s , the equilibrium being reached instantaneously (in one generation) for $s = 1$. In RNA viruses, the statistical distribution of selection coefficients for new mutations has been characterized using site-directed mutagenesis. This has revealed that RNA viruses show very high s -values (on average, $s > 0.1$) compared to other microorganisms. Hence, selection against deleterious mutations acts strongly on RNA viruses. This implies that the mutation-selection balance tends to be reached in few generations and that, at equilibrium, most genetic variants have low population frequencies. Supporting this view, high-fidelity Illumina sequencing of experimental poliovirus populations has revealed a large number of deleterious variants at frequencies that are not detectable by conventional next-generation sequencing. A likely explanation for why s -values are high in RNA viruses is that their genomes are extremely compact, with few and often multifunctional and overlapping genes, many of which

are essential. These genome features also apply to small DNA viruses and, consequently, mean s -values are also very high for DNA viruses such as microviruses (e.g., phage Φ X174) and inoviruses (e.g., phage M13).

Viral Quasispecies

Quasispecies theory is primarily a theory of mutation-selection balance at extremely high mutation rates. Quasispecies were originally studied within the framework of physical chemistry as models for the dynamics of self-replicating sequences in a hypothetical primitive RNA world. However, due to the high mutation rates exhibited by RNA viruses, the theory subsequently became popular amongst virologists. Despite using different terminology, quasispecies and classical mutation-selection models essentially deal with the same process and are largely equivalent. Yet, quasispecies models make a more explicit approach to mutant dynamics by explicitly considering the replicative rate of each sequence as well as the probability of each sequence mutating to another specific sequence. A transition matrix is thus defined, whose eigenvector with the highest eigenvalue defines a stable mutation-selection balance. This approach defines fitness as a property of the ensemble of sequences present at equilibrium (the quasispecies). As a consequence, fitness is not merely determined by the initial sequence, but also by its mutational neighborhood. This means that a quasispecies containing a maximally fit individual sequence might be outcompeted by another quasispecies that does not contain the fittest sequence but shows higher average fitness. This phenomenon is called the survival of the flattest. A related prediction of quasispecies models is that, beyond a certain mutation rate, the fittest sequence within a given quasispecies can be outnumbered by its own cloud of deleterious mutants, a process known as error catastrophe. How this process occurs depends on the topology of the fitness landscape. Original quasispecies models assumed a landscape with a single high-fitness sequence (master) in which all mutants had the same fitness value. In this particular system, beyond the error threshold all mutants reach a frequency similar to that of the master and, since there is an extremely large number of such possible mutants, the actual frequency of the master sequence becomes vanishingly small.

Error Catastrophe and Lethal Mutagenesis of Viruses

The theory of error catastrophe suggested that viral populations could be effectively extinguished by increasing the viral mutation rate sufficiently. Treatment of RNA virus experimental populations with chemical mutagens has provided ample support for the feasibility of mutation-driven extinction in cell cultures and even in animal models. However, error catastrophe models only deal with the fitness of certain sequences relative to others, and thus ignore the fact that extinction is an absolute, demographical process that takes place when the basic reproductive number of a population decays below one. It has been thus suggested that extinction of experimental RNA populations following mutagen treatments is better defined as lethal mutagenesis process. Regardless of the models used and the mechanisms underlying extinction, empirical evidence strongly support the view that RNA viruses naturally exist at the verge of mutation-driven extinction, since extinction can be induced by modest increases in mutation frequencies. In contrast, the lower rates of spontaneous mutation shown by most DNA viruses makes them less likely to undergo lethal mutagenesis or error catastrophe.

Adaptation via Directional Selection of Beneficial Mutations

The high mutation rates exhibited by RNA viruses reduce the waiting times required for the emergence of selectively beneficial mutations compared to other microorganisms. The emergence of beneficial mutations is also favored by the extremely high population sizes of RNA viruses, which is on the order of billions of infectious particles per infected host. For a mutation rate of 10^{-5} substitutions per nucleotide per round of genome copying, replication of a population of 10^9 viral particles should produce as many as 10^4 mutations, which approaches the total number of possible single-point mutations in a typical 10 kb RNA virus genome. This makes RNA viruses extremely responsive to host adaptive immunity, antiviral drugs, and other rapidly-changing selective pressures. Yet, despite favoring the appearance of beneficial alleles, high mutation rates can hamper their fixation process in the population. First, at high mutation rates, many genomes carry several deleterious mutations, and beneficial mutations may fail to be selected if they occur in low-fitness genetic backgrounds. Alternatively, if the beneficial effect of a mutation exceeds the disadvantageous effect of other mutations present in the same genome, deleterious mutations can undergo fixation, a process known as genetic hitch-hiking. Another “traffic issue” in fast-mutating systems is clonal interference, a process whereby multiple beneficial mutations coexist in the same population and only the fittest variant is selected, leading to extinction of other beneficial variants. Since lower mutation rates increase waiting times for the emergence of beneficial mutations but higher mutation rates reduce the efficiency with which beneficial mutations are selected, adaptation should be maximal for some intermediate, optimal mutation rate. A similar optimality argument can be used for neutral evolution. Analysis of empirical data suggests that RNA viruses replicate at nearly-optimum mutation rates. Lastly, clonal interference, counter-selection of beneficial alleles in low-fitness genetic backgrounds and hitch-hiking can all be dissipated by recombination, since recombination can break disadvantageous genetic combinations and create new ones.

Frequency-Dependent Selection

The fitness of a given virus can be modified by other viruses of the same population. Such interactions can be cooperative or competitive, and are more likely to occur among viruses infecting the same cell because this allows them to share proteins. Coinfection of a cell with independent virions requires a high local density of viral particles. However, many viruses spread as collective infectious units composed of multiple particles, which promotes coinfection even at low overall viral densities. When a situation in which each cell receives multiple viral particles is maintained for several infection cycles, defective viruses tend to invade the population. Defective viruses, produced by illegitimate recombination, lack large portions of the genome and can only replicate as hyper-parasites at the expense of complete viral genomes present in the same cell. In addition to replicating faster than the wild type because their genomes are shorter, defective viruses often interfere with some steps of the infectious cycle. Hence, the fitness values of both the wild type and the defective viruses are frequency-dependent. Viral interference can produce sharp drops in fitness, which has led to suggest defective interfering viruses as possible antivirals. The emergence of interfering viruses can also be promoted by chemically induced mutagenesis in RNA viruses. On the other hand, coinfection of cells with different virus variants may allow for cooperative virus-virus interactions. In principle, two viruses each carrying a different genetic defect can mutually increase their fitness values if they share gene products (genetic complementation). It is also possible that each variant has optimized a different aspect of the infection cycle, and that such variants achieve higher fitness in coinfection than alone. This type of cooperative interaction has been suggested for measles virus and influenza virus, and might contribute to the maintenance of viral genetic diversity because fitness would be maximal at intermediate frequencies of each variant in the population. Therefore, diversity can lead to negative or positive virus-virus interactions. In turn, these interactions feedback on diversity by modifying the population frequency of certain variants.

Random Genetic Drift

Despite their enormous population sizes, viruses experience significant genetic drift. This is because the strength of drift depends on the effective population size, not on the census size. Demographical bottlenecks have a strong and long-lasting effect on effective population sizes, and such bottlenecks occur extensively during inter-host transmission in nearly all viruses. The size of the founder population following inter-host transmission has been estimated using genetic markers, and is typically lower than 10 genome copies. This number, though, can vary widely depending on transmission route (aerosol, fecal-oral, sexual, insect-borne, etc.). Additionally, viral infections show extensive intra-host spatial structure caused by restrictions to viral trafficking among tissues or organs (organ compartmentalization) as well as by the limited diffusion of virions in the extracellular milieu (localized infection foci). Spatial structure also reduces effective population sizes, hence promoting drift. The immediate consequence of drift is the loss of genetic diversity, because only a subset of the population contributes to the next generation. A longer-term consequence is the random fixation of alleles in the population. In highly variable populations such as those of RNA viruses and at sufficiently low population sizes, drift can result in the loss of the wild-type sequence. In the absence of back mutation or recombination, this loss is irreversible and hence leads to the accumulation of random mutations in the population, most of which are neutral or deleterious. This process, known as Muller's ratchet, reduces mean population fitness and can endanger the survival of small RNA virus populations.

Implications for Viral Disease

Viral population genetic diversity plays a major role in the ability of viruses to cause disease. In general, pathogens evolve faster than their hosts owing to their shorter generation times and higher population sizes. In addition, in their evolutionary arms race against hosts, most viruses benefit from a faster mutation rate, particularly RNA viruses. On the other hand, hosts have the ability to deploy complex antiviral responses such as adaptive immunity. As a result, viruses are subject to continual selective pressure for avoiding antiviral defenses.

Short-Term Pathogenesis

Different early processes including efficient entry, local replication, and rapid spread determine pathogenesis. The association between RNA virus population diversity and pathogenesis has been examined using polymerase fidelity variants. Increased-fidelity variants are typically attenuated in animal models, suggesting a role for viral population diversity in causing disease. This was first shown using a variant of poliovirus carrying a single amino acid replacement in the polymerase that confers an approximately three-fold increase in replication fidelity. Since these experiments involved acute infections, adaptive immunity was probably not responsible for the observed association between population diversity and pathogenesis, and the mechanism linking intra-host viral diversity with pathogenesis has not been elucidated. Importantly, low-fidelity variants have also been tested and found to be attenuated. A candidate alternative explanation for these results is that in RNA viruses, high mutation rates have been selected as a byproduct of selection for high replicative speed. Increasing fidelity would come at the cost of slower growth, hence producing an attenuated phenotype. On the other hand, decreasing fidelity further would elevate the mutational load of the population, reducing mean population fitness at mutation-selection balance or promoting error catastrophe.

Immune Escape

In chronic infections such as those caused by HIV or hepatitis B and C, among others, adaptive immune responses involving neutralizing antibodies and cytotoxic T cells select for escape mutants during the course of infection. In HIV-1, the ability of the virus to escape immunity partially correlates with its ability to generate a genetically diverse intra-host population. DNA viruses such as, for instance, herpes simplex virus also cause chronic infections and, hence, have to cope with the adaptive immune response as well. Yet, their strategy for evading immune responses appears to be based on latency rather than on the continual production of immune escape variants. In viruses causing acute infections, adaptive immunity escape mutants are also selected, but this process occurs mainly at the inter-host population level. For this selection to occur, hosts need to recover from infection. This creates a host subpopulation immunized against the virus variant that originally infected them, thus favoring escape variants. This process has been amply documented for influenza virus, and is called antigenic drift or antigenic shift, depending on the diversity-generating process at play. Antigenic drift is mainly based on point mutations that change epitope sequences, whereas antigenic shift occurs as a result of segment reassortments that produce new subtype combinations. Finally, immune escape is not restricted to vertebrate viruses. For instance, bacteria use clustered regularly interspaced short palindromic repeats (CRISPR) and associated proteins as an adaptive immune system. Bacteriophages can escape CRISPR via point mutations and have also evolved anti-CRISPR proteins (Acrs) of variable strength.

Drug Resistance

Antiviral drugs impose a strong selective pressure on viruses causing chronic diseases such as HIV-1. The first attempts to treat AIDS using base analogs failed due to the emergence of drug-resistant mutants. Monotherapies were subsequently replaced with combination therapies aimed at preventing drug resistance, which have been successful despite the remarkable genetic variability of HIV-1 intra-host populations. A similar scenario applies to hepatitis C virus, in which resistance against protease inhibitor-based monotherapies emerges rapidly through one or few mutations, thus requiring the use of combination therapies. Targeting different steps of the viral infection cycle, such as for instance nucleoside analog-based polymerase inhibitors and protease inhibitors reduces the risk of resistance compared to combination therapies aiming at a single target. An alternative antiviral treatment consists in taking advantage of the highly error-prone replication of RNA viruses to promote viral extinction via lethal mutagenesis using nucleoside analogs or other mutagenic compounds. This type of treatment can select for high-fidelity polymerases that partially compensate the mutagenic effect of the drugs, although this might not be sufficient for preventing viral extinction. In other cases, drug-resistance against mutagenesis may fail to evolve. Finally, yet another alternative treatment strategy that has been designed to minimize the evolution of drug resistance consists in targeting host factors such as molecular chaperones, the viral receptor, or proteins involved in the immune response.

Conclusions

Viruses are a highly heterogeneous group of molecular parasites showing varied replication strategies, genome organizations, mutability, and evolutionary properties. RNA viruses are characterized by their elevated mutation rates and extremely high genetic diversity, but reverse-transcribing DNA viruses, single-stranded DNA viruses, and some small double-stranded DNA viruses can also achieve high levels of population genetic diversity. Here, we have shown how viral diversity is generated by widely different molecular mechanisms, including lack of proofreading, avoidance of DNA repair, host-mediated viral genome editing, and specific diversity-generating modules, as well as different mechanisms of recombination. However, viral diversity does not only depend on these purely genetic factors. Ecological and demographic processes determine the action of natural selection and random genetic drift, producing a myriad of different population genetic processes, of which pathogenesis, immune escape, and drug resistance are observable consequences. Therefore, viral population diversity and evolution should help us develop more efficient antiviral control and treatment strategies.

Further Reading

- Andino, R., Domingo, E., 2015. Viral quasispecies. *Virology* 479–480, 46–51. doi:10.1016/j.virol.2015.03.022.
- Cvijović, I., Nguyen Ba, A.N., Desai, M.M., 2018. Experimental studies of evolutionary dynamics in microbes. *Trends in Genetics* 34, 693–703. doi:10.1016/j.tig.2018.06.004.
- Dolan, P.T., Whitfield, Z.J., Andino, R., 2018. Mechanisms and concepts in RNA virus population dynamics and evolution. *Annual Review of Virology* 5, 69–92. doi:10.1146/annurev-virology-101416-041718.
- George, C.X., Gan, Z., Liu, Y., Samuel, C.E., 2011. Adenosine deaminases acting on RNA, RNA editing, and interferon action. *Journal of Interferon Cytokine Research* 31, 99–117. doi:10.1089/jir.2010.0097.
- Gutiérrez, S., Michalakis, Y., Blanc, S., 2012. Virus population bottlenecks during within-host progression and host-to-host transmission. *Current Opinion in Virology* 2, 546–555. doi:10.1016/j.coviro.2012.08.001.
- Harris, R.S., Dudley, J.P., 2015. APOBECs and virus restriction. *Virology* 479–480, 131–145. doi:10.1016/j.virol.2015.03.012.
- Irwin, K.K., Renzette, N., Kowalik, T.F., Jensen, J.D., 2016. Antiviral drug resistance as an adaptive process. *Virus Evolution* 2, vew01. doi:10.1093/ve/vew014.
- Kautz, T.F., Forrester, N.L., 2018. RNA virus fidelity mutants: A useful tool for evolutionary biology or a complex challenge? *Viruses* 10, E-600. doi:10.3390/v10110600.
- McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T., 2016. Reassortment in segmented RNA viruses: Mechanisms and outcomes. *Nature Reviews Microbiology* 14, 448–460. doi:10.1038/nrmicro.2016.46.

- Perales, C., Domingo, E., 2016. Antiviral strategies based on lethal mutagenesis and error threshold. *Current Topics in Microbiology and Immunology* 392, 323–339. doi:10.1007/82_2015_459.
- Retel, C., Märkle, H., Becks, L., Feulner, P.G.D., 2019. Ecological and evolutionary processes shaping viral genetic diversity. *Viruses* 11, 220. doi:10.3390/v11030220.
- Sanjuán, R., 2017. Collective infectious units in viruses. *Trends in Microbiology* 25, 402–412. doi:10.1016/j.tim.2017.02.003.
- Sanjuán, R., Domingo-Calap, P., 2016. Mechanisms of viral mutation. *Cellular and Molecular Life Sciences* 73, 4433–4448. doi:10.1007/s00018-016-2299-6.
- Simon-Lorière, E., Holmes, E.C., 2011. Why do RNA viruses recombine? *Nature Reviews Microbiology* 9, 617–626. doi:10.1038/nrmicro2614.
- Weller, S.K., Sawitzke, J.A., 2014. Recombination promoted by DNA viruses: Phage λ to herpes simplex virus. *Annual Review of Microbiology* 68, 237–258. doi:10.1146/annurev-micro-091313-103424.

Mechanisms of RNA Virus Evolution

Lisa M Bono and Siobain Duffy, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Evolvability The potential for a biological entity (such as an RNA virus) to create new, beneficial phenotypes, typically through rapid genetic adaptation. RNA viruses are often considered very evolvable because their high mutation rates create many different genomes that can have different phenotypes, a small number of which may be beneficial.

Fitness The ability of a genotype to produce descendants (grandchildren and beyond) in a particular environment.

Genomic architecture The structure of a genome, which can involve its nucleic acid, whether it is single- or double-stranded, whether or not it is segmented or has a single genomic molecule, gene content and order and secondary structures that form in the genome.

Genotype The set of alleles carried in an individual.

Homologous Adjective describing similarity between two items due to common descent. Used here in the context of homologous sequences, which would be similar or identical regions of two or more RNA viral genomes.

Life history The sequence of events relevant to survival and reproduction from birth to death. Life history traits of viruses including adsorption rate, lysis time, and burst size, have a direct bearing on fitness and have been observed to vary widely. Since resources are not unlimited, life histories strategies are often constrained by trade-offs, for instance, increases in fitness of one trait decrease the fitness of another trait.

Nonsynonymous A genetic mutation that leads to a change in amino acid, which causes a change to the encoded protein. For example a uracil-to-cytosine mutation, causing a codon to change from guu to gcu, changes a valine into an alanine.

Phenotype The physical manifestation of the genotype in a particular environment. It is the set of observable

characteristics, which are influenced by the genotype, the environment, and any interactions between genotype and environment.

Phylogeny A representation of the evolutionary relationships among lineages, such as isolates of a viral species descended from a common ancestor. Evolution leads to branching patterns as lineages diverge. Viral phylogenies are based on genomic similarity, which can be used to infer evolutionary history of a lineage. Phylogenies are useful in molecular epidemiology of outbreaks.

Reassortment A form of genetic exchange that occurs in RNA viruses when multiple segmented viruses co-infect the same host cell; the progeny virion can package a mix of the parental segments into a single progeny virion.

Recombination A form of genetic exchange that occurs in RNA viruses when the RdRp falls off of one template and reattaches to another template.

RNA-dependent RNA polymerase (RdRp) The enzyme all RNA viruses have that complements a single-stranded RNA genomic segment for replication and to produce transcripts.

Secondary structure Regular structures formed by the primary sequence of a macromolecule, in this case the RNA genome, which folds into hydrogen bonded stems with regions of unpaired loops.

Substitution rate The rate at which mutations come to very high frequency (ideally present in 100% of individuals) in a population or a lineage, typically measured in years.

Synonymous A “silent” genetic mutation that does not lead to a change in amino acid, preserving the encoded protein. For example a uracil-to-cytosine mutation, causing a codon to change from guu to guc, does not change the amino acid – both call for valine.

RNA viruses have a well-deserved reputation for fast evolution compared to DNA-based life, and most DNA viruses. They are known to have very high mutation rates, change and adapt quickly, are linked to the bulk of host-shifting zoonotic diseases and often confound human abilities to make durable vaccines. However, while RNA viruses may make use of strategies that many other viruses and cells do not, they evolve according the same principles that govern the rest of nature; evolution occurs through mutation, recombination (including the migration of alleles from other lineages into a viral genome), selection and drift. The fact that RNA viruses show different evolutionary patterns than DNA organisms comes from a combination of their genetics, life history and environmental factors. In addition, not all RNA viruses employ the same strategies and evolve in the same way or at the same rate: there is tremendous diversity among RNA viruses.

RNA Viruses Have High Mutation Rates

Evolution is fueled by mutations – changes to genomes that occur during genomic replication and at other times. They make up the bulk of the variation upon which natural selection and drift (see below) acts. These mutations can change the proteins coded by a virus, the regulation of a gene, affect a secondary structure important for function, or may not have any effect on the phenotype of the virus at all. Since any virus capable of successful infection has sufficient fitness, most mutations decrease fitness, harming the virus that carries them.

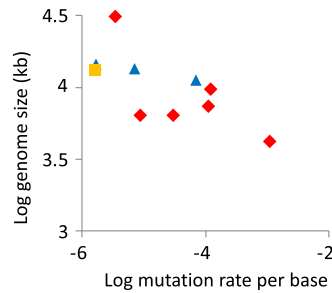


Fig. 1 RNA viruses with rigorously measured mutation rates (per base, per infection), plotted by their genome size. Mutation rates from positive sense ssRNA viruses (bacteriophage Qbeta, human rhinovirus 14, poliovirus 1 and tobacco mosaic virus, hepatitis C virus, tobacco etch virus, murine hepatitis virus) are shown as red diamonds, negative sense ssRNA viruses, (vesicular stomatitis virus, influenza A virus and influenza B virus) are shown as blue triangles and the sole measured mutation rate from a dsRNA virus (bacteriophage phi6) is shown as a yellow square. Data are from Rafael Sanjuán's viral mutation rate compendium (see relevant web pages, below).

Some mutations do not have much of an effect on the phenotype (are neutral or nearly neutral), and a small fraction of mutations are beneficial. Even though most mutations are not beneficial, biological entities like viruses with short generation times and many offspring per infection can sample a lot of mutations and stumble across beneficial mutations faster than cellular organisms.

Much of that ability to sample many mutations in a short period of time is due to RNA viruses having high mutation rates. All RNA viruses have higher mutation rates than their hosts – a hundred to a million times higher. However, not all RNA viruses have the same mutation rate. Viruses with single-stranded genomes have faster mutation rates than the sole double-stranded RNA virus with a measured mutation rate (bacteriophage phi6, 2×10^{-6} mutations/base/infection). The mutation rates of RNA viruses are thought to be as high as they possibly can be without creating too many mutations per generation to maintain the information encoded in their genomes. Too many mutations lead to a build-up of deleterious mutations, dramatically reducing fitness and even resulting in a critical and irreversible loss of genomic information. Viruses of roughly similar genome sizes have mutation rates that vary by almost 1000-fold, which means that there is no common, or optimal mutation rate for all RNA viruses (Fig. 1). One of the main reasons RNA viruses have such high mutation rates is the RdRps that all RNA viruses use to replicate their genomes typically lacks any ability to proofread and correct errors during replication. Exceptions to this are larger RNA viruses (family *Nidovirales*), which have an RdRp that is phylogenetically distinct from the RdRp shared by most RNA viruses, which have genomes ≤ 20 kb. These larger RNA viruses (up to ~ 41 kb) have a unique part of their RdRps which has the ability to cut out nucleotides from the newly synthesized RNA. This proofreading reduces the mutation rate compared to just the polymerization of the RdRp, and permitted genome sizes of these viruses to grow up to double the size of those RNA viruses without it (note the data point from murine hepatitis virus, a member of *Nidovirales*, which has the largest genome size of all viruses in Fig. 1).

One of the reasons RNA viruses have such high mutation rates is because more careful polymerization takes time, and faster replication is inherently sloppier replication. All else being equal, a virus with faster replication, and therefore a higher mutation rate, can have more offspring in the same time as a virus with slower replication, and therefore a lower mutation rate. Or the faster replicating virus can burst its cell earlier and its offspring will have an advantage finding their next host cell compared to the slower replicating virus. More of the offspring of the faster and sloppier replicating virus may have mutations that inactivate the virus, or reduce its fitness, but some of their offspring will be successful. Overall, this has been a more successful strategy for RNA viruses than a slower and more careful replication speed.

The mutation rate of RNA viruses is evolvable, but mostly in the direction of a lower mutation rate – higher mutation rates are not tolerated well, which is one of the reasons scientists feel RNA virus mutation rates are quite close to the highest they possibly can be. In the laboratory, some RNA viruses have evolved lower RdRp mutation rates. Multiple groups who have passaged poliovirus in the presence of a mutagen to increase the extrinsic mutation rate have evolved a more fidelitous RdRp, which has a G64S mutation in its polymerase. When scientists have used mutagens to increase RNA virus mutation rates, the increasing number of mutations per genome causes the viruses to poorly replicate and make few viable offspring, eventually killing the population. This lethal mutagenesis is one of the ways some of the drugs clinically used against RNA viruses and retroviruses can work. Ribavirin is an antiviral that is an RNA base analog that can be incorporated into RNA viral genomes and complements both uracil or cytosine, increasing the virus' mutation rate.

In addition to mutagens like ribavirin, many hosts make enzymes that can increase the mutation rate of RNA viruses. The best studied of these is the cytidine deaminase APOBEC3G that is part of the innate antiviral defense of some mammals, including humans. This enzyme targets cytosines in viral genomes, both the DNA form of retroviruses and hepatitis B and the RNA genomes of viruses such as measles, mumps and respiratory syncytial virus. These enzymes can turn many cytosines into uracils on a single viral genome, and its effect was termed "hypermutation" when discovered in the early days of HIV sequencing. Host cells also make adenine deaminases (ADAR) that have been observed editing the bases of RNA viruses, including vesicular stomatitis virus (VSV) and measles. Both of these mechanisms are thought to be the cell using an increased mutation rate as a way of inactivating RNA viruses – encouraging lethal mutagenesis.

Finally, just as RNA viruses themselves have a wide range of mutation rates, varying over more than 100-fold, mutation rates and frequencies of RNA viruses vary even within a single species. Host environment is an important factor for mutation, with plant viruses like cucumber mosaic virus having different numbers (and kinds) of mutations depending on whether it infects a pepper or tobacco plant. The same effect has been seen in the arbovirus VSV, which infects both mosquitos and mammals: VSV has a lower mutation rate in insect cells than in mammalian cell culture. Understanding how and why hosts influence the mutation rate of their invading viruses is an open area of research, but a similar effect has been seen in long-term viral substitution rates. Scientists measure mutation rate as numbers of mutations per base, per round of genomic replication or per infected cell. However, they measure genomic evolution in numbers of very high frequency mutations (ideally present in all members of a population), per base, per year or other unit of time. RNA viruses with different mutation rates often have similar rates of evolution through time, and scientists have endeavored to find out why some RNA viruses do evolve faster than others. A meta-analysis showed that the sole determinant of faster evolution in mammalian RNA viruses was primarily infecting epithelial cells – and slower evolution was associated with infecting neural cells. There may be a relationship between the replicative ability of the host cells and evolutionary rate – epithelial cells divide the fastest of all mammalian tissues, and neurons the slowest – but this is another example of host cells influencing RNA virus evolvability.

Genetic Exchange

The ultimate source of most novel genetic variation is mutation, but genetic exchange through recombination and reassortment brings together different variants to make potentially novel genotypes. Additionally, there is the potential for recombination to make novel alleles of genes not yet generated by viral mutation, such as when viruses recombine with homologous sequences in transgenic plants. Recombination occurs when a viral genomic segment has more than one parental sequence; reassortment occurs when different genomic segments in a segmented or multipartite virus have different parents. Both kinds of viral gene exchange have the potential to create progeny with much higher levels of variation than a single generation of replication with the virus' error-prone RdRp alone. However, gene exchange need not create novel variation – identical viruses can undergo the processes of recombination and reassortment and produce the identical offspring to the parents (Fig. 2).

Recombination

While most people are familiar with recombination as the reciprocal gene exchange that occurs in diploid, sexually reproducing eukaryotes, viruses do not undergo reciprocal gene exchange. Instead, as is the case throughout the microbial world, one viral genome is the donor to another, which is the acceptor. In RNA viruses, recombination is usually homologous, meaning that the sequences of the donor and acceptor parent strains are very similar at the site of recombination, which usually leaves the structure and protein-coding capacity of the recombination genome unchanged (Fig. 2). Recombination can also be non-homologous (illegitimate), which includes both unrelated viruses without homologous regions recombining, or the site at which related viruses are recombining is not one of high sequence similarity. Such non-homologous recombination is usually deleterious, which is perhaps why it is observed less frequently than the homologous type. Recombination between viruses always occurs during co-infection of the same cell by multiple viruses. Very rarely viral RNA can recombine with host RNA, but recombination with host genetic material is much more common among DNA viruses.

Copy-choice is the most widely accepted mechanism for RNA viral recombination and is also termed replicative recombination. Originally described in poliovirus, recombination occurs when the RNA polymerase falls off the donor template and reattaches to an acceptor template – thus switching which parent it copies from for the offspring sequence – during genome replication. The site at which the polymerase switches templates is called the breakpoint, and homologous recombination is encouraged by the partially

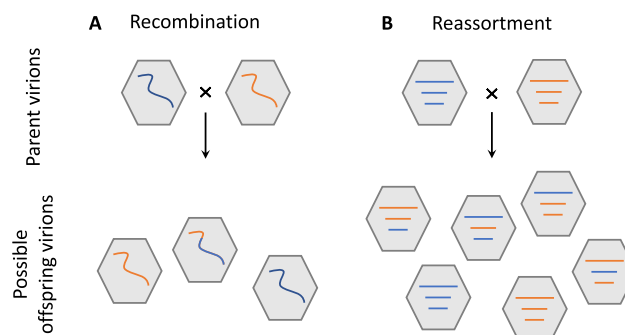


Fig. 2 Forms of genetic exchange among RNA viruses. During co-infection, multiple parental genomes can be replicated simultaneously, and offspring virions package with a combination of the parental genomes. (A) Recombination produces offspring that have a genome segment that is a hybrid of the parental genomes. Here, we depict homologous recombination. (B) Reassortment produces offspring that have a combination of genome segments from the parental genomes.

sequenced offspring genome (or genomic segment) complementarily pairing to the new template. The RdRp binds to the new template, continuing to synthesize the progeny RNA, creating a chimera of the parental templates. Sequence similarity between the templates at the breakpoint can encourage recombination, helping reduce the probability of non-homologous recombination, which is more likely to produce deleterious genotypes. The kinetics of replication and the secondary structure of RNA may also influence the likelihood of switching templates, such as when a part of the genome is covered in nucleoproteins and is less accessible for replication. Secondary structures are also known to create hotspots of RNA virus recombination.

There are other mechanisms of non-replicative recombination that follow a breakage-joining model, where genetic fragments from different viruses join to form hybrid genomes. As breaks can theoretically occur anywhere in the genome, the mechanism can be either homologous or non-homologous, but it is more likely to be non-homologous given that it can occur randomly. Although non-replicative recombination *in vivo* is possible, it has only been observed in cell-free laboratory settings and it is not known what, if any, contribution non-replicative recombination makes in nature.

Reassortment

Reassortment is another form of genetic exchange that can occur in segmented viruses—viruses that have their genome split into multiple segments. Reassortment only occurs when multiple viruses co-infect the same cell, and replicate their progeny segments in the same cytoplasm. The progeny virions are assembled from a mix of segments from the parental virions, creating a novel assortment of genome segments – a hybrid – into a capsid (Fig. 2). In most segmented viruses, such as those from the families *Cystoviridae*, *Orthomyxoviridae*, and *Reoviridae*, segments usually neatly replace each other, so each progeny virion gets a single copy of each segment. Unlike recombination, which can occur nearly anywhere in the genome, creating chimeric genomic segments, reassortment swaps entire genes or sets of genes with no chance of disturbing gene coding. In multipartite viruses – viruses for which each genome segment is packaged in a separate virion – reassortment is stochastic, with segments from parents resulting in infection events with a random mix of segments. Interestingly, reassortment in multipartite viruses need not occur in infected cells, or even when multiple viruses coinfect the same multicellular host; it can also occur during vector transmission. Also, segmented and multipartite viruses can also undergo both reassortment and recombination, *e.g.*, rotavirus, influenza A, hantavirus, flock house virus, and plant viruses, although one form of genetic exchange may be more frequent than the other depending on the virus.

Even though the exact details of how reassortment occurs across viruses differ, some generalizations about the process can be made. After the genome segments have been replicated, packaging signals indicate how the genome segments should be incorporated into the viral capsid, regardless from which parent the segments are derived. Although the parental virions may differ, the packaging signals need to be compatible enough for the segments to assemble in a single virion. The exact mechanisms of reassortment vary among different viral families. The process of reassortment is well understood in the bacteriophage $\phi 6$ (*Cystoviridae*), which is described here as an example. This phage has a genome divided into three segments, each of which is individually inserted into the procapsid in sequence. Encapsidation of each segment triggers a conformational change, revealing the binding site specific for the next segment to be packaged. Once encapsidation is complete, the procapsid expands, triggering the segments to synthesize the negative-sense strand of their dsRNA genome and completing virion morphogenesis. The poster child for a frequently reassorting virus would be Influenza A, which has eight genomic segments. Its reassortants are closely watched by public health researchers because successful, human-infecting reassortants can lead to epidemics and the failure of that year's flu shots, which were designed based on influenza evolution in the absence of reassortment.

Rates of Recombination

Recombination and reassortment rates vary across viruses, but nearly all viruses recombine, and segmented and multipartite viruses are all thought to reassort. Some families show higher rates than others, *e.g.*, (–)ssRNA viruses tend to have low rates of recombination, and (monopartite) segmented viruses tend to exhibit lower levels of recombination. On the other hand, multipartite viruses have been observed to have higher recombination rates in addition to exchanging genes *via* reassortment. Furthermore, recombination rates vary according to environmental factors. Since recombination between different viral strains requires that the host cell be co-infected by those strains simultaneously, factors that influence the incidence of co-infection, *e.g.*, virion particle density in the environment surrounding the host cell, can influence how often recombination occurs. Reassortment rates among the family *Cystoviridae* have been observed to vary markedly by geographic location.

Detecting Recombination

All recombination detection methods look for similarity between regions of related genomes, and then identify disruptions in the pattern of the relationships between regions. This can be achieved by comparing phylogenetic trees that have been constructed for different components on a set of virus genomes, though not all methods require phylogenetic incongruence. The incongruences between phylogenetic trees can be visualized using tanglegrams (Fig. 3). A link to a Windows-native freeware program that unites many recombination detection methods together in one graphic interface is given in relevant web links (below).

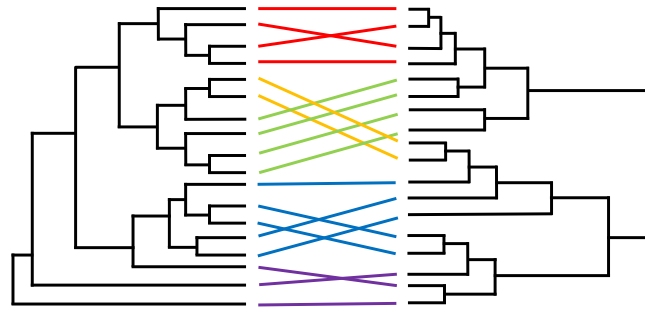


Fig. 3 A cartoon example of a tanglegram, a way of visualizing phylogenetic incongruence, which can provide insights into genetic exchange between different viral groups. The left and right phylogenies represent two different genes in a viral species. The colored lines connect the two genes from the same individual virus. Because the phylogenies are not identical, the colored lines sometimes cross each other, which is an indication of gene exchange. The two genes do not have an identical evolutionary history in this virus.

Evolutionary Outcomes for Recombination and Reassortment

Like mutation, most gene exchanges are not thought to be beneficial, even when the recombination is homologous. However, there are many well-studied examples of the rare beneficial outcomes of gene exchange, where recombination led to host range expansion, transmission *via* new vectors, increases in virulence and pathogenesis, evasion of host immunity, antiviral resistance, or the ability to infect new tissue types. Like winning the lottery, hybrid offspring may receive a highly beneficial mix of their parental genomes and also be in an environment where this genetic combination allows them to have a selective advantage. One of the best-described examples of this is in the Western equine encephalomyelitis (WEE) complex of viruses, an alphavirus that formed as the result of a recombination event between Eastern equine encephalomyelitis virus (EEEV) and Sindbis-like viruses that probably occurred within the last 2000 years. This recombination event led to the evolution of a predominantly EEEV-like virus that inherited a capsid from its Sindbis-like ancestor. Similarly, reassortment can confer great fitness advantages, such as when reassortment among circulating strains of influenza has repeatedly led to competitively superior strains of influenza A. Several instances of strains endemic to swine or birds have been able to jump into humans after a reassortment event, *e.g.*, the 1957 'Asian', 1968 'Hong Kong', and 2009 'swine flu' pandemics. The first two pandemics were the result of reassortment between a circulating human and avian strains, while the latter pandemic arose from multiple reassortment events between highly divergent North American and Eurasian swine, avian, and human strains. Such cross-species reassortment events have often been observed in concert with the emergence of pandemic strains of influenza. The introduction of alleles to a different influenza strain is a form of migration, which is rarely discussed in virology as separate from recombination and reassortment.

In addition to hybrids, recombination and reassortment can also generate defective interfering particles – a type of social cheater that can only replicate in the presence of another virion that can supply the missing functions. These particles can usually tolerate large deletions while maintaining crucial genes for replication and packaging but rely on the co-infecting virus to supply those missing functions. These cheaters benefit disproportionately from this relationship, gaining an intracellular fitness advantage due to their small size and fast replication while interfering with the replication of the co-infecting virus.

Natural Selection in Viruses

After a population of RNA viruses has generated genetic diversity through mutation and recombination, the varied viral genotypes in the population will be subjected to selection. If a viral genotype makes a phenotype with a fitness advantage – an advantage over other viral genotypes in the population – then it will leave more offspring, and over time its descendants will increase in frequency in the population. Otherwise, if the phenotype has a lower fitness level, that genotype will likely be pruned from the population over time. The fitness of an individual is its survival and reproductive success relative to the rest of the population. Evolution by natural selection occurs in all biological entities, and is an inevitable consequence of the three following conditions: (1) variation in phenotypic traits must exist, (2) the phenotypic variation must correlate with genetic variation, and (3) individuals with these different traits have differential reproduction. For instance, imagine there is variation within a population of viruses in how a virus attaches to a receptor on the surface of a cell. The variation for attachment has an underlying genetic basis, *i.e.*, different genotypes attach differently. The genotypes that attach more securely are able to produce more progeny virions and be proportionally more numerous in the next generation, and therefore are of higher fitness. Which genotypes are most fit in a population is a function of a large number of biotic and abiotic factors, and the same population of viruses would have different genotypes positively selected in different environments.

Three common kinds of effects of natural selection can be discussed in terms of the interactions of viruses and host immune systems: positive (adaptation), purifying, and diversifying (or disruptive) selection (**Fig. 4**). Positive selection is directional – phenotypes of increasing extremes of a trait are selected as genotypes at the other end of the distribution of the trait leave fewer and fewer offspring. One example of this is the CpG content of Influenza A genomes since the host-shift of this virus into humans from

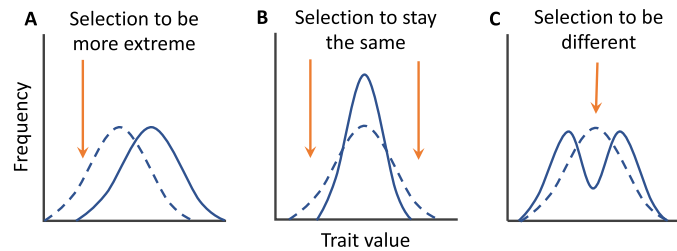


Fig. 4 Forms of selection and how they impact the distribution of trait values. The dashed line indicates the distribution prior to selection. The solid line shows the distribution after selection has acted on the population. The arrows indicate which portion of the trait values were selected against. (A) Positive selection acts against trait values from one extreme, resulting in a distribution shift away from that extreme. (B) Purifying selection acts against trait values from both extremes of the distribution, resulting in a narrower yet higher peak. (C) Diversifying selection acts against the mean of the distribution, favoring extreme trait values. In this single-axis example, it results in a bimodal distribution of trait values.

birds in the early 1900s. Mammals, but not birds, recognize cytosines next to guanines in RNA (CpG sites) as a sign of foreign genes in the cell, which triggers an innate immune reaction. Influenza A strains that first encountered humans in the 1918 pandemic had many CpG sites, and one theory of why the pandemic killed healthy adults more often than the young and old is that healthy adults mounted a huge immune response to these genomes, resulting in frequently fatal cytokine storms. Over the last century, the CpG content of circulating flu strains has steadily decreased – evidence of selection against high CpG strains and for lower CpG strains in a directional manner.

While positive selection is the most intensively studied of the kinds of selection discussed here, purifying selection is the most common kind of pressure on viruses (and all biological entities). Since most mutations and gene exchange events hurt fitness, most phenotype-affecting mutations will cause their genotypes to leave fewer and fewer offspring over time, and the affected genotypes will drop out of the population. This is the strong selection to maintain the exact function of all the proteins in a virus by maintaining the exact protein sequences. Purifying selection is exemplified by arboviruses that use an insect vector between mammalian hosts. The exterior proteins of arboviruses, such as the capsid proteins, are under stronger purifying selection and must maintain their exact protein sequences more than related viruses that do not use vectors. This is because the coat proteins must be functional in two very different animal cell types – and mutations have a very high probability of disrupting fitness in at least one of the hosts.

Diversifying selection also plays a significant role in viral evolution. Mammalian immune systems recognize viral epitopes and develop antibodies to neutralize these viruses upon further encounters (or in the case of vaccine-induced immunity, upon first encounter). For instance, after a flu season, most people are immune to the most prevalent flu strain's epitope, and that exact strain will not re-infect those who have already developed antibodies to it. However, any differences in the epitope, even slight ones, may allow a virus to break that immunity. Variants are more successful than the most prevalent strain of the prior year – natural selection favors strains that are different, but not in a particular, directional way.

The most common way to detect whether a gene is under selection is d_N/d_S : the ratio of the rate of nonsynonymous changes to the rate of synonymous changes. More nonsynonymous mutations persisting over time means that changes to the protein were not deleterious, possibly even beneficial. A high d_N/d_S ratio indicates positive or diversifying selection. A higher rate of synonymous substitution means that changes to the protein were deleterious and quickly removed from the population, which indicates purifying selection – changes to the protein sequence of the gene do not survive long in the population. Sometimes virologists will find a d_N/d_S ratio near unity, indicating that both nonsynonymous and synonymous mutations are equally likely to survive, which often means that the region is evolving neutrally and is not a gene that is under selection to maintain function. A popular website that can calculate d_N/d_S from aligned datasets is given in relevant web links (below).

Coevolution With Hosts

Hosts and their viruses can exert intense selective pressure on each other. Viral infections may lead to high fitness costs or death for the host, while the viruses depend on their hosts for reproduction. RNA viruses evolve quickly, and there are endless examples of viruses evolving to better exploit their hosts. Recent human genomics research suggests defending against viral infection has been very important for human evolution over longer stretches of evolutionary time, with RNA viruses Hepatitis C and Influenza among the 10 viruses having had the greatest impact on human genetic variation. Coevolution occurs when both the virus evolves in response to the host and the host evolves in response to the virus, for instance creating an arms race of successive rounds of selection for beneficial mutations for more virulent viruses and well as stronger defenses in the host. Coevolution does not require the number of evolutionary changes to be equal, as fast-evolving RNA viruses are expected to respond to selection on hosts more and more often than slower evolving hosts can. Reciprocal cycles of coevolution of RNA viruses and mammals are difficult to observe experimentally, given the hosts' slow generation time, but have been seen through experimental evolution: in the laboratory, and in the release of rabbit hemorrhagic disease virus as a biocontrol agent for rabbits in Australia.

On the other hand, coevolution can be detected through inferential methods. Mutually beneficial relationships between viruses and hosts imply a long-term interaction that was honed through rounds of natural selection. One example would be the

dependence of a tropical panic grass on a dsRNA virus-infected fungus, which allows the grass to grow at high soil temperatures, far above what the plant can grow alone or even with the uninfected fungus. When the fungus was cleared of the viral infection, the plants died, indicating a clear evolutionary advantage to maintaining viral infection.

More commonly, scientists take a computational approach to discern long-term relationships between viruses and hosts that could indicate coevolution. It is clear that primates and the closely related simian and human immunodeficiency viruses (SIV and HIV) are locked in an arms race, exerting strong selection on each other over long-term coevolution. The primate innate immune systems use APOBEC3G to hypermutate the viruses, and the viruses fight this with a specific virion infectivity factor (Vif) which targets APOBEC3G for destruction. However, a previously common method of looking for evidence of coevolution, matching host and virus phylogenies, which was once thought to be strong evidence of long-term relationships between RNA viruses and their hosts, has faltered with increased investigation. Fast evolving RNA viruses tend to host-jump, and often to phylogenetically related hosts, so RNA virus phylogenies can match host phylogenies for other reasons, and with increased sampling many RNA virus trees no longer match those of their hosts. Overall, the majority of measurable viral adaptation is to hosts that are not appreciably coevolving, given the timescale of observation. Most viral evolutionary biology focuses on the evolution of the virus, and considers the host more static. An important exception to this is adaptive immune systems that can and do change on timescales relevant to observable viral evolution.

Viruses are Under Selection by Factors Other Than Host Cells

In addition to the essential interactions with their hosts, viruses interact with each other at the population level (intraspecific) or with other viral species (interspecific). For instance, strong intraspecific competition for host cells among the bacteriophage $\phi 6$ drove the evolution of a phenotype that evolved to infect novel bacterial strains (host range expansion). This new phenotype could escape the intense competition by infecting the novel strains, and was favored even though the novel host was a less productive host and the expanded host range mutant virus had fewer offspring in that host. Viruses have evolved mechanisms to outcompete each other within the same cell as well. For instance, superinfection exclusion occurs when a preexisting infection prevents the host from being infected by the same or a closely related virus. This has been broadly observed across host taxa, *e.g.*, Hepatitis C virus (HCV) infection can prevent infection with another strain of HCV. Cooperative interactions can also occur during infection. Coinfection by potato virus Y (PVY) and potato virus X (PVX), which despite their similar names come from different viral families, is synergistic, both exacerbating disease symptoms and dramatically increasing the titer of PVX when compared to single infections. These 'social' interactions among viruses can also influence viral evolution.

The environment beyond the host cell can also select for different viral phenotypes. During transmission between hosts, there can be strong, even lethal, selection pressures from desiccation, pH changes, UV radiation, and high temperatures. In human environments, sanitation of surfaces also exerts a selective pressure on RNA viruses. Experimental evolution studies have shown that viruses can evolve in response to selection on survival. For instance, VSV can evolve increased thermostability outside of the cell and human echovirus 11 can become more tolerant to UV light by lowering its mutation rate.

Genetic Drift

While it is tempting to assume that beneficial mutations will always increase and deleterious mutations will always disappear from populations, this assumption is flawed. Allele frequencies can fluctuate due to sampling effects in bottlenecks by the process of genetic drift, rather than as a consequence of selection on fitness. From a mathematical perspective, the current generation is essentially a sample of the previous generation (plus mutation). In the absence of selection, the frequency of alleles present in the current generation would be solely based on their frequency in the previous generation. The presence of selection increases the frequency of alleles associated with higher fitness and reduces the frequency of alleles associated with lower fitness. However, random chance leads to sampling error, producing deviations between the expected and the observed allele frequencies. The smaller the population, such as in a population bottleneck, the greater these deviations are likely to be. Through this sampling error, genetic drift can cause beneficial alleles to decrease in frequency and even bring deleterious alleles to high frequency.

With their high mutation rates and often large population sizes, one might expect RNA viruses to escape the effects of genetic drift. However, RNA viral populations are only intermittently large, and it is during those times when population size is restricted that genetic drift can be detected. RNA viral populations experience bottlenecks at multiple steps during infection, for instance, during transmission between hosts, between tissues in a multicellular host, and in vectors. These small population sizes allow sampling error to overwhelm the effects of natural selection, which can have lasting effects for the virus population post-bottleneck. This effect is well-studied in vector-borne viruses. The bottleneck from aphids to new plant hosts has been quantified in potato virus Y and the multipartite cucumber mosaic virus, and both showed only 1 or 2 complete viral genomes pass on to a new host. The longer-term effects of drift have been studied in cucumber mosaic virus, where plant-to-plant transmission significantly reduced genetic variation. In mammalian RNA viruses like Venezuelan equine encephalitis virus and dengue, the transmission from mammal to vector is associated with tight population bottlenecks, and drift (not selection by vector cells) primarily explains the pattern of genotypes that move into the vector.

As a note on nomenclature, genetic drift has nothing to do with "antigenic drift," a term most commonly used in Influenza evolution, which is the result of natural selection to avoid immune recognition. Genetic drift is the explanation for the random change in allele frequencies – the complement to natural selection, which is responsible for non-random changes in allele frequencies.

The Quasispecies Concept

Since RNA viruses have such high mutation rates, a population of RNA viruses descended from a single ancestor is more diverse than one founded by, say, a dsDNA herpesvirus. A parent RNA virus may produce offspring where none, or only a few, are genetically identical to the parent. This has prompted some to argue that RNA virus populations (and other genomic parasites with certain high mutation rates, such as viroids, or single-stranded DNA viruses) should be considered the unit of selection instead of an individual genotype. That is, a central, or master, sequence and all of its near mutational neighbors (one or a few mutations away from the master sequence) would be the “individual” that natural selection could affect. This shift in the unit of selection allows for the cooperation of multiple genotypes as part of the same “individual” population. There is developed mathematical theory to model and describe the behavior of RNA virus quasispecies populations. However, there is significant controversy about whether or not RNA viruses have high enough mutation rates to achieve quasispecies dynamics without additional mutagens and strong mathematical arguments showing that the same population genetics (a branch of evolutionary biology) equations that describe cellular life explain the dynamics of RNA virus populations. Therefore, there is not a necessity to invoke a different frame for evolution for RNA viruses, and doing so might not even be applicable. Perhaps as a means of satisfying these two camps within evolutionary virology, the word “quasispecies” will be frequently invoked as describing the mutant cloud of descendant sequences expected during the growth of an RNA virus population – as a synonym for genetic diversity. While some RNA virologists use the conclusions from quasispecies theory to drive their RNA virus research forward, much of the community remains to be convinced by one or more experiments that demonstrate quasispecies dynamics that cannot be otherwise explained with current, widely accepted evolutionary biology theory.

Additional Constraints on RNA Virus Evolution

Genomic architecture can affect the evolution of all biological entities, but it has noticeable and significant effects on viruses. Single-stranded RNA viral genomes form stem-loop structures, where some of the genome becomes double-stranded. This alters the evolvability of the stem regions, which evolve more slowly than loops. Regions that frequently form stems are therefore more conserved, and if a rare positively selected mutation occurs in a stem, typically its complementing base also has to change – with likely neutral or deleterious effects – for the beneficial mutation to rise in frequency in the population. This constraint on stem evolvability has been demonstrated in HIV, one of a handful of RNA and retroviruses that have had their secondary structures fully experimentally mapped, and not just inferred from sequence.

RNA viruses, like DNA viruses and cellular life, can also have overlapping reading frames: where the same stretch of RNA have multiple coding regions, offset by one or two nucleotides so that different codons are read by the ribosome. Overlapping reading frames are thought to help RNA viruses maximize the utility of their limited genome sizes, though most overlapping regions are not substantial in length (<100 bases) and involve the overlap of the ends of two genes. The less common overlaps where one gene is overprinted into another reading frame for the bulk of a gene poses a strong constraint on the evolution of the region that does double duty – encoding two proteins that are likely both essential for virus function. In eukaryotic RNA viruses that express limited numbers of cistrons, some of these overlapping reading frames are caused by ribosomal frameshifts. This can lead virologists to discovering a novel ORF in otherwise frequently sequenced and studied viruses. Two examples would be the fourth open reading frame, *vf1*, discovered in murine norovirus, and the pretty interesting potyvirus open reading frame (*pipo*) which is entirely nested within one of the genes in potyviruses’ genomic polyprotein. In the latter case, *pipo* was discovered because of its constraining effect on potyvirus evolution – there was a noticeably lower d_N/d_S ratio in the same place in multiple potyvirus species, and researchers sought an explanation for the stronger purifying selection.

Conclusions

Like all areas of science our understanding of the mechanisms of RNA virus evolution is tempered by knowing that virologists and evolutionary biologists do not yet know everything. The RNA viral biosphere is undersampled relative to cellular life, and relative to DNA viruses. There may well be RNA viruses lurking in understudied hosts or environments that will expand the ranges of RNA virus mutation rates, genome length or recombination proclivity. Novel host defense mechanisms that alter viral genomes may be discovered, contributing to additional mutation biases. However the details change, the mechanisms that underlie all biological evolution will still hold in RNA viruses: variation that arises in a population through mutation, recombination or migration rises or falls in frequency due to selection and drift.

Further Reading

- Arenas, M., Araujo, N.M., Branco, C., *et al.*, 2018. Mutation and recombination in pathogen evolution: relevance, methods and controversies. *Infection, Genetics and Evolution* 63, 295–306.
- Belshaw, R., Gardner, A., Rambaut, A., Pybus, O.G., 2008. Pacing a small cage: Mutation and RNA viruses. *Trends in Ecology & Evolution* 23, 188–193.
- Geoghegan, J.L., Holmes, E.C., 2018. Evolutionary virology at 40. *Genetics* 210, 1151–1162.

- Holmes, E.C., 2009. The Evolution and Emergence of RNA Viruses. New York: Oxford University Press.
- McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T., 2016. Reassortment in segmented viruses: Mechanisms and outcomes. *Nature Reviews Microbiology* 14, 448–460.
- Pérez-Losada, M., Arenas, M., Galán, J.C., Palero, F., González-Candelas, F., 2015. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics, and Evolution* 30, 296–307.
- Simon-Lorière, L., Holmes, E.C., 2011. Why do RNA viruses recombine? *Nature Reviews Microbiology* 9, 617–626.
- Solé, R., Elena, S.F., 2019. *Viruses as Complex Adaptive Systems*. Princeton: Princeton University Press.
- Varsani, A., Lefeuve, P., Roumagnac, P., Martin, D., 2018. Notes on recombination and reassortment in multipartite/segmented viruses. *Current Opinion in Virology* 33, 156–166.
- Weaver, S.C., Denison, M., Roossinck, M., Vignuzzi, M., 2016. *Virus Evolution: Current Research and Future Directions*. Poole: Caister Academic Press.

Relevant Websites

- <https://www.uv.es/rasanve2/virmut.htm>
Rafael Sanjuan, University of Valencia, Spain.
- <http://web.cbio.uct.ac.za/~darren/rdp.html>
RDP home page.
- <https://stevenweaver.github.io/hyphy-site/methods/selection-methods/>
Selection
HyPhy
Steven Weaver.
- <https://evolution.berkeley.edu/evolibrary/home.php>
Understanding Evolution.

Mechanisms of DNA Virus Evolution

Moriah L Szpara, Pennsylvania State University, University Park, PA, United States

Koenraad Van Doorslaer, University of Arizona, Tucson, AZ, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Adaptation vs. Evolution Here we use the term “adaptation” to refer to events that occur within a host, and “evolution” to refer to those that occur over much longer spans of time. This allows us to highlight that local adaptation within a host is due to different selective pressures than those that impact transmission to new hosts, or that act across multiple generations of hosts.

Consensus genome This term refers to a genome derived by selecting the most commonly observed allele detected at each genomic position in a sequencing-based analysis of a virus sample. Compare this term to “minor variant” below.

Fossilized or endogenized viral genome When viral DNA becomes integrated into the host's germline, these endogenous viruses can be vertically transmitted. These endogenized viruses represent a molecular fossil record of past viral invasions.

Horizontal gene transfer (HGT) Refers to the movement of a fragment of genetic material between unrelated species. Viral HGT can occur between host and virus, between two viruses, or between a virus and a coincident species that enters the same host cell. Viruses are thought to be major mediators or vectors of HGT, due to their ability to introduce genetic material into new host cells and to infect multiple closely-related host species.

In vivo vs. in vitro These terms are used here to distinguish between experiments conducted within a complex host organism (*in vivo*), vs. within cells in culture (*in vitro*).

Latency A phase where a herpesvirus is present as an episome in a host cell nucleus, mostly quiescent, and not producing any lytic viral progeny.

Lysogeny A phase where bacteriophage or archaeal viruses integrate into their host genome and are propagated along with the host genome as the cell divides.

Minor variant A sequence variant which is not the most common allele in a given virus population (e.g., within an infected host). Compare this term to “consensus genome” above.

Persistent or chronic infection This term is used to refer to a long-lasting viral infection, i.e., one that exceeds the time frame of an acute infection for that virus species.

Recombination (homologous vs. non-homologous) This term refers to the joining of DNA segments after a break. Homologous recombination encompasses several mechanisms such as strand invasion, single-strand annealing, and microhomology-mediated end-joining. Non-homologous recombination involves end-joining without any homology required.

Single nucleotide polymorphism (SNP) This term is used here to denote a single nucleotide difference (allele), which is observed when comparing sequenced isolates of a given viral species.

Standing variation This refers to a viral population that contains more than one allele or variant at a given locus, or at multiple loci in the genome (e.g., within a single infected host or within a group of hosts). See also the term “minor variant” above.

Tandem repeats Short repetitive elements found in any nucleotide sequences. These are categorized based on the length of their repeating unit, n , as follows: homopolymers ($n = 1$ base pair, bp), microsatellites ($n < 10$ bp), macrosatellites ($n \geq 10$ bp), minisatellites ($n \geq 100$ bp).

Transposable elements (TEs) Transposons are segments of DNA that can move, as a unit, from one location in the genome to another.

Introduction

A historical view of viral evolution might suggest that the evolutionary processes of RNA and DNA viruses adhere to distinct and non-overlapping rules. RNA virus evolution, as covered elsewhere in this volume, involves error-prone polymerases, an inability to perform error-correction (except in rare cases such as the coronaviruses), the existence of viral quasispecies, and a constant interplay of mutation and fitness-based selection. In contrast, DNA virus evolution is often discussed in more sweeping historical terms, with a focus on how evolution has led to speciation through the slow accumulation of genetic drift and relatively rare fixation of recombination-based genetic shifts. However, there is actually much in common between the mechanisms of evolution for both RNA and DNA viruses. For instance, while the polymerases used by DNA viruses are less error-prone and can perform error-correction, the larger size of many DNA virus genomes still leaves room for the accumulation of genetic variation in every round of viral replication. Furthermore, evidence from multiple DNA viruses suggests that rather than being rare, recombination between DNA virus genomes is rampant. The progeny of these genetic exchanges go unnoticed when recombination occurs between identical or highly similar genomes, or if the progeny do not survive fitness-based selection. Host-linked evolution or co-divergence may also contribute to the apparent low mutation rates in DNA viruses. Understanding the factors that determine the rate at which viral genomes generate and fix mutations provides essential insights into their evolutionary mechanisms. We

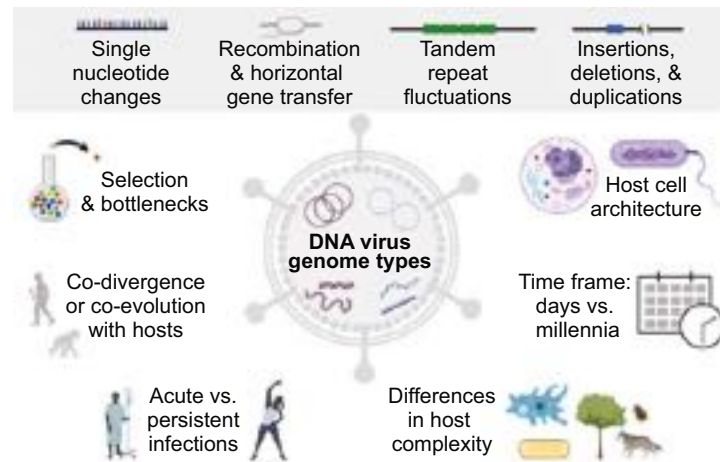


Fig. 1 DNA virus evolution relies on molecular mechanisms (top, shaded gray) which are impacted by host biology (arrayed below). DNA viruses exist in a range of genome formats (center) and sizes, each of which has a different propensity to evolve via these mechanisms. Viral genome formats include circular and linear DNA that is either single- or double-stranded, with lengths ranging from ~ 2 to > 2000 kbp. The molecular mechanisms that underlie DNA virus evolution include single nucleotide changes, recombination and horizontal gene transfer, fluctuations in tandem repeat length, and sequence gain or loss through insertions, deletions, and segment duplications. Host impacts on DNA virus evolution (listed clockwise) include host cell architecture (e.g., nucleated vs. non-nucleated host cells), the time frame being considered (e.g., one round of infection or many generations), the host complexity (single-cells vs. complex organisms), an acute vs. long-term persistent duration of host infection, selective pressures and bottlenecks that act on each virus population, and co-divergence with host species over millennia. Image created using BioRender.com and Adobe Illustrator.

cover these topics in greater detail below, after introducing a number of additional considerations to the discussion of how DNA viruses evolve (see Fig. 1 for summary).

Diversity of DNA Virus Genome Types

A simplistic division of evolutionary mechanisms for viruses is generally split based on whether the genome being considered is RNA or DNA. While a single-stranded RNA virus and a double-stranded DNA virus might be considered typical exemplars of each group, these are by no means the only genome types — there are numerous variations on these themes. The prototypic double-stranded DNA (dsDNA) viruses exist in both linear and circular forms. These dsDNA viruses run the gamut in terms of size, from tiny (~ 5 – 8 kilobase pairs, kbp) papillomaviruses and polyomaviruses, to large bacteriophage, adenovirus, and herpesvirus genomes (ranging from ~ 30 – 250 kbp), to the over-sized nucleocytoplasmic large DNA viruses (NCLDV) such as poxviruses and phycodnaviruses (~ 130 – 400 kbp), and finally the giant viruses found in algae and amoeba (upwards of ~ 1 – 2 megabase pairs, Mbp). There are also unusual genome formats among these dsDNA viruses, for instance, the covalently-closed ends of linear poxvirus genomes, or the partially gapped circular dsDNA genome of hepadnaviruses (e.g., hepatitis B virus). Also, there are abundant examples of single-stranded DNA (ssDNA) viruses, which include both linear (e.g., parvovirus and densovirus) and more numerous circular forms (e.g., circovirus, nanovirus, and geminivirus; these are also known as Circular Rep-Encoding Single-Stranded or CRESS DNA viruses). In each case, these genome formats lead to particular constraints and opportunities for the evolutionary mechanisms discussed here. We describe the evolutionary mechanisms below in light of the most common dsDNA virus examples, and where possible, we note those areas where other DNA virus genome formats may differ.

Host Cell Biology and Availability of Host Enzymes Constrains Virus Evolution

It is possible – though not advisable – to discuss the mechanisms of DNA virus evolution without considering host cell biology. This simplification is enabled by the fact that all known hosts for these viruses are DNA-based life forms, with the concomitant presence of the requisite machinery of a DNA polymerase for replication, RNA polymerase for transcription, and ribosomes for translation. The most apparent distinctions among potential hosts for DNA viruses fall along the known bifurcations of the tree of life – namely bacteria, archaea, and the major groups of eukaryotes (i.e., plants, animals, fungi, and protists). In bacterial and archaeal hosts, the absence of a nucleus removes any distinction in where DNA virus replication occurs. However, in eukaryotes, many host enzymes are constrained to the nucleus, including host DNA and RNA polymerases as well as the RNA splicing machinery, whereas translation is limited to the cytoplasm. Viruses that utilize the host DNA polymerase to copy their genomes, such as members of the *Polyomaviridae* and *Papillomaviridae*, must therefore replicate in the nucleus. Likewise, while the *Herpesviridae* and *Adenoviridae* encode their own DNA polymerase, they use the host RNA polymerase and splicing functions, restricting their replication to the nucleus. In contrast, members of the *Poxviridae* and *Mimiviridae* that replicate in the cytoplasm

encode their own DNA and RNA polymerases, whose fidelity can, therefore, evolve on a separate trajectory from that of the host. Finally, while ssDNA viruses use host DNA polymerases, their observed mutation rate far exceeds that detected in their host-cell genomes or in dsDNA viruses, suggesting that other sources of mutation such as oxidative damage and/or lack of DNA repair may be at play. For these reasons, knowledge of the host cell biology and the usage of host enzymes by a given virus species is a requirement for understanding the constraints on viral evolution.

Time Frames: Viral Adaptation Within a Host vs. Evolution Over Multiple Generations

Any discussion of the mechanisms of virus evolution needs to begin by defining the time scale under consideration. At the shortest end of this spectrum lies the time frame of a single round of viral infection. As noted below, the first infected cell may be anything from a single-celled organism to the first cellular entry point into a complex human host. From a clinical perspective, viral infection and disease are often considered on the time frame of a single individual's infection – often a human or animal subject. As described below, the virus population within a given host may undergo adaptation within the relatively short time frame of the host's infection. Mechanisms that enable diversification or speciation of a given virus usually require thousands of viral replication cycles, encompassing multiple host generations. At the grandest scale, the origins of viruses and specific lineages thereof spans the history of life on earth. The origins of viruses as we know them are covered elsewhere in this volume, so here we focus solely on the mechanisms that form the foundation of all viral adaptation and evolution. As such, we focus mostly on the time scale of an individual cell and/or host infection, which can include the contributions of virus populations that are more diverse and/or less fit than those which we see preserved over longer sweeps of evolutionary time.

DNA Virus Hosts Vary From Single Cells to Complex Multi-Cellular Organisms

An understanding of DNA virus adaptation and evolution requires a consideration of the host as a single-cell versus a complex multi-cellular organism. A basic theoretical model of viral replication would include productive viral replication in a single cell, followed by spread to nearby uninfected cells, potentially over multiple generations. This model may well apply to bacterial and archaeal cells, and to single-celled eukaryotic species such as marine alga or amoeba. However in most cases, more complex eukaryotic organisms, from plants to animals and humans, require a complicated series of steps for successful virus propagation and spread. These steps include entry via an accessible portal of the organism, dissemination within the organism to reach susceptible cells, evasion of host defensive responses (including innate and adaptive immunity), and egress to allow for potential spread to new hosts. There is ample evidence that evolution acts within a single host, although for the sake of clarity we will refer to these intra-host events as “adaptation” rather than evolution. Using these terms allows us to highlight the distinction that local adaptation within a host is due to selective pressures that differ from those that impact transmission to new hosts, or that act across multiple generations of hosts. Also, the virus population within a complex organism may partition into distinct environmental niches within the host. For instance, the genomic diversity of human cytomegalovirus (HCMV) in patient samples is often analyzed from blood samples, and yet this viral population does not directly represent a common source of natural virus transmission between hosts (e.g., saliva). Studies of virus evolution need to carefully consider the source material used in examinations of viral diversity, and how this choice may influence the resulting observations of evolutionary fitness.

The Contributions of DNA Virus Persistence and Chronic Infections

We referred above to a theoretical model of DNA virus replication that involved productive replication in a single cell and spread into nearby uninfected cells, across multiple viral generations. An underlying assumption in such a model is that multiple rounds of productive infection occur sequentially. However, the lifecycle of many if not most DNA viruses exhibit other phases of existence, namely through persistence and chronic infections. For many bacteriophage and archaeal viruses, a common strategy is the well-known cycle of lysis versus lysogeny. For these viruses, the productive and often cell-destructive strategy of lytic replication is interleaved with phases of lysogeny, when the viral genome integrates into the host genome and is propagated as part of the host genome during cell division. A similar strategy exists for the large family of herpesviruses that infect most animal species and humans, with the long-term non-lytic phase being termed latency instead of lysogeny. An important distinction is that with a few notable exceptions, integration into the host genome is not a normal part of herpesvirus latency. Instead, these herpesviruses remain episomal in the host nucleus during lifelong latency. At the molecular level latency can be defined by the absence of significant viral replication and limited viral gene expression. Herpesvirus episomes can undergo sporadic reactivation to produce new viral progeny, which is followed by additional cycles of latency and reactivation. Similar to herpesviruses, certain members of the *Adenoviridae* can progress from a lytic infection of epithelial cells to a latent infection in T-lymphocytes of the tonsils and other adenoid tissues. The ability to establish a long-term infection is thus a vital part of the viral lifecycle of many DNA viruses, which contrasts with the acute infectious period of many RNA viruses (e.g., influenza virus or rotavirus). Persistence and chronic infections motivate the need to explore the contributions of within-host variation and adaptation to the evolutionary mechanisms of DNA viruses.

In addition to latency and lysogeny, virus persistence or chronic infection includes a whole class of DNA virus infections where viral replication is readily detected in the host, but the infection is not cleared for a significant length of time. Many smaller DNA

viruses such as papillomaviruses, polyomaviruses, and certain members of the *Circoviridae* use this “low-and-slow” approach. These viruses replicate in actively dividing cells, but have evolved to avoid detection by the host immune system. Interestingly, many of these viruses appear to be pathogenic only if the virus persists for an extraordinarily long time. For example, in most cases, the host immune system will eventually clear human papillomavirus infections. This process typically spans several months if not years. However, a long-term infection (> 2 years) dramatically increases the risk of virus-induced cancer. Similarly, while polyomavirus infections in humans are typically asymptomatic, long-term persistence of JC polyomavirus causes complications in immunocompromised hosts. In these hosts, the otherwise benign infection can spread into the nervous system, where the viral infection can then induce significant damage (as discussed further below). The duration of animal lifespans, as opposed to single-celled hosts, means that long-term persistent viruses of animal cells have evolved to have significantly more interactions with their host’s immune system during lifelong latency, than are observed during bacteriophage or archaeal virus lysogeny. Recent advances in high-throughput sequencing technology are now enabling researchers to interrogate whether mutations in viral genomes are specifically correlated with disease progression in these chronic infection settings.

Co-Divergence With Hosts as a Driver of DNA Diversification

A common perception is that RNA viruses mutate rapidly while DNA viruses are slow and stable. This may stem from the view that the diversity of many DNA viruses can be explained by co-divergence with host species, thus placing viral evolution on a timescale of millions of years. Long-term co-divergence and consequently low rates of nucleotide substitution have been supported in some DNA viruses; however, this is likely only part of the equation. The development of new sequencing technologies and the ability to include temporal information into molecular clock models allows us to estimate the rate and timescale of virus evolution independent of the (strong) assumption of co-divergence. Indeed, many DNA viruses show evolutionary rates close to those of RNA viruses, which themselves span a range of mutation rates. It is important to note that time-structured sequence data spanning years or decades often contain short-lived polymorphisms. Researchers should thus use caution when comparing mutation rates at such distinct evolutionary scales.

Nonetheless, for many viruses, it is essential to acknowledge that both short and long timescales may provide valuable information. While there is strong evidence supporting co-divergence of the *Polyomaviridae* with their hosts, recent studies have demonstrated the need to account for faster evolution within this virus family. In immunocompromised patients, mutations in the JC polyomavirus capsid protein allow it to escape neutralizing antibodies and invade the central nervous system, causing an opportunistic brain disease called progressive multifocal leukoencephalopathy (PML). The ability to evade the immune system – while remaining extraordinarily stable over longer timeframes – suggests that the *Polyomaviridae* evolve at two distinct rates. In the case of ssDNA parvoviruses, researchers seeking to understand the determinants of host range variation have tested the outcome of culturing several closely related viruses ($>98\%$ nucleotide identity) in cells derived from phylogenetically distinct hosts. The authors found that canine parvovirus (CPV-2) underwent extensive mutation during passage in non-native host cells, while no mutations arose in cells from the native host. These data indicate that the virus was well-adapted to its current host species, but that multiple mutations in its surface protein were needed for it to infect diverse host species efficiently. These data illustrate how long-term host dependency may constrain evolutionary rates in many DNA viruses.

Single Nucleotide Differences as a Measure of Evolutionary Change

Specific mutations such as single nucleotide polymorphisms (SNPs), insertions, and deletions (together termed in/dels) are likely to experience different selection dynamics, which impact the chances that these variations become fixed in the population. However, unlike for nucleotide substitutions (i.e., SNPs), the methods for measuring the evolutionary rate of insertions and deletions (in/dels) are not well developed. Because of this limitation, our understanding of viral evolution is primarily based on measuring the accumulation of SNPs over time, which ignores the potentially critical influence of other sources of variation, such as in/dels, tandem repeat fluctuations, and recombination (discussed further below). Recent studies have also provided evidence that viral evolutionary rate estimates decrease as their measurement timescales increase. This is evident in the field of paleovirology. For example, “fossilized” hepadnavirus DNA integrated into bird genomes suggests that these viruses are at least 19 million years old. In turn, this implies a significantly slower evolutionary rate than what was predicted based solely on extant viruses.

Early studies on the mutation rate of DNA viruses using single-gene or single-locus analyses estimated a mutation rate on the order of 1×10^{-7} to 1×10^{-8} substitutions/site/year. These values have been further supported by genome-wide comparisons for a handful of large DNA viruses. For instance, a recent study used a high-fidelity high-throughput sequencing (HTSeq) technique called duplex sequencing to detect spontaneous mutations in clonal lineages of human adenovirus 5, and the authors found that these occurred at a rate of 1.3×10^{-7} per base, per infection cycle. This rate matches well to a genome-wide estimate of the *in vitro* and *in vivo* mutation rates for murine CMV, which was obtained by shotgun Sanger approaches just before the development of HTSeq ($\sim 1 \times 10^{-7}$ mutations per bp per day). These low mutation rates are often cited by those wishing to contrast DNA virus stability with RNA virus diversity. However, data from both modeling and newer HTSeq-based comparative genomics studies have indicated that large DNA viruses may have mutation rates closer to 1×10^{-5} or 1×10^{-6} . In our comparisons of sub-clones generated from a parental population of herpes simplex virus 1 (HSV-1), we observed 3%–4% variation between sub-clones,

genome-wide. Other studies of HSV have shown that antiviral drug resistance mutations can be selected from a naïve virus population in just one round of viral passage *in vitro*. These data suggest that at least under certain circumstances, standing variation is maintained in DNA virus populations. An alternative or additional theory is that *de novo* mutations may occur at specific genomic regions more often than others (e.g., hot spots). The wider application of genome-wide measurements of viral variation will help to elucidate these possibilities.

***In Vivo* Observations of Within-Host Diversity and Adaptation of DNA Viruses**

Recent advances in high-throughput sequencing have now enabled the detection of minor variants within a single viral isolate or patient. These minor alleles can manifest as a new dominant allele or genotype after population bottlenecks or selective pressures such as antiviral therapy. Evidence of sequential takeover by distinct HCMV strains has been observed in immunocompromised adult patients, demonstrating both the existence of co-infections as well as the opportunities for recombination and/or subsequent selection. Studies of vaccine-associated rashes for varicella-zoster virus (VZV), and of congenital infections by HCMV, have demonstrated the potential for niche-specific adaptation or segregation of viral variants within specific body sites of infected hosts. For human papillomavirus 16 (HPV16), a recent study of several thousand women used a combination of PCR and Illumina-based HTSeq to reveal an unexpectedly high level of viral genetic variability. Of note, there was higher HPV16 genetic variability between patients than within a single patient, suggesting that many of the identified sequence differences were specific to each patient. Interestingly, women with pre-cancerous lesions had significantly less variation than women with a productive (early stage) HPV16 infection, confirming that cellular transformation by HPV represents a genetic bottleneck. This high level of inter-patient variability demonstrates that, at least within some settings, the mutation rate for HPV must be significantly higher than the previously estimated 2×10^{-8} nucleotide substitutions/site/year for the viral coding genome. Importantly, the higher-than-expected rate of inter-host evolution argues against the notion that a subset of (oncogenic) human papillomaviruses were acquired by archaic hominins during their migration out of Africa. Together these data indicate that many DNA virus populations may contain and/or generate standing variation following infection. It also appears that this variation is not often transmitted to a new host. The lack of successful transmission of these minor variants suggests that the standing variation in viral populations only becomes phenotypically apparent after population bottlenecks or selection. Importantly, these studies provide corroborating evidence that the molecular mechanisms of DNA virus evolution which have been demonstrated *in vitro*, also operate *in vivo*.

Fluctuations in Tandem Repeat Copy Number as a Mechanism of Evolution

Changes in the length or copy number of tandem repeats (TRs) provide another mechanism of virus evolution. Short TRs are usually categorized into three groups: homopolymers, which are sequential repeats of a single base (e.g., 5 or more C's in a row); microsatellites, which have a repeating unit of <10 base pairs (bp); and mini- or macrosatellites, which include repeating units of 10–500 bp. The mechanisms of repeat expansion or contraction vary by the repeat size. Homopolymer-based length variants are presumed to arise primarily through polymerase slippage, whereas larger TRs may arise either by template looping during polymerase progression or through recombination as discussed below. The repeating units of TRs may be perfect copies or include minor imperfections in the repeating sequence, and these repeats can occur in both coding and non-coding regions. In coding sequences, repeated elements may contribute to structural units of protein folding (e.g., turns of an alpha-helix) or provide variable lengths of unstructured regions within a multi-domain protein. Noncoding repeats have been shown to include promoter elements, chromatin or insulator binding motifs, as well as secondary structural elements such as quadruplexes and other motifs.

For many tandem repeats, the only viral data available is their conservation of position in the genome of a given species, and perhaps data on the degree to which a given TR varies in length across different virus isolates of the species. Functional roles have been demonstrated for select TRs in just a handful of DNA viruses. In the few herpesvirus species that have been shown to integrate into a host genome, there are viral telomeric repeats that function in their integration into the host. In other non-integrating herpesviruses, length variations at homopolymeric tracts in the thymidine kinase (TK) and polymerase genes are a common route of viral escape from the antiviral drug acyclovir. Ribosomal frameshifting of defective transcripts in these drug-resistant genomes allows the translation of a low level of functional TK or polymerase, enabling viral survival even in the face of an otherwise disabling mutation. Fluctuations in TR lengths have also been described for JC polyomavirus populations in patients. In this case a predominant polyomavirus genotype, or archetype, is shed in the urine of most infected individuals, while rearranged forms with deletions and TR variations are found in the brains of patients with PML disease. For poxviruses and other large DNA viruses, restriction fragment length polymorphisms (RFLPs) have often been used to track changes in the dominant virus genotypes and TRs over time. In a recent study of myxoma virus (a *Poxviridae* member), the predominant RFLP type was observed to change each year. Expansion of the inverted terminal repeat boundaries appears to provide myxoma virus with an opportunity for evolution. Likewise, the genome of the vaccinia poxvirus shows similar heterogeneity of the terminal repeats. Repeated plaque based purifications have shown that heterogeneity in the terminal repeats can evolve rapidly from the DNA of a single vaccinia virion. As technologies to track fluctuations in the length of TRs improve, it will no doubt become easier to examine these changes and gain a better understanding of their contribution to virus adaptation and evolution.

Large DNA Viruses Undergo Frequent Recombination

Recombination can serve as a driving force for evolutionary shifts in DNA viruses, akin to the genetic shifts that result from reassortment in segmented RNA viruses. Recombination can be classified as homologous recombination – between like sequences – or as illegitimate or non-homologous recombination. For most large DNA viruses, the potential of the viral genome to recombine has been studied by analyzing phylogenetic relationships between naturally circulating viral genomes. Among the adenoviruses, which include seven species (human adenovirus A-G) and multiple serotypes, recent studies applying HTSeq-based comparative genomics have demonstrated both intra-species and interspecies recombinants – often in association with pathogenic infections. For instance, a naturally circulating intratypic recombinant of human adenovirus subtype C was found to be the etiologic agent of severe acute respiratory infections in children in China. There are also examples of both historical and recent isolates of pathogenic adenoviruses that appear to have arisen from zoonotic transmission and recombination between simian and human adenoviruses. For the beta-herpesvirus HCMV, multiple studies have demonstrated a history of rampant recombination between the genomes of different isolates. Particular sections or islands of the HCMV genome appear to have co-segregated, while widespread recombination between strains has created a mixture of alleles elsewhere in the genome. It is thought that genes in these islands are co-dependent, thus placing a fitness cost on any recombination events that occur inside these regions. Similar levels of within-species recombination have been shown for most herpesviruses with sufficient genome sequence availability to make these comparisons. Recently, data supporting potential inter-species recombination among these viruses have been observed as well, with HSV-1-like DNA detected in several loci of the HSV-2 genome. Likewise, a virulent avian herpesvirus that created an outbreak in Australian poultry was revealed to be a spontaneous recombinant derived from two live-attenuated vaccines in use in the area. For large DNA viruses such as herpesviruses and poxviruses, laboratory co-infection studies and analysis of recombinant progeny by HTSeq have further defined the genome-wide potential for recombination and begun to define hot spots or regions with a higher propensity to recombine. Together these data demonstrate the extensive role of recombination in the evolution of both nuclear- and cytoplasmic-replicating large DNA virus genomes.

Recombination at Different Frequencies for Small DNA Virus Genomes

Large DNA viruses appear to recombine more readily than the small dsDNA viruses of the *Papillomaviridae* and *Polyomaviridae*. Even under controlled experimental conditions, no conclusive evidence for recombination within these two virus families has been described. One theory for this lack of observable recombination is that smaller viruses have fully optimized the usage of their genomic real-estate, such that recombination events would be highly likely to interrupt co-dependent genes or regulatory sequences – and thus carry too high a fitness cost to survive. However, phylogenetic analyses have identified evidence for several recombination events within the *Papillomaviridae*. As in HCMV, it appears that ancient recombination has segregated functional regions of the viral genome, separating the genes coding for non-structural proteins from the structural genes. Recombination does not appear to play a significant role in the short-term adaptation of the papillomaviruses, implying that recombined daughter viruses are not as fit as the parental genomes. Supporting this hypothesis, even when evidence of HPV16 recombination was detected within a single patient, these recombinant genomes were incapable of sustained replication within the host. Similarly, while phylogenetic analysis can detect evidence for ancient recombination near the root of the *Polyomaviridae* phylogenetic tree, recombination does not appear to be a significant component of ongoing polyomavirus evolution. However rare recombination events can and do contribute to virus evolution. For instance, conservation efforts to prevent the extinction of the western barred bandicoot have been hampered by an outbreak of the bandicoot papillomatosis carcinomatosis virus type 1 (BPCV1), a recombinant between an ancestral papillomavirus and polyomavirus. This virus is a hybrid that appears to have recombined the structural genes of the *Papillomaviridae* with the non-structural genes of the *Polyomaviridae*. These examples illustrate how rare and unusual recombination events can enable the dramatic expansion of viral evolutionary sequence space.

Despite being roughly the same size as the *Polyomaviridae*, single-stranded DNA viruses recombine relatively efficiently. Single-stranded parvoviruses have shown an ability to jump to new hosts rapidly, and recombination along with a relatively high mutation rate has been hypothesized to underlie this ability. Parvoviruses have also been demonstrated to readily recombine in cell culture. Although the mechanism of parvovirus recombination is not known, a role for viral secondary structure has been proposed. Indeed, the parvovirus origin of replication forms a hairpin structure that is a recombination hot spot, potentially due to stalling of DNA polymerase at this secondary structure. Template swapping before re-initialization of replication could then result in the formation of a chimeric genome. Alternatively, parvovirus replication may create intermediate concatemers. Resolving these concatemers may activate DNA repair enzymes, leading to the creation of mosaic viruses through the homologous recombination repair system. While recombination appears to play an essential role in the evolution of ssDNA viruses, these viruses appear to have adapted to minimize combinations of incompatible regulatory elements. For example, the gene encoding the replication protein (Rep) and the cis-acting elements that interact with the replication protein are usually within 100 nucleotides of one another. This ensures that the replication machinery is highly likely to remain together and compatible following any recombination events. A detailed comparison of recombination patterns within ssDNA viruses also found that breakpoints tend to fall outside of known genes. These observations imply that viruses expressing recombinant proteins are not usually tolerated.

Duplication and Deletions of Genes and Genome Segments

The outcome of recombination within identical or highly similar genomes is rarely noticed, except for occasions where this event leads to gene duplication or loss. Evidence of gene duplication and subsequent divergence is prevalent in adenovirus genomes. Ancient incidents of gene capture presumably produced those adenoviral gene products with similarity to host genes or those of other viruses, which are found across many adenoviral genera. Other more evolutionarily-recent duplications are found in smaller subsets of adenoviral species. The phenomenon of gene loss has been well-documented in herpesviruses, where across the diverse alpha-, beta-, and gamma-subfamilies of the *Herpesviridae*, many examples of gene loss have been found during viral propagation *in vitro*. The phenomena of genetic drift and gene loss were first detected in laboratory-passaged strains of the beta-herpesvirus HCMV, where the gene regions lost *in vitro* were later found to have functions associated with cell tropism and immune evasion *in vivo*. The extremely large mimivirus dsDNA genome has also been shown to undergo gene loss from both its termini during repeated passage in an amoebal host. In mimivirus, this gene loss was associated with a phenotypic change in virions, which was visible as a loss of fibrils on the virion surface.

In contrast to gene loss, the duplication of genetic segments – a gene accordion – has been best demonstrated by a series of elegant studies in poxviruses grown *in vitro*. These studies showed that expansion of gene copy number could provide functional fitness recovery after deletion of a core viral gene, by driving higher expression of a less-efficient gene version. This expansion also enabled the adaptation and eventual evolution of improved function, via mutations that occurred in the redundant copies of this gene. Whether or not this type of gene accordion occurs for DNA viruses that replicate in the nucleus remains to be determined. The segregated nature of nuclear replication and transcription, followed by translation in the cytoplasm, means that nuclear-replicating viruses will complement defects in co-replicating genomes *in trans*, since proteins made in the cytoplasm can be utilized by all progeny genomes.

Among the small DNA viruses, a subset of human papillomaviruses is associated with recurrent respiratory papillomatosis (RRP). Interestingly, these RRP-associated viruses are not typically considered as oncogenic viruses. However while RRP is considered a benign neoplasm of the larynx, involvement of the lungs is almost invariably fatal. Whole genome sequencing efforts have implicated a duplication of the viral promoter and a subset of viral genes in the RRP progression towards lung invasion. While the expansion of these loci in the papillomaviral genome is likely not important during a normal viral lifecycle, these data illustrate how duplications can provide a powerful adaptation mechanism for otherwise slow-evolving viruses.

Host-Virus Exchange via Horizontal Gene Transfer and Transposable Elements

Horizontal gene transfer (HGT) provides another avenue for evolutionary adaptation of both viruses and their hosts. HGT has been well-documented between bacterial and archaeal host species, often vectored by large DNA bacteriophages or archaeal viruses. Recent data have demonstrated that HGT may also take place between eukaryotic hosts and their viruses. For example, transposable elements (TEs) found in the moth genome have also been detected in the genomes of baculoviruses that infect these moths. Since this baculovirus infects several species of sympatric, co-occurring moths, it may well be the historical vector that moved TEs among these different host species. Other host-derived sequences were also detected in about 5% of progeny baculovirus genomes, although the co-opted host DNA was not carried beyond a few cycles of viral replication. Most of the integrated host sequences were TEs, but others appeared to result from recombination at sites of microhomology between the host and viral genomes. Most large DNA viruses are not known to integrate into the host genome as part of their overall replication strategy. Select herpesviruses of the alpha- and gamma- subfamilies do integrate into the host genome, although for these viruses it appears to be a reversible process that can lead to later excision and non-integrative replication. Marek's disease virus, an alpha-herpesvirus of poultry, and human herpesvirus (HHV) 6A and 6B, two gamma-herpesviruses of humans, integrate into host telomeres as a central part of their lifecycle. The germline or chromosomal integration of human herpesviruses (ciHHV), usually HHV6A, is detected in about 1% of the human population, although the clinical consequences of ciHHV are as yet unknown. These examples recommend the use of genome-wide HTSeq of viral populations as a means to detect horizontal gene transfer in action.

For the small DNA polyoma- and papillomaviruses, integration of all or a fragment of the viral genome into the host cell DNA is an evolutionary dead end, with an outcome that is nonetheless well-known for having the potential to induce dramatic outcomes of dysregulated cell division and tumor formation. In a recent study of HTSeq data from HPV-positive head and neck cancers, evidence was found to suggest that the HPV genome can replicate as an independent viral-human hybrid mini-chromosome, at least in some instances. These data implied that following an integration event, the viral genome may be excised from the human chromosome, creating a viral-human hybrid circular episome. Under particular circumstances, these hybrid genomes could theoretically get packaged into infectious virions. However, considering the tight regulation of papillomavirus replication, it appears unlikely that these hybrid genomes would be able to establish an infection in the next host.

Conclusions

Much remains to be resolved about the dichotomy between the measurably low rate of polymerase error in most DNA viruses, and their ability to undergo rapid genetic change in the face of intense selective pressures. However as discussed here, the multiple mechanisms of DNA virus evolution beyond single nucleotide substitutions likely provide the resources to confer this level of

evolutionary adaptability. Researchers have long agreed that ancient events of recombination and horizontal gene transfer, as well as gene duplications and subsequent divergence, could explain many aspects of virus origins. The breadth of new insights offered by high-throughput and deep viral sequencing, as well as by virus discovery and metagenomic approaches, have begun to broaden and clarify this picture. Deep sequencing has revealed the level and ubiquity of standing variation in virus populations, which provides fodder for future adaptation and selection. Metagenomic approaches and viral discovery have allowed researchers to detect novel viruses and recombinants that would have been missed using prior methods, which tracked viral presence using single-point genetic markers. These data provide ample assurance that the textbook explanation of mechanisms of virus evolution will need continued revision in the years to come, as more examples are brought forward and we expand our knowledge of how viral diversity arises and fuels virus evolution.

Acknowledgments

We appreciate the contributions of Molly Rathbun and other members of the Szpara and Van Doorslaer labs for their helpful input, as well as our many colleagues whose research and insights have been incorporated into this article. M.L.S. acknowledges support from the Eberly College of Science and the Huck Institutes of the Life Sciences at Pennsylvania State University, the Pennsylvania Department of Health Commonwealth Universal Research Enhancement (CURE) Program, as well as from NIH grants R01 AI132692, R21AI130676, and R21 AI140443. KVD is supported by a State of Arizona Improving Health TRIF, and by a USDA Hatch grant NC229.

Further Reading

- Allison, A.B., Kohler, D.J., Ortega, A., *et al.*, 2014. Host-specific parvovirus evolution in nature is recapitulated by *in vitro* adaptation to different carnivore species. *PLoS Pathogens* 10, e1004475.
- Buck, C.B., Van Doorslaer, K., Peretti, A., *et al.*, 2016. The ancient evolutionary history of polyomaviruses. *PLoS Pathogens* 12, e1005574.
- Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: Patterns and determinants. *Nature Reviews Genetics* 9, 267–276.
- Elde, N.C., Child, S.J., Eickbush, M.T., *et al.*, 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell* 150, 831–841.
- Firth, C., Kitchen, A., Shapiro, B., *et al.*, 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Molecular Biology and Evolution* 27, 2038–2051.
- Gilbert, C., Feschotte, C., 2018. Horizontal acquisition of transposable elements and viral sequences: Patterns and consequences. *Current Opinion in Genetics & Development* 49, 15–24.
- Greenbaum, B.D., Ghedin, E., 2015. Viral evolution: Beyond drift and shift. *Current Opinion in Microbiology* 26, 109–115.
- Houldcroft, C.J., Beale, M.A., Breuer, J., 2017. Clinical and biological insights from viral genome sequencing. *Nature Reviews Microbiology* 15, 183–192.
- Ismail, A.M., Cui, T., Dommaraju, K., *et al.*, 2018. Genomic analysis of a large set of currently—and historically—important human adenovirus pathogens. *Emerging Microbes & Infections* 7, 1–22.
- Jansen, A., Gemayel, R., Verstrepen, K.J., 2012. Unstable microsatellite repeats facilitate rapid evolution of coding and regulatory sequences. *Genome Dynamics* 7, 108–125.
- Koonin, E.V., Yutin, N., 2019. Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Advances in Virus Research* 103, 167–202.
- Mirabello, L., Yeager, M., Yu, K., *et al.*, 2017. HPV16 E7 genetic conservation is critical to carcinogenesis. *Cell* 170, 1164–1174.
- Renner, D.W., Szpara, M.L., 2018. The impacts of genome-wide analyses on our understanding of human herpesvirus diversity and evolution. *Journal of Virology* 92, e00908–17.
- Renzette, N., Pfeifer, S.P., Matuszewski, S., Kowalik, T.F., Jensen, J.D., 2017. On the analysis of intrahost and interhost viral populations: Human cytomegalovirus as a case study of pitfalls and expectations. *Journal of Virology* 91, e01976–16.
- Stedman, K., 2013. Mechanisms for RNA capture by ssDNA viruses: Grand theft RNA. *Journal of Molecular Evolution* 76, 359–364.
- Van Doorslaer, K., 2013. Evolution of the papillomaviridae. *Virology* 445, 11–20.
- Zhao, L., Rosario, K., Breitbart, M., Duffy, S., 2019. Eukaryotic circular rep-encoding single-stranded dna (CRESS DNA) viruses: Ubiquitous viruses with small genomes and a diverse host range. *Advances in Virus Research* 103, 71–133.

Introduction

Paleovirology can be defined as the study of ancient viruses and the way they have shaped the evolution of host genomes and phenotypes. Ancient viruses can be found in the form of viral genes or entire viral genomes integrated into the genome of their hosts. Some viruses, such as the retroviruses that infect vertebrate animals (family *Retroviridae*), encode all the machinery necessary to integrate their genome into that of their host. Prior to integration, the retrovirus' RNA genome is reverse-transcribed into a DNA molecule. Integration is an essential step of the replication cycle for retroviruses, which generally takes place in the somatic cells of the host. Sometimes, however, integration can occur into the genome of host germ cells. Gametes bearing such an integration may be involved in the formation of a zygote and if the integration does not impede normal development of this zygote, then all somatic and germline cells of the resulting host individual will bear the viral sequence integrated into their genome. The viral sequences, also called endogenous viral elements (EVE), can then be inherited vertically from one generation to the next and the frequency of this EVE can increase until it eventually reaches fixation in the host population, that is, it becomes present in all individuals of the population (Fig. 1). This phenomenon, often called "endogenisation", has occurred recurrently, seeding a large number of retrovirus-derived EVE in the genomes of vertebrates. For example, it has been estimated that no less than 8% of the human genome derives from retroviral EVE.

Until about ten years ago, the vast majority of EVE known in eukaryotes were retroviral sequences integrated into the genomes of vertebrate animals. Since then, many eukaryote genomes have been sequenced and it became possible to uncover a large number of new EVE originating from non-retroviral viruses integrated into the genomes of a variety of eukaryote species. Importantly, viruses are not found in the geological fossil record, such that before the discovery of EVE, it was very difficult to study viral evolution over long evolutionary periods of time. Paleovirological studies of EVE thus offer unique opportunities to trace the origin of currently circulating viral families, as well as to better understand how viruses have shaped the biology of their hosts over dozens of millions of years.

Patterns and Mechanisms Underlying Viral Integration

Endogenous viral elements are typically identified within the sequenced genome of their host using algorithms designed to detect similarities at the nucleotide or amino acid level with known viral genes. It is becoming standard to perform systematic characterization of EVE in a given set of genomes using all sequenced viral proteins as baits. There are 5,119,722 of such proteins available in the Genbank database of the NCBI as of January 2019. Searching for similarities between these proteins and eukaryote genomes usually returns a very large number of host sequences resembling viral genes, many of which are false positives that can be filtered out by implementing several filters. It is then important to ensure that the remaining candidate EVE sequences are indeed integrated into the host genome instead of deriving from circulating viruses, the genome of which could have been extracted and sequenced together with that of the host. Among other ways to verify integration, finding a DNA version of an RNA virus can be taken as a good indication of the integrated nature of this virus. Depending on the quality of the genome assembly under study, it may be useful to further verify the integrated nature of EVE by amplifying loci of interest using PCR and to resequence them using targeted sequencing.

Systematic screenings for EVE performed during the last 10 years have revealed that virtually all types of RNA and DNA viruses can integrate into the genome of their eukaryotic hosts. Endogenous viral elements have been identified in a large variety of both uni- and multi-cellular eukaryotes. These discoveries are remarkable in the sense that the majority of viruses are not able to actively integrate into the genome of their host. It is thus believed that most endogenisation events are the product of fortuitous, illegitimate recombination events between host and viral genomes that are catalyzed by host enzymes. Several analyzes of the genomic context in which EVE lie within a given genome have revealed that they are often directly flanked by transposable elements (TE). Transposable elements are sequences that have the ability to move from one locus to another in a given genome and to duplicate themselves in the process. They are found in the genomes of all organisms, often reaching high numbers of copies. The presence of such TE around EVE coupled with detailed analyzes of the molecular signatures associated with EVE integration suggest that integration of non-retroviral RNA viruses often occurs through reverse-transcription of viral RNA genomes or transcripts, which is mediated by reverse-transcriptases encoded by retroelements residing in the host genome (Fig. 1). Furthermore, some parvoviruses can integrate into the host genome by a Rep-mediated mechanism and it has been suggested that

[☆]This Chapter is partly adapted from Gilbert, C., Feschotte, C., 2016. Endogenous viral elements: Evolution and impact. *Virologie*. 20, 158-173.

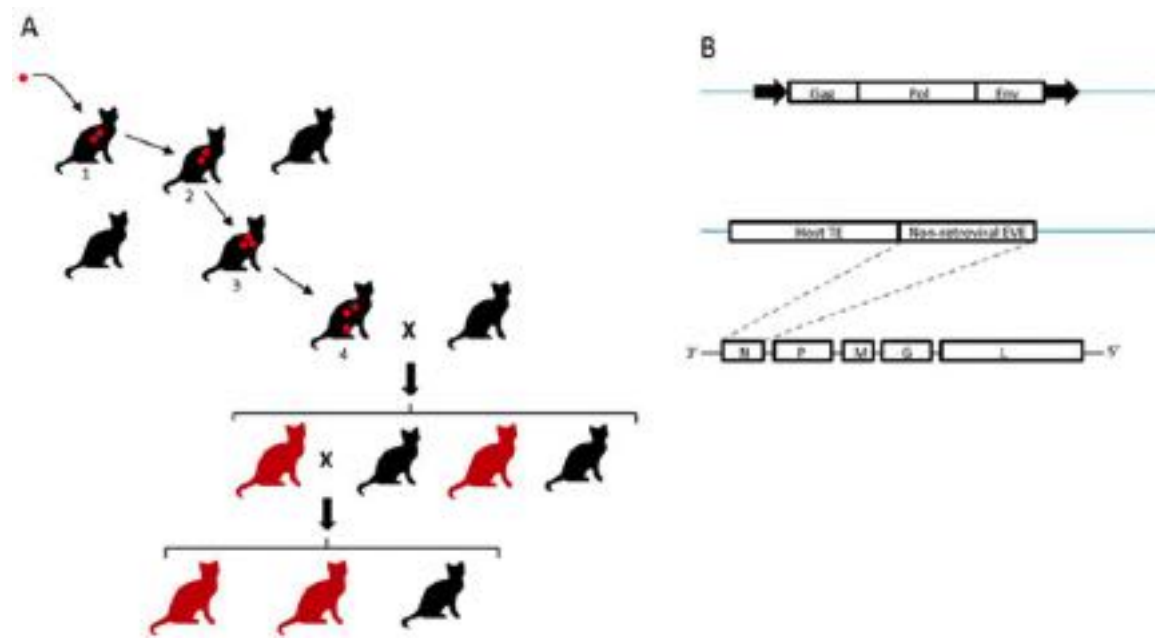


Fig. 1 Viral endogenisation. (A) A virus (red dots) circulates horizontally within a population of cats and replicates in somatic cells. In one of the infected cats (#4), the virus is able to penetrate into germ cells and a fragment of its genome is integrated into that of the cat's gametes. The integration may have been facilitated by the reverse-transcription of the viral RNA catalyzed by the reverse-transcriptase encoded by the cat's retrotransposons. The cat bearing the viral integration reproduces with another cat devoid of integration. This crossing produces cats which bear the integrated virus in all their somatic and germline cells. If the endogenous virus does not have any deleterious effect on the cat bearing it, then its frequency can eventually increase in the cat population until it reaches fixation. (B) Top panel: structure of an endogenous retrovirus. The two black arrows correspond to the long terminal repeats (LTR) flanking each new copy of endogenous retrovirus. Gag: Group-specific antigen, processed into matrix, capsid, nucleocapsid and, in some viruses, additional core proteins. Pol: Polyprotein encoding reverse-transcriptase, RNase H and integrase functions. Env: envelope protein. Lower panel: structure of a non-retroviral endogenous viral element. This example illustrates the integration of the N gene of the genome of a mononegavirus (-ssRNA virus). The integrated N gene lies next to a host transposable element (TE), indicating that it was likely accidentally reverse-transcribed into the host genome together with this TE via template switching of the TE reverse-transcriptase.

DNA viruses in general can integrate into host chromosomes when DNA double strand breaks are repaired, through processes such as non-homologous end-joining.

Endogenous viral elements often correspond only to fragments of viral genomes. While the fragmentary nature of EVE may generally be best explained by the fact that endogenisation involved only a fragment of a viral genome, it is also possible that portions of an EVE corresponding to an entire viral genome were lost through mutations and deletions. In most cases, such degradation may affect randomly any region of an integrated viral genome. In rare instances, however, some viral genes may be recycled, or domesticated, by the host. Given the new beneficial cellular function fulfilled by such domesticated viral genes, they are less likely to be degraded than non-domesticated viral genes. The fragmentary nature of EVE may also be explained by the likelihood with which a given transcript or viral genome fragment may be reverse-transcribed, which may depend on its abundance in the host cell. Some segmented viruses such as partitiviruses, for example, typically package different amount of each genome segment in their capsid. Other non-segmented viruses such as those belonging to the *Mononegavirales* order produce six transcripts in different quantities, none of which correspond to the entire viral genome. The quantity of each transcript is determined by a 5'-to-3' gradient, with the N gene encoding the nucleocapsid located at the 5' extremity of the genome being the most transcribed. In agreement with this, the majority of EVE deriving from endogenisation of *Mononegavirales* in animal genomes correspond to the N gene (Fig. 1(B)).

Another reason which may have more to do with methodological limitations linked to the detection of EVE, might sometimes explain why we observe some viral genes more often than others under the form of EVE. For example, a study performing a systematic search of EVE in the genomes of six crustacean species revealed that the vast majority of the 210 annotated EVE corresponded to viral genes encoding non-structural proteins. It is unlikely that the observed higher abundance of non-structural proteins be explained solely by a mechanistic process because the bias was observed for all six viral groups to which the 210 crustacean EVE belong (both RNA and DNA viruses), which differ greatly in their replication cycle. Rather, the observed gene bias in endogenisation may be caused by differences in the mode and rate of evolution among viral genes. Indeed, it is known that viral genes encoding structural proteins tend to evolve faster and be less conserved than genes encoding non-structural proteins because they are more directly involved in evolutionary arms races with their host lineages. Endogenised viral genes encoding structural proteins may thus be often too distantly related to viral proteins used as baits to be detected in

similarity searches. It follows that the number of EVE that we are able to recognize in the genome of eukaryotes likely is an underestimation of the true number of EVE that have entered these genomes, and that we will be able to discover more EVE as we discover more extant viruses.

Calibrating the Long-Term Evolutionary History of Extant Viral Families

Perhaps one of the most interesting features of endogenisation is that once a virus gene enters the genome of its eukaryotic host, it basically freezes in time. In other words, it replicates only at the rate of the host genome and ceases to mutate very rapidly as it did when it was replicated by a highly error-prone viral polymerase. Given that eukaryote mutation rates are two to five orders of magnitude slower than most viral mutation rates, it becomes possible to retrieve very old EVE. Studies of the short-term evolution of viruses, including estimates of the time at which extant viruses have entered the human population, have suggested that most viruses that infect humans today may be at most one million years (my) old. Human RNA viruses would even be much younger than that and be characterized by extremely fast mutation rates. On the other hand, several authors have tried to trace back the origin of all viruses by comparing the secondary structure of bacterial and eukaryotic virus proteins. Based on these comparisons, a consensus has emerged suggesting that viruses are likely very old, probably as old or even older than cellular organisms, that is, more than 1 billion years old. According to this view, viruses would have even played a major role in the origin and evolution of cellular organisms. Between the very recent time scale of extant virus evolution and the very ancient origin of viruses, our understanding of the mode and tempo of virus evolution, including the antiquity of the families to which extant viruses belong, have long remained mysterious. The discovery of EVE is remarkable in this respect in that many EVE can be dated and used as molecular fossils to calibrate the origin and tempo of evolution of their cognate viral families.

Several dating approaches can be used to infer the age of EVE (Fig. 2). The first one is specific to retroviruses, which upon integration into the host genome generate identical repeated sequences on both sides of their genome, called long terminal repeats (LTR). Since mutations accumulate in each LTR at the mutation rate of the host, the age of a given retroviral EVE can be calculated by counting mutations that differentiate the two LTR of this EVE and dividing this number by two and by the mutation rate of the host. The second approach can be applied to all types of viruses. It relies on assessing the taxonomic distribution of a given EVE within the host lineage. The chances for a virus to integrate exactly the same portion of its genome exactly at the same genomic locus independently in two host lineages are extremely low. Therefore, finding a given EVE shared at the same genomic locus by two host species strongly suggests that endogenisation of this EVE took place in an ancestor of the two species (Fig. 2). If the time at which the two species diverged has been calculated, based on geological fossils for example, then one can infer that the age of the EVE is at least as old as this divergence time. Finally, one can also assign a relative age to a given EVE when this EVE has been hit by a TE. Since landing of the TE inside the EVE can only have taken place after the EVE entered the host genome, it is possible to infer that the EVE is older than the TE. If the time at which the TE has landed into the EVE can be calculated, based on comparison with other TE copies found in the host genome for example, then one can assign a minimal absolute age to the EVE.

The dating of EVE related to extant viral families can yield new, unique information that can be used in various ways (Fig. 3). For example, when a dated EVE falls within the diversity of an extant viral family, then the age of the entire family can be inferred to be at least as old as that of the EVE. This approach has led to drastic shifts in our perception of the time scale over which current viral lineages have evolved, showing that many viral families are much older than previously thought. For example, EVE belonging to the Circoviridae family have been found at the same genomic locus in the dog and the cat, which have diverged more than 50 my ago. This finding directly indicates that this family of small single-stranded DNA viruses is more than 50 my old, which sharply contrasts with the age of about 500 years which had previously been inferred for the origin of circoviruses currently circulating in pigs and birds. Similarly, while it was thought that Hepadnaviridae were at most 30 thousand years old, the discovery of endogenous hepadnaviruses shared between turtle lineages that diverged more than 200 my ago suggests that these viruses have co-evolved with amniotes since the beginning of the Mesozoic. The age of an EVE can also be used to calibrate the phylogeny of a viral lineage in order to estimate the age of the other nodes of this phylogeny. For example, the age of endogenous viruses related to nudiviruses found in the genomes of some parasitoid wasps has been used to infer divergence times among insect large double-stranded DNA viruses. The approach yielded an age of about 310 my for the origin of these viruses, suggesting that the first insects, which appeared at the end of the Devonian and beginning of the Carboniferous, were already infected by this type of viruses.

It is also possible to use EVE to study the evolutionary dynamics of viral lineages over timescales much larger than when including only circulating viruses in the analysis. For example, long-term viral substitution rates can be calculated on the branches separating an EVE from its closest circulating virus by counting the number of mutations that occurred on these branches, removing those that occurred on the EVE branch after endogenisation and dividing the remaining number of mutations by the time separating the EVE from circulating viruses (i.e., the age of the EVE). Such an approach has shown that viral substitution rates calculated over several dozens of my were several orders of magnitude slower than those calculated based only on currently circulating viruses. The discrepancy between substitution rates calculated over short and long timescales is particularly strong for viruses but it is also observed in cellular organisms. A large part of it can be explained by methodological limitations linked to the calculation of rates, both on short and long timescales. Importantly, it shows that rates calculated over different timescales depict different biological processes and that a rate calculated over a given timescale cannot be used to extrapolate the evolutionary dynamics of a viral lineage to a different timescale.

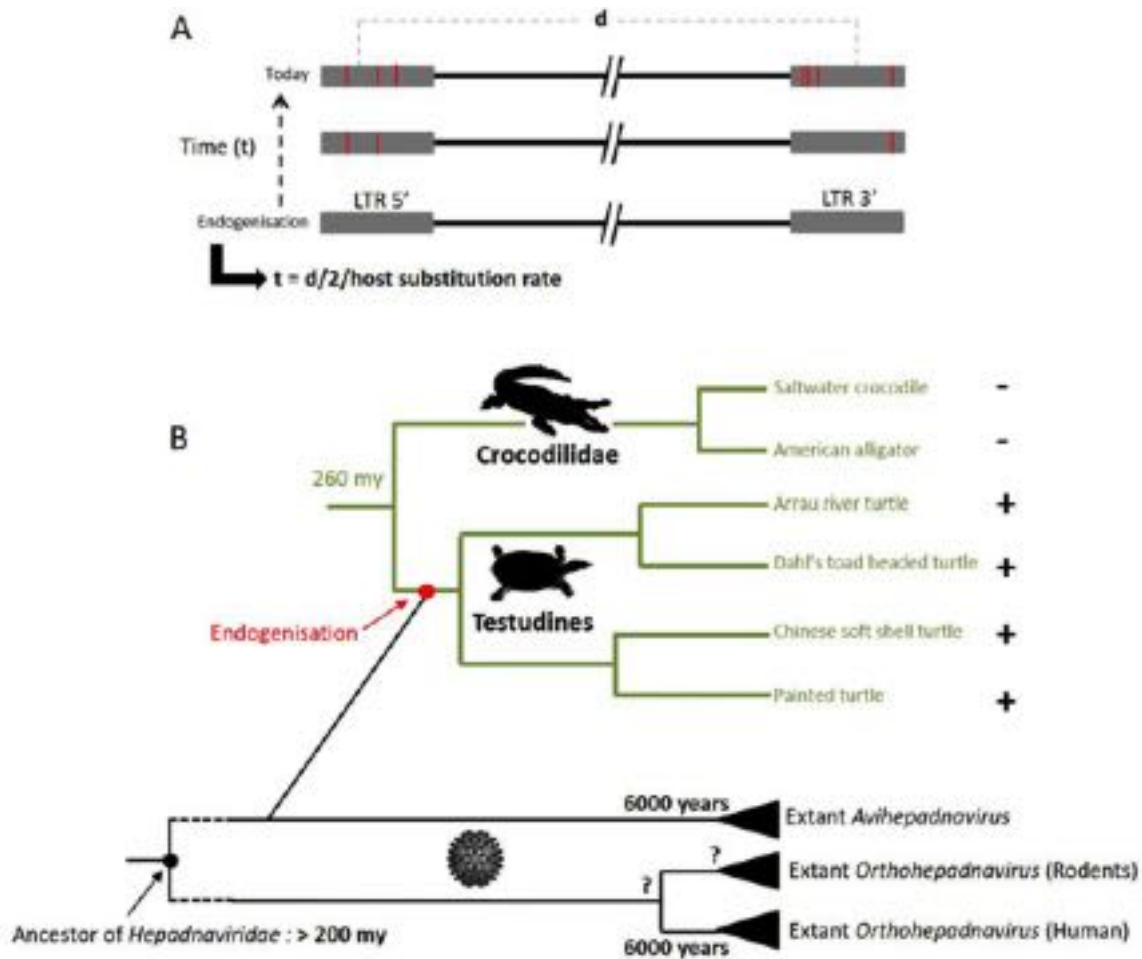


Fig. 2 Dating endogenous viruses. (A) Endogenous retroviruses can be dated by dividing the genetic distance (d) calculated between their two long terminal repeats (LTR) by 2 and by the host substitution rate. This is because when a retrovirus integrates, its two LTR are identical. The number of mutations differentiating the two LTR at any given time is proportional to the time since endogenisation. (B) All endogenous viruses (EVE) can be dated by searching for orthologous EVE copies in species other than the one in which they have been first identified. In this example, an endogenous hepadnavirus has been identified at the same orthologous locus in four turtles (+) and this EVE is absent from two crocodile species. Divergence times between turtles are known so it is possible to infer that the EVE is at least as old as the ancestor of the four turtles (200 my old). This age contrasts sharply with the age inferred for viruses currently circulating in mammals and birds.

Evolutionary Impact of Endogenous Retroviruses

The presence of many EVE in the genome of eukaryotes begs the question of the fate of these sequences, as well as of the impact they may have on the evolution of their host. Most of what we know about these questions can be drawn from the studies of vertebrate retroviruses, which are both the most numerous and most studied EVE. After endogenisation, some copies of endogenous retroviruses can be transcribed and generate new copies in the host genome. The increase in copy number can occur either through retrotransposition in germline cells, a process that does not involve the formation of infectious viral particles, or through reinfection of germline cells by infectious viral particles produced by endogenised copies. About half of the endogenous retrovirus copies found today in the genome of vertebrates are old, degraded and unable to produce proteins involved in viral replication. Yet, several recent integrations of retroviruses able to produce new copies have been described in various species. For example, the genome of the mouse contains several thousands of an endogenous retrovirus called IAP (Intracisternal A Particle), among which hundreds are still able to duplicate via retrotransposition. In koalas, a circulating retrovirus called KoRV is currently invading the genome of some populations located in northern Australia, while southern populations are completely devoid of the virus. Copies of KoRV have been dated at less than 50,000 years, but it is unknown whether they are transcribed and able to generate infectious particles. In human, a study published in 2015 has shown that copies of the most recently endogenised retrovirus (HERVK) constitutively produce viral particles during early embryonic stages. The function of these particles, if any, remains to be elucidated. While it is possible that they may simply be relics of benign infections without any detrimental effect, they may also be involved in the priming of the immune system which is essential during early embryonic development.

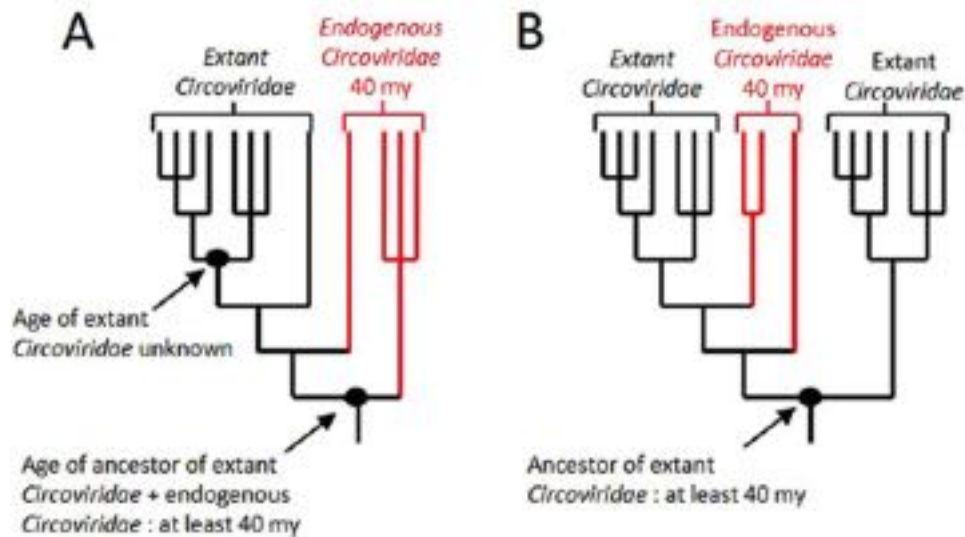


Fig. 3 Phylogenetic position of endogenous viruses and age of cognate viral families. (A) In this tree, endogenous 40 my old circoviruses fall outside of the clade containing all modern members of the *Circoviridae*. Thus the age of endogenous circoviruses does not inform us on the age of extant ones. In (B), endogenous circoviruses fall within the diversity of extant *Circoviridae* such that it is possible to infer that the ancestor of all known circoviruses is at least as old as the age of endogenous circoviruses, that is, 40 my old.

The presence of multiple copies of endogenous retroviruses in the genomes of vertebrates can have multiple negative consequences, whether these copies are still active or not. New copies generated by active EVEs can land within genes, affecting transcription or gene splicing and leading to the formation of defective mRNA. In addition, each endogenous retrovirus copy contains numerous regulatory elements necessary to recruit all the components of the transcription machinery. Integration of a new copy of endogenous retrovirus at a given locus can thus severely disturb the chromatin structure and the regulation of transcription in the vicinity of this locus. Host genes located within the affected region may be negatively impacted. In the mouse, it has been estimated that no less than 10% of all mutations inducing visible phenotypic consequences are due to new integrations of endogenous retroviruses that modify the regulation of neighboring genes. Another type of negative consequences endogenous retroviruses may have on their host genome is through mediating events of non-allelic homologous recombination between two identical copies located at different loci in the genome. Several human diseases are known to be the consequence of such rearrangements. For example, a deletion of 729 kb produced by homologous recombination between two copies of endogenous retroviruses located on the Y chromosome is associated to some forms of male sterility.

The impact of endogenous retroviruses on their host can also be positive. A large fraction of the genetic material added to vertebrate genomes by retrovirus endogenisation has indeed been recycled, or domesticated, and is now involved in multiple cellular functions. Numerous examples of human or mouse genes have been characterized for which transcription is initiated by a promoter derived from the LTR of an endogenous retrovirus copy. A number of studies have even shown that regulatory elements found in the LTR of some endogenous retroviruses have been recurrently domesticated and are now pivotal for the coordination of some gene regulatory networks, such as those involved in immune responses. Ironically, these EVE-derived regulatory sequences, which used to be essential for the virus to fulfill its replication cycle at the detriment of the host, are now involved in the adaptation of the host antiviral immune response.

The contribution of endogenous retroviruses to vertebrate genes at the protein level is less extensive, in terms of number of genes, than the contribution of EVE-derived regulatory elements. Yet, some domesticated viral proteins are now fulfilling major cellular functions that have probably been crucial to the evolutionary success of mammals. The first of these functions is the formation of a specific membrane of the placenta, called syncytiotrophoblast, which facilitates exchanges of nutrients, gases and other material between the mother and the embryo. Some of the key genes involved in the formation of this membrane have been acquired by mammals through domestication of the envelope gene of several endogenous retroviruses. Interestingly, domestication did not happen only once in the ancestor of all mammals, but repeatedly and independently on at least seven occasions during the evolution of mammals. Each of these domestication events, which took place in primates, rodents, rabbits, carnivores, ruminants, tenrecs and marsupials, involved a different copy of endogenous retrovirus, originating from distinct endogenisation events. The second function deriving from domestication of an endogenous retrovirus protein is linked to antiviral immunity. Several genes deriving from the envelope of specific copies of endogenous retroviruses have been characterized in cat, mouse and sheep, that provide immunity against currently circulating retroviruses. Circulating and endogenous retroviruses interact both with the same cellular receptors via the surface domain of their envelope protein. The regulated expression of domesticated endogenous envelope proteins interferes with the interaction between circulating infectious viruses and cell receptors, preventing viral entry into host cells.

Evolutionary Impact of Non-Retroviral EVE

Non-retroviral EVE are less numerous and have been less studied than retroviral EVE. While their global impact on eukaryotic genomes does not seem to be systematically as important as that of retroviral EVE, an increasing number of examples tend to show that non-retroviral EVE are not simply neutral passengers of eukaryotic genomes. In plants, for example, the recurrent endogenisation of florendoviruses, which are related to members of the *Caulimoviridae* family, produced a large number of copies in the genome of several species such as grapevine, rice, banana, *Petunia* or tomato, reaching 1.3% of the total genome size in the castor bean (*Ricinus communis*). It has also been shown that the expression of some copies of these endogenous florendoviruses could be induced by several types of stresses, including wound, heat or hybridization, ultimately leading to the production of infectious viral particles. In human, we know that the HHV6 herpesvirus is able to integrate its genome into the telomeric regions of several chromosomes and that it can be vertically transmitted from parents to their children. It has been estimated that about 0.5% of the human population bears at least one copy of endogenous HHV6 that can be activated and form infectious viral particles under various types of stressful circumstances. Some evidence even exists supporting a link between the presence of endogenous HHV6 and the development of angina pectoris.

Much like what has been observed for endogenous retroviruses, some non-retroviral EVE have been domesticated by their host and are now fulfilling crucial cellular functions. One of the most spectacular example of domesticated non-retroviral EVE is found in parasitoid wasps. Most parasitoid wasps (Hymenoptera, Ichneumonoidea) are koinobionts, meaning that their larval development takes place inside the arthropod hosts they parasitize, which remain alive until the wasps enter the nymphosis stage. Several wasp families have domesticated EVE-derived genes, which allow them to produce virus-like particles that they inject together with their eggs into host larvae. These particles contain virulence factors (circular wasp DNAs and/or wasp proteins) which decrease the intensity of the host immune response and facilitate the development of wasp larvae. The EVE from which wasp domesticated viral particles derive result from multiple events of endogenisation of large double stranded DNA viruses that took place at different time points and in several branches of the wasp phylogeny. It is likely that these endogenisation events, one of which has been dated at 100 my ago, have been pivotal for the evolutionary success of wasps.

A beneficial role of non-retroviral EVE linked to antiviral immunity is also suspected in an increasing number of cases. The way EVE-derived antiviral immunity is thought to occur is via the production of EVE-derived small RNA that trigger RNAi mediated recognition and destruction of cognate circulating viral RNA genomes and/or transcripts. In the tiger mosquito for example, no less than 472 EVE have been identified, deriving from 8 viral families. Mapping of small RNA libraries on the mosquito genome revealed that many EVE produce piRNA, a type of small RNA that, among other roles, are involved in controlling the activity of TE. While a definitive demonstration of the antiviral role of EVE-derived piRNA is still lacking, such piRNA have been identified in a large number of other species such as plants, many other insects, ticks and rodents. Based on these and other findings, a general model of EVE-derived antiviral immunity has been proposed which posits that repeated viral assault of a given host lineage might be accompanied by recurrent birth and death cycles of domesticated immunity-conferring EVE. This model predicts that when the antiviral activity of a given EVE becomes obsolete, either because the host is no longer exposed to the virus, or the host has become tolerant, or the virus has gone extinct, the EVE may either slowly decay by accumulating mutations or it may be co-opted to perform a new cellular function.

Ecology of Host-Virus Interactions

Our understanding of host-virus interactions has long been biased by the methodological limitations preventing us from discovering non-pathogenic viruses. Today, we know that viruses are the most numerous organisms on the planet, that they make up the largest and most diversified gene reservoir, and that they play key roles in ecosystem functioning. For example, viruses contribute substantially to carbon cycling through their decay and the pressure they exert on marine microbial communities. It is becoming clear that viruses are ubiquitous both in the environment and in cellular organisms and that they can no longer be only considered as pathogenic organisms having deleterious consequences on their hosts. Indeed, some plant and insect viruses are regularly discovered that are beneficial to their host, either by boosting their immune system, by allowing them to produce higher numbers of offspring, or even allowing them to colonize new ecological niches.

A large fraction of our new perception of the place viruses occupy in ecosystems derives from the development of viral metagenomics, a technique allowing to sequence the genomes of all viruses present at a given time in a given environment without having to isolate and cultivate them. Much like viral metagenomics, paleovirology offers a new way to discover and study viruses without a priori suspicion of their presence based on pathological symptoms. Just by discovering an EVE in a given host, one can deduce that the circulating virus from which the EVE derives was infecting this host, that the cellular tropism of this virus included germline cells and that the virulence of this virus was low enough for the host to survive, ensuring vertical transmission of the EVE via reproduction. Many EVE have been discovered in the genomes of species that are not known to be infected by cognate circulating viruses today. It is true for hepadnaviruses, for example, which were found endogenised in the genomes of passerine birds, snakes, turtles and crocodiles. Endogenous lentiviruses have been found in the genome of various mammals (e.g., rabbits and hares, malagasy lemurs, dermopterans) in which related circulating viruses have never been found. Whether these species are no longer infected by these viruses or whether they are still infected but we have so far not been able to detect such circulating viruses is an open question. Given the role some EVE play in thwarting cognate circulating viruses, one could expect a link between

the presence of EVE in a given species and the absence of cognate circulating viruses, or at least the absence of pathology induced by such viruses, rendering them difficult to detect.

The combination of approaches such as viral metagenomics and paleovirology holds great promise in furthering our understanding of host-virus interactions. Sometimes, however, paleovirology by itself offers unique opportunities to shed light into long-held conundrums. For example, the discovery of EVE in the genome of an algae species (*Bigelowiella natans*; Rhizaria) helped deciphering the intriguing relationships between giant viruses, viruses infecting these giant viruses (called virophages), and unicellular eukaryotes infected by giant viruses. The Sputnik virus, discovered in 2008, was the first to be classified in the new category of viruses parasitizing other viruses called virophages because of its inhibitory effect on the replication of giant viruses belonging to the putative “Megavirales” order. Since then, other virophages have been characterized and it has been proposed that eukaryotic TE belonging to the Maverick/Polinton group derive from ancient virophages that have acquired the capacity to integrate into and transpose within eukaryotic genomes. The discovery of intact virophage genomes in the genome of *B. natans* (called provirophages) was thus remarkable in this context, as it provided strong support for the evolutionary link between these viruses and Maverick/Polinton TE. Interestingly, the study performed in *B. natans* also characterized genes derived from giant viruses endogenised in the *B. natans* genome and showed that many provirophage genes were transcribed. The results thus confirmed that virophages and giant viruses can co-exist in *B. natans* cells and they suggest that endogenisation of virophages could be beneficial not only to the virophage itself but also to its host. Being integrated and transcribed in the genome of its host, the virophage increases its chances to encounter the giant viruses it replicates on, and by parasitizing these viruses, the virophage offers protection against giant viruses to the algae. More recently, another elegant study combining paleovirology and experimental assays verified the predictions formulated in the *B. natans* study, which were based only on a paleovirological approach. This second study showed that the Mavirus virophage is able to integrate at multiple sites into the genome of its host, the marine protozoan *Cafeteria roenbergensis*. The genes encoded by integrated virophages are specifically activated when the host is infected by a giant virus called CroV (*Cafeteria roenbergensis* virus). Remarkably, it was demonstrated that the virophage particles released by infected cells are able to suppress CroV replication in other cells, thus offering an antiviral protection to the protozoan.

Other Contributions of Paleovirology to Our Understanding of Host-Virus Interactions

Paleovirology has contributed to bolster our understanding of long-term host-virus interactions in a number of ways other than those already exposed in the preceding sections. For example, several research groups have successfully attempted to reconstruct portions or even fully functional and infectious versions of ancient retroviruses using a collection of endogenous retroviral sequences. The resurrection of ancient primate retroviruses such as HERK found in human and CERV1 and CERV2 found in chimpanzee showed that these viruses were unlikely targeted by TRIM5 α , a primate restriction factor known today to provide immunity against various modern retroviruses. On the contrary, the replication of these ancient viruses was shown to be actively repressed by mRNA editing through APOBEC3G, another restriction factor which has been shown to evolve under constant purifying selection during primate evolution. It has even been possible to reconstruct the crystal structures of the capsid proteins of two prehistoric lentiviruses identified in rabbits and Malagasy primates that are respectively about 12 and 4 my old. These studies revealed a remarkable structural conservation between ancient and modern lentiviruses despite a very low level of primary sequence identity. They also showed that interaction between the capsids of modern lentiviruses and the host factor cyclophilin A, known to be essential for extant lentiviruses to be infectious, already existed at the time at which the two reconstructed lentiviruses were infecting mammals.

A lot can be learnt on the long-term evolutionary arms races that have been taking place between viruses and their hosts not only by studying EVE but also through the inference of the forces that have governed the evolution of host restriction factors. Viruses and their hosts are often caught in a cycle of adaptation/counter-defense whereby the virus constantly finds ways to thwart host immune strategies. Such dynamics translates in a high rate of amino acid change in some regions of the viral genome, driven by positive selection. After reconstructing the phylogenetic tree of a given restriction factor, it is possible to assess which part of the evolutionary history of this factor has been characterized by intense battles with some viruses by mapping inferred ratios of synonymous versus non-synonymous substitution rates on each branch of the tree. Such studies have not only revealed that extant primate restriction factors were shaped by long-lasting conflicts with viruses, they also helped to pinpoint which protein domains of these factors were at the very interface of these conflicts.

Finally, the study of EVE offers unique opportunities to understand how the structure of modern viral genomes has evolved. In particular, EVE may allow us to decipher what mechanisms underlie the birth of new viral accessory genes, which has often remained mysterious. For example, the finding that an accessory gene lying between the pol and env genes of an ancient primate lentivirus was similar to the 3' end of the reverse-transcriptase (RT) domain of this virus suggested that the accessory gene might have arisen through duplication of a fraction of the RT gene. While none of the known accessory genes of modern retroviruses are similar to any other protein domain encoded by their respective viral genome, it is possible that at least some of them first emerged through such duplication and later lost any trace of their origin because of the numerous mutations they accumulated. The characterization of ancient hepadnaviruses also allowed to untangle the origin of the X gene found in mammalian orthohepadnaviruses and known to be involved in the formation of tumors in chronic infection by these viruses. An X-like gene has been characterized in some bird avihepadnaviruses suggesting that the birth of the X gene may predate the split between bird and mammal hepadnaviruses. Yet, the absence of this gene in all endogenous bird hepadnaviruses, coupled to the different genomic

location of the X-like gene and X gene have challenged this view. Much like what has been inferred for the origin of lentiviral accessory genes, a fine dissection of the structure of endogenous and exogenous hepadnaviruses led to the conclusion that the X gene evolved in the ancestor of orthohepadnaviruses, after the split between mammalian and bird hepadnaviruses, likely through segmental duplication of the viral genome.

Further Reading

- Aswad, A., Katourakis, A., 2012. Paleovirology and virally derived immunity. *Trends in Ecology and Evolution* 27, 627–636.
- Babaian, A., Mager, D.L., 2016. Endogenous retroviral promoter exaptation in human cancer. *Mobile DNA* 7, 24.
- Blanc, G., Gallot-Lavallée, L., Maumus, F., 2015. Provirophages in the *Bigeloviella* genome bear testimony to past encounters with giant viruses. *Proceedings of the National Academy of Science of the United States of America* 112, E5318–E5326.
- Chuong, E.B., Elde, N.C., Feschotte, C., 2017. Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics* 18, 71–86.
- Dewannieux, M., Heidmann, T., 2013. Endogenous retroviruses: Acquisition, amplification and taming of genome invaders. *Current Opinion in Virology* 3, 646–656.
- Emerman, M., Malik, H.S., 2010. Paleovirology—Modern consequences of ancient viruses. *PLOS Biology* 8, e1000301.
- Feschotte, C., Gilbert, C., 2012. Endogenous viruses: Insights into viral evolution and impact on host biology. *Nature Reviews Genetics* 13, 283–296.
- Fischer, M.G., Hackl, T., 2016. Host genome integration and giant virus-induced reactivation of the virophage mavirus. *Nature* 540, 288–291.
- Gilbert, C., Feschotte, C., 2016. Endogenous viral elements: Evolution and impact. *Virologie* 20, 158–173.
- Herniou, E.A., Huguet, E., Theze, J., *et al.*, 2013. When parasitic wasps hijacked viruses: Genomic and functional evolution of polydnaviruses. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20130051.
- Katourakis, A., Gifford, R.J., 2010. Endogenous viral elements in animal genomes. *PLOS Genetics* 6, e1001191.
- Koonin, E.V., Dolja, V.V., Krupovic, M., 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 479–480, 2–25.
- Lavialle, C., Cornelis, G., Dupressoir, A., *et al.*, 2013. Paleovirology of “syncytins”, retroviral env genes exapted for a role in placentation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20120507.
- Pybus, O.G., Rambaut, A., 2009. Evolutionary analysis of the dynamics of viral infectious disease. *Nature Reviews Genetics* 10, 540–550.
- Roossinck, M.J., 2011. The good viruses: Viral mutualistic symbioses. *Nature Reviews Microbiology* 9, 99–108.
- Suh, A., Weber, C.C., Kehlmaier, C., *et al.*, 2014. Early Mesozoic coexistence of amniotes and hepadnaviridae. *PLOS Genetics* 10, e1004559.
- Whitfield, Z.J., Dolan, P.T., Kunitomi, M., *et al.*, 2017. The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. *Current Biology* 27. (3511–3519.e7).

Evolution Steered by Structure

Nicola GA Abrescia, Center for Cooperative Research in Biosciences, Basque Research and Technology Alliance, Derio, Spain; Ikerbasque, Basque Foundation for Science, Bilbao, Spain; and Center for Biomedical Research in the Liver and Digestive Diseases Network, Carlos III Health Institute, Madrid, Spain

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

dsDNA double-stranded DNA

dsRNA double-stranded RNA

ssDNA single-stranded DNA

(+)ssRNA positive-sense single-stranded RNA

(-)ssRNA negative-sense single-stranded RNA

Glossary

Analogy Similarity observed between traits that arose independently in different lineages and that resulted from convergent evolution. Also termed coincidence.

Fusogen Any substance that mediates the fusion of two biological membranes.

Homology Similarity observed between traits that descended from a common ancestor by divergent evolution.

Horizontal gene transfer The exchange of genes between organisms of the same or different species, or between

organisms and viruses, in contrast to vertical gene transfer in which parental genes are inherited by progeny.

Jelly-roll A protein fold composed of eight β strands arranged in two four-stranded sheets ("double β -barrel" and "double jelly-roll" are used here as synonyms).

Phylogenetics The study of how epidemiological, immunological, and evolutionary processes act and potentially interact to shape viral phylogenies.

Virosphere The ensemble of all viruses that populate the Earth.

The Evolution of Viruses

Viruses permeate the entire biosphere and are by far the most abundant biological entities, outnumbering all cellular organisms in the biosphere ten-fold. A recognition of their evolutionary relationships helps us to understand how viral infections impact host populations. Moreover, untangling the complexity of viral diversity is an essential component of the human effort to comprehend the natural phenomena of which we are a part.

The raw materials for viral evolution are the heritable genetic changes that viral genomes accumulate during the virus life cycles. The maintenance of these changes in the viral population can be an adaptive response to environmental pressures, pressures that include antiviral pharmacological therapies and the host's immune response. Without such adaptation, that virus would disappear. Viruses

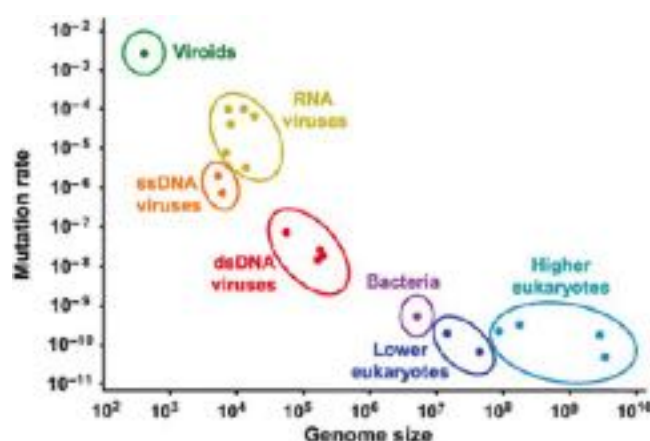


Fig. 1 Estimation of mutation rates in viruses, viroids, and cellular organisms. Diagram plotting the per-site mutation rate against the genome size of various entities: (1) a viroid (*Chrysanthemum chlorotic mottle*); (2) RNA viruses (tobacco mosaic virus, human rhinovirus, poliovirus, vesicular stomatitis virus, bacteriophage $\Phi 6$, and measles virus); (3) ssDNA viruses (bacteriophage $\Phi X174$ and bacteriophage m13); (4) dsDNA viruses (bacteriophage λ , herpes simplex virus, bacteriophage T2, and bacteriophage T4); (5) bacteria (*Escherichia coli*); (6) lower eukaryotes (*Saccharomyces cerevisiae* and *Neurospora crassa*); higher eukaryotes (*Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus*, and *Homo sapiens*). Reproduced with permission from Gago, S., Elena, S.F., Flores, R., Sanjuan, R., 2009. Extremely high mutation rate of a hammerhead viroid. *Science* 323, 1308.

evolve rapidly (**Fig. 1**), so rapidly that evidence of inherited similarity in viral genomes can be obscured, or even lost entirely. Inherited traits with major consequences for the virus life cycle and structure may persist longer than others and therefore provide a more powerful means of assessing viral evolutionary relationships. Even with this approach, the making of such assessments is a complex challenge that draws upon numerous diverse, yet complementary, disciplines including virology, epidemiology, computational biology, genomics, structural biology, evolutionary biology, and others.

When studying the origin and evolution of cellular organisms, the genes encoding ribosomal proteins, which are present in all organisms, have been used as genetic fossils that preserve a record of evolutionary history. Not only do viruses not have ribosomes, but there is no other gene that is present in all viruses. Possible similar universal hallmarks that could be used to infer evolutionary relationships among viruses are still under intense debate. The emerging concept of viruses as “capsid-encoding organisms” underlines that all of these entities need to couple the genetic information for a replicon (nucleic acid capable of autonomous replication) with that for a capsid to produce a virion. This definition is in line with the notion proposed a few years ago by Bamford, Burnett, and Stuart that each virus has an innate “self” which is constituted by structural and functional traits that recapitulate the virion architecture and the major components of the virion. This concept has been applied to viral classification and phylogeny through the analysis of the conformational space adopted by the major capsid proteins (MCPs).

The description of viruses as “capsid-encoding organisms” attempts to draw a parallelism with the definition of cells as “ribosome-encoding organisms”. In so doing, the denigration of viruses as pickpockets of cellular genes fades and is replaced by appreciation of the intricate relationships between viruses and their host cells. These relationships and the capacity of viruses to subvert cellular functions for the production of their virions have led some to wonder whether viruses should be considered to be alive. While viruses remain parasitic entities, their evolution is tightly linked to cellular evolution and the origin of Life, and vice versa. Even a minimal understanding of their evolution demonstrates their impact on several areas of importance ranging from host-virus co-evolution to virus-mediated cross-species transmission, as well as vaccine escape, the evolution and spread of resistance of anti-viral drugs, and so on.

To help organize the known viruses – a small part of the entire virosphere – a hierarchical taxonomic system has been widely used. This classification system, which is based mainly on similarities at the genome level and in virion morphology, groups the viruses by order, family, genus, and species, and more recently expanded to include higher-level taxonomic ranks of class, phylum, kingdom and realm (equivalent to the domain rank for cellular organisms). Another virus classification scheme that has also persisted was proposed by Baltimore in 1971, which groups viruses based on their genome’s molecular structure and replication mechanism. This yielded five groups with straightforward replication of dsDNA, ssDNA, dsRNA, (+)ssRNA, and (–)ssRNA genomes. Two additional groups utilize reverse transcriptase in replication, one of which replicates their ssRNA genome via a DNA intermediate, while the other replicates dsDNA genomes via an RNA intermediate. Currently, the ICTV taxonomy system employs the information available about the viral genome combined with the virus’s cellular host (animal, plant, fungus, bacterium, or archaeon). Although this approach tries to utilize the sheer wealth of data coming from recent viral nucleotide sequencing projects to make sense of the vast viral diversity, viral classification on this basis becomes difficult when sequence similarity at genomic (and protein) level is questionable or has been lost over time.

Thus, some of the basic questions underlying the concept of viral evolution are: First, what tools do we have in our hands to decipher viral evolution? And second, how far can our tools trace viral phylogenies?

In the following sections, I will succinctly touch on some of the tools and viral traits that have been employed in efforts to understand the virosphere and its evolutionary history.

Viral Evolutionary Relationships From Sequence Analysis

When we compare two genes or proteins, the higher the percentage of sequence similarity between them, the lower the probability that they have originated independently. On this basis, we can infer that highly similar sequences are homologous, i.e., that they descended from a common ancestor (homology; divergent evolution). However, this interpretation based on high sequence similarity cannot exclude an alternative scenario driven by convergent evolution (analogy or coincidence); for an elegant explanation of the two different evolutionary scenarios, please see [Koonin and Galperin \(2003\)](#) in Further Reading. Viruses evolve fast (**Fig. 1**). As a result, sequence similarity assessed at the genomic (DNA or RNA) or protein level is quickly lost. Valid inference of common ancestry among viruses requires application of other criteria beyond the single number representing the sequence similarity. The diverse viruses assemble virions that are of different shapes and sizes, and that encapsulate genomes of different types that were replicated by different mechanisms. These, too, are inherited traits whose similarities can also be considered.

Cheaper and more powerful computing, combined with the explosion of viral genomics and metagenomics, has accelerated extensive sequence analysis which, in turn, has established viral relationships that were previously unnoticeable. Analysis of viral sequences has led to two important outcomes, among others: (1) the ability to group viruses in lineages based on sequences of cardinal viral genes/proteins; and (2) the establishment of phylodynamics, a field that draws upon multiple disciplines including epidemiology and evolutionary virology.

When applying the usual distance-based methods of sequence analysis, one must choose whether to compare entire genomes or only selected parts. Whole genome comparisons can be misleading when tracing the evolutionary history of viruses. Careful selection of subgenomic regions for comparison can be more appropriate to this purpose (see below). Another fundamental issue when establishing viral evolutionary relationships from sequence analysis is how to define the minimal similarity threshold required for

two sequences to be recognized as similar. The appropriate threshold depends on the sensitivity of our tools for detection, as well as on the further constraints applied by the sequence alignment strategies and algorithms employed. These constraints may include the position of the sequence within the genome, conserved mutations, frequencies of appearance, and/or any relevant biological information available.

Wise selection of genes for comparison must take into account the eventuality of horizontal gene transfer between viruses and the corresponding hosts. One result of these transfers is that viral genomes can contain modules with different evolutionary histories. To minimize this potential source of error we are obliged to select genes that are more likely to be transferred vertically rather than horizontally. Horizontal gene transfer with the host, for example, can bias an analysis and yield erroneous viral groupings. Of course, a wise assessment demands that the biological function of the genes being considered for analysis must be known. In fact, accurate analysis is even more challenging. Whether considering a gene or a protein, the analysis is limited to comparisons made between primary sequence(s), i.e., between two strings of letters representing bases or amino acids. Each of these one-dimensional strings encodes three-dimensional spatial information and function, neither of which can be decoded from the sequence *a priori*.

Koonin and colleagues have profitably classified viruses using specific viral hallmark genes such as the one that encodes the RNA-dependent RNA polymerase (RdRp) enzyme (Fig. 2). This enzyme catalyzes the synthesis of RNA from an RNA template and is unique to RNA viruses that must replicate the viral genomic RNA. This viral group includes the majority of viruses that infect eukaryotes. Comparative analysis of RdRp sequences has been used to classify these viruses as well as to address their origins. This approach led, for example, to the proposal that the picorna-like superfamily possesses a monophyletic origin. Initially, this sparked some debate about the adequacy of selecting a single gene to represent virus identity as opposed to the structure-based approach that uses the viral capsid fold as a “fingerprint” for mapping viral phylogeny (see below).

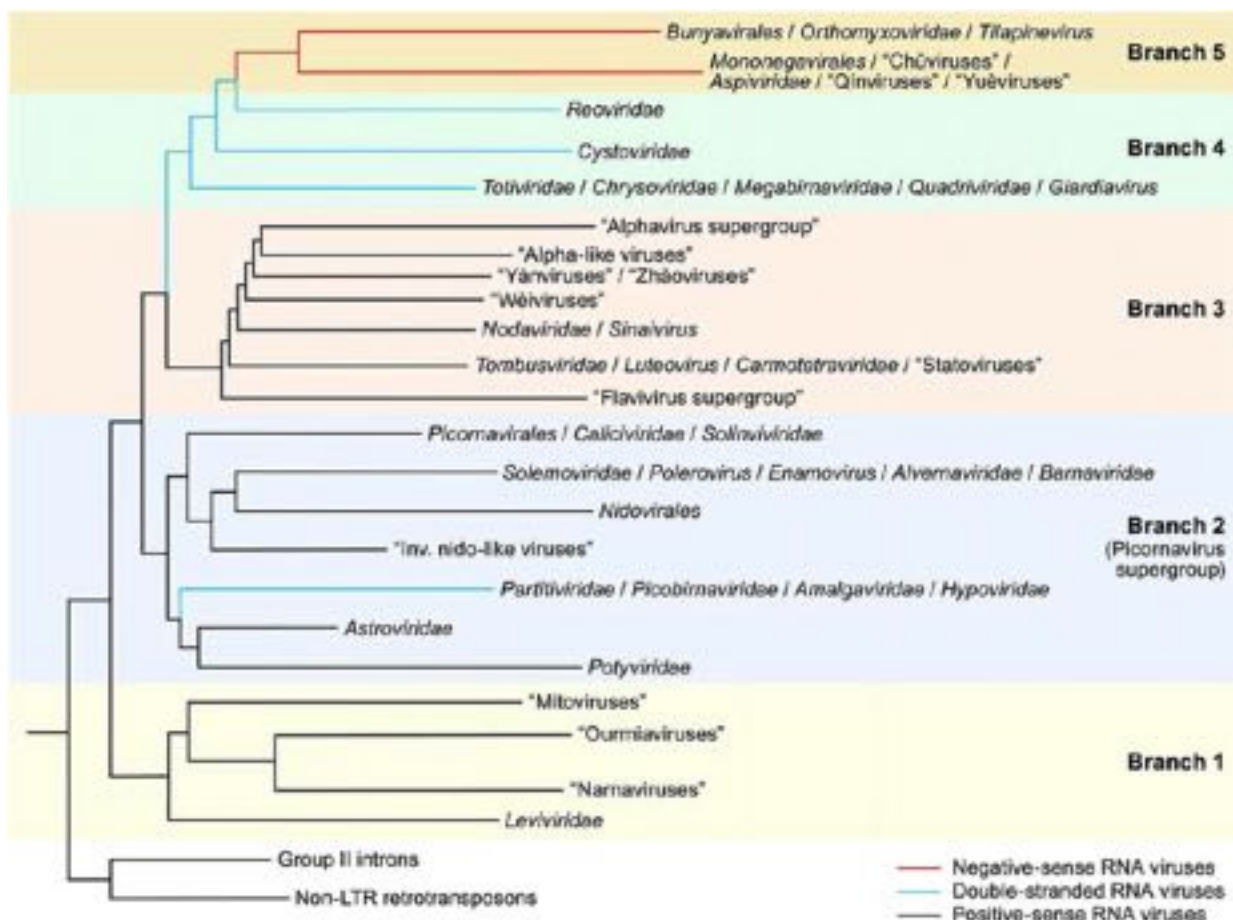


Fig. 2 Phylogenetic tree of RNA viruses using RNA-dependent RNA polymerase (RdRps) and reverse transcriptase (RTs) sequences. The different colors highlight the five main branches emerging from the iterative clustering-alignment-phylogeny procedure. Branch-1: leviviruses and their eukaryotic relatives; Branch-2: (+)ssRNA and dsRNA viruses of eukaryotes; Branch-3: distinct subset of (+)ssRNA viruses; Branch-4: dsRNA viruses; Branch-5: (–)ssRNA viruses. A set of RTs from group II introns and non-long-terminal-repeat (non-LTR) retrotransposons was used as an outgroup. Reproduced and adapted with permission from Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9.

This apparent conceptual and methodological dichotomy is soon resolved by defining a virus as a replicon capable of producing a virion. Within this framework, both approaches manifest strengths and limitations as tools for virus classification. Viral sequence analysis has proven itself to be a powerful means to establish relatedness between genes/proteins and, importantly, to trace the complex intertwined evolutionary relationship between viruses and cellular organisms. Further, the power of comparative viral genomics can also be used to rein in virus emergence. Virus genome sequences paired with their geographical distribution allows prediction of the rise of novel strains or species and also establishment of a molecular clock for the spread of different strains. Although this is not, strictly speaking, the assessment of common ancestry, nevertheless, the predictive power of phylodynamics over short evolutionary time is indisputable. Already it has served to infer spatiotemporal information during the diffusion of the HIV, Ebola, and Zika viruses, among others.

In addition to the viral genomic data, there is an immense amount of metagenomic information currently available thanks to the increased speed and lower cost of high-throughput sequencing. Viral metagenomics retrieves sequences directly from an environment without knowing a priori whether they are derived from viruses, viral hosts, or other sources. Comparison of these metagenomic libraries against known and assigned genome sequences can sometimes reveal both sequence source and function. For sequences from cellular organisms, the source can be identified based on the sequence of the ribosomal genes that all organisms possess. No such universal fingerprint gene exists for viruses. Despite this difficulty, viral metagenomics offers a unique window into viral communities and their ecological niches. It also can help decipher viral phylogeny provided the distance-based tools are sensitive enough to identify at least those viral genetic traits that are less susceptible to the passage of time. Given this wealth of genomic and metagenomic sequence data, the question arises whether it should be integrated into virus taxonomic classification and, if so, how.

A Change of Paradigm: Structure-Based Virus Classification

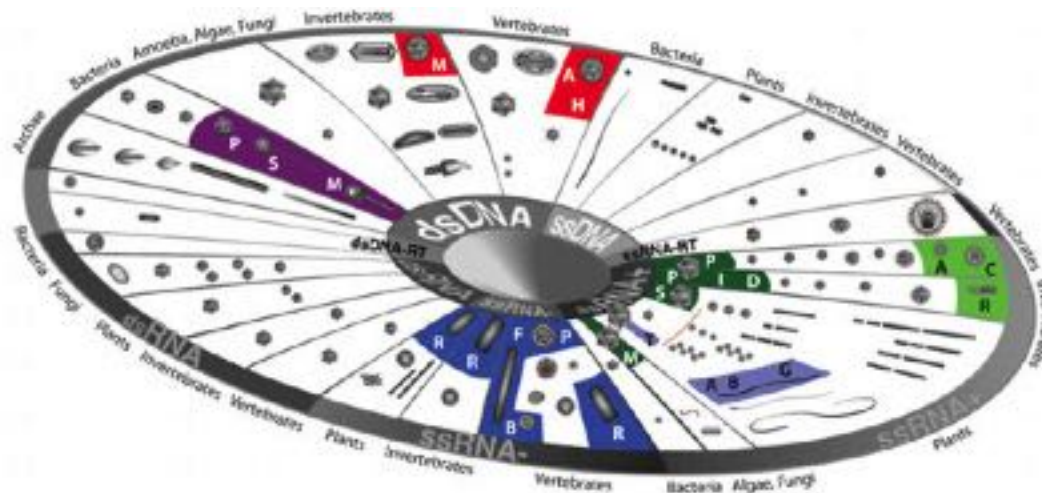
Comparative analysis of viral proteins based on their primary sequence is limited to comparison of monodimensional arrays of twenty letters, each letter identifying one of the twenty common amino acids. By contrast, comparison of three-dimensional (3D) protein structures captures increased complexity. In the eighties, structural analyses of plant and animal-infecting viruses of the *Picornaviridae* family revealed similarities in the corresponding MCP folds despite their lack of sequence similarity. These MCPs possessed single β -barrels arranged with the long axis of the β -barrel tangential to the virus surface. At the time, the dearth of available virus structures made it impossible to determine whether the detected similarities were the remains of an ancestral relationship or were the result of structural necessity, being the only way to form a suitable viral capsid.

With further research came the realization that the capsid protein fold and the arrangement of protein subunits within the capsid can be preserved over evolutionary time. Thus preserved, the same ancient structures can be found today even in viruses that belong to different taxonomic families and that infect divergent hosts. Pivotal here was the determination of the structure of the MCP of the enteric, membrane-containing bacteriophage PRD1 (family *Tectiviridae*) by X-ray crystallography. Strikingly, it was found that the PRD1 MCP possessed the so-called double β -barrel fold (or double jelly-roll fold) equivalent to the human adenovirus hexon MCP (Uniprot code P03277) with which it was superimposable with 3.6 Å rmsd [192 C α s aligned between PRD1 P3 (PDB 1HX6) and adenovirus hexon (PDB 1P30)], but with which it shared only 12.5% sequence similarity (8.3% sequence identity) in a pairwise sequence alignment. This, in conjunction with the initial cryo-electron microscopy (cryo-EM) studies of PRD1, made it clear that both PRD1 MCP P3 and the adenovirus hexon were arranged in similar trimers, with the long axis of the β -barrels orthogonal to the virus surface, to form pseudo-hexagonal capsomer structures. Both viruses displayed the same triangulation number: pseudo- $T = 25$.

This emerging concept that structure was preserved longer than sequence similarity suggested a possible method for demonstrating even more distant evolutionary relationships across diverse viruses. More evidence of this was soon obtained. Simple manual fitting of the four P3 capsomers making up for the PRD1 icosahedral asymmetric unit into the low-resolution 3D cryo-EM derived map of the large chlorella virus PBCV-1 (1,900 Å diameter) (family *Phycodnaviridae*) by Bamford and colleagues supported the hypothesis that PBCV-1 belonged to the same lineage as PRD1 and adenovirus. This hypothesis was then experimentally proven when the Rossmann Lab determined the crystal structure of the PBCV-1 MCP Vp54 and found that it, too, displayed a double jelly-roll fold. Use of the conservation of MCP folds to classify viruses beyond short-distance relationships not only prompted vigorous debate, but also further fuelled the systematic biochemical and structural study of these viral proteins and novel viruses.

Conceptualization of a structure-based viral phylogeny took another step forward with the determination of the crystal structure of the whole PRD1 virus to 4.2 Å resolution. To the compelling structural equivalences in the MCP fold was now added similarity in the architectural building principles of larger viruses such as adenovirus. These observations provided the foundation for definition of the PRD1-adenovirus lineage. In the course of time, bacteriophage PRD1 has turned out to be a remarkable model system for studying important viral processes in lipid-containing viruses. Its structure made clear the virion assembly pathway and also offered a unique scalable mechanism for determining particle size by using a molecular tape measure (see [Abrescia et al., 2004](#) in Further Reading).

However, the concept of the MCP as the defining ‘viral self’ structure, based as it was on comparison of 3D atomic models, led inevitably to debate about homology versus analogy (for a deeper discussion of this topic, I direct the more curious reader to the Further Reading section). When using a structure-based approach in virus classification one must keep in mind that a virus not only has to build a virion, but also to package its genome within. The packaging mechanism, in turn, is tightly linked to how the genome-containing safe box, the capsid shell, is constructed. Therefore, if similarity of the “self” characters (MCP folds and assembly principles) leaves doubts as to the relatedness of two viruses and further constraints are needed, the genome encapsidation mechanism is the next criterion in line. If doubts still remain, comparisons should be made on a case-by-case basis.



Order: Caudovirales

M: Myoviridae
P: Podoviridae
S: Siphoviridae

Order: Herpesvirales

A: Alieherpesviridae
H: Herpesviridae
M: Malacoherpesviridae

Order: Mononegavirales

B: Bornaviridae
F: Filoviridae
P: Paramyxoviridae
R: Rhabdoviridae

Order: Picornavirales

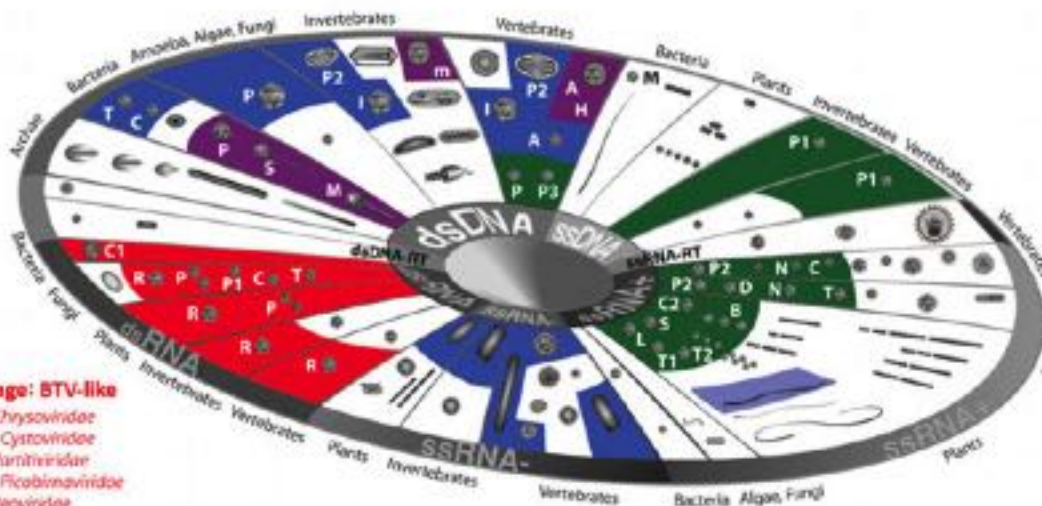
D: Dicistroviridae
I: Ioviridae
M: Moraxviridae
P: Picornaviridae
S: Secoviridae

Order: Tymovirales

A: Alphatymoviridae
B: Betafymoviridae
C: Gammafymoviridae
T: Tymoviridae

Order: Nidovirales

A: Arteriviridae
C: Coronaviridae
R: Roniviridae



Lineage: BTV-like

C: Chrysoviridae
C1: Cystoviridae
P: Parvoviridae
P1: Picrobiviridae
R: Reoviridae
T: Tetraviridae

Lineage: PRD1/Adeno

A: Adenoviridae
C: Corticoviridae
I: Iridoviridae
P: Polyomaviridae
P2: Papillomaviridae
T: Tectiviridae

Lineage: HK97-like

M: Myoviridae
P: Podoviridae
S: Siphoviridae
A: Alieherpesviridae
H: Herpesviridae
M: Malacoherpesviridae

Lineage: Picorna-like

B: Bornaviridae
C: Caliciviridae
C2: Comoviridae
D: Dicistroviridae
L: Luteoviridae
M: Microviridae
N: Nodaviridae
P: Papillomaviridae
P1: Parvoviridae
P2: Picornaviridae
P3: Polyomaviridae
S: Sequoviridae
T: Tetraviridae
T1: Tombusviridae
T2: Tymoviridae



During the past fifteen years, structures of several more viral MCPs and viruses have been determined by different laboratories around the world. This deeper sampling of viral MCP folds, in conjunction with structural comparisons between them all, showed that (so far) many of the viral families within the viral universe can be classified into four lineages: the PRD1-adenovirus lineage, the HK97 lineage (comprising the largest number of representatives), the BTV-lineage, and the picorna-like lineage (Fig. 3). Inevitably, the complexity of ordering the viral universe through this structure-based approach will require constant revisiting of the work done until now, particularly when more viral structures emerge. Unanswered questions still remain. For example, enveloped viruses, for which a consensus “self” has not yet been identified, remain unclassified so far. Moving forward now calls for more viral structures at atomic resolution to deepen and extend the comparisons across different virus families.

Cryo-EM is supplementing X-ray crystallography now more and more thanks to its capability to provide high-resolution (HR) imaging. Recent examples of its usefulness include the structure determination of plant filamentous viruses and the elucidation of the assembly mechanism of viruses whose MCPs contain two vertical single β -barrels. In the former, the cryo-EM structure of helical plant viruses with (+)ssRNA belonging to the *Alphaflexiviridae* and *Potyviridae* families at 3.9 Å resolution unexpectedly revealed that their MCPs are structurally similar to the nucleoprotein of certain non-segmented (–)ssRNA enveloped animal-infecting viruses of the *Bunyaviridae* and *Orthomyxoviridae* families. This observation could be interpreted as evidence that the MCP of these plant viruses was transferred to the animal kingdom and there hijacked for use as a chromosome condensing agent that later, over evolutionary time, acquired a membrane envelope. Should we conclude that these plant and animal viruses share common ancestry? Caution should be used in this case. While the structural homology of the corresponding MCP and nucleoprotein is apparent, no other traits are shared between those viruses, e.g., pepino mosaic virus *versus* Rift Valley Fever virus. This fact undermines the hypothesis of their common descent, but does evidence a possible gene transfer between the plant and animal kingdoms.

In the case of vertical single β -barrel viruses belonging to the *Sphaerolipoviridae* family, the discovery that different viruses using two MCPs displayed capsomers with a pseudo-hexameric footprint but with a distinct number of protruding turrets (even within the same icosahedral asymmetric unit) challenged the known assembly mechanisms. Cryo-EM of whole virions at 3.7 Å resolution showed that the two vertical single β -barrel MCPs (one of which turreted) formed heterodimers that assembled as pseudo-hexameric capsomers to form the capsid shell. This step relies on two proteins with different folds located proximal to the membrane vesicle, as well as on the organizing membrane protein complex at the vertices. The individual vertical single β -barrels whose hetero-dimerization replicates the double β -barrel MCP, can be structurally classified as clade within the vertical double β -barrel PRD1-adenovirus lineage (Fig. 4). Most importantly, determination of these virus structures has revealed how the fusion during evolution of the two vertical single jelly-roll MCPs led to the assembly of membrane-less members of the PRD1-adenovirus lineage such as adenovirus.

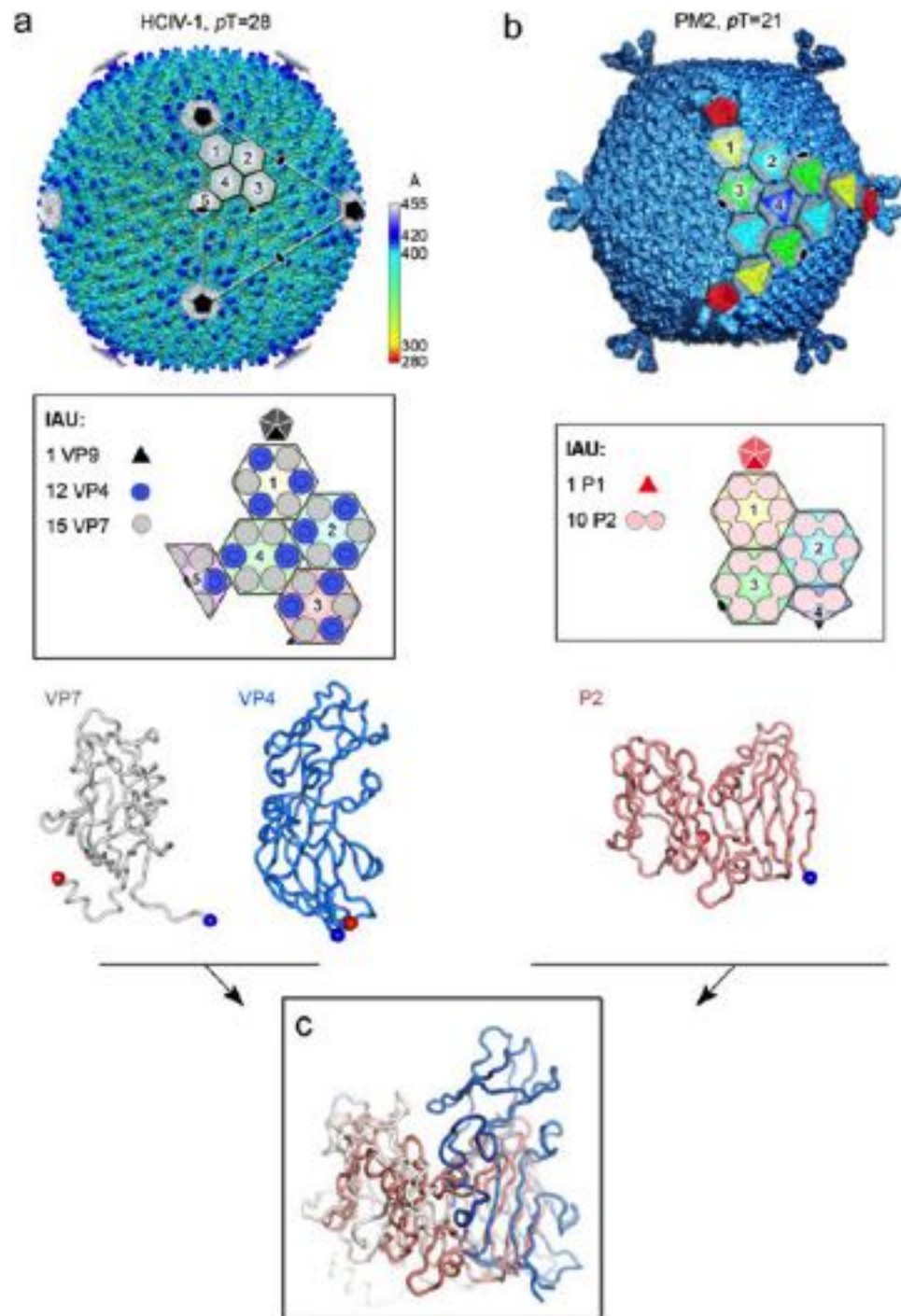
In conclusion, changes in the 3D structure seem to occur at a slower rate than do changes in the sequence, indicating that the sequence space is larger than the fold space. Possible folds of these proteins are more constrained by their function, while sequence similarity is more easily lost without loss of function during evolution. Therefore, similarity of viral sequences is a less reliable means for mapping distant viral phylogeny. A better tactic here is to compare the structures of viral “self” proteins (e.g., proteins involved in virion assembly and those coupled to the genome packaging processes) whose alteration might compromise the virus identity and/or the production of competent progeny. In this view, the structural-functional requirements placed on the viral “self” act as a biological constraint on evolutionary changes, thus “steering” evolution.

Inferring Viral Evolutionary Relationship Through Structure-Based Classification

Identifying homology in protein structures requires more than simply interrogating the virosphere. Development of tools for this next step and their refinement represents one of the ongoing challenges of modern biology.

To infer evolutionary relationships using the 3D structure of viral proteins entails four important concepts: (1) targeting the “correct” viral protein, i.e., one that reflects the invariant character of a virus (“self”) through evolution; (2) solution of the 3D structure of this viral protein; (3) development and application of “specific” tools for structural comparison; and (4) translation

Fig. 3 The virosphere and the representative MCP fold for each structure-based lineage. (Above) The International Committee on Taxonomy of Viruses (ICTV) scheme (adapted from the Universal Virus Database 2005) is shown with the currently defined orders and the virus families within each defined by letters. Families that in 2005 were not assigned to an order are drawn in black (*Adenoviridae*, *Arenaviridae*, *Ascoviridae*, *Asfarviridae*, *Astroviridae*, *Baculoviridae*, *Barnaviridae*, *Birnaviridae*, *Bromoviridae*, *Bunyaviridae*, *Caliciviridae*, *Caulimoviridae*, *Chrysoviridae*, *Circoviridae*, *Closteroviridae*, *Corticoviridae*, *Cystoviridae*, *Flaviviridae*, *Fuselloviridae*, *Geminiviridae*, *Guttaviridae*, *Hepadnaviridae*, *Hypoviridae*, *Inoviridae*, *Iridoviridae*, *Leviviridae*, *Lipothirixviridae*, *Luteoviridae*, *Metaviridae*, *Microviridae*, *Nanoviridae*, *Narnaviridae*, *Nimaviridae*, *Nodaviridae*, *Orthomyxoviridae*, *Papillomaviridae*, *Partitiviridae*, *Parvoviridae*, *Phycodnaviridae*, *Plasmaviridae*, *Polydnaviridae*, *Polyomaviridae*, *Potyviridae*, *Poxviridae*, *Pseudoviridae*, *Reoviridae*, *Retroviridae*, *Rudoviridae*, *Tectiviridae*, *Tetnaviridae*, *Togaviridae*, *Tombusviridae*, *Totiviridae*). (Below) MCP structure-based viral lineages mapped onto current ICTV taxonomy with each lineage colored differently. Individual viral families within each lineage are labeled and colored according to the key. Beneath that, representative ribbon diagrams depict the 3D structural models for each of the lineages. Structures are drawn to scale, colored in rainbow from the N-terminus (blue) to the C-terminus (red), and viewed as it would sit on the virion surface. Abbreviations: BTV, bluetongue virus (PDB ID 2BTV); PRD1/Adeno, PRD1/Adenovirus (PDB ID 2VVF); HK97, bacteriophage Hong Kong 97 (PDB ID 1OHG); Picorna, picornavirus (PDB ID 5NE4, VP2 subunit). Adapted with permission from Abrescia, N.G., Bamford, D.H., Grimes, J.M., Stuart, D.I., 2012. Structure unifies the viral universe. *Annual Review of Biochemistry* 81, 795–822.



of the result of a structural comparison into an evolutionary distance. Once the viral “self” has been identified or hypothesized (see previous sections), the limiting factor in ordering the viral universe through a structure-based approach is the availability of a sufficient pool of target proteins for which high resolution 3D structures have been determined.

With such a pool of structural data in hand, the next step is to employ this data in unraveling ancient viral evolutionary history. The fold of (viral) proteins is more than the manifestation in the three-dimensional space of the one-dimensional primary sequence. It is also the product of optimization of its biological function during evolution through the establishment of unique intra- and inter-molecular interactions. When molecular structures are available, their comparison has the power to make evident what constraints have been applied during evolution in order to maintain essential biological functions, even when similarity between the linear amino acid sequence appears to have been lost. This became apparent when the double β -barrel was found to be shared by the MCPs of membrane-containing bacteriophage PRD1 and adenovirus. With the viral MCPs we have proteins with a quintessential role in protecting the viral genome when outside the host cell. Their vital importance implied that a three-dimensional alteration could jeopardize the virus’s very existence. This, in turn, supported the hypothesis that in this case the structural similarity represented homology rather than a coincidental event. This view was later reinforced by the facts that virion architecture is, of necessity, coupled to genome packaging and that the genes for the capsid and for genome packaging often evolve together.

But how does one quantify the degree of the three-dimensional similarity when comparing viral ‘self’ proteins? The root-mean-square-deviation (rmsd) is one obvious measure. However, the choice of the method to be employed to structurally superimpose the two 3D atomic models constitutes an important decision as this will influence the meaningfulness of the resulting rmsd.

The Structural Homology Program (SHP), written by Stuart and colleagues and based on the technique developed by Argos and Rossmann, has been widely used for structural comparison of viral proteins. SHP is sensitive to “weak similarities” and relies on the probability of equivalence between alpha carbons (C α s) in pairwise structure comparisons. When comparing two proteins, designated A and B, the structure of protein A is rotated to explore all possible orientations about its center of mass. For each orientation relative to the structure of protein B, the probabilities of structural equivalence between the C α s are calculated. Specifically, these calculations for residues *i* and *j* in structures A and B take into account the closeness in space of their C α atoms and the local shape of the two polypeptide chains around these residues. This yields an estimate of the likelihood/probability (P_{*i,j*}) that these two residues (*i* and *j*) are structurally equivalent to each other. Once this has been estimated for all C α s, a matrix of P_{*i,j*} values for all *i* and *j* is built. The best path across this matrix is then defined as the one that fulfils the largest set of equivalences between A and B and maximizes the total probability, (Σ) P_{*i,j*}, for all residue pairs. From this an rmsd can also be produced for these two proteins (Fig. 5(a)).

At this point, having defined the total probability as an overall measure of the structural similarity of two proteins, the next step is to transform this probability into an evolutionary distance. In SHP the logarithm function is used for this transformation. The value of this function is 0 when its argument is 1; when the argument is less than 1, the value approaches minus infinity. Thus, if the probability of equivalence is 1, i.e., the structures are 100% equivalent (the same structure), then their evolutionary distance (D) is 0. Similarly, when the probability of equivalence approaches 0, the distance approaches infinity (there is a change in sign in the definition of D) (Fig. 5(a) and (b)). In this manner one can calculate the pairwise distances between various 3D protein structures and from this generate a distance matrix. This can then be fed into the PHYLIP phylogenetic analysis software package for transformation into a phylogenetic tree. A more exhaustive analysis of the classification methods based on 3D structures can be found elsewhere in the encyclopedia.

Fig. 4 Capsid organization of vertical single β -barrel MCPs and vertical double β -barrel MCP of the PRD1-adenovirus lineage. (a) Top, cryo-EM map at 3.7 Å resolution of *Haloarcula californica* icosahedral virus 1 (HCIV-1), a lipid-containing archaeal virus, pseudo-*T* = 28, color-coded by radius as indicated in key, with one virus facet delimited by white lines. Black pentagons, ovals, and triangle indicate the five-fold, two-fold, and three-fold axes of icosahedral symmetry, respectively. The white capsomers numbered from 1 to 3 yield a pseudo-hexameric, three-turreted morphology, while capsomer 4 and half-capsomer 5 (sitting on the two-fold axis of symmetry) yield a pseudo-hexameric, two-turreted morphology. Capsomers 1–5, together with one copy of the penton protein, compose the icosahedral asymmetric unit (IAU). Middle, schematic representation of the IAU with the arrangement of the individual MCPs depicted as circles (12 copies of turreted VP4 in blue and 15 copies of VP7 in gray) forming the differently-colored pseudo-hexameric capsomers, one copy of the penton protein VP9 as a black triangle, and the remaining four copies of VP9 sitting on the five-fold axis in dark gray. Bottom, tube representation of the individual vertical single β -barrel MCPs VP7 and VP4. Blue and red spheres denote the N-terminus and C-terminus, respectively. (b) Top, crystallographically-derived electron density at 7 Å resolution (slate-blue) of a member of the PRD1-adenovirus lineage: marine, lipid-containing bacteriophage PM2 (outer diameter 570 Å between icosahedral facets; pseudo-*T* = 21). One triangular virus facet is represented by the grey hexagons with on top triangles (coloured in yellow, green, cyan and blue) indicating, respectively, the displayed pseudo-hexameric morphology of the capsomers and the oligomerization state of the vertical double β -barrel MCP P2. The capsomers that compose the IAU are numbered from 1 to 4. Capsomer 4 sits on the icosahedral three-fold symmetry axis, red pentagons sit on the five-fold vertices, and the black ovals mark the two-fold axes of icosahedral symmetry. Middle, schematic representation of the IAU with the arrangement of the vertical double β -barrel MCP P2 depicted as pink dumbbells. Three copies of P2 form each of the pseudo-hexameric capsomers 1–3 and one copy forms capsomer 4 at the icosahedral three-fold axis (black triangle). The copy of the penton protein P1 participating in this IAU is depicted as a red triangle (with the remaining four copies of P1 sitting on the five-fold axis shown in dark-pink). Bottom, tube representation of the vertical double β -barrel P2 MCP with blue and red spheres indicating the N-terminus and C-terminus, respectively. (c) Superimposition of HCIV-1 MCP VP7-VP4 heterodimer (white-blue, 403 residues, PDB ID 6h9c), the building block for HCIV-1 capsomer assembly, onto the double β -barrel MCP P2 of bacteriophage PM2 (pink; 269 residues, PDB ID 2VVf), to yield a 4.2 Å rmsd with 189 C α equivalences. The relative angular orientation between the individual vertical HCIV-1 VP7-VP4 β -barrels practically replicates that of the double β -barrel MCP.

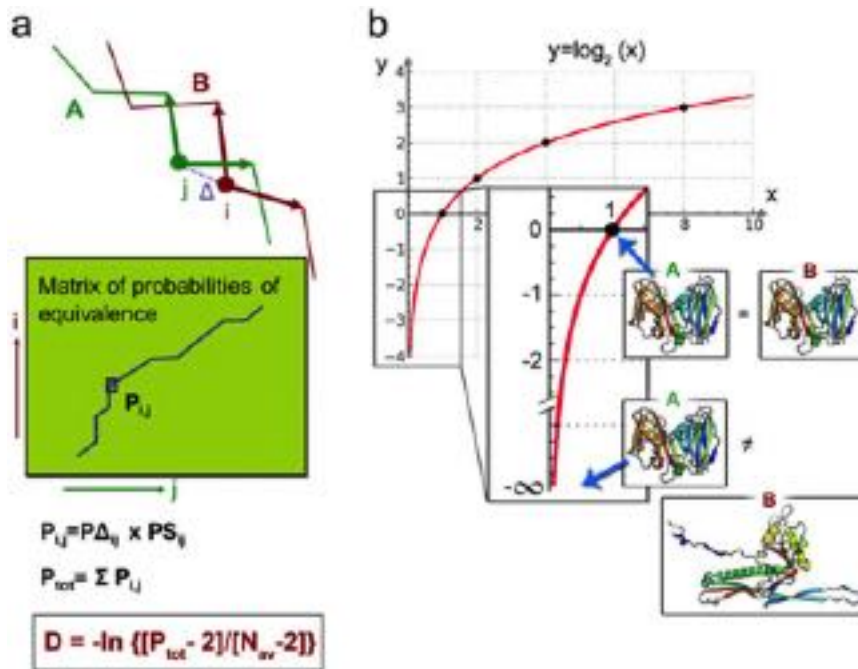


Fig. 5 Schematic representation of the concepts of structural equivalence and evolutionary distance. **(a)** Diagrammatic representation of a matrix of probabilities of equivalence calculated between two protein structures: A (green) and B (brick red). The i denotes the C α of residue i in A, while j denotes the C α of residue j in B. P_{ij} is the probability of equivalence for two given C α s, where $P\Delta_{ij}$ is a probability factor representing the distance between the two residues and PS_{ij} is a probability factor that represents the similarity of shape between the two chains in the regions surrounding residues i and j . The total probability ($P_{\text{tot}} = (\sum)P_{ij}$) defines the best 'path', i.e., the conformation that contains the largest set of correspondences and maximizes the sum of the individual probabilities. D is the evolutionary distance between the two structures (A and B) defined as the natural logarithm of the total probability minus two (the corresponding probability of equivalence of matching the N- and C-termini), normalized over the average numbers of C α s of A and B, N_{av} , with the removal of the N- and C-terminal C α s. **(b)** Properties of the logarithmic function that are exploited for the definition of the evolutionary distance (for visual clarity, this is shown for base 2 but holds for any base > 1). For the limiting case when $A = B$ (e.g., ribbon diagram of PM2 MCP P2 colored as a rainbow from blue N-terminus to red C-terminus), the argument of the logarithm is 1 and its result, the corresponding evolutionary distance D , is 0. When the two 3D structures are increasingly dissimilar (e.g., PM2 MCP P2 versus HK97 MCP), the argument of the logarithm approaches 0 as the evolutionary distance increases towards infinity. Please note the negative sign in the definition of D in (a) that inverts the $-\infty$ to ∞ .

The Fold of Some Viral Envelope Glycoproteins Recapitulates That of Cellular Proteins

Different scenarios have been postulated for the origin of viruses: (1) the "virus early" hypothesis which posits that viruses emerged from a pool of replicative elements that preceded the first protocell; (2) the 'regression' hypothesis which posits that viruses emerged from the degeneration of cells; (3) the "escaped genes" hypothesis which suggests that viruses are derived from cellular genes that acquired the capabilities required for the parasitic viral lifestyle; and (4) the "chimeric" hypothesis which proposes that viruses arose from replicative elements that existed prior to the first protocell and which later acquired virion structural elements from primitive cells (Krupovic and Koonin, 2017 in Further Reading). All four scenarios inescapably involved an initial exchange of genetic material with a (proto)cell, with further exchanges during the subsequent evolutionary steps.

Establishing virus relationships requires one to distinguish viral genes that are transferred vertically at infection from those exchanged horizontally with the host. Identification of structural homologs by comparison of 3D structures of cellular and viral proteins may help to identify those that have been exchanged between them. For instance, consider the three known classes of viral fusion glycoproteins (class I, II, III). All three classes mediate fusion with a cell membrane. Class II fusion glycoproteins are found in some viruses, including members of the *Togaviridae* and *Flaviviridae*. These viral class II fusogens are structural homologs of cell-cell fusogens such as epithelial fusion failure 1 (EFF-1) that is responsible for syncytia formation during *C. elegans* embryogenesis and gamete fusogen HAPLESS 2 (HAP2). Although viral and cellular fusogens share the same domain architecture (with three domains: DI, DII, DIII) and protein fold, their mode of action has been specifically adapted to the different tasks at hand. Fusion in the case of the viruses is unidirectional (heterotypic), while for cellular proteins fusion is bidirectional (homotypic) (Fig. 6).

These glycoproteins evidence the co-evolution of virus and host, but this co-evolution leaves us with the question as to the origin of these glycoproteins. Although the original directionality of the transfer might never be established, the broad distribution of HAP2 in cellular organisms, including some basal lineages, in contrast to the rather narrow distribution in viruses suggests that these viruses had recruited the corresponding cellular genes. In summary, this example adds further evidence that viral evolution cannot only conserve a useful fold, whatever its origin, but also fine tune it further to better serve viral purposes.

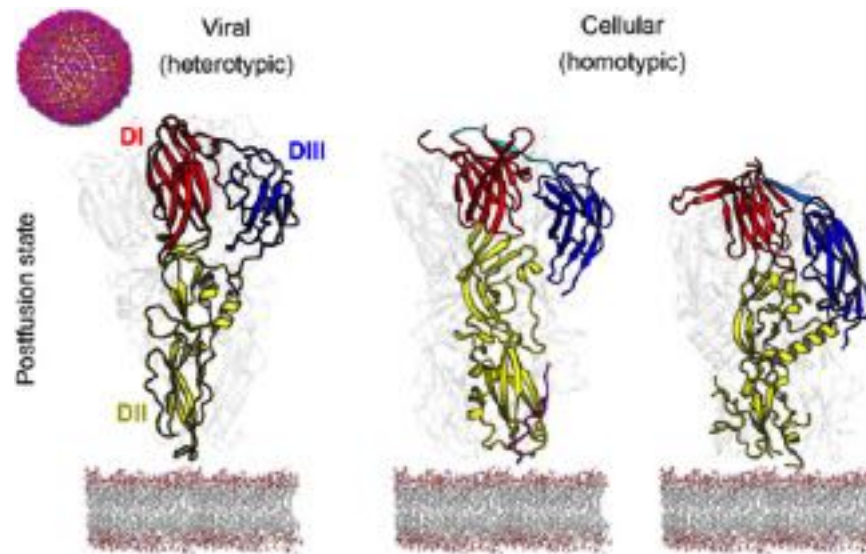


Fig. 6 Structural and functional conservation of viral and cellular proteins involved in membrane fusion. Side-by-side comparison of viral and cellular membrane fusogens in the post-fusion state. These cartoons represent the overall protein shapes and the arrangement and architecture of their domains, color-coded according to the classical color scheme originally given to the envelope protein of Tick-borne encephalitis virus (TBEV) E protein (PDB ID 1URZ) on the left. At the top left, the icosahedral enveloped TBEV virion (member of the *Flaviviridae* family) in a pre-fusion state (EMDB EMD-3752) (not to scale). At the center cellular protein EFF1 (PDB ID 4OJE), and on the right, HAP2 (PDB ID 5MF1). Each of the three proteins forms a trimer in the post-fusion state. Their corresponding partnering molecules are depicted in transparent white.

Would comparison of these viral fusion glycoproteins with one another be a useful tool to probe the ancient evolutionary relationships among the enveloped viruses? These glycoproteins in the viral envelope bind to the cellular receptor and facilitate membrane fusion with the host cell membrane. Host receptors evolve rapidly under pressure to escape viral infection. In response, viral fusogens are likely to also evolve more rapidly and thus to have only limited value as a signature indicating an ancient common ancestry. By contrast, evolutionary change to components such as the nucleoprotein might be more constrained because, in this instance, the protein being studied must interact with the viral genome. Although far from being agreed upon, currently there is an inclination to consider the nucleoprotein fold as a more informative character for the establishment of phylogeny for the enveloped viruses. This nucleoprotein is intimately associated with the genome which it enfolds, condenses, and encapsidates. However caution should be exercised when extrapolating a common ancestry for any two viruses based solely on the structural homologies detected for any single protein. One confounding factor is that many extant capsid protein and nucleoprotein folds appear to have originated from diverse cellular antecedents on independent occasions. These folds became “self” characters of a viral lineage when architecture, fold, and coupled functions were vertically transmitted together to viral descendants. For the above reasons, establishment of a valid structure-based viral phylogeny for the enveloped viruses awaits structure determination for many more enveloped virus proteins, including glycoproteins and nucleoproteins among others. Nevertheless, the correspondence between the folds of viral and cellular fusion glycoproteins reiterates that comparison of protein 3D structures is a powerful means to detect protein homology even when the sequence similarity has been lost.

Architectural and Functional Similarities With Cellular Structures: Are Exosomes Reminiscent of Enveloped Viruses?

Overall shape, domain architecture and fold are defining spatial properties not only of proteins but also of the structures of which they are a part. The assembly mechanism of a virion from its constituent biological components (e.g., nucleic acids, proteins, lipids) and the mechanical properties of the resultant virion rely on the specific 3D structures and physicochemical properties of its building blocks. These virus-specific assembly mechanisms, however, evolved in symbiotic relationship with the mechanisms active in the cellular host. This raises the question whether similar assembly processes are used by cellular organisms.

This brings us to exosomes, relatively small vesicles released by cells that carry different sets of proteins, nucleic acids, and lipids depending on the cell type of origin (Fig. 7(a) left). One type of exosome characterized early by cryo-EM was derived from rat liver. These were spherical structures that ranged from ~35 to ~80 nm in diameter (Fig. 7(b)). Subsequent analyses have visualized exosomes derived from other cell types. All so far share the same architecture including a lipid bilayer (composed mainly of cholesterol, ceramide, and glycerophospholipids) fenestrated by cellular membrane proteins (including integrins and tetraspanins, among others). Orientation of these membrane proteins relative to the lipid membrane is not always clear. Some exosomes have also been shown to

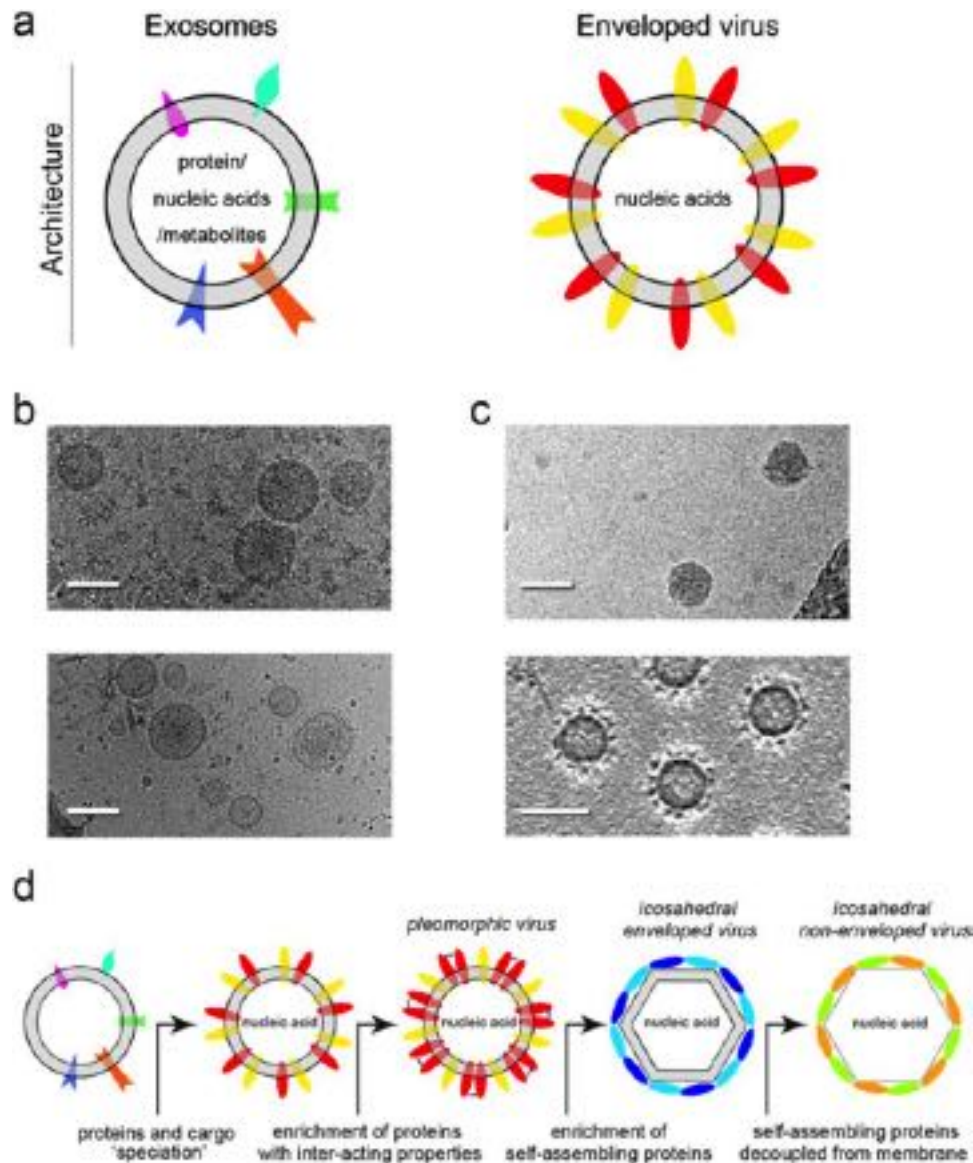


Fig. 7 Architectural similarities between exosomes and enveloped viruses. **(a)** Left, schematic of an exosome showing the vesicle (gray) harboring varied proteins (various shapes, differently colored). The orientation of the membrane proteins relative to the bilayer is not known. Right, schematic of a spherical, enveloped virus with membrane-bound glycoproteins (yellow and red). Their position on the envelope shown here is arbitrary. **(b)** 2D cryo-EM images of exosomes isolated from (above) supernatant of a human hepatic adenocarcinoma cell line (SK-HEP-1) and (below) from supernatant of a murine liver progenitor cell line (MP29); scale bar = 100 nm. (Source: obtained in-house on a JEOL JEM2200FS with a CCD detector. Courtesy of Dr. Falcón-Pérez at CIC bioGUNE). **(c)** (Above) 2D cryo-EM image of icosahedral enveloped Zika virus (member of the *Flaviviridae* family) produced, purified, and imaged in our laboratory; scale bar = 50 nm. (Below) slice through a 3D tomogram of the archaeal *Halorubrum pleomorphic virus 6* with only one glycoprotein species (VP5) fenestrating the envelope; scale bar = 40 nm. Reproduced under the terms of the Creative Commons CC BY license from Omari, K.E., Li, S., Kotecha, A., *et al.*, 2019. Nature Communications 10, 846. **(d)** Schematic diagram of speculative structural events leading to the emergence of enveloped, icosahedral enveloped and non-enveloped viruses.

incorporate viral proteins, for example during infection and transmission of hepatitis C virus (HCV), as well as during heterologous overexpression of HCV polyproteins (*unpublished results*). Others participate in the regulation of various pathophysiological processes including angiogenesis, inflammation, and metastasis. Fundamentally, exosomes function in intercellular communication and intercellular transport of biological material, including viruses such as hepatitis A, HIV, and poliovirus, within and between tissues. Their specific protein composition reflects the cell type from which they were purified and its pathophysiological state. Active research is currently ongoing to explore their potential use as biomarkers for human diseases and to investigate possible specificity of cargo or a pre-established ratio of structural components.

Exosomes originate from endosomal compartments, specifically from the multivesicular bodies (MVBs) that contain membrane-bound intraluminal vesicles. These (proteo-)vesicles assemble by budding into the lumen of the MVBs and can be released into the extracellular space via fusion with the plasma membrane. Architecturally exosomes are reminiscent of enveloped viruses (Fig. 7(a) right and 7(c)). Their morphogenesis is suggestive of pleomorphic members of the *Flaviviridae* family (e.g., hepatitis C virus, bovine viral diarrhea virus) rather than those of the *Alphaviridae* family whose biogenesis resembles the pinching-off of vesicles from the plasma membrane. However, both the vesicles formed as intraluminal vesicles and released via exocytosis (termed delayed exosome biogenesis) and those released by budding at the plasma membrane (termed immediate exosome biogenesis) can be defined as “exosomes”. The resemblance of the architecture and release mechanism between the budding exosomes and members of the *Retroviridae* family appears even stronger now than when originally postulated in 2003 in “The Trojan exosome hypothesis” by Gould and co-workers. They hypothesized not only that retroviruses use the pre-existing exosome biogenesis pathway for their assembly, but also that they take advantage of the pre-existing exosome uptake pathway for infection.

Exosomes – as we understand them today – appear to be byproducts of fundamental cellular physiology and housekeeping, and as such lack the specificity and the replicative character of viruses. Nevertheless, due to their architecture and their inherent self-assembly capability, exosomes might represent a suitable framework for the subsequent incorporation of specialized functions over evolutionary time. Interestingly, two of the simplest enveloped viruses structurally described so far, *Halorubrum pleomorphic virus* 2 (HRPV-2) and 6 (HRPV-6), strongly resembles a vesicle (Fig. 7(d)). Both are ssDNA viruses whose replication mechanism is not yet clear and they infect halophilic archaeal organisms (Archaea have also been shown to secrete exosomes). Their corresponding virions possess protruding spikes composed solely of a monomeric membrane-associated protein, VP5, whose fold has not been observed in any of the known viral class I-III fusogens.

Thus, can we consider certain enveloped viruses as being evolutionarily adapted exosomes? Might the evolution of the structure/architecture of the exosome in this direction have been steered by the relative abundance of a specific protein or protein class? A general or conclusive answer is not possible. Any claimed similarities between exosomes and viruses have always to be questioned. More convincing than current structural similarities may be the evident commonalities between retroviral and exosome biogenesis and release. In addition, the architectural mold offered by exosomes morphogenesis lend them exquisitely well as natural benchtops for experimenting how different 3D structures of cellular proteins associated to the envelope might modulate exosomes assembly and oversee recruitment of the internal cargo. Thus, while exosomes are extremely heterogeneous in size, in both their protein and nucleic acid compositions and tuneable by the cellular physio-pathological conditions, they engrave the archetypical composite architecture whose further evolution steered by the origination or enrichment of certain protein folds and motifs, might have led to the appearance of enveloped viruses.

These primordial and likely pleomorphic enveloped viruses may have been enriched for envelope-binding proteins and also for proteins with increased self-interactions. These same proteins might have come to form a proteinaceous shell (capsid) that could both protect the cargo engulfed by the vesicle and define the size of the resultant particles.

In subsequent steps, the incorporation of proteins with specialized capabilities, such as coating of an exterior membrane, self-assembly, and interaction with the host, might have rendered the vesicle membrane dispensable, thus giving rise to non-enveloped viruses (Fig. 7(d)). Today icosahedral capsid architecture is common in virions of diverse but discrete sizes, either enveloped or not. This development would reflect the selection of only those combinations of these viral coating proteins able to generate viable viral particles, thus leading to a discretization of the size of icosahedral virions. This discretization is currently described by the so-called triangulation number T , an integer number of defined values which increases with increasing virion size and which, for dsDNA viruses for example, broadly echoes the larger size of the viral genome packaged within.

Conclusions

During recent years, huge efforts have been made toward a comprehensive ordering of the viral universe through the use of sequence-based, and increasingly, structure-based methodologies. Nevertheless, deciphering viral phylogenies remains daunting. The vast diversity of viruses populating the biosphere remains mostly uncharacterized. More viral protein and virion structures, in particular, are needed! By their example, our few successes so far in using structural comparisons to establish viral evolutionary relationships portend the significant impact that this information can have across multiple scientific disciplines. Moreover, combined with elucidation of the modes of virion assembly and packaging, structural analyses may even provide a window on the earliest steps in viral evolution.

Acknowledgments

I am indebted to Prof. Dennis Bamford for bringing to my attention gaps in my knowledge – many of which still exist – and for the insightful comments on the manuscript. I am also grateful to Prof. David I Stuart for allowing me to use his original sketch of Fig. 5(a) and introducing me to the viral structure-based phylogeny field. I wish to thank also the past and present members of the Abrescia Lab for continuous support and discussions and in particular Ane Martinez-Castillo for carefully reading and commenting upon the manuscript.

N.G.A.A. is supported by the Spanish *Ministerio de Ciencia, Innovación y Universidades* (RTI2018–095700-B-I00) and by the Basque *Departamento de Educación, Política Lingüística y Cultura* (Refs: PRE_2016_2_0151, PRE_2018_1_0102). MICINN is also thanked for the Severo Ochoa Excellence Accreditation to the CIC bioGUNE (SEV-2016–0644).

See also: Classification of the Viral World Based on Atomic Level Structures

References

- Abrescia, N.G., Cockburn, J.J., Grimes, J.M., *et al.*, 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* 432, 68–74.
- Koonin, E.V., Galperin, M.Y., 2003. Evolutionary concept in genetics and genomics. *Sequence-Evolution-Function: Computational Approaches in Comparative Genomics*. Boston: Kluwer Academic.
- Krupovic, M., Koonin, E.V., 2017. Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences of the United States of America* 114, E2401–E2410.
- Bamford, D.H., Burnett, R.M., Stuart, D.I., 2002. Evolution of viral structure. *Theoretical Population Biology* 61, 461–470.
- Conde-Vancells, J., Rodríguez-Suárez, E., González, E., *et al.*, 2010. Candidate biomarkers in exosome-like vesicles purified from rat and mouse urine samples. *PROTEOMICS – Clinical Applications* 4, 416–425.
- Dietmann, S., Holm, L., 2001. Identification of homology in protein structure classification. *Nature Structural & Molecular Biology* 8, 953–957.
- Forrester, P., Prangishvili, D., 2009. The origin of viruses. *Research in Microbiology* 160, 466–472.
- Gould, S.J., Booth, A.M., Hildreth, J.E., 2003. The Trojan exosome hypothesis. *Proceedings of the National Academy of Sciences of the United States of America* 100, 10592–10597.
- Holmes, E.C., 2010. Evolution in health and medicine Sackler colloquium: The comparative genomics of viral emergence. *Proceedings of the National Academy of Sciences of the United States of America* 107 (Suppl 1), 1742–1746.
- Holmes, E.C., 2011. What does virus evolution tell us about virus origins? *Journal of Virology* 85, 5247–5251.
- Kolodny, R., Pereyaslavets, L., Samson, A.O., Levitt, M., 2013. On the universe of protein folds. *Annual Review of Biophysics* 42, 559–582.
- Koonin, E.V., Dolja, V.V., 2013. A virocentric perspective on the evolution of life. *Current Opinion in Virology* 3, 546–557.
- Krupovic, M., Dolja, V.V., Koonin, E.V., 2019. Origin of viruses: Primordial replicators recruiting capsids from hosts. *Nature Reviews Microbiology* 17, 449–458.
- Nolte-T Hoën, E., Cremer, T., Gallo, R.C., Margolis, L.B., 2016. Extracellular vesicles and viruses: Are they close relatives? *Proceedings of the National Academy of Sciences of the United States of America* 113, 9155–9161.
- Oksanen, H.A., Abrescia, N.G., 2019. Membrane-containing icosahedral bacteriophage PRD1: The dawn of viral lineages. In: GREBER, U. (Ed.), *Physical Virology - Virus Structure and Mechanics*. Springer Nature Switzerland.
- Podbilewicz, B., 2014. Virus and cell fusion mechanisms. *Annual Review of Cell and Developmental Biology* 30, 111–139.
- Rossmann, M.G., Argos, P., 1976. Exploring structural homology of proteins. *Journal of Molecular Biology* 105, 75–95.
- Santos-Perez, I., Charro, D., Gil-Carton, D., *et al.*, 2019. Structural basis for assembly of vertical single beta-barrel viruses. *Nature Communications* 10, 1184.
- Valle, M., 2018. Structural homology between nucleoproteins of ssRNA viruses. *Subcellular Biochemistry* 88, 129–145.

Relevant Websites

- <https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses (ICTV).
- <http://viprdb.scripps.edu/>
The Scripps Research Institute.

Pairwise Sequence Comparison in Virology

Tao Zhang, Zheng Gong, Tongkun Guo, and Zhuang Xiong, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Yiming Bao, Beijing Institute of Genomics, Chinese Academy of Sciences; University of Chinese Academy of Sciences, Beijing, China

© 2021 Elsevier Ltd. All rights reserved.

Virus Classification Methods, the Increase Use of Sequences for Virus Classification

Viruses are the most abundant and genetically diverse acellular biological entities on the Earth. There are at least 10^{31} virus particles spread across most environments of the world, including marine and freshwater habitats and gastrointestinal tracts of metazoan, where the number of detectable viral particles exceeds 10–100 times that of cells. Viruses not only are large in number, but also have very important functions. Marine viruses, for example, play significant roles in inducing microbial mortality, shaping microbial community structure, and mediating the horizontal gene transfer between different host cells. There still exists an extraordinary amount of uncharacterized viral “dark matter” in the world. Therefore, studying and classifying these unknown viruses will help us make better use of virus resources and will lead to a better understanding of the ecology, history, and impact of the global viromes.

Virus classification is the process of naming and placing viruses accurately into a taxonomic system. This taxonomic system places viruses that are similar to each other in a hierarchy of relationships, and can describe viral diversity and the genetic and evolutionary relationships between different viruses, helping us understand the world of viruses. For international norms on virus classification, the formal taxonomic classification of viruses is the responsibility of the International Committee on Taxonomy of Viruses (ICTV) system which began to devise and implement virus naming and classification rules as early as the 1970s. The current classification of viruses is primarily based on biologically meaningful criteria, which divide the viruses into orders, families, subfamilies, genera and species.

Virus classification is an extremely difficult task for many families. Traditionally, virus classification relied on properties such as virion morphology, genome organization, replication mechanism, serology, natural host range, mode of transmission, and pathogenicity. Yet viruses sharing the above properties can reveal tremendous differences at the genome level. For example, classification of many phages used to be based on the presence, structure, and length of a tail, and this approach has been shown not to correlate with genomic information, leading to a very difficult situation and hundreds of mis-classified phages.

With the continuous development of technology, high-throughput sequencing and metagenomics have fundamentally changed virology. The number of viral genome sequences in the public databases is increasing rapidly, and these sequences are playing a significant role in virus classification. The most commonly used sequence comparison methods include multiple sequence alignment and phylogenetic analysis. In recent years, many molecular classification methods have been developed, and one that has drawn more and more attention from virologists is pairwise sequence comparison (PASC). In this article, we briefly describe various sequence comparison methods, introduce the PASC tool, and compare it with other methods.

Sequence Comparison Methods

Most sequence comparison algorithms can be broadly categorized by the scope of their application on sequences (local vs. global), by the number of sequences involved (pairwise vs. multiple), and by the involvement of sequence or sub-sequence alignment. In bioinformatics, sequence alignment is an approach to arrange sequences of DNA, RNA, or protein. Alignment of biological sequences helps to identify regions of similarity that may be a consequence of functional, structural, and evolutionary relationships between the sequences. The growing number of sequence data demands more efficient and faster alignment algorithms. A variety of algorithms and programs are available to suit a wide range of problems requiring sequence alignments as parts of their solutions. Depending on the specifics of a problem, different types of algorithms and their combinations may be the optimal choice.

Each pairwise alignment can be viewed as an array of per-residue operations transforming one sequence to the other. These operations are substitutions (called matches and mismatches in nucleotide alignments), insertions and deletions. In pairwise sequence alignment methods, scientists focus on finding the best-matching piecewise alignments of two query sequences by comparing the two DNA or protein sequences and finding the largest similarity matches between them. By inserting gaps, the two sequences with the same or similar residues are aligned up and down, and the number of matching residues is statistically analyzed to get a similarity score, which serves as a basis for determining the relationship between the two sequences. A generalization of this concept to multiple alignments is possible. Given a set of sequences, an alignment is called optimal if it has the maximal score over all possible alignments. Optimal alignments are not necessarily unique, two or more alignments can be tied with the same score.

Local alignments are useful for dissimilar sequences that are suspected to contain regions of similarity between arbitrary parts of sequences within their larger sequence context. There are many applications involving local alignment, including the search for orthologous genes or conserved protein domains. The alignment generated using local algorithms is easy to analyze mathematically,

which allows the construction of tools to access the statistical significance of the alignment. The Smith–Waterman algorithm is a general local alignment method based on the dynamic programming scheme but with additional choices to start and end at any place. This algorithm runs in a time proportional to the product of the lengths of sequences. Since this is too slow for large-scale searches, many heuristics have been developed that allow typical queries to be matched to gigabase-sized sequence archives in seconds.

BLAST is one of the most widely used bioinformatics programs for sequence searching. The heuristic algorithm it uses is much faster than other approaches, such as calculating an optimal alignment. However, algorithms like BLAST are not suitable for all applications. Although they are capable of picking up segments of high similarity, fragments may be inconsistently placed in the resulting alignment, and some fragments may belong to more than one individual alignment. Additional post-processing steps are usually required to generate consistent sets of local alignments, which makes the use of local alignment complex in applications involving uniform computing of identities.

Global sequence alignment attempts to match as many characters as possible, from beginning to the end of both the sequences to find the best possible alignment. The Needleman-Wunsch algorithm is a general global alignment technique, which is based on the dynamic programming. Global alignment is mainly used to find closely related sequences and its running time is the same as that of the local alignment. Additionally, hybrid methods, known as semi-global, search for the best possible partial alignment of the two sequences. This is especially useful when the downstream part of one sequence overlaps with the upstream part of the other sequence. Another case where semi-global alignment is useful is when one sequence is short and the other is very long. In this case, the short sequence should be aligned globally but only a local alignment is required for the long sequence.

Multiple sequence alignment is an extension of pairwise alignment to incorporate more than two sequences at a time. However, they are often used for different purposes, such as detecting weak or scattered similarities on a set of sequences known to share a common structure or function. And thus, multiple sequence alignment is the basis for evolution analysis, consensus generation and structure prediction. Computing an optimal multiple sequence alignment is a computationally costly task, and most implementations use various heuristics to approximate the alignment in a reasonable time. So there is no available optimal solution for multiple sequence alignment and all methods are heuristics, such as progressive/hierarchical alignment, and iterative alignment. There are many multiple sequence alignment tools available: Clustal Omega, Clustal W, MAFFT, MUSCLE, T-Coffee, ProbCons, ProDA, MSAProbs, GLProbs, and so on.

Phylogenetic analysis is also a popular method for the comparison of more than two sequences. A phylogeny or evolutionary tree is a mathematical structure which is used to model the historical relationships between groups of organisms or sequences. The types of methods used to construct phylogenetic trees include distance-based methods (such as neighbor-joining), parsimony, maximum likelihood, and other probabilistic inference techniques. The most common distance-based methods utilize multiple sequence alignments to estimate the evolutionary distance between each pair of sequences and reconstruct the tree from the distances. Either protein sequences or DNA sequences can be used. Phylogenetic analysis was used to support the classification within the vast majority of virus families described in the ICTV Report on Virus Taxonomy. Nearly all methods of phylogenetic analysis share a number of fundamental assumptions, including that the sequences in the multiple sequence alignment are homologs. So it can also be used to the classification of a large group of distantly related viruses. For example, phages consist of many different families. Therefore, the conventional phylogenetic analysis that uses genomic sequences or individual protein sequences would not work for the classification of phages as a whole group. A ‘phage proteomic tree’ was developed to classify phages by the overall similarities of all protein sequences present in the phage genomes.

Alignment-free approaches to sequence comparison can be defined as any method of quantifying sequence similarity/dissimilarity that does not use or produce alignment (assignment of residue-residue correspondence) at any step of algorithm application. The alignment-free approaches are mathematically well founded in the fields of linear algebra, information theory, and statistical mechanics, and calculate pairwise measures of dissimilarity or distance between sequences. The rationale behind this method is that similar sequences share similar words/k-mers (subsequences of length k) and mathematical analysis of the k-mers’ occurrences can give a good relative measure of dissimilarity. For example, for two virus genome sequence, collect the common k-mers, calculate an array of numbers (by counting the number of times each particular k-mer appears within the sequences) and use this to quantify the dissimilarity between sequences. Conveniently, most of these measures can be directly used as an input into standard tree-building software, such as Phylip or MEGA.

Although phylogenetic analysis is well established as a tool for virus classification, it is usually computationally intensive and requires expertise to perform the analysis and interpret the results. For the alignment-free method, it will need more memory as the k-mer length increases. Furthermore, there are still many problems that need to be solved although the method has been applied in a few programs.

Sequence-Based Virus Classification Methods Review

With the advances in sequencing technologies, viral genome sequences in public databases are increasing rapidly. Many of these sequences are from metagenomics sequencing, and there is a lack of other information about the viruses associated with them. The traditional virus classification method, therefore cannot be applied to them. As a result, virus classification methods based on viral genome sequences have become popular and indeed essential, in recent years. Virus classification methods solely based on viral genome sequences have been accepted by the international community because they usually produce very similar results to those obtained by the traditional methods.

DEmARC

DEmARC (DivErsity pArtitioning by hieRarchical Clustering) is a quantitative, genetics-based framework for classification of a virus family. This approach performs a multiple alignment of conserved proteins within a virus family, and estimates the pairwise evolutionary distances (PEDs) by maximum likelihood inference. The distribution of PEDs is used to quantify genetic divergence and classify viruses at hierarchical levels in the family. DEmARC has been applied to the classification in the families *Picornaviridae*, *Filoviridae*, and *Mesoniviridae*. However, because DEmARC requires the manual selection of protein sequences and optimization of the results of multi-sequence comparison, it is not well suited for high-throughput applications.

Natural Vector

The Natural Vector method is an alignment-free tool to characterize and analyze virus sequences and classify viral genome sequences. This approach constructs a one-to-one correspondence between genome sequences based on a 12-dimensional genome space. The vector contains the quantities of each of the four bases, their mean positions in the sequence and the normalized variances of the position of each. The distance between two genome sequences is defined as the distance between their natural vectors. This recasts a variable length nucleic acid sequence as a fixed-length low dimensional vector and has been expanded to various similar methods such as 60-dimensional protein space and an 18-dimensional natural vector which also considered covariance between nucleotides. This method provides a very fast way to characterize phylogenetic relationships amongst viruses.

SDT

The Sequence Demarcation Tool (SDT) is a stand-alone program with graphical interface that allows classification of virus sequences based on sequence pairwise identity (a robust Needleman-Wunsch pairwise-alignment). It takes DNA or protein sequences as input and aligns each unique pair of sequences, calculates pairwise similarity scores, and displays a color-coding matrix for those scores. It also generates the corresponding plot and a text file containing the results of the analysis. SDT has been recently used for the reclassification of viruses in the family *Geminiviridae*, in the classification of viruses in the families *Circoviridae*, *Nanoviridae*, *Amalgaviridae*, *Smacoviridae*, and *Alphasatellitidae*. The virus sequences can be classified in batches if computational power is limited.

GRAViTy

Genome Relationships Applied to Virus Taxonomy (GRAViTy) is a stable algorithm that classifies viruses based only on sequences. GRAViTy first uses the known viral proteins and genome sequences to establish the Protein Profile Hidden Markov Model (PPHMM) and Genomic Organization Model (GOM) databases. Then, by calculating the distance between the PPHMM and GOM signatures of two virus sequences, the evolutionary distance between the two viruses can be inferred, and used for virus sequence classification. This approach can be effectively applied to the family-level classification of eukaryotic viruses.

ViPTree

ViPTree is a web-based classification tool to generate viral proteomic trees for the classification of viruses based on genome-wide similarities. Specifically, the proteomic tree is generated by BIONJ based on a genomic distance matrix computed from the normalized tBLASTx scores. Users can upload viral genomic sequences from genomics or metagenomics studies and generate proteomic trees using their own sequences and selecting reference viral genomes from the GenomeNet/Virus-Host database. For the uploaded viral genomes, ViPTree also provides visualization of comparative genomes and automatically annotates gene functions. In addition, they also developed a command line tool, ViPTreeGen, for viral proteomic tree generation.

ViCTree

ViCTree is a bioinformatics framework which can automatically synchronize with viral sequences in the National Center for Biotechnology Information (NCBI) GenBank and generates an interactive maximum-likelihood phylogenetic tree combined with distance data. Specifically, all known viral protein sequences available in GenBank are downloaded and compared with a curated set of seed protein sequences by BLAST. Significant matches are clustered by using CD-HIT with a user-specified identity threshold parameter, followed by multiple sequence alignment and RAxML maximum likelihood tree generation. ViCTree is capable of automatically building new phylogenies when novel viral species data are available in public databases. The built-in ViCTreeView visualization tool enables users to visualize and explore distances in the context of the tree in a web browser. This approach is currently set up for the *Herpesviridae*, *Parvoviridae*, and *Densoviridae* families but is flexible and can be adapted for any other virus family.

PASC Description and Application

PASC, one of the most widely used sequence-based tools for virus classification, was developed by NCBI. PASC provides a web interface ("See Relevant Websites section"), which makes it easy to navigate and perform analyses. For a given virus family/group, complete genome sequences are retrieved from the NCBI viral genomes collection, which includes both reference sequences and genome sequences of other members of the same species. These sequences, together with their NCBI taxonomy lineages, are stored in a database. The tool's database now stores pairwise identities for complete genomes/segments of 66 virus families/groups. Data in the system are updated daily to reflect changes in virus taxonomy and additions of new virus sequences to the public database.

The latest version of PASC uses a process of automated BLAST alignment to identify regions with significant sequence similarity. To be specific, two rounds of BLAST are performed on each pair of genome sequences. In the first round, the protein sequences of one genome, translated in six reading frames, are searched against the other genome using tblastn. The amino acid alignments in the tblastn results are converted back to nucleotide alignments. In the second BLAST round, pairwise blastn is carried out on the nucleotide sequences of the genomes. And then, a consistent set of hits are selected from the two sets of BLAST results, giving preference to higher identity hits and trimming overlaps out of lower identity hits, to generate a set of hits that do not overlap in any region on any genome. Pairwise identities are calculated as the total number of identical bases in local hits divided by the average length of the genome pair. Furthermore, the number of virus pairs at each percentage is plotted. A chart is plotted showing the distribution of identities between all members of the selected virus family or group, and can help determine the demarcations at strains, species, genera, and subfamilies level. If the two genomes belong to the same species according to their assignment in the NCBI taxonomic database, the pair is represented in green; if the two genomes belong to different species but belong to the same genus, it is yellow; and if they belong to different genera, it is peachy. Both linear and log scales are available for the y-axis (number of pairs). The gap sizes between the peaks indicate the possibility that these regions will be filled with sequences from novel viruses. The larger the gaps, the more likely that they are the true threshold to separate species/genera. When given a novel virus, PASC will begin to compute the alignments, or to extract them from the database if the query is a member of the group. At the end of the process, a user is presented with a list of closest matches. Matches can be selected to visualize their positions on the identity distribution chart. A new virus sequence can be tested with this system within a few minutes to suggest the taxonomic position of the virus in these families.

PASC has been successfully applied to many viral families. First, frequency distributions of pairwise genetic identity scores generated by PASC can be used to manually identify taxonomically optimal species or genus demarcation thresholds.

The mononegaviral family *Filoviridae* has eight members assigned to three genera and seven species. Previously, arbitrarily chosen filovirus genome sequence divergence values ($\approx 50\%$ for genera, $\approx 30\%$ for species) and phenotypic virus or virion characteristics were used to decide genus and species demarcation. After comparison of 152 distinct near-complete, coding-complete or complete filovirus genome sequences using PASC, optimal genus demarcation at the 55%–58% sequence diversity threshold range for genera and at the 23%–36% sequence diversity threshold range for species were revealed (Fig. 1). These thresholds do not change the current official filovirus classification, so they were implemented as filovirus taxon demarcation criteria that may solely be used for filovirus classification in case additional data are absent. In addition, PASC can be applied to multi-segmented viruses. For example, the mammalian arenavirus genome consists of two single-stranded ambisense RNA molecules. The two RNA segments are designated Small (S) and Large (L). PASC of arenavirus genomes and NP amino acid pairwise distances support the modification of the present classification. For the S segment, the pairwise nucleotide sequence

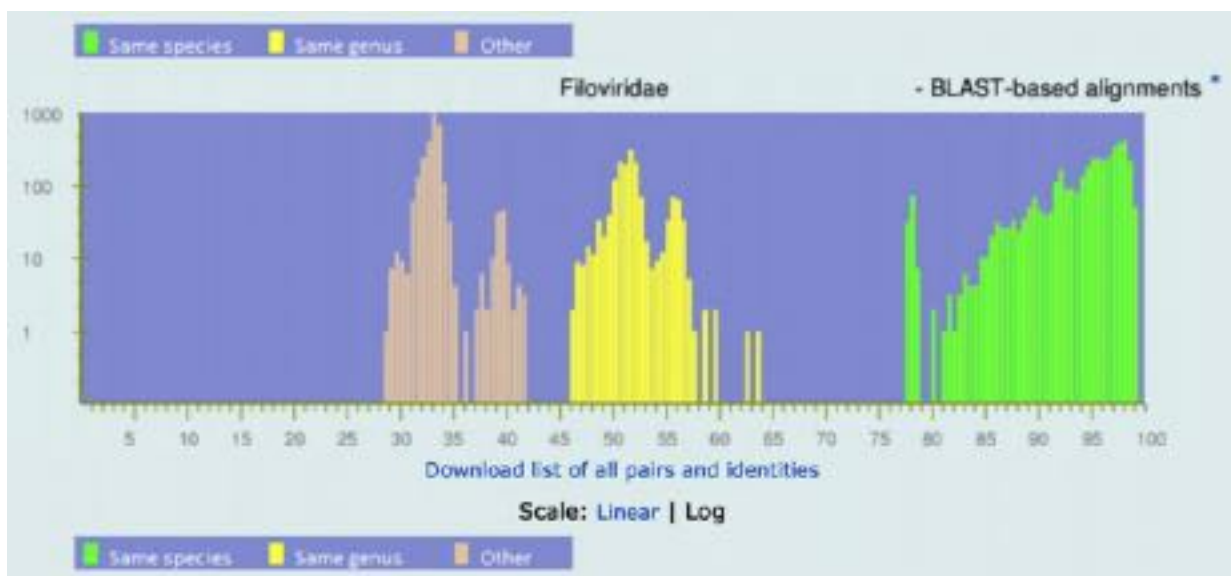


Fig. 1 Screenshot of PASC result in log scale after comparing filovirus genome sequences.

identities within the same proposed genus are higher than 40%, while those from different proposed genera are lower than 29%. PASC analysis indicates that an ideal genus separation cutoff range of 29%–40% sequence divergence for the S segment, and 30%–35% for the L segment. In order to introduce the fewest changes and minimal inconsistency with the current arenavirus classification scheme, nucleotide sequence identity > 80% in the S segment and > 76% identity in the L segment were chosen as values for viruses that should belong to the same species.

Newly sequenced viruses can be placed into the correct taxonomy group using PASC. Given a novel virus genome sequence, PASC compares it to a defined set of publicly available sequences and gives a list of matches within the family from the highest to the lowest based on BLAST identity. According to top pairwise identities and species or genus demarcation thresholds revealed by PASC, users can easily assign a novel sequence to the most appropriate genus, species or other taxonomic unit. Dot matrix and text views of pairwise alignment between genome sequences can also be accessed. Lymphocystis disease, which is provoked by the iridovirus lymphocystis disease virus (LCDV), affects marine and freshwater fish species worldwide. Direct sequencing of the virome of lymphocystis lesions from affected *S. aurata* fish was used to obtain the complete genome of a new LCDV-Sa species (GenBank accession number KX643370) that is the largest vertebrate iridovirus sequenced to date. Pairwise identities were calculated using PASC and it was found that LCDV-Sa shared only 54.67% identity with LCDV-C (NC_005902.1) and 38.95% identity with LCDV-1 (NC_001824.1) while LCDV-1 and LCDV-C shared 41.53% identity with each other. Accepted viral species within other genera of the *Iridoviridae* family show identity values above 85%. Therefore, LCDV-Sa represents a novel species of the genus *Lymphocystivirus*. Additionally, users can find novel genera with the help of PASC. Pteromalus puparum negative-strand RNA virus 1 (PpNSRV-1) is a novel negative-stranded RNA virus from a parasitoid wasp. It mediates the offspring sex ratio by decreasing female offspring numbers. PASC analysis of the PpNSRV-1 genome (Genbank accession number KX431032), reveals 13.8% identity to Soybean cyst nematode virus 1 (SbCNV-1; nyamiviral genus Socyvirus; GenBank accession number NC_024702), 12.1% to Midway virus (MIDWV; nyamiviral genus Nyavirus; GenBank accession number NC_012702), 12.3% to Nyamanini virus (NYMV; nyamiviral genus Nyavirus; GenBank accession number NC_012703), and 10.9% to Sierra Nevada virus (SNVV; nyamiviral genus Nyavirus; GenBank accession number NC_024376), with those viruses being the closest known relatives of PpNSRV-1. The 13.8% identity of the closest relative is below the 17% identity between nyavirus and socyvirus genomes, showing the need for a novel nyamiviral genus. The ICTV Nyamiviridae Study Group has agreed with the assessment.

Finally, PASC can also perform simulation of taxonomy changes. In other words, it allows users to test ideas for genus/species demarcation and see what taxonomy changes are needed using existing sequences to achieve user-selected demarcations. Even though NCBI uses the official ICTV taxonomy names whenever possible, there are times when GenBank sequences are assigned to incorrect virus names by submitters, which will lead to a mixture of colors in the peaks. For example, green or yellow bars are mixed with the dominating peach bars. To explore different demarcations, users can try different numbers for “Merge species above” and “Separate species below”, and for “Merge genera above” and “Separate genera below” as indicated at the lower left section of “Taxonomy changes:”, see for instance Fig. 2. In addition, PASC also lists recommended taxonomy changes needed for such a result, some of which are shown at the bottom of the webpage. Another taxonomy change simulation tool in PASC splits genomes into subgroups, which called “Split genomes below” (at the lower left of Fig. 2). After this is applied, users can see how genome groups are formed.

Advantages of PASC Compared to Other Methods

In recent years, many methods and tools have been applied to the classification of viruses (Table 1). Since for many viruses only sequence information is available, the tools and methods developed now aim to achieve rapid sequence-based classification and identification. These tools fall into three major categories: phylogenetic analyses such as ViCTree and ViPTree; alignment-based such as PASC and SDT; and alignment-free such as NVR and GRAVITY. PASC, one of the first such methods, was brought online in 2008. Its core classification algorithm and visualization method were updated in 2014.

Unlike classification methods based on the phenotypic properties of viruses, PASC is a quantitative tool. For those virus families that are suitable for PASC analysis, demarcations can be easily determined and new viruses can be clearly placed into the correct taxonomy. However, there are times when PASC alone cannot give a definite classification, and other viral properties have to be considered.

The most commonly used sequence-based virus classification tool is phylogenetic analysis. Compared to this approach, PASC is less computationally intensive and can be easily updated with new sequence data. In addition, PASC results are relatively easy to interpret, and this could be done potentially by a computer program without human intervention. It would therefore be possible to set up an automatic system for high throughput classification.

PASC was greatly improved by using the BLAST-based alignment method. Previously, PASC used a pairwise global alignment algorithm to calculate viral genome similarity. Although this method was applicable to some virus families/groups, such as papillomaviruses and *potyviruses*, the results were not very good for others. Circular genome sequences are deposited with inconsistent starting and ending coordinates, which will cause trouble for the global alignment approach. The BLAST-based alignment (combination of blastn and tblastn) method selects the better of the two alignments for the same region in the genomes, which effectively applies the most appropriate blast program automatically for protein coding and non-coding regions. This approach can find all possible regions of similarity in the genome and improved the results for almost all viral families/groups.

The PASC tool is one of NCBI's web tools synchronized with NCBI's viral genome collection and taxonomy database that are updated regularly. It runs very fast, and results can usually be obtained within minutes. The tool is online, so there is no software



Fig. 2 Frequency distribution of pairwise identities from the complete genome sequence comparison of 737 viruses in Rhabdoviridae, and the simulation of taxonomy changes using proposed species and genus demarcations. A good taxonomy can be achieved when 75% is used as the species demarcation. Black arrowheads on the x-axis of the upper and lower plots show the difference after doing simulation of taxonomy changes.

Table 1 Comparison of sequence-based virus classification methods

Methods	Core Algorithm	Web/Package	Resource	LUD ^a	Version	Citations ^b
PASC	BLAST	web	RefSeq	2014/8/14	2.0	79
DEmARC	Pairwise Evolutionary Distances	package	public data ^c	2012/1/25	1.0	54
NV	Alignment-free Natural Vector Method	web	VirusDB	2014/11/1	1.0	62
SDT	Needleman-Wunsch	package	public data ^c	2014/8/27	1.2	398
GRAViTy	Protein Profile Hidden Markov Model Genomic Organization Model	package	public data ^c	2018/2/20	1.0	11
ViPTree	Proteomic Tree	web	Virus-Host DB	2018/12/25	1.7	12
ViCTree	Maximum Likelihood Phylogeny	package	GenBank	2018/10/7	1.0	0

^aLast Update Date (LUD).

^bData from Google Scholar on March 31, 2019.

^cData are downloaded from public database (need manual update).

to download/install, no parameters to set, and everybody uses the same algorithm and same dataset, so that the results are directly comparable. PASC is one of the more important sequence-based alignment virus classification tools and provides support to the ICTV study groups for the families of *Polyomaviridae*, *Arenaviridae*, *Bornaviridae*, *Arteriviridae*, and *Filoviridae*.

Limitations of PASC

Although PASC has been updated in 2014 and applied successfully to more virus families, the approach has some limitations.

First of all, PASC is not suitable for virus families whose current classification is largely based on virus morphologies, such as phages in the families of *Siphoviridae* and *Podoviridae*.

Second, although demarcation values were established in PASC for several virus families, for most virus families no cutoff values in sequence identity percentages are currently suggested to separate species and genera. The separation points between

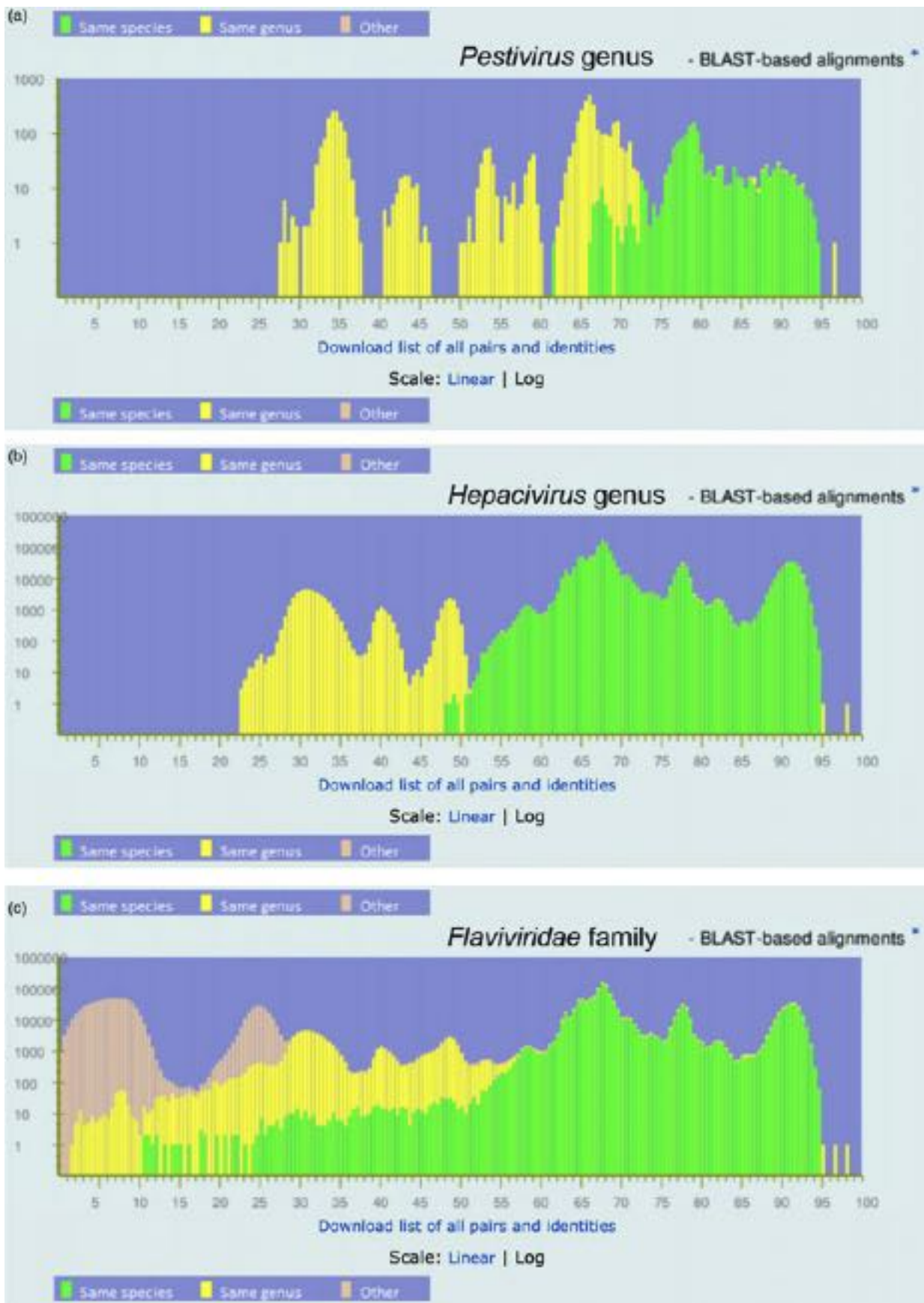


Fig. 3 Limitations of PASC in virus families with highly diverse sequences. PASC results in the *Pestivirus* genus (a), the *Hepacivirus* genus (b), and the *Flaviviridae* family (c).

different classification levels need to be manually selected, which would result in the decrease of efficiency and accuracy. When the distribution suggests that one cutoff could be selected as demarcation in PASC, other criteria (e.g., host) should be considered to select the suitable one.

Furthermore, the use of the whole genome in PASC may not work well for virus families with highly diverse sequences. One assumption of PASC is that the taxonomic thresholds of genera and species are the same among different evolutionary lineages of a family. Thus, if the threshold (BLAST score) for species assignment in one genus is 85, then it would also be the threshold for other genera in the same family. However, in many virus families, uniform threshold definitions cannot be applied. For example, the *Pestivirus* genus of the virus family *Flaviviridae* shows mean species threshold of 74 (range 72–76) between nucleotide sequences of the genomes (Fig. 3(a)). In contrast for the *Hepacivirus* genus, it is 51 (range 48–53) (Fig. 3(b)). When applied to the entire *Flaviviridae* family, PASC shows a broad range of pairwise distances, with no clear divide between inter- and intra-species pairwise distances (Fig. 3(c)).

Conclusion

The current standard method of viral classification has been challenged by the large amount of data generated by high throughput sequencing. There is an urgent need to develop novel approaches to classify viruses using their sequences only. Various sequence-based methods currently exist, amongst which PASC is frequently used due to its good performance and ease of use. PASC is an online tool, and there is no need for users to install any software/package, to download datasets or to set parameters. It therefore avoids the inconsistency of results between different users who choose different protocols to calculate the pairwise identities. In addition, it runs quickly and a new virus sequence can usually be tested within a few minutes to suggest optimal taxonomic position in selected family. The tool has already been successfully applied to a number of virus families and increasingly species/genus demarcation criteria revealed by PASC are being adopted by ICTV Study Groups. However, PASC, and all sequence-based virus classification methods, have shortcomings. For example, PASC may not work well for virus families with highly diverse sequences. Hence, further improvement or a combination of different algorithms need to be investigated in the future.

Further Reading

- Bao, Y., Chetvernin, V., Tatusova, T., 2014. Improvements to pairwise sequence comparison (PASC): A genome-based web tool for virus classification. *Archives of Virology* 159, 3293–3304.
- Edgar, R.C., Batzoglou, S., 2006. Multiple sequence alignment. *Current Opinion in Structural Biology* 16, 368–373.
- Felsenstein, J., 2004. *Inferring Phylogeny*. Sunderland, MA: Sinauer Associates.
- Gusfield, D., 1997. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge: Cambridge University Press.
- King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., 2011. *Virus Taxonomy-ninth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier/Academic Press.
- Page, R.D., Holmes, E.C., 1998. *Molecular Evolution: A Phylogenetic Approach*. Oxford: Blackwell Science.
- Simmonds, P., Adams, M.J., Benko, M., *et al.*, 2017. Consensus statement: Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology* 15, 161–168.
- van Regenmortel, M.H.V., 2007. Virus species and virus identification: Past and current controversies. *Infection, Genetics and Evolution* 7, 133–144.
- van Regenmortel, M.H.V., Bishop, D.H., Fauquet, C.M., *et al.*, 1997. Guidelines to the demarcation of virus species. *Archives of Virology* 142, 1505–1518.
- Zielezinski, A., Vinga, S., Almeida, J., Karlowski, W.M., 2017. Alignment-free sequence comparison: Benefits, applications, and tools. *Genome Biology* 18, 186.

Relevant Websites

<https://www.ncbi.nlm.nih.gov/sutils/pasc/>
 PASC.
 NCBI.
 NIH.

Computational Analysis of Recombination in Viral Nucleotide Sequences

Miguel Arenas, Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Spain and CINBIO (Biomedical Research Center), University of Vigo, Vigo, Spain

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Genetic recombination consists of an exchange of genetic fragments that shuffles mutations to generate genetic variants upon which natural selection operates. Different types of recombination have been described (e.g., [Scheel et al., 2013](#)). Homologous recombination occurs in the same location in both parental strands and it is often observed in viruses (i.e., in movements of genetic material by horizontal gene transfer, HGT). By contrast, non-homologous recombination exchanges genetic fragments placed at different locations and can lead to aberrant genetic data ([Galli and Bukh, 2014](#)) but also it can be useful for virus populations, for example, providing new enzyme cleavage sites that increase virulence ([Godoy et al., 2014](#)). Some viruses (i.e., influenza) with segmented genomes can also exchange complete genome segments in a recombination-based reassortment process ([McDonald et al., 2016](#)).

Recombination occurs in a large variety of organisms, predominantly in bacteria and viruses ([Perez-Losada et al., 2015](#); [Simon-Loriere and Holmes, 2011](#)). Concerning the latter, the recombination rate broadly varies among known viruses ([Perez-Losada et al., 2015](#)). Recombination is particularly frequent in dsDNA viruses where it is involved in replication and DNA repair and, in some RNA viruses (especially in retroviruses such as HIV) where the recombination rate per nucleotide can surpass the mutation rate ([Jetzt et al. 2000](#)) it can purge deleterious variants and accelerate the emergence of advantageous variants.

Recombination has been associated with the expansion of viral populations, emergence of outbreaks and new viruses, increase of fitness in new environments, modification of transmission specificities, increase of virulence and pathogenesis, escape from the immune system and antiviral therapies, among others ([Perez-Losada et al., 2015](#); [Simon-Loriere and Holmes, 2011](#); [Arenas et al., 2016](#)). The evolutionary innovations required to acquire these achievements in viral populations would be slow or inaccessible by only the mutation process, thus favouring the establishment of genetic recombination. At the inter-host level, understanding recombination can produce epidemiologically-relevant circulating recombinant forms (CRFs) that may become stable over time ([Lihana et al., 2012](#)) (a list of common HIV-1 CRFs is referenced in the section Relevant Websites) and that can favor failure of vaccines against viruses such as the HIV ([Barouch et al., 2013](#)). At the intra-host level, recombination can increase viral diversity and shuffle resistance mutations, which overall increase the diversity of drug resistance genotypes, allowing viral adaptation to the immune system and antiviral therapies ([Charpentier et al., 2006](#)). In this concern, modeling recombination, together with the viral life cycle and viral fitness, could be useful to improve the design of antiviral therapies. However, models that accurately mimic Virus Evolution are still challenging due to the complexity of parameterizing these processes accounting for a temporal variation of the environment ([Gog et al., 2015](#)).

The relevance and impact of recombination in Virus Evolution is clear. Consequently, evolutionary biologists and molecular virologists require the analysis of recombination, at qualitative and quantitative levels, from genetic data. For this reason, a variety of computational methods have been developed and implemented in currently available frameworks, to analyze recombination from genetic sequences. The influence of recombination on phylogenetic tree inference and derived evolutionary analyses, including guidelines on how recombination should be considered to avoid estimation biases will also be described and illustrative case studies of viral genetic data that evolved with recombination presented.

Computational Estimation of Recombination

The large variety of consequences derived from recombination in viruses leads to the requirement of accurate analyses of recombination, including the estimation of recombination rates and recombination breakpoints. Currently, diverse methods (some of them implemented in computational frameworks) can be used to perform these analyses. These methods present differences concerning their performance ([Posada and Crandall, 2001](#)) and so selecting a particular method can be complex. The choice is usually based on the kind of analysis desired (i.e., detect presence/absence of recombination, map recombination breakpoints, quantify constant or heterogeneous recombination rate along the genetic sequences), genetic marker or available computational resources ([Martín et al., 2011](#)). An updated list of well-established computational frameworks to detect recombination breakpoints and estimate recombination rates from genetic sequences (together with their main features) is shown in [Table 1](#). It is recommended to explore the comprehensive list of recombination analysis frameworks provided by David Robertson's Lab (section Relevant Websites). In the following subsections the most used methods and frameworks to analyze recombination in genetic sequences will be summarized.

Rapid Tests to Detect the Presence of Recombination

The presence or absence of recombination can be evaluated with statistical tests ([Wiuf et al., 2001](#)) that are computationally rapid. In general, these tests provide accurate predictions in data with large genetic diversity and recombination ([Posada, 2002](#)) (particularities that are frequently observed in viruses) but there are exceptions. While tests based on homoplasy are more accurate

Table 1 Commonly used frameworks to estimate recombination rate and recombination breakpoints in a multiple alignment of genetic sequences. For each framework, the table indicates the underlining statistical approach (ML and ABC refer to maximum-likelihood and approximate Bayesian computation, respectively), type of genetic marker of the input dataset (SNP refers to single nucleotide polymorphism), estimation of homogeneous and/or heterogeneous recombination rate along the sequences, prediction of recombination breakpoints, operating systems where the framework can run and the corresponding references. In italic frameworks are highlighted that have been frequently used or that are recommended based on the author's experience. Additional software to analyze recombination is shown in a web page provided by David Robertson's Lab (section Relevant Websites)

Program	Statistical approach	Genetic marker	Recombination rate along sequences		Recombination breakpoints	Operating systems	References
			Homogeneous	Heterogeneous			
<i>RDP</i>	Bayesian	DNA	Yes	Yes	Yes	Windows	(Martin <i>et al.</i> , 2015)
<i>OmegaMap</i>	Bayesian	Coding DNA	Yes	Yes ^a	No	Linux, Mac, Windows	(Wilson and McVean, 2006)
<i>LDhat</i>	Bayesian	DNA, SNP	Yes	Yes	No	Linux	(McVean <i>et al.</i> , 2004)
<i>PhyML_Multi</i>	Bayesian ^b	DNA	No	No	Yes	Linux, Mac	(Boussau <i>et al.</i> , 2009)
<i>jpHMM</i>	Bayesian ^b	DNA ^c	No	No	Yes	Linux, Online ^d	(Schultz <i>et al.</i> , 2009)
<i>DualBrothers</i>	Bayesian	DNA	No	No	Yes	Linux, Java	(Minin <i>et al.</i> , 2005)
<i>LDhelmet</i>	Bayesian	DNA, SNP	Yes	Yes	No	Linux	(Chan <i>et al.</i> , 2012)
<i>Lamarc</i>	Bayesian, ML	DNA	Yes	No	No	Linux	(Kuhner, 2006)
<i>TOPALI</i>	Bayesian, ML	DNA	No	No	Yes	Linux, Mac, Windows	(Milne <i>et al.</i> , 2009)
<i>Hyphy/Datamonkey</i>	ML	DNA	No	No	Yes	Linux, Mac, Windows, Online ^d	(Kosakovsky Pond <i>et al.</i> , 2005; Weaver <i>et al.</i> , 2018)
<i>Recco</i>	Cost minimization algorithm	DNA	No	No	Yes	Linux, Mac, Windows	(Maydt and Lengauer, 2006)
<i>CodABC</i>	ABC	Coding DNA	Yes	No	No	Linux	(Arenas <i>et al.</i> , 2015; Lopes <i>et al.</i> , 2014)

^aIt can estimate recombination rates for user-specified partitions.

^bIt is based on a Hidden Markov Model (HMM).

^cIt can only be used to study HBV and HIV-1 data.

^dThe link to the online web sever is given in the section Relevant Websites.

analyzing data with low genetic diversity, other tests [like those based on the maximum chi-square and the tests implemented in the framework *RDP* (Martin *et al.*, 2015)], are more confident in data presenting high genetic diversity (Bruen *et al.*, 2006). An interesting test is the pairwise homoplasy index (PHI), implemented in the *PhiPack* software (Bruen *et al.*, 2006), because it can distinguish recurrent mutation from recombination. Because of the different performance of these recombination tests, it is recommended to use several tests and make conclusions from the overall findings (Posada, 2002). In this concern, well-established frameworks implementing multiple recombination tests are *RDP* and *PhiPack*.

Additionally, some recombination tests have been developed considering the bias that recombination can make on the phylogenetic tree reconstruction (Schierup and Hein, 2000a). In particular, a dataset of genetic sequences presenting recombination can be split into fragments that have evolved under different evolutionary histories (Box 1 and Fig. 1), showing a phylogenetic tree incongruence that can be statistically quantified. Phylogenetic incongruence tests (e.g., Strimmer and Rambaut 2002) have been developed using this peculiarity and can be considered as a complement of the previously presented recombination tests. A limitation of phylogenetic incongruence tests is that recombination does not always lead to phylogenetic incongruence [the recombinant fragments could share a most recent common ancestor (MRCA) and could present a similar evolutionary history], which can result in false negatives.

Identification of Recombination Breakpoints

Generally, recombination events are heterogeneously distributed along genomes. Certain regions of a genome (recombination hotspot regions) can present a high frequency of recombination respect to the neutral expectation while other regions (recombination coldspot regions) can display lack of recombination with respect to the neutral expectation (Smyth *et al.*, 2014). In this concern, the identification of recombination breakpoints has been used in viruses to map recombination hotspots (e.g., Dalmon *et al.*, 2017) but also to characterize recombination forms (e.g., Archer *et al.*, 2008), detect adaptive recombination (e.g., Arenas *et al.*, 2016; Monjane *et al.*, 2011), understand molecular mechanisms of recombination (Perez-Losada *et al.*, 2015) and perform evolutionary analyses accounting for recombination (Box 1).

Most of the methods to detect recombination breakpoints are based on the statistical quantification of a particular property (i.e., phylogenetic tree) that differs at both sides of the recombination breakpoint. Basically, these methods explore the genetic location (breakpoint) with the highest variation of the cited property along the sequences (Martin *et al.*, 2011). The searching for recombination breakpoints along the sequences was originally explored with sliding windows where the breakpoint is moved along the sequences and it is statistically evaluated at every location (e.g., Holmes *et al.*, 1999), but more recent methods optimized the procedure by considering a probabilistic/heuristic search along the sequences (e.g., Westesson and Holmes, 2009).

Box 1. The influence of recombination on phylogenetic tree reconstruction and inferences derived

The genetic segments involved in a recombination event could have evolved under different evolutionary processes and present different evolutionary histories as phylogenetic trees (Fig. 1). Therefore, a recombination event can result in two phylogenetic trees (one for each fragment), which may differ among them (phylogenetic incongruence). The combination of these phylogenetic trees produces a phylogenetic network that is traditionally called an ancestral recombination graph (ARG) (Fig. 1) (Arenas, 2013; Griffiths and Marjoram, 1997). Consequently, the reconstruction of a single phylogenetic tree from genetic sequences that suffered recombination can be biased resulting in a phylogenetic tree with an incorrect topology and branch lengths (Schierup and Hein, 2000a).

If the phylogenetic tree is biased, subsequent evolutionary analyses could also be biased. For example, the estimation of the nonsynonymous (dN) to synonymous (dS) substitution rates ratio (dN/dS), which is a well-established measure of selection in protein-coding sequences (Hurst, 2002) and that is frequently used to study Virus Evolution at the molecular level (e.g., Poon *et al.*, 2007; Arenas, 2015), can be biased if recombination is ignored. In particular, this bias consists of estimating false positively selected sites (PSS) (Arenas and Posada, 2014; Anisimova *et al.*, 2003). Another bias caused by recombination is related to the ancestral sequence reconstruction (ASR). Incorrect ancestral sequences of the internal nodes of a phylogenetic tree can be inferred in data that suffered recombination (Arenas and Posada, 2010b). The ASR error increases with the recombination rate and the genetic diversity (Arenas and Posada, 2010b). Other evolutionary analyses can also be affected by recombination. For example, ignored recombination can lead to a loss of the molecular clock, apparent homoplasies and false substitution rate heterogeneity (Schierup and Hein, 2000a,b; Arenas *et al.*, 2018). Altogether, if recombination is found [i.e., recombination breakpoints are detected and delimit recombinant fragments], one should perform the evolutionary analyses (i.e., dN/dS estimation or ASR) by one of the following procedures. (1) Separately analyze each recombinant fragment by using the phylogenetic tree inferred for the corresponding genetic fragment (e.g., Perez-Losada *et al.*, 2011; Perez-Losada *et al.*, 2009; Arenas and Posada, 2010b). (2) Apply an analytical method that performs the estimation accounting for recombination [i.e., for dN/dS see (Arenas *et al.*, 2015; Lopes *et al.*, 2014; Wilson and McVean, 2006) or for the reconstruction of a full phylogenetic history (i.e., recombination network) see (Huson, 1998)].

Several well-established frameworks to detect recombination breakpoints are presented in Table 1. RDP is one of the most used frameworks to perform this task probably because it implements a variety of algorithms and includes a graphical user interface (GUI). Another well-established framework is Hyphy (Pond *et al.*, 2005) [together with its online web server Datamonkey (Weaver *et al.*, 2018)] through the algorithm GARD (Pond *et al.*, 2006).

Estimation of Recombination Rates

The rate at which recombination events occur during the evolutionary history of a genetic dataset can also be estimated. Instead of estimating the recombination rate per generation per site (r), most methods estimate the population recombination rate $\rho = 2nNrL$, where $n = 1$ or 2 for haploid or diploid populations, respectively, N is the effective population size and L is the sequence length in nucleotides. Since ρ depends on N , it must be carefully interpreted (i.e., when comparing ρ estimated in two populations with different sizes or under variable population sizes) (e.g., Kuhner *et al.*, 2000). The estimation of the recombination rate, either at the molecular or population level, has been developed using different statistical approaches, including maximum likelihood (ML) (e.g., Kuhner, 2006), Bayesian (e.g., Kuhner, 2006; Wilson and McVean, 2006) and approximate Bayesian computation (ABC) (Lopes *et al.*, 2014), among others (Martin *et al.*, 2011). A variety of software implements the estimation of the recombination rate using these approaches (Table 1). Some well-established frameworks to perform this estimation are LDHAT (McVean *et al.*, 2004), Lamarc (Kuhner, 2006), RDP and OmegaMap (Wilson and McVean, 2006). The program LDHAT implements the coalescent composite likelihood method developed by Hudson (2001) into a Bayesian approach and can estimate variable recombination rate along sequences under different genetic markers. Another traditionally used framework is Lamarc, which implements ML and Bayesian approaches to estimate a population recombination rate that is assumed constant along the sequences. Lamarc can estimate additional evolutionary parameters such as the population growth rate and the migration rate, which can be useful to facilitate the understanding of the evolutionary process. However, both LDHAT and Lamarc can be problematic in reaching convergence between independent runs, especially with large datasets. The software RDP has been frequently updated, it is quite robust in reaching convergence between independent runs and includes a user-friendly GUI for Windows. Today it seems the most widely used framework to analyze recombination (more than 5000 citations from its four versions) and it has been widely used to analyze genetic data in diverse viruses (e.g., Faria *et al.*, 2017; Castelhanos *et al.*, 2017). Another very useful framework is OmegaMap. It is based on the product of approximating conditionals (PAC) likelihood and estimates recombination, dN/dS and substitution rates that can vary among user-specified partitions of protein-coding sequences. Finally, CodABC (Arenas *et al.*, 2015) is a framework developed by this author and others to accurately estimate recombination, substitution and dN/dS rates from protein-coding sequences and has already been applied to analyze a variety of HIV-1 datasets (Lopes *et al.*, 2014). It is based on ABC and its goal is derived from considering interactions between different evolutionary parameters (i.e., dN/dS is estimated accounting for recombination or the recombination rate is accurately estimated in data with low genetic diversity, outperforming other methods). On the other hand, ABC requires a large number of computer simulations [which are computationally intensive for coding data due to the large size of the codon exchangeability matrix (61×61 , excluding stop codons) (Arenas and Posada, 2012)] and this results in long computer times. Conveniently, CodABC can run in parallel on a multiprocessor machine allowing a reduction in computational time.

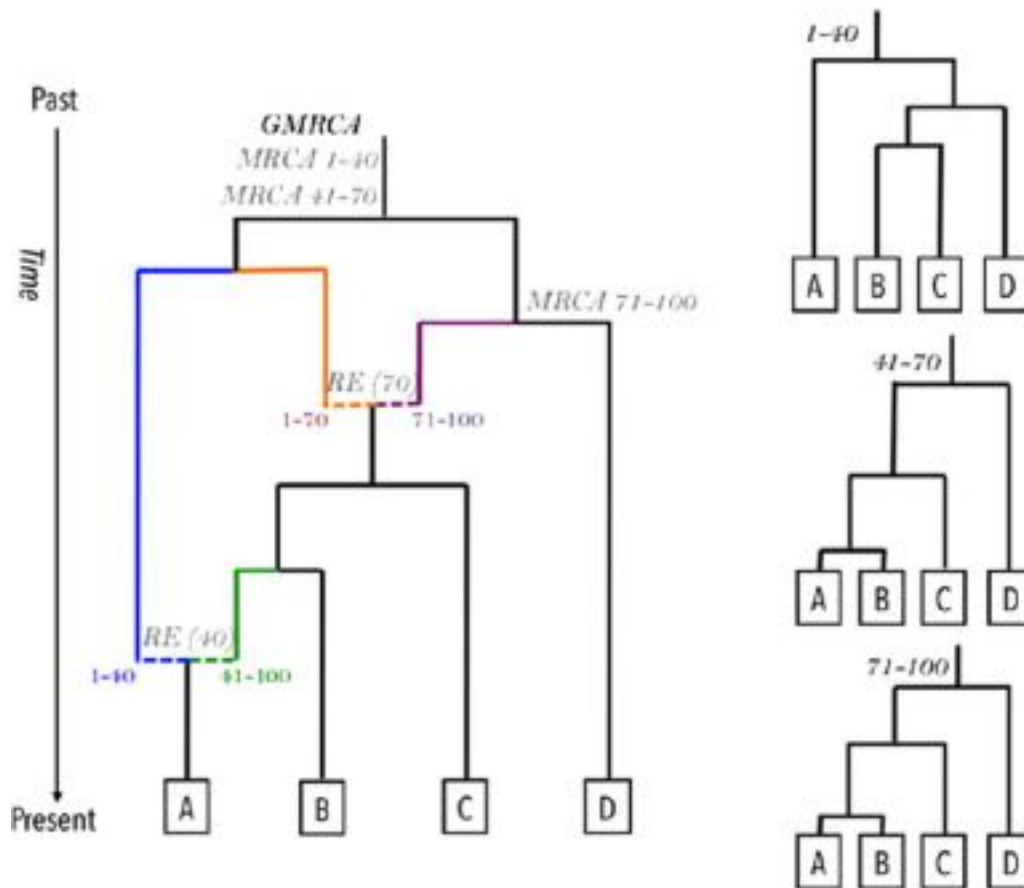


Fig. 1 Illustrative example of an ancestral recombination graph based on two recombination breakpoints. The figure shows an ancestral recombination graph (ARG) for a sample of 4 sequences (A, B, C, D) with 100 sites (i.e., nucleotides). Two recombination events (RE, dashed line) occurred with breakpoints at position 40 and 70, leading to three recombinant fragments 1–40, 41–70 and 71–100. Going backwards in time, a recombination event generates two lineages each (shown with different colors), for every recombinant fragment. Each recombinant fragment presents a specific phylogenetic tree (shown on the right) that can be built from the ARG by following the nodes and branches containing the corresponding fragment until reaching its most recent common ancestor (MRCA). The phylogenetic trees of the recombinant fragments can differ in topology, branch lengths and MRCA. The oldest MRCA is the root of the network and it is usually referred to as the grand most recent common ancestor (GMRCA).

Reconstruction of Phylogenetic Recombination Networks

The genetic regions on both sides of a recombination breakpoint can present different evolutionary histories (Box 1 and Fig. 1). Consequently, the inference of the evolutionary history of a genetic dataset that has undergone recombination should not be done by reconstructing a single phylogenetic tree (Mallo *et al.*, 2016). Instead, one should reconstruct a phylogenetic tree for each recombinant fragment or a phylogenetic network (the ARG) where each recombination event constitutes a reticulate node (Fig. 1). The ARG was originally simulated from the coalescent with recombination (Hudson and Kaplan 1988) and this modeling allowed its implementation in likelihood functions. However, calculating a likelihood function that considers recombination requires huge computational costs due to the large number of ARGs that should be explored, and thus leads to the use of methods (i.e., composite-likelihood models) to simplify the likelihood function (McVean *et al.*, 2002). The inference of the ARG has been implemented in several evolutionary frameworks. The software BEAST (Bouckaert *et al.* 2019) estimates the ARG together with other population genetic parameters under a Bayesian approach. Another software program to infer recombination networks is *SplitsTree* (Huson 1998), which has been frequently applied to analyze the evolutionary history of viruses (e.g., Castelhamo *et al.*, 2017). This tool includes several algorithms (based on traditional genetic distances) to efficiently infer phylogenetic networks and it includes a user-friendly GUI.

Analyzing Recombination in Sequences Derived From Next Generation Sequencing

Next generation sequencing (NGS, also known as high-throughput sequencing) allows an in-depth study of the diversity and structure of genomes, including virus genomes (e.g., Perez-Losada *et al.*, 2018; Marston *et al.*, 2013). However, the evaluation of recombination in sequences derived from NGS should be carefully performed because the assembly of reads may result in artifacts that create false positive recombination (Prosperi *et al.*, 2011; Zanini *et al.*, 2017). In order to account for this possible source of

error, a few complex strategies using experimental (e.g., [Laureau et al., 2016](#)) and computational (e.g., [Zhang, 2013](#)) methods have been proposed to estimate recombination from NGS data. However, methods to accurately estimate recombination from NGS data, including a proper evaluation and software implementation, are still demanded.

Illustrative Case Studies of Analysis of Recombination in Virus Genetic Data

The following case studies focus on why and how recombination was analyzed:

Recombination in the Design of HIV-1 Centralized Vaccines

Molecular evolution can allow viruses to escape from vaccine mediated immunity. In order to take into account this aspect, centralized sequences from HIV-1 have been explored to develop vaccines (centralized vaccines) ([McBurney and Ross, 2007](#)). These centralized sequences include consensus (CON, based on the abundance of nucleotide states) (e.g., [Ellenberger et al., 2002](#)), ancestral (ANC, sequence of the root of a phylogenetic tree) (e.g., [Doria-Rose et al., 2005](#)) and center-of-tree (COT, ancestral sequence at the place of a phylogenetic tree with the smallest evolutionary distance to the tip nodes) (e.g., [Rolland et al., 2007](#)) sequences. HIV-1 vaccines based on these centralized sequences were only partially effective in producing T-cell responses (e.g., [Frahm et al., 2008](#)). One of the possible causes of the incomplete immune response could be the computational inference of the centralized sequences ([Arenas and Posada, 2010a](#)). In particular, studies developing centralized HIV-1 vaccines inferred phylogenetic trees (required to obtain ANC and COT sequences) ignoring recombination, despite recombination being commonly observed in this virus ([Jetzt et al., 2000](#)). Ignoring recombination (if it is present) can lead to incorrect phylogenetic tree reconstructions ([Box 1](#)). In a previous study it was found that ancestral sequences inferred ignoring recombination (when it is present) display sequence errors that lead to HIV-1 epitopes and N-glycosylation sites [fundamental for the protein activity and predicted with the tool *ELF* (Epitope location finder), referenced in the section Relevant Websites] different to those derived from considering recombination ([Arenas and Posada, 2010b](#)). Altogether, future studies inferring centralized sequences should take into account recombination ([Box 1](#)).

The Influence of Recombination on HIV-1 Fitness Recovery

The specific influence of recombination on viral fitness has not been totally clear. Recombination can generate, but also break-down, better adapted variants ([Hadany and Beker, 2003](#)). While some studies found that variants derived from recombination overall present a decrease of viral fitness (e.g., [Bretscher et al., 2004](#)), other studies indicated the opposite (e.g., [Moradigaravand et al., 2014](#)). The influence of recombination on viral fitness was investigated by monitoring the fitness of several HIV-1 biological clones during an in vitro evolutionary process that included a large number of population plaque-to-plaque passages ([Arenas et al., 2016](#)). It was found that the fitness increase coincided with the increase of the recombination rate (estimated with *OmegaMap*), but not for all the studied clones. The conclusion drawn was that the fitness increase was driven by additional processes. Firstly, after recombination some mutations should have occurred to optimize the recombinant variant and further, a new recombinant variant requires enough frequency in the viral population to present a detectable effect on fitness ([Iglesias-Sanchez and Lopez-Galindez, 2002](#)). In addition, some clones increased fitness without requiring recombination. Altogether, the fitness effects of mutation and recombination events seemed to follow a heterogeneous and complex fitness landscape ([Lorenzo-Redondo et al., 2014](#)). When moving through the fitness landscape, a mutation event could lead to a small fitness variation (as a tuning) while a recombination event could imply a big change (i.e., generating a variant with a totally different fitness). In other words, the need for either mutation or recombination events could be related with the shape of the fitness landscape.

Recombination Among Hepatitis B Virus Genotypes

The Hepatitis B virus (HBV) presents a large diversity among its 10 so far detected genotypes (A-J). Interestingly, this virus has shown extensive recombination between genotypes (e.g., [Araujo, 2015](#)) where most of the recombinant forms (60%) include B/C and A/D hybrids [other recombinant forms include A/B/C, A/C, A/C/G, A/D, A/E, A/G, B/C/U (U = unknown genotype), C/F, C/G, C/J, D/E, D/F and F/G] ([Araujo, 2015](#)). The frequency of recombination was evaluated (through the recombination rate estimated with the program *RDP*) within each HBV genotype in more than 4700 HBV full genome sequences ([Castelhamo et al., 2017](#)). It was found that the recombination rate varied among HBV genotypes (genotypes B, C and E presented high levels of recombination while genotype G presented negligible recombination) and these differences could be explained by the epidemic and genetic characteristics of every genotype ([Castelhamo et al., 2017](#)). A recombination phylogenetic network was also inferred (with the program *SplitsTree*) showing the evolutionary relationships among HBV genotypes. The phylogenetic network could distinguish every genotype into a particular group and showed connections between genotypes. The network also presented recombination nodes within and among genotypes that quantitatively correlated with the estimated recombination rates (i.e., genotypes with high recombination rate presented a large number of reticulate nodes and vice versa).

Analyzing Adaptation of HIV-1 gp120 Under Recombination

The creation of an effective vaccine against HIV still remains as a challenging task (McMichael *et al.*, 2013). One of the multiple attempts made in the development of a vaccine against HIV consisted of a recombinant envelope glycoprotein subunit 120 (rgp120) vaccine (Flynn *et al.*, 2005). The vaccine was evaluated in more than 5000 volunteers and, unfortunately, it did not prevent HIV-1 acquisition. Interestingly, the results varied among volunteers according to ethnic groups [i.e., nonwhite volunteers showed a higher prevention of infection respect to other ethnic groups (further details in Flynn *et al.*, 2005)] and a possible cause could be differences in the immune system among these groups. This hypothesis was explored by analyzing the molecular adaptation (through dN/dS) of the gp120 viruses in hosts belonging to different ethnic groups (Perez-Losada *et al.*, 2009). Although this analysis should be straightforward by directly estimating dN/dS , the presence of recombination in the data was detected with the framework *HYPHY* (through its algorithm *GARD*). Note that recombination can bias the phylogenetic tree reconstruction and inferences derived, including the estimation of dN/dS (Box 1). Consequently, a methodology to estimate dN/dS was applied that considers the recombination breakpoints identified with *GARD*. In particular, the recombination fragments delimited by the recombination breakpoints were separately analyzed by inferring a phylogenetic tree specific for each fragment and by estimating dN/dS for each fragment accounting for the corresponding phylogenetic tree. The parameter dN/dS was estimated with well-established frameworks including *HYPHY*, *PAML* (Yang, 2007) and *SNAP* (Korber, 2000). The results showed differences of dN/dS in viruses from different ethnic groups, where in general the Black ethnic group presented a higher dN/dS (with values above 1 that suggested positive or diversifying selection) than other ethnic groups (including Whites, Hispanic and Asians, presenting values below 1 that suggested negative or purifying selection) (further details in Perez-Losada *et al.*, 2009). The results showed viral genetic consequences that could be derived from a different immune response among races and, consequently, multiple ethnicities should be taken into account to design future trials and vaccines.

Concluding Remarks

Recombination is a crucial evolutionary process in multiple viruses allowing a rapid adaptation to changing environments. The major importance of recombination motivated the development of a number of computational methods (together with software implementation) to analyze it in terms of evaluating its presence/absence, predicting breakpoints and estimating rates. However, only a few computational frameworks (especially those implementing a GUI) have been well-established in the field. In practice, estimates of recombination can disagree between frameworks and, for this reason, it could be interesting to find more studies evaluating the accuracy of the currently available frameworks (i.e., using expectations and computer simulations) in the analysis of typical DNA sequences under different levels of diversity and recombination. An attempt has been made to address this concern evaluating the recombination rate estimated with three frameworks (Lopes *et al.*, 2014) but a more comprehensive study is necessary.

In this article it has been highlighted that ignored recombination can bias the phylogenetic tree reconstruction and other evolutionary analyses based on such a phylogenetic tree (Box 1). However, in the bibliography one can easily find studies presenting recombination during the evolution of virus populations and phylogenetic tree reconstructions that ignore recombination. In this article two methodologies have been presented to properly perform a phylogenetic inference (and other evolutionary analyses affected by recombination) under the presence of recombination and it is recommended to apply them to formally avoid biases. Hopefully, future studies on Virus Evolution undergoing recombination will benefit from the technical information reviewed in this article.

Acknowledgments

This study was supported by the Grants “RYC-2015–18241” from the Spanish Government and “ED431F 2018/08” from the Xunta de Galicia.

References

- Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164, 1229–1236.
- Araujo, N.M., 2015. Hepatitis B virus intergenotypic recombinants worldwide: An overview. *Infection, Genetics and Evolution* 36, 500–510.
- Archer, J., Pinney, J.W., Fan, J., *et al.*, 2008. Identifying the important HIV-1 recombination breakpoints. *PLOS Computational Biology* 4, e1000178.
- Arenas, M., 2013. The importance and application of the ancestral recombination graph. *Frontiers in Genetics* 4, 206.
- Arenas, M., 2015. Genetic consequences of antiviral therapy on HIV-1. *Computational and Mathematical Methods in Medicine* 2015, 9.
- Arenas, M., Araujo, N.M., Branco, C., *et al.*, 2018. Mutation and recombination in pathogen evolution: relevance, methods and controversies. *Infection, Genetics and Evolution* 63, 295–306.
- Arenas, M., Lopes, J.S., Beaumont, M.A., Posada, D., 2015. CodABC: A computational framework to coestimate recombination, substitution, and molecular adaptation rates by approximate Bayesian computation. *Molecular Biology and Evolution* 32, 1109–1112.

- Arenas, M., Lorenzo-Redondo, R., Lopez-Galindez, C., 2016. Influence of mutation and recombination on HIV-1 in vitro fitness recovery. *Molecular Phylogenetics and Evolution* 94, 264–270.
- Arenas, M., Posada, D., 2010a. Computational design of centralized HIV-1 genes. *Current HIV Research* 8, 613–621.
- Arenas, M., Posada, D., 2010b. The effect of recombination on the reconstruction of ancestral sequences. *Genetics* 184, 1133–1139.
- Arenas, M., Posada, D., 2012. Simulation of coding sequence evolution. In: Cannarozzi, G.M., Schneider, A. (Eds.), *Codon Evolution*. Oxford: Oxford University Press.
- Arenas, M., Posada, D., 2014. The influence of recombination on the estimation of selection from coding sequence alignments. In: Fares, M.A. (Ed.), *Natural Selection: Methods and Applications*. Boca Raton: CRC Press/Taylor & Francis.
- Barouch, D.H., Stephenson, K.E., Borducchi, E.N., et al., 2013. Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. *Cell* 155, 531–539.
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., et al., 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* 15, e1006650.
- Boussau, B., Gueguen, L., Gouy, M., 2009. A mixture model and a hidden markov model to simultaneously detect recombination breakpoints and reconstruct phylogenies. *Evolutionary Bioinformatics Online* 5, 67–79.
- Bretscher, M.T., Althaus, C.L., Muller, V., Bonhoeffer, S., 2004. Recombination in HIV and the evolution of drug resistance: for better or for worse? *Bioessays* 26, 180–188.
- Bruen, T.C., Philippe, H., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. *Genetics* 172, 2665–2681.
- Castellano, N., Araujo, N.M., Arenas, M., 2017. Heterogeneous recombination among Hepatitis B virus genotypes. *Infection, Genetics and Evolution* 54, 486–490.
- Chan, A.H., Jenkins, P.A., Song, Y.S., 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. *PLOS Genetics* 8, e1003090.
- Chapentier, C., Nora, T., Tenaillon, O., Clavel, F., Hance, A.J., 2006. Extensive recombination among Human immunodeficiency virus type 1 quaspecies makes an important contribution to viral diversity in individual patients. *Journal of Virology* 80, 2472.
- Dalmon, A., Desbiez, C., Coulon, M., et al., 2017. Evidence for positive selection and recombination hotspots in Deformed wing virus (DWV). *Sci Rep* 7, 41045.
- Doria-Rose, N.A., Learn, G.H., Rodrigo, A.G., et al., 2005. Human immunodeficiency virus type 1 subtype B ancestral envelope protein is functional and elicits neutralizing antibodies in rabbits similar to those elicited by a circulating subtype B envelope. *Journal of Virology* 79, 11214–11224.
- Ellenberger, D.L., Li, B., Lupo, L.D., et al., 2002. Generation of a consensus sequence from prevalent and incident HIV-1 infections in West Africa to guide AIDS vaccine development. *Virology* 302, 155–163.
- Faria, N.R., Quick, J., Claro, I.M., et al., 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 546, 406–410.
- Flynn, N.M., Forthal, D.N., Harro, C.D., et al., 2005. Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *The Journal of Infectious Diseases* 191, 654–665.
- Frahm, N., Nickle, D.C., Linde, C.H., et al., 2008. Increased detection of HIV-specific T cell responses by combination of central sequences with comparable immunogenicity. *Aids* 22, 447–456.
- Galli, A., Bukh, J., 2014. Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. *Trends Microbiol* 22, 354–364.
- Godoy, M.G., Suarez, R., Lazo, E.S., et al., 2014. Genetic analysis and comparative virulence of infectious salmon anemia virus (ISAV) types HPR7a and HPR7b from recent field outbreaks in Chile. *Virology Journal* 11, 204.
- Gog, J.R., Pellis, L., Wood, J.L., et al., 2015. Seven challenges in modeling pathogen dynamics within-host and across scales. *Epidemics* 10, 45–48.
- Griffiths, R.C., Marjoram, P., 1997. An ancestral recombination graph. In: Donnelly, P., Tavaré, S. (Eds.), *Progress in Population Genetics and Human Evolution*. Berlin: Springer-Verlag.
- Hadany, L., Beker, T., 2003. On the evolutionary advantage of fitness-associated recombination. *Genetics* 165, 2167–2179.
- Holmes, E.C., Worobey, M., Rambaut, A., 1999. Phylogenetic evidence for recombination in dengue virus. *Molecular Biology and Evolution* 16, 405.
- Hudson, R.R., 2001. Two-locus sampling distributions and their application. *Genetics* 159, 1805–1817.
- Hudson, R.R., Kaplan, N.L., 1988. The coalescent process in models with selection and recombination. *Genetics* 120, 831–840.
- Hurst, L.D., 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet* 18, 486.
- Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14, 68–73.
- Iglesias-Sanchez, M.J., Lopez-Galindez, C., 2002. Analysis, quantification, and evolutionary consequences of HIV-1 in vitro recombination. *Virology* 304, 392–402.
- Jetz, A.E., Yu, H., Klarmann, G.J., et al., 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *Journal of Virology* 74, 1234–1240.
- Korber, B., 2000. HIV signature and sequence variation analysis. In: Rodrigo, A.G., Learn, G.H. (Eds.), *Computational Analysis of HIV Molecular Sequences*. Dordrecht, Netherlands: Kluwer Academic Publishers.
- Kosakovsky Pond, S.L., Frost, S.D., Muse, S.V., 2005. HYPHY: Hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679.
- Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D., 2006. GARD: A genetic algorithm for recombination detection. *Bioinformatics* 22, 3096–3098.
- Kuhner, M.K., 2006. LAMARC 2.0: Maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics* 22, 768–770.
- Kuhner, M.K., Yamato, J., Felsenstein, J., 2000. Maximum likelihood estimation of recombination rates from population data. *Genetics* 156, 1393–1401.
- Laureau, R., Loeillet, S., Salinas, F., et al., 2016. Extensive Recombination of a Yeast Diploid Hybrid through Meiotic Reversion. *PLOS Genetics* 12, e1005781.
- Lihana, R.W., Ssemwanga, D., Abimiku, A., Ndembu, N., 2012. Update on HIV-1 diversity in Africa: a decade in review. *AIDS Reviews* 14, 83–100.
- Lopes, J.S., Arenas, M., Posada, D., Beaumont, M.A., 2014. Coestimation of recombination, substitution and molecular adaptation rates by approximate Bayesian computation. *Heredity* 112, 255–264.
- Lorenzo-Redondo, R., Delgado, S., Moran, F., Lopez-Galindez, C., 2014. Realistic three dimensional fitness landscapes generated by self organizing maps for the analysis of experimental HIV-1 evolution. *PLoS One* 9, e88579.
- Mallo, D., Sánchez-Cobos, A., Arenas, M., 2016. Diverse considerations for successful phylogenetic tree reconstruction: Impacts from model misspecification, recombination, homoplasy, and pattern recognition. In: Elloumi, M., Iliopoulos, C., Wang, J., Zomaya, A. (Eds.), *Pattern Recognition in Computational Molecular Biology*. John Wiley & Sons, Inc.
- Marston, D.A., McElhinney, L.M., Ellis, R.J., et al., 2013. Next generation sequencing of viral RNA genomes. *BMC Genomics* 14, 444.
- Martin, D.P., Lemey, P., Posada, D., 2011. Analysing recombination in nucleotide sequences. *Molecular Ecology Resources* 11, 943–955.
- Martin, D.P., Murrell, B., Golden, M., Khosla, A., Muhire, B., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 1, vev003.
- Maydt, J., Lengauer, T., 2006. Recco: Recombination analysis using cost optimization. *Bioinformatics* 22, 1064–1071.
- McBurney, S.P., Ross, T.M., 2007. Developing broadly reactive HIV-1/AIDS vaccines: A review of polyvalent and centralized HIV-1 vaccines. *Curr Pharm Des* 13, 1957–1964.
- McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T., 2016. Reassortment in segmented RNA viruses: Mechanisms and outcomes. *Nature Reviews Microbiology* 14, 448–460.
- McMichael, A., Picker, L.J., Moore, J.P., Burton, D.R., 2013. Another HIV vaccine failure: Where to next? *Nature Medicine* 19, 1576–1577.
- Mcvean, G., Awadalla, P., Fearnhead, P., 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160, 1231–1241.
- Mcvean, G.A., Myers, S.R., Hunt, S., et al., 2004. The fine-scale structure of recombination rate variation in the human genome. *Science* 304, 581–584.
- Milne, I., Lindner, D., Bayer, M., et al., 2009. TOPALI v2: A rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* 25, 126–127.
- Minin, V.N., Dorman, K.S., Fang, F., Suchard, M.A., 2005. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* 21, 3034–3042.
- Monjane, A.L., Van Der Walt, E., Varsani, A., Rybicki, E.P., Martin, D.P., 2011. Recombination hotspots and host susceptibility modulate the adaptive value of recombination during maize streak Virus Evolution. *BMC Evolutionary Biology* 11, 350.

- Moradigaravand, D., Kouyos, R., Hinkley, T., *et al.*, 2014. Recombination accelerates adaptation on a large-scale empirical fitness landscape in HIV-1. *PLOS Genetics* 10, e1004439.
- Perez-Losada, M., Arenas, M., Castro-Nallar, E., 2018. Microbial sequence typing in the genomic era. *Infection, Genetics and Evolution* 63, 346–359.
- Perez-Losada, M., Arenas, M., Galan, J.C., Palero, F., Gonzalez-Candelas, F., 2015. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. *Infection, Genetics and Evolution* 30C, 296–307.
- Perez-Losada, M., Jobes, D.V., Sinangil, F., *et al.*, 2011. Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. *PLOS One* 6, e16902.
- Perez-Losada, M., Posada, D., Arenas, M., *et al.*, 2009. Ethnic differences in the adaptation rate of HIV gp120 from a vaccine trial. *Retrovirology* 6, 67.
- Poon, A.F., Kosakovsky Pond, S.L., Richman, D.D., Frost, S.D., 2007. Mapping protease inhibitor resistance to human immunodeficiency virus type 1 sequence polymorphisms within patients. *Journal of Virology* 81, 13598–13607.
- Posada, D., 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Molecular Biology and Evolution* 19, 708–717.
- Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* 98, 13757–13762.
- Prosperi, M.C., Prosperi, L., Bruselles, A., *et al.*, 2011. Combinatorial analysis and algorithms for quasispecies reconstruction using next-generation sequencing. *BMC Bioinformatics* 12, 5.
- Rolland, M., Jensen, M.A., Nickle, D.C., *et al.*, 2007. Reconstruction and function of ancestral center-of-tree human immunodeficiency virus type 1 proteins. *Journal of Virology* 81, 8507–8514.
- Scheel, T.K., Gall, A., Li, Y.P., *et al.*, 2013. Productive homologous and non-homologous recombination of hepatitis C virus in cell culture. *PLOS Pathogens* 9, e1003228.
- Schierup, M.H., Hein, J., 2000a. Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156, 879–891.
- Schierup, M.H., Hein, J., 2000b. Recombination and the molecular clock. *Molecular Biology and Evolution* 17, 1578–1579.
- Schultz, A.K., Zhang, M., Bulla, I., *et al.*, 2009. jpHMM: improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Research* 37, W647–W651.
- Simon-Loriere, E., Holmes, E.C., 2011. Why do RNA viruses recombine? *Nature Reviews Microbiology* 9, 617–626.
- Smyth, R.P., Schlub, T.E., Grimm, A.J., *et al.*, 2014. Identifying recombination hot spots in the HIV-1 genome. *Journal of Virology* 88, 2891–2902.
- Strimmer, K., Rambaut, A., 2002. Inferring confidence sets of possibly misspecified gene trees. *Proceedings of Royal Society London Series B – Biological Sciences* 269, 137–142.
- Weaver, S., Shank, S.D., Spielman, S.J., *et al.*, 2018. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution* 35, 773–777.
- Westesson, O., Holmes, I., 2009. Accurate detection of recombinant breakpoints in whole-genome alignments. *PLOS Computational Biology* 5, e1000318.
- Wilson, D.J., Mcvean, G., 2006. Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* 172, 1411–1425.
- Wiuf, C., Christensen, T., Hein, J., 2001. A simulation study of the reliability of recombination detection methods. *Molecular Biology and Evolution* 18, 1929–1939.
- Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24, 1586–1591.
- Zanini, F., Brodin, J., Albert, J., Neher, R.A., 2017. Error rates, PCR recombination, and sampling depth in HIV-1 whole genome deep sequencing. *Virus Research* 239, 106–114.
- Zhang, Y., 2013. A dynamic Bayesian Markov model for phasing and characterizing haplotypes in next-generation sequencing. *Bioinformatics* 29, 878–885.

Relevant Websites

- <https://www.datamonkey.org/>
Datamonkey Adaptive Evolution Server.
- https://www.hiv.lanl.gov/content/sequence/ELF/epitope_analyzer.html
ELF Epitope Location Finder.
- <https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>
HIV Circulating Recombinant Forms (CRFs).
- <http://bioinf.man.ac.uk/robertson/recombination/programs.shtml>
Links to Recombinant Software Detection/Analysis Software.
- <http://jpHMM.gobics.de/submission.html>
Online submission.
jpHMM.

Phylogeny of Viruses

Alexander E Gorbalenya, Leiden University Medical Center, Leiden, The Netherlands

Chris Lauber, TWINCORE - Center for Experimental and Clinical Infection Research, Hannover, Germany

© 2021 Published by Elsevier Ltd.

This is an update of Alexander E. Gorbalenya, C. Lauber, Phylogeny of Viruses, Reference Module in Biomedical Sciences, Elsevier Inc., 2017, doi:10.1016/B978-0-12-801238-3.95723-4.

Introduction: Evolution, Phylogeny, and Viruses

Biological species, including viruses, change through generations and over time in the process known as evolution. These changes are first fixed in the genome due to selection or by chance (due to genetic drift or as a result of ‘founder effect’) that give rise to genetic lineages. Due to either limited fidelity of the replication apparatus copying the genome or adverse physico-chemical activity of the environment, nucleotides may be changed, inserted, or deleted. Genomes of other origin may also be a source of innovation for a genome through the use of specially evolved mechanisms of genetic exchange (recombination). Accepted changes, known as mutations, may be neutral, advantageous, or deleterious, and depending on the population size and environment, the mutant lineage may proliferate or go extinct. Overall, advantageous mutations and large population size increase the chances for a lineage to succeed. The fitness of a lineage is constantly re-assessed in the ever-changing environment and lineages that, due to mutation, became a success in the past could be unfit in new environments. Due to the growing number of mutations accumulating in the genomes that may repeatedly affect the same genomic position, lineages diverge over time, although occasionally, due to stochastic reasons or under similar selection pressure, they may converge locally.

The relationship between biological lineages related by common descent is called phylogeny; the same term also embodies the methodology of reconstructing these relationships. Phylogeny deals with past events and, therefore, it is reconstructed by quantification of differences accumulated between lineages. Due to the lack of physical fossils and (relatively) high mutation rates, the utility of phylogeny was considered limited in the application to viruses until the advent of molecular data and discovery of endogenous viral elements integrated into host genomes proved otherwise. Comparison of nucleotide and amino acid sequences, and, occasionally, other quantitative characteristics such as distances between three-dimensional structures of biopolymers, have been used to reconstruct virus phylogenies to different depths. Results of a phylogenetic analysis are typically depicted in the form of a tree that may be used as a synonym for phylogeny. For instance, the all-inclusive phylogeny of cellular species is known as the Tree of Life (ToL) (Fig. 1(A)). More recently, two techniques, networks and forests of trees, are used to illustrate the complexity of phylogenetic relations and the uncertainty of phylogenetic inference, respectively.

With few exceptions, virus phylogeny follows the theory and practice developed for phylogeny of cellular life forms. For inferring phylogeny, differences between the sequences of species members, assumed to be of a discernable common origin, are analyzed. If species in all lineages evolve at a uniform and constant rate, like clocks tick, their evolution conforms to a molecular clock model. The utility of this model in relation to viruses may be very limited. Rather, related virus lineages may evolve at different and fluctuating rates and some sites may mutate repeatedly, including reverse substitutions. As a result, reconstruction of a full record of change at all sites may be infeasible for the currently available methods and is associated with ever increasing uncertainty with each new mutation as the distance increases. Furthermore, the accumulation of inter-species residue differences may progress nonlinearly with the time elapsed. At present, our understanding of these parameters of virus evolution is poor and this limits our ability to assess the fit between a reconstructed phylogeny and the true phylogeny, with the latter practically remaining unknown for most virus isolates. This gap in our knowledge does not undermine the conceptual strength and utility of phylogenetic analysis for reconstructing the relationships between biological species including the origins of viruses.

The ultimate goal of virus phylogeny is reconstructing the relationships between ‘all’ virus isolates and species. For instance, cellular species form three compact domains (kingdoms) and their origin can be traced back to a common ancestor in the ToL, using either ribosomal RNA or a common set of single-copy genes. Such inference is not feasible for viruses due to their diversity and the lack of a universal molecular denominator (trait). Thus, reconstructing a comprehensive virus phylogeny may require comparisons that involve genomes of viral and cellular origins. This formidable task remains largely ‘work in progress’. In fact, most efforts in virus phylogeny are invested in reconstructing the relationships at the micro, rather than grand, scale and they focus on well-sampled lineages that have practical (e.g., medical) relevance. Most recently, due to the advent of high-throughput next generation sequencing (NGS) and metagenomics, phylogeny of distant relations to characterize diverse viromes and the entire virosphere has become an active area of research. Phylogeny itself or in combination with other data may provide a deep insight into virus evolution and diverse aspects of virus life cycles, including virus interactions with their hosts.

Our knowledge about contemporary virus diversity has been steadily advancing with new viruses being constantly described by systematic efforts as well as occasional discoveries. In addition, descriptions of viral integrations into host genomes grant insights into some historic viral diversity that may now be extinct. These developments indicate that only small fractions of the currently circulating viruses and those that went extinct have so far been unraveled and have become available for phylogenetic studies. In any case, the continuing virus discovery effort will help to further refine viral phylogenies and resolve relationships that are unresolvable with our current virus sampling.

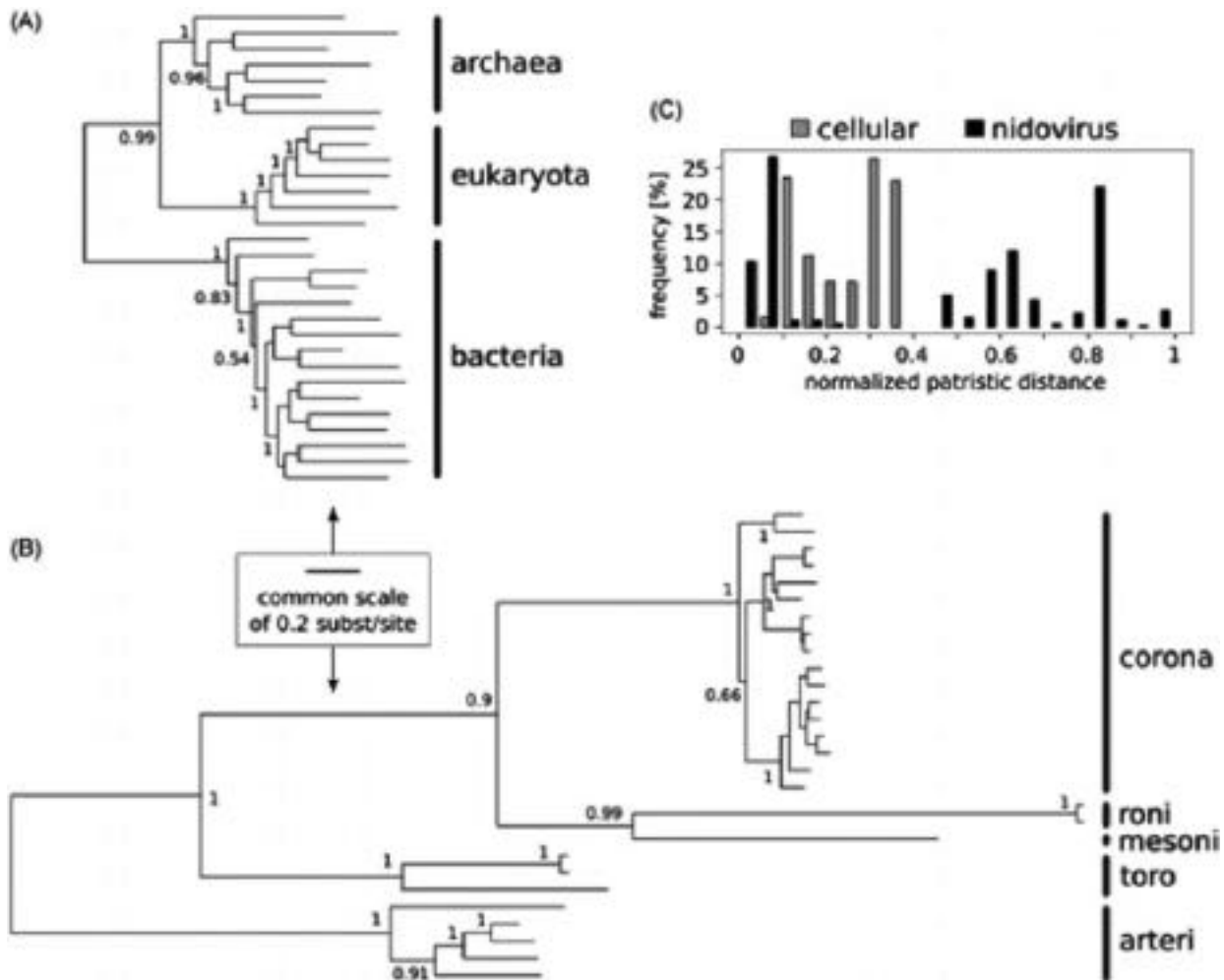


Fig. 1 Phylogeny of nidoviruses in comparison to the Tree of life (ToL) using the most conserved single-copy orthologous proteins. Bayesian phylogenies under the WAG amino acid substitution model with rate heterogeneity across sites and relaxed molecular clock with log-normal distribution of nidoviruses (A) and ToL (B) are drawn to a common scale of 0.2 amino acid substitutions per position. Major lineages are indicated by vertical bars and names; arteri: *Arteriviridae*, mesoni: *Mesoniviridae*, roni: *Roniviridae*, toro: *Toroviridae* (*Tobaniviridae*, from 2018), corona: *Coronaviridae*. Support values at basal internal nodes are posterior probability support values. (C) Distributions of pair-wise patristic distances extracted from (A) and (B). The combined set of distances was normalized relative to the largest distance that was set to one. Figure adapted from Lauber, C., Goeman, J., Parquet, M.D.C., *et al.*, 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathogens* 9 (7), e1003500. including update of the family taxonomy as of 2020.

Tree Definitions

Species share similarity that varies depending on the rate of evolution and time of divergence. The entire process of generating contemporary species diversity from a common ancestor is believed to proceed through a chain of intermediate ancestors specific for different subsets of the analyzed species (Fig. 2). Typically, these ancestral sequences are estimated internally during the tree building process or are not required at all, depending on the method used. The relationship between the common ancestor, intermediate ancestors, and contemporary species may be likened to the relationship between, respectively, root, internal nodes, and terminal nodes (leaves) of a tree, an abstraction that is widely used for the visualization of this relationship (Fig. 2). Alignment of the contemporary sequences with the reconstructed tree side by side, like shown for the toy example in Fig. 2, may reveal the full chain of sequence changes that have happened during evolution which, however, is rarely the case for real data sets due to repeated substitutions and incomplete species sampling. Trees are also part of graph theory, a branch of mathematics, whose apparatus is used in phylogeny. Formally and due to a strong link between phylogeny and taxonomy, leaves may be called operational taxonomy units (OTUs) and internal nodes and roots, since they have not been directly observed, are known as hypothetical taxonomy units (HTUs). Nodes are connected by branches or edges.

The tree may be characterized by topology, length of branches, shape, and the position of the root (Fig. 3). The topology is determined by relative positions of internal nodes, including the root, and terminal nodes; it defines branching events leading to

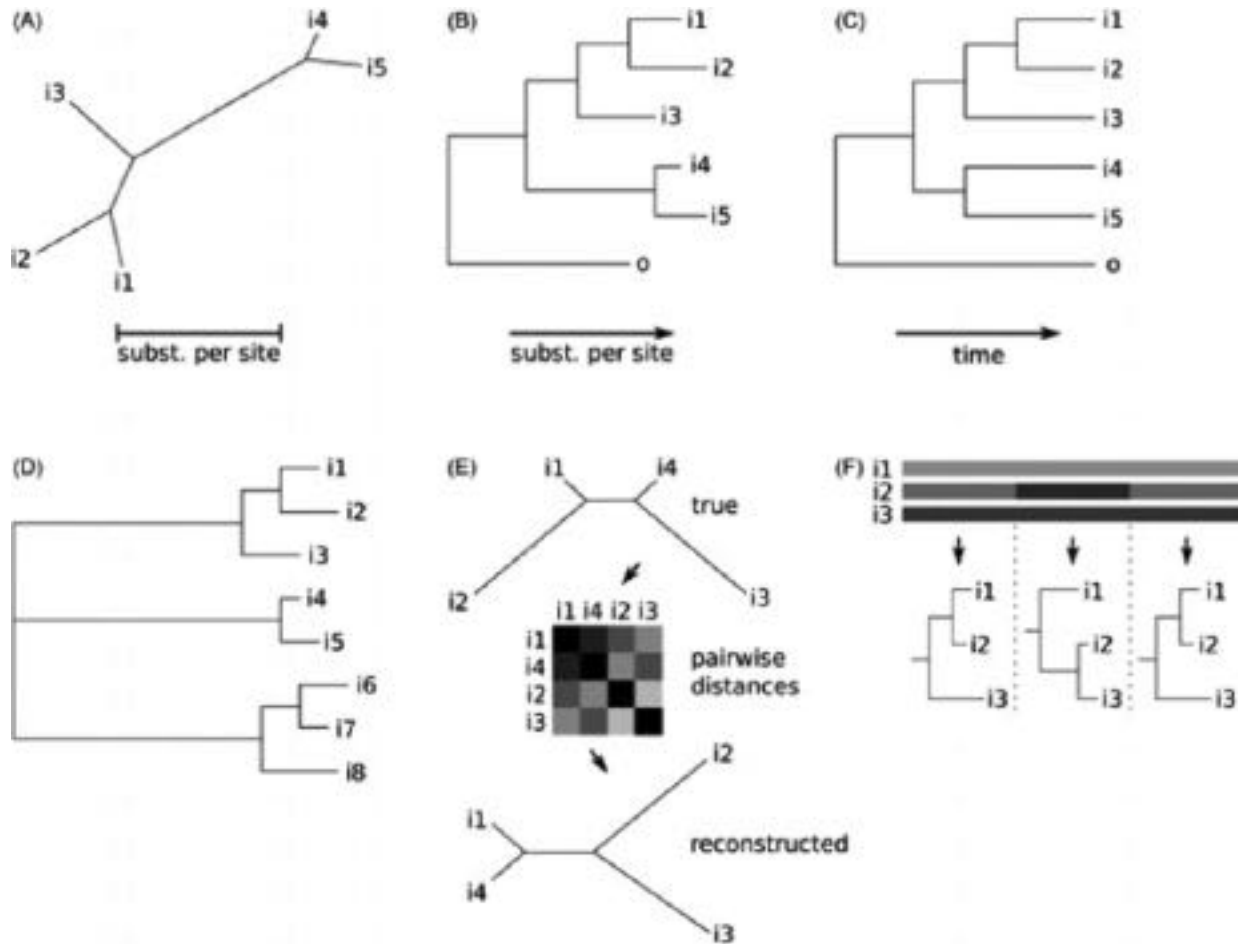


Fig. 3 Tree types and pitfalls of phylogeny reconstruction. (A) Unrooted tree for five hypothetical viruses that was reconstructed based on their gene or protein sequences. Branch lengths represent the amount of genetic change between two viruses typically measured in units of substitutions per site. The direction of evolution is undetermined. (B) The tree for the five ingroup viruses in A was rooted using an outgroup (o). The direction of evolution is from left to right. (C) The same tree as in B but with branches calibrated to represent time. Note that time does not necessarily correlate with the amount of genetic change (for instance, compare length of the branch leading to the cluster joining i4 and i5 with that of the same branch in B). (D) The relative positions of three highly divergent lineages is unresolved by the phylogeny (polytomy). (E) The true relationship of four hypothetical sequences (top) is not recovered by the phylogenetic reconstruction (bottom) due to long branch attraction involving i2 and i3. (F) Phylogenetic trees (bottom) reconstructed for three adjacent genomic regions (top) are different with respect to the position of i2 which was subject to a recombination event in middle genome region.

acceleration of computing power (see also below). Distance methods are often designed to converge on a unique phylogeny by clustering, with none others being even considered. The unweighted pair group method with arithmetic means (UPGMA) in which a constantly recalculated distance matrix is used to define the hierarchy of similarities through systematic and stepwise merging of most similar pairs at a time was the first technique introduced for clustering. The neighbor-joining (NJ) method uses a more sophisticated algorithm of clustering that minimizes branch lengths, and is the most popular among distance methods.

Although different trees may be compared in how they fit a distance matrix, it is character-based methods that are routinely used to assess numerous alternative phylogenies in search for the best one in a computationally intensive process. Due to the calculation time involved, assessing all possible phylogenies is found to be impractical for data sets including more than 10 sequences; for larger data sets different heuristic approximations are used that may not guarantee a recovered phylogeny to be the best overall. There are two major criteria for selecting the best phylogeny using character-state based information through either maximum parsimony (MP) or maximum likelihood (ML). In MP analysis, a phylogeny with a minimal number of substitutions separating the analyzed species is sought. The ML analysis offers a statistical framework for comparing the likelihood of fitting different trees to the data under competing models of evolution with parameters including population size change and rate of mutation in search for one with the best fit. The latter approach is mathematically robust and its statistical power may also be used in combination with other techniques of tree generation. A Bayesian variant of the ML approach has gained popularity in recent years. It can utilize prior knowledge about the evolutionary process, like known substitution rates or clustering of species subsets or dates of species isolation, in combination with repeated sampling from subsequently derived hypotheses. The result of a Bayesian

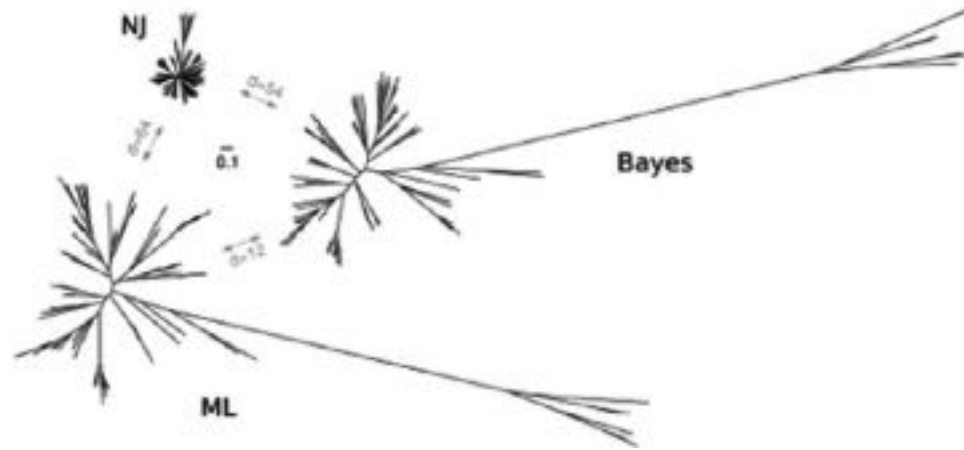


Fig. 4 Comparison of phylogenetic results between three methods. Shown are unrooted NJ, ML and Bayesian trees reconstructed for the same dataset of 287 aligned VP1 protein positions of 93 polyomaviruses. The LG amino acid substitution model with site heterogeneity modeled by a gamma distribution with four categories, as selected by ProtTest, was used. In the Bayesian analysis, a relaxed molecular clock approach with log-normally distributed rate was applied. The trees are drawn to the same scale of average amino acid substitutions per site, as indicated by the bar in the middle. Note the considerably shorter branch lengths of the NJ tree compared to the other two trees. Robinson-Foulds distances measuring the topological differences between tree pairs are shown in gray.

analysis is thus a forest of trees that reflects the uncertainty associated with the reconstructed phylogeny and which forms the basis to derive a consensus tree and statistical support for its branches and other estimated evolutionary parameters. In phylogenetic analysis of viruses the dates of species isolation are often used to date the MRCA of the analyzed viruses under a Bayesian framework, while fossil information is routinely used to time-calibrate trees of cellular organisms. The use of endogenous viral elements integrated into host genomes as genetic ‘fossils’ for time-calibration of virus phylogenies has also been explored. Bayesian methods have the highest computational cost due to their sampling approach and thus show the lowest speed, while realization of the similarly advanced ML algorithm may be largely comparable in speed to distance methods, allowing for the phylogenetic analysis of very large data sets like genome-wide tree reconstructions of cellular organisms or many thousands of viruses.

One should keep in mind that different methods for phylogeny reconstruction can produce different trees, concerning both topology and branch lengths, for the same data set, although better agreement between ML and Bayesian trees is common, especially in respect to branch lengths (Fig. 4). None of the methods is considered superior to the other methods with respect to all aspects of phylogeny reconstruction, and which method to use under what circumstances is often a point of debate. A valid approach to gain further confidence in phylogenetic results is to apply several methods on the data and to only trust HTUs that are inferred by more than one method.

After a tree is chosen, it is common to assign support values to internal nodes through assessing the nodes’ persistence in trees related to the chosen tree. One particular technique, called bootstrap analysis, in which trees are generated for numerous randomly modified derivatives of the original data set, is most frequently used in distance-based as well as MP and ML analysis. Each internal node in the original tree is characterized by a so-called bootstrap value that is equal to the number of times a node appears in all tested trees. Although the relationship between bootstrap and statistical values is not linear, nodes with very high bootstrap values are considered to be reliable. In a Bayesian analysis, the support of internal nodes is quantified through posterior probability values.

If species evolve according to a molecular clock model, the root position in a tree could directly be calculated from the observed inter-species differences as a midpoint of cumulative inter-species differences. Alternatively, the root position may be assigned to a tree from knowledge about the analyzed species that was gained independently from phylogenetic analysis. Commonly, this knowledge comes in the form of a single or more lineages which are assumed (or known) to have emerged before the ‘birth’ of the analyzed cluster. These early diverged species are collectively defined as ‘outgroup’, while the analyzed species may be called ‘ingroup’ (Fig. 3(B) and (C)). Also, a tree may be generated unrooted, a common practice in phylogenetic analysis of viruses for which the applicability of the molecular clock model remains largely untested and reliable outgroups may not be routinely available (Fig. 4). In an unrooted tree, grouping of species in separate clusters may be apparent, although these clusters may not be treated as monophyletic as long as the direction of evolution has not been defined. These challenges are addressed by the development of new approaches that infer rooted trees without artificially restricting species evolution to a constant rate (known as relaxed molecular clock models commonly used within the Bayesian framework).

Virus phylogeny can be inferred using either genomes or distinct genes, and each of these approaches, standard in phylogenomics, may be considered as complementary. Under the first approach, genome-wide alignments are used for analysis. Due to complexities of the evolutionary process that may be region specific, reliable genome-wide alignments can routinely be built only for relatively closely related viruses whose analysis, however, may be further complicated by recombination events (see below). Using the second approach, genes with no evidence for recombination may be merged (concatenated) in a single data set that may

be used to produce a superior phylogenetic signal compared to those generated for distinct genes or entire genomes. For viruses with small genomes or for a diverse set of viruses, it is common practice to use a single gene to infer virus phylogeny. Although the results produced may be the best models describing evolutionary history of a group of viruses, the validity of this gene-based approach for the genome-wide extrapolation remains a point of debate. Recently, network methods were used to infer and depict evolutionary relationships of multigene virus genomes taking into account gene-specific sequence affinities.

When the gene tree is used as representing the phylogeny of the entire genome, an underlying most common assumption is that its topology but not branch lengths holds for different genomic regions in reflection of their coevolution with potentially different rates of substitution. This assumption may be violated due to several evolutionary processes, including orthologous gene exchange between (closely) related viruses, gene duplication and horizontal gene transfer (HGT), all involving one or another form of recombination, or incomplete lineage sorting. In phylogenetic terms, this violation may be revealed through incongruency of trees built for different genome regions (Fig. 3(F)). Trees may also become incongruent due to various technical reasons related to the size and diversity of a virus data set. These characteristics complicate interpretation of the congruency test, which is widely used in different programs to identify recombination in viruses. Other pitfalls of phylogenetic reconstruction include the inability to resolve basal branching patterns of highly divergent lineages (Fig. 3(D)) and the relatively close clustering of lineages that are only distantly related and do not form a monophyletic group in the true (unknown) phylogeny (Fig. 3(E)). The latter phenomenon is known as long branch attraction (LBA) and the phylogenetic artifacts produced by LBA are most frequently observed for isolated, that is, long branches in the tree which represent distant lineages with no close relatives known.

Applications of Phylogeny in Virology

Phylogenetic analysis is used in a wide range of studies to address both applied and fundamental issues of virus research, including epidemiology, diagnostics, forensic studies, phylogeography, origin, evolution, and taxonomy of viruses. The first questions to be answered during an outbreak of a virus epidemic concern the virus identity and origin. Answers to these questions form the basis for implementing immediate practical measures and prospective planning, enabling specific and rapid virus detection and epidemic containment, which may include the use and development of antiviral drugs and vaccines. Among different analyses performed for virus identification at the early stage of a virus epidemic, the phylogenetic characterization is used for determining the relationship of a newly identified virus with all other previously characterized and sequenced viruses.

Results of this analysis may be sufficient to provide answers to the questions posed, as regularly happens with closely monitored viruses that include most human viruses of high social impact, for example, influenza virus, human immunodeficiency virus (HIV), hepatitis C virus (HCV), poliovirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and others that belong to different species. For these viruses, there exist large databases of previously characterized isolates and strains that comprehensively cover the so far characterized natural diversity. Should a newly identified virus belong to one of the respective species, chances are that it has evolved from a previously sampled isolate or a close variant and this immediately becomes evident in the clustering of these viruses in the phylogenetic tree. Combining the results of gene-specific and genome-wide phylogenetic analysis allows one to determine whether recombination contributed to the isolate origin. For instance, recombination was found to be extremely uncommon in the evolution of HCV, but not for poliovirus lineages that recombine promiscuously, also with closely related human coxsackie A viruses, both of which belong to the same virus species of human enteroviruses known as *Enterovirus C*.

When an emerging infection is caused by a new never-before-detected virus, the phylogenetic analysis is instrumental for classification of this virus and in the case of a zoonotic infection, for determining the dynamic of virus introduction into the (human) population and initiating the search for the natural virus reservoir. This was the case with many emerging infections including those caused by Nipah virus, SARS-CoV, Middle East respiratory syndrome coronavirus (MERS-CoV), ebolavirus, and Zika virus. In the case of SARS-CoV, very limited sampling of the coronavirus diversity at the time, some uncertainty over the relationship between phylogeny and taxonomy of coronaviruses, and the complexity of phylogenetic analysis of a virus data set including isolated distant lineages led to considerable controversy over the exact evolutionary position of SARS-CoV among coronaviruses, does it prototype a new genus (group at the time) or not. Since then, the matter has fully been resolved but this experience illustrates some challenges in inferring virus phylogeny. That experience informed the subsequent classification of MERS-CoV and SARS-CoV-2, most recently.

The search for a zoonotic reservoir of an emerging virus may involve a significant and time-consuming effort that requires numerous phylogenetic analyses of ever-expanding sampling of the virus diversity generated in pursuit of the goal. In this quest, phylogenetic analysis canalizes the effort and provides crucial information for reconstructing parameters of major evolutionary events that promoted the virus origin and spread. For instance, intertwining HIV and simian immunodeficiency virus (SIV) lineages in the primate lentivirus tree led to the postulation that the existing diversity of HIV in the human population originated from several ancestral viruses independently introduced from primates over a number of years. Similar phylogenetic reasoning was used to trace the origin of a local HIV outbreak to a common source of HIV introduction through dental practice (known as 'HIV dentist' case). These are typical examples illustrating the utility of phylogenetic analysis for epidemiological and forensic studies.

Geographic distribution of places of virus isolation is another important characteristic relative to which virus phylogeny may be evaluated. This field of study belongs to phylogeography. The evolution of human JC polyomavirus provides an example of confinement of circulation of virus clusters to geographically isolated areas, represented by three continents. Identification of West

Nile virus in the USA illustrates a geographical expansion of an Old World virus into the New World. Analysis of phylogenies of field isolates of rabies virus of the family *Rhabdoviridae* sampled from different animals across Europe led to the recognition that interspecies virus expansion occurs faster when compared to geographical expansion.

Phylogenies can also reveal information about the relative strength of the virus–host association over time. In some virus families (e.g., the *Coronaviridae*) host-jumping events may be relatively frequent in establishing new species or diversifying existing species; they include the emergence of successfully circulating human coronavirus OC43 (HCoV-OC43), and repeated introduction of strains from the species *Middle east respiratory syndrome-associated coronavirus* (ongoing MERS-CoV epidemic) and *Severe acute respiratory syndrome-associated coronavirus* (SARS-CoV and SARS-CoV-2 pandemics) into the human population, the latter providing examples of two independent and temporally separated introductions of closely related viruses from the same viral species. At the other end of the spectrum one finds the family *Herpesviridae*. Extensive phylogenetic analysis of herpesviruses and their hosts showed a remarkable congruency of topologies of trees indicating that this virus family may have emerged some 400 million years ago and that herpesviruses largely cospeciate with their hosts. Likewise, members of the family *Hepadnaviridae*, which are reverse-transcribing small DNA viruses, are known for their strong host association and low frequency of cross-speciation events. Moreover, through phylogenetic analysis one can show that most viruses, and in particular RNA viruses, evolve at rates that are orders of magnitude faster than those of cellular organisms. For instance, even the most conserved enzymes encoded by nidoviruses, comprising only few RNA virus families, accumulated more than twice as many substitutions during evolution than their counterparts across the ToL, as estimated through branch lengths of the respective phylogenetic trees (Fig. 1). Taking into account that the MRCA of all cellular organisms likely predates the nidovirus MRCA, this reveals that most residues of viral proteins changed repeatedly and more frequently than cellular protein residues during long-term evolution. In fact, this high evolutionary rate seems to be a prerequisite for RNA viruses to stay fit in the ever-changing environment considering their tiny genomes that would otherwise not be able to produce enough genetic variation.

Phylogenetic analysis becomes increasingly important in virus classification (taxonomy), especially after the acceptance of genomic sequences by the International Committee on Taxonomy of Viruses (ICTV) as the sole basis of virus classification in 2017, which incorporated metagenomic sequences into the taxonomic framework. Viruses discovered by metagenomics now greatly outnumber viral genomes from reference databases. For viruses united in taxa above the genus rank, phylogenetic clustering for most conserved replicative and occasionally structural genes is commonly observed and has been used in the decision making process for many years. For instance, human hepatitis E virus, originally classified as a calicivirus using largely virion properties, was eventually expelled from the family *Caliciviridae* due to poor fit of genome characteristics, including results of phylogenetic analysis; it now forms a separate family *Hepeviridae*. Phylogenetic considerations also played an important role in establishing many new families, for example, the *Marnaviridae* and *Dicistroviridae*. Now these sequence-based analyses are transforming the taxonomy of large DNA phages which was, until recently, developed in such a way that existing families may have united phages with different gene layouts and phylogenies. Phylogenetic analyses of very distant relationships were used to populate the newly established ranks above the order rank, which prompted discussion about the reliability of such inferences. The interaction between phylogeny and taxonomy is dynamic and efforts were also made in extracting taxa and rank structures from monophyletic clusters in trees using analysis of pairwise evolutionary distances. In future, one might hope for important advancements of virus taxonomy that improve cross-family consistency in relation to phylogeny. Genome-based phylogenetic analysis offers a robust and systematic approach to incorporate these viruses into the taxonomy of viruses to foster their easy access and use by the research community.

Acknowledgments

AEG research was partially supported by Leiden University Fund and EU Horizon2020 project EVAg 653316. CL is supported by the Deutsche Forschungsgemeinschaft (DFG) through the Cluster of Excellence RESIST (EXC 2155) project 390874280.

Further Reading

- Dolja, V.V., Koonin, E.V., 2006. Comparative genomics and evolution of complex viruses. *Virus Research* 117, 1–184.
- Domingo, E., 2007. Virus evolution. In: Knipe, D.M., Howley, P.M., Griffin, D.E., *et al.* (Eds.), *Fields Virology*, fifth ed. Philadelphia, PA: Wolters Kluwer, Lippincott Williams and Wilkins, pp. 389–421.
- Domingo, E., Webster, R.G., Holland, J.J., 1999. *Origin and Evolution of Viruses*. San Diego: Academic Press.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29, 1969–1973.
- Felsenstein, J., 2004. *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates, Inc.
- Gibbs, A.J., Calisher, C.H., Garcia-Arenal, F., 1995. *Molecular Basis of Virus Evolution*. Cambridge: Cambridge University Press.
- Guindon, S., Dufayard, J.F., Lefort, V., *et al.*, 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* 59, 307–321.
- Holmes, E.C., Duchêne, S., 2019. Can sequence phylogenies safely infer the origin of the global virome? *mBio* 10, e00289-19.
- King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., 2012. *Virus Taxonomy: The 9th Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier, Academic Press.
- Laubert, C., Goeman, J., Parquet, M.D.C., *et al.*, 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathogens* 9 (7), e1003500.

- Lauber, C., Kazem, S., Kravchenko, A.A., Feltkamp, M.C.W., Gorbalenya, A.E., 2015. Interspecific adaptation by binary choice at de novo polyomavirus T antigen site through accelerated codon-constrained Val-Ala toggling within an intrinsically disordered region. *Nucleic Acids Research* 43 (10), 4800–4813.
- Lauber, C., Seitz, S., Mattei, S., *et al.*, 2017. Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. *Cell Host & Microbe* 22 (3), 387–399.
- Moya, A., Holmes, E.C., Gonzalez-Candelas, F., 2004. The population genetics and evolutionary epidemiology of RNA viruses. *Nature Reviews Microbiology* 2, 279–288.
- Page, R.D., Holmes, E.C., 1998. *Molecular Evolution: A Phylogenetic Approach*. Boston: Blackwell Publishing.
- Salemi, M., Vandamme, A.M., 2003. *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*. Cambridge: Cambridge University Press.
- Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Villarreal, L.P., 2005. *Viruses and Evolution of Life*. Washington, DC: ASM Press.
- Vision, T., Brenner, E., Bowers, J.E., *et al.*, 2006. Taking the first steps towards a standard for reporting on phylogenies: Minimum Information about a Phylogenetic Analysis (MIAPA). *OMICS: A Journal of Integrative Biology* 10 (2), 231–237.
- Weaver, S.C., Danison, M., Roosinck, M., Vignuzzi, M., 2016. *Virus Evolution: Current Research and Future Directions*. Caister: Academic Press.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9, e02329-18.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2019. Reply to Holmes and Duchêne, “Can sequence phylogenies safely infer the origin of the global virome?”: Deep phylogenetic analysis of RNA viruses is highly challenging but not meaningless. *mBio* 10, e00542-19.

Virus Bioinformatics

Nikolaos Pappas, Utrecht University, Utrecht, The Netherlands

Simon Roux, US Department of Energy Joint Genome Institute, Walnut Creek, CA, United States

Martin Hölzer, Kevin Lamkiewicz, Florian Mock, and Manja Marz, University of Jena, Jena, Germany

Bas E Dutilh, Utrecht University, Utrecht, The Netherlands and Radboud University Medical Center, Nijmegen, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

What is up in Virus Bioinformatics

The virosphere may contain the greatest diversity known to mankind. It has been estimated that there are 10^{31} viruses on Earth, and for billions of years their ongoing proliferation and mutation has contributed to an unparalleled genomic diversity globally. Viral mutation rates range from 10^{-8} to 10^{-6} substitutions per nucleotide per cell infection for DNA viruses and from 10^{-6} to 10^{-4} substitutions per nucleotide per cell infection for RNA viruses. The only way to efficiently analyse this biodiversity is by applying powerful computational tools to (1) identify viral sequences and their encoded functional elements, (2) predict, annotate, and compare their functions, and (3) structure the data to move from measuring to understanding. Until recently, our full understanding of viruses was based on a few hundred viruses that were isolated and could be studied in detail. With recent bioinformatic developments, thousands of new viruses can be readily discovered in all natural and host-associated biomes (see also Section “Viral Metagenomics” below). Including these naturally occurring viruses in comparative analyses opens up possibilities for *de novo* computational predictions, including about the structure and function of viral genes.

Technology and Bioinformatics Drive Discoveries

The past decades have been characterized by technological innovations that revolutionized the way we do science, ranging from the development of computers and the internet, to high-throughput measurement technologies including DNA sequencing, mass spectrometry, and imaging. New fields were built based upon these developments, including bioinformatics, machine learning, and omics. These advances have expanded the scope in all scientific fields, not least in virology. One of the most profound impacts is a new view of the virosphere that is one of an unparalleled diversity. To illustrate, the number of recognized deep viral taxonomic groups has been greatly expanded and the International Committee for Taxonomy of Viruses (ICTV) has recently approved an expansion of the resolution of the viral taxonomy to 15 ranks: realm, subrealm, kingdom, subkingdom, phylum, subphylum, class, subclass, order, suborder, family, subfamily, genus, subgenus, and species.

Bioinformatic analyses of omics and other biological datasets depend on specialized computational tools. The development of these tools begins with basic analyses that are then incrementally used to create more complex applications. Examples of basic applications include software to validate the data derived from next-generation sequencing machines, build alignments of gene or protein sequences, and perform statistical tests. Higher-level analyses may include pipelines for metagenomic analysis, genome annotation, or genotype-phenotype association. Taken together, bioinformatics is arguably one of the subdisciplines in the life sciences with the broadest applicability. When calculated as the amount of computer time allotted to computational analyses, the largest consumer in virology is the analysis of omics datasets. Omics analyses are characterized as high-throughput, untargeted, and generally quantitative, and their application opens the door to systems level analysis of viruses and their effects on their hosts. For example, comparative genomics allows thousands of viruses to be analyzed, identifying important viral genes, their functions, and their evolution; metagenomics allows viruses to be discovered and identified with high throughput; and phylogenetics and phylogenomics allow new viral taxonomic groups to be identified. Some of these applications are presented and discussed in the article below.

Tools for Diagnostics

Viral infections can form a significant burden not only for human health but also for the health livestock and plants. The direct detection of viruses in clinical and other samples include microscopy, antigen detection such as ELISA, and molecular detection of the viral genomic material by PCR. Popular molecular diagnostic techniques including qPCR or RT-qPCR also allow quantification of viral loads. While these techniques are highly sensitive for the detection of specific viruses in a sample, they can only identify viral sequences that match a pre-defined search image that matches the designed PCR primers. Thus, these established diagnostic tests frequently yield negative results when a patient presents a clinical phenotype, but no virus is detected. This can be either because an uncommon variant of a known pathogen is present in the sample, or because a novel virus is the causative agent of the disease. Notably, the difference between these two possibilities is continuous, reflecting increasing evolutionary distances along the viral phylogeny.

Bioinformatic approaches allow PCR panels to be designed that capture an increasingly diverse array of viruses, but these assays will always remain limited to detecting viruses within a known range, and cannot extrapolate to identify completely novel ones. This may be resolved by untargeted (shotgun) sequencing of isolated viruses or complete sample DNA (metagenomics). Variants of

known viruses may be detected by aligning the reads derived from the sample to the reference sequence of the known virus that was originally used for designing the primers. If enough high-quality reads span the regions where the primer sequences should anneal with the target, specialized variant detection tools can call the variant with a high degree of confidence, and new PCR primers can be designed to capture them. For example, a recent PCR-based investigation of the widespread human gut-associated bacteriophage crAssphage designed globally applicable primers by screening an alignment of sequencing reads from a range of publicly available metagenomes and identifying highly variable regions of the appropriate size (1000–1400 nucleotides) that were flanked by conserved regions which could be targeted by primers, and were present in $\geq 90\%$ of all metagenomic samples ($< 10\%$ gaps). These primers allowed a range of collaborating laboratories to independently detect crAssphage in samples from 62 different localities on six continents. Detection of completely novel viruses from metagenomic datasets is less straightforward and will be discussed below.

Genome Sequencing

Obtaining the genome sequence of viral isolates is a highly standardized approach where second (massively parallel) and third (single molecule) DNA sequencing technologies have allowed immense progress. An important first step prior to any downstream analyses based on raw sequencing data is quality control. Several metrics can be used to estimate read quality. First, the sequencing machine provides quality scores for the individual nucleotides that estimate the probability that a nucleotide was wrongly measured. These are based on the logarithmic Phred scores and range from 0 (nucleotide measured with 0% accuracy) to > 40 ($> 99.99\%$ accuracy). Second, several heuristics have proven useful in the identification of potentially spurious sequencing reads that may be removed from the data, including the presence of any remaining primer, index, or adapter sequences, the over-representation of specific nucleotide subsequences (*k*-mers), divergent GC content, and the presence of duplicates of the sequence in the sequencing dataset. There is a wide array of bioinformatic tools that have been developed to calculate these metrics and produce quality reports that summarize the results in useful graphical interfaces, such as FastQC, multiQC, and PRINSEQ (Table 1). Once potential issues with the data have been identified, short read sequences can be pre-processed to eliminate these sources of technical variation or errors. Typically, bases that fall under a set threshold are trimmed off along with any leftover primer or adapter sequences. Depending on the downstream application, remaining reads shorter than a length threshold are also discarded. Alternatively, dedicated tools can perform error correction on the short reads themselves.

An additional step that is specific for viral datasets is the removal of any remaining host sequences. If a genome sequence of the host is available, host-derived reads may be detected by mapping the reads against the host genome. Reads that confidently map to the host may be removed to ensure that the remaining part of the read set reflects the viral fraction of the sample. In the clinical context and for human patient samples, in particular, removal of reads mapping to the human reference is essential in order to abide to established international guidelines for safeguarding the individual's privacy. Correct identification of viral sequences that might have integrated in the host genome, as is the case with retroviruses and prophages, still remains a challenge. In these cases, viral reads may still align to the sequences of proviruses or prophages in the host genome if they are sufficiently similar, and may thus be removed from the sequencing dataset. Potential solutions include masking these proviruses in the host reference genome, or postprocessing the removed putative host reads by comparing them to another database of known viral sequences that includes the sequences of the integrated viruses. Detecting integrated viral sequences in the genome sequences of cellular hosts, and accurately determining their integration boundaries remains an ongoing bioinformatic challenge.

A typical sequencing effort targeting a viral genome with current second generation sequencing technologies will result in millions of short sequences of the order of $\sim 10^2$ nucleotides in length. Typically, these sequencing reads are generated from random fragments of the genome or genomes in the sample (hence the term “shotgun sequencing”). Because viral genomes range from $\sim 10^3$ – 10^5 nucleotides, for most practical purposes these short reads need to be assembled, unless a very closely related reference genome sequence is available. Sequence assembly is the process whereby short reads are combined into longer stretches of contiguous sequence (contigs). In case a single non-segmented genome is assembled, the end result optimally consists of a single string of nucleotides representing the complete genome sequence. Since most viral genomes are smaller and simpler in their structure than those of cellular organisms, assembling a full viral genome is relatively straightforward. Still, sequencing errors, repetitive and low complexity regions, and especially quasispecies diversity that results from high mutation rates may pose specific hurdles into obtaining a complete genome sequence. Several virus specific genome assemblers have been developed to address these issues including VICUNA and IVA (Table 1).

The latest advances in long read sequencing technologies promise high quality viral genomes. Recently, long read sequencing was shown to allow whole viral genomes to be captured in a single read, for example by direct sequencing of influenza and coronavirus genomes. Long read sequencing technologies still come with a higher error rate than their short counterparts. Hybrid approaches leveraging the advantages of long reads (ability to span low complexity and coverage regions) and short reads (low error rates) produce high quality, full viral genomes. Successful long read-based genome assemblies have been reported for the human cytomegalovirus and the pig pseudorabies virus. In principle, similar pre-processing steps apply to long as to short sequencing reads, but dedicated tools are used that take into consideration their specific limitations. Extensive quality summaries can be obtained with Poretools or nanoOK. The relatively error-prone long-read sequencing data may be corrected either without the use of additional short-read sequences (i.e., non-hybrid) or with a hybrid approach. Examples of tools performing non-hybrid error-correction include Nanocorrect and PoreSeq, while hybrid methods include Nanocorr and NaS. Similarly, non-hybrid assemblers include Canu and Miniasm, with SPAdes and Unicycler performing hybrid assemblies.

Table 1 List of selected software tools and resources for virus bioinformatics tasks

Read processing tools		
<i>Quality check</i>	FastQC, PRINSEQ, mulitQC Poretools, nanoOK	Checks read sequencing quality Quality checks for nanopore long reads
<i>Raw reads pre-processing</i>	Cutadapt, Trimmomatic, BBduk Nanocorrect, PoreSeq Nanocorr, NaS	Quality trimming, artefacts removal on short reads Non-hybrid error correction for nanopore long reads Hybrid error correction for nanopore long reads
Genome assembly tools		
<i>Single genomes</i>	VICUNA IVA SPAdes Canu, Miniasm Unicycler	Produces population consensus genome assembly Assembler designed for RNA viruses Generic genome assembler Non-hybrid assemblers for nanopore long reads Hybrid assembly pipeline for nanopore long reads with the use of short reads
<i>Metagenomes</i>	MEGAHIT, metaSPAdes, Ray-meta, IBDA-UD crAss	Assemblers optimized for metagenomics data Cross-assembly analysis of multiple metagenomes
Read mapping		
	BWA, Bowtie, BBmap STAR GraphMap, LAST	Align short read sequences to a reference Splice-aware aligner for RNA-seq data Align long read sequences to a reference
Gene Prediction		
	ORF Finder Prodigal VIGOR	Searches for open reading frames in the provided sequence A protein-coding gene prediction software tool Annotation program for small viral genomes
Similarity searches		
	BLAST HHpred HMMER	A suite of tools to find regions of similarity between DNA and protein sequences Sensitive protein homology detection, function, and structure prediction Homology based search
Multiple Sequence Alignment		
	MAFFT, ClustalW MUSCLE	Multiple sequence alignment for DNA and protein sequences Multiple sequence alignment for protein sequences
Sequence taxonomic annotation		
	CAT, Kraken, Centrifuge, Kaiju	Assign taxonomic labels to reads or assembled contigs
Phylogenies		
	RaxML, PhyML BEAST	Inference of large phylogenetic trees A software package for phylogenetic analysis with an emphasis on time-scaled trees
Taxonomy and classification		
	GRAViTy vConTACT VICTOR DEmARC	Classification of eukaryotic viruses Classification of double stranded DNA viruses of bacteria and archaea Genome based phylogeny and classification of prokaryotic viruses Classification of viruses based on genetic divergence
RNA secondary structures		
	mfold/UNAFold ViennaRNA package LocARNA	RNA secondary structure prediction Suite of tools to perform RNA structures prediction and comparison Structure-guided multiple sequence alignment of RNA sequences
Transcriptomics		
	DESeq2, Sleuth	Statistical analysis of RNA-seq data
Databases		
	ViralZone Virus Variation Resource Virus Pathogen Database and Analysis Resource (ViPR)	Link specific knowledge for each virus family with viral protein and genomic sequences A community portal for viral sequence data An integrated repository of data and analysis tools for multiple virus families

After genome assembly, annotation is an important computational analysis that is required for interpreting the functionality of the virus in its environment. This includes a prediction of functional features such as protein-coding genes and tRNAs, and a classification step where each feature is characterized by similarity to known proteins or RNAs. Importantly, most genes predicted from novel viruses are distant from known references, and common similarity detection tools like BLAST often cannot provide relevant information in this context. Conversely, the use of more sensitive tools relying on the detection of conserved residues and the representation of protein sequence diversity as Hidden Markov Model (HMM) profiles (e.g., HMMER, HH-PRED, PSI-BLAST), is much more useful when analyzing novel virus genomes. These tools are able to detect distant homologies between distantly related viral proteins, which is often the only way to detect homology and, consequently, suggest potential functions in the light of the rapid viral sequence evolution.

RNA Secondary Structures in Viruses

RNA secondary structures play an important role in the life cycle of viruses, especially RNA viruses. These are formed either via the interaction of nucleotides located at close proximity to each other or at distances of several thousand bases (i.e., long-range RNA-RNA interactions, LRI). Local RNA structures were shown to be involved in translation initiation in Hepacivirus and Tombusvirus, while LRIs between the 5' and 3'-UTRs of *Flaviviridae* family genomes promote replication. Moreover, a network of intra- and intersegment RNA-RNA interactions facilitates reassortment between Influenza A genomic segments from different co-infecting strains. This genomic reshuffling may have important effects, including the loss of vaccine efficacy.

Current algorithms for *in silico* prediction of RNA secondary structures mainly employ thermodynamic methods and can be applied to single sequences with software tools like mfold or its successor UNAFold. Furthermore, functional RNA secondary structures are conserved among different viral strains and species and it has been shown that this conservation is higher on the structure level than on the sequence level. Improved bioinformatic methods for *in silico* predictions use several different sequences and their covariances. These covariances, originating from different viruses, are used to generate structure-guided multiple sequence alignments and increase the accuracy of these predictions. Such approaches have been implemented in tools like LocaRNA. Both single and multiple sequence-based predictions can be made using various tools included in the ViennaRNA package.

The *in silico* prediction of structures is limited by some assumptions of the underlying models. First, unpaired regions of two (or more) structures that interact with each other (i.e., pseudo-knots) are usually neglected. Second, the length of the input sequence is limited. The number of all possible structures increases exponentially with the length of the RNA sequence. In other words, the longer the sequence, the less confident the *in silico* prediction. Third, interaction sites that do not follow the canonical base-pairing model are usually not included in the prediction algorithms. New specific tools and analysis pipelines are constantly being developed and existing ones improved, in order to address known challenges. *In vitro* or *in vivo* validation of the presence of predicted structures in viruses and their biological function is of paramount importance. To this end, the close cooperation of virologists and bioinformaticians is indispensable for achieving new knowledge in the field of secondary structures in viruses.

Viral Metagenomics

Recently, a new source of viral genomic sequences has become increasingly important. Metagenomics samples genomic material directly from the environment, allowing for the reconstruction of complete viral sequences without cultivation. Early metagenomes did not allow for the assembly of large genome fragments, mostly because of a limited capacity in sequencing depth and assembly software available at the time. Hence, most analyses focused on individual marker genes or global comparison between datasets, i.e., “all-versus-all” similarity. While providing important information on the overall genetic diversity of viruses, these gene-level analyses suffered from major limitations. Specifically, gene-based approaches can only target specific groups of viruses since no universal viral marker gene exists and are thus limited in their ability to discover novel viral diversity and draw inference at the scale of whole viral communities.

An early database independent tool for cross-metagenomic comparison at the level of sequencing reads was crAss, for cross-Assembly, an approach that exploits sequence assembly to identify shared elements in different metagenomic samples. Greater sequencing depth per sample and improved bioinformatics now enable the assembly of large genome fragments and even complete genomes from metagenomes. These genomes, termed “uncultivated virus genomes” to distinguish them from genomes obtained from virus isolates, are now becoming the primary unit of most virome analyses. The community has thus recently established a set of standards and guidelines to identify, analyze, and report these genomes of uncultivated viruses in a manuscript entitled “Minimum Information about an Uncultivated Virus Genome (MIUViG)”, so that these sequences can contribute to a comprehensive mapping of viral diversity on Earth. In addition, the different tools commonly used in virus genome analysis are progressively being made available on free online data analysis platforms so that all researchers can incorporate uncultivated virus genomes into their analysis.

One of the areas that strongly depends on a thorough understanding of uncultivated viruses is viral ecology. Briefly, the first step in the analysis of viral diversity from metagenomics is to identify which of the assembled sequences are derived from viral genomes.

This step is required even when processing purified viromes, where most of the data is expected to be viral, because these can still contain a substantial fraction of contaminating cellular sequences. The second requirement is to evaluate whether an assembled contig corresponds to a fragment of a larger genome, or represents the majority and possibly the entirety of a genome sequence. This information is critical to correctly interpret these data, especially for analyses such as functional potential and taxonomic classification of the identified viruses. The current standards, comparable to the ones used for uncultivated genomes of bacteria and archaea, comprise three categories defined based on estimated genome completeness and the level of genome annotation provided: “genome fragments” are sequences representing <90% of the full genome, “high-quality draft genomes” represent $\geq 90\%$ of the genome with minimal functional annotation, and “finished genomes” are complete genomes with comprehensive annotations of the encoded functional elements. When these considerations are addressed, the relative abundance of different viral groups can then be assessed through “read mapping”, i.e., the reads sequenced in a metagenome are compared to the viral genomes, and the number of reads matching each genome is interpreted as a measure of the number of copies of this genome in the initial sample. Using this approach, traditional microbial ecology approaches can be applied to assess alpha and/or beta diversity of the viral community. Moreover, species-species interaction networks can be inferred based on the correlation of viral and/or microbial groups across samples, where network nodes reflect species and edges reflect their correlated abundance or occurrence patterns across samples. Thus, these networks summarize the information in individual metagenomes that represent temporary or spatial snapshots of ecosystems. Network-based approaches have been applied in the study of viruses and the interactions with their potential hosts, but their interpretation remains challenging. Nevertheless, they can help elucidate temporal and spatial dynamics of viral diversity as well as unravel the role of viruses in important ecological processes such as the carbon cycle.

The application of metagenomic deep sequencing and *de novo* assembly for virus diagnostics allows also distantly related viruses to be identified, provided that sensitive homology detection tools are used. The unbiased, high-resolution view of the viral diversity present in a sample that is offered by metagenomics, allows the identification of viruses in patient samples directly through agnostic sequencing. Advances have been made towards the application of metagenomics outside of the research context and into the clinic, with a growing number of studies evaluating metagenomics as a tool for animal and zoonotic diseases detection. These include discoveries of new viruses associated with deaths after organ transplantations in humans, polyomaviruses associated with Merkel cell carcinomas, encephalitis-causing viruses in cattle and other difficult-to-diagnose cases. However, these studies often reveal an array of viruses in most samples, and predicting for each of them whether they impact on human health remains an unresolved issue. There is still some ground to be covered for metagenomics to be established as a routine diagnostic test. Decreasing sequencing costs, improvements in the underlying bioinformatics tools, as well as standardization of protocols and well-defined guidelines for laboratory personnel will render this option even more viable in the future.

Evolution and Phylogenetics

Phylogenetic inference based on molecular data starts with the alignment of multiple homologous sequences that allows mutations to be identified. Next, several approaches exist that transform the knowledge of the identified mutations into a phylogenetic tree, including distance-based approaches like neighbor-joining that can be rapidly calculated, and more advanced approaches that build reliable phylogenetic trees based on an evolutionary model, including maximum likelihood or Bayesian optimization. Established statistical and mathematical models of evolution and the estimation of parameters such as substitution rates, divergence times, and other population genetics patterns are incorporated in various bioinformatic software packages such as RAXML, BEAST, and PhyML that are widely used in virus bioinformatics (Table 1).

Phylogenetic analyses of viruses encounters several challenges. First, efforts to reconstruct the “deep phylogeny” of viruses are hampered by the inability to calculate genetic distances between the highly divergent sequences of distant families. Thus, phylogenetics and phylogenomics are most successful in the context of narrowly defined groups of related genomes, whose members share a set of core genes that allow all members to be compared in a common framework. For example, the gene encoding the RNA-dependent RNA polymerase (RdRp) is the only universal gene among RNA viruses and phylogenetic reconstructions based on this have shed light into the origins and evolution of the global RNA virome. Second, phylogenetic trees of different viral genes often yield inconsistent phylogenies due to the high frequency of genomic recombination in viruses. Conventional phylogenetic trees used for graphically representing viral phylogenies are challenged by variable evolutionary rates, lack of physical “fossil records” of viruses, confounding evolutionary relationships between viruses and their hosts, high rates of horizontal gene transfer and rampant genomic rearrangements. An alternative approach to visualizing distant relationships are genome-level networks. In this context, network nodes represent virus genomes and edges are drawn between them if they share at least one gene. Using formal analytical tools the network topology can be interrogated. Such analyses have given insights in host range of phages. Furthermore, bipartite networks may also be used to depict the links between homologous gene families and genomes.

One application of phylogenetics in virology is the study and tracking of transmission networks and epidemics. Understanding the relationships between viruses can provide us with a wealth of information about when, where and how viruses are transmitted, and in what ecological or clinical context. Phylogenetic analyses can be employed to infer phylogenetic relationships between different strains, such as building a clearer picture of a viral outbreak or for the reconstruction of the demographic history of the pathogen. Transmission networks allow the analysis of evolutionary trajectories of very recently diverged strains of the range of less than years thanks to high viral mutation rates. This enables near real-time monitoring of virus outbreaks, as was shown in the

2013–2016 West African Ebola outbreak that was monitored “live” by nanopore sequencing of 142 Ebola virus genomes from patients in Guinea. Moreover, geographic mapping of 1610 Ebola virus genomes allowed the dispersal, proliferation and decline outbreak to be analyzed, revealing a heterogeneous and spatially dissociated epidemic consisting of different transmission clusters. In an epidemiological study of the 2013–2014 Zika virus outbreak in the Americas, molecular clock estimates suggested that its introduction into Brazil predated the 2014 World Cup soccer tournament and a canoeing event, potentially pointing to its introduction during the 2013 Confederations Cup soccer tournament. The integration of genomic, epidemiologic, and mobility data has led to the blossoming of the field of phylodynamics.

Another important application of phylogenetics in virology is taxonomy. Virus taxonomy is a field in flux. No standard automated viral taxonomy framework currently exists, but several computational tools have been developed that allow clustering of viral sequences and objective demarcation of the boundaries between taxonomic levels, such as DEmARC, vConTACT, VICTOR, and GRAViTy. In this context, the ICTV is now exploring genome-based taxonomy methods for different types of viruses, which would enable an integration of high-quality and finished metagenome-assembled virus genomes in the official taxonomy, and thus a better representation of viral diversity in the official ICTV classification. Taxonomic classification is grounded in phylogenetics, and different phylogenetic characters are suitable for distinguishing recent and ancient taxonomic groups, in accordance to their rates of evolution. The emerging consensus is that gene content methods similar to those developed in the beginning of the genomic era for cellular organisms are the method of choice for resolving ancient taxa. For defining recent taxa, alignment-based methods are appropriate for taxa with widely shared marker genes.

Virus-Host Interactions

In 1973, an early breakthrough in the understanding of co-evolution came with the definition of a law known as Red Queen, named after the Alice in Wonderland character. Applied to viruses, the law describes a co-evolutionary steady state in which hosts evade the viruses that infect them by mutating certain interaction molecules, while viruses also mutate to retain virulence. Thus, for both parties “it takes all the running [in genome sequence space] they can do, to keep in the same place”. Both viruses and their hosts have evolved evolvability mechanisms that boost mutations in genomic regions containing important genes involved in virus-host interaction. For example, some bacteria and bacteriophages encode mechanisms of targeted genomic hyper-variation that diversify receptor-binding proteins (RBPs). Others encode genomic islands that contain proteins involved in cell decoration, or anti-phage defense systems that may be readily gained and lost from the genome. These and other mechanisms of accelerated evolution, combined with rapidly fluctuating selection pressures, make virus-host interaction genes among the most variable elements of bacterial and phage genomes.

Traditionally, viruses were always discovered and analyzed in the context of a host, either because the host showed symptoms of the viral infection, or because the viruses were isolated by growing them in a cell culture of their host. This has changed with the advent of metagenomics, as viral genomic sequences can be identified directly from their environment (see also Section “Viral Metagenomics” above). The direct sampling of viruses without host information complicates the interpretation of the ecological roles of viruses, including fulfillment of Koch’s postulates in the case of samples from diseased organisms. Thus, host prediction is an important current challenge in understanding the role of viruses identified from metagenomics. Metagenomics may reveal sequences that are distinct from those of known viruses; thus, their hosts cannot be predicted based on similarity to viruses that have been experimentally characterized. Recent advances in machine learning hold promise for predicting virus-host interactions, including several approaches that are based on the genome sequences alone. These approaches exploit genomic signals including, for example (1) the nucleotide usage profile of the genome sequence that is adapted in viruses as a result of co-evolution with their hosts; (2) regions of sequence similarity between virus and host genomes, which could reflect integrated proviruses, horizontally transferred genes, or other mechanisms; (3) CRISPR spacers in the bacterial genome matching the genomes of bacteriophages that infected that host lineage in the past; and several other signals. Nevertheless, most virus genomes assembled from metagenomes remain without any predicted host at this point, and designing new approaches to establish these linkages remains a major computational challenge in the field.

Machine Learning as an Opportunity

As new technologies allow diverse aspects of biological systems to be measured at unprecedented scale and resolution, the rapidly increasing complexity and dimensionality of the data are increasingly challenging to interpret. Emerging analysis and processing methods based on machine learning have the ability to deal with such large, complex data sets. The main power of such approaches is their ability to identify signals and patterns in the data, enabling predictions to be made by using statistical models. Machine learning describes the process of gaining general knowledge to effectively perform a specific task by analyzing samples.

Machine learning algorithms can be divided into different types based on their strategy and the kind of problem they address, primarily including supervised and unsupervised machine algorithms. Unsupervised learning algorithms analyze high-dimensional input data for patterns without a search image. This information can be extracted by clustering the input data and can be used to determine the importance of each dimension of the input data. Common approaches including PCA and t-SNE do

this by reducing the dimensionality of the input data while preserving the information. This can lead to insightful visualizations of clusters or patterns in complex data. For example, the k-means clustering algorithm is an unsupervised method that clusters the data into k different groups. This is achieved by combining the samples with the highest similarities into one group. If clear clusters are observed, this indicates variations in the data that may need further analysis before drawing general conclusions about the complete dataset. This can help to identify potential pitfalls of an experiment represented in the data. In contrast to unsupervised learning, supervised learning approaches like random forest, gradient boosting trees and support vector machines analyze features of the input data for solutions that correspond to a pre-defined pattern represented by training data. In a genomic context, such features could include the length of the genome sequence, the amino acid distribution of the encoded proteins, the age of a sampled patient, etc. From the methodological perspective, a feature can be any property of the data that can be numerically or categorically represented. The machine learning algorithm analyses the predefined features of each sample and searches for similarities which help to solve the predefined task. The predefined task is the transformation of the input data to the desired output data. For example, given a metagenomic sequence as input, a supervised machine learning algorithm could determine whether it is derived from a virus.

A major challenge in supervised learning is to identify generalizable patterns that are predictive of new “unseen” cases. To assess the performance of such algorithms, it is good practice to split the full dataset into three parts, including (1) training data that is used to identify significant patterns; (2) validation data that is used to optimize the parameters of the machine learning algorithm; and (3) testing data that is used to assess the performance of the approach on unseen data. For fair comparison, it is important that the testing data is in no way used to tweak or optimize the procedure. The amount of data used for training, validation, and testing may differ, but could represent e.g., 80:10:10 ratio, where it is important that the data points represent independent measurements. This may be especially challenging when predicting virus-host interactions (see Section “Virus-Host Interactions” above), because the viruses with known hosts that are present in the database are highly skewed for a few well-described groups. If left unaccounted for, this database bias leads to inflated performance statistics for virological machine learning predictors.

Recently, representation learning methods have gained a lot of attention due to their extraordinary ability to represent complex information in statistical models. Approaches including deep learning and transfer learning have the advantage that very complex data can be analyzed and processed, with only a minimal requirement for the user to define features. Such approaches are ideal for analyzing big omics scale datasets, but require large amounts of training data and concomitantly heavy computing power, orders of magnitude more than the classical approaches. To summarize, machine learning enables us to analyze, understand, and evaluate the huge amounts of data that are becoming available through technological innovations.

Host Transcriptomics

Viruses only come to life after infecting their cellular host. Understanding the host response is of utmost importance for the investigation of viral infections. Once the virus enters the host system (either a living organism or a virus-responsive cell line), highly specific regulators identify the threat and then stimulate the expression of a cascade of genes. For example, in the case of an RNA virus infection, antiviral type I interferons (IFN- α /beta) bind to their receptors, thus activating specific transcription factors and promoting the expression of several IFN-stimulated genes (ISGs) with antiviral and immunomodulatory activity.

Today, transcriptomics (RNA-Seq) is widely used to study the host response to viral infections. To this end, total RNA extracted from, e.g., uninfected (mock) and virus-infected host cells can be reverse transcribed into cDNA (complementary DNA), fragmented, and sequenced. Short-read sequencing technologies generate a high-resolution expression profile consisting of millions of short sequencing reads that represents the composition and relative amount of RNA molecules, together making up the transcriptome. Different computational approaches are combined to build bioinformatics pipelines to process and analyze these short-read data comprehensively. In the case of an available host (or closely related) reference genome, the reads can be mapped back to identify their origin and thus the origin of their corresponding RNA expression.

If no reference genome is available, RNA-Seq data can also be assembled *de novo* and subsequently characterized, to identify differentially expressed genes. A promising avenue in transcriptomics are long-read sequencing technologies such as offered by PacBio Iso-Seq and Oxford Nanopore technologies. The latter has already been used for sequencing and investigation of full-length host and virus transcripts in their native RNA form.

A gene that is more strongly transcribed during viral infection statistically yields a higher number of short reads after sequencing. Therefore, RNA-Seq not only allows the identification of transcribed host genes but also provides a quantitative value. The number of reads of the same gene derived from different conditions (e.g., mock versus infected) can be compared to identify differentially expressed genes. After normalization (taking into account different sequencing library depths and/or gene lengths), fold changes and their significance are calculated for each gene and between replicated conditions. To this end, tools such as DESeq2 and Sleuth, provided as R packages, are used to statistically evaluate the quality checked, mapped, and quantified RNA-Seq data. **Fig. 1** shows an exemplary RNA-Seq bioinformatics pipeline for the calculation of differentially expressed genes between mock, IFN-treated, and virus-infected cell lines of a microbat species (*Myotis daubentonii*) at two different times after treatment. Since no genome of this bat species was publicly available, the genome of a close relative (*M. lucifugus*) served as a reference for mapping and quantification.

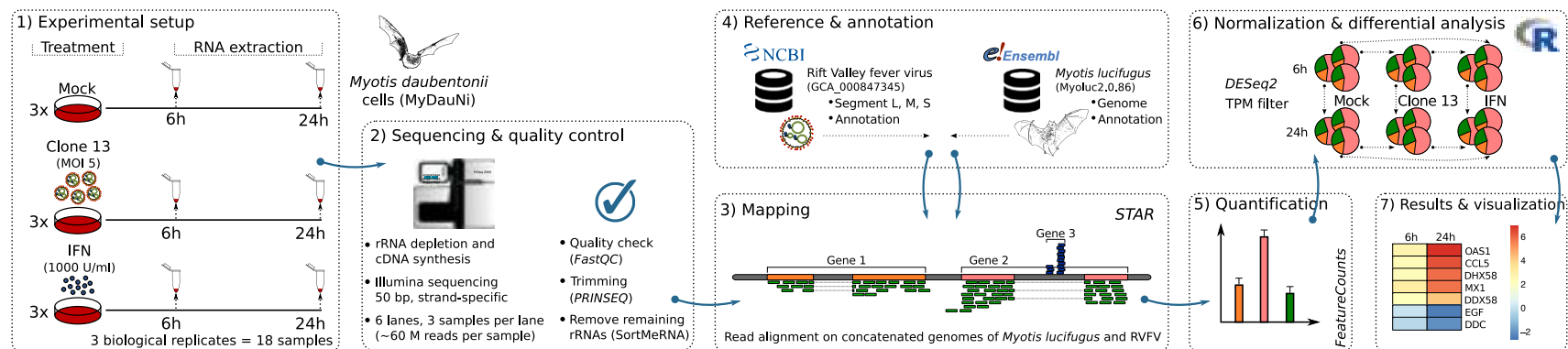


Fig. 1 Overview of an RNA-Seq bioinformatics pipeline for the identification of differentially expressed genes based on a study of an IFN-treated and virus-infected (RVFV Clone 13) bat cell line (*Myotis daubentonii*). Samples were taken in triplicates at two time points and sequenced with Illumina. Reference genomes for the virus and a close relative bat species were obtained from NCBI and Ensembl, respectively.

Conclusions

Computers are not only indispensable to analyze data in virology, but also to store and distribute the large volumes of data generated in a reproducible way. Efforts into making the unprecedented amounts of data Findable, Accessible, Interoperable and Reusable (FAIR) should also be applicable in the field of virology. Several resources are currently available that are dedicated to viral specific sequence information and their associated metadata (Table 1). Currently, most analyses require a reference database, be it sequence based similarity searches for the identification of viruses, functional annotation of protein sequences or genome based phylogeny and classification. Making data publicly available in databases is important, not only as part of their general interest. As new methodologies are being developed, the available data can be further mined or reanalyzed to extract new information. A prominent example is the discovery of a highly abundant phage in the gut of humans, crAssphage; another is the recent suggestion that small circular single-stranded DNA smacoviruses infect Archaea instead of humans. The timely deposition of genomic data and their availability to the public domain is also crucial in the epidemiological context where this information can be used for continuous surveillance, designing effective diagnostics, vaccines and antibody-based therapies.

Bioinformatics opens up a vast range of possibilities for new analyses and interpretations of viruses. While computational predictions always need to be validated by relevant *in vitro* experimental follow-up, the unprecedented availability of big omics datasets in the public domain already allow bioinformaticians to perform many initial validations *in silico*. These best practices can be used to estimate the accuracy of diverse bioinformatics tools, providing an important focus for wet laboratory experiments and saving valuable time and resources. Thus, bioinformatics has already become an integral and transformative component of virus research, much like techniques such as culturing, microscopy, and molecular biology have done in the past.

Further Reading

- De Jonge, P.A., Nobrega, F.L., Brouns, S.J.J., Dutilh, B.E., 2018. Molecular and evolutionary determinants of bacteriophage host range. *Trends in Microbiology* 27, 51–63.
- Edwards, R.A., McNair, K., Faust, K., Raes, J., Dutilh, B.E., 2016. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiology Reviews* 40, 258–272.
- Hall, R.J., Draper, J.L., Nielsen, F.G.G., Dutilh, B.E., 2015. Beyond research: A primer for considerations on using viral metagenomics in the field and clinic. *Frontiers in Microbiology* 6, 224.
- Hölzer, M., Marz, M., 2016. Differential transcriptional responses to Ebola and Marburg virus infection in bat and human cells. *Scientific Reports* 6, 34589.
- Hölzer, M., Marz, M., 2017. Software dedicated to virus sequence analysis: "Bioinformatics goes viral". *Advances in Virus Research* 99, 233–257.
- Jaafar, Z.A., Kieft, J.S., 2019. Viral RNA structure-based strategies to manipulate translation. *Nature Reviews in Microbiology* 17, 110–123.
- Marz, M., Beerenwinkel, N., Drosten, C., *et al.*, 2014. Challenges in RNA virus bioinformatics. *Bioinformatics* 30, 1793–1799.
- Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., Koopmans, M.P.G., 2018. Overview of virus metagenomic classification methods and their biological applications. *Frontiers in Microbiology* 9, 749.
- Roux, S., Adriaenssens, E.M., Dutilh, B.E., *et al.*, 2019. Minimum information about an uncultivated virus genome (MIUViG). *Nature Biotechnology* 37, 29–37.
- Roux, S., Emerson, J.B., Eloe-Fadrosh, E.A., Sullivan, M.B., 2017. Benchmarking viromics: An *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5, e3817.
- Siddell, S.G., Walker, P.J., Lefkowitz, E.J., *et al.*, 2018. Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018). *Archives of Virology* 164, 943–946.
- Simmonds, P., Adams, M.J., Benkő, M., *et al.*, 2017. Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology* 15, 161–168.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9, e02329-18.

Relevant Websites

- <http://metavir-meb.univ-bpclermont.fr/>
Analysis of viromes.
MetaVir.
- <https://www.cyverse.org/>
CyVerse: Home.
- <https://img.jgi.doe.gov/cgi-bin/vr/main.cgi>
Ecosystems.
- <https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses (ICTV).
- <http://dmk-brain.ecn.uiowa.edu/pVOGs/>
The pVOGs Database.
- Prokaryotic Virus Orthologous Groups.
- <https://www.ncbi.nlm.nih.gov/genome/viruses/>
Viral Genomes.
NCBI.
- NIH.
- <https://viralzone.expasy.org/>
ViralZone root.
- <http://viromes.org/>
Viromes.
- <http://kbase.us/>
Welcome to KBase Predictive Biology | KBase.

Metagenomics in Virology

Simon Roux, Department of Energy Joint Genome Institute, Walnut Creek, CA, United States

Jelle Matthijnssens, Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

Bas E Dutilh, Utrecht University, Utrecht, The Netherlands and Radboud University Medical Center, Nijmegen, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Metagenomics The study of genetic material (DNA or RNA) extracted from an environmental sample. Recent studies use “shotgun” metagenomics, i.e., the untargeted sequencing of all genomes from all members in the sampled community.

Viromics Viral metagenomics, i.e., shotgun metagenomics applied specifically to the encapsidated fraction of DNA and/or RNA from a sample. The encapsidated fraction is typically obtained through a combination of filtration, precipitation, and DNase/RNase treatment.

Metagenomics Applied to Viruses

Historically, viruses have been primarily explored using laboratory cultivation: new viruses were obtained from clinical or environmental samples through propagation and isolation on cell cultures. This process is, however, biased and challenging to apply at large scales because (i) many viruses depend on host cells that are difficult to maintain as clonal culture in the laboratory, and (ii) even if the cells are available, propagating viruses may require specific conditions distinct from those used to cultivate the cells. These considerations are especially meaningful for viruses with microbial hosts, the vast majority of which remain uncultivated to date.

Metagenomics bypasses this requirement for cultivation and instead relies on the sequencing of viral genomic material extracted directly from a sample (see [Box 1](#) and [Fig. 1](#)). Thus far, the history of viral metagenomics has seen two major phases. Initially, entire communities of viruses were assayed by analyzing and comparing short sequencing reads obtained from diverse environments. Because of the fragmented nature of these data, most of these studies had to be conducted at the community scale, and identifying and distinguishing individual viruses in these datasets remained challenging. More recently, bioinformatic advances have enabled the reconstruction of individual viral genome sequences from metagenomes, allowing naturally occurring viruses to be identified and studied at high, genomic resolution. Using a metagenomics approach, entirely new types of viruses can now be discovered, surveyed, and characterized even without cultivation. The unique ability offered by metagenomics to study uncultivated viruses led to the emergence of two parallel and interconnected fields: a clinical one, where metagenomics promises to be a catch-all method for the unbiased surveillance and diagnosis of viral pathogens, and one focused on natural biomes, that aims to describe the diversity of the viral world and understand its ecological and evolutionary drivers and impacts.

Pioneering Viral Metagenomics, One Gene at a Time

When the field of shotgun environmental metagenomics was pioneered in 2002 by the laboratory of Forest Rohwer at San Diego State University, the first datasets consisted of three viral metagenomes (viromes) that, together comprised just under 2,500 short genomic fragments derived from two natural marine viral communities and one human feces sample. While limited in scope and resolution, these and other early viromes provided an unprecedented view of complex viral communities in nature. Both oceanic and human fecal viromes pointed toward the existence of an extensive virus diversity. This diversity of the virosphere was estimated by comparing the sequencing reads within each metagenome, and observing that almost every fragment was unique. Moreover, comparing the short sequencing reads to a reference database of known viral genomes sequences revealed that up to 99% was not similar to any known virus, suggesting that most of the virosphere was yet to be discovered. This uncharted genomic biodiversity became popularly known as “viral dark matter”.

In the years that followed, a broader range of environments was progressively surveyed using viromics including freshwater lakes, hot springs, agricultural soils, or human skin, saliva, and gut samples. Improvements in DNA sequencing technologies, especially the advent of the popular pyrosequencing platform, that has since been surpassed and discontinued, increased the scale of these datasets by providing hundreds of thousands of short genomic fragments for each sample. By directly comparing the sequences across these datasets, several studies indicated that virus genes tend to structure by environment rather than by sample location, implying that some of these genes may be globally distributed. In addition, when sampled from the same freshwater and hypersaline ponds across several days, weeks, and months, viral metagenomes revealed that the genetic composition of viral communities was coherent at a broad level, but some individual viral genes experienced rapid changes in relative abundance.

Scaling up From Fragmented Genes to Complete Genomes

While the analyses outlined above were foundational for our current understanding of virus diversity, they were limited by the short length of next-generation sequencing reads which fragmented the view of viral genomes. These limitations were progressively

Box 1 Use of complementary methods to target different types of viruses

A number of approaches have been developed to specifically select and survey the genetic material contained by virus particles in a given sample. Alternatively, viral genomes can also be analyzed from “bulk” metagenomes which include both virus particles and microbial cells. Virus sequences obtained from “bulk” metagenomes will typically reflect viruses infecting their host cell at the time of sampling, either actively replicating or not, while viromes enables a deeper and more focused exploration of the virus diversity in a specific site or sample.

Regardless of the type of sample, viromes are most often generated through a combination of centrifugation, filtration and DNase/RNase treatment, aiming at removing as much of the cellular genomes as possible (Fig. 1). A typical protocol will notably include a filtration through 0.22, 0.45, or 0.8 µm membrane filters to remove bacteria and larger cells. Depending on the initial concentration of virus particles, a concentration step using e.g., iron chloride (FeCl₃), PolyEthyleneGlycol (PEG), or tangential flow filtration step(s), may be necessary to obtain enough material for sequencing library preparation. Cesium chloride density gradients ultracentrifugation can also be used to further separate viruses from extracellular DNA and large particles in complex samples, although this step can also lead to a substantial loss of viral material. Finally, the virus particles obtained are typically treated with DNase or RNase to remove free DNA and/or RNA. Depending on the type of virus studied, the corresponding protocols for RNA or DNA extraction and sequencing library preparation are then applied, after releasing the genetic material from the virus particle through e.g., a heat shock if necessary.

A critical step in this process is to recover enough material for sequencing. While micrograms of DNA were initially needed, several protocols are now available which only require ~ 1 ng of DNA. In addition, a DNA/RNA random amplification step, called “whole genome amplification”, can also be conducted in order to gather enough input material. This type of approach was initially used in almost every virome study, and revealed important information for example on the unsuspected diversity of ssDNA genome viruses in the environment (see below). However, the whole genome amplification process is inherently biased, and these datasets are not quantitative, i.e., one cannot draw any conclusion about the relative abundance of the viruses identified in these amplified metagenomes. Thus, whole genome amplification methods have now been often replaced by advanced library preparation protocols which require nanogram-scale input but enable quantitative datasets well suited for ecological studies. Alternatively, for cases in which target viruses represent a minor part of the templates, targeted sequence capture approaches have been used, mainly in a clinical framework as they can only be applied to viruses with known genomes but can detect these viruses with a very high sensitivity.

The recovery of virus genomes from bulk metagenomes and from viromes each have their own limitations. For bulk metagenomes, viruses typically represent only a minor fraction of all sequences compared to cellular genomes. This means that the virus genomes obtained this way will tend to be restricted to abundant viruses found in their host cells, while viruses that are not infecting at the time of sampling, viruses with a low frequency of infected hosts, or viruses infecting rare hosts will likely be missed. Viromes provide a deeper description of the virus community, since most of the sequencing data will be obtained from virus genomes. In addition, virus particles will not represent only current infections but a more integrated sampling of all recent successful infections, the timing of which depending on the type of sample and the individual virus decay rate. Yet viromes still suffer from several biases. Notably, the size-based selection of virus particles excludes most of the larger viruses such as the “giant viruses”, and viromes also tend to be dominated by viruses with high burst size while under-sampling viruses with low burst size and long infection time. All metagenomes (bulk and viromes) will struggle with very rare viruses, as well as hypervariable viruses which genome will not assemble well. Hence complementary approaches such as targeted capture approach for the former, and long read sequencing for the latter, are being developed (Fig. 1).

Overall, the different methods developed over the last decade to sequence genomes from uncultivated viruses are mostly complementary and can be individually tailored for specific applications. Virus discovery can be achieved through bulk metagenomes or viromes, while viral ecology studies will tend to rely more on viromes as a reflection of virus activity and transport, and metagenomics used as a diagnostic tool in the clinic would be the most likely to use sequence capture. Nevertheless, all these complementary approaches will be needed for achieving a comprehensive picture of viral diversity.

overcome through an increase in sequencing throughput associated with improvements in sequence assembly and analysis tools. The first large-scale assemblies of viral genomes from short metagenomic fragments were published around 2010, and quickly became a standard analysis in the viral metagenomic field so that by the year ~ 2015, complete or near-complete virus genomes were routinely reported and analyzed in viromics studies.

In only a couple of years, metagenomics has thus transformed the way scientists can identify and study viruses in the environment, as illustrated by the quick rise of virus genomes and genome fragments assembled from metagenomes available in public databases (Fig. 2). In 2010, only 84 viral genomes (fragments) assembled from metagenomes were publicly available, while this number reached 35,000 in 2016, and 775,000 in 2018. Genome sequences of uncultivated viruses are frequently obtained not only from viral metagenomes, i.e., metagenomes from virus-targeted samples, but also from “bulk” metagenomes in which virus particle were not enriched and viral and microbial sequences are mixed. Combined with genome sequences obtained from isolates, these “uncultivated virus genomes” represent the foundation of an extensive mapping of the viral sequence space.

Viral Metagenomics in the Clinic

Over the past few decades, a number of molecular techniques, such as (q)PCR or ELISA, have been developed and used to detect pathogenic viruses in clinical samples. However, these techniques can only detect previously known viruses, and often require

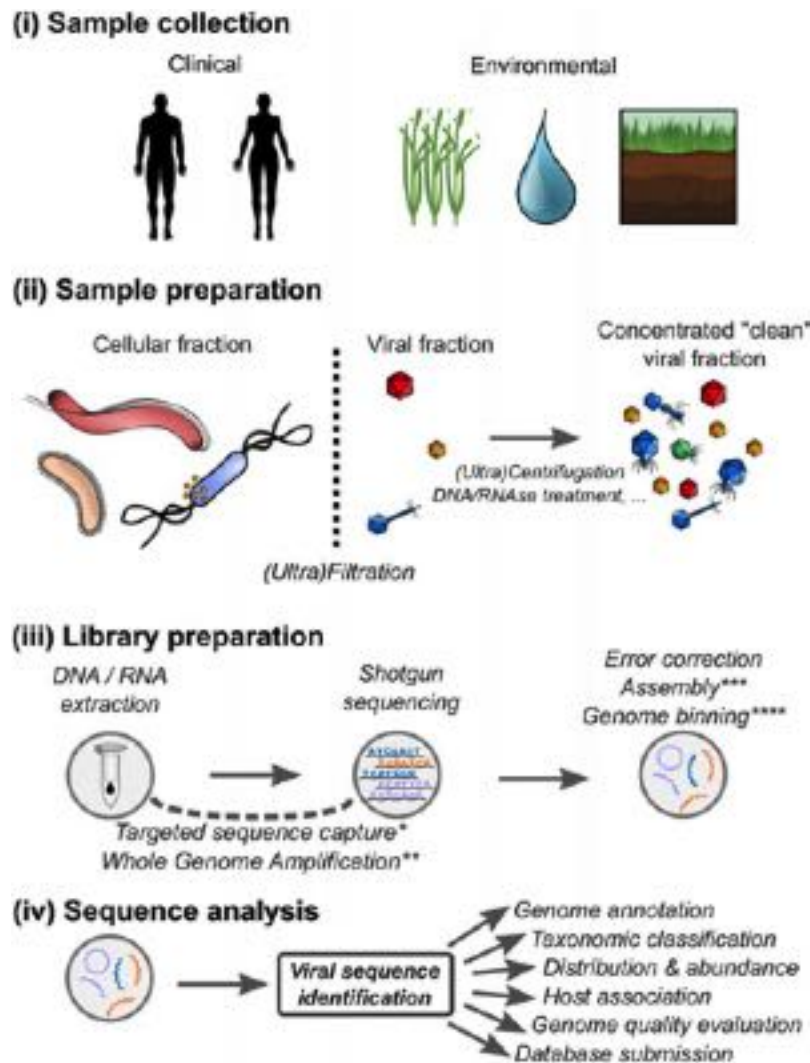


Fig. 1 Overview of the viral metagenomics workflow. The overall process used to generate and analyze viral metagenomes can be divided into four major steps: (i) Collection of environmental and/or clinical sample, (ii) sample preparation, (iii) library preparation, and (iv) sequence analysis. The sample preparation step can target either the cellular fraction (left) or the viral fraction (right) in which case viral particles are often further concentrated and purified to remove free nucleic acids. *Targeted sequence capture can be applied to the extracted DNA/RNA to enrich for a specific virus. **While whole genome amplification was initially used routinely for viral metagenomes, it has now been supplanted by methods enabling the preparation of more quantitative libraries from low input (~ 1 ng), hence whole genome amplification is now primarily used in single-cell or single-virus-particle experiments. ***The genome assembly can be bypassed if using long-read sequencing technologies, although these long-read datasets require a more careful error correction. ****Genome binning, i.e., the identification of multiple contigs assembled from a metagenome and corresponding to the same genome, is typically only used for large genomes (e.g., > 500 kb), and individual contigs are directly analyzed instead for most viruses.

specific assays for each pathogen. Metagenomics instead offers the possibility to detect known and novel viruses without prior knowledge from a single analysis, and is thus well suited and already applied to study emerging and/or rare viruses, as well as cases which remain negative using the available diagnostic tests (see below). A number of challenges remain however for viral metagenomics to become a standard clinical procedure. First, given the current cost associated with sample processing and sequencing, metagenomes are still more expensive and slower than ELISA's or qPCR assays. Second, there is no generally validated bioinformatics pipelines than can perform a rapid, sensitive, and specific analysis of the obtained data on a bench top computer. Finally, physicians will have to be trained and guided to deal with the obtained breadth of data. Specifically, it is becoming clear that each individual is chronically infected with a dozen or more eukaryotic viruses (many of which have not been associated with any disease, e.g., anelloviruses), and that many known viral pathogens can also cause asymptomatic infections. Therefore, a physician might get a list of viruses (and other potentially pathogenic or unknown organisms), and it will be a challenge to identify the actual cause of a particular disease. Nevertheless, the price of sample preparation and high throughput sequencing has declined enormously in the last decade including with the development of smaller and faster machines, while automatic virome

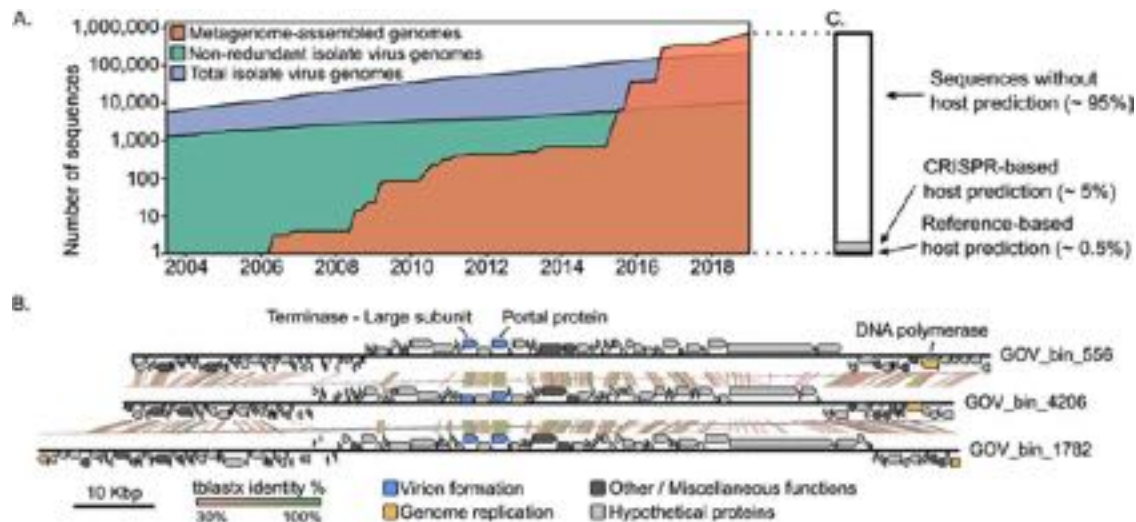


Fig. 2 Size of virus genome databases over time, host linkage information, and examples of uncultivated virus genomes. A. The total number of genomes from isolates was based on queries to the NCBI nucleotide database portal, while the number of uncultivated virus genomes was estimated by compiling data from the literature and from the IMG/VR database. The number of sequences is displayed on a log₁₀ scale. B. Comparison of 3 complete viral genomes assembled from viral metagenomes sampled from the Indian, Pacific, and Atlantic oceans, through the *Tara* Oceans expedition. These sequences were identified and analyzed as part of the “Global Ocean Virome” dataset (GOV). Predicted genes are colored by functional annotation. C. Overview of the host predictions available for uncultivated virus genomes in the IMG/VR database. Host prediction was based on signals including sequence similarity with isolate viruses, prophages, and CRISPR spacers derived from known bacterial and archaeal genomes.

analysis pipelines are being actively developed, so that metagenomics will likely be available in the near future as a routine test allowing physicians to get a viral diagnosis from a biological sample in a matter of minutes to hours in their home office or on the clinic bedside.

Metagenomic Discovery of New Viral Pathogens

Currently, metagenomics is most often used in a diagnostic context when both conventional and enhanced molecular testing fail to identify a causative agent in a sample. These cases can represent a significant fraction of patients for diseases such as acute diarrhea, for which an etiological agent is identified in only ~ 60% of cases. In this framework, metagenomic analysis can lead to the discovery of unexpected or novel viruses that are associated with a specific set of symptoms.

First, metagenomics can successfully identify known viruses in unexpected sample types. These studies include the detection of enterovirus D68 in clinical samples (rectal, throat, and oral swabs as well as blood samples) in cases of acute flaccid paralysis, the detection of herpes simplex virus 1 (HSV-1) in cerebrospinal fluid samples of a patient with encephalitis, and the detection of mumps vaccine virus from the brain biopsy of a patient with chronic encephalitis. In addition, new human pathogens only distantly related to known viruses have also been discovered with metagenomics. These include the Bas-Congo virus, a rhabdovirus that was associated with a 2009 hemorrhagic fever outbreak in the African Congo, as well as novel rhinovirus, bocavirus, arenavirus, and parechoviruses. Finally, entirely novel types of potentially pathogenic viruses have been described through metagenomics, including previously unknown cycloviruses, cosaviruses, and klasseviruses.

Diagnostics through viral metagenomics has also been applied to non-human animals as well as plants, and similarly revealed new potential viral pathogens in organisms showing unexplained symptoms. Multiple new virus types including novel parvoviruses, polyomaviruses, sapoviruses, and picomaviruses were for example identified in livestock samples (porcine and bovine), while a large diversity of persistent RNA viruses were newly identified across several groups of plants. However, it is important to note that the detection of a (novel) virus in a sample from a patient with an illness of unknown etiology does not prove causation, even in cases of a demonstrated significant association between the presence of the virus sequence and the observed symptoms. Hence, metagenomics will often be the first step of a longer process involving attempts to propagate the virus in culture, or monitoring healthy individuals exposed to the suspected pathogen (see below “Future of viral metagenomics: major challenges and upcoming innovations”).

Epidemiological Surveillance and Environmental Monitoring

In parallel to the diagnosis application, metagenomics is also very well suited for environmental surveillance. Species representing important reservoirs of viruses with high zoonotic pandemic potential such as mosquitoes, rodents, and bats have been specifically targeted in this context. A recent study investigating the virome of more than 200 invertebrate species (a fraction of known invertebrate species), identified more than 1,400 novel RNA viruses, exemplifying that the diversity of unknown eukaryotic viruses

in the environment is enormous and only poorly characterized. Since the majority of human pandemics have a zoonotic origin, one hope is that such metagenomic surveillance will allow a faster identification of novel pandemic viruses during outbreaks, as well as identify their natural reservoirs. This knowledge is crucial for an appropriate and fast response from a medical and global health perspective. As an example, in the last two decades zoonotic coronaviruses were able to jump from bats to humans and pigs. Both the SARS (Severe Acute Respiratory Syndrome virus) and MERS (Middle East Respiratory Syndrome) viruses caused large-scale disease outbreaks in humans, whereas SADS (Swine Acute Diarrhea Syndrome) caused an epidemic in the swine industry. Ongoing efforts to characterize the virome of such reservoir animals will facilitate the implementation of control measure to prevent epidemics or enforce appropriate actions to stop ongoing epidemics. In an ideal situation, obtained environmental virome data in combination with biochemical experiments could help with the early identification of candidate viruses with the potential to transfer to a human host. For instance, a combination of metagenomics and DNA synthesis-based experiments revealed that a novel coronavirus (WIV1-CoV) initially detected in bat samples could be prime for transfer and emergence into human hosts.

Metagenomic analysis can also be leveraged in response to viral outbreaks, for example to rapidly determine viral subtypes in a novel infection source. This has been applied to cases of influenza infections as well as for a novel wild type Ebola virus outbreak, for which metagenomic approaches could correctly identify the causative agent, even in cases where traditional methods were unsuccessful because the wild type virus was too distantly related to known Ebola viruses. A correct and rapid identification of these viruses could enable the application of the correct therapeutics and guide preventive efforts against potential epidemics.

Characterizing the Global Viral Diversity

While viruses of humans, animals, and plants may have direct clinical or economic relevance, the vast majority of the (estimated) 10^{31} virus particles on Earth infect micro-organisms, including bacteria, archaea, protists, fungi, and other environmental microbes. Initial studies of environmental viral diversity focused on human feces, coastal and open ocean, freshwater lakes, as well as hypersaline and hot geothermal ponds, because protocols for efficient separation of virus particles from microbial cells were first developed for aquatic samples. Importantly though, recent innovations and technology improvements now enable application of viromics to more complex samples such as soil, groundwater, or ice cores, helping to expand our view of global viral diversity both in the human microbiome and in the environment.

Identifying Globally Dominant Bacteriophages

A striking example of a viromics discovery is that of a highly abundant bacteriophage, named “crAssphage”, that was assembled from a set of human fecal viromes in 2014. The crAssphage genome was identified by combining information from 12 individual viromes, which yielded a high-confidence 97 kb sequence with matching 5' and 3' ends, suggesting that it represented a complete circular genome. This crAssphage genome was mostly unrelated to any isolated phage genome known at the time: from the 80 predicted proteins, less than half (39) were even remotely similar to known proteins or domains, and only 25 had a predicted function, such as “phage structural protein” or “DNA helicase”. While clearly novel, crAssphage was also found to be uniquely abundant and ubiquitous: its genome was detected across 940 metagenomes, primarily from human feces, at average levels that were six times higher than all other known phages combined. By applying several independent computational host-prediction approaches, a bacterial host (*Bacteroides*) was predicted. Thus, in this instance, metagenomics revealed what remains to date the most abundant and widespread phage associated with the human gut microbiome, which had until then evaded detection through classical approaches like laboratory cultivation and PCR.

Assembling genomes of uncultivated viruses can not only identify some of the most abundant and widespread viruses in an ecosystem, but these sequences also represent foundational data for targeted follow-up experiments aimed at further characterizing these novel viruses. In the case of crAssphage, two major studies leveraged this initial genome sequence to better understand the diversity and host of these phages. First, predicted proteins from the original crAssphage genome were used as “bait” to identify related phages in a broad range of metagenomes. This revealed an extensive and diverse group of “crAss-like” phages predicted to represent a new family within the *Caudovirales* order, that may be related to *Podoviridae*. Genome comparisons within this new family also enabled the identification of conserved structural and replication gene modules. Meanwhile, another study was able to isolate a member of the crAssphage-like family through broth enrichment on *Bacteroides intestinalis* strains isolated from human gut samples, confirming the computational predictions from bioinformatic analyses that these phages were likely infecting *Bacteroidetes* hosts and had a *Podoviridae*-like morphology.

In 2016, a comprehensive effort to chart viral diversity across the global oceans yielded a similar observation. This study detected more than 15,000 viral genome fragments, and grouped them into clusters of closely related viruses, approximately consistent with genera in the viral taxonomy. Two out of the four most highly abundant and ubiquitous clusters were entirely novel and had not been described before, while the other two were similar to known bacteriophages. With viral metagenomics being applied to a larger set of samples and environments, and with bioinformatic analyses including genome assembly and interpretation constantly improving, novel groups of dominant and widespread viruses may thus be progressively revealed across many environments.

Unveiling New Uncultivated Giant Viruses

Another group of viruses whose known diversity has been vastly expanded through metagenomics are the so-called “giant viruses”, dsDNA viruses with a uniquely large virion ($\sim 0.5\text{--}1\ \mu\text{m}$) and genome (often $> 1\ \text{Mb}$), blurring the boundaries between “simple” viruses and “complex” cellular life. Following the isolation and characterization of the first giant virus in 2004 (“*Acanthamoeba polyphaga* mimivirus”), around 50 other members of this group have been isolated, the vast majority by using an *Acanthamoeba* host. However, metagenome analyses suggest that the true diversity of giant viruses vastly exceeds the number of isolates.

As early as 2013, an analysis of 17 metagenomes revealed that giant viruses could be found in the ocean at concentrations of $\sim 10^4$ genomes/ml. These initial studies were based on the detection of marker genes, since the technologies available at the time did not enable the assembly of complex and large genomes like those of giant viruses. More recently, four complete or near-complete giant virus genomes could be assembled from metagenomes of a wastewater treatment plant. This revealed a new subgroup of giant viruses named Klosneuviruses which comprised some genomes with the largest set of translation system components found at the time in any virus, including aminoacyl transfer RNA synthetases with specificity for all 20 amino acids. Undoubtedly, as our collective ability to assemble large genomes from metagenomes increases, the giant virus diversity will keep expanding.

Revealing the Extraordinary Diversity of ssDNA and RNA Viruses

While most sequencing technologies are designed for dsDNA templates (see [Box 1](#)), our knowledge of single-stranded DNA (ssDNA) and RNA viruses has also been transformed by metagenomics. In both cases, specific sample processing steps are required to access these genomes, however their relatively short length (usually $< 20\ \text{kb}$) means that complete genomes are routinely assembled from total community shotgun metagenomes that target all the nucleic acids in an environment. As for dsDNA viruses, metagenomics revealed that these ssDNA and RNA viruses were much more diverse and broadly distributed than previously inferred from isolation and cultivation approaches.

Enrichment for circular ssDNA viruses can be achieved through phi29-based whole genome amplification, which is known to over-amplify small circular ssDNA templates. Pragmatically, this translates into viral metagenomes that are dominated by ssDNA viruses with circular genomes, which helped shed a new light on the diversity of two major groups: bacteriophages from the *Microviridae* family, and eukaryotic viruses from the CRESS DNA (Circular REp-encoding ssDNA) supergroup. The latter saw the more striking expansion: until 2009, these viruses were known exclusively in plants and vertebrates, specifically pigs and birds, yet in less than a decade, CRESS DNA viruses were detected in metagenomes sampled from primates, arthropods, and unicellular protists, as well as diverse aquatic, terrestrial, and man-made ecosystems. Hence, while the exact host range and impact of these viruses remain to be fully characterized, metagenomics already revealed that ssDNA viruses are ubiquitous and can be found associated with all types of cellular hosts.

For RNA viruses, several additional sample processing steps have to be performed to preferentially sequence viral RNA, typically including reverse transcription and random amplification. The most comprehensive study of RNA virus diversity to date included samples from > 220 invertebrate species across 9 phyla, and led to the discovery of nearly 1,500 novel viruses across the 13 major clades of RNA viruses. In addition, the assembly of complete genomes provided new insights on the recombination patterns of these viruses, highlighting a remarkable propensity of RNA viruses to exchange or acquire genes horizontally, both with other viruses and with their host. RNA viruses were also detected in a much broader host range than currently known from isolates, although these host associations now have to be confirmed through laboratory experiments since virus detection in metagenomes does not equate active infection.

Leveraging Time Series to Track Virus Populations Dynamics

Improvements in metagenomics protocols post ~ 2015 enabled the analysis of dozens of samples in parallel. In the field of viral metagenomics, this increased capacity has been leveraged specifically to analyze viral signal along time series and thus investigate virus-host dynamics in nature. Such datasets have notably been obtained from freshwater lakes, for which recurrent sampling across months or years can be done, and which usually harbor a high concentration of viruses. These first explorations of environmental viral diversity across months and seasons indicated that viruses display a large range of relative abundance patterns, from “ephemeral” ones with a single peak in abundance to “constitutive” ones detected in virtually all samples. Some of these patterns were seasonal and possibly linked to similar patterns of abundance for their microbial hosts, while other viruses displayed drastic changes from one year to the next. For instance, although longitudinal virome studies of the human gut are scarce, available data suggests a rather stable population of gut viruses (almost exclusively phages) in adults over time, whereas the infant gut virome is much more variable and may be dominated by eukaryotic viruses at particular time point coinciding with an acute enteric infection.

Time series metagenomes are especially interesting to discover and predict virus-host associations, and to analyze dynamics of known virus-host pairs. The former approach already provided host prediction for several giant viruses that are so far known exclusively from metagenome assemblies, and suggested that these may be linked to uncultivated protist hosts. The latter raised the intriguing possibility of complex and diverse virus-host relationships occurring in nature: while the expected patterns would be a strong correlation between virus and host abundances with possibly a short lag in the virus signal in a typical predator-prey

fashion, these large-scale metagenome time series instead suggested that some of the viruses could peak prior to a peak in abundance of their host, while other virus-host pairs showed no similarity in relative abundance at all. These conflicting results likely reflect the complex interactions at play between viruses and microbes in nature, including variable host ranges, from viruses infecting a unique host strain to others infecting multiple host species sometimes across different genera, as well as the spectrum of infection dynamics from fast-acting lytic viruses to slower, temperate, and even chronic ones, and the development of resistance to the virus among the host population. Despite these numerous challenges in their analysis, time series metagenomes are poised to become a key approach to complement laboratory experiments and untangle the intricate relationships between viruses and their hosts.

Future of Viral Metagenomics: Major Challenges and Upcoming Innovations

Metagenomics has quickly become a major tool for exploring viral diversity, yet several challenges need to be addressed in order to fully leverage the potential of these methods. First, metagenomes built from limited input material are still difficult to reliably obtain and interpret, and do not yet provide a comprehensive and quantitative view of the viral community present in the sample. This includes environments with low biomass such as some human tissues, hydrothermal vents, ice cores, or ancient samples, but also samples with a thick substrate or matrix to which cells and virus particles tend to adhere such as human lung mucus or coral samples. Improvement in the recovery of cells and virions from this type of substrates and in the generation of quantitative libraries from sub-nanogram input will help better survey these viral communities.

The second major challenge lies in the absence of direct host information for genomes assembled from metagenomes. In a clinical context, this means that one of Koch's postulates, which requires that the candidate etiological agent be isolated from a diseased organism and grown in pure culture, cannot be fulfilled. Already, several smacoviruses which had been detected in human samples metagenomes and suspected to represent new human viral pathogens have been found to likely infect prokaryotic cells from the human microbiome instead. In a similar way, evidence is emerging that picobirnaviruses, which are believed to be eukaryotic viruses, might actually infect bacterial cells. These examples should thus serve as a cautionary tale when trying to detect entirely new viral pathogens from mixed samples containing both human and microbial cells. A modified Koch's postulate for the metagenomic era has been proposed in which potential new pathogens must first be present and more abundant in the diseased subject compared to matched control. Then, experiments using either a sample from a disease subject or an artificial virus obtained through DNA synthesis and expression in cell cultures must be performed to demonstrate that this agent induces disease in another healthy subject. While not trivial, these additional experiments based on metagenomic results could still lead to the identification of viral pathogens much more quickly than classic culture techniques.

In an ecological context, associating uncultivated viruses to their host is also critical to understand their impact on microbial communities and to meaningfully integrate viruses into ecosystem models. Because viral ecology studies typically include hundreds to thousands of viruses of interest, these host associations are typically derived from *in silico* approaches based on various types of genome sequence comparison. While methods for *in vitro* confirmation of these metagenome-derived virus-host pairs are currently being developed, they will need to improve both in terms of scale and resolution to provide meaningful host association for the vast diversity of uncultivated viruses.

Among the expected technological improvements, two stand out as likely to benefit the field of viral metagenomics in the near future. First, long-read sequencing technologies are progressively amenable to the sequencing of environmental viral communities. Pragmatically, this means that instead of having to assemble virus genomes from short reads, a process which can yield potentially erroneous and/or incomplete genome sequences, a complete viral genome could be sequenced as a single read. Once broadly available, these long-reads metagenomes will not only bypass assembly issues but also provide valuable information about virus genome evolution by enabling whole-genome phasing of polymorphisms. Meanwhile, in an epidemiological context, long-read sequencing technologies associated with miniaturized devices, streamlined sample preparation, and live scanning of the sequencing results offers unique possibility for real-time surveillance or diagnostics. This is especially the case for the MinION sequencer based on Nanopore sequencing technology, allowing the identification of viral pathogens from a patient sample in less than 6 h, compared to more than 20 h for other sequencing technologies. The computational framework to analyze and share these types of data in a timely, safe, and meaningful way remains to be built, however it is likely that metagenomics through portable genome sequencers will become a major component of the epidemiological toolkit in the near future.

Complementarily, the throughput of sample preparation protocols and short-read sequencing approaches is likely to keep increasing at a fast pace. Concretely, these technological improvements will translate into a lower cost per sample, and an increased ability to process hundreds of samples in parallel in a timely fashion, in particular through laboratory robotics automation. For the detection of viral pathogens as well as the exploration of viral diversity and virus-host interactions in nature, this increased throughput will provide the opportunity to generate e.g., high-resolution time-series, possibly including paired cellular and viral size fractions with multiple replicates per sample, enabling more robust and sensitive data analyses.

Eventually, a fully developed virus metagenomics toolkit will enable the accurate identification of viruses in natural, clinical, and biotechnological samples for monitoring and diagnostics purposes. Moreover, as bioinformatics analysis tools advance, the reconstruction of full viral genome sequences will allow predictions to be made for the most important viruses in different environments, leading to the reconstruction of environmental virus-host networks and, when combined with other 'omics' approaches, the comprehensive evaluation of viral activity across an entire ecosystem. Collectively, these studies should lead to a

deeper understanding of viral impacts on ecological, evolutionary, and metabolic processes as well as information on potentially new viral pathogens and putative molecular virus-host interactions which could then be further characterized through targeted laboratory experiments. Hence viral metagenomics will remain a central and fundamental way to interrogate the viral world in many research fields.

Further Reading

- Bibby, K., 2013. Metagenomic identification of viral pathogens. *Trends in Biotechnology* 31, 275–279.
- Breitbart, M., Bonnain, C., Malki, K., Sawaya, N.A., 2018. Phage puppet masters of the marine microbial realm. *Nature Reviews Microbiology* 3, 754–766.
- Conceição-Neto, N., *et al.*, 2015. Modular approach to customise sample preparation procedures for viral metagenomics: A reproducible protocol for virome analysis. *Scientific Reports* 12 (5), 16532.
- Dutilh, B.E., Reyes, A., Hall, R.J., Whiteson, K.L., 2017. Virus discovery by metagenomics: The (im)possibilities. *Frontiers in Microbiology* 8, 5–8.
- Gardy, J., Loman, N.J., Rambaut, A., 2015. Real-time digital pathogen surveillance – The time is now. *Genome Biology* 16, 15–17.
- Greninger, A.L., 2018. A decade of RNA virus metagenomics is (not) enough. *Virus Research* 244, 218–229.
- Hall, R.J., Draper, J.L., Nielsen, F.G.G., Dutilh, B.E., 2015. Beyond research: A primer for considerations on using viral metagenomics in the field and clinic. *Frontiers in Microbiology* 6, 224.
- Mokili, J.L., Rohwer, F., Dutilh, B.E., 2012. Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology* 2, 63–77.
- Racaniello, V., 2016. Moving beyond metagenomics to find the next pandemic virus. *Proceedings of the National Academy of Sciences of the United States of America* 113, 2812–2814.
- Rosario, K., Duffy, S., Breitbart, M., 2012. A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Archives of Virology* 157, 1851–1871.
- Roux, S., *et al.*, 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology* 37, 29–37.
- Roux, S., Brum, J.R., 2019. A viral reckoning: Viruses emerge as essential manipulators of global ecosystems. *Environmental Microbiology Reports* 11, 1–6.
- Shkoporov, A.N., Hill, C., 2019. Bacteriophages of the human gut: The 'known unknown' of the microbiome. *Cell Host Microbe* 25, 195–209.
- Sullivan, M.B., 2014. The phage metagenomic revolution. In: Rohwer, F., Youle, M., Maughan, H., Hisakawa, N. (Eds.), *Life in Our Phage World*. San Diego, CA: Wholon, pp. 2–55. (p. 2–55–70).
- Williamson, K.E., Fuhrmann, J.J., Wommack, K.E., Radosevich, M., 2017. Viruses in soil ecosystems: An unknown quantity within an unexplored territory. *Annual Review of Virology* 4, 201–219.
- Zhang, Y.-Z., Shi, M., Holmes, E.C., 2018. Using metagenomics to characterize an expanding virosphere. *Cell* 172, 1168–1172.

Relevant Websites

- <https://img.jgi.doe.gov/cgi-bin/vr/main.cgi>
IMG/VR – Collection of viral genomes assembled from metagenomes.
- <http://ivirus.us>
iVirus – Analysis of viromes.
- <http://metavir-meb.univ-bpclermont.fr/>
MetaVir – Analysis of viromes.
- <https://www.protocols.io/groups/verve-net>
VERVE Net – Viral ecology collaboration network.
- <http://viromes.org>
VIROME – Analysis of viromes.

Database and Analytical Resources for Viral Research Community

Sujal Phadke and Saichetana Macherla, J. Craig Venter Institute, La Jolla, CA, United States

Richard H Scheuermann, J. Craig Venter Institute, La Jolla, CA, United States; University of California, San Diego, CA, United States; La Jolla Institute for Immunology, La Jolla, CA, United States; and Global Virus Network, Baltimore, MD, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of K. McLeod, C. Upon, Virus Databases, In Reference Module in Biomedical Sciences, Elsevier Inc., 2017, doi:10.1016/B978-0-12-801238-3.95728-3.

Significance of Viral Databases

Viral disease outbreaks are an ongoing threat to public health. Every few years, viral pathogens, including various influenza strains, SARS and MERS coronaviruses, Ebola, and most recently Zika virus, have caused considerable personal and economic loss (see “Relevant Websites section”). Identification of causative agents, clinical reporting of infections, and epidemiological surveillance are all critical during these outbreaks. Such efforts to identify the causes of these infectious outbreaks lead to a wealth of information about the viruses and their host. Databases allow storage and analysis of this information to fuel further wet-lab experimentation on the causative agents and comparative analyses of their genomes. Such research efforts are important for predicting, preventing and limiting future outbreaks.

Comprehensive databases such as Viral Pathogen Resource (ViPR, see “Relevant Websites section”) and Influenza Research Database (IRD, see “Relevant Websites section”) have been instrumental in providing a one-stop-shop for data and analytical tools for basic and applied research in virology. The significance of such databases is evident in multiple use cases that demonstrate the utility of the resources. For example, data and bioinformatics tools from ViPR and IRD have facilitated research into detection and diagnostics of viral pathogens, prediction of viral hosts and environmental reservoirs, viral evolution, development of vaccines, discovery of genomic determinants of virulence, and anti-viral drug development. Importantly, these resources allow investigators at all levels of training and expertise to easily perform their desired analyses and to contribute critical information about infectious disease outbreaks.

Overview of Viral Databases and Analytical Tools

Viruses infect all kingdoms of life. In this article, we focus on database resources for viruses that infect humans and other animals. We call attention to databases such as Plant Viruses Online (see “Relevant Websites section”) and the Prokaryotic Virus Ortholog Groups (pVOGs) (see “Relevant Websites section”) for readers interested in viruses that infect other host organisms, which are out of scope of this article. The landscape of databases and analytical tools available for human virology research is guided by research and development goals for priority pathogens. The available resources can be categorized as databases that store specific data types and bioinformatics webtools that offer specific analytical capabilities. These two essential functions have also been combined and integrated in comprehensive resources such as ViPR and IRD.

Types of Databases

Several types of databases are available for virology research that can be distinguished based on the type of data they contain or the pathogen area of focus (Table 1). For instance, many popular databases focus on storing information about specific biomolecules, such as gene and protein sequences, immune epitopes, or protein structures. These databases can be further distinguished as sequence archives such as GenBank (see “Relevant Websites section”), and UniProt (see “Relevant Websites section”), where data is deposited by the primary investigators and curated DBs such as RefSeq (see “Relevant Websites section”) that integrate additional knowledge (e.g., annotations) with sequence records to provide an enhanced knowledgebase. Biomolecule information other than sequences is also stored in other databases, including the Protein Data Bank (PDB; see “Relevant Websites section”), which stores 3D structural data, the Immune Epitope Database (IEDB; see “Relevant Websites section”), which catalogs experimental data on B cell and T cell epitopes studied in humans and other animals and the Virus Particle Explorer (VIPERdb; see “Relevant Websites section”), which stores the structures of viruses with icosahedral virions.

Virology databases have also been designed to focus on particular taxa of viral pathogens. For example, recognizing hepatitis B virus as a major public health problem worldwide, the Hepatitis B Virus Database (HBVDb; see “Relevant Websites section”) has been designed to facilitate research on the genetic variability of HBV and its resistance to treatment. HBVDb allows the analysis of annotated sequences for genotyping and drug resistance profiling. Similarly, a collection of databases for research on the Human Immunodeficiency Virus (HIV) are available (see “Relevant Websites section”) that contain comprehensive data on genome and protein sequences and immunological epitopes. Because influenza virus poses perhaps the most persistent major global public health threat, several databases are dedicated to research on influenza. For instance, the Global Initiative on Sharing All Influenza Data (GISAID; see “Relevant Websites section”) is an access-controlled resource of influenza sequence information and related epidemiological data. FluNet (See Relevant Websites section) is a global web-based influenza surveillance data collection, maintained at the World Health Organization (WHO) and available for tracking the movement of flu viruses globally. The Influenza Virus Resource (see “Relevant Websites section”) supports the search and analysis of

Table 1 List of databases and webtools

Category	Name	Types of data/services	Weblink
Databases	GenBank	Gene and genome sequences	https://www.ncbi.nlm.nih.gov/genbank/
	UniProt	Protein sequences	https://www.uniprot.org
	RefSeq	Curated genome and protein sequences	https://www.ncbi.nlm.nih.gov/refseq/
	Protein Data Bank (PDB)	3D protein structures	www.pdb.org
	Immune Epitope Database (IEDB)	Experimental data on B cell and T cell epitopes	www.iedb.org
	Virus Particle Explorer (VIPERdb)	Structures of viruses with icosahedral virions	http://viperdbscripps.edu/
	Viral Pathogen Resource (ViPR)	Comprehensive collection of multiple data types on high priority human pathogenic and related viruses and an integrated suite of analytical and visualization capabilities	https://www.viprbrc.org/
	Influenza Research Database (IRD)	Comprehensive collection of influenza virus-related data and an integrated suite of analytical and visualization capabilities	https://www.fludb.org
	Global Initiative on Sharing All Influenza Data (GISAID)	Access-controlled resource of influenza sequence information and related epidemiological data	https://www.gisaid.org
	FluNet	Global web-based influenza surveillance data collection	https://www.who.int/influenza/gisrs_laboratory/flunet/en/
	Influenza Virus Resource	Influenza genomic and protein sequences	https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database
	Hepatitis B Virus Database (HBVDb)	Nucleotide and protein sequence information for analysis of drug resistance profiling	https://hbvdb.ibcp.fr/HBVdb/
	Human Immunodeficiency Virus (HIV)	Genome and protein sequences and immunological epitopes	https://www.hiv.lanl.gov/content/index
	ViralZone	highly curated and extensive virus knowledgebase	https://viralzone.expasy.org
	Prokaryotic Virus Ortholog Groups (pVOGs)	Bacteriophage protein orthology information	http://dmk-brain.ecn.uioowa.edu/VOG/
Webtools	Plant Viruses Online	Plant virus gene and protein sequences and structures	http://sdb.im.ac.cn/vide/refs.htm
	IDSeq	Real-time pathogen detection from metagenomes	https://idseq.net
	Virome	Viral detection from environmental metagenomes	http://virome.dbi.udel.edu
	VirusDetect	viral detection from small RNA datasets using both <i>de novo</i> and reference-based assemblies	http://virusdetect.feilab.net/cgi-bin/virusdetect/index.cgi
	Viral Genome ORF Reader (VIGOR)	Homology-driven viral gene prediction and genome annotation	https://github.com/JCVenterInstitute/VIGOR4
	STRING-Viruses	Assessment of viral-host protein-protein interactions using visualization tools such as Cytoscape	http://viruses.string-db.org
	NextStrain	Rapid, real-time tracking and prediction of spatio-temporal spread of infection during infectious outbreaks	https://nextstrain.org

influenza genomic and protein sequences at National Center for Biotechnology Information (NCBI). The Influenza Research Database (IRD; see “Relevant Websites section”) provides the most comprehensive collection of influenza virus-related data and an integrated suite of analytical and visualization capabilities for research on influenza virus.

In contrast to the aforementioned resources that are focused on a particular data type or virus, the ViPR resource (see “Relevant Websites section”) is unique in that it provides cross-referenced data of multiple types on all high priority human pathogenic viruses that pose a threat to public health, except HIV. Each virus family has a dedicated portal within ViPR that offers intuitive, customized search interfaces and analytical options tailored for each of the virus families. ViralZone (see “Relevant Websites section”) provides access to a highly curated and extensive knowledgebase about a wide range of viruses.

Types of Bioinformatics Webtools

Research in virology is heavily dependent on data mining using sophisticated bioinformatics tools. With the foresight into the importance of such capabilities, several dedicated webtools are available for the users to conduct various types of analyses on viral genomes. For instance, tools such as IDSeq (see “Relevant Websites section”), Virome (see “Relevant Websites section”) and VirusDetect (see “Relevant Websites section”) allow detection of viruses from deep-sequencing of metagenomics samples. IDSeq is designed with the aim of real-time pathogen detection from metagenomes. Virome focuses on environmental metagenomes, whereas VirusDetect specifically uses small RNA datasets to detect viruses using both *de novo* and reference-based assemblies.

Once a novel virus isolate or variant is detected, tools such as the Viral Genome ORF Reader (VIGOR; see “Relevant Websites section”) enable genome annotation. VIGOR is a homology-driven viral gene prediction program that yields predicted proteins and mature peptides for newly sequenced isolates and variants of human virus. The software uses a set of highly curated databases enabling VIGOR to annotate a given viral genome. Currently VIGOR supports gene prediction and annotation of about 25 different virus taxonomic groups.

Tools are also available to study the viral pathogen in the context of its host environment. For instance, STRING-Viruses (see “Relevant Websites section”) is a webtool available as part of the STRING database that allows assessment of protein-protein interactions using visualization tools such as Cytoscape. This webtool is particularly important for studying how viral proteins interact with host proteins during various stages of infection. Likewise, NextStrain (see “Relevant Websites section”) is a webtool that enables rapid, real-time tracking of evolving pathogen populations during infectious outbreaks. NextStrain is an open source system that tracks mutation marker data on pathogen phylogenies to make inferences about epidemiologically-relevant parameters such as spatio-temporal spread of the infection within a host population.

Virus Pathogen Database and Analysis Resource (ViPR) and Influenza Research Database (IRD)

For the remainder of this article, we focus on describing two related database and analytical resources available for research on human viral pathogens – ViPR and IRD – as examples for how these types of resources are developed and used. For a more comprehensive list of other available virus database and analysis resources, we encourage the reader to explore additional information about resources listed at ViralZone (see “Relevant Websites section”).

The National Institute of Allergy and Infectious Diseases (NIAID) at the U.S. National Institutes of Health (NIH) implemented the Bioinformatics Resource Centers (BRCs) for Infectious Diseases program to support research on priority pathogens of humans. As a result, the BRC focused on viral pathogens has developed the ViPR and IRD resources as publicly-accessible online repositories for viruses that adversely affect public health with the aim of integrating research and surveillance data. ViPR (see “Relevant Websites section”) is unique amongst viral-centered databases in offering a wealth of information on a large number of viral families. In contrast, IRD (see “Relevant Websites section”) is a parallel resource that is focused exclusively on Influenza virus. The objective of both resources is to provide virus data and analytical capabilities to advance the understanding of virus transmission, pathogenesis, and host range, and to support the development of diagnostics and therapeutic interventions.

Sources of Data

The ViPR and IRD databases integrate data from three sources ([Table 2](#)):

Data Aggregated From Public Data Archives

The ViPR and IRD databases capture various data types from multiple publicly-accessible data archives. ViPR and IRD integrate genomic sequence information from GenBank (see “Relevant Websites section”), protein sequences from UniProt (see “Relevant Websites section”), protein structures from the Protein Data Bank (PDB; see “Relevant Websites section”), experimentally determined T-cell and B-cell epitopes from the Immune Epitope Database (IEDB; see “Relevant Websites section”), and Gene Ontology annotations from the GO database (GO, see “Relevant Websites section”). All data types are regularly updated and are searchable using their original accession numbers within intuitive web-based user interfaces.

Direct Submission of Novel Data

In some cases, active research projects supported by the U.S. National Institutes of Health and other interested parties submit data and related metadata directly to ViPR and IRD. For instance, NIAID-funded Systems Biology Consortium for Infectious Diseases research programs submit a variety of different transcriptomic, proteomic, and metabolomic datasets that investigate in vivo and in vitro host responses to viral infections. The Genomic Sequencing Centers for Infectious Diseases (GCID) program submit detailed structured metadata, including clinical information such as disease symptoms, severity, and diagnostic test outcomes, that are linked with sequence records of the corresponding virus isolate obtained from GenBank. IRD serves as the repository for the influenza human and animal surveillance data collected by the Centers of Excellence for Influenza Research and Surveillance (CEIRS) program.

Derived and Predictive Data

The IRD and ViPR development team generates and integrates unique derived data from bioinformatics analysis pipelines performed in-house, tailored specifically for a given taxonomic groups. Derived data include improved and consistent metadata annotations including strain name, clade and genotype information, virus taxonomy, host and country of isolation, and collection date. For instance, the ViPR annotation process extends information available in the representative RefSeq strain for each species. The process uses multiple sequence alignment to map homologous regions across related viral genomes to map mature peptide cleavage sites on

Table 2 Sources of IRD and ViPR data

	<i>Data source/algorithm</i>	<i>Data type</i>	<i>Component</i>
<i>Imported public data</i>	NCBI - GenBank	Genome sequences/annotations	ViPR and IRD
	NCBI - RefSeq	Genome sequences/annotations	ViPR and IRD
	Immune Epitope Database (IEDB)	Curated epitopes	ViPR and IRD
	UniProt	Protein annotations	ViPR and IRD
	RCSB Protein Data Bank (PDB)	Protein 3D structures	ViPR and IRD
	Catalytic site atlas	Active sites	ViPR and IRD
	PATRIC & VBRC Bioinformatics Resource Centers	Orthologs	ViPR
	AVIBase	Bird taxonomy	IRD
<i>Data submitted directly</i>	NIAID Genome Sequencing Centers	Clinical metadata	ViPR and IRD
	NIAID Systems Biology program	Host factor data	ViPR and IRD
	ViPR-funded driving biological projects	Host factor data	ViPR
	NIAID Centers of Excellence for Influenza Research and Surveillance (CEIRS)	Surveillance records,	IRD
	NIAID Centers of Excellence for Influenza Research and Surveillance (CEIRS)	Serology test records	IRD
<i>ViPR/IRD generated data</i>	NCBI BlastP	Sequence similarities	ViPR and IRD
	InterProScan	Domains/motifs	ViPR and IRD
	NetCTL	Predicted CTL epitopes	ViPR and IRD
	ViPR pipeline	Mature peptides	ViPR
	ViPR custom algorithm	Isoelectric point and molecular weight	ViPR
	ViPR curation	Sequence feature variant types	ViPR
	IRD pipeline	SNP/consensus sequence	IRD
	IRD custom algorithm	Isoelectric point and molecular weight	IRD
	IRD curation	Sequence feature variant types	IRD
	IRD curation	PCR primers & probes	IRD
	IRD algorithm	PA-X protein annotation	IRD
	IRD tool	H5N1 clade classification	IRD
	IRD curation	Flu season assignment	IRD
	IRD tool	2009 pH1N1 sequence classification	IRD

polyproteins. Likewise, a custom annotation pipeline is used in IRD to predict open reading frames and sequences for variants of influenza proteins including PA-X, PA-N155, PA-N182, M42, NS3 and PB1-40. The predicted variant proteins can be retrieved from the Nucleotide and Protein Sequence Search pages. Various tree-based clade classification tools are also available and have been used to predict clades and genotypes of pathogenic strains of several viruses including Zika, rotaA, and Hepatitis C virus in ViPR and H1N1, H5N1 and swine H1 strains in IRD. Furthermore, Sequence Features (SFs) are derived using information integrated from UniProt, GenBank, IEDB and the scientific literature followed by inspection and validation by domain experts. SFs are protein regions with important structural, functional, immune epitopes, or sequence alteration characteristics. Once the SF protein regions are defined, the extent of sequence variation observed in each region is determined as a series of Variant Types (VTs). Lastly, the Host factor component of IRD/ViPR contains a variety of derived data that gives insights about the systems-level infection dynamics. For instance, host factor biosets are group of genes/proteins/metabolites that are significantly differentially expressed/abundant at different times post infection. Data models derived using Weighted Gene Coexpression Network Analysis (WGCNA) are available to aid identification of co-expressed genes that may be functionally related, tightly co-regulated or members of similar pathway. The set of co-expressed genes can also be visualized as Cytoscape networks where nodes represent genes and edges represents the strength of co-expression.

Data Summary

Table 3 ViPR and IRD offer frequent updates on all data types. Genome sequence data are updated daily (IRD) or weekly (ViPR) while all other data types are updated with each bimonthly release. As of September 23, 2019, ViPR provides data on 667,249 virus strains from nearly 6126 viral species belonging to 20 families including *Arenaviridae*, *Caliciviridae*, *Coronaviridae*, *Fimoviridae*, *Filoviridae*, *Flaviviridae*, *Hantaviridae*, *Hepeviridae*, *Herpesviridae*, *Nairoviridae*, *Paramyxoviridae*, *Peribunyaviridae*, *Phasmaviridae*, *Phe-nuoviridae*, *Picornaviridae*, *Poxviridae*, *Reoviridae*, *Rhabdoviridae* and *Togaviridae*. It contains sequences from nearly 883,170 genomes, out of which upwards of 110,742 are complete genome sequences. Sequence data on > 2,100,000 proteins are also available and contains various attributes including annotations, mature peptide data, experimentally determined epitopes, etc. ViPR contains a total of 16,945 3D protein structures from PDB and 61,816 experimentally-determine immune epitopes. **Table 3** displays a breakdown of available data; details may be found at the link (see “Relevant Websites section”).

Table 3 Data summary. Numbers of various data types available in ViPR and IRD as of September 23, 2019 are shown. All data types are regularly updated

Data category	Attribute	ViPR	IRD
Genome information	Species	6,126	
	Genomes/segments	883,170	751,002
	Complete genome segments	110,742	419,482
	Proteins	2,143,646	1,184,929
	Mature peptides	243,538	
	Strains	667,249	161,216
	Strains with predicted genotypes	140,817	
	Strains with predicted segments	28,570	
	Genomes with clinical metadata (NIAID GSCID, manual curation)	3,931	
	Orthology group	9,385	
	Functional annotation	1,242,728	
	PubMed references	350,036	180,692
	Sequence feature	1,659	5,629
Number of proteins with specified annotations	Proteins	2,143,646	1,184,929
	Epitopes from IEDB	61,816	3,885
	PDB Files	16,945	1,748
	Pfam domains	1,699,884	1,104,118
	Other domains/motifs	1,295,044	760,672
	GO IDs biological process	155,156	599,355
	GO IDs molecular function	177,460	575,346
	GO IDs localization	262,063	663,293
	GO IDs		792,355
	EC numbers	33,636	
	Proteins with predicted epitopes*	1,769,976	1,089,613
	BlastP Alignments*	1,914,679	
Number of strains with specified annotation	Strains with predicted pH1N1 classification		69,668
	Strains with predicted H5 clade classification		9,042
Number of segments with specified annotation	Total segment sequences		751,002
	Polymorphism data*		437,685
	PubMed references		180,692
Number of Surveillance Samples	Hosts sampled		1,055,511
	Samples		1,233,703
	Samples tested for flu		1,211,204
	Flu-positive samples		80,040
	Samples with sequence data		9,686
	Samples with serology data		339
	Samples with structured metadata		1,233,703

The IRD holds 751,002 total influenza genome segment sequences, 1748 PDB structures and 1,184,929 proteins with predicted epitopes. Also, the IRD is unique in providing host factor datasets generated from experimental infections of host organism and cell lines with various viral strains. These cover a range of pathogens in the *Orthomyxoviridae* and *Coronaviridae* families. Currently, 66 datasets from four types of “omics” experiments (transcriptomics, proteomics, lipidomics, and metabolomics) are provided. Out of 66 datasets 34 have been analyzed for the WGCNA data models and 25 have the Cytoscape network visualization implemented. IRD is also unique in offering human and animal surveillance data collected by Centers of Excellence for Influenza Research and Surveillance (CEIRS). Additionally, in collaboration with the Global Animal Disease Information System the of Food and Agricultural Organization of United Nations, IRD has established links between sequence records in IRD and disease outbreak event records in EMPRES-I (see “Relevant Websites section”).

KEY DATA FEATURES of ViPR and IRD

- Comprehensive and up-to-date
- Consistent annotations
- Well-curated Influenza variant proteins
- Mature peptides for *Flaviviridae*, *Coronaviridae*, and *Picornaviridae*
- Well-curated metadata about geographic locations, host species, date of isolation, etc.
- Unique data types including host factor, sequence features and animal and human surveillance information

Data Curation

ViPR and IRD offer highly-curated data that has been vetted using computational and manual curation strategies. For instance, an in-house curation and annotation pipeline provides curated sequences from which sequence anomalies have been detected for potential removal during downstream analysis. Along with the sequence data, the ViPR team has manually-curated the scientific literature to provide improved and consistent annotations of metadata including the geographic location, year, and host for many clinically-relevant taxonomic groups. The highly curated strain level data are displayed with a Genome Map image and a Protein Information table from which detailed structural and functional information for a given gene/protein can be obtained. ViPR utilizes RefSeq strains to extend the manually-curated annotations to strains belonging to the same taxon. Furthermore, RefSeq sequences are used to construct virus ortholog groups and their associated annotations, which enable identification of proteins with similar function within a given virus taxon. ViPR and IRD also offer curated data on T-cell and B-cell immune epitopes and their predicted positioning on protein structures from the IEDB. Data curation in ViPR and IRD continues to grow and expand beyond sequence and strain level information. For example, both databases offer curated antiviral drug data from the DrugBank (see “Relevant Websites section”), including the descriptive drug information, 3D structures for target complexes, interaction sites as sequence features and antiviral resistance mutations to aid in assessing the risk of anti-viral drug resistance development.

Data Retrieval

Search Interface

ViPR offers customized search interfaces to allow for the retrieval of selected genomic, structural and other data records using specific metadata for different virus families. Users initiate the search by selecting a virus family on the home page. A user can narrow the search data specific to a virus strain by querying the database using genus or species, geographical location and date of isolation, virus host, and clinical or experimental data. The user also has an option to cast a wider net using keywords with further narrowing using advanced search options. Once a strain or set of strains is selected, detailed genomic and protein sequence information and associated annotation can be accessed. These data can then be directly analyzed using any of the appropriate tools available from within ViPR.

Because IRD is dedicated to influenza viruses, the search interface design is guided by the availability of influenza strain-specific data. Users can query the database using the branching logic inherent in the database. For instance, users can search for complete or partial genome and segment sequences, and proteins by directly entering the name of the strain(s) of interest. Users can also choose amongst the several metadata fields such as host, geographic location and the date of pathogen isolation. Once a particular taxon, strain or metadata category is selected as a search criterion, additional search criteria appear dynamically to allow the users to perform more focused searches. Moreover, users can also use the advanced search options to refine the search results with the more fine-grained search criteria. For instance, users can choose to view data on strains isolated in specific months of a given year(s) or limit search to specific host attributes such as gender and age and choose to limit their search to specific specimen type, laboratory strains and organism detection method. Lastly, users can customize their viewing options to specific display fields through advanced search menu. An example of the various search options is shown in [Fig. 1](#).

Application Programming Interfaces (API)

ViPR and IRD provide users an option to retrieve certain data types using command line utilities via Application Programming Interfaces (APIs). Specifically, the sequence search API allows users to retrieve sequences and associated metadata using GenBank and protein accession IDs. The retrieved sequences can be obtained in either FASTA or JSON formats with user-defined associated metadata. The surveillance API allows retrieval of surveillance records and metadata from host surveillance samples. IRD allows users to submit sequences for large phylogenetic analysis jobs through an API to the high-performance computing environment provided by the NSF-sponsored Cyber-Infrastructure for Phylogenetic Research (CIPRES) Gateway. Tree calculations are made using the high-performance computing environment and the resulting phylogenetic tree is returned for visualization using the Archaeopteryx tool in IRD.

Analysis and Visualization Capabilities

ViPR and IRD host a comprehensive suite of bioinformatics tools for data analysis and visualization, closely integrated with the supported data. These include popular webtools in bioinformatics constructed by the ViPR team or contributed by users/collaborators. Examples of the types of analysis that can be performed and the webtools that are available are described below. For a complete list of analytical tools, the reader is directed to the ViPR (see “Relevant Websites section”) and IRD (see “Relevant Websites section”) homepages.

Sequence Annotation

The sequence annotation pipelines allow users to upload and annotate genomic sequences to predict segment type, CDS location, and genotype information, and to identify possible sequencing artifacts.

Protein Sequence Search ¹

Search for influenza sequences, proteins, and strains using two types of searches. Use the advanced search to allow you to refine your search with the more fine grained search, and you can pick your viewing options.

Results matching your criteria: 0

DATA TYPE

- ☐ Genome Segments
- ☒ Protein
- ☐ Strain

VIRUS TYPE

- ☒ A
- ☐ B
- ☐ C
- ☐ Provisional Influenza D (PMD:24595369)

SUBTYPE

* Use commas to separate multiple entries.
Ex: H1N1, H7, H3N2.

STRAIN NAME

* Use commas to separate multiple entries.
Ex: A/Chicken/Ireland/1025/2008, A/Chicken/Lao/16/2006.

DATE RANGE

From: 2018 To: 2019
To add month to search, see Advanced Options: Month Range

ADVANCED OPTIONS Show All

Select Advanced Option

- ☒ Select An Advanced Option
- Keyword search
- Flu Season
- Month Range
- Submission Date

'CLASSICAL' PROTEINS

All

- 1 PB2
- 2 PB1
- 3 PA
- 4 HA
- 5 NP
- 6 NA
- 7 M1
- 7 M2
- 8 NS1
- 8 NS2

Complete?

- ☐ All
- ☐ PB2
- ☐ PB1
- ☐ PA
- ☐ HA
- ☐ NP
- ☐ NA
- ☐ M1
- ☐ M2
- ☐ NS1
- ☐ NS2

'VARIANT' PROTEINS (SOP)

- 2 PB1-F2
- 2 PB1-N40
- 3 PA-N155
- 3 PA-N182
- 3 PA-X
- 7 M42
- 8 NS3

CLADE CLASSIFICATION

- ☒ None
- ☐ Global H1 Clade (SOP) Open Source code here
- ☐ US H1 Clade (SOP) Open Source code here
- ☐ H5 Clade (SOP) Open Source code here
- ☐ 2009 pH1N1 Sequence Similarity (SOP) Open Source code here

HOST

Avian x

AVIAN

African Starling x

African Stonechat x

GEOGRAPHIC GROUPING

North America x

COUNTRY

Barbados x

Costa Rica x

Tip: To select multiple or deselect, Ctrl-click (Windows) or Cmd-click (MacOS)

Clear Search

Fig. 1 Protein search interface in IRD. The search page supports queries based on “classical” as well as “variant” proteins and associated metadata. A search query can be made more specific by choosing various query features. For example, users can search for specific strain(s) by entering the strain name and subtypes in the appropriate search fields. Additionally, choosing a type of host, such as avian, brings a drop down menu from which the user can choose one or more species to make the search criterion more specific. Users may also choose to limit their search results by geographic region(s) by choosing one or more countries in the dropdown menu. Search results may be limited to a specific date range by putting in the years or by choosing a month range through advanced options. Multiple other search criteria such as keyword search, submission date, host gender and age etc. are available through advanced options to make the search results more specific.

Sequence Search and Alignments

Users can use popular tools such as BLAST and MUSCLE within ViPR and IRD. Sequences can be selected from a search result or a working set in their personal workbenches. Users can also perform manual exploration and curation of sequence alignments including relabeling the sequences and adding sequence features. After an alignment is completed, users have an option to download the input sequences and output files in a variety of formats or pass the alignment to another tool including SNP analysis or meta-CATS.

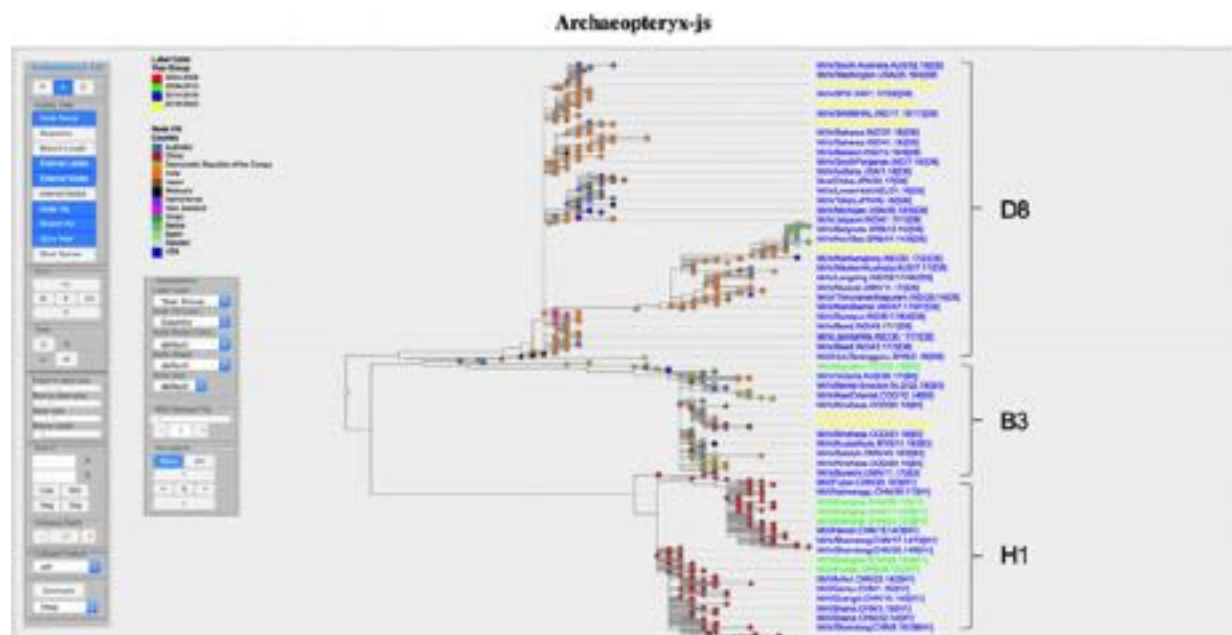


Fig. 2 Example of a phylogenetic tree constructed in ViPR. The search interface was used to retrieve unique sequences belonging to the 450 bp region that codes for the C-terminal 150 amino acids of the N nucleocapsid protein of the human measles *Morbillivirus*. A total of 554 unique sequences were obtained. Multiple sequence alignments were performed with MUSCLE and phylogenetic relationships inferred using RAxML for visualization with Archaeopteryx.js. The legend shows options for users to customize the tree visuals and highlight desired metadata in Archaeopteryx. For instance, the phylogram display with aligned labels has been chosen from the top left panel for improved readability. Likewise, the “Dyna Hide” option on the left panel has been selected to only show representative sequence names. Names of the nodes have been color coded to indicate the year of isolation. The node color indicates the country of origin. The available sequences separate into 3 clades belonging to subtypes D8, B3 and H1. The subtypes are represented in the parentheses following the names of the sequences. While the subtypes D8 and B3 are more globally circulating, causing infections across different countries, the subtype H1 is predominantly found in China. Moreover, in a given country, strains belonging to the same subtype have repeatedly caused measles infections over several years indicating pathogen persistence in the population, likely due to imperfect and inadequate vaccination practices.

Phylogenetic Tree Reconstruction

Users can infer phylogenetic relationships using RAxML and FastME algorithms, and visualize the results using a customized visualization tool developed in house called Archaeopteryx, which allows users to color-code and annotate various tree branches and nodes using available metadata, such as geographical location, host, isolation year, and amino acid residues at user-selected positions (e.g., [Fig. 2](#)).

Metadata-Driven Comparative Genomics

The meta-CATS tool provides a statistical analysis of sequences to identify genome and protein positions that show significantly-different residue distributions between groups of sequences using the Chi-squared statistic. Sequences can be segregated into groups manually or automatically based on selected metadata values, such as year of isolation, geographical location, host species, etc. Thus, a user can put the analysis of sequences in the context of infection and infer association of variations in a genomic region with a particular infection characteristic.

Analysis and Visualization of 3D Protein Structures

Users can search for protein structures using multiple types of queries, including PDB IDs, gene symbol, Entrez ID, UniProt accession, and gene product names. Furthermore, search can be restricted to include only proteins with experimentally determined epitopes, experimentally determined active sites and proteins with sequence features. Additionally, users can use advanced options to query the database using theoretical structures. Once a particular structure from the search results is selected, users can customize the general appearance of protein structures. For example, users can highlight ligands, active sites, epitopes and sequence features on the 3D structures. Individual residues within the protein structures are mapped to homologous positions from UniProt records, which allows comparison between protein structures. Annotating a 3D structure with important residues and regions of interest can yield testable hypotheses about the functional relevance of the protein. Lastly, users can download the highlighted protein structure as a publication quality image file or a structure movie.

Genome Annotation Using VIGOR

Users can use the VIGOR software tool along with its collection of highly-curated reference databases for different viruses to predict viral protein open reading frames and sequences, and to identify typical viral transcriptional and translational exceptions including RNA editing, stop codon read-throughs and ribosomal slippage.

Virus Genotype/Clade Classification

ViPR and IRD offer two types of user/community contributed tools for virus classification. A clade classification tool infers clades for a query sequence from its position within a reference phylogenetic tree. Currently, clade classification is available for Zika virus, and Hepatitis C Virus (HCV) in ViPR and swine H1 and H5N1 influenza viruses in IRD. The H5N1 classification tool uses phylogenetic analysis to classify HA sequences according to the WHO H5 classification scheme. The H5N1 classifier has been verified to have > 98% accuracy for sequences of at least 300 nucleotides of HA1. On the other hand, the H1N1 classification tool in IRD is a robust application of BLAST to recognize sequences closely related to pandemic sequences. BLAST-based classification is also available for classifying rotaA virus sequences in ViPR.

HA Subtype Numbering Conversion

IRD has implemented an HA subtype numbering conversion tool that allows users to convert HA sequence coordinates among any selected subtypes based on protein structure alignment rather than sequence-based alignment. Using this tool, the user can convert the coordinates of an HA protein sequence to the corresponding coordinates in other subtypes, to compare substitutions associated with phenotypic changes and to identify cross-reactive immune epitopes. The tool can also be integrated with sequence variation analysis and meta-CATS.

KEY ANALYSIS FEATURES

- Seamless integration of data and analysis/visualization tools
- Analysis of user data in combination with database data

TOP ANALYSIS TOOLS BASED ON USAGE

- Multiple sequence alignment
- BLAST
- HA numbering
- SNP analysis
- Phylogenetic tree reconstruction
- H5N1 classifier
- H1 classifier

Workbench

Users can establish personal workspaces under the “workbench” feature within the IRD and ViPR. This tool provides an interface that allows users to save previous search or analysis results, which enables users to re-use their work without re-running the analysis. It also allows users to combine multiple analyses. Users can upload and save their own private data and metadata to their



Fig. 3 Links to multifaceted user support available in ViPR.

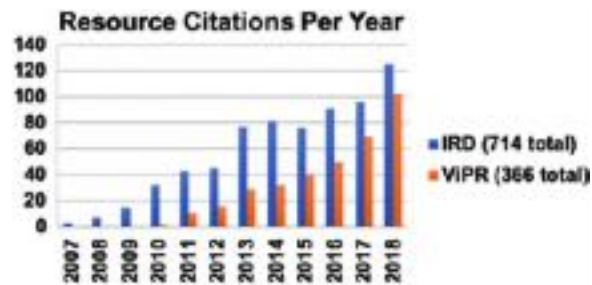


Fig. 4 The number of peer-reviewed citations for the ViPR and IRD resources.

workspace to be analyzed using the analytical and visualization tools provided by IRD and ViPR. The saved data and analysis results can be shared with collaborators through their workbench accounts.

User Support

The IRD and ViPR databases are open access resources and can be used and shared without restrictions. The databases offer multifaceted user support (Fig. 3). Users can report a problem or ask a question using the forms provided online. The development and management teams of both IRD and ViPR are responsive to questions from the helpdesk and to suggestions for enhancements. Users can join a newsletter mailing list to get information about updates of the resources.

Both IRD and ViPR provide extensive tutorials, training modules and manuals. For additional support, the development and management teams engage in outreach sessions that include webinars, tutorials, and training workshops at various geographical locations. For an expert user, the analytical tools developed by the ViPR/IRD team are also available on GitHub, which avails the user with an option of using the tools outside of the IRD and ViPR resources on their preferred platform.

Usage Statistics

The ViPR and IRD databases continue to provide critical resources in several research studies as evident by the increasing number of citations in the scientific literature (Fig. 4). Together, the two databases have been cited in 1080 publications as of May 10, 2019. The number of new sessions initiated per week in 2018 (Google Analytics) tallies at 1488 at ViPR and 1482 at IRD. Importantly, these sessions have been documented from 181 countries for ViPR and 174 countries for IRD.

Summary and Conclusions

Virology research is dependent on timely availability of reliable data on viral pathogens, their hosts and the infection/outbreak dynamics. ViPR and IRD offer comprehensive, highly curated data on human viral pathogens along with an intuitive search interface and seamless integration of the data with analytical and visualization tools. The resources are available freely without restrictions. The availability of such resources streamlines and expedites experimental discovery advancing the ultimate goal of developing improved diagnostics and therapeutics for priority pathogenic viruses.

Further Reading

- Adhikari, U.K., *et al.*, 2018. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging Oropouche Virus. *Journal of Immunology Research*. doi:10.1155/2018/6718083.
- Afelt, A., *et al.*, 2018. Bats, Bat-Borne Viruses, and Environmental Changes. *IntechOpen*.
- Andreani, J., *et al.*, 2019. Atypical cowpox virus infection in smallpox-vaccinated patient, France. *Emerging Infectious Diseases* 25 (2), 212–219.
- Babayan, S.A., *et al.*, 2018. Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science* 362 (6414), 577–580. doi:10.1126/science.aap9072.
- Brown, D.M., *et al.*, 2018. Contemporary circulating enterovirus D68 strains show differential viral entry and replication in human neuronal cells. *mBio* 9 (5), doi:10.1128/mBio.01954-18.
- Carter, K., *et al.* Anti-Chikv Antibodies and Uses Thereof, US Patent. US20180127487A1.
- Claes, *et al.*, 2014. The EMPRES-i genetic module: A novel tool linking epidemiological outbreak information and genetic characteristics of influenza viruses. *Database*. bau008.
- Dutta, S.K., *et al.*, 2018. Chikungunya virus: Genomic microevolution in Eastern India and its in-silico epitope prediction. *3 Biotech* 8, 318.
- Greene, J.M., *et al.*, 2007. National institute of allergy and infectious diseases bioinformatics resource centers: New assets for pathogen informatics. *Infection and Immunity* 75, 3212–3219.
- Langfelder, P., Hovrath, S., 2008. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559–572.
- Lee, A.J., *et al.*, 2017. Identification of diagnostic peptide regions that distinguish zika virus from related mosquito-borne flaviviruses. *PLoS One* 12 (5), e0178199. doi:10.1371/journal.pone.0178199.
- Lee, *et al.*, 2015. Diversifying selection analysis predicts antigenic evolution of 2009 pandemic H1N1 influenza A virus in humans. *Journal of Virology* 89, 5427–5440. doi:10.1128/JVI.03636-14.

- Mette, *et al.*, 2005. An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *European Journal of Immunology* 35 (8), 2295–2303. doi:10.1002/eji.200425811.
- Miller, M.A. *et al.*, 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: *Proceedings of the Gateway Computing Environments Workshop (GCE)*, pp. 1–8. New Orleans, LA.
- Mock, F., *et al.*, 2019. Viral host prediction with deep learning. *bioRxiv*. doi:10.1101/575571.
- Peng, M., *et al.*, 2018. Luteolin escape mutants of dengue virus map to prM and NS2B and reveal viral plasticity during maturation. *Antiviral Research* 154, 87–96.
- Pickett, B.E., *et al.*, 2011. ViPR: An open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research* 40 (Database issue), D593–D598. (PMID: 22006842).
- Pickett, B.E., *et al.*, 2012. Virus pathogen database and analysis resource (ViPR): A comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses*. 4 (11), 3209–3226.
- Pickett, B.E., *et al.*, 2013. Metadata-driven comparative analysis tool for sequences (meta-CATS): An automated process for identifying significant sequence variations that correlate with virus attributes. *Virology* 447 (1–2), 45–51. doi:10.1016/j.virol.2013.08.021.
- Pinsky, B.A., *et al.*, 2019. Methods and Reagents for Detection of Chikungunya Virus and Zika Virus, United States Patent Application. 20190024195.
- Zhang, Y., *et al.*, 2017. Influenza research database: An integrated bioinformatics resource for influenza research. *Nucleic Acids Research* 45 (Database issue), D466–D474. doi:10.1093/nar/gkw857.
- Zou, C., *et al.*, 2019. Virulence difference of five type I dengue viruses and the intrinsic molecular mechanism. *PLOS: Neglected Tropical Diseases* 13 (3), e0007202.

Relevant Websites

- <https://emergency.cdc.gov/recentincidents/index.asp>
Centers for Disease Control and Prevention.
- <http://www.drugbank.ca>
DrugBank.
- <http://empres-i.fao.org/empres-i>
EMPRES-i - FAO.
- https://www.who.int/influenza/gisrs_laboratory/flunet/en/
FluNet - WHO.
- <https://www.ncbi.nlm.nih.gov/genbank/>
GenBank Overview - NCBI - NIH.
- www.geneontology.org
Gene Ontology Resource.
- <https://www.gisaid.org>
GISAID - Global Initiative on Sharing All Influenza Data.
- <https://idseq.net>
IDseq.
- www.iedb.org
IEDB.org: Free epitope database and prediction resource.
- <https://www.fludb.org>
Influenza Research Database.
- www.fludb.org
Influenza Research Database.
- <https://github.com/JCVenterInstitute/VIGOR4>
JCVenterInstitute/VIGOR4: VIGOR4 - GitHub.
- <https://www.hiv.lanl.gov/content/index>
Los Alamos National Laboratory.
- <https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database>
National Center for Biotechnology Information.
- <https://nextstrain.org>
Nextstrain.org.
- <http://sdb.im.ac.cn/vide/refs.htm>
Plant Viruses Online: Descriptions and Lists from the VIDE Database.
- <http://dmk-brain.ecn.uiowa.edu/VOG/>
pVOGs - Prokaryotic Virus Orthologous Groups.
- <https://www.ncbi.nlm.nih.gov/refseq/>
RefSeq: NCBI Reference Sequence Database - NIH.
- <http://viruses.string-db.org>
STRING Viruses.
- <https://hbvdb.ibcp.fr/HBVdb/>
The Hepatitis B Virus database.
- <https://www.uniprot.org>
UniProt.Org.
- <http://viprdb.scripps.edu/>
VIPERdb - The Scripps Research Institute.
- <https://www.viprbrc.org/brc/dataSummary.spg?decorator=vipr>
ViPR.
- <https://viralzone.expasy.org>
ViralZone root - ExPASy.
- <https://viralzone.expasy.org/677>
Virology links ~ ViralZone page.
- <http://virome.dbi.udel.edu>
VIROME.

<http://virusdetect.feilab.net/cgi-bin/virusdetect/index.cgi>

VirusDetect.

<https://www.viprbrc.org/>

Virus Pathogen Database and Analysis Resource.

www.viprbrc.org

Virus Pathogen Database and Analysis Resource.

www.pdb.org

wwPDB: Worldwide Protein Data Bank.

Classification of the Viral World Based on Atomic Level Structures

Janne J Ravantti, University of Helsinki, Helsinki, Finland

Nicola GA Abrescia, Center for Cooperative Research in Biosciences, Basque Research and Technology Alliance, Derio, Spain; Ikerbasque, Basque Foundation for Science, Bilbao, Spain; and Center for Biomedical Research in the Liver and Digestive Diseases Network, Carlos III Health Institute, Madrid, Spain

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

cryo-EM electron cryo-microscopy

C α Alpha carbon atoms

DJR Double jelly-roll (or double β -barrel)

dsDNA Double-stranded DNA

ICTV International Committee on Taxonomy of Viruses

MCP Major Capsid Protein

NMR Nuclear Magnetic Resonance

PDB Protein Data Bank

RMSD Root-mean-square deviation

ssDNA Single-stranded DNA

X-ray X-ray crystallography

Å Ångström, 0.1 nm.

Glossary

Protein Data Bank A database for the three-dimensional structural data of large biological molecules (including protein, nucleic acids and viruses).

Similarity score A numerical measure to quantify how similar two or more objects of study are.

Introduction

Classification is the process of discovering what is similar and what is different in a set of objects. In biology, in particular, the classification of organisms has been a major focus since Aristotle, who created a complete order of living things (*"Scala naturae"*). The modern standardized biological classification of organisms was established in the 18th century by Linnaeus, who showed, that a meaningful classification can be established by comparing only selected parts of plants. Then, the first "tree of life" (*"Arbre botanique"*) relating members of the plant kingdom was published by Augustin Augier. Later, taking into account this concept of a tree-of-life, Darwin proposed the theory of natural selection and the theory of descent, which attempted to explain how species emerge. Finally, in 1953, the determination of the structure of the DNA double helix by Rosalind Franklin, James Watson and Francis Crick, and how it carries the genetic information from one generation to the next, enabled a unified basis for the classification at the molecular level of all biological entities on earth.

A simple viral particle consists of the genetic material, DNA or RNA, which is shielded from hostile environments by a protein structure called the capsid which, in turn, can be enveloped by a membrane. A regular sized capsid is formed by one or several types of proteins. As viruses have to pack their entire genome into the capsid, the number of genes encoding capsid forming proteins is severely constrained. Thus, multiple copies of the same protein or proteins are used to build the capsid of a required size. This intricate way of forming capsids from repeating subunits locks the capsid's protein structures from evolving in such a way that the assembly of the whole capsid is jeopardized.

Viruses are obligatory parasites that need their hosts to proliferate, but they do not often offer any direct benefits to their infected hosts. Thus, there is a constant arms race for host cells to evolve and escape viral infections whereas viruses have to adapt to the adapting hosts to generate a progeny. This tug-of-war can be seen in the receptor binding proteins that are under constant evolutionary pressure to change enough to be able to detect and bind the host cells.

Viruses have been classified like any other biological entities ever since their discovery. However, since they are "pseudo-living", their placement within the originally proposed tree of life has been problematic. The earliest proposed virus classification was based on the Linnaean hierarchical method based on virus' shared properties and genome types. In 1966 the International Committee on Taxonomy of Viruses (ICTV) was officially set up to formalize naming conventions and classification guidelines. ICTV defines virus species as: "a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche". Thus, viruses belonging to same species would share common traits like the nature of the genome, morphology and host.

There are also other possible ways to classify viruses such as the Baltimore classification that uses exclusively genome type (*e.g.*, group I: dsDNA viruses and group II: ssDNA viruses) and the replication method as a basis for the classification. Finally, as the virus capsid is considered by some to be the hallmark of the virus ("no capsid, no virus"), the structure-based virus classification has been proposed. In this view, those viruses that share the same fold in the major capsid proteins (MCPs) and the capsid architecture, belong to the same viral lineage regardless of their genome type or the hosts. All in all, the different classification schemes complement each other and highlight the diversity of the virosphere.

Viruses, despite undergoing constant environmental pressure and intrinsic genomic mutations that effectively sample the sequence space, keep the functionalities of their proteins intact, particularly those forming the capsids. Major capsid proteins provide a unique opportunity to develop and test computational methods for structural classifications, as well as to address the question of how the primary structure dictates the three-dimensional functional protein structure.

Aside from the pure academic interest, knowing how different (viral) proteins are related allows the development of new drugs or the repurposing of old ones, *e.g.* for tackling viral diseases. It is also possible to infer functions of newly discovered proteins by comparing them against known proteins. The primary structure – the amino acid sequence – can be used for comparing closely related proteins as the evolutionary trajectory is constrained by protein's function, but for more distant relatives, the tertiary structure – the three-dimensional shape of the protein – is usually far more conserved than the sequence, since the three-dimensional structure, whether being part of the virus or the host is the functional unit. Thus, the tertiary structure allows both (1) long-distance evolutionary studies, *e.g.*, to find homologous proteins where the sequence conservation is lost, and (2) an opportunity to increase our knowledge about proteins' physical properties, *e.g.*, to understand thermodynamic parameters that govern the protein folding process.

With the technological advances to structurally probe biological entities (*e.g.*, X-ray crystallography, high-throughput sequencing and electron cryo-microscopy [cryo-EM]) as well as the greatly increased computational power to process the data, we have now significantly improved our understanding of the living world. This has come with a price: it is no longer straightforward to classify organisms like bacteria into a single tree with a simple top-down hierarchy. This, in fact, does not necessarily depict the real relationships between the bacterial species, which might have acquired whole gene families and new metabolic pathways by horizontal gene transfer. Thus, the current problem of biological classification is that it is based on *hierarchies*, *i.e.*, all objects to be categorized must form a tree-like structure.

Once the central dogma of the molecular biology was established, the early idea of “one sequence – one protein – one function” was quickly discarded as more and more structural data became available and it was observed that proteins with dissimilar sequences might encode highly similar structures. As viruses have to constantly maintain the ability to infect their hosts that are in turn evolving in new ways to evade infections, virus' protein sequences and structures have been essential for understanding the complex sequence-structure relationships. One of the early hallmarks was the observation by Michael Rossmann that the structure of the coat protein of human rhinovirus 14 is surprising similar to that of the plant virus, despite the lack of detectable sequence similarity between the two. However, at the same time, it was observed that the naturally occurring (viral and cellular) protein fold-space seemed to be quite limited. Thus, the biological (macro-)molecules are formed as combinations of common three-dimensional folds.

Ultimately, if one would know how the protein's primary structure (sequence) dictates the tertiary structure, the classification could be directly sequence-based. Currently, the best methods to classify a set of proteins are based on using both sequence and structural information. For the distantly related proteins especially, the ultimately redundant information encoded in their 3D-structures enhances the weak evolutionary signals.

Methods to Compare (Viral) Proteins

The comparable 3D structures are derived using X-ray, NMR, cryo-EM or, in some cases, computational methods. Historically, X-ray has been the main source of atomic resolution structures for proteins (and icosahedral viruses). NMR has been limited to smaller molecules, whereas it is only in recent years that cryo-EM has reached the atomic level.

In sequence comparisons, the nucleotides and amino acids represented by single letters are always in register, *i.e.*, the movement of the primary structures in respect to each other is one-dimensional (left or right) and constrained to have either matching letters denoting the same nucleotide or amino acid, mismatching letters denoting point mutations or gap-letter *versus* nucleotide/amino acid letter denoting insertion or deletion events (called indels) for each position in both sequences. However, even the simplest sequence comparisons can report good matches by chance, in particular if comparing nucleotide sequences. Computationally, sequence comparison is a well understood problem since it can be directly formulated as a dynamic programming problem, where scores of each residue-residue comparison are strictly based on the scoring matrix and previous neighboring comparisons and, the best alignment score can be calculated by progressing sequentially through each individual comparison. However, as the rules of how sequence dictates the three-dimensional structure of a protein are yet to be discovered, the comparisons with the current methods tend to break down with highly diverse proteins, often found in viruses.

The tertiary structures (with the assumption that they are at the same scale) have six degrees of freedom (three axes to move structures relative to each other and three rotational axes) making the alignment computationally more demanding and the exhaustive search impractical. On the other hand, structural similarities last longer in the evolutionary timescales, thus it is possible to study remote homologs with structural comparisons even when the sequence similarities cannot be detected using current methods. **Fig. 1** shows an example of how viral coat proteins with rather different protein sequences can be aligned accurately based on their structural properties.

The earliest method of comparing two protein structures was visual comparison by eye. There, the protein structure coordinates were visualized together and then aligned to each other by moving and rotating structures. The proper alignment needed an expert to decide when the equivalent parts of the proteins were aligned. Usually, the C α atoms were considered for the alignment for simplicity and efficiency. After the alignment, one could calculate a Root Mean Square Deviation (RMSD, the square root of the averaged sum of the squared differences of the C α distances) between the relevant parts of the structures. The alignment by eye becomes cumbersome and error-prone when there are more than a few structures to compare.

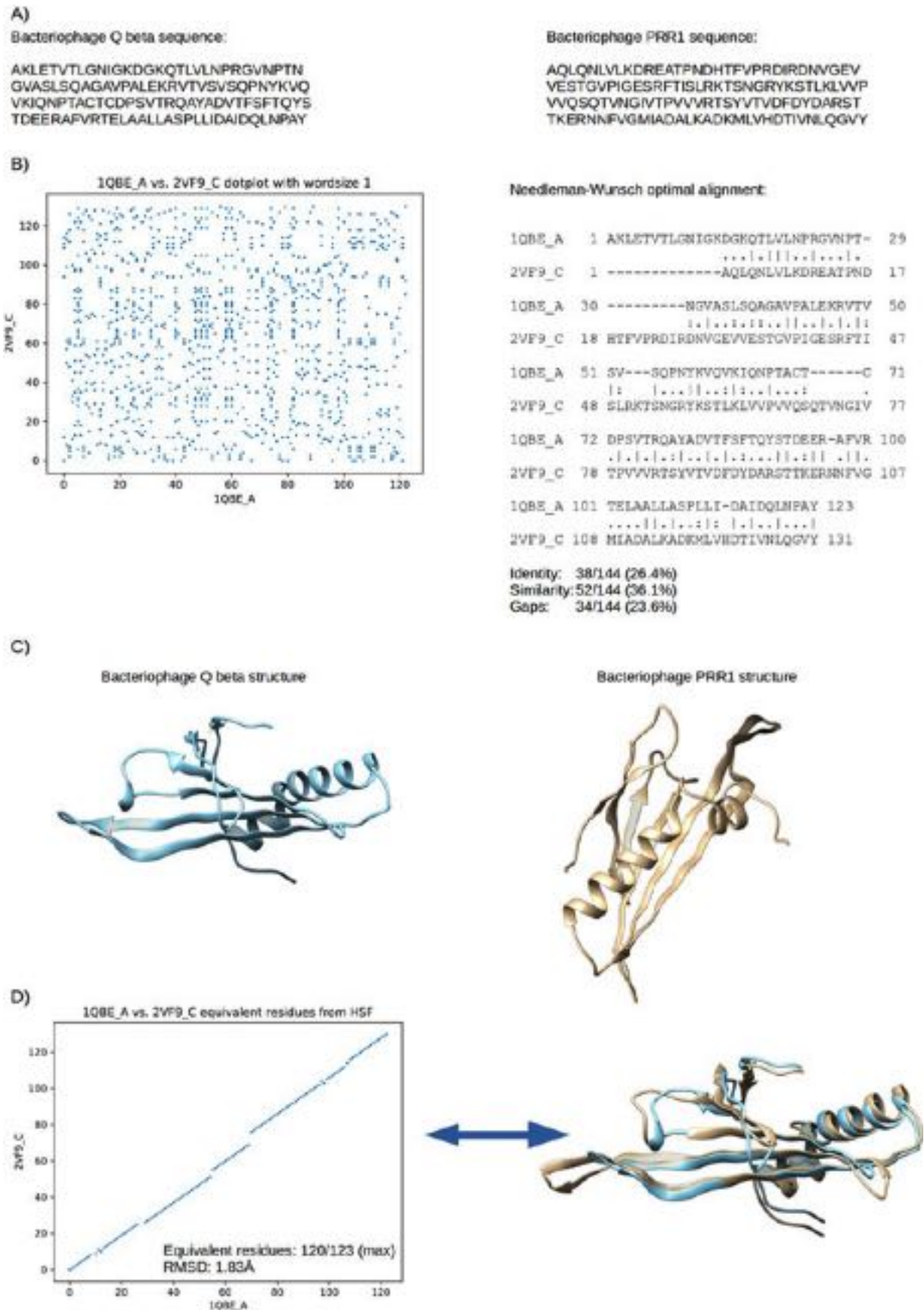


Fig. 1 Sequence vs. structure alignment. How sequence alignment differs from structural alignment. (a) protein sequences of bacteriophage Q beta (family *Leviviridae*; genus *Allolevivirus*; PDBid 1QBE_A) and bacteriophage PRR1 (family *Leviviridae*; genus *Levivirus*; PDBid: 2VF9_C). Viruses in the *Leviviridae* are small RNA-viruses known to have very high mutation rate. (b) left: dotplot of protein sequences with wordsize 1. If there were significant sequence identity, the diagonal would be filled with points. Right: Needleman-Wunsch sequence alignment. Sequence similarity is 36.1% (52/144). (c) Secondary structure visualizations of the 1QBE_A and 2VF9_C. (d) left: dotplot of the equivalent residues based on HSF-program. Right: the superimposition of the structures.

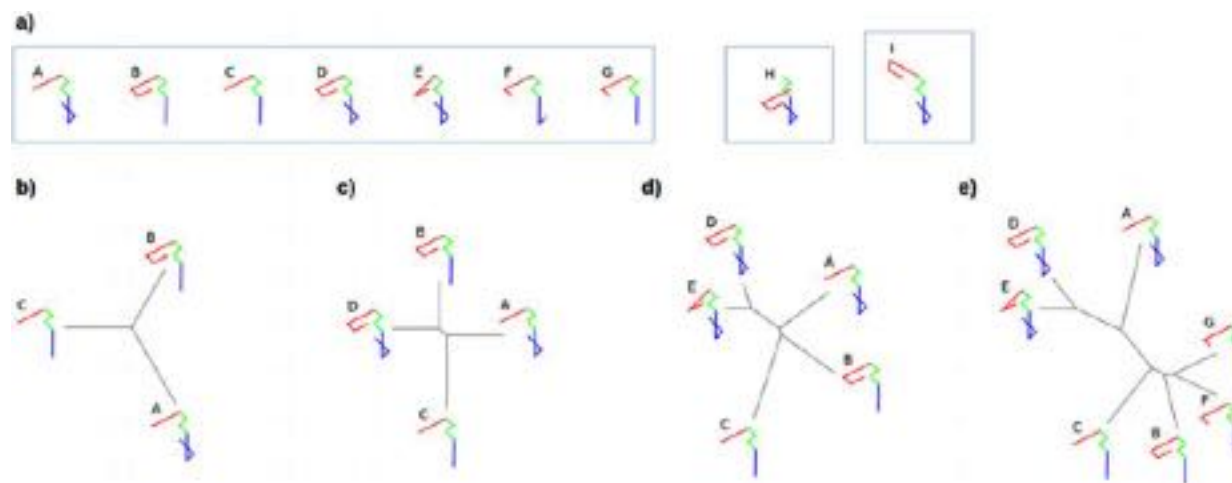


Fig. 2 Clustering of multiple structures and effects of adding structures into the data set. (a) line representation of hypothetical protein structures to be clustered. Each “structure” (e.g., structure A) has several “domains” colored separately. (b)–(e) “phylogenetic” trees based on structural similarities when more and more structures are added and cluster boundaries/tree branches change. In the beginning (panel b)) the structures A, B and C are equally apart. In (c) adding structure D causes clustering of B and D together and A and C together. In (d) adding structure E causes structure B to radically change from being closest to D to being closest to structure C. In (e) added structures F and G push structure A farther away from original structures B and C, but at the same time two sides of the tree (E, D and A vs. C, B, F and G) have enough close relatives that the distances and the local topology stabilizes. Structures H and I show difficulties when comparing to already clustered structures. Structure H has rearranged domains, thus making it difficult to place into the final tree. Structure I in turn shows how movement of parts of the protein structure makes especially rigid body alignment hard to solve.

The first automated methods for structural alignment were devised in the early 70's and the fundamental question for comparing a set of proteins became evident: what is similar and what is different and how to quantify those similarities and differences. The similarity measures between two proteins need some definition of the concept of *equivalence*. In the geometrical case, equivalent residues can be thought to be those that superimpose well on top of each other when the rigid structures are aligned. How to define which residues in two structures are equivalent algorithmically is similar to the “point pattern matching problem” in image processing whose exact solutions are computationally very hard to find [*i.e.*, the non-deterministic polynomial acceptable problems (NP-hard) in computer science]; in other words, neither fast nor universal methods are known. Thus, all practical methods to establish equivalence between residues use heuristic techniques.

There will not be scaling or perspective to worry about if we are at the atomic resolution level. The three dimensional matching even without considering scaling is still a hard problem computationally. There are again six degrees of freedom (three rotational and three translational ones) assuming that the matching points between different point clouds (e.g., C α positions of protein residues in space per compared structure) are known. If there is ambiguity in point-matching or if there are more points in one cloud compared to another, the problem becomes even harder.

It is important to understand that the geometry (*i.e.*, position of the atoms in the 3D-space) does not entirely cover the question of the similarity (see Fig. 2). Proteins are highly dynamic in the cell and their conformations depend on the circumstances e.g., possible binding of cofactors. A typical example, calmodulin, has two distinct conformations when binding calcium ions: an open one without calcium ions (PDBid: 1CFD) and a closed one with those ions bound to the protein (PDBid: 1CLL). The amino acid sequence for both conformations of the protein is of course the same, but a rigid body structural alignment taking into account only their geometries could only align one of the two globular domains per structure when comparing the open and closed form structures to each other.

Thus, the biologically relevant computational methods dedicated to protein comparison cannot rely only on the global structural superimpositions. An early step towards more realistic comparisons was to compare the intramolecular distances instead of the intermolecular ones (e.g., equivalent C α distances). Regarding the intramolecular methods, the idea is to calculate distance maps from each structure to compare and then calculate a similarity score based on the similarity of the relevant parts of those maps. The distance maps are rotationally and translationally invariant *i.e.*, independent of the orientation of the molecule, therefore, the structures do not need to be superimposed in 3D in order to find similarities. For example, in the calmodulin case, both globular domains retain their internal distances at the open and closed conformations, thus giving higher similarity scores compared to simple rigid body superimposition and RMSD. There are several different approaches to calculate and compare intramolecular distances. Some methods (e.g., DALI) use the complete distance matrices to find regions of similarity whereas other methods (e.g., Combinatorial Extension) combine and compare larger sets of residues using a dynamic programming approach.

It is also possible to “scale back” from the purely atomic level comparisons and to compare derived properties that comprise larger, functionally important, subunits of proteins like secondary structures (e.g., helices and β -sheets) up to “super secondary structures” *i.e.*, continuous folds (e.g., whole domains). The property-based classification methods can be either based on

superimposition of the subunits or they can be alignment free (*e.g.*, based on comparing frequencies of the properties). The currently best methods for protein structure comparisons are using hybrid methods, where different properties of the proteins are considered together.

There are many different methods to compare structures. Similarly, there are several distinct methods to evaluate the quality of the alignment (*i.e.*, scores). The most used ones are calculated from aligned C α -atoms: the earlier mentioned RMSD and the Z-score, developed for the DALI-method that uses distance matrix element differences from compared structures.

As with sequence alignments, once pairwise alignment is defined, multiple alignment remains problematic. It is possible to use the multidimensional dynamic programming approach for exact multiple sequence alignments, but it is an infeasible task for more than three sequences of modest length in terms of computational resources (time and memory).

One particular method that combines several approaches and levels of representations is implemented in the Homologous Structure Finder (HSF) software. HSF integrates several different ideas of the structural alignment and classification of multiple protein structures in a single extensible framework. It has been inspired by earlier software such as COMPARE and Structure-Homology-Program (SHP). COMPARE is the first software to take into account several levels of data. However, the levels (*e.g.*, the secondary structures or C α -positions) are handled separately. On the other hand, SHP effectively uses Argos & Rossmann's idea of probabilities to find an optimal alignment of the C α -atoms through a path of dynamic programming matrix. A distinctive feature of the HSF, when deriving the alignment of multiple structures, is the construction of a consensus structural *core*. This core may not resemble the true 3D (sub)structure of any of the structures in the data set and it does not have to follow the linear representation of any of the protein sequences, but it is rather a set of all residues that are equivalent for all structures. Logically, the core should closely resemble the original 3D-structures, if the structures themselves are very similar to each another. The concept of core also enables structural classification of very localized, yet highly conserved, structural features like ion-binding sites, where the position, rather than the type of amino acid, is critical for the establishment of the binding site.

The HSF calculates the final core by doing all-against-all pairwise comparisons of cores in an iterative manner. After each step, the cores yielding the best score are merged to a new core which then represents merged cores in the next iteration. Initially, each structure is considered as a single individual core comprised of all residues. The core can then be used to calculate scores between different structures and the scores then can be converted, in turn, to distances to calculate the final "phylogenetic" tree. Using the core allows one to compare only those parts of the structures that are guaranteed to be present in all structures.

There are two main categories of comparable properties in HSF: invariant and dynamic. Invariant properties (*e.g.*, electric charge and hydrophobicity of a residue) are such that they do not change when structures are moved with respect to each other in the 3D space, whereas dynamic properties (*e.g.*, C α -C α distances) change accordingly. All comparisons of properties are represented as individual matrices of equivalent residues. The way properties are compared depends on the property itself. For example, local geometry or simply sequence similarity can be compared, but understandably both have very different comparison methods. The property matrices can be further used to derive new matrices *e.g.*, a matrix representing secondary structure comparisons between two cores.

As at the start there is no spatial information about distances between residues, HSF takes an unsupervised approach for the initial choice of the equivalent residues by selecting the most probable equivalent residues based only on the invariant properties. After the initial assignment, the first superposition is performed per equivalent residues. After that, the new equivalences are assigned based on the weighted sum of invariant and dynamical properties with an iterative process of superposition and assignment. HSF has been able to classify automatically the PRD1-adenovirus lineage structures (see below) as well as to define a conserved structural core for the non-catalytic ion-binding site in multiple viral RNA-dependent RNA polymerases.

Incorporating Different Levels of Data to Structural Alignment and Classification

Practically, all modern tools for classifying atomic structures use several different layers of data (*e.g.*, sequence-, secondary structure- and 3D-information). The HSF software is not an exception and it already incorporates several properties derived from both sequences (*e.g.*, physiochemical properties of amino acids) and 3D structures (*e.g.*, local dihedral angles and secondary structure information). However, it is still possible to further constrain the classification process with atomic-level information by using for example the overall resolution of the structure and the B-factors per atom. There is also a wealth of other data available from laboratory experiments like protein-protein interactions, cellular localization, biochemical functions and other experimental evidence of biological processes and pathway associations. Indeed, pre-existing annotations are being taken into account, for example, to select the structures in the initial data sets, but since annotations might be also misleading and lead to circular reasoning in separating the classes, they should be used carefully. This is particularly relevant when considering viruses, because the taxonomical assignments can change radically when new viruses are discovered (Fig. 2).

Databases for Structure-Based Classification

Although there are no dedicated structural classification databases for viruses alone, several databases (*e.g.*, SCOP, CATH, HOMSTRAD, FSSP, PDBFold, PFAM) exist that use various methods to classify protein structures or single folds into hierarchical manner. The most widely used ones are SCOP (Structural Classification of Proteins) and CATH (name based on the four levels of classification: Class, Architecture, Topology/fold and Homologous superfamily). Both are viewed as gold standards and they rely

on both expert curation and also automated methods being able to keep up with the current data accumulation. The PDB itself provides classification for known structures with closest structural neighbors by using the jFATCAT-rigid algorithm.

As an example of viruses' uniqueness in structural databases, bacteriophage PRD1's P2 receptor binding protein (PDB ID: 1N7V) has currently no structurally similar entities in the PDB, apart from the same structure in a different crystal form, and no structural neighbors for any of the domains in CATH database. By contrast, *Rattus norvegicus*' Neuroligin-1 protein (PDB ID: 3BIX) has 39 structurally similar entities in PDB and 252 representative domains in CATH.

It is an open question whether protein folds can be classified in discrete hierarchical classes at all, since there is evidence that suggests that the fold space might overall be continuous. A recent version of SCOP (SCOP2) has tackled this problem by replacing the simple taxonomy-like classification with directed acyclic graphs to better capture possible different relationships (*e.g.*, structural and evolutionary relationships are separated).

Inferring Phylogeny Based on Structural Data

Once the classification has been performed for a set of structures, the question is then how to utilize this information to decipher possible evolutionary relationships among the proteins. The most common route is to mimic how sequence alignments are used to generate phylogenies. Like with pair-wise or multiple sequence alignment scores, the structural alignment scores can be converted to estimate "evolutionary distances". The distances can then be used to build an unrooted tree using for example the *de facto* standard phylogenetic analysis package PHYLIP.

Nevertheless, it is not yet clear how accurately the calculated structural distances truly correspond to evolutionary distances, since there is no fossil or archeological data to calibrate the molecular clock for macromolecular evolution. Applied to previously characterized protein families, the structural classification-based phylogenies correspond well to studies using sequence-based protein coding genes. Furthermore, structural phylogenies have been able to unveil distant evolutionary relationships, even when the protein sequences have mutated beyond detectable similarity.

Case Study: Past and Present of the Structure-Based PRD1-Adenovirus Viral Lineage

The finding that both lipid-containing bacteriophage PRD1 and eukaryotes-infecting adenoviruses possess an MCP with a vertical double jelly-roll (DJR) fold has been instrumental in establishing the structure-based approach to infer viral phylogeny. This fold entails two single jelly-rolls connected by a short linker; each jelly-roll consists of eight beta strands arranged in two four-stranded antiparallel beta sheets packed together (topology of one single jelly-roll shown in Fig. 3(a)). The power of the structure-based classification to infer viral phylogeny has been strengthened by the addition of several viral protein 3D structures determined during the last 15 years. The first viral MCP structure-based phylogenetic tree published by Bamford, Grimes and Stuart (see Section "Further Readings") for the PRD1-adenovirus lineage contained three virus members: PRD1, adenovirus and PBCV-1 (Fig. 3(b)). The latter virus was the largest of the three with a diameter of ~ 1900 Å. The assumption that coat protein topology and virion architecture could be used as fossil fingerprints led some of us to explore the use of this approach in grouping other viruses with MCPs with a different fold (a dedicated article can be found elsewhere in the Encyclopedia). It was postulated that viruses could be structurally classified in four lineages initially called PRD1-like, HK97-like, BTV-like and picorna-like. Currently, the classification of the virosphere using the above four structure-based viral lineages accounts for more than 30 viral families, in contrast to the grouping in "order" and "families" used by the International Committee on Taxonomy of Virus (ICTV) until 2019 which accounted for 22. Very recently, however, due to the accumulated evidence of distant viral relationships the ICTV classification scheme has been substantially expanded.

Focusing on the first established PRD1-adenovirus lineage, over the years the members of this lineage have been expanding in some cases leading to unexpected insights. Chronologically, after the structure determination of the hexon MCP (adenovirus), P3 MCP (PRD1) and Vp54 MCP (PBCV-1) for which also the complete virion structures were available and showing the use of pseudo-hexameric capsomers (trimers of the MCP) as their assembly building blocks, two more MCPs were determined by X-ray crystallography: the MCP of membrane-containing Sulfolobus turreted icosahedral virus (STIV), the first virus infecting an archaeal host displaying a DJR capsid protein and the major coat protein P2 from the marine lipid-containing bacteriophage PM2. STIV MCP not only exquisitely superimposed onto the available MCPs of the PRD1-adenovirus lineage but its modeling into the virion map at the time supported a similar principle for virus assembly.

P2-PM2 was the first – and still remains the only – MCP with a minimalist double jelly-roll serving as the simplest prototype for the PRD1-adenovirus lineage. Structural superimposition of the two individual jelly-rolls of P2 (V1, residues 1–141; V2, residues 142–269) shows that they are more similar to each other (2.4 Å rmsd for 116 C α s; 14% sequence identity) than in the corresponding single jelly-rolls from other members of this lineage (Fig. 4). This supports the proposal that the P2 protein arose by gene duplication, suggesting that PM2 virion is more similar to the last common ancestor for this lineage.

It came as a surprise when the structure of vaccinia virus (VACV) D13 protein was solved by X-ray crystallography. Vaccinia virus belongs to the *Poxviridae* family and is characterized by a brick-shaped morphology (not icosahedral as in all previous mentioned viruses) with a lipoprotein surface. D13 is not part of the mature virion but is used as a scaffolding protein during VACV morphogenesis. It was noticed that during virus assembly and maturation, D13 is capable of forming a transitory spherical honeycomb structure providing curvature and rigidity to the convex forming membrane. The building blocks of this layer were trimers of D13 with an apparent pseudo-

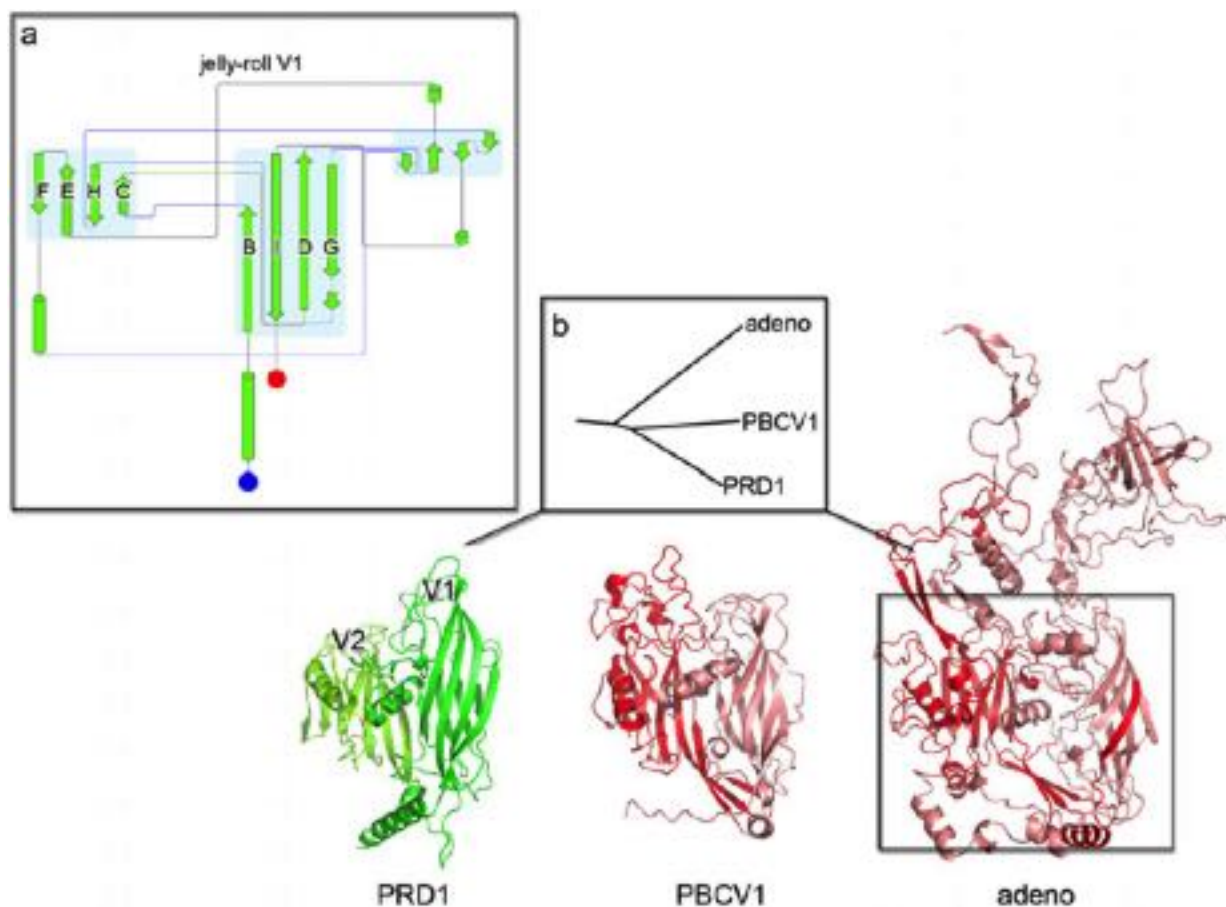


Fig. 3 (a) Example of topology of one of the two single jelly-rolls of the DJR PRD1 P3 MCP: the V1 jelly-roll (strands labeled as BIDG/CHEF), the dark-blue and red circles being respectively the N-terminus and C-terminus of V1 (residues 15–245). (b) First structure-based phylogenetic tree published in 2005 for the definition of the PRD1-adenovirus lineage and obtained using the SHP and PHYLIP software; the sequence identity across the three MCPs is at the most 18% (in the case PRD1 vs PBCV-1). On PRD1 P3 MCP the two single jelly-rolls are marked as V1 and V2 (residues 246–383); across the three MCPs the V1 and V2 are colored in distinct shades of green and red. Adapted from Bamford, D.H., Grimes, J.M., Stuart, D.I., 2005. What does structure tell us about virus evolution? *Current Opinion in Structural Biology* 15, 655–663.

hexameric morphology. This morphology recapitulated the organization of the capsomers formed by the viral DJR MCPs onto the virus facets. Ultimately, the crystal structure revealed that D13 possesses the DJR fold with a “turret” domain as an elaboration of a loop at the top of the V2 domain. Its structure was strikingly similar to the MCPs of the members of the PRD1-adenovirus lineage (Fig. 4).

The structural sampling of the viral world also led to the determination of the MCP structure of Sputnik virophage by cryo-EM to 3.5 Å resolution. Sputnik is a satellite (non membrane-containing) virus of the giant Mimivirus which infects *Acanthamoeba polyphaga* cells. While the MCP of Mimivirus has yet to be determined (there is the proposal that it might possess the DJR fold), the MCP of Sputnik was resolved thanks to the high quality of the virion cryo-EM map. Its overall structure displays a DJR fold with the classical BIDG and CHEF strands and, despite its variation on the structural motifs placed at the top of the single jelly-rolls, it fits nicely in the PRD1-adenovirus lineage (Fig. 4). With the identification of Mimivirus, about 5000 Å in diameter, the perception of viruses as small entities faded away. To date, several even larger viruses have been identified grouping into the assemblage informally known as nucleocytoplasmic large DNA viruses (NCLDVs) and recently officially classified into a phylum *Nucleocytoviricota*. Faustovirus, about 2600 Å in diameter, infects amoeba and its overall structure was resolved by cryo-EM at 15 Å resolution. Strikingly it showed two almost juxtaposed proteinaceous layers. The outermost capsid is organized in capsomers with a pseudo-hexameric appearance consisting of trimers of the MCP which possess a double jelly-roll fold as determined by X-ray crystallography. The capsomers are crowned by large and structured insertions in loops connecting the DE and FG β -strands above the two V1 and V2 β -barrels that protrude toward the outside of the virus.

Exploring novel ecological niches, specifically a Finnish boreal fresh water habitat, the identification and structural characterization of *Flavobacterium*-infecting, lipid-containing phage (FLiP) described a phage with a circular ssDNA with an MCP with a DJR fold structurally closer to PRD1 and PM2 MCPs (Fig. 4). This represented a clear evidence of evolutionary relatedness between icosahedral ssDNA and dsDNA viruses. FLiP rolling-circle replication and structure testify the possible combination of different replication mechanisms with architecture and assembly principles recapitulating those of the PRD1-adenovirus lineage. The latest addition to this lineage has been African swine fever virus (ASFV). ASFV is the etiological agent of a swine viral disease present in

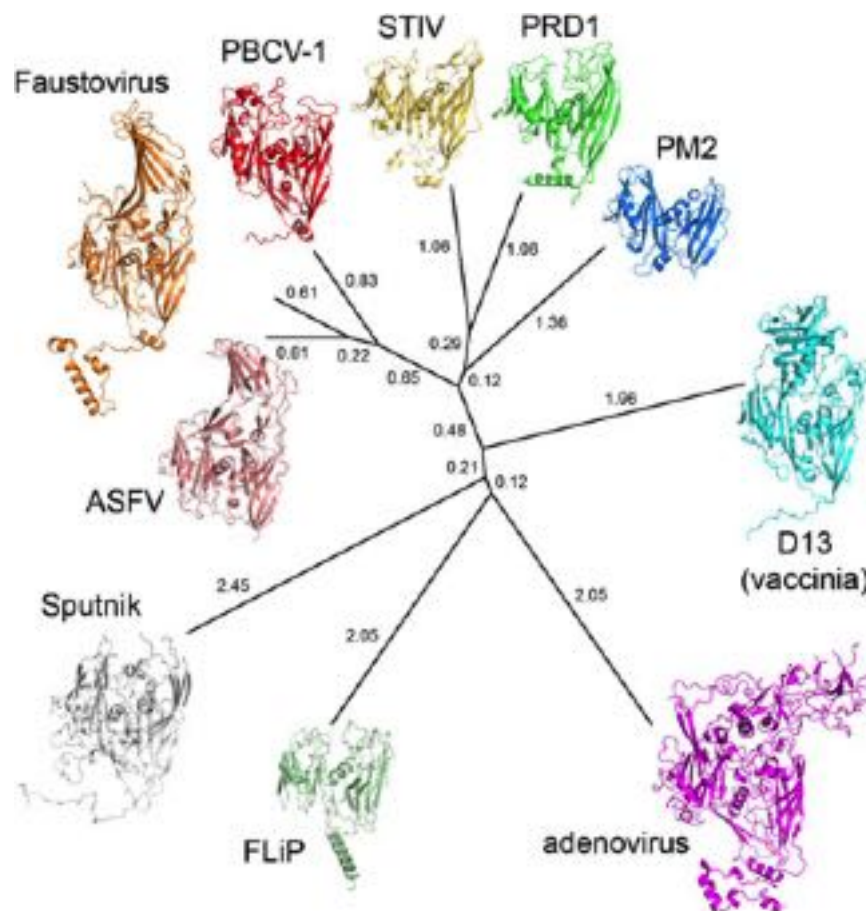


Fig. 4 Current structure-based phylogenetic tree of the PRD1-adenovirus lineage. Structural alignment and evolutionary distances performed with HSF and unrooted phylogenetic tree depicted using PHYLIP. In cartoon and to scale the MCPs corresponding to the different viruses (PRD1: PDB ID 1HX6; PBCV-1: 1M3Y; adenovirus: 1P2Z; STIV: 2BBD; Vaccinia D13: 2YGB; Sputnik: 3J26; Faustovirus: 5J70; FLiP: 5OAC; ASFV p72: 6KU9; PM2: 2W0C). Numbers next to each branch indicate their corresponding lengths (evolutionary distances).

many of the countries in Africa, Europe, and Asia. Due to its recent devastating effects on worldwide pig population, several laboratories have independently determined by cryo-EM its 3D structure and the structure of its major protein components. Strikingly, ASFV virion, with a radial diameter of ~ 2080 Å, displays a unique architecture comprising two icosahedral capsids and two lipoprotein membranes. The outermost capsid shell is composed of 8280 copies of MCP p72, which possesses a DJR fold (Fig. 4). Interestingly, p72 MCP clusters closer with the MCPs of the other NCLDV, and in particular, faustovirus. This NCLDV clustering indicates that structural differences exist in the core consensus DJR and this difference might recapitulate the requirement of these MCPs to stably assemble the largest capsids of this lineage.

In conclusion, the conceptualization of the structure-based PRD1-adenovirus lineage has served to define a new approach to identifying common ancestry, not only within the member of this same clade but across the different known viruses populating the virosphere. The PRD1-adenovirus lineage is nevertheless steadily expanding and providing further evidence for an evolutionary descent across different viruses. In the new ICTV ranking all DJR are now classified as Kingdom (official taxon) carrying the taxon name of *Bamfordvirae*. It is likely that this lineage (or Kingdom) will accrue more members as recently it has been suggested that a new tentative family of viruses called *Autolykiviridae* may also possess MCP with a DJR fold. Further, the recent structure determination by high-resolution cryo-EM of novel viruses with two vertical single jelly-roll MCPs (*Spherolipoviridae* family) forming capsomers with a pseudo-hexameric footprint is also contributing to deciphering the evolutionary relatedness between these viruses and those of the PRD1-adenovirus lineage. In this light, it is enticing to speculate the existence of an ancestral virus with only one vertical single jelly-roll MCP capable to assemble into pseudo-hexameric capsomers that would anticipate the assembly of viruses of the PRD1-adenovirus lineage. The search is on!

Conclusions and Future Perspectives – New Methods for Classifying Biological Entities Including Viruses

The known protein universe is highly diverse from the sequence point-of-view, especially for viral proteins, whereas the actual fold-space is constrained to a handful of common folds. Having atomic level information of proteins therefore enables

classification at the longer evolutionary timescales. This, in turn, is essential to understand protein-protein interactions within the viral capsids and to recognize active sites when only the amino acid position in 3D structure is conserved.

The current high-throughput data acquisition methods and initiatives like Structural Genomics initiative produce enormous amounts of both sequence and structural data. In particular, cryo-EM methods have progressed to be able to routinely reach atomic resolutions and the derived protein structures can now be used along with the crystal structures. With sequence based viral phylogenetics, it is possible to estimate molecular clock for highly mutating RNA-viruses by yearly sampling and sequencing. So, theoretically, similar method could be also used for generating data to calibrate molecular clock for structural phylogenetics.

Currently, there are tens of tools comparing and aligning protein structures. Thus, the alignment problem and especially the alignment of multiple structures is an active and important area of research with real life implications *e.g.*, for understanding the molecular principles of disease causing mutations and for developing antiviral drugs. Some of the challenges that lie ahead for computational classification based on structural data are – to mention a few: (1) how to compare intrinsically disordered proteins; (2) how to perform alignments of flexible domains efficiently and (3) to score such alignments. To infer structural phylogenies, one of the questions is whether it would make sense to describe some of the evolutionary relationships between multidomain structures as set of trees or networks of protein domains instead of a single tree-like hierarchy.

There are also completely new computational methods: for example the Helsinki-Okinawa sequence similarity (HOSS) tool, that allows us to penetrate further into the ‘Twilight Zone’ of the sequence similarity and so being able to reconstruct the PRD1-adenovirus lineage from sequences directly. Similarly the AlphaFold algorithm, a recent breakthrough that can predict in certain cases protein structures from sequences alone.

In conclusion, with the current experimental and computational methods we are approaching a level that at least allows us to efficiently use the rules that govern protein folding. It is also foreseeable that in the future new computational methods relying on vast sequence and structural databases and advanced machine learning approaches (*e.g.*, deep learning with recurrent neural networks) will finally bring us to a point where a complete protein structure can be inferred from the sequence alone.

Acknowledgments

We are grateful to Gabriel Soares (Abrescia Lab) for useful comments on the article. J.J.R. is supported by the Faculty of Biological and Environmental Sciences/Molecular and Integrative Biosciences research program, University of Helsinki. N.G.A.A. is supported by the Spanish *Ministerio de Ciencia, Innovación y Universidades* (RTI2018-095700-B-I00), by the Basque *Departamento de Educación, Política Lingüística y Cultura* (Refs: PRE_2016_2_0151, PRE_2018_1_0102) and by “La Caixa” Foundation INPhINIT programme. MICINN is also grateful for the Severo Ochoa Excellence Accreditation to the CIC bioGUNE (SEV-2016-0644).

Further Reading

- Abrescia, N.G., Bamford, D.H., Grimes, J.M., Stuart, D.I., 2012. Structure unifies the viral universe. *Annual Review of Biochemistry* 81, 795–822.
- Bamford, D.H., Grimes, J.M., Stuart, D.I., 2005. What does structure tell us about virus evolution? *Current Opinion in Structural Biology* 15, 655–663.
- Mönttinen, H.A.M., Ravantti, J.J., Poranen, M.M., 2012. Evidence for a non-catalytic ion-binding site in multiple RNA-dependent RNA polymerases. *PLoS One* 7 (7), e40581. doi:10.1371/journal.pone.0040581.
- Ng, W.M., Stelfox, A.J., Bowden, T.A., 2020. Unraveling virus relationships by structure-based phylogenetic classification. *Virus Evolution* 6 (1), veaa003. doi:10.1093/ve/veaa003.
- Ravantti, J.J., Bamford, D., Stuart, D.I., 2013. Automatic comparison and classification of protein structures. *Journal of Structural Biology* 183 (1), 47–56. doi:10.1016/j.jsb.2013.05.007.
- Senior, A.V., Evans, R., Jumper, J., *et al.*, 2020. Improved protein structure prediction using potentials from deep learning. *Nature* 577, 706–710.
- Sinclair, R.M., Ravantti, J.J., Bamford, D.H., 2017. Nucleic and amino acid sequences support structure-based viral classification. *Journal of Virology* 91 (8), e02275. doi:10.1128/JVI.02275–16.
- Vogel, C., Bashton, M., Kerrison, N.D., Chothia, C., Teichmann, S.A., 2004. Structure, function and evolution of multidomain proteins. *Current Opinion in Structural Biology* 14 (2), 208–216. doi:10.1016/j.sbi.2004.03.011.

Relevant Websites

- <http://ekhidna2.biocenter.helsinki.fi/dali/>
Dali server
ekhidna.biocenter.helsinki.fi.
- <https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses (ICTV).
- <http://scop.mrc-lmb.cam.ac.uk/SCOP>
Structural Classification of Proteins.
- <http://viperdbscripps.edu/>
Welcome to VIPERdb
Scripps Research.
- <https://www.wwpdb.org/>
wwPDB: Worldwide Protein Data Bank.

Isolating, Culturing, and Purifying Viruses With a Focus on Bacterial and Archaeal Viruses

Katri Eskelin, University of Helsinki, Helsinki, Finland

Hanna M Oksanen, Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AF4 Asymmetrical flow-field flow fractionation

dsDNA double-stranded DNA

EOP Efficiency of plating

g-force Force of gravity

ICTV International Committee on Taxonomy of Viruses

MOI Multiplicity of infection

PEG Polyethylene glycol

pfu Plaque forming unit

prokaryotes Bacteria and archaea

RCF Relative centrifugation force

RPM Revolutions per minute

S Svedberg unit

ssDNA single-stranded DNA

w/v Weight per volume

Glossary

Lysogeny Non-productive virus life cycle where viral DNA is integrated into host genome or replicated independently while the host growth is not affected and no new virus progeny made. See also provirus.

Lytic life cycle Productive virus life cycle where new virus progeny is made and released by host cell lysis.

Plaque Originates from a single virion initially infecting a single cell. Visible area on dense host lawn caused by virus replication and result of host cell lysis or growth inhibition of host cells.

Provirus A viral genome integrated into the host genome.

Replication cycle Refers to events where viruses use host's cellular machinery and metabolites to make copies of

themselves. It involves host recognition, attachment, entry, uncoating, genome multiplication, mRNA and protein synthesis, assembly, and release.

Sedimentation rate Of a particle is the speed of the travel in solution subjected to high force of gravity (g-force) upon centrifugation.

Svedberg unit (S) is a measure of the sedimentation rate of a particle when subjected to centrifugation.

Temperate virus A virus capable to replicate by lysogenic and lytic cycles.

Virion Infectious mature virus particle.

Virulent virus A virus replicating via lytic cycle.

Bacterial and Archaeal Viruses Come in Different Sizes and Form a Polyphyletic Group of Viruses

Viruses are large macromolecular complexes capable of transmitting their genomes from a host cell to another. Bacterial viruses (i.e., bacteriophages) and archaeal viruses infect prokaryotes (bacteria and archaea) that are single-cell organisms lacking a nucleus and organelles. Viruses of bacteria and archaea are polyphyletic in their origin and form a diverse group of viruses differing in terms of virion morphologies, genome types, sequences, life cycles, and host ranges. There are also a substantial number of proviruses found in the chromosomal sequences of bacteria and archaea. Currently the archaeal and bacterial viruses i.e., prokaryotic viruses, have been classified by the International Committee on Taxonomy of Viruses (ICTV) into ~30 families, but there are also a significant number of unclassified prokaryotic viruses. The recent burst in sequence data of prokaryotic virus genomes has revealed that viruses are more diverse than previously recognized. Thus, the current official taxonomic framework of the ICTV does not adequately reflect their evolutionary relationships.

Bacteriophages come in many sizes and shapes that include icosahedral, enveloped, pleomorphic, and helical morphologies, and combinations of these, whereas the virion morphotypes of isolated archaeal viruses are even more versatile (Fig. 1). Some of the virus morphotypes such as icosahedral tailed and icosahedral internal membrane-containing virus types are shared by archaeal and bacterial viruses, whereas certain morphotypes, such as spindle-shaped, bottle-shaped, and droplet-shaped, are specific for archaeal viruses only and have not been discovered for viruses infecting bacteria or for eukaryotic viruses (Fig. 1). The size range of the prokaryotic virions varies from ~25 nm (e.g., bacteriophage ϕ X174 in the family *Microviridae*) to large and complex jumbo bacteriophages (e.g., *Bacillus megaterium* phage G, an icosahedral tailed virus in the order *Caudovirales*) with a capsid size of ~160 nm and a tail length of ~450 nm. Filamentous single-stranded (ss) DNA phages can be long and their length is mainly dependent on the genome size. Lengths of inoviruses range from ~700 nm to ~3700 nm for *Pseudomonas* viruses Pf3 and Pf4, respectively. Additional diversity to the virus morphologies is obtained due to the variations in appendices on the virion surface e.g., spikes and tails that play a role in host recognition and thus vary a lot between viruses. Viruses with highly symmetric icosahedral particles (e.g., family *Corticoviridae* and order *Caudovirales*) produce a very homogenous population of progeny virions. In some virus systems that include pleomorphic or

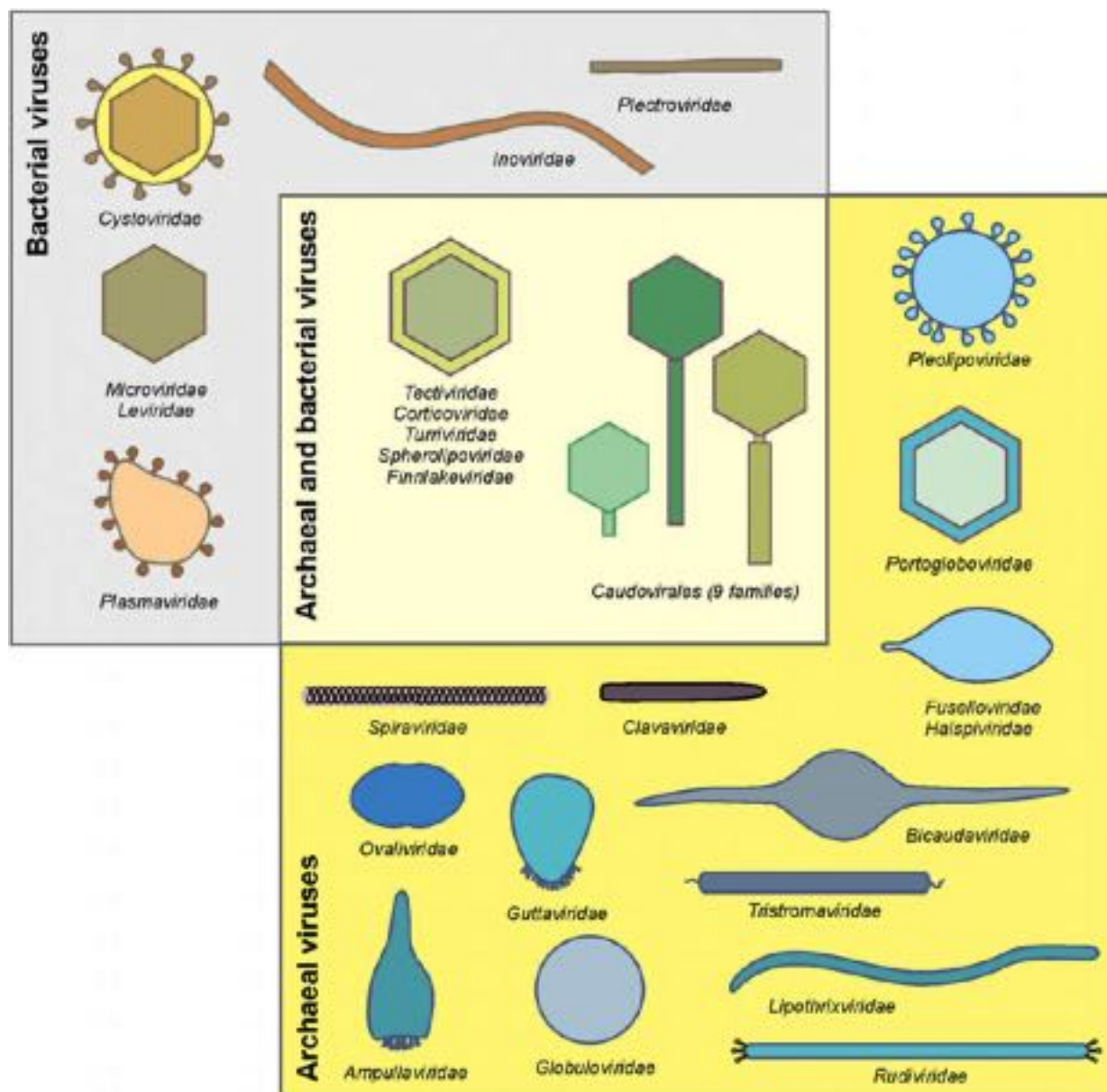


Fig. 1 Virion morphotypes of bacterial and archaeal viruses. The virus family or order based on the ICTV for each morphotype are indicated. The schematic presentations are not drawn on scale.

spindle-shaped archaeal viruses (families *Pleolipoviridae* and *Fuselloviridae*) particle sizes and morphologies can be more variable due to the floppy appearance of the formed particles and the random distribution of surface structures.

At the simplest, virions i.e., the infectious mature virus particles, consist of a protective shell made of protein and nucleic acid that is encapsidated inside the shell. Virus proteins are mostly encoded by viral genes and can have very different physicochemical properties based on their amino acid compositions and tertiary structure. Some prokaryotic viruses have lipids in their virion. Viruses typically do not have lipid metabolism and the lipids are acquired from the host cell membranes during virus assembly. Viral lipid composition can be identical to that of host, but some viruses favor selective uptake of the host lipid species. Viral lipids are either in the form of an internal membrane vesicle underlying the icosahedral protein capsid (e.g., viruses in the families *Tectiviridae*, *Corticoviridae*, *Turriviridae*, *Fimlakeviridae*, and *Spherolipoviridae*) or as an outer envelope e.g., *Pseudomonas* virus $\phi 6$ in the family *Cystoviridae* (enveloped bacteriophages with a double-layered protein capsid) and *Acidianus* virus AFV1 in the family *Lipotrixviridae* (filamentous archaeal viruses). Lipid modifications can also be found as in the case of the major capsid protein of the spindle-shaped *Haloarcula* virus His1 (family *Halspiviridae*). Pleomorphic virions of the viruses in the families *Plasmaviridae* (bacteriophages) and *Pleolipoviridae* (archaeal viruses) resemble simple extracellular membrane vesicles. Bacteriophage genomes can be either DNA or RNA – single-stranded (ss) or double-stranded (ds). The known DNA genomes of bacteriophages are circular or linear, whereas ssRNA genomes are circular and dsRNA

genomes are segmented. No RNA genome is known for archaeal viruses, their genomes are either dsDNA (circular or linear) or ssDNA (circular). The most common genome type among the archaeal and bacterial viruses is dsDNA.

Virions have a dual role. They have to be stable enough to protect the genome but also capable to attach to an appropriate host cell, to disassemble, and to deliver the viral genome into the cell. Consequently, virions are metastable structures able to change their conformation. Viruses utilize various strategies to be able to fulfill the two opposite tasks. Thus, the particles respond differently to fluctuating physico-chemical conditions and are dependent on variable compounds e.g., divalent cations or a certain pH for integrity and optimal infectivity. Archaeal and bacterial viruses originate from moderate to very extreme environments and have adapted to “live” with and within their host cells. Consequently, the requirements for viral growth and maintaining the stability and infectivity of the virions are very diverse. The sensitivity of the virus and its host to pressure, oxidation, temperature, pH, high or low ionic strength etc. has a major impact on the virus production methods that can be applied and also affects the feasibility of the scale up process.

Isolation and Culturing of Bacterial and Archaeal Viruses

Viruses are ubiquitous in microbial ecosystems and their abundance (estimated to be over $\sim 10^{31}$) is a hundred-fold higher than the total cell number on Earth. Metagenomics approaches have demonstrated that viruses represent the largest reservoir of genetic diversity on the planet. Viruses play significant roles in microbial ecosystems and in microbial evolution and pathogenesis, and influence host population dynamics and their mortality, hence affecting microbial processes and nutrient cycling. Bacterial and archaeal viruses can be found wherever their host cells inhabit. Water, soil, and sediments are full of prokaryotes and their viruses. Some, often extreme biotopes, such as highly saline or high temperature acidic environments, are enriched with archaea.

The basic method of phage and archaeal virus isolation has remained unchanged since the discovery of bacteriophages by Felix d’Herelle 1917. The first archaeal virus was found later in 1974. During recent years, there has been a burst of new archaeal viruses described especially for euryarchaea and crenarchaea. The simplest way to isolate bacteriophages and archaeal viruses is to mix a virus-containing environmental or clinical sample with a potential host strain(s) in appropriate growth conditions typically selected based on the host strain growth requirements. After incubation of the virus-host mixture (enrichment culture), the cells are removed by centrifugation or filtration or a combination of both and the presence of viruses assayed from the supernatant or filtrate flow-through. The most evident sign of a new virus is a plaque formed on a lawn of susceptible host strain cells (**Fig. 2(a)** and **(b)**). The plaques can be clear or hazy depending on the way the viruses are released from the cell (see below). Sometimes enrichment cultures are not necessary and samples with a high concentration of viruses can be plated directly with a susceptible host strain to isolate new viruses. In a hunt for new viruses, it is important to control that the strains used for isolation are not releasing temperate viruses in the applied experimental conditions.

Currently available techniques limit the diversity of viruses that can be isolated, since all of them induce a bias towards those viruses that can multiply most efficiently under the experimental conditions selected. In addition, it has been estimated that less than 10% of all microbes can be cultivated in laboratory conditions. However, several parameters can be optimized including the composition of the growth media, growth temperature and aeration to permit greater success in virus isolation. To obtain further insights into the real diversity of viruses and to initiate in-depth studies on virus-host interactions of ecologically relevant virus-host systems, future work should focus on improving growth methods and developing new microbial cultivation protocols. Importantly, experimental laboratory work on well-defined virus-host model systems isolated from environmental samples provides tools to understand the immense genetic diversity and the role of viruses in the ecosystems and enables studies of virus function at a molecular level and structures up to high-resolution.

There is no single method to isolate a new virus. The best way to start is to go where the potential host strains are. The virus isolation can be accomplished on a single prokaryotic strain, but utilization of a mixture of multiple host strains or multiple strains of the same species usually increases the probability of finding a virus. In the case of multiple hosts, however, it is essential to test them for antagonism that could suppress or interfere with the growth of other microbial species or strains. Use of autochthonous host strain isolates originating from the same environmental samples might also increase the probability for successful virus isolation. Traditionally, most viruses have been isolated by first culturing their hosts and afterwards isolating viruses for the established archaeal or bacterial culture. Obviously, selected growth conditions and isolation hosts set the limits for viruses to be found. In addition, practical reasons play a role when choosing the isolation host: challenging laboratory growth conditions or pathogenic bacterial strains limits their easy use. Importantly, even though a virus was originally isolated from a certain host, it may not be optimal for virus growth and higher efficiency of plating (EOP) may be possible on alternative host strains that can be advantageous in subsequent purification attempts. Non-pathogenic variants of the host species are often beneficial when setting up model system for laboratory experiments.

Prior to virus isolation, the autochthonous cells are usually removed from samples to avoid multiplication of the cells during the enrichment thus allowing specific selected host strains to be used as baits for viruses. However, sometimes the endogenous host strains present in the sample can be beneficial for multiplication of viruses during enrichment culturing. In general, any bias introduced by sample processing should be avoided whenever possible. For samples with low virus concentration, utilizing filtration, precipitation or a combination of both (see below), before the virus isolation may be advantageous. Chloroform treatment of samples is occasionally used during virus production and storage to remove the cellular organisms from samples or to release some viruses from the cells. However, extra caution is needed when using chloroform, since its usage limits the potential to isolate and grow both membrane-containing and filamentous viruses. Even some icosahedral tailed viruses are inactivated or their infectivity is reduced in the presence of chloroform.

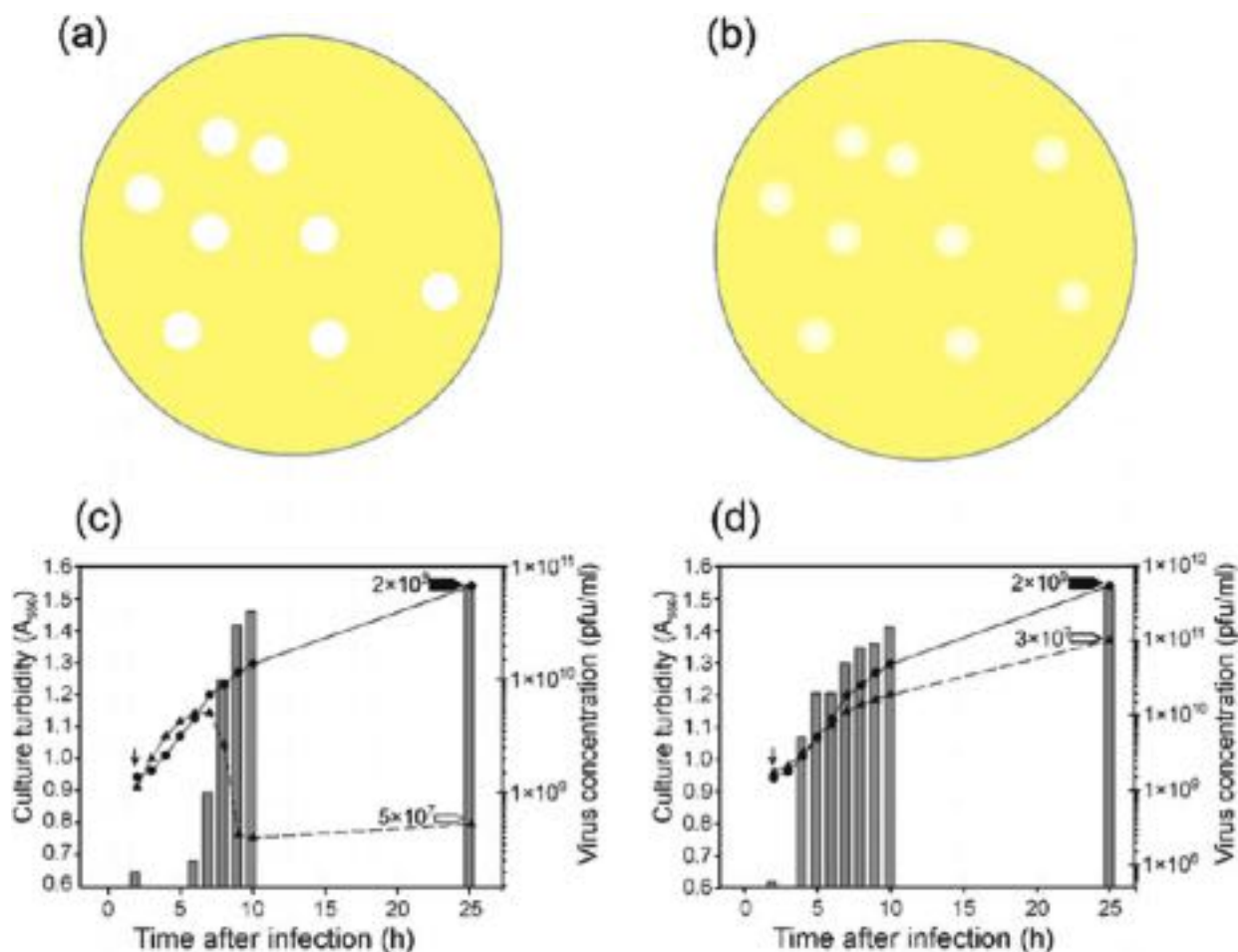


Fig. 2 Plaque morphologies. (a) Clear plaques and (b) hazy plaques on a host cell lawn (yellow). Each plaque originates from a single virus infection and can be used to enumerate the number of infectious viruses in the sample. Virus infection can induce (c) host lysis (the life-cycle of the archaeal icosahedral tailed virus HHTV-1 in *Haloarcula hispanica* as an example) or (d) retardation of host growth in the absence of lysis (the non-lytic life-cycle of the archaeal pleomorphic virus His2 in *Haloarcula hispanica* as an example). Consequently, viruses causing host lysis can be typically observed as clear plaques (see Panel a), whereas viruses inhibiting the host growth can be observed as turbid plaques (see Panel b). The onset of lysis and the amount of host growth retardation is specific for each virus-host system and can vary from tens of minutes to several days depending on the host physiology as well. Turbidity of the uninfected (black circles) and infected (black triangles) cultures; the number of free progeny viruses (plaque forming units/mL; pfu/mL) in the infected cultures (gray bars); and the number of viable cells (colony forming units/mL) at 25 h post infection in the uninfected (black arrow head) and infected cultures (white arrow head) are shown. The Panels c and d are modified from Svirskaitė, J., Oksanen, H.M., Daugelavičius, R., Bamford, D.H., 2016. Monitoring physiological changes in haloarchaeal cell during virus release. Viruses 8, 59. doi:10.3390/v8030059. Copyright © 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license <http://creativecommons.org/licenses/by/4.0/>.

After successful virus isolation, pure virus cultures are obtained by a subsequent single-plaque purification technique. Prokaryotic viruses have two major life cycle strategies. They can be virulent and capable of conducting their replication cycle in a cell releasing progeny virions by host cell lysis (lytic replication cycle). Temperate viruses can follow two alternative replication cycles: lysogenic and lytic. During the lysogenic cycle, the virus genome is reversibly integrated into the host chromosomes or replicates as a plasmid (or episome) in such a way that it does not lead to virus progeny production. This “silent” viral genetic element that is part of the host chromosome is called a provirus. They are numerically very significant components of bacterial and archaeal genomes. Spontaneously or under certain conditions, proviruses can switch to a lytic virus replication cycle. To isolate such viruses, proviruses can be induced from bacterial and archaeal strains by physiological stimuli (e.g., elevated temperature for psychrophiles or lowered ionic strength for halophiles) or exposing the strains to inducing agents that impair DNA integrity and cause DNA damage (e.g., mitomycin C and ultraviolet light). Culture-based experimental work on lysogenized strains has resulted in a number of new recent virus isolates e.g., the temperate *Aeropyrum pernix* spindle-shaped virus 1 and *Aeropyrum pernix* ovoid virus 1 of the hyperthermophilic archaeon *Aeropyrum pernix* and the temperate archaeal pleolipovirus SNJ2 originating from the halophilic archaeal *Natrialba* strain. The classical example of a temperate bacterial virus is the coliphage λ . Importantly, currently

available public sequence databases provide a significant culture-independent way to find new potential proviruses based on a homology search. However, the isolation of such viruses is still requires finding sensitive and cultivable hosts.

Progeny archaeal and bacterial viruses are released from host cells either by cell lysis at once or continuously through the cell envelope only causing host cell growth retardation (non-lytic life cycle) (Fig. 2). The difference between such virus exit strategies can be seen in synchronized one-step growth experiments by using high multiplicity of infections (MOI) to ensure that most of the cells are infected (Fig. 2(a) and (b)). In both cases by using optimized conditions, a high number of virus particles can be produced for certain virus systems. Production of prokaryotic viruses on solid media by using a double agar overlay plaque assay method is often more efficient than a liquid culture production method. However, some temperate viruses replicate poorly in liquid medium resulting in incomplete lysis, whereas some bacteria or archaea do not form lawns on solid culture media, and consequently the use of a plaque assay to determine the number of infectious viruses or solid media based virus production techniques cannot be used. Solid-media-based virus production systems are more effortless to set up and require typically less optimization than virus production in liquid medium. However, the liquid-medium-based growth systems are easier to scale up. Optimization of the culturing conditions for efficient virus production start from finding the optimal host for the virus and the optimal growth conditions for the host strain to support the maximal virus production. Several variables can be optimized for efficient virus production. Host cell growth, growth phase, cell density at the time of infection, MOI value, temperature, growth media composition, and optimal amount of the inducing agent are the major parameters to be optimized.

Purification of Viruses

Methods and techniques used for purification of bacterial and archaeal viruses are universal and applicable for eukaryotic viruses and other large macromolecular biocomplexes as well. The biophysical and biochemical properties of the virus particles such as size, shape, density, composition, homogeneity, net charge, hydrophobicity as well as sensitivity to shear, osmotic pressure, temperature, pH, and ionic strength affect the purification methods that can be applied (Fig. 3). For instance, enveloped viruses are more sensitive to shear-induced damage than are viruses having proteinaceous capsids. Filamentous particle morphology or the presence of long and fragile appendices e.g., spikes or tails, may make viruses sensitive to shear as well. Lipid components render viruses sensitive to organic solvents and detergents, whereas halophilic or thermophilic viruses living in acidic environments are often sensitive to low ionic strength or basic conditions.

Purification aims to remove impurities that originate from the host, culture media, and purification process itself. The heterogeneity of the produced virus particles (e.g., in size and/or density), the ratio of viruses to other sample components, the complexity of the starting material, and the initial concentration of viruses affect the purification methods that can be applied. In addition, the intended use of the virus specimen determines the needed quality, quantity, and concentration of the final purification product. In general, the more purification steps there are, the less the final yield will be and the higher the production cost. Luckily, many techniques are rapidly evolving in directions that require less sample. Typically, the initial purification steps aim to remove the most abundant impurities and to concentrate the sample to enable subsequent purification steps, where small sample volumes are beneficial, allowing higher throughput. Subsequent purification steps typically aim to remove the remaining impurities and to concentrate the sample.

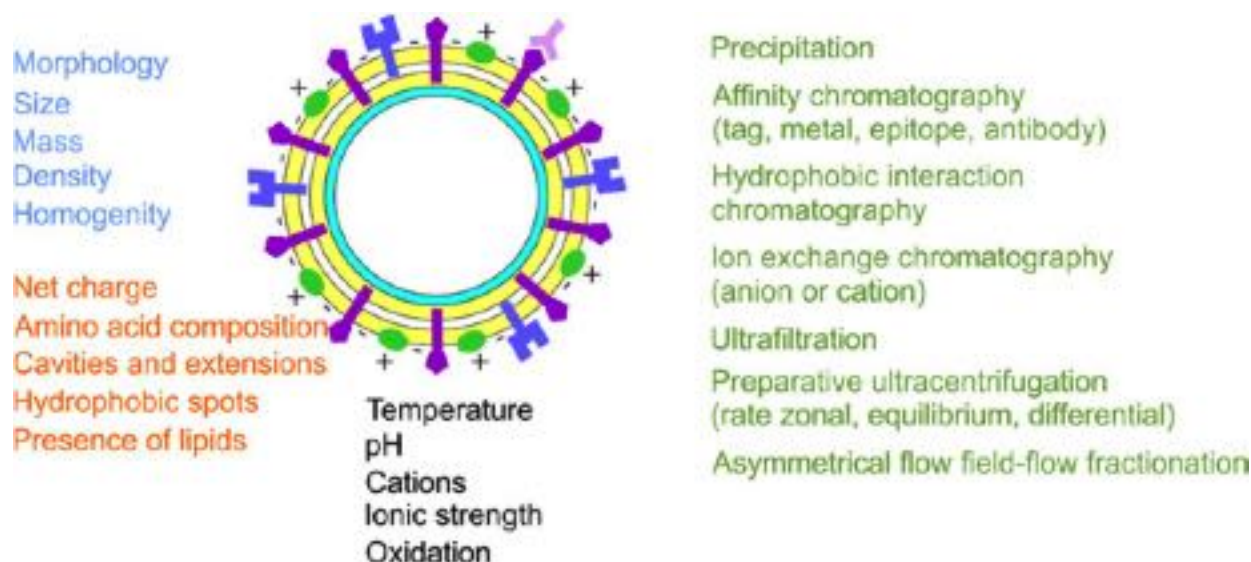


Fig. 3 Different physico-chemical properties of virus particles and experimental conditions (blue, orange, and black) that have an influence on the selection of the purification methods (green). Depending on the virion properties, various chromatography methods can be combined with precipitations, ultrafiltration and ultracentrifugation methods to result in a highly concentrated virus specimen of high quality.

Whether the progeny virions are released by host cell lysis or in a non-lytic way e.g., budding (Fig. 2(c) and (d)) has an impact on the complexity of the starting material. Starting materials that result from viruses that are released from the cells without host lysis often have fewer impurities. In addition, the released viruses mostly represent intact mature functional viruses. In contrast, upon virus-induced cell lysis, mature viruses are released concomitantly with other cellular components including immature virus particles e.g., procapsids and other subassemblies. Thus, lysates are often more complex and heterogeneous starting materials for purification. However, archaeal and bacterial cells are devoid of cellular organelles which is an advantage when compared to the purification of eukaryotic viruses. Virus particles are generally relatively easy to separate from large-sized cellular debris and intact cells by high-speed centrifugation, microfiltration or a combination of these techniques. The biggest challenges in the purification arise from other macromolecular complexes having properties similar to the virus of interest. Selective removal of certain unwanted sample components at the beginning of the purification process typically improves separation in subsequent steps. For instance, the removal of contaminating nucleic acids (chromosomal DNA and ribosomal RNA) from lysates reduces the viscosity of the sample and dissolves complexes of nucleic acids and proteins. Large amounts of chromosomal DNA can hinder the use of precipitation methods as polyethylene glycol (PEG) precipitation in the presence of NaCl also efficiently precipitates DNA causing problems in the down-stream purification processes (see below). Addition of detergents usually increases the solubility, but may be detrimental for the integrity of membrane-containing and filamentous viruses. Importantly, the subsequent purification methods have to ensure that the added chemical components can be successfully removed if needed.

Ultrafiltration Methods

Ultrafiltration is an easy non-selective purification method, where separation is based on the size (Fig. 4). It requires no specialized equipment and little optimization and is thus rather inexpensive. Successful filtration purifies and concentrates the virus specimen. Filtration is driven by hydrostatic pressure that forces the liquid (buffer) and small sample components to pass through the semipermeable membrane. Separation capacity is determined by the pore size of the ultrafiltration membrane. Sample components larger than the pores are retained in the concentrate, whereas smaller components are transferred to the flow through (Fig. 4). Ultrafiltration is applicable to viruses that are smaller than ~100 nm in diameter. However, due to the size-based separation it cannot separate viruses from impurities that are larger than the pores of the ultrafiltration membrane that is used.

Ultrafiltration is a relatively fast purification method. Its speed depends on the pressure difference between the surface of the membrane and the flow through, and is affected by membrane porosity. Adsorption of the sample components to the membrane or blockage of the pores may slow down the filtration speed and result in low yields. Reduced permeability of the membrane can also increase the ion concentration and osmotic pressure of the concentrate that can be harmful for viruses. In such cases, testing different membrane materials and pore sizes, may improve the purification speed and yield.

The above described dead-end filtration, where the flow of the solution is perpendicular to the membrane surface, is the most typical filtration technique used. However, cross flow filtration (or tangential flow filtration) can also be applied. In cross-flow filtration the buffer flows parallel to the membrane surface reducing the accumulation of the sample components at the membrane surface and resulting in improved yields. Passive dialysis through membranes with specified pore sizes is also a valid and simple method to separate small and large sample components from each other. Importantly, filtration techniques are useful, if the buffer composition of the sample needs to be changed.

Precipitation Methods

Precipitations are simple to perform, require no specialized equipment, are inexpensive, and have generally good yields. Successful precipitation results in purification and concentration of the sample. The most commonly used precipitant for viruses is high

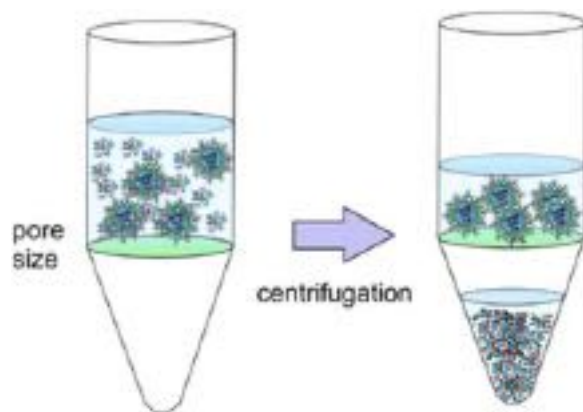


Fig. 4 Ultrafiltration. The pore size of the ultrafiltration membrane determines the size of components that are retained in the concentrate and those that are filtered through. Ultrafiltration speed is often enhanced by low speed centrifugation.

molecular weight polyethylene glycol (PEG 6000 or 8000) as it works for dilute and concentrated virus solutions, at variable pH (6–8 at least), and up to 2 M ionic strength. It is usually used in concentrations of 2–10% (w/v). Typically, virus-sized particles require lower concentrations of salt or polymer for their precipitation than small soluble proteins. Efficient PEG precipitation requires the presence of at least 0.5 M NaCl. Filamentous and helical viruses precipitate with lower PEG concentrations than the spherical ones. Importantly, PEG concentrations above 5% (w/v) precipitate cellular DNA, ribosomes, and vesicles. This needs to be taken into account of in the sample preparation before or after precipitation. Other polymers and salts such as ammonium sulfate can also be successfully utilized for virus precipitation. Concentration of the precipitating agent, time, and temperature affect the precipitation process. Sometimes step-wise precipitations with increasing amounts of the precipitant may improve purification. Precipitation as the first purification step is usually advantageous as it concentrates the virus preparation and removes several host-derived components.

Preparative Ultracentrifugation Methods

All particles gradually sediment downwards due to gravity. This process is slow for virus-sized particles and can be enhanced by using ultracentrifugation. Ultracentrifuges can accelerate an appropriate rotor up to the speed of 60,000–100,000 revolutions per minute (RPM), generating a relative centrifugal force (RCF) higher than $350,000 \times g$, where $1 \times g$ is equal to the force of gravity at the Earth's surface. Available techniques are analytical and preparative ultracentrifugation. Analytical ultracentrifugation is a powerful method for quantitative analysis of macromolecules in solution, whereas preparative ultracentrifugation is commonly used for purification of large biocomplexes, although it can also be used for certain analytical purposes. Preparative ultracentrifugation is highly selective and widely used for purification of viruses. The separation that can be achieved is based on the size, shape, mass, and density of the virus particles which determines the sedimentation coefficient specific for each virus. Obviously, the sedimentation speed is proportional to the applied relative centrifugal force. In general, larger sample components with higher masses sediment faster than the smaller ones. By using optimized protocols, the specimen obtained is typically pure, homogenous, and of high quality. Importantly, the same running conditions are usually suitable for viruses of similar sizes, densities, and morphologies. Filamentous material such as bacterial flagella and archaeal archaella, and extracellular vesicles are common impurities and challenges for virus purification by ultracentrifugation. Combining consecutive ultracentrifugation separation techniques (rate zonal and equilibrium) typically results in good recoveries and purities.

Two different preparative ultracentrifugation separation techniques can be distinguished: differential centrifugation and density gradient centrifugation (Fig. 5). Density gradient ultracentrifugation can be further divided into rate zonal and equilibrium (isopycnic) centrifugation. Differential ultracentrifugation is typically performed using fixed angle rotors (typical angles 15° to 45°), whereas rate zonal and equilibration centrifugation uses swing-out rotors (Fig. 5). The centrifugal force obtained during centrifugation depends on the rotor and more specifically on the distance between its center and the position of the sample components

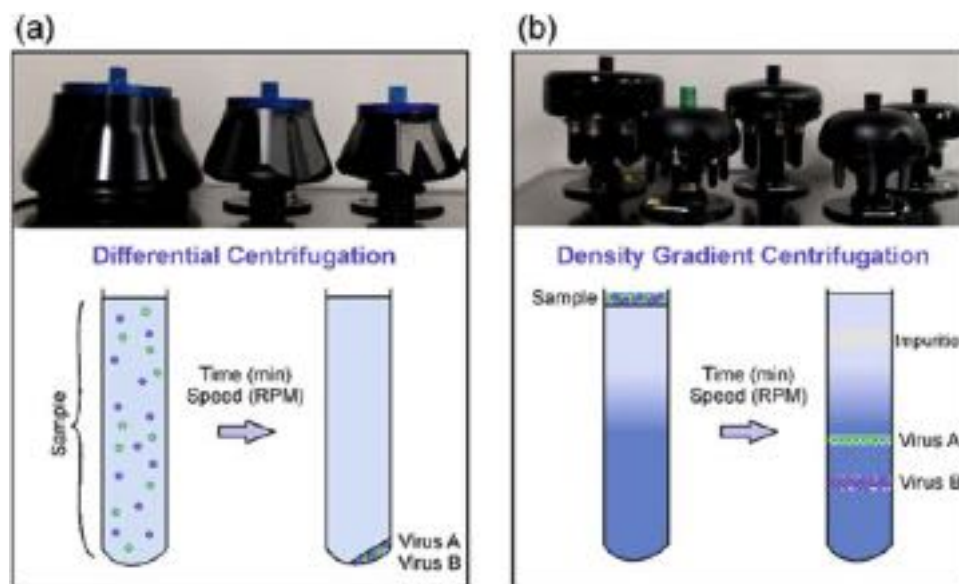


Fig. 5 Different modes of preparative ultracentrifugation. (a) Separation in differential centrifugation is based on the size (diameter, mass) of the particles. (a, top) Fixed-angle rotors are typically used for this separation technique. (b) There are two different modes of density gradient centrifugation: rate zonal centrifugation and equilibrium centrifugation. Both separation techniques use density gradient solutions and the sample is typically loaded on top of the gradient material. In rate zonal centrifugation, particles move at different rates depending on their sedimentation coefficients. In equilibrium centrifugation, particles form zones at the position in the gradient that is equivalent to their buoyant density. (b, top) Swing-out rotors of different volumetric and maximal speed limits are available and typically used for density gradient centrifugation.

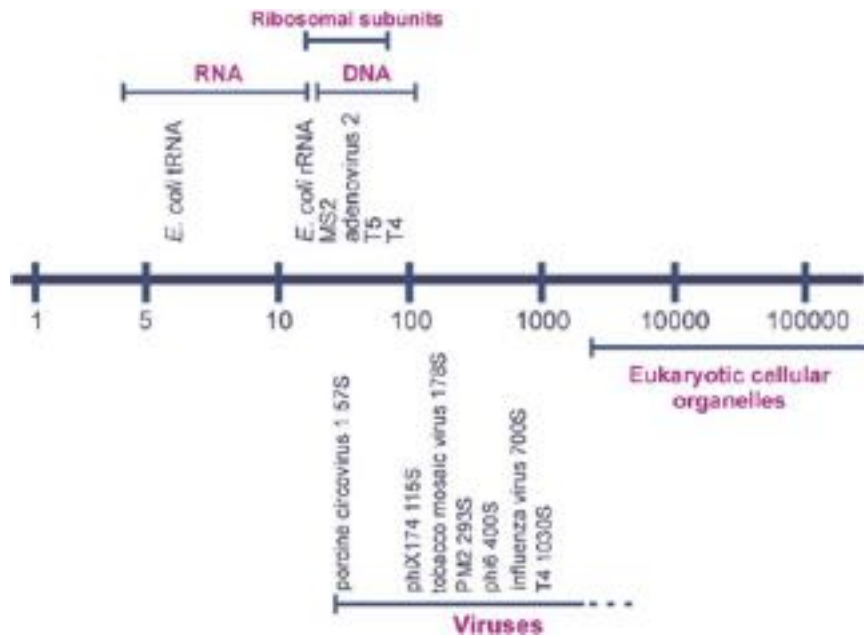


Fig. 6 Comparison of S-values of some bacteriophages (phiX174, PM2, phi6, and T4) to other cell components and some eukaryotic viruses.

in the tube during the run. Importantly, the centrifugal force is different along the length of the tube, being the highest at the bottom of the tube. This difference is compensated for by the use of density gradient solutions (see below). The volumetric capacity of a single ultracentrifuge rotor can be limiting. Thus, large-scale preparative approaches require several ultracentrifuges and rotors with high volume capacities. The use of preparative ultracentrifugation may be limited by the costs of ultracentrifuges and their rotors.

Differential centrifugation, i.e., pelleting, separates sample components based on their sedimentation rates and depends on the size, shape, density, the applied centrifugal force (speed), and the used medium (Fig. 5(a)). In general, large particles tend to pellet before small ones, dense particles sediment before light ones, and symmetrical particles (e.g., globular) sediment before the asymmetrical ones (e.g., particles with long extensions). Optimal conditions for pelleting always need to be experimentally verified. The sedimentation rate of a particle is its speed of travel in solution when subjected to a high gravitational force (g-force) upon centrifugation. The Svedberg unit (S) is a measure of this sedimentation rate. Typical S-values of prokaryotic viruses are in the range of 100 to 1000 S (Fig. 6), the larger the virus, the larger the S-value. During differential centrifugation, the sample components move towards the bottom of the tube as long as the force is applied. The separation power is rather limited due to the even dispersal of the sample components in the sample buffer before the run (Fig. 5(a)). Consequently, small sample components locating close to the bottom of the tube can pellet alongside large components. Thus, the formed pellets always represent an enrichment of different sizes of samples components – larger ones forming the majority. Sometimes step-wise removal of large sample components followed by repeated pelleting of the resulting supernatant containing the majority of small sample components with higher speed may result in a specimen with adequate purity. However, due to the limited separation power, differential centrifugation is often used as a final step to concentrate the viruses and transfer them to an appropriate buffer. Recovery yields are usually good, if the virus particles tolerate high shear forces. However, viruses can aggregate in high concentration solutions and resuspension of viruses from pellets can sometimes be difficult. Use of cushions (small volumes of high-density solution) might help to keep the particles in solution upon differential centrifugation and improve separation.

Density ultracentrifugation can be divided into rate zonal and equilibration centrifugation. Rate zonal and equilibrium centrifugations are among the few methods that enable the separation of particles that have the same size but different mass e.g., separation of the procapsids lacking the genome from the mature virions. They are performed by using continuous or step-wise gradients made of viscous solutions of various concentrations with the densities increasing towards the bottom of the tube. The gradients can be self-forming during the applied force (e.g., CsCl) or preformed (e.g., sucrose) before the ultracentrifugation. There are commercial gradient formers for making continuous gradients, but also continuous gradients can be formed by allowing a step-wise gradient to diffuse to linearity with time (e.g., overnight in a cold room). Selection of the gradient material depends on the density and other properties of the virus studied (Fig. 3). Importantly, the gradient must not affect the biological activity or stability of the virus. Whether ultracentrifugation results in good yields depends on the sensitivity of the viruses to the viscous and hyperosmotic nature of the gradient materials used, to high shear forces, and to the osmotic pressure present during centrifugation. Gradient materials that enable low viscosity gradient runs are also available. Sucrose is usually well-tolerated with viruses, but it enables experiments to be performed at densities up to 1.3 g/mL only, whereas dense salts such as CsCl can provide densities up to 1.9 g/mL. Other widely used gradient materials for viruses are glycerol and iodixanol. Some gradient materials such as CsCl may be toxic to cells, and this may hinder their application in virus purification. When working with CsCl, it is essential to ensure that the

applied forces are low enough so that CsCl is not precipitated at the bottom of the gradient during the run. The density of crystallized CsCl (4 g/mL) produces stress far in excess of the limits of the rotors causing serious safety risk. Manufacturers provide CsCl safety curves for all rotor types.

Upon density ultracentrifugation, virus particles are subjected to buoyant and sedimentation forces that are affected by the size, shape, mass, and density of the virus particles and the viscosity of the gradient material. They are separated and they form zones that are visible to the eye due to light-scattering under certain virus particle concentration thresholds (Fig. 5). In rate zonal ultracentrifugation, the density gradient is selected so that its highest density is lower than the density of the separated viruses allowing the separation of different components to occur according to their sedimentation coefficients. The viscosity of the gradient solution improves particle resolution as it prevents the broadening of the particle sedimentation zone by acting as a counterforce against the relative centrifugal force that is higher towards the bottom of the tube. Using an optimized combination of time and speed a powerful separation can be achieved.

In equilibrium ultracentrifugation, the density of the gradient must cover a wide range of densities allowing separation, and the maximum density must exceed that of the densest particle. Virus particles migrate until they reach the position in the gradient that is equivalent to their buoyant density and remain there forming a narrow concentrated zone. Thus, the applied equilibrium centrifugation time needs to be long enough to establish the equilibrium. The sample can be loaded either on the top of the gradient or on the bottom from where the sample components can reach their equilibrium positions by floating upwards (flotation centrifugation). When using self-forming gradient material (CsCl), the sample can be dissolved in CsCl and equilibrium positions will be achieved during the run simultaneously with the gradient formation. However, this approach is time consuming. Typically, the density gradient is selected allowing viruses to equilibrate at a position approximately two-thirds from the top of the gradient. Viruses are complexes of nucleic acid (DNA $\rho = 1.7$ g/mL; RNA $\rho = 1.8$ g/mL in CsCl) and protein ($\rho = 1.3$ g/mL in CsCl), and in some cases also lipids ($\rho < 1.0$ g/mL), and their buoyant densities usually range between 1.3–1.5 g/mL (in CsCl), the membrane-containing viruses typically having lighter buoyant densities.

Chromatographic Methods

Successful chromatography enables efficient but reversible binding of virus particles to a chromatography matrix and subsequent elution, in a controlled manner, results in a pure virus preparation with high biological activity separated from other sample components. In optimal conditions, chromatography offers high selectivity leading to good resolution, purity, and yield, accompanied by concentration of the sample.

Traditional chromatography columns are packed with porous particles and most of the adsorptive surface area resides inside the pores (usually ~50–100 nm in diameter). Binding of sample components to porous particles occurs mainly via diffusion. Although traditional matrixes can work for small-sized viruses, many viruses are too large to enter into the pores. Limited binding results in low dynamic binding capacity and reduced yields. Furthermore, due to the small diffusion coefficients of virus particles, low flow rates must be used to enable slow penetration and elution from the porous particles that can broaden peaks and lengthen the processing time. Virus particles are also subjected to shear due to eddy dispersion. However, traditional chromatography matrixes can sometimes be useful to separate virus-sized components in the void volume, while the smaller sample components are retained in the column.

There are non-conventional resins and matrixes available that enable purification of virus-sized complexes. Monoliths, for example, are made of a continuous single-piece stationary phase that has wide 1–5 μm interconnected channel networks distributed throughout the column. Sample components are transported in the column via convective mass transport in a laminar flow, and no diffusion is involved. Shear is minimal. All sample components have equal velocities irrespective of their size. Consequently, resolution and binding are independent of the flow rate, and immediate responses to changes in the mobile phase composition are obtained that result in narrow and sharp peaks. This enables the use of high flow rates promoting high processing speed. The channels have high surface area that is directly accessible for sample components. Consequently, efficient binding and high yields can be obtained and smaller columns and less buffer can be used for purification. Several bacteriophages with various morphologies have been successfully purified with monoliths (Fig. 7). Even though input samples have usually undergone some pre-purification steps, virus particles have also been successfully purified with monoliths directly from cell lysates.

Stacked membranes resemble monoliths in many aspects. They have large accessible channels and sample components are transported via convective mass transport that enables high dynamic binding and capacity. Thus, stacked membranes are applicable to purification of viruses, but as the bed height is usually lower than in monoliths, less sample can be purified.

Chromatographic methods where the separation is based on the charge, affinity, hydrophobic interaction, or size are applicable for viruses (Fig. 3). The surface properties of virions, the number and distribution of charge, the presence of hydrophobic spots, and their accessibility determine the type(s) of chromatography that can be applied. Importantly, binding of virus to the matrix can occur via multiple different contacts on the virion surface and can thus involve different chemistries depending on the pH and ionic strength. For instance, enveloped viruses have a negatively charged lipid bilayer that makes them potentially amenable for purification using anion exchange chromatography. In contrast, many halophilic viruses require high ionic strength for stability and their binding to ion exchange chromatography columns at low ionic strength is not possible. Virions that have exposed histidine, tryptophan, or cysteine residues, may be amenable to metal affinity chromatography that is based on their specific interaction with nickel, cobalt, or copper. Availability of specific antibodies enables affinity chromatography purification that is

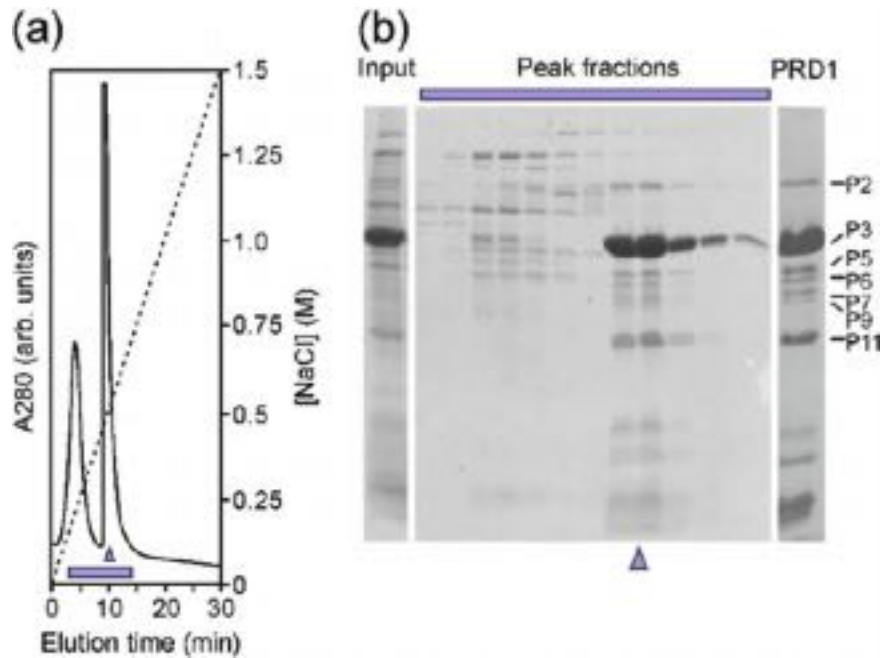


Fig. 7 Virus purification by monolithic anion exchange chromatography with quaternary amine matrix. (a) Icosahedral internal membrane-containing PRD1 bacteriophages were first precipitated by PEG-NaCl from the lysate supernatant (input in b) and bound to monolithic column and eluted with a linear 0–1.5 M NaCl gradient (dashed line). Optical density (absorbance) is indicated by a solid line. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis (Coomassie blue staining) of the peak fractions (marked by purple bars in a and b) collected during elution. The virus peak is indicated by an arrow head in a and b. Positions of the major PRD1 structural proteins are marked on the right. The figure is modified and reprinted from Oksanen, H.M., Domanska, A., Bamford, D.H., 2012. Monolithic ion exchange chromatographic methods for virus purification. *Virology* 434, 271–277. Copyright (2012), with permission from Elsevier.

based on the specific interaction between the antigen present on the virion surface and the antibody bound to the stationary phase. Size exclusion chromatography can work for small viruses. The optimal separation power needs to be established experimentally in conditions supporting the virus integrity. For instance, elution from ion exchange columns by increased ionic strength can be detrimental for some sensitive viruses, whereas elution due to change in pH can be damaging to others. Membrane-containing viruses are typically sensitive to buffers containing detergents. The composition of the elution buffer may also bring needs for additional purifications. Obtaining good separation requires optimization of the loading amount, eluent composition, and elution conditions (e.g., linear or step-wise gradients), as well as column chemistry. However, a single chromatography purification is usually fast to perform and can be easily scaled up by increasing the matrix amount and column size. Obviously, various chromatography methods can be tandemly applied to obtain better purity. Chromatography equipment, columns, as well as matrixes are relatively expensive, although columns and matrixes can be washed and reused.

Flow Field-Flow Fractionation Methods

Asymmetrical flow field-flow fractionation (AF4) is a method where sample components are separated based on their hydrodynamic sizes in the absence of a solid stationary phase (Fig. 8). Separation takes place in a thin channel of ~190–500 µm. The bottom of the channel (i.e., accumulation wall) is permeable and covered with an ultrafiltration membrane of selected material with defined molecular weight cut-off determining the size range of the sample components retained in the channel for separation. Separation is achieved by applying cross flow but is also influenced by the velocity of the channel flow that goes through the channel from the inlet to the outlet. Cross flow is perpendicular to the channel flow, and it pushes the sample components towards the accumulation wall. However, the sample components diffuse against the cross flow based on their diffusion coefficients. After equilibrium between the cross-flow force and diffusion is reached, the elution order of the sample components is dictated by their position in the channel and thus their access to different channel flow velocities that are the highest in the center of the channel and slowest at the walls (Fig. 8). Components having a diameter smaller than 0.5 µm are eluted in order with the small sample components eluting before the large ones.

AF4 is a gentle separation method since there is no stationary phase. As separation is based on hydrodynamic size, the same methods and running parameters can be applied to various viruses of similar sizes. However, optimization of the flow rates, gradients, eluent composition, and loading amounts is required for the best separation. AF4 was originally designed for analytics rather than preparative applications, and thus the loading amounts that still lead to good separation are relatively low. Despite this limitation, several viruses have been successfully purified by AF4, establishing it as one of the potential virus purification methods.

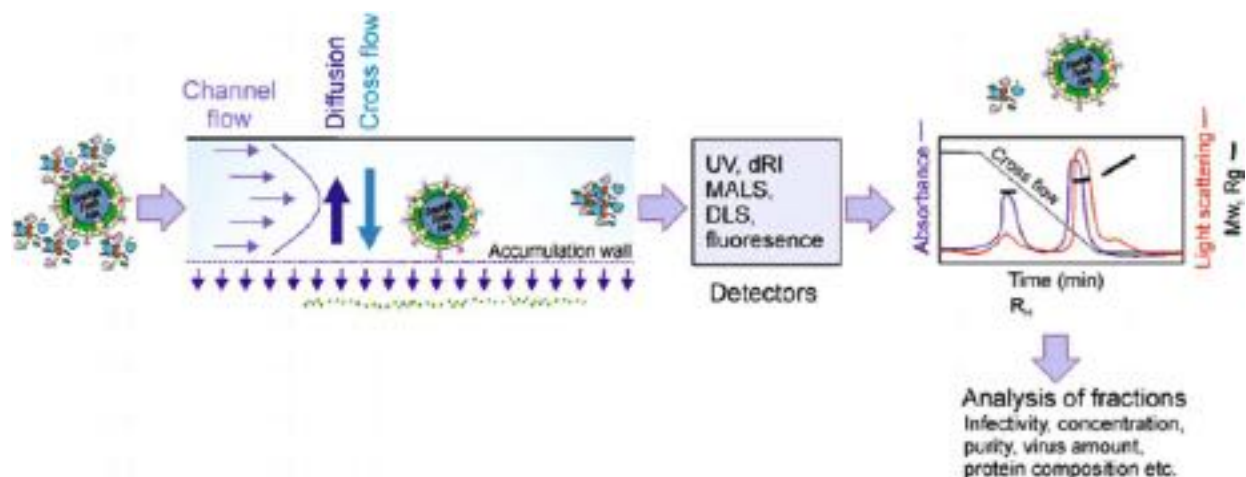


Fig. 8 Principles of AF4 fractionation. Applied cross flow pushes the sample components towards the accumulation wall that consists of an ultrafiltration membrane and a permeable ceramic frit through which the cross flow is pumped out. Sample components that are smaller than the pores of the ultrafiltration membrane pass through (green spots). Remaining sample components diffuse against the cross-flow force depending on their diffusion coefficients. Channel-flow rate is slowed down at the channel walls due to friction and the eluent moves with the highest pace in the middle of the channel. Consequently, smaller sample components with higher diffusivity reach higher flow rates and elute first. Viruses, proteins, and channel height are not to scale. In practice, the sample components occupy only $\sim 1\%$ of the channel thickness. Separation can be coupled to on-line detectors such as UV, multi angle light scattering (MALS), dynamic light scattering (DLS), fluorescence etc. to obtain information for instance on particle concentration, size and size distribution. Fractions can also be collected for off-line analysis.

The running time for a single AF4 experiment is typically short, less than one hour. Consequently, several experiments can be performed in a row to improve the total yield. Recovery yields are generally good, and the integrity and infectivity of viruses is well maintained. AF4 can be coupled to multiple detectors (Fig. 8) that enable simultaneous collection of data on for instance the size and size distribution of the viruses. AF4 is broadly compatible with a variety of buffers. It has been successfully utilized for purification of archaeal viruses in buffers containing moderate (0.6 M) and high (1.6 M) total ionic strengths. Initial investment costs for AF4 are high and the theory is rather complicated.

Assessing Purification Performance

Plaque assay is an extremely sensitive and easy method to determine the number of infectious viruses (plaque forming units, pfu) and hence the success of the purification of infectious viruses. However, not all virus systems produce plaques and the efficiency of the plating (EOP) can be very low. Protein concentration can be estimated by multiple methods (for example the Lawry method or Bradford assay) that depend on the amino acid composition of the proteins and vary in their sensitivity as well as compatibility to various chemicals. The easiest and most sensitive method to estimate the amount of protein is based on absorbance at 280 nm (A_{280}). Additional absorbance measurement at other wavelengths such as 260 nm (A_{260}) provides further information about the nucleic acid content. Depending on the ratio of protein and nucleic acid as well as the amino acid and nucleotide composition, each virus has a typical A_{260} to A_{280} ratio. Based on the infectivity and protein concentration or absorbance values, specific infectivity of the virus preparation (pfu/mg of protein; pfu/one A_{280} unit; pfu/one A_{260} unit) throughout the purification can be calculated and compared. Each virus has its own optimal specific infectivity value which is dependent on the EOP for the host used, and on the size and composition of the virus. In one virus system or between similar virus systems specific infectivity is a good indicator of the quality of the virus preparation. A high ratio between the number of infectious viruses and quantity of protein indicates a high-quality specimen. As an example, purified bacteriophage PRD1 has specific infectivity values over 10^{13} pfu/mg of protein (Fig. 9(b)).

Denaturing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is commonly used to analyze the quality and composition of the purified virus sample. The protein composition can be visualized by staining the gels with protein-specific stains such as Coomassie blue (Fig. 9(a)). The presence of lipids can be detected from the SDS-PAGE gels by lipid-specific stains, but more detailed information on the viral lipid composition can be obtained by lipid extraction followed by thin layer chromatography or lipidomics. The presence of nucleic acids can be visualized in the stacking gels of SDS-PAGE gels, but agarose gels coupled with nucleic acid specific stains and nucleic acid specific techniques and sequencing are required for obtaining information on the viral genome. Transmission electron microscopy (TEM) and negative staining is one way to analyze the purity of the samples, but it also gives information on virus morphology and can be used to enumerate viruses especially when a plaque assay cannot be applied. The specimen preparation for cryo-EM results in the preservation of biological specimens at near native conditions within a thin amorphous ice film allowing more detailed analysis of the virus morphology and sample purity, and potentially an atomic resolution structure of the virus particles (Fig. 9(c)).

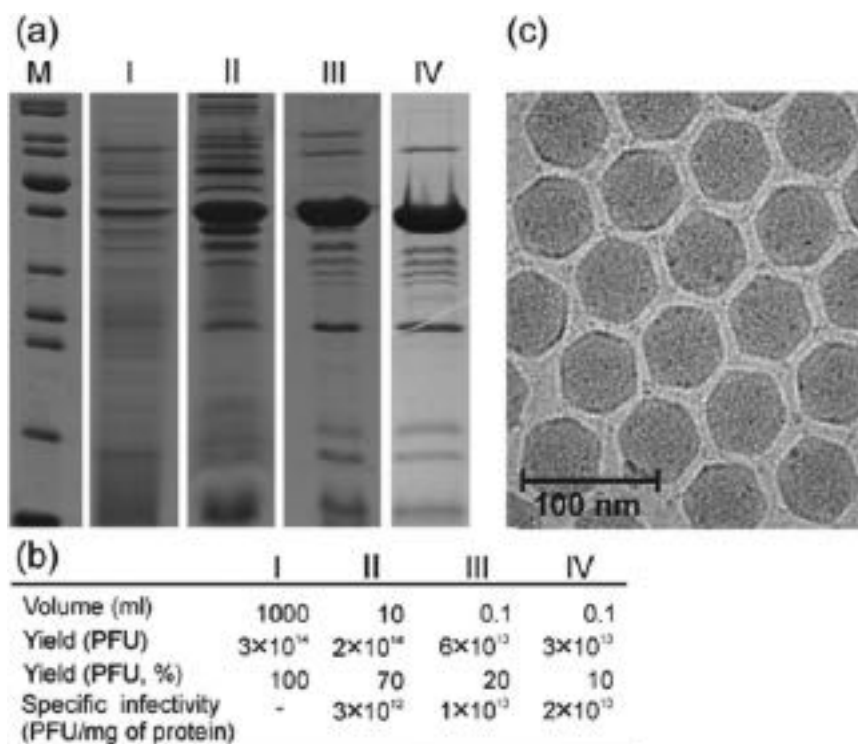


Fig. 9 PRD1 virus purification by precipitation, density gradient ultracentrifugation in sucrose and concentration by differential centrifugation. (a) SDS-PAGE analysis followed by Coomassie blue staining of different purification stages of bacteriophage PRD1: (I) production as a lysate; (II) PEG-NaCl precipitate of the virus; (III) purified virus after 5%–20% rate zonal sucrose gradient and differential ultracentrifugation; (IV) purified virus after 5%–20% rate zonal and 20%–70% equilibrium ultracentrifugations followed by differential ultracentrifugation. Molecular mass marker is from 10 kDa to 100 kDa. The PRD1 major capsid protein P3 with the highest copy number in the virion is 43 kDa. (b) Purification statistics derived from purification of 1 litre of lysate. See the explanations for I–IV. (c) Cryo-electron microscopy micrograph of bacteriophage PRD1 purified by PEG-NaCl precipitation, rate zonal and equilibrium ultracentrifugation in sucrose, and concentrated by differential centrifugation (see IV in a and b). Panel c: Courtesy of Ane Martinez-Castillo and Nicola G.A. Abrescia, CIC bioGUNE, Spain.

Conclusions

More than 100 years have passed since the isolation of the first bacteriophage. Archaeal viruses were recognized over 60 years later at the time when archaea were discovered and acknowledged as the third domain of life. Bacteriophages and archaeal viruses are the most abundant group of biological entities on Earth, outnumbering their host cells by at least ten-fold thus influencing global biogeochemical cycles and driving microbial evolution. These parasites thrive in various ecological niches all over the world. Since their discovery, bacterial and archaeal viruses have served as model systems to unravel the fundamental aspects of virology, microbiology, structural biology, and ecology. Methods and technologies that provide a high number of viruses and virus-like particles purified to homogeneity are required not only for basic research but also applications of viruses in biotechnology, medicine and industry. Success in obtaining high quality samples in quantities relies on the ability to produce high titer virus stocks and is largely dependent on the particular virus and its host cell. There are certain principles and methods that can be broadly applied, but experimentation of the optimal storage, production, and buffer conditions for each virus of interest is typically needed to maximize the yield. Virus purification techniques involve filtration, dialysis, precipitation, ultracentrifugation, chromatographic and flow field-flow fractionation methods that are optimized and used together and sequentially in different combinations. There is a growing need for new virus purification technologies to meet the purity criteria for phage and virus-based gene therapy, vaccine developments and nanotechnology, paving the way for new innovations and applications.

Further Reading

- Adams, M., 1959. Bacteriophages. New York: Interscience Publishers, pp. 1–592.
- Atanasova, N.S., Senčilo, A., Pietilä, M.K., *et al.*, 2015. Comparison of lipid-containing bacterial and archaeal viruses. *Advances in Virus Research* 92, 1–61.
- Eskelin, K., Lampi, M., Meier, F., *et al.*, 2016. Asymmetric flow field flow fractionation methods for virus purification. *Journal of Chromatography A* 1469, 108–119.
- Eskelin, K., Poranen, M.M., Oksanen, H.M., 2019. Asymmetrical flow field-flow fractionation on virus and virus-like particle applications. *Microorganisms* 7, 555.

- Lewis, A.C., 2015. Chromatographic techniques. eLS. Chichester: John Wiley & Sons Ltd. Available online at: <https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0002705.pub2> (accessed 21.05.19).
- Mohr, H., Völkl, A., 2017. Ultracentrifugation. eLS. Chichester: John Wiley & Sons Ltd. Available online at: <https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0002969.pub3> (accessed 02.05.19).
- Oksanen, H.M., Domanska, A., Bamford, D.H., 2012. Monolithic ion exchange chromatographic methods for virus purification. *Virology* 434, 271–277.
- Pietilä, M.K., Demina, T.A., Atanasova, N.S., Oksanen, H.M., Bamford, D.H., 2014. Archaeal viruses and bacteriophages: Comparisons and contrasts. *Trends in Microbiology* 22 (6), 334–344.
- Podgornik, A., Yamamoto, S., Peterka, M., Krajnc, N.L., 2013. Fast separation of large biomolecules using short monolithic columns. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences* 927, 80–89.
- Williams, S.K.R., Caldwell, K.D., 2012. Field-Flow Fractionation in Biopolymer Analysis. Wien: Springer-Verlag Wien, pp. 1–306.
- Wolf, M.W., Reichl, U., 2011. Downstream processing of cell culture-derived virus particles. *Expert Review of Vaccines* 10, 1451–1475.
- Yamamoto, K.R., Alberts, B.M., Benzinger, R., Lawhorne, L., Treiber, G., 1970. Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology* 40, 734–744.

Relevant Websites

<https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses (ICTV).

High Throughput Sequencing and Virology

Graham L Freimanis and Nick J Knowles, The Pirbright Institute, Pirbright, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Nanopore sequencing A technology of single molecule sequencing enabling direct and real-time sequencing of DNA and RNA strands using nanopores (either biological or solid state).

Next generation sequencing/High throughput sequencing New methods for high speed, low cost sequencing of genomes, inclusive of both second and third generation sequencing technologies.

Second generation sequencing Second generation methods are all characterized by the requirement to prepare amplified sequencing libraries before the process of sequencing of amplified DNA clones.

Sequencing library Any piece of nucleic acid that is attached to adapter barcodes for sequencing. Libraries may

also contain indices enabling the multiplexing of libraries from different samples on the same sequencing run.

Shotgun sequencing A technique for determining the sequence of entire genomes based on sequencing of random fragments of DNA that are then assembled by *de novo* assembly into fragments.

Single cell sequencing The sequencing of information e.g., transcriptome, at the level of individual cells.

Single molecule sequencing Techniques that can read the base sequence directly from individual strands of DNA or RNA present in the sample of interest, without the requirement for amplification.

Introduction

Over the last two decades, significant progress has been made in the field of DNA sequencing. Sanger (dideoxy) sequencing, or what may be thought of as traditional sequencing, has been around for more than 40 years. This platform produces well-trusted, up to 1-kbp long reads which are still considered the gold standard for sequencing and are routinely used for clinical gene tests or validations, although their low throughput limits them to single or small batches of targets. Next generation sequencing technologies however offer a significant step forward in our capability to dissect virus-host interactions and explore the viral evolutionary mechanisms that underpin transmission and pathogenicity. The diversification of these new technologies have enabled the diagnostic evaluation of gene panels, whole genome sequences and viral transcriptomics analysis, which have become commonplace within the scientific literature (Fig. 1). Importantly, such new chemistries have enabled a dramatic reduction in cost per gigabase, thereby increasing the potential for sample throughput. All of this has had a significant impact upon the field of virology.

First Sequencing Methods

The first short DNA sequences were obtained in the early 1970s using a variety of different methods. However, in 1977, three new sequencing methods were published; (1) Allan Maxam and Walter Gilbert, working at Harvard University, published a DNA sequencing method based on chemical modification of DNA and subsequent cleavage at specific bases. Also known as chemical sequencing, this method allowed purified samples of double-stranded DNA to be directly sequenced. Maxam-Gilbert sequencing requires radioactive labeling at one 5' end of the DNA and purification of the DNA fragment to be sequenced. Chemical treatment then generates breaks at a small proportion of one or two of the four nucleotide bases in each of four reactions (G, A + G, C, C + T) generating a series of fragments, from the radiolabelled end to the first "cut" site in each molecule. The fragments in the four reactions are electrophoresed side by side in denaturing acrylamide gels for size separation and thus determination of the base sequence; (2) a method for the direct determination of RNA sequences, based on digestion of radiolabelled RNA with different ribonucleases, was developed by George Brownlee and co-workers at Cambridge University, UK; (3) Fred Sanger and co-workers (Cambridge, UK) published the chain-termination method which used fewer toxic chemicals and lower amounts of radioactivity than the Maxam-Gilbert method. Separate reactions are set up, each with a small amount of one of the four (A,T,G,C) dideoxynucleotides and larger amounts of deoxynucleotides. The reactions produce chains complementary to the DNA to be sequenced, until extension is terminated by incorporation of a dideoxynucleotide, producing fragments which are analyzed by electrophoresis. Sanger sequencing has been applied to RNA using a reverse transcriptase to generate cDNA fragments, or DNA using a DNA polymerase and was flexible in that it could incorporate either a radiolabelled oligonucleotide primer or a radiolabelled deoxynucleotide. Due of its comparative ease, the Sanger method was soon automated and was the method used in the first generation of DNA sequencers. Sanger sequencing is the method which prevailed from the 1980s until the mid-2000s. Over that period, great advances were made in the technique, such as fluorescent labeling, capillary electrophoresis, and general automation. These developments allowed much more efficient sequencing, leading to lower costs. The Sanger method is the technology which produced the first human genome in 2001, ushering in the age of genomics. The first two methods have been used to sequence the

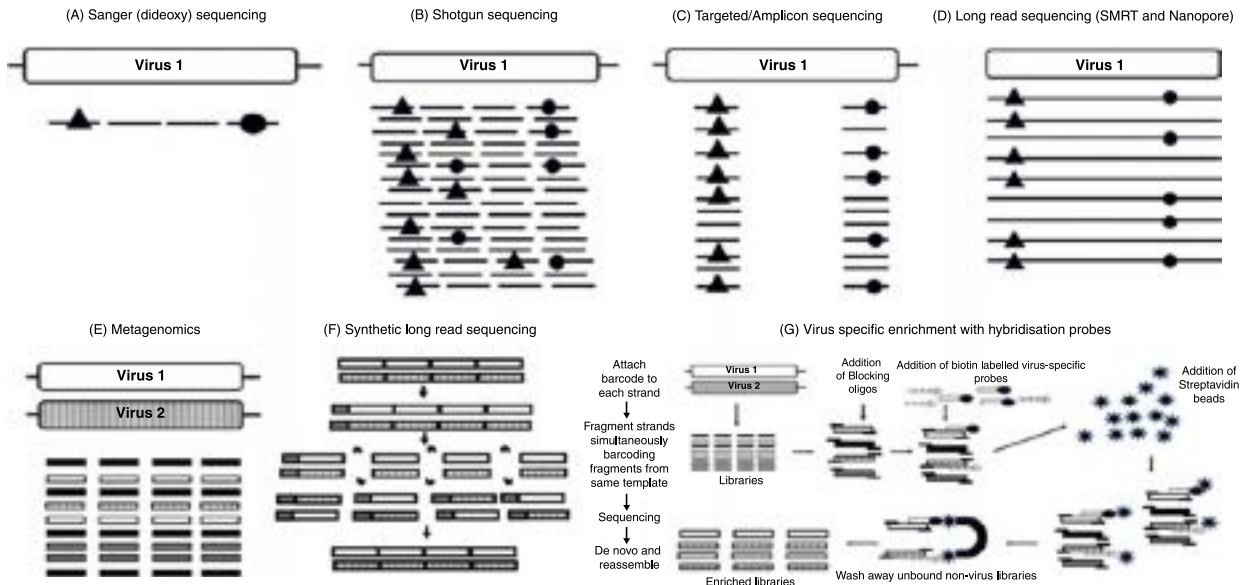


Fig. 1 Comparison of advantages of different types of enrichment and high throughput sequencing library preparation methodologies. (A) *Sanger sequencing*: A low throughput methodology where reads are sequenced that match the consensus level sequence of populations within the sample. (B) *Shotgun sequencing*: A protocol, with minimal intervention and biases, that sequences all dsDNA templates present within a sample (either native or converted from RNA using reverse transcription). (C) *Amplicon sequencing*: A targeting and enrichment protocol that directly targets specific regions with gene specific primers and enriches via amplification. Incurs biases associated with amplification, but generally produces deep coverage across targeted regions. (D) *Third generation (long-read) sequencing*: Produced using single molecule sequencing by either Single Molecule Real-Time (SMRT) based platforms (including Pacific Biosystems) or by use of synthetic nanopores (presently available platforms include Oxford Nanopore technologies). Results in single long reads representative of the original template. (E) *Metagenomics*: Employs shotgun sequencing or RNA-Seq protocols to characterise all species of pathogen present within populations. (F) *Synthetic long/linked reads*: A methodology able to tag reads originating from the same template with specific unique molecular identifiers denoting the original template from which they originated. After sequencing, these indexed short reads sharing unique indices can be combined to produce long reads. (G) *Probe hybridization enrichment*: Probes can be used to enrich for nucleic acids derived from viral material either directly after nucleic acid extraction (prior to library preparation) or after the amplification stage of library preparation. The homology of the probes to the pathogen is key to the success of the method.

unusually long (~80–250 residues) poly(C) tracts of foot-and-mouth disease virus (family *Picornaviridae*, genus *Aphthovirus*) and encephalomyocarditis virus (family *Picornaviridae*, genus *Cardiovirus*), a feat which Sanger sequencing is unable to achieve.

Next Generation Sequencing and Applications in Virology

All second-generation sequencing technologies utilize chemistries incorporating an amplification step prior to the actual process of sequencing, resulting in millions of short clonal copies of each sequencing library being produced, which are then sequenced in massive parallel sequencing reactions. Although there are distinct differences in the way second generation sequencing libraries are produced between platform chemistries, essentially a sequencing library is a short piece of DNA (typically between 50 and 600 bp in length) generated from a sample of interest. This is bracketed by platform-specific barcodes or adapter sequences, and unique barcodes enabling multiple samples to be combined (multiplexed) on the same sequencing run. Any template lacking appropriate adapter sequences would not be recognized as a library. The four main second generation sequencing platforms i.e., 454, Illumina, Ion torrent and SOLiD, all employ short reads (< 600 bp) and share an amplification stage prior to sequencing to drive their chemistries.

454 (Pyrosequencing)

Originally marketed by 454 Life Sciences (now owned by Roche), 454 was the first commercial HTS platform released in 2005 and acted as one of the foundational platforms of the technology, however, the chemistry was discontinued in 2013. The 454 technology uses the sequencing by synthesis chemistry in which DNA molecules are sequenced upon the addition of individual nucleotides. Initially, DNA molecules are attached to nanobeads in an emulsion of oil and PCR reagents. An excess of beads favors the statistical probability that there should be at least a single bead for each nucleic acid template and the emulsion leads to a formation of droplets, each containing a bead, ssDNA template and PCR reagents. Each reaction generates multiple ssDNA clones of an original template DNA present within the droplet enriching clonal copies on the surface of the nanobeads. Once the droplets are broken the ssDNA-coated beads within the droplet are released and dispersed into nanowells across the surface of a microtiter plate. Each well contains a single bead, in addition to luciferase and sulfurylase molecules. As nucleotides are washed across the plate a single unlabeled nucleotide at a time. Each time a nucleotide matches the dNTP in the cloned template, the base is added to the DNA

chain, releasing a pyrophosphate. The sulfurylase processes the pyrophosphate, creating available ATP, which is processed by luciferase to generate a pulse of light. Higher error rates for such datasets arise from homopolymeric regions and artificial amplification artefacts. 454 was discontinued by Roche in 2013. Ion-torrent (marketed by Thermo/Life Technologies) utilizes a similar chemistry to the 454 system with emulsion PCR as a central part of the library preparation, however, it has important differences in its sequencing reaction, which is detected by a pH semiconductor rather than fluorescence, with the release of protons during nucleotide incorporation being detected using an ion sensor to measure shifts in pH through changes in conductance. Ion torrent can use different sized chips for sequencing, which adds flexibility to the maximum read numbers produced during each sequencing run, e.g., a run requiring a larger number of reads or an increased number of samples would require a higher specification chip.

SOLiD sequencing developed by Applied Biosystems shares similar elements of the 454 library preparation in which DNA is fragmented onto beads in an emulsion PCR, however the sequencing centers around the principle of ligating of di-basic probes to a complementary template resulting in fluorescence.

Illumina Sequencing

Illumina sequencing utilizes an alternative method of on-board amplification called “bridge amplification”. As with other short read chemistries, libraries are composed of short fragments of the dsDNA of interest, with adapters and indices attached to either end. Libraries are dispersed across the flow cell surface and allowed to bind to a lawn of primers. Once bound to the flow cell, libraries undergo “bridge amplification”, whereby clusters of clonal copies of each original library are produced. These clusters of libraries are then sequenced using a process whereby reversible terminators incorporated into fluorescently labeled nucleotides added at each step block elongation after incorporation of a single nucleotide, in a process analogous to Sanger sequencing. Any incorporated nucleotides are then detected by a colored light emission, before denaturation solutions remove the reversible terminator, enabling elongation and attachment of the next fluorescently labeled/blocked nucleotide. Illumina have a number of different models of sequencer available from small benchtop machines including the iSeq-100, MiniSeq and MiSeq, to the larger scale machines (HiSeq 4000, HiSeq X and NovaSeq). Newer versions of the technology including the iSeq, MiniSeq NextSeq use reduced numbers of colors to streamline chemistries and reagents, combined with patterned flow cells to increase sequencing accuracy.

It has been the introduction of the benchtop level machines to the market that has driven the genomics revolution, enabling access to the platforms which had previously not been accessible due to the high upfront costs for such technologies.

Applications of High-Throughput Sequencing in Virology

The high throughput sequencing of viruses produces a snapshot of the total diversity within a virus population, below the level of consensus that is possible by Sanger sequencing. This is particularly useful when investigating RNA viruses, which due to an error prone RNA-dependent RNA polymerase produce large populations of progeny virions which differ from the parental virus. Such diversity can be characterized by the assembling of sequencing reads and characterization of polymorphisms in the assembled contigs (constructs of overlapping reads). During the analysis, the greater the number of reads mapped to each position (known as coverage depth), the higher the accuracy or level of confidence in the conclusions that can be drawn from data. This will proportion confidence to predictions made from the data regarding minority variants that are truly present within the population, as opposed to artefacts of library preparation or analysis. Genome sequencing can also be taken as consensus (present in more than 50%) of reads at a given position. These principles underlie how workflows such as metagenomics, PCR amplicon sequencing and enrichment strategies can be successfully tailored to specific applications.

Approaches to high throughput sequencing (HTS) in virology, generally take one of three main approaches. The term “metagenomics” describes all processes by which sequences can be identified without any prior knowledge. This is most useful in studies where virus discovery, or the characterization of different types of virus e.g., characterising the virome within a sample, are the main objectives. DNA sequencing or amplicon sequencing describes the enrichment of a sample using amplification (mainly PCR), providing ultra-deep coverage datasets on targeted parts of a viral genome in order to identify minority variants involved in drug resistance, transmission or virulence studies. Finally, the targeted enrichment of libraries during library preparation describes the specific enrichment of viral nucleic acid in order to increase the number of virally derived reads after sequencing thereby reducing host background and increasing overall specificity. Such protocols are particularly relevant in clinical samples where the amount of virus may be limited, but the identity of the virus known. Although enrichment for viral reads will reduce background read counts and increase sensitivity for viral reads in hard to sequence samples, this comes at an increased cost in terms of reagents and labor.

Metagenomic Approaches

The term metagenomics can be defined as the characterization of genetic information directly from the clinical or environmental sample in the absence of culture. Metagenomic approaches have been used extensively for pathogen discovery and the characterization of microbial diversity within samples and thus constitute some of the most widely used applications of high throughput sequencing in virology. Whilst the workflow presents considerable flexibility in tailoring to the objectives of any screening study, it is particularly effective in characterising unknown pathogens without a priori knowledge of genomes, or when identifying multiple virus types within samples, e.g., virome analysis.

In such studies, total DNA/RNA is extracted from samples, libraries prepared and sequenced by either shotgun sequencing or RNA sequencing protocols. Metagenomic workflows have the additional benefit of being effective when nothing is known about the pathogens present in the sample, thereby circumventing the need for expensive primers or probes which makes it particularly useful when dealing with outbreak situations, as demonstrated recently during the Zika virus (genus *Flavivirus*, family *Flaviviridae*) outbreak, where the causative virus was previously unsuspected. Whilst the use of metagenomics to produce viral whole genome sequences from outbreak samples can provide an enhanced level of resolution, the overall sensitivity of such protocols is significantly less than more targeted methods. However, metagenomic approaches have been used effectively in previous studies of outbreaks of norovirus (genus *Norovirus*, family *Caliciviridae*) and parainfluenza virus (genus *Respirovirus*, family *Paramyxoviridae*), enabling tracing of transmission in real-time. Such information can be invaluable in front line healthcare settings, particularly when identifying transmission clusters between patients and clinical staff. Metagenomic protocols have the added advantage of being able to generate both host and viral derived data, thus potentially informing on virus associated cancer formation and diversity of integrated virus genomes, which can provide an important additional level of detail to tailored treatment programs in patients.

Whilst shotgun metagenomic protocols provide a representation of nucleic acid present within a sample, its sensitivity is limited. It can be further enriched by incorporating random amplification of nucleic acid whether using Sequence Independent Single Primer Amplification (SISPA), Single Primer Isothermal Amplification (SPIA) or using RNA-Seq protocols to probe samples for viral transcripts, although these introduce their own biases. The application of enrichment protocols to these approaches, although useful can result in costs becoming prohibitively high. Random priming and metagenomics workflows have previously produced genome sequences for a novel mink astrovirus (genus *Mamastrovirus*, family *Astroviridae*), paramyxoviruses (family *Paramyxoviridae*), influenza viruses (family *Orthomyxoviridae*), dengue virus (genus *Flavivirus*, family *Flaviviridae*) in infected mosquitos, viruses in stool and avian viruses. SISPA is one of the more commonly employed protocols for metagenomics investigations, particularly virus discovery and involves the DNase digesting a sample to reduce host nucleic acid, in order to specifically enrich for viral RNA before an amplification step with tagged primers. SISPA was used to identify the Schmallerberg virus where a small number of reads led to the identification of a novel virus (Schmallerberg virus; genus *Orthobunyavirus*, family *Peribunyaviridae*) infecting dairy cattle. Although RNA-Seq based protocols are effective in identifying the presence of RNA viruses in samples, it should be underlined that the technical and analytical requirements for transcriptomics studies are very different to those requirements in metagenomics studies. The downside of such protocols is the large percentages of reads deriving from either host or other organisms present in or contaminating the sample. As a result, the read depth is often variable, with the reagent cost high. Other applications include RNA-Seq-based approaches in which data can be used to identify extraneous viral reads in those unmapped reads after host depletion. Examples of previous application of such datasets include the detection of adventitious agents within vaccine seed stocks or the detection of viral contaminants in bioreactor cultures. Virome and microbiome diversity are also areas where metagenomics has provided significant contributions to our understanding of viruses within our environment with investigations of the RNA virosphere within samples taken from several locations including the Mediterranean and underground lakes on the Western Antarctic peninsula.

RNA Sequencing

More traditional applications for RNA sequencing in virology (i.e., outside of metagenomics) include small whole genome sequencing and transcriptomic profiling, from differential gene expression (poly-A enriched) to mapping sites of transcript initiation (Cap analysis gene expression sequencing - CAGE-Seq). Many viral infections suppress host cell expression, resulting in the sequencing of viral mRNAs which are instead produced by the hijacked cellular machinery. The inclusion of enrichment protocols such as ribosomal depletion (for total RNA-Seq), enable the sequencing of non-polyA species of RNAs such as non-coding regions, whereas enrichment of poly-A for differential gene expression studies would reveal the impact of viral infection upon cellular expression. RNA-Seq is particularly powerful for investigating host-pathogen interactions and has been used to characterise host transcriptional responses to Ebola virus, identify interactions between hepatitis virus associated hepatocellular carcinoma and investigate virus infected plant transcriptome profiles.

The development of advanced RNA sequencing methodologies in recent years has made it possible to monitor viral gene expression at single nucleotide resolution. Protocols for the sequencing of total RNA (RNA-Seq) and ribosome-protected RNA fragments (RPFs) (Ribo-Seq, also known as ribosome profiling) of hepatitis C virus (HCV; genus *Hepacivirus*, family *Flaviviridae*), influenza, Epstein-Barr virus (EBV) and murine leukemia virus (MuLV; genus *Gammaretrovirus*, family *Retroviridae*) have been established. Other advanced RNA-Seq processes including ATAC-Seq and DIVA-Seq that investigate chromatin accessibility for vRNA have also been performed in human immunodeficiency virus (HIV; genus *Lentivirus*, family *Retroviridae*) and herpes simplex virus (HSV; genus *Simplexvirus*, family *Herpesviridae*) infections, respectively, enabling the identification of transcriptionally active regions of viral and cellular nucleic acids.

Due to their level of sensitivity, contamination can be an issue with RNA-Seq-based protocols, with results and analysis needing to be taken in the context of the nature of the sample and library preparation.

DNA Sequencing – Amplicon Sequencing

If the pathogen's identity is known prior to testing or the whole genome is not required, a more suitable and cost-effective approach may be to enrich for an informative part of the viral genome prior to sequencing, using amplification with primers complementary to

known nucleotide sequences within that region. This is a common approach for enriching small viral genomes although the use of PCR introduces certain well-established biases into the resultant data sets. In many clinical settings whole genome sequencing may not be feasible, whilst targeted sequencing techniques of specific regions of the virus e.g., non-structural genes, provides an alternative that has been employed to identify variants involved in disease pathogenesis and prognosis.

Several different PCR techniques can be applied, including long-range PCR, multiplex PCR and microdroplet-based PCR. Multiplex PCR for NGS can include the parallel amplification of large numbers of targets simultaneously as in AmpliSeq or microdroplet-based PCR as used as part of the RainDance platform. Long-range PCR can facilitate the sequencing of entire genes or even genomes with primers generating amplicons with an average length of 10 kb. Examples of PCR amplified approaches to whole viral genome sequencing have been reported for influenza virus, HCV, equine infectious anemia virus (genus *Lentivirus*, family *Retroviridae*), dengue virus and porcine reproductive and respiratory syndrome virus – PRRSV (genus *Betaarterivirus*, family *Arteriviridae*). The advantage here is that sequencing of amplicons will result in a sequence reads being highly specific for the chosen amplicons (high specificity) and with even coverage across the amplicon (dependent upon the nucleotide sequence). The variability of some viruses, including HCV, has necessitated the use of overlapping primer sets in order to capture total sequence heterogeneity. Influenza virus sequencing protocols have used PCR enrichment, as multiple genome segments can be amplified using conserved primers.

Often specific gene panels or hotspot mutation panels can be used when working with multiple targets such as HIV drug resistance genes or can enable the characterization of tumor associated viruses, in parallel with the mutational profiles of their associated tumors as has been used with human papillomavirus (HPV; family *Papillomaviridae*) and human herpesvirus 8 (HHV8; Kaposi's sarcoma-associated virus; genus *Rhadinovirus*; family *Herpesviridae*). This can contribute important information for patient screening and tailoring treatment regimes. Such panels are useful in institutions that process large numbers of samples and offer more efficient and cost-effective approaches to testing. Few comparative studies have been performed using different methods to sequence the same material. One study compared to the performance of different sequencing methods using HCV infected patient material and spiked in-vitro transcripts in serum. Compared to other protocols (including metagenomics, RNA-Seq and enrichment), amplicon sequencing generated greater sequence depth and increased sensitivity, particularly in low viral load samples. PCR amplification can also be useful when working with samples with degraded nucleic acid, often found in DNA extracted from formalin-fixed paraffin-embedded (FFPE) samples. PCR amplification also has its disadvantages, due to its very nature, i.e., amplification. The inclusion of amplification in the library preparation is particularly important when considering SNPs and minority variants as it can result in preferential amplification of specific variants or sequences, resulting in a loss of diversity for those variants that may not be amplified as efficiently. It can also lead to the occurrence of jackpot mutations, which are introduced in early amplification cycles and become dominant in the later cycles as they are carried forward (and therefore indistinguishable from the true viral sequences). Issues with primer bias in primer regions should also not be discounted.

Enrichment Strategies

The purpose of enrichment is to use a selective method to increase the ratio of virus sequences to non-virus sequences, or to reduce the frequency of non-viral sequences, maximizing the number of viral reads to be sequenced, while remaining unbiased towards the natural composition of the genome. This results in an increased specificity of sequencing, whilst eliminating the need for additional culture or amplification, however, it is dependent upon homogeneity between the target and probe/primer regions. Target enrichment is often used with clinical samples and can produce increased genome coverage from limited clinical samples in an economic way, with methods involving the use of DNA or RNA probes, which are potentially overlapping and complementary to the pathogen reference sequence. Unlike PCR, these processes can be performed in a single tube format containing overlapping probes covering the entire genome. In hybridization, such probes are bound to a solid phase and, as in enrichment of rotavirus, can be performed in a streptavidin coated ELISA plate, whereas probes can also be added in liquid phase and be bound later. Success, however, is based on both the a priori knowledge of the viral genome sequence/reference or conserved nature of the viral genomes.

Positive Selection-Based Enrichment

Positive selection-based capture enriches samples for viruses using PCR, microarray or virus capture (in solution-based hybridization), binding and retaining the viral reads within the sample. Poly (A) enrichment is one of the more commonly used modes of positive selection for viruses, with commercial kits readily available. Sequence specific enrichment probes have been used previously to study the human virome with VirCapSeq and VirCap and viroFIND. Although resulting in increased virus derived read numbers, both were limited in the identification of highly divergent viruses. Such probe capture systems have also been used to improve sequencing of viral sequences from FFPE tissue samples resulting in up to 37,000-fold coverage of regions covered by probes. Such probe-based strategies have proven useful in the characterization of enteric viruses in sewage, compared to metagenomic based strategies where viral concentrations were low, paired with high sample complexity. Adapted protocols have employed multiplexing steps to enable high throughput sequencing of such samples on the MinION and Illumina platforms.

Negative Selection-Based Enrichment

Negative selection approaches include ribosomal depletion and involve the binding and removal of non-viral reads from the sample. In addition to being used in gene differential expression protocols, ribosomal depletion has been used to sequence several

viruses including HHV8, GB virus C (genus *Pegivirus*, family *Flaviviridae*) and transfusion transmitted (or torque teno) virus (TTV; genus *Alphatorquevirus*, family *Anelloviridae*).

Single Cell Sequencing

Since their introduction in 2009, single cell sequencing technologies have presented a step forward in the capability to investigate virus-host interactions, with the majority of studies focussing on expression profiling in infected vs non-infected cells. This usage hints at its real potential to dissect the complex relationships between host and virus at the level of individual cells or pathogens. Several alternate approaches have been developed, based on cells sorted either by fluorescence-activated cell sorting (FACS) or using droplet-based workflows. These include Smart-Seq and SmartSeq2 for processing FACS-sorted cells and Drop-Seq, inDrop and 10X Genomics Chromium for droplet-based technologies. Droplet-based methods have advantages including increased throughput and reduced reagent costs, but frequently require more specialist equipment to place nucleic acids and reagents inside droplets.

Viruses are dependent on the host cell to replicate and thus heterogeneity in the host cell population will be a significant factor in viral infection outcome. Achieving and maintaining a 100% infection rate of cells is a challenging task for an infecting virus, due to; (1) cellular heterogeneity – different stages of the cell cycle, different states of cell activation or different subsets of cell, and (2) heterogeneity in the virus particles. Single cell technologies can dissect out these heterogeneities and their impact upon the outcome of infection. Such experimental strategies, however, are not without their challenges, with three of the most significant being; (1) the ability to isolate cells in an appropriate way; (2) the ability to generate comprehensive sequence data from minute amounts of nucleic acid; and (3) how to discriminate between true variation and background noise.

Using Single Cell Sequencing to Investigate Cellular Heterogeneity and its Impact on Viral Infection

Single cell sequencing of viruses has predominantly focussed upon establishing cell to cell variability or cellular heterogeneity and its impact upon viral infection rates. Whilst early studies used fluorescently sorted cells and time lapse microscopy to establish the presence and expression of virus encoded reporter genes, more recent investigations have employed molecular methods including high throughput sequencing protocols to establish cellular expression. Studies of the impact of cellular heterogeneity upon viral infection and outcome have predominantly focussed upon differences in susceptibility to viral infection. Such studies have explored the permissiveness of mouse T cells and human macrophages to human influenza A virus (genus *Alphainfluenzavirus*) infection, identified cellular marker subsets associated with susceptibility to HIV-1 infection and characterized genes associated with latent HIV-1 infection, cell cycle status and transcriptional heterogeneity. These technologies have also been used to study the impact of immune gene expression upon viral infection for HCV, vesicular stomatitis virus (VSV; genus *Vesiculovirus*, family *Rhabdoviridae*) and West Nile virus (genus *Flavivirus*, family *Flaviviridae*). The impact of flavivirus vaccination upon the host transcriptome, notably focussing upon cellular immunity has also been explored. Of note, is a study that investigated cellular heterogeneity of dendritic cell subsets taken from elite controller patients, previously described to have more efficient immune responses to HIV-1 infection. Single cell analysis enabled researchers to characterise markers in a subset of dendritic cells with an improved capability for T cell proliferation. The utilization of a dual RNA-Seq approach in single cells, thereby characterising both host cell and virus expression during HSV-1 infection looked at how the antiviral state in infected cells was reset after virus infection. The study concluded that whilst most cells were able to establish an antiviral state, HSV-1 was able to induce cellular reprogramming in a small number of highly infected cells causing them to revert to embryonic transcriptional states, effectively resetting the cells transcriptome. The elucidation of such novel mechanisms makes single cell sequencing a powerful technique for defining and understanding underlying key events that occur within a cell during viral infection.

Using Single Cell Sequencing to Investigate Viral Heterogeneity

The impact of viral diversity upon the outcome of cell infection has also been the focus of studies, characterising the level of viral diversity within single infected cells during infection. For example the impact of viral diversity upon drug resistance markers during HCV replicon infection was investigated, revealing up to 113 copies of RNA derived from a single HCV replicon present within a single infected cell. Other studies include VSV, FMDV, HIV-1 and Influenza and their impact upon cells during infection, in addition to investigating the role of defective interfering particles (DIPs) in generating diversity during infection and the impact they have upon productive infections of cells.

Two principles are becoming clear from this research which will influence how the field develops – (1) increasingly, evidence suggests that for several viral infections, multiple copies of vRNA from the same virus within single infected cells maintain pools of diversity for subsequent transmission; and (2) there exists a role for transcriptionally distinct cell clusters in response to virus infections. It is possible that these types of study will be able to answer more fundamental questions regarding viral infection, such as identifying molecular switches between lytic and latent infection types.

Single Molecule Sequencing (Pacific Biosciences and Nanopore)

One of the main limitations with short read sequencing (SRS) platforms comes from the requirement for amplification of individual libraries prior to sequencing. This can preferentially amplify specific sequence motifs, resulting in an underrepresentation of bases in

areas of high or low GC content. Amplification can also result in a loss of data concerning RNA and DNA modifications which can convey a significant amount of information regarding the *epi*-transcriptome. Furthermore, during early cycles of amplification, base incorporation errors in individual molecules within a cluster can occur (termed jackpot mutations). These factors make assembly and any related analyses challenging, with theoretical modeling suggesting that decreasing read lengths from 1000 to 100 bp can lead to a minimum of six-fold more decrease in contiguity. This can lead to misalignments and misassemblies in areas of low or high complexity, and impair phasing of variants. For these reasons, SRS of regions of repetitive DNA, high sequence homology, or extreme GC content remains challenging. Finally, with specific impact on virus genome sequencing, both Sanger sequencing and NGS shotgun sequencing are population-based detection (as opposed to single molecule/genome detection) so that their resolution is difficult at haplotype/swarm level, i.e., it is difficult to conclude that any two mutations are linked or associated with one another upon the same genome. Several modifications are now commercially available combining the use of SRS to produce longer synthetic reads, thereby attempting to take advantage of the high throughput and accuracy of short reads, and the benefits of longer reads. This is particularly important for viruses enabling the association of two mutations on the same template/genome. Approaches such as 10x Genomics Chromium use microfluidics to assign barcodes to shorter reads derived from a single template, although the sequencing reaction itself is still amplification driven. Additionally, utilizing long read sequencing (LRS) and SRS in parallel can provide greater accuracy and benefit of both approaches.

“Single molecule” sequencing represents the next evolution for high-throughput sequencing, sometimes termed “third generation sequencing”. Third generation sequencing platforms differ from SRS platforms by directly producing long reads with no amplification during sequencing. Currently two third generation technology platforms technologies are commercially available – Pacific Biosciences (PB) and Oxford Nanopore Technologies (ONT). Both platforms have been developed around novel biochemistries that enable the capture of long read DNA sequencing or cDNA from full length transcripts. Both technologies have higher rates of error (3%–15%) and different error profiles (mostly insertions or deletions) compared to short-read platforms; however, since such errors are random due to the nature of single-molecule sequencing, extremely accurate consensus can be achieved by aligning multiple concatenated copies of the same library, enabling the derivation of a consensus from a single sequence read containing repeated copies of the original DNA molecule. This is defined as “closed circular consensus” (CCS) or “read of insert” (ROI).

Pacific Biosciences

The early PB platform, the RS-II (released in 2010) had relatively short read lengths (~1–2 kb) (for a third-generation sequencer) and an average error rate of (~13%). Since then the technology has improved culminating in the release of a new machine in 2015 – the Sequel, which was also accompanied by a major improvement in single molecule real-time (SMRT) chemistry, providing an almost tenfold difference in output between the two machines. Furthermore, short templates ~1–2 kb can be circularized and sequenced repeatedly, in order to draw a CCS from the multiple copies of the same library. Briefly, hairpin adapters were ligated to either end of the template DNA molecule for generation of capped templates (SMRT-bell). DNA polymerases [otherwise known as Zero-mode waveguides (ZMW)] were located at the base of each nanowell, with a single molecule of DNA. All of this is located on a glass base, so that when fluorescently labeled nucleotides were flushed across the nanowells, during polymerization, fluorescent tags are cleaved off and released, with events being recorded in real-time. If the input DNA is relatively short (less than a few kilobase pairs), the molecule can be sequenced many times, creating a highly accurate consensus.

PB strength is in provision of linkage between individual SNPs, notably in the clinical setting. Studies have included tracing HCV drug resistance mutations, to tracking changes in the envelope genomic region to provide phylogenetic tracing between sexually linked clusters of cases. Additional studies have also applied this targeted protocol to HIV-1 and prevalence of drug resistance mutations. Transcriptomics studies have identified different patterns of splicing events between different families of viruses, with such events being rare in alpha-herpesviruses (Pseudorabies), baculoviruses, and orthomyxoviruses, but more common in beta- and gamma-herpesviruses, retroviruses and hepadnaviruses. The strength of PB in generating large datasets to understand viral transcriptomics was also demonstrated in a study using used 17 RSII SMRT cells and 1 Sequel SMRT cell, complemented with ONT MinION flow cells to produce a novel comprehensive long-read data reference dataset with which to explore transcriptomic variation in HSV-1 infection.

Oxford Nanopore Technologies

In 2014, Oxford Nanopore released a nanopore based sequencer called the MinION. This technology operated by monitoring the electrical conductance changes as nucleic acid (DNA/RNA) was passed through a molecular pore by a motor protein, whose main role was to slow the speed of translocation across the pore sufficiently that the changes were measurable. The smaller ONT platforms, such as the USB stick-sized MinION that costs only around US\$1000, were unique in this age of large, expensive machines since they offered relatively cheap start up investments and unparalleled portability, making in-the-field sequencing a possibility. As a result, the main appeal of ONT technologies has been its potential for mobile or frontline applications, which has been reflected in the number of studies focussing upon this approach with viruses including FMDV, Zika virus and Ebola virus and several plant pathogens. Since the MinION's release two larger models have been commercially marketed with the aim of increasing output by combining multiple nanopore flow cells: the GridION (up to 5 flow cells) and the PromethION (up to 48 flow cells).

This sequencing process generates 1D and 2D reads in which both “1D” strands can be aligned to create a consensus sequence “2D” read. The 1D raw reads have error rates of more than 10%, similar to PB raw reads. Although the error rates of 2D reads are somewhat improved, these were still higher than the consensus reads generated by SMRT sequencing. However, 2D reads were replaced, in 2017, by 1D² where sequencing adapters were added at either end of a piece of nucleic acid, meaning both strands i.e., complement, and template can translocate through the pore providing additional accuracy.

Since its release, there has been significant interest in using nanopore technology to sequence viruses, again, like PB, due to the ability to infer SNPs onto the same read and its portability, however, despite this interest, successful publications have been slow to appear due to the challenges of meeting the significant initial template requirements, which can be a particular issue when working with RNA viruses. There have however been notable successes with DNA viruses, including poxviruses and herpesviruses, due to greater ease in meeting such input requirements. Papers have used nanopore platforms to investigate metagenomics of the aquatic virosphere and transmission in outbreaks, including Ebola virus and Zika virus. Direct RNA sequencing (DRS), however, still has proven challenging, although the first publications are now appearing with Influenza, Coronavirus and PRRSV, demonstrating the potential of the method.

Merging SRS and LRS

Whilst LRS technologies mature and associated applications become established, SRS remains more cost effective, accurate and higher throughput. This is driving several groups to combine both platforms, thus gaining the benefits from both. Whilst the short-read platforms have high base accuracy and coverage, facilitating the identification of transcriptional start sites (TSSs), transcriptional end sites (TSEs), splice junctions and RNA editing, long-read sequencing is better suited to determining 5' and 3' UTR variants, splice isoforms, long non-coding transcripts and the detection of overlapping and embedded transcripts. Studies focussed upon merging both platforms have looked to identify linkage between SNPs in yellow fever virus (genus *Flavivirus*, family *Flaviviridae*), circoviruses (genus *Circovirus*, family *Circoviridae*) and African swine fever virus (ASFV; genus *Asfivirus*, family *Asfarviridae*). Of the above, the ASFV genome poses a unique challenge for NGS, particularly given its current importance as outbreaks spread across the globe, threatening the global porcine population. Its genome is composed of extensive homopolymer and repeat regions including inverted repeats (ITRs) of variable length, in addition to platform-specific limitations to homopolymer frequency should be considered. One study used ASFV-specific target enrichment, Illumina sequencing and long read Nanopore sequencing to generate high quality ASFV genome sequences which was used to correct a previously annotated genome prior to being used as a reference to investigate an outbreak in Moldova. Combined SRS and LRS approaches have also been applied to innate immunity activation in cells after infection with Influenza virus using Illumina-based transcriptomics and full-length PB sequencing of virus genes from single cells, characterising the transcriptomes of baculovirus (family *Baculoviridae*) and pseudorabies (genus *Varicellovirus*, family *Herpesviridae*) virus by combining the high accuracy Illumina reads onto third generation datasets.

Discussion/Conclusions

Next generation sequencing has now been available for 15 years, following the introduction of the 454 Genome Sequencer and pyrosequencing in 2005. Since then, numerous applications have been developed, which have revolutionised the way we view and understand viruses, both as pathogens and how they interact with their environments and hosts. Whilst short read technologies are still maturing, notably in the field of single cell sequencing, the emergence of third generation sequencing platforms are enabling researchers to identify hitherto unknown avenues of research, for instance the interactions of the virome and microbiome, and what implications this has for the health of the host. Additionally, HTS is responsible for the generation of a vast amounts of data which have ushered in new methods in computational biology. With single cell technologies and fourth generation advances, such as spatial transcriptomics now on the horizon it is an exciting time as genomics continues to push the cutting edge in virology. There is ongoing discussion of the increasing role of NGS in clinical diagnostics, with metagenomics type assays potentially not only taking the place of routine nucleic acid testing for blood donations and screening, but also carving out a niche in frontline diagnostic applications.

Further Reading

- Briese, T., Mishra, N., Jain, K., *et al.*, 2014. Middle East respiratory syndrome coronavirus quasispecies that include homologues of human isolates revealed through whole-genome analysis and virus cultured from dromedary camels in Saudi Arabia. *mBio* 5 (3), e01146-14.
- Carroll, M.W., Matthews, D.A., Hiscox, J.A., *et al.*, 2015. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. *Nature* 524 (7563), 97–101.
- Cristinelli, S., Ciuffi, A., 2018. The use of single-cell RNA-Seq to understand virus-host interactions. *Current Opinion in Virology* 29, 39–50.
- Faria, N.R., Kraemer, M.U.G., Hill, S.C., *et al.*, 2018. Genomic and epidemiological monitoring of yellow fever virus transmission potential. *Science* 361 (6405), 894–899.
- Forth, J.H., Forth, L.F., King, J., *et al.*, 2019. A deep-sequencing workflow for the fast and efficient generation of high-quality African swine fever virus whole-genome sequences. *Viruses*. 11.
- Freimanis, G.L., Di Nardo, A., Bankowska, K., *et al.*, 2016. Genomics and outbreaks: Foot and mouth disease. *Revue scientifique et technique* 35 (1), 175–189.
- Martin-Gayo, E., Buzon, M.J., Ouyang, Z., *et al.*, 2015. Potent cell-intrinsic immune responses in dendritic cells facilitate HIV-1-specific T cell immunity in HIV-1 elite controllers. *PLOS Pathogens* 11 (6), e1004930.

- Quick, J., Loman, N.J., Duraffour, S., *et al.*, 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530 (7589), 228–232.
- Russell, A.B., Elshina, E., Kowalsky, J.R., Te Velthuis, A.J.W., Bloom, J.D., 2019. Single-cell virus sequencing of influenza infections that trigger innate immunity. *Journal of Virology* 93 (14).
- Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74 (12), 5463–5467.
- Ziegenhain, C., Vieth, B., Parekh, S., *et al.*, 2017. Comparative analysis of single-cell RNA sequencing methods. *Molecular Cell* 65 (4), 631–643. e4.

Single-Virus Genomics: Studying Uncultured Viruses, One at a Time

Manuel Martínez-García and Francisco Martínez-Hernández, University of Alicante, Alicante, Spain

Joaquín Martínez Martínez, Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Metagenomics The study of genetic material recovered directly from environmental samples. Traditionally, a natural sample is collected and subjected to nucleic acid extraction and sequencing. Thus, the term metagenome refers to the collection of genes sequenced from the environment.

Multiple-displacement amplification A non-PCR, isothermal based DNA amplification technique that uses

random hexamer primers and the high fidelity Φ 29 DNA polymerase. Sufficient DNA for genomics analysis can be obtained from minute amounts of starting DNA template.

Whole-genome amplification A molecular method to amplify the genetic material. In biology, there are different contrasting technologies that allow the amplification of minute amounts of nucleic acids. One such method is multiple-displacement amplification

Single-Virus Genomics: A New Tool for Viral Discovery

Viruses are by far the most abundant biological entities in the oceans, outnumbering any unicellular or multicellular organisms, and represent a major reservoir of genetic diversity that hides an enormous complexity across all marine habitats (Breitbart *et al.*, 2018; Paez-Espino *et al.*, 2016; Roux *et al.*, 2016; Suttle, 2007). Since the discovery by fluorescence microscopy of an unusual amount of ‘virus-like particles’ (VLP) in aquatic ecosystems in the advent of the 90’s, viral ecology has experienced a deep revolution and ‘shaken’ our understanding of marine science. Cultivation techniques for identifying, establishing, and maintaining virus-host systems in the laboratory, though very informative and necessary, are biased since the majority of microbial hosts remain uncultivable (Pedrós-Alió, 2012; Rappé and Giovannoni, 2003). Consequently, over the last decade much of the work carried out for characterizing marine viral diversity has employed a suite of culture-independent approaches, such as metagenomics (Brum *et al.*, 2015; Hurwitz and Sullivan, 2013; Roux *et al.*, 2016). Recently, thanks to researches at local and global scale, such as the Pacific Ocean Virome (POV), or the *Tara* expedition, global oceanic patterns of double-stranded DNA viruses (dsDNA) in surface oceans have been uncovered (Brum *et al.*, 2015; Hurwitz and Sullivan, 2013; Mizuno *et al.*, 2013; Roux *et al.*, 2016), albeit that the majority of the oceans, below mesopelagic, are mostly unexplored (Mizuno *et al.*, 2016; Roux *et al.*, 2016; De Corte *et al.*, 2019), and the documentation of their diversity has only just begun.

Metagenomics – the study of sequenced nucleic acids obtained from bulk environmental samples – provides a culture-independent way to discover novel microbial and viral genetic diversity. Commonly, metagenomic data deliver millions of short DNA reads (150–300 bp) from a viral community and later that complex ‘puzzle’ has to be assembled to deliver complete or draft viral genomes. In contrast, single-virus genomics simplifies this puzzle, as viral genomes are assembled from one virus at a time. In single-virus genomics (SVG), similar to single-cell genomics (Stepanauskas *et al.*, 2017), virus particles are fluorescently stained, sorted by flow cytometry and deposited individually in multi-well plates (Allen *et al.*, 2011; Martínez-Hernández *et al.*, 2017). Later, the capsid of the sorted single-viruses is ‘lysed’ and the genetic material is whole-genome amplified (WGA) and downstream sequenced. The most commonly employed WGA method is the multiple-displacement amplification (MDA). The chance of sorting and recovering a viral particle using this technique is directly dependent on its abundance. Therefore, genomes of abundant virus taxa in a given sample have a greater probability of being recovered. As discussed above, marine virologists have focused on unveiling global patterns of dsDNA viral populations at the family/genus and species levels in the oceans (Brum and Sullivan, 2015; Roux *et al.*, 2016). Despite the power of viral metagenomics, in 2017, the application of single-virus genomics to marine viral samples demonstrated that viral metagenomics had overlooked abundant and cosmopolitan marine viral populations in our planet (Martínez-Hernández *et al.*, 2017). Indeed, that study demonstrated that microdiversity hinders optimal metagenomic assembly, which could explain why the genomes of these abundant viral populations discovered by SVGs have not been identified before. That study, showed that the uncultured single-virus vSAG 37-F6 represented the, possibly, most abundant viral population in ocean surface waters and it was also present in the deep ocean. Now, we know that this uncultured virus is a pelagiphage (Martínez-Hernández *et al.*, 2019). The first single-virus genomic study was a few years before, in 2011 (Allen *et al.*, 2011), but it took a while to apply this technique to a natural environmental sample. In that pioneering study, the authors used fluorescence activated virus sorting to separate single viral particles of bacteriophages lambda and T4 and demonstrated that multiple-displacement amplification was suitable for achieving enough genetic material from a single virus for sequencing and assembling complete genomes. Since that report, so far, although incipient, this technique has been applied to several natural environmental samples, mainly marine, although recently, it was successfully used to sequence uncultured salivary viruses from humans (de la Cruz Peña *et al.*, 2018; Martínez-Hernández *et al.*, 2017; Stepanauskas *et al.*, 2017; Wilson *et al.*, 2017; Bhattacharjee *et al.*, Unpublished). We envision SVGs has great potential to accelerate discovery in the field of viral ecology. The following two sections

discuss and address best practices for applying single-virus genomics and provide all methodological details for laboratory implementation.

Best Practices and Protocol for Single-Virus Genomics

Sample Size and Preservation

Sample volumes used in single-virus genomics are typically in the order of milliliters. There are no agreed-upon standards as to what constitutes a useful sample size to study natural viral assemblages. Instead, sampling decisions should be guided by specific research questions and knowledge of the environment (e.g., estimated relative viral abundance). One milliliter of seawater can provide an insight into viral diversity that is comparable, or in instances superior, to the information derived from viromes produced from tens to hundreds of liters (de la Cruz Peña *et al.*, 2018; Martínez-Hernández *et al.*, 2017; Martínez Martínez *et al.*, 2014; Wilson *et al.*, 2017; Bhattacharjee *et al.*, Unpublished).

After collection, we recommend to store the samples at 4°C and proceed with fluorescence-activated virus sorting (FAVS) as soon as possible. If the sample cannot be processed within a few hours from collection, we strongly suggest preserving samples as described below in detail by following either cryopreservation or a mild fixation with glutaraldehyde. Samples for flow cytometry viral enumeration are usually fixed with 0.5% glutaraldehyde (final concentration) and stored at –80°C (Brussaard, 2004). However, that concentration of glutaraldehyde in the sample hinders WGA. Reducing the fixative to a final 0.1% concentration does not prevent DNA amplification and allows optimum flow cytometry (FCM) resolution of virus-like-particle groups (Martínez-Hernández *et al.*, 2017; Martínez Martínez *et al.*, 2014; Wilson *et al.*, 2017). Alternatively, we recommend, as in single-cell genomics for microbial cells (Martínez-García *et al.*, 2012; Stepanauskas, 2012), the addition of 10% glycerol Tris HCl-EDTA (glyTE, final concentration) prior to cryo-conservation at –80°C. Recent unpublished data from our laboratories indicate this method might be more suitable for SVG than glutaraldehyde fixation by providing a higher percentage of genome recovery (García-Heredia *et al.*, 2020).

Staining and Single-Virus Separation (Sorting)

Several methods are available to individually sort single cells, e.g., micropipetting (Zong *et al.*, 2012), microwell dilution (Gole *et al.*, 2013), or microfluidics (Prakadan *et al.*, 2017). Nevertheless, flow cytometry fluorescence-activated cell sorting (FACS) is the most commonly used method (Rinke *et al.*, 2014). Likewise, flow cytometry sorting is the preferred method for sorting individual viruses (FAVS) for SVG (Allen *et al.*, 2011; de la Cruz Peña *et al.*, 2018; Martínez-Hernández *et al.*, 2017; Martínez Martínez *et al.*, 2014; Stepanauskas *et al.*, 2017; Wilson *et al.*, 2017). However, microfluidics has also been successfully employed for sorting single viruses (Han *et al.*, 2015) and it is regarded as a promising technique for wider implementation (Schmidt and Hawkins, 2016).

Employing fluorescent dyes other than the commonly used DNA-specific SYBR® Green I (Brussaard, 2004) can aid in the detection and discrimination of, at least, certain free virus types. For example, we have noticed that SYBR® Gold dye performs better than SYBR® Green I for fresh unfixed marine samples. Also, for unfixed samples, washing steps with Tris-EDTA (TE) buffer before and after staining may improve detection and sorting of VLPs (Martínez-Hernández *et al.*, 2017), possibly by removing free environmental DNA (eDNA) which can create FCM background when stained with the fluorescence dye. Additionally, a DNase digestion step in solution is recommended to degrade the eDNA and DNA bound to cell debris, prior to adding the fluorescent stain. A control sample in which viruses have been removed (e.g., filtering through 0.02 µm) prior to adding the dye, may aid in discriminating VLP populations from instrument noise and eDNA background during FCM.

When possible, the flow cytometer has to be set in the most accurate mode for sorting single particles e.g., the single sort mode for the BD Influx sorter (Becton Dickinson, San Jose, CA). In addition, sorting at a very low flow rate helps avoiding sorting of doublets (two viruses in the same droplet; see recommended values for sorting viral marine sample in (Martínez-Hernández *et al.*, 2017; Wilson *et al.*, 2017)). Nanobeads with a 220 nm diameter (e.g., Sphero nano fluorescent particles, yellow 0.22 µm, Spherotech Inc.) should be used for instrument fine-tuning (e.g., voltage for fluorescence detection, flow rate; see details for fine-tuning in (Martínez-Hernández *et al.*, 2017)) to improve resolution of nano-particle detection.

As a final consideration, we should highlight the necessity of removing any traces of contaminating DNA within all reagents and material used during FAVS (Martínez-Hernández *et al.*, 2017; Rinke *et al.*, 2014; Woyke *et al.*, 2011). Even very small amounts of DNA contaminants may be amplified in downstream WGA masking, or prevent amplification of the targeted viral DNA which is in the order of femtograms.

Viral Capsid Lysis and Whole Genome Amplification

As a first step, viral capsids have to be broken down – lysed – to release the DNA so it is accessible for enzymatic amplification. Importantly, genome damage/degradation must be avoided during the lysis process. Virus particle lysis is typically achieved via temperature and/or pH shocks (Allen *et al.*, 2011; de la Cruz Peña *et al.*, 2018; Martínez-Hernández *et al.*, 2017; Martínez Martínez *et al.*, 2014; Stepanauskas *et al.*, 2017; Wilson *et al.*, 2017). Based on mixed results across previous published studies and recent

unpublished results from our laboratories (Garcia-Heredia *et al.*, 2020), the exact lysis treatment may necessitate tweaking for different virus types, possibly depending on specific protein capsid structures and the presence/absence of an envelope.

Multiple displacement amplification using Phi29 DNA polymerase has been the most common method for WGA. It has been used in the first application of SVG with cultures of lambda and T4 bacteriophages (Allen *et al.*, 2011), in marine environmental based studies (Martinez-Hernandez *et al.*, 2017; Stepanauskas *et al.*, 2017; Wilson *et al.*, 2017), and also in the first application of SVG to human-derived samples (de la Cruz Peña *et al.*, 2018).

Recently, EquiPhi29 (Thermo Fisher Scientific) – a thermostable mutant of the phi29 polymerase – has been proven to significantly improve genome recovery and to reduce amplification reaction time (Stepanauskas *et al.*, 2017). Another improvement toward enabling near-complete WGA from single cells may be TruePrime® RCA (Sygnis/Expedeon), which combines the Phi29 polymerase with the *Thermus thermophilus* (Tth) PrimPol's primase activity preferring dNTPs as substrates for synthesizing DNA primers (Picher *et al.*, 2016).

Sequencing

In the single-virus genomic studies mentioned above, single-amplified virus genomes were sequenced using mainly Illumina technology (pair-end 150–300 bp $\times$ 2). We recommend at least a sequencing coverage of $\times$ 1000 per single virus (assuming a mean viral genome size of 40 kb). Very recently, some single-viruses from our lab have been successfully sequenced by Nanopores proving that this technology is compatible with MDA products. Thus, the novel long-read sequencing platforms, such as Pacbio or Nanopore will undoubtedly have potential for improving genome assembly and completeness of single viruses in further studies.

Detailed Protocol

The following protocol is best suitable for aquatic samples but may be applicable to samples of another nature, e.g., soil, tissue, blood or other bodily fluids, following extraction of the viral particles from other materials in the sample, purification, and resuspension in a suitable solution, such as water or Tris HCl-EDTA buffer. A similar workflow can be applied to solid samples, such as soils, as long as viruses have been previously resuspended.

Sample Collection and Preservation

Step 1: Collect 5–50 mL of your sample in sterile tubes.

- If viral abundance is low, sampling volume might need to be adjusted.

Step 2: Optionally, samples can be filtered through a 0.2 μ m filter to remove cellular fractions or a natural sample can be processed unfiltered. The latter procedure is recommended since it ensures that large viruses remain in samples and are not removed in the filtration step. However, in the former, the viral fraction has increased purity and is free of potential co-contamination and co-sorting of single-cells during sorting. In any case, after sampling, always keep samples at 4°C until sorting.

- Bear in mind that this filtration step could remove some viruses, especially large viruses. Depending upon the objectives of the study and the target viruses, other methods such as centrifugation, can be used to separate viruses from cells, or even filtration can be omitted. In this case, FAVS must be done as soon as possible to avoid changes in the community structure. However, as discussed below, preservation of samples can be applied.

Step 3: If sorting cannot be performed on the same sampling day, the sample could be preserved in 10% GlyTE buffer as detailed below:

- (1) Prepare GlyTE mixing 20 mL 100 $\times$ TE pH 8.0, 60 mL deionized water and 100 mL molecular-grade glycerol.
- (2) Filter GlyTE through a 0.2 μ m filter.
- (3) Add 100 μ L of GlyTE buffer to 1 mL of sample.
- (4) Mix for 1 min.
- (5) Store at -80°C until use.

Alternatively, samples can also be preserved in formaldehyde at 0.1% v/v final concentration and the sample stored at -80°C until required.

Viral Staining

If the sample has been cryopreserved, thaw it on ice and begin the staining procedure as follows:

- (1) Add sample (1 mL approx.) to a 10 kDa centrifugal filter (e.g., Nanosep 10 kDa (OMEGA, Pall Life Sciences)).
- (2) Centrifuge at maximum speed for 2 min, to concentrate the viral sample.
- (3) Add 500 μ L of sterile TE buffer to the sample in the centrifugal filter device.
- (4) Mix and resuspend by inversion and vortex strongly for 15 s

- (5) Transfer viruses in TE to a new Eppendorf tube.
 - (6) Add 2.5 μL of SYBR-Gold working solution (1/10 dilution of stock) 0.02 μm filtered to the 500 μL of viral sample.
 - (7) Stain at room temperature, in the dark for 20 min.
 - (8) Transfer staining viral sample to a 10 kDa centrifugal tube.
 - (9) Centrifuge at maximum speed for 2 min up to 50 μL approximately.
 - (10) Discard the eluate and add 500 μL of sterile 0.02 μm -filtered TE buffer to the viruses.
 - (11) Mix by inversion and centrifuge at maximum speed for 2 min, to wash the remaining free SYBR Gold dye from the sample.
 - (12) Repeat washing step 2 more times.
 - (13) Add 500 μL of sterile 0.02 μm -filtered TE buffer to the sample.
 - (14) Mix carefully and recover the stained viral fraction into a new Eppendorf tube.
 - (15) The sample is ready for sorting.
- Autoclaved TE buffer used in the staining procedure is previously ultraviolet treated for 16 h and 0.02 μm -filtered through Anotop filters (Whatman).
 - We advise filtering all solutions through 0.02 μm syringe filters before using.
 - An identical staining procedure can be done with a 0.02 μm -filtered, UV-treated TE buffer, to use as a blank before the viral sample sorting.

Fluorescence-Activated Virus Sorting

DNA decontamination and fluorescence-activated virus sorting preparation

- (1) For a DNA decontamination procedure for a flow cell sorter, we recommend strictly following the recommendations of [Rinke et al. \(2014\)](#). In brief, if the sample is from a marine environment, prepare at least 4 L of sterile 1 $\times$ phosphate buffered saline (PBS) one day before sorting to use as sheath fluid. Decontaminate the sheath fluid and the sheath tank of the sorter by overnight UV exposition by opening the tank and placing it into a laminar hood with a UV-lamp (for more details see [Rinke et al., 2014](#)). On the day of sorting, run 1 L of sodium hypochlorite (bleach) 0.3%–0.8% final concentration, in deionized water, through the fluidic line of the sorter for 2 h. Clear the hypochlorite solution by running 1 L of sterile water through the fluidic tubing for at least 30 min as per the recommendation ([Rinke et al., 2014](#)). If the sorter is within a laminar flow hood with a UV lamp, bleach all the internal accessible surface of the sorter and hood, and UV treat for 30 min prior to analyses and sorting.

Preparation of 384-well plates for sorting

- (1) Use autoclaved 384-well plates for sorting (note: some commercially available multiwell plates are not sterile).
 - (2) Add 0.6 μL of the UV-treated sterile TE buffer to each well.
 - (3) UV-treat the 384-well plates containing TE buffer for 10 min.
 - (4) Seal the plates and keep at low temperature (4°C) until sorting.
- Plates not used for sorting can be stored at -20°C for further use. In some cases, after freezing, plates can become slightly distorted, which can affect the deposition of the droplet containing the single-virus in the center of the well.
 - Immediately before sorting, the plates should be centrifuged at low speed to make sure the TE buffer is in the bottom of the well.

Sorting

- (1) Stained viral samples should be screened through a 35–40 μm mesh-size cell strainer (e.g., BD Biosciences) to ensure the removal of large particles that can potentially clog the fluidic line during sorting.
- (2) Analyze your stained viral sample in green fluorescence detecting side scatter (SSC) and/or forward scatter (FSC). Stained blanks will aid in defining the target sorting population gate. At this point, the experience of the flow cytometer sorter operator is paramount. Please refer to a previous reference protocol developed by ([Brussaard, 2004](#)) to detect viruses by flow cytometry. Nanoscale fluorescent beads (e.g., 220 nm 1-peak yellow beads (Sphero Nano Fluorescent Particles (SpheroTech Inc.)) may be run to set up and fine tune the sorter (for details see [Martinez-Hernandez et al., 2017](#)).
- (3) Sort an individual viral particle per well in the plate. Leave 2–3 columns or rows without sorted viruses to use as a negative control during MDA. Use several wells (e.g., 4–10) for sorting multiple virus particles (e.g., 10 per well, 20 per well and 100 per well). ([Fig. 1](#))
- (4) Seal the plate immediately after sorting and store it at -80°C .

Whole-Genome Amplification

Preliminary notes:

- The whole genome process and reagent preparation are done in a dedicated laminar flow hood.
- A lab coat, gloves, sleeves cover, protection glasses, and a mask should be worn. Bleach your gloves frequently.

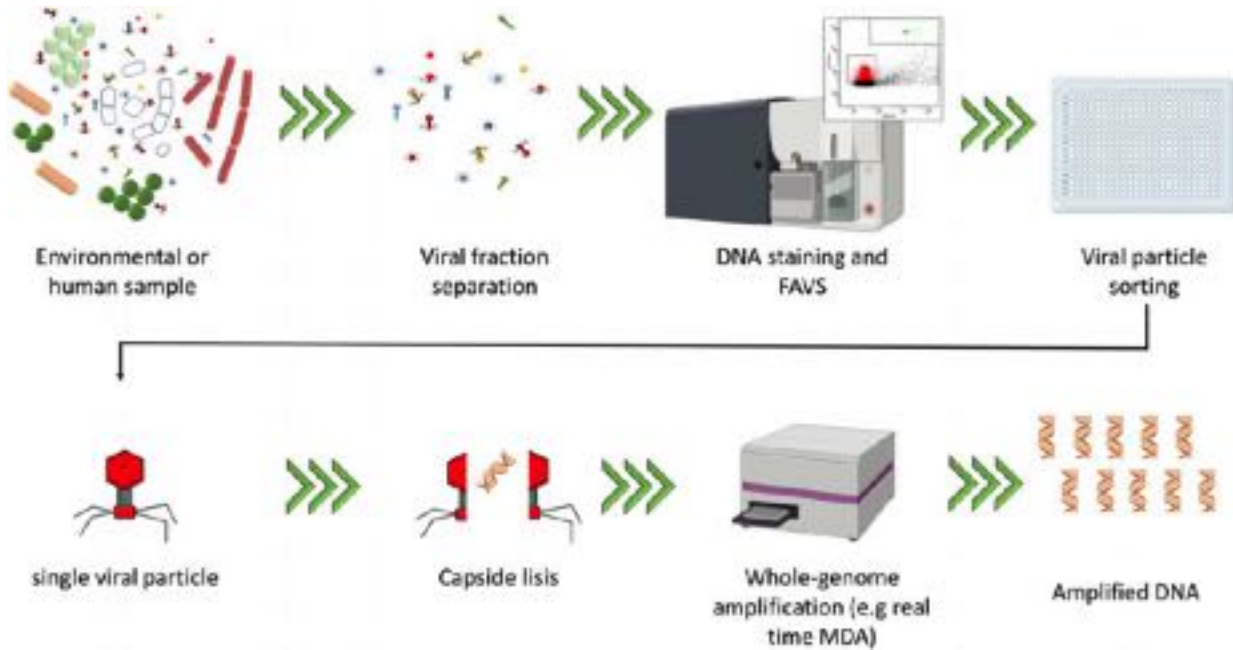


Fig. 1 *Single-virus genomics workflow.* Viruses are fluorescently stained (e.g., SYBR Gold for dsDNA viruses) and sorted one at a time from the natural sample using fluorescence-activated virus sorting (FAVS). Sorted single viral particles are deposited in multi-well plates and then the capsid is lysed by a combination of pH and temperature shock. Free viral DNA is then subjected to whole genome amplification (e.g., real time multiple-displacement amplification, MDA). After amplification, enough DNA is available for DNA sequencing or target screening.

- Always handle all reagents using coolers to maintain low temperatures. Bleach the surface of tubes and reagents with a wipe.
- Water used in whole genome amplification has to be sterile deionized water and UV-treated for 16 h.

For more details, see [Rinke *et al.* \(2014\)](#).

DNA decontamination

- (1) Carefully clean all laminar flow hood surfaces, and all eppendorfs and tubes employed for the WGA amplification, with new prepared 10% bleach in deionized water.
- (2) Before WGA, UV-treat the laminar flow hood for 1 h.

Reagent preparation

- (1) DLB buffer (Quiagen 1031206) is used for viral lysis.
- (2) Stop solution (Quiagen 1032393) is used to neutralize the effects of the direct lysis buffer (DLB). It comes ready for use.

DLB preparation

- (1) Add 500 μ L deionized UV-treated water to the lyophilized DLB.
 - (2) Thaw 1 M DTT, vortex and add 45.5 μ L.
- You need 800 μ L of DLB buffer per 384-well plate.
 - Any leftover can be stored at -20°C and used in the next whole genome amplification (Only one thaw cycle is allowed).
- (1) Vortex both tubes and seal using parafilm.

Stop solution

- (1) Add 900 μ L of stop solution to a new Eppendorf tube.
- Any leftover can be stored at -20°C and used in the next whole genome amplification (Only one thaw cycle is allowed).
- (1) Seal the Eppendorf tube with the 900 μ L of Stop solution using parafilm.

DNA decontamination of DLB and stop solution

- UV crosslinker CL-1000 UVP was used to irradiate the reagents ($999.9 \mu\text{J cm}^{-2}$). Other UV crosslinkers can be used, but we strongly suggest to estimate the UV dose according to the UV lamp used.
- (1) Put the Eppendorf tubes containing DLB and Stop solution floating on sterile deionized water in a cooler unit at 8 cm distance to the UV bulbs.
 - (2) UV-treat the tubes for 30 min on one side.
 - (3) Vortex the tubes.
 - (4) UV-irradiate for 30 more minutes on the other side.

Master mix preparation

The whole genome amplification described here is done applying the modified WGA-X method (Stepanauskas *et al.*, 2017) which uses the novel thermotolerant EQUIPhi29 DNA Polymerase (Thermo Fisher Scientific).

8.0 μL of master mix should be added to each well, containing:

- 5.50 μL of sterile deionized and UV-treated (16 h) water
- 1.00 μL of Reaction buffer (supplied with the EQUIPhi29 DNA pol)
- 0.4 μL of dNTPs 10 mM each
- 0.1 μL of DTT 1 M
- 0.8 μL of heptamers 0.5 mM
- 0.2 μL of EQUIPhi29 DNA polymerase ($10 \text{ U } \mu\text{L}^{-1}$)

Calculate the total volume needed for your sample, considering the pipetting error.

e.g., For a 384-well plate and considering a 20% pipetting error, total volumes would be:

- 2534.86 μL of sterile deionized and UV-treated (16 h) water
- 460.80 μL of Reaction buffer (supplied with the EQUIPhi29 DNA pol)
- 184.32 μL of dNTPs 10 mM each
- 46.08 μL of DTT 1 M
- 368.64 μL of heptamers 0.5 mM
- 92.16 μL of EQUIPhi29 DNA polymerase ($10 \text{ U } \mu\text{L}^{-1}$)

Divide to the total volume in Eppendorf tubes (maximum 750 μL per tube), seal them with parafilm and irradiate them as for DLB and Stop solution, but for a total time of 30 min (vortex and turn them after the first 15 min).

After the UV irradiation step, add SYTO-9, 1.25 μM final concentration, and store at 4°C until use. SYTO-9 dye is UV photosensitive.

The whole genome reaction procedure

- (1) Thaw and spin down the plate with the sorted single-viruses.
- (2) Place the plate in a cooler.
- (3) Add 0.7 μL of DLB to each well and mix thoroughly to “lyse” the viruses.
- (4) Incubate for 10 min at room temperature. Seal the plate and spin it down.
- (5) Add 0.7 μL of stop solution, and mix thoroughly again, to neutralize the DLB.
- (6) Seal and spin down the plate.
- (7) Add 8 μL of the master mix (with SYTO-9 added) to all samples and mix thoroughly.
 - (a) Carefully add 1 ng and 10 ng of viral DNA positive control to some empty wells with no viral sorted particles. Dedicate at least 4 wells for the positive control. *Positive controls should be placed in peripheral wells to minimize the risk of contamination.*
- (8) Seal and spin down the plate.
- (9) Run the multiple displacement amplification (MDA reaction) at 45°C for approximately 4 h in a plate reader. Time can be adjusted according to the efficiency of the reaction by monitoring the positive and negative control wells. Positive control wells should amplify in 30 min, while negative start to amplify after 4 h approx. Single-viruses should amplify in between those 0.5–3 h approximately.
- (10) After 4 h stop the MDA reaction incubating samples at 75°C for 15 min.
- (11) Make a 1:10 dilution in a new plate adding 1 μL of MDA product and 10 μL of sterile TE 0.02 μm filtered and UV-treated buffer.
- (12) Store the MDA product and dilution plate at -80°C .

References

- Allen, L.Z., Ishoey, T., Novotny, M.A., *et al.*, 2011. Single virus genomics: A new tool for virus discovery. *PLoS ONE* 6, e17722. doi:10.1371/journal.pone.0017722.
- Bhattacharjee, A.S., Schulz, F., Woyke, T., Orcutt, B.N., Martínez Martínez, J., (unpublished). Giant viruses from Juan de Fuca Ridge flank oceanic crustal fluids. In prep.
- Breitbart, M., Bonnain, C., Malki, K., Sawaya, N.A., 2018. Phage puppet masters of the marine microbial realm. *Nature Microbiology* 3, 754–766. doi:10.1038/s41564-018-0166-y.

- Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., *et al.*, 2015. Patterns and ecological drivers of ocean viral communities. *Science* 348, 1261498. doi:10.1126/science.1261498.
- Brum, J.R., Sullivan, M.B., 2015. Rising to the challenge: Accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* 13, 147–159. doi:10.1038/nrmicro3404.
- Brussaard, C.P.D., 2004. Optimization of procedures for counting viruses by flow cytometry. *Applied and Environmental Microbiology* 70, 1506–1513. doi:10.1128/AEM.70.3.1506–1513.2004.
- de la Cruz Peña, M., Martínez-Hernández, F., García-Heredia, I., *et al.*, 2018. Deciphering the human virome with single-virus genomics and metagenomics. *Viruses* 10, 113. doi:10.3390/v10030113.
- De Corte, D., Martínez Martínez, J., Cretoiu, M.S., *et al.*, 2019. Viral Communities in the Global Deep Ocean Conveyor Belt Assessed by Targeted Viromics. *Frontiers in Microbiology* 10. doi:10.3389/fmicb.2019.01801.
- García-Heredia, I., Bhattacharjee, A., Fornas, O., *et al.*, 2020. Benchmarking of Single-Virus Genomics: A new tool for uncovering the virosphere. *Environmental Microbiology*, 1462–2912. doi:10.1111/1462-2920.15375.
- Gole, J., Gore, A., Richards, A., *et al.*, 2013. Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells. *Nature Biotechnology* 31, 1126–1132. doi:10.1038/nbt.2720.
- Han, H.-S., Cantalupo, P.G., Rotem, A., *et al.*, 2015. Whole-genome sequencing of a single viral species from a highly heterogeneous sample. *Angewandte Chemie International Edition* 54, 13985–13988. doi:10.1002/anie.201507047.
- Hurwitz, B.L., Sullivan, M.B., 2013. The Pacific Ocean Virome (POV): A marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* 8, e57355. doi:10.1371/journal.pone.0057355.
- Martínez-García, M., Brazel, D.M., Swan, B.K., *et al.*, 2012. Capturing single cell genomes of active polysaccharide degraders: An unexpected contribution of verrucomicrobia. *PLoS ONE* 7, e35314. doi:10.1371/journal.pone.0035314.
- Martínez-Hernández, F., Fornas, O., Lluésma Gómez, M., *et al.*, 2017. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nature Communications* 8, 15892. doi:10.1038/ncomms15892.
- Martínez-Hernández, F., Fornas, O., Lluésma Gómez, M., *et al.*, 2019. Single-cell genomics uncover Pelagibacter as the putative host of the extremely abundant uncultured 37-F6 viral population in the ocean. *The ISME Journal* 1. doi:10.1038/s41396-018-0278-7.
- Martínez Martínez, J., Swan, B.K., Wilson, W.H., 2014. Marine viruses, a genetic reservoir revealed by targeted viromics. *The ISME Journal* 8, 1079–1088. doi:10.1038/ismej.2013.214.
- Mizuno, C.M., Ghai, R., Saghai, A., López-García, P., Rodríguez-Valera, F., 2016. Genomes of abundant and widespread viruses from the deep ocean. *mBio* 7, e00805–e00816. doi:10.1128/mBio.00805-16.
- Mizuno, C.M., Rodríguez-Valera, F., Kimes, N.E., Ghai, R., 2013. Expanding the marine virosphere using metagenomics. *PLoS Genetics* 9, e1003987. doi:10.1371/journal.pgen.1003987.
- Paez-Espino, D., Elie-Fadrosh, E.A., Pavlopoulos, G.A., *et al.*, 2016. Uncovering Earth's virome. *Nature* 536, 425–430. doi:10.1038/nature19094.
- Pedrés-Alió, C., 2012. The rare bacterial biosphere. *Annual Review of Marine Science* 4, 449–466. doi:10.1146/annurev-marine-120710-100948.
- Picher, Á.J., Budeus, B., Wafzig, O., *et al.*, 2016. TruePrime is a novel method for whole-genome amplification from single cells based on TthPrimPol. *Nature Communications* 7, 13296. doi:10.1038/ncomms13296.
- Prakadan, S.M., Shalek, A.K., Weitz, D.A., 2017. Scaling by shrinking: Empowering single-cell “omics” with microfluidic devices. *Nature Reviews Genetics* 18, 345–361. doi:10.1038/nrg.2017.15.
- Rappé, M.S., Giovannoni, S.J., 2003. The uncultured microbial majority. *Annual Review of Microbiology* 57, 369–394. doi:10.1146/annurev.micro.57.030502.090759.
- Rinke, C., Lee, J., Nath, N., *et al.*, 2014. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nature Protocols* 9, 1038–1048. doi:10.1038/nprot.2014.067.
- Roux, S., Brum, J.R., Dutilh, B.E., *et al.*, 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537, 689–693. doi:10.1038/nature19366.
- Schmidt, H., Hawkins, A.R., 2016. Single-virus analysis through chip-based optical detection. *Bioanalysis* 8, 867–870. doi:10.4155/bio-2016-0004.
- Stepanuskas, R., 2012. Single cell genomics: an individual look at microbes. *Current Opinion in Microbiology* 15, 613–620. doi:10.1016/j.mib.2012.09.001.
- Stepanuskas, R., Fergusson, E.A., Brown, J., *et al.*, 2017. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. *Nature Communications* 8, 84. doi:10.1038/s41467-017-00128-z.
- Suttle, C.A., 2007. Marine viruses – Major players in the global ecosystem. *Nature Reviews Microbiology* 5, 801–812. doi:10.1038/nrmicro1750.
- Wilson, W.H., Gilg, I.C., Moniruzzaman, M., *et al.*, 2017. Genomic exploration of individual giant ocean viruses. *The ISME Journal* 11, 1736–1745. doi:10.1038/ismej.2017.61.
- Woyke, T., Sczyrba, A., Lee, J., *et al.*, 2011. Decontamination of MDA reagents for single cell whole genome amplification. *PLoS ONE* 6, e26161. doi:10.1371/journal.pone.0026161.
- Zong, C., Lu, S., Chapman, A.R., Xie, X.S., 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* 338, 1622–1626. doi:10.1126/science.1229164.

Further Reading

- Castillo, Y.M., Mangot, J., Benites, L.F., *et al.*, 2019. Assessing the viral content of uncultured picoeukaryotes in the global-ocean by single cell genomics. *Molecular Ecology* 28, 4272–4289. doi:10.1111/mec.15210.
- Labonté, J.M., Swan, B.K., Poulos, B., *et al.*, 2015. Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *The ISME Journal* 9, 2386–2399. doi:10.1038/ismej.2015.48.
- Martínez-García, M., Swan, B.K., Poulton, N.J., *et al.*, 2012. High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *The ISME Journal* 6, 113–123. doi:10.1038/ismej.2011.84.
- Paez-Espino, D., Roux, S., Chen, I.-M.A., *et al.*, 2019. IMG/VR v2.0: An integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Research* 47, D678–D686. doi:10.1093/nar/gky1127.
- Roux, S., Emerson, J.B., Elie-Fadrosh, E.A., Sullivan, M.B., 2017. Benchmarking viromics: An in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5. doi:10.7287/peerj.preprints.3053v1.
- Sieracki, M.E., Poulton, N.J., Jaillon, O., *et al.*, 2019. Single cell genomics yields a wide diversity of small planktonic protists across major ocean ecosystems. *Scientific Reports* 9, 6025. doi:10.1038/s41598-019-42487-1.
- Swan, B.K., Tupper, B., Sczyrba, A., *et al.*, 2013. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proceedings of the National Academy of Sciences* 110, 11463–11468. doi:10.1073/pnas.1304246110.

Biophysical Characterizations in the Solution State

Robert P Rambo and Katsuaki Inoue, Diamond Light Source, Didcot, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Structural investigations in the solution state provide the opportunity to study biological particles in near native conditions. Solution-state small angle X-ray scattering (SAXS) has a long history as a complementary technique to high resolution structural methods such as X-ray crystallography (MX), nuclear magnetic resonance spectroscopy (NMR) and now, cryo-electron microscopy (EM). SAXS can efficiently validate structural hypotheses, answering the basic question “Does the high-resolution atomistic model exist in solution?” Additionally, SAXS is often used to model missing tails or loops from high resolution structures as everything in the path of the incident X-ray beam will scatter X-rays. Consequently, atomistic interpretation of a macromolecular SAXS experiment relies critically on constraints provided by prior chemical knowledge (e.g., secondary structure assignments, dihedral angle restraints) and molecular dynamics. The coupling of SAXS with computational approaches allows SAXS experiments to be interpreted at the residue-level and has led to insightful structure-function relationships.

A solution state SAXS measurement is a resolution-limited, picture of the conformational landscape determined at a given condition (i.e., salt, pH, temperature). For flexible systems, the modeling of SAXS data requires extensive computational searches. Here, flexibility is inferred by the requirement of multiple conformations to satisfy the SAXS observations. Residues enabling chain dynamics must be assigned during the computational searches which are often identified by limited-proteolysis, NMR, sequence alignments and secondary structure predictions. The validation of flexibility by SAXS is challenging as the time-resolution of SAXS has often been too slow to capture the micro- to milli-second (ms) time scales of many biological functions. Alternative methods must be pursued such as synchrotron-based X-ray Footprinting and Mass Spectrometry (XFMS) and Diffracted X-ray Tracking (DXT).

XFMS and DXT are solution-state methods capable of informing on sub-millisecond reaction time-scales. XFMS is a destructive method that interrogates the solvent accessibility of residues using X-ray induced radiolysis of water. Residues that are highly accessible to a water molecule will form hydroxyl adducts that are subsequently identified using mass spectrometry. Many internal waters will change their occupancies during a conformational change and XFMS has allowed identification of functionally critical waters protein structures. XFMS reactions occur within the sub micro-second to micro-second time window enabling interrogation of biological relevant time-scales. Likewise, DXT is a true, real-time, time-resolved technique that informs directly on dynamics. DXT requires the attachment of a gold nanocrystal to label the biological particle. Polychromatic X-rays will diffract from the gold nanocrystal producing a single diffracted spot that moves as the protein wiggles about in solution. High-speed X-ray cameras are used to track changes in the diffracted spot at nano-second time resolutions. Both XFMS and DXT can also be performed using *in vivo* systems, allowing for direct observations of the biological particles in native conditions.

Solution-State Small Angle X-Ray Scattering

Small angle X-ray scattering in the solution state examines the scattering of X-rays by particles in solution. The particles are assumed to be randomly oriented and at sufficiently dilute conditions such that there does not exist any interparticle interactions or correlations. These conditions comprise the prototypical bioSAXS experiment and the SAXS signal is recovered as a background subtracted measurement. Contemporary bioSAXS is routinely performed using size-exclusion chromatography coupled instruments where SAXS measurements are made during purification of the sample that ensures both sample homogeneity and ideal background subtraction.

The scattering of X-rays is by electrons and it is the spatial arrangement of the particle's electrons that provide the observed SAXS signal. Here, the particle's distribution of electrons is known as the particle form-factor, $A(q)$, which is a function of scattering-vector, q . Changes in the particle's conformation will cause a comparable change in the particle's form-factor and consequently in the shape of the SAXS curve. Since SAXS curves can be readily calculated from atomistic or high-resolution models, SAXS offers an efficient method for validating the solution state.

Solution-state SAXS is an ideal investigative tool for studying viruses as assembled viruses are hierarchical structures with dimensions much larger than their constituent capsid proteins. Time-resolved SAXS experiments allow for the simultaneous detection and monitoring of the smaller structures that are required to nucleate and form the much larger capsid structure. Furthermore, SAXS is often used to detect and model conformational changes. SAXS studies on the C-terminal transmembrane domain (TMD) from parainfluenza virus (PIV) demonstrated lipid dependent conformational changes. The effect of this conformational change was further explored using model membranes that showed that the TMD induces a negative topological curvature in membranes suggesting an active role in membrane fusion.

SAXS studies of the matrix (m) protein from the Newcastle disease virus (NDV) revealed the protein exists as a stable dimer in solution that self assembles at neutral pH into a hollow helical structure by recruiting tetramers of m-NDV protein. A shift to acidic environment disassembles the hierarchical structure and stabilizes dimer formation. These observations were readily made by performing SAXS experiments at different pH conditions. Furthermore, since SAXS measures the molecular form-factor, shape retrieval algorithms could be applied which consequently revealed the shape of the m-NDV dimer and its hierarchical helical structure.

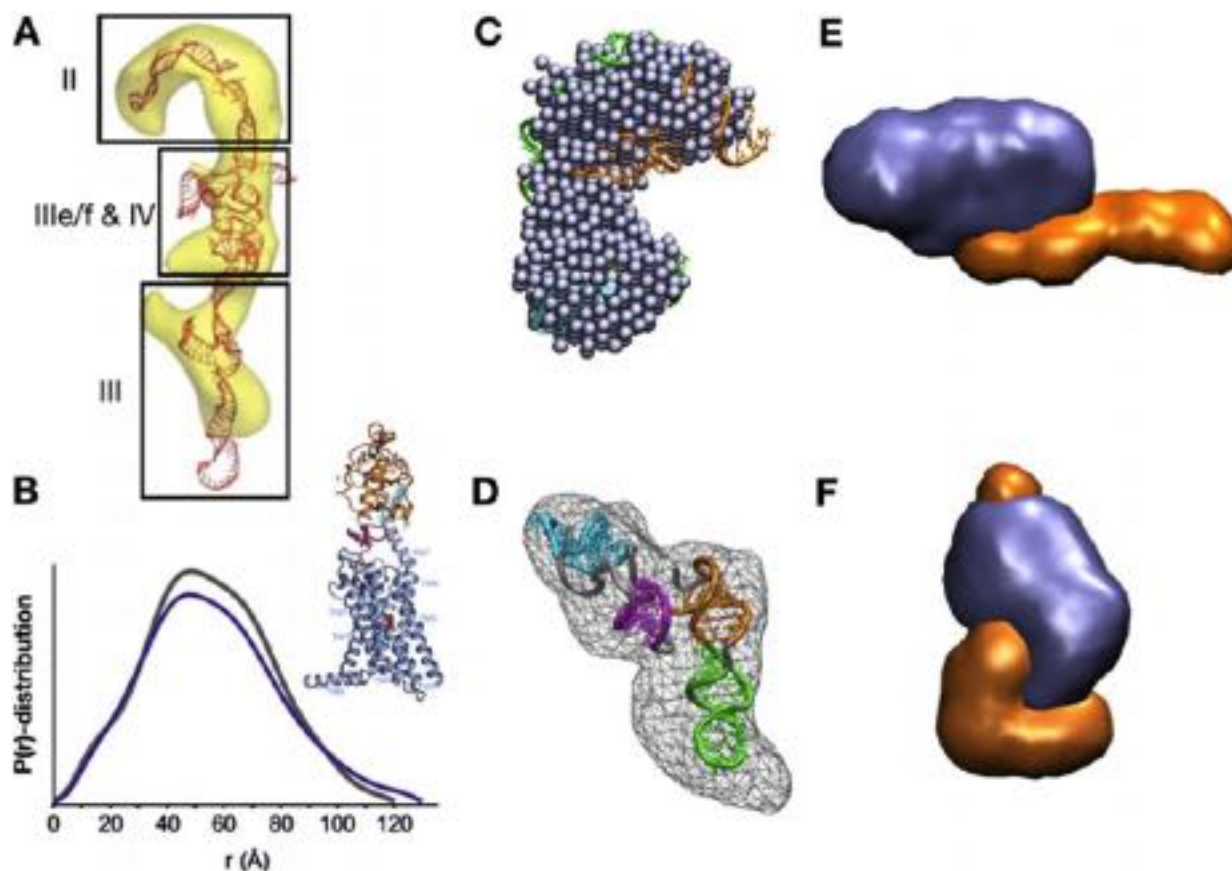


Fig. 1 Popular applications of bioSAXS experiments. (A) All atom solution-state SAXS model of HCV IRES (red) containing domains II, III and IV. Model is overlaid with low resolution cryo-EM density (yellow) extracted from the IRES bound to large ribosomal subunit. (B) Detection of conformational change within amphipol-stabilized G-protein coupled receptor in presence (purple) and absence (ligand) of ligand. P(r)-distribution is the set of all paired-distances with the particle as measured by SAXS. A comparable change in the P(r)-distribution suggests conformational change (image adapted from). (C) SAXS based reconstruction (purple) of a protein RNA complex (green, orange, cyan) using IKETAMA (Diamond Light Source, UK). (D) Volumetric *ab initio* model of Tetrahymena telomerase p65 RNA (mesh). Atomistic model suggest two of the four helices are co-axial stacked in solution. (E) and (F) Differential contrast volumetric models using MONSA software (EMBL Hamburg). (E) protein DNA complex SMARCAL1, (F) Splicing related protein RNA complex. Protein phase in purple and nucleic acid phase in orange. Image adapted from: (A) Pérard, J., Leyrat, C., Baudin, F., Drouet, E., Jamin, M., 2013. Structure of the full-length HCV IRES in solution. *Nature Communications* (4), 1612–1623. (B) Byrne, E.F.X., Sircar, R., Miller, P.S., *et al.*, 2016. Structural basis of smoothened regulation by its extracellular domains. *Nature* 535, 517–522.

Throughout their life-cycle, viruses form a variety of complexes with host nucleic acids, proteins and small molecules, and SAXS has been a core technique for validating oligomeric states and shapes, providing low-resolution models and suggesting new functions (Fig. 1). Combined SAXS and molecular modeling on RNA fragments from HIV-1 genome specifically containing the transactivation (TAR)/polyadenylation (polyA), primer-binding site (PBS), and Psi-packaging domains were able to propose atomistic models that allowed direct interpretation of the SAXS data at the residue level. Notable changes in the tRNA-like element loop suggested a function required for release of the tRNA primer in genome replication. Similarly, a SAXS-based approach determined a full-length model of the Hepatitis C Virus (HCV) internal ribosome entry site (IRES) that revealed the structure of the IRES free in solution. SAXS based modeling is an effective strategy for determining highly structured RNA or DNA elements derived from viral genomes. Furthermore, combined NMR and SAXS studies on viral chaperone phosphoprotein P from Nipah virus showed P is globally disordered and persists as a disordered protein when bound to its target viral nucleoprotein, N⁰. The disorder is hypothesized to be required for P to act as a tether for enriching N⁰ at the site of viral RNA synthesis.

SAXS as a Structural Tool (Theoretical Considerations)

There are various mathematical formalisms to describe SAXS including atomistic modeling, form-factor fitting and indirect Fourier transform (IFT). An IFT relates the set of all electron-electron paired distances, P(r)-distribution, within the macromolecule to the SAXS intensities (Fig. 2). The P(r)-distribution is a histogram of all intermolecular distances and is equivalent to the squared particle form-factor. The P(r)-distribution informs on the maximum dimension of the particle, the radius-of-gyration (distribution of mass about the particle center) and likewise, will demonstrate conformation dependent changes. The P(r)-distribution is

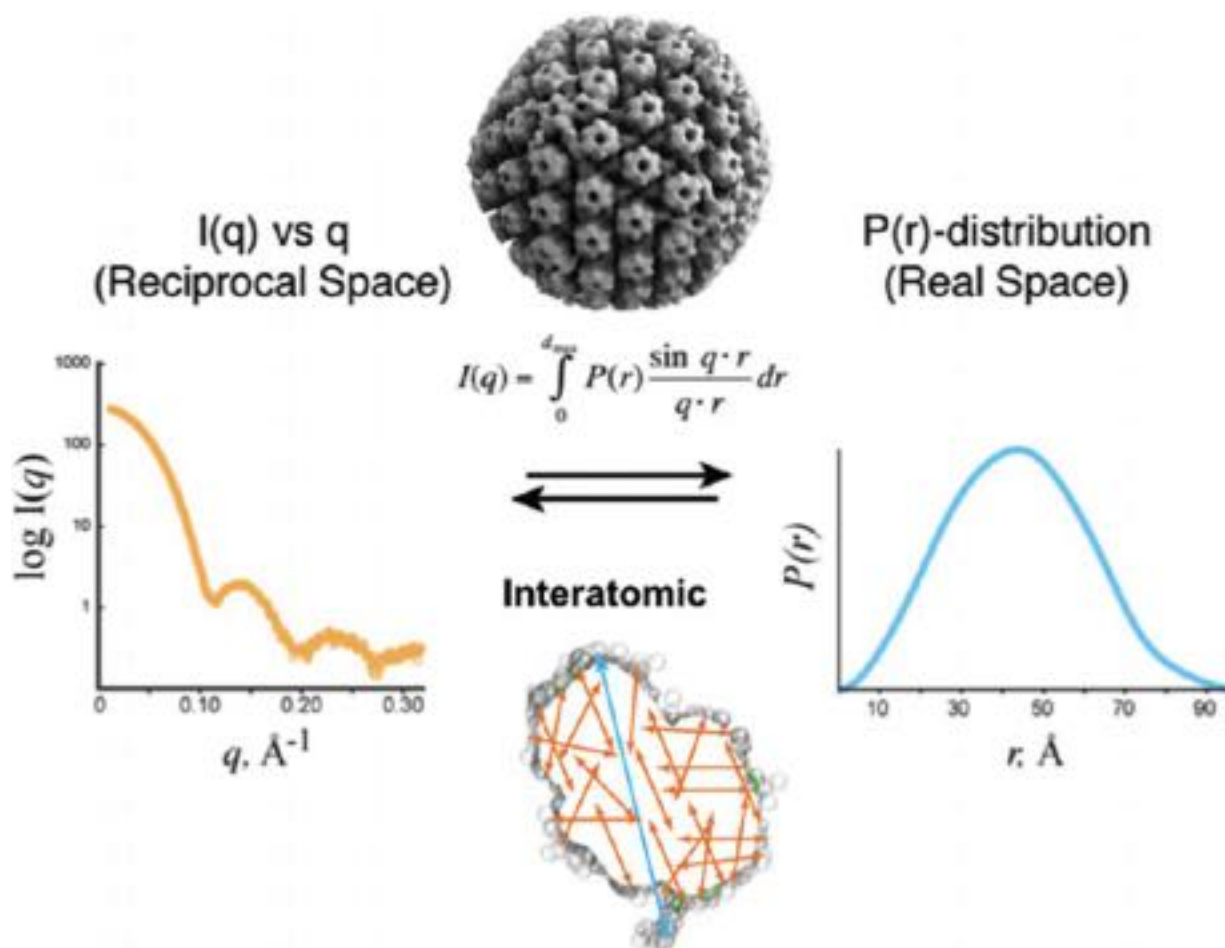


Fig. 2 SAXS basics. A measured SAXS curve (orange curve, left) is made in reciprocal space where the scattering vector, q , is in reciprocal nm or Å. SAXS is a resolution limited measurement meaning scattering at higher q values delivers higher resolution. At low resolution, SAXS of a particle can be approximated by its overall shape. The Fourier sine-integral transform of the SAXS curve produces the paired-distance, $P(r)$ -, distribution function (blue curve, right) measured in real-space units. The $P(r)$ -distribution is a histogram of all interatomic paired-distances within the particle. The $P(r)$ -distribution determines the maximum dimension, d_{max} , of the particle. The shape of the SAXS curve is determined by the shape of the scattering particle. Here, a nearly spherical shape will demonstrate oscillating features in the SAXS curve.

resolution-limited, implying SAXS data collected to higher resolution will support additional features within the distribution. Since a bioSAXS measurement is made from the ensemble of billions of molecules, it is more accurate to describe SAXS as a structural measurement of the ensemble and the $P(r)$ -distribution is a sum of all members in the ensemble.

Alternatively, form-factor fitting attempts to describe SAXS data using simplistic models such as spheres, cylinders, or ellipsoids (Fig. 3). These models assume homogeneous electron densities (ρ) but can be extended to heterogeneous electron densities using a core-shell approach. All viruses can be coarsely described as either a two-phase (protein and nucleic acid) or three-phase (protein, nucleic acid and lipid) particle. The nucleic acid phase (ρ_{core}) has a greater electron density than the protein (ρ_{shell}) and likewise, the protein has a greater electron density than lipid. Form-factor fitting of filled virions determines shell thickness (capsid coat), diameter of the virus and possibly, the thickness of the lipid envelope. These are all fitted parameters and can be used to follow overall structural changes in a virus (such as swelling) at different pH levels and salt concentrations. Furthermore, assuming that the electron density contrasts (ρ) for protein and solvent are constant, form-factor fittings allows for the quantitative determination of nucleic acid content within formed virions.

The core-shell formalism (Fig. 3) shows that the magnitude of the SAXS scattering due to each respective phase is dependent on the magnitude of the ρ difference. Since $\rho_{solvent}$ can be effectively modulated through increases in either salt or monovalent salt concentration, there will be a match point concentration where the effective contrast difference between $\rho_{solvent}$ and ρ_{shell} is zero. At the match point, the measured SAXS signal will only be due to the contents of the inner core (Fig. 3). If sucrose is used as the contrasting agent, the match point will be greater than 50% sucrose. High sucrose concentrations support many proteins and complexes in solution, however, reducing the contrast of the protein will also reduce the contrast of nucleic acid. This places a high demand on sample concentration to mitigate errors in background matching. For nucleic acid filled capsids, concentrations greater than 10 mg per ml can be expected.

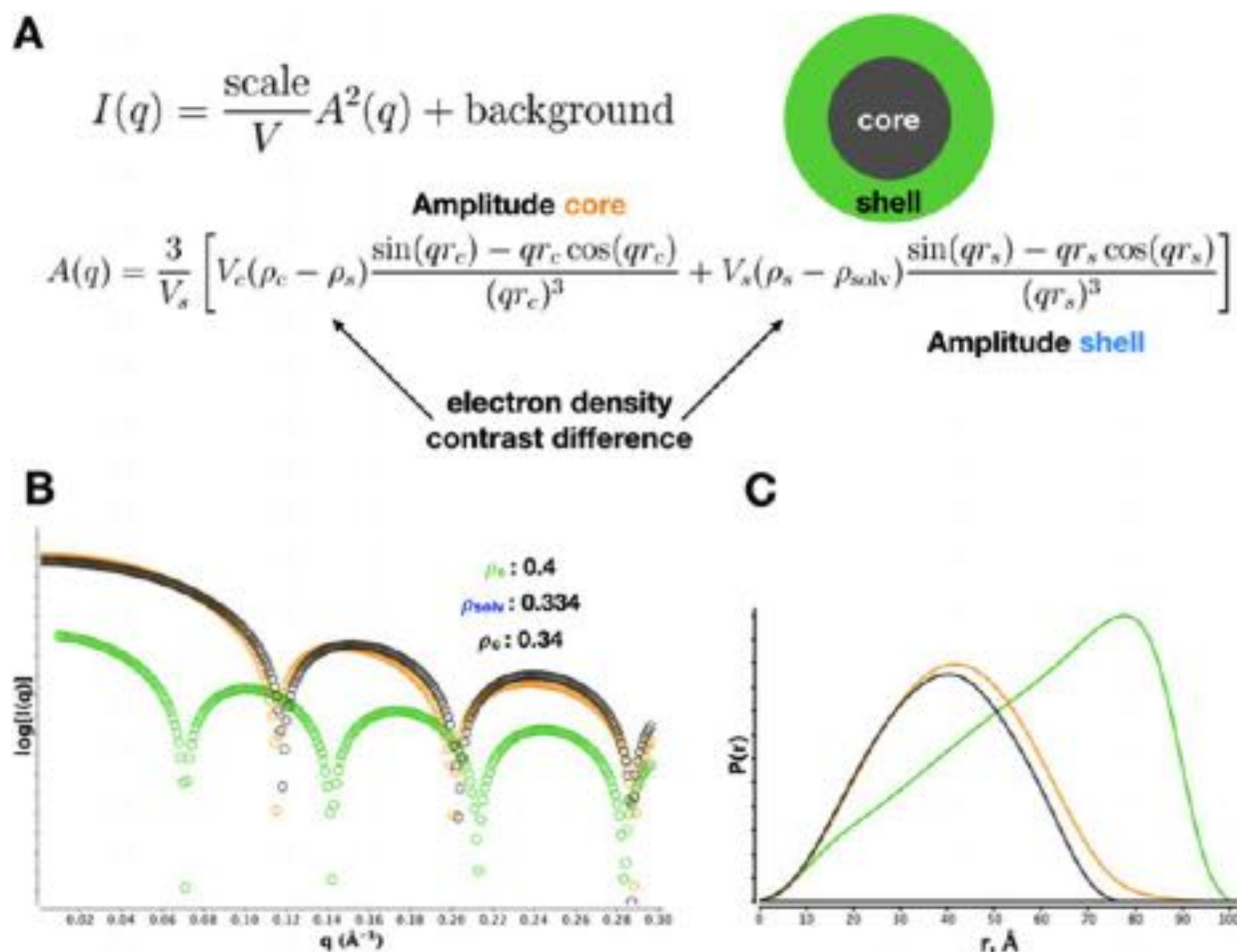


Fig. 3 Core-shell modeling. (A) Form factor fitting describes the scattering of the particle by amplitude function where the square of the amplitude is proportional to $I(q)$. A spherical core shell model is described by two spheres of different radii and electron density contrast (ρ_{shell} and ρ_{core}). The amplitude contributions of the core and shell are weighted by their respective volumes, V_c and V_s and effective contrast differences (ρ_{core} and ρ_{shell}) and (ρ_{shell} and ρ_{solvent}). (B) SAXS of a two phase particle with a d_{max} of 100 Å modeled with a shell thickness of 22 Å (orange). Contrast matching the shell (dark gray) causes a shift in the SAXS curve due to the apparent smaller diameter. (C) Real-space $P(r)$ -distribution demonstrates a reduction in d_{max} . An empty virus will have a core that has the same electron density as the solvent. This shifts the $P(r)$ -distribution to the right (green) as most of the particles' distances are contained within the shell.

Recent application of contrast variation to nucleic acid filled MS2 virions have allowed for the visualization of the RNA genome in intact virus showing an asymmetric, compact shape. Comparison of the RNA genome free of the capsid shows the viral RNA unpacks into a discrete extended structure that is not a random coil. It suggests the final packing of the RNA is facilitated by interactions with MS2 coat protein. Exploiting the contrast differences between protein and nucleic acids in complexes has been a popular application in SAXS where the difference allows for the separate modeling of each phase in *ab initio* reconstructions (Fig. 1(E) and (F)).

Time-Resolved SAXS

Studying virus particles by SAXS will typically require $\sim 10^{10}$ X-rays photons in a single exposure to provide meaningful low-resolution information ($q_{\text{max}} < 0.2 \text{ \AA}^{-1}$). For most synchrotron-based monochromatic X-ray sources, this implies the shortest possible time-scale that can be measured is around 1–10 ms using a stopped-flow kinetics sample environment. However, for studying kinetics of virus assembly, at least 10^{11} photons will be required to follow capsid nucleation events which may not be possible for short exposures of less than 10 ms. To access smaller time-scales ($< 1 \text{ ms}$) by SAXS using a monochromatic X-ray source, continuous mixing experiments can be performed where a reaction is initiated in a mixing chamber and flowed into the SAXS sample cell. Reaction time points are determined by the distance from the point of mixing and flow rates. Extended exposures can be performed if the flow rate is fixed, substantially increasing the signal-to-noise of the measurement (Fig. 4).

Time-resolved stopped flow experiments were performed examining the assembly of Hepatitis B capsids in the presence and absence of RNA. Exposures were limited to 5 ms however, the mixing time was much slower at 60 ms suggesting the earliest possible time point that can be measured is $\sim 65 \text{ ms}$ after mixing. This was sufficient to observe the condensation of the capsid proteins onto the RNA

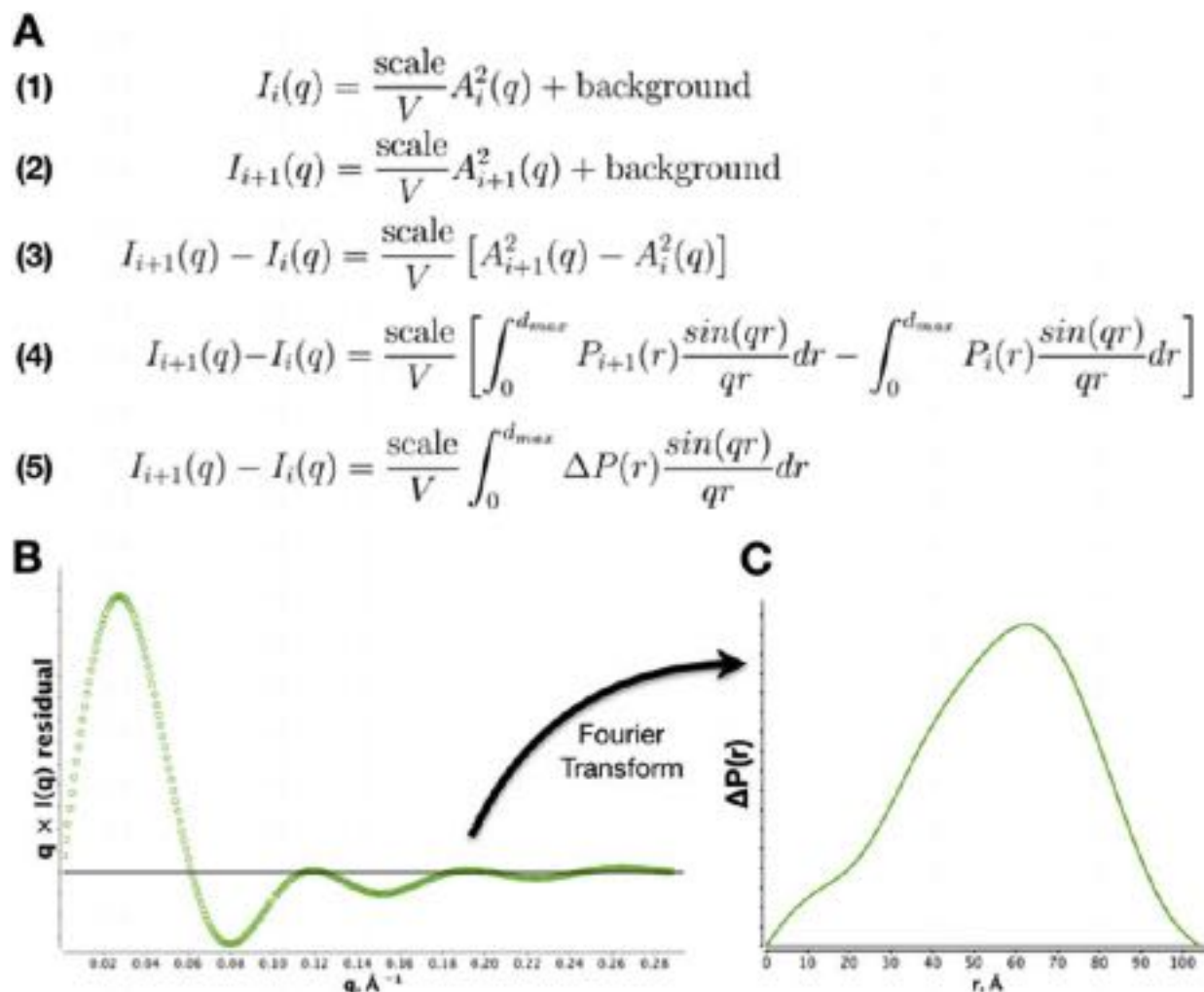


Fig. 4 Time-resolved SAXS analysis. (A) Each measured time point, i , corresponds to a SAXS frame that contains both the particle amplitude, $A(q)$, and corresponding background term. Given each subsequent measurement ($i + 1$) corresponds to a small change in time assuming negligible background differences, the effective difference between each time point measures the change in the particle amplitude or $P(r)$ -distribution. (B) Intensity difference curve modeled from swelling of a two-phase core-shell model (in Fig. 3). Protein phase was allowed to swell by 5 Å. (C) Difference curve transformed to real-space showing swelling shifts peak distances to right.

followed by isomerization to the mature capsid (Fig. 5). The SAXS camera covered the length scales from 14 to 785 Å during each measurement. During the time-resolved assembly of the virion, the contributions of each component (free RNA and protein, partially assembled RNA-protein complexes, and substates of the virion) to the SAXS signal is additive. At the earliest time point, the mixture will mostly contain free RNA, protein and partially assembled complexes whereas at the later time points, the SAXS signal will be due to the assembled virion. The signal deconstruction technique, multivariate curve resolution (MCR), can then be applied to the entire set of time-resolved measurements to determine a set of basis SAXS curves that describe the component species in solution. MCR helps reveal the folding pathway of the protein-RNA complex during assembly of the virion.

Diffracted X-Ray Tracking

Diffracted X-ray tracking (DXT) monitors the tilting and twisting motions of individual polymers to be followed within the microsecond to millisecond time scales (Fig. 6). For biological applications, DXT uses a small gold nanocrystal as a labeling probe that is covalently attached to the protein or nucleic acid through an accessible thiol group. The thiol group is often engineered through standard site-specific mutation methods. Under dilute conditions, the biological particles are subsequently adhered to a surface substrate such as Kapton. Each electron dense gold nanocrystal will diffract X-rays and by using a wide energy band-width X-ray source (Laue diffraction), any motion of the nanocrystal will consequently cause a movement in the observed diffracted X-ray spot.

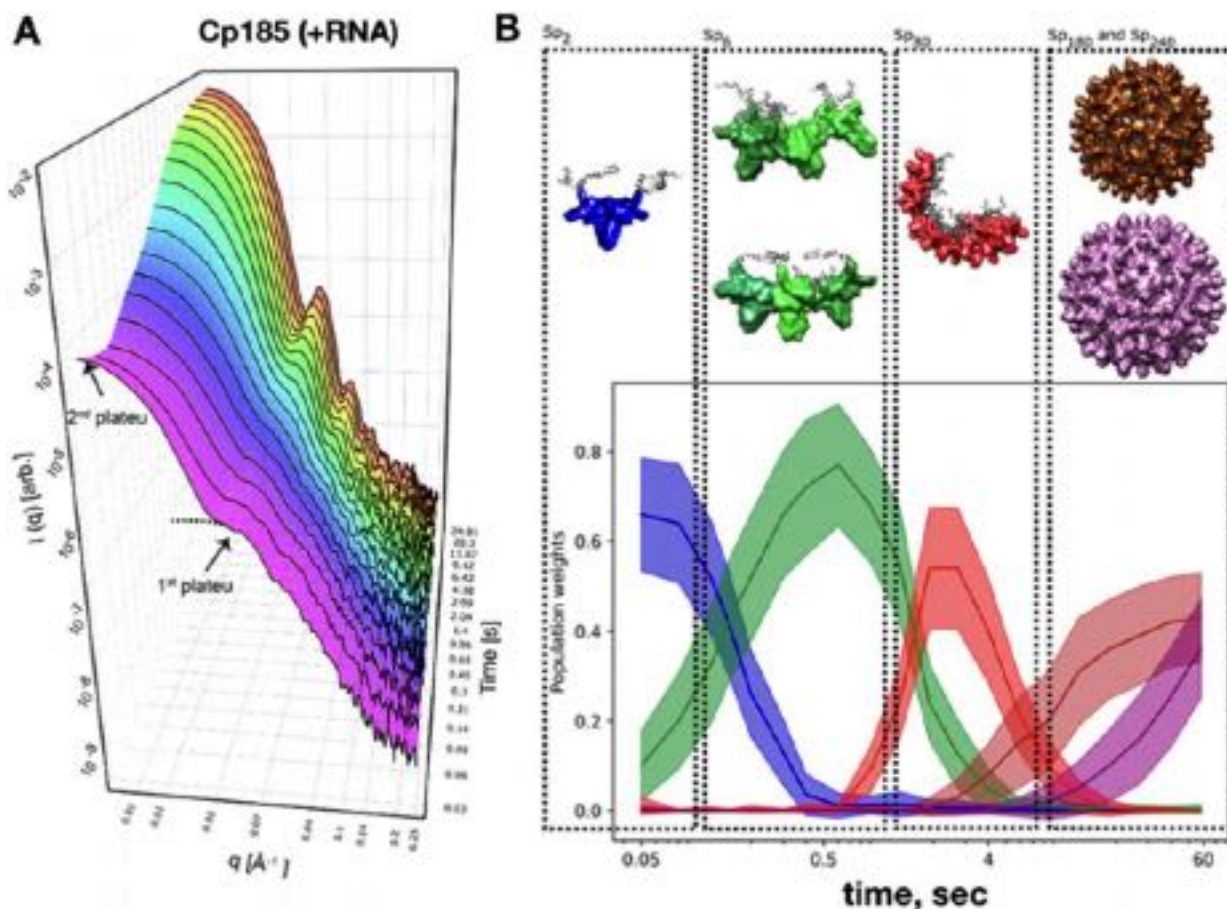


Fig. 5 Time-resolved SAXS of Hepatitis B virus capsid assembly. (A) SAXS curves from ~ 65 ms to 30 s following assembly of viral protein Cp185 with 40 nucleotide RNA. RNA contains preferred binding sites of Cp185. (B) Multivariate curve resolution of the time-resolved SAXS curves suggest 5 major species evolve through the assembly of the capsid. Initial stages suggest dimerization of proteins on the RNA followed by stable hexamer formation. The hexamers condense into a stable 80-mer complex followed by complex capsid formation. Images adapted from Oliver, R.C., Potrzebowski, W., Najibi, S.M., *et al.*, 2020. Assembly of capsids from hepatitis B virus core protein progresses through highly populated intermediates in the presence and absence of RNA. *ACS Nano* 14 (8), 10226–10238.

Each diffracted spot corresponds to a single protein molecule. Therefore, the movement of the spot is a single molecule event that traces a track in the set of recorded images using a high-speed X-ray camera (> 1 MHz). Each track is resolved into two time-resolved angular components χ (twisting) and θ (tilting) that describe the motion of the nanocrystal. DXT will record tens of thousands of tracks that form an empirical distribution of χ and θ . These distributions describe the non-random degrees-of-freedom accessible to the protein. The distributions can change in the presence of ligand, when bound to a cofactor or partner or due to a mutation.

DXT is a novel technique with considerable potential for virology, although to date it has been applied to only a handful of macromolecules. One example is the group II chaperonin, a large macromolecular machine that converts the energy in ATP to useful work. Time-resolved studies were initiated using caged-ATP that followed the mechanical cycle of the chaperonin. DXT demonstrated non-concerted motions between the subunits where ATP binding promotes twisting and closing of a subunit. The time-resolved, single-molecule studies proposed an asynchronized closing motion of the eight subunits challenging the synchronization model proposed by previous cryo-EM studies.

Furthermore, DXT has been applied to several membrane protein systems including bacterio rhodopsin (bR), KcsA potassium channel, nicotinic acetylcholine receptor (nAChR), major histocompatibility complex (MHC) and G-protein coupled receptors (GPCRs). Studies on nAChR in the presence and absence of ligand (ACh) revealed a fundamental dynamic in the gating mechanism of nAChR, which showed that the first motion related to the ligand (ACh) binding initiated vigorous molecular fluctuations in nAChR that subsequently triggered the gating mechanism. The nAChR channel was labeled through an engineered thiol group, however, in studies of the MHC complex, the recognition peptides were labeled with the gold nanocrystal. This allowed peptides to be flowed in such that dynamics of peptide binding could be followed in real-time.

DXT is a highly specialized technique that is gaining support. Sample quantities are 1 μ L at < 0.1 mg per ml and the method has not been applied to viral systems. However, DXT could be an invaluable scientific tool in virology. The single molecule tracking, small sample requirements and ease-of-use would afford microsecond time-scales to studying structural changes in

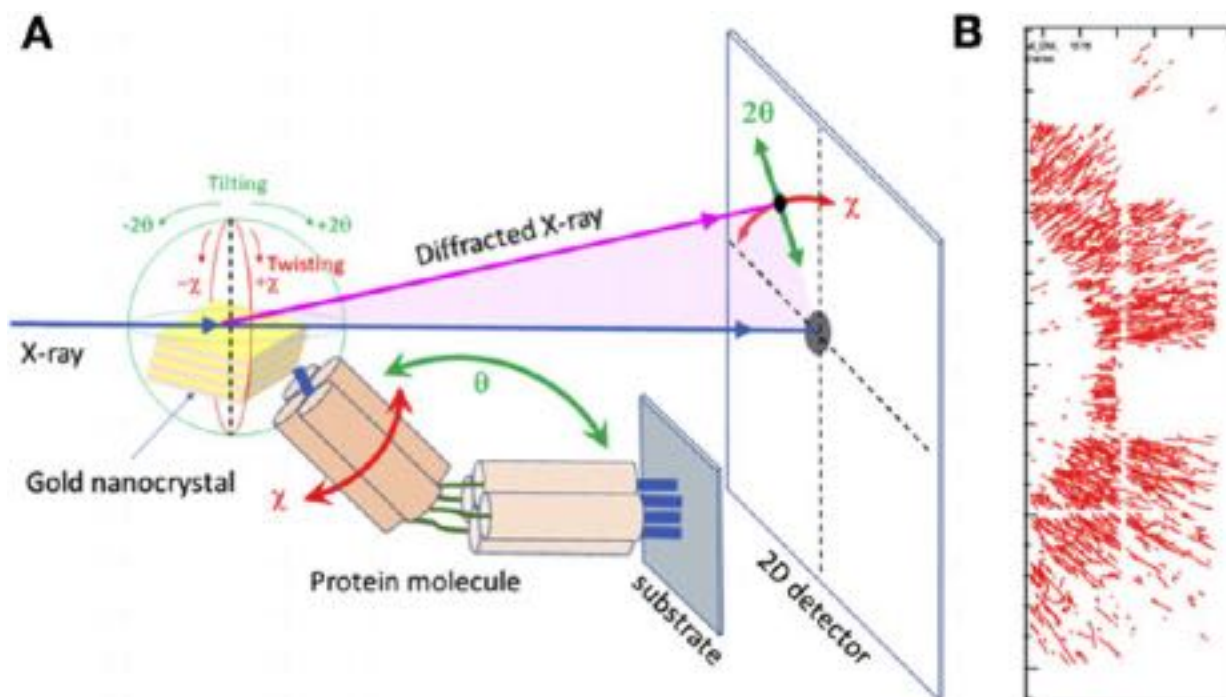


Fig. 6 *Diffracted X-ray Tracking (DXT).* (A) Gold nanocrystals coupled to proteins via thiol chemistry will wiggle in solution. The intrinsic motions of the protein are non-random and will move the gold nanocrystal in either a twisting (χ) or tilting (θ) motion. The [111] diffraction condition is detected using a broad band X-ray source and motions are recorded with a high-frame rate detector (> 100 kHz). Protein coupled motions of the crystal will be recorded as moving diffracted spots (tracks). (B) Each diffracted spot corresponds to a single-molecule and tens of thousands of tracks are recorded in a typical DXT experiment. Overlay of DXT measurement of coil-coiled protein showing all recorded tracks at beamline B16 using a Tristan 1 M detector. *Source:* Diamond Light, Didcot UK.

component proteins of the capsid, packaging of the genome, receptor interactions required for infection and a new assay for intrinsically disordered proteins. For example the pH initiated conformational changes of virus fusion machines could, for the first time, be directly studied using photo-release of caged protons as a trigger. Current research has also applied DXT to living cells where target GPCRs are transiently expressed in HEK 293 cells. The cells are grown on the Kapton support and used directly in the DXT measurement. The ability to monitor structural changes at biological relevant time-scales *in vivo* can help resolve or validate difficult hypotheses. In SAXS, flexibility is an inferred property and DXT has been used to directly test the flexible hypothesis of a coil-coiled protein and demonstrate stiffening due to a mutation which causes severe errors in meiosis.

X-Ray Footprinting Mass Spectrometry

X-ray footprinting is a technique that activates water molecules through radiolysis producing a highly reactive hydroxyl radical chemical probe. A synchrotron X-ray source is used to provide a broad-energy bandwidth at extremely high flux ($> 10^{15}$ photons $\cdot$ second $^{-1}$) in < 100 μ m-square spot. This induces rapid generation of hydroxyl radicals in the sample. The duration of the X-ray bursts controls concentration of hydroxyl radicals in the sample which is adjusted to minimize side reactions such that only the residues in the protein are reacting with the chemical probe. The detector for XFMS is a mass spectrometer and modified amino acid residues are subsequently identified by peptide fragmentation and mass spectrometry.

For each sample, a series of short exposures (< 10 ms) is performed to determine the hydroxyl reaction rates for the amino acids in the structure. Variation in the reaction rate is related to how buried or solvent accessible a residue is in the folded structure. More importantly, reaction rates can be followed as a function of treatment (ligand, pH, temperature). Functionally important residues in a binding reaction may become more buried showing a decrease in reaction rate. Likewise, some residues may become more solvent exposed demonstrating an increase in reaction rate. XFMS has been applied to a wide variety of systems including protein nucleic acid complexes, membrane proteins (GPCRs and ion channels) and *in vivo* mapping of the ribosome.

XFMS is a high information technique that not only identifies which residues are solvent exposed but also which atoms in the residue are reacting. This can be appreciated in a cryo-EM investigation of the SARS-CoV-2 spike protein that investigated the binding modes of various high affinity nanobodies. Not all nanobody-spike complexes were suitable for cryo-EM structure determination and XFMS allowed direct mapping of the residues involved in the interaction. The spike protein residues mapped far from the host-cell angiotensin converting enzyme (ACE) 2 binding site suggesting neutralization by the nanobody occurred through an indirect, non-competitive mechanism.

Concluding Remarks

Biophysical characterizations in the solution-state are critical to validating structural hypotheses derived from high resolution structure determination methods. As described, SAXS is a powerful tool that informs on the solution state of macromolecules. SAXS provides a comprehensive and composite picture of all the conformations in the ensemble. This allows SAXS to be used to complete partial high-resolution models, refine homology models, determine the structure of complexes and reliably test the structural hypothesis from MX and cryo-EM. More importantly, SAXS can reveal novel conformations hidden from high-resolution methods, essential for studying assembly/disassembly pathways of a virus. The information in a SAXS measurement is sufficient for modeling conformational changes when coupled with advanced computational methods such as molecular dynamics. More direct validation of dynamics can be achieved using DXT. DXT reveals, in real-time, the intrinsic motions of single molecules at microsecond time-scales. These observations show how a protein “wiggles” in solution allowing functional identification of intrinsically disordered regions. More importantly, DXT and XFMS can be performed on living cells providing opportunities for hypothesis testing *in vivo*. It is supposed that a virus will have fast and large morphological changes during cell entry with specific staged motions/conformations related to its life cycle. The significance of describing fundamental inter- and intra-particle interactions of a virus is obvious. In this context, approaches utilizing SAXS, DXT and XFMS can detect structural changes with high accuracy and precision.

Further Reading

- Adilakshmi, T., Soper, S.F.C., Woodson, S.A., 2009. Structural analysis of RNA in living cells by *in vivo* synchrotron X-ray footprinting. *Methods in Enzymology* 468, 239–258.
- Chevreuil, M., Law-Hine, D., Chen, J., *et al.*, 2018. Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte. *Nature Communications* 9, 3071–3080.
- Choi, K.H., Morais, M., 2014. Use of small-angle X-ray scattering to investigate the structure and function of dengue virus NS3 and NS5. *Methods in Molecular Biology* 1138, 241–252.
- Emeterio, J.S., Pollack, L., 2020. Visualizing a viral genome with contrast variation small angle x-ray scattering. *Journal of Biological Chemistry*.
- Gupta, S., Feng, J., Chan, L.J.G., Petzold, C.J., Ralston, C.Y., 2016. Synchrotron X-ray footprinting as a method to visualize water in proteins. *Journal of Synchrotron Radiation* 23, 1056–1069.
- Hiroshi Sekiguchi, H., Sasaki, Y.C., 2019. Dynamic 3D. visualization of active protein's motion using diffracted X-ray tracking. *Journal of Applied Physics* 58, 120501–120508.
- Jones, C.P., Cantara, W.A., Olson, E.D., Musier-Forsyth, K., 2014. Small-angle X-ray scattering-derived structure of the HIV-1 5' UTR reveals 3D tRNA mimicry. *Proceedings of the National Academy of Sciences of the United States of America* 111, 3395–3400.
- Josts, I., Gao, Y., Monteiro, D.C.F., *et al.*, 2020. Structural kinetics of MsbA investigated by stopped-flow time-resolved small-angle X-ray scattering. *Structure*. 28, 348–354.
- Khaykelson, D., Raviv, U., 2020. Studying viruses using solution X-ray scattering. *Biophysical Reviews* 24, 41–48.
- Kuzmanovica, D.A., Elashvili, I., O'Connell, C., Krueger, S., 2008. A novel application of small-angle scattering techniques: quality assurance testing of virus quantification technology. *Radiation Physics and Chemistry* 77, 215–224.
- Kuzmanovica, D.A., Elashvili, I., Wick, C., O'Connell, C., Krueger, S., 2006. Quantification of RNA in bacteriophage MS2-like viruses in solution by small-angle X-ray scattering. *Radiation Physics and Chemistry* 75, 359–368.
- Mason, A.C., Rambo, R.P., Greer, B., *et al.*, 2014. A structure-specific nucleic acid-binding domain conserved among DNA repair proteins. *Proceedings of the National Academy of Sciences of the United States of America* 111, 7618–7623.
- Matsui, T., Tsuruta, H., Johnson, J.E., 2010. Balanced electrostatic and structural forces guide the large conformational change associated with maturation of T = 4 virus. *Biophysical Journal* 98, 1337–1343.
- Mio, K., Ishihara, M., Fujimura, S., *et al.*, 2020. X-ray-based living-cell motion analysis of individual serotonin receptors. *Biochemical and Biophysical Research Communications* 529, 306–313.
- Pérard, J., Leyrat, C., Baudin, F., Drouet, E., Jamin, M., 2013. Structure of the full-length HCV IRES in solution. *Nature Communications* 4, 1612–1623.
- Rambo, R.P., Tainer, J.A., 2013. Super-resolution in solution X-ray scattering and its applications to structural systems biology. *Annual Review of Biophysics* 42, 414–441.
- Rambo, R.P., 2015. Resolving individual components in protein–RNA complexes using small-angle x-ray scattering experiments. *Methods in Enzymology* 558, 363–390.
- Ravishankar, H., Pedersen, M.N., Eklund, M., *et al.*, 2020. Tracking Ca²⁺ ATPase intermediates in real time by x-ray solution scattering. *Science Advances* 6, 1–10.
- Schoof, M., Faust, B., Saunders, R.A., *et al.*, 2020. An ultra-potent synthetic nanobody neutralizes SARS-CoV-2 by locking Spike into an inactive conformation. *BioRxiv*. doi:10.1101/2020.08.08.238469.
- Sekiguchi, H., Suzuki, Y., Nishino, Y., *et al.*, 2014. Real time ligand-induced motion mappings of AChBP and nAChR using X-ray single molecule tracking. *Scientific Reports* 4, 6384.
- Shimizu, H., Iwamoto, M., Konno, T., *et al.*, 2008. Global twisting motion of single molecular KcsA potassium channel upon gating. *Cell* 132, 67–78.
- Yabukarski, F., Lawrence, P., Tarbouriech, N., *et al.*, 2014. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nature Structural & Molecular Biology* 21, 754–759.
- Yamamoto, Y.Y., Uno, Y., Sha, E., *et al.*, 2017. Asymmetry in the function and dynamics of the cytosolic group II chaperonin CCT/TRiC. *PLoS One* 12, e0176054.
- Yao, H., Lee, M.W., Waring, A.J., Wong, G.C.L., Hong, M., 2015. Viral fusion protein transmembrane domain adopts β -strand structure to facilitate membrane topological changes for virus–cell fusion. *Proceedings of the National Academy of Sciences of the United States of America* 112, 10926–10931.

Virus Crystallography

Jonathan M Grimes, University of Oxford, Oxford, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Although the first X-ray crystallographic structures of small plant viruses were solved in the late 1970s for tomato bushy stunt virus (TBSV) and Southern bean Mosaic virus (SBMV), the history of virus crystallography extends back to the 1930s and 40s, with the work of Bernal and Fankuchen on the X-ray fiber diffraction analysis on concentrated solutions of tobacco mosaic virus and the work of Dorothy Hodgkin in 1945 on single crystals of tobacco necrosis virus. These early investigations established that many of these viruses were spherical and about 300 Å in diameter. This led to the realization that there was not enough capacity in the viral genomes to code for a single protein sufficiently large to protect the genome and there had to be multiple copies of a smaller capsid protein that would have to assemble to protect the cargo of viral RNA. As all the subunits would be identical, the prediction was that they would have to have identical environments, and this led to the hypothesis that virus capsids would be symmetrical. This was shown to be the case by Caspar in an X-ray study on TBSV which proved that the virus capsid had axes of icosahedral symmetry. Icosahedral particles have axes of 2-fold, 3-fold and 5-fold symmetry and this means that there are 60 copies of the icosahedral building block (or asymmetric unit) that are required for the construction of an icosahedral viral particle. Due to the physical requirements on viruses to enclose sufficient volume to house their large genomes, most viruses contain more than 60 protein subunits and thus the icosahedral asymmetric unit or building block contains more than one chemically identical subunit. Caspar and Klug proposed an extension to the rules of strict symmetry underpinning virus structure with the idea of quasi-equivalence where chemically identical subunits would adapt to similar but non-identical environments. In this model the triangular icosahedral asymmetric unit is further divided into n sub triangles, n being denoted the triangulation number (T number). Although the underlying chemical justification of quasi-equivalence is naive the theory has been extremely useful in interpreting low resolution structural results.

The structures of TBSV and SBMV revealed how 180 copies of the capsid protein had positioned themselves in quasi-equivalent environments – the three chemically identical subunits (A, B and C) were arranged around a quasi-threefold axis in a $T = 3$ structure. A further surprise was the realization that although there was no recognizable sequence similarity between the capsid proteins, they had very similar structures, known now as the “jelly-roll” fold. Not only were the capsid proteins very similar in structure but their architectural arrangement in forming the closed icosahedral capsid was identical, indicating that these two viruses diverged from a common ancestor and were evolutionarily related.

In the 1980s the structures of the first mammalian viruses were solved. These were of picornaviruses, a group of small animal viruses about 300 Å in diameter. Their viral capsids are composed of 60 copies of four viral proteins VP1, VP2, VP3 and VP4, unlike the plant viruses previously solved which have 180 copies of a single protein. The revelations from the structures of rhinovirus, polio virus and foot-and-mouth disease virus (FMDV) were that VP1, VP2 and VP3 all had the same jelly-roll fold and that the fold was identical to that of the capsid proteins of TBSV and SBMV, despite having no obvious amino acid sequence similarity (Fig. 1). Furthermore, these picornavirus proteins were organized in the viral capsid with the same architecture as the plant RNA viruses, forming a pseudo $T = 3$ (or $P = 3$) structure.

Since then the number of structures of viruses has grown exponentially, driven by the developments in the crystallographic methodologies, the technologies of X-ray production at synchrotrons, and of X-ray detectors and by the astronomical advances

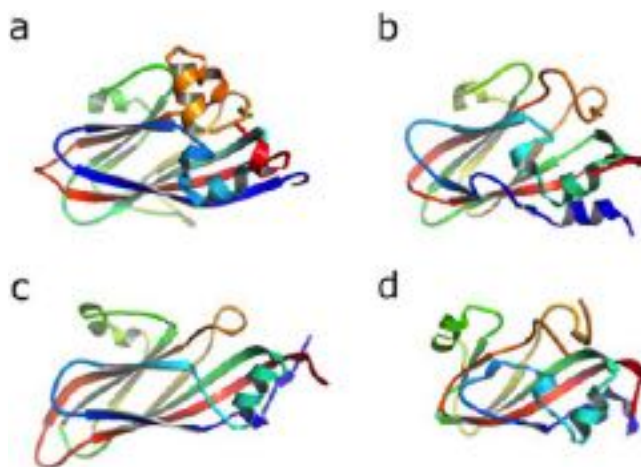


Fig. 1 Comparison of the jelly roll capsid proteins of (a) Southern bean mosaic virus (PDB ID 4sbv), and those of Foot-and-Mouth disease virus (5ne4) (b) VP1 (c) VP2 and (d) VP3 colored as a rainbow from the N-terminus to the C-terminus. N- and C-terminal extensions have been removed for clarity.

made in the speed and performance of computing. Although the fundamental principles and ideas behind the crystallographic analysis of a virus particle are no different from those for protein crystallography – the determination of the amplitude and phase of each diffracted X-ray – there is a significant scaling up required of many aspects of the process when compared to a standard crystallographic analysis of a protein. This is due mainly to size – large particles like viruses will crystallize in large unit cells, which means that, for an analysis at a fixed resolution, there are many thousands more reflections that need to be recorded and the density of these reflections on the detector will be equally high. This problem is compounded by the issue that the intensity of these reflections will be many thousand times weaker on average. The weakness of each reflection and the high density and large number of reflections place demands not only on the X-ray sources and detectors but on computer hardware and software.

The seeming structural complexity, might seem problematic in a crystallographic analysis, but the icosahedral symmetry present in virus capsids is of enormous importance providing extremely powerful constraints that facilitate many aspects of the crystallographic analysis (which will be discussed later). In crystals of icosahedral viruses there will be at least 5-fold non-crystallographic symmetry, which gives rise to a 5-fold redundancy in the data. This is because mathematically, a 5-fold axis cannot be found throughout a repeating crystal lattice but can only be present as local symmetry.

Many of the ideas in modern macromolecular crystallography, have been developed and driven by crystallographers having to optimize methods for data collection, by maximizing the signal to noise in the data, as well as developing new software and algorithms as part of their standard toolkit, when confronted by the challenges of the structural analysis of viruses. These include the development of the molecular replacement method and the use of self-rotation functions to determine the orientation of viruses in the crystal unit cell. The fixed relationship between the various icosahedral symmetry axes and the non-crystallographic 5-fold axis means that it is often possible to solve the orientation problem by simple analysis of the diffraction data, in the absence of a model, by use of the self-rotation function. Additionally, the development of the use of non-crystallographic averaging in real space, as well as the use of constraints in the refinement of the model in reciprocal space, was driven largely by the exploitation of the non-crystallographic redundancy that the viral icosahedral symmetry provided. Remarkably Rossmann *et al.* showed that, starting from low resolution experimental phases, high levels of non-crystallographic symmetry provided, through averaging, enough redundancy to allow phase extension, yielding a detailed high resolution map. In the case of FMDV Stuart *et al.* further enhanced the method and were able to obtain an excellent high resolution map without experimental phase information, with minimal (5-fold) non-crystallographic symmetry.

The structural analysis of viruses also led to the development of hybrid methods, particularly the fitting of high-resolution structures of proteins derived by crystallography into lower resolution cryo-EM maps, to generate models, richer in information content, for use in molecular replacement but also that could be tested in functional assays. An example is the structure determination of the Bluetongue virus (BTV) core using a molecular replacement hybrid model generated by the fitting of 260 trimers of the crystal structure of BTV VP7 into a 22 Å cryo-EM map of the BTV core.

Crystallography has also played a fundamental role in structure-based drug discovery (SBDD), as virus enzyme structures allowed the identification of potential inhibitor-binding sites and the optimization of hit molecules against the target structure. During the past decades there have been numerous successful drugs designed using SBDDs against both viral and cancer targets.

One of first landmark examples of SBDD was in the design of pocket factor drugs against human rhinoviruses (HRV). Elucidation of the structure of HRV revealed a canyon formed at the interface of VP1 and VP3. Beneath this canyon, within VP1, there is a hydrophobic pocket that is occupied by a molecule proposed to be a fatty acid. It was shown that this pocket was the likely binding site for a drug previously shown to inhibit picornavirus infections and identification of this site led to the development of anti-virals in collaboration with Sterling Winthrop Pharmaceuticals. Although these anti-virals were highly potent they were not licensed by the FDA due to significant side effects in pregnant women.

In the first few months of the COVID-19 crisis a huge effort has been made across the globe to solve the structures of SARS-CoV2 proteins and use SBDD approaches to develop both anti-viral drugs against validated drug targets, and understand the role of neutralizing antibodies by solving the structure of the F fusion protein complexed to Fab fragments. Although single-particle-analysis (SPA) cryo-EM is now playing a more substantive role with some 109 structures deposited in the PDB database, over 370 structures have been solved by X-ray crystallography, and even for the receptor binding domain of the fusion F spike close to 50% of the structures solved in complex with neutralizing Fabs or nanobodies have been done by crystallography (See “Relevant Websites section”). Alongside SBDD, whole capsid crystallography and the development of recombinant methods for producing virus-like-particles (VLPs) have allowed a more rational basis for the design of vaccines (structural vaccinology), based on improvements in capsid stability and optimization of antigenicity.

The Nitty Gritty: How do you do Virus Crystallography

The basic route to solving a crystal structure is (1) growth of single crystals that diffract to suitable resolution, (2) collection of diffraction intensities from single crystals to give a data set of structure factor amplitudes, (3) initial phase determination, and (4) phase refinement in real and reciprocal space to allow the generation of an atomic model, with minimal errors. Virus crystallography is somewhat more complex but identical in its essentials to protein crystallography. This complexity reflects the large size of even the smallest viruses compared to proteins, which means that there are differences in experimental approaches to deal with the issue of scale. Current methodologies that have been implemented in virus crystallography will be discussed in the following sections:

Targets, Virus Families and Predisposition to Crystallization (Biosafety Levels 2 and 3)

Viruses come in all shapes and sizes – from the simplest virions such as porcine circoviruses with a simple genome coding for the viral replicase and the capsid protein for making the icosahedral viral shell, to the most complex viruses that are larger than the smallest cells; for example, mimiviruses have a virion diameter of about 0.65 μm (some 400 times the size of circoviruses) with a genome encoding >900 genes. Unsurprisingly many of the best studied viruses that directly affect human health and prosperity, are animal and human pathogens, such as human immunodeficiency virus (HIV), measles and now Coronaviruses, but the vast majority of viruses in the biosphere are as yet ill-defined and infect microbes.

At its simplest a virus consists of a genome, which can be RNA or DNA, encapsulated and protected by a protein or proteolipid shell. Viruses are extremely genetically efficient and the use of symmetry, almost invariably icosahedral, helical, or a combination of the two, underpins the structure of many viruses. In viruses that are pleomorphic (with a lipid envelope), there is usually some structural component of the virus, like the matrix protein and the nucleocapsid protein of Influenza virus, that uses multiple copies of a symmetrically arranged protein. Although lipid containing viruses, where the icosahedral protein shell sits on a labile lipid bilayer such as PRD1 and PM2, have been studied by crystallography, the only viruses that have yielded routinely to high resolution single crystal analysis, are viruses that are isometric with highly ordered protein capsids.

Crystallization and Automation (Additives and Aids)

The crystallization of viruses is not any different from that of proteins – in fact given their isometric highly symmetric and ordered capsids one could argue that their tendency to form crystals is unsurprising. All of the caveats and rules surrounding protein crystallization apply to virus crystallization; for a start highly pure, concentrated solutions of structurally homogeneous material are required. One of the reasons why the plant viruses were the earliest viruses studied structurally was that grams of virus could be purified easily from leaf material. It is now routine to use ultracentrifugation with CsCl or sucrose gradients to purify structurally homogeneous capsid particles from genetically homogeneous viruses grown in tissue culture. This purified virus material can then be concentrated up using 100-kDa microconcentrators. The concentrations appropriate for crystallization are usually between 2 and 10 mg/ml, although nowadays a pre-crystallization test (PCT, Hampton) is done to check if the concentration is suitable for crystallization. Over the last twenty years developments in liquid dispensing robotics, driven by structural genomics, have revolutionized automated high-throughput crystallization procedures. Crystallizations are now set-up in 96-well crystallization plates using standard commercial crystallization screening kits (Hampton Research, CA, USA and Emerald Bio-Structures, deCode Genetics, WA, USA), where 100 nl of sample is mixed with 100 nl of precipitant solution.

The reduction of the protein and reservoir solution volumes required for screening crystallization conditions, coupled with improvements in efficiency and reproducibility has effectively replaced hand-pipetting methods for crystallization screening and optimization. The reduction in protein/virus volumes required by a factor of 10 has meant that seemingly intractable targets for crystallization, due to difficulties in producing sufficient purified material, are now tractable. One of the well-known issues that can prevent the growth of well diffracting single crystals is aggregation, sometimes caused by particle instability. If aggregation proves to be a problem, it may be relieved by the addition of detergents, whilst particle stability can be addressed by the use of additives like divalent cations known to interact with the capsid, or changing the pH of the crystallization solution, if the virus is known to be pH sensitive. For viruses, such as ones with an internal lipid bilayer, that are highly fragile and which have a tendency to fall apart when purified, cross-linkers such as glutaraldehyde have been used to stabilize the virus covalently linking proteins prior to crystallization. Even in these problematic cases, for viruses like PRD1 and PM2, it was then necessary to grow the crystals in quartz capillaries to eliminate the mechanical handling of the crystals during data collection. The final step to produce well diffracting crystals was to then dehydrate the crystals in the capillaries with high-molecular-weight PEG, effectively reannealing the crystals in-situ.

Crystal Handling and Cryoprotection (Notes for Enveloped and Non-Enveloped Viruses)

Whilst it is perhaps not surprising that robust spherical virus capsids crystallize, it is astonishing that crystals held together by only a few noncovalent interactions, are so beautifully ordered. Virus crystals tend to be more fragile than protein crystals, due to the relatively smaller surface area to mass ratio of the interfaces that hold the crystal together. Whereas protein crystals are cryogenically cooled to liquid nitrogen temperatures prior to data collection, the cooling of virus crystals is still uncommon. The major problem is that the mosaic spread of crystals tends to increase on cooling, meaning that the spot size of the recorded reflections becomes larger, frequently resulting in the overlap of reflections. The traditional approach for mounting virus crystals was to use the method of sealing hydrated crystals in capillaries – this was the standard method for protein crystallography prior to the now routine method of cryogenically cooling of crystals in loops. Due to disease security issues this method also had the advantage of being suitable for crystals of highly contagious viruses, like FMDV.

An implication of room temperature (RT) data collection is that a large number of crystals are often required to collect sufficient data to solve a virus structure. Thus, for challenging projects such as bluetongue virus (BTV) more than 1000 crystals were needed to solve the structure.

Data Collection (In-House and Synchrotron) and Automation

It was commonplace for previous generations of macromolecular crystallographers to have their own in-house X-ray sources for data collection. For virus data collection this was a laborious process requiring the careful focussing of the beam and many hours of exposure of the crystal to collect a single X-ray diffraction image (constituting a tiny fraction of a complete data set). However, one of the fundamental drivers that has powered the revolutionary advances in macromolecular crystallography over the last 30 years has been the availability of synchrotron radiation sources to macromolecular crystallographers. In the 1980s access to second generation synchrotrons, like Daresbury (UK), greatly speeded up and improved the process of data collection, revolutionizing how diffraction data were collected not only for viruses but also for all crystals of macromolecules by bringing the streets of Baltimore to the synchrotron with the introduction of the so-called “American Method” of data collection (shoot first, ask questions later), now rebadged as serial crystallography. Historically, as part of the data collection process, the crystal was set into a known orientation relative to the detector axes by a series of short X-ray exposures. Virus crystals are highly sensitive to intense synchrotron X-ray beams and so are normally only aligned optically before data collection, meaning that the auto-indexing algorithms that are now routine had to be developed to analyze the diffraction pattern. Further gains in crystal lifetime and signal-to-noise ratio were achieved by tuning the wavelength to below 1 Å (compared to 1.5 Å from in-house sources). Analysis by Fry *et al.* 1999 demonstrated that this method of data collection, where crystals are randomly orientated in the beam, is surprisingly efficient, and for high symmetry space groups like I23, a 64% complete data set can be collected from only 8° of data.

Historically in the structure analysis of small picornaviruses, like human rhinovirus (HRV) or FMDV, for a 3 Å resolution data set, the crystal would be oscillated in the beam for 0.3–0.5°, for several seconds, depending on the lattice properties and the image would be recorded on X-ray film, requiring significant manual intervention in “dark rooms” to develop the films, prior to scanning and image rasterization.

Further gains were made with the development of third-generation sources such as the European Synchrotron Radiation Facility (ESRF) in France and Diamond Light Source in the UK. Such synchrotrons provide highly brilliant, tuneable, and parallel X-ray beams that allow the collection of densely-packed but weak diffraction data – X-ray beams that are ideally suited optically to the large unit cells of virus crystals. As an aside it is interesting to note that the power of synchrotron beams fueled the explosion in cryo-crystallography which allowed the hundred-fold increase in crystal lifetime to be fully exploited in modern macromolecular crystallography.

These highly intense beams have meant that exposure times have been reduced by at least an order of magnitude. This in turn has driven the development of faster X-ray detectors, such as the Dectris pixel-array detectors which match the speed of data collection and this increase in speed has in turn driven the development of automation and the remote control of beamlines. For cryo-crystallography of macromolecules these developments have brought the time to collect a complete 360° data set down to a few seconds.

For room temperature analysis bottle necks remain in the logistics of presenting crystals to the X-ray beam. Over the past ten years, the installation of high-precision goniometers at beamlines I03 and I24 at Diamond Light Source capable of holding crystallization plates and presenting each of the crystallization drops to the X-ray beam, has allowed the rapid collection of in-situ room temperature data (Fig. 2). Collecting diffraction data in-situ eliminates the need to mount virus crystals in capillaries, a process that can often result in physical damage to the crystals, and maintains the integrity of virus crystals which are often very fragile. An additional consideration in virus crystallography is that in-situ analysis in sealed plates, as opposed to collecting data from crystals in open loops with the inherent issues of disease-security, provides a safe alternative for these pathogenic samples.

As an example of the latest state of the art, the structures of bovine enterovirus 2 (BEV2), enterovirus 71 (EV71) and FMDV virus like particles (VLPs) have all been solved on I24, from crystals grown in CrystalQuick X plates. These plates reduce the background by around a factor of 2.5, compared to Greiner CrystalQuick SW plates. Data were collected on a Pilatus 6MF detector operating in shutterless mode at 294K, the ambient temperature of the beamline, with individual frames of either 0.1° or 0.05° and exposure times of typically 0.1 s or less, enough to ensure the data were of sufficient strength for relatively routine processing. Under the regime described here the crystal life-time was typically 0.4 s. In the example of BEV2, a beam size of 20 × 20 µm allowed 76 exposed positions from 26 crystals, resulting in a 66% complete data set to 2.1 Å (Fig. 2).

Many synchrotrons now have automated pipelines for data analysis and data processing is now standard for crystals with large unit cells. To those who have not collected and analysed RT data, there will be a large number of partially recorded reflections that cannot be summed with their complementary parts on successive images (unlike cryo-data collection) and for these data the knowledge of the fraction of each partial reflection recorded allows the observed intensity to be scaled up to the value it would be if the reflection were fully recorded.

Structure Solution and Model Building

Once diffraction data have been gathered the next stage in the structure analysis breaks into two stages, initial phase determination and then subsequent phase refinement and extension.

Phases estimation requires defining the orientation and the position of the virus particle(s) in the crystal cell. This information is essential for phase refinement. Because of the inevitable non-crystallographic redundancy (a minimum of 5-fold) and the fixed relationship between the various icosahedral symmetry axes, it is often possible to solve the orientation problem by analysis of the diffraction data in the absence of a model, usually by use of a self-rotation function.

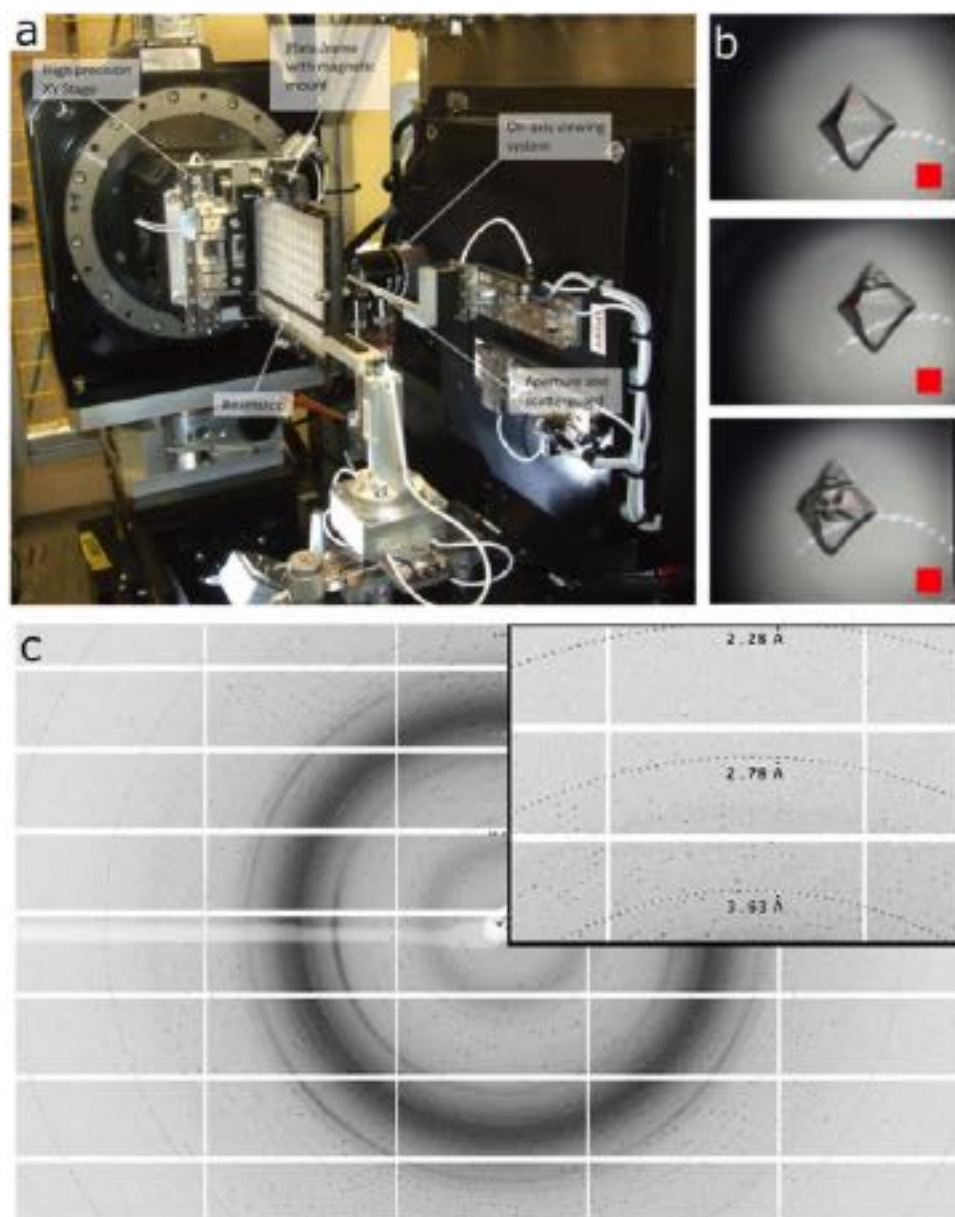


Fig. 2 Data collection on virus crystals. (a) The sample environment at beamline I24 for in-situ X-ray diffraction. (b) Images of a crystal of BEV during an X-ray diffraction experiment using a 20 μm beam size. (c) Typical diffraction pattern from a crystal of BEV with an exposure of 0.1 s and an oscillation on 0.1°.

Phases need to be derived for each of the measured reflections before electron density maps can be calculated, and these are usually generated by molecular replacement with a model structure, perhaps derived by cryo-EM. The non-crystallographic symmetry provides powerful phase constraints so that even a poor model, with poor phase estimates can be improved dramatically by the cyclic imposition of icosahedral symmetry. The procedure is relatively straightforward. The non-crystallographically related portions of the electron density map of the virus are averaged, and the regions not defined as virus are flattened to an average electron density value. This effectively cleans up the map, which is then back-transformed to generate improved calculated amplitudes and phases. The calculated structure factors are then recombined, suitably weighted, with the observed amplitudes and a new map is calculated. The procedure is then iterated to convergence. If the initial model is particularly inaccurate it may be necessary to solve the structure at low resolution, and then use the constraints in real space to extend and refine the phases to the resolution of the data.

If phases are to be extended, the map is back-transformed to a slightly higher resolution than it was calculated at, which provides phase estimates (albeit crude) to this new limit, which can then be fed back into the cyclical procedure. There are features in the process that are not standard and require thought, such as adding in calculated data for missing observed

amplitudes, and careful analysis of the averaging and solvent masks, which will need to be recalculated at intervals during the phase refinement and extension.

Once the real space phase refinement has converged, the map should be of sufficient quality that the polypeptide chain can be readily interpreted and a model rapidly build for the icosahedral asymmetric unit. Standard crystallographic refinement packages can then be used to refine the model against the data, normally imposing strict non-crystallographic constraints to maximize the observation to parameter ratio.

Understanding Evolution through Structural Anatomy

The structural analysis of the first plant and mammalian viruses demonstrated that the capsid protein fold and arrangement of the capsid subunits was conserved between the viruses although they belong to different families and infect different hosts. These striking similarities were assumed to signify descent from a common ancestor, but it was acknowledged at the time that it may reflect convergence, from separate ancestors, to a suitable stable capsid. Although comparison of genome and protein sequences allows the comparison of closely related viruses this approach becomes meaningless as sequences diverge. The observation that structure changes more slowly than sequence (due to the constraints of chemistry) provided a basis for comparison of seemingly unrelated viruses.

The principle when comparing structure, as with classical anatomy, is that similarity in structure can be taken as evidence of homology. Features such as atomic coordinates, chain direction, secondary structure similarity, topological similarity and loop length similarity are built into the comparison such that the argument for homology becomes compelling.

As the number of virus structures solved increased, it became apparent that structure could be used to uncover evolutionary relationships between seemingly diverse viruses, and that these viruses had an ancient history, spread across the 3 kingdoms of life.

A key observation was made by Burnett *et al.*, when they solved the PRD1 major coat protein structure and showed that its closest structural relative was the coat protein of adenovirus. This work revealed that the major viral capsid protein was a trimer with a common coat fold (the so-called double-barrel) arranged in hexagonal capsomer structures which provided direct evidence linking the bacterial virus (PRD1 infecting gram-negative bacterial hosts) to a human pathogen (the dsDNA adenovirus).

Two examples which provide further conclusive evidence for this are the capsid protein of sulfolobus turreted icosahedral virus (STIV), an archaeal virus, which was shown to have the same fold and architecture as the capsid proteins of PRD1 and adenovirus. In addition, the human pathogen Herpes simplex virus was shown to be structurally related to the bacteriophages HK97, P22 and T4, both in terms of capsid architecture but also in the molecular portals that package the genome into the preformed viral capsid.

Since then it has become widely accepted that the comparison of coat protein structure and virion architecture can provide fundamental insights into the evolutionary grouping of viruses previously considered to be unrelated, although further work is needed to improve the statistical basis and the robustness of the computational procedures. It is clear that icosahedral viruses fall into a small number of lineages, the members of which infect all 3 domains of life, but for enveloped viruses without a defined capsid protein, it is unclear what aspects retain an evolutionary signature.

Synergies

Our understanding of the biology of viruses through structure is best addressed by a range of different techniques that reveal both the static and dynamic aspects of the virus and on different scales.

The resolution revolution in cryo-EM now means that the one can obtain atomic information by SPA of flexible assemblies but also now, with cryo-electron tomography and in-situ imaging, structural details of viruses and viral assemblies within cells. With the developments in FIB milling, to produce thin lamella of a suitable thickness to allow electrons to penetrate through the sample, it is feasible to get to sub 10 Å resolution, where atomic models derived by crystallography and SPA cryo-EM, can be refined as rigid bodies into cellular maps. A recent beautiful example of this is the structure of an assembly intermediate of a mammalian dsRNA orthoreovirus, solved at a resolution of 5.6 Å, by imaging of cryo-preserved infected cells, which revealed a striking similarity with an assembly intermediate of a dsRNA bacteriophage, suggesting a common mechanism of viral assembly.

Alongside EM, soft X-ray imaging when combined with cryo-structured illumination microscopy (cryo-SIM) provides a powerful technique to image whole cells (unlike EM), due to the penetrative quality of X-rays, and the potential to correlate cellular remodeling that occurs in response to viral infection, with fluorescently labeled viruses or other biomarkers. These techniques combined with live cell imaging can provide orthogonality to the molecular detail by providing temporal information on the biology of viruses.

At the more extreme end of what is now possible in diffraction, serial femtosecond crystallography (SFX) at an X-ray free-electron laser (XFEL) now provides the opportunity, due to the extreme brightness of the X-rays, to image viral crystalline arrays within cells as well as the possibilities of capturing stages in viral disassembly.

Vaccines, Anti-Virals and Therapeutics

Structural biology has played an increasing role in the development of vaccines and anti-virals. Conventional vaccines have had a radical impact in improving human health, and are one of the most effective tools for the prevention of infectious diseases. The

discovery 200 years ago that vaccinia virus could be used to protect people from small pox, has led to the eventual eradication of the disease, and the development of vaccines that protect against polio and measles infections.

The current vaccines on the market are based largely on inactivated or attenuated live virus, with some recent developments in virus-vector vaccines. Whilst these live attenuated or inactivated vaccines have been successful in protecting people from infectious diseases there are significant drawbacks such as the cost and risks associated with the production of large volumes of highly pathogenic viruses under high biosafety containment along with the significant risk of attenuated forms reverting back to pathogenic forms. Traditional vaccine development is a long complex process and can take 10–15 years before the vaccine is approved for use. The Ebola outbreak in 2014–2016 in West Africa and the COVID 19 pandemic of 2020 highlighted the need of different approaches where targeted vaccines can be rapidly produced in sufficient quantities.

The large amount of knowledge now available on the three-dimensional structure of viral capsids and viral glycoproteins provides tremendous opportunities for structure-based vaccine design, based on the rational engineering of the antigenic and biophysical properties. The adaptive and innate immune systems have evolved to recognize highly repetitive structures, such as viral capsids and bacterial flagella rather than soluble antigen, and it is apparent that capsid integrity, and stabilization of antigenic forms of viral proteins is key to generating a protective immune response.

Two approaches to vaccine design that have benefitted from structural data are of viral vectors and the use of viral-like-particles (lacking any genomic material). Viral vectors are a non-pathogenic or non-replicating virus which is used as a vector to deliver the gene, encoding the viral antigen, into the host cell, where the antigen is expressed and induces an immune response. A very current example of this is ChAdOx1, currently under trial as a vaccine against SARS-CoV-2 which is a harmless chimpanzee adenovirus encoding the gene for the surface spike protein of SARS-CoV-2. This spike protein is expressed on cells in the immunized patient and the hope is that the immune response generated against this antigen is sufficient for protection. One problem that can occur is that the presenting antigen may not be sufficiently immunogenic. An example of this is the vaccine development against Respiratory syncytial virus (RSV), which is an enveloped RNA virus second only to malaria as a cause of infant mortality worldwide. Antibodies against RSV from infected patients mainly target the trimeric F surface glycoprotein of RSV which drives the membrane fusion that occurs on viral infection. However, F is metastable and rapidly converts to its post-fusion form which induces the production of antibodies that offer less protection against the virus. The structure of RSV F allowed the rational design of mutations that stabilised F in its prefusion form. Initial experiments using the stabilised F as a vaccine showed that it elicits antibodies that target prefusion epitopes. Recently the design of self-assembling nano-particles, where the stabilised F proteins are displayed on an icosahedral scaffold, has been shown to induce a 10-fold higher response of neutralizing antibodies compared to the protein alone.

An example of the development of VLPs in vaccine design is FMDV. FMDV is a highly contagious disease of cattle and sheep and the cost of outbreaks can be huge due to the requirement to slaughter infected animals. In the outbreak of 2001 over 6 million animals were slaughtered at a cost in excess of £8 billion. Current vaccines for FMDV are based on inactivated virus, and vaccine production is expensive and technically challenging. The vaccines are thermally unstable, and the capsids rapidly breakdown into pentamers which are not immunogenic. FMDV vaccines thus require a cold chain due to the warm climate in those areas of the world where they are used. There are also difficulties in distinguishing vaccinated animals from infected animals which has significant economic implications, due to restrictions that are imposed on the movement of infected animals. Development of novel vaccines for FMDV has been based on the rational introduction of mutations to increase thermostability. For VLPs a disulfide across the icosahedral 2-fold axis between adjacent pentamers effectively locks the VLP capsid, whilst for FMDV O and SAT2 serotype viruses which are unstable vaccines, the introduction of a hydrophobic residue at the same position conferred increased thermal stability, and retained immunogenicity over long term storage unlike wild type inactivated virus which rapidly becomes non-protective when stored (Fig. 3). Recent work on enteroviruses has revealed a positively-charged surface depression at a protomer interface on the capsid surface that is bound by glutathione known to be essential for the assembly of many enteroviruses. This depression has been identified as potentially druggable and amenable to SBDD (Fig. 3).

Whilst this review has focussed on whole capsid crystallography, there have been numerous successes in the rational design of anti-viral inhibitors targeting functionally important proteins of significant human pathogens such as HIV, Hepatitis C and Influenza virus. The rationale being that if you can design small molecules that inhibit these conserved mechanisms then the viral lifecycle is disrupted. Some of the earliest successful examples of SBDD were antiviral drugs that targeted the viral neuraminidase of Influenza virus (oseltamivir, zanamivir) and the peptide-based inhibitors of the HIV protease. Other validated viral targets include proteins involved in replication of the viral genome, such as integrases, methyltransferases as well as host cell factors.

These successes were followed by the development of nucleoside inhibitors (NI) and non-nucleoside inhibitors (NNI) against the reverse transcriptase of HIV (HIV-1 RT) – which is now one of the most successful approaches for the treatment of HIV infections. NIs are incorporated into the newly synthesized genome by RT, and prevent further incorporation of incoming nucleotides, acting as obligate chain terminators. The NNIs act by binding to or close to the active site of RT and blocking replication. RNA dependent polymerases are attractive targets for anti-virals and successful drugs have been developed against human pathogens such as Hepatitis C virus and Influenza virus. Recently new compounds targeting the influenza virus polymerase complex have recently progressed to or through late phase clinical trials. The polymerase complex of influenza viruses is a heterotrimer and both replicates the viral genome as well as initiating transcription using short capped RNAs snatched from RNA polymerase II (Pol II). The polymerase has both a cap binding domain as well as an endonuclease domain which capture cellular mRNAs and subsequently cleave the mRNAs 10–12 nucleotides from the cap. This multifunctional enzyme is an attractive target for drug development and recently a number of drugs have been developed that target the different active sites – Pimodivir blocks cap binding, Baloxavir inhibits the cap-dependent endonuclease

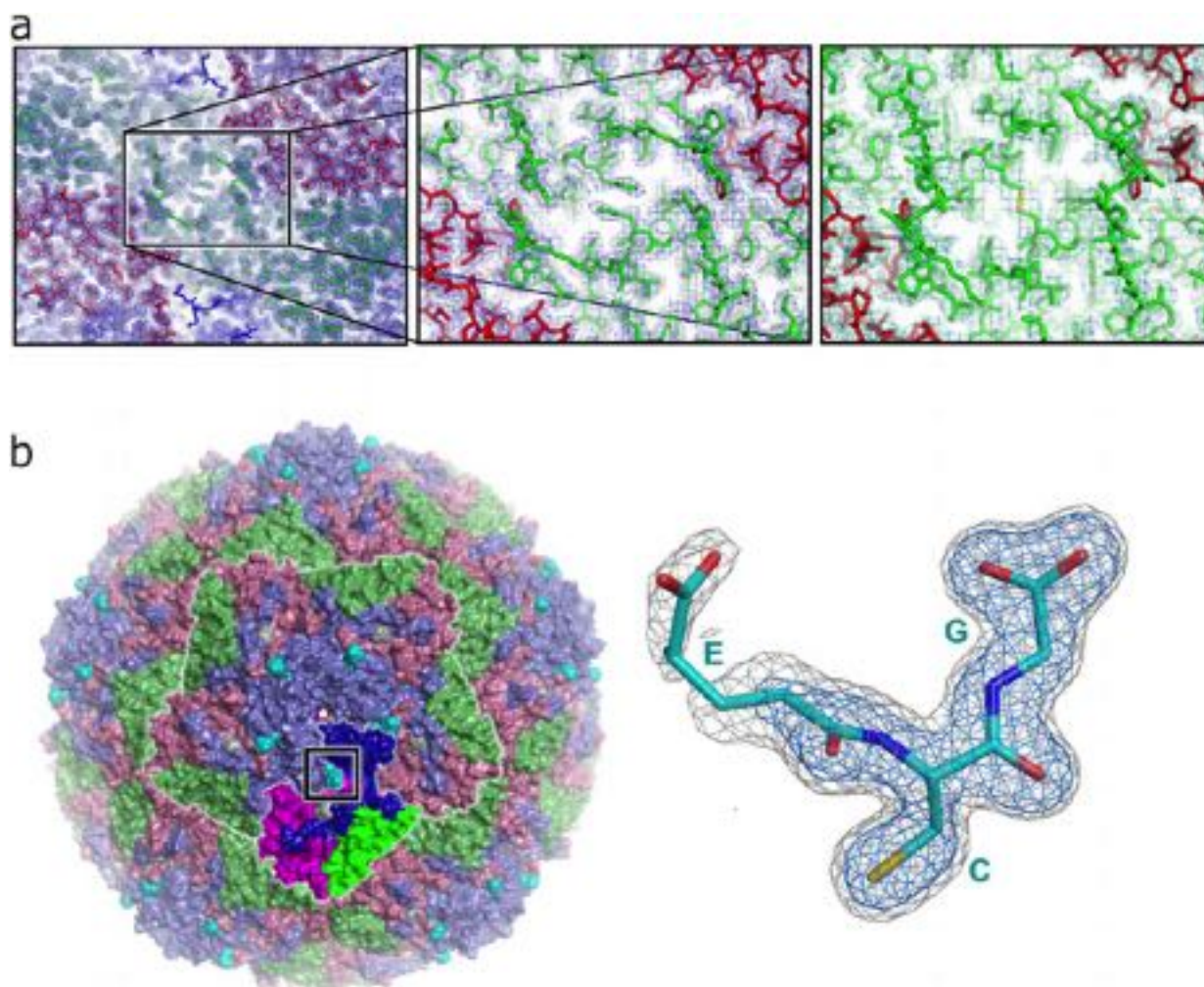


Fig. 3 Structural analysis of FMDV A22 empty particles and BEV. (a) Electron density maps for FMDV A22 (left and zoomed up middle panel) and the A22-H2093C mutant (right panel) showing the disulfide bond forming across the icosahedral 2-fold axis. (b) Structure of BEV capsid with one of the bound glutathione molecules highlighted and in the right panel the 15-fold NCS-averaged electron density for the glutathione.

activity thus inhibiting RNA cleavage and Favipiravir acts as a purine nucleoside which is recognized as an alternative substrate by the viral polymerase resulting in errors during viral RNA synthesis. The current anti-virals in clinical use or in late clinical trials target functionally key steps in the viral lifecycle, but unfortunately due to the high error rates in RNA virus replication, escape mutations drive drug-resistance in HIV, Influenza and other RNA viruses. Hence there is a great clinical need to discover additional strategies and targets as part of the development of novel antivirals.

Future Outlook

We have come a long way over the last thirty years since the first high resolution crystallographic analyzes of virus structures. Although now routine, these pioneering efforts were laborious and required innovative approaches to both collection of the data and the subsequent downstream analysis, which at the time was computationally demanding. The developments in technologies in cryo-EM means that for many samples, where there is heterogeneity in both conformation and composition, it has become the method of choice for structural analysis. However, crystallography, by selecting out a particular single conformation (or several with NCS) in the crystal, very often gives higher resolution structures than cryo-EM where most particles imaged must be thrown away or classified into a plethora of conformations. For structural analyzes of many of these fragile macromolecular complexes we will see hybrid models produced by the fitting of high-resolution structures of fragments or domains, derived by crystallography, into lower resolution cryo-EM maps produced by SPA or tomography. Looking into the future where imaging will take us to structural cellular biology these models will be key to building up an understanding of the complete virus life-cycle – crystallography will still have a key role to play.

Acknowledgments

I would like to thank Dr Harkiolaki and Dr Fry for critical reading of the manuscript, and Drs Duyvesteyn, Ren and Axford for help with production of the figures.

Further Reading

- Abrescia, N.G.A., Cockburn, J.J.B., Grimes, J.M., *et al.*, 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* 432, 68–74.
- Axford, D., Owen, R.L., Aishima, J., *et al.*, 2012. In situ macromolecular crystallography using microbeams. *Acta Crystallographica Section D Biological Crystallography* 68, 592–600.
- Baker, M.L., Jiang, W., Rixon, F.J., Chiu, W., 2005. Common ancestry of herpesviruses and tailed DNA bacteriophages. *Journal of Virology* 79, 14967–14970.
- Crowfoot, D., Schmidt, G.M.J., 1945. X-ray crystallographic measurements on a single crystal of a tobacco necrosis virus derivative. *Nature* 155, 504–505.
- Duyvesteyn, H.M.E., Ginn, H.M., Pietilä, M.K., *et al.*, 2018. Towards in cellulo virus crystallography. *Scientific Reports* 8, 3771.
- Esnouf, R., Ren, J., Ross, C., *et al.*, 1995. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. *Nature Structural & Molecular Biology* 2, 303–308.
- Fry, E., Acharya, R., Stuart, D., 1993. Methods used in the structure determination of foot-and-mouth disease virus. *Acta Crystallographica Section A: Foundations and Advances* 49 (Pt 1), 45–55.
- Grimes, J.M., Burroughs, J.N., Gouet, P., *et al.*, 1998. The atomic structure of the bluetongue virus core. *Nature* 395, 470–478.
- Harrison, S.C., Olson, A.J., Schutt, C., Winkler, F., Bricogne, G., 1978. Tomato bushy stunt virus at 2.9 Å resolution. *Nature* 276, 368–373.
- Kotecha, A., Wang, Q., Dong, X., *et al.*, 2017. Rules of engagement between $\alpha\beta6$ integrin and foot-and-mouth disease virus. *Nature Communications* 8, 15408.
- Kounatidis, I., Stanifer, M.L., Phillips, M.A., *et al.*, 2020. 3D correlative cryo-structured illumination fluorescence and soft X-ray microscopy elucidates reovirus intracellular release pathway. *Cell* 182, 515–530.
- Rossmann, M.G., Arnold, E., Erickson, J.W., *et al.*, 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317, 145–153.
- Rossmann, M.G., Blow, D., 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* 15, 24–31.
- Wang, X., Peng, W., Ren, J., *et al.*, 2012. A sensor-adaptor mechanism for enterovirus uncoating from structures of EV71. *Nature Structural & Molecular Biology* 19, 424–429.
- Winter, G., McAuley, K.E., 2011. Automated data collection for macromolecular crystallography. *Methods* 55, 81–93.

Relevant Websites

www.pdb.org-wwPDB
Worldwide Protein Data Bank.

Advanced Light and Correlative Microscopy in Virology

Sergi Padilla-Parra, University of Oxford, Oxford, United Kingdom; Department of Infectious Diseases, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom; and Randall Division of Cell and Molecular Biophysics, King's College London, London, United Kingdom

Charles A Coomer and Irene Carlon-Andres, University of Oxford, Oxford, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

CLEM Correlative Light and Electron Microscopy, a technique that consists on acquiring fluorescence microscopy micrographs and normally Cryo Electron Tomography to obtain data that is complementary from both techniques.

EM-CCD camera Electron Multiplied Charged Coupled Device sensor, Integrated circuit onto a silicon surface with light sensitive pixels. When photons are detected in the semiconductor oxide interface they generate charge read by electronics and translated into a digital image. Ideal for low intensities and noise characterization.

FCCS Fluorescence Cross-Correlation Spectroscopy, is a derivative from FCS that utilizes two channels to cross-correlate fluorescence fluctuations of two fluorescent particles that might co-diffuse together through a confocal volume. The technique assumes that if two particles co-diffuse together through a confocal volume these particles are interacting and therefore the technique is useful to detect protein-protein interactions in live cells at low concentrations.

FCS Fluorescence Correlation Spectroscopy, is a microscopy technique that employees the fluorescence fluctuations resolved in time coming from fluorescent particles diffusing through a confocal volume. A correlation analysis is performed on the fluctuation of the fluorescence intensity providing information on the concentration of fluorescent particles and their diffusion coefficient.

FLIM Fluorescence Lifetime Imaging Microscopy, a microscopy technique able to recover the time that an electron resides in the Lowest Unoccupied Molecular Orbital (LUMO) before relaxing down to the Highest Occupied Molecular Orbital (HOMO) through radiative (emitting a photon) and non-radiative pathways (emitting heat). The inverse of the kinetic constants regulating these process is the mathematical definition of the Lifetime for a particular fluorophore.

Fluorescent Proteins Ever since the discovery in the 1960s of the green fluorescent protein (GFP), composed of 238 amino acids and able to fluoresce when exposed to blue light (~488 nm) other variants and colors have been derived or discovered from other organisms. There is a rich

palette of different genetically encodable fluorescent proteins with attractive properties for imaging from the blue to the far-red.

FRET Forster Resonance Energy Transfer, a phenomenon that consists on the non-radiative emission of energy between two dipoles situated at the right distance (~10 nm), the right relative orientation and the right spectral properties between the donor and the acceptor of non-radiative energy (the donor absorption spectral must overlap with the acceptor's emission).

N&B Number and Brightness, is a correlation technique based on moment analysis to measure the average number of fluorescent particles diffusing through a confocal volume (that can be scanned to create an image) and the molecular brightness (in counts per molecule per unit time) of each fluctuation trace (i.e., pixel by pixel when analyzing a whole stack of images).

Pseudovirus A recombinant particle that contains its core and envelope proteins from different viruses. They may also contain altered or modified genes so that they will be unable to produce surface proteins on the infected host.

Quantum Dot A nanometer sized semiconductor crystal with distinctive conductive properties and able to fluoresce with high intensities.

sCMOS camera scientific Complementary Metal-Oxide-Semiconductor, these image sensors offer low noise, big field of view, high quantum yield and high speed. CMOS circuitry dissipates less power than other logic families of sensors.

Single Molecule Localization Microscopy A super resolution technique that consists of isolating single emitters and fitting the corresponding point spread function with a mathematical model that determines its location with a precision limited by its intensity (number of photons).

Stokes shift Difference between the absorption and emission fluorescence spectra for the same electronic transition caused by a loss of vibrational energy when electrons relax from the LUMO to the HOMO (see above).

Synthetic Fluorescent Dyes A fluorescent chemical compound that re-emits light upon exposure to excitation source. Fluorophores contain several combined aromatic groups.

Introduction

Live-cell imaging of virus-infected cells has served as an innovative platform to revolutionize and transform our understanding of virus infection and pathogenesis. But it was centuries after the visualisation of microorganisms with Anton van Leeuwenhoek's microscope in the 17th century, and decades after Abbe's optical diffraction limit of light microscopes was postulated, that

Table 1 Use of different microscope techniques to visualize single virus entry

	<i>3D acquisition</i>	<i>High-throughput</i>	<i>Super-resolution</i>	<i>Deep tissue</i>	<i>Thin sections</i>	<i>FRET compatible</i>	<i>FLIM compatible</i>	<i>Immunofluorescence compatible</i>	<i>SPT Compatible</i>
Wide Field		+				+	+	+	+
Spinning Disk	+/-	+			+/-	+	+	+	+
2 Photon	+		+	+	+/-	+	+	+	+
Structured Illumination			+		+			+	
STED			+		+			+	
SMLM			+			+	+	+	
TEM	+				+				
TIRF			+			+	+	+	+
Hi Lo			+			+	+	+	+

nanometer-scale virus particles were finally resolved. This revolution in virology was powered by imaging systems harnessing electromagnetic waves (i.e., electron microscopy) to visualize virion structure and the early and late stages of their life-cycles. In parallel, groundbreaking advances in light microscopy, in particular fluorescence and quantum dot labeling, combined with single-particle tracking (SPT) and quantitative spectral imaging approaches (e.g., FCS, FCCS and Number and Brightness), advanced numerous studies probing virion structure, entry and pathogenesis. Importantly, virology investigations now rest on the complementarity between electron and light microscopy approaches, where light microscopy sensitivity and resolution can now border single-molecules and 1 nm scales, respectively.

This cohesiveness between light- and electron-based microscopy platforms in answering critical questions surrounding virion structure, entry and pathogenesis are exemplified by recent advances in our understanding of many viruses. Human Immunodeficiency Virus (HIV-1) has been extensively studied and the application of different microscopy approaches has allowed us to understand many aspects of its pathogenicity and the availability of genetically-encoded fluorescent proteins (FPs) tagged to proteins of interests has yielded key insights regarding its lifecycle, including intracellular trafficking, nuclear import and capsid oligomer formation and budding in live single cells. Because of the rapid development of these sophisticated fluorescent probes and the ability to tag various different viral components, the detection of and ability to track single virus particles (i.e., SPT) in live cells has also provided a mechanistic understanding of key steps of virus infection, especially virus entry. In particular, SPT has created intense discussion concerning HIV-1 entry mechanisms (i.e., plasma membrane fusion versus endocytosis), and has been shown to differ in part according to the host cell studied. Additionally, our group utilized fluorescence fluctuation approaches (i.e., N&B) to uncover the evolution of the HIV-1 receptor (i.e., CD4) and coreceptor (i.e., CCR5 or CXCR4) stoichiometry during virion entry with single molecule spatial and temporal resolution. These studies are examples of how investigators are currently creating platforms to dissect interactions between host and virus-specific factors with high spatio-temporal resolution. In parallel, cryo-electron tomography (cryo-ET)-based methods have been able to link these live, dynamic observations between the virus and host with high-resolution structural data. For example, after cryo-ET was used to visualize the surprisingly few trimeric Env glycoproteins in their native environment on the virion surface, it was soon revealed that these glycoproteins cluster in an “entry claw” like fashion that is maturation-dependent, corroborating quantitative live-cell light microscopy studies.

In this article, our goal is to summarize the methodologies we believe are most useful to uncover mechanisms of virus entry, pathogenesis and structure in both live cells and tissue with single molecule resolution. In particular, we believe that recent innovations in virus labeling, SPT, quantitative spectral imaging and multiphoton microscopy will lead to unprecedented advances in virology studies at the single-molecule level in physiologically-relevant models, such as tissue explants, organoids and in vivo imaging. In addition, advances in cryo-ET and its role in advancing our understanding of virus structure and its influence on viral protein function will be discussed with a particular emphasis on HIV-1. Finally, we conclude with a brief discussion on the use of correlative live-cell and cryo-ET approaches to visualize and understand dynamic virus-host interactions during entry. For a comprehensive summary of common imaging modalities, see [Table 1](#).

Virus Labeling

The use of different methods to label enveloped and non-enveloped viruses is crucial for light microscopy. Overall, two approaches have been employed to label viruses during the last two decades. Either one or more viral proteins can be fused with a fluorescent protein (FP) or direct chemical labeling is employed utilising small dye molecules such as cyanine or Alexa dyes, or lipophilic dyes for enveloped viruses such as DiD or DiO.

Since the discovery of green fluorescent protein (GFP), many variants have been developed utilising directed evolution techniques that present attractive features in terms of emission spectra (ranging from the blue to the far red), brightness, quantum yield and photostability. Importantly the pKa of different proteins can be employed to measure pH, and these measurements can be very accurate using FRET-based biosensors incorporated in single viruses such as pHlameleons. The process of fusing a FP into a viral protein does not drastically differ from simple cloning employed to label a protein of interest. The corresponding FP is incorporated into the open

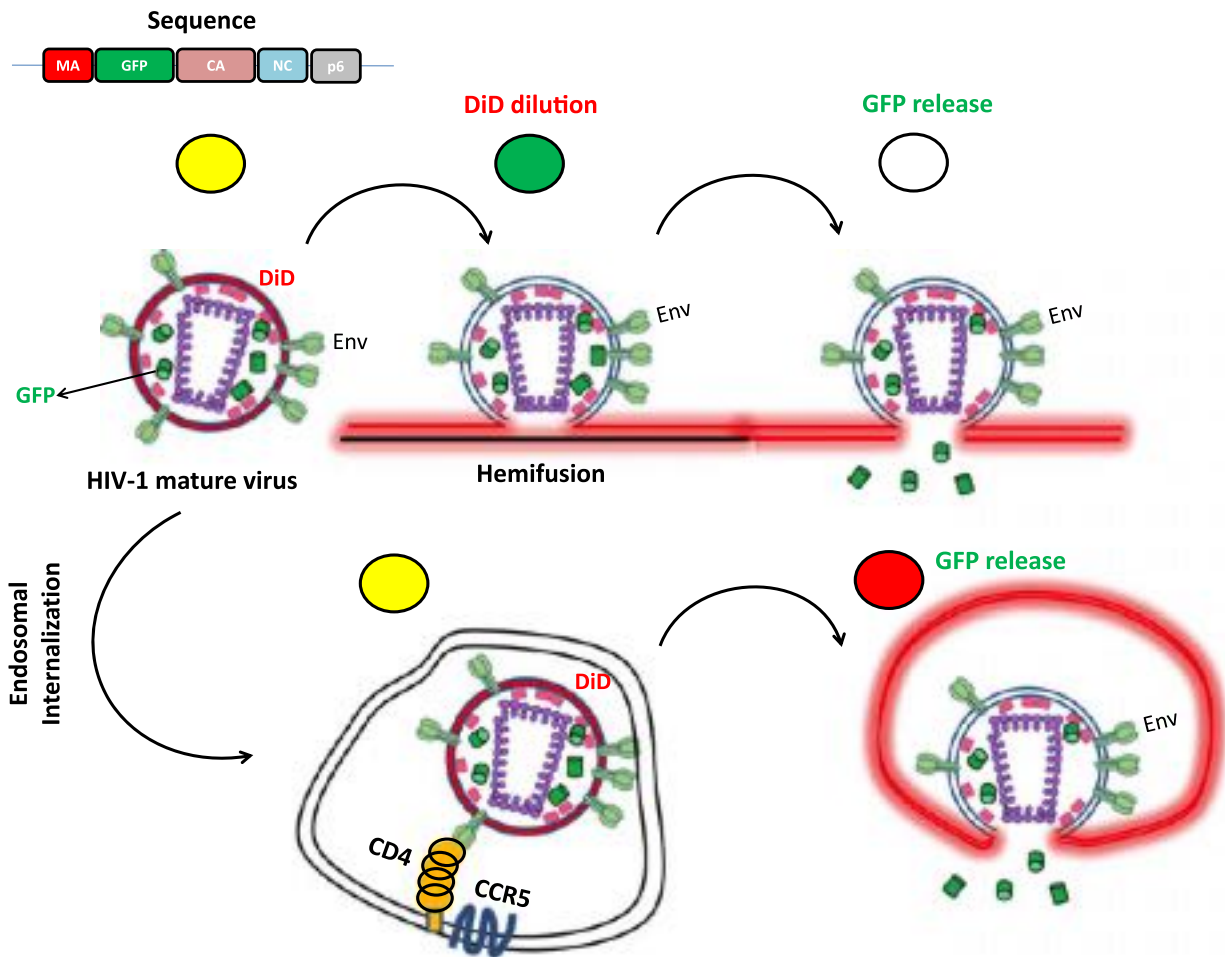


Fig. 1 Labeling of HIV-1 virions to ascertain the point of fusion via single virus tracking. Cartoon depicting double labeled HIV-1 mature virions with a strategy that identifies the point of fusion: membrane versus endosomal. DiD lipophilic dye (far red) stains the viral membrane and the capsid is labeled in the Gag (Gag-GFP). In mature viruses, cleavage releases the GFP (green fluorescent protein) that resides in the matrix of the virus. When hemifusion occurs the far red dye DiD is diluted in the plasma membrane of the host (yellow signal shifts to green). Subsequently, full fusion and pore opening allows the cleaved GFP to be released in the cytosol (green signal disappearance). If the primed double labeled virus is internalized via endocytosis and fusion occurs within endosomes, the DiD lipophilic dye redistributes around the endosomal lumen and only when fusion occurs there is a shift in color from yellow to red.

reading frame of the target viral protein, for example in the case of retroviruses Gag is a common target for FP labeling (Fig. 1). Nevertheless, the detection of single viruses requires large numbers of labeled viral proteins, and depending on the sensitivity of the light microscope employed, one can resolve single viruses when labeling a few Env glycoproteins or hundreds of cleaved nucleocapsid proteins, such as Vpr in the case of HIV.

Chemical labeling can also be employed to label virus structures. The type of bonds created between the dye and the viral protein can be covalent or non covalent. Thus capsids of non-enveloped viruses are often labeled with amino-reactive dyes. Another type of labeling is related to cell-permeable biarsenical dyes that specifically attach to viral proteins with tetracysteine peptide sequences. Importantly, the use of antibodies is not recommended as they might interfere with viral protein function before or after binding. In that regard, it will be interesting to witness how nanobodies can be employed to label viral proteins and how they affect their function, as nanobodies were utilised to target the HIV-1 Env in our recent study on HIV-1 entry.

Advanced Light Microscopy Approaches

Quantitative light microscopy has been employed to understand with molecular detail the life cycle of many viruses such as HIV, Ebola, Avian Sarcoma and Leukosis Virus, or Vesicular Stomatitis Virus. Light microscopy is a particularly attractive technique for virology as it can resolve, with single molecular resolution, the temporal dimension of different stages of infection. Importantly, the combined use of quantitative spectral imaging such as fluorescence lifetime imaging microscopy with single virus tracking, offers the opportunity to recover on-line information of single virus entry and the particular response of the host molecules that

are modulating this process. Here, we focus on advanced light microscopy techniques with considerable potential: real-time single virus tracking, spectral imaging and how to correlate these approaches with light super-resolution techniques and cryo-electron tomography. We have decided not to talk about super-resolution light microscopy as recent literature focuses on this aspect in the context of virus entry.

Real-Time Single Virus Tracking in Live cells

Real-time single virus tracking (SVT) has been extensively employed for many viruses during the last two decades. Indeed, high temporal resolution allows researchers to track the different stages of the infection process for single viruses. In a seminal work utilising this technique, Seisenberger *et al.* studied the infection pathway of Adeno-Associated Virus (AAV). The authors covalently attached Cy5 to viral proteins analyzing thousands of trajectories of single AAV-Cy5 particles in living HeLa cells. The authors established that single particles were able to diffuse randomly outside the cell; this slowed down when AAV approached the cell surface and around 13% of the particles internalized via endocytosis. Furthermore, this SVT-approach showed that endosome diffusion was directed toward the nuclei and that virus particles within the nucleolus underwent directed motion and were unidirectional. This report on the application of SVT on AAV opened up a number of possibilities to study the molecular environment and entry pathways of many other viruses. This work was carried out with wide field illumination and a CCD camera, an affordable set up that has since evolved with developments in optics and detection units (e.g., scientific CMOS and back illuminated CMOS cameras).

Combining the right strategy for virus labeling with SVT has proven to be beneficial to establish the point of entry and fusion for many other viruses. In a seminal paper describing the use of SVT to understand virus entry, Melikyan *et al.* examined individual retroviral fusion events. Avian Sarcoma and Leukosis Virus (ASLV) was labeled with a lipophilic dye and palmitoylated enhanced yellow fluorescent protein (YFP) incorporated in the inner leaflet of the viral membrane. Melikyan *et al.* found that hemifusion and small pore formation are key intermediates of ASLV fusion in CV-1 cells. In another paper, Padilla-Parra *et al.* took advantage of the combined use of FRET-based biosensors to quantitatively measure pH and SVT to detect single virus fusion (Gag-mCherry). This approach allowed the quantification of the pH sensed by individual ASLV virus particles during priming, endocytosis and fusion in live CV-1 cells. For the majority of ASLV virus tracked to the point of endosomal fusion, fusion occurred on average at 6–6.1 pH units. This article demonstrated the usefulness of combining SVT with FRET-based biosensors incorporated in single retroviruses. As mentioned above, there are now reports utilising SVT to dissect the point of entry for many other viruses. For example, recent applications of SVT allowed the detection of entry and fusion of HIV in macrophages via endocytosis as opposed to T cells that occurred on the plasma membrane.

Quantitative Spectral Imaging

The use of spectral imaging applied to virology offers the possibility to obtain time-resolved molecular information for a number of host-virus interactions. Various spectral-based techniques such as Fluorescence Lifetime Imaging Microscopy (FLIM), Number and Brightness (N&B) (Fig. 2) and Fluorescence Correlation Spectroscopy and derivatives (FCS) have been employed. In a recent report, our group combined SVT with N&B to resolve the time-resolved stoichiometry of HIV-1 pre-fusion reaction. We employed HIV-1 pseudoviruses labeled with Gag-mCherry exposed to live COS-7 cells co-expressing CD4 and CCR5 or CXCR4 coreceptors labeled with mTFP1 and mOrange respectively. The HIV-1 Env pre-fusion reaction was shown to consist of a three-step mechanism with different stoichiometry (i.e., number of CD4 and coreceptors engaged in the fusion complex) depending on HIV tropism. In another report our group also demonstrated the role of dynamin during HIV-1 entry and fusion by combining functional assays for HIV fusion with FLIM and N&B in live TZM-bl and primary CD4⁺ T cells. FLIM was employed to detect Förster Resonance Energy Transfer (FRET) in TZM-bl cells co-expressing dynamin labeled with GFP and mCherry. The addition of HIV pseudoparticles decorated with different envelope proteins (either VSV-G or JR-FL) induced a different response in the extent of dynamin interactions suggesting a different role for dynamin during entry for HIV and VSV. These results were confirmed when combining fast Total Internal Reflection Fluorescence Microscopy (TIRFM) with N&B with TZM-bl cells expressing Dynamin-GFP, where it was illustrated that the oligomeric state of dynamin differs when the cells are exposed to VSV-G or JR-FL pseudoparticles. Further experiments based on co-localization, SVT and cell-cell fusion confirmed the role of dynamin during HIV fusion, where it was shown that dynamin stabilizes the fusion pore during HIV entry with a low oligomeric state, (i.e., a tetramer). These results were confirmed by Aggarwal *et al.* utilising functional assays and SVT. Spectral imaging has been traditionally applied almost exclusively in cell biology. Nevertheless, there are a number of groups combining the power of these approaches to better understand different stages of the virus infection cycle. We are convinced of the general usefulness of these techniques when elucidating virus entry.

Multiphoton Imaging

Two-photon microscopy is a platform rooted in nonlinear interactions between photons and matter, in which two photons from an ultrashort, pulsed infrared laser each deliver approximately one-half of the energy necessary to excite the same fluorophore “instantaneously” to its lowest unoccupied molecular orbital. Combined with a high numerical aperture, an extraordinarily high photon flux is delivered to a fractional volume of the specimen. Furthermore, because a multiphoton absorption event is only likely at the focal point in the excitation volume, nonlinear photon interactions exist only in the focal plane. Moreover,

two-photon decay processes occur identically to a single-photon absorption event, leading to a similar fluorescence profile, albeit an “inverse” Stokes shift as the excitation is completed by infrared light.

In addition to two-photon microscopy, other multiphoton microscopy platforms can be utilised for label-free and non-invasive functional imaging in organoids and tissue, such as second harmonic generation (SHG), third harmonic generation (THG) and three-photon microscopy. Briefly, in three-photon microscopy three photons are absorbed by the fluorophore to achieve the excited state and in SHG and THG, high photon densities are optically scattered such that they combine to form a single photon with double the energy of the incident photon in either spatially-ordered or heterogenic specimens, respectively.

Most multiphoton modalities utilise near-infrared excitation, taking advantage of either endogenous fluorescence, SHG or THG to image biological phenomena in living tissue with high spatiotemporal resolution. For example, endogenous fluorescence is often the result of the excitation of metabolic redox reaction cofactors NADH, NAD(P)H or FAD. These molecules may absorb UV light and subsequently fluoresce; however, this is often photo-toxic for the living tissue examined. Therefore, by combining multiphoton microscopy with imaging (FLIM), researchers may be able to specifically study NAD(P)H and FAD dynamics to report on the metabolic redox activity of individual cells within tissue with exceptional environmental sensitivity. In particular, the fluorescence lifetimes of NAD(P)H and FAD change according to their environment; that is upon protein binding during the electron transport chain (ETC). Therefore, label-free lifetime imaging these molecules informs the ratio of free and bound NAD(P)H, in addition to the redox states of FAD/FADH₂ and NADH/NAD⁺ of single cells within the tissue. This information is in turn, related to the glycolytic and the metabolic states of the cells, which can change during and influence the course of viral entry and infection.

Excitingly, this technology has been employed in different biological and pathological contexts. For example, two-photon FLIM was recently employed in hepatocytes to demonstrate how hepatitis C virus core protein D2 decreases free and bound NAD(P)H lifetimes with a concomitant increase in free NAD(P)H levels, indicative of D2-induced disruption of the ETC, subsequently triggering a shift in glycolysis and lipid biomass accumulation in human hepatoma cell line Huh-7. Our group recently found using this approach the metabolic shift of primary T cells in the context of HIV-1 entry or infection. Indeed, the importance of cellular metabolism and infection

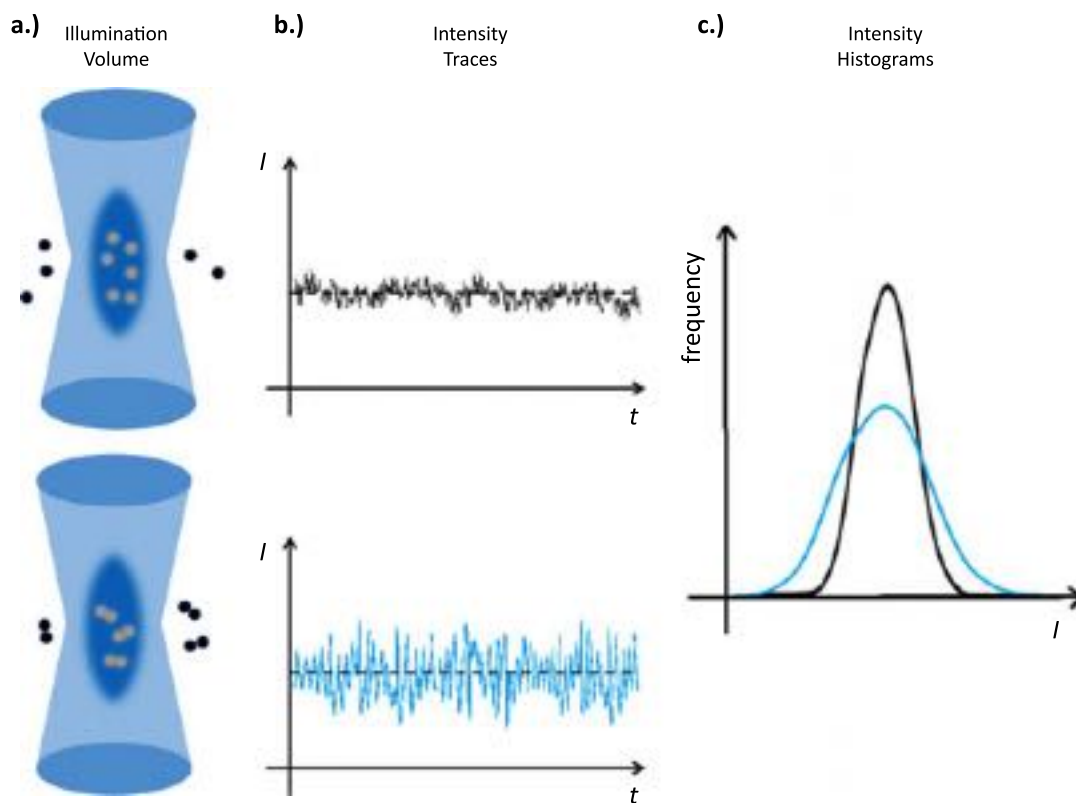


Fig. 2 *Mathematical theory of number and brightness.* Let us consider a system such that a monomeric protein (single circles) undergo the process of dimerization (paired circles). (a) Top row: before dimer formation, an average of six monomers are optically excited in the confocal volume. Bottom row: once dimers have been formed, an average of three dimers are optically excited in the confocal volume. Therefore, the concentration of fluorophores in the confocal volume is equivalent, such that the (b) intensity (I) means in both scenarios are the same. However, it is important to realise that after dimerization, the variance in intensity between the two rows are dissimilar, where the dimers (blue intensity trace, bottom row) possess a higher variance in intensity when compared to the monomers (black intensity trace, top row), due to the fact the fluorophores must enter and leave the confocal volume in pairs. This constant mean and difference in variance is depicted in (c) as intensity histograms (monomers in black, dimers in blue), such that the widening of the histogram shows the increase in variance in the dimer scenario. These intensity and variance values are utilised to calculate the number and brightness of entities in live cells for mobile particles.

has been shown in numerous publications. In particular, studies investigating the role of glycolysis and glucose uptake in HIV-1 infection illustrate that the host capacity for increased glucose transport could condition CD4⁺ T cell and macrophage susceptibility to HIV-1 infection. The possibility of label-free imaging of individual cell's metabolism within infected tissue would aid in understanding many questions related to HIV-1 pathogenesis. These types of quantitative measurements are further warranted because CD4⁺ T cells sitting in observation chambers most likely do not behave the same as CD4⁺ T cells residing in a 3D matrix. It is therefore very likely that their behavior will be very different also in tissue and organoids.

In addition to combining FLIM with multiphoton microscopy, on-line 3D single particle tracking originally developed by Hellriegel and Gratton has recently been multiplexed with two-photon microscopy, giving rise to: 3D Multi-resolution Microscopy (3D-MM). Currently, 3D-MM has been harnessed to investigate primarily coated nano-particles (100 nm) in limited cellular contexts, particularly endocytosis and transcytosis. Nevertheless, because these particles are of similar size to HIV-1 and other virions, and the imaging platform is equipped with a multiphoton laser, it is theoretically possible to image and track single HIV-1 particles in living tissue. However, the probability of detecting and tracking single HIV-1 labeled particles that would be subsequently infectious and productive in tissue is very low. This drawback necessitates automation of these experiments (both acquisition and analysis) to provide crucial information about the whereabouts and specific trajectories of infectious particles in physiologically relevant environments.

Encouragingly, a real-time 3D single-particle tracking approach using a dynamic moving laser spot, called 3D-DyPLOT, has been recently produced. Moreover, the authors were strikingly able to track VSV-G-eYFP pseudotyped lentiviral particles. The large observation area of this method (1 $\mu\text{m} \times 1 \mu\text{m} \times 4 \mu\text{m}$) and the improvement in detection sensitivity as compared to 3D-MM makes this technique a potential candidate to track fast single virions in tissue (Fig. 3). However, scattering, bleaching and limited photon counts coming from single virions could challenge long-term observations and these prospects need to be validated.

Although two-photon microscopy offers several advantages to elucidate previously unknown information about virus infection and metabolic pathogenesis in thick specimens (i.e., tissue and organoids), there are often several drawbacks. Firstly, the required ultrafast laser is prohibitively expensive, secondly the augmented photobleaching from the higher photon density relative to standard confocal microscopy precludes multiple excitation protocols. Additionally, naturally occurring chromophores often present in tissue samples must be considered when deciding if two-photon microscopy is the optimal method for an investigation.

Cryo Electron Tomography

Cryo-tomography (cryo-ET) is a 3D imaging technique that, when combined with advanced computational methods, aims to construct 3D images of cells and virus architectures in situ, as there is no requirement for fixation, dehydration or sectioning in sample preparation. Briefly, cryo-ET images are obtained from samples applied to glow-discharged grids with gold fiducial particles, which are subsequently vitrified by plunge-freezing. During imaging, after deciding on an optimal choice of electron dose and defocus (which generate contrast), a tilt-axis series of images are prepared and aligned to compute a tomogram of the desired area of interest, revealing its 3D structure.

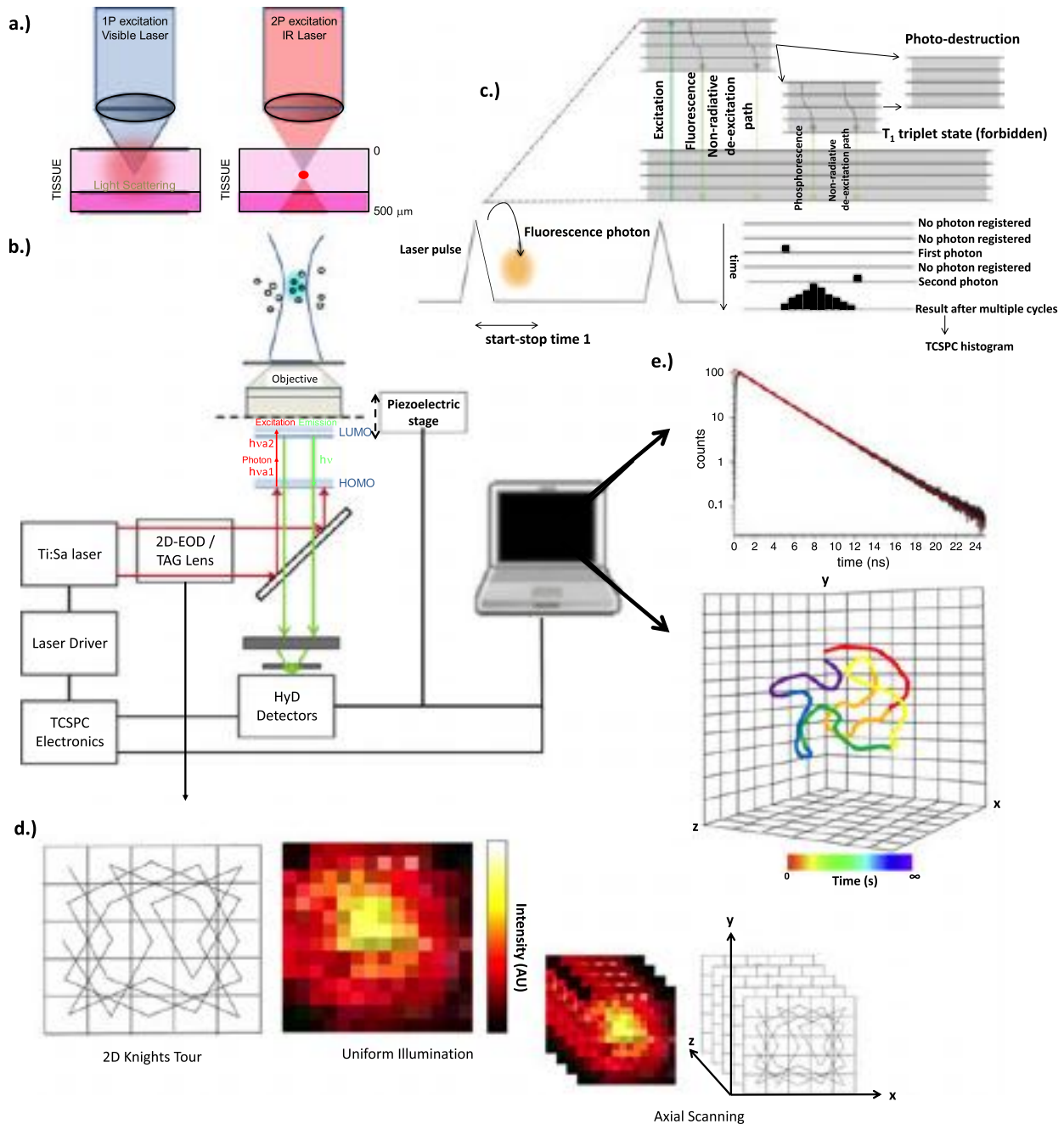
Until recently, studies employing cryo-ET platforms were plagued by suboptimal labeling and low signal-to-noise ratios. However, the implementation of novel direct electron detectors, focused ion beam milling, phase plates and optimized subvolume averaging methods are expanding the capability of cryo-ET to investigate macromolecular complexes in the context of virus-host interactions. The first wave of these advances have been critical players in imaging viral glycoproteins of enveloped viruses. With these capabilities, cryo-ET has now been used to construct molecular models of viral glycoproteins and their interactions with receptors and antibodies, complementing investigations targeting novel pharmacological approaches or rational-vaccine design strategies against HIV-1.

The utilization of cryo-ET has been paramount in understanding the architecture of native HIV-1 Env in hopes of designing novel preventative strategies. In fact, cryo-ET was initially utilised to uncover the original 3D structure and numbers of HIV-1 Env trimer on the virion surface, ranging from 7 to 20 Env trimers across all isolates. Shortly thereafter, Liu and colleagues discovered the flexible mobility of the Env trimer upon binding of receptor CD4 and broadly neutralizing antibody b12 at a resolution close to 2 nm, paving the way for subsequent cryo-ET studies to uncover the multiple conformational changes HIV-1 Env must undergo to reach its prefusion conformation. These studies not only suggested strategies to generate antibodies against the trimeric nature of the Env spike, but also contextualized how some antibodies could stabilize closed pre-fusion conformations, explaining their broad and potent neutralization. These broadly neutralizing antibodies (bNAbs) generally target five sites against HIV-1 Env, and now the architecture of these bNAb-Env interactions have been determined through cryo-ET.

Although cryo-ET has been paramount in extracting structural information about HIV-1 glycoproteins interactions with receptors and neutralizing antibodies, much of this information is poorly resolved, and important parameters of neutralization (i.e., epitope exposure) must be inferred. Consequently, it is imperative for structural studies which utilize cryo-ET to be contextualized and/or complemented with other microscopy-based methods to confirm their results. For example early observations with cryo-ET illustrated HIV-1 Envs typically clustered on the envelope surface and this clustering of Env trimers correlated with infectivity. Furthermore, other observations of Env trimers showed their mobility on the virion surface upon approach to the host CD4⁺ T cells, forming a namesake "entry claw". Interestingly, subsequent studies have illustrated with STED microscopy that this entry claw is maturation dependent, a pioneering study unveiling the interdependence of the exterior virion surface Env mobility required for entry with the maturation-dependent Gag lattice rearrangement.

This utilization of both fluorescent imaging and cryo-ET has been a recurring motif in HIV-1 Env investigations and illustrates the importance of combining both approaches to answer questions regarding virus-host interactions to inform rational-vaccine

design. For example, single-molecule FRET (smFRET) may resolve individual macromolecular conformational rearrangements of viral glycoproteins, often revealing hidden conformational states undetected by alternative approaches due to their insufficient accumulation. Recently, the dynamic structural rearrangements of trimeric HIV-1 Env in native virions, visualised by smFRET, demonstrated how the Env trimer engaged with receptor CD4 and co-receptor CCR5 in different smFRET states whilst sampling different intrinsic structural configurations. These experiments showed that both unliganded laboratory-adapted and clinical isolate HIV-1 Env preferred the closed conformation (i.e., a low-FRET state), where the CD4 binding site was shielded by the V1/V2 loop at the trimer's apex, and that introduction of soluble CD4 domains or antibody fragments stabilized other smFRET states (i.e., high-FRET state) corresponding to open conformations depicted in cryo-ET studies. Furthermore, Markov modeling of the smFRET conformations revealed that the open conformation is achieved via a previously uncharacterized intermediate smFRET state, which has not been structurally characterized, yet may represent a critical motif for neutralization. Given that this intermediate state can be stabilized in some isolates by soluble CD4, it will be critical to characterise this state with the nanometer resolution provided by cryo-ET to uncover its significance. Finally, by utilising antibodies which bound the CD4-binding site, the



V1/V2 region, the V3 loop and fusion loop, it was shown that these antibodies stabilized the closed conformation previously suggested by cryo-ET-based investigations, demonstrating that a general mechanism of Env inhibition occurs by stabilizing the closed conformation (i.e., high-FRET state) such that the co-receptor binding site is obstructed.

Cryo-ET has provided structural breakthroughs to increase our understanding of the architecture of HIV-1 Env and has provided the preliminary structural knowledge to inspire further functional investigations to capture its dynamic nature and targets for neutralization. In addition to virus entry, cryo-ET has illuminated other important aspects of the virus lifecycle, such as virus maturation or integration, pinpointing the mechanism of novel maturation inhibitors (i.e., Bevirimat) and refining drug design of HIV-1 integrase inhibitors. Therefore, it can be expected that cryo-ET will continue to inform structural-based vaccine and therapeutic design as novel technologies increase its capability to image at near-atomic resolution.

Correlative Light and Cryo Tomography in Cells

Live, biological specimens are dynamic and pleomorphic, transforming the process of unequivocally identifying and localizing a molecule imaged in cryoEM into a difficult task. This is further complicated by the rarity and isolated nature of some biological events such as viral entry. However, this challenge of correlating structural and dynamic live-cell data can be resolved by the exquisite combination of the spatiotemporal information gained from fluorescence microscopy and structural data extracted from cryoET, and typically referred to as correlative light and electron microscopy (CLEM) (Fig. 4).

CLEM-approaches were first investigated by proof-of-principle experiments utilising HeLa cells to study the course of adenovirus infections and the subsequent changes to the nucleus during infection. Since then, a plethora of recent advances have propelled CLEM into a wealth of studies analyzing the virus-host interaction at multiple steps of the viral life-cycle. These advances include strategies to better manipulate and preserve viruses and cells in their near-native state, in addition to advanced labeling of macromolecular structures to localize tagged molecules both inside the cell and at its surface, a drastic improvement to canonical immunogold-labeling which limited labeling to cell surface and viral glycoproteins. Importantly, the development of the cryo-fluorescence light microscopy stage has made it possible to correlate the same vitrified, near-native fluorescently labeled sample at both the light and electron microscope. Because of these advances, in addition to both microscope software and hardware developments over the last decade the structural insights gained from cryoEM are merging with investigations in cellular physiology.

Advances in cryo-EM have been key to understanding various HIV-1-host interactions. For example, in a proof-of-principle experiment, a post-embedding CLEM approach was utilised to localize individual, non-fusogenic eGFP-MA labeled HIV-1 particles among filopodia-like protrusions of MDCK cell surfaces in electron tomograms. This study was expanded upon by Jun and colleagues, who combined high-speed confocal live-cell imaging and cryoET to create a novel platform which could locate ROIs for cryoET data collection for dynamic, diffraction-limited infectious virus particles, as well as characterise the spatiotemporal behavior of fusion-capable HIV-1 particles in living cells prior to and after membrane fusion. Specifically, samples which had been imaged live were imaged again by cryo-light microscopy after vitrification, enabling the HIV-1 particles to be relocated and visualised in 3D in situ, in addition to tracking the dynamics underpinning post-entry processes, such as HIV-1 capsid disassembly after membrane fusion. Because of these studies, it has been clearly demonstrated that CLEM is a feasible and robust method to analyze viral entry, and has been applied to other viruses recently such as vaccinia and herpes simplex viruses.

Perhaps one of the most profound examples of virus-host interactions can be surmised from virus restriction, and with a resurgence of interest in restriction factor biology, understanding the interactions between host and viral proteins can be

Fig. 3 Multiphoton microscopy and its capabilities to perform fluorescent lifetime imaging (FLIM) or online 3D single particle tracking (SPT).

(a) Comparison of conventional single photon confocal microscopy (left) with canonical two-photon microscopy. It should be noted that the focal point generated by the two photon beam generates considerably less light scattering and enables deeper penetration depth into the tissue specimen whilst creating a finer focal volume. (b) Model of a two-photon FLIM system capable of 3D SPT modeled after 3D-DyPlot. A pulsed titanium: sapphire laser without utilization of the 2D-EOD/Tag lens excites the sample for FLIM (explained further in c). For online 3D SPT, a 2D electric optical deflector deflects the laser beam to generate a 5×5 knight's tour pattern, achieving equivalent illumination over the $1 \mu\text{m} \times 1 \mu\text{m}$ square area. Simultaneously, to create a 3D dynamically moving laser point in the axial direction, the same beam is deflected by a TAG lens. Photons are collected by a detector (shown as HyD detectors), but can also be collected by avalanche photodiode and monitor cameras (i.e., sCMOS), where the real-time position of the particle is determined by the photon arrival times. This information is subsequently transferred to the 3D piezoelectric stage in order to keep the particle within the focal volume of the objective lens by counteracting its motion. (c) A general description of time-correlated single photon counting (TCSPC), where a pulsed femtosecond laser (i.e., Ti: Sa laser in b directed towards the dichroic) is used as the fluorophore's excitation source. Fluorophore excitation is followed by the emission of a photon with a given start-stop time. The excitation of the fluorophore can be depicted by the Perrin-Jablonski diagram and its various de-excitation paths after excitation. This excitation and fluorophore detection start-stop cycle is repeated many times, which leads to the construction of a TCSPC histogram relating the number of photon counts to the length of time needed to detect the fluorescence photon after laser pulse excitation (i.e., the lifetime of the fluorophore). Subsequently, the histogram is fitted with an exponential decay curve to extract the lifetime from the slope. (d) The 5×5 knight's tour pattern (left) generated by the 2D-EOD lens to ensure equal illumination across the specimen in the XY plane and its corresponding intensity image (right). These images are created over a long axial distance by the TAG lens without loss of fidelity via the deflection the 2D knight's tour above and below the center focal plane (far right), enabling approximation of the particle's axial position despite the fast tracking and low photon counts. (e) Final reconstruction of the TCSPC histogram (top) or the 3D trajectory of the tracked particle (e.g., labeled virus) in a tissue sample.

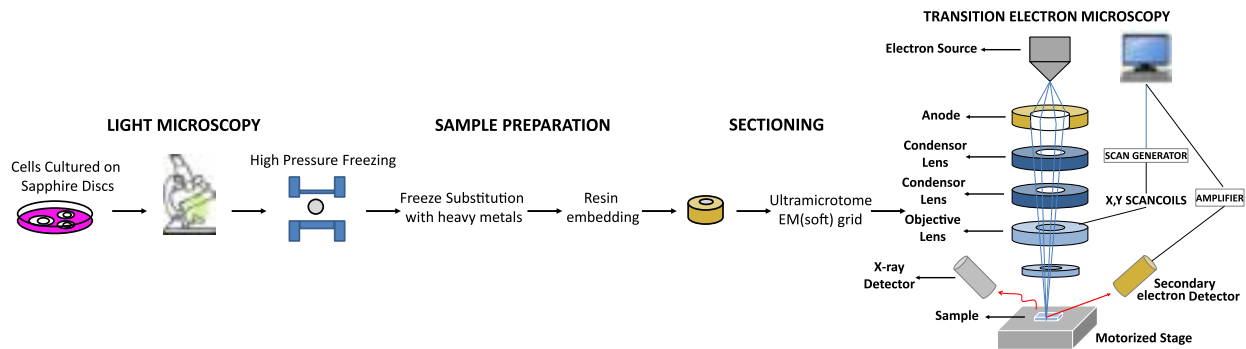


Fig. 4 Workflow of analysis of cells via CLEM scheme. Cells can be grown on sapphire disks and first analyzed under the light microscope. Cells labeled with fluorescent proteins or other means can help to localize proteins of interest and organelles but also quantitative imaging could be performed at this stage (this includes lifetime imaging, FLIM or super-resolution approaches that preserve the sample with mild illumination approaches). After this, cells are fixed and embedded in resin. The sapphire disks are then removed from the resin (polymerized) block. Trimming of the resin can then be performed and cells sectioned with a diamond knife. The very thin sections are then collected on slot grids and examined in a transmission electron microscope (TEM) to obtain ultrastructural information.

informed by their structure and activity in living cells. Recently, the arrangement of HIV-1 particles anchored to the host plasma membrane via tetherin, a restriction factor which abrogates virion release, has been uncovered via a CLEM-based approach. Before this study, investigations aiming to analyze the ultrastructure of HIV-1 tethered to human cells were limited to membrane-disrupting procedures such as heavy-metal staining and chemical fixation for TEM or SEM imaging. However, by imaging mCherry-Gag labeled HIV-1 VLPs tethered to HT1080 human cells via CLEM, it was shown for the first time at high-resolution in 3D that HIV-1 VLPs and virions were connected to the plasma membrane and each other by approximately 1–8 rod-like tethers over a range of distances between 7 and 37 nm which separated tethered HIV-1 virions from each other or from the plasma membrane.

Although CLEM has bridged high-resolution structural studies with the visualisation of dynamic biological events, several limitations of this approach to visualize virus-host interactions exist. For example, because the electron beam of many TEMs is unable to penetrate sample depths beyond approximately one micron, cryo-CLEM is limited to specimens with a thickness below ~ 750 nm, therefore excluding much of the cell's interior. However, this limitation is now being circumvented by multiplexing sectioning techniques with CLEM, such as cryo-ultramicrotomy, focused ion-beam machining to cut sections, and scanning TEM-tomography in order to visualize and preserve internal cellular structures and their fluorescent markers.

In addition to sample thickness, there remains a resolution gap between cryo-fluorescence microscopy and the sub-nanometer resolution obtained via cryoEM, which often impedes the interpretation of CLEM data. This is often due to the suboptimal optical properties of the resin of embedded samples, which influence fluorophore behavior via the cross-linking them and altering their environment, creating a challenge to introducing super-resolution to CLEM platforms. Furthermore, these single-molecule localization methods often rely on high excitation intensities on the sample to obtain optimum photon yields, which can cause sample devitrification and damage. Nevertheless, a general approach to cryo-super-resolution microscopy that enables subsequent cryoEM investigations has recently been envisaged utilising photoswitchable fluorescent proteins which remain active after vitrification, need not require cryoprotectants, and were utilised to image lipid nanotubes on intact mammalian cells for the first time at a localization precision of 30 nm. Undoubtedly, in combination with focused ion-beam milling to image intracellular structures, super-resolution CLEM will become a pioneering technique that will be widely available and applicable via the use of commercially-available tools and fluorescent proteins to localize region of interest (ROIs) with high precision.

Discussion and Outlook

Today, both electron and light microscopic imaging techniques are vital tools which can unveil the intricacies of different steps in viral replication cycles with unprecedented spatiotemporal resolution, particularly in the context of viral entry. In this article, we have highlighted seminal investigations which have illustrated the utility of an array of complex microscopy platforms while addressing their benefits, novelty and drawbacks for quantitative analysis of the processes under observation. Nevertheless, improvements continue to flood the literature, including the continual development of labeling strategies, semi-automated or automated data acquisitions as well as correlation methods and algorithms for processing SPT, FCS/FCCS and super-resolution imaging data are becoming increasingly unified and easily accessible for unfamiliar users. In this context, we firmly believe that novel insights revealing the mechanistic underpinnings of the entry and replication of many viruses will be ascertained. Undoubtedly, these observations will become invaluable for the development of novel and potent antivirals or vaccine strategies.

Further Reading

- Briegleb, A., *et al.*, 2010. Correlated light and electron cryo-microscopy. *Methods in Enzymology* 481, 317–341.
- Giepmans, B., Adams, S., Ellisman, M., Tsien, R., 2006. The fluorescent toolbox for assessing. *Science* 217–225.
- Iliopoulou, M., *et al.*, 2018. A dynamic three-step mechanism drives the HIV-1 pre-fusion reaction. *Nature Structural & Molecular Biology* 25, 814–822.
- Hellriegel, C., Gratton, E., 2009. Real-time multi-parameter spectroscopy and localization in three-dimensional single-particle tracking. *Journal of the Royal Society Interface* 6, S3–S14.
- Hoffman, D.P., *et al.*, 2020. Correlative three dimensional super-resolution and block-face electron microscopy of whole vitreously frozen cells. *Science* 367, 6475.
- Miyauchi, K., Kim, Y., Latinovic, O., Morozov, V., Melikyan, G.B., 2009. HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell* 137, 433–444.
- Munro, J.B., *et al.*, 2014. Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science* 346, 759–763.
- Zhu, P., *et al.*, 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 441, 847–852.

Atomic-Force Microscopy (AFM) Investigation of Viruses

Alexander McPherson, University of California, Irvine, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

Atomic Force Microscopy Principles and Technology

A direct imaging technology that promises to have a significant impact on structural virology, and which is, in most ways, complementary to X-ray diffraction and electron microscopy is atomic force microscopy (AFM). The method has seen only limited use in the study of viruses, but its relatively low expense and simple principle makes it an attractive option. With AFM, the general, and sometimes specific features of a virus, even in impure preparations, may be swiftly determined. The images, which are three dimensional, not projections, are also rather easy to interpret, and much quantitative information may also be gleaned. The microscope itself is a small device, which has a size about that of a common household coffee pot, but may be accompanied by substantial ancillary computing support.

An AFM device and the principles by which it operates are shown schematically in Fig. 1(a). AFM instruments can be operated in either contact mode, or what is referred to as tapping mode. In contact mode, a probe made of silicon or silicon nitride is placed in near contact with the surface of interest, say the capsid of a virus, and then translated in a systematic raster mode over the surface. The AFM probe is a sharp stylus, like those pictured in Fig. 1(b), similar in function to a minute phonograph needle. The tip ideally has a single point, with a very small radius of curvature. The probe is mounted at the end of a short cantilever, typically 100–250 μm in length.

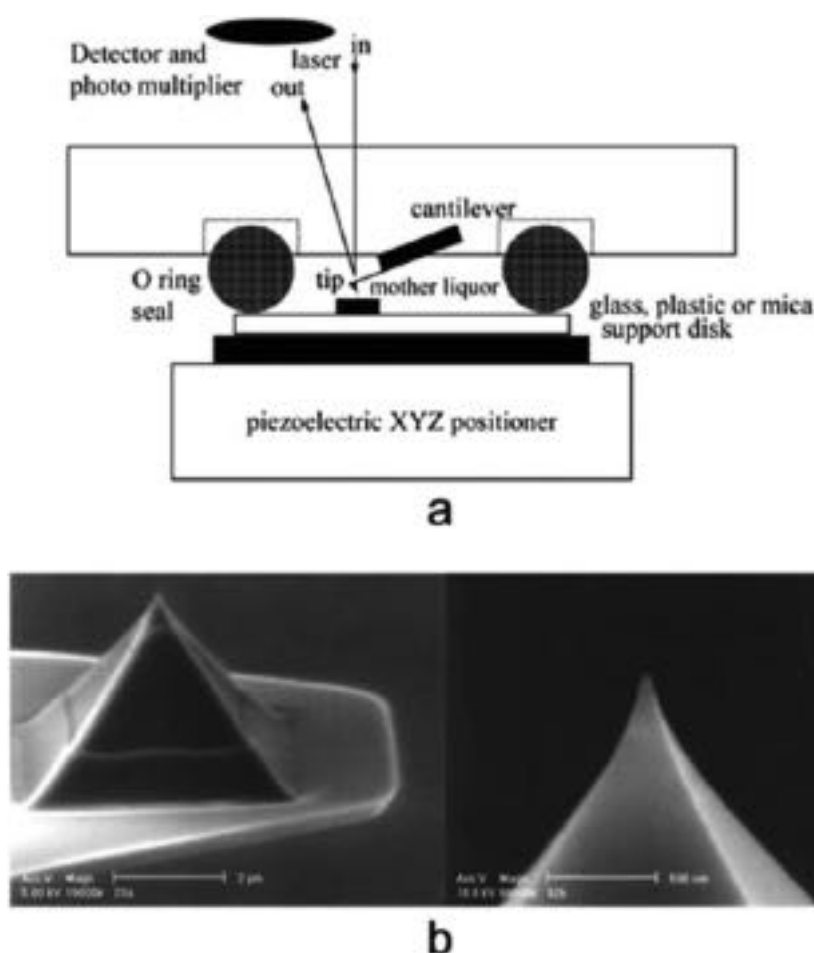


Fig. 1 Schematic drawing illustrating the principles of an atomic force microscope. (a) The vertical deflection that the cantilever tip experiences upon encountering some topological feature on a specimen is amplified through a reflected laser beam, which is tracked and reported by a split-diode photoelectric detector. Scanning takes place in a fluid-filled (or dry if preferred) cell of about 75 μL in volume. The sample is translated in a raster manner by piezoelectric positioners upon which the fluid cell is mounted. (b) Scanning electron micrographs of high-quality AFM tips etched from silicon.

Scanning is achieved by translating the sample beneath the probe, using a piezoelectric positioned x–y stage, along a continuous sequence of raster lines. As the probe tip passes over the surface, it interacts through “aggregate atomic forces”, which remain somewhat mysterious, with structural features on the surface. Encounters with these substructures cause the probe to be displaced vertically as the tip rides across. Exceedingly small displacements of the tip are amplified by deflection of a laser beam that is reflected from the upper surface of the cantilever, and these deflections are detected and tracked by a split photodiode. Photoelectric circuitry converts the deflections into height information. The resulting scan data, recorded as a digital topographical image, can then be presented in a number of visual formats.

In contact mode of operation, the data may be acquired as either “height” or “deflection” information, or the data from both modes may be obtained simultaneously. In “height” mode the sample surface is maintained at a constant distance from the probe tip by the piezoelectric positioner below, through a feedback mechanism. The cantilever deflection in this case is very small. In “deflection” mode the sample is stationary and actual cantilever deflection data are collected. Microfabricated cantilevers exert a force on the substrate surface of about 10^{-9} to 10^{-12} N/m, and, as one might anticipate, the resolution of the technique depends on the degree of force employed. Up to a point, the greater the force between probe and surface, the more sensitive the probe is to surface variations. On the other hand, the greater the force, the more the probe may perturb the surface.

Sample perturbation and other problems arising from unfavorable probe – surface interactions have been obviated to a great extent by the development of “tapping” mode instruments. With tapping mode, the probe tip is not in continuous contact with the sample surface, but rapidly oscillates up and down as it is scanned over the surface, essentially “tapping” its way and gently sensing the heights of obstacles it encounters. In tapping mode, the vertical position of the sample is continually adjusted by a feedback mechanism to maintain the amplitude of the freely oscillating probe constant. Image acquisition time ranges from 0.5 to 4 min, with shorter scan times usually associated with greater tip – specimen interaction. Tapping mode minimizes contact between the probe tip and the sample surface and greatly reduces lateral forces.

The “tapping mode” approach has proven a significant boon in biological investigations as it has allowed the characterization of samples that would otherwise be too soft or too fragile to withstand contact mode examination. Operating with tapping mode in a liquid environment presents some complications due to fluid dynamics, but these are not severe. A constraint that sometimes presents obstacles during analysis in a liquid medium is that the specimen under study must be fixed to, or made to adhere firmly to the substrate surface of the fluid cell, which may be glass, cleaved mica, plastic, or any other hard material. To achieve this, it may be necessary to treat the substrate with various reagents, such as poly-L-lysine, to induce better adhesion of samples. If this condition is not met then the specimen will move due to interaction with the probe and no useful information is gathered.

One particular feature of AFM must be borne in mind whenever one is interpreting images. The one or two-dimensional trace obtained of any object, or surface substructure, is the convolution of the tip shape with that of the feature being scanned. This is illustrated in Fig. 2. An image of an object scanned with a broad, dull tip is not the same as that acquired with a sharper tip. In particular, while the height of the object will be the same regardless of the tip shape (because the maximum vertical deflection of the cantilever tip would be the same), the lateral dimensions will not. The broader tip yields a broader object, the sharper tip produces the more accurate size. Because one does not, in general, know the tip shape one is working with at the time, the image cannot be easily deconvoluted to provide the true dimensions. Hence, height information is almost always trustworthy, but lateral measurements are suspect. The reliability of lateral measurements can, however, be increased if some standard having defined spatial features is first scanned and its known spacings or cell dimensions compared with those in the image. Such standards may be etched grids on silicon, or the surfaces of protein crystals.

The areas of scanning fields may range from 20 nm^2 to $150 \mu\text{m}^2$, with a spatial resolution on biological samples approaching 1 nm in the best of cases. Thus it provides precise visual detail over a size range that eludes most other techniques. Lateral resolution varies depending on the prominence of features and the deformability of the specimen. For small isolated samples on mica, such as macromolecular assemblies and single virus particles, the resolution is most limited by the size and structure of the tip. Commercially available tips have radii of curvature in the 5–20 nm range and provide resolution at fractions of those dimensions.

AFM of Biological Specimens

On large soft samples, such as living animal cells lateral resolution may be more limited by the motion and deformation of the cell surface in response to tip pressure than by tip structure. Height resolution for all samples is typically better than 1 nm. Because visualization can be carried out in a fluid environment, specimens may suffer no dehydration as is generally the case with electron microscopy, and they usually require no fixing or staining. Indeed, specimens can be observed over long periods, so long as they stay relatively unchanged and immobilized during a single frame interval. For the most part, even living cells seem oblivious to the presence of the probe tip.

Specimens, however, are not always best visualized under physiological conditions, particularly when high resolution is desired. Because cantilever tip pressure, even in “tapping mode”, may produce deformation, for example of a cell membrane, in some cases fixation is the better option. This, as with light microscopy histological procedures, usually relies on glutaraldehyde, paraformaldehyde, or osmium tetroxide fixation, followed by dehydration and imaging in water-alcohol mixtures, or in air. These methods have been developed by microscopists for more than a century to preserve the natural morphology of a sample but still allow high-resolution imaging. While not as ideal as *in situ* observation (the cells of course are no longer alive or viruses infective), fine details of their structures can be visualized that would otherwise be obscured by membrane flexion.

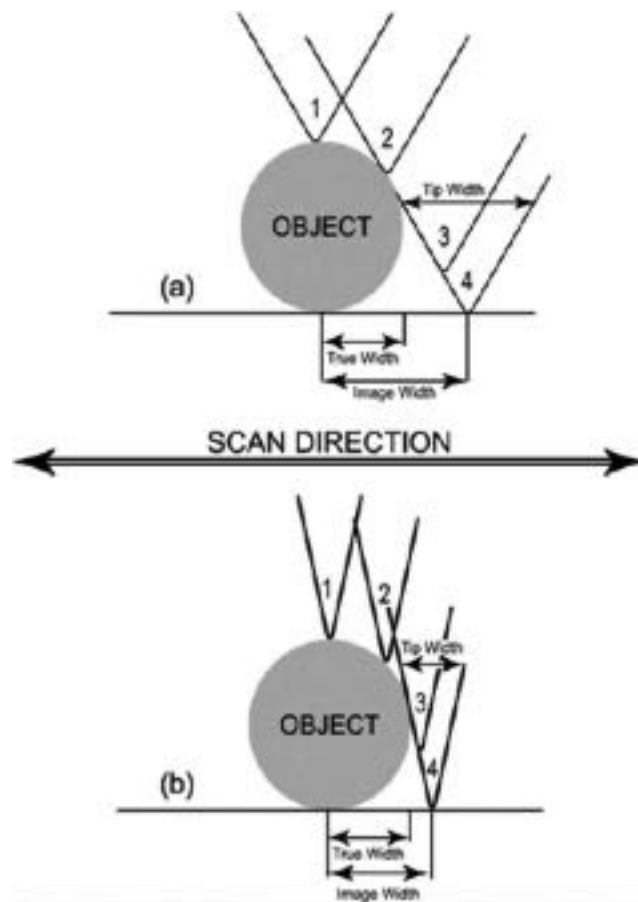


Fig. 2 Schematic illustration of the convolution of the shape of the AFM tip with the shape of the feature or particle being scanned. The side of the cantilever tip contacts the object and begins to produce a deflection of the cantilever before the tip apex actually reaches the object. Similarly, the opposite side of the tip is still in contact with the object even after the apex itself has passed. Thus, the total deflection implies a virtual lateral dimension for the object that is greater than its actual dimension. The difference between the virtual and actual dimensions is a function of the width of the cantilever tip. The sharper the tip, the more accurate the observed dimensions and the greater the resolution attainable.

Although AFM was initially employed to study the crystals and the crystallization of small icosahedral plant viruses, atomic force microscopy (AFM) does not require that the virus be crystallized or otherwise ordered. It is applied to individual particles and does not yield an average structure over an entire population. It does not require that the virus have a symmetrical or uniform architecture, or even that all particles be the same in structure. Thus, it is equally applicable to small icosahedral viruses such as Brome Mosaic Virus and satellite tobacco mosaic virus, to helical viruses such as Tobacco Mosaic Virus, to completely irregular, complex viruses like Vaccinia, or the retroviruses. There is no size restriction. It has been used to analyze small plant viruses to massive icosahedral viruses such as PBCV-1, an algal virus and mimi virus, one of the largest viruses known.

Initial Observations From AFM

A valuable qualitative result that emerges almost immediately from AFM images is what the virus looks like, what is its overall architecture, and how similar are particles to one another. Are they uniformly the same in appearance, or are there a variety of forms? Even a cursory AFM investigation may quickly reveal certain general features that allow rapid classification, as illustrated by the various structural classes of viruses shown in Fig. 3. The virions may be spherical, cylindrical, or filamentous. They may have symmetrically arranged capsomeres or other surface units, fibers, protruding vertices, prolate or icosahedral shapes, unusual morphologies, pleiomorphic character, etc. Tail assemblies may be observed directly, as on phages for example. AFM is, therefore, a useful tool for simply deducing the kind of virus one is dealing with, whether more than one kind of virus is present in a population, and the general level of contamination that may accompany the virus; cellular material, degraded virions, and macromolecular impurities of all sorts.

A fundamental parameter for virus particles is their diameter if they are spherical viruses, or their diameter and length if they are helical. AFM can provide measures of these in both the hydrated and dried state, which also gives an estimate of the degree of shrinkage they undergo as a result of dehydration. Because of the finite tip size, and tip-to-tip variation in radius of curvature, it is

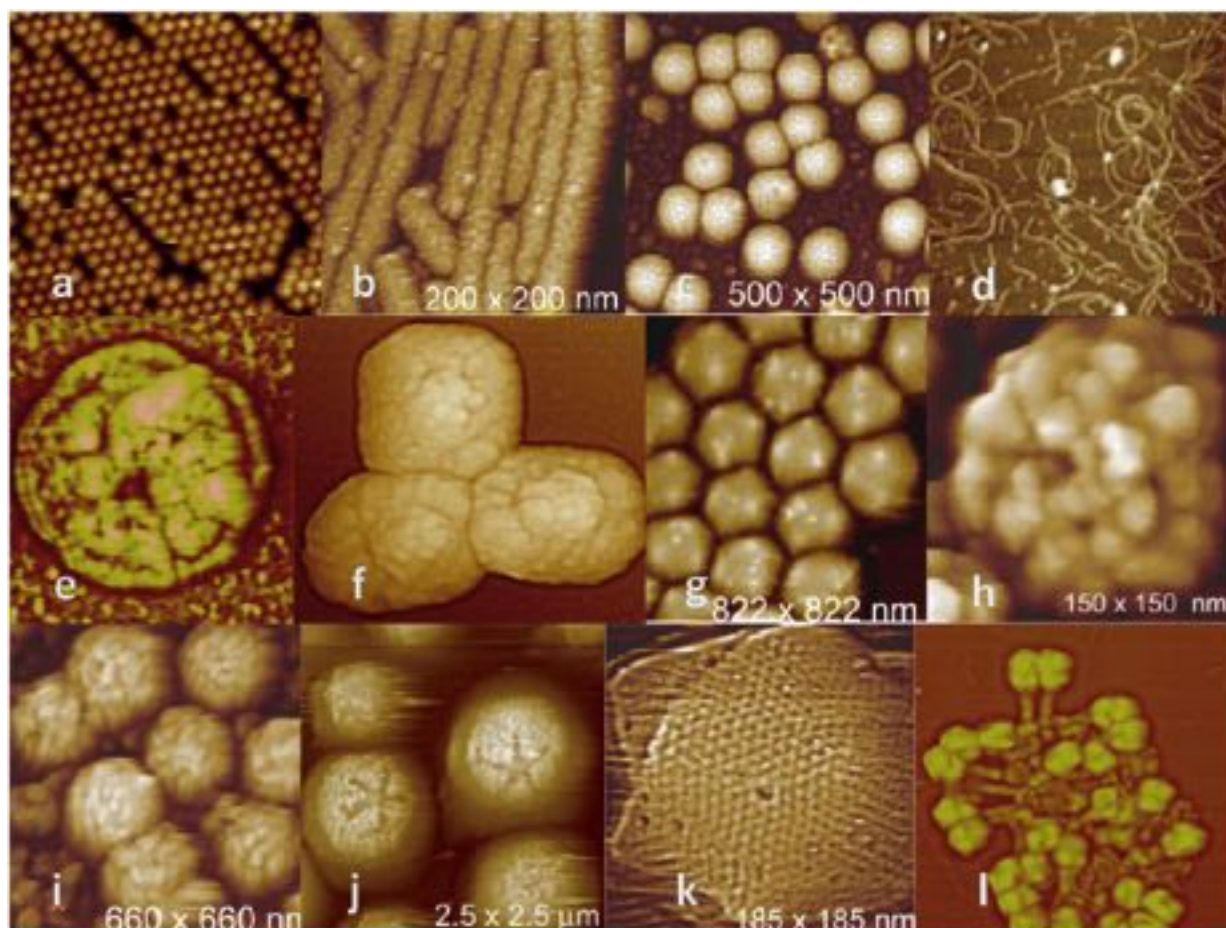


Fig. 3 *Virus shapes and forms.* (a) Crystal of satellite tobacco mosaic virus, a $T = 1$ icosahedral plant virus. (b) Rod shaped tobacco mosaic virus having a helical architecture. (c) Ty3 retrotransposons, a proto-retrovirus of icosahedral design. (d) Filamentous bacteriophage that infect bacteria in the Pacific Ocean. (e) A single virion of the membrane encapsidated, RNA containing SARS virus responsible for severe human respiratory disease. (f) Three virions of vaccinia virus with their prominent "lateral bodies" and coatings of hair-like fibers. (g) Tipula Iridescence Virus, a large membrane encapsidated, icosahedral virus that infects insects. (h) A single virion of HIV that produces AIDS in humans exhibiting a surface of Gp120 protein puffs. (i) Virions of Moloney-Mouse Leukemic virus. (j) Virions, nearly 800 nm in diameter, of the largest virus known, mimi virus. The virus is immediately recognizable by the remarkable star shaped structures it displays at one vertex of its icosahedral capsid. (k) PBCV-1 algal virus, a large virus of icosahedral design. (l) Tailed cyanophage that infect photosynthetic bacteria in both fresh and salt water bodies. Here they are seen clustered about and infecting a single bacterial cell.

risky to measure linear dimensions directly by AFM as was discussed above. It is, however, safe to measure the heights of objects above the substrate plane, and the distances between the points of maximum elevation (e.g., capsomere to capsomere) on particles, or center-to-center distances (e.g., particles in a crystal or in a cluster). This last approach can be applied to noncrystalline or paracrystalline arrays of viruses, both helical, rod shaped, and spherical viruses.

For spherical and cylindrically symmetric particles, measurements of particle heights above the substrate plane yield reasonably accurate values for their diameters and individual measurements are usually accompanied by rather modest error, generally of the order of 5% or less. By repeating measurements for a number of particles in the field, and using different scan directions, good statistics can be obtained, and histograms of size distributions compiled. Precision of a few angstroms is possible. Histograms of particle sizes are often informative. If the distribution is a simple Gaussian, then it can be presumed that particles of only one general morphology, or icosahedra of only one triangulation number are present, but that their diameters vary to some degree about the mean, perhaps due to physiological state or degree of maturation. If, on the other hand, a more complex distribution is observed, one having multiple peaks and shoulders, then particles of separate classes may be present. From such a histogram analysis it was deduced, for example, that Ty3 retrotransposon particles existed as icosahedra having triangulation numbers 3, 4, and 7.

As shown by **Fig. 3**, the surfaces of virus particles vary topologically as a function of their composition and architectures. Plant viruses, for example, generally exhibit protein capsids with few embellishments, and this is true of many simple animal viruses, such as poliovirus, and filamentous, rod shaped, and non-tailed bacteriophages as well. These geometrical capsids are generally based on icosahedral or helical architectures, and clusters of coat protein subunits, or capsomeres, are symmetrically distributed. Complex animal viruses on the other hand, though they may contain an icosahedral capsid in their interior, often have either a lipid

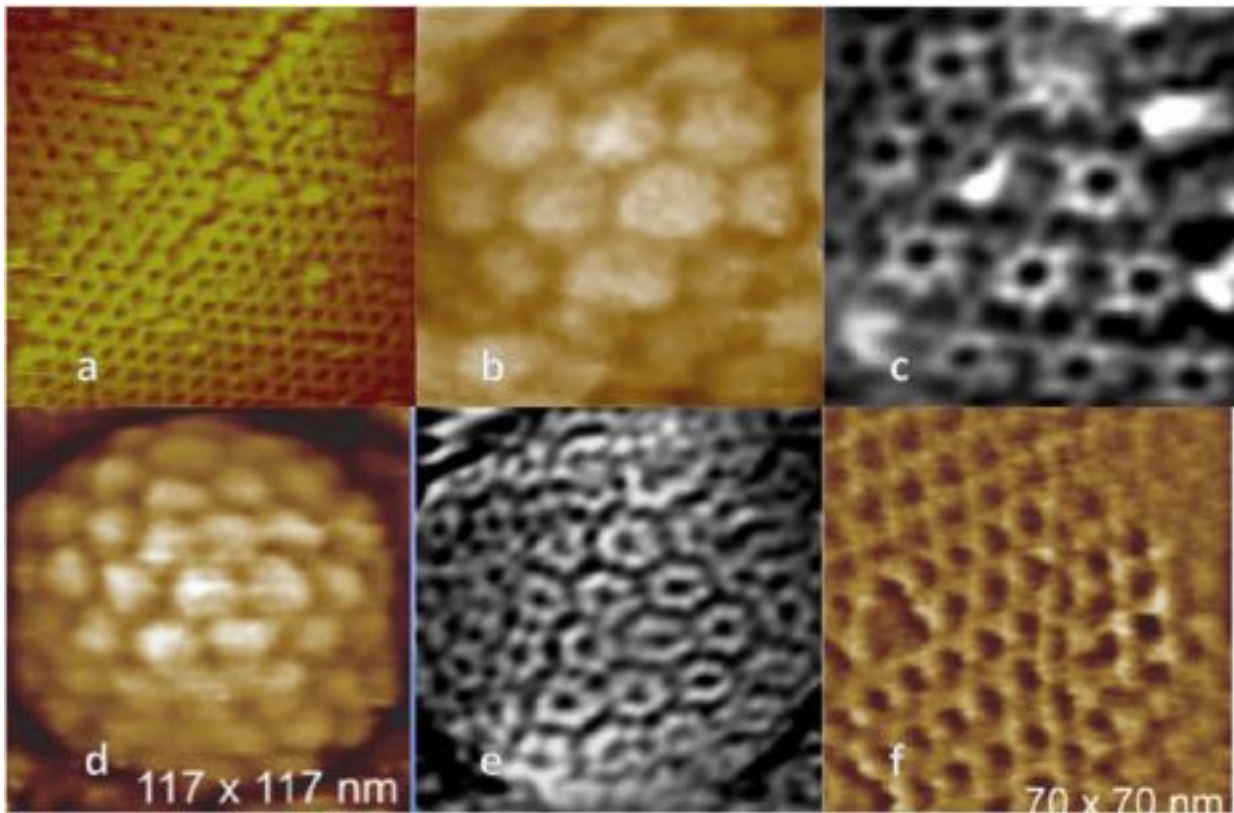


Fig. 4 Icosahedral arrays on virion surfaces. (a) A honeycomb arrangement of trimeric major capsid protein making up the capsid of mimivirus. (b) A single virion, about 30 nm across of the icosahedral plant virus Turnip Yellow Mosaic Virus displaying a $T = 3$ arrangement of hexameric and pentameric capsomeres. (c) Hexameric capsomeres are separated from one another by trimeric proteins on the surface of the icosahedral capsid of a cyanophage. (d) The icosahedral virion of a Ty3 retrotransposon, a proto-retrovirus. (e) The icosahedral DNA filled head of a T4 bacteriophage. (f) Trimeric protein capsomeres composing the large faces of the icosahedral PBCV-1 algal virus.

membrane over their surface, a covering of protein clusters, or even a fur like coating of fibers. These characteristic surfaces are usually apparent by AFM. With the aid of some histological procedures, such as osmium tetroxide fixation, or protease treatment, can be identified and delineated with a high degree of precision. Some examples of icosahedral capsids and cores are presented in Fig. 4.

Icosahedral capsids, or bullet shaped or elongated capsids based on that symmetry, can be characterized in terms of the structure of the fundamental capsomere, along with the icosahedral triangulation number, T . In many cases the exterior shell of a virus may not be icosahedral, but it might possess an inner capsid that is. For example, though membrane covered and of pleiomorphic external shape, herpes simplex possesses a nucleic acid containing capsid of icosahedral form $T = 16$. Mimivirus exhibits a complex outer surface coated with a forest of fibers, but it too contains an icosahedral core with a triangulation number lying between $T = 324$ and 381 . The T number, then, contains half of the information one needs to describe an icosahedral capsid.

Triangulation Numbers of Icosahedra

The triangulation numbers of icosahedral viruses can often be deduced from AFM images like those in Fig. 4. In the case of Ty3 retrotransposons, a proto retrovirus, it was found that virions existed in three different diameters, and three different architectures corresponding to T numbers 7 (the largest fraction of the population), 4, and 3. AFM images of the surface capsomere distribution could be triangulated visually, as in Fig. 5, by defining the arrangement of hexamers with respect to pentamers. The T number was thereby defined. Of particular interest were the $T = 7$ particles which could conceivably exist in either of two enantiomorphs, d and l . Because height information is preserved in AFM images, so is handedness. Thus it was possible in the case of the $T = 7$ Ty3 virions to determine that their actual T number was $7d$.

A somewhat different approach must be taken with very large icosahedral capsids, like some of those in Fig. 4, which include PBCV-1 and mimivirus. For these viruses it is necessary to determine two indices h and k , which define T ($T = h^2 + hk + k^2$), by following a row of hexagonal capsomeres from one pentagonal vertex to the next icosahedral edge, and counting the number of capsomeres along one edge h and the other k (the h and k coordinates of the intersection point on the icosahedral edge). This approach was used to determine the T numbers of the iridovirus PBCV-1 and the capsid architecture of mimivirus. While the

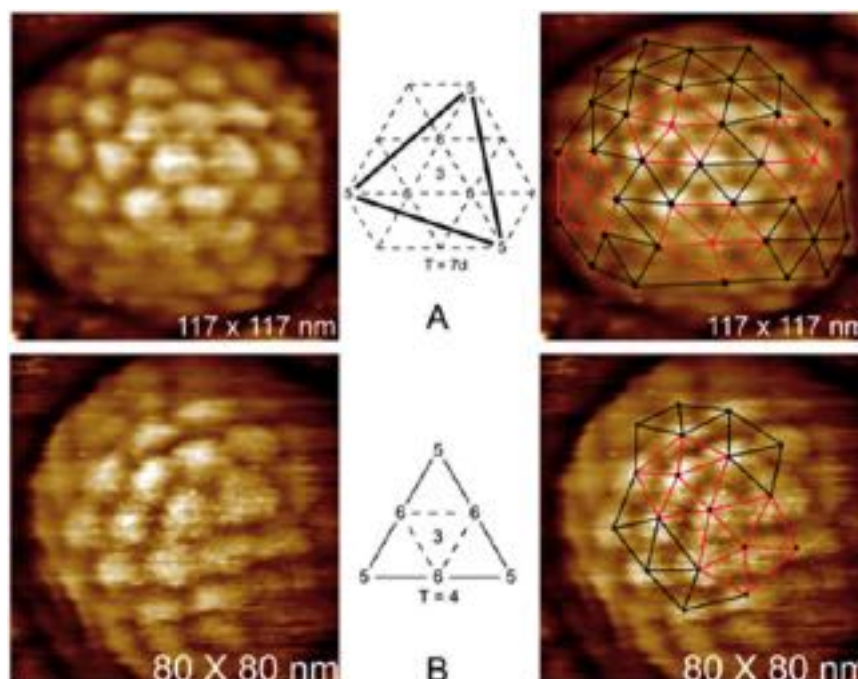


Fig. 5 The icosahedral symmetry of the Ty3 retrotransposon, a protoretrovirus that infects yeast, was established by AFM. It was further shown that Ty3 existed in three different icosahedral forms corresponding to T numbers of 3, 4, and 7. (A) The left panel shows a Ty3 particle, and the right panel shows the same particle with a T7 icosahedral net overlay. Pentagonal vertices are in red. The center diagram is that of a triangular face of an icosahedron showing positions of pentagonal and hexagonal capsomeres for T7. (B) Another, smaller particle of Ty3 and the equivalent overlay and diagram showing it to have T4 icosahedral symmetry.

T number describes the overall distribution of capsomeres on the surface of an icosahedral capsid the distribution of subunits within capsomeres can also sometimes be deduced by AFM analysis.

Analyses like this were used in the case of the large algal virus PBCV-1, shown in **Figs. 3(k)** and **4(f)**. Knowing the diameters of capsomeres is often of considerable importance, even when individual subunits cannot be resolved. In mimivirus, seen in **Figs. 3(j)** and **4(a)**, for example, capsomere diameter provided a crucial clue in delineating the capsid architecture, and permitting subsequent detailed analysis and reconstruction by cryo-EM. Although capsids of native HIV have yet to be visualized by AFM, helical tubes of capsid protein reassembled *in vitro* have, and one is shown in **Fig. 6(a)** along with examples of other reassembled virus particles (VLPs). In these tubes a hexagonal arrangement of coat proteins could be clearly seen, and this provided support for a capsid model based on modified icosahedral architecture.

Recombinant and Mutant Virus Particles

Not only native viruses, but also recombinant and mutant particles can, and are frequently recorded in AFM investigations. **Fig. 6** presents an array of examples. In **Fig. 6(a)** is seen a helical tube that is occasionally generated when the recombinant capsid protein of HIV is made and allowed to assemble in bacterial cells. In **Fig. 6(b)** are particles of recombinant Gag protein from Ty3 retrotransposons produced and assembled in *E. coli* that fail to assemble into icosahedral shells, and similar results with Mason-Pfizer Monkey Virus. **Fig. 6(c)** is a structure, rather common in preparations of T4 bacteriophage, a twin tailed virion next to a normal, single tailed virus. In **Fig. 6(d)** a lymphocyte infected with a mutant of Murine Leukemia Virus (MuLV) that lacks the ability to separate itself from the host cell membrane. It forms, instead, the extended comet-like protrusions that decorate the host cell surface. **Fig. 6(e)** is another mutant of MuLV that fails to make its normal (Gp120) surface protein, giving rise to what are termed “bald” particles. In **Fig. 6(f)** is a crystal of Brome Mosaic Virus, a small (28 nm, T = 3) icosahedral plant virus. What is striking is the inclusion in the crystal of a mutant of smaller diameter (likely a T = 1 particle), and a mutant of larger diameter (likely a T = 7 particle), illustrating the somewhat promiscuous nature of icosahedral shell assembly.

The helical tubes reassembled from HIV Gag protein in **Fig. 6(a)** reminds us that, although we have emphasized spherical particles to this point, helical and rod shaped structures having periodic sub structure are also excellent specimens for AFM analysis. These frequently appear in investigations of intact viruses, and even in studies of spherical viruses when their interiors are explored. While we tend to think of virus particles as geometric solids such as icosahedrons, or layered structures such as mimi, herpes, or vaccinia, they often include or produce substructures that represent completely different material forms. Among these are fibers of various sorts, and membranes.

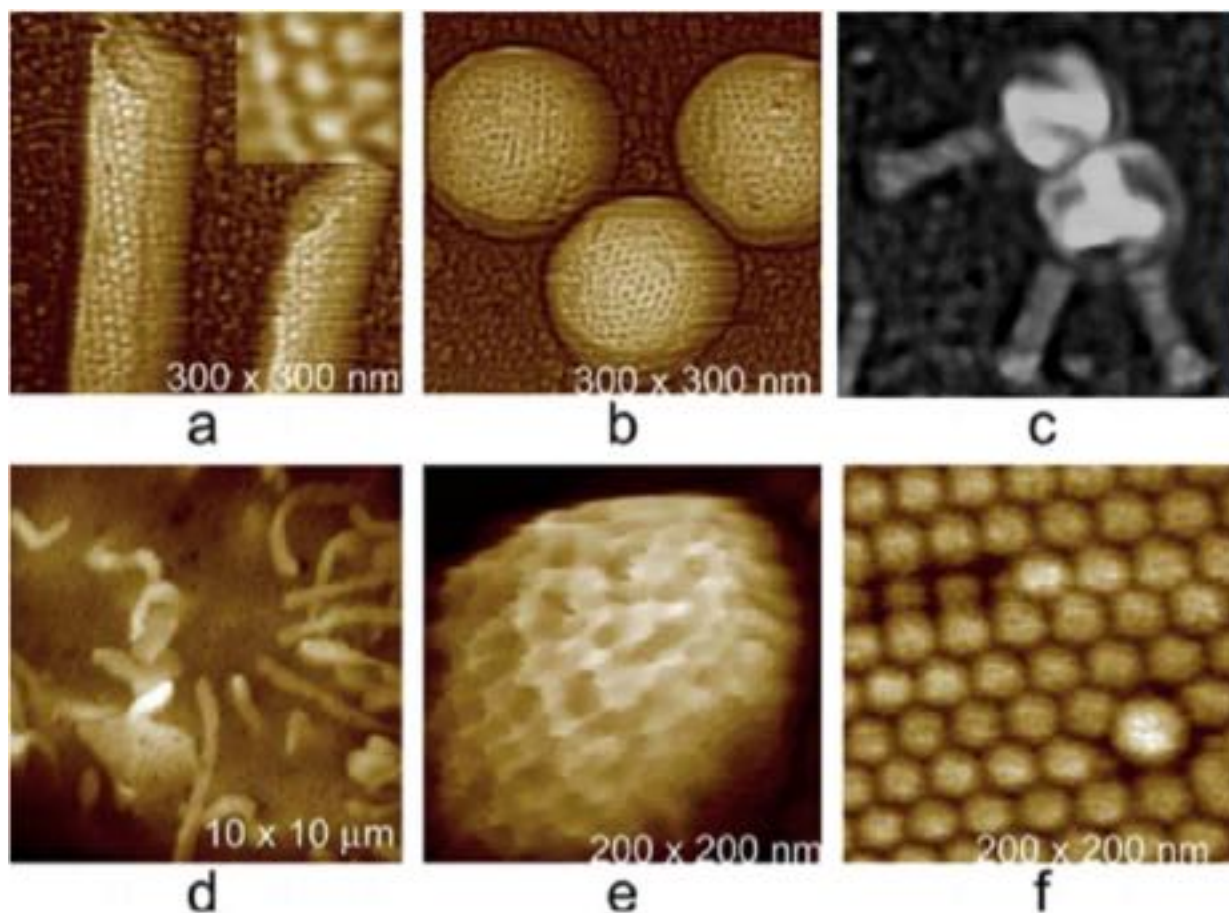


Fig. 6 *Mutants and recombinant particles.* (a) Helical tubes, a product of in vitro self-assembly of the Gag protein of human immunodeficiency virus. (b) Particles reassembled in vitro from a truncated form of the Gag protein from Mason-Pfizer monkey virus. (c) Preparations of T4 bacteriophage were found to commonly contain mutant virions like the double tailed example seen here next to a normal virion. (d) AFM image of 3T3 cells in culture that are infected with a mutant form (gPr80^{gag}) of Moloney murine leukemia virus. The virus, upon budding, is unable to separate completely from the host cell membrane and forms long, comet-like extensions from the cell surface. (e) A mutant of Moloney murine leukemia virus that, genetically, lacks the capacity to make envelope protein. As a consequence, it appears as a “bald” virus that exposes its limiting lipid membrane to the exterior. The undulations and variations of the membrane surface are an effect produced by local movement of the membrane in response to AFM tip pressure. (f) A crystal of brome mosaic virus. Though most of the crystal is composed of conformist T3 particles with the standard diameter of 30 nm. The crystal also contains an exceptionally small virion, probably a T = 1 icosahedra lacking RNA, and one exceptionally large virion, probably T4 or 7, likely containing multiple copies of RNA.

Virus Fibers and Membranes

In **Fig. 7** are examples of fibers, all of viral origin. They include the six tail fibers of a bacteriophage emerging from its hexagonal base plate, fibers that first engage the host cell surface, **Fig. 7(b)** are the surface fibers of mimi virus, and **Fig. 7(c)** those of vaccinia, a coating that gives the intact viruses the appearance of fur covered particles in AFM images. In **Fig. 7(d)** are fibers of unknown function that are prolific in AFM images of disrupted mimi virus. They resemble collagen fibers, but exhibit a different helical periodicity of 7 nm. The 3 + 1 decorated fibers of some cyanophages in **Fig. 7(e)**, though prominent, remain of mysterious but intriguing function that may select “life at sea” and the search for hosts in a dispersed environment. **Fig. 7(f)** contains the DNA strands from a T4 bacteriophage.

Fig. 8 presents three examples of viral membranes encountered in AFM investigations of viruses. In **Fig. 8(a)** the icosahedral core of a mimi virus is encapsidated by a thin, presumably lipid membrane surrounding it like a corona. In **Fig. 8(b)** the membrane sheet of a herpes virus has been partially pulled away to reveal the underlying T=25 icosahedral core. In **Fig. 8(c)**, following expulsion of cores, the membrane coverings of vaccinia, heavily studded with protein aggregates, litter the AFM substrate.

With AFM, it is not essential that highly purified virus particles be used as samples, although that might be ideal. Moloney Mouse Leukemia Virus particles are accompanied by a background of protein molecules in **Fig. 3(i)**, for example, but they are nonetheless distinct in their features. The same was true of the TMV in **Fig. 3(b)** and the phage particles in **Fig. 3(l)**. Because individual particles can be investigated whenever a good specimen is spatially distinct from the surrounding rubble of proteins,

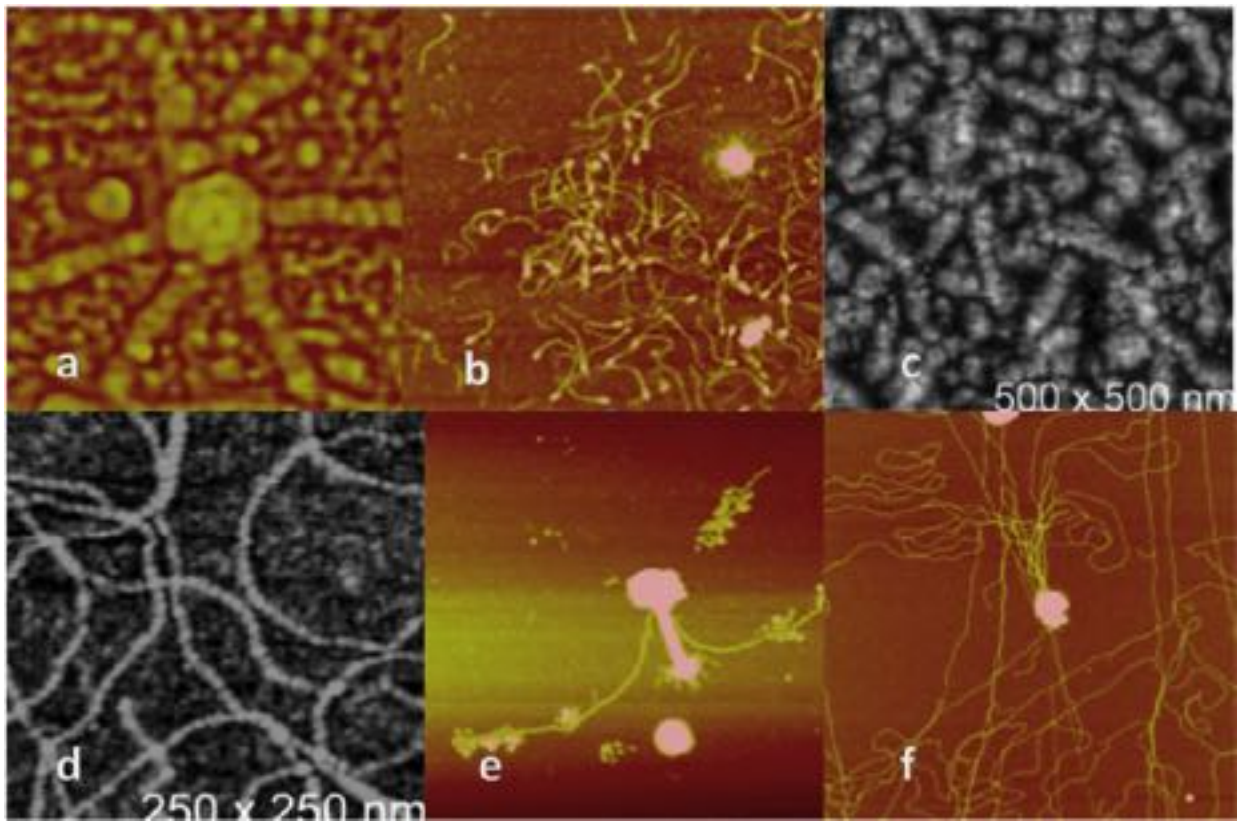


Fig. 7 *Viral fibers.* (a) A view along the axis of a bottom segment of a cyanophage tail assembly that is standing upright on the AFM substrate. The tail fibers are played out, showing their segmented character. (b) The surfaces of mimi virus are coated with fibers of uniform length having a distinctive head like those shown here. (c) Vaccinia virus is coated with the short, thick fibers seen here, discarded from the virions as they decompose. The fibers give the vaccinia the appearance of being coated with fur. (d) AFM images of unique fibers, having a pronounced 7-nm repeat along their lengths, that were observed in abundance when mimivirus capsids were disrupted. (e) A class of cyanophage has two fibers originating at the collar between the icosahedral head and the helical tail. The fibers are always decorated by the 3 + 1 arrangement of distinctive, attached protein masses near their distal termini. (f) The bright mass near the center of the image is a cyanophage head that has burst. The strands of DNA are seen emerging from the head and spreading on the substrate.

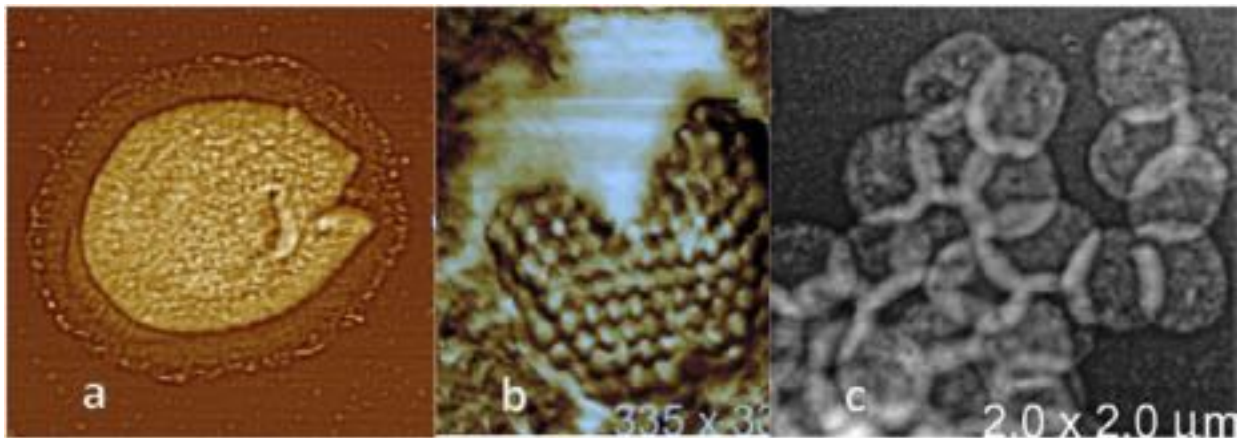


Fig. 8 *Viral membranes.* (a) A mimi virus, with its fivefold aperture in the process of opening to release its sac of DNA, is seen enclosed in a thin membrane envelope. (b) The white sheet of the membrane coating a herpes simplex virus has "splashed" to one side of the capsid and lies on the substrate plane. (c) Vaccinia virus cores are encapsidated within membrane envelopes, which are heavily decorated with proteins and their aggregates. Upon expulsion of the cores the empty envelopes remain and are seen here.

cellular debris, and biological detritus, it may still yield excellent images. A problem, however, is that biological debris often adhere to, and foul the AFM tip, and this severely degrades the quality of images.

Infection and Budding From Host Cells

Viruses on the surfaces of host cells, or within the cytoplasm, can be visualized as well as free particles, and sometimes with superior acuity, because they are better immobilized. They may be seen entering cells upon infection, or budding from cells after replication and assembly. This often provides valuable insights into which cells in a population are producing virus, the distribution of virus particles on the surface of the cells (e.g., are there preferred sites for budding?), and some details of the budding process itself.

In **Fig. 9(a)** and **(d)**, for example, human lymphocytes in culture are seen virtually coated with HIV budding from their plasma membranes. Investigation of a large population of cells, however, showed that this was an unusual occurrence, in that most cells exhibited only a dozen or less HIV on their surfaces. Thus only a few cells seemed to account for the great majority of virus. Algal cells have both a cell membrane and a cell wall that must be breached in order for cyanophage to enter. In **Fig. 9(b)** an algal cell is under attack from a host of cyanophage. **Fig. 9(c)** is an AFM image of a disrupted amoeba cell infected with mimivirus, its nucleus in the lower left. The cell contains on the order of a thousand new particles.

Specific and Special Structural Features

Special features abound when more complex viruses are imaged by AFM, and these often, but not always, are found at specific, specialized vertices of icosahedral viruses, or their cores. In **Fig. 10(a)** is shown the pentagonal star at a unique fivefold vertex of a PBCV-1 algal virus and in **Fig. 10(b)**, the far more complicated star shaped structure that fastens shut the internal cavity of mimivirus that contains its DNA. In this case the function of the star has been deduced primarily through AFM investigation and

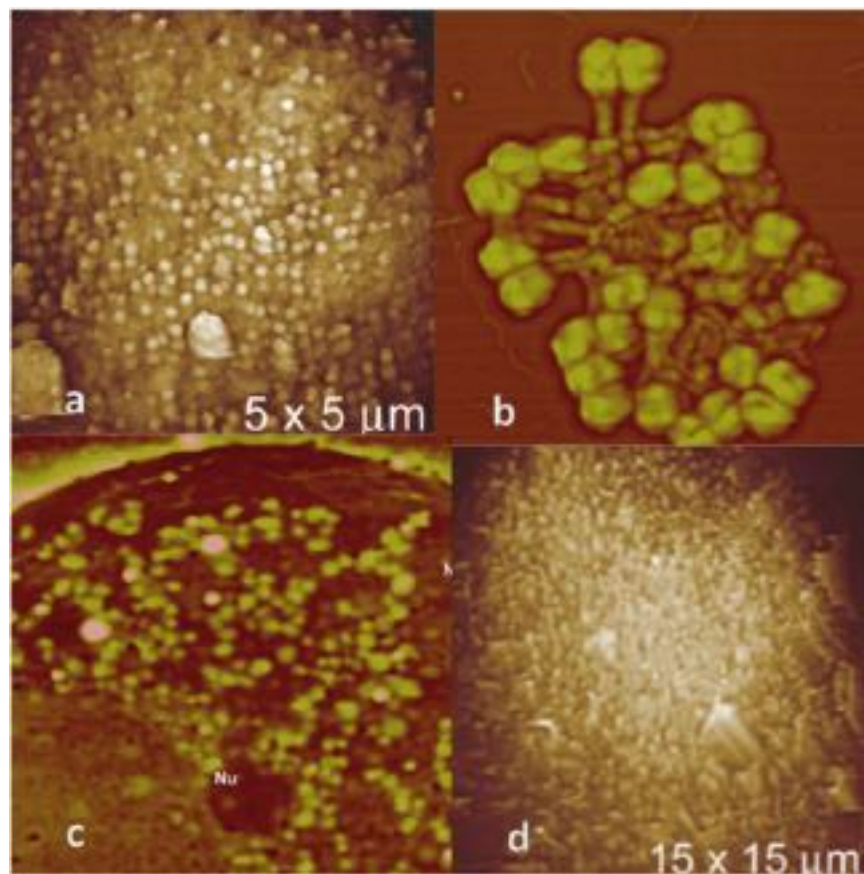


Fig. 9 Virus infected cells. (a) and (d) are human lymphocytes grown in culture that are infected with HIV. The budding virus appears to literally coat the surfaces of these super-infected cells. (b) The icosahedral heads of cyanophage are seen at the surface of an infected bacterium, their tails engaged with the cell surface. (c) An amoeba cell infected with mimi virus has been disrupted allowing the AFM to record the contents. The nucleus of the cell can be seen occupying the lower left. The cell is filled with new virus particles emerging from the several “virus factories” in the cytoplasm.

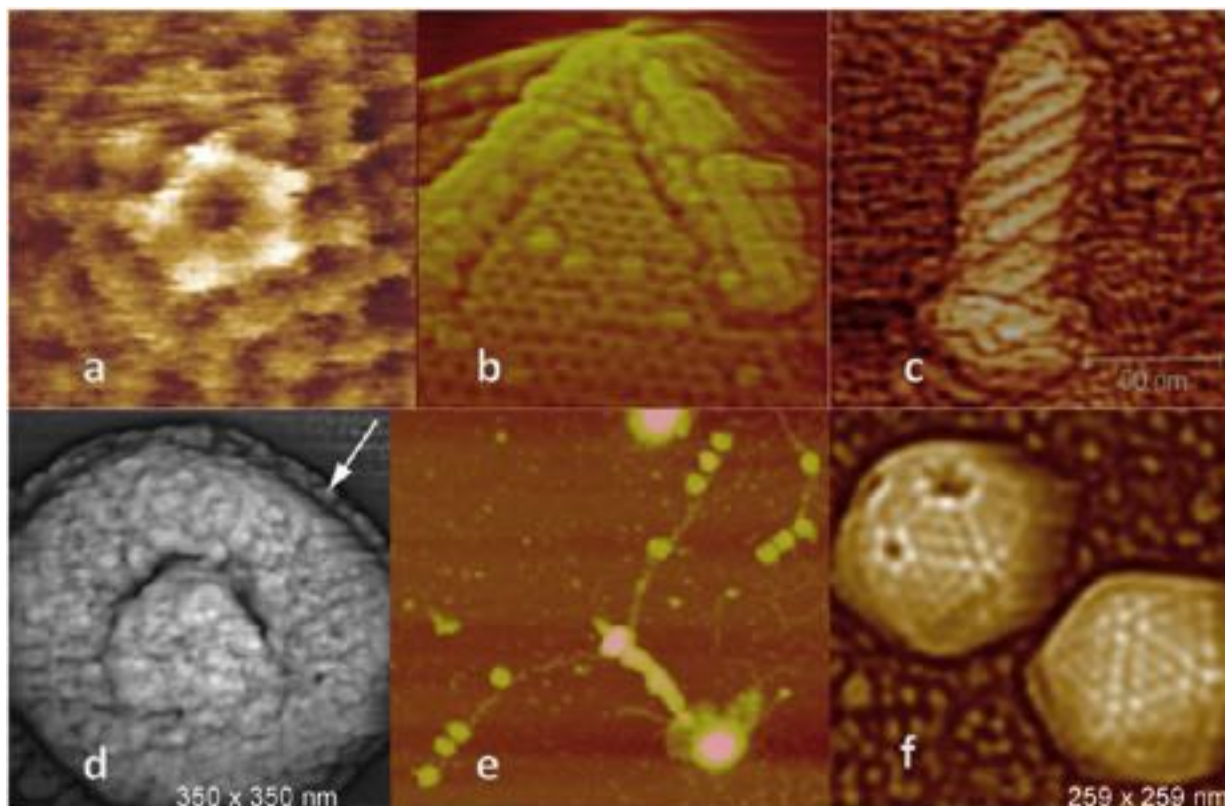


Fig. 10 *Special features of complex viruses.* (a) A specialized cluster of five proteins, with another protein at the center, that occupy the 12 unique vertices on the surface of PBCV-1. All of the proteins in these clusters are clearly different in structure from the normal capsid protein (b) Image of the “stargate,” five-vane apparatus found at only a single pentameric vertex of mimivirus. The apparatus provides a mechanism for the release of the encapsidated DNA. (c) Helical tail assembly from bacteriophage T4 that is responsible for injection of the phage DNA into its host. (d) Air-dried vaccinia virus. Retraction of the membrane and overall shrinkage emphasize the lateral body at center. A unique feature of vaccinia virus is the presence of two such lateral bodies associated with each virion. (e) A cyanophage displays two unique fibers decorated in a $3 + 1$ fashion with large protein aggregates. The fibers always have the same lengths, four aggregates, and consistent intervals between the aggregates. (f) AFM image of adenovirus. A noteworthy feature of adenovirus is that degradation of the virions invariably initiates and proceeds with loss of pentons at the 5-fold vertices, leaving particles perforated in an icosahedrally symmetric manner.

shown to be the gateway for delivery of the nucleic acid. The helical tail of prominent pitch, a tube essential for injection of nucleic acid and that is joined to a singular five fold vertex on the head of a $T = 7$ icosahedral head of a bacteriophage is shown in [Fig. 10\(c\)](#). In [Fig. 10\(d\)](#) is an example of the prominent, embedded “lateral body” of a vaccinia virus, and in [Fig. 10\(e\)](#) the curious $3 + 1$ decorated fibers attached to the head of a cyanophage. In [Fig. 12\(f\)](#) a surface fiber important in host recognition, has been pulled from a fivefold vertex of an adenovirus.

Dissection of Complex Viruses

One might think that because AFM provides images of the surfaces of objects, and does not peer into their interiors, as do X-ray diffraction and electron microscopy, that they would be of little value in delineating the interior structure of viruses, the layers beneath the external surface. This is not true, however, as we can apply the same technique that has been used by anatomists for centuries, dissection. With the aid of chemical, enzymatic, and physical tools, we can systematically pare a complex entity, including a virus, down to its core, layer by layer. At each stage, AFM may then be used to visualize what remains, and what has been removed as well.

This approach is particularly effective with large, complex viruses such as Vaccinia Virus, as illustrated by [Fig. 11](#), or Mimivirus. With these large assemblies, ordered and disordered protein shells, lipid membranes, and the nucleic acid within can be revealed and analyzed. By deconstruction, the architecture of particles is revealed, and, at the same time, the kind of biochemical interaction that maintains each level of structure is delineated as well.

Among the most useful agents for chemical dissection have been detergents, usually 0.5%–2% of some non-ionic detergent such as NP40, and reducing agents such as DTT or DTE. The former causes protein structure to gradually unravel, and detergents strip away lipid membrane. The latter reduce disulfide bonds and liberate polypeptides otherwise bonded to one another.

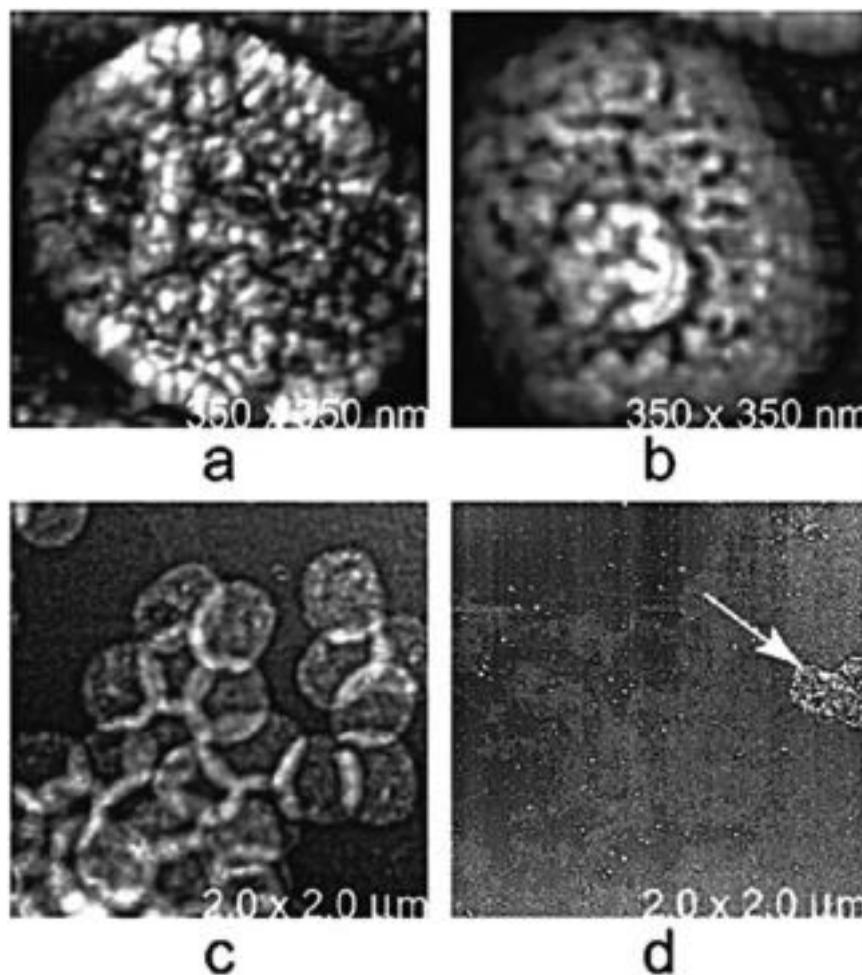


Fig. 11 Four stages in the chemical and enzymatic dissection of vaccinia virus. (a) An intact vaccinia virus virion. (b) The core of the virus obtained by treatment with a nonionic detergent (NP-40) and dithiothreitol, a reducing agent. (c) The ghosts, or capsules, that remain after the cores have been treated with proteases, which also produces release of the viral nucleic acid. (d) Masses of DNA released onto the AFM substrate by disrupted viral cores. The arrow indicates a disrupted vaccinia virus core adjacent to the DNA.

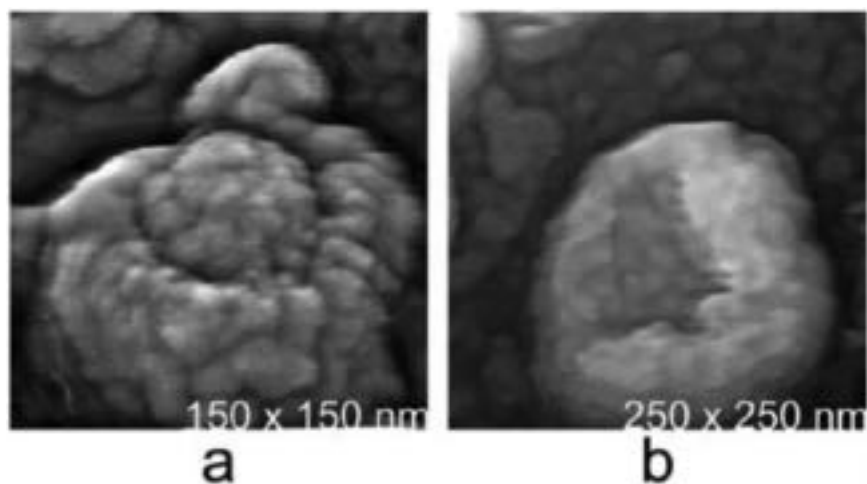


Fig. 12 Damaged retroviral virions. Retroviruses are fragile and often are disrupted or damaged by the shear forces of preparative centrifugation alone. These damaged particles frequently expose, for AFM visualization, internal structure and reveal unique elements of the virus architecture. (a) A virion of Moloney murine leukemia virus where the outer shell of envelope protein, lipid membrane, and matrix protein have been partly torn away to reveal, like the kernel of a nut or the pit of a fruit, the nucleic acid-containing capsid that it enshrouds. From such images we can obtain quantitative measures of the thickness of the virus shells and the diameters and structural character of the capsid within. (b) A human immunodeficiency virus that has been shorn during centrifugation and has completely lost the protein capsid that it contained.

Disulfide bond reduction appears to be particularly important in large, complex viruses where such covalent linkages cross-link coat proteins and stabilize capsids.

In some cases, non-ionic detergents are insufficient to disrupt structure and more vigorous ionic detergents such as SDS must be used. There is difficulty with SDS, however. It tends to have an all or none effect, so that upon reaching a concentration sufficient to disrupt viruses, it completely degrades them uncontrollably. SDS can also produce artifacts due to drying on the substrate.

The most effective enzymatic tools have been proteases that degrade polypeptides. These are particularly useful because they have a range of activities and a spectrum of specificities. As a consequence, a whole variety of proteases have been employed, and these include trypsin, bromelain, proteinase K, subtilisin, and mixtures of pancreatic proteases. Viruses are usually exposed to the proteases for anywhere from 15 min to several hours, or even overnight, at concentrations 0.5 mg/ml to as high as 5 mg/ml. The proteases must be washed from the virions with buffer or water before imaging as they otherwise produce a dense, irregular background that makes imaging problematic, and they foul the cantilever tip.

Physical forces have also been used to disrupt viruses, and often fortuitous perturbations, resulting simply from preparation and handling, have proven structurally illuminating. Heat, for example, was used to open TYMV to release its encapsidated RNA, and direct physical pressure on mimivirus sandwiched between two layers of atomically smooth, cleaved mica, as well. There are also instances where “hammering” of individual particles with the AFM tip has been utilized, taking advantage of the fact that AFM can serve as a tool as well as an imaging device.

It is occasionally unnecessary to actually treat viruses with any chemical or biochemical agent to see inside, as the physical stress of preparation and purification may result in damaged or partially degraded particles. These may expose interior structural features otherwise not apparent. Retroviruses, in particular, are physically fragile. Some MuLV, as shown in Fig. 12, when subjected to the shear forces of centrifugation, lose portions of the shell surrounding the capsid. This permits direct visualization of the virus core still embedded within the layers of envelope and matrix protein.

HIV is another example where even the mildest procedures produce some damaged virions. Although the cores of HIV have not yet been visualized by AFM, likely due to their fragility, the remainder of the virus, those without the cores have. Such partially disrobed particles, both MuLV and HIV, provide specimens that can be subjected to quantitative examination and thereby yield the dimensions, the thicknesses of internal layers of structure, and they give some clues as to their components as well.

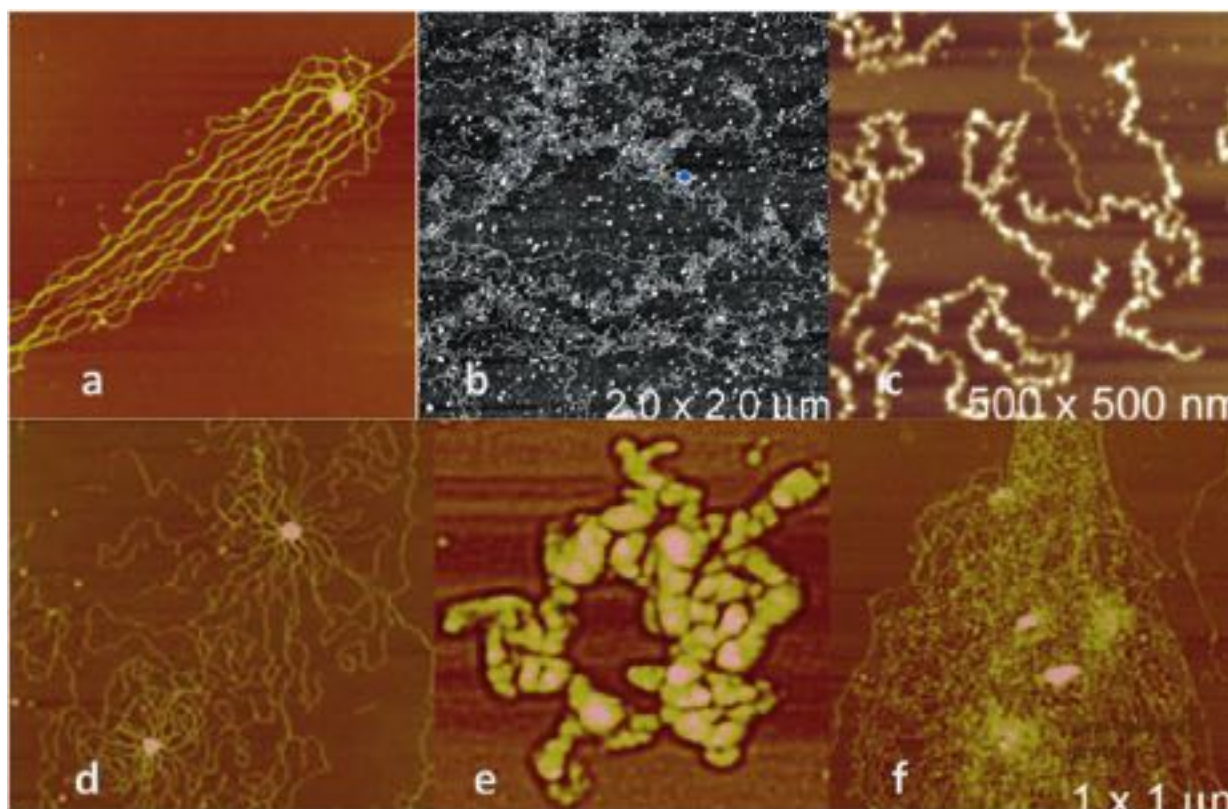


Fig. 13 *Nucleic acid.* (a) Strands of DNA spreading on the AFM substrate from a broken icosahedral head of a cyanophage. (b) Masses of DNA released by vaccinia cores upon their decomposition. (c) RNA lost from chemically degraded satellite tobacco mosaic virus. (d) DNA emerging from burst heads of T4 bacteriophage. (e) RNA heavily complexed with proteins is seen to form the “pinwheel” structures shortly after decomposition of the restraining shells of the SARS virus. (f) A mass of DNA studded with proteins pours from a single mimi virus after it disgorges its sac of the nucleic acid.

The best example of a complete dissection of a complex virus using AFM is that of Vaccinia virus, a pox virus of about 300 nm diameter that is delimited by a lipid membrane. It contains a double stranded DNA genome bounded by several protein shells. It also has associated with its inner core two unusual protein assemblies of still unknown function, known as lateral bodies.

As illustrated in Fig. 11, vaccinia could be sequentially degraded with 0.5% NP40 non-ionic detergent combined with 0.05 M DTT, followed by exposure to this same mixture but containing either trypsin or proteinase K, or to the proteases alone. Four stages in this process were presented in Fig. 11. In the end, the innermost core was breached and the DNA was exposed. From images of the DNA emerging from the core, it was deduced that while some portion of the encapsidated DNA was heavily integrated with protein, the vast majority was largely naked, with only occasional associations with protein.

Visualization of Viral Nucleic Acid

The nucleic acids of viruses, from a structural standpoint, are of considerable interest, and in particular, how they are condensed and packaged inside capsids and cores. Clearly, packaging is accomplished in different ways by specific families of viruses. It is unlikely, for example, that bacteriophage and pox virus package their genomic double stranded DNA the same way. The packing densities of the nucleic acid differ by more than tenfold. Nor is it likely that large, single stranded RNA containing viruses, such as retroviruses, package their genomes the same way as do $T = 1$ or $T = 3$ icosahedral viruses. Certainly, helical and filamentous viruses use entirely different mechanisms.

AFM investigations have been conducted on RNA extracted by phenol from a series of small icosahedral viruses, and from Tobacco Mosaic Virus, the classical rod shaped, helical virus. The spherical viruses included poliovirus, Satellite Tobacco Mosaic Virus, Turnip Yellow Mosaic Virus, and Brome Mosaic Virus. In these studies, the gradual unraveling of the tertiary structure of the RNA, and ultimately the secondary structure as well, could be produced in stages simply by heating. A counter example was provided by Tobacco Mosaic Virus RNA, which appeared initially as a thread, a completely extended molecule lacking any

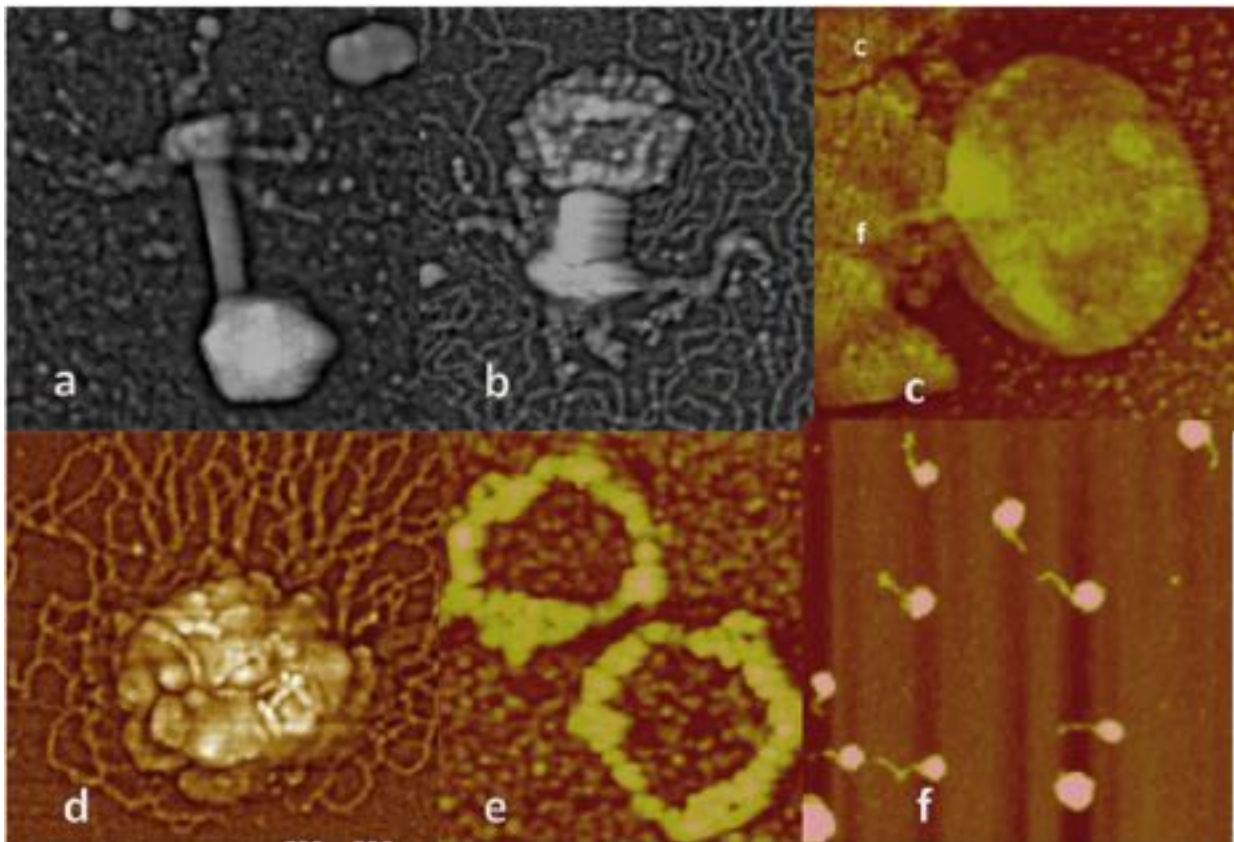


Fig. 14 *Delivery of viral nucleic acid.* (a) and (b) show cyanophage before (a) and after delivery through the tail injection mechanism. The tail is fully extended in (a) and compressed in (b). After ejection, the particle in (b) is surrounded by stands of DNA on the AFM substrate. (c) A remarkable image of a mimi virus icosahedral core disgorging its sac of DNA through the open "stargate". (d) The core of a vaccinia virus is seen in a state of decomposition which allows subsequent release of massive quantities of, initially, highly self entwined, cables of DNA. (e) Distinctive rings of RNA complexed with proteins that form as the SARS virus releases its genome. (f) Simultaneous emission of RNA from satellite tobacco mosaic virus upon gentle treatment with phenol.

secondary structure. With time, it began forming local secondary structural elements and eventually condensed into forms similar to those seen for the RNA from the icosahedral viruses. The upshot of the study was that the single strands of RNA spontaneously condensed as linear arrangements of stem – loop substructures following synthesis, that the condensed RNA bound coat protein to it, and the two cooperatively coalesced into the completed particle. In those studies, AFM proved itself an able technique for directly visualizing nucleic acid structure, demonstrating its fluidity, and suggesting the mechanisms by which it is encapsidated.

DNA and RNA have quite different appearances in AFM images, and this is evident in **Fig. 13**, which presents both kinds of nucleic acid. In **Fig. 13(a)** and **(d)** are seen the genomes of T4 and cyanophage, respectively, emerging from viral phage heads. In **Fig. 13(b)** are enormous masses of DNA released from the cores of only a few vaccinia virus. In **Fig. 13(f)** DNA pours forth from a single mimi virion, the nucleic acid highly complexed with proteins. In **Fig. 13(c)** RNA from disrupted satellite tobacco mosaic virus (1058 nucleotides) are displayed on the AFM substrate. In **Fig. 13(e)** The RNA genome of a SARS virus, heavily associated with protein, forms a characteristic pinwheel structure as the containing shell disintegrates. DNA, in general, looks like strands and coils of rope lacking any higher levels of structure. RNA on the other hand appears as complicated, linear sequences of self-involved secondary structure. Sometimes, however, the distinction is not entirely clear and further evidence may be needed to show whether a filament or strand or complex is DNA or RNA through the use of specific enzymes.

In investigating virus infected cells and viruses engaged in the infection of host cells, occasionally virus can be observed passing their nucleic acid into the host cell. Sometimes it is possible to infer mechanistic features of DNA or RNA delivery. **Fig. 14** presents some examples. Most familiar is the injection of DNA into a host bacterial cell by a tailed bacteriophage. **Fig. 14(a)** shows a cyanophage with its head still filled with DNA and its tail sheath fully extended. In **Fig. 14(b)** another cyanophage has a fully compressed tail sheath, the head now an empty shell, the nucleic acid has been ejected, and all around the phage are strands of its DNA.

Fig. 14(c) shows a remarkable event captured by AFM. A mimi virus core is seen expelling an encapsidated, membrane sac containing its DNA through the now wide open, fivefold mouth. The opening is produced by the dissolution of the mimi star structure (**Fig. 10(b)**) and the folding back of the leaves of the vertex. In **Fig. 14(d)** a near totally disassembled vaccinia core is expunging its massive load of DNA. In **Fig. 14(e)** the RNA of a SARS virion, still heavily complexed with proteins, forms a circular,

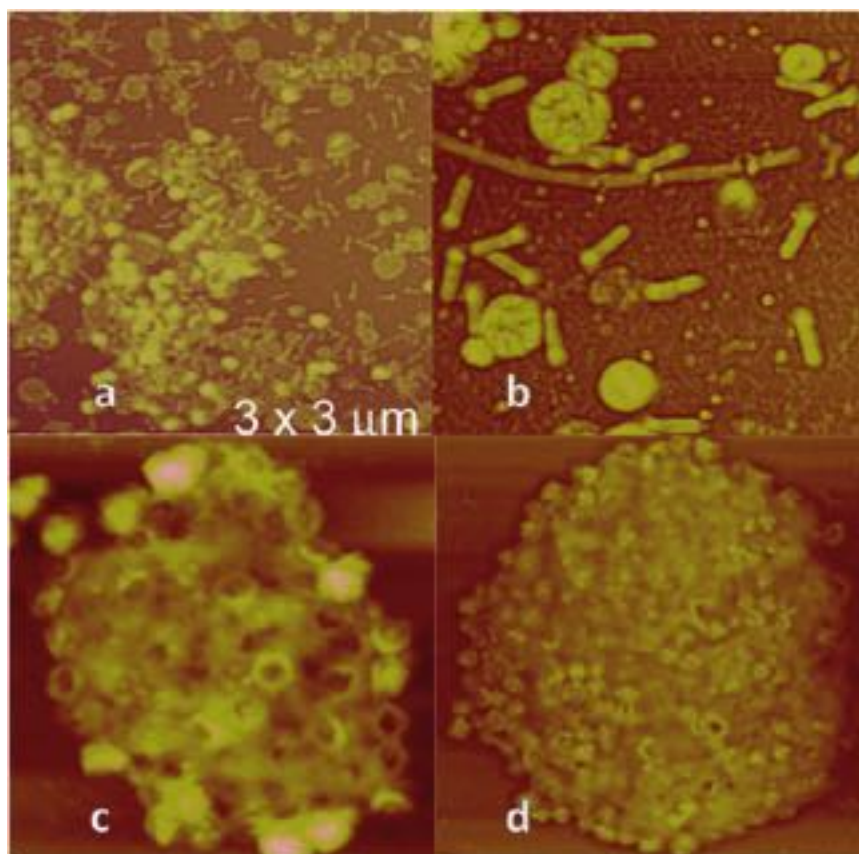


Fig. 15 *Viral assembly.* (a) and (b) are AFM images showing the simultaneous production of icosahedral heads and helical tails of future cyanophage within a single bacterial cell. (c) and (d) are AFM images of “virus factories” found in mimi virus infected amoeba cells. The factories each produce hundreds of virus particles.

distinctive structure as its outer shell dissolves into pieces. These circular forms later restructure as the “pinwheel” in **Fig. 13(e)**. In **Fig. 14(f)**, virions of the very small satellite tobacco mosaic virus (T-1, 17 nm diameter), treated lightly with phenol, are seen with striking synchrony extruding their single stranded RNA complements that contain but a single gene.

Viral Assembly

In rare instances, when host cells that are actively engaged in producing virus are disrupted by mechanical or chemical means, the actual production and/or assembly of particles can be observed. **Fig. 15** presents two examples. In **Fig. 15(a)** and **(b)** are infected cyanobacterial cells in the process of manufacturing bacteriophage. Apparent in the images is the separate construction in the cytoplasm of icosahedral heads and helical tails. It is not, however, apparent from the observations how the two substructures are subsequently brought together and joined, a process on the nanoscale level that must resemble the coupling of boxcars in composing a train.

Of great interest in the genesis of large, complex viruses is the role of “viral factories” that, like a bubbling fountain, gush out masses of newly formed virus particles. **Fig. 15(c)** and **(d)** are AFM images of “virus factories” producing, at their surfaces, remarkable amounts of mimi virus cores. The prolific factories are found in substantial numbers in the interiors of infected amoeba host cells that have been mechanically disrupted.

Further Reading

- Kuznetsov, Y.G., Daijogo, S., Zhou, J., Semler, B.L., McPherson, A., 2005. Atomic force microscopy analysis of icosahedral virus RNA. *Journal of Molecular Biology* 347, 41–52.
- Kuznetsov, Y.G., Datta, S., Kothari, N.H., *et al.*, 2002. Atomic force microscopy investigation of fibroblasts infected with wild type and mutant murine leukemic virus (MuLV). *Biophysical Journal* 83, 3665–3674.
- Kuznetsov, Y.G., Gershon, P.D., McPherson, A., 2008. Atomic force microscopy investigation of vaccinia virus structure. *Journal of Virology* 85, 7551–7566.
- Kuznetsov, Y.G., Low, A., Fan, H.Y., McPherson, A., 2005. Atomic force microscopy investigation of isolated virions of murine leukemia virus. *Journal of Virology* 79, 1970–1974.
- Kuznetsov, Y.G., Malkin, A.J., Lucas, R.W., Plomp, M., McPherson, A., 2001. Imaging of viruses by atomic force microscopy. *Journal of General Virology* 82, 2025–2034.
- Kuznetsov, Y.G., Martiny, J.B.H., McPherson, A., 2010. Structural analysis of a *Synechococcus myovirus* S-CAM4 and infected cells by atomic force microscopy. *Journal of General Virology* 91, 3095–3104.
- Kuznetsov, Y.G., McPherson, A., 2011. Atomic force microscopy in imaging of viruses and virus-infected cells. *Microbiology and Molecular Biology Reviews* 75 (2), 268–285.
- Kuznetsov, Y.G., Rossmann, M., McPherson, A., 2013. Morphogenesis of mimivirus and its viral factories: An atomic force microscopy study of infected cells. *Journal of Virology* 87 (20), 11200–11213.
- Kuznetsov, Y.G., Ulbrich, P., Haubova, S., Ruml, T., McPherson, A., 2007. Atomic force microscopy investigation of Mason–Pfizer monkey virus and human immunodeficiency virus type 1 reassembled particles. *Virology* 350 (2), 434–446.
- Kuznetsov, Y.G., Victoria, J.G., Robinson Jr., W.E., McPherson, A., 2003. Atomic force microscopy imaging of human immune deficiency. *Virus* 77 (22), 11896–11909.
- Kuznetsov, Y.G., Xiao, C., Sun, S., *et al.*, 2010. Atomic force microscopy investigation of the giant mimivirus. *Virology* 404, 127–137.
- Kuznetsov, Y.G., Zhang, M., Menees, T.M., McPherson, A., Sandmeyer, S., 2005. Investigation by atomic force microscopy of the structure of Ty3 retrotransposon particles. *Journal of Virology* 79, 8032–8045.
- Low, A., Datta, S., Kuznetsov, Y.G., *et al.*, 2007. Mutation in the glycosylated Gag protein of murine leukemia virus results in reduced in vivo infectivity and a novel defect in viral budding or release. *Journal of Virology* 81 (8), 3685–3692.
- McPherson, A., Kuznetsov, Y.G., 2011. Atomic force microscopy investigation of viruses. In: Braga, P.C., Ricci, D. (Eds.), *Atomic Force Microscopy in Biomedical Research, Methods and Protocols: Methods in Molecular Biology* 736. Humana Press, pp. 171–195.
- Plomp, M., Rice, M.K., Wagner, E.K., McPherson, A., Malkin, A.J., 2002. Rapid visualization at high resolution of pathogens by atomic force microscopy - structural studies of herpes simplex virus – 1. *American Journal of Pathology* V160 (6), 1959–1966.

Cryo-Electron Microscopy (CEM) Structures of Viruses

David Chmielewski and Wah Chiu, Stanford University, Stanford, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

Historical Background

The first direct visualization of virus particles, an electron micrograph of tobacco mosaic virus, came half a century after the existence of organisms smaller than bacteria was established (Kausche *et al.*, 1939). The advent of electron microscopy was shortly followed by a period of discovery whereby viruses related to many diseases were observed for the first time: variola, vaccinia and varicella, and poliovirus (Nagler and Rake, 1948; Van Rooyen and Scott, 1948; Reagan and Brueckner, 1952). However, during this period resolution was limited to gross particle shapes, often distorted due to chemical fixation and negative staining of the virus prior to imaging in the electron microscope. Embedding biological samples in ice was a critical advance, preserving high resolution information that is lost in the air-dried state (Adrian *et al.*, 1984; Taylor and Glaeser, 1974). While vitrification traps biological molecules in a thin film of amorphous ice, electron micrographs of these particles are inherently limited to low signal due to the sensitivity of biological molecules to radiation damage imparted by the electron beam. Because of the large depth of field in the electron microscope, each micrograph is a projection through the specimen, yielding 2D information (De Rosier and Klug, 1968). While direct observation of virus particles in 2D micrographs is useful for many applications, such as rapid diagnosis of clinical samples from virus outbreaks, the true power of cryo-EM relies on statistical averaging of many particle copies and interpretation of high-resolution 3D information (Kapikian *et al.*, 1972; Johnson *et al.*, 1977; Chiu *et al.*, 1997).

A 3D density map of the specimen can be generated from a set of particle views in different orientations. Before averaging the contrast transfer function (CTF) information, each projection image needs to be deconvoluted for its defocus setting. The averaging of multiple, equivalent copies of the molecule in a specific view serves to increase the signal-to-noise ratio of that orientation (Crowther *et al.*, 1970; Frank, 1975). Given sufficient views of a particle taken with enough defocus settings, a 3D volume can be reconstructed from 2D images without missing information. The combination of vitrification (to preserve high-resolution features of the specimen), large particle numbers (to overcome low-SNR in each image), and methods to reconstruct a 3D density map from multiple particle orientations, allowed for the first sub-nanometer reconstruction of tobacco mosaic virus by cryo-EM (Jeng *et al.*, 1989). The first 3D reconstructions of single particles to reach sufficient resolution to model the capsid polypeptide backbone utilized high-symmetry specimens, namely the icosahedral viruses epsilon-15, cytoplasmic polyhedrosis virus and rotavirus (Yu *et al.*, 2008; Jiang *et al.*, 2008; Zhang *et al.*, 2008). Each icosahedral virus particle has significant protein and nucleic acid mass (typically tens of megadaltons), with protein capsids arranged in 60 identical asymmetric units. A combination of electron scattering mass and structural symmetry allows for accurate determination of particle orientations and significant averaging power. For many of the same reasons that viruses proved ideal specimen in reaching near-atomic resolution for single-particle cryo-EM, sub-volume averaging of viruses is leading the progression of *in-situ* cryo-electron tomography (cryo-ET) to increasingly higher resolutions today (Schur *et al.*, 2016).

In this article we first describe cryo-EM methods for solving and analyzing virus structure at near-atomic resolution. Next, we describe information within cryo-EM maps of different viruses and discuss the implications for such processes as virus assembly, genome packaging, infection, and host-interactions. Finally, we briefly describe exciting recent advancements in directly visualizing stages of the virus life cycle *in-situ* using cryo-ET.

Cryo-EM at Near-Atomic Resolution

In recent years, many virus morphologies have proven amenable to high-resolution structure analysis by cryo-EM (Fig. 1). The vast majority of these PDB-deposited structures utilize global symmetry inherent in the particles during the image processing steps. Sorting of conformational heterogeneity both between particle populations and within individual particles has proven a valuable method to increase resolution of virions lacking exact icosahedral symmetry. The rapid increase of near-atomic resolution structures in the field can be attributed to several advances in both experimental and computational techniques.

Data Collection and Image Reconstruction

The prospects of achieving near-atomic resolution structure of viruses by cryo-EM now largely relies on a few notable factors: particle count, particle homogeneity, structural flexibility, particle size and ice thickness. Much of the recent proliferation in near-atomic resolution structures of virus capsids can be attributed to the improved quality of images recorded on a new generation of direct electron detectors with increased detective quantum efficiency (DQE) and high frame-rate acquisition of “movie-mode” images (McMullan *et al.*, 2009). The improvement in detector DQE corresponds to a higher SNR contained in each image, boosting contrast and increasing the accuracy of particle alignment during the orientation refinement steps (described below). Additionally, images collected as a “movie” of sub-frames allow for alignment of the same specimen area in each individual frame

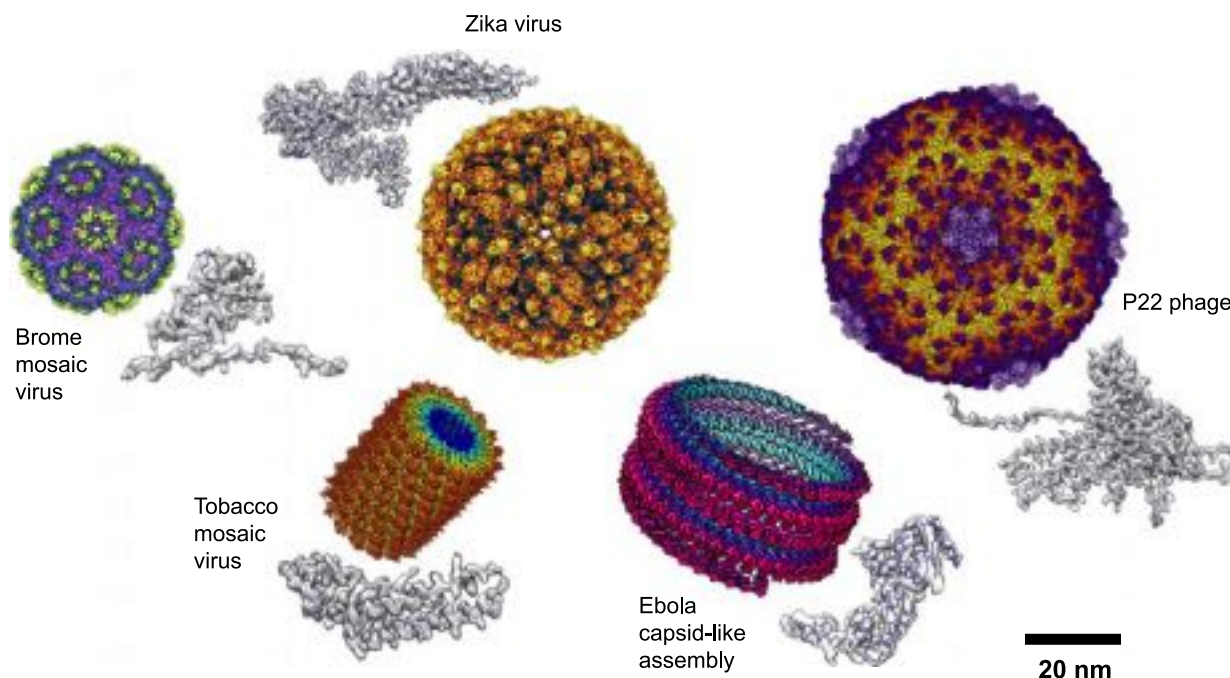


Fig. 1 Structures of diverse virus capsids have been solved by cryo-electron microscopy to near-atomic resolutions. Representative examples include plant viruses brome mosaic virus (EMD-6000, PDB-3J7L) and tobacco mosaic virus (EMD-2833, PDB-4UDV), bacteriophage P22 (EMD-8606, PDB-5UU5), and significant human-disease causing agents Zika virus (EMD-8116, PDB-5IRE) and the capsid-like assembly of Ebola virus (EMD-7343, PDB-6C54) (Wang *et al.*, 2014; Fromm *et al.*, 2015; Hryc *et al.*, 2017; Sirohi *et al.*, 2016; Su *et al.*, 2018). At sub-4 Å resolution, an atomic model of the capsid polypeptide (dark blue) can be reliably fit into the density (gray). Virus capsids to scale.

to correct for beam-induced movements of the stage and/or specimen, further improving image SNR and resolution (McMullan *et al.*, 2009; Wang *et al.*, 2014; Brilot *et al.*, 2012).

Single-particle image processing pipelines used to compute cryo-EM density maps involve assessing micrograph quality, selection of particles, image defocus estimation and correction, particle orientation estimation and refinement, image reconstruction and map validation. A large number of different software packages, often utilizing different computational techniques, have been used to solve 300+ virus structures between 1.9 and 4.0 Å resolution (Tan *et al.*, 2018). Recently, additional methods such as 3D-classification of particle conformations, refinement of sub-particle orientations and correction of Ewald-sphere effects for large viruses have been used to overcome limitations in the conventional data processing pipelines (Walls *et al.*, 2016; Wolf *et al.*, 2006). Moreover, the computational capabilities of facilities with large computing clusters has allowed for a remarkable increase in both throughput of cryoEM structures and the ability to handle large datasets that now routinely reach millions of individual particle images for a single experiment.

In addition to the microscope hardware and computational improvements, success of single-particle cryo-EM experiments in reaching near-atomic resolutions is also dependent on the specimen preparation. Since the invention of vitrification of biological samples, as first demonstrated with virus particles, the process of plunge-freezing thin films of specimens into liquid ethane has remained incredibly consistent (Vogel *et al.*, 1986). Today, a variety of commercial plunge-freezer apparatuses are available for this task, allowing for high throughput of cryoEM grids prepared with different parameters. To complete a successful sample preparation, the investigator is responsible for optimizing the sample purity and distribution of particles within image acquisition areas, as well as optimizing thickness of the embedding ice. Recently, there have been major innovations for measuring and optimizing ice thickness, both before vitrification and during cryo-EM imaging (Rice *et al.*, 2018; Dandey *et al.*, 2018). In addition, the development of techniques for rapid application of sample has shown promise for improving particle distribution, capturing transient reaction intermediates, reducing denaturing effects of the air-water interface, and limiting the amount of specimens needed for cryo-EM experiments (Dandey *et al.*, 2020).

Resolution in cryo-EM maps is determined by computing the Fourier-shell coefficient between two independent half-maps, at a specific threshold (0.143), termed the “gold-standard” procedure (Henderson *et al.*, 2012; Scheres and Chen, 2012). This is practically achieved by randomly splitting the data into two random sets at the beginning of the data processing procedure and computing two independent reconstructions. The resolution is then a measure of the reproducibility of features in the two maps. As resolution improves, the amount of biological information that can be extracted from the map increases. At sub-nanometer resolution, (6–9 Å), protein subunit boundaries can be defined, while β -strands appear as continuous slabs and helices as flat “sausages”. At 4.5 Å resolution and better, β -strands are separated, the pitch of α -helices can be seen, and some bulky side chains become apparent. An atomic model, representing the full atomic coordinates of the protein, can be fit into the density map by

taking stereochemical constraints and minimization of atomic clashes into consideration. Typically, at 3.5 Å resolution, the C α polypeptide backbone can be traced and most side chains accurately positioned with proper stereochemistry (Hryc *et al.*, 2017). Interpretation of the atomic model can be extended beyond peptide and side chain placements at resolutions better than 2.8 Å, where densities such as bound waters, lipids, and sugars can be accurately modeled with proper stereochemistry relative to nearby side-chains, without the need for orthogonal validation experiments (Roh *et al.*, 2020).

Interpretation of Near-Atomic Resolution Maps

Side Chain Interactions and Virus Assembly

Many viruses utilize conformational changes between morphogenic states as part of their assembly and maturation pathway. This is exemplified by DNA phages that transition from a genome-free “procapsid” state to a final, genome-packaged, metastable virion through a series of conformational rearrangements in the major capsid protein (Guo *et al.*, 2014; Jiang *et al.*, 2003). Viruses of the enveloped ssRNA family *Flaviviridae* begin as rough, spiky particles composed of E glycoprotein surface trimers and finish as a smooth coat of E proteins arranged as dimers in herringbone pattern (Hryc *et al.*, 2017; Zhang *et al.*, 2007; Zhang *et al.*, 2013). When possible, the purification of viruses in distinct assembly states allows for comparison of global morphologies and observation of subtle changes in polypeptide conformations and side-chain interactions essential to the structural transitions. Classification of particle conformations in 3D during data processing offers the potential to separate distinct populations and solve structures of multiple virus assembly states in a single experiment without the need for absolute sample homogeneity.

A striking example of the transition between virus conformations is explored in the near-atomic resolution cryo-EM maps of phage P22 procapsid and mature virion. Comparison of sub-nanometer structures first revealed the procapsid is more rounded in global conformation relative to the expanded, more angular virion capsid (Scheres, 2016; Chen *et al.*, 2011). At 3.8 Å, the C α backbone models of gp5 coat protein and gp8 scaffolding protein reveal electrostatic interactions between the negatively-charged N-arm region of gp5 and positive charged C-terminus of gp8. In the mature virion model of gp5, derived from the 3.3 Å resolution map, the N-arm region of every subunit forms a β -sheet with a protomer in a neighboring asymmetric unit, and the neighboring subunits in the same asymmetric unit form salt bridges (residues Glu5 and Glu15) with residues Lys31 and Arg42, respectively (Hryc *et al.*, 2017). The rearrangement of the flexible gp5 N-terminal arm from interacting with scaffold gp8 in the procapsid state to then stabilizing inter-subunit interfaces in the metastable virion, is proposed to serve as a driving signal for global capsid maturation.

Beyond Global Symmetry

The widespread utilization of icosahedral capsids across diverse viral lineages is a striking display of structure convergence to maximize evolutionary fitness. It is only when taking a more detailed view of these particles that the structural asymmetries essential to individual viral life cycles are revealed. Until now, discussion has been limited to averaging of icosahedral viral capsids, where 60-fold symmetry is applied during reconstruction to achieve near-atomic resolution structure. In icosahedral viruses, structural gene products range from arrangement in perfect symmetry to quasi-equivalence and full asymmetry. In addition, structural proteins are often incorporated into the mature viral particles or structural intermediates at sub-stoichiometric amounts relative to the major capsid proteins. To resolve these features, the highly accurate orientation of a single asymmetric unit in the icosahedral capsid typically serves as a starting point for searching over all possible asymmetric orientations and choosing the correct one based on alignment to a symmetry-breaking feature. The use of so-called “expanded symmetry”, in combination with 3D classification on focused regions (available through a popular image reconstruction software called Relion), has proven extremely valuable in resolving complexes small in size and/or with different symmetries relative to the protein capsid (Scheres, 2016). The relationship between symmetry and resolvability in capsid structures is exemplified by viral gene products such as portal and tail complexes in DNA viruses and viral genomes in RNA viruses, while local symmetry of membrane glycoproteins in pleomorphic enveloped viruses aids their study at near-atomic resolutions.

Symmetry Mismatch in DNA Portals

Double-stranded DNA (dsDNA) bacteriophages package and often eject pressurized DNA through a portal complex located at one unique vertex on the virus capsid. The portal connector hub forms an attachment site for phage tail machinery and closure proteins. Phage portals and tail complexes show great variability in morphology and complexity: some are mostly-contained inside of the capsid shell while others are long extensions that greatly exceed the length of the capsid head. Because the portal is located at a single five-fold vertex on the capsid, traditional icosahedral-reconstruction will fail to resolve this non-icosahedral component. Instead, the orientation of the tail relative to the dominant icosahedral capsid must be found. A typical asymmetric reconstruction of the virus capsid, required to resolve the position of the unique portal, will contain 1/60th the signal at each asymmetric unit compared to an icosahedral-symmetrized reconstruction. It is the lack of cumulative signal that explains why asymmetric reconstructions routinely fail to reach the resolutions achieved in icosahedral reconstructions of epsilon15 and P22 phages (Chang *et al.*, 2006; Jiang *et al.*, 2006).

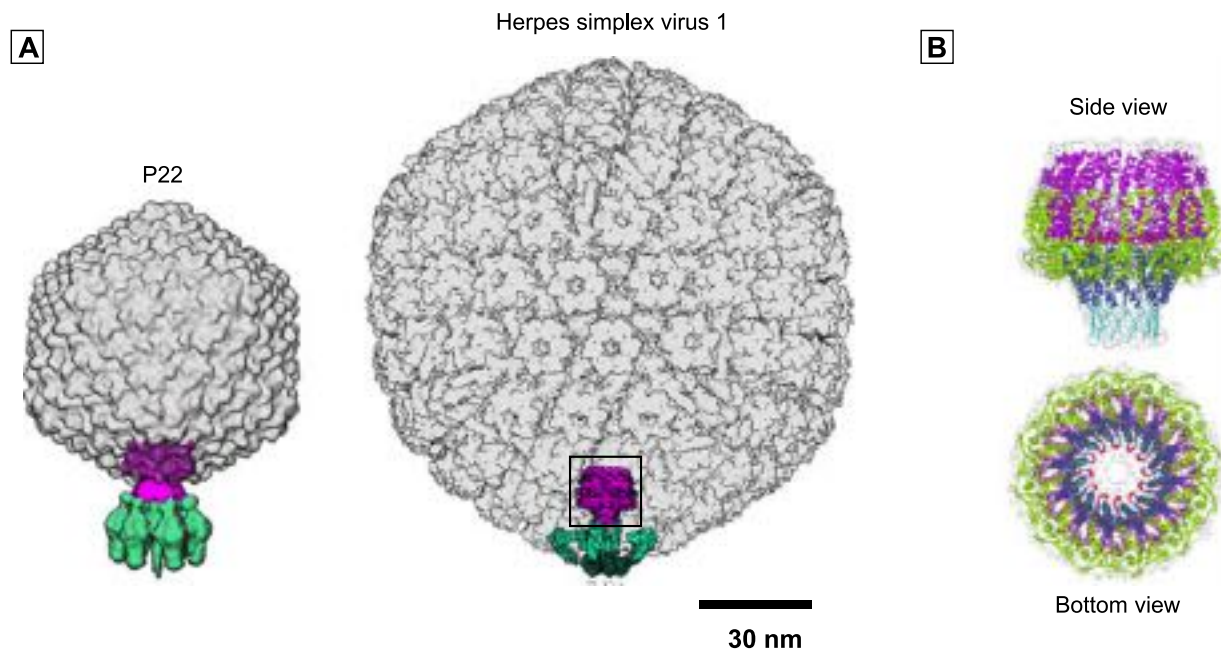


Fig. 2 Protein machines responsible for DNA packaging and exit are located at one unique vertex on the virus capsid in (A) bacteriophage P22 (EMD-1222) and Herpes simplex virus 1 (HSV-1) (EMD-9864). The dodecameric portal assemblies, breaking the icosahedral symmetry of the capsid, are revealed by asymmetric reconstructions (Chang *et al.*, 2006; McElwee *et al.*, 2018) (Liu *et al.*, 2019) (Chang *et al.*, 2006; McElwee *et al.*, 2018). With further focused-classification techniques, the structure of all components of the HSV-1 portal complex were determined at sufficient resolution to build a (B) model of the polypeptide-backbone of each protein and resolve interactions with the dsDNA (PDB-60D7) (Liu *et al.*, 2019).

To overcome the difficulty of aligning virion particles without imposed icosahedral symmetry, the predominant strategy is to use the coordinates of an icosahedral refinement, rotate the particle to each of the 60 equivalent icosahedral positions, and find the one position that best matches an asymmetric model without refining the orientation. In some cases, where the tail provides significant signal, the orientation of the tail complex can be directly found without using the capsid orientation, or by searching over a constrained orientation space using the above procedure as a starting point. The first visualizations of portal complexes in asymmetric reconstructions revealed conclusively that the portal is organized as a dodecamer, typically with a fixed orientation relative to the capsid (Liu *et al.*, 2010; Tang *et al.*, 2008; Jiang *et al.*, 2006). Symmetry mismatches often exist between the portal protein and external tails and/or internal connector stacks (Chang *et al.*, 2006; Guo *et al.*, 2013). In P22 phage, the asymmetric reconstruction of the unique portal revealed a sixfold hexamer tail attached to a dodecamer connector complex, while a long C-terminal tail inside the capsid appears to serve a similar function as an internal connector stack (Chang *et al.*, 2006).

The insights into DNA packaging and genome egress gained from model bacteriophages can be extended to eukaryotic systems wherever shared structural mechanisms exist. One such example is the conservation of the characteristic HK-97 protein fold and capsid assembly pathway between prokaryote-infecting tailed bacteriophages (*Caudoviridae*) and eukaryote-infecting herpesviruses (*Herpesviridae*) (Fig. 2). Where no sequence similarity could predict common ancestry, cryoEM maps of phage P22 and HSV-1 revealed common capsid protein fold and virion architecture (Baker *et al.*, 2005). Interestingly, a protein portal was subsequently discovered in HSV-1 B capsids, answering the outstanding question of why the Herpes capsid is located asymmetrically in close proximity to one side of each virus envelope (Rochat *et al.*, 2011). Similarly, following the elucidation of an internal portal complex in model phage PRD1, re-evaluation of the structurally-related human Adenovirus resulted in identification of a unique icosahedral fivefold vertex by immunogold labeling of protein IVa2 (Ahi *et al.*, 2017; Strömsten *et al.*, 2003).

The study of portal complexes in the context of the native particle by cryo-EM has revealed common folds, common ancestry among lineages of DNA viruses infecting prokaryotes and eukaryotes, and common strategies for incorporating symmetry mismatches into functional protein machines. The study of portals in intermediate states has also led to a model whereby these complexes act as essential nucleators of procapsid assembly. Current data processing techniques offer the potential to better understand portal assemblies and portal-genome interactions at near-atomic resolutions (Liu *et al.*, 2019; McElwee *et al.*, 2018).

Envelope Glycoproteins

Enveloped viruses utilize membrane-anchored proteins that mediate cell entry *via* binding to receptor molecules on the host cell surface and facilitating fusion between the viral and host membrane bilayers. These membrane protein assemblies are often visible in electron micrographs of enveloped virus particles, existing as distinctive “spikes” protruding outward from the virus surface. For

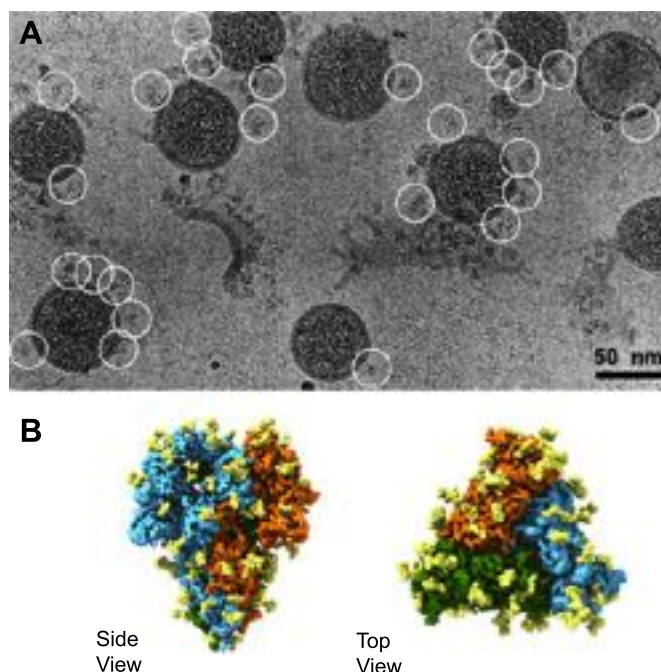


Fig. 3 Cryo-EM structure of a human coronavirus S trimer computationally extracted from intact particles. (A) Representative cryo-EM micrograph of purified human coronavirus NL63 (HuCoV-NL63) particles with marked positions of spike (S) trimers (white circles). (B) Reconstructed cryo-EM map of HuCoV-NL63 S trimer in closed, pre-fusion conformation as side and top views with glycans depicted (yellow).

many viruses, these envelope proteins are heavily glycosylated, with N- or O-linked glycan species serving critical functions in protein folding and assembly, cellular tropism and immune evasion. Due to the pleomorphic nature of the lipid envelope, enveloped viruses generally lack global symmetry useful for determining highly accurate orientations of virus capsids as previously discussed. Instead, structural studies of envelope glycoproteins reaching near-atomic resolution typically require *in vitro* expression and purification of the soluble protein sequence exterior to the viral membrane (Lee *et al.*, 2016; Walls *et al.*, 2016). In many cases, stabilizing mutations are required to produce suitable glycoprotein complexes for these structural studies (Kirchdoerfer *et al.*, 2018; de Taeye *et al.*, 2015). Cryo-ET of membrane viruses and averaging of envelope proteins computationally extracted from tomograms of virus particles allows for study of these complexes *in-situ*, both representing the native chemical state of the glycoproteins and providing spatial context relative to other components of the virion. While the resolvability of these glycoprotein subtomogram averages are typically limited to 5–10 Å due to small particle size, conformational variability and loss of high-resolution information during tilt series acquisition, recent developments in subtomogram averaging show promise in pushing on to higher resolutions.

The study of envelope glycoproteins in the native state by single particle cryo-EM is exemplified by structures of human coronavirus spike (S) trimers that are computationally extracted from images of intact virions (Ke *et al.*, 2020; Zhang *et al.*, 2020). Previously, *in vitro* structural studies of purified coronavirus S trimers, the major focus of current vaccine design, relied on protein expression of the S1/S2 “crown” region stabilized by introduced mutations in the S2 fusion machinery (Kirchdoerfer *et al.*, 2018; Pallesen *et al.*, 2017; Wrapp *et al.*, 2020). However, in the study of human alphacoronavirus NL63 (HuCoV-NL63), full-length S trimers, including the C-terminal stalk connecting the crown to viral envelope, are clearly visible in micrographs of vitrified virus particles (Fig. 3) (Zhang *et al.*, 2020). In this case, the sample preparation was performed without utilizing any of the chemical fixatives required for handling and freezing of more pathogenic coronaviruses. Surprisingly, though the majority of extracted particles are “side” views projecting outward from the viral envelope, the c3-symmetric average of the S1/S2 “crown” region in the closed, pre-fusion conformation reached 3.4 Å resolution. This exciting result displays the potential of single particle cryo-EM methods to resolve near-atomic resolution structures of glycoprotein complexes in the context of the virus particle, without the need for *in vitro* purification or mutations. In the near future, this method will likely be applied to additional viral envelope proteins, producing *in-situ* structures at resolutions suitable for drug optimization and vaccine design studies.

Genome Structures

Due to the dominant, symmetric signal provided by viral protein capsids, many insights into viral assembly from cryo-EM structures have naturally involved capsid protein-protein interactions and changes in polypeptide conformations during capsid maturation. An underappreciated aspect of viral assembly pathways is the encapsidated nucleic acid genome and its interactions with the protein capsid. In dsDNA phages of family *Podoviridae*, including phage P22, interactions between DNA and an

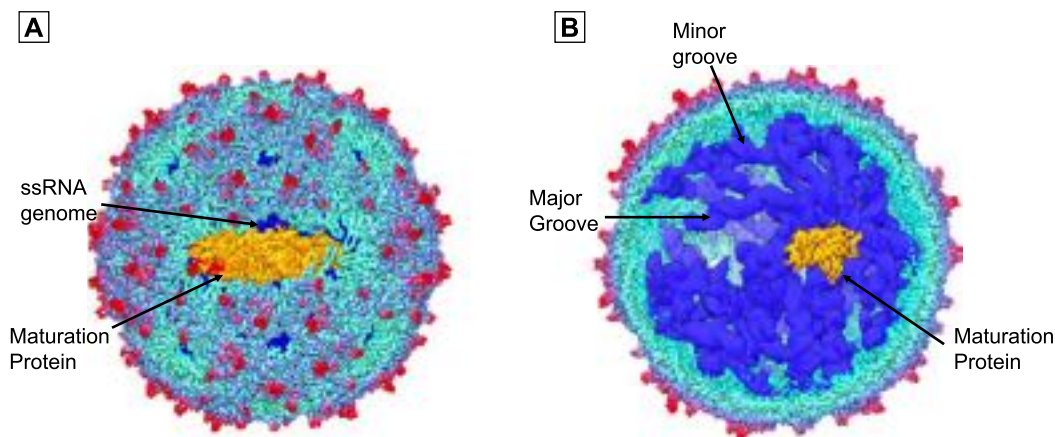


Fig. 4 Cryo-EM and asymmetric reconstruction allows for the visualization of viral genomes *in situ*. The (A) cryo-EM density map of phage MS2 (EMD-8397) reveals a single copy of maturation protein on the viral capsid, providing a strong asymmetric density (Dai *et al.*, 2017). The (B) cut-open view of the virion reveals the entirety of the viral genome (blue), where the 3' end interacts with the maturation protein in a relatively fixed orientation. The genome density map revealed the short- and long-range genomic RNA interactions as well as contacts with the interior surface of the capsid protein.

asymmetric portal complex state are proposed to induce a conformational change in the portal proteins and signal an end to genome packaging (Lokareddy *et al.*, 2017). Instead of injecting genome into preformed protein shells, RNA viruses typically co-assemble with the genome. For viruses of family *Orthoreoviridae*, interactions between 10 ssRNA gene segments within the host cytosol are believed to facilitate encapsulation of the viral genome by an inner protein shell at the earliest point of assembly (Borodavka *et al.*, 2017).

The structural organization of the viral genome can be best resolved in cases where it exists in a fixed conformation relative to an asymmetric feature on the virus capsid (Fig. 4). This was exhibited in the stunning maps of ssRNA phages MS2 and Q β , where the majority of the full RNA genome could be modeled into sub-nanometer resolution density maps based on RNA base-pairing constraints (Dai *et al.*, 2017; Gorzelnik *et al.*, 2016). In both cases, the conformationally homogenous ssRNA genome exists in a fixed orientation relative to a significant asymmetric protein density on the capsid surface termed “maturation protein.” In addition to revealing the *in situ* organization of the viral genome, these maps allowed the identification of interactions between genomic secondary structure and recognition elements on the interior surface of the protein capsid. In less ideal cases where the genome is not organized relative to a strong asymmetric feature on the virus capsid, methods to subtract the symmetric capsid signal from each raw particle image and determine the asymmetric genome orientation from the remaining information have proven useful (Liu and Cheng 2015; Zhang *et al.*, 2015).

In-Situ Structural Virology

Cellular Electron Cryotomography

As discussed so far, the basis for single-particle averaging is the existence of a large number of equivalent molecules in distinct orientations. Many situations exist when the specimen of interest does not exhibit conformational homogeneity, particle numbers, or stability required for single-particle analysis: some of these cases include pleomorphic enveloped virions, virus assembly and disassembly intermediates, and virus-induced subcellular structures, among others. When presented with these challenges, tomography is the only way to generate 3D structures of individual molecules. In cryo-electron tomography (cryo-ET), a series of projection images are collected by tilting the specimen and combining the resultant “tilt-series” into a 3D volume (Crowther *et al.*, 1970; De Rosier and Klug, 1968). In contrast to individual projection images, these reconstruction volumes provide valuable information regarding the 3D orientation and spatial organization of features. This 3D information is especially useful in applications such as analyzing glycoprotein assemblies in pleomorphic virus envelopes, asymmetric virus-receptor interactions, and virus intermediates in the context of the host cell.

The structural information within a reconstructed tomogram is limited by three notable factors: low signal in each tilt image to limit accumulated electron dose often compounded by thick specimen, missing information due to mechanical limitations in the microscope stage (termed the “missing wedge”), and accuracy of alignments between each successive tilt image prior to reconstructing a 3D volume (Lučić *et al.*, 2013). Recovery of high-resolution information from individual features, the “N of 1” situation, has not been achieved without the use of averaging (Kaelber *et al.*, 2017). Because the entire tomographic field of view is reconstructed in 3D, homogenous copies of an object or consistent substructures within heterogenous objects, can be extracted and aligned to generate a 3D density map.

Study of native viral intermediates is often limited by the fast dynamics of assembly and sensitivity of these conformations to purification. For this reason, stable genetic mutants that resemble the corresponding transient intermediate state are often studied *in-vitro* to gain insight into the cellular processes. An elegant solution to this problem was provided in the study of native Syn5 dsDNA phage assembly *in situ* (Dai *et al.*, 2013). Whole-cell imaging of cyanobacteria containing Syn5 intermediates using Zernike phase plate for increased low-resolution contrast allowed for visualization of native phage progeny within the cytosol (Dai *et al.*, 2014). Analysis of these progeny *via* extraction and 3D classification resulted in the identification of five distinct Syn5 conformations. This provided conclusive evidence of an assembly order in which Syn5 first acquires its DNA genome through a portal at one icosahedral vertex, a tail is added to the portal vertex, and finally a protein complex termed the “horn” is attached to the vertex opposite the tail to complete the virion assembly process. In this way, cryo-ET studies of virus infection and assembly in whole bacteria cells offer incredible amounts of spatial and temporal information regarding the virus replication cycle and viral infection-induced effects on the host in the natural environment.

Analysis of human cells by cryo-ET is limited to the areas where the electron beam can penetrate the sample. For this reason, cell lines that spread on the cryo-EM grid surface are preferred, allowing for visualization of features deeper into the cell body. With many enveloped viruses, rich information conveniently exists at the plasma membrane, where particle budding and assembly occurs for a number of ssRNA virus families including *Retroviridae*, *Filoviridae*, and *Alphaviridae*. Cryo-ET imaging of human u2os cells infected with *Chikungunya* (CHIKV) virus of family *Alphaviridae*, and subsequently treated with a potent human anti-CHIKV antibody, provided a stunning visual understanding of the mechanism of antibody neutralization at the cell surface (Fig. 4) (Jin *et al.*, 2018). Antibody C9 can be seen bound to glycoproteins exposed at the cell surface, while pleomorphic nucleocapsids containing the ssRNA genome are trapped beneath the membrane in the cytosol. Further, the close proximity of RNA replication spherules to the sites of inhibited virus budding provide valuable information related to the CHIKV assembly pathway starting with positive-strand RNA synthesis and progressing to genome packaging, virus assembly and budding at the plasma membrane (Fig. 5).

Due to the high copy number of progeny virus in infected cells, virus assembly events in the cellular context are most easily located for native imaging by cryo-ET. However, it is also possible to study virus infection of host cells, though capturing this process often requires significant searching of the cryo-EM grid and a sharp eye! In the case of HSV-1 infection at the plasma membrane, distinct 3D entry intermediates revealed the morphological changes in virus and plasma membrane at stages of virus attachment, glycoprotein-driven membrane fusion and entry (Maurer *et al.*, 2008). With a sufficient number of captured events, it is possible to gain insights into the reaction dynamics of transient processes. Correlating the occupancy of distinct 3D states with the underlying process dynamics remains a rich area of development to bring *in situ* cryo-ET together with molecular simulations. Identifying points of dynamic pauses or rate-limiting steps in virus assembly and infection by *in situ* cryo-ET offers the potential to target these intermediate structures *via* antiviral strategies (Sutton *et al.*, 2020).

While cryo-ET analysis of human cells is mostly limited to the cell surface, advancements in focused ion beam (FIB) milling of thin cell lamella under cryo-conditions offers the potential to look deep within the cell at areas of interest at medium resolution

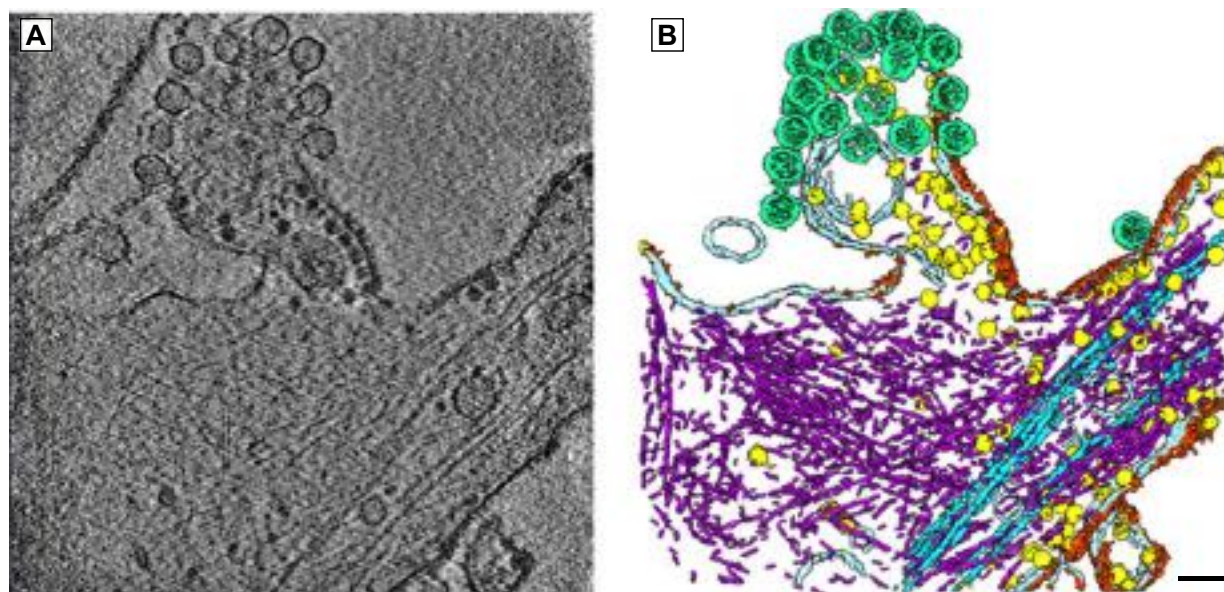


Fig. 5 Neutralizing antibodies (Nabs) bind to Chikungunya virus (CHIKV) glycoproteins exposed on the cell surface to block budding and egress of nascent virions. The blocking of viral egress is an additional Nab function beyond classical entry inhibition. Electron cryotomography imaging of CHIKV-infected human cells *in situ* reveal (A) budding-arrested CHIKV nucleocapsids docked to the plasma membrane in the Nab-treated condition (Jin *et al.*, 2018). (B) Automated annotation of cellular features using a convolutional neural network reveals the arrested nucleocapsids (yellow), membrane (light blue), actin filaments (purple), microtubules (cyan), CHIKV RNA replication spherules (green) and Nab C9 (orange) (Chen *et al.*, 2017). Scale bar 100 nm.

(Wu *et al.*, 2020). Locating the position of fluorescent-tagged virus protein and correlating to the position on the cryo-EM grid within the microscope prior to imaging offers the potential to study events such as enveloped virus assembly and genome replication on internal membranes, viral fusion in endosomes, and disassembly and transport processes in the cytosol and nucleus. Recent developments in cryo-ET, including the use of phase plates for improved image contrast near-to-focus, FIB milling areas of interest, and a renewed interest in high-voltage electron microscopes and detectors, offer the potential to push resolution of specimen in a cellular context to near-atomic resolutions.

Acknowledgment

This research has been supported by NIH grant (P41GM103832). We thank Dr. Jing Jin, Dr. Stella Sun and Dr. Greg Pintilie for providing original content used in figure preparation.

References

- Adrian, M., Dubochet, J., Lepault, J., McDowell, A.W., 1984. Cryo-electron microscopy of viruses. *Nature* 308, 32–36. doi:10.1038/308032a0.
- Ahi, Y.S., Hassan, A.O., Vemula, S.V., *et al.*, 2017. Adenoviral E4 34K protein interacts with virus packaging components and may serve as the putative portal. *Scientific Reports* 7 (1), 7582.
- Baker, M.L., Jiang, W., Rixon, F.J., Chiu, W., 2005. Common ancestry of herpesviruses and tailed DNA bacteriophages. *Journal of Virology* 79 (23), 14967–14970.
- Borodavka, A., Dykeman, E.C., Schrimpf, W., Lamb, D.C., 2017. Protein-mediated RNA folding governs sequence-specific interactions between rotavirus genome segments. *eLife* 6 (September), doi:10.7554/eLife.27453.
- Brilot, A.F., James, Z.C., Cheng, A., *et al.*, 2012. Beam-induced motion of vitrified specimen on holey carbon film. *Journal of Structural Biology* 177 (3), 630–637.
- Chang, J., Weigele, P., King, J., Chiu, W., Jiang, W., 2006. Cryo-EM asymmetric reconstruction of bacteriophage P22 reveals organization of its DNA packaging and infecting machinery. *Structure* 14 (6), 1073–1082.
- Chen, D.-H., Baker, M.L., Hryc, C.F., *et al.*, 2011. Structural basis for scaffolding-mediated assembly and maturation of a dsDNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 108 (4), 1355–1360.
- Chen, M., Dai, W., Sun, S.Y., *et al.*, 2017. Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nature Methods* 14 (10), 983–985.
- Chiu, W., Burnett, R.M., Garcea, R.L., 1997. *Structural Biology of Viruses*. United States: Oxford University Press.
- Crowther, R.A., Amos, L.A., Finch, J.T., de Rosier, D.J., Klug, A., 1970. Three dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature* 226, 421–425. doi:10.1038/226421a0.
- Crowther, R.A., DeRosier, D.J., Klug, A., 1970. The reconstruction of a three-dimensional structure from projections and its application to electron microscopy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 317, doi:10.1098/rspa.1970.0119.
- Dai, W., Fu, C., Khant, H.A., *et al.*, 2014. Zernike phase-contrast electron cryotomography applied to marine cyanobacteria infected with cyanophages. *Nature Protocols* 9 (11), 2630–2642.
- Dai, W., Fu, C., Raytcheva, D., *et al.*, 2013. Visualizing virus assembly intermediates inside marine cyanobacteria. *Nature* 502 (7473), 707–710.
- Dai, X., Li, Z., Lai, M., *et al.*, 2017. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* 541 (7635), 112–116.
- Dandey, V.P., Budell, W.C., Wei, H., *et al.*, 2020. Time-resolved cryo-EM using spotiton. *Nature Methods* 17 (9), 897–900.
- Dandey, V.P., Wei, H., Zhang, Z., *et al.*, 2018. Spotiton: New features and applications. *Journal of Structural Biology* 202 (2), 161–169.
- De Rosier, D.J., Klug, A., 1968. Reconstruction of three dimensional structures from electron micrographs. *Nature* 217, 130–134. doi:10.1038/217130a0.
- de Taeye, S.W., Ozorowski, G., de la Peña, A.T., *et al.*, 2015. Immunogenicity of stabilized HIV-1 envelope trimers with reduced exposure of non-neutralizing epitopes. *Cell* 163 (7), 1702–1715.
- Frank, J., 1975. Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy* 1 (2), 159–162.
- Fromm, S.A., Bharat, T.A.M., Jakobi, A.J., Hagen, W.J.H., Sachse, C., 2015. Seeing tobacco mosaic virus through direct electron detectors. *Journal of Structural Biology* 189 (2), 87–97.
- Gorzelnik, K.V., Cui, Z., Reed, C.A., *et al.*, 2016. Asymmetric cryo-EM structure of the canonical allovirus Q β reveals a single maturation protein and the genomic ssRNA in situ. *Proceedings of the National Academy of Sciences of the United States of America* 113 (41), 11519–11524.
- Guo, F., Liu, Z., Fang, P.-A., *et al.*, 2014. Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *Proceedings of the National Academy of Sciences of the United States of America* 111 (43), E4606–E4614.
- Guo, F., Liu, Z., Vago, F., *et al.*, 2013. Visualization of uncorrelated, tandem symmetry mismatches in the internal genome packaging apparatus of bacteriophage T7. *Proceedings of the National Academy of Sciences of the United States of America* 110 (17), 6811–T16.
- Henderson, R., Sali, A., Baker, M.L., *et al.*, 2012. Outcome of the first electron microscopy validation task force meeting. *Structure* 20 (2), 205–214.
- Hryc, C.F., Chen, D.-H., Afonine, P.V., *et al.*, 2017. Accurate model annotation of a near-atomic resolution cryo-EM map. *Proceedings of the National Academy of Sciences of the United States of America* 114 (12), 3103–3108.
- Jeng, T.-W., Jeng, T.-W., Crowther, R.A., Stubbs, G., Chiu, W., 1989. Visualization of alpha-helices in tobacco mosaic virus by cryo-electron microscopy. *Journal of Molecular Biology* 205, 251–257. doi:10.1016/0022-2836(89)90379-3.
- Jiang, W., Chang, J., Jakana, J., *et al.*, 2006. Structure of epsilon15 bacteriophage reveals genome organization and DNA packaging/injection apparatus. *Nature* 439 (7076), 612–616.
- Jiang, W., Li, Z., Zhang, Z., *et al.*, 2003. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nature Structural Biology* 10 (2), 131–135.
- Jiang, W., Matthews, L.B., Jakana, J., *et al.*, 2008. Backbone structure of the infectious ν 15 virus capsid revealed by electron cryomicroscopy. *Nature* 451, 1130–1134. doi:10.1038/nature06665.
- Jin, J., Galaz-Montoya, J.G., Sherman, M.B., *et al.*, 2018. Neutralizing antibodies inhibit chikungunya virus budding at the plasma membrane. *Cell Host & Microbe* 24 (3), 417–428.
- Johnson, K.M., Lange, J.V., Webb, P.A., Murphy, F.A., 1977. Isolation and partial characterisation of a new virus causing acute haemorrhagic fever in Zaire. *The Lancet* 309, 569–571. doi:10.1016/s0140-6736(77)92000-1.
- Kaelber, J.T., Corey, F.H., Chiu, W., 2017. Electron cryomicroscopy of viruses at near-atomic resolutions. *Annual Review of Virology* 4, 287–308. doi:10.1146/annurev-virology-101416-041921.

- Kapikian, A.Z., Wyatt, R.G., Dolin, R., *et al.*, 1972. Visualization by immune electron microscopy of a 27-Nm particle associated with acute infectious nonbacterial gastroenteritis. *Journal of Virology* 10 (5), 1075–1081.
- Kausche, G.A., Plankuch, E., Ruska, H., 1939. Die Sichtbarmachung von Pflanzlichem Virus Im bermikroskop. *Die Naturwissenschaften* 27, 292–299. doi:10.1007/bf01493353.
- Ke, Z., Oton, J., Ou, K., *et al.*, 2020. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature*. doi:10.1038/s41586-020-2665-2.
- Kirchdoerfer, R.N., Wang, N., Pallesen, J., *et al.*, 2018. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Scientific Reports* 8 (1), 15701.
- Lee, J.H., Ozorowski, G., Ward, A.B., 2016. Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science* 351 (6277), 1043–1048.
- Liu, H., Cheng, L., 2015. Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. *Science* 349, 1347–1350. doi:10.1126/science.aaa4938.
- Liu, Y.-T., Jih, J., Dai, X., Bi, G.-Q., Zhou, Z.H., 2019. Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome. *Nature* 570 (7760), 257–261.
- Liu, X., Zhang, Q., Murata, K., *et al.*, 2010. Structural changes in a marine podovirus associated with release of its genome into *prochlorococcus*. *Nature Structural & Molecular Biology* 17, 830–836. doi:10.1038/nsmb.1823.
- Lokareddy, R.K., Sankhala, R.S., Roy, A., *et al.*, 2017. Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nature Communications* 8 (January), 14310.
- Lučić, V., Rigort, A., Baumeister, W., 2013. Cryo-electron tomography: The challenge of doing structural biology in situ. *The Journal of Cell Biology* 202 (3), 407–419.
- Maurer, U.E., Sodeik, B., Grünwald, K., 2008. Native 3D intermediates of membrane fusion in herpes simplex virus 1 entry. *Proceedings of the National Academy of Sciences of the United States of America* 105 (30), 10559–10564.
- McElwee, M., Vijaykrishnan, S., Rixon, F., Bhella, D., 2018. Structure of the herpes simplex virus portal-vertex. *PLOS Biology* 16, e2006191doi:10.1371/journal.pbio.2006191.
- McMullan, G., Chen, S., Henderson, R., Faruqi, A.R., 2009. Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy* 109 (9), 1126–1143.
- Nagler, F.P., Rake, G., 1948. The use of the electron microscope in diagnosis of variola, vaccinia, and varicella. *Journal of Bacteriology* 55 (1), 45–51.
- Pallesen, J., Wang, N., Corbett, K.S., *et al.*, 2017. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proceedings of the National Academy of Sciences of the United States of America* 114, E7348–E7357. doi:10.1073/pnas.1707304114.
- Reagan, R.L., Brueckner, A.L., 1952. Morphological observations by electron microscopy of the Lansing strain of poliomyelitis virus after propagation in the Swiss albino mouse. *Texas Reports on Biology and Medicine* 10 (2), 425–428.
- Rice, W.J., Cheng, A., Noble, A.J., *et al.*, 2018. Routine determination of ice thickness for cryo-EM grids. *Journal of Structural Biology* 204 (1), 38–44.
- Rochat, R.H., Liu, X., Murata, K., *et al.*, 2011. Seeing the portal in herpes simplex virus type 1 B capsids. *Journal of Virology* 85 (4), 1871–1874.
- Roh, S.H., Shekhar, M., Pintlilie, G., *et al.*, 2020. Cryo-EM and MD infer water-mediated proton transport and autoinhibition mechanisms of V_0 complex. *Science Advances* 6 (41), eabb9605. doi:10.1126/sciadv.abb9605.
- Scheres, S.H.W., 2016. Processing of structurally heterogeneous cryo-EM data in RELION. *Methods in Enzymology* 579 (May), 125–157.
- Scheres, S.H.W., Chen, S., 2012. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* 9, 853–854. doi:10.1038/nmeth.2115.
- Schur, F.K.M., Obr, M., Hagen, W.J.H., *et al.*, 2016. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* 353 (6298), 506–508.
- Sirohi, D., Chen, Z., Sun, L., *et al.*, 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* 352 (6284), 467–470.
- Strömsten, N.J., Dennis, H.B., Bamford, J.K.H., 2003. The unique vertex of bacterial virus PRD1 is connected to the viral internal membrane. *Journal of Virology* 77 (11), 6314–6321.
- Su, Z., Wu, C., Shi, L., *et al.*, 2018. Electron cryo-microscopy structure of Ebola virus nucleoprotein reveals a mechanism for nucleocapsid-like assembly. *Cell* 172 (5), 966–978.
- Sutton, G., Sun, D., Fu, X., *et al.*, 2020. Assembly intermediates of orthoreovirus captured in the cell. *Nature Communications* 11, 4445. doi:10.1038/s41467-020-18243-9.
- Tang, J., Olson, N., Jardine, P.J., *et al.*, 2008. DNA poised for release in bacteriophage $\phi 29$. *Structure*. doi:10.1016/j.str.2008.02.024.
- Tan, Y.Z., Aiyer, S., Mietzsch, M., *et al.*, 2018. Sub-2 Å Ewald curvature corrected structure of an AAV2 capsid variant. *Nature Communications* 9 (1), 3628.
- Taylor, K.A., Glaeser, R.M., 1974. Electron diffraction of frozen, hydrated protein crystals. *Science* 186, 1036–1037. doi:10.1126/science.186.4168.1036.
- Van Rooyen, C.E., Scott, G.D., 1948. Smallpox diagnosis with special reference to electron microscopy. *Canadian Journal of Public Health* 39 (12), 467–477.
- Vogel, R.H., Provencher, S.W., von Bonsdorff, C.-H., Adrian, M., Dubochet, J., 1986. Envelope structure of semliki forest virus reconstructed from cryo-electron micrographs. *Nature* 320, 533–535. doi:10.1038/320533a0.
- Walls, A.C., Tortorici, M.A., Bosch, B.-J., *et al.*, 2016. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* 531 (7592), 114–117.
- Wang, Z., Hryc, C.F., Bammes, B., *et al.*, 2014. An atomic model of brome mosaic virus using direct electron detection and real-space optimization. *Nature Communications* 5 (September), 4808.
- Wolf, M., De Rosier David, J., Grigorieff, N., 2006. Ewald sphere correction for single-particle electron microscopy. *Ultramicroscopy* 106 (4–5), 376–382.
- Wrapp, D., Wang, N., Corbett, K.S., *et al.*, 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi:10.1101/2020.02.11.944462.
- Wu, G.H., Mitchell, P.G., Galaz-Montoya, J.G., *et al.*, 2020. Multi-scale 3D Cryo-Correlative Microscopy for Vitrified Cells. *Structure* 3 28 (11), 1231–1237. e3. doi:10.1016/j.str.2020.07.017.
- Yu, X., Jin, L., Zhou, Z.H., 2008. 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature* 453, 415–419. doi:10.1038/nature06893.
- Zhang, X., Ding, K., Yu, X., *et al.*, 2015. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* 527 (7579), 531–534.
- Zhang, X., Ge, P., Yu, X., *et al.*, 2013. Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. *Nature Structural & Molecular Biology* 20 (1), 105–110.
- Zhang, Y., Kaufmann, B., Chipman, P.R., Kuhn, R.J., Rossmann, M.G., 2007. Structure of immature west nile virus. *Journal of Virology* 81 (11), 6141–6145.
- Zhang, K., Li, S., Pintlilie, G., *et al.*, 2020. A 3.4-Å cryo-EM structure of the human coronavirus spike trimer computationally derived from vitrified NL63 virus particles. *bioRxiv* (Preprint). doi:10.1101/2020.08.11.245696.
- Zhang, X., Settembre, E., Xu, C., *et al.*, 2008. Near-atomic resolution using electron cryomicroscopy and single-particle reconstruction. *Proceedings of the National Academy of Sciences of the United States of America* 105, 1867–1872. doi:10.1073/pnas.0711623105.

Further Reading

- Chiu, W., Johnson, J.E., 2003. *Virus Structure (Advances in Protein Chemistry)*, vol. 64. Academic Press. pp. xi–xii.
- Glaeser, R.M., 2019. How good can single-particle cryo-em become? What remains before it approaches its physical limits? *Annual Review of Biophysics* 48 (May), 45–61.
- Joachim, F., 1996. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Elsevier.
- Prasad, B.V.V., Michael, F.S., 2012. Principles of virus structural organization. *Advances in Experimental Medicine and Biology* 726, 17–47.
- Zhou, Z., Hardt, S., Wang, B., *et al.*, 1996. CTF determination of images of ice-embedded single particles using a graphics interface. *Journal of Structural Biology* 116, 216–222. doi:10.1006/jsbi.1996.0033.

Analysis of Viruses in the Cellular Context by Electron Tomography

Peijun Zhang, University of Oxford, Oxford, United Kingdom and Electron Bio-Imaging Centre, Diamond Light Source, Didcot, United Kingdom

Luiza Mendonça, University of Oxford, Oxford, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

3DEM Three-dimensional electron microscopy

CEMOVIS Cryo-Electron Microscopy Of Vitreous Samples

CLEM Correlative Light and Electron Microscopy

Cryo-EM Cryo-electron microscopy

Cryo-ET Cryo-electron tomography

Cryo-FIB/SEM Cryo-Focused Ion Beam Scanning Electron Microscopy

EM Electron Microscopy

ET Electron Tomography

FIB/SEM Focused Ion Beam Scanning Electron Microscopy

HPF High Pressure Freezing

HPF/FS High Pressure Freezing and Freeze Substitution

RT-EM Room temperature-electron microscopy

SEM Scanning Electron Microscopy

TEM Transmission Electron Microscopy

Introduction

Viruses are parasites which hijack and depend on cellular functions for their own replication, indeed, the viral replicative cycle obligatorily takes place inside a host cell. Therefore, studying viruses in the cellular context is the most biologically relevant way to study the processes occurring during viral infection. The viral replication cycle can be generalized to the following steps: attachment, entry, replication, assembly and release. Some viruses also undergo a maturation step. All these steps are amenable to imaging by electron microscopy, using one or more techniques.

The development of electron microscopy contributed significantly to the advancement of virology (Kausche and Ruska, 1939), as viruses are beyond the diffraction limit of light microscopy. That being the case, the evolution of virology has gone hand-in-hand with advances in electron microscopy. In fact, electron microscopy was and still is used as a powerful diagnostic tool in the clinical setting, in some cases capable of identifying an etiological viral agent at the genus level (Goldsmith, 2014).

Electron microscopy relies, as the name implies, on electrons as the illumination source. While electron microscopy can yield a wealth of information at nanometer and subnanometer levels of detail, it also has limitations. The electron beam path in the microscope column needs to be a vacuum. As a result, an ultra-low pressure is present inside the column and, without preventative measures, this would cause evaporation of any water that is present in the sample. For this reason, the sample needs to be treated to withstand the vacuum. The two main kinds of treatment for this purpose are (1) chemical fixation followed by dehydration, resin embedding and (2) physical-fixation/cryo-immobilization (which may or may not be followed by freeze substitution, resin embedding). Both methods are discussed in the following sections, and, together, they divide the field of electron microscopy into two major “flavors”: room temperature-electron microscopy (RT-EM) and cryo-electron microscopy (cryo-EM).

Electrons also have a very limited penetration ability. Therefore, if the specimen is to be analyzed by Transmission Electron Microscopy (TEM), it must be very thin, usually less than 300 nm thick (Lučič *et al.*, 2013). Cell thickness varies from cell to cell and from one point of the cell to another (e.g., the nucleus is much thicker than filopodia). Depending on the cellular region of interest, thinning of the specimen may be required. This can be achieved by microtome sectioning or ion-beam milling of the sample into ultrathin slices 70–200 nm thick (Romero-Brey and Bartenschlager, 2015).

In the case of a virus-infected cell, the ultrathin sections may be rich in information but must be considered in the context of the much larger cellular volume; this can be difficult. Further, 2D projection images recorded from ultrathin sections may be hard to interpret and even mis-interpreted. As an illustrative example, for many years, it was debated whether HIV can bud into endosomes in infected macrophages or if HIV exclusively buds from the cell plasma membrane. It took 3D electron microscopy to finally settle the matter by describing a new type of compartment, Intracellular Plasma-Membrane Connected Compartments (IPMCs) which may resemble endosomes in 2D cross-section micrographs, but are in fact connected to the plasma membrane through complex membrane folds (Bennett *et al.*, 2009; Welsch *et al.*, 2011).

Several microscopy techniques offer 3D representation of an imaged object. In serial sectioning EM, successive microtomy slices of a sample are imaged and stitched together to reconstruct a thicker volume (Sjostrand, 1958). Serial block face imaging works on a similar concept, but instead of having a slice ribbon mounted on an EM grid, the sample block is mechanically sliced directly inside the microscope with a microtome, and the newly exposed surface is imaged (Denk and Horstmann, 2004). In FIB/SEM, the same principle is applied, instead of using a microtome, successive surface layers are exposed by Ion Abrasion (Heymann *et al.*, 2006). These 3D electron imaging techniques are applied to study biological samples at different spatial scales, which can be combined and integrated with optical imaging modalities through correlative light and electron microscopy (Jun *et al.*, 2011; Moser *et al.*, 2019). Here we focus discussions on electron tomography.

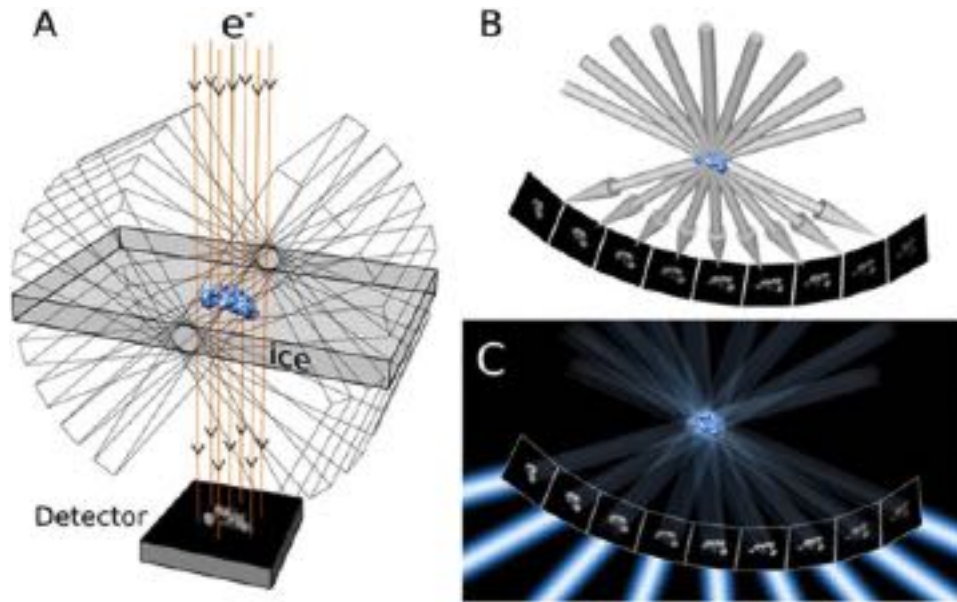


Fig. 1 From a tilt-series to a tomogram. (A) In the electron microscope, a fixed sample (either in vitreous ice or resin) is imaged multiple times, each time tilted to a different and known angle. (B) This collection of images is called a tilt-series. (C) From this series of projection images, a 3D object, called a tomogram, can be calculated (Koning *et al.*, 2018).

Electron tomography is one of many 3D imaging techniques, by which the same region of interest is imaged multiple times, but each time at a different angle, resulting in a tilt-series. These images can be processed computationally to produce a 3D object that represents the original sample volume (Crowther *et al.*, 1970) and is called a tomogram (Fig. 1).

Limitations of electron tomography include electron radiation damage and the missing-wedge effect (Radermacher, 1988). Since the same region is imaged multiple times, at different specimen tilt angles, the accumulated electron dose may induce alterations to the structures imaged. For that reason, careful electron dose management is necessary, particularly in cryo-fixed samples (Hagen *et al.*, 2017). The missing wedge is an artifact of electron tomography. It is a product of both the low penetrating power of electrons and limited mechanics of the sample stage. As the sample is tilted, the sample becomes increasingly thicker, to a point where the transmitted signal becomes too low. In addition, microscope sample stages are typically limited to 70 degrees tilting due to stage stability and the limited space available inside the microscope column. Due to the missing wedge problem, the 3D representation of the imaged object is not perfect, but elongated in the direction parallel to the beam path (Radermacher, 1988).

Sample Preparation

As mentioned, the manner in which a sample is fixed and treated to withstand the vacuum of the microscope column divides electron tomography into two types: Room-temperature electron tomography and cryo-electron tomography. This division is not always mutually exclusive. Some techniques, such as High Pressure Freezing and Freeze Substitution require initial cryo-fixation of the sample, but the imaging is done in room-temperature (after the sample is freeze-substituted to resin and sliced).

Chemical Fixation

Chemical fixation uses cross-linking chemical reagents (formaldehyde and glutaraldehyde are the most commonly used) to stabilize the biological components in the cell. Traditionally, to enhance contrast and improve interpretability of images, stains containing heavy atoms (e.g., Uranium, Osmium, Molybdenum) are used. After fixation, the water present in the sample needs to be replaced by resin, allowing the sample to withstand the vacuum. The resins used in electron microscopy (e.g., epoxy) are very hydrophobic, so an intermediate step of sample dehydration is necessary. The sample undergoes a series of incubations with solvents at increasing concentrations, to gradually replace all the water in the sample with solvent (most commonly ethanol or acetone). The solvent is then replaced by the resin, also in step-wise fashion. After polymerization of the resin, the sample block is sectioned in thin slices (70–300 nm thick), attached to an electron microscopy grid, and imaged in the electron microscope (Griffiths, 1993). Although this time-consuming and labor-intensive procedure can introduce artifacts in the samples and/or destroy some sensitive ultrastructural details (Small, 1981), much of what is currently known in virology was the product of this technique, and the field of virology owes a lot to it.

Cryo-Immobilization

In the last three decades technological advances have enabled the imaging of biological specimens in their near-native state in the absence of stains, especially the development of cryo-fixation through vitrification and cryo-electron microscopy (cryo-EM) (Taylor and Glaeser, 1976; Dubochet and McDowell, 1981). At very low temperatures (below -150°C), the issue of water evaporation in the vacuum of the microscope column is eliminated, and ice serves as fixative. However, only ice in a vitreous form can preserve a biological sample in a near-native state; formation of crystalline ice perturbs the ultrastructural details in the sample. Vitrification of thin specimen ($<2\ \mu\text{m}$) can be achieved by rapid plunge-freezing into liquid ethane or a liquid ethane/propane mixture (Dubochet and McDowell, 1981).

Vitrification of thicker and larger sample volumes (μm^3 to mm^3), such as tissue, can be achieved by High Pressure Freezing (HPF) (Moor *et al.*, 1980). Such samples, therefore, need to be thinned prior to imaging in the electron microscope. This can be achieved by maintaining cryo-preservation using either CEMOVIS (Cryo Electron Microscopy of Vitrified Samples) (Hsieh *et al.*, 2002) or cryo-FIB/SEM milling (Marko *et al.*, 2007). Alternatively, the sample may be slowly dehydrated by a technique called freeze substitution (FS), then resin-embedded and sliced in plastic sections (Hawes *et al.*, 2007).

Room Temperature-Electron Tomography

In room temperature-electron tomography (RT-ET), heavy-metal stained sections of resin embedded sample are typically imaged (Romero-Brey and Bartenschlager, 2015). In this technique, a cell culture or tissue is chemically fixed, stained, dehydrated, resin-embedded and sliced. The slices are then mounted on an EM grid and loaded onto the electron microscope. Tilt-series are collected at the regions of interest and then computationally converted to tomograms (Fig. 2, top left panel). This technique allows imaging of any part of the cell or tissue. The judicious use of chemical fixatives and immunocytochemistry-friendly resins may preserve antigenicity of the sample and allows direct immunostaining and gold-labeling, enabling identification of target structures in the electron micrographs (Griffiths, 1993).

An extended version of RT-ET is serial tomography, where consecutive sections of a cell or tissue are imaged by ET and then combined to build a larger volume. This is a powerful technique as it combines the high-resolution ET and the possibility of reconstructing up to mm^3 of a sample. The sample may be chemically fixed or be processed by HPF/FS (High Pressure Freezing/Freeze Substitution) before resin embedding and sectioning.

The application of RT-ET is exemplified by imaging of virus infected tissues. In Kieffer *et al.*, 2017, the authors pursued multiscale imaging of gut-associated lymphoid tissue, the female reproductive tracts and spleens of humanized mice that had been infected with HIV (Kieffer *et al.*, 2017). By combining 3D immunofluorescence with electron tomography they described systemic viral spread during early infection. Apart from the wealth of environmental context contained in the tissue samples, such as the presence of blood vessels near infected cells in multiple tissues, or the collagen fibrils surrounding pools of viruses in spleen samples, the study also allowed observations regarding the timing and progression of tissue infection and the routes of HIV spread after the initial infection.

Cryo-Electron Tomography

In cryo-ET, the sample is held at a temperature below -160°C and the electron dose used for imaging is kept low (typically $60\text{--}120\ \text{e}^- \text{\AA}^{-2}$) to reduce radiation damage to the specimen. Since the sample is preserved and imaged in a frozen-hydrated state, no additional sample manipulation is necessary and little ultrastructural detail is compromised (McDowell *et al.*, 1989). Due to the absence of stain and the use of low electron doses, the contrast is poor. Nonetheless, the structural information is present in its entirety. If a tomogram depicts multiple copies of the same object, they may be computationally combined to enhance the information of the object of interest. This is particularly useful in virology. Viruses, as very simple pathogens, rely on small genomes that encode few ORFs (Open Reading Frames) but can use multiple copies of the same protein to build a larger structure. A good example of this is the viral capsid. Many copies of morphological building blocks (capsomers) assemble to form a protein shell that protects other viral components. Capsomers in a tomogram (or in multiple tomograms) can be averaged to produce a high-resolution depiction of the capsomer unit. This strategy may yield maps at near-atomic resolution (Himes and Zhang, 2018). This computational approach is called sub-tomogram averaging, and may be used not only for capsomers, but also in any other case where repeated units of the same structure are present and can be averaged, such as envelope spikes and viral portals (Cardone *et al.*, 2007; Liu *et al.*, 2008).

A major caveat of cryo-ET is that it is not amenable to direct immunofluorescence or gold-labeling. This could prevent unambiguous protein identification in macromolecular complex structures. Identification of protein components in cryo-microscopy may be achieved by comparative studies of wild-type and knock-out structures, or wild-type and fusion proteins. Extra density (in the case of fusion proteins) or an absence density (in the case of knock-out of a protein present in a macromolecular complex) can aid the identification of a target protein (Chang *et al.*, 2017). Another way of identifying a structure is to achieve a high-resolution map that enables confident fitting of existing atomic models.

Cryo-ET of Cell Periphery

The easiest way to comply with the requirement for thin samples is to study the thin periphery of cells. Fortunately, many stages of viral replication happen at or near the plasma membrane. Adhesion, fusion, and entry are early stages of viral replication that are

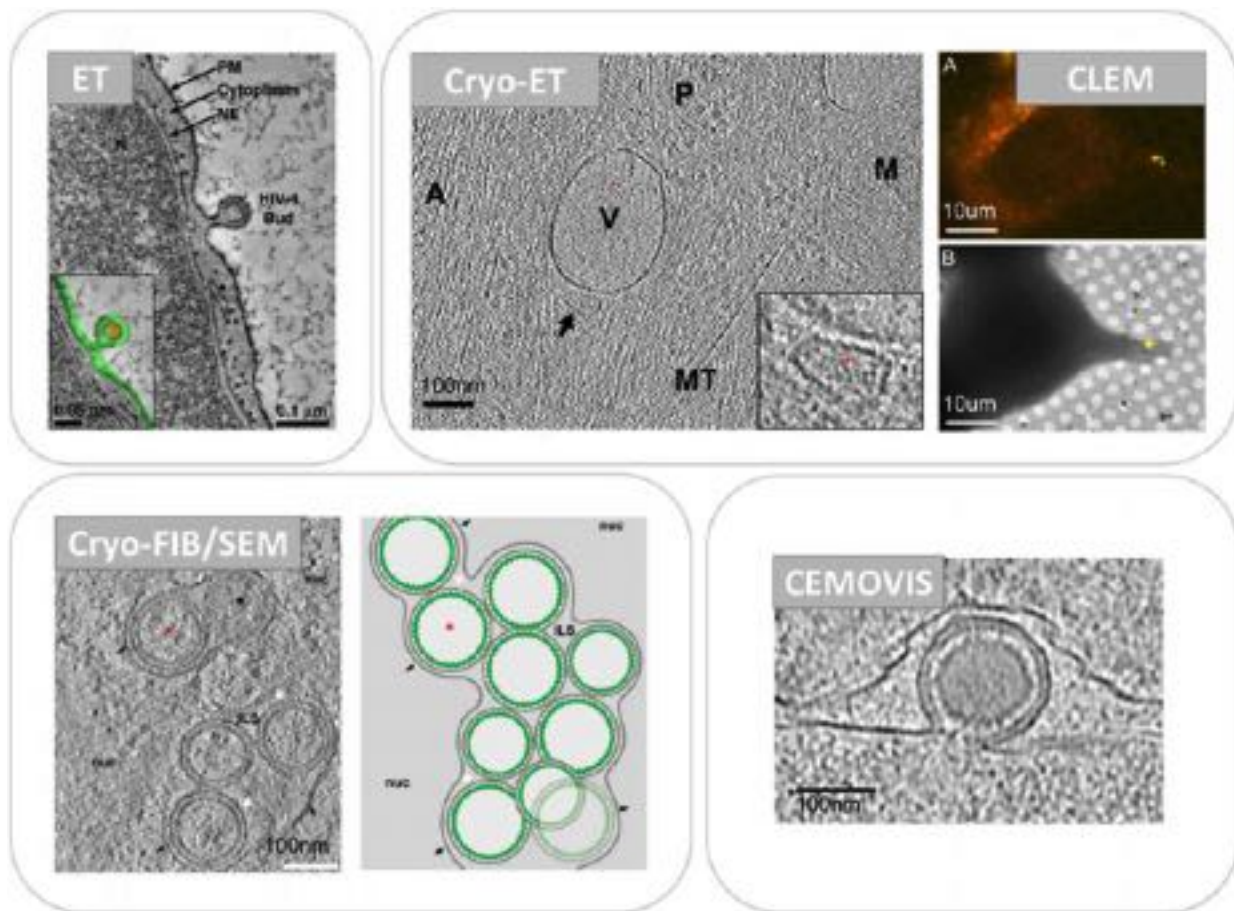


Fig. 2 Examples of different techniques used to study viruses in the cellular context. Top Left: Electron Tomography. Tomographic slice of an HIV particle budding from a T-lymphocyte. PM – Plasma Membrane, NE – Nuclear envelope, N – Nucleus. Inset shows a segment of the bud with a viral immature lattice in yellow and the lipid membrane in green (from [Ladinsky et al., 2019](#)). Top Right: Cryo-electron tomography at the cell periphery. Tomographic slice of an HIV core (arrow) inside a HeLa cell. A – actin filaments, MT – microtubule, V – vesicle, M – mitochondria, P – protein complexes. An enlarged view of the HIV core structure is in the inset. The viral core location was found using CLEM (Correlated Light and Electron Microscopy). Panel A shows the path taken by a fluorescently-tagged viral particle inside a HeLa cell (yellow tracing). The same region is imaged by cryo-EM at low magnification in panel B (from [Jun et al., 2011](#)). Bottom Left: Cryo-FIB/SEM. Tomographic slice of a cell lamella depicting the HSV-1 nuclear egress complex in the nuclear membrane intraluminal space. Black arrows point to inner nuclear membrane and white arrows to vesicle membranes. Nuc – Nucleus, ILS – Intraluminal space (from [Hagen et al., 2015](#)). Bottom Right: CEMOVIS. Tomographic slice of an HSV-1 infected Vero cell depicting a viral capsid budding into the perinuclear space (from [Hagen et al., 2015](#)).

well suited for study without the need of thinning ([Jun et al., 2011](#); [Riedel et al., 2017](#)), as are the late stages of assembly and release (when happening at the cellular membrane). These can be imaged directly in cells growing on EM grids ([Bharat et al., 2011](#)). Spread and dissemination (either between cells or within one cell, such as retrograde transport) may also be processes well-suited for study at cell periphery ([Ibircu et al., 2011](#); [Mueller et al., 2014](#)).

Target cells can be grown directly on the surface of a disinfected carbon-coated EM grid, typically a gold grid which is not toxic to cells. Some cells may attach directly to the carbon film support of the grid, but coatings like Poly-L-Lysine and fibronectin may be used to improve cell adherence. The cells can be infected with the virus of choice and incubated for a period of time appropriate to study the desired viral replication stages. Then, EM grids can be plunge-frozen in a cryogen (ethane or ethane/propane mixture). The thin periphery region of the cell (or filopodia, or axons) can be imaged by cryo-ET, resulting in a 3D representation of the imaged volume ([Fig. 2](#), top left panel) ([Ibircu et al., 2011](#); [Jun et al., 2013](#)).

Using cryo-ET, Riedel and colleagues were able to characterize the interaction between the viral envelope spike of Murine Leukemia Virus and the host receptor mCAT ([Riedel et al., 2017](#)). By imaging virus particles attached to the target cells and analyzing the contact extent between viral and cell membranes they were able to determine that viral attachment was in its early to intermediary stage. They also determined the average number of receptors engaged in MLV entry. Clustering of the receptors at the interface of virus and cell membrane was observed, indicating that lateral mobility of membrane proteins in the viral envelope is possible. Sub-tomogram averages of the spike show that there are large conformational changes between the pre and post fusion conformations. This is a good illustrative example of all the insights that may be derived from 3D imaging of viral processes at the cell periphery.

Cryo-ET of Cell Lamellae or Vitreous Sections

Some viral processes occur deep inside the cell, such as virus trafficking, nuclear entry, and assembly in viroplasma/viral factories. In these cases, two techniques are typically used to generate cell lamellae or sections from frozen-hydrated virus-infected cells for cryo-ET: Cryo-FIB/SEM (Focused Ion Beam/Scanning Electron Microscopy) milling and CEMOVIS.

In Cryo-FIB/SEM milling, a dual-beam system with a cryo-sample stage is used to process frozen-hydrated samples. An electron beam is used to visualize the sample and monitor the milling process. An Ion Abrasion beam is used to remove material from frozen-hydrated cells or tissues. Successive milling is made until a lamella, a thin sliver of sample typically 100–300 nm thick, is produced, which is then imaged by cryo-ET ([Fig. 2](#), bottom left panel) ([Marko et al., 2007](#)).

In CEMOVIS, a sample is cryo-fixed by HPF and is sectioned mechanically by a diamond knife in a cryo-ultramicrotome to produce thin vitreous sections (around 200 nm thick). The sections are transferred to EM grids and imaged by cryo-ET ([Al-Amoudi et al., 2004](#)). Successive sections may be imaged to increase the volume ([Fig. 2](#), bottom right panel). Nonetheless, this process is very challenging and may introduce artifacts such as knife marks and compression in the direction of the cutting ([Al-Amoudi et al., 2005](#)).

A great illustrative example of both techniques being used to study viral processes in the cellular context can be found in [Hagen et al., 2015](#) ([Hagen et al., 2015](#)). The authors used both milling and CEMOVIS to study the nuclear egress of Herpes Simplex 1. By observing viral egress in the native cellular environment, the authors were able to conclude that the nuclear egress complex (NEC) proteins were lost after the transfer of the viral capsid from the nucleus to the cytoplasm, and that the capsid proceeded to further assembly thereafter, correcting a notion born from observations of stained plastic-embedded samples. They also showed the NEC lattice organization and ultrastructure by sub-tomogram averaging and describe the mechanism by which the protein curvature drives the membrane curvature. This work is also a good example of sub-tomogram averaging of repeating NEC subunits.

Acknowledgments

We thank Dr. Teresa Brosenitsch for critical reading the manuscript.

References

- Al-Amoudi, A., Norlen, L.P., Dubochet, J., 2004. Cryo-electron microscopy of vitreous sections of native biological cells and tissues. *Journal of Structural Biology* 148 (1), 131–135.
- Al-Amoudi, A., Studer, D., Dubochet, J., 2005. Cutting artefacts and cutting process in vitreous sections for cryo-electron microscopy. *Journal of Structural Biology* 150 (1), 109–121.
- Bennett, A.E., et al., 2009. Ion-abrasion scanning electron microscopy reveals surface-connected tubular conduits in HIV-infected macrophages. *PLoS Pathogens* 5 (9), e1000591.
- Bharat, T.A., et al., 2011. Cryo-electron tomography of Marburg virus particles and their morphogenesis within infected cells. *PLoS Biology* 9 (11), e1001196.
- Cardone, G., et al., 2007. Visualization of the herpes simplex virus portal in situ by cryo-electron tomography. *Virology* 361 (2), 426–434.
- Chang, Y.W., et al., 2017. Architecture of the *Vibrio cholerae* toxin-coregulated pilus machine revealed by electron cryotomography. *Nature Microbiology* 2, 16269.
- Crowther, R.A., et al., 1970. Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs. *Nature* 226 (5244), 421–425.
- Denk, W., Horstmann, H., 2004. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biology* 2 (11), e329.
- Dubochet, McDowell, 1981. Vitrification of pure water for electron microscopy. *Journal of Microscopy* 124 (3), 2.
- Goldsmith, C.S., 2014. Morphologic differentiation of viruses beyond the family level. *Viruses* 6 (12), 4902–4913.
- Griffiths, G., 1993. *Fine Structure Immunocytochemistry*. Springer.
- Hagen, W.J.H., Wan, W., Briggs, J.A.G., 2017. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of Structural Biology* 197 (2), 191–198.
- Hagen, C., et al., 2015. Structural basis of vesicle formation at the inner nuclear membrane. *Cell* 163 (7), 1692–1701.
- Hawes, P., et al., 2007. Rapid freeze-substitution preserves membranes in high-pressure frozen tissue culture cells. *Journal of Microscopy* 226 (Pt 2), 182–189.
- Heymann, J.A., et al., 2006. Site-specific 3D imaging of cells and tissues with a dual beam microscope. *Journal of Structural Biology* 155 (1), 63–73.
- Himes, B.A., Zhang, P., 2018. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nature Methods* 15 (11), 955–961.
- Hsieh, C.E., et al., 2002. Electron tomographic analysis of frozen-hydrated tissue sections. *Journal of Structural Biology* 138 (1–2), 63–73.
- Ibricic, I., et al., 2011. Cryo electron tomography of herpes simplex virus during axonal transport and secondary envelopment in primary neurons. *PLoS Pathogens* 7 (12), e1002406.
- Jun, S., et al., 2011. Direct visualization of HIV-1 with correlative live-cell microscopy and cryo-electron tomography. *Structure* 19 (11), 1573–1581.
- Jun, S., et al., 2013. Correlative microscopy for 3D structural analysis of dynamic interactions. *Journal of Visualized Experiments*. (76).
- Kausche, Ruska, P., 1939. Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop. *Naturwissenschaften* 27 (18), 8.
- Kieffer, C., et al., 2017. Longitudinal imaging of HIV-1 spread in humanized mice with parallel 3D immunofluorescence and electron tomography. *eLife* 6.
- Koning, R.I., Koster, A.J., Sharp, T.H., 2018. Advances in cryo-electron tomography for biology and medicine. *Annals of Anatomy* 217, 82–96.
- Ladinsky, M.S., et al., 2019. Mechanisms of virus dissemination in bone marrow of HIV-1-infected humanized BLT mice. *eLife* 8.
- Liu, J., et al., 2008. Molecular architecture of native HIV-1 gp120 trimers. *Nature* 455 (7209), 109–113.
- Lučić, V., Rigort, A., Baumeister, W., 2013. Cryo-electron tomography: the challenge of doing structural biology in situ. *Journal of Cell Biology* 202 (3), 407–419.
- Marko, M., et al., 2007. Focused-ion-beam thinning of frozen-hydrated biological specimens for cryo-electron microscopy. *Nature Methods* 4 (3), 215–217.
- McDowell, A., et al., 1989. The structure of organelles of the endocytic pathway in hydrated cryosections of cultured cells. *European Journal of Cell Biology* 49 (2), 281–294.
- Moor, H., et al., 1980. The influence of high pressure freezing on mammalian nerve tissue. *Cell and Tissue Research* 209 (2), 201–216.
- Moser, F., et al., 2019. Cryo-SOFI enabling low-dose super-resolution correlative light and electron cryo-microscopy. *Proceedings of the National Academy of Sciences of the United States of America* 116 (11), 4804–4809.

- Mueller, J., *et al.*, 2014. Electron tomography and simulation of baculovirus actin comet tails support a tethered filament model of pathogen propulsion. *PLoS Biology* 12 (1), e1001765.
- Radermacher, M., 1988. Three-dimensional reconstruction of single particles from random and nonrandom tilt series. *Journal of Electron Microscopy Technique* 9 (4), 359–394.
- Riedel, C., *et al.*, 2017. Native structure of a retroviral envelope protein and its conformational change upon interaction with the target cell. *Journal of Structural Biology* 197 (2), 172–180.
- Romero-Brey, I., Bartenschlager, R., 2015. Viral infection at high magnification: 3D electron microscopy methods to analyze the architecture of infected cells. *Viruses* 7 (12), 6316–6345.
- Sjostrand, F.S., 1958. Ultrastructure of retinal rod synapses of the guinea pig eye as revealed by three-dimensional reconstructions from serial sections. *Journal of Ultrastructure Research* 2 (1), 122–170.
- Small, J.V., 1981. Organization of actin in the leading edge of cultured cells: influence of osmium tetroxide and dehydration on the ultrastructure of actin meshworks. *Journal of Cell Biology* 91 (3 Pt 1), 695–705.
- Taylor, K.A., Glaeser, R.M., 1976. Electron microscopy of frozen hydrated biological specimens. *Journal of Ultrastructure Research* 55 (3), 448–456.
- Welsch, S., *et al.*, 2011. Architecture and regulation of the HIV-1 assembly and holding compartment in macrophages. *Journal of Virology* 85 (15), 7922–7927.

Further Reading

- Risco, C., de Castro, I.F., Sanz-Sánchez, L., *et al.*, 2014. Three-dimensional imaging of viral infections. *Annual Review of Virology* (1), 453–473. doi:10.1146/annurev-virology-031413-085351.

Relevant Websites

- https://en.wikipedia.org/wiki/Viral_disease
Viral disease - Wikipedia.
- https://www.onhealth.com/content/1/viral_infections
What's a Virus Viral Infection Types, Symptoms, Treatment.

Mathematical Modeling of Virus Architecture

Reidun Twarock, University of York, York, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

CKT Caspar-Klug theory

CP Capsid protein

HP Hamiltonian path

HPA Hamiltonian path analysis

PS Packaging signal

T The triangulation number

T_t The trihexagonal number

VTT Viral tiling theory

Glossary

Affine extension A geometric transformation that preserves lines and parallelism (but not necessarily distances and angles).

Archimedean lattices (also known as uniform tilings) Tessellations in terms of regular polygons that have the same vertex environment throughout, i.e., the sequence and types of polygons surrounding each vertex are identical across the lattice (and are therefore used to label the lattice).

Asymmetric unit (also called: fundamental domain) The smallest unit from which a symmetric structure can be generated using the operations of the symmetry group.

Capsomers The units of capsid assembly, typically clusters of two (dimers), three (trimers), five (pentamers) or six (hexamers) capsid proteins.

Caspar-Klug theory (in short: CKT) Mathematical models of virus architecture based on the principle of quasi-equivalence, positing that capsid proteins should form the same local interactions at all positions in the capsid surface.

Gillespie-type algorithms Generate a statistically correct trajectory (possible solution) of a stochastic equation system for which the reaction rates are known.

Goldberg polyhedral Polyhedra with 12 pentagonal and otherwise hexagonal faces.

Gyrated lattice A gyrated version of the trihexagonal lattice is obtained if the edge lengths of triangular and hexagonal tiles are distinct.

Hamiltonian Paths Analysis (in short: HPA) Geometric constraints on the conformation of the packaged viral genome inside its capsid are used, in combination with bioinformatics, to identify multiple dispersed packaging signals in viral genomes.

Icosahedral Symmetry Symmetry defined by the rotational symmetries of an icosahedron, given by an ensemble of six 5-fold, ten 3-fold, and fifteen 2-fold axes.

Laves lattice Duals to Archimedean lattices, with vertices at the centers of its faces, and edges connecting vertices in neighboring faces.

Levinthal's Paradox in protein folding The conundrum of how proteins achieve their biologically functional native state swiftly via specific folding pathways, instead of a random exploration of all combinatorially possible ones. Virus assembly poses a viral equivalent to Levinthal's Paradox – how do viruses navigate the complex landscape

of combinatorially possible assembly pathways efficiently to ensure swift completion of viable capsid geometries.

Packaging signal mediated assembly An assembly mechanism reliant on the collective action of multiple dispersed packaging signals.

Packaging signal Nucleotide sequence/structure element in a viral genome that directs the packaging of the viral genome.

Polyhedron A three-dimensional shape with flat polygonal faces, straight edges and sharp corners called vertices.

Principle of genetic economy A principle introduced by Crick and Watson, arguing that viral genomes dedicate as little as possible coding capacity to viral proteins. As a consequence, viral capsids are formed from multiple copies of identical capsid proteins.

Quasi-equivalence A term used to describe protein positions in viral capsids in Caspar-Klug theory, meaning that protein positions are locally equivalent.

Traveling salesman problem The combinatorial problem of finding the shortest connected path that visits a given number of cities (nodes on a graph) precisely once before returning to the city (node) of origin.

Triangulation number series (in short: T-number series) A series of polyhedra used as models of capsid architecture in Caspar-Klug theory. The polyhedra are derived from a hexagonal lattice and are labeled in terms of the triangulation number *T*.

Triangulation Tessellations formed exclusively from triangles. The duals of the Goldberg polyhedra – triangulations called geodesic polyhedra – are used in Caspar-Klug theory as models of capsid architecture.

Trihexagonal lattice series (in short: T_t-number series) A series of polyhedra that are derived from the trihexagonal lattice and are labeled in terms of the trihexagonal number *T_t*. They provide alternative models to the surface lattices in Caspar-Klug theory, accounting also for virus architectures formed from more than one type of capsid protein, such as a combination of a major and minor capsid protein, and non-quasiequivalent capsid architectures.

Viral Capsid Protein container enclosing the genetic material of a virus.

Viral Tiling theory (in short: VTT) A mathematical theory representing viral capsids as surface tessellations called tilings. It generalizes Caspar-Klug theory and accommodates non-quasiequivalent surface architectures such as those of the polyoma- and papillomaviruses.

Introduction

Viruses package their genetic material into protein containers, called capsids, that provide protection between rounds of infection. In the vast majority of cases viral capsids adopt polyhedral shapes with icosahedral symmetry. Mathematical modeling can therefore be used to predict and classify virus architecture. Starting with an exposition of the classical Caspar Klug theory, further developments in the modeling of virus architecture are covered, that culminate in an overarching theory of icosahedral virus architecture. This theory contains Caspar Klug theory as a special case, and also covers its outliers, in particular larger and more complex viruses that are ill described by Caspar and Klug's polyhedral models. Demonstrations are given of how such mathematical models of virus architectures impact on our understanding of viral life cycles and viral evolution, and inform applications in nanotechnology.

Icosahedral capsid architecture also has implication for genome organization, mediated via multiple dispersed contacts between capsid protein and the packaged genome. Hamiltonian Paths Analysis (HPA) enables the formulation of geometric constraints on genome organization arising from these contacts. The predictive power of HPA is illustrated via its role in the discovery of capsid assembly instructions that occur in many viral genomes in the form of multiple dispersed sequence/structure motifs with affinity for capsid protein, termed packaging signals (PSs). HPA has played a major role in their discovery and characterization, and in the analysis of the PS-mediated assembly mechanism, that resulted in a paradigm shift in our understanding of virus assembly. The asymmetric organization of the viral genome inside the capsid is only one form of asymmetry in virus architecture. Other asymmetric components often occur as an integral part of the capsid shell, and such symmetry breaking in icosahedral capsids has functions that are discussed from a modeling perspective.

The Principle of Genetic Economy

A reason for the polyhedral nature of viral capsids has been given in a seminal paper by Crick and Watson. They propose that viral capsids are either helical or polyhedral due to the *principle of genetic economy* (Crick and Watson, 1956). By repeatedly synthesizing capsid protein subunits from the same genomic segment, viruses minimize the portion of their genomes dedicated to the coding of the capsid. The repeated use of the same type of capsid protein, on the other hand, results in architectures that are highly symmetrical, either helical or polyhedral.

Viral Geometry and Icosahedral Symmetry

The majority of viral capsids have polyhedral shapes, exhibiting the same rotational symmetries as an icosahedron. This means that the organization of their protein subunits in the capsid surface respects the rotational symmetry axes of icosahedral symmetry. The locations of four of these axes are indicated for the example of Satellite Tobacco Necrosis virus (STNV; PDB-id 4BCU), a small plant virus, with respect to an icosahedral reference frame in Fig. 1(a). There are in total 12 axes of 5-fold rotational symmetry through opposite vertices of the icosahedron; 20 3-fold axes through the midpoints of opposite faces; and 30 axes of 2-fold symmetry through the midpoints of opposite edges. Icosahedral symmetry implies that the capsid surface must be tessellated by 60 identical copies of the so-called asymmetric unit, or fundamental domain, of the symmetry group, that is given by the kite-shaped area indicated in green in Fig. 1(a).

If a virus capsid is formed from 60 capsid protein subunits, like STNV, then there is precisely one capsid protein per copy of the asymmetric unit, and the positions of all capsid proteins are therefore fully determined by symmetry. As the icosahedral symmetry group is the largest rotational symmetry group in three dimensions, i.e., has the largest number of elements, it guarantees the largest number of repeats of the asymmetric unit in the capsid surface. Polyhedral capsids with icosahedral symmetry therefore optimize container volume. This facilitates the packaging of the genomic cargos, which provides an explanation for the prevalence of icosahedral symmetry in virology.

Mathematical Models of Icosahedral Capsid Architecture

For viruses formed from more than 60 capsid proteins, there is more than one protein per asymmetric unit. The placement of these proteins within the asymmetric unit is not constrained by symmetry, and other geometric principles are therefore required in order to fully characterise virus architecture in these cases.

Quasi-Equivalence Theory

The first principle introduced to predict the structures of larger capsids is *quasi-equivalence* (Caspar and Klug, 1962). It posits that the identical capsid proteins forming larger capsids must make the same types of local interactions with each other in the capsid shell, that is, they must occupy similar local environments. From a mathematical point of view, this implies that their positions must conform to the organization of a lattice in terms of one type of tile, that can tessellate an icosahedral surface without gaps and overlaps. Caspar and Klug therefore modeled capsid architectures by embedding icosahedral surfaces into a hexagonal lattice (Fig. 1(b) and (c)).

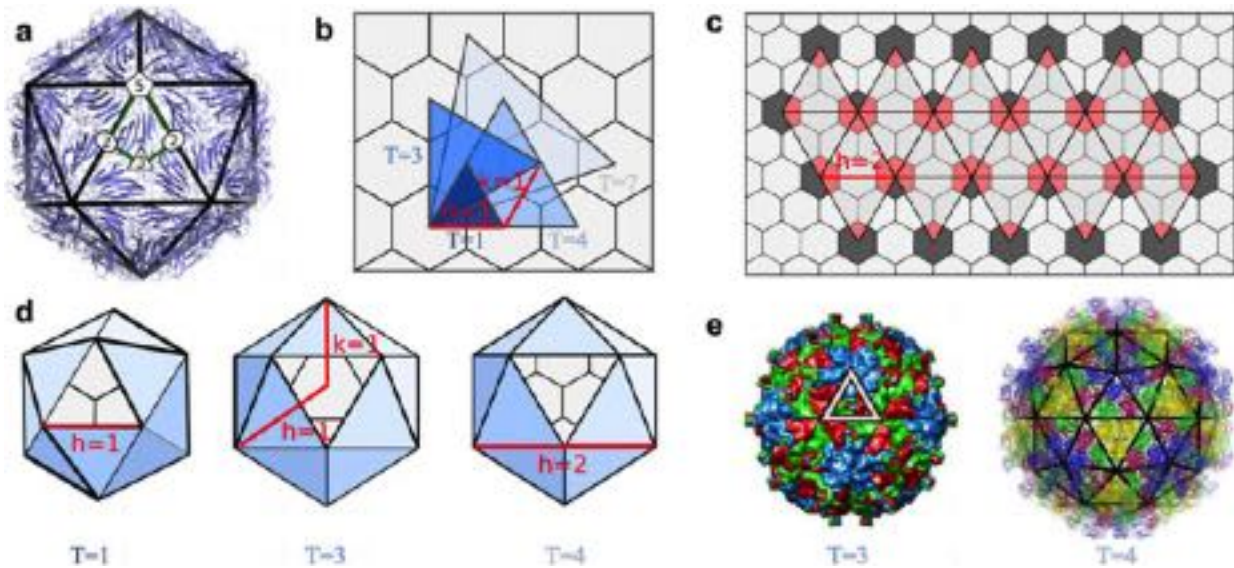


Fig. 1 *Icosahedral Symmetry and Caspar-Klug theory.* (a) Four axes of icosahedral symmetry are shown superimposed on an icosahedral reference frame. The asymmetric unit of the symmetry group is a kite (green) with vertices at these symmetry axes. (b) Different options of embedding an icosahedral surface into a hexagonal grid are labeled in terms of the T -number. Triangles characterising one triangular face of this embedding are shown for the smallest options, showing the pair $(h, k) = (1, 1)$ defining the $T(1, 1) = 3$ architecture explicitly. (c) 20 identical copies of such triangles then define the icosahedral surface (shown here for the $T(2, 0) = 4$ option). (d) By identifying edges and vertices appropriately, this surface can be closed up into a 3D structure. (e) Pariacoto virus (left) and Providence virus (right) are examples of viruses organized according to the dual triangulations of such polyhedral models.

Since the icosahedral surface is formed from 20 equilateral triangles, this embedding is fully characterized once the position of one of the icosahedral edges has been fixed. Any edge connecting the midpoints of two hexagons in the lattice defines a hexagonal surface lattice with icosahedral symmetry. Characterising such edges by counting steps (denoted as h and k) between hexagonal midpoints along two vectors h and k at an angle of 60 degrees allows all possible such surface lattices to be enumerated. The example of $h = k = 1$ is shown in Fig. 1(b). The polyhedra obtained by reconstructing the 3D structures from these surface lattices (cf. Fig. 1(d)) must have 12 pentagonal faces, and $10(T-1)$ hexagonal faces, where the T -number $T = n^2 + nk + k^2$ is given in terms of the positive integers (or zero in at most one case) n and k . Such polyhedra are known as Goldberg polyhedra.

The T -number has a geometric interpretation in terms of the dual polyhedra – triangular surface lattices called geodesic polyhedra. It indicates the number of triangular faces that, by area, cover each icosahedral face, and is therefore also called the *triangulation number*. The case $T = 1$ corresponds to the icosahedron itself. Given that an icosahedron has 20 faces, the next larger and larger values describe triangulations of the icosahedral surface with $20T$ triangular faces.

Predictions and Limitations of Quasi-Equivalence Theory

Caspar and Klug interpret these polyhedral blueprints as follows: Since each triangular face has 3-fold symmetry, they assign three protein subunits within it. Therefore, the $20T$ faces accommodate precisely $60T$ capsid proteins. This implies that the protein numbers in the viral capsid models in Caspar Klug theory (CKT) are quantized. The smallest ‘allowed’ protein numbers are 60 ($T = 1$), 180 ($T = 3$), and 240 ($T = 4$). Pariacoto virus and Providence virus, shown in Fig. 1(e), exhibit such architectures; many other examples can be found on the ViPER website (Ho *et al.*, 2018), that has a function enabling the user to identify virus structures by T -numbers. The predictions of CKT were long thought to be universally true, and for over half a century, virus architectures have been classified in terms of T -numbers. However, with the development of refined imaging techniques, many outliers to quasi-equivalence theory were discovered with protein numbers violating this restriction, instigating the development of new mathematical approaches.

Viral Tiling Theory

Prominent examples of viruses with capsid protein numbers violating the Caspar and Klug rule are polyoma- and papilloma-viruses. Their capsids are formed from 360 protein subunits, that would correspond to a “disallowed” T -number of 6. The reason for this is that their capsids are formed from 72 clusters of five proteins (called pentamers), rather than the characteristic combination of 12 pentamers and otherwise clusters of six proteins (hexamers) as in Caspar and Klug’s approach. As planar lattices formed from pentagons do not exist without gaps – a result known as the crystallographic restriction – a simple adaptation of the Caspar-Klug construction is not possible in these cases. This conundrum is reminiscent of the mathematical challenges faced in the modeling of quasicrystals, alloys exhibiting long range order but lacking periodicity, that were discovered in 1984 and won Dan

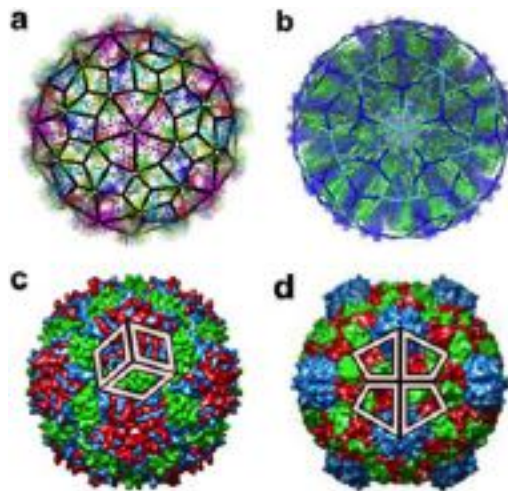


Fig. 2 *Viral tiling theory.* (a) Tilings in terms of kites and rhombs explain the surface architectures of the polyoma- and papillomaviruses, shown here for SV40. (b) An adaptation of Viral Tiling theory to applications in nanotechnology reveals the structure of self-assembling protein nanoparticles used in malaria vaccine design. (c) A rhomb tiling encoding the surface architecture of bacteriophage MS2; and (d) a kite tiling for Tobacco ringspot virus.

Shechtman the Nobel Prize in Chemistry in 2011. Adapting techniques from tiling theory used in the modeling of quasicrystals, in particular Penrose tilings, a tiling approach for the modeling of virus architecture has been developed, that, among others, accounts for the surface structures of the polyoma- and papillomaviruses. The tiling for the polyoma and papilloma capsids is shown in [Fig. 2\(a\)](#), with symmetry equivalent protein subunits color-coded and organized into 72 pentamers. This is the first example ([Twarock, 2004](#)) of a theory now known as Viral Tiling theory (VTT).

In VTT, capsids are represented by icosahedral tilings formed from multiple copies of a finite set of tiles, that have a biological interpretation in terms of interactions between the protein subunits associated with that tile. These can either be interactions *between* capsomers on *within* capsomers. The kite-and-rhomb tiling in [Fig. 2\(a\)](#) is an example of the former. Here, the distinct tile types indicate different types of interactions between proteins in neighboring pentamers. In particular, interactions between pentamers occur either in groups of two (dimer interactions) or three (trimer interactions), and are represented by rhombs and kites in the tiling, respectively. As in CKT, these tiles indicate the positions of the capsid proteins. Note that, however, if tiles represent interactions between capsomers then it is not necessarily the symmetry of the tile that determines the number of proteins represented by it, but the number of edges in the tiling meeting at its vertices. For example, in the polyoma- and papillomavirus tiling, proteins are associated with those corners of a tile that correspond to a 5-coordinated vertex in the tiling.

VTT also includes the option of representing interactions within capsomers by tiles. In this case, the tile must have a footprint that represents the stoichiometry of the capsomer. Examples are shown in [Fig. 2\(c\)](#) and [\(d\)](#): a rhomb tiling representing the positions of the 90 protein dimers in bacteriophage MS2 (c); and a kite tiling representing the 3-domain architecture of Tomato ringspot virus (d). The latter is a pseudo $T = 3$ ($pT = 3$) virus formed from 60 identical copies of a protein, the three domains of which occupy the positions of a $T = 3$ polyhedral model. Note that VTT distinguishes their architectures from the $T = 3$ triangulation in CK theory, i.e., from the example shown in [Fig. 1\(e\)](#). VTT thus discriminates between different capsid layouts with the same number of capsid proteins, that would be represented by the same surface lattice in CKT.

Practical Applications

The tiling models can be used in an interdisciplinary context in order to analyze experimental data on particle assembly. An example of this is the characterization of self-assembling protein nanoparticles (SAPNs) in nanotechnology ([Indelicato et al., 2016](#)). Given information on the protein building blocks in the experiment – in this example a pair of fused helices that form contacts with other copies in groups of five (green helix) and three (blue helix) – the tiling approach enables a classification of all symmetric particle types. These can then be compared with mass spectrometry data of capsid assembly, thus identifying the likely particle morphology in an interdisciplinary context. The model applied to a protein nanoparticle used in the design of Malaria vaccines is shown in [Fig. 2\(b\)](#).

An Overarching Framework for Icosahedral Architectures

These examples beg the question whether there is a geometric principle that can be used to classify all possible icosahedral capsid geometries in virology. In order to address this, it is necessary to revisit the principle of quasi-equivalence from a biological point of view. Quasi-equivalence is based on two assumptions: (i) that capsids are formed from multiple copies of one type of protein subunit; and (ii) that these subunits form the same type of local interaction with each other. Whilst this is the case for a large

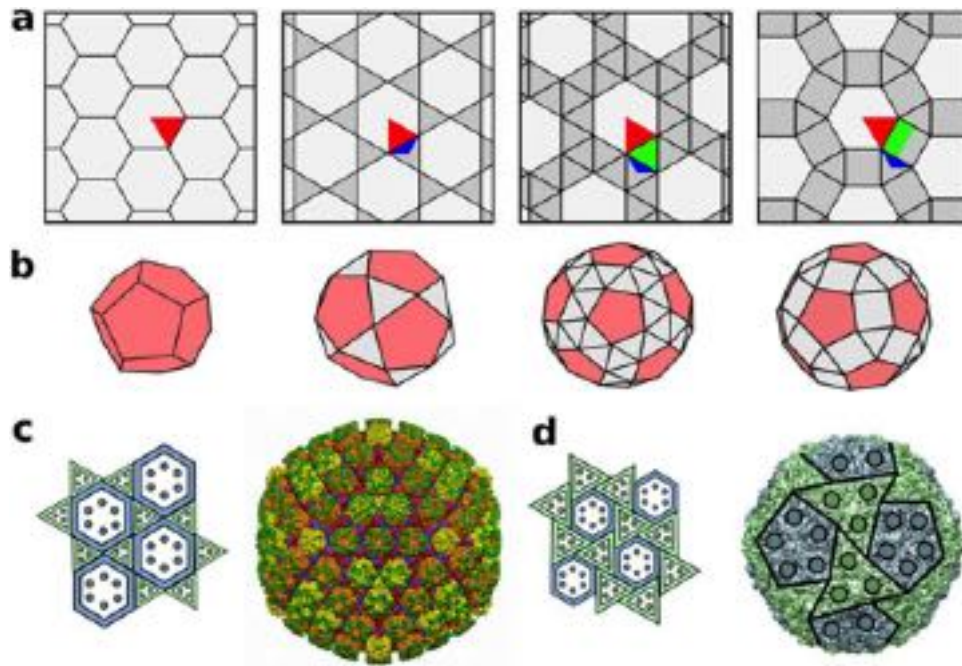


Fig. 3 An overarching design principle of viral architecture. (a) The Archimedean lattices, from left to right the hexagonal, trihexagonal, snub hexagonal and rhombihexagonal lattice, provide the basis for the construction of capsid layouts that contain the surface architectures of Caspar Klug theory as a special case. (b) The first elements of the four polyhedral series derived via an embedding of the icosahedral surface into the lattices in (a). These polyhedral series provide models for: (c) viruses formed from more than one type of capsid protein, such as Herpes Simplex Virus (right) with a trihexagonal lattice architecture (left) in which minor capsid proteins occupy the triangular lattice positions; and (d) outliers to Caspar Klug theory with protein stoichiometries disallowed by the quasi-equivalence principle, such as the 120 CP architecture of Phi6 (right), that follows a gyrated version of the trihexagonal lattice (left).

number of viruses – explaining the success of CKT – there are many examples of viruses that violate these assumptions. In particular larger and more complex viruses are often formed from more than one type of capsid protein, such as a major and minor capsid protein, thus violating (1). Such distinct types of protein subunits do not interact with each other in the same way, resulting in different types of interactions between and within protein types, violating (2). Moreover, as the example of polyoma- and papillomavirus in VTT in Fig. 2(a) demonstrates, even identical protein units do not always exhibit the same type of local interactions, and therefore cannot be expected to result in quasi-equivalent capsid architectures. This suggests that quasi-equivalence must be a special case of a more general biological principle.

Such a generalized principle of quasi-equivalence has been introduced in (Twarock and Luque, 2019a). It is based on the assumption that the local rules according to which different types of proteins (e.g., major and minor capsid proteins) interact with each other should be universal across the capsid, i.e., they assemble locally into the same type of configuration. This minimal assembly principle thus requires that different units organize into surface lattices that are formed from one type of vertex environment (one vertex type), at which a set of (potentially different) polygons meet. Each type of polygon is specific to a given type of capsid protein, and their relative sizes must reflect the footprints of the corresponding protein types in the capsid surface. For example, the areas of the polygons representing major and minor capsid proteins should correlate with their respective sizes.

Planar lattices with this property have been classified by Kepler in his *Harmonices Mundi* in 1619, and are also known as Archimedean lattices or uniform tilings. The hexagonal lattice and its dual, the triangular lattice, that are used in CKT, are special cases, and correspond to the only two examples of such lattices that are formed from a single type of polygon. Mimicking the Caspar and Klug construction for other Archimedean lattices (Fig. 3(a)), we constructed four infinite series of icosahedral polyhedra that contain the Caspar-Klug series of T -number geometries as a special case; the smallest polyhedron in each series is shown in Fig. 3(b). These polyhedra are defined by the equation $T_j(h, k) = c_j T(h, k)$, where the scaling factor c_j is related to the four lattice types in Fig. 3(a), and $T(h, k)$ is the triangulation number from CKT. In particular, c_j takes on the values: $c_h = 1$; $c_t = 4/3$; $c_s = 7/3$; and $c_r = 4/3 + 2/\sqrt{3}$ for the hexagonal, trihexagonal, snub hexagonal, and rhombihexagonal lattices, respectively; a full classification is provided in (Twarock and Luque, 2019a).

Together with their duals, the Laves lattices, these polyhedral series provide blueprints for virus architectures that also encompass the outliers of CKT and VTT, thus expanding the repertoire of allowed protein numbers in capsid architecture. A number of examples of viruses with trihexagonal lattice architectures are provided in (Twarock and Luque, 2019a), including herpes simplex virus type 1 (HSV-1) (Fig. 3(c)). HSV-1, a $T_t(4, 0) = 64/3$ architecture in the extended classification scheme, had previously been classed as a $T(4, 0) = 16$ structure in CKT. Whilst the latter correctly represents the number of major capsid

proteins (MCPs; in this case 960 VP5s) forming the hexamers and pentamers, it is at variance with their orientations in the capsid lattice. The trihexagonal lattice architecture moreover also accounts for the locations of the trimeric complexes formed from the minor capsid proteins (mCPs; here Tr1, Tr2a, and Tr2b) at the interstitial triangular lattice positions (cf. the green triangles in [Fig. 3\(c\)](#), left). Other examples presented in ([Twarock and Luque, 2019a](#)) include Halorubrum sodomense tailed virus 2 (HSTV-2), that has been classified as $T(2,1)=7$, but better agrees with a $T_t(2,1) = 28/3$ trihexagonal structure that also accounts for the gpD-like trimers occupying the interstitial trimeric positions between hexamers and pentamers. Similarly, the thermophilic bacteriophage P23-45 is better described by a $T_t(2,1) = 28/3$, rather than a $T = 7$, surface lattice due to its capsid radius, that reflects the additional surface area occupied by the triangular tiles.

Non-Quasi-Equivalent Trihexagonal Lattice Architectures

The trihexagonal (T_t) lattice geometry predicts the relative sizes of the areas occupied by MCP and mCP as 3:1, as a consequence of the size ratio of the areas covered by a hexagonal (A_H) and triangular (A_T) tile. However, in addition to the T_t lattice itself, the mathematical construction allows for a gyrated version ([Fig. 3\(d\)](#), left) in which the size of the triangles is enlarged relative to that of the hexagons. There is precisely one scaling for which A_H and A_T are such that the hexagonal and triangular positions are predicted to be occupied by the same protein subunit (i.e., the case $A_H:A_T=2$). Note, however, that this arrangement is not quasi-equivalent, as some subunits are in hexagonal and others in the triangular positions. The non-quasi-equivalent architecture of the inner capsid of Pseudomonas phage phi6 with the forbidden protein stoichiometry of 120 is an example of a gyrated $T_t(1,0)$ lattice ([Fig. 3\(d\)](#), right).

Other forms of non-quasi-equivalent architectures are higher order rhomb tilings. As Archimedean lattices have a single type of vertex environment, their duals, the Laves lattices, are formed from a single type of tile. Revisiting the generalized principle of quasi-equivalence reveals two distinct ways in which it can be realised in nature: (1) with the minimal number of protein subunits per tile (i.e., two, three, five or six proteins for rhomb, triangle, pentagonal or hexagonal tiles, respectively); or (2) with multiple identical copies of that subunit organized in a configuration that respects the symmetry of that tile. The former recovers the rhomb and kite tilings in VTT (cf. [Fig. 2\(c\)](#) and [\(d\)](#)), where each rhomb is occupied by two, and every kite by three, subunits. However, the second option presents a different way in which identical protein subunits can be arranged in a non-quasi-equivalent capsid geometry. Examples are Picobirnavirus and Zika virus ([Twarock and Luque, 2019a](#)), for which the area of each rhomb in the capsid surface is occupied by two or respectively three, dimers. These are the smallest geometries in two series of capsid architectures with $120T(h,k)$ and $180T(h,k)$ CPs, respectively. The example of Picobirnavirus is another type of pseudo $T = 2$ architecture with a CP number of 120, that is “forbidden” in the CK framework. Moreover, giant phages also present a hierarchical organization according to principle (2), but based on the gyrated lattices in [Fig. 3\(d\)](#) (left). The triangular faces (trisymmetrons) are much larger than the pentagonal ones (pentasymmetrons), and each accommodates multiple hexamers.

Implications for Viral Evolution

The importance of capsid architecture for understanding viral evolution has been pointed out by Bamford and Stuart in their visionary approach to phylogeny. They are grouping viruses into lineages based on homology of the capsid protein core folds, often combining different viral families that lack any significant sequence similarity, which is the usual measure of evolutionary relatedness. The observation in ([Twarock and Luque, 2019a](#)) that viruses with similar capsid protein folds follow the same types of capsid architectures, as exemplified for viruses using the HK97 fold, suggests that capsid geometry is indeed a good indicator for evolutionary relatedness. It also suggests that the limited spectrum of different types of capsid protein folds might be a consequence of the limited number of geometrically possible surface lattice organizations, pointing towards geometric constraints acting as drivers of convergent evolution.

A scan for a match of viral structures on the PDB-data bank with the different possible lattice types reveals that the overwhelming majority follow the hexagonal and trihexagonal lattice organizations and their duals. This could, of course, be because viruses following the other lattice types have not yet been discovered. More likely, however, this is due to deeper geometric reasons. For example, distinct lattice types imply different numbers of interfaces between capsomers, which impacts on capsid stability, as well as assembly and disassembly. Moreover, there could be a profound link with viral evolution at evolutionary timescales. It is not clear how viruses may have bridged the size gap from smaller to larger and more complex capsid structures during early stages of evolutionary history, as such transitions between different capsid structures in the CK series would correspond to significant evolutionary steps. The overarching classification scheme in ([Twarock and Luque, 2019a](#)) suggests that capsid proteins may have evolved additional domains occupying the triangular positions of a trihexagonal lattice in order to bridge those gaps. It is conceivable that this may have occurred by CPs evolving additional domains that become larger, using the geometric degree of freedom of expanding the triangular faces with respect to the pentagonal face via gyrated lattice architectures. Ultimately, or immediately, such additional domains may become independent capsid protein units in their own right. Remnants of such a mechanism are still seen, for example in the pseudo $T=3$ architectures in the family of Picornaviridae, where the polyprotein is cleaved only at a later stage in the lifecycle; or in the pseudo $T=2$ architectures exhibiting the gyrated lattice type. Note that from a mathematical point of view a gyrated version of the lattice only exists for the trihexagonal lattice. This might explain why this

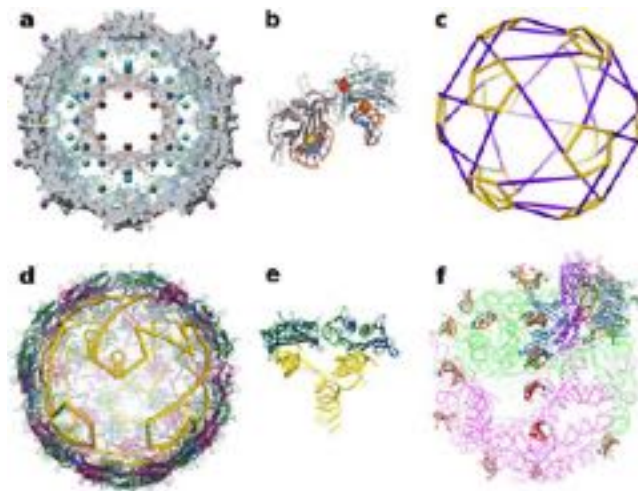


Fig. 4 *Geometric constraints on genome organization in Hamiltonian Paths Analysis.* (a) A geometric constraint set derived via affine extension of the icosahedral symmetry group, encoded by a nested point array, is shown superimposed on a cross-sectional view of a cryo-EM reconstruction of bacteriophage MS2. (b) Connecting points corresponding to the binding sites of genomic RNA at neighboring capsid proteins (yellow and orange dots) results in the polyhedral shell in (c). (d) Its vertices mark potential positions of stem-loops (packaging signals, PSs) at the inner capsid shell. Connecting vertices along polyhedral edges if the corresponding PSs occur consecutively in the linear genomic sequence from the 5' to the 3' end results in a path; if the maximal number of sites are occupied, this path is called a Hamiltonian path (HP). (e) HPs are abstract book-keeping devices that do not show the exact position of the genome, which can extend into the capsid interior as shown in (e). (f) HPA has correctly predicted both the organization of the packaged MS2 genome into two hemispheres (pink and green) and the 15 PSs identified via cryo-EM.

lattice type is preferred over other Archimedean lattices, and is therefore seen as the dominant alternative to the hexagonal lattice architectures in CKT.

Virus Structure in 3D – The Hamiltonian Paths Approach

The planar surface lattices discussed above provide models for capsid organization, but ignore structural information at different radial levels. Such detail is important, however, as the structure of the inner capsid surface impacts on the conformation of the packaged genome. A mathematical approach has been developed to model 3D capsid architecture via affine extensions of the icosahedral symmetry group, resulting in a library of nested point arrays that each provides constraints on virus architecture at different radial levels (Keef, 2013). Using the outermost features of a viral capsid of interest, a best-fit point array is selected from this library. This is exemplified for bacteriophage MS2 in Fig. 4(a), where the array is shown superimposed on a cross-sectional view of the capsid and its packaged genome obtained via cryo-electron microscopy.

Note that array points map around material boundaries of the virus, and that arrays also contain points at the multiple dispersed contacts between the genomic RNA and CP, as shown as yellow and orange spheres in Fig. 4(b). Connecting such sites if they are associated with proximal CPs results in a polyhedron (Fig. 4(c)), that is a mathematical book-keeping device for the possible connections between potential genome binding sites at the inner capsid surface. However, in any given particle, each vertex can be occupied at most once. Connecting vertices along a polyhedral edge if the corresponding contact sites are consecutive in the linear genomic sequence results in a path on the polyhedron. If all vertices are occupied, this is known as a Hamiltonian path; an example of such a path is shown in yellow in Fig. 4(d). Note that if only a subset of the vertices are occupied, then the constraints are encoded by fragments of a Hamiltonian path. Such constraints on genome organization can be used, in combination with bioinformatics, to interrogate viral genomes for potential binding sites with capsid protein. This is known as the Hamiltonian Paths Approach (HPA).

Applications of the Hamiltonian Paths Approach

An Assembly Code Embedded Within the Viral Genetic Message

HPA has been instrumental in the discovery of virus assembly instructions – a “virus assembly” code – embedded within the genetic message of viral pathogens, because it provides a way of interrogating viral genomic RNAs for secondary structure elements (such as the two connected stem-loops in Fig. 4(e)) in contact with the inner capsid surface (Prevelige, 2016). Given an ensemble of candidate contact sites, typically containing more than the maximal number of possible binding sites, the Hamiltonian path provides geometric constraints that allow the identification of geometrically compatible subsets, thus narrowing down the options (Dykeman *et al.*, 2013). HPA has revealed multiple dispersed sites within the genomic RNA of bacteriophage MS2 with a

consensus sequence/structure motif recognizing CP, that we termed packaging signals (PSs) due to their roles in genome packaging and capsid assembly. Further analysis strategies have since been developed in an interdisciplinary context to identify these virus assembly instructions in a number of viruses, including major human pathogens (Twarock and Stockley, 2019b).

Note that Hamiltonian paths do not only identify PSs, but also make predictions regarding the occupancy of binding sites at the inner capsid surface. For example, the MS2 genome was predicted to be organized into two separate hemispheres (Fig. 4(f)), with one harboring more of the higher affinity sites, as was subsequently confirmed experimentally (Twarock, 2018a).

Virus Assembly Mechanisms

Hamiltonian paths also provide a means of describing the formation of viruses as a traveling salesman problem. Teaming up the geometric understanding of Hamiltonian paths with stochastic simulations of capsid assembly based on Gillespie-type algorithms has revealed fundamental insights into the mechanism of PS-mediated assembly (Twarock *et al.*, 2018b). In particular, using Hamiltonian paths as book-keeping devices for the geometries of the assembly intermediates along virus assembly pathways, it has been possible to demonstrate how PS-mediated assembly enables efficient and selective packaging of the viral genome: The cooperative action of multiple dispersed PSs, distributed along the viral genome according to a specific affinity profile, is key for selecting specific assembly pathways amongst the vast number of geometrically possible ones, and for outcompeting cellular competitor RNA in packaging (Dykeman *et al.*, 2014). PS-mediated assembly is thus essentially a solution to a viral equivalent of Levinthal's Paradox in protein folding – the conundrum of how proteins achieve their biologically functional native state swiftly via specific folding pathways, instead of a random exploration of all combinatorially possible ones. The modeling approach also demonstrates that the effects of PS-mediated assembly can only be observed at the low protein concentrations typical of a viral infection *in vivo*, but would be obscured in *in vitro* experiments in which the full CP aliquot is added at the start of the assembly reaction, as is typically the case. This need for experimental probing of the mechanism at low CP concentration, taken together with the strong variation of PSs around a sparse consensus motif, perhaps explains why the mechanism is so difficult to detect by experiment alone, and illustrates the important role of mathematical modeling in virology.

Beyond Icosahedral Symmetry

Even though the majority of viruses exhibit icosahedral symmetry, there are many other forms of structural organization. These include helical architectures, capsid shapes in the form of fullerene cones as in HIV, lemon shaped architectures among archaeal viruses, and roughly spherical shapes with a lipid membrane taken from the host cell as in influenza virus. Even many capsids classed as icosahedral have asymmetric capsid features that distort their icosahedral lattice geometry. Examples are portal proteins that are required for genome packaging into DNA viruses, or asymmetric components with a role in genome release such as the single copy of a maturation protein in MS2. It is likely that such components have multiple functions. For example, they may also play roles in nucleation of capsid assembly. Such symmetry breaking via asymmetric capsid components has implications for viral dynamics. Among others, it impacts on the pathways of capsid lattice rearrangements during maturation (Indelicato *et al.*, 2016), that occur in many viruses as a prerequisite for particles becoming infective. Apart from asymmetric capsid components, genome organization introduces asymmetry into the overall virion architecture. As capsid symmetry impacts on genome organization in viruses assembling via PS-mediated assembly, it imposes constraints that result in a certain degree of order in genome organization. As HPA demonstrates, mathematical models of capsid symmetry are therefore avenues into the characterization of such asymmetric features and their functional roles.

Conclusion

Implications of viral geometry, symmetry and symmetry breaking percolate through all aspects of virology, and are key to understanding how viruses form, evolve and infect their hosts. Geometric insights, twinned with experiment, can also act as drivers of discovery, as demonstrated by the role of HPA in the discovery of PS-mediated assembly. Modeling of viral geometry therefore has a central role in virology, revealing mechanistic insights that can be exploited in novel forms of antiviral intervention.

References

- Caspar, D., Klug, A., 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposia on Quantitative Biology* 27, 1–24.
- Crick, F., Watson, J., 1956. Structure of small viruses. *Nature* 177, 473–475.
- Dykeman, E.C., Stockley, P.G., Twarock, R., 2013. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *Journal of Molecular Biology* 425, 3235–3249.
- Dykeman, E.C., Stockley, P.G., Twarock, R., 2014. Solving a Levinthal's paradox for virus assembly suggests a novel anti-viral therapy. *Proceedings of the National Academy of Sciences of the United States of America* 111, 5361–5366.
- Ho, V.P.T., Montiel-Garcia, D.J., Wong, J.J., *et al.*, 2018. VIPERdb: A tool for virus research. *Annual Review of Virology* 5, 477–488.
- Indelicato, G., Wahome, N., Ringler, P., *et al.*, 2016. Principles governing the self-assembly of coiled-coil protein nanoparticles. *Biophysical Journal* 110, 646–660.

- Keef, T., Wardman, J.P., Ranson, N.A., Stockley, P.G., Twarock, R., 2013. Structural constraints on the three-dimensional geometry of simple viruses: Case studies of a new predictive tool. *Acta Crystallographica Section A* 69, 140–150.
- Prevelige, P., 2016. Follow the yellow brick road: A paradigm shift in virus assembly. *Journal of Molecular Biology* 428, 416–418.
- Twarock, R., 2004. A tiling approach to virus capsid assembly explaining a structural puzzle in virology. *Journal of Theoretical Biology* 226, 477–482.
- Twarock, R., Leonov, G., Stockley, P.G., 2018a. Hamiltonian path analysis of viral genomes. *Nature Communications* 9, 2021.
- Twarock, R., Bingham, R.J., Dykeman, E.C., Stockley, P.G., 2018b. A modelling paradigm for RNA virus assembly. *Current Opinion in Virology* 31, 74–81.
- Twarock, R., Luque, A., 2019a. Structural puzzles in virology solved with an overarching icosahedral design principle. *Nature Communications* 10, 4414.
- Twarock, R., Stockley, P.G., 2019b. RNA-mediated virus assembly: Mechanisms and consequences for viral evolution and therapy. *Annual Review of Biophysics* 48, 495–514.

Further Reading

- Twarock, R., Keef, T., 2010. Viruses and geometry: Where symmetry meets function. *Microbiology Today* 37, 24–27.
- Goetschius, D.J., Parrish, C.R., Hafenstein, S., 2019. Asymmetry in icosahedral viruses. *Current Opinion in Virology* 36, 67–73.

Relevant Websites

<http://viperdbscripps.edu>
VIPERdb.

Principles of Virus Structure

Madhumati Sevvana, Thomas Klose, and Michael G Rossmann[†], Purdue University, West Lafayette, IN, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Ångstrom A unit of length equal to one hundred-millionth of a centimeter to measure wavelengths and distances (1 Å = 0.1 nm).

Capsid The protein shell of a virus particle.

Capsomere The basic structural unit of a capsid which can self-assemble to form the complete capsid.

Envelope A host derived lipid membrane separating the genome from the viral capsid.

MCP Major capsid protein.

Nucleocapsid The protein capsid and the genome.

Spike Glycoprotein surface projections of varying length at regular intervals on a viral surface.

Structural proteins Proteins present in the virus particle.

Virion A virus particle capable of infecting a host cell.

VLP Virus like particle (VLP) consisting of the outer capsid without the genome.

Introduction

Viruses are nanomolecular assemblies tailored to parasitize a wide variety of hosts ranging from prokaryotes to eukaryotes. A virus particle consists of a genome, which can comprise either single- or double-stranded RNA or DNA molecule(s) protected by a protein or a proteolipid outer capsid. Viruses insert their genome into a host organism where they can replicate and assemble. They therefore have a strong influence on protein adaptation and are key drivers of molecular evolution (Enard *et al.*, 2016). Viruses occur in distinct shapes, sizes and symmetries designed for effective transmission of their genetic material. They are classified based on the type and organization of the encapsulated genome and the presence of a lipid envelope derived from the host cell surrounding the viral nucleocapsid.

Viruses occur in three predominant shapes: filaments, spheres and pleomorphic (Fig. 1). The spheres are 20-sided regular icosahedra. The size of icosahedral viruses varies from a diameter of 170 Å (porcine circovirus) to 5000 Å (mimivirus). However, pithovirus and pandoravirus, the largest known viruses, are oval and measure about 10,000 Å and 15,000 Å in size, respectively. Mimivirus, a double-stranded DNA (dsDNA) virus with a genome size of 1.2 megabase pairs has an icosahedral shape with a diameter of 5000 Å and is associated with 1200 Å-long fibers (Klose *et al.*, 2010; Raoult *et al.*, 2004; Xiao *et al.*, 2005,2009). Most plant viruses are small and are either filamentous or polygonal, as are also many bacterial viruses (Johnson and Speir, 1997; Rossmann, 2013; Solovyev and Makarov, 2016). The larger and more complex dsDNA containing bacteriophages, which infect *E. coli*, combine both filamentous and polygonal shapes (Ackermann, 1999). For example, the T4 bacteriophage is composed of an icosahedral head (1150 Å long, 210 Å in diameter) (Fokine *et al.*, 2004) with an associated 1000 Å rod-shaped long tail (Kostyuchenko *et al.*, 2005), a 460 Å-diameter baseplate and six 1450 Å-long tail fibers (Kostyuchenko *et al.*, 2003). Structures such as these are unique to bacteriophages. Animal viruses, however exhibit extreme variations in size and shape. For example, parvoviruses are icosahedrons and measure about 180–250 Å in diameter. In contrast, members of *Poxviridae* and *Togaviridae* are about 250–750 Å in diameter and are either polygons, filamentous or pleomorphic.

The viral genome encodes structural proteins, which constitute the virus particle and non-structural proteins essential for the replication of the virus inside the host cell. These proteins aid in several stages of a typical virus life cycle that includes host cell recognition, replication, assembly, maturation and the release of progeny virus particles. The capsid protects the viral genome against harsh conditions like variation in pH, temperature and degradation by enzymes during infection. The capsid also plays a role in host cell recognition and entry. Although the number and nature of the capsid building blocks vary among viruses, there are certain common principles of capsid organization that apply to most viruses.

Several excellent reviews on the principles of virus structure organization have been published over the years (Caspar and Klug, 1962; Harrison, 2007; Johnson and Speir, 1997; Klug and Caspar, 1960; Lee and Johnson, 2003; Prasad and Schmid, 2012; Rossmann, 2013; Rossmann and Johnson, 1989). Here, we give a short history of structural virology, a discussion of virus symmetry and the methods used to study virus structures followed by the current understanding of the principles of virus architecture. We will be discussing a series of example virus structures to show: (1) how virus capsid subunits are assembled, (2) what triggers their disassembly to release their genome, (3) why certain protein folds occur frequently in virus capsids, (4) how viruses recognize and enter host cells, and (5) how structures aid in the development of vaccines and anti-viral therapeutics.

General Principles of Capsid Architecture

The first significant development in structural virology was in 1935, when Stanley crystallized the infectious particles of tobacco mosaic virus (TMV) (Stanley, 1935) and subsequently Bawden and Pirie crystallized tomato bushy stunt virus (TBSV)

[†]Deceased.

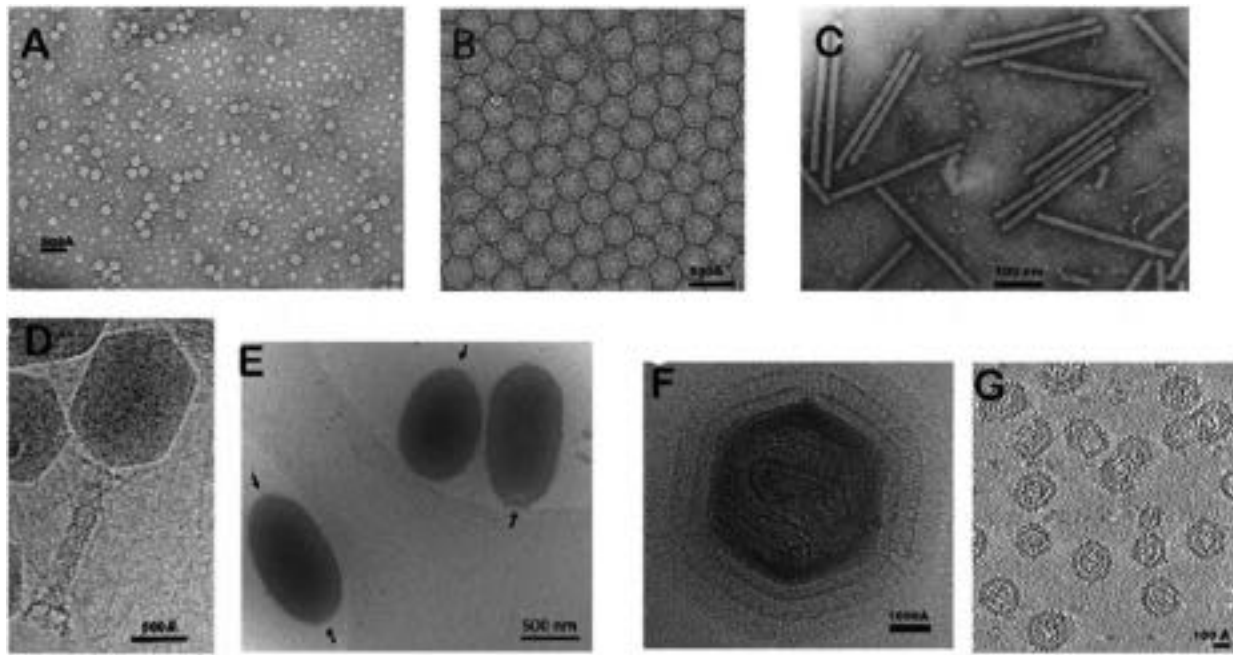


Fig. 1 Morphology of viruses. Micrographs showing icosahedral capsids of parvovirus (A), enterovirus 71 (B), the bacteriophage T4 head (D) and mimivirus (F). Micrographs of the filamentous structures in TMV (C) and the bacteriophage T4 tail (D). Micrographs of the oval-shaped capsids of pithoviruses are shown in (E) and of the pleomorphic capsids of rubella in (G). The scale bars in each figure are shown with their respective units. The micrographs were adapted from: (A) Hafenstein, S., Palermo, L.M., Kostyuchenko, V.A., *et al.*, 2007. Asymmetric binding of transferrin receptor to parvovirus capsids. *Proceedings of the National Academy of Sciences of the United States of America* 104, 6585–6589. (B) Plevka, P., Lim, P.Y., Perera, R., *et al.*, 2014. Neutralizing antibodies can initiate genome release from human enterovirus 71. *Proceedings of the National Academy of Sciences of the United States of America* 111, 2134–2139. (C) <http://www.dpvweb.net/notes/showem.php?genus=Tobamovirus>. (D) Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., Leiman, P.G., 2004. The bacteriophage T4 DNA injection machine. *Current Opinion in Structural Biology* 14, 171–180. (E) Okamoto, K., Miyazaki, N., Song, C.H., *et al.*, 2017. Structural variability and complexity of the giant Pithovirus sibericum particle revealed by high-voltage electron cryo-tomography and energy-filtered electron cryo-microscopy. *Scientific Reports* 7. (F) Xiao, C., Kuznetsov, Y.G., Sun, S., *et al.*, 2009. Structural studies of the giant Mimivirus, *PLOS Biology* 7, e1000092. (G) Mangala Prasad, V., Klose, T., Rossmann, M.G., 2017a. Assembly, maturation and three-dimensional helical structure of the teratogenic rubella virus. *PLOS Pathogens* 13, e1006377.

(Bawden and Pirie, 1938). In 1941 Bernal and Fankuchen used X-ray diffraction to establish the size and shape of TMV and TBSV particles and presented the first detailed description of any virus structure (Bernal and Fankuchen, 1941a,b). They established that the TMV virus particles consisted of a protein capsid that protected the viral genome. This work was then extended to near-atomic resolution by Franklin (Franklin, Holmes, 1958; Franklin *et al.*, 1957), Holmes (Holmes *et al.*, 1975), Stubbs (Stubbs *et al.*, 1977), and Bloomer (Bloomer *et al.*, 1978; Champness *et al.*, 1976). The description of the assembly of TMV capsid subunits around the RNA genome was among the foremost studies of virus capsid self-assembly (Butler and Klug, 1972).

In 1953, James Watson and Francis Crick reached their ground-breaking conclusion, using Franklin's X-ray diffraction images, that DNA exists in the form of a three-dimensional (3D) double helix with the two strands connected by hydrogen bonds between the nucleotides (Watson and Crick, 1953b). They also postulated the 'Central Dogma', which implied that the information flow in biological systems is from DNA to RNA to protein (Watson and Crick, 1953a). They observed that a much larger genome than observed in all known viruses would be required to code for a protein of sufficient size to completely cover the virus genome. Thus, the capsid must be constructed of multiple copies of the same protein. As each of these subunits have the same sequence and therefore the same structure, the environment of each subunit must be the same. Hence, they suggested that the capsid structure of viruses have to be regular polyhedrons. Of the five possible polyhedra, the icosahedron would have the greatest number of subunits and would, therefore, represent the most conservative way of using the least amount of genome to code for the largest possible virus (Crick and Watson, 1956; Crick and Watson, 1957). This was verified by Caspar (Caspar, 1956), who observed in an X-ray crystallographic study of TBSV that there appeared to be rotation axes at angles to each other consistent with icosahedral symmetry. The icosahedral symmetry in TBSV was also confirmed using electron microscopy (Home *et al.*, 1959; Williams and Smith, 1958).

Icosahedral Symmetry

Caspar and Klug provided the first evidence for the presence of icosahedral symmetry by analyzing the X-ray diffraction patterns of TBSV and turnip yellow mosaic virus (TYMV). An icosahedron has 12 vertices with 5-fold rotational symmetry, 20 triangular faces

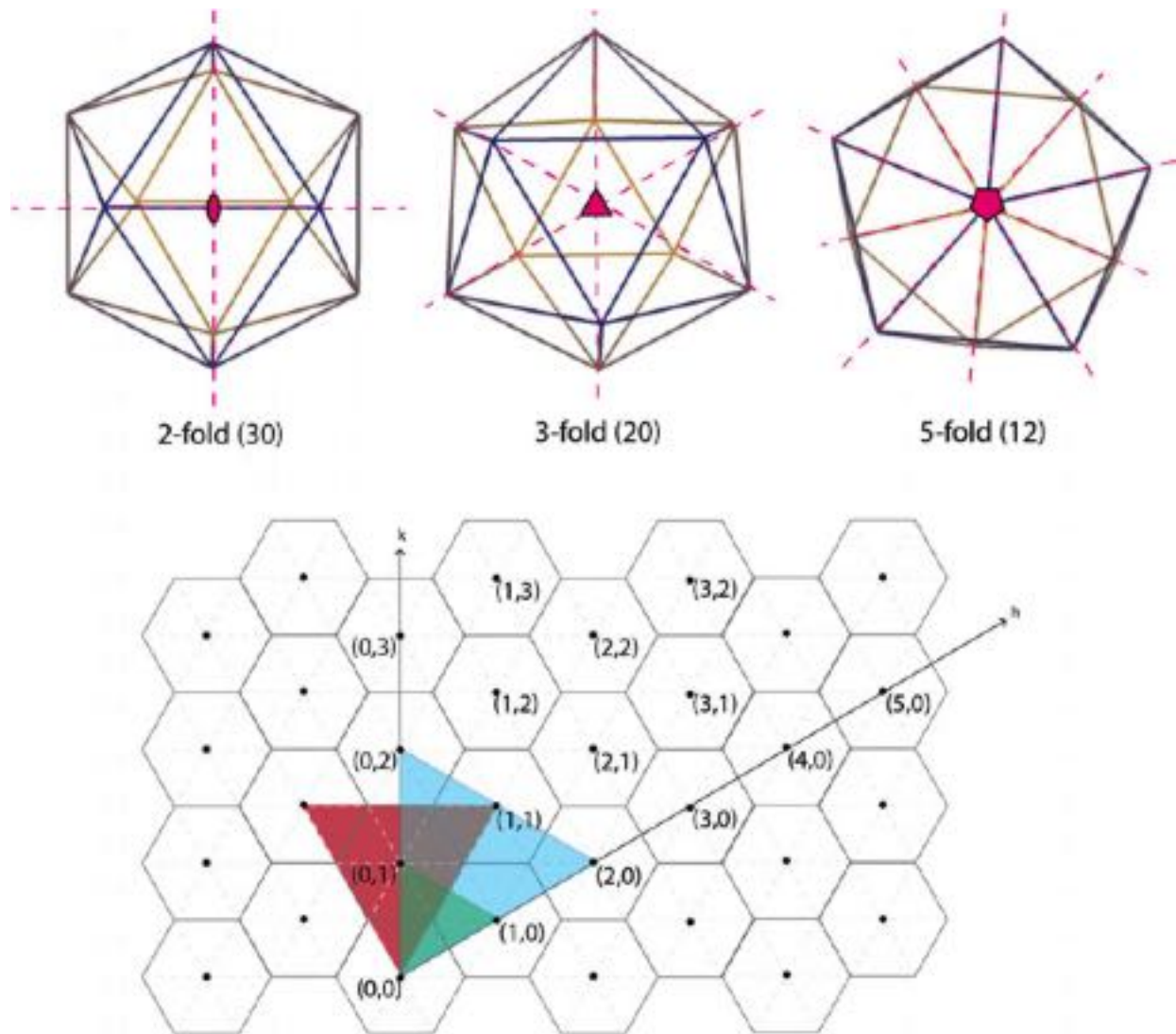


Fig. 2 *Icosahedral symmetry and triangulation.* An icosahedron displayed along the 2-fold (A), 3-fold (B) and 5-fold (C) axes of symmetry. An icosahedron has twelve 5-fold axes of symmetry along each of the vertices, twenty 3-fold axes of symmetry passing through the center of triangular faces and thirty 2-fold axes through the middle of each edge. (D) The hexagonal lattice system to define triangulation (T number) of an icosahedron. The hexagonal lattice is constructed with h and k axes crossing at 60° angles. Each hexagon can be decomposed into six equilateral triangles. The coordinates on the hexagonal lattice system (h, k) are labeled. Examples of three triangles with T numbers 1, 3 and 4 are shown in green, red, and blue respectively.

with 3-fold rotational symmetry and 30 edges with 2-fold rotational symmetry (Fig. 2). This allows the placement of 60 identical subunits decorated with capsid proteins with equivalent contacts between them. However, there are many spherical viruses where the number of capsid subunits far exceed the 60 subunits required by icosahedral symmetry. To address this question, Caspar and Klug proposed an extension to Watson and Crick's strict icosahedral symmetry concept in viruses (Caspar and Klug, 1962). They suggested that capsid subunits in an icosahedron might adapt to slightly different inter-subunit interfaces and introduced the principle of quasi-equivalence for larger virus structures with more than 60 capsid subunits. They predicted a virus assembly process in which the subunits assembled into hexagonal arrays. This would give rise to a scheme where an icosahedral virus can be constructed from pentamers of triangular faces (pentons) forming the 12 vertices and hexamers of triangular faces (hexons) covering the 20 icosahedral faces. The assumption that the monomers that make a hexamer would have quasi-equivalent environments increased the number of subunits in an icosahedron from 60 to $60T$, with T representing the triangulation number.

The triangulation number (T) is the number of quasi-similar subunits per icosahedral asymmetric unit with possible values of T , because of geometric restrictions, dictated by the equation $T = h^2 + hk + k^2$. Here, h and k are the number of hexons between pentons along the h and k axes of a hexagonal array, where h and k are either zero or positive integers. To further elaborate, the formation of an icosahedron can be visualized using a uniform hexagonal lattice, where each hexagon can be decomposed into six equilateral triangles. The hexagonal lattice is constructed with h and k axes crossing at 60° angles. A pentagon can be constructed in 3D by removing one triangle from the hexagon (Figs. 2 and 3). By arbitrarily choosing one of the 5-fold vertices as the origin ($h, k = 0, 0$),

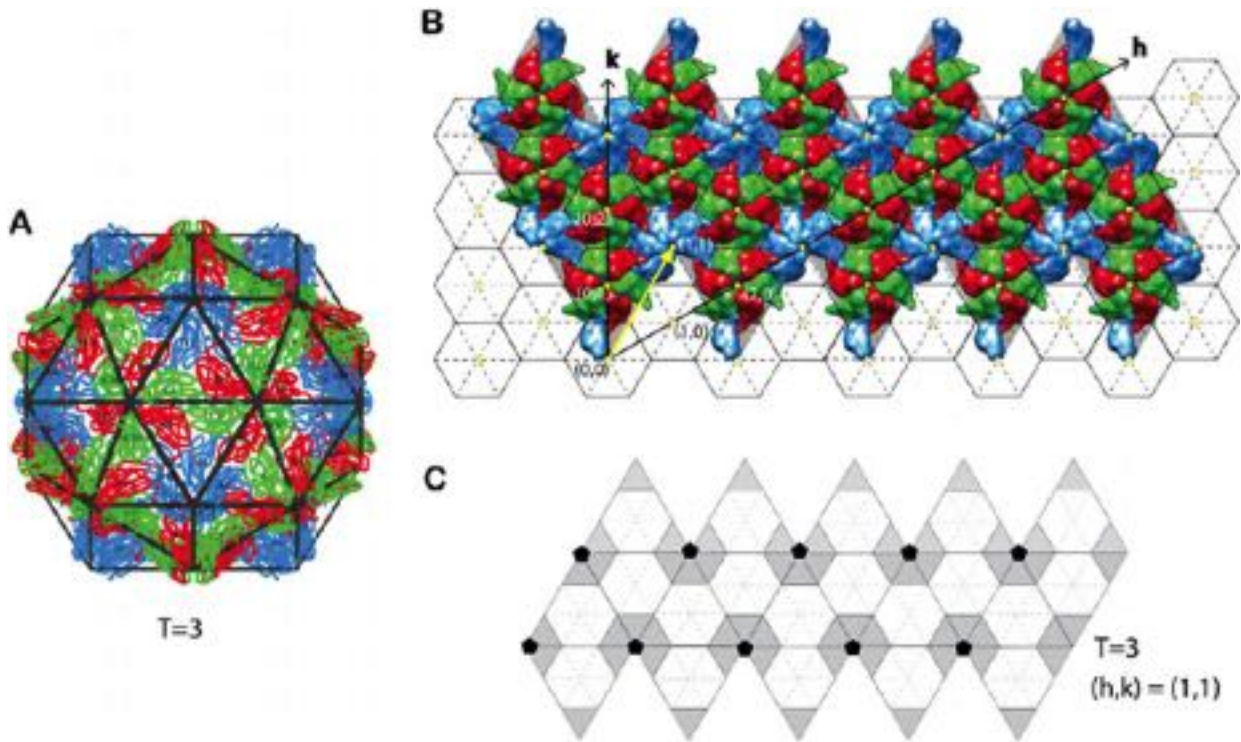


Fig. 3 The construction of an icosahedron from arrays of hexagons. $T = 3$ structure of turnip yellow mosaic virus, where (A) shows the icosahedral capsid assembly in TYMV (PDB ID: 2FZ2) and (B) the planar sheet of twenty equivalent triangles superposed on the hexagonal lattice. The $T = 3$ arrangement with $h = 1$ and $k = 1$ is indicated by a yellow arrow. (C) Schematic showing a planar sheet of 20 triangles which can then be folded up to form a closed icosahedron. Figures (A) and (B) were prepared using the Viper data base.

the positions h and k of the neighboring 5-fold vertex closest to the origin gives the T -number of that icosahedron. For example, $T = 1$ represents an icosahedron with the closest 5-fold vertex positioned at either $h = 1$ and $k = 0$ or at $h = 0$ and $k = 1$. $T = 3$, however, represents an icosahedron with the closest 5-fold vertex positioned at $h = 1$ and $k = 1$ (Fig. 3). Thus, a simple icosahedron with 60 identical subunits would have $T = 1$ symmetry. The structures of tobacco bushy stunt virus (TBSV) (Harrison *et al.*, 1978) and southern bean mosaic virus (SBMV) (Abad-Zapatero *et al.*, 1980), the first near-atomic resolution structures of icosahedral viruses, had $T = 3$ quasi-symmetries consistent with the predictions of Caspar and Klug (subsection “Non-enveloped Icosahedral RNA Viruses”). Although the amino acid sequence of the capsid proteins was rather different, both structures were very much alike with three identical protein subunits constituting the icosahedral asymmetric unit. The first animal virus structure, that of human common cold virus strain 14 (HRV14) (Erickson *et al.*, 1983; Rossmann *et al.*, 1985), mimicked the structures of TBSV and SBMV plant viruses. However, in HRV14, the three independent subunits in the icosahedral asymmetric unit were different proteins, referred to as viral proteins VP1, VP2, and VP3 (subsection “Non-enveloped Icosahedral RNA Viruses”).

Based on the equation $T = h^2 + hk + k^2$, only certain T -numbers are allowed. Therefore, the icosahedra can be further divided into three classes (Caspar and Klug, 1962; Prasad and Schmid, 2012). These classes are calculated using the formula $T = Pf^2$, where f is the largest common divisor between h and k . The first class, $P = 1$ ($T = 1, 4, 9, 16, 25, 36$ etc.), where h is greater than or equal to 1 and $k = 0$, the lattice lines run parallel to the edges of the triangular icosahedral facet. In the second class, $P = 3$ ($T = 3, 12, 27, 48, 75, 108$ etc.), the icosahedra have T numbers along the line that bisects the h and k axes with $h = k$. The third class, called the skew class, includes icosahedra with other allowed T numbers (7, 13, 19, 21, 28, 31 etc.) with the possibility of having enantiomorphic configurations. In $T = 7$ icosahedra, the asymmetric unit is made of 7 proteins giving a total of 420 capsid proteins per virus capsid. However, depending on the values of h and k , the capsid can either have a configuration of dextro (d) for $h = 1, k = 2$ or laevo (l) for $h = 2, k = 1$. For example, polyomavirus has a non-enveloped capsid with $T = 7d$ pseudo-symmetry because it is only composed of pentamers and lacks the expected hexamers in its capsid whereas Rotavirus, a non-enveloped, icosahedral virion with a double capsid structure has a $T = 13l$ symmetry for the outer capsid and $T = 1$ for the inner capsid. Another form of icosahedron, called the prolate icosahedron where the icosahedron is stretched along one axis, is found in viruses like T-even bacteriophages (Fokine *et al.*, 2004; Rao and Black, 2010). Here, the icosahedron has a central cylindrical body made of 10 triangular faces with two caps at the bottom and the top. The prolate icosahedra are characterized by the combination of a T number and an elongation number called Q according to the formula $n = 30(T + Q)$, where T is the triangulation number of the central body, Q is any positive integer and n refers to the number of subunits. The caps follow the triangulation T , whereas the main cylindrical body follows the Q number. When $T = Q$, the icosahedra becomes isometric, when $Q > T$, the icosahedron is referred to as prolate and when $Q < T$, the icosahedra is referred to as oblate.

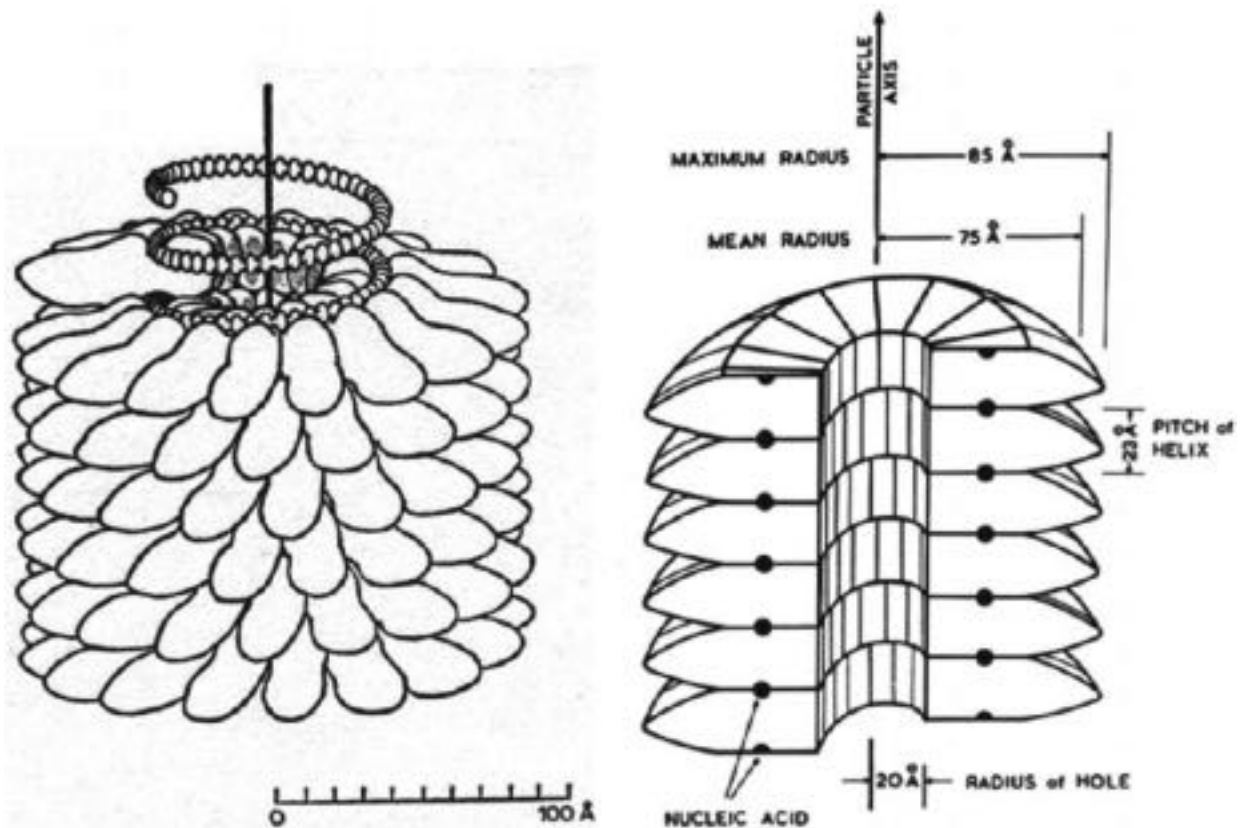


Fig. 4 Structure of tobacco mosaic virus (TMV). Figure adapted from Bernal, J.D., Fankuchen, I., 1941a. X-ray and crystallographic studies of plant virus preparations: I. Introduction and preparation of specimens. II. Modes of aggregation of the virus particles. III. (1) The structure of the particles and (2) biological implications. *Journal of General Physiology* 25, 111–165.

Helical Symmetry

Helical symmetry is another prevalent symmetry occurring in the capsids of numerous plant viruses, tails of bacteriophages, and flexible capsids of negative-stranded RNA genome containing viruses. Helical organization of capsids in plant viruses has the advantage that it imposes no limitations on the size of the packed RNA genome compared to icosahedral viruses which are limited by their internal volume. Helical symmetry combines rotation and translation, giving rise to a screw axis where the pitch of the helix (P) is defined as the axis rise. A helical virus structure is described by the location of subunits with respect to the helix axis, the number of subunits per helix turn (n) and the axial rise per subunit (h).

In TMV, one of the earliest and best characterized helical virus structures, the genomic RNA of 6400 bases is encapsidated by 2140 identical viral capsid proteins giving rise to a rod-like virion, 3000 Å long and 180 Å in diameter with a central hole of 20 Å in radius (Fig. 4). After the preliminary structural studies of TMV using X-ray fiber diffraction to 2.9 Å (Namba *et al.*, 1989a), a more recent cryoEM structure has been determined to a resolution of 3.3 Å (Ge and Zhou, 2011). In TMV the capsid protein subunits form a right-handed helix with a pitch of 23 Å and 16 and 1/3rd subunits in each turn. Each capsid subunit interacts with three nucleotides. The structure of the TMV genome follows the symmetry of the capsid providing details on capsid-RNA interactions and the assembly of the virus.

Methods for Studying Virus Structures

The principal techniques used to study intact virus structures are X-ray crystallography, X-ray fiber diffraction, cryo-electron microscopy (cryoEM) and cryo-electron tomography (cryo-ET). To date, according to the statistics on the VIPERdb database (Carrillo-Tripp *et al.*, 2009) about 404 virus structures have been determined using X-ray crystallography and 345 structures have been determined using cryoEM. Major accomplishments in understanding the structure of viruses were closely associated with methodological, technological and computational advances in X-ray crystallography and cryoEM and developments in molecular virology techniques.

X-ray crystallography has been the most successful technique for determining the structure of individual icosahedral viruses since the 1970s. The first of the structures determined using crystallography were of icosahedral plant viruses (Abad-Zapatero *et al.*, 1980;

Harrison *et al.*, 1978) followed by more complex animal and human viruses. X-ray fiber diffraction was used to study viruses with helical symmetry. A key issue in X-ray crystallography is the solution of the phase problem, which in the case of viruses is helped enormously by the presence of non-crystallographic symmetry (NCS). Several crystallographic milestones that have also assisted in solving many significant virus structures have been (1) the use of high intensity synchrotron radiation for automated collection of diffraction images (Rossmann, 1979,1984,1999; Rossmann *et al.*,1979); (2) determining the orientation of virions in the unit cell (Rossmann and Blow, 1962; Tong and Rossmann, 1990); (3) using non crystallographic symmetry based methods to solve the phase problem (Åkervall *et al.*, 1972; Rossmann and Blow, 1963) and the use of NCS averaging for improving the phases (Buehner *et al.*, 1974; Champness *et al.*, 1976). If, for instance, isomorphous replacement had been used to obtain a 3.5 Å structure, then averaging that structure will increase the quality of the phases of the current map which can then be used to extend the resolution of the map (usually about one reciprocal lattice point in all directions). The slightly improved map can be used again for averaging and so forth.

The above real space considerations can also be expressed in reciprocal space and by doing so give insight into why phasing can be very successful by gradual resolution increase at each stage by less than one reciprocal lattice point given knowledge of the NCS. This had been under considerable dispute until the structure of HRV14 in 1985 (Arnold and Rossmann, 1986; Rossmann *et al.*, 1985) when the initial 5 Å resolution isomorphous replacement phases were extended to a beautiful 3 Å resolution map using NCS (Rossmann *et al.*, 1992a,b).

Electron microscopy of negatively stained virus samples was among the first techniques used to recognize virus shapes and sizes (Brenner and Horne, 1959; Horne *et al.*, 1959). Subsequent development of computational techniques to reconstruct the 3D structure of viruses from 2D projection images (Crowther *et al.*, 1970) and sample freezing techniques (Adrian *et al.*, 1984; Dubochet *et al.*, 1988) to preserve the native structure, revolutionized intact virus structure determination using cryoEM (Haas and Rossmann, 1970). For the next three decades, many virus structures were determined to a resolution of 25 Å. However, the more recent development of direct electron detectors as well as other technological advances in cryoEM has made it easier and more common to obtain cryoEM electron potential maps of virus samples to a resolution better than 3 Å. Prior to this, pseudo-atomic models of the virus capsid were derived by exploiting cryoEM and X-ray crystallography-based hybrid techniques. This required high resolution structures of viral components, obtained by using X-ray crystallography to be fitted into the low resolution cryoEM capsid map of the whole virus (Kuhn *et al.*, 2002; Zhang *et al.*, 2002). CryoEM has the advantage that it can be used where crystal growth is limiting especially with heterogeneous samples of enveloped viruses and viruses in complex with antibodies, receptors or anti-viral compounds. For example, crystals of West Nile virus diffracted to about 16 Å resolution (Kaufmann *et al.*, 2010), whereas cryoEM gave a map to a resolution of 10 Å (Zhang *et al.*, 2013). In the case of pleomorphic viruses such as Rubella (Battisti *et al.*, 2012; Mangala Prasad *et al.*, 2013; 2017a), cryo-electron tomography (cryo-ET) has been the method of choice to study the 3D structure. Tomography uses a series of collected images, where the microscope stage has been tilted to give a series of 2D projections that can be combined to form a 3D image (Frank, 1992; Lucic *et al.*, 2005; Subramaniam *et al.*, 2007). However, the resolution limitation for this technique (in the absence of averaging) has been about 20 Å because of the need to reduce the dose of each image to be able to collect a 3D data set before the sample has been destroyed by repeated exposure to electrons.

Structural Folds of Capsid Proteins

One of the characteristic features of viral capsid proteins lies in the folded topology of the capsid monomers, allowing them to assemble into large, symmetrical and geometrically sophisticated architectures (Chapman and Liljas, 2003; Cheng and Brooks III, 2013). The canonical jelly-roll β -barrel, as a core structural motif, is the most common capsid fold occurring in virus shells of varying sizes (Abad-Zapatero *et al.*, 1980; Chelvanayagam *et al.*, 1992; Harrison *et al.*, 1978; Johnson and Chiu, 2000). Other common protein folds occurring in viruses are the Greek key β -barrel with six strands (Choi *et al.*, 1991a; Tong *et al.*, 1992,1993), the helix bundle (Bloomer *et al.*, 1978; Conway *et al.*, 1997; Gamble *et al.*, 1997), the immunoglobulin-like fold (Rey *et al.*, 1995) and the HK97 fold (Suhanovsky and Teschke, 2015). The structurally unique capsid proteins are also functionally unique in terms of their folding topology because of the evolutionary constraints imposed by protein-protein interactions and are most likely a consequence of geometric requirements of the building blocks to form cage-like assemblies. Here, we describe the commonly occurring virus capsid folds.

Jelly-Roll β -Barrel

Both TBSV and SBMV, the first determined icosahedral virus structures, were similar with respect to their capsid structures and organization. It was then obvious that the two virus capsid structures with no apparent sequence similarity must have diverged from a common ancestral fold, later termed as a 'jelly-roll' fold by Jane S. Richardson in 1981 (Richardson, 1981), reflecting its resemblance to a Swiss roll cake. The jelly-roll fold is also referred to as a wedge shaped fold, an RNA virus capsid β -barrel or a β -sandwich (Cheng and Brooks, 2013).

A jelly-roll fold consists of eight antiparallel β -strands that are named from B to I along the polypeptide sequence (Fig. 5). The two four-stranded sheets that form the opposite sides of a β -barrel consist of strands BIDG and CHEF, folded such that strand B packs opposite strand C, I opposite H, etc. The end strands almost form a closed barrel structure, however there is no hydrogen-bonding between the end strands of facing sheets. A single tight turn between strands E and F leads to the pairing of strands D and G, C and H, and B and I, which then roll up, Swiss cake style. Therefore, the BI and DG pairs combine to form the BIDG sheet, and

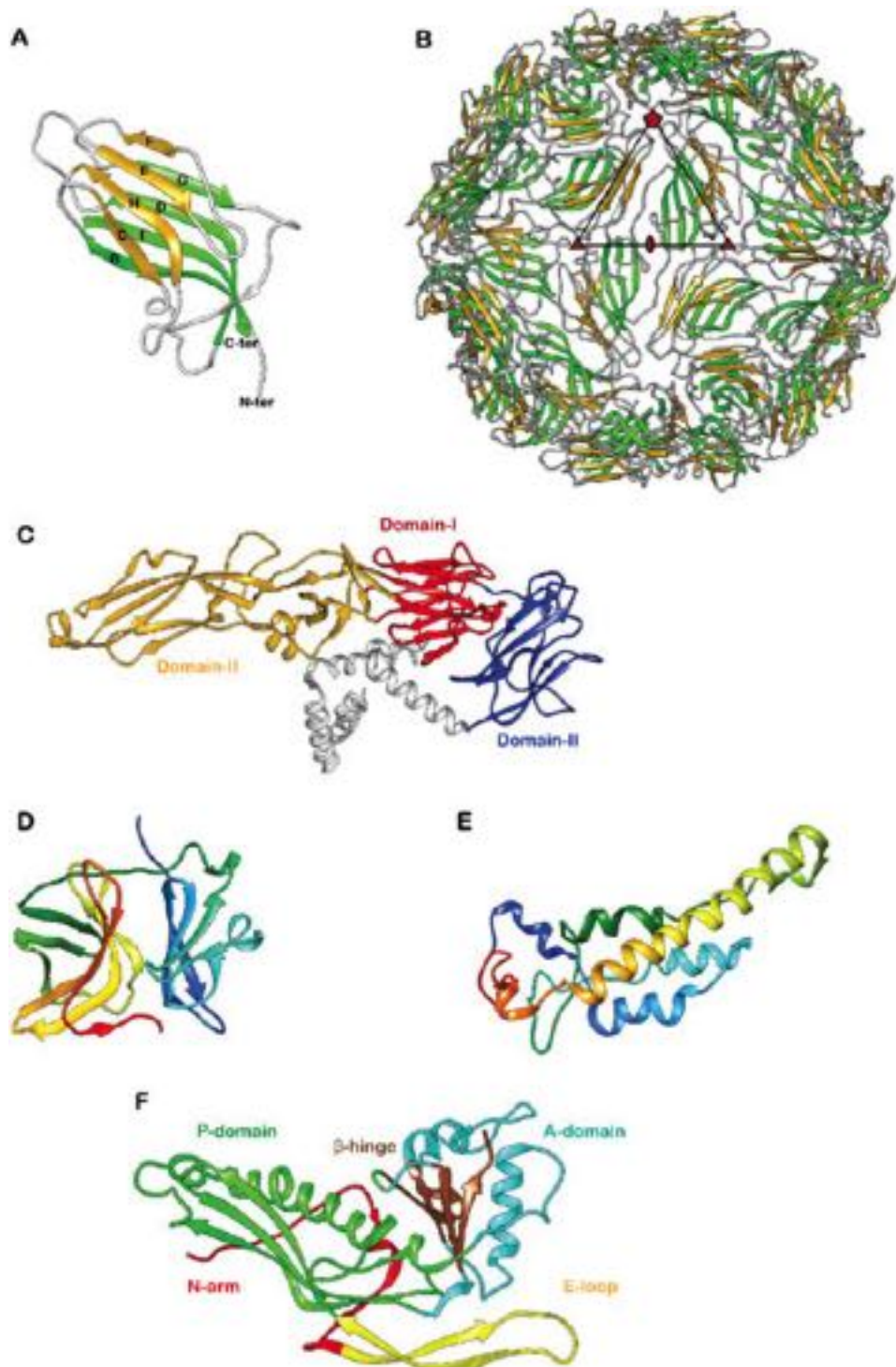


Fig. 5 *Virus capsid structure folds.* (A) and (B) show the jelly-roll β -barrel fold in satellite tobacco mosaic virus (STMV) (PDB ID: 40Q9). The β -sheets BIDG are colored green and the β -sheets CHEF are colored golden yellow. (B) The T = 1 icosahedral assembly of the subunits to form the STMV virus capsid. The icosahedral asymmetric unit is drawn as a black triangle while the 2-, 3- and 5-fold symmetry axes are drawn in red. (C) Immunoglobulin fold in the domain-III (blue) of ZIKV envelope protein E (PDB ID: 6C08). Domain I, domain II and the membrane associated helices are colored red, yellow, and gray, respectively. (D) The serine protease-like fold in the capsid of Sindbis virus (PDB ID: 1SVP) and (E) the four-helix bundle in TMV (PDB ID: 1VTM) are shown in rainbow colors (blue to red) from N- to C-terminus. (F) The HK97 coat protein in the immature state (PDB ID: 3E8K). The common secondary structural elements observed in HK97 fold are labeled.

the CH and EF pairs combine to form the CHEF sheet. However, variations to the common core might occur in different viruses. There may be one or two additional strands like the β -strand A in TBSV outside the capsid's common fold. The length of the strands might also differ depending on triangulation number and capsid architecture. The largest variations occur in the size of the loops connecting the β -strands. In some viruses, these external loops come together to form antigenic sites (Rossmann *et al.*, 1985) and in others the loops account for the increase in capsid size (McKenna *et al.*, 1992a,b). Most of the jelly-roll folds are about 180 amino acids and can be as big as 580 amino acids in parvoviruses (Tsao *et al.*, 1991).

A large number of viruses build their capsids using either single or double jelly-roll fold capsid subunits. Double jelly-roll proteins have presumably evolved from single jelly-roll proteins by gene duplication. Single jelly-roll capsid folds are found primarily in many icosahedral RNA and DNA viruses. However, the majority are positive-sense single-stranded RNA viruses and the only dsDNA virus with single jelly-roll folds are the members of *Papillomaviridae* and *Polyomaviridae*. The double jelly-roll capsid proteins are found exclusively in dsDNA viruses spanning a large capsid size range. Most members of the double jelly-roll fold group are icosahedral, a few families such as *Poxviridae* and *Ascoviridae* have oval or brick-shaped mature virions.

The Immunoglobulin-like Fold

The immunoglobulin fold is found in a vast variety of proteins of different functions such as antibodies, cellular adhesion molecules, the receptor tyrosine kinase and many other molecules. The canonical immunoglobulin fold consists of a 7-stranded antiparallel β -sheet sandwich within which there is extensive hydrogen bonding. One example in a virus is one of the domains of a flavivirus envelope glycoprotein (E) (Rey *et al.*, 1995). The flavivirus protein E consists of three domains: The N-terminal domain-I, which is an antiparallel β -barrel quite different from the jelly-roll fold, a β -sheet rich domain-II and the C-terminal domain-III with an immunoglobulin fold. In flavivirus E protein domain-III, the three- (β -sheet CFG) and four- (β -sheet ABED) member sheets face each other and are commonly linked by a disulfide bond. The domain topology resembles a Greek key (Fig. 5).

The Serine Protease Fold

The serine protease fold is observed in chymotrypsin, many other proteases and viral proteins. It is again an anti-parallel Greek key β -barrel with six-strands. Examples of viruses with this domain as core protein can be found in the capsid proteins of the members of *Togaviridae* such as Chikungunya virus (CHIKV) (Sun *et al.*, 2013b), Sindbis virus (Choi *et al.*, 1991a), Ross River virus, Rubella (Mangala Prasad *et al.*, 2013) and Eastern equine encephalitis virus (Hasan *et al.*, 2018b) (Fig. 5). In Sindbis virus, the capsid polypeptide is folded into two similar 'Greek key' β -barrel domains with the substrate site situated between the domains. The capsid structure of Sindbis virus identified Ser215, His141 and Asp163 as the essential catalytic triad as in other serine proteases (Choi *et al.*, 1991a).

The Four-helix Bundle

The four-helix bundle is a common motif in which antiparallel α -helices are packed side by side with a close packed hydrophobic core (Fig. 5). It is also found in a wide range of proteins with diverse functions like ferritin, various cytochromes, cytokines, growth hormones etc. Viral examples include the coat protein of TMV, capsid of hepatitis B and the C-terminal domain of the HIV-1 capsid (Bloomer *et al.*, 1978; Conway *et al.*, 1997; Gamble *et al.*, 1997). Each capsid subunit of TMV consists of 158 amino acids assembled into a four-helix bundle, which are joined by a loop proximal to the virion axis. Each coat protein interacts with 3 RNA nucleotides (Namba *et al.*, 1989b).

The HK97 Fold

HK97 capsid fold is termed after the first capsid protein structure solved by X-ray crystallography of bacteriophage HK97 (Helgstrand *et al.*, 2003; Suhanovsky and Teschke, 2015; Wikoff *et al.*, 2000). Most of the viral capsids using the HK97-fold are found in bacteriophages. Additionally, this fold is also found in the capsids of Herpes Simplex virus type 1, murine cytomegalovirus, pseudorabies virus and the rhesus monkey rhadinovirus. Viruses using the HK97-fold share unexpected similarities beyond the fold of their coat proteins including the architecture of their virions, an internal scaffolding protein-mediated assembly, active dsDNA packaging, and capsid maturation events (Suhanovsky and Teschke, 2015).

The HK-97 fold is characterized by having an N-arm (sometimes α -helical rich), an E-loop with variable lengths that is a two-stranded anti-parallel β -sheet (β 2 and β 3), a peripheral P-domain with a long helix called the "spine helix" and an unusually long β -sheet; and an A-domain (axial domain) with a central β -sheet referred to as the β -hinge surrounded by short helices and loops (Fig. 5).

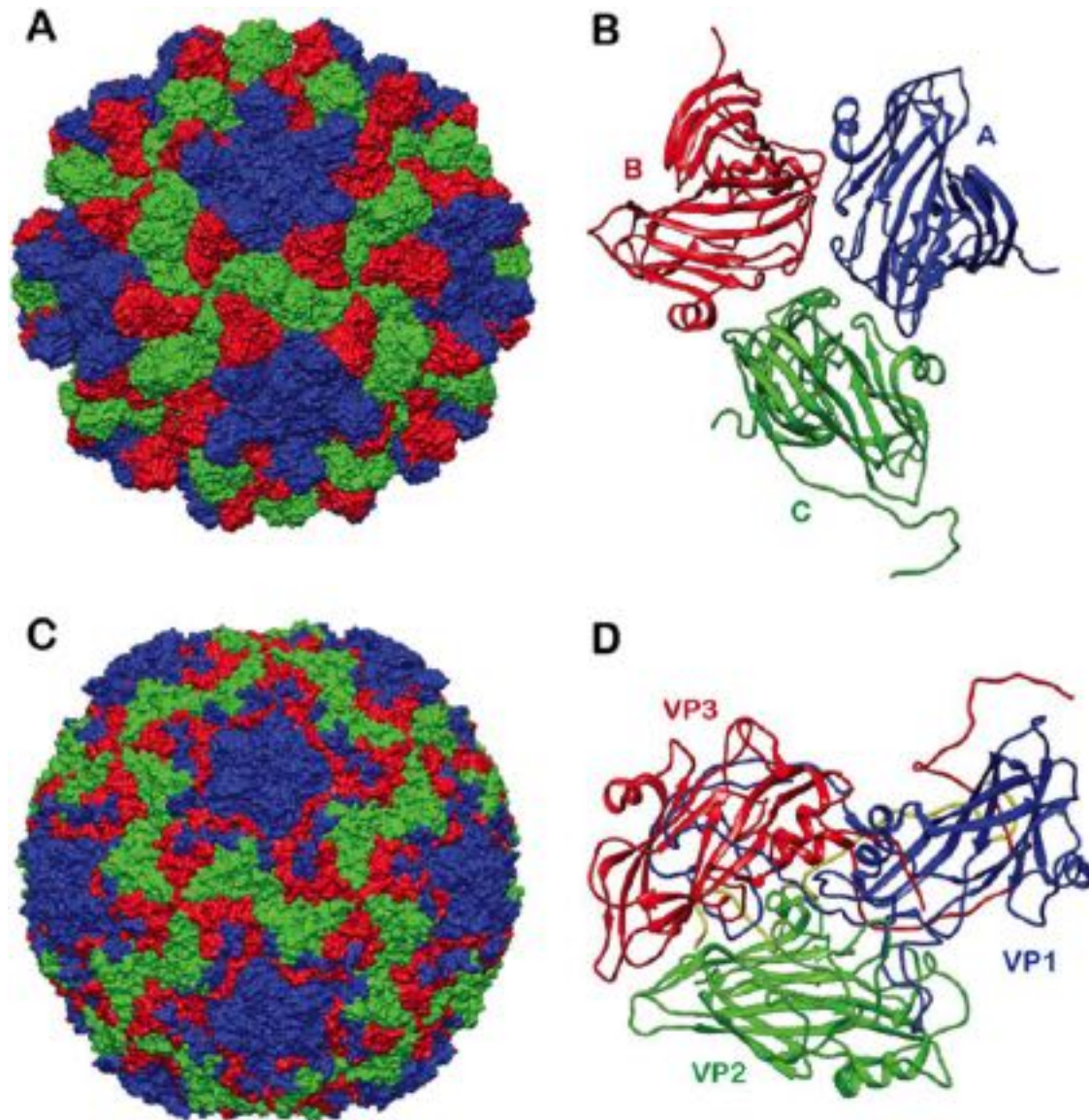


Fig. 6 $T = 3$ icosahedral capsid structure of non-enveloped RNA viruses. (A) The capsid structure of tomato bushy stunt virus (TBSV) (PDB ID: 2TBV). (C) A surface representation of human rhinovirus 14 (HRV14) (PDB ID: 4RHV). The icosahedral asymmetric unit of TBSV (B) consists of three similar subunits colored in red, blue and green, whereas in HRV14 (D) the subunits are formed by different proteins and labeled as VP1 (blue), VP2 (red), VP3 (red), and VP4 (yellow).

Capsid Assembly

Non-Enveloped Icosahedral RNA Viruses

The first icosahedral RNA plant virus structures, TBSV (Harrison *et al.*, 1978; Winkler *et al.*, 1977) and SBMV (Abad-Zapatero *et al.*, 1981), had similar capsid folds (jelly-roll fold) and organization with three independent subunits (A, B, and C) in the icosahedral asymmetric unit arranged around a quasi-3-fold axis (as predicted by Caspar and Klug). These structures clearly established that virus capsids with more than 60 subunits were organized based on a triangulated icosahedral lattice as described by Caspar and Klug (Caspar and Klug, 1962). The capsid is composed of 180 capsid subunits and exhibited a $T = 3$ capsid organization with rings of pentons and hexons. The assembly of these viruses occurred via AB and CC dimers. However, the AB dimer had a disordered amino-terminal β -strand compared to the one of the CC dimer, which was ordered, giving the two types of dimers a different curvature on the virus surface (Fig. 6).

The first animal virus structure was that of HRV14 (Arnold *et al.*, 1987; Erickson *et al.*, 1983), which mimicked the structures of TBSV and SBMV plant viruses. HRV14 belongs to the family of picornaviruses, which are small icosahedral animal viruses of about 300 Å diameter with a positive-sense eight thousand base RNA genome. However, while in the plant viruses the three protein

subunits in the icosahedral asymmetric unit were identical, the icosahedral asymmetric unit of HRV14 was composed of three independent subunits of the viral proteins: VP1, VP2, and VP3 (and VP4 which is internal in the mature virus) (Fig. 6). The capsid is composed of 60 copies of VP1, VP2, and VP3 (and internal VP4) arranged with pseudo $T = 3$ icosahedral symmetry. Nevertheless, each subunit had the same fold referred to as the “jelly-roll” fold. Shortly after the determination of the HRV14 structure and its extraordinary evolutionary implication, Hogle was able to determine the structure of poliovirus serotype 1, using the same phase extension technique as had been pioneered for HRV14 (Hogle *et al.*, 1985). This further confirmed the remarkable similarity of simple plant and animal RNA viruses.

The structure determination of HRV14 was not only a significant advancement in the method used for the analysis of X-ray diffraction data but was also informative of the functions required by the virus for infecting a cell and for replication. Viral escape mutations for a series of neutralizing monoclonal antibodies were all found to be on the virus surface and thus were likely the site of antibody binding (Sherry *et al.*, 1986; Sherry and Rueckert, 1985). The surface of HRV14 had a deep depression (“canyon”) around each pentameric vertex (Rossmann *et al.*, 1985). None of the MAb binding sites were in the canyon. This, therefore, suggested that the canyon might be the site where cellular receptor molecules could bind without experiencing any inhibiting viral mutations (Rossmann *et al.*, 1985; Rossmann and Palmenberg, 1988). Thus, by hiding the receptor binding site in the canyon (which would require a long thin receptor molecule), the virus could evade the host’s immune system while still being able to bind to the same receptor molecule (Rossmann, 1989). Subsequently it was established that most HRVs use intercellular adhesion molecule 1 (ICAM1) as receptor (Greve *et al.*, 1989; Staunton *et al.*, 1989). In one of the earlier applications of cryoEM, Rossmann and colleagues verified that indeed ICAM1 bound into the canyon as predicted (Olson *et al.*, 1993).

Another significant discovery related to the structure of HRV14 was that a series of anti-rhino-viral compounds bound into a hydrophobic pocket within VP1 (Smith *et al.*, 1986). This pocket was below and close to the canyon. It was shown that binding one of the antiviral compounds being developed by the Sterling Winthrop Company caused an enlargement of the pocket and a change of structure in the canyon, explaining why the binding of these compounds inhibited cellular attachment. Furthermore, it was found that this pocket was occupied by a small stabilizing lipid in HRVs and poliovirus. Thus, binding of the infectious virus to ICAM1 would displace the stabilizing “pocket factor” and prepare the virus for the release of its genome into the cytoplasm of the cell containing the ICAM1 (Filman *et al.*, 1989; Oliveira *et al.*, 1993; Rossmann, 1994; Smyth *et al.*, 1995). These results stimulated extensive work by the ViroPharma Company to create an inhibitor that was effective over a wide range of HRV serotypes, resulting in the development of “Pleconaril” (Hayden *et al.*, 2003; Xiao *et al.*, 2011). Although this compound was successful in clinical trials it put women on contraceptive hormones at risk of conception. Thus, the drug never reached the pharmaceutical market.

Enveloped Icosahedral RNA Viruses

Many major viral pathogens such as human immunodeficiency virus (HIV), hepatitis B virus (HBV), influenza viruses, herpes simplex virus (HSV), coronaviruses as well as various poxviruses, alphaviruses and flaviviruses contain host-derived lipid membranes studded with glycoproteins. Among these, members of *Hepadnaviridae* (ex: HBV) and *Herpesviridae* have an internal icosahedral capsid shell surrounded by an amorphous tegument and a lipid envelope embedded with glycoproteins. Flaviviruses and alphaviruses, however, have an icosahedral outer shell and an internal lipid membrane separating the outer glycoprotein shell from the genome (Mukhopadhyay *et al.*, 2005).

Alphaviruses have a positive-sense ssRNA genome that is about 11 kb in length and codes for 9 proteins (Kuhn, 2007). The 3' end of the genome is transcribed for subsequent translation into a polyprotein precursor containing the three structural proteins PE2 (the precursor of E3 and E2), E1, and the capsid protein. The structures of a number of alphaviruses have been determined using cryoEM to resolutions better than 10 Å, including the structure of Chikungunya virus-like particles to 5.3 Å resolution (Sun *et al.*, 2013b), Venezuelan equine encephalitis virus to 4.6 Å resolution (Zhang *et al.*, 2011a), Eastern equine encephalitis virus to 4.4 Å resolution (Hasan *et al.*, 2018b) and Sindbis virus to 3.5 Å resolution (Chen *et al.*, 2018). Alphaviruses have an external diameter of about 700 Å and are icosahedral with quasi- $T = 4$ symmetry (Fuller, 1987; von Bonsdorff and Harrison, 1975; von Bonsdorff and Harrison, 1978). They have a nucleocapsid core that is completely surrounded by a lipid envelope, derived from a host membrane, into which is embedded an icosahedral array of glycoproteins. A single virus particle contains 240 copies each of the E1 and E2 glycoproteins in the mature virus (in addition 240 copies of E3 in the immature virus) that form 20 “i3” spikes situated on the icosahedral 3-fold axes and 60 quasi-3-fold “q3” spikes at general positions (Cheng *et al.*, 1995; Zhang *et al.*, 2002). There are 240 copies of the capsid protein on the internal cytoplasmic side of the viral lipid membrane, arranged as 12 pentamers about the 5-fold vertices and 30 hexamers about the icosahedral 2-fold vertices, consistent with the $T = 4$ symmetry of the glycoprotein on the external side of the membrane. The capsid protein of alphaviruses is similar to trypsin-like proteases and participates in the assembly process (Choi *et al.*, 1991b) (Fig. 7).

The E1 glycoprotein of CHIKV forms three β -barrel domains (I, II, and III). The CHIKV E2 glycoprotein is arranged into three immunoglobulin-like domains A, B and C (Gibbons and Kielian, 2002; Li *et al.*, 2010; Voss *et al.*, 2010). Domain A contains the receptor binding site, domain B is at the distal end of each spike protecting the fusion loop on DII of E1 and domain C is situated closest to the viral membrane. The E1 glycoprotein contains a hydrophobic fusion loop that is responsible for membrane fusion with an endosomal membrane when initiating infection. The E2 protein can bind to cellular receptors and protects the E1 fusion loop at neutral pH. During the initial stages of infection, the host cellular receptor is recognized by the surface glycoprotein E2.

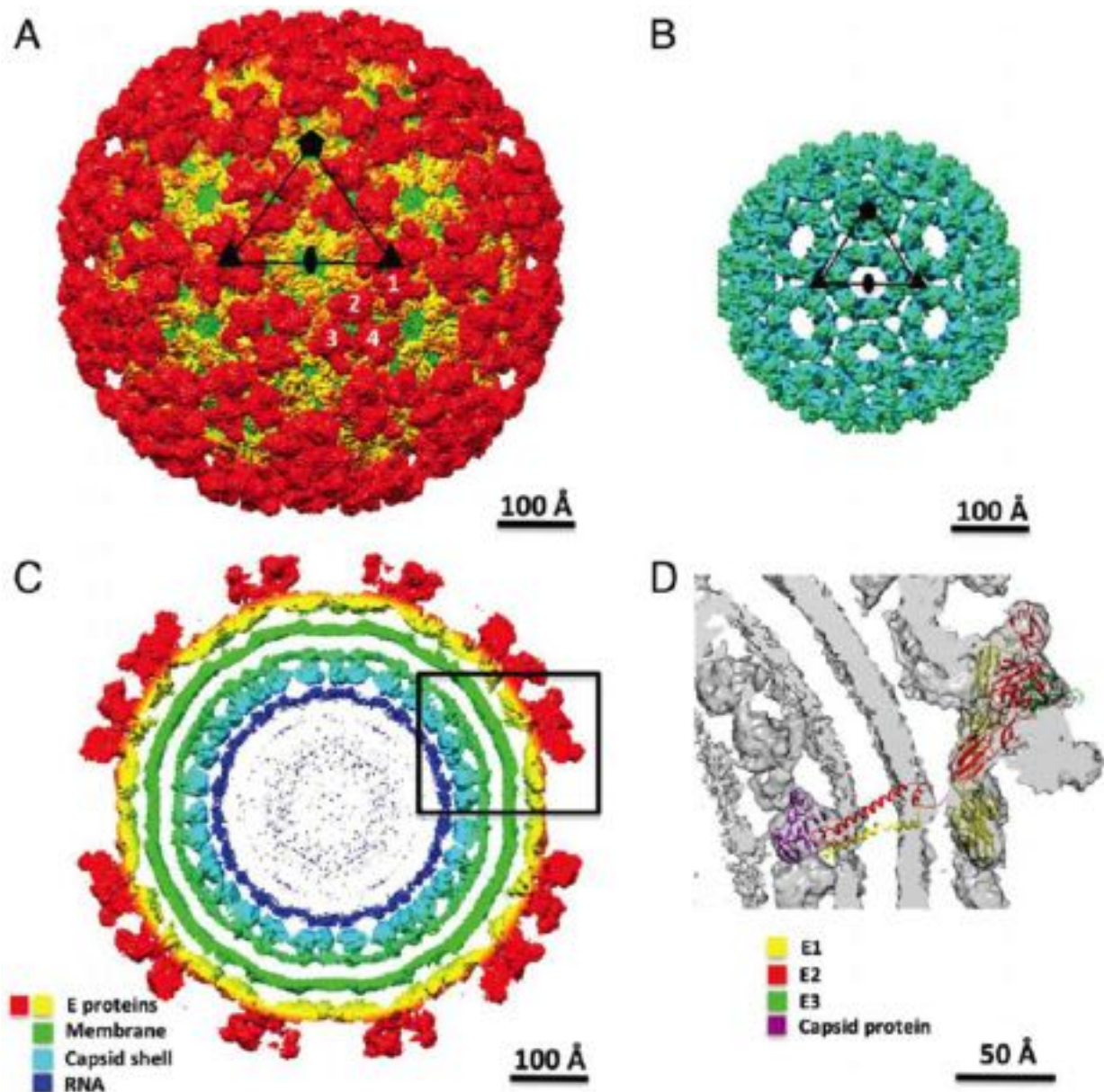


Fig. 7 $T = 4$ icosahedral capsid architecture in immature Chikungunya virus (CHIKV). (A) Electron potential map of immature CHIKV viewed down an icosahedral 2-fold axis. The icosahedral asymmetric unit is drawn as a black triangle and 2-, 3- and 5-fold symmetry elements are represented as oval, triangle and pentagon, respectively. The four subunits of icosahedral asymmetric units are labeled from 1 to 4 in white text. (B) Internal capsid shell of immature CHIKV. (C) Central cross-section of (A) with the represented components colored as labeled. (D) The fitting of E1-E2-E3 and C coordinates into the region outlined in (C). Figure adapted from Yap, M.L., Klose, T., Urakami, A., *et al.*, 2017. Structural studies of Chikungunya virus maturation. *Proceedings of the National Academy of Sciences of the United States of America* 114, 13703–13707.

Once the virus has been internalized, the low pH of the endosome causes the virion to undergo an irreversible conformational change resulting in the disassociation of E2 from E1 and the formation of E1 trimers. Upon E2 disassociation, the fusion loop then binds to and fuses with the endosomal membrane. Finally, the viral genome is released into the host cytoplasm, where replication of new viral particles can begin (Kielian and Rey, 2006; Lescar *et al.*, 2001; Lu *et al.*, 1999; Marsh and Helenius, 1989).

The presence of the lipid causes some heterogeneity, making it often difficult to crystallize these viruses (Harrison *et al.*, 1992). Instead the structure of many of these larger, lipid containing viruses have been determined by using cryoEM to provide a lower resolution structure of the virus and higher resolution X-ray crystal structures of the components. The latter could then be fitted into the EM structure of the whole virus. A particularly good early example was the determination of the mature dengue virus structure (Kuhn *et al.*, 2002). However, more recently it has been possible to determine the structure of lipid containing viruses to about 3 Å resolution with the development of direct electron counting detectors for electron microscopes and general

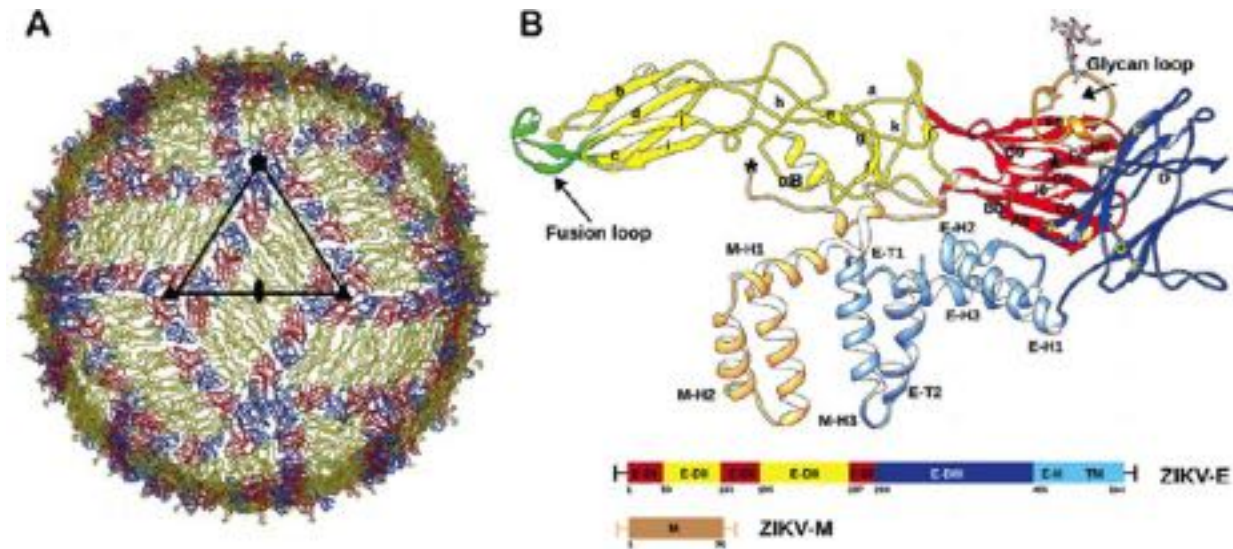


Fig. 8 $T = 3$ icosahedral capsid architecture in flaviviruses. (A) ZIKV capsid formed by herringbone pattern of 6 E-M heterodimers. The E proteins are shown as $C\alpha$ backbone and the icosahedral asymmetric unit is outlined by a black triangle. The E protein domains E-DI, E-DII, E-DIII and the fusion loop are colored in red, yellow, blue and green, respectively. (B) E-M heterodimer with labeled secondary structure elements. The stem and transmembrane helices (E-H1-3, E-T1-T2) of the E protein and the M protein (M-H1-3) are colored in light blue and light brown, respectively. Residue numbering and domain definitions are shown as a linear peptide. Figure adapted from Sevvana, M., Long, F., Miller, A.S., et al., 2018. Refinement and analysis of the mature Zika virus cryo-EM structure at 3.1 Å resolution. Structure 26, 1169–1177.

improvements to microscopes and techniques. For instance, the structure of mature Zika virus (ZIKV) was determined to about 3 Å resolution using cryoEM (Sevvana et al., 2018), thus also confirming the earlier hybrid techniques.

Similar to other flaviviruses, ZIKV is an enveloped, single-stranded, positive-sense RNA virus (Hasan et al., 2018a; Sevvana et al., 2018; Sirohi et al., 2016). The 11 kb RNA genome is translated into a long polyprotein. It is post-translationally cleaved by host and viral proteases into three structural proteins: pre-membrane (prM), envelope (E), and the capsid (C) as well as seven non-structural proteins (Sirohi and Kuhn, 2017). The E and prM proteins form a protective capsid around the genome. The E-protein mediates the assembly of virus, virus entry and fusion with host membrane and contains putative receptor binding sites. It forms a complex with prM in the endoplasmic reticulum shortly after its synthesis. The E-prM complex is arranged as 60 trimeric spikes ($T = 3$) on the surface of the immature virus (Mangala Prasad et al., 2017b). The pr domain of prM prevents premature fusion to membranes during virus production. Immature virions undergo pH-induced conformational changes to form the mature virion. Mature virus particles are formed after the pr domain is cleaved from the prM protein by furin in the trans-Golgi network. The icosahedral mature ZIKV consists of 180 copies of E and M proteins arranged in 60 asymmetric units (quasi- $T = 3$). Each icosahedral asymmetric unit consists of three E-M oligomers. Three parallel dimers from adjacent asymmetric units form a “raft”. Therefore, there are 30 rafts arranged in a herringbone pattern on the surface of the mature virus (Fig. 8). Flavivirus membrane fusion is initiated by E-protein interaction with a potential receptor molecule and the endocytosis of the virus. This is followed by low pH-induced conformational changes in the endosome, exposure of the fusion loop and its interaction with the endosomal membrane. The E-proteins then rearrange into trimeric structures (fusogenic trimer) leading to the fusion of viral and endosomal membrane and subsequent release of viral genome into the host cytosol.

Though, most of the flavivirus structures have icosahedral symmetry, a recent study has shown that immature Kunjin and Zika viruses are not accurately icosahedral (Therkelsen et al., 2019). Their nucleocapsid core approaches close to one side of the external glycoprotein shell, indicating the presence of a special, unique region that may have been created while interacting with glycoproteins to promote budding from the endoplasmic reticulum (ER) in infected cells and may be required for genome packaging in flaviviruses. Furthermore, on the opposite face on the exterior glycoprotein surface, there is a departure from exact icosahedral symmetry that may reflect the final step in budding. Similar asymmetric properties may have been overlooked in the usual averaging procedures assumed in the study of most essentially icosahedral viruses.

Small Icosahedral ssDNA Viruses

Two examples of small icosahedral ssDNA viruses are ΦX174 (McKenna et al., 1992b), which belongs to the *Microviridae* family, with a single-stranded circular DNA and parvoviruses with a linear genome. Parvoviruses have a 5000 nucleotide ssDNA genome, and typically have a $T = 1$ icosahedral capsid that is about 260 Å in diameter (Fig. 9). Most mammalian parvoviruses have 60 subunits of VP1, VP2, and VP3 in the viral capsid. Many parvovirus structures such as human B19 virus-like particles, canine parvovirus, feline parvovirus, porcine VLP's and murine VLP's were determined to near atomic resolution using X-ray crystallography (Agbandje et al., 1993; Kaufmann et al., 2005; 2008; 2004; Simpson et al., 2000; Simpson et al., 2002; Tsao et al., 1991;

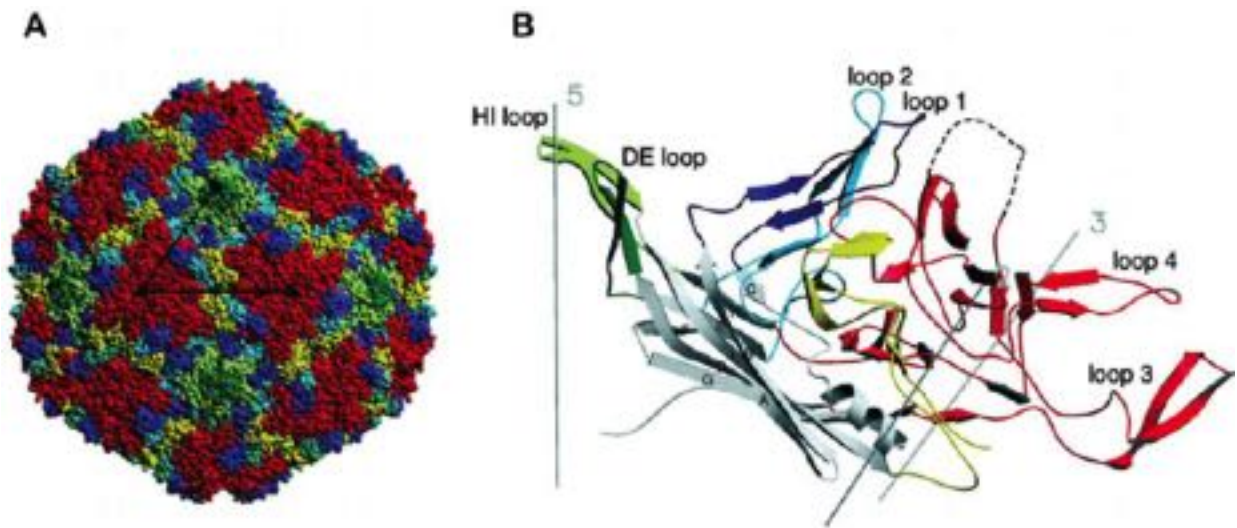


Fig. 9 Structure of icosahedral ($T = 1$), human parvovirus B19 VP2 particles. (A) Surface topography of B19 color-coded according to the ribbon diagram in (B). (B) Secondary structure of parvovirus B19 VP2 showing the jelly-roll β -barrel fold. The surface loops connecting the strands of the β -barrel are colored as follows: dark blue, BC loop; dark green, DE loop; light blue, EF loop; red, GH loop; light green, HI loop; yellow, C-terminal amino acids. Figure adapted from Kaufmann, B., Simpson, A.A., Rossmann, M.G., 2004. The structure of human parvovirus B19. *Proceedings of the National Academy of Sciences of the United States of America* 101, 11628–11633.

Wu and Rossmann, 1993; Xie and Chapman, 1996). Adeno-associated parvoviruses (AAV-2) are of particular interest because of their potential use in gene therapies (DiMattia *et al.*, 2012; Hueffer *et al.*, 2003b; Lerch *et al.*, 2012; Padron *et al.*, 2005; Walters *et al.*, 2004; Xie *et al.*, 2002). The capsid proteins of parvoviruses have large insertions on their jelly-roll structure. These account for most of the, interactions between neighboring subunits as well as receptor interactions (Chapman and Rossmann, 1996; Tsao *et al.*, 1992). For example, canine and murine parvoviruses can use transferrin (TfR) receptor for binding and infection (Hueffer *et al.*, 2003a). However, the TfR molecule only binds to a few of the 60 available binding sites on the virus thereby inducing asymmetry in the icosahedral virus (Hafenstein *et al.*, 2007).

Bacteriophage Φ X174 is a tailless phage that contains a single-stranded, closed and circular DNA with about 5000 nucleotides (Dokland *et al.*, 1999; 1998; McKenna *et al.*, 1992b). The $T = 1$ mature icosahedral virus consists of four structural proteins, J, F, G, and H. The capsid of 260 Å diameter is formed by 60 F subunits. The 12 pentameric spikes formed by 60 copies of the G protein rise about 30 Å above the capsid surface and both F and G subunits have jelly-roll folds. There are 12 copies of the H protein, which form a tail tube similar to the tail tube of the common tailed dsDNA phage during infection for the insertion of DNA into the host cell (Sun *et al.*, 2013a). Structural studies shed light on the assembly of Φ X174, which proceeds via virally encoded scaffolding proteins (B and D) both of which are discarded when assembly is complete (Fig. 10).

Large Icosahedral dsDNA Viruses

Nucleocytoplasmic large dsDNA viruses (NCDLVs) infecting a wide variety of organisms' span members of the *Asfaviiridae*, *Ascoviridae*, *Marseilleviridae*, *Mimiviridae*, *Pandoraviridae*, *Phycodnaviridae*, *Iridoviridae*, and *Poxviridae*. Most NCDLVs are roughly icosahedral in shape. However, members of Poxviruses, Ascoviruses, and Pandoraviruses are pleomorphic. CryoEM reconstructions of several icosahedral NCDLVs infer similar capsid organization with large T numbers (Andreani *et al.*, 2017; Cherrier *et al.*, 2009; Klose *et al.*, 2016; Okamoto *et al.*, 2018; Reteno *et al.*, 2015; Xiao *et al.*, 2017; Yan *et al.*, 2005; 2000; 2009; Zhang *et al.*, 2011b). An inner membrane that separates the genome from the external capsid is observed in most NCDLVs except Faustoviruses, which have an internal capsid.

The capsomeres in icosahedral NCDLVs are arranged in trisymmetrons (20 triangular arrays) and pentasymmetrons (12 pentagonal arrays) (Fig. 11) (Klose and Rossmann, 2014). The capsid is assembled from pseudo-hexagonal, closely arranged trimeric capsomeres where each subunit is formed by a double jelly-roll fold protein. Each capsomere has a diameter of about 75–85 Å and a height of 75 Å. Because of their size range, the capsomeres can have two orientations related by a six-fold rotation. First, the capsomeres within each trisymmetron can have the same orientation or, second, they can be rotated by 60° relative to neighboring capsomeres which creates cleavage planes between neighboring trisymmetrons (Wrigley, 1969). The five-fold vertices are formed by a pentameric single jelly-roll fold protein. The pentasymmetrons have a constant size and consist of 30 trimeric double jelly-roll capsomeres and one pentameric single jelly-roll protein. It was postulated that this constant size is ideal to relieve the strain induced around the five-fold axis for icosahedral NCDLVs with large diameters.

Another characteristic feature of NCDLVs is the presence of minor capsid proteins found underneath the major capsid proteins towards the inside of the virus capsid as can be seen in high resolution structures of faustovirus (Klose *et al.*, 2016), Chilo iridescent virus (CIV) and Paramecium bursaria chlorella virus 1 (PBCV-1) (Zhang *et al.*, 2011b). The minor capsid proteins

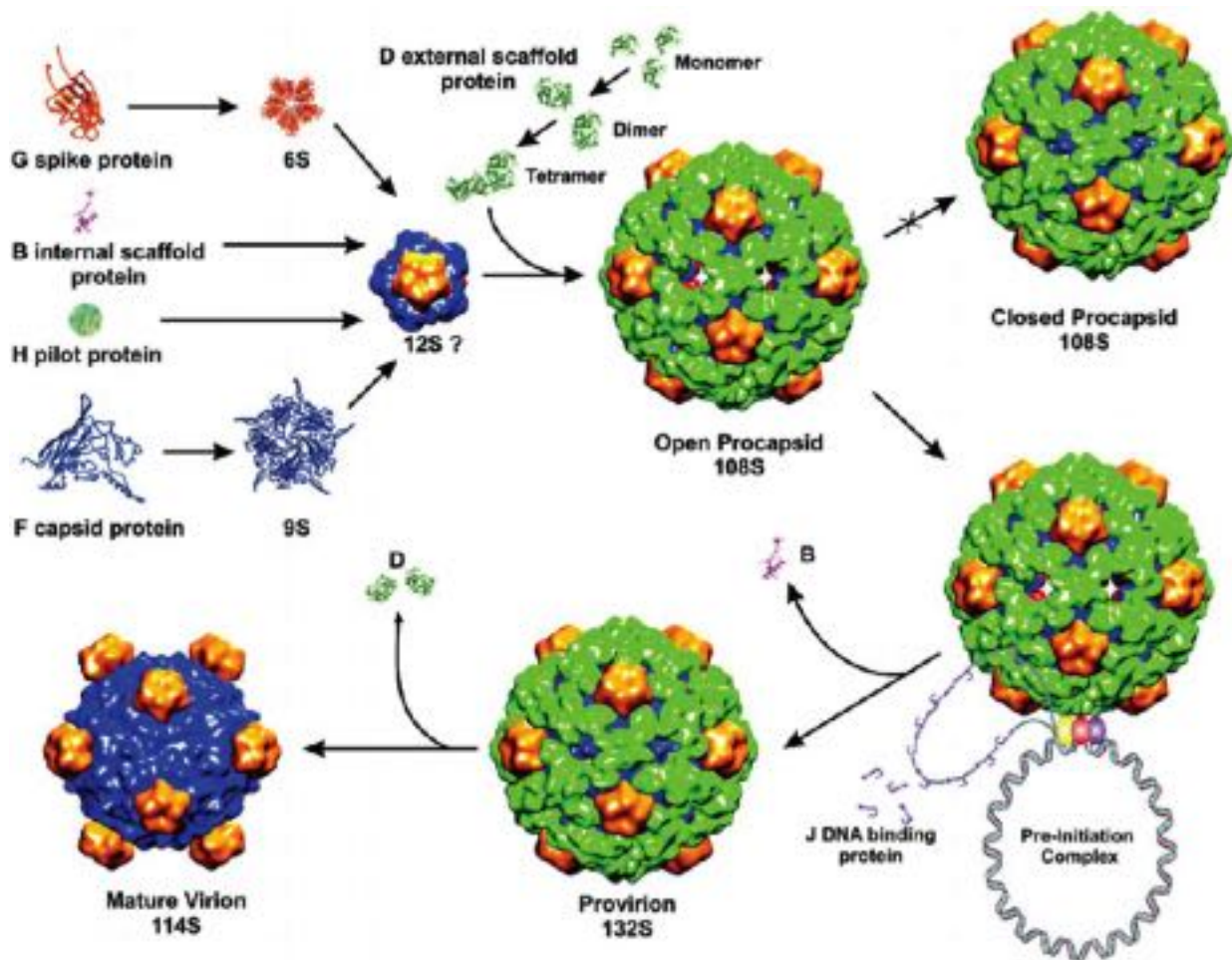


Fig. 10 Capsid assembly pathway in *Microviridae*. The first intermediates are formed by pentameric assemblies of F capsid in 9 s and G spike proteins in 6 s particles. With the help of internal and external scaffolding proteins (B, D, and H), these are later assembled into an empty protein capsid shell called the procapsid. The ssDNA is concurrently synthesized and packaged into the procapsid in complex with DNA-binding protein J, therefore forming the mature virion. Figure adapted from Bernal, R.A., Hafenstein, S., Olson, N.H., *et al.*, 2003. Structural studies of bacteriophage ϕ 3 assembly. *Journal of Molecular Biology* 325, 11–24.

support the outer capsid shell, connect the neighboring trisymmetrons and/or connect the outer capsid to the inner membrane. Some NCLDV s like mimivirus (Xiao *et al.*, 2005,2009; Klose *et al.*, 2010) and PBCV-1 also have a unique vertex where the membrane becomes separated from the capsid. Many NCLDV s have decoration proteins and surface fibers attached to the viral capsid. Mimivirus is covered by a dense layer of surface fibers except near the unique vertex called the stargate, which is tightly associated with the adjacent capsomers.

One of the most studied NCLDV s is PBCV-1, where a recent near-atomic resolution structure identified several of the minor capsid proteins and cast light on the capsid assembly pathway (Fang *et al.*, 2019). PBCV-1 has a 330 kbp genome that encodes 416 proteins, of which 149 different proteins were identified in the mature virion. The nucleocapsid is surrounded by a membrane, which in turn is surrounded by a roughly icosahedral capsid shell giving it an overall size of 190 nm. It also has a unique vertex with a spike structure required for host entry. The outer shell consists of the MCP (Vp54) and the penton protein P1. Each tri- and pentasymmetron contains 66 (MCP) and 31 (30 MCP + 1 P1) capsid proteins respectively, arranged in an icosahedral lattice with a triangulation number of $T = 169d$. The high resolution cryoEM structure led to the identification of the penton protein, P1, and 13 minor capsid proteins (named P2 through P14). The minor capsid proteins form a hexagonal network below the outer capsid shell. Most of the minor capsid proteins are primarily located at the interface between neighboring capsomeres. The size of the viral capsid is determined by a tape-measure protein, P2, and the zip protein P11 which plays a crucial role in cementing neighboring symmetrons (Fig. 11).

Faustovirus, on the other hand uses two concentric capsid shells to protect its genome (Klose *et al.*, 2016). Consequently, two distinct types of particles were observed in cryoEM images of purified virus. The larger icosahedral particle constituting the outer and inner capsids measured about 2600 Å in diameter, whereas the smaller particles with only an inner capsid measured between 1600–1900 Å. The capsomere on the outer capsid is arranged with a triangulation number of $T = 277$ ($h = 7$ and $k = 12$). The

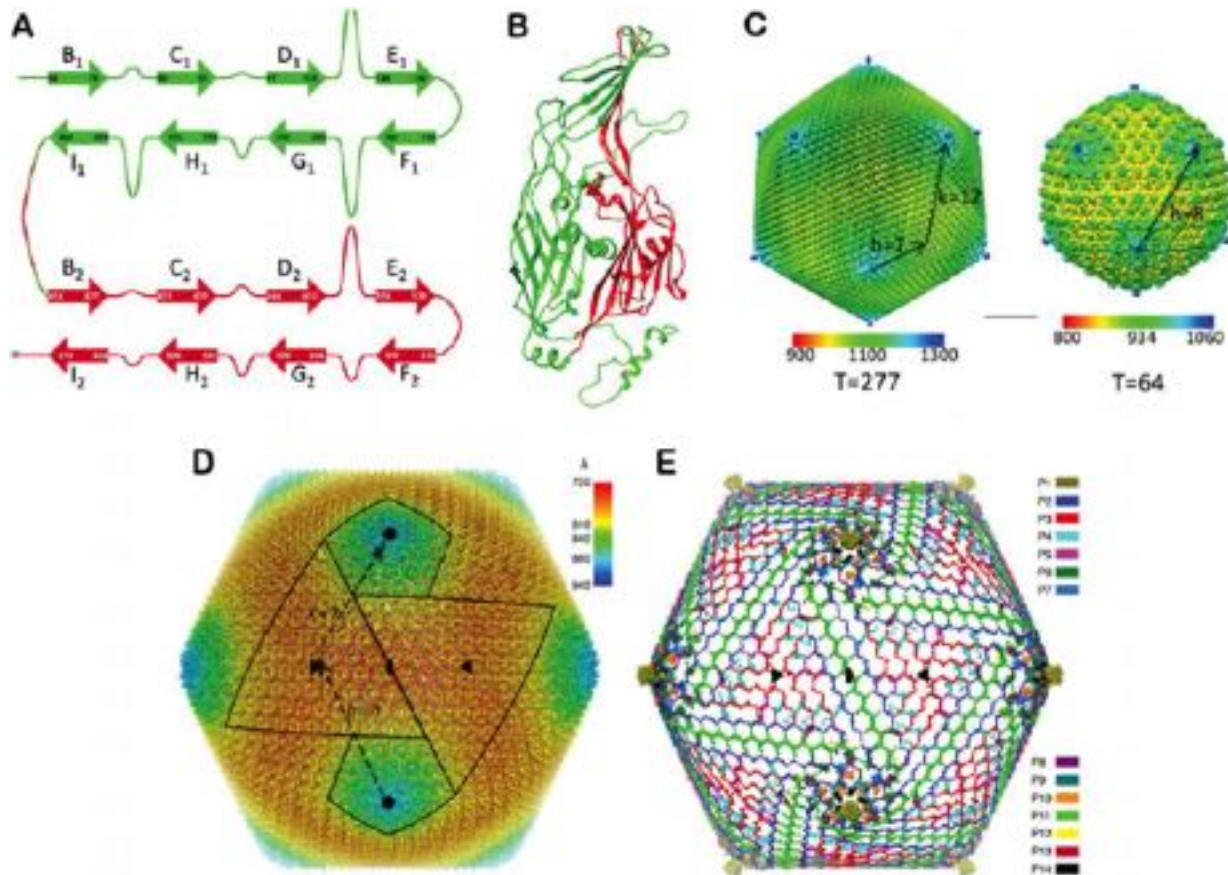


Fig. 11 Capsid organization in NCLDVs. (A) Topology plot and (B) cartoon representation of the double jelly-roll fold of the MCP in faustovirus. The secondary structure elements are labeled. (C) $T = 277$ and $T = 64$ capsid organizations of faustovirus external and internal capsid shells respectively. (D) CryoEM map of the overall structure of the PBCV-1 capsid. The boundaries of neighboring trisymmetrons and pentasymmetrons are outlined in black. (E) The cryoEM density of the minor capsid proteins and the penton proteins after subtracting the density for MCP. Each protein is depicted in a different color as indicated to the right of the image. The h and k indices are labeled on the icosahedron in panels (C) and (D). The radial color gradient scale bar is in Å. Figures adapted from: (A), (B), and (C) Klose, T., Reteno, D.G., Benamar, S., *et al.*, 2016. Structure of faustovirus, a large dsDNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 113, 6206–6211. (D) and (E) Fang, Q.L., Zhu, D.J., Agarkova, I., *et al.*, 2019. Near-atomic structure of a giant virus. *Nature Communications* 10, 388.

icosahedral asymmetric unit is composed of one penton protein and 276 MCPs. Each tri- and pentasymmetron of the outer capsid shell contains 360 (MCP) and 31 (30 MCP + 1 P1) proteins, respectively (Fig. 11).

Tailed Bacteriophages

The structure of dsDNA bacteriophages is unique in that most of them have tails attached to a special pentameric vertex of their capsids (Fokine and Rossmann, 2014). The unique vertex in most bacteriophages is occupied by a dodecameric portal protein, which forms a channel for genome packaging during virion assembly and subsequent exit during infection. Tailed bacteriophages belong to the order *Caudovirales*, which are further divided into three families: *Myoviridae* (i.e., T4 and ϕ 92 have long, straight contractile tails), *Siphoviridae* (i.e., lambda and HK97) have long flexible non-contractile tails, and *Podoviridae* (i.e., T7) have short non-contractile tails.

Capsid heads of tailed bacteriophages come in a large variety of sizes and range in diameter from about 400–1700 Å with either icosahedral or prolate heads. The capsid contains linear double-stranded DNA packaged to a high density of about 500 g/l and exerts an internal pressure of tens of atmospheres on the capsid walls. The capsid precursor, called the procapsid or prohead, is assembled at the portal vertex by copolymerization of the MCP and scaffolding proteins. The scaffolding protein regulates the correct geometry of the capsid. The prohead consists of the portal protein, the outer MCP, the internal scaffolding protein and in some cases other minor capsid proteins. The scaffolding protein and the internal core are degraded in the procapsids of many phages by head maturation proteases, therefore making space for DNA packaging. The genome is packaged by a DNA translocation motor via the portal vertex and is ATP-dependent. DNA packaging is followed by large structural rearrangements, an increase in internal volume and attachment of other proteins (for example decoration proteins) to form the mature capsid. After DNA packaging, the DNA translocation motor is removed and the capsid assembly is completed by sealing the portal gate, the

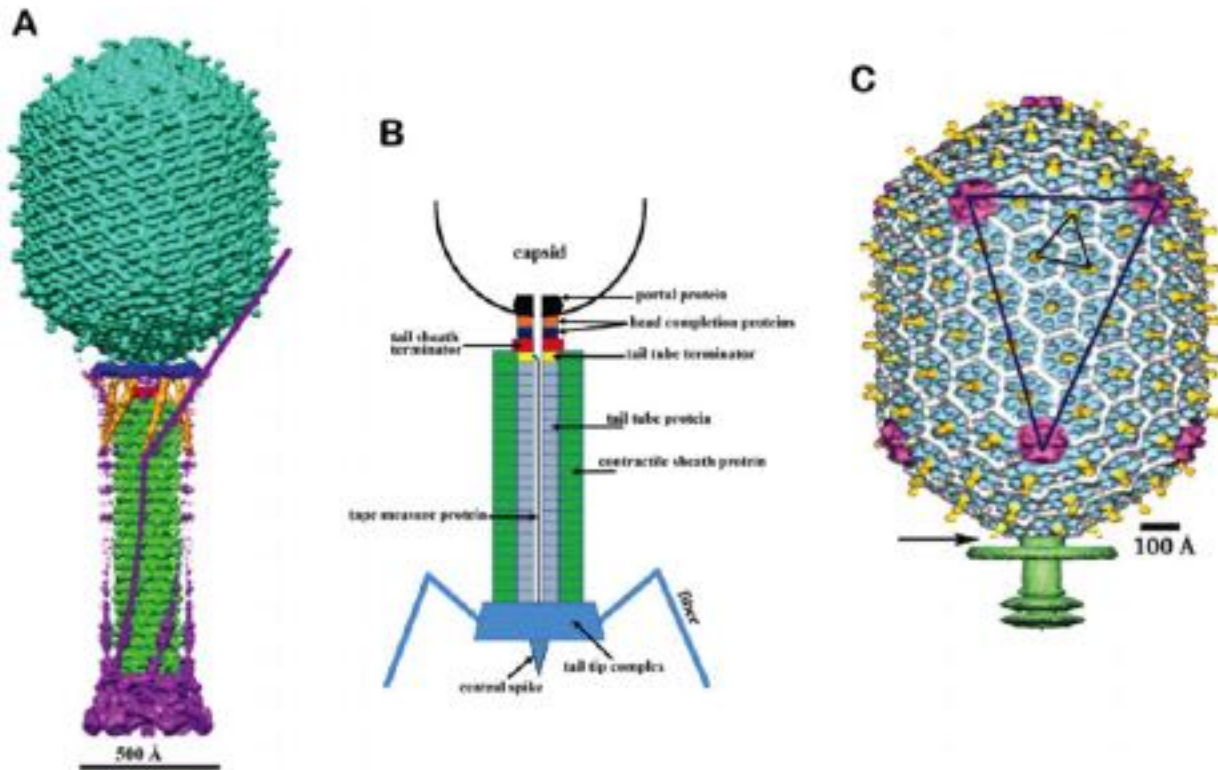


Fig. 12 Structure of T4 bacteriophage ($T = 13$ laevo, $Q = 20$, $h_1 = 3$, $k_1 = 1$, $h_2 = 4$, and $k_2 = 2$). (A) CryoEM reconstruction of the T4 capsid and extended tail. The capsid, tail sheath, phage collar and whiskers are colored cyan, green, blue and gold respectively. The density corresponding to the baseplate and long tail fibers is colored magenta. (B) Schematic representation of the bacteriophage T4 neck and tail components. (C) CryoEM reconstruction of the T4 head viewed perpendicular to the 5-fold axis. The capsid proteins, gp23 and gp24, are shown in blue and magenta respectively, *soc* is in white, *hoc* is in yellow, and the tail is in green. The facet triangle is shown in blue and the basic triangles in black. Adapted from: (A) Fokine, A., Zhang, Z., Kanamaru, S., *et al.*, 2013. The molecular architecture of the bacteriophage T4 neck. *Journal of Molecular Biology* 425, 1731–1744. (B) Fokine, A., Rossmann, M.G., 2014. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* 4, e28281. (C) Fokine, A., Chipman, P.R., Leiman, P.G., *et al.*, 2004. Molecular architecture of the prolate head of bacteriophage T4. *Proceedings of the National Academy of Sciences of the United States of America* 101, 6003–6008.

attachment of head completion proteins and the assembly of the tail. Bacteriophage tails are designed to recognize host cells, penetrate the cell envelope and deliver the genome. Their length varies from ~ 100 Å in *Podoviridae* to ~ 8000 Å in some *Siphoviridae*. Most tails of the members of *Myoviridae* and *Siphoviridae* consist of a tail tip complex for host cell recognition, a tail tube, which makes a conduit for genomic DNA with a contractile sheath around it in some phages and the terminator proteins near the tail-head interface (Fig. 12).

One of the most studied bacteriophage structures is of T4 bacteriophage, which infects *E. coli*. The dsDNA genome encodes about 40 structural proteins. The mature virus consists of an 1150 Å-long and 850 Å-wide prolate head encapsidating the genomic DNA. The head is attached to a 1000 Å-long and 210 Å-diameter cylindrical contractile tail and is terminated with a 460 Å-diameter baseplate (Fokine *et al.*, 2004; Kostyuchenko *et al.*, 2003) (Fig. 12). The baseplate is associated with six 1450 Å-long fibers (Aksyuk *et al.*, 2009; Kostyuchenko *et al.*, 2005). The distal ends of the long tail fibers (made of gp37) recognize the host cell (Bartual *et al.*, 2010) and then transfer the signal to the baseplate to unfold the short tail fibers (gp12) (Thomassen *et al.*, 2003; van Raaij *et al.*, 2001). The short tail fibers subsequently bind to the lipopolysaccharides on the surface of the *E. coli* host. The unfolding of the short tail fibers results in large conformational changes of the dome-shaped baseplate to a star-shape as the tail sheath (gp18) starts to contract (Aksyuk *et al.*, 2009; Aksyuk and Rossmann, 2011; Crowther *et al.*, 1977; Kostyuchenko *et al.*, 2003; Leiman *et al.*, 2004). The tip of the tail tube (gp5) punctures the cell, enters the periplasm and digests the cell wall (Browning *et al.*, 2012; Kanamaru *et al.*, 2002). The genome is then ejected into the host's cytoplasm by the tail tube. In phage T4 the pentameric protein forming 11 of the 12 vertices (gp24) has an HK97 fold, which has a high sequence similarity to T4 MCP (gp23) (Jiang *et al.*, 2003; Morais *et al.*, 2005).

Conclusion

Viruses are complex organisms that build extraordinary structures out of a limited number of building blocks. It is remarkable that the principles established by Watson and Crick (Watson and Crick, 1953b) and expanded by Caspar and Klug (Caspar and Klug,

1962) in the 1960s have been validated and still hold true for most virus structures discovered to date. Furthermore, it is worth noting that the large variety of virus structures that have been described in the scientific literature so far utilize a fairly small number of protein folds, most of which have been covered in this review. Improvements in sequencing technologies and less stringent isolation techniques have led to an immense increase in the number of viruses that have been described in the literature and it will be interesting to see if this increase eventually correlates with the discovery of new folds or new assemblies. It is already becoming clear that the future of structural virology will be dominated by cryoEM. Most of the recently determined virus structures were solved using this technique and one would expect this number to continue increasing. X-Ray crystallography will still play an important role, but will likely be confined to solving the structures of smaller components of virions or assemblies. In addition, cryoEM and cryoET will allow us to look into deviations of the strict symmetry requirements imposed onto most viral structures to date and help us to learn more about assembly pathways and unique structural components that might otherwise be overlooked. This might show us that viruses are not that perfect after all and may lead to important biological discoveries and potentially new therapeutic approaches.

Acknowledgments

MGR would like to thank all his many collaborators, post-docs, graduate students and colleagues who made it possible to accomplish many of the findings reported here. MGR is also grateful to the National Science Foundation, the National Institutes of Health and Purdue University for their generous support over many years. We thank Sheryl Kelly for her help in preparing the paper. MS and TK would like to thank MGR for the opportunity to co-write this article. This was the last paper MGR worked on and submitted a month before his tragic demise. It was an exciting time-travel through several decades of the history of structural virology.

References

- Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., *et al.*, 1980. Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature* 286, 33–39.
- Abad-Zapatero, C., Abdel-Meguid, S.S., Johnson, J.E., *et al.*, 1981. A description of techniques used in the structure determination of southern bean mosaic virus at 2.8 Å resolution. *Acta Crystallographica B* 37, 2002–2018.
- Ackermann, H.-W., 1999. Tailed bacteriophages: The order Caudovirales. *Advances in Virus Research* 51, 135–201.
- Adrian, M., Dubochet, J., Lepault, J., McDowell, A.W., 1984. Cryo-electron microscopy of viruses. *Nature* 308, 32–36.
- Agbandje, M., McKenna, R., Rossmann, M.G., Strassheim, M.L., Parrish, C.R., 1993. Structure determination of feline panleukopenia virus empty particles. *Proteins* 16, 155–171.
- Åkervall, K., Strandberg, B., Rossmann, M.G., *et al.*, 1972. X-ray diffraction studies of the structure of satellite tobacco necrosis virus. *Cold Spring Harbor Symposia on Quantitative Biology* 36, 469–483.
- Aksyuk, A.A., Leiman, P.G., Kurochkina, L.P., *et al.*, 2009. The tail sheath structure of bacteriophage T4: A molecular machine for infecting bacteria. *EMBO Journal* 28, 821–829.
- Aksyuk, A.A., Rossmann, M.G., 2011. Bacteriophage assembly. *Viruses* 3, 172–203.
- Andreani, J., Khalil, J.Y.B., Sevana, M., *et al.*, 2017. Pacmanvirus, a new giant icosahedral virus at the crossroads between *Asfarviridae* and *Faustoviruses*. *Journal of Virology* 91, e00212–e00217.
- Arnold, E., Rossmann, M.G., 1986. Effect of errors, redundancy, and solvent content in the molecular replacement procedure for the structure determination of biological macromolecules. *Proceedings of the National Academy of Sciences of the United States of America* 83, 5489–5493.
- Arnold, E., Vriend, G., Luo, M., *et al.*, 1987. The structure determination of a common cold virus, human rhinovirus 14. *Acta Crystallographica Section A* 43, 346–361.
- Bartual, S.G., Otero, J.M., Garcia-Doval, C., *et al.*, 2010. Structure of the bacteriophage T4 long tail fiber receptor-binding tip. *Proceedings of the National Academy of Sciences of the United States of America* 107, 20287–20292.
- Battisti, A.J., Yoder, J.D., Plevka, P., *et al.*, 2012. Cryo-electron tomography of rubella virus. *Journal of Virology* 86, 11078–11085.
- Bawden, F.C., Pirie, N.W., 1938. Crystalline preparations of tomato bushy stunt virus. *British Journal of Experimental Pathology* 19, 251–263.
- Bernal, J.D., Fankuchen, I., 1941a. X-ray and crystallographic studies of plant virus preparations: I. Introduction and preparation of specimens. II. Modes of aggregation of the virus particles. III. (1) The structure of the particles and (2) biological implications. *Journal of General Physiology* 25, 111–165.
- Bernal, J.D., Fankuchen, I., 1941b. X-ray and crystallographic studies of plant virus preparations. III. *Journal of General Physiology* 25, 147–165.
- Bloomer, A.C., Champness, J.N., Bricogne, G., Staden, R., Klug, A., 1978. Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. *Nature* 276, 362–368.
- Brenner, S., Horne, R.W., 1959. A negative staining method for high resolution electron microscopy of viruses. *Biochimica et Biophysica Acta* 34, 103–110.
- Browning, C., Shneider, M.M., Bowman, V.D., Schwarzer, D., Leiman, P.G., 2012. Phage pierces the host cell membrane with the iron-loaded spike. *Structure* 20, 326–339.
- Buehner, M., Ford, G.C., Moras, D., Olsen, K.W., Rossmann, M.G., 1974. Structure determination of crystalline lobster α -glyceraldehyde-3-phosphate dehydrogenase. *Journal of Molecular Biology* 82, 563–585.
- Butler, P.J., Klug, A., 1972. Assembly of tobacco mosaic virus in vitro: Effect of state of polymerization of the protein component. *Proceedings of the National Academy of Sciences of the United States of America* 69, 2950–2953.
- Carrillo-Tripp, M., Shepherd, C.M., Borelli, I.A., *et al.*, 2009. VIPERdb(2): An enhanced and web API enabled relational database for structural virology. *Nucleic Acids Research* 37, D436–D442.
- Caspar, D.L.D., 1956. Structure of bushy stunt virus. *Nature* 177, 475–476.
- Caspar, D.L.D., Klug, A., 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposia on Quantitative Biology* 27, 1–24.
- Champness, J.N., Bloomer, A.C., Bricogne, G., Butler, P.G., Klug, A., 1976. The structure of the protein disk of tobacco mosaic virus to 5 Å resolution. *Nature* 259, 20–24.
- Chapman, M.S., Lijias, L., 2003. Structural folds of viral proteins. *Advances in Protein Chemistry* 64, 125–196.
- Chapman, M.S., Rossmann, M.G., 1996. Structural refinement of the DNA-containing capsid of canine parvovirus using RSRef, a resolution-dependent stereochemically restrained real-space refinement method. *Acta Crystallographica Section D: Biological Crystallography* 52, 129–142.
- Chelvanayagam, G., Heringa, J., Argos, P., 1992. Anatomy and evolution of proteins displaying the viral capsid jellyroll topology. *Journal of Molecular Biology* 228, 220–242.
- Cheng, S., Brooks III, C.L., 2013. Viral capsid proteins are segregated in structural fold space. *PLOS Computational Biology* 9, e1002905.

- Cheng, R.H., Kuhn, R.J., Olson, N.H., *et al.*, 1995. Nucleocapsid and glycoprotein organization in an enveloped virus. *Cell* 80, 621–630.
- Chen, L.H., Wang, M., Zhu, D.J., *et al.*, 2018. Implication for alphavirus host-cell entry and assembly indicated by a 3.5 Å resolution cryo-EM structure. *Nature Communications* 9, 5326.
- Cherrier, M.V., Kostyuchenko, V.A., Xiao, C., *et al.*, 2009. An icosahedral algal virus has a complex unique vertex decorated by a spike. *Proceedings of the National Academy of Sciences of the United States of America* 106, 11085–11089.
- Choi, H.K., Tong, L., Minor, W., *et al.*, 1991a. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature* 354, 37–43.
- Choi, H.K., Tong, L., Minor, W., *et al.*, 1991b. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature* 354, 37–43.
- Conway, J.F., Cheng, N., Zlotnick, A., *et al.*, 1997. Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* 386, 91–94.
- Crick, F.H.C., Watson, J.D., 1956. Structure of small viruses. *Nature* 177, 473–475.
- Crick, F.H.C., Watson, J.D., 1957. Virus structure: General principles. In: Wolstenholme, G.E.W., Millar, E.C.P. (Eds.), *Ciba Foundation Symposium on "The Nature of Viruses"*. Chichester, UK: John Wiley & Sons, Ltd, pp. 5–13.
- Crowther, R.A., Amos, L.A., Finch, J.T., DeRosier, D.J., Klug, A., 1970. Three-dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature* 226, 421–425.
- Crowther, R.A., Lenk, E.V., Kikuchi, Y., King, J., 1977. Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *Journal of Molecular Biology* 116, 489–523.
- DiMattia, M.A., Nam, H.J., Van Vliet, K., *et al.*, 2012. Structural insight into the unique properties of adeno-associated virus serotype 9. *Journal of Virology* 86, 6947–6958.
- Dokland, T., Bernal, R.A., Burch, A., *et al.*, 1999. The role of scaffolding proteins in the assembly of the small, single-stranded DNA virus ϕ X174. *Journal of Molecular Biology* 288, 595–608.
- Dokland, T., McKenna, R., Sherman, D.M., *et al.*, 1998. Structure determination of the ϕ X174 closed procapsid. *Acta Crystallographica Section D: Biological Crystallography* 54, 878–890.
- Dubochet, J., Adrian, M., Chang, J.J., *et al.*, 1988. Cryo-electron microscopy of vitrified specimens. *Quarterly Reviews of Biophysics* 21, 129–228.
- Enard, D., Cai, L., Gwennap, C., Petrov, D.A., 2016. Viruses are a dominant driver of protein adaptation in mammals. *eLife* 5, e12469.
- Erickson, J.W., Frankenberger, E.A., Rossmann, M.G., *et al.*, 1983. Crystallization of a common cold virus, human rhinovirus 14: "Isomorphism" with poliovirus crystals. *Proceedings of the National Academy of Sciences of the United States of America* 80, 931–934.
- Fang, Q.L., Zhu, D.J., Agarkova, I., *et al.*, 2019. Near-atomic structure of a giant virus. *Nature Communications* 10, 388.
- Filman, D.J., Syed, R., Chow, M., *et al.*, 1989. Structural factors that control conformational transitions and serotype specificity in type 3 poliovirus. *EMBO Journal* 18, 1567–1579.
- Fokine, A., Chipman, P.R., Leiman, P.G., *et al.*, 2004. Molecular architecture of the prolate head of bacteriophage T4. *Proceedings of the National Academy of Sciences of the United States of America* 101, 6003–6008.
- Fokine, A., Rossmann, M.G., 2014. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* 4, e28281.
- Frank, J., 1992. *Electron Tomography: Methods for Three-dimensional Visualization of Structures in the Cell*. New York: Springer Science + Business Media LLC.
- Franklin, R.E., Holmes, K.C., 1958. Tobacco mosaic virus: Application of the method of isomorphous replacement to the determination of the helical parameters and radial density distribution. *Acta Crystallographica* 11, 213–220.
- Franklin, R.E., Klug, A., Holmes, K.C., 1957. X-ray diffraction studies of the structure and morphology of tobacco mosaic virus. In: Wolstenholme, G.E.W., Millar, E.C.P. (Eds.), *Ciba Foundation Symposium on the Nature of Viruses*. J. & A. Churchill, London: Ciba Foundation, pp. 39–55.
- Fuller, S.D., 1987. The T=4 envelope of Sindbis virus is organized by interactions with a complementary T=3 capsid. *Cell* 48, 923–934.
- Gamble, T.R., Yoo, S., Vajdos, F.F., *et al.*, 1997. Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* 278, 849–853.
- Ge, P., Zhou, Z.H., 2011. Hydrogen-bonding networks and RNA bases revealed by cryo electron microscopy suggest a triggering mechanism for calcium switches. *Proceedings of the National Academy of Sciences of the United States of America* 108, 9637–9642.
- Gibbons, D.L., Kielian, M., 2002. Molecular dissection of the Semliki Forest virus homotrimer reveals two functionally distinct regions of the fusion protein. *Journal of Virology* 76, 1194–1205.
- Greve, J.M., Davis, G., Meyer, A.M., *et al.*, 1989. The major human rhinovirus receptor is ICAM-1. *Cell* 56, 839–847.
- Haas, D.J., Rossmann, M.G., 1970. Crystallographic studies on lactate dehydrogenase at -75°C . *Acta Crystallographica Section B: Structural Science* 26, 998–1004.
- Hafenstein, S., Palermo, L.M., Kostyuchenko, V.A., *et al.*, 2007. Asymmetric binding of transferrin receptor to parvovirus capsids. *Proceedings of the National Academy of Sciences of the United States of America* 104, 6585–6589.
- Harrison, S., 2001. *Principles of virus structure*. In: *Fields Virology*. Philadelphia: Lippincott Williams & Wilkins Publishers.
- Harrison, S.C., Olson, A.J., Schutt, C.E., Winkler, F.K., Bricogne, G., 1978. Tomato bushy stunt virus at 2.9 Å resolution. *Nature* 276, 368–373.
- Harrison, S.C., Strong, R.K., Schlesinger, S., Schlesinger, M.J., 1992. Crystallization of Sindbis virus and its nucleocapsid. *Journal of Molecular Biology* 226, 277–280.
- Hasan, S.S., Sevana, M., Kuhn, R.J., Rossmann, M.G., 2018a. Structural biology of Zika virus and other flaviviruses. *Nature Structural & Molecular Biology* 25, 13–20.
- Hasan, S.S., Sun, C., Kim, A.S., *et al.*, 2018b. Cryo-EM structures of Eastern equine encephalitis virus reveal mechanisms of virus disassembly and antibody neutralization. *Cell Reports* 25, 3136–3147.e5.
- Hayden, F.G., Herrington, D.T., Coats, T.L., *et al.*, 2003. Efficacy and safety of oral pleconaril for treatment of picornavirus colds in adults: Results of two double-blind, randomized, placebo-controlled trials. *Clinical Infectious Diseases* 36, 1523–1532.
- Helgstrand, C., Wikoff, W.R., Duda, R.L., *et al.*, 2003. The refined structure of a protein catenane: The HK97 bacteriophage capsid at 3.44 Å resolution. *Journal of Molecular Biology* 334, 885–899.
- Hogle, J.M., Chow, M., Filman, D.J., 1985. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* 229, 1358–1365.
- Holmes, K.C., Stubbs, G.J., Mandelkow, E., Gallwitz, U., 1975. Structure of tobacco mosaic virus at 6.7 Å resolution. *Nature* 254, 192–196.
- Horne, R.W., Brenner, S., Waterson, A.P., Wildy, P., 1959. The icosahedral form of an adenovirus. *Journal of Molecular Biology* 1, 84–86.
- Hueffer, K., Govindasamy, L., Agbandje-McKenna, M., Parrish, C.R., 2003a. Combinations of two capsid regions controlling canine host range determine canine transferrin receptor binding by canine and feline parvoviruses. *Journal of Virology* 77, 10099–10105.
- Hueffer, K., Parker, J.S.L., Weichert, W.S., *et al.*, 2003b. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. *Journal of Virology* 77, 1718–1726.
- Jiang, W., Li, Z., Zhang, Z., *et al.*, 2003. Coat protein fold and maturation transition of bacteriophage P22 seen at subnanometer resolutions. *Nature Structural Biology* 10, 131–135.
- Johnson, J.E., Chiu, W., 2000. Structures of virus and virus-like particles. *Current Opinion in Structural Biology* 10, 229–235.
- Johnson, J.E., Speir, J.A., 1997. Quasi-equivalent viruses: A paradigm for protein assemblies. *Journal of Molecular Biology* 269, 665–675.
- Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., *et al.*, 2002. Structure of the cell-puncturing device of bacteriophage T4. *Nature* 415, 553–557.
- Kaufmann, B., Baxa, U., Chipman, P.R., *et al.*, 2005. Parvovirus B19 does not bind to membrane-associated globoside in vitro. *Virology* 332, 189–198.
- Kaufmann, B., Chipman, P.R., Kostyuchenko, V.A., Modrow, S., Rossmann, M.G., 2008. Visualization of the externalized VP2 N-termini of infectious human parvovirus B19. *Journal of Virology* 82, 7306–7312.

- Kaufmann, B., Plevka, P., Kuhn, R.J., Rossmann, M.G., 2010. Crystallization and preliminary X-ray diffraction analysis of West Nile virus. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 66, 558–562.
- Kaufmann, B., Simpson, A.A., Rossmann, M.G., 2004. The structure of human parvovirus B19. *Proceedings of the National Academy of Sciences of the United States of America* 101, 11628–11633.
- Kielian, M., Rey, F.A., 2006. Virus membrane-fusion proteins: More than one way to make a hairpin. *Nature Reviews Microbiology* 4, 67–76.
- Klose, T., Kuznetsov, Y.G., Xiao, C., *et al.*, 2010. The three-dimensional structure of Mimivirus. *Intervirology* 53, 268–273.
- Klose, T., Reteno, D.G., Benamar, S., *et al.*, 2016. Structure of faustovirus, a large dsDNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 113, 6206–6211.
- Klose, T., Rossmann, M.G., 2014. Structure of large dsDNA viruses. *Biological Chemistry* 395, 711–719.
- Klug, A., Caspar, D.L., 1960. The structure of small viruses. *Advances in Virus Research* 7, 225–325.
- Kostyuchenko, V.A., Chipman, P.R., Leiman, P.G., *et al.*, 2005. The tail structure of bacteriophage T4 and its mechanism of contraction. *Nature Structural & Molecular Biology* 12, 810–813.
- Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., *et al.*, 2003. Three-dimensional structure of bacteriophage T4 baseplate. *Nature Structural Biology* 10, 688–693.
- Kuhn, R.J., 2007. *Togaviridae*: The viruses and their replication. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, Fifth Lippincott Williams & Wilkins, pp. 1001–1021.
- Kuhn, R.J., Zhang, W., Rossmann, M.G., *et al.*, 2002. Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion. *Cell* 108, 717–725.
- Lee, K.K., Johnson, J.E., 2003. Complementary approaches to structure determination of icosahedral viruses. *Current Opinion in Structural Biology* 13, 558–569.
- Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V., Rossmann, M.G., 2004. Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* 118, 419–429.
- Lerch, T.F., O'Donnell, J.K., Meyer, N.L., *et al.*, 2012. Structure of AAV-DJ, a retargeted gene therapy vector: Cryo-electron microscopy at 4.5 Å resolution. *Structure* 20, 1310–1320.
- Lescar, J., Roussel, A., Wein, M.W., *et al.*, 2001. The fusion glycoprotein shell of Semliki Forest virus: An icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105, 137–148.
- Li, L., Jose, J., Xiang, Y., Kuhn, R.J., Rossmann, M.G., 2010. Structural changes of envelope proteins during alphavirus fusion. *Nature* 468, 705–708.
- Lucic, V., F. F., Baumeister, W., 2005. Structural studies by electron tomography: From cells to molecules. *Annual Review of Biochemistry* 74, 833–865.
- Lu, Y.E., Cassese, T., Kielian, M., 1999. The cholesterol requirement for Sindbis virus entry and exit and characterization of a spike protein region involved in cholesterol dependence. *Journal of Virology* 73, 4272–4278.
- Mangala Prasad, V., Fokine, A., Battisti, A.J., *et al.*, 2013. Rubella virus capsid protein structure and its role in virus assembly and infection. *Proceedings of the National Academy of Sciences of the United States of America* 110, 20105–20110.
- Mangala Prasad, V., Klose, T., Rossmann, M.G., 2017a. Assembly, maturation and three-dimensional helical structure of the teratogenic rubella virus. *PLOS Pathogens* 13, e1006377.
- Mangala Prasad, V., Miller, A.S., Klose, T., *et al.*, 2017b. Structure of the immature Zika virus at 9 Å resolution. *Nature Structural & Molecular Biology* 24, 184–186.
- Marsh, M., Helenius, A., 1989. Virus entry into animal cells. *Advances in Virus Research* 36, 107–151.
- McKenna, R., Xia, D., Willingmann, P., *et al.*, 1992a. Atomic structure of single-stranded DNA bacteriophage ΦX174 and its functional implications. *Nature* 355, 137–143.
- McKenna, R., Xia, D., Willingmann, P., Ilag, L.L., Rossmann, M.G., 1992b. Structure determination of the bacteriophage ΦX174. *Acta Crystallographica Section B: Structural Science* 48, 499–511.
- Morais, M.C., Choi, K.H., Koti, J.S., *et al.*, 2005. Conservation of the capsid structure in tailed dsDNA bacteriophages: The pseudoatomic structure of ϕ29. *Molecular Cell* 18, 149–159.
- Mukhopadhyay, S., Kuhn, R.J., Rossmann, M.G., 2005. A structural perspective of the flavivirus life cycle. *Nature Reviews Microbiology* 3, 13–22.
- Namba, K., Pattanayek, R., Stubbs, G., 1989a. Visualization of protein-nucleic acid interactions in a virus: Refined structure of intact tobacco mosaic-virus at 2.9-Å resolution by X-ray fiber diffraction. *Journal of Molecular Biology* 208, 307–325.
- Okamoto, K., Miyazaki, N., Reddy, H.K.N., *et al.*, 2018. Cryo-EM structure of a *Marseilleviridae* virus particle reveals a large internal microassembly. *Virology* 516, 239–245.
- Oliveira, M.A., Zhao, R., Lee, W., *et al.*, 1993. The structure of human rhinovirus 16. *Structure* 1, 51–68.
- Olson, N.H., Kolatkar, P.R., Oliveira, M.A., *et al.*, 1993. Structure of a human rhinovirus complexed with its receptor molecule. *Proceedings of the National Academy of Sciences of the United States of America* 90, 507–511.
- Padron, E., Bowman, V., Kaludov, N., *et al.*, 2005. Structure of adeno-associated virus type 4. *Journal of Virology* 79, 5047–5058.
- Prasad, B.V.V., Schmid, M.F., 2012. Principles of virus structural organization. *Viral Molecular Machines* 726, 17–47.
- Raoult, D., Audic, S., Robert, C., *et al.*, 2004. The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350.
- Rao, V.B., Black, L.W., 2010. Structure and assembly of bacteriophage T4 head. *Virology Journal* 7, 356.
- Reteno, D.G., Benamar, S., Khalil, J.B., *et al.*, 2015. Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae. *Journal of Virology* 89, 6585–6594.
- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375, 291–298.
- Richardson, J.S., 1981. The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry* 34, 167–339.
- Rossmann, M.G., 1979. Processing oscillation diffraction data for very large unit cells with an automatic convolution technique and profile fitting. *Journal of Applied Crystallography* 12, 225–238.
- Rossmann, M.G., 1984. Synchrotron radiation studies of large proteins and supramolecular structures. In: Diakun, G.P.G., D.C. (Eds.), *Biological Systems: Structure and Analysis*. Daresbury: Science and Engineering Research Council, pp. 28–40.
- Rossmann, M.G., 1989. The canyon hypothesis. Hiding the host cell receptor attachment site on a viral surface from immune surveillance. *Journal of Biological Chemistry* 264, 14587–14590.
- Rossmann, M.G., 1994. Viral cell recognition and entry. *Protein Science* 3, 1712–1725.
- Rossmann, M.G., 1999. Synchrotron radiation as a tool for investigating virus structures. *Journal of Synchrotron Radiation* 6, 816–821.
- Rossmann, M.G., 2013. Structure of viruses: A short history. *Quarterly Reviews of Biophysics* 46, 133–180.
- Rossmann, M.G., Arnold, E., Erickson, J.W., *et al.*, 1985. Structure of a human common cold virus and functional relationship to other picornaviruses. *Nature* 317, 145–153.
- Rossmann, M.G., Blow, D.M., 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* 15, 24–31.
- Rossmann, M.G., Blow, D.M., 1963. Determination of phases by the conditions of non-crystallographic symmetry. *Acta Crystallographica* 16, 39–45.
- Rossmann, M.G., Johnson, J.E., 1989. Icosahedral RNA virus structure. *Annual Review of Biochemistry* 58, 533–573.
- Rossmann, M.G., Leslie, A.G.W., Abdel-Meguid, S.S., Tsukihara, T., 1979. Processing and post-refinement of oscillation camera data. *Journal of Applied Crystallography* 12, 570–581.
- Rossmann, M.G., McKenna, R., Tong, L., *et al.*, 1992a. Molecular replacement real-space averaging. *Journal of Applied Crystallography* 25, 166–180.
- Rossmann, M.G., McKenna, R., Tong, L., *et al.*, 1992b. Molecular replacement real space averaging. In: Dodson, E., Gover, S., Wolf, W. (Eds.), *Molecular Replacement*. Proceedings of the CCP4 Study Weekend, 31 January – 1 February 1992. Daresbury: Science and Engineering Research Council, pp. 33–48.
- Rossmann, M.G., Palmenberg, A.C., 1988. Conservation of the putative receptor attachment site in picornaviruses. *Virology* 164, 373–382.
- Sewana, M., Long, F., Miller, A.S., *et al.*, 2018. Refinement and analysis of the mature Zika virus cryo-EM structure at 3.1 Å resolution. *Structure* 26, 1169–1177.
- Sherry, B., Mosser, A.G., Colonna, R.J., Rueckert, R.R., 1986. Use of monoclonal antibodies to identify four neutralization immunogens on a common cold picornavirus, human rhinovirus 14. *Journal of Virology* 57, 246–257.

- Sherry, B., Rueckert, R., 1985. Evidence for at least two dominant neutralization antigens on human rhinovirus 14. *Journal of Virology* 53, 137–143.
- Simpson, A.A., Chandrasekar, V., Hébert, B., *et al.*, 2000. Host range and variability of calcium binding by surface loops in the capsids of canine and feline parvoviruses. *Journal of Molecular Biology* 300, 597–610.
- Simpson, A.A., Hebert, B., Sullivan, G.M., *et al.*, 2002. The structure of porcine parvovirus: Comparison with related viruses. *Journal of Molecular Biology* 315, 1189–1198.
- Sirohi, D., Chen, Z., Sun, L., *et al.*, 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* 352, 467–470.
- Sirohi, D., Kuhn, R.J., 2017. Zika virus structure, maturation, and receptors. *Journal of Infectious Diseases* 216, S935–S944.
- Smith, T.J., Kremer, M.J., Luo, M., *et al.*, 1986. The site of attachment in human rhinovirus 14 for antiviral agents that inhibit uncoating. *Science* 233, 1286–1293.
- Smyth, M., Tate, J., Hoey, E., *et al.*, 1995. Implications for viral uncoating from the structure of bovine enterovirus. *Nature Structural Biology* 2, 224–231.
- Solovvey, A.G., Makarov, V.V., 2016. Helical capsids of plant viruses: Architecture with structural lability. *Journal of General Virology* 97, 1739–1754.
- Stanley, W.M., 1935. Isolation of a crystalline protein possessing the properties of tobacco-mosaic virus. *Science* 81, 644–645.
- Staunton, D.E., Merluzzi, V.J., Rothlein, R., *et al.*, 1989. A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. *Cell* 56, 849–853.
- Stubbs, G., Warren, S., Holmes, K., 1977. Structure of RNA and RNA binding site in tobacco mosaic virus from 4 Å map calculated from X-ray fibre diagrams. *Nature* 267, 216–221.
- Subramaniam, S., Bartesaghi, A., Liu, J., Bennett, A.E., Sougrat, R., 2007. Electron tomography of viruses. *Current Opinion in Structural Biology* 17, 596–602.
- Suhanovsky, M.M., Teschke, C.M., 2015. Nature's favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* 479–480, 487–497.
- Sun, S., Xiang, Y., Akahata, W., *et al.*, 2013b. Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *eLife* 2, e00435.
- Sun, L., Young, L.N., Boudko, S.P., *et al.*, 2013a. Icosahedral bacteriophage FX174 forms a tail for DNA transport during infection. *Nature* 505, 432–435.
- Therkelsen, M.D., Klose, T., Vago, F., *et al.*, 2019. Flaviviruses have imperfect icosahedral symmetry. *Proceedings of the National Academy of Sciences of the United States of America* 115, 11608–11612.
- Thomassen, E., Gielen, G., Schutz, M., *et al.*, 2003. The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *Journal of Molecular Biology* 331, 361–373.
- Tong, L., Choi, H.K., Minor, W., Rossmann, M.G., 1992. The structure determination of Sindbis virus core protein using isomorphous replacement and molecular replacement averaging between two crystal forms. *Acta Crystallographica Section A* 48 (Pt 4), 430–442.
- Tong, L., Rossmann, M.G., 1990. The locked rotation function. *Acta Crystallographica Section A* 46, 783–792.
- Tong, L., Wengler, G., Rossmann, M.G., 1993. Refined structure of Sindbis virus core protein and comparison with other chymotrypsin-like serine proteinase structures. *Journal of Molecular Biology* 230, 228–247.
- Tsao, J., Chapman, M.S., Agbandje, M., *et al.*, 1991. The three-dimensional structure of canine parvovirus and its functional implications. *Science* 251, 1456–1464.
- Tsao, J., Chapman, M.S., Wu, H., *et al.*, 1992. Structure determination of monoclinic canine parvovirus. *Acta Crystallographica Section B* 48, 75–88.
- van Raaij, M.J., Schoehn, G., Burda, M.R., Miller, S., 2001. Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *Journal of Molecular Biology* 314, 1137–1146.
- von Bonsdorff, C.H., Harrison, S.C., 1975. Sindbis virus glycoproteins form a regular icosahedral surface lattice. *Journal of Virology* 16, 141–145.
- von Bonsdorff, C.H., Harrison, S.C., 1978. Hexagonal glycoprotein arrays from Sindbis virus membranes. *Journal of Virology* 28, 578–583.
- Voss, J.E., Vaney, M.-C., Duquerry, S., *et al.*, 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468, 709–712.
- Walters, R.W., Agbandje-McKenna, M., Bowman, V.D., *et al.*, 2004. Structure of adeno-associated virus serotype 5. *Journal of Virology* 78, 3361–3371.
- Watson, J.D., Crick, F.H.C., 1953a. Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171, 964–967.
- Watson, J.D., Crick, F.H.C., 1953b. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 171, 737–738.
- Wikoff, W.R., Liljas, L., Duda, R.L., *et al.*, 2000. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science* 289, 2129–2133.
- Williams, R.C., Smith, K.M., 1958. The polyhedral form of the Tipula iridescent virus. *Biochimica et Biophysica Acta* 28, 464–469.
- Winkler, F.K., Schutt, C.E., Harrison, S.C., Bricogne, G., 1977. Tomato bushy stunt virus at 5.5-Å resolution. *Nature* 265, 509–513.
- Wrigley, N.G., 1969. An electron microscope study of the structure of Sericesthis iridescent virus. *Journal of General Virology* 5, 123–134.
- Wu, H., Rossmann, M.G., 1993. The canine parvovirus empty capsid structure. *Journal of Molecular Biology* 233, 231–244.
- Xiao, C., Chipman, P.R., Battisti, A.J., *et al.*, 2005. Cryo-electron microscopy of the giant Mimivirus. *Journal of Molecular Biology* 353, 493–496.
- Xiao, C., Fischer, M.G., Bolotaulo, D.M., *et al.*, 2017. Cryo-EM reconstruction of the Cafeteria roenbergensis virus capsid suggests novel assembly pathway for giant viruses. *Scientific Reports* 7, 5484.
- Xiao, C., Kuznetsov, Y.G., Sun, S., *et al.*, 2009. Structural studies of the giant Mimivirus. *PLOS Biology* 7, e1000092.
- Xiao, C., McKinlay, M.A., Rossmann, M.G., 2011. Design of capsid-binding antiviral agents against human rhinoviruses. In: Agbandje-McKenna, M., McKenna, R. (Eds.), *RSC Biomolecular Sciences Series*, No. 21, Structural Virology. London, England: Royal Society of Chemistry, pp. 321–339.
- Xie, Q., Bu, W., Bhatia, S., *et al.*, 2002. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proceedings of the National Academy of Sciences of the United States of America* 99, 10405–10410.
- Xie, Q., Chapman, M.S., 1996. Canine parvovirus capsid structure, analyzed at 2.9 Å resolution. *Journal of Molecular Biology* 264, 497–520.
- Yan, X., Chipman, P.R., Castberg, T., Bratbak, G., Baker, T.S., 2005. The marine algal virus PpV01 has an icosahedral capsid with T=219 quasi-symmetry. *Journal of Virology* 79, 9236–9243.
- Yan, X., Olson, N.H., Van Etten, J.L., *et al.*, 2000. Structure and assembly of large lipid-containing dsDNA viruses. *Nature Structural Biology* 7, 101–103.
- Yan, X., Yu, Z., Zhang, P., *et al.*, 2009. The capsid proteins of a large, icosahedral dsDNA virus. *Journal of Molecular Biology* 385, 1287–1299.
- Zhang, R., Hryc, C.F., Cong, Y., *et al.*, 2011a. 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. *EMBO Journal* 30, 3854–3863.
- Zhang, W., Kaufmann, B., Chipman, P.R., Kuhn, R.J., Rossmann, M.G., 2013. Membrane curvature in flaviviruses. *Journal of Structural Biology* 183, 86–94.
- Zhang, W., Mukhopadhyay, S., Pletnev, S.V., *et al.*, 2002. Placement of the structural proteins in Sindbis virus. *Journal of Virology* 76, 11645–11658.
- Zhang, X., Xiang, Y., Dunigan, D.D., *et al.*, 2011b. Three-dimensional structure and function of the *Paramecium bursaria* chlorella virus capsid. *Proceedings of the National Academy of Sciences of the United States of America* 108, 14837–14842.

Further Reading

- Bernal, R.A., Hafenstein, S., Olson, N.H., *et al.*, 2003. Structural studies of bacteriophage a3 assembly. *Journal of Molecular Biology* 325, 11–24.
- Chiu, W., Burnett, R.M., Garcea, R.L., 1997. *Structural Biology of Viruses*. Oxford University Press.
- Fokine, A., Zhang, Z., Kanamaru, S., *et al.*, 2013. The molecular architecture of the bacteriophage T4 neck. *Journal of Molecular Biology* 425, 1731–1744.
- Horne, R.W., 1974. *Virus Structure*. New York and London: Academic press.
- Mateu, M.G., 2013. *Structure and Physics of Viruses*. Heidelberg, New York, London: Springer Dodrecht.
- Nermut, M.V., Steven, A.C., 1987. *Animal Virus Structure*. Amsterdam, New York: Elsevier Academic Press.
- Okamoto, K., Miyazaki, N., Song, C.H., *et al.*, 2017. Structural variability and complexity of the giant Pithovirus sibericum particle revealed by high-voltage electron cryo-tomography and energy-filtered electron cryo-microscopy. *Scientific Reports* 7.

- Plevka, P., Lim, P.Y., Perera, R., *et al.*, 2014. Neutralizing antibodies can initiate genome release from human enterovirus 71. *Proceedings of the National Academy of Sciences of the United States of America* 111, 2134–2139.
- Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F., Leiman, P.G., 2004. The bacteriophage T4 DNA injection machine. *Current Opinion in Structural Biology* 14, 171–180.
- Roy, P., 2005. *Virus Structure and Assembly*. California: Elsevier Academic Press.
- Yap, M.L., Klose, T., Urakami, A., *et al.*, 2017. Structural studies of Chikungunya virus maturation. *Proceedings of the National Academy of Sciences of the United States of America* 114, 13703–13707.

Relevant Websites

<https://pdb101.rcsb.org/browse/viruses>

Browse: Viruses.

PDB-101.

<https://youtu.be/KoJWuWzVgqQ>

Stephen Harrison (Harvard) Part 1: Virus structures: General principles.

<https://www.ebi.ac.uk/pdbe/emdb/>

The Electron Microscopy Data Bank (EMDB) at PDBe.

<https://viralzone.expasy.org>

ViralZone root.

<https://www.viprbrc.org/brc/home.spg?decorator=vipr>

Virus Pathogen Database and Analysis Resource.

<http://viperdb.scripps.edu>

Welcome to VIPERdb.

Scripps Research.

Structures of Small Icosahedral Viruses

Elizabeth E Fry, Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Jingshan Ren, University of Oxford, Oxford, United Kingdom

Claudine Porta, The Pirbright Institute, Pirbright, United Kingdom and University of Oxford, Oxford, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Icosahedral asymmetric unit The smallest part of the structure from which the complete structure of the virus can be built using a specific set of 60 rotational matrices that describe the 5:3:2 symmetry of the virus particle.

ICTV International Committee on Taxonomy of Viruses.

Quasi equivalence The interacting environment between subunits in an icosahedron with triangulation number

greater than 1 cannot be strictly equivalent, thus they are referred to as “quasi-equivalent” subunits.

SELEX A combinatorial chemistry technique in molecular biology for producing oligonucleotides of either single-stranded DNA or RNA that specifically bind to a target ligand or ligands and are commonly referred to as aptamers.

Triangulation number The number of facets (unique environments that subunits occupy) per triangular face of an icosahedron.

Introduction

Basic Structure

The proteins which form a protective shell around the viral genome (capsid) of a non-enveloped virus are multifunctional. The capsid must protect the genome integrity and release it into the target host cell, furthermore capsids may, depending on the host, mediate systemic transport across the host or between cells as well as host gene regulation and receptor-mediated endocytosis. In addition, for viruses of vertebrates the capsid outer surface is also the face the virus presents to the humoral immune system and largely defines its antigenic properties. Predicted in the early 1950s by Crick and Watson, the innate icosahedral symmetry of spherical capsids was established in the late 1950s when Don Caspar, Rosalind Franklin, John Finch and Aaron Klug obtained X-ray diffraction patterns of spherical plant viruses showing the presence of 5-fold symmetry. This was confirmed by the first high resolution structures of the small single-stranded (ss)RNA spherical plant viruses tomato bushy stunt virus (TBSV) and southern bean mosaic virus (SBMV), in 1979 and 1980 respectively.

An icosahedron has 12 vertices with 5-fold rotational symmetry, 20 triangular faces with 3-fold symmetry and 30 edges with 2-fold symmetry allowing the placement of 60 identical sub-units. Caspar and Klug proposed that the icosahedral asymmetric unit may be further triangulated to accommodate more subunits. The level of triangulation is denoted by a T number where the number of subunits is 60T. TBSV and SBMV adopted $T = 3$ architecture with 180 copies of the capsid protein forming three quasi-equivalent subunits (A, B, C) per icosahedral asymmetric unit. This gives rise to A/B dimers on the quasi 2-fold axes (around the 5-fold) which are bent and C/C dimers on the 2-fold axes which are flat, together generating the capsid curvature from what would otherwise amount to flat hexagonal close-packing. The dimer conformation is mediated by a conformational change in the flexible N-terminal arm which extends from the subunit core. The subunit core adopts a canonical eight-stranded jelly-roll β -barrel motif comprising 2 twisted β -sheets of 4 anti-parallel strands. This structure is pervasive across very many small viruses, and is often referred to as the viral jelly-roll. The strands of the two sheets are labeled (according to convention) alphabetically, following the amino-acid sequence from the N- to C-terminus as CHEF and BIDG; for the latter, strand labeling starts at B since TBSV has an additional strand outside the core fold. In the spherical capsid, the amino termini are on the inside, and particularly in plant viruses possess an N-terminal arginine rich motif which interacts with the negatively charged genome, the jelly-roll core lies parallel to the surface of the capsid, whilst the carboxy-termini are on the outside.

Picornaviruses were the first human viruses for which atomic resolution structures were obtained. Published in 1985, the structures of human rhinovirus 14 and poliovirus were a landmark in virology, being not only similar to each other but instantly perceived to have many core features in common with the plant virus structures, with the viral subunits comprised of a core jelly-roll domain. In these structures each of the 60 icosahedral ‘triangular’ subunits are sub-divided into 3, corresponding to the triangulation number $T = 3$, in exactly the same way as the plant viruses, but as each sub-triangle now contains a different protein (VP1–3) this is referred to as pseudo $T = 3$. The similarities in topology, detailed positioning of secondary structural elements and overall architecture led to the realization that structure was a more sensitive indication of evolutionary origin in viral proteins than their sequence. The structure of the insect virus, cricket paralysis virus provided a link between the plant and vertebrate viruses and confirmed evolutionary divergence. This underpins the basis for a higher-order classification of viruses and has led to the concept of a small ssRNA virus lineage ([Table 1](#)).

Diversity

The small ssRNA viruses infect organisms from different kingdoms but the same domain of life (Eukarya). The order *Picornavirales* classified according to the ICTV ([Table 1](#)) forms the major subset of this lineage. Members are largely non-enveloped (although

Table 1 Family members assigned by the ICTV for the Order *Picornavirales* and those assigned to a Picorna-like lineage by structure-based phylogeny

ICTV 2019 Order <i>Picornavirales</i>	Structure based Picorna-like lineage
<i>Caliciviridae</i>	<i>Birnaviridae</i>
<i>Dicistroviridae</i>	<i>Bromoviridae</i>
<i>Iflaviridae</i>	<i>Caliciviridae</i>
<i>Marnaviridae</i>	<i>Comoviridae</i>
<i>Picornaviridae</i>	<i>Dicistroviridae</i>
<i>Polycipiviridae</i>	<i>Luteoviridae</i>
<i>Secoviridae</i> = <i>Comoviridae</i> + <i>Sequiviridae</i>	<i>Sequiviridae</i>
<i>Soliniviridae</i>	<i>Microviridae</i>
	<i>Nodaviridae</i>
	<i>Papillomaviridae</i>
	<i>Parvoviridae</i>
	<i>Picornaviridae</i>
	<i>Polyomaviridae</i>
	<i>Tetraviridae</i>
	<i>Tombusviridae</i>
	<i>Tymoviridae</i>

picornaviruses hepatitis A and E are quasi-enveloped, usually exiting the cell in vesicles), icosahedral, have diameters ~ 300 Å and pseudo $T = 3$ architecture. These viruses have one or a few positive sense ssRNA(s) genomes usually without overlapping open reading frames although the latter does not hold true for *Caliciviridae* and new families infecting arthropods (*Polycipiviridae* and *Soliniviridae*).

A major characteristic of this order includes the presence of a replication block comprised of 3 elements: a superfamily III helicase, a chymotrypsin-like proteinase and an RNA-dependent RNA polymerase. Whilst in the families *Picornaviridae* and *Iflaviridae*, the structural proteins are in the N-terminal region and replicative proteins in the C-terminal region of the polyprotein, the genomic organization is reversed in *Marnaviridae* (monocistronic) and *Dicistroviridae*, (dicistronic). In the family *Secoviridae* the replication block is either at the N-terminus of a monopartite genome or on a separate strand in a bipartite genome.

The topology of the capsid subunits is similar in all cases, with the termini forming a network of interactions maintaining particle integrity. However, there are distinctions in the assembly/disassembly pathways. Variations reflect different infection mechanisms e.g., via insect vectors or means to evade immune surveillance in animal hosts.

To derive phylogenies a similarity matrix is usually constructed from available sequences but in the case of the picornavirales which have single-stranded RNA(s) serving directly as mRNA the sequences change rapidly due to the error-prone replication mechanism and the process of selection. An alternative method for classification is defined by the Linnean approach of structure-based homology. The protein tertiary and quaternary structure being more constrained than its sequence, remains highly conserved despite substantial sequence variation and provides a complementary approach to deciphering the evolution of viruses. In a structure-based phylogenetic analysis, a similarity matrix is constructed based on pairwise superimposition of the various structures, with evolutionary distance being derived from the root mean square deviation between the matched residues divided by the fraction of the structure that is matched.

The structure of tobacco ringspot virus (TRSV), a plant nepovirus classified in subfamily *Comovirinae* (formerly family *Comoviridae*) of family *Secoviridae* revealed that the capsid protein comprises three β -barrel domains covalently joined by extended polypeptides reflecting the pre-cleaved states of the polyprotein of comoviruses which is cleaved into 2 capsid proteins and that of picornaviruses which is cut into 3 distinct coat proteins: this further confirms the notion of divergent evolution of the capsid polypeptides of nepoviruses, comoviruses and picornaviruses from a common ancestor. Structural similarity is highest with the cowpea mosaic virus (CPMV), another comovirus. CPMV particles are made up of a small protein equivalent to the picornavirus VP1 and a large coat protein corresponding to VP2 and VP3 joined into a single polypeptide.

Structure-based phylogenetic analyses have allowed assignment of further virus families to the single jelly-roll lineage that reaches beyond ssRNA + sense icosahedral viruses. This allowed the inclusion of the family *Birnaviridae* (capsids with $T = 14$ symmetry), *Bromoviridae* (3 genomic and 1 subgenomic ssRNA), *Luteoviridae* (ribosomal frameshifting expression strategy), *Microviridae* (ssDNA genome), *Nodaviridae* (bipartite ssRNA and a subgenomic RNA), *Papillomaviridae* (circular dsDNA and $T = 7$ dextro (right hand-skewed) symmetry capsids), *Parvoviridae* (ssDNA either + or -), *Polyomaviridae* (circular dsDNA, $T = 7$ d symmetry capsids), *Tetraviridae* (capsids with $T = 4$ symmetry), *Tombusviridae* (2 genomic and 2 subgenomic ssRNAs) and *Tymoviridae* (1 genomic and 1 subgenomic ssRNA) into the picorna-like lineage (Table 1).

Picornaviruses as an Exemplar Family

Picornaviruses (a family of vertebrate-infecting viruses with a positive ssRNA genome) have been well-studied structurally (Fig. 1) due to their importance to human and animal health. The largest genus within this family is that of the enteroviruses that use as

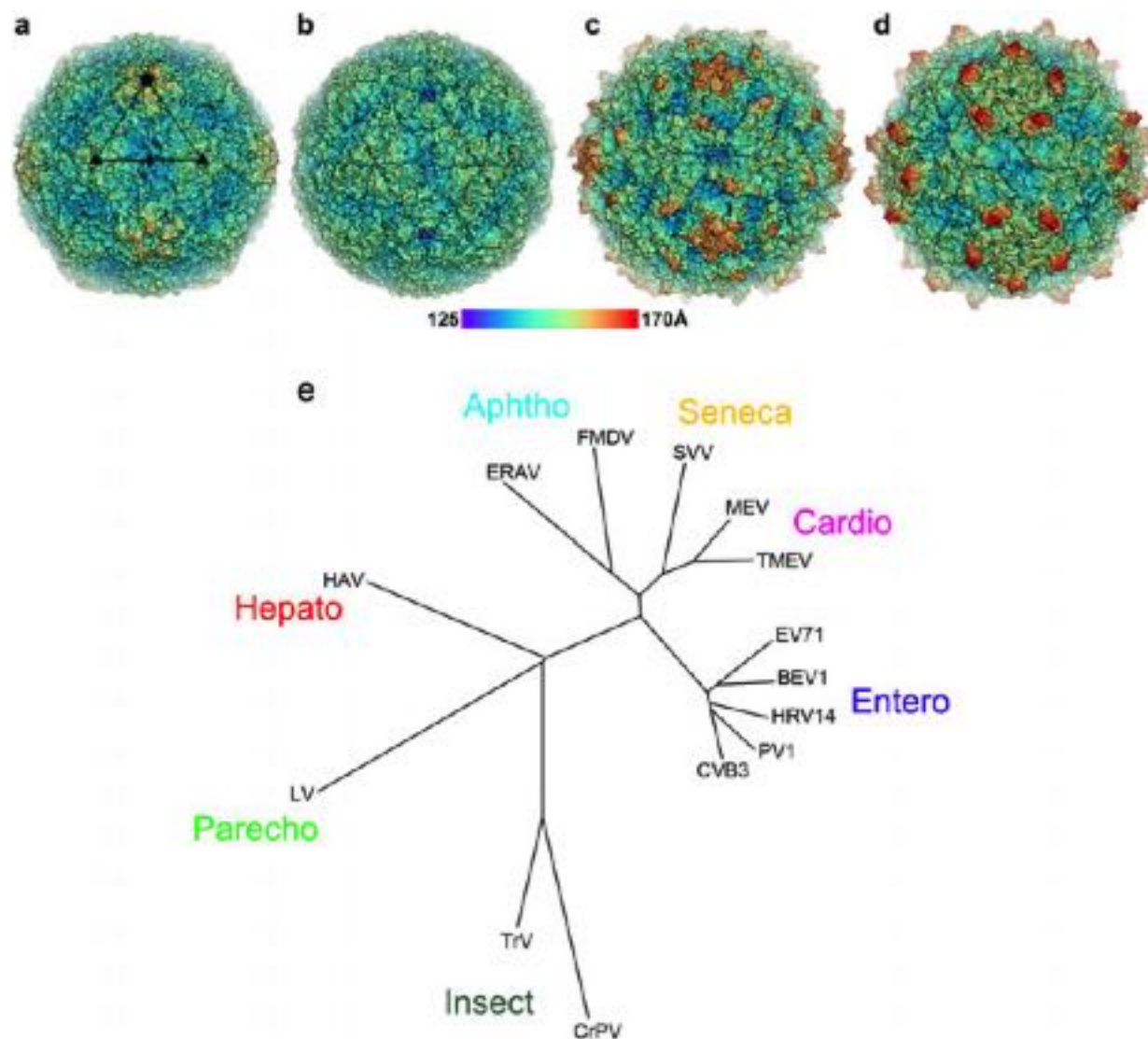


Fig. 1 Picorna-like virus capsids. a to d, HAV, FMDV, PV1 and TrV (PDB IDs, 4QPI, 1ZBA, 1HXS and 2NAP respectively). All capsids are shown in a surface representation colored radially as shown. An icosahedral asymmetric unit is outlined in black on a) with the 5-fold, 3-fold and 2-fold axes denoted by a pentagon, triangle and ellipse, respectively. e, Unrooted structure-based phylogenetic tree based on a selection of picornavirus structures representative of various genera from the PDB together with two insect picornaviruses (cripaviruses). CVB3, coxsackievirus B3; PV1, poliovirus type 1; HRV14, human rhinovirus 14 (RV-B14); BEV1, bovine enterovirus type 1 (EV-E1); EV71, enterovirus A 71 (EV-A71); TMEV, Theiler's murine encephalitis virus; MEV, mouse Elberfield virus also known as Mengo virus; SVV, Seneca valley virus; FMDV, foot-and-mouth disease virus; ERAV, equine rhinitis A virus; HAV, hepatitis A virus; AiV, Aichi virus; TrV, triatoma virus; CrPV, cricket paralysis virus.

the primary site of replication either the gastrointestinal tract (e.g., polio-, echo-, and coxsackieviruses), or the upper respiratory tract (rhinoviruses). Of the other genera, aphthoviruses include the causative agent of foot-and-mouth disease (FMDV) which infects split-hoofed animals, the genus name being derived from the Greek *aphtha*- meaning vesicles in the mouth. Cardioviruses cause myocarditis and encephalomyelitis in rodents which are considered to be their natural reservoir but their host range includes elephants, monkeys, pigs and humans. Hepatoviruses replicate in the liver of mammals and hepatitis A virus (HAV) the prototype of this genus is the only species that affects humans and primates. Parechoviruses and kobuviruses are more recently established genera; the first includes two species: Ljungan virus (LV), a virus of rodents and the human parechoviruses whilst the second is associated with gastroenteritis in humans, cattle, swine, sheep, goats, dogs, cats and mice with the genus deriving its name from the lumpy appearance of virus particles in electron microscopy; “kobu” meaning “knob” in Japanese. Overall enteroviruses are the most elaborate and best studied picornaviruses whilst hepatoviruses are likely to be much closer to the last common ancestor of picornaviruses.

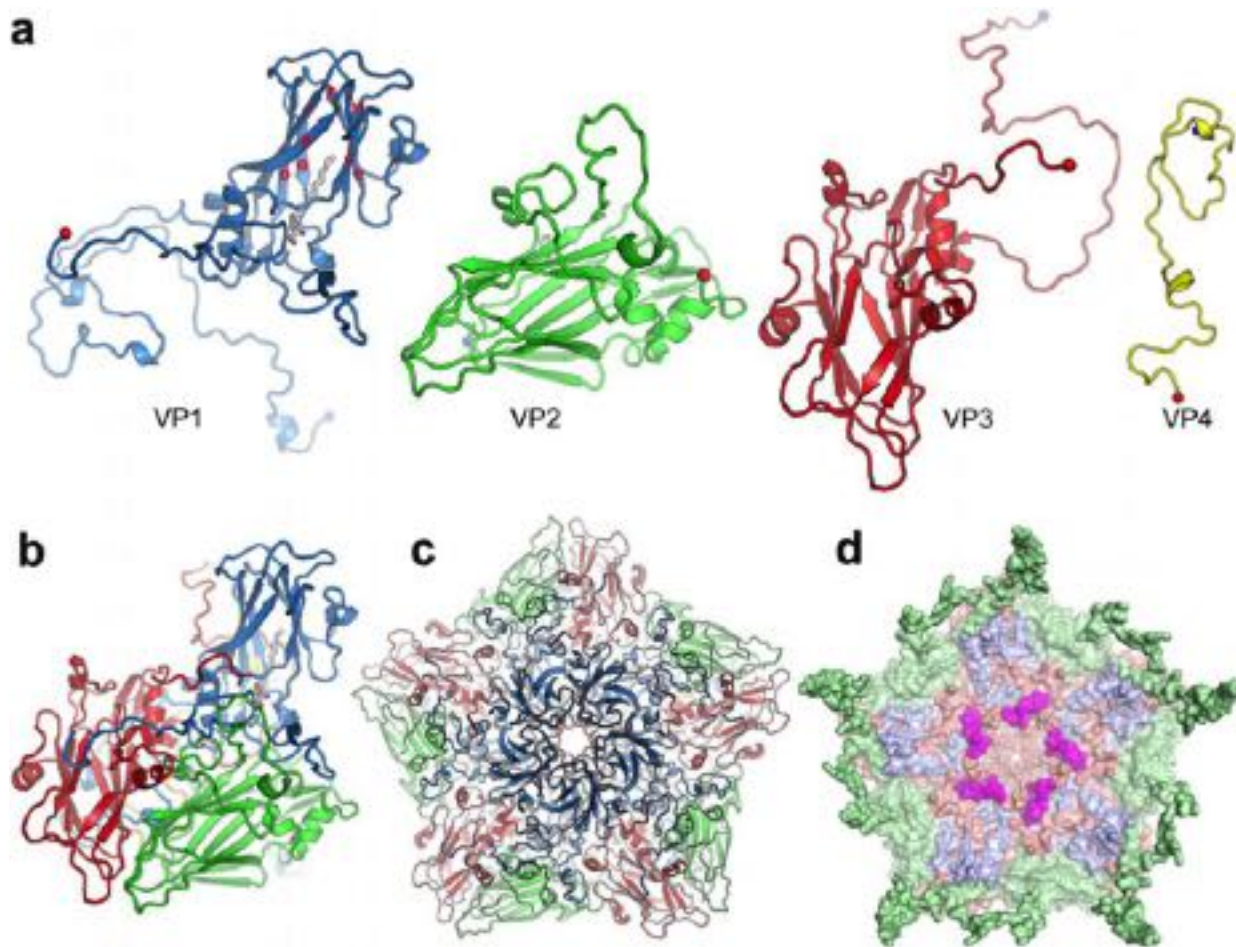


Fig. 2 Picornavirus assembly. a, Capsid proteins of EV-A71 VP1–4 (PDB ID, 3VBH). The proteins are drawn as cartoons, with conventional protein coloring, to show the secondary structure elements in particular the β -barrel fold with the strands labeled CHEF and BIDG on VP1. The N- and C-termini are marked with blue and red spheres respectively. A lipid ‘pocket-factor’ is shown (in grey) bound within the hydrophobic pocket of VP1 b, The biological protomeric subunit (proteins depicted as in (a)). c, The pentameric subunit. d, A parechovirus pentamer drawn in surface representation with the proteins color coded as for EV-A71, viewed from the inside. Ordered RNA genome is shown in magenta around the 5-fold axis (PDB ID, 5MJV).

Capsid Assembly

The picornaviral positive-sense ssRNA genome (approx. 8K bases) is translated as a polyprotein and cleaved into the structural and replication proteins by encoded proteases. The polyprotein comprises three functional regions: P1, P2 and P3 with P3 proteins participating directly in genome replication and P2 proteins contributing to the host cell interactions required for virus multiplication, whilst the structural proteins form the P1 section at the 5′ end of the open reading frame. The polyprotein is divided into 1ABCD-2ABC-3ABCD units, where the numbers refer to the functional region and the letters to individual proteins. P1 adopts a three-dimensional fold prior to assembly (although no structure has thus far been determined for a picornaviral P1). This is then cleaved into VP0, VP3 and VP1 by the 3C protease (or 3CD precursor) to form 5S protomers. The cellular chaperone Hsp90 has been shown to be required for this processing in both entero- and aphthoviruses. The released protein termini associate to link five copies of each 5S protomer into a $\sim 14S$ pentamer and 12 such pentamers assemble to form a spherical shell 27–30 nm diameter (Fig. 2). Once the RNA has been encapsidated, VP0 undergoes an additional cleavage to produce VP4 (N-terminal) and VP2 to form mature virus particles.

VP1 proteins abut around the icosahedral 5-fold, and VP2 and 3 around the 2- and 3-fold axes whilst VP4 is internal. VP1–3 adopt the viral jelly-roll fold with the loops connecting the strands contributing to the external antigenic surface. Across the picornavirus family VP1–3 vary in size from approx. 24 kDa in FMDV to approx. 30 kDa in many other picornaviruses due to longer surface loops in the latter and more surface structure. This also results in variation in the capsid thickness between 33 Å for FMDV (excluding VP4) and 42.5 Å for human rhinovirus 14 (HRV14). In particular VP1 dominates the accessible surface forming extended loops around the 5-fold axis and in enteroviruses binds a lipid (pocket-factor) within a hydrophobic pocket in the β -barrel.

VP4 is much smaller than VP1–3 and has no core structure, rather it contributes to the extensive protein network on the inside of the capsid. The N-termini of VP1–3 also contribute to this network, knitting the protein cores together, often forming scaffold structures: they intertwine around the 5- and 3-fold axes with VP3 proteins contributing most towards the stability of the pentamer, through their N-termini associating around the 5-fold axes to form β -annuli. Extensions of VP4, VP2 (and VP1 in enteroviruses) contribute to stabilizing the interface between pentameric subunits by forming an extended β -sheet. In aphthoviruses this comprises six strands, the VP2 N-terminal hairpin (strands β A1 and β A2) of one pentamer together with the CHEF sheet of the β -barrel of a VP3 from an adjacent pentamer. Enteroviruses elaborate this further - the N- and C-termini of VP1 from the same protomer as the 'core' CHEF strands of VP3, sandwich the VP2 N-terminal hairpin from the adjacent pentamer. In parechoviruses a domain swap in VP2 sees this hairpin form on the two-fold related subunit instead of the VP2 from which it derives. This may explain why enteroviruses do not dissociate so readily into pentameric subunits but tend to form modified particles whilst for cardio and aphthoviruses, capsid integrity is vulnerable to low pH. The extreme lability of aphthovirus capsids also correlates with a greater number of histidines ($pK_a \sim 6$) at the pentamer interface.

The myristylation of the N-terminus of VP4 is a feature almost entirely conserved in picornaviruses (barring hepatoviruses and parechoviruses). In poliovirus, a hydrogen bond has been observed between the myristate carbonyl and the hydroxyl side chain of threonine 28 of VP4 (not conserved across all picornaviruses) between two 5-fold symmetry-related copies of VP4, thereby stabilizing the pentameric interaction. Ions have been visualized on the 3- and 5-fold icosahedral symmetry axes of a number of picornaviruses e.g., the DE loop (between β -strands D and E) of VP1 is involved in the interaction with a putative Ca^{2+} ion on the viral 5-fold axes in all analyzed rhinoviruses, which may play a role in regulation of stability although no conformational changes were observed in EGTA (chelating agent) treated virus structures.

In hepatitis A virus (HAV), 3 structural proteins are involved in capsid assembly: VP0, VP3 and VP1pX. pX a C-terminal extension of VP1 is required for efficient structural protein processing and virion morphogenesis is cleaved off by host proteases trypsin and cathepsin L in vitro to produce mature capsids. This is mirrored in CPMV a picornavirus-like plant virus, where the C-terminus of the protein equivalent to VP1 (called S protein) has been demonstrated to be required for RNA encapsidation and is also subsequently shed from mature capsids. Cryo-EM showed that the 24 amino acid peptide occupies the cleft between S protein subunits in the pentameric capsomere and stabilizes its structure by functioning as a dab of molecular glue: it acts as a molecular chaperone, that enhances the efficiency of folding/assembly but does not remain part of the final structure. In the case of HAV which is released in a quasi-enveloped form (eHAV) pX has additionally been shown to promote the secretion of virions through exosome-like vesicles by acting as a sorting signal for interaction with the host endosomal sorting complexes required for transport (ESCRT).

Genome Encapsidation

The picornavirus ssRNA genome (between 7.1 and 9.7 kb) is around 2.5 μ m in length and tightly packaged within the capsid (at a concentration of ~ 700 mg/ml). The precise details of genome encapsidation in picornaviruses are yet to be unraveled but variation is seen across the family. Whereas in most viruses no more than a few well-ordered RNA bases have been visualized in association with mature capsids (e.g., stacking with VP4 residues Tyr 20 and Phe 46 in poliovirus type 2 Lansing), the structures of the parechoviruses: human parechovirus 1 (HPeV1) and LV, revealed a stretch of relatively well-ordered RNA associated with the N-termini of VP3 (15 bases in LV) near the five-fold axes (Fig. 2). The VP3 N-termini of these viruses are uniquely rich in positively charged residues (arginine/lysine-rich motif) and a passive genome encapsidation mechanism has been proposed whereby VP3 acts as an RNA chaperone, organizing the genome into a condensed structure compatible with the icosahedral symmetry of the particle through numerous interactions with short RNA segments with pre-formed stem-loop structures (although largely electrostatic there is also specificity in these interactions). Such a mechanism would appear to be at variance with that put forward for enterovirus packaging which for poliovirus involves interactions between capsid and non-structural proteins in an RNA replication complex, but is not inconsistent with the observation of a 5'-terminal RNA stem-loop packaging signal (PS) in Aichi virus (genus Kobuvirus) and with that of a series of 5 degenerate RNA binding motifs in the 5'-terminal fragment of the plant virus, satellite tobacco necrosis virus. In the case of poliovirus, no PS could be identified in either the 5'- and 3'-non-translated regions, P1 or P2 regions of the genome, although rather weak PSs dispersed through the genome have not been ruled out.

The classical PS hypothesis assumes that a single PS site with affinity for cognate CP forms an RNA-CP assembly initiation complex. Through SELEX and bioinformatics analyses a series of up to 60 palindromic sequences could be identified throughout the genome of HPeV1 that bind to sites on the inside of VP3 and these are proposed to help direct assembly in a pathway mediated by the 2C ATPase, which can also act as a viral helicase. In vitro assembly studies showed that the CP does not bind spontaneously to RNA; the latter needs to actively fold into an assembly-competent conformation first. This could be achieved during replication when local secondary structures are favored as incomplete transcripts emerge from viral polymerases and in some picornaviruses, only nascent genomic transcripts are encapsidated. A complementary role for 2C would consist of RNA chaperoning and generation of the PSs. In summary whilst genome packaging remains poorly understood, for most picornaviruses it likely combines the use of dispersed PSs with specific protein chaperone interactions at the replication complex to achieve rapid genome collapse and encapsidation.

Many picornaviruses also assemble empty particles as part of a natural infection; these so-called natural empties are thought to be either a storage form of pentamers or a dead-end by-product of virion morphogenesis. Notably, natural empty particles are not

observed for parechoviruses which is in line with a distinct assembly mechanism. Recombinant empty particles, also called virus-like particles (VLPs) can be generated notably by co-expression of the capsid precursor polyprotein and the viral protease in heterologous systems and this is the focus of much research destined to produce new or safer alternatives to existing vaccines against e.g., FMDV, poliovirus and coxsackie viruses. One of the challenges is to compensate for the absence of internal RNA 'glue' and/or lack of VP0 processing by empirically discovered or rationally designed capsid protein mutations aimed at increasing VLP stability.

The maturation cleavage of VP0 into VP4 and VP2 upon RNA encapsidation occurs in all picornaviruses except kobuviruses and parechoviruses which may be related to the presence of ordered RNA density. This enigmatic cleavage is still assumed to be autocatalytic, being inaccessible to exogenous proteases. The subsequent rearrangement of the VP4 C-terminus and VP2 N-terminus confers stability on the particle and empty capsids containing VP0 are less resistant to heat and pH. The structure of natural empty capsids of poliovirus 1/Mahoney shows rearranging or disordering of the network on the inner surface. The VP0 cleavage site is located some 21 Å from the location observed in virions, the scissile bond being adjacent to a hydrophobic depression on the inner surface analogous to the binding site for icosahedrally ordered segments of the genomic RNA in the plant virus beanpod mottle virus. A possible mechanistic model for VP0 cleavage involves a conserved histidine residue located close to the VP0 cleavage site. Histidine 195 of VP2 (2195H) is hypothesized to activate local water molecules, thus initiating a nucleophilic attack at the scissile bond. 2195H mutants were nonviable and analyses showed that these particles contain genomic RNA and uncleaved VP0. In poliovirus immature empty capsid where the cleavage of VP0 has not occurred, VP0 residues near the cleavage site prevent the N-terminus of VP1 from accessing its position in the mature particle where it would interact with the inner surface of VP2 and VP3 contributing to particle stability. In contrast, empty capsids of FMDV type A22 Iraq 24/64 contain VP2 and VP4. The disordering of the internal network in these FMDV empty capsids is less extensive than that seen in the poliovirus empty capsid, which has VP0 intact. Thus, VP0 cleavage confers stability on the picornavirus capsid over and above that attributable to RNA encapsidation. His 2145 in FMDV VP2 appears to be analogous to His 2195 in poliovirus and similarities in the putative active site support the cleavage hypothesis. This final stage in assembly is thought to establish a metastable state, priming the particle to initiate the entry process when receptor interactions and/or reduced pH trigger the conformational transition to a lower energy state.

Host Interactions

Virus capsids recognize susceptible cells by attachment to specific receptors on the plasma membrane thereby determining the host range and tropism of infection. Although many cellular receptors have been identified and structural analyses of virus-receptor complexes have provided insight into the initial stages of this interaction, subsequent events remain unclear (Fig. 3). Picornaviruses generally gain entry to the cytoplasm via an endocytic vesicle and often require complex receptor/co-receptor interactions to deliver them to appropriate sites on the cell surface for the induction of endocytosis.

Clues about receptor interactions came from the very first enterovirus structures. As one of the loops around the five-fold axis, the BC loop of VP1 forms a projecting ridge in enteroviruses. The depression beyond it encircles the five-fold axis and is bounded at the southern side by the GH loop and C-terminus of VP1. For HRV14, this approximately 12 Å depression was named the 'canyon' and hypothesized to correspond to the receptor binding site since its dimensions made it inaccessible to antibodies, suggesting a mechanism by which the virus could evade immune surveillance. Rhinoviruses are classified into major and minor

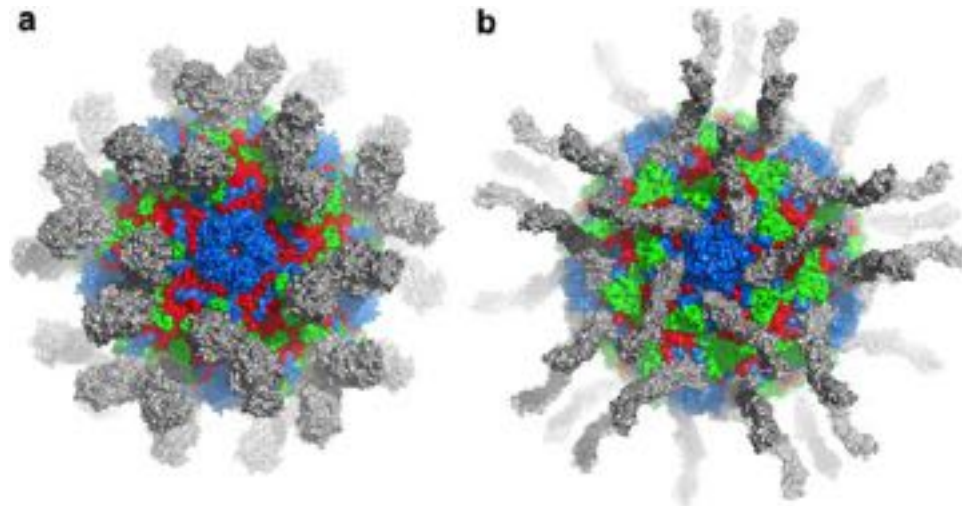


Fig. 3 Picornavirus receptor complexes. a, A surface representation of the EV-A71-SCARB2 (6I2K) complex with the capsid proteins colored as in Fig. 2. SCARB2 is shown in gray binding at the southern edge of the canyon depression and does not occlude the canyon. b, A similar view of the poliovirus-CD155 (3J8F) complex, note the receptor inserts into the canyon.

receptor-binding groups. Major group rhinoviruses e.g., HRV14 and HRV16 recognize the ICAM-1 receptor, and structural results show that it binds within the canyon. However, key viral amino acid residues involved in binding ICAM-1 are also accessible to antibodies; i.e., the receptor site is not effectively hidden and several antibody/virus complexes show antibodies binding across or partially within the canyon.

Similarly, the binding site of the poliovirus receptor (PVR), the nectin CD155 has been visualized in the canyon (Fig. 3). Cryo-electron microscopy and image reconstruction of receptor-poliovirus complexes reveal PVR binding to the 'wall' of surface protrusions surrounding the 'canyon' in a more tangential orientation than ICAM-1. The greater exposure of the receptor-binding sites in polioviruses compared to rhinoviruses makes them less protected from immune surveillance but well suited for receptor-mediated virus uncoating and entry at the cell surface.

The coxsackievirus B3-receptor (CAR; coxsackie and adenovirus receptor) also binds in the canyon, with the distal end of its N-terminal Ig-like domain, but its binding orientation relative to the viral surface is quite different. In the coxsackievirus B3-full length CAR complex the transmembrane and cytoplasmic regions of two adjacent CAR molecules related by icosahedral two-fold axes associate with each other which increases the receptor binding avidity to the virus without altering the orientation of the CAR domains on the viral surface. Coxsackie B viruses bind both decay-accelerating factor (DAF) and CAR and as their respective footprints on the viral surfaces do not overlap each other, they could be used simultaneously in cell entry. In contrast in the echovirus 7-DAF receptor complex, the DAF molecules bind across the icosahedral two-fold axes, not in the canyon. Since the binding of echovirus 7 to DAF is reversible, this is consistent with DAF being an initial attachment receptor triggering signal-dependent transport of the receptor-virus complex to CAR located in the tight junctions of epithelial cells so that binding and cell-entry can occur.

For enterovirus 71 (EV-A71) responsible for hand, foot-and-mouth disease (HFMD), two receptors have likewise been identified: human P-selectin glycoprotein ligand-1 (PSGL-1), a mucin-like protein that serves as a receptor for the infection of leukocytes and scavenger receptor class B member 2 (SCARB2; lysosomal integral membrane protein-2) which mediates uncoating at low-pH. A complex with the latter was obtained using a low-pH stabilized mutant of EV-A71 further stabilized using a drug candidate to replace the lipid pocket-factor. SCARB2 binds on the southern rim of the canyon rather than across the canyon as predicted. Whilst 12 of the 14 residues involved in receptor binding are conserved across EV-A71 genotypes, only 10 are conserved in other viruses that use SCARB2 as a receptor, likely arising from antigenic variation. Another enterovirus causing HFMD, Coxsackievirus A10 (CV-A10) recognizes Kringle-containing transmembrane protein-1 (KREMEN-1). This rather large molecule binds across the canyon (although it doesn't fully reach the canyon floor) spanning two adjacent asymmetric units via its KR (Kringle) and WSC (carbohydrate binding) domains. CV-A10 makes interactions using the VP1 EF and GH loops and C-terminus together with the VP2 EF loop. The residues involved in receptor binding are mostly conserved in Kremen-1 dependent enteroviruses suggesting a similar mode of binding. Echovirus 30 (E30) which causes human aseptic meningitis, utilizes DAF (CD55) and also FcRn (Neonatal Fc receptor) as attachment and uncoating receptors. Structural studies of complexes showed CD55 binding outside the canyon extending south (akin to EV-A71/SCARB2) with little change in the viral capsid. FcRn in contrast, bound within the canyon to the VP1 BC, VP1 GH and VP2 EF loops. Enteroviruses can be sub-divided on the basis of their capsid structure, in particular the canyon morphology which correlates with their receptor usage suggesting that receptor switching drives evolution in these viruses.

In cardiociruses, e.g., Mengovirus, a depression similar to the 'canyon' exists except that it is not continuous but broken into five deep pits located in the region homologous to the deepest portions of the rhinovirus canyon. The remainder of the canyon is filled by two insertions (loops I and II between β C and β D of VP1). Another cardiocirus, persistent Theilers Murine Encephalitis Virus (TMEV) uses sialic acid as a receptor moiety. Crystallographic analysis shows the receptor mimic sialylactose to bind to a negatively charged pocket on the viral surface composed mainly of residues of puff B from capsid protein VP2, in the vicinity of the VP1 loop II and VP3 C-terminus. The putative receptor binding site for cardiociruses, the 'pit' (analogous to the canyon) is about 15 Å from this sialic acid binding site.

In contrast to the rhinovirus major group, minor group viruses bind members of the LDL-receptor family including the very low-density lipoprotein receptor (VLDL-R). A 15 Å resolution cryo-electron microscopy analysis of human rhinovirus 2 (HRV2 or RV-A2) complexed with the first three (of eight) ligand-binding repeats of VLDL-R showed the receptor to bind to the star shaped dome on the icosahedral 5-fold axis rather than in the canyon. The footprint of the first two domains of VLDL-R on the viral surface (domain three does not contact the capsid) overlays the BC- and HI- loops of VP1. X-ray structures of HRV2/VLDL V3 concatamers demonstrated multi-modular binding around the five-fold axes. Receptor specificity is thought to arise from hydrophobic interactions between a conserved lysine on the virus and a tryptophan in the receptor modules together with coulombic attraction.

FMDV differs from the entero and cardiociruses in having a relatively smooth surface, with one exceptionally long exposed loop, the GH-loop of VP1 (residues 134–160) which not only constitutes a major antigenic site but also contains a conserved RGD (Arg–Gly–Asp) sequence which is essential for cell attachment via an integrin receptor, predominantly $\alpha_v\beta_6$. This loop is highly disordered in the native virus structure of all serotypes examined to date. When the disulfide bond that links the base of the loop in strain FMDV-O₁ (Cys134 to Cys130 of VP2) is reduced, the loop adopts a more highly ordered position lying along the viral surface. The fully open form of the integrin has been visualized in different poses attached to the VP1 GH loop which is extended away from the virus surface. The integrin binds almost perpendicular to the virus surface. Whilst the flexibility of the VP1 GH loop precludes a high resolution structure determination the binding can be inferred to occur via interactions with the RGD motif and

downstream hydrophobic residues. In addition, an N-linked sugar on the integrin attaches to a heparan sulfate binding site on the virus.

FMDVs repeatedly passaged in tissue culture adapt to use heparan sulfate, a glycosaminoglycan (GAG), as a receptor. Crystal structures of both serotype O1BFS and A1061 with this GAG show that it binds with high affinity by immobilizing a specific highly abundant motif of sulfated sugars in a shallow depression on the virion surface, located at the junction of the three major capsid proteins, VP1, 2 and 3. In both serotypes, residue 56 of VP3, an arginine, is critical to this recognition, switching from a histidine in field isolates in adaptation to tissue culture. In the infected animal there may be a biological advantage to low affinity, or more selective, interactions with GAG receptors.

Uncoating

There appear to be two modes of receptor interactions with picornaviruses (1) that exemplified by most enteroviruses where receptor binding facilitates, by a quasi-mechanical effect, the formation of an altered capsid which is thought to commit the virus to uncoating and (2) that exemplified by cardio- and aphthoviruses where the receptor interaction is more surface orientated and appears to be a simple attachment with low pH instigating dissociation. All picornaviral capsids are balanced between stability and instability - they must be sturdy enough to allow the virus to pass from host to host through a hostile environment (especially enteroviruses which infect via the gut) and able to readily uncoat and release their RNA upon cell entry.

In enteroviruses the interaction with receptor initiates irreversible structural changes starting with dislodging the fatty acid pocket factor and ultimately leading to the exit of the N-terminus of VP1 and VP4 to form the expanded intermediate or 'A-particle'. This appears to correspond to the dominant form of the virus found in cells early in infection and is thought to be a necessary transitional state in the cell entry process. This particle exhibits changes in sedimentation behavior (from 160S to 135S), antigenicity and protease sensitivity, has increased hydrophobicity and readily attaches to liposomes. The A-particle is endocytosed and at some point, engages fully with the vesicle membrane to deliver the RNA to the cell, leaving an empty 80S or B particle. In recent years knowledge of the structure of the 135S particle and its formation has advanced significantly. For example, in EV-A71 the virus VP1 GH and VP2 EF loops interact with helices $\alpha 5$ and $\alpha 7$ of receptor SCARB2, part of a helix bundle that undergoes low-pH dependent conformational changes. The GH loop of VP1 is proposed to act as a sensor in a sensor-adaptor uncoating mechanism for enteroviruses and it seems likely that a low-pH induced conformational change in the receptor acts as a trigger for this mechanism. Likewise, KREMEN-1 binding by CV-A10 at low pH elicits virus uncoating *in vitro* utilizing a similar mechanism where the receptor induces flexibility in the VP1 GH sensor loop facilitating pocket factor release and the structural re-arrangements leading to the formation of the 135S particle. This intermediate particle for CV-A16 has been captured in a crystal structure with the N-terminus of VP1 in egress from the particle (Fig. 4). As for other 135S particles, the capsid is expanded ($\sim 4\%$), the pocket factor absent and the VP1 β -barrel collapsed, VP4 and the N-terminus of VP1 are missing from the interior, the paired helices on the 2-fold axis each from a 2-fold related VP2 molecule are pulled apart opening a '2-fold channel' and a smaller channel forms close-by at the base of the canyon, the 'off-axis channel'. However, in this structure, the 'off-axis channel' is plugged by residues 71 to 65 of VP1 passing through the capsid with residue 62 the last seen on the particle surface, the first 61 residues assumed disordered. The 'off-axis channel' is formed by a conformational change in the GH-loop of VP3 from a helix in the mature capsid to a β -hairpin in the expanded state. The enlargement of the '2-fold channel' is achieved by re-orientation of side-chains triggered by reorganization of the C-terminus of VP2. Residues separating the two channels are also seen to become much less well-ordered. The RNA appears condensed and packed in layers, rather than evenly distributed as observed in mature virions.

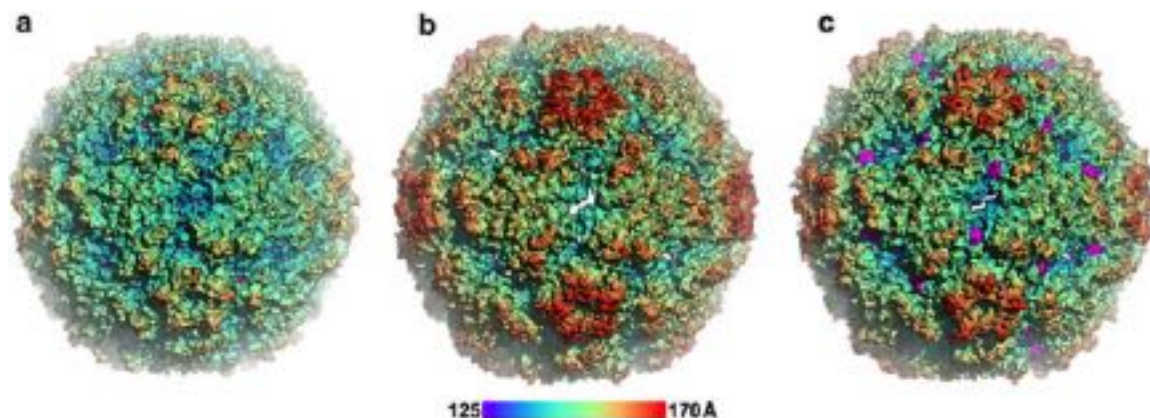


Fig. 4 Picornavirus uncoating. a to c, coxsackievirus A16 (CV-A16) virion (5C4W), expanded CV-A16 empty capsid (6LHC) and CV-A16 A-particle (4JGY) with extruded N-terminus colored in magenta. The capsids are depicted as in Fig. 1. The use of consistent radial coloring across all panels shows the capsid expansion.

There is a 15 Å gap between the capsid and RNA where the N-terminal protein network existed and bridging electron density is seen between the capsid and RNA apparently close to the N-terminal end of VP1 and N-terminus of VP2. The extruded segment of VP1 harbors an amphipathic helix and in conjunction with the myristoyl-VP4, is proposed to facilitate cell entry by forming a pore in, or disrupting, the endosomal membrane for egress of the RNA. At least a sub-set of the changes that occur on 135S conversion are reversible since there is evidence that mature poliovirus particles can 'breathe' leading transiently to the externalization of internal polypeptides, as captured by electron microscopy using an Fab directed against the N-terminus of VP1. For these studies, entry intermediates have been produced *in vitro*, for instance by incubating the virus at high temperatures or with excess solubilized receptor molecules. An alternative approach has been used to induce the coxsackievirus B3 entry intermediate by triggering a conformational change in the virus with full-length receptors embedded in lipid bilayer nanodiscs. Thus, rather than decorating a particle such that many or all of the receptor binding sites are occupied, asymmetrically decorated particles more akin to *in-vivo* cell-entry are formed. A-particles generated in this way have been reconstructed using cryo-electron microscopy with and without imposing icosahedral symmetry at 3.9 and 7.8 Å, respectively. In contrast to the A-particles described above, asymmetric receptor binding leads to minimal global capsid expansion but more concerted conformational changes at the site of receptor interaction with viral proteins extruded from the capsid only at this unique site adjacent to the nanodisc.

For cardio- and aphthoviruses, in contrast, there is no evidence for receptor-binding induced conformational changes and it has been thought that these viruses dissociate directly to pentamers at low pH. Disruption into pentamers under the influence of reduced pH in the endosome would render safe transfer of RNA to the cytoplasm very problematic. An apparent link to the enterovirus mechanism was revealed by the analysis of the aphthovirus equine rhinitis A virus (ERAV), which showed that acid-driven disassembly into pentamers proceeds via a quasi-stable 80S empty particle. This putative intermediate in the cell entry process is not a peculiarity of ERAV since intact, empty FMDV particles missing both RNA and VP4 have also been observed, at least transiently. A low pH ERAV crystal structure presumed to correspond to the 80S particle observed biochemically, shows changes restricted to internal features affecting particle stability e.g., there is the movement of the N-terminus of VP1 (corresponding to residues 44–56 in poliovirus) to the pentamer interface, displacing the N-terminus of VP2, which becomes disordered. The structure of a low pH form of the acid-dissociable Mengovirus, a cardiovirus, also shows very few conformational changes compared to the native structure: it appears that there are simply changes in the receptor binding 'pit' which would be consistent with a loss of receptor binding prior to direct involvement with the membrane. Loss of receptor binding has also been reported for enteroviruses HRV2 and poliovirus and may be a generic feature of picornavirus cell entry.

In poliovirus the final structural rearrangements to form the mature capsid involve similar structures to those externalized reversibly when the virus 'breathes' and irreversibly in receptor-mediated conformational rearrangements early in the entry process. That the changes in ERAV correlate strongly with this suggests that cleavage and reorganization to prime the virus for the conformational changes required for cell entry is a general principle in all picornaviruses. Also, the ability to eject the genome while maintaining icosahedral integrity is a feature common to both aphtho- and enteroviruses and suggests that there may be a general mechanism by which all picornaviruses protect their genome within intact capsids until the moment it is safely transported into the cytoplasm. However, there is still much to be learnt about the structural basis of cell entry, not least since none of the structures observed to date explain how a heavily structured RNA genome is transported intact, without expending chemical energy, into the cytoplasm of the host cell and the uncoating of highly stable picornaviruses such as hepatitis A virus remains an enigma.

Antiviral Drugs

The VP1 β -barrel of enteroviruses contains a hydrophobic pocket which, in some structures is occupied by endogenous lipids (cofactors known as 'pocket factors') that have been variously modeled as e.g., sphingosine, lauric acid, myristic acid. Crystallographic analyses revealed that hydrophobic antiviral 'WIN' compounds bind tightly to this pocket, displacing any natural pocket factor. The entrance to the pocket is at the bottom of the canyon and the antiviral agents lie roughly parallel to the floor of the canyon, the inner end of the pocket being lined by residues more hydrophobic than those towards the pore. In addition to major and minor receptor binding groups, rhinoviruses are classified into antiviral groups A and B according to their sensitivity to 15 antiviral compounds, group A being more sensitive to long antiviral compounds and group B to short antiviral compounds. Minor group viruses all belong to group B and the similar amino-acid composition in the pocket suggests that these viruses may all contain pocket factors. These hold the pocket "open" and in the case of human rhinovirus 1A (HRV1A) (group B) on binding of antiviral compounds only small conformational changes are seen. Larger conformational changes occur with all anti-viral compounds on binding to HRV14 (group A) - principally involving a movement of the VP1 GH loop upwards into the canyon. Potential antiviral inhibitors block virus replication in two ways: (1) they prevent virus uncoating and the release of viral RNA, by filling the hydrophobic pocket within the β -barrel of VP1 and stabilizing the virion, (2) they reduce receptor binding. WIN resistance mutations occur within the pocket, usually involving a change to a larger hydrophobic residue, blocking antiviral entry. Only enteroviruses seem to be susceptible to this class of drugs since in aphthoviruses and cardioviruses the pocket is filled by bulky sidechains. These studies have led to the discovery of potential drugs, for example Pleconaril (WIN 63843) which was aimed at the prevention of asthma exacerbations and common cold symptoms and completed phase two clinical trials in 2007 but is not licensed for use owing to the potential for multiple drug interactions.

A novel 'pocket-factor mimic' compound was rationally designed based on analysis of complexes of EV-A71 with four 3-(4-pyridyl)-2-imidazolidinone derivatives with varying anti-EV-A71 activities. Key structure-activity correlates were used as guides for

potentially beneficial substitutions and methods developed to reliably triage these compounds by quantum mechanics-enhanced ligand docking. Two candidates were synthesized, and structural analyses and in vitro assays confirmed the predicted binding modes and their ability to block viral infection. One ligand (with IC₅₀ of 25 pM) was an order of magnitude more potent than the best previously reported inhibitor and also more soluble. This structure-based approach may be useful for further design of effective drugs for enterovirus infections.

A recent cell-based antiviral screen against Coxsackievirus B3 (CVB3) led to the identification of compound 17, a benzene sulfonamide derivative, as a potent in vitro inhibitor. Its binding site was identified by cryo-EM as being a pocket formed by two VP1 and one VP3 protein at the interface between protomers. This positively-charged surface depression is conserved across many enteroviruses and has been identified as a binding site for glutathione which has an essential role in the formation of the pentameric subunit since bound glutathione makes strong interactions with adjacent protomers. This explains the underlying biological role of this binding pocket and opens the possibility for the rational design of antiviral drugs targeting this site. In fact, since this second druggable pocket is separated from the VP1 hydrophobic pocket by 16 Å, complementary antiviral effects might be achievable.

Antiviral drugs may also target the non-structural proteins; the 3C protease and 2C ATPase have been targets for the design of broad-spectrum drugs. However, the first line of defense against a viral attack is the immune response directed at epitopes at the capsid surface.

Antigenicity and Antibody Interactions

Locating the sites of escape mutations for resistance to neutralisation by monoclonal antibodies was traditionally used to map epitopes to the virus surface. Escape mutations often occur repeatedly at only a few residue positions and for FMDV-C, structural studies have shown that they address amino acids not involved in interactions with other capsid residues, suggesting that they are not required for maintenance of the capsid structure and since the virus remains viable, they are not required for another essential biological function. This applies to receptor binding sites in exposed locations such as the FMDV GH loop where amino acids adjacent to the key receptor binding residues can readily mutate without compromising the viability of the virus particle, although they may even switch the conformation of this immunodominant loop. Many picornaviruses possess 4–6 distinct antigenic regions and there is some correspondence between different viruses. It is, however, likely that the majority of the virus surface is antigenic. Whilst epitopes may comprise residues from disparate parts of the same, or different proteins, they cluster in surface patches, usually corresponding to variable surface accessible sequences. The flexibility of continuous epitopes located in antigenic loops provides a vehicle whereby, by switching loops from one virus into another, it has been possible to produce particles that possess antigenic properties in common with two different viruses. Such chimeric viruses may present possibilities for more broadly protecting vaccines.

Understanding the mechanisms by which antibodies neutralize picornaviruses and, conversely, how viruses escape from antibody recognition requires a detailed knowledge not only of the virus structure but also of the virus–antibody interactions at the amino acid level. Studies of viral peptide–antibody complexes and picornavirus–Fab complexes by X-ray crystallography, cryo-electron microscopy or a combination of the two techniques, have built on the data from escape mutant analyses.

The surface area on HRV14 in contact with Fab17-IA, the antibody footprint, was estimated to be approximately 500 Å² - much larger than the NIIm-IA neutralizing immunogenic epitope defined by escape mutants: in addition to antigenic residues which are tolerant to change through escape mutation, some invariant residues that may be critical for virus structure also make antibody contacts.

Antibodies neutralize virus infectivity by a variety of mechanisms other than simply interfering with infection by binding to cell attachment sites on the virion, for example, altering virus stability, preventing uncoating, or cross-linking viruses causing aggregation. Neutralization titration curves indicate three categories: strong, intermediate and weak. The majority of antibodies are weak neutralizers that function by using the two arms of the antibody to cross-link different virus particles giving rise to aggregation. The strong/intermediate neutralizers show a monotonic decrease in the amount of surviving virus as the antibody to virus ratio increases and may neutralize by blocking the receptor binding site, interfering with cell attachment.

For the strongly neutralizing mAb17-IA binding to HRV14, 5–6 mAbs/virion are sufficient to block cell attachment by 50%. The complex structure shows that the Mab binds bivalently across an icosahedral 2-fold symmetry axis. Since all 60 receptor binding sites cannot be blocked, the mechanism of neutralization may be by preventing structural changes necessary for receptor attachment and indeed no significant changes were seen in capsid structure. The C3 neutralizing antibody for poliovirus binds to the analogous site, a peptide corresponding to the BC loop of VP1 (though varying in length, structure and sequence from that in HRV14). Both structural and biochemical studies indicate that there are differences in the modes of binding and mechanisms of neutralization of the two antibodies. C3 neutralizes poliovirus 1 at a low Ab: virus ratio and no reduction in attachment to cells is seen. It may neutralize by causing significant local alterations in the antigenic site and prevent structural changes subsequent to receptor attachment that are necessary for cell entry or uncoating.

In the SD6 Fab/FMDV-CS8 complex, the Fab projects almost radially from the viral surface in an orientation only compatible with monovalent binding of the mAb to its GH-loop epitope in what is termed the 'up' position close to the capsid 5-fold axis, the 'down' position being closer to the capsid 2-fold axis. Even taking into account antibody flexibility, it was not possible to model bivalent binding without severe distortion. Although this might suggest that SD6 driven aggregation is the mechanism of

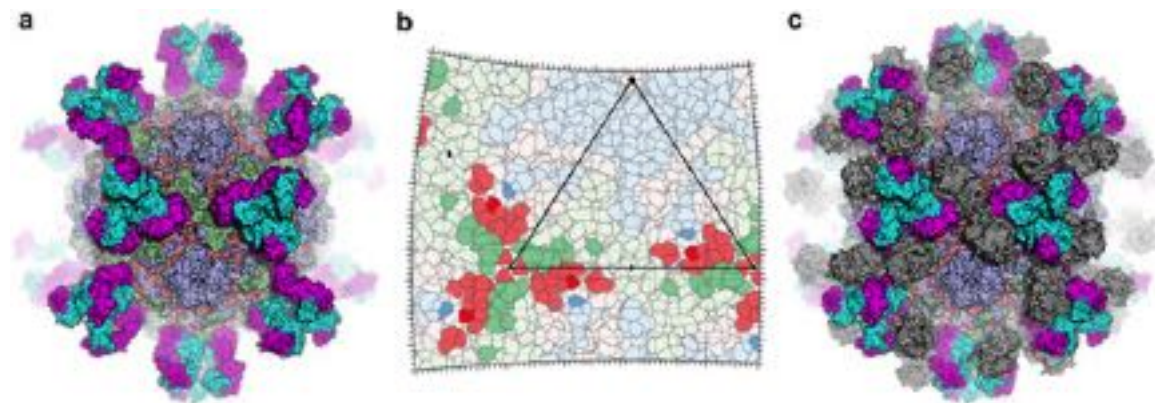


Fig. 5 EV-A71 and antibody 38-3-11A Fab complex (6Z3P). a) The virus capsid is shown as a surface representation with the proteins colored as for Fig. 2(a). The Fabs are also shown as a surface representation with the heavy chains in magenta and light chains in cyan. b) A roadmap (flattened surface representation of the virus surface where residues are drawn to size according to their accessible surface area). The Fab footprints are shown by enhancing the color of the capsid proteins to which they bind. c) As a) but showing SCARB2 (gray) binding in the context of the Fab complex.

neutralization, this is not the full story since the SD6 GH-loop epitope includes the integrin binding motif and the Fab alone acts as a strong neutralizer, presumably by interfering with cell attachment. The SD6 Fab shows induced fit conformational changes on binding the FMDV GH-loop epitope, to create a complementary pocket. This feature is also observed in the complex between the weakly neutralizing antibody 8F5 and a contiguous epitope at immunogenic site N1m B, from VP2 of HRV2. This complex shows bivalent binding across the icosahedral 2-fold axis and demonstrates that bivalent binding does not ensure strong neutralization.

In recent years, the development of techniques for single cell sorting followed by RT-PCR and sequencing have allowed immunoglobulin (Ig) heavy (H) and light (L) chain matching from isolated peripheral blood mononuclear cells (PBMCs). This enables recombinant expression and characterization of large numbers of antibodies derived from the real hosts of viruses. Six complementarity determining regions (CDRs), 3 each in the variable (V) domain of the H and L chains determine the specificity and affinity of an antibody. Of these, VH CDR3 is the most variable and plays a dominant role in recognizing and binding antigens; in mice this region is shorter and potentially less able to explore key areas of human epitopes, emphasizing that Ig repertoire studies using animal models can be misleading. When a large panel of mAbs derived from antibody secreting cells of several EV-A71-infected children were analyzed, potency and breadth of neutralisation of different virus genotypes could be correlated to epitope localization on the virus capsid: antibodies binding to the floor and rims of the capsid canyon were more effective than those mapped to the 3- and 2-fold so called plateau epitopes on the margin of pentamers (Fig. 5).

Conclusion

Structural analyses have now been completed for many small ssRNA viruses, and in particular studies of picornaviruses have highlighted similarities and differences between the genera relating to their biological properties. Indeed, structure-based phylogenetic analysis reveals the relative distance of the parecho and kobu genera from the entero, aphtho and cardio genera and places HAV within the wider picornavirales between typical 'modern' picornaviruses and the more primitive precursor insect viruses reflecting its enigmatic properties. These advances in our understanding of this lineage of viruses take us from the fundamental biology of the virus life-cycle and evolution to the potential for rationally designed antivirals and vaccines.

Further Reading

- Abrescia, N.G.A., Bamford, D., Grimes, J.M., Stuart, D.I., 2012. Structure unifies the viral universe. *Annual Review of Biochemistry* 81, 795–822. doi:10.1146/annurev-biochem-060910-095130.
- De Colibus, L., Wang, X., Spyrou, J.A.B., *et al.*, 2014. More-powerful virus inhibitors from structure-based analysis of HEV-A71 capsid-binding molecules. *Nature Structural & Molecular Biology* 21 (3), 282–288. doi:10.1038/nsmb.2769.
- Ho, I.Y., Bunker, J.J., Erickson, S.A., *et al.*, 2016. Refined protocol for generating monoclonal antibodies from single human and murine B cells. *Journal of Immunological Methods* 438, 67–70. doi:10.1016/j.jim.2016.09.001.
- Huang, K.A., Zhou, D., Fry, E.E., *et al.*, 2020. Structural and functional analysis of protective antibodies targeting the threefold plateau of enterovirus 71. *Nature Communications* 11, 5253. doi:10.1038/s41467-020-19013-3.
- Kotecha, A., Wang, Q., Dong, X., *et al.*, 2017. Rules of engagement between $\alpha v\beta 6$ integrin and foot-and-mouth disease virus. *Nature Communications* 8, 15408. doi:10.1038/ncomms15408.
- Lee, H., Shingler, K.L., Organtini, L.J., *et al.*, 2016. The novel asymmetric entry intermediate of a picornavirus captured with nanodiscs. *Science Advances* 2 (8), e1501929. doi:10.1126/sciadv.1501929.

- Marsian, J., Fox, H., Bahar, M.W., *et al.*, 2017. Plant-made polio type 3 stabilized VLPs-a candidate synthetic polio vaccine. *Nature Communications* 8, 245. doi:10.1038/s41467-017-00090-w.
- McPherson, A., Larson, S.B., 2015. A guide to the crystallographic analysis of icosahedral viruses. *Crystallography Reviews* 21 (1–2), 3–56. doi:10.1080/0889311X.2014.963572.
- Porta, C., Kotecha, A., Burman, A., *et al.*, 2013. Rational engineering of recombinant picornavirus capsids to produce safe, protective vaccine antigen. *PLoS Pathogens* 9 (3), e1003255. doi:10.1371/journal.ppat.1003255.
- Prasad, B.V., Schmid, M.F., 2012. Principles of virus structural organization. *Advances in Experimental Medicine and Biology* 726, 17–47. doi:10.1007/978-1-4614-0980-9_3.
- Rossmann, M.G., He, Y., Kuhn, R.J., 2002. Picornavirus-receptor interactions. *Trends in Microbiology* 10 (7), 324–331. doi:10.1016/s0966-842x(02)02383-1.
- Tuthill, T.J., Groppelli, E., Hogle, J.M., Rowlands, D.J., 2010. Picornaviruses. *Current Topics in Microbiology and Immunology* 343, 43–89. doi:10.1007/82_2010_37.
- Wang, X., Peng, W., Ren, J., *et al.*, 2012. A sensor-adaptor mechanism for enterovirus uncoating from structures of EV-A71. *Nature Structural & Molecular Biology* 19 (4), 424–429. doi:10.1038/nsmb.2255.
- Zhou, D., Zhao, Y., Kotecha, A., *et al.*, 2019. Unexpected mode of engagement between enterovirus 71 and its receptor SCARB2. *Nature Microbiology* 4 (3), 414–419. doi:10.1038/s41564-018-0319-z.
- Zhao, Y., Zhou, D., Ni, T., *et al.*, 2020. Hand-foot-and-mouth disease virus receptor KREMEN1 binds the canyon of Coxsackie Virus A10. *Nature Communications* 11 (1), 38. doi:10.1038/s41467-019-13936-2.

Relevant Websites

<https://www.picornaviridae.com>
 Picornavirus Home.
www.pdb.org
 wwPDB: Worldwide Protein Data Bank.

Structural Principles of the Flavivirus Particle Organization and of Its Conformational Changes

Stéphane Duquerroy, University of Paris-Saclay, Orsay, France and Institut Pasteur, Paris, France
Arvind Sharma and Félix A Rey, Institut Pasteur, Paris, France

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Flaviviruses cause some of the most devastating arthropod-borne diseases around the world, including dengue hemorrhagic fever, yellow fever, Zika disease, Japanese encephalitis, West Nile encephalitis and tick-borne encephalitis (Huang *et al.*, 2019). The emergence and spread of flaviviruses into new geographical areas pose substantial public health challenges for multiple countries. Important progress has been made in understanding the structural organization of the flavivirus particles and the way their envelope proteins function during the virus life cycle, providing crucial knowledge for the development of rational approaches to cure and/or prevent these diseases. Yet many aspects of the flavivirus particle assembly process remain not understood.

The flavivirus molecular biology era began with the determination of the full nucleotide sequence of the yellow fever virus genome in 1985 (Rice *et al.*, 1985) and was pursued with structural biology ten years later with the high-resolution X-ray structure of the envelope protein E of the tick-borne encephalitis virus (TBEV) (Rey *et al.*, 1995) and later that of dengue virus type 2 (DENV2) (Modis *et al.*, 2003). These structures corresponded to the ectodomain of the pre-fusion E dimer, i.e., the form present at the surface of infectious particles. The structures of the E trimer ectodomain soon followed (Bressanelli *et al.*, 2004; Luca *et al.*, 2013; Modis *et al.*, 2004) – i.e., in the conformation adopted after the functional reorganization of E required to drive the fusion of the viral envelope with an endosomal membrane for entry into a target cell. A further important landmark was the X-ray structure of the DENV2 prM/E heterodimer present in immature flavivirus particles (Li *et al.*, 2008). The first cryo-electron-microscopy (cryo-EM) structure of a mature virion to a resolution sufficient to resolve the individual E dimers at the particle surface was reported in 2002 (Kuhn *et al.*, 2002). This pioneering work was followed by the cryo-EM structures of mature DENV2 particles at a nominal resolution of 3.5 Å reported in 2013 (Zhang *et al.*, 2013) – at a time when the direct electron-counting detectors (McMullan *et al.*, 2016) that led to the present day “resolution revolution” in single-particle cryo-EM (Henderson, 2015) were not yet available. A number of important cryo-EM structures have been reported since then, such as the structures of the mature particles of DENV4 (Kostyuchenko *et al.*, 2014), Zika virus (ZIKV) (Kostyuchenko *et al.*, 2016; Sirohi *et al.*, 2016), Japanese encephalitis virus (JEV) (Wang *et al.*, 2017) and TBEV (Fuzik *et al.*, 2018) at around 4 Å resolution, culminating in 2018 with the refined structure of the ZIKV mature particle at 3.1 Å resolution (Sevvana *et al.*, 2018). The structures of immature flavivirus particles were also determined by cryo-EM, albeit to lower resolution, the best defined one was that of DENV1 at 6 Å resolution (Kostyuchenko *et al.*, 2013).

In this article, we examine the interactions between the viral glycoproteins at the particle surface, with a particular emphasis on the immature particle at neutral pH – the form in which it accumulates in the ER of the infected cell, before transport through the secretory pathway. We describe this icosahedral surface lattice as an intertwined network of E triskelions that are crosslinked by prM, and analyse the major conformational change it undergoes at acid pH. We then examine the geometric construction leading to the peculiar icosahedral neutral pH immature particle, and identify that a key feature of this organization is an empty position in the corresponding icosahedral lattice. We further show that the thermodynamic relation between the two immature lattices – at neutral and acid pH – can be explained in terms of the 532 symmetry point group within the framework of the Landau “weak crystallization theory” as implemented by (Rochal *et al.*, 2016).

The Flavivirus Precursor Polyprotein and the Derived Structural Proteins

The flavivirus genome (Fig. 1) is a single-stranded RNA molecule about 11,000 nucleotides long containing a single open reading frame (ORF) flanked by 5′ and 3′ untranslated regions (UTRs) that control its replication in the infected cell (Pierson and Diamond, 2013). In a first step, the genome acts as a messenger RNA directing its translation by cellular ribosomes. In a subsequent step, it directs the production of full-length antigenomic copies by the newly translated viral polymerase to serve as replication template. The translation of the flavivirus genome by cellular ribosomes yields a single, virtual polyprotein precursor about 3400 amino acids (aa) long (Fig. 1(A)) that crosses the ER membrane several times. This polyprotein is virtual because it undergoes autocleavage before translation is complete, together with processing by host proteases such as ER membrane resident signal peptidases and by an additional virally-encoded protease to yield the individual viral proteins (Pierson and Diamond, 2013). The three proteins composing the flavivirus virion – i.e., the “structural” proteins – are derived from the N-terminal third of the precursor, while its remainder yields the non-structural (NS) proteins, which are essential viral proteins present in the infected cell (including the replication enzymes, Fig. 1(A)) but which are not incorporated into particles.

Some NS proteins are integral membrane proteins spanning the ER membrane multiple times, while others are peripherally associated with the membrane via protein-protein interactions on the cytosolic side. The latter group includes the two main replication enzymes, the viral protease/RNA helicase NS3 and the methyl transferase/RNA polymerase NS5 – reviewed by (Tay and Vasudevan, 2018). The only non-cytoplasmic NS protein is NS1, which forms membrane-associated dimers (Akey *et al.*, 2014) that

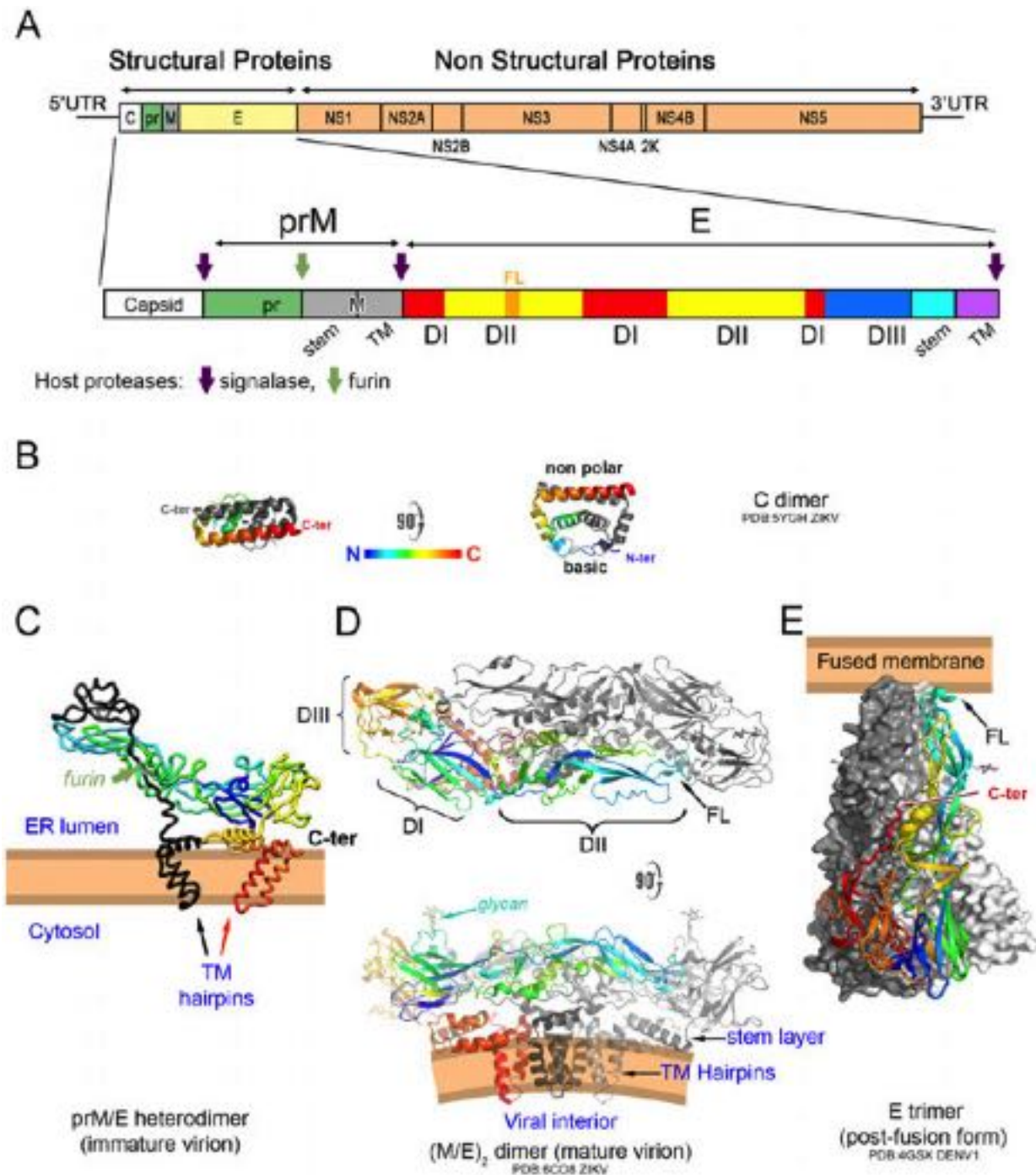


Fig. 1 Flavivirus polyprotein precursor and derived structural proteins. (A) The flavivirus genome is a messenger RNA molecule with a single open reading frame (ORF) (top, boxed) coding for a polyprotein precursor of all viral proteins. The ORF is flanked by highly structured untranslated regions (UTRs) that are required for genome replication. The mature viral proteins are indicated within the ORF box, roughly to scale. The structural proteins, colored differently in the top diagram, are enlarged in the second panel to show the various domains of protein E. (B) The flavivirus capsid protein C dimer. One of the chains is colored gray while the other is ramp-colored from N- to C-ter according to the color-key. Note that the α -helices from both chains are involved in creating a single tertiary structure. There is a non-polar face predicted to contact the viral membrane, and a basic face predicted to interact with the genomic RNA. The structure displayed corresponds to the 1.9 Å X-ray structure of ZIKV C (Shang *et al.*, 2018). (C) The prM/E heterodimer derived by docking the X-ray structure of the DENV2 pr/E complex at 2.2 Å resolution (Li *et al.*, 2008) into a new reconstruction of immature flavivirus particles to 4.4 Å resolution (PDB:XXXX, F. Coulibaly, personal communication). prM is represented in gray whereas E is shown ramp colored according to the key in panel B. The furin cleavage site in prM is indicated. (D) The mature (M/E)₂ dimer, as extracted from the refined 3.1 Å resolution cryo-EM structure of the ZIKV virion (Sevvana *et al.*, 2018), presented in top (upper panel) and side (lower panel) views. The individual domains of E are labeled, and the location of the fusion loop (FL), buried at the E dimer interface, is indicated. In the side view, the viral membrane is schematized in brown/beige. (E) The post-fusion E trimer. The subunit in the foreground is shown as colored ribbons (as in D) and the two subunits in the background in surface representation in two different shades of gray. The predicted location of the fused membrane is indicated, into which the fusion loops are inserted. Reproduced from Klein, D.E., Choi, J.L., Harrison, S.C., 2013. Structure of a dengue virus envelope protein late-stage fusion intermediate. *Journal of Virology* 87 (4), 2287–2293. doi:10.1128/JVI.02957-12.

participate in the replication complex – albeit from the ER luminal side – via interactions with other TM NS proteins such as NS4B (Mackenzie *et al.*, 1996; Youn *et al.*, 2013, 2012). A fraction of NS1 is secreted as hexamers from infected cells (Gutsche *et al.*, 2011). Once sufficient viral proteins have accumulated in the cell, the viral genome is replicated by the viral polymerase and associated NS proteins. Although the molecular details are not understood, genome replication occurs within invaginations of the ER membrane (Junjhon *et al.*, 2014; Welsch *et al.*, 2009).

The three structural proteins of the flaviviruses are depicted in Fig. 1(B)–(D): the capsid (C) protein, which is roughly 100 aa long (depending on the flavivirus), and two transmembrane glycoproteins: the precursor Membrane protein (prM, ~170 aa, Fig. 1(C)) and the Envelope protein (E, ~500 aa, Fig. 1(C)–(E)). The cytosolic C protein folds as an α -helical dimer (Dokland *et al.*, 2004; Jones *et al.*, 2003; Ma *et al.*, 2004; Shang *et al.*, 2018) exposing a non-polar face predicted to contact the lipids of the internal leaflet of the viral membrane, and a basic face contacting the RNA genome (Fig. 1(B)). The C dimer, which is initially sequestered by lipid droplets while replication takes place (reviewed by (Byk and Gamarnik, 2016)), subsequently binds the newly replicated genomic RNA molecules to form a ribonucleoprotein (RNP) complex. The RNPs do not accumulate in the cell during infection, and their formation appears to be concomitant to particle assembly and budding.

Glycoproteins prM and E are co-translationally translocated into the ER lumen, where they form a prM/E heterodimer. They both feature a β -sheet rich ectodomain: prM has a single globular β -sandwich encompassing the N-terminal half, while E displays a typical class II membrane fusion protein fold (Kielian and Rey, 2006; Modis, 2013), with three characteristic β -sheet rich domains, termed I, II and III (Fig. 1(D)) making roughly the 4/5th of the intact E protein (Fig. 1(A), second panel). In both glycoproteins, the ectodomain is followed by a flexible region termed “stem” that connects to a C-terminal transmembrane (TM) α -helical hairpin. The E stem forms three mostly amphipathic α -helices oriented tangentially and sandwiched between the viral membrane and the E ectodomain layer (Fig. 1). The prM stem adopts mainly an extended conformation followed by an amphipathic α -helix preceding the TM hairpin. A furin cleavage site is present at its junction with the globular “pr” domain (Fig. 1(C)). The prM/E heterodimer accumulates at the luminal side of the ER membrane until the budding of new particles, and multimerizes via lateral interactions that induce the required membrane curvature to make closed particles even in the absence of genomic RNPs (Allison, 1995b). Incorporation of the viral RNA to form complete virions presumably occurs via interactions of the “turn” in the helix-turn-helix of the TM hairpins (thin arrows in Fig. 1(C)) – the only site exposed at the cytosolic side – and C dimers in the genomic RNPs (Tan *et al.*, 2020). Particle assembly and budding, which also requires specific interactions of the prM/E heterodimers with NS1 (Scaturro *et al.*, 2015), results in the accumulation of immature spiky virions in the neutral environment (pH ~7.2) of the ER lumen, forming para-crystalline arrays. Interaction with the KDEL receptor of the host results in their subsequent transport to the cell surface through the TGN (Li *et al.*, 2015).

The Immature Flavivirus Particle

The surface of the immature particles at neutral pH is formed by 180 prM/E protomers making a network of intertwined interactions with overall icosahedral symmetry, as illustrated in Fig. 2(A), center-left panel. Examination of the available structures revealed a set of key interactions made by the TM helical hairpins of prM and E, which associate to make a trimeric α -helical bundle in the membrane (Fig. 3(C)). The resulting (prM/E)₃ trimer of heterodimers has the form of a propeller – referred to as “triskelion” – projecting three blades (Fig. 3(A)). Sixty triskelions make up the particle interacting about the icosahedral axes as outlined in Fig. 3(B). The molecular axis of the triskelion appears derived from an initial 3-fold axis that is distorted by non-equivalent interactions made by its three blades to form the particle, resulting in significant deviation from 3-fold symmetry (Fig. 3(B)). The three blades, termed I5, I3 and I2 by reference to the closest icosahedral symmetry axis (5-fold, 3-fold and 2-fold), are respectively colored with prM/E in cyan/blue, pink/red and yellow/green in Figs. 2 and 3(B). In the particle, each triskelion has the blades of three adjacent ones interacting directly on top of its central TM helical bundle (Fig. 2(A), center-left), forming prominent trimeric spikes. These spikes were the only distinguishable morphological features in the initial low resolution cryo-EM studies (Zhang *et al.*, 2003; Zhang *et al.*, 2007). The interaction between the three blades forming the spike are not 3-fold symmetric: rather, a local dimer axis relates I3 and I2 blades (red and green in the Figures) with about 650 Å² buried surface area (BSA) per blade, in an interaction involving mainly prM (Table 1). The I5 blade leans on this dimer, with a BSA of about 450 Å², with prM contacting only the E moiety of the I3 blade. The buried surface of the I5 blade is mostly the same prM surface involved in the 2-fold interaction between I2 and I3 blades. The overall arrangement results in a lattice with a highly intertwined pattern, in which most of the lateral interactions are made by prM (colored cyan, mauve and yellow in the I5, I3 and I2 blades, respectively, Figs. 2 and 3(B)). Importantly, very recent cryo-EM data showed that prM directs its TM hairpin from the top of the spike directly into the helical bundle underneath, whereas the E TM hairpin is part of three adjacent bundles. This connectivity results in a heterodimer “swapping” with respect to the initial connectivity deduced from lower resolution cryo-EM maps. The implied pattern is now that each triskelion blade has the prM and E TM hairpins making part of different TM helical bundles (F. Coulbaly, Personal communication; PDB:XXXX). This particular organization of the prM TM hairpins results in a highly interconnected lattice, and is likely to be responsible for the reversible transition between the two lattices when going from neutral to acid pH, and vice-versa.

A cryo-EM structure of the immature ZIKV particle to 9 Å resolution showed internal density directly underneath the α -helical bundle of the triskelions (Fig. 3(C) and (D)) (Prasad *et al.*, 2017). The location of this density supports the notion that C dimers in the genomic RNP interact with the cytosolic side of the TM bundles, although there is no clear density to resolve the C protein in the structure. More recently, an asymmetric reconstruction of the immature particle of Kunjin virus (a flavivirus very close to West

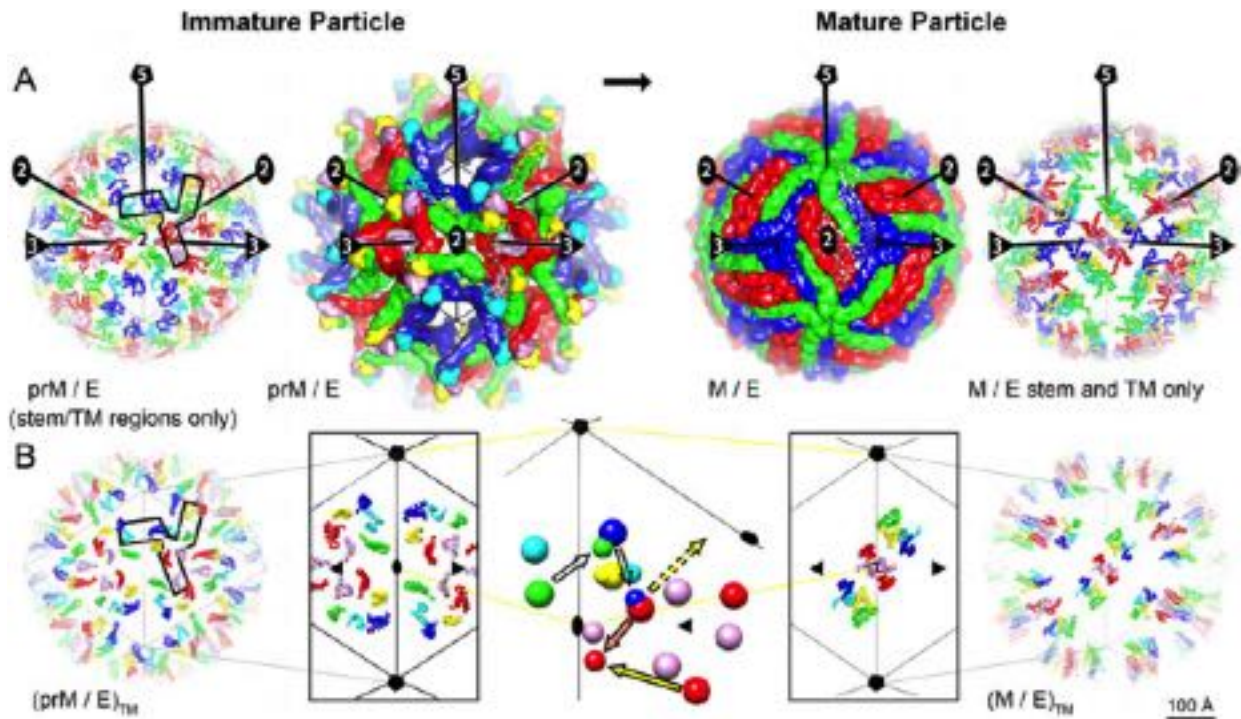


Fig. 2 Flavivirus Particle Organization. (A) Immature (left panels) and mature (right panels) flavivirus particles viewed down an icosahedral 2-fold axis. The mature particle is based in the immature lattice at acid pH, with the difference that prM has been cleaved by furin and the pr moiety is shed upon exposure to neutral pH. Three independent immature prM/E (left) or mature M/E (right) protomers make up the icosahedral asymmetric unit, respectively colored cyan/blue, pink/red or yellow/green according to the nearest icosahedral axis (I5, I3 or I2, as described in the text). The central panels display a surface representation, except for one asymmetric unit shown with the protomers in ribbons. The immature particle displays a highly intertwined arrangement, in which each triskelion is centered underneath a spike formed by the blades of three adjacent ones. The central triskelion in turn contributes one blade each of its three adjacent spikes but the prM moiety of each blade connects directly to the TM helical bundle located underneath each spike. We define a “local” (prM/E)₂ dimer (yellow/green and cyan/blue protomers) within which the TM hairpins of prM are swapped (i.e., the pr “head” of prM interacts with domain II of one E polypeptide, and the prM TM hairpin interacts with the TM hairpin of the opposite E polypeptide; see also Fig. 4(D)). This local dimer appears to give rise to the local dimers in the low pH lattice (upon dissociation of the TM helical bundles). This reorganization is illustrated by using the same color-coding in the center-right panel, where the mature particle is shown, whose surface lattice corresponds also to that of the acid pH immature particle. At neutral pH, the pink/red prM/E protomers form rings about the 20 I3 axes of the particle, with a circular connectivity within each ring via the swapped prM TM hairpins. These same protomers are proposed to re-organize into the 30 (prM/E)₂ dimers sitting on the I2 axes in the low pH lattice. The lateral panels show the organization of the “stem” and transmembrane segments. In the left panel, an outline indicates the three prM/E protomers constituting one triskelion, with the prM TM segment swapped into adjacent alpha-helical TM bundles. (B) The outer panels show the prM/E TM helix bundles in the immature (left) and mature (right) particle, with a zoom as inset. In the central inset, only the center of mass (COM) of the TM hairpin of each subunit is represented, color-coded and with big and small spheres respectively indicating their position in the neutral and acid pH lattice. White arrows show the COM displacements of the prM TM in the local dimer defined above; pale red and yellow arrows display alternative possibilities for its displacement to make the icosahedral dimer.

Nile virus), was determined to 19 Å resolution, showing the presence of an RNP core interacting only on one side with the viral membrane (Therkelsen *et al.*, 2018). This eccentricity of the RNP, which explains the lack of well-defined density in the icosahedrally averaged reconstructions, further suggests that its interactions with only a subset of the 60 trimeric α -helical bundles of the particle may be sufficient for its incorporation into a budding particle. Recent evidence has confirmed the interactions of the triskelion TM bundle with the genomic RNP (Tan *et al.*, 2020).

Acid Sensitivity

Flavivirus virions are lipid-enveloped, pH sensitive macromolecular machines tailored to efficiently deliver the viral genome into target cells upon receptor-mediated endocytosis. As such, they are exquisitely sensitive to acid pH, to which they will be exposed twice in the viral cycle: the first-time during transport to the cell surface, upon exposure to the mildly acidic environment (pH ~6.0) of the trans-Golgi network (TGN). The second (and final) exposure to acid pH occurs in the endosome of a different cell, leading to the irreversible disassembly of the particle with concomitant fusion with the endosomal membrane, thereby releasing the genomic RNP into the cytoplasm of a target cell.

The first acid exposure triggers a major reorganization of the immature particle. The surface lattice disassembles as the trimeric α -helical TM bundle of the triskelion comes apart and the blade/blade contacts forming the spikes are broken. The 180 prM/E

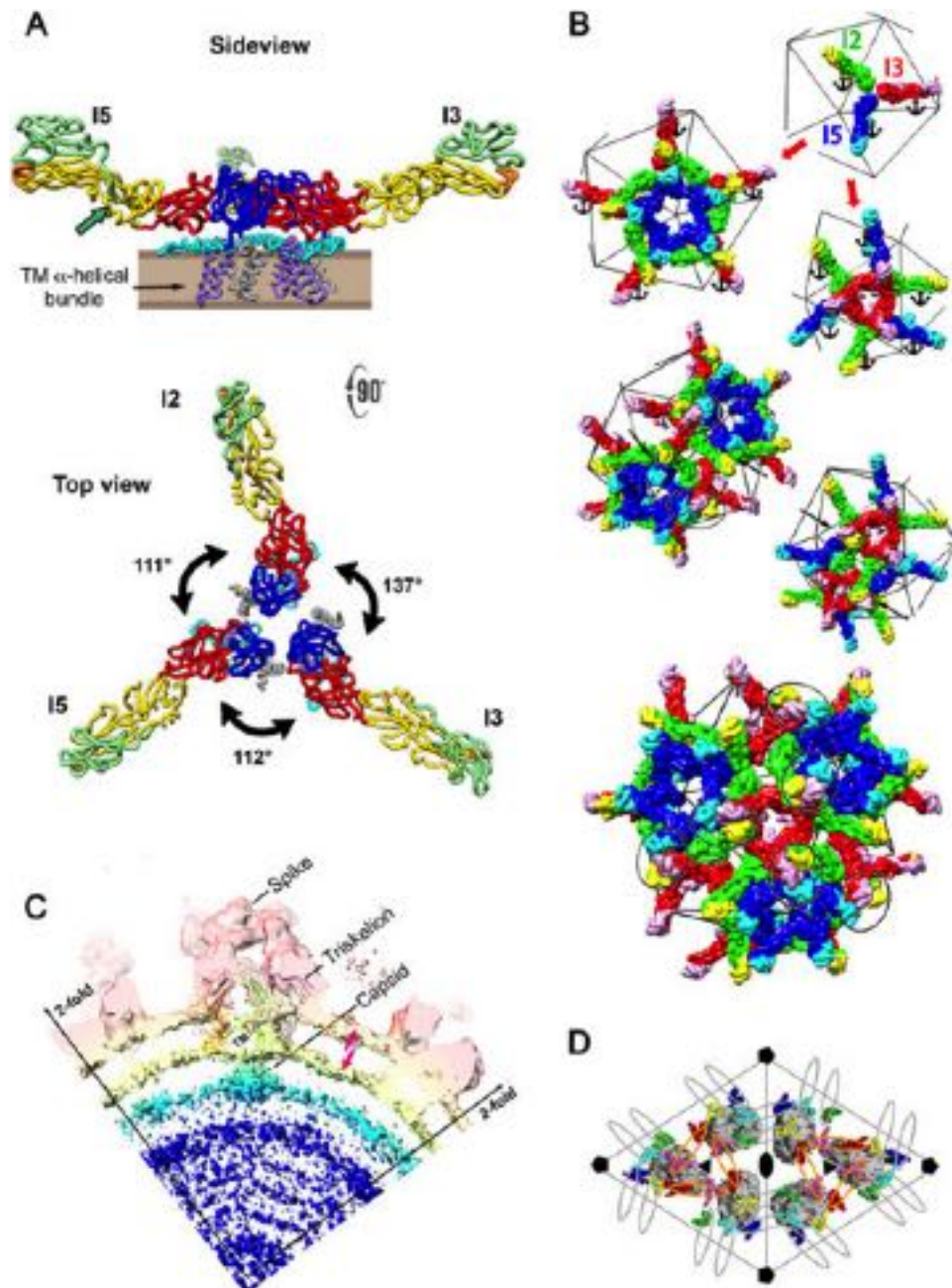


Fig. 3 The building block of immature virions. (A) A triskelion shown in side (top) and top view (bottom), with its three "propeller" blades labeled. prM is in green while E is colored according to domains (domains I, II, III, stem and TM hairpin in red, yellow, blue, cyan and violet, respectively, as in the linear diagram of Fig. 1(A), bottom panel). The prM stem and TM hairpin are shown in grey, as they belong to the heterodimers from adjacent triskelions. The E fusion loop (FL) is in orange. A green arrow points to the region of prM connecting to its TM hairpin in an adjacent alpha-helical bundle. The viral membrane is schematically outlined in brown. The bottom view shows that the triskelion deviates substantially from 3-fold symmetry. (B) Potential assembly pathways. At the top right, a single triskelion is shown in surface representation and colored by blades as in Fig. 2. Marine anchor symbols indicate the connections to adjacent triskelions via the prM TM hairpins. The triskelions interact laterally around the 5-fold axes (as shown to the left) and about the 3-fold axes of the particle (as shown to the right). These subsets in turn associate with each other, as shown in the lower panels. In the spikes, the I2 and I3 propeller blades (red and green) make a near 2-fold interaction, as described in the text, while the I5 blade (blue), leans against the I3 blade only, without contacting the I2 blade. These contacts are shown by thin black arrows in the middle panels. Note that full spikes are only formed upon interaction of pentameric triskelion assemblies, forming the spikes outlined in the lower panel. (C) Low resolution internal density map for the immature ZIKV particle (Prasad et al., 2017), showing the density of the TM bundle and membrane proximal region of one triskelion, with density above corresponding to the blades of three adjacent triskelions making a spike (as labeled). The contours were chosen to show continuous density for the lipid headgroup region of the membrane (indicated by the two red arrows). At this contour level, there is internal density (cyan) that can be interpreted as being a C dimer underneath, intercalated by weak density. Internal density attributed to the RNA genome is shown in dark blue. (D) Representation of the TM segments in two icosahedral facets of the immature particle (i.e., 6 asymmetric units) at the level of the membrane. The TM hairpins are shown as ribbons, colored as in Fig. 2, and the gray surfaces correspond to the cyan density in (C) viewed from the outside of the particle. Grey and orange ellipses indicate swapped prM TM segments around the 5-fold and 3-fold axes. Pairs of grey ellipses correspond to the local dimers defined in the legend to Fig. 2. Small black polygons mark the icosahedral symmetry axes of the particle, as a guide.

Table 1 Buried accessible surface areas (BASA) per prM/E blade in the spike

blade	I2	I3	I5	Total
I2-E	/	10.1	0	10.1
I2-pr	/	654.0	0	654.0
total	/	664.1	0	664.1
I3-E	6.8	/	583.8	590.6
I3-pr	641.2	/	74.5	715.7
total	648.0		658.3	1305.3
I5-E	0	141.0	/	141.0
I5-pr	0	390.4	/	390.4
total	0	431.4	/	431.4

Note: BASA were calculated using Areaimol (CCP4) (Lee and Richards, 1971) and Zika immature model 5U4W. The colors correspond to the E moiety in the triskelion blades displayed in Figs. 2 and 3(B). Note that the I5 and I2 blades do not contact each other in the spikes.

Table 2 Protein rearrangements in the viral membrane

E-TM	Δcom (Å)	prM-TM	Δcom (Å)
I5	40	I5	56
I3	39 (64)	I3	48 (83)
I2	44	I2	4

Note: Translation of the E and prM trans-membrane hairpins during the transition from the neutral form to the acid pH immature form of the DENV1 particle. E-TM corresponds to E aa450–495, and M-TM to aa 130–166 of prM. Comparison done using entries PDB:4B03 (immature particle)- and PDB:4CCT (mature particle) (Kostyuchenko *et al.*, 2013). For the I3 protomers, the values correspond to the shift indicated by the pale red arrow in Fig. 2(B). The values in parenthesis correspond to an alternative shift, indicated by the yellow arrow in Fig. 2(B). For the I2 and I5 protomers, the shifts correspond to the white arrows in Fig. 2(B). (these shifts preserve the local dimer).

protomers reorganize into 90 (prM/E)₂ dimers forming a smooth shell completely cloaking the viral membrane (Li *et al.*, 2008; Yu *et al.*, 2008). The TM hairpins forming the trimeric bundle undergo different translations in the membrane, as listed in Table 2. Although the fate of the individual protomers is not established with certainty, the local dimer (prM/E)₂ defined in Fig. 2 suggest a potential pathway for the transition. In this dimer, the E surfaces directed towards each other are the same that form the local dimer in the mature form (Fig. 2(A)). Indeed, as shown in Fig. 4, there is essentially a translation of the prM/E protomers in opposite directions to form the mature-like dimer, allowing the E ectodomains to lie flat on the membrane. As shown for DENV2, this conformational change is reversible, and re-exposure to neutral pH results in dissociation of the (prM/E)₂ dimers of the acid pH lattice (Fig. 4, right panels) and re-formation of the triskelions and the spikes (Yu *et al.*, 2008). The stem regions, tethering the ectodomains to the TM anchor, play a crucial role in controlling the pH dependent reversible conformational changes. Yet the dynamic transitions in both prM and E stems that regulate the drastic functional rearrangement observed are not understood. The particle reorganization is rendered irreversible by furin cleavage (Stadler *et al.*, 1997) of the pr “head” from the stem “tether”, which removes the cross-linking between triskelions. The head remains bound in the resulting (prM/E)₂ dimers of the smooth particle as long as the environment is acidic (Li *et al.*, 2008; Yu *et al.*, 2008). In this way, the release into the neutral pH extracellular milieu after prM cleavage entails no reversion to the spiky form. A concomitant drop in affinity of the pr head for the (M/E)₂ dimer at neutral pH leads to its dissociation and shedding, resulting in a smooth mature virion exposing 90 metastable (M/E)₂ heterodimers at its surface (Fig. 2(A), middle-right panels). The mature particle is primed for fusogenic activation upon subsequent exposure to low pH. In the absence of pr, the second exposure to acid pH triggers an irreversible conformational change of protein E (Allison *et al.*, 1995a; Stiasny *et al.*, 2001), in which it dissociates from the (M/E)₂ dimers (Fig. 1(D)) so that E monomers can insert into the target membrane. The membrane-inserted monomers then re-associate to form E homotrimers (Fig. 1(E)) (Liao *et al.*, 2010; Stiasny *et al.*, 2007; Zheng *et al.*, 2014) that drive the fusion of the viral envelope with the endosomal membrane.

Mature Particle Organization

As discussed above, the mature virion maintains the organization of the immature particle after the acid-induced conformational change. We refer to this arrangement as the “mature” lattice for convenience, as there are no high-resolution structural studies of

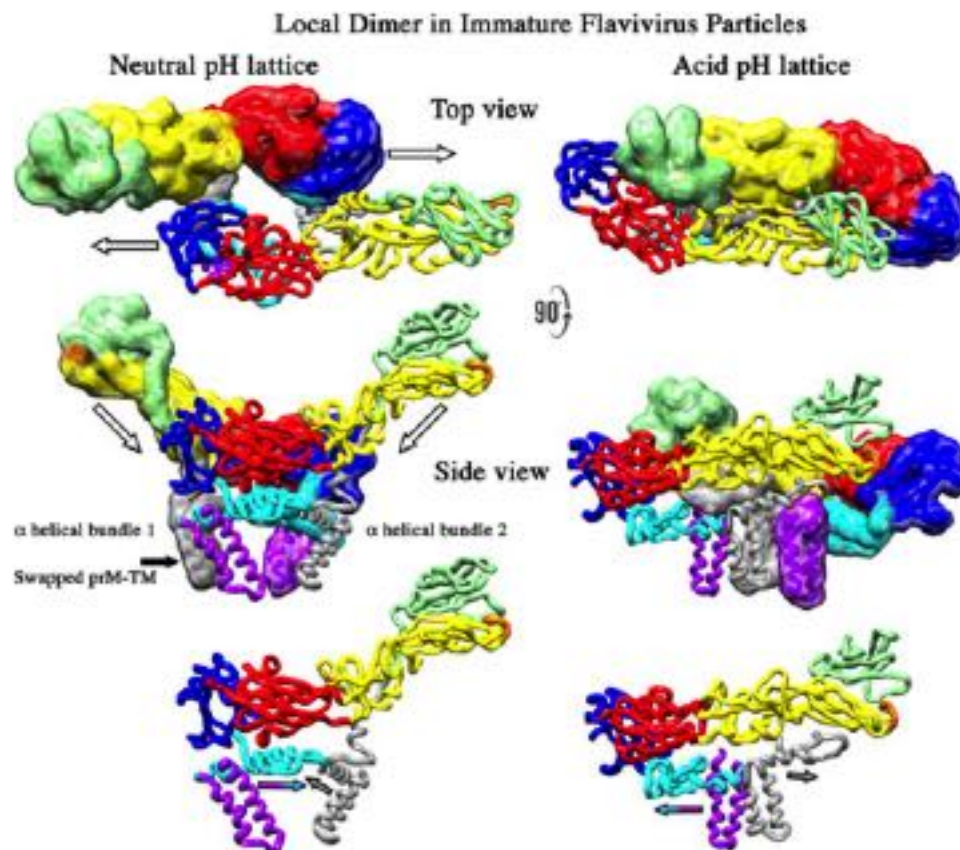


Fig. 4 Reorganization of the local (prME)₂ dimer during the transition to the low pH lattice. To highlight the swapping of the prM TM hairpin, one protomer in the dimer is shown in surface representation and the other as ribbons, colored as in Fig. 3(A). The (prME)₂ dimers in the neutral pH (left) and the acid pH (right) lattice are shown with their local 2-fold symmetry axes perpendicular to (top panels) and in the plane of (middle panels) the Figure. In the neutral pH lattice, the two protomers belong to two different triskelions (as labeled). The swapping of the prM TM hairpin is indicated with a black arrow in the side view (note that the E and prM chains displayed in ribbons interact with each other at the top, whereas in the membrane the interaction is between prM in ribbons and E in surface, and vice-versa). The third row shows the front protomer alone to better visualize the swapping, with colored arrows showing the rearrangement of the TM hairpins. Table 2 shows the transformation undergone by the E ectodomain in all three protomers of the triskelion.

the immature lattice at acidic pH. The only available cryo-EM map is at 25 Å resolution (accession EMD5006) (Yu *et al.*, 2008), and at this resolution its organization appears identical to that of mature virions. It is important, however, to recall that furin maturation occurs only after the transition between the two lattices, and that once prM is cleaved, the resulting pr/M/E heterotrimeric protomers are no longer capable of driving the reversible switch to the spiky lattice when exposed to neutral pH.

The organization of the mature flavivirus particle has been thoroughly described (Hasan *et al.*, 2018; Mukhopadhyay *et al.*, 2005). The (M/E)₂ dimers associate by sets of three, making parallel interactions and forming what is called a “raft”, in which a central dimer (in green) is related by an icosahedral 2-fold symmetry axis of the particle, while two symmetrical dimers at each side display only local 2-fold symmetry (blue/red dimers, Fig. 2(A), third panel). Thirty such rafts coat the viral surface to completely cover the viral membrane. The raft arrangement is reminiscent of a tricontahedron – a polyhedron with 532 point-group symmetry and 30 diamond shaped faces normal to the 2-fold icosahedral axes of the particle. Yet the particle is not polyhedral, as the curvature of the individual rafts results in a spherical virion. The glycan chains (one or two per E polypeptide, depending on the flavivirus) project from the smooth surface. Domain III, which has been proposed to be responsible for receptor binding (although there is no direct experimental evidence), protrudes slightly from the smooth particle. The viral membrane, in contrast, was shown to be highly distorted by the short TM α-helical hairpins, and the inner leaflet shows a polygonal shape that contrasts with the spherical outer surface of the virus particle. A 35 Å resolution non-symmetrized cryo-EM reconstruction of the mature particle of Kunjin virus showed that, contrary to immature particles at neutral pH, the density of the internal RNP is not eccentric (Therkelsen *et al.*, 2018). These data are in line with the dissociation of the trimeric α-helical bundles postulated to interact with the RNP core, freeing it to adopt a membrane-equidistant, central location in the particle.

Physical Principles of the Flavivirus Particle Organization

The organization of the flavivirus particle clearly does not follow the quasi-equivalence principle enunciated by Caspar and Klug (Caspar and Klug, 1962). This principle served to explain the organization of multiple viruses, postulating quasi-equivalent

interactions between the viral coat proteins (CPs) at the particle surface to make a closed shell. Although quasi-equivalence accounted for the organization of many spherical viruses, it did not for many others. Yet, a recent generalization of the geometrical concepts of Caspar and Klug (Rochal *et al.*, 2016) allowed its extension to understand the organization of virus particles not obeying the principle of quasi-equivalence. This basic formalism can indeed be adapted to understand the organization of both lattices, at neutral and at acid pH, of the immature flavivirus particle. The Caspar and Klug formulation for building an icosahedral capsid can be described as starting from a two-dimensional (2D) hexagonal lattice in which the individual CPs display identical lateral interactions to make a planar protein shell with P6 symmetry. In such a lattice, the 6-fold axes also generate parallel 3-fold and 2-fold axes, all normal to the lattice plane. Applying the architectural concepts for building a geodesic dome developed first by Walther Bauersfeld and expanded later by Walter Gropius and Buckminster Fuller, Caspar and Klug converted the planar protein shell into an icosahedral coat by eliminating one of the six proteins surrounding a 6-fold axes and rearranging the contacts to make a 5-fold axis instead, as outlined in Fig. 5(A). Each hexagonal node of the lattice is thus transformed into one of the 12 vertices of an icosahedron. This procedure allows the generation of icosahedral particles of increasing size built from similarly sized CPs by leaving intact increasing numbers of 6-fold nodes in between the nodes that are converted into 5-fold axes. The icosahedral lattice generated by the above procedure maintains the 3-fold and 2-fold nodes, but they now intersect with the 5-fold axes at the center of the icosahedron to yield 532 point-group symmetry. Although the interactions made between CPs at the 5-fold axes are not equivalent to those they make at the 6-fold nodes – termed quasi 6-fold (Q6) axes as they are now local – they were termed quasi-equivalent as they involve the same interacting surfaces of the protein. Fig. 5(B) shows that there are other ways of transforming a planar lattice into a spherical lattice with 532 point-group symmetry. In particular, the distortions introduced to the planar lattice in Fig. 5(B) indicate a way of transforming a planar lattice of triskelions in which each subunit displays equivalent interactions, into a spherical lattice in which the interactions are non-equivalent. The resulting pattern, instead of corresponding to a truncated icosahedron, corresponds to a polyhedron termed rhombicosidodecahedron (Twarock and Luque, 2019) that recapitulates all the observed interactions of the neutral pH lattice.

The triangulation number T of the icosahedral lattice was defined by Caspar and Klug as $T = h^2 + k^2 + hk$, i.e., as the square of the distance along the edges of the planar hexagonal lattice of consecutive 6-fold nodes transformed into 5-fold vertices to yield an icosahedral lattice. h and k are thus integers counting the number of Q6 nodes in between adjacent vertices along the two directions of the triangulated sphere (Fig. 6(A)). Lattices in which the vertices are along a single direction (i.e., one of the indices is null, h or $k = 0$, or are along the diagonal, $h = k$), are centrosymmetric, whereas all other combination of indices lead to a chiral lattice. As pointed out by Rochal and collaborators (Rochal *et al.*, 2016), the Q6 nodes are not real nodes of the lattice, yet the Caspar and Klug construction required each individual CP not to occupy these “special” positions. As illustrated in Fig. 6(B), such nodes are not forbidden, and an icosahedral lattice can be constructed by placing CPs over Q6 nodes. The resulting pattern is termed “spherical lattice” (SL).

An illustration of the principles of Rochal *et al.* is provided in Fig. 6. In the basic $T = 28$ l and $T = 19$ d icosahedral lattices, respectively corresponding to the spherical lattices SL<4,2> and SL<3,2> (Fig. 6(A)), the original 6-fold nodes of the planar P6 lattice used to derive it are shown as empty blue circles. The Caspar and Klug formulation allows tiling the lattices by placing a capsid protein only at general positions, thereby accounting for 6 CPs around each Q6 node and five about the I5 vertices, all displaying quasi-equivalent interactions (Fig. 6(B)). The resulting tiling has 60T copies of the CP making a closed particle (i.e., 1680 proteins on the $T = 28$ l lattice, and 1140 in the $T = 19$ d lattice). Positioning instead a CP on the Q6 nodes yields an icosahedral lattice in which four CPs (instead of 28) interact to form a closed particle in the 28 l lattice, and three instead of 19 in the 19 d lattice. The CPs with a unique environment are represented in Fig. 6(C) in different colors and occupy the nodes containing a full CP hexamer in the Caspar and Klug construction (panel B). The flavivirus surface lattice can be understood in terms of the two lattices represented in Fig. 6, the immature particle as a $T = 28$ l lattice in which the yellow position about the 5-fold axes remains vacant, and the mature lattice (or the immature lattice at acid pH) as a fully occupied $T = 19$ d lattice. The organization of the flavivirus envelope proteins on these lattices is indicated with a purple outline in Fig. 6(D), in which each CP is assimilated to a prM/E heterodimeric protomer. In Fig. 6(E), a (prM/E)₃ triskelion is positioned occupying a third of one of the 20 icosahedral facets, while three (M/E)₂ dimers are positioned to cover one of the 30 diamond-like facets of the tricontahedron corresponding to the mature flavivirus particle.

Thermodynamic Transition Between Icosahedral Surface Lattices

In their generalization of the Caspar and Klug geometric model, Rochal and colleagues went a step further, and showed that Landau’s “weak crystallization theory” could be applied to “crystallization” on an icosahedral surface lattice. This theory uses thermodynamic considerations to describe the first order transition from a liquid state to a crystalline state, and can be decomposed in terms of wave vectors (Kats *et al.*, 1993). Rochal and colleagues showed that the 532 point-symmetry group imposes important restrictions to the wave number l such that the spherical harmonics corresponding only to certain l values are not null. They showed that the allowed wave numbers are given by the series $l = 15 + 6i + 10j$, where i and j are integers or zero. The first l values in this series are: $l = (15, 21, 25, 27, 31, 33, 35, \text{etc.})$, each of which leads to a characteristic “irreducible icosahedral density function” representing the weak crystallization energy Δ for the corresponding icosahedral lattice. In particular, Rochal *et al.* showed that the lattices SL<4,2> and SL<3,2> presented in Fig. 6 are characterized by wave numbers $l = 31$ and $l = 25$, respectively. The density function corresponding to these two wave numbers is represented as a function of the angular variables

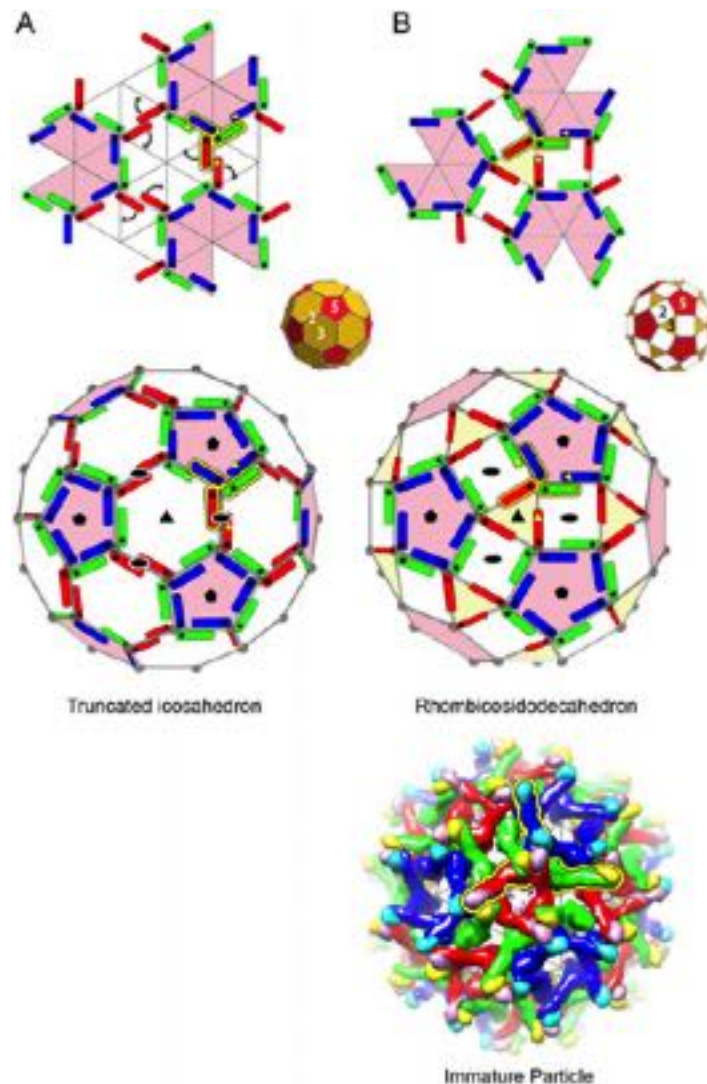


Fig. 5 Derivation of the icosahedral surface lattice of the immature flavivirus particle. (A) The top panel shows an example of the Caspar and Klug geometrical construction to generate an icosahedral lattice from a planar hexagonal lattice with triangulation $T = 3$ – i.e., with three CPs in the icosahedral asymmetric unit (IAU). Each CP is represented by a rectangle colored differently in the IAU, where they are arranged by groups of three as in a triskelion. A black dot on each rectangle indicates the side where the α -helical TM hairpins cluster. One triskelion is highlighted with a yellow outline, and small yellow triangles mark the blades of three adjacent triskelions that form a spike on top of the highlighted one. Selected nodes of the planar lattice (shown in a pink background) are converted into 5-fold axes, creating a 5-fold vertex. Applied in a regular way – in this case, leaving one 6-fold node (or Q6 axis) in between – this process results in the construction of a symmetrical closed particle. In this case, the Q6 and I3 axes coincide. The resulting closed particle is associated to the truncated icosahedron illustrated in between the two panels – with 20 hexagonal facets normal to the Q6 axes, and 12 pentagonal facets normal to the I5 axes. The I2 axes intersect the edges between hexagonal facets. In the planar lattice, all the interactions between the CPs are equivalent: the 2-fold interactions between green and blue are the same as those between two red ones. In the truncated icosahedron drawn below, only the red CPs maintain strict 2-fold interactions while the green and blue interact about a local 2-fold axis. The interactions between blue/green CP dimers along the I5 pentamer edges are quasi-equivalent to those between red/red and blue/green dimers at the edges of the Q6 hexamers. This lattice corresponds to triangulation $T = 3$, as defined in the legend to Fig. 6. The curved arrows in the top panel show the changes required to reach the arrangement of panel B. (B) Alternative distortions to the planar hexagonal lattice introduced to generate a particle organized as the immature flavivirus virions. In this case, before turning the pink hexagonal facets into pentamers, they are brought out of plane by turning the 6-fold node relating them in the P6 lattice into a 3-fold axis. For this purpose, the pink hexagonal facets are rotated by 30 degrees anticlockwise, breaking the 2-fold inter-vertex contacts between red CPs in the planar lattice and bringing them into forming different inter-vertex contacts relating triskelions about the I3 axes. The required transformation in 3-dimensions is more complex, as it brings out of plane the interactions about I3, I2 and I5 axes. The overall transformation is partially indicated by the curved arrows in panel A. The resulting interactions between CPs become non-equivalent. The triskelions in the planar lattice are postulated to have 3-fold symmetry, while those in the icosahedral lattice deviate from it, as shown in Fig. 3(A). The Archimedean solid associated with the resulting lattice is termed “rhombicosidodecahedron”, illustrated in the inset and in the second panel. It has 12 faces normal to the 5-fold axes (pink), 20 triangular faces normal to the 3-fold axes (yellow), and 30 rhombic faces normal to the 2-fold axes (white). The lower panel displays the immature particle in the same orientation, with the same triskelion highlighted, buried under a spike made from the blades indicated by the yellow triangles in the middle panel. Note that a similar arrangement of the triskelions could be reached using the $T = 3$ lattice as in (A), except that it would leave a large space surrounded by 6 triskelions at the particle’s 3-fold axes, in an arrangement that would contain blue pentamers about the I5 and red/green hexamers about the I3 axes. Also, the triskelions would display quasi-equivalent interactions about the 2-fold axes (red-red, at the edges between hexamers) and quasi 2-fold (blue-green) about the edges between pentamers and hexamers. These interactions are clearly absent in the flavivirus immature particle.

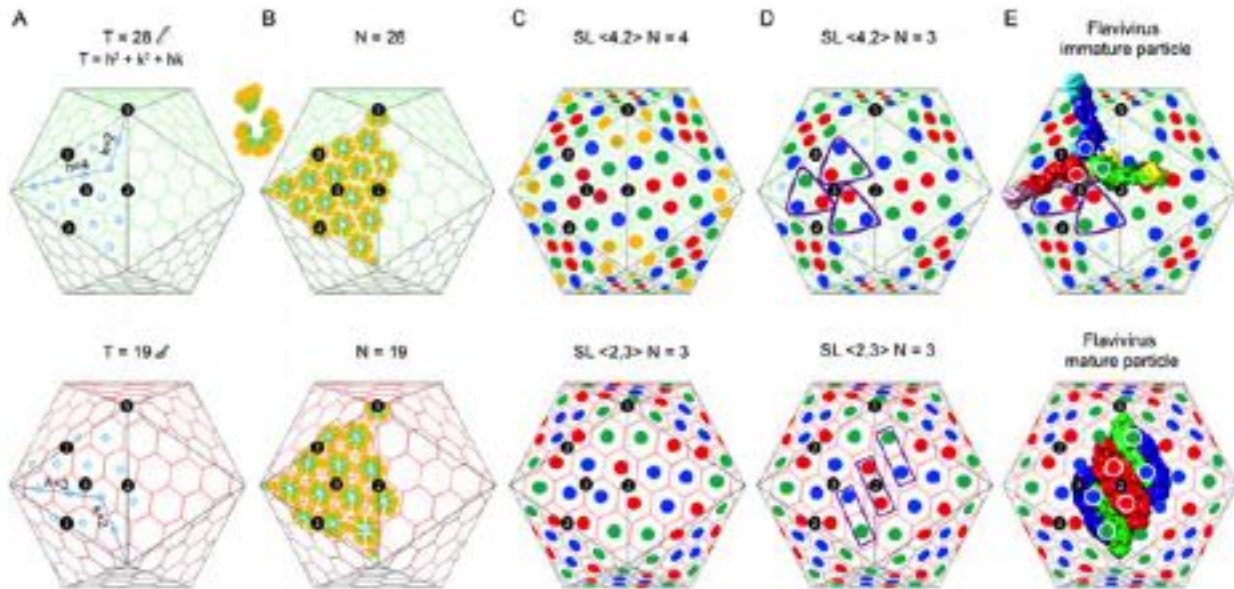


Fig. 6 Triangulation and derivation of spherical lattices. (A) Definition of triangulation and two examples. The top panel shows an icosahedral lattice with $T = 28$ with indices $h = 4$ and $k = 2$ (top panel) and $T = 28$ with $h = 3$ and $k = 2$. The triangulation T number, $T = h^2 + k^2 + hk$, is indicated. Each unit cell in the virtual planar lattice contains 6 CPs, and all CPs are in an equivalent environment. Empty blue circles denote the 6-fold nodes that remain as Q6 axes. The examples show two chiral icosahedral lattices: in the top panel, it is laevo as the indices leading from one vertex to the next are 4 nodes along h and 2 to the left along k . Similarly, in the bottom panel, the lattice is dextro as there are 3 nodes along h followed by 2 to the right along k . The corresponding operations are shown by thin blue arrows in both panels, and the indices are labeled. (B) The geometric model of Caspar and Klug with 6 CPs per unit cell of the virtual planar lattice and 5 around the I5 vertices, making CP hexamers and pentamers via quasi-equivalent interactions. Each CP is represented by an asymmetric shape colored yellow and green (see inset at the top-left), with a green thin end pointing toward the Q6 or I5 axes. One of the 20 facets of the resulting icosahedron is tiled in this way, the IAU being 1/3rd of the facet. In the top panel, the IAU contains 28 CPs organized as 4 full hexamers, one half-hexamer by the I2 axes and one fifth of a pentamer (1 CP) by the I5 axes. In the bottom panel, the IAU contains 19 CPs organized as three full hexamers and one fifth of a pentamer by the I5 axes. (C) The revised Caspar and Klug geometric construction proposed by Rochal and colleagues (Rochal *et al.*, 2016) uses the same considerations to generate an icosahedral lattice, which is defined as a spherical lattice ($SL < h, k >$) corresponding to the same triangulation indices. In the example of this Figure, the lattices are termed $SL < 4, 2 >$ (top) and $SL < 2, 3 >$ (bottom), as labeled. Note that the order of the indices affects the chirality, as $SL < h, k >$ denotes a lattice that is the $SL < k, h >$ enantiomorph. The difference in the new construction is that the individual CPs are placed on the nodes of the planar lattice, except in those nodes that become special positions (for instance, the node coincident with the I2 axes in Fig. 6(C), top panel). The CPs are shown as colored dots on the lattice: there are 4 CPs in the IAU of the $SL < 4, 2 >$ ($N=4$, corresponding to the four full hexamers in B, top panel) and 3 in the $SL < 2, 3 >$ ($N=3$, corresponding to the three full hexamers in B, bottom panel). The special positions, i.e., the nodes coinciding with the I2 axis, occupied by half a hexamer in the Caspar and Klug model, or the I5 axis, are left vacant (panel B). The icosahedrons displayed here are drawn at the same scale, but the one in panel B would be substantially larger, given the number of CPs per facet. Note that the SL construction is equivalent to the Caspar and Klug $T = 4$ icosahedral lattice. Similarly, an $SL < 4, 1 >$, resulting in $N=3$ (not shown in the Figure,) would be equivalent to a $T = 3$ in the Caspar and Klug construction (displayed in Fig. 5(A)). The Caspar and Klug construction and the quasi-equivalence principle thus represent a special case, encompassed by the construction presented by Rochal *et al.* (D) The organization of the flavivirus immature particle corresponds to an $SL < 4, 2 >$ lattice in which one of the allowed positions is vacant (the node around the 5-fold axes, top row), i.e., with $N = 3$. The base of the triskelions in one icosahedral facet is outlined. The mature particle corresponds to the $SL < 2, 3 >$, as noted previously (Rochal *et al.*, 2016). The rectangles drawn on this lattice indicate the (M/E)₂ dimers forming the mature particle (lower panel). (E) Top panel: one (prM/E)₃ triskelion (corresponding to one icosahedral asymmetric unit) displayed in surface representation and placed to scale on the $SL < 4, 2 >$ lattice. Bottom panel: three (M/E)₂ dimers (forming two icosahedral asymmetric units) placed to scale on the $SL < 2, 3 >$ lattice.

on a spherical coordinate system in Fig. 7(A). They correspond to the probability distribution of the CPs “crystallizing” in this lattice. As a guide, we have outlined with a black triangle the location of one triskelion in the $l = 31$ plot, and a magenta rhombic tile comprising the three dimers of a raft in the mature particle in the $l = 25$ plot. The middle top panel shows a superposition of the two plots, with arrows indicating the transitions the CP (in our case, a prM/E protomer) would undergo to switch from the probability maximum in one lattice to that in the other.

Inspection of the immature lattice at neutral pH further suggests an order of events for the transition to the low pH lattice, in particular because the blades of the I3 protomers are trapped in the spike in between the I2 (green/yellow) and I5 (blue/cyan) blades (Table 1) of the local dimer (Fig. 7), which need to come apart for the transition to proceed. As the contacts in this lattice become destabilized by protonation, the local dimer can re-organize as indicated in Fig. 4 to reach an alternative energy minimum illustrated by the different pattern of the probability distribution displayed in Fig. 7(A). The I3 blades (red/pink) are left to rearrange around the icosahedral 2-fold axes to form the icosahedral dimers of the acid pH lattice. The overall shift is circular: the I5 protomers (cyan/blue) become I3, the I2 ones (yellow/green) become I5, and the I3 (pink/red) become I2, as shown by the permutation of colors around the icosahedral axes in Fig. 7B, left compared to right.

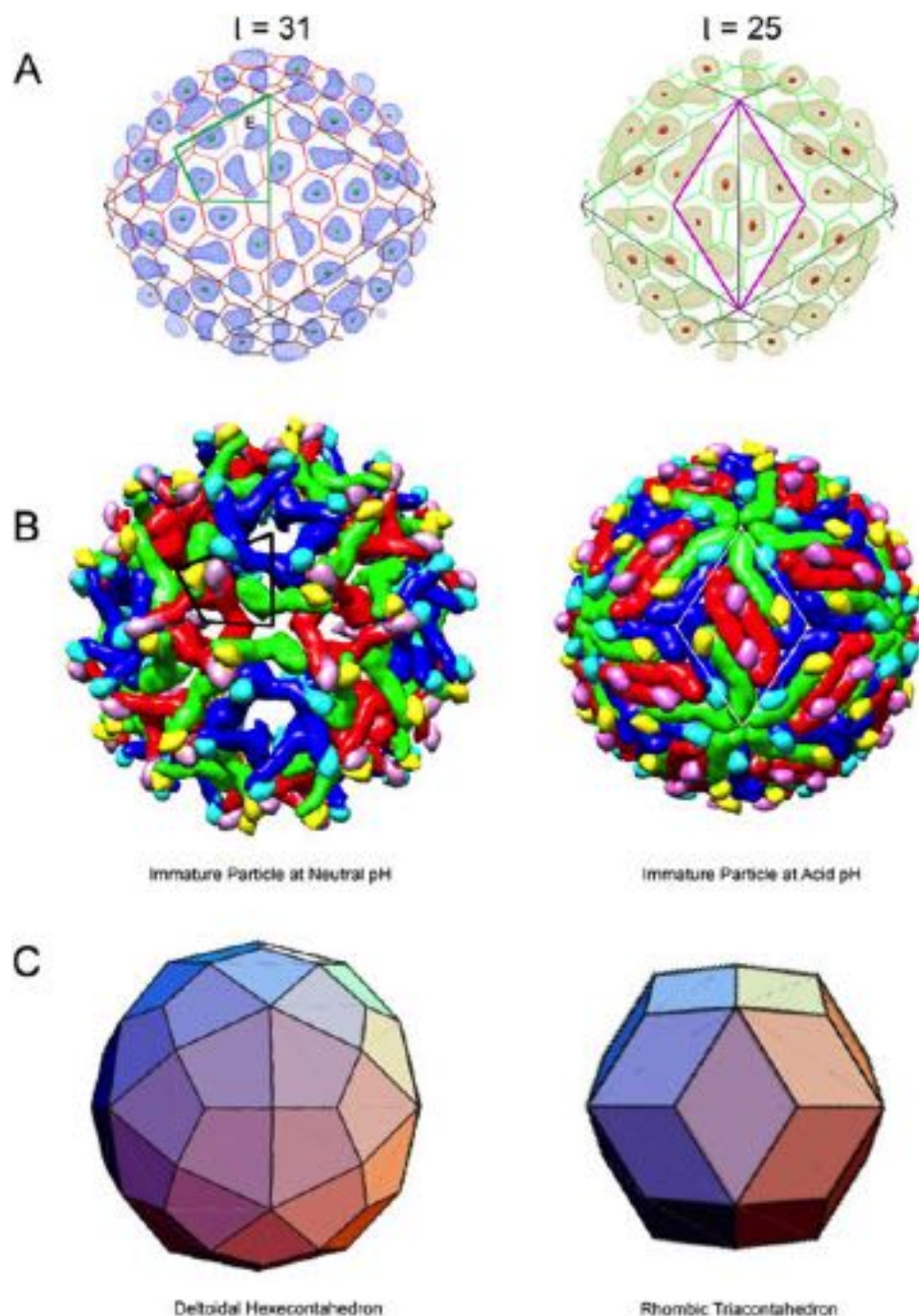


Fig. 7 Thermodynamics of the low pH-induced lattice transition. (A) Plot of the icosahedral density functions $\Delta\rho(\theta, \varphi)$ for wave numbers $l = 31$ (left) and $l = 25$ (right), respectively superposed to the matching $SL\langle 4,2\rangle$ and $SL\langle 2,3\rangle$ lattices. A mesh shows the probability distribution of protomers in the lattice, and dots indicate the highest peaks. In the left panel, a deltoidal green outline indicates the asymmetric unit of the neutral pH lattice (an "E" labels the peak corresponding to the position left empty in the lattice (see Fig. 6(C)-(D))). In the right panel, a magenta diamond shape outlines two asymmetric units about the I_2 axis, matching the rafts of the mature flavivirus particles. (B) The left and right panels show the immature particle at neutral pH and at acid pH, respectively, with the protomers shown in surface representation and colored as in Fig. 2. The asymmetric units are outlined as in A. (C) Polyhedra matching the icosahedral organization of the immature lattice at neutral pH (left) and at acid pH (right). Note the deltoidal and rhombic faces matching the arrangements shown in A and B. The deltoidal hexecontahedron is the dual of the rhombicosidodecahedron drawn in Fig. 5(B). (Two polyhedra are dual when the faces of one correspond to the vertices of the other, and vice versa).

Discussion

Despite remarkable progress in understanding the structural organization of the flavivirus particles achieved in the past few decades, multiple questions remain unanswered. Understanding in detail the conformational change undergone by the immature particle upon exposure to acid pH remains a major challenge. The identified close relation between the two corresponding icosahedral lattices provides new clues as it suggests that upon protonation, the destabilization of the contacts made a neutral pH puts each prM/E protomer down the slope to reach a lowest energy location – corresponding to a maximum in the order parameter density function plotted in Fig. 7 to make the required contacts in the alternative lattice. And vice versa when the pH is raised again in the absence of furin cleavage. The fact that the initial SL<4,2> lattice leaves one position vacant likely facilitates the transition from one lattice to the other by avoiding potential steric interference between protomers.

In summary, the flavivirus exit pathway involves a fascinating set of macromolecular interactions which are not understood today. We hope that this review will have helped bring up the relevant questions to be addressed in the years to come. The advent of correlative light and electron microscopy, allowing the visualization of complex processes within cells, will be key to provide further insight into these complex macromolecular processes.

Acknowledgement

We thank Dr. J. Navaza for helpful discussions and providing bioinformatic tools, and Dr. F. Coulibaly for sharing unpublished data.

References

- Akey, D.L., Brown, W.C., Dutta, S., *et al.*, 2014. Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Science* 343, (6173), 881–885. doi:10.1126/science.1247749.
- Allison, S.L., Schlich, J., Stiasny, K., *et al.*, 1995a. Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. *Journal of Virology* 69, 695–700.
- Allison, S.L., Stadler, K., Mandl, C.W., Kunz, C., Heinz, F.X., 1995b. Synthesis and secretion of recombinant tick-borne encephalitis virus protein E in soluble and particulate form. *Journal of Virology* 69 (9), 5816–5820. doi:10.1128/JVI.69.9.5816-5820.1995.
- Bressanelli, S., Stiasny, K., Allison, S.L., *et al.*, 2004. Structure of a flavivirus envelope glycoprotein in its low-pH-induced membrane fusion conformation. *EMBO Journal* 23, 728–738.
- Byk, L.A., Gamarnik, A.V., 2016. Properties and functions of the dengue virus capsid protein. *Annual Review of Virology* 3, 263–281.
- Caspar, D.L., Klug, A., 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposia on Quantitative Biology* 27, 1–24.
- Dokland, T., Walsh, M., Mackenzie, J.M., *et al.*, 2004. West Nile virus core protein; tetramer structure and ribbon formation. *Structure* 12, 1157–1163.
- Fuzik, T., Formanova, P., Ruzek, D., *et al.*, 2018. Structure of tick-borne encephalitis virus and its neutralization by a monoclonal antibody. *Nature Communications* 9, 436.
- Gutsche, I., Coulibaly, F., Voss, J.E., *et al.*, 2011. Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein. *Proceedings of the National Academy of Sciences of the United States of America* 108, 8003–8008.
- Hasan, S.S., Sevwana, M., Kuhn, R.J., Rossmann, M.G., 2018. Structural biology of Zika virus and other flaviviruses. *Nature Structural & Molecular Biology* 25, 13–20.
- Henderson, R., 2015. Overview and future of single particle electron cryomicroscopy. *Archives of Biochemistry and Biophysics* 581, 19–24.
- Huang, Y.S., Higgs, S., Vanlandingham, D.L., 2019. Emergence and re-emergence of mosquito-borne arboviruses. *Current Opinion in Virology* 34, 104–109.
- Jones, C.T., Ma, L., Burgner, J.W., *et al.*, 2003. Flavivirus capsid is a dimeric alpha-helical protein. *Journal of Virology* 77, 7143–7149.
- Junjhon, J., Pennington, J.G., Edwards, T.J., *et al.*, 2014. Ultrastructural characterization and three-dimensional architecture of replication sites in dengue virus-infected mosquito cells. *Journal of Virology* 88, 4687–4697.
- Kats, E.I., L, V.V., M, A.R., 1993. Weak crystallization theory. *Physics Reports* 228, 1–91.
- Kielian, M., Rey, F.A., 2006. Virus membrane-fusion proteins: More than one way to make a hairpin. *Nature Reviews Microbiology* 4, 67–76.
- Kostyuchenko, V.A., Chew, P.L., Ng, T.S., Lok, S.M., 2014. Near-atomic resolution cryo-electron microscopic structure of dengue serotype 4 virus. *Journal of Virology* 88, 477–482.
- Kostyuchenko, V.A., Lim, E.X., Zhang, S., *et al.*, 2016. Structure of the thermally stable Zika virus. *Nature* 533, 425–428.
- Kostyuchenko, V.A., Zhang, Q., Tan, J.L., Ng, T.S., Lok, S.M., 2013. Immature and mature dengue serotype 1 virus structures provide insight into the maturation process. *Journal of Virology* 87, 7700–7707.
- Kuhn, R.J., Zhang, W., Rossmann, M.G., *et al.*, 2002. Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion. *Cell* 108, 717–725.
- Lee, B., Richards, F.M., 1971. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology* 55, 379–400.
- Li, L., Lok, S.M., Yu, I.M., *et al.*, 2008. The flavivirus precursor membrane-envelope protein complex: Structure and maturation. *Science* 319, 1830–1834.
- Li, M.Y., Grandadam, M., Kwok, K., *et al.*, 2015. KDEL receptors assist dengue virus exit from the endoplasmic reticulum. *Cell Reports* 10, 1496–1507.
- Liao, M., Sanchez-San Martin, C., Zheng, A., Kielian, M., 2010. In vitro reconstitution reveals key intermediate states of trimer formation by the dengue virus membrane fusion protein. *Journal of Virology* 84, 5730–5740.
- Luca, V.C., Nelson, C.A., Fremont, D.H., 2013. Structure of the St. Louis encephalitis virus postfusion envelope trimer. *Journal of Virology* 87, 818–828.
- Ma, L., Jones, C.T., Groesch, T.D., Kuhn, R.J., Post, C.B., 2004. Solution structure of dengue virus capsid protein reveals another fold. *Proceedings of the National Academy of Sciences of the United States of America* 101, 3414–3419.
- Mackenzie, J.M., Jones, M.K., Young, P.R., 1996. Immunolocalization of the dengue virus nonstructural glycoprotein NS1 suggests a role in viral RNA replication. *Virology* 220, 232–240.
- McMullan, G., Faruqi, A.R., Henderson, R., 2016. Direct electron detectors. *Methods in Enzymology* 579, 1–17.
- Modis, Y., 2013. Class II fusion proteins. *Advances in Experimental Medicine and Biology* 790, 150–166.
- Modis, Y., Ogata, S., Clements, D., Harrison, S.C., 2003. A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America* 100, 6986–6991.
- Modis, Y., Ogata, S., Clements, D., Harrison, S.C., 2004. Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427, 313–319.
- Mukhopadhyay, S., Kuhn, R.J., Rossmann, M.G., 2005. A structural perspective of the flavivirus life cycle. *Nature Reviews Microbiology* 3, 13–22.
- Pierson, T., Diamond, M., 2013. Flaviviruses. In: Howley, D.M.K.A.P.M. (Ed.), *Fields Virology*. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 747–794.

- Prasad, V.M., Miller, A.S., Klose, T., *et al.*, 2017. Structure of the immature Zika virus at 9 Å resolution. *Nature Structural & Molecular Biology* 24, 184–186.
- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375, 291–298.
- Rice, C.M., Lenches, E.M., Eddy, S.R., *et al.*, 1985. Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. *Science* 229, 726–733.
- Rochal, S.B., Konevtsova, O.V., Myasnikova, A.E., Lorman, V.L., 2016. Hidden symmetry of small spherical viruses and organization principles in "anomalous" and double-shelled capsid nanoassemblies. *Nanoscale* 8, 16976–16988.
- Scaturro, P., Cortese, M., Chatel-Chaix, L., Fischl, W., Bartenschlager, R., 2015. Dengue virus non-structural protein 1 modulates infectious particle production via interaction with the structural proteins. *PLoS Pathogens* 11, e1005277.
- Sevvana, M., Long, F., Miller, A.S., *et al.*, 2018. Refinement and analysis of the mature Zika virus cryo-EM structure at 3.1 Å resolution. *Structure* 26, 1169–1177.
- Shang, Z., Song, H., Shi, Y., Qi, J., Gao, G.F., 2018. Crystal structure of the capsid protein from Zika virus. *Journal of Molecular Biology* 430, 948–962.
- Sirohi, D., Chen, Z., Sun, L., *et al.*, 2016. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* 352, 467–470.
- Stadler, K., Allison, S.L., Schalich, J., Heinz, F.X., 1997. Proteolytic activation of tick-borne encephalitis virus by furin. *Journal of Virology* 71, 8475–8481.
- Stiasny, K., Allison, S.L., Mandl, C.W., Heinz, F.X., 2001. Role of metastability and acidic pH in membrane fusion by tick-borne encephalitis virus. *Journal of Virology* 75, 7392–7398.
- Stiasny, K., Kossli, C., Lepault, J., Rey, F.A., Heinz, F.X., 2007. Characterization of a structural intermediate of flavivirus membrane fusion. *PLoS Pathogens* 3.
- Tan, T.Y., Fibriansah, G., Kostyuchenko, V.A., *et al.*, 2020. Capsid protein structure in Zika virus reveals the flavivirus assembly process. *Nature Communications* 11 (1), 895. doi:10.1038/s41467-020-14647-9.
- Tay, M.Y.F., Vasudevan, S.G., 2018. The transactions of NS3 and NS5 in flaviviral RNA replication. *Advances in Experimental Medicine and Biology* 1062, 147–163.
- Therkelsen, M.D., Klose, T., Vago, F., *et al.*, 2018. Flaviviruses have imperfect icosahedral symmetry. *Proceedings of the National Academy of Sciences of the United States of America* 115, 11608–11612.
- Twarock, R., Luque, A., 2019. Structural puzzles in virology solved with an overarching icosahedral design principle. *Nature Communications* 10 (1), 4414. doi:10.1038/s41467-019-12367-3.
- Wang, X., Li, S.H., Zhu, L., *et al.*, 2017. Near-atomic structure of Japanese encephalitis virus reveals critical determinants of virulence and stability. *Nature Communications* 8, 14.
- Welsch, S., Miller, S., Romero-Brey, I., *et al.*, 2009. Composition and three-dimensional architecture of the dengue virus replication and assembly sites. *Cell Host & Microbe* 5, 365–375.
- Youn, S., Ambrose, R.L., Mackenzie, J.M., Diamond, M.S., 2013. Non-structural protein-1 is required for West Nile virus replication complex formation and viral RNA synthesis. *Virology Journal* 10, 339.
- Youn, S., Li, T., McCune, B.T., *et al.*, 2012. Evidence for a genetic and physical interaction between nonstructural proteins NS1 and NS4B that modulates replication of West Nile virus. *Journal of Virology* 86, 7360–7371.
- Yu, I.M., Zhang, W., Holdaway, H.A., *et al.*, 2008. Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* 319, 1834–1837.
- Zhang, X., Ge, P., Yu, X., *et al.*, 2013. Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. *Nature Structural & Molecular Biology* 20, 105–110.
- Zhang, Y., Corver, J., Chipman, P.R., *et al.*, 2003. Structures of immature flavivirus particles. *EMBO Journal* 22, 2604–2613.
- Zhang, Y., Kaufmann, B., Chipman, P.R., Kuhn, R.J., Rossmann, M.G., 2007. Structure of immature West Nile virus. *Journal of Virology* 81, 6141–6145.
- Zheng, A., Yuan, F., Kleinfelter, L.M., Kielian, M., 2014. A toggle switch controls the low pH-triggered rearrangement and maturation of the dengue virus envelope proteins. *Nature Communications* 5, 3877.

Reoviruses (*Reoviridae*) and Their Structural Relatives

Liya Hu, Mary K Estes, and B V Venkataram Prasad, Baylor College of Medicine, Houston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

Introduction

The double-stranded RNA (dsRNA) viruses are non-enveloped icosahedral viruses with multiple concentric icosahedral capsid layers. These viruses are classified into twelve families displaying a wide host range, including bacteria, fungi, plants, animals, and humans. The *Reoviridae* represent the largest and most diverse dsRNA virus family. Based on the structural organizations of the viral particles, members of the *Reoviridae* are grouped into two subfamilies, the non-turreted *Sedoreovirinae* and the turreted *Spinareovirinae*. The members of *Sedoreovirinae* have a smooth outermost layer short viral spikes and includes six genera (*Cardoreovirus*, *Mimoreovirus*, *Orbivirus*, *Phytoreovirus*, *Rotavirus*, and *Seadornavirus*). The characteristic feature of the members of *Spinareovirinae* is that they have large turreted structures at the 5-fold axes of the icosahedral particles. This subfamily consists of nine genera (*Aquareovirus*, *Coltivirus*, *Cypovirus*, *Dinovernavirus*, *Fijivirus*, *Idnoreovirus*, *Mycoreovirus*, *Orthoreovirus*, and *Oryzavirus*) (Dermody, 2013). Structures of many dsRNA viruses have been determined by X-ray crystallography and/or cryo-electron microscopy (cryo-EM) techniques, providing molecular insights into their assembly and replication mechanisms. Here, we review the similar and contrasting structural characteristics of dsRNA viruses with a focus on the members of the *Reoviridae*, including some novel features that are revealed by recent asymmetric analysis using cryo-EM.

Overall Structure of the *Reoviridae* Family Viruses

The Non-Turreted Subfamily *Sedoreovirinae*

Rotavirus

Rotavirus, the leading cause of diarrheal diseases in animals and young children worldwide, is a large icosahedral virus with a diameter of ~100 nm (Fig. 1(A)). The rotavirus genome is composed of 11 dsRNA segments encoding six viral structural proteins (VP1–VP4, VP6–VP7) (Table 1) and six non-structural proteins (NSP1–NSP6) (Estes, 2013). The VPs form three concentric capsid layers, while the NSPs are not present in the mature viral particles yet are critical for virus replication. The outermost layer that is composed of glycosylated capsid protein VP7 and the middle layer composed of VP6 both assemble with $T = 13$ (*laevo*) icosahedral symmetry (Prasad *et al.*, 1988). The protease-sensitive spike protein VP4 spans the VP7 and VP6 capsid layers and protrudes from the virion surface. The VP7 and VP4 proteins define the G- and P-genotypes of rotaviruses, respectively. The innermost capsid layer is formed by 120 copies of VP2 and exhibits $T = 1$ icosahedral symmetry (Prasad *et al.*, 1996; Li *et al.*, 2009; McClain *et al.*, 2010; Estrozi *et al.*, 2013; Settembre *et al.*, 2011) (Fig. 2(A)). The RNA-dependent RNA polymerase (RdRp) VP1 and the capping enzyme VP3 form a heterodimer that is anchored near the 5-fold vertices beneath the VP2 core layer.

Orbivirus

Bluetongue virus (BTV), the prototype member of the genus *Orbivirus*, is an important insect vectored pathogen of wild ruminants and livestock that causes significant economic losses in many counties (Roy, 2013). BTV has a genome of 10 dsRNA segments, which encodes seven structural proteins (VP1–VP7) (Table 1) and four non-structural proteins (NS1, NS2, NS3, and NS3A). The overall organization of the virion that is 90 nm in diameter with three capsid layers is similar to that of rotavirus except for the outer capsid layer (Fig. 1(B)). The outer capsid layer is formed by trimers of VP2 and VP5 interdigitating with one another; the middle layer that is similar to the rotavirus VP6 layer contains VP7 trimers; the innermost core layer constitutes 120 subunits of the capsid protein VP3 with the RdRp VP1, the mRNA capping enzyme VP4, and the putative helicase VP6 attached at the inner surface of VP3 (Stauber *et al.*, 1997; Sutton *et al.*, 2007; Roy, 2017; Zhang *et al.*, 2016; Grimes *et al.*, 1998; Nason *et al.*, 2004; He *et al.*, 2019).

Phytoreovirus

Rice dwarf virus (RDV) is a plant virus of the genus *Phytoreovirus* and causes rice dwarf disease that reduces crop production resulting in economic loss in Asia. The viral genome consists of 12 dsRNA segments encoding seven structural (P1, P2, P3, P5, P7, P8, and P9) (Table 1) and five non-structural (Pns4, Pns6, Pns10, Pns11, and Pns12) proteins (Nakagawa *et al.*, 2018). In contrast to rotavirus and orbivirus, the RDV particles are slightly smaller, with ~70 nm in diameter (Fig. 1(C)) and contain a double-layered capsid shell (Lu *et al.*, 1998). The outer shell is formed by the major protein P8 and the minor proteins P2 and P9. There are 780 copies of the P8 protein, which as 260 trimers, follow the $T = 13$ icosahedral symmetric organization (Fig. 2(B)). The inner shell, similar to VP2 and VP3 capsid layers in rotavirus and orbivirus, respectively, is composed of 120 molecules of P3, the RdRp P1, the capping enzyme P5, and the nucleic acid-binding protein P7. The 3.5 Å crystal structure of an RDV particle has provided the structural details of the capsid proteins (Nakagawa *et al.*, 2003).

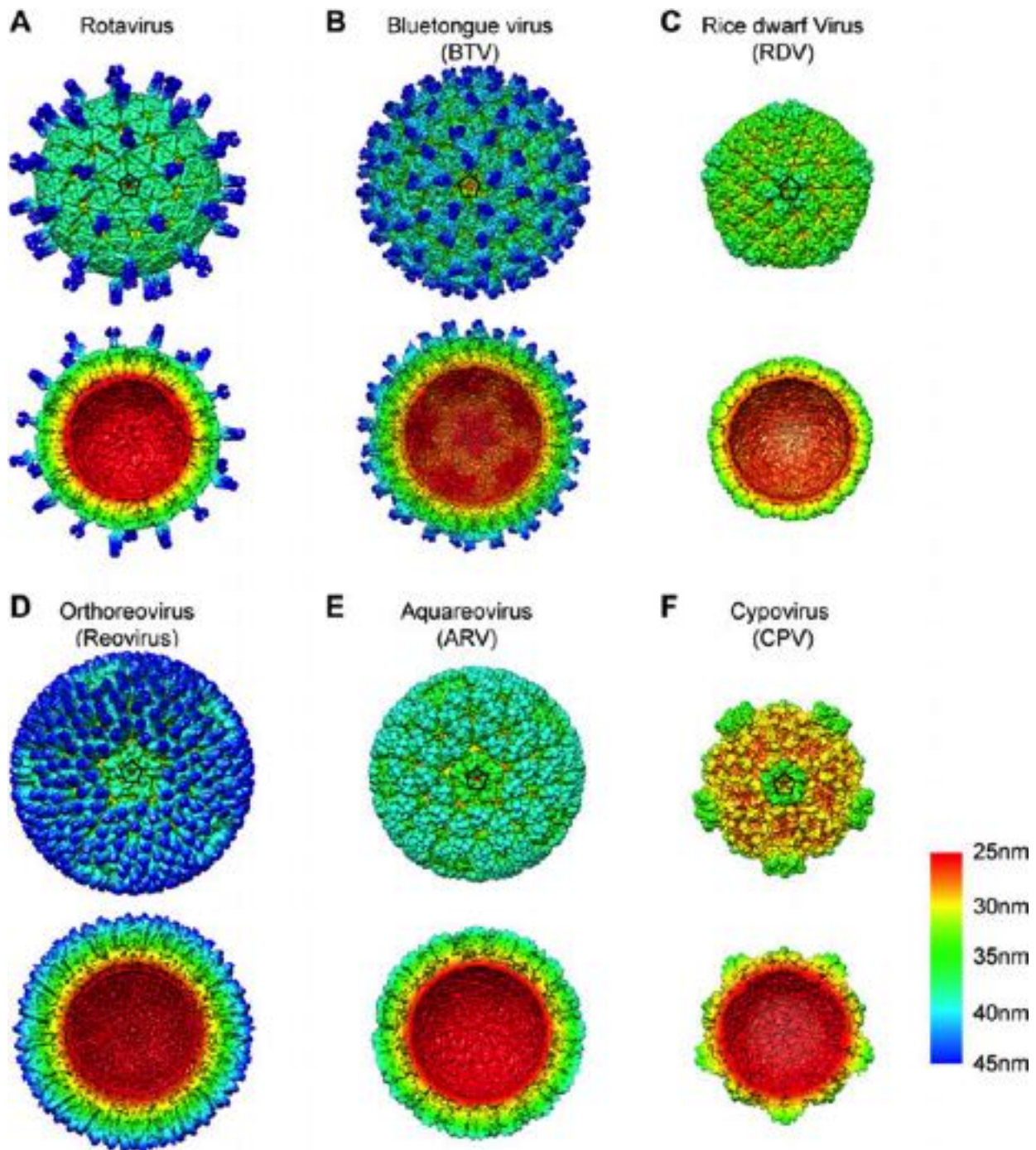


Fig. 1 Structures of dsRNA viruses. The surface representation of each dsRNA virus was viewed along a 5-fold axis, as indicated by the black pentagon symbol. The viruses are radially colored using rainbow colors, from red to blue (25–45 nm), with a cutaway view shown below to reveal the internal density of the virus. (A) Atomic model of rotavirus particle (3.8 Å, Cryo-EM, PDB 4V7Q). (B) Bluetongue virus structure (7 Å, Cryo-EM, EMD-5147). (C) Atomic structure of rice dwarf virus (RDV) (3.5 Å, X-ray diffraction, PDB 1UF2). (D) Reovirus T1L (8.2 Å, Cryo-EM, EMD-8916). (E) Structure of grass carp reovirus, a pathogenic aquareovirus (3.3 Å, Cryo-EM, PDB 5ZVT). (F) Atomic model of transcribing cytoplasmic polyhedrosis virus (CPV) (3 Å, Cryo-EM, PDB 3JAY). Adapted from: (A) Settembre, E.C., Chen, J.Z., Dormitzer, P.R., Grigorieff, N., Harrison, S.C., 2011. Atomic model of an infectious rotavirus particle. *EMBO Journal* 30, 408–416. (B) Zhang, X., Boyce, M., Bhattacharya, B., *et al.*, 2010a. Bluetongue virus coat protein VP2 contains sialic acid-binding domains, and VP5 resembles enveloped virus fusion proteins. *Proceedings of the National Academy of Sciences of the United States of America* 107, 6292–6297. Zhang, X., Jin, L., Fang, Q., Hui, W.H., Zhou, Z.H., 2010b. 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* 141, 472–482. (C) Nakagawa, A., Miyazaki, N., Taka, J., *et al.*, 2003. The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure* 11, 1227–1238. (D) Snyder, A.J., Wang, J.C., Danthi, P., 2019. Components of the reovirus capsid differentially contribute to stability. *Journal of Virology* 93 (2). doi:10.1128/JVI.01894-18 (E) Wang, X., Zhang, F., Su, R., *et al.*, 2018. Structure of RNA polymerase complex and genome within a dsRNA virus provides insights into the mechanisms of transcription and assembly. *Proceedings of the National Academy of Sciences of the United States of America* 115, 7344–7349. (F) Yu, X., Jiang, J., Sun, J., Zhou, Z.H., 2015. A putative ATPase mediates RNA transcription and capping in a dsRNA virus. *Elife*. 4, e07901. doi:10.7554/eLife.07901.

Table 1 Structural proteins of selected reoviruses

Capsid layer	Non-turreted <i>Sedoreovirinae</i>			Turreted <i>Spinareovirinae</i>		
	<i>Rotavirus</i>	<i>Bluetongue virus</i>	<i>Rice dwarf virus</i>	<i>Mammalian orthoreovirus</i>	<i>Aquareovirus</i>	<i>Cypovirus</i>
Outer	VP4 spike	VP2 spike		Attachment protein $\sigma 1$		
	Major capsid protein VP7	Major capsid protein VP5	Major capsid protein P8 and minor capsid proteins P2 and P9	Major capsid proteins $\sigma 3$ and $\mu 1$	Major capsid proteins VP7 and VP5	
Middle	Major capsid protein VP6	Major capsid protein VP7				
Core	Major capsid protein VP2	Major capsid protein VP3	Major capsid protein P3	Major capsid proteins $\lambda 1$ and $\sigma 2$	Major capsid protein VP3 and VP6 clamp	Major capsid protein VP1
	RdRp VP1	RdRp VP1	RdRp P1	RdRp $\lambda 3$ and its cofactor $\mu 2$	RdRp VP2 and its cofactor VP4	RdRp VP2 and NTPase VP4
	RNA capping enzyme VP3	RNA capping enzyme VP4 and the helicase VP6	RNA capping enzyme P5 and RNA binding protein P7	RNA capping/ turret $\lambda 2$	RNA capping/ turret VP1	RNA capping/ turret VP3 and VP5 clamp

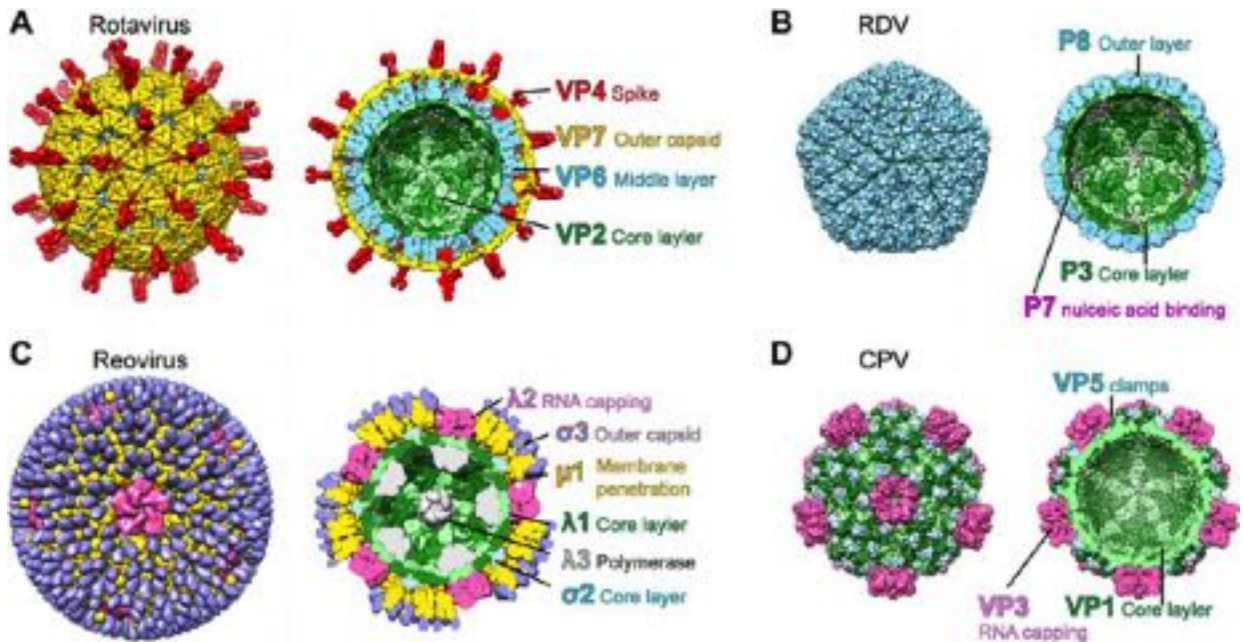


Fig. 2 The structural organization of non-turreted and turreted dsRNA viruses. (A) The triple-layered particle of rotavirus (PDB 4V7Q). The outer capsid protein VP4 and VP7 are colored in red and yellow, respectively. The VP6 protein of the middle layer is in light blue. The VP2A and VP2B of the core layer are shown in light green and forest green, respectively. (B) The double-layered particle of RDV (PDB 1UF2). The core layer protein P3 and the outer layer protein P8 are colored in green and light blue, respectively. A fragment of nucleic acid-binding protein P7 is shown in magenta. (C) Reovirus T1L (PDB 2CSE) (Zhang *et al.*, 2005). The viral proteins are labeled in a cutaway view on the right of the panel. (D) Single-layered cytoplasmic polyhedrosis virus (CPV) (PDB 3JAY). The pentamer of turret protein VP3 is colored in pink. The conformer of the core-shell protein VP1A and VP2B are colored in light green and forest green, respectively. The clamp protein VP5 is in light blue.

The Turreted Subfamily *Spinareovirinae*

Orthoreovirus

Mammalian orthoreovirus (reovirus) infects a variety of mammals, including humans. It has been linked to celiac disease and used as an oncolytic virus (Tenorio *et al.*, 2019; Dermody, 2013). The reovirus genome consists of 10 dsRNA segments: three large (L1–L3), three medium (M1–M3), and four small (S1–S4) segments, encoding lambda (λ), mu (μ), and sigma (σ) proteins, respectively (Dermody, 2013). There are twelve reovirus proteins, including the eight structural proteins ($\lambda 1$, $\lambda 2$, $\lambda 3$, $\mu 1$, $\mu 2$, $\sigma 1$, $\sigma 2$, and $\sigma 3$)

(Table 1) and four non-structural proteins (μ NS, μ NSC, σ 1s, and σ NS). The virus particles are ~ 85 nm in diameter and consist of two capsid layers, an outer capsid layer and an inner core (Fig. 1(D)). The outer capsid layer with a quasi $T = 13$ icosahedral symmetry consists of the membrane-penetration μ 1 protein and its protective cap σ 3 assembled as μ 1 $_3\sigma$ 3 $_3$ heterohexamers (Liemann *et al.*, 2002). The capping enzyme λ 2 forms a pentameric turret at the icosahedral vertices of the reovirus particle that spans through both the inner core and the outer capsid (Fig. 2(C)). Extending from a λ 2 turret at each vertex of the virion is a trimer of the cellular attachment protein σ 1. The reovirus core is ~ 70 nm in diameter and formed by five of the eight viral proteins of the virion: λ 1, λ 2, λ 3, σ 2, and μ 2. The shell contains λ 1 dimers and σ 2 monomers (Reinisch *et al.*, 2000). Minor core proteins, the RdRp λ 3 and its cofactor μ 2, are attached on the inner surface of the core shell around the icosahedral 5-fold vertex.

Aquareovirus

Grass carp reovirus (GCRV), a member of the genus *Aquareovirus* (ARV), causes hemorrhagic diseases in grass carp (*Ctenopharyngodon idella*), resulting in significant economic loss in the aquaculture industry. The genome of ARV contains 11 dsRNA segments that code for seven structural proteins (VP1–VP7) (Table 1) and five non-structural proteins (NSP1–NSP5). The structure of ARV is remarkably similar to that of orthoreovirus with an overall diameter of ~ 75 nm (Fig. 1(E)). The outermost layer is formed by heterodimers of the membrane penetration protein VP5 and the protection cap protein VP7 with an incomplete icosahedral $T = 13$ symmetry. Twelve pentameric VP1 turrets are located at the 5-fold vertices. The core shell is composed of 120 VP3 subunits, which are arranged with $T = 1$ icosahedral symmetry and are clamped together by 120 VP6 monomers. The RdRp VP2 and its cofactor NTPase VP4 associate with each other and within capsid vertices (Cheng *et al.*, 2010; Zhang *et al.*, 2010b; Ding *et al.*, 2018; Wang *et al.*, 2018).

Cypovirus

Cytoplasmic polyhedrosis virus (CPV), or cypovirus for short, has 10 dsRNA segments which encode 10–12 proteins. Unlike other viruses in the *Reoviridae*, which have 2 or 3 capsid layers, CPV with a diameter of ~ 65 nm has a single capsid shell, which makes it a simplified model for the dsRNA viruses (Fig. 1(F)). The structure of CPV has been extensively studied (Hill *et al.*, 1999; Cheng *et al.*, 2011; Yang *et al.*, 2012; Liu and Cheng, 2015; Zhang *et al.*, 2015; Li *et al.*, 2017; Cui *et al.*, 2019). The virus particle contains three major capsid proteins VP1, VP3, and VP5 (Table 1). VP1 forms the capsid shell; VP3 is the capping enzyme; VP5 forms clamps that stabilize the virus capsid, and it also has RNA chaperone-like activity that destabilizes RNA helices and promotes strand annealing in an ATP-independent manner (Yang *et al.*, 2014) (Fig. 2(D)). A transcriptional enzyme complex (TEC) is located at the inner surface of the capsid at the 5-fold axes and consists of two extensively interacting subunits: the RdRp VP2 and the NTPase VP4 (Hill *et al.*, 1999; Zhang *et al.*, 2015).

General Features of the Viral Genome

The genomes of dsRNA viruses generally contain multiple segments except for the members in the *Totiviridae* (e.g., L-A virus), which have a single segment enclosed within the virus core. Members of the *Reoviridae* contain a genome of 9–12 dsRNA segments that are generally monocistronic with each segment encoding one viral protein. However, some segments have additional in-frame initiation codons resulting in the expression of additional viral proteins. The gene segments are packaged in equimolar proportions and vary in length. For example, in rotavirus, the lengths of 11 dsRNA segments range from 0.7–3.3 kb. The positive-stranded (+)RNA of the dsRNA genome and the messenger RNA (mRNA) have a 5'-end cap and a non-polyadenylated 3'-end tail, and the negative-stranded (–)RNA has a diphosphate at its 5'-end. All gene segments have a conserved short 5'- and 3'-end untranslated regions (UTRs), which are important in gene transcription, assortment, and packaging.

Using icosahedral averaging methods, cryo-EM structural studies have shown that a significant portion of the viral genome is ordered and forms concentric layers of density in rotavirus (Prasad *et al.*, 1996; Pesavento *et al.*, 2001), BTV (Gouet *et al.*, 1999), and orthoreovirus cores (Reinisch *et al.*, 2000). More recently, in rotavirus, spherically averaged density shows 8 concentric layers of the dsRNA genome with an average spacing of 28 Å between the layers (Jenni *et al.*, 2019). Pesavento *et al.* observed that the genome of rotavirus could undergo pH-dependent reversible compaction, and their studies suggested that each gene segment is spooled around a transcriptional enzyme complex at the 5-fold vertices such that the segments can be transcribed simultaneously. By contrast, recent cryo-EM using asymmetry reconstruction analyses of CPV and aquareovirus particles revealed that the non-spooled RNA genome forms a liquid-crystalline array within the virus (Liu and Cheng, 2015; Zhang *et al.*, 2015; Wang *et al.*, 2018). With anticipated further technical advances in cryo-EM, future structural studies will likely reveal higher-resolution asymmetric features of the dsRNA genome in other dsRNA viruses.

Structural Organization of the Capsid Layers

The Outer Capsid Layer

Consistent with differing host specificities, the overall structural organization of the outer capsid layer in the *Reoviridae* members show marked variations. In rotavirus, the outermost layer consists of 260 VP7 trimers assembled on a $T = 13$ icosahedral lattice capping the layer below formed by VP6 trimers following the same icosahedral symmetry in such a way that there are solvent channels at all the icosahedral 5-fold and quasi 6-fold axes (Prasad *et al.*, 1988). Another component of the outermost layer is the 60 spikes of VP4 trimers, which are locked in place by VP7 and the VP6 trimers at the quasi 6-fold axes (Fig. 2(A)) (Li *et al.*, 2009;

Settembre *et al.*, 2011). VP7 and VP4 are the main targets for neutralizing antibodies (Prasad *et al.*, 1990; Shaw *et al.*, 1993). The crystal structure of rotavirus VP7 in complex with a Fab of a neutralizing antibody (mAb 4F8) showed how the neutralizing 4F8 Fab stabilizes the VP7 trimers and revealed two Ca^{2+} binding sites at the VP7 inter-subunit contacts (Aoki *et al.*, 2009). The core of rotavirus VP7 folds into two domains, a “Rossmann fold” (domain I) and a jellyroll β sandwich (domain II). The cryo-EM structure of the rotavirus particle showed VP7 uses its the N-terminal arm ($\sim$ residues 51–70) to clamp into the VP6 trimer of the middle capsid layer. It has been proposed that VP7 might bind to $\alpha\beta 2$ integrins using its integrin-binding motif (GPR, residues 253–255). In BTV, the organization of the outer layer, composed of 60 triskelion-shaped VP2 trimers and 120 globular VP5 trimers, is not based on $T = 13$ icosahedral symmetry, and is noticeably different from that of the rotavirus (Zhang *et al.*, 2016).

In turreted reovirus, the outer capsid protein $\mu 1$ and its protective cap $\sigma 3$ form $\mu 1_3\sigma 3_3$ heterohexamers, with three $\sigma 3$ monomers bound on the upper half of $\mu 1$ trimer (Liemann *et al.*, 2002). These heterohexamers are organized with quasi $T = 13$ symmetry because the pentameric $\lambda 2$ turret substitutes the lattice around each of the twelve icosahedral 5-fold axes (Fig. 2(C)). The cellular attachment protein $\sigma 1$ exists as filamentous trimers and extends from a $\lambda 2$ turret at each vertex of the virion (Chappell *et al.*, 2002). Due to the flexibility of $\sigma 1$, only limited densities attributable to the base of $\sigma 1$ trimer have been observed in the cryo-EM map of the virion (Zhang *et al.*, 2005). Among the *Reoviridae*, CPV, with a single capsid layer, exhibits the simplest $T = 1$ structural organization with pronounced turrets at the 5-fold vertices.

Cell Entry of the dsRNA Viruses

The outer capsid proteins initiate the virus infections by recognizing and attaching to specific cellular receptors on the host cell surface (Fig. 3). For instance, the outer capsid proteins, rotavirus VP4, BTV VP2, and orthoreovirus $\sigma 1$ recognize specific cellular receptors, thereby mediating the host specificity, tissue tropism, and pathogenesis of the viruses. These proteins are usually the least conserved structural proteins in the dsRNA viruses and might have evolved to provide distinct mechanisms to recognize specific receptors of different hosts and tissues. Recent structural studies have furthered our understanding of the molecular mechanisms by which the capsid proteins bind to host receptors.

In non-turreted dsRNA viruses, such as rotavirus, the spike protein VP4 is cleaved into VP8* and VP5* domains, which remain associated with the virions. Proteolytic cleavage, which stabilizes the spike into a ‘dimer-looking’ state with a significant portion of one of the subunits in the trimer disordered, is known to enhance the virus infectivity (Crawford *et al.*, 2001; Pesavento *et al.*, 2006). The “root” of VP5* along with a portion of the N-terminal region of VP8* is anchored into the middle capsid layer composed by VP6, while the major portion of VP8* is distally located at the tip of the VP4 spike (Fig. 4(A)). The structure of VP8* has a conserved characteristic

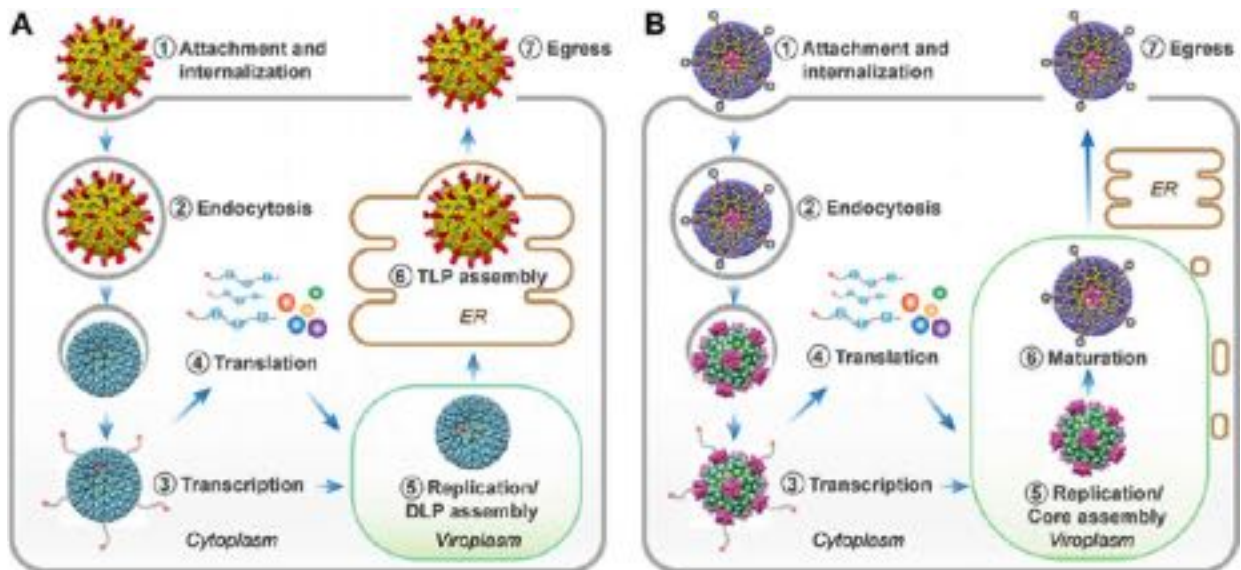


Fig. 3 The replication cycles of rotavirus (A) and reovirus (B). (A) Rotavirus attach to the host cell surface and get internalized into cells via endocytic pathways. The low Ca^{2+} level in the endosome triggers the loss of the outer capsid proteins and release of the transcriptionally active double-layered particles into the cytoplasm. The mRNA transcripts are used for viral protein translation and as template for negative strand RNA synthesis in the viroplasm. The dsRNA genome of rotavirus is assorted and packaged into DLPs, which bud into ER for TLP assembly. The progeny TLPs are released from host cells via cell lysis or by a Golgi-dependent non-classical vesicular transport mechanism. (B) Reovirus enters host cells through receptor-mediated endocytosis. The outer capsid layer is cleaved by host cell proteases in the endosomes, releasing the transcriptionally active cores into the cytoplasm. The newly transcribed mRNAs are used to viral protein synthesis and as templates for negative strand RNA synthesis in the viral inclusions (VIs). The dsRNA genome is packed into viral cores, followed by addition of outer-capsid proteins to form mature progeny viral particles. The cartoon of capped mRNAs and reovirus attachment protein $\sigma 1$ were drawn at the 5-fold axis of the viral particles. A schematic of viral protein translation processes is shown.

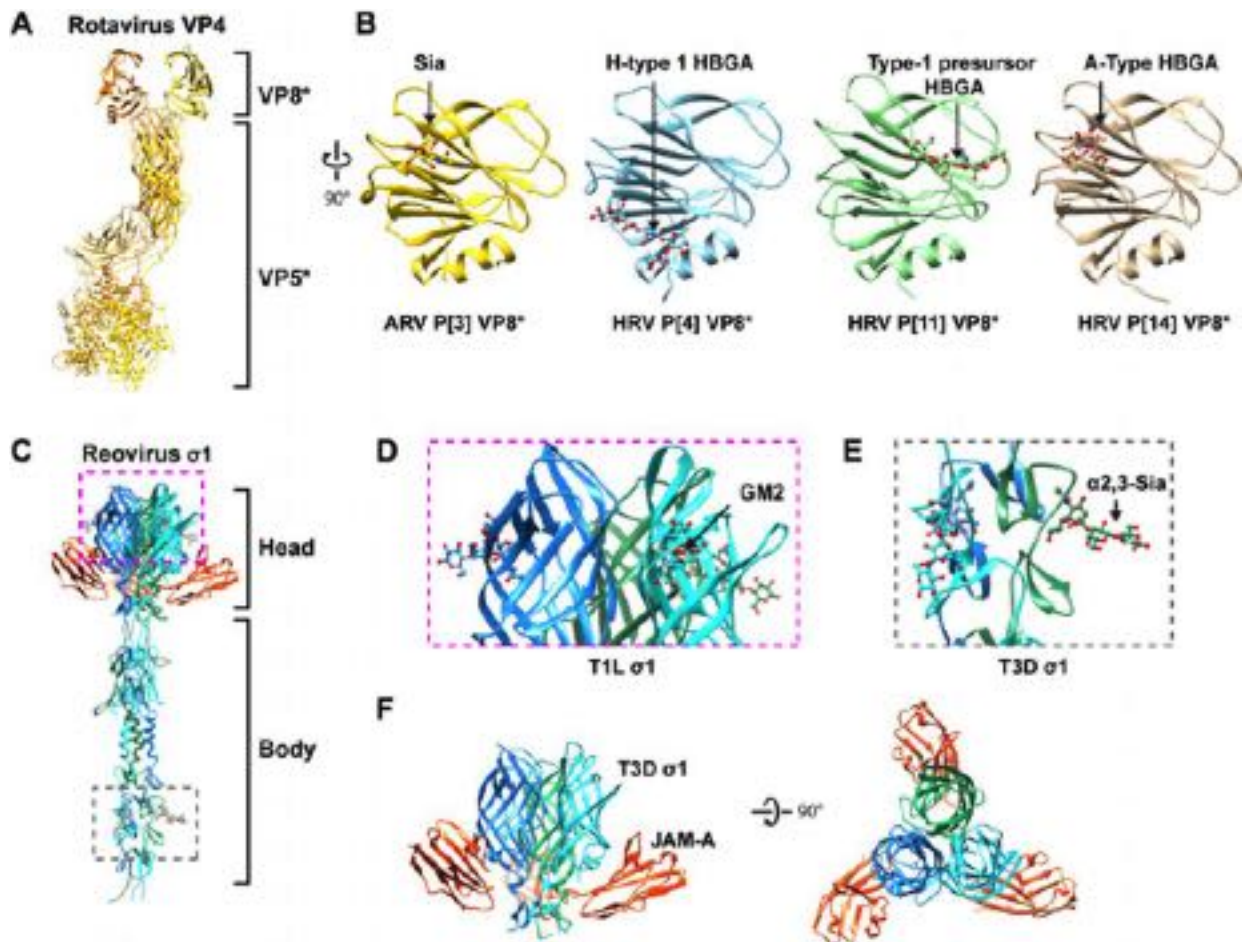
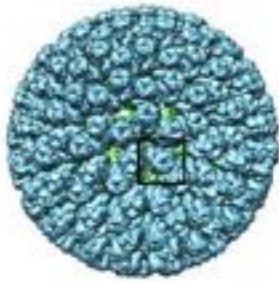


Fig. 4 Structure of cell attachment protein of dsRNA viruses in complex with its receptors. (A) Ribbon representation of rotavirus spike protein VP4 (PDB 4V7Q) that is composed of VP8* and VP5* domains. (B) Structures of VP8* proteins in complex with sialo-glycan or HBGA. The VP4 genotypes and the bound glycans are labeled: ARV P³ VP8*/Sia (PDB 1KQR); HRV P⁴ VP8*/H-type 1 pentasaccharide (PDB 5VX5); HRV P¹¹/Type-1 precursor (PDB 4YFZ); HRV P¹⁴/A-type HBGA trisaccharide (PDB 4DRV). The protein and the glycan are shown in ribbon representation and ball-and-stick models, respectively. (C) Structural overlay of reovirus attachment protein $\sigma 1$ in complex with receptors. The glycan receptor binding site on the head and body domains are noted by magenta and gray box, respectively. (D) The structure of reovirus T1L $\sigma 1$ in complex with the GM2 glycan (PDB 4GU3). (E) The structure of reovirus T3D $\sigma 1$ in complex with α -2,3-sialyllactose (PDB 3S6X). (F) The structure of reovirus T3D $\sigma 1$ in complex with JAM-A (PDB 3EOY). The $\sigma 1$ trimer (green, cyan, and blue) and JAM-A (red) are shown in ribbon representation.

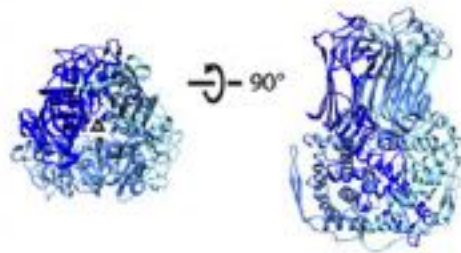
galactin-like fold with two twisted β -sheets separated by a shallow cleft. Rotaviruses VP8* recognizes cellular glycans, including sialic acid (Sia) (Dormitzer *et al.*, 2002) and histo-blood group antigens (HBGAs) (Hu *et al.*, 2018, 2015, 2012a). HBGA are genetically determined glycans present on epithelial cells, red blood cells, and in mucosal secretions. The VP8* domain of VP4 spike recognizes cellular glycans in a VP4 genotype (P genotype)-dependent manner to mediate the initial cellular attachment (Fig. 4(B)). Following cell attachment, during the cell entry process, VP4 spike, particularly VP5*, is suggested to undergo drastic membrane-coupled conformational changes to permeabilize host cells (Yoder *et al.*, 2009; Trask *et al.*, 2010; Pesavento *et al.*, 2001). The outer capsid VP7 is a Ca^{2+} -stabilized protein with two Ca^{2+} ion binding sites at each subunit interface of VP7 trimers (Aoki *et al.*, 2009). The loss of Ca^{2+} inside the host cell triggers the dissociation of VP4 and VP7 and the release of DLPs into the cytosol (Salgado *et al.*, 2018). In BTV, the VP2 trimer with identifiable sialic acid-binding pockets is suggested to play a role in cell attachment and receptor engagement, whereas VP5 with distinct α -helical regions is implicated in membrane fusion activity (Forzan *et al.*, 2007; Zhang *et al.*, 2010a). Distinct zinc-binding sites in VP2 and a cluster of histidine residues in VP5 are suggested to control pH-dependent conformational changes required for cell entry of BTV and the subsequent detachment of the outer layer (Zhang *et al.*, 2016).

In turreted reovirus, two outer layer proteins, the filamentous cell attachment protein $\sigma 1$ and the membrane-penetration protein $\mu 1$, are critical for viral entry. The trimers of $\sigma 1$ protrude out from the virus surface from the $\lambda 2$ turret at each vertex of the virion. $\sigma 1$ attaches to cell-surface glycans containing Sia and the immunoglobulin superfamily member junctional adhesion molecule-A (JAM-A). The $\sigma 1$ protein is composed of three domains, including a globular head domain comprised of eight antiparallel β -strands (residues 310–455), a body domain of β -spiral repeats (residues 170–309), and a tail domain (residues 1–160) that has a α coiled-coil (Fig. 4(C)) (Dietrich *et al.*, 2018). The type-1 Lang (T1L) and type-3 Dearing (T3D) orthoreovirus

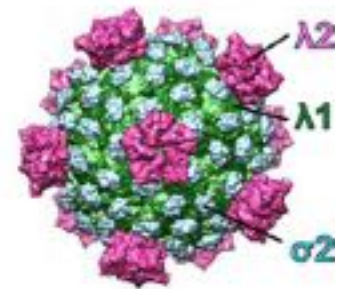
A Rotavirus DLP



VP6 trimer



B Reovirus core

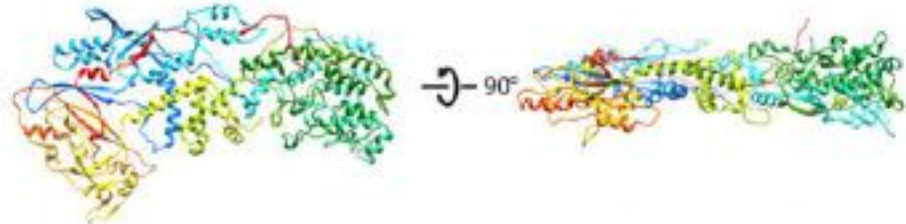
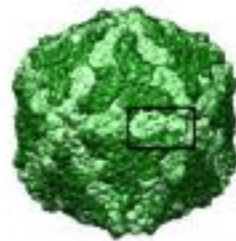


C

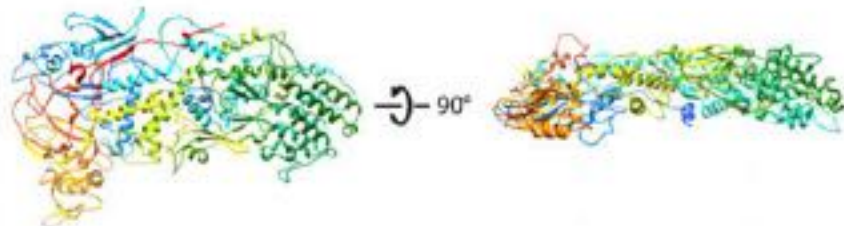
Rotavirus VP2



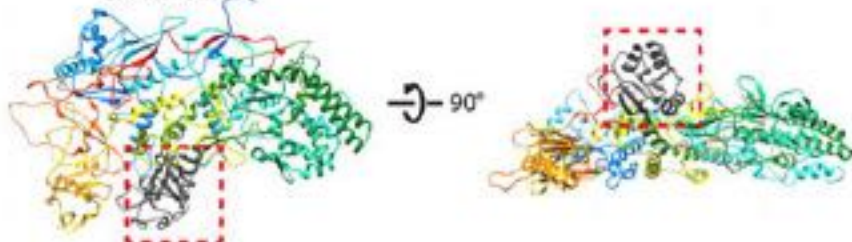
RDV P3



Reovirus λ 1



CPV VP1



bind to sialo-glycans using the head domain and the body domain, respectively (Fig. 4(D) and (E)) (Reiss *et al.*, 2012; Reiter *et al.*, 2011). After attachment to the cellular glycans, all human reoviruses bind to the tight junction component JAM-A (Fig. 4(F)) (Kirchner *et al.*, 2008). After initial attachment, the viruses are internalized through binding of $\beta 1$ integrins to reovirus protein $\lambda 2$, which contains integrin-binding motifs, RGD and KGE. The chaperone protein $\sigma 3$ is then removed in cathepsin proteases in the endosome, resulting in the formation of the infectious subviral particle (ISVP). The membrane-penetration protein $\mu 1$ is also proteolytically processed and undergoes conformational changes to rupture the endosomal membrane and release the transcriptionally active virus core into the cytosol (Danthi *et al.*, 2010, 2013).

The dsRNA Virus Core and Endogenous Transcription

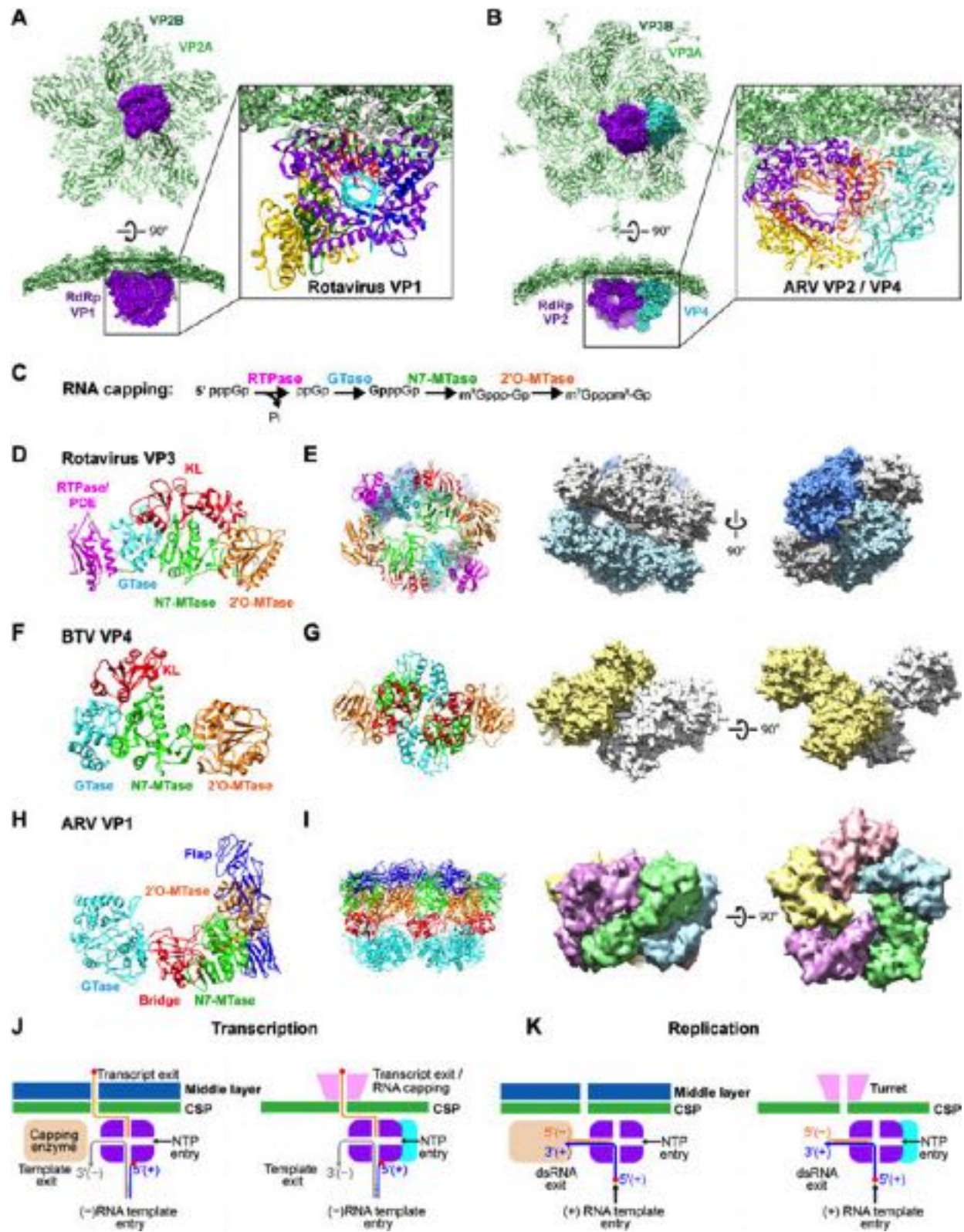
A distinguishing feature of the dsRNA viruses in general and of *Reoviridae* in particular is the process of endogenous transcription (Fig. 3). In this process, the genomic dsRNA segments are transcribed within the particle interior and the transcripts are extruded. This process is typically activated by the removal of the outer capsid layer during the entry process (Fig. 3). For instance, the outer layer of rotavirus is removed, resulting in a double-layered particle (DLP) (Fig. 3(A)). The outer layer of DLP is a $T = 13$ icosahedral lattice of 260 VP6 trimers, which are anchored on the $T = 1$ innermost core layer of VP2 (Fig. 5(A)). When incubated with NTPs, Mg^{2+} , and S-adenosylmethionine (SAM), the rotavirus DLPs undergo rounds of transcription, with the mRNA transcripts emerging from the pores/channels at the 5-fold vertices of DLPs (Lawton *et al.*, 1997). The process of transcription is very similar in the case of BTV with the removal of the outer layer consisting of VP2 and VP5, the core particle (the equivalent of rotavirus DLP) with a $T = 13$ VP7 and $T = 1$ VP3 capsid layers is the transcriptionally active form. Similarly, in orthoreovirus, endogenous transcription is activated upon the removal of the outer capsid layer, and the nascent mRNA transcripts exit through the turrets at the 5-fold axes of the resulting infectious subviral particles (Figs. 3(B) and 5(B)).

Viral Innermost Capsid Layer That Encloses the Genome

The innermost capsid layer has a $T = 1$ (also referred to as " $T = 2$ ") icosahedral symmetry with 60 asymmetric dimers and encloses the dsRNA genome (McClain *et al.*, 2010). This structural organization is highly conserved in all dsRNA viruses, including those of bacterial and fungal origins (Luque *et al.*, 2018). X-ray crystallographic and cryo-EM structures of the dsRNA virus cores showed that the innermost capsid proteins share a conserved α -helical domain even despite little sequence identity (Fig. 5(C)). The VP1 of CPV exhibits a slight variation with the addition of a small domain that protrudes from the outer capsid surface. It has been hypothesized that such a conserved organization of the innermost layer has evolved to facilitate and regulate the endogenous transcription by appropriately packing the segmented dsRNA genome and anchoring the RdRp and/or capping enzyme in the interior (Prasad and Privilege, 2003).

Cryo-EM studies of several members of *Reoviridae*, including rotavirus, bluetongue virus, and orthoreovirus, have provided molecular insight into how the unique organization of the innermost layer supports the fascinating process of endogenous transcription. In rotavirus, two VP2 conformers, VP2-A and VP2-B, assembled around the fivefold axis to form a decamer, and 12 of VP2 decamers pack together to form the innermost layer (Fig. 5(C)). Rotavirus VP2 contains the N-terminal tether domain (residues 1–125) and the C-terminal scaffold domain (McClain *et al.*, 2010). The N-terminal domain is flexible. There was no density structurally resolved for the first 99 residues of the VP2-A and the first 80 residues of the VP2-B in the 3.8 Å crystal structure of the inner capsid particle. N-terminal deletion of the VP2 mutant failed to encapsidate VP1 and VP3 into the virus-like particles in vitro (Zeng *et al.*, 1998; Estrozi *et al.*, 2013). Thus, it has been thought that the N-terminal domain tethers the VP1/VP3 complex at the fivefold vertices. The recent in situ cryo-EM structure of rotavirus VP1 revealed that it binds to the inner surface of the VP2 shell with complementary interactions and with three N-terminal arms of VP2, which form a tentacle-like structure (Fig. 6(A)) (Jenni *et al.*, 2019). A similar structural organization of the inner shells has also been observed in other members of the *Reoviridae* family. For instance, in BTV, the RdRp VP1 is anchored inside the core capsid shell via five asymmetrically arranged N termini of the capsid shell protein VP3A around the fivefold axis (He *et al.*, 2019). In ARV, the transcriptional enzyme complex of RdRp VP2 and its cofactor protein VP4 is anchored via four of the five N-termini of the core capsid protein at the fivefold vertices (Wang *et al.*, 2018) (Fig. 6(B)).

Fig. 5 The dsRNA virus cores. (A) Rotavirus double-layered particle (DLP) (PDB 3KZ4). The VP6 trimer is shown in ribbon representation. (B) Reovirus core (PDB 1EJ6). (C) Structures of $T = 1$ core capsid proteins: Rotavirus VP2; Rice dwarf virus (RDV) P3; Orthoreovirus $\lambda 1$; Cytoplasmic polyhedrosis virus (CPV) VP1. The CPV VP1 has an extra protrusion domain (residues 825–962) marked with a red box with the dotted line. Adapted from: (A) McClain, B., Settembre, E., Temple, B.R., Bellamy, A.R., Harrison, S.C., 2010. X-ray crystal structure of the rotavirus inner capsid particle at 3.8 Å resolution. *Journal of Molecular Biology* 397, 587–599. (B) Reinisch, K.M., Nibert, M.L., Harrison, S.C., 2000. Structure of the reovirus core at 3.6 Å resolution. *Nature* 404, 960–967. (C) McClain, B., Settembre, E., Temple, B.R., Bellamy, A.R., Harrison, S.C., 2010. X-ray crystal structure of the rotavirus inner capsid particle at 3.8 Å resolution. *Journal of Molecular Biology* 397, 587–599. Nakagawa, A., Miyazaki, N., Taka, J., *et al.*, 2003. The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure* 11, 1227–1238. Reinisch, K.M., Nibert, M.L., Harrison, S.C., 2000. Structure of the reovirus core at 3.6 Å resolution. *Nature* 404, 960–967. Yu, X., Jiang, J., Sun, J., Zhou, Z.H., 2015. A putative ATPase mediates RNA transcription and capping in a dsRNA virus. *Elife*. 4, e07901. doi:10.7554/eLife.07901.



Viral RNA-Dependent RNA Polymerase (RdRp)

Due to the lack of cellular enzymes that can synthesize mRNA transcripts using the (–)RNA strand of the dsRNA genome as the template, the dsRNA viruses encode their own viral RdRps and capping enzymes that carry out the endogenous transcription of the dsRNA segments. During the process of endogenous transcription, the virus core remains structurally intact, such that the dsRNA genome is protected from triggering the activation of cellular antiviral interferon responses by dsRNA. In several viruses of the *Reoviridae* family, the RdRp is directly anchored to capsid protein at the inner surface at the fivefold vertices. Crystal structures of the *Reoviridae* RdRps, such as reovirus λ 3 and rotavirus VP1 in isolation have been determined (Tao *et al.*, 2002; Lu *et al.*, 2008). The RdRps are largely structurally conserved with a cage-like organization that includes an N-terminal domain, a right-hand polymerase domain, and a C-terminal “bracelet” domain (Fig. 6(A) and (B)). There are four tunnels in the RdRp for template entry, NTP entry, dsRNA/(–)RNA exit, and mRNA exit. The mRNA cap-binding site is located on the surface of RdRps and has been thought to anchor the 5′-end of (+)RNA, which will guide 3′-end of (–)RNA to re-enter the RdRp for another round of transcription (Fig. 6(J) and (K)).

Recent advances in cryo-EM and single-particle analysis, especially in using focused classification and asymmetric reconstructions, have provided in situ structures of RdRp inside the rotavirus, BTV, ARV, and CPV (Wang *et al.*, 2018; Ding *et al.*, 2018; He *et al.*, 2019; Cui *et al.*, 2019; Ding *et al.*, 2019; Jenni *et al.*, 2019). These structures reveal the conformational changes of RdRp upon transcription and the trajectories of mRNA transcripts. In rotavirus, Jenni *et al.* showed that dissociation of the outer layer of rotavirus leads to RNA synthesis without introducing structural changes in VP1. However, the addition of NTPs, Mg^{2+} , and S-adenosylmethionine to trigger transcription results in significant conformational changes in VP1 and capsid shell protein in DLPs (Jenni *et al.*, 2019) (Fig. 6(A)). These studies have also shown that VP1 attaches to the VP2 shell stochastically at one of the five possible positions at each of the fivefold vertices. In the turreted ARV, the RdRp protein VP2 forms a complex with its cofactor VP4 and genomic RNA near the 5-fold axes (Fig. 6(B)). The in situ cryo-EM structures of CPV RdRp with its associated RNA have revealed the transcription steps by capturing the quiescent, initiation, early elongation, elongation, and abortive states under different conditions (Cui *et al.*, 2019).

Viral RNA Capping

The 5′-cap of mRNA is critical for RNA stability, gene translation, and immune evasion. This cap not only prevents RNA degradation by cellular exonucleases but also promotes RNA translation by specific recognition of the eukaryotic translation initiation factor 4E (eIF4E). Moreover, it is crucial for evading host innate immune response mediated by RIG-I, which recognizes the 5′-triphosphate of uncapped RNA (Hornung *et al.*, 2006). The formation of the 5′-cap is catalyzed by viral capping machineries, which include RNA triphosphatase (RTPase), guanylyltransferase (GTase), 7-N-methyltransferase (N7-MTase), and 2′-O-methyltransferase (2′O-MTase) activities (Fig. 6(C)). One major structural difference between turreted and non-turreted dsRNA viruses of the *Reoviridae* family is the location of their capping enzymes. In non-turreted dsRNA viruses, such as rotavirus and BTV, the capping enzyme likely associates with the RdRp inside the virus core (Fig. 6(J)). In contrast, in turreted dsRNA viruses, such as orthoreovirus, ARV, and CPV, the capping enzyme is external to the core layer and forms the turret-like feature (Fig. 6(J)).

The in situ cryo-EM structure of rotavirus showed that the mRNA exit channel faces the pore at the fivefold axes. How does the RNA transcripts get capped by the capping enzyme? One hypothesis is that the base-paired (–)RNA template/(+)RNA product emerges through the template exit channel and encounters VP3 for RNA capping. The capped transcript then re-threads and enters the transcript exit channel (Jenni *et al.*, 2019). It might be because rotavirus VP3 does not anchor to a fixed location on VP2 capsid or VP1 RdRp. However, the in situ cryo-EM structure of rotavirus did not resolve any density for VP3 to allow determination of the structure of rotavirus VP3 and how it coordinates with VP1 to cap the nascent transcript exiting from VP1. Recently, cryo-EM and X-ray crystallography analyses of recombinant full-length VP3 revealed the atomic details of how the catalytic domains for mRNA capping are assembled in line (Fig. 6(D)) (Kumar *et al.*, 2020). Intriguingly, the rotavirus VP3 forms a stable tetramer both in solution and in crystals by the association of antiparallel dimers (Fig. 6(E)), which may have functional roles in mRNA capping and anti-interferon response mediated by the C-terminal PDE domain of VP3 (Ogden *et al.*, 2015; Kumar *et al.*, 2020). In BTV, the crystal structure of RNA-capping enzyme VP4 has been determined, and it assembles all four enzymatic activities with a directional layout (Fig. 6(F)) (Sutton *et al.*, 2007). Similar to the rotavirus VP3, BTV VP4 forms dimers in solution and also in crystals along a

Fig. 6 The transcription and replication mechanism of non-turreted and turreted dsRNA viruses. (A) In situ structure of rotavirus RdRp VP1 that is colored in purple (PDB 6OJ6) (Jenni *et al.*, 2019). The core shell protein VP2A and VP2B are shown in light green and forest green ribbon diagrams, respectively. The close-up view of the inset shows the ribbon representation of VP1: the N-terminal domain (residues 1–332) is in gold; the palm subdomain (residues 489–523, 596–685) is in red; the fingers (residues 333–488, 524–595) is in blue; the thumb (residues 686–778) is in forest green; the bracelet (residues 779–1089) is colored purple. The dsRNA is colored cyan. (B) In situ structure of grass carp reovirus (Aquareovirus) RdRp VP2 (purple) and its cofactor protein VP4 (cyan) (PDB 5ZVS). The close-up view of the inset shows the ribbon representation of VP2/VP4 complex: the N-terminal domain (residues 1–386) in gold, a central polymerase domain (residues 387–897) in red, and a C-terminal “bracelet” domain (residues 898–1273) in purple. (C) Schematic mRNA capping mechanism by dsRNA viruses. (D) Structure of rotavirus VP3 monomer (PDB 6O6B). Each domain has been colored and labeled. (E) Tetramer assembly of rotavirus VP3 shown in along the 2-fold axes of the D2 symmetry. The VP3 domains in the ribbon representation is colored as in (D). Four VP3 chains in surface representation is colored in white, gray, light blue, and cornflower blue. (F) Structure of BTV VP4 monomer (PDB 2JH8). (G) Dimer of BTV VP4 formed along a crystallographic two-fold axis and viewed in ribbon and surface representations as in (E). (H) Structure of ARV turret protein VP1 (PDB 5ZVT). (I) Pentamer of ARV turret protein VP1 viewed perpendicular to or along the 5-fold axis in ribbon and surface representations as in (E). (J)–(K) Mechanisms of RNA transcription and replication of non-turreted (left panel) and turreted (right panel) dsRNA viruses.

crystallographic two-fold axis (**Fig. 6(G)**), suggesting that the oligomeric formation of the capping enzyme might be a common structural mechanism in the non-turreted dsRNA viruses.

Viruses of the turreted *Spinareovirinae* subfamily have the capping enzyme located at the 5-fold vertices. Reovirus $\lambda 2$ forms a pentameric turret with the active sites facing the interior tunnel of the turret. The nascent mRNA exits through these tunnels while being capped by $\lambda 2$ (Reinisch *et al.*, 2000) (Figs. 3(B) and 5(B)). Similarly, in ARV and CPV, the pentameric turret contains the capping domains, which are connected by unique channels to allow the sequential GTase, 7 N7-MTase, and 2'O-MTase enzymatic reactions (**Fig. 6(H)** and **(I)**) (Cheng *et al.*, 2011; Wang *et al.*, 2018; Ding *et al.*, 2018; Cui *et al.*, 2019). However, the turrets themselves do not contain the RTPase activity, which cleaves the 5'- γ -phosphate of nascent (+)RNA to allow the addition of GMP and formation of the 3'-GpppG 5'-cap structure. Instead, the associated NTPase likely catalyzes the hydrolysis of the 5'- γ -phosphate.

Genome Replication and Packaging

Viroplasm or Replication Factories

In viruses of the *Reoviridae* family, mRNAs are released from viral particles and translated to synthesize viral proteins that form condensed inclusions termed viroplasms, viral factories (VFs), or viral inclusions (VIs) in the cytosol (Tenorio *et al.*, 2019; Borodavka *et al.*, 2018; Roy, 2017) (**Fig. 3**). These inclusions contain viral genomic RNA and structural proteins that package into progeny viruses, as well as NSPs that regulate viral genome replication and packaging (Hu *et al.*, 2012b). In rotavirus, NSP2 and

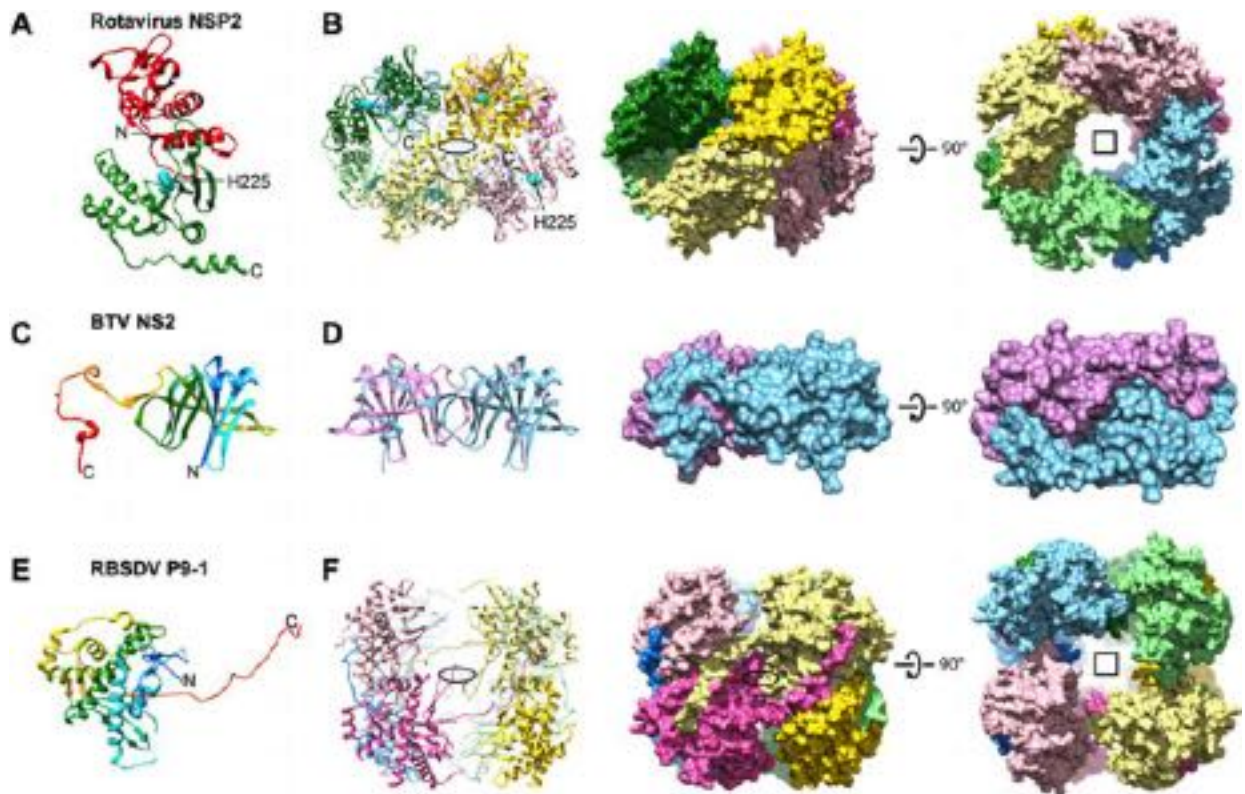


Fig. 7 Structure of nonstructural RNA binding proteins that form the viroplasms. (A) Structure of rotavirus NSP2 (PDB 1L9V). The N-terminal and C-terminal domains are colored in red and green, respectively. The catalytic residue His225 between two domains is shown as cyan sphere. (B) The octamer of rotavirus NSP2 viewed along a 2-fold axis in ribbon and surface representation with each chain colored differently. Surface representation of NSP2 octamer along a 4-fold axis is shown on the right of the panel. (C) Structure of the RNA-binding domain (RBD) of BTV NS2 protein shown as ribbon and colored with rainbow color from N-terminal to C-terminal end (blue to red). (PDB 1UTY). (D) Dimer of NS2 RBD shown in ribbon and surface representations. Surface representation of NS2 RBD dimer is also shown along a 2-fold axis. (E) Structure of P9-1 protein of rice black streaked dwarf virus presented in rainbow colored ribbon (PDB 3VJJ). (F) The P9-1 octamer viewed along a 2-fold axis in ribbon and surface representations. Each chain is colored differently. Surface representation of P9-1 octamer is also shown along a 4-fold axis. Adapted from: (A) Jayaram, H., Taraporewala, Z., Patton, J.T., Prasad, B.V., 2002. Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. *Nature* 417, 311–315. (C) Butan, C., Van Der Zandt, H., Tucker, P.A., 2004. Structure and assembly of the RNA binding domain of bluetongue virus non-structural protein 2. *Journal of Biological Chemistry* 279, 37613–37621. (E) Akita, F., Higashiura, A., Shimizu, T., *et al.*, 2012. Crystallographic analysis reveals octamerization of viroplasm matrix protein P9-1 of Rice black streaked dwarf virus. *Journal of Virology* 86, 746–756.

NSP5 are essential for viroplasm formation, and genome replication and packaging. NSP2 is a multifunctional protein that exhibits nucleoside triphosphatase (NTPase), nucleoside diphosphate kinase (NDP kinase), RNA triphosphatase (RTPase), single-stranded RNA binding, and helix destabilizing activities (Hu *et al.*, 2012b). The crystal structure of rotavirus NSP2 revealed that the catalytic center of NSP2 resembles the histidine triad (HIT) of cellular nucleotidyl hydrolases (Jayaram *et al.*, 2002; Taraporewala *et al.*, 2002; Taraporewala and Patton, 2001; Taraporewala *et al.*, 1999) (Fig. 7(A)). The full-length recombinant NSP2 protein exists as octamers in solution (Fig. 7(B)). Cryo-EM studies of NSP2 alone or in complex with either NSP5 or RNA showed that NSP2 exists as octamers in solution and binds to NSP5 and RNA at the same site (Jiang *et al.*, 2006). In addition, NSP2 also interacts with RdRp VP1 and the N-terminal hub of the core capsid protein VP2 (Viskovska *et al.*, 2014). These data support a model that NSP2 may function as a molecular motor using energy from NTP hydrolysis to destabilize the mRNA template and feed it to the RNA template to VP1 during genome replication. The existing biochemical data suggest that rotavirus NSP2 is functionally homologous to NS2 of BTV (Fillmore *et al.*, 2002; Taraporewala *et al.*, 2001; Zhao *et al.*, 1994), and σ NS of orthoreovirus (Gillian and Nibert, 1998). The NS2 of BTV is critical for viral inclusion body formation and recruits viral RNA and viral proteins for virus core assembly. The crystal structure of the RNA-binding domain of BTV NS2 (residues 1–177) has a β -sandwich structure and assembles into dimers in crystal packing (Butan *et al.*, 2004) (Fig. 7(C) and D). Small-angle X-ray scattering analysis of the C-terminal domain of NS2 showed that it forms an elongated dimer. Electron microscopy of the full-length protein showed some ring-like structures, providing a basis for the proposed model for a decamer of the full-length protein (Mumtsidu *et al.*, 2007). Interestingly, in the turreted Rice black streaked dwarf virus (RBSDV), a member of the genus *Fijivirus* in the family *Reoviridae*, the P9–1 protein also accumulates in viroplasm inclusions. Crystallographic analysis of RBSDV P9–1 protein revealed that it assembled into octamers, remarkably resembling the octamer formation of rotavirus NSP2 (Fig. 7(E) and (F)) (Akita *et al.*, 2012). In cells infected by turreted reovirus, viral non-structural proteins μ NS and σ NS and structural protein μ 2 regulate the formation of VIs (Broering *et al.*, 2002; Becker *et al.*, 2003). However, the structural basis of how the replication factories are formed remains unknown due to the lack of structures of reovirus μ NS or σ NS.

Replication and Packaging

The viral mRNA transcripts are used not only for the translation of viral proteins but also for the synthesis of (–)RNA to produce dsRNA genome within viroplasm (Figs. 3 and 6(K)). It remains unclear how the dsRNA viruses precisely assort, assemble, and package a complete set of gene segments into progeny viruses. Although the detailed structural information is still unknown, it has been proposed that specific recognitions via RNA–RNA interactions result in the clustering of a set of the dsRNA segments during genome packaging (Pesavento *et al.*, 2006). In rotavirus, the RdRp VP1 specifically recognizes the 3′-consensus sequence (3′-CS) in the template-entry channel (Lu *et al.*, 2008). In addition, VP1 extensively interacts with VP2 of the innermost layer and stabilizes the VP2 decamers (Jenni *et al.*, 2019). This interaction with VP2 initiates the synthesis of (–)RNA by RdRp VP1, thereby coordinating RNA replication with the assembly of the virus core. *In vitro* biochemical assays using recombinant proteins showed VP1 is only active when VP2 is present (Patton *et al.*, 1997). Using cell-free reconstitution and reverse genetics systems, it has been suggested that BTV (+)RNAs assort and assemble into large complexes before being incorporated into the core particles (Periz *et al.*, 2013). The 5′- and 3′-UTRs mediate the trans-acting RNA–RNA interactions (tRRIs) between gene segments and facilitate the packing of the entire genome.

Other dsRNA Virus Families

Most dsRNA viruses that have been structurally characterized are icosahedral non-enveloped particles, except the *Cystoviridae* family, which has enveloped viruses with a genome of three dsRNA segments and a double-layered protein capsid. For instance, the bacteriophage ϕ 6 has three structurally defined layers. Its outermost layer contains a lipid bilayer envelope formed by viral integral membrane proteins (P6, P9, P10, and P13) and cellular phospholipids. The cellular attachment spike protein P3 is anchored onto the envelope via a fusogenic protein P6. The nucleocapsid is formed by two concentric icosahedrally organized protein layers. The outer layer of the nucleocapsid is formed by P8 trimers with $T = 13$ icosahedral symmetry (Poranen *et al.*, 2017) generally similar to the organization of the middle layers in the members of *Reoviridae* such as rotavirus and BTV. The core capsid layer is formed by the capsid protein P1, the RdRp P2, the NTPase P4 with a hexameric clustering, and the assembly cofactor P7. The asymmetric P1A–P1B dimers arranged with $T = 1$ symmetry, enclosing the genome of three dsRNA segments which adopt a dsDNA-like single-spoiled genome organization (Ilca *et al.*, 2019). Another dsRNA virus that is structurally well characterized is the λ -A virus, which belongs to the *Totiviridae* family. Interestingly, it has a single capsid layer that encloses a single dsRNA segment genome. Remarkably, this capsid layer with 120 capsid protein subunits follows the same organization of the innermost capsid layers observed in the *Reoviridae*, suggesting an evolutionary hierarchy in dsRNA viruses (Naitow *et al.*, 2002; Bamford *et al.*, 2005).

Conclusions

Members of the *Reoviridae*, the largest and most diverse dsRNA virus family, exhibit enormous structural diversity in their outer capsid proteins, which might be the molecular determinants of the host specificity, tissue tropism, and pathogenesis of the viruses.

Nonetheless, they also display remarkable structural conservation particularly in the organization of the inner capsid layers to support the common requirement of endogenous transcription. Combined with the structural information gained through X-ray crystallography and cryo-EM, the structural basis of cellular attachment mediated by the outer capsid proteins, the capping mechanisms of mRNA, endogenous transcription, and viroplasm formation have begun to emerge in recent years. However, many fascinating questions still remain that perhaps can be approached by further advances in the structural techniques particularly in cryo-EM techniques, which probably are more suitable to address such questions. Through these studies, we can anticipate significantly more detailed insight into receptor(s) recognition, mechanisms of virus entry into cells, how the outer layer is removed to deposit the transcriptionally active form of the particle in the cytoplasm, formation of the replication factories, mechanism of immune evasion, genome reassortment and packaging, and exiting from infected cells. Further structural studies will be needed to understand how the capping enzymes, such as rotavirus VP3 and BTV VP4, are located in the non-turreted dsRNA viruses and how they coordinate with the viral RdRps during transcription and virion assembly. Also, the atomic models of many viral non-structural proteins, for example, the rotavirus NSP1, NSP4, NSP5, and NSP6, the reovirus μ NS and σ NS, are still unknown. These NSPs are critical players during virus replication and may serve as potential candidates for antiviral development. Another aspect of the replication of the dsRNA viruses not discussed here is how these viruses are released from infected host cells. Although biochemical studies suggest that there are multiple mechanisms by which the dsRNA virus egress from host cells, little structural information about this process is known. For instance, the progeny of rotavirus is released from non-polarized kidney epithelium cells via direct cell lysis but traffics to the apical surface and is secreted from polarized intestinal epithelial host cells. Future in situ structural studies of the infected cells using electron cryotomography (cryoET) may yield more molecular and mechanistic details about these processes.

Acknowledgments

We acknowledge support from the National Institutes of Health (NIH) grants AI36040 (to B.V.V.P.) and AI080656 (to M.K.E), and the Robert Welch Foundation (Q1279) to B.V.V.P.

References

- Akita, F., Higashiura, A., Shimizu, T., *et al.*, 2012. Crystallographic analysis reveals octamerization of viroplasm matrix protein P9-1 of Rice black streaked dwarf virus. *Journal of Virology* 86, 746–756.
- Aoki, S.T., Settembre, E.C., Trask, S.D., *et al.*, 2009. Structure of rotavirus outer-layer protein VP7 bound with a neutralizing Fab. *Science* 324, 1444–1447.
- Bamford, D.H., Grimes, J.M., Stuart, D.I., 2005. What does structure tell us about virus evolution? *Current Opinion in Structural Biology* 15, 655–663.
- Becker, M.M., Peters, T.R., Dermody, T.S., 2003. Reovirus sigma NS and mu NS proteins form cytoplasmic inclusion structures in the absence of viral infection. *Journal of Virology* 77, 5948–5963.
- Borodavka, A., Desselberger, U., Patton, J.T., 2018. Genome packaging in multi-segmented dsRNA viruses: distinct mechanisms with similar outcomes. *Current Opinion in Virology* 33, 106–112.
- Broering, T.J., Parker, J.S., Joyce, P.L., Kim, J., Nibert, M.L., 2002. Mammalian reovirus nonstructural protein microNS forms large inclusions and colocalizes with reovirus microtubule-associated protein micro2 in transfected cells. *Journal of Virology* 76, 8285–8297.
- Butan, C., Van Der Zandt, H., Tucker, P.A., 2004. Structure and assembly of the RNA binding domain of bluetongue virus non-structural protein 2. *Journal of Biological Chemistry* 279, 37613–37621.
- Chappell, J.D., Prota, A.E., Dermody, T.S., Stehle, T., 2002. Crystal structure of reovirus attachment protein sigma1 reveals evolutionary relationship to adenovirus fiber. *EMBO Journal* 21, 1–11.
- Cheng, L., Sun, J., Zhang, K., *et al.*, 2011. Atomic model of a cypovirus built from cryo-EM structure provides insight into the mechanism of mRNA capping. *Proceedings of the National Academy of Sciences of the United States of America* 108, 1373–1378.
- Cheng, L., Zhu, J., Hui, W.H., *et al.*, 2010. Backbone model of an aquareovirus virion by cryo-electron microscopy and bioinformatics. *Journal of Molecular Biology* 397, 852–863.
- Crawford, S.E., Mukherjee, S.K., Estes, M.K., *et al.*, 2001. Trypsin cleavage stabilizes the rotavirus VP4 spike. *Journal of Virology* 75, 6052–6061.
- Cui, Y., Zhang, Y., Zhou, K., Sun, J., Zhou, Z.H., 2019. Conservative transcription in three steps visualized in a double-stranded RNA virus. *Nature Structural & Molecular Biology* 26, 1023–1034.
- Danthi, P., Guglielmi, K.M., Kirchner, E., *et al.*, 2010. From touchdown to transcription: The reovirus cell entry pathway. *Current Topics in Microbiology and Immunology* 343, 91–119.
- Danthi, P., Holm, G.H., Stehle, T., Dermody, T.S., 2013. Reovirus receptors, cell entry, and proapoptotic signaling. *Advances in Experimental Medicine and Biology* 790, 42–71.
- Dermody, T.S., Parker, J.S., Sherry, B., 2013. Orthoreoviruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins.
- Dietrich, M.H., Ogden, K.M., Long, J.M., *et al.*, 2018. Structural and functional features of the reovirus sigma1 tail. *Journal of Virology* 92.
- Ding, K., Celma, C.C., Zhang, X., *et al.*, 2019. In situ structures of rotavirus polymerase in action and mechanism of mRNA transcription and release. *Nature Communications* 10, 2216.
- Ding, K., Nguyen, L., Zhou, Z.H., 2018. In Situ Structures of the Polymerase Complex and RNA Genome Show How Aquareovirus Transcription Machineries Respond to Uncoating. *Journal of Virology* 92.
- Dornitzer, P.R., Sun, Z.Y., Wagner, G., Harrison, S.C., 2002. The rhesus rotavirus VP4 sialic acid binding domain has a galectin fold with a novel carbohydrate binding site. *EMBO Journal* 21, 885–897.
- Estes, M.K., Greenberg, H.B., 2013. Rotaviruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins.
- Estrozi, L.F., Settembre, E.C., Goret, G., *et al.*, 2013. Location of the dsRNA-dependent polymerase, VP1, in rotavirus particles. *Journal of Molecular Biology* 425, 124–132.
- Fillmore, G.C., Lin, H., Li, J.K., 2002. Localization of the single-stranded RNA-binding domains of bluetongue virus nonstructural protein NS2. *Journal of Virology* 76 (2), 499–506. Epub 2001/12/26. doi:10.1128/jvi.76.2.499-506.2002. PubMed PMID: 11752140; PMCID: PMC136823.

- Forzan, M., Marsh, M., Roy, P., 2007. Bluetongue virus entry into cells. *Journal of Virology* 81, 4819–4827.
- Gillian, A.L., Nibert, M.L., 1998. Amino terminus of reovirus nonstructural protein sigma NS is important for ssRNA binding and nucleoprotein complex formation. *Virology* 240 (1), 1–11. Epub 1998/02/04. doi:10.1006/viro.1997.8905. PubMed PMID: 9448684.
- Gouet, P., Diprose, J.M., Grimes, J.M., *et al.*, 1999. The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. *Cell* 97, 481–490.
- Grimes, J.M., Burroughs, J.N., Gouet, P., *et al.*, 1998. The atomic structure of the bluetongue virus core. *Nature* 395, 470–478.
- He, Y., Shivakoti, S., Ding, K., *et al.*, 2019. In situ structures of RNA-dependent RNA polymerase inside bluetongue virus before and after uncoating. *Proceedings of the National Academy of Sciences of the United States of America* 116, 16535–16540.
- Hill, C.L., Booth, T.F., Prasad, B.V., *et al.*, 1999. The structure of a cypovirus and the functional organization of dsRNA viruses. *Nature Structural & Molecular Biology* 6, 565–568.
- Hornung, V., Ellegast, J., Kim, S., *et al.*, 2006. 5'-Triphosphate RNA is the ligand for RIG-I. *Science* 314, 994–997.
- Hu, L., Crawford, S.E., Czako, R., *et al.*, 2012a. Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen. *Nature* 485, 256–259.
- Hu, L., Crawford, S.E., Hyser, J.M., Estes, M.K., Prasad, B.V., 2012b. Rotavirus non-structural proteins: Structure and function. *Current Opinion in Virology* 2, 380–388.
- Hu, L., Ramani, S., Czako, R., *et al.*, 2015. Structural basis of glycan specificity in neonate-specific bovine-human reassortant rotavirus. *Nature Communications* 6, 8346.
- Hu, L., Sankaran, B., Laucirica, D.R., *et al.*, 2018. Glycan recognition in globally dominant human rotaviruses. *Nature Communications* 9, 2631.
- Ilca, S.L., Sun, X., El Omari, K., *et al.*, 2019. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. *Nature* 570, 252–256.
- Jayaram, H., Taraporewala, Z., Patton, J.T., Prasad, B.V., 2002. Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold. *Nature* 417, 311–315.
- Jenni, S., Salgado, E.N., Herrmann, T., *et al.*, 2019. In situ structure of rotavirus VP1 RNA-dependent RNA polymerase. *Journal of Molecular Biology* 431, 3124–3138.
- Jiang, X., Jayaram, H., Kumar, M., *et al.*, 2006. Cryoelectron microscopy structures of rotavirus NSP2-NSP5 and NSP2-RNA complexes: Implications for genome replication. *Journal of Virology* 80, 10829–10835.
- Kirchner, E., Guglielmi, K.M., Strauss, H.M., Dermody, T.S., Stehle, T., 2008. Structure of reovirus sigma1 in complex with its receptor junctional adhesion molecule-A. *PLoS Pathogens* 4, e1000235.
- Kumar, D., Yu, X., Crawford, S.E., *et al.*, 2020. 2.7 Å cryo-EM structure of rotavirus core protein VP3, a unique capping machine with a helicase activity. *Science Advances* 6, eaay6410.
- Lawton, J.A., Estes, M.K., Prasad, B.V., 1997. Three-dimensional visualization of mRNA release from actively transcribing rotavirus particles. *Nature Structural & Molecular Biology* 4, 118–121.
- Liemann, S., Chandran, K., Baker, T.S., Nibert, M.L., Harrison, S.C., 2002. Structure of the reovirus membrane-penetration protein, Mu1, in a complex with its protector protein, Sigma3. *Cell* 108, 283–295.
- Liu, H., Cheng, L., 2015. Cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. *Science* 349, 1347–1350.
- Li, Z., Baker, M.L., Jiang, W., Estes, M.K., Prasad, B.V., 2009. Rotavirus architecture at subnanometer resolution. *Journal of Virology* 83, 1754–1766.
- Li, X., Zhou, N., Chen, W., *et al.*, 2017. Near-atomic resolution structure determination of a cypovirus capsid and polymerase complex using cryo-EM at 200 kV. *Journal of Molecular Biology* 429, 79–87.
- Luque, D., Mata, C.P., Suzuki, N., Ghabrial, S.A., Caston, J.R., 2018. Capsid Structure of dsRNA Fungal Viruses. *Viruses* 10.
- Lu, X., McDonald, S.M., Tortorici, M.A., *et al.*, 2008. Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. *Structure* 16, 1678–1688.
- Lu, G., Zhou, Z.H., Baker, M.L., *et al.*, 1998. Structure of double-shelled rice dwarf virus. *Journal of Virology* 72, 8541–8549.
- McClain, B., Settembre, E., Temple, B.R., Bellamy, A.R., Harrison, S.C., 2010. X-ray crystal structure of the rotavirus inner capsid particle at 3.8 Å resolution. *Journal of Molecular Biology* 397, 587–599.
- Mumtsidu, E., Makhov, A.M., Roessle, M., Bathke, A., Tucker, P.A., 2007. Structural features of the Bluetongue virus NS2 protein. *Journal of Structural Biology* 160, 157–167.
- Naitow, H., Tang, J., Canady, M., Wickner, R.B., Johnson, J.E., 2002. L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nature Structural & Molecular Biology* 9, 725–728.
- Nakagawa, A., Miyazaki, N., Higashiura, A., 2018. Hierarchical structure assembly model of rice dwarf virus particle formation. *Biophysical Reviews* 10, 659–665.
- Nakagawa, A., Miyazaki, N., Taka, J., *et al.*, 2003. The atomic structure of rice dwarf virus reveals the self-assembly mechanism of component proteins. *Structure* 11, 1227–1238.
- Nason, E.L., Rothagel, R., Mukherjee, S.K., *et al.*, 2004. Interactions between the inner and outer capsids of bluetongue virus. *Journal of Virology* 78, 8059–8067.
- Ogden, K.M., Hu, L., Jha, B.K., *et al.*, 2015. Structural basis for 2'–5'-oligoadenylate binding and enzyme activity of a viral RNase L antagonist. *Journal of Virology* 89, 6633–6645.
- Patton, J.T., Jones, M.T., Kalbach, A.N., He, Y.W., Xiaobo, J., 1997. Rotavirus RNA polymerase requires the core shell protein to synthesize the double-stranded RNA genome. *Journal of Virology* 71, 9618–9626.
- Periz, J., Celma, C., Jing, B., *et al.*, 2013. Rotavirus mRNAs are released by transcript-specific channels in the double-layered viral capsid. *Proceedings of the National Academy of Sciences of the United States of America* 110, 12042–12047.
- Pesavento, J.B., Crawford, S.E., Estes, M.K., Prasad, B.V., 2006. Rotavirus proteins: structure and assembly. *Current Topics in Microbiology and Immunology* 309, 189–219.
- Pesavento, J.B., Lawton, J.A., Estes, M.E., Venkataram Prasad, B.V., 2001. The reversible condensation and expansion of the rotavirus genome. *Proceedings of the National Academy of Sciences of the United States of America* 98, 1381–1386.
- Poranen, M.M., Mantynen, S., ICTV Report Consortium, 2017. ICTV virus taxonomy profile: Cystoviridae. *Journal of General Virology* 98, 2423–2424.
- Prasad, B.V., Burns, J.W., Marietta, E., Estes, M.K., Chiu, W., 1990. Localization of VP4 neutralization sites in rotavirus by three-dimensional cryo-electron microscopy. *Nature* 343, 476–479.
- Prasad, B.V., Rothnagel, R., Zeng, C.Q., *et al.*, 1996. Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. *Nature* 382, 471–473.
- Prasad, B.V., Wang, G.J., Clerx, J.P., Chiu, W., 1988. Three-dimensional structure of rotavirus. *Journal of Molecular Biology* 199, 269–275.
- Prasad, B.V., Prevelige, P.E. Jr., 2003. Viral genome organization. *Advances in Protein Chemistry* 64, 219–258. doi:10.1016/S0065-3233(03)01006-4. PMID: 13677049.
- Reinisch, K.M., Nibert, M.L., Harrison, S.C., 2000. Structure of the reovirus core at 3.6 Å resolution. *Nature* 404, 960–967.
- Reiss, K., Stencel, J.E., Liu, Y., *et al.*, 2012. The GM2 glycan serves as a functional coreceptor for serotype 1 reovirus. *PLoS Pathogens* 8, e1003078.
- Reiter, D.M., Frierson, J.M., Halvorson, E.E., *et al.*, 2011. Crystal structure of reovirus attachment protein sigma1 in complex with sialylated oligosaccharides. *PLoS Pathogens* 7, e1002166.
- Roy, P., 2013. Orbiviruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Roy, P., 2017. Bluetongue virus structure and assembly. *Current Opinion in Virology* 24, 115–123.
- Salgado, E.N., Garcia Rodríguez, B., Narayanaswamy, N., Krishnan, Y., Harrison, S.C., 2018. Visualization of calcium ion loss from rotavirus during cell entry. *Journal of Virology* 92.
- Settembre, E.C., Chen, J.Z., Dormitzer, P.R., Grigorieff, N., Harrison, S.C., 2011. Atomic model of an infectious rotavirus particle. *EMBO Journal* 30, 408–416.
- Shaw, A.L., Rothnagel, R., Chen, D., *et al.*, 1993. Three-dimensional visualization of the rotavirus hemagglutinin structure. *Cell* 74, 693–701.
- Staubert, N., Martínez-Costas, J., Sutton, G., Monastyrskaya, K., Roy, P., 1997. Bluetongue virus VP6 protein binds ATP and exhibits an RNA-dependent ATPase function and a helicase activity that catalyze the unwinding of double-stranded RNA substrates. *Journal of Virology* 71, 7220–7226.
- Sutton, G., Grimes, J.M., Stuart, D.I., Roy, P., 2007. Bluetongue virus VP4 is an RNA-capping assembly line. *Nature Structural & Molecular Biology* 14, 449–451.

- Tao, Y., Farsetta, D.L., Nibert, M.L., Harrison, S.C., 2002. RNA synthesis in a cage—structural studies of reovirus polymerase lambda3. *Cell* 111, 733–745.
- Taraporewala, Z., Chen, D., Patton, J.T., 1999. Multimers formed by the rotavirus nonstructural protein NSP2 bind to RNA and have nucleoside triphosphatase activity. *Journal of Virology* 73, 9934–9943.
- Taraporewala, Z.F., Patton, J.T., 2001. Identification and characterization of the helix-destabilizing activity of rotavirus nonstructural protein NSP2. *Journal of Virology* 75, 4519–4527.
- Taraporewala, Z.F., Schuck, P., Ramig, R.F., Silvestri, L., Patton, J.T., 2002. Analysis of a temperature-sensitive mutant rotavirus indicates that NSP2 octamers are the functional form of the protein. *Journal of Virology* 76, 7082–7093.
- Tenorio, R., Fernandez Decastro, I., Knowlton, J.J., *et al.*, 2019. Function, architecture, and biogenesis of reovirus replication neoorganelles. *Viruses* 11.
- Trask, S.D., Kim, I.S., Harrison, S.C., Dormitzer, P.R., 2010. A rotavirus spike protein conformational intermediate binds lipid bilayers. *Journal of Virology* 84, 1764–1770.
- Viskowska, M., Anish, R., Hu, L., *et al.*, 2014. Probing the sites of interactions of rotaviral proteins involved in replication. *Journal of Virology* 88, 12866–12881.
- Wang, X., Zhang, F., Su, R., *et al.*, 2018. Structure of RNA polymerase complex and genome within a dsRNA virus provides insights into the mechanisms of transcription and assembly. *Proceedings of the National Academy of Sciences of the United States of America* 115, 7344–7349.
- Yang, J., Cheng, Z., Zhang, S., *et al.*, 2014. A cypovirus VP5 displays the RNA chaperone-like activity that destabilizes RNA helices and accelerates strand annealing. *Nucleic Acids Research* 42, 2538–2554.
- Yang, C., Ji, G., Liu, H., *et al.*, 2012. Cryo-EM structure of a transcribing cypovirus. *Proceedings of the National Academy of Sciences of the United States of America* 109, 6118–6123.
- Yoder, J.D., Trask, S.D., Vo, T.P., *et al.*, 2009. VP5* rearranges when rotavirus uncoats. *Journal of Virology* 83, 11372–11377.
- Zeng, C.Q., Estes, M.K., Charpilienne, A., Cohen, J., 1998. The N terminus of rotavirus VP2 is necessary for encapsidation of VP1 and VP3. *Journal of Virology* 72, 201–208.
- Zhang, X., Boyce, M., Bhattacharya, B., *et al.*, 2010a. Bluetongue virus coat protein VP2 contains sialic acid-binding domains, and VP5 resembles enveloped virus fusion proteins. *Proceedings of the National Academy of Sciences of the United States of America* 107, 6292–6297.
- Zhang, X., Ding, K., Yu, X., *et al.*, 2015. In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* 527, 531–534.
- Zhang, X., Jin, L., Fang, Q., Hui, W.H., Zhou, Z.H., 2010b. 3.3 Å cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry. *Cell* 141, 472–482.
- Zhang, X., Ji, Y., Zhang, L., *et al.*, 2005. Features of reovirus outer capsid protein mu1 revealed by electron cryomicroscopy and image reconstruction of the virion at 7.0 Å resolution. *Structure* 13, 1545–1557.
- Zhang, X., Patel, A., Celma, C.C., *et al.*, 2016. Atomic model of a nonenveloped virus reveals pH sensors for a coordinated process of cell entry. *Nature Structural & Molecular Biology* 23, 74–80.
- Zhao, Y., Thomas, C., Bremer, C., Roy, P., 1994. Deletion and mutational analyses of bluetongue virus NS2 protein indicate that the amino but not the carboxy terminus of the protein is critical for RNA-protein interactions. *Journal of Virology* 68 (4), 2179–2185. Epub 1994/04/01. doi:10.1128/JVI.68.4.2179-2185.1994. PubMed PMID: 8139002; PMCID: PMC236693.

Structures of Tailed Phages and Herpesviruses (*Herpesviridae*)

Montserrat Fàbrega-Ferrer and Miquel Coll, Institute for Research in Biomedicine, Barcelona, Spain and Institute for Molecular Biology of Barcelona, Barcelona, Spain

© 2021 Published by Elsevier Ltd.

Nomenclature

CATC Capsid-associated tegument complex
cryo-EM Cryo-electron microscopy
dsDNA Double-stranded DNA
EBV Epstein-Barr virus
HCMV Human cytomegalovirus
HK97 Hong Kong 97

HSV Herpes simplex virus
KSHV Kaposi's sarcoma-associated herpesvirus
MCP Major capsid protein
MTP Major tail protein
SCP Smallest capsid protein
TMP Tape measure protein
VZV Varicella-zoster virus

Glossary

Capsid or head Icosahedral shell that encloses the genomes of tailed phages and herpesviruses.

Concatamer Continuous nucleic acid molecule made up of multiple head-to-tail linked unit-length viral genomes.

Cryo-electron microscopy (cryo-EM) Image reconstruction-based structural biology technique that has been used in recent years to describe the structure of viruses and viral macromolecules at atomic or quasi-atomic resolution.

Genome packaging Viral cycle process by which the genome is encapsidated into preformed procapsids or proheads.

Portal Dodecameric ring-like assembly located at a unique capsid vertex that plays a key role during genome ejection, procapsid morphogenesis and genome packaging processes.

Procapsid or prohead Immature form of the capsid or head that is a necessary precursor for the genome packaging process.

Tail Appendage machinery of the *Caudovirales* structure involved in host recognition and genome delivery.

Tegument In mature herpesvirus virions, proteinaceous cluster located in the space between the capsid and the envelope.

Triangulation (T) number Number of smaller triangles (facets) contained in each of the 20 faces of the icosahedral capsid. The number of coat protein subunits in an isometric virus is calculated as 60 T, because there are three subunits per facet.

Virion Mature viral particle. In bacteriophages, it comprises the capsid filled with the genome and the tail. In the case of herpesviruses, it is formed by the lipidic envelope, the tegument and the capsid with the genome.

Introduction

This article reviews the structure of tailed phages and herpesviruses. Although their respective hosts are very distant from a phylogenetical perspective, these two viral groups share some morphology features that suggest an evolutionary linkage between them with a putative ancient common ancestor.

The chapter starts with a separate explanation of the architecture of tailed phages and herpesvirus virions, and then both are compared from a structural point of view. In this regard, it is important to mention that in recent years the “resolution revolution” provided by cryo-electron microscopy (cryo-EM) has made a significant contribution to the structural virology field, with many new atomic or quasi-atomic models available for both whole viral particles and specific protein complexes. The combination of such information with that yielded by X-ray crystallography offers an unprecedented amount of knowledge about tailed phage and herpesvirus virion architecture. Finally, the last section of the article gives an insight into the structural basis of key processes during the viral infection cycle.

Tailed Phages

Prokaryote-infecting viruses, which are called bacteriophages or phages, present enormous diversity in terms of morphology and type of nucleic acids that form their genome. Among these viruses, the *Caudovirales* order comprises the so-called double-stranded DNA (dsDNA) tailed phages, probably the most abundant type of organisms on Earth and found in a wide range of ecosystems. First discovered independently by Frederick W. Twort and Felix d’Herelle (1915 and 1917, respectively), some *Caudovirales* species have been extensively used as model organisms in biochemistry, genetics and molecular biology studies. Moreover, phage therapy emerged as a potential treatment for bacterial diseases, but the discovery of antibiotics reduced research in that field. However, the appearance of multidrug-resistant bacteria is now reigniting interest in therapeutic applications for phages.

Caudovirales are thought to be among the oldest microorganisms on Earth because they are able to infect both Bacteria and Archaea. Genome comparison studies show a huge genetic diversity among tailed phages, revealing a mosaic nature caused by frequent homologous recombination and horizontal gene transfer events. Nevertheless, all *Caudovirales* species present a shared morphology, with an overall architecture of the virion or mature viral particle outside host cells consisting of two main structural elements, namely a head, which encloses the genome, and a tail appendage, which is in charge of host recognition and dsDNA delivery into the cell.

Icosahedral Head or Capsid

Phage heads are typically icosahedral, which is the cubic point symmetry arrangement that offers a lowest surface/volume ratio and therefore the most economic from a genetic point of view. Icosahedrons are formed by 20 triangular faces and contain three types of symmetry axes (two-fold, three-fold and five-fold) that yield 60 equivalent positions. However, heads may contain more than 60 structural subunits in not strictly equivalent positions when each face of the icosahedron is built by more than one smaller triangle or facet. The triangulation number (T) is used to define the specific geometric arrangement of each head, as it reflects the number of facets present per face. There is a relationship between the number of structural subunits present in a virus and its T number, the former being around 60 times the latter because there are 20 faces per virion and three equivalent structural subunits per facet. It is important to mention that around 15% of tailed phage species have heads that do not present isometric icosahedral symmetry, but rather show a prolate or elongated morphology with a tubular middle section between two icosahedral ends. In these cases, previous estimation of the number of major capsid proteins (MCPs) present is not correct, and structural subunits in the middle section should also be included in the calculation.

The structural subunits that comprise the head are called capsomers, which are either hexamers or pentamers of the MCP, named hexons and pentons, respectively. Hexons form the capsid faces and edges, while pentons are located at eleven of the five-fold vertices, all but one holding the portal protein. In phages with prolate heads, the portal vertex is located at one end of the long axis of the head. Given that the portal vertex also has five-fold symmetry and the portal is a ring-shaped dodecamer, there is a symmetry mismatch, which means that there cannot be identical interactions between portal and MCP subunits. In some phages, other proteins associate with the MCP's, sometimes stabilizing the head and building connections between capsomers. Finally, it is important to mention that some phages contain head fibers protruding from the capsid.

T4 phage, for instance, presents a prolate head comprised mainly of the MCP gp23, which forms the hexons. Although pentamers are usually formed by the same MCP, T4 pentons are built with the capsid vertex protein gp24, which is found in eleven vertices. The N-terminal ends of both gp23 and gp24 are cleaved in mature capsids. A gp20 dodecamer is located at the unique vertex, building the portal. Moreover, the small outer capsid (Soc) and highly immunogenic outer capsid (Hoc) proteins are also found at the outer surface of the mature head. The former provides the head with stability against extreme temperatures and pHs, while the latter is probably involved in the infection of the *Escherichia coli* host (Fig. 1(A)).

The main function of phage heads or capsids is to contain the viral genome and protect it from the environment when virions are outside the host cells. Given that tailed phages do not need to disassemble upon cell infection in order to release the genome, because it is ejected from a unique capsid point, heads can be built as highly stable shells. Moreover, capsids are extremely resistant since the genome is packaged inside them at near-crystalline densities (> 500 mg/mL), which causes huge internal pressures. Apart from the genome, the head of some phages also contains an inner core of proteins, usually assembled at the unique vertex interacting with the portal. The core proteins typically form a cylindrical structure with a central channel continuous with that of the portal. Core proteins are thought to have important functions during genome ejection.

Several models suggested distinct arrangements of the genome in the form of solenoids, liquid crystals or folded toroids inside phage heads. High-resolution structures have now revealed concentric shells at regular intervals within the capsid corresponding to the genome. In phages with an inner core, the dsDNA is spooled around it and in some cases density corresponding to DNA is observed in the center of the core protein channel. Moreover, some structures also indicate the presence of dsDNA within the portal and proximal part of tail channel, with the genome ready to be injected upon infection of the host cell.

The unique vertex where the portal is found is associated with a number of specific functions. It is where the DNA packaging and ejection processes occur and the docking point where neck and/or tail proteins attach. Therefore, the phage portal is sometimes also referred to as connector, because in mature virions it links the head to the rest of the particle.

Tail

The main function of the tail is to contact the host cell during infection and to deliver the genome into its cytoplasm. Most tails present protruding elongated proteins called fibers whose tip domains are key for host cell recognition and binding. The tail is assembled at the portal vertex and it attaches to the head by specific protein complexes called collar, gatekeeper or adapter. Directly interacting with the portal, these are proteins with a dodecameric architecture containing four α -helices per monomer, an arrangement that is conserved in the different phage families. In some cases, a second ring formed by β -sheets called stopper has been described to build a plug that closes the channel.

Classical taxonomic studies of phages were strongly based on the virion morphology when observed by negative stain electron microscopy. Therefore, the *Caudovirales* order comprises three families depending on the morphology and complexity of their tails:

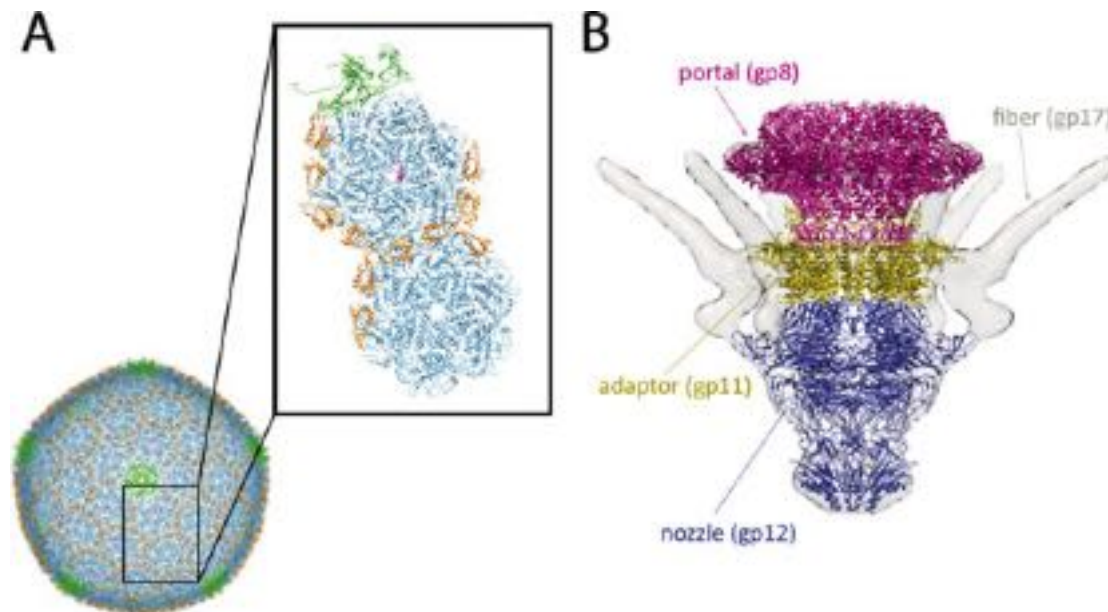


Fig. 1 Structure of tailed phages virions. (A) T4 isomeric head (PDB 5VF3) formed by the MCP building the hexons (blue), the capsid vertex protein (green), the small outer capsid protein (orange) and the highly immunogenic outer capsid protein (pink). The structural assembly depicted in the zoom is repeated 60 times in order to build the whole isometric capsid. (B) Fiberless T7 tail complex (PDB 6R21) comprising the portal (magenta), adapter (yellow) and nozzle (dark blue) proteins fitted into the tail volume (EMD-5689). Regions corresponding to the fibers are also indicated.

Podoviridae have short non-contractile tails, *Siphoviridae* long non-contractile tails and *Myoviridae* long contractile tails. However, as tail structure must be able to adapt to each specific host cell, they are quite variable even among the same family. At this point, it is important to comment that new insights into phage evolution might question the utility of the classical taxonomic families. Nevertheless, the aforementioned classification based on the tail morphology is still the one most widely accepted.

Podoviridae

Viruses of this family present short simple tails that are assembled directly into the viral head and attached to the portal by the adapter protein. Most of the known *Podoviridae* tail structures are formed by a central nozzle that interacts with the fiber appendages.

Analysis of the available structures shows specific differences depending on the phage. For instance, P22, which infects *Salmonella typhimurium*, has a plug for the DNA channel with a trimeric needle-like structure that is released to penetrate the cell envelope during infection. Another example is the *Bacillus subtilis* phage Φ 29, where 12 trimeric fibers are bound to its collar region. *Podoviridae* also includes T7 phage, which infects *E. coli* and whose fiberless tail structure in complex with the portal protein was solved (**Fig. 1(B)**). The cryo-EM 3.3 Å resolution structure shows a dodecameric adapter that interacts directly with the hexameric nozzle, whose fold is organized around six central seven-bladed β -propellers, one per monomer. The β -propellers are placed with their planes radial and are essential to tightly close the channel gate in the mature phage. The remaining domains of the protein are organized around them.

Siphoviridae

This family is characterized by a relatively well conserved long non-contractile flexible tail formed mainly by a long-tube built by copies of a major tail protein (MTP) organized in the form of a ring of homohexamers. Tail tube length is determined by the tape measure protein (TMP), located at the tube channel, and the terminator protein, which is found at the end proximal to the head. These two proteins act together to stop the polymerization of the MTP once the required size of the tail has been reached. Finally, the absorption apparatus is a variable host recognition device found at the distal end of the tail. Some *Siphoviridae* contain a baseplate structure, although this complex architecture is more typical of the *Myoviridae* family.

One of the most studied *Siphoviridae* is the *E. coli*-infecting λ phage, which presents a 150 nm long tail while its head has a diameter of only around 63 nm. Although this phage presents side tail fibers, most laboratory strains carry a mutation that eliminate them. Its MTP, gpV, has two domains, and its N-terminal one has a fold which is conserved among proteins of different phage families. Another well-known *Siphoviridae* phage is SPP1, which infects *Bacillus subtilis*. The structure of its tail has been studied by electron microscopy, showing a rotation of around 21° between adjacent hexameric rings of the MTP. Other members of this family are T5 and Hong Kong 97 (HK97), which infect *E. coli*, and the thermophilic phage P23-45, which infects *Thermus thermophilus*.

Myoviridae

Viruses belonging to the *Myoviridae* family have long contractile tails, which are formed by the following parts: the baseplate, which is the part in charge of the contact with the surface of the host cell; the sheath, which builds the outer protein cylinder of the tail and provides a contractile capacity; the tail tube, which assembles the inner part of the channel; and fibers, which recognize the host. Moreover, TMP and terminator proteins have also been described to regulate the length of the tail, and some *Myoviridae* terminators are homologous to that of the λ phage.

A prototypical member of *Myoviridae* is the *E. coli*-infecting T4, which has a complex 100 nm long tail and a head width of 85 nm. Its head and tail are connected by a neck, which presents a thick collar formed by hexameric rings with protruding whiskers that control the extension of the fibers. The tail tube is the scaffold for sheath assembly, which is formed by 23 stacked hexameric rings. The T4 baseplate contains at least 16 different proteins and also has 6-fold symmetry. Long and short tail fibers protrude out of the baseplate, but the short ones are folded under it until the long ones contact the primary cellular receptor. Only then do short fibers also unfold and recognize a secondary receptor. Other well-known phages of the *Myoviridae* family include Mu and P1, which also infect *E. coli*.

Herpesviruses

Herpesviridae is a family of enveloped dsDNA viruses able to infect eukaryotic cells. Herpesviruses infect many animal species, including mammals, fishes, birds, amphibians, reptiles and some types of invertebrates. Although herpesvirus-caused diseases have been known for centuries, these viruses were not detected until the beginning of the 20th century. The following nine herpesviruses belonging to three distinct subfamilies are known human pathogens: herpes simplex virus types 1 and 2 (HSV-1, HSV-2) and varicella zoster virus (VZV) are *Alphaherpesvirinae*; human cytomegalovirus (HCMV) and roseolovirus (human herpesvirus 6A, human herpesvirus 6B and human herpesvirus 7) are *Betaherpesvirinae*; and Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated virus (KSV) are *Gammapherpesvirinae*. Herpesviruses have a wide range of pathogenic effects in humans, from inapparent or mild to life-threatening, and some have been linked to specific types of cancer. For instance, EBV is associated with lymphomas and gastric and nasopharyngeal carcinomas. After primary infection, herpesviruses remain latent in host tissues throughout lifetime and retain the capacity to reactivate under certain conditions, causing recurrent diseases.

Mature herpes virions are enveloped spherical particles with a diameter of around 200 nm. Inside the lipidic envelope there is a capsid containing the genome (also called nucleocapsid) and covered by a proteinaceous tegument.

Herpesvirus genomes in mature virions are single linear dsDNA molecules containing 120–300 kbp. They are made up of two covalently linked components, named U_L and U_S , and contain terminal repeats. During the latency phase, viral genomes are usually maintained as circular episomes replicated by the cellular machinery, and at low frequencies they may be found integrated in the host chromosomes. Within the capsid, they are found in a near-crystalline state, in a left-handed spool forming concentric shells around a disordered, ellipsoidal dsDNA core.

The capsids of herpes virions are icosahedral and they have a diameter of around 120–130 nm. They have a structure similar to that of non-prolate phage heads and are formed by 161 hollow capsomers (150 hexons and 11 pentons) plus the portal located at a unique five-fold vertex. The MCP subunits are joined in 3-fold symmetry points by triplex structures, which contain two copies of one protein plus a single copy of another one. These triplex structures serve to join capsomers and stabilize the viral shell. Capsids may contain small amounts of other viral proteins.

The tegument, which is made of partially amorphous proteinaceous globular material, surrounds the capsid. It has an asymmetrical distribution, with the capsid not necessarily located at its center, and presents poor structural definition except for a certain degree of icosahedral arrangement in the region closest to the capsid. The pleomorphic tegument is formed by around 20 different proteins encoded in the viral genome, and several of them are related to herpesvirus gene expression. Tegument proteins are less conserved among species than capsid proteins.

Both the nucleocapsid and the tegument are surrounded by the envelope, a lipidic bilayer decorated with glycoproteins. Such macromolecules are important for the subsequent infectivity of the virions, as they interact with cell receptors prior to the fusion of the host membrane cell with the viral envelope. Glycoproteins can be virus-encoded or have a cellular origin, and when virions are visualized by negative stain microscopy they look like projecting spikes on the surface of the particle.

Regarding *Alphaherpesvirinae*, a cryo-EM structure of the HSV-1 capsid with proteins of the capsid-associated tegument complex (CATC) is available at 4.2 Å resolution (Fig. 2(A)). The $T = 16$ capsid contains pentons, three types of hexons (peripentonal, edge and center) and six types of triplexes (Ta to Tf). Atomic models are available for the MCP (VP5), the triplex dimer protein (VP23), the triplex monomer protein (VP19c) and the smallest capsid protein (SCP, VP26). At capsid vertexes, there are five copies of the star-shaped heteropentameric CATC, containing two copies of pUL25 and pUL36, respectively, and a single copy of pUL27. Each CATC is formed by a region contacting with the triplexes, an extended five-helix bundle, and a head region placed between two MCP subunits of the penton. The C-terminal α -helix of pUL36 is thought to mediate the linking of the capsid to the outer tegument and envelope, as it interacts with other tegument proteins that have contact with envelope glycoproteins. An equivalent structure of HSV-2 has also been solved by cryo-EM at 3.75 Å, and in both cases CATC seems to contribute to the exceptional stability of HSV virions, thereby reinforcing the most stressed parts of capsids, the vertexes.

HCMV, which belongs to the *Betaherpesvirinae* subfamily, has a large 235 kbp genome, although the size of its capsid is similar to that of the herpesviruses previously described. Therefore, the effect of the highly pressured genome is even more remarkable in this case. A 3.9 Å resolution cryo-EM structure reveals the presence of capsid stabilization mechanisms conserved across herpesviruses, based on

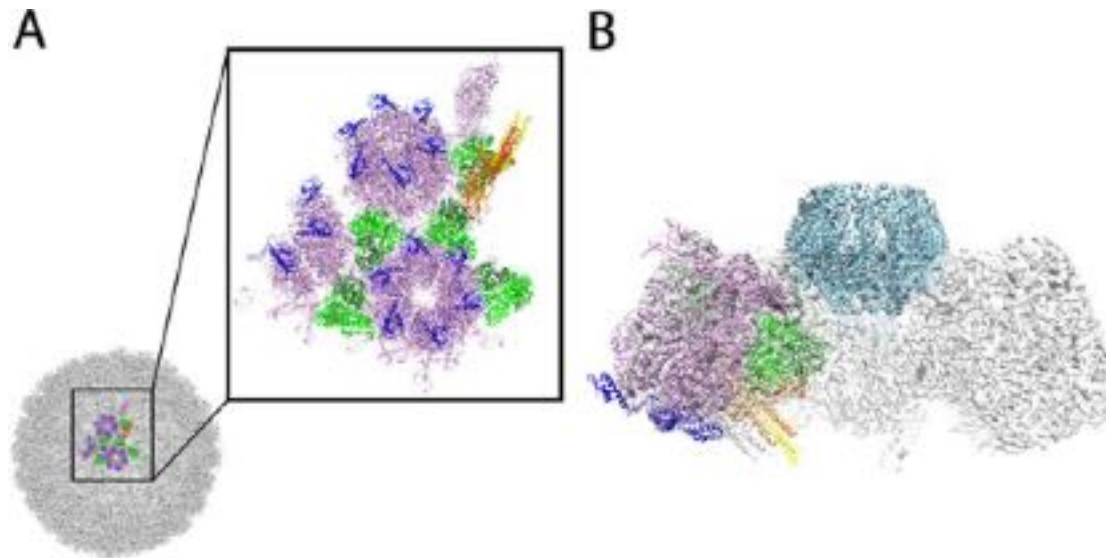


Fig. 2 Structure of herpesviruses virions. A. HSV-1 capsid (EMD-7472 and PDB 6CGR). The cryo-EM map is in gray, and the zoomed region shows the atomic structures of the MCP (pink), the small capsomere-interacting protein (blue), the triplex capsid proteins 1 and 2 (green forest and green, respectively), the capsid vertex components 1 and 2 (orange and orange-red) and the large tegument protein deneddylase (yellow). The structural assembly depicted in the zoom is repeated 60 times in order to build the whole capsid. B. Structure of the unique vertex portal (EMD-9861, PDB 6OD7 and PDB 6ODM). The map appears in gray, and atomic models of the same proteins as in panel A with the same color code are shown for one of the 5-fold symmetry components building the vertex. The atomic model of the dodecameric portal is depicted in light blue.

the role of MCP subunits and triplexes. Moreover, some specific features have been described, such as the presence of hexon channels accommodating parts of the genome and facilitating the packaging of DNA, and the role of a *Betaherpesvirinae* specific tegument protein, pp150, which interacts with the SCP and builds a layer that further stabilizes the capsid. The structure of human herpesvirus 6B shows that the pp150 homologous protein, pUL11, is also present on the surface of the capsid but in lower amount. This observation suggests that the significantly smaller genome size of the virus (162 kbp) diminishes the need for stabilization.

CATC homologs have also been described in *Gammaherpesvirinae*, as shown by the structure of KSV. However, they present a lower occupancy compared to those present in the other herpesviruses, with a long-range asymmetric attachment determined by the unique portal vertex. CATCs are preferentially associated with this vertex and the portal-proximal sides of the penton vertices. Therefore, the binding of CATC seems to be penton-independent and would depend on contacts with peripentonal hexons and triplexes.

Structures of the portal vertex obtained using symmetry-relaxation cryo-EM strategies have also provided more insights into the architecture of the unique capsid vertex, which has sometimes been described to present a tail-like structure. Tentacle-like helices extend from the HSV-1 portal base to an outer vertex-capping density that anchors the genome terminus, which extends outside of the portal protein channel (**Fig. 2(B)**). Both helical structures present a five-fold symmetry like the capsid vertex, which produce a symmetry mismatch with the dodecameric portal.

Structural Comparison of Tailed Phages and Herpesviruses

As mentioned at the beginning of the article, structural and functional similitudes between tailed phages and herpesviruses suggest a phylogenetic linkage between them. Both share certain replication and transcription patterns, and in some cases plasmid-type and integrative lysogeny processes.

Regarding the mature architecture of tailed phages and herpesviruses, the main difference is that while the former are non-enveloped particles, the latter do have an envelope. Therefore, the crucial role of protecting the genome from the environment is performed entirely by the capsid shell in the case of tailed phages while in herpesviruses the envelope and tegument also play an important role. Nevertheless, phage heads and herpes capsids share the icosahedral arrangement built by equivalent organizations of their capsomers and have a portal, which is crucial during capsid assembly and genome encapsidation and release processes. The heads of tailed phages show significant differences in their T numbers, while herpesviruses have always a T number of 16, which is within the range of known arrangements described for tailed phages.

MCP Fold

Sequence similarity matches to assess evolutionary relationships may fail when homologous proteins have diverged significantly. However, in such cases, structural analysis can be useful to detect evolutionary relationships. The most widely studied structural feature shared between mature virions of tailed phages and herpesviruses is the conserved fold of the MCP. Structural data have shown that

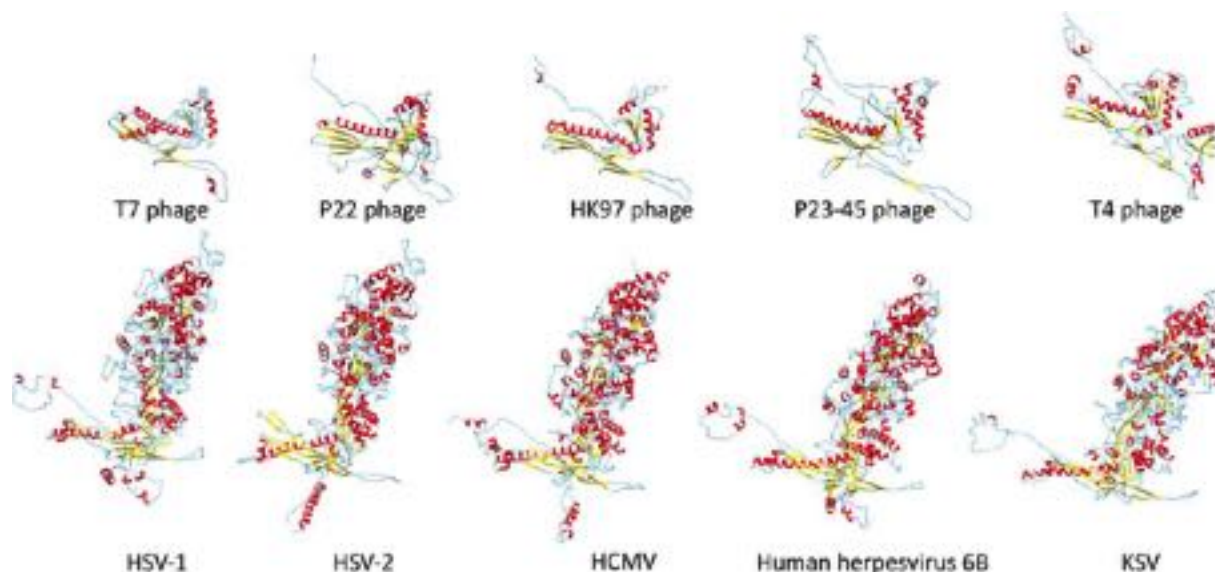


Fig. 3 MCP HK97-fold conservation. Monomeric MCP structures. First line corresponds to phage proteins: T7 (PDB 3J7V), P22 (PDB 5UU5), HK97 (PDB 1OHG), P23-45 (PDB 6I9E) and T4 (PDB 5VF3). Second line corresponds to herpesvirus proteins: HSV-1 (PDB 6CGR), HSV-2 (PDB 5ZZ8), HCMV (PDB 5VKU), human herpesvirus 6B (PDB 6Q1F) and KSV (PDB 6PPB). Structures are depicted with the same orientation with α -helices in red, β -strands in yellow and loops in light blue. The tailed phages HK-97 fold is conserved in the floor domain of the herpesvirus proteins, at the bottom region of their structure.

MCP topology is conserved, and both types of viruses form the so-called HK97 lineage, as they share the basic fold of the phage HK97 MCP, the first homolog described (Fig. 3). The structure of the HK97 MCP was determined by X-ray crystallography, and although it does not share sequence similarity with the MCP of other phages, available structures have shown that proteins from phages belonging to different families, including T4, P22 or Φ 29, share the same fold and arrangement. In the case of T4, both the MCP and capsid vertex proteins have the HK97 fold. Basically, this fold consists of three domains: the A domain, the P domain and the E loop. The A domain, with a V-shape, builds the apex of the capsomers. The P domain typically consists of a long α -helix plus a β -sheet formed by three β -strands, and it is found at the capsomer periphery. Finally, the E loop expands from the P domain and performs contacts between neighboring subunits of the MCP in the capsid shell. However, the conserved fold of each viral species has modifications, thereby conferring the viruses specific characteristics and abilities.

In the case of HK97, crosslinking between MCP subunits has been described, with the E loop covalently bound to the P domain of another MCP placed on an adjacent capsomer. In the case of T4, the capsid vertex protein contains an extra insertion domain (I-domain) that binds to neighboring subunits from the same penton and MCP subunit from another capsomers by non-covalent interaction in equivalent location to those where crosslinking occurs in the case of phage HK97.

The phage P23-45 presents an increased spacing between capsomers in the icosahedral lattice, that allows the virus to build larger capsids able to fit bigger genomes. MCP presents an extended A-domain and a longer E loop. In mature P23-45 virions, the E-loops become well-defined β -strands that create a binding site for the N-terminal part of auxiliary surface proteins, which stabilize three-fold axes.

Regarding herpesviruses, the MCP is much larger, and although they contain the HK97 fold, they also have other domains. It is the floor domain that has the overall shape of the structural signature with conserved disposition of the secondary structure elements, while the middle and upper domains that form the capsomer towers are more divergent, even among herpesvirus families.

The canonical HK97 fold is not limited to viruses infecting bacteria and eukaryotes but has also been described in archaeal head-tailed viruses, like in the 8.9 Å resolution cryo-EM structure of the *Haloarcula sinatiensis* tailed virus 1 (HSTV-1). Therefore, viruses from archaea, bacteria and eukaryotic domains are evolutionarily connected, suggesting that viruses appeared before the three domains split a few billion years ago. Structural data suggest that this fold is a basic element on which to build stable capsids while at the same time it allows some conformational flexibility required during capsid morphogenesis and maturation.

Commonality between tailed phages and herpesvirus capsids is not restricted to the MCP fold. The architectures of triplex proteins also have a conserved fold with respect to the gpD auxiliary cementing protein from λ phage, which has β -strand identical topology. It has been suggested that evolution may have diversified the strategies of different viruses depending on the size of their genomes. While the cementing protein may have been replaced by the crosslinking process in the case of HK97, the size of the herpesviruses genomes requires the presence of triplexes.

Portal Protein

The portal proteins of different species differ in size and sequence, but there is a common structural architecture composed of 12 subunits of a single protein arranged in a ring-shaped dodecamer with a central channel (Fig. 4).

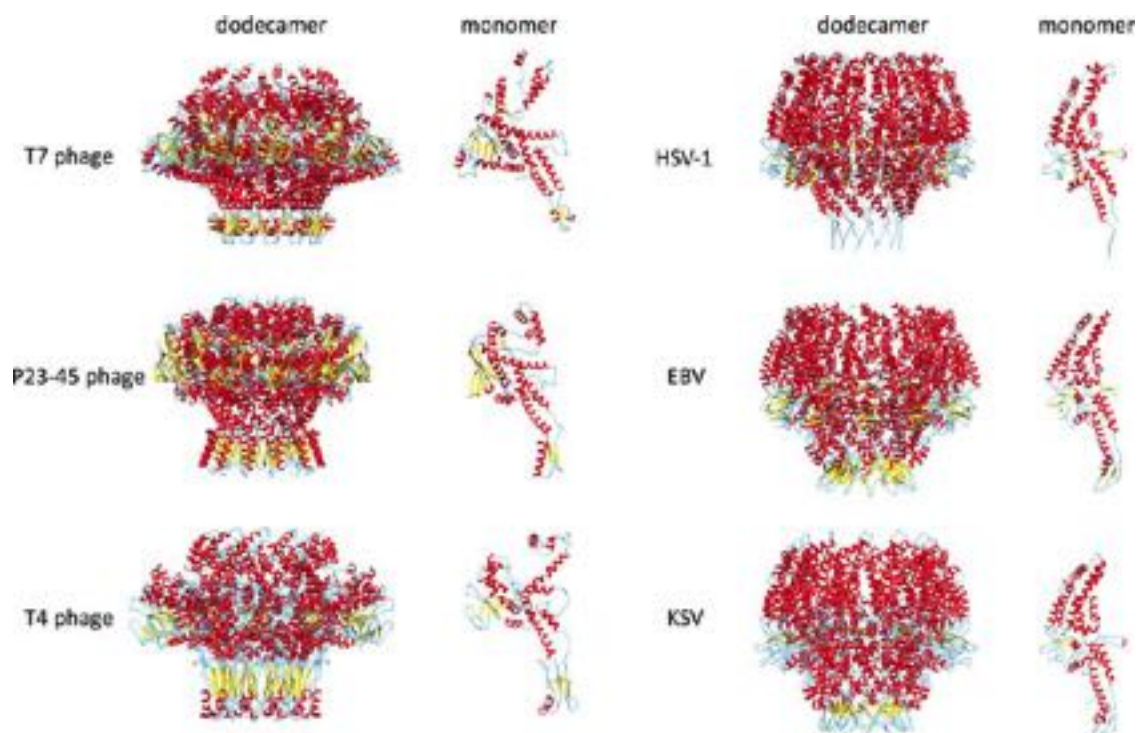


Fig. 4 Portal protein fold conservation. Portal structures are shown both as dodecamers and monomers. First column corresponds to phage proteins: T7 (PDB 6QX5), P23-45 (PDB 6IBG) and T4 (PDB 3JA7). Second column corresponds to herpesvirus proteins: HSV-1 (PDB 6OD7), EBV (PDB 6RVS) and KSV (PDB 6PPI). Structures are depicted with the same orientation with α -helices in red, β -strands in yellow and loops in light blue.

The following portal domains have been classically described for the portal of tail phages: the crown, inserted into the head interacting with the inner core if present; the wing, at the outermost part interacting with the capsid proteins; the stem, which builds the middle section of the channel wall, and the clip, located at the distal part with respect to the capsid and interacting with the adapter/gatekeeper in mature virions and with the terminase during packaging. Available structural data obtained both by X-ray crystallography and cryo-EM show that the most conserved region comprises the 24 helices of the stem domain, which forms the central part of the channel, while the wing and crown present a significant variability in size and shape. Given that the portal serves as a conduit for dsDNA passage, one of the most interesting parts of its structure is the channel region. An α -helix from the wing protrudes into the channel, and a flexible tunnel loop connects the protruding α -helix to one of the stem α -helices.

In the case of herpesvirus portals, the atomic structures of HSV-1, EBV and KSV were recently solved by cryo-EM, revealing a similar structural arrangement, also containing the characteristic stem 24 α -helices, but with some differences. Regarding the channel region, the two narrowest points have been defined for these three viruses. First, in an equivalent position to that of the protruding α -helix from tailed phages, there is a β -hairpin with a small tunnel loop between the two β -strands, forming an aperture perpendicular to the portal channel axis. Second, the region equivalent to the clip holds a β -barrel, formed by three β -strands per monomer arranged almost parallel to the channel axis. Each group of β -strands contains two parallel β -strands from the same monomer while the third one is antiparallel and comes from the clockwise neighbor. The portal structures of these three herpesviruses have a missing fragment, which should be located extending from the β -barrel clip. It has been hypothesized that the missing residues build the five sets of tentacle-like helices extending towards the cap, but this would imply a symmetry mismatch inside the same protein.

Structural Insights Into the Infective Viral Cycle

Similarities between tailed phages and herpesviruses go beyond their mature structures. Interestingly, they share some similar mechanisms along the whole viral infection cycle. The following sections review the main stages of this cycle, highlighting the differences from a structural and functional perspective when present.

Genome Ejection

The process by which the linear dsDNA molecule exits the viral particle when it infects a new host cell is called genome ejection. The genome of tailed phages and herpesviruses is released from the icosahedral capsid via the portal vertex, which acts as a channel

for dsDNA passage. However, many differences arise apart from that, given that the whole architecture of the virus and host cells to infect are very different.

In the case of phages with a tail, they do not enter the cell and therefore require sophisticated machineries to overcome the cell wall and plasmatic membranes. It is important to take into account that such appendages do not act like syringes, in fact dsDNA entry into the host cells is mainly passive as a result of dense packaging under very high pressure. In some phages, like T7 and $\Phi 29$, there is an initial passive ejection followed by a pause, after which injection resumes when proteins expressed from the first part of the genome ejected into the cytoplasm pull the rest of the genome.

In *Myoviridae* phages, like T4, some conformational changes occur after attachment of the short fibers to the secondary host receptor. The baseplate reorganizes as it binds to the cell, changing its shape from a hexagon to a star. This conformational change triggers the contraction of the sheath, whose subunits slide over each other and drive the tail tube through the membrane of the host cell, creating a channel for genome transport. The tip of the inner tail tube is the first part to pass through the outer membrane, and then lysozyme molecules from the tail are released to create a passage in the peptidoglycan layer of the cell wall. Finally, the tip of the tail tube also reaches the inner membrane, and a channel is created for dsDNA passage into the cell. Regarding *Siphoviridae*, for instance, SPP1 presents a rotation of the internal domain of the tail during genome ejection, propagating the signal from the tip to the gatekeeper. The tip is released after receptor binding, opening the basal part of the tail, and the TMP exits, allowing DNA to pass. Also in *Podoviridae*, some main conformational changes have been observed related to the position of fiber proteins and the opening of the end of the tail tube.

In addition to injecting DNA into the host cell, many viruses also inject proteins. For instance, T4 injects a protein that modifies host RNA polymerase function, which is not essential for phage infection. T7 bacteriophage injects large proteins of the inner core, which may build channels in the periplasmatic space for the passage of DNA and other proteins. In this case, these core proteins have been described as essential for phage growth.

If we now focus on herpesviruses, the first thing to note is that they do not have a tail machinery, and after the first contact with the host cell their capsid enters it via membrane fusion, a process that leaves it uncoated. The capsid and tegument are then released into the cytoplasm, and it is known that the capsids are transferred via microtubules to the nuclear pores. It is there where genome ejection occurs, with the capsid docked at the nuclear pore and the genome entering the nucleus. This process is largely unknown, but the tegument proteins from the portal vertex tail-like structure may be involved. Some tegument proteins are thought to reach the interior nucleus as well, possibly independently of the genome.

Therefore, the strategy by which herpesviruses release their genomes into the nucleus of the host cell may appear in some aspects analogous to that used by tailed phages to infect cells. Given the endosymbiotic theory, this similarity could reflect the conservation of an ancient infection pathway.

Genome Replication

Once released into the prokaryotic cell or the eukaryotic nucleus, the genome is replicated, usually via the rolling-circle mechanism, which produces a long head-to-tail concatamer. The size of genomes ranges from approximately 15 kbp to 500 kbp, without conserved genomic structure in most cases, this feature being common in only some genera.

There are some exceptions regarding the concatemeric form of genome replication. For instance, $\Phi 29$ replicates its genome in unit-lengths with terminal proteins, P1 phage has a circular genome, λ uses first the θ replication mechanism and the chromosome of Mu is surrounded by bacterial-derived sequences because it is inserted into the host genome and directly cleaved from there and packaged. For some herpesviruses a first phase of θ -type replication has been described to initiate replication before the rolling-circle mechanism starts.

Procapsid Morphogenesis

Procapsids, the precursors of mature capsids, contain no dsDNA and are not icosahedral but spherical. In some cases, they are smaller than mature capsids, although in other cases the volume does not increase upon maturation. Procapsids, also called proheads in the case of tailed phages, contain the portal, the MCP and scaffolding proteins. The interaction between these three components is key for determining the correct shape and size of the procapsid. The assembly process takes place in an ordered manner, with the building of some subassemblies, which are later joined. MCP and scaffolding proteins co-assemble and arrange around the portal protein, which acts as nucleation point or assembly initiator complex. The MCS proteins build a complete shell, devoid of DNA but filled with scaffolding proteins (**Fig. 5(A)**). Inner scaffolding proteins are present inside procapsids before dsDNA encapsidation and they facilitate morphogenesis of the particle, providing stability and preventing enclosure of inappropriate molecules. Some viruses have independent scaffolding proteins, while in others the role of these proteins is performed by a disposable part of the MCP. In many cases, the scaffolding protein is cleaved by proteases also present in the procapsid and expelled after assembly during genome translocation. In the case of some herpesviruses, it has been described that around 1200 copies of the scaffold protein are found on a single procapsid, and 150 of them are an extended form with a N-terminal protease domain. The scaffold protein is cleaved and lost, except for this protease domain, which remains in the mature capsid after DNA is packaged. When present, proteins of the inner core may also be in the procapsid, undertaking a largely unknown role during its morphogenesis but presenting structural changes during maturation.

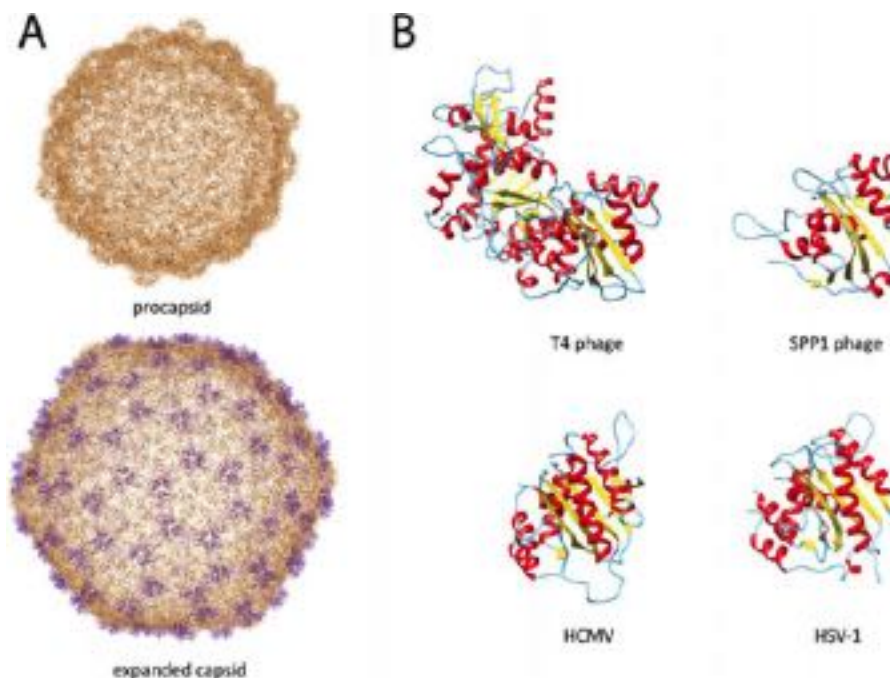


Fig. 5 *Procapsid morphogenesis and genome packaging.* (A) P23-45 procapsid and expanded capsid (PDB 6IBC and PDB 6I9E). Structures are shown at the same scale. The MCP is depicted in orange and the auxiliary protein in magenta. (B) Large terminase monomeric structures solved by X-ray crystallography. The structure of the T4 phage (PDB 3CPE) corresponds to the complete protein, while nuclease domains are shown in the case of SPP1 phage (PDB 2WC9), HCMV (PDB 3N4Q) and HSV-1 (PDB 4INOX). Structures are depicted with the same orientation, showing the common conserved fold, with α -helices in red, β -strands in yellow and loops in light blue. In the case of SPP1 and HCMV structures manganese atoms are represented in purple.

During capsid maturation, which occurs in parallel to genome packaging, other changes have been described, in addition to the cleavage and release of scaffolding proteins. A reorganization of the MCP domains occurs, with some conformational changes that confer an increase in stability throughout the formation of the mature capsids. MCP domain A is tilted around 20° with respect to domain P, and there is an increase in the sixfold symmetry of hexons. As mentioned before, in some cases there is an overall increase in the head volume, for instance, the diameter of the head of T7 increases by 12%–15% and the head has a thinner and smoother shell with respect to the prohead. As the outline of mature capsids is more symmetrical than those of procapsids, the bonding matrix between intersubunits changes, because the subunits of the MCP move in the plane and also along radial axes, thereby reinforcing ionic and hydrophobic interactions and producing a more stable and angularized capsid. After dsDNA packaging, additional decoration or stabilization proteins may also be added to the shell, further reinforcing it or conferring other properties.

Genome Packaging

Regarding dsDNA packaging, it is an energy-driven process that fills the procapsid with the genome, yielding the final mature capsid. The genome is recognized by the terminase complex, and the terminase-dsDNA complex docks at the procapsid portal vertex. The terminase has ATPase activity, which provides the energy for the translocation of the genome inside the capsid through the portal channel. This machinery is among the most highly energy-demanding described in biology. It has to push dsDNA into a closed space against a genome concentration gradient, exerting a considerable force, reaching 60 pN.

Although terminases provide an example of sequence similarity between tailed phages and herpesviruses, it is not yet clear whether they have different stoichiometries. In bacteriophages, terminases are formed by a small and large subunit: the small subunit binds the DNA and recognizes the specific sites, while the large subunit contains the ATPase and nuclease activity. Available structures show that terminases assemble at the portal vertex, where they form ring-shaped complexes. Phage $\Phi 29$, which is replicated in unit-length genomes with terminal proteins, has only one terminase protein and a packaging RNA (pRNA). In systems where no cut is required, packaging ATPase complex is a more accurate name than terminase, as the terminase complex receives its name because it cuts the DNA after encapsidation of a single unit genome. In herpesviruses, the terminase is a complex comprising at least three proteins. There is little structural information available about the complex, but the crystallographic structure of a nuclease domain of HCMV pUL89 and HSV-1 pUL15 shows an RNaseH/integrase-like fold, whose activity is dependent on manganese ions (Fig. 5(B)).

There are two different mechanisms to ensure that complete genomes are packaged: either specific recognition sites that are recognized by the packaging machinery, or terminally redundant genomes that are packaged with a headful system that fills

capsids with slightly more than one genome. Phages such as λ and herpesviruses employ the first strategy, while P22 and SPP1 use the second one.

Regarding the role of the portal during packaging, it has been suggested that it participates in signaling the end of the translocation. The portal may sense the pressure in the interior of the capsid and transmit this information to the packaging enzymes in order to trigger dsDNA cleavage and detachment of the terminase-dsDNA complex from the portal vertex. Synchronized movements of the wing and clip domains could allow a cross-talk mechanism between the interior and exterior of the capsid. Some portal mutants in SPP1 and P22 present altered chromosome lengths. An additional function of the portal may be the temporary retention of the dsDNA before the tail assembly, as has been described in the case of T7 bacteriophage. T7 portal presents two distinct conformations in its channel valve region, formed by the tunnel loop plus the α -helix protruding into the channel, which is able to bend and tilt 90°. The open state would allow passage of the genome, while the closed state would not. The closed state may serve to seal the capsid once the genome has been packaged and the terminase detached but the tail has not yet been assembled, thereby preventing dsDNA leakage.

As viral genomic translocation has no counterpart in cellular processes, the inhibition of packaging has been highlighted as a target for antiviral therapies.

Final Virion Assembly

When the terminase-dsDNA complex detaches from the portal, the final virions must be assembled, and at this point the process differs between tailed phages and herpesviruses. In bacteriophages, the neck and/or tail proteins bind at the portal vertex, yielding directly the mature virion. It has been proposed that collar, gatekeeper or adapter proteins may participate in the retention of the genome before the incorporation of the whole tail. In the case of *Siphoviridae* and *Myoviridae* like λ and T4, neck proteins assemble first, and then the whole pre-assembled tail binds to them. Tail morphogenesis usually begins with the formation of the initiator complex, which is the distal part of the tail, the tip or baseplate, plus the TMP complex. This substructure initiates the polymerization of the tubular section of the tail around the TMP, in the case of *Siphoviridae* built by the MTP, and in the *Myoviridae* also by the outer contractile sheath. Finally, when the whole TMP is enclosed, the terminator or capping proteins complete the assembly of the tail. The proteins, which stabilize the complex and bind it to the neck, present structural similarity between the two subfamilies. In contrast, *Podoviridae* phages like Φ 29 and P22 use a stepwise addition mechanism of tail components to the prepackaged head. The same occurs for the T7 bacteriophage, for which it has been described how first the dodecameric adapter oligomerizes upon binding to the head portal, and then the hexameric nozzle is assembled, building the tubular end of the tail. Finally, the incorporation of the six trimeric fibers generates the mature virion, which is released to the media after lysis of the bacterial cell.

In the case of herpesviruses, they use a non-lytic mechanism for exiting the infected cell. After genome encapsidation and assembly of some tegument proteins in the nucleus, the envelope is acquired when nucleocapsids bud through cell membranes during the late stages of viral assembly. The origin of the first viral envelope is the inner nuclear membrane, and it is formed immediately after the nucleocapsid and primary tegument are assembled at the nucleus. Afterwards, the particle is de-enveloped upon transit across the outer nuclear membrane. Once in the cytoplasm, the particle enters the Golgi apparatus, where additional tegument proteins are added and the final envelope is acquired by budding into post-Golgi cellular compartments. Finally, virions are released into the media after a reverse endocytosis process.

Conclusions

Structural analysis of the mature virions and viral infection cycles of tailed phages and herpesviruses reveals that, despite the multiple differences between these two viral groups, they share significant characteristics and traits. Given the similarities, it is reasonable to assume that they had a common ancestor a long time ago. Since herpesviruses are now much more uniform than tailed phages, it has been hypothesized that either their architecture represents the ancestral form before evolutionary divergence, or that only one among the various tailed phage types was able to colonize eukaryotic cells. The spectacular advances in structural biology will allow future studies to provide more valuable information that could answer some of the questions remaining.

Further Reading

- Aksyuk, A.A., Rossmann, M.G., 2011. Bacteriophage assembly. *Viruses* 3 (3), 172–203.
- Chen, Z., Sun, L., Zhang, Z., *et al.*, 2017. Cryo-EM structure of the bacteriophage T4 isometric head at 3.3-Å resolution and its relevance to the assembly of icosahedral viruses. *Proceedings of the National Academy of Sciences of the United States of America* 114 (39), E8184–E8193.
- Cuervo, A., Daudén, M.I., Carrascosa, J.L., 2013. Nucleic acid packaging in viruses. In: Mateu, M. (Ed.), *Structure and Physics of Viruses (Subcellular Biochemistry)* 68. Dordrecht: Springer, pp. 361–394.
- Cuervo, A., Fàbrega-Ferrer, M., Machón, C., *et al.*, 2019. Structures of T7 bacteriophage portal and tail suggest a viral DNA retention and ejection mechanism. *Nature Communications* 10 (1), 3746.
- Dai, X., Zhou, Z.H., 2018. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. *Science* 360 (6384), eaao7298.
- Gong, D., Dai, X., Jih, J., *et al.*, 2019. DNA-packing portal and capsid-associated tegument complexes in the tumor herpesvirus KSHV. *Cell* 178 (6), 1329–1343.

- Liu, Y.T., Jih, J., Dai, X., Bi, G.Q., Zhou, Z.H., 2019. Cryo-EM structures of herpes simplex virus type 1 portal vertex and packaged genome. *Nature* 570 (7760), 257–261.
- Machón, C., Fàbrega-Ferrer, M., Zhou, D., *et al.*, 2019. Atomic structure of the Epstein-Barr virus portal. *Nature Communications* 10 (1), 3891.
- Mettenleiter, T.C., Klupp, B.G., Granzow, H., 2009. Herpesvirus assembly: An update. *Virus Research* 143 (2), 222–234.
- Rixon, F.J., Schmid, M.F., 2014. Structural similarities in DNA packaging and delivery apparatuses in herpesvirus and dsDNA bacteriophages. *Current Opinion in Virology* 5, 105–110.
- Wang, J., Yuan, S., Zhu, D., *et al.*, 2018. Structure of the herpes simplex virus type 2C-capsid with capsid-vertex-specific component. *Nature Communications* 9 (1), 3668.
- Yu, X., Jih, J., Jiang, J., Zhou, Z.H., 2017. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. *Science* 356 (6345), eaam6892.

Adenoviruses (*Adenoviridae*) and Their Structural Relatives

Gabriela N Condezo, National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Natalia Martín-González, Autonomous University of Madrid, Madrid, Spain

Marta Pérez-Illana, Mercedes Hernando-Pérez, José Gallardo, and Carmen San Martín, National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AU Icosahedral asymmetric unit

CAR Coxsackievirus and adenovirus receptor

Cryo-EM Cryo-electron microscopy

DJR Double jelly roll

dsDNA Double stranded DNA

EMD Electron microscopy databank

GON Group of nine

GOS Group of six

HAdV Human adenovirus

ICTV International Committee on Taxonomy of Viruses

INT Integrase

kbp Kilo base pairs

MCP Major capsid protein

MDa MegaDalton

PBCV-1 *Paramecium bursaria* chlorella virus 1

PDB Protein databank

pI Isoelectric point

POL Polymerase

Glossary

Cryo-electron microscopy Technique allowing to image macromolecular structures in frozen-hydrated conditions with a transmission electron microscope. Images obtained can be combined via image processing and reconstruction algorithms to produce a three-dimensional map of the object.

Double jelly roll Protein fold formed by two β -barrels. In adenoviruses and structurally related viruses, the β -barrels are perpendicular to the capsid surface. Double jelly roll capsid proteins form pseudo-hexagonal capsomers, allowing trimeric proteins to fill in the six-fold coordinated positions of the icosahedral capsid.

Four-helix bundle Protein fold consisting of four α -helices coiled together with a hydrophobic core in the center.

Icosahedral asymmetric unit The minimal part of an icosahedral object which contains enough information to

build the complete object by repetition following the icosahedral symmetry transformations. There are 60 asymmetric units per icosahedron.

Polintons Eukaryotic dsDNA transposons encoding a protein-primed DNA polymerase (POL) and a retroviral-like integrase (INT).

Structural lineage A group of viruses which share common structural elements and assembly principles.

β -barrel fold Protein fold consisting of eight antiparallel β -strands organized in two sheets that form the opposite sides of the barrel.

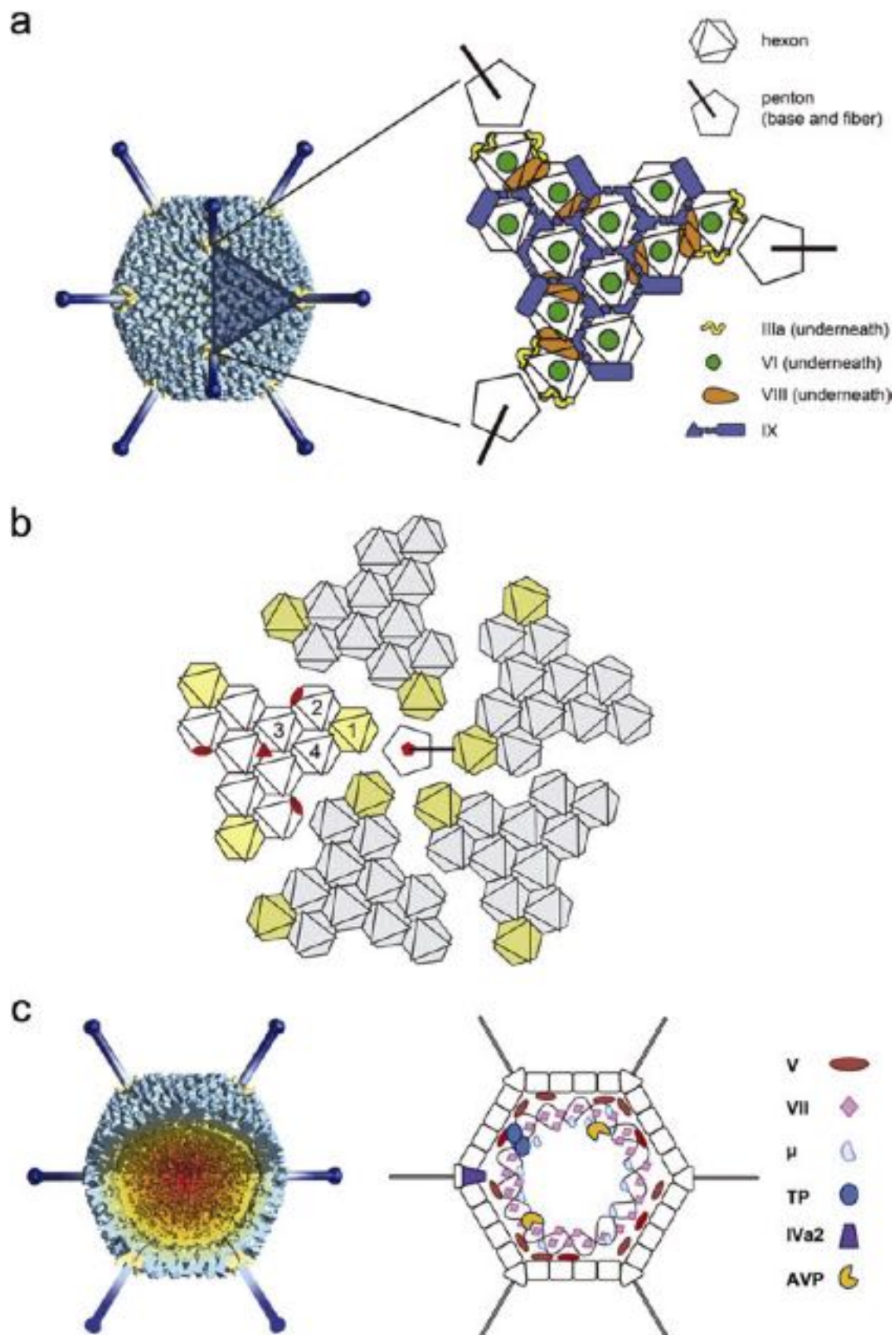
β -helix Protein fold consisting of parallel β -strands in a solenoid pattern, usually with three faces.

β -spiral Protein fold consisting of repetitions of two anti-parallel β -strands connected by a β -turn.

General Features of Adenoviruses

Adenoviruses were discovered in 1953 in human adenoid tissue, and since then have been found in all types of vertebrates, from fish to humans. They are currently grouped in five genera. Mastadenoviruses and aviadenoviruses infect mammals and birds, respectively. So far only a single member of the Ichtadenovirus genus has been reported, isolated from white sturgeon. The other two genera have broader host ranges. Atadenoviruses have been found in reptiles, ruminants, and birds, and siadenoviruses in amphibians, birds, and reptiles. More than 80 different human adenovirus types have been reported, causing mainly mild respiratory, gastrointestinal, and ocular diseases. Adenoviruses are also widely used as experimental vectors for gene transfer, vaccination, and oncolysis.

Adenoviruses have a linear, dsDNA genome packed within a non-enveloped icosahedral capsid of 95 nm diameter and triangulation number $pseudoT = 25$. The virion has a mass of 150 MDa and presents a complex protein composition (**Fig. 1(a)**). The major coat protein, hexon, forms most of the icosahedral shell. Penton capsomers (formed by penton base and fiber) sit at the vertices. At least three different minor coat proteins (IIIa, VI, VIII) are required to complete the icosahedral shell, and more are present in some genera. To describe the complex adenovirus structure, it is useful to consider some geometrical landmarks. Each capsid facet is formed by 12 hexon trimers. Four hexon trimers, plus one penton base monomer, form the icosahedral asymmetric unit (AU). Nine hexon trimers form the central plate of each facet or GON ("Group of Nine"). The group formed by five peripentonal hexon trimers, together with the penton base, has been designed as GOS ("Group of Six") (**Fig. 1(b)**). Inside the capsid, the adenovirus genome is bound to a large amount of positively charged, virus encoded proteins, forming the so-called core, which does not follow icosahedral symmetry (**Fig. 1(c)**). Genome length varies between adenovirus genera, ranging from 26 to 45 kbp.



Structure of Human Adenovirus Type 5

For historical reasons and because of its extended use in the viral vector field, human adenovirus type 5 (HAdV-C5) is the best characterized member of the adenovirus family. The structure of the HAdV-C5 virion was solved by cryo-electron microscopy (cryo-EM) at 3.5 Å resolution in 2010, in an impressive achievement given the size and complexity of the viral particle. The structure has been improved to 3.2 Å resolution in 2017, showing clearer details for some of the minor proteins.

Hexon

Hexon (also called polypeptide II in early literature) is the main building block in the icosahedral shell. There are 720 hexon monomers (forming 240 trimers) in the icosahedral particle. The 952 residue-long monomer contains two eight-stranded β barrels V1 and V2 (also called jellyroll domains) which in the virion are oriented perpendicularly to the capsid surface. A small β -sheet Vc (connector domain) holds the two barrels apart (**Fig. 2(a)**). The jelly roll domains form the hexon trimer pseudo-hexagonal base, which occupies the 6-fold coordinated positions in the icosahedral shell. Long loops connecting the β -strands form the hexon towers, which appear as triangular-shaped protrusions on the capsid surface. Some of the loops correspond to the so-called hypervariable regions, involved in defining the specific serological response to the virus. In the trimer, the three hexon monomers are heavily intertwined, resulting in a highly stable molecule. The amino and carboxy terminal regions adopt different conformations in each monomer, facilitating the hexon interactions with its neighboring proteins in the different capsid environments.

Penton Base

The penton base protein, or polypeptide III, occupies the 5-fold coordinated positions in the icosahedral network. There are 12 penton base pentamers in the capsid, one in each vertex.

The 571 amino acid-long penton base monomer folds into two domains: a single jellyroll perpendicular to the capsid surface, completing the spaces left by hexons in the shell, and an upper insertion facing the solvent exposed exterior (**Fig. 2(b)**). This upper domain contains an Arg-Gly-Asp sequence motif, needed to interact with cell integrins for viral internalization during infection. This motif forms part of a long (80 residues), flexible loop which could not be resolved in the structure. For each penton base monomer, an N-terminal arm (residues 37–51) extends away from the main body of the protein towards the viral core, interacting with two monomers of the inner coat protein IIIa (see below), and therefore contributing to anchor the penton within the surrounding hexons in the GOS. Regulation of penton stability is crucial during the infectious cycle. Pentons are the first elements to dissociate from the capsid under stress, and it is thought that their release at the initial stages of uncoating opens the way for the adenovirus membrane lytic peptide to exit the capsid, and disrupt the endosomal membrane. In this way, the virus particle can travel along the microtubular network and reach the nuclear pore, where it will release its genome for transcription and replication.

Fiber

Adenovirus fiber, or polypeptide IV, is responsible for host receptor recognition. One trimer of the 582 residue long fiber is bound to each pentamer of penton base at the vertices. Because of this symmetry mismatch, and because fibers are flexible, they have not been clearly visualized in cryo-EM structures of the complete virion, which use icosahedral averaging. The available fiber structures have been obtained from the analysis of recombinant protein fragments. Each fiber contains an amino-terminal penton base attachment domain; a long, thin central shaft; and a carboxy-terminal head (or knob) domain (**Fig. 2(c)**). A crystal structure of recombinant penton base bound to the N-terminal fiber peptide showed a 10-residue long fiber fragment bound at the cleft formed by two adjacent penton monomers. In the crystal, all five clefts were occupied by the fiber peptide. In the virion, there are only three fiber monomers per vertex. It is not known what mechanism decides which three of the five available binding clefts in penton base are used to bind fiber.

Beyond the N-terminal tails, the shaft extends away from the capsid surface and ends at the head domain. The sequence of the shaft domain consists of several repetitions of a 15-residue sequence motif. Each one of the repeats forms two anti-parallel

Fig. 1 Overall adenovirus structure and components, as exemplified by current knowledge on human adenovirus structure. (a) Icosahedral shell organization. The left hand side panel is a model built from a low resolution cryoEM map, with penton bases highlighted in yellow, and fibers built from the crystal structure of the knob and distal shaft (PDB ID 1QIU) in dark blue. The shaded triangle indicates one facet. The right hand side is a cartoon representing the location of major and minor coat proteins forming the icosahedral shell, as indicated by the adjacent legend. Each hexon trimer is depicted as a hexagon with an overlaid triangle indicating the position of the towers. (b) Geometrical landmarks in the adenovirus icosahedral net. One penton and the five adjacent facets are represented, spread out for clarity. The four hexon trimers forming the asymmetric unit (AU) are numbered in one facet. Peripentonal hexons, forming the Group of Six (GOS) together with the penton, are highlighted in yellow. Hexons in the Groups of Nine (GON) are depicted in white for the reference facet, gray in the rest. Red symbols indicate the icosahedral symmetry axes. (c) Non-icosahedral components. A segment has been removed from the cryoEM map to show the inner capsid contents. The schematics on the right hand side indicate tentative positions, as little is known about the structure and organization of the genome and accompanying proteins in the virion. A small fragment of protein VII has recently been identified in the icosahedral shell (see main text, and **Fig. 3**). Figure modified from San Martín, C., 2012. Latest insights on adenovirus structure and assembly. *Viruses* 4, 847–877.

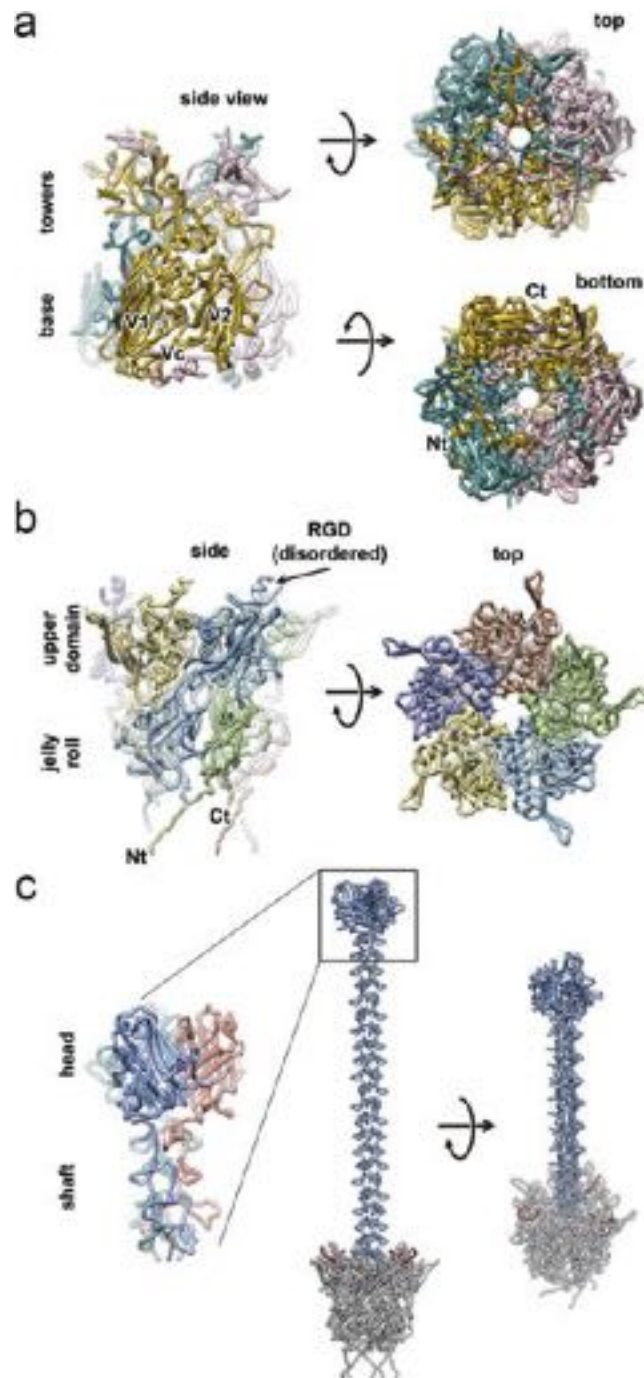


Fig. 2 High resolution structures of HAdV-C5 capsid proteins. (a) The hexon trimer is depicted in a ribbon representation, with each monomer in a different color, in side, top (as seen from outside the capsid) and bottom (from inside) views. The two vertical jelly rolls in one monomer are labeled (V1 and V2). A β -sheet that separates them is labeled Vc (for Viral Connector domain). The amino (Nt) and carboxy (Ct) termini of one monomer are indicated. (b) The penton base pentamer depicted in side and top views. Each monomer is rendered in a different color. The position of the disordered loop containing the integrin binding Arg-Gly-Asp sequence motif, as well as the amino (Nt) and carboxy (Ct) termini of one monomer, are indicated. (c) Structure of the vertex complex, composed by the pentameric penton base and the trimeric fiber. The fiber knob and distal shaft structures solved by protein crystallography are shown in the left hand side zoom, with each monomer in a different color. A simple model of the fiber shaft has been generated by drawing 20 shifted copies of one shaft repeat. Three N-terminal fiber peptides (red) are shown, bound to the penton base pentamer (gray). Center: side view; right, oblique view. The protein databank identifiers for the structures depicted here are: 1QIU for the fiber and shaft, 3IZO for the N-terminal fiber peptides, 6B1T for hexon and penton.

β -strands connected by a β -turn, in a fold known as β -spiral. The β -strands are oriented along the fiber axis, and each repeat represents a rise of 13 Å in the fiber length. The HAdV-C5 fiber shaft contains 22 such repeats, resulting in a 330 Å long fiber (including the ~45 Å tall head). A breakdown of the consensus sequence in the third repeat causes a kink in the shaft. The average diameter of the shaft, excluding the surface loops, is 15 Å. At the C-terminal end of the shaft we find the knob domain, responsible for attachment to the adenovirus receptor at the host cell membrane. The head domain in each fiber monomer folds as an anti-parallel β -sandwich formed by two four-stranded sheets. Loops connecting the β -strands are exposed on the surface of the trimeric fiber head, available to bind the cell receptor.

Minor Coat Proteins

Apart from hexons and pentons, the HAdV-C5 virion contains four minor coat proteins: IIIa, VI, and VIII on the inner capsid surface (Fig. 3(a)), and protein IX on the outer one (Fig. 3(b)).

Protein IIIa

Five copies of protein IIIa (585 amino acids) are located beneath each vertex (Fig. 3(a)). The structure of a region comprising residues 7–300 has been solved, while the rest of the polypeptide chain (residues 300–585) is not icosahedrally ordered, and thus remains unresolved. The IIIa fold is predominantly α -helical. The N-terminal domain, or GOS-glue domain, helps stabilize the GOS by bridging every penton to its surrounding hexons. This domain is linked by a long helix to another region (“VIII-binding domain”) which interacts with another minor coat protein, polypeptide VIII, joining the GOS to the hexons in the central plate of the facet (GON). The unresolved C-terminal region may be interacting with the non-icosahedral core. Apart from stabilizing the vertices, protein IIIa contributes to the specificity of genome packaging via interactions with one of the adenovirus packaging proteins, L1 52/55 kDa.

Protein VI

Polypeptide VI is synthesized as a 250 residue long precursor (pVI) and cleaved at two sites by the adenovirus protease during maturation: between residues 33 and 34 at the N-terminus, and 239–240 at the C-terminus. There are approximately 360 copies of protein VI in the virion. This protein binds to hexon and to dsDNA, bridging capsid and core. Only two fragments of pVI have been solved in cryo-EM studies of the virion (Fig. 3(a), inset). The cleaved N-terminal peptide (residues 5–33) lines the wall of the hexon trimer cavity that opens towards the core. Another segment, comprising residues 109–143, crosses the opening of the cavity. Only three copies of the N-terminal peptide and one of the central fragment have been traced in the AU, indicating that most of the protein is not ordered. The hexon:VI stoichiometry is also uncertain. There are too many copies of VI (360) to have one per hexon trimer (240), but not enough to have one per hexon monomer (720). Protein VI carries out multiple functions during the adenovirus infectious cycle. It acts as a shuttle to transport newly synthesized hexons to the nucleus for assembly; its C-terminal peptide is a cofactor required to activate the maturation protease; and it contains the adenovirus lytic peptide, an amphipathic helix (residues 34–54) that alters the curvature of the endosome membrane, so that the virus can escape into the cytosol during entry. After internalization, protein VI needs to be released from its internal position in the capsid for the lytic activity to take place.

Protein VIII

The precursor form of protein VIII (pVIII, 227 amino acids) is cleaved by the viral protease during maturation at residues 112, 131, and 157. The larger fragments (residues 2–112 and 157–227) stay together and ordered in the virion. The 45 central residues have not been identified in the cryo-EM HAdV-C5 studies, indicating that they are either removed upon proteolytic processing, or disordered. There are two independent monomers of polypeptide VIII in the AU. One of them is wedged between protein IIIa and the hexons at the periphery of the GOS, cooperating with IIIa to bind each GOS to its five surrounding GONs. The second copy is located around the icosahedral 3-fold symmetry axis, contributing with polypeptide IX (see below) to the stabilization of each GON (Fig. 3(a)). Each copy of VIII interacts with four hexon trimers. Interestingly, some of these interactions occur by β -sheet augmentation. That is, the interaction is mediated by a β -strand from polypeptide VIII binding to the edge of a β -sheet in hexon.

Protein IX

Polypeptide IX is the only minor coat protein located on the outer surface of the HAdV-C5 capsid (Fig. 3(b)). This 140 amino acid long polypeptide has a remarkable extended conformation, reaching 185 Å in length. The amino- and carboxy-terminal domains are connected by a central “rope domain” with little secondary structure. The amino-terminal domains of three IX monomers join via hydrophobic interactions at the icosahedral and local 3-fold axes in the GONs, forming a triskelion-shaped structure (four triskelions per facet). The rope domain of each monomer forming the triskelion runs in a different direction towards the facet edges. There, the C-terminal α -helices of four IX monomers join to create a 4-helix bundle. Three of the helices in the bundle are parallel and belong to IX monomers forming different triskelions in the same facet. The fourth is antiparallel, coming from a copy of IX in the adjacent facet (Figs. 3(b) and 5(b)). The 240 copies of protein IX form a sort of hairnet on the virion surface, keeping the hexons in each GON together and binding GONs to neighboring GONs across the icosahedral edges. Protein IX is dispensable for adenovirus assembly, but IX-deletion mutants assemble viral particles with low thermostability. The triskelion N-terminal domain of IX is enough to confer capsid thermostability. Protein IX has also been proposed to participate in the final stages of uncoating during adenovirus entry. After

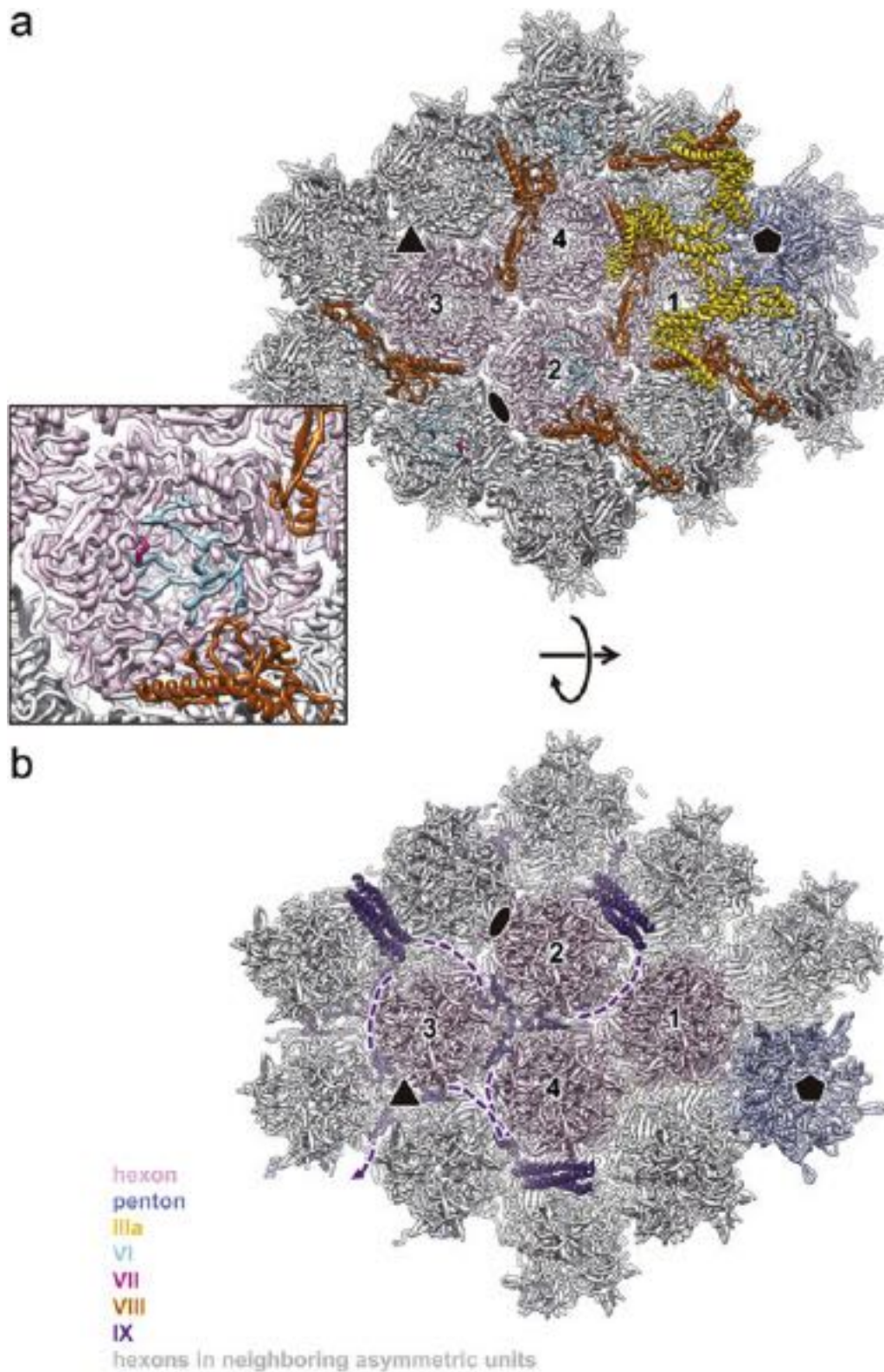


Fig. 3 Structure of the complete HAdV-C5 icosahedral asymmetric unit. The proteins in one AU and their closest neighbors are depicted as ribbons as seen from inside (a) and outside (b) the virion. Proteins are colored according to the legend at the bottom left. The four hexon trimers in the AU are numbered 1–4 as in Fig. 1(b). A pentagon, a triangle and an oval indicate the positions of the icosahedral 5-fold, 3-fold and 2-fold axes respectively. The inset in (a) shows a zoom in the hexon 2 region, where fragments of protein VI and VII are traced. Dashed purple lines in (b) indicate the path followed by disordered regions of protein IX. The dashed arrow indicates where a monomer of IX would continue towards the other side of the icosahedral facet. The structure depicted corresponds to protein databank entry 6B1T.

escaping from the endosome, adenovirus particles travel along the microtubule network towards the nuclear pore, to release the viral genome into the nucleus. Capsids are anchored to the nuclear pore complex via hexon, while simultaneously being pulled away by microtubule motor kinesin-1 which binds to protein IX, resulting in capsid dismantling. However, mutants lacking IX are still infectious, so alternative mechanisms must exist to complete capsid disassembly.

Core Proteins

The HAdV-C5 genome is organized within the capsid in a nucleoprotein core composed by a 35 kbp long, linear dsDNA molecule, bound to over 20 MDa of viral proteins (Fig. 1(c)). The copy number of the proteins present in the core is not well defined, as different estimations have been reported. Core proteins V (366 amino acids, 150 copies), VII (198 amino acids, 500–800 copies), and μ (also called X, 80 amino acids, 100–300 copies) have a large amount of positive charges and are thought to condense the viral genome by compensating for the electrostatic repulsion between the negative charges of the dsDNA molecule.

Protein V has a theoretical isoelectric point (pI) of 10.3. It interacts with penton base and minor coat protein VI, and part of it is released in the endosome during the first stages of uncoating. It seems therefore that protein V is located at the outermost layers of the core, bridging it to the capsid. *Protein VII* (pI 12.3) is the mature form of polypeptide pVII, resulting from cleavage by the maturation protease at residues 13 and 24. Polypeptide VII participates in virus assembly by interacting with packaging proteins IVa2 and L1 52/55 kDa. However, particles without protein VII can be assembled, although they are not infectious. One of the pVII segments cleaved by the protease (residues 14–24) occupies a position equivalent to that of the cleaved N-terminal peptide of minor coat protein VI, lining the hexon cavity wall (Fig. 3(a), inset). *Protein μ* (pI 12.9) is also synthesized as a precursor form, cleaved by the maturation protease at residues 32 and 51. It has been suggested that μ helps condense DNA in the virus core acting as a bridge between DNA strands.

Because the core does not follow icosahedral symmetry, its architecture has not been elucidated in high resolution cryo-EM studies, and structural data for the individual proteins are scarce. When the nucleoprotein complex is extracted from the virion, objects consistent with a “beads-on-a-string” organization are observed, supporting the idea that core proteins condense the genome by a strategy similar to that of histones. Protein VII, the most abundant core protein, remains bound to the genome in the beaded structure even after harsh chemical treatment. It is therefore thought that protein VII wraps the dsDNA to form the 9.5 nm diameter nucleosome-like beads. Cryo-electron tomography analyses allow observing the core in individual virus particles without imposing icosahedral symmetry, but so far the resolution they provide has been limited. A physical model for the core organization proposed that the organization of DNA and associated proteins is ruled by a soft electrostatic repulsive potential. This model is consistent with the fact that the positive charges of all the core proteins are not enough to compensate for all the negative charges provided by the dsDNA.

Also bound to the HAdV-C5 genome in the core are the terminal protein and the maturation protease. The terminal protein is involved in initiation of genome replication, and one copy remains bound to the 5' termini of the dsDNA molecule in the infectious particle. The maturation protease enters the particle bound to the genome in a largely inactive form. After the protease cleaves pVI, the carboxy-terminal released peptide (pVIC) remains covalently bound to the enzyme and enables it to slide on the dsDNA molecule to reach all its targets. These include: minor coat proteins IIIa, VI and VIII, core proteins VII and μ , the terminal protein, and the packaging protein L1 52–55 kDa. Proteolytic maturation makes the HAdV-C5 particle metastable, priming it to start the programmed uncoating cascade, release the lytic peptide in the endosome, and the genome at the nuclear pore. Particles deficient in maturation are not infectious, remain trapped in the endosome and are destroyed further along the endocytic pathway.

Variants: Other Adenoviruses

Structural studies on the complete virion are available for only two of the five adenovirus genera: mastadenoviruses, and atadenoviruses. Mastadenoviruses infect mammals and include the human adenoviruses (HAdV), classified into seven species (A to G). High resolution cryo-EM structures are available for HAdV-C5 (3.2 Å resolution) and HAdV-D26 (3.7 Å). Structural characterization of non-human mastadenoviruses is limited to bovine (BAdV-3) incomplete capsids at high resolution (4.5 Å) and medium resolution structures of bat (BtAdV-250A) (13.7 Å) and canine (CAdV-2) (12 Å) adenovirus. As for atadenoviruses, two medium resolution structures are available: ovine adenovirus OAdV-7 (10.6 Å) and snake adenovirus SnAdV-1 (10.9 Å). The latter is the only structure for an adenovirus infecting non-mammalian hosts analyzed so far. Overall, structural studies demonstrate that all adenoviruses share a common general organization of the virion, yet certain differences are noteworthy as they might have relevance in features such as structural stability or tropism. The main differences observed concern the structure and number of fibers, and the minor coat proteins on the outer capsid surface.

Fiber Shafts and Heads

The main tropism determinant in adenovirus is the fiber-penton base complex. A considerable number of fiber head structures are available, obtained from recombinant proteins. All adenovirus fibers described so far are trimeric, with an N-terminal tail, a central shaft made of repeating sequences, and a C-terminal globular knob domain as in HAdV-C5. Although the main structural features are conserved, fiber sequences vary considerably between adenovirus types, reflecting the ability to bind different receptors for reaching different target tissues.

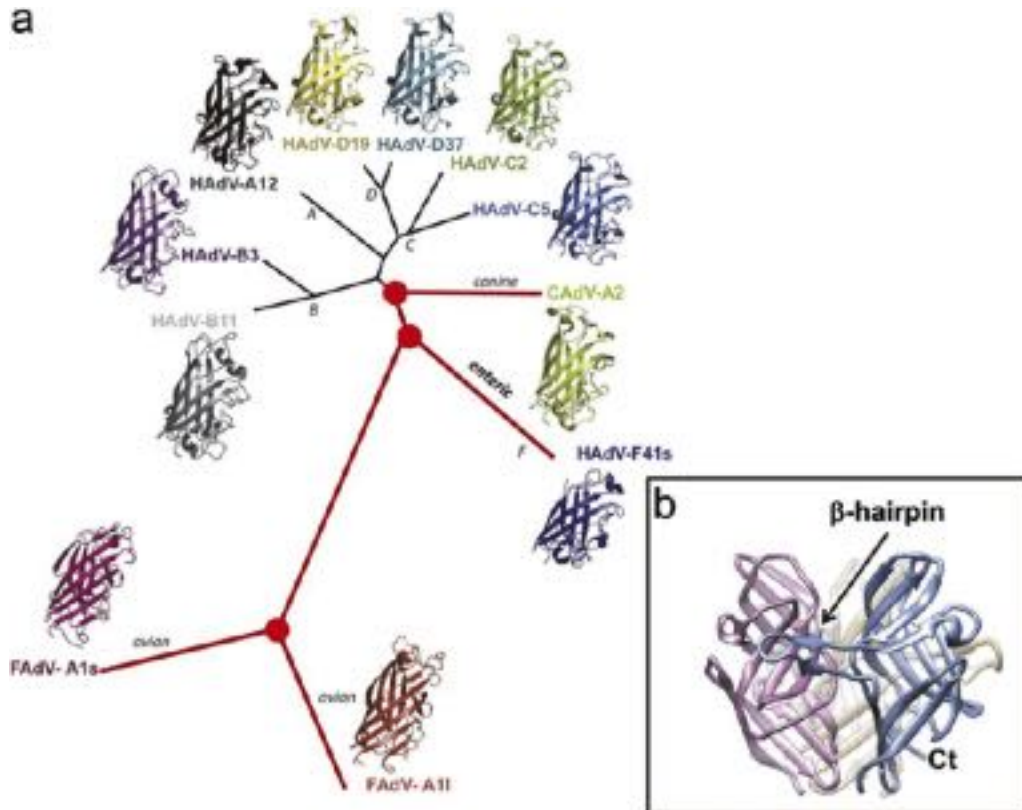


Fig. 4 Variations in the structure of adenovirus fiber heads. (a) Phylogenetic tree of adenovirus fiber head domains based on structural comparison. A total of 11 fiber head structures are compared, with protein data bank identifiers as follows. Human adenoviruses: HAdV-C2 (1QHV); HAdV-C5 (1KNB); HAdV-A12 (1KAC); HAdV-B11 (2O39); HAdV-B3 (1H7Z); HAdV-D37 (1UXE); HAdV-D19p (1UXB); HAdV-F41s short fiber head (2BZV). Non-human adenoviruses: canine adenovirus CAAdV-A2 (2J2J); fowl adenovirus FAdV-A1l long fiber head (2IUM); and FAdV-A1s short fiber head (2VTW). Capital letters by the tree branches refer to the human adenovirus species A-F. Red lines (branches) and circles (nodes) highlight the relationship between canine, enteric and avian adenovirus fiber heads. (b) Structure of a siadenovirus fiber head, highlighting the unique β -hairpin that embraces the neighboring subunit in the trimer. The structure depicted corresponds to the turkey adenovirus type 3 (also known as turkey hemorrhagic enteritis virus, THEV) protein (PDB: 3ZPF). The carboxy (Ct) terminus of one monomer is indicated. (a) Figure modified from El Bakkouri, M., Seiradake, E., Cusack, S., Ruigrok, R.W., Schoehn, G., 2008. Structure of the C-terminal head domain of the fowl adenovirus type 1 short fibre. *Virology* 378 (1), 169–176, with permission from Elsevier.

In human adenoviruses, the first 45 residues of the fiber protein correspond to the penton binding tail, and are highly conserved (typically over 50% identity with the HAdV-C5 prototype). The rest of the protein (central shaft and knob domain), however, differs in key structural and chemical features. The number of repeats in the shaft determines the fiber length, and disruptions in the repeat sequence determine the location of kinks in the otherwise rigid fibers. Fiber lengths in human adenoviruses range from 130 Å (300–350 amino acids, 5–7 repeats) in species B, to more than 300 Å (500–600 amino acids, over 20 repeats) in species C and A. Fibers in human adenovirus species B and D lack the 2- to 4- residue insertion that disrupts the repeating sequences of the shaft, resulting in rigid fibers. Notably, the fiber of bovine BAdV-3 is extremely long (976 amino acids). Fiber head surface charge plays a role in the selectivity of the different receptors in the cell membrane. For example, members of species C tend to have negatively charged fiber heads ($pI \sim 5$) and use the membrane protein CAR as a receptor, while most members of species D have positively charged fiber heads ($pI \sim 9$) and bind sialic acid. Most adenovirus fiber heads engage their receptors with lateral interactions, usually involving two monomers of the head per receptor, and binding 3 receptor molecules per fiber trimer. However, frontal interactions have recently been reported for HAdV-B3, with asymmetric interactions (either one or two receptors per fiber trimer).

Apart from shaft length, fibers may also vary in numbers. Human adenoviruses belonging to species F (HAdV-F40 and HAdV-F41), that show enteric tropism, have two fiber genes of different length, but only incorporate one fiber per vertex, either a long or a short one. All known aviadenoviruses have two fibers per vertex. Fowl adenovirus type A (FAdV-A) has two fiber genes of different length, whereas two fiber genes of equal length are found in FAdV-C. FAdV-B, FAdV-D, and FAdV-E contain just one fiber gene. Structure and sequence alignments suggest that, among human adenoviruses, HAdV-F41 is the closest to the animal viruses. In fact, in phylogenetic trees calculated from such alignments, the HAdV-F41 short fiber head appears closer to the avian fiber heads than the canine adenovirus fiber head (Fig. 4(a)). In the Atadenovirus genus, ovine (OAdV-7) and snake (SnAdV-1) adenoviruses have only one fiber gene and present only one fiber per vertex. However, in a lizard atadenovirus (LAdV-2), two fiber genes were found. Interestingly, the two different fiber proteins assemble as either one short or three long fiber projections per

vertex (present in one or two vertices per virion). Overall, fiber heads in atadenoviruses are smaller (107 amino acids in SnAdV-1; 123 and 120 amino acids in LAdV-2) than in mastadenoviruses (183 amino acids for HAdV-C5) and aviadenoviruses (248 and 209 amino acids for the two FAdV-A1 fibers). Finally, only one fiber gene has been found in the siadenoviruses studied so far. Siadenovirus fiber heads have two distinctive features. First, their closest structural homologs are reovirus fiber heads, rather than those of other adenoviruses. Second, an insertion in one of the β -strands creates a β -hairpin that embraces the neighboring monomer, a unique feature among all adenovirus fiber heads characterized so far (Fig. 4(b)).

External Minor Coat Proteins

Protein IX, the only minor coat protein located on the outer surface of the HAdV-C5 capsid, is unique to mastadenoviruses. Even within the mastadenoviruses, polypeptide IX considerably varies between the different virus types. In particular, the central region defined in the cryo-EM structure of HAdV-C5 as the rope domain is conserved only among human and simian adenoviruses (Fig. 5(a)). Canine (CAdV-1), bovine (BAdV-3) and bat (BtAdV-250A) adenoviruses have a shorter IX protein lacking this rope domain. This results in the carboxy-terminal helical bundle of IX being located directly on top of the N-terminal triskelions, instead of at a distance at the facet edges as in HAdV-C5. In human adenoviruses the bundle is formed by four helices, three of them belonging to monomers of protein IX forming three different triskelions in one icosahedral facet, and the fourth one coming from a copy of protein IX in the adjacent facet, assembling in an antiparallel fashion to the other three. In non-human adenoviruses the organization is simpler, and the IX monomers do not reach across different AUs or facets: the bundle is formed by just the three helices corresponding to the IX molecules forming the triskelion located directly below (Fig. 5(b)).

In atadenovirus capsids, a genus specific protein called LH3 has been found at the icosahedral and local three-fold axes, the same location occupied by IX in mastadenoviruses. LH3 (~380 amino acids) is twice the size of protein IX (140 residues) and forms prominent knobs on the surface of the capsid, interacting with the towers of three surrounding hexon trimers (Fig. 6). Remarkably, protein LH3 has a trimeric β -helix fold typical of bacteriophage tailspikes, very different from the helix clusters observed in protein IX. The extensive interaction between LH3 and hexons may have a role in the higher capsid stability of atadenoviruses, that can withstand temperatures at which HAdV-C5 particles are completely disrupted (over 48°C). Because bacteriophage tailspikes are responsible for virus attachment to the host, and LH3 configures the outermost surface of the atadenovirus shell, it has been hypothesized that, apart from a cementing function, LH3 may have some role in interaction with the host.

Adenovirus Structural Relatives

During the last two decades, structural biology analyses have shown that viruses infecting hosts far apart in evolution share similar architectural features, prompting a new virus classification based on *structural lineages*. Four icosahedral virus structural lineages are recognized at present. Viruses included in one of these lineages share their major coat protein (MCP) fold with adenovirus.

The Double Jelly Roll Lineage

For a long time, adenovirus was the only known virus where the six-fold coordinated positions in the capsid were occupied by trimers of a major coat protein folding as a double jelly roll perpendicular to the capsid surface. In 1999, the structure of the MCP of a tailless bacteriophage called PRD1 (a tectivirus) was solved, surprisingly showing the same kind of fold and trimeric oligomerization. This finding demonstrated that viruses with none or little sequence homology, infecting very different hosts (human and bacteria), use the same kind of architectural solution to build their capsids, and inspired the idea of classifying viruses on the basis of their MCP structure.

The structural lineage including adenovirus is variously referred to as the PRD1-adenovirus lineage (for its first two described members), the PRD1-like lineage, or the double jelly roll (DJR) lineage. Since the discovery of the same fold in adenovirus and PRD1, dsDNA viruses belonging to 11 different families have been shown to share this MCP fold (*Corticoviridae*, *Iridoviridae*, *Lavidaviridae*, *Marseilleviridae*, *Mimiviridae*, *Phycodnaviridae*, *Poxviridae*, *Sphaerolipoviridae*, *Tectiviridae*, and *Turriviridae*), as well as other unclassified viruses (Table 1). Members of the DJR lineage infect organisms across the evolutionary tree: bacteria (tequiviruses, corticoviruses); archaea (turriviruses); unicellular animals (giant viruses like mimivirus and their relatives, also their virophages) and algae (phycodnaviruses); insects, fish, amphibians, and reptiles (iridoviruses); and vertebrates in general including humans (adenoviruses). The infectious particles of these viruses have a wide range of sizes and triangulation numbers (Table 1 and Fig. 7), but they are all built from trimeric DJR capsomers. It is notable that poxviruses, which do not have an icosahedral capsid, also encode a trimeric DJR protein called D13. D13 is not present in the mature poxvirus particle, but is required for assembly, and its structure suggests that poxviruses may have evolved from an icosahedral ancestor.

Variants of the Double Jelly Roll Fold: PRD1-Like Viruses With Authentic Hexameric Capsomers (Paired Orthogonal Single Jelly Rolls)

The double β -barrel fold is likely to have arisen through gene duplication. In line with this hypothesis is the recent identification of PRD1-like viruses that have two MCPs instead of one, for example bacteriophage P23-77 and several archaea viruses: *Haloarcula californiae* icosahedral virus 1 (HHIV-1), *Haloarcula hispanica* icosahedral virus 2 (HHIV-2), and its close relative virus SH1. In these

viruses the two MCPs (called VP4 and VP7 in the archaea viruses) are morphologically distinct. VP4 is composed of two vertical single jelly rolls, one standing on top of the other. One of the jelly rolls forms the capsomer base, while the other forms the tower. VP7 consists of only one vertical single jelly roll. These proteins share a high structural homology despite their low sequence identity (<20%) with the individual barrels of the double β -barrel MCP of *Pseudoalteromonas* phage PM2, which has the simplest DJR protein of the PRD1-adenovirus lineage (Fig. 7). VP4 and VP7 are ordered in heterodimers which produce two morphologically distinct capsomers, one with two and the other with three towers. The three-tower capsomers are composed of three copies of VP4 and three copies of VP7. The two-tower capsomers are composed of two copies of VP4 and four copies of VP7. The relative orientation of the individual β -barrels in the VP7-VP4 heterodimer anticipates that of the double β -barrel arrangement in PM2. On the basis of common structural and assembly principles, it has been proposed that single vertical β -barrel viruses should be grouped together with the double β -barrel viruses into a vertical β -barrel superlineage. Such a superlineage would include all viruses that use similar virion architecture regardless of whether the hexagonal capsomers are composed of six monomers or three dimers.

Vertex Structures in the Double Jelly Roll Lineage

Apart from the major coat protein which forms most of the capsid, members of the DJR lineage have also a single perpendicular jelly roll protein forming the pentameric vertex capsomers (pentons). The PRD1 penton is essentially a simplified version of the MCP, consisting of a single minimal jelly-roll. In adenovirus, the penton base protein possesses a PRD1-like lower core, adorned with an elaborate upper domain, rich in long loops (Fig. 2(b)). Penton proteins have been found in all PRD1-like viruses for which high-resolution structural information is available, including corticovirus PM2, archaeal turrevirus STIV (*Sulfolobus* turreted icosahedral virus), and virophage Sputnik. Thus far, an equivalent of the penton protein has not been identified among the Megavirales, except for phycodnaviruses (PBCV-1). However, considering the importance of this protein for the formation of icosahedral virions, it is expected that these giant viruses also encode a penton homolog.

Viruses of the DJR lineage tend to possess spikes protruding from the vertices, with a variety of sizes and shapes (Fig. 8). Similarly to the adenovirus fiber, these spikes have been shown to be involved in host recognition and attachment for some of the PRD1-like characterized viruses. As in adenovirus, spikes are often flexible or assemble via a symmetry mismatch on the penton, making their detailed structural characterization much more difficult. PRD1 has two different spikes (proteins P2 and P5) attached to a single pentamer of the vertex protein P31 (penton). Moreover, in PRD1 each spike has a different oligomerization state. P5 is a trimer resembling the structure of the adenovirus fiber, while P2 is a monomer with a pseudo- β propeller head. Sequence similarity and low resolution structural analyses suggest that the N-terminal domain of P5 could be embedded in the capsid by forming a heteropentameric base with P31. The role of each spike in PRD1 host recognition and attachment is not fully known.

It is notable that a trimeric spike attached to a pentameric base is also observed for the dsRNA reovirus attachment protein, $\sigma 1$. What is more, $\sigma 1$ is composed of a filamentous N-terminal tail and a globular C-terminal head folding as a compact β -barrel, with topology identical to the adenovirus fiber head. This observation points to a distant evolutionary relationship between the two proteins, nowadays coded by virus genomes with no other similarity.

Membranes and Membrane Proteins

Most members of the DJR lineage have an internal membrane, except adenovirus and the virophages Sputnik and Mavirus. The internal membrane forms a vesicle that contains the virus genome. Membrane proteins connect the lipid bilayer to the capsid, altering its curvature to follow the icosahedral contours of the shell. In the largest members of the lineage such as mimiviruses, multiple layers of proteins and membranes may exist. Internal membranes seem to be important for different aspects of the infectious cycle, such as genome delivery or assembly. In PRD1, the membrane remodels into a proteo-lipid tube that provides a conduit for passage of the ejected linear dsDNA through the cell envelope. A similar structure has been observed in PBCV-1, where a membrane tunnel is generated by fusion of the virus internal membrane with the host membrane. In Mimivirus, the

Fig. 5 Differences in external minor coat proteins between mastadenoviruses. (a) Alignment of representative polypeptide IX sequences for different species in the Mastadenovirus genus. The extent of the N-terminal (N), rope (r) and helical (H) domains from the HAdV-C5 structure are indicated with letters colored according to the ribbon depiction of the protein shown at the top. Notice that the rope domain is conserved only in human and simian viruses (sequences within a black rectangle). (b) Comparison of protein IX organization in human and non-human mastadenovirus capsids. Left: capsid organization based on the cryo-electron microscopy structures of human adenovirus HAdV-D26 (PDB: 5TX1) and bovine adenovirus BAdV-3 (PDB: 3ZIF). Penton base and fibers are shown in magenta and gold, while the 4 hexon trimers in the AU are colored blue, pink, green, and khaki, respectively. Hexon trimers in one of the asymmetric units are represented as hexagons and labeled 1–4 as in Fig. 1(b). One hexon from a neighboring AU is labeled 2'. A schematic of the icosahedron edges is shown as yellow rods. Right: magnified view of a single facet of the capsid with hexons and internal minor coat proteins hidden to visualize the two different protein IX organizations. The four structurally distinct IX molecules in each AU are colored blue, light blue, cyan, and purple, respectively. In HAdV-D26, only one of the four independent copies of IX (cyan) could be traced in its entirety. Disordered regions which could not be traced are depicted as red lines. In both viruses the triskelions can be seen at both the local and icosahedral 3-fold symmetry axes, identified by red and orange triangles, respectively. While the tetrameric helical bundles in HAdV-D26 are located at the facet edges (light gray filled ovals), the trimeric bundles in BAdV-3 are positioned directly on top of the triskelions. Panel (a) modified from San Martín, C., 2012. Latest insights on adenovirus structure and assembly. *Viruses* 4, 847–877. Panel (b) modified from Matteson, N.L., Barry, M.A., Reddy, V.S., 2018. Structure-based assessment of protein-protein interactions and accessibility of protein IX in adenoviruses with implications for antigen display. *Virology* 516, 102–107, with permission from Elsevier.

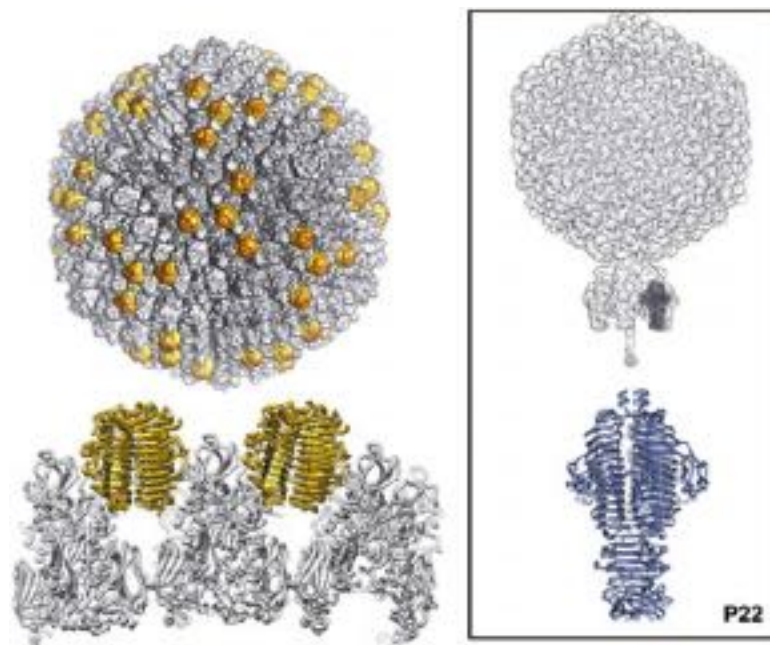


Fig. 6 External minor coat protein in atadenoviruses. Top left: the cryo-EM map of snake adenovirus type 1 is shown in gray, with the genus specific minor coat protein LH3 highlighted in gold. Top bottom: the crystal structure of LH3 (gold, PDB: 5G5N) and its position with respect to hexons (modeled by homology with the HAdV-C5 hexon as template, gray) is shown in a view across the particle. Bacteriophage P22 capsid (EMD-8005) with the tailspike protein in blue (PDB: 1TSP) is shown at the right hand side for comparison. The SnAdV-1 cryo-EM map identifier at the electron microscopy databank is EMD-3599.

internal membrane fuses with the host endocytic vacuole to release the viral genome to the cytoplasm. For the marine bacteriophage PM2, it has been proposed that membrane proteins may play a scaffolding role, recruiting capsid proteins to the lipid bilayer during assembly.

Genomes in the Double Jelly Roll Lineage

Most members of the DJR lineage have a dsDNA genome, either linear (PRD1, Mimivirus, adenovirus) or circular (PM2, STIV, Sputnik). The only known exception so far is the *Flavobacterium* phage FLiP, with a circular ssDNA genome. A characteristic of DJR viruses infecting eukaryotic hosts (adenovirus, PBCV-1, poxviruses, asfarviruses) is the encoding of positively-charged, histone-like proteins that are packaged together with the genome. These proteins condense the dsDNA and also play a role in protecting the virus genome from the host defense machinery from entry to replication. In the algae virus PBCV-1, the capsid remains outside the cell while the viral genome is injected. This genome has to reach the nucleus to start the first replication phase. It has been observed that the PBCV-1 genome is translocated as a condensed structure. This form would facilitate its trafficking in the dense cytoplasm milieu by facilitating the bypass of the cellular obstacles.

Asymmetric Features

Apart from the symmetry mismatches mentioned above between penton proteins and spikes, conspicuous features that do not follow icosahedral symmetry have been observed in some of the largest DJR viruses. PBCV-1 has a unique vertex, with a pocket on the inside and a spike structure on the outside of the capsid (Fig. 8). Comparison of shape, size, and location of the spike with similar features in tailed bacteriophages suggests that it might be a cell-puncturing device. In Mimivirus, a starfish-like structure (the stargate) is present at one of the vertices. The stargate defines the first part of the mimivirus capsid to be recruited on the membrane during assembly, and the exit conduit of the genome during uncoating. Inside the icosahedral shell, the Mimivirus core has a concave depression beneath the stargate. Non-icosahedral features have also been found in one of the smallest members of the DJR lineage. Bacteriophage PRD1 also has a unique vertex, where the viral DNA enters and exits the capsid. At the unique vertex, the packaging complex replaces the regular 5-fold structure and crosses the lipid bilayer. This complex consists of a packaging ATPase (P9), a packaging efficiency factor (P6), heterodimers of proteins P20 and P22, and the genome terminal protein P8 which functions as a valve, closing the channel once the genome is inside.

Table 1 Double jelly-roll virus families for which the major capsid protein and/or capsid structures have been studied

Example virus and family name	Host	Capsid diameter	Triangulation number	Genome type and approximate size	Accession number in structure databank
Prokaryotic host - bacteria					
PRD1, <i>Tectiviridae</i>	Gram-negative bacteria	70 nm	$T = 25$	linear dsDNA, 15 kbp	PDB 1W8X
PM2, <i>Corticoviridae</i>	<i>Pseudoalteromonas</i>	60 nm	$T = 21d$	circular dsDNA, 10 kbp	PDB 2W0C
<i>Salisaeta</i> Icosahedral phage 1 (SSIP – 1), <i>Sphaerolipoviridae?</i> , Unclassified	<i>Salisaeta sp</i>	100 nm (single jelly roll?)	$T = 49$	circular dsDNA, 44 kbp	EMD 2061
Flavobacterium-infecting, lipid-containing phage (FLiP), Unclassified	<i>Flavobacterium sp.</i>	55 nm	$T = 21d$	circular ssDNA, 9 kb	EMD 3771
Prokaryotic host – archaea					
STIV, <i>Turriviridae</i>	<i>Sulfolobus solfataricus</i>	96 nm (with turrets) 73 nm (without)	$T = 31d$	circular dsDNA, 18 kbp	PDB 3J31
HHIV – 2, <i>Sphaerolipoviridae</i> ^a	<i>Haloarcula hispanica</i>	80 nm (single jelly roll)	$T = 28d$	linear dsDNA, 30 kbp	EMD 0172
Eukaryotic host					
Paramecium bursaria chlorella virus 1 (PBCV – 1), <i>Phycodnaviridae</i>	<i>Chlorella variabilis</i>	190 nm	$T = 169d$	dsDNA with covalently closed hairpin termini, 330 kbp	PDB 1M4X
<i>Phaeocystis pouchetii</i> virus 1 (PpV01) <i>Phycodnaviridae</i> or <i>Mimiviridae</i> : under debate.	<i>Phaeocystis pouchetii</i> (phytoplankton)	220 nm	$T = 219$	485 kbp	–
Cafeteria roenbergensis virus (genus Cafeteriavirus, <i>Mimiviridae</i>)	<i>Cafeteria roenbergensis</i> (zooplankton)	300 nm	$T = 499$	730 kbp	EMD 8748
<i>Acanthamoeba polyphaga</i> Mimivirus (APMV), genus Mimivirus, <i>Mimiviridae</i>	<i>Acanthamoeba polyphaga</i>	500 nm	$T = 972\text{--}1200$	linear dsDNA, 1180 kbp	EMD 10623
Sputnik, <i>Lavidaviridae</i>	Amoebae/Mimivirus (virophage)	75 nm	$T = 27$	circular dsDNA, 18 kbp	EMD 5495
Melbournevirus, <i>Marseilleviridae</i>	<i>Acanthamoeba castellanii</i>	230 nm	$T = 309$	circular (?) dsDNA, 369 kbp	EMD 3868
Faustovirus, unclassified (distantly related to <i>Asfarviridae</i>)	<i>Vermamoeba vermiciformis</i>	260 nm	$T = 277$	circular dsDNA, 466 kbp	EMD 8144/8145
Pacmanvirus, unclassified (distantly related to Faustovirus and <i>Asfarviridae</i>)	<i>Acanthamoeba castellanii</i>	250 nm	$T = 309$	dsDNA, 395 kbp	–
Chilo iridescent virus (CIV), <i>Iridoviridae</i>	Invertebrates, amphibians, fish	185 nm	$T = 147$	linear dsDNA, 212 kbp	EMD 1580
African swine fever virus, <i>Asfarviridae</i>	Invertebrates, mammals	240nm	$T = 277$	linear dsDNA, 170–194kbp	EMD 0846 PDB 6KU9
Adenovirus, <i>Adenoviridae</i>	Vertebrates	95 nm	$T = 25$	linear dsDNA, 26–45 kbp	EMD 5172
Vaccinia virus, <i>Poxviridae</i>	Vertebrates	200–300 nm	Non-icosahedral	linear dsDNA, 130–375 kbp	PDB 3SAQ

^aSphaerolipoviruses have also been isolated from extremophile bacterial hosts.

Note: Modified from San Martín, C., van Raaij, M.J., 2018. The so far farthest reaches of the double jelly roll capsid protein fold. *Virology Journal* 15, 181.

Capsid Size Determination in the DJR Lineage

A large variety of capsid sizes can be built with DJR capsomers of essentially the same diameter (Fig. 7 and Table 1). To build a larger particle using building blocks (capsomers) of a similar size, viruses in the DJR lineage adopt larger triangulation numbers. That is, they build their icosahedrons with more capsomers per facet. An intriguing question is how the capsid size is determined for each particular virus, because in principle their capsomer size and shape would allow building particles of different geometries. Obviously, there is a relationship between the genome size and the internal capsid volume, since larger genomes require larger capsids to hold them. However, some of these viruses (e.g., PRD1) assemble an empty capsid into which the genome is later packaged. Therefore, at least in this case, the genome itself cannot direct the assembly of a correctly sized particle.

A possible mechanism to determine capsid size is the use of minor coat proteins acting as a “tape measure”. Similarly to adenovirus, other members of the DJR lineage have minor coat proteins on the inner and outer capsid surfaces, helping to assemble and stabilize the virion. One of these proteins, called P30, has been proposed to act as a tape measure in bacteriophage PRD1. With an extended conformation, P30 runs from the icosahedral fivefold region to the two-fold symmetry axes, following the inner surface of the icosahedron edges. This arrangement suggested that P30 could be the main player in determining particle size and, consequently, the T number. A tape measure protein has also been found in phage Bam35, a *pseudoT* = 25 virus that infects Gram-positive hosts (whereas PRD1 infects Gram-negative hosts).

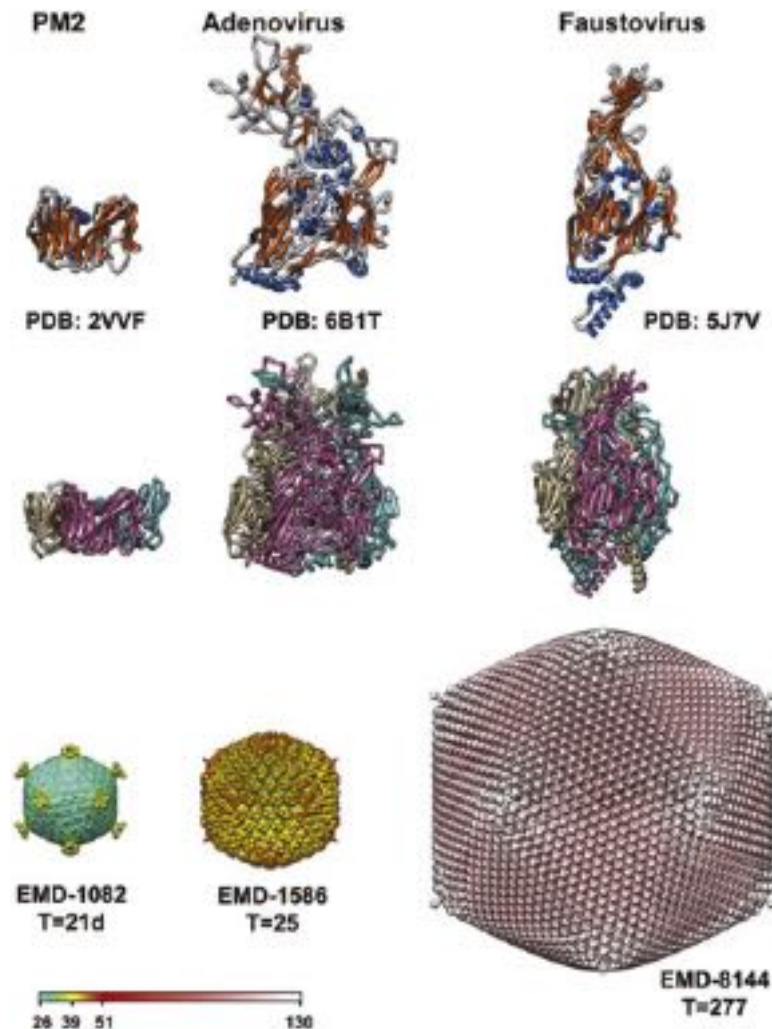


Fig. 7 From the simplest to the most complex double jelly roll virus structures solved so far. The structures of the major capsid protein monomer (top row) and trimer (middle row) are shown, together the complete capsid (bottom row) of bacteriophage PM2, human adenovirus, and Faustovirus. These viruses represent the simplest and most complex examples for which both the high resolution structure of the major coat protein and at least the general capsid organization have been solved. While the PM2 major coat protein is formed by the double jelly roll motif with no more elaborations, the adenovirus and Faustovirus proteins have extensive tower domains which establish intricate interlacing in the trimer. Database identifiers and triangulation numbers are shown by each structure. The color bar indicates capsid radii, in nm. Figure reproduced from San Martín, C., van Raaij, M.J., 2018. The so far farthest reaches of the double jelly roll capsid protein fold. *Virology Journal* 15, 181.

An exception to the use of larger triangulation numbers for increasing capsid size is found in adenovirus. With the same triangulation number (*pseudo*T = 25) as the smaller lineage member PRD1, adenovirus packages more than double the genome length (35 kbp vs 15 kbp in PRD1). One way to achieve a larger capacity in adenovirus is to dispense with the internal membrane. The other is to increase the diameter of the capsomer itself. In the adenovirus hexon monomer, a small β -sheet parallel to the capsid surface is intercalated between the two orthogonal jelly rolls (Fig. 2(a) and Fig. 7), moving them apart to produce a wider pseudo hexagonal base. No tape measure element has been found in adenovirus. The principles that determine adenovirus capsid size are unknown.

Polintons and the Evolutionary Pathway of DJR Viruses

A remarkable evolutionary connection has been found between viruses in the DJR lineage and large (15–20 kbp) eukaryotic dsDNA transposons called Polintons. Polintons are so named because they encode a protein-primed DNA polymerase (to sustain self-replication, POL) and a retroviral-like integrase (INT). Many also include genes for a DNA-packaging ATPase and a maturation protease like those found in DJR lineage viruses. Exhaustive sequence analyses revealed that these transposable elements also encode genes that could translate into double or single jelly roll proteins, suggesting that at some point in time, or in certain conditions, they could form icosahedral capsids. In the light of these findings, an evolutionary model has been proposed in which



Fig. 8 Different architectures of vertex spikes in DJR viruses. Colors refer to the type of host (archaea: red; bacteria, yellow; eukaryotic, green). Database identifiers for the maps used are as follows: HHIV-2 (EMD-3110), SH1 (EMD-1498), STIV (EMD-5584); bacteriophage P23-77 (EMD-1525), PBCV-1 (EMD-1597), and HAdV-D26 (EMD-8471). The scale bar represents 100 Å.

a primordial, PRD1-like phage encoding a DJR capsid protein, a protein-primed DNA polymerase and a packaging ATPase would have invaded a proto-eukaryotic host with a bacterial endosymbiont (a mitochondrion, for example). This primordial phage would somehow have reached the nucleus, and recombined with a eukaryotic transposable DNA element carrying the integrase and maturation protease. This “polintovirus” element would have then evolved in separate ways to produce the polintons (transposable, capsid-less integrating elements), and a variety of eukaryotic “free-standing” viruses, all the way from adenovirus to mimiviruses.

Further Reading

- Abrescia, N.G., Bamford, D.H., Grimes, J.M., Stuart, D.I., 2012. Structure unifies the viral universe. *Annual Review of Biochemistry* 81, 795–822.
- Berk, A.J., 2013. Adenoviridae. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- Butcher, S.J., Manole, V., Karhu, N.J., 2012. Lipid-containing viruses: Bacteriophage PRD1 assembly. *Advances in Experimental Medicine and Biology* 726, 365–377.
- Dai, X., Wu, L., Sun, R., Zhou, Z.H., 2017. Atomic structures of minor proteins VI and VII in human adenovirus. *Journal of Virology* 91.
- Greber, U.F., Flatt, J.W., 2019. Adenovirus entry: From infection to immunity. *Annual Review of Virology* 6, 177–197.
- Hernando-Pérez, M., Martín-González, N., Pérez-Illana, M., *et al.*, 2020. Dynamic competition for hexon binding between core protein VII and lytic protein VI promotes adenovirus maturation and entry. *Proceedings of the National Academy of Sciences of the United States of America* 117, 13699–13707.
- Hong, C., Oksanen, H.M., Liu, X., *et al.*, 2014. A structural model of the genome packaging process in a membrane-containing double stranded DNA virus. *PLoS Biology* 12, e1002024.
- Krupovic, M., Bamford, D.H., 2008. Virus evolution: How far does the double beta-barrel viral lineage extend? *Nature Reviews Microbiology* 6, 941–948.
- Krupovic, M., Koonin, E.V., 2015. Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. *Nature Reviews Microbiology* 13, 105–115.
- Liu, H., Jin, L., Koh, S.B., *et al.*, 2010. Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks. *Science* 329, 1038–1043.
- Ma, H.C., Hearing, P., 2011. Adenovirus structural protein IIIa is involved in the serotype specificity of viral DNA packaging. *Journal of Virology* 85, 7849–7855.
- Mangel, W.F., San Martín, C., 2014. Structure, function and dynamics in adenovirus maturation. *Viruses* 6, 4536–4570.
- Matteson, N.L., Barry, M.A., Reddy, V.S., 2018. Structure-based assessment of protein-protein interactions and accessibility of protein IX in adenoviruses with implications for antigen display. *Virology* 516, 102–107.
- Menéndez-Conejero, R., Nguyen, T.H., Singh, A.K., *et al.*, 2017. Structure of a reptilian adenovirus reveals a phage tailspike fold stabilizing a vertebrate virus capsid. *Structure* 25, 1562–1573. [e5].
- Nicklin, S.A., Wu, E., Nemerow, G.R., Baker, A.H., 2005. The influence of adenovirus fiber structure and function on vector development for gene therapy. *Molecular Therapy* 12, 384–393.
- Ortega-Esteban, A., Condezo, G.N., Pérez-Berná, A.J., *et al.*, 2015. Mechanics of Viral Chromatin Reveals the Pressurization of Human Adenovirus. *ACS Nano* 9, 10826–10833.
- Ostapchuk, P., Suomalainen, M., Zheng, Y., *et al.*, 2017. The adenovirus major core protein VII is dispensable for virion assembly but is essential for lytic infection. *PLoS Pathogens* 13, e1006455.
- Pérez-Berná, A.J., Marion, S., Chichón, F.J., *et al.*, 2015. Distribution of DNA-condensing protein complexes in the adenovirus core. *Nucleic Acids Research* 43, 4274–4283.
- San Martín, C., 2012. Latest insights on adenovirus structure and assembly. *Viruses* 4, 847–877.
- Santos-Perez, I., Charro, D., Gil-Carton, D., *et al.*, 2019. Structural basis for assembly of vertical single beta-barrel viruses. *Nature Communications* 10, 1184.
- van Raaij, M.J., Mitraki, A., Lavigne, G., Cusack, S., 1999. A triple β -spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* 401, 935–938.
- Vassal-Stermann, E., Eftantin, G., Zubieta, C., *et al.*, 2019. CryoEM structure of adenovirus type 3 fibre with desmoglein 2 shows an unusual mode of receptor engagement. *Nature Communications* 10, 1181.
- Xiao, C., Rossmann, M.G., 2011. Structures of giant icosahedral eukaryotic dsDNA viruses. *Current Opinion in Virology* 1, 101–109.
- Zubieta, C., Schoehn, G., Chroboczek, J., Cusack, S., 2005. The structure of the human adenovirus 2 penton. *Molecular Cell* 17, 121–135.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/93/adenoviridae
 Adenovirus.
 dsDNA Viruses
<https://sites.google.com/site/adenoseq/>
 Adenovirus sequences.
 Google Sites

<https://www.emdataresource.org/>

EMDataResource

<https://www.visual-science.com/projects/adenovirus/web-application/>

Human adenovirus structure.

Scientific 3D model.

Visual Science

<https://www.wwpdb.org/>

wwPDB: Worldwide Protein Data Bank

Negative Single-Stranded RNA Viruses (*Mononegavirales*): A Structural View

Juan Reguera, Aix-Marseille University, French National Center for Scientific Research, Marseille, France and French National Institute of Health and Medical Research, Marseille, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

F Fusion protein
G Glycoprotein
gRNA Genomic RNA
HA Hemagglutinin
mRNA Messenger RNA

NC Nucleocapsid
NP Nucleoprotein
NSRV Negative single stranded RNA virus
P Phosphoprotein
RdRpol RNA dependent RNA polymerase
vRNA Viral RNA

Glossary

3' RNA Refers to the RNA end with a free hydroxyl 3'OH group. This end is the last incorporated by the polymerase and the first read for the synthesis of the complementary RNA strand.

5' RNA Is the RNA end with a free phosphate 5'-ppp triphosphate base. Is the first incorporated by the polymerase during RNA synthesis.

Alpha helix A common motif in the secondary structure of proteins as a consequence of the helical arrangement of the amino acids. This secondary structure is stabilized by hydrogen bonds maintained by the amino acid backbone between consecutive helical turns.

Beta sheet A common motif in the secondary structure of proteins as a consequence of the lateral packing of extended amino acid stretches (beta strands). These structures are stabilized by hydrogen bonds maintained between the backbones of the amino acids forming parallel or antiparallel strands.

Budding Process of enveloped virus assembly by which the viral proteins and RNA are packed into a membrane envelope borrowed from the cell membrane.

C terminus Refers to the carboxy end of the polypeptide chain, belonging to the last residue incorporated by the ribosome during protein synthesis.

Hydrophobic Are those compounds or amino acids that are unable to maintain interactions with water molecules, thus tend to cluster together or to interact with hydrophobic environments, like the interior of lipid bilayers.

N terminus Refers to the amino end of the polypeptide chain, belonging to the first residue incorporated by the ribosome during protein synthesis.

Phospholipids Molecules consisting of two hydrophobic fatty acids and a hydrophilic head consisting of a phosphate group. They generate lipid bilayers and are the major components of cell membranes.

Protein domain Is the protein folding unit. Each domain forms a compact three-dimensional structure that can often be independently stable.

Protein loops Protein fragments with no secondary structure and often flexibly deployed from the protein.

Prologue

Negative single stranded RNA viruses (NSRV) are a vast group of plants, animals and human pathogens that includes many relevant human pathogens with serious impact on public health at a global scale. Examples of well-known viruses belonging to this group are Influenza virus which causes yearly pandemic infections around the globe; Ebola and Lassa virus, which cause hemorrhagic fever and cause high mortality among infected patients; Respiratory syncytial virus, which is responsible of severe respiratory disease in newborns and a major concern in pediatrics; and Rabies virus, a zoonotic virus causing severe nervous system infections and that is still present in under-developed countries representing an real health risk, although the first viral vaccine against Rabies was developed by Pasteur in the 1880s. These viruses, among others, have been veritable flagships for the research on NSRV providing critical clues on viral molecular biology, mechanisms for triggering and evasion of the immune response or their viral epidemics. However, these are only the tip of an iceberg that includes hundreds of viral strains with rich diversity in the molecular mechanisms of infection at all levels including the transmission routes, the unique mechanisms of genome replication and transcription, the genome encapsidation in viral particles, the ways of counteracting the immune response and their virion structures.

In this article, we will focus on the current knowledge of the structural basis of NSRVs virions, tackling the structural components associated to different processes of the viral cycle. The aim is to give the reader insight on the mechanisms by which the different virion components, namely proteins, lipids and nucleic acids, together perform the essential processes of virus infection leading to disease. Through techniques such as nuclear magnetic resonance (NMR), X-ray crystallography and electron Cryo-microscopy (Cryo-EM), a large number of proteins and RNA structures of NSRVs have been reported. This structural information

provides, at the atomic level, a large body of knowledge illustrating how infection occurs through the molecular interplay of the different factors forming the virions. Based on this information we can more accurately understand the effect of mutations in a viral strain, explain, to a certain extent, the reasons for conservation or variability of the viral genomic sequences and understand the evolution of the viruses by gain or loss of function. In addition, the atomic structures provide the necessary information for the development of small molecules that can be used as effective treatments for viral diseases by blocking essential catalytic functions or protein interactions. In conclusion, structure determines the function, and through the structural understanding of biological processes we can reach an in-depth comprehension of life and, in this particular case, of infection.

Classification of NSRVs and Genome Organization

NSRVs are included in the Phylum *Negarnaviricota*. This classification is based on the configuration of their genomes and RNA and protein sequence similarities. Briefly, two main NSRV sub-groups are defined by their genomic segmentation. The non-segmented NSRVs that have a single RNA genomic segment (Subphylum *Haploviricotina*), and the segmented NSRVs (Subphylum *Polyploviricotina*) in which the genome is split in two to eight segments. Most of non-segmented NSRVs are included in the Order *Mononegavirales*; Important pathogens like Ebola virus (Family *Filoviridae*), Respiratory Syncytial Virus and Metapneumovirus (Family *Pneumoviridae*), Rabies (Family *Rhabdoviridae*) and Measles (Family *Paramyxoviridae*) are included in this order of viruses. The segmented RNA viruses are mainly classified in two groups, the order *Articulavirales*, including all strains of influenza (family *Orthomyxoviridae*), with genomes split into seven or eight segments, and the large order of *Bunyavirales*, including Rift Valley fever (Family *Phenuiviridae*), Crimean Congo Fever (Family *Nairoviridae*) or Lassa viruses (Family *Arenaviridae*), whose genomes are split in mainly two or three genomic fragments. An important reclassification of these viruses occurred in the recent years, thus in old bibliography this classification varies. What is important to retain is that NSRVs can be segmented or not and this has a direct impact on their biology, from the mechanisms of transcription and replication to the genome packing and structure of the virions.

General Features of NSRV Virion Structures

The common intrinsic feature of NSRVs is that their genomes are composed of RNA of negative polarity, meaning that the nucleotide sequence of the RNA is complementary to the messenger RNAs that are translated by the cellular ribosomes into proteins. In contrast, positive stranded RNA viruses can, just by delivering their positive sense RNAs into the infected cell, directly promote the synthesis of the viral proteins that will take over the cell. In other words, the positive RNA is infectious but the negative RNA is not and needs transcription to become infectious. This strategy for genetic storage challenges the NSRVs to carry all the machinery within the viral particles required to perform transcription after entering the infected cells, leading to translation of viral proteins.

NSRVs have membrane enveloped virions containing the viral genome assembled with the replicational-transcriptional machinery (Fig. 1). These assemblies are usually called nucleocapsids (NCs) or ribonucleoproteins (RNPs), the latter term being used for orthomixoviruses (e.g., Influenza virus) and bunyavirales (e.g., La Crosse virus). The virus envelope is covered by the viral glycoproteins and eventually accessory proteins. All are integral membrane proteins that cross the lipid bilayer and have a part of the protein inside the virion and a part exposed at the surface. Protein glycosylation is achieved by adding long sugar chains at specific sites by the Golgi apparatus of the cell. Glycoproteins recognize the target cell receptors and are generally the major target of neutralizing antibodies, so they aim, as far as possible, to avoid being recognized by immune system antibodies, whilst facilitating the transit through the host organisms and reach the target cells. The glycoproteins specifically recognize target cell receptors and mediate membrane fusion between the viral envelope and the target cell membrane. The part of the glycoprotein inside the virion is small and often maintains interactions with a shield of matrix protein between the lipid envelop and the nucleocapsid or RNPs. This shield underneath the lipid envelope is present in non-segmented viruses and in Influenza-like viruses but is absent in bunyavirales where the glycoproteins are in direct contact with the RNPs. These networks of interactions confer a dynamic stability to the virion that is important for its correct assembly and disassembly during the infection cycle.

The basic configuration of NCs or RNPs consists on the viral RNA genome assembled with multiple copies of the viral NP, that shield the RNA molecule, and at least one viral polymerase. Thus, the non-segmented genomes include one big NC while the segmented include several smaller RNPs into the viral particle. Importantly, non-segmented viruses include an additional viral phosphoprotein (P) in their NCs, that is essential for replication activity. NC assemblies confer stability to the RNA molecule and can determine the shape of the viral particle. For instance, the long nucleocapsids of filoviruses (e.g., Ebola), that package one ~18 Kb (~18,000 nucleotides) long RNA molecule in helicoidal nucleocapsids, confer to the virion a filament shape that can reach ~900 nm in length and 80 nm broad. Rhabdoviruses (e.g., Rabies) package their ~11–15 Kb long genomes into broad helicoidal nucleocapsids, conferring to the virion a bullet-like shape of 180 nm long and 75 nm of diameter in the basis. The rest of non-segmented NSRVs have normally a spherical shape. Paramyxoviruses (e.g., Measles) or pneumoviruses (e.g., RSV) package their genomes (RNAs between 13 and 15 Kb long) in spherical virions of ~150 nm diameter. On the other hand, all segmented NSRVs virions have a globular shape. Orthomyxoviruses (e.g., Influenza) and bunyaviruses (e.g., RVFV) package their RNPs in virions ~80–120 nm diameter. The RNA segments of Influenza comprise in total 13,5 Kb of RNA, and for the different tri-segmented bunyaviruses the total length varies between 11 and 19 Kb.

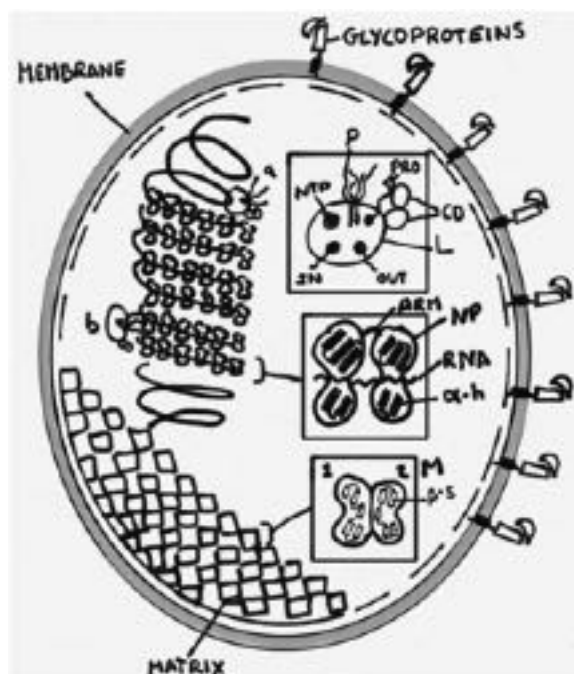


Fig. 1 Schematic drawing of a negative stranded virus taking paramyxovirus as an example. The draw shows the virus enveloped by a lipidic membrane. In the membrane are shown the viral glycoproteins responsible of cell binding and membrane fusion, they are anchored to the membrane by a transmembrane region and are in contact with the matrix proteins which form a shield of square shaped dimers of matrix proteins, the monomers are folded into beta sheets (β -S in the inset panel M). Inside the matrix we have the nucleocapsid made of the vRNA genome of negative sense completely covered with nucleoproteins (partially un-covered in the drawing for clarity) and one polymerase (L) in complex with phosphoproteins. The nucleoproteins are folded in alpha helixes (α -h in the inset panel NP) and are bilobular with a central positively charged crevice with the RNA bound. NPs are forming an helical mega complex by NP–NP contacts through flexible arms and the RNA. The polymerase, L, is linked to the NC through the tetrameric phosphoprotein (P) which binds both the L protein and the NPs (marked at b). The top inset panel shows the polymerase in complex with the tetrameric phosphoprotein. vRNA accesses the polymerase through the IN channel and exits through the OUT channel. To allow RNA synthesis the L protein has in addition an NTP entry channel and a product exit channel (PRO) from where the newly synthesized RNA comes out. In case of replication the product is a new vRNA and in case of transcription is a capped mRNA after its labeling by the capping domains (CD). When the polymerase is synthesizing RNA in the cell cytoplasm it is attached to the NC and harbors RNA inside (shown at a). It can also (e.g., in the context of the virion) be simply attached through contacts bridged by the phosphoprotein (shown at b).

Structure of the Virion Surface

Structure and Function of Virion Glycoproteins

Glycoproteins coating NSRV virions are very diverse in their structure, assembly, receptor recognition and membrane fusion mechanisms. In all cases the glycoproteins mediate the proper virion attachment to the target cell and trigger the internalization of the viral particles into the cell. The two main structural elements of the glycoproteins are the region specifically binding the cellular receptor and the region that fuses the virus and cell membranes for delivering the viral RNPs or NCs into the cell. These two regions are often in the same glycoprotein but can also exist in different glycoproteins, as in the case of paramyxoviruses, where receptor binding is mediated by the hemagglutinin-neuraminidase (HN) protein and membrane fusion by the fusion protein F. Nevertheless, the two proteins work tightly together ensuring that binding to the receptor is coupled to subsequent membrane fusion.

Target cell binding is achieved by specific interactions of certain exposed regions of the glycoproteins that have stereochemical complementarity with the cell receptor, like keys that specifically fit into a particular lock. These protein regions are called receptor binding domains and are the main target for the neutralizing antibodies, which are produced by the immune system and compete with the receptor for binding in order to block viral entry. The glycoprotein's affinity to cellular receptors is usually not very high but the accumulation of high numbers of glycoproteins in the virion surface results in the interesting phenomenon of avidity by which the glycoproteins act as a sort of "velcro", increasing the affinity of an otherwise weak interaction. The virus is thus able to through more or less specific interactions with cellular surface factors. For instance, cellular glycoproteins as DC sign (phlebovirus, Ebola) TIM1 and 4 (Ebola) Sialic acid (Influenza, Ebola) have been identified as receptors and co-receptors for NSRVs. The interaction with the right combination of specific receptors is a major determinant of cellular tropism for the virus and triggers the internalization of the virion through cellular endocytic pathways, the membrane protein recycling system of the cell. The endosomes progressively increase the proton concentration inside, this reduction of pH is detected by the glycoproteins triggering a drastic conformational change (from prefusion to post-fusion states) that allows the insertion of the so-called fusion peptides

into the endosomal membrane. The fusion peptides are short hydrophobic loops of the polypeptide chain that anchor the glycoprotein to the endosomal membrane, allowing the subsequent drawing together of the viral and endosomal membranes leading to their fusion. Upon fusion, the viral replication machinery is released into the cellular cytoplasm.

Interestingly, the three different classes of fusion proteins identified in nature, can be found among NSRV glycoproteins. Influenza virus Hemagglutinin and Ebola virus glycoproteins are class I fusion proteins. They are generated by a single polypeptide precursor that is subsequently cleaved by a cellular protease into two subunits. The subunits stay covalently linked by a disulfide bond. One is responsible for receptor recognition (HA₁ and GP₁ for influenza and Ebola respectively) and the second (HA₂ and GP₂) is responsible for membrane fusion and anchors the protein to the virion membrane by a transmembrane region. The protein cleavage, carried out by furin-like cellular proteases, is required for fusion activity. The virion is coated by these activated glycoproteins forming metastable trimers.

Bunyavirales glycoproteins are class II fusion proteins, which are heterodimers of glycoproteins Gn and Gc. They are also produced as glyco-polypeptides that, after cleavage by the cellular protease (signalase) in the endoplasmic reticulum, become heterodimers. Further processing may occur depending on the viral strain. The Gn/Gc complex associates into larger complexes driving virion budding and inducing the necessary membrane curvature. Gn is responsible for receptor recognition and Gc is a class II fusion protein. Unlike the class I fusion glycoproteins of Influenza and Ebola virus or other class II fusion proteins in nature (e.g., in positive stranded RNA viruses Flavivirus and Alphavirus), the Gn/Gc complex does not need proteolytic processing in the cell surface for fusion activation. These glycoproteins tend to form a tight lattice on the virion surface. For Hantavirus (a family of bunyavirales) has been nicely shown the tetrameric spikes (form by four heterodimers) arranged into a tight lattice stabilized by contacts between spikes.

Finally, Rhabdovirus glycoprotein G is a class III fusion protein. It is monomeric and forms trimers in the virion surface. In this case the only factor that triggers the fusion state are differences in protonation (pH) of the environment. Unlike the other classes, the pre-fusion and post-fusion states are reversible, conferring a dynamic nature to these fusion proteins.

The fusion mechanisms are determined by the structure of the glycoproteins. Class I proteins (Influenza, Ebola) have a central stem consisting of mostly alpha helices shrouded by the receptor binding domains. Lowering the pH of the activated glycoprotein triggers a drastic conformational change that extends the central stem, casts aside the receptor binding regions, and inserts hydrophobic residues at the tip into the lipid bilayer of the endosome membrane. Bunyavirus Class II fusion dimers, in contrast, have two stems (one for each monomer), bridged by a domain rich in beta sheets. The low pH triggers the stems to move closer, forcing the extension of the beta sheet domain harboring the hydrophobic loops (fusion peptides) at its tip, to penetrate the endosomal membrane. Class III proteins are more dynamic and have a central trimeric stem surrounded by domains rich in beta sheets that points towards the virion. The low pH causes a 180-degree rotation of the beta-rich domains towards the endosome membrane, joining together the fusion peptides from the tip of the three subunits and allowing their insertion in the endosomal membrane. In all cases the glycoproteins in the post-fusion state bring together the two membranes, leading to membrane fusion.

Overall, the glycoproteins are more-or-less sensitive to different markers of the virion life cycle. The need of proteolytic activation of the protein fusion prevents the virion from prematurely fusing to the cell of origin, so it can be ready to infect the next one. The interaction with specific receptors allows the internalization of the viral particle and the decrease of pH in the endosome triggers the post-fusion state. The viral glycoproteins act as sensors that control a well-programmed mechanism for virion spreading. It is remarkable that whilst there is no homology between the glycoproteins of NSRVs influenza, bunyavirus or Rhabdovirus these instead show homology with fusion machines found in other less related viruses and organisms. Thus class I fusion proteins are found in retroviruses as HIV, Bunyavirus class II fusion glycoproteins have homologs in positive stranded RNA viruses as flavivirus and Alphavirus, and even in the cell fusion protein of *C. elegans* EEF-1, while the Rhabdovirus class III fusion proteins are homologous to glycoproteins in herpesvirus (a DNA virus). These reflect the apparently independent evolution of the glycoproteins and viruses and illustrates the genetic exchange between organisms during evolution or, in other words, "the mosaic nature of viral genomes, which are a patchwork of genes having different origins (Guardado-Calvo and Rey, 2017).

Accessory Proteins in the Virion Surface

NSRVs can incorporate additional viral proteins in the virion surface. The most remarkable example is the neuraminidase of influenza virus. This protein cleaves the sialic acid (the receptor moiety recognized by the viral glycoprotein) of cellular glycoproteins. This cleavage allows the release of the virions that otherwise would be retained in the cell of origin, rendering the cleavage essential for influenza infectivity. Paramyxovirus also incorporates neuraminidase activity but in the same protein that recognizes the receptor, the hemagglutinin-neuraminidase glycoprotein. Other NSRVs lack any neuraminidase.

Influenza virus also incorporates a small membrane protein (M2) in the virion membrane. This protein forms tetrameric pores that function as a pH-sensitive proton conducting channel essential for virus infectivity. In addition, M2 interacts with the matrix protein coating the inner side of the virion membrane.

Other viruses incorporating auxiliary proteins are those from the family pneumoviridae (e.g., Metapneumovirus). The G protein and sh proteins are present in the virion surface, but are not essential for the infection. The heavily glycosylated G protein is disordered and modulates the innate and adaptive immune responses by providing a heavily glycosylated shield to the virion, thus preventing their recognition by neutralizing antibodies. As a consequence, the incomplete immunization of the host permits reinfections. The hydrophobic sh is of unknown function, has nevertheless been shown to be important for the infection in vivo.

Inside the Virion

The Matrix Proteins

Matrix proteins are essential for virion formation and stability. They generate a two-dimensional network underneath the virion envelope, connecting the different components of the virion by simultaneously interacting with membrane lipids, glycoproteins and NPs. In consequence matrix proteins are directly involved in the recruitment of all the above factors that are necessary for the budding of the RNPs or NC into virions at the cell membrane. With the exception of bunyavirales, all NSRVs have matrix proteins buttressing the viral envelope. In mononegavirales the matrix proteins are small and folded in two very similar beta-strand rich domains. These matrix protein monomers form square shaped dimers (e.g., Ebola VP40, Paramyxovirus M and Borna disease virus M proteins), which are the building blocks of a two-dimensional lattice by interdimer contacts through their square vertices. One side of the protein is very positively charged, generating an extended positively charged network that is important for binding to the negative phosphate moieties of the envelope phospholipids. Through this mechanism the matrix proteins confer a dynamic stability to the viral particle. Orthomyxoviruses have a different matrix protein (M1) which however meets the same functional needs. As reported for the infectious salmon anemia virus, M1 folds in two domains exclusively composed of alpha helices. One domain has a positively charged side facing the viral envelope whilst the other is oriented towards the virion core facing the RNPs. M1 oligomerizes to generate a very tight matrix network, much denser than in the mononegavirales. Once again, the diversity of protein structures addressing the same functional needs is a characteristic of this group of viruses, as if all possible alternatives were explored during their evolution.

The Viral Ribonucleoproteins/Nucleocapsids Assemblies

To incorporate all the replication machinery into compact small virions NSRVs have adopted the strategy of coating their RNA with NPs, small RNA binding proteins capable of self-associating into linear scaffolds that bind and cover the length of the gRNA. In all cases the inter-NP interactions generate a more-or-less pronounced helical arrangement. In addition, the RNA-NP assemblies incorporate at least one copy of the viral RNA dependent RNA polymerase, responsible for the transcription and replication of the genome. The flexibility of the inter NP interactions allows the vRNA to be compacted for packing into virions and relaxed to allow the polymerase to access and read the RNA template. The nucleocapsids or RNPs are thus functional units where all the protein factors act in an intimate and coordinated fashion to carry out replication and transcription. In addition, the interactions of the NC and RNPs with the internal tails of the matrix proteins or the glycoproteins guides the correct budding of the genomes into the virions.

The mononegavirales NCs are linear and, in addition to the gRNA, the Polymerase (L protein) and the nucleoproteins, incorporate phosphoproteins (P) that, by interacting with the polymerase and NPs, play a key role in coordinating the NC function. Conversely, each genomic segment of segmented NSRVs is assembled in circular RNPs. The circularization was thought to be caused by the RNA base-pairing of the genome ends, since all genomic segments have a strong sequence complementarity. Instead, the reported structures on Influenza polymerase and La Crosse bunyavirus L protein in complex with vRNA showed that the circularization is mediated by the interaction of the polymerase with the RNA ends in two different binding sites, specific to each end. In addition to this circularization, the RNPs also maintain a helical arrangement ranging from the highly coiled double helix of Influenza to circular flexible helices of bunyaviridae.

The Nucleoproteins

NSRV NPs are broadly diverse, as is the case for other proteins of the virion. The NP function is to stabilize and linearize the gRNA preventing, to an extent, the formation of secondary structures. The NPs are generally kidney-shaped globular proteins folded in alpha helices that are able to coat the RNA by defining two important surfaces in their structure, the RNA binding groove and the oligomerization interface. The positively charged groove, formed by the two lobes of the kidney-like structure, accommodates the negatively charged RNA. Multimerization is often achieved by N and C terminal extensions of the polypeptide chain acting as arms to grab the neighbor interacting proteins in a head to tail fashion. Upon multimerization, the NP grooves generate a long positively charged channel that accommodates the long gRNA molecules. The interaction of NP and vRNA is sufficiently non-specific to permit the encapsidation of the entire gRNAs independently of their sequence. However, the NPs specifically cover the gRNA and not viral mRNA or any other RNA in the cell. The mechanisms of specific encapsidation are related to the capacity of RNA-binding inhibition of the NPs that can be mediated by interactions with the P protein (mononegavirales) or self-inhibition of the monomeric NP (bunyavirales). This inhibition is eliminated by specific RNA sequences and/or the polymerase during the synthesis of the newly synthesized gRNAs. Further experimental evidence is required to clarify these mechanisms.

The arrangement of NPs on the gRNA is different for segmented and non-segmented NSRVs. The structure of measles or Ebola virus nucleocapsids shows the NPs closely interacting with each other side to side. They form a broad helical assemblies with the RNA binding sites facing the exterior. The flexible extensions on both sides of the NP interact with the neighboring NPs conferring further stability to the assembly. The gRNA is tightly bound to the NPs but remains accessible to the polymerase from the exterior of the helix. Contacts between NPs of adjacent helical turns compact and rigidify the assembly. The sizes of these helices are variable; Ebola nucleocapsid helices have a diameter of ~ 295 Å and ~ 19 NP per turn, compared to the ~ 190 Å diameter and ~ 12 NP per turn helices of measles. In both cases 6 nucleotides of the viral RNA are accommodated per NP. Conversely, the

segmented NSRV NPs tend to form narrower helical assemblies with the RNA binding groove facing the interior of the helix. This is the case for members of the families *Peryorthobunyaviridae* (e.g., Bunyamvera or La Crosse viruses) and *Phenuiviridae* (e.g., Rift Valley fever virus). The RNA, inside the helical structure, would be inaccessible to the polymerase, which relies on flexible NP arms connecting the proteins to allow the eventual flipping of the NPs, transiently exposing the RNA to the exterior. However, more rigid RNP structures are found, as in the case of Hantaviridae family members (e.g., Hantaan or Andes virus) where the NPs weave a net of interactions building a very closed and rigid helical assembly with the RNA inaccessible in the interior. How RNA is synthesized by the polymerase in these structures is still an open question.

Not surprisingly, we can find interesting variations of the NP structure. Arenavirus (members of the family *Arenaviridae*, order *bunyavirales*) have larger NPs that incorporate an additional domain consisting of an exonuclease able to cleave RNA. The function of this endonuclease is related to the suppression of the interferon response and, although its importance in viral replication is limited, mutant viruses lacking this activity are not able to infect animals, providing another example of the importance of viral interference with the host immune response for successful infection. Another peculiar NP belongs to the Influenza virus (and other viruses included in the family *Orthomyxoviridae*). These NPs coat circular gRNAs that coil further in a flexible double helix assembly of variable diameter ($\sim 150\text{--}200\text{ \AA}$). NPs cover the RNA head to tail and constitute a bunyavirus-like circular assembly with the polymerase holding both RNA ends. In addition, influenza NPs interact in an antiparallel fashion generating the supercoil resulting in a double helix formation. The inter-strand NP–NP contacts are very flexible allowing the displacement of one NP–RNA stream over the other. This very sophisticated mechanism allows the polymerase to read the whole genome in a compact RNP assembly, while keeping both RNA ends attached (see relevant web pages for explanatory videos).

The Polymerase

sNSRV polymerases are large multifunctional proteins able to transcribe and replicate the viral RNA. They are monomeric (single polypeptide chains), referred to as L proteins, with the notable exception of the heterotrimeric Influenza polymerase (made of three different proteins). Their size is generally around 2250 amino acid residues but viruses belonging to certain families (e.g., *Nairoviridae*) have polymerases of around 4500 amino acids. The polymerase perform RNA dependent RNA synthesis that, for replication, is coordinated with the assembly of new RNPs by coupling the RNA synthesis to the incorporation of NPs and, in segmented viruses, at least one polymerase on the nascent strand. Transcription is achieved by incorporating the molecular labels rendering the viral mRNAs recognizable by the cellular translational machinery. These modifications are the polyadenylation, capping and methylation of the RNA. Thus these polymerases adopt two modes of RNA synthesis, one incorporating NPs in the nascent RNA (replication) and the other labeling the uncoated RNA to resemble a cellular mRNA (transcription).

The polymerases of NSRV have all the basic canonical RdRpol fold with palm, fingers and thumb subdomains. In addition, similarly to polymerases of double stranded RNA viruses (e.g., reovirus), they incorporate N-terminal and C-terminal extensions forming an enclosed chamber connected to the exterior by four tunnels. Inside the chamber, the emerging product and template RNA strands are separated at an early stage of RNA synthesis (Fig. 1). Template entry and exit tunnels are close to each other facilitating reading the genome with minimal RNP disruption, while the product exit tunnel guides the nascent transcript or replicate towards the RNA transcription processing machinery, which is flexibly linked to the core. This flexibility is important for adopting different functional conformational states that determine the transcription or replication mode of the polymerase.

Beside these common features, segmented and non-segmented NSRV have substantial differences in their mechanisms of transcription and this is reflected by different structures of their transcription processing machinery. Non-segmented viruses perform transcription by capping and methylating the RNA with a polyribonucleotidyltransferase (PRNTase) and a methyltransferase domain included in their polymerases at the C terminus. These two domains respectively transfer the GTP moiety to the nascent RNA and methylate this GTP in the N7 position. The PRNTase is included in the core of the polymerase and the methyltransferase is flexibly linked at the exit of the product exit channel, allowing the RNA to be capped upon the exit from the polymerase core.

Segmented NSRVs polymerases perform transcription by cap-snatching. They “snatch” 5′ short capped fragments from the cellular mRNAs using domains with distinct cap-binding and endonuclease activities. These domains are at the N and C terminus of the RdRpol core respectively. Interestingly, despite the lack of amino acid sequence conservation, all these polymerases have a similar overall architecture. Another striking feature of segmented NSRVs polymerases is their interactions with the vRNA. Two RNA binding sites specific for the 5′ and 3′ vRNA ends hold the RNA molecule and cause its circularization. 5′ binding allosterically regulates the active site of the RdRpol and 3′ binding tightly holds the RNA during the pre-initiation and the elongation of the RNA synthesis, maintaining the circular structure of their RNPs. This elaborate mechanism allows vRNA to regulate the polymerase activity and keeps it always bound to the RNP.

Phosphoproteins

Perhaps because of the absence of the polymerase–RNA interactions seen in segmented viruses, non-segmented NSRVs include an extra protein in their nucleocapsids: the phosphoprotein. This protein is mostly disordered, can be phosphorylated and forms tetramers. A central domain made of alpha helices mediates tetramerization as well as the interaction with the L protein, resulting in the formation of a functional complex with L protein with a 4:1 ratio. The N terminus interacts with the viral NP preventing its nonspecific binding to cellular RNAs and thus rendering it available for the nucleocapsid assembly. The C terminus interacts with the N terminal tail of NP engaged in the NP–NP multimerization interface in the nucleocapsid, assuring attachment to the

nucleocapsid. However, the N and C termini have also been shown to interact with the L protein inducing certain conformational changes regulating its activity. Thus, P proteins have several ways of interacting with the L and the NP proteins and this network of interactions is a major regulatory factor of transcription and replication. The P protein thus facilitates the access of the L protein to the gRNA template in the NP, allows the incorporation of RNA-free NPs to the nascent nucleocapsids during gRNA synthesis and regulates L protein activity.

Concluding Remarks

As we have seen, NSRV biology is complex because of the diversity in the molecular mechanisms they have developed. Despite maintaining some basic principles, NSRVs have adopted at all levels (e.g., transcription, replication, membrane fusion) an original set of solutions to face the biological challenges for infection. In comparison to prokaryotic or eukaryotic organisms, which share highly similar basic molecular processes, NSRVs can be regarded as veritable bio-reservoirs of alternative molecular pathways. A single chapter does not allow an in-depth description of all the biology known for them. Furthermore, the study of poorly known NSRV viral families will reveal new variations on the mechanisms described here. Thus, studying NSRVs will not only be crucial for facing possible future pandemic challenges but will also expand our very basic molecular understanding of life.

Further Reading

- Cox, R.M., Plemper, R.K., 2017. Structure and organization of paramyxovirus particles. *Current Opinion in Virology* 24, 105–114.
- Desfosses, A., *et al.*, 2019. Assembly and cryo-EM structures of RNA-specific measles virus nucleocapsids provide mechanistic insight into paramyxoviral replication. *Proceedings of the National Academy of Sciences of the United States of America* 116, 4256–4264.
- Guardado-Calvo, P., Rey, F.A., 2017. The envelope proteins of the bunyavirales. *Advances in Virus Research* 98, 83–118.
- Kranzusch, P.J., Whelan, S.P., 2012. Architecture and regulation of negative-strand viral enzymatic machinery. *RNA Biology* 9, 941–948.
- Radzimanowski, J., Effantin, G., Weissenhorn, W., 2014. Conformational plasticity of the Ebola virus matrix protein. *Protein Science* 23, 1519–1527.
- Reguera, J., Cusack, S., Kolakofsky, D., 2014. Segmented negative strand RNA virus nucleoprotein structure. *Current Opinion in Virology* 5, 7–15.
- Reguera, J., Gerlach, P., Cusack, S., 2016. Towards a structural understanding of RNA synthesis by negative strand RNA viral polymerases. *Current Opinion in Structural Biology* 36, 75–84.

Relevant Websites

- <https://www.cell.com/cms/10.1016/j.cell.2020.03.061/attachment/b4534876-c9c3-4baa-8e15-62eb6c00f82d/mmc4>
CellPress.
- <https://talk.ictvonline.org/taxonomy/>
International Committee on Taxonomy of Viruses (ICTV).
- https://static-content.springer.com/esm/art%3A10.1038%2Fs41564-020-0675-3/MediaObjects/41564_2020_675_MOESM3_ESM.mp4
Video illustrating the Influenza transcription in double helical RNPs.
- <https://viralzone.expasy.org/>
ViralZone.

Structure of Retrovirus Particles (*Retroviridae*)

David K Stammers and Jingshan Ren, University of Oxford, Oxford, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

The presence of virions is characteristic of the basic requirement for the propagation of viruses to be sustained (Vogt, 1997). Knowledge of the structure of virions or their components represents an important means of understanding the functioning of such systems. Enveloped RNA viruses are made up of a range of constituent parts that must carry out five main functions: (1) an enzyme for copying the viral RNA in the newly infected cell (in the case of –ve strand RNA viruses and retroviruses), (2) formation of the genome in complex with proteins, (3) packaging of this RNA-protein complex inside a capsid shell, (4) an envelope added to the capsid in the form of membrane lipids and (5) addition to the envelope of proteins that contain the molecular recognition determinants for binding to a host cellular target and promoting membrane fusion (Vogt, 1997). The assembly of the virion structure together with the encapsidation of the genomic RNA form a ‘late stage’ in the virus life cycle.

Retroviruses are RNA viruses, which are enveloped, and have been shown to target a wide range of species. A retrovirus is so-called because of the flow of information from the genetic material is in a backward (retro) direction i.e., from RNA → DNA → RNA → protein rather than the more usual route, namely DNA → RNA → protein. A key property of retroviruses is the integration of a DNA copy of the viral genomic RNA into the host genome (the provirus), using the virally encoded integrase. The activity of the provirus thus makes an infection persistent, potentially requiring lifelong therapy in the case of a virus causing human disease such that caused by Human immunodeficiency virus (HIV). In contrast, for many other viruses the infection is only transient or may flair up intermittently in which case drug treatment, if available, is required only for a limited period.

Retroviruses are thought to have evolved from Long Terminal Repeat (LTR)-retrotransposons (Dodonova *et al.*, 2019), which are transposable elements that can form a significant proportion of an organism’s DNA. In the case of the human genome it amounts to ~8% of the total DNA. Retrotransposons do not have an extracellular phase of their life cycle and thus are less easily able to spread to other cells. The evolution of retrotransposons into retroviruses appears to be via the acquisition of a lipid coat together with an *env* gene, the product of which recognizes a cellular receptor on a target tissue of particular organisms infected and promotes fusion of lipid membranes from the virus and host, hence propagation of the virus (Dodonova *et al.*, 2019).

The first retroviruses to be identified were Rous sarcoma virus (RSV) and mouse mammary tumor virus (MMV), both tumor inducing viruses found in mice. It was also later shown that a long known disease giving rise to anemia in horses was also caused by a retrovirus, in this case equine infectious anemia virus (EIAV) (Vogt, 1997).

Retroviruses are important human pathogens. Human T-cell lymphotropic virus I (HTLV1), which can cause a T-cell leukemia, is an early example. A more recent threat to human health is the global pandemic of Acquired Immuno-Deficiency Syndrome (AIDS) which was caused by a retrovirus, HIV. In this syndrome the infection by HIV results, if untreated, in an eventual catastrophic depletion of CD4⁺ lymphocytes thereby making individuals susceptible both to opportunistic infections as well as to various cancers.

A further factor in making HIV-1 infection so persistent is the rapid turnover of the virus with a cycle time of 1–2 days. Such rapid replication together with the error prone viral reverse transcriptase lacking any editing or corrective functions results in the generation of many mutant forms of the viral genomes (Coffin, 1995). As a result drug treatment will potentially rapidly select resistant virus from the array of quasi-species present, an effect that can be countered to a greater extent by the use of multi-drug cocktails. New drugs acting on novel target sites are required to maintain the ability to suppress HIV infections. Only a limited number of anti-HIV drugs target the virion structural proteins (as opposed to the vast majority of anti-HIV drugs that inhibit the pol enzymes) including the fusion inhibitor enfuvirtide and the co-receptor antagonist Maraviroc. Currently there are experimental compounds that target the HIV-1 capsid protein but as yet none have been approved for clinical use (Purdy *et al.*, 2018).

HIV-1 has thus been intensively studied in order to provide a better understanding of the properties of the virus. Such research has yielded much structural data which have been invaluable in assisting in the development of new drugs against a range of viral targets particularly the pol enzymes. A continued further long term aim is to utilize structural data to aid in the design of vaccines, which to date have proved extremely difficult to develop against HIV due to the ability of the virus to evade the immune system. A useful vaccine is still seen as an important goal in order to prevent further spread and eventual elimination of HIV and AIDS (Ward and Wilson, 2017).

On account of this focus there has been a very significant number of studies of the structural aspects of HIV-1 which have included the whole virus in both immature and mature forms, as well as protein components. Use of X-ray crystallography or solution NMR of individual proteins and cryo-EM of macromolecular assemblies of virus components have been combined to greatly improve the understanding of the relation of structure to function. HIV-1 is by far the most studied retrovirus we thus use this virus as the main example in describing retroviral virion structure. Due to space limitations it is not possible to refer to all the research papers in this field, rather the many excellent review articles which describe both the overall retrovirus virion structure as well as components should be referred to for more detailed information (Conte and Matthews, 1998; Briggs and Krausslich, 2011; Sundquist and Krausslich, 2012; Mattei *et al.*, 2016b; Perilla and Gronenborn, 2016; Pornillos and Ganser-Pornillos, 2019; Zhang *et al.*, 2015; Chen, 2019; Wilen *et al.*, 2012; Ward and Wilson, 2017).

Gag Proteins and Retroviral Virion Structure

The assembly of the immature retrovirus, following viral gene expression which in turn leads to virion maturation, occurs towards the end of the phase of retrovirus infection. These events follow the earlier phases of cell entry, reverse transcription and DNA integration into the host genome.

The assembly of retroviruses is directed by the viral protein, Gag, the Group-specific antigen protein which forms the main structural component. The expression of Gag (p55 for HIV) is sufficient for the production of non-infectious virus-like particles (VLPs), (Briggs and Krausslich, 2011). Gag is a polypeptide that has several protein domains interspersed by flexible linker regions. Three of the Gag domains show conservation across retroviruses (Fig. 1). Firstly the matrix protein (MA) which directs Gag to the plasma membrane and also assists in the viral glycoprotein Env being incorporated into the assembling virions (Murakami, 2008). Secondly the capsid protein (CA) which is involved in forming oligomers of Gag leading to capsid formation and thirdly the nucleocapsid protein (NC) which binds to nucleic acids including the genomic RNA allowing it to be packaged into the newly formed virus. The assembly of these components leads eventually to the mature HIV virion (Fig. 2). Details of the structural changes going from the immature to mature HIV virion are shown in Fig. 3.

The exact composition of Gag varies amongst different retrovirus genomes as seen in comparison of the gag genes of HIV-1, RSV, Mason-Pfizer monkey virus (*m*-PMV) and murine leukemia virus (MLV). In some cases there are short linker peptides as is the case for HIV, where SP1 and SP2 as well as p6 are released by the viral protease (Mattei *et al.*, 2018) (Fig. 1). For HIV-1 the pol gene

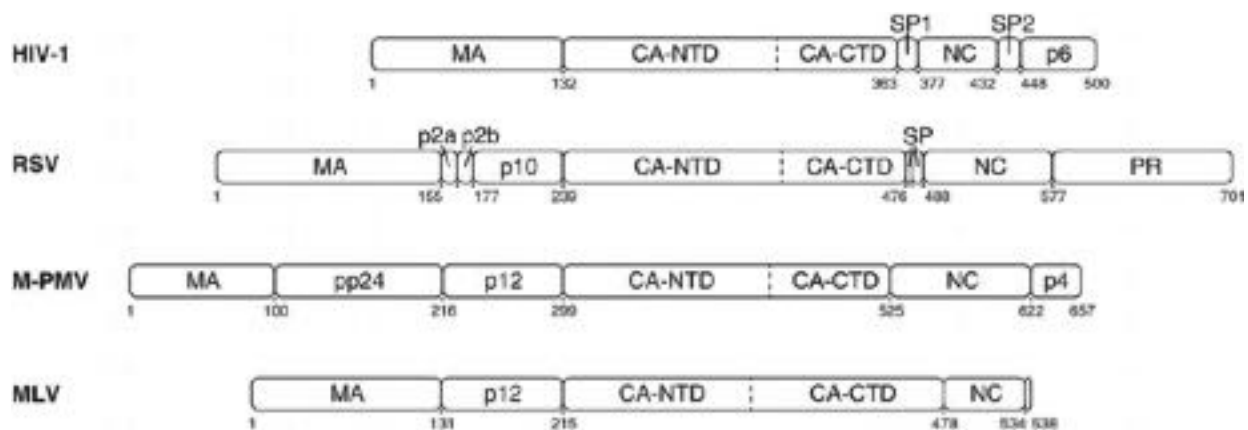


Fig. 1 A schematic depiction of Gag domain architecture: - in human immunodeficiency virus type 1 (HIV-1), Rous sarcoma virus (RSV), Mason-Pfizer monkey virus (*m*-PMV) and murine leukemia virus (MLV). The three canonical domains MA, CA and NC are conserved. Outside of these domains additional domains and peptides are present in the different retroviruses. Reproduced from Mattei, S., Schur, F.K., Briggs, J.A., 2016b. Retrovirus maturation an extra-ordinary structural transformation. *Current Opinion in Virology* 18, 27–35, with permission.

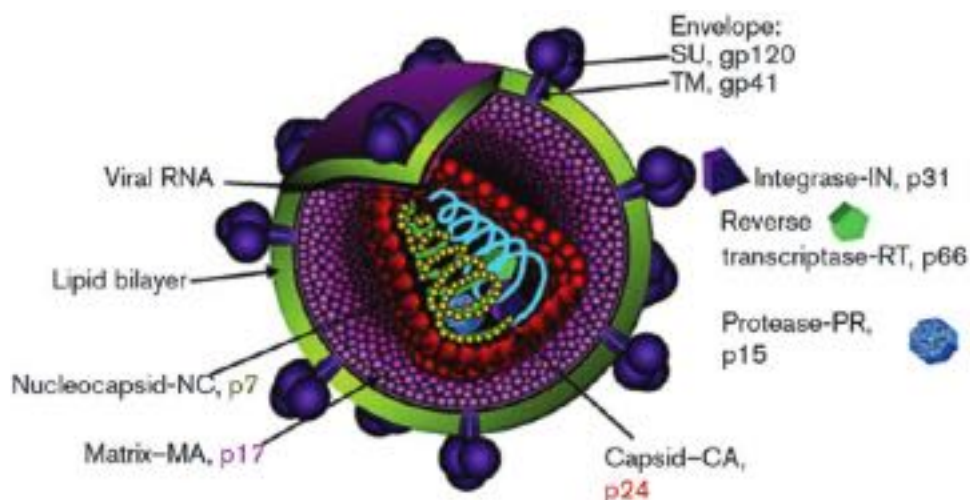


Fig. 2 HIV virion structure. Schematic representation of a mature HIV virion detailing the localization of viral proteins and the approximate virion structure. The representation is not to scale. Reproduced from Steckbeck, J.D., Kuhlman, A-S., Montelaro, R.P., 2013. C-terminal tail of human immunodeficiency virus gp41: Functionally rich and structurally enigmatic. *Journal of General Virology* 94, 1–19, with permission.

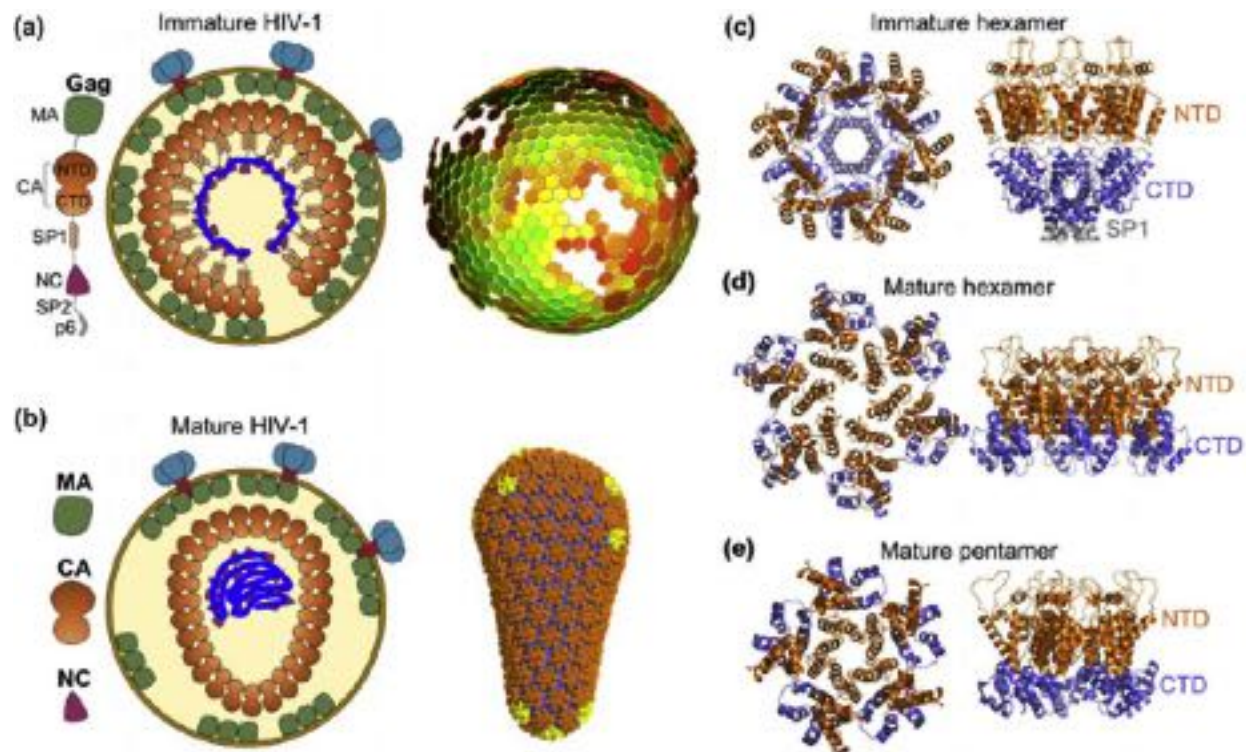


Fig. 3 Structural organization of immature and mature HIV-1 particles. (a) The immature virion is organized by the Gag polyprotein and its multiple domains. Gag assembles into a spherical immature lattice made up of interlinked Gag hexamers. (b) The mature virion is organized by the mature structural proteins – MA, CA, and NC – which are derived from Gag. The genome is encapsidated in a fullerene capsid made up of CA hexamers and pentamers. (c) Structure of the immature hexamer, made of the CA subdomains (NTD colored in orange and CTD in blue) and downstream SP1 spacer (colored in gray). (d) Structure of the mature hexamer. (e) Structure of the mature pentamer. Reproduced from Pornillos, O., Ganser-pornillos, B.K., 2019. Maturation of retroviruses. *Current Opinion in Virology* 36, 47–55, with permission.

encodes enzymes required for replication of the virus (including reverse transcriptase and integrase) and is translated as a Gag-pol fusion protein following a (–1) ribosomal frameshift event. The latter occurs with a frequency of 1 in 20, hence providing the means for producing the correct ratio of Gag to pol proteins as required for virion production and genomic RNA transcription into DNA which is subsequently integrated into the host DNA. During the virus maturation phase these individual proteins of the gag-pol polyprotein are released by the action of the HIV protease. The protease also further processes the reverse transcriptase to form the active p66/p51 heterodimer, retaining one RNase H domain in the p66 subunit and additionally yielding an isolated p15 RNaseH domain which is apparently non-functional.

There are two types of Gag proteins as seen in the pathway followed for morphogenesis of new virus. For lentiviruses, type C viruses and most other retroviruses (Mammalian & Avian) the Gag proteins are targeted to the plasma membrane where immature particle formation followed by budding occurs. For Type D and Type B viruses (M-PMV and MMTV, respectively) the Gag protein precursors first assemble into a core particle which is subsequently transported to the plasma membrane for release (Conte and Matthews, 1998; Briggs and Krausslich, 2011; Perilla and Gronenborn, 2016).

Gag proteins form the principal components of the HIV-1 virion, accounting for 50% of the mass, whilst viral membrane lipids make up approximately 30% of the virion mass (Sundquist and Krausslich, 2012). Other viral and cellular proteins amount to ~20%, whilst the two copies of the genomic RNA together with other small RNAs account for ~2.5% of virion mass. Virally encoded proteins such as reverse transcriptase, integrase, protease and the nucleocapsid protein are also present in the virion, as is t-RNA-lys3 which is required as a primer for the reverse transcription of the genomic RNA into DNA. We will focus on the structures of the Gag proteins as these form the main constituents of the virion and are largely responsible for the assembly of new virus.

Structures of Individual Gag Domains

It has not proved possible to determine the structure of the full length Gag protein due to the presence of flexibility in the interdomain linker regions which results in a heterogeneous protein conformation precluding structure determination. Instead structures of the larger individual Gag domains have been determined to high resolution using crystallography or NMR.

Matrix Protein (MA)

The Matrix protein (MA), also referred to as p17 in HIV, forms the outer shell of the virion just beneath the lipid membrane (Figs. 2 and 3(a)). In HIV, MA is co-translationally modified by the addition of a myristoyl group on the N-terminus. Such a modification is essential for trafficking of the Gag protein to the plasma membrane which is the site of virion release. Also required are a group of positively charged residues that form a basic patch interacting with the acidic head groups of various phospholipids located in the inner leaflet of the plasma membrane including phosphatidylinositol 4,5-bisphosphate (PIP2) (Lingappa *et al.*, 2014).

In none of the numerous cryo-EM studies of the structure of HIV-1 virions, whether from the mature or immature virus, did the MA domain give any indications of forming an organized structure relative to the capsid domain (Mattei *et al.*, 2016b; Pornillos and Ganser-Pornillos, 2019). As a result of this lack of order, detailed structural information of MAs from different retroviruses has relied upon the study of isolated domains.

Retroviral MAs can vary significantly in length from 100 residues for M-PMV to 150 residues for the protein from RSV. The 3-D structures of several recombinant retrovirus MAs have been determined by use of NMR or X-ray crystallography. These structures include examples from HIV-1, simian immunodeficiency virus (SIV), human T-cell leukemia virus type II (HTLVII), bovine leukemia virus (BLV), M-PMV & MLV (Conte and Matthews, 1998). MAs from different viruses share little sequence identity, yet the protein folds all consist of mainly α -helices. In some cases similarities in structure have been noted by visual inspection (Conte and Matthews, 1998). However, in another example it did not prove possible to overlap HIV-1 MA with M-PMV MA, hence they contain distinctive folds (Fig. 4(a) and (b)). Analysis of the packing in the crystal structures of SIV MA showed the presence of a trimer that is related by a non-crystallographic three fold axis (Rao *et al.*, 1995). Later work on HIV-1 MA revealed the same trimeric assembly in several crystal forms, in one case this was located on a crystallographic three-fold axis (Hill *et al.*, 1996). In the MA trimer the conserved basic residues of lysine and arginine present in the N-terminal region form a surface which in conjunction with the myristol group could orientate the trimer towards the membrane (Fig. 4(c)) (Rao *et al.*, 1995; Hill *et al.*, 1996; Conte and Matthews, 1998). It is also interesting to note that MA trimerisation is known to be necessary for the incorporation of the trimers of gp120/gp41 into the HIV virion surface (Murakami, 2008; Tedbury *et al.*, 2016).

Capsid Protein (CA)

The capsid protein (CA) forms the basic building block of immature and mature retroviruses (Perilla and Gronenborn, 2016). It is positioned inside the MA layer of the virion forming the outer region of the core (Fig. 3). CA is a flexible molecule and can readily undergo the conformational changes necessary for the transition between the immature and mature states. A further important property of CA is its ability to form different states of oligomerisation including hexamers and pentamers which are necessary for the formation of the capsid (Perilla and Gronenborn, 2016). CA is reported to have essentially identical three-dimensional structures for most retroviruses (Perilla and Gronenborn, 2016) of which the protein from HIV-1 is the most studied example.

CA consists of about 230 amino acids and is organized into two lobes or domains called CA-NTD (N-terminal domain) and CA-CTD (C-terminal domain) which are both largely α -helical. CA-NTD consists of ~ 150 amino acids and has seven helices. CA-CTD contains four α -helices and additionally a 3_{10} helix (Fig. 5). A region around residues 150–170 is referred to as the major homology region (MHR) and is highly conserved amongst retroviruses. Deletion of the MHR inhibits virus assembly.

One of the main properties of CA is its flexibility, as demonstrated by the linker between NTD and CTD allowing different relative domain orientations which relate to the assembly state of the virion (Fig. 3). CA can oligomerise into a number of different aggregation states, which have particular functional roles. Thus hexamerisation of CA yields the basic unit for capsid formation but

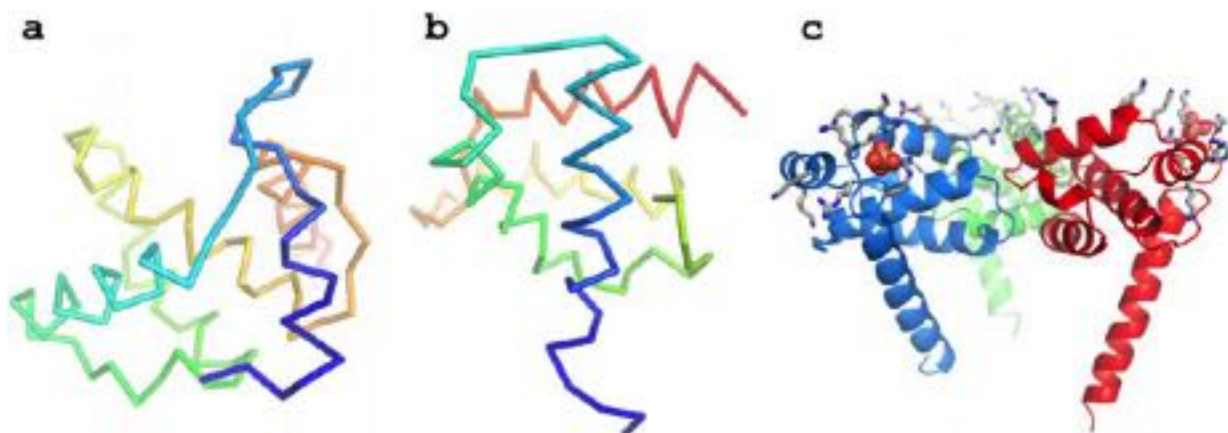


Fig. 4 Comparison of the protein folds of retrovirus matrix proteins (MA). (a) HIV-1 matrix. protein (PDB:1HIW). (b) M-P M virus matrix protein (PDB:1BAT) (c) SIV matrix protein showing trimer (PDB:1ECW). The conserved positively charged residues that are thought to interact with phospholipids in the membrane are marked in gray.

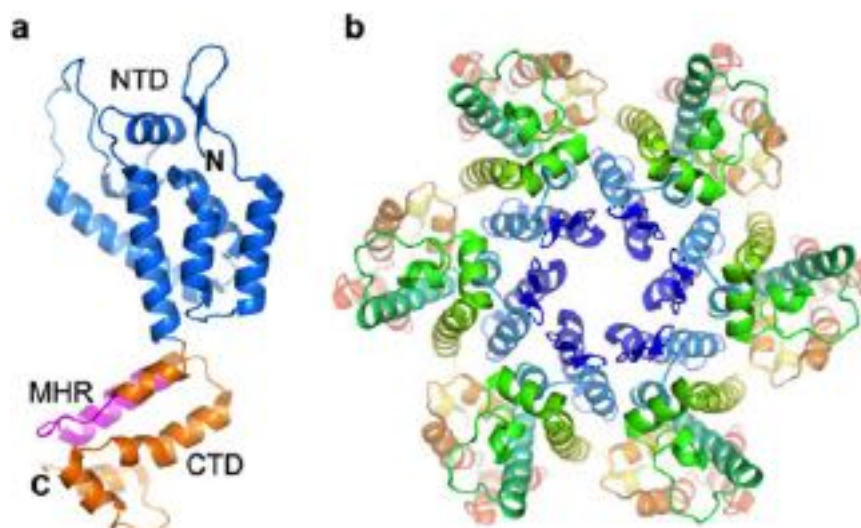


Fig. 5 HIV capsid protein structure CA, (a) monomer showing N-terminal domain (NTD) in blue and C-terminal domain (CTD) in orange (PDB: 3NTE). The major homology region (MHR) consisting of residues 153–172 is colored pink. (b) Hexamer native form (PDB:4XFX).

in order to enclose space, pentamers must also be incorporated, a total of 12 pentamers being required in this case. Early studies introduced disulfide bridges into CA to stabilize the hexamer or pentamer forms (Pomillos *et al.*, 2011). More recently full length unmutated CA has been crystallized as a hexamer leading to the visualization of the native contacts (Gres *et al.*, 2015).

Nucleocapsid Protein (NC)

NC is of variable length depending on the particular retrovirus and contains one or two zinc fingers as well as positively charged regions. NC is a multi-functional protein, its primary role is to act as a chaperone for nucleic acids, remodeling them in order to give the most thermodynamically stable conformations (Levin *et al.*, 2010). The protein is responsible for recruiting two copies of the viral RNA genome into the newly forming Gag complex and hence is required for virion assembly. In the case of HIV the RNA genome amounts to ~10,000 bases per copy. Further functions for NC include the placement of primers such as the annealing of a tRNA molecule (tRNA^{lys3} in the case of HIV-1) onto the primer binding site on the viral RNA, required for the initiation of the reverse transcriptase reaction.

The main structural feature of the HIV NC protein is the presence of two zinc finger domains (also called zinc knuckles) each containing the metal-ion binding motif CX₂CX₄HX₄C. The two domains are interconnected by a flexible basic linker region which is largely disordered in solution. The NMR structure with a bound RNA fragment consisting of the stem-loop s12 of the psi-RNA packaging signal has been determined (Amarasinghe *et al.*, 2000) (Fig. 6).

Linker Peptides

In HIV-1, three additional short peptides are released from Gag following protease processing (SP1, SP2 and p6). SP1 has been shown to have a role in CA assembly whilst the function of SP2 is unclear (Mattei *et al.*, 2016b). The short linker peptide following SP2 and thus positioned at the C-terminus of HIV-1 Gag is p6, which has been shown to be a site for docking various cellular as well as viral binding partners and appears important for the production of infectious viruses. Most of its known functions are suggested to occur under hydrophobic conditions near the cytoplasmic membrane and some formation of helical regions at the N- and C-termini have been demonstrated in micelles but it is otherwise largely unstructured (Solbak *et al.*, 2013).

Other components incorporated into the virion include viral accessory as well as various cellular proteins, small RNA molecules, and specific lipids. The relevance of the presence of these various factors is generally not fully understood (Sundquist and Krausslich, 2012).

Assembly of the Immature Retroviral Virion Leading to Maturation and Release of Virus

The mechanisms underlying the assembly of immature retrovirus virions followed by the transition to the mature form have been intensively studied. Many excellent reviews have been published on the subject particularly focussing on HIV and the extensive primary scientific literature is fully documented therein (Briggs and Krausslich, 2011; Sundquist and Krausslich, 2012; Mattei *et al.*, 2016a,b).

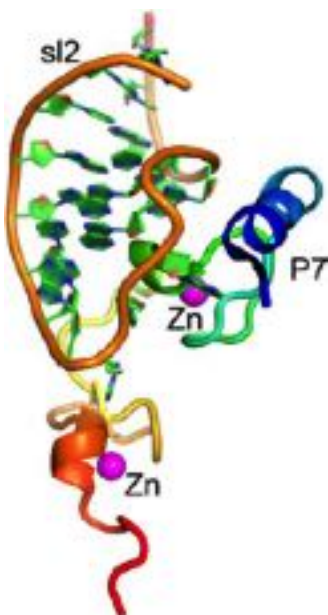


Fig. 6 NMR solution structure of HIV-1 Nucleocapsid protein (p7). The zinc ions of the 2 p7 zinc fingers are shown in purple. The bound RNA is the stem loop SL2 of the psi RNA packaging signal (PDB: 1F6U).

Many structure determination methodologies have been used in studying retrovirus virion proteins. Thus crystallography and NMR have been used for high resolution studies of isolated Gag domains as described above, such structures are useful for fitting into lower resolution maps of the whole virion derived from cryo-electron microscopy (cryo-EM) imaging. Cryo-EM is pivotal for structural studies of both assemblies of CA and the whole virus but has not proved straight forward as for both immature and mature HIV-1 particles since the virions are pleomorphic. Such variability in the immature capsid structure is demonstrated by various spherical forms that contain defects (de Marco *et al.*, 2010). For the mature virion the capsid adopts various morphologies (Fig. 3(a)) (Mattei *et al.*, 2016b). This variability is in contrast to the requirements of single particle cryo-EM where identical particles are picked for averaging. However the development of more powerful microscopes together with improvements in detector technology, data acquisition and processing have contributed to significant progress being made. Cryo electron tomography has been used whereby a number of 2D images collected at differing tilt angles are combined to give a medium resolution image. A further pivotal development is sub tomogram averaging where the resolution of the structure determination can be significantly improved to the sub nanometer level (Mattei *et al.*, 2016b).

The initial assembly of the immature virus is directed by Gag which yields a spherical shell that MA anchors to the viral membrane. The co-translational modification consisting of an N-terminal myristoyl chain in the case of HIV MA, as described previously, is placed in the inner part of the membrane bilayer and the positively charged region binds to the head group of PIP2 (Fig. 3(a)). Membrane budding occurs after the assembly of a curved Gag lattice (Pornillos and Ganser-Pornillos, 2019).

The next step is the auto-activation of the protease, cleaving Gag at the recognition sites (Fig. 1) which in turn leads to the Gag lattice disassembling and the formation of a ribonucleoprotein particle from the newly released NC in complex with the genomic RNA in a form ready for reverse transcription (Pornillos and Ganser-Pornillos, 2019). The cone like capsid is formed (Fig. 3(b)) with the dense NC-RNA particle at one end. The cone of HIV is thought to utilize fullerene type geometry where largely hexameric units require the incorporation of exactly twelve pentamers in order to close the capsid shell (Sundquist and Krausslich, 2012). Other retroviruses have differing core morphologies from HIV. Thus the M-PMV core has a cylindrical shape whilst MLV is more spherical. In each case twelve pentamers are present in the capsid but their relative positions are thought to vary (Mattei *et al.*, 2016b; Zhang *et al.*, 2015).

It can be seen that the immature hexamer has extensive contacts made by α -helices from CTD of CA and SP1 close to the six-fold axis (Fig. 3(c)) (Pornillos and Ganser-Pornillos, 2019; Schur *et al.*, 2015; Mattei *et al.*, 2016b). On transition to the mature form of HIV a significant reorganization of CA occurs such that NTD helices form the main contacts donating 3 helices per CA giving a ring of 18 helices (Fig. 3(d)) (Mattei *et al.*, 2016a). The mature pentamer structure is shown in Fig. 3(e), the contacts formed in hexameric and pentameric oligomers have been shown to be quasi equivalent (Mattei *et al.*, 2016b).

Accompanying the transition from immature to mature HIV virions two potentially significant structural changes that might be determinants for the structural maturation were observed. These changes are (1) a β -hairpin formation at the N-terminus of the CA protein and (2) destabilization of a 6-helix bundle made up from the C-terminus of CA and the adjoining SP1. (Mattei *et al.*, 2016b; Pornillos and Ganser-Pornillos, 2019). It was not clear as to how particular proteolytic cleavages contribute to the alterations in the protein structure. The structures of a series of mutants at the sites of cleavage were determined and further analysis indicated different cleavages can lead to destabilization of the 6-helix bundle which therefore acts as the switch of HIV virions to the infectious form (Mattei *et al.*, 2018).

CA is a potential site for the development of ‘maturation inhibitors’ as anti-HIV drugs. Compounds that bind to the six-helix bundle and prevent it from unwinding have been identified (Purdy *et al.*, 2018; Pornillos and Ganser-Pornillos, 2019).

Structure of ENV

The *env* gene product or “spike protein” is a glycoprotein located on the virion surface that plays an essential role in viral transmission by firstly promoting attachment to cellular surfaces through the recognition of specific receptors. Subsequent entry into the cell is promoted by fusion of the cellular and viral membranes, leading to the release of the RNA genome into the host, thereby giving rise to a further cycle of virus infection.

For HIV-1 the initially synthesized polypeptide, gp160, is organized into a homotrimer. gp160 consists of a number of motifs and domains spanning intracellular and extracellular regions that have different functions (Fig. 7). There are 5 variable loops and 5 conserved loops in the extracellular region which is then followed by the fusion peptide, 2 heptad repeats, the trans membrane region and the cytoplasmic tail.

A further post translational modification occurs when gp160 is subjected to proteolytic cleavage by a cellular enzyme, furin. The resulting fragments, gp120 and gp41 form a trimer of heterodimers that remain associated in a non-covalent manner. gp120 forms the receptor binding site and for this role is positioned on the surface of the virus. gp41, which is responsible for membrane fusion, is more closely associated with the lipid bilayer, possessing a trans membrane domain in addition to the ectodomain (Fig. 7). The attachment of gp120 to target cells is by binding to one domain of the CD4 receptor located on immune cells (e.g., T helper cells, monocytes and macrophages). Additional interactions generally are also made to a co-receptor which is either of two chemokine receptors, CCR5 or CXCR4 (Fig. 8). gp120 is heavily glycosylated with >80 sites being modified (Fig. 7), thus forming a “glycan shield” to help fend off the immune system (Ward and Wilson, 2017).

The envelope protein of HIV-1 is a major viral target for the host immune system. HIV hides from the immune system in a number of ways including the presence of flexible loops or “variable” loops in gp120 that can easily mutate. An important objective in the structural studies of gp160 was the understanding of the basis for the action of “broadly neutralizing antibodies” (which can target gp160 from many different clades) in order to be able to design new immunogens for a novel vaccine.

The availability of detailed three-dimensional structural information on gp160 was considered to be of crucial importance in the understanding of underlying mechanisms which would in turn contribute to vaccine design. Considerable efforts by many

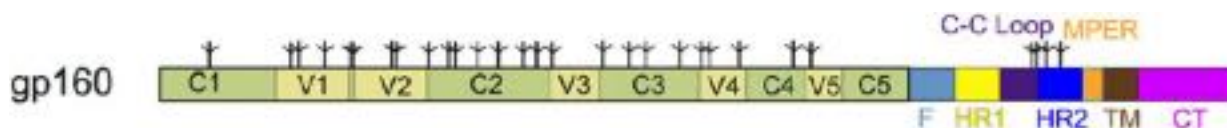


Fig. 7 HIV-1 Env The full-length HIV-1 Env, gp160. Segments of gp120 and gp41 include: C1–C5, conserved regions 1–5; V1–V5, variable regions 1–5; F, fusion peptide; HR1, heptad repeat 1; C–C loop, the immunodominant loop with a conserved disulfide; HR2, heptad repeat 2; MPER, membrane-proximal external region; TM, transmembrane anchor; CT, cytoplasmic tail; tree-like symbols, glycans. Reproduced from Chen, B., 2019. Molecular mechanism of HIV-1 entry. Trends in Microbiology 27, 878–891, with permission

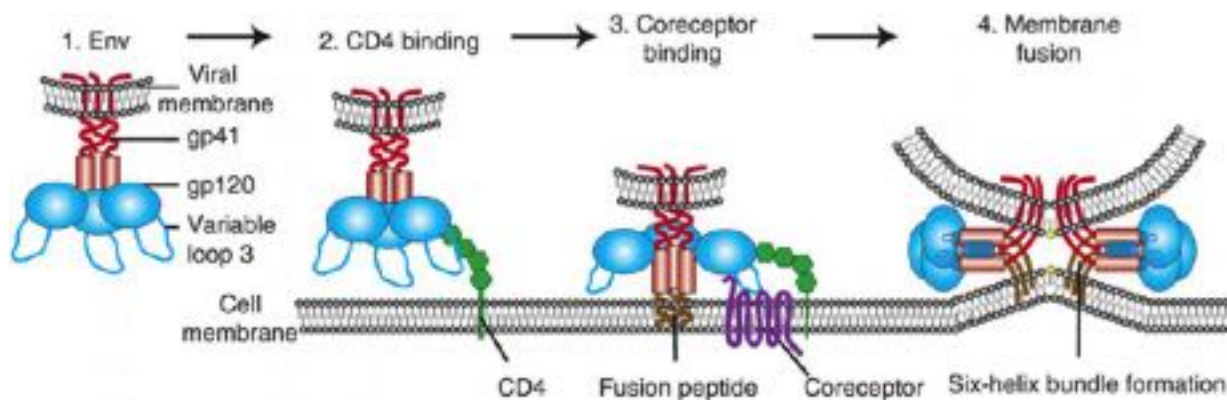


Fig. 8 Overview of HIV entry. To deliver the viral payload into cells, HIV Env, comprised of gp120 and gp41 subunits (1), first attaches to the host cell, binding CD4 (2). This binding causes conformational changes in Env, allowing coreceptor binding, which is mediated in part by the V3 loop of Env (3). This step initiates the membrane fusion process as the fusion peptide of gp41 inserts into the target membrane, followed by six-helix bundle formation and complete membrane fusion (4). Reproduced from Wilen, C.B., Tilton, J.C., Doms, R.W., 2012. HIV: Cell binding and entry. Cold Spring Harbor Perspectives in Medicine 2, with permission.

groups over several years has now yielded very significant progress. In this context the major advances in higher resolution cryo-EM structure determination made over the last decade have greatly contributed to the success of such research efforts.

The first structure determined of a fragment of gp160 was part of the ectodomain of gp41 complexed with a solubilising peptide (Weissenhorn *et al.*, 1997) which showed an all helical structure consisting of an extended, triple-stranded α -helical coiled coil with the amino terminus at its tip. A carboxy-terminal α -helix packs in the reverse direction against the outside of the coiled coil, as seen in other viruses therefore suggesting a common mechanism for initiating fusion (Figs. 8 and 9(a)).

Crystallization of the cleaved gp120/gp41 has proved problematic as the heterodimer is unstable when located outside of the virion. Focussing initially on the gp120 core domain some of the flexible surface loops (V1-V5) were truncated and the heterogeneity of the glycosylation sites reduced in order to give a more homogeneous protein that was more likely to crystallize (Ward and Wilson, 2017). Further efforts included co-crystallization with various receptor fragments as well as with antibody Fabs. A huge effort to survey many conditions lead eventually to the crystallization and structure determination of the core gp120 in

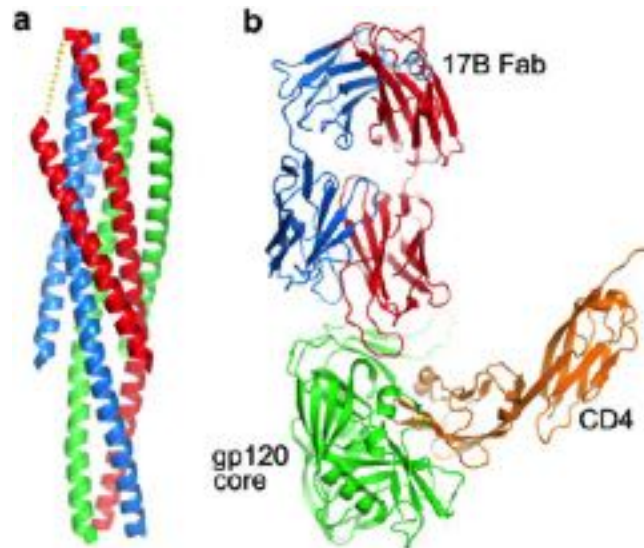


Fig. 9 Structures of HIV gp41 and gp120 domains.(a) Structure of gp41 ecto domain in post fusion state showing helix bundle (PDB:1ENV) (b) gp120 core domain (green) in complex with CD4 soluble domain (orange) and Fab fragment (domains blue and red) (PDB:1GC1).

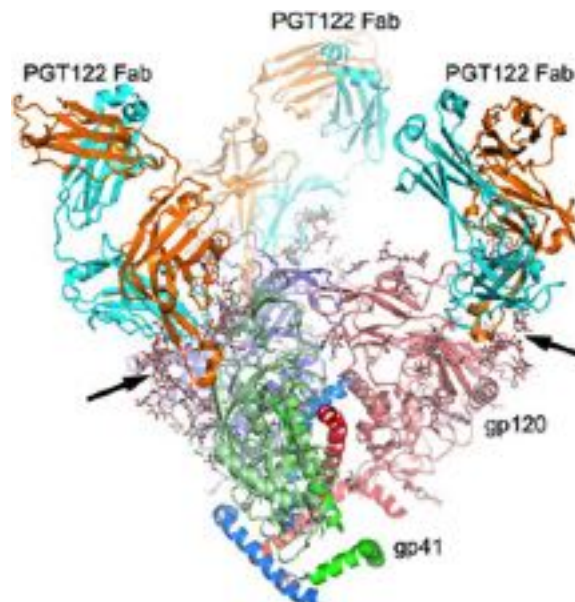


Fig. 10 The structure of a cleaved HIV-1 envelope glycoprotein in complex with a neutralizing Fab PGT122. Chains of the gp41 trimer are shown in red, green and blue, and the corresponding chains in gp120 in salmon, pale green and slate respectively. The glycans on the glycoprotein are drawn as sticks. The Fabs are shown in orange (heavy chains) and cyan (light chains). Arrows indicate the gp120 and Fab interface (PDB : 4NCO).

deglycosylated form, with variable loops truncated and in complex with the soluble domain of its receptor (sCD4) and the Fab of a CD4i neutralizing antibody (Kwong *et al.*, 1998). The structure of the deglycosylated core gp120 is a prolate ellipsoid organized into a novel fold consisting of two major domains containing 25 β -strands and 5 α -helices and 6 disulfide bridges (Fig. 9(b)) (Kwong *et al.*, 1998).

A further major step in obtaining the crystal structure of the trimer of the homodimeric gp120/gp41 complex was the introduction of a disulfide bridge between gp120 and gp41 to hold the subunits together. In conjunction with a single point mutation in the HR1 region of gp41 that improves stable trimer formation, the construct (SOSIP) (Sanders *et al.*, 2013) in a particular stable strain (BG505) together with a suitable Fab fragment and some deglycosylation gave protein that yielded well ordered crystals leading to structure determination at 4.7 Å (Julien *et al.*, 2013) and also a cryo-EM structure at 5.8 Å in which full glycosylation was retained (Lyumkis *et al.*, 2013; Ward and Wilson, 2017). The new structures revealed the protein including the gp41 in some detail together with its interactions with gp120 and the spatial arrangement of V1/V2 and V3 loops (Fig. 10). Of particular interest for vaccine design was the interactions of the Fab of a broadly neutralizing antibody. A long extension from the antibody projects through the coating of sugar chains and into a conserved region of gp160.

Conclusions

Much new information has been derived for the structure of retrovirus virions particularly for HIV. Combining structural results from cryo-EM, crystallography and NMR has shown atomic resolution detail of the Gag domains, Matrix, Capsid and Nucleocapsid, the principle components of the HIV virion. Studies of the immature and mature forms of HIV have revealed the mechanism of the transition between the two states. A future challenge is the utilization of these structural data for developing novel therapies for HIV. Whilst prototype compounds are available for inhibiting the maturation phase there are still many other potential sites on the Gag domains and other virion proteins that can potentially be exploited for drug design (Tedbury and Freed, 2015).

For ENV, extensive protein engineering has allowed crystallization of ENV domains leading to detailed structures in conjunction with cryo-EM studies which should be of value in immunogen design for more effective vaccines against HIV.

References

- Amarasinghe, G.K., De Guzman, R.N., Turner, R.B., *et al.*, 2000. NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition. *Journal of Molecular Biology* 301, 491–511.
- Briggs, J.A., Krausslich, H.G., 2011. The molecular architecture of HIV. *Journal of Molecular Biology* 410, 491–500.
- Chen, B., 2019. Molecular mechanism of HIV-1 entry. *Trends in Microbiology* 27, 878–891.
- Coffin, J.M., 1995. HIV population dynamics in vivo: Implications for genetic variation, pathogenesis, and therapy. *Science* 267, 483–489.
- Conte, M.R., Matthews, S., 1998. Retroviral matrix proteins: A structural perspective. *Virology* 246, 191–198.
- De Marco, A., Muller, B., Glass, B., *et al.*, 2010. Structural analysis of HIV-1 maturation using cryo-electron tomography. *PLOS Pathogens* 6, e1001215.
- Dodonova, S.O., Prinz, S., Bilanchone, V., Sandmeyer, S., Briggs, J.A.G., 2019. Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. *Proceedings of the National Academy of Sciences of the United States of America* 116, 10048–10057.
- Gres, A.T., Kirby, K.A., Kewalramani, V.N., *et al.*, 2015. X-ray crystal structures of native HIV-1 capsid protein reveal conformational variability. *Science* 349, 99–103.
- Hill, C.P., Worthylake, D., Bancroft, D.P., Christensen, A.M., Sundquist, W.I., 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: implications for membrane association and assembly. *Proceedings of the National Academy of Sciences of the United States of America* 93, 3099–3104.
- Julien, J.P., Cupo, A., Sok, D., *et al.*, 2013. Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 342, 1477–1483.
- Kwong, P.D., Wyatt, R., Robinson, J., *et al.*, 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 393, 648–659.
- Levin, J.G., Mitra, M., Mascarenhas, A., Musier-Forsyth, K., 2010. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biology* 7, 754–774.
- Lingappa, J.R., Reed, J.C., Tanaka, M., Chutiraka, K., Robinson, B.A., 2014. How HIV-1 Gag assembles in cells: Putting together pieces of the puzzle. *Virus Research* 193, 89–107.
- Lyumkis, D., Julien, J.P., De Val, N., *et al.*, 2013. Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* 342, 1484–1490.
- Mattei, S., Glass, B., Hagen, W.J., Krausslich, H.G., Briggs, J.A., 2016a. The structure and flexibility of conical HIV-1 capsids determined within intact virions. *Science* 354, 1434–1437.
- Mattei, S., Schur, F.K., Briggs, J.A., 2016b. Retrovirus maturation—an extraordinary structural transformation. *Current Opinion in Virology* 18, 27–35.
- Mattei, S., Tan, A., Glass, B., *et al.*, 2018. High-resolution structures of HIV-1 Gag cleavage mutants determine structural switch for virus maturation. *Proceedings of the National Academy of Sciences of the United States of America* 115, E9401–E9410.
- Murakami, T., 2008. Roles of the interactions between Env and Gag proteins in the HIV-1 replication cycle. *Microbiology and Immunology* 52, 287–295.
- Perilla, J.R., Gronenborn, A.M., 2016. Molecular architecture of the retroviral capsid. *Trends in Biochemical Sciences* 41, 410–420.
- Pornillos, O., Ganser-Pornillos, B.K., 2019. Maturation of retroviruses. *Current Opinion in Virology* 36, 47–55.
- Pornillos, O., Ganser-Pornillos, B.K., Yeager, M., 2011. Atomic-level modelling of the HIV capsid. *Nature* 469, 424–427.
- Purdy, M.D., Shi, D., Chrustowicz, J., *et al.*, 2018. MicroED structures of HIV-1 Gag CTD-SP1 reveal binding interactions with the maturation inhibitor bevirimat. *Proceedings of the National Academy of Sciences of the United States of America* 115, 13258–13263.
- Rao, Z., Belyaev, A.S., Fry, E., *et al.*, 1995. Crystal structure of SIV matrix antigen and implications for virus assembly. *Nature* 378, 743–747.
- Sanders, R.W., Derking, R., Cupo, A., *et al.*, 2013. A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLOS Pathogens* 9, e1003618.
- Schur, F.K., Hagen, W.J., Rumiova, M., *et al.*, 2015. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. *Nature* 517, 505–508.
- Solbak, S.M., Reksten, T.R., Hahn, F., *et al.*, 2013. HIV-1 p6 – A structured to flexible multifunctional membrane-interacting protein. *Biochimica et Biophysica Acta* 1828, 816–823.

- Sundquist, W.I., Krausslich, H.G., 2012. HIV-1 assembly, budding, and maturation. Cold Spring Harbor Perspectives in Medicine 2, a006924.
- Tedbury, P.R., Freed, E.O., 2015. HIV-1 gag: An emerging target for antiretroviral therapy. Current Topics in Microbiology and Immunology 389, 171–201.
- Tedbury, P.R., Novikova, M., Ablan, S.D., Freed, E.O., 2016. Biochemical evidence of a role for matrix trimerization in HIV-1 envelope glycoprotein incorporation. Proceedings of the National Academy of Sciences of the United States of America 113, E182–E190.
- Vogt, V.M., 1997. Retroviral virions and genomes. In: Coffin, J.H., Hughes, S.H., Varmus, H.E. (Eds.), Retroviruses. New York: Cold Spring Harbor Laboratory Press.
- Ward, A.B., Wilson, I.A., 2017. The HIV-1 envelope glycoprotein structure: Nailing down a moving target. Immunological Reviews 275, 21–32.
- Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., Wiley, D.C., 1997. Atomic structure of the ectodomain from HIV-1 gp41. Nature 387, 426–430.
- Wilen, C.B., Tilton, J.C., Doms, R.W., 2012. HIV: Cell binding and entry. Cold Spring Harbor Perspectives in Medicine 2.
- Zhang, W., Cao, S., Martin, J.L., Mueller, J.D., Mansky, L.M., 2015. Morphology and ultrastructure of retrovirus particles. AIMS Biophysics 2, 343–369.

Structure of Helical Viruses

C Martin Lawrence, Montana State University, Bozeman, MT, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AFV1 Acidianus Filamentous Virus 1
APBV1 *Aeropyrum pernix* Bacilliform Virus 1
ATSV *Acidianus* Tailed Spindle Virus
BaMV Bamboo Mosaic Virus
CGMMV Cucumber Green Mottle Mosaic Virus
HLSV Hibiscus Latent Singapore Virus

ORSV *Odontoglossum* Ringspot Virus
PepMV Pepino Mosaic Virus
RMV Ribgrass Mosaic Virus
SFV1 *Sulfolobus* Filamentous Virus 1
SIRV2 *Sulfolobus islandicus* Rod-shaped Virus 2
TMGMV Tobacco Mild Green Mosaic Virus
TMV Tobacco Mosaic Virus

Glossary

4-Helix bundle A family of protein folds consisting of 4 α -helices oriented with the long axes of the helices roughly parallel to each other. The direction of the helices with respect to each other (parallel versus antiparallel) and the order in which the polypeptide connects them determines the specific type, or fold of 4-helix bundle.

Capsid The protein coat that encloses the viral genome.

Cryoelectron microscopy A method for visualizing biological structures by suspending unstained objects in a thin sheet of non-crystalline (vitreous) ice. Computer assisted averaging is utilized to reconstruct 3-dimensional images of the object. In favorable cases, the technique can reach high enough resolution to allow de novo modeling of protein and nucleic acid structure.

Envelope A lipid membrane, usually derived from the cellular membrane, with inserted viral glycoproteins. Envelopes are present in many viruses as an external layer

wrapped around the nucleocapsid, though some viruses may instead utilize an internal lipid envelope. The envelope is usually a lipid bilayer, but in hyperthermophiles, membranes are frequently a monolayer of cyclic lipids, with head groups on both faces of the membrane.

Helical symmetry Symmetry defined by the combination of rotational and translational displacement about a single helix axis.

Icosahedral symmetry Symmetry of a regular icosahedron, with 12 axes of 5-fold symmetry, 20 axes of 3-fold symmetry, and 30 axes with 2-fold symmetry.

Nucleocapsid Viral capsid containing the viral nucleic acid genome.

Virion A complete, infectious virus particle.

X-ray diffraction The diffraction of x-rays by oriented molecular assemblies, usually in crystals but sometimes in fibers or “sols”, used to determine the atomic structure of the object.

Introduction

The proteins that form the shells, or capsids, that encase viral nucleic acid are involved in a variety of functions, including viral assembly, cellular egress, extracellular genome protection, host recognition, entry and disassembly. Due to the need for genetic economy, these capsids are constructed of smaller protein subunits, or building blocks, that assemble through a well-defined set of interactions that are repeated over and over again to construct these beautiful, complex containers. Whenever a repeated set of interactions are used to assemble a larger structure, whether a viral capsid or some other object, the result is always symmetric and this symmetry can be either open or closed. In the case of closed symmetry, the subunit interactions will result in an assembly with a defined number of subunits arranged about a single point (point group symmetry). In contrast, open symmetry includes a translational component that instead results in an extended helical assembly.

Interestingly, nature exploits each of these principles to package viral nucleic acid, and thus most viral nucleocapsids are either closed spherical shells with icosahedral symmetry or long, filamentous tubes with helical symmetry. Among the helical viruses, some are rigid rod-shaped virions while others are flexible filamentous assemblies, and the packaged genome can be either RNA or DNA. In addition, for some helical viruses, their helical nucleocapsid assemblies are enclosed within a lipid membrane envelope. A potential advantage of the nucleoprotein helix is its ability, in principle, to package a genome of any size. However, unlike icosahedral viruses, helical viruses are recalcitrant to crystallization. Thus, high resolution structural studies of intact helical virions have historically relied on X-ray fiber diffraction and, increasingly, on electron microscopy.

Rod-Shaped Helical Plant Viruses

Helical viruses are found in each of the 3 domains of life, Eukarya, Bacteria and Archaea. Among the Eukarya, filamentous non-enveloped positive strand RNA viruses are particularly common in plants, both rigid (*virgaviridae*), and flexible (*potyviridae*,

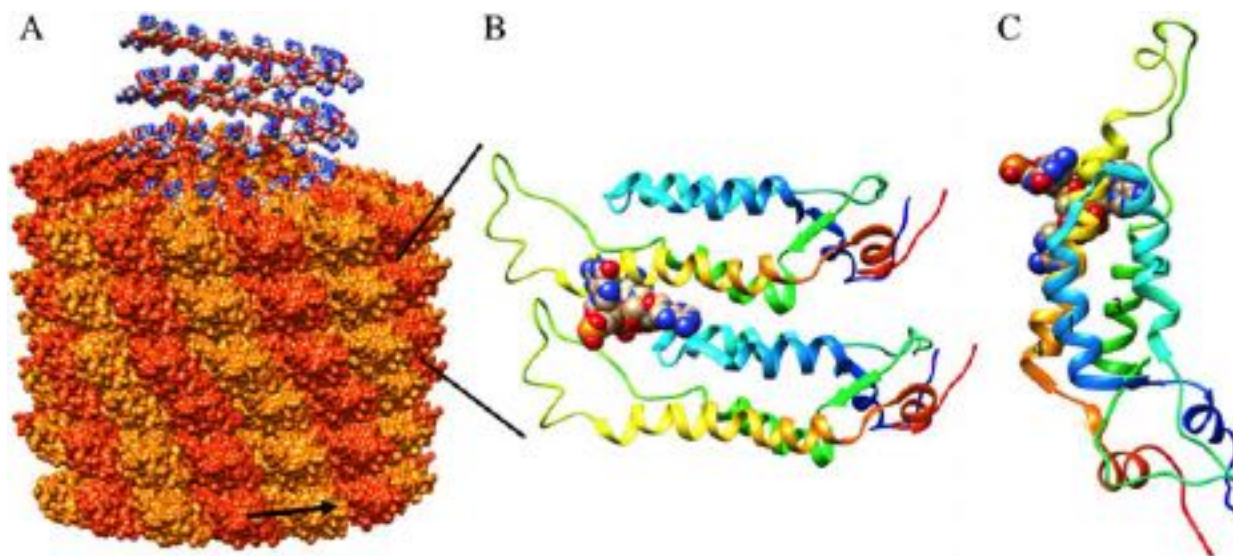


Fig. 1 Tobacco Mosaic Virus (TMV, PDB ID 2TMV) (A) A short section of the TMV single-start, right handed helical assembly is shown. Subunits are colored in alternating shades of orange and pale-orange, and the right-handed or clockwise twist is indicated by the black arrow at the bottom of the assembly. Subunits at the top of the assembly are not shown in order to reveal the helical path of the encapsidated RNA, with the non-hydrogen atoms shown as spheres (C: beige, N: blue, O: red, P: orange). (B) Two TMV subunits stacked one on top of the other from successive strands of TMV, equivalent to an orange/orange subunit interaction in panel A. Each subunit is depicted in rainbow colors with the N-terminal residues in blue, the C-terminal residues in red, and intervening residues a gradient over the blue, cyan, green, yellow, orange and red gradient. The RNA is bound at the subunit interface towards the interior of the capsid. (C) The TMV subunit is a decorated, right-handed up and down 4-helix bundle. The 4 α -helices forming the 4-helix bundle are numbered and shown in light blue, cyan, green and yellow/orange, respectively. The 1st α -helix in light blue is at the front left side, running from bottom to top. When the polypeptide reaches the top of the 1st α -helix, it then makes a "right turn" into the 2nd α -helix (cyan), which runs in the opposite direction. Thus, the "right-handed" 4-helix bundle designation. The 2nd α -helix then connects via a long loop into the 3rd α -helix (green, back right), running bottom to top, and then into the 4th α -helix (yellow/orange) running top to bottom.

alphaflexiviridae, *betaflexiviridae*, *closteroviridae*). A prominent example is tobacco mosaic virus, or TMV, a member of the *virgaviridae*. Historically, TMV has been at the forefront of structural studies of helical viruses, and the subject of X-ray fiber diffraction studies as early as 1941. Nearly 50 years later, following a host of methodological developments and low to medium resolution structures, the structure of TMV was finally determined at 2.9 Å resolution (PDB ID 2TMV).

The earlier, lower resolution studies revealed a TMV nucleocapsid approximately 300 nm in length and 18 nm in diameter, with a hollow core roughly 4 nm in diameter running along the length of the virion. Overall, the nucleocapsid is composed of approximately 2130 identical coat protein subunits arranged in a right-handed helical assembly. A single RNA strand is nestled within a groove that follows the viral helix at a radius of 4 nm, with each protein subunit bound to three nucleotides of RNA (Fig. 1). Helical structures are characterized by their pitch (P), which is the distance along the helix axis corresponding to one full turn of helix. The pitch is equal to the number of subunits per turn of helix (u) multiplied by the relative translation or displacement of each subunit along the helical axis (p). Specifically, $P = u \times p$. For the TMV nucleocapsid, there are 16.33 subunits per turn of helix (49 in 3 turns), with each subunit displaced by 0.14 nm along the helical axis, giving a pitch of 2.3 nm (Fig. 1).

The 154 residue (17.5 kDa) TMV subunit is a right handed, antiparallel 4-helix bundle. As one follows the first helix up the fold, there is a "right turn" into the second helix (Fig. 1(C) and Table 1). In the literature, these 4 helices are designated as the LS, RS, RR and LR helices, respectively. Within the viral capsid, the long axis of the 4-helix bundle is roughly perpendicular to the long axis of the TMV capsid (Fig. 1), such that the extended loop connecting helices 3 and 4 (LR/RR) projects into the interior of the capsid. In contrast, the exterior end of the bundle is decorated by a small mixed β -sheet and 2 small α -helices, with the N- and C-termini exposed on the exterior of the capsid. There are extensive subunit interactions, both laterally to neighboring subunits within the same helical turn, and to subunits in the preceding and subsequent helical turns as well.

The RNA follows a helical path at a radius of 4 nm, running along a groove between successive helical turns of the coat protein. Thus, each 3 base segment is sandwiched between two subunits. The 1st and 3rd bases are in an *anti* conformation, with the second base in the less common *syn* conformation. The bases pack against hydrophobic side chains "underneath" helix 4 (LR) of the "upper" subunit, and the *syn* base and one of the *anti* bases stack together in a cavity between 2 inter-subunit salt bridges. For the most part, the interactions are not base specific, but there is a preference for a guanine at position 1. The charge on the phosphate groups is offset in part by Arg residues in the coat protein, as well as a Ca^{2+} ion interdigitated between 2 successive phosphate moieties and Asp116.

Table 1 High resolution structures of helical viruses

<i>Viral Family</i>	<i>Viruses</i>	<i>Genome</i>	<i>Domain</i>	<i>Rigid/Flexible</i>	<i>Twist</i>	<i>Capped</i>	<i>Enveloped</i>	<i>Subunit</i>
<i>Virgaviridae</i>	TMV TMGMCGMMV RMV ORSV	+ RNA	Eukarya (Plants)	Rigid Rod	Right-Handed, Single-Start Helical Assembly	No	No	Decorated, Right- handed 4-helix Bundle
<i>Alphaflexiviridae</i>	PepMV BaMV	+ RNA	Eukarya (Plants)	Flexible Filament	Left-Handed, Single-Start Helical Assembly	No	No	Multi-helix Bundle, With N- and C-Terminal Arms
<i>Filoviridae</i>	Ebola Virus	– RNA	Eukarya (Primates)	Flexible Filament	Left-Handed, Single Start	Yes	Yes	2-domain Multi- helix Bundles
<i>Inoviridae</i>	Fd Ike and Pf1	Single-stranded DNA	Bacteria	Semi-flexible Filament	Right-Handed, 5-Start and 6-Start Helical Assemblies	Yes	No	Single Helix
<i>Rudiviridae</i>	SIRV2	Linear Double- stranded DNA (A-form)	Archaea	Rigid Rod	Right-Handed, Single-Start Assembly	Yes	No	Homo-dimer, Decorated, Left-handed 4-helix Bundle
<i>Lipothrixviridae</i>	AFV1 SFV1		Archaea	Flexible Filament	Right-Handed Single-Start Assembly	Yes	Yes	Hetero-dimer, Decorated, Left-handed 4-helix Bundle
<i>Claviviridae</i>	APBV1	Circular Double- stranded DNA	Archaea	Rigid Rod	Left-Handed, 5-Start Assembly	Yes	No	Helix-Turn-Helix
<i>Bicaudaviridae</i>	ATSV	Circular Double- stranded DNA	Archaea	Lemon-shaped Capsid/Rigid Tail	Left Handed Multi-Start	Yes	No	Decorated, Right-handed 4-helix Bundle

The Structures of 5 additional tobamoviruses (TMGMV, CGMMV, RMV, ORSV, HLSV) have also been determined by X-ray fiber diffraction. These viruses share the same helical symmetry (16.33 subunits per turn of helix), and their respective coat proteins show sequence identity to TMV ranging from 36% to 72%. Thus, it is not surprising that they show the same overall fold. Notably, each of these structures show conservation of the Ca^{2+} mediated Asp116/phosphate interaction. Outside of these tobamoviruses, high resolution structures for other rod-shaped plant viruses are lacking. They generally exhibit helical symmetry, though the helical parameters do differ. It is expected that they will utilize subunits based on a core 4-helix bundle, but sequence differences and differences in the size of the coat proteins make it clear that there is still much to be learned.

Flexible, Filamentous Plant Viruses

While TMV and other virgaviridae are stiff, rod-shaped helical assemblies, the capsids of many plant viruses are flexible, filamentous assemblies. These include the *alphaflexiviridae*, *betaflexiviridae*, *potyviridae* and *closteroviridae*. These flexible nucleocapsids also show helical symmetry, though the helical parameters differ from the tobamoviruses. For example, Potato virus X (PVX) has 8.9 subunits per turn and a pitch of 34.5 Å. Similarly, the coat proteins show high helical content, and it was expected for some time the N- and C-termini were on the outer surface of the nucleocapsids, and that their structures would be built around a core 4-helix bundle. One difference, however, is the incorporation of additional proteins at the 5'-end of the RNA. The potyvirus protein VPg is covalently linked to the 5' end of the viral RNA and is therefore present in the virion. And in the *closteroviridae*, both a major and a minor capsid protein are utilized. The minor capsid protein coats a short segment of the genome (less than 5%), at the 5' end of the RNA, giving these viruses a "rattlesnake" like appearance.

While the earliest studies of helical virus structure relied on X-ray fiber diffraction, more recently electron cryo-microscopy (cryo-EM) has seen enormous advances in instrumentation and software that allow helical virus reconstruction by single particle analysis (SPA) or iterative helical reconstruction (IHR). These techniques have enabled structural studies of helical viruses, including flexible filamentous viruses, and are now the predominant method used for high-resolution structural studies. IHR has been employed to determine two higher resolution structures for members of the *alphaflexiviridae*, Bamboo Mosaic Virus (BaMV) and Pepino Mosaic Virus (PepMV), at resolutions of 5.6 and 3.9 Å respectively. In each case the inner and outer diameter of the capsid, the number of subunits per turn and the pitch are nearly identical to PVX (above), but the high-resolution data surprisingly revealed left-handed, rather than right-handed helical assemblies (Fig. 2 and Table 1). In addition, while the individual coat protein subunits did indeed reveal high helical content, their structures are not antiparallel, up and down 4-helix bundles like those seen for TMV. Instead, the subunit shows a large N-terminal helical bundle with uncommon 3D topology, and a small C-terminal helical lobe (Fig. 2). Importantly, the structures also reveal extended N- and C-terminal arms on the exterior and

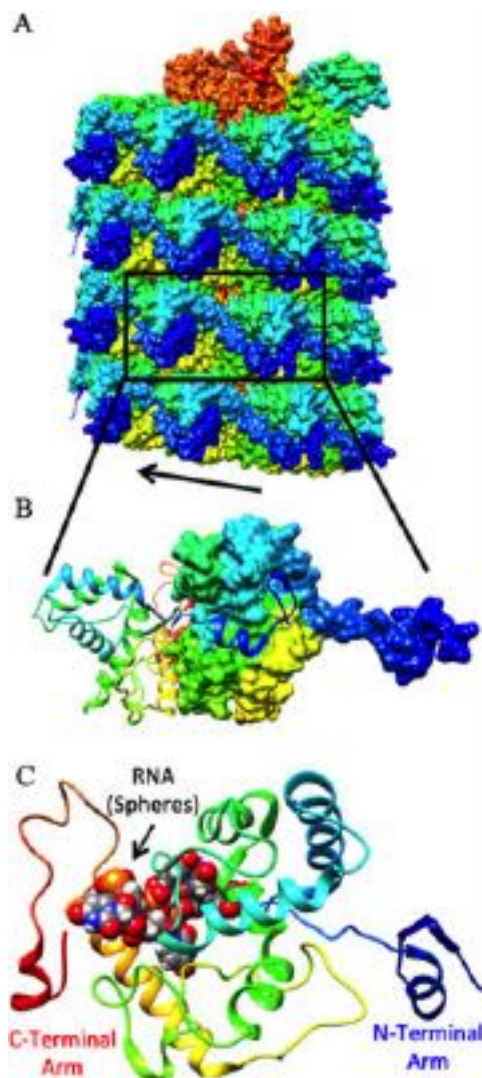


Fig. 2 Pepino Mosaic Virus (PepMV, PDB ID 5FN1) (A) A short section, ~ 4 -turns, of the PepMV single-start, left-handed helical nucleocapsid. Each subunit is colored identically as a rainbow gradient with the N-terminus in blue and the C-terminus in red. The left-handed or counter-clockwise twist is indicated by the black arrow at the bottom of the assembly. The N-terminal extensions (dark blue) on the outside of the virus reach laterally to interact with the preceding subunit. The C-terminal extensions (red) on the interior surface of the capsid protrude from the top of the structure, where they interact axially with the next helical turn of the assembly (not shown). (B) An enlarged view of two adjacent capsid proteins. Each subunit again colored with a rainbow gradient. The left subunit (i) depicted as a ribbon, the preceding subunit on the right ($i-1$) as a rendered surface. The N-terminal arm of subunit i extends laterally to interact with a hydrophobic cleft on subunit $i-1$. (C) A single capsid protein subunit, a 2-lobed multi-helix bundle. The orientation is rotated $\sim 40^\circ$ about the vertical axis to show the RNA accommodated between the major (N-terminal) and minor (C-terminal, yellow and red helices) lobes.

interior of the nucleocapsid, respectively. The cleft between the two lobes of the core structure is utilized to accommodate 5 bases per subunit of the + sense RNA genome, which mirrors the helical symmetry of the capsid.

A key question is how does the capsid form a stable helical assembly while at the same time allowing for flexibility. In this regard, the structures show reduced lateral and vertical contacts between the helical bundles, but at the same time the extended N- and C-terminal arms make substantial interactions with neighboring subunits. Specifically, the N-terminal arm makes hydrophobic interactions with the preceding subunit ($n-1$), providing the main lateral interactions between subunits. In contrast the C-terminal arm builds structure along the interior of the capsid, first interacting with 3 subunits in the turn above ($i+7$, $i+8$, $i+9$), before pointing toward the center of the virion where the extreme C-terminal residues contact subunit $i-7$ in the turn below, to build the axial interactions that support helix formation. The flexible linkers present in the coat protein between the helical core and the N- and C-terminal extensions then allow relative movements between the coat proteins, explaining the flexible, yet stable nature of this helical virion. Surprisingly, it was recognized that the structure of the PepMV coat protein showed similarity to the nucleoprotein in the *phleboviridae*, suggesting horizontal gene transfer between these evolutionarily distant groups of eukaryotic RNA viruses.

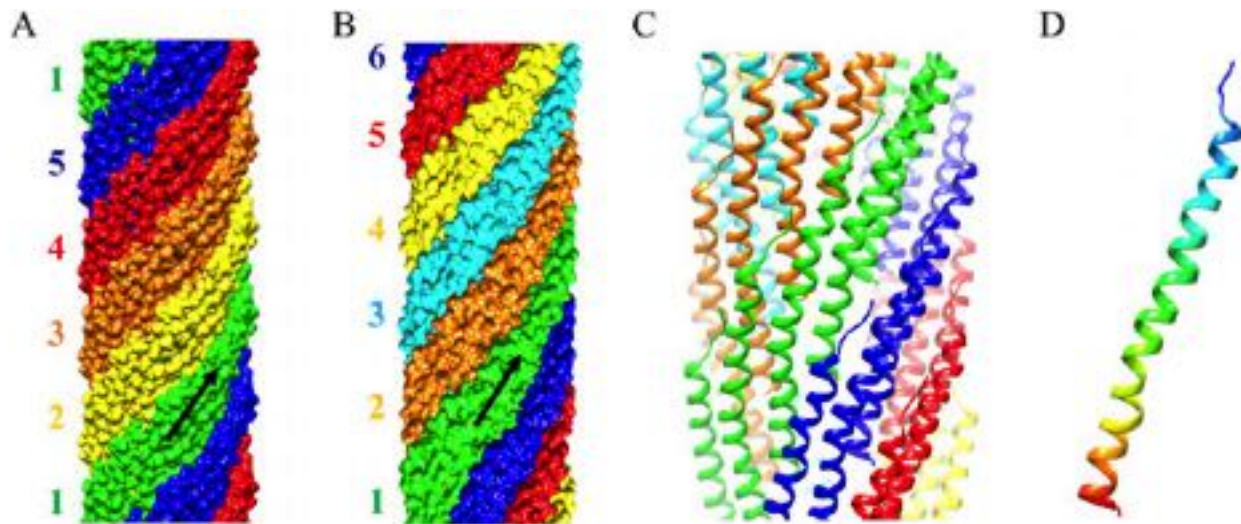


Fig. 3 Bacteriophage fd (PDB ID 2COW) and Pf1 (PDBID 1QL2). (A) A short section of the fd, right-handed (black arrow) 5-start helical assembly is shown. The successive helicoid components are numbered and colored in green, yellow, orange, red and blue, respectively. Filamentous bacteriophage f1 and M13 show similar architectures. (B) A short section of the Pf1 capsid, illustrating its right handed (black arrow) 6-start helical architecture with successive helicoids colored in green, orange, cyan, yellow, red and blue. (C) An enlarged view of the base of the Pf1 capsid rendered to show secondary structural elements and the orientation of the subunits with the capsid. Each subunit is a single α -helix of ~ 50 amino acid residues. (D) An enlarged view of a single Pf1 subunit, its N-terminus is at the top (blue), and the C-terminus at the bottom in red. Within the capsid, the N-terminus is exposed on the outside surface of the nucleocapsid, while basic residues at the C-terminus are on the interior where they presumably neutralize the negative charge on the packaged RNA.

Filamentous Bacteriophage

X-ray fiber diffraction, solid state NMR and low-resolution electron microscopy have also been used to derive models for filamentous bacteriophage. These viruses package circular, single stranded DNA genomes within a semi-flexible nucleocapsid 6 nm in diameter and 800–2000 nm in length. Like the filamentous plant viruses, they also utilize minor structural proteins to cap the ends of the virion, with 3–5 copies of p3 and p6 at one end, and p7 and p9 at the other (Table 1). The major capsid protein is small, only about 50 amino acids (6 kDa). Solid state NMR and X-ray fiber diffraction each suggest it forms a simple, slightly curved alpha-helix (Fig. 3). The subunits form an overlapping, interdigitated helical array, utilizing hydrophobic interactions between apolar regions in the central portion of the helix. The negatively charged N-terminal region is on the outer surface and the C-terminal region is at the inner surface, where positively charged (basic) residues help neutralize the negative charge on the DNA. While the DNA has not been visualized at high resolution, it runs through the center of the helical array, and apparently does not adopt the helical symmetry of the capsid. The helical capsid thus acts as a protective sheath around the circular, single stranded DNA genome.

Two different, but related symmetry classes of filamentous bacteriophage have been identified. Phage fd, along with f1 and M13, are specific for bacteria carrying F-pili, and are thus known as the Ff or F-specific filamentous phage group, and belong to class I. The Pf1 and Pf3 strains belong to class II. The repeating unit for fd (Class I) consists of 10 subunits arranged in two interdigitated rings of five subunits each. Subunits within a ring do not contact each other, instead they interact with subunits in the second ring. The two rings are related to each other by a translation of 1.6 nm along the viral axis, and a rotation of 36° . Alternatively, one can begin with a single subunit and consider a set of repeating lateral interactions as one moves from subunit to subunit. When a single such subset is examined, it is seen that they form an open helicoid structure. In the case of fd, 5 such helicoid structures are used to complete the intact virion. Thus, rather than a single helical array as seen for the tobamoviruses, fd is a “5-start” helical assembly (Fig. 3). Similarly, the repeating subunit for Pf1 (Class II) is described as 27 subunits in 5 turns of helix with 5.4 subunits per helical turn, but can also be visualized as a 6-start helix (Fig. 3). While not discussed here, bacteriophage tail sheaths have also been described as 6-start helical assemblies. Ready examples of multistart helices outside the arena of protein quaternary structure include the right-handed threads on glass bottles and plastic beverage containers.

Hemorrhagic Filoviridae

While helical viruses are common in plants, only one family of filamentous animal viruses is known, the *filoviridae* (Table 1). Examples are Ebola virus and the related Marburg virus, which cause hemorrhagic fever with high fatality rates in humans and non-human primates. Ebola is an enveloped virus with a linear, negative sense, single stranded RNA genome of approximately 19,000 bases, that encodes 7 proteins, a nucleoprotein (NP), along with a glycoprotein (GP), a polymerase (L) and several

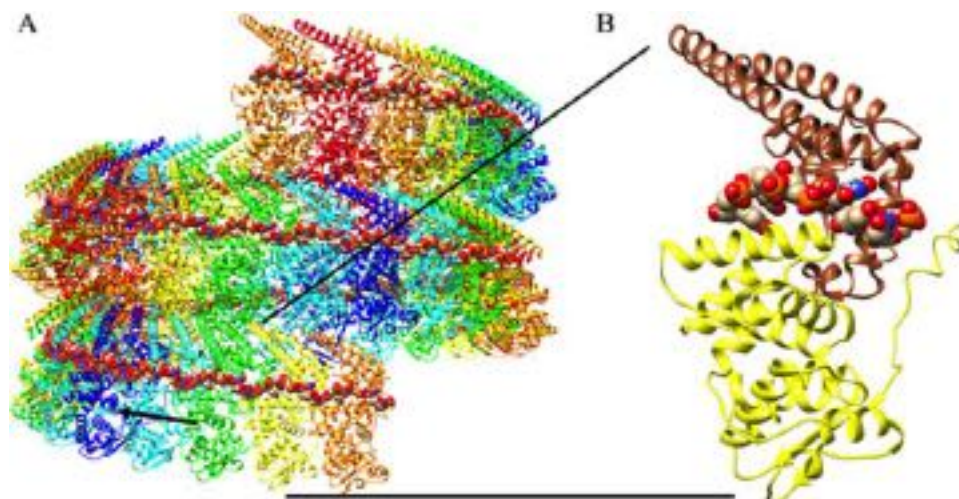


Fig. 4 Ebola Virus (EBV) Nucleocapsid (PDB ID 5Z9W). (A) A short section, ~ 2 -turns, of the EBV single-start, left handed helical nucleocapsid assembly is shown. Subunits are colored in a gradient of colors, and the left-handed or counter clockwise twist is indicated by the black arrow at the bottom of the assembly. The outer lipid envelope is not shown. The DNA, again rendered as spheres, is colored as in Fig. 1. (B) An enlarged view of the yellow subunit at the bottom of panel A. The EBV subunit is composed of two domains (yellow and brown), and each is a large α -helical bundle (see text). The DNA is bound between the two domains, towards the outside edge of the nucleocapsid.

“minor” structural proteins (VP24, VP30, VP35, VP40). The nucleoprotein condenses with the RNA to form a helical nucleocapsid, which forms the core of the virion. The outer viral envelope is assembled as the nucleocapsid buds through the host cell membrane into which the glycoprotein “spikes” have been inserted. VP40 and VP24 are “matrix” proteins found in the space between the envelope and the nucleocapsid, VP30 is a minor nucleoprotein, while VP35 (a chaperone to keep NP in an RNA free state) and the polymerase are present at the 3′ end of the viral RNA.

The structure of the Ebola virus NP/RNA nucleocapsid has been determined using cryo-EM single particle analysis. It assembles as a left-handed, single start helix with a pitch of 73.56 Å and 24.44 subunits per turn (Fig. 4). It has inner and outer diameters of 175 and 295 Å, respectively, and an overall length of 950 μm to accommodate a single copy of the full RNA genome.

The Ebola virus nucleoprotein, composed of 739 amino acids, is significantly larger than the capsid proteins in the helical viruses described above; its structure was first determined in the RNA free state by X-ray crystallography. It is a 2-domain protein comprised of N- and C-terminal lobes, plus an N-terminal arm and C-terminal tail. The fold of the N-terminal lobe is not easily described in detail, but is a bundle of 12 α -helices with a few minor β -strands. The C-terminal lobe, also a helical bundle, is comprised of 6 α -helices and a critical β -finger ($\beta 3/\beta 4$). The cryo-EM structure shows clear density for the RNA, which is firmly secured within a positively charged ellipsoidal cleft between the N- and C-terminal lobes. Binding of the RNA results in conformational changes in the NP, giving rise to the RNA bound “closed” conformation. The conformational change is not just a simple hinge motion between the 2 domains, but includes rearrangement of the secondary structural elements within the C-terminal domain, including substantial rotations of the α -helices and a 90° rotation of the β -finger. The RNA follows an undulating path along the nucleoprotein strand with 6 nucleotides per subunit. These nucleotides are present in a “3-bases-inward, 3-bases-outward” configuration, in which the first 3 bases point towards the exterior of the nucleocapsid, and the next three towards the interior (Fig. 4). Though they are not considered as helical viruses, this is similar to the nucleoprotein interactions seen for measles virus and parainfluenza virus.

Lateral interactions between subunits include hydrophobic interactions between the parallel $\alpha 17$ and $\alpha 18$ helices, giving rise to a “zipper” like arrangement, as well as an extensive role for the N-terminal arm on the inside of the nucleocapsid (Fig. 4). In contrast, the inter-strand interface is primarily electrostatic, which may explain why the nucleoprotein-RNA complex can easily uncoil, yet remain stable in solution.

Viruses of the Archaea, *Rudoviridae* and *Lipothirixviridae*

Helical viruses are common in the Archaeal, or 3rd domain of life, especially among thermophiles (Table 1). Thus, many of the structurally characterized archaeal viruses replicate in high temperature ($> 80^\circ\text{C}$), acidic ($\text{pH} < 4$) environments, and the virion morphologies and dynamics should be considered in the context of these unusual environments. One of the best characterized viral families in the Archaea are the *Rudoviridae*, where the structure of *Sulfolobus islandicus* Rod-shaped Virus 2 (SIRV2) has been determined using cryo-EM iterative helical reconstruction.

SIRV2 is a non-enveloped, rod-shaped virus with a linear, double stranded DNA genome. The nucleoprotein complex forms a single start, right handed helical assembly with a pitch of 42 Å and 14.67 subunits per turn (Fig. 5). In contrast to tobacco mosaic

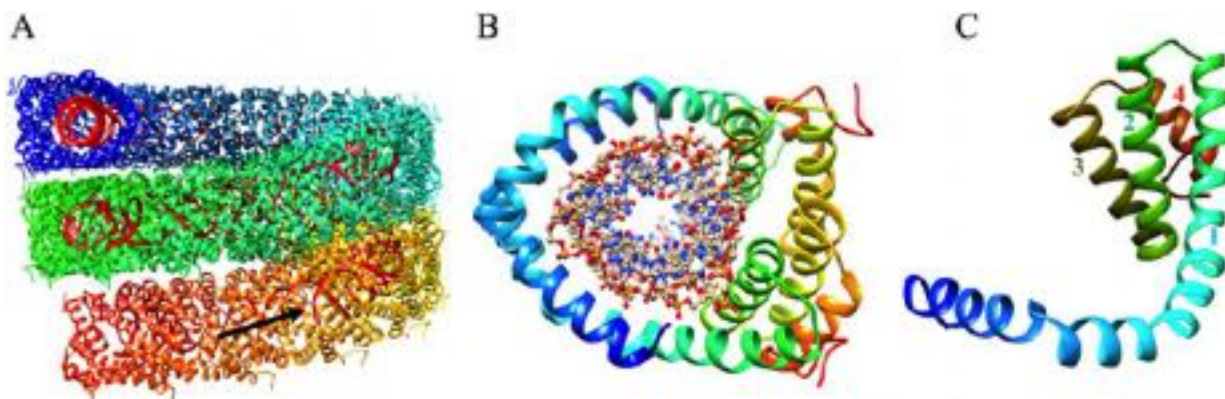


Fig. 5 *Sulfolobus islandicus* Rod Shaped Virus 2 (SIRV2, PDB ID 3J9X) (A) A short section, ~ 2 -turns, of the SIRV2 single-start, right-handed helical nucleocapsid. The subunits are colored in a gradient of colors from orange to red, and the right-handed or clockwise twist is indicated by the black arrow at the bottom of the assembly. The A-form, double stranded DNA is rendered as red ribbons. The blue subunits at the top nicely illustrate how thoroughly the DNA is encased by the protein subunits. (B) An enlarged view of a major capsid protein homodimer. Each subunit is colored with a rainbow gradient, blue at the N-terminus, red at the C-terminus. The core 4-helix bundles are on the right side of the DNA, on the exterior of the capsid. The long, kinked N-terminal α -helices (blue, light blue, cyan) arc around the DNA to dehydrate and encase it. The A-form DNA is shown as balls and sticks, and colored as in Fig. 1. (C) A single subunit of the capsid protein colored as in panel B. The 1st α -helix of the 4-helix bundle is on the front right side. It runs up the page, and at the top of the α -helix, makes a left-hand turn into the 2nd α -helix, and is thus a left-handed 4-helix bundle. The polypeptide chain then runs into α -helices 3 and 4, and each α -helix runs antiparallel to both the preceding and following α -helices.

virus, which utilizes a right handed 4-helix bundle, the major capsid protein in SIRV2 adopts a left handed 4-helix bundle. In addition, helix 1 shows a long, kinked, N-terminal helical extension, which can be described as a helix-kink-helix motif, with a curved appearance (Fig. 5). This extension is disordered in the monomeric form of the protein, becoming ordered only when incorporated into the virion. Otherwise, this left-handed 4-helix bundle core is highly compact with short loops connecting each helix.

Within the virion, the capsid protein binds dsDNA as a set of repeating homodimers. The long axis of the 4-helix bundle is parallel to the capsid surface, with the N-terminal helix-kink-helix motif projecting inward, where it wraps around and completely encircles the dsDNA, such that the DNA is bound within a solvent inaccessible surface. Subunit interactions within the dimer are primarily hydrophobic, as are the lateral interactions between dimeric subunits. In contrast, the inter-strand interactions are largely polar. So too are the interactions with the DNA, which are mediated by conserved, positively charged (basic) side chains. Remarkably, the DNA is bound in the A-form, rather than B-form DNA typically seen in biological structures. The transition to A-form DNA is thought to stabilize the DNA genome in the acidic extracellular environment. This is likely a common mechanism for protecting DNA in harsh environments, as A-form DNA is also seen in bacterial spores.

Enveloped filamentous viruses, the *Lipothrixviridae*, are also present in the Archaea. Currently, high-resolution structural information exists for two of them, *Acidianus* filamentous Virus 1 (AFV1) and *Sulfolobus* Filamentous Virus 1 (SFV1). Their nucleocapsids show distinct similarities to SIRV2, although in contrast to SIRV2, they employ a heterodimer of paralogous proteins. AFV1 shows a variable pitch ranging from 39 to 47 Å centered on 43 Å with 9.3 heterodimers per turn, and SFV1 exhibits a 47 Å pitch with 17.14 units per turn. The major capsid proteins again adopt the right handed, antiparallel 4-helix bundle fold with a curved N-terminal helical extension. The SFV1 structure is further elaborated by a C-terminal extension, encompassing a 2-stranded antiparallel β -finger that packs against helices 3 and 4 on the outside of the nucleoprotein complex, followed by an extended random coil and two short α -helices that are collectively involved in subunit interactions with an adjacent super-helical turn. While the lateral interactions between subunits are strong, tight contacts between helical turns are lacking, potentially allowing the virion to bend. Again, the DNA is fully encased within the heterodimer as dehydrated A-form DNA, protecting it from acid catalyzed hydrolysis in the acidic hot spring environment.

Overall, however, the main feature distinguishing the *lipothrixviridae* from the *rudiviridae* is the presence of the outer lipid envelope. The AFV1 envelope is composed of cyclic glycerol dialkyl glycerol tetraether (GDGT) lipids, primarily GDGT-0, while the SFV1 envelope is dominated by archaeol, a diether lipid. Thus, both viruses specifically recruit minor host lipids, each building a membrane different from that of their host and from each other. In this context, the high-resolution helical reconstruction of AFV1 revealed an additional surprise; the envelope is half the thickness of the host membrane. In the viral envelope, molecular dynamics simulations indicated that the individual tetraether GDGT-0 lipids are likely folded back on themselves into a horseshoe-shaped configuration. Cyclic tetraether lipids have previously been observed to adopt this configuration at air water interfaces, however this represents the first indication of a biological membrane composed of this thinner monolayer of “horseshoe-shaped” lipids. Critically, the tetraether GDGT-0 lipids lack cyclopentane rings present in GDGT-4, allowing the flexibility needed to adopt this unique configuration and explaining the preferential recruitment of GDGT-0 into the viral envelope. Similarly, SFV1 shows an external lipid envelope thinner than expected. While the thickness of the envelope is consistent with an archaeol (C_{20}) monolayer thought to include a third capsid protein (VP3), it was concluded that the archaeol: VP3 viral envelope could not be reliably modeled with the information at hand. Interestingly, during budding the fusellovirus SSV1 also incorporates a GDGT-0 envelope

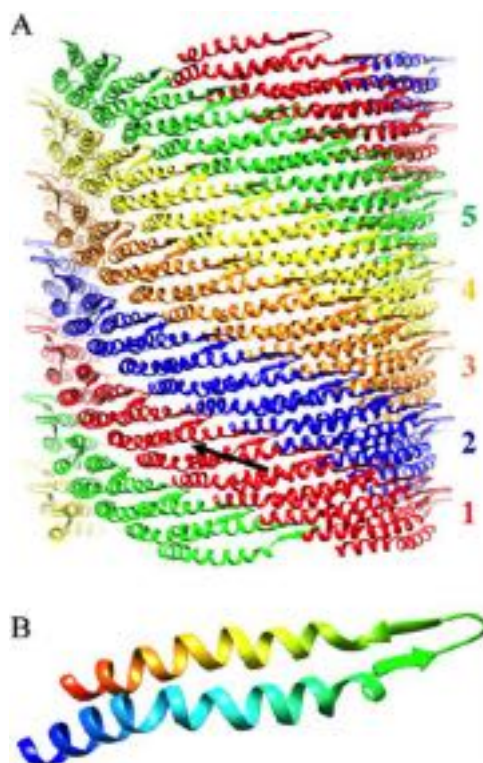


Fig. 6 *Aeropyrum pernix* Bacilliform Virus 1 (APBV1, PDB ID 5OXE) (A) A short section of the APBV1 capsid, illustrating its left-handed (black arrow) 5-start helical architecture with successive helicoids colored in red, blue, orange, yellow and green. The envelope is not depicted. (B) An enlarged view of the APBV1 subunit with its helix, reverse-turn, helix motif. The N-terminus in blue and the C-terminus in red. Within the capsid, the reverse-turn is exposed on the outside surface of the nucleocapsid, while basic residues at the N- and C-terminus are on the interior, where they presumably neutralize the negative charge on the packaged B-form DNA. While not resolved at high resolution, the DNA appears to adopt the 5-start symmetry of the capsid, though there must be symmetry mismatches in places.

with reduced thickness suggesting a related membrane structure, at least during this stage of the viral life-cycle. Thus, lipid monolayers of single-leaflet dimensions may play a critical role in the life cycles of thermophilic viruses.

Claviviridae

Aeropyrum pernix Bacilliform Virus 1 (APBV1) is the only representative of the *Claviviridae* family (Table 1). Although its nonenveloped rod-shaped morphology is reminiscent of the *Rudiviridae*, the 3 Å iterative helical reconstruction and *de novo* modeling of the VP1 coat protein distinguish it as an unrelated virus family. Subunits are organized as a left-handed 5-start helical assembly with 21.75 subunits per turn and a pitch of 133.7 Å (Fig. 6). The glycosylated major capsid protein uses a helix-turn-helix motif with a type-1 β -hairpin to form the turn. Overall the virion is highly compact with extensive helix-helix packing and an abundance of hydrophobic contacts. Relative to the abundance of 4-helix bundles described for other helical viruses, the packing of one helix-turn-helix motif upon its neighbor gives a similar structure, although this gives neither a right, nor left-handed up and down 4-helix bundle topology. The axis of each VP1 helix lies along the capsid surface with the helix curvature dictating the overall curvature of the virion. Minor structural proteins assemble to cap each end of the virion, with 5 terminal fibers used for cell adhesion extending from one end.

The circular double stranded DNA genome was not visualized at similarly high resolution, but it is known that it is nicked and therefore in a relaxed, as opposed to supercoiled form. Low resolution density that follows the symmetry of the 5-start helical coat protein is seen on the interior of the nucleocapsid. This symmetry gives rise to positively charged tracks formed by Lys9 and Arg79 with spacings that mirror the distance between phosphate groups in B-form DNA. In addition the genome appears fully hydrated, which is also consistent with B-form DNA. While the *rudiviridae* and *lipothrixviridae* inhabit thermo-acidic environments, APBV1 was isolated from a near neutral pH hot spring, potentially explaining the presence of B-form DNA in this archaeal virus.

Spindle-Shaped Viruses: *Bicaudaviridae* and *Fuselloviridae*

While clearly not classical helical viruses, the archaeal spindle-shaped viruses (Table 1) do show interesting parallels to helical viruses. These lemon-shaped viruses are divided into 2 families, the *bicaudaviridae* (large tailed spindles, 50–100 kbp ds DNA



Fig. 7 *Acidianus* Tailed Spindle Virus (ATSV, PDB ID 5EQW) (A) Large tailed spindle-shaped viruses like ATSV generally exit the cell as lemon- or spindle shaped particles that mature into smaller lemon-shaped particles with extended tail structures at one or both ends. Here, ATSV is modeled as a left-handed multi-start helix. (B) The ATSV subunit is depicted in rainbow colors, with the N-terminus in blue and the C-terminus in red. It is a decorated, right-handed 4-helix bundle, with a small capping β -sheet and extended hydrophobic loop on top, and a 5th helix at the C-terminal end. Within the capsid, the long axis of the 4-helix bundle runs $\sim$ parallel to the long axis of the capsid. Lateral interactions within a helicoid strand are extensive, but the inter-strand interactions are more modest, potentially allowing the rope-like strands to “slip” past each other as it transitions from the metastable lemon-shaped capsid to the tubular structure.

genomes) and the fuselloviridae (smaller tailless spindles, 15–20 kbp ds DNA genomes). For the Bicaudaviridae, as the virus matures, the lemon-shaped capsid transitions into an elongated cylindrical form at one or both ends of the capsid, giving the appearance of tail growth. Similarly, the fuselloviridae are seen to transform into empty cylinders in response to chemical perturbation. While we lack structural information on the fuselloviridae, coat protein structures have been determined for two members of the bicaudaviridae: *Acidianus* Two-tailed Virus and *Acidianus* Tailed Spindle Virus (ATSV). Like tobacco mosaic virus, the central fold consists of a classic right-handed antiparallel 4-helix bundle (Fig. 7). In addition, the helix is capped end on by a small, mixed β -sheet, although the β -cap is at the opposite end of the helical bundle relative to the cap on TMV, and a short 5th helix is found at the C-terminus.

The ATSV capsid protein crystalizes as a left-handed four-start helical assembly with 10 subunits per turn of a single helical strand, suggesting a similar structure for the virus tail (Fig. 7). The long axis of the helical bundle lies parallel to the long axis of the helical assembly, not radially as in tobacco mosaic virus. Combined with radially symmetric striations on the surface of the spindle-shaped capsid in cryo-EM micrographs, this led to a proposed architecture in which a multi-start helix of varying width forms the lemon-shaped capsid that then extends into the tail, with the dsDNA genome packaged asymmetrically within the lemon-shaped capsid. However, it was instead modeled as a 6-start helical assembly in order to fit the diameter of the observed tail to the model. The true helical parameters remain unknown.

Subunit interactions in the crystal reveal extensive intra-strand contacts, with minimal inter-strand contacts. The combination of strong intra-strand and weak inter-strand forces could allow the capsid coils to slip past one another, smoothly altering the capsid diameter as it transitions to the lower energy cylindrical state. It is likely that the dynamic nature of multi-start helical assemblies will be a unifying feature of spindle-shaped virions, and that the transition from the metastable spindle-shape to the lower energy tailed helical structures are the driving force for genome delivery into the host cell.

Overall, lipid envelopes are so far found in 11 of the 18 archaeal virus families, many of them from high temperature acidic environments. Archaeal lipids composed of fused diethers or tetraether components show extremely low proton permeability indicating that one function in the virion could be the protection of the packaged viral genome from the acidic environment. Intriguingly, a number of these viruses selectively acquire a subset of host lipids. The fuselloviridae and lipothrixviridae envelopes are composed primarily of GDGT-0 (SSV1 and AFV1) or archaeol (SFV1), which have much higher flexibility than the GDGT-4 (GDGT with 4 cyclopentane groups) that predominates in the membranes of their hosts. These flexible lipids are likely required to facilitate the high membrane curvature of fuselloviridae and lipothrixviridae capsids. Increased environmental stability is another rationale for selected lipid acquisition. SH1 enriches its membrane with PGP-Me which is known to increase membrane stability in the high salt environments inhabited by this virus.

While the unique properties of these envelopes solve one problem, they raise new ones. Specifically, SSV1 has been shown to bud from the host cell in a manner reminiscent of eukaryotic viruses, and it seems likely that of the many enveloped archaeal viruses, at least some of them will fuse with host cell membranes to deliver their genomes. How do these viruses bud from or fuse with membranes composed of membrane spanning cyclic tetraether lipids that are frequently found in thermophilic archaea? The presence of such membrane spanning lipids would seem to prohibit the separation of the membrane into the individual leaflets characteristic of the hemi-fusion intermediate proposed for eukaryotic viruses, suggesting unique strategies and protein machinery might be needed to drive the process. In this light, one possibility is that the horseshoe-shaped lipid monolayer could serve as an intermediate equivalent to a single membrane leaflet in the hemi-fusion model.

Conclusion

The last decade has seen a major expansion in the level of high-resolution structures available for helical viruses. It is now clear that viral capsids with helical symmetry show multiple variations on several common themes. They may be rigid rods or flexible filaments composed of one or more parallel strands exhibiting either right- or left-handed twist, with open or capped ends, either naked or membrane enveloped. The protein subunits are often decorated 4-helix bundles, but subunits with 1 or 2 α -helices, and one or two domain subunits composed of larger helical bundles are also common. Nucleic acid is packaged independent of sequence, and can be bound within a single subunit or between subunits in successive helical strands, and may or may not share the same helical symmetry.

Despite recent progress, high-resolution structures of helical viruses are still scarce. Structural studies of the hemorrhagic *filoviridae*, including Marburg virus are clearly a priority. So too are studies of the *potyviridae*, *betaflexiviridae* or *closteroviridae*. These flexible plant viruses are of extreme agricultural importance and we lack high-resolution structural information for any member of these viral families. Recent structural studies of viruses infecting the Archaea also make clear that much remains to be learned about the origins, evolution and diversity of these viruses. Indeed, it seems each new high-resolution structural study identifies important new variations on the general themes described above, and in many cases reveals new architectural principles for the construction of viral capsids. Finally, high-resolution structural studies of the capping structures of helical viruses are generally lacking and sorely needed. These studies will certainly reveal further structural diversity among helical viruses and provide major insight into viral assembly, egress, maturation, attachment and entry. In short, much remains to be done, but recent advances in cryo-EM methodology clearly suggest that these are tractable problems and this structural information is now within reach.

Further Reading

- Agirrezabala, X., Mendez-Lopez, E., Lasso, G., *et al.*, 2015. The near-atomic cryoEM structure of a flexible filamentous plant virus shows homology of its coat protein with nucleoproteins of animal viruses. *eLife* 4, e11795.
- Dimaio, F., Chen, C.C., Yu, X., *et al.*, 2015. The molecular basis for flexibility in the flexible filamentous plant viruses. *Nature Structural & Molecular Biology* 22, 642–644.
- Dimaio, F., Yu, X., Rensen, E., *et al.*, 2015. Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* 348, 914–917.
- Harrison, S.C., 2015. Viral membrane fusion. *Virology* 479–480, 498–507.
- Harrison, S.C., 2011. Virus structure and assembly. In: Acheson, N.A. (Ed.), *Fundamentals of Molecular Virology*, second ed. John Wiley & Sons, Inc, pp. 18–30.
- Hochstein, R., Bollschweiler, D., Dharmavaram, S., *et al.*, 2018. Structural studies of acidianus tailed spindle virus reveal a structural paradigm used in the assembly of spindle-shaped viruses. *Proceedings of the National Academy of Sciences of the United States of America* 115, 2120–2125.
- Kasson, P., Dimaio, F., Yu, X., *et al.*, 2017. Model for a novel membrane envelope in a filamentous hyperthermophilic virus. *eLife* 6.
- Liu, Y., Osinski, T., Wang, F.B., *et al.*, 2018. Structural conservation in a membrane-enveloped filamentous virus infecting a hyperthermophilic acidophile. *Nature Communications* 9.
- Marvin, D.A., Welsh, L.C., Symmons, M.F., Scott, W.R., Straus, S.K., 2006. Molecular structure of fd (f1, M13) filamentous bacteriophage refined with respect to X-ray fibre diffraction and solid-state NMR data supports specific models of phage assembly at the bacterial membrane. *Journal of Molecular Biology* 355, 294–309.
- Namba, K., Pattanayek, R., Stubbs, G., 1989. Visualization of protein-nucleic acid interactions in a virus - Refined structure of intact tobacco mosaic-virus at 2.9-Å resolution by X-ray fiber diffraction. *Journal of Molecular Biology* 208, 307–325.
- Ptchelkine, D., Gillum, A., Mochizuki, T., *et al.*, 2017. Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. *Nature Communications* 8, 1436.
- Stubbs, G., Kendall, A., 2012. Helical viruses. *Viral Molecular Machines* 726, 631–658.
- Sugita, Y., Matsunami, H., Kawaoka, Y., Noda, T., Wolf, M., 2018. Cryo-EM structure of the Ebola virus nucleoprotein-RNA complex at 3.6 Å resolution. *Nature* 563, 137–140.
- Welsh, L.C., Symmons, M.F., Marvin, D.A., 2000. The molecular structure and structural transition of the alpha-helical capsid in filamentous bacteriophage Pf1. *Acta Crystallographica Section D Biology Crystallographica* 56, 137–150.
- Xu, J., Dayan, N., Goldbourt, A., Xiang, Y., 2019. Cryo-electron microscopy structure of the filamentous bacteriophage IKe. *Proceedings of the National Academy of Sciences of the United States of America* 116, 5493–5498.

Relevant Websites

- <https://viralzone.expasy.org/750>
dsRNA virion ~ ViralZone page.
- <https://www.rcsb.org/>
RCSB PDB: Homepage.
- <https://www.ebi.ac.uk/pdbe/emdb/>
The Electron Microscopy Data Bank (EMDB) at PDB.
- <https://viralzone.expasy.org/>
ViralZone root.

Giant Viruses and Their Virophage Parasites

Rodrigo AL Rodrigues, Ana CdSP Andrade, Grazielle P Oliveira, and Jônatas S Abrahão, Federal University of Minas Gerais, Belo Horizonte, Brazil

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Megavirales A proposed taxon (order) to comprise all nucleocytoplasmic large DNA viruses, given its hypothetical common origin.

MIMIVIRE A hypothetical mimivirus defense system against virophage infection.

Polintovirus Large DNA transposons found in cellular organisms that encodes up to 10 proteins, including a homolog of viral capsid, possibly generating infectious particles.

Provirophage A virophage genome integrated in the host (giant virus) DNA.

Transpoviron Named after the “transposon,” this is a small linear plasmid-like DNA element found in mimivirus genomes.

Viral factory A viral-induced region in the host cytoplasm, wherein genome replication and viral morphogenesis of giant viruses occur.

Virophage A new family of viruses associated with giant viruses, depending on the viral host for replication.

Introduction

Viruses are traditionally perceived as filterable agents capable of passing through membranes with pore sizes of 0.22 μm , containing small genomes encoding only a few proteins. However, with the discovery of giant viruses, these concepts have been broken down, and a new era of virology has arisen. Giant viruses are part of a group of nucleocytoplasmic large DNA viruses (NCLDV), in which the viral particle size and structure and the genome length and complexity make them special within the group. In the original NCLDV proposal, the group consisted of the viral families *Poxviridae*, *Asfarviridae*, *Iridoviridae* and *Phycodnaviridae*. These families are composed of viruses that contain large double-stranded DNA (dsDNA) genomes, which replicate partially or entirely in the host cytoplasm, and which also share genetic similarities that points to a hypothetical common origin, thus comprising a monophyletic group. However, NCLDV only became a more intense topic of discussion and evolutionary study in 2003, due to the discovery of the first giant virus of amoebae named *Acanthamoeba polyphaga* mimivirus (APMV). Phylogenetic analyses indicated that APMV could be categorized with other viruses belonging to the NCLDV group, and the fact that this virus occupied an isolated branch in the phylogenetic tree that was associated with distinct morphological features led to the establishment of a new family, named *Mimiviridae*. The discovery of mimiviruses paved the way for the isolation of other giant viruses, with the development of new isolation techniques that were no longer constrained by the classical definitions of viruses based on particle size. As a consequence, new NCLDV members were discovered, including marseilleviruses, faustoviruses, pandoraviruses, cedratviruses, pithoviruses, orpheovirus, mollivirus, pacmanvirus and kaumoebavirus and new definitions of viruses are currently being debated. These viruses exhibit many distinctive features, along with some common characteristics that classify them as part of the NCLDV group, which was proposed to comprise a new viral order named “Megavirales.” In this article, we describe how the giant viruses were discovered and describe the basic features of this expansive group, including their diversity, replication cycles, genomics and evolution. Finally, we outline their genetic mobilome, including virophages, the parasites of giant viruses.

The Serendipitous Discovery of Giant Viruses of Amoebae

The isolation of the first mimivirus accidentally occurred during studies of amoebae-associated pathogens from environmental water samples in the midst of a pneumonia outbreak in Bradford, England. This virus was initially considered to be a bacterium infecting amoebae, mainly due to its Gram-positive staining back in 1992 that revealed a microorganism referred to as “Bradford coccus.” In addition, molecular attempts had failed to identify the new isolate. It was years later, when the sample was taken to Aix Marseille University where it was characterized and described in 2003 as the largest virus known at that time. Its classification as a virus was confirmed by the observation of the eclipse-phase during its replication and the visualization by transmission electron microscopy (TEM) of a likely isometric particle, which is typical of viruses, that was covered with long fibrils. The virus was named after its ability to mimic microbes and the host cell it was first isolated from, *Acanthamoeba polyphaga*. Mimivirus particles measure about 750 nm in diameter and are composed of a peculiar structure, characterized by a protein capsid containing an internal lipid membrane and surrounded by a dense glycoprotein fibril layer that is important for viral attachment to the host cell and other microorganisms (Fig. 1(A)). The capsid does not exhibit the typical icosahedral symmetry, presenting instead a modified vertex in starfish shape, named *stargate*, a portal from where the viral genome is released into the hosts’ cytoplasm (Fig. 1(B)). Its genome consists of a long linear double-stranded DNA molecule of approximately 1.2 Mbp, encoding over 1000 genes with many different

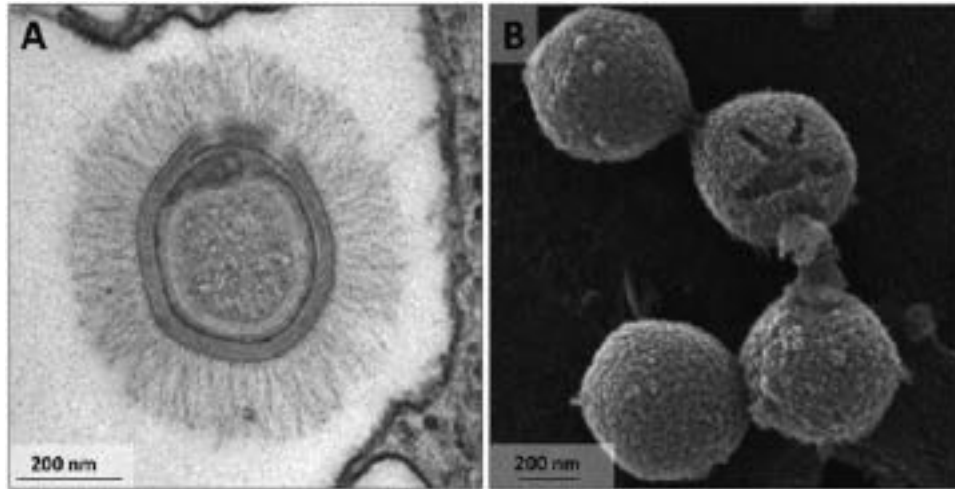


Fig. 1 Electron microscopy images of mimivirus particles. (A) Transmission electron microscopy image demonstrating the capsid covered in fibrils and an internal membrane; (B) scanning electron microscopy image evidencing the stargate structure.

functions, such as enzymes involved in DNA replication, recombination and repair (e.g., DNA polymerase and DNA topoisomerase), transcription (RNA polymerase and mRNA capping enzyme) and translation (translation factors and aminoacyl-tRNA synthetases), among others. The mimivirus particles penetrate into the host cell (free-living amoebae) by phagocytosis; this leads to rounding of the infected cell due to cytopathic effects, followed by lysis of the amoebae. Similar to other viruses belonging to the NCLDV group, mimiviruses replicate in the cytoplasm of the host cell, where they form large viral factories.

Since the discovery of APMV in the context of pneumonia, discussions about the possible relationship between mimiviruses and this disease have been raised. One study showed that laboratory mice inoculated intracardially with mimivirus displayed histopathologic symptoms, and mimiviruses were re-isolated from lung samples of these animals. Moreover, patients with pneumonia presented antibodies against mimivirus, and a high positivity of mimivirus in hospital facilities was recorded. Years later, the isolation of mimivirus from stool and bronchoalveolar lavage samples from patients with pneumonia strengthened the relationship between mimiviruses and this disease. However, human cells capable of being infected by these viruses, wherein the virus could establish a productive replication cycle, have not been found yet. Recent studies found no evidence of mimivirus replication in human peripheral blood mononuclear cells, despite its ability to enter and interfere with antiviral responses. Furthermore, other studies revealed the absence of mimiviruses in patients with respiratory disease by means of molecular biology and/or serology. However, it is important to consider the possibility that mimivirus might be a member of the vertebrates' virome, since their genome has been detected in humans, as well as both in domesticated and wild mammals. Nonetheless, this association needs more evidence to be confirmed. It is worth noting that a mimivirus-related virus infecting sturgeons has been described in the last years, causing lethal disease in the integumentary systems of these animals. Further studies are required to better evaluate the putative pathogenic role of mimiviruses and other giant viruses in humans and other vertebrates.

Diversity and Distribution of Giant Viruses in the World

After the discovery of mimiviruses, other giant viruses of amoebae have been described over the past decade in different places around the world, revealing an increasing complexity of this fraction of the virosphere. These viruses have a great structural diversity, with particles of different sizes and shapes, ranging from icosahedral capsids of about 200 nm to spherical, ovoid or tail-containing particles that exceed 1000 nm in length. Four distinct groups, named marseilleviruses (family *Marseilleviridae*), kaumobavirus, pacmanvirus and faustovirus (all phylogenetically related to *Asfarviridae* members), are composed of smaller viruses that present icosahedral capsids of 175–250 nm. Marseilleviruses were the second group of amoeba viruses to be discovered and were associated with *Acanthamoeba* cells in water samples from Paris, France. Over the years, other isolates were described, expanding the new viral family that is currently composed of five different lineages (A–E), with viruses discovered in different regions of the planet, such as Brazil, Australia, Tunisia and Japan, among others. The other groups, despite infecting amoeba cells, are most closely related to the African swine fever virus, a mammalian parasite. Pacmanvirus also infects *Acanthamoeba* cells, but kaumobavirus and faustoviruses are associated with *Vermamoeba vermiformis* cells, a different amoebal host. Similar to mimiviruses, these smaller viruses seem to have an internal lipid membrane surrounding their genetic material. Electron microscopy images demonstrate an overall similar structure between these different viruses (Fig. 2(A,B)). Although in-depth structural analyses of giant virus particles are scarce, some studies have revealed that certain viruses display exclusive features. Among them, faustovirus has the peculiarity of being formed by double concentric icosahedral capsids that connect to each other by protrusions of the internal capsid surface, which was not observed for other giant viruses until now.

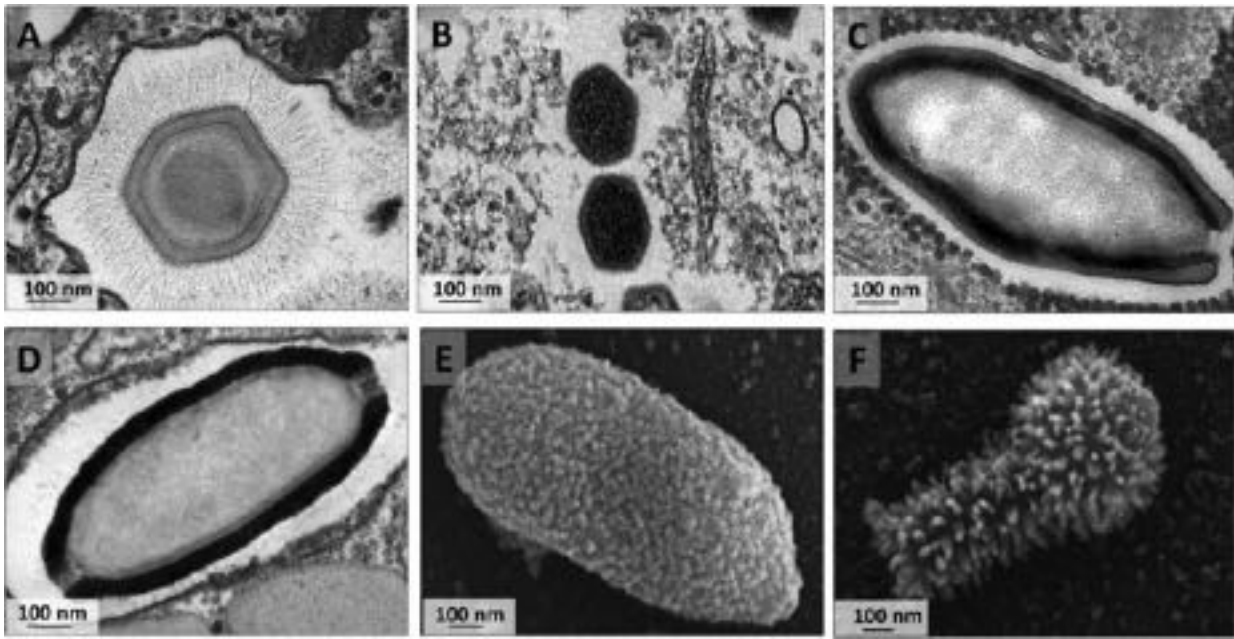


Fig. 2 Diversity of giant viruses. Transmission and scanning electron microscopy images of different viral groups, representing the large variety of particle sizes and shapes among giant viruses. (A) mimivirus; (B) marseillevirus; (C) pandoravirus; (D) cedratvirus; (E) orpheovirus; (F) tupanvirus.

Other groups have larger particles that are ovoid in shape. In 2013, the description of pandoraviruses set a new limit to the viral particles' sizes, with ovoid-shaped virions with 1.0 μm presenting an ostiole-like apex at one end of its capsid (Fig. 2(C)). These viruses were first isolated in Chile, Australia and Germany, but novel isolates have been found in France, New Caledonia and Brazil, constituting an expanding putative *Pandoraviridae* family. Curiously, the pandoravirus capsid composition remains unknown, since no protein even remotely similar to the major capsid protein (MCP) was detected in the genomes of these viruses. A year later, the previous size limit was pushed once again by the discovery of the giant Pithovirus sibericum, isolated from 30,000-year-old permafrost soil samples from Siberia, composed of ovoid particles reaching more than 1.5 μm in length, making them the largest viral particles known so far. The particles have a striated capsid wall and a hexagonal grid-like structure, called a cork, which is generally found at one extremity of the particle and is an analogous structure to the apex of pandoraviruses. Curiously, a new virus was isolated from the same ancient sample, named Mollivirus sibericum, which is a spherical-shaped virus that is approximately 600 nm in diameter, but no further characterization of this virus has been performed to date. A contemporary pithovirus, named Pithovirus massiliensis, was isolated in France that exhibited a similar structure and a genomic conservation. Due to these structurally and genomically distinct features, a new viral family called "*Pithoviridae*" was suggested. In 2016, cedratvirus, a virus with a morphology similar to the pithovirus, was described, but it contained smaller particles with a mean size of 1.2 μm . Like the pithovirus, cedratvirus has apical corks, but at both ends of the particles. Other cedratviruses have been isolated in France and Brazil, with the Brazilian isolate exhibiting differences in particle and genome size and probably constituting a new lineage among cedratviruses (Fig. 2(D)). Cedratviruses and pithoviruses are phylogenetically close and might constitute a new single family. A distantly related *Pithoviridae* member named Orpheovirus was recently isolated using *V. vermamoeba* cells, which is morphologically similar to pandoraviruses, however, have more electron-dense capsid layers that are covered with short fibrils on their outer surface (Fig. 2(E)). More studies should be performed to get a clear picture of the diversity and distribution of these new giant viruses.

Currently, the family *Mimiviridae* encompasses the largest number of representatives, which have considerable differences in both structure and genomic features. After the discovery of APMV, several other mimivirus-like viruses were isolated from different parts of the world that exhibited structural similarities, but with genomic differences, contributing to the formation of three different lineages, informally called lineage A (represented by APMV), B (represented by Moumouvirus) and C (represented by Megavirus). Distantly related mimiviruses were also described infecting marine flagellates, namely, Cafeteria roenbergensis virus (CroV). CroV exhibits large particles with many genes shared with mimiviruses, which placed it within the family *Mimiviridae*, albeit within a separate genus, the *Cafeteriavirus*. The Bodo saltans virus, a kinetoplastid-infecting virus, constitutes a new member of the family, the first isolated member of the putative subfamily "*Klosneuvirinae*", whose members were initially identified by metagenomics analysis, revealing a new group of mimiviruses with unexpected genomic features, such as a much larger translation-related gene set. More recently, the structural and genomic complexity of the family was expanded with the discovery of Tupanviruses in samples collected from extreme places in Brazil, which are the viruses with the most complete translational apparatus in the virosphere. The viral particles are formed by a capsid similar to that of APMV and are associated with a cylindrical

tail of about 550 nm in length and 450 nm in diameter, with the whole particle covered in fibrils (Fig. 2(F)). The size of these particles is approximately 1.2 μm on average, although some particles may be up to 2.3 μm , due to the plasticity of the tail. Moreover, a new mimivirus-like parasite was isolated from the bark of a sycamore tree using *Saccamoeba* sp. as a host organism, referred to as Mimivirus KSL-5, exhibiting polyhedral virions of ~ 450 nm. Curiously, this virus was able to infect only *Saccamoeba* cells, suggesting the existence of a new group of mimiviruses with a novel host range.

The discovery of the first viruses associated with amoebas at the beginning of 21st century, with their diverse and complex structural characteristics, led to an intense search for new giant viruses in distinct places around the world. Since then, the number of new isolates has grown every year. These viruses have already been isolated using samples collected on each continent from various conditions, such as polluted environments, those with extreme conditions of temperature and pH, slightly anthropized regions and clinical samples. The constant discovery of giant viruses in wide-ranging types of samples reinforces the idea that these viruses are ubiquitous, as well as their protist hosts. Metagenomic studies reinforce their ubiquity, further indicating that many amoeba-related viruses have yet to be discovered. Remarkably, a large metagenomic analysis of soil samples from the Harvard forest (Massachusetts, USA) led to the assembly of the complete genomes of 16 new giant viruses, representing new lineages of klosneuviruses, other members of the family *Mimiviridae* and hundreds of MCP fragments belonging to giant viruses, thus indicating a huge diversity of these viruses in soil ecosystems that remains to be characterized. The host spectrum for most of these viruses is restricted to a few species of amoebae of the genus *Acanthamoeba* and/or *Vermamoeba*. However, recent discoveries of new members of the *Mimiviridae* family have shown that the spectrum may be broader, since they can infect different hosts, including other groups of protozoa and fish species (e.g., Bodo saltans virus, tupanviruses and namao virus). Novel studies involving exploration in unexplored regions using different cell platforms and isolation techniques could reveal exciting new viral groups, contributing to an increase in our knowledge about the diversity, ecology and evolution of these viruses.

The Life Style of the Giants

The replication cycle of giant viruses begins with viral entry into the host cells. For most of these viruses, this process has been proposed to occur by phagocytosis. Drugs such as cytochalasin, a phagocytosis inhibitor, have already been used to demonstrate the impact on this entry mechanism for mimivirus, cedratvirus and large membranous vesicles containing marseillevirus particles. For pandoravirus, pithovirus, mollivirus, orpheovirus, faustovirus, pacmanvirus and kaumoebavirus, entry by phagocytosis was suggested by electron microscopy image analyses. The phagocytosis process requires the recognition of particles of at least 500 nm, so the assumption that giant viruses enter by this mechanism seems reasonable. However, other smaller viruses, such as marseilleviruses (~ 250 nm) do not fulfill this criterion; thus, other mechanisms are likely involved. A thorough investigation of the replication cycle of this virus demonstrated that it is able to penetrate into cells by two alternative routes: by phagocytosis of grouped particles inside giant vesicles, or by exploiting other endocytic pathway for single particles. It is not clear yet if a similar process could occur for other large (yet not giant) viruses, such as faustoviruses and kaumoebavirus. After entry, the viral genome is released into the host cytoplasm. It has been observed that the uncoating process of mimivirus, mollivirus, pandoravirus, pithovirus, faustovirus and cedratvirus occurs after the fusion of the inner viral membrane with the endosomal membrane, allowing the formation of a channel through which the genome is released into the host cell. For other giant viruses, the uncoating process has not been described in detail yet. Mimiviruses release their genomes after stargate opening, and it was verified that endosome acidification is important for the induction of this step. The genomes of pandoraviruses, orpheovirus, pithovirus and cedratvirus are released by their respective ostiole-like apex and corks at the extremities of the particles.

Following the release of the genome, a typical viral eclipse phase is observed, wherein no viral particles are detected in the host cytoplasm, followed by the formation of small viral factories (VFs) in the cytoplasm. Throughout the infection, these VFs increase in size and occupy a large portion of the cellular cytoplasm in the later stages of infection. VFs of most giant viruses are non-delimited electron-lucent areas that contain particles at different stages of morphogenesis. Nevertheless, mimivirus VFs are well-defined structures, with at least two areas described: the central area, where replication of the genome and assembly of the structures necessary to form the viral capsid occurs, and a less electron-dense region of apparent fibrillar nature in the periphery. This outermost portion was named the fibrils acquisition area, since it was observed that is in this region the newly formed particles acquire the fibrils. Such compartmentalization of VFs was not observed for other giant viruses so far. Previous studies proposed that capsid morphogenesis of mimiviruses, pandoraviruses and cedratvirus is initiated by crescent-shaped lamellar structures that can be observed inside the VFs. The morphogenesis of cedratvirus involves crescent-shaped structures that assume staple-shaped and horseshoe conformations, depending on the section of TEM. Initially, only one of the corks is observed; then the capsid is filled and closed with the emergence of the second cork, and the capsid wall becomes thicker. Regarding pandoravirus, the capsid and internal contents seem to be assembled simultaneously. A pattern of where the assembly of the particle begins was not observed, since it can be at the extremity that presents ostiole-like apex or at the opposite end. For mimiviruses, capsid precursors increase in complexity and assume a pseudo-icosahedral symmetry. Then genome acquisition occurs at the stargate on the opposite side, and this step may occur simultaneously with the acquisition of surface fibrils.

During the morphogenesis step, defective particles have been recurrently observed for different groups of viruses, including mimiviruses, pandoraviruses, cedratviruses, among others. For mimiviruses, these unusual particles were initially associated with

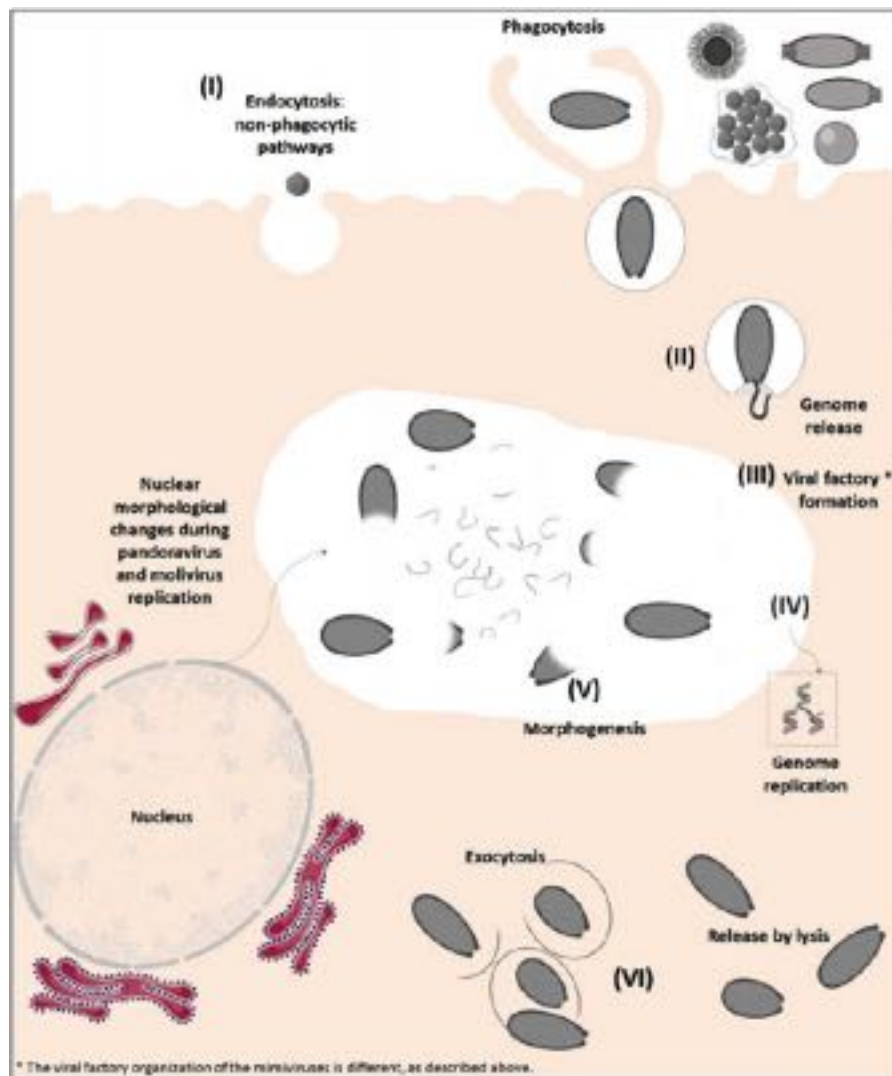


Fig. 3 Representative scheme of the amoeba giant viruses replication cycle. The particles can enter host cell by phagocytosis or endocytosis for individual particles of marseillevirus (i). Subsequently, the genome is released into the host cytoplasm (ii). A viral factory is established in the cytoplasm of the host cell (iii) wherein the genome replicates (iv) and the morphogenesis of new particles occurs (v). During morphogenesis of mollivirus and pandoravirus, nuclear morphological changes are observed. The cycle ends with the release of progeny, which can occur by exocytosis or after cell lysis (vi). In this illustration, the pandoravirus was chosen in a representative way among the other groups of giant viruses. *Mimivirus viral factories organization is different from that illustrated in the figure, being more electron-dense than the cytoplasm and divided into at least two distinct areas.

the presence of virophages, parasites of giant viruses (described later in this article). However, recent analysis demonstrated the formation of defective particles, even in the absence of these parasites, thus, making clear that malformed particles are a natural process during the replicative cycle of these viruses. Once viral particles are fully assembled, release of most of these viruses occurs after cell lysis. Marseillevirus can release many particles into giant vesicles after cell lysis, which boosts its entry into another host cell by exploiting different entry strategies. Moreover, it has been suggested that the exocytosis process could be used for releasing mollivirus, pandoravirus, orpheovirus and cedratvirus particles. However, the viral release through exocytosis still needs to be further investigated for the majority of these viruses. The real impact of this viral release mechanism is still not well known.

Overall, in spite of some peculiarities of each viral group, a general picture of the replication cycle of giant viruses comes to light (Fig. 3). Viruses enter host cells through endocytosis or phagocytosis (i), release the genome into the cytoplasm (ii), establish a viral factory (iii) where the genome is replicated (iv), new particles are assembled during the morphogenesis step (v), and finally, the new progeny is released by cell lysis or exocytosis (vi). It is still not clear if these viruses depend on the host nucleus to complete their replication cycle, but some evidence for this has been described for marseilleviruses, pandoraviruses, mimivirus and mollivirus. Additional analyses are needed to better elucidate this important step of the life cycle for the different groups of giant viruses.

Genomics and Proteomics of Giant Viruses

A striking feature of giant viruses is the presence of an extensive dsDNA genome containing hundreds of genes. Some viruses have a linear genome, such as mimiviruses, tupanviruses, pandoraviruses, mollivirus and pacmanvirus, while others have a circular genome, such as marseilleviruses, faustoviruses, kaumobebavirus, pithoviruses, cedratviruses and orpheovirus. For the most part, the genome of these viruses is A/T-rich, ranging from ~55% in marseillevirus to ~75% in mimiviruses. Pandoravirus and mollivirus differ from the others in this regard, presenting genomes with G/C content above 60%. The size of the genome varies among the different viral groups between, ~350 kb and ~2500 kb, with Pandoravirus salinus containing the largest viral genome described so far. These viruses present, on average, 1 gene/kb and short intergenic regions of approximately 200 nucleotides, although large regions of thousands of nucleotides can occur between genes.

Even more impressive than the size of the genome is the genetic content of these viruses. Sequencing of the APMV genome in 2004 revealed the presence of genes rarely found in the virosphere, some of which are considered exclusive to the cellular world. Giant viruses present several genes involved in DNA replication, recombination and repair, nucleotide and carbohydrate metabolism, transcription (including RNA polymerases and various transcription factors) and protein synthesis. In the latter category, members of the family *Mimiviridae* are characterized by exhibiting several transfer RNAs (tRNAs), many translation factors (involved in the three stages of the process, i.e., initiation, elongation and release) and aminoacyl-tRNA synthetases (aaRS), which are key enzymes in the process of protein synthesis and have never been observed before for other viruses. In this context, the tupanviruses are the viruses that present the largest arsenal of genes involved in the translation processes observed in the virosphere, with up to 70 tRNAs and 20 aaRS (one for each proteinogenic amino acid encoded by the standard genetic code). Other giant viruses also have components of the translational apparatus, but in smaller quantity and diversity. Although several translation-related genes exist in these viruses, no ribosomal RNA (rRNA) gene has been detected in a virus thus far, making them still dependent on their hosts to synthesize proteins. In spite of several gene novelties observed, a common feature in all giant viruses at the time of their discovery is the large number of genes with unknown functions, without any similarity to other genes deposited in databases, called ORFans, with values varying from 31% to 84% of ORFans in the genome. Pandoravirus salinus has the larger fraction of ORFans among the giant viruses (Fig. 4). This massive number of new genes reflects how different these viruses are and suggests that they are a valuable source of genetic diversity.

Little is known about the transcriptional profile of these viruses and how the regulation of gene transcription occurs. Conserved A/T-rich sequences were identified in most of the mimivirus, marseillevirus, faustovirus and kaumobebavirus genes. Analyses based on RNA sequencing for APMV indicated the existence of a temporal profile of gene expression, with the AAAATTGA promoter sequence being associated with early expressed genes. Other members of the family *Mimiviridae* exhibit the same promoter motif, such as CroV, Bodo saltans virus and tupanvirus, which suggests it being a typical feature of this viral family. Marseillevirus also has genes expressed at different times in the replication cycle, but the previously identified motif promoter (AAATATTT) does not appear to be specifically related to any temporal class of genes, as is observed for mimiviruses. Most genes involved in cell signaling (e.g., serine/threonine kinases) and DNA replication and repair (e.g., DNA polymerase and DNA topoisomerases) are expressed at earlier times in the viral replication cycle, whereas transcription, translation and especially those involved in viral morphogenesis, such as constituents of the capsid (e.g., MCP and DNA packaging ATPase), are expressed later during the infectious cycle. In addition, transcripts of these viruses are polyadenylated, and hairpin polyadenylation signals have been identified throughout the genome of the mimivirus; however, there is still no information for the other giant viruses, being a broad field for new investigations.

Proteomics studies have revealed that giant virus particles are composed of a large amount of proteins, with more than 100 proteins found in the particles of APMV, CroV, Tupanvirus soda lake, Pithovirus sibericum, Mollivirus sibericum, pandoraviruses and E12 faustovirus. In the particles of all these viruses proteins with different functions were detected that were involved in DNA replication, oxidative pathways, lipid modification, transcription processes, in addition to the already expected structural genes. It is worth noting that in all viral particles, products from some ORFans were also identified, demonstrating that new genes discovered in these entities are truly expressed and translated, although their exact function remains unknown. Many genes identified in viral particles, as well as others expressed during the viral replication cycle, have multiple origins (i.e., from viruses, eukaryotes, bacteria and archaea), making giant DNA virus true genetic mosaics. This feature is observed in all groups of giant viruses discovered so far (Fig. 4). This genomic mosaicism is attributed in large part to the sympatric lifestyles of these organisms (i.e., from the same environment as several other microorganisms), since they infect amoebas, which are considered to be melting pots for the emergence of new organisms.

Origin and Evolution: An Intriguing Enigma

Giant amoeba viruses share several genes with other members of the NCLDV group, which is a possibly monophyletic group within the virosphere. With the discovery of different viral groups, the amount of so-called core genes has been drastically reduced, with only three genes currently present in members of all viral groups, including the D5 primase-helicase, viral late transcription factor 3 (VLTf3) and DNA polymerase B family. Some core genes appear to have been lost in some specific groups throughout evolution, such as the MCP genes in pandoravirus and DNA packaging ATPase in pithovirus and cedratvirus. Phylogenetic analyses based on these core genes tend to maintain the same topology, with the presence of three major clades, with the different groups of amoeba viruses scattered among them (Fig. 5). Although they share a common ancestor with the other *Megavirales* members, it is

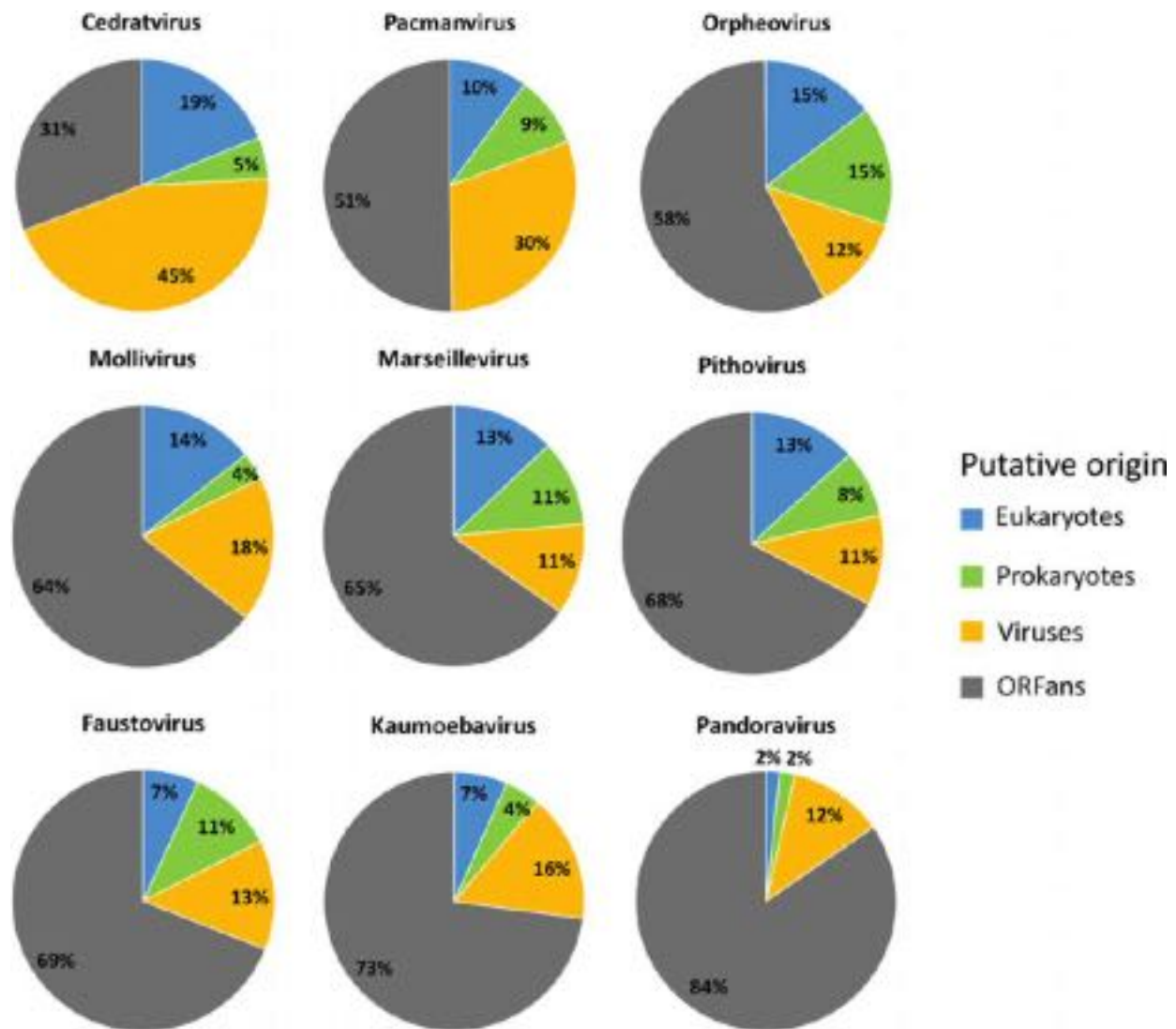


Fig. 4 Genome composition of giant viruses. Pie-chart representing the proportion of genes presenting no homology in databases (ORFans) and genes with the best hits with eukaryotes, prokaryotes and viruses for different groups of giant viruses. The data was considered at the time the first isolate for each group was described. Representatives include Cedratvirus A11, Pacmanvirus A23, Orpheovirus LCC2, Mollivirus sibericum, Marseillevirus T19, Pithovirus sibericum, Faustovirus E12, Kaumoebavirus isolate Sc and Pandoravirus salinus.

possible that the gigantism observed for some groups has arisen independently throughout evolution, with the host acting as a selection agent of this characteristic. However, the origin and evolution of these viruses are at the heart of a heated debate that has lasted more than a decade, and there is still no scientific consensus.

The discovery of the presence of an extensive genome (> 1.0 Mb) encoding hundreds of proteins (~ 1000) in mimiviruses, with particular attention to translation genes like aaRS, contributed to reviving a long-standing debate about the nature of viruses (i.e., whether they are living organisms or not). Initial analyses suggested that mimiviruses could be remnants of a fourth domain of life, originating from a more complex organism (possibly a cell), which would have evolved through genomic reduction to adapt to the intracellular parasitism lifestyle. Additional phylogenomic analyses, associated with phylogenetic reconstructions based on protein fold conservation, reinforced this initial hypothesis, even inferring that giant viruses would have coexisted with primitive cellular life forms. However, alternative analyses involving conserved genes, as well as those related to the translation process, suggested an opposite scenario, in which the giant viruses would have originated from simpler entities and which would expand the genome mainly through gene gain and duplication. These analyses used sequence sampling and varied methods of phylogenetic reconstruction and have indicated that the evolution of these viruses would have followed an accordion-like model, where there were losses and gains of genes along the process of evolution. A positive balance of this relation would have occurred, resulting in viruses with extensive genomes and giant particles. Thus, the two opposite scenarios are in force and are constantly debated among specialists.

The accordion hypothesis has gained support from comparative analyses of Polintons/polintoviruses, a superfamily of virus-like DNA transposons commonly identified in the genome of eukaryotes and which have genes homologous to some giant virus

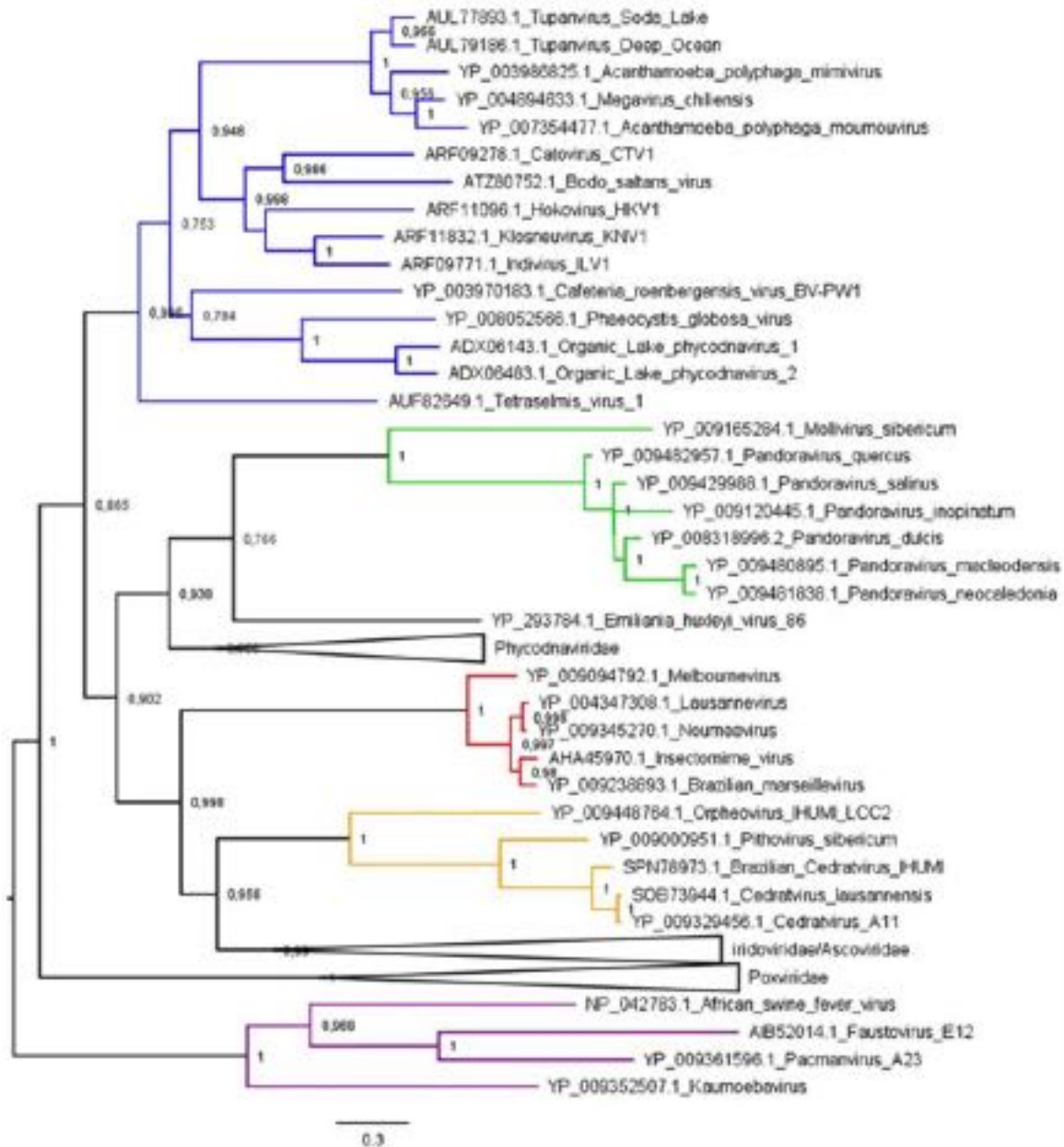


Fig. 5 Phylogeny of NCLDV. Phylogenetic reconstruction based on amino acid sequences of DNA polymerase B family of different members of NCLDV group. Viral families containing viruses not associated with amoebae were collapsed. Different colors represent distinct groups of amoebae giant viruses. Blue: mimiviruses, green: pandoraviruses and mollivirus, red: marseilleviruses, orange: pithoviruses, purple: asfarviridae-related viruses. Alignment was performed using MUSCLE, and the tree was built using the Maximum Likelihood method in FastTree software with 1000 bootstrap replicates. Only bootstrap values > 50 are shown. Scale bar represents rate of evolution.

genes, including those encoding the major capsid protein, the genome packaging ATPase and the virion maturation protease conserved in most NCLDVs. Recent analyses have suggested these mobile elements are the ancestors of NCLDVs, hence the various groups of giant amoeba viruses. These elements form a complex network of evolutionary interactions together with transpovirons and proviophages, elements that are part of the giant virus mobilome, something never before observed in other viruses.

Parasites of Viruses: The Nature of Virophages and the Genetic Mobilome

A hallmark of giant viruses is the presence of their own genetic mobilome, which includes introns, transposable elements and even viruses, which are the so-called virophages. Virophages are a new group of viruses associated with giant viruses, the first isolate of

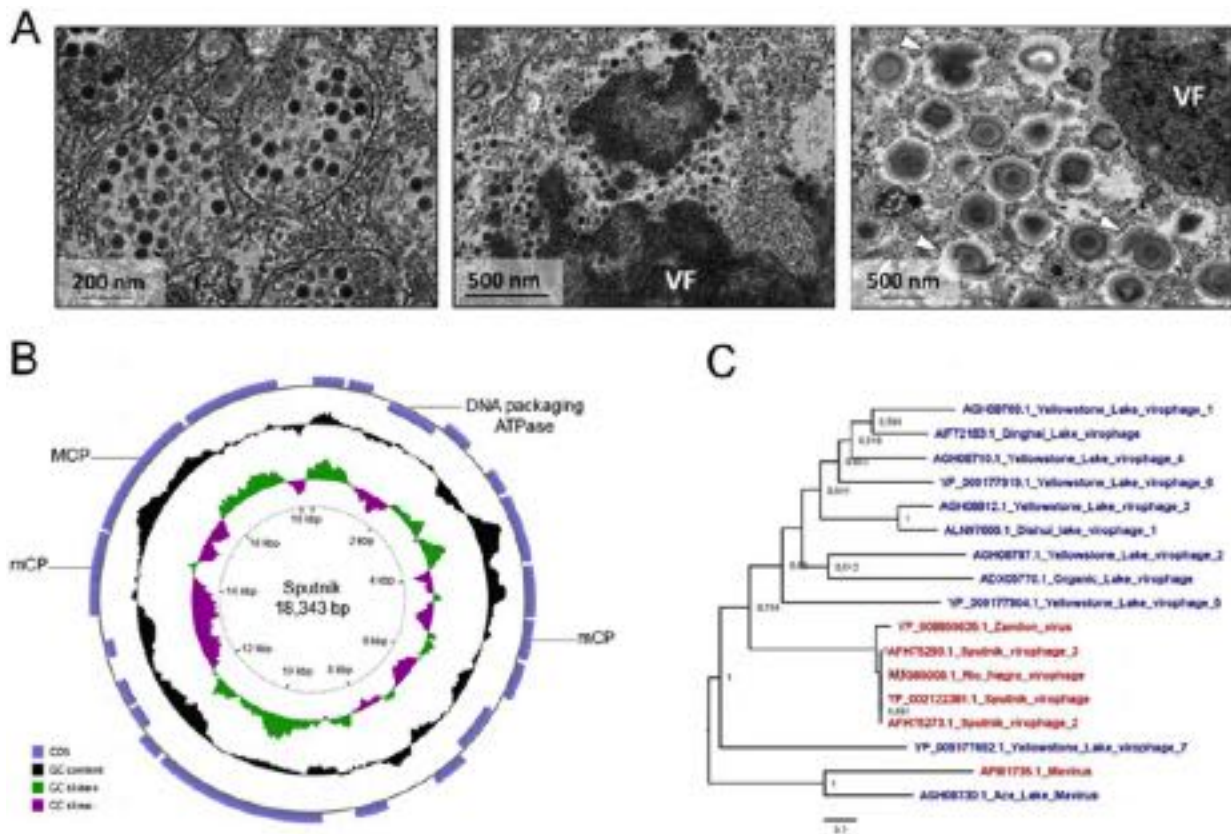


Fig. 6 Virophage characteristics. (A) Transmission electron microscopy images of the Rio Negro virophage, a sputnik-related virus associated with the Samba virus (*Mimiviridae*), showing the virion shape within vesicles (left image), and replication cycle inside viral factory (middle image), and effect on giant virus particles (right image). White arrows indicate defective particles. VF: Viral factory; (B) genome features of the Sputnik virus, the first virophage to be discovered, indicating the genome size, ORF position, GC content and skew, and the conserved genes found in members of the *Lavidaviridae* family; (C) phylogenetic reconstruction based on MCP gene of virophages. Colors represent the virophages identified by culture methods (red) and metagenomics methods (blue). Alignment was performed using MUSCLE and the tree was built using the Maximum Likelihood method in FastTree software with 1000 bootstrap replicates. Only bootstrap values > 50 are shown. Scale bar represents rate of evolution.

which was described in 2008 and was associated with the *Acanthamoeba castellanii* mamavirus, one of the mimivirus isolates. The smaller virus, called Sputnik virus, has icosahedral particles of 75 nm and a circular A/T-rich genome of 18 kb encoding 21 ORFs (Fig. 6(A,B)). This virus was initially observed to be associated with mimivirus fibrils and this was initially suggested to serve as the mode of virophage entry into amoebae cells. However, analyses based on other isolated virophages have suggested that penetration may also occur through clathrin-mediated endocytosis. These viruses replicate in the viral factory of the giant virus, hijacking the transcription machinery of the viral host (Fig. 6(A)). These viruses do not have their own RNA polymerase, and *in silico* analyses have demonstrated the presence of promoter motifs similar to those found in mimivirus, reinforcing the hypothesis that these viruses depend on the giant virus host to have their genomes transcribed, and subsequently, to form the new progeny. The fact that the presence of the associated virus causes the reduction of the multiplication of the giant virus by ~70%, coupled with other peculiar characteristics of these parasites, was suggested to differentiate them from the already known satellite viruses, leading to the proposal of a new category of viruses, the virophages.

In the years following the discovery of the first isolate, other virophages were isolated, such as the Mavirus that is associated with the *Cafeteria roenbergensis* virus, Sputnik 2, Sputnik 3, Zamilon and Rio Negro virus, and all were associated with different isolates of mimiviruses. Interestingly, Sputnik 2 was identified integrated into the genome of a host mimivirus (Lentille virus), which led to the introduction of the term “provirophage.” Recently, the presence of provirophages in the genome of a unicellular algae (*Bigelowiella natans*), whose genes are actively transcribed, has been described, and their presence has led to the hypothesis that they would act as a defense system against giant virus infection. Mavirus was found integrated in multiple sites of the genome of the protozoan *Cafeteria roenbergensis*, and its activation was observed upon CroV infection, which led to the generation of a progeny with subsequent cellular lysis. The newly produced virophages lead to the suppression of the giant virus replication and increased the survival of the cellular host community, reinforcing the hypothesis of an antiviral system triggered by the presence of a provirophage.

Interestingly, the Zamilon viroplage is able to replicate with mimiviruses from lineage B and C, but not when in the presence of viruses from lineage A. Extensive analysis of the genome of different mimivirus isolates revealed that members from lineage A contain the insertion of a repeated Zamilon sequence within an operon, which was named the “mimivirus viroplage resistant element” or

MIMIVIRE. This system is composed of interspaced, repeated virophage sequences, along with nuclease genes, in a way analogous to the CRISPR-Cas system found in prokaryotes. By silencing the MIMIVIRE genes and the repeat sequences, the mimivirus becomes susceptible to the virophage infection, which led to the proposition of a new nucleic-acid-based immunity system against virophage infection. An alternative explanation was further proposed, stating that the defense system is not CRISPR-like, but instead a protein interaction model would be responsible for the phenotype. Recently, a pivotal component in this system, the R354 gene of mimivirus encoding a nuclease, had its structure defined, with analysis evidencing that the viral nuclease is functionally similar to the Cas4 protein, thus providing additional evidence that the MIMIVIRE is a new innate immune defense system.

Several other virophages have had their genomes identified by means of metagenomic analyses in different places of the world, including Yellowstone lake virophages (1–7, USA), Organic lake virophage and Ace lake mavirus (Antarctic), and Dishui lake virophage and Qinghai lake virophage (China). Other virophage genomes and gene fragments have been recovered from metagenomics datasets from Lake Mendota and Trout Bog Lake (Wisconsin, USA), identified as “freshwater virophage candidate genus.” Moreover, virophages have also been identified in datasets from sheep rumen metagenomics, constituting a hybrid group of virophages and polintons (linear genomes with inverted terminal repeats and genes with a strong similarity to polinton genes), named “sheep rumen virophages.” Some of these virophages were associated with algae viruses, which would be responsible for controlling the dynamics between the virus and algae-host in the environment. These new viruses have a genome that ranges from 17 to 30 kb, encoding 16–34 ORFs and containing several genes without known functions, but also containing some conserved genes, such as the major and minor capsid proteins, protease and genome packaging ATPase, which are used as genetic markers for these viruses and for phylogenetic reconstructions (Fig. 6(C)). The presence of at least some of these genes, along with the association or dependence on viruses of the NCLDV group to fully replicate, characterize a virus as a virophage and are thus classified in the family *Lavidaviridae*, which currently includes the genera *Sputnikvirus* and *Mavirus*. Recently, a new putative virophage was discovered associated with a mimivirus infecting *Saccamoeba* cells, which seems to affect the replication of the giant virus host, but no genomic data are available to date. It is possible that this new virophage and others yet to be found constitute new taxa in the family *Lavidaviridae*.

Virophages and provirophages are part of the genetic mobilome of giant viruses. Along with these elements are the transpovirons, a linear sequence of 7 kb capable of integrating randomly into regions of the genomes of giant viruses, and which have been identified in different isolates of the three lineages of the genus *Mimivirus*. In addition, introns have already been detected in genes of different giant viruses, such as mimivirus and faustovirus, and together with inteins complete the known viral mobilome so far. Genomic analyses have indicated similarities between the mobile elements of giant viruses and polintons/mavericks, large transposable DNA elements found in various cellular organisms. Phylogenetic reconstructions suggest that virophages, as well as giant viruses, may have originated from polintoviruses, which constitute a complex network of evolutionary interaction.

Perspectives

Giant viruses and their parasites represent a new and vast field within virology. Since the discovery of the giant mimivirus, new viral groups have been described, constantly expanding our appreciation of the complexity of the virosphere as we know it, and making it necessary to advance the taxonomy of these viruses. New isolation techniques for subsequent viral characterization associated with metagenomic techniques have strongly contributed to the advancement of our understanding of this complex group of eukaryoviruses and their parasites. Finally, studies aimed at better understanding the biology of these viruses are necessary and may reveal new features that could be exploited from the biotechnological point of view, given the enormous genomic and structural variety presented by these viruses, and the singular features of these giants of the virosphere.

Further Reading

- Abergel, C., Legendre, M., Claverie, J.M., 2015. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. *FEMS Microbiology Reviews* 39, 779–796.
- Abrahão, J., Silva, L., Silva, L.S., *et al.*, 2018. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nature Communications* 9, 749.
- Claverie, J.M., Abergel, C., 2009. Mimivirus and its virophage. *Annual Review of Genetics* 43, 49–66.
- Colson, P., Levasseur, A., La Scola, B., *et al.*, 2018. Ancestrality and mosaicism of giant viruses supporting the definition of the fourth TRUC of microbes. *Frontiers in Microbiology* 9, 2668.
- Desnues, C., La Scola, B., Yutin, N., *et al.*, 2012. Provirophages and transpovirons as the diverse mobilome of giant viruses. *Proceedings of the National Academy of Sciences of United States of America* 109, 18078–18083.
- Filée, J., 2013. Route of NCLDV evolution: The genomic accordion. *Current Opinion in Virology* 3, 595–599.
- Krupovic, M., Kuhn, J.H., Fischer, M.G., 2016. A classification system for virophages and satellite viruses. *Archives of Virology* 161, 233–247.
- La Scola, B., Desnues, C., Pagnier, I., *et al.*, 2008. The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104.
- Raoult, D., Audic, S., Robert, C., *et al.*, 2004. The 1.2-Megabase genome sequence of Mimivirus. *Science* 306, 1344–1350.
- Xiao, C., Kuznetsov, Y.G., Sun, S., *et al.*, 2009. Structural studies of the giant Mimivirus. *PLoS Biology* 7, 0958–0966.
- Yutin, N., Raoult, D., Koonin, E.V., 2013. Virophages, polintons, and transpovirons: A complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. *Virology Journal* 10, 1.
- Yutin, N., Wolf, Y.I., Koonin, E.V., 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* 466–467, 38–52.

Viral Replication Cycle

Alan J Cann, University of Leicester, Leicester, United Kingdom

© 2021 Elsevier Ltd. All rights reserved

This is a reproduction of A.J. Cann, Replication of Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00486-6.

Glossary

(–)-sense RNA (minus-sense RNA) A virus with a single-stranded RNA genome of the opposite polarity ('sense') as mRNA.

(+)-sense RNA (plus-sense RNA) A virus with a single-stranded RNA genome of the same polarity ('sense') as mRNA.

Assembly The stage of replication during which all the structural components come together at one site in the cell and the basic structure of the virus particle is formed.

Attachment The binding of a virus particle to a specific receptor on the surface of a host cell.

Capsid A protein shell comprising the main structural unit of a virus particle.

Envelope A lipid membrane enveloping a virus particle.

Fusion protein The protein(s) on the surface of a virus particle responsible for fusion of the virus envelope with cellular membranes.

Gene expression An important stage of viral replication at which virus genetic information is expressed: one of the major control points in replication.

Genome replication The stage of viral replication at which the virus genome is copied to form new progeny genomes.

Matrix protein A structural protein of a virus particle which underlies the envelope and links it to the core.

Maturation The stage of viral replication at which a virus particle becomes infectious.

Molecular epidemiology The use of nucleotide sequence information to study the diversity and distribution of virus populations.

mRNA Messenger RNA, translated on ribosomes to produce proteins.

Nucleocapsid The core of a virus particle consisting of the genome plus a complex of proteins.

Penetration The stage of viral replication at which the virus genome enters the cell.

Polyprotein A long polypeptide encoding several mature proteins which are subsequently released by protease cleavage.

Receptor A specific molecule on the surface of a cell which is used by a virus for attachment.

Release The stage of viral replication at which virus particles escape the infected cell.

Tropism The ability of a virus to infect specific cell or tissue types.

Uncoating The stage of viral replication at which structural proteins are lost and the virus genome is exposed to the replication machinery.

Virions Structurally mature, extracellular virus particles.

Virus attachment protein The protein on the surface of a virus particle responsible for binding the receptor.

Unlike cellular organisms, which 'grow' from an increase in the integrated sum of their components and reproduce by division, virus particles are produced from the assembly of preformed components. Once manufactured, virus particles (virions) do not grow or undergo division. This alone makes the process of virus replication distinct from the growth of all other biological agents, and although the term 'grow' is sometimes used in the vernacular to refer to propagation of viruses, it is best to avoid this word when referring to the processes of virus replication.

Although this article will attempt to paint a general picture of the process of virus replication, the type of host cell infected by the virus has a profound effect on the replication process. There are many examples of viruses undergoing different replicative cycles in different cell types. However, the coding capacity of the genome determines the basic replication strategy used by different viruses. This strategy may involve heavy reliance on the host cell, in which case the virus genome can be very compact and need only encode the essential information for a few proteins, for instance, in parvoviruses. Alternatively, large and complex virus genomes, such as those of poxviruses, encode most of the information necessary for replication, and the virus is only reliant on the cell for the provision of energy and the apparatus for macromolecular synthesis, such as ribosomes. Viruses with RNA genomes have no apparent need to enter the nucleus, although during the course of replication, some do. DNA viruses, as might be expected, mostly replicate in the nucleus where host cell DNA is replicated and the biochemical apparatus necessary for this process is located. However, some viruses with DNA genomes (e.g., poxviruses) have evolved to contain sufficient biochemical capacity to be able to replicate in the cytoplasm, with minimal requirement for host cell functions.

Virus replication can be divided into eight stages, as shown in Fig. 1. It should be emphasized that these are arbitrary divisions, used here for convenience in explaining the replication cycle of a theoretical, 'typical' virus. Regardless of their hosts, all viruses must undergo each of these stages in some form to successfully complete their replication cycle. Not all the steps described here are detectable as distinct stages for all viruses; often they blur together and appear to occur almost simultaneously. Some of the individual stages have been studied in great detail and a considerable amount of information is known about them. Other stages have been much harder to study, and less information is available.

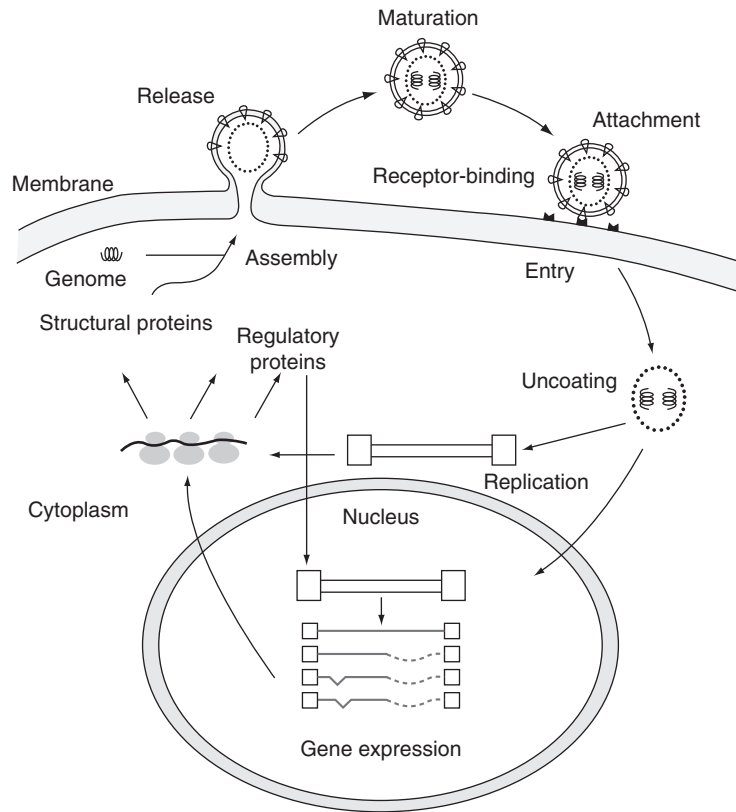


Fig. 1 Schematic overview of a generalized scheme of virus replication. Reproduced from Cann AJ (2004) *Principles of Molecular Virology*, 4th edn. Amsterdam: Elsevier, with permission from Elsevier.

Attachment

The attachment phase of replication comprises specific binding of a virus-attachment protein (or 'antireceptor') to a cellular receptor molecule. Virus receptors on cell surfaces may be proteins (usually glycoproteins) or carbohydrate residues present on glycoproteins or glycolipids. Some complex viruses (e.g., in the *Poxviridae* or *Herpesviridae*) use more than one receptor and therefore have alternative routes of uptake into cells. Most bacteriophage receptors are on the bacterial cell wall, although certain phages use cellular appendages (pili, flagella) as primary adsorption sites. Attachment is an automatic docking process and the kinetics of receptor binding are controlled by the chemical and thermodynamic characteristics of the molecules involved, that is, their relative concentrations and availability.

In most cases, the expression (or absence) of receptors on the surface of host cells determines the tropism of a particular virus, that is, the types of cell in which it is able to replicate. The attachment phase of infection therefore has a major influence on viral pathogenesis and in determining the course of a virus infection. Plant viruses must overcome different problems to animal viruses in initiating infection. The outer surfaces of plants are composed of protective layers of waxes and pectin, and each cell is surrounded by a thick wall of cellulose overlying the cytoplasmic membrane. No known plant virus uses a specific cellular receptor of the type that animal and bacterial viruses use to attach to cells and plant viruses must rely on mechanical breaks in the cell wall to directly introduce a virus particle into a cell.

Some virus receptors consist of more than one protein and multiple interactions are required for virus entry. An example of this is human immunodeficiency virus-1 (HIV-1), the primary receptor for which is the T-cell antigen, CD4. The binding site for the HIV-1 attachment protein (antireceptor), gp 120, has been mapped to the first variable region of CD4, although additional amino acids of the second variable domain also contribute toward binding. The sequences important for CD4 binding have also been mapped in gp120. Deletions in this region or site substitutions abolish CD4 binding. In addition to CD4, there is at least one accessory factor which is necessary to form a functional HIV-1 receptor. These factors have now been identified as a family of proteins known as β -chemokine receptors. Multiple members of this class of proteins have been shown to play a role in the entry of HIV-1 into cells, and their distribution in the body is the primary control for the tropism of HIV-1 for different cell types.

Occasionally, the specificity of receptor binding can be subverted by nonspecific interactions between virus particles and host cells. Virus particles may be taken up by cells by pinocytosis or phagocytosis. However, without some form of physical interaction which holds the virus particle in close association with the cell surface, the frequency of these events would be very low. In addition, the fate of viruses absorbed into endocytic vacuoles is usually to be degraded, except in cases where the virus particle enters cells by this route.

On occasion, binding of antibody-coated virus particles to Fc receptor molecules on the surface of monocytes and other blood cells can result in virus uptake. The presence of antiviral antibodies can result in increased virus uptake by cells and increased pathogenicity, rather than virus neutralization, as would normally be expected. The significance of such mechanisms *in vivo* is not known.

Entry

Entry of the virus particle into the host cell normally occurs a short time after attachment of the virus to the receptor. Unlike attachment, cell entry is generally an energy-dependent process, that is, the cell must be metabolically active for this to occur. Three main mechanisms are observed:

1. Translocation of the entire virus particle across the cytoplasmic membrane of the cell. This process is relatively rare among viruses and is poorly understood. It is mediated by proteins in the virus capsid and specific membrane receptors.
2. Endocytosis of the virus into intracellular vacuoles. This is probably the most common mechanism of virus entry into cells. It does not require any specific virus proteins (other than those already utilized for receptor binding) but relies on the normal formation and internalization of coated pits (term to be explained) at the cell membrane. Receptor-mediated endocytosis is an efficient process for taking up and concentrating extracellular macromolecules.
3. Fusion of the virus envelope (where present) with the cell membrane, either directly at the cell surface or following endocytosis in a cytoplasmic vesicle. Fusion requires the presence of a specific fusion protein in the virus envelope, for example, influenza A virus hemagglutinin or the transmembrane glycoproteins of retroviruses. These proteins promote the joining of the cellular and virus membranes which results in the nucleocapsid being deposited directly in the cytoplasm. There are two types of virus-driven membrane fusions: pH-dependent and pH-independent.

The process of endocytosis is almost universal in animal cells and requires the formation of clathrin-coated pits which results in the engulfment of a membrane-bounded vesicle by the cytoplasm of the cell. At this point, any virus contained within these structures is still cut off from the cytoplasm by a lipid bilayer and therefore has not strictly entered the cell. As endosomes fuse with lysosomes, the environment inside these vessels becomes progressively more hostile as they are acidified and the pH falls, while the concentration of degradative enzymes rises. This means that the virus must leave the vesicle and enter the cytoplasm before it is degraded. There are a number of mechanisms by which this occurs, including membrane fusion and rescue by transcytosis. The release of virus particles from endosomes and their passage into the cytoplasm is intimately connected with (and often impossible to separate from) the process of uncoating.

Uncoating

Uncoating describes the events which occur after host cell entry, during which the virus capsid is partially or completely degraded or removed and the virus genome exposed, usually still in the form of a nucleic acid–protein complex. Uncoating occurs simultaneously with or immediately after entry and is thus difficult to study. In bacteriophages which inject their genome directly into the cell, entry and uncoating are the same process.

The removal of a virus envelope during membrane fusion is the initial stage of the uncoating process for enveloped viruses. Uncoating may occur inside endosomes, being triggered by the change in pH as the endosome is acidified, or directly in the cytoplasm. Entry into the endocytic pathway is a hazardous process for viruses because if they remain in the vesicle too long, they will be irreversibly damaged by low pH or lysosomal enzymes. Hence, some viruses have evolved proteins to control this process; for example, the influenza A virus M2 protein is a membrane channel which allows entry of hydrogen ions into the nucleocapsid, facilitating uncoating. The M2 protein is multifunctional, and also has a role in virus uncoating. In the piconaviruses, penetration of the cytoplasm by exit of virus from endosomes is tightly linked to uncoating. The acidic environment of the endosome causes a conformational change in the particle at around pH 5 that reveals hydrophobic domains not present on the surface of mature virus capsids. These hydrophobic patches interact with the endosomal membrane and form pores through which the RNA genome passes into the cytoplasm of the host cell.

The ultimate product of uncoating depends on the structure of the virus genome/nucleocapsid. In some cases, the resulting structure is relatively simple; for example, piconaviruses have only a small basic protein of approximately 23 amino acids covalently attached to the 5' end of the RNA genome. In other cases, the virus core which remains is highly complex; for example, in the poxviruses uncoating occurs in two stages – removal of the outer membrane as the particle enters the cell and in the cytoplasm, followed by further uncoating as the core passes into the cytoplasm. In this case, the core still contains dozens of proteins and at least 10 distinct enzymes.

The structure and chemistry of the nucleocapsid determines the subsequent steps in replication. Reverse transcription can only occur inside an ordered retrovirus core particle and does not proceed to completion with the virus RNA free in solution. Eukaryotic viruses which replicate in the nucleus, such as members of the *Herpesviridae*, *Adenoviridae*, and *Polyomaviridae*, undergo structural changes following penetration, but overall remain largely intact. This is important because these capsids contain nuclear localization sequences responsible for attachment to the cytoskeleton and this interaction allows the transport of the entire capsid to the nucleus. At the nuclear pores, complete uncoating occurs and the nucleocapsid passes into the nucleus.

Transcription and Genome Replication

The replication strategy of a virus depends, in large part, on the structure and composition of its genome. For viruses with RNA genomes in particular, genome replication and transcription are often inextricably linked, and frequently carried out by the same enzymes. Therefore, it makes most sense to consider both of these aspects of virus replication together.

Group I: Double-Stranded DNA

This class can be further subdivided into two as follows:

1. Replication is exclusively nuclear or associated with the nucleoid of prokaryotes. The replication of these viruses is relatively dependent on cellular factors. In some cases, no virus-encoded enzymes are packaged within these virus particles as this is not necessary, whereas in more complex viruses numerous enzymatic activities may be present within the particles.
2. Replication occurs in cytoplasm. These viruses have evolved (or acquired from their hosts) all the necessary factors for transcription and replication of their genomes and are therefore largely independent of the cellular apparatus for DNA replication and transcription. Because of this independence from cellular functions, these viruses have some of the largest and most complex particles known, containing many different enzymes.

Group II: Single-Stranded DNA

The replication of these virus genomes occurs in the nucleus, involving the formation of a double-stranded intermediate which serves as a template for the synthesis of new single-stranded genomes. In general, no virus-encoded enzymes are packaged within the virus particle since most of the functions necessary for replication are provided by the host cell.

Group III: Double-Stranded RNA

These viruses all have segmented genomes, as each segment is transcribed separately to produce individual monocistronic messenger RNAs. Replication occurs in the cytoplasm and is largely independent of cellular machinery, as the particles contain many virus-encoded enzymes essential for RNA replication and transcription since these processes (involving copying RNA to make further RNA molecules) do not normally occur in cellular organisms.

Group IV: Single-Stranded (+)-Sense RNA

These viruses can be subdivided into two groups.

1. *Viruses with polycistronic mRNA such as flaviviruses and picornaviruses.* As with all the viruses in this group, the genome RNA represents mRNA which is translated after infection, resulting in the synthesis of a polyprotein product, which is subsequently cleaved to form the mature proteins.
2. *Viruses with complex transcription such as coronaviruses and togaviruses.* In this subgroup, two rounds of translation are required to produce subgenomic RNAs which serve as mRNAs in addition to the full-length RNA transcript which forms progeny virus genomes. Although the replication of these viruses involves copying RNA from an RNA template, no virus-encoded enzymes are packaged within the genome since the ability to express genetic information directly from the genome without prior transcription allows the virus replicase to be synthesized after infection has occurred.

Group V: Single-Stranded (-)-Sense RNA

The genomes of these viruses can also be divided into two types.

- *Segmented.* The first step in the replication of these viruses (e.g., orthomyxoviruses) is transcription of the (-)-sense RNA genome by the virion RNA-dependent RNA polymerase packaged in virus particles to produce monocistronic mRNAs, which also serve as the template for subsequent genome replication.
- *Nonsegmented.* Monocistronic mRNAs for each of the virus genes are produced by the virus transcriptase in the virus particle from the full-length virus genome. Subsequently, a full-length (+)-sense copy of the genome is synthesized which serves as a template for (-)-sense progeny virus genomes (e.g., paramyxoviruses and rhabdoviruses).

Group VI: Single-Stranded RNA with DNA Intermediate

Retrovirus genomes are composed of (+)-sense RNA but are unique in that they are diploid and do not serve directly as mRNA but as a template for reverse transcription into DNA. A complete replication cycle involves conversion of the RNA form of the virus genetic material into a DNA form, the provirus, which is integrated into the host cell chromatin. The enzyme reverse transcriptase needs to be packaged into virus particles to achieve this conversion, as virus genes are only expressed from the DNA provirus and not from the RNA genome found in retrovirus particles of retroviruses.

Group VII: Double-Stranded DNA with RNA Intermediate

This group of viruses also relies on reverse transcription, but unlike the retroviruses, this occurs inside the virus particle during maturation. On infection of a new cell, the first event to occur is repair of the gapped genome, followed by transcription. As with group VI viruses, a reverse transcriptase enzyme activity is present inside virus particles, but in this case, the enzyme carries out the conversion of virus RNA into the DNA genome of the virus inside the virus particle. This contrasts with retroviruses where reverse transcription occurs after the RNA genome has been released from the virus particle into the host cell.

Assembly

During assembly, the basic structure of the virus particle is formed as all the components necessary for the formation of the mature virion come together at a particular site in the cell. The site of assembly depends on the pattern of virus replication and the mechanism by which the virus is eventually released from the cell and so varies for different viruses. Although some DNA virus particles form in the nucleus, the cytoplasm is the most common site of particle assembly. In the majority of cases, cellular membranes are used to anchor virus proteins, and this initiates the process of assembly.

For enveloped viruses, the lipid covering is acquired through a process known as budding, where the virus particle is extruded through a cell membrane. Lipid rafts are membrane microdomains enriched in glycosphingolipids (or glycolipids), cholesterol and a specific set of associated proteins. Lipid rafts have been implicated in a variety of cellular functions, such as apical sorting of proteins and signal transduction, but they are also used by viruses as platforms for cell entry (e.g., for HIV-1, SV40, and the rotaviruses), and as sites for particle assembly, budding and release from the cell membrane (e.g., in influenza A virus, HIV, measles virus, and rotaviruses).

As with the earliest stages of replication, it is often not possible to identify the assembly, maturation, and release of virus particles as distinct and separate phases. The site of assembly has a profound influence on all these processes. In general terms, rising intracellular levels of virus proteins and genomes reach a critical concentration and this triggers assembly. Many viruses achieve high levels of newly synthesized structural components by concentrating these into subcellular compartments known as inclusion bodies. These are a common feature of the late stages of infection of cells by many different viruses. Alternatively, local concentrations of virus structural components can be boosted by lateral interactions between membrane-associated proteins. This mechanism is particularly important in enveloped viruses which are released from the cell by budding (see above).

Maturation

Maturation is the stage of the replication cycle at which virus particles become infectious. This often involves structural changes in the newly formed particle resulting from specific cleavages of virus proteins to form the mature products or from conformational changes in proteins which occur during assembly (e.g., hydrophobic interactions). Protein cleavage frequently leads to substantial structural changes in the capsid. Alternatively, internal structural alterations, for example, the condensation of nucleoproteins with the virus genome, often result in changes visible by electron microscopy.

Proteases are frequently involved in maturation, and virus-encoded enzymes, cellular proteases or a mixture of the two may be used. Virus-encoded proteases are usually highly specific for particular amino acid sequences and structures, only cutting a particular peptide bond in a particular protein. Moreover, they are often further controlled by being packaged into virus particles during assembly and only activated when brought into close contact with their target sequence by the conformation of the capsid, for example, by being placed in a local hydrophobic environment, or by changes of pH or cation cofactor concentrations inside the particle as it forms. Retrovirus proteases are good examples of enzymes involved in maturation which are under tight control. The retrovirus core particle is composed of proteins from the *gag* gene and the protease is packaged into the core before its release from the cell on budding. During the budding process, the protease cleaves the *gag* protein precursors into the mature products – the capsid, nucleocapsid, and matrix proteins of the mature virus particle. Other protease cleavage events involved in maturation are less closely controlled. Influenza A virus hemagglutinin must be cleaved into two fragments (HA₁ and HA₂) to be able to promote membrane fusion during infection. Cellular trypsin-like enzymes are responsible for this process, which occurs in secretory vesicles as the virus buds into them prior to release at the cell surface; however, this process is controlled by the virus M2 protein, which regulates the pH of intracellular compartments in influenza virus-infected cells.

Release

For lytic viruses (most nonenveloped viruses), release is a simple process – the infected cell breaks open and releases the virus. The reasons for lysis of infected cells are not always clear, but virus-infected cells often disintegrate because viral replication disrupts normal cellular function, for example, the expression of essential genes. Many viruses also encode proteins that stimulate (or in some cases suppress) apoptosis, which can also result in release of virus particles.

Enveloped viruses acquire their lipid membrane as the virus buds out of the cell through the cell membrane, or into an intracellular vesicle prior to subsequent release. Virion envelope proteins are picked up during this process as the virus particle is extruded. This process is known as budding. As mentioned earlier, assembly, maturation, and release are usually simultaneous processes for viruses which are released by budding. The release of mature virus particles from their host cells by budding presents a problem in that these particles are designed to enter, rather than leave, cells. Certain virus envelope proteins are involved in the release phase of replication as well as in receptor binding. The best-known example of this is the neuraminidase protein of influenza virus. In addition to being able to reverse the attachment of virus particles to cells via hemagglutinin, neuraminidase is also believed to be important in preventing the aggregation of influenza A virus particles and may well have a role in virus release. In addition to using specific proteins, viruses which bud have also solved the problem of release by the careful timing of the assembly-maturation-release pathway. Although it may not be possible to separate these stages by means of biochemical analysis, this does not mean that spatial separation of these processes has not evolved as a means to solve this problem.

Further Reading

- Cann, A.J., 2004. *Principles of Molecular Virology*, 4th edn. Amsterdam: Elsevier.
- Freed, E.O., 2004. HIV-1 and the host cell: An intimate association. *Trends in Microbiology* 12, 170–177.
- Kasamatsu, H., Nakanishi, A., 1998. How do animal DNA viruses get to the nucleus? *Annual Review of Microbiology* 52, 627–686.
- Lopez, S., Arias, C.F., 2004. Multistep entry of rotavirus into cells: A Versaillesque dance. *Trends in Microbiology* 12, 271–278.
- Moore, J.P., Kitchen, S.G., Pugach, P., Zack, J.A., 2004. The CCR5 and CXCR4 coreceptors – central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. *AIDS Research and Human Retroviruses* 20, 111–126.
- Rossmann, M.G., He, Y., Kuhn, R.J., 2002. Picornavirus–receptor interactions. *Trends in Microbiology* 10, 324–331.
- Schneider-Schaulies, J., 2000. Cellular receptors for viruses: Links to tropism and pathogenesis. *Journal of General Virology* 81, 1413–1429.

Viral Receptors

José M Casasnovas, National Center for Biotechnology, Spanish National Research Council (CSIC), Madrid, Spain

Thilo Stehle, University of Tuebingen, Tuebingen, Germany and Vanderbilt University School of Medicine, Nashville, TN, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of D.J. Evans, Viral Receptors, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00531-8.

Nomenclature

Ab	Antibody	IgSF	Immunoglobulin Superfamily
ACE2	Angiotensin-converting enzyme 2	IgV	Ig-like variable
Ad	Adenovirus	JAM-A	Junctional adhesion molecule-A
APN	Aminopeptidase N	JCPyV	JC polyomavirus
BKPyV	BK polyomavirus	LDLR	low density lipoprotein receptor
CAR	Coxsackievirus and adenovirus receptor	LSTa	Lactoseries sialyl tetrasaccharide a
CAV	Coxsackie A virus	LSTc	Lactoseries sialyl tetrasaccharide c
CCP	Complement control proteins	MERS	Middle east respiratory syndrome
CoV	Coronavirus	MV	Measles virus
cryo-EM	Cryo-electron microscopy	Neu5Ac	5-N-acetylneuraminic acid
DAF	Decay-accelerating factor	PtdSer	Phosphatidylserine
DPP4	Dipeptidyl peptidase 4	PV	Poliovirus
EBV	Epstein–Barr virus	PVR	Poliovirus receptor
EchoV	Echovirus	PyV	Polyomavirus
EV	Enterovirus	RBD	Receptor binding domain
GAG	Glycosaminoglycan	RCA	Regulators of complement activity
GalNAc	N-acetylgalactosamine	S	Spike
GlcNAc	N-acetylglucosamine	SARS	Severe acute respiratory syndrome
HBGA	Histo-blood group antigen	SCR	Short consensus repeats
HRV	Human rhinovirus	Sia	Sialic acid
HS	Heparan sulfate	SLAM	Signaling lymphocyte-activation molecule
ICAM-1	Intercellular adhesion molecule-1	TIM	Transmembrane, Immunoglobulin and Mucin domain

Glossary

Attachment factor Cell surface molecule that allows a virus to adhere to a cell, but it is dispensable for virus cell entry.

Entry receptor Cell membrane molecule necessary for virus penetration into cells and that participates in the virus cell entry process. Its expression makes a cell susceptible to virus infection.

Lipidome (virus) Set of lipids in a virus envelope membrane.

Metastable Refers to virus particles that are in a relatively high energy state, and capable of transitioning to more stable conformations.

Uncoating The process by which the virus genome exits from a virus particle to be transferred inside of the cell for its replication and generation of the virus progeny. Although it is commonly applied to non-enveloped viruses, virus-cell fusion can be considered an uncoating process.

Introduction

The majority of viruses have an extracellular phase in their life cycle, which is a necessary part of virus transmission and dissemination. Virus particles are metastable structures that transport the viral genome from cell to cell. To initiate the replication cycle for the production of progeny virions, the virus must first enter a cell. Cell entry is mediated by the specific interaction of the virus with molecules on the cell surface – the receptor – resulting in attachment, and the subsequent internalization of the virus particle, or the genome. A cell that expresses the cognate cell surface receptor for a virus is termed susceptible. However, not all susceptible cells to which the virus binds are capable of supporting infection and replication, perhaps due to the absence of necessary intracellular components. Therefore, cells that support virus binding, entry, replication, and progeny virion release are also termed permissive.

With certain notable exceptions, such as viruses of fungi that exhibit no extracellular stage in the life cycle or plant viruses introduced mechanically by insect vectors, the virus receptor is the primary determinant of virus tropism. The identification and

characterization of the receptor therefore provides important insights into the early stages of the virus life cycle, and has been a major research focus of molecular virologists worldwide.

Virus Receptors in Virus Cell Entry

Our enhanced understanding of virus–receptor interactions and the early events in cell entry have demonstrated that this process is, in many cases, considerably more complicated than originally thought, involving multiple cell-surface or intracellular molecules with different roles in the entry events. A classical definition of a virus receptor would be the cell-surface molecule that mediates virus attachment and cell entry. Virus receptor expression in a cell is necessary for virus infection. It should be distinguished from attachment factors, which are dispensable for cell entry and whose expression can enhance infection but it does not render a cell susceptible. Our current understanding of these processes demonstrates that many viruses use different molecules to mediate attachment and post-attachment entry events. For this reason, the term co-receptor is often used to indicate an accessory molecule implicated in cell membrane penetration. Some viruses can use several receptors for cell entry or evolve to use alternate receptors, which changes their tropism, facilitates host-to-host transmission and generates zoonosis.

Attachment

Viruses initially bind to cell surface receptors or attachment factors. Receptor binding is more virus-specific than binding to attachment factors. Numerous viruses attach to lectins using virus envelope carbohydrates or benefit from envelope phosphatidylserine (PtdSer) to bind to PtdSer-binding proteins. Nonetheless, virus–receptor interactions are determined by unique virus protein surfaces, which restrict receptor selection. Although highly distinct viruses can select common receptors, some of which are abundant on the cell surface, there is a large diversity in receptor usage due to differences in virus structures. In many viruses, receptor binding surfaces have low accessibility, which decreases their association binding rates and the monovalent affinity. Virus binding sites for glycan receptors are typically more surface-exposed, with low affinities that often lie in the millimolar range. In all cases, the presence of many identical binding sites on a virus allows for multivalent binding, and this can increase its avidity for the cell by several orders of magnitude.

Receptor-mediated virus uncoating

Receptors can play an active role during cell entry, initiating the translocation of the genome from the virus particle to the cytosol. Virus binding to receptor molecules can mediate conformational changes in the particles that initiate genome uncoating, which require multivalent receptor engagement. In enveloped viruses, receptors also facilitate the engagement of co-receptors for membrane fusion. Mature virions and virus proteins are susceptible to temperature-induced motions, which generate metastable virus conformations. The receptors can thus sustain certain virus states primed for uncoating at physiological temperatures without external energy. In line with this, it is accepted that certain receptors function as catalyzers, maintaining metastable virus structures that facilitate genome release, with or without involvement of other host factors such as the low endosomal pH.

Receptor-mediated endocytosis

Using receptor molecules, viruses can subvert the endocytic machinery to gain access into the cell using a variety of mechanisms. In addition, virus-mediated receptor crosslinking can trigger signaling events that regulate endocytosis. Certain receptors move to endosomes and ferry bound viruses to intracellular compartments, where the viruses come across a pH gradient (7 to 5) that in many cases initiates genome translocation. Mild low pH (5.5–6.0) is a relevant trigger of membrane penetration events, such as the fusion of virus and cell membranes that commonly occurs at endosomes. Even viruses that do not need low pH for entry penetrate in endosomes. Others use endocytosis to access entry receptors located in the endolysosomal system.

The Identification of Virus Receptors

Since the virus receptor is a cell-specific molecule involved in cell entry, the identification of the receptor requires demonstrating that a cell becomes susceptible to a virus upon expression of the receptor. In cases where several receptors are required for cell infection, detecting virus attachment can be used to identify tentative virus receptors.

Several approaches have identified virus receptors based on the detection of virus binding, which can require labeled viruses or antibodies (Abs) against virus proteins. Indirect methods either use Abs that target the receptor, proteases or glycosidases to block binding. Direct methods use virus proteins to identify and/or purify cell surface receptors. Recent advances in glycan microarray screening technology have accelerated the identification of glycan receptors for many viruses. Once the attachment molecule is identified, its ability to support virus cell entry or infection is assessed. Genetic methods are based on the transformation of a non-susceptible cell with DNA encoding the protein. For these methods, cDNA libraries from permissive cells are used. Following expression on the cell surface, the acquisition of a virus binding phenotype is tested, perhaps combined with the demonstration of specificity by the inhibition with monoclonal Abs. As an example, a novel CRISPR-Cas9 based screen has been recently followed to identify an alphavirus receptor. Typically, a combination of the approaches outlined above is necessary to unequivocally demonstrate the identity of a virus receptor. Multiple confirmatory approaches are required as each technique has inherent uncertainties.

Structural knowledge on the interactions between viruses and protein receptors is available for many viruses. In some cases, such knowledge has led to sophisticated models involving conformational changes in viral proteins and receptors as a result of receptor engagement. By contrast, the role of protein-glycan interactions in viral attachment and entry is less well understood, in part because cell-surface glycans form a heterogeneous mixture of complex carbohydrate moieties that are difficult to classify. For many viruses, only fragments of glycan receptors such as terminal sialic acid (Sia), sialyllactose, or sulfated oligosaccharides have been identified, and it is entirely unknown to which cell-surface glycoconjugates these fragments belong. As a consequence, it is largely unclear how glycan-binding impacts post-attachment events in the life cycle of most viruses, such as cell entry and viral uncoating.

Rather than aim to provide a complete catalog of viruses and receptors, the focus of this article is to use key examples to illustrate cell surface receptor families engaged in virus cell entry. The examples used below illustrate a range of different types of viral receptors, either carbohydrates or proteins, and those in which multiple receptors and co-receptors have been implicated. Due to their pivotal role in determining tissue and host tropism, and hence influence on pathogenesis, particular efforts have been made in identifying receptors for important human and animal virus pathogens, some of which are described here.

Glycan Receptors, Interactions with Viruses

Many viruses attach to glycan structures during the first step of an infection (**Fig. 1**). In some cases, the binding of carbohydrates can prime the virus particle for conformational changes. Glycan-virus interactions also affect other aspects of viral infections, such as the exit of virus particles from host cells and the efficiency of viral spread, illustrated by influenza viruses. For most viruses, the biological and medical consequences of the glycan-virus interplay, such as viral tropism, cell entry and pathogenesis, are not well understood. The use of glycans as biological tools or as targets for drug discovery has likewise been hindered by the inability to easily synthesize, sequence and study their biological functions. Below, we review examples of viruses binding to sialic acids, histo-blood group antigens (HBGAs) and glycosaminoglycans (GAGs) (**Fig. 1**).

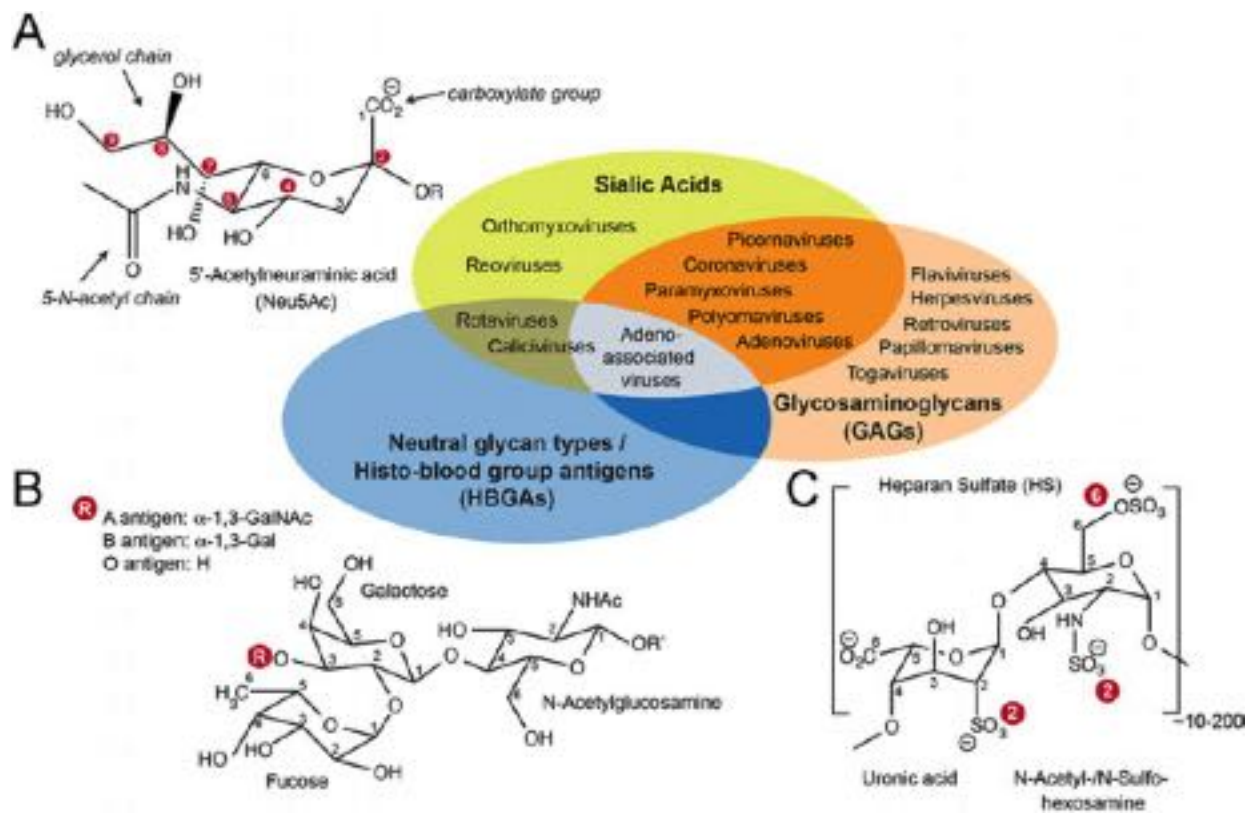


Fig. 1 Glycans as viral receptors. Host- and cell-specific variants of sialic acids (**A**), neutral oligosaccharides (**B**) such as histo-blood group antigens (HBGAs) and glycosaminoglycans (GAGs) (**C**) used as receptors/attachment factors by many viruses. Examples of virus families with members that are known to interact with one or more glycan types are indicated in the respective overlapping ellipses of the Venn diagram.

A. The most common form of sialic acid in humans, N-Acetylneuraminic acid, is shown in the α -conformation. This monosaccharide often caps N-, O-glycans and glycolipids (R), which are attached via different glycosidic linkages to position C2. Various modifications can occur at carbon positions C4, C5 and C7-C9 (marked with red spheres). **B.** Schematic representation of the type 2 core antigen, which is part of the ABO blood group determinants on N- and O-glycans. R: heterosaccharide core attached to a glycoprotein. **C.** Heparan sulfate (HS), an example for GAGs. HS domains along the oligosaccharide chain usually vary in their modification contents by N-sulfation/N-acetylation at carbon positions C2 and C6 of the hexosamine, and by sulfation at C2 of the uronic acid.

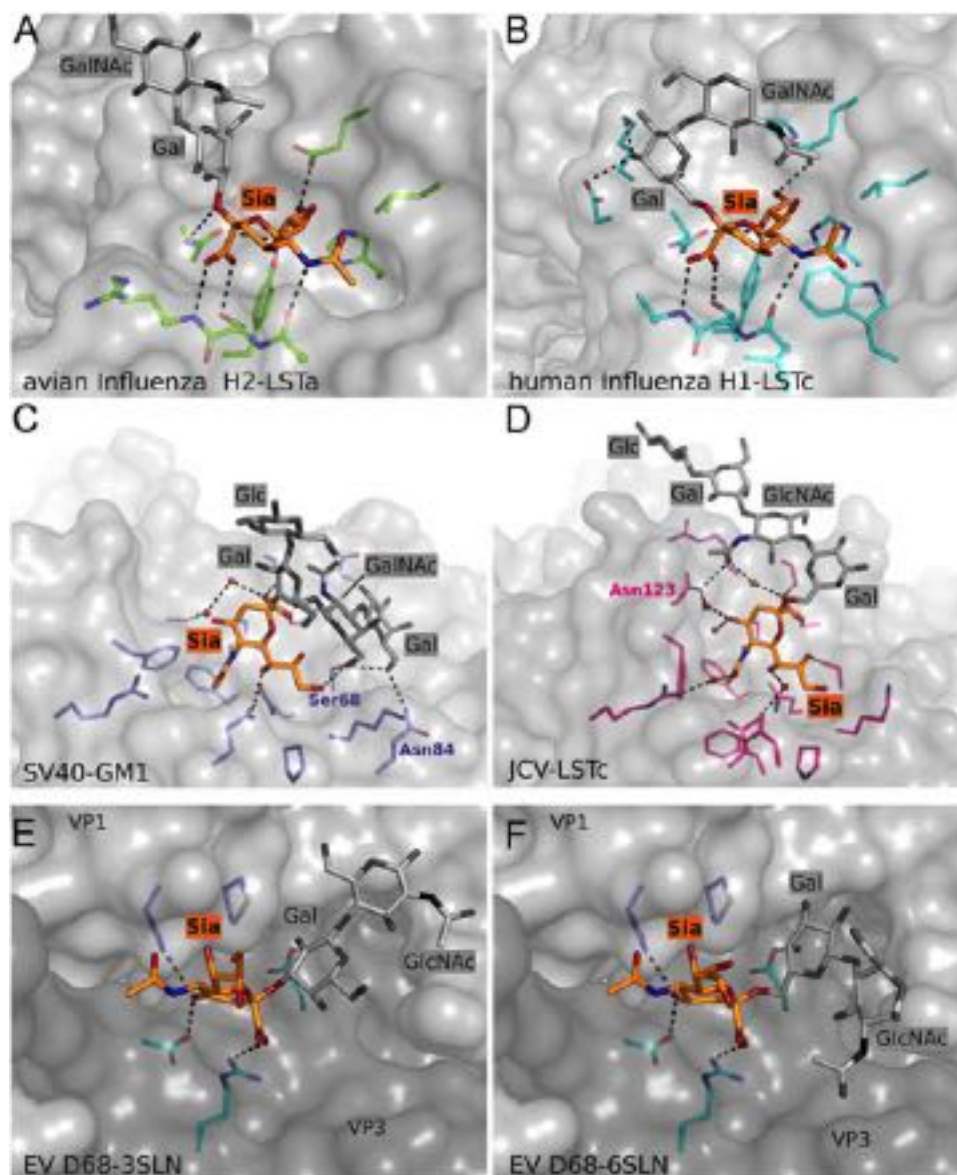


Fig. 2 Viruses engage different sialic acid conformations. Examples of how closely related (or identical) viruses engage α 2,3-linked (left column) or α 2,6 linked (right column) sialylated glycans. **A, B.** Close-up views of the glycan-binding sites of avian (PDB ID 2WR2) and human (PDB ID 1RVZ) influenza virus haemagglutinins bound to the LSTa and LSTc glycans, respectively. **C, D.** Close-up views of the glycan binding sites of polyomavirus VP1 proteins (PDB IDs 3BWR for the SV40-GM1 complex and 3NXD for the JCPyV-LSTc complex). The GM1 glycan terminates in α 2,3-linked sialic acid. **E, F.** Enterovirus D68 structures bound to α 2,3-sialyllactose (PDB ID 5BNP) and α 2,6-sialyllactose (PDB ID 5BNO). In all panels, the viral capsid proteins are shown in surface representation, with glycans and contacting protein residues represented with sticks. Sialic acid residues are shown in color, while the remaining glycan chains are gray. Hydrogen bonds and salt bridges between glycans and proteins are indicated with black dashed lines.

Sialic Acid Receptors

Cell-surface glycans featuring terminal sialic acids can serve as functional receptors for enveloped and nonenveloped viruses (Fig. 1). Sialic acids are ubiquitously expressed in higher vertebrates and are attached at the terminal ends of *N*- and *O*-glycans as well as glycolipids. The most common sialic acid in humans is α 5-*N*-acetylneuraminic acid (Neu5Ac). Sialic acids are usually connected via α 2,3 or α 2,6 glycosidic linkages to galactose (Gal) or *N*-acetylgalactosamine (GalNAc), but they can also be linked to one another through α 2,8 or α 2,9 linkages. Here, we review the sialic acid-binding strategies of influenza viruses, polyomaviruses, and enteroviruses (Fig. 2).

Influenza Viruses

Influenza viruses are segmented, single-stranded RNA viruses in the orthomyxovirus family that infect mammals and birds; they can be classified into types A, B and C. Influenza A and B viruses primarily cause serious disease in humans, as evidenced by yearly

outbreaks worldwide. Influenza viruses engage α 2,3- and α 2,6-linked Neu5Ac attached to a penultimate galactose of the glycan receptor. The trimeric viral hemagglutinin (HA) initiates cell contact by binding to Neu5Ac. The carbohydrate-binding site of HA is conserved in all influenza subtypes and located in a shallow groove near the top of the protein. The orientation of Neu5Ac and its interactions with HA are also conserved among influenza virus strains (**Fig. 2(A)** and **(B)**). Nonetheless, the context of sialic acids in the recognized glycan structures varies. Avian influenza virus HA primarily binds to α 2,3-linked Neu5Ac in a glycan that typically assumes an elongated *trans* conformation (**Fig. 2(A)**), whereas human influenza viruses preferentially attach to α 2,6-linked Neu5Ac in an oligosaccharide with a folded-back *cis* conformation (**Fig. 2(B)**). Structural analysis of avian H5 and H7 strains demonstrated that an HA point mutation leads to an increased affinity for α 2,6-linked sialic acid.

Glycans containing α 2,3-linked Neu5Ac display restricted conformational freedom and form a cone-like glycan structure. Conversely, glycans containing α 2,6-linked sialic acid have greater conformational flexibility, and such glycans assume umbrella-like shapes. The link between sialic acid and galactose in the receptor molecules defines the glycan topology and thus determines the affinity of a given HA.

Polyomaviruses

Polyomaviruses (PyVs) form a rapidly expanding family of small, non-enveloped dsDNA viruses that include several recently discovered human pathogens and cause a range of diseases in humans, including cancer. PyV capsids comprise 72 VP1 pentamers, termed capsomers. Each capsomer has five low-affinity glycan binding sites, facilitating multivalent binding to receptors and high avidity interactions that are critical for PyV targeting to the endocytic pathways and cell entry. Typically, VP1 binds terminal Neu5Ac as well as neighboring sugar residues in its glycan receptors, whereas the human BK (BKPyV) and JC (JCPyV) polyomaviruses engage the ganglioside GT1b and the lactoseries sialyl tetrasaccharide c (LSTc), respectively.

Some PyVs share a common binding site for sialic acid, with largely conserved contacts to the protein capsid. This is for example the case in simian virus 40 (SV40) and JCPyV, which bind sialic acids in the same orientation and with largely identical contacts (**Fig. 2(C)** and **(D)**). Nevertheless, glycan microarrays have shown that the two viruses are able to specifically bind to only one sialyloligosaccharide, and specificity for this glycan is encoded by a small number of residues that lie outside of the sialic acid binding pocket, and that make contacts with other components of the receptor. In SV40, several residues contact the terminal Gal residue in the branched GM1 glycan, which features an α 2,3-linked sialic acid (**Fig. 2(C)**). The monosaccharides Neu5Ac, GalNAc and Gal are arranged in a more linear conformation in this glycan. In JCPyV, the side chain of Asn123 contacts the acetyl group of the third monosaccharide, GlcNAc, ensuring specific recognition of the L-shaped pentasaccharide receptor LSTc, which contains α 2,6-linked sialic acid (**Fig. 2(D)**). As in influenza viruses, the longer α 2,6-linkage allows a L-shaped receptor conformation that enables the interactions with Asn123.

Enteroviruses

Enteroviruses (EVs) are very common and cause infections with a diverse array of clinical features. They belong to the picornavirus family and form non-enveloped, small (22–30 nm) icosahedral particles that encase an RNA genome. Although most members cause gastrointestinal and respiratory illnesses, often in children, some EVs have unique disease associations, including acute hemorrhagic conjunctivitis (AHC) caused by EV-70 and the group A coxsackieviruses A24 (CVA24). Recent structural studies on the EV-D68 and a CVA24 variant (CVA24v), show that both viruses engage sialylated glycan receptors, albeit in different regions of the capsid.

Interestingly, both EVs preferentially bind α 2,6-linked sialic acid, but for different reasons. In the case of EV-D68, specificity can be attributed to the architecture of the binding site. Binding of the linear α 2,3-linked sialoglycans would introduce steric hindrance, while the binding site can easily accommodate the bent α 2,6-linked sialylated structures (**Fig. 2(E)** and **(F)**). In CVA24v, the structures only provide details about the sialic acid moieties, while the remaining saccharide chains are not visible in the electron density maps. However, molecular dynamics simulations show that the α 2,6-linked glycans are more flexible and can establish more contacts with the viral capsid, suggesting a reason for the virus selection of this linkage.

Histo–Blood Group Antigen Receptors

The wide range of glycans used for attachment by viruses also includes the histo-blood group antigens (HBGAs) (**Fig. 1**), which are complex carbohydrates expressed on the surfaces of red blood and epithelial cells. HBGAs are also found in soluble form in mucosal secretions. The HBGA phenotypes in humans are polymorphic and complex, and they are associated with multiple gene families. HBGAs can serve as attachment factors for noroviruses and rotaviruses.

Noroviruses

Human noroviruses are non-enveloped, positive-sense RNA viruses of the calicivirus family that interact with HBGAs (**Fig. 1**) and cause acute gastroenteritis outbreaks worldwide. Binding studies with norovirus virus-like particles (VLPs) and specific Abs to blood-group antigens, soluble blood-group antigens or glycosidases, have confirmed a role for HBGAs in norovirus attachment. Based on the capsid gene sequence, noroviruses are divided into seven genogroups (GI to GVII) and over 30 genotypes. Their capsids are formed by a single protein, which can be divided into shell (S) and protruding (P) domains. The P domain is highly

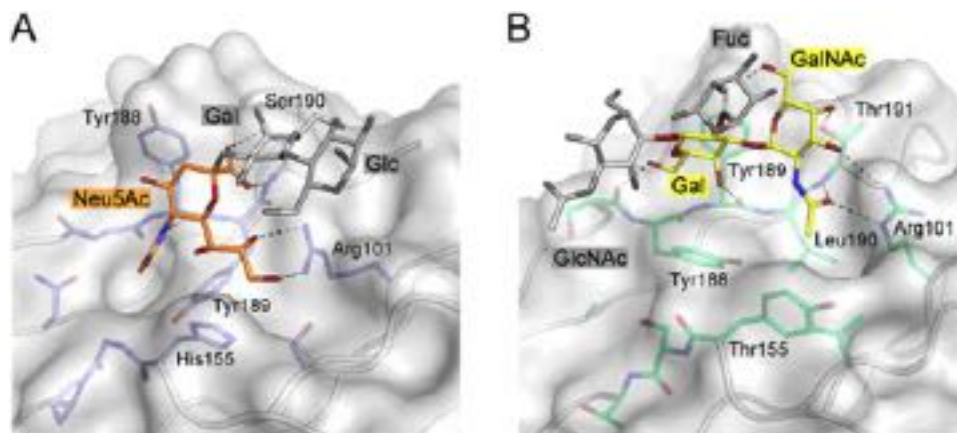


Fig. 3 The receptor-binding region on the rotavirus spike VP8* protein accommodates different glycan types. Panel **A** shows a view into the ligand-binding site of porcine rotavirus CRW-8 VP8* bound to the GM3 glycan (PDB ID 3SIT), and panel **B** shows the same view for human rotavirus strain HAL1166 VP8* bound to a tetrasaccharide corresponding to the terminal structure of A-type HBGA (PDB ID 4DS0). The VP8* structures are shown in semitransparent surface representation, and key contact residues and bound glycans are depicted as sticks. Carbon atoms of GM3 residues that interact with the protein are colored in orange and those in A-type HBGA residues in yellow; unbound residues gray. Direct hydrogen bonds between VP8* and glycans are black dashed lines, to indicate the different classes of the two glycans.

variable, but contains determinants for cell attachment and entry. Capsid proteins from distinct human norovirus strains show different HBGA binding patterns, suggesting that capsid-glycan interactions are major determinants for norovirus evolution. Host susceptibility to specific human norovirus strains and pathogenesis are likely influenced by genetically controlled expression of different HBGA structures on host cells.

Rotaviruses

Rotavirus infections are a leading cause of severe, dehydrating gastroenteritis in young children. Despite the global introduction of a rotavirus vaccine over a decade ago, rotavirus infections still result in >200,000 deaths annually worldwide. A human rotavirus was recently found to bind HBGAs, and this is of particular interest as most known rotaviruses had been reported to engage sialylated glycan receptors. In this study, a combination of glycan microarray screening and crystallographic analysis demonstrated that the attachment protein VP8* from the human P[14] rotavirus strain HAL1166 specifically binds HBGAs at the same site that interacts with sialic acid in other rotavirus strains (Fig. 3). The VP8* protein of HAL1166 rotavirus exhibits subtle modifications in its binding site that lead to inability to bind sialic acid and engagement of HBGAs. The specificity switch is particularly relevant, as more recent analyses have shown that the globally dominant human rotavirus genotypes bind HBGAs; in fact, they recognize a subset within the HBGA family that represents the most abundant glycans in the intestinal epithelial cells.

Glycosaminoglycan Receptors

The third class of virus glycan receptors or attachment factors is the polyanionic GAGs (Fig. 1). They are linear glycans assembled with β 1,4-linked disaccharide blocks, which contain an N-acetylated or N-sulfated amino sugar (GlcNAc or GalNAc) linked to either a uronic acid (glucuronic or iduronic acid) or galactose (Fig. 1). GAGs are components of proteoglycans and thus are in the extracellular matrix as well as in the cell membranes of most eukaryotic tissues.

Given their prominent location on the surface of cells, it is not surprising that a large number of viruses have been reported to use GAGs as attachment or entry receptors. However, our understanding of the rules underlying virus-GAG interactions at the molecular level remains limited because of few reliable high-resolution structures of viral proteins in complex with GAGs. Furthermore, observations that several viruses, such as foot-and-mouth disease virus (FMDV), undergo cell culture adaptation related to an increased binding to HS without any selective pressure, questioned the utilization of HS by clinical, non-culture-adapted strains. Finally, it is also becoming clear that GAGs act as decoy receptors for some viruses.

The picornavirus FMDV was one of the first examples for which high-resolution studies of GAG-virus interactions became available. Crystal structures of the human papillomavirus (HPV) capsid protein L1 bound to oligomeric heparin have more recently been reported. However, our inspection of the available coordinates and electron densities (PDB ID 3OAE and 3OFL) shows that the ligands are not well defined in the complexes, likely because of low occupancies. The reported contacts between the HPV L1 proteins and heparin must therefore be treated with caution. A more recent cryo-EM reconstruction of the entire HPV capsid bound to heparin shows that the glycans bind to recessed regions between the L1 pentamers. The resolution of the structure is however not sufficient to assign atomic-level interactions. A number of low-resolution structures of complexes with GAGs or analogs have also been determined for different adeno-associated viruses.

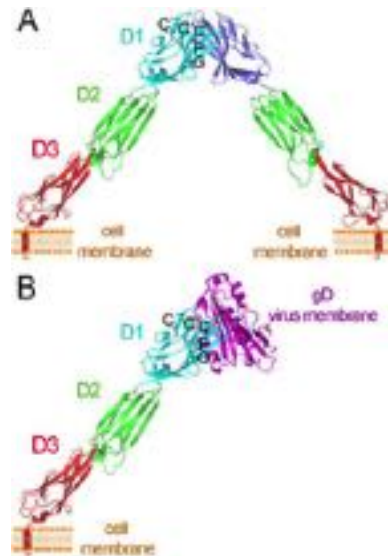


Fig. 4 The IgSF Nectin-1 in cell adhesion and virus recognition. **A.** Homophilic nectin-1 cell adhesion interactions. Ribbon diagram of the extracellular nectin-1 region with the three Ig-like domains (D1-D3), which forms a dimer in the crystal structure (PDB ID 3U83). The interacting N-terminal domains (D1) in the homophilic dimer are blue and cyan, with the GFCC'C'' β -sheet engaged in the adhesion interaction labeled. **B.** The HSV gD glycoproteins bound to nectin-1. Ribbon diagram of the gD-nectin complex structure (PDB ID 3U82), with the gD Ig-like domain (magenta) bound to the N-terminal D1 (cyan) of nectin-1. The virus protein interacts with the GFCC'C'' β -sheet of the membrane distal D1.

Glycan-Based Antiviral Strategies

To date, very few examples for antiviral strategies that exploit knowledge about virus-glycan interactions exist. The best-known of these are probably the influenza virus inhibitors oseltamivir and zanamivir, which target the viral neuraminidase by mimicking host glycan structures. Both compounds act as inhibitors of influenza neuraminidase, and prevent the emergence of progeny virions from infected cells. Additional examples include recently developed potent trivalent glycan-based inhibitors of some human adenoviruses, as well as multivalent entry-inhibitors against norovirus infection. In these cases, the inhibitors simultaneously target several glycan-binding sites on the viral capsid with high affinity, efficiently blocking attachment. These studies demonstrate the general feasibility of designing glycan receptor-based viral inhibitors.

Protein Receptors, Interactions With Viruses

Immunoglobulin Superfamily Proteins

Perhaps reflecting their relative abundance at the cell surface, proteins belonging to the immunoglobulin superfamily (IgSF) are implicated in the cell entry of many viruses. The IgSF comprises numerous cell surface proteins, most of which function as cell adhesion molecules (CAM). They are type I membrane proteins with concatenated Ig-like domains at the extracellular region, and viruses commonly recognize the membrane-distal, N-terminal domains (Fig. 4). The Ig fold has two β -sheets bridged by a conserved disulfide bond, although some domains contain additional disulfides. They are classified in different structural sets according to the number of β -strands in each of the two β -sheets.

The IgSF proteins can be grouped in subfamilies based on structure-function similarities. Members of the Nectin and Nectin-like (Necl) subfamilies mediate cell entry of poliovirus (PV), alphaherpesviruses and measles virus (MV). The nectins contain three (D1 to D3) tandem extracellular Ig-like domains (Fig. 4); the N-terminal D1 has an Ig-like variable (IgV) fold, whereas D2 and D3 belong to the constant (C2) Ig set. Nectins mediate homophilic and heterophilic cell adhesion interactions in different tissues. Cell-cell and virus-cell interactions similarly engage the GFCC'C'' β -sheet in the nectin D1 (Fig. 4).

The PV receptor (PVR, CD155), also known as Necl-5, was the first nectin described as a virus receptor. Subsequently, poliovirus receptor-related proteins (PRR) were identified as alphaherpesviruses entry receptors. PRR-1 or Nectin-1 is the entry receptor of herpes simplex virus (HSV) 1 and 2. Nectin-2 functions also as the receptor of certain HSV variants, whereas Nectin-4 is an epithelial cell MV receptor, necessary for movement across the airway epithelium and transmission to new hosts. MV also uses another IgSF member, signaling lymphocyte-activation molecule (SLAM), to initiate the infection in lymphoid cells.

Other cell surface IgSF proteins bearing N-terminal virus-binding IgV domains function as virus receptors: The T-cell specific CD4 protein for HIV-1, the coxsackievirus and adenovirus receptor (CAR), junctional adhesion molecule-A (JAM-A) for reovirus, CD300f for norovirus, carcinoembryonic antigen-like cell adhesion molecule-1 (CEACAM-1) for coronavirus, the Hepatitis A virus cellular receptor-1 (HAVcr-1 or TIM-1), etc. In these receptors, the GFCC'C'' β -sheet is engaged in virus binding. Nevertheless, the virus-binding domains in

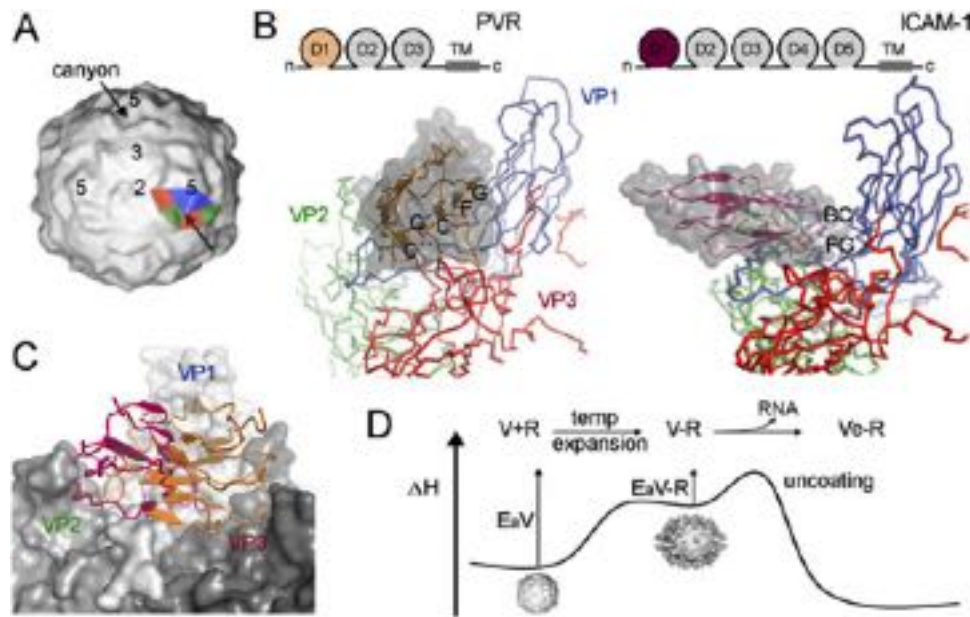


Fig. 5 The picornavirus "canyon" in receptor recognition. **A.** Surface representation of the poliovirus (PV) particle (PDB ID 1ASJ). Numbers indicate location of the icosahedral two-, three- and five-fold axes. Two capsid biological protomers are shown with the VP1 (blue), VP2 (green) and VP3 (red), and the canyon region at the interprotomer junction marked with arrows. **B.** IgSF receptors bind to the canyon. Top: Scheme of the PVR and ICAM-1 picornavirus receptors, with the extracellular domains (D) outlined and the N-terminal virus binding D1 colored. Transmembrane (TM), n and c-terminus labeled. Bottom: Representation of the PV and CAV cryo-EM structures bound to PVR (PDB ID 3J8F) and ICAM-1 (PDB ID 6EIT), respectively. Lateral view of the canyon region, with the virus proteins colored as in A. Ribbon and surface representations of the receptor domains and the main virus binding motifs labeled: GFCC'C'' β -sheet in PVR D1 and the BC and FG loops in ICAM-1. **C.** Frontal view of the canyon in the superimposed structures. Surface (gray) and ribbons representations of the virus and the receptors, respectively. **D.** Canyon-binding receptors catalyze picornavirus uncoating. The endothermic canyon-receptor interaction generates high energy (ΔH) virus-receptor (V-R) complexes with an expanded capsid at physiological temperature, which assists RNA exit and accelerates uncoating. Receptor binding lowers the activation energy (E_a) of the virus uncoating process (vertical thin arrows): $E_aV-R < E_aV$. Ve-R: Empty complexes.

other IgSF members such as intercellular adhesion molecule-1 (ICAM-1), a rhinovirus receptor, do not belong to the V set and its narrower conformation facilitates binding into a recessed virus surface (see below). In the recently identified Mxra8 alphavirus receptor, the two virus-binding IgV domains of Mxra8 arrange head-to-head, are disulfide bridged and contain the first β -strands swapped.

Even though IgSF protein structures are similar, viruses recognize specific molecular features in the Ig-like domains, which determine the virus-receptor specificity.

Virus interactions with IgSF receptors

PV, the causative agent of paralytic poliomyelitis, uses an IgSF member for cell entry. Murine cell expression of PVR renders them fully permissive for poliovirus infection, and is the basis for the generation of PVR-transgenic mouse lines for neurovirulence studies. PV is a representative member of the picornavirus family, genus enterovirus; it is a non-enveloped virus with an icosahedral capsid that encloses a positive-sense single-stranded RNA genome (Fig. 5(A)). A depressed capsid surface or canyon surrounding the five-fold icosahedral vertices binds to the membrane distal D1 of PVR (Fig. 5(A) and (B)). Receptor binding initiates a series of irreversible conformational changes in the virus particle, which results in the externalization of capsid polypeptides and the virus genome through the capsid 2-fold axis after expansion. Exit of internal capsid polypeptides occurs in other picornavirus entry intermediates, and it is thought important for capsid-membrane interaction and RNA translocation from the capsid to the cytosol. Indeed, studies with PVR-loaded membrane vesicles showed a membrane pore for viral genome delivery into the cytosol.

Human rhinoviruses (HRV) comprise three species of EVs. The majority of HRVs attach to the membrane-distal Ig-like domain of ICAM-1 (Fig. 5(B)), a member of the IgSF that binds to leukocyte integrins, whereas about 12 HRV serotypes use the low-density lipoprotein receptor (LDLR) and the HRV C viruses use cadherin-related family member 3 as entry receptor. ICAM-1 is also the entry receptor of Coxsackie A virus (CAV). HRV and CAV similarly use the canyon for binding to the N-terminal domain of ICAM-1, such as shown by cryo-EM structures of virus particles in complex with soluble receptors (Fig. 5(B)). The bound ICAM-1 D1 is more perpendicular to the capsid than the PVR D1 (Fig. 5(B) and (C)). The structures indicate that receptor-binding residues are more exposed to Ab neutralization in PV than in CAV or HRV.

Like poliovirus, receptor binding to the canyon initiates conformational changes in picornavirus particles that are considered essential for infection. These canyon-receptor interactions are endothermic and generate metastable complexes with an expanded

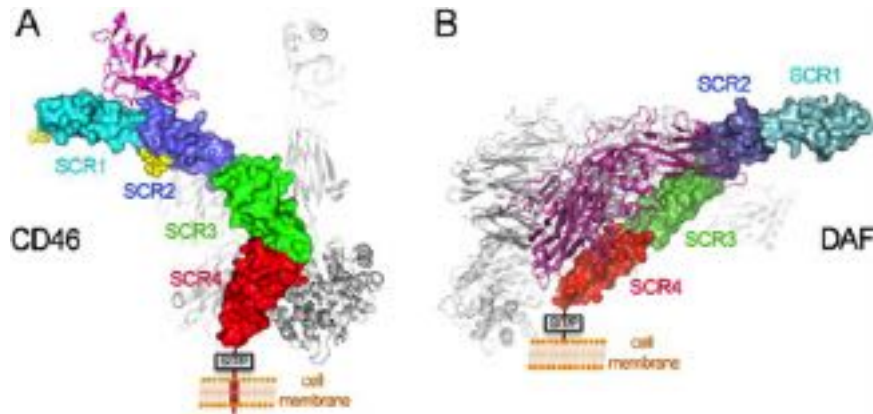


Fig. 6 Complement control proteins in virus recognition. **A.** CD46 in complex with the adenovirus (Ad) type 11 knob (magenta) and the complement C3b protein (gray). Surface representation of the CD46 extracellular short consensus repeats (SCR) followed by the serine, threonine and proline rich region and the transmembrane domain. The superimposed CD46-Ad (PDB ID 3O8E) and CD46-C3b (PDB ID 5FO8) complex structures show that Ad and C3b binds to distinct SCR. **B.** DAF binding to the EchoV6 capsid protomer (magenta) and the complement C3b protein (gray). SCR shown as in CD46. The superimposed DAF-EV6 (PDB ID 6ILK), DAF-C3b (PDB ID 5FOA) and DAF alone (PDB ID 1OJV) structures show virus and complement binding to overlapping receptor sites. Structures superimposed based on the receptor molecules.

capsid, a feature of entry intermediates. Capsid expansion occurs by the concerted action of temperature-driven capsid “breathing” and receptor binding, and precedes RNA exit. In certain picornaviruses, capsid expansion in virus-receptor complexes initiates at low pH, with the release of lipid molecules attached to VP1. Canyon-binding receptors mediate uncoating by maintaining the virus capsid in an intermediate open state, which decreases the activation energy of the uncoating process (Fig. 5(D)). Rossmann proposed that the canyon is too narrow to allow the binding of virus neutralizing Abs, thereby allowing the virus to retain a relatively invariant receptor-binding domain in the presence of immunoselection. Nonetheless, structural studies demonstrated that the canyon instead evolved in response to receptor binding requirements and to accommodate structural changes that occur post-binding.

CD4, the primary receptor for HIV, is composed of four Ig-like domains; the HIV envelope glycoprotein (env) binds the most membrane-distal domain (see below). This interaction induces conformational changes in the virus gp120 moiety that mediate attachment to chemokine receptors, which then trigger virus-cell membrane fusion and cell entry. Conformational flexibility, sequence and gp120 glycosylation diversity hide the CD4-binding site (CD4bs) from neutralizing Abs. However, several CD4bs-specific Abs have been identified from HIV-infected elite donors. They exhibit very high breadth and potency, and neutralize over 90% of HIV-1 strains by interaction with highly conserved residues engaged in receptor recognition.

The identification of potent HIV-neutralizing Ab enables prophylactic and therapeutic trials using single or Ab cocktails; CD4bs Ab administration reduced viraemia in HIV-1 infected individuals. Alternatively, significant effort has been expended in generating vaccine candidates capable of inducing broadly neutralizing Ab against HIV env vulnerable sites. The current understanding of Ab affinity maturation in HIV-infected individuals indicates that successful immunization should induce a precise Ab lineage evolution toward a defined class of broadly neutralizing molecules that target the CD4bs or specific antigenic sites.

Complement Control Proteins

A number of viruses have been shown to bind to complement control proteins, also termed the regulators of complement activity (RCA). They protect the host from uncontrolled complement activation and lysis of intact self-cells. The RCA are soluble or membrane bound proteins composed of concatenated domains termed short consensus repeats (SCR) (Fig. 6), modules of about 60 amino acid residues and four invariant cysteines that form a compact barrel structure.

Epstein-Barr virus (EBV) attachment to B lymphocytes occurs by the interaction of the virus gp350/220 envelope glycoprotein with the cell-surface complement receptor 2 (CR2, CD21) or CR1 (CD35). CR2 possesses 16 extracellular SCR, the two most membrane-distal being the site of gp350/220 binding. The EBV receptor-binding region is targeted by neutralizing Abs, which indicates that blocking EBV-receptor interaction is a major neutralization mechanism. Nanoparticles that display the gp350-receptor recognition site elicit Abs that neutralize EBV, and are relevant vaccines candidates.

CD46 is a four-SCR protein expressed in all nucleated cells that acts as a cofactor in the proteolytic inactivation of C3b/C4b (Fig. 6(A)). It serves as the entry receptor of some MV genotypes, group B adenoviruses (Ad) and herpesvirus 6. The MV envelope hemagglutinin and the Ad fiber knob bind to a glycan-free surface formed by the most membrane-distal SCR1 and SCR2 (Fig. 6(A)), initiating virus cell entry. In MV, a unique concave surface with a socket accommodates CD46 and two other IgSF receptors, SLAM and Nectin-4, which facilitates virus dissemination among hosts and confers an extraordinary transmissibility to MV. CD46 virus and complement binding regions are not overlapping (Fig. 6(A)).

Decay-accelerating factor (DAF, CD55), an RCA-family protein with four SCR, is the receptor for a significant number of human EVs, including the group B coxsackieviruses (CVB) and many echoviruses (EchoV). The atomic structure of DAF shows a rod-shaped protein with four SCRs (**Fig. 6(B)**), separated from the membrane by a heavily glycosylated Ser/Thr-rich stalk. In contrast to many of the other examples presented here, where binding is to the membrane-distal domain of the receptor, the majority of the known DAF-binding EchoV interact with SCR3 (**Fig. 6(B)**), and with the second and/or fourth SCR domains. Virus and complement binding regions overlap (**Fig. 6(B)**). The interaction of DAF with EchoV occurs at a surface-exposed region or “puff” region, outside the canyon. Recently, the neonatal Fc receptor (FcRn) has been identified as an EchoV uncoating receptor that binds to the canyon.

Ectoenzymes

Ectoenzymes are membrane proteins with their catalytically active sites in the extracellular environment. They play diverse functions relevant to the regulation of homeostasis, and they are linked to numerous diseases and infections. Several ectoenzymes are entry receptors of Coronavirus (CoV), a large family of enveloped positive RNA viruses that infect a wide range of mammals and birds.

Aminopeptidase N (APN, CD13) was the first ectoenzyme to be related to CoV cell entry. It is a type II surface metalloprotease with a large ectodomain that forms dimers and contains a zinc ion at its active site (**Fig. 7(A)**). APN is termed a “moonlighting enzyme” because of its link to many cell functions. Another zinc metalloprotease, human angiotensin-converting enzyme 2 (ACE2), functions as the receptor for important human pathogens, the severe acute respiratory syndrome coronaviruses (SARS-CoVs). ACE2 is a type I membrane glycoprotein with an N-terminal extracellular domain built of two α -helical lobes (**Fig. 7(B)**); the catalytic site with a coordinated zinc ion is located between the lobes. More recently, a third ectoenzyme, dipeptidyl peptidase 4 (DPP4, CD26), was identified as the receptor of the middle east respiratory syndrome coronavirus (MERS-CoV). DPP4 is a type II membrane protein composed of two domains and forms homodimers. APN, ACE2 and DPP4 catalyze hydrolysis of short peptides and function in angiogenesis, cell adhesion and blood pressure regulation. Nonetheless, CoV selection of ectoenzymes as entry receptors is not determined by their enzymatic activity, which is not required for cell entry and infection.

CoV-ectoenzyme interactions

The major envelope spike (S) glycoprotein mediates receptor recognition and fusion in CoV. The trimeric S is composed of the S1 and S2 fragments and several recent structures have determined its domain architecture. The receptor-binding domains (RBD) that attach to ectoenzymes are at the most membrane distal pole of the S. Several crystal structures have described the molecular basis for CoV recognition of ectoenzymes.

The APN ectodomain is composed of four domains (DI-DIV), is heavily glycosylated and forms dimers through extensive DIV-DIV interactions (**Fig. 7(A)**). The CoV RBD interacts with a single APN monomer and preferentially contacts the membrane distal region of DIV. Human and animal CoV that belong to the alpha-CoV genus use extended loops that protrude at one edge of the RBD to attach to APN (**Fig. 7(A)**). Immune pressure has provoked marked differences on the receptor-binding loops generating different APN recognition among CoV.

The SARS- and MERS-CoV belong to the beta-CoV genus and are highly pathogenic in humans. The RBDs of these beta-CoVs contain an external subdomain specialized in receptor recognition, inserted in a central β -stranded core (**Fig. 7(B)** and **(C)**). In SARS-CoVs, the central concave subdomain surface cradles the ACE2 N-terminal α -helix, whereas the terminal side of the subdomain interacts with other ACE2 regions (**Fig. 7(B)**). Similarly, the MERS-CoV RBD engages the DPP4 molecule through the solvent-exposed side of its external subdomain (**Fig. 7(C)**). It contacts the edges of DPP4 β -propeller blades IV and V. N-linked glycosylations are engaged in CoV binding to DPP4 and APN and they are key determinants in CoV-receptor binding specificity. Subtle residue substitutions in the SARS and MERS receptor-binding subdomains increase their affinities for the human receptors, favor cross-species transmission and cause zoonosis.

Co-Receptors for Infection

Virus binding and entry is often a multistage process involving a cascade of interactions between the virus and host factors. The initial attachment is usually to the receptor, with subsequent interactions mediated by co-receptors. Commonly, specific virus interactions with the primary receptor and the co-receptor are necessary for virus penetration into the cell.

CBVs bind DAF at the cell surface, but infection also requires CAR, an IgSF protein with two extracellular domains. Differing from DAF, CAR engages the CBV canyon and triggers capsid conformational changes needed for entry. However, CAR is naturally located in tight junctions where it mediates homotypic cell adhesion, and is therefore not immediately accessible for virus attachment at the apical cell surface. Elegant studies have demonstrated that virus binding to DAF at the latter site results in Abl-kinase mediated signaling events which cause actin rearrangements that allow virus movement to the tight junctions where uncoating occurs. DAF binding also activates Fyn-kinase, which results in the phosphorylation of caveolin and subsequent virus entry in caveolin vesicles. CBVs therefore exploit the signaling capacity of their low-affinity primary DAF receptor to traffic to the high-affinity, uncoating CAR receptor and thereby cross the epithelial barrier.

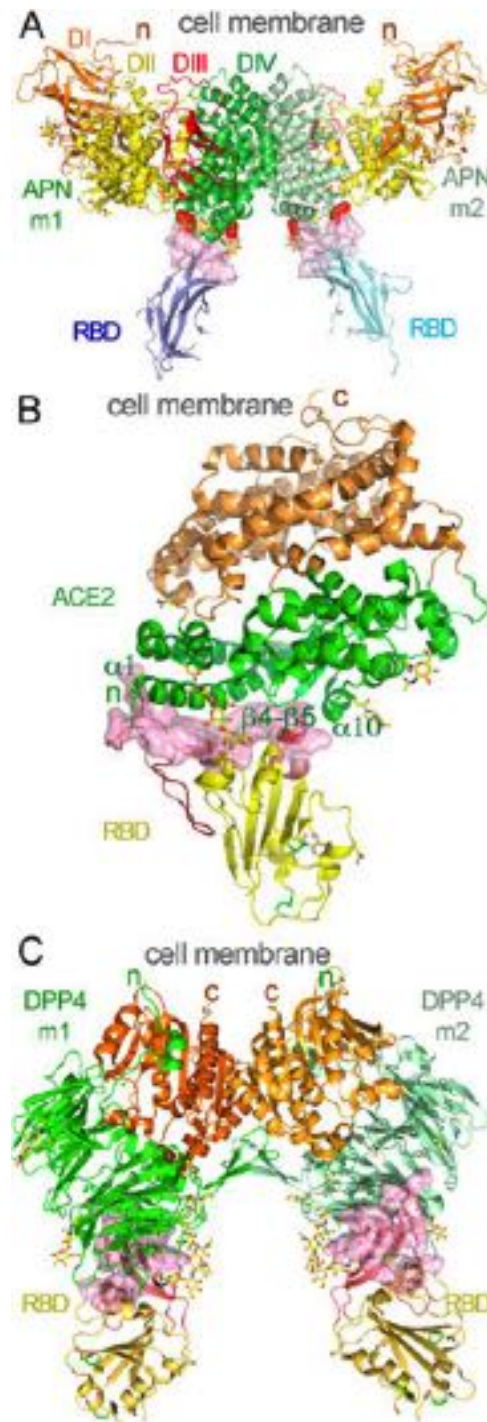


Fig. 7 Ectoenzymes as coronavirus receptors. **A.** Dimeric APN-CoV structure. Ribbon drawing of the pig APN in complex with a porcine CoV RBD (PDB ID 4F5C). APN monomers (m) are shown with domains in orange (N-terminal DI), yellow (DII), red (DIII) and green (C-terminal DIV), as well as the N-terminal ends near the cell membrane. The bound RBDs are shown in blue and cyan, with the APN-binding surface in pink and two key residues in the domain tip in red. **B.** The SARS-CoV RBD bound to its ACE2 receptor. Ribbon drawing of the ACE2-RBD complex structure (PDB ID 2AJF). The RBD is shown with the core in yellow, the receptor-binding subdomain in dark red and the ACE2-binding surface in pink. Disulfide bonds as green cylinders. ACE2 is shown with the virus binding domain in green and the membrane-proximal lobe in orange. The three main ACE2 regions recognized by SARS-CoV are labeled in green. **C.** MERS-CoV binding to DPP4. Ribbon drawing of the dimeric DPP4-RBD complex structure (PDB ID 4KRO). The DPP4 monomers are shown with the virus binding, N-terminal-propeller domain in green and the C-terminal hydrolase domain in orange. The RBD core and the receptor-binding subdomain as described for SARS in panel B. Asn residues at glycosylation sites and the attached glycans are shown as sticks, with carbons in yellow in all the panels.

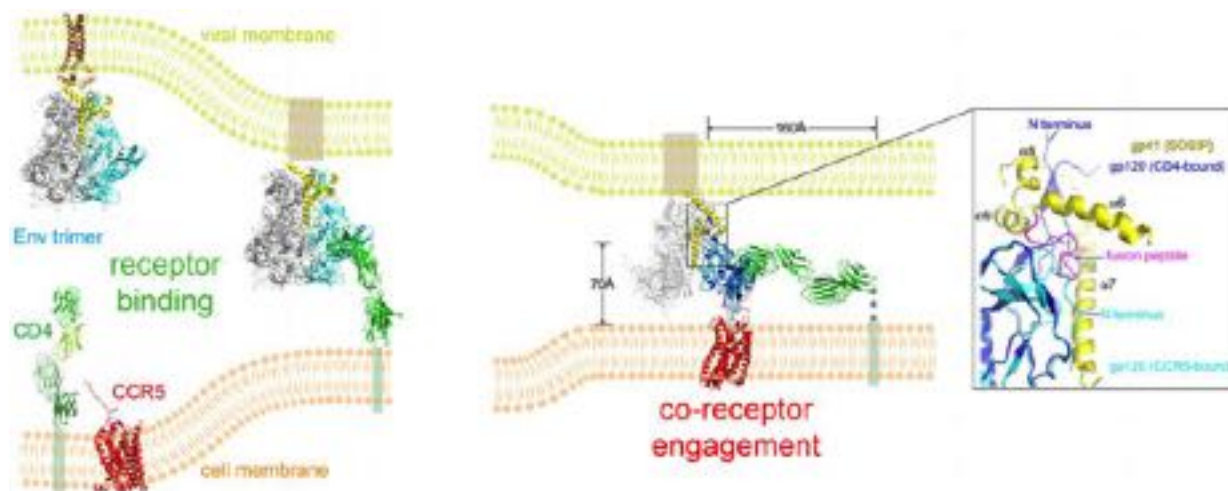


Fig. 8 The pathway of HIV-1 entry. Structure of the prefusion HIV-1 Env trimer ectodomain in the viral membrane, and the structures of the primary CD4 receptor (green) and the CCR5 co-receptor (red) in the target cell membrane. One gp120/gp41 heteromonomer in the trimer is shown in cyan and yellow, respectively, the other two gray. Binding of gp120 to the primary CD4 receptor (left) exposes the co-receptor-binding region and facilitates gp120 binding to CCR5 (right). Co-receptor engagement brings the trimer close to the cell membrane, initiates dissociation of gp120 and gp41 and the virus-cell membrane fusion. Inset: Conformational changes in gp120 and gp41 related to its dissociation and the insertion of the fusion peptide (magenta) into the cell membrane. Reproduced with permission from Cheng, B., 2019. Molecular mechanism of HIV-1 entry. Trends in Microbiology 27, 878–891.

Like poliovirus and rhinovirus, HIV binding to CD4 initiates conformational changes in the trimeric env that are a prerequisite for the subsequent fusion events (Fig. 8). However, unlike the two picornaviruses, interaction with the primary receptor exposes conserved residues that interact with a co-receptor, a necessary precursor to the fusion of viral and cellular membranes (Fig. 8). This requirement for a co-receptor explains why HIV can bind to human CD4 expressed on murine cells – which lack the co-receptor – without undergoing membrane fusion. The chemokine receptors CXCR4 or CCR5 function as HIV co-receptors; they belong to the seven-transmembrane-domain G-protein-coupled receptor superfamily. The importance of the chemokine co-receptor in HIV infection is supported by the observed genetic resistance to HIV of individuals with a deletion and frameshift within CCR5. Finally, the identification of chemokine receptors as critical co-receptors of HIV infection has also contributed to our understanding of virus tropism and evolution in vivo. CCR5-tropic strains of HIV are predominant for transmission and during the asymptomatic phase of virus infection. CCR5 is expressed at high levels in activated/memory cells and in gut-associated lymphoid tissue, the primary site of HIV replication. As infection progresses, CXCR4-tropic viruses evolve, able to exploit the wider tissue expression of the CXCR4 receptor.

Attachment Factors, Molecules That Facilitate Infection

Certain cell surface molecules cooperate in virus infection by promoting attachment to host cells and virus transfer to entry receptors on the cell surface or in endosomes. Attachment factors or receptors allow viruses to bind to the cell membrane, but additional interactions, for example with entry receptors, are required to mediate delivery of the virus or viral genome into the cytoplasm. Frequently, virus proteins, glycosylations or the virus lipidome recognize attachment factors – glycans or proteins – that facilitate cell infection.

For some viruses, the attachment factors that promote binding to permissive cells are quite well defined. Since they decorate cell surfaces, sulfated GAGs or HBGAs frequently serve as attachment factors for a range of viruses, including members of the flavivirus, picornavirus, papillomavirus, foamy virus, and calicivirus families. Sialylated glycans are also virus attachment factors that do not promote entry, for example in the case of Merkel cell PyV, while they function as entry receptors for many other viruses. The predominant display of specific glycan motifs on surfaces of different cells and tissues contributes to the host restriction and cell/tissue tropism of the virus.

HIV and Ebola virus (EBOV) envelope glycoproteins bear N-linked high mannose glycans that bind to the lectin dendritic cell-specific ICAM-grabbing non-integrin (DC-SIGN) (Fig. 9). These viruses employ DC-SIGN with different purposes. HIV-1 uses dendritic cells (DC) as a “Trojan horse” to move from mucosa to secondary lymphoid tissues. DC-SIGN bound viruses are stored in DC and presented to CD4⁺ cells, which mediates its trans-infection (Fig. 9(A)). Nonetheless, EBOV subverts DC-SIGN to penetrate into cell endosomes (Fig. 9(B)), where the EBOV envelope glycoprotein (GP) is processed and subsequently binds to its Niemann-Pick C1 (NPC1) entry receptor, which triggers membrane fusion and entry.

An alternative way of HIV-1 and EBOV attachment to DC through the virus lipidome has been identified. The HIV-1 envelope membrane contains gangliosides with a sialyllactose – sialic acid on galactose – moiety that binds to sialic-acid-binding

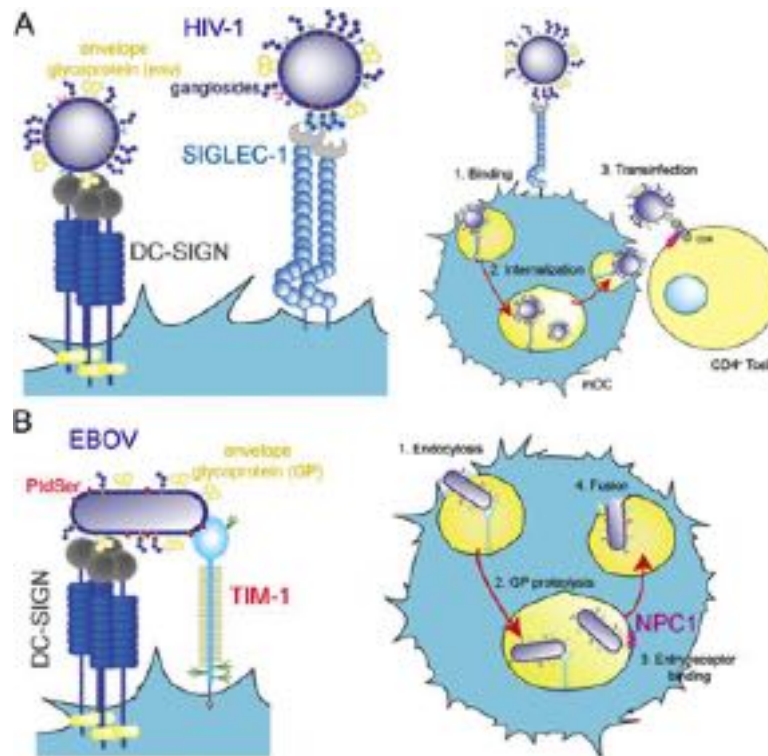


Fig. 9 Attachment factors in virus cell entry. **A.** Factors that mediate HIV-1 transinfection. HIV-1 uses membrane env and gangliosides bearing sialyllactose to bind to DC-SIGN and Siglec-1, respectively. Right: Siglec-1-bound virions in mature DC (mDC) are internalized; in secondary lymphoid tissues mDC present HIV-1 particles to CD4⁺ T cells and infection initiates. A similar process is described with DC-SIGN. **B.** Attachment factors facilitate the EBOV transfer to endosomes for membrane penetration. EBOV can use envelope GP or PtdSer to attach respectively to cell surface DC-SIGN or PtdSer-receptors such as TIM-1. Adhered particles are internalized by macropinocytosis and reach late endosomes, where proteases remove the GP mucin-like domain and the glycan cap and expose the receptor binding region. The particles can then bind to the NPC1 entry receptor, which mediates the fusion of viral and cell membranes, and the release of the nucleoprotein in the cell cytoplasm. Based on Izquierdo-Useros, N., Lorizate, M., McLaren, P.J., *et al.*, 2019. HIV-1 capture and transmission by dendritic cells: The role of viral glycolipids and the cellular receptor Siglec-1. *PLoS Pathogens* 10, e1004146.

immunoglobulin-like lectin-1 (Siglec-1) on the DC cell surface (Fig. 9(A)). The HIV-1 particles are captured by mature DC that express high Siglec-1 levels, and presented to CD4⁺ T cells in lymphoid tissues (Fig. 9(A), right). This trans-infection route is thought more relevant for HIV-1 transmission than engaging DC-SIGN because of the high amounts of Siglec-1 in different DC subsets. A similar retrovirus trans-infection route is also described in macrophages. Siglec-1 can function as an EBOV attachment receptor that facilitates virus penetration into DC endosomes.

Numerous studies with different viruses have demonstrated the relevance of virus membrane phosphatidylserine (PtdSer) in cell entry, which was first identified as a cofactor for HIV-1 infection of macrophages. PtdSer is a danger signal recognized by specific receptors that mediate the uptake of apoptotic cells. The implication of viral PtdSer in virus entry was first described as apoptotic mimicry with vaccinia virus, which penetrates into cells by macropinocytosis. More recent studies demonstrated that PtdSer receptors enhance entry and infection of alphavirus, baculovirus, flavivirus, filovirus and NewWorld arenavirus. It is proposed that EBOV can use PtdSer receptors such as TIM-1 to gain access to its entry receptor in late endosomes (Fig. 9(B)). Nonetheless, the enhancing activity of PtdSer-receptors on virus entry is mostly described *in vitro*, and the relevance of these receptors on virus infection is still under discussion.

Evolution of Virus-Receptor Interactions and Viral Tropism

Due to the possession of error-prone replication strategies, many viruses exhibit an exquisite ability to adapt to the environment in which they are grown. Virus-receptor interactions are constantly subject to changes because of evolutionary pressure to increase the virus infection efficiency and to escape from immune surveillance. Numerous modifications on the receptor-binding regions lead to the emergence of a new pathogen with altered infectivity, tissue tropism, or host range.

Using a conserved binding site, many human polyomaviruses bind to N-acetyl neuraminic acid, a carbohydrate that is prominently present on the surface of human cells. In contrast, the non-human polyomavirus simian virus 40 (SV40) binds to N-glycosyl neuraminic acid, a modified sugar that is present in monkeys (SV40 host) but that is absent in humans. The two

neuraminic acid variants differ only by a single hydroxyl group, and the sialic acid binding pocket of SV40 differs from that of closely related human polyomaviruses at only very few positions that enable SV40 to accommodate the additional hydroxyl group present in the glycolyl group. Thus, subtle modifications in a binding pocket can enable a virus to change its host range.

Adenoviruses have evolved to recognize a range of different protein and glycan receptors (CAR, CD46, the GD1a glycan and desmoglein-2) using the same protein, the fiber knob. Here, more extensive amino acid substitutions lead to changes in surface loop structure that endow a given knob with the ability to interact specifically with one of these receptors. Adenoviruses that infect the eye and cause Epidemic Keratoconjunctivitis, for example, use the GD1a glycan as their receptor, and since this glycan is heavily expressed on human corneal cells the ability of the virus to bind GD1a is linked to tissue tropism.

In CoVs, the evolution in receptor recognition can lead to receptor specificity switches as well as to inter-species transmission and zoonosis. The APN recognition mode in the human CoV-229E diverged from related animal CoV of the alpha genus, whereas the human NL63-CoV switched to the recognition of ACE2. These divergences among related CoVs are due to immune pressure over the receptor binding region, which causes conformational changes in the receptor-interacting motifs to escape from Ab neutralization. The emergence of the SARS- and MERS-CoVs, which are responsible for severe respiratory syndromes, generated greater interest on CoV evolution. These viruses originated in bats, their natural reservoir, adapted to intermediate hosts – civets for SARS and dromedary camels for MERS – from which they transmitted to humans. The efficient transfer of SARS- and MERS-CoVs from bats to humans is due to increased affinity for their human orthologue receptors. These CoVs contain large receptor binding subdomains (Fig. 7(B) and (C)), which accommodate subtle residue substitutions that increase their affinity for the human receptor proteins and facilitate its transmission to man, where they cause severe diseases. Adaptive mutation in the CoV S also facilitates their spread among the human population.

EBOV is transmitted from bats to primates and generates recurrent outbreaks in Central Africa. It is thought that efficient EBOV transmission to humans is related to the receptor binding properties of the EBOV GP. Whereas most EBOV in Central Africa affected a limited number of individuals, the Ebola epidemic that began in West Africa at the end of 2013 spread rapidly and reached an unprecedented scale. It was caused by a variant named Makona, which evolved during the epidemic toward an enhanced primate cell infectivity. Some of the acquired residue substitutions mapped at the Makona GP receptor-binding region, and it is hypothesized that they facilitate the interaction with cell surface receptors. Evolution of the EBOV-receptor binding features in the human population can thus lead to an increased human-to-human transmission and to epidemics.

Summary

Virus receptors are key components of the early events involved in cell infection by mediating attachment as well as conformational changes in the virus that lead to cell entry. In addition, they provide mechanisms for internalization in which the virus may subvert a natural receptor cycling process. Furthermore, the identification of virus receptors contributes significantly to our understanding of host, tissue, and cell tropism, and helps explain aspects of virus pathogenesis. Insights into virus-receptor interactions can inform the rational design of vaccines that elicit virus immune neutralization.

The selected bibliography illustrates relevant and recent research on different aspect of viruses and their receptors.

Further Reading

- Baggen, J., Hurdiss, D.L., Zocher, G., *et al.*, 2018. Role of enhanced receptor engagement in the evolution of a pandemic acute hemorrhagic conjunctivitis virus. *Proceedings of the National Academy of Sciences of the United States of America* 115, 397–402.
- Blaum, B.S., Stehle, T., 2019. Sialic acids in nonenveloped virus infections. In: *Advance in Carbohydrate Chemistry and Biochemistry* 76. Elsevier, pp. 65–111.
- Bostina, M., Levy, H., Filman, D.J., Hogle, J.M., 2011. Poliovirus RNA is released from the capsid near a twofold symmetry axis. *Journal of Virology* 85, 776–783.
- Cui, J., Li, F., Shi, Z.-L., 2019. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* 17, 181–192.
- Izquierdo-Useros, N., Lorzate, M., McLaren, P.J., *et al.*, 2014. HIV-1 capture and transmission by dendritic cells: The role of viral glycolipids and the cellular receptor Siglec-1. *PLoS Pathogens* 10, e1004146.
- Kwong, P.D., Mascola, J.R., 2018. HIV-1 vaccines based on antibody identification, B cell ontogeny, and epitope structure. *Immunity* 48, 855–871.
- Liu, Y., Sheng, J., Baggen, J., *et al.*, 2015. Sialic acid-dependent cell entry of human enterovirus D68. *Nature Communications* 6, 8865.
- Moller-Tank, S., Maury, W., 2014. Phosphatidylserine receptors: Enhancers of enveloped virus entry and infection. *Virology* 468–470, 565–580.
- Narpala, S., Todd, J.-P., Rao, Srinivas, S., *et al.*, 2015. Rational design of an Epstein-Barr virus vaccine targeting the receptor-binding site. *Cell* 162, 1090–1100.
- Nasr, M.L., Baptista, D., Strauss, M., *et al.*, 2017. Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nature Methods* 14, 49–52.
- Reguera, J., Mudgal, G., Santiago, C., Casasnovas, J.M., 2014. A structural view of coronavirus-receptor interactions. *Virus Research* 194, 3–15.
- Ren, J., Wang, X., Hu, Z., *et al.*, 2013. Picornavirus uncoating intermediate captured in atomic detail. *Nature Communications* 4, 1929.
- Sok, D., Burton, D.R., 2018. Recent progress in broadly neutralizing antibodies to HIV. *Nature Immunology* 19, 1179–1188.
- Yuan, Y., Cao, D., Zhang, Y., *et al.*, 2017. Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nature Communications* 8, 15092.
- Zhao, X., Zhang, G., Liu, S., *et al.*, 2019. Human neonatal Fc receptor is the cellular uncoating receptor for enterovirus B. *Cell* 177, 1553–1565.

Bacterial and Archeal Virus Entry

Minna M Poranen and Aušra Domanska, University of Helsinki, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

dsDNA Double-stranded DNA
dsRNA Double-stranded RNA

ssDNA Single-stranded DNA
ssRNA Single-stranded RNA

Glossary

Adsorption Initial interaction between a virus particle and a cellular receptor molecule.

Archaeal virus A virus that infects archaea.

Bacterial virus, bacteriophage or phage A virus that infects bacteria.

Capsid The protective protein coat of a virus particle.

Capsule External polysaccharide layer outside the bacterial cell envelope.

Cell envelope Plasma membrane and cellular structures located outside the plasma membrane.

Cell wall Cellular structures located outside the plasma membrane.

Lipopolysaccharide A unique glycolipid of the outer membrane of gram-negative bacteria.

Peptidoglycan A polymer consisting of long glycan chains cross-linked via peptide bridges and forming a

homogenous layer outside the plasma membrane of eubacteria.

Receptor A specific molecule or molecular assembly exposed on the surface of a cell to which a virus entering the cell attaches.

Receptor-binding protein A virion protein responsible for the interaction of a virion with a specific cellular receptor molecule.

Surface layer (s-layer) A cell surface layer composed of proteins or glycoproteins commonly found from archaea and sometimes also from bacteria.

Vertex Five-fold symmetry position of the icosahedra; one icosahedral particle has 12 five-fold symmetry positions.

Viral envelope An outer lipid-protein bilayer of a virus.

Virion A virus particle, the extracellular form of a virus.

Introduction

Viruses are intracellular parasites that are dependent on the metabolic apparatus of the cell. Unlike other parasitic self-replicating systems, such as plasmids and viroids, viruses possess an extracellular phase that allows spread from one infected cell or organism to another. Consequently viruses have to have means to infect new host cells. The entry into a suitable host cell is a key event for the viral reproduction and survival. These processes are understood in depth for several bacterial viruses. Information on archaeal virus entry is also emerging and discussed here alongside bacterial virus entry strategies.

Host Cell Barriers

The nature of the host cell envelope has a great influence on the viral entry strategy. Most bacteria and archaea have a multilayered cell envelope forming a complex physical barrier that the entering virus has to pass. Gram-positive bacteria have a single internal lipid bilayer and a thick cell wall made of peptidoglycan (also known as murein) while Gram-negative bacterial cells are covered by an internal membrane, a thin layer of peptidoglycan, and an outer membrane. The outer membrane is an asymmetric bilayer containing lipopolysaccharides in its outer leaflet. In addition, bacterial cells may secrete polysaccharides (e.g., alginate, hyaluronic acid, or polysialic acids) that make a protective extracellular capsule on the surface of the cell envelope. The extracellular capsule masks the cell surface receptors and efficiently restricts phage access. Archaeal cell surface structures are more diverse than those described for bacteria. The cell wall of the best-characterized archaea is composed of a proteinaceous surface (S-) layer made of pseudocrystalline protein or glycoprotein arrays. Archaeal cell wall may also contain peptidoglycan-like polysaccharide, called pseudomurein, or other types of polymers (e.g., methanochondroitin). In addition, wall-less forms of bacteria and archaea exist.

The relatively strong and inert cell envelope of bacteria and archaea efficiently restricts the passage of macromolecules. Moreover, bacterial and archaeal cells do not employ endocytic-like uptake systems, which are commonly exploited by eukaryotic viruses. These host features strongly influence the strategies utilized by bacterial and archaeal viruses to gain access into the host cytoplasm. In fact, the capsids of most bacteriophages are not internalized into the host cell, and only the viral genome, often with some protein factors, is delivered across the host envelope to the cytoplasm. In this way, small opening in the bacterial or archaeal cell envelope is sufficient for successful virus entry.

Virion as a Genome Delivery Device

Virions represent an extracellular form of the virus. It is a vehicle, which allows the virus to resist the harsh environment outside the cell. In addition to its protective nature, the main task of the virion is to recognize the host and to deliver the viral genome with necessary accessory factors to the new host cell. The mechanism of the genome delivery is typically reflected in the structure of the virion (see below). Thus, viruses sharing similar virion morphology often have similar entry pathways.

Furthermore, the nature of the viral genome influences the mechanisms of virus entry. Viruses that have genomes, which cannot be expressed using the enzymatic apparatuses of the host, need to also bring viral polymerases into the host cell. This applies to those bacterial viruses which have dsRNA genomes. Regardless of the type of the genome, many viruses deliver some accessory protein factors inside the host. These are required in the early stage of the infection, prior to the viral genome expression, either to complete the entry process or for successful genome replication and expression.

Host Recognition and Adsorption

Adsorption of viruses to their host cells often involves three steps: (1) initial virus contact with the cell, (2) reversible virus adsorption, and (3) subsequent irreversible adsorption. The first step, which is non-specific, brings the virus close to the cell and occurs through random events such as Brownian motion or diffusion. By reversible adsorption, the virus binds to the outermost cell surface components (i.e., the primary receptor) and then can detach and bind again, thereby searching for a specific cellular receptor (the secondary receptor). Irreversible adsorption takes place, when the virus encounters and binds to a specific cellular receptor.

The recognition and adsorption occurs via receptor-binding proteins located on the surface of the virion. These proteins recognize specific receptor molecules exposed on a susceptible cell. The receptor-binding proteins are often localized in the vertices of the icosahedral virions, the tips of the helical virions, or on the surface of the viral envelope. In tailed bacteriophages the initial recognition (reversible adsorption) is carried out by the fibers that are connected to the tail (Fig. 1(a)). The specific recognition of the cellular receptor leads to irreversible structural rearrangements in the virion components; the viral receptor-binding complexes overcome an energy barrier and fold into a minimal energy state. These conformational changes lead to more tight (irreversible) attachment and eventually trigger the entry process. The rigid structure of the virion or some part of it (e.g., one vertex of an icosahedral virion) is destabilized so that the genome delivery can be accomplished.

In terms of cellular receptors, many bacteriophages, including icosahedral ssRNA (e.g., MS2, Q β), filamentous ssDNA (e.g., Ff), and enveloped dsRNA (ϕ 6) phages, utilize a bacterial pilus as their primary receptor. Similarly, attachment to pilus-like filamentous structures has been documented for rod-shaped rudivirus (SIRV2) and enveloped filamentous lipothrixvirus (AFV1) infecting archaeal hosts. Binding to the pilus enables efficient capture of the virion at a distance from the cell surface, and the retraction of the pilus translocates the bound viruses to the host envelope (Fig. 1(b)). This allows virions to get access to the cell surface regardless of the polysaccharide capsule, which could restrict the easy access to the cell surface. Some bacteriophages (e.g., K1–5, K5, K1E), however, may also use the polysaccharide capsule as the initial site of recognition and binding, and by enzymatic degradation of the capsule get access to the cell surface (see below).

Other receptor sites for bacteriophages are cell surface lipopolysaccharides, various cell envelope and flagellar proteins, as well as cell wall carbohydrates. The tailed phages infecting Gram-positive bacteria typically attach to the cell wall teichoic acids, while those infecting Gram-negative bacteria use either the lipopolysaccharide moieties (e.g., T2, T4, T7) or envelope proteins, such as porins and transporters (e.g., PP01, T1, T5, λ) as their receptors. Among archaeal viruses, the acidianus two-tailed virus also utilizes archaeal cell surface transporter protein as its receptor.

Viral Enzymes in Capsule and Cell Wall Penetration

The virions of many bacteriophages contain enzymes, which degrade cell wall polymers and thereby facilitate virion access to the plasma membrane. These depolymerizing enzymes are encoded by the viral genome and, after expression in the cell, incorporated into the virions during their assembly. A number of different types of depolymerases have been identified in the genomes of bacteriophages, including endosialidases, hyaluronidases, levanases, xylosidases, dextranases, peptidases and pectin lyases. Beside their importance in the viral replication cycle, bacterial cell wall degrading enzymes have great biotechnological and biomedical potential.

The tail spikes of tailed phages that use exopolysaccharides for their attachment, have exopolysaccharide degrading activity which assist in the penetration of the bacterial capsule. Depending on the phage and the chemical composition of the host extracellular polysaccharide these enzymes are typically either endosialidases (e.g., *Escherichia coli* phages K1–5 and K1E) or hyaluronidase (e.g., *Streptococcus pyogenes* phage SF370.1). Other types of phage encoded exopolysaccharide depolymerases have been also described. Due to the sequential action of the polysaccharide depolymerases, the phage virion can drill through the capsule layer and reach its secondary receptors at the cell envelope.

O-polysaccharide specific enzymes are found in the virions binding to the O-polysaccharide moieties of the outer membrane lipopolysaccharides found in Gram-negative bacteria. Endorhamnosidase present, for example, in the tailed *Salmonella* phage P22 and O-antigen deacetylase present in *Escherichia* phage G7C are able to modify the O-polysaccharide chain length or its composition, respectively. The enzymatic activity of the O-polysaccharide specific enzymes apparently facilitates the irreversible binding to the host outer membrane and primes the virion for genome delivery.

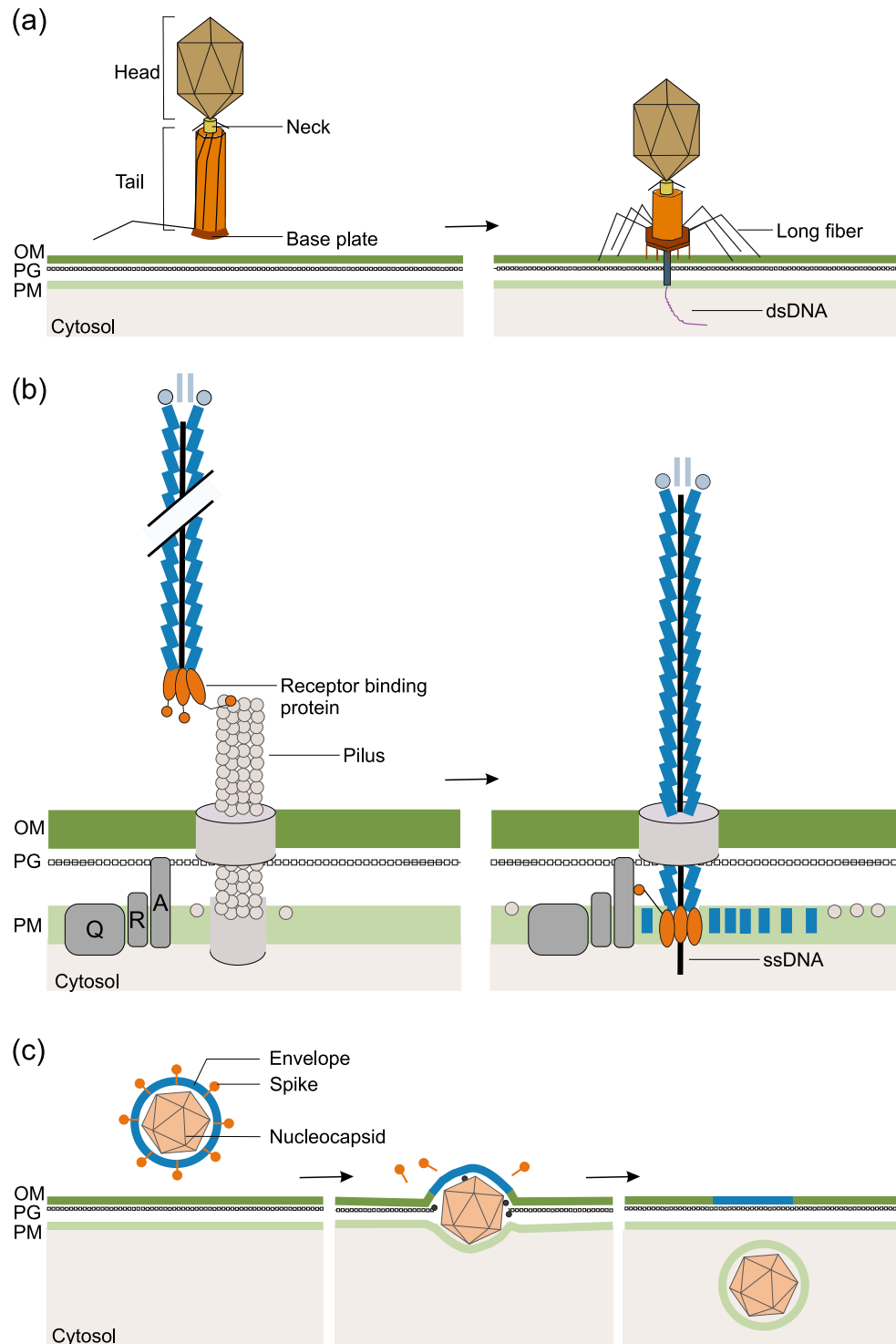


Fig. 1 Schematic presentation of the main entry strategies utilized by bacterial viruses. OM, outer membrane; PG, peptidoglycan; PM, plasma membrane. (a) Genome delivery through an icosahedral vertex; given example is a phage with long-contractile tail from the family *Myoviridae* (e.g., T4). Primary interaction between the phage and the host cell is mediated by the long tail fibers (black) and short tail fibers (brown). Contraction of the tail sheath (orange) facilitates the penetration of the cell envelope. The viral genome (purple) is released from the head of the phage virion through the tail tube (dark blue) into the host cytosol. The protein capsid remains outside the cell. (b) Dissociation of filamentous phage capsid at the cell envelope. The receptor binding protein (orange) located at the tip of the helical virion interacts with the bacterial pilus (gray). Pilus retraction brings the virion on the cell surface enabling interaction with the co-receptor molecule TolQRA complex (dark gray). The viral genome (black) is released into the host cytosol as the virion capsid proteins (blue) are inserted into the plasma membrane (PM). (c) Penetration of the nucleoprotein complex of an enveloped dsRNA virion. The viral spike proteins (orange) on the virion surface mediate the interaction between the host and the virion. Fusion between the host outer membrane (OM, green) and phage envelope (blue) takes place leading to the mixing of host and viral lipids. The nucleoprotein assembly (nucleocapsid; light orange) released into the bacterial periplasm penetrates the peptidoglycan network with the aid of a lytic enzyme (black) located on the nucleocapsid surface. Subsequently, the host plasma membrane is penetrated via an endocytic-like process.

In addition to the above discussed depolymerases, the virions of many bacteriophages contain peptidoglycan-hydrolyzing or modifying enzymes. These are specialized proteins that locally and temporarily disrupt the peptidoglycan network, thus allowing the penetration of the cell wall. Similarly to exopolysaccharides and lipopolysaccharides degrading enzymes, the peptidoglycan-hydrolyzing enzymes of tailed bacterial viruses are often associated with the tail structures or, in the case of some short-tailed phages (e.g., T7), with internal head proteins that are ejected at the beginning of the infection cycle. Icosahedral dsDNA phages with an internal membrane and no tail (e.g., PRD1, Bam35) as well as the enveloped dsRNA phage $\phi 6$ also possess peptidoglycan-hydrolyzing enzymes in their virions (see below). The most common cell-wall-degrading enzymes found in bacteriophage virions are lysozymes (N-acetylmuramidases) and lytic transglycosylases.

Genes encoding cell wall degrading enzymes, pseudomurein endoisopeptidases, have also been identified in a few tailed archaeal viruses of methanogens (e.g., ψ M2 and provirus ψ M100). However, it is not clear whether these enzymes are structural proteins of the virion, and thus could participate the cell wall penetration during virus entry, or if they contribute to the host cell lysis and release of progeny virions in the end of the infection cycle. Furthermore, most archaea do not have pseudomurein cell wall but rather have an S-layer, which displays a substantial diversity between different species. No S-layer specific proteases, which could assist in S-layer penetration have been currently identified in archaeal viruses.

Despite the wide distribution and variety of phage encoded enzymes, not all viruses require specialized enzymes for the penetration of the host cell wall during entry. Several phages are known to rely on preexisting channels within the host envelope. This applies to icosahedral and filamentous ssDNA bacteriophages (e.g., ϕ X174 and M13, respectively) as well as icosahedral ssRNA bacteriophages (e.g., MS2). For example, infection of filamentous ssDNA phages absolutely depends on a protein complex assembled in the bacterial plasma membrane spanning the peptidoglycan layer (Fig. 1(b)).

Genome Delivery Mechanisms of Phages

In general, entry strategies of bacterial viruses fall into three main categories: (1) genome delivery through an icosahedral vertex, (2) virion dissociation at the cell envelope, and (3) virion penetration via membrane fusion and an endocytic-like event (Fig. 1). Tailed dsDNA phages (e.g., T4, T5, T7), dsDNA phages with an internal membrane (tequiviruses), as well as icosahedral ssDNA and ssRNA phages deliver their genome through a genome delivery apparatus located at one of the capsid vertices (Fig. 1(a)). Capsids of filamentous phages (e.g., M13) and membrane-containing icosahedral phage PM2 disassemble completely at the cell envelope (Fig. 1(b)), whereas the enveloped dsRNA bacterial viruses utilize a unique membrane fusion-type uncoating at the outer membrane and virus subparticle internalization through the plasma membrane (Fig. 1(c)).

Icosahedral Tailed dsDNA Bacterial Viruses

Tailed dsDNA bacteriophages represent the most abundant morphotype among bacterial viruses. The tailed viruses are also common among known isolates of archaeal viruses. The virion of such viruses is comprised of an icosahedral capsid (head) enclosing highly condensed viral DNA, and a helical tail attached to a unique vertex of the capsid. The tail is a multiprotein tubular apparatus responsible for the host cell recognition, attachment and cell envelope penetration. Despite the numerous isolates of tailed viruses, they all possess one of the following three types of tails: long contractile (family *Myoviridae*, e.g., T4; *Ackermannviridae*; *Herelleviridae*), long noncontractile (family *Siphoviridae*, e.g., T5), or short noncontractile tail (family *Podoviridae*, e.g., T7 and ϕ 29). The type of tail influences the genome delivery mechanism utilized by the tailed viruses.

Bacteriophage T4 infecting Gram-negative bacteria is a classic example of a tailed phage. The contractile T4 tail is composed of four major parts: the terminator, the rigid central tube, the contractile sheath around the tail tube, and the base plate complex. The central tube is made of many copies of the tail tube protein, which are stacked around an extended tape measure protein present in few copies. A contractile sheath built of many copies of the tail sheath protein surrounds the tail tube. The proximal end of the tail ends with the tail tube terminator and the sheath terminator, and attaches to the head via neck and portal structures. The portal is a specialized vertex, through which the viral DNA is packaged into the head. The distal end of the tail tube is attached to the base plate, a complex structure decorated with long and short tail fibers terminating the tail (Fig. 1(a)). The long tail fibers act as sensors to detect the host bacterium and serve for the primary attachment. The short tail fibers, initially folded beneath the base plate, unfold and bind irreversibly to the host bacterium. Interaction between the tail fibers and their surface receptors leads to conformational changes in the base plate (base plate transition from dome-shaped to star-shaped) which triggers contraction of the tail sheath to about half of its original length (Fig. 1(a)). Due to the contraction of the tail sheath and the irreversible attachment of the tail fibers to the cell surface, the tail tube moves towards the cell envelope and penetrates it (Fig. 1(a)). The peptidoglycan-hydrolyzing enzyme associated with the tip of the tail tube facilitates the penetration of the peptidoglycan layer. Eventually the tail tip reaches the plasma membrane and fuses with it. The force required for penetration of the tail tube through the cell envelope is associated with the contraction of the tail sheath. Some viral proteins enter the cell along with the phage DNA in order to protect it from host exonuclease activities and are implicated in the early stages of phage genome transcription.

The long noncontractile tail of siphoviruses such as λ , T5, HK97 and SPP1, is composed of trimers (e.g., T5) or hexamers (e.g., SPP1) of a tail tube protein polymerized around few copies of the tape measure protein. The tape measure proteins in their elongated conformation span the whole tail tube and the C-termini of these proteins together with several other proteins form the tail tip complex at the distal end of the tube. The receptor binding protein of T5 is part of the tail tip. The relatively flexible tails of

siphoviruses may also possess the long tail fibers serving for reversible attachment to the host cell as exemplified by T5 and λ phages. In contrast to myoviruses, tails of siphoviruses do not possess a contractile sheath and therefore do not contract during DNA transfer from the virion to the cell. The tail tube protein of T5 does not undergo conformational changes upon receptor binding and thereby plays no role in the signal transmission from the tail tip to the head. The conformational changes mostly concern the tail tip and therefore it is suggested that the tape measure protein is involved in the signal transmission from the distal end of the tail to the head upon receptor binding. The tape measure proteins are released from the tail tube into the periplasm of Gram-negative host bacterium after irreversible binding to the receptor (FhuA in the case of T5 or LamB in the case of HK97). The host cell proteins such as periplasmic chaperones and inner membrane proteins are involved in the later stages of the siphovirus genome translocation to the cell cytoplasm. The periplasmic chaperones (FkpA in HK97 infection) might aid the tape measure proteins in adopting a conduit conformation suitable for DNA transfer through the cell envelope. The inner membrane proteins of the host cell (PtsG in HK97 infection) possibly facilitate the phage entry by anchoring the viral protein conduit to the inner membrane or by providing a pore in the inner membrane for DNA transfer.

Podoviruses possess short noncontractile tails and, compared to other tailed viruses, exhibit simplest tail morphology. The tails of podoviruses, like P22, ϕ 29 and T7, are formed by a nozzle also known as central knob, which is decorated with trimeric appendages called fibers or spikes. The tail is attached to the unique vertex of the head via adapter proteins forming a dodecameric ring. The podovirus tails are usually too short to span the host cell envelope and therefore additional proteins are needed for DNA translocation across the cell envelope. In the intact virion these proteins either form the internal core structure of the head (like in the case of T7 and P22) or are located in the tail (for example the terminal loops of the ϕ 29 tail knob). Bacteriophage T7 infecting Gram-negative bacteria uses tail fibers to recognize the lipopolysaccharide moieties of the host rough strains (the lipopolysaccharide molecules of such strains lack the O-antigen). Irreversible adsorption of T7 tail to the outer membrane leads to the ejection of several proteins from the virion to the cell's periplasm. After ejection, these proteins associate with the inner and outer membranes of bacterial envelope creating a channel for DNA translocation spanning the periplasm. This process has been visualized by cryo-electron tomography showing conduit in the bacterial periplasm made of viral core proteins, which were ejected from the head upon irreversible T7 adsorption. Moreover, one of the proteins ejected from the virion possess the transglycosylase activity, which helps to overcome the peptidoglycan layer of the cell envelope. After conduit is assembled, the transfer of T7 DNA from the virion to the cytoplasm occurs in a stepwise manner. First, a small portion of the viral DNA (approximately 1 kb out of 40 kb) exits from the virion spontaneously. Subsequently, the rest of the genome is internalized enzymatically via transcription by host RNA polymerase and then by viral RNA polymerase. The stepwise mode of viral genome translocation provides effective means for gene regulation, as genes that first enter the cell are expressed first; genes encoding structural proteins are required late in the infection and are expressed late because they are not present in the cell at early stages of infection.

Podoviruses infecting Gram-positive bacteria such as ϕ 29, have to penetrate the thick layer of peptidoglycan constituting the cell wall of these bacteria. The ϕ 29 entry initiates when appendages of the virion attach to the cell wall teichoic acids protruding on the cell surface. The enzymatic activity present in the appendages digests the teichoic acid chains thereby pulling the virion closer to the cell wall. When the ϕ 29 tail touches the peptidoglycan layer, the enzymatic activity present in the tail tip hydrolyzes the peptidoglycan thus clearing the path for the virion towards the cytoplasmic membrane. Upon interaction with the membrane, the tail tip undergoes structural changes and transmembrane pore is formed through which viral DNA is delivered to the cytoplasm. The hydrophobic loops present at the end of the tail knob of ϕ 29 were suggested to form a transmembrane pore for DNA translocation.

The archaeal tailed viruses share similar morphology with the bacterial myoviruses, siphoviruses and podoviruses. It is expected that they use the tail for genome delivery and that their entry pathways are similar to the above described tailed bacterial viruses.

Icosahedral DNA Viruses With an Internal Membrane

Several icosahedral bacterial and archaeal viruses have lipid-protein membrane residing beneath the protein capsid. Despite of the structural similarity, the genome of such viruses can be either linear or circular dsDNA (PRD1 and PM2, respectively), or circular ssDNA (FLiP). The role of the internal membrane in the entry process is best characterized for bacteriophage PRD1, the type member of the *Tectiviridae* family.

Bacteriophage PRD1 infects Gram-negative hosts which harbor a multiple drug resistance conjugative plasmid. This plasmid encodes the receptor complex that is recognized by the spikes located at the icosahedral vertices of the PRD1 virions. Specific interactions between the phage and the host induce dissociation of the spike complex from one of the capsid vertices and subsequent formation of a pore in the viral capsid. The internal membrane is transformed into a proteo-lipidic tube, which protrudes from one of the capsid vertices and provides a conduct for passage of the viral genome through the outer layers of the host cell, analogous to the tails of the dsDNA bacteriophages discussed above. Lytic transglycosylase associated with the PRD1 membrane assist the tube in the peptidoglycan penetration. Once the tip of the tube has reached the host cytoplasm it unplugs resulting in the translocation of the viral genome via the tube to the host interior. The rigid protein capsid stays intact outside the cell. The members of *Tectiviridae* that infect gram-positive bacteria (e.g., Bam35) attach directly to the host cell peptidoglycan structure. The internal membrane of these viruses can also transform into a tubular structure suggesting similar mechanism for DNA entry into the host cell.

Bacteriophage PM2 (*Corticoviridae*) infects gram-negative marine bacteria from genus *Pseudoalteromonas*. Although PM2 share a similar structural organization with PRD1 and other tectiviruses, it uses a distinct entry pathway. The capsid of PM2 dissociates on the host cell surface releasing the internal lipid core enclosing the viral DNA. A current model suggests that the viral membrane then fuses

with the bacterial outer membrane delivering the highly supercoiled dsDNA genome into the periplasm, from where it travels to the host cytosol. Whether the recently isolated membrane containing ssDNA phage FLiP (family *Finnlakeviridae*), or the archaeal dsDNA viruses having integral membrane layer in their virions (e.g., SH1, HCIV-1, and HHIV-2 of the family *Sphaerolipoviridae*) have entry pathways similar to PRD1 or PM2, or use alternatively some yet unknown entry mechanisms, remains to be discovered.

Pleomorphic DNA Viruses

The quasi-spherical virions of pleomorphic dsDNA bacterial viruses (*Plasmaviridae*) are nucleoprotein complexes within a lipid-protein membrane. The only classified member of the family *Plasmaviridae* is Acholeplasma phage L2. Acholeplasma are wall-less bacteria; thus, the phage adsorbs directly on the plasma membrane. It is assumed that adsorption leads to fusion of viral and host cell membranes resulting in entry of the nucleoprotein complexes into the cell. Similar entry strategy has also been proposed for haloarchaeal pleomorphic ssDNA and dsDNA viruses of the family *Pleolipoviridae* including Halorubrum pleomorphic virus 2 and 6 (HRPV-2 and HRPV-6). The virion of these viruses is decorated by a monomeric envelope surface protein which can mediate the fusion between the viral membrane and the host cytoplasmic membrane.

Icosahedral ssDNA Bacterial Viruses (With no Membrane)

The icosahedral ssDNA bacterial viruses of the family *Microviridae* infect Gram-negative bacteria (e.g., Escherichia phages ϕ X174, G4, and α 3) or obligate intracellular parasites lacking a cell wall, like chlamydia (e.g., Chp2) and spiroplasma (SpV4).

Icosahedral ssDNA bacterial viruses infecting Gram-negative bacteria contain large pentameric spike complexes (made of protein G in ϕ X174) at the vertices of the virion. These spikes are responsible for the initial binding of the phage to the host cell envelope. Binding to the lipopolysaccharide is mediated by one of the twelve spikes and it results in the dissociation of the interacting spike. This triggers conformational changes in the exposed major capsid protein resulting in its attachment to the lipopolysaccharide, and opens a gate in the capsid vertex bound to the cell surface. Additional changes in the pilot protein (H) and the ssDNA genome within the capsid further prime the virion for genome delivery. The pilot protein oligomerizes into super-helical coiled-coil structure which forms a tube-like conduit for genome translocation through the cell envelope. Both the genome and the pilot protein are eventually internalized into the host cell cytoplasm while the viral capsid lacking DNA and one of 12 minor spike proteins remains outside the cell. Entry pathways of microviruses infecting wall-less parasitic bacteria is less well studied and therefore not discussed here.

Filamentous DNA Viruses

Most of the filamentous non-enveloped ssDNA viruses from family *Inoviridae* identified to date infect Gram-negative bacteria, like Escherichia phages M13, F1, and Fd (known as Ff phages), or wall-less bacteria like Acholeplasma virus L51 and Spiroplasma virus C74. The Propionibacterium virus B5 is the only classified filamentous ssDNA phage infecting Gram-positive bacteria. Filamentous viruses infecting archaea usually possess dsDNA genomes and are more diverse in their morphology. Filamentous archaeal viruses that belong to families *Lipotrixviridae* and *Tristromaviridae* are enveloped and those that belong to *Rudiviridae* and *Clavaviridae* are non-enveloped viruses. The Ff phages infecting Gram-negative bacteria are the best studied filamentous viruses. The virion of Ff phages is composed of thousand copies of major coat protein pVIII helically arranged around the circular ssDNA. Both ends of the virion are terminated by two different minor coat proteins, pVII and pIX at one end, and pIII and pVI at the other end. The receptor-binding protein of Ff phages, pIII, is located at one end of the phage filament. The primary binding site of these viruses is bacterial pilus, a long thin proteinaceous structure on the surface of bacteria. The Ff phages adsorb to the tip of F pilus (conjugative type IV secretion pilus), other filamentous phages may also adsorb to the type IV pili (T4P). Binding of pIII protein to the tip of F pilus induces conformational changes in the pIII uncovering the receptor binding domain. Both type of pili can retract thereby bringing attached phages close to the receptor located in the periplasm of bacterial cells. How exactly filamentous phages cross the outer membrane is not known. Possibly they can use pilus structures spanning the outer membrane to reach the periplasm. To traverse the periplasmic space, most filamentous phages utilize a trans-envelope protein complex, Tol system. The exposed receptor binding domain in pIII of Ff phages interacts with the periplasmic part of TolA subunit, which is anchored to the inner membrane via its transmembrane helix. The pIII interaction with TolA leads to the insertion of pIII and the major coat protein pVIII subunits into the bacterial inner membrane resulting in stepwise dissociation of the filamentous capsid and concomitant uncoating of the phage genome at the cytoplasmic membrane. Eventually, this uncoating process results in the delivery of the viral ssDNA genome into the cytoplasm of the host cell (Fig. 1(b)). The TolQ and TolR subunits making a complex with TolA in the inner membrane can facilitate this process, though their exact role is not clear. The entry of filamentous viruses infecting Gram-positive bacteria and archaea has not been elucidated to date.

Icosahedral Enveloped dsRNA Bacterial Viruses

The dsRNA bacteriophages of the family *Cystoviridae* have a unique entry mechanism among bacteriophages. Similarly to eukaryotic dsRNA viruses, cystoviruses deliver their genome into the host cytoplasm within a large protein capsid, which protects the enclosed dsRNA genome throughout the infection cycle.

The enveloped virion of dsRNA bacteriophages contains three structural layers that sequentially assist in the penetration of the host outer membrane, the peptidoglycan layer, and the plasma membrane. The spikes protruding from the virion surface are involved in the receptor recognition and binding (**Fig. 1(c)**). The dsRNA bacteriophages adsorb either to a bacterial pilus (e.g., *Pseudomonas* phage $\phi 6$), which then retracts, or to lipopolysaccharide on the surface of the gram-negative host bacterium (e.g., *Pseudomonas* phage $\phi 13$). After phage adsorption at the outer membrane, the viral envelope fuses with the host membrane, thus uncoating the virion and placing the resulting viral nucleoprotein complex (or nucleocapsid) into the periplasm (**Fig. 1(c)**). The fusion between the viral envelope and bacterial outer membrane is driven by the transmembrane proteins of the viral envelope, whose fusogenic properties are activated after virus adsorption. The lytic enzyme of the virus located between the envelope and nucleocapsid locally digests the host cell-wall peptidoglycan, thereby allowing the nucleocapsid to reach the plasma membrane (**Fig. 1(c)**). The nucleocapsid then binds to the plasma membrane lipids inducing the formation of plasma membrane invagination. Eventually, the nucleocapsid penetrates the plasma membrane via an endocytic-like route. This involves the formation of the plasma membrane curvature at the contact site resulting in an intracellular vesicle, which then pinches off from the plasma membrane (**Fig. 1(c)**). This process is dependent on the energized host plasma membrane. The transcriptionally active viral particle is released from the vesicle into the host cytosol.

Icosahedral ssRNA Bacterial Viruses

Icosahedral ssRNA viruses that infect bacteria belong to a single family *Leviviridae*. Thus far there are no known RNA viruses infecting archaea. The best studied leviviruses are *Escherichia* phage MS2 and Q β . The virion of MS2 contains ssRNA genome enclosed by 89 coat protein dimers and a single copy of maturation protein, which replaces one of the coat protein dimers in the capsid. The maturation protein strongly interacts with the 3' end of the viral RNA and is also responsible for the attachment to the bacterial F pilus. Similarly to the other viruses utilizing pili, retraction of a pilus brings the virion to the surface of the bacterial cell. Binding to the pilus does not trigger the genome release which suggests that the energy required to pull the genome out of the capsid comes from the pilus dynamics, rather than the attachment. Eventually, the maturation protein dissociates from the capsid and is delivered to the cell cytoplasm together with the tightly bound RNA. The empty capsid is left outside the cell. Bacteriophages with ssRNA genomes do not possess peptidoglycan-hydrolyzing enzymes present in many other bacterial viruses to facilitate the peptidoglycan penetration during the entry process. Instead, ssRNA phages rely on the host cell pilus to reach the cytoplasm.

Summary

Many bacterial viruses encode different types of cell wall degrading or modifying enzymes that are incorporated into the virion. These enzymes help viruses to get access to the host plasma membrane during entry. Typically, the viral genome is delivered into the host via a specific tubular structure which is part of the virion or is formed upon virion binding to the host cell surface. Regardless of the origin and composition of the tube, it is used to deliver the genome from the icosahedral viral capsid to the host cytoplasm, while most of the other parts of the virion remain outside the cell. Alternatively, the virion dissociates at the cell envelope resulting in the uncoating of the genome and its delivery into the host cell. In rare cases, the genome enters into the host cell within an icosahedral viral capsid.

Further Reading

- Bertozi, S.J., Storms, Z., Sauvageau, D., 2016. Host receptors for bacteriophage adsorption. *FEMS Microbiology Letters* 363, fnw002.
- Casjens, S.R., Molineux, I.J., 2012. Short noncontractile tail machines: Adsorption and DNA delivery by podoviruses. In: Rossmann, M., Rao, V. (Eds.), *Viral Molecular Machines: Advances in Experimental Medicine and Biology* 726. Boston, MA: Springer, pp. 143–179.
- Dai, X., Li, Z., Lai, M., *et al.*, 2017. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* 541, 112–116.
- Duché, D., Houot, L., 2019. Similarities and differences between colicin and filamentous phage uptake by bacterial cells. *EcoSal Plus* 8, EPS-0030-2018.
- El Omari, K., Li, S., Kotecha, A., *et al.*, 2019. The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. *Nature Communications* 10, 846.
- Fernandes, S., São-José, C., 2018. Enzymes and mechanisms employed by tailed bacteriophages to breach the bacterial cell barriers. *Viruses* 10, 396.
- Mäntynen, S., Sundberg, L.R., Oksanen, H.M., Poranen, M.M., 2019. Half a century of research on membrane-containing bacteriophages: Bringing new concepts to modern virology. *Viruses* 11, 76.
- Peralta, B., Gil-Carton, D., Castaño-Díez, D., *et al.*, 2013. Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLoS Biology* 11, e1001667.
- Poranen, M.M., Daugelavicius, R., Bamford, D.H., 2002. Common principles in viral entry. *Annual Review of Microbiology* 56, 521–538.
- Quesin, E.R.J., Quax, T.E.F., 2015. Archaeal viruses at the cell envelope: Entry and egress. *Frontiers in Microbiology* 6, 552.
- Sun, Y., Roznowski, A.P., Tokuda, J.M., *et al.*, 2017. Structural changes of tailless bacteriophage Φ X174 during penetration of bacterial cell walls. *Proceedings of the National Academy of Sciences of the United States of America* 114, 13708–13713.
- Xu, J., Xiang, Y., 2017. Membrane penetration by bacterial viruses. *Journal of Virology* 91, e00162.

Nonenveloped Eukaryotic Virus Entry

Ian M Jones, University of Reading, Reading, United Kingdom and London School of Hygiene and Tropical Medicine, London, United Kingdom

Polly Roy, Department of Pathogen Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom and University of Reading, Reading, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Preamble

Non-enveloped viruses represent a diverse range of infectious agents found across the phyla. They include well known viruses, largely as a result of the pathology they cause, such as poliovirus, rotavirus and the papilloma viruses, to silent yet widely distributed viruses such as transfusion transmitted (TT) virus. Some non-enveloped viruses are classified as emerging viruses such as Deformed Wing Virus of the honeybee, enterovirus D68 in man and circovirus infection in mammals and birds. Indeed, any list of non-enveloped viruses of concern would date rapidly as more and more members are added. The diversity of non-enveloped viruses is mirrored by the pathology they cause which in turn is related to the type and range of tissues infected. Their mode of transmission, airborne, blood borne or by direct contact is also dictated in many cases by their sites of replication. Central to this diversity and its consequences is the target cell for virus entry and the mechanisms used by the virus to gain access into it. With the exception of plant viruses, which require physical breakage of the hard cell wall to enter, the outline process is the same for the majority of viruses, non-enveloped included; locate to the cell surface using a specific or general receptor, enter the cell in an invaginated vesicle (no non-enveloped virus fuses at the plasma membrane on entry), become wholly vesiculated and finally breach the vesicle membrane wholly or partially to gain entry to the cytoplasm or organelle (**Fig. 1**).

However, this simple outline belies a complexity of mechanism that is still the subject of active investigation: what does the virion sense, using which capsid components, to initiate vesicle exit, what conformational changes accompany entry, how is the necessarily unstable state of being primed for uncoating balanced with the need for virion stability, and not least, how does a large, often charged virus particle deform, dissolve or pass through a hydrophobic lipid bilayer. In addition, the act of uncoating exposes the virus nucleic acid to the cellular environment where it may be degraded or engaged by sensors of the innate immune system. This step must therefore be as quick as possible to minimize the potential danger to virus replication. The overall process is quite well understood for some viruses although as ever in virology, diversity abounds. For example, extensive membrane breaching studies have been completed with small, easily grown model viruses, such as Flock House Virus (FHV), an insect virus widely studied for the relationship between its structure and function. However, the identity of the receptor for this virus has never been defined, as is the case for most, if not all, viruses of insects. It is clear that the virus undergoes morphological change after contact

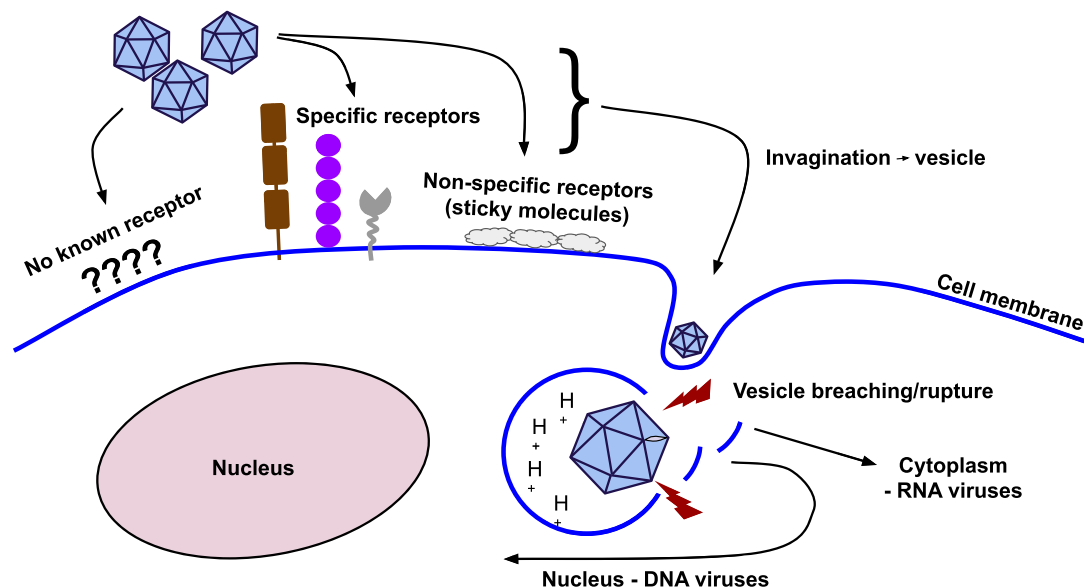


Fig. 1 Overview of non-enveloped virus cell entry. Virus attachment to the cell surface occurs via a range of receptors. Highly specific receptors tend to be tissue restricted cell surface proteins whereas less specific interactions occur on molecules that are widely distributed, such as sialic acid, heparins or integrins. In many cases the receptor remains unknown. Trapping to the cell surface leads to virus entry by one of the defined mechanisms of endocytosis which ultimately delivers the virus into a membrane bound vesicle. Many, but not all, vesicles become as acidic as they mature, triggering a conformational change in the virus which then breaches the membrane to enter the cell.

with a permissive cell surface but the identity of the molecules contacted has not been clearly described. By contrast, the entry reaction, once triggered, is well characterized. Thus, the precise understanding in one part of the entry process is offset by a lack of detail in another, a situation that is typical across the non-enveloped virus group as a whole.

Simple Non-Enveloped Viruses

While no virus is simple, non-enveloped viruses with a single capsid shell arguably have less complexity to deal with during the entry process as once the capsid is disrupted the genome can leave. For many classically studied examples, initial cell contact is via a specific receptor, an example being use of members of the immunoglobulin superfamily in the case of many picornaviruses (small round viruses with positive sense single stranded RNA genomes). Poliovirus uses CD155 whereas classical rhinoviruses use intercellular adhesion molecule-I, ICAM-I. The enteroviruses (those picornaviruses that infect in the gastrointestinal tract) use a range of receptors, for example both human scavenger receptor class B, member 2 (SCARB2) and P-selectin glycoprotein ligand-1 (PSGL1) are ligands for EV71 with SCARB2 acting as the critical entry molecule while both a general receptor, sialic acid, and the neuron specific intercellular adhesion molecule 5 (ICAM-5) act as receptors for enterovirus D68. Rhinovirus C, an emerging respiratory pathogen, uses cadherin-related family member 3 (CDHR3) and, in a very clear example of the link between receptor use and pathology, single nucleotide polymorphisms (SNP) in CDHR3 associated with asthma and wheezing in children also mediate the level of virus binding. Similarly, SNPs in the presumed hepatitis A virus receptor, T-cell immunoglobulin and mucin domain 1 (TIM-1), have been linked to enhanced virus entry and the severity of infection although in this case TIM-1 also functions as a known immune signaling molecule so the link to solely virus induced pathology is less clear cut. A specific receptor, junctional adhesion molecule 1 (JAM-1) can act as receptor for some caliciviruses but for others a general carbohydrate moiety is used while for other related small round RNA viruses, for example astroviruses, no defined receptors have yet been described. In other non-enveloped viruses too, a range of both highly specific and more general receptors are utilized to locate the virus to the cell surface. Parvoviruses, small round DNA viruses, use both carbohydrate binding, principally to sialic acid, and distinct protein receptors such as the transferrin receptor (TfR) to attach to the cell, and other small DNA viruses such as the papillomaviruses also demonstrate binding to both general and specific receptors, heparan sulfate and members of the integrin family of cell surface proteins respectively. There remain also many cases of unknown receptor usage across the non-enveloped virus group as a whole, sometimes through lack of experimentation but in other cases because no clear candidate has emerged, despite considerable experimental effort.

Notwithstanding the diversity of primary contacts, location to and active engagement with the cell surface of a permissive cell leads to uptake of virions into the cell by one of three well described mechanisms, macropinocytosis, clathrin mediated endocytosis and caveolin mediated endocytosis. Macropinocytosis is an actin-dependent uptake mechanism in which a membrane ruffle engulfs the virus particle, clathrin mediated endocytosis uses clathrin coated pits at the plasma membrane, often containing large receptor molecules, while endocytosis in caveolae use the protein caveolin as mediator of invagination of a cholesterol rich membrane, also enriched with large protein receptors (Fig. 2). Viruses using defined protein receptors will more commonly use one of the latter two mechanisms of entry although a mixed route of entry is not uncommon. Vesicles trafficked by one of the former two mechanisms will migrate, mature or merge with others in the cell cytoplasm and change the environment, notably the intra-endosomal pH, that surrounds the vesiculated virus, to trigger conformational changes that enable vesicle escape. Viruses entering via the caveolae route tend not to become acidified but are nonetheless trafficked to other organelles on cell entry.

In contrast to the diversity of entry routes, there is an emerging consensus on the shared features that describe the conformational change that many viruses undergo as a result of the changed environment of the endosome. To bind to, traverse or rupture the cellular membrane the engulfed virus must become more hydrophobic and this is achieved by the revelation of a peptide or protein containing hydrophobic residues often arranged as an amphipathic helix. Here, the canonical studies on FHV have provided a model. FHV assembles from 120 copies of the ~400 amino acid capsid protein, protein α , arranged as an icosahedral shell of 60 trimers. Following assembly the α protein cleaves autocatalytically to form capsid protein β and peptide γ , ~360 and ~40 residues respectively, the latter of which contains an amphipathic helix in the N-terminal half of the molecule and three phenylalanine residues at the C terminus. The γ peptide is located on the inside of the capsid following its release from protein α although it is transiently exposed by a process termed particle breathing, a reversible process by which the virus samples the environment before committing fully to uncoating. When it encounters acidic pH however, or exceptionally following contact with the receptor itself, the γ peptide is fully externalized via changes in capsid conformation and acts to destabilize the vesicle membrane. The mechanism of membrane destabilization is likely to be detergent-like as holes of 50–500 nm have been detected in the membrane following incubation with isolated peptide γ , easily sufficient to allow the two positive strand RNA molecules of 4.5 kilobase pairs, which constitute the virus genome, to pass through. This model, a maturation dependent cleavage of the capsid to generate a protein that is required for cell entry applies also to other groups of small non-enveloped viruses. In the picornaviruses, assembly requires 60 copies of each of three related proteins, VP0, VP1 and VP3, but within the newly assembled shell the VP0 protein is cleaved further, again autocatalytically, to generate VP2 and VP4, the latter of which is required for membrane permeabilization. VP4 is myristoylated and this post translational mechanism, which adds a long chain hydrophobic moiety to the N-terminus of the protein, is required for the entry reaction. As for the γ peptide of FHV, VP4 is predominantly found on the inside of the mature virus particle although its N-terminus is also transiently exposed by virion breathing. Following receptor triggering and/or low pH however, the particle undergoes irreversible change, which results in an opening at the 2-fold axis of symmetry and the emergence of VP4, which locates to the proximal membrane to form a pore through which the genome RNA can

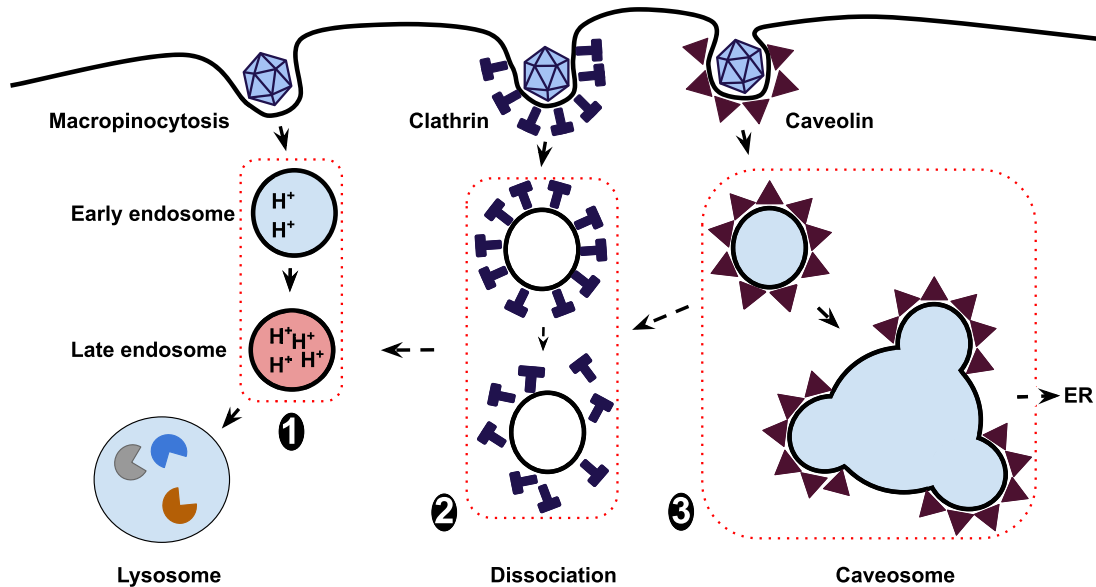


Fig. 2 The defined uptake mechanisms of virus following receptor binding. Viruses enter cells to become enclosed in a vesicle by the same processes cells use to capture other cargo. They may use more than one pathway depending on the receptor bound. Pathway 1 involves the canonical acidified vesicle that follows macropinocytosis. Pathway 2, possibly the most widely used virus entry pathway, uses clathrin coated pits which subsequently may also become acidified. In both cases viruses must escape the vesicle during its maturation and before it fuses with the lysosome. Pathway 3 uses membrane patches enriched in certain lipids and different adapter proteins, caveolins, to mediate the inward budding process. Although some caveolae may also follow the endocytic route to a lower pH, an alternate non pH-mediated pathway via the formation of large caveosomes can lead to delivery of part disassembled viruses to other cellular compartments, such as the endoplasmic reticulum.

pass. The N-terminus of VP1 also emerges during this process and may serve to anchor the opening virion to the membrane allowing VP4 to function wholly in genome release. A somewhat similar mechanism was recently shown to operate also in the caliciviruses, small round viruses causing respiratory and gastrointestinal infections, where a minor capsid protein, VP2, forms a funnel-like pore extending away from the 3-fold axis of symmetry of the viral icosahedral structure following receptor engagement. The distal end of the VP2 structure is very hydrophobic and is assumed to embed in the host membrane to facilitate genome release into the cytoplasm through the funnel.

For some small non-enveloped DNA viruses the precise mechanisms of entry remain debated but the underlying principles leading to disassembly remain recognizable. Entry for viruses such as human papilloma virus, and the related polyoma viruses, is via the caveolae pathway leading to an endosome that also becomes acidic as it matures. Virus escape is not directly into the cytoplasm but rather into the endoplasmic reticulum (ER) as the virus must access the nucleus for replication and the continuity of the ER with the outer nuclear membrane provides a directly accessible route. There is conflicting data on exactly when and to what extent the virus dissociates but it is clear that there is a role for pH as well as for proteolytic enzymes, all present in the caveolin mediated vesicles. A role for host chaperones, particularly those that function as redox exchangers, and possibly a number of other trafficking molecules, has also been described. Together, their actions result in the loss of the major capsid protein VP1, leaving the minor capsid protein VP2 still associated with the genome. It is the VP2-genome complex that completes the trafficking process and enters the nucleus, probably when the nuclear membrane dissolves at cell division. The somewhat smaller and simpler parvoviruses, a member of which, adeno-associated virus (AAV), is in widespread use as a gene delivery vector, follow a similar pathway. A mix of specific (large protein) and general (sialic acid) receptors mediate the initial contact, depending on the strain of virus used, leading to virus uptake into endosomes predominantly via the clathrin dependent pathway. As for the papillomaviruses the precise stage at which endosomal escape takes place and the exact triggers that enable it, are poorly defined. Nevertheless, roles for a low pH acid activated proteolysis have been reported which together modify the capsid to allow membrane breaching. In a manner similar to the particle breathing noted for FHV and the picornaviruses, the N-terminus of capsid protein VP2 tests the environment, progressively increasing virus porosity in the endosomal environment until the N-terminus of VP1 protrudes from the five-fold vertices of the capsid. The virus does not encode a distinct lytic peptide, perhaps as a consequence of the limited coding potential associated with a small genome, but the VP1 protein contains a phospholipase A2-like domain whose activity is essential for endosomal escape. The overall process appears to be inefficient however, with many virus particles failing to escape the endosome and being delivered to the lysosome where they accumulate and degrade. Processing of the parvovirus capsid in the endosome is necessary as it equips it for engagement with cellular factors required for transport to the nucleus, principally the microtubule network and its associated motors, and direct injection of the virus into the cellular cytoplasm fails to recapitulate the entry reaction and so does not lead to a productive infection. As far as has been described, entry for circoviruses, small viruses with circular single stranded DNA genomes, is similar, clathrin mediated uptake after which viral antigen is found accumulated in

the lysosome. This implies that escape of the circular DNA genome must occur before this point but there is little detail of the mechanism involved beyond the mapping of a nuclear localization signal in the N-terminus of the single capsid protein of the virus.

Complex Non-Enveloped Viruses

Non-enveloped viruses with multiple layers face an additional dimension of challenge in that uncoating is often only partial, either to protect the nucleic acid from the sensors of innate immunity or because a second step is required before the replication cycle can begin, for example transport to the nucleus. In these cases the virus must arrest the process of disassembly before the genome is exposed and a core or sub-core is what enters the cytoplasm or other cellular compartment. As for the simpler viruses, primary contact with the cell can be highly specific via defined receptors or general via electrostatic charge or glycan interaction. For most human adenoviruses, large icosahedral DNA viruses with a linear double-stranded DNA genome, the human receptor is CAR, the Coxsackie and Adenovirus Receptor, a member of the immunoglobulin superfamily involved in cell to cell contact. A subset of adenoviruses also use CD46, another Ig superfamily member and some reports have also noted other receptors or attachment factors, including sialic acid. For the *Reoviridae*, complex multi-layered viruses with segmented double-stranded RNA genomes, several receptors have been described. For orthoreovirus (10 dsRNA segments) a single cellular tight junction protein, JAM-A, acts as viral receptor followed by association with various carbohydrates, including sialic acid, present on cell surface glycoproteins and glycolipids, which act as co-receptors. For the related rotaviruses (11 dsRNA segments), widely studied as they are a primary cause of childhood gastrointestinal disease, general and specific receptors again feature. Binding to the cell surface by some strains of rotavirus can be prevented by treatment with neuraminidase, which strips sialic acid from glycan chains, showing that sialic acid is a binding factor, but other sugars including those that occur internally can also act as binding sites. Some of these glycans make up the histo-blood group antigens and human rotaviruses can be distinguished by which of these antigens, H-type or A-type, they bind. More specific receptors are also used and many candidates have been described for various rotavirus strains including cell surface integrins and the ubiquitous chaperone heat shock cognate protein, hsc70. Tight-junction proteins, including JAM-A as used by reovirus, are also used by some rotavirus strains. For the orbivirus genus of the *Reoviridae*, insect viruses that nonetheless infect mammalian species, no clear receptors have emerged. For Bluetongue virus, the type species of the genus and responsible for economically important disease in sheep and cattle, lectins that bind to cell surface sugars reduce infectivity consistent with use of a sugar as receptor and potential sialic acid binding sites have been noted in the structure of the outermost capsid protein, VP2. However, sugar blocking or removal does not abolish infection suggesting that other, so far undiscovered, receptors must exist including those on insect cells, which remain wholly unknown.

Following binding to the cell surface complex non-enveloped viruses are taken up by the same mechanisms as those discussed for the simple non-enveloped viruses with a tendency more towards the use of clathrin and caveolin mediated pathways rather than the macropinocytosis pathway. For adenovirus, an example of a large dsDNA virus, the precise entry process is complex. The receptor is contacted by the penton fiber, a distinct structure on the virus particle that extends from the vertex of the icosahedral shell and which, with the penton base, mediates the overall process of cell entry. The distal knob domain of the fiber binds to the initial receptor, CAR or CD46 while conserved RGD residues in a loop of the penton base protein bind to cell surface integrins which act as co-receptors. Engagement with integrins leads to fiber shedding, in part at the cell surface but more completely as the virus completes the clathrin-mediated endocytic pathway of entry. The role of endosomal acidification for adenovirus entry is equivocal with evidence both for and against and the final triggers of disassembly remain unclear but the process of membrane breaching is carried out by protein VI which becomes active at this time. Mirroring the simpler FHV example, protein VI is packaged into the assembling virus particle as a pre-protein attached to the hexon protein via its N-terminus. Cleavage of the pre-protein releases protein VI as part of the viral maturation process and, among other functions, it acts as the membrane lytic agent after receptor and integrin binding. It is a partially uncoated adenovirus that then enters the cytoplasm, engaging with the microtubule network to traverse the cell to the nuclear membrane where final disassembly occurs after docking with a nuclear pore.

For members of the *Reoviridae* too it is a core particle rather than the viral genome that enters the cytoplasm. Not, as in the example of adenovirus, to facilitate later association with other cellular components or organelles, but to prevent the dsRNA genome from triggering the innate immune response. Double-stranded RNA is one of the most powerful inducers of the host interferon response, being recognized by multiple sensor proteins including toll-like receptor (TLR) 3 in the endosome and protein kinase R, MDA-5 and RIG-I in the cytoplasm. Clearly, revelation of the dsRNA genome by any member of the *Reoviridae* on entry to the cell would curtail infection very quickly. Receptor mediated location of *Reoviridae* members to the cell surface by one or more of the receptors discussed leads to internalization, in most cases via the clathrin mediated entry pathway as suggested by inhibition of virus replication by clathrin pathway inhibitors, although macropinocytosis mediated entry has also been observed. For orthoreovirus, the virus outer capsid protein sigma 1 mediates cell attachment and uptake. The subsequent acidification of the endosome does not lead to immediate conformational change in the capsid but to activation of an acid dependent cysteine protease, which degrades another major outer capsid protein sigma 3, to reveal the hydrophobic protein mu 1, also a component of the outer capsid. Further degradation of mu 1 essentially removes the outer capsid completely and in doing so, generates mu 1 derived hydrophobic peptides, which then mediate endocytic escape of the now denuded core into the cytoplasm. In this case the mechanism of membrane breaching must be by pore formation as the core is still large, 50–60 nm, and passes intact into the

cytoplasm, the dsRNA genome still safely surrounded by core protein lambda 1. Rotaviruses use a somewhat similar pathway although human and animal rotaviruses differ in their uptake mechanisms, clathrin mediated or not, and the receptors they use. Within the endosome rotavirus VP4 protein, which with VP7 makes up the two outer capsid proteins of the triple layered virus, is cleaved to form VP5* and VP8* with VP5* acting as the membrane active component. Through the use of tracker dyes that respond to the immediate cellular environment significant detail of the process has been accrued. Following engulfment, the particle begins the morphological changes required to escape the endosome almost immediately and particles that are still resident when the endosome becomes acidified do not complete the process. Calcium concentrations vary markedly between the endosomal space and the cytoplasm and calcium ions are also present in the complete virus particle where they stabilize the VP7-VP7 interactions in the outer capsid. The concentration of these ions drops quickly in a virus loaded endosome as a result of VP5* mediated membrane perturbation and the lower levels lead to a destabilized VP7, loosening the entire outer capsid layer. The loss of both VP4 and VP7 produces the double layered particle, the DLP, which is the form of the virus that enters the cytoplasm, the whole process taking as little as 10 min to complete. Entry is accompanied by activation of virus transcription so the loss of the outer capsid must transmit a signal to the virion interior to boot-up the resident polymerase. Recent cryo-EM structures of the rotavirus polymerase in situ come close to explaining how this is achieved although some detail remains unclear.

In the case of Bluetongue virus, a multilayered arbovirus with a ten segment dsRNA genome, no clear receptor has emerged beyond the observations that lectins antagonize virus uptake by mammalian cells and that sialic acid binding sites occur in the outermost capsid protein, VP2. However, these occur in the body of the large trimeric protein, not in the distal tip which is presumed to contact the cell surface first so initial contacts is likely to be with other cell surface molecules. In addition, insect cells, which are also a permissive host for virus replication, do not generally encode sialyltransferases, thus sialic acid binding cannot be the means of the location to the insect cell surface. Here, integrins have been suspected to play a role and integrin binding motifs occur in VP2 and in the inner capsid proteins, VP7, but overall the details of engagement with the insect cell surface remain unclear. A role for metal ions, in this case zinc, in capsid stability has also been shown following the description of a zinc finger motif in the structure determined for VP2. Mutation of the conserved residues within the motif destabilizes the virus, consistent with a physiological role. Uptake, at least in mammalian cells, is via the clathrin pathway and VP2 is shed in the early endosome, plausibly aided by the loss of the chelated zinc ions as the lower pH protonates the ligand residues of the zinc finger (Fig. 3). As the endosome matures and the pH drops further the second outer capsid protein VP5 undergoes extensive conformational change to promote the transfer of the inner particle, the virus core, into the cytoplasm. The VP5 structure obtained by high-resolution cryo-EM of virus particles at neutral pH is broadly globular with extensive amphipathic alpha helices that include a central coiled-coil

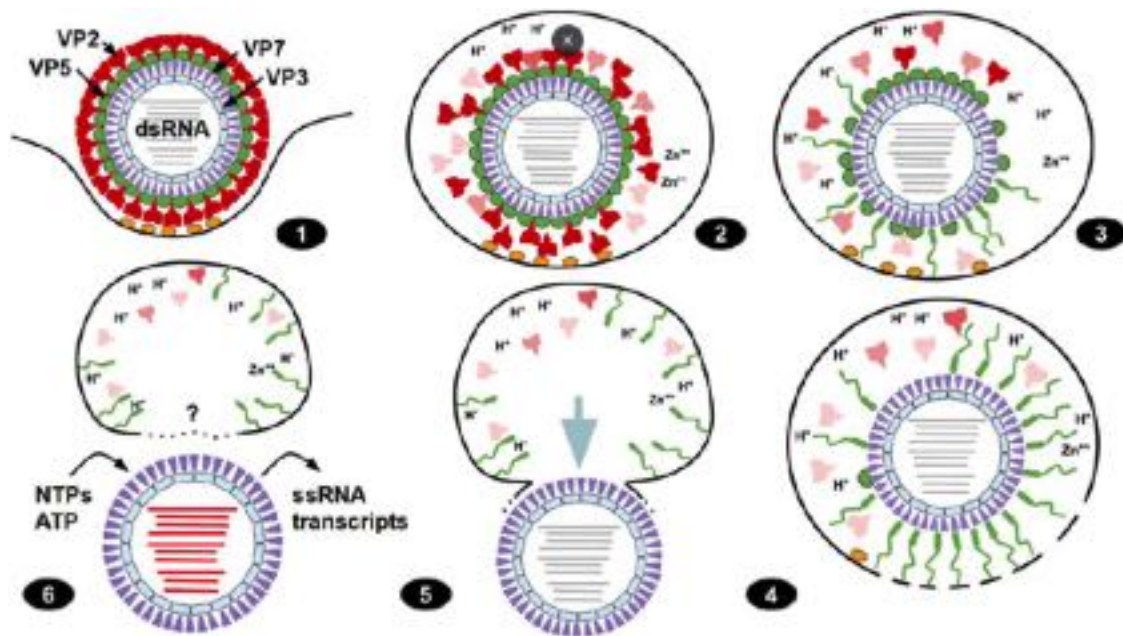


Fig. 3 Complex, layered virus entry as exemplified by the entry of Bluetongue virus. (1) The virus binds to unspecified receptors including sialic acid via its outermost capsid protein VP2 and the process of clathrin mediated uptake begins. (2) In the early endosome the pH drops and the VP2 protein loses structural integrity through the loss of chelated zinc ions, disengaging from the particle and possibly degrading. (3) Further acidification in the late endosome triggers a substantial conformational change in the VP5 protein which now binds the endosomal membrane. (4) Sufficient VP5 binding occurs to trigger membrane disruption and the beginnings of a pore in the vicinity of the virus particle. (5) The pore enlarges and the virus core passes through by an as yet undefined mechanism. The VP5 conformational change is irreversible and it is shed during the passage of the core. (6) The core is released to the cytoplasm where it immediately becomes transcriptionally active, triggered by a minor change in the inner capsid protein VP3. The fate of the residual endosome is unclear but the pore could close and the endosome continue its progress to the lysosome where any remaining virus proteins would be degraded.

alpha-helical bundle, reminiscent of the structures found in the fusion proteins of many enveloped viruses. The structure of the virus at low pH however fails to identify VP5 beyond a stub which remains in contact with the core, indicating that the majority of the VP5 chain is extended and flexible, consistent with extensive interaction with the endosomal membrane. Indeed, translocation of VP5 to the cell surface by an expression vector followed by an acid pulse leads to syncytium formation formally demonstrating its membrane fusion potential. As for rotavirus, recent high resolution cryo-EM structures of the virus core have located the virion polymerase, VP1, which similarly becomes active once the outer capsid has detached and the core particle is in the cytoplasm. In part this is the result of unblocking of the pores at the vertices so that solutes can enter and transcripts leave but the structure also suggests a more active signal in that a distinct contact between the N-terminus of the inner core protein VP3 and the polymerase acts to stimulate the resting structure into action (Fig. 3). This active cross talk between the final stages of uncoating and the initiation of virus replication confirms how important the entry pathway is. It is not enough just to enter the cell, that entry reaction must follow certain defined steps if the infection is to become a reality.

Exploitation of the Virus Entry Pathway

As the description of cell entry for a range of viruses has indicated, the mature virion, that is a virion that is capable of undergoing the entry reaction, is metastable. Given the right cues the virus will disassemble and there is very little wriggle room between the two extreme states, that is, primed for entry and entered. This presents a known difficulty in some virus applications, notably vaccine preparation and storage. Here the long-term storage of virus can lead to premature disassembly as the dynamic particle breathing that almost all viruses undergo can eventually tip the balance in favor of full activation and lead to an irreversible loss of infectivity or immunogenicity. In poliovirus for example, the preparation of vaccine involves a measure of the H (heated) and N (neutralizing) antigenic forms with only those preparations with high N levels deemed acceptable for vaccine use. The H form, which can be fully induced by heating the virus, is a slightly expanded particle, which has initiated the early stages of the entry reaction and, in doing so, abolished the capsid features that are necessary to induce neutralizing antibodies. As noted, the maturation reaction for many viruses, often a protein cleavage which primes the virus for entry, is a very late event in the replication pathway and usually accompanies genome packaging. This means that recombinant routes to vaccine production, by the formation of virus-like-particles for example, run the risk of failing to mimic the final maturation event as no genome is ever packaged. As a result of this concern a form of recombinant vaccine, safe and manufacturable outside of containment facilities, may not be as immunogenic on a weight for weight basis as the authentic virus. Recombinant approaches to the production of other picornavirus vaccines, Foot and Mouth Disease Virus for example, face similar challenges and active capsid engineering to increase the stability of the empty capsids is required to ensure that the recombinant product is fully immunogenic. Understanding the details of the entry reaction can therefore have important consequences in a number of areas of applied virology (Fig. 4). Beyond vaccine development, a number of other areas of application similarly depend on a detailed understanding of the entry reaction. Several non-enveloped viruses are in active development as oncolytic agents and as mediators of gene transfer. In these cases the recombinant virus, acting purely as a vector in the latter case, must enter target cells efficiently to fulfill its purpose. As noted, parvoviruses are often destroyed on entry as they fail to escape the endosome before fusion with the lysosome meaning that their genetic cargo never reaches the nucleus to enable transcription and the production of a therapeutic effect. In consequence, very large doses of gene delivery vectors such as AAV are needed to ensure the delivery of a sufficient genetic payload for therapy and, as a result, the yields of a number of gene therapy vectors represent a current bottleneck, for which improvements in the cell entry pathway could, in part, address.

A final area of application where study of the entry mechanism of non-enveloped viruses may benefit application is in the development of antiviral drugs. As described, the entry pathway is, in essence, a series of programmed virus specific changes

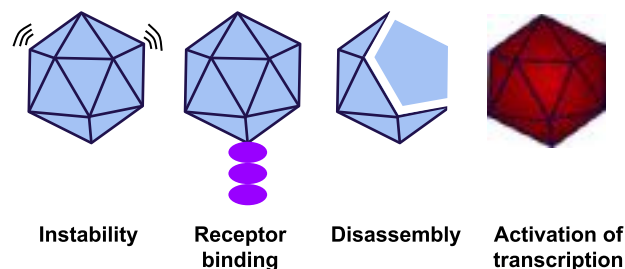


Fig. 4 Points at which the cell entry pathways of viruses could be open to exploitation. The metastable state of the virus can be modified by recombinant capsid engineering to improve vaccine stability. Receptor binding can be modified to repurpose viruses as gene delivery agents with altered tissue tropism. Disassembly inhibitors are already a reality and more could be developed, locking the virus in a form that cannot progress to infection. As revealed by recent structural studies the activation of a complex virus core depends on the completion of the correct entry process leading to the possibility that this too could be a target for inhibition once detailed knowledge of the activation steps have been obtained.

engendered at the appropriate time by contact with host factors. It follows that blocking any one of these obligatory changes would block virus infection. In the picornaviruses precisely such a block has been demonstrated by the isolation of compounds that prevent the stages of the entry reaction that follow initial receptor binding. In some, but not all, picornaviruses one of the three proteins that make up the final capsid shell, VP1, contains a hydrophobic cavity that is naturally occupied by a long chain fatty acid molecule picked up from the cell during assembly. These “pocket” factors provide a stuffer function which prevents the cavity from collapsing and lie immediately underneath the VP1 residues that contact the receptor. The ejection of such factors follows engagement with the cell receptor and initiates the disassembly reaction. A series of synthetic compounds have been isolated that bind the pocket far more tightly than their natural counterparts, such that they cannot be jettisoned when the virus engages with the receptor. As a result the entry reaction is aborted and the drugs concerned exhibit powerful antiviral activity. To a greater or lesser extent all picornaviruses which make use of pocket factors can be inhibited by such compounds, which have broad antiviral properties for this family of viruses. The exemplar illustrated by pocket factor mimic compounds demonstrates that the entry pathway of non-enveloped viruses is a *bona fide* target for antiviral drug development and that it offers the possibility of broadly reactive compounds, one of the most sought after goals in the field of antivirals.

Concluding Remarks

The diversity of the global virome might suggest a myriad of entry pathways to allow infection of any cell of any organism in any phylum. Perhaps surprisingly, this is not the case and while the precise details of entry naturally vary depending on the virus, the principles of the process are shared. Recently, transient retention of a membrane was demonstrated for some non-enveloped viruses, Hepatitis A virus serving as the exemplar, but these membranes are not impregnated with any virus protein so cannot act in membrane fusion. Rather they appear to be a cloaking mechanism to hold off immune recognition for as long as possible, enabling the virus to disseminate within the host more efficiently. Thus, for the non-enveloped viruses the possibility of simply merging membranes together to enter a cell is ruled out and particular mechanisms must have evolved to deal with crossing what is otherwise a controlled and impermeable barrier. Immobilization on the cell surface is a necessary first requirement that is met by engagement with a number of molecules that act as virus receptors. A range of molecular types serve this purpose with a loose correlation between the selectivity of binding and virus tropism, high affinity protein interaction being used to gain access to a particular tissue while common cell surface components tend towards a distributed infection. In some cases the binding to the receptor may trigger virus conformational change but more commonly this requires entry into, and detection of the environment within, an endosome. The key triggers, depending on the endocytic route taken, are a lowering of pH and or the activation of proteolysis. The type of lipid within the endosomal membrane may also serve as a cue to alert the virus to a suitable environment for disassembly. Disassembly invariably results in the exposure of a more hydrophobic protein or peptide that is lipophilic, a process that is irreversible. The precise mechanism of membrane disruption that these proteins enable remains unclear in many cases. For some, such as the picornaviruses, an intermediate of the capsid attached to the endosomal membrane and an asymmetrical release of the genome through a portal is relatively well defined but in other cases a residual virion structure, often close to the size of the original virus, must traverse the membrane intact. How the necessary pore is formed and whether it closes behind, to maintain immunological elusiveness perhaps, is not known. The complexity of biological membranes is only partly reproduced by the use of *in vitro* systems, limiting the experimental approaches that can be applied but the increasing use of newer technologies, such as cryo-EM tomography, which is capable of visualizing the act of fusion, will surely shed light on these intricate and fast interactions. Many details of entry are known and as the pathways and triggers become better described so the opportunity to exploit them will increase. For the virus, the potential for replication lies at the end of a successful entry reaction, for the observer the opportunity to prevent it. It is a curious fact that the start of something for one could spell an opportunity for its control by another.

Further Reading

- Assoy, P., Gottschalk, E.Y., Meneses, P.I., 2017. HPV entry into cells. *Mutation* 772, 13–22.
- Blaum, B.S., Stehle, T., 2019. Sialic acids in nonenveloped virus infections. *Advances in Carbohydrate Chemistry and Biochemistry* 76, 65–111.
- Bochkov, Y.A., Watters, K., Ashraf, S., *et al.*, 2015. Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proceedings of the National Academy of Sciences of the United States of America* 112 (17), 5485–5490.
- Callaway, H.M., Feng, K.H., Lee, D.W., *et al.*, 2017. Parvovirus capsid structures required for infection: Mutations controlling receptor recognition and protease cleavages. *Journal of Virology* 91 (2).
- Conley, M.J., McElwee, M., Azmi, L., *et al.*, 2019. Calicivirus VP2 forms a portal-like assembly following receptor engagement. *Nature* 565 (7739), 377–381.
- Ding, K., Celma, C.C., Zhang, X., *et al.*, 2019. In situ structures of viral polymerase in action and mechanisms of mRNA transcription and release. *Nature Communications* 10, 2216. doi:10.1038/s41467-019-10236-7.
- Du, J., Bhattacharya, B., Ward, T.H., Roy, P., 2014. Trafficking of bluetongue virus visualized by recovery of tetracycline-tagged virion particles. *Journal of Virology* 88 (21), 12656–12668.
- He, Y., Shivakoti, S., Ding, K., *et al.*, 2019. In situ structures of RNA-dependent RNA polymerase inside bluetongue virus before and after uncoating. *PNAS* 116 (33), 16535–16540.
- Jenni, S., Salgado, E.N., Herrmann, T., *et al.*, 2019. In situ structure of rotavirus VP1 RNA-dependent RNA polymerase. *Journal of Molecular Biology* 431 (17), 3124–3138.
- Maginnis, M.S., 2018. Virus-receptor interactions: The key to cellular invasion. *Journal of Molecular Biology* 430 (17), 2590–2611.

- Nemerow, G.R., Stewart, P.L., 2016. Insights into adenovirus uncoating from interactions with integrins and mediators of host immunity. *Viruses* 8 (12).
- Palmenberg, A.C., 2017. Rhinovirus C, asthma, and cell surface expression of virus receptor CDHR3. *Journal of Virology* 91 (7).
- Rivera-Serrano, E.E., Gonzalez-Lopez, O., Das, A., Lemon, S.M., 2019. Cellular entry and uncoating of naked and quasi-enveloped human hepatoviruses. *eLife* 8.
- Ros, C., Bayat, N., Wolfisberg, R., Almendral, J.M., 2017. Protoparvovirus cell entry. *Viruses* 9 (11).
- Snyder, A.J., Danthi, P., 2019. Selection and characterization of a reovirus mutant with increased thermostability. *Journal of Virology* 93 (9).
- Wu, W., Celma, C.C., Kerviel, A., Roy, P., 2019. Mapping the pH sensors critical for host cell entry by a complex nonenveloped virus. *Journal of Virology* 93 (4).
- Zhang, X., Patel, A., Celma, C.C., *et al.*, 2016. Atomic model of a nonenveloped virus reveals pH sensors for a coordinated process of cell entry. *Nature Structural & Molecular Biology* 23 (1), 74–80.

Enveloped Virus Membrane Fusion

Aurélien A Albertini and Yves Gaudin, Institute for Integrative Biology of the Cell (I2BC), French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, Paris-Sud University, University of Paris-Saclay, Gif-sur-Yvette, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ASLV Avian sarcoma-leukosis virus
BMP Acid bis(monoacylglycero)phosphate
BVDV Bovine viral diarrhea virus
DENV Dengue virus
EBOV Ebola virus
EFF1 Epithelial fusion failure 1
FD Fusion domain
HA Hemagglutinin
HCV Hepatitis C virus
HIV Human immunodeficiency virus
HR Heptad repeat
HRPV Haloarchaeal pleomorphic viruses

HSV-1 Herpes simplex virus 1
LBPA Lysobisphosphatidic
MERS-CoV Middle East respiratory syndrome coronavirus
PHD Pleckstrin homology domain
RVFV Rift valley fever virus
SARS-CoV Severe acute respiratory syndrome coronavirus
SFV Semliki forest virus
TBEV Tick-borne encephalitis virus
TGN Trans-Golgi network
TM Trans membrane
trD Trimerization domain
VSV Vesicular stomatitis virus

Glossary

Enveloped viruses Viruses that are surrounded by a lipid bilayer derived from host cell membranes generally acquired at the budding stage.

Glycoprotein Proteins which contain oligosaccharide chains covalently attached to amino acid side-chains. In general, fusion glycoproteins are N-glycosylated on asparagine residues.

Endocytosis A cellular process in which material is internalized by the cell. The cargo to be internalized is progressively wrapped by a cellular membrane which then buds off inside the cell forming a vesicle containing the material.

Fusion peptide or loop Short hydrophobic motifs in a viral fusion glycoprotein which interacts with the cellular membrane to initiate the fusion process. When they are amino terminal, they are referred as a fusion peptide whereas when they are internal, they are referred as a fusion loop.

Hemifusion An intermediate stage during membrane fusion that is characterized by the merger of the outer monolayers but not the two distal (inner) monolayers.

Fusion pore Small opening that allows flux between two membrane bound compartments. Fusion pores form at a stage of membrane fusion that follows hemifusion and finally widen, leading to full fusion.

Prefusion conformation Conformation of the fusion glycoprotein at the viral surface after priming but before fusion triggering.

Postfusion conformation Conformation of the fusion glycoprotein at the end of the fusion reaction.

Priming Proteolytic cleavage in the fusion glycoprotein precursor or in an accompanying protein which brings the fusion glycoprotein into a state capable of responding to a fusion trigger.

Triggering Triggering induces the structural transition from the prefusion to the postfusion conformation. Usual triggers are receptor interaction and low-pH environment.

Introduction

Enveloped viruses are surrounded by a membrane made of a lipidic bilayer. They are encountered in living organisms of the three domains of life. In most cases, their envelope is acquired at the end of the cellular viral cycle at the budding stage and is derived from the host cell membrane. It contains phospholipids and cholesterol, and includes virally encoded transmembrane proteins. In general, cellular membrane proteins are largely excluded from the viral envelope.

Entry of enveloped viruses into host cells requires binding of the virus to one or more receptors present at the cell surface, followed by fusion of the viral envelope with a cellular membrane which allows the release of the inner content of the virion, including its genome, into the host cell cytoplasm for the subsequent steps of infection.

Membrane fusion, being key in several biological processes, has been well studied and, in all the systems that have been characterized, it has been shown that the fusion of two membranes facing each other proceeds via local dehydration and formation of an initial stalk that is a local lipidic connection between the outer leaflets of the fusing membranes. Radial expansion of the stalk induces the formation of a transient hemifusion diaphragm (i.e., a local bilayer made by the two initial inner leaflets). The next step is the formation of a pore in the fusion diaphragm. The initial pore is small and may open and close repeatedly

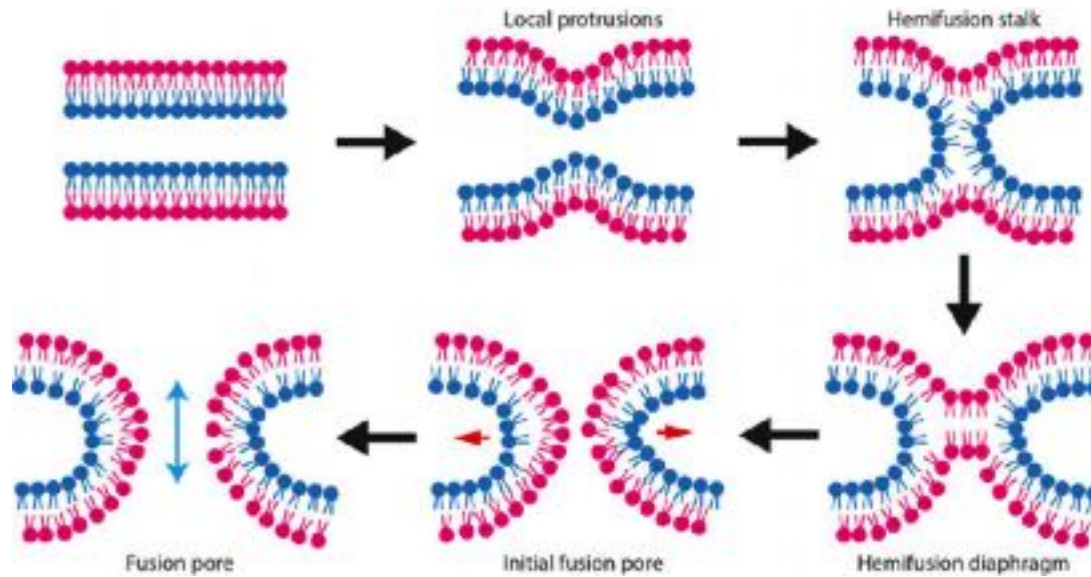


Fig. 1 Stages of membrane fusion of two lipid bilayers according to the stalk-pore model.

before its enlargement that leads to complete fusion (**Fig. 1**). Although membrane fusion can occur spontaneously, there are energetic barriers that have to be overcome at different steps of the process, the highest one being encountered during the expansion of the initial fusion pore. Therefore, in living organisms, there are fusion machineries which catalyze membrane fusion so that it occurs at the right place and time.

In the case of enveloped viruses, this machinery is constituted by virally encoded transmembrane proteins. The best characterized viral fusion machineries are the fusion glycoproteins of eukaryotic viruses which are transmembrane type I glycoproteins. After cleavage of their amino terminal signal peptide, the mature fusion glycoproteins have an N-glycosylated aminoterminal ectodomain anchored in the membrane by a single α -helical transmembrane segment. The C-terminal intraviral domain is of variable size and may interact with internal proteins of the virion and/or regulate the fusion activity.

Upon interaction with specific triggers such as viral receptors or a low-pH environment, viral fusion glycoproteins undergo a huge conformational change from their prefusion (or native) toward their postfusion (or final) conformation. During the structural transition, they expose hydrophobic motifs (the so-called fusion peptide or fusion loops) that interact with one or both of the participating membranes, resulting in their destabilization and initiation of the fusion reaction. When the hydrophobic motif is amino terminal, it is referred as a fusion peptide whereas when it is internal, it is referred as a fusion loop.

Based on their structure and common structural motifs, viral fusion proteins have been classified into three distinct classes.

General Considerations

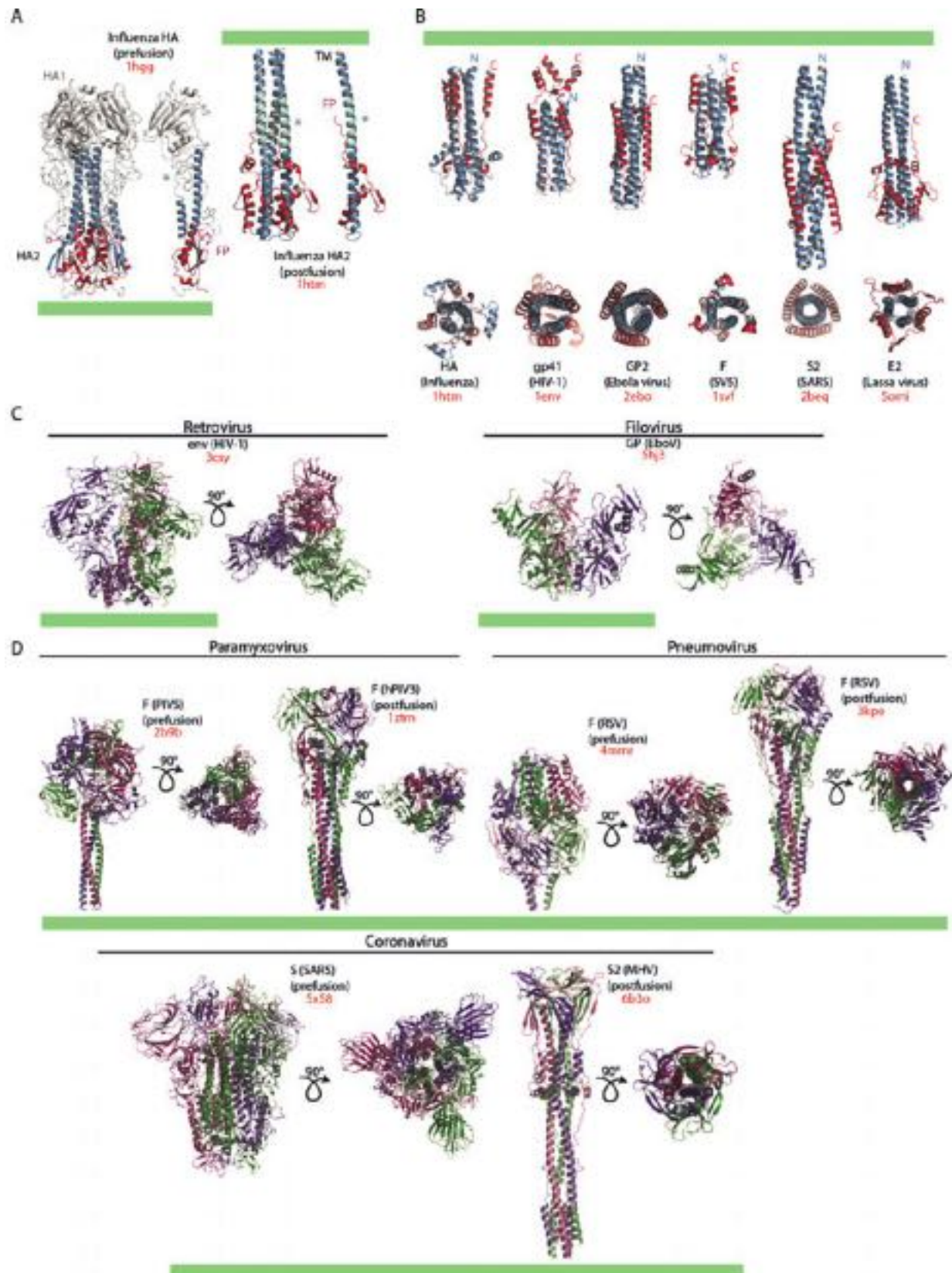
Viral Fusion Site in the Cell

For eukaryotic viruses, fusion can take place either at the plasma membrane or after endocytosis of the viral particle in the endosome.

Fusion of several viruses belonging to the Herpesviridae, Retroviridae, Paramyxoviridae family and some members of the Coronaviridae occurs at the plasma membrane. For those viruses, the fusogenic conformational change is triggered by receptor recognition. For Retroviridae and Coronaviridae, the fusion glycoprotein is also involved in receptor binding whereas, for Paramyxoviridae and Herpesviridae, another viral glycoprotein promotes receptor recognition.

For other viruses, fusion occurs after particle endocytosis. In the endosome, fusion is often triggered by the acidification of the lumen of the vesicle. Indeed, low pH alone is able to trigger fusion of rhabdoviruses (such as rabies virus or vesicular stomatitis virus-VSV-), orthomyxoviruses (such as influenza virus), bunyaviruses (such as Ukuniemi virus), alphaviruses (such as Semliki forest virus-SFV-), flaviviruses (such as Dengue virus-DENV- and tick borne encephalitis virus-TBEV-) and baculoviruses. Depending on the pH threshold required for triggering the fusion glycoprotein conformational change, fusion occurs in the early endosomes (pH $\sim$ 6.5) or in late endosomes (whose pH can be as low as 5). As in other pH-sensing proteins, histidine residues often play a key role as their side chain pKa is around 6.5 and their protonation destabilizes the prefusion conformation.

For some other endocytosed viruses, such as avian sarcoma leucosis virus (an alpharetrovirus), Ebola virus (EBOV, a filovirus) or Lassa virus (an arenavirus), fusion activation requires both receptor binding and low-pH. It is worth noting that, in some cases, the cellular protein triggering the conformational change of the fusion glycoprotein is only located in the endosome and, therefore,



is different from the receptor leading to virion endocytosis. Finally, for viruses including severe acute respiratory syndrome coronaviruses 1 and 2 (SARS-CoV-1 and SARS-CoV-2), Middle East respiratory syndrome coronavirus (MERS-CoV) and respiratory syncytial virus, low pH is not required for fusion but endocytosis is necessary for the activation of the fusion glycoprotein by endosomal proteases (such as furin or cathepsin).

Finally, the dichotomy between viruses fusing at the plasma membrane and those fusing in the endosome is not that strict inside a viral family (e.g., Retroviruses and coronaviruses) and even for a single virus the place of fusion may depend on the cell type (e.g., for some herpesviruses).

Specific Lipid Requirements

In vitro fusion systems between viral particles and artificial membranes of defined composition have revealed the absolute requirement of some lipids for some viruses. The best example is SFV for which both cholesterol and sphingomyelin have to be present in the target membrane for fusion to occur.

In several other cases, although not an absolute requirement, presence of some lipids enhances the fusion reaction. This is the case for VSV, DENV and Uukuniemi virus for which fusion is enhanced by presence of anionic lipids such as lysobisphosphatidic acid (LBPA) or bis(monoacylglycerol)phosphate (BMP) that are both late endosome specific phospholipids.

Structure of Viral Fusion Glycoproteins

Class I fusion glycoproteins

Influenza HA, the prototype of class I fusion glycoproteins

The first crystal structure of a viral fusion glycoprotein ectodomain to be determined was that of the influenza virus hemagglutinin (HA), isolated after limited bromelain digestion from purified virions. HA is synthesized as HA0, a precursor molecule cleaved by a cellular protease to produce HA1, a sialic acid binding domain involved in receptor recognition, and HA2, a membrane-anchored fusion domain. HA0 cleavage primes the protein for the low pH-induced fusogenic conformational change.

In its prefusion conformation, HA is a homotrimer, each protomer consisting of an HA1 and an HA2 chain connected through a single disulfide bridge. In the prefusion structure, most of HA2 forms an elongated trimeric coiled coil to the top of which HA1 is tethered. The amino-terminal extremity of HA2, which has been generated by cleavage of HA0, is highly hydrophobic and constitutes the fusion peptide which is buried at the trimeric interface (Fig. 2(A)).

After low-pH exposure, HA undergoes a huge structural transition which results in the loss of HA1–HA2 interactions and dissociation of the HA1–HA1 interface which stabilizes the prefusion structure. Globally, HA low-pH-induced conformational changes result in a dramatic refolding of HA2 while HA1 keeps its tertiary structure. During the refolding, a long loop in native HA2 converts to a helix, increasing the length of the central coiled coil at its N-terminal end. This movement translocates the fusion peptide toward the head of the molecule, where is located the target host membrane. The structural transition is completed by a relocation of the carboxyterminal membrane anchor through a foldback mechanism to the amino-terminal end of the rod-shaped molecule. This results in the so-called hairpin conformation in which the transmembrane segment and the fusion peptide are in close proximity at the same end of the elongated molecule, as expected at the end of the fusion process when the viral and endosomal membranes have merged (Fig. 2(A)).

Class I common characteristics

Several fusion glycoproteins have similar features to those of HA. Together, they define the characteristics of class I fusion glycoproteins. Besides HA, this class includes the fusion glycoproteins of Paramyxoviridae (e.g., parainfluenza and measles viruses), Pneumoviridae (e.g., respiratory syncytial virus), Retroviridae (e.g., HIV, Avian sarcoma leukemia virus-ASLV-, Murine leukemia virus), Filoviridae (e.g., EBOV), Arenaviridae (e.g., Lassa virus) and Coronaviridae (e.g., SARS-CoV, MERS-CoV), these viruses are usually not icosahedral.

Class I fusion glycoproteins are trimeric in both their pre- and post-fusion conformation. They have an internal, frequently furin-like, cleavage site. The cleavage is required for priming them to the prefusion conformation that will be capable of undergoing the fusogenic

Fig. 2 Class I fusion glycoproteins. (A) Influenza hemagglutinin (HA) structures. Left part: Ribbon representation of HA trimer and HA2 protomer in the prefusion conformation. The sialic acid binding domain (HA1) subunits are colored in gray and the HA2 fusion proteins are colored in blue and red. The amino-terminal extremity of HA2 constitutes the hydrophobic fusion peptide (FP, colored in magenta) which is buried at the trimeric interface in this conformation. Right part: Structure of HA2 trimer and protomer in postfusion conformation. The positions of the TM domain and of the fusion peptide (FP) are indicated on the postfusion protomer. The HA2 segment that refolds into an helix is colored in cyan in both conformations. (B) Trimeric hairpins structures of several class I viral fusion proteins. Top and side views of the fusion proteins from 6 different viruses are shown: The N-terminal coiled-coil core is colored in blue and the C-terminal domain reaching the transmembrane domain is in red. In all those structures, the trimer-of-hairpins conformation brings the N- and C-termini (corresponding to the position of the fusion peptides and transmembrane domain respectively) into close proximity. (C) Ribbon representation of the prefusion structures of HIV-1 env and EBOV GP fusion proteins (side and top views, each protomer is depicted in a different color). (D) Ribbon representation of Paramyxovirus, Pneumovirus and Coronavirus viral fusion proteins structures in their pre and postfusion conformation. For each structure, a side and a top view are presented and each protomer is in a different color. All those structures have an extended common core indicating that they evolved from a common ancestor. On each panel, the position of the membrane with respect to the side view of the fusion glycoprotein is indicated by a green line. For each structure, its PDB code is indicated in red.

structural transition in response to a specific fusion trigger. The transmembrane C-terminal fragment resulting from the cleavage is the fusion subunit (corresponding to influenza HA2) which bears at or near its N-terminal end a stretch of about 15 hydrophobic amino acids which constitutes the membrane interacting motif.

In the prefusion state, the N-terminal and C-terminal fusion subunits remain held together by either a disulfide bond (as in influenza HA) or non-covalent interactions (e.g., in HIV Env). The fusion peptide is buried at a protein-protein interface.

In the postfusion state, the fusion subunit is refolded into a trimeric coiled coil at the N-terminal end of which are located the three fusion peptides, now solvent exposed, and in the grooves of which are packed, in an antiparallel manner, the segments abutting the transmembrane region. In this conformation, the protomer shape is thus an elongated hairpin-like structure with the fusion peptide and the transmembrane domain located at the same end (**Fig. 2(B)**).

A class I signature is the presence of heptad repeats (HRs) in the amino-acid sequence of the glycoprotein. HRs are made of a repeating pattern of seven amino acids *abcdefg* where *a* and *d* are hydrophobic residues. A first HR region (named HRA or HR1) is located immediately downstream of the fusion peptide and forms the central trimeric coiled coil. It is often accompanied by a second (named HRB or HR2) located in the C-terminal part of the ectodomain that constitutes the helices positioned in the groove of the central coiled coil.

Except for this common structural motif, the structure of class I glycoproteins are quite different and it is not clear whether these proteins have evolved from a distant common ancestor (**Fig. 2(C, D)**). However, the structures of the fusion glycoproteins of Coronaviridae, Pneumoviridae and Paramyxoviridae share an extended common core implying that they have an evolutionary connection (**Fig. 2(D)**).

Class II Fusion Glycoproteins

TBEV E, the prototype of class II fusion glycoproteins

Glycoprotein E from TBEV, a member of the Flavivirus genus in the Flaviviridae family, was the first class II fusion glycoprotein of which a crystal structure was been solved.

The E protein folds co-translationally with a companion protein called prM. The heterodimeric association between E and prM is absolutely required for correct folding and transport of E. Immature virions are formed by budding of the genome-containing capsid into the lumen of the endoplasmic reticulum. prM and E are organized into an icosahedral lattice at the viral surface. Virion maturation involves a dramatic rearrangement of their organization at the particle surface. Finally, proteolytic cleavage of prM in the trans-Golgi network by a furin-like protease primes E for low pH-induced fusogenic conformational change and is therefore required for virion infectivity.

E displays a molecular architecture completely different from that of HA and other class I fusion glycoproteins (**Fig. 3(A)**). In its prefusion conformation, it forms an elongated antiparallel dimer, lying flat at the viral surface, and organized in an icosahedral lattice. The polypeptide chain of a protomer folds into three distinct domains. Domain I is a central β -barrel. Domain II is the fusion domain, it is elongated and also implicated in E dimerization. Domain III is located in the carboxyterminal part of the ectodomain and has an immunoglobulin-like fold. The internal fusion loop is located between two β -strands at the tip of the fusion domain and buried at the dimeric interface. Upon exposure to low pH, dimers dissociate and the protomers reassociate in a trimeric structure (**Fig. 3(A)**). The conformational change of a protomer is less impressive than that observed for class I viral fusion glycoproteins as there is little change in the secondary and tertiary structures of the three domains. Nevertheless, the interactions between the domains constituting a protomer are markedly different in the pre- and postfusion conformations. In particular, domain III, which is connected to the C-terminal part of the molecule, moves by about 35 Å toward the fusion loop. This movement redirects the polypeptide chain, so that the transmembrane segment is located near the fusion loop region at the same end of an elongated trimer that is now perpendicular to the membrane. Therefore, despite the different architecture of the proteins, the postfusion structure of E is reminiscent of the postfusion hairpin structure of class I fusion proteins.

Class II common characteristics

Other flaviviruses E glycoproteins (including DENV, West Nile, Japanese encephalitis and Zika viruses E) were subsequently shown to adopt a very similar structure. E1 glycoproteins of alphaviruses (including SFV, Sindbis and Chikungunya viruses) (**Fig. 3(B)**) and of rubella virus (previously considered together with alphaviruses as a member of the togaviridae family) (**Fig. 3(C)**) also have the same fold despite the absence of sequence similarity. This revealed that E and E1 derived from a common ancestor. The simplest explanation for this homology is that flaviviruses and alphaviruses, having both positive-stranded RNA genomes with very similar organization, share an ancient common ancestor.

More unexpectedly, the fusion glycoprotein Gc of some viruses belonging to the unrelated bunyavirales order (including Rift valley fever virus, RVFV, from the Phenuiviridae family and Hantaan and Puumala viruses from the Hantaviridae family), were also shown to be class II fusion glycoproteins (**Fig. 3(D)**). The prefusion structure of RVFV Gc is remarkably similar to that of flaviviruses E glycoprotein as it also forms head-to-tail dimers (**Fig. 3(D)**). This extreme similarity is not found in the spike architecture of the Peribunyaviridae (such as La Crosse, Schmallenberg and Bunyamwera viruses) which also belong to the bunyavirales order. Indeed, in addition to the class II fold, Peribunyaviridae Gc glycoproteins have two additional aminoterminal domains that are involved in the trimerization of the prefusion conformation.

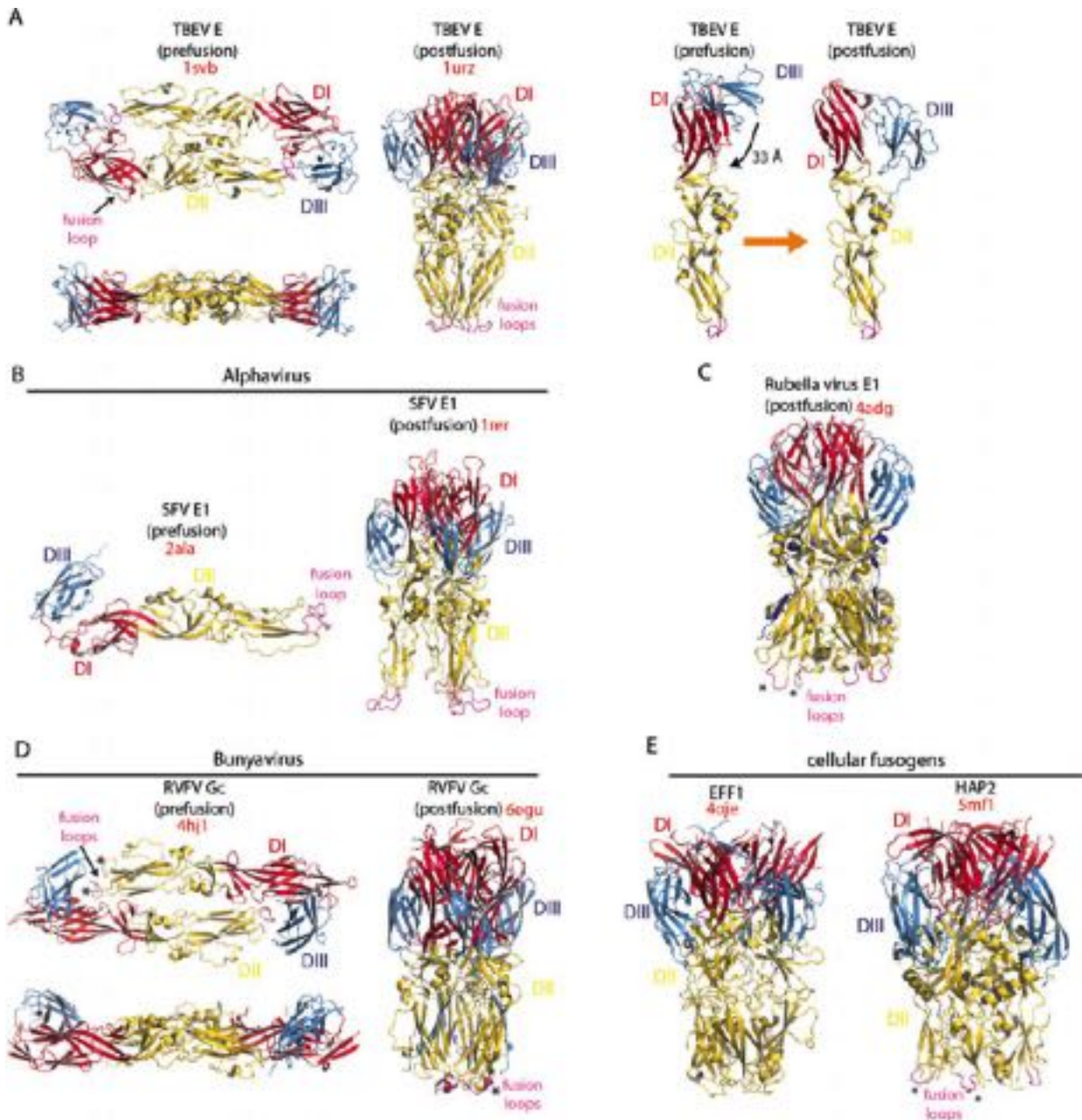


Fig. 3 Class II fusion glycoproteins. (A) Ribbon representations of the dimeric prefusion conformation of TBEV E (left part) and of its trimeric postfusion conformation (mid. part). Right part: TBEV E protomer conformational change. The arrow indicates DIII movement toward the fusion loop in DII. (B) Ribbon representation of SFV E1 in its pre and postfusion conformations (C) Ribbon representation of rubella virus E1 in its trimeric postfusion conformation. Note that rubella virus E1 has two fusion loops (indicated by asterisks). (D) Ribbon representations of RVFV Gc head-to-tail prefusion dimer and of its trimeric postfusion conformation (right part). The two fusion loops are indicated by asterisks. The prefusion organization of RVFV Gc is remarkably similar to that of flaviviruses E glycoprotein. (E) Ribbon representations of EFF1 protein from *C. elegans* (left part) and structure of HAP2 from *Chlamydomonas reinhardtii* gamete fusion protein (right part) in their postfusion states. Note the absence of fusion loop at the tip of DII on EFF1. All class II glycoproteins have the same fold, made of a central domain DI (in red), a fusion domain DII (in yellow) and C-terminal domain DIII (in blue), which reveals that they derived from a common ancestor. Fusion loops, located at the tip of DII, are in pink. For each structure, its PDB code is indicated in red.

All class II viral fusion glycoproteins identified so far have the same structural organization into three domains as TBEV E. The target membrane interacting region is located at the tip of the fusion domain (domain II). It is in general monopartite (bearing a single fusion loop) but sometimes bipartite (bearing two fusion loops) as in the case of rubella virus E1 and RVFV Gc (Fig. 3(C and D)) and even tripartite in the case of hantavirus Gc. In the prefusion conformation, fusion loops are in general buried at a dimer interface.

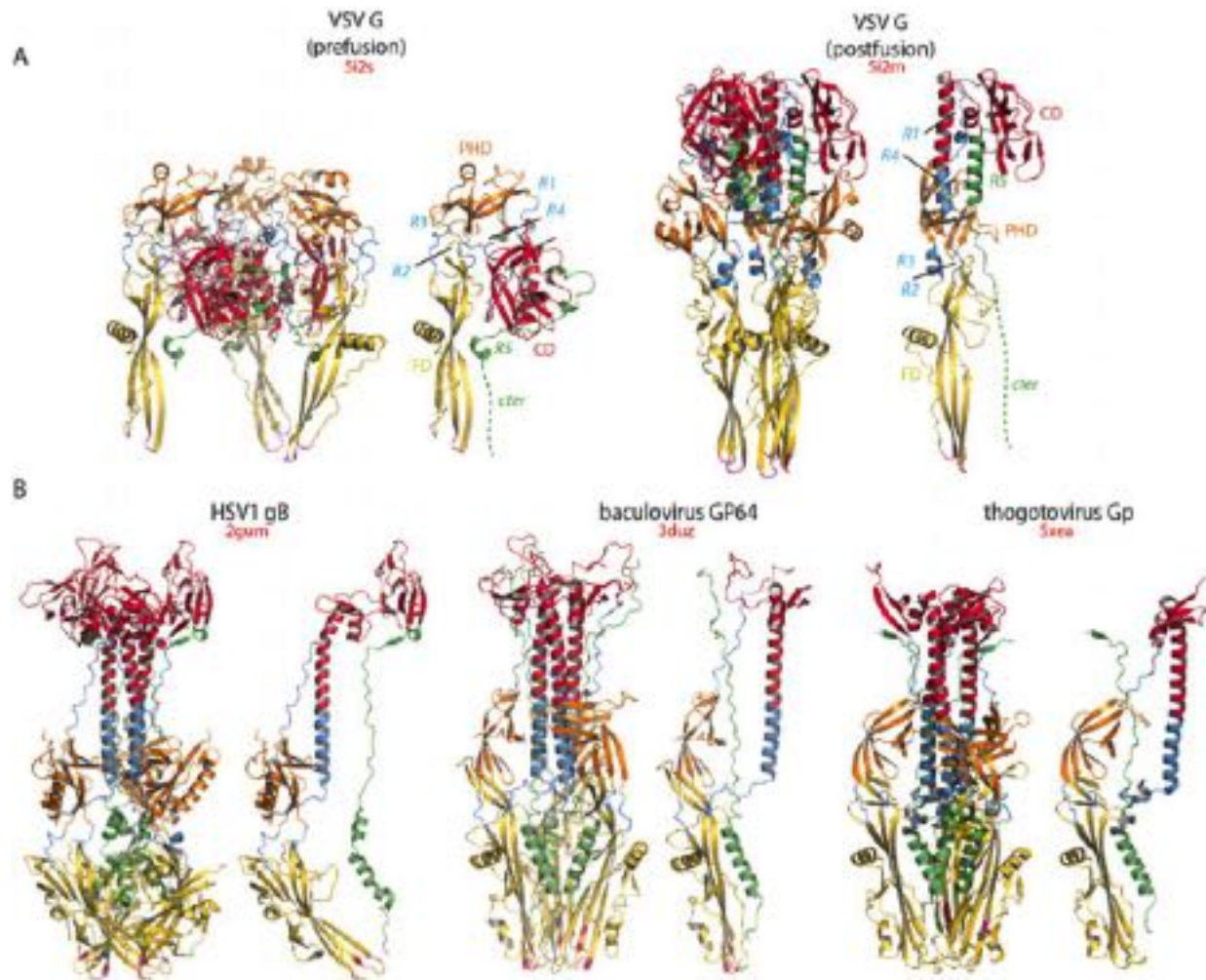


Fig. 4 Class III fusion glycoproteins. (A) Ribbon representations of VSV G prefusion trimer and protomer (left part) and of VSV G postfusion trimer and protomer (right part). (B) Ribbon representation of HSV-1 gB, baculovirus GP64 and thogotovirus Gp in their post-fusion conformation (trimer and protomer). Color code: central domain (CD) is in red, pleckstrin homology domain (PHD) is in orange, fusion domain (FD) is in yellow, refolding segments R1-R4 are in cyan, C-terminal segment reaching the transmembrane anchor (R5) is in green. The position and length of segments R1-R5 for gB and GP64 and Gp are only postulated as the extent of the conformational change remains unknown. For each structure, its PDB code is indicated in red.

Class II viral fusion glycoproteins are synthesized and folded as a complex with a second viral envelope protein (prM for flaviviruses, E2 for alphaviruses, Gn for bunyaviruses) that plays a chaperone role. Proteolytic cleavage of this companion protein primes the fusion glycoprotein so that it is capable of undergoing the low-pH-induced fusogenic structural transition. In their native conformation, they form homo- (flaviviruses and phenuiviridae) or hetero- (alphaviruses) dimers that lie flat or nearly flat at the viral surface but also sometimes trimers (Peribunyaviridae) having a tripodal organization.

In most cases, in their prefusion state, class II fusion glycoproteins are organized in an icosahedral assembly at the viral surface. However, no icosahedral symmetry is detected on rubella virus particles although it cannot be excluded that the symmetry is disrupted during purification and manipulation of the virions. Icosahedral symmetry is also highly improbable for hantaviruses: at the surface of the virion, ordered patches of square-shaped assemblies with 4-fold symmetry are detected which are not compatible with icosahedral symmetry.

Finally, upon low pH exposure, class II prefusion oligomers dissociate and the fusion glycoprotein subunits reassociate in a trimeric structure. Similar to the structure of post-fusion class I proteins, the fusion loops and the transmembrane domains are then located at the same end of an elongated molecule that is now perpendicular to the membrane.

Several cellular fusogens have the class II fold

Remarkably, several eukaryotic proteins involved in cell-cell fusion have been found to possess the class II fold and, therefore, share a common ancestor with all class II viral fusion glycoproteins. This was initially demonstrated for the EFF1 (epithelial fusion failure 1) protein which catalyzes cell-cell fusion events required for the development of *Caenorhabditis elegans* (Fig. 3(E)). Later, the

class II fold was also found in *Chlamydomonas*, *Arabidopsis*, *Plasmodium* and *Tetrahymena* HAP2 proteins which mediate gamete fusion (Fig. 3(E)). In their post-fusion state, all those cellular fusogens form a trimer organized as the post-fusion trimer of class II viral fusion glycoproteins (Fig. 3(E)).

Despite the conservation of their folds, it is not clear whether those proteins all act similarly. Indeed whilst several HAP2 proteins have one or several fusion loops and probably mediate fusion in a manner similar to viral fusogens, EFF1 does not seem to have such a loop and has to be present in both membranes to catalyze their merger.

It is unclear whether the gene encoding the original class II fusion glycoprotein was present in a cellular or a viral genome. However, it is certain that there have been horizontal transfers of genes encoding class II fusion glycoproteins between viruses and their host (whatever the direction of this transfer) and/or between viruses during co-infections.

Class III Fusion Glycoproteins

VSV G, the prototype of class III fusion glycoproteins

The third class of fusion proteins was identified when the structures of the ectodomains of the fusogenic glycoproteins G of VSV and glycoprotein B (gB) of herpes simplex virus 1 (HSV-1) were determined. Their comparison revealed an unanticipated similarity between the structures of the two proteins, although no sequence similarity had been detected previously.

VSV G is involved in both membrane fusion and receptor recognition. It is the only class III fusion protein for which the X-ray structures of both the pre-fusion and post-fusion trimeric states have been determined (Fig. 4(A)). The polypeptide chain of G ectodomain folds into three distinct domains: the fusion domain (FD) is inserted in a loop of a pleckstrin homology domain (PHD) that is itself inserted in a trimerization domain (TrD) which is involved in G trimerization in both states. The domains are connected to each other by segments that change their conformation during G structural transition. Segments R1 and R4 connect TrD to PHD, while R2 and R3 connect PHD to FD. The TrD comprises a β -sheet rich region connected to the C-terminal segment (R5) of the ectodomain (Fig. 4(A)).

During the low pH-induced structural transition from pre- to post-fusion state, FD, PHD and TrD retain their tertiary structure but undergo large rearrangements in their relative orientation. In the context of VSV G trimer, the pre- and post-fusion states are related by flipping both FD and the C-terminal segment around TrD. The conformational change is driven by refolding of segments R1 to R5. At the end of the transition, in the post-fusion state, the core of the structure is organized in a 6-helix bundle, the trimeric core of which is made by the three TrD central helices (extended by the refolding of segment R4), whose grooves accommodate three lateral helices resulting in the refolding of R5 (Fig. 4(A)). This structural organization is obviously very similar to that of the central core of the post-fusion conformation of some class I viral such as EBOV gp2 and HIV-1 gp41.

On the other hand, the organization of VSV G FD is very similar to that of class II fusion proteins, although they have distinct topologies. Similarities include an extended β -sheet structure terminating in a three-stranded sheet at the tip of which are located two hydrophobic fusion loops. In the pre-fusion structure, in contrast to class I and class II fusion proteins, those fusion loops are not buried at an oligomeric interface but point toward the viral membrane.

A particular feature of VSV G which is shared with other rhabdovirus glycoproteins is that the low pH-induced conformational change is reversible. In fact, there is a pH-dependent equilibrium between the different states of G (i.e., the prefusion state, the post-fusion state and also monomeric conformations that are intermediates on the pathway of the structural transition) that is shifted toward the post-fusion state at low pH. This is a major difference with class I and class II viral fusion glycoproteins for which the pre-fusion conformation is metastable and, therefore, the fusogenic structural transition is irreversible.

Class III common characteristics

Besides VSV G and HSV-1 gB (Fig. 4(B)), fusion glycoproteins of other herpesviruses (including Epstein-Barr virus and cytomegalovirus) and other rhabdoviruses (such as rabies virus) also belong to the class III of fusion glycoproteins. The fusion glycoprotein gp64 of baculovirus and of thogoto virus Gp (an arbovirus member of the Orthomyxoviridae as influenza virus) have also been demonstrated to be class III fusion glycoproteins (Fig. 4(B)) and, on the basis of amino acid sequence alignment, it seems that it is also the case of the glycoprotein of bornaviruses.

All class III viral fusion glycoproteins identified so far are homologous and have the same domain organization as VSV G. They are encountered in viruses which have no common origin (negative-strand RNA and double stranded DNA viruses) which suggests that, as for class II fusion glycoproteins, at least some of them have been acquired by horizontal transfers. Thogoto virus Gp and baculovirus gp64 are extremely similar, suggesting a relatively recent gene transfer between those two arboviruses, whereas rhabdovirus glycoproteins are more distantly related to gB and gp64 suggesting a much more ancient separation.

As already mentioned, we do not know the prefusion conformation of class III glycoproteins except for VSV G. However, low resolution envelopes of herpesviruses gB in their putative prefusion state have been obtained by cryotomography. They strongly suggest that the prefusion state of gB has an organization which is similar to that of VSV G.

Finally, whereas rhabdovirus G, Thogoto Gp and baculovirus gp64 are triggered at low pH, herpesvirus gB only functions in association with two other glycoproteins gH and gL which form a complex that regulate its activity. It is only when a fourth protein (gD for HSV-1 and gp42 for EBV) binds a cellular receptor, that the complex made by gH and gL undergoes a structural transition which itself triggers gB refolding.

Other Viral Fusion Glycoproteins

There are viral fusion proteins that cannot be classified in one of the three previously described classes.

Among unclassified fusion proteins from eukaryotic viruses are those of the genera pesti- and hepacivirus which both belong to the flaviviridae family. Unlike members of the flavivirus genus (which possess only one envelope glycoprotein, E), pestiviruses and hepaciviruses encode two envelope glycoproteins, E1 and E2. Based on the disulfide-bonds pattern, it was thought that the fusion protein was E2 and that it had a class II fold. However, structural studies performed on E2 of bovine viral diarrhea virus (BVDV, a pestivirus) and hepatitis C virus (HCV, a hepacivirus) revealed that neither BVDV E2 nor HCV E2 adopt a class II fold. Furthermore, BVDV E2 and HCV E2 structures, which otherwise are unrelated, appear to lack an N-terminal fusion peptide and internal fusion loops indicating that they are probably not viral fusion proteins. Finally, the structure of the amino terminal domain of HCV E1 reveals a complex network of covalently linked intertwined homodimers which definitively demonstrate that HCV has a novel fusion machinery of which the basic characteristics remain to be determined.

There are also many families of eukaryotic viruses for which little is known about the fusion machinery. As an example, the fusion machinery of poxviruses (of which the prototype is vaccinia virus) is not well characterized. Genetic and biochemical studies have provided evidence for an entry fusion complex, made up of about ten viral proteins, whose precise functions in the process are unknown. Similarly, the structure and working of the fusion machinery of the members of the Hepadnaviridae family (of which the prototype is Hepatitis B virus) is poorly understood.

Finally, in a very general manner, the fusion machineries of prokaryotic virus have not been characterized. A notable example is the determination of the fusion protein, called VP5, of two haloarchaeal pleomorphic viruses (HRPV-2 and HRPV-6) infecting archaeal hosts living in hypersaline environments and belonging to the Pleolipoviridae family. VP5 has a novel V-shaped fold and is monomeric on the viral surface. It may be a representative of a new class of prokaryotic virus fusion proteins.

Regulating the Conformational Change to Avoid Premature Triggering

As mentioned above, in most cases, the prefusion conformation of viral fusion proteins is metastable and the conformational change is irreversible. Therefore, premature triggering of the structural transition of the fusion protein is deleterious. This is particularly problematic for fusion glycoproteins that are triggered at low pH because, during their maturation and transport, they cross Golgi compartments that are slightly acidic. It appears that those viruses have evolved the ability to protect their fusion protein from undergoing irreversible transition toward their postfusion state and/or to avoid undesirable fusion reactions.

Some viruses have evolved the ability to regulate the pH of the transport vesicles. For influenza virus, this is achieved by the viral M2 protein which is a proton channel.

Furthermore, for all class I protein, the cleavage that primes the protein for the low pH-induced fusogenic conformational change occurs at a late stage of transport or even after viral budding in the extracellular medium.

Similarly, for class II fusion proteins, the cleavage of the companion protein is required to prime the fusion glycoprotein. The maturation process has been largely characterized for flaviviruses and particularly DENV. Immature DENV particles bud into the ER as spiky virions with 60 icosahedrally arranged protrusions, each consisting of a trimer of prM-E heterodimers. The immature particles are transported into the trans-Golgi network (TGN), where low pH environment induces a conformational change in the virion, resulting in a rearrangement of the surface glycoproteins and particles having a much smoother surface with E proteins arranged in the so-called herringbone pattern. This transition exposes the furin cleavage site allowing the proteolysis of prM into pr peptide and M to take place in the TGN. After cleavage, pr peptide remains associated with the virion. It stabilizes the metastable E dimer, preventing its dissociation and the formation of the postfusion homotrimers. It also masks the fusion loop of E, preventing interaction with host cell membranes. Finally, pr dissociates from the virion when the particle is released into the slightly alkaline extracellular medium. This definitively primes the metastable E dimer for subsequent low-pH-induced fusogenic conformational change.

For rhabdoviruses such as VSV or rabies virus, the native, prefusion conformation is not metastable and, as a consequence, the low-pH-induced structural transition is reversible. This allows G to be transported through the acidic compartment of the Golgi apparatus and to recover its native prefusion state at the viral surface in the extracellular environment.

Working of the Fusion Machinery

A Common Mechanism of Fusion

Experimental data suggest that the mechanisms for the catalysis of fusion are very similar for all the enveloped viruses studied so far whatever the organization of their fusion machinery.

Indeed, even though the structures of fusion proteins from the different classes are unrelated, the mechanisms for refolding share two key common features. First, the fusion peptides/loops are exposed and projected toward the viral membrane. Second, the folding-back of the C-terminal region onto a trimeric N-terminal region leads to the formation of a post-fusion protein conformation with the outer regions zipped up against the inner trimeric core. This refolding pathway constitutes the basis of the model presented for influenza HA in [Fig. 5](#).

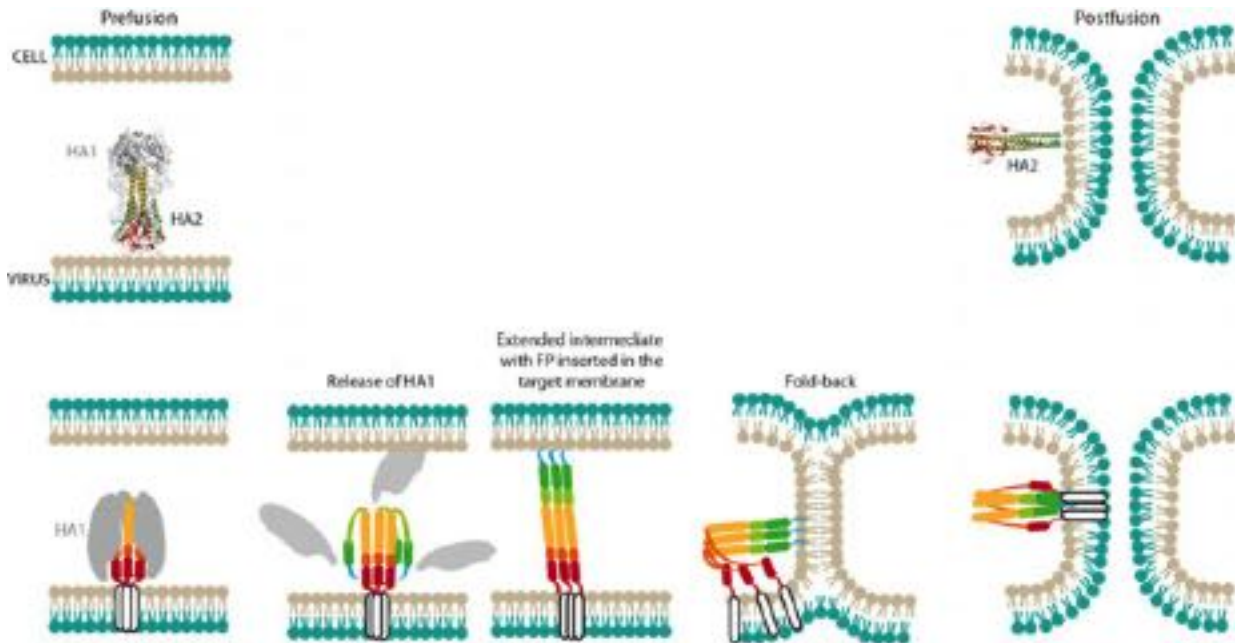


Fig. 5 Schematic diagram of the fusion process between viral and cellular membranes mediated by influenza HA. In the prefusion conformation, the fusion peptide is buried in the trimer interface. Then, after release of HA1 subunits, the formation of a trimeric extended intermediate creates an initial bridge between the two bilayers. This extended intermediate is unstable and refolds into the more energetically favorable postfusion conformation. It is suggested that during this refolding stage the transmembrane domains are involved in the opening of the fusion pore. Similar schemes have been proposed for class II fusion glycoproteins.

Interaction With the Target Membrane

Interactions between fusion peptides and target membranes have been demonstrated for both rhabdoviruses and influenza virus using hydrophobic photolabeling. This interaction is an early step of the fusion process. This indicates that an early intermediate state during the structural transition exposes the fusion peptides that become accessible and are orientated toward the target membrane.

Polar aromatic residues (Tyr and Trp) located in the fusion loops of class II and class III fusion glycoproteins have been shown to be essential to the fusion process by directed mutagenesis experiments. They are supposed to position themselves at the interface between the fatty acid chains and the polar head group layers of lipids. For those glycoproteins, any deep penetration of the fusion domain inside the membrane is precluded by the presence of charged residues and polar groups close to the fusion loops. Such an interaction which only involves a few residues most probably perturbs the outer leaflet of the target bilayer and facilitates the formation of point-like protrusions that have been suggested to be stalk precursors.

Some class I fusion glycoproteins such as Ebov GP and ASLV Env also have fusion loops (located close to the amino-terminal end of the fusion subunit) which are enriched in aromatic residues. However, most class I fusion glycoproteins have a fusion peptide. Structures of synthetic fusion peptides of influenza virus HA2 and of HIV gp41 in membrane mimicking environments have been characterized by several biophysical approaches (including NMR and EPR spectroscopies) and molecular dynamics. In the case of HA2, the fusion peptide adopts a kinked helical amphipathic structure which, here again, does not seem to penetrate deep into the bilayer structure but rather induces local membrane curvature.

Structural Intermediates During the Conformational Changes

An elongated trimeric intermediate for class I and class II fusion glycoproteins

For class I fusion glycoproteins, it is assumed that the target-membrane-interacting conformation is in a trimeric extended conformation in which the central trimeric coiled-coil (made by HR1 heptad repeats) of the fusion subunit is already formed but the relocation of the C-terminal segments against this central core has not yet occurred. In this conformation, the fusion peptide is located at one end and interacts with the target membrane while the transmembrane domain, anchored in the viral membrane, is at the other end. This creates an initial bridge between the two bilayers.

Class II fusion glycoproteins form monomers during the structural transition. Although, here again, it is assumed that, at some stage, an extended trimeric intermediate is bridging the viral and target membranes, the oligomeric status of the glycoprotein when it interacts with the target membrane remains a matter of debate and may vary between class II glycoproteins. For TBEV E, the initial interaction with the target membrane seems to involve the insertion of multiple copies of E monomers. However, for SFV, it has been proposed that trimer formation is rapid and might precede membrane insertion.

For class I and class II fusion glycoproteins, the transition from the trimeric extended intermediate to the trimeric post-fusion state cannot maintain strict threefold symmetry. This symmetry is disrupted by the refolding -back of the C-terminal portion of the molecule. It is worth noting that during the structural transition from the putative intermediate to the post-fusion state, the TM segments cannot remain associated as they have to rotate independently of each other to achieve the transition.

Monomeric intermediates for vesiculovirus G

For VSV and Chandipura virus glycoprotein G, it has been demonstrated that both in solution (after proteolysis or detergent solubilization) and at the viral surface, there was an equilibrium between pre-fusion and post-fusion trimers and flexible monomers. The sequential appearance of the different species when lowering the pH suggested that the monomers were intermediates during the conformational change.

Two distinct intermediate monomeric conformations have been trapped in a single crystal form. The first one resembled the prefusion protomer and corresponded to an early intermediate. However, in this conformation, the partial refolding of the C-terminal segment of the ectodomain renders G able to bridge the viral and target membranes. The second one corresponded to a late intermediate on the transition pathway. Those two structures offer a plausible pathway for the conformational change and indicate that most of the structural transition occurs in a monomeric form and that retrimerization into the post-fusion state is a late event.

It is not clear whether this pathway can be generalized to other class III fusion glycoproteins. However, it is worth noting that independent refolding of G monomers before re-association to form the post-fusion trimer overcomes some of the problems encountered during class I and class II fusion glycoprotein refolding from their extended trimeric intermediate to their post-fusion state.

Cooperativity

Whatever the class of fusion glycoproteins, it has been shown that the fusion process requires the cooperation of several trimers. For influenza HA, West Nile Virus E and VSV G, the cooperation of 3–5 adjacent trimers is required to reach the hemifusion state. Cryo-electron tomography has suggested that they are organized around a lipid neck joining the viral and the liposome membranes.

It has also been suggested that glycoproteins located outside the contact zone might be involved in the late stages of the fusion process (i.e., the enlargement of the initial fusion pore). Indeed, several fusion glycoproteins in their post-fusion conformation have a tendency to cluster at the viral surface and sometimes form regular arrays. This is the case for several class II fusion glycoproteins of which spikes in their post-fusion conformation self-associate to form more or less regular networks that are different from the initial organization of the prefusion state. Similarly, VSV glycoproteins in their post-fusion conformation form helical arrays around the viral particle. The concerted formation of such networks might explain how glycoproteins surrounding the contact zone between fusing membranes exert a role at the late stages of the fusion process.

Conclusions

The structures of viral fusion glycoproteins and the deduced phylogeny illustrate nicely that viral genomes are a combination of gene modules that can be exchanged during evolution by horizontal transfer. Such transfers of genes could occur during co-infection of cells by distinct viruses. However, the homology between class II viral fusion glycoproteins and cellular fusogens suggest that viral fusion proteins genes may have been acquired from the genome of infected cells by viruses from unrelated viral families.

The field of virus fusion has greatly benefited from the progress of structural biology. This has revealed similar principles of action for fusion machineries made up of proteins having very different shapes. The next challenge is to determine how these fusion machineries operate when they interact with the target membrane and to get access to their dynamics at the highest possible spatial and temporal resolution.

Finally, viral fusion can constitute a target for antiviral strategies. several antiviral drugs that block enveloped virus fusion are already used or in development. A challenge for the future is to develop molecules that inhibit fusion machineries from a broad spectrum of viral families.

Further Reading

- Bullough, P.A., Hughson, F.M., Skehel, J.J., Wiley, D.C., 1994. Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 371, 37–43.
- Chernomordik, L.V., Kozlov, M.M., 2008. Mechanics of membrane fusion. *Nature Structural Molecular Biology* 15, 675–683.
- El Omari, K., Li, S., Kotecha, A., *et al.*, 2019. The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. *Nature Communications* 10, 846.
- Guardado-Calvo, P., Rey, F.A., 2017. The envelope proteins of the bunyavirales. *Advances in Virus Research* 98, 83–118.
- Harrison, S.C., 2015. Viral membrane fusion. *Virology* 479–480, 498–507.
- Heldwein, E.E., Lou, H., Bender, F.C., *et al.*, 2006. Crystal structure of glycoprotein B from herpes simplex virus 1. *Science* 313, 217–220.

- Julien, J.P., Cupo, A., Sok, D., *et al.*, 2013. Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 342, 1477–1483.
- Lescar, J., Roussel, A., Wien, M.W., *et al.*, 2001. The fusion glycoprotein shell of Semliki forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105, 137–148.
- Modis, Y., Ogata, S., Clements, D., Harrison, S.C., 2004. Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427, 313–319.
- Rey, F.A., Heinz, F.X., Mandl, C., Kunz, C., Harrison, S.C., 1995. The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375, 291–298.
- Roche, S., Bressanelli, S., Rey, F.A., Gaudin, Y., 2006. Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* 313, 187–191.
- Roche, S., Rey, F.A., Gaudin, Y., Bressanelli, S., 2007. Structure of the prefusion form of the vesicular stomatitis virus glycoprotein G. *Science* 315, 843–848.
- Weissenhorn, W., Dessen, A., Harrison, S.C., Skehel, J.J., Wiley, D.C., 1997. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* 387, 426–430.
- White, J.M., Whittaker, G.R., 2016. Fusion of enveloped viruses in endosomes. *Traffic* 17, 593–614.
- Wilson, I.A., Skehel, J.J., Wiley, D.C., 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* 289, 366–373.

Genome Replication of Bacterial and Archaeal Viruses

Česlovas Venclovas, Vilnius University, Vilnius, Lithuania

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Negative sense (–) strand A negative-sense DNA or RNA strand has a nucleotide sequence complementary to the messenger RNA and cannot be directly translated into protein.

Positive sense (+) strand A positive sense DNA or RNA strand has a nucleotide sequence, which is the same as that of the messenger RNA, and the RNA version of this sequence is directly translatable into protein.

Protein-primed DNA replication DNA replication whereby a DNA polymerase uses the 3'-OH group provided by the specialized protein as a primer to synthesize a new DNA strand.

RNA-primed DNA replication Conventional DNA replication used by all cellular organisms whereby a primase synthesizes a short RNA primer with a free 3'-OH group which is subsequently elongated by a DNA polymerase.

Rolling-circle DNA replication DNA replication whereby the replication initiation protein creates a nick in the circular double-stranded DNA and becomes covalently attached to the 5' end of the nicked strand. The free 3'-OH group at the nick site is then used by the DNA polymerase to synthesize the new strand.

Genomes of Prokaryotic Viruses

At present, all identified archaeal viruses have either double-stranded (ds) or single-stranded (ss) DNA genomes. Although metagenomic analyses suggested the existence of archaeal viruses with RNA genomes, this finding remains to be substantiated. Bacterial viruses, also referred to as bacteriophages or phages for short, have either DNA or RNA genomes, including circular ssDNA, circular or linear dsDNA, linear positive-sense (+)ssRNA or segmented dsRNA (Table 1). So far, no bacteriophages with negative sense (–)ssRNA genomes have been identified. Both archaeal and bacterial viruses with dsDNA genomes are most abundant, whereas those with ssDNA genomes represent a smaller group. The genome size of prokaryotic viruses is clearly linked with the structure of the genome. Thus, genomes approximately up to 25 kb in size are represented by both types of nucleic acids (DNA or RNA) and various topologies (circular, linear, segmented). All genomes larger than that are represented only by circular or linear dsDNA (Fig. 1). This size-dependent choice of the carrier of genomic information likely reflects physical constraints related to the genome stability, which is the highest for dsDNA.

Genome Replication of Prokaryotic dsDNA Viruses

Genomes of prokaryotic dsDNA viruses range from ~5 to ~500–600 kb, they can be either circular or linear. At present, dsDNA bacteriophages comprise the largest fraction of prokaryotic viruses that have their genomes sequenced. According to the data on complete viral genomes available at the NCBI database (see “Relevant Websites section”) currently the smallest dsDNA genome (10079 nt) of a bacterial virus is represented by the tailless *Pseudomonas* virus PM2. Phage PM2 is a representative of family *Corticoviridae* and the first lipid-containing phage to be isolated. The largest genome among officially classified bacterial dsDNA viruses belongs to the tailed *Bacillus* virus G (Order *Caudovirales*, Family *Myoviridae*) reaching nearly 500 kb (497513 nt). Recent metagenomics studies have identified phages with even larger dsDNA genomes that are in the 540–600 kb range. This points to the hidden diversity and abundance of very large phages and suggests that phages carrying even larger genomes might still be discovered. Notably, largest bacteriophage genomes are comparable in size with genomes of small bacteria capable of autonomous growth such as *Mycoplasma genitalium* with the 580 kb genome and exceed those of many symbiotic and parasitic bacteria. Archaeal dsDNA viruses range in size from 5.3 kb in the *Aeropyrum pernix* bacilliform virus 1 (APBV1), a member of the family *Clavaviridae*, to 144 kb for *Halogramma* tailed virus 1 (HGTV-1), a representative of the *Caudovirales*.

DNA replication has been studied in detail for only a handful of prokaryotic dsDNA viruses such as model bacteriophages phi29, T7 and T4. For most prokaryotic dsDNA viruses, however, there is little or no experimental data on their genome replication mechanisms. Therefore, whatever is inferred about DNA replication machineries for majority of prokaryotic dsDNA viruses typically comes from *in silico* analysis of viral genomes and corresponding proteomes. Based on both experimental characterization and computational analyses of the repertoire of putative DNA replication genes several major types of dsDNA genome replication systems can be distinguished. (1) A common type is the one that mirrors major molecular functions associated with RNA-primed DNA replication in cellular organisms. However, this type of DNA replication machineries may differ greatly in completeness in various prokaryotic viruses. Some, such as phage T4, have essentially autonomous DNA replication machinery, which does not require assistance from the host. Others encode only few components of their own and rely on the host to provide the remaining ones (e.g., phage lambda). (2) Protein-primed DNA replication system specific to viruses and other mobile genetic elements (MGEs). (3) Rolling circle DNA replication (RCR) system, in which the key component is a viral replication initiation protein

Table 1 Classification of prokaryotic viruses

Genome type	Archaeal viruses	Bacterial viruses
dsDNA	Order: Caudovirales Families: <u>Myoviridae</u> (HGTV-1) <u>Siphoviridae</u> (HCTV-1) Order: Ligamenvirales Families: <u>Lipothirixviridae</u> (AFV1) <u>Rudiviridae</u> (SIRV2) Other families: <u>Ampullaviridae</u> (ABV) <u>Bicaudaviridae</u> (ATV) <u>Clavaviridae</u> (APBV1) <u>Fuselloviridae</u> (SSV1) <u>Globuloviridae</u> (PSV) <u>Guttaviridae</u> (APOV1) <u>Ovaliviridae</u> (SEV1) <u>Pleolipoviridae</u> (His2) <u>Portogloboviridae</u> (SPV1) <u>Sphaerolipoviridae</u> (SNJ1) <u>Tristromaviridae</u> (PFV1) <u>Turriviridae</u> (STIV) <u>Pleolipoviridae</u> (HRPV-1) <u>Spiraviridae</u> (ACV)	Order: Caudovirales Families: <u>Ackermannviridae</u> (Limestone) <u>Herelleviridae</u> (SP01) <u>Myoviridae</u> (T4, P2) <u>Podoviridae</u> (T7, phi29) <u>Siphoviridae</u> (λ) Other families: <u>Corticoviridae</u> (PM2) <u>Plasmaviridae</u> (L2) <u>Sphaerolipoviridae</u> (IN93) <u>Tectiviridae</u> (PRD1)
ssDNA		<u>Inoviridae</u> (M13)
dsRNA	N/A	<u>Microviridae</u> (phiX174)
ssRNA (+)	N/A	<u>Cystoviridae</u> (phi6)
ssRNA (–)	N/A	<u>Leviviridae</u> (MS2)
		N/A

Note: Viral families according to the ICTV Master Species List 34, 2018b with representatives indicated in parentheses. Families common to both archaeal and bacterial viruses are underlined. N/A, viruses with the indicated genome structure have not been identified (not available).

(Rep) whereas all or nearly all other proteins required for DNA replication are recruited from the host. (4) DNA replication systems that use other strategies to propagate viral genome (see below).

RNA-Primed DNA Replication

DNA replication systems in many prokaryotic viruses, particularly in bacteriophages, follow the same organizational principles as the corresponding systems of their cellular hosts. Archaea and bacteria use RNA-primed DNA replication systems that are made up of components which may not always be evolutionarily related, but perform the same function. At the most general level, cellular DNA replication machineries include the following key functional components: replicative helicases, primases, replicative DNA polymerases, DNA sliding clamps and clamp loaders, single-stranded DNA binding proteins, primer removal nucleases, DNA ligases and topoisomerases (Fig. 2). Typically, viral RNA-primed DNA replication machineries represent hybrid systems that are composed of proteins encoded by the viral genome and those provided by the host. According to the ratio of their own/host-provided DNA replication proteins viruses may differ considerably. Some viruses encode DNA replication systems that are essentially autonomous or nearly autonomous, whereas others encode only one or two DNA replication proteins and are largely dependent on the host to provide the rest. There is a continuum between these extreme cases. A global survey of viral-encoded DNA replication proteins has shown that both archaeal viruses and bacteriophages most frequently encode their own replicative helicases and not, as might be expected, DNA polymerases. Primases are also more frequent than DNA polymerases in both archaeal viruses and bacteriophages.

In addition, it was observed that certain functional categories of viral DNA replication proteins are tightly coupled. One such tightly coupled group comprises replicative helicase, primase and DNA polymerase. Viruses encoding either a DNA polymerase or primase nearly always encode a replicative helicase. The inverse co-occurrence is not nearly as stringent. Thus, in many cases, viruses, which encode replicative helicases, lack the genes for primases and DNA polymerases and likely rely on the corresponding proteins of the host. This asymmetry suggests that viral primases and DNA polymerases are constrained to work together with viral helicases. Such a coupling reflects intimate physical and functional interactions between these three proteins. Primase and helicase are often fused into a single polypeptide chain as exemplified by the primase-polymerase encoded by gene 4 of phage T7. There are

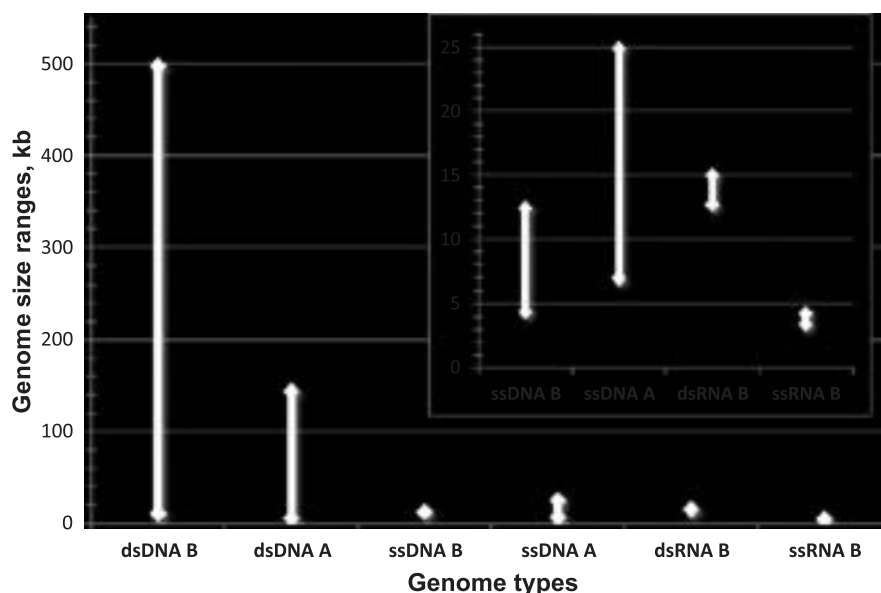


Fig. 1 Genome size ranges of prokaryotic viruses with different genome structure. B, bacteriophages, A, archaeal viruses. The inset shows a zoom-in on viruses with ssDNA and RNA genomes.

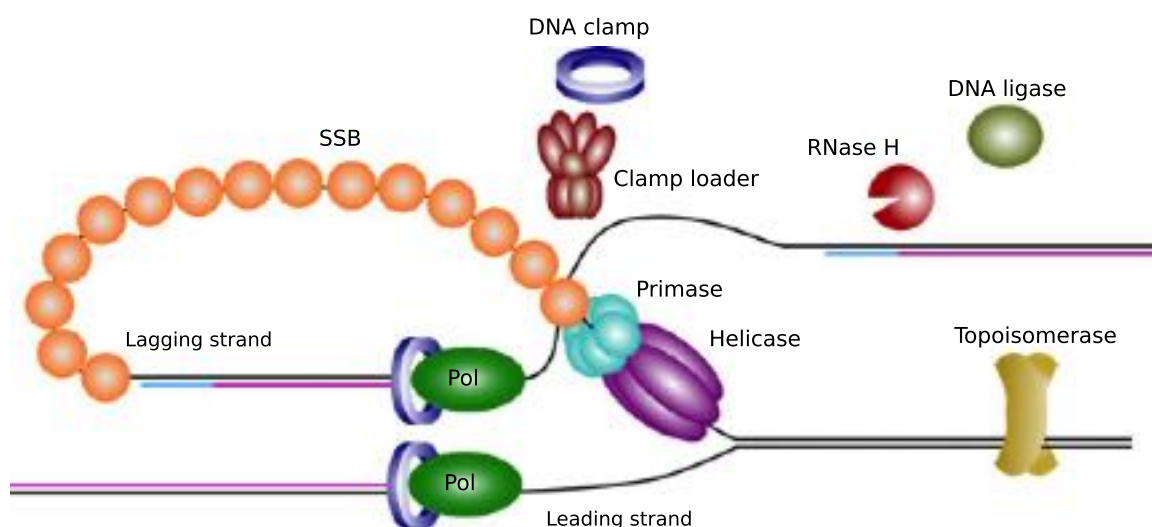


Fig. 2 Schematic representation of DNA replication fork based on bacteriophage T4 RNA-primed DNA replication system. The leading strand is synthesized continuously whereas the lagging strand is synthesized as Okazaki fragments. Newly synthesized strands are shown in magenta, RNA primers are represented as cyan-colored regions.

also cases where all three proteins (helicase, primase and polymerase) are fused together. One such example is the gp55 protein (Acc: YP_001828703) of *Lactococcus* phage 1706, and another one is the Pas55 protein (Acc: YP_024841) of *Actinoplanes* phage phiAsp2. Unfortunately, no experimental data are available on the molecular mechanism of these multifunctional proteins.

Another strong, unidirectional functional link was observed between viral DNA polymerases and their accessory factors, DNA sliding clamps and clamp loaders. It was found that all viruses which encode clamp loaders also encode DNA sliding clamps and DNA polymerases. This co-occurrence pattern is a direct reflection of steps involved in the assembly of a processive DNA replicase. The clamp loader loads the clamp, which in turn tethers DNA polymerase to the DNA enabling processive DNA synthesis. Again, the opposite link between the presence of a viral polymerase and accessory proteins is weak. The reason for this is that some viral DNA polymerases have high intrinsic processivity whereas others utilize host proteins to increase the processivity of DNA synthesis. A well-known example of host-supplied processivity factor is the case of *Escherichia coli* phage T7. By itself the T7 DNA polymerase is a low-processivity enzyme, however, upon binding the *E. coli* thioredoxin the processivity of the DNA polymerase increases by about 1000-fold.

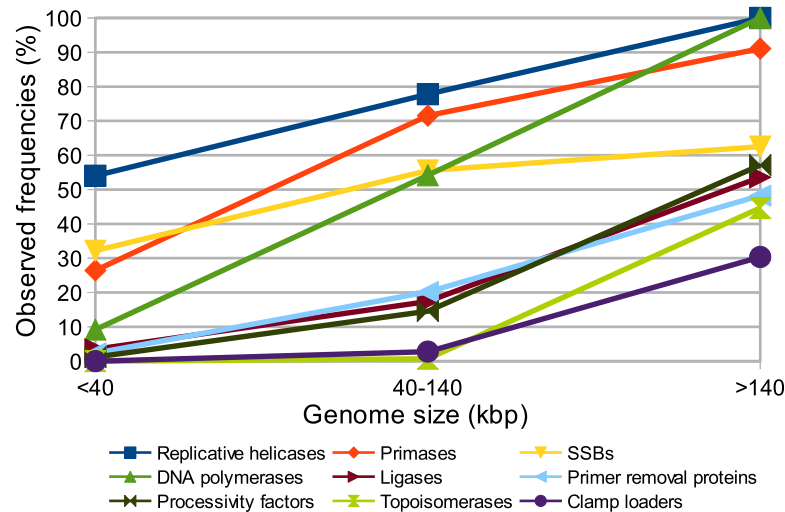


Fig. 3 Observed frequencies of DNA replication proteins encoded in genomes of prokaryotic dsDNA viruses. The figure is adapted from Kazlauskas, D., Krupovic, M., Venclovas, Č., 2016. The logic of DNA replication in double-stranded DNA viruses: Insights from global analysis of viral genomes. *Nucleic Acids Research* 44, 4551–4564.

Table 2 Prokaryotic dsDNA viruses having the most complete DNA replication machineries

Taxon/Virus	Genome size (kb)	DNA polymerase	DNA clamp	Clamp loader	Helicase	Primase	SSB	Primer removal protein	Ligase	Topoisomerase
Bacillus phage G	498	■	□	□	■	■	■	■	■	■
T4-like	135-359	■	■	■	■	■	■	■	▨	▨
Ralstonia phage RSL1	231	■	■	■	■	■	■	■	■	■
Sphingomonas phage PAU	219	■	■	■	■	■	■	■	■	■
Clostridium phage c-st	186	■	□	□	■	■	■	□	■	■
Microcystis aeruginosa phage Ma-LMM01	162	■	■	■	■	■	□	□	□	□
Cyanophage S-TIM5	161	■	□	□	■	■	□	■	□	■
Halovirus HVTv-1	102	■	■	■	■	■	□	■	□	□
Vibrio phage vB_VpaS_MAR10	79	■	■	■	■	■	■	■	■	□
Clostridium phage phiCTP1	59	■	■	■	■	■	■	□	■	□
T7-like	37-42	■	□	□	■	■	■	■	■	□

Note: DNA replication proteins are marked by their presence in the genome (present, black square; absent, white square; present in some members, black-and-white square). Viruses are arranged by the genome size and colored by their host (yellow, bacteriophages; pink, archaeal viruses). Adapted from Kazlauskas, D., Krupovic, M., Venclovas, Č., 2016. The logic of DNA replication in double-stranded DNA viruses: Insights from global analysis of viral genomes. *Nucleic Acids Research* 44, 4551–4564.

What determines the completeness of the viral genome-encoded DNA replication machinery? It appears that one of the key factors is the size of the viral genome. In a sense, viral genome may be compared to a computer disk – the larger the disk, the more data it can hold (Fig. 3). For example, both archaeal and bacterial viruses with dsDNA genomes exceeding 140 kb all encode their own replicative helicases and DNA polymerases. However, the correlation between the genome size and the completeness of the viral-encoded DNA replication machinery is quite noisy and far from perfect. The increased coding capacity of viral genome is not always a decisive factor. This can be clearly seen in the list of prokaryotic viruses with the most complete DNA replication machineries (Table 2). For example, phage T4 with the 169 kb genome has all the components needed for the genome replication, whereas a phage with the 498 kb genome (*Bacillus* phage G) lacks the recognizable accessory subunits (DNA sliding clamp and clamp loader) of a DNA polymerase.

Not only the completeness of viral DNA replication machineries is highly variable, but also components making up these machineries are highly diverse. In particular, DNA replication proteins used by bacteriophages often differ from those used by their bacterial hosts. For chromosomal DNA replication bacteria use the DnaB-type superfamily 4 (SF4) helicases, TOPRIM domain-containing primase (i.e., *E. coli* DnaG), C-family DNA polymerase, processivity β -clamp and bacterial clamp loader, OB-fold-containing SSB protein. RNA primers in bacteria are excised by RNase HI and Pol I 5'-3' exonuclease domain homologous to FEN1. Nicks are then sealed by a NAD⁺ – dependent DNA ligase. DNA replication machineries encoded by bacteriophages represent a mosaic of proteins typical of bacterial hosts, proteins of archaeo-eukaryotic type and those specific to viruses and other MGEs. Thus, bacteriophages are not simply mirroring DNA replication systems employed by their bacterial hosts. There are a number of components notably different from the bacterial ones. Among these are Superfamily 3 (SF3) helicases that are specific to MGEs and that are not involved in chromosomal replication of cellular organisms. Bacteriophages also frequently encode archaeo-eukaryotic primases (AEP) that have unrelated structural fold to bacterial TOPRIM primases. Most interestingly, C-family DNA polymerases that represent replicative polymerases in bacteria, are rarely found in bacteriophages. Instead, most abundant DNA polymerases in bacteriophages are represented by A- and B-family enzymes, both featuring the same structural fold of their catalytic domain. A-family polymerases in bacteria are ubiquitous, but their role is limited to participation in gap-filling during replication and in DNA repair. B-family polymerases in eukaryotes and archaea represent replicative enzymes whereas in bacteria their presence is sporadic. Thus, B-family polymerases provide another link between bacteriophage and archaeal/eukaryotic replication machineries. Some bacteriophages have DNA polymerase accessory proteins that also appear to be of archaeal/eukaryotic origin. The case in point is DNA sliding clamps and clamp loaders of T4-like bacteriophages. Both archaeal/eukaryotic (PCNA) and T4 (gp45) clamps are homotrimers, unlike the dimeric bacterial β -clamps. In addition, the crystal structure of T4 clamp loader (gp44/62) revealed that it represents a minimal version of the archaeal/eukaryotic RFC clamp loader. Archaeal/eukaryotic proteins participating in the lagging strand synthesis, most notably ATP-dependent DNA ligase, are also common in bacteriophage genomes.

Components of DNA replication systems encoded by archaeal viruses generally are homologous with the archaeal counterparts. However, there are fewer complete genomes of archaeal viruses, they tend to be smaller than in the case of bacteriophages and they are extremely diverse. Therefore, it is difficult just yet to make broader generalizations regarding components of RNA-primed replication systems in archaeal dsDNA viruses. Notably, some previous analyzes of complete archaeal genomes failed to find typical SSB proteins featuring OB-fold. However, a more recent metagenomics study identified archaeal members of the *Caudovirales*, dubbed Magroviruses, that possess dsDNA genomes of 65–100 kb in size and encode a nearly complete replication apparatus of apparent archaeal origin including SSB proteins. In addition to typical archaeal/eukaryotic proteins, just like bacteriophages, archaeal viruses were found to encode SF3 helicases, specific to MGEs.

Thus, there are some DNA replication proteins of prokaryotic viruses that are spread across both bacterial and archaeal domains. The 'universal' proteins include SF3 helicases, AEP primases, B-family DNA polymerases, RNase H and ATP-dependent DNA ligases. Only RNase H is commonly found in both bacteria and archaea, the remaining proteins, except SF3 helicases, are of archaeal/eukaryotic origin. If the similarity is considered at the structure level, the set of common archaeal/eukaryotic proteins could be extended even further into DNA sliding clamps and clamp loaders (PCNA/T4 gp45 and RFC/(T4 gp44/gp62)).

Protein-Primed DNA Replication

Some bacterial viruses with relatively small linear dsDNA genomes utilize the so-called protein-primed DNA replication. Cellular replicases use the 3'-hydroxyl (3'-OH) group provided by a nucleic acid primer (RNA or DNA) to initiate DNA synthesis. In contrast, protein-primed DNA polymerases utilize the OH group presented by a specific serine, threonine or tyrosine residue of a terminal protein (TP). Protein-primed DNA replication systems are distinctly different from those utilizing nucleic acid primers. Unlike the latter, protein-primed DNA replication systems require only few components to be fully functional. Such systems typically include a distinct B-family DNA polymerase, TP and an atypical single-stranded DNA binding protein (SSB). The DNA replication machinery of a protein-primed DNA replication system may be exemplified by one of the best studied such systems from *Bacillus subtilis* phage phi29 (Fig. 4). Phage phi29 genome is a linear approximately 19 kb-long dsDNA with covalently attached TP to each 5'-end. DNA replication is initiated by binding of the heterodimer of phi29 DNA polymerase and the free TP to the genomic DNA ends. Phi29 DNA polymerase uses the OH group of a specific serine residue as a primer bypassing the need for a primase. During the extension stage the phi29 DNA polymerase continues to synthesize DNA in a standard DNA-primed fashion and the displaced single-stranded regions are covered by SSB. Once the replication of a single DNA duplex is completed, there are two newly synthesized linear DNA duplexes with TP covalently attached to the 5'-ends. The structure of the phi29 DNA polymerase, so far the only experimentally determined 3D structure of a protein-primed DNA polymerase, has been instrumental in understanding structural and functional features of protein-primed DNA polymerases. Phi29 and other protein-primed DNA polymerases have two additional subdomains, Terminal Protein Region 1 (TPR1) and 2 (TPR2), that distinguish them from B-family members involved in conventional RNA-primed DNA replication. The TPR1 subdomain is involved in interaction with the TP. The palm, thumb and TPR2 subdomains form an internal sliding clamp-like structure that encircles the upstream duplex DNA, providing the enzyme with its inherently high processivity without the assistance of processivity factors. In addition, the TPR2 subdomain of the phi29 DNA polymerase couples processive DNA synthesis with the strand displacement in downstream dsDNA, making the function of a replicative helicase unnecessary. Based on identified genes for protein-primed DNA polymerase,

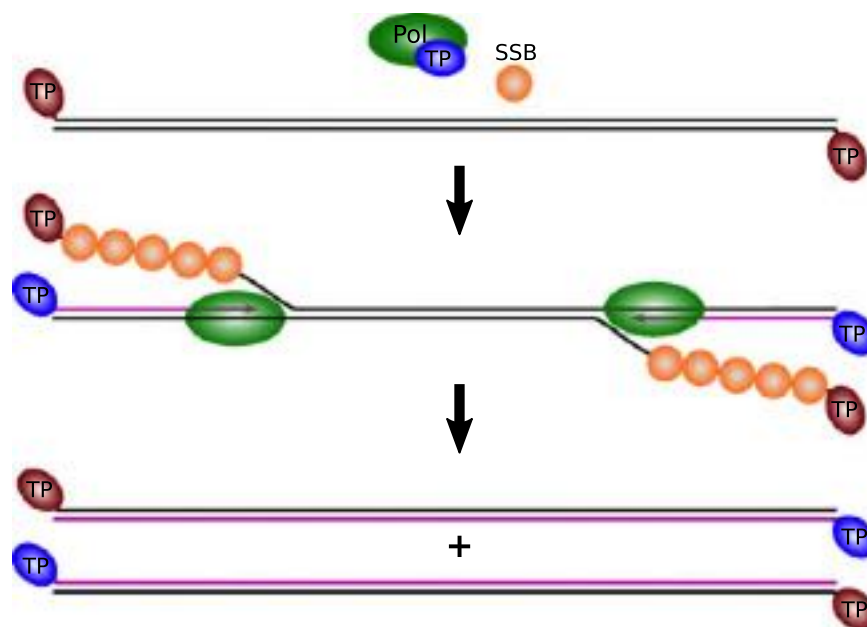


Fig. 4 Schematic representation of protein-primed replication of bacteriophage phi29. Pol, polymerase; TP, terminal protein; SSB, single-stranded DNA binding protein. Newly synthesized strands are shown in magenta.

five groups of archaeal viruses are also believed to replicate their genomes in this manner. These groups include bottle-shaped ampullaviruses, spindle-shaped salterprovirus His1, pleolipovirus His2, *Sulfolobus ellipsoid virus 1* and *Methanosarcina* Spherical Virus (MetSV). However, in these cases protein-primed DNA replication systems have not been studied experimentally. Apparently, the protein-primed DNA replication system is optimal for small-sized viral genomes, because these systems have only been identified in viruses with dsDNA genomes of no more than 50 kb.

Rolling Circle DNA Replication

Some archaeal and bacterial dsDNA viruses lacking typical components of either RNA-primed or protein-primed DNA replication systems utilize the so-called rolling circle DNA replication (RCR) mode. The signature of RCR is the presence of the multifunctional replication initiation protein (RCR-Rep). In many cases, RCR-Rep is the only viral protein needed for viral DNA replication. Therefore, RCR replication systems are prevalent in small viral genomes. The RCR replication mode is not specific to dsDNA prokaryotic viruses. In fact, RCR was first discovered and characterized in *E. coli* bacteriophage phiX174 and other ssDNA bacteriophages. Subsequently, RCR was identified as the replication mode of many other genetic elements such as bacterial and archaeal plasmids as well as a mechanism used by plasmids for conjugative DNA transfer. DsDNA bacteriophages that utilize the RCR replication mode include bacteriophages such as P2, 186, HP1 and PM2. The general mechanism of dsDNA bacteriophage genome replication using the RCR mode can be illustrated with that of bacteriophage P2 (**Fig. 5**). Following the entry of linear P2 DNA into the host cell, DNA circularizes due to the cohesive DNA ends and is sealed by the cellular DNA ligase. The RCR-Rep protein (protein A) of bacteriophage P2 nicks a circular DNA at the replication origin generating the covalent protein-DNA intermediate and a free 3'-OH. The latter primes leading strand DNA synthesis by the bacterial replicase, which uses the non-cleaved strand as a template. After the full circle is replicated, protein A cuts the newly generated origin and also acts as a ligase, producing a covalently closed circular ssDNA molecule. In this way, RCR-Rep is able to perform multiple rounds of cleavage and ligation at the origin by what has been termed a "flip-flop" mechanism. The circular ssDNA is converted to dsDNA by the replication machinery of the host.

Some archaeal dsDNA viruses are also known to encode RCR-Rep homologs. Among those archaeal viruses are rudiviruses that all have a candidate RCR-Rep protein. The structure of the RCR-Rep protein (gp119) of *Sulfolobus islandicus* rod-shaped virus 1 (SIRV1) has been solved and it revealed similarity to the HUH endonuclease superfamily. DNA replication of another rudivirus, SIRV2, has been studied experimentally. These studies revealed that although SIRV2 RCR-Rep protein can both nick and seal DNA like canonical RCR-Rep proteins, the RCR replication mode cannot fully explain the experimental observations. Based on the observed multimeric and highly branched DNA replication intermediates, it was suggested that SIRV2 could employ a combination of strand-displacement, rolling-circle and strand-coupled genome replication mechanisms. Whether all of these mechanisms occur during each cycle of viral DNA replication remains unclear. In another archaeal virus, sphaerolipovirus SNJ1 (*Sphaerolipoviridae*), the RCR-Rep protein of the HUH endonuclease superfamily has been experimentally shown to be essential for

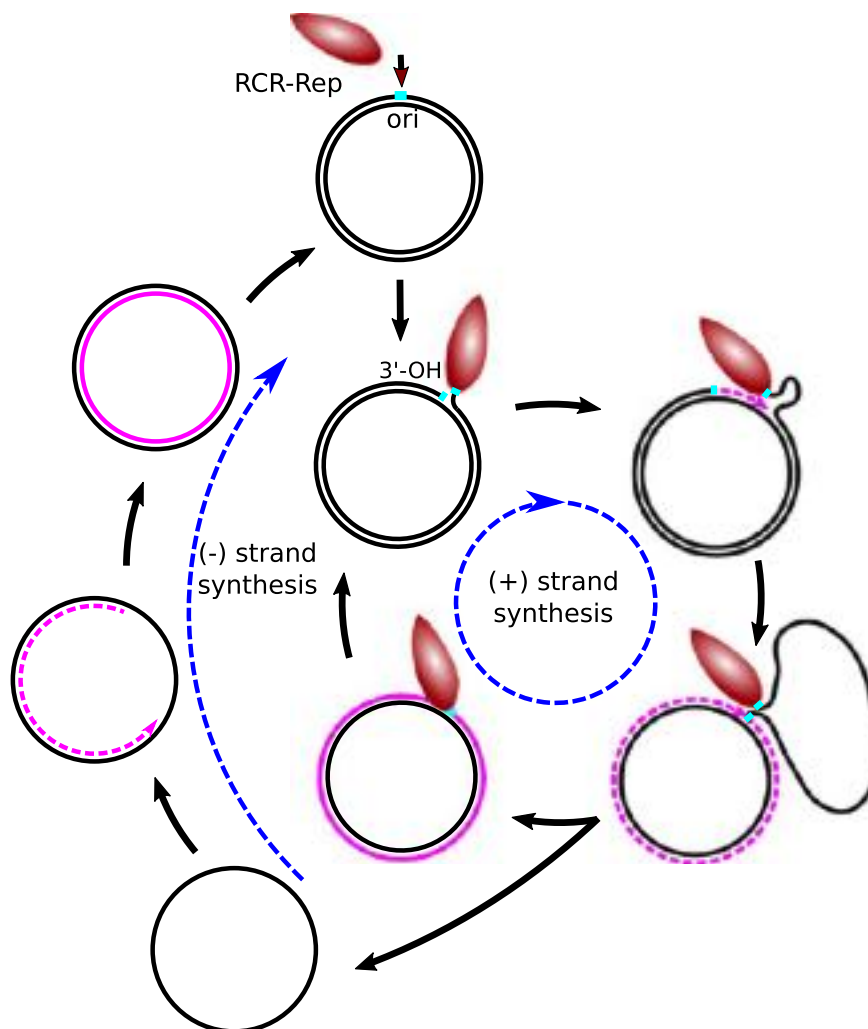


Fig. 5 Schematic illustration of the rolling-circle DNA replication based on the bacteriophage P2 replication mechanism. Phage RCR-Rep protein cuts the circular dsDNA at the replication origin (cyan) generating the covalent protein-DNA intermediate and a free 3'-OH group. Host's replicase utilizes the 3'-OH group as a primer and synthesizes the new (+) strand (magenta) while displacing the original strand. After the full circle is replicated, RCR-Rep nicks the newly generated origin at the same time sealing the displaced strand to produce a closed ssDNA circle. Host's replication machinery converts circular ssDNA to dsDNA by synthesizing complementary (-) strand (magenta). In the case of ssDNA viruses replicating through RCR, the latter process is the initial step in DNA replication.

viral genome replication. Furthermore, point mutations confirmed that the inferred catalytic residues of RCR-Rep are essential for its functionality. As pointed out above, the RCR mode of replication can be used by either dsDNA or ssDNA genomes. This is nicely illustrated in the case of haloarchaeal pleomorphic viruses. For example, *Haloarcula hispanica* pleomorphic virus 1 (HHPV-1) with circular dsDNA genome is closely related to *Halorubrum* pleomorphic virus 1 (HRPV-1) with circular ssDNA genome and both encode a homolog of RCR-Rep protein.

Since RCR is not the main replication mode in dsDNA prokaryotic viruses, but is prevalent in ssDNA viruses, the detailed discussion on the RCR mechanism and RCR-Rep proteins is provided in the section devoted to the replication of ssDNA viruses (see below).

Other DNA Replication Mechanisms

There are prokaryotic dsDNA viruses lacking any of the components indicative of a specific DNA replication system such as a replicative helicase (suggestive of RNA-primed DNA replication), a protein-primed DNA polymerase (protein-primed DNA replication) or an RCR-Rep homolog (suggestive of the RCR replication mode). In such cases dsDNA bacteriophages frequently encode either integrases, DNA recombination proteins or various replication initiation proteins, including homologs of bacterial replisome organizer DnaA and helicase loader DnaC. These observations suggest that in these cases viral genome is either

integrated into the host genome and propagated by the replication of the host genome or other proteins in the replication initiation stage are used to recruit cellular replication proteins. There are also cases when no typical DNA replication proteins were detected in the viral genomes. In particular this is true in the case of archaeal viruses. These viruses might either have highly diverged DNA replication proteins that cannot be detected by current computational approaches or use new replication strategies. The existence of the latter possibility is exemplified by *Acidianus* filamentous virus 1 (AFV1), which appears to use a novel strand displacement mechanism.

Genome Replication of Prokaryotic ssDNA Viruses

Viruses with ssDNA genomes are among the smallest viruses and encode as few as two proteins – one for capsid formation and the other one for genome replication. ssDNA bacteriophages have circular genomes and are classified into two families, *Microviridae* and *Inoviridae*. The *Microviridae* family comprises ssDNA phages with small icosahedral capsids including archetypal phage phiX174. Members of this family have circular ssDNA genomes in the 4.4–6.1 kb range. The *Inoviridae* family comprises filamentous ssDNA phages such as the M13 phage. Inoviruses also have circular ssDNA genomes and their size is in the 4.5–12.4 kb range. Officially recognized ssDNA bacteriophages represent only a small fraction in comparison with dsDNA bacteriophages. However, recent metagenomic studies found that ssDNA bacteriophages are present in various environments and that their abundance and diversity have been underestimated.

Archaeal viruses that have ssDNA genomes belong to two officially recognized families, *Pleolipoviridae* and *Spiraviridae*. Members of *Pleolipoviridae* similarly to the ssDNA bacteriophages have rather small ssDNA genomes ranging from 7 to 10.6 kb. Strikingly, the family *Pleolipoviridae* comprises viruses not only with ssDNA genomes, but also with either linear or circular dsDNA genomes. This observation complicates the traditional virus classification based on genome type. Not all archaeal ssDNA viruses are small. The sole member of the *Spiraviridae* family, *Aeropyrum* coil-shaped virus (ACV) infecting an extreme aerobic hyperthermophile (*A. pernix*) was found to have an unusually large (24.9 kb) ssDNA genome. So far, the ssDNA genome of ACV is the largest among all known ssDNA viruses.

Rolling Circle Replication

Most bacteriophages possessing ssDNA genome replicate using the RCR mechanism. In general, genome replication of ssDNA bacteriophages is similar to the RCR-using dsDNA viruses (see Fig. 5). The key difference is that in the case of ssDNA phages there is an additional stage, the conversion of ssDNA into dsDNA. The whole process can be generally divided into three stages as exemplified by the phage phiX174 replication. During the first stage, the infecting (+)ssDNA is converted by host replication proteins into a covalently closed dsDNA called replicative form DNA (RF). During the second stage, RF DNA is amplified. RCR-Rep (protein gpA) nicks the DNA at the origin of replication forming the covalent 5'-phosphotyrosine intermediate and a free 3'-OH. The 3'-OH serves as a primer for host DNA replicase which uses (–) strand as a template to synthesize DNA by “peeling off” the (+) strand. After one round of rolling circle synthesis, RCR-Rep cuts the newly generated origin and acts as a ligase, producing a covalently closed circular (+)ssDNA molecule. Newly generated (+)ssDNA genomes are again converted into dsDNA circular molecules. During the last stage of DNA synthesis, the ssDNA genome is concurrently synthesized and packaged into the viral procapsid.

The key role in ssDNA replication belongs to a multifunctional RCR-Rep protein. Early on it was noticed that RCR-Rep proteins are not monophyletic and that they often cluster with corresponding proteins from rolling-circle plasmids found in bacteria and archaea. Bacteriophages with ssDNA that replicate through the RCR mechanism use RCR-Rep proteins belonging to at least two evolutionarily and structurally distinct superfamilies, the HUH endonucleases and the TATA-box binding protein-like enzymes.

The superfamily of HUH endonucleases is named after the conserved metal binding HUH motif, consisting of two His residues separated by a bulky hydrophobic residue (U). In addition, HUH endonucleases have either one or two catalytic Tyr residues that form the covalent 5'-phosphotyrosine intermediate during nicking of ssDNA. The phage phiX174 RCR-Rep (protein gpA) represents the founding member of the RCR-Rep HUH endonucleases. It uses two catalytic Tyr residues to initiate and terminate rolling-circle replication and to spin off multiple circles of phiX174 ssDNA. Remarkably, despite decades of studies the structure of phiX174 gpA, the founding member of HUH endonuclease superfamily, remains unsolved. Peculiarly, the closest homolog with the solved three-dimensional structure appears to be a protein (PDB id: 2X3G) from dsDNA archaeal virus SIRV1. HUH endonucleases have the ferredoxin-like fold which is also known as the RNA recognition motif (RRM). HUH endonucleases also share similarity with origin binding proteins of small dsDNA viruses suggesting that there are intricate and ancient evolutionary relationships between these proteins.

RCR-Rep proteins of the second class, typified by GP2 of phage M13, belong to the Pfam family *Phage_CRI* (PF05144) and the related *Rep_trans* (PF02486) family of plasmid replication initiation proteins. Only recently the first three-dimensional structures of *Rep_trans* family members have been solved. These structures revealed that RCR-Rep proteins of *Rep_trans* family feature TATA-box binding protein-like (TBP-like) structural fold, which is entirely different from the RRM fold of HUH endonucleases. The metal binding site is also organized differently. *Rep_trans* representatives do not have the HUH motif. Instead, the metal binding is mediated by three acidic residues. However, despite structural differences and differently organized metal binding sites of the two types of RCR-Rep proteins, both types use an active site tyrosine residue as the nucleophile during the nicking/religation reactions. Moreover, the side chains of the corresponding catalytic tyrosine residues adopt a similar orientation with respect to the metal ions.

Therefore, these two different types of RCR-Rep proteins may represent a case of convergent evolution for catalyzing the same nicking and religation reactions. Interestingly, even within the same phage family RCR-Rep proteins may be of different types as exemplified by *Inoviridae*. Thus, the RCR-Rep protein from *Xanthomonas* phage Lf is a member of the HUH endonuclease superfamily, whereas the one from phage M13 belongs to the TBP-like superfamily.

Most ssDNA archaeal viruses, similarly to ssDNA bacteriophages, likely replicate their genomes via RCR or related rolling hairpin replication mechanisms as implied by the identified RCR-Rep homologs. A candidate replication initiator (Rep) family gene was identified in HRPV-1, a representative of the *Pleolipoviridae* family and the first characterized archaeal virus having ssDNA genome. The HRPV-1 Rep gene is homologous to RCR-Rep HUH endonucleases. On the other hand, no potential candidate showing significant sequence homology or sharing conserved motifs with known RCR-Rep proteins could be identified in ACV, a virus with the largest ssDNA genome, suggesting that it might use a unique mechanism of genome replication.

Genome Replication of Prokaryotic RNA Viruses

RNA viruses are divided into three distinct classes depending on the nature of their genome: positive-sense (+) single-stranded RNA (ssRNA) viruses, double-stranded (ds) RNA viruses, and negative-sense (–) ssRNA viruses. Although RNA viruses infecting eukaryotes are very common, the known representation of prokaryotic RNA viruses is quite narrow. Currently, there are only two recognized families of RNA bacteriophages, the *Leviviridae* family consisting of (+)ssRNA bacteriophages and the *Cystoviridae* family comprising segmented dsRNA bacteriophages. So far, no (–)ssRNA bacteriophages have been isolated. At present, there are no known RNA viruses that infect archaea. Although the putative presence of archaeal (+)ssRNA viruses has been reported, this finding remains unsubstantiated. Recent metagenomics surveys in diverse environments have substantially expanded known host range and spectrum of genome architectures of RNA bacteriophages. Nonetheless, it appears that the prokaryotic RNA virome is both significantly smaller and less diverse than the DNA virome.

The *Leviviridae* family comprises bacteriophages with (+)ssRNA genomes. The family includes four recognized species (Enterobacteria phage Q β , Enterobacteria phage F1, Enterobacteria phage MS2, and Enterobacteria phage BZ13). Bacteriophages of the *Leviviridae* family have a monopartite (+)ssRNA genome of 3.3–4.3 kb in size and are among the simplest and smallest of known viruses. Notably, the genome of phage MS2 was the first ever genome to be fully sequenced.

The *Cystoviridae* family for over two decades was represented by only one recognized virus species, *Pseudomonas* phage ϕ i6. Recently, a number of additional dsRNA phages have been isolated from various environmental samples and six of them (*Pseudomonas* phages ϕ i8, ϕ i12, ϕ i13, ϕ i2954, ϕ iNN and ϕ iYY) have been fully sequenced. All seven ICTV-recognized species of the *Cystoviridae* family have a dsRNA genome, which is divided into three separate segments: large (L), medium (M) and small (S). Individual segments range in size from 2.9 to 6.4 kb. The total genome size varies from 12.7 kb (ϕ i2954) to 15.0 kb (ϕ i8).

Replication Using an RNA-Dependent RNA Polymerase

Bacterial hosts lack the capability to synthesize complementary strands using the RNA strand as a template. Thus, all characterized RNA phages encode their own RNA-dependent RNA polymerase (RdRp). Notably, RdRp is the only universal protein not only in RNA phages, but in all known RNA viruses. In (+)ssRNA phages, the genome serves simultaneously as a genome template and as messenger RNA (mRNA). Since replication and translation run in the opposite directions, there is a competition between the two processes. Viral RdRp assembles with host proteins (ribosomal protein S1, translation elongation factors EF-Tu and EF-Ts) to form the active RNA polymerase holoenzyme (replicase). The replicase then initiates synthesis of the negative RNA strand by replicating through the (+)ssRNA genome. In turn, the newly synthesized (–)ssRNA strand is used to produce new (+)ssRNA genomes for the viral progeny.

In the case of dsRNA phages, RdRp (P2) is encoded in the largest genome fragment. The cystoviral RdRp first initiates synthesis of the plus-strands by unwinding each of the dsRNA segments and using the minus-strands as template. Newly synthesized (+)ssRNA segments (+S, +M, +L) are utilized as mRNAs by the host translational machinery. Once phage proteins are produced, the empty procapsids are self-assembled and the three (+)ssRNA segments are packaged into each of them. Following packaging, P2 initiates a single round of minus-strand RNA synthesis to recapitulate the dsRNA phage genome.

RdRps of (+)ssRNA and dsRNA phages as well as those of eukaryotic RNA viruses are evolutionary related. Viral RdRps belong to the class of right-hand polymerases that share the same RRM structural fold of their “palm” domain. This class also includes single-subunit DNA-dependent RNA polymerases, reverse transcriptases and DNA polymerases of A, B and Y families. Based on both sequence and structure analysis it was suggested that RdRps may represent the most ancient group of right-hand polymerases in agreement with the RNA World hypothesis, because RdRps could serve as replicases of RNA genomes. Phylogenetic analyses have further suggested that RdRps of dsRNA viruses evolved from (+)ssRNA viruses pointing to their ancient origin.

Concluding Remarks

Genome replication strategies employed by prokaryotic viruses are extremely diverse. This diversity in part can be explained by differences in type (DNA or RNA) and topology (linear or circular, double- or single-stranded, contiguous or segmented) of nucleic

acids encoding viral genomic information. However, even those prokaryotic viruses that, like cellular organisms, have dsDNA genomes, show remarkable diversity both in replication strategies and in composition of their DNA replication machineries. Interestingly, bacterial viruses often encode DNA replication proteins typical not of bacteria, but of archaea and eukaryotes, pointing to an ancient and complex evolutionary history of dsDNA replication systems of bacteriophages. Other viral DNA replication strategies such as protein-primed DNA replication and rolling circle replication evolved within a broader context of mobile genetic elements and apparently also have deep roots. Viral RNA-dependent RNA polymerases are believed to represent the most ancient group of right-handed polymerases in accord with the RNA world theory. Therefore, studies of viral proteins involved in genome replication might hold important clues for the emergence and the evolution not only of viral but also of cellular DNA replication machineries. A very fast pace at which genomic and metagenomic data for viruses are currently accumulating provides a solid basis for such studies.

Acknowledgments

The author thanks Darius Kazlauskas and Mart Krupovic for comments and suggestions. This work was in part supported by the Research Council of Lithuania [09.3.3-LMT-K-712-01-0080].

Further Reading

- Callanan, J., Stockdale, S.R., Shkoporov, A., *et al.*, 2018. RNA phage biology in a metagenomic era. *Viruses* 10.
- Carr, S.B., Phillips, S.E., Thomas, C.D., 2016. Structures of replication initiation proteins from staphylococcal antibiotic resistance plasmids reveal protein asymmetry and flexibility are necessary for replication. *Nucleic Acids Research* 44, 2417–2428.
- Černý, J., Černá Bolířková, B., De, A.Z.P.M., Grubhoffer, L., Růžek, D., 2015. A deep phylogeny of viral and cellular right-hand polymerases. *Infection Genetics and Evolution* 36, 275–286.
- Chandler, M., De La Cruz, F., Dyda, F., *et al.*, 2013. Breaking and joining single-stranded DNA: The HUH endonuclease superfamily. *Nature Reviews Microbiology* 11, 525–538.
- Dellas, N., Snyder, J.C., Bolduc, B., Young, M.J., 2014. Archaeal viruses: Diversity, replication, and structure. *Annual Review of Virology* 1, 399–426.
- Depamphilis, M., Bell, S., 2010. *Genome Duplication*. Taylor & Francis Group.
- Devoto, A.E., Santini, J.M., Olm, M.R., *et al.*, 2019. Megaphages infect *Prevotella* and variants are widespread in gut microbiomes. *Nature Microbiology* 4, 693–700.
- Kazlauskas, D., Krupovic, M., Venclovas, Č., 2016. The logic of DNA replication in double-stranded DNA viruses: Insights from global analysis of viral genomes. *Nucleic Acids Research* 44, 4551–4564.
- Koonin, E.V., Dolja, V.V., 2013. A virocentric perspective on the evolution of life. *Current Opinion in Virology* 3, 546–557.
- Kornberg, A., Baker, T.A., 2005. *DNA Replication*. University Science.
- Krupovic, M., 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Current Opinion in Virology* 3, 578–586.
- Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., Koonin, E.V., 2018. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Research* 244, 181–193.
- Krupovic, M., Forterre, P., 2015. Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. *Annals of the New York Academy of Sciences* 1341, 41–53.
- Malathi, V.G., Renuka Devi, P., 2019. ssDNA viruses: Key players in global virome. *Virusdisease* 30, 3–12.
- Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., *et al.*, 2016. Uncovering Earth's virome. *Nature* 536, 425–430.
- Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature Reviews Microbiology* 15, 724–739.
- Prangishvili, D., Koonin, E.V., Krupovic, M., 2013. Genomics and biology of Rudiviruses, a model for the study of virus-host interactions in Archaea. *Biochemical Society Transactions* 41, 443–450.
- Rumnieks, J., Tars, K., 2018. Protein-RNA interactions in the single-stranded RNA bacteriophages. *Subcellular Biochemistry* 88, 281–303.
- Salas, M., De Vega, M., 2016. Protein-primed replication of bacteriophage Phi29 DNA. *Enzymes* 39, 137–167.
- Székely, A.J., Breitbart, M., 2016. Single-stranded DNA phages: From early molecular biology tools to recent revolutions in environmental microbiology. *FEMS Microbiology Letters* 363.
- Wawrzyniak, P., Plucienniczak, G., Bartosik, D., 2017. The different faces of rolling-circle replication and its multifunctional initiator proteins. *Frontiers in Microbiology* 8, 2353.
- Weigel, C., Seitz, H., 2006. Bacteriophage replication modules. *FEMS Microbiology Reviews* 30, 321–381.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9.

Relevant Websites

- <http://ictv.global>
International Committee on Taxonomy of Viruses (ICTV).
- <https://www.ncbi.nlm.nih.gov/genome/viruses/>
NCBI Viral Genomes.
- <http://dmk-brain.ecn.uiowa.edu/VOG/>
The database of Prokaryotic Virus Orthologous Groups, or pVOGs.
- <https://viralzone.expasy.org/>
Viral Zone.
- <http://www.virology.ws/>
Virology blog.

Viral Transcription

David LV Bauer, Francis Crick Institute, London, United Kingdom
Ervin Fodor, University of Oxford, Oxford, United Kingdom

© 2021 Published by Elsevier Ltd.

Introduction

While viruses vary in their genome organization and genetic makeup, they all face the same fundamental task of producing a mRNA that is recognized by the host ribosome and translated to produce viral proteins. This process – viral transcription – varies depending on the virus family and lifecycle, though a number of general principles are common to nearly every virus. In the case of RNA viruses, transcription requires a virus-encoded enzyme, RNA-directed RNA polymerase (RdRp) that can use RNA, rather than DNA, as a template for mRNA synthesis.

General Principles of Viral Transcription

Degree of Dependence on Host Transcription Machinery

Two factors determine the extent to which viruses may rely on host machinery for transcription: their genome type (DNA vs. RNA) and, for viruses that infect eukaryotic hosts, the cellular compartment in which the viral genome is transcribed. DNA viruses that transcribe in the nucleus (e.g., most DNA viruses and reverse transcribing RNA viruses in their DNA provirus stage) can effectively mimic the host genome and utilize the host transcription machinery fully – including RNA Polymerase II (Pol II), which is the DNA-dependent RNA polymerase responsible for mRNA transcription in eukaryotes (Fig. 1(A)). DNA viruses that transcribe their genome in the cytoplasm (e.g., poxviruses) must encode their own RNA polymerase. In contrast, RNA viruses cannot rely on host Pol II, since Pol II is unable to use their RNA genomes as a template for transcription. Instead, all RNA viruses encode an RNA-dependent RNA polymerase (RdRp), which is used for both genome replication and transcription (Fig. 1(B)). RNA viruses that replicate in the nucleus, however, may still exploit other aspects of host RNA processing.

Features of Viral Transcripts at the 5' and 3' Ends Promote Translation

In eukaryotes, translation of mRNA is typically dependent on the presence of a 5'-cap and a 3' poly(A) tail. Viruses that depend on host Pol II for transcription (e.g., nuclear-transcribing DNA viruses) follow the canonical pathway to generate these structures: the 5'-cap is added by the host capping machinery following transcription initiation by Pol II, and the poly(A) tail is added by poly(A) polymerase following recognition of a poly(A) signal and cleavage of the nascent transcript by the host cleavage and polyadenylation specificity factor (CPSF).

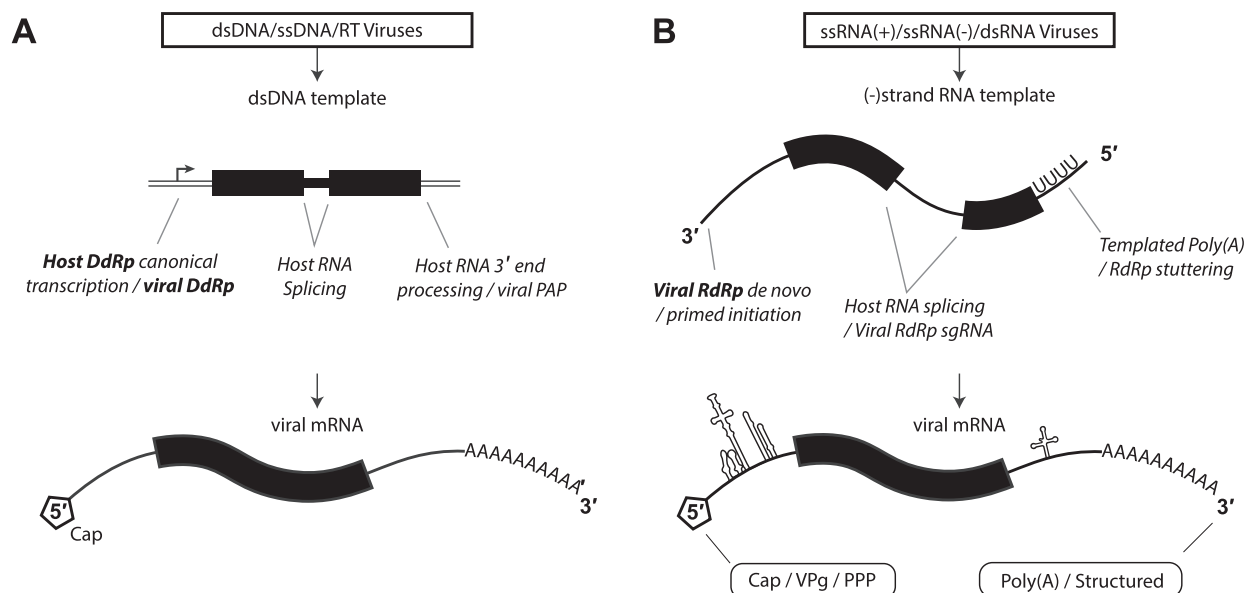


Fig. 1 Mechanisms of viral transcription in eukaryotic viruses that transcribe (A) from a dsDNA template, or (B) from a RNA template, as discussed in this text. Open reading frames (ORF) are represented as solid-filled shapes.

A broad range of viruses that do not depend on host Pol II, including cytoplasmic-transcribing DNA viruses (e.g., poxviruses) and the majority of RNA viruses, have adopted the strategy of effectively generating this canonical RNA structure on their own. Such a strategy involves either encoding enzymes to synthesize a 5'-cap *de novo*, or enzymes to cleave host mRNAs at their 5' ends in order to "snatch" 5'-caps for use in viral transcripts. Similarly, such viruses may generate a poly(A) tail by encoding their own poly(A) polymerase, by transcribing directly from a long poly(U) tract, or by stimulating the viral RdRp to "stutter" and repeatedly add non-templated A bases to the nascent RNA transcript.

RNA viruses that neither depend on host Pol II, nor encode their own capping or cap-snatching machinery, have adopted an alternate strategy in which the function of the canonical 5'-cap and 3' poly(A) tail are supplanted by virus-specific elements.

In the canonical host mRNA pathway, mRNA is exported from the nucleus and translation initiation is driven by recognition of the 5' cap by eukaryotic initiation factor (eIF) 4E, leading to assembly of the eukaryotic initiation complex and the ribosome upstream of the translation start codon. Some ssRNA(+) viruses contain a highly structured RNA motif at their 5' end of their genome, which acts to recruit the ribosome independently of a 5' cap. This RNA motif, termed an internal ribosome entry site (IRES), may interact with the eIF complex or the ribosomal subunits directly in order to drive translation of the viral RNA. Additionally, some ssRNA(+) viruses covalently link a viral protein to the 5' end of their genome (Viral Protein genome, VPg). While the VPg protein serves to protect the 5' end from degradation and to prime RNA synthesis by the RdRp, in some viruses (e.g., astroviruses and caliciviruses), it interacts directly with the eIF complex to promote translation.

The 3' end of mRNA also facilitates translation. In canonical translation of mRNA, the poly(A) tail is bound by poly(A) binding protein (PABP), which in turn binds the eIF complex that is bound to the 5'-cap of the same mRNA. The resulting circularization of RNA facilitates translation by promoting ribosome "recycling" following each round of translation. Some RNA viruses (e.g., flaviviruses) lack a poly(A) tail, but promote this process by encoding 3' RNA structures that bind PABP directly and also contain sequence complementary to the 5' region that promotes base-pairing between the two ends of the genome – though it is not clear whether these promote translation or, rather, replication by the RdRp. Some RNA viruses (most notably those that infect plants), contain a tRNA-like structure in their 3' untranslated region (UTR) that can be aminoacylated, promoting both RNA stability and translation. More broadly, the 3' UTR of mRNAs contain sequence elements that determine their stability, which are also exploited by viral RNAs.

Transcription Strategies Determine Mechanisms of Host Shutoff

Most viruses have adopted strategies to repress host gene transcription and translation, broadly termed "host shutoff", which serve to both conserve cellular resources for viral functions, and to prevent the host from expressing antiviral genes. The mechanisms of host shutoff that are available to a virus are determined by the degree to which it is dependent on host transcription. For example, viruses that generate transcripts with canonical host mRNA structures at their 5' and 3' end do not broadly inhibit canonical translation mechanisms – whereas viruses without these structures may target the components of the eIF complex or PABP. Similarly, viruses that do not rely on host transcription machinery, for example, may inhibit Pol II directly, or broadly target RNA splicing, 3' end processing, or mRNA export. Host shutoff does not have to be a binary process: specific viruses that use components of these pathways may modulate them so as to enhance their function, while also encoding mechanisms to block unused aspects of these pathways.

Overall Genome Organization Orchestrates Viral Transcriptional Program

Viral transcription takes place in the context of the lifecycle of a given virus: genes necessary early in infection are not necessarily needed late in infection – and vice versa. Viral genomes are therefore often arranged so that the order of transcription matches the temporal requirements for gene expression. This fundamental feature was first observed in bacteriophages, and is a common feature of many viruses that divide their genes approximately into those expressed early, middle, and late in infection. Broadly speaking, early genes are associated with establishing infection and suppressing host antiviral responses; middle genes are associated with genome replication or establishing latency; and late genes are associated with the production of the structural components of new virions.

Such transcriptional patterns can be achieved by diverse mechanisms: DNA viruses may encode motifs that are recognized by existing cellular transcription factors to activate or repress a given gene under certain conditions – or encode their own – to generate a transcriptional cascade. Similarly, they may also control the use of alternate splice or alternate poly(A) sites through the expression of viral proteins that change downstream expression as infection progresses. RNA viruses that exploit host RNA splicing factors may do the same. Some DNA viruses also encode their own viral DNA-dependent RNA polymerase in order to control the transcription of middle- and late-stage viral genes while removing their dependence on the host RNA polymerase – a notable feature of bacteriophage T7, for example.

RNA viruses follow a similar pattern, often arranging their genomes so that genes needed early in infection are transcribed (and translated) first. This arrangement, common in ssRNA(–) viruses, allows control of transcript abundance by regulating the proportion of RdRps that continue to transcribe from one block to another. This arrangement is also present to a certain extent in simple ssRNA(+) viruses for which viral genome transcription and replication are not differentiated: these viruses may produce a polyprotein that is cleaved into N-terminal proteins that are more typical of "early" genes and C-terminal proteins more typical of "middle" and "late" genes, though the relative levels of each protein cannot be controlled transcriptionally.

More complex ssRNA(+) viruses (e.g., coronaviruses) separate their “late” genes encoding viral structural proteins at the 3′ end of the viral genome, and require transcription by the viral RdRp to produce subgenomic RNA (sgRNA), directed by specific RNA motifs. Segmented RNA viruses follow a more complex pattern, often using RNA structures to control the level of transcription of each segment – indeed, the exploitation of the ability of RNA to form complex secondary and tertiary structures is exploited by nearly all RNA viruses to modulate the activity of the RdRp.

RNA Viruses Encode a RdRp that Transcribes and Replicates Their Genome

While DNA viruses may take advantage of existing host polymerases to replicate and transcribe their genes, RNA viruses must encode their own replicase and transcriptase since host cells do not encode enzymes that can generally use RNA as a template for RNA synthesis – i.e., they lack RNA-directed RNA polymerases (RdRps).

The simplest RNA viruses, such as picorna- and flaviviruses, take the form of an infectious mRNA that encodes the necessary proteins to replicate and transmit its genomic material from an infected cell to a new host cell. In these viruses, the function of the RdRp is to replicate the viral genome, and there is no distinct “transcription” step. As discussed above, specialization of RNA synthesis for mRNA production vs. genome replication permits control over protein levels at the level of transcripts.

While more complex ssRNA(+) viruses carry out transcription to generate sgRNAs, transcription is still not required to express genes at the 5′ end of the genome. In contrast, all dsRNA and ssRNA(−) viruses must first carry out a distinct transcription step to generate (+)-sense RNA that is translated into viral proteins. As a result, RdRps are packaged into virions in order to initiate transcription (and thus infection) in new host cells. In all cases, the RdRp must also be able to shift between transcription and replication activities, which are often mechanistically distinct.

RdRps Contain an Evolutionarily Conserved Architecture

All viral RdRps share common structural motifs that are characteristic of template-dependent nucleic acid polymerases, including an overall domain architecture that resembles a cupped right hand, composed of a “thumb” and “fingers” that bind RNA and a “palm” domain that contains the highly-conserved polymerase active site. At a sequence level, this active site (comprised of three distinct, conserved amino acid motifs) and four additional motifs located outside of the active site show a high degree of sequence conservation.

Advances over the past fifteen years in protein production (notably eukaryotic cell culture) and protein structural characterization (most recently, cryoEM) has led to an abundance of polymerase and RdRp structures. Analysis of these structures has shown that extensive *structural* conservation exists that extends beyond the core regions of sequence alignment alone, and that such conservation extends to other classes of related polymerases, for example RNA-dependent DNA polymerases (RdDps – i.e., reverse transcriptases) from group II introns/retrotransposons and retroviruses (e.g., HIV-1).

The advent of high-throughput sequencing, virus surveillance, and broad metagenomic sampling has led to an explosion of the number of sequences arising from RNA viruses – many of which remain uncultured, uncharacterised, and formally unclassified. Such data has been used – albeit controversially – to carry out large-scale sequence alignment and phylogeny analysis in order to reconstruct the evolutionary history of RNA viruses via the conserved RdRp. This analysis supported the hypothesis that the first RNA viruses were ssRNA(+) viruses that evolved from an ancestral retrotransposon when its reverse transcriptase acquired the ability to synthesize RNA directly, thus removing the need to replicate via a DNA intermediate. The phylogenetic analysis also suggests that dsRNA viruses evolved independently twice from ssRNA(+) viruses, with further evolution of one group of dsRNA viruses giving rise to ssRNA(−) viruses.

From the perspective of viral transcription, two observations can be made if the broad groupings of viruses hold true, irrespective of the precise finer-level tree structures: first, that a “transcription-first” lifecycle has emerged independently at least twice, in the form of dsRNA viruses that package their viral RdRp and undergo an obligatory transcription step to generate mRNA for translation. Second, that RdRps which carry out specialized transcription functions, distinct from genome replication, are widely distributed across ssRNA(+) virus branches. Such a distribution would suggest that specialized transcription (vs. replication) is intrinsic to viral RdRps – either in its ability to evolve independently, or as a common function of ancestral RdRps that was lost in virus families that do not utilize this step.

Different RdRp and Template Conformations Distinguish Transcription from Replication

The initiation of RNA synthesis *de novo* is typically rate-limiting for templated nucleic acid polymerases, since it requires binding a single initiating NTP and holding it in the active site until a second NTP may be bound and polymerized into a nascent dinucleotide transcript. In order to initiate transcription, RdRps may either prime transcription using viral or host-derived RNA fragments, or may attempt to facilitate *de novo* synthesis themselves.

There are distinct requirements for either strategy: RdRps that use a primer to initiate transcription must be able to accommodate double-stranded primer:template RNA within the core of the polymerase during initiation – whether RNA or RNA-protein (U-VPg) – and typically have a thumb and fingers arrangement to accommodate this. They may also position residues at the core of the RdRp to facilitate correct primer alignment (or re-alignment) against the RNA template. In contrast, RdRps that initiate transcription *de novo* usually contain a “priming loop” that is inserted into the active site and stabilises the template: incoming

NTP, often by providing an aromatic residue against which the NTP may stack. Intriguingly, while this approach of *de novo* initiation is conserved, the actual priming loop structures themselves are not: RdRp priming loops may be attached to the thumb in some ssRNA(+) and ssRNA(−) viruses, or the palm in some dsRNA viruses. This situation implies convergent evolution gave rise to these structures – a scenario that is mirrored in a comparison between transcription initiation by prokaryotic and eukaryotic DdRps: the finger domain of the *E. coli* RNA polymerase transcription factor σ^{70} and the B reader loop of eukaryotic RNA Polymerase II transcription factor II B (TFIIB) perform the same function, but are positioned overall in different orientations.

In segmented ssRNA(−) viruses, both primer-dependent and *de novo* initiation are employed by the RdRp, and used to distinguish between transcription (using host-derived primers containing a 5' cap) and viral genome replication (initiated *de novo* with the assistance of a priming loop). This dual function requires structural flexibility in the RdRp in order to accommodate both functions – reflected in the considerable rearrangements observed in polymerase and accessory domains between transcriptase and replicase structures.

Conformational flexibility is also required of RdRps from ssRNA(+) viruses that carry out transcription through the generation of sgRNAs, during which the RdRp must pause transcription, accommodate and hold the nascent transcript while disengaging from the template and re-engaging downstream to resume transcription. This process, often considered a form of intramolecular RNA recombination, can also be imagined as a form of primer-dependent (re-)initiation of transcription downstream of the original transcription start site.

In all cases, multiple factors exist that promote the RdRp to “switch” between replication and transcription functions – a process which can be RNA-directed, protein-directed, or a combination of the two. In the case of ssRNA(+) viruses, RNA structure and sequence complementarity drive sgRNA generation, while in segmented ssRNA(−) viruses, a distinct RNA structure in the RdRp-bound promoter of the genome's complementary (+) sense replication intermediate alters RdRp conformation to prevent transcription and promote replication.

Protein factors that regulate RdRp behavior can be virus- or host-encoded, and are fundamentally linked to the progression and regulation of the viral lifecycle. In alphaviruses, a ssRNA(+) virus, for example, proteolytic processing of the polyprotein cofactor of RdRp triggers transcription. In VSV, a non-segmented ssRNA(−) virus, the RdRp initially binds to host elongation factor 1- α (EF1- α), leading to transcription of viral proteins. Once expressed, viral N protein displaces EF1- α in order to generate a replication-competent RdRp complex. Similarly, in influenza A virus, a segmented ssRNA(−) virus, the RdRp initially binds to host Pol II, locking the RdRp in its transcription-competent conformation. Subsequently, RdRp binding to a second newly-expressed RdRp in complex with a host protein (ANP32) drives formation of a replication-competent RdRp conformation.

RdRps Can Coordinate Transcription, Capping, and Polyadenylation

The adoption of a specific transcription-competent conformation by RdRps is coupled with rearrangements that promote core transcriptional behaviors, and, for viruses that require them, is also coupled with the addition of 5' caps and 3' poly(A) tails to viral mRNA. Core changes in RdRp activity often include fundamental changes to properties of polymerization, such as processivity. These changes can act to promote internal transcription initiation and termination on non-segmented ssRNA(−) genomes, to permit RNA editing by the RdRp, and to favor RdRp pausing and “jumping” to generate sgRNAs on ssRNA(+) genomes.

In eukaryotes, the addition of a 5'-cap to nascent mRNAs is physically coupled to Pol II transcription – a process that has convergently evolved in many RNA viruses. Viruses that encode their own capping machinery may either recruit viral capping enzymes to the RdRp via specific interactions, or encode the capping machinery and the RdRp as a single polypeptide. Such an arrangement (Fig. 2) is present in the RdRps of some ssRNA(+) and ssRNA(−) viruses that synthesize their caps *de novo* using

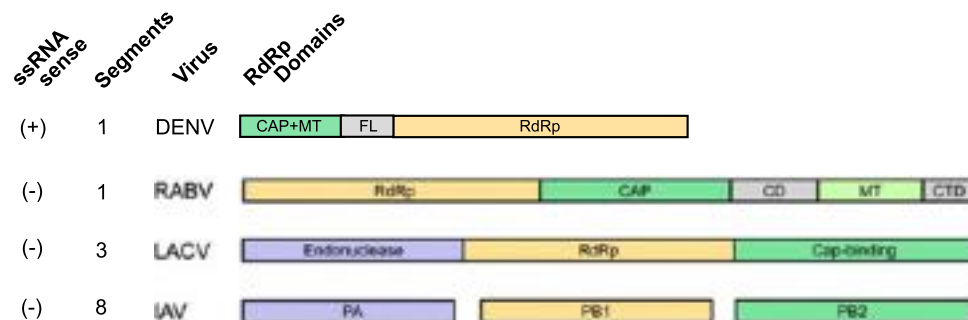


Fig. 2 Domain structures of cap-synthesizing and cap-snatching polymerases of positive-strand and negative-strand RNA viruses. Positive-strand viruses such as dengue virus (Flaviviridae, DENV), encode single-polypeptide cap-synthesizing polymerases, which are composed of a combined capping (CAP/PNRTase) and methyltransferase (MT) domain, a flexible linker (FL) that recruits additional viral proteins, and a RNA-dependent RNA polymerase (RdRp) domain. Nonsegmented negative-strand RNA viruses such as rabies virus (Rhabdoviridae, RABV) encode single-polypeptide cap-synthesizing polymerases, which are composed of the RNA-dependent RNA polymerase (RdRp) domain, capping (CAP / PNRTase) domain, connector domain (CD), methyltransferase (MT) domain, and C-terminal domain (CTD). Segmented negative-strand RNA viruses such as La Crosse virus (LACV) and influenza A virus (IAV) encode single-polypeptide (LACV) or multisubunit (IAV) cap-snatching polymerases with RdRp, cap-binding, and endonuclease activities. The IAV polymerase is composed of polymerase acidic (PA) (endonuclease), polymerase basic 1 (PB1) (RdRp), and polymerase basic 2 (PB2) (cap-binding) subunits.

polyribonucleotidyl transferase (PRNTase) and methyltransferase (MTase) activities, as well as in those ssRNA(–) viruses that instead acquire already-capped RNA fragments from the 5′ end of host RNAs to use as primers for viral transcription via a capping domain and an endonuclease domain – a process referred to as “cap-snatching”.

Recent advances in protein structure determination have shown that in some viruses, capping and transcription are so closely coupled that a single channel between the RdRp active site and the capping active site is generated – again, an arrangement adopted by the (structurally-unrelated) host Pol II and capping machinery. A closely-coordinated mechanism is also used by cap-snatching RdRps, which must position their newly-acquired 5′-capped primers precisely to the RdRp active site following endonuclease cleavage.

Polyadenylation can also be achieved by multiple mechanisms. In some ssRNA(+) viruses, the poly(A) tail can be fully encoded in genome sequence itself, whereas in others the mechanism appears to be more elaborate, with separate proteins possessing intrinsic polyadenylation capabilities. In human coronavirus 229E, for example, nsp8 has 3′-terminal adenylyl-transferase (TATase) activity – given that nsp7 & nsp8 complex with the RdRp (nsp12) to form a replicase, these activities are likely coupled. In ssRNA(–) viruses, however, the RdRp synthesizes the poly(A) tail itself by “stuttering” on a U-rich tract at the end of each gene. For segmented ssRNA(–) viruses such as influenza A virus, this process is facilitated by RNA/RdRp contacts: as the RdRp approaches the 5′ end of the template during transcription, the 5′-terminal promoter remains tightly bound to the RdRp, sterically preventing transcription beyond the poly(U) tract that is positioned in the RdRp active site.

Prospective

As we have discussed here, viruses adopt a wide range of transcription strategies to ensure efficient expression of their genes. The exact strategy employed by a virus depends in part on its lifecycle and how it interacts with its host – a process that is not strictly one-directional, since both viruses and hosts have co-existed and co-evolved. Indeed, it is tempting to speculate that transcription strategies available to a given virus may be a direct product of its evolutionary history. Poxviruses, for example, contain large genomes that encode for many genes and functions also found in host cells (e.g., DdRp, PAP), leading some to suggest that poxviruses originate from an intracellular eukaryotic parasite that lost its own protein-production capacity – while others have suggested that the ancestral poxvirus evolved to efficiently carry out horizontal gene transfer and simply accumulated a large genome as a result. Such evolution might also apply at the level of individual genomic elements: for example, it has been posited that RNA motifs such as IRESes are remnants of a (pre-transcriptional) “RNA world”, rather than a more recently-evolved mechanism to subvert canonical transcription and translation.

The theme of viral subversion is recurrent: while DNA virus transcription broadly resembles host transcription, RNA viruses (which by their very existence subvert the central dogma of molecular biology) adopt a diverse set of approaches to generate mRNA. As is so often the case in virology, there are further “exceptions to these exceptions”: Hepatitis delta virus (a circular ssRNA (–) satellite virus that depends on co-infection with hepatitis B virus) uses host Pol II (normally a DdRp) to replicate rather than an RdRp; Infectious bursal disease virus (IBDV, a dsRNA virus) has a rearranged order of RdRp catalytic motifs, which are otherwise absolutely conserved.

Our understanding of viruses and transcription has increased considerably in recent years, driven in a large part by the advent of high-throughput sequencing and powerful structural imaging that offer unprecedented insights into the atomic details of macromolecular transcription complexes and their mechanisms. Concurrently, our understanding of host cell transcription and gene expression has increased, leading to the discovery of host gene-encoded IRESes and 3′-end structures to enhance expression, as well as RNA editing and slippage behaviors by host RNA polymerases. Such findings suggest that the diverse mechanisms of viral transcription discussed here are not quite as unique as previously imagined.

Further Reading

- Choi, K.H., 2012. Viral polymerases. *Advances in Experimental Medicine and Biology* 726, 267–304.
- Coffin, J.M., Hughes, S.H., Varmus, H.E., 1997. *Transcription*. Cold Spring Harbor Laboratory Press.
- Holmes, E.C., 2013. Virus evolution. In: Knipe, D.M., Howley, P. (Eds.), *Fields Virology*. Philadelphia, UNITED STATES: Wolters Kluwer Health, pp. 286–313.
- Holmes, E.C., Duchêne, S., 2019. Can sequence phylogenies safely infer the origin of the global virome? *mBio* 10 (2).
- Ogino, T., Green, T.J., 2019. RNA synthesis and capping by non-segmented negative strand RNA viral polymerases: Lessons from a prototypic virus. *Frontiers in Microbiology* 10, 1490.
- Picard-Jean, F., Tremblay-Létourneau, M., Serra, E., *et al.*, 2013. RNA 5′-end maturation: A crucial step in the replication of viral genomes. In: Romanowski, V. (Ed.), *Current Issues in Molecular Virology*. Rijeka: IntechOpen.
- Sainsbury, S., Niesser, J., Cramer, P., 2013. Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* 493 (7432), 437–440.
- Sola, I., Almazán, F., Zúñiga, S., Enjuanes, L., 2015. Continuous and discontinuous RNA synthesis in coronaviruses. *Annual Review of Virology* 2 (1), 265–288.
- te Velhuis, A.J.W., 2014. Common and unique features of viral RNA-dependent polymerases. *Cellular and Molecular Life Sciences* 71 (22), 4403–4420.
- Venkataraman, S., Prasad, B.V.L.S., Selvarajan, R., 2018. RNA dependent RNA polymerases: Insights from structure, function and evolution. *Viruses* 10 (2).
- Wandzik, J.M., Kouba, T., Karuppasamy, M., *et al.*, 2020. A structure-based model for the complete transcription cycle of influenza polymerase. *Cell* 181 (4), 877–893.
- Werner, F., Grohmann, D., 2011. Evolution of multisubunit RNA polymerases in the three domains of life. *Nature Reviews Microbiology* 9 (2), 85–98.
- Whelan, S., 2013. Viral replication strategies. In: Knipe, D.M., Howley, P. (Eds.), *Fields Virology*. Philadelphia, United States: Wolters Kluwer Health, pp. 105–126.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9 (6).
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2019. Reply to Holmes and Duchêne, “Can sequence phylogenies safely infer the origin of the global virome?”: Deep phylogenetic analysis of RNA viruses is highly challenging but not meaningless. *mBio* 10 (2).

Translation of Viral Proteins

Martin D Ryan and Garry A Luke, University of St. Andrews, St. Andrews, United Kingdom

© 2021 Published by Elsevier Ltd.

Nomenclature

eEF Eukaryotic elongation factor.

eIF Eukaryotic initiation factor.

eRF Eukaryotic release (termination) factor.

kb Kilobases

kDa Kilo Daltons (molecular mass)

Introduction

Viruses are obligate intracellular parasites and are absolutely dependent upon the protein synthesis apparatus of the host-cell. Upon infection and the initiation of the virus replication cycle, the biogenesis of virus proteins requires the sequestration of the infected-cell resources (translational apparatus, aminoacyl-tRNAs, rNTPs) - resulting in a 'competition' for these resources between viral and host-cell mRNAs. Viruses have evolved many, diverse, mechanisms to supplant the 'canonical' translation of host-cell proteins (see below), including mRNA processing/export from the nucleus, mRNA stability and primarily the initiation, but also the elongation and termination stages of translation. Dominance of the virus mRNAs in this competition serves to both maximize virus particle production and to inhibit the innate immune system: secretion of cytokines from the infected-cell signals surrounding, uninfected, cells to enter the 'antiviral state' and recruits immune cells to sites of infection. Virus genomes have also evolved to encode (1) *cis*-acting RNA elements which confer a 7-methylguanosine (m^7G) cap-independent mode of translation, (2) translational enhancers, (3) elements that increase the coding capacity of their genomes, or, (4) to regulate virus protein biogenesis. Translation imposes a high bioenergetic cost upon cells and cellular stress (e.g., metabolite starvation, hypoxia, endoplasmic reticulum (ER) cargo overload, toxins and virus infection) has led to the cell evolving mechanisms to reduce canonical translational activity during stress. Proteins required during stress or involved in stress recovery may be preferentially translated *via* non-canonical translational mechanisms: there are clear parallels between the strategies that host-cells and viruses employ to shut-down canonical host-cell mRNA translation and the use of alternative, non-canonical, translational mechanisms. In this chapter we will provide an over-view of the wide range of mechanisms viruses employ to radically alter the translational 'landscape' of the infected cell to maximize the biogenesis of virus replication and encapsidation proteins.

Canonical or Cap-Dependent mRNA Translation

To discuss the various mechanisms by which viruses subvert the host-cell translational apparatus, it is worthwhile outlining the main features of host-cell mRNA canonical translation. This refers to the process by which the large majority of cellular mRNAs, which comprise a m^7G 'cap' structure at the 5' terminus and a poly(A) tail at the 3' terminus, are translated. The first stage in translation is initiation - a key stage in which the cell responds to various forms of stress to reduce translational activity.

The Initiation Phase

The mRNA 5' cap structure is bound by eukaryotic initiation factor eIF4E. This complex binds to eIF4G - referred-to as a 'scaffolding' protein. This new complex, in turn, binds eIF4A (an RNA helicase). eIF4F is the term for the complex comprising eIF4A, eIF4E, and eIF4G. eIF4B (4A co-factor) is then bound. eIF4G also binds poly(A) binding protein (PABP), bound to the 3' poly(A) mRNA tail, leading to the circularization of the mRNA such that the sites of initiation and termination become proximal. The activity of eIF4E is regulated by phosphorylation at Ser²⁰⁹ by the MAPK-interacting serine/threonine-protein kinase (MNK). The association of MNK with eIF4G is essential for eIF4E phosphorylation. This complex of mRNA and initiation factors then binds the 43S preinitiation complex comprising the small ribosomal subunit (40S) bound to the initiation factors eIF3, eIF1, eIF1A, eIF5 and the ternary complex [eIF2•Met•tRNA_i•GTP] ([Fig. 1, Panel A](#)).

Protein Kinase R (PKR)

Initiation is regulated by the cell at two key stages, the phosphorylation of eIF2 α and the sequestration of eIF4E by eIF4E binding proteins (see below). The GDP-bound form of eIF2 cannot bind Met-tRNA_i and is converted by eIF2B (a guanine exchange factor) to the GTP-bound form, which now binds Met-tRNA_i. Phosphorylation of the alpha subunit of eIF2 (eIF2 α) converts it from a substrate for eIF2B to a competitive inhibitor. Since eIF2B is only present in limiting amounts (10–20 fold less than eIF2), proportionally small changes in eIF2 α phosphorylation have a large effect on initiation: the phosphorylation status of eIF2 α is a key regulator of initiation. eIF2 α is phosphorylated by four different cellular kinases, notably here protein kinase R (PKR or eIF2 α K2). This kinase is activated by double-stranded RNA (dsRNA) - a pathogen associated pattern (PAMP) produced during viral infection. Viruses have evolved proteins/protein domains which may either inhibit PKR activation or dimerization, degrade PKR, sequester dsRNA, synthesize PKR pseudo-

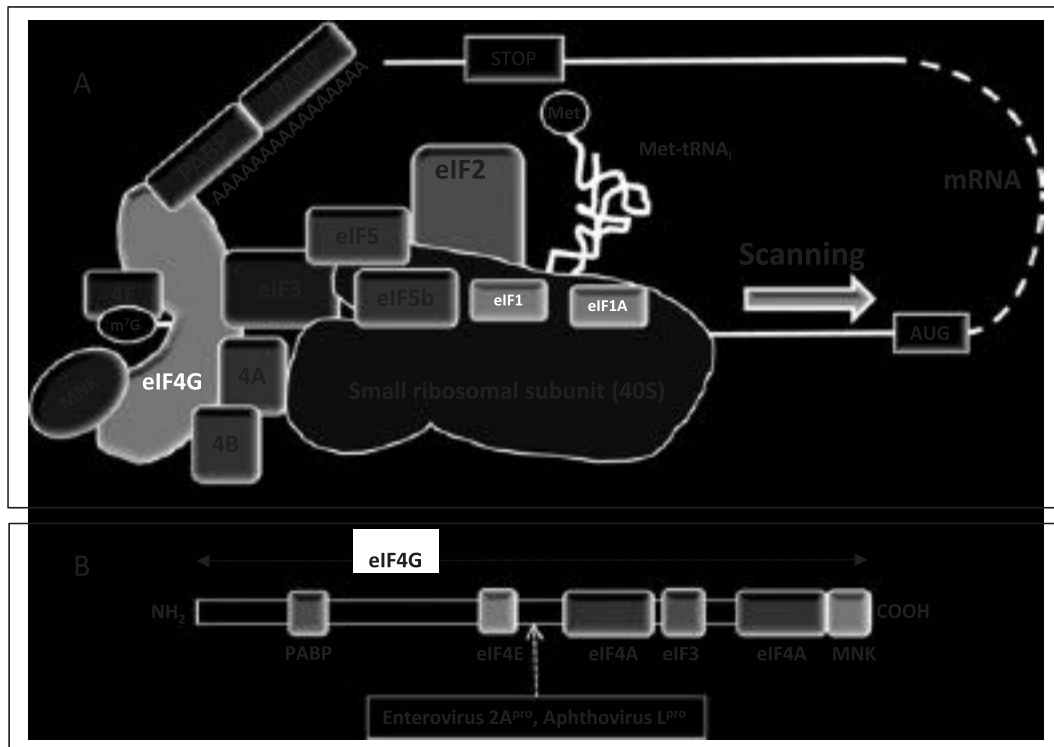


Fig. 1 Schematic showing the association of the various eukaryotic initiation factors (eIFs) in association with the mRNA and the 40S ribosomal small subunit – comprising canonical initiation of translation (Panel A). The binding sites of poly(A) binding protein (PABP), eIFs and MNK on the ‘scaffolding’ protein eIF4G are shown, together with the site where the picornavirus proteinases cleave eIF4G (Panel B).

substrates or activate antagonist phosphatases (reviewed by [García *et al.*, 2006](#); [Langland *et al.*, 2006](#)). Conversely, in the case of the positive (+ve) sense, single-stranded (ss) RNA alphaviruses Sindbis and Semliki Forest viruses, eIF2 is highly phosphorylated within infected cells, shutting off cellular mRNA translation - although virus mRNA is translated. RNA structures in the 5' region of the sub-genomic mRNA are involved in conferring eIF2-independent translation within infected mammalian cells, whereas mutation of RNA structures in the long 3' UTR had little effect upon translation of the genomic and sub-genomic RNAs in mammalian cells, but was highly suppressed in mosquito cells (reviewed by [Carrasco *et al.*, 2018](#)).

Initiation factor eIF4E

A second mechanism of regulation of initiation is the sequestration of eIF4E by eIF4E binding proteins (eIF4E-BPs). The binding of eIF4E-BPs to eIF4E does not inhibit eIF4E binding mRNA cap structures, but inhibits the binding of the mRNA-bound eIF4E to eIF4G: the binding of eIF4G and eIF4E-BP to eIF4E is mutually exclusive. The eIF4E binding activity of eIF4E-BPs is regulated by phosphorylation: in their hyperphosphorylated form eIF4E-BPs cannot bind eIF4E. eIF4E-BPs become hyperphosphorylated at multiple sites following treatment with a variety of extracellular stimuli. Both eIF4E and eIF4E-BPs are targets for viruses. The 100K adenovirus viral protein blocks MNK1 phosphorylation of eIF4E, directly regulating eIF4E activity ([Cuesta *et al.*, 2000](#)). Infection by the picornavirus Encephalomyocarditis Virus (EMCV; cardiovirus A) leads to dephosphorylation of eIF4E-BP1 - which then sequesters eIF4E ([Gingras *et al.*, 1996](#)). Infection by another picornavirus, enterovirus 71 (EV-A71), induces the synthesis of microRNA 141 (miRNA-141) which inhibits the expression of eIF4E ([Ho *et al.*, 2011](#)). Indeed, eIF4E can be replaced or displaced by virus proteins to promote their translation: the arenavirus Junin virus (JUNV) N protein replaces eIF4E, the lentivirus human immunodeficiency virus 1 (HIV-1) promotes the binding of the dead box helicase DDX3 to the eIF4F complex thereby displacing eIF4E, whilst the RING finger Z protein of another arenavirus, lymphocytic choriomeningitis virus (LCMV), directly binds to and blocks eIF4E function (reviewed by [Montero *et al.*, 2015](#)).

Scanning and assembly of the 80S ribosome

The preinitiation complex then scans from the 5' position along the mRNA 5' untranslated region (5' UTR) until encountering the first initiation (AUG) codon. RNA secondary structures present in the 5' UTR are unwound by eIF4A/eIF4B. The mRNA channel is kept in an open conformation by eIF1/eIF1A which also plays an important role in recognition of the AUG initiation codon and flanking nucleotides. The complex pauses at the AUG and eIF5 triggers hydrolysis of GTP bound to eIF2 and the 40S subunit adopts a closed conformation: the ribosomal A-site is empty with Met-tRNA_i occupying the P-site. The large 60S ribosomal subunit associates (catalyzed by eIF5B/GTP) to form the 80S ribosome, this assembly process displacing factors eIF2•GDP, eIF5, eIF3 and eIF1.

Elongation and Termination Phases

Translation now enters the second phase - elongation. Elongation factor eEF1A•GTP-bound aminoacyl-tRNAs enter the vacant A-site where the cognate aminoacyl-tRNA: mRNA codon interaction binds the correct aminoacyl-tRNA into the A-site. GTP is hydrolyzed and eEF1A•GDP released. The peptide bond is formed and translocation occurs: (1) the peptidyl-tRNA (now in the A-site) is translocated from the A- to P-site, concomitant with (2) the deacylated tRNA translocated from the P- to E-site, and, (3) the ribosome translocated one codon along the mRNA. The A-site is now vacant allowing ingress of the next (cognate) aminoacyl-tRNA. Translocation is catalyzed by eEF2•GTP: GTP is hydrolyzed and, following translocation, eEF2•GDP is released. The activity of eEF2 is regulated by a specific calcium/calmodulin dependent kinase, eEF2 kinase (eEF2-K). The activity of eEF2-K is regulated by a series of stress signaling pathways: increased stress signaling produces increased eEF2 phosphorylation and leads to reduced ribosome processivity. The ribosome continues the codon by codon progression through the mRNA open reading frame (ORF) until a termination codon (UAA, UAG or UGA) enters the A-site. In this third phase of translation, all three stop codons are recognized by eukaryotic release factor eRF1 which uses an activated water molecule to hydrolyze the ester linkage between the (completed) peptide and tRNA, located in the P-site. The nascent protein is released with the release of eRF1 being accelerated by eRF3•GTP.

Canonical (Cap-Dependent) Initiation of Virus mRNA Translation

In the case of viruses whose mRNAs do utilize 5'-end cap-dependent (canonical) translation, this may be accomplished simply by the transcription of individual genes (DNA viruses), or, in the case of certain RNA viruses the presence of internal promoters/terminators allows the transcription by the RNA-dependent RNA polymerase (RdRp) of monocistronic mRNAs encoded by individual genes - typified by paramyxoviruses and pneumoviruses (-ve sense, ssRNA genomes). Another strategy is encoding different proteins on separate (generally monocistronic) genomic RNA segments - typified by orthomyxoviruses (-ve sense, ssRNA genomes), reoviruses and rotaviruses (dsRNA genomes).

Cellular mRNAs are processed within the nucleus (capping/splicing/polyadenylation). Viruses that replicate within the cytoplasm cannot utilize this cellular capping activity. Such viruses have evolved various strategies to generate m⁷G capped mRNAs: their polymerases comprise domains or subunits with capping activities such as guanylyltransferase (GTase)/RNA guanine-N⁷-methyltransferase (N⁷MTase)/nucleoside 2'-O-methyltransferase (2'OMTase)/RNA: GDP polyribonucleotidyltransferase (PRNTase).

In the case of certain -ve sense ssRNA viruses (e.g., influenza, lassa, hantaa and rift-valley fever viruses) replication occurs within the nucleus and viral mRNAs become capped by a 'cap-snatching' (or 'cap-stealing') mechanism in which the virus polymerase binds to the m⁷G structure of nascent cellular mRNAs, cleaves the mRNA some 10–20 nucleotides downstream, then uses this capped mRNA tract as a primer for transcription of virus mRNAs. Arenaviruses, which replicate within the cytoplasm, are also thought to cap-snatch, structural studies suggesting that the N-terminal domain of the L polymerase comprises a functional cap-snatching endonuclease (Rosenthal *et al.*, 2017).

Virus mRNAs may be polyadenylated by mechanisms such as re-iterative stuttering of the viral polymerase on a poly(U) tract (e.g., influenza, paramyxoviruses), or, by the virus encoding a protein with poly(A) polymerase activity (e.g., DNA pox viruses). In the case of rotaviruses (dsRNA genomes; 11 segments), virus mRNAs from each segment are capped, but not polyadenylated. Rotavirus non-structural protein 3 (NSP3) is, however, a functional analog of cellular PABP. Rotavirus mRNAs have a 3' UTR consensus sequence which is bound by NSP3. The C-terminal domain of this protein binds to eIF4G, competing with PABP for binding to the same site (Fig. 1, Panel B). NSP3, therefore, circularizes rotavirus mRNA translational complexes. However, NSP3 also binds the cellular protein RoXan. The PABP displaced by NSP3 from translational complexes accumulates in the nucleus, a process which requires NSP3-RoXan interaction and, as infection proceeds, increasingly inhibits the translation of host-cell mRNAs (Harb *et al.*, 2008).

Given that (1) both virus and host-cell mRNAs conform to cap-dependent (canonical) initiation of translation, and (2) there is competition between the two types of mRNA for translational resources, how does translation of the former supplant the latter? Naturally, the cap-snatching and PABP displacement mechanisms outlined above contribute to host-cell shut-off, but preferential degradation of cellular mRNA (Beloso *et al.*, 1992), preferential translation of viral mRNA (Park and Katze, 1995), degradation of cellular RNA polymerase II (RNA pol II; Rodriguez *et al.*, 2007), inhibition of the generation cellular pre-mRNA (Nemeroff *et al.*, 1998) and the retention ('imprisonment') of host-cell mRNA in the nucleus (see below) have been proposed (reviewed by Mohr, 2016). In addition, Herpes Simplex Virus 1 (HSV-1) protein ICP27 highly stimulates the export of intronless virus mRNAs: both ICP27 and Human Cytomegalovirus (HCMV) protein UL69 bind directly to viral mRNA and either the export factors Aly/REF or UAP56, consequently directing the export of virus mRNAs (Koffa *et al.*, 2001; Toth and Stamminger, 2008).

In some cases the pool of mRNA inside cells is remodeled: the genomes of poxviruses, herpesviruses and coronaviruses encode proteins that stimulate mRNA decay. The poxvirus vaccinia virus (VACV; dsDNA) encodes proteins which target all stages in cellular protein biogenesis: host-cell mRNA transcription, mRNA processing (splicing/polyadenylation) and mRNA degradation (reviewed by Walsh, 2017). With regards the latter, can virus proteins involved in mRNA degradation differentiate between virus and host-cell mRNAs? VACV expresses two de-capping enzymes (D9 and D10) that remove the cap from both cellular and viral mRNAs (Shors *et al.*, 1999; Parrish and Moss, 2007). As the replication cycle progresses, increasing the relative abundance of virus mRNAs, this lack of discrimination increasingly favors the translation of viral mRNA. Furthermore, degrading virus mRNAs serves

to promote the control of gene expression during the progressive change in transcription from early to intermediate to late virus genes. VACV infection also leads to changes in both the post-translational modification and sub-cellular distribution of certain translation initiation factors, conferring advantages to the translation of VACV over host-cell mRNAs (reviewed by [Dhungel et al., 2020](#)). Notably, RNA viruses have smaller genomes with a much smaller coding capacity. However, in the case of influenza A virus (IAV) it has been shown that the virus-encoded RNA endoribonuclease PA-X plays a major role in the shut-off of host-cell translation ([Hayashi et al., 2015](#)) – probably by stimulating degradation of RNA pol II transcribed (host-cell) mRNA ([Khapersky et al., 2016](#)). In the case of coronaviruses (+ve sense, ssRNA), non-structural protein nsp1 binds to the 40S ribosome subunit of all active translational complexes inducing cleavage of mRNAs – thought to be by activation of a cellular mRNA surveillance pathway (reviewed by [Abernathy and Glaunsinger, 2015](#)).

‘Imprisonment’ of Cellular mRNAs Within the Nucleus

Inhibiting the nuclear mRNA export pathway to bring about the shut-off of translation of cellular mRNAs is a strategy adopted by a wide range of viruses which may use either canonical and non-canonical modes of initiation. DNA viruses replicate within the nucleus (except pox-, asfar- and phycoviruses) and encode genes without introns: virus mRNAs are, therefore, not spliced. Herpes virus protein ICP27 inhibits cellular mRNA splicing, hijacks the RNA export factor REF plus the export receptor NXF1 thereby facilitating the export of virus mRNAs from the nucleus. Other DNA viruses exploit components of the export pathway to promote export of virus mRNAs whilst inhibiting export of cellular mRNAs (reviewed by [Kuss et al., 2013](#)). The strategy of mRNA imprisonment is also used by a wide range of RNA viruses replicating in either the nucleus or the cytoplasm. In the case of the former, the NS1 protein of Influenza A virus was shown to block cellular mRNA export from the nucleus ([Fortes et al., 1994](#)). NS1 affects multiple stages in cellular mRNA processing: NS1 binds to poly(A)-binding protein nuclear 1 (PABPN1) leading to the accumulation of mRNAs with short poly(A) tails which are not exported to the cytoplasm. The rhabdovirus Vesicular Stomatitis Virus (VSV; -ve sense RNA) replicates within the cytoplasm. The VSV matrix M protein bears nuclear localization signals: it is transported into the nucleus where it binds the mRNA export factor Rae1/Mmp41, blocking mRNA export by the disruption of Rae1 function ([Faria et al., 2005](#)). The nsp1 β protein of the arterivirus porcine reproductive and respiratory syndrome virus (PRRSV; +ve sense, ssRNA) brings about the retention of cellular mRNAs within the nucleus and the shut-off of host-cell translation. Mutations within the SAP (SAF-A/B, Acinus and PIAS) motif of nsp1 β destroyed this phenotype ([Han et al., 2017](#)). The L protein of the picornavirus Theiler's Murine Encephalomyelitis Virus (TMEV; cardiiovirus B) imprisons cellular mRNAs by bringing about the hyperphosphorylation of nucleoporins 62 and 98 (Nup62, Nup98; [Delhaye et al., 2004](#); [Ricour et al., 2009](#)). Imprisonment is also brought about by the L protein of the related EMCV, which interacts with Ran GTPase and brings about the hyperphosphorylation of Nup 62, Nup 153 and Nup 214 ([Lidsky et al., 2006](#); [Porter et al., 2006](#); [Bardina et al., 2009](#)). In the case of viruses within the enterovirus genus of the picornaviruses, the polyprotein comprises two proteinase domains – the 2A proteinase (2A^{pro}) and 3CD proteinase (3CD^{pro}). Both proteinases cleave components of the nuclear pore complex: 2A^{pro} degrades Nup358 and Nup214, whilst 3CD^{pro} degrades Nup62 (reviewed by [Flather and Semler, 2015](#)). Since the virus mRNAs discussed above use canonical initiation, the number of targets to suppress the translation of host-cell mRNAs is constrained. It can be seen, however, that viruses have evolved a surprisingly wide range of mechanisms to inhibit manifold steps in the biogenesis, nuclear export and translation of host-cell mRNAs.

Non-Canonical (Cap-Independent) Initiation of Virus mRNA Translation

Internal Ribosome Entry Sites (IRESes)

Although first discovered in picornaviruses, IRESes have subsequently been characterized in other virus families and certain cellular mRNAs. IRESes are RNA secondary structural features which bind ribosomes not at the mRNA 5' cap, but at an internal site within the 5' UTR. Cellular IRESes predominantly occur in the mRNAs of proteins that are important for cell proliferation, differentiation, regulating apoptosis or responding to stress conditions - notably when cap-dependent translation is repressed. The RNA secondary structures of virus IRESes are formed by intra-strand base-pairing within single, contiguous, tracts of RNA - some hundreds of nucleotides long. They are well studied and bioinformatics (RNA folding *in silico*) together with biochemical RNA structural probing studies and mutagenesis reveal four distinct structural classes (see below). In contrast, IRESes in the 5'-UTRs of certain cellular mRNAs are much more heterogeneous: RNA structures may be formed from non-contiguous tracts of RNA and do not fall into clear structural categories as is the case for virus IRESes. By directly interacting with host translational components, virus IRESes by-pass the canonical method of ribosome recruitment: this creates more targets for the virus to suppress canonical initiation and also by-passes many of the constraints of host-cell regulation. This confers a highly effective advantage to virus mRNAs, allowing the complete shut-off of host-cell mRNA translation.

Type I IRESes ([Fig. 2](#)) are encoded by the enterovirus genus of picornaviruses – typified by poliovirus ([Pelletier and Sonenberg, 1988](#)). The 2A proteinase of enteroviruses (2A^{pro}) cleaves eIF4G ([Fig. 1](#), Panel B) such that the eIF4E and eIF4A binding sites are now on separate cleavage products – shutting off canonical initiation. Interestingly, quite a different type of proteinase, the L proteinase (L^{pro}) encoded by the aphthovirus genus of picornaviruses (typified by Foot-and-Mouth Disease Virus; FMDV), cleaves eIF4G at a site highly proximal to enterovirus 2A^{pro} ([Fig. 1](#), Panel B; [Devaney et al., 1988](#)). Type I IRESes interact with the C-terminal cleavage product of eIF4G and require eIF5B, eIF2/Met-tRNA_i, with activity being stimulated by eIF1, eIF1A and eIF4B. IRES *trans*-acting factors (ITAFs) are RNA-binding proteins that alter IRES conformation and promote binding to eIF4G. Ribosome recruited by type I IRESes scan through a ‘spacer’ region in the 5' UTR (~160nts) until the initiating AUG is encountered.

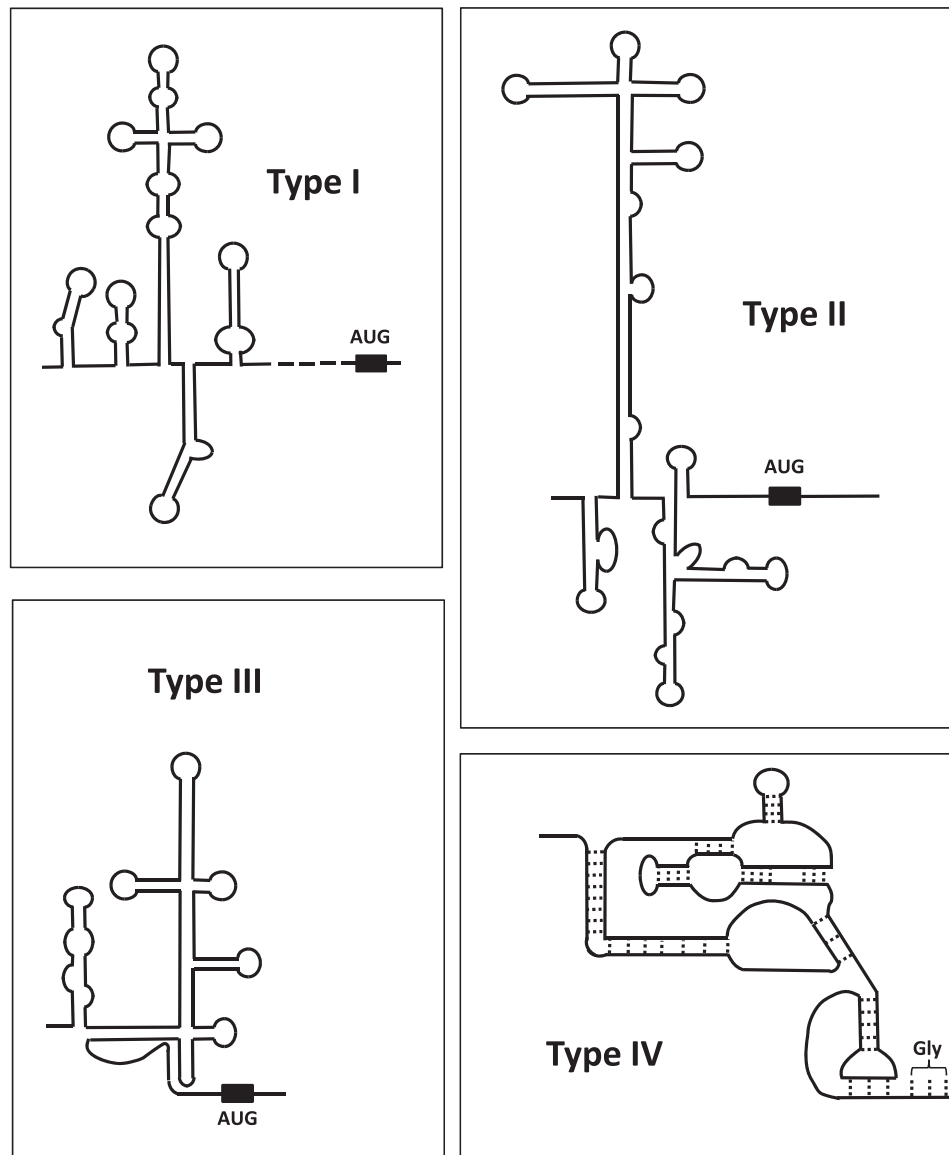


Fig. 2 The RNA structures of the different types of internal ribosome entry sites (IRESes) are shown together with the site of initiation of translation (black boxes). In the case of the type IV IRES, a CCU sequence (part of pseudoknot 1; PK1) is located in the ribosome P-site and acts as a molecular mimic of tRNA, whilst a GGC codon (encoding glycine) is located in the ribosome A-site.

Type II IRESes are encoded by the cardiovirus genus of picornaviruses - typified by EMCV (Jang *et al.*, 1988). In this case eIF4G is not cleaved: like type I IRESes the translation factors/ITAFs mentioned above are required. Interestingly in another picornavirus genus – the hepatoviruses (typified by hepatitis A virus; HAV), IRES activity in hepatoma cell extracts is higher in comparison to extracts derived from a non-hepatic line (Sadahiro *et al.*, 2018). Ribosomes recruited by type II IRESes bind proximal to the initiating AUG and little, if any, scanning takes place.

Type III IRESes encoded by pestiviruses (typified by Hepatitis C Virus; HCV) interact with both eIF3 and the 40S ribosome thereby by-passing eIF4F altogether (Pestova *et al.*, 1998): a process analogous to ribosome binding by Shine–Dalgarno sequences in bacteria. Generally, type III IRESs require eIF3, eIF5B and the eIF2/Met-tRNA_i ternary complex. Here the initiating AUG codon is directly placed into the ribosomal P-site with no scanning (reviewed by Lukavsky, 2009).

Type IV IRESes encoded by the *Dicistroviridae* (typified by Cricket Paralysis Virus; CrPV) are located in an intergenic region (IGR) between OR1 (non-structural polyprotein) and ORF2 (structural polyprotein), referred-to as an IGR-IRES. Here, direct interaction with the 40S subunit occurs removing the necessity of canonical initiation factors (Wilson *et al.*, 2000; Jan and Sarnow, 2002). This binding leads to changes in the conformation of the 40S subunit such that the 60S subunit associates to form 80S ribosomes. A CCU sequence (part of pseudoknot 1; PK1) is located in the ribosome P-site – acting as a molecular mimic of tRNA, whilst a GGC codon (encoding glycine) is located in the ribosome A-site (Butcher and Jan, 2016). When cognate Gly-tRNA•GTP•eEF1A enters the A-site, base-pairing occurs, then a ‘pseudo-translocation’ of the Gly-tRNA into the P-site allows the elongation phase to

proceed: the N-terminal residue of the nascent ORF2 polypeptide being glycine rather than methionine. In a single additional base-pairing within PK1, the first guanine of the GGC codon is sequestered and a GCG (encoding alanine) occupies the A-site: pseudo-translocation occurs as before and now a new ORF (ORF_x) is translated with an N-terminal alanine (Ren *et al.*, 2012).

In contrast to the complex RNA secondary structures outlined above, simpler structures may also promote cap-independent translation. The 5' UTR leader of Tobacco Etch Virus (TEV; potyvirus) is 143 nucleotides long comprising two domains each forming an RNA pseudoknot (PK1 and PK2; Gallie, 2001), although bioinformatic analyses suggest the presence of a third pseudoknot (Domashevskiy and Cheng, 2015). PK1 binds eIF4G, promoting cap-independent translation. Complementarity was noted between the loop 3 of PK1 and a region within 18S ribosomal RNA (rRNA). Mutations introduced within loop 3 (designed to disrupt potential base-pairing with 18S rRNA) reduced cap-independent translation, whereas mutations which preserved base-pairing had minimal effects (Zeenko and Gallie, 2005). Together, these data indicate the TEV 5' UTR promotes cap-independent translation by providing a functional substitute for a 5'-cap structure.

Cap-Independent Translation Enhancers (CITEs)

The genomes of many +ve-stranded RNA plant viruses comprise neither an m⁷G cap structure at their 5' termini nor a poly(A) tail at their 3' termini. Here, unusual mechanisms have evolved to recruit ribosomes to the 5'-end of the genomes and compete with host-cell mRNAs. One strategy employed by tombus- and luteoviruses involves a range of RNA structures sited within, or proximal to, the 3' UTR - termed 3' cap-independent translation enhancers (3' CITEs), greatly enhancing the initiation of translation. Initially characterized in Satellite Tobacco Necrosis Virus (STNV), a 93 bp tract forms a simple RNA structure - a long stem-loop termed a translation enhancer domain (TED; Fig. 3; Danthinne *et al.*, 1993). Other RNA structures are more complex but together fall into six major types, each with individual properties (Fig. 3). All types appear, however, to follow the same mechanism. Firstly, the 3' CITE binds to components of the translation machinery (e.g., eIF4E, eIF4G or 60S subunit: type dependent) and secondly, the RNA structures in the 3' CITE base-pair to elements within the virus 5' UTR - concomitantly translocating these translation components to the 5' UTR (reviewed by Nicholson and White, 2011; Simon and Miller, 2013).

Virus Alternatives to Components of Initiation

An Alternative to the m⁷G Cap Structure: Virus Protein, Genome Linked (VPg)

Many +ve stranded RNA viruses (both animal and plant) do not bear a m⁷G cap, but have either an oligopeptide or a protein covalently bound to the 5' terminus - VPg. In the case of picornaviruses VPg is ~22 amino acids long and acts as a primer for RNA synthesis, but is not essential for viral RNA (vRNA) infectivity (reviewed by Goodfellow, 2011). In the case of caliciviruses, VPg is much larger (13–15 kDa) and removal of VPg from calicivirus RNA decreases infectivity and translation in cell-free systems. Calicivirus VPg interacts with eIF4E, but does not inhibit neither eIF4E-BP1 binding to eIF4E, nor m⁷G caps binding to eIF4E, indicating a unique VPg binding site for eIF4E. It appears, therefore, calicivirus VPg functions as a m⁷G cap 'substitute' in the initiation of calicivirus translation (Goodfellow *et al.*, 2005; reviewed by Goodfellow, 2011). In the case of plant potyviruses, VPg is some 20–22 kDa, but in members of the genus Sobemovirus VPg is smaller (9–13 kDa). Potyvirus VPg binds eIF4E: however, in plants two isoforms of eIF4E are present - eIF4E and eIF(iso)4E. In general, potyviruses utilize only one specific isoform to initiate

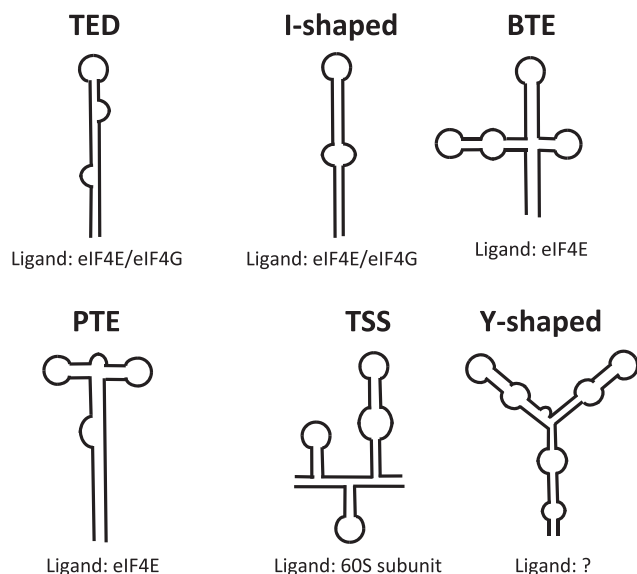


Fig. 3 The RNA structures of the different types of cap-independent translation enhancers (CITEs) are shown; translation enhancer domain (TED), barley yellow dwarf virus (BYDV) BYDV-like translation element (BTE), panicum mosaic virus (PMV) PMV-like translation element (PTE), T-shaped structure (TSS) plus the I- and Y-shaped structures.

translation, but some potyviruses can utilize both isoforms. Trimeric complexes of potyvirus VPg-eIF4E-eIF4G are detected suggesting that these plant virus VPgs, like calicivirus VPg, act as m⁷G cap substitute (reviewed by Jiang and Laliberté, 2011).

An alternative to eIF4F

The expression of various reporter genes was enhanced by the bunyavirus nucleocapsid (N) protein. Further work showed: (1) high affinity binding of N to the 5' cap of cellular mRNAs - replacing eIF4E, (2) N binds directly to 43S pre-initiation complexes - functionally replacing eIF4G and (3) N obviates the requirement for the helicase, eIF4A. Bunyavirus N protein replaces, therefore, the cellular eIF4F complex (Mir and Panganiban, 2008).

Ribosome 'Shunting' (Discontinuous Scanning)

First characterized in the plant pararetrovirus Cauliflower mosaic virus (CaMV; Caulimoviridae), ribosome shunting occurs in the 5'-leader region of CaMV pregenomic RNA (pgRNA). In the current model of ribosome shunting, a 43S preinitiation complex binds at the cap structure of pgRNA, scans through the 5'-leader until the initiation codon of a short upstream ORF (uORF) is encountered: the 60S subunit associates and elongation commences. Translation terminates and the 40S subunit shunts over a long RNA secondary structural feature, then 'lands' in a region (characteristically U/A rich) immediately downstream of the RNA structure. Translation then reinitiates at the start codon of the downstream ORF (Fig. 4). Ribosome shunting is observed in all

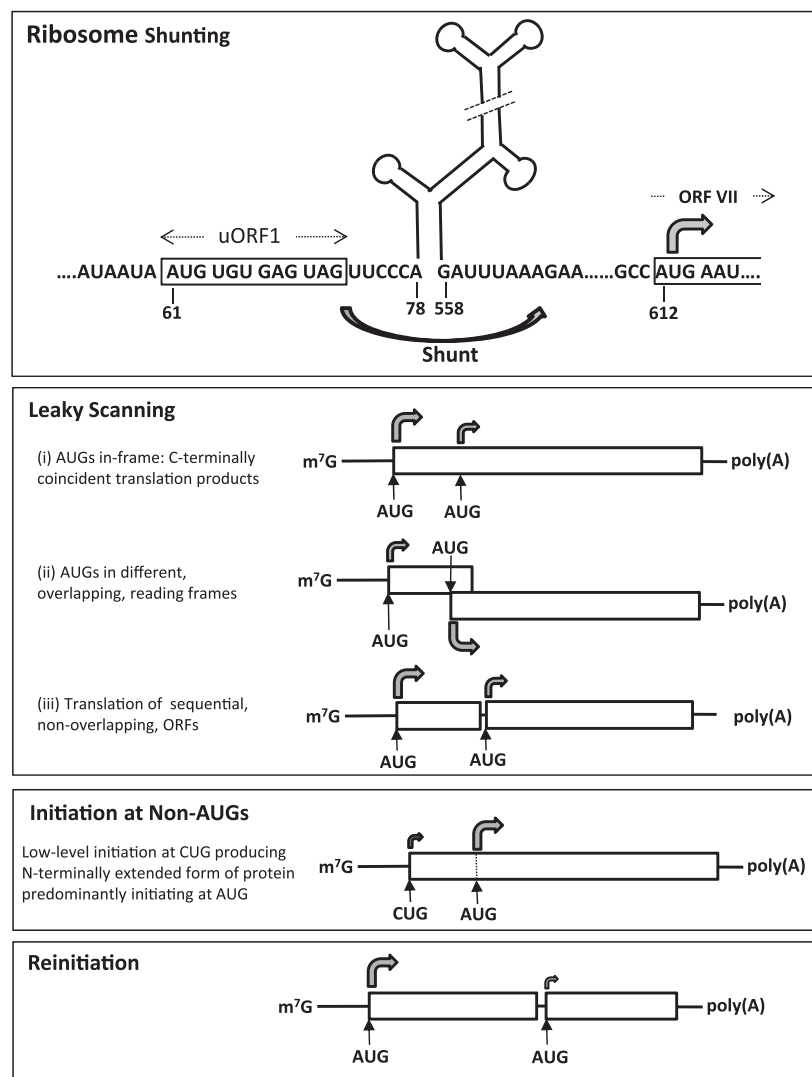


Fig. 4 Various forms of alternative initiation are shown. Ribosome shunting in which following termination at the short upstream ORF, the 40S subunit shunts over a long RNA secondary structural feature, then 'lands' downstream. Leaky scanning in which proportion of ribosomes do not initiate at the first AUG but continue to scan for the next AUG downstream. Initiation at non-AUG codons may occur at near-cognate codons (e.g., ACG and CUG), recognized by Met-tRNA_i. Reinitiation may occur if a very short ORF (normally <30 codons) is translated and the 40S subunit does not dissociate and resumes scanning.

genera of the Caulimoviridae, but also in the RNA picorna-like virus Rice tungro spherical virus (RTSV; Secoviridae) and certain animal viruses (reviewed by Firth and Brierley, 2012; Pooggin and Ryabova, 2018).

Leaky Scanning

In the 'classical' model of initiation outlined above, the preinitiation complex scans from the 5' cap along the 5' UTR until encountering the first AUG. However, the efficiency of initiation at a particular AUG is affected by the surrounding context, notably bases at the +4 and -3 positions ($X^{-3}nn \underline{A}^{+1}\underline{U}^{+2}\underline{G}^{+3}X^{+4}$; potential initiation codon underlined), with a purine at -3 and a G at +4 being a 'strong Kozak' context. Sub-optimal contexts produce a significant proportion of ribosomes that do not initiate at the first AUG but continue to scan for the next AUG downstream (Kozak, 1986). This is termed 'leaky scanning' and produces a series of potential outcomes. Firstly, if the next AUG initiation codon is in the same reading frame as the first, then initiation at both AUGs produces both a full-length and N-terminally truncated form of the same protein: C-terminally coincident. Secondly, the second AUG may be in a different, but overlapping, reading frame to the first AUG – thereby producing two distinct proteins from the same coding region. Thirdly, translation of sequential, non-overlapping, ORFs (Fig. 4). Leaky scanning is used by a wide range of RNA viruses – both +ve and -ve sense, plant and animal viruses – to produce multiple proteins from a single message. Besides the context proximal to the first AUG, a number of other factors promote leaky scanning (reviewed by Firth and Brierley, 2012).

Initiation at Non-AUGs

The elongation phase commences predominantly, but not exclusively, at AUG codons. Under some circumstances near-cognate codons (e.g., ACG and CUG) can be recognized by Met-tRNA_i. Such recognition normally requires a strong Kozak context and is enhanced by an RNA secondary structural feature (e.g., stem-loop) forming ~14 nts downstream of the initiation codon such that (1) the structure is located at the entrance to the mRNA entry channel and (2) the near-cognate initiation codon is located in the ribosomal P-site (Kozak, 1990). The group of near-cognate codons comprise CUG (most efficient), GUG, ACG, AUU, AUA, AUC and UUG with initiation levels between 2%–30%, creating an appreciable diversity of protein isoforms (reviewed by Touriol *et al.*, 2003). Since initiation at non-AUG codons is relatively inefficient, ribosomes predominantly scan through a non-AUG codon and initiate either at a downstream AUG or near-cognate codon. Thus, RNA viruses may translate single, or multiple, forms of N-terminally extended protein isoforms, or, multiple different proteins from alternative reading frames (Fig. 4). For example, in the paramyxovirus murine respirovirus (formerly Sendai Virus; SeV) a C' protein (215 amino acids) is translated from an in-frame, upstream, ACG codon whilst the C protein (204 amino acids) is translated from a downstream AUG codon. Indeed, the SeV C ORF produces a nested set of four co-C-terminal proteins, C', C, Y1, and Y2, all of which use the same termination codon (UAA; 728), but arise from four different ribosomal initiation sites, C' = ACG (nt 81), and C = AUG (nt 114), Y1 = AUG (nt 183), and Y2 = AUG (nt 201) (Takeuchi *et al.*, 2008).

Ribosome Reinitiation

Following termination, the ribosomal subunits dissociate from one another and the mRNA. If a very short ORF (normally <30 codons) is translated, however, the 40S subunit may not dissociate from the mRNA and resume scanning. If a downstream initiation codon is encountered reinitiation may occur (Fig. 4). This is thought to depend upon certain initiation factors remaining associated with the ribosome whilst the short ORF is translated. Immediately following termination, the 40S subunit cannot reinitiate and during this new scanning phase must re-acquire specific initiation factors – notably the eIF2•Met•tRNA_i•GTP ternary complex to become competent to reinitiate (reviewed by Jackson *et al.*, 2012). Calicivirus genomes comprise a long ORF1, encoding replication proteins, but transcribe a sub-genomic RNA (sgRNA) which encodes ORF2, the major capsid protein, and ORF3 which encodes a small basic protein essential for infectivity. ORFs 2 and 3 overlap: the ORF2 stop codon is proximal to the ORF3 AUG codon, the length of the overlap being species dependent (1–14 nts): in some cases actually overlapping UAA**AUG** or AUG**A** (stop codons underlined, start codons in bold). ORF3 is translated by reinitiation following ORF2 translation. NOTE: this form of reinitiation is quite distinct from reinitiation following translation of an upstream short uORF outlined above. It has been shown that the calicivirus termination-reinitiation process relies upon mRNA motifs normally 40–49nts upstream of the ORF2 termination codon termed the TURBS (termination upstream ribosome-binding site; Meyers, 2003; Luttermann and Meyers, 2007; Pöyry *et al.*, 2007; Napthine *et al.*, 2009). In the case of Feline Calicivirus (FCV) the efficiency of ORF3 translation is also modulated by RNA sequences downstream of the reinitiation site (Habeta *et al.*, 2014). Reinitiation is observed in a wide range of viruses: the orthomyxovirus Influenza Virus B (IVB; segment 7 – M1 and BM2 proteins; Horvath *et al.*, 1990), the pneumovirus Human Respiratory Syncytial Virus (HRSV; M2-1 and M2-2 proteins; Ahmadian *et al.*, 2000), the totivirus Victorivirus (Gag and Pol proteins; Huang and Ghabrial, 1996), the hypoviruses Cryphonectria Hypoviruses 1 and 2 (CHV-1, CHV2; ORFA and ORFB proteins; Shapira *et al.*, 1991), and cauliflower mosaic virus (CaMV; multiple proteins; Fütterer and Hohn, 1991). Detailed discussions of the mechanisms of virus reinitiation are given in Firth and Brierley, 2012 and Gunišová *et al.*, 2018.

Translational 'Recoding': Non-Canonical Elongation and Termination

Translational 'recoding' is taken here to mean the observed translation products do not correspond with predictions made by simple inspection of the nucleotide sequence of the ORF. In general, cellular canonical initiation of translation only allows the synthesis of

one protein from a specific mRNA. Viruses must, however, translate multiple, different, types of protein (encapsidation functions, replication functions etc.) – often from a single ORF – to maximize coding capacity. Furthermore, the various mechanisms involved in translational recoding often provide sophisticated controls over protein biogenesis during the infectious cycle.

Ribosomal Bypassing ('Hopping')

Normally co-linearity exists between the mRNA coding sequences and the translated protein sequence. In the case of bacteriophage T4 *gene 60* (encoding a subunit of the phage topoisomerase), however, ribosomes translate the first 46 codons, bypass (do not translate) the next 50 nucleotides, then resume translation (Huang *et al.*, 1988). Whilst ~100% of ribosomes appear to commence bypassing, only ~50% resume translation of the second ORF. A diverse range of elements contribute to bypassing: (1) the codons at the 'take-off' and 'landing' sites must be identical (GGA), (2) a stop codon (UAG) must be present immediately 3' of the take-off codon, (3) an RNA stem-loop structure formed by the 3' end of ORF1 and the 5' region of the intervening non-coding sequence (Fig. 5) and (4) a cis-acting peptide signal within the nascent ORF1 peptide - positioned within, and thought to interact with, the ribosome exit tunnel. Additional RNA stem-loop structures are formed by sequences encoding the C-terminal region of ORF1 and those encoding the N-terminal region of ORF2 (for simplicity not shown in Fig. 5 - see Agirrezabala *et al.*, 2017). Recent, detailed, analyses of reconstituted translation systems suggested rather than hopping, the ribosome 'slides' over the gap in a rapid and uniformly processive manner – consistent with the observation that elongation factor G (EF-G; the bacterial equivalent to eukaryotic eEF2, mentioned above) promotes bypassing at the expense of GTP hydrolysis (Klimova *et al.*, 2017). Structural analyses showed that during this process the peptidyl-tRNA^{Gly} anticodon remains base-paired with the mRNA codon - the nascent peptide/ribosome exit tunnel interaction anchoring peptidyl-tRNA^{Gly} in the P-site and 'locking' the peptidyl transferase center (PTC) into an inactive conformation, together with a rotation of the 30S small subunit opening the A-site. The conformation of the PTC and the mRNA secondary structure inhibit termination at, and by reading-through, the UAG stop codon and the altered structure of the ribosome is commensurate with sliding along the intervening non-coding sequences (Samatova *et al.*, 2014; Chen *et al.*, 2015; Agirrezabala *et al.*, 2017; Klimova *et al.*, 2017).

Ribosomal 'Frame-Shifting'

The *gag* and *pol* ORFs of avian alpharetrovirus Rous Sarcoma Virus (RSV) overlap by 58 nucleotides prior to the *gag* termination codon, with *pol* in the –1 reading frame with regards *gag*. Translation products comprised both *gag* and a *gag-pol* fusion protein, with *gag* in a ~20 fold excess over *gag-pol* (Jacks and Varmus, 1985; Jacks *et al.*, 1988). Since this discovery, programmed –1 ribosomal frameshifting has been described in many different viruses (reviewed by Firth and Brierley, 2012; Atkins *et al.*, 2016). Eukaryotic –1 frame-shifting mRNA signals comprise two cis-acting elements. Firstly, a 'slippery' sequence with the consensus X-XXY-YYZ (XXX representing identical nucleotides, YYY representing a triplet of A or U whilst Z represents A, C or U) – the –1 frame shift site. Secondly, an RNA structure, some 5–9nts downstream of the slippery sequence, may comprise either a stem-loop or pseudoknot and is required for efficient frame-shifting. It is thought that tandem slippage occurs in which tRNAs in both the ribosomal A- and P-sites detach from the mRNA zero frame (-XXY-YYZ-) and re-base pair with the mRNA in the –1 frame (-XXX-YYY-). In summary, the processivity of the majority of ribosomes is not impeded by the RNA secondary structure and continue elongation until the stop codon is encountered – translating Gag. The processivity of other ribosomes is impeded by the RNA structure, pausing the ribosome over the slippery sequence where a –1 frame-shift switches the ribosome into the *pol* ORF - translating a Gag-Pol fusion protein. This mechanism facilitates the generation of a defined, fixed, ratio of Gag: Gag-Pol optimal for replication and virion assembly. Furthermore, the Gag-Pol fusion protein facilitates targeting of this replication protein into the virion particle (also obviates the need for a separate ORF encoding Pol). This strategy is employed by many +ve strand RNA viruses to produce an optimal ratio of polymerase in comparison with other virus replication proteins (reviewed by Ahlquist, 2006).

Picornavirus particles are non-enveloped icosahedrons formed of 60 copies each of 4 different capsid proteins (VP1–4) encapsidating a single RNA molecule (~7.5–8.5 kb). These capsid proteins, together with the RNA replication proteins, are encoded by a single, long, ORF. It would appear, therefore, that a uniform stoichiometry in the synthesis of all of these proteins would occur throughout the infectious cycle: no possibility of 'early' (replication proteins)/'late' (capsid proteins) control over protein biogenesis. In contrast to the cis-acting elements described above producing a fixed ratio of –1 frame-shifting, recently it has been shown that EMCV and TMEV have evolved a trans-acting –1 frame-shifting mechanism driven by the 2A protein (Fig. 5; Naphthine *et al.*, 2017, 2019). In the early stages of infection the level of 2A is low, but increases throughout the infectious cycle. The 2A protein binds a stem-loop RNA structure present in the genomic region encoding protein 2B, immediately downstream of protein 2A. At the C-terminus of 2A a 'StopGo' recoding event (see below) produces a discontinuity in the polypeptide backbone thereby separating 2A from 2B. However, a shift site (-CCG-GUU-UUU-; underlined) and the stem-loop structure is formed by sequences encoding the N-terminal region of protein 2B. The stem-loop is bound by protein 2A inducing a –1 frame-shift, in the case of EMCV producing a short protein (2B'; 128 amino acids), rather than the remainder of the polyprotein – some 1249 amino acids. In the latter stages of infection, therefore, the cell's resources are increasingly devoted to the synthesis of capsid (rather than RNA replication) proteins thereby increasing the yield of virus particles.

'StopGo'/'Stop Carry-On'/Ribosomal 'Skipping'

A group of oligopeptide sequences collectively known as 2A, mediate a translational recoding event *in cis* known as "ribosome skipping," "StopGo" or "Stop Carry-on" translation (Donnelly *et al.*, 2001a,b; Atkins *et al.*, 2007; Doronina *et al.*, 2008). The function of the 2A sequence was first characterized from the +ve-stranded RNA picornavirus Foot-and-Mouth Disease Virus (FMDV). The 2A oligopeptide

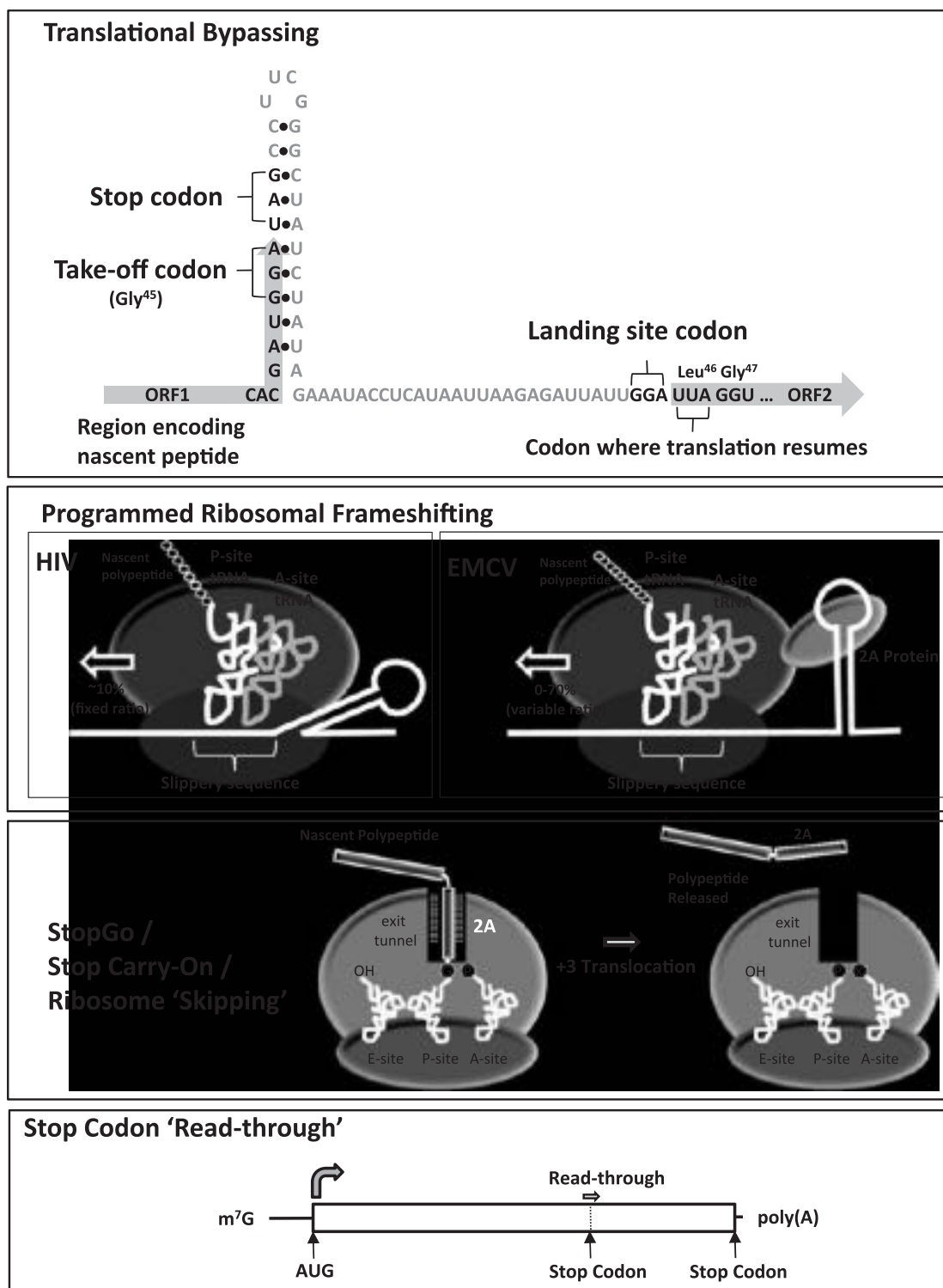


Fig. 5 Translation bypassing occurs when ribosomes translate the first 46 codons of the T4 gene 60 ORF, bypass (do not translate) the next 50 nucleotides, then resume elongation downstream of the 'take-off' RNA stem-loop structure. Programmed ribosomal frameshifting may occur when the ribosome encounters an RNA structure such that the ribosome is paused over the slippery sequence. In the case of EMCV, the 2A virus protein binds a stem-loop increasingly promoting frameshifting as the infectious cycle progresses. In the case of StopGo, the 2A sequence is thought to interact with the ribosome exit tunnel such that the stereochemistry in the peptidyl transferase center is altered: the peptide bond is not formed. The model proposes that prolyl-tRNA exits the A-site, eRF1/3 enters the A-site (at a sense codon) and terminates translation. After exit of eRF1, prolyl-tRNA then re-enters the A-site and is then (pseudo) translocated such that the A-site is now vacant and elongation can resume. Stop codon read-through is influenced by both the particular stop codon and its nucleotide context permitting low-level (0.3%–5%) read-through. The reverse-transcriptase of MuLV binds to eRF1 enhancing read-through of the stop codon at the end of the gag ORF - leading to the translation of a gag-pol fusion protein. Figure panel adapted from Napthine, S., Ling, R., Finch, L.K., *et al.*, 2017. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nature Communications* 8, 15582.

is only 18 amino acids long (-LLNFDLLKLAGDVESNPG-), delineated by 3C^{pro} post-translational cleavage at its N-terminus and co-translational 'cleavage' at its C-terminus (reviewed by Belsham, 1993). The sequence at the C-terminus of 2A is strongly conserved in FMDV comprising the canonical motif D(V/I)E(S/T)NPG¹P (arrow indicates 'cleavage' site; underlined proline representing the conserved N-terminal residue of protein 2B). The less conserved part of the 2A sequence, located upstream of the motif, was shown to be essential for 2A function (Ryan and Drew, 1994; Sharma *et al.*, 2012). Further analyses showed a similar recoding activity in just the C-terminal region of the longer cardiovirus 2A protein (Donnelly *et al.*, 1997). Early studies showed the FMDV 2A region was not simply a substrate for an FMDV proteinase (L^{pro}, 3C^{pro}), nor a substrate for a host-cell proteinase – 'self-cleavage' was a novel translational effect (Ryan *et al.*, 1989, 1991; Palmenberg *et al.*, 1992).

The model of translational recoding proposes that when a ribosome encounters 2A the nascent 2A oligopeptide sequence interacts with the ribosome exit tunnel altering the conformation of peptidyl-tRNA^{Gly} in the peptidyl transferase center such that the prolyl-tRNA present in the P-site is unable to form a peptide bond. Prolyl-tRNA exits the P-site allowing entry of eukaryotic release factor 1 (eRF1) to enter at the sense codon (proline) and terminate translation. The nascent polypeptide is released and eRF1 exits the A-site, accelerated by eRF3 (Doronina *et al.*, 2008). Prolyl-tRNA then re-enters the A-site and a 'pseudo-translocation' event occurs, mediated by eEF2, translocating the (deacylated) Glycyl-tRNA from the P-site to the exit site (E-site) plus the prolyl-tRNA from the A- to P-site positioning the next in-frame codon in the A-site (Fig. 5). Elongation of the downstream sequences may then proceed. In this manner a discontinuity in the polypeptide backbone (between 2A and 2B) is produced. Active '2A-like' sequences have been shown to be encoded by a wide range of viruses (Donnelly *et al.*, 2001b; Luke *et al.*, 2008). Since 2A/2A like sequences are active in all eukaryotic (but not prokaryotic) systems, they are used for protein co-expression (concatenating ORFs into self-processing polyproteins) in a wide range of biomedical/biotechnological purposes (Luke and Ryan, 2013, 2018).

Ribosome Stop Codon 'Read-Through'

Stop codons (UAA, UAG or UGA) are recognized by release factors which bind into the A-site and use an activated water molecule to hydrolyze the ester linkage between the completed polypeptide and tRNA. In general, this is a highly efficient process, but may be influenced by both the particular stop codon and nucleotide context permitting low-level (0.3%–5%) read-through (Fig. 5; Bertram *et al.*, 2001). Read-through arises from stop codons being read by near cognate or suppressor tRNAs. Viruses use all three types of motif to promote read-through (Beier and Grimm, 2001). The type I motif comprises the six nucleotides immediately downstream of the stop codon (-UAG-CAR-YYA-; R = purine, Y = pyrimidine), utilized by beny- and pomoviruses to produce C-terminally extended forms of the coat protein. Type II motifs comprise a UGA stop codon immediately followed by a triplet of CUA or CCG (Firth *et al.*, 2011) and are utilized by alphaviruses to read-through a stop codon in the replicase gene, although some alphaviruses do not have this internal stop codon. The replicase gene of other viruses tobria-, peclu-, furo- and pomoviruses all have internal stop codons and utilize the type II motif to promote read-through (reviewed by Firth and Brierley, 2012). Type III motifs are more heterogenous, comprising a UAG stop codon, a 3'-adjacent G or purine-rich octanucleotide tract plus a 3' RNA secondary structure. In gammaretrovirus replicase genes, an internal UAG codon together with (1) certain base sequences comprising an 8 nt spacer region separating the UAG from (2) a pseudoknot RNA structure brings about relatively efficient read-through (Alam *et al.*, 1999; Houck-Loomis *et al.*, 2011). However, influential RNA structures may be located some distance from the stop codon: in luteo- and poleroviruses the structure may be 700–750 nts downstream. Indeed, in viruses such as carnation italian ringspot tomosvirus and turnip crinkle carmovirus (CIRV, TCV; Tombusviruses) efficient read-through relies upon a long-range (3.5 kb) interaction between a proximal RNA stem-loop structure and a distal element (Cimino *et al.*, 2011).

In murine leukemia viruses (MuLVs) the polymerase is expressed as a gag-pol precursor by suppression of termination at the UAG stop codon at the end of the gag ORF. The MuLV reverse transcriptase (RT) has been shown to interact with eRF1 thereby enhancing the suppression of termination: the interaction of RT with eRF1 enhances stop codon read-through to control the biogenesis of the gag-pol fusion protein (Orlova *et al.*, 2003).

Virus-Encoded Proteinases

Collectively, viruses encode molecular mimics of serine, thiol and aspartyl cellular proteinases. These virus-encoded proteinases serve to control virus protein biogenesis (temporally and/or spatially) by 'processing' of virus precursor forms into mature products (reviewed in Ryan and Flint, 1997; Tong, 2002; Rodamilans *et al.*, 2018). Importantly, certain proteinases have also evolved to degrade key host-cell proteins – notably initiation factors and proteins involved in the innate immune system. Enterovirus polyproteins comprise two proteinase domains, 2A^{pro} and 3C^{pro}. In the enteroviruses the uncleaved 3C proteinase/3D polymerase (3CD^{pro}) is also a proteinase and in this form, rather than 3C^{pro}, is required for proteolytic processing of the capsid protein precursor P1 (Jore *et al.*, 1988). Alternative polyprotein processing pathways can, therefore, be used to control protein biogenesis. As outlined above, processing leads to the generation of either 3C^{pro} and 3D^{pol} as individual entities, or, 3CD^{pro} where the 3D^{pol} component of the uncleaved fusion protein acts to modify the substrate specificity of 3C^{pro}.

Another important aspect of virus encoded proteinases is the degradation of key host-cell proteins. As mentioned earlier, both enterovirus 2A^{pro} and 3CD^{pro} proteinases degrade proteins within the nuclear pore complex. Critically, enterovirus 2A^{pro} degrades eIF4G (Fig. 1, Panel B) bringing about 'shut-off' but also degrades PABP (Joachims *et al.*, 1999; Kerekatte *et al.*, 1999). Factors eIF4G and eIF5B are also degraded by 3C^{pro} (de Breyne *et al.*, 2008). FMDV polyproteins also comprise two proteinase domains:

L^{Pro} (a thiol-type proteinase) and 3C^{Pro}, both also cleaving eIF4G (Fig. 1, Panel B; Devaney *et al.*, 1988). Similarly, the calicivirus 3C-like proteinase degrades PABP (Kuyumcu-Martinez *et al.*, 2004). Again, eIF4G and PABP are also degraded by the lentivirus HIV-1 (aspartyl-type) proteinase (Ventoso *et al.*, 2001; Alvarez *et al.*, 2006). It is quite apparent, therefore, that both eIF4G and PABP are common 'targets' for virus-encoded proteinases (reviewed in Smith and Gray, 2010).

Virus proteinases may also control protein biogenesis in both a temporal and spatial manner. HIV RNA contains many post-transcriptional modifications such as N6-methyladenosine (m⁶A), which is thought to reduce viral replication. The m⁶A reader protein YTHDF3 is incorporated into HIV particles and reduces viral infectivity in the next cycle of infection. The Gag-Pol polyprotein is incorporated into the assembling particle through Gag-Gag interactions: concomitant with virus budding, the proteinase is activated to facilitate particle maturation (reviewed by Ganser-Pornillos *et al.*, 2008). Furthermore, the HIV proteinase has recently been shown to cleave YTHDF3 incorporated within virions, increasing the infectivity of the mature particles (Jurczyk *et al.*, 2020).

Another highly important function of virus proteinases is to degrade components of the innate immune response. For example, enterovirus 2A^{Pro} degrades melanoma differentiation-associated gene 5 (MDA5) and mitochondrial antiviral signaling protein (MAVS; Feng *et al.*, 2014), whereas retinoic acid-inducible gene-I (RIG-I) cleavage is observed during infection by various enteroviruses and cardioviruses (Barral *et al.*, 2009; Papon *et al.*, 2009). FMDV L^{Pro} degrades the p65 subunit of NF- κ B (de Los Santos *et al.*, 2007), reduces IRF-3/7 expression (Wang *et al.*, 2010), degrades the Laboratory of Genetics and Physiology 2 (LGP2) protein (Rodríguez Pulido *et al.*, 2018) and MDA5 (Rodríguez Pulido *et al.*, 2020).

MicroRNAs (miRNAs)

miRNAs are small, non-coding RNAs, with an average length of 22 nucleotides that play important roles in the post-transcriptional regulation of cellular gene expression. The viral replication cycles of both DNA and RNA viruses may be enhanced by the evolution of mechanisms that can either degrade, boost, or even hijack cellular miRNAs - one example of each being given below.

In marmoset T-cells latently infected with Herpesvirus saimiri (HVS; gammaherpesvirus), the most abundant virus transcripts are seven small non-coding RNAs termed HSURs (Herpesvirus saimiri U-rich RNAs). HSUR 1 base-pairs with the host-cell miRNA-27 which brings about degradation of miRNA 27 (Cazalla *et al.*, 2010). What advantage does this confer on the virus? miRNA-27 has been shown to decrease the levels of the cell surface signaling protein semaphorin 7A (SEMA7A), the effector cytokine gamma interferon (IFN- γ) and the adapter protein growth factor receptor-bound protein 2 (GRB2; Guo *et al.*, 2014). The repression of T-cell activation by miRNA-27 is relieved by HSUR 1-induced miR-27 degradation.

In contrast, in Epstein-Barr virus (EBV; gammaherpesvirus) latently infected B-cells the expression of host-cell miR-155 is increased ~1000 fold (Lu *et al.*, 2008; Yin *et al.*, 2008). miR-155 plays a number of important immunomodulatory roles in B-cells, T-cells, macrophages, dendritic cells and is an oncogenic miRNA. (O'Connell *et al.*, 2007; Thai *et al.*, 2007). In lymphoblastoid cell-lines or diffuse large B-cell lymphomas positive for EBV, the inhibition or deletion of miR-155 reduces cell growth and promotes apoptosis (Linnstaedt *et al.*, 2010). EBV induced upregulation of host miR-155 may, therefore, be a key aspect in virus-induced oncogenesis.

miR-122 is a highly abundant liver-specific miRNA regulating fatty acid and cholesterol biosynthesis. miR-122 facilitates the replication of hepatitis C virus (HCV) and base-pairs with two conserved binding sites in the HCV IRES. Indeed, miR-122 expression is sufficient to convert non-hepatic cells to become permissive for HCV (Fukuhara *et al.*, 2012; Kambara *et al.*, 2012).

References

- Abernathy, E., Glaunsinger, B., 2015. Emerging roles for RNA degradation in viral replication and antiviral defense. *Virology* 479–480, 600–608.
- Agirrezabala, X., Samatova, E., Klimova, M., *et al.*, 2017. Ribosome rearrangements at the onset of translational bypassing. *Science Advances* 3, e1700147.
- Ahlquist, P., 2006. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nature Reviews Microbiology* 4, 371–382.
- Ahmadian, G., Randhawa, J.S., Easton, A.J., 2000. Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. *European Molecular Biology Organization Journal* 19, 2681–2689.
- Alam, S.L., Wills, N.M., Ingram, J.A., Atkins, J.F., Gesteland, R.F., 1999. Structural studies of the RNA pseudoknot required for readthrough of the gag-termination codon of murine leukemia virus. *Journal of Molecular Biology* 288, 837–852.
- Alvarez, E., Castello, A., Menendez-Arias, L., Carrasco, L., 2006. HIV protease cleaves poly(A)-binding protein. *Biochemical Journal* 396, 219–226.
- Atkins, J.F., Loughran, G., Bhatt, P.R., Firth, A.E., Baranov, P.V., 2016. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Research* 44, 7007–7078.
- Atkins, J.F., Wills, N.M., Loughran, G., *et al.*, 2007. A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA* 13, 803–810.
- Bardina, M.V., Lidsky, P.V., Sheval, E.V., *et al.*, 2009. Mengovirus-induced rearrangement of the nuclear pore complex: hijacking cellular phosphorylation machinery. *Journal of Virology* 83, 3150–3161.
- Barral, P.M., Sarkar, D., Fisher, P.B., Racaniello, V.R., 2009. RIG-I is cleaved during picornavirus infection. *Virology* 391, 171–176.
- Beier, H., Grimm, M., 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Research* 29, 4767–4782.
- Beloso, A., Martínez, C., Valcárcel, J., Santarén, J.F., Ortín, J., 1992. Degradation of cellular mRNA during influenza virus infection: Its possible role in protein synthesis shutoff. *Journal of General Virology* 73, 575–581.
- Belsham, G.J., 1993. Distinctive features of the foot-and-mouth disease virus, a member of the picornavirus family; aspects of virus protein synthesis, protein processing and structure. *Progress in Biophysics and Molecular Biology* 60, 241–260.

- Bertram, G., Innes, S., Minella, O., Richardson, J., Stansfield, I., 2001. Endless possibilities: translation termination and stop codon recognition. *Microbiology* 147, 255–269.
- Butcher, S.E., Jan, E., 2016. tRNA-mimicry in IRES-mediated translation and recoding. *RNA Biology* 13, 1068–1074.
- Carrasco, L., Sanz, M.A., González-Almela, E., 2018. The regulation of translation in alphavirus-infected cells. *Viruses* 10, 70.
- Cazalla, D., Yario, T., Steitz, J.A., 2010. Down-regulation of a host microRNA by a herpesvirus saimiri noncoding RNA. *Science* 328, 1563–1566.
- Chen, J., Coakley, A., O'Connor, M., *et al.*, 2015. Coupling of mRNA structure rearrangement to ribosome movement during bypassing of non-coding regions. *Cell* 163, 1267–1280.
- Cimino, P.A., Nicholson, B.L., Wu, B., Xu, W., White, K.A., 2011. Multifaceted regulation of translational readthrough by RNA replication elements in a tombusvirus. *Public Library of Science Pathogens* 7, e1002423.
- Cuesta, R., Xi, Q., Schneider, R.J., 2000. Adenovirus-specific translation by displacement of kinase Mnk1 from cap-initiation complex eIF4F. *European Molecular Biology Organization Journal* 19, 3465–3474.
- Danthinne, X., Seurinck, J., Meulewaeter, F., Van Montagu, M., Cornelissen, M., 1993. The 3' untranslated region of satellite tobacco necrosis virus RNA stimulates translation in vitro. *Molecular and Cellular Biology* 13, 3340–3349.
- de Breynne, S., Bonderoff, J.M., Chumakov, K.M., Lloyd, R.E., Hellen, C.U.T., 2008. Cleavage of eukaryotic initiation factor eIF5B by enterovirus 3C proteases. *Virology* 378, 118–122.
- de Los Santos, T., Diaz-San Segundo, F., Grubman, M.J., 2007. Degradation of nuclear factor kappa B during foot-and-mouth disease virus infection. *Journal of Virology* 81, 12803–12815.
- Delhay, S., van Pesch, V., Michiels, T., 2004. The leader protein of Theiler's virus interferes with nucleocytoplasmic trafficking of cellular proteins. *Journal of Virology* 78, 4357–4362.
- Devaney, M.A., Vakharia, V.N., Lloyd, R.E., Ehrenfeld, E., Grubman, M.J., 1988. Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap-binding protein complex. *Journal of Virology* 62, 4407–4409.
- Dhunge, P., Cantu, F.M., Molina, J.A., Yang, Z., 2020. Vaccinia virus as a master of host shutoff induction: Targeting processes of the central dogma and beyond. *Pathogens* 9, 400.
- Domashevskiy, A., Cheng, S., 2015. Thermodynamic analysis of binding and enzymatic properties of pokeweed antiviral protein (PAP) toward Tobacco Etch Virus (TEV) RNA. *Journal of Nature And Science* 1, e82.
- Donnelly, M.L.L., Gani, D., Flint, M., Monaghan, S., Ryan, M.D., 1997. The cleavage activities of aphthovirus and cardiovirus 2A proteins. *Journal of General Virology* 78, 13–21.
- Donnelly, M.L.L., Hughes, L.E., Luke, G.A., *et al.*, 2001a. The "cleavage" activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring "2A-like" sequences. *Journal General Virology* 82, 1027–1041.
- Donnelly, M.L.L., Luke, G.A., Mehrotra, A., *et al.*, 2001b. Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal "skip". *Journal General Virology* 82, 1013–1025.
- Doronina, V.A., Wu, C., de Felipe, P., *et al.*, 2008. Site-Specific release of nascent chains from ribosomes at a sense codon. *Molecular and Cellular Biology* 28, 4227–4239.
- Faria, P.A., Chakraborty, P., Levay, A., *et al.*, 2005. VSV disrupts the Rae1/Mrp41 mRNA nuclear export pathway. *Molecular Cell* 17, 93–102.
- Feng, Q., Langereis, M.A., Lork, M., *et al.*, 2014. Enterovirus 2A^{pro} targets MDA5 and MAVS in infected cells. *Journal of Virology* 88, 3369–3378.
- Firth, A.E., Brierley, I., 2012. Non-canonical translation in RNA viruses. *Journal of General Virology* 93, 1385–1409.
- Firth, A.E., Wills, N.M., Gesteland, R.F., Atkins, J.F., 2011. Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. *Nucleic Acids Research* 39, 6679–6691.
- Flather, D., Semler, B.L., 2015. Picornaviruses and nuclear functions: Targeting a cellular compartment distinct from the replication site of a positive-strand RNA virus. *Frontiers in Microbiology* 6, 594.
- Fortes, P., Beloso, A., Ortín, J., 1994. Influenza virus NS1 protein inhibits pre-mRNA splicing and blocks mRNA nucleocytoplasmic transport. *European Molecular Biology Organization Journal* 13, 704–712.
- Fukuhara, T., Kambara, H., Shiokawa, M., *et al.*, 2012. Expression of microRNA miR-122 facilitates an efficient replication in nonhepatic cells upon infection with hepatitis C virus. *Journal of Virology* 86, 7918–7933.
- Fütterer, J., Hohn, T., 1991. Translation of a polycistronic mRNA in the presence of the cauliflower mosaic virus transactivator protein. *European Molecular Biology Organization Journal* 10, 3887–3896.
- Gallie, D.R., 2001. Cap-independent translation conferred by the 5' leader of tobacco etch virus is eukaryotic initiation factor 4G dependent. *Journal of Virology* 75, 12141–12152.
- Ganser-Pornillos, B.K., Yeager, M., Sundquist, W.I., 2008. The structural biology of HIV assembly. *Current Opinion in Structural Biology* 18, 203–217.
- García, M.A., Gil, J., Ventoso, I., *et al.*, 2006. Impact of protein kinase PKR in cell biology: from antiviral to antiproliferative action. *Microbiology and Molecular Biology Reviews* 70, 1032–1060.
- Gingras, A.C., Svitkin, Y., Belsham, G.J., Pause, A., Sonenberg, N., 1996. Activation of the translational suppressor 4E-BP1 following infection with encephalomyocarditis virus and poliovirus. *Proceedings of the National Academy of Sciences USA* 93, 5578–5583.
- Goodfellow, I., 2011. The genome-linked protein VPg of vertebrate viruses – A multifaceted protein. *Current Opinion in Virology* 1, 355–362.
- Goodfellow, I., Chaudhry, Y., Gioldasi, I., *et al.*, 2005. Calicivirus translation initiation requires an interaction between VPg and eIF4E. *European Molecular Biology Organization Reports* 6, 968–972.
- Gunišová, S., Hronová, V., Mohammad, M.P., Hinnebusch, A.G., Valášek, L.S., 2018. Please do not recycle! Translation reinitiation in microbes and higher eukaryotes. *Federation of European Microbiology Societies Microbiology Reviews* 42, 165–192.
- Guo, Y.E., Riley, K.J., Iwasaki, A., Steitz, J.A., 2014. Alternative capture of noncoding RNAs or protein-coding genes by herpesviruses to alter host T cell function. *Molecular Cell* 54, 67–79.
- Habeta, M., Luttermann, C., Meyers, G., 2014. Feline calicivirus can tolerate gross changes of its minor capsid protein expression levels induced by changing translation reinitiation frequency or use of a separate VP2-coding mRNA. *Public Library of Science One* 9, e102254.
- Han, M., Ke, H., Zhang, Q., Yoo, D., 2017. Nuclear imprisonment of host cellular mRNA by nsp1 β protein of porcine reproductive and respiratory syndrome virus. *Virology* 505, 42–55.
- Harb, M., Becker, M.M., Vitour, D., *et al.*, 2008. Nuclear localization of cytoplasmic poly(A)-binding protein upon rotavirus infection involves the interaction of NSP3 with eIF4G and RoXaN. *Journal of Virology* 82, 11283–11293.
- Hayashi, T., MacDonald, L.A., Takimoto, T., 2015. Influenza A virus protein PA-X contributes to viral growth and suppression of the host antiviral and immune responses. *Journal of Virology* 89, 6442–6452.
- Horvath, C.M., Williams, M.A., Lamb, R.A., 1990. Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. *European Molecular Biology Organization Journal* 9, 2639–2647.
- Houck-Loomis, B., Durney, M.A., Salguero, C., *et al.*, 2011. An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* 480, 561–564.
- Ho, B.-C., Yu, S.-L., Chen, J.J.W., *et al.*, 2011. Enterovirus-induced miR-141 contributes to shutoff of host protein translation by targeting the translation initiation factor eIF4E. *Cell Host and Microbe* 9, 58–69.

- Huang, W.M., Ao, S.Z., Casjens, S., *et al.*, 1988. A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science* 239, 1005–1012.
- Huang, S., Ghabrial, S.A., 1996. Organization and expression of the double-stranded RNA genome of Helminthosporium victoriae 190S virus, a totivirus infecting a plant pathogenic filamentous fungus. *Proceedings of the National Academy of Sciences USA* 93, 12541–12546.
- Jackson, R.J., Hellen, C.U., Pestova, T.V., 2012. Termination and post-termination events in eukaryotic translation. *Advances in Protein Chemistry and Structural Biology* 86, 45–93.
- Jacks, T., Madhani, H.D., Masiarz, F.R., Varmus, H.E., 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55, 447–458.
- Jacks, T., Varmus, H.E., 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. *Science* 230, 1237–1242.
- Jang, S.K., Kräusslich, H.G., Nicklin, M.J., *et al.*, 1988. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *Journal of Virology* 62, 2636–2643.
- Jan, E., Sarnow, P., 2002. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *Journal of Molecular Biology* 324, 889–902.
- Jiang, J., Laliberté, J.-F., 2011. The genome-linked protein VPg of plant viruses—a protein with many partners. *Current Opinion in Virology* 1, 347–354.
- Joachims, M., van Breugel, P.C., Lloyd, R.E., 1999. Cleavage of poly(A)-binding protein by enterovirus proteases concurrent with inhibition of translation in vitro. *Journal of Virology* 73, 718–727.
- Jore, J., De Geus, B., Jackson, R.J., Pouwels, P.H., Enger-Valk, B.E., 1988. Poliovirus protein 3CD is the active protease for processing of the precursor protein P1 in vitro. *Journal of General Virology* 69, 1627–1636.
- Jurczyszak, D., Zhang, W., Terry, S.N., *et al.*, 2020. HIV protease cleaves the antiviral m⁶A reader protein YTHDF3 in the viral particle. *Public Library of Science Pathogens* 16, e1008305.
- Kambara, H., Fukuhara, T., Shiokawa, M., *et al.*, 2012. Establishment of a novel permissive cell line for the propagation of hepatitis C virus by expression of microRNA miR122. *Journal of Virology* 86, 1382–1393.
- Kerekatte, V., Keiper, B.D., Badoff, C., *et al.*, 1999. Cleavage of poly(A)-binding protein by coxsackievirus 2A protease in vitro and in vivo: Another mechanism for host protein synthesis shutoff? *Journal of Virology* 73, 709–717.
- Khaperskyy, D.A., Schmaling, S., Larkins-Ford, J., McCormick, C., Gaglia, M.M., 2016. Selective degradation of host RNA polymerase II transcripts by influenza A virus PA-X host shutoff protein. *Public Library of Science Pathogens* 12, e1005427.
- Klimova, M., Senyushkina, T., Samatova, E., *et al.*, 2017. EF-G-induced ribosome sliding along the noncoding mRNA. *Science Advances* 5 (6), eaaw9049.
- Koffa, M.D., Clements, J.B., Izaurralde, E., *et al.*, 2001. Herpes simplex virus ICP27 protein provides viral mRNAs with access to the cellular mRNA export pathway. *European Molecular Biology Organization Journal* 20, 5769–5778.
- Kozak, M., 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292.
- Kozak, M., 1990. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proceedings of the National Academy of Sciences USA* 87, 8301–8305.
- Kuss, S.K., Mata, M.A., Zhang, L., Fontoura, B.M.A., 2013. Nuclear imprisonment: viral strategies to arrest host mRNA nuclear export. *Viruses* 5, 1824–1849.
- Kuyumcu-Martinez, M., Belliot, G., Sosnovtsev, S.V., *et al.*, 2004. Calicivirus 3C-like proteinase inhibits cellular translation by cleavage of poly(A)-binding protein. *Journal of Virology* 78, 8172–8182.
- Langland, J.O., Cameron, J.M., Heck, M.C., Jancovich, J.K., Jacobs, B.L., 2006. Inhibition of PKR by RNA and DNA viruses. *Virus Research* 119, 100–110.
- Lidsky, P.V., Hato, S., Bardina, M.V., *et al.*, 2006. Nucleocytoplasmic traffic disorder induced by cardioviruses. *Journal of Virology* 80, 2705–2717.
- Linnstaedt, S.D., Gottwein, E., Skalsky, R.L., Luftig, M.A., Cullen, B.R., 2010. Virally induced cellular microRNA miR-155 plays a key role in B-cell immortalization by Epstein-Barr virus. *Journal of Virology* 84, 11670–11678.
- Lukavsky, P.J., 2009. Structure and function of HCV IRES domains. *Virus Research* 139, 166–171.
- Luke, G.A., de Felipe, P., Lukashev, A., *et al.*, 2008. The occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes. *Journal of General Virology* 89, 1036–1042.
- Luke, G.A., Ryan, M.D., 2013. The protein co-expression problem in biotechnology and biomedicine: Virus 2A and 2A-like sequences provide a solution. *Future Virology* 8, 983–996.
- Luke, G.A., Ryan, M.D., 2018. Therapeutic applications of the 'NPGP' family of viral 2As. *Reviews in Medical Virology* 28, e2001.
- Luttermann, C., Meyers, G., 2007. A bipartite sequence motif induces translation reinitiation in feline calicivirus RNA. *Journal of Biological Chemistry* 282, 7056–7065.
- Lu, F., Weidner, A., Liu, C.G., *et al.*, 2008. Epstein-Barr virus-induced miR-155 attenuates NF- κ B signaling and stabilizes latent virus persistence. *Journal of Virology* 82, 10436–10443.
- Meyers, G., 2003. Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism. *Journal of Biological Chemistry* 278, 34051–34060.
- Mir, M.A., Panganiban, A.T., 2008. A protein that replaces the entire cellular eIF4F complex. *European Molecular Biology Organization Journal* 27, 3129–3139.
- Mohr, I., 2016. Closing in on the causes of host shutoff. *eLife* 5, e20755.
- Montero, H., García-Román, R., Mora, S.I., 2015. eIF4E as a control target for viruses. *Viruses* 7, 739–750.
- Napthine, S., Bell, S., Hill, C.H., Brierley, I., Firth, A.E., 2019. Characterization of the stimulators of protein-directed ribosomal frameshifting in Theiler's murine encephalomyelitis virus. *Nucleic Acids Research* 47, 8207–8223.
- Napthine, S., Lever, R.A., Powell, M.L., *et al.*, 2009. Expression of the VP2 protein of Murine Norovirus by a translation termination-reinitiation strategy. *Public Library of Science One* 4 (12), e8390.
- Napthine, S., Ling, R., Finch, L.K., *et al.*, 2017. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nature Communications* 8, 15582.
- Nemeroff, M.E., Barabino, S.M., Li, Y., *et al.*, 1998. Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3' end formation of cellular pre-mRNAs. *Molecular Cell* 1, 991–1000.
- Nicholson, B.L., White, K.A., 2011. 3' Cap-independent translation enhancers of positive-strand RNA plant viruses. *Current Opinion in Virology* 1, 373–380.
- Orlova, M., Yueh, A., Leung, J., Goff, S.P., 2003. Reverse transcriptase of Moloney murine leukemia virus binds to eukaryotic release factor 1 to modulate suppression of translational termination. *Cell* 115, 319–331.
- O'Connell, R.M., Taganov, K.D., Boldin, M.P., Cheng, G., Baltimore, D., 2007. MicroRNA-155 is induced during the macrophage inflammatory response. *Proceedings of the National Academy of Sciences USA* 104, 1604–1609.
- Palmenberg, A.C., Parks, G.D., Hall, D.J., *et al.*, 1992. Proteolytic processing of the cardioviral P2 region: 2A/2B cleavage in clone-derived precursors. *Virology* 190, 754–762.
- Papon, L., Oteiza, A., Imaizumi, T., *et al.*, 2009. The viral RNA recognition sensor RIG-I is degraded during encephalomyocarditis virus (EMCV) infection. *Virology* 393, 311–318.
- Park, Y.W., Katze, M.G., 1995. Translational control by influenza virus. Identification of cis-acting sequences and trans-acting factors which may regulate selective viral mRNA translation. *Journal of Biological Chemistry* 270, 28433–28439.
- Parrish, S., Moss, B., 2007. Characterization of a second vaccinia virus mRNA-decapping enzyme conserved in poxviruses. *Journal of Virology* 81, 12973–12978.
- Pelletier, J., Sonenberg, N., 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320–325.
- Pestova, T.V., Shatsky, I.N., Fletcher, S.P., Jackson, R.J., Hellen, C.U., 1998. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes and Development* 12, 67–83.

- Pooggin, M.M., Ryabova, L.A., 2018. Ribosome shunting, polycistronic translation, and evasion of antiviral defenses in plant pararetroviruses and beyond. *Frontiers in Microbiology* 9, 644.
- Porter, F.W., Bochkov, Y.A., Albee, A.J., Wiese, C., Palmenberg, A.C., 2006. A picornavirus protein interacts with Ran-GTPase and disrupts nucleocytoplasmic transport. *Proceedings of the National Academy of Sciences USA* 103, 12417–12422.
- Pöyry, T.A., Kaminski, A., Connell, E.J., Fraser, C.S., Jackson, R.J., 2007. The mechanism of an exceptional case of reinitiation after translation of a long ORF reveals why such events do not generally occur in mammalian mRNA translation. *Genes and Development* 21, 3149–3162.
- Ren, Q., Wang, Q.S., Firth, A.E., *et al.*, 2012. Alternative reading frame selection mediated by a tRNA-like domain of an internal ribosome entry site. *Proceedings of the National Academy of Sciences USA* 109, E630–E639.
- Ricour, C., Delhaye, S., Hato, S.V., *et al.*, 2009. Inhibition of mRNA export and dimerization of interferon regulatory factor 3 by Theiler's virus leader protein. *Journal of General Virology* 90, 177–186.
- Rodamilans, B., Shan, H., Pasin, F., García, J.A., 2018. Plant viral proteases: Beyond the role of peptide cutters. *Frontiers in Plant Science* 9, 666.
- Rodríguez Pulido, M., Martínez-Salas, E., Sobrino, F., Sáiz, M., 2020. MDA5 cleavage by the leader protease of foot-and-mouth disease virus reveals its pleiotropic effect against the host antiviral response. *Cell Death and Disease* 11, 718.
- Rodríguez Pulido, M., Sánchez-Aparicio, M.T., Martínez-Salas, E., *et al.*, 2018. Innate immune sensor LGP2 is cleaved by the leader protease of foot-and-mouth disease virus. *Public Library of Science Pathogens* 14, e1007135.
- Rodríguez, A., Pérez-González, A., Nieto, A., 2007. Influenza virus infection causes specific degradation of the largest subunit of cellular RNA polymerase II. *Journal of Virology* 81, 5315–5324.
- Rosenthal, M., Gogrefe, N., Vogel, D., *et al.*, 2017. Structural insights into replectenavirus cap-snatching machinery. *Public Library of Science Pathogens* 13, e1006400.
- Ryan, M.D., Belsham, G.J., King, A.M., 1989. Specificity of enzyme-substrate interactions in foot-and-mouth disease virus polyprotein processing. *Virology* 173, 35–45.
- Ryan, M.D., Drew, J., 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *European Molecular Biology Organization Journal* 134, 928–933.
- Ryan, M.D., Flint, M., 1997. Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* 78, 699–723.
- Ryan, M.D., King, A.M., Thomas, G.P., 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *Journal of General Virology* 72, 2727–2732.
- Sadahiro, A., Fukao, A., Kosaka, M., *et al.*, 2018. Translation of hepatitis A virus IRES is upregulated by a hepatic cell-specific factor. *Frontiers in Genetics* 9, 307.
- Samatova, E., Konevega, A.L., Wills, N.M., Atkins, J.F., Rodnina, M.V., 2014. High-efficiency translational bypassing of non-coding nucleotides specified by mRNA structure and nascent peptide. *Nature Communications* 5, 4459.
- Shapira, R., Choi, G.H., Nuss, D.L., 1991. Virus-like genetic organization and expression strategy for a double-stranded RNA genetic element associated with biological control of chestnut blight. *European Molecular Biology Organization Journal* 10, 731–739.
- Sharma, P., Yan, F., Doronina, V., *et al.*, 2012. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Research* 40, 1–9.
- Shors, T., Keck, J.G., Moss, B., 1999. Down regulation of gene expression by the vaccinia virus D10 protein. *Journal of Virology* 73, 791–796.
- Simon, A.E., Miller, W.A., 2013. 3' Cap-independent translation enhancers of plant viruses. *Annual Review of Microbiology* 67, 21–42.
- Smith, R.W., Gray, N.K., 2010. Poly(A)-binding protein (PABP): A common viral target. *Biochemical Journal* 426, 1–12.
- Takeuchi, K., Komatsu, T., Kitagawa, Y., Sada, K., Gotoh, B., 2008. Sendai virus C protein plays a role in restricting PKR activation by limiting the generation of intracellular double-stranded RNA. *Journal of Virology* 82, 10102–10110.
- Thai, T.H., Calado, D.P., Casola, S., *et al.*, 2007. Regulation of the germinal center response by microRNA-155. *Science* 316, 604–608.
- Tong, L., 2002. Viral proteases. *Chemical Reviews* 102, 4609–4626.
- Toth, Z., Stamminger, T., 2008. The human cytomegalovirus regulatory protein UL69 and its effect on mRNA export. *Frontiers in Bioscience* 13, 2939–2949.
- Touriol, C., Bornes, S., Bonnal, S., *et al.*, 2003. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the Cell* 95, 169–178.
- Ventoso, I., Blanco, R., Perales, C., Carrasco, L., 2001. HIV-1 protease cleaves eukaryotic initiation factor 4G and inhibits cap-dependent translation. *Proceedings of the National Academy of Sciences USA* 98, 12966–12971.
- Walsh, D., 2017. Poxviruses: slipping and sliding through transcription and translation. *Public Library of Science Pathogens* 13, e1006634.
- Wang, D., Fang, L., Luo, R., *et al.*, 2010. Foot-and-mouth disease virus leader proteinase inhibits dsRNA-induced type I interferon transcription by decreasing interferon regulatory factor 3/7 in protein levels. *Biochemical and Biophysical Research Communications* 399, 72–78.
- Wilson, J.E., Pestova, T.V., Hellen, C.U., Sarnow, P., 2000. Initiation of protein synthesis from the A site of the ribosome. *Cell* 102, 511–520.
- Yin, Q., McBride, J., Fewell, C., *et al.*, 2008. MicroRNA-155 is an Epstein-Barr virus-induced gene that modulates Epstein-Barr virus-regulated gene expression pathways. *Journal of Virology* 82, 5295–5306.
- Zeenko, V., Gallie, D.R., 2005. Cap-independent translation of tobacco etch virus is conferred by an RNA pseudoknot in the 5'-leader. *Journal of Biological Chemistry* 280, 26813–26824.

Further Reading

- Atkins, J.F., Gesteland, R.F., 2010. *Recoding: expansion of decoding rules enriches gene expression*. New York: Springer.
- Guo, Y.E., Steitz, J.A., 2014. Virus meets host microRNA: The destroyer, the booster, the hijacker. *Molecular and Cellular Biology* 34, 3780–3787.
- Hoang, H.-D., Neault, S., Pelin, A., Alain, T., 2020. Emerging translation strategies during virus–host interaction. *Wiley Interdisciplinary Reviews RNA* 5, e1619.
- Jaafar, Z.A., Kieft, J.S., 2019. Viral RNA structure-based strategies to manipulate translation. *Nature Reviews Microbiology* 17, 110–123.
- Jan, E., Mohr, I., Walsh, D., 2016. A cap-to-tail guide to mRNA translation strategies in virus-infected cells. *Annual Review of Virology* 3, 283–307.
- Jopling, C.L., Schutz, S., Sarnow, P., 2008. Position-dependent function for a tandem microRNA miR-122-binding site located in the hepatitis C virus RNA genome. *Cell Host and Microbe* 4, 77–85.
- Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M., Sarnow, P., 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science* 309, 1577–1581.
- Leppek, K., Das, R., Barna, M., 2018. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology* 19, 158–174.
- Lloyd, R.E., 2006. Translational control by viral proteinases. *Virus Research* 119, 76–88.
- Martínez-Salas, E., Piñeiro, D., Fernández, N., 2012. Alternative mechanisms to initiate translation in eukaryotic mRNAs. *Comparative and Functional Genomics* 2012, 391546.
- Walsh, D., Mathews, M.B., Mohr, I., 2013. Tinkering with translation: Protein synthesis in virus-infected cells. *Cold Spring Harbor Perspectives in Biology* 5, a012351.
- Walsh, D., Mohr, I., 2011. Viral subversion of the host protein synthesis machinery. *Nature Reviews Microbiology* 9, 860–875.

Relevant Websites

<https://viralzone.expasy.org/1579>

Eukaryotic host translation shutoff by virus.
ViralZone page.

<https://viralzone.expasy.org/867>

Viral initiation of translation.
ViralZone page.

Recombination

Jozef J Bujarski, Northern Illinois University, DeKalb, IL, United States and Polish Academy of Sciences, Poznan, Poland

© 2008 Elsevier Ltd. All rights reserved

This is a reproduction of J.J. Bujarski, Recombination. In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00545-8.

Introduction

Genetic recombination of viruses could be defined as the exchange of fragments of genetic material (DNA or RNA) among parental viral genomes. The result of recombination is a novel genetic entity that carries genetic information in nonparental combinations. Biochemically, recombination is a process of combining or substituting portions of nucleic acid molecules. Recombination has been recognized as an important process leading to genetic diversity of viral genomes upon which natural selection can function. Depending on the category of viruses, recombination can occur at the RNA or DNA levels. Since these processes are different for DNA and RNA viruses, they are described separately.

Recombination in DNA Viruses

In many DNA viruses, genetic recombination is believed to occur by means of cellular DNA recombination machinery. Cellular DNA recombination events are of either homologous (general recombination) or nonhomologous types. The nonhomologous recombination events occur relatively rarely and are promoted by special proteins that interact with special DNA signal sequences. In general, homologous recombination events occur much more often and they are most commonly known as genetic crossing-over that happens in every DNA-based organism during meiosis.

The biochemical pathways responsible for DNA crossing-over are well established. General elements involved in general recombination include DNA sequence identity, complementary base-pairing between double-stranded DNA molecules, heteroduplex formation between the two recombining DNA strands, and specialized recombination enzymes. The best-studied recombination system of *Escherichia coli* involves proteins such as *recA*, and *RecBCD*, and it has led to a large amount of literature. Interestingly, related DNA recombination proteins have been characterized in eukaryotes, including yeast, insects, mammals, and plants.

Yet certain DNA virus species encode their own recombination proteins, and some of these viruses serve as model system with those to study the recombination processes. One of the best-known systems is of certain bacteriophages that recombine independently from the host mechanisms. These independent pathways are used for repairing damaged phage DNA and for exchanging DNA to increase diversity among the related phages. In Enterobacteria phage M13, high recombination frequency was observed within the origin of phage DNA replication in the *E. coli* host. There, the crossovers have occurred at the nucleotide adjacent to the nick at the replication origin, because of joining to a nucleotide at a remote site in the genome. These results implicated a breakage-and-religation mechanism of such apparently illegitimate cross-overs.

Importantly enough, many of these phage recombination mechanisms are analogous to the pathways operating in the host bacteria. For instance, *Rec* proteins of phages T4 and T7 are analogous to bacterial *RecA*, *RecG*, *RuvC*, or *RecBCD* proteins, while *RecE* pathway in the *rac* prophage of *E. coli* K-12 or the phage 1 red system influenced the studies of bacterial systems. A correlation of different stages of DNA recombination with transcription and DNA replication during Enterobacteria phage T4 growth cycle is shown in Fig. 1.

Phage lambda (λ) has a recombination system that can substitute for the *RecF* pathway components in *E. coli*. The Enterobacteria phage λ moves its viral genome into and out of the bacterial chromosome using site-specific recombination. Based on crystal structures of the reaction intermediates, it is clear how the Enterobacteria phage λ integrase interacts with both core and regulatory DNA elements (Fig. 2).

Recombination between viral DNA and host genes can lead to acquisition of cellular genes by DNA viruses. For instance, tRNA genes are present in Enterobacteria phage T4. Interestingly, these tRNA sequences contain introns suggesting that Enterobacteria phage T4 must have passed through a eukaryotic host during evolution. Similar viral–host recombination events were observed for retroviruses in eukaryotic cells.

Genetic recombination in DNA viruses is often studied using functional marker mutations. In single-component DNA viruses recombination occurs by exchanging DNA fragments, whereas in segmented DNA viruses, additional events rely on reassortment of the entire genome segments. This complicates the recombination behavior observed among mutants. One method of recombination analysis utilizes so-called conditional-lethal types, where the cells are infected with two variants and the recombinants are selected after application of nonpermissive conditions (two-factor crosses). This allows the mutants to be organized into complementation groups with the relative positions of mutations being placed on a linear map. Another method is called three-factor crosses. Here three mutations are employed, with crossing-over occurring between two mutations while the third mutation is not selected. This allows for determination of linkage relationships among mutants and of the order of marker mutations. Due to reassortment, both the two-factor and the three-factor crosses are of less use in segmented DNA viruses.

DNA viruses of eukaryotes also recombine their genomic material. For instance, herpes simplex virus (HSV) was found to support recombination while using pairs of temperature-sensitive mutants (two-factor crossings). In fact, a recombination-

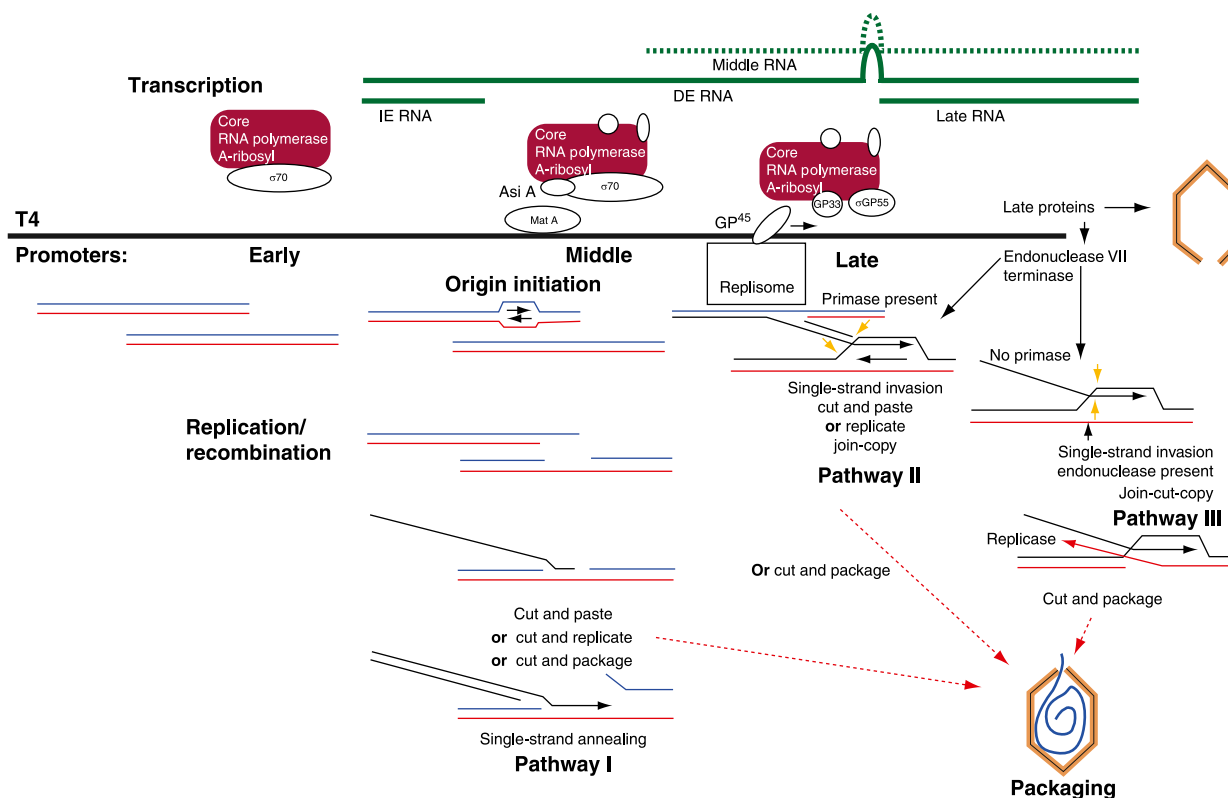


Fig. 1 Diagram of the relationship between the Enterobacteria phage T4 transcriptional pattern and the different mechanisms of DNA replication and recombination. (a) Shows the transcripts initiated from early, middle, and late promoters by sequentially modified host RNA polymerase. Hairpins in several early and middle transcripts inhibit the translation of the late genes present on these mRNAs. (b) Depicts the pathways of DNA replication and recombination. Hatched lines represent strands of homologous regions of DNA, and arrows point to positions of endonuclease cuts. Reproduced from Mosig G (1998) Recombination and recombination-dependent DNA replication in enterobacteria phage T4. *Annual Review of Genetics* 32: 379–413, with permission from Annual Reviews.

dependent mechanism of HSV-1 DNA replication has been described. The recombination frequency was proportional to the distance between mutations which suggested the lack of specific signal sequences responsible for the crossing-over. By using three-factor crossing, the HSV system involved two *ts* mutants and a syncytial plaque morphology as an unselectable marker. Similarly, in case of adenoviruses, the host range determined by the helper function of two mutations has been used as a third marker between *ts* mutants. Here, intertypic crosses between *ts* mutants have been identified based on segregation patterns and the restriction enzyme polymorphism.

Epstein-Barr virus (EBV) is a member of the family *Herpesviridae*, and it carries a long double-stranded genomic DNA, that shows a high-degree variation among strains. These variations include single base changes, restriction site polymorphism, insertions, or deletions. Based on tracking these mutations, it was found that some EBV strains arose due to DNA recombination.

Poxviruses represent the largest DNA viruses known (except those of algae and the mimivirus). Homologous recombination was detected in the genome of vaccinia virus (VV), based on the high frequency of intertypic crossovers, the marker rescue, and the sequencing of recombinants. These processes could be both intra- and intermolecular, and they depend on the size of the DNA target. It has been suggested that either viral DNA replication itself or the activity of the viral DNA polymerase might participate in VV DNA recombination. Indeed, some VV proteins with DNA strand transfer activity have been identified.

The DNA genome of Simian virus 40 (SV40, *Papovaviridae*) was found to recombine in somatic cells. The artificially constructed recombinant circular oligomers were used to find high general recombination frequency of SV40 DNA. However, homologous recombination events were rare.

Among plant DNA viruses, genetic recombination was studied in case of geminiviruses and caulimoviruses. The geminiviruses carry a single-stranded DNA genome, composed of either one or two circular DNA molecules. Frequent intermolecular crossing-over events were observed by using mutant combinations. Homologous crossovers were detected to occur intramolecularly between tandem repeats of a geminivirus DNA using agro-infected tobacco plants. The mechanism may involve either homologous crossing-over events or copy-choice processes that rely on template switching by DNA replicase. Moreover, deletions, insertions, and more profound rearrangements have been detected in the geminivirus DNA. These are the illegitimate recombination processes that may involve aberrant breakage-and-religation events or errors in DNA replication, that could occur either inter- or intramolecularly.

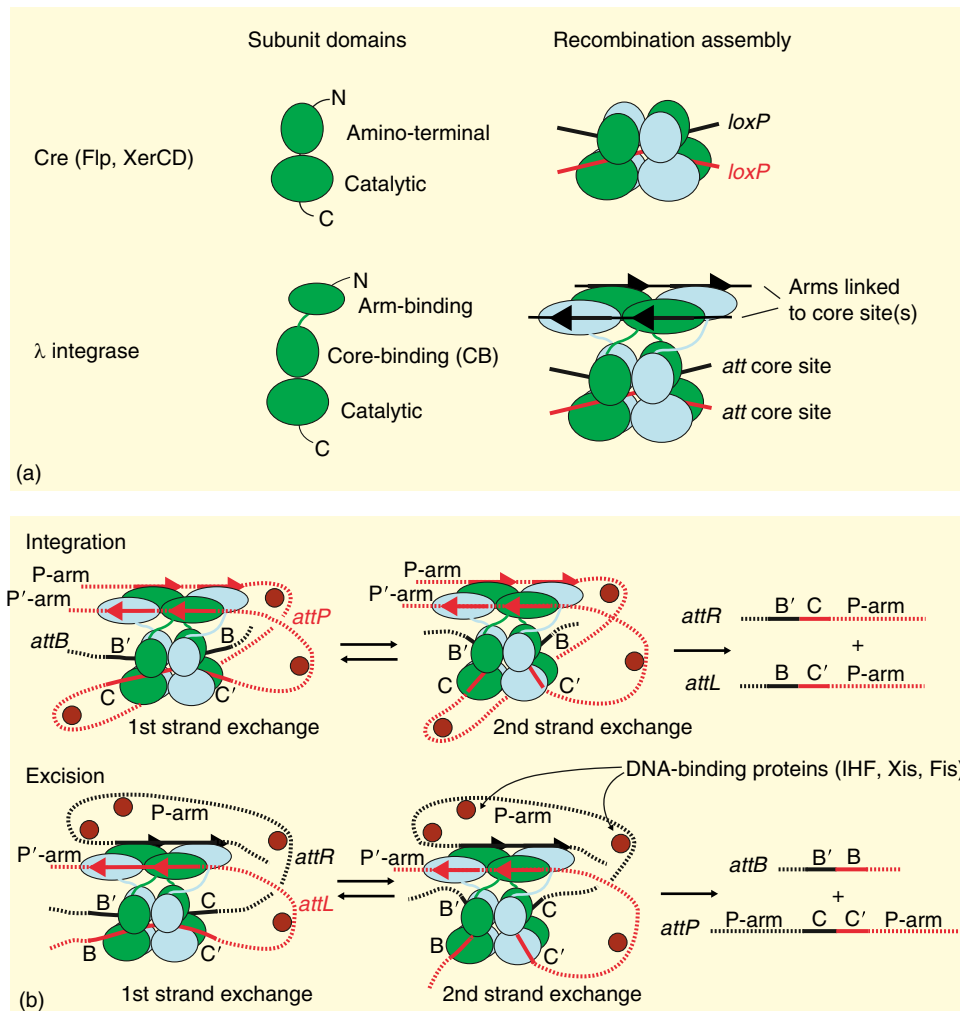


Fig. 2 (a) Enterobacteria phage λ integrase compared to the simpler recombinases. Tyrosine recombinases such as Cre have two domains that bind the core recombination sites and carry out recombination on their own. Enterobacteria phage λ integrase has a third, amino-terminal 'arm binding' domain that binds to the arm region of the attachment site. The DNA complex cartoon for Enterobacteria phage λ integrase (lower right) represents the new crystal structures. (b) Integration and excision by Enterobacteria phage λ integrase. The first and second strand exchange cartoons represent the first and second halves of the recombination reaction, respectively. In the first half of integration, for example, Enterobacteria phage λ integrase brings *attP* and *attB* sites together and exchanges the first pair of strands to generate a Holliday junction intermediate. In the second half of the reaction, the Holliday intermediate has isomerized to form a distinct quaternary structure and exchange of the second pair of strands generates recombinant *attL* and *attR* products. Reproduced from Van Duyne GD (2005) Enterobacteria phage λ integrase: Armed for recombination. *Current Biology* 15: R658–R660, with permission from Elsevier.

Cauliflower mosaic virus (CaMV) belongs to a family of plant double-stranded (ds) DNA pararetroviruses that replicate via reverse transcription. A high recombination rate was observed during CaMV infection *in planta*. These crossovers could occur at the DNA level (thus in the nucleus) or at the RNA level (thus more likely during reverse transcription in cytoplasm). However, features such as recombinational hot spots and mismatch repair might indicate replicative (i.e., RNA) step, whereas mismatch repair can occur due to the formation of heteroduplex intermediates and thus suggest DNA recombination. These data further suggest that CaMV has the recombination mechanisms available at both steps of its life cycle. Recombination between CaMV variants and the CaMV transgenic mRNAs has been reported and this is believed to represent the RNA–RNA recombination events that happen during reverse transcription.

Recombination in RNA Viruses

RNA is the genetic material in RNA viruses, and a high mutation rate has been observed for the viral RNA genome. This likely occurs during RNA replication by means of action of an RNA-dependent RNA polymerase enzyme due to either replication errors or because of the replicase switching among viral RNA templates. The terms of classic population genetics do not describe RNA

viruses. A better description of RNA viral populations is provided with a term 'quasispecies' that has been proposed to address a distribution of RNA variants in the infected tissue.

Many of the RNA viruses limit their life cycle to cytoplasm and thus the observed recombination events among RNAs of plus-stranded RNA viruses must occur outside the nucleus. In general, the RNA crossing-over processes are categorized as being either homologous or nonhomologous, but some earlier authors proposed that there are homologous, aberrant homologous, and nonhomologous RNA recombination types. Aberrant homologous recombination involves crossovers between related RNAs, but the crosses occur at not-corresponding sites leading to sequences insertions or deletions. More recently, mechanistic models were utilized to define the following RNA–RNA recombination classes: (1) The 'similarity-essential' recombination, where substantial sequence similarity between the parental RNAs is required as the major RNA determinant; (2) The 'similarity-nonessential' recombination does not require sequence similarity between the parental RNAs, although such regions may be present; and (3) There is the 'similarity-assisted' recombination where sequence similarity can influence the frequency or the recombination sites but additional RNA determinants are also critical.

Genetic RNA recombination has been described in many RNA virus groups. In particular, sequence data reveal RNA rearrangements reflecting RNA–RNA crossover events during RNA virus evolution. For instance, RNA rearrangements were demonstrated in the genomes of dengue virus-type I, flock house virus, hepatitis D virus, bovine viral diarrhea virus, and equine arthritis virus RNA. For plant RNA viruses, this has been demonstrated in potyviruses such as yam mosaic virus, sugarcane yellow leaf virus, and luteoviruses. Experimentally, RNA recombination has been shown to occur in picornaviruses, coronaviruses, or alphaviruses and in the following plant viruses: plum pox virus, cowpea chlorotic mottle virus, alfalfa mosaic virus, cucumber mosaic virus, tobacco mosaic virus, turnip crinkle virus (TCV), and tomato bushy stunt virus (TBSV). It has also been demonstrated in enterobacteria phage Qbeta, in negative RNA viruses, in double-stranded RNA viruses, and in retroviruses, as well as during formation of defective-interfering (DI) RNAs.

Recombination by reassortment was demonstrated for multisegmental animal RNA viruses, such as influenza virus, and in double-stranded reoviruses and orbiviruses. Specifically, the interpretation of two-factor crosses (using, e.g., ts mutants) in reoviruses turned out to be difficult due to recombination. The mutant sites cannot be ordered on a linear map and often no linkage between mutants could be detected.

Interestingly, there are examples of viral RNA recombination with host-derived sequences. These include the presence of unigenic-coding region in bovine diarrhea virus, a sequence from 28S rRNA found in the hemagglutinin gene of influenza virus or a tRNA sequence in Sindbis virus RNA. Also, in plant viruses the host-derived sequences were found in potato leaf-roll virus isolates that carry sequences homologous to an exon of tobacco chloroplast. Chloroplast sequences were found in the actively recombining RNAs of brome mosaic virus (BMV). Several plant viruses were also confirmed to recombine with viral RNA fragments expressed in transgenic plants, including cowpea chlorotic mottle virus, red clover necrotic mottle virus, potato virus Y virus, and plum pox virus.

The existence of several RNA virus recombination systems has made possible the studies of the molecular mechanisms of RNA recombination. The majority of RNA recombination models predict copy-choice mechanisms, either due to primer extension (in flaviviruses, carmoviruses), at the subgenomic promoter regions (BMV, poliovirus), or by strand translocation (in nidoviruses). In retroviruses, there are three copy-choice mechanisms: (1) forced (strong stop) strand transfer, (2) pause-driven strand transfer, and (3) pause-independent (RNA structure-driven) strand transfer. However, in enterobacteria phage Qbeta a breakage-and-religation mechanism has been described. Details of some of these systems are discussed below.

The molecular mechanisms of the formation of both nonhomologous and homologous RNA recombinants have been studied using an efficient system of BMV. In order to increase recombination frequency, the BMV RNA3-based constructs were generated where the 3' noncoding region was extended, while carrying partial deletions. This debilitated the replication of RNA such that the sequence got repaired by recombination with the sequences of other BMV RNA segments. It appeared that short base-paired regions between the two parental BMV RNA molecules could target efficient nonhomologous crossovers. A proposed model predicted that the formation of local RNA–RNA heteroduplexes could function because they brought together the RNA substrates and because they slowed down the approaching replicase enzyme complexes.

These early studies also analyzed the molecular requirements of homologous recombination by inserting the BMV RNA2-derived sequences into the recombination vector. This revealed the accumulation of both precise and imprecise RNA2–RNA3 recombinants and that the recombination frequencies depended upon the composition of nucleotide sequences within the region of recombination. The crossovers tended to happen at stretches of GC-rich regions alternating with AU-rich sequences suggesting the RNA replicase switching between RNA templates. Elements capable of forming strand-specific, stem-loop structures were inserted at the modified 3' noncoding regions of BMV RNA3 and RNA2 in either positive or negative orientations, and various combinations of parental RNAs were tested for patterns of the accumulating recombinant RNA3 components. This provided experimental evidence that homologous recombination between BMV RNAs more likely occurred during positive-rather than negative-strand synthesis.

True homologous recombination crossing-over has been observed among the RNA molecules of the same segments during BMV infections. By using nonselective marker mutations at several positions, it was demonstrated that RNA1 and RNA2 segments crossed-over at 5–10% frequency, whereas the intercistronic region in RNA3 supported an unusually high recombination frequency of 70%. The subsequent use of various deletion constructs has revealed that the high-frequency crossing-over mapped to the subgenomic promoter (sgp) region, and in particular to its internal polyA tract. Further studies have shown that sgp-mediated crossing-over has occurred at the minus-strand level (i.e., during plus-strand synthesis), most likely by discontinuous process,

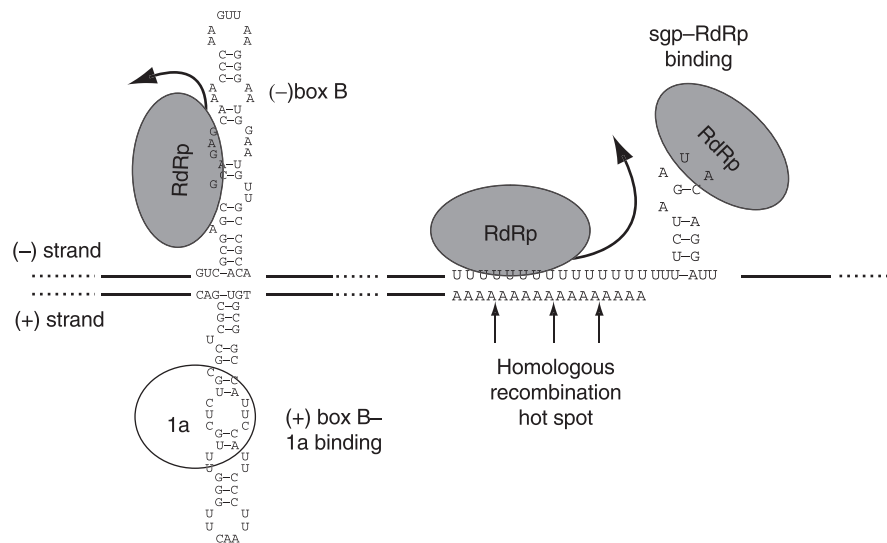


Fig. 3 Model illustrating the synthesis of sg RNA3a in view of multiple functions of the intergenic region in (–) RNA3. The BMV RdRp enzyme complex (represented by gray ovals) migrates alongside the (–)-strand RNA template and pauses (represented by curved arrows) at the secondary structure or, most notably, at the oligoU tract, leading to the formation of subgenomic sgRNA3a. Yet another molecule of the RdRp enzyme binds to the sgp and initiates the *de novo* synthesis of sgRNA4. Also, the rehybridization of the sgRNA3a oligoA tail to the RNA3 (–) template can resume full-length copying which primes the observed RNA3–RNA3 recombination (5, 69). The (+) and (–) RNA strands are represented by thick lines and both the oligoU tract in the (–) strand template and the oligoA 3' termini are exposed. The stem-and-loop structures adopted by the (+) and the (–) strands upstream to their oligoU (A) tracts are shown. The binding region to protein 1a via the box B of the stem-loop structure in (+) strands is shown. Reproduced from Wierchoslawski R, Urbanowicz A, Dzianott A, Figlerowicz M, and Bujarski JJ (2006) Characterization of a novel 5' subgenomic RNA3a derived from RNA3 of brome mosaic bromovirus. *Journal of Virology* 80: 12357–12366, with permission from American Society for Microbiology.

where the replicase complex detached from one strand and reinitiated on another strand. This process is most likely primed by a sg RNA3a intermediate, which prematurely terminates on the polyU (in minus strand) tract and re-anneals to this region on another minus RNA3 template (Fig. 3). Also, it turned out that the frequency of crossing-over and the process of initiation of transcription of sg RNA4 were reversely linked, suggesting a competition between these two reactions.

The role of replicase proteins in BMV RNA recombination has been studied by using well-characterized 1a and 2a protein mutants. A ts mutation in protein 1a 5' shifted the crossover sites indicating the participation of helicase domain of 1a. Likewise, mutations at several regions of 2a affected the frequency of nonhomologous recombination. The relationship between replication and recombination was studied by using BMV variants that carried mutations in 1a and 2a genes. This revealed that the 1a helicase and the 2a N-terminal or core domains were functionally linked during both processes *in vivo* and *in vitro*. Also, it was shown that the characteristics of homologous and nonhomologous crossovers could be modified separately by mutations at different protein sites. All these studies confirmed the involvement of replicase proteins in recombination and supported the template-switching model.

More recently, the role of host factors in BMV recombination was addressed by using both yeast and *Arabidopsis* systems. In yeast, transient co-expression of two derivatives of BMV genomic RNA3 supported intermolecular homologous recombination at the RNA level but only when parental RNAs carried the *cis*-acting replication signals. The results implied that recombination occurred during RNA replication. In *Arabidopsis*, the use of gene-knock-out mutations in the RNA interference pathway revealed that BMV can recombine according to both the copy-choice template-switching and to the breakage-and-religation mechanisms.

The role of replicase proteins in RNA recombination has also been studied in other RNA viruses. For TCV, a small single-stranded RNA virus, a high-frequency recombination was observed between satellite RNA D and a chimeric subviral RNA C. The crossing-over most likely relied on viral replicase enzyme switching templates during plus-strand synthesis of RNA D which reinitiated RNA elongation on the acceptor minus-strand RNA C template. The participation of replicase proteins was demonstrated *in vitro*, where a chimeric RNA template containing the *in vivo* hot-spot region from RNA D joined to the hot-spot region from RNA C. Structural elements such as a priming stem in RNA C and the replicase binding hairpin, also from RNA C, turned to play key roles during recombination, probably reflecting late steps of RNA recombination such as strand transfer and primer elongation. The host factors related to the host-mediated viral RNA turnover have been found to participate in tombusvirus RNA recombination. The screening of essential yeast genes mutants identified host genes that affected the accumulation of TBSV recombinants, including genes for RNA transcription/metabolism, and for protein metabolism/transport. Suppression of TBSV RNA recombination was observed by the yeast Xrn1p 5'–3' exoribonuclease, likely due to rapid removal of the 5' truncated RNAs, the substrates of recombination. These 5' truncated viral RNAs are generated by host endoribonucleases, such as the Ngl2p endoribonuclease.

Coronavirus RNAs were found to recombine between the genomic and DI RNA molecules. It was postulated that recombination has occurred due to the nonprocessive nature of the coronavirus RNA polymerase enzyme (Fig. 4) and an efficient protocol for targeted recombination has been developed.

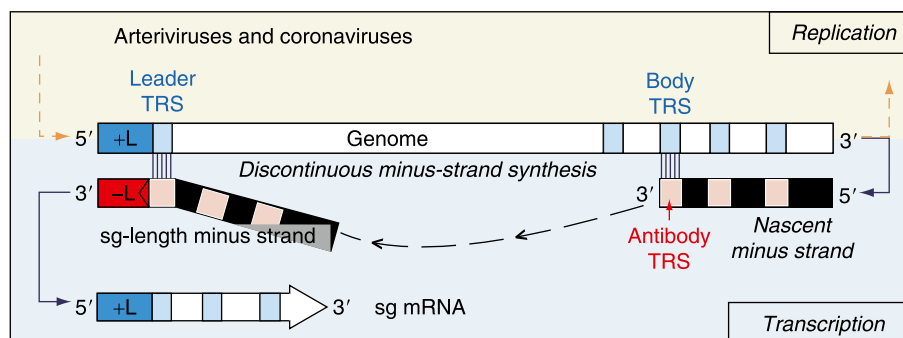


Fig. 4 Models for discontinuous transcription from minus-strand sg-length templates in arteriviruses and coronaviruses. These viruses have a common 59 leader sequence on all viral mRNAs. Discontinuous extension of minus-strand RNA synthesis has been proposed as the mechanism to produce sg-length minus-strand templates for transcription. The replicase/transcriptase can attenuate at one of the body TRSs in the 3'-proximal part of the genome, after which the nascent minus strand extends with the anti-leader ('L') sequence. Next, the completed sg-length minus strands serve as templates for transcription. Reproduced from Pasternak AO, Spaan WJM, and Snijder EJ (2006) Nidovirus transcription: How to make sense...? *Journal of General Virology* 87: 1403–1421, with permission from Society for General Microbiology.

Similarly, in nodaviruses, the two-partite RNA viruses, recombination processes were found to occur between RNA segments at a site that potentially could secure base pairing between the nascent strand and the acceptor template. The recombination sites might have been chosen based on factors such as the similarity to the origin of replication or special secondary structures. A postulated model implies the polymerase to interact directly with the acceptor nodavirus RNA template.

A double-stranded RNA *Pseudomonas* phage Phi6 was hypothesized to recombine its RNA based on a copy-choice template switching mechanism, where the crossovers would have occurred inside the virus capsid structures at regions with almost no sequence similarity. Interestingly, the frequency of recombination was enhanced by conditions that prevented the minus-strand synthesis. Experiments were designed to reveal the effects of drift on existing genetic variation by minimizing the influence of variation on beneficial mutation rate. The segmented genome of the *pseudomonas* phage Phi6 has allowed to present the first empirical evidence that the advantage of sex during adaptation increases with the intensity of drift.

The enterobacteria phage Qbeta, a small single-stranded RNA virus, could recombine both *in vivo* and *in vitro*. Here, the mechanism of recombination was not based on a template-switching by the replicase, but rather via a replicase-mediated splicing-type religation of RNA fragments. The system produced nonhomologous recombinants, whereas the frequency of homologous crossovers was low. These data suggested an RNA trans-esterification reaction catalyzed by a conformation acquired by enterobacteria phage Qbeta replicase during RNA synthesis. In summary, the results on various plus-strand RNA virus systems demonstrate the availability of a variety of template-switch mechanisms, the mutual-primer-extension on two overlapping RNA strands, the primer-extension on one full-length RNA strand, as well as both replicative and nonreplicative trans-esterification mechanisms where a piece of another RNA is added to the 3' terminus of an RNA either by viral RdRp or by other enzymes (e.g., RNA ligase), respectively.

The recombination events in retroviruses contribute significantly to genetic variability of these viruses. The crossovers do occur by reverse transcriptase jumpings between the two genomic RNA molecules inside virion capsids. Apparently, the virally encoded reverse transcriptase enzymes have been evolutionarily selected to prone the jumpings between templates during reverse transcription. It turns out that the recombinant jumpings between RNA templates are responsible for both inter- and intramolecular template switchings and also for the formation of defective retroviral genomes. It has been found that the most stable interactions between two copies of retrovirus RNAs were within the 5' nucleotides 1–754. There is experimental evidence demonstrating that the template 'kissing' interactions effectively promote recombination within the HIV-1 5' untranslated region. The possibilities of recombination in retroviruses at the DNA level (of the integrated provirus sequences) were discussed earlier in this article.

Defective-Interfering RNAs

There is a variety of subviral RNA molecules that are linked to viral infections. Those derived from the viral genomic RNAs and interfering with the helper virus accumulation or symptom formation are called as DI RNAs. First reports (in 1954) about DI RNAs coexisting with viral infection was provided by Paul von Magnus with influenza virus. Thereafter, numerous both animal and plant viruses were found to generate DI RNAs. Naturally occurring DI RNAs have been identified in coronavirus infections. These molecules appear to arise by a polymerase strand-switching mechanism. The leader sequence of the DI RNAs was found to switch to the helper-virus derived leader sequence, indicating that helper virus-derived leader was efficiently utilized during DI RNA synthesis. Also, the leader switching likely occurred during positive-strand DI RNA synthesis, and the helper-virus positive-strand RNA synthesis tended to recognize double-stranded RNA structures to produce positive-strand DI RNAs. The parts of the coronavirus RNA required for replication and packaging of the defective RNAs were investigated, with both the 5'- and the 3'-terminal sequences being necessary and sufficient. The coronavirus DI RNAs have been utilized to study the mechanism of site-specific RNA recombination. This process relies on the acquisition of a 5' leader that is normally used for production of numerous

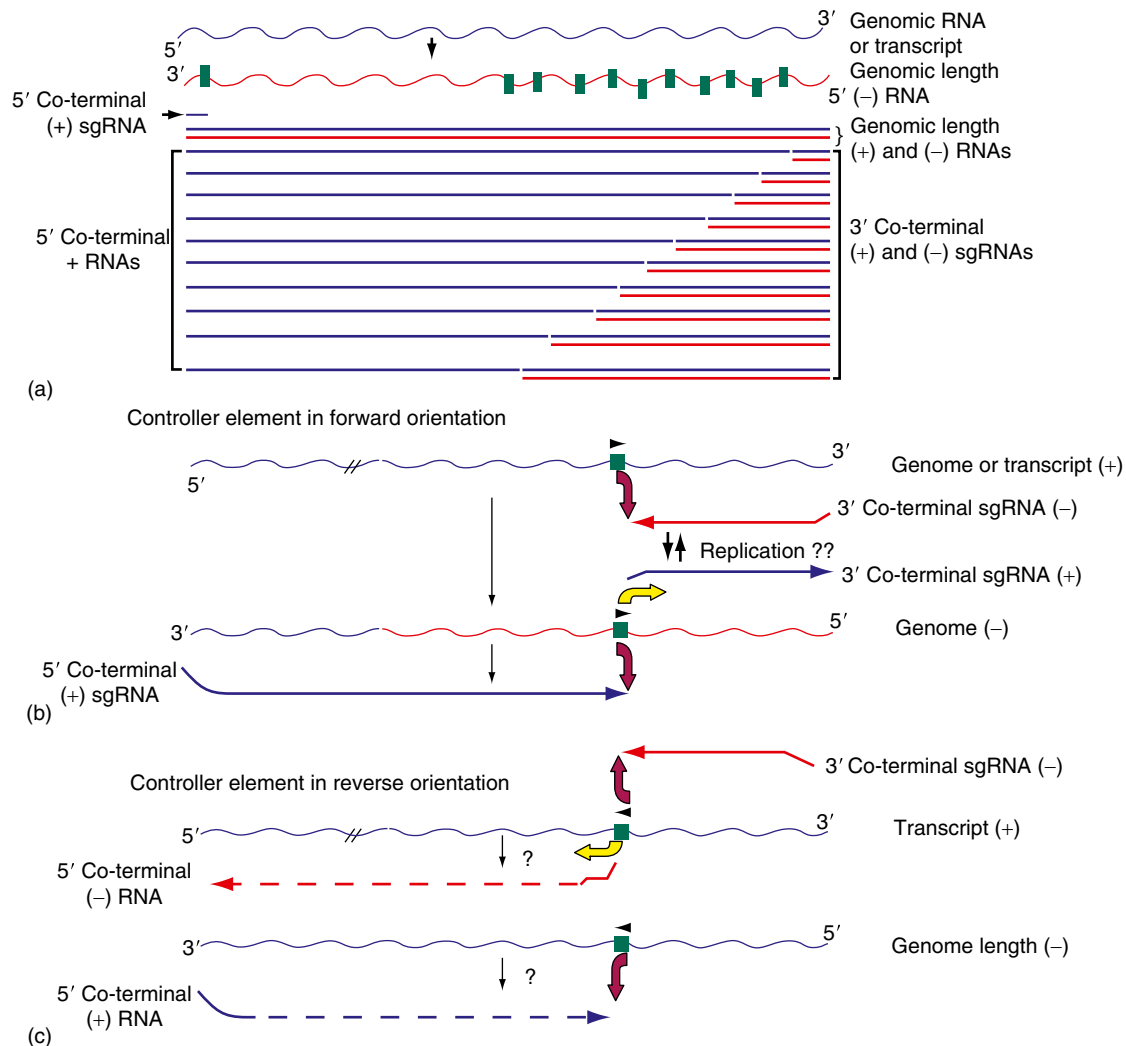


Fig. 5 (a) The outline of different species of genomic RNA and 59 and 39 terminal sgRNAs potentially produced in CTV-infected cells. The positive-sense RNAs are shown in blue and the negative-sense RNAs are shown in red. The wavy line represents the genomic RNA or the plus-sense transcript (blue) and the genomic length minus-sense RNA (red) produced from the plus-sense RNA. The solid green boxes on the genomic negative-stranded RNA represent the sgRNA controller elements. The solid lines represent the full array of plus- and minus-stranded genomic and 39- and 59-terminal sgRNAs potentially produced during replication of CTV. (b, c) Models predicting the generation of 59- and 39-terminal positive- and negative-sense sgRNAs with the controller element present in normal and reverse orientation. One control region is shown for clarity. The wavy blue line represents the transcript (blue) containing the control region (green box) in normal and reverse orientation (the direction of the arrowheads above the controller element indicates the orientation of the controller element). The thick curved arrows represent the transcription termination (vertical direction, red) or promotion (horizontal direction, yellow). The solid blue lines with arrowheads represent the positive-sense 39- and 59-terminal sgRNAs and the solid red line with arrowhead indicates the 39-terminal negative-sense sgRNA. The dashed lines with arrowheads indicate the potential 59 terminal positive (blue)- and negative (red)-sense sgRNAs. Reproduced from Gowda S, Satyanarayana T, Ayllon MA, *et al.* (2001) Characterization of the *cis*-acting elements controlling subgenomic mRNAs of *Citrus tristeza virus*: Production of positive- and negative-stranded 39-terminal and positive-stranded 59 terminal RNAs. *Virology* 286: 134–151, with permission from Elsevier.

coronavirus sg RNAs. Also, these DI RNAs have been used as vehicles for the generation of designed recombinants from the parental coronavirus genome.

In case of plant viruses, tombusviruses and carmoviruses were found to accumulate DI RNAs, which maintain a consistent pattern of rearranged genomic sequences flanked by the 5' and 3' unchanged replication signals. In some cases, the base pairing between a partial nascent strand and the acceptor template can lead to the appearance of the rearranged regions in DI RNAs.

In addition to rearranged DI RNAs, some RNA viruses accumulate defective RNAs due to a single internal deletion in the genomic RNA of the helper virus. Such examples include beet necrotic wheat mosaic furovirus, peanut clump furovirus, clover yellow mosaic potyvirus, sonchus yellow net rhabdovirus, and tomato spotted wilt tospovirus. Features such as the ability to translate or the magnitude of the defective RNA seem to affect the selection of the best-fit sizes of DI RNAs during infection.

Another type of single-deletion DI RNAs are produced during broad bean mottle bromovirus (BBMV) infections from the RNA2 segment. A model has been proposed where local complementary regions bring together the remote parts of RNA2 which then facilitates the crossover events. Similar RNA2-derived DI RNAs have been reported to accumulate during the cucumovirus infections.

The closteroviruses, the largest known plant RNA viruses, form multiple species of defective RNAs, including the citrus tristeza virus defective RNAs that arise from the recombination of a subgenomic RNA with distant 5' portion of the virus genomic RNA (Fig. 5). Apparently, closteroviruses can utilize sg RNAs for the rearrangement of their genomes.

Negative-strand RNA viruses also form DI RNAs. For instance, in vesicular stomatitis virus (VSV), a rhabdovirus, the *cis*-acting RNA replication terminal elements participate in the formation of the 5'-copy-back DI RNAs, reflecting likely communication between distant portions of the VSV genome.

Summary and Conclusions

Genetic recombination is a common phenomenon among both DNA and RNA viruses. The recombination events have been observed based on natural rearrangements of the sequenced viral genomes. Also, experimental systems demonstrate the occurrence of recombination events that play important roles in securing the genetic diversity during viral infection. Different molecular mechanisms are involved in DNA versus RNA viruses. Many DNA viruses utilize host cellular machinery of general homologous recombination (such as meiotic crossing-over), whereas some encode their own proteins that are responsible for recombination. In addition, certain groups of DNA viruses support site-specific (nonhomologous) recombination events. In general, the virus DNA recombination mechanisms seem to involve post-DNA replication molecular events.

For RNA viruses the majority of known homologous and nonhomologous RNA recombination events appear to be integrally linked to RNA replication machinery. Various types of copy-choice (template-switching) mechanisms were proposed to describe the easy formation of RNA recombinants in numerous RNA virus systems. The roles of both special RNA signal sequences and viral proteins have been elucidated, reflecting the variety of the recombination strategies used by RNA virus groups. The involvement of host cell genes in RNA virus recombination has begun to get elucidated in several RNA viruses. Besides replicational copy-choice mechanisms, some RNA viruses use the breakage-and-religation mechanism where viral RNA gets regenerated by religation from RNA fragments, as shown experimentally for Enterobacteria phage Qbeta. New venues of RNA recombination research just emerge including our better understanding of the involvement of RNA *cis*-acting signals, the role of RNA replication, and the importance of cellular host genes such as RNA ribonucleases or RNA interference.

See also: Bean Common Mosaic Virus and Bean Common Mosaic Necrosis Virus (*Potyviridae*). Brome Mosaic Virus (*Bromoviridae*). Cotton Leaf Curl Disease (*Geminiviridae*). Evolution Steered by Structure. Mechanisms of DNA Virus Evolution. Mechanisms of RNA Virus Evolution. Tomato Yellow Leaf Curl Viruses (*Geminiviridae*)

Further Reading

- Agol, V.I., 2006. Molecular mechanisms of poliovirus variation and evolution. *Current Topics in Microbiology and Immunology* 299, 211–259.
- Briddon, R.W., Stanley, J., 2006. Subviral agents associated with plant single-stranded DNA viruses. *Virology* 344, 198–210.
- Chetverin, A.B., 2004. Replicable and recombinogenic RNAs. *FEBS Letters* 567, 35–41.
- Cromie, G.A., Connelly, J.C., Leach, D.R., 2001. Recombination at double-strand breaks and DNA ends: Conserved mechanisms from phage to humans. *Molecular Cell* 8, 1163–1174.
- Figlerowicz, M., Bujarski, J.J., 1998. RNA recombination in brome mosaic virus, a model plus strand RNA virus. *Acta Biochimica Polonica* 45 (4), 847–868.
- Galetto, R., Negroni, M., 2005. Mechanistic features of recombination in HIV. *AIDS Reviews* 7 (2), 92–102.
- Gowda, S., Satyanarayana, T., Ayllon, M.A., *et al.*, 2001. Characterization of the *cis*-acting elements controlling subgenomic mRNAs of *Citrus tristeza virus*: Production of positive- and negative-stranded 39-terminal and positive-stranded 59 terminal RNAs. *Virology* 286, 134–151.
- Koonin, E.V., Senkevich, T.G., Dolja, V.V., 2006. The ancient Virus World and evolution of cells. *Biology Direct* 1 (29), 1–27.
- Masters, P.S., Rottier, P.J., 2005. Coronavirus reverse genetics by targeted RNA recombination. *Current Topics in Microbiology and Immunology* 287, 133–159.
- Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Ruger, W., 2003. Bacteriophage T4 genome. *Microbiology and Molecular Biology Reviews* 67, 86–156.
- Miller, W.A., Koev, G., 1998. Getting a handle on RNA virus recombination. *Trends in Microbiology* 6, 421–423.
- Mosig, G., 1998. Recombination and recombination-dependent DNA replication in Enterobacteria phage T4. *Annual Review of Genetics* 32, 379–413.
- Nagy, P.D., Simon, A.E., 1997. New insights into the mechanisms of RNA recombination. *Virology* 235, 1–9.
- Noueiry, A.O., Ahlquist, P., 2003. Brome mosaic virus RNA replication: Revealing the role of the host in RNA virus replication. *Annual Review of Phytopathology* 41, 77–98.
- Pasternak, A.O., Spaan, W.J.M., Snijder, E.J., 2006. Nidovirus transcription: How to make sense...? *Journal of General Virology* 87, 1403–1421.
- Poon, A., Chao, L., 2004. Drift increases the advantage of sex in RNA bacteriophage Phi6. *Genetics* 166, 19–24.
- Van Duynne, G.D., 2005. Lambda integrase: Armed for recombination. *Current Biology* 15, R658–R660.
- Weigel, C., Seitz, H., 2006. Bacteriophage replication modules. *FEMS Microbiology Reviews* 30, 321–381.
- White, K.A., Nagy, P.D., 2004. Advances in the molecular biology of tombusviruses: Gene expression, genome replication, and recombination. *Progress in Nucleic Acid Research and Molecular Biology* 78, 187–226.
- Wierzboslawski, R., Urbanowicz, A., Działot, A., Figlerowicz, M., Bujarski, J.J., 2006. Characterization of a novel 5' subgenomic RNA3a derived from RNA3 of Brome Mosaic Virus. *Journal of Virology* 80, 12357–12366.
- Wilkinson, D.E., Weller, S.K., 2003. The role of DNA recombination in herpes simplex virus DNA replication. *IUBMB Life* 55 (8), 451–458.
- Worobey, M., Holmes, E.C., 1999. Evolutionary aspects of recombination in RNA viruses. *Journal of General Virology* 80, 2535–2543.

Assembly of Viruses: Enveloped Particles

CK Navaratnarajah, Ranjit Warriar, and Richard J Kuhn, Purdue University, West Lafayette, IN, United States

© 2008 Elsevier Ltd. All rights reserved

This is a reproduction of C.K. Navaratnarajah, R. Warriar, and R.J. Kuhn, Assembly of Viruses: Enveloped Particles, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00667-1.

Glossary

Core-like particle (CLP) Subviral particles assembled from recombinantly expressed capsid protein are referred to as core-like particles. They are morphologically similar to authentic cores isolated from viruses or infected cells.

Cryo-EM Cryo-electron microscopy and image reconstruction techniques are used to elucidate the structures of viruses and other macromolecular structures.

Immature virion Viruses usually produce noninfectious particles which require a maturation step in order to form the infectious, mature virus. The maturation step often involves the proteolytic cleavage of a precursor protein.

Nucleocapsid core (NC) Capsid protein packages the nucleic acid genome to form a stable protein–nucleic acid complex which is then enveloped with a lipid bilayer.

Introduction

Viruses have long been distinguished by their physical features, usually visualized by electron microscopy or analyzed biochemically. One feature that has been frequently used to categorize viruses is the presence or absence of a lipid bilayer. Many animal viruses are surrounded by a lipid bilayer that is acquired when the nucleocapsid buds through cell membranes, usually at a late stage of virus assembly. While the protein coat of nonenveloped viruses plays a crucial role in protecting the genome from the environment, for enveloped viruses the lipid membrane partially fulfills this role. The lipid membranes are decorated with virus-encoded envelope proteins that are important for the subsequent infectivity of the virus, although some viruses also incorporate cellular proteins in their membrane. Virus envelopment can take place after the assembly of an intact nucleocapsid structure (betaretroviruses) or capsid assembly and envelopment can occur concomitantly (orthomyxovirus). Specific or nonspecific interactions between the viral envelope glycoproteins and the proteins that make up the nucleocapsid mediate the envelopment of the core or nucleoprotein–nucleic acid complex. Enveloped viruses acquire their lipid bilayer from a variety of locations within the cell, but a given virus will usually bud from one specific cellular membrane (Table 1). Enveloped viruses can take advantage of the cellular secretory pathway in order to assemble and bud out of the cell. In contrast, nonenveloped viruses usually exit infected cells by disrupting the plasma membrane. Thus, budding provides enveloped viruses with a nonlytic method of exiting infected cells, and they must do so while the cell is still alive.

Viral Envelope

The main component of the viral envelope is the host-derived lipid bilayer. The precise composition of this lipid membrane varies, as different viruses acquire their envelopes from different cellular membranes. The choice of membrane from which the virus buds is often determined by the specific targeting and accumulation of the envelope proteins at a particular site in the secretory pathway (Table 1). There are examples of viruses that bud from the plasma membrane (togaviruses, rhabdoviruses, paramyxoviruses, orthomyxoviruses, and retroviruses), endoplasmic reticulum (ER) (coronaviruses and flaviviruses) and the Golgi complex (bunyaviruses). There are also examples of viruses that undergo transient envelopment and reenvelopment (herpesvirus).

Viral proteins are found embedded in the lipid membrane. The majority of these proteins are transmembrane glycoproteins. The viral envelope glycoproteins mediate the interaction of the virus with cell receptors and promote the fusion of the viral and cellular membranes during infection of susceptible cells. Viral glycoproteins are also crucial for the assembly of the virion. They can make important lateral contacts with each other, thus driving oligomerization and also capturing other viral components such as the capsid or matrix protein. The majority of enveloped viruses contain one or more glycoproteins that are usually found as oligomers embedded within the lipid membrane. High-resolution structural information is available for many glycoproteins such as the hemagglutinin and neuraminidase proteins of influenza A virus, the gp120 of HIV, and the E protein of dengue virus. Based on these structural and biochemical studies, it has been shown that most glycoproteins are primed for the conformational changes that are required in order to gain entry to the host cell during an infection.

Some enveloped viruses contain integral membrane proteins that have multiple membrane-spanning regions that oligomerize to form channels in the membrane. The influenza A virus M2 protein forms an ion channel and plays an important role in the assembly and entry of the virus particle. Some viruses such as the retroviruses also incorporate cellular membrane proteins into the viral envelope. In a majority of the cases the host proteins that are present at the sites of assembly or budding are incorporated in a passive, nonselective manner. However, there are examples where the virus actively recruits specific host proteins that may help in evading the defenses of the immune system or enhance infectivity.

Table 1 List of enveloped virus families and the origin of the envelope

Virus family ^a	Membrane
<i>Arenaviridae</i>	Plasma membrane
<i>Arterivirus</i>	Endoplasmic reticulum
<i>Asfarviridae</i>	Endoplasmic reticulum and plasma membrane
<i>Baculoviridae</i>	Plasma membrane
<i>Bunyaviridae</i> ^b	Golgi complex
<i>Coronaviridae</i>	ER Golgi intermediate compartment
<i>Cystoviridae</i> ^c	Plasma membrane
<i>Deltavirus</i>	Endoplasmic reticulum
<i>Filoviridae</i>	Plasma membrane
<i>Flaviviridae</i>	Endoplasmic reticulum
<i>Fuselloviridae</i> ^d	Plasma membrane
<i>Hepadnaviridae</i>	Endoplasmic reticulum
<i>Herpesviridae</i>	Nuclear envelope
<i>Hypoviridae</i>	Plasma membrane
<i>Iridoviridae</i>	Plasma membrane
<i>Lipothruxviridae</i> ^d	Plasma membrane
<i>Orthomyxoviridae</i>	Plasma membrane
<i>Paramyxoviridae</i>	Plasma membrane
<i>Plasmaviridae</i> ^c	Plasma membrane
<i>Polydnaviridae</i>	Plasma membrane
<i>Poxviridae</i>	ER Golgi intermediate compartment
<i>Retroviridae</i>	Plasma membrane
<i>Rhabdoviridae</i> ^b	Plasma membrane
<i>Togaviridae</i>	Plasma membrane

^aAnimal viruses unless otherwise noted.

^bThis group also includes plant viruses.

^cBacteriophage.

^dArchaea.

Icosahedral Enveloped Viruses: Alphaviruses and Flaviviruses

Alphaviruses, and more recently, flaviviruses have served as model systems to study the assembly and budding of simple enveloped viruses. These positive-strand RNA viruses consist of a single RNA genome that is encapsidated by multiple copies of a capsid protein to form the nucleocapsid core (NC). The envelopment of the NC is mediated by the interaction between the envelope glycoproteins and this core. The assembly and budding of these two simple enveloped viruses will be described in detail in order to present common themes in the assembly of icosahedral enveloped viruses.

Alphavirus Assembly

Alphavirus life cycle

Alphaviruses are members of the family *Togaviridae*, which also includes the genus *Rubivirus*. Alphaviruses enter the host cell by receptor-mediated endocytosis via the clathrin-coated endocytic pathway. Following fusion at low pH with the endosomal membrane, the NC is released into the cytoplasm. The NC has been proposed to uncoat by transfer of capsid proteins (CPs) to ribosomes. This releases the genome RNA into the cytoplasm which is translated to produce the nonstructural proteins. The nonstructural proteins transcribe a negative-sense copy of the genome RNA. This RNA serves as template for genomic and subgenomic RNA. The subgenomic RNA, which is synthesized in greater amounts than the genomic RNA, codes for the structural proteins of the virus. CP is found at the N-terminus of the structural polyprotein, followed by proteins PE2 (E3 + E2), 6K, and E1. Two hundred and forty copies of CP, E2, and E1 assemble to form the alphavirus virion (Fig. 1(a)). The transmembrane E1 glycoprotein functions during entry to mediate the fusion of the viral membrane with that of the endosomal membrane, while the transmembrane E2 glycoprotein is responsible for cell receptor binding. E1 and PE2, a precursor of E2 and E3, are processed together as a heterodimer in the ER and Golgi, and are transported to the cell surface in the form of spikes that are each composed of three heterodimers of E1/E2. E3 serves as a chaperone to promote the correct folding of E2, as well as to prevent the premature fusion of E1 in the acidic environment of the late Golgi. The maturation cleavage of PE2 to generate E3 and E2 by a furin-like protease in a late Golgi compartment primes the glycoprotein spike complex for subsequent fusion during virus entry. The function of 6K is unclear but it does promote infectivity of the particle. A single copy of the genome RNA is packaged by 240 copies of the CP to form an icosahedral NC in the cytoplasm of infected cells. The NC interaction with the E1/E2 trimeric spikes at the plasma membrane results in the budding of the mature virus from the cell membrane.

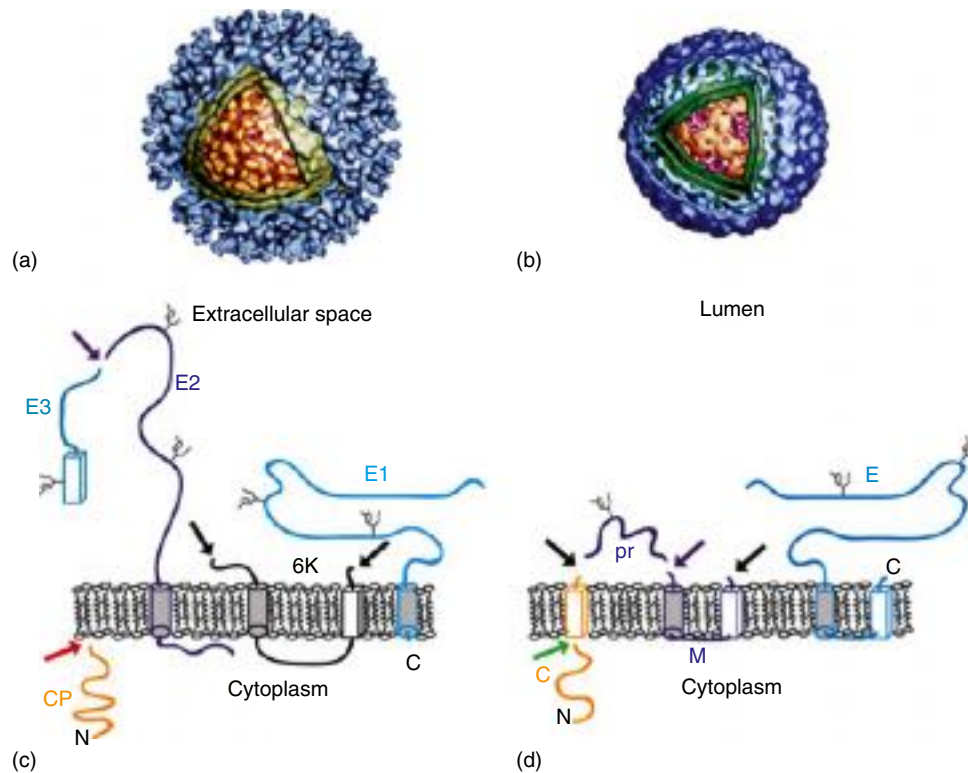


Fig. 1 Cryo-EM reconstructions and glycoprotein topology of alphaviruses and flaviviruses. (a) Surface shaded view of a Cryo-EM reconstruction of Ross River virus (alphavirus). The glycoprotein envelope shown in blue has been cut away to reveal the lipid bilayer (green) and the nucleocapsid core (orange). (b) Cryo-EM reconstruction of dengue virus (flavivirus). The glycoprotein envelope shown in blue has been cut away to reveal the lipid bilayer (green) and the nucleocapsid core (orange/purple). (c) The topology of the alphavirus glycoproteins is shown. The rectangular cubes represent signal sequences while the cylinders represent stop transfer sequences. The glycosylation sites are represented by branched structures. The arrows indicate cleavage sites. The black arrows indicate signalase cleavage sites, the purple arrow indicates the furin cleavage site, and the red arrow indicates the CP cleavage site. (d) The topology of the flavivirus structural proteins is shown. The rectangular cubes represent signal sequences while the cylinders represent stop transfer sequences. The glycosylation sites are represented by branched structures. The arrows indicate cleavage sites. The black arrows indicate signalase cleavage sites, the purple arrow indicates the furin cleavage site, and the green arrow indicates the NS2B/NS3 (viral protease) cleavage site.

Alphavirus virion structure

Cryo-electron microscopy (cryo-EM) and image reconstruction techniques have revolutionized the understanding of the molecular architecture of alphaviruses (Fig. 1(a)). Studies with Ross River, Semliki Forest, Venezuelan equine encephalitis, Aura, and Sindbis have shown that these viruses consist of an outer protein layer made up of the glycoproteins E1 and E2 (Fig. 1(a) and (c)). The membrane spanning regions of these glycoproteins traverse a host-derived lipid bilayer that surrounds the NC of the virus. The CP and glycoprotein layers interact with one another and are arranged symmetrically in a $T = 4$ icosahedral configuration. Fitting of the atomic coordinates of the ectodomain of E1 and amino acids 106–264 of the CP into the cryo-EM density of Sindbis virus allowed a pseudo-atomic model of the virus to be generated. The fitting of the E1 structure into the cryo-EM density reveals that E1 forms an icosahedral scaffold on the surface of the viral membrane. E1 is positioned almost tangential to the lipid bilayer, whereas E2 has a more radial arrangement. The bulk of E2 lies on top of E1 and caps the fusion peptide, thereby preventing premature fusion with cell membranes. This arrangement of the glycoproteins is in agreement with the function of each protein, where the surface-exposed E2 interacts with cellular receptors and protects E1 until it is required for fusion. The fusion peptide is only exposed when the E1–E2 heterodimer dissociates in the presence of low pH in the endosome. Fitting of amino acids 106–264 of the CP into the cryo-EM density of Sindbis virus showed that each subunit of the projecting pentamers and hexamers (known as capsomeres) observed in the NC layer is made up of the CP protease domain consisting of amino acids 114–264. There is very little interaction between amino acids 114–264 of the CP either within the capsomere or in between capsomeres. Thus, the major contributors to the stability of the NC in the absence of glycoproteins are CP–RNA and RNA–RNA interactions that take place in the RNA–protein layer below the projecting capsomeres.

Alphavirus assembly and budding

Alphavirus virions always contain an NC and it is likely that this promotes and is required for budding through direct interactions with the glycoproteins. Thus, the first step in assembly is for the alphavirus CP to specifically recognize and encapsidate the

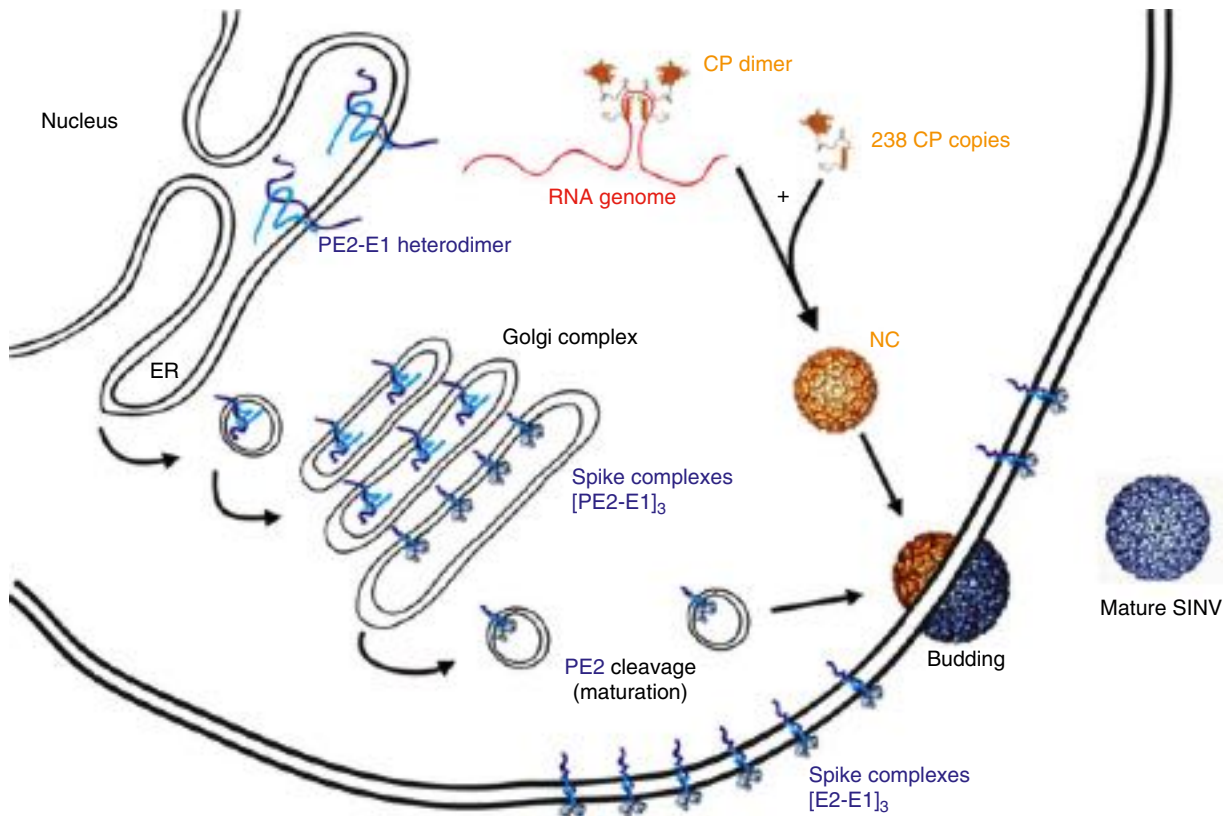


Fig. 2 Alphavirus assembly. Two capsid proteins (CPs) bind the encapsidation signal of the genome RNA to form a CP dimer-RNA complex in the cytoplasm. The CP dimer is stabilized by coiled-coil helix I interactions. The subsequent steps of nucleocapsid core assembly have not been elucidated but cores form and accumulate in the cytoplasm. The glycoproteins PE2 and E1 form heterodimers in the endoplasmic reticulum (ER). The glycoproteins are folded, glycosylated, and palmitoylated as they are transported through the ER and Golgi. The PE2-E1 heterodimers form spike complexes [PE2-E1]₃ in the Golgi. E3 is cleaved from the spikes by a furin-like protease before they are transported to the plasma membrane. The NC interacts with the cytoplasmic domain of E2 in the spike complexes, driving the budding of the mature virus from the plasma membrane.

genome RNA to form NCs in the cytoplasm of infected cells (Fig. 2). The N-terminus of the CP (amino acids 1–80, SINV numbering) is largely basic and thought to be involved in charge neutralization of the genome RNA. Amino acids 38–55 are conserved uncharged residues that form a coiled-coil alpha-helix (helix I) important in dimerization of the CP during the assembly process. While the process of virus assembly is difficult to study in the complex cellular environment, the development of an *in vitro* assembly system based on bacterial expression of CP has led to advances in understanding NC assembly. These studies suggest that the initial event of NC assembly is the binding of CP amino acids 81–112 to the encapsidation signal on the genome RNA corresponding to nucleotides 945–1076 (SINV numbering) (Fig. 2). This interaction exposes a second site on the encapsidation signal where another molecule of CP binds and forms a dimer with the first CP molecule. Amino acids 114–264 constitute the previously mentioned chymotrypsin-like serine protease that autoproteolytically cleaves CP from the nascent structural polyprotein. This region is involved in binding residues from the cytoplasmic domain of E2, thus linking the outer icosahedral glycoprotein shell with the NC across the lipid bilayer (Fig. 1(a)).

Other lines of evidence support the dimer model of NC assembly. Helix I of CP, which is required for core accumulation in infected cells, may be functionally substituted by a GCN4 helix that forms dimeric coiled-coil interactions but not by a GCN4 helix that has a propensity to form trimeric coiled-coil interactions. In addition, helix I acts as a checkpoint in NC assembly whereby incompatible helices prevent the formation of core-like particles (CLPs) *in vitro*. Furthermore, a portion of CPs in either NCs or CLPs may be cross-linked into dimers by DMS, a lysine specific cross-linker with a 12 Å cross-linking distance. Cross-linking enabled an assembly deficient helix mutation of CP to assemble into NCs, suggesting that the cross-link can functionally replace the helix interaction.

While assembly of the CP into NCs proceeds in the cytoplasm, the processing and assembly of the glycoproteins occur in the ER and Golgi (Figs. 1(c) and 2). The autocatalytic cleavage of the CP reveals a signal sequence on the N-terminus of the newly cleaved structural polyprotein that directs it to the ER (Fig. 1(c)). PE2 is translocated into the ER until it reaches a 26-amino-acid stop transfer signal which anchors PE2 in the membrane. The C-terminal 33 residues of E2 then act as a second signal sequence to direct the next protein, 6K (55 amino acids), into the ER. 6K possesses a stop transfer sequence which anchors it in the membrane, and the C-terminus of 6K acts as signal sequence for the translocation of E1. E1 is anchored in the ER membrane by a final stop transfer

sequence close to its C-terminus. The release of PE2 and E1 by cellular signalase cleavage allows the formation of PE2-E1 heterodimers in the ER (**Fig. 2**). PE2 and E1 are each glycosylated in all alphaviruses, but the number and location of the modifications vary. In addition to glycosylation, the glycoproteins are palmitoylated in the Golgi apparatus. As the heterodimers are processed and transported through the ER and Golgi, they undergo a series of folding intermediates that are mediated by disulfide exchange and chaperones. Ultimately, they associate to form spikes which are composed of trimers of PE2-E1 dimers (**Fig. 2**).

The final maturation event is the cleavage of PE2 into E3 and E2 by a furin-like protease (**Fig. 2**). This cleavage occurs in a late Golgi or post-Golgi compartment and results in the destabilization of the heterodimer enabling the mature virus to fuse more readily with the target membrane. In most alphaviruses including Sindbis, E3 is released and not found in the mature virion. The final destination for the spike complexes is the plasma membrane, where the cytoplasmic domain of E2 (cdE2) recruits NCs assembled in the cytoplasm (**Fig. 2**). Structural studies show that cdE2 residues Tyr400 and Leu402 bind into a hydrophobic pocket of the CP. Mutation of Tyr400 negatively impacted virus budding, while protein translation and core accumulation were at wild-type levels. This interaction of cdE2 with the hydrophobic pocket in the CP is thought to drive the budding of the mature virus at the cell membrane.

Flavivirus Assembly

Flavivirus life cycle

Flaviviruses belong to the family *Flaviridae* of positive-strand RNA viruses which also consist of the pestiviruses and the hepaciviruses. The flaviviruses comprise more than 70 members including important human pathogens such as yellow fever virus, dengue virus, and West Nile virus. Flaviviruses enter cells via receptor-mediated endocytosis. The low pH environment of the endosomal membrane triggers the conformational change of the envelope glycoprotein which results in the fusion of the viral and endosomal membranes releasing the genome into the cytoplasm. The viral proteins are translated from the RNA genome as a single polyprotein. Signal sequences and stop transfer sequences result in the translocation of the nascent polyprotein to the ER membrane. The polyprotein is processed by a combination of cellular and viral proteases to produce the mature structural and nonstructural proteins. Genome replication and virion assembly occur in ER membrane-bound vesicles. The structural proteins and the genome bud into the lumen of the ER to form the immature virion which is transported through the secretory pathway. Prior to secretion of the virion, a furin cleavage converts the immature virus into the mature, infectious form of the virus.

Flavivirus virion structure

The flavivirus virion is made up of three structural proteins: capsid (C), pre-membrane (prM), and envelope (E) that are translated from the 5' one-third of the RNA genome (**Fig. 1(b)** and **(d)**). Signal sequences at the C-terminus of the C protein and prM serve to translocate prM and E respectively into the ER (**Fig. 1(d)**). The role of the highly basic C protein (12 kDa) is to encapsidate the viral genome during virion assembly. In contrast to the alphavirus CP which exhibits no membrane association, the flavivirus C protein is anchored to the membrane, at least transiently (**Fig. 1(d)**). However, Rubella virus, the sole member of the genus *Rubivirus* within the family *Togaviridae*, has a membrane anchored CP, perhaps indicating a common origin for the *Togaviridae* and *Flaviviridae*. prM is a glycoprotein that associates with the E protein and serves as a chaperone to facilitate the proper folding of E. The immature virions that bud into the ER consist of prM-E heterodimers. prM prevents premature fusion from occurring in the acidic environment of the ER and Golgi. Thus, prM has several functions analogous to the E3 glycoprotein of the alphaviruses. Cleavage of prM into pr and M by a furin-like protease triggers the rearrangement of the prM-E heterodimers into E-E homodimers, resulting in a radical change in size and shape required for the formation of the mature virus particle. The E glycoprotein is responsible for host cell receptor binding and for fusion of the viral and cellular membranes. The E glycoprotein is also critical for the assembly of the virion. High-resolution structures of the ectodomains of several flavivirus E proteins are available. The ectodomain is divided into three domains. Domain II constitutes the dimerization domain as it contains most of the intradimeric contacts between E-E homodimers. Domain II also contains the fusion peptide, a glycine-rich hydrophobic sequence that initiates fusion by insertion into the target cell membrane. Domain III comprises the immunoglobulin-like domain responsible for receptor binding. In addition to the dramatic conformational and translational changes that the E protein undergoes during the virion maturation process, it also changes conformation during membrane fusion. The low pH of the endosome during infection triggers a conformational change which results in the formation of E homotrimers. In this arrangement, the fusion peptides are exposed and available to insert into cellular membranes. Interestingly, the structure of the E protein was found to be very similar to the structure of the Semliki Forest virus E1 protein, the fusion protein of the alphaviruses.

The structures of two flaviviruses, dengue and West Nile virus, have been solved by cryo-EM and image reconstruction techniques and have been shown to be similar (**Fig. 1(b)**). The mature virion is ~50 nm in diameter and exhibits a smooth outer surface in contrast to the alphaviruses which have distinctive spike structures (cf. **Fig. 1(a)** and **(b)**). The E proteins are arranged parallel to the surface of the virus, with 90 E dimers arranged in groups of three to form a 'herringbone' pattern on the viral surface. This arrangement of the E proteins completely covers the surface of the virus, thus rendering the lipid bilayer inaccessible. Domain III of E protrudes slightly from the viral surface, allowing interaction with cell receptors. The membrane-spanning regions of E and M proteins form antiparallel helices while the stem regions are arranged parallel to the membrane.

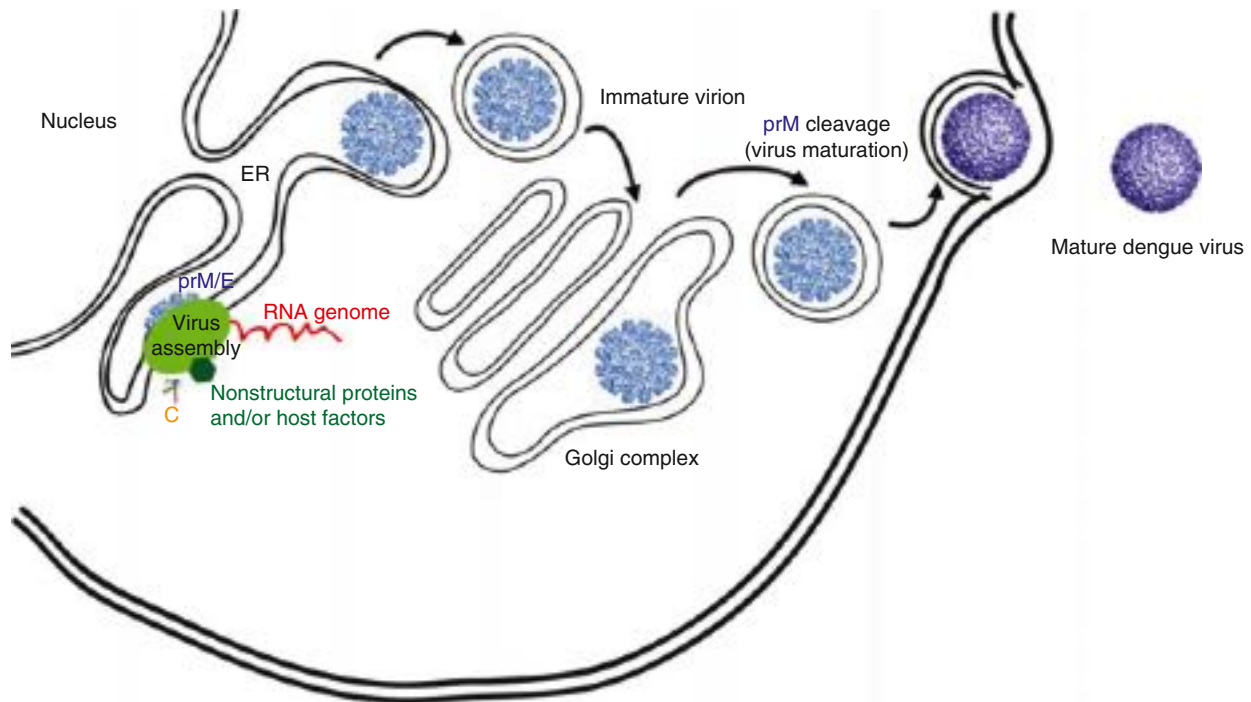


Fig. 3 Flavivirus assembly. Flavivirus assembly occurs on ER-associated membranes known as vesicle packets. Assembly and genome replication are coupled and the sites of assembly consist of the capsid proteins (C), the glycoproteins prM and E, the RNA genome, and one or more nonstructural proteins and/or host factors. The immature virion buds into the ER and is transported to the Golgi and trans-Golgi network. The glycoproteins are post-translationally modified as the immature virus is transported through the secretory pathway. Furin cleavage of prM results in the formation of the mature virus which then exits the cell by exocytosis.

The immature virus particle exhibits a dramatically different glycoprotein organization compared to the mature virion. Cryo-EM and image reconstruction of dengue and yellow fever virus immature virions have revealed that these particles are larger (~60 nm) and have spikes that protrude from the surface of the virus. These spikes are composed of trimers of prM-E heterodimers. The pr peptide covers the fusion peptide of E in this arrangement, similar to E2 covering E1 in alphaviruses, thus protecting it from premature fusion as the immature particle is transported through the acidic environment of the secretory pathway.

The NC is found below the viral envelope and is composed of a single copy of the genome RNA and multiple copies of the C protein. Cryo-EM reconstructions of the virion have shown that in contrast to the alphaviruses, there is no apparent organization to the flavivirus NC. This may be because there is no direct interaction between the C proteins in the core and the glycoproteins in the viral envelope since the E and M proteins do not penetrate below the inner leaflet of the membrane. Furthermore, no NCs have been observed in the cytoplasm of infected cells and attempts to establish an *in vitro* assembly system analogous to the alphavirus *in vitro* assembly system have failed. The lack of coordination between the C protein and the viral envelope proteins suggests that the assembly of virions is driven by the lateral interactions of the E and M proteins in the viral envelope and not by the C protein. This is supported by the observation that flavivirus infections result in the production of noninfectious subviral particles which are composed of just the viral envelope (E and M) and the lipid bilayer. Thus, the flavivirus glycoproteins are sufficient to induce particle budding.

Flavivirus assembly and budding

Virus-induced membrane structures called vesicle packets, which are continuous with the ER membrane, are the sites of flavivirus replication and assembly (Fig. 3). Within these structures the structural proteins are in intimate contact with the genome RNA. The C protein associates with the genome RNA via interactions between the positive charges distributed throughout the protein and the negatively charged phosphate backbone of the RNA. It is not yet clear how the C protein specifically recognizes the genome RNA; unlike for alphaviruses, a packaging signal has not been conclusively identified for flaviviruses. Coupling between genome replication and assembly within the vesicle packets has been proposed as a mechanism to ensure the specific encapsidation of the genome RNA. It has been shown that one or more nonstructural proteins (NS2A and NS3) are involved in genome packaging and NC assembly. The NC lacks a defined icosahedral structure as described above. Therefore, core formation is probably concomitant with the association of the C protein and RNA genome with the viral glycoproteins and budding into the ER lumen, thus giving rise to the immature particle (Fig. 3). The immature virion is transported from the ER to the Golgi where the viral glycoproteins are post-translationally modified. The cleavage of the prM protein in the trans-Golgi network triggers the dramatic reorganization of the viral glycoproteins that results in the formation of the mature virion (Fig. 3). The mature virion is then released from the host cell by exocytosis.

Conclusion

Following from the discussion of alphavirus and flavivirus assembly, it is apparent that the assembly of even the simplest enveloped viruses requires the complex interaction of viral and host factors in order to produce a virus particle which is at once stable and at the same time primed for disassembly. The whole range of the cell's machinery including the translation apparatus, polymerases, chaperones, and post-translational modification enzymes are co-opted by viruses in order to replicate the viral components necessary for assembly. Enveloped viruses have evolved to utilize different cellular membranes and cellular compartments for assembly and they take advantage of the secretory pathway to produce their viral glycoproteins. A majority of viruses bud from the plasma membrane (**Table 1**). This is the case with alphaviruses, where NC assembly occurs in the cytoplasm and the final assembly of the mature virion occurs at the plasma membrane. The high concentration of viral proteins, often concentrated at specific sites allows for the efficient interaction and assembly of virions. In contrast to alphaviruses, NC assembly and glycoprotein assembly is coupled in the flaviviruses and occurs in vesicle packets associated with the ER. Thus, the whole flavivirus virion is transported through the ER and Golgi while in the case of the alphaviruses only the glycoproteins are transported through the secretory pathway. These exit strategies are not unique and thus serve as model systems to study enveloped virus assembly and release.

Proteolytic cleavage of glycoproteins in order to convert them from stable oligomeric structures to metastable structures primed for fusion are common themes in enveloped virus structure and assembly. Cleavage of PE2 into E3 and E2 by a furin-like protease primes the alphavirus spike complex for fusion. A similar cleavage of prM triggers a dramatic conformational change of the flavivirus glycoproteins resulting in the formation of the mature virion which is now infectious.

Alphavirus budding requires the specific interaction of the NC with the E1–E2 spike complexes at the plasma membrane, thus ensuring that all virions have a genome packaged into them. However, the flaviviruses only require the interaction of the envelope proteins for budding, giving rise to subviral particles devoid of the C protein and genome RNA. Thus, the flavivirus envelope proteins alone are sufficient to drive budding of virus particles and the close coupling of genome replication and the C protein (perhaps mediated by replication proteins and host factors) is required to package the genome into virus particles. A third strategy for budding is exhibited by the retroviruses where capsid assembly has been shown to be sufficient to drive budding of the virus. In this case, targeting of the envelope proteins to these sites of CP assembly is essential to ensure the incorporation of the glycoproteins into the virion.

Although much has already been discovered about enveloped virus assembly, there are still many processes yet to be described. There is an increasing interest in the assembly pathway of viruses partly fueled by the potential to develop successful therapeutic agents targeting virus specific assembly processes. Advances in the field of structural biology will further help attempts to understand the assembly pathway of this important class of viruses.

See also: Assembly of Viruses: Nonenveloped Particles

Further Reading

- Garoff, H., Hewson, R., Opstelten, D.J., 1998. Virus maturation by budding. *Microbiology and Molecular Biology Reviews* 62 (4), 1171–1190.
- Harrison, S.C., 2006. Principles of virus structure. In: Knipe, D.M., Howley, P.M., Griffin, D.E., *et al.* (Eds.), *Fields Virology*, 5th edn. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 53–98.
- Hunter, E., 2006. Virus assembly. In: Knipe, D.M., Howley, P.M., Griffin, D.E., *et al.* (Eds.), *Fields Virology*, 5th edn. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 141–168.
- Mukhopadhyay, S., Kuhn, R.J., Rossmann, M.G., 2005. A structural perspective of the flavivirus life cycle. *Nature Reviews Microbiology* 3 (1), 13–22.

Assembly of Viruses: Nonenveloped Particles

M Luo, University of Alabama at Birmingham, Birmingham, AL, United States

© 2008 Elsevier Ltd. All rights reserved

This is a reproduction of M. Luo, Assembly of Viruses: Nonenveloped Particles, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00666-X.

Architecture of Viruses

Nonenveloped viruses have two essential components: protein and nucleic acid. The protein forms a coat called 'capsid' that packages the nucleic acid, which may be DNA or RNA. This complex constitutes a virion or a virus particle. The nucleic acid is the viral genome that encodes all the virus-specific genes required for viral replication. The protein capsid packages the viral genome during replication, and transmits it for the next round of infection. When the virion reaches the host cell, the capsid usually recognizes a specific receptor that helps the virion to enter. Once inside the host cell, the capsid has to release the viral genome so replication can begin. The size of a viral genome is usually limited, so only a few genes can be encoded. It is more efficient if only one or a small number of genes encode for capsid proteins that can self-assemble into a complete shell by use of many copies of the same proteins. The assembly of the capsid proteins follows a specific type of symmetry that allows a small protein unit to assemble into a large particle. The protein capsid can have a helical (filamentous virus) or icosahedral (spherical virus) symmetry. Helical symmetry is described by the diameter, d , the pitch, P , and the number of subunits per turn. There are as many capsid proteins as necessary for completely covering the nucleic acid genome. Icosahedral symmetry is defined by 12 fivefold axes, 20 threefold axes, and 30 twofold axes. A number, T , called the 'triangulation number', indicates how many quasi-symmetrical subunit interactions are within one asymmetrical region of the icosahedron. There are a total of 60 T copies of the capsid proteins in one icosahedral capsid.

For example, a picornavirus (pico-(small)-RNA virus) has a single positive-stranded RNA genome of about 8000 nucleotides. The RNA genome encodes a long polypeptide that is processed into individual viral proteins after translation directly from the positive RNA genome. Three of the 10 proteins are capsid proteins: viral protein 1 (VP1), viral protein 3 (VP3), and viral protein 0 (VP0). VP0 is processed to VP2 and VP4 after virus assembly. VP1, VP2, and VP3 are the three capsid proteins that form the body of the coat while VP4 is entirely inside the coat. There are 60 copies of each capsid protein in the coat. A picornavirus is, therefore, a $T = 1$ particle. However, the structure of VP1, VP2, and VP3 is highly homologous. If the small differences in each major capsid protein are ignored, the three proteins can be considered as the same building block. A picornavirus is thus called a pseudo- $T = 3$ particle (Fig. 1). The interactions between the capsid proteins in the coat are similar despite their different symmetry locations. For instance, VP1 (blue) interacts around the fivefold symmetry axis, whereas VP2 (green) and VP3 (red) interact with each other around the threefold symmetry axis. Since VP2 and VP3 are similar, the symmetry around the threefold axis is like a sixfold symmetry axis. This symmetry axis is therefore called a quasi-sixfold axis. The interactions between the neighboring VP1 proteins are considered quasi-equivalent (means more or less similar) to the interactions between VP2 and VP3 even though VP1 subunits are around a fivefold axis, and VP2/VP3 subunits are around a quasi-sixfold axis. This quasi-equivalence allows the closed capsid to be assembled using the same building block, the capsid proteins. It should be emphasized, however, that the true differences that exist in the capsid proteins are critically correlated with functions such as virus entry and release or packaging the viral genome.

Methods of Structure Determination

X-ray diffraction is the common technique used for studying the atomic structure of proteins and nucleic acids. When X-rays strike on the electrons of the atoms in a stationary specimen, a diffraction pattern of spots with different intensities is generated and recorded. By analysis of the diffraction pattern and the spot intensities, a three-dimensional electron density map (EDM) can be calculated by Fourier transformation. A three-dimensional chemical structure could be built based on the interpretation of the EDM. Two types of X-ray diffraction experiments are useful for virus structure studies: fiber diffraction (for filamentous viruses) and crystallography (for spherical viruses and globular viral proteins).

Another common technique is electron microscopy. Recent advances in electron microscopy allow researchers to determine relatively high-resolution three-dimensional structures of viral particles by use of image reconstruction of cryo-electron micrographs and electron tomography. This technique is particularly suitable for large viral particles that are difficult to crystallize. Electron microscopy and X-ray crystallography are therefore complementary to each other.

Atomic Structure of Helical Viruses

The disk of the tobacco mosaic virus (TMV) coat protein has been crystallized and its atomic structure resolved by X-ray crystallography. The intact TMV structure containing its nucleic acid could only be determined by X-ray fiber diffraction

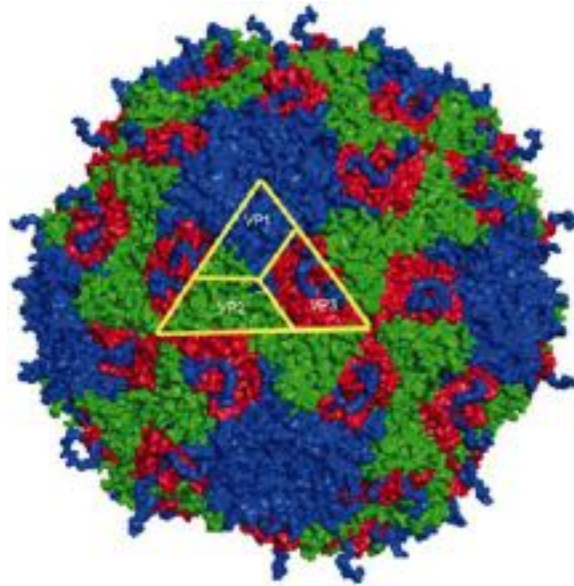


Fig. 1 The capsid of Theiler's murine encephalomyelitis virus (TMEV), a picornavirus. The triangle outlines one set of 60 copies of each capsid protein.

experiments, as also was that of Pf2 phage. The coat proteins of TMV and Pf2 contain mainly α -helices and the nucleic acid interacts with the coat protein by one base (Pf2) or three bases (TMV) per protein unit. The protein subunit is arranged in a super-helical structure coincident with a super-helical structure of the nucleic acid. The protein subunits form the outer layer of the helix to protect the nucleic acid from the environment. The nucleic acid is embedded between the protein layers that are stacked up in the spiral structure. The axis of the coat protein helix coincides with that of the nucleic acid. Since the symmetry of the super-helix puts no limit on the length of the helix, the virus particle can grow as long as the length of the nucleic acid genome; in this way all bases in the genome are covered by the coat protein. The coat proteins of TMV have many aggregation forms, depending on pH or ionic strength. In natural conditions, the TMV coat protein forms a disk with two layers of 17 protein subunits in each. The subunit of the TMV capsid protein contains four antiparallel α -helices with connecting long loops. Each protein subunit takes the shape of a shoe. When the disks of the capsid proteins are stacked, the RNA fits into the grooves formed by the protein subunits. The backbone of the RNA has charged interactions with the side chains of the amino acids while the bases of the RNA are accommodated by the hydrophobic space. The TMV RNA genome is inserted into the center of the coat protein disk to begin virus assembly. More and more disks are added to the top of the growing virus particle to pull the RNA through the center of the super-helix. The assembly is completed when the complete RNA is pulled into the virus particle. The coat protein of Pf2 has a different shape that contains an extended α -helix. The coat protein is added one by one to the DNA helix emerging from the bacterial membrane. The length of the α -helix is in the direction of the virus particle.

Atomic Structure of Spherical Viruses

Nonenveloped spherical viruses form large single crystals under proper conditions. Their atomic structure can be determined by X-ray crystallography with aid of fast computers and synchrotron X-ray sources. Since 1978, there are numerous atomic structures of viruses reported. Today, atomic structures have been determined for every major family of nonenveloped viruses, especially human pathogens.

Most capsid proteins of these viruses contain an antiparallel, eight-stranded β -barrel folding motif (Fig. 2). The motif has a wedge-shaped block with four β -strands (BIDG) on one side and another four (CHEF) on the other. There are also two conserved α -helices (A and B): one is between β C and β D, and the other between β E and β F. In animal viruses, there are large loops inserted between the β -strands. These loops form the surface features of individual viruses. The common presence of the β -barrel motif in viral capsid proteins is the result of structural requirements for capsid assembly. It also points to a common ancestor of different virus families.

A virus capsid may contain multiple copies of the β -barrel fold with the same amino acid sequence (such as $T = 3$ calicivirus or $T = 1$ canine parvovirus) or with different amino acid sequences (such as pseudo $T = 3$ picornavirus). In some cases, there are two β -barrel folds in a single polypeptide (such as adenovirus hexon). Capsid proteins of spherical viruses can have other motifs such as α -helices in reovirus and hepadnavirus.



Fig. 2 The β -barrel fold found in spherical viruses. The fold contains eight essential β -strands: BIDG on one side and CHEF on the other side. The figure was generated with rhinovirus 16 VP1 (PDB code 1AVM).

Nucleic Acid–Protein Interaction

The viral nucleic acid genome is always packaged inside the protein capsid. Positively charged patches formed by the side chains of arginines and lysines are found on the interior surface of the protein capsid. These positively charged areas are the preferred binding sites for the nucleotides. Usually the structure of the nucleic acid cannot be observed in a single-crystal X-ray diffraction experiment because of the random orientation of the icosahedral particles in the crystal. However, in rare cases, the nucleic acid might assume icosahedral symmetry by interacting with the protein capsid. Fragments of the complete genome assume the same conformation, although with different nucleotide sequences, at locations related by icosahedral symmetry. Such structures have been seen in bean pod mottle virus, tobacco mosaic virus, pariacoto virus and flock house virus (RNA viruses), and canine parvovirus (DNA virus). The bases are stacked either as an A-type RNA helix or form a coiled conformation to fit the interactions with the protein capsid. In the case of pariacoto virus, the RNA genome forms a cage that reassembles the icosahedral symmetry of the capsid (**Fig. 3**). These viruses readily form empty virus particles and have a hydrophobic pocket on the interior surface of the capsid. The nucleic acid generally interacts nonspecifically with the capsid protein.

Evolution

The highly conserved β barrel motif of the viral capsid protein indicates that many viruses must have evolved from a single origin. The unique three-dimensional structure of this motif is required for capsid assembly and it is generally conserved over a longer period of time than the amino acid sequence. The superposition of the capsid proteins from different viruses can be used to estimate the branch point in the evolutionary tree for each virus group. The structure alignment not only relates plant viruses to animal viruses, RNA viruses to DNA viruses, but also viruses to other proteins such as concanavalin A that has a similar fold and competes with poliovirus for its cellular receptor. The evolutionary relationship of these viruses is supported by amino acid sequence alignment of more conserved viral proteins such as the viral RNA polymerase. The structural similarities of the capsid proteins support the notion of a common evolutionary origin among nonenveloped viruses.

Assembly

The icosahedral capsid is assembled from smaller units made of several protein subunits. For picornaviruses, a protomeric unit is first formed with one copy of each polypeptide after translation. The termini of the subunits are intertwined with each other to hold the subunits together in the protomer. The protomers are then associated as pentamers which in turn form the complete icosahedral virion while encapsidating the viral RNA. In $T = 3$ or $T = 1$ plant RNA viruses, the pentamers are formed by dimers of the capsid proteins. In adenovirus and SV40, the capsid proteins form hexon units (three polypeptides, each has two

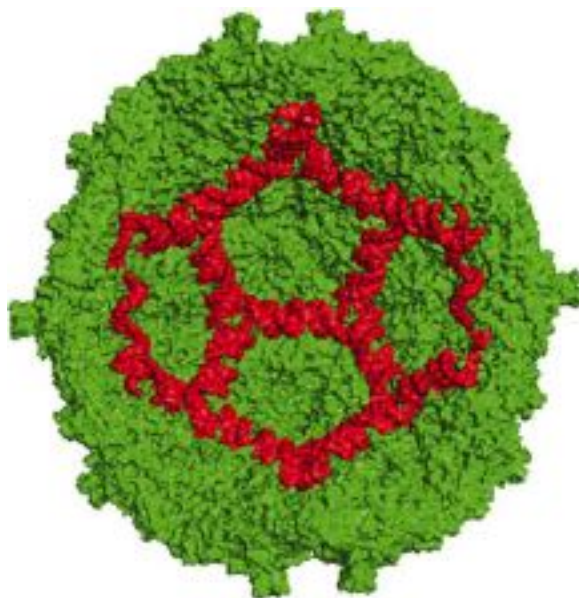


Fig. 3 The interior of the pariacoto virus capsid (green). The structured RNA genome is shown in red. The figure was generated with coordinates of pariacoto virus (PDB code 1F8V).

β -barrels) or pentamers before they assemble into an icosahedral shell. The nucleic acid could be packaged at different stages of assembly.

Host Receptor Recognition Site

Animal viruses have to recognize a specific host cellular receptor for entry during infection. Host receptor binding is the initial step of virus life cycle and could be an effective target for preventing virus infection. Based on the atomic structure of animal viruses, it was found that the receptor recognition site is located in an area surrounded by hyper-variable regions of the antigenic sites. Usually, the area is in a depression (called the 'canyon') on the viral surface that may be protected from recognition by host antibodies. This structural feature is, for instance, present in human rhinovirus (also known as the common cold virus), and the active site of influenza virus hemagglutinin (HA). The receptor-binding site on influenza virus HA does not have a deep depression, but it is surrounded by antigenic sites. The receptor-binding area on the surface of the viral capsid is conserved for recognition by the receptor, whereas the sites recognized by antibodies are distinct from the receptor-binding site and keep changing from strain to strain. By this mechanism, the virus can escape the host immune system by mutating the antibody epitopes, and at the same time maintain a constant receptor-binding site to continue its infection of the host cells. Evidence supports that this is a general mechanism that viruses use to evade the host immune defense.

Antigenic Sites

Antibodies are the first line of defense by the immune system against a viral infection. The epitopes combined with the neutralizing antibodies are mapped on a few isolated locations on the surface of viral proteins. The structure of human rhinovirus complexed with Fab fragments showed that the antibody makes contact with an area about 6 nm^2 and that the epitope spans different discontinuous polypeptides. Therefore, an effective vaccine usually needs to include a complete viral protein or a large fragment. The binding of the antibodies does not significantly change the structure of the antigen. The exact mechanism by which antibodies neutralize antigens is dependent upon the binding site and processes of the virus replication.

Antiviral Agents

Viral infectious diseases can be cured if an agent can be administered to stop viral infection. Such agents have been synthesized and shown to bind to the capsid of rhinovirus in the crystal structure. The compounds were inserted into the hydrophobic pocket

within the β -barrel of the major capsid protein VP1. Binding of the compounds stops uncoating of the virion and the receptor binding, which resulted in the failure of releasing the viral RNA into the cytoplasm. These compounds inhibit infections of several other RNA viruses and may be effective against other viruses after modification since the β -barrel structure exists in many viruses.

The most successful antiviral drugs are the HIV protease inhibitors which are developed based on the atomic structure of the protease. Through iterative cycles of computer modeling, chemical synthesis and structural studies of the protein-inhibitor complexes, a panel of clinical effective drugs has been brought to the market and has shown great benefits to patients. Inhibitors of influenza virus neuraminidase have also been developed by the same method and marketed as antiviral drugs.

See also: Assembly of Viruses: Enveloped Particles

Further Reading

- Knipe, D.M., Howley, P.M., Griffin, D.E., *et al.* (Eds.), 2002. Fields Virology, 4th edn. Philadelphia, PA: Lippincott Williams and Wilkins.
- Rossmann, M.G., Johnson, J.E., 1989. Icosahedral RNA virus structure. *Annual Review of Biochemistry* 58, 533–573.
- Schneemann, A., 2006. The structural and functional role of RNA in icosahedral virus assembly. *Annual Review of Microbiology* 60, 51–67.

Virion Assembly: From Small Picornaviruses (*Picornaviridae*) to Large Herpesviruses (*Herpesviridae*)

Ling Zhu, Nan Wang, and Xiangxi Wang, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Viruses are infectious agents composed of proteins and a viral genome, and occasionally, other components, such as lipids and carbohydrates. Upon infecting a host cell, viruses are able to use their own components and the cellular machinery to produce copies of the viral genome and proteins, which then assemble into progeny viruses that can eventually infect other host cells. The assembly of the virion is an essential step in the viral life cycle and thus understanding the mechanisms underlying the assembly of the viral particles could facilitate identification of targets or opportunities for novel antiviral therapies. More recently, the process of assembly of viruses has come into sharp focus due to renewed interest in self-assembled structures and their applications. With regards to self-assembly, viruses are remarkably sophisticated biological machines. Therefore, insights obtained from the characterization of the mechanism of the processes of self-assembly can be harnessed for use in nano- and bio-technology-based industries. In fact, we are already beginning to see the application of this knowledge; for example, viral capsids have been successfully used for guiding the synthesis of inorganic and organic nanostructures and as vectors for gene therapies.

Viruses present remarkable differences in (1) size, ranging from nanometers to hundreds of nanometers; (2) morphology, including spherical icosahedral particle, cylindrical- or bullet-shaped helical particle and filament- or brick-shaped or irregular particle; (3) molecular composition; (4) structural organization; and (5) complexity (Fig. 1). The simplest nonenveloped virions are made up of just an oligomeric protein shell (capsid) that encapsidates the viral genome. The capsid itself contains only one or a few types of protein subunits [some examples are shown in Fig. 1]. As opposed to this, in some large virions, the particles not only contain the viral genome, but also proteins and/or other macromolecules, which may be organized in internal subassemblies. Furthermore, proteins and/or protein subassemblies like portals, tails, fibers, etc. with specific functions may be externally attached to the genome-containing capsid (head) (e.g., bacteriophage ϕ 29 in Fig. 1). Lastly, enveloped viruses present an additional layer of complexity, where the capsid is surrounded by a lipid bilayer (envelope) in which proteins are embedded [e.g., Japanese encephalitis virus (JEV), and human herpes simplex virus (HSV) in Fig. 1]. Some enveloped virions have an intricate multi-layered structure made up of lipid, protein and/or nucleoprotein structured layers [e.g., African swine fever virus (ASFV) in Fig. 1].

Although viral morphologies are structurally diverse, most viral capsids have either icosahedral or helical symmetry. Roughly half of all viruses are spherical, usually icosahedral. The focus of this review is the assembly of icosahedral viruses. The formation of a simple icosahedral capsid involves tens of copies of a single type of capsid protein (CP). However, most capsids have complex structures, which are formed by using hundreds or thousands of one or several different types of CPs that may adopt various conformations within the capsid. Adaptation to different hosts and microenvironments has led to many variations in the modes of assembly of viruses. Here, we provide a general overview of an icosahedral viral capsid assembly and comment on the features of the stability of the capsid.

Structural Principles in Icosahedral Capsid Assembly

Typical icosahedral capsids must be made up of exactly 60 identical structural elements in order to fulfill intrinsic geometric constraints. The simplest icosahedral capsids, e.g., parvoviruses, are indeed composed of 60 identical CP subunits with the same

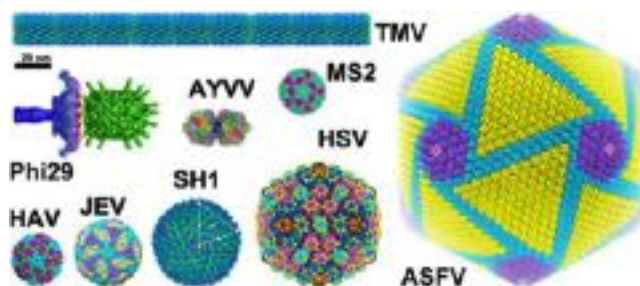


Fig. 1 Some representative virus particles. Particles are reproduced at the same scale (indicated by the bar on the top left) to clearly show the size and morphology. From left to right and top to bottom: TMV (Tobacco mosaic virus, PDB accession: 2OM3), Phi29 (bacteriophage, EMDB accession: EMD-1420), AYVV (Ageratum yellow vein virus, EMDB accession: EMD-4174), MS2 (bacteriophage, PDB accession: 1ZDH), HAV (hepatitis A virus, PDB accession: 4QPI), JEV (Japanese encephalitis virus, PDB accession: 5WSN), SH1 (Haloarcula hispanica SH1 virus, PDB accession: 6QT9), HSV (herpes simplex virus type 2, PDB accession: 5ZAP), ASFV (African swine fever virus, EMDB accession: EMD-0815).

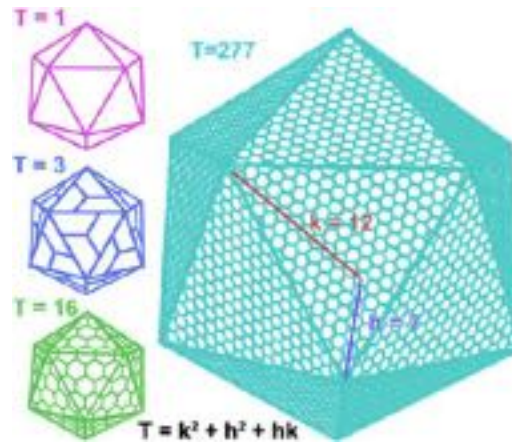


Fig. 2 Icosahedral capsids with different triangulation number. Top left, $T = 1$ (no quasi-equivalence) such as MVM (Minute virus of mice); middle left, $T = 3$, such as JEV; bottom left, $T = 16$, such as HSV; right, $T = 277$, such as ASFV.

conformation (Fig. 2). Similar geometric constraints dictate the assembly of larger icosahedral capsids. Larger viruses must be made by using multiples of 60 subunits of a single CP type that adopt slightly different or quasi-equivalent conformations. This theory on the rules of the assembly of icosahedral capsids was firstly proposed by Caspar and Klug in 1962. The Caspar and Klug theory explained how some multiples of 60 identical subunits can be arranged with similar or quasi-equivalent interactions, following the $T = h^2 + hk + k^2$ rule, where h and k are integers and T is called the triangulation number. Icosahedral capsids made up of identical subunits can be thought of as geometrically composed of pentameric and hexameric capsomers, in which subunits in one configuration participate in pentamers, while subunits in any of the other conformations shape into hexamers (Fig. 2). The capsids mentioned previously, formed by 60 identical CP subunits in the same conformation are defined as $T = 1$ with 12 pentameric capsomers and no hexameric capsomers (Fig. 2). The larger, $T = 3$ capsids, e.g., those of flaviviruses, are made up of $60 \times 3 = 180$ identical CP subunits, each adopting one of the three quasi-equivalent conformations during assembly such that these capsids can be regarded as composed of 12 pentameric capsomers and 20 hexameric capsomers. A special case is observed in the picornaviruses where the capsid is built from 60 copies of a fundamental subunit (the protomer) composed of three paralogous proteins (termed VP1, VP2, and VP3) and a small extended polypeptide (VP4, located inside) with $T=1$ (pseudo $T=3$) symmetry. Some complex capsids have higher T numbers, e.g., the $T=16$ capsids of herpesviruses and the $T=277$ capsid of the African swine fever virus (Figs. 1 and 2). Interestingly, not all icosahedral capsids adhere to the quasi-equivalence rules of Caspar and Klug. For example, the $T=7$ capsids of polyomaviruses are made from 72 CP pentamers instead of the 12 pentamers and 60 hexamers as proposed by the quasi-equivalence theory.

Although many viruses, such as parvoviruses, picornaviruses, polyomaviruses, and some large DNA viruses (ASFV and mimiviruses), are not related in amino acid sequences, the core of their major CPs fold into an eight antiparallel β -strands containing structure termed as the 'jelly roll' fold (Fig. 3). The structure of the CP subunit in some viruses; particularly in large DNA viruses, consists of two jelly roll domains. However, the single jelly roll structure is more common, as it also occurs in the CP subunits of small icosahedral viruses (Fig. 3). Except for the loops connecting the strands of the jelly roll, the overall topology of all the CP subunits is similar. The loops form the major epitopes and confer virus-specific differences. The remarkable similarity of these jelly rolls among viral CPs that do not share amino acid sequences and belong to different families of viruses, suggests a common ancestral evolutionary history.

Synthesis of Nascent Capsid Assembly and Assembly Intermediates

For successful assembly, the synthesis of the CPs is regulated for the following purposes: (1) synthesis quantity, the accumulation of high amounts of CPs in the host cells is necessarily required to compete with the vast amount and diversity of pre-existing cellular proteins; (2) synthesis speed, CPs synthesis must reach the maximum levels by the time the viral genome is being replicated, so that it can be efficiently packaged; and (3) synthesis stoichiometry, CPs must be synthesized in the proper ratio to ensure the correct assembly of viral particles.

Viral capsids are assembled from capsid building blocks (CBB). CBBs are built using either identical CPs or non-identical CPs that are related by two-fold, three-fold, and five-fold symmetry axes. The stable oligomeric CBBs may possibly act as assembly intermediates during the capsid assembly process and the regular interactions between the CBBs and their neighbors could probably facilitate the further assembly of the capsid. Results of experiments performed using in vitro assembly systems in combination with diverse theoretical approaches have shed light on the fundamental principles of viral capsid assembly starting from CBBs. Some studies also support a nucleation-driven cooperative process for capsid assembly, which begins with a lag phase during which a critical concentration of the CBB is built. This phase is followed by a build-up of an assembly line of intermediate structures. The contacts and structural changes in CBBs as well as assembly intermediates drive the assembly of viral capsids.

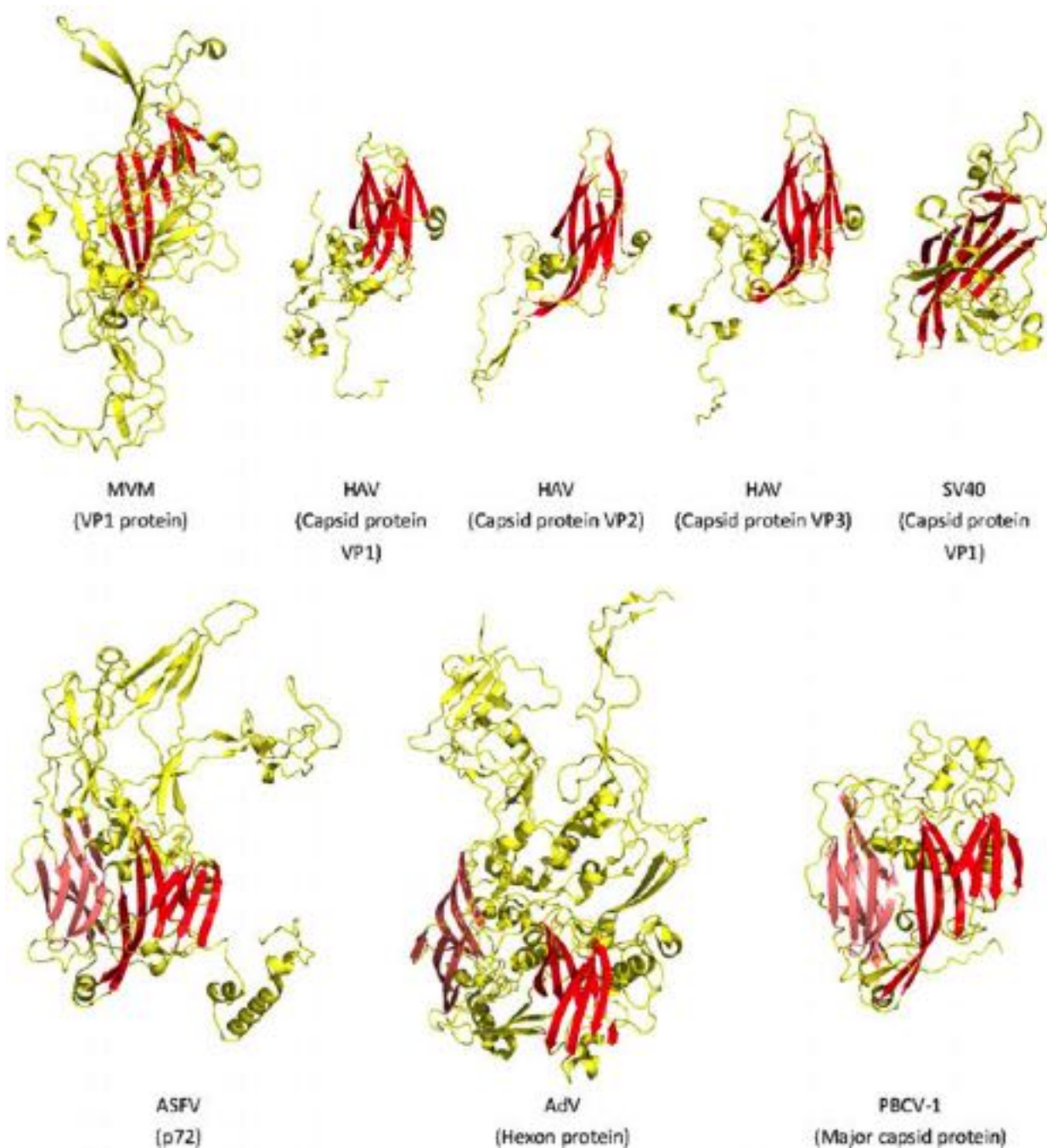


Fig. 3 Structures of representative CPs with single or double jelly roll fold domains are described here. Ribbon diagrams of the VP1 subunit in the MVM (PDB accession: 4ZPY), the VP1, VP2, and VP3 subunits in the hepatitis A virus (HAV, PDB accession: 4QPI), the VP1 subunit in the SV40 (PDB accession: 3BWQ), the major capsid proteins in AdV (adenovirus, PDB accession: 4V4U), ASFV (PDB accession: 6L2T), and PBCV-1 (Paramecium bursaria Chlorella virus 1, PDB accession: 6NCL). The jelly roll fold domain is colored in red, other regions are presented in yellow color. The figures are prepared using the Pymol program (<http://www.pymol.org>).

Modes of Assembly

Assembly of a viral capsid is a complex oligomerization process that involves a sophisticated pathway, which is regulated by ordered interactions between the CPs or CBBs. Three main modes of assembly for the viral capsids have been recognized. These are: (1) Self-assembly, where the CP subunits can spontaneously assemble themselves into capsids under appropriate conditions; (2) Scaffolding protein-assisted capsid assembly, where not only the CP subunits, but also the scaffolding proteins from viruses or host cells are required for the capsid assembly; (3) Viral genome-assisted capsid assembly, where specific and simultaneous

interactions between the CP subunits and the viral genome drive the capsid assembly. Details of each of the assembly modes with specific examples are described below.

Capsid Self-Assembly

Empty capsids of many icosahedral RNA viruses are known to exist during the viral life cycle or under specific conditions, such as those of the cowpea chlorotic mosaic virus (CCMV), which can aggregate into capsid-like particles at high ionic strength. The natural empty particles are produced in cells infected with some, but not all picornaviruses; weak protein-protein interactions are sufficient to drive assembly of Hepatitis B virus capsid. When properly isolated, these empty particles are antigenically indistinguishable from virions and can be completely dissociated into CBBs or CPs in vitro by treatment under special conditions. To some extent, the capsid self-assembly resembles the viral assembly. As a representative example, the capsid self-assembly for picornaviruses is summarized below.

The *Picornaviridae* are an important family of human and animal pathogens, including poliovirus (PV), human rhinovirus (HRV), enterovirus 71, hepatitis A virus, and foot and-mouth disease virus. The 30-nm-diameter virion comprises 60 copies of each of the structural proteins, VP1, VP2, VP3, and VP4, (VP1, VP3, and VP0 in empty particle, VP0 is the precursor of VP2 and VP4) arranged in a pseudo- $T = 3$ ($T = 1$) icosahedral lattice (Figs. 1 and 2). The viral genome is translated in a single open reading frame to produce a polyprotein that is cleaved co-translationally by viral proteases to yield the viral proteins. An early cleavage of the polyprotein releases a precursor protein P1, which is further processed by the viral 3CD protease into a heterotrimeric protomer (VP0–VP3–VP1), sedimenting at 5S (Fig. 4). It is believed that 5S protomers self-associate to form 14S particles (VP0–VP3–VP1)₅ and that 12 copies of 14S particles can self-assemble into an 75S empty capsid [(VP0–VP3–VP1)₅]₁₂ (Fig. 4). Given the fact that empty particles are generally stable enough under physiological conditions, it suggests that empty particles are not a storage form of 14S particles, at least for some enteroviruses (Fig. 4). However, it has still been debatable whether empty particles are either a dead-end product or direct precursor into which viral RNA is packaged by as-yet-unidentified machinery. Regardless, the encapsidation appears to be tightly linked to RNA replication, as in vitro assembly assays reveal an absolute dependence of encapsidation on *de novo* synthesis of progeny RNA. Encapsidation leads to the formation of the provirion, which consists of the RNA and 60 copies each of VP0, VP1, and VP3. The maturation of the provirion is signaled with the cleavage process of VP0 into VP2 and VP4 by viral RNA, which significantly increases the stability of the particle; albeit, the detailed mechanism of VP0 cleavage remains elusive. Viral genome-assisted capsid assembly in picornaviruses will be discussed further in the following paragraphs.

Scaffolding Protein-Assisted Capsid Assembly

The correct assembly of many icosahedral viral capsids requires assistance of scaffolding proteins. This requirement of scaffolding proteins has been observed in dsDNA bacteriophages, herpesviruses, and some ssDNA phages. Typically scaffolding proteins taking part in some steps of the assembly process by establishing specific but transient protein-protein interactions with the CP subunits are later digested away during virus maturation, being absent in the mature virion. Some viruses encode only one type of scaffolding protein. This protein acts as an assembly chaperone, forming an internal scaffolding core in the procapsid. Other viruses encode more than one scaffolding protein, forming both the external and internal scaffold to drive the capsid assembly. In some viruses, the scaffolding proteins alone are capable of clustering together to form a scaffold core, that is subsequently used as a nucleus for capsid assembly by recruiting CP subunits. However, the scaffolding proteins interact with CP subunits in a co-assembly mode in which the scaffolding proteins facilitate CP subunits' associations and in turn the CP subunits promote the scaffolding proteins interactions by triggering conformational changes in them. As a typical example, the scaffolding protein-assisted capsid assembly for herpesviruses is described below.

Herpesviridae are a large family of dsDNA viruses that cause a number of diseases such as encephalitis, oral and genital blisters, congenital disorders and even cancer, and whose members include herpes simplex virus (HSV), human cytomegalovirus (HCMV), Epstein–Barr virus (EBV), and Kaposi sarcoma herpesvirus (KSHV). The herpesvirus capsid initially assembles as a precursor procapsid that subsequently matures during DNA packaging facilitated through a portal channel. To form a herpesvirus precursor capsid, at least four types of components are required: VP5, triplex, scaffold, and portal complex, in which the portal assembles a unique dodecamer at a five-fold vertex. Scaffold proteins are reported to trigger the portal oligomerization and incorporation, determining the correct curvature and ultimate closure of the capsid shell. The recently reported high-resolution structure of HSV-2 capsid reveals a pronounced assembly pattern for triplexes, exhibiting a twofold symmetry organization pattern, on the capsid shell, together with a structurally rigid unit comprising a triplex and its covalently linked lasso triangle (probably function as a basic assembly unit), indicative of the possibility of triplexes acting as external scaffold proteins to regulate the correct assembly. Based on our more recent asymmetrical reconstructions of the HSV portal vertex (unpublished data), we propose that the formation of a unique “portal complex” with the requisite scaffold proteins, triplexes and VP5 is probably the first event for initiating proper precursor capsid assembly (Fig. 5). Subsequent assembly proceeds by the addition of assembly units to form the portal vertex. We envisage that five assembly units cluster to make up a “penton complex”, followed by the formation of the penton vertex by addition of assembly units (Fig. 5). With the mediation of the scaffold core and triplexes, eleven penton vertices are attached to the nucleation of the portal vertex, leading to a complete precursor capsid. After that, the cleavage of the scaffold

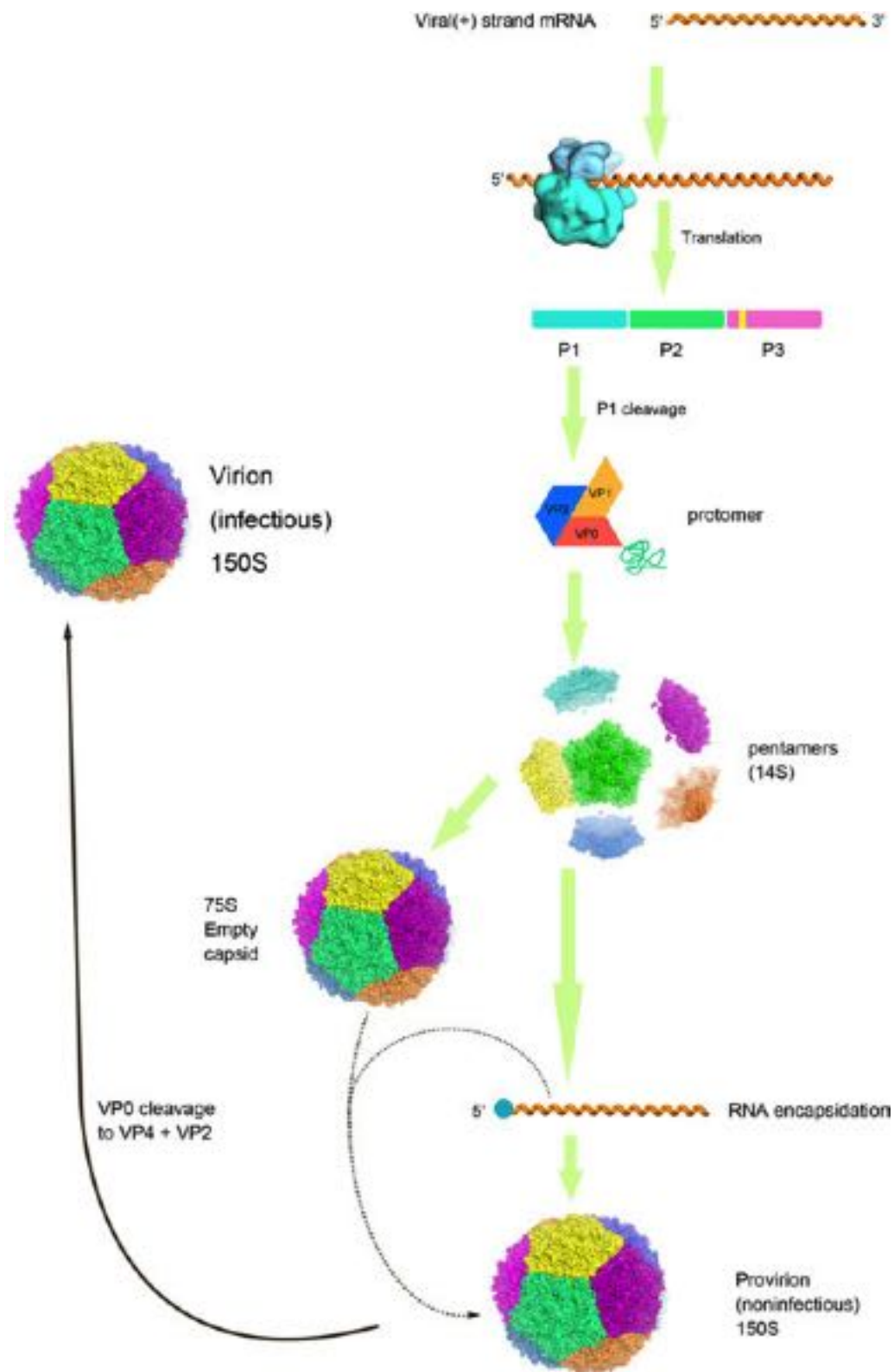


Fig. 4 Proposed model for poliovirus assembly. During the virus life cycle, viral translation produces a long polyprotein that is processed by viral proteases. Assembly of the virus is linked to processing of the polyprotein and proceeds through a series of assembly intermediates, including protomer, pentamer, natural empty particle, provirion and mature virion.

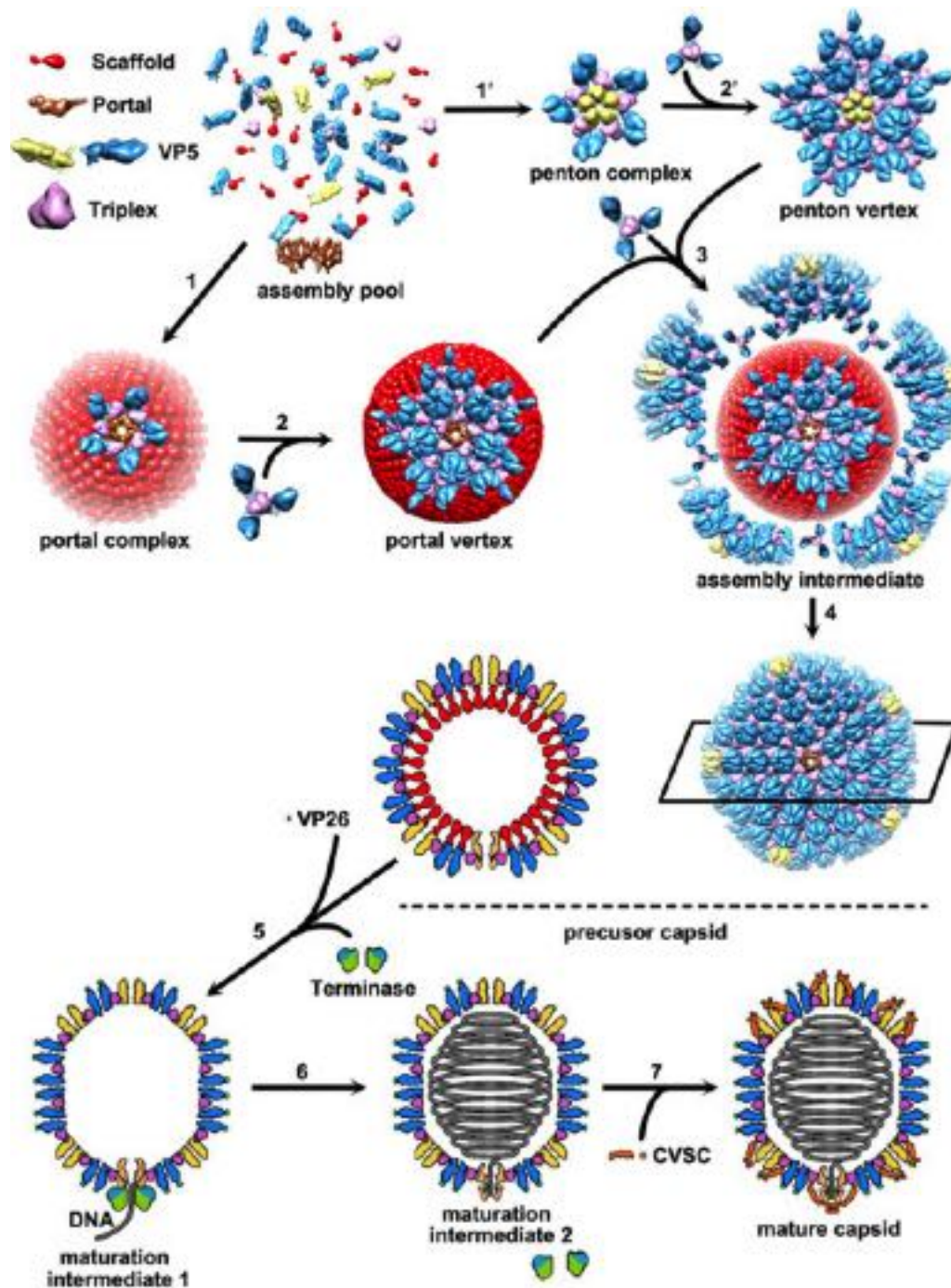


Fig. 5 Proposed model for herpesvirus capsid assembly. Herpesvirus capsid assembly starts from an assembly pool containing VP5, scaffolding proteins, triplexes (two copies of VP23 and a VP19C), and UL6. Three copies of VP5 and a triplex could form a triangular complex as an assembly unit. Scaffolding proteins trigger 12 UL6 to make up a dodecameric portal and associate with the portal, together with VP5 and triplexes, forming the portal complex. After this, the assembly units probably bind to the portal complex for assembling the portal vertex. Meanwhile, 5 assembly units could self-assemble into a penton complex, which recruits more assembly units to form the penton vertex. As soon as the formation of the two types of vertices is completed, 11 penton vertices and 20 free assembly units (Tf) attach to the portal vertex to form a precursor capsid in which the scaffold core formed by scaffolding protein ensures the correct curvature. The precursor capsid matures via a major structural transformation triggered by limited proteolysis of the scaffolding proteins. Angularization might make the capsid expose the VP26 binding site. The angularized capsid encounters the terminase complex and dsDNA is packaged into the shell, after which the portal changes its configuration to disassociate from the terminase and the CVSC are attached to the vertices to form a mature capsid.

proteins causes the precursor capsid to undergo substantial conformational rearrangement such that the outer shell changes from roughly spherical to icosahedral. On conditions that the packaging machinery (terminase complex) is engaged properly, DNA is pumped into the angularized shell. Then, newly packaged DNA triggers the portal conformational switch from an immature to a mature state, recruiting the CVSC and ensuring the retention of DNA, resulting in the formation of C-capsids (Fig. 5).

Viral Genome-Assisted Capsid Assembly

Encapsidation of the viral genome leading to the assembly of infectious progeny virions is an essential step in the life cycle of viruses. The most commonly employed strategies for packaging viral genomes involve either filling a preformed procapsid assembled via self-assembly or by assistance from scaffolding proteins with progeny genome or by nucleation of capsid assembly around viral nucleic acids (viral genome-assisted capsid assembly). We have a convincing picture of how some bacteriophages and herpesviruses package their dsDNA genome into a preassembled procapsid. This process involves an ATP-driven packaging motor complex and a portal complex located at one of the vertices of the icosahedral capsid (see the section of scaffolding protein-assisted capsid assembly). However, many other viruses, in particular ssRNA viruses, package their genome concomitantly with capsid assembly, in which the viral genome plays key roles in recruiting CPs or CBBs by charge interactions, yielding genome-containing virions. In these cases, no natural empty particles are produced during the virus life cycle. The process of viral genome-assisted capsid assembly is reviewed next, by focusing on some well-studied examples.

For many ssRNA icosahedral viruses, the initiation of virion assembly involves the recognition of a specific nucleotide sequence in the viral genome called the packaging signal by CP subunits. The length, stiffness, and the secondary structure of the viral RNA play essential roles in determining capsid morphology and assembly efficiency. Even in the absence of the viral RNA, the CP subunits of some single-stranded RNA viruses have abilities to self-assemble *in vitro*, but the assembly efficiency is much higher in the presence of genomic RNA; as observed in the cases of the assembly of bacteriophage MS2 and most picornaviruses. Structural and experimental evidences reveal that binding of the translational repressor (TR, a stem-loop in the viral RNA, known as a packaging signal) to the CP dimer triggers a conformational switch from the symmetric dimer, the dominant form in solution/cell cytoplasm, to its asymmetric configuration, constructing the capsid with a ratio of 1:2 (symmetric dimer: asymmetric dimer). Sequence-specific interaction between CP dimer and a 19 nt-length RNA stem-loop (TR) that has been well characterized is believed to function as an assembly initiation site *in vivo*, ensuring in this case cognate packaging of viral RNA even in the presence of competing RNA phases.

However, previous studies suggest that picornaviruses might be an exception to exploiting high affinity CP: RNA interactions for initiating virion assembly. Many studies including extensive nucleotide recoding of large sections of the polio genomes have failed to identify any RNA determinants involved in picornavirus assembly. Recently, multiple genome regions in parechoviruses, belonging to the family of *Picornaviridae*, have been reported to possess high affinity for CP subunits. Remarkably, these bindings of the genomic sequences to CPs are sequence-specific. Several high-resolution parechovirus structures (including Ljungan virus) have been reported that do contain 60 extensive densities for RNA fragments in close contact with the overlaying CPs, amounting to some 12% of the viral genome. These data support the existence of multiple RNA-based PSs in some of the picornaviruses and these PSs contribute differentially to assembly efficiency as expected for a PS-mediated process. Apart from PSs' mediation, there are also competing RNA conformations that regulate replication, translation and assembly. In line with this, in picornaviruses, the virally encoded 2C protein induces nascent viral RNAs with special local secondary structures into the pentameric CBBs ((VP0-VP3-VP1)₅). The above observations have led to a suggestive scenario for viral genome- assisted capsid assembly in many ssRNA viruses. In combination with the possible mediations led by viral encoded proteins and host cell factors, the binding of CP subunits to the viral RNA might trigger the structural changes of viral genome from an extended, secondary structures enriched fold to a highly compacted configuration. In turn, the folded RNA can facilitate CP polymerization and determine the size and geometry of the particle-to-be, ensuring the right assembly for virions.

Concluding Remarks

Studies on molecular mechanisms for viral capsid assembly are gaining more attention and importance, which also reflects how much we do not know yet about these processes. Recent dramatic improvements and new developments in structural, biophysical and biochemical techniques, and the discovery of novel theoretical and experimental approaches have provided impetus for a better understanding of these fundamental aspects of the viral assembly.

Acknowledgments

We thank Prof. D. Stuart for his valuable comments on the review. Work was supported by the National Key Research and Development Program (2018YFA0900801), the Key Programs of the Chinese Academy (KJZD-SW-L05), the Strategic Priority Research Program (XDB08020200) and National Science Foundation of China (31800145 and 31570717). Xiangxi Wang was supported by Ten Thousand Talent Program.

Further Readings

- Acharya, R., *et al.*, 1989. *Nature* 337, 709.
- Adeyemi, O.O., *et al.*, 2019. *Journal of Virology* 93.
- Aksyuk, A.A., Rossmann, M.G., 2011. *Viruses-Basel* 3, 172.
- Albertini, A.A.V., *et al.*, 2006. *Science* 313, 360.
- Ashfaq, U.A., Javed, T., Rehman, S., Nawaz, Z., Riazuddin, S., 2011. *Virology Journal* 8, 161.
- Basnak, G., *et al.*, 2010. *Journal of Molecular Biology* 395, 924.
- Beniac, D., Booth, T.F., 2017. *Scientific Reports* 7.
- Beren, C., Dreesens, L.L., Liu, K.N., Knobler, C.M., Gelbart, W.M., 2017. *Biophysical Journal* 113, 339.
- Cao, L., *et al.*, 2019. *PLoS Biology* 17.
- Caspar, D.L.D., 1984. *Citation Classic/Life Science* 15.
- Ceres, P., Zlotnick, A., 2002. *Biochemistry* 41, 11525.
- Chen, X., Goncalves, M.A.F.V., 2016. *Molecular Therapy* 24, 447.
- Coleman, J.R., *et al.*, 2008. *Science* 320, 1784.
- Cuconati, A., Molla, A., Wimmer, E., 1998. *Journal of Virology* 72, 6456.
- Dykeman, E.C., Stockley, P.G., Twarock, R., 2013. *Journal of Molecular Biology* 425, 3235.
- Fu, C.Y., Morais, M.C., Battisti, A.J., Rossmann, M.G., Prevelige, P.E., 2007. *Journal of Molecular Biology* 366, 1161.
- Harvey, S.C., Petrov, A.S., Devkota, B., Boz, M.B., 2009. *Physical Chemistry Chemical Physics* 11, 10553.
- Hogle, J., Chow, M., Filman, D., 1985. *Science* 229, 1358.
- Kang, S., *et al.*, 2010. *Biomacromolecules* 11, 2804.
- Khayat, R., *et al.*, 2011. *Journal of Virology* 85, 7856.
- Lavelle, L., Michel, J.P., Gingery, M., 2007. *Journal of Virological Methods* 146, 311.
- Liu, Y., *et al.*, 2010. *PLoS Pathogens* 6.
- Li, S.Y., Erdemci-Tandogan, G., van der Schoot, P., Zandi, R., 2018. *Journal of Physics: Condensed Matter* 30.
- Li, C.L., Wang, J.C.Y., Taylor, M.W., Zlotnick, A., 2012. *Journal of Virology* 86, 13062.
- Mateu, M.G., 2013. *Archives of Biochemistry and Biophysics* 531, 65.
- Medrano, M., *et al.*, 2016. *Journal of the American Chemical Society* 138, 15385.
- Mettenleiter, T.C., 2002. *Journal of Virology* 76, 1537.
- Motwani, T., *et al.*, 2017. *Science Advances* 3.
- Nugent, C.I., Kirkegaard, K., 1995. *Journal of Virology* 69, 13.
- Perlin, M., Phillips, B.A., 1973. *Virology* 53, 107.
- Perlmutter, J.D., Qiao, C., Hagan, M.F., 2013. *eLife* 2.
- Phillips, B.A., 1971. *Virology* 44, 307.
- Prevelige, P.E., Fane, B.A., 2012. *Viral Molecular Machines* 726, 325.
- Ren, J., *et al.*, 2013. *Nature Communications* 4, 1929.
- Ren, J.S., *et al.*, 2015. *Journal of Virology* 89, 10500.
- Rolfsson, O., Toropova, K., Ranson, N.A., Stockley, P.G., 2010. *Journal of Molecular Biology* 401, 309.
- Rossmann, M.G., *et al.*, 1985. *Nature* 317, 145.
- Salas, M.L., Andres, G., 2013. *Virus Research* 173, 29.
- Shakeel, S., *et al.*, 2017. *Nature Communications* 8.
- Simmons, N.S., Blout, E.R., 1960. *Biophysical Journal* 1, 55.
- Stilwell, J.L., Samulski, R.J., 2004. *Molecular Therapy* 9, 337.
- Stockley, P.G., *et al.*, 2013. *Journal of Biological Physics* 39, 277.
- Twarock, R., Bingham, R.J., Dykeman, E.C., Stockley, P.G., 2018. *Current Opinion in Virology* 31, 74.
- Villamor, D.V., Eastwell, K.C., 2012. *Phytopathology* 102, 128.
- Wang, X., *et al.*, 2012. *Nature Structural & Molecular Biology* 19, 424.
- Wang, X., *et al.*, 2015. *Nature* 517, 85.
- Wang, X., *et al.*, 2017. *Nature Communications* 8, 14.
- Wang, J.L., *et al.*, 2018. *Nature Communications* 9.
- Xiao, C., *et al.*, 2009. *PLoS Biology* 7, 958.
- Xu, J.W., Wang, D.H., Gui, M., Xiang, Y., 2019. *Nature Communications* 10.
- Young, M., Willits, D., Uchida, M., Douglas, T., 2008. *Annual Review of Phytopathology* 46, 361.
- Yuan, S.A., *et al.*, 2018. *Science* 360, 48.
- Zhang, W., *et al.*, 2001. *Virology* 279, 471.
- Zhou, Z.H., *et al.*, 2000. *Science* 288, 877.
- Zhu, L., *et al.*, 2015. *Nature Communications* 6.
- Zhu, L., *et al.*, 2016. *Nature Microbiology* 1, 16150.

Genome Packaging

Richard J Bingham and Reidun Twarock, University of York, York, United Kingdom
Carlos P Mata and Peter G Stockley, University of Leeds, Leeds, United Kingdom

© 2021 Published by Elsevier Ltd.

Nomenclature

ATP Adenosine 5'-triphosphate

CP Capsid protein

HPA Hamiltonian path analysis

PS Packaging signal

Glossary

Adenosine 5'-triphosphate (ATP) The principal molecule for storing and transferring energy in cells.

ATP-driven genome packaging A genome packaging mechanism reliant on ATP to package the viral genome into a preformed capsid.

Energy-independent genome packaging A genome packaging mechanism reliant of a co-assembly mechanism in which genome packaging and virus assembly are

concomitant. It is known as “energy independent” as no external energy sources, such as ATP, are required for packaging.

Packaging signal Nucleotide sequence/structure element in a viral genome that directs the packaging of the viral genome.

Packaging signal mediated assembly An assembly mechanism reliant on the collective action of multiple dispersed packaging signals.

Introduction

Viruses employ a wide range of distinct genome packaging strategies that depend on both the sizes and chemical nature of their genomes. Smaller viruses typically assemble their nucleocapsids around their genomes in a co-assembly process combining genome compaction with capsid formation. Larger viruses, on the other hand, form empty capsids into which the viral genome is subsequently packaged, e.g., with the help of a packaging motor. Packaging strategies are adapted to the biophysical properties of the nucleic acids to be packaged, such as their persistence lengths, with smaller viruses favoring nucleic acids that are more compactable. We therefore organize this article by genome size ([Fig. 1](#)).

Energy-independent Genome Packaging in Small Single-Stranded Viruses

Positive-sense RNA virus genomes typically comprise three to ten genes, including the structural proteins and an RNA-dependent RNA polymerase (RdRp). Viruses with the shortest genomes, ranging from ~ 1 to ~ 3.5 kilobases (kbs) in length, are typically packaged into icosahedral capsid architectures formed from between 60 and 240 coat proteins (T -numbers in the range $T = 1$ to $T = 4$). Such viruses co-assemble their capsids around their genomes, and sequence/structure elements in the genomic RNA called packaging signals (PSs) are known to aid both genome compaction and capsid assembly ([Twarock and Stockley, 2019](#)). Well-studied examples of this genome packaging mechanism occur in satellite tobacco necrosis virus (STNV), that packages its ~ 1.3 kb-long genome into a $T = 1$ capsid formed from 60 coat proteins ([Fig. 2\(A\)](#)). It has been shown that the genome packaging mechanism in this virus relies both on a 5' assembly initiation cassette formed from five PSs, and multiple dispersed other PSs distributed along the entire genome ([Patel et al., 2015, 2017a](#)). Its packaging properties can be enhanced via variation of the PS sequences/structures. This implies that packaging properties can be engineered and controlled in a laboratory environment, and potentially harnessed for applications in nanotechnology and medicine.

Another example of PS-mediated genome packaging is bacteriophage MS2 ([Fig. 2\(B\)](#)). Its ~ 3.5 kb-long genome becomes packaged into a $T = 3$ capsid formed from 89 coat protein dimers, 29 symmetric and 60 asymmetric ones, and a single copy Maturation Protein that takes the place of one of the symmetric dimers (see [Fig. 2](#)). PSs have been identified in this virus using Hamiltonian Paths Analysis ([Twarock et al., 2018](#)), resulting in predictions that are in excellent agreement with experiment. The genome packaging mechanism can also involve other proteins. For example, in Picornaviruses packaging-signal mediated assembly occurs in the context of an assemblysome that also involves a non-structural protein in complex with the viral polymerase ([Shakeel et al., 2017](#)). It appears that PSs in these viruses are conserved across different strain variants ([Shakeel et al., 2017](#)). For Picornaviruses that undergo a maturation event, such as Rhinoviruses and Enteroviruses, there is more than one type of packaging signal ([Chandler-Bostock et al., 2020](#)), and it is possible that PSs act at different life-cycle stages.

In all cases discussed above, specific interactions between the viral genome and capsid protein form an integral part of the genome packaging mechanism, and are the common characteristic of the genome packaging mechanism. It is still debated how many other small viruses share this mechanism. For example, plant viruses like Cowpea Chlorotic Mottle virus (CCMV) package

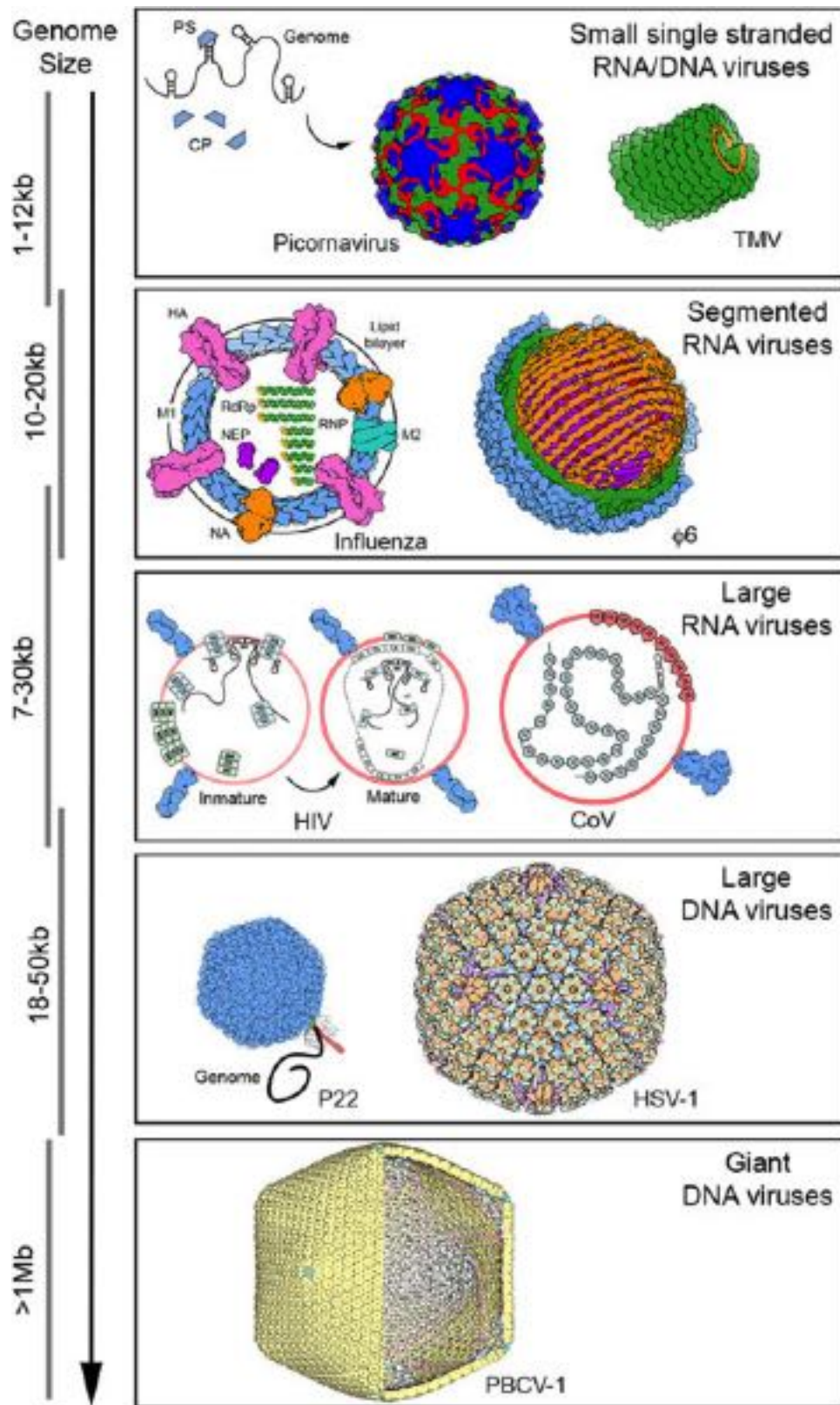


Fig. 1 Overview of scale-dependent changes in genome packaging. Viruses with increasing genome sizes employ distinct genome packaging mechanisms. Whilst smaller viruses favor an energy-independent genome packaging mechanism, larger viruses use an ATP-driven motor to package their genomes into a preformed capsid. Virus structures have been rendered as a cartoon using UCSF Chimera based on PDB 1AYM (HRV16), 3J06 (TMV), 4BBL, 6QX8, 3LBW, 6Z5L, 6NHP, 1PD3, 5HUG (Influenza RNP, RdRp, M2, M1, HA, NEP, NA respectively), 1GC1 (HIV), 6VXX (CoV), 2XYZ (P22), 6CGR (HSV-1), 6HY0 ($\phi 6$), (6NCL (PBCV-1); and EMD-0300, EMD-0294, EMD-0295 and EMD-0296 ($\phi 6$ genome).

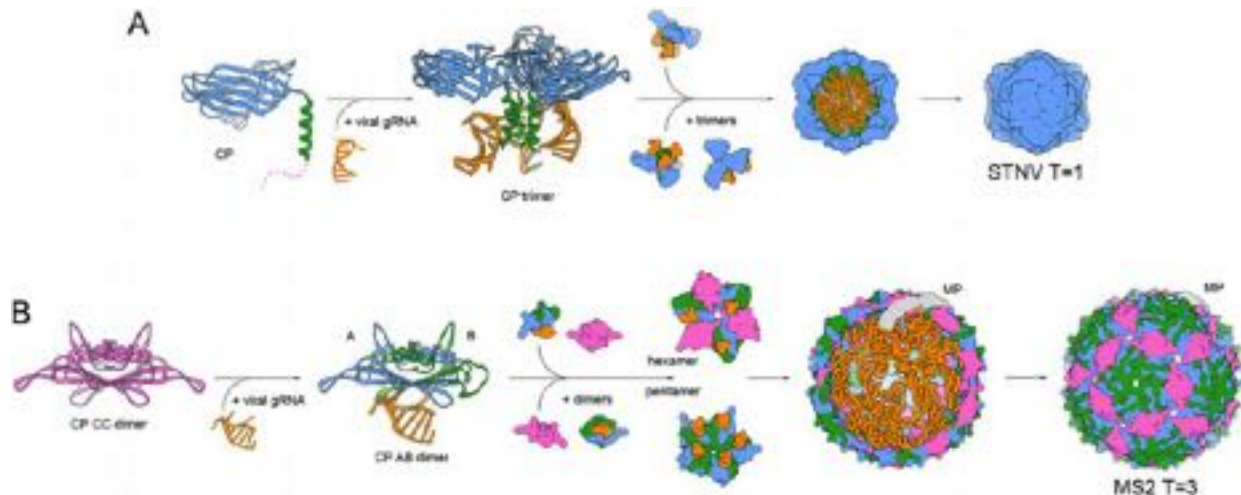


Fig. 2 Examples of an energy-independent co-assembly mechanism. Smaller viruses combine genome compaction and capsid assembly into a single step. This is facilitated by multiple dispersed sequence/structure motifs called packaging signals (PSs), that act collectively in a mechanism known as packaging signal-mediated assembly. (A) In Satellite Tobacco Necrosis Virus (STNV) PSs promote formation of CP complexes that become the particle 3-fold axes in the $T = 1$ capsid shell. (B) In bacteriophage MS2, PSs promote the switch from a symmetric to an asymmetric CP dimer, that is required around the 5-fold axes for particle formation. Virus structures have been rendered as a cartoon using UCSF Chimera based on PDB 4BCU (STNV), 1ZDH and 5TC1 (MS2).

one of three distinct genomic RNA segments into separate versions of the same type of capsid shell, which may suggest that packaging is non-specific. However, it is also possible that all three distinct segments present similar signals as a means of ensuring that all three are packaged with the required frequency. In any case, the assembly efficiency that PSs confer on the virus assembly process make it highly likely that this mechanism has evolved widely among small RNA viruses.

A similar mechanism is also employed by small DNA viruses that package their genomes in the form of pre-genomic, single-stranded RNA. An example is Hepatitis B Virus (HBV) (Patel *et al.*, 2017b), that packages its ~ 3.2 kb-long pre-genome into nucleocapsids with $T = 3$ or $T = 4$ capsid architectures. HBV is an example of a virus that undergoes a reverse transcription step following genome packaging, perhaps accounting for the fact that the number of PSs is typically smaller in these viruses than for ssRNA viruses, thus permitting the reverse transcription step to occur *in capsa*.

Nodaviruses are icosahedral viruses with ~ 4.5 kb genomes that are packaged into $T = 3$ capsids 29–35 nm in diameter. Examples are Pariacoto and Flockhouse virus. Packaging is achieved by a bipartite organization into a large (RNA1, ~ 3.1 kilobases in length) and small (RNA2, ~ 1.4 kilobases in length) segment. Whilst RNA1 encodes a protein with multiple functional domains, including a transmembrane domain and an RdRp domain, RNA2 encodes only a viral capsid protein precursor, which is auto-cleaved into two mature proteins during virus assembly. Roughly 35% of their single-stranded RNA genomes are organized into dodecahedral cages composed of RNA duplex, and there are extensive interactions between portions of the viral RNA genome and the icosahedral capsid, reminiscent of the packaging signal-mediated assembly mechanism.

Single-stranded RNA and DNA viruses with genomes of size ~ 6 kb appear to employ a different genome packaging mechanism. These viruses typically form helical structures around their genomes, resulting in rod-shaped capsid architectures. An example of a positive-sense single-stranded RNA virus is the Tobamovirus tobacco mosaic virus (TMV) that infects a wide range of plants. Its ~ 2 k coat protein subunits form two-layer disk structures that assemble around its ~ 6.4 kb long genomic RNA using the “lock-washer” mechanism, forming helical particles ~ 300 nm in length and ~ 18 nm in diameter. Genome packaging specificity is ensured by a single 69-nucleotide long packaging initiation signal, but there is no evidence for multiple dispersed signals in this virus.

Single-stranded DNA viruses of comparable length also form rod-shaped structures. An example is the ~ 6.4 kb long circular ssDNA of the filamentous bacteriophage M13. It is encapsidated by ~ 2.7 k coat proteins (the major protein P8), and this nucleoprotein rod becomes capped with five copies of different minor proteins. These are five copies of the surface exposed P9 and a more buried companion protein P7 at one end of the filament, and five copies of P3 and its less exposed accessory protein P6 at the other one. In contrast to the TMV case, the genome packaging mechanism is still poorly understood, but it is at least partially considered electrostatic.

Energy-independent Genome Packaging in Larger (ss)RNA Viruses

The flaviviridae, which have genomes between 9 and 12 kb in length, and the alphaviruses, with genomes 11–12 kbs long, are among the larger RNA viruses self-assembling into icosahedral capsids without an energy-driven mechanism. Both of these families contain

human pathogens of significant clinical interest. These viruses package their genomes within icosahedral capsids, but unlike RNA viruses with shorter genomes, they also surround their capsids with lipid bilayer envelopes. The longer genomes encode additional structural proteins to mediate this process, primarily envelope (E) proteins. However, flaviviridae also encode smaller membrane (M) proteins. Given the relatively small increase in genome length in comparison to the picornaviruses, it is perhaps expected that these viruses should share similar assembly and packaging mechanisms. The packaging of flaviviruses including dengue, Zika and Hepatitis C have been suggested to rely upon regulation of the viral genome, even though this has not been definitively verified (Mazeaud *et al.*, 2018). Cryo-electron microscopy (cryo-EM) reconstructions of Zika virus show ordered density of RNA seemingly in contact with the capsid surface, suggesting the presence of RNA-protein contacts, whilst Hepatitis C virus has been shown to use multiple-dispersed packaging signal motifs to ensure reliable, specific assembly, similar to picornaviruses.

Packaging of alphaviruses has been intensively studied in recent years, perhaps due to the growing spread of Chikungunya virus (Mendes and Kuhn, 2018). The alphavirus Venezuelan equine encephalitis virus (VEEV) was used in mutational studies of alphavirus assembly to identify a region of the genome required for efficient assembly. Reports suggest that their packaging signals take the form of 4–6 stem-loops containing a triple guanine motif at the start of the loop, and more than one stem-loop was shown to be required for efficient assembly. Phylogenetic analysis shows that this packaging signal is conserved across the majority of the alphaviruses. This single-site packaging signal was initially suggested as a contrast to the multiple-dispersed packaging signal paradigm, in part because this packaging signal is not a determinant of assembly, but does greatly improve the efficiency of assembly (Mendes and Kuhn, 2018). However, a recent study using the alphavirus Semliki Forest virus (SFV) identified multiple capsid protein binding sites dispersed across the genome, strongly suggesting that multiple-dispersed packaging signals play a role in genome recognition and packaging in this viral family.

With increasing genome size, the capsids required to package them must also increase in size. As larger capsids are less likely to form by diffusive aggregation of the protein subunits, even if facilitated by multiple-dispersed PSs, evolution has resulted in multiple strategies to ensure that viral genetic material is packaged specifically and efficiently. One solution to the genome packaging problem for larger viral genomes is its segmentation. Viruses in the family *Orthomyxoviridae*, for example, have negative-sense, single-stranded RNA genomes that are in between 10k and 14.6k nucleotides in length, and organized into six to eight segments. A prominent example is influenza A virus, with a 13.5 kilobase long genome distributed over eight segments. These present as ribonucleoproteins, each composed of a distinct RNA segment bound by the viral polymerase and oligomeric viral nucleoprotein (Bolte *et al.*, 2019). A specific organization of the segments, called the 1 + 7 configuration, is important for influenza A virus genome packaging. The morphology of *Orthomyxoviridae* is either ellipsoidal or filamentous, with particles 80–120 nm in diameter and up to 20 µm long. Viruses in this family face the challenge of ensuring the packaging of the correct number and types of segments. This is achieved via packaging sequences that are unique to individual segments in combination with specific nucleoproteins, and relies on a redundant and plastic network of RNA-RNA and potentially RNA-nucleoprotein interactions (Bolte *et al.*, 2019). A similar mechanism reliant on an interaction network of multiple RNA signals in the genome segments occurs also in other multi-segmented viruses, such as Bluetonguevirus (AlShaikhahmed *et al.*, 2018) (see below). Whilst it is widely accepted specific sequences within the viral RNA segments are required for genome packaging, many mechanistic details are to date not yet fully understood.

As genome size increases even further, so does also the diversity in capsid morphology when compared with smaller RNA viruses. A common strategy is to separate two of the roles of the capsid; genome compaction and genome packaging. In RNA viruses with longer genomes, the genome is often selectively bound by a protein called *nucleocapsid* (NC), which is distinct from the protein controlling the conformation of the virus particle and aids genome compaction. This approach is exemplified by the Ebolavirus, which has a negative-sense RNA genome of around ~19k nucleotides. The RNA is protected by multiple copies of NC, which bind to the genome via non-specific interactions. The protein-genome complex, or ribonucleoprotein (RNP), forms the core of the virion, while the envelope protein (among other proteins) forms the filamentous capsid along with cellular lipids. The same organization is seen also in the closely related Marburg virus, and the plant-infecting aspiroviridae whose genomes are ~11 kb long.

Retroviruses have shorter genomes, typically 7–10 kb in length, and the virus packages two copies of the genome into each virion. Whilst the evolutionary advantages of packaging two genome copies are still debated, they might be related to the high degrees of recombination in these viruses, and the mechanism favors packaging of two genome copies of similar type. Retroviruses have evolved a mechanism to ensure that precisely two copies of their genome are packaged. In human immunodeficiency virus (HIV) this is achieved via the dimerization initiation sequence, a palindromic sequence of six nucleotides which enables an inter-strand interaction to form between two genomes. Genome specificity is ensured using a mechanism similar to that in smaller RNA viruses, via a packaging signal sequence/structure motif encoded in the genomic RNA. In HIV, the most well-characterized retrovirus, this packaging signal is denoted as *Psi*, which is now widely assumed to be the stem-loop SL3 in the 5' leader of the genome. It specifically binds the nucleocapsid domain of the Group-specific antigen, or Gag protein. A GGAG motif in a tetraloop enables a sequence-specific interaction of the SL3 stem-loop with the zinc-knuckles of the NC domain of Gag. This interaction is often referenced as 'the' packaging signal of HIV. However, it is not clear that the SL3-NC interaction is sufficient to ensure packaging, especially given that the neighboring stem-loop, SL2, also has a very high affinity for NC. Subsequent experiments have determined that multiple copies of NC bind to the 5' leader of HIV, suggesting a similarity with the multiple PS sites in the genomes of smaller RNA viruses. Prior to assembly of the complete virion, Gag forms complexes on the surface of the membrane which will later become part of the virion. The interaction between the NC domain and the RNA ensures that the genomes are localized before packaging. It is unknown whether the RNA-Gag interaction occurs prior to dimerization, but it is logical to assume that genomes localized on the membrane surface are more likely to join together into a dimerized complex than

two genomes in solution (Rein, 2019). Following dimerization, Gag undergoes proteolytic cleavage into matrix (MA), capsid (CA) and nucleocapsid (NC) domains, where the RNA-bound NC forms an RNP, while the CA domain forms the conical, mature capsid enveloped by a lipid membrane containing MA.

The coronaviridae genomes are among the longest of all viral single-stranded RNA genomes, ranging between 26 and 31 kb in length. In order to achieve the compaction required for packing into the confines of the viral particles, these viruses again delegate the stages of genome recognition and virion formation to distinct viral proteins. The coronavirus family contains viruses of clinical relevance, most notably SARS-CoV 1 & 2, MERS, as well as the ‘common-cold’ causing HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1. Experiments with murine hepatitis virus (MHV), a lineage A betacoronavirus, involving packaged and non-packaged defective interfering RNA identified a region essential for packaging roughly 20 kb from the 5′ end of the genome. Further studies refined this to a 190nt sequence, which forms an extended stem-loop with a repeating sequence-structure motif on the 3′ side of its stem. This stem-loop is frequently referred to as the coronavirus PS. During replication, the coronaviruses produce multiple subgenomic fragments, to act as templates for translation of the various protein components. However, the PS ensures selectivity of the full-length genome, as the PS is not present in these shorter RNA fragments. Phylogenetic analyses suggest that other coronavirus lineages and genera do not contain this PS, although similar sequences have been found at other sites in the genomes (Masters, 2019). Alphacoronaviruses, such as TGEV, have been suggested to have an alternative packaging motif consisting of multiple stem-loops in the 5′ leader. However, these have not been well characterized and it is unclear how these would avoid the issue of packaging subgenomic fragments, given that these contain the 5′ leader. While the MHV PS is understood to be required to ensure packaging selectivity, the RNA-protein interactions it mediates are not understood. Coronaviral RNA forms an RNP with the nucleocapsid (N) protein, but these contacts are dispersed across the genome. Several hypotheses have been suggested involving the membrane (M) protein, but it is still debated which mechanism is employed. Given the relative simplicity of the repeating motif in the PS, it is also possible that multiple RNA-protein contacts could occur (Masters, 2019).

Double-Stranded RNA Viruses

The vast majority of these viruses have multi-segmented genomes that share similarities in replication strategies. These genomes are encapsidated into a multi-shelled icosahedral capsid. The two most studied viral families (*Reo-* and *Cystoviridae*) have revealed evolutionary pathways with similar results despite differences in the replication process, thus creating a competent viral progeny which includes a complete genomic component (Borodavka *et al.*, 2018).

Members of the *Reoviridae* are characterized by encapsidating 9–12 (+)ssRNA genomic segments into a structurally unaltered innermost $T = 1$ core particle, built from 60 asymmetric dimers of a single CP. This core particle nucleates around larger and ordered RNA complexes made from those (+)ssRNA genomic segments as a consequence of previous assortment, which occurs in cytoplasmic inclusion bodies called viroplasm. In Bluetongue virus (BTV), efficient core assembly is produced when all of the 10 (+)ssRNA segments are included, as well as the viral RdRp and RNA capping enzyme (CAP), both located near the 5-fold vertices on the inner surface. Additionally, this packaging process has been shown to occur sequentially, and is mediated by a combination of specific (1) PSs localized close to both UTRs, and (2) segment assortment signals (SAS) promoting trans interactions between different segments (AlShaikhahmed *et al.*, 2018). Finally, when all of the required components are inside the container of the $T = 1$ particle, this acts as a scaffold to build a surrounding $T = 13$ capsid and an additional final outer triple layer. In Rotavirus, the third layer is lost during cell entry, and allows the double-layered particles (DLPs) to synthesize (+)ssRNAs. Assembly of this DLP is mediated by cooperative interactions of disordered N-terminal capsid shell protein (CSP) protruding domains with both the RdRp and (+)ssRNAs. RdRp has a high affinity for the 3′-terminus of rotavirus (+)ssRNAs and entails the specific packaging of only those segments that are bound to RdRp molecules, thus ensuring the consequent replication. In addition to this, each individual genome segment is associated with RdRp complex that will allow later transcription and exit of a unique type of (+)ssRNA through each of the 5-fold pores. Furthermore, the non-structural protein NSP2 has been found to have chaperone-like activities and to be involved in sequence-specific inter-molecular contacts during the assortment of the 11 (+)ssRNAs.

Viruses of the *Cystoviridae* family share the same multi-layered capsid architecture of the *Reoviridae*, and the (+)ssRNAs segments are also encapsidated into an inner $T = 1$ core where they will be subjected to dsRNA synthesis by the RdRp. However, in contrast to the *Reoviridae*, these (+)ssRNAs are actively translocated into pre-assembled cores or procapsids in a similar way to dsDNA viruses (see below). This family is well represented by the phage $\phi 6$, whose genome is based on three dsRNA segments termed Large (L), Medium (M) and Small (S). In the case of $\phi 6$, the procapsid consists of a dodecahedron with the presence of RNA helicase molecules located at its outer vertices and the RdRp facing the inner surface. In contrast to reoviruses, and because of the presence of a helicase that confers the ability to package the viral genome, $\phi 6$ RdRp is not involved in this function. Genome packaging is mediated by specific PSs present at the 5′ ends of the three (+)ssRNAs that form unique secondary structures. These are recognized close to the helicase, which allows packaging to occur in a 5′ to 3′ directionality and sequentially starting with the smallest segment. This first event starts a chain reaction of procapsid expansion coupled to exposition of new recognition sites for the other (+)ssRNAs. Once the procapsid is completely expanded because of the presence of all segments, RdRp produces new dsRNA molecules which adopt a single-spoiled genome organization similar to dsDNA viruses (Ilca *et al.*, 2019).

Infectious bursal disease virus (IBDV), a virus from the Birnaviridae family, constitutes a remarkable singularity within dsRNA viruses as it encompasses diverse features from the different viral groups. Particles contain a variable amount of their genome, organised into RNP complexes similar to those in (−)ssRNA viruses. As an exception to all other dsRNA viruses, the IBDV particle

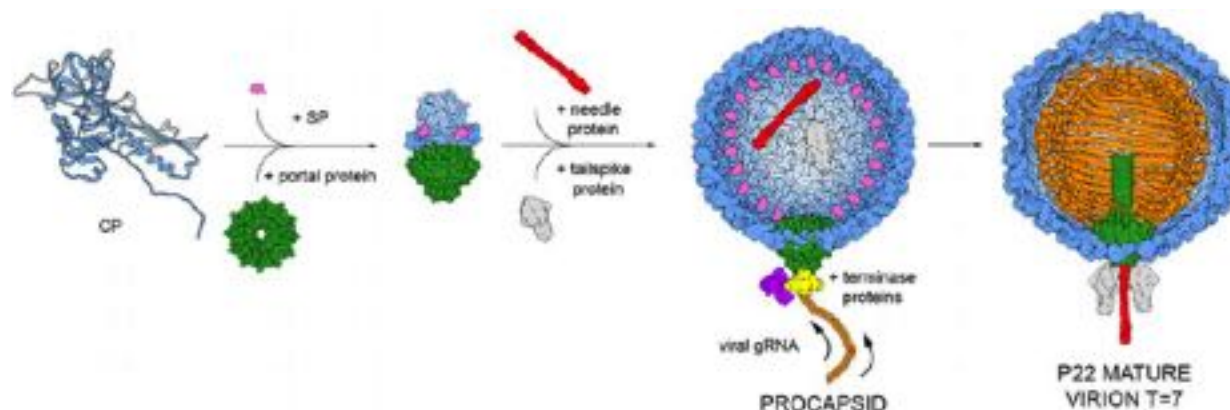


Fig. 3 Example of an ATP-driven genome packaging mechanism. Larger viruses, in particular dsDNA viruses, form an empty capsid prior to packaging. Genome packaging is facilitated by an ATP-driven packaging motor through the portal as illustrated here for P22. Virus structures have been rendered as a cartoon using UCSF Chimera based on PDB 2XYY (procapsid), 2XYZ (mature capsid), 2GP8 (Scaffold Protein), 4DKW (large terminase protein), 3P9A (small terminase protein), 5JJ1 (portal protein procapsid), 5JJ3 (portal protein mature capsid), 2POH (needle protein), 1TYW (tailspike protein); and EMD-0304 (genome in mature virion).

is built from a single-shelled $T = 13$ capsid, which instead shares similarities with the CP of (+)ssRNA picornaviruses. Additionally, the assembly of this $T = 13$ capsid is produced via a procapsid composed of an internal transitory scaffold shell similar to dsDNA herpesvirus. Finally, although viral assembly remains uncertain in most of fungal dsRNA viruses (chryso-, partiti-, toti- and quadrvirus), some of them have a multipartite genome, thus encapsidating each of the distinct dsRNA segments that constitute the complete genome separately into different single-shelled $T = 1$ particles.

Energy-Driven Genome Packaging in Double-Stranded DNA and Giant Viruses

The most abundant morphologies among bacteriophages are tailed phages with icosahedral capsids known as “heads”. Packaging of the genome into these viruses occurs via an ATP-driven packaging motor (Rao and Feiss, 2015), that couples ATPase to DNA translocation and fills the head to near-crystalline density. The mechanism of such packaging motors has been dissected in many viruses. For example, a nucleotide-dependent DNA gripping mechanism has been described in bacteriophage T4, bacteriophage $\phi 29$, bacteriophage P22, and thermophilic bacteriophage P23-45. Packaging motors are very sophisticated, adapting DNA rotation and step size to preserve subunit coordination as the capsid fills (Fig. 3).

Whilst this mechanism has been studied predominantly in dsDNA phages, variants have also been reported in animal virus descendants. An example is herpesvirus, which retains the basic characteristics of the phage packaging mechanism described above. Its dsDNA genome is packaged into an icosahedral nucleocapsid, that is structurally similar to that of tailed bacteriophages in the order of Caudovirales. There also are double-stranded RNA viral motors (Guo and Lee, 2007) that share similar characteristics (see above).

Mimivirus is a nucleo-cytoplasmic large DNA virus (NCLDV). Like other complex giant viruses in this family, such as megavirus and pandoravirus, it employs a packaging mechanism that is distinct from the ones described above (Chelikani *et al.*, 2014), i.e., the energy-independent packaging mechanisms of the RNA and DNA viruses with smaller genomes (<20 kb), or the portal-terminase packaging systems of the larger viruses. The precise determinants of this mechanism are still under investigation.

Conclusion

Genome packaging is a vital stage in any viral life cycle. A better understanding of the mechanistic details of this process opens up novel avenues to prevent formation of viable virus particles. The packaging signals identified in smaller RNA and DNA viruses are examples of such novel therapeutic targets. A better understanding of packaging mechanisms also enables their exploitation in bionanotechnology, where viral capsids are repurposed for cargo delivery and storage.

References

- AlShaikhahmed, K., *et al.*, 2018. Dynamic network approach for the modelling of genomic sub-complexes in multi-segmented viruses. *Nucleic Acids Research* 46, 12087–12098. doi:10.1093/nar/gky881.
- Bolte, H., Rosu, M.E., Hagelauer, E., Garcia-Sastre, A., Schwemmler, M., 2019. Packaging of the influenza virus genome is governed by a plastic network of RNA- and nucleoprotein-mediated interactions. *Journal of Virology* 93. doi:10.1128/JVI.01861-18.

- Borodavka, A., Desselberger, U., Patton, J.T., 2018. Genome packaging in multi-segmented dsRNA viruses: Distinct mechanisms with similar outcomes. *Current Opinion in Virology* 33, 106–112. doi:10.1016/j.coviro.2018.08.001.
- Chandler-Bostock, R., Mata, C.P., Bingham, R.J., *et al.*, 2020. Assembly of infectious Enteroviruses depends on multiple, conserved genomic RNA-Coat protein contacts. *PLoS Pathogens* 16 (12): e1009146. doi:10.1371/journal.ppat.1009146.
- Chelikani, V., Ranjan, T., Kondabagil, K., 2014. Revisiting the genome packaging in viruses with lessons from the “Giants”. *Virology* 466–467, 15–26. doi:10.1016/j.virol.2014.06.022.
- Guo, P., Lee, T.J., 2007. Viral nanomotors for packaging of dsDNA and dsRNA. *Molecular Microbiology* 64, 886–903. doi:10.1111/j.1365-2958.2007.05706.x.
- Ilca, S.L., *et al.*, 2019. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. *Nature* 570, 252–256. doi:10.1038/s41586-019-1229-9.
- Masters, P.S., 2019. Coronavirus genomic RNA packaging. *Virology* 537, 198–207. doi:10.1016/j.virol.2019.08.031.
- Mazeaud, C., Freppel, W., Chatel-Chaix, L., 2018. The multiples fates of the flavivirus RNA genome during pathogenesis. *Frontiers in Genetics* 9, 595. doi:10.3389/fgene.2018.00595.
- Mendes, A., Kuhn, R.J., 2018. Alphavirus nucleocapsid packaging and assembly. *Viruses* 10, 138. doi:10.3390/v10030138.
- Patel, N., *et al.*, 2015. Revealing the density of encoded functions in a viral RNA. *Proceedings of the National Academy of Sciences of the United States of America* 112, 2227–2232. doi:10.1073/pnas.1420812112.
- Patel, N., *et al.*, 2017a. Rewriting nature’s assembly manual for a ssRNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 114, 12255–12260. doi:10.1073/pnas.1706951114.
- Patel, N., *et al.*, 2017b. HBV RNA pre-genome encodes specific motifs that mediate interactions with the viral core protein that promote nucleocapsid assembly. *Nature Microbiology* 2, 17098. doi:10.1038/nmicrobiol.2017.98.
- Rao, V.B., Feiss, M., 2015. Mechanisms of DNA packaging by large double-stranded DNA viruses. *Annual Review of Virology* 2, 351–378. doi:10.1146/annurev-virology-100114-055212.
- Rein, A., 2019. RNA packaging in HIV. *Trends in Microbiology* 27, 715–723. doi:10.1016/j.tim.2019.04.003.
- Shakeel, S., *et al.*, 2017. Genomic RNA folding mediates assembly of human parechovirus. *Nature Communications* 8, 5. doi:10.1038/s41467-016-0011-z.
- Twarock, R., Leonov, G., Stockley, P.G., 2018. Hamiltonian path analysis of viral genomes. *Nature Communications* 9, 2021. doi:10.1038/s41467-018-03713-y.
- Twarock, R., Stockley, P.G., 2019. RNA-mediated virus assembly: Mechanisms and consequences for viral evolution and therapy. *Annual Review of Biophysics* 48, 495–514. doi:10.1146/annurev-biophys-052118-115611.

Relevant Websites

<http://viperd.b.scripps.edu>
VIPERdb.

Virus Factories

Isabel Fernández de Castro,¹ Raquel Tenorio,¹ and Cristina Risco, Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Aggresome Perinuclear inclusion body assembled around the microtubule organizing center (MTOC) and surrounded by a cage of vimentin where the cell encloses misfolded proteins. Some DNA viruses hijack this pathway to concentrate materials and build their factories.

Autophagy Mechanism for degradation and recycling of cellular components. Expendable constituents are isolated inside a double-membrane organelle known as autophagosome that later fuses with a lysosome where materials are degraded. Some RNA viruses induce and use components of this pathway for building their factories.

Correlative light and electron microscopy Group of methods that combine light and electron microscopy imaging of the same specimen. Light microscopy offers a large field of view and imaging the dynamics of living cells.

Particular cells with interesting features are selected for high resolution electron microscopy (EM). This approach is very useful to study the biogenesis of virus factories.

Replication organelle Membranous structure that harbors viral replication complexes inside the viral factory. It is usually made of single-membrane vesicles or spherules, double-membrane vesicles, multi-membraned vesicles, tubule-vesicular cubic membranes or planar oligomeric arrays.

Viral replication complex Macromolecular structure constituted by viral polymerases, other viral non-structural proteins and cell co-factors that makes multiple copies of the viral genome and transfers them to the assembly sites.

Virus factory or viral factory Intracellular compartment that contains the viral replication organelles and the sites for viral particle assembly and maturation.

Introduction

Viruses are obligate intracellular parasites. In the extracellular environment, viral particles exist as rather simple macromolecular entities made of protein shells also known as capsids, and/or membranes, also known as envelopes, that protect and transport the viral genome from cell to cell. It is only during the intracellular phase of their life that viruses synthesize distinct molecules and perform activities that are characteristic of living entities. One of these activities is the construction of a platform for genome replication and morphogenesis, also known as virus factory (VF). The central element of the factory is the replication organelle (RO) where the viral replication complexes (VRC) make multiple copies of the viral genome. In many cases, factories are made of remodeled cell membranes and contain functional compartments with specific roles in virus replication, assembly and egress. Other cell elements such as mitochondria and the cytoskeleton are often recruited to viral factories and make contacts with the RO. In most cases, the specific role of these contacts is not known. VFs can be rather big, with a diameter of several microns and occupy large intracellular regions. VFs are dynamic structures that maintain their communication with intracellular transport pathways to incorporate essential components and to facilitate the exit of newly formed viral particles (Fig. 1). Furthermore, VFs change over time with the assembly and disassembly of particular structures needed for specific steps of the virus life. In some cases, VFs or only the ROs are mobile compartments that move inside the infected cell, like in the case of the early ROs of Vaccinia Virus, Hepatitis C Virus, rotaviruses and the mouse hepatitis coronavirus. In the case of some plant viruses, ROs are transported from cell to cell through intercellular channels known as plasmodesmata.

To build their replication neo-organelles and factories, viruses make use of cellular materials and signaling pathways. The study of VFs and the way they hijack cellular pathways give also insight in the pathway itself. Researchers have identified two main pathways hijacked by viruses to build their factories: The aggresome formation for DNA viruses and autophagy for RNA viruses. Moreover, viruses manipulate lipid synthesis and flows to create membranes with particular lipid composition and biophysical properties that are ideal for the activation and efficient performance of the macromolecular complexes that work in viral genome replication and assembly.

Although present in many VFs, it is not known why and how mitochondria are recruited to viral factories. It has been proposed that mitochondria would provide the energy storage molecule adenosine triphosphate (ATP) necessary for virus replication, but direct evidence is lacking. Alternatively, viruses may need mitochondrial components such as particular proteins that incorporate into the replication organelles where they would enhance viral genome replication. Another possibility is that viruses recruit mitochondria around VFs to block the innate antiviral immunity response mediated by these organelles.

Methods to Study Virus Factories

The study of virus factories includes the characterization of their structure, biogenesis, functions, and evolution over time (Fig. 1). Virus factories were identified with light and electron microscopy imaging techniques. The first evidences of the existence of VFs

¹These authors contributed equally to this manuscript.

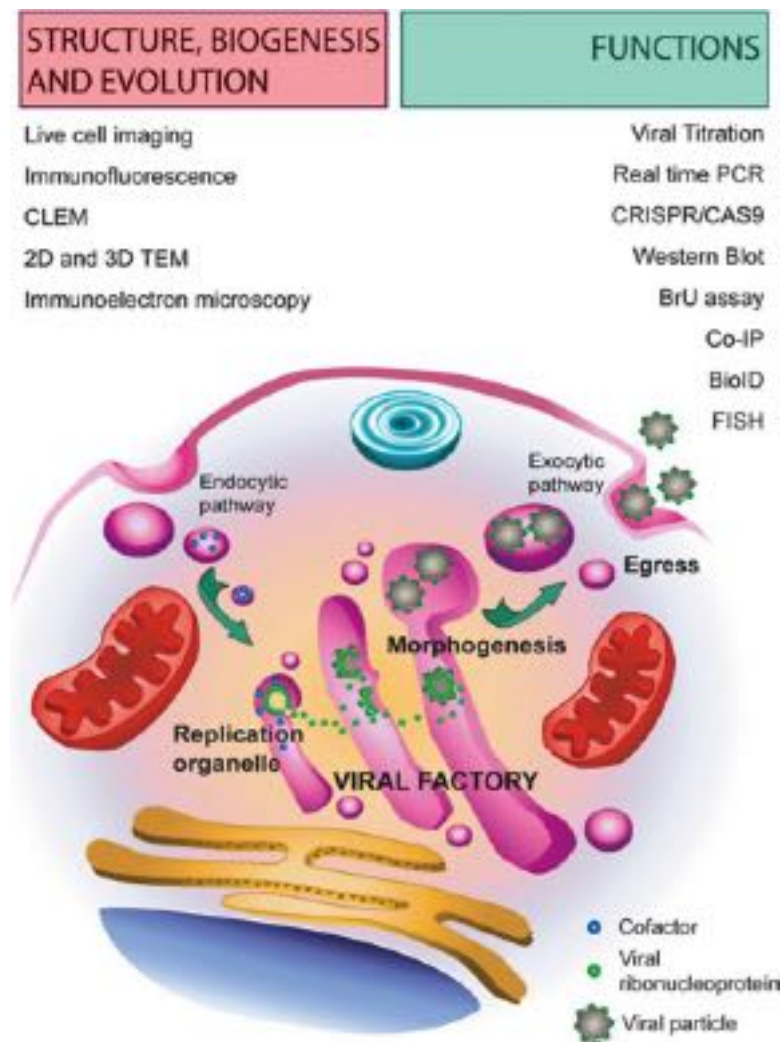


Fig. 1 Main components and functions of virus factories and methods for their characterization. Techniques for characterizing the structure, biogenesis and evolution of VFs are listed on the left, those for studying their functions are listed on the right.

came from studies of thin-sections of infected cells visualized by transmission electron microscopy (TEM). These images showed viral particles associated with macro-structures made of organelles that had been remodeled and transported to particular intracellular locations. Light microscopy (LM) then showed how viruses change the organization of whole infected cells to build their VFs. Nowadays, advances in fluorescent probes and LM methods are providing new insights into the biogenesis and functional architecture of VFs. Time-course assays combined with immunofluorescence and confocal microscopy provide a general view of the progression of infection. For these studies, cells at different times post-infection (tpi) are processed for immunolabeling with primary antibodies specific for viral proteins and secondary antibodies conjugated with fluorophores. When fluorescent viruses are available, real time, live cell microscopy shows an accurate view of all steps of the viral infection, including the early stages of VF assembly (**Fig. 2(a)–(c)**). Moreover, simultaneous labeling of cell compartments with specific markers allows the identification of particular organelles subverted by the virus and provides a dynamic view of their transformation during VF assembly.

For a detailed analysis of VFs and their components, the high resolution of transmission electron microscopy is needed. The size of most viruses is beyond the resolution limit of optical microscopy and EM provides the necessary resolution to study viruses and their VFs. The most common approach is the study of thin sections of virus-infected cells. These sections must be thin enough (~50–100 nm) for electrons to go through and generate a projection image in the TEM. With advances in sample preparation methods for TEM we are getting close to the visualization of the VF native-like structure. Cryo-preparation techniques such as high-pressure freezing followed by freeze-substitution or frozen-hydrated sections are some of the options to visualize VFs close to their natural state. In addition, ultrastructural imaging can be done either in two dimensions (2D) or three dimensions (3D). The images a TEM generates from a thin section are 2D projections of three-dimensional, much larger structures. In these projections, we can observe considerable details about the elements that build the factory (**Fig. 2(d)**) but the spatial information within the volume of the structure is lost. Different 3D imaging techniques for biological samples are now available. They are unveiling how

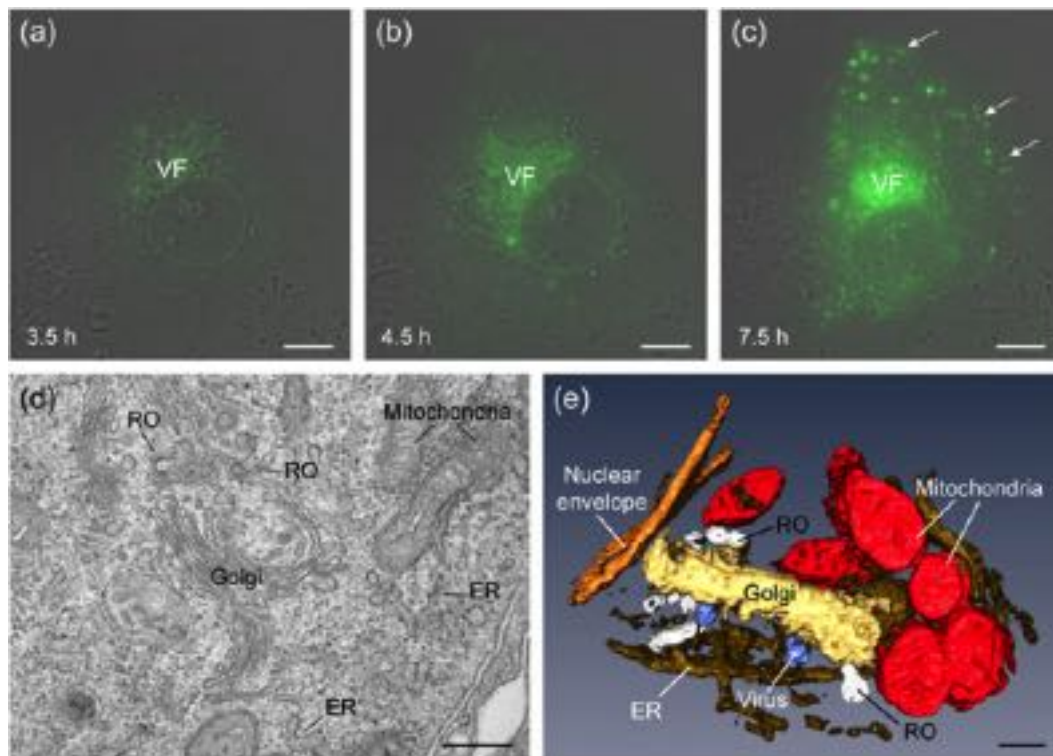


Fig. 2 Visualization of a viral factory by live cell imaging, 2D TEM and 3D TEM. (a, b, c) Snapshots from a live cell microscopy movie showing the biogenesis and evolution of a bunyavirus factory (VF). BHK-21 cells were infected with the fluorescent bunyavirus eGFP-Gc-BUNV and transferred to a TCS-SP5 microscope (Leica Microsystems). Images were recorded every 15 min from 1 to 18 h post-infection (hpi). (a) Initial stage of factory assembly. (b) The viral factory grows over time with perinuclear localization. (c) Late stage of infection with the large factory and some secretory vesicles (arrows) with viruses at the plasma membrane. (d) Ultrathin section of viral factories observed after conventional embedding and TEM. BHK-21 cells were adsorbed with bunyavirus and fixed at 10 hpi; Golgi stacks and replication organelles (RO) that contain the replication complexes are shown with mitochondria and endoplasmic reticulum (ER) membranes surrounding the factories. (e) Factory as visualized by TEM of serial sections, 3D reconstruction and image processing. Vero cells were adsorbed with bunyavirus for 10 h. The modified Golgi stack (gold) with spherules/ROs (white) and newly formed viral particles (blue) is surrounded by mitochondria (red) and ER (brown), adjacent to the nuclear envelope (orange). Scale bars = 10 μ m in (a, b, c); 500 nm in (d); and 200 nm in (e).

different elements compose the VF macro-structure inside cells. Methods such as 3D reconstruction of serial sections (Fig. 2(e)) and electron tomography (ET) are currently applied to study the three-dimensional architecture of viral factories. 3D studies of VFs discovered inter-organelle contacts that do not exist in non-infected cells.

The group of techniques known as correlative light and electron microscopy (CLEM) combines both types of microscopy and allows the selection of particular cells with interesting features detected by (live) light microscopy imaging to study them at higher resolution by TEM. This strategy is very useful to study the different steps of the infection, such as the assembly of the VF and the heterogeneous response of different cells to the virus.

Microscopy studies show when, where and how the VF is built and therefore provide useful information about the most adequate tpi for other studies, such as the quantification of viral genome synthesis with real-time polymerase chain reaction (RT-PCR), viral protein production by western-blot and the production of intracellular and extracellular infectious viruses with titration assays. All these methods work together to identify VF functions in the infected cell. However, to assign particular activities to VF sub-compartments, an *in situ* molecular mapping of the structures is necessary. Detection of viral and cellular macromolecular complexes in VFs can be done with fluorescent probes and super-resolution light microscopy techniques. Stimulation emission depletion (STED), stochastic optical reconstruction (STORM) and total internal reflection fluorescence (TIRF) microscopy are some of these techniques that produce images at higher resolution than the one imposed by the diffraction limit. Their application in virology is still limited but they have a great potential to study the macromolecular architecture of VFs. At the ultrastructural level, molecular mapping is done with specific primary antibodies and secondary antibodies conjugated with colloidal gold particles in immunogold labeling assays. Recently, clonable tags for EM have been developed. The method known as metal-tagging TEM (METTEM) uses the metal-binding protein metallothionein (MT) coupled with a gold nano-cluster as an electron-dense tag. MT has unveiled the 3D organization of the Tombusvirus polymerase molecules in ROs and the movement of newly synthesized influenza virus ribonucleoproteins from factories to the plasma membrane before viral particle assembly and egress.

To identify the active ROs in infected cells, we count with assays of brome-uridine or brome-deoxyuridine incorporation for RNA or DNA viruses, respectively. These assays can be combined with immunofluorescence and confocal microscopy or immunogold labeling and electron microscopy to localize the sites where the viral polymerases are making new copies of the viral genome. In addition, probes for imaging specific viral RNA molecules are powerful tools to study viral replication and assembly in live cells. For example, fluorescence in situ hybridization (FISH) shows where the viral genome localizes and assembles to form new viral particles. Live FISH is a new technique under development. Never used for VF studies, it might show a dynamic view of viral genome synthesis, transport and assembly in virions.

A very active field of research is the identification of cell factors used by viruses to remodel compartments and build their factories. Gene expression microarrays can give us some clues by showing the genes that are over- or under-expressed in infected cells. Also, because viral non-structural (NS) proteins are usually involved in the biogenesis of ROs and VFs, techniques that detect protein-protein interactions such as yeast two-hybrid (Y2H), co-immunoprecipitation and proximity-dependent biotin identification (Bio-ID) followed by proteomics can catch relevant cell factors that interact with NS viral proteins early in infection. Validation of candidates can be laborious and involves methods such as gene silencing with small interference RNA (siRNA) or gene deletion using the CRISPR-Cas9 technology. Transient expression of viral and cell proteins after transfection with plasmids together with the generation of stable cell lines to control the expression of specific (tagged)-proteins are very useful techniques for these studies. The impact of the over-expression or elimination of the selected candidates in VF assembly and function is analyzed with the morphological and functional studies of VFs described above. When a fluorescent virus is available, another strategy is possible. Cells are infected with the recombinant fluorescent virus and at different tpi fluorescent and non-fluorescent cells are separated by cell sorting. Different cell populations can be then studied by EM, Proteomics, Transcriptomics, and Lipidomics to search for cell factors participating in viral infection in general and VF biogenesis in particular.

Representative Examples of Virus Factories

With the examples of selected DNA and RNA viruses, this section revises some typical viral factories. Most currently known DNA viruses carry out replication and transcription either entirely or partially within the nucleus of the host cell. For these viruses, the nucleus provides the machinery required for particular steps of the viral life cycle. Due to the limited comprehension of the functional architecture of the cell nucleus, the organization of nuclear factories is poorly understood.

Some of the best studied nuclear VFs are the ones of Polyomaviruses (PyV). PyVs are small, non-enveloped DNA viruses that infect mammals and birds. They have also been associated with the development of cancers in their hosts. During PyV infection, viral DNA and capsid proteins concentrate in nuclear bodies, suggesting that these sites may function as virus factories. However, PyV active DNA replication has been located adjacent to these bodies, associated with the recruitment of cellular factors such as DNA repair-related proteins. By electron microscopy and electron tomography, PyV factories are seen as collections of proteinaceous tubular structures and clusters of viruses. The current model proposes that PyV assembles from the ends of these tubules by a budding process and that nuclear bodies may provide the necessary architectural foundation for PyV tubular structures.

Some DNA viruses use also the cytosol for some steps of their life cycle. This is the case of herpesvirus, which causes many different diseases and is associated with human tumors. After entering cells, viral particles are transported to the nuclear envelope and the viral DNA is injected into the nucleus. Viral DNA replication, protein expression and core assembly occur inside nuclear factories. Then, immature viral cores exit the nucleus and enter the *trans*-Golgi network (TGN), where they acquire their final envelope. Thus, herpesvirus needs the cytosol and Golgi complex to complete the maturation of the virus progeny.

In contrast to most DNA viruses, the infection cycles of poxvirus and mimivirus take place exclusively within the host cytoplasm. Poxviruses such as vaccinia virus (VACV), are large, enveloped DNA viruses that infect vertebrate and invertebrate species. VACV replication requires rearrangement of vimentin filaments to build a vimentin cage, structure that assembles close to the microtubule organizing center (MTOC) and resembles the cellular aggresome. Before the formation of an aggresome-like structure, VACV factories are made of ER-derived membranes that build a mini-nucleus where viral genome replication takes place. Thus, VACV assembles two different factories, the first one is a mini-nucleus for viral genome replication and the second one is an aggresome-like compartment for viral morphogenesis.

Unlike DNA viruses, almost all RNA viruses form factories exclusively in the cytoplasm. These factories are in tight association with host membranes and depending on the virus, they use different cell organelles. Flaviviruses such as hepatitis C virus (HCV), dengue virus (DENV), or Zika virus (ZIKV) are positive-sense RNA viruses and important human pathogens. HCV is the major causative agent of chronic liver disease leading to cirrhosis and cancer. Early in infection, HCV induces a remodeling of ER membranes forming double-membraned vesicles (DMVs), and later on, it induces the formation of multi-membraned vesicles (MMVs) that are composed of several concentric membrane bilayers and accumulate at late tpi. These membranous rearrangements are produced by the action of viral nonstructural proteins. Viral RNA replication seems to be associated with DMVs which constitute the major component of HCV factories. DENV, the causative agent of dengue fever, is another flavivirus that induces rearrangements of ER membranes. EM has shown ER-derived structures with distinct morphologies: convoluted membranes (CMs) and vesicle packets (VPs). Using ET, it was demonstrated that these structures form part of a single ER-derived network. Furthermore, ET has allowed the study of the 3D architecture of DENV-induced ER membrane arrangements, revealing vesicles with pores that could enable the release of newly synthesized viral RNA. In addition, these studies showed the DENV particles budding into ER membranes, directly joined to vesicle pores. ZIKV is an emerging flavivirus that has been linked to abortion and birth defects associated

with severe infection of the central nervous system (CNS). Like other flaviviruses, ZIKV induces rearrangements of the ER forming characteristic vesicles and other membranous structures where viral replication and particle assembly occur in a coordinated manner. In addition, newly synthesized viral particles aggregate forming arrays of viruses in the lumen of the ER.

Coronaviruses (CoV) and arteriviruses are also positive-sense RNA viruses that generate membrane alterations in the host cytoplasm. There are common features between their virus factories, despite the significant differences in their morphology. The CoV group includes important human pathogens such as the severe acute respiratory syndrome (SARS) viruses. ET was decisive to reveal that coronaviruses form a unique reticulovesicular membrane network from the ER, where viral replicase subunits and double-stranded RNA (dsRNA), a viral replication intermediate, are both located. The membranous platform is continuous with the rough ER and contains in its lumen numerous vesicles that strongly stain for anti-dsRNA antibody. VP and CM structures have been also detected in association with the ER network. Equine arteritis virus (EAV) is the prototype of the family *Arteriviridae*, which includes important pathogens of animals. Arteriviruses also modify ER membranes, forming a network characterized by the accumulation of double-membraned sheets and DMVs. The diameter of these DMVs is smaller compared to those induced by CoV, and CMs are not observed. In addition, ET has shown that DMVs associate with each other in the membranous web. All these studies show that flaviviruses, coronaviruses and arteriviruses induce a single ER-derived endomembrane platform where viral replication and assembly are interconnected.

Many positive-sense RNA viruses replicate in single membrane invaginations named spherules that assemble in a variety of cell organelles. For example, togaviruses assemble spherules in lysosomes. The *Togaviridae* family is composed of two genera which include important animal and human pathogens, such as the Semliki Forest virus (SFV) and rubella virus (RUBV). Togavirus spherules are first formed at the plasma membrane, internalized by endocytosis and finally fused with late endosomes and lysosomes. These modified lysosomes, called cytopathic vacuoles (CPVs), are the central element of the VF that harbors viral replication and morphogenesis. Mitochondria and ER are recruited around the CPVs. In the case of RUBV, the Golgi complex is also recruited and participates in virus morphogenesis.

Other viruses assemble spherules in the ER, the Golgi complex, mitochondria, chloroplasts or peroxisomes. For example, the nodavirus Flock House virus (FHV), which can replicate in insect, plant, mammalian and yeast cells, assembles spherules in mitochondria. ET and 3D visualization have shown that spherules form by invaginations of the outer mitochondrial membrane and communicate with the cytoplasm through a "neck". Tombusviruses that are frequently used as a model to study virus-host interactions, assemble their spherules in the peroxisomes of plant and yeast cells. With advanced EM techniques, it was demonstrated that a cellular protein, the ESCRT (endosomal sorting complexes required for transport) component Vps4 controls the traffic of macromolecules between the spherule and the cytosol.

Negative-sense RNA viruses also build replication factories although their structure and biogenesis are less characterized. Bunyaviruses are an exception because their large VF has been studied in detail by light and electron microscopy. Bunyaviruses constituted a large order of enveloped viruses that includes emerging pathogens of humans, animals and plants. Bunyavirus factories use the Golgi complex, where the ROs are formed and consist of typical spherules connected with atypical cylindrical structures, surrounded by mitochondria and ER membranes. Viral replication takes place in spherules but the role of the cylinders is not understood. Serial sectioning and 3D reconstruction have shown the interactions between the different elements of this complex factory. The analysis of volumes together with immunogold labeling and TEM, suggested that the cylindrical structures probably mediate the transport of essential viral and cell factors throughout the factory. The virus assembly sites locate near spherules with VRCs. Viral particles bud in *cis*-Golgi membranes and mature in the *trans*-Golgi and *trans*-Golgi network (TGN) subcompartments.

From the examples described in this section, we can conclude that the ER that is the largest organelle in eukaryotic cells, is also the favorite compartment for viruses to build the VFs. One of the most spectacular viral-induced ER remodeling is triggered by the human reovirus. This dsRNA virus causes disease in children and young mammals, has been involved in the etiology of the celiac disease and is currently being used as an oncolytic in anti-cancer therapies. The human reovirus uses the ER to build their VFs also known as viral inclusions. To that purpose, two non-structural viral proteins, σ NS and μ NS bind to the peripheral ER elements and fragment them to make a collection of small membranous tubules and vesicles that aggregate and remain attached to the ER framework. These groups of membranes concentrate and protect the viral core-associated replication complexes and possibly provide also physical support for viral particle assembly.

Viral Factories as Targets for Antiviral Therapies

Emerging and re-emerging viruses constitute a serious public health problem and the need for new broad-spectrum antivirals is a top priority. Generally, antiviral drug development has focused on targeting specific viral components or on the modulation of the host immune response. Other strategies aim to develop molecules that target host cellular factors needed for viral replication. In this case emergence of resistant viruses is less likely. Moreover, the approaches focused on host cellular pathways used by viruses can favor the development of broad spectrum antivirals. Because different viruses use the same cell compartments and signaling pathways to build their factories, the identification of factors involved in VF biogenesis can provide new ways for antiviral design. For example, molecules targeting lipid transfer proteins (LTPs), enzymes of lipid metabolism, organelle remodeling proteins, and those involved in intracellular transport are good candidates for antiviral therapies. In addition, the current strategy known as drug repurposing is an attractive way to save time by finding molecules already approved to treat other diseases and use them to treat viral infections.

Further Reading

- Altan-Bonnet, N., 2017. Lipid tales of viral replication and transmission. *Trends in Cell Biology* 27, 201–213.
- Erickson, K.D., Bouchet-Marquis, C., Heiser, K., *et al.*, 2012. Virion assembly factories in the nucleus of polyomavirus-infected cells. *PLoS Pathogens* 8, e1002630.
- Fernández de Castro, I., Volonté, L., Risco, C., 2013. Virus factories: Biogenesis and structural design. *Cellular Microbiology* 15, 24–34.
- Fernández de Castro, I., Sanz-Sánchez, L., Risco, C., 2014. Metallothioneins for correlative light and electron microscopy. *Methods in Cell Biology* 124, 55–70.
- Fernández de Castro, I., Tenorio, R., Risco, C., 2016. Virus assembly factories in a lipid world. *Current Opinion in Virology* 18, 20–26.
- Fernández-Oliva, A., Ortega-González, P., Risco, C., 2019. Targeting host lipid flows: Exploring new antiviral and antibiotic strategies. *Cellular Microbiology* 21, e12996.
- Fridmann-Sirkis, Y., Milrot, E., Mutsafi, Y., *et al.*, 2016. Efficiency in complexity: Composition and dynamic nature of mimivirus replication factories. *Journal of Virology* 90, 10039–10047.
- Harak, C., Lohmann, V., 2015. Ultrastructure of the replication sites of positive-strand RNA viruses. *Virology* 479–480, 418–433.
- Kopeck, B.G., Perkins, G., Miller, D.J., Ellisman, M.H., Ahlquist, P., 2007. Three-dimensional analysis of a viral RNA replication complex reveals a virus-induced mini-organelle. *PLoS Biology* 5, e220.
- Mutsafi, Y., Fridmann-Sirkis, Y., Milrot, E., Hevroni, L., Minsky, A., 2014. Infection cycles of large DNA viruses: Emerging themes and underlying questions. *Virology* 466–467, 3–14.
- Risco, C., Fernández de Castro, I., Sanz-Sánchez, L., *et al.*, 2014. Three-dimensional imaging of viral infections. *Annual Review of Virology* 1, 453–473.
- Romero-Brey, I., Bartenschlager, R., 2014. Membranous replication factories induced by plus-strand RNA viruses. *Viruses* 6, 2826–2857.
- Sachse, M., de Castro, I.F., Fournier, G., Naffakh, N., Risco, C., 2018. Metal-tagging transmission electron microscopy and immunogold labeling on Tokuyasu cryosections to image influenza A virus ribonucleoprotein transport and packaging. *Methods in Molecular Biology* 1836, 281–301.
- Sachse, M., Fernández de Castro, I., Tenorio, R., Risco, C., 2019. The viral replication complexes in cells studied by electron microscopy. *Adv. Virus Res.* 105, 1–33.
- Strating, J.R., van Kuppeveld, F.J., 2017. Viral rewiring of cellular lipid metabolism to create membranous replication compartments. *Current Opinion in Cell Biology* 47, 24–33.
- Wileman, T., 2007. Aggresomes and pericentriolar sites of virus assembly: Cellular defense or viral design? *Annual Review of Microbiology* 61, 7905–7912.

Relevant Websites

- <https://youtu.be/NKtTv5xVJ0U>
Annual Review of Virology.
Viral Manipulation of Plant Host Membranes.
- <https://youtu.be/57RwV2nkSkU>
Flaviviridae replication organelles.
- https://youtu.be/_GrXSClpbDw
Three-dimensional imaging of viral infections.
- <https://viralzone.expasy.org/1951>
Viral factories.
ViralZone page.

Release of Phages From Prokaryotic Cells

Jesse Cahill, Sandia National Labs, Albuquerque, NM, United States

Ry Young, Texas A&M University, College Station, TX, United States

© 2021 Published by Elsevier Ltd.

Glossary

IM Cytoplasmic membrane or inner membrane.

i-spanin The inner membrane spanin subunit. A type-II integral membrane protein (N-in, C-out topology). Forms a complex with the o-spanin that spans the periplasmic space. Both subunits are required for outer membrane disruption during lysis.

lipobox The lipobox is a four aa sequence at the C-terminus of SPII signal sequences. The consensus structure is {L, V, I}-X-{G, A, S}-C, where X is any aa except Pro or a charged residue. In *E. coli*, after secretion of the lipoprotein, the Cys residue of the lipobox is lipoylated by Lgt, before SPII cleavage and by Lnt, after SPII cleavage, resulting in attachment of the lipoprotein to the outer leaflet of the IM. If the residue following Cys is any residue but Asp, the fully-lipoylated mature lipoprotein is sorted by the LOL system to the inner leaflet of the OM.

MGL Multi-gene lysis. These systems feature two components (holins and endolysins) to lyse Gram-positive bacteria. Gram-negative bacteria require three components: holins, endolysins, and spanins for disruption of the IM, PG, and OM, respectively.

OM Outer membrane.

OMM Outer mycolate membrane.

o-spanin The outer membrane spanin subunit. A lipoprotein anchored in the inner leaflet of the outer membrane.

PG Peptidoglycan.

SAR endolysins Signal anchor release endolysins.

Accumulates as an IM protein with a single TMD and an N-in, C-out topology. Unlike a typical TMD, the SAR domain is rich in glycine and alanine. SAR endolysins are inactive until the SAR domain exits the bilayer. This occurs when the membrane is depolarized or can occur spontaneously at a low rate.

SGL Single gene lysis.

SPI The product of the *lep* gene in *E. coli*, Signal Peptidase I is the enzyme that cleaves the common secretory signal for periplasmic and outer membrane proteins after export through the Sec or Tat systems. SPI signal sequences are very diverse but have a common domain structure, including a basic N domain, a hydrophobic H domain and a hydrophilic domain of 3–7 aa, leading up to the cleavage site.

SPII The product of the *lsp* gene in *E. coli*, Signal Peptidase II (Lsp) is an enzyme involved in the processing of lipoprotein precursors. SPII signals resemble SPI signals except that the former ends in a lipobox motif (see below). SPII cleaves at the terminal Cys residue of the lipobox, so all mature lipoproteins have an N-terminal Cys residue.

TMD Transmembrane domain.

u-spanin A unimolecular spanin. In contrast to the two-component spanin systems, which use both an i-spanin and an o-spanin, the u-spanin is a single peptide with a lipoprotein OM anchor and a C-terminal transmembrane domain. U-spanins disrupt the outer membrane during lysis.

Introduction

Viruses of prokaryotic cells generally face two problems in achieving release from the infected host, at least in planktonic populations. First, the cell typically has a tough cell envelope evolved to withstand chemical, mechanical and thermal insult. Eubacteria, for example, have a cross-linked peptidoglycan (PG) arranged in a mesh-like layer with interstices that are normally far too small to allow virion release. Second, since the availability of host cells for the progeny is uncertain the timing of lysis may be important. In other words, continuing an infection cycle rather than causing cellular disruption may be a better strategy if the infected cell is in a host-poor environment, so appropriate regulation of the lytic event should confer a significant fitness advantage. In this review, the molecular strategies that have evolved to deal with these issues will be discussed. As the reader should soon discover, the strategies that have already been discovered cover the spectrum of complexity. These stories come mostly from studies focused on the phages of *E. coli* and other model eubacteria and are consequently well documented and genetics-based. However, most of the new information about viral life cycles is derived from “omics” and high-resolution microscopy; moreover, the host is often a poorly characterized pathogenic bacterium or an exotic archaeal species. Far less is established in these systems but even so, enough is already known to guarantee that the lysis mechanisms will be even more diverse and unusual.

In terms of organization, this review will focus first on “Multigene Lysis” (MGL) (Fig. 1), the general strategy used by the *Caudovirales*, complex double-stranded DNA (dsDNA) phages that dominate the prokaryotic virosphere and have genomes containing up to ~700 genes. Phages that use MGL encode at least one lysis protein to target each structural layer of the cell envelope. The minimal set is two proteins: A holin and an endolysin, targeting the cytoplasmic membrane and cell wall (PG) layer, respectively. However, in many cases, other proteins are encoded for regulation of lysis at the functional level or to target additional envelope components, like the outer membrane (OM) of Gram-negative hosts or the mycolate outer layer of some Actinobacteria. The primary platform for unraveling the molecular nature of these lysis systems has been phage lambda, because of

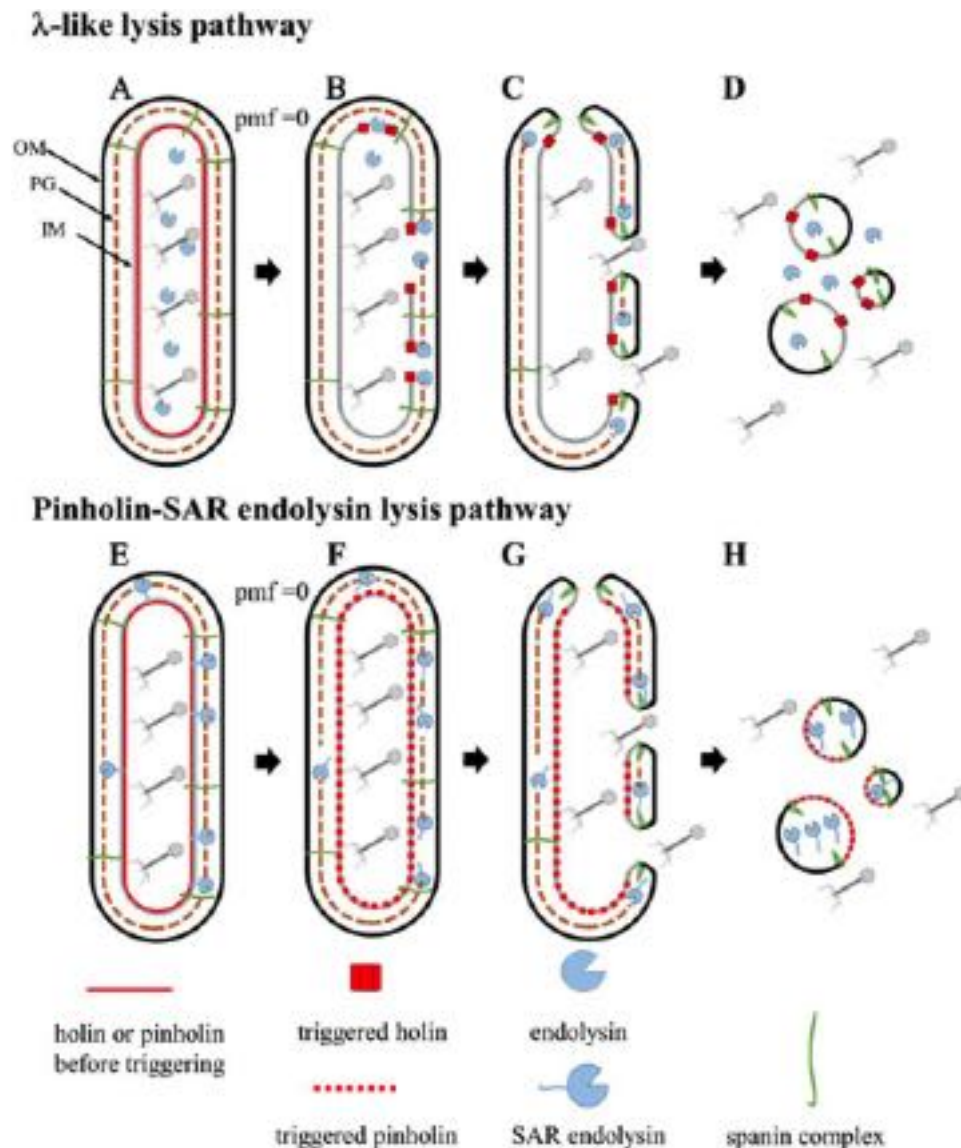


Fig. 1 (A)–(D) The lysis model for lambda-like MGL systems. During phage replication the holin, endolysin and spanins accumulate in the indicated cellular compartments. Holin triggering (B) results in the sudden formation of micron-scale holes in the inner membrane which releases the endolysin into the periplasm where it degrades its substrate peptidoglycan. Spanins are activated by degradation of the surrounding peptidoglycan and cause outer membrane disruption by fusing the inner and outer membrane (C)–(D) at the last step of lysis. (E)–(H): The lysis model for pinholin-SAR endolysin MGL systems. The SAR-endolysin accumulates in an inactive form in the inner membrane. Pinholin triggering (F) results in the formation of ~2 nm “pinholes” that collapse the PMF. Upon depolarization, the SAR endolysin is released from the bilayer and activated to degrade the peptidoglycan (G). The spanins then fuse the IM and OM at the last step of lysis (H). Reprinted from Cahill, J., Young, R., 2019. Phage lysis: Multiple genes for multiple barriers. *Advances in Virus Research* 103, 33–70.

its unparalleled genetic facility, so considerable space is devoted to recounting the work that originally defined lysis protein function in this system. The next section, standing in stark contrast to the complex pathways and sophisticated coordination evident in the MGL systems, is devoted to Single Gene Lysis (SGL) (Fig. 2), the strategy used by the simplest lytic phages, the ssDNA *Microviridae* (~5 kb; ~10 genes) and the ssRNA *Leviviridae* (~4 kb; 3–4 genes). In these systems, a single polypeptide, usually small (50–100 residues), is produced during infection. In the few cases where the molecular basis has been established, the SGL protein acts as an inhibitor of cell wall biosynthesis, causing the term “protein antibiotic” to be invoked. For one class of phages of eubacteria, the filamentous *Inoviridae*, phage release is achieved without lysis, by virtue of extrusion of the viral particle through the intact cell envelope. Finally, a brief overview of how virions are released in archaeal infections will be presented. We hope the archaeal community will forgive our inclusion of these viruses in a phage review; indeed, since many viruses of archaea are morphologically indistinguishable from phages of eubacteria, we will use the term phage for any virus of a prokaryotic cell, rather than switching back and forth from phage to virus in terms of both shared and different lysis features.

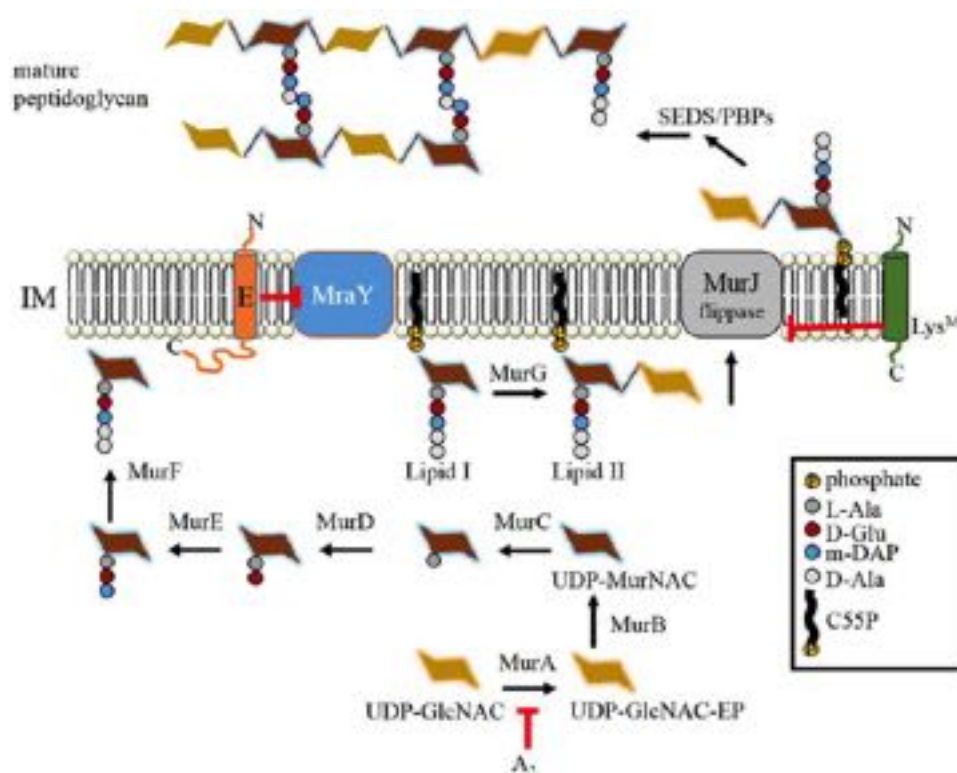


Fig. 2 Some Sgl proteins act as “protein antibiotics” by inhibiting different steps of the PG synthesis pathway. The lysis proteins E, Lys^M, and A₂ of phages ϕ X174, M, and Q β inhibit MurA, MurJ, and MurA, respectively.

The Definition of Phage Lysis

It is important to define what is meant here by “phage lysis”. Generally, absent the lysogenic response or some sort of cell defense event, the vegetative phase of a phage infection (i.e., the intracellular production of progeny virions) is invariably lethal to the host. There are reports of alternative outcomes in which virions are produced but the cell remains viable in some sort of carrier state, but these have not been sufficiently well-characterized genetically or physiologically and may reflect heterogeneous outcomes within the infected cell population. The lethality, in terms of loss of genetic viability of the infected bacterial cell, is detectable as soon as irreversible adsorption has occurred, but physiological viability, in terms of macromolecular synthesis and energy metabolism, is maintained until the phage lysis proteins overtly terminate the infection by destruction of the cell envelope. It seems likely that lysis genes were the last essential genes evolved by phages. The primordial phage presumably had no lysis genes and relied on the generalized failure of the cell envelope that would eventually occur as a result of re-direction of the resources of the cell into production of virus particles and the consequent neglect of homeostatic processes needed for envelope maintenance. Thus even in the absence of functional lysis genes, eventually infected cells will simply fall apart, freeing the progeny virions. Among devotees of lysis functions, this type of outcome is called cellular “rotting”, associated with the gradual loss of optical density in an infected culture over a time scale of hours. In contrast, overt lysis imposed by the phage lytic program occurs on a time scale of minutes. Indeed, there is evidence that adjusting the timing of lysis by just a few minutes is a valuable fitness adjustment (see below), sufficient to cause the evolution of regulatory proteins called antiholins. In any case, unlike any other “essential” gene, the holin-negative phenotype includes a dramatic increase in virion accumulation in the infected cell; this can complicate phenotypic analysis but also be of value when assigning holin function.

The Cell Envelope

A brief overview of the cell envelope and the sorting of envelope proteins, is in order, focusing on Gram-negative cell where most of our knowledge of lysis mechanisms lies. The minimum envelope consists of the cytoplasmic or inner membrane (IM), the cell wall or peptidoglycan layer (PG), and the outer membrane (OM); the compartment between the IM and OM is the periplasm (Fig. 3). The IM is a classic phospholipid bilayer but the OM is asymmetric, with an inner phospholipid leaflet and an outer leaflet composed of lipid A decorated with polysaccharide chains featuring phosphate and other anionic moieties. The OM is rich in non-specific porins, pore-forming beta-barrel proteins that allow free diffusion of soluble molecules $< \sim 600$ Da, so in general the pH,

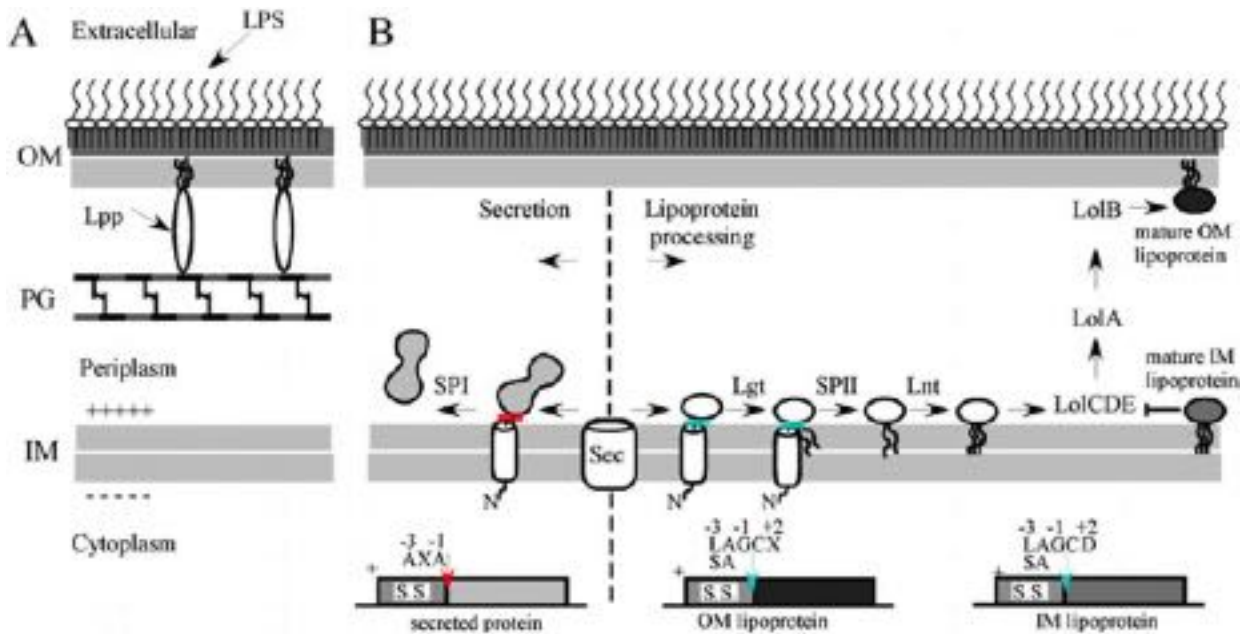


Fig. 3 (A) The outer membrane (OM), peptidoglycan (PG), and inner membrane (IM) are labeled to the left. Lpp and lipopolysaccharide (LPS) are denoted with arrows. (B) Pathways for Sec-mediated secretion (left of the dashed line) or lipoprotein processing (right of dashed line). The precursor forms of periplasmic proteins pass through the Sec translocon and are cleaved by Signal peptidase I (SPI; site denoted by red dash) and released into the periplasm. The lipoprotein cleavage site is denoted with a blue dash (for the gene product) and an arrow (position of the coded cleavage site in the gene). After passing through the Sec translocon, prolipoprotein diacylglyceryl transferase (Lgt) transfers a diacylglyceride group to the sulfhydryl of the Cys marked for lipoylation. Signal peptidase II (SPII) cleaves the N-terminal TMD at the -1 position. Apolipoprotein N-acetyltransferase (Lnt) transfers an acyl group to the free amino terminus of the Cys. The presence of Asp at the $+2$ or $+3$ position determines will result in Lol avoidance for an IM lipoprotein. Conversely, small polar residues at this position will result in the transport of the lipoprotein to the inner leaflet of the OM by the Lol system.

ionic, and small nutrient content of the periplasm reflects the contents of the external medium. Nevertheless, the periplasm is maintained in an iso-osmotic state to the cytoplasm by the regulated production of polymeric membrane-derived oligosaccharides of a size ($MW > 600$) that cannot escape through the porins. The peptidoglycan is a single covalent meshwork made up of glycosidic chains averaging ~ 30 disaccharide units cross-linked by oligopeptide chains and oriented, in *E. coli*, in a generally orthogonal orientation to the long axis of the cell.

In the healthy cell, a proton-motive force (PMF) of ~ 200 mV is maintained across the IM (positive outside), due to the pH difference between the periplasm (external medium) and the cytoplasm ($pH \sim 7.5$) and electrochemical potential. There is no ATP or other energy currency molecule in the periplasm, so energy-requiring steps in the OM require transduction of free energy from the IM to the OM. Where relevant to lysis, the differences in the envelopes of Gram-positive bacteria and in archaeal cells will be discussed below.

Multigene Lysis (MGL) Systems

The Lambda Prophage as a Lysis Platform

The premier MGL system has been that of coliphage lambda, for two main reasons. First, lambda lysis phenotypes are “clean”, mostly because the infection cycle largely spares the host from undue physiological stress. For example, lambda does not interfere with host gene expression or chromosome integrity, whereas other genetic paradigm phages like T4 degrade host DNA and block both host transcription and translation. Thus, rapid onset of the “rotting” state described above does not occur in the lambda system when lysis genes are inactivated. The second main reason is experimental convenience, largely because of a famous repressor allele, cl_{857} . A culture of lysogenic cells carrying a λcl_{857} prophage can be grown at 30°C and induced by a brief heat shock at 42°C . Every cell thus enters the lytic cycle synchronously, and, if variables such as culture to flask volume ratio, shaker speed, media and growth phase are controlled, reproducible “lysis curves”, the plot of culture turbidity versus time, can be recorded. For example, a 25 ml LB culture of the standard *E. coli* K-12 strain MG1655(λcl_{857}) induced at $A_{550} = 0.2$ ($\sim 10^8$ cells/ml) in a 250 mL flask at 250 RPM rotations per min undergoes lysis at 50 min after induction, releasing ~ 100 plaque-forming units (virions) per induced cell. Using these rigorously defined conditions, an experimenter can discriminate between holin alleles that differ in lysis timing by as little as 2 min. Because of the facile nature of this system, it has generally been the practice to characterize the lysis

genes of any particular lytic phage by engineering them into lambda, using recombineering or other recombination-based methods on the prophage, or by using an inducible plasmid clone to complement the lytic defect associated with a lambda lysis mutant. Indeed, this is how most of our knowledge of the T4 lysis system was obtained.

The Lambda Lysis Cassette

The lambda lysis genes are clustered together into a cassette of four genes: *SRRzRz1* (Fig. 4(A)). These are the first four genes of the polycistronic late transcript produced after activation of the single late promoter at 8–10 min after induction. The genes are conveniently arranged in order of action in lysis. *S* encodes the lambda holin, S105, which is an integral inner membrane protein with 3 transmembrane domains (TMDs) in an N-out, C-in orientation. The *R* gene encodes the lambda endolysin, a 15 kDa soluble lytic transglycosylase that cleaves the glycosidic bonds in the PG. The *Rz-Rz1* architecture is striking, with the latter occupying 60 codons in the +1 reading frame of the former. Despite sharing the same DNA, the two cistrons encode entirely different proteins with different envelope destinations. *Rz* is a 153 aa integral IM protein with type II (N-in, C-out) topology, whereas after secretion and processing, *Rz1* is a 40 aa lipoprotein attached to the inner leaflet of the OM (Fig. 4(A)).

Operational Outline of Lambda MGL

Extensive molecular, genetic, biochemical and fluorescence microscopy studies have resulted in the following operational narrative concerning the roles of the lambda lysis proteins. Throughout the morphogenesis period, both the *S* and *R* gene products accumulate harmlessly. GFP-fusion and cross-linking experiments indicate that S105 accumulates as freely-mobile homodimers dispersed uniformly in the IM, while the *R* endolysins are synthesized as soluble, fully active monomers in the cytoplasm (Fig. 4(B)). The *Rz* and *Rz1* proteins accumulate in bizarre heterotetrameric (Rz_2-Rz1_2) complexes formed by heterotypic C-terminal interactions (Fig. 4(A) and (B)). The complex is called the “spanin” since it spans the entire periplasm; the *Rz* and *Rz1* proteins are thus the i-spanin and o-spanin (IM and OM) subunits, respectively. Necessarily, these complexes are engaged in the surrounding PG meshwork from the moment of assembly. For almost all of the latent period, there is no detectable effect of any of the lysis proteins, expressed singly or together, on cell growth and, more relevantly, intracellular virion morphogenesis. This changes dramatically when the holins trigger, which occurs when $\sim 5 \times 10^3$ molecules of S105 have accumulated, which, in the induced wt lysogen, is at ~ 50 min; the accumulation of the other lysis proteins is comparable. Operationally, triggering means that the holin population is suddenly redistributed from a uniform, mobile state into micron-scale lesions, a.k.a. the “S-holes”, detectable in cryo-electron tomographs as enormous interruptions in the lipid bilayer (Fig. 5). These large, non-specific holes allow the endolysin molecules to escape to the periplasm and degrade the PG; this is step two. This in turn liberates the spanin complexes from the encaging PG meshwork, leading to catastrophic outer membrane disruption, the third step in lysis (Fig. 4(B)). The entire time from triggering to overt lysis requires only a few seconds.

Looking Under the Hood of Lambda Lysis

Lambda lysis is the prototype MGL, so a mechanistic understanding would be crucial for phage biology in general. The most interesting questions, and the current views, are:

- (1) What is the structure of the micron-scale “S-hole”?

As noted above, the S-holes are seen as interruptions in the IM in cryoEM tomographs. Typically, there are 1–3 large holes per induced cell, with an average diameter of > 300 nm. Holes of the same character have been observed for two other holins from paradigm coliphages: *gpt* of T4 and *Y* of P2. Nothing is known about the detailed structure of the holes. Chemical accessibility studies have shown that, after triggering, essentially all S105 molecules have one face of TMDs 1 and 3 exposed to the aqueous phase, presumably the lumen of the hole (Fig. 6). Moreover, the total number of S105 molecules roughly corresponds to the number needed to line the holes, assuming two TMDs (~ 1 nm) per holin in the wall, so the available data are consistent with the S-hole having a single-molecule wall of holins, with TMD1 and TMD3 facing the lumen. However, dynamic and “swiss-cheese”-like structures that intermittently expose the TMDs to the aqueous interface of the lumen are also possible. *In vivo*, the holes are big enough to allow the passage of ~ 0.5 MDa homotetrameric R- β gal chimeras (i.e., the entire *R* protein fused to the ~ 1000 aa ω domain of β -galactosidase). Purified S105 has been shown to permeabilize liposomes; moreover, mutant lysis phenotypes, including temperature-sensitive function, are faithfully reproduced in these assays. Despite exhaustive mutant hunts, no host factor has been implicated in S105 function. Moreover, S105 retains lethal function in yeast and mammalian cells, supporting the notion that S105 acts alone *in vivo*. Nevertheless, high-resolution fluorescence microscopy aimed at directly visualizing the holes has not yet been reported for S105 or any holin.

- (2) What is the mechanism of triggering and hole-formation and how is it regulated temporally?

This is the most important question because it is fundamental to the evolutionary role of lysis. Other than the lysis-lysogeny decision for temperate phages, the decision of when to terminate the infection cycle and lyse the host is the only decision made in the infection cycle. The key observable is that for lambda, and for all other defined MGL systems, artificial depolarization of the cytoplasmic membrane causes premature lysis. This applies to addition of energy poisons like cyanide or dinitrophenol or even to the sudden (and usually unintentional) anaerobiosis imposed by briefly interrupting the aeration

of an induced culture (i.e., turning off the shaker for a minute). A reduction of as little as 30% of the PMF is sufficient to trigger lambda S105, so the holin system can be regarded as an exquisitely tuned sensor of membrane integrity and energy metabolism. Another fundamental property of the lambda holin is that highly conservative single-missense changes throughout the length of S105 can advance, retard or abrogate triggering; similar genetic fragility has been observed for other holins. These and other observations have led to a model for timing in which, at a critical concentration, the holin population nucleates the formation of two-dimensional aggregates, designated as “death rafts” (Fig. 6). These rafts are thought to be lipid-depleted due to intimate intra- and inter-molecular TMD-TMD helix interactions. It is proposed that the lipid-depleted rafts have much higher conductance to protons, thus causing instantaneous reduction or collapse of the PMF. The loss of the ~ 0.2 V potential, which translates to a $\sim 10^5$ v/cm electric field across the ~ 3 nm IM, forces a conformational change in the holin leading immediately to formation of the holes. Note that all of this occurs in 10–20 s after a latent period of ~ 50 min throughout which the PMF is maintained at normal levels, as determined by experiments using flagella rotation as a non-invasive readout.

In this model, the extreme mutational fragility of S, and other holins is due to effects of the missense changes on the multiple helix-helix interactions involved in the pathway from mobile homodimer to raft. The notion of a critical concentration for aggregation is derived from studies of the formation of the purple membrane in halobacteria. After induction, the bacteriorhodopsin monomer initially accumulates as a freely mobile monomer before, at a critical concentration, suddenly nucleating the two-dimensional light harvesting array. Similarly, conservative changes at intermolecular contact points in the some of the seven TMDs elicit unpredictable changes in the timing of purple membrane nucleation.

(3) Why is the spanin necessary and how does it disrupt the OM?

Although the holin-endolysin pathway has been recognized for decades, the requirement for OM disruption was only recently discovered. The key result was that inductions of a lambda lysogen carrying nonsense mutations in both the *Rz* and *Rz1* genes terminated not in lysis but instead in a spherical cell morphology. Cryo-EM images show that the PG layer was missing, so the loss of rod shape reflected the capacity of the OM to withstand the osmotic pressure of the cell. Structural predictions indicate that *Rz* has two coiled-coil alpha helices in its periplasmic domain, whereas *Rz1* largely unstructured due to its high proline content (Fig. 4(A)). Biochemical experiments support these predictions. Both *Rz* and *Rz1* form homodimers that are covalently linked by homotypic intermolecular disulfide bonds, based on Cys residues at *Rz* positions 99 and 152 and *Rz1* position 29 (Fig. 4(A)). Although the Cys at position 99 is dispensable, the lytic function of the complex requires at least one disulfide bond in either of the other two positions.

Based on the similarities of spanins to membrane fusion systems (coiled-coil structure and membrane-spanning complex) it was hypothesized that spanins disrupt the OM by causing the IM and OM to fuse during lysis (Figs. 1(C–D) and 4(B)). Genetic analyses lent support to this notion, implicating two coiled-coil helix domains in *Rz* and the proline-rich region in *Rz1* as key elements in the lytic function. Proline-rich stretches have been shown to be essential for the reovirus FAST fusion proteins and coiled-coil domains are a conserved feature across many fusion systems. Forward mutation and suppressor analysis suggested that most mutants blocked the ordered conversion from the extended form of the complex to the hairpin conformation.

A Second Strategy: Pinholins and SAR Endolysins

A second general type of MGL is represented by the lysis system of the lambdoid coliphage $\phi 21$. Although the $\phi 21$ lysis cassette is arranged syntentically to lambda, which resulted in its genes being assigned homologous names, there are fundamental differences in these two MGL systems. First, the *S*²¹ gene encodes a 68 residue “pinholin” (*S*²¹68), with two TMDs in N-in, C-in topology. Biochemical and microscopic studies indicate that this pathway is operationally similar to the lambda pathway, in that the pinholins accumulate as freely-mobile homodimers in the IM. However, after triggering, the pinholins form many more ($\sim 10^3$) and much smaller holes (~ 2 nm) “pinholes”. Moreover, pinholin function requires unprecedented topological dynamics, in that TMD1 must exit the bilayer before triggering (Fig. 4(C)). Indeed, TMD1 is dispensable and thus can be considered an intramolecular negative-regulatory domain.

Clearly, the pinholes are too small to allow the release of folded endolysins, so the *R*²¹ endolysin has a special secretory properties conferred by its N-terminal SAR domain. This signal engages the host *sec* translocon like a typical N-terminal TMD, so that *R*²¹ accumulates in the periplasm tethered to the IM (Fig. 4(C), left). Importantly, this form is enzymatically inactive, thus avoiding premature lysis. Equally important is that the SAR domain is metastable in the bilayer, so that it escapes the membrane

Fig. 4 (A) Left: Cartoon of lambda spanins *Rz* and *Rz1* shown within the cell envelope. The homotypic intermolecular disulfide linkages are shown. Right: The lambda lysis cassette is shown. Expression of these genes is under control of the late promoter *pR'*. The *Rz1* gene is embedded in the +1 reading frame of *Rz* and the following functional domains for spanin gene products have been identified by genetic analysis: NTMD = N-terminal transmembrane domain, JM = juxtamembrane region, CC1 = alpha helix with predicted coiled-coil domains, CC2 = distal coiled-coil domain, PRR = proline-rich region. (B) Model for MGL systems encoding a soluble endolysin: The holin forms holes in the inner membrane that allow the endolysin to access the peptidoglycan. After the endolysin degrades the peptidoglycan, the spanins activate and undergo a conformational change that causes fusion of the IM and OM. (C) Model for MGL systems encoding a pinholin-SAR endolysin: During triggering the pinholin TMD1 exits the bilayer and TMD2 oligomerizes into ~ 2 nm pinholes in the IM. The formation of ~ 1000 pinholes causes depolarization of the membrane, resulting in the release and activation of the SAR endolysin from the bilayer. Reprinted from Cahill, J., Young, R., 2019. Phage lysis: Multiple genes for multiple barriers. *Advances in Virus Research* 103, 33–70.

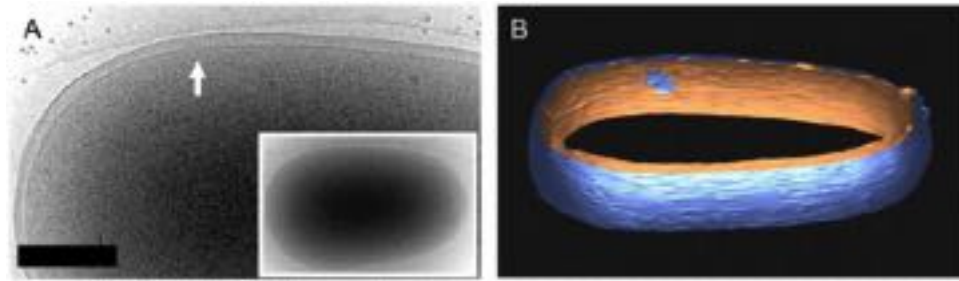


Fig. 5 The S105 holin lesion visualized by CryoEM tomography. (A) A culture expressing *S105* was concentrated and frozen at the time of holin triggering. The arrow indicates the lesion in the inner membrane. (B) 3D reconstruction: The outer membrane is shown in blue and the inner membrane is shown in orange. Scale bar is 250 nm. Reprinted from Dewey, J.S., Savva, C.G., White, R.L., *et al.*, 2010. Micron-scale holes terminate the phage infection cycle. *Proceedings of the National Academy of Sciences* 107, 2219–2223.

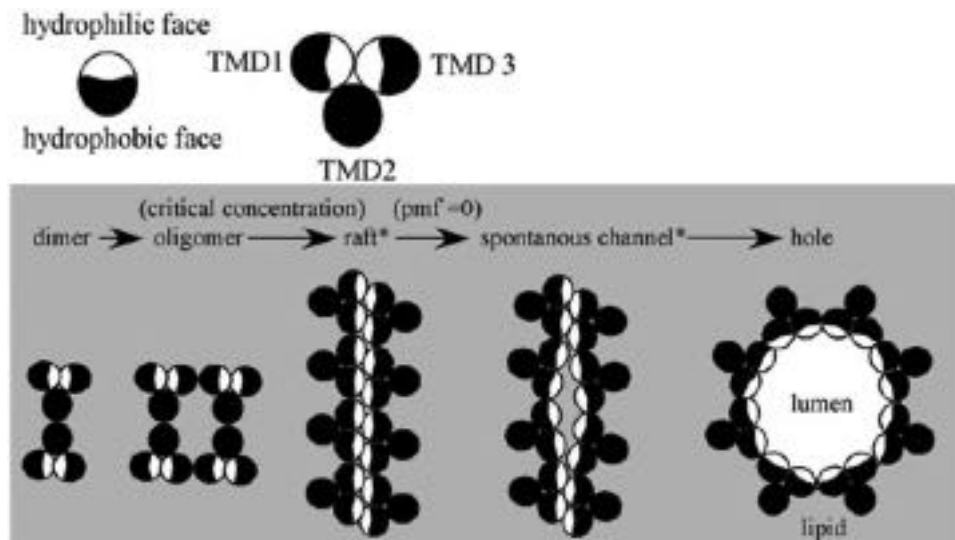


Fig. 6 Model of lambda holin function. Top-down view of the S105 holin, depicting a hydrophobic face (black) and a hydrophilic face (white) for TMDs 1 and 3. The holin accumulates as a dimer with the hydrophilic faces sequestered intramolecularly. As the holin accumulates in the inner membrane there is an increase in the oligomeric state (dimers to oligomers). At a critical concentration the holin molecules nucleate into rafts with the hydrophilic faces oriented toward each other. The rafts are thought to be large enough to disrupt the PMF of the cell. When the cell depolarizes the holin transitions to hole-forming structure, with lipid outside (gray) and the hydrophilic faces of TMDs 1 and 3 facing an aqueous lumen (white). Although rafts have been visualized using holin-GFP chimeras, the arrangement of holin molecules in rafts is inferred (asterisks) from the final hole-forming arrangement. The cartoons of the rafts and the hole are not drawn to scale. Nothing is known about the two-dimensional organization of the rafts, so the linear arrangement is shown purely to indicate and intramolecular sequestration of the hydrophilic surfaces of TMDs 1 and 3. Reprinted from Cahill, J., Young, R., 2019. Phage lysis: Multiple genes for multiple barriers. *Advances in Virus Research* 103, 33–70.

when the pinholin triggers and, as a result of the massive formation of pinholes, the PMF is completely collapsed. Release from the bilayer causes refolding of the SAR endolysin to an active form, and the PG is subsequently degraded (**Figs. 2** and **4(C)**).

A Second Type of Spanin

Coliphage T1 has a special place in the history of phage biology, being the first member of the “Seven Dwarves” (phages T1–T7) designated by the Delbrück phage group in the 1940s as the sole permissible experimental systems. I-spanin and o-spanin genes were easily identified in T2 through T7 by virtue of the distinct secretory signals (an N-terminal TMD for the former and an OM SPII signal sequence for the latter) and the fact that the genes were always closely linked (either embedded as in lambda, overlapped as in P2, or adjacent as in T4). In the T1 lysis cassette, the cistron in the position normally occupied by the i-spanin/o-spanin genes encoded a protein, gp11, with unprecedented dual targeting signals: an N-terminal OM lipoprotein signal and a C-terminal TMD that would span the IM (**Fig. 7**). Genetics and biochemical approaches confirmed the dual targeting, indicating that gp11, once synthesized and processed by the LOL OM lipoprotein export pathway (**Fig. 3(B)**), connects the OM and IM as a single polypeptide. Since gp11 fully complemented the lysis defect of an induced lambda *Rz_{am}Rz1_{am}* lysogen, it has been

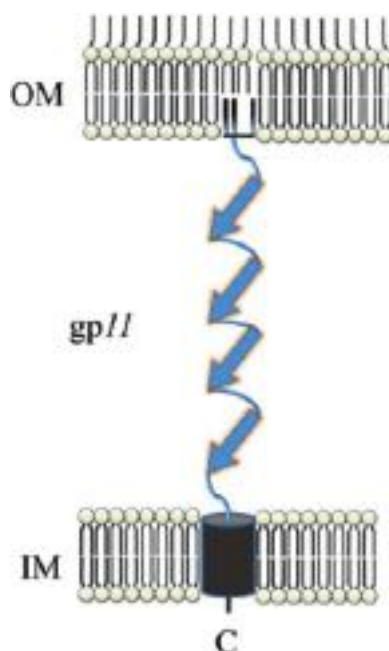


Fig. 7 Cartoon of the T7 u-spanin gp11, attached to the OM with an N-terminal lipid anchor and embedded in the IM by a helical C-terminal TMD.

designated as a unimolecular spanin (*u*-spanin), in contrast to the *i*-spanin and *o*-spanin subunits of the two-component spanin. Unlike the *i*-spanin, which is rich in coiled-coil helical structure, the *u*-spanin periplasmic domain is dominated by stretches of predicted beta-strand conformation. The length of the periplasmic domain, if fully extended, closely matches the observed width of the periplasm.

Nothing is known about the molecular mechanism by which the *u*-spanin disrupts the OM. However, both the two-component spanin and the *u*-spanin appear to be negatively regulated by the intact PG. This is based on experiments in which the spanin complex was capable of causing lysis in an endolysin-independent fashion if the PG layer was sufficiently compromised by starvation for the precursors of the cross-linking polypeptides. This common feature suggests that the *u*-spanins also disrupt the OM by fusing it with the IM; in this regard, gp11 resembles class-II viral fusion proteins, primarily due to the dominance of β -structure in its periplasmic domain. Presumably, the step of bringing the IM and OM together is facilitated by collapse of the periplasmic domain from its original extended form into beta-sheet structures. The C-terminal TMD plays no role specific other than anchoring in the IM, because an artificial TMD can replace the original TMD without affecting lytic function.

Regulation of MGL

Dual start holin genes: Holin-antiholin pairs

For both canonical (large-hole) holins and pinholin systems, holin accumulation controls the timing of lysis. However, in all the phage systems that have received sufficient study to allow a thorough analysis of gene expression, the MGL genes are co-transcribed with all the morphogenesis genes, ruling out transcriptional regulation. The best studied example of lysis regulation is in lambda, where the *S* gene encodes a second protein, S107, by virtue of an upstream translational start. Although S107 and S105 differ only by the two extra N-terminal residues in the former, S107 acts as a negative regulator, or “antiholin” of S105 (**Fig. 8(A)**). The root of this antagonistic behavior is that the extra positive-charged Lys2 residue of S107 prevents TMD1 from entering the bilayer, rendering S107 into a non-functional topology with only TMD2 and TMD3 in the bilayer. S107 preferentially heterodimerizes with S105, and, since the normal ratio of S105:S107 is ~ 2 , only 1/3 of the freely diffusing dimers are S105:S105 homodimers constituting the mass action pool that leads to triggering. The ratio of S105 to S107 is controlled by an RNA stem-loop structure that overlaps the ribosome binding sites for Met1 and Met3. The potential effect of S107 regulation is mainly negative; eliminating the S107 gene start has only a minor effect on lysis timing, advancing lysis by ~ 5 min, but increasing S107 to or beyond the level of S105 blocks spontaneous lysis entirely. There is some evidence for host factor control of the stem-loop formation and thus of the S105–S107 ratio, but the molecular signal, if any, that might increase the proportion of antiholin starts has not been identified. In any case, when the S105 population reaches its critical concentration and the PMF is collapsed, the S107 molecules are converted to the N-out, C-in topology. At least in principle, this instantly triples the pool of functional holin molecules available for hole formation, contributing to the saltatory, or “all or nothing”, character of lysis regulation. However, how much these the converted S105–S107 heterodimers actually contribute to hole structure has not been determined.

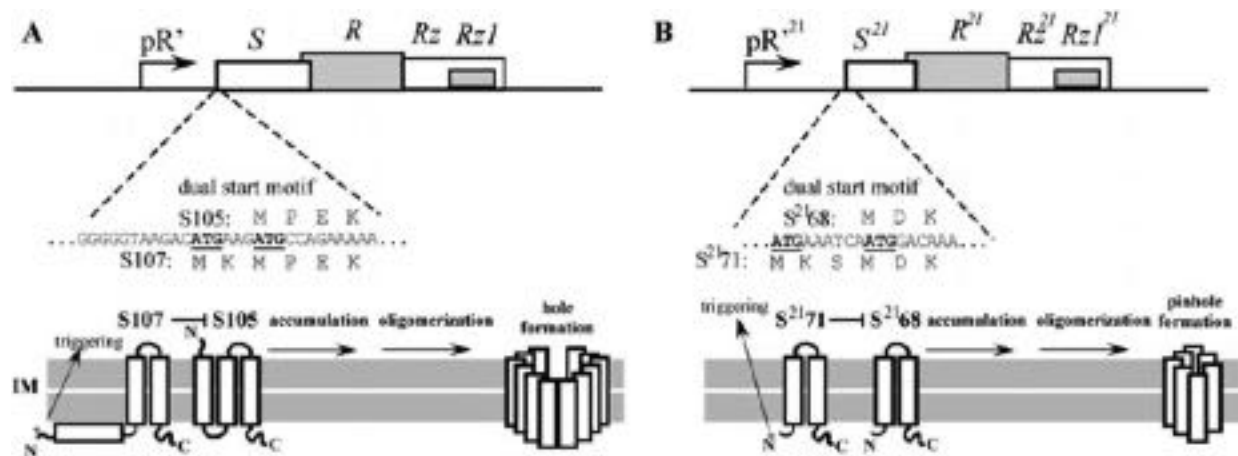


Fig. 8 Comparison of the dual start holin/pinholin motifs and a model for the inhibitory role of the antiholin for lambda (left) and phage 21 (right). The start codons are underlined for *S105*, *S107*, *S²¹71* and *S²¹68* and the first residues of the translational sequence for the holin/pinholin (*S105* and *S²¹68*) or antiholin/antipinholin (*S107* and *S²¹71*) are shown aligned to their respective codon. For both antiholins, the positive charge on the N-terminal extension blocks a topological change for TMD1. Therefore, *S107* is unable to enter the bilayer and *S²¹71* is unable to exit the bilayer. In this topology *S107* and *S²¹71* are inhibitory to their cognate holin. Upon triggering, the topological blockage is removed and the antiholins functionally become holins, joining the triggering pathway toward hole formation. Reprinted from Cahill, J., Young, R., 2019. Phage lysis: Multiple genes for multiple barriers. *Advances in Virus Research* 103, 33–70.

The prototype pinholin gene, *S²¹*, has a similar “dual start” motif, with opposite topological consequences. The antiholin product, *S²¹71*, is produced from translational initiation three codons upstream of the pinholin start and has an extra positive-charge residue that prevents TMD1 from exiting the bilayer. As noted above, TMD1 of the pinholin is a built-in inhibitory domain that must exit the bilayer before TMD2 can proceed to hole formation (Fig. 8(B)). The presence of the extra basic residue at the cytoplasmic end of TMD1 blocks its release into the periplasm and thus inactivates the heterodimer in terms of pinhole formation. Again, the PMF collapse deriving from triggering by the uninhibited pinholins should, in principle, allow TMD1 of *S²¹71* to escape from the membrane and thus activate the inhibited heterodimers. Based on the rigorous characterization of the lambda and ϕ 21 systems, other dual-start holins have been noted and some of them confirmed, usually by cloning into inducible expression vectors in *E. coli*. However, neither of the well-characterized dual start systems can be considered real-time regulation, since the ratio of the holin and pinholin appears to be set and no physiological condition for altering it is known.

Real-Time Regulation of Holin Function: Phage T4 Lysis Inhibition

The T4 holin, endolysin, i-spanin and o-spanin are encoded by genes *t*, *e*, *pseT.3*, and *pseT.2*, respectively. When not under LIN regulation (discussed below), the lysis pathway is analogous to lambda. The T holin forms holes at the time of triggering, which releases soluble endolysin E into the periplasm where it degrades the PG. Like lambda, the last step of lysis is the disruption of the OM by the two-component spanin complex assembled from the i-spanin PseT.3 and o-spanin PseT.2. However, the topology and primary structure of the T4 holin is grossly dissimilar to that of lambda. Unlike other holins, *gpt* has a single TMD, a large C-terminal periplasmic domain (163 aa) and a substantial (32 aa) N-terminal cytoplasmic domain (Figs. 9 and 10). Nevertheless, despite the disparity in holin primary structure, cryoEM studies indicate that there is little difference in the size and number of T-holes compared to lambda S-holes. The T4 system has an additional level of regulation, reflected in the phenomenon of “lysis inhibition” (LIN), first documented in the 1940s by the Delbrück group. If a T4-infected cell is super-infected at 5 min or more after the initial infection, lysis is blocked, at least temporarily. Continued superinfection can extend the LIN state for hours, leading to hyper-accumulation of progeny virions intracellularly, before spontaneously collapsing. The LIN state can be prematurely collapsed by addition of uncouplers or energy poisons, indicating that, as expected, LIN reflects inhibition at the level of holin triggering. Accordingly, there are many missense mutants of *gpt* that do not support LIN and instead trigger lysis at ~25 min after infection, irrespective of super-infection. The requirement for at least 5 min post-infection before superinfection stems from the need to synthesize four early T4 proteins: RI, RIII, RIV (also Spackle or Sp), and Imm. It is thought that RIV and Imm both function to block the tail tube of the super-infecting T4 from penetrating the IM and ejecting the capsid contents into the cytoplasm. RIV inhibits the tail lysozyme responsible for penetrating the PG layer; Imm is a membrane protein that may interfere with the temporary permeabilization of the IM. As a result, the contents of the capsid head are ejected ectopically into the periplasm. The contents include the 170 kb gDNA and > 1000 protein molecules. It is thought that something in this mélange is the signal for LIN (Fig. 9), since superinfection with T4 “ghosts”, in which the T4 capsid contents have been released osmotically, does not cause LIN.

Recent work addressing the molecular basis for LIN have demonstrated that RI and RIII are antiholins and effect LIN by binding to the *gpt* holin and inhibiting the triggering pathway. The key LIN component is RI, which has a SAR domain and a 75 aa C-terminal

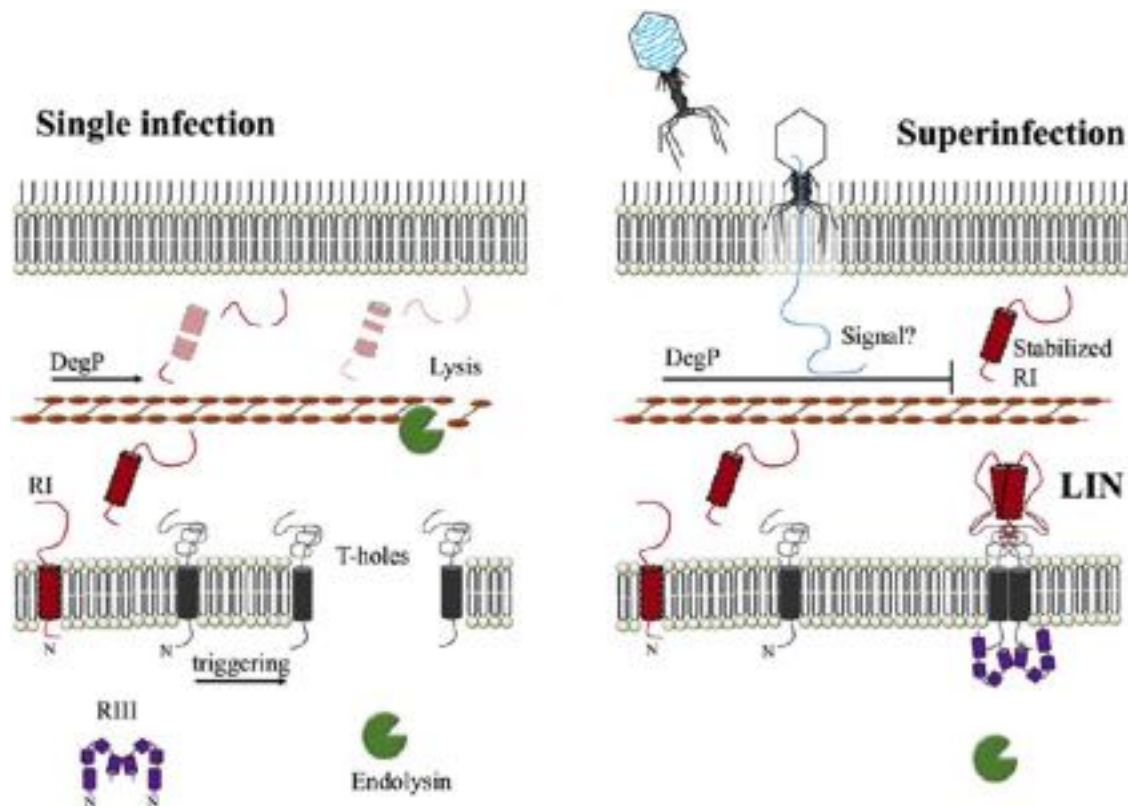


Fig. 9 Model for T4 lysis and lysis inhibition. Left: During a single infection, RI is degraded by DegP. The T holin triggers at the normal ~25 min time (after infection), which results in the release of endolysin into the periplasm to degrade the peptidoglycan. The subsequent spanin-mediated outer membrane disruption step follows (not shown). Right: Superinfection at >5 min after infection results in ectopic localization of the capsid contents into the periplasm, in the inhibition of DegP-mediated degradation of RI. Stabilized RI dimerizes and binds T dimers, blocking it from further oligomerization. In this state, a symmetric binding site is provided for dimeric RIII binding. Formation of this heterohexameric complex stably maintains the LIN state.

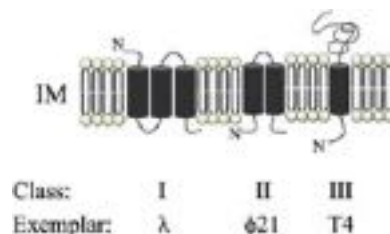


Fig. 10 The founding members of each class of holins are shown along with their membrane topology.

periplasmic domain (CTD). After synthesis, RI transits to the periplasm where it is rapidly degraded if expressed alone or during single infections, whereas superinfection from a second T4 particle stabilizes RI. The rapid inactivation of RI is caused by the presence of its SAR domain. Replacement of the RI SAR domain with a normal secretory signal sequence resulted in the secretion of a proteolytically stable truncated form of RI (sRI) to the periplasm; this species can impose LIN on *gpt* irrespective of superinfection. LIN requires the formation of equimolar complexes between the periplasmic domains of RI and *gpt*; these complexes block progression of *gpt* into the triggering pathway. The RIII protein is a 95 aa soluble cytoplasmic species that binds to the N-terminal cytoplasmic domain of *gpt*. RIII alone has only a minor effect on *gpt*-mediated lysis but expression of RI and RIII together can establish a stable LIN state without the need for superinfection. The simplest model incorporating these observations is that RI and RIII act as periplasmic and cytoplasmic antiholins, respectively, with RIII acting to stabilize the RI-*gpt* complex (Fig. 9).

T4 lysis and the LIN response constitute the oldest genetically-defined phenomena in phage biology and arguably in modern molecular genetics. With the progress achieved in identifying the molecular components and their interaction domains, the remaining mystery occluding a clear picture of this elegant regulatory scheme is the nature of the LIN signal generated by the super-infecting phage. In addition, the *gpt* holin is itself an enigma. It is unclear how a large globular periplasmic domain, roughly

5–6 nm in diameter, can be arranged in two-dimensions to allow the single TMD (~1 nm diameter) to oligomerize and make the micron-scale holes observed by cryo-EM. One possibility is that the N-terminal domain, which at 32 residues is much larger than the cytoplasmic domains of other holins, may rearrange to form a second TMD that could be assembled into the wall of the hole.

MGL Diversity

There are thus two distinct molecular strategies for each step in phage-mediated lysis of Gram-negative hosts: holins and pinholins targeting the IM, soluble endolysins and SAR endolysins targeting the PG, and the two types of spanins, two-component (i-spanin/o-spanin) and unimolecular (u-spanin). With the exception that pinholins require SAR endolysins, there are six possible combinations, all of which have either been encountered in nature or constructed in the laboratory and found to be functional, at least in terms of achieving lysis with kinetics that would be useful for a phage. The term “canonical” has been applied to the large-hole forming holins and soluble endolysins to distinguish them from pinholins and SAR endolysins. Each functional class of lysis proteins has its own profile in diversity.

Holin Diversity

The term holin was coined to denote proteins that form lethal and non-specific holes in the IM. As established in coliphage systems, the holin has two roles, serving to release the endolysin into the periplasm as well as to function as the “lysis clock”. The clock function refers to the fact that holins control the initiation of lysis and thereby govern the length of the vegetative phase. Holin homologs have been found in all three domains of life and there are more than 60,000 unique entries for holins in the NIH identical protein database. In recent years, there have been efforts to categorize this very diverse group of proteins as well as understand the functional role of holin homologs outside of phage lysis.

Holins are classified based on the number of TMDs: Class I, II, and III holins have three, two, and one TMD(s), respectively (Fig. 10), and there is bioinformatic evidence for holins with four TMDs. The immense diversity makes them difficult to identify through homology. Holin gene calling is limited to searching for a small integral IM protein with up to four TMDs located near an endolysin. In 2013, Reddy and Saier identified 52 families (there are now 67) of established or putative holins, which can be viewed on the TCDB (transporter classification database). Using statistical analyses, 21 of the 52 families were found to be made up of seven superfamilies; i.e., the members are homologous but distantly related.

The number of different holin families with distinct topologies suggests that the diversity is a product of convergent evolution. The most striking demonstration of how holins that are found in one domain could also function in another was done with Bax and Bak, which are involved in the apoptotic response in mitochondria. Bax/Bak are encoded in the nucleus and cause permeabilization of the mitochondrial OM. Interestingly, these proteins were shown to function as holins in *E. coli*. The authors showed that active Bax or Bak formed large enough holes to release lambda endolysin into the periplasm, causing lysis. It should be noted however, that these proteins do not replicate the clock function of “real” holins, and the size of the lesions produced is unknown.

Some phages of Gram-positive and Mycobacterial hosts require two separate genes for holin function, the first report of which was for the defective prophage PBSX of *Bacillus subtilis*. In this system, both *xhlB* (predicted class II holin) and *xhlA* (class III but with large N-terminal domain) are required for endolysin release. In another example, *Bacillus* phage SPP1 holin activity requires genes 24.1 and 26, encoding proteins with class I and class II topologies, respectively. In an unusual twist on lysis gene regulation, it was reported that gp24.1 is constitutively expressed with the endolysin, thus allowing gp26, which is a late transcription product, to serve as the clock.

The mycobacteriophage Ms6 lysis system has been reported to use two holins, gp4 and gp5 (discussed below). Taken together, there is increasing evidence for co-holin systems. It would be interesting to address why these systems evolved to require two-components; e.g., perhaps one holin component functions to alter the timing of lysis in response to cellular or environmental signals.

Endolysin Diversity

Endolysins have four different mechanisms of targeting linkages essential to the integrity of the cell wall. Amidases and endopeptidases break oligopeptide crosslinks, whereas glycosylases and transglycosylases target glycosidic linkages. The lambda endolysin is transglycosylase that cleaves the β -1,4 glycosidic linkage between MurNAc and GlcNAc, resulting in the formation of a cyclic 1,6-anhydromuramic acid cleavage product. In contrast, the endolysin of T4 and phage 22 are “true lysozymes,” which hydrolyze the bond between MurNAc and GlcNAc using a glycosidase catalytic triad (E-xxxxx-C/D-xxx-T, or a variant). The hydrolysis reaction releases a disaccharide product with reducing ends. The T7 endolysin is a product of gene 3.5 and the best-characterized amidase. Interestingly, gp3.5 has been shown to bind and inhibit T7 RNA polymerase. Despite the variety of mechanisms, the difference in enzymatic activities is not critical to the overall lysis reaction; complementation studies have demonstrated that endolysins will work with any holin. Moreover, the endolysin is usually produced in great excess over what is needed for lysis; thus some amber nonsense alleles of lambda *R* do not block lysis, due to <5% read-through of the UAG stop codon.

Recently, a comprehensive analysis was done on endolysins from 890 phage genomes that were isolated from 64 different bacterial genera. Endolysins were shown to be very diverse, perhaps owing to the large variety of PG chemotypes. The 723 putative endolysins ranged from 72 to 578 residues in length, using 24 different catalytic domains and 18 different cell wall-binding domains organized with 89 different architectures. From this list, there were 103 endolysins identified with a SAR domain. Interestingly, except for four endolysins from phages of lactococcal hosts, endolysins with predicted SAR domains were exclusively from phages infecting Gram-negative bacteria (especially *Enterobacteriaceae*). Moreover, there are significant differences in canonical endolysins from Gram-negative versus Gram-positive bacteria. In general, endolysins from phages of Gram-negative hosts are smaller with a single domain responsible for both binding and catalysis of the PG. In contrast, phages infecting Gram-positive bacteria produce endolysins with the cell-wall-binding domain separated from the catalytic domain. It has been suggested that the separated domain architecture functions to anchor the endolysin to the PG and prevent collateral damage during lysis. In other words, lysing cells might release endolysin which would destroy the exposed peptidoglycan of neighboring cells, which are potential hosts for the next infection cycle.

Spanin Diversity

Recently, a comprehensive bioinformatic study was done to examine the diversity and conservation of spanins for phages with Gram-negative hosts. Most phages (528 out of 677 genomes) employ two-component spanins, i.e., encoding an i-spanin and an o-spanin. Within this set, 34% were arranged like lambda, with the o-spanin gene totally embedded within the i-spanin. The most common arrangement was overlapped (like phage P2), accounting for 43% of the genomes and the remainder of the genomes had separated architecture (like T4). Spanins were then grouped into families so that members shared >40% identity over >40% of the length of their sequence. Surprisingly, this revealed >100 families and there were 99 and 65 singletons for i-spanin and o-spanins, respectively. The trend continued for the 58 genomes with u-spanins, with 13 families and 6 singletons. The large number of families, especially singletons, shows that spanins are very diverse proteins.

An interesting question that stemmed from the study above is why no spanin genes could be identified in 13% of the genomes. One problem is that annotation of the spanins depends on the identification of the signal peptidase II (SPII) processing site, or lipobox, in either the o-spanin or u-spanin sequence. The algorithm for assignment of SPII sites is problematic for protein sequences of less than 70 residues, and utterly dependent on accurate determination of the translational start. We have found the latter is problematical in many phage genomes in the database, presumably because of the extensive gene crowding and codon overlap. However, some phages, like the T7-like podophages, have genome organizations that are relatively easy to annotate and where the existence of any lipobox-containing sequences can be ruled out. For example, the T7-like phage Petty of *Acinetobacter baumannii* was shown to cause the rapid and explosive lysis characteristic of MGL. The lysis morphology is significant because it indicates that the lysis pathway of Petty does not stall at the OM disruption step. Petty encodes a holin and an endolysin but careful analysis of all possible reading frames revealed that no lipobox-containing reading frame exists. This supports the idea that Petty and presumably other spanin-less phages disrupt the OM by a yet-to-be discovered OM disrupting protein.

MGL in Gram-Positive Hosts

Of all the various lysis proteins, the one class that has attracted the vast majority of enzymological and biotechnological interest in lysis has been focused on the endolysins of phages with Gram-positive hosts. This stems from the fact that the endolysins of these phages can be applied from the outside and cause efficient bacteriolysis, as documented in early work using lysates generated by *Streptococcus* phage C1 and later with purified C1 endolysin. The use of such lysins as anti-bacterials has since become a vibrant field of its own. Nevertheless, how endolysin activity is coordinated with the infection cycle has not been studied extensively in Gram-positive systems. In the relatively few cases where the functional identification of holins in phages of Gram-positive hosts has been accomplished, it was usually done either by complementation of the lysis defect of lambda S mutants (see above), or by cloning the putative holin and endolysin genes in an inducible expression vector in *E. coli*. In principle, MGL in most Gram-positive systems should be simple, with only the holin and an endolysin needed. For example, the *B. subtilis* phage phi29 has a small genome and has been subjected to extensive genetic and molecular analysis, revealing gene 15 to be the endolysin gene. The adjacent gene 14 was predicted to be an integral membrane protein with 3 TMDs and a potential dual start, which would yield a 131 aa antiholin and 129 aa class I holin; these predictions were unambiguously confirmed by inducing clones of this simple lysis cassette in *E. coli*. However, many Gram-positive species are capable of autolytic physiological responses, often as part of sporulation responses, natural transformation pathways and stationary phase adaptation. Accordingly, holin-endolysin gene cassettes have been identified in all pneumococcal phages and some have been shown to be functional in *E. coli*, but there are indications that the lessons may not be so simple. For example, it has been reported that in pneumococcus phage SV1, the endolysin does not require the holin for escape from the cytoplasm and localization to the PG, despite lacking any secretion signal. Instead the holin is required to activate the endolysin by collapsing the membrane potential. In addition, efficient lysis in infections by SV1 requires also the participation of the powerful host autolysin LytA, which is also activated by the depolarization of the membrane. The holin-independence of the endolysin may reflect the use of the cryptic secretion system that enables cytoplasmic enzymes to partition to the cell surface and "moonlight" as adhesins, as has been documented in *Streptococcus* for the glycolytic enzyme GAPDH and other proteins. However, the subcellular distribution of LytA during different growth phases has been quantified,

leading to the conclusion that its partial (~5%) extracellular localization derives from “fratricidal” lysis events in a fraction of the cells during exponential phase, rather than a cryptic secretion pathway. In this model, lysis is prevented at the substrate level, in that the healthy PG biosynthesis machinery blocks access of the endolysin to sensitive sites. Entrance into stationary phase or other stressful conditions lead to incapacitation of the machinery, exposing the PG at critical sites, especially the developing septum. This leads to a lysis cascade, with the system primed by the low level of externalized LytA in exponential phase. This rationale provides a common basis for the apparent activation of LytA, and by extension, to the SV1 endolysin, by antibiotics that interfere with PG biosynthesis and depolarization by holin triggering.

There are other reports of holin-independent localization of endolysin in phages of Gram-positive hosts, including what appears to be a SAR endolysin in the *Lactobacillus* phage phiPYB5. The earliest, pre-dating the discovery of SAR endolysins, was for the *Oenococcus* phage fOg44, where the endolysin was shown to be secreted by a standard SPI signal sequence. Nevertheless, the secreted mature protein did not immediately cause lysis but, instead, became muralytically active when the holin triggered to permeabilize the membrane. It should be noted that the lysis cascade scheme mentioned above might apply here too; i.e., the depolarization may interrupt normal PG synthesis, exposing otherwise masked substrate structures to the secreted fOg44 endolysin. Indeed, this perspective provides a molecular rationale for the proposition, promulgated recently by Fernandes and Sao-Jose, that, in infections of Gram-positive cells, the central role of holins is the depolarization of the cytoplasmic membrane rather than the release of endolysin.

Lysis in Mycobacteria

The members of the order *Corynebacteriales* have a three-layer envelope, like the Gram-negatives, but instead of the asymmetric phospholipid-LPS OM, have an outer surface layer of mycolic acid, the “mycomembrane” (OMM), connected to the PG by arabinogalactan. Thus phages with hosts of this order should require something operationally analogous to the spanins for disruption of the OMM. To this end, most mycophages encode a protein called LysB, after the prototype first studied in phage Ms6. LysB has esterase activity that cleaves the linkages between the mycolic acids and the arabinogalactan. Genetic studies in *M. smegmatis* phages showed that defects in LysB are associated with substantial losses of progeny virions due to entrapment in the host debris, although deletion of the *lysB* gene does not abrogate plaque-formation. Presumably LysB is secreted through the holin lesion, although this has not been explicitly demonstrated. Moreover, studies of LysB localization and coordination with respect to the endolysin have not yet been reported.

One might view the finding that mycophage use a lysis protein to disrupt the OMM as an indication that MGL in the Mycolata closely parallels MGL in Gram-negative hosts. However, the same Ms6 lysis system, which has been subjected to rigorous genetic and molecular analysis, has unique features that dispel any such notion. The first surprises concern the LysA endolysin gene, which has two translational starts 143 codons apart, thus encoding full-length LysA384 and a shorter product, LysA241. Both forms have amidase activity and PG binding motifs and both are required for efficient lysis and release of progeny. The gene immediately upstream of *lysA*, gene 1, encodes a 78 aa acidic polypeptide that forms oligomeric complexes with the endolysin and acts as a chaperone for its *secA*-dependent externalization in Ms6 infections of *M. smegmatis* and in inductions in *E. coli*. There is nothing resembling a signal sequence in LysA, so this is just as mysterious from the perspective of the well-characterized function of the Sec translocon as it is from our general understanding of how lysis proteins are localized. In addition, the shorter *lysA* product lacks the N-terminal domain required for gp1 binding, so there is no model for how it transits the membrane. Moreover, the apparent lysis cassette terminates in two genes encoding small membrane proteins, gp4 and gp5. The former has predicted topology that fits a standard class II (2 TMDs, N-in/C-in) holin, whereas the latter has a single TMD with unambiguous N-out, C-in topology defined by extraordinarily hydrophilic N- and C-terminal domains. Experiments with clones in *E. coli* indicated the two proteins oligomerize and function together in lysis, but single-step growth experiments with Ms6 mutants suggested that gp5 has antiholin character. Taken together, these rather heretical conclusions indicate that MGL systems in the mycolata and other Gram-positive hosts may have many new working parts and regulatory features that deserve attention.

Single Gene Lysis (SGL) Systems

Although the dsDNA phages dominate the virosphere, there are two classes of small lytic phages with single-stranded genomic nucleic acids that are widespread in nature: the single-strand circular DNA *Microviridae* and the ssRNA *Leviviridae*. Despite the vast biological and evolutionary differences, they all achieve lysis without deployment of a muralytic enzyme. Instead each of these phages encodes a single lysis protein that causes the host to undergo autolysis. Here we will refer to each lysis protein as an Sgl, but it must be noted that the SGL proteins are in general unrelated to each other except in sharing the same operational strategy.

ϕ X174 E: The First “Protein Antibiotic” Sgl

The ϕ X174 microvirus is a phage of many “firsts.” It was the first DNA genome to be sequenced and the first genome to be synthesized *in vitro*, and the first to have its Sgl to be characterized. The life cycle is simple; after penetration of the 5 kb circular, single-stranded gDNA into the cytoplasm, all 10 genes are transcribed constitutively from multiple promoters. The Sgl of the prototype microvirus ϕ X174 is protein E, encoded in a reading frame embedded in the +1 register in gene D (Fig. 11), which

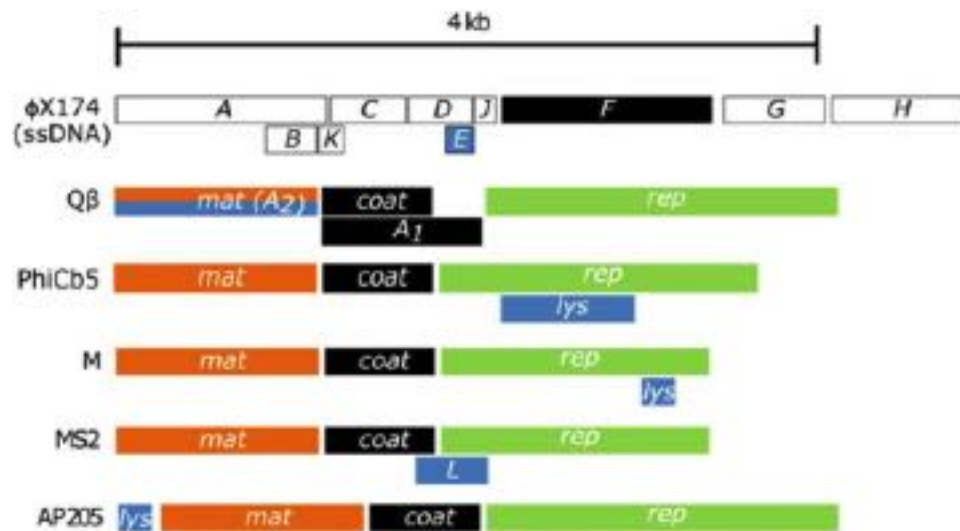


Fig. 11 Sgl phage gene maps. All the known architectures of SGL phages are shown, including the ssDNA coliphage ϕ X174 and five *Leviviridae*: Qb and MS2, which are specific for the F conjugation pilus; PhiCb5, specific for the *C. crescentus* TAD polar pilus; M, specific for the IncM conjugation pilus; and AP205, specific for the type IV pilus of *Acinetobacter*. The *sgl* genes are colored blue. The Qb Sgl protein is also the Mat protein, designated A_2 historically. The A_1 protein is a minor capsid protein generated by read-through of the coat stop codon. Capsid proteins are colored black. All L-like *sgl* genes are situated overlapping the *coat-rep* junction, as in MS2, except for AP205. The PA-type *sgl* genes (ϕ X174, Qb, and M) have diverse locations. The lytic mode of phiCb5 is unknown but thought to be PA-type. Originally published in Chamakura, K., and Young, R., 2019. Phage single-gene lysis: Finding the weak spot in the bacterial cell wall. *Journal of Biological Chemistry* 294, 3350–3358.

encodes the scaffolding protein essential for capsid morphogenesis. The *E* product is a 91-residue type-I integral membrane protein in which the N-terminal 34 residues (E_{NTD}) are conserved in other *Microviridae* and essential for lytic function, whereas the highly basic (+11 net charge), proline-rich C-terminal domain (E_{CTD}) can be replaced by heterologous domains like β -galactosidase or GFP without impairing lytic function. *E* exerts its lytic function by acting as a non-competitive inhibitor of *MraY*, a conserved membrane protein which catalyzes the formation of Lipid I, the first lipid-linked intermediate in the pathway for biosynthesis of PG precursors (Fig. 2). *MraY* has 11 TMDs, and resistance to *E* can arise by mutations in TMDs 5 and 9 that reduce binding affinity. Conservative mutations in hydrophobic residues along one face of the TMD in *E* and in residues near its presumed periplasmic interface can affect either *E* binding or the capacity for non-competitive inhibition. These data are most consistent with a model in which one face of the single TMD of *E* interacts directly with TMDs 5 and 9 of *MraY*.

E lysis requires a host factor, *SlyD*, a FKBP-type peptidylprolyl-cis-trans-isomerase cytoplasmic chaperone. In a *slyD* background, *E* is extremely unstable. This likely reflects the requirement of folding of the E_{CTD} , because *E* fusion proteins carrying heterologous CTDs do not require the chaperone. It is possible that the E_{CTD} has evolved as a regulatory domain responsive to the host *SlyD* activity. However, nothing is known about how *SlyD* levels or activity vary with host physiology.

E lysis, either in ϕ X174 infections or from inductions of a plasmid clone of *E*, occurs only in growing cells that are undergoing cell division; consistent with the biochemical studies, *E*-lysis begins about 20 min after the cessation of PG biosynthesis. The morphology of *E*-mediated lysis is striking, in that nearly all the lysing cells undergo mid-cell blebbing. The production of these septal catastrophes is indistinguishable from the cell fate imposed by cell wall antibiotics. *E* was thus the founding member of a general class of Sgls designated as the “protein antibiotics” (PA). ϕ X174 is the prototype of the *Bullavirinae* subfamily of the *Microviridae*, the only group of this ubiquitous phage type that has been the subject of significant experimentation. *E* is conserved, especially in the N-terminal domain essential for lysis, in all the 15 other *Bullavirinae*, each of which is an enteric phage demonstrated to have lytic function and plaque-forming ability. To assess the diversity and distribution of *E*-like lysis proteins in the other *Microviridae*, we have restricted analysis to the curated RefSeq database. There are 45 other *Microviridae* genomes in RefSeq, divided into the *Gokushovirinae* subfamily (15 phages) and a group of 30 unclassified phages, including members from hosts as diverse as *Chlamydia* and *Bdellovibrio*. No protein related to *E* is encoded by an annotated gene in any of these genomes or in any genome outside the *Bullavirinae*. Of these 45 phage genomes, 36 are derived from environmental metagenome, so not only is the host unknown but it is not established that any of these phages effect lysis at all. Of the 9 established lytic phages, five have chlamydial hosts, which lack PG and are not relevant to this discussion. The four bacteriolytic phages comprise the *Bdellovibrio* phage phiMH2K, the *Spiroplasma* phage SPV4, and two related phages of *Cellulophaga*, phi12.2 and phi12a.1. In the latter two phages, two proteins of comparable size to *E* with strongly predicted N-out TMDs are encoded, but both have been assigned to be virion proteins, so the identity of the Sgl is a mystery. SPV4 has two small integral membrane proteins but the topology is N-in for both. *Bdellovibrio* phage MH2K encodes a 102 protein with a single N-out TMD; although it has no detectable similarity with *E*, it is an attractive Sgl candidate because, like *E*, it is located in the +1 reading frame of a conserved gene.

The SGLs of the *Leviviridae*

The *Leviviridae* are ssRNA phages with extremely simple genetic organization. They all have three core genes, encoding the essential Mat, Coat, and Rep proteins (Fig. 11). Coat is the major capsid protein that is assembled into a T=3 shell around the ~4 kb RNA; Mat is a ~45 kDa “single molecule tail” that replaces two Coat proteins in the shell. All *Leviviridae* adsorb to a retractable host pilus as the first step in the infection cycle. Most of the genetics and molecular biology studies on *Leviviridae* have been with two phages that recognize the F conjugational pilus: MS2 and Q β . The two phages differ drastically in terms of lysis. MS2 uses a fourth protein, L, encoded in a cistron straddling the *coat-rep* junction, as its Sgl; L-like Sgls are the most common lysis proteins in the diverse *Leviviridae* and will be considered separately below. Q β does not encode a separate lysis gene; instead, lysis is a second function of the Mat protein, which, for historical reasons, is called A₂ in Q β . A₂ has a PA-type mechanism in that it acts as a non-competitive inhibitor of MurA, the conserved enzyme that catalyzes the first step of the PG precursor pathway (Fig. 2). In the MurA catalytic cycle, the two halves of the globular protein close over the catalytic cleft when the two substrates, PEP and UDP-GlcNAc, are bound; genetic and biochemical studies showed that A₂ binds across the cleft after ligand binding, thus locking the enzyme in a stable, inactive form. Recently, cryo-EM studies have revealed the structure of the A₂ protein, mounted on the virion, in complex with MurA; in effect, the Q β particle is the world’s largest enzyme inhibitor.

The Mysterious L

The target of A₂ was identified with the same simple genetic approach used for E: selecting survivors of the induction of the cloned A₂ gene and mapping dominant mutations conferring A₂-resistance to the *murA* locus. In contrast, the target of L has remained a mystery for decades. Indeed, in perhaps the last small triumph of classical phage genetics over genomics, the L gene was not noticed in the published MS2 genome sequence, despite the fact that this was the first genome ever determined and MS2 was known to be a lytic phage. In any case, after nonsense-suppressor analysis revealed that L was the essential lysis gene, many efforts were made to unravel its lytic mechanism. Unlike E and A₂, expression of L does not cause inhibition of net PG biosynthesis, which instead continues unabated until lysis. Moreover, cells undergoing lysis by MS2 infection or by induction of a cloned L gene do not develop the characteristic septal bleb catastrophes. Rigorous mutational and physiological analysis has revealed that L is the prototype of a very diverse family of Sgls, sharing a common domain structure (Fig. 12). L is almost a “mirror-image” of the E Sgl, in that L has a large, dispensable, highly basic N-terminus and an essential, more hydrophobic C-terminus. The N-terminal Domain I confers a requirement for the function of the chaperone DnaJ, strikingly reminiscent of the requirement for SlyD that is conferred by the highly basic E_{CID}. Again, this dependence on a host chaperone may be the key to regulation of lysis under differing physiological conditions. Domain II is extremely hydrophobic but, unlike a traditional transmembrane domain, has positions that are hypersensitive to conservative changes in the side chains. Domain III is the universally conserved dipeptide sequence Leu-Ser, and Domain IV is a species-specific region (Fig. 12).

Although progress has been made in identifying the important parts of L, its target is still mysterious. This is all the more frustrating because leviviral genomics, although ludicrously limited in totaling <100 kb, most of which are male-specific coliphages closely related to MS2 and Q β , indicate that the majority of the ssRNA phages use L-like Sgls. There are only 8 *Leviviridae* phages known outside of the ~30 male-specific coliphages of the MS2 and Q β clans (Table 1). Six of them have L-like Sgl sequences, five of which evolved, like L, at the *coat-rep* junction and the fifth, from *Acinetobacter* phage AP205, at a distinct

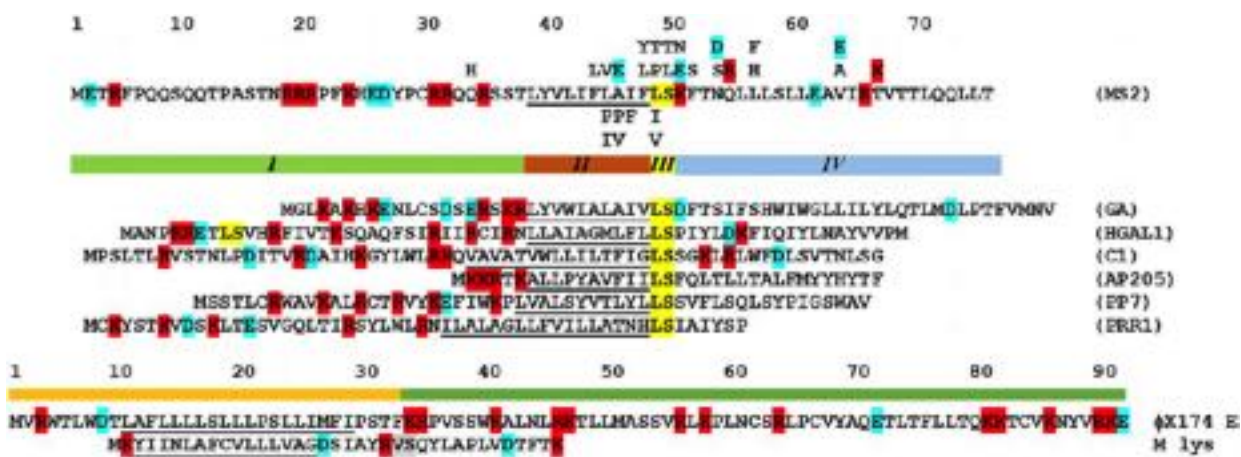


Fig. 12 Sgl diversity. Top: L-like Sgls. Domain structure of MS2 L is shown. Missense mutations that abolish lytic function or are without effect are shown above and below the L sequence. Representative L-like Sgls are aligned below the domain diagram. Sources: MS2 and GA (F-specific); HGAL-1 and C1 (IncH and IncC conjugational pili, respectively); AP205 (*Acinetobacter* type IV pilus); PP7 (*Pseudomonas* polar pilus); PRR1 (conjugational pilus of broad host range plasmid RR1). Bottom: PA-type Sgls. N-terminal domain (NTD; light brown) and C-terminal domain (CTD; green) are shown. Sources: ssDNA coliphage ϕ X174 and levivirus M (IncM conjugational pilus). Originally published in Chamakura, K., and Young, R., 2019. Phage single-gene lysis: Finding the weak spot in the bacterial cell wall. *Journal of Biological Chemistry* 294, 3350–3358.

Table 1 The known Levivirus genomes and *sgl* characteristics. All known Levivirus genomes are listed, except for the ~30 relatives of MS2 and Qb. No Sgl has been identified for LeviOR01

Phage	Pilus	Host	Sgl location	Sgl target	Aa
MS2	F	<i>E. coli</i>	coat-rep junction	L-target	75
Q	F	<i>E. coli</i>	A ₂	MurA	420
phiCB5	polar Tad	<i>C. crescentus</i>	middle rep	?	136
PP7	twitching	<i>P. aeruginosa</i>	coat-rep junction	L-target	55
PRR1	IncP-1	<i>P. aeruginosa</i> , others	coat-rep junction	L-target	54
M	IncM	<i>E. coli</i>	distal rep	MurJ	37
Hgal1	IncH	<i>E. coli</i>	coat-rep junction	L-target	65
C-1	IncC	<i>E. coli</i>	coat-rep junction	L-target	65
AP205	twitching	<i>Acinetobacter</i>	5' end	L-target	35
LeviOR01	twitching	<i>P. aeruginosa</i>	?	?	?

Note: Originally published in Chamakura, K., and Young, R., 2019. Phage single-gene lysis: Finding the weak spot in the bacterial cell wall. *Journal of Biological Chemistry* 294, 3350–3358.

site at the 5' end of the gRNA. The other two are phage M, which infects the retractable pilus encoded by IncM plasmids, and CB5, which infects the retractable TAD pili of the marine bacterium (and developmental model system) *C. crescentus*. These two genes encode type I IM proteins with predicted N-out topology like E. Although there is no detectable sequence similarity with E, the Occam's Razor hypothesis would be that both of these Sgls would be MraY inhibitors. This hypothesis was invalidated when it was shown that Lys^M was an inhibitor of MurJ and caused lysis by inhibition of PG biosynthesis (Fig. 2). Indeed, experiments that showed that Lys^M bound MurJ and caused cytoplasmic accumulation of Lipid I, not only settled the target of Lys^M but unambiguously settled the role of MurJ as the long-sought Lipid II flippase.

Nothing has been reported for Lys^{CB5}, but the addition of Lys^M to Qβ A₂ and phiX174 E brings the total number of different PA-type SGLs to 3, each targeting a different step in PG biosynthesis. Considering the diversity of these Sgls, it is a reasonable and exciting hypothesis that any step in the PG biosynthesis pathway, including periplasmic steps, would be a suitable target for a PA-type Sgl. However, after “solving” the CB5 SGL system, which is underway in our laboratory, the sad truth is that there are no more *Leviviridae* with sequenced genomes. Recently some relief from this dismal situation has been brought about by two meta-virome studies revealing a total of ~80 new leviviral sequences. Only one of these genomes had an annotated lysis gene, which turned out to encode an L-like sequence. It will be interesting to see if a rich library of PA-type Sgls can be derived from analysis of these sequences. The first obstacle, of course, is identifying the Sgl in each of these genomes. Considering how small some already identified Sgls are, identifying candidates by genome analysis is going to be problematic. Also, the meta-virome sourcing means that the hosts corresponding to these Sgls are not known, so testing of each Sgl will have to be done in *E. coli* using inducible plasmid systems, with no guarantee that the cognate *E. coli* target will be sensitive to binding and inhibition.

Extrusion

Filamentous phages (*Inoviridae*) produce chronic infections of a culture and release phage progeny by extrusion rather than lysis. The majority of work on filamentous phage has been done on the closely-related *E. coli* Ff phages f1, M13, and fd. These virions have distinct filamentous morphology derived from a single molecule of the circular single-stranded gDNA surrounded in a sheath-like structure made of ~3000 of copies of the major coat protein pVIII. The ends of the filament are made up of minor coat proteins pVII and pIX and pIII and pVI. During replication, all of the structural proteins of Ff phages are targeted to the IM before being assembled in a secretion-like process. Three proteins, pI, pIV, and pXI are not part of the virion structure but form a trans-envelope complex with pI and pXI at the IM and pIV forming a pore-like structure in the OM. In the later stage of infection, positive ssDNA is coated with pV, forming a precursor virion complex with a 32-bp hairpin packaging signal at one end. Genetic analysis indicates that assembly is initiated when the packaging signal interacts with the C-termini of pVII and pIX. The pV-coated ssDNA is brought to the pI-pXI complex and on the cytoplasmic face of the IM. Then pV dissociates from the complex and is replaced by a helical shell of pVIII as the ssDNA is translocated across the IM. Ff assembly requires ATP and the N-terminal region of pI has an essential Walker A motif. Therefore, pI is probably a molecular motor that drives the assembly process. After all of the Ff DNA is covered with pVIII, the filament is released when pVI and pIII bind to the terminal cap of the virion. It is important to note here, although Ff phage do not kill the host cell, they still form plaques. In this case, the clearings are a result of a reduced growth rate that reflects the burden of Ff phage production and secretion on the physiology of infected cells.

Phage Egress in Archaea

The main load-bearing component of the bacterial envelope is the peptidoglycan, so a muralytic enzyme is always a key component of MGL systems. The cell envelope structures of the archaea are extremely diverse and differ from bacteria in profound

ways, chief among which is that, except for a few species of methanogens, archaea lack a load-bearing polysaccharide sacculus. In the methanogens a sacculus is formed from pseudomurein, which resembles bacterial murein in having disaccharide repeating units containing GlcNAc and cross-linking oligopeptides. However, in pseudomurein, the cross-linking oligopeptides contain no D-amino acids and originate in amide links to L-N-acetylalosaminuronic acid rather than MurNAc. Several *Methanobacterium* phages, including the siphophage PsiM2, has been shown to encode and produce an endoisopeptidase (PeiP) that cleaves the epsilon-Ala-Lys isopeptide bond in pseudomurein. Gene *peiP* encodes no secretion signal but is surrounded by five genes that encode proteins with one or more TMDs. The Occam's Razor argument is that one or more of these membrane proteins provide temporally controlled permeabilization of the cytoplasmic membrane, rather than having to invoke a novel secretion mechanism. Whether this mechanism involves holin-like triggering and formation of micron-scale holes in the cytoplasmic membrane remains to be addressed.

As the primary and outer-most component of the envelope, most archaea have S-layers that can not only confer tolerance for extremely harsh conditions but in some cases seem to determine cell shape. Many archaeal viruses appear to conduct chronic infections, released like the bacterial *Inoviridae* by an extrusion process without reducing host viability. Operationally, the extrusion process appears to be similar to the process of extrusion of enveloped viruses in eukaryotes, in which the ESCRT machinery pinches off the viral-containing membrane buds. Even in these cases, release of the virus still requires disruption of the S-layer, by unknown mechanisms. Lytic viruses of the *Crenarchaea* have evolved a spectacular solution for release of the progeny virions, the lysis pyramids. These stunning structures are 7-sided pyramids formed by homo-oligomerization of a single phage protein. When mature, the pyramids protrude through the surface of the cell, penetrating the membrane and dislodging the S-layer array, before opening up to form ~100 nm holes through which the virions are released. These release structures were observed in lytic infections of *Sulfolobus* by the small phages STIV and SIRV2, which are unrelated except for having similar pyramid proteins. One perspective is that in the pleomorphic *Sulfolobus* cells lacking a load-bearing sacculus, there is insufficient osmotic differential between the cytoplasm and the environment to allow explosive events like those effected by MGL.

Extrusion and pyramid-scheme viral release mechanisms would seem to be incompatible with *Caudovirales* structure, and, accordingly no tailed phages have been isolated for the *Crenarchaeota*. However, many phages of the *Euryarchaeota* are *Caudovirales*, comprising all three classic morphologies that are morphologically indistinguishable from their bacterial brethren. In the curated RefSeq database, there are 33 complete genomes of phages of *Euryarchaeota*, of which 20 are *Caudovirales* and two more are *Sphaerolipoviruses*, which are evolutionarily related to the Corticoviruses. All of these phages presumably use lytic mechanisms, rather than extrusion and have been subjected to the evolutionary pressures that led to the development of the complex, multi-step, highly regulated MGL systems in the phages of bacteria. Other than the case of PsiM2, where the presumed MGL-type endolysin has been identified (PeiP), there are no genes related to any lysis gene found in phages of bacteria. Considering that archaea, once thought to be restricted to extreme environmental niches, are now being identified in most environments that have bacteria and may be major players in the ecosystem, it would seem timely to investigate the lysis mechanisms of archaea phages.

Further Reading

- Cahill, J., Rajaure, M., Holt, A., Moreland, *et al.*, 2017. Suppressor analysis of the fusogenic lambda spanins. *Journal of Virology* 91, e00413–e00417.
- Cahill, J., Young, R., 2019. Phage lysis: Multiple genes for multiple barriers. *Advances in Virus Research* 103, 33–70.
- Chamakura, K., Young, R., 2019. Phage single-gene lysis: Finding the weak spot in the bacterial cell wall. *Journal of Biological Chemistry* 294, 3350–3358.
- Chen, Y., Young, R., 2016. The last *r* locus unveiled: T4 RIII is a cytoplasmic antiholin. *Journal of Bacteriology* 198, 2448–2457.
- Dewey, J.S., Savva, C.G., White, R.L., *et al.*, 2010. Micron-scale holes terminate the phage infection cycle. *Proceedings of the National Academy of Sciences* 107, 2219–2223.
- Fernandes, S., São-José, C., 2018. Enzymes and mechanisms employed by tailed bacteriophages to breach the bacterial cell barriers. *Viruses* 10, 396.
- Fischetti, V., 2018. Development of phage lysins as novel therapeutics: A historical perspective. *Viruses* 10, 310.
- Hernandez-Morales, A.C., Lessor, L.L., Wood, T.L., *et al.*, 2018. Genomic and biochemical characterization of *Acinetobacter* podophage petty reveals a novel lysis mechanism and tail-associated depolymerase activity. *Journal of Virology* 92, e01064.
- Kongari, R., Rajaure, M., Cahill, J., *et al.*, 2018. Phage spanins: diversity, topological dynamics and gene convergence. *BMC Bioinformatics* 19, 326.
- Oliveira, H., Melo, L.D., Santos, S.B., *et al.*, 2013. Molecular aspects and comparative genomics of bacteriophage endolysins. *Journal of Virology* 87, 4558–4570.
- Quemin, E.R., Quax, T.E., 2015. Archaeal viruses at the cell envelope: Entry and egress. *Frontiers in Microbiology* 6, 552.
- Rajaure, M., Berry, J., Kongari, R., Cahill, J., Young, R., 2015. Membrane fusion during phage lysis. *Proceedings of the National Academy of Sciences* 112, 5497–5502.
- Rakonjac, J., Russel, M., Khanum, S., Brooke, S.J., Rajič, M., 2017. Filamentous phage: Structure and biology. *Recombinant Antibodies for Infectious Diseases*, 2017. Cham: Springer, pp. 1–20.
- Saier, M.H., Reddy, B.L., 2015. Holins in bacteria, eukaryotes, and archaea: Multifunctional xenologues with potential biotechnological and biomedical applications. *Journal of Bacteriology* 197, 7–17.
- White, R., Chiba, S., Pang, T., *et al.*, 2011. Holin triggering in real time. *Proceedings of the National Academy of Sciences* 108, 798–803.

Virus Budding

Lara Rheinemann and Wesley I Sundquist, University of Utah, Salt Lake City, UT, United States

© 2021 Elsevier Ltd. All rights reserved.

Overview

To spread infection, viruses must exit producer cells and transmit their genetic material into target cells. Viruses have evolved two general strategies for cellular egress: (1) Enveloped viruses acquire a host-derived lipid membrane as they breach the limiting membranes of the cell. During this process, the viral membrane typically also acquires one or more viral glycoproteins that bind target cell receptors and facilitate the membrane fusion step required for viral entry. (2) Non-enveloped viruses, by contrast, exit cells when the plasma membrane is disrupted, typically by cell lysis. Non-enveloped viruses then infect target cells either by disrupting limiting membranes to gain access to the cytoplasm or by pumping their nucleic acids directly into the cell. In recent years, the distinction between enveloped and non-enveloped viruses has been blurred by the discovery that some viruses, traditionally thought of as non-enveloped, can exit cells within vesicles. These are termed “quasi-enveloped” viruses.

The process of enveloped virus release comprises a series of coordinated steps, which are illustrated for human immunodeficiency virus type 1 (HIV-1) in **Fig. 1**:

- (1) **Assembly:** Viral proteins and other essential components co-assemble to form virions. Many viruses assemble at the plasma membrane, but others assemble at internal membranes or in the cytoplasm before trafficking to the plasma membrane or exiting via the secretory system.
- (2) **Envelopment:** The host membrane is bent and wrapped around the nascent virion.
- (3) **Budding:** The membrane stalk connecting the virion to the host membrane is constricted and severed to release the enveloped particle.
- (4) **Maturation:** Most enveloped viruses undergo further proteolytic and conformational maturation steps during or after budding. Maturation converts the assembly-competent virion into an infectious virus that can enter, uncoat, and replicate in the new target cell.

The complex series of events that accompany enveloped viral egress must be coordinated spatially and temporally, and these events are typically orchestrated by virally-encoded, multifunctional structural proteins. These proteins bind and remodel the membrane, self-assemble into virions, package other essential components such as nucleic acids into the nascent virion, and contain or recruit all of the activities necessary for budding and maturation. This article will describe general principles of enveloped virus assembly and release using the well-characterized HIV-1 Gag protein as a paradigm for a viral structural protein. Important principles employed by other viral families will also be discussed.

Assembly

All retroviruses, including HIV-1, express a Gag polyprotein that coordinates assembly of the immature virion (Sundquist and Krausslich, 2012; Meng and Lever, 2013; Lingappa *et al.*, 2014). When expressed alone, Gag is released within virus-like particles (VLPs), indicating that it contains or recruits all activities necessary for assembly, envelopment, and budding. An exciting new

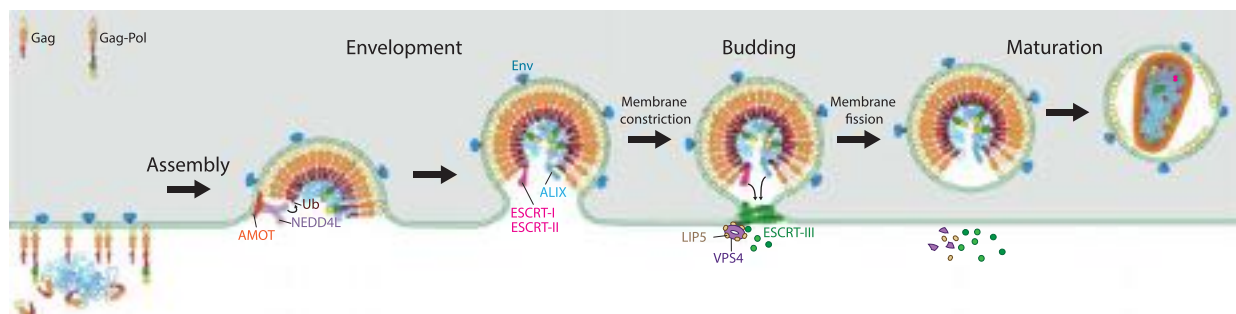


Fig. 1 Stages of HIV-1 virion formation. HIV-1 virion formation is coordinated by the multifunctional structural polyprotein Gag. Complexes of Gag with the viral RNA genome (blue) traffic to the plasma membrane where Gag assembles and binds other virion components, including the longer Gag-Pol polyprotein, which contains the viral protease, reverse transcriptase and RNase H enzymes. Gag also recruits three different host factors to facilitate membrane constriction and fission: AMOT-NEDD4L (orange and light purple, respectively), and the early-acting ESCRT-associated factors, ESCRT-I/II (pink) and ALIX (blue), which in turn recruit ESCRT-III proteins (green). ESCRT-III proteins form polymeric filaments that constrict the bud neck with the help of the VPS4 AAA + ATPase and its cofactor LIP5. Protease activation during budding leads to Gag and Gag-Pol processing and formation of the mature, infectious virion.

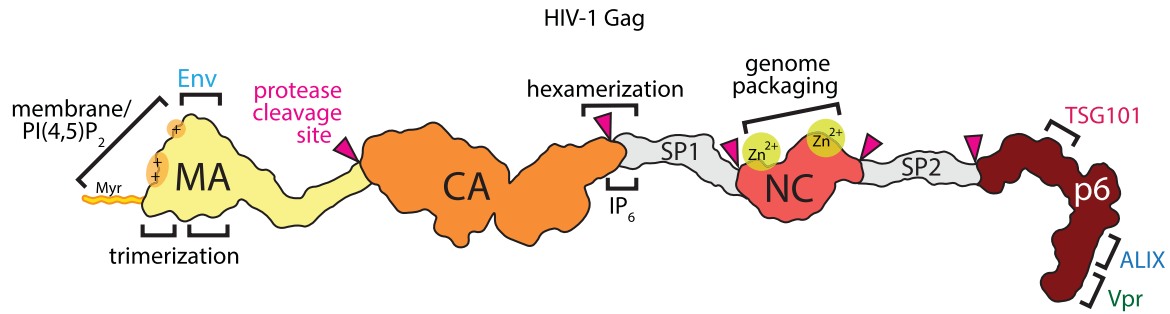


Fig. 2 The HIV-1 Gag polyprotein. HIV-1 Gag comprises four functional elements connected by two spacer peptides SP1 and SP2 (gray). MA (yellow) facilitates membrane binding and Env incorporation. CA (orange) mediates assembly of the immature capsid and, after proteolytic processing, forms the mature conical capsid. NC (red) binds the viral RNA genome through two zinc finger motifs. p6 (brown) binds Vpr and recruits early-acting ESCRT proteins TSG101 (a subunit of the ESCRT-I complex) and ALIX to facilitate membrane fission. Pink arrowheads denote proteolytic cleavage sites during maturation.

variation on this same theme is the discovery of Gag-related cellular proteins that can self-assemble, exit producer cells, and transfer RNA molecules between cells (Pastuzyn *et al.*, 2018; Ashley *et al.*, 2018).

As illustrated in **Fig. 2**, HIV-1 Gag comprises four functional elements: MA, CA, NC and p6, connected by two spacer peptides: SP1 and SP2. The MA domain in Gag (MA^{Gag}) targets assembly to the plasma membrane, NC^{Gag} binds the dimeric viral RNA genome, the CA^{Gag} and SP1^{Gag} elements mediate spherical particle assembly, and p6^{Gag} recruits cellular ESCRT (Endosomal Sorting Complexes Required for Transport) pathway factors required for virus budding. Gag also packages several other components required for full infectivity, including the viral enzymes, which are packaged as co-assembling Gag-Pol fusion proteins, the viral Env protein, whose intracellular domain interacts with MA^{Gag} , and Viral Protein R (Vpr), which binds p6^{Gag} .

- (1) *RNA packaging*: Following synthesis in the cytoplasm, Gag monomers or low-order multimers bind two associated copies of the full-length viral genomic RNA through a highly structured packaging site located near the 5' end of the genome (Jouvenet *et al.*, 2009; Kutluay and Bieniasz, 2010; Keane *et al.*, 2015). This interaction is mediated by NC^{Gag} and enables specific selection of the HIV-1 genome over cellular RNAs and viral RNAs undergoing translation (Kaddis Maldonado and Parent, 2016; Dubois *et al.*, 2018; Kuzembayeva *et al.*, 2014; Brown *et al.*, 2020). Several host proteins, including ABCE1, Staufen1, and DDX6, can associate with Gag-RNA complexes and have been proposed to promote Gag trafficking, multimerization, and/or genome encapsidation, although these activities are not yet fully defined (Reed *et al.*, 2012; Zimmerman *et al.*, 2002; Chatel-Chaix *et al.*, 2007).
- (2) *Membrane trafficking*: Gag and Gag-RNA complexes move to the plasma membrane and associate with cholesterol-rich microdomains, commonly termed “lipid rafts” (Ono and Freed, 2001). Binding of Gag to the plasma membrane is mediated by a bipartite signal in the MA^{Gag} domain that comprises a post-translational myristic acid modification at the amino-terminus and a cluster of basic residues, the highly basic region (HBR). Prior to membrane binding, the myristate is sequestered in a hydrophobic cavity within the globular MA^{Gag} domain. The HBR interacts with acidic head groups of the plasma membrane-specific phospholipid phosphatidylinositol-(4,5)-bisphosphate ($\text{PI}(4,5)\text{P}_2$) and this interaction, perhaps in concert with Gag multimerization, exposes the myristate moiety. $\text{PI}(4,5)\text{P}_2$ and the conformational “myristoyl switch” enable targeting and stable Gag association with the plasma membrane (Chukkapalli and Ono, 2011; Tang *et al.*, 2004). In addition to $\text{PI}(4,5)\text{P}_2$, MA^{Gag} can also bind cellular RNA, especially tRNAs. The MA-tRNA interactions likely prevent non-specific binding of Gag to cellular membranes other than the plasma membrane (Chukkapalli *et al.*, 2010; Alfadhli *et al.*, 2011; Kutluay *et al.*, 2014; Gaines *et al.*, 2018).
- (3) *Virion assembly*: At the plasma membrane, Gag molecules multimerize and assemble into a hexagonal lattice through lateral protein-protein interactions that are mediated by the CA-SP1 region (Schur *et al.*, 2016; Wagner *et al.*, 2016). Host-derived inositol hexakisphosphate (IP_6) binds within the CA-SP1 hexamers and facilitates formation of the immature Gag lattice (Dick *et al.*, 2018; Mallery *et al.*, 2018). In addition to the genomic RNA, Gag interacts directly with other viral proteins that are incorporated into the nascent virion, including the multifunctional accessory protein Vpr, which binds a leucine-rich element located within p6^{Gag} (Kondo *et al.*, 1995), and the cytoplasmic tail of the gp41 subunit of the heterotrimeric transmembrane Env protein, which is incorporated within holes in the associating Gag lattice created by MA trimerization (Tedbury and Freed, 2014; Tedbury *et al.*, 2016; Hill *et al.*, 1996; Pezeshkian *et al.*, 2019). In addition to the Gag polyprotein, the full-length viral RNA also encodes the Gag-Pol polyprotein. The longer Gag-Pol protein is translated by a ribosomal frameshifting mechanism, contains the viral enzymes, and is incorporated into the nascent virion through interactions with Gag (Smith *et al.*, 1993).

Envelopment

The multifunctional structural proteins that mediate assembly and membrane targeting also appear to facilitate virion envelopment by inducing membrane curvature. The hexagonal HIV-1 Gag lattice contains small discontinuities that accommodate

Table 1 *Viral proteins that mediate assembly and budding of enveloped viruses.* Virus families and species are listed as described in the Virus Taxonomy: 2018b Release by the International Committee on Taxonomy of Viruses (ICTV). To be included in the table, virus families must be enveloped and contain at least one species that infects mammals

<i>Virus family</i>	<i>Example species</i>	<i>Protein(s) mediating assembly and budding</i>
Arenaviridae	Lassa virus, lymphocytic choriomeningitis virus	matrix protein Z
Arteriviridae	Equine arteritis virus, porcine reproductive and respiratory syndrome virus	envelope proteins GP5 and M
Asfarviridae	African swine fever virus	p72 (outer capsid layer), polyproteins pp220 and pp62 (inner core shell)
Bornaviridae	Mammalian 1 orthobornavirus	matrix protein M
Coronaviridae	Middle East respiratory syndrome-related coronavirus, severe acute respiratory syndrome-related coronavirus,	membrane protein M, envelope protein E
Filoviruses	Zaire ebolavirus, Marburg marburgvirus	matrix protein VP40
Flaviviridae	Dengue virus, West Nile virus, Yellow fever virus, hepatitis C virus	envelope proteins prM and E, aided by non-structural proteins
Hantaviridae	Sin Nombre orthohantavirus	envelope proteins Gn and Gc
Hepadnaviridae	Hepatitis B virus	envelope proteins S and L
Herpesviridae	human alphaherpesvirus	Nuclear egress complex (pUL31 and pUL34, primary envelopment), pUL7, pUL51 and other tegument proteins (secondary envelopment)
Matonaviridae	Rubella virus	envelope proteins E1 and E2
Nairoviridae	Crimean-Congo hemorrhagic fever orthonairovirus	envelope proteins Gn and Gc
Orthomyxoviridae	Influenza A virus, influenza B virus	matrix protein M1
Paramyxoviridae	Hendra henipavirus, Nipah henipavirus, measles morbillivirus	matrix protein M
Peribunyaviridae	Bunyamwera orthobunyavirus, La Crosse virus	envelope proteins Gn and Gc
Phenuiviridae	Rift Valley fever virus	envelope proteins Gn and Gc
Pneumoviridae	Human metapneumovirus, human orthopneumovirus	matrix protein M
Poxviridae	Vaccinia virus, variola virus	membrane proteins A14 and A17, scaffold protein D13 and other viral membrane assembly proteins (formation of mature virions); A27, B5, and F13 (formation of enveloped virions)
Rhabdoviridae	Rabies lyssavirus	matrix protein M
Retroviridae	Human immunodeficiency virus-1, human immunodeficiency virus-2,	Gag polyprotein
Tobnaviridae	Equine torovirus, porcine torovirus	membrane protein M
Togaviridae	Chikungunya virus, eastern equine encephalitis virus	envelope proteins E1 and E2

declination and allow the immature lattice to bend the membrane and create a spherical virion (Schur *et al.*, 2016). The host factor angiomin (AMOT) has also been implicated in HIV-1 virion envelopment because fully enveloped spherical particles are not formed efficiently in the absence of AMOT (Mercenne *et al.*, 2015). The Bar domain of AMOT likely contributes to this activity as this domain has been shown to bend and tubulate membranes in other contexts (Nishimura *et al.*, 2018).

Unlike HIV-1, betaretroviruses and spumaviruses assemble in the cytoplasm before being trafficked to the plasma membrane. Thus, assembly and budding are spatially and temporally separated in these viruses. For example, Gag proteins from prototypal foamy virus, Mason-Pfizer monkey virus, and mouse mammary tumor virus preassemble into immature virions near the pericentriolar region and are then trafficked on microtubules to associate with membranes and the viral Env protein (Müllers, 2013; Hütter *et al.*, 2013; Swanstrom and Wills, 1997). Hepatitis B virus (HBV), a hepadnavirus, similarly forms capsids in the cytoplasm, which then associate with envelope proteins in cellular membranes to mediate particle envelopment and budding (Prange, 2012; Blondot *et al.*, 2016).

More broadly, most enveloped viruses have multifunctional structural proteins that mediate assembly and envelopment. Often, this role is fulfilled by matrix proteins that bind to the viral membrane, assemble into higher-order structures, and link internal ribonucleoprotein complexes to external envelope glycoproteins. In viruses that lack classical matrix proteins, such as hantaviruses (Hepojoki *et al.*, 2012; Cifuentes-Munoz *et al.*, 2014; Muyangwa *et al.*, 2015) and alphaviruses (Brown *et al.*, 2018), cytoplasmic tails of envelope glycoproteins act as matrix protein surrogates by directly interacting with nucleoproteins. Larger, more complex DNA viruses such as pox viruses (Roberts and Smith, 2008; Liu *et al.*, 2014) and herpes viruses (Lv *et al.*, 2019) divide these functions between multiple viral proteins (see Table 1).

Budding

ESCRT-dependent budding

As membrane envelopment proceeds, the membrane is constricted until nascent virions are connected to the plasma membrane by a thin stalk that must be severed to separate the viral and cellular membranes. Many enveloped viruses accomplish membrane constriction and fission by recruiting the machinery of the cellular ESCRT pathway (Votteler and Sundquist, 2013).

ESCRT-independent enveloped viruses also exist, however, and these viruses must therefore either recruit other, as yet unidentified, host factors or encode viral proteins that mediate budding.

The ESCRT pathway mediates cellular membrane fission events throughout eukarya and also in some archaeal species (McCullough *et al.*, 2018; Christ *et al.*, 2017; Scourfield and Martin-Serrano, 2017; Henne *et al.*, 2013). The pathway was initially identified as the machinery that mediates intraluminal vesicle budding into specialized late endosomes, termed multivesicular bodies (MVB) (Hanson and Cashikar, 2012), but is now known to act at many other cellular membranes, including during cytokinetic abscission (Carlton and Martin-Serrano, 2007; Morita *et al.*, 2007), resealing of the post-mitotic nuclear envelope (Olmos *et al.*, 2015; Vietri *et al.*, 2015), membrane repair (Scheffer *et al.*, 2014; Jimenez *et al.*, 2014; Skowrya *et al.*, 2018), closure of autophagosomes (Takahashi *et al.*, 2018; Zhou *et al.*, 2019), and neuronal pruning (Zhang *et al.*, 2014; Loncle *et al.*, 2015; Issman-Zecharya and Schuldiner, 2014). Notably, all of these membrane fission events involve constricting membranes toward a cytoplasm-filled neck and are therefore topologically equivalent to virus budding from the plasma membrane.

ESCRT-mediated membrane fission events are catalyzed by a common core machinery (McCullough *et al.*, 2018; Christ *et al.*, 2017; Scourfield and Martin-Serrano, 2017; Banjade *et al.*, 2019), which is recruited to different membranes by adapter proteins. These membrane-specific adapters recruit early-acting ESCRT proteins, which help to stabilize membrane curvature and also nucleate assembly of late-acting ESCRT-III proteins, which form the core fission machinery. ESCRT-III proteins can be recruited by three known mechanisms: (1) Adapters can recruit Bro1 domain-containing proteins such as ALIX, which serves as a bridge to the ESCRT-III proteins, (2) Adapters can bind the ESCRT-I complex, which in turn recruits ESCRT-III proteins via intermediate ESCRT-II complexes, and (3) The nuclear LEM2 adapter binds CHMP7, a hybrid ESCRT-II/ESCRT-III protein, which then binds other ESCRT-III proteins.

Humans express 12 related ESCRT-III proteins that are divided into eight subfamilies, CHMP1–7 and IST1, with some subfamilies comprising several homologs. ESCRT-III proteins can adopt “closed” and “open” conformations. In the autoinhibited closed state, ESCRT-III proteins are soluble and monomeric. When autoinhibition is relieved, ESCRT-III subunits open and can then bind membranes and form curved helical filaments. These filaments constrict membranes and recruit VPS4 AAA + ATPases. VPS4 enzymes dynamically remodel and disassemble ESCRT-III filaments, using the energy of ATP hydrolysis to extract individual ESCRT-III subunits and release them back into the cytoplasm. These enzymes thereby power the virus budding cycle, although the precise mechanism by which ESCRT-III filaments and VPS4 enzymes collaborate to mediate fission is not yet fully understood.

Enveloped viral structural proteins recruit the ESCRT pathway using motifs that mimic cellular ESCRT adapters (Votteler and Sundquist, 2013; Hurley and Cada, 2018; Lippincott-Schwartz *et al.*, 2017). These motifs were initially discovered in retroviral Gag proteins (Gottlinger *et al.*, 1991; Huang *et al.*, 1995; Parent *et al.*, 1995; Xiang *et al.*, 1996) and were termed “late domains” because they exerted their effects at a late stage of assembly. Analogous late domains were subsequently identified in many other enveloped viruses. Late domains can frequently function from different positions within viral structural proteins and can be swapped between viruses (Parent *et al.*, 1995; Yuan *et al.*, 2000), consistent with the idea that although they have different primary binding partners, they all ultimately converge on common downstream ESCRT-III proteins. Several different classes of late domains are now well understood, and others have been identified but remain to be linked to ESCRT binding partners.

P(S/T)AP: The P(S/T)AP late domain (where the second residue can be either a serine or a threonine) was first identified in the p6 polypeptide of HIV-1 Gag (Gottlinger *et al.*, 1991; Huang *et al.*, 1995), and subsequently identified in structural proteins of other retroviruses, filoviruses, arenaviruses and reoviruses (Votteler and Sundquist, 2013). The P(S/T)AP motif recruits the four-protein ESCRT-I complex by binding the UEV domain of the TSG101 subunit (Demirov *et al.*, 2002; Garrus *et al.*, 2001; VerPlank *et al.*, 2001; Martin-Serrano *et al.*, 2001). P(S/T)AP motifs are found in several cellular ESCRT adapter proteins, such as the early endosomal protein HRS (Bache *et al.*, 2003).

YPX_nL: YPX_nL late domains (where X_n can vary in sequence and length) recruit ALIX by binding the central V domain (Strack *et al.*, 2003; Martin-Serrano *et al.*, 2003; von Schwedler *et al.*, 2003). HIV-1 contains a YPX_nL late domain, although this motif is less critical for budding than the PTAP motif in most cell types. Other viruses rely exclusively or primarily on YPX_nL domains for budding, including other retroviruses, paramyxoviruses, flaviviruses, and possibly herpesviruses (Votteler and Sundquist, 2013). Divergent structural proteins that lack a readily detectable YPX_nL motif, yet still bind to ALIX, have also been described (Boonyaratanakornkit *et al.*, 2013; Lee *et al.*, 2012), suggesting that ALIX-recruiting sequence motifs may accommodate more variability than has been documented to date. Cellular YPX_nL-containing ESCRT adapters recruit ALIX during exosome biogenesis and lysosomal sorting (Baietti *et al.*, 2012; Dores *et al.*, 2012).

PPXY: The PPXY late domain (where X can be any residue but is often a proline) binds WW domains in NEDD4-like HECT ubiquitin E3 ligases. PPXY late domains are found in retroviruses, filoviruses, arenaviruses, rhabdoviruses, and hepadnaviruses (Votteler and Sundquist, 2013). HIV-1 Gag does not contain a recognized PPXY motif. Nevertheless, overexpression of the ubiquitin E3 ligase NEDD4L induces budding of HIV-1 Gag constructs that lack PTAP and YPX_nL late domains (Usami *et al.*, 2008; Chung *et al.*, 2008). In this context, NEDD4L is recruited by a cellular PPXY-containing protein AMOT, which also binds HIV-1 Gag (Mercenne *et al.*, 2015). The related angiomin-like 1 (AMOTL-1) recruits NEDD4 family members to paramyxovirus M proteins, which also lack discernible PPXY domains (Pei *et al.*, 2010; Ray *et al.*, 2019). NEDD4 E3 enzymes transfer K63-linked ubiquitin chains onto PPXY-containing cellular proteins, and they regulate endocytosis, ESCRT-dependent MVB cargo selection and protein trafficking through ubiquitination (Rotin and Kumar, 2009).

It is not yet fully understood how NEDD4 family members promote budding, but it has been suggested that ubiquitination of viral structural proteins, or other proteins within the budding site, recruits the ESCRT pathway because ALIX and TSG101 both contain ubiquitin-binding domains that can recognize ubiquitinated cargos for MVB incorporation and lysosomal degradation

(Shields and Piper, 2011). Consistent with this idea, retroviral Gag proteins are typically ubiquitinated (although this requirement is not absolute (Zhadina *et al.*, 2007)), and ubiquitin depletion or mutations that prevent ubiquitin ligase recruitment inhibit retrovirus budding (Patnaik *et al.*, 2000; Strack *et al.*, 2000).

FPIV: The paramyxovirus SV5 employs an FPIV motif to facilitate budding. SV5 release is ESCRT-dependent and is augmented by AMOTL1, but the binding partner for the FPIV motif remains to be defined (El Najjar *et al.*, 2014).

Viral structural proteins typically encode multiple late domains that function in synergy. For example, HIV-1 p6^{Gag} contains both a P (S/T)AP and a YPX_nL motif, the HTLV-I Gag and Ebola virus (EBOV) structural protein VP40 proteins contain adjacent PPPY and PTAP late domains that bind both TSG101 and NEDD4, and murine leukemia virus Gag contains all three canonical late domains (Votteler and Sundquist, 2013). Nevertheless, some viruses contain a single (or at least dominant), late domain. For example, the retroviral EIAV Gag protein appears to recruit ALIX exclusively through a single YPX_nL motif (Votteler and Sundquist, 2013).

All modes of viral ESCRT pathway recruitment ultimately converge on ESCRT-III, the machinery that catalyzes membrane fission. In the best-studied cases such as HIV-1, multiple different mammalian ESCRT-III proteins have been shown to localize to the bud neck (Jouvenet *et al.*, 2011), but only CHMP2 and CHMP4 family members seem to perform indispensable functional roles (Sandrin and Sundquist, 2013; Morita *et al.*, 2011). ESCRT-III proteins form helical filaments in the bud neck and progressively constrict it with the help of VPS4 enzymes, as described above. Ultimately, a membrane fission reaction severs the neck, releasing the virion from the cell.

ESCRT-independent budding

Some enveloped viruses bud independently of the ESCRT pathway, including alphaviruses (Brown *et al.*, 2018), some paramyxoviruses (Salditt *et al.*, 2010; Utley *et al.*, 2008), and influenza A virus (Rossman and Lamb, 2011). ESCRT-independent membrane scission mechanisms are generally not well understood, but appear to involve as yet unidentified cellular factors, virally encoded proteins, or a combination.

Some RNA viruses, such as alphaviruses, contain an outer glycoprotein shell that completely covers the exterior of the viral membrane. The formation of this external protein coat has been suggested to play a role in virus budding, both in ESCRT-independent and ESCRT-dependent viruses. The alphaviral transmembrane glycoproteins E1 and E2 are embedded in the viral lipid envelope and form heterodimers that further trimerize into a continuous icosahedral lattice. These interactions are required to complete the budding step, and completion of the E1-E2 protein lattice, together with the nucleocapsid interactions across the viral membrane may be sufficient to drive both membrane curvature and fission (Brown *et al.*, 2018; Weissenhorn *et al.*, 2013). In contrast, flaviviruses and hepaciviruses also contain an outer protein coat, but still depend on the ESCRT pathway to complete their replication cycle (see below). Thus, external protein coats can apparently either replace or act in concert with the ESCRT pathway.

Influenza A is another ESCRT-independent virus (Bruce *et al.*, 2009; Watanabe and Lamb, 2010; Chen *et al.*, 2007), but in this case, the viral transmembrane protein M2 appears to mediate membrane fission (Rossman *et al.*, 2010). During particle assembly, the hemagglutinin (HA) and neuraminidase (NA) glycoproteins are targeted to the plasma membrane and cluster in lipid rafts. The matrix protein M1 interacts with the cytoplasmic tails of HA and NA, polymerizes against the membrane, and apparently acts in concert with HA to induce membrane curvature. M1 also recruits M2 to the bud neck, and an amphipathic helix in the cytoplasmic M2 tail inserts, deforms and promotes plasma membrane fission (Martyna and Rossman, 2014; Chlanda and Zimmerberg, 2016). M2 also functions as a pH-regulated ion channel that facilitates the release of viral ribonucleoprotein complexes from the endosome into the cytoplasm, but ion channel and membrane fission appear to be independent activities (Rossman *et al.*, 2010).

Intracellular budding

Some viral families envelop and bud into internal cellular membranes rather than at the plasma membrane. Budding into the lumen of a cellular organelle is topologically equivalent to budding from the plasma membrane, but in these cases the viral particle is temporarily surrounded by two membranes, the viral envelope and the organelle membrane. Virion release into the extracellular space therefore requires transport to and fusion of the virion-containing organelle with the plasma membrane.

One such case is the hepatitis B virus (HBV), which co-opts the MVB pathway for egress (Prange, 2012; Blondot *et al.*, 2016; Patient *et al.*, 2009). The three viral envelope proteins S, M, and L form budding sites at the MVB membrane that recruit mature cytoplasmic nucleocapsids. The assembling HBV particles bud into the MVB lumen in an ESCRT-dependent reaction that creates intraluminal virions. MVBs then fuse with the plasma membrane to release the enveloped viral particles from the cell, a process that resembles exosome release.

Herpesviruses are released via the secretory pathway in a complex series of events that requires several viral and cellular proteins (Lv *et al.*, 2019; Fradkin and Budnik, 2016; Owen *et al.*, 2015). Herpesviral genome replication, capsid assembly, and genome packaging all take place in the nucleus. The fully assembled nucleocapsids are too large to escape into the cytoplasm through nuclear pores and instead exit the nucleus and cell by undergoing several steps of envelopment and de-envelopment at multiple cellular membranes. During primary envelopment, the nucleocapsids bud through the inner nuclear membrane into the perinuclear space, thereby acquiring a lipid envelope. This process requires the virus to remodel the nuclear lamina. Cellular and viral kinases phosphorylate components of the nuclear lamina, leading to its disassembly. Two viral proteins, pUL31 and pUL34, then assemble into a cage-like nuclear egress complex, that carries the virion across the inner nuclear membrane (Bigalke and Heldwein, 2015). There is some evidence that the ESCRT machinery may also be involved in facilitating the membrane fission step required

to release the enveloped virion into the intermembrane space (Arii *et al.*, 2018). The primary envelope then fuses with the outer nuclear membrane in a process termed de-envelopment. The viral glycoproteins are necessary for de-envelopment, likely because they mediate fusion with the outer nuclear membrane. Once in the cytoplasm, nucleocapsids associate with tegument proteins, which in the mature virion occupy the space between the nucleocapsid and the envelope. The nucleocapsids then bud into vesicles, whose origins have been variously described as the trans-Golgi network, endosomes, or autophagic membranes. Recruitment to sites of secondary envelopment is promoted by tegument proteins through interactions with vesicle membranes and viral glycoproteins. Early-acting ESCRT proteins are not required for this process, but there are reports that ESCRT-III and VPS4 activity are required for virion release (Crump *et al.*, 2007; Calistri *et al.*, 2007; Pawliczek and Crump, 2009), and an exciting new structure of the herpes simplex virus pUL7: pUL51 complex, which is required for efficient virion assembly, reveals that the N-terminal region of pUL51 adopts a CHMP4B-like fold that may function as a viral ESCRT-III-like protein (Butt *et al.*, 2020). Following membrane fission, the enveloped virions end up inside intracellular vesicles and are released into the extracellular space when the vesicles fuse with the plasma membrane.

In a related process, flaviviruses and hepaciviruses are released through the secretory pathway (Chatel-Chaix and Bartenschlager, 2014; Falcon *et al.*, 2017; Gerold *et al.*, 2017). Both genera of viruses induce extensive remodeling of endoplasmic reticulum (ER) membranes to form replication compartments. These vesicle-like structures remain connected to the ER, enclose viral proteins and the viral genome, and serve as protected compartments where almost all steps of the life cycle are carried out. Assembled nucleocapsids then bud into the ER lumen and are released through the secretory pathway.

Quasi-enveloped viruses

Historically, viruses have been divided into enveloped and non-enveloped classes based on the presence or absence of a host-derived membrane envelope, and it was thought that non-enveloped viruses were released exclusively by host cell lysis. This simple paradigm has now been overturned by studies showing that many different classes of non-enveloped viruses can acquire host-derived lipid envelopes and exit cells within vesicles. These extracellular vesicles resemble exosomes, and viruses that use this egress method are termed “quasi-enveloped” (Feng *et al.*, 2014).

The picornavirus hepatitis A virus (HAV) was the first virus definitively shown to be quasi-enveloped (Feng *et al.*, 2013), and HAV still serves as a paradigm for the process. Quasi-enveloped HAV particles (eHAV) are released in exosome-like vesicles that typically contain 1–4 particles per vesicle. These vesicles are formed when HAV capsids bud into endosomes in an ESCRT-dependent manner. To promote budding, the VP2 capsid protein recruits ALIX, apparently using tandem YPX₃L domains that become buried in the fully assembled virion (Gonzalez-Lopez *et al.*, 2018; McKnight *et al.*, 2017). Virion-containing multivesicular bodies then fuse with the plasma membrane and release eHAV particles into the extracellular space. There is now good evidence that this is the primary mode of HAV release from hepatocytes *in vivo*, and that HAV circulates in the blood exclusively within small vesicles (Feng *et al.*, 2013). HAV can then shed its envelope in the biliary tract, which produces a non-enveloped particle that may be more stable in harsher environmental conditions (Feng *et al.*, 2014). Importantly, HAV is highly infectious in both its enveloped and non-enveloped states.

The capacity for quasi-envelopment has since been described for several other viruses that were traditionally considered to be non-enveloped, including many other picornaviruses (Chen *et al.*, 2015; Mutsafi and Altan-Bonnet, 2018), Hepatitis E virus (Qi *et al.*, 2015), rotaviruses and noroviruses (Santiana *et al.*, 2018). Furthermore, some picornaviruses, including poliovirus and coxsackievirus, differ from the HAV paradigm in that they form quasi-enveloped virions by subverting the autophagy pathway. In these cases, double-membraned autophagosomes engulf multiple naked viral particles, which then release quasi-enveloped viruses when the outer autophagosomal membrane fuses with the plasma membrane (Mutsafi and Altan-Bonnet, 2018; Bird *et al.*, 2014). A final variation on this theme is the exosomal transfer of viral nucleic acids between cells, which apparently can, in some viruses like hepatitis C, spread productive infections without requiring full viral assembly (Ramakrishnaiah *et al.*, 2013; Bukong *et al.*, 2014).

The membrane appears to perform several important functions for quasi-enveloped viruses, including protecting the capsid from antibody-mediated neutralization (Feng *et al.*, 2013), and clustering together of multiple virions so that they can enter target cells as a swarm or “quasi-species” that can cooperate genetically through cross-complementation. The later activity may be most important for enteroviruses, whose larger vesicles can each contain tens or even hundreds of viral particles (Chen *et al.*, 2015; Santiana *et al.*, 2018).

The outer membranes of quasi-enveloped viruses lack viral glycoproteins, and therefore cannot fuse with target cell membranes. Instead, eHAV particles are taken up into the host cells by endocytosis and trafficked toward the lysosome, where the membrane is degraded, and the released naked virions can cross into the cytoplasm by disrupting the endolysosomal membrane (Rivera-Serrano *et al.*, 2019).

Maturation

Most enveloped viruses undergo additional maturation steps during and after budding. Before maturation, the virion functions as an assembly machinery that can package components and leave the producer cell. Conformational changes, typically triggered by proteolytic cleavage or pH changes, then convert the virion into a particle that is capable of entering and replicating in a new target cell (Veesler and Johnson, 2012; Steven *et al.*, 2005).

In the case of HIV-1, the viral protease (PR) is activated by autoproteolysis as the virus assembles and buds, and it cleaves the Gag polyprotein at five different sites, producing three new proteins (MA, CA and NC) and three smaller peptides (SP1, SP2, and p6). Gag processing drives a series of major rearrangements in which the CA protein forms a conical internal capsid that surrounds viral RNA in complex with NC protein and viral enzymes. Gag cleavage is a sequential, ordered process, and each processing event appears to perform a different function. Cleavage at the SP1-NC junction releases the NC-RNA complex to condense to the center of the virion, cleavage at the MA-CA junction promotes folding of the CA N-terminus into a β -hairpin that will ultimately form an NTP-permeable pore in the assembled capsid, and cleavage at the CA-SP1 junction destabilizes the immature lattice and promotes formation of the mature capsid lattice. The NC-SP2 and SP2-p6 cleavages are also required for infectivity, as is cleavage of the longer Gag-Pol polyprotein, which liberates the viral enzymes. The mature conical capsid is a fullerene cone, with a curved hexagonal lattice comprising CA hexamers, and the cone ends closed through the incorporation of 12 CA pentamers. CA hexamers are stabilized by binding IP₆ (Dick *et al.*, 2018; Mallery *et al.*, 2018), and differential placement of the hexamers and pentamers produces a variety of related capsid structures that each differ slightly in length and shape (Sundquist and Krausslich, 2012; Mattei *et al.*, 2016; Freed, 2015).

Viral glycoproteins and their fusion peptides that enable entry into target cells must also typically be proteolytically processed to be functional. For example, the HIV-1 Env glycoprotein is synthesized as a polyprotein precursor (gp160), which is inserted into the endoplasmic reticulum membrane co-translationally. Env is glycosylated and then proteolytically cleaved by the host Golgi-associated protease furin as it traffics through the secretory pathway, producing the mature surface gp120 and transmembrane gp41 glycoprotein subunits, which remain non-covalently associated as heterotrimeric spikes. Proteolytic processing exposes the fusion peptide at the gp41 N-terminus and is required for fusogenic activity (Checkley *et al.*, 2011).

In other viruses, proteolytic activation of viral fusion proteins can occur following entry into the target cell. Activation of the EBOV glycoprotein is a particularly well-understood case. EBOV particles associate with the host cell surface by interactions with host receptors that bind to glycans on the viral glycoprotein GP and phosphatidylserine in the viral envelope. After internalization through macropinocytosis, endosomal cysteine proteases such as cathepsins L and B proteolytically process GP to remove a mucin-like subdomain and the glycan cap and expose the receptor-binding site (RBS). The RBS binds the late endosomal/lysosomal protein NPC1, which induces a conformational change in GP, insertion of a fusion loop into the endosomal membrane, fusion of viral and endosomal membranes and release of the nucleocapsid into the cytoplasm (Lee and Saphire, 2009; Carette *et al.*, 2011; Cote *et al.*, 2011; Gong *et al.*, 2016; Wang *et al.*, 2016).

Cell-to-Cell Transmission

After budding, viruses can spread in two different ways; through cell-free transmission and cell-to-cell transmission. Cell-free virions diffuse freely through the extracellular space, and even between organisms, before entering target cells. This process can promote dissemination over long distances, to new tissues, and between hosts. However, untargeted diffusion through aqueous media is relatively inefficient, and free viruses are susceptible to immune recognition. In contrast, viral spread through direct sites of cell-to-cell contact increases transmission efficiency and can help evade antibody recognition. To promote cell-to-cell transmission, viruses often subvert cellular structures that are normally used for cell-cell communication or cargo transfer.

Retroviruses such as HIV-1 and MLV actively promote the formation of adhesive structures between donor and target cells. These stable contact sites are termed virological synapses because they resemble the immunological synapses that mediate antigen presentation, and even employ some of the same molecular components. Virological synapse formation requires interactions between the viral glycoprotein on the donor cell and its cognate target cell receptor. Coreceptors and adhesion molecules are then recruited to stabilize and further organize these contact sites, and the producer cell cytoskeleton is repolarized towards the synapse to promote directional viral assembly and release. Virions bud directly into the intersynaptic space and are transferred efficiently to the closely opposed target cell (Agosto *et al.*, 2015; Bracq *et al.*, 2018; Dufloo *et al.*, 2018; Nejmeddine and Bangham, 2010).

Other viruses hijack existing cell-cell contacts for transmission. For example, neurotropic viruses such as herpesviruses and rabies viruses are transported along axons and spread across synaptic contacts (Koyuncu *et al.*, 2013). Viruses can also achieve targeted release by exploiting membrane protrusions such as nanotubes and filopodia, which normally transmit information and cargoes between cells (Agosto *et al.*, 2015; Bracq *et al.*, 2018; Dufloo *et al.*, 2018).

Conclusions

The principles of enveloped virus budding are remarkably conserved between different virus families, presumably owing to evolutionary history and common functional requirements. Viral egress is typically orchestrated by multifunctional structural proteins that recruit components, assemble the virion, bend host membranes, and facilitate membrane fission. In many, but not all cases, the cellular ESCRT pathway is recruited to mediate the final membrane fission step. Recent studies have also revealed that many traditional “non-enveloped” viruses can be released within vesicles as quasi-enveloped viruses and that viruses frequently alter cellular pathways to promote directional release and synapse formation.

Although the general strategies for enveloped virus egress are increasingly well understood, important challenges remain, including characterizing the release mechanisms of ESCRT-independent viruses, the biology and entry mechanisms of quasi-enveloped viruses, and the molecular mechanisms and pathogenesis associated with cell-to-cell viral spread. These and other

advances will help reveal the best approaches for inhibiting virus release for therapeutic benefit and harnessing release activities in new systems that can be used to deliver biomolecular cargoes into target cells in vivo.

Acknowledgments

We thank Moona Huttunen and Jörg Votteler for helpful comments on the manuscript, and Janet Iwasa for creating the figures. This work was funded by NIH R37 AI051174. We apologize to our many colleagues whose primary literature references we were unable to cite owing to space limitations.

References

- Agosto, L.M., Uchil, P.D., Mothes, W., 2015. HIV cell-to-cell transmission: Effects on pathogenesis and antiretroviral therapy. *Trends in Microbiology* 23 (5), 289–295.
- Alfadhli, A., *et al.*, 2011. HIV-1 matrix protein binding to RNA. *Journal of Molecular Biology* 410 (4), 653–666.
- Arii, J., *et al.*, 2018. ESCRT-III mediates budding across the inner nuclear membrane and regulates its integrity. *Nature Communications* 9 (1), 3379.
- Ashley, J., *et al.*, 2018. Retrovirus-like gag protein arc1 binds RNA and traffics across synaptic boutons. *Cell* 172 (1–2), 262–274.
- Bache, K.G., *et al.*, 2003. Hrs regulates multivesicular body formation via ESCRT recruitment to endosomes. *Journal of Cell Biology* 162 (3), 435–442.
- Baietti, M.F., *et al.*, 2012. Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nature Cell Biology* 14 (7), 677–685.
- Banjade, S., Tang, S., Emr, S.D., 2019. Genetic and Biochemical Analyses of Yeast ESCRT, *Methods in Molecular Biology* 1998, 105–116.
- Bigalke, J.M., Heldwein, E.E., 2015. The great (nuclear) escape: New Insights into the role of the nuclear egress complex of herpesviruses. *Journal of Virology* 89 (18), 9150–9153.
- Bird, S.W., *et al.*, 2014. Nonlytic viral spread enhanced by autophagy components. *Proceedings of the National Academy of Sciences of the United States of America* 111 (36), 13081–13086.
- Blondot, M.L., Bruss, V., Kann, M., 2016. Intracellular transport and egress of hepatitis B virus. *Journal of Hepatology* 64 (1 Suppl), S49–S59.
- Boonyaratnakornkit, J., *et al.*, 2013. Alix serves as an adaptor that allows human parainfluenza virus type 1 to interact with the host cell ESCRT system. *PLoS One* 8 (3), e59462.
- Bracq, L., *et al.*, 2018. Mechanisms for cell-to-cell transmission of HIV-1. *Frontiers in Immunology* 9, 260.
- Brown, R.S., Wan, J.J., Kielian, M., 2018. The alphavirus exit pathway: What we know and what we wish we knew. *Viruses* 10 (2), 1–12.
- Brown, J.D., *et al.*, 2020. Structural basis for transcriptional start site control of HIV-1 RNA fate. *Science* 368 (6489), 413–417.
- Bruce, E.A., *et al.*, 2009. Budding of filamentous and non-filamentous influenza A virus occurs via a VPS4 and VPS28-independent pathway. *Virology* 390 (2), 268–278.
- Bukong, T.N., *et al.*, 2014. Exosomes from hepatitis C infected patients transmit HCV infection and contain replication competent viral RNA in complex with Ago2-miR122-HSP90. *PLoS Pathogens* 10 (10), e1004424.
- Butt, B.G., Owen, D.J., Jeffries, C.M., *et al.*, 2020. Insights into herpesvirus assembly from the structure of the pUL7:pUL51 complex. *Elife* 9, e53789.
- Calistri, A., *et al.*, 2007. Intracellular trafficking and maturation of herpes simplex virus type 1 gB and virus egress require functional biogenesis of multivesicular bodies. *Journal of Virology* 81 (20), 11468–11478.
- Carette, J.E., *et al.*, 2011. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* 477 (7364), 340–343.
- Carlton, J.G., Martin-Serrano, J., 2007. Parallels between cytokinesis and retroviral budding: A role for the ESCRT machinery. *Science* 316 (5833), 1908–1912.
- Chatel-Chaix, L., *et al.*, 2007. The host protein Stau1 participates in human immunodeficiency virus type 1 assembly in live cells by influencing pr55Gag multimerization. *Journal of Virology* 81 (12), 6216–6230.
- Chatel-Chaix, L., Bartenschlager, R., 2014. Dengue virus- and hepatitis C virus-induced replication and assembly compartments: The enemy inside—caught in the web. *Journal of Virology* 88 (11), 5907–5911.
- Checkley, M.A., Luttmann, B.G., Freed, E.O., 2011. HIV-1 envelope glycoprotein biosynthesis, trafficking, and incorporation. *Journal of Molecular Biology* 410 (4), 582–608.
- Chen, B.J., *et al.*, 2007. Influenza virus hemagglutinin and neuraminidase, but not the matrix protein, are required for assembly and budding of plasmid-derived virus-like particles. *Journal of Virology* 81 (13), 7111–7123.
- Chen, Y.H., *et al.*, 2015. Phosphatidylserine vesicles enable efficient en bloc transmission of enteroviruses. *Cell* 160 (4), 619–630.
- Chlanda, P., Zimmerberg, J., 2016. Protein-lipid interactions critical to replication of the influenza A virus. *FEBS Letters* 590 (13), 1940–1954.
- Christ, L., *et al.*, 2017. Cellular functions and molecular mechanisms of the ESCRT membrane-scission machinery. *Trends in Biochemical Sciences* 42 (1), 42–56.
- Chukkappalli, V., Oh, S.J., Ono, A., 2010. Opposing mechanisms involving RNA and lipids regulate HIV-1 Gag membrane binding through the highly basic region of the matrix domain. *Proceedings of the National Academy of Sciences of the United States of America* 107 (4), 1600–1605.
- Chukkappalli, V., Ono, A., 2011. Molecular determinants that regulate plasma membrane association of HIV-1 Gag. *Journal of Molecular Biology* 410 (4), 512–524.
- Chung, H.Y., *et al.*, 2008. NEDD4L overexpression rescues the release and infectivity of human immunodeficiency virus type 1 constructs lacking PTAP and YPX late domains. *Journal of Virology* 82 (10), 4884–4897.
- Cifuentes-Munoz, N., Salazar-Quiroz, N., Tischler, N.D., 2014. Hantavirus Gn and Gc envelope glycoproteins: Key structural units for virus cell entry and virus assembly. *Viruses* 6 (4), 1801–1822.
- Cote, M., *et al.*, 2011. Small molecule inhibitors reveal Niemann-Pick C1 is essential for Ebola virus infection. *Nature* 477 (7364), 344–348.
- Crump, C.M., Yates, C., Minson, T., 2007. Herpes simplex virus type 1 cytoplasmic envelopment requires functional Vps4. *Journal of Virology* 81 (14), 7380–7387.
- Demirov, D.G., *et al.*, 2002. Overexpression of the N-terminal domain of TSG101 inhibits HIV-1 budding by blocking late domain function. *Proceedings of the National Academy of Sciences of the United States of America* 99 (2), 955–960.
- Dick, R.A., *et al.*, 2018. Inositol phosphates are assembly co-factors for HIV-1. *Nature* 560 (7719), 509–512.
- Dores, M.R., *et al.*, 2012. ALIX binds a YPX(3)I motif of the GPCR PAR1 and mediates ubiquitin-independent ESCRT-III/MVB sorting. *Journal of Cell Biology* 197 (3), 407–419.
- Dubois, N., *et al.*, 2018. Retroviral RNA dimerization: From structure to functions. *Frontiers in Microbiology* 9, 527.
- Dufloo, J., Bruel, T., Schwartz, O., 2018. HIV-1 cell-to-cell transmission and broadly neutralizing antibodies. *Retrovirology* 15 (1), 51.
- Falcon, V., *et al.*, 2017. Ultrastructural and biochemical basis for hepatitis C virus morphogenesis. *Virus Genes* 53 (2), 151–164.
- Feng, Z., *et al.*, 2013. A pathogenic picornavirus acquires an envelope by hijacking cellular membranes. *Nature* 496 (7445), 367–371.
- Feng, Z., *et al.*, 2014. Naked viruses that aren't always naked: Quasi-enveloped agents of acute hepatitis. *Annual Review of Virology* 1 (1), 539–560.
- Fradkin, L.G., Budnik, V., 2016. This bud's for you: Mechanisms of cellular nucleocytoplasmic trafficking via nuclear envelope budding. *Current Opinion in Cell Biology* 41, 125–131.
- Freed, E.O., 2015. HIV-1 assembly, release and maturation. *Nature Reviews Microbiology* 13 (8), 484–496.

- Gaines, C.R., *et al.*, 2018. HIV-1 Matrix Protein Interactions with tRNA: Implications for Membrane Targeting. *Journal of Molecular Biology* 430 (14), 2113–2127.
- Garrus, J.E., *et al.*, 2001. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. *Cell* 107 (1), 55–65.
- Gerold, G., *et al.*, 2017. Protein Interactions during the flavivirus and hepatitis virus life cycle. *Molecular & Cellular Proteomics* 16 (4 suppl 1), S75–S91.
- Gong, X., *et al.*, 2016. Structural insights into the Niemann-Pick C1 (NPC1)-mediated cholesterol transfer and ebola infection. *Cell* 165 (6), 1467–1478.
- Gonzalez-Lopez, O., Rivera-Serrano, E.E., Hu, F., *et al.*, 2018. Redundant late domain functions of tandem VP2 YPX₃L motifs in nonlytic cellular egress of quasi-enveloped hepatitis A virus. *Journal of Virology* 92 (23), e01308–18.
- Gottlinger, H.G., *et al.*, 1991. Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proceedings of the National Academy of Sciences of the United States of America* 88 (8), 3195–3199.
- Hanson, P.I., Cashikar, A., 2012. Multivesicular body morphogenesis. *Annual Review of Cell and Developmental Biology* 28, 337–362.
- Henne, W.M., Stenmark, H., Emr, S.D., 2013. Molecular mechanisms of the membrane sculpting ESCRT pathway. *Cold Spring Harbor Perspectives in Biology* 5 (9), Hepojoki, J., *et al.*, 2012. Hantavirus structure – Molecular interactions behind the scene. *Journal of General Virology* 93 (Pt 8), 1631–1644.
- Hill, C.P., *et al.*, 1996. Crystal structures of the trimeric human immunodeficiency virus type 1 matrix protein: Implications for membrane association and assembly. *Proceedings of the National Academy of Sciences of the United States of America* 93 (7), 3099–3104.
- Huang, M., *et al.*, 1995. p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *Journal of Virology* 69 (11), 6810–6818.
- Hurley, J.H., Cada, A.K., 2018. Inside job: How the ESCRTs release HIV-1 from infected cells. *Biochemical Society Transactions* 46 (5), 1029–1036.
- Hütter, S., Zurnic, I., Lindemann, D., 2013. Foamy virus budding and release. *Viruses* 5 (4), 1075–1098.
- Issman-Zecharya, N., Schuldiner, O., 2014. The PI3K class III complex promotes axon pruning by downregulating a Ptc-derived signal via endosome-lysosomal degradation. *Developmental Cell* 31 (4), 461–473.
- Jimenez, A.J., *et al.*, 2014. ESCRT machinery is required for plasma membrane repair. *Science* 343 (6174), 1247136.
- Jouvenet, N., Simon, S.M., Bieniasz, P.D., 2009. Imaging the interaction of HIV-1 genomes and gag during assembly of individual viral particles. *Proceedings of the National Academy of Sciences of the United States of America* 106 (45), 19114–19119.
- Jouvenet, N., *et al.*, 2011. Dynamics of ESCRT protein recruitment during retroviral assembly. *Nature Cell Biology* 13 (4), 394–401.
- Kaddis Maldonado, R.J., Parent, L.J., 2016. Orchestrating the selection and packaging of genomic RNA by retroviruses: An ensemble of viral and host factors. *Viruses* 8 (9), Keane, S.C., *et al.*, 2015. Structure of the HIV-1 RNA packaging signal. *Science* 348 (6237), 917–921.
- Kondo, E., *et al.*, 1995. The p6gag domain of human immunodeficiency virus type 1 is sufficient for the incorporation of Vpr into heterologous viral particles. *Journal of Virology* 69 (5), 2759–2764.
- Koyuncu, O.O., Hogue, I.B., Enquist, L.W., 2013. Virus infections in the nervous system. *Cell Host & Microbe* 13 (4), 379–393.
- Kutluay, S.B., Bieniasz, P.D., 2010. Analysis of the initiating events in HIV-1 particle assembly and genome packaging. *PLOS Pathogens* 6 (11), e1001200.
- Kutluay, S.B., *et al.*, 2014. Global changes in the RNA binding specificity of HIV-1 gag regulate virion genesis. *Cell* 159 (5), 1096–1109.
- Kuzembayeva, M., *et al.*, 2014. Life of psi: How full-length HIV-1 RNAs become packaged genomes in the viral particles. *Virology* 454–455, 362–370.
- Lee, C.P., *et al.*, 2012. The ESCRT machinery is recruited by the viral BFRF1 protein to the nucleus-associated membrane for the maturation of Epstein-Barr Virus. *PLOS Pathogens* 8 (9), e1002904.
- Lee, J.E., Saphire, E.O., 2009. Ebolavirus glycoprotein structure and mechanism of entry. *Future Virology* 4 (6), 621–635.
- Lingappa, J.R., *et al.*, 2014. How HIV-1 gag assembles in cells: Putting together pieces of the puzzle. *Virus Research* 193, 89–107.
- Lippincott-Schwartz, J., Freed, E.O., van Engelenburg, S.B., 2017. A consensus view of ESCRT-mediated human immunodeficiency virus type 1 abscission. *Annual Review of Virology* 4 (1), 309–325.
- Liu, L., *et al.*, 2014. From crescent to mature virion: Vaccinia virus assembly and maturation. *Viruses* 6 (10), 3787–3808.
- Loncle, N., *et al.*, 2015. An ESCRT module is required for neuron pruning. *Scientific Reports* 5, 8461.
- Lv, Y., Zhou, S., Gao, S., Deng, H., 2019. Remodeling of host membranes during herpesvirus assembly and egress. *Protein Cell* 10 (5), 315–326.
- Mallery, D.L., Márquez, C.L., McEwan, W.A., *et al.*, 2018. IP6 is an HIV pocket factor that prevents capsid collapse and promotes DNA synthesis. *eLife* 7, e35335.
- Martin-Serrano, J., Zang, T., Bieniasz, P.D., 2001. HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. *Nature Medicine* 7 (12), 1313–1319.
- Martin-Serrano, J., *et al.*, 2003. Divergent retroviral late-budding domains recruit vacuolar protein sorting factors by using alternative adaptor proteins. *Proceedings of the National Academy of Sciences of the United States of America* 100 (21), 12414–12419.
- Martyna, A., Rossman, J., 2014. Alterations of membrane curvature during influenza virus budding. *Biochemical Society Transactions* 42 (5), 1425–1428.
- Mattei, S., Schur, F.K., Briggs, J.A., 2016. Retrovirus maturation—an extraordinary structural transformation. *Current Opinion in Virology* 18, 27–35.
- McCullough, J., Frost, A., Sundquist, W.I., 2018. Structures, functions, and dynamics of ESCRT-III/Vps4 membrane remodeling and fission complexes. *Annual Review of Cell and Developmental Biology* 34, 85–109.
- McKnight, K.L., *et al.*, 2017. Protein composition of the hepatitis A virus quasi-envelope. *Proceedings of the National Academy of Sciences of the United States of America* 114 (25), 6587–6592.
- Meng, B., Lever, A.M., 2013. Wrapping up the bad news: HIV assembly and release. *Retrovirology* 10, 5.
- Mercenne, G., Alam, S.L., Arij, J., Lalonde, M.S., Sundquist, W.I., 2015. Angiomotin functions in HIV-1 assembly and budding. *eLife* 4, e03778.
- Morita, E., *et al.*, 2007. Human ESCRT and ALIX proteins interact with proteins of the midbody and function in cytokinesis. *The EMBO Journal* 26 (19), 4215–4227.
- Morita, E., *et al.*, 2011. ESCRT-III protein requirements for HIV-1 budding. *Cell Host & Microbe* 9 (3), 235–242.
- Müllers, E., 2013. The foamy virus gag proteins: What makes them different? *Viruses* 5 (4), 1023–1041.
- Mutsaers, Y., Altan-Bonnet, N., 2018. Enterovirus transmission by secretory autophagy. *Viruses* 10 (3),
- Muyangwa, M., *et al.*, 2015. Hantaviral proteins: Structure, functions, and role in hantavirus infection. *Frontiers in Microbiology* 6, 1326.
- El Najjar, F., Schmitt, A.P., Dutch, R.E., 2014. Paramyxovirus glycoprotein incorporation, assembly and budding: A three way dance for infectious particle production. *Viruses* 6 (8), 3019–3054.
- Nejmeddine, M., Bangham, C.R., 2010. The HTLV-1 virological synapse. *Viruses* 2 (7), 1427–1447.
- Nishimura, T., Morone, N., Suetsugu, S., 2018. Membrane re-modelling by BAR domain superfamily proteins via molecular and non-molecular factors. *Biochemical Society Transactions* 46 (2), 379–389.
- Olmos, Y., *et al.*, 2015. ESCRT-III controls nuclear envelope reformation. *Nature* 522 (7555), 236–239.
- Ono, A., Freed, E.O., 2001. Plasma membrane rafts play a critical role in HIV-1 assembly and release. *Proceedings of the National Academy of Sciences of the United States of America* 98 (24), 13925–13930.
- Owen, D.J., Crump, C.M., Graham, S.C., 2015. Tegument assembly and secondary envelopment of alphaherpesviruses. *Viruses* 7 (9), 5084–5114.
- Parent, L.J., *et al.*, 1995. Positionally independent and exchangeable late budding functions of the Rous sarcoma virus and human immunodeficiency virus gag proteins. *Journal of Virology* 69 (9), 5455–5460.
- Pastuzyn, E.D., *et al.*, 2018. The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. *Cell* 172 (1–2), 275–288.
- Patient, R., Hourieux, C., Roingeard, P., 2009. Morphogenesis of hepatitis B virus and its subviral envelope particles. *Cellular Microbiology* 11 (11), 1561–1570.
- Patnaik, A., Chau, V., Wills, J.W., 2000. Ubiquitin is part of the retrovirus budding machinery. *Proceedings of the National Academy of Sciences of the United States of America* 97 (24), 13069–13074.

- Pawliczek, T., Crump, C.M., 2009. Herpes simplex virus type 1 production requires a functional ESCRT-III complex but is independent of TSG101 and ALIX expression. *Journal of Virology* 83 (21), 11254–11264.
- Pei, Z., Bai, Y., Schmitt, A.P., 2010. PIV5 M protein interaction with host protein angiomin-like 1. *Virology* 397 (1), 155–166.
- Pezeshkian, N., Groves, N.S., van Engelenburg, S.B., 2019. Single-molecule imaging of HIV-1 envelope glycoprotein dynamics and Gag lattice association exposes determinants responsible for virus incorporation. *Proceedings of the National Academy of Sciences of the United States of America* 116 (50), 25269–25277.
- Prange, R., 2012. Host factors involved in hepatitis B virus maturation, assembly, and egress. *Medical Microbiology and Immunology* 201 (4), 449–461.
- Qi, Y., *et al.*, 2015. Hepatitis E virus produced from cell culture has a lipid envelope. *PLoS One* 10 (7), e0132503.
- Ramakrishnaiah, V., *et al.*, 2013. Exosome-mediated transmission of hepatitis C virus between human hepatoma Huh7.5 cells. *Proceedings of the National Academy of Sciences of the United States of America* 110 (32), 13109–13113.
- Ray, G., Schmitt, P.T., Schmitt, A.P., 2019. Angiomin-like 1 links paramyxovirus M proteins to NEDD4 family ubiquitin ligases. *Viruses* 11 (2), 1–12.
- Reed, J.C., *et al.*, 2012. HIV-1 gag co-opts a cellular complex containing DDX6, a helicase that facilitates capsid assembly. *Journal of Cell Biology* 198 (3), 439–456.
- Rivera-Serrano, E.E., *et al.*, 2019. Cellular entry and uncoating of naked and quasi-enveloped human hepatoviruses. *eLife* 8, e43983.
- Roberts, K.L., Smith, G.L., 2008. Vaccinia virus morphogenesis and dissemination. *Trends in Microbiology* 16 (10), 472–479.
- Rossman, J.S., Lamb, R.A., 2011. Influenza virus assembly and budding. *Virology* 411 (2), 229–236.
- Rossman, J.S., *et al.*, 2010. Influenza virus M2 protein mediates ESCRT-independent membrane scission. *Cell* 142 (6), 902–913.
- Rotin, D., Kumar, S., 2009. Physiological functions of the HECT family of ubiquitin ligases. *Nature Reviews Molecular Cell Biology* 10 (6), 398–409.
- Salditt, A., *et al.*, 2010. Measles virus M protein-driven particle production does not involve the endosomal sorting complex required for transport (ESCRT) system. *Journal of General Virology* 91 (Pt 6), 1464–1472.
- Sandrin, V., Sundquist, W.I., 2013. ESCRT requirements for EIAV budding. *Retrovirology* 10, 104.
- Santiana, M., *et al.*, 2018. Vesicle-cloaked virus clusters are optimal units for inter-organismal viral transmission. *Cell Host & Microbe* 24 (2), 208–220.
- Scheffer, L.L., *et al.*, 2014. Mechanism of Ca(2+)-triggered ESCRT assembly and regulation of cell membrane repair. *Nature Communications* 5, 5646.
- Schur, F.K., *et al.*, 2016. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* 353 (6298), 506–508.
- von Schwedler, U.K., *et al.*, 2003. The protein network of HIV budding. *Cell* 114 (6), 701–713.
- Scourfield, E.J., Martin-Serrano, J., 2017. Growing functions of the ESCRT machinery in cell biology and viral replication. *Biochemical Society Transactions* 45 (3), 613–634.
- Shields, S.B., Piper, R.C., 2011. How ubiquitin functions with ESCRTs. *Traffic* 12 (10), 1306–1317.
- Skowrya, M.L., *et al.*, 2018. Triggered recruitment of ESCRT machinery promotes endolysosomal repair. *Science* 360, 6384.
- Smith, A.J., *et al.*, 1993. Requirements for incorporation of Pr160gag-pol from human immunodeficiency virus type 1 into virus-like particles. *Journal of Virology* 67 (4), 2266–2275.
- Steven, A.C., *et al.*, 2005. Virus maturation: dynamics and mechanism of a stabilizing structural transition that leads to infectivity. *Current Opinion in Structural Biology* 15 (2), 227–236.
- Strack, B., *et al.*, 2000. A role for ubiquitin ligase recruitment in retrovirus release. *Proceedings of the National Academy of Sciences of the United States of America* 97 (24), 13063–13068.
- Strack, B., *et al.*, 2003. AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. *Cell* 114 (6), 689–699.
- Sundquist, W.I., Krausslich, H.G., 2012. HIV-1 assembly, budding, and maturation. *Cold Spring Harbor Perspectives in Medicine* 2 (7), a006924.
- Swanstrom, R., Wills, J.W., 1997. Synthesis, assembly, and processing of viral proteins. In: Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), *Retroviruses*. NY: Cold Spring Harbor.
- Takahashi, Y., *et al.*, 2018. An autophagy assay reveals the ESCRT-III component CHMP2A as a regulator of phagophore closure. *Nature Communications* 9 (1), 2855.
- Tang, C., *et al.*, 2004. Entropic switch regulates myristate exposure in the HIV-1 matrix protein. *Proceedings of the National Academy of Sciences of the United States of America* 101 (2), 517–522.
- Tedbury, P.R., Freed, E.O., 2014. The role of matrix in HIV-1 envelope glycoprotein incorporation. *Trends in Microbiology* 22 (7), 372–378.
- Tedbury, P.R., *et al.*, 2016. Biochemical evidence of a role for matrix trimerization in HIV-1 envelope glycoprotein incorporation. *Proceedings of the National Academy of Sciences of the United States of America* 113 (2), E182–E190.
- Usami, Y., *et al.*, 2008. Efficient and specific rescue of human immunodeficiency virus type 1 budding defects by a Nedd4-like ubiquitin ligase. *Journal of Virology* 82 (10), 4898–4907.
- Utey, T.J., *et al.*, 2008. Respiratory syncytial virus uses a Vps4-independent budding mechanism controlled by Rab11-FIP2. *Proceedings of the National Academy of Sciences of the United States of America* 105 (29), 10209–10214.
- Veesler, D., Johnson, J.E., 2012. Virus maturation. *Annual Review of Biophysics* 41, 473–496.
- VerPlank, L., *et al.*, 2001. Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). *Proceedings of the National Academy of Sciences of the United States of America* 98 (14), 7724–7729.
- Vietri, M., *et al.*, 2015. Spastin and ESCRT-III coordinate mitotic spindle disassembly and nuclear envelope sealing. *Nature* 522 (7555), 231–235.
- Votteler, J., Sundquist, W.I., 2013. Virus budding and the ESCRT pathway. *Cell Host & Microbe* 14 (3), 232–241.
- Wagner, J.M., Zadrozny, K.K., Chruszowicz, J., *et al.*, 2016. Crystal structure of an HIV assembly and maturation switch. *eLife* 5, e17063.
- Wang, H., *et al.*, 2016. Ebola viral glycoprotein bound to its endosomal receptor Niemann-Pick C1. *Cell* 164 (1–2), 258–268.
- Watanabe, R., Lamb, R.A., 2010. Influenza virus budding does not require a functional AAA + ATPase, VPS4. *Virus Research* 153 (1), 58–63.
- Weissenhorn, W., *et al.*, 2013. How to get out: ssRNA enveloped viruses and membrane fission. *Current Opinion in Virology* 3 (2), 159–167.
- Xiang, Y., *et al.*, 1996. Fine mapping and characterization of the Rous sarcoma virus Pr76gag late assembly domain. *Journal of Virology* 70 (8), 5695–5700.
- Yuan, B., *et al.*, 2000. Infectivity of Moloney murine leukemia virus defective in late assembly events is restored by late assembly domains of other retroviruses. *Journal of Virology* 74 (16), 7250–7260.
- Zhadina, M., *et al.*, 2007. Ubiquitin-dependent virus particle budding without viral protein ubiquitination. *Proceedings of the National Academy of Sciences of the United States of America* 104 (50), 20031–20036.
- Zhang, H., *et al.*, 2014. Endocytic pathways downregulate the L1-type cell adhesion molecule neuroglian to promote dendrite pruning in drosophila. *Developmental Cell* 30 (4), 463–478.
- Zhou, F., *et al.*, 2019. Rab5-dependent autophagosome closure by ESCRT. *Journal of Cell Biology* 218 (6), 1908–1927.
- Zimmerman, C., *et al.*, 2002. Identification of a host protein essential for assembly of immature HIV-1 capsids. *Nature* 415 (6867), 88–92.

Vesicle-Mediated Transcytosis and Export of Viruses

Efraín E Rivera-Serrano and Stanley M Lemon, Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

© 2021 Published by Elsevier Ltd.

Glossary

Apical membrane Surface of the plasma membrane of polarized cells that faces inward to the lumen.

Autophagy A highly regulated process by which cytoplasmic contents become engulfed in membranes resulting in a double-membrane vesicle (autophagosome) that fuses with lysosomes for degradation of its cargo.

Basolateral membrane Surface of the plasma membrane that faces towards the interstitium and forms the basal and lateral sides of polarized cells.

Caveolae Microdomains of the plasma membrane characterized by invaginations of the lipid bilayer in a flask-like structure that is involved in endocytosis and signal transduction pathways.

Cell Polarity Asymmetric organization of cellular components including its plasma membrane, cytoskeleton, and/or organelles into functionally specialized domains. For epithelial and endothelial cells, this usually reflects polarization across the apical-basal axis.

Clathrin-coated pits Regions of the plasma membrane coated with clathrin on its cytoplasmic surface involved in receptor-mediated endocytosis and vesicle transport.

Dynamin Enzyme responsible for pinching off the cargo-loaded vesicles from the parent membrane during the early stages of endocytosis.

Endocytosis Process of cellular internalization of molecules through budding of vesicles derived from the plasma membrane into the cytoplasm.

Endothelial cell Cells that line the interior surface of all blood and lymphatic vessels.

Epithelial cell Cells that line the surface and cavities of tissues and organs and are involved in protection, absorption, secretion, and intercellular transport of molecules.

ESCRT Endosomal complexes required for transport, a series of multiprotein complexes that sort cargo for loading

into multivesicular endosomes (MVE), and mediate membrane abscission events in MVE and autophagosome formation.

Exocyst Octameric protein complex that mediates the tethering of secretory vesicles to the plasma membrane prior to fusion and release of cargo.

Exocytosis Process of release of material from the cytoplasm to the extracellular space mediated by intracellular vesicles.

Exosome A small (50–150 nm diameter) extracellular vesicle that shuttles protein or RNA cargo between cells and contributes to important intercellular communications and that originates within multivesicular endosomes.

Intraluminal vesicle (ILV) A cargo-laden vesicle located within a multivesicular endosome.

Macropinocytosis Actin-dependent type of endocytosis involved in the non-selective engulfment of macromolecules.

Microvesicle Large extracellular vesicle (> 150 nm diameter) shed from the plasma membrane of cells.

Multivesicular endosome (MVE) A large endosome containing multiple intraluminal vesicles (ILV) slated either for delivery to lysosomes for degradation or export as exosomes.

Polymeric immunoglobulin receptor (pIgR) Receptor that mediates the transcytosis of the soluble dimeric immunoglobulin A isoforms and immune complexes from the basolateral to the apical cell surface.

Retrotranscytosis Apical-to-basolateral retrograde transport of secretory immunoglobulin-antigen complexes from the lumen to the subepithelial space.

Transcytosis Vectorial transport of molecules enclosed within membrane-bound vesicles from one side of a polarized cell to the other.

Introduction

The transport of viruses through cells in membrane-limited vesicles plays an important role in many viral infections. Transcytosis describes the vectorial intracellular transport of macro- and micro-molecules (*i.e.*, cargo) within membrane-bound vesicles from one side of a polarized cell to the other (**Fig. 1**, left). The existence of this directional process of intracellular trafficking was first postulated by George Palade in the 1950s, but it was not until the late 1970s that the term ‘transcytosis’ was coined (**Tuma and Hubbard, 2003**). The identification of caveolae, originally termed ‘plasmalemma vesicles’, around the same time represented an important discovery that provided a mechanistic explanation for the movement of cargo through layers of cells in situations where there was minimal paracellular permeability. While these and other early studies described this phenomenon almost exclusively in endothelial cells lining the inner surface of the microvasculature, it is now known to occur in a variety of other polarized cell types.

From a cellular point of view, viruses undergoing transcytosis can be considered as macromolecular cargo capable of hijacking this normal mechanism of vesicular transport given the correct cues and context. From the virus point of view,

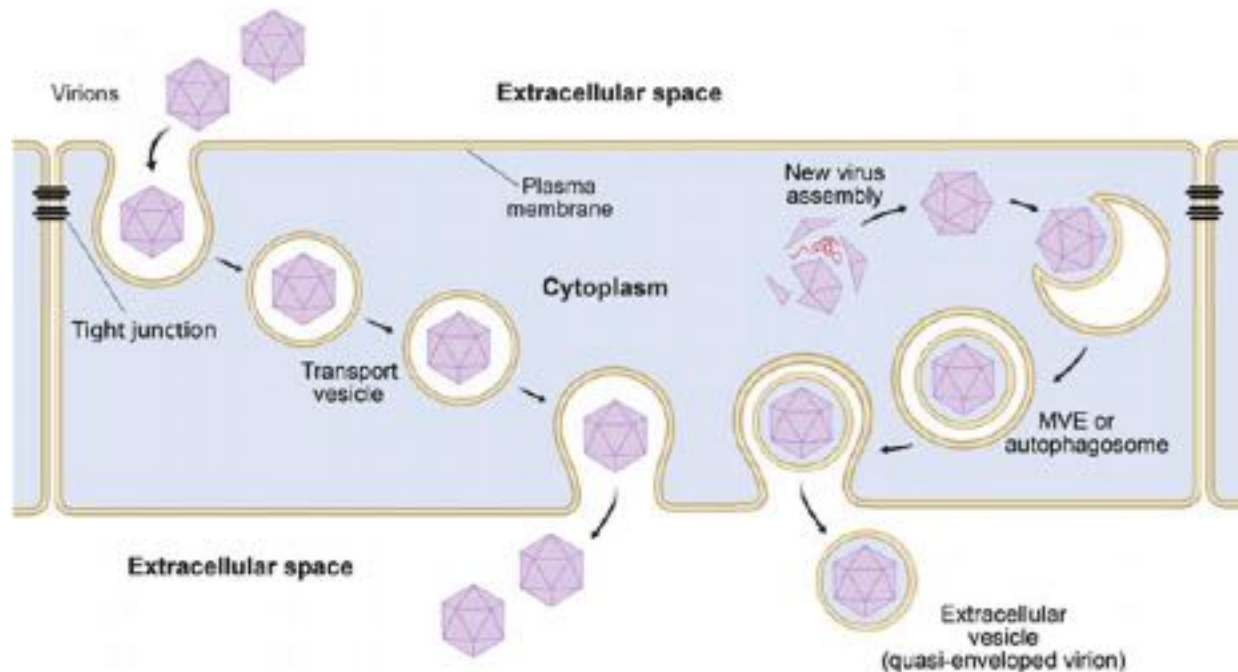


Fig. 1 Membrane topology in vesicle-mediated transcytosis and export of viruses. Transcytosis is displayed on the left. Viruses are internalized through endocytosis, transported intracellularly in membrane-bound vesicles, and secreted towards the extracellular space on the opposite side of a polarized cell. Traffic can be in either direction. Vesicle-mediated virus release is shown on the right. Newly assembling virus particles are recruited to and bud into endosomal membranes, forming multivesicular endosomes, or become engulfed in double-membrane autophagosomes. Fusion of the outer membrane of these transport vesicles with the plasma membrane results in the extracellular release of vesicles containing viral capsids.

transcytosis represents an elegant mechanism for crossing physical barriers that normally provide an important intrinsic defense mechanism. Transcytosis has been documented not only for mammalian viruses, but also for invertebrate viruses such as those that need to cross the midgut of insects. Re-purposing this method of directional transport to facilitate invasion of the host offers several advantages, including the potential for a lack of dependence on specific cellular receptors and limiting the induction of signals capable of triggering innate immune antiviral responses. By definition, transcytosis does not result in productive infection of the involved cell. Nonetheless, by allowing viruses to traverse barriers posed by layers of protective epithelial cells to reach specific permissive types of cells supporting replication, transcytosis may be critical for virus spread within the host, pathogenesis, and viral transmission.

Certain viruses also usurp cellular mechanisms of vesicle-mediated transport to facilitate their release from cells. This form of viral egress is becoming increasingly recognized among viruses that are canonically considered to be nonenveloped but that are released from cells in the absence of cell lysis. It is best characterized for hepatitis A virus (HAV) and enteroviruses, both members of the picornavirus family, and hepatitis E virus (HEV), an unrelated positive-strand RNA virus. Newly assembled progeny virions either bud into endosomes, forming intraluminal vesicles within multivesicular endosomes, or are engulfed in autophagosomes, in both cases resulting in multi-layered vesicles that traffic to the plasma membrane where fusion of the outer membrane and plasma membrane discharges their contents and results in the release of virus within extracellular vesicles (Fig. 1, right). While sharing in common with transcytosis several features of vesicle-mediated transport, including the involvement of Rab-GTPases facilitating directional transport of virus-containing vesicles within cells and possibly also mechanisms of vesicle fusion at the plasma membrane, there are fundamental differences and different molecular machinery involved in these pathways.

Transcytosis

Transcytosis can be divided into three general steps: (1) the internalization of virus on one side of a polarized cell within a membrane-bound vesicle, (2) directional transport of the cargo-laden vesicle across the cell, and (3) release of the vesicle contents towards the extracellular space on the opposite side of the cell (Tuma and Hubbard, 2003) (Fig. 1, left). The first step in transcytosis of viruses commonly involves their endocytic uptake through invaginations of the plasma membrane in the form of clathrin-coated pits, caveolae, or macropinosomes. Once the virions have been internalized within membrane-bound vesicles, the second step requires the proper sorting and transport of these virion-loaded vesicles to the opposing side of the cell rather than, for example, endosomal recycling or a degradative lysosomal pathway. The final step involves release of the intraluminal contents of the vesicles to the extracellular space on the opposite side of the cell. Several signaling cues are required for this highly coordinated process to occur, and many aspects of it are not understood. Transcytosis requires cells to be highly polarized in nature, typically

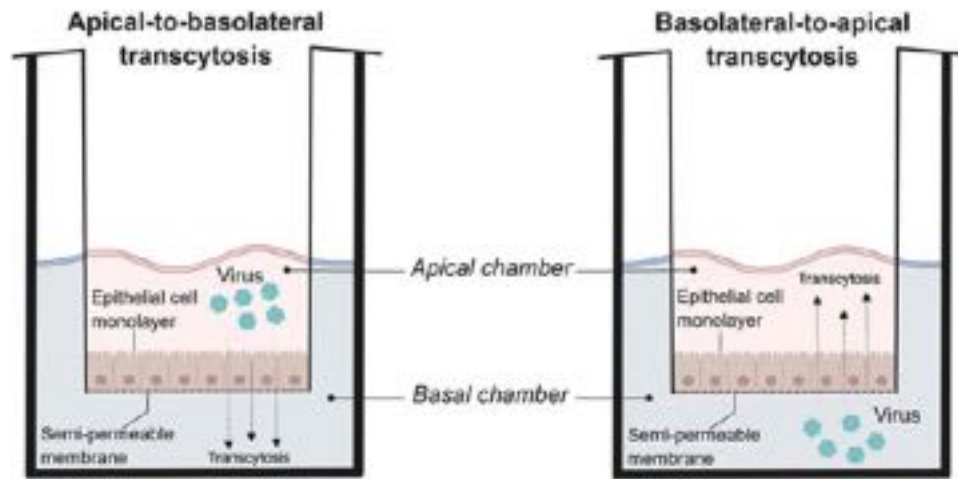


Fig. 2 Experimental set-up to study viral transcytosis in cell culture. Culture cells are grown at high confluency on a semi-permeable membrane and allowed to form a tight monolayer and establish polarity. For studying apical-to-basolateral viral transcytosis, virus inoculum is added to the media bathing the top of the cell monolayer (apical chamber). For basolateral-to-apical transcytosis, the virus inoculum is added to the basal chamber. Apical and basolateral fluids are collected at specific times post-infection and viral titers are measured. The integrity of the monolayer must be assessed throughout the experiment by measuring transepithelial electrical resistance or paracellular diffusion of marker molecules to ensure transcytotic movement.

across an apical-basal axis, and it is thus restricted to specific cell types. While most commonly studied in endothelial and epithelial cells, transcytosis occurs in other types of polarized cells, including hepatocytes and syncytiotrophoblasts of the placenta.

***In vitro* Models of Viral Transcytosis**

Cell culture models have been used widely to recapitulate transcytosis in the context of a viral infection. These systems can be conditionally manipulated to study effects of environment or genetic factors on the efficiency and kinetics of directional virus transport across cells. Most frequently used is the 'Transwell' assay that involves use of a nonpermeable monolayer of epithelial cells grown on a porous surface as substratum, with media bathing both sides of the cells to promote cell polarity (**Fig. 2**) (**Oztan et al., 2008**). A wide array of different porous materials (*e.g.*, polycarbonate, polystyrene, polyethylene-terephthalate) with distinct pore sizes and coatings that mimic the extracellular matrix are commercially available. In these assays, cells seeded at high density will establish polarity that can be assessed by determining the localization of proteins expressed on either the apical or basolateral membranes using fluorescence microscopy. The integrity of the cell monolayer is a critical factor for proper interpretation of these assays. To assess this, measurements of transepithelial electrical resistance (TER) and/or paracellular diffusion of a tracer molecule of known molecular mass, such as radiolabeled inulin, are routinely used to reflect tight junction integrity prior to and during viral infection.

The characteristics of cell lines used most often in Transwell assays (*e.g.*, cell-specific TER) have been well established (**Oztan et al., 2008**). Culture conditions in the two chambers can be modified to recapitulate *in vivo* properties, for example differences in pH between apical and basolateral fluids. Assays can also be modified to determine how different populations of cells impact transcytosis, by co-culturing different cell types, and to answer specific questions such as directionality of infection, infectivity of transcytosed virus, and contact-dependent transcytosis. Thus, in addition to documenting the absence of excessive paracellular diffusion, a proper viral transcytosis assay will assess (1) viral uptake from one side of the monolayer, (2) its appearance on the opposite side, (3) any intracellular accumulation of virus-derived material, and (4) any possible modification or degradation of viral material during the overall transcytosis process.

Several cell lines have been extensively used for these polarized cell culture models. These include Madin-Darby canine kidney (MDCK) cells (**Oztan et al., 2008**), which are easily cultured and establish polarity in less than a week, are readily infected by many types of viruses, and can be modified genetically. These cells can also be engineered to stably express the polymeric immunoglobulin receptor (pIgR), and thus can be used to study the role of IgA-dependent transcytosis during viral infections as discussed below. However, MDCK cells almost never represent the natural target cell for the virus in question. Viral transcytosis across intestinal epithelial cells has commonly been studied using Caco-2 cells, which differentiate along a crypt-to-villus axis and are easy to transfect. Efforts have also been made to model transcytosis of hepatitis viruses in hepatocytes using HepG2-N6 cells that express many liver-specific genes and can establish simple polarity in Transwell assays (**Snooks et al., 2008**). However, recapitulating the natural polarity of hepatocytes is difficult to accomplish in cell culture. The use of collagen gels employed in a sandwich configuration (*e.g.*, Matrigel) has improved efforts to establish polarization of primary hepatocyte cultures. 'Organoids' derived from animal or human tissue provide novel opportunities for achieving polarized cell function *in vitro* and are likely to open up new avenues in the future for studying directional viral entry, transcytosis, and release.

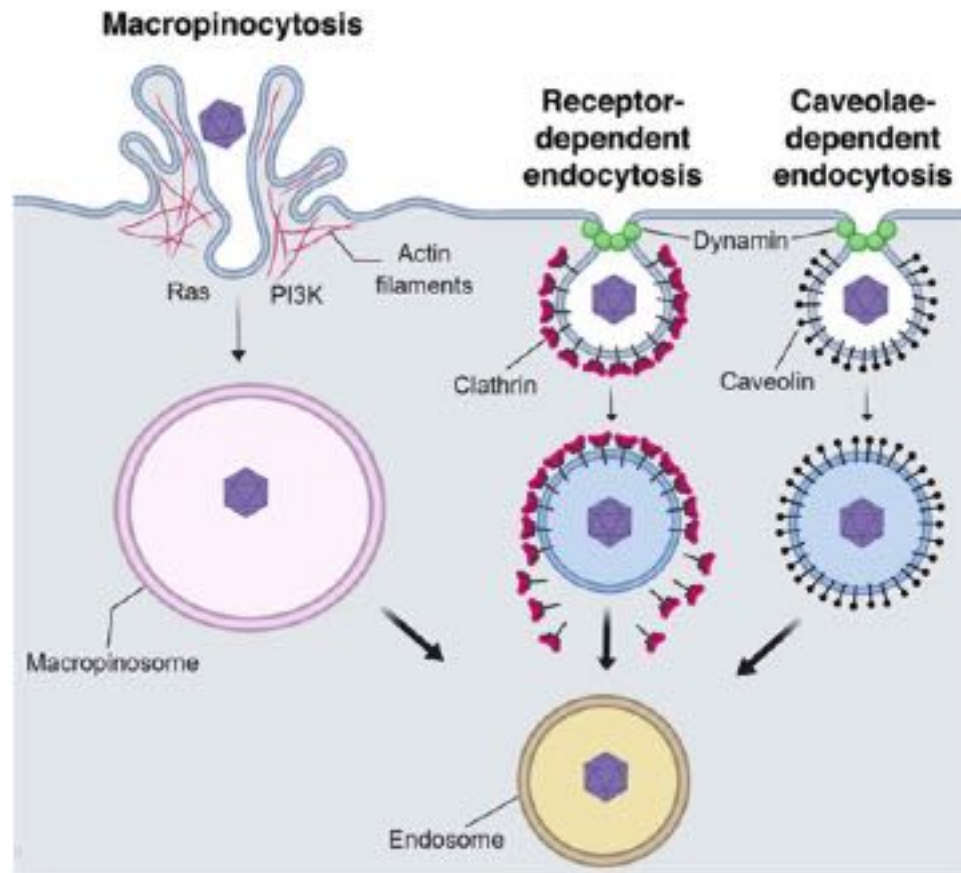


Fig. 3 Common pathways of endocytosis involved in transcytosis of viruses. Viruses can be internalized through different mechanisms of endocytosis. Macropinocytosis is an actin-dependent process regulated by Ras and PI3K in which molecules and fluid are engulfed from the extracellular space into a macropinosome. Receptor-mediated endocytosis involves the recognition of a ligand, such as a viral protein or immunoglobulin, through specific binding events and the subsequent formation of a clathrin-coated pit that internalizes the ligand in the form of a basket. Dynamin facilitates the scission of the pit off the plasma membrane, resulting in a clathrin-coated vesicle that will eventually uncoat. Caveolae-dependent endocytosis occurs at specific areas of the plasma membrane known as lipid rafts. Caveolin proteins and specific lipids are the main regulators of caveolae formation. In all cases, these vesicles typically deliver their contents to endosomes that will transport the cargo across the cell.

Endocytic Mechanisms Leading to Transcytosis

Endocytosis is a critical step in the internalization of viruses destined for transcytosis, and three distinct mechanisms of endocytosis are known to be involved. These different mechanisms share several common features, including the binding of the virion to attachment factors and/or receptors on the plasma membrane, invagination of the local plasma membrane, followed by membrane scission and internalization of a free vesicle primed for transport with its cargo (Elkin *et al.*, 2016) (Fig. 3). Endocytosis remains an active field of investigation, however, and other mechanisms may come to light in the future (Hemalatha and Mayor, 2019).

Caveolin-Dependent Endocytosis

Caveolin-dependent endocytosis entails internalization of extracellular material through flask-shaped pits of 50–80 nm in diameter known as caveolae that are present on apical and basal membranes of endothelial cells (Parton, 2018). Biochemically, caveolae are characterized by the presence of caveolin-1 (CAV-1), an integral transmembrane cholesterol-binding protein, and an enrichment in specific lipids such as cholesterol, sphingolipids, and phosphatidylinositol (4,5)-bisphosphate [PIP₂], all of which are required for the formation and/or stability of caveolae. CAV-1 binding to cholesterol promotes its oligomerization and the subsequent formation of caveolae, and thus chemicals that bind to or deplete cholesterol from membranes such as filipin and methyl- β -cyclodextrin, respectively, inhibit caveolae function. Studies using simian virus 40 (SV40) were among the first to reveal that the entrapment of virions in pre-existing caveolae triggers local tyrosine kinase-dependent signaling events, transient depolymerization of the cortical actin skeleton, and recruitment of dynamin to facilitate the fission of virion-loaded caveolae vesicles (known as caveosomes) into the cytoplasm for subsequent sorting (Pelkmans *et al.*, 2002). Epstein-Barr virus (EBV) utilizes caveolae to initiate transcytosis and

translocate from the basolateral to apical (mucosal) membrane of oral epithelial cells (Tugizov *et al.*, 2013). The apical release of EBV virions results in their secretion into saliva, which is likely important for the oral transmission of this herpesvirus.

Macropinocytosis

EBV also undergoes transcytosis in the reverse direction, from apical to basolateral membranes. In the early stages of infection, this may facilitate the transport of virus to B lymphocytes lying beneath the tonsillar epithelium which are permissive for productive infection. Apical-to-basolateral transcytosis of EBV is relatively rapid, and complete within 4 h, prior to the appearance of any progeny virions. It is not dependent on caveolae, but mediated instead by macropinocytosis, a distinct mechanism of endocytosis involving the non-selective ingestion of extracellular solutes and fluids by cells (Tugizov *et al.*, 2013). Macropinocytosis is characterized by the formation of a hollow ring of actin polymerization under the plasma membrane forming a circular ruffle. The closure of this ruffle and dissociation of actin filaments, steps regulated by the Ras protein and accumulation of phosphatidylinositol (3,4,5)-trisphosphate (PIP₃), leads to the formation of a relatively large 'macropinosome' vesicle (0.2–5 µm in diameter) primed for intracellular trafficking (Bloomfield and Kay, 2016). In addition to mediating apical-to-basolateral transcytosis of EBV, macropinocytosis has been implicated in the cellular entry of several other herpesviruses. Thus, endocytic mechanisms resulting in viral entry and productive infection overlap those associated with transcytosis. Why EBV uses two distinct endocytic mechanisms to initiate transcytosis in different directions (caveolae for basolateral-to-apical transcytosis, and micropinocytosis for the reverse) remains to be determined, but differences in plasma membrane protein and lipid composition may be at least partially responsible. In contrast to EBV, caveolae appear to mediate the apical-to-basolateral transcytosis of some adeno-associated viruses (AAV) (Di Pasquale and Chiorini, 2006), highlighting the complexity of the process and the diversity of signals that are likely involved in regulating transcytosis. This diversity is further highlighted by the fact that mice that are genetically incapable of forming caveolae are able to support transcytosis (Drab *et al.*, 2001).

Receptor-Mediated, Clathrin-Dependent Endocytosis

In addition to being nearly ubiquitous among different cell types, caveolae and macropinocytosis both share an independence from the requirement for a specific cellular receptor to trigger their activity. As such, these endocytic pathways are commonly referred to as receptor-independent. While both are undoubtedly involved in viral transcytosis, a third, clathrin-dependent endocytosis has been the most extensively documented and characterized (McMahon and Boucrot, 2011). This process involves the recognition of extracellular ligands – including viral proteins and immunoglobulin (Ig) receptors – by transmembrane receptors and the subsequent recruitment of clathrin and adapter proteins (APs) to the cytoplasmic face of the plasma membrane for the formation of vesicles. As such, this mechanism of entry provides higher specificity than receptor-independent endocytosis. Adapter complexes, such as that involving adapter protein 2 (AP2), recognize signals in the cytoplasmic tails of plasma membrane-anchored receptors and recruit clathrin molecules to initiate the polymerization and growth of a lattice that eventually coats the invaginating vesicle in the form of a basket known as clathrin-coated pit (McMahon and Boucrot, 2011). As for caveolae and some specific cases of macropinocytosis, dynamin is required for abscission of the membrane of the clathrin-coated vesicle from the plasma membrane. While clathrin-dependent endocytosis is directly involved in the canonical entry pathways of many diverse viruses, its role in viral transcytosis is almost exclusively linked to usage of Ig and Ig receptors as shuttles.

Immunoglobulins as Mediators of Transcytosis

Transcytosis plays an important role in the production of secretory immunoglobulins A and M (SIgA and SIgM) that provide protection against viruses at mucosal barriers (Turula and Wobus, 2018). These polymeric immunoglobulins (pIg) are produced by plasma cells within the lamina propria underlying the epithelial cell barrier. They undergo basolateral-to-apical transcytosis across this barrier upon binding the polymeric immunoglobulin receptor (pIgR). pIgR-pIg complexes are internalized by clathrin-mediated endocytosis, followed by vesicle-mediated transport from early endosomes near the basolateral membrane of the epithelial cell to Rab17-positive recycling endosomes at the apical membrane. Proteolytic cleavage of pIgR at its transmembrane domain results in release of the secretory component of pIgR bound to pIg as SIgA or SIgM into the lumen of the gut or bronchus. These secretory pIg complexes protect the epithelial surface but can also neutralize virus within endosomal compartments (Mazanec *et al.*, 1992). SIgA has been shown to mediate protection against many viruses, including Sendai virus, influenza A virus, measles virus, human immunodeficiency virus (HIV), reovirus and rotavirus (Turula and Wobus, 2018). Some viruses, such as simian immunodeficiency virus (SIV), downregulate pIgR expression in mucosal cells to impede Ig-mediated neutralization.

Although virus-specific immunoglobulins are typically protective against viruses at mucosal surfaces, in some cases antibodies promote dissemination of the virus within the host by facilitating transcytosis. The binding of virus-specific secretory IgA to EBV promotes internalization of the virus in epithelial cells expressing the pIgR, and this has been shown to allow the virus to enter and initiate replication within cells that are otherwise refractory to direct infection (Sixbey and Yao, 1992). Acting similarly,

EBV-specific immunoglobulins have been shown to mediate basolateral-to-apical transcytosis of EBV across polarized cultures of MDCK cells (Gan *et al.*, 1997), and may promote the apical-to-basolateral transcytosis of EBV across the tonsillar epithelium by micropinocytosis, as discussed above (Tugizov *et al.*, 2013). IgA-mediated transcytosis has also been described for human cytomegalovirus (HCMV), HAV and other viruses, as discussed in greater detail below.

Endocytic Sorting and Vectorial Transport of Vesicles

Endocytic vesicles can have one of several fates, including recycling back to the original plasma membrane, delivery of cargo to the lysosome for degradation, or transcytotic shuttling (Elkin *et al.*, 2016). This decision is controlled by a myriad of cellular proteins and lipids that must act in concert to ensure proper delivery of cargo and that are only partly understood. A detailed discussion of vesicle trafficking in polarized epithelial cells is well beyond the scope of this review. However, important regulators of trafficking can be grouped generally into three categories: the SNARE (Soluble NSF [N-ethylmaleimide-sensitive factor] Attachment Protein Receptor) proteins, Munc-18/Sec1 proteins, and Rab proteins.

SNARE proteins constitute a family of membrane-associated proteins present on vesicles (*v*-SNAREs) or target vesicles (*t*-SNAREs) that are involved in promoting membrane docking and fusion through interactions with NSF and its attachment factor, alpha-soluble NSF attachment protein (α -SNAP) (Wang *et al.*, 2017; Han *et al.*, 2017). In general, SNARE-dependent membrane targeting requires (1) SNARE activation and recruitment of specific Rab proteins to the proper membrane, (2) formation of cognate membrane attachment complexes provided by specific pairwise coupling of SNARE isoforms, and (3) membrane fusion and cargo mixing or release (Han *et al.*, 2017). The cytosolic NSF adapter protein is required for membrane fusion and is a critical regulator of polarized plasma membrane targeting, including IgA transcytosis (Apodaca *et al.*, 1996; Rothman, 1987). Following its discovery, the SNAP family of NSF receptors was shown to recruit NSF to specific organelles and to activate its ATPase activity. While SNARE proteins were initially thought to contribute to the specificity required for vesicle and target membrane fusion, the localization of specific SNAREs – 38 of which are known to date – to specific organelles and plasma membrane domains appears to be cell type-specific. Nonetheless, these proteins all function during tethering of a vesicle (*e.g.*, secretory vesicle) to a target membrane (*e.g.*, plasma membrane) and assist with the biophysical and energetic requirements for membrane fusion of the independent lipid bilayers. This is accomplished by interactions between three or four *v*- and *t*-SNAREs that result in the formation of a complex that bridges the membranes in the form of a molecular zipper and ultimately fuses the membranes. Thus, SNARE proteins and their adapters are critical for both the proper endocytic sorting and fusion of virion-loaded endosomes to secretory vesicles and the subsequent secretion of these vesicles upon fusion with the plasma membrane.

SNARE proteins function in concert with Munc-18/Sec1 that regulates vesicle docking and are associated with the plasma membrane through interactions with SNAREs. These proteins promote stability and spatially correct assembly of SNAREs complexes in vesicles, and may further assist in promoting SNARE-dependent membrane fusion (Wang *et al.*, 2017; Han *et al.*, 2017). The small GTP-binding Rab proteins constitute a third group of endocytic regulators involved in intracellular vesicle movement during transcytosis (Bhuin and Roy, 2014). At least 60 Rabs have been identified in mammalian cells and their expression – as with the SNAREs – is dependent on the cell type. Some of these proteins localize to specific domains of the plasma membrane in polarized cells, highlighting their function in transcytosis. For example, Rab11 and Rab25 regulate basolateral-to-apical transcytosis and inhibition of their function leads to accumulation of IgA in the apical recycling endosome and poor delivery to the apical plasma membrane. Rab proteins appear to facilitate vectorial trafficking of vesicles through associations with the cytoskeleton. As for Munc-18 proteins, Rabs are also involved in promoting vesicle docking and priming; and some Rabs may even recruit NSF and SNAREs to drive formation of the complex.

The cytoskeleton also plays a key role in the trafficking, and ultimately secretion, of endocytic vesicles involved in transcytosis (Apodaca, 2001). While microtubule function is not directly required for transcytosis, it does provide a molecular highway for vesicles to travel from one side of the cell to the other and is essential for the maintenance of cell polarity. RhoA, Rac1 and other Rho GTPases functionally integrate the organization of the cytoskeleton for membrane trafficking (Aspenström, 2014). Specific lipids with signaling properties – such as phosphoinositides, cholesterol, and glycosphingolipids – contribute as well. Reductions in phosphoinositide 3-phosphate (PI3P) resulting from pharmacological inhibition of phosphoinositide 3-kinase (PI3K) by inhibitors such as wortmannin and LY294002 have well known detrimental effects on endocytic trafficking. Inhibition of PI3K impairs basolateral-to-apical transcytosis of dimeric IgA and pIgR (Vergés *et al.*, 2007).

Secretion of Cargo

The final step in transcytosis involves the tethering of the secretory vesicle to the plasma membrane, SNARE-mediated fusion of the two membranes, and secretion of cargo. In eukaryotes, the exocyst complex is responsible for the initial tethering interaction between the secretory vesicle and the plasma membrane (Martin-Urdiroz *et al.*, 2016). Exocyst function is aided by Munc-18/Sec1 and is believed to bridge the SNAREs on the secretory vesicle and the plasma membrane. This evolutionarily conserved octameric complex – composed of Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70, and Exo84 – localizes to the cytoplasm, the trans-Golgi network, recycling endosomes, and specific domains of the plasma membrane. In the current model, Exo70 recruits other components of the exocyst to apical or basolateral membranes through interactions with specific lipids in the inner leaflet. This interaction is critical for assembly of the exocyst complex at the plasma membrane and for the docking and fusion of secretory vesicles (He and Guo, 2009).

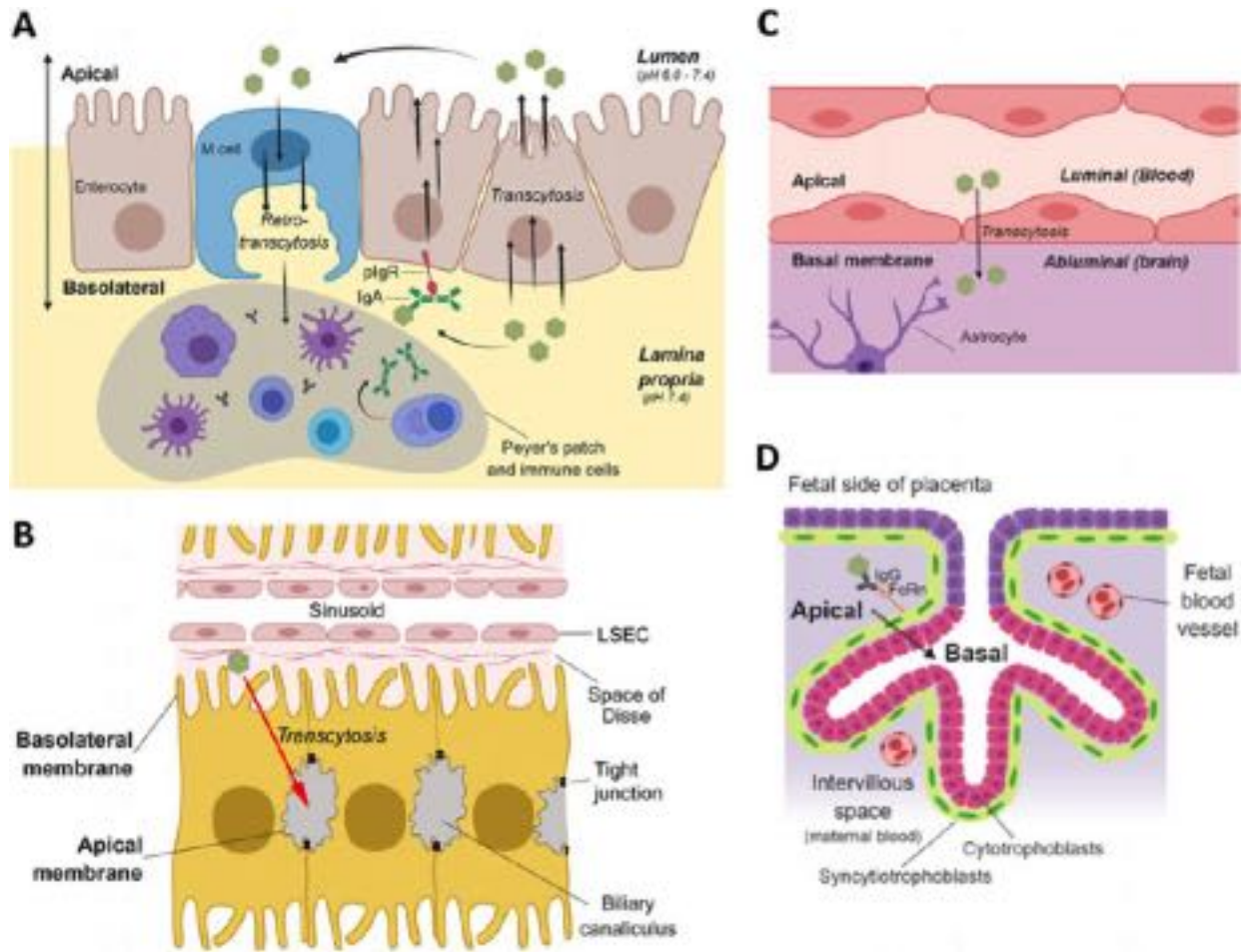


Fig. 4 Transcytosis contributes to viral pathogenesis in diverse tissues. (A) Viral transcytosis and retrotranscytosis across the gastrointestinal tract. (B) Hepatocyte polarity and basolateral-to-apical transcytosis resulting in virus release into the biliary tract. LSEC = liver sinusoidal endothelial cell. (C) Structure of the blood-brain barrier and apical-to-basolateral transcytosis of Virus Researchulting in neuroinvasion. (D) Structure of a placental villous tree and apical-to-basolateral transcytosis of viruses from the maternal blood to fetal tissue.

Transcytosis and Viral Pathogenesis

Cell polarity is essential for the three-dimensional development of tissues and the creation of protective barriers between cells and the external environment. Transcytosis itself contributes to cell polarity through directional transport of *de novo* synthesized proteins to specific domains of the plasma membrane (Garcia-Castillo *et al.*, 2017). Polarized cells possess distinct populations of endocytic vesicles, such as 'apical' and 'basolateral' endosomes, depending on their location and biochemical composition. Although transcytosis is defined as the directional movement of cargo within polarized cells, several cell types that lack clear apical-basal polarity, such as neurons and osteoclasts, employ similar strategies to transfer macromolecules between distinct extracellular environments (Coxon and Taylor, 2008). Transcytosis allows for directional viral entry and secretion from apical and/or basolateral membranes and is a critical factor in the pathogenesis of many viral infections. This section presents some additional examples of viruses that use transcytosis to cross cellular barriers in diverse organs and discusses the cell types involved in this process.

Breaching Epithelial Barriers

Enteric viral infections constitute a major burden in public health and several related viruses have been demonstrated to use transcytosis to facilitate dissemination within the gastrointestinal tract. Human noroviruses (HuNoVs) are responsible for a large number of gastroenteritis outbreaks around the world and are a leading cause of severe childhood diarrhea in the United States. While the precise mechanisms by which HuNoVs infect the intestinal tract and induce pathology remain debated, both HuNoVs and murine noroviruses (MuNoVs) undergo transcytosis across intestinal epithelial cells *in vitro* in the absence of productive infection (Karst and Wobus, 2015). Noroviruses are transcytosed from the gut lumen to the lamina propria through M-like cells where they may infect immune cells resident within Peyer's patches (Fig. 4(A)). Recent studies using jejunal enteroids show that HuNoVs also replicate in enterocytes, which may be more important target cells in the pathogenesis of the infection (Ettayebi *et al.*, 2016).

Secretory immunoglobulin-virus complexes in the lumen of the gut can be sampled and endocytosed by specialized microfold cells (M cells) overlying mucosal-associated lymphoid tissue by a process known as 'retrotranscytosis' (Fig. 4(A)). The apical-to-basolateral retrograde transport of SIg complexes and other antigens to immune cells residing beneath the epithelium in Peyer's patches within the lamina propria is important for normal immune responses and the maintenance of tissue homeostasis (Kobayashi *et al.*, 2019). M cells may serve as portals for viral entry and dissemination, as transcytosis of virus across M cells may allow subsequent infection of epithelial cells from their basolateral surface, or invasion of cells residing within the lamina propria. Although it is not clear that transcytosis is required for infection of enterocytes, MuNoV titers were reduced in mice in which M cells were depleted (Gonzalez-Hernandez *et al.*, 2014). Poliovirus undergoes transcytosis in Transwell experiments utilizing Caco-2 human intestinal epithelial cells, and apical-to-basolateral transcytosis across M cells may allow it to breach the gastrointestinal epithelial barrier (Ouzilou *et al.*, 2002). Support for this hypothesis comes from electron microscopy studies that have documented the presence of poliovirus virions within villous M-like cells in tissue from humans and experimentally-challenged rhesus monkeys (Takahashi *et al.*, 2008; Siciński *et al.*, 1990). The mechanistic details of transcytosis are uncertain, including which endocytic pathway is utilized and whether a specific cellular receptor is required.

A particularly interesting example of viral transcytosis is provided by HIV. Upon contacting the apical surface of an epithelial cell, HIV-infected peripheral blood leukocytes have been shown to release virus in a vectorial fashion towards the epithelial surface (Bomsel, 1997). The released virus undergoes rapid transcytosis, emerging on the basolateral side of the epithelial barrier within 30 min where it can then infect macrophages and lymphocytes resident within the submucosa.

Neuroinvasion

The blood-brain barrier (BBB) shields the brain both physically and metabolically, and acts as a highly selective portal for nutrient delivery to the brain. It is comprised primarily of brain microvascular endothelial cells (BMECs), supported by a network of underlying astrocytes and pericytes embedded in the basement membrane surrounding the vasculature (Hou *et al.*, 2016; Engelhardt and Sorokin, 2009). The BBB is essential for maintenance of homeostasis within the brain microenvironment. It is oriented with its apical surface facing the systemic circulation, and the basolateral surface facing the abluminal (brain) side of the capillary endothelium. Barrier function is provided by tight junctions composed of transmembrane proteins including occludin, claudins, and junctional adhesion molecules. Positioned between adjacent cells, the tight junctions limit free fluid exchange and paracellular diffusion of solutes (Fig. 4(B)). Transcytosis across BMECs is highly restricted and actively regulated. Both caveolae- and clathrin-mediated endocytosis occur in BMECs, with caveolae-dependent endocytosis more prevalent (Ayloo and Gu, 2019). Micropinocytosis is less evident.

Despite the strong physical barrier provided by the BBB, a lengthy list of viruses is capable of breaching it and invading the central nervous system. Neuroinvasion can be achieved by multiple mechanisms, including non-hematogenous entry through peripheral nerves, but transcytosis across the BBB likely contributes to pathogenesis of some infections. Fluorescently-labeled Zika virus, a mosquito-borne flavivirus with severe consequences for the developing brain, has been shown to undergo energy-dependent transcytosis across polarized Transwell cultures of choriocarcinoma cells with impermeable tight junctions (Chiu *et al.*, 2020). Transcytosis was reduced by small molecule inhibitors of both clathrin-dependent and caveolae-mediated endocytosis. However, Zika virus has also been shown to productively infect BMEC, and to be released on the basolateral side of an impermeable model BBB without destroying its integrity (Mladinich *et al.*, 2017). Thus, the importance of transcytosis in neuroinvasion by Zika virus remains uncertain. More convincing evidence exists for encephalitic alphaviruses. Electron microscopy demonstrated the presence of immunogold-labeled Venezuelan equine encephalitis virus (VEEV) particles within cortical BMECs of infected mice, and VEEV was shown to undergo caveolae-mediated apical-to-basolateral transcytosis in Transwell cultures of primary murine BMECs (Salimi *et al.*, 2020). Transcytosis was inhibited by both interferon and small molecule inhibitors of RhoA GTPases. Consistent with these *in vitro* observations, VEEV titers were reduced in the brains of CAV-1 knockout mice without reductions in the titer in other organs (Salimi *et al.*, 2020). These latter findings provide strong support for caveolae-mediated transcytosis across the BBB playing an important role in neuroinvasion by alphaviruses.

AAV has been considered as a possible vector for gene therapy in the central nervous system, resulting in considerable interest in its potential for transcytosis across the BBB. AAV9 has been shown to be transported across polarized monolayer cultures of primary human BMECs without increases in paracellular diffusion, with release occurring on the basolateral side of the monolayer within 3 h (Merkel *et al.*, 2017). AAV9 transcytosis was notably more efficient than AAV2 transcytosis, and also possessed greater gene transduction activity within the central nervous system. By contrast, AAV2 demonstrated greater transduction activity within the BMEC monolayer. Differences in AAV9 and AAV2 may reflect utilization of different cell attachment factors, with AAV2 binding to heparan sulfate proteoglycans and AAV9 suggested to bind galactose moieties on the apical BMEC membrane (Merkel *et al.*, 2017).

Viral Hepatitis

The liver is the largest metabolic organ in the human body and responsible for synthesis of bile acids, lipoproteins, clotting factors and other multiple plasma proteins. It has a unique vascular circuit in which venous blood coming from the intestines *via* the portal vein and arterial blood from the hepatic artery pass through sinusoids aligned along linear arrays of hepatocytes, the major cell type in the liver (Fig. 4(C)). The hepatocytes are separated from the sinusoids by rows of liver sinusoidal endothelial cells

(LSECS) with fenestrae that provide access from the sinusoids into the space of Disse that exists between LSECS and hepatocytes (Shetty *et al.*, 2018). Hepatocytes are of epithelial origin and are highly polarized, with their basolateral surface facing onto the space of Disse, and with tight junctions on their lateral surfaces bracketing a much smaller apical membrane surface that opens into small canals (biliary canaliculi) that drain bile into the biliary tract (Gissen and Arias, 2015). The apical and basolateral membranes of hepatocytes are thus separated by the tight junctions that seal the canalicular lumen. The apical membrane is metabolically highly active and responsible for the secretion of bile salts that are ultimately delivered to the gastrointestinal tract (Fig. 4(C)). Given this organization, viruses typically reach hepatocytes hematogenously, through sinusoidal blood, and enter hepatocytes through the basolateral membrane.

Productive infection of hepatocytes results in either the basolateral release of viral particles back to the sinusoidal blood for spread within the liver, apical release of virus into the biliary tract for eventual shedding in feces, or a combination of both. Five very different viruses have primary tropism for the hepatocyte, only two of which retain significant infectivity in the bile: HAV (a picornavirus) and HEV (a hepevirus). Hepatitis B virus (HBV, an hepadnavirus), hepatitis C virus (HCV, a flavivirus), and the satellite-like hepatitis D virus, each possess lipid envelopes that are essential for their infectivity and that are destroyed by the high concentrations of unbuffered bile salts present in the proximal biliary canaliculus. Transcytosis can thus contribute only to the pathogenesis of HAV or HEV infection in hepatocytes, and only in a basolateral-to-apical (sinusoids to biliary canaliculus) direction.

Although distantly related phylogenetically, HAV and HEV share an unusual lifecycle in which they are shed from infected persons in feces as naked virus particles lacking envelopes, but circulate within the blood during acute infection cloaked in host membranes (Feng *et al.*, 2014). In the case of HAV, these ‘quasi-enveloped virions’ (eHAV) are small membranous vesicles resembling exosomes that contain one or more capsids (Feng *et al.*, 2013). eHAV vesicles are as infectious as the naked virus (nHAV), and enter cells through a complex endocytic pathway in which the membranes surrounding the capsid are degraded in late endolysosomes (Rivera-Serrano *et al.*, 2019; Das *et al.*, 2020). Quasi-enveloped virions are the dominant (perhaps only) form of these viruses released from cells, and both eHAV and eHEV are released from the basolateral membrane as well as the apical membrane of infected hepatocytes (Capelli *et al.*, 2019; Hirai-Yuki *et al.*, 2016). This results in their secretion into both blood and bile. Bile salts strip the membranes from virus released into the biliary canaliculus, explaining why only naked virions are shed in feces. This unusual lifecycle allows for stealthy spread of virus within the host, in a membrane-cloaked form sequestered from the immune system, coupled with efficient transmission of stable, non-enveloped particles to naïve hosts through the environment. Transcytosis across polarized, nonpermeable cell monolayers occurs only at very low rates for both HAV and HEV in Transwell assays (Capelli *et al.*, 2019; Blank *et al.*, 2000). However, transcytosis of antibody-virus complexes has been suggested to contribute to the pathogenesis of hepatitis A in humans.

Studies by Dotzauer *et al.* (2000) demonstrated that HAV can gain entry into otherwise poorly permissive mouse hepatocytes when complexed with virus-specific IgA. Endocytosis of the IgA-virus complex may have been driven in part in these studies by binding to the asialoglycoprotein receptor (ASGPR) expressed by hepatocytes. However, the pIgR was subsequently shown to mediate apical-to-basolateral transcytosis of IgA-coated HAV particles across polarized epithelial cells (Dotzauer *et al.*, 2005). Elegant experiments done much later in mice revealed that intravenously administered HAV undergoes transcytosis across the liver, from the circulation to the biliary tract and thence to feces (Counihan and Anderson, 2016). Transcytosis occurred at a low basal rate in mice, but was increased to ~9% over a 4 h period when virus was bound to a poorly neutralizing IgA antibody. Viral uptake into hepatocytes was dependent upon the murine pIgR, as transcytosis was not observed in mice in which this gene was knocked out (Counihan and Anderson, 2016). Transcytosis and fecal shedding were also not observed when virus was complexed with a potent neutralizing IgA antibody. Despite their elegance, the relevance of these observations to fecal shedding of HAV during acute hepatitis A is doubtful, as virus circulating in the blood and passing through the hepatic sinusoids is mostly if not entirely quasi-enveloped and incapable of binding HAV-specific antibodies (Feng *et al.*, 2013). Moreover, most fecal shedding of virus occurs prior to the development of humoral immunity to the virus, and the transcytosis rate observed in the absence of antibody cannot account for the large amounts of virus shed in feces (Counihan and Anderson, 2016; Lanford *et al.*, 2011). On the other hand, transcytosis of naked HAV virions across the gastrointestinal mucosa, perhaps in M cells, provides an attractive explanation for how virus might initially reach the liver in the absence of productive replication within the gut, which has never been demonstrated.

Fetal Infections

The placenta provides the growing fetus with physical and immunological protection against blood-borne pathogens during pregnancy, while also ensuring delivery of essential nutrients and oxygen. The main cell type in the placenta, a specialized epithelial cell known as the trophoblast, exists in two distinct layers (Fig. 4(D)): an inner layer composed of mononucleated cytotrophoblasts positioned at the fetal interface that retain proliferative activity and serve as progenitor cells, and an outer layer consisting of a syncytium of terminally-differentiated syncytiotrophoblasts facing onto the maternal circulation (Turco and Moffett, 2019). The latter are highly polarized and characterized by a dense layer of microvilli with expression of the neonatal IgG receptor (FcRn) on the apical membrane facilitating apical-to-basolateral transport of maternal IgG to the fetus by transcytosis (Pyzik *et al.*, 2019).

Several viruses that can be transmitted from the mother to the fetus *in utero* have been suggested to breach this barrier by undergoing apical-to-basolateral transcytosis across the syncytiotrophoblast epithelium. These include HIV and other lentiviruses, HBV, human cytomegalovirus (HCMV), and Zika virus (Lagaye *et al.*, 2001; Bhat and Anderson, 2007; Maidji *et al.*, 2006). With HIV,

close cellular contact between infected macrophages and the syncytiotrophoblast may be important for the initiation of transcytosis. In other cases, transcytosis may be receptor-dependent and involve virus-antibody complexes. For example, transcytosis of HCMV appears to be mediated by the interaction of immunoglobulin-coated virions with FcN expressed by syncytiotrophoblasts in placentas from women with low levels of neutralizing antibodies to the virus (Maidji *et al.*, 2006). Studying the transcytosis of viruses across the placental barrier has been historically challenging due to limitations of *in vitro* models that fail to accurately recapitulate the structure of the human placenta. Moreover, it may be difficult to discriminate between transcytosis and other mechanisms that may lead to breaching of the placental barrier, such as the productive infection of trophoblasts by Zika virus (Aagaard *et al.*, 2017). Decreased tight junction function leading to increased paracellular permeability of the syncytiotrophoblast layer also has been observed in Zika virus-infected women (Miranda *et al.*, 2019). The recent development of trophoblast organoids is likely to provide new opportunities to determine how viruses cross the barrier provided by the placenta.

Extracellular Vesicle-mediated Export of Viruses

Nonlytic Release of Virus in Extracellular Vesicles

Many viruses that have been classified canonically as ‘nonenveloped’ are released nonlytically from infected cells in extracellular vesicles (EVs), a phenomenon that has potentially strong consequences for the course of infection and the host immune response to it (Feng *et al.*, 2013; Robinson *et al.*, 2014; Santiana *et al.*, 2018; van der Grein *et al.*, 2019). As discussed above, both HAV and HEV, positive-strand RNA viruses that infect the liver, are released from the basolateral surface of hepatocytes into the bloodstream where they circulate cloaked in lipid membranes that protect the capsid from neutralizing antibodies (Feng *et al.*, 2013; Feng *et al.*, 2014). Quasi-enveloped eHAV virions possess a specific infectivity similar to naked HAV virions, are ~50–110 nm in diameter, and have a buoyant density of ~1.100 gm/cm³ in iodixanol gradients. eHAV enters cells to initiate infection *via* clathrin-dependent endocytosis, enabled in part by receptors such as TIM1 on the cell surface that recognize phosphatidylserine displayed on the exterior of eHAV (Feng *et al.*, 2015; Das *et al.*, 2017; Rivera-Serrano *et al.*, 2019). The membrane is degraded by enzymes and cholesterol transporters in the late endolysosome, allowing the exposed capsid to bind gangliosides in the endolysosomal membrane that mediate later steps in entry resulting in release of the genome into the cytoplasm to initiate translation (Rivera-Serrano *et al.*, 2019; Das *et al.*, 2020).

Several other picornaviruses have been shown to be released from cultured cells in EVs, including coxsackievirus B3, poliovirus, and encephalomyocarditis virus (Robinson *et al.*, 2014; Chen *et al.*, 2015; van der Grein *et al.*, 2019; Yang *et al.*, 2020). These EVs are larger than eHAV vesicles, and contain large numbers of capsids. Infection with vesicles containing multiple viral capsids may result in the simultaneous introduction of multiple viral genomes into a newly infected cell. This ‘*en bloc*’ transmission of virus effectively raises the infection multiplicity and may provide some fitness advantages to the virus (Santiana *et al.*, 2018; Aguilera and Pfeiffer, 2019). Vesicle-mediated export has also been demonstrated for norovirus and rotavirus, and infectious vesicles containing rotaviruses and norovirus have been observed in fecal samples (Santiana *et al.*, 2018).

The release of virus in EVs involves membrane budding that is topologically the inverse of endocytosis (Fig. 1, right). Whereas vesicles resulting from endocytosis exclude cytoplasm and contain extracellular fluid, EV release results from budding in an outward direction from the cytoplasm with the resulting vesicle containing cytoplasmic contents. This difference in topology mandates the use of different cellular machinery to seal the membrane during abscission of the vesicle, with EV release being dependent upon endosomal sorting complexes required for transport, otherwise known as ESCRT (Vietri *et al.*, 2020), rather than the membrane remodeling activity of dynamin used in endocytosis. This outward budding of cytoplasmic contents can occur at the plasma membrane, resulting in the release of large ‘microvesicles’ from cells. However, such budding also occurs on endosomal membranes, resulting in intraluminal vesicles (ILVs) carrying cytoplasmic cargo, and ultimately large multivesicular endosomes (MVEs, sometimes called multivesicular bodies) containing multiple ILVs (Gruenberg, 2020). MVEs can fuse with lysosomes leading to degradation of their cargo, or traffic to the cell surface where fusion with the plasma membrane results in the release of ILVs to the external environment as ‘exosomes’, small vesicles capable of shuttling proteins and microRNAs between different cells (Colombo *et al.*, 2014).

Exosome-Like Release of Quasi-enveloped Viruses

In many ways, including their size and protein composition, eHAV vesicles resemble exosomes (McKnight *et al.*, 2017; They *et al.*, 2018). Both eHAV and eHEV appear to originate within MVEs and their release is highly dependent upon ESCRT and conserved protein motifs (‘late domains’) present in the structural proteins that mediate interactions with ALIX and TSG101, ESCRT-associated proteins. These late domains are similar to those present in structural proteins of canonical enveloped viruses (PPXY, P(S/T)AP, GPPX₃Y, or YPX₁₀₃L) that mediate interactions with ESCRT machinery (Chen and Lamb, 2008; Ren and Hurley, 2011). Quasi-envelopment of these hepatitis viruses thus shares several features in common with the budding of canonical enveloped viruses, which is considered elsewhere in detail in this volume. A major difference, however, at least with eHAV, is that virus assembly (assembly of the capsid and encapsidation of the genome) occurs prior to and independent of recruitment of the virus as cargo to be loaded into ILVs (Fig. 1, right). By contrast, assembly of canonical enveloped viruses typically occurs at the membrane and is coupled to budding.

ESCRT complexes play a key role in eHAV and eHEV release. Three distinct ESCRT protein complexes act sequentially to sort and load cargo into MVBs (Hurley, 2015; Votteler and Sundquist, 2013; Votteler *et al.*, 2016). ESCRT-I and ESCRT-II are core ESCRT machinery that sort cargo and sequester it at the endosomal membrane, creating an inward bud (Hurley, 2015). ESCRT-III oligomers are subsequently recruited and form a complex that pinches the neck of the membrane invagination, releasing the bud into the endosomal lumen and sealing the membrane, thus forming an ILV. In a later step, the ATPase VPS4 acts to dissociate the ESCRT-III oligomers. As indicated above, two ESCRT-associated proteins are important for the release of eHAV or eHEV in vesicles: TSG101 and ALIX. TSG101 is associated with ESCRT-I, whereas ALIX is an accessory ESCRT protein that feeds into ESCRT-III complexes in parallel with ESCRT-I/II. Interactions with it represent a second pathway for cargo selection. The HAV capsid interacts with ALIX through tandem YPX₃L late domain motifs in VP2 as well as in the C-terminal pX domain of the VP1pX protein, both of which contribute to eHAV release (Gonzalez-Lopez *et al.*, 2018; Jiang *et al.*, 2020). The HEV capsid interacts with TSG101 through a PSAP 'late domain' motif in its ORF3 protein which is similarly essential for release of quasi-enveloped virus (Nagashima *et al.*, 2011). siRNA knockdown of ALIX, CHMP2A (part of the ESCRT-III complex) or VPS4B strongly inhibit eHAV release, and both ALIX and VPS4 are physically present in extracellular eHAV vesicles (Feng *et al.*, 2013; McKnight *et al.*, 2017). Following formation of MVEs containing intraluminal eHAV or eHEV, the cues that result in their trafficking to the plasma membrane, where fusion results in release of quasi-enveloped virus to the extracellular milieu, are unknown.

Autophagy-Related Vesicle-Mediated Release of Virus

Autophagy is a complex and tightly regulated cellular process whereby unwanted cytoplasmic content becomes engulfed within double-membrane vesicles (autophagosomes) that fuse with lysosomes for degradation of their cargo (Levine and Kroemer, 2008). In some cases, however, autophagy does not result in lysosomal degradation but rather noncanonical secretion of cytoplasmic cargo isolated by the autophagic machinery (Ponpuak *et al.*, 2015). Secretory autophagy involves the movement of the autophagosome to the cell surface and fusion of its outer membrane with the plasma membrane. The membrane topology of autophagy has long been recognized to offer the potential for nonlytic release of viruses that assemble within the cytoplasm of infected cells (Jackson *et al.*, 2005). The topology is essentially the same as that described above for exosome-like release of quasi-enveloped viruses (Fig. 1, right). Autophagy is similarly ESCRT dependent (Vietri *et al.*, 2020; Zhou *et al.*, 2019).

While incompletely understood, it is becoming increasingly clear that multiple RNA viruses with cytoplasmic replication cycles subvert different facets of autophagy to promote both amplification of their genomes as well as their nonlytic release from cells in EVs. Viral particles assembling within the cytoplasm become engulfed in autophagosomes and subsequently undergo non-canonical secretion from the cell. Autophagic signaling is required for both the replication of poliovirus and other enteroviruses as well as their release from cells in EVs (Robinson *et al.*, 2014; Bird *et al.*, 2014). Coxsackievirus B3 is released in EVs associated with the autophagy-related protein, LC3. Poliovirus utilizes LC3 as well as a second autophagy-related protein, ATG9, to promote its intracellular replication, possibly contributing to membrane rearrangements required for the assembly of replication organelles that mediate viral RNA synthesis (Abernathy *et al.*, 2019). Autophagic signaling is induced in a noncanonical fashion, independent of the ULK1 kinase complex normally engaged in induction of autophagy (Corona Velazquez *et al.*, 2018). The enterovirus EV-D68, which is likely responsible for recent outbreaks of acute flaccid paralysis in children, also activates autophagy, but it also suppresses later steps in autophagy resulting in decreased autophagic flux (Corona *et al.*, 2018). A protease expressed by the virus cleaves SNAP29, a SNARE protein engaged in regulating autophagosome-lysosome fusion, thereby redirecting autophagic trafficking and limiting degradative autophagic flux.

References

- Aggaard, K.M., Lahon, A., Suter, M.A., *et al.*, 2017. Primary human placental trophoblasts are permissive for Zika virus (ZIKV) replication. *Scientific Reports* 7, 41389.
- Abernathy, E., Mateo, R., Majzoub, K., *et al.*, 2019. Differential and convergent utilization of autophagy components by positive-strand RNA viruses. *PLOS Biology* 17, e2006926.
- Aguilera, E.R., Pfeiffer, J.K., 2019. Strength in numbers: Mechanisms of viral co-infection. *Virus Research* 265, 43–46.
- Apodaca, G., 2001. Endocytic traffic in polarized epithelial cells: Role of the actin and microtubule cytoskeleton. *Traffic* 2, 149–159.
- Apodaca, G., Cardone, M.H., Whiteheart, S.W., Dasgupta, B.R., Mostov, K.E., 1996. Reconstitution of transcytosis in SLO-permeabilized MDCK cells: Existence of an NSF-dependent fusion mechanism with the apical surface of MDCK cells. *EMBO Journal* 15, 1471–1481.
- Aspenström, P., 2014. Atypical Rho GTPases RhoD and Rif integrate cytoskeletal dynamics and membrane trafficking. *Biological Chemistry* 395, 477–484.
- Ayloo, S., Gu, C., 2019. Transcytosis at the blood-brain barrier. *Current Opinion in Neurobiology* 57, 32–38.
- Bhat, P., Anderson, D.A., 2007. Hepatitis B virus translocates across a trophoblastic barrier. *Journal of Virology* 81, 7200–7207.
- Bhuin, T., Roy, J.K., 2014. Rab proteins: The key regulators of intracellular vesicle transport. *Experimental Cell Research* 328, 1–19.
- Bird, S.W., Maynard, N.D., Covert, M.W., Kirkegaard, K., 2014. Nonlytic viral spread enhanced by autophagy components. *Proceedings of the National Academy of Sciences of the United States of America* 111, 13081–13086.
- Blank, C.A., Anderson, D.A., Beard, M., Lemon, S.M., 2000. Infection of polarized cultures of human intestinal epithelial cells with hepatitis A virus: Vectorial release of progeny virions through apical cellular membranes. *Journal of Virology* 74, 6476–6484.
- Bloomfield, G., Kay, R.R., 2016. Uses and abuses of macropinocytosis. *Journal of Cell Science* 129, 2697–2705.
- Bomse, M., 1997. Transcytosis of infectious human immunodeficiency virus across a tight human epithelial cell line barrier. *Nature Medicine* 3, 42–47.
- Capelli, N., Marion, O., Dubois, M., *et al.*, 2019. Vectorial release of hepatitis E virus in polarized human hepatocytes. *Journal of Virology* 93.
- Chen, Y.H., Du, W., Hagemeijer, M.C., *et al.*, 2015. Phosphatidylserine vesicles enable efficient *en bloc* transmission of enteroviruses. *Cell* 160, 619–630.
- Chen, B.J., Lamb, R.A., 2008. Mechanisms for enveloped virus budding: Can some viruses do without an ESCRT? *Virology* 372, 221–232.

- Chiu, C.F., Chu, L.W., Liao, I.C., *et al.*, 2020. The mechanism of the Zika virus crossing the placental barrier and the blood-brain barrier. *Frontiers in Microbiology* 11, 214.
- Colombo, M., Raposo, G., Thery, C., 2014. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annual Review of Cell and Developmental Biology* 30, 255–289.
- Corona Velazquez, A., Corona, A.K., Klein, K.A., Jackson, W.T., 2018. Poliovirus induces autophagic signaling independent of the ULK1 complex. *Autophagy* 14, 1201–1213.
- Corona, A.K., Saulsbery, H.M., Corona Velazquez, A.F., Jackson, W.T., 2018. Enteroviruses remodel autophagic trafficking through regulation of host SNARE proteins to promote virus replication and cell exit. *Cell Reports* 22, 3304–3314.
- Counihan, N.A., Anderson, D.A., 2016. Specific IgA enhances the transcytosis and excretion of hepatitis A virus. *Scientific Reports* 6, 21855.
- Coxon, F.P., Taylor, A., 2008. Vesicular trafficking in osteoclasts. *Seminars in Cell and Developmental Biology* 19, 424–433.
- Das, A., Barrientos, R., Shiota, T., *et al.*, 2020. Gangliosides are essential endosomal receptors for quasi-enveloped and naked hepatitis A virus. *Nature Microbiology* 5, 1069–1078.
- Das, A., Hirai-Yuki, A., Gonzalez-Lopez, O., *et al.*, 2017. TIM1 (HAVCR1) Is not essential for cellular entry of either quasi-enveloped or naked hepatitis A virions. *mBio* 8.
- Di Pasquale, G., Chiorini, J.A., 2006. AAV transcytosis through barrier epithelia and endothelium. *Molecular Therapy* 13, 506–516.
- Dotzauer, A., Brenner, M., Gebhardt, U., Vallbracht, A., 2005. IgA-coated particles of Hepatitis A virus are translocated antievectorially from the apical to the basolateral site of polarized epithelial cells *via* the polymeric immunoglobulin receptor. *Journal of General Virology* 86, 2747–2751.
- Dotzauer, A., Gebhardt, U., Bieback, K., *et al.*, 2000. Hepatitis A virus-specific immunoglobulin A mediates infection of hepatocytes with hepatitis A virus *via* the asialoglycoprotein receptor. *Journal of Virology* 74, 10950–10957.
- Drab, M., Verkade, P., Elger, M., *et al.*, 2001. Loss of caveolae, vascular dysfunction, and pulmonary defects in caveolin-1 gene-disrupted mice. *Science* 293, 2449–2452.
- Elkin, S.R., Lakoduk, A.M., Schmid, S.L., 2016. Endocytic pathways and endosomal trafficking: A primer. *Wien Med Wochenschr* 166, 196–204.
- Engelhardt, B., Sorokin, L., 2009. The blood-brain and the blood-cerebrospinal fluid barriers: Function and dysfunction. *Seminars in Immunopathology* 31, 497–511.
- Etayebi, K., Crawford, S.E., Murakami, K., *et al.*, 2016. Replication of human noroviruses in stem cell-derived human enteroids. *Science* 353, 1387–1393.
- Feng, Z., Hensley, L., Mcknight, K.L., *et al.*, 2013. A pathogenic picornavirus acquires an envelope by hijacking cellular membranes. *Nature* 496, 367–371.
- Feng, Z., Hirai-Yuki, A., Mcknight, K.L., Lemon, S.M., 2014. Naked viruses that aren't always naked: Quasi-enveloped agents of acute hepatitis. *Annual Review of Virology* 1, 539–560.
- Feng, Z., Li, Y., Mcknight, K.L., *et al.*, 2015. Human pDCs preferentially sense enveloped hepatitis A virions. *Journal of Clinical Investigation* 125, 169–176.
- Gan, Y.J., Chodosh, J., Morgan, A., Sixbey, J.W., 1997. Epithelial cell polarization is a determinant in the infectious outcome of immunoglobulin A-mediated entry by Epstein-Barr virus. *Journal of Virology* 71, 519–526.
- Garcia-Castillo, M.D., Chinnapan, D.J., Lencer, W.I., 2017. Membrane transport across polarized epithelia. *Cold Spring Harbor Perspectives in Biology* 9.
- Gissen, P., Arias, I.M., 2015. Structural and functional hepatocyte polarity and liver disease. *Journal of Hepatology* 63, 1023–1037.
- Gonzalez-Hernandez, M.B., Liu, T., Payne, H.C., *et al.*, 2014. Efficient norovirus and reovirus replication in the mouse intestine requires microfold (M) cells. *Journal of Virology* 88, 6934–6943.
- Gonzalez-Lopez, O., Rivera-Serrano, E.E., Hu, F., *et al.*, 2018. Redundant late domain functions of tandem VP2 YPX3L motifs in nonlytic cellular egress of quasi-enveloped hepatitis A virus. *Journal of Virology* 92, 1308–1318.
- Gruenberg, J., 2020. Life in the lumen: The multivesicular endosome. *Traffic* 21, 76–93.
- Han, J., Pluhackova, K., Böckmann, R.A., 2017. The multifaceted role of SNARE proteins in membrane fusion. *Frontiers in Physiology* 8, 5.
- Hemalatha, A., Mayor, S., 2019. Recent advances in clathrin-independent endocytosis. *F1000Research* 8.
- He, B., Guo, W., 2009. The exocyst complex in polarized exocytosis. *Current Opinion in Cell Biology* 21, 537–542.
- Hirai-Yuki, A., Hensley, L., Whitmire, J.K., Lemon, S.M., 2016. Biliary secretion of quasi-enveloped human hepatitis A virus. *mBio* 7, e01998-16.
- Hou, J., Baker, L.A., Zhou, L., Klein, R.S., 2016. Viral interactions with the blood-brain barrier: Old dog, new tricks. *Tissue Barriers* 4, e1142492.
- Hurley, J.H., 2015. ESCRTs are everywhere. *EMBO Journal* 34, 2398–2407.
- Jackson, W.T., Giddings JR., T.H., Taylor, M.P., *et al.*, 2005. Subversion of cellular autophagosomal machinery by RNA viruses. *PLOS Biology* 3, e156.
- Jiang, W., Ma, P., Deng, L., *et al.*, 2020. Hepatitis A virus structural protein pX interacts with ALIX and promotes the secretion of virions and foreign proteins through exosome-like vesicles. *Journal of Extracellular Vesicles* 9, 1716513.
- Karst, S.M., Wobus, C.E., 2015. A working model of how noroviruses infect the intestine. *PLOS Pathogens* 11, e1004626.
- Kobayashi, N., Takahashi, D., Takano, S., Kimura, S., Hase, K., 2019. The roles of Peyer's patches and microfold cells in the gut immune system: Relevance to autoimmune diseases. *Frontiers in Immunology* 10, 2345.
- Lagaye, S., Derrien, M., Menu, E., *et al.*, 2001. Cell-to-cell contact results in a selective translocation of maternal human immunodeficiency virus type 1 quasiespecies across a trophoblastic barrier by both transcytosis and infection. *Journal of Virology* 75, 4780–4791.
- Lanford, R.E., Feng, Z., Chavez, D., *et al.*, 2011. Acute hepatitis A virus infection is associated with a limited type I interferon response and persistence of intrahepatic viral RNA. *Proceedings of the National Academy of Sciences of the United States of America* 108, 11223–11228.
- Levine, B., Kroemer, G., 2008. Autophagy in the pathogenesis of disease. *Cell* 132, 27–42.
- Maidji, E., McDonagh, S., Genbacev, O., Tabata, T., Pereira, L., 2006. Maternal antibodies enhance or prevent cytomegalovirus infection in the placenta by neonatal Fc receptor-mediated transcytosis. *American Journal of Pathology* 168, 1210–1226.
- Martin-Urdiroz, M., Deeks, M.J., Horton, C.G., Dawe, H.R., Jourdain, I., 2016. The exocyst complex in health and disease. *Front Cell Dev Biol* 4, 24.
- Mazanec, M.B., Kaetzel, C.S., Lamm, M.E., Fletcher, D., Nedrud, J.G., 1992. Intracellular neutralization of virus by immunoglobulin A antibodies. *Proceedings of the National Academy of Sciences* 89, 6901–6905.
- Mcknight, K.L., Xie, L., González-López, O., Chen, X., Lemon, S.M., 2017. Protein composition of the hepatitis A virus quasi-envelope. *Proceedings of the National Academy of Sciences of the United States of America* 114, 6587–6592.
- Mcmahon, H.T., Boucrot, E., 2011. Molecular mechanism and physiological functions of clathrin-mediated endocytosis. *Nature Reviews Molecular Cell Biol* 12, 517–533.
- Merkel, S.F., Andrews, A.M., Lutton, E.M., *et al.*, 2017. Trafficking of adeno-associated virus vectors across a model of the blood-brain barrier: a comparative study of transcytosis and transduction using primary human brain endothelial cells. *Journal of Neurochemistry* 140, 216–230.
- Miranda, J., Martín-Tapia, D., Valdespino-Vázquez, Y., *et al.*, 2019. Syncytiotrophoblast of Placentae from Women with Zika Virus Infection Has Altered Tight Junction Protein Expression and Increased Paracellular Permeability. *Cells* 8.
- Mladinich, M.C., Schwedes, J., Mackow, E.R., 2017. Zika virus persistently infects and is basolaterally released from primary human brain microvascular endothelial cells. *mBio* 8.
- Nagashima, S., Takahashi, M., Jirintai, *et al.*, 2011. A PSAP motif in the ORF3 protein of hepatitis E virus is necessary for virion release from infected cells. *Journal of General Virology* 92, 269–278.
- Ouzilou, L., Caliot, E., Pelletier, I., *et al.*, 2002. Poliovirus transcytosis through M-like cells. *Journal of General Virology* 83, 2177–2182.
- Oztan, A., Rondanino, C., Apodaca, G., 2008. Transcytosis of polymeric immunoglobulin A in polarized Madin-Darby canine kidney cells. *Methods in Molecular Biology* 440, 157–170.
- Parton, R.G., 2018. Caveolae: Structure, function, and relationship to disease. *Annual Review of Cell and Developmental Biology* 34, 111–136.
- Pelkmans, L., Püntener, D., Helenius, A., 2002. Local actin polymerization and dynamin recruitment in SV40-induced internalization of caveolae. *Science* 296, 535–539.
- Ponpuak, M., Mandell, M.A., Kimura, T., *et al.*, 2015. Secretory autophagy. *Current Opinion in Cell Biology* 35, 106–116.
- Pyzik, M., Sand, K.M.K., Hubbard, J.J., *et al.*, 2019. The neonatal Fc receptor (FcRn): A misnomer? *Frontiers in Immunology* 10, 1540.
- Ren, X., Hurley, J.H., 2011. Proline-rich regions and motifs in trafficking: From ESCRT interaction to viral exploitation. *Traffic* 12, 1282–1290.

- Rivera-Serrano, E.E., Gonzalez-Lopez, O., Das, A., Lemon, S.M., 2019. Cellular entry and uncoating of naked and quasi-enveloped human hepatoviruses. *eLife* 8, e43983.
- Robinson, S.M., Tsueng, G., Sin, J., *et al.*, 2014. Coxsackievirus B exits the host cell in shed microvesicles displaying autophagosomal markers. *PLOS Pathogens* 10, e1004045.
- Rothman, J.E., 1987. Transport of the vesicular stomatitis glycoprotein to trans Golgi membranes in a cell-free system. *Journal of Biological Chemistry* 262, 12502–12510.
- Salimi, H., Cain, M.D., Jiang, X., *et al.*, 2020. Encephalitic alphaviruses exploit caveola-mediated transcytosis at the blood-brain barrier for central nervous system entry. *mBio* 11.
- Santiana, M., Ghosh, S., Ho, B.A., *et al.*, 2018. Vesicle-cloaked virus clusters are optimal units for inter-organismal viral transmission. *Cell Host & Microbe* 24, 208–220. [e8].
- Shetty, S., Lalor, P.F., Adams, D.H., 2018. Liver sinusoidal endothelial cells – Gatekeepers of hepatic immunity. *Nature Reviews Gastroenterology & Hepatology* 15, 555–567.
- Siciński, P., Rowiński, J., Warchol, J.B., *et al.*, 1990. Poliovirus type 1 enters the human host through intestinal M cells. *Gastroenterology* 98, 56–58.
- Sixbey, J.W., Yao, Q.Y., 1992. Immunoglobulin A-induced shift of Epstein-Barr virus tissue tropism. *Science* 255, 1578–1580.
- Snooks, M.J., Bhat, P., Mackenzie, J., *et al.*, 2008. Vectorial entry and release of hepatitis A virus in polarized human hepatocytes. *Journal of Virology* 82, 8733–8742.
- Takahashi, Y., Misumi, S., Muneoka, A., *et al.*, 2008. Nonhuman primate intestinal villous M-like cells: An effective poliovirus entry site. *Biochem Biophys Res Commun* 368, 501–507.
- Thery, C., Witwer, K.W., Others, A., 2018. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): A position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *Journal of Extracellular Vesicles* 7, 1535750.
- Tugizov, S.M., Herrera, R., Palefsky, J.M., 2013. Epstein-Barr virus transcytosis through polarized oral epithelial cells. *Journal of Virology* 87, 8179–8194.
- Tuma, P., Hubbard, A.L., 2003. Transcytosis: Crossing cellular barriers. *Physiological Reviews* 83, 871–932.
- Turco, M.Y., Moffett, A., 2019. Development of the human placenta. *Development* 146.
- Turula, H., Wobus, C.E., 2018. The role of the polymeric immunoglobulin receptor and secretory immunoglobulins during mucosal infection and immunity. *Viruses* 10.
- Van Der Grein, S.G., Defourny, K.A.Y., Rabouw, H.H., *et al.*, 2019. Picornavirus infection induces temporal release of multiple extracellular vesicle subsets that differ in molecular composition and infectious potential. *PLOS Pathogens* 15, e1007594.
- Vergés, M., Sebastián, I., Mostov, K.E., 2007. Phosphoinositide 3-kinase regulates the role of retromer in transcytosis of the polymeric immunoglobulin receptor. *Experimental Cell Research* 313, 707–718.
- Vietri, M., Radulovic, M., Stenmark, H., 2020. The many functions of ESCRTs. *Nature Reviews Molecular Cell Biology* 21, 25–42.
- Votteler, J., Ogohara, C., Yi, S., *et al.*, 2016. Designed proteins induce the formation of nanocage-containing extracellular vesicles. *Nature* 540, 292–295.
- Votteler, J., Sundquist, W.I., 2013. Virus budding and the ESCRT pathway. *Cell Host & Microbe* 14, 232–241.
- Wang, T., Li, L., Hong, W., 2017. SNARE proteins in membrane trafficking. *Traffic* 18, 767–775.
- Yang, J.E., Rossignol, E.D., Chang, D., *et al.*, 2020. Complexity and ultrastructure of infectious extracellular vesicles from cells infected by non-enveloped virus. *Scientific Reports* 10, 7939.
- Zhou, F., Wu, Z., Zhao, M., *et al.*, 2019. Rab5-dependent autophagosome closure by ESCRT. *Journal of Cell Biology* 218, 1908–1927.

Further Reading

- Fung, K.Y.Y., Fair, G.D., Lee, W.L., 2018. Transcellular vesicular transport in epithelial and endothelial cells: Challenges and opportunities. *Traffic* 19, 5–18.

Vector Transmission of Animal Viruses

Houssam Attoui and Fauziah Mohd Jaafar, UMR1161 Virologie, INRAE - French National Research Institute for Agriculture, Food and Environment, ANSES, Ecole Nationale Vétérinaire d'Alfort, University of Paris-Est, Maisons-Alfort, France

Rennos Fragkoudis, University of Nottingham, Sutton Bonington, United Kingdom and University of Edinburgh, Edinburgh, United Kingdom

Peter PC Mertens, University of Nottingham, Sutton Bonington, United Kingdom

© 2021 Published by Elsevier Ltd.

Glossary

Anthropogenic factors Refers to effects, processes, objects, or materials that are derived from human activities, as opposed to occurring in natural environments. They can include the effects of human trade, commerce, farming, land use, urbanisation, effects on wild animals and even climate change.

Arboviruses The term “Arboviruses” is an abbreviation of “arthropod borne viruses” and includes viruses that are transmitted between their host species by arthropod vectors.

Epizootic An epizootic is a disease occurrence in a nonhuman animal population that is analogous to an epidemic in humans. It may be restricted to a specific locale (an “outbreak”), can be general (an “epizootic”), or more widespread (“panzootic”).

Haematophagous arthropods These are arthropods that can mediate the transmission of infectious pathogens, and can include, ticks, mosquitoes, sand flies and *Culicoides* midges and biting flies.

Mechanical transmission This describes the simple physical movement of a virus from one infected host to

another susceptible host, usually by an arthropod vector, allowing infection to occur, without infection or replication in the vector.

Naïve A term used to describe animals or animal populations that have not previously come into contact with a pathogen and therefore have no acquired immunity to the disease agent.

Persistent infection Many arboviruses will infect their vectors and vector cells for life, often with little discernible effect on the individual vector. As a result, once infected the vector can usually transmit the virus throughout its life.

Vector capacity This capacity of an arthropod population reflects its ability to transmit the pathogen to the host population, within a specific ecosystem-setting and measures the efficiency of virus transmission between hosts.

Vector competence It is an assessment of the capability of a given vector to transmit a specific virus to its host.

Zoonosis An infection or disease that can be transmitted from animals to humans under natural conditions, also: an infection or disease that can be transmitted between animals and humans (bidirectional zoonosis).

Introduction

The term “Arboviruses” is an abbreviation of ‘arthropod borne viruses’ and includes viruses that are transmitted between their host species by arthropod vectors. For many years after the discovery of yellow fever virus (YFV) by Walter Reed in 1900, the confirmation of Carlos Finlay’s findings that YFV is transmitted by mosquitoes, and the discovery of the first African horsesickness virus by M’Fadyean in 1900, the arboviruses were considered to be a singular unique group. It was even supposed that arboviruses formed a natural taxonomic group.

The arboviruses include viruses that are mechanically transmitted, for example on the mouth parts of haematophagous arthropods, as well as viruses that are ingested, before passing through the arthropod vectors system, before being transmitted in its saliva during feeding on another host. However, in a strict sense, ‘true’ arboviruses are regarded as those agents that infect and replicate in both the host species and the arthropod vector.

In the 1960s and 1970s it became clear that the arboviruses fall into different taxonomic groups. In 1968–1971 the first biophysical, biochemical and morphological characterization of these viruses permitted their initial taxonomic classification. The development of better serological and/or molecular tools, including sequencing and genomics, is at the heart of the identification of different taxonomic groups of viruses and a rapid expansion of the number of identified virus species, genera, families and orders.

Arbovirus vectors include a wide range of arthropods. The first indication that an arthropod could transmit a pathogen to humans, came after the work of Sir Patrick Manson in 1878, when he identified mosquitoes to be the ‘nurse’ or vector of filarial worms.

WHO defines arboviruses as “viruses that share the characteristic of being naturally maintained through biological transmission between susceptible vertebrate hosts by haematophagous arthropods, or transovarial transmission from infected female arthropods to her progeny”. However, the arboviruses also include viruses that can specifically infect and replicate in their arthropod vectors and are then transmitted by them between individual susceptible plants or animals. The plant viruses can be transmitted by a range of vectors including nematodes, fungi and arthropods (aphids, leafhoppers, treehoppers, thrips, whiteflies, mealybugs,

beetles, mites), which can be either persistently or non-persistently infected. The biting-chewing herbivorous insects, like thrips, or beetles and piercing-sucking insects like aphids, are the most common vectors of the plant arboviruses.

This article focuses on the transmission of those arboviruses that infect animals, which includes those transmitted by haematophagous arthropods (ticks, mosquitoes, sandflies, biting midges and flies). Although certain animal arboviruses can be transmitted mechanically, the principal mode of transmission of the animal arboviruses, involves infection of individual vectors by virus ingested as part of a blood-meal from an infected host, followed by virus replication, and after an extrinsic incubation period, the ability to transmit the virus to other hosts (persistent transmission).

The "genomics era" has unravelled some of the complexity of the arboviruses, their evolutionary relationships and their close associations with specific groups of vectors. For example, links were revealed between the flaviviruses and a group of segmented "flavi-like" viruses identified in ticks or mosquitoes such as Jingmeng tick virus and Guaico *Culex* virus (with 4 ssRNA genome segments). Homology was detected particularly in the RNA-dependent RNA polymerase, with the genes of the non-segmented flaviviruses.

To date, at least 500 animal arboviruses have been identified, almost all of which are RNA viruses. However, African swine fever virus (ASFV) is a large dsDNA virus that can be transmitted by *Ornithodoros* ticks, although it can also be transmitted in the absence of ticks by other routes. The animal Arboviruses can be transmitted by ticks, mosquitoes, sandflies or *Culicoides* biting midges, approximately half by mosquitoes, a third by ticks and the rest by sandflies or biting midges. These arboviruses are often dependent on their ability to replicate in two very different species (their vectors and vertebrate hosts) both of which can act as an amplifier and/or reservoir, for their continued transmission and survival.

The arboviruses can have important impacts on human and animal health. They are either epizootic or zoonotic. Emergence and/or re-emergence of the arboviruses is a challenge that humanity will face during the current century. Incursions into novel geographical regions are becoming more frequent. For example, bluetongue disease of ruminants caused by bluetongue virus infection was considered for a long time as a tropical disease, based on its geographical distribution between approximately 40°N and 35°S. However, during the second half of the 20th century, it emerged all around the world, including for the first time in northern Europe, disrupting trade and causing severe economic damage.

Arthropod Vectors of Animal Arboviruses

Except for those arboviruses that can be transmitted mechanically, or by a direct animal to animal route, the animal arboviruses ingested in a contaminated blood meal must infect and further replicate within the arthropod-vector, in order to be transmitted during a later blood meal. Transmission of arboviruses by arthropod vectors can be influenced by a number of variables, which include the vectorial capacity/vector competence of the vector population /individual and the entomological inoculation rates. Vectorial capacity of an arthropod population reflects an ability to transmit the pathogen to the host population within a specific ecosystem-setting and measures the efficiency of a virus transmission between hosts. Vector competence is an assessment of the capability of a given vector to transmit a specific virus. Vectorial capacity is measured by the formula $C = ma^2VP^n / -\log_e P$, where C is the vectorial capacity (the number of infectious bites that the host is subjected to per day), m is the vectors' density relative to the host's density, a is number of vectors feeding on a host divided by the duration of the gonotrophic cycle (gonotrophic cycle: blood-feeding, egg maturation and oviposition) in days, V is the vector competence, P is the daily survival of vectors and n is the extrinsic incubation period. Another variable which helps estimating risk of transmission and disease is the entomological inoculation rate. Entomologic inoculation rate is calculated by multiplying the mosquito biting rate and relative number of infected mosquitoes in a population.

Mosquitoes are major vectors of arthropod-borne infectious diseases and are considered as the most dangerous vectors in the world. There are over 3500 known mosquito species which have been classified within 112 genera. They belong to subfamilies *Anophelinae* and *Culicinae* within family *Culicidae*. Examples of mosquito-borne viruses are shown in Table 1. *Aedes aegypti* and *Aedes albopictus*, are considered as two of the most important vectors of arboviruses in the world.

Originating from Sub-Saharan Africa, *Aedes aegypti* was introduced into Asia in the late 19th century and into the Americas during the slave trade in 1500s–1600s. Up until the late 20th century, *Aedes albopictus* was confined to Asia but has spread to the western hemisphere during a shipment of used tyres from Japan in 1985. These two mosquito species are vectors of important arboviruses, including yellow fever, dengue, Zika, chikungunya and Rift Valley fever (RVF) virus. Over the past thirty years, the distribution of *Aedes albopictus* has expanded to cover almost 40% of the world's landmass. Mathematical modelling predicts that the coming decades will see *Aedes* mosquitoes colonising major cities both in the United States of America and in Europe.

Culex and *Anopheles* mosquitoes are also important vectors for a range of arboviruses such as West Nile virus (WNV), Japanese encephalitis virus (JEV), St. Louis encephalitis virus (SLEV) and Rift Valley fever virus (RVFV). Mosquito vectors transmit a wide range of pathogenic viruses belonging to several genera (*Alphavirus*, *Flavivirus*, *Orthobunyavirus*, *Phlebovirus*, *Orbivirus*, *Seadornavirus*, *Vesiculovirus*). More than 90% of known human arboviruses are transmitted by mosquitoes.

Over the past two decades, insect-only or insect-specific viruses were also found in mosquito populations. These viruses include insect-specific flaviviruses (such as the cell-fusing agent virus and *Culex* flavivirus), insect specific phenuiviruses (such as Jonchet or Ferak virus), the insect-specific reoviruses (such as dinovernaviruses and cypoviruses) and the insect specific alphaviruses (Eilat virus), which can be all transmitted vertically (transovarial transmission) in these insect populations. The Insect-specific viruses cannot infect mammalian systems, and are restricted by a number of factors including temperature, the innate immune responses, or incorrect interaction with specific cell-cofactors. It has been suggested that these viruses could be used to control arbovirus

Table 1 Order of virus containing arboviruses with examples of viruses of medical or veterinary importance

Order	Family	Genus	Species	Biological vector	Genome organisation	Particle structure/symmetry
Asfuvirales	Asfarviridae	Asfivirus	African swine fever asfivirus	Ticks	dsDNA	Internal core, internal lipid layer, icosahedral capsid, outer lipid envelope
Articulavirales	Orthomyxoviridae	Thogotovirus	Thogoto thogotovirus	Ticks	6 ssRNA segments	Enveloped
		Quarantavirus	Quaranfil quarantavirus	Ticks	8 ssRNA segments	Enveloped
Bunyavirales	Nairoviridae	Orthonairovirus	Cribean Congo haemorrhagic fever orthonairovirus	Ticks	3 ssRNA segments	Spherical enveloped
	Peribunyaviridae	Orthobunyavirus	La Cross orthobunyavirus	Mosquitoes	3 ssRNA segments	Spherical enveloped
			Schmallenberg orthobunyavirus	Culicoides spp	3 ssRNA segments	Spherical enveloped
			Akabane orthobunyavirus	Culicoides spp	3 ssRNA segments	Spherical enveloped
			Rift Valley fever Phlebovirus	Mosquitoes	3 ssRNA segments	Spherical enveloped
Amarillovirales	Phenuiviridae	Phlebovirus	Rift Valley fever Phlebovirus	Mosquitoes	3 ssRNA segments	Spherical enveloped
	Flaviviridae	Flavivirus	Yellow fever flavivirus	Mosquitoes	Linear ssRNA	Enveloped, icosahedral nucleocapsid
			Dengue flavivirus	Mosquitoes	Linear ssRNA	Enveloped, icosahedral nucleocapsid
			Japanese encephalitis flavivirus	Mosquitoes	Linear ssRNA	Enveloped, icosahedral nucleocapsid
			Chikungunya alphavirus	Mosquitoes	Linear ssRNA	Spherical, enveloped icosahedral capsid
Martellivirales	Togaviridae	Alphavirus	Sindbis alphavirus	Mosquitoes	Linear ssRNA	Spherical, enveloped icosahedral capsid
			Middleburg alphavirus	Mosquitoes	Linear ssRNA	Spherical, enveloped icosahedral capsid
			Bleutongue orbivirus	Culicoides spp	10 segments of linear dsRNA	Icosahedral non-enveloped
			African horsesickness orbivirus	Culicoides spp	10 segments of linear dsRNA	Icosahedral non-enveloped
			Kemerovo orbivirus	Ticks	10 segments of linear dsRNA	Icosahedral non-enveloped
Reovirales	Reoviridae	Orbivirus	Colorado tick fever coltivirus	Ticks	12 segments of linear dsRNA	Icosahedral non-enveloped
			Eyach coltivirus	Ticks	12 segments of linear dsRNA	Icosahedral non-enveloped
			Banna seadornavirus	Mosquitoes	12 segments of linear dsRNA	Icosahedral non-enveloped
			Nyamanini nyavirus	Ticks	Linear ssRNA	Enveloped, with spherical geometry
			Fukuoka ledantavirus	Culicoides punctatus	Linear ssRNA	Membrane, bullet shaped
Mononegavirales	Nyamiviridae	Nyavirus	Nyamanini nyavirus	Ticks	Linear ssRNA	Enveloped, with spherical geometry
	Rhabdoviridae	Vesiculovirus	Vesicular stomatitis virus	Ticks	Linear ssRNA	Membrane, bullet shaped

replication in the insects in a similar manner as *Wolbachia*. For instance, the cell fusing agent virus (CFAV) interferes with the replication of the 17D vaccine strain of YFV when mosquito cells are infected with CFAV prior to infection with YFV17D (unpublished observations) and the double-stranded (ds) RNA viruses can shut-down replication of single-stranded RNA viruses upon co-infecting the same cell.

Ticks are distributed worldwide. About 900 tick species were identified to date. They are classified into 3 families which are *Argasidae* (soft ticks, 183 species), *Ixodidae* (hard ticks, 683 species) and *Nuttalliellidae* (one species). Ticks are haematophagous ectoparasites which feed on vertebrate hosts, affecting both animal and human health. Ticks are vectors for a wide range of pathogens including bacteria, parasites and viruses. They are second to mosquitoes in terms of their importance as vectors of human arthropod-borne pathogens, but they represent the primary vectors for pathogens in animal health. In order to be transmitted by ticks, a pathogen must be trans-stadially transmitted, thus nymphs and adults are both major vectors. The tick population size is an important criterion which influences vector capacity. Less than 10% of known ticks are involved in field transmission of viruses. These ticks belong to genera *Ornithodoros* and *Argas* (soft ticks) within family *Argasidae* and genera *Ixodes*, *Haemaphysalis*, *Hyalomma*, *Amblyomma*, *Dermacentor*, *Rhipicephalus*, and *Boophilus* (hard ticks) within family *Ixodidae*. Certain ticks can transmit several viral species. One hundred and seventy tick-borne viruses have been identified to date (examples are given in [Table 1](#)).

Culicoides biting midges are haematophagous insects. They are members of the family *Ceratopogonidae*, order *Diptera*. Over 1368 species of *Culicoides* are known to date and they are found in geographical areas ranging from the temperate zones to the tropics. During their lifespan, female *Culicoides* midges can feed multiple times on hosts, allowing them to become infected and transmit arboviruses. The lifespan of biting midges is relatively short and only few adults survive longer than 10 days to 3 weeks. Most of the information known about the vector competence of *Culicoides* midges comes from studies of wild caught insects and laboratory studies mainly with established colonies of *Culicoides sonorensis* and *Culicoides nubeculosus*.

Culicoides midges are important vectors of economically important viral pathogens particularly the orbiviruses. Indeed, viruses belonging to the species *bluetongue virus*, *African horsesickness virus* and *Epizootic haemorrhagic disease virus* are all transmitted by *Culicoides* spp. vectors, although these viruses can also be transmitted vertically in their vertebrate hosts, or by an oral route and can cause fatal infections in certain carnivores. Some recently identified bluetongue viruses (serotypes 25 and 26) can be transmitted in the absence of adult *Culicoides*, which they appear to be incapable of infecting. Orbivirus infection of arthropods has little or no evident effect. Infection in vertebrates can vary from inapparent to fatal, depending on both the virus and the host species, age, immune status and stress factors. Severe BT disease is mostly restricted to certain breeds of sheep and some species of deer.

Depending on the strain, BTV may cause death or a variety of pathologies in sheep. These include haemorrhage, lameness, oedema, a transitory cyanotic appearance of the tongue (giving rise to the species name), nasal and mouth lesions, etc.; still others cause no overt pathology. BTV infection of cattle may show no signs of disease but can involve long-lived viraemias. AHSV and EHDV can cause severe pathology in their respective vertebrate hosts. Mortality rates in serologically naive populations can be over 98% (AHSV). BTV is teratogenic and can cause severe deformities (including dummy calf syndrome) due to a failure of the central nervous system to develop correctly after infection *in utero* in cattle and sheep. *Culicoides* midges are also vectors for orthobunyaviruses that are important in animal health (examples are shown in [Table 1](#)), including Akabane (AKA) and Schmallenberg (SB) virus.

Akabane virus was first isolated in Japan in the 1950s. Vectors of AKAV include *Culicoides imicola*, *C. brevitarsis*, *C. wadai*, and *C. minlei*. The geographical region affected by AKAV expanded to include Asia, the Middle East, Australia and Africa. Pathological manifestations on AKAV infections are observed during pregnancies leading to congenital malformations of the foetus in ruminants or encephalitis in new-borne calves. SBV was first identified in 2011 and spread rapidly throughout Europe. The virus has been reported in cattle, sheep and goats. It is believed to be transmitted by a range of *Culicoides* including *C. imicola*, *C. obsoletus*, *C. punctatus*, *C. dewulfi*, *C. lupicaris*, *C. newsteadi*, *C. nubeculosus*, *C. pulicaris*. SBV is responsible for congenital malformations of the foetus, stillbirths and failed pregnancies in cattle, sheep and goats. Adult *Culicoides* are also vectors of Ganjam virus in India (an isolate of *Nairobi sheep disease virus*, *Orthonairovirus*).

A biting midge identified as *Culicoides anophelis* has been documented since the 1950s in China and India. This biting midge feeds on mosquitoes of the genera *Aedes*, *Anopheles*, *Armigeres* and *Culex*. The wide host range of this midge makes it ideal to act as a biological vector, spreading viruses among various mosquito populations.

Sandflies which transmit viruses belong to genera *Lutzomyia* and *Phlebotomus* within subfamily *Phlebotominae*, in family *Psychodidae*. Sandflies of the genus *Phlebotomus* transmit important human pathogens such as Sandfly fever Naples virus (Mediterranean regions) or Chandipura virus (India). Sandflies of the genus *Lutzomyia* transmit arboviruses such as vesicular stomatitis virus (*Vesiculovirus*, *Rhabdoviridae*) or Changuinola virus (*Orbivirus*, *Reoviridae*).

Stomoxys flies transmit a number of viruses mechanically, particularly on their mouth parts and potentially by regurgitation of contaminated blood upon an onward blood meal. They have been incriminated in the mechanical transmission of a number of viruses including African swine fever, Equine infectious anemia (EIAV), West Nile and Rift Valley fever virus.

Co-Evolution of Viruses and Vectors: Vector-Only, Vector-Borne Viruses and Non-Vectored Viruses

There are different views as to whether arboviruses have co-evolved, or not, with their arthropod vectors. Looking at the genomes of the arboviruses in genera where tick-borne, mosquito-borne, sandfly-borne and/or *Culicoides*-borne viruses are encountered

(such as the flaviviruses or the orbiviruses) different observations are made. For the flaviviruses, the genomes of the insect-only viruses have a G + C of 50%–53%, the mosquito-borne flaviviruses have 47%–53%, and the tick-borne viruses 53%–54%. The “no known vector” flaviviruses (NKV) have a G/C content of 43%–48%, with the exception of Tamana bat virus (TABV) where the G + C content is ~38%. In the orbiviruses, the G + C content of the mosquito-borne orbiviruses range from 35% to 41%. The G + C content of the *Culicoides*-borne viruses range from 40% to 45% and the tick-borne orbiviruses have a G + C content of 52%–58%. The G + C content of arthropod vector genomes is ~56% for ticks, ~39%–42% for *Culicoides* and 35%–38% for mosquitoes. There are clear differences between the two genera. Only the tick-borne flaviviruses have a G + C content that is similar to that of their tick vectors (~57%). The G + C content of the insect only flaviviruses is very similar to that of tick-borne viruses and the G + C content of the mosquito-borne flaviviruses is largely overlapping with that of both the insect-only and tick-borne flaviviruses. This contrasts with orbiviruses where G + C content is very similar to that of their respective arthropod vector groups. These observations, among other findings, suggest that the orbivirus ancestors have coevolved with their vectors and may have originally been arthropod viruses that have adapted to their current vertebrate hosts.

Phylogenetic studies have classified the flaviviruses within four groups: two mosquito-borne groups, a tick-borne group and a group where viruses are considered as non-vectorized or as having no known vectors (NKV). Phylogenetic studies suggest that the mosquito-borne flaviviruses ‘root’ all of the other flaviviruses. Other studies suggest that NKV flaviviruses do not form a single phylogenetic group and that some of the NKV such as *Yokose* (YOK), *Sokoluk* (SOK), and *Entebbe bat* (ENT) virus cluster with the mosquito-borne flaviviruses, while other NKV viruses cluster with the tick-borne flaviviruses. It has also been suggested that YOKV, SOKV and ENTV are likely arboviruses, based dinucleotide frequencies in their genomes, as they do not seem to be specifically adapted to vertebrate hosts. Only *Tamana bat virus* (TABV) seems to be a vertebrate-specific virus, having a significantly lower CpG dinucleotide frequency than the insect-specific flaviviruses or arthropod/vertebrate flaviviruses.

Whether flaviviruses co-evolved with their arthropod vectors (i.e., originated before the split of invertebrates and vertebrates during evolution) or they originated in one or other group of hosts, followed by co-adaptation, remains debatable. Yet the high G + C content of most flavivirus genomes (including the insect-only viruses) raises questions as to whether the flaviviruses originated in ticks, or if this trait provided an advantage for the dispersal of flaviviruses among various arthropod vectors.

Certain insect-borne viruses are capable of infecting and replicating in tick cell lines. Semliki Forest virus (SFV) and Venezuelan equine encephalitis virus (VEEV), two mosquito-borne alphaviruses, can replicate well in tick cell lines *Rhipicephalus decoloratus* BDE/CTVM16 and *Rhipicephalus appendiculatus* RAE/CTVM1 cells. The G + C content of SFV genome is ~53% and that of VEEV is ~50%. Replication of 13 flaviviruses including Dengue virus (DENV), WNV, SLEV, YFV, tick-borne encephalitis virus (TBEV) Powassan virus (POWV), Louping ill virus (LIV), Negishi virus (NGV) and Langat virus (LGTV) was compared in mosquito and tick cell lines. The mosquito-borne viruses replicate efficiently in the mosquito cells and it was shown that WNV is capable of infecting and replicating in four tick cell lines. The tick-borne viruses replicate only in tick cell lines, except LGTV where signs of infection were observed in C6/36 cells. A previous study using Singh’s non-cloned *Ae. albopictus* cells (from which C6/36 were derived) failed to show replication of LGTV. NKV flaviviruses do not replicate in any arthropod cell lines. The G + C contents of DENV, SLEV and YFV are comprised between 45% and 50% whereas those of WNV, TBEV, POWV and LIV are up to 51%. Larvae and nymphs, but no adults, of *Amblyomma* species were shown experimentally to be susceptible for infection by WNV or SLEV. These results highlights important differences between different flaviviruses in terms of their capacity to potentially infect a distinct arthropod vectors.

Colorado tick fever virus (CTFV) and *Eyach virus* (EYAV) are two species within the genus *Coltivirus* (family *Reoviridae*). Replication of CTFV or EYAV in mosquito and/or tick cell lines has been explored. CTFV was grown in *Ae. albopictus* cells over 7 weeks and the titre reached ~10⁶ pfu/ml after 6 weeks, declining afterwards. As compared to titres of the same virus grown in mammalian cells (~10⁸ pfu/ml), the difference represents approximately 99% less virus in mosquito cells. Limited replication of EYAV-Fr578 was shown in C6/36 cells. A strain of Kemerovo virus (KEMV, genus *Orbivirus*) was grown both in *Ae. albopictus* C6/36 cells and *I. ricinus* tick cells. While titres of approximately 10⁷–10⁸ pfu/ml were observed in tick cells, titres of 10⁴ or less were obtained with C6/36 (unpublished data), that is to say 99.9%–99.99% less virus in mosquito cells. In summary, the lower levels of replication of a given animal arbovirus in a heterologous cell culture system suggest that a heterologous arthropod would not act as a likely vector for that particular virus.

Environmental/Anthropogenic Factors Contributing to Vector-Borne Virus Spread

Emerging/re-emerging viruses have been responsible for severe economic losses and heavy losses in terms of human lives. Over the past four decades the world has seen a number of epidemics, epizootics and pandemics which are thought to be due to emerging viruses.

The SARS-CoV2 pandemic, which has caused almost 1.2 million deaths to date, is one the most serious recent examples, yet human immunodeficiency virus remains the most devastating in terms of number of fatalities. Besides these examples, the bulk of emerging viruses are arboviruses. Their spread is linked to a number of factors and in particular, anthropogenic factors are among the most important including international trade and travel (movement of populations, animals and commercial goods), deforestation, afforestation, urbanisation and climate change induced by human activities.

The emergence of specific viruses can be facilitated by optimal climate conditions and the presence of local competent vectors. Tangible examples of arbovirus emergence are the epidemic of Chikungunya virus (CHIKV) in Italy in 2007 and south-eastern

France in 2010 and the return of autochthonous circulation of this virus in both countries in 2017. The Zika virus emerged causing an outbreak in the federated states of Micronesia in 2007 and in Brazil in 2015, while DENV emerged in India in 2015. Perhaps one of the most striking examples of economic damage caused by an arbovirus is bluetongue virus introduction into Europe. During the 20th century, BTV was endemic in sub-Saharan Africa. Since 1998, multiple introductions of BTV have occurred, initially in southern Europe, culminating in the BTV-8 outbreak during the summer of 2006 in Northern Europe, where transmission by the local midge populations, led to a devastating disease in the naive sheep and bovine herds in the region.

Replication and Transmission of Arboviruses in Their Arthropod Vector and the Role of Arthropods in Maintenance and Spread of the Virus

Arthropod vectors usually become persistently infected by the arboviruses they transmit (leading to persistent transmission). This is usually designated as “biological transmission” (persistent circulative propagative transmission in haematophagous arthropods). However, some viruses can be transmitted mechanically by biologically “non-competent” vectors.

Mechanical transmission of animal arboviruses by biologically non-competent arthropods shares some similarities with the non-circulative transmission of plant arboviruses. Both are non-persistent forms of transmission which are lost at a certain point after acquisition of the virus. However, specific interactions have been suggested between the animal virus and a haematophagous arthropod during mechanical transmission.

The poxviruses are poorly transmitted mechanically. Lumpy skin disease virus (LSDV) has been shown to be transmitted mechanically by *Aedes aegypti*. This mosquito harbours LSDV for up to 7 days. A distinct study showed that LSDV is not transmitted mechanically to susceptible animals by other insects including *Anopheles stephensi*, *Stomoxys calcitrans*, *Culicoides nubeculosus*, or *Culex quinquefasciatus*, although from a historical point of view *Stomoxys* flies were considered to play an important role in transmission. However, sheep and goat poxvirus, which is related to LSDV has been shown to be mechanically transmitted to susceptible hosts by *Stomoxys*. In summary, several studies suggest that transmission of LSDV by *Aedes aegypti*, is due to a far more elegant mode of transmission than a mere “dirty-pin” virus transfer. There have been suggestions that this parallels transmission of some plant viruses where specific associations and molecular interactions have been observed. *Stomoxys* flies can also mechanically transmit other animal arboviruses (ASFV, EIAV, WNV, RVFV).

Persistent transmission can occur when a virus crosses the gut wall, with or without replication within the gut cells, into the haemolymph, ending in salivary glands. Circulative non-propagative transmission is exclusively observed with plant viruses, in particular those belonging to families *Geminiviridae*, *Nanoviridae*, and *Luteoviridae*. It is believed that no known animal arboviruses use this mode of transmission. More than one study concluded the absence of replication of LSDV in *Amblyomma* and/or *Rhipicephalus* ticks, or derived cell lines. However, the virus was detected in the saliva of *Amblyomma hebraeum* and *Rhipicephalus appendiculatus*. Using anti-LSDV monoclonal antibodies, viral antigens were detected in the salivary glands, haemocytes, ganglia, ovaries, testes, fat bodies, and midgut of nymphs and adult ticks fed on LSDV infected cattle. Furthermore *R. appendiculatus* males (from a laboratory bred colony) fed on viremic cattle were found to transmit the virus to healthy cattle. Finally, of LSDV failed to replicate in *Rhipicephalus* tick cell lines. In conclusion, LSDV can be transmitted by *Rhipicephalus*, does not replicate in tick cells, even though LSDV antigen was detected in several tick organs. This situation is clearly distinct from simple mechanical transmission and points towards a form of circulative non-propagative transmission.

In biological transmission of animal viruses or the circulative propagative transmission of plant viruses, the pathogen is ingested as part of the meal (blood or sap). The virus then infects the midgut epithelium (single layer of cells) and replicates in the midgut cells, spreading into the haemocoel and other organs/tissues, including the salivary glands, where it undergoes further replication. The virus passes into the saliva of the vector and is delivered to the host upon feeding. While in insects blood meals are digested in the lumen of the gut, ticks are heterophagous, meaning that digestion of blood is intracellular within the midgut cells. Plant viruses which are transmitted by a circulative propagative mode, include those belonging to families *Reoviridae*, *Rhabdoviridae* and *Tospoviridae*. Like the animal arboviruses which are biologically transmitted by their haematophagous arthropods, these plant viruses replicate in their insect vectors.

The virus must cross/escape several barriers in the arthropod. The midgut in particular is the site of entry into the biologically competent vector. The “gut barrier”, is a major determinant of vector competence, controlling the virus both qualitatively and quantitatively and defines the “permissivity” and vector competence of an arthropod for a given virus. When environmental conditions favour rapid development of the arthropod exoskeleton, while internal organs are not fully developed, a ‘leaky gut’ may occur leading to viruses bypassing the gut barrier. Replication within midgut cells does not suffice for a virus transmission by haematophagous arthropods. The virus must also disseminate to the other tissues in the vector, in particular to the salivary glands.

In studies conducted with yellow fever virus vaccine strain YFV-17D in *Aedes aegypti* mosquitoes, it has been shown that the virus replicates in the mid gut cells (with 10 times less viral RNA of YFV-17D produced, compared to wild type YFV) but fails to disseminate onwards. Upon intrathoracic injection of mosquitoes with YFV-17D, the virus efficiently infects secondary tissues including legs and salivary glands. Thus, the outcome of the interactions between the virus and the gut barrier determines the chances of dissemination and onward transmission. Not all vaccine arbovirus strains are non-transmissible by arthropod vectors. For example, *Culicoides* midges can transmit live-attenuated vaccines of the orbiviruses. Although attenuated, these vaccines act like wild type viruses in terms of their transmission from inoculated (vaccinated) to healthy animals and have been shown to persist, become disseminated and can cause infection/disease in the field.

The virus quasispecies found in infectious blood meals are subjected to a “bottleneck” in the midgut and it is supposed that infection of the midgut will require a relatively high titre of virus to infect the midgut cells. Disseminated viruses from the midgut are subjected to a second bottleneck in the salivary glands, due to particular genotypes of the virus being maladapted to salivary glands, or the viral population is subjected to non-selective reduction.

Identification of receptors of arboviruses on gut cells and other organs is work in progress. Data concerning receptors of arboviruses in their arthropod vectors come mainly from cell culture studies. Poorly characterized receptors have been identified, such as those for DENV (polypeptides ranging from 40 to 80 kDa) and VEEV (a major polypeptide of 32 kDa). For DENV one of the most promising potential receptors in mosquito cells is a 35 kDa protein known as prohibitin. Treating insect cells with anti-prohibitin antibodies or silencing its expression using siRNAs, significantly reduced DENV binding. Yet arthropod cell cultures are polyclonal with poorly characterized tissue origins. Receptors of CHIKV have been identified in larvae and adults of *Aedes aegypti* to be in the midgut brush border membrane. They were identified as glycoproteins ranging in size from 24 to 62 kDa. Identified receptors in arthropod cells are distinct from their counterparts in vertebrate cells. The outer capsid proteins (VP2 and VP5) of bluetongue virus are responsible for cell attachment and infection of mammalian cells. However, in *Culicoides* cell cultures, or adult *Culicoides* (via an oral route), proteolytic cleavage of VP2 can significantly enhance the infectivity of the virus. Complete removal of these outer capsid proteins massively reduces or abolishes the infectivity of the resulting BTV core particles for the mammalian BHK or CHO cells respectively. However, BTV core particles retain the same infectivity as the intact virus particles for *Culicoides* cells, suggesting the availability of alternative receptors and cell entry mechanisms in the insect vectors.

Viruses must survive the arthropods’ behavioral characteristics. In contrast to mosquitoes that can take several blood meals during the same life stage (adult female), hard ticks feed only once before moulting, meaning that viruses need to survive moulting, in order to persist into the next life stage and the opportunity to be transmitted. Several months can pass between two tick life-stages. Tick-borne viruses can persist by infecting tissues which do not undergo histolysis during moulting. Some tick-borne viruses can also be transmitted transovarially and/or trans-sexually to the next generation. Vertical transmission in ticks and insects and transmission by co-feeding between ticks contribute to the natural survival and spread of the arboviruses.

The local abundance of vectors and the genetics of the virus can both influence transmissibility, as seen for example when CHIKV switched vectors from *Aedes aegypti* to *Aedes albopictus* in 2004. A single amino acid substitution A226V in the E1 envelope glycoprotein of CHIKV helped the virus to move to its new vector. The mutation in the E1 glycoprotein lowers the threshold of viraemia necessary to infect *Ae. albopictus* mosquitoes. A change at this position in Semliki Forest virus (P226S) was previously reported to free the virus from dependence on cholesterol for cell entry.

Certain arboviruses can be sexually (horizontally) transmitted between arthropods. The male mosquitoes acquire the virus from an infected female. Upon mating, the male further transmits the virus to non-infected females, in which the virus infects the oocytes, leading to vertical transmission to the offspring. This was shown for La Crosse virus in *Aedes triseriatus* and Dengue and Zika virus in *Aedes albopictus*.

Vertical transmission has been proposed as a maintenance mechanism during adverse conditions for some arboviruses. Dengue, Japanese encephalitis and St. Louis encephalitis virus are vertically transmitted when the fully formed egg of *Aedes albopictus* is oviposited. Vertical transmission of RVFV (*Phlebovirus*) was observed through the eggs of *Aedes* spp (subgenus *Neomelanimon*). It has been suggested that this mechanism, known as trans-ovum transmission could be common to all mosquito-borne flaviviruses. Transovarial transmission, a variant form of vertical transmission, occurs when germinal cells of a female mosquito are infected with a given virus, hence the majority of offspring will be infected. The lifespan of mosquitoes is dependent on different factors. Male mosquitoes are not haematophagous and feed on plant nectar. They live up to 7 days on average. The average lifespan of a female mosquito is around 6 weeks. Female mosquitoes are haematophagous and depending on abundance and adequacy of hosts, female mosquitoes can live up to 5 months.

Virus–Vector Interactions and Role of the Arthropod Innate Immune Response

Virus–vector interactions dictate the outcome of infections, replication, dissemination and onward transmission.

Salivary components facilitate vasodilation in order to promote blood flow, promote replication, suppress host immunity and pain and prevent blood coagulation and haemostasis. Arthropod saliva increases disease severity. The physiological activities of salivary components have been linked to promoting viral replication and dissemination and pathogenesis in the host bitten by the arthropod. Mosquito saliva promotes a TH2 immune response while induction of this type of response is less prominent by tick saliva. Mosquito saliva suppresses type-I IFN signalling in the host thus promoting replication. Enhancement of replication of arboviruses by arthropod salivary components was observed with orthobunyaviruses (La Crosse encephalitis and Cache Valley virus), phleboviruses (RVFV), ASFV, and flaviviruses (DENV). The sensitivity response to insect bites attracts cellular components of the immune system to the site of biting and infection. Bluetongue virus, can infect these cells, promoting virus replication and the early dissemination within the vertebrate host.

Arboviruses seem to contribute to the modification of arthropods’ behavior. For example, *Aedes aegypti* infected with DENV experience longer probing times and feed for longer durations. This contrast with studies of RVFV-infected *Culex pipiens*, where mosquitoes with non-disseminated infection feed for significantly longer times than mosquitoes with fully disseminated infection.

Determinants of the vector competence include the innate immune response within the infected arthropod. These responses include RNA interference, Jak-STAT pathway (Vago), Nf-κB, Imd and Toll and autophagy. RNA interference is the predominant

pathway which controls virus replication in infected arthropod cells. Double-stranded RNA (dsRNA) triggers the RNAi response. Long dsRNA is processed by the mediator Dicer-2 which cleaves the dsRNA into 21–23 long short interfering RNAs designated siRNA or viral interfering RNAs (viRNAs). The guide strand of the viRNA binds to its foot-print in the viral mRNA. Upon binding, the dsRNA region is processed by Ago-2 which cleaves the mRNA inactivating it. RNAi is a common response in various arthropods. Double-stranded RNA presence in cells indicates viral replication. viRNAs have been detected in arthropod infected cells with a wide range of viruses as diverse as CHIKV (*Alphavirus*), DENV (*Flavivirus*), or BTV (*Orbivirus*) and St. Croix River virus (*Orbivirus*).

Environmental conditions such as the extrinsic temperature influence the RNAi response. Thus, lower temperatures supposedly reduce the efficacy of the RNAi response. For single stranded RNA viruses, the virus replication mechanism includes the formation of a dsRNA replication intermediate. Virus with dsRNA genomes have their dsRNA sheltered at all times within the core particles, shielding it from detection by the innate immune system. Unlike plants, which express RNA-dependent RNA polymerase (RDRP), insects lack an RDRP. The RDRP is necessary for the amplification of the RNAi response, by continuously generating viRNAs. dsRNA intermediates of ssRNA viruses may be accessible to Dicer, yet for dsRNA viruses where the genome is sheltered in the compact core particles, an accessible form of dsRNA might be generated from positive sense mRNA by the viral RDRP.

The genomes of arboviruses can become incorporated as cDNA copies into the genomes of their arthropod vectors. There is potential for these copies to act as sources of viral RNA sequences that can initiate a silencing response to infecting virus. If these sequences become incorporated into the arthropod germ line they can in practice act to generate a trans generational immune memory response.

Facing this efficient innate immune response, arboviruses have evolved strategies to counter RNAi response. One of the well-known strategies is the synthesis of decoy RNAs such as the flavivirus (such as DENV and WNV) sRNAs which interact with the RNAi machinery, sequestering RNAi mediators such as Dicer and Ago. Plant viruses have evolved other strategies by expressing proteins designated as viral suppressors of RNA silencing (VSRs), which bind to and sequester viRNAs from the RNAi machinery. Many arbovirologists do not support the presence of such a mechanism in animal arboviruses, despite that NS4B of DENV has been shown to act as a VSR using a reporter gene assay.

Other innate immune pathways include Jak-STAT, Toll and IMD. These pathways have been shown to be activated in response to Sindbis virus and DENV in *Ae. Aegypti* and WNV in *Culex pipiens*. Infection of *Aedes albopictus*-derived U4.4 mosquito cells by SFV activated STAT/IMD pathway, but not Toll. CHIKV repressed the activity of Toll signalling pathway in vitro. Neither JAK/STAT, IMD nor Toll pathways were found to mediate antiviral activities to CHIKV infection.

BTV replicates efficiently in *Culicoides* biting midges or derived cells lines. Until 2008 there were 24 known serotypes of BTV. However, since 2008 several novel serotypes have been identified some of which do not replicate in live adult *Culicoides* or a derived cell line. We have identified that their restriction in *Culicoides* cells is mediated by 4 genomes segments (segments 1,2,3 and 7) each of which individually restricts replication when included by reverse genetics in the backbone of a vector transmitted BTV. This group of non-vectored BTVs is transmitted horizontally between animals by direct contact, likely by an oral-fecal mechanism.

Co-Infection of Arthropods With Symbiotic Bacteria and How It Influences Virus Replication

Wolbachia is an intracellular gram-negative bacteria which was first identified in 1924 in *Culex pipiens*. It naturally infects more than 65% of arthropod species and *Wolbachia* genetic lineages are typically associated with a single host species. Despite infecting their hosts, a number of *Wolbachia* strains protect their host against several viruses including DENV, CHIKV, WNV and ZIKV in mosquitoes, *Drosophila C virus* (DCV) and *flock house virus* in *Drosophila*, although the mechanisms involved are not known. *Wolbachia* has been assessed as a novel alternative biological control agent in mosquitoes, to limit their ability to transmit flaviviruses and potentially reduce their life span. Of the two main dengue vectors, only *Ae. albopictus* has a natural *Wolbachia* infection in the wild, with two strains: wAlbA and wAlbB. No natural association is known of *Wolbachia* with *Ae. Aegypti* mosquito (primary species responsible for transmitting DENV). However, there are currently three *Wolbachia* transinfections of *Ae. aegypti*, with wAlbB (from *Aedes albopictus*) and another two strains wMel and wMelPop (from *Drosophila melanogaster*). These strains have been used to infect a variety of mosquitoes, and the resulting host manipulations are diverse, with some proving useful for mosquito control purposes. Strain wMel is a non-virulent strain while wMelPop is a virulent strain shortening the life in *Aedes aegypti*. Both wMel and wMelPop successfully inhibited the ability of *Ae. Aegypti* to transmit DENV (replication and dissemination); the more virulent wMelPop showed higher efficacy in limiting DENV replication.

There are two potential explanations as to how *Wolbachia* interferes with virus replication. The first being a competition for host resources between *Wolbachia* and viruses. Previous studies showed that *Wolbachia* modulates the host cholesterol levels in *D. melanogaster* which leads to protection against DCV. *Wolbachia* blocks less effectively virus replication when cholesterol levels are significantly higher in *Drosophila*. During DENV infection, specific lipid classes become enriched in the mosquitoes. *Wolbachia* causes certain lipid classes to be depleted and likely produces cellular lipid environment that inhibits replication of mosquito-borne viruses.

The second explanation involves activation of immune genes. *Wolbachia* induces reactive oxygen species (ROS)-dependent activation of the Toll pathway. Oxidative stress is induced upon infection with *Wolbachia* and an increased level of reactive oxygen is observed. Antimicrobial peptides defensins and cecropins are activated by ROS and are involved in the inhibition of DENV replication. *Wolbachia* downregulates chromodomain helicase DNA-binding protein of *Ae. aegypti*. Previous studies showed

significant downregulation of AeCHD7 in the presence of *Wolbachia* in female mosquitoes only. Levels of AeCHD7 are significantly increased during DENV infection, while depletion of AeCHD7 significantly reduces DENV replication.

Further studies identified induction of Vago by *Wolbachia* wMelPop as a mechanism of inhibition of DENV replication in *Aedes aegypti* cells. *Wolbachia* suppresses the replication of WNV in *Culex quinquefasciatus*, that of DENV, CHIKV, and ZIKV in *Ae. Aegypti* but has no significant effect of YFV in *Ae. aegypti* or WNV in *Culex tarsalis*.

Role of the Vector in the Genetic Diversity of Animal Arboviruses

Mosquitoes, sandflies, *Culicoides* and ticks do not have the same lifecycles and lifestyles. The tick-borne viruses are longer associated with their vectors than the insect-borne viruses, due to a much longer vector life span. The arbovirus subpopulations that are generated during infection of mammals/arthropods are likely to influence the outcome of infection and onward transmission. Although alternation between host and vector constrains the evolution of the majority of arboviruses, serially passaged animal arboviruses in their vectors leads to host specialisation. For example, VEEV serially passaged in live mosquitoes, induced low levels of virameia in mice after feeding on the animals. However, the serially passaged virus resulted in higher rates of infection in mosquitoes, as compared to the first passage and had a single synonymous substitution in the non-structural protein nsP1 (viral capping enzyme). The frequency of the viral population having this substitution increased by the eight passage becoming frequent at the tenth passage.

Serial passages of RVFV in *Aedes aegypti* mosquito cells led to deletion of large parts of the virulence factor NSs protein and this was observed by the 10th passage. This virus was non-virulent in mice acting like a protective vaccine upon challenge with wild type RVFV.

Hyalomma marginatum tick-adapted TBEV (17 serial passages) is less virulent to mice than the parental strain. Fifteen nucleotide substitutions, six of which were non-synonymous (resulting in changes in protein sequences for E, prM, NS2A and NS4A) were identified in the genome of the tick-adapted virus. Changes were also linked to stronger/lower affinities of the virus to cellular heparan sulphate and the binding of virions to cellular glycosaminoglycans.

The genome plasticity of CCHFV was assessed in a tick vector and an animal host. Nymphs of *Hyalomma marginatum* tick were fed on CCHFV-infected mice and conserved for one year after moulting into the adult stage. NGS sequencing identified substitutions only in CCHFV from ticks but not mice. Mutations were distributed on all three genome segments and were common to more than one tick. The genetic diversity of CCHFV is larger in ticks than in mice.

Passaging orbiviruses sequentially in mammalian cells has been the basis of attenuating these viruses. For instance, serially passaged BTV in BHK-21 cells significantly reduces its virulence in mice. Passaging this virus once in cells derived from the embryos of *Culicoides* midges is enough to restore virulence and this has been associated with expanding the genetic variability in insect cells.

Further Reading

- Agboli, E., Leggewie, M., Altinli, M., Schnettler, E., 2019. Mosquito-specific viruses-transmission and interaction. *Viruses* 11 (9), 873.
- Baldacchino, F., Muenworn, V., Desquesnes, M., *et al.*, 2013. Transmission of pathogens by Stomoxys flies (Diptera, Muscida). *Parasite* 20, 26.
- Beerntsen, B.T., James, A.A., Christensen, B.M., 2000. Genetics of mosquito vector competence. *Microbiology and Molecular Biology Reviews* 64 (1), 115–137.
- Bellone, R., Failloux, A.B., 2020. The role of temperature in shaping mosquito-borne viruses transmission. *Frontiers in Microbiology* 11, 584846.
- Benelli, G., 2020. Manipulating tick behavior-through a glass, darkly. *Pathogens* 9 (8), 664.
- Bhowmick, B., Han, Q., 2020. Understanding tick biology and its implications in anti-tick and transmission blocking vaccines against tick-borne pathogens. *Frontiers in Veterinary Science* 7, 319.
- Blair, C.D., Olson, K.E., 2015. The role of RNA interference (RNAi) in arbovirus-vector interactions. *Viruses* 7 (2), 820–843.
- Blitvich, B.J., Firth, A.E., 2017. A review of flaviviruses that have no known arthropod vector. *Viruses* 9, 1–25.
- Ciota, A.T., Lovelace, A.O., Ngo, K.A., *et al.*, 2007. Cell-specific adaptation of two flaviviruses following serial passage in mosquito cell culture. *Virology* 357, 165–174.
- Conway, M.J., Colpitts, T.M., Fikrig, E., 2014. Role of the vector in arbovirus transmission. *Annual Review of Virology* 1 (1), 71–88.
- Girard, M., Nelson, C.B., Picot, V., Gubler, D.J., 2020. Arboviruses: A global public health threat. *Vaccine* 38 (24), 3989–3994.
- Hemati, B., Contreras, V., Urien, C., *et al.*, 2009. Bluetongue virus targets conventional dendritic cells in skin lymph. *Journal of Virology* 83 (17), 8789–8799.
- Johnson, K.N., 2015. The impact of *Wolbachia* on virus infection in mosquitoes. *Viruses* 7, 5705–5717.
- Lawrie, C.H., Uzcátegui, N.Y., Armesto, M., Bell-Sakyi, L., Gould, E.A., 2004. susceptibility of mosquito and tick cell lines to infection with various flaviviruses. *Medical and Veterinary Entomology* 18, 268–274.
- Lequime, S., Paul, R.E., Lambrechts, L., 2016. Determinants of arbovirus vertical transmission in mosquitoes. *PLoS Pathogens* 12 (5), e1005548.
- Liang, G., Gao, X., Gould, E.A., 2015. Factors responsible for the emergence of arboviruses: Strategies, challenges and limitations for their control. *Emerging Microbes and Infection* 4 (3), e18.
- Madison-Antenucci, S., Kramer, L.D., Gebhardt, L.L., Kauffman, E., 2020. Emerging tick-borne diseases. *Clinical Microbiology Reviews* 33 (2), e00083.
- Mertens, P., Baylis, M., Mellor, P., 2008. *Bluetongue*. London: Academic Press.
- Migné, C.V., Moutailler, S., Attoui, H., 2020. Strategies for assessing arbovirus genetic variability in vectors and/or mammals. *Pathogens* 9, 915.
- Mohd Jaafar, F., Belhouchet, M., Belaganahalli, M., *et al.*, 2014. Full-genome characterisation of Orungo, Lebombo and Changuinola viruses provides evidence for co-evolution of orbiviruses with their arthropod vectors. *PLoS ONE* 9 (1), e86392.
- O'Neal, S.T., Samuel, G.H., Adelman, Z.N., Myles, K.M., 2014. Mosquito-borne viruses and suppressors of invertebrate antiviral RNA silencing. *Viruses* 6 (11), 4314–4331.
- Ölund, P., Lundén, H., Blomström, A.E., 2019. Insect-specific virus evolution and potential effects on vector competence. *Virus Genes* 55, 127–137.
- Osei-Poku, J., 2012. The evolution and genetics of vector competence in mosquito disease vectors. PhD Thesis, Clare College, University of Cambridge.
- Purse, B.V., Mellor, P.S., Rogers, D.J., *et al.*, 2005. Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews in Microbiology* 3 (2), 171–181.

- Rückert, C., Ebel, G.D., 2018. How do virus-mosquito interactions lead to viral emergence? *Trends in Parasitology* 34 (4), 310–321.
- Sick, F., Beer, M., Kampen, H., Wernike, K., 2019. Culicoides biting midges – Underestimated vectors for arboviruses of public health and veterinary importance. 11 (4), 376.
- Sim, S., Jupatanakul, N., Dimopoulos, G., 2014. Mosquito immunity against arboviruses. *Viruses* 6 (11), 4479–4504.
- Tassetto, M., Kunitomi, M., Whitfield, Z.J., *et al.*, 2019. Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements. *eLife* 8, e41244.
- Vasilakis, N., Deardorff, E.R., Kenney, J.L., *et al.*, 2009. Mosquitoes put the brake on arbovirus evolution: Experimental evolution reveals slower mutation accumulation in mosquito than vertebrate cells. *PLoS Pathogens* 5 (6), e1000467.
- Young, P.R., 2018. Arboviruses: A Family on the move. *Advances in Experimental Medicine and Biology* 1062, 1–10.

Relevant Websites

<http://btv-glue.cvr.gla.ac.uk/#/home>

BTV-GLUE: University of Glasgow.

<https://talk.ictvonline.org/>

International Committee on Taxonomy of Viruses (ICTV).

<http://viperdbscripps.edu/>

VIPERdb.

The Human Virome

Alexia Bordigoni, Aix-Marseille University, CNRS, IRD, Mediterranean Institute of Oceanography, Marseille, France and Aix-Marseille University, IRD 257, Assistance-Publique des Hôpitaux de Marseille, UMR Microbes, Evolution, Phylogeny and Infections (MEPHI), IHU Méditerranée Infection, Marseille, France

Sébastien Halary, National Museum of Natural History, UMR 7245 CNRS/MNHN Molécule de Communication et Adaptation des Micro-organismes, Paris, France

Christelle Desnues, Aix-Marseille University, University CNRS, IRD, Mediterranean Institute of Oceanography, Marseille, France and Aix-Marseille University, IRD 257, Assistance-Publique des Hôpitaux de Marseille, UMR Microbes, Evolution, Phylogeny and Infections (MEPHI), IHU Méditerranée Infection, Marseille, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

BAM Bacteriophage adherence to mucus

HERVs Human endogenous retroviruses

HMP Human microbiome project

HPV Human papillomavirus

IGLP Immunoglobulin-like proteins

LPS Lipopolysaccharide

NGS Next-generation sequencing

Glossary

Dysbiosis Alteration of the composition of the normal microbiome.

Human microbiome The collection of bacteria, archaea, micro-eukaryotes and viruses that colonize the human body.

Human virome The entire collection of viruses that inhabit our body. The virome is part of the human microbiome.

Next generation sequencing New technologies that allow for massively parallel sequencing of nucleic acids (also known as high-throughput sequencing).

Prophage A phage genome that is incorporated into the bacterial host chromosome.

Shotgun sequencing The sequencing of entire chromosomes and entire genomes based on the production of random fragments of DNA.

Viral metagenomics The analysis of the whole viral genomes in a particular sample.

Definition of the Human Virome

The human body shelters a complex, dynamic and abundant microbial community composed of bacteria, archaea, microeukaryotes (fungi and protozoa) and viruses that contribute to human health and disease. These microorganisms constitute what is called the microbiome and by extension, the viral component of this microbiome is called the virome. The human virome is basically made of all the viruses that are found at the surface or inside our body. It thus includes eukaryotic viruses infecting our cells or members of our eukaryotic microbiome, viruses of archaea and viruses of bacteria (also called bacteriophages or phages). Beside free virions, the human virome also includes all viruses and virus-derived elements that are either integrated within the host chromosome (human endogenous retroviruses, proviruses and prophages) or maintained extrachromosomally within the cells (episomes).

Eukaryotic viruses infecting our cells have evident role in human health and disease. Viral infections lead to a wide range of clinical manifestation, from asymptomatic to severe, or even fatal conditions. During lytic infection, the viral genome is actively replicated and the release of new progeny virus particles often leads to the lysis of the host cell. In case of resolutive infection, virus is cleared by host immunity whereas in case of persistent infection, virus continues to replicate (chronic infection) or is staying quiescent without replication (latent infection). Latent viruses have the ability to reactivate into a lytic form. Human papillomaviruses (HPV), polyomaviruses, adenoviruses or herpesviruses are persistent viruses with the capacity to promote cell transformation and induce cancer development. Viral integration may also have significant impact on host evolution if the virus infects the germ lines. Indeed, Human Endogenous Retroviruses (HERVs), which represent more than 8% of the human genome, are remnants of exogenous retroviruses that have been fixed in the human genome. They are transmitted through generations and ubiquitously transcribed in normal tissues. Some viral proteins of endogenous viruses have given important physiological functions to the host such as syncytins that are essential for placental morphogenesis in mammals.

Bacteriophages constitute the major fraction of the human virome. They are particularly abundant at body barrier sites such as the gastrointestinal tract or the oropharynx where the density of their bacterial host is high ([Table 1](#)).

Bacteriophages infect bacterial cells either via lytic (for virulent viruses) or lysogenic (for temperate viruses) cycles. Lytic cycle leads to host cell lysis and viral production through diversion of the cellular host machineries. In contrast, lysogenic cycle leads to the integration of viral genome into bacterial genome. The latent bacteriophage is then called a prophage and is vertically transmitted after cell division. Environmental stressors can induce lytic activation of integrated prophages. Viral infection thus

Table 1 Abundances of virus-like particles in different body systems. Abundances of free virions in different compartments of the human body have been measured experimentally by direct counts using epifluorescence microscopy (eFM) or flow cytometry (FC) after filtration and staining of the viral nucleic acids protected by the capsid or *in-silico* quantification using spiked exogenous virus as internal control (IC)

Body system	Sample	Number of virus-like particles	Pathology	Method
Gastrointestinal tract	Stools	10^8 – 10^{10} /g	–	eFM, IC
	Gut mucosa	10^9 /biopsy	–	eFM
Circulatory	Blood	10^5 /ml	–	eFM
Oropharynx	Saliva	10^8 /ml	–	eFM
	Dental plaque	10^7 /mg	–	eFM
	Maxillary sinus flush	10^9 /ml	Chronic rhinosinusitis	FC
Urinary tract	Urine	10^7 /ml	Controls and patients with urinary tract infection	eFM
Tegumentary	Skin	10^6 /cm ²	–	eFM

regulates bacterial community composition and structure as well as bacterial function by providing genes that may be involved in bacterial fitness, pathogenesis and adaptation to changing environments. Bacteriophages can also indirectly influence human health by maintaining homeostasis of the indigenous bacterial community through the lysis of bacterial pathogens or in contrast, by inducing dysbiosis through predation on beneficial bacterial strains.

Only scarce information is currently available for viruses that infect other members of the human microbiome such as protozoans, fungi or archaea in the body compartments of asymptomatic individuals. These correspond to a largely unexplored area of the human virome and are therefore not included in this article.

Technological Development and Limits for the Description of the Human Virome

Advances in virology have often been driven by technological developments (Fig. 1) and nearly 40 years after the first description of viruses, it became possible to visualize their structure with electron microscopy. Cell culture for growing viruses was another technological breakthrough as it allowed sufficient viral material to be recovered for the development of biochemical virology in the 60's and subsequent molecular biology in the 80's.

In the last 10 years, the study of the human virome has been largely facilitated by the development of next-generation sequencing (NGS) technologies even though it still lags far behind those addressing the bacterial component of the human microbiome, as done within the Human Microbiome Project (HMP) (see “Relevant Websites section”). This is partly due to the fact that viruses, in contrast to bacteria, lack universal phylogenetic markers such as the 16S rRNA gene. Current protocols for generating a viral metagenome then rely on the high-throughput shotgun sequencing of viral nucleic acids extracted after enrichment/purification of circulating viral particles by using filtration and (ultra)centrifugation steps and further digestion of the non-encapsidated contaminant nucleic acids (e.g., human, bacterial) with nucleases. Using such protocols, latent viruses (prophages and proviruses) that stably reside within cells (in integrated or episomal forms) are largely missed and their diversity/functionality is probably underestimated. Other poorly described members of the human virome in healthy individuals are large dsDNA viruses whose recovery is prevented by filtration and RNA viruses, because most viral metagenomic studies conducted on RNA viruses were targeting diseased individuals. Thus, many eukaryotic viruses (both DNA and RNA viruses), including potential pathogens, have yet to be discovered.

Finally, there is still a high proportion of the reads obtained after sequencing that do not have homologs in public databases. This represents a major obstacle for accurate description of the human virome based on taxonomic composition. As more viruses are being isolated and functionally characterized using more traditional methods (cell culture, molecular biology, Fig. 1) and their genomes are added to the databases, a more precise and complete overview of the human virome will emerge.

Composition and Diversity of the Human Virome by Body Compartments

So far, most metagenomic works have focused on the analysis of the composition and the diversity of viruses at individual body sites such as the gastrointestinal, respiratory and genito-urinary tracts, the oral cavity, skin and blood. The distribution of the main viral families detected in these major human body sites is presented in Fig. 2.

As done for microbial communities, further initiatives analyzing the differences in viral communities between body sites (body barriers or systemic compartments), geographic sites, disease states or age groups should be largely developed in the future. Results from such studies will certainly deepen our knowledge of the factors that influence the composition, diversity and dynamics of the human virome in healthy individuals.

Gastrointestinal Tract Virome

Gut microbiome is so far the most studied microbial community in healthy human, due to its known or suspected roles in a broad range of homeostatic physiological functions and inflammatory bowel diseases, and its large biomass which makes possible reliable

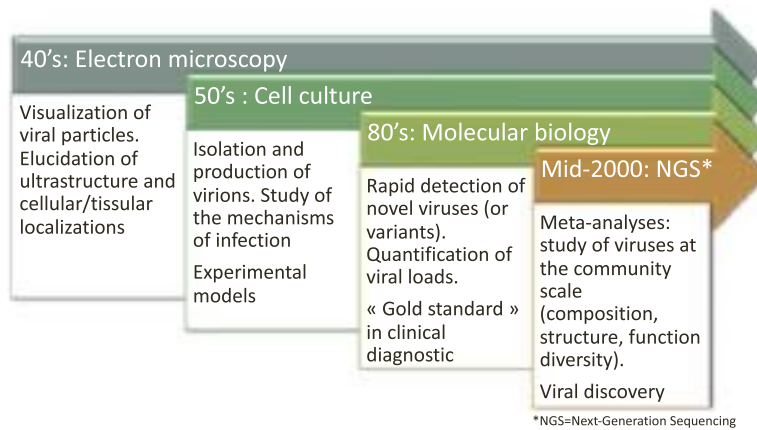


Fig. 1 Timeline of the main technological advances in virology.

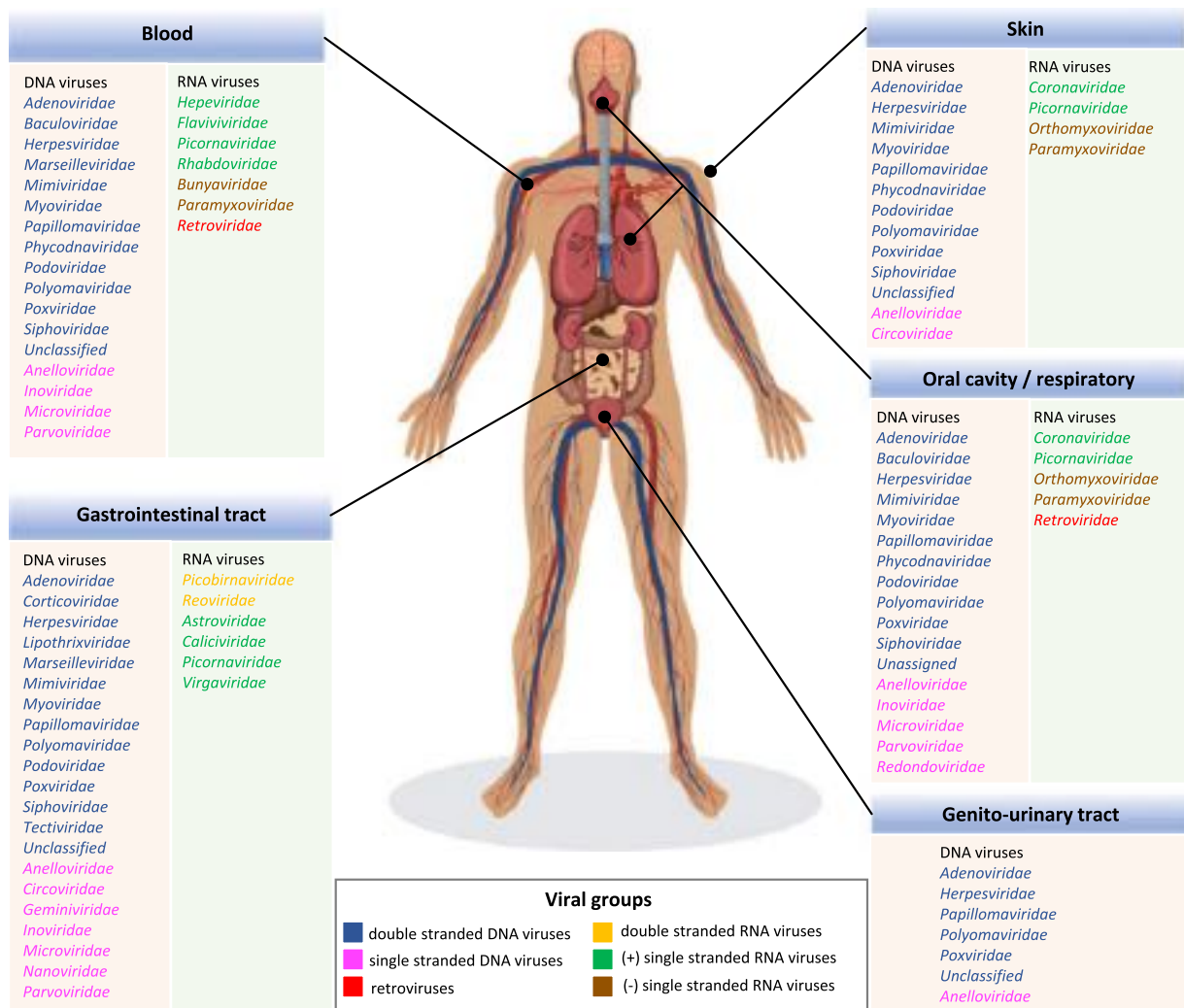


Fig. 2 Distribution of the main viral families found in major human body sites. Each viral group is represented with a unique color. Modified from Popgeorgiev, N., Temmam, S., Raoult, D., Desnues, C., 2013. Describing the silent human virome with an emphasis on giant viruses. Intervirology 56, 395–412, with permission.

sampling. The gastrointestinal tract is a highly heterogeneous environment harboring different microbial and viral communities both across its length and between the mucosa and lumen. Most analyses of gut viromes used samples from feces, due to obvious convenience of sampling. However, the beginning of the gastrointestinal tract in humans and pigs display a much higher and more stable proportion of bacteriophages. At the mucosal epithelium, phages accumulate at a concentration 10 times higher than bacteria. Viruses are highly abundant in the gut, with a large predominance of bacteriophages compared to eukaryotic viruses. Phage communities are primarily composed of tailed bacterial viruses (order Caudovirales), of the *Myoviridae*, *Podoviridae*, *Siphoviridae* families, and members of the *Microviridae* family. Other bacteriophage families, such as *Corticoviridae*, *Inoviridae* and *Tectiviridae*, have been identified with much lower abundances, as well as rare archaeal viruses (mainly from *Lipothrixviridae*). But nowadays, the majority of the gut virome (up to 85%) is still composed by unclassified or unknown viruses. Nevertheless, recent studies demonstrate that such difficulties could be overcome through bioinformatics efforts, as confirmed by the recent discovery within this unknown fraction of the novel bacteriophage, named α Assphage. Since then, data mining showed that α Ass-like phages are very abundant and highly prevalent in human populations worldwide especially those living in industrial areas. Successful isolation of a α Assphage was achieved in 2018 on *Bacteroides intestinalis*. Although this α Assphage isolate is a virulent podophage, it does not hamper proliferation and stably replicates along with its bacterial host for many generations. Going further, recent investigations demonstrate that α Assphage is a long-term human partner that has co-evolved with the human lineage.

Phage colonization of the gut occurs very early after (and possibly before) the birth, dynamically with their bacterial hosts/preys, coming from the diet and the environment, with primo-colonizers principally composed by *Caudovirales*. A recent study, using high depth sequencing technology, described high richness and diversity of first viral populations, coupled with high taxonomic instability, certainly due to the low colonization of the corresponding microbial hosts. During the first months of life, this diversity dramatically decreases, and the virome evolves toward a composition characterized by a higher relative abundance of *Microviridae* around 24 months of life/age. In healthy adult, gut virome is stable for long periods, even though phage communities are very different between individuals. Despite this individual specificity, a tiny core of bacteriophage core “species” seems to be widely distributed and conserved within the human population. Environmental exposure, diets, habits, and even genetics are suspected to play a role in shaping these communities, but the mechanisms and processes contributing to interpersonal phage dynamics are still unclear.

In contrast to bacteriophages, the eukaryotic virome is acquired after birth through environmental exposures and, as a consequence, eukaryotic virus diversity increases over time. Between 0 and 2 years, the most prevalent RNA viruses are enterovirus, parechovirus, tobamovirus and sapovirus and anelloviruses for DNA viruses. Anelloviruses are rarely detected earlier than 3 months of age and their abundance increase until peaking at 6–12 months and then decrease at 18 months of age. Others eukaryotic DNA viral families that have been less frequently detected by metagenomic studies are *Geminiviridae*, *Herpesviridae*, *Nanoviridae*, *Papillomaviridae*, *Poxviridae*, *Parvoviridae*, *Adenoviridae*, *Polyomaviridae*, *Circoviridae* and families of large dsDNA viruses such as *Marseilleviridae* and *Mimiviridae*. Eukaryotic RNA viruses in the gut of healthy adults have been thus far largely ignored by metagenomic sequencing analyses but this community seems to be dominated by plant- and insect-viruses, likely originating from the food.

Skin Virome

The human skin represents one of the first line of defense against external pathogens. It is composed of diverse microenvironments with distinct physicochemical parameters (temperature, moisture, pH, *etc.*) that strongly influence and shape microbial composition and function. Healthy skin usually shows low viral biomass and, depending on the study, either phages or eukaryotic DNA viruses dominate the viral community. Overall, the skin virome is highly variable across anatomical location and individuals with a relative stability of the phage community across time whereas eukaryotic DNA viruses are the most transient members of the skin community. Most phage sequences detected from the skin virome cannot be classified, even at the family level. Among those that can be taxonomically identified, most belong to the *Myoviridae* and *Siphoviridae* families. A core phageome of *Propionibacterium*, *Staphylococcus* and *Streptococcus* phages infecting the main bacterial genera present on human skin, is often detected between and within individuals. Their detection may be linked to the lysogenic nature of the infection as the majority of the skin bacteriophages were predicted to be temperate. Dominant families of eukaryotic DNA viruses were *Papillomaviridae*, *Polyomaviridae* and *Circoviridae*. Other DNA virus families that can be detected are *Poxviridae*, *Phycodnaviridae*, *Mimiviridae*, *Herpesviridae*, *Adenoviridae* and *Anelloviridae*. In few individuals and at particular body sites, a bloom of potential human pathogens can be observed in absence of clinical lesions. In addition, an expansion of the eukaryotic skin virome has been observed in immunodeficient patients compared to controls. This suggests, to some extent, an immune control of the skin colonization and infection. Finally, only scarce studies have described RNA viruses from the skin of healthy individuals so this group remains largely uncharacterized. In a recent study, RNA viromes were generated but most of the sequences mapped to DNA viruses of the *Papillomaviridae* and *Polyomaviridae* families, suggesting active replication state of these viruses. Only few other sequences were detected and these were related to *Coronaviridae*, *Picornaviridae*, *Orthomyxoviridae* and *Paramyxoviridae*.

Oral Cavity and Respiratory Tract Viromes

Similarly to the skin, the oral cavity is a highly heterogeneous environment displaying aerobic and anaerobic environmental niches made of soft (palatine, tongue) and hard (teeth) tissues, mucosa and biofilms that are all cleaned by saliva. The virome from the

oral cavity is dominated by phages of the *Siphoviridae*, *Myoviridae* and *Podoviridae* families and the composition of the oral virome is highly stable over time and personalized among individuals and, to some extent, gender specific due to host factors, such as hormones. Oral viromes were more similar between people having the same diet and between people sharing the same household or family. The core oral virome consisting of phages and prophages infecting *Streptococcus* species and human herpesvirus 7 has been identified, at least in the Western society. Compared to phages, the proportion of eukaryotic viruses in the oral cavity virome of healthy individuals is far less abundant. Apart from herpesviruses, sequence homologs of papillomaviruses, anelloviruses, poxviruses, phycodnaviruses, mimiviruses and few RNA viruses have been occasionally detected and their abundance might increase in particular stressful situations such as antibiotic intake or periodontal disease.

The virome from the respiratory tract has been mainly studied in the context of naso-pharyngeal and pulmonary pathologies as sampling procedures might be invasive particularly those targeting the lower airway. It thus mostly focused on the detection of potentially pathogenic eukaryotic viruses whose presence and abundance were highly variable among studies mostly due to the season of sampling, the viruses that were circulating in those particular geographical regions and the sample type. Very recently, novel circular Rep-encoding single-stranded DNA (CRESS) viruses were selectively identified in the human oro-respiratory tract and proposed as a new *Redondoviridae* family. Although the redondovirus prevalence among healthy and diseased subjects was similar, viral burden was observed in case of periodontal diseases.

Blood Virome

Early metagenomic studies have demonstrated that the human blood may represent a substantial reservoir of novel viruses that may represent a risk for the safety of blood products. Recently, a large screening of virus sequences in shotgun sequencing data from human blood cells from 8240 individuals identified the presence of 94 different viruses including 19 human viruses in 42 % of the participants. High prevalence of human herpesviruses and anelloviruses was detected and differences in viral prevalence depended on age, ancestry and sex with a greater prevalence of viruses in men than in women. This observation may be related to gender susceptibility to infection or to different innate and adaptive immune responses. Anelloviruses are ssDNA viruses belonging to the *Anelloviridae* family and include Torque Teno Viruses (TTVs), TTV-like viruses, SEN viruses, TTV midi viruses and TTV-like mini viruses. Anelloviruses are commonly detected in peripheral blood samples and present a high level of genomic heterogeneity. They are transmitted by transfusion and their abundance could serve as a biomarker of immune competence and as predictor of adverse outcomes after organ transplantation. Their role in human physiology and disease is still unclear but they are often considered as non-pathogenic commensals. Other eukaryotic DNA viruses are episodically detected in the blood of healthy humans and these belong to the *Papillomaviridae*, *Polyomaviridae*, *Parvoviridae*, *Adenoviridae*, *Poxviridae*, *Mimiviridae*, *Marseilleviridae* and *Phycodnaviridae* families, and Mollivirus genus.

Among eukaryotic RNA viruses, human pegivirus (HPgV) of the *Flaviviridae* family, formerly known as GB virus C (GBV-C) is highly prevalent among human populations and can be transmitted through parenteral, sexual, or mother-to-infant routes. Co-infection of HPgV with immunodeficiency virus (HIV) delays progression to AIDS as well as it improves survival to Ebola infection possibly by preventing aberrant immune activation. Other RNA viruses detected in the blood of asymptomatic individuals belong to the *Picornaviridae* and *Rhabdoviridae* families.

Phage sequences belonging to the *Myoviridae*, *Podoviridae*, *Siphoviridae*, *Microviridae*, and *Inoviridae* families are also frequently detected in blood samples of healthy individuals and higher sequence abundances have been detected in case of pathology. Although phage sequences in the human circulation may derive from contaminations, phage translocation and phage transcytosis events support the establishment of the intrabody phageome. How these phages might have reached the circulation and their role in the human body is discussed in the next article.

Genital and Urinary Tract Viromes

Viruses from the genito-urinary environment are only poorly described. The vaginal eukaryotic DNA virome of pregnant and non-pregnant women is dominated by viruses from the *Papillomaviridae* family, with half of them being only found in this body compartment whereas the others are shared with the skin, gastrointestinal tract and oral cavity. Other viral families such as *Polyomaviridae*, *Herpesviridae*, *Poxviridae*, *Adenoviridae* and *Anelloviridae* are also occasionally detected. A higher vaginal viral diversity is associated with pre-birth term in pregnant women. Human papillomaviruses, herpesviruses and polyomaviruses are also found in urine samples without significant differences in prevalence between men and women. The urine virome also contains a large proportion of bacteriophage sequences whose majority are probably lysogenic based on the high proportion of integrases identified in their sequences. According to a recent study on women with or without lower urinary tract infection symptoms, 86% of urinary bacterial genome contained at least one phage sequence most of them being novel. The detection of similar phage sequences among different women suggests the presence of a core bladder phageome whose role in health and disease remains to be elucidated.

Implication of the Virome in Human Health and Immunity

Numerous reports have highlighted the potential role of the virome in resilience and human health although the molecular and physiological mechanisms underlying these interactions remain largely unknown. Besides the evident role

of bacteriophages in regulating the bacterial microbiome, there is increasing evidence that this fraction of the virome may also interact with the human immune system by inducing a humoral response and triggering innate and adaptive immune responses. Circulating bacteriophages can interact directly with human antibodies, be engulfed by dendritic cells *in vitro*, and inhibit activation and proliferation of human T cells *in vitro*. Prophages may also change bacterial recognition by the immune system by modifying the outer membrane lipopolysaccharide (LPS). Few years ago, the bacteriophage adherence to mucus (BAM) has been proposed as a model of non-host derived innate immunity mediated by bacteriophages. This model relies on the observation that bacteriophages accumulate to high densities at the mucosal surfaces. According to this model, lytic phages able to transiently bind the mucin-glycoproteins composing the mucus with immunoglobulin-like proteins (IGLP) harbored on their capsids have been shown to drastically reduce bacterial colonization of the epithelium. These transient interactions are not directly responsible for the phage accumulation but rather result in a subdiffusive motion of bacteriophages that increases the frequency of phage–host encounters when bacterial densities are low. BAM thus promotes the replication of lytic phages at mucosal surfaces and allows preventing entry of pathogens into the organism by killing bacteria approaching the epithelial cells by viral lysis. BAM may also promote lysogenic conversion of bacteria by temperate phages at these sites. Temperate phages carry functional genes whose expression increases the fitness of infected bacterial cells. Among known adaptive functions found in healthy human virome, one could cite phage superinfection protection, multi-stress resistance and polysaccharide and carbohydrate metabolism genes. By promoting the adaptation and the spread of infected hosts, lysogenic conversion provides to the virus widespread multiplication when the conditions change, leading to the shift of viral replication strategy from lysogeny to lysis. The impact of lysogeny on the bacterial population regulation is thought to be fundamental, and the fact that temperate phages dominate human viromes and are stable over time within individuals suggest that lysogeny plays a major role in the stability and maintenance of the microbiome. Last, phage sequences have also been found in other body compartments that have long been considered as sterile, such as organs and fluids. Several routes for phages entry into the human body have been proposed such as a passage in favor of permeability at cell junctions in inflammatory gut epithelium. But recently, epithelial cells have been shown to be able to translocate phages from their apical to their basal pole, through a way involving the Golgi apparatus, in the absence of bacterial endotoxins that could alter epithelium permeability. The frequency of such transcytosis event is low ($\sim 0.1\%$) but would be sufficient to transport around 31 billion phages per day across the brain-blood barrier, allowing their continuous dissemination all over the human body. Thus, lytic phages would play a systemic role in immunity, protecting all the organs against pathogenic or opportunistic bacteria that could cross the mucosal barrier.

Beside this large majority of phages, each healthy individual harbors, in addition to integrated chromosomal viruses, at least ten permanent chronic systemic viral infections that drive continuous activation of the immune system. These resident viruses include members of the *Herpesviridae*, *Polyomaviridae*, *Anelloviridae*, *Adenoviridae*, *Papillomaviridae*, *Genomoviridae* and *Circoviridae* families (Fig. 2). While some are recognized human pathogens, many others have no known association with human pathology and could even provide benefit to the human host either by viral interference or cross-immunity. For example, it has been hypothesized that ubiquitous human β -papillomavirus may persist on human skin due to its potential to promote skin healing whereas murine norovirus compensates for the absence of intestinal bacteria in germ-free mice and protects against intestinal injury and inflammatory lesions during superinfection with bacterial pathogens. Other murine models have shown that herpesvirus infections may improve the outcome of secondary infection by priming the immune system for protection against bacteria and other viruses. The large prevalence and particularly high genetic diversity observed among members of the *Anelloviridae*, *Papillomaviridae*, *Genomoviridae* or *Circoviridae* families suggest that they may act as human health promoters, by priming the immune system and providing cross-protection against pathogenic strains as established for commensal bacteria. In contrast, viral infection can promote immunosuppression that facilitates secondary infection as it is the case for the depletion of CD4⁺ T cells by HIV whereas the chronic inflammation state maintained by viral infection may increase the risk of inflammatory response due to allergens, of secondary infection or carcinogenesis, as shown for certain members of *Papillomaviridae*, *Herpesviridae* and *Polyomaviridae* families.

Viral interactions with humans are complex and several eukaryotic viruses have established persistent interactions with their host that go beyond simple parasitism. The physiological role of most of these viruses in humans has yet to be defined and the fine balance between adverse or beneficial effects of these interactions probably depends on the virus itself (*e.g.*, viral genotype, replication mode, host range, abundance), the host (*e.g.*, genetic background, age, immune system) and its environment (*e.g.*, geographic location, demographic distribution, animal proximity).

Conclusion

The recent advances in next-generation sequencing over the past decade, have led to the discovery that viruses are an important component of the microbial communities that colonize every ecological niche of the human body. The study of the human virome is an emerging research field and our understanding of how these viruses interact with each other, with the other microbes and their host genetics and immune systems is still fragmented. The next ten years will certainly pave the way for the development of new techniques and integrative approaches that will allow drawing a more precise picture of the human virome and deciphering its role in shaping health and disease.

Further Reading

- Cadwell, K., 2015. The virome in host health and disease. *Immunity* 42, 805–813.
- De Paepe, M., Leclerc, M., Tinsley, C.R., Petit, M.A., 2014. Bacteriophages: An underestimated role in human and animal health? *Frontiers in Cellular and Infection Microbiology* 4, 39.
- Duerkop, B.A., Hooper, L.V., 2013. Resident viruses and their interactions with the immune system. *Nature Immunology* 14 (7), 654–659.
- Haynes, M., Rohwer, F., 2011. The human virome. In: Nelson, K.E., Peterson, J.L., Garges, S. (Eds.), *Metagenomics of the Human Body*. New York, NY: Springer-Verlag, pp. 63–77.
- Manrique, P., Bolduc, B., Walk, S.T., *et al.*, 2016. Healthy human gut phageome. *Proceedings of the National Academy of Sciences of the United States of America* 113, 10400–10405.
- Moya, A., Pérez Brocal, V. (Eds.), 2018. *The Human Virome*. *Methods in Molecular Biology*, vol. 1838. New York, NY: Humana Press.
- Parker, M.T., 2016. An ecological framework of the human virome provides classification of current knowledge and identifies areas of forthcoming discovery. *Yale Journal of Biology and Medicine* 89 (3), 339–351.
- Popgeorgiev, N., Temmam, S., Raoult, D., Desnues, C., 2013. Describing the silent human virome with an emphasis on giant viruses. *Intervirology* 56, 395–412.
- Rascovan, N., Duraisamy, R., Desnues, C., 2016. Metagenomics and the human virome in asymptomatic individuals. *Annual Review of Microbiology* 70, 125–141.
- Van Bellegghem, J.D., Dąbrowska, K., Vaneechoutte, M., Barr, J.J., Bollyky, P.L., 2019. Interactions between bacteriophage, bacteria, and the mammalian immune system. *Viruses* 11 (1), 10.
- Virgin, H.W., Wherry, E.J., Ahmed, R., 2009. Redefining chronic viral infection. *Cell* 138, 30–50.
- Virgin, H.W., 2014. The virome in mammalian physiology and disease. *Cell* 157, 142–150.
- Wylie, K.M., Weinstock, G.M., Storch, G.A., 2012. Emerging view of the human virome. *Translational Research* 160, 283–290.
- Wylie, K.M., 2017. The virome of the human respiratory tract. *Clinics in Chest Medicine* 38, 11–19.
- Zarate, S., Taboada, B., Yocupicio-Monroy, M., Arias, C.F., 2017. Human virome. *Archives of Medical Research* 48, 701–716.
- Zou, S., Caler, L., Colombini-Hatch, S., Glynn, S., Srinivas, P., 2016. Research on the human virome: Where are we and what is next. *Microbiome* 4, 32.

Relevant Websites

<https://hmpdacc.org>
NIH Human Microbiome Project.

Epidemiology of Human and Animal Viral Diseases

Michael Edelstein, Faculty of Medicine, Bar Ilan University, Ramat Gan, Israel

© 2021 Published by Elsevier Ltd.

This is an update of F.A. Murphy, Epidemiology of Human and Animal Viral Diseases, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00390-3.

Glossary

Airborne transmission Transmission via small aerosolized droplets emitted from infected persons from environmental sources.

Case-control study An epidemiological study in which the risk factors of humans or animals with a disease are compared with those without the disease.

Cohort study Attempt to identify the cause of a disease by comparing exposed and nonexposed (control) populations in terms of outcome.

Common source transmission Pertains to contamination of food and water supplies.

Digital epidemiology the use of online data sources to study the determinants, dynamics, and distribution of diseases in populations.

Direct contact transmission Involves actual physical contact between an infected subject and a susceptible subject (e.g. touching, kissing, biting, sexual activity).

Droplet transmission Transmission via large droplets typically traveling a few meters, emitted from infected persons during coughing or sneezing or from environmental sources.

Epidemic (Outbreak) Increase in disease incidence above what is expected in a particular place during a particular time period.

Epidemiology The study of the determinants, dynamics, and distribution of diseases in populations.

Fomite An inanimate object that may be contaminated with virus and become the vehicle for transmission.

Herd immunity The immune status of a population that affects viral transmission rates. Often used in describing the elimination of a virus from a population when there are too few susceptible hosts remaining to sustain a transmission chain.

Horizontal transmission The transfer of infectious virus from one human or animal to another by any means other than vertical transmission.

Iatrogenic transmission Transmission via health care procedures, materials, and workers (e.g., physicians, nurses, dentists, veterinarians).

Incidence risk (or attack rate) A measure of the occurrence of infection or disease in a population over time – it refers to the proportion of a population contracting a particular disease during a specified period.

Mathematical model (epidemiological) A means to convey quantitative information about a host-virus interaction, such as an epidemic or an emerging disease episode, by the construction of a set of predictive mathematical algorithms.

Nosocomial transmission Pertains to infections acquired while a patient, human or animal, is in hospital.

Prevalence The ratio, at a particular point in time, of the number of cases currently present in the population divided by the number of subjects in the population at risk; it is a snapshot of the occurrence of infection or disease at a given time.

Species jumping (or host range extension) Referring to a virus that derives from an ancient reservoir life cycle in animals, but has subsequently established a new life cycle in humans or a different animal species and no longer uses, the original animal reservoir.

Transmission The process by which a pathogen is shed from one host and infects the next.

Vector-borne transmission Involves the bites of arthropod vectors (e.g., mosquitoes, ticks, sandflies).

Vertical or transplacental transmission Transmission of virus from mother to fetus prior to or during parturition, either across the placenta, when the fetus passes through the birth canal, or via colostrum and milk.

Zoonosis Disease which is naturally transmitted to humans from an ongoing reservoir life cycle in animals or arthropods, without the permanent establishment of a new life cycle in humans.

Introduction

Viral disease epidemiology is the study of the determinants, dynamics, and distribution of viral diseases in populations. The risk of infection or disease in a population is determined by characteristics of the pathogen, the host and the environment, as well as behavioral, societal, environmental, and ecological factors that affect virus transmission from one host to another. Epidemiology attempts to meld these factors into a unified whole. The depiction of the interaction of factors favoring the emergence of a viral disease (Fig. 1), called “the convergence model”, is taken from the US Institute of Medicine study, Microbial Threats to Health, Emergence, Detection and Response (National Academy Press, 2003). At the center is a box representing the convergence of factors leading to “the black box”, reflecting the reality that many unknown interactions are important virologically and epidemiologically. Complementing the convergence model, improvements in our understanding of host susceptibility and propensity to transmit viral disease has led to the concept of the “exposome”, the totality of human environmental exposures from conception

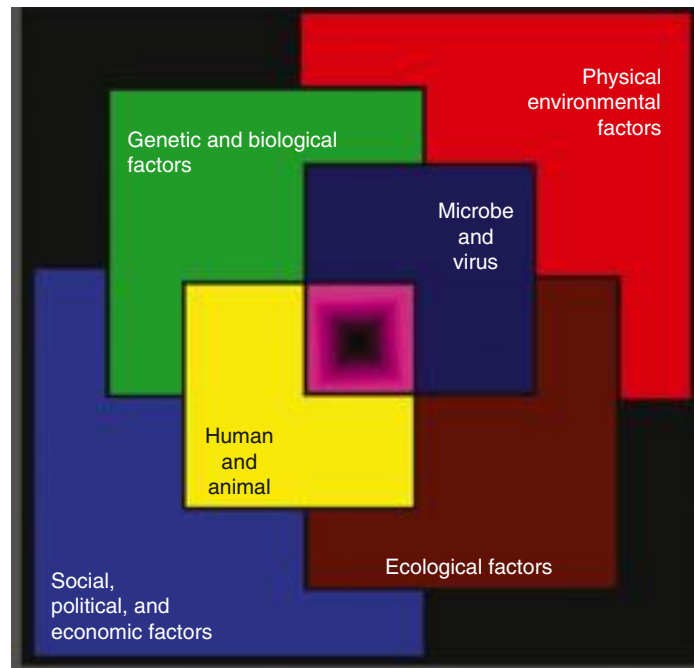


Fig. 1 The Convergence Model. Factors contributing to the emergence of viral diseases. At the center is a box representing the convergence of factors leading to the emergence of a viral disease. The black center represents unknown interactions, “the black box”. Reproduced from Smolinski, M.S., Hamburg, M.A., Lederburg, J., (Eds.), 2003. *Microbial Threats to Health, Emergence, Detection and Response*, Fig. 1 p. 3. Washington, DC: National Academies Press. Reprinted with permission from the National Academy Press, copyright (2003), National Academy of Sciences.

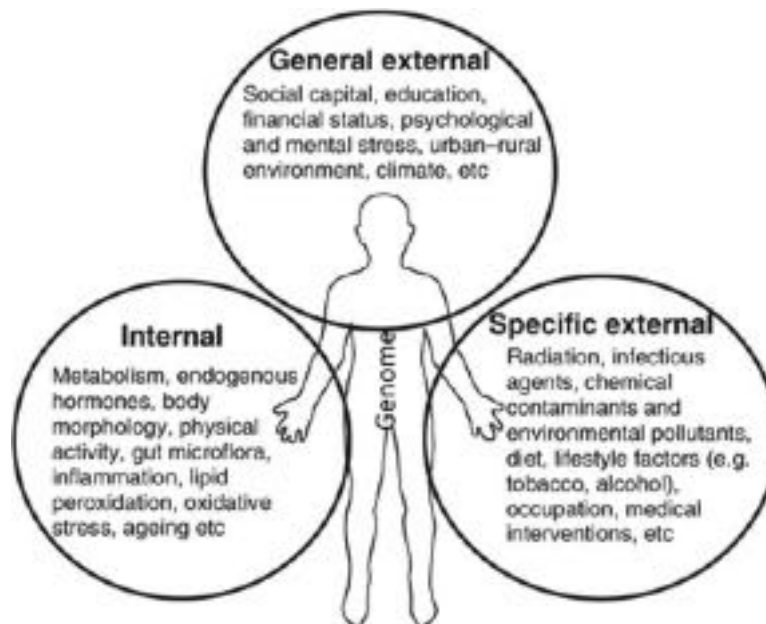


Fig. 2 Three different domains of the exposome are presented diagrammatically with non-exhaustive examples for each of these domains. Reproduced from Wild, C., 2012. *Int. J. Epidemiol.* 41, 24–32.

onwards, complementing the genome. The combination of internal and external factors influence the susceptibility of individuals to be infected and transmit viral disease, and in turn of viruses to spread in human populations (Fig. 2).

The foundations of epidemiology predate the microbiological and virological sciences, starting with Hippocrates, the Greek physician and father of medicine, who in the fourth century BC made important epidemiologic observations on infectious

diseases. John Snow is called the father of modern epidemiology because he developed excellent quantitative methods while studying the source of a cholera outbreak at the Broad Street pump in London in 1849. Snow was followed by William Farr, who in the 1870s advanced the use of vital statistics and clarified many of the principles of risk assessment and retrospective and prospective studies. Their vision is reflected in the fast-changing science of epidemiology which is now supported by increased availability and quantity of health and non health data ("big data"), advances in data processing in the form of Artificial intelligence, machine learning and natural language processing, sophisticated statistical methods, and very sensitive and specific diagnostic systems. Improvements in our understanding of the dynamics of disease transmission has placed epidemiology in the center of a complex multidisciplinary network that includes public health, biostatistics, data science, microbiology as well as sociology, anthropology and psychology. Epidemiology in the 21st century is therefore firmly collaborative and sits within a multidisciplinary ecosystem that includes medicine and biology as well as social and computer sciences.

Factors Influencing the Spread of Viral Diseases

Perpetuation of Viruses in Nature

Perpetuation of a virus in nature depends upon the maintenance of serial infections, that is, a chain of transmission; Viruses survive only if they are able to be transmitted from one host to another, whether of the same or another species. Transmission cycles require virus entry into the body, replication, and shedding with subsequent spread to one or several hosts. The potential for the spread of a virus in a population is measure by the basic reproductive number (R_0) defined as the average number of cases directly generated by one case in a population where all individuals are susceptible to infection. A virus will continue to spread in a population (at a speed determined by the R_0) until it runs out of susceptible individuals or control measures that stop or reduce transmission are implemented. The occurrence of clinical disease is neither required nor necessarily advantageous for viruses from a species survival perspective.

Virus Entry

Portals of virus entry into the body include the skin, respiratory tract, intestinal tract, oropharynx, urogenital tract, and conjunctiva. In some cases, viruses use a particular portal of entry because of particular environmental or host-behavior factors and in other cases because of specific viral ligands and host-cell receptors. In many cases, disruption of normal host-defense mechanisms leads to entry that might otherwise be thwarted; for example, papillomaviruses may enter the deep layers of the skin via abrasions, acid-labile coronaviruses may enter the intestine protected by the buffering capacity of milk, and influenza viruses may enter the lower respiratory tract because a drug has dampened ciliary action of the respiratory epithelium.

Virus Shedding

The exit of virus from an infected host is just as important as entry in maintaining its transmission cycle. All portals used by viruses to gain entry are used for exit. The important elements in virus shedding are virus yield (from the standpoint of the virus, the more shedding the better) and timeliness of yield (again, the earlier the shedding the better). Viruses that cause persistent infections often employ remarkable means to avoid host inflammatory and immune responses so as to continue shedding. For example, the epidemiologically important shedding of herpes simplex viruses 1 and 2 that perpetuates the viruses in populations requires recrudescence of persistent ganglionic infection, centrifugal viral genomic transit to peripheral nerve endings, and productive infection of mucosal epithelium, all in the face of established host immunity.

Modes of Virus Transmission

Virus transmission within a species may be "horizontal" or "vertical". The vast majority of transmission is horizontal, that is, between individuals within the population at risk. Modes of horizontal transmission of viruses can be through direct contact, fomites (inert objects that can carry viruses like surfaces, clothes, utensils etc.), airborne transmission, droplet transmission, vector-borne, iatrogenic, and nosocomial. In addition to transmission between individuals, outbreaks of viral disease may occur through a common contaminated source such as food or water. Vertical transmission occurs between the mother and her fetus or newborn. Some viruses are transmitted in nature via several modes, others exclusively via one mode (see [Table 1](#)).

'Direct contact transmission' involves actual physical contact between an infected subject and a susceptible subject (e.g., kissing for the Epstein-Barr virus, the cause of mononucleosis; biting for rabies; sexual activity (genito-genital, anogenital, oro-genital) for herpes).

Indirect contact transmission occurs via "fomites", such as shared eating utensils, improperly sterilized surgical equipment, or improperly sterilized non-disposable syringes and needles.

"Common source transmission" pertains to contamination of food and water supplies (e.g., norovirus diarrhea).

"Airborne transmission" typically results in respiratory infections (and less typically in intestinal infections), Airborne transmission occurs via very small droplet nuclei (aerosols) emitted from infected persons during coughing, sneezing, singing or talking loudly (e.g., influenza, COVID-19) or from environmental sources.

Table 1 Examples of human and animal virus transmission patterns

<i>Infectious agent/disease</i>	<i>Mode of transmission</i>	<i>Portal of entry</i>
Influenza virus/influenza	Contact/direct/indirect via droplets and droplet nuclei	Respiratory tract
Rhinoviruses/common cold	Contact/direct/indirect via droplets and droplet nuclei and fomites	Respiratory tract
Rubella virus/congenital rubella	Contact/direct/indirect via droplets and droplet nuclei	Respiratory tract
Rotaviruses/diarrhea	Vertical/congenital	Transplacental
Poliovirus/poliomyelitis	Contact/direct/indirect via fomites	Intestinal tract (oral)
Norovirus/diarrhea	Contact/direct/indirect via fomites	Intestinal tract (oral)
Hepatitis A virus/hepatitis	Common vehicle/fecal contamination of water	Intestinal tract (oral)
Variant Creutzfeldt–Jakob disease prion/prion disease (spongiform encephalopathy)	Common vehicle/fecal contamination of food/Contact/direct (sexual)	Intestinal tract (oral)
Herpes simplex virus/genital herpes	Common vehicle/bovine spongiform encephalopathy prion contamination of beef or beef products	Intestinal tract (oral)
Human immunodeficiency virus 1/acquired immunodeficiency syndrome (AIDS)	Contact/direct (sexual)	Genital tract
Rabies virus/rabies	Contact/direct (sexual), contact/direct (blood)	Genital tract, bloodstream, transplacental, at birth and via breast feeding
Russian spring summer encephalitis virus/encephalitis	Vertical/congenital	Skin (bite wound)
Dengue viruses/dengue	Zoonotic/contact/direct (saliva)	Skin (tick bite)
Sin Nombre and related viruses	Zoonotic/arthropod-borne	Skin (mosquito bite)
Lassa virus	Zoonotic/contact/direct (rodent urine, saliva and feces)	Respiratory tract
Ebola and Marburg viruses	Zoonotic/contact/direct (rodent urine, saliva and feces)	Respiratory tract and intestinal tract (oral)
Leukemia viruses/leukemias (proven only in animals)	Zoonotic/reservoir host unknown; secondary cases contact/direct/nosocomial and iatrogenic	Index cases unknown, likely respiratory tract or skin and mucous membranes; secondary cases, contact and iatrogenic (injection)
	Vertical/germ-line	Transmitted as genetic trait

Droplet transmission occurs via large droplets ($> 10 \mu\text{m}$ in diameter) that settle quickly. Droplet nuclei can evaporate forming dry particles ($< 5 \mu\text{m}$ in diameter) which remain suspended in the air for extended periods. Droplets may travel only short distances (a few meters) while droplet nuclei may travel over much longer distances.

“Vector-borne transmission” involves an arthropod vector (e.g., mosquitoes, ticks, and sandflies) acquiring the pathogen when biting an infected person and transmitting the pathogen when biting someone else.

“Iatrogenic transmission” involves health care procedures, materials, and workers (e.g., physicians, nurses, dentists, and veterinarians).

“Nosocomial transmission” pertains to infections acquired while a patient, human or animal, is in hospital.

“Vertical or transplacental transmission” occurs from mother to fetus prior to or during parturition. Certain retroviruses are vertically transmitted in animals via the integration of viral DNA directly into the DNA of the germline of the fertilized egg. Other viruses are transmitted to the fetus across the placenta; yet others are transmitted when the fetus passes through the birth canal. Another vertical transmission route is via colostrum and milk. Vertical transmission of a virus may or may not be associated with “congenital disease” (i.e., disease that is present at birth) which may be lethal (and the cause of abortion or stillbirth) or the cause of congenital abnormalities. The herpesviruses, especially cytomegaloviruses, and rubella virus cause important congenital diseases in humans, and pestiviruses, such as bovine viral diarrhea virus, in animals. Viruses acquired vertically can lead to chronic illness later in life, e.g., hepatitis B virus and HIV.

Influence of the Clinical Status of the Host

It is possible for an individual to be infected but present no symptoms (asymptomatic infection) or symptoms so mild that the individual does not identify as ill (subclinical infection). Infected individuals who are asymptomatic or subclinical may however still transmit the disease. This presents a challenge when trying to identify contacts of cases in order to break chains of

transmission, even with the help of laboratory diagnostics. Persistent infections, whether or not they are associated with episodes of clinical disease, also play an important role in the perpetuation of many viruses in nature. For example, prolonged virus shedding can reintroduce virus into a population of susceptibles all of which have been born since the last clinically apparent episode of infection. This is important in the survival of rubella virus in some isolated populations. Sometimes the persistence of infection, the production of disease, and the transmission of virus are dissociated; for example, togavirus and arenavirus infections may have little adverse effect on their reservoir hosts (arthropods, birds, and rodents), but transmission may be very efficient. On the other hand, the persistence of infection in the central nervous system, as with measles virus in subacute sclerosing panencephalitis (SSPE), is of no epidemiological significance, since no infectious virus is shed from this site.

Influence of Virulence of the Virus

The virulence of the infecting virus may directly affect the probability of its transmission, with both extremes providing a disadvantage from a survival perspective. The classic example of this is rabbit myxomatosis. In Australia, mosquito-borne transmission of myxoma virus was found to be most effective when infected rabbits maintained highly infectious skin lesions for several days before death. Highly virulent strains of the virus were found to kill rabbits so quickly that transmission did not occur, and naturally attenuated strains were found to produce minimal lesions that healed quickly and did not permit transmission. In human populations viruses with a very high case fatality rate and a high probability of transmission only when cases are very unwell (such as Ebola or Middle East Respiratory syndrome, MERS) have less potential for global spread than viruses that result in mild or asymptomatic cases and where transmission occurs where symptoms are mild or absent (such as sexually transmitted herpesvirus for example).

Influence of Host Population Immunity

With most viruses, endemic or epidemic transmission leads to increasing levels of immunity in the host population until population immunity is high enough to slow down and eventually interrupt further transmission- the proportion of the population that needs to be immune in order to stop viral transmission is called the herd immunity threshold. It is always below 100% and its exact value varies between viruses according to how transmissible they are (measured by the R_0). The “herd immunity” effect is countered in some cases by viral antigenic variation that causes viruses to evade immunity in previously infected individuals. For example, influenza viruses undergo genetic variations such that persons immune to previously circulating virus strains are susceptible to new strains. Assessing these genetic changes is an important objective of laboratory-based influenza surveillance programs, which in turn are the basis for decisions on the formulation of each year's influenza vaccine.

Influence of Population Size

Long-term survival of a virus requires that it be continuously transmitted from one host to another. In general, for rapidly and efficiently transmitted viruses such as many respiratory viruses, local survival of the virus requires that the susceptible host population be very large. A virus may disappear from a population because it exhausts its potential supply of susceptible hosts as they acquire immunity to reinfection with the same virus. Depending on duration of immunity and the pattern of virus shedding, the “critical population size” varies considerably with different viruses and with different host species. The most precise data on the importance of population size in acute nonpersistent infections come from studies of measles. Persistence of measles virus in a population depends upon a continuous supply of susceptible children. Analyses of the incidence of measles in large cities and in island communities have shown that a population of about half a million persons is needed to ensure a large enough annual input of new susceptible hosts, by birth or immigration, to maintain measles virus in the population. Because infection depends on respiratory transmission, the duration of epidemics of measles is correlated inversely with population density. If a population is dispersed over a large area, the rate of spread is reduced and the epidemic may last longer, so that the number of susceptible persons needed to maintain transmission chains is reduced. On the other hand, in such a situation a break in the transmission chain is much more likely. When a large proportion of the population is initially susceptible, the intensity of the epidemic builds up very quickly and attack rates are almost 100% (“virgin-soil epidemic”). On the other hand, when measles vaccination programs achieve over 95% coverage, the virus disappears almost completely from the population.

Influence of Zoonotic Transmission Cycles

Because most viruses are host-restricted, most viral infections are maintained in nature within populations of the same or related species. However, there are a number of viruses that may have multiple hosts and spread naturally between several different species of vertebrate host, for example, rabies and eastern equine encephalitis viruses. The term “zoonosis” is used to describe multiple-host infections that are transmissible from animals to man. Approximately 70% of emerging infections are zoonoses. Recent example of zoonotic infections that have led to outbreaks or pandemics include Ebola, MERS and COVID-19. Once viruses jump the species barrier, they may cause an outbreak and disappear, or establish themselves in human populations and become endemic.

Influence of Arthropod Transmission Cycles

Many viral zoonoses are caused by arboviruses. Arboviruses have two classes of hosts, vertebrate and invertebrate. Over 500 arboviruses are known, of which about 100 cause disease in humans and 40 in domestic animals; some of these are transmitted by ticks, some by mosquitoes, and yet others by phlebotomine flies (sandflies) or *Culicoides* spp. (midges). Arthropod transmission may be “mechanical”, where the arthropod acts as a “flying pin”, or more commonly, “biological”, involving replication of the virus in the arthropod vector. The arthropod vector acquires virus by feeding on the blood of a viremic person or animal. Replication of the ingested virus, initially in the arthropod's gut, and its spread to the salivary glands takes several days; the interval varies with different viruses and is influenced by ambient temperature. Virions in the salivary secretions of the vector are injected into human or animal hosts during subsequent blood meals. Most arboviruses have localized natural habitats in which specific receptive arthropod and vertebrate hosts are involved in the viral life cycle. Vertebrate reservoir hosts are usually wild mammals or birds; humans are rarely involved in primary transmission cycles, although the exceptions to this generalization are important (e.g., Venezuelan equine encephalitis, yellow fever, and dengue viruses). Humans are in most cases infected incidentally, for example, by the geographic extension of a reservoir vertebrate host and/or a vector arthropod. Ecological changes produced by human activities disturb natural arbovirus life cycles and have been incriminated in the geographic spread or increased prevalence of arbovirus diseases.

Assessment of Disease Occurrence and Outcome

By introducing quantitative measurements of disease trends, epidemiology has come to have a major role in improving our understanding of the overall nature of disease and in alerting and directing disease control activities. Epidemiology is also effective in (1) clarifying the role of particular viruses and viral variants as the cause of disease, (2) clarifying the interaction of viruses with environmental determinants of disease, (3) determining factors affecting host susceptibility, (4) unraveling modes of transmission, (5) evaluating the effectiveness of control measures and (6) field testing of vaccines and antiviral drugs.

At its core, epidemiology concerns itself with describing the occurrence of disease in populations in time (when disease occurs), place (where it occurs) and person (the characteristics of cases). With regards to viral disease, quantifying disease implies either calculating the occurrence of new cases in a population within a specified time period (interchangeably called Incidence, incidence risk, attack rate or cumulative incidence) or the proportion of the population infected at a point in time or over a time period (respectively point and period prevalence). Although of these measures are risks (probabilities), they are sometimes incorrectly referred to as rates. Incidence rate is a separate concept in epidemiology, mainly used in cohort studies (see below) where the denominator is no longer the population but the person time at risk, i.e., the sum of the time-at-risk of infection that each individual in the population contributes. The term “case-fatality rate” (technically a risk or probability) is used to indicate the percentage of subjects with a particular disease that die from the disease. In order to improve understanding of the burden of disease, the measurements (incidence, prevalence, case-fatality) can be calculated and presented stratified by attributes of interest such as age, sex, genetic constitution, immune status, pregnancy, nutritional status, and various behavioral and medical care and patient management parameters. Understanding disparities in health outcomes using measures of disease occurrence stratified by attributes of interest is becoming increasingly important as reducing inequalities in health becomes an explicit objective in global health policy related to infectious diseases.

A viral disease is characterized as “endemic” when there are multiple or continuous chains of transmission resulting in continuous occurrence of disease in a population over a period of time. “Epidemics” or “Outbreaks” are peaks in disease incidence that exceed the endemic baseline or expected rate of disease. A “pandemic” is an epidemic that spreads to more than one region of the globe.

Epidemiological Studies

The description of an epidemiological situation should include the dimensions of time, place and person. A proper description of an outbreak of disease or an epidemic must include the parameters of “person (or subjects in the case of animals), place, and time”. Such descriptive information is a necessary first step in describing the occurrence, distribution, and course of disease, enabling to assess the level of threat, and anticipated action response to the initial recognition of a cluster of cases of disease. In many instances, robust descriptive epidemiology is sufficient to provide the data and evidence required to guide the response to a viral threat. When it is not, a range of analytical approaches can be used to better understand the source of an outbreak and inform the implementation of a control strategy.

Case-control studies and cohort studies

These observational study designs are often used in infectious disease epidemiology to identify associations between an exposure of interest and an outcome. Case-control studies are always retrospective and compare two groups of individuals in terms of exposure: those who have developed the outcome of interest (cases) and those who have not (controls). Case control studies are relatively cheap and quick to run and are particularly well suited to investigate rare outcomes, or when the population of interest is not easily identifiable. One of the key challenges in case control studies is the recruitment of controls that are representative of the

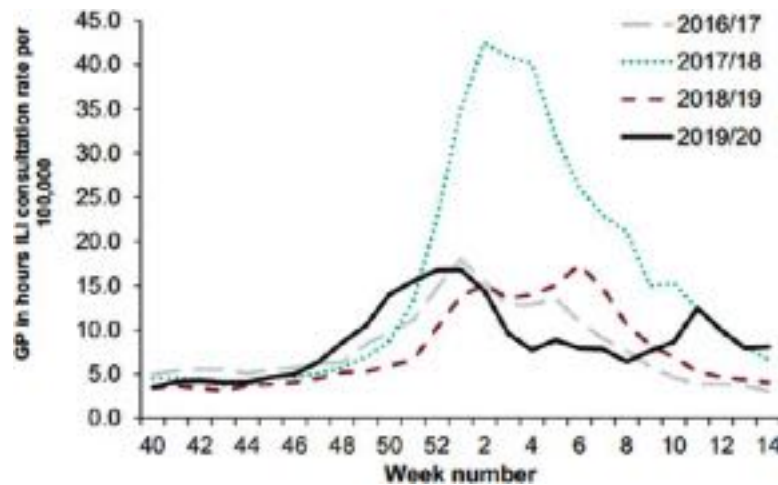


Fig. 3 An example of the use of time series in epidemiology: Weekly all age GP consultations for Influenza like illness in England, 2016–2020 (Source: Surveillance of influenza and other respiratory viruses in the UK Winter 2019–2020 Public Health England).

population of interest in order to avoid bias. Incidence cannot be calculated in case-control studies, which uses odds instead. Odds are the ratio of those developing an outcome (or having an exposure) to those who don't. Typically a case control study measure the ratio of the odds of exposure among cases and controls. Cohort studies compare individuals exposed and not exposed to an exposure of interest in terms of outcome of interest. Cohort studies can be either retrospective (when both exposure and outcome have already occurred) or prospective (when the outcome of interest has not yet occurred). Retrospective cohort studies are commonly used to investigate foodborne outbreak for example where the population is closed and well defined, such as during a wedding or on a cruise ship. Prospective cohort studies are long and expensive to run but provide powerful evidence of association between exposure and outcome. A classic example is the British Doctors study, which followed over 34,000 doctors in total between 1951 and 2001 and established an association between smoking (the exposure) and lung cancer (the outcome).

Time series analysis

These are observational studies widely used in the epidemiology of viral diseases, and in particular in disease surveillance. The two key roles of such an approach are (1) to follow and (2) describe trends of disease overtime establishing seasonality or cyclical patterns spreading over several years (Fig. 3). Examples include establishing and monitoring the seasonality of influenza activity, and characterizing the 2–3 yearly peaks in measles incidence in the absence of vaccine and the lengthening of the cycle as vaccine coverage increases. The second common use of time series is the detection of outbreaks or the beginning of seasonal activity of a particular viral disease through the use of exceedance thresholds. Retrospective time series data provides a baseline, and complex statistical algorithms can detect departure from this baseline, indicating the emergence of an outbreak or the beginning of the season of a viral disease. This approach is commonly used to detect the onset of the influenza or norovirus for example, triggering public health actions such as vaccination and infection control measures in healthcare settings and the community.

Qualitative studies

In recent years the role of quantitative epidemiology to understand the spread and determinants of viral disease has been gradually complemented by qualitative approaches to understanding the epidemiology of viral diseases. While quantitative studies focus on describing the extent of the spread disease and associated risk factors in populations and their subgroups, qualitative approaches help understanding why specific populations are affected by the exposure, the outcome or both. An optimal understanding of the epidemiological of viral disease and the generation of evidence to inform public health action therefore requires a combination of both quantitative and qualitative approaches to understand the spread of viral diseases and its barriers and enablers.

Molecular Epidemiology

The term “molecular epidemiology” is used to denote the use of any of a large number of molecular biological methods in support of epidemiological investigations. The combination of molecular techniques with epidemiology now routinely provides previously unavailable information relevant to the control of infectious diseases such as the ability to ascertain the geographical origin of a viral strain or the ability to link cases in an outbreak investigation with homologous viral sequences even in the absence of an epidemiological link. For example, with herpesviruses, restriction endonuclease mapping has provided a means of identification of unique viral genotypes – in an epidemiologic study recognized as the first based upon viral molecular characterization, the source of herpes simplex virus 1 causing disease in a hospital newborn nursery was traced to one persistently infected nurse rather than any of several other possible shedders. With rotaviruses and bluetongue viruses, polyacrylamide gel electrophoresis of the

segmented viral RNA has been used epidemiologically, for example, to unravel outbreaks involving multiple viral variants. Panels of monoclonal antibodies have been used to distinguish virus variants for epidemiologic purposes; they have been particularly useful in elucidating host-range and geographic variants of rabies virus. Today, partial sequencing has become the most commonly used molecular epidemiologic methodology; partial sequencing of poliovirus isolates recovered from patients indicates whether they are wild type (even local or introduced wild type), attenuated vaccine type, or a vaccine type that has reacquired neurovirulence during human passage. Partial sequencing of foot-and-mouth disease viruses can offer the same kind of geographic information of virus movement as has proved so useful in polio control and eradication programs, but because of political sensitivities in some countries a robust international reference laboratory system has not been established that could provide the same kind of practical disease control information as has been the case with polio. Whole genome sequencing (WGS) is becoming more commonplace and decreases in cost and increases in speed means that WGS is now used as part of routine surveillance activities for many viruses in many countries, enabling unprecedented abilities to map and track the spread of viruses in human and animal populations. Thus, with many human and animal viruses the use of molecular epidemiological techniques are flourishing and expanding, and for some diseases such as influenza have become part of a global surveillance activities. There are many more viral diseases, especially animal diseases, in need of this kind of development.

Seroepidemiologic Studies

Seroepidemiology is useful in public health and animal health investigations and in research to determine the prevalence or incidence of particular infections, to estimate population immunity (either through natural infection or vaccination) in a population or sub groups of a population, and to assess past exposure when a virus is newly identified. Seroepidemiological studies generally aim to detect either immunoglobulin M (IgM), indicating recent infection, or IgG, indicating infection in the past. When paired serum specimens are obtained from individuals several weeks apart, IgG can also be used to detect new infections if IgG antibodies are absent from the first specimen and appear in the second, or if antibody titers rise between the first and second sample. Correlation of serologic tests with clinical observations makes it possible to determine the ratio of clinical to subclinical infections and therefore to determine the proportion of asymptomatic cases. In the COVID-19 pandemic that emerged in 2019, this approach has been essential to identify the existence of asymptomatic carriers and to determine their role in the spread of the pandemic.

Digital Epidemiology

In the past two decades the volume and speed of online data available has increased exponentially. Data from online news sites, news-aggregation services, social networks, web searches, and crowdsourced data from volunteers among others – sometimes referred to as big data- , have been increasingly used to describe the epidemiology of viral diseases and have gradually moved away from experimental research to integrate the formal surveillance landscape. Data-aggregation systems such as the World Health Organization's Epidemic Intelligence from open sources (EIOS) uses advanced analytics such as natural language processing and machine learning to process large quantities of data and provide insights that complement traditional sources of epidemiological data such as laboratory and clinical data. The principal added value of digital epidemiology data is that it can detect the emergence or trends in disease without relying on healthcare seeking behavior. It can therefore describe the epidemiology of diseases for which healthcare is not typically sought, for example acute gastro intestinal illness or acute respiratory infections. It can also detect the emergence of viral threats before they are detected by surveillance systems embedded in healthcare. Several of these systems for examples have claimed to report the emergence of COVID-19 before any official statement from the World health Organization.

Sentinel Studies

Because of advanced diagnostic/serologic methods, sentinel studies can yield many valuable data in timely fashion about impending disease risks. For example, sentinel chicken flocks are set out for the early detection of the presence of arboviruses such as West Nile virus in the United States. These flocks are bled and tested weekly for the presence of virus or antiviral antibody; they provide an early warning of the levels of virus amplification that occur before epidemics.

Vaccine Trials

The immunogenicity, potency, safety, and efficacy of vaccines are first studied in laboratory animals, followed by small-scale closed trials on humans focusing on safety and immunogenicity, and finally in large-scale open trials focusing on efficacy. Such studies use randomized control trials to estimate the efficacy of vaccines. In most cases, there is no alternative way to evaluate new vaccines, and the design of trials has now been developed so that they yield maximum information with minimum risk and acceptable cost. Because trial size is limited by cost and logistical barriers observational studies are required post vaccine implementation to monitor the coverage, safety, effectiveness and impact of vaccines.

Table 2 Criteria for disease causation: A unified concept appropriate for viruses as causative agents of disease, based on the Henle–Koch postulates, and modified by A. S. Evans

1. Prevalence of the disease is significantly higher in subjects exposed to the putative virus than in those not so exposed.
2. Incidence of the disease is significantly higher in subjects exposed to the putative virus than in those not so exposed (prospective studies).
3. Evidence of exposure to the putative virus is present more commonly in subjects with the disease than in those without the disease.
4. Temporally, the onset of disease follows exposure to the putative virus, always following an incubation period.
5. A regular pattern of clinical signs follows exposure to the putative virus, presenting a graded response, often from mild to severe.
6. A measurable host-immune response, such as an antibody response and/or a cell-mediated response, follows exposure to the putative virus. In those individuals lacking prior experience, the response appears regularly, and in those individuals with prior experience, the response is anamnestic.
7. Experimental reproduction of the disease follows deliberate exposure of animals to the putative virus, but nonexposed control animals remain disease free. Deliberate exposure may be in the laboratory or in the field, as with sentinel animals.
8. Elimination of the putative virus and/or its vector decreases the incidence of the disease.
9. Prevention or modification of infection, via immunization or drugs, decreases the incidence of the disease.
10. “The whole thing should make biologic and epidemiologic-sense.”

Mathematical Modeling

From the time of William Farr, who studied epidemic disease problems in the 1870s, mathematicians have been interested in “epidemic curves”, secular trends in the incidence of infectious diseases, and forecasting how the spread of a disease may evolve. Advanced computational power has enabled increasingly complex modeling of the spread of a disease, with an increasing number of parameters being taken into account. Despite the availability of computational power, high profiles models attempting to predict the long term evolution of outbreaks of Ebola (2014–16) or COVID19 (2019–2020) have been consistently wrong and despite the aspiration to predict how outbreaks evolve on the long run, forecasting beyond a few weeks remains very difficult. Models are however an important tool to understand how diseases spread in population in the present, and to assess the likelihood of future scenarios. Mathematical models are also the key decision-making tools to inform the introduction of new vaccines in population as they estimate impact and cost effectiveness.

Data Sharing, Data Privacy and Ethics

As the volume and velocity of data increase, the possibilities of more granular insight about the epidemiology of viral diseases, data linkage and innovative approaches to analytics such as machine learning and natural language processing have vastly increased. Yet these approaches are only as good as the data they rely on. Many datasets remain inaccessible due to privacy and security concerns, and quality is variable due to a lack of standardization and completeness. A global consensus to share quality data in a proportionate, ethical and privacy-preserving manner, where it can improve the health of population is still lacking. Successful control of infectious diseases benefits individuals and communities globally and can be considered a global public good and data is a critical tool that helps achieve this. Governments should therefore improve availability and accessibility of their epidemiological data in order to maximize the generation of evidence to inform control of infectious diseases.

Proof of Causation

One of the landmarks in the history of infectious diseases was the development of the Henle-Koch postulates which established the evidence required to prove a causal relationship between a particular infectious agent and a particular disease. These simple postulates were originally drawn up for bacteria, but were revised in 1937 by Rivers and again in 1982 by Evans in attempts to accommodate the special problem of proving disease causation by viruses (Table 2). With a similar purpose, Austin Bradford Hill established a series a criteria for causation in 1965 (Table 3). While the limit of such criteria remain a debate among epidemiologists, using a set of criteria to determine causality remains a relevant approach since statistical approaches are focused on testing association (rather than causation) between exposure and outcome. In many cases, virologists have had to rely on indirect evidence, “guilt by association”, with associations based on epidemiologic data and patterns of serologic positivity in populations. Today, many aspects of epidemiologic investigation play roles, especially in trying to distinguish an etiological, rather than coincidental or opportunistic relationship between a virus and a given disease. The availability and combination of epidemiological evidence combined with molecular techniques can greatly help with determining the likelihood of causation between exposure and outcome where a statistical association is found (Table 3).

For example, early in the investigation of human acquired immunodeficiency syndrome (AIDS), before its etiology was established, many kinds of viruses were being isolated from patients and many candidate etiologic agents were being advanced. Prediction that the etiologic agent would turn out to be a member of the family *Retroviridae* was based upon years of veterinary research on animal retroviruses and animal retroviral diseases. This prediction was based upon recognition of common biologic and pathogenetic characteristics of AIDS and animal retroviral diseases. This prediction guided many of the early experiments to find the etiologic agent of AIDS; later, after human immunodeficiency virus (HIV1) was discovered, its morphological similarity to equine infectious anemia virus, a prototypic member of the genus *Lentivirus*, family *Retroviridae*, was the key to unraveling

Table 3 Bradford Hill criteria for causation

1. **Strength** (effect size): A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.
2. **Consistency** (reproducibility): Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.
3. **Specificity**: Causation is likely if there is a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.
4. **Temporality**: The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).
5. **Biological gradient** (dose-response relationship): Greater exposure should generally lead to greater incidence of the effect. However, in some cases, the mere presence of the factor can trigger the effect. In other cases, an inverse proportion is observed: greater exposure leads to lower incidence.^[1]
6. **Plausibility**: A plausible mechanism between cause and effect is helpful (but Hill noted that knowledge of the mechanism is limited by current knowledge).
7. **Coherence**: Coherence between epidemiological and laboratory findings increases the likelihood of an effect. However, Hill noted that "... lack of such [laboratory] evidence cannot nullify the epidemiological effect on associations".
8. **Experiment**: "Occasionally it is possible to appeal to experimental evidence".
9. **Analogy**: The use of analogies or similarities between the observed association and any other associations.

confusion over the fact that the human virus killed host lymphocytes rather than transforming them as typical oncogenic retroviruses would do. Ever since, this essence of comparative medicine has been guiding HIV/AIDS research in many areas, including drug design, diagnostics, and vaccine development. HIV/AIDS epidemiologic research has often been intertwined with research on the several simian immunodeficiency viruses (SIVs).

Conclusion: Implications for Disease Prevention

Knowledge of the epidemiology and modes of transmission of infectious diseases is critical to the development and implementation of prevention and control strategies. Data on incidence, prevalence, and mortality contribute directly to the establishment of priorities for prevention and control programs while knowledge of viral characteristics and modes of transmission are used in deciding prevention and control strategies including vaccine development and delivery, environmental improvements, enhancement of nutritional status, improvement in personal hygiene, and behavioral changes. Epidemiology has become part of a wide multidisciplinary network of disciplines that include laboratory, data and social sciences in order to achieve this objective.

See also: Zoonosis, Emerging and Re-Emerging Viral Diseases

Further Reading

- Evans, A.S., Kaslow, R.A. (Eds.), 1997. *Viral Infections of Humans: Epidemiology and Control*, fourth ed. New York: Plenum Medical Book Company.
- Hawker, J., Begg, N., Reintjes, R., *et al.*, 2019. *Communicable Disease Control and Health Protection Handbook*, fourth ed. Hoboken: John Wiley and Sons.
- Heymann, D.L. (Ed.), 2015. *Control of Communicable Diseases Manual*, twentieth ed. Washington, DC: American Public Health Association Press.
- Mandell, G.L., Bennett, J.E., Dolin, R. (Eds.), 2000. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, fifth ed. New York: Churchill Livingstone.
- Monto, A., 2005. The epidemiology of viral infections. In: Mahy, B.W.J., ter Meulen, V. (Eds.), *Virology: Topley and Wilson's Microbiology and Microbial Infections* 1. London and Washington, DC: Hodder Arnold and ASM Press.
- Murphy, F.A., Gibbs, E.P.J., Horzinek, M.C., Studdert, M.J., 1999. *Veterinary Virology*, third ed. New York: Academic Press.
- Nathanson, N., 2006. Epidemiology. In: Knipe, D.M., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A. (Eds.), *Fields Virology*, fifth ed. Philadelphia: Lippincott Williams and Wilkins.
- Nathanson, N., Murphy, F.A., 1997. Evolution of viral diseases. In: Nathanson, N., Ahmed, R., Griffin, D.E., *et al.* (Eds.), *Viral Pathogenesis*. Philadelphia: Lippincott-Raven Press.
- Nowak, M.A., May, R.M., 2001. *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford: Oxford University Press.
- Smolinski, M.S., Hamburg, M.A., Lederberg, J., 2002. *Emerging Microbial Threats to Health in the 21st Century*. Washington, DC: National Academy Press, Institute of Medicine/National Academy of Sciences.
- Smolinski, M.S., Hamburg, M.A., Lederberg, J. (Eds.), 2003. *Microbial Threats to Health, Emergence, Detection and Response*. Washington: National Academies press, p. 367.
- Thrusfield, M.V., Bertolo, G., 2005. *Veterinary Epidemiology*, third ed. London: Blackwell Publishing.

Zoonosis, Emerging and Re-Emerging Viral Diseases

Janet M Daly, University of Nottingham, Sutton Bonington, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

BCE Before the common era
HBV Hepatitis B virus
HCMV Human cytomegalovirus
HCV Hepatitis C virus
HIV Human immunodeficiency virus
HPV Human papilloma virus
HSV Herpes simplex virus
IAV (IBV, ICV) Influenza A (B, C) virus

MERS Middle East respiratory syndrome
NNRTI Non-nucleoside reverse transcriptase inhibitor
NRTI Nucleoside reverse transcriptase (RT) inhibitor
RSV Respiratory syncytial virus
SARS-CoV Severe acute respiratory syndrome coronavirus
SBV Schmallenberg virus
VZV Varicella zoster virus
WHO World Health Organization
WNV West Nile virus

Glossary

Endemic Infection is constantly/normally present in a geographic area.
Epidemic Rapid spread of infection in a large number of people.
Outbreak Greater number of cases of disease than would normally be expected in a defined population, geographical area or season.
Pandemic An epidemic involving a large proportion of the world population.

Reassortment The process of swapping whole gene segments between two viruses with segmented genomes during replication when a cell is infected by more than one virus strain.
Recombination The process of swapping sections of genome between two viruses during replication when a cell is infected by more than one virus strain.
Virulence Harm a pathogen (virus) does to its host.
Zoonosis A disease that can be transmitted from veterinary species to humans.

Introduction

Emergent viruses include viruses that were not previously identified (emerging) or had either not occurred in a species before or had occurred previously but were limited geographically (re-emerging). The potential for emergent viruses to have a devastating impact on human populations is not a novel phenomenon. In the 1500s, smallpox was introduced into the Americas, resulting in an estimated 10–15 million deaths and the downfall of the Aztec civilization. Nonetheless, even after controlling for confounding factors such as increased awareness and improved surveillance, it is estimated that the worldwide annual occurrence of viral outbreaks has increased substantially in the past 40 years. The 21st century has witnessed the emergence of viral hemorrhagic fevers such as Ebola and Lassa, arthropod-borne virus (arbovirus) infections such as Zika and chikungunya and respiratory infections such as the SARS (severe acute respiratory syndrome) (Addendum added during preparation for publication: This article was written shortly before the emergence of SARS-CoV-2. It therefore refers to SARS-CoV-1 as ‘SARS’) and MERS (Middle East respiratory syndrome) coronaviruses (CoV), to name just a few. Despite improvements in prevention, early detection, control and treatment, emergent viruses remain one of the main causes of human deaths globally and can also have a dramatic impact on livestock species. Viral emergence is typically perceived as a cause of epidemics in which viruses spread quickly in a naïve population. However, not all viruses will adapt to the new host. In general, as the number of susceptible individuals in a population decreases, the epidemic will die out, but as long as the source or reservoir for the virus is still around, there is always a risk that the disease could return.

Although viral zoonoses account for the majority of emerging infectious diseases in the human population, the biological and epidemiological barriers to interspecies transmission are high. Five stages have been proposed to represent the transition from a pathogen that infects only veterinary species (stage 1) to a human-only pathogen (stage 5; Fig. 1). In stage 2, only primary human infection occurs with no onward transmission – examples include Nipah, rabies and West Nile viruses. In stage 3, there are pathogens such as Ebola virus that undergo a few cycles of secondary transmission so that occasional human outbreaks occur. Stages 2 and 3 equate to “spillover infection”; an infection that results in a dead-end infection with no onward transmission or a stuttering chain of limited transmission in the new host. Stage 4 includes viruses that have a natural (sylvatic) cycle of primary transmission from the non-human host but can also cause larger outbreaks with secondary human transmission. The relative importance of primary and secondary transmission can vary. For example, with yellow fever virus, sylvatic transmission is more important than human-to-human spread, while for human-adapted influenza A virus (IAV) there is sustained human transmission.

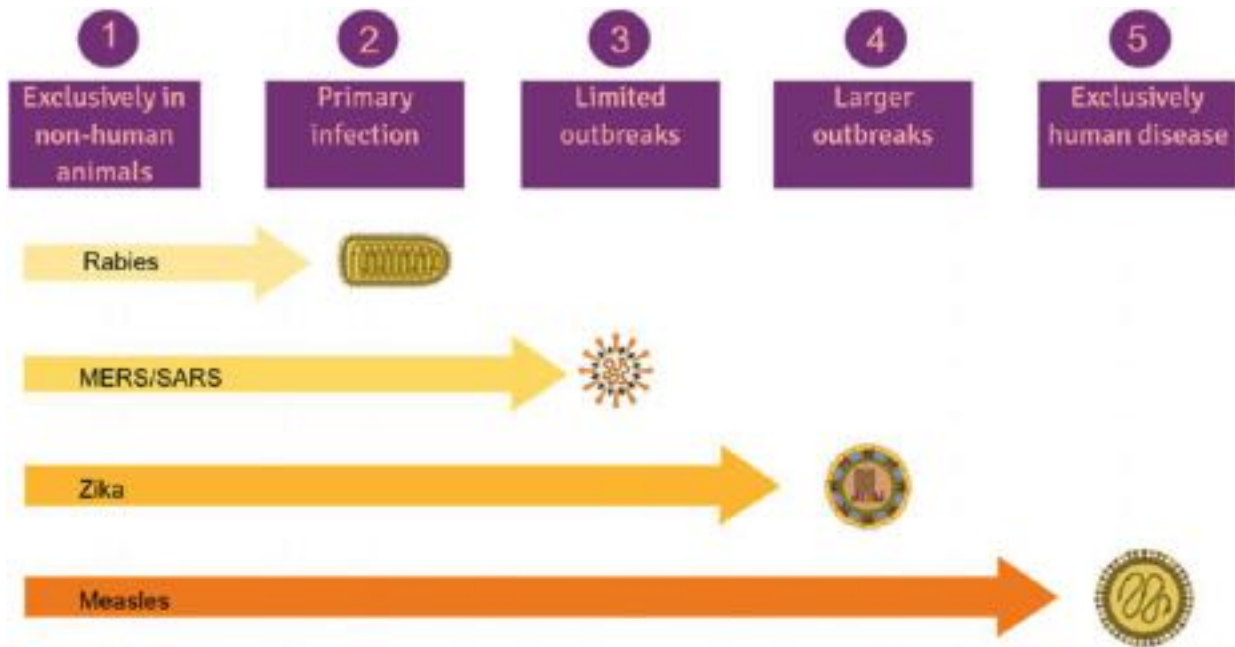


Fig. 1 The stages of transition of a zoonotic virus from infection of a veterinary species only (stage 1) to becoming a human-only pathogen (stage 5).

Factors in Interspecies Transmission

Viral emergence is often a result of a host shift, which is a relatively rare event and will usually only reach stage 2 due to lack of adaptation of the virus to the novel host. If a virus does not cause severe disease in a novel species, the majority of interspecies transmission events will go unnoticed. There are multiple factors influencing the occurrence of interspecies transmission (or “host jumps”).

Epidemiological or Ecological Barriers

Sustained contact between reservoir hosts (often bats, rodents or birds) and other species increases the likelihood of the emergence of a virus adapted to infect and replicate in a new host. Most of the human-only (stage 5) infectious diseases in temperate zones are thought to have arisen in the Old World (Africa, Asia and Europe), often acquired from domesticated animals. For example, measles is thought to have originated from rinderpest virus of cattle, which has now been eradicated. The most common domesticated animals, including cattle and other livestock species (goats, pigs and sheep) and cats and dogs were domesticated between 8000 and 13,000 BCE, with poultry and equids domesticated slightly more recently (3500–6000 BCE). As a consequence, most novel interspecies transmissions now occur from wildlife species, although these may involve domesticated animals as intermediate hosts, for example pigs act as an intermediate host for the transmission of influenza A and Nipah viruses. Shifts in the epidemiology and ecology of infectious diseases play a significant role in the increasing occurrence of viral emergence (see factors in increasing emergence of viral diseases). There has also been speculation that elimination of a virus from a species might leave a vacant ecological niche that could be filled by a related virus. For example, since rinderpest virus was eradicated from cattle by vaccination, there has been an upsurge of another member of the same virus family, peste des petits ruminants (PPRV), possibly due to the lack of cross-reactive immunity previously provided by rinderpest virus.

Host–Pathogen Interactions

More closely-related species tend to be more susceptible to viruses crossing species barriers because they are more likely to share similar receptors recognized by the virus or have similar innate immune response mechanisms that the virus is already adapted to circumvent.

Viral Factors

Receptor specificity

Receptor specificity is usually the first barrier encountered by a virus during infection. Influenza A virus (IAV) provides a classic example of the role played by receptor specificity in viral emergence. The viral surface protein responsible for entry into host cells is the hemagglutinin (HA). Notwithstanding the recent identification of novel IAV in bats, wild aquatic ducks, particularly in the family *Anatidae*, are regarded as the main reservoir host of IAV. In these hosts, the HA recognizes sialic acids with an $\alpha 2,3$ -linkage to

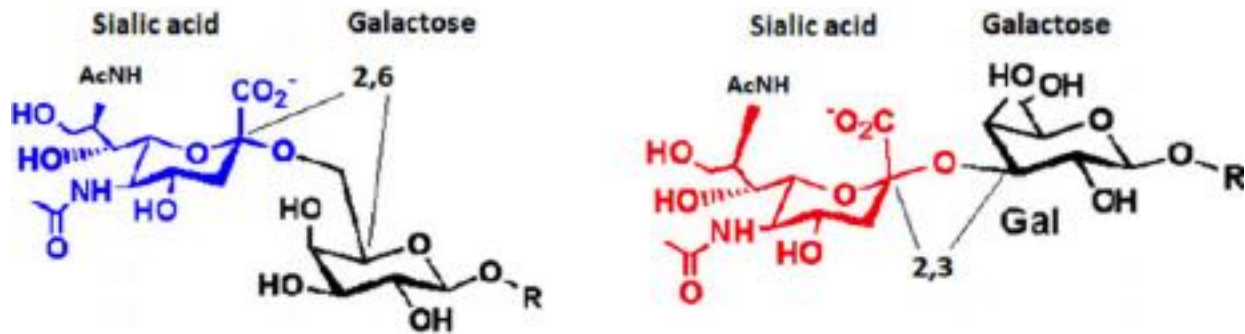


Fig. 2 The α 2,3 or α 2,6 glycosidic linkage between sialic acid and galactose on cell-surface glycoproteins.

Table 1 Summary of classification of (re-)emergent zoonotic viruses described in this text (International Committee on Taxonomy of Viruses Virus Taxonomy: 2018b Release)

Family-viridae	Genus	Species	Genome
Arena ~	<i>Mammarenavirus</i>	Lassa virus	(-) ssRNA - segmented
Corona ~	<i>Betacoronavirus</i>	Severe acute respiratory syndrome-related coronavirus	(+) ssRNA
Filo ~	<i>Ebolavirus</i>	Zaire Ebolavirus	(-) ssRNA
Flavi ~	<i>Flavivirus</i>	Dengue virus	(+) ssRNA
		Zika virus	
Hanta ~	<i>Orthohantavirus</i>	Hantaan virus	(-) ssRNA - segmented
Orthomyxo ~	<i>Alphainfluenzavirus</i>	Influenza A virus	(-) ssRNA - segmented
Paramyxo ~	<i>Henipavirus</i>	Hendra henipavirus	(-) ssRNA
		Nipah henipavirus	
Peribunya ~	<i>Orthobunyavirus</i>	Schmallenberg virus	(-) ssRNA - segmented
Phenui ~	<i>Phlebovirus</i>	Rift Valley fever phlebovirus	(-) ssRNA - segmented
Reo ~	<i>Orbivirus</i>	Bluetongue virus	dsRNA - segmented
Retro ~ (<i>Orthoretrovirinae</i>)	<i>Lentivirus</i>	Human immunodeficiency virus	(+) ssRNA - RNA reverse transcribing

a galactose (Fig. 2), which are expressed in abundance on both respiratory and gastrointestinal epithelial cells in ducks. Sialic acids with an α 2,3-linkage also predominate in the equine upper respiratory tract; therefore, IAV are believed to have become established in horses as a result of direct transmission from the avian reservoir. In contrast, the sialic acids on cells of the human upper respiratory tract predominantly have an α 2,6-linkage, limiting the likelihood of direct transmission of avian influenza viruses to people. However, the 1918 influenza pandemic appears to have involved direct transmission of an avian IAV that adapted to recognize receptors with α 2,6-linkage.

Ability to adapt

Many of the (re-)emerging viruses have an RNA genome (Table 1). In DNA replication, proofreading enzymes remove misincorporated nucleotides from newly synthesized DNA. This does not occur in RNA replication which, combined with the high rates of replication in RNA viruses, results in a much greater capacity for rapid evolution and hence adaptation to changing environments and new hosts. Furthermore, several of the viral families associated with (re-)emerging disease have segmented RNA genomes. A major consequence of this genome structure is that reassortment can occur, in which genome segments are swapped during replication. When a host cell is infected at the same time with distinct variants, this can result in a virus that combines properties that were previously unique to each parent, which may have a neutral effect or be disadvantageous, but may confer a fitness advantage on the new novel virus. The three most recent pandemics of human IAV involved reassortment. The 1957 pandemic strain was a reassortant of a human H1N1 virus with three gene segments (HA, NA and PB1) from an avian H2N2 virus. The 1968 pandemic strain was a reassortant of the H2N2 human virus in which the HA and PB1 genome segments were from an H3 avian virus. Finally, the 2009 pandemic H1N1 virus appears to have originated from at least three parental strains (avian H1N1, swine H1N1 and human H3N2) through a series of reassortment events in pigs.

Recombination is another larger scale mechanism of genetic variation that can occur in viruses with non-segmented RNA or DNA genomes. The frequency of recombination varies greatly between viruses – it is less common in negative-stranded RNA viruses, but frequent in retroviruses. Recombination, like reassortment, can result in new genome combinations with a fitness advantage, including more efficient replication in a new host. Coronaviruses have the largest single-stranded, positive-sense RNA genome currently known and it is thought to be as a consequence of this that they uniquely encode RNA processing and editing

enzymes that increase the fidelity of replication. However, they have a particularly high rate of homologous recombination (a result of template switching where the corresponding sequence is exchanged between co-infecting viruses). It is thought that recombination contributes to emergence of novel coronaviruses such as turkey coronavirus and SARS-CoV.

Some emerging viruses, such as West Nile virus (WNV) only cause “dead-end” infections in incidental hosts such as equids and humans because there is insufficient viremia for onward transmission by the mosquito vector. Arthropod-borne viruses (arboviruses) such as WNV may be less able to adapt rapidly to new hosts, as changes that are beneficial to replication in the novel host may be detrimental to replication in the insect host.

Virulence

Viruses are often particularly virulent in a new host. It is generally believed that long-term co-evolution results in viruses that are less pathogenic to their natural or reservoir host, either due to tolerance on the part of the host and/or adaptation of the virus to an optimal level of virulence to maximize transmission. A classic example of viral adaptation was observed when myxoma virus, which causes mild disease in its natural host (American rabbits in the genus *Sylvilagus*), was introduced into Australia to control European rabbits (*Oryctolagus cuniculus*). Initially, fatality was close to 100%, but within a few years, the virus became attenuated. The attenuated virus strains have higher rates of transmission due to the host surviving for longer, providing more opportunity for transmission to occur. High levels of virulence in a new host are more likely due to an inappropriate immune response than high levels of virus replication.

In addition to viruses often being particularly virulent in a new host, variation in the ability of a virus to cross species barriers may also be related to changes in its virulence before transmission. This may be illustrated by the interspecies transmission of an equine-adapted IAV to dogs. This was associated with a more virulent strain of the equine virus that caused more frequent and prolonged coughing, which in turn could increase transmissibility of the virus.

Factors in Increasing Emergence of Viral Diseases

There are numerous factors associated with the apparent increase in emergence of viral diseases, most of which have a human element and many of which are inter-related. These include increasing population densities, encroachment on wilderness, dispersal of vectors and pathogens through trade, transport, and migration, and climate change. Growth of the human population has led to increasing population densities and an increase in intensive farming. Intensification of farming can result in high rates of contact between individuals, which may also be more susceptible to infection due to stress, resulting in more opportunities for virus transmission leading to adaptation. On the other hand, backyard farming is often associated with housing different species together and more opportunity for contact with wild animals, increasing the risk of interspecies transmission. A more certain factor in increasing risk of interspecies transmission resulting from the growing population is human encroachment into wilderness habitats bringing them and domesticated species into closer contact with reservoir hosts including monkeys and bats.

Dispersal of vectors and pathogens through trade, transport, and migration has been a factor in the spread of viruses to new regions for centuries. Global distribution of *Aedes aegypti*, the mosquito that transmits dengue, chikungunya, Zika, Mayaro and yellow fever viruses, began with spread from Africa to the Americas associated with the slave trade between the 15th and 19th centuries. During the 18th and 19th centuries, it was spread by trading with Asia with further spread by troop movements after the second-world war. Since the 1980s, the role in dispersal of mosquitoes of worldwide trading of used tires, which when piled in the open can collect small puddles of water provide a mosquito breeding habitat, has been recognized. Historical trade routes connected European countries with Asian and African countries and North America. However, new air routes between countries in sub-Saharan Africa and the Caribbean and Central America could underpin the spread of viruses such as Zika and chikungunya. Whereas conditions are only likely to be favorable for exotic mosquito species to survive in Europe (at least for a short time) if they arrive in summer, conditions are more suitable in the Caribbean and Central America all year round. In the 21st century, viral diseases can be rapidly translocated from one side of the world to the next as billions of people fly between countries annually. The implications of this for dissemination of viral diseases were dramatically illustrated during the emergence of severe SARS in 2002–2003. The virus, which is transmitted in droplets generated by sneezing or coughing, spread to 25 other countries within a few months of its emergence in China with evidence for person-to-person transmission occurring during four international flights. Initially, public health authorities worldwide were ill-prepared for containing international spread of SARS, but no further cases of in-flight transmission were documented after the World Health Organization (WHO) introduced recommendations for screening measures at airports.

Some viruses that were previously controlled by vaccination have re-emerged for a variety of reasons, such as breakdown in public health systems because of civil strife. Others, such as measles, are re-emerging because of vaccine hesitancy.

Prevention and Control of Emerging Viruses

Appropriate measures to prevent and control emerging viruses are highly dependent on the individual situation. For example, what is known about the mode of transmission (either directly or by inference from related viruses) and whether diagnostic tests and vaccines are already available dictate the possible courses of action. Where vaccines are not yet available, or to augment vaccines, prevention of disease focuses on avoiding exposure to infection. Thus may involve education of the public and healthcare

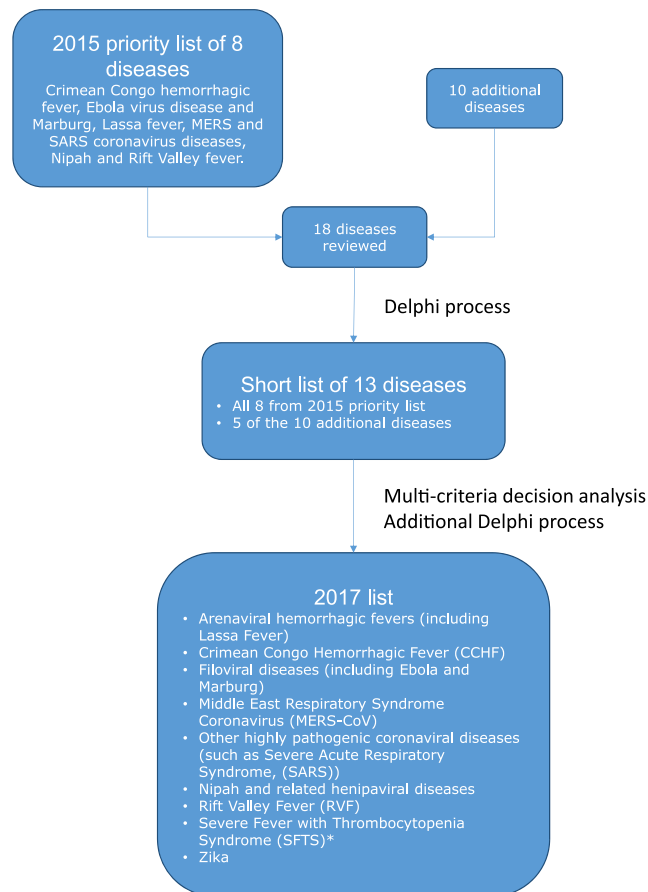


Fig. 3 World Health Organization list of priority diseases for the WHO R&D Blueprint.

workers to avoid behaviors that lead to disease transmission. For example, avoidance of contact with rats, which transmit Lassa fever virus, in Nigeria. More drastic measures may be taken to control outbreaks of an emerging virus in livestock, including culling infected animals (“stamping out”). It is generally easier to impose movement restrictions for livestock than people. More extreme and costly control measures can only be achieved with substantial investment from governments and non-governmental organizations. Increasingly, prevention of spread of emerging viruses requires international co-ordination – a role chiefly undertaken by the WHO and the World Organization for Animal Health (OIE).

In 2015, the WHO first published a priority list of eight emerging diseases likely to cause major epidemics for which current prevention and control measures are inadequate – all viral diseases. In 2017, the WHO refined the process for selecting these priority diseases, assembling a panel of experts to undertake a Delphi process to first create a shortlist then to rank the viruses on the shortlist using a multi-criteria decision analysis and finally to further refine the list with a second Delphi round (Fig. 3). The final list of 10 diseases proposed in 2017 formed the WHO “blueprint” of diseases requiring prioritized research and development efforts. The current list (after a review in 2018) contains Crimean-Congo hemorrhagic fever (CCHF), Ebola virus disease and Marburg virus disease, Lassa fever, Middle East respiratory syndrome coronavirus (MERS-CoV) and Severe Acute Respiratory Syndrome (SARS), Nipah and henipaviral diseases, Rift Valley fever (RVF), Zika and “disease X”. Inclusion of disease X, a pathogen not yet known to cause human disease, exemplifies the need for platforms for developing interventions that can be readily (and rapidly) adapted to deal with future epidemics caused by unforeseen pathogens.

There has been extensive investment of resources in efforts to predict the next pandemic, but increasingly it is recognized that it is critical to ensure enhanced ability to identify, through surveillance, and rapidly and appropriately respond to an emerging virus threat. However, as more data are acquired from emerging disease outbreaks, there is greater scope to develop risk-based models that can inform authorities where to focus surveillance efforts and adaptive mathematical models that can be used to help make appropriate decisions during the course of an outbreak.

Diagnosis

Diagnosis of viral disease by detection of virus was traditionally very resource-intensive with virus isolation requiring cell culture, inoculation of embryonated hen’s eggs or intracerebral inoculation of suckling mice or, for non-cultivable viruses, the use of electron microscopy. Attempts to isolate virus remain important for comprehensive viral characterization, and development and

testing of vaccines and antiviral drugs. However, more rapid diagnostic techniques are now available such as ELISA for detection of viral proteins and molecular methods such as PCR. The emergence of Schmallenberg virus (SBV), a midge-borne virus affecting ruminants that was hitherto unheard of, highlighted how technological advances have accelerated our capacity to identify emerging viruses. In the summer and autumn of 2011, the etiologic agent of a disease causing fever, decreased milk yield and diarrhea in cattle in Germany and Holland could not be identified using existing diagnostic tests. Metagenomic analysis was performed on samples taken in October 2011 and viral sequences most similar to viruses in the Orthobunyavirus genus of the family *Bunyaviridae* (according to the classification at the time, now the *Peribunyaviridae*), were identified, and SBV was subsequently isolated. This was the first example of next-generation sequencing being used to identify an emerging virus. Furthermore, the first commercial ELISA to detect antibodies to SBV entered the market in April 2012. The company was able to rapidly develop the indirect ELISA following the same principles used for a kit already developed for Rift Valley fever virus. Nonetheless accurate diagnosis of some groups of viruses continues to pose a challenge. For example, clinical diagnosis of important human flavivirus infections such as Zika, Japanese encephalitis, and dengue viruses is difficult because of the overlap in their geographical distribution (defined by their mosquito hosts) and their similar clinical disease spectrum. Transient viremia means that laboratory diagnosis by PCR-based methods frequently returns a false negative result and serological diagnosis is complicated by high levels of cross-reactivity of antibodies. Therefore, there is a need to improve the availability of reliable and cost-effective diagnostic tests that can be deployed wherever required to enable effective control and prevention measures to be implemented before an emerging virus has the opportunity to develop into an epidemic.

Treatment

Particularly when an emergent virus has not been identified previously, or did not cause sufficient human disease to warrant investment in vaccine development, antiviral drugs may be important in the control of disease. The first antiviral drug to be approved for human use in 1963 was Idoxuridine, a 5-substituted 2'-deoxyuridine analog active against herpes simplex virus. Since then, antiviral drugs with various modes of action (see [Table 2](#) for summary) have been approved for use against several human viral diseases. In the absence of an effective vaccine, vast amounts have been invested in development of anti-viral drugs for human immunodeficiency virus (HIV) and, more recently, hepatitis C virus (HCV). The development of antiviral drugs is a lengthy and costly process, but broad-acting antiviral drugs such as ribavirin may be effective against emerging viruses. In addition, interferons, immunostimulators, oligonucleotides, and antimetabolic inhibitors can be used against various human viruses.

In general, antiviral drugs are not widely used to treat livestock animals due to their expense, but there may also be a desire to restrict antiviral drugs to treatment of severe human cases to limit the opportunity for development of widespread drug-resistant variants. For example, it is thought that injudicious use of the IAV antiviral amantadine in poultry in Egypt led to a high proportion of resistance to this drug among H5N1 avian influenza virus isolates and veterinary pharmaceuticals containing antiviral compounds are now banned in Egypt. There are also concerns that the active metabolite of the IAV antiviral oseltamivir is not removed by conventional sewage treatment and can be found in river water at concentrations that might be sufficiently high to drive selection of drug-resistant variants of IAV in ducks.

Table 2 Summary of licensed antiviral drugs

Drug group	Approved use	Mechanism (s) of drug action
5-substituted 2'-deoxyuridine analogs	HSV, VZV	Inhibit viral DNA polymerase
Nucleoside analogs	HBV	Inhibit viral DNA polymerase
Pyrophosphate analogs	HCMV, (acyclovir resistant HSV)	Inhibit viral DNA polymerase
NRTIs	HIV, HBV	Target HIV RT to competitively inhibit DNA synthesis
NNRTIs	HIV-1	Bind directly to HIV RT and inhibits DNA synthesis
Protease inhibitors	HIV, HCV genotype 1	Block active site of protease to prevent cleavage of viral precursor proteins
Integrase inhibitors	HIV	Target HIV integrase to inhibit the integration of viral DNA into human chromosomes
Entry inhibitors	RSV, HSV, HIV, VZV	Specifically block binding of viral surface proteins to entry receptors
Acyclic guanosine analogs	HSV, VZV, HCMV	Inhibit viral DNA polymerase activity
Acyclic nucleoside phosphonate analogs	HCMV, HIV, HBV	Inhibit the activity of HCMV DNA polymerase
Direct-acting antivirals (DAAs) against HCV	HCV	Four classes used in combination that inhibit NS3/4A protease, inhibit NS5A, and nucleoside analogs or non-nucleoside analogs that inhibit NS5B polymerase
Direct-acting antivirals (DAAs) against influenza viruses	IAV (IBV and ICV)	Three classes that target matrix protein 2 to inhibit viral uncoating, neuraminidase to inhibit virus release from host cells or activity of RNA polymerase
Ribavirin	IAV, HCV, RSV, some hemorrhagic fever viruses	Target viral RNA polymerase to inhibit mRNA synthesis

Key: HBV, hepatitis B virus; HCMV, human cytomegalovirus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; HPV, human papilloma virus; HSV, herpes simplex virus; IAV (IBV, ICV), influenza A (B, C) virus; NRTI, nucleoside reverse transcriptase (RT) inhibitor; NNRTI, non-nucleoside reverse transcriptase inhibitor; RSV, respiratory syncytial virus; VZV, varicella zoster virus.

Because direct-acting antivirals (DAA) are often associated with the rapid development of antiviral resistance, there is increasing interest in the potential of drugs that target cellular pathways required for viral replication rather than the virus itself. A drawback of this approach is that drugs that target key host cell pathways are more likely to be associated with adverse side-effects than virus-specific drugs. Nonetheless, there are opportunities to “repurpose” drugs that have already been through phase I and II clinical trials, for example as potential cancer therapeutics. Combining *in silico* structure-based screening and high throughput screening assays of the large chemical libraries now available may identify potential compounds for treatment of emerging viruses.

Vaccination

Vaccination is generally considered the most effective method to prevent infectious diseases and strategic vaccination programs were central to the eradication of smallpox and rinderpest, the only two viral diseases to have been eradicated to date. However, development and safety testing of vaccines against novel viruses takes time and financial resources. Furthermore, it may not be apparent in the early stages of an outbreak whether investment in vaccine development will ultimately be profitable. Therefore preparedness for emerging and re-emerging viruses is largely contingent on international organizations such as GAVI, the vaccine alliance founded by the Bill and Melinda Gates Foundation. For example, GAVI funds a continually replenished emergency stockpile of 6 million doses of yellow fever virus (YFV) vaccine. Compassionate use of an investigational vaccine that had not yet undergone the full testing required to be licensed was recently permitted for the control of Ebola.

“Ring vaccination”, in which only those directly exposed or most at risk of infection are vaccinated, is a strategic approach to contain a virus outbreak when, for example, limited doses of vaccine are available. The approach is particularly effective in the control of outbreaks in livestock when used in combination with a diagnostic test that allows a DIVA (differentiation of infected from vaccinated animals) strategy to be employed. This approach was used when equine IAV first emerged in Australia in 2007 by vaccinating with a canarypox-vectored vaccine expressing only the IAV hemagglutinin and screening with an ELISA that detects antibodies to the nucleoprotein, which will only be expressed in infected animals. However, in the case of a novel emerging virus disease of livestock, inactivated virus vaccines are more often the first to be licensed. When there is no capacity to distinguish antibodies that result from vaccination rather than infection, it can be difficult to declare a region free of disease leading to prolonged trade restrictions being imposed.

Control of Mosquito Vectors

The *Aedes aegypti* mosquito has several behavioral features that make it particularly difficult to avoid contact with. It generally lives indoors and near people, moves very quickly and may bite repeatedly, often around the feet and ankles. Because the bite is not particularly painful, it can go unnoticed, and it also feeds during daylight hours, therefore using bed nets is not effective.

Traditional mosquito control programs involving the use of insecticides have had limited long-term success due to the development of resistance. The release of large numbers of radiation-sterilized insects has been successful for the control of several insect species that are agricultural pests, but less so for mosquitoes. “Transgenic” sterile male mosquitoes have been developed. For example, genetically-modified male mosquitoes that are homozygous for a late-acting dominant lethal transgene that is transmitted to the embryo *via* the sperm and leads to death of the zygote. Disadvantages of this method are that it only places a temporary restriction on disease transmission by reducing mosquito numbers, but population levels will return to normal levels when release of sterile insects stops. A novel approach is to release mosquitoes infected with an intracellular bacterium called *Wolbachia*. *Wolbachia* is a genus of gram-negative bacteria that infects many different insect species. Introducing strains from fruit flies into mosquitoes shortens the life span of female mosquitoes, which are responsible for disease transmission, but, importantly, some strains also increase the resistance of mosquitoes to infection with various pathogens. *Wolbachia* is a maternally inherited endosymbiont; therefore, release of *Wolbachia*-infected mosquitoes replaces indigenous mosquito populations with mosquitoes that block pathogen transmission. Early trials of the *Wolbachia* approach have focused on the control of dengue. Dengue is the most significant human arboviral disease with an increasing number of countries at risk for dengue infection. A field trial in an Australian city that involved the release of around 4 million *Wolbachia*-infected *Aedes aegypti* mosquitoes over an area of 66 square kilometers has shown promising results. Further trials are being conducted in Indonesia, where there is a much higher incidence of dengue, and areas where no mosquitoes are released are included. Release of *Wolbachia*-infected mosquitoes in large cities in Brazil and Colombia will also determine whether the strategy is effective in densely-populated areas with high incidence of dengue.

Concluding Remarks

Recent human epidemics caused by Ebola and Zika viruses have fueled increased research efforts and attracted media attention to emerging viruses. With increased globalization, density of populations, climate change and changes in habitat, there is increased risk of interaction within and between species. This provides more opportunities for the emergence of viruses by increased contact and interspecies transmission as well as spread of viruses into geographic regions where they have not been found before. Whilst a major driver of emerging virus disease research is public health concern over zoonotic infections, emerging viral diseases may have devastating effects on veterinary species with consequent impact on security of food production and the survival of wildlife species.

Increasingly, there is also the potential for emerging viruses to affect pet animals, causing welfare issues and distress to owners. This article has focussed exclusively on animal viruses, but emerging plant viruses also pose a serious threat to agricultural production, and similar factors, such as changes in agricultural practices and global transport of plant materials, have led to increased emergence of plant viruses. Understanding the drivers of viral emergence is vital to inform efforts to mitigate against its impact and this requires a one-health approach involving collaboration between the human and animal health and environmental and social fields.

Further Reading

- De Clercq, E., Li, G., 2016. Approved antiviral drugs over the past 50 years. *Clinical Microbiology Reviews* 29, 695–747.
- Jones, K.E., Patel, N.G., Levy, M.A., *et al.*, 2008. Global trends in emerging infectious diseases. *Nature* 451, 990–993.
- Lauring, A.S., Frydman, J., Andino, R., 2013. The role of mutational robustness in RNA virus evolution. *Nature Reviews Microbiology* 11, 327–336.
- Peeling, R.W., 2018. Epidemic preparedness: Why is there a need to accelerate the development of diagnostics? *The Lancet Infectious Diseases* 19, 172–178.

Antiviral Innate Immunity: Introduction

Friedemann Weber, FB 10 – Institute for Virology, Justus Liebig University Giessen, Giessen, Germany

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Apoptosis A form of programmed cell death.

Apoptotic bodies Remnants of cells which underwent apoptosis.

Complement system A pathogen-triggered cascade of biochemical reactions involving more than 20 soluble and cell-bound proteins. Complement activation results in opsonization, priming of humoral immune responses, and perforation of membranes.

Cytokines Proteins which mediate cell–cell communication related to pathogen defense. Secreted by immune cells or tissue cells.

Innate immunity Physical and chemical barriers, cells, cytokines, and antiviral proteins which exclude, inhibit, or slow down infection with little specificity and without much adaptation or generation of a long-lasting protective memory.

Interferons (IFNs) Cytokines mediating antiviral activity. Distinguished into type I (IFN- α/β), type II (IFN- γ), and type III (IFN- λ). Type I and type III IFNs directly mediate antiviral activity in responding cells, whereas type II IFN is more immunomodulatory.

Interferon-stimulated response element (ISRE)

A promoter element common to all type I IFN-stimulated genes.

Opsonization Tagging of infected cells or pathogens for destruction by phagocytic cells.

Pathogen-associated molecular patterns (PAMPs) Molecular signatures of pathogens used by the innate immune system to distinguish self from non-self. Often highly repetitive patterns.

Pattern recognition receptors (PRRs) Intracellular and extracellular receptors recognizing specific PAMPs.

Phagocytosis Uptake of particles by cells.

Introduction

Viruses attempting to conquer a mammalian body are faced with an impressive array of hurdles. “Innate immunity” in a wider sense comprises all sorts of factors which exclude, inhibit, or slow down infections in a rapid manner but with little specificity and without adaptation or generation of a long-lasting protective memory. Many of these efficient and not at all primitive defenses are evolutionarily old and can be found in all metazoans. For the sake of brevity, however, the discussion in this article is mostly restricted to mammals. RNA interference, the innate immune system of plants and non-vertebrates, is not covered.

Mammalian innate immune defenses against virus infections can be divided into several distinct parts such as mechanical and chemical barriers (not further mentioned here), defensins, complement system, phagocytic/cytolytic cells of the immune system which act in a nonspecific manner, and cytokines.

Defensins

Defensins are small, cysteine-rich cationic, amphipathic peptides with broad activity against bacteria, fungi, and viruses. They are produced by immune cells and skin and mucosal epithelial cells, and are present on epithelia and in body fluids. On top of their constitutive expression levels, defensin genes can be induced by viral infection. Their most common antimicrobial function is the formation of destructive pores in membranes of pathogens including enveloped viruses. Defensins can however also block infection by enveloped and non-enveloped viruses alike by aggregating the particles, blocking receptor binding, inhibiting virus entry, particle uncoating or intracellular trafficking, interfere with essential cell signaling or viral gene expression, or act by other, ill-understood mechanisms. Moreover, besides these direct antiviral activities, defensins were shown to attract immune cells and modulate adaptive immune responses. Since defensins activities are mostly studied in cell culture, it is thought that their immunomodulatory action is currently underestimated.

Complement System

The complement system (which “complements” the adaptive immune system in the defense against pathogens) primes the adaptive immune response and is also directly effective against pathogens. Complement activation is achieved by specific receptors recognizing pathogens or immunocomplexes. Three different pathways are being distinguished which are termed the classical pathway (triggered by antigen–antibody complexes), the mannan-binding lectin pathway (triggered by lectin binding of pathogen surfaces), and the alternative pathway (triggered by complement factor C3b-coated pathogen surfaces). They all activate a cascade of reactions involving more than 20 soluble and cell-bound proteins, thus resulting in a rapid and massive response. The

complement system is able to (1) tag infected cells and pathogens for destruction by phagocytic cells (opsonization), (2) prime humoral immune responses, and (3) perforate membranes of infected cells by the membrane-attack complex. In response, viruses have evolved effective countermeasures such as incorporation of cellular complement-regulatory proteins into particles or expressing specific inhibitors in infected cells.

Cellular Innate Immunity

Macrophages/monocytes, granulocytes, neutrophils, natural killer cells, and dendritic cells belong to the cellular branch of the innate immune system. Monocytes circulate in the bloodstream for several hours before they differentiate into macrophages. These potent phagocytic cells either continue patrolling or they permanently settle in particular tissues (i.e., the Kupffer cells of the liver), being able to rapidly remove viral particles and apoptotic bodies. Activated macrophages also synthesize inflammatory cytokines such as IFN- γ and tumor necrosis factor (TNF)- α , thus triggering an adaptive immune response. Granulocytes are also able to remove viral particles and apoptotic bodies by phagocytosis. They are rapidly attracted to inflammatory sites and enter the tissue by transendothelial migration. Both macrophages and granulocytes cleave the ingested viral proteins into fragments and present them to T lymphocytes. Neutrophils (polymorphonuclear cells) are also phagocytosing pathogens, among other directly acting activities. Their most fascinating ability is however the release of neutrophil extracellular traps (NETs), which are networks of proteins and chromatin that immobilize and inactivate extracellular pathogens. Natural killer (NK) cells are able to recognize infected cells in an antigen-independent manner and destroy them by their cytotoxic activity. Also, they rapidly produce large amounts of IFN- γ to activate the adaptive immune system. NK cells are regulated by a fine balance between stimulatory and inhibitory receptors. One of their prominent features is their ability to destroy cells which lack MHC I molecules on their surface. As many viruses downregulate MHC expression in order to avoid an adaptive immune response, NK surveillance represents an important early warning and attack system against virus infections.

A key connection between the innate and the adaptive immune system is provided by dendritic cells (DCs). These specialized immune cells sample antigen at the site of infection, activate themselves and the surrounding tissue cells by cytokine synthesis, and then migrate to secondary lymphatic organs in order to mobilize T cells against the presented antigen. The differentiation of DCs into efficient antigen-presenting cells (APCs) is achieved by cytokine production which, in turn, is triggered by stimulation of receptors recognizing pathogen-specific molecular patterns (PAMPs). Two main types of DCs are being distinguished: myeloid DCs (mDCs) and plasmacytoid DCs (pDCs). mDCs are an early split-off of the myeloid bone marrow precursors, that is, the stem cells which are also at the root of macrophage/monocyte and granulocyte differentiation, among others. Depending on the location, several subsets of mDCs such as Langerhans cells (residing in epidermis and epithelia) or interstitial cells are being distinguished. pDCs, which are not segregated into subpopulations, are derived from lymphatic precursor cells, that also generate the B and T cells, for example. Both mDCs and pDCs can sense viral infection by several intra- and extracellular PAMP receptors (see below). Depending on the DC type, high levels of interleukins or IFNs are being produced which coin the subsequent immune reaction. pDCs are potent producers of the antiviral type I IFNs.

Antiviral Cytokines: The Type I Interferons

Isaacs and Lindenmann discovered in 1957 that cells which had been in contact with inactivated virus particles secrete a soluble factor which confers cellular resistance to influenza viruses, a phenomenon called “interference”. In the subsequent years, it became more and more clear that the so-called type I IFN (encompassing IFN- β and a set of IFN- α subtypes) system is our primary defense mechanism against viral infections. In fact, humans with genetic defects in the IFN signaling pathway have a bad prognosis as they die at an early age of viral diseases which would otherwise pose little problems. Similarly, knockout mice with a defective type I IFN system quickly succumb to viral pathogens of all sorts although they have an intact adaptive immune system.

In response to virus infection, pDCs are particularly well equipped to synthesize and secrete IFN- α/β , but in principle all nucleated cells are able to do so. In an autocrine and paracrine manner, IFNs trigger a signaling chain leading to the expression of genes for potent antiviral proteins which limit further viral spread. In addition, IFNs initiate, modulate, and enhance the adaptive immune response. The signaling events which culminate in the direct IFN-dependent restriction of virus growth can be divided into three steps, namely (1) transcriptional induction of IFN synthesis, (2) IFN signaling, and (3) antiviral mechanisms.

New Kids in the Gut: The Type III Interferons

Type III IFNs (IFN- λ s), a relatively recent discovery, are distinct antiviral cytokines with many features in common with type I IFNs. IFN- $\lambda 1$, $\lambda 2$, and $\lambda 3$ are induced by virus PAMPs and normally signal through the JAK/STAT cascade, but use a separate receptor. They are able to activate expression of interferon-stimulated genes (ISGs) and have been shown to inhibit replication of viruses. However, whereas all nucleated cells express the receptor for type I IFNs, the type III IFN receptor is limited to epithelial cells on mucosal barriers. IFN- λ s are therefore important for limiting the infection and transmission by respiratory and gastrointestinal viruses. Moreover, while type I IFN signaling activates ISG expression in a fast, strong and transient manner, type III IFN signaling is characterized by a weaker,

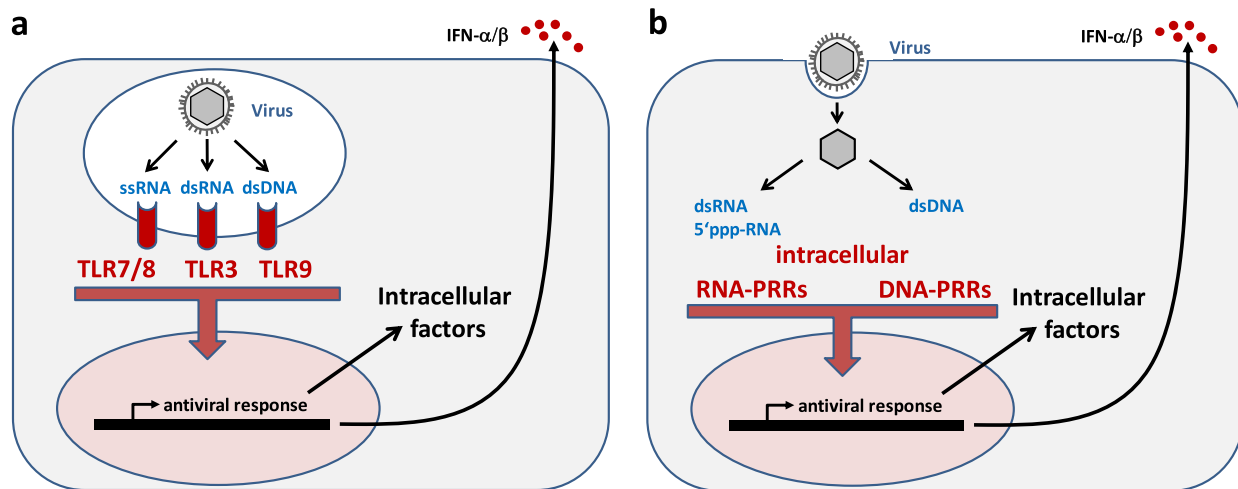


Fig. 1 Depending on the virus, ssRNA, dsRNA, 5'ppp-RNA, dsDNA, or combinations thereof represent characteristic by-products of infection (i.e., PAMPs) which are recognized by PRRs to induce production of antiviral IFN- α/β and intracellular factors with antiviral or regulatory function. The viral PAMP signature molecules are recognized in the endosome by TLRs (a), and after entry into the cytoplasm by intracellular PRRs (b).

delayed, but longer-lasting ISG response. Besides the direct antiviral effect mediated by ISGs, IFN- λ s are stimulating adaptive immune responses. Interestingly, type III IFNs are also responsible for the elevated immunogenicity of live attenuated vaccines as compared to inactivated vaccines, since only the former are triggering IFN- λ production. Thus, IFN- λ s are complementing the IFN response by conferring a lasting, local protection at anatomical sites which are most exposed to viral intruders.

Interferon Induction

Nucleic acids are the main PAMPs of viruses, being recognized by a number of pattern recognition receptors (PRRs) to initiate induction of IFN genes (see Fig. 1). Classes of virus-triggered PRRs can be divided into the endosomal toll-like receptors (TLRs) and various intracellular (mostly cytoplasmic) receptors. It is thought that the TLRs can be activated by viral nucleic acids that had been released from virus particles (or endocytosed remnants of infected cells) that were dissolved by the endosomal low pH and/or degradative enzymes. Major PAMPs are double-stranded RNA (dsRNA), single-stranded RNA (ssRNA), 5'-triphosphorylated (ppp) RNAs, and double-stranded DNA (dsDNA). dsRNA and 5'ppp-RNA are underrepresented in uninfected cells, thus predestining them as immunorelevant markers of non-self. In the case of ssRNA and dsDNA, it is thought that unusual locations (e.g., endosome or cytoplasm for DNA), a lack of cell-typical modifications (e.g., methylation), or specific secondary structures are responsible for triggering antiviral responses. Moreover it should be noted that a full-blown infection converts the heterogeneous population of cellular nucleic acids into a pool of largely homogeneous, highly abundant viral RNAs and/or DNAs. Thus, immunogenicity could also be caused by over-representation of uniform RNA or DNA species.

dsRNA is an almost ubiquitous by-product of virus infection that is sensed by a multitude of PRRs. In the endosome it is recognized by the TLR3, and in the cytoplasm by the RNA helicases RIG-I and MDA-5 (collectively termed RIG-I-like receptors, RLRs) and the protein kinase PKR. Substantial amounts of dsRNA are produced by viruses with a positive-strand ssRNA or with a dsRNA genome (e.g., coronaviruses or reoviruses, respectively) during genome transcription and replication. For DNA viruses (e.g., Herpes viruses or poxviruses), it is thought that dsRNAs are formed either by hybridization of read-through transcripts from promoters on opposite DNA strands, or from secondary structures present on particular viral RNAs. Viruses with a negative-strand ssRNA genome (e.g., influenza virus) are unique in that they do not produce substantial amounts of long dsRNA, since their genome and antigenome RNAs are always encapsidated by viral nucleoprotein. In the cytoplasm, their genomes are recognized by RIG-I in a 5'-triphosphate-dependent manner. In addition a short dsRNA region, formed by the annealing of complementary 5' and 3' ends of the RNA genome (the so-called "panhandle"), is essential for RIG-I to be activated.

Viral ssRNAs can be recognized in the endosome by TLR7 and -8.

The third important PAMP, viral dsDNA, is again recognized both by an endosomal receptor, TLR9, and a series of intracellular receptors such as e.g., IFI16, DDX41, RNA polymerase III, and cGAS. dsDNA recognition by RNA polymerase III actually represents a crosstalk between the RNA-PRRs and the DNA-PRRs, since the polymerase transcribes viral DNA into 5'ppp-dsRNA which then activates RIG-I. A similar second messenger principle is realized by cGAS (cGAMP synthase) which produces cyclic di-GMP-AMP (cGAMP) molecules in response to cytoplasmic dsDNA. cGAMP, in turn, activates the adapter protein STING (stimulator of interferon genes) and hence downstream antiviral signaling.

Besides nucleic acids, some viral proteins can provoke a TLR response such as the envelope proteins of respiratory syncytial virus and measles virus by activating TLR4 and TLR2, respectively.

All PRRs are triggering signaling chains (with partial crosstalk and usage of common adapters and kinases) which canonically culminate in activation the IFN regulatory factor (IRF) -3, the general immune-regulatory transcription factor NF- κ B, and the stress-activated transcription factor AP-1. In cooperation, they upregulate IFN gene expression. This leads to a “first wave” of IFN production (IFN- β and IFN- α 4 in mice) which triggers the expression of the transcription factor IRF-7. IRF-7 is a master regulator of IFN gene expression cooperating with IRF-3 for full activity. IRF-7 can be activated in the same way as IRF-3 and is responsible for a positive-feedback loop that initiates the synthesis of several IFN- α subtypes as the ‘second-wave’ IFNs.

While all cells with a nucleus are thought to be equipped with the set of intracellular PRRs, expression of TLRs is more restricted to epithelial and immune cells. mDCs, for example, can sense dsRNA by the classic intracellular RLR pathway and, in addition, by TLR3. pDCs sense the presence of viral ssRNA or dsDNA by TLR7, TLR8, and TLR9 to transcriptionally activate multiple IFN- α genes. This broad and strong IFN induction is due the presence of constitutively expressed IRF-7 in pDCs, which enables them to bypass the dependency on IRF-3 and to directly launch a “second-wave”-like IFN response. IRF-7 is further upregulated in response to IFN and generates a positive-feedback loop for high IFN- α and IFN- β production. Furthermore, TLR7 and TLR9 are retained in the endosomes of pDCs to allow prolonged IFN induction signaling.

Type I IFN Signaling

IFN- β and the multiple IFN- α subspecies activate a common type I IFN receptor (IFNAR) which signals to the nucleus through the so-called JAK-STAT pathway (Fig. 2). The signal transducer and activator of transcription (STAT) proteins are latent cytoplasmic transcription factors which become phosphorylated by the Janus kinases JAK1 and TYK2. Phosphorylated STAT1 and STAT2 recruit a third factor, IRF9, to form a complex known as IFN-stimulated gene factor 3 (ISGF3) which translocates to the nucleus and binds to the IFN-stimulated response element (ISRE) in the promoter region of ISGs.

The collective term “ISGs” implies a common, uniform mode of regulation. In fact, however, there are three different classes of ISGs, first of all those classical ones responding to IFN signaling (STAT dependent), those responding either to IFN signaling or to PRR activation by PAMPs (universal ISGs), and those that exclusively respond to PRR activation (IRF dependent). The latter ones are not ISGs in a proper sense (but often called so), since they are stimulated by infection rather than by IFN. Moreover, for some ISGs regulation can differ between animal species or between cell types, and by far not all ISGs are entirely characterized with respect to the ISG class they belong to.

Direct Antiviral Effects of Type I IFNs

Type I IFNs activate the expression of several hundred STAT dependent ISGs of which only a fraction has been studied in great detail. Well characterized examples with broad antiviral activity are the 2’–5’ OAS/RNaseL system, protein kinase R (PKR), RNA-specific adenosine deaminase 1 (ADAR1), ISG20, IFN-induced tetratricopeptide repeat (IFIT) protein 1, the Mx proteins, viperin, and tetherin. 2’–5’ OAS and PKR are constitutively expressed in normal cells in a latent, inactive form. Basal mRNA levels are upregulated by IFN- α/β and these enzymes need to be activated by viral dsRNA. The 2’–5’ OAS catalyzes the synthesis of short 2’–5’ oligoadenylates that activate

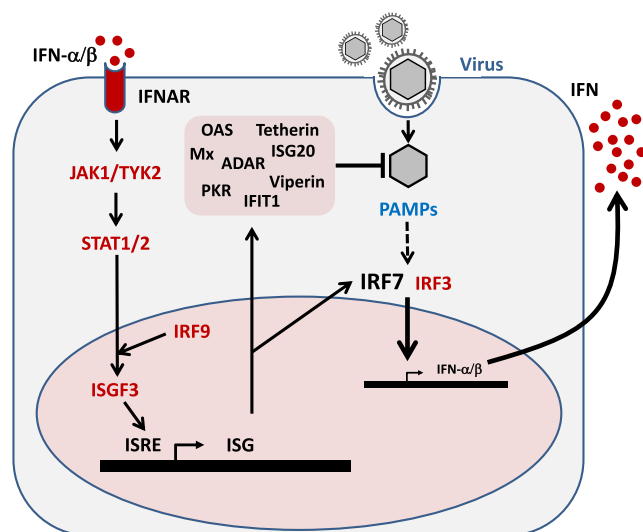


Fig. 2 IFN- α and IFN- β bind to the type I IFN receptor (IFNAR) and activate the expression of numerous ISGs via the JAK/STAT pathway. OAS, ISG20, Mx, ADAR, PKR, Tetherin, Viperin and IFIT1 are examples of proteins with antiviral activity. IRF-7 amplifies the IFN response by upregulating PAMP-dependent expression of several IFN subtypes.

the endoribonuclease RNaseL. RNaseL, an ISG in most species (humans are an exception expressing it constitutively), degrades viral and cellular RNAs. PKR is a serine-threonine kinase that phosphorylates – among other substrates – the α -subunit of the eukaryotic translation initiation factor eIF2 to block translation of cellular and viral mRNAs. PKR also plays a role in virus-induced NF- κ B activation, as described above. ADAR1 catalyzes on viral dsRNAs the deamination of adenosine to inosine. As a result the secondary structure is destabilized due to a change from an AU base pair to the less stable IU base pair, and the accumulating mutations are hampering viral replication. Recent research showed that ADAR1 is important to suppress the formation of dsRNA structures by endogenous Alu retroelements. In ADAR1-deficient animals, endogenous RNAs are not edited, and the arising dsRNAs activate the PRRs MDA5 and PKR, leading to chronic IFN induction, inhibition of mRNA translation, and eventually autoimmunity. ISG20 is on one hand an exonuclease that degrades viral ssRNA, but in the other hand it was recently found to downmodulate translation of non-self mRNAs in an RNase-independent manner. IFIT1 is expressed at extremely high levels after IFN stimulation or PRR signaling and sequesters viral RNA with a 5'ppp end or an unmethylated cap. Mx proteins are enwrapping viral nucleocapsids, thus preventing the viral polymerase from elongation of transcription. Viperin (Radical S-adenosyl methionine domain containing 2, RSAD2) is an enzyme that converts the nucleotide CTP to a chain terminator for viral RNA-dependent RNA polymerases. Tetherin (bone marrow stromal cell antigen 2; BST2) is a transmembrane protein able to restrict the release of enveloped virus particles from the plasma membrane.

The antiviral profiles of the IFN effectors listed above are distinct but often overlapping. Mx proteins, for example, mainly inhibit segmented negative-strand RNA viruses and also Semliki Forest virus, whereas the 2'-5' OAS/RNaseL system appears more important against positive-strand RNA viruses. Moreover, only rarely the presence of one particular IFN effector explains host resistance. Rather, it is the sum of antiviral factors affecting, for example, genome stability, genetic integrity, transcription, and translation that confers the full antiviral power of IFN.

Indirect Antiviral Effects of Type I IFNs

Besides the effector proteins listed above, several ISGs contribute in a more indirect manner to the enhancement of both innate and adaptive immune responses. Virus-sensing (and in part antiviral) PRRs such as TLR3, PKR, RIG-I, and MDA5 are by themselves upregulated in an IFN-dependent manner. Similarly, IRF-7 and STAT1, the key factors of type I IFN and ISG transcription, respectively, are ISGs. The strong positive-feedback loop mediated by the upregulation of these PRRs and transcription factors is counterbalanced by several negative regulators (e.g., LGP2, SOCS, PIAS), which are either ISGs or depend on IFN signaling for their suppressive action.

Type I IFNs can directly enhance clonal expansion and memory formation of CD8⁺ T cells. Also, IFNs promote NK cell-mediated cytotoxicity and trigger the synthesis of other cytokines such as IFN- γ or IL-15 which modulate the adaptive immune response, enhance NK cell proliferation, and support CD8⁺ T-cell memory. Moreover, by upregulating TLRs, MHCs, and costimulatory molecules, IFNs enable APCs (most prominently DCs) to become competent in presenting viral antigens and stimulating the adaptive immune response.

Innate Immunity Memory

Since long, the lack of any type of memory is used as a feature distinguishing innate from adaptive immunity. It became meanwhile clear, however, that IFNs can reprogram cell responses by leaving epigenomic signatures on promoters, mostly by post-translational modifications of histones. This memory-like mechanism can last for weeks, and explains the long-known phenomenon of “priming”, i.e., that a first small shot of IFNs is potentiating the effect of a second IFN treatment. A phenomenon related to priming is “innate immunity training”. Here, prior inflammation or infection enhances a second innate response and inflammation, again via changing epigenomic signatures.

Tonic IFN Levels

Although the classical model outlined above depicts that PRR-triggered upregulation of IFN- β starts from level zero, at least at barrier tissues basal levels of IFN expression are maintained constantly. This “tonic” IFN expression is thought to be a reaction to commensal bacteria, and primes the organism for rapidly reacting to pathogenic invaders.

Viral Counterstrategies

Given the massive direct and indirect antiviral effects of type I IFNs, it comes as no surprise that viruses had evolved efficient countermeasures. In fact, most viruses investigated so far were found to actively inhibit either IFN induction, IFN signaling, antiviral ISG action, or combinations thereof. A common strategy to avoid IFN induction seems to be the targeting of treacherous dsRNA or 5'ppp structures by binding, modifying, or degradation through viral factors, the so-called IFN antagonists. Moreover, many viral features such as encapsidating genomic RNA by nucleoprotein (negative-strand RNA viruses), hiding replication

complexes in intracellular membrane compartments (positive-strand RNA viruses) or multiple protein sheets (dsRNA viruses), as well as replicating in the nucleus (orthomyxoviruses) can be regarded as passive strategies to minimize generation and exposure of PAMPs to PRRs. Active strategies include sequestration or degradation of key factors of antiviral signaling like PRRs, kinases, IRFs, STATs or even RNA polymerase II itself. The bundle of these measures, i.e., the individual anti-IFN profile of a particular virus, can represent a major marker for host range, cell tropism, and virulence. Weak anti-IFN capabilities can render a virus unfit in a given host, whereas strong IFN suppression enables productive replication. Fine-tuning the IFN antagonistic activities may allow the adaption to optimal host-to-host transmission. Depending on the particular transmission mode and host population density, either massive, damaging viremia (like e.g., for arthropod- or aerosol-transmitted pathogens) or lower-level, locally restricted or chronic infection (e.g., herpes, papilloma or hepatitis C viruses) are positively selected for. In the case of the persistent infections, long-term production of IFN can even be beneficial for the virus as it results in an immunosuppression aimed at limiting immunopathology. Thus, the balance and timing of activation versus suppression of the IFN response can determine the outcome of infection, ranging from a straight fending off (host wins by early antiviral IFN action) to acute disease (virus wins by suppressing early IFN action) to chronic viral disease (constant IFN levels maintain infection by suppressing immune responses). In any case, the simple equation: strong IFN antagonism = virulent virus, weak IFN antagonism = harmless virus, should be applied with some caution.

Good Cop–Bad Cop

Given their massive impact on the cellular gene expression profile, it is quite expected that type I IFNs have not only antiviral, but also antiproliferative and immunomodulatory effects. Treatment with IFNs is an established therapy against several viral and malignant diseases such as hepatitis B, hepatitis C, Kaposi's sarcoma, papillomas, multiple sclerosis, and several leukemias and myelomas. However, the strong and systemic effects of IFNs do not come without a price. Administration of IFN can locally produce inflammation, and systemically cause fever, fatigue, malaise, myalgia, and anemia. It is no coincidence that these latter are "flu-like" symptoms, since in many acute infections IFNs play a dominant role. The effects of IFN which are desired and beneficial if restricted to the site of first infection can turn into a life-threatening "cytokine storm" if it becomes systemic. Severe acute respiratory syndrome (SARS) and human infections with H5N1 influenza viruses are examples of such out-of-control innate immune responses. Moreover, long-term and high IFN levels are facilitating persistent infections, as discussed above, and IFN therapy can aggravate autoimmune disorders. Another "dark side" aspect is the so-called interferonopathies. Patients with mutated PRRs or IRFs exhibit type IFN gene expression signatures and are prone to autoimmune diseases. Chronic production of IFNs causes maturation of mDCs, which in turn activate autoreactive T and B cells.

Concluding Remarks

The concept of innate immunity certainly comprises more than the IFN system (see above), but type I and type III IFNs represent a central part. These cytokines not only have direct antiviral effects but also orchestrate the first defense reactions and the subsequent adaptive immune response, thus determining the course of infection. The fact that basically every virus appears to have evolved one or several countermeasures for controlling the IFN response is testament to its importance. In addition, IFNs are not only antiviral, but also effective tumor suppressors. Tumor cells often eliminate the IFN system during the transformation process. The payoff is an increased susceptibility to infection, an Achilles heel which is exploited by the therapeutic concept of oncolytic viruses. Tumor selectivity of such viruses can be even more increased by using IFN-sensitive mutants. The inability of those mutants to fight the IFN response is complemented by the mutations of the tumor cells, thus allowing virus growth. At the same time, these viruses are unable to infect the IFN-competent healthy cells.

Cells had to cope with viruses since the early days in the primordial pond. No wonder innate immune responses are so astonishingly multi-faceted, consisting of a wide array of cells, signaling chains and effector molecules solely dedicated to the elimination of infectious intruders. Viruses are the most abundant biological entities on earth, but most accidental contacts are not even noticed by us. Only those viruses which had evolved tailor-made counterstrategies can break through and establish infection for long enough to be multiplied and transmitted further. The innate immune system may be old, but as long as there are viruses (and tumors), it will never come out of fashion.

Further Reading

- Ablasser, A., Hur, S., 2019. Regulation of cGAS- and RLR-mediated immunity to nucleic acids. *Nature Immunology* 21, 17–29.
- Barrat, F.J., Crow, M.K., Ivashkiv, L.B., 2019. Interferon target-gene expression and epigenomic signatures in health and disease. *Nature Immunology* 20, 1574–1583.
- Holers, V.M., 2014. Complement and its receptors: New insights into human disease. *Annual Review of Immunology* 32, 433–459.
- Reizis, B., 2019. Plasmacytoid dendritic cells: Development, regulation, and function. *Immunity* 50, 37–50.
- Schmid, S., Mordstein, M., Kochs, G., García-Sastre, A., Tenover, B.R., 2010. Transcription factor redundancy ensures induction of the antiviral state. *Journal of Biological Chemistry* 285, 42013–42022.
- Schoggins, J.W., 2019. Interferon-stimulated genes: What do they all do? *Annual Review of Virology* 6, 567–584.

- Versteeg, G.A., García-Sastre, A., 2010. Viral tricks to grid-lock the type I interferon system. *Current Opinion in Microbiology* 13, 508–516.
- Wilson, S.S., Wiens, M.E., Holly, M.K., Smith, J.G., 2016. Defensins at the mucosal surface: Latest insights into defensin-virus interactions. *Journal of Virology* 90, 5216–5521.
- Ye, L., Schnepf, D., Staeheli, P., 2019. Interferon- λ orchestrates innate and adaptive mucosal immune responses. *Nature Reviews Immunology* 19, 614–625.

Humoral and T Cell-Mediated Immunity to Viruses

Ane Ogbe and Lucy Dorrell, University of Oxford, Oxford, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Role of the Innate Immune System in the Initiation of the Adaptive Immune Response

Introduction to Barrier Immunity

On a daily basis, we encounter numerous micro-organisms. We are, however, protected from the majority of them by physical and chemical barriers, consisting of epithelial cells of the skin and the mucus layer covering internal mucosal surfaces, which together provide the first line of defense. These surfaces host a microbial ecosystem or microbiome comprising largely bacteria and fungi, which collectively outnumber the cells in the human body and provide a hostile environment for infectious pathogens. The skin and mucous membranes serve not only as a passive physical barrier but also as a site of active immune surveillance, which is provided by diverse immune effector cell populations. Their function is to send the first warning signals upon detection of potential pathogens. Viruses, along with other microbes, gain entry to the body either when the skin barrier is broken or via contact with exposed mucous membranes, such as those of the upper respiratory and gastrointestinal tracts. Because viruses can only replicate by co-opting host nucleic acid and protein synthetic machinery, they have evolved specific mechanisms to attach to and gain entry into their target cells.

Innate Immune Effectors are Poised to Respond to Viruses That Overcome Barrier Protection

The first immune responses to viral infections are mediated by effector cells of the innate immune system. In contrast to adaptive immune responses, they are initiated within hours, are non-specific and are coordinated via signaling pathways that are common to many cell types. The effector populations are mononuclear (macrophages) and polymorphonuclear (neutrophils) phagocytes, innate lymphoid cells, natural killer (NK) cells and dendritic cells. As discussion of each of these in depth is beyond the scope of this article, we highlight the innate cell populations that play a key role in the induction of humoral and cellular immune responses to viruses. Innate immune responses may be associated with local and systemic inflammation, a tightly regulated, physiological process mediated by chemokines and cytokines that leads to increased blood flow, enhanced permeability of the vasculature and recruitment of phagocytic cells to the site of infection. Phagocytosis is an evolutionarily conserved physiological process in which invading pathogens are recognized via cell surface receptors, internalized into a phagosome and subsequently destroyed. While thought to be primarily a defense mechanism against extracellular pathogens, macrophages and neutrophils are capable of eliminating virus-infected cells by phagocytosis. Along with other invading pathogens, they may be tagged with molecules known as opsonins, which serve to enhance the efficiency of the phagocytic process.

Natural killer (NK) cells are a subset of innate lymphoid cells that are typified by the expression of CD56 and absence of CD3. They can be further segregated based on FCγRIIIA (CD16) expression. Their antiviral functions can be broadly categorized as: (1) cytotoxicity, mediated by the release of cytolytic granules, perforin and granzyme, or in some circumstances, the engagement of death receptors on target cells and (2) the secretion of pro-inflammatory cytokines such as interferon- γ (IFN- γ) and tumor necrosis factor- α (TNF- α). A unique antiviral effector function of NK cells is antibody-dependent cellular cytotoxicity (ADCC), mediated by CD16 by recognition of antibody-decorated ('opsonized') virally infected cells. The secretion of cytolytic granules and cytokines upon activation is immediately triggered by encounter with virus-infected cells and does not depend on prior exposure to antigen. Because of this, NK cell cytotoxicity is tightly regulated by the balance of inhibitory and activating signals that are transduced by killer immunoglobulin (Ig) receptors (KIRs) when they detect loss of major histocompatibility complex (MHC) molecules on the infected cell surface. NK cells also express several other classes of receptor, including C-type lectin receptors that detect stressed cells via MHC homologs (MICA/B). While NK cell responses to viruses are generally non-specific, the concept of NK cell memory, or trained immunity, has been proposed in the light of evidence of a long-lived NK cell subset that shows a rapid response to re-encounter with viruses, which is a feature of adaptive immune responses.

Barrier immunity, early onset inflammatory responses, microbial phagocytosis, and NK cell effector responses may be sufficient to eliminate viral infections. However, viruses have evolved various strategies to evade the innate immune system. Under these circumstances, elimination of the infection requires engagement of the adaptive immune system to provide a highly specific and potent response that can be recalled if necessary, i.e. upon re-exposure to the same or similar pathogens. Dendritic cells (DC) play a central role in priming adaptive immune responses by B and T cells. Originating in the bone marrow, after release into the circulation they migrate to tissues where they continually sample the extracellular environment, particularly at sites of pathogen entry. They capture pathogens by endocytosis, either in their native form, or by engulfing infected cells or apoptotic cell debris. Antigen capture in tandem with danger signals from pathogens causes DC maturation, manifested by upregulation of co-stimulatory molecules and migration to lymphoid tissue, where they encounter and activate naïve T cells. DC are further categorized as myeloid (mDC) and plasmacytoid (pDC). mDC are the most potent antigen-presenting cells (APC) while pDC secrete inflammatory cytokines and type I interferons when activated by detection of viral nucleic acids.

Innate Immune Recognition of Viruses

Immune sensing of viral components is mediated by germline-encoded pattern recognition receptors (PRRs) that recognize broadly conserved pathogen-associated molecular patterns (PAMPs) such as viral DNA, single stranded RNA, 5'-triphosphate-containing RNA, double-stranded RNA and viral proteins. A vast array of PRRs has been described; these can be classified by their cellular location. Cell surface receptors that recognize viral components include C-type lectin receptors and some Toll-like receptors (TLRs), namely TLR 2 and 4. Intracellular receptors include endosomal TLRs (TLR 3, 7, and 9) and cytoplasmic NOD-like receptors (NLRs), RIG-I-like receptors (RLRs) and double-strand DNA sensors (interferon-inducible protein-16 and cyclic GMP-AMP synthase) (Table 1). Ligation of viral components by PRRs activates signaling cascades, resulting in the production of type I and type III interferons (IFNs) and other proinflammatory cytokines and chemokines, which act as a chemoattractant for phagocytic cells. In addition, the expression of co-stimulatory molecules (e.g., CD40 and the B7 family, CD80 and CD86) on these cells is increased. Secreted IFNs bind to receptors on infected and uninfected cells, inducing the expression of numerous IFN-stimulated genes. Type I interferons (IFN- α and IFN- β) encode proteins that induce an antiviral state, for example, enzymes that edit viral RNA and DNA (e.g. adenosine and cytidine deaminases) or directly antagonize viral proteins involved in replication (e.g., The tripartite motif-containing protein (TRIM) family). Type III interferon (IFN- λ) also induces an antiviral state, although this may be more limited due to the restricted expression of IFN- λ receptors.

Crosstalk Between Innate and Adaptive Immune Cells

The process of antigen presentation bridges the innate and adaptive immune systems and is a crucial step in the initiation of the adaptive immune response. Professional APCs, namely mature DC, macrophages and B cells, display processed intracellular protein antigens on the cell surface in the form of peptides bound to of MHC molecules (human leukocyte antigen, HLA, in humans). MHC/peptide complexes are recognized by T cell receptors (TCRs) expressed on CD4+ and CD8+ T cells. MHC molecules are encoded by MHC genes on chromosome 6 and are divided into two major classes, I and II. Structural differences between MHC class I and class II molecules underlie differential preferences in the lengths of peptides that are accommodated in the peptide binding groves of these molecules. MHC class I molecules bind peptides that are 8–11 amino acid (aa) long while MHC class II bind longer peptides between 13 and 25 aa length. MHC class I is expressed on the surface of all nucleated cells while MHC class II expression is restricted to professional APCs. Viral antigens access MHC class I and II molecules via segregated pathways: phagocytosed viruses or viral products enter an exogenous pathway in which they are degraded in endosomes before transport into the endoplasmic reticulum (ER) and loading onto MHC class II molecules; proteins that are generated in the cytosol in infected cells as a result of viral gene transcription enter an endogenous pathway in which they are degraded in the proteasome and transported by cellular chaperones to the ER, where they are loaded onto MHC class I molecules. Peptide-loaded MHC complexes (pMHC) are transported via Golgi apparatus to the surface of the cells for cognate recognition by TCRs.

Exceptions to these rules are: (1) the phenomenon of cross-presentation in professional APCs, whereby phagocytosed virions or viral proteins escape from endosomal compartments into the cytosol and enter the endogenous pathway, leading to presentation by MHC class I; (2) access of cytosolic peptides to MHC class II molecules as a result of autophagy, the process of recycling of organelles and proteins in lysosomes as observed in the presentation of Epstein-Barr virus nuclear antigen (EBNA1) and Influenza matrix protein (MP1) to CD4 T cells.

The importance of antigen presentation in the development of the adaptive immune response to viruses is evident in the primary immunodeficiency syndrome, Bare lymphocyte syndromes I and II, an autosomal recessive genetic disorder where gene defects lead to a lack of expression of MHC class I and II respectively and carriers are known to suffer from multiple persistent viral infections. Furthermore, downregulation of MHC presentation is one of the most well-characterized mechanisms used by viruses to avoid T cell recognition. This is achieved by diverse mechanisms operating at different stages of the processing and presentation pathways described. For example, herpes simplex virus type 1 and cytomegalovirus encode proteins (ICP47, US6 and US3) that interfere with peptide loading in the ER, while human immunodeficiency virus type 1 (HIV-1) Nef diverts MHC class I molecules to the trans-Golgi network and Epstein-Barr virus (EBV) BLZF1 gene represses the transcription of MHC class II transcriptional activator (CIITA).

The Adaptive Immune Response to Viral Infections

Overview

The adaptive or acquired immune response to all pathogens, including viruses, provides protection from infection upon re-exposure to the same or similar micro-organisms. It is mediated by T and B lymphocytes that express antigen-specific cell surface receptors. Key elements of the adaptive immune response are: antigen specificity that results from the selection of TCRs and B cell receptors (BCRs) from a vast repertoire of receptor variants; receptors are generated by gene rearrangements and are selected on the basis of self/non-self-discrimination; responses are honed by a process of maturation in specialized lymphoid organs; after resolution of the infection, a proportion of T cell and B cell clones survive and are able to re-expand rapidly upon re-encounter with cognate antigen, providing immunological memory. In contrast to the innate immune response, the adaptive immune

Table 1 Pattern recognition receptors and their activating ligands

<i>PRR family</i>	<i>Cellular location</i>	<i>PRR</i>	<i>Ligand</i>	<i>Adapter proteins</i>	<i>Viral targets</i>
<i>RLR</i>	Intracellular (cytoplasm)	RIG-I	short dsRNA < 1 kb and uncapped RNA 5'-Triphosphate	MAVS	Sendai Virus, IAV, IBV, Nipah virus, Measles virus, RSV, VSV, Rabies virus, Ebola virus, Lassa virus, LCMV, HCV, West Nile virus, Dengue Virus
		MDA5	Long dsRNA > 2 kb		Rabies virus, West Nile virus, Sendai virus, Dengue virus, Murine hepatitis virus
		LGP2	dsRNA	No intrinsic signaling described	VSV, Encephalomyocarditis virus
<i>TLR</i>	Intracellular (endosomal vesicles)	TLR3, TLR7 ^a TLR8 and TLR9 ^a	TLR 3 – dsRNA TLR7 and TLR9 – ssRNA and DNA with CpG motif	TLR 3 – TRIF TLR 7/8/9 – MyD88	TLR3 – RSV, IAV, West Nile virus, Poliovirus; TLR7 – IAV, VSV, HIV, Dengue virus, Sendai virus, MMTV; TLR8 – HIV
	Transmembrane	TLR2 and TLR4	Bacteria protein, Lipoprotein and LPS	MyD88	TLR3 – RSV, IAV, West Nile virus, Poliovirus; TLR7 – IAV, VSV, HIV, Dengue virus, Sendai, MMTV; TLR8 – HIV
<i>NLRs</i>	Intracellular (cytoplasm)	NLRP3, NLRC2, NLRC5	NLRP3 and NLRC5 – Indirect sensing of viral invasions through stress signals induced by invasion of viruses, endogenous DAMPs, and ROS dsDNA, ssRNA NLRC2 – direct sensing of ssRNA from viruses	ASC	NLPR3 and NLRC5 – NLRC2 – RSV, IAV, Parainfluenza
<i>cytosolic DNA sensors</i>	Intracellular (cytoplasm)	cGAS	dsDNA > 45 bp	STING	CMV, HPV, HSV1 and 2, Adenoviruses, VACV
	Intracellular (cytoplasm and nuclear)	AIM2	dsDNA	ASC	VACV, MCMV
		IFI16	ssDNA	ASC	HSV-1, CMV

^aExpressed on pDC s and B cells.

Abbreviations: PRR, pattern recognition receptor; DAMPs, death associated molecular patterns; ROS, reactive oxygen species; LGP2, laboratory of genetics and physiology 2; LPS, Lipopolysaccharide; ds, double stranded; ss, single stranded; kb, kilobase; MDA5, Melanoma differentiation associated gene 5; cGAS, cyclic GMP–AMP synthase; STING, signaling effector stimulator of interferon genes; AIM2, absent in melanoma 2; ASC, apoptosis-associated speck-like protein containing a CARD; IFI16, IFN-inducible protein 16; NLR, Nucleotide-binding oligomerization domain-containing (NOD)-like receptors; TLR, Toll-like receptors; RLR, Retinoic acid-inducible genes I (RIG-I)-like receptors.

response evolves over time and therefore does not provide immediate protection against invading pathogens. However, protection is durable and in some cases, lifelong, due to the longevity of a subset of T cell and B cell clones.

T Cell Recognition of Viral Antigens

Lymphoid progenitor cells (thymocytes) develop from hematopoietic stem cells in the bone marrow and migrate to the thymus where they undergo sequential selection steps to ensure that only functional non-autoreactive T cells enter the circulation. Here, thymocytes first begin to express TCRs as a result of random recombination of TCR V(D)J gene segments. TCRs are heterodimeric membrane-bound proteins, comprising a pair of disulfide-linked polypeptide chains. The vast majority of TCRs comprise paired alpha and beta chains ($\alpha\beta$ TCR) while a minority consists of paired gamma and delta chains ($\gamma\delta$ TCR). During the process of TCR generation, thymocytes become committed to either CD4 or CD8 T cell lineage through the expression of CD4 or CD8 co-receptors, which confer distinct specificities for MHC molecules. Positive selection ensures that only thymocytes capable of interacting with MHC molecules survive: those that bind strongly to MHC class II become CD4⁺ T cells while those that bind to MHC class I become CD8⁺ T cells. Positively selected thymocytes that show strong reactivity to self antigens are subsequently deleted, or undergo clonal diversion or receptor editing in a process termed central tolerance. Eventually, single-positive CD4 and CD8 T cells exit the thymus as naïve T cells that circulate between blood and peripheral lymphoid organs, including the lymph nodes and tonsils, which they constantly patrol in order to detect their specific antigens.

The α and β chains of the TCR each consist of an extracellular portion that structurally resembles an Ig antigen-binding fragment, a transmembrane domain and a cytoplasmic tail. The extracellular portion of each chain comprises a variable (V) domain, which binds to the peptide-MHC complex, and a membrane-proximal constant (C) domain. The V regions contain 3 segments of highly variable amino acid sequences, or complementarity-determining regions (CDR), that are generated by random gene rearrangements, as described earlier. The third hypervariable region (CDR3) directly contacts the peptide-MHC complex and has the highest level of diversity, due in part to the random addition of non-germline encoded nucleotides. The $\alpha\beta$ TCR forms part of the TCR complex, together with non-covalently associated signal-transducing molecules, CD3 and the zeta chain, which initiate T cell activation upon recognition of antigen. These signals are amplified by the interactions of the co-receptors, CD4 and CD8, which make contact with the β 2 region of MHC class II and the α 3 region of MHC class I respectively.

T Cell Effectors in the Response to Viral Infections

CD4⁺ T Cells

CD4⁺ T cells were originally known as helper T (Th) cells due to recognition of their essential roles in providing help to cytotoxic T cells and B cells, which were identified in experimental models of viral infections. CD4⁺ T cells are activated upon interaction with professional APCs presenting specific peptides bound to MHC class II and receive signals that drive their differentiation into one of several effector populations with distinct functions. The factors that govern this differentiation program include the quality and dose of antigen, the duration of TCR engagement with the peptide-MHC complex, the co-stimulatory signals provided by the APC, and the cytokines present in the environment during T cell activation. Together, these induce the expression of key transcription factors and signaling transducer and activator of transcription proteins (STATs) that drive commitment to specific Th lineages. There are at least 5 major and well-described CD4⁺ T cell subsets: Th1, Th2, Th17, T follicular helper (Tfh), and regulatory T (Treg) cells. Viral antigens favor Th1 and Tfh differentiation, and Tregs play a crucial role in modulating virus-specific effector responses, therefore, for brevity we have focused on these subsets. Th1 cells in particular have been implicated in the response to virtually all viral infections. While CD4⁺ T cells have long been considered to play an indirect role in the adaptive immune response to viruses, namely, orchestrating the recruitment and effector programs of B cells and CD8⁺ T cells, a subset of cytotoxic CD4⁺ T cells may play a direct role by lysing virally infected cells. Recent evidence has revealed an important role for these cytotoxic CD4⁺ T cells in the response to diverse viruses (discussed below). Finally, it should also be noted that there is plasticity in CD4⁺ T cell differentiation and lineage commitment is not necessarily irreversible (Fig. 1).

T Helper 1 (Th1) Cells

Commitment to this lineage is driven by the secretion of IFN γ and interleukin-12 (IL-12) and the expression of the transcription factor T-bet in activated CD4⁺ T cells. A defining feature of Th1 effector response is the secretion of cytokines IFN- γ , IL-2 and tumor necrosis factor- α (TNF- α). These cytokines are crucial in the CD4⁺ T helper-mediated activation of cytotoxic CD8⁺ T cells, as was demonstrated in early experimental mouse models using lymphocytic choriomeningitis virus (LCMV): infection of CD4-depleted mice resulted in a reduction in cytotoxic CD8⁺ T cell responses. The secretion of IFN- γ and TNF- α was also shown to have direct antiviral effects in the LCMV model and in viral infections such as herpes simplex type 1 (HSV-1) and measles virus (MV).

T Follicular Helper (Tfh) Cells

One of the most well described T helper functions is the provision of help to B cells by Tfh cells, a specialized subset of CD4⁺ T cells, that drive germinal center (GC) formation in lymphoid follicles, antibody isotype (class) switching, the production of

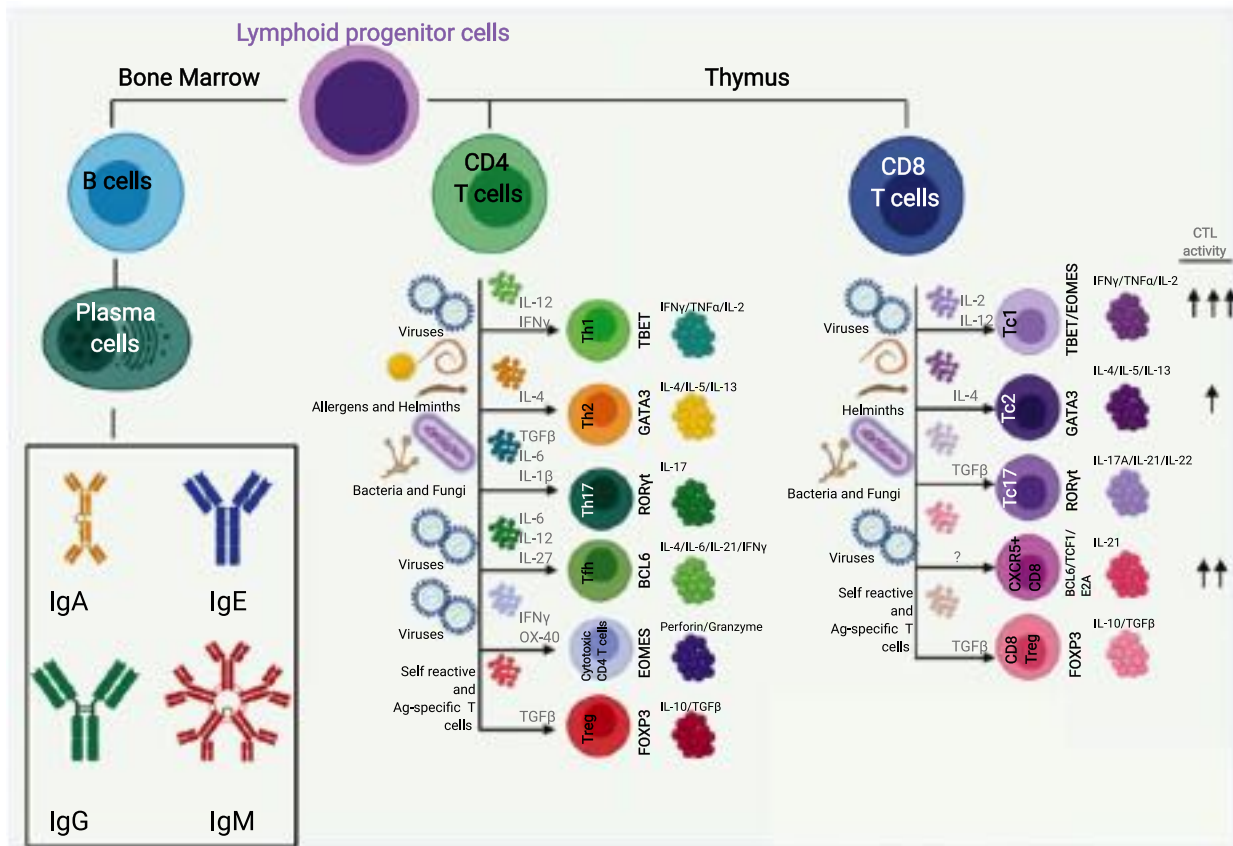


Fig. 1 Subsets of cells within the adaptive immune system. T and B lymphocytes are composed of various cell subsets that differentiate from lineage-naïve cells upon antigen priming. They produce various cytokines and antibodies in response to stimulation. FOXP3, Foxhead box p3; ROR γ t, retinoic acid receptor-related orphan receptor γ t; GATA3, GATA binding protein 3; EOMES, Eomesodermin; TBET, T box protein expressed in T cells; BCL6, B cell lymphoma 6; TCF1, T cell factor 1; Ig, Immunoglobulin; TGF- β , Transforming growth factor β ; TNF- α , Tumor necrosis factor α ; IFN- γ , Interferon- γ ; IL, Interleukin.

durable, potent and high affinity antibodies and the generation of memory B cells in the GCs. Tfh differentiation involves a multi-step program, initiated by IL-6, IL-12, and IL-23, which upregulate the transcription factor B cell lymphoma protein 6 (Bcl-6), and is followed by increased surface expression of the chemokine, CXCR5, together with IL-21 secretion. This leads to CXCL13-dependent migration of Tfh cells to the GC. Here, encounter with cognate antigens on B cells drives their maturation into GC Tfh cells that coordinate the GC reaction.

Even before the formal characterization of this cell subsets in the 2000s, there was an appreciation for CD4 + T cells help in the generation of potent antibody responses. In the murine vesicular stomatitis virus model, CD4 T cells were shown to be essential for the generation of virus-specific neutralizing antibodies. Furthermore, it was observed that chronic LCMV infection could skew the Th differentiation program towards Tfh cell differentiation. Tfh frequency and/or function was also found to correlate with viral control in LCMV and Zika virus infections. In addition, virus-specific Tfh cells have been detected in patients with acute hepatitis C (HCV) infection and their functional activity correlated with the development of antibodies over time.

Regulatory T Cells (Tregs)

Tregs negatively regulate the immune response to prevent tissue damage caused by uncontrolled inflammation. Natural Tregs (nTreg) are generated in the thymus via recognition of self-peptides and upregulate the transcription factor, Foxp3. Induced Tregs arise in the periphery in response to IL-10-mediated induction of transforming growth factor- β (TGF- β), which may accompany the inflammatory cytokine cascade that is provoked by viral infections. Tregs exert suppression of T cells and other effector cells via multiple mechanisms, including the secretion of cytokines such as IL-10 and TGF- β , metabolic disruption leading to Th cell nutrient deprivation or apoptosis and inhibition of DC maturation. An example of their regulatory function is the attenuation of the T cell-mediated inflammatory response to HSV-1 in the corneal stroma of the eye, that can otherwise lead to blindness. However, Treg-mediated regulation of T effector responses can also hinder viral clearance, as has been observed in cervical dysplastic lesions caused by human papillomavirus (HPV).

Cytotoxic CD4 + T Cells

These non-canonical CD4 + T cells recognize antigens in association with MHC class II, yet display cytolytic activity in a similar manner to CD8 + T cells, either by the release of cytotoxic granules or by engagement of death receptors on infected cells. Although their differentiation from naïve CD4 + T cells is poorly defined, they express CD57 and lectin-like receptors that are more commonly found on NK cells, which suggests that they arise in the context of chronic antigen exposure. A protective role for this cell subset has been described in numerous viral infections including HIV-1, influenza A virus (IAV), hepatitis B (HBV), HCV, cytomegalovirus (CMV), and EBV infections. They may compensate for the loss or exhaustion of cytotoxic CD8 + T cells in chronic infection.

CD8 + Cytotoxic T Cells (CTL)

CD8 + T cells are able to recognize and kill virus-infected cells and to secrete cytokines and chemokines that inhibit viral replication and entry into neighboring cells. Their indispensable role in protection from disease and in the clearance of viral infections was first demonstrated in adoptive transfer and CD8 + T cell depletion experiments in mouse models. Similar to CD4 + T helper cells, CD8 + T cells differentiate into diverse cytotoxic T cell subsets depending on the initial signals they receive from APCs, the antigen load and the cytokine environment. APC-derived IL-12 and IFN- γ are important for the induction of cytolytic proteins, perforin and granzymes. They can be defined as Tc1, Tc2, Tc17, and Tregs in an analogous fashion to Th cells, ie according to their cytokine secretion profile and effector functions. The acquisition of cytolytic capacity is under the control of the transcription factors T-bet, eomesodermin. Blimp-1 is also implicated in efficient killing (Fig. 1).

A population of CD8 T cells that express the chemokine CXCR5, follicular cytotoxic T cells (TFC) have recently been described in chronic viral infections including murine LCMV, simian immunodeficiency virus (SIV), HIV-1 and EBV infections. These CXCR5 + CD8 + T cells are known to migrate to follicles and lyse virus-infected cells. They may also provide 'help' to B cells to generate class-switched antibodies.

Mounting an Effective T Cell-Mediated Immune Response to Viral Infections

T Cell Priming

Naïve T cells that encounter APCs displaying a specific viral peptide can be activated to respond to the infection, provided they receive the appropriate signals. The process of T cell priming by DCs, which encompasses activation, clonal expansion and differentiation, takes between 3 and 7 days. According to the 3-signal paradigm, signal 1 is provided by the interaction of the TCR with the peptide-MHC complex; signal 2 requires the binding of co-stimulatory molecules on the surface of T cells to their respective ligands on APCs, while signal 3 is mediated by APC-derived cytokines that are induced by PRR activation. The interaction of CD28 on T cells and B7 molecules on APCs provides co-stimulatory signals that are crucial for T cell survival and induces the production of IL-2, which drives clonal expansion and effector cell differentiation. The absence of co-stimulation leads to a state of anergy, in which the T cell is refractory to further antigenic stimulation.

As stated earlier, DCs and macrophages are professional APCs that upregulate MHC class II as they mature, enabling peptide presentation to CD4 + T cells. In addition, they express MHC class I and are capable of direct presentation of viral peptides to naïve CD8 + T cells if they become infected by viruses. This has been demonstrated in the context of many viral infections, including vaccinia virus (VACV), murine and human CMV, human lymphotropic virus type 1 (HTLV-1), Dengue virus (DV) and HSV. They are also capable of cross-presentation, which may be important to the effectiveness of the response to IAV, HIV-1, HSV-1, and EBV.

Following activation, T cells undergo clonal expansion of several thousand-fold and, depending in part on the third signal they receive, they differentiate into the specialized subsets previously described, giving rise to virus specific effector T cells that are poised to respond to viral infections and to migrate to the site of infection.

T Cell Help

CD4 + T cells provide help to cytotoxic CD8 + T cells in two principal ways: (1) via the secretion of cytokines such as IL-2 and IL-21, that may promote the survival and lytic capacity of CD8 + T cells and (2) the interaction of DCs with CD4 T cells via CD40-CD40 ligand (CD40L) interactions, which license DCs to prime naïve CD8 + T cells. In addition, CD4 + T cells secrete cytokines and chemokines that attract CD8 + T cells to the site of infection and promote their migration towards DCs. Although CD8 + T cells can be activated directly by DCs, CD4 + T cell help is considered to be essential for optimal CD8 + T cell effector function and for durable memory responses.

The help provided by CD4 + Tfh cells to B cells constitutes several functions, including the provision of signals that promote survival and proliferation, plasma cell differentiation and the induction of AID, which drives somatic hypermutation and Ig class switch recombination. These are primarily mediated by CD40L-CD40 interaction and by the cytokines IL-4 and IL-21. They are also dependent on cell-to-cell contact, which is facilitated by adhesion molecules expressed on Tfh cells. Subpopulations of Tfh cells

influence the development of different antibody subclasses, for example, IFN- γ -producing Tfh1 cells favor the class switching of B cells to the production of the IgG subclass, IgG2a, which has antiviral effects.

Cytotoxicity

The specific killing of virally infected cells by activated CD8 + T cells involves the formation of an immunological synapse (IS), which ensures that neighboring uninfected cells are unharmed. Initial contact between the infected target cell and the effector CD8 + T cell is stabilized by the interaction of adhesion molecules on each cell (intracellular adhesion molecule-1 or ICAM-1 and lymphocyte function-associated antigen-1 or LFA-1, respectively) and is followed by engagement of the TCR with peptide-MHC. This triggers the accumulation of TCR complexes at the point of contact, together with rearrangement of the cytoskeleton and mobilization of perforin and granzymes towards the IS. Perforin forms holes in the target cell membrane, through which lysosomes containing granzymes are delivered. Granzymes induce apoptosis of the target cell by activating caspase-3, which directs the degradation of target cell DNA. After this process is complete, the effector cell detaches from the target and may engage in serial killing of other targets by the same mechanism. The importance of perforin/granzyme-mediated cytotoxicity in the control and eventual clearance of viral infections has been demonstrated in various perforin-deficient mouse models, in which animals failed to clear LCMV or Ebola virus.

While granule exocytosis is the dominant mechanism of cytotoxicity, it can also be mediated by the interaction of death ligands (Fas-ligand and TNF-related apoptosis-inducing ligand, TRAIL) on CD8 + and CD4 + T cell effectors with the death domain on receptor-bearing target cells, leading to caspase activation and apoptosis. This process is slower than granule exocytosis. Its role in the control of human viral infections is uncertain, however, *in vitro* studies have suggested that diverse viruses can trigger death receptor-mediated responses (examples being IAV, HPV, EBV, HSV, HCV, West Nile virus (WNV), and HIV-1).

Contraction of Cellular Immune Responses and the Formation of Immune Memory

After clearance or control of a viral infection the effector immune response must contract in order to avoid excessive inflammation and tissue damage. This self-limiting process is fundamental to the prevention immune-mediated disorders. It is achieved by the elimination of the majority (>90%) of effector CD4 + and CD8 + T cells via programmed cell death (or activation-induced cell death). In addition, up-regulation of co-inhibitory receptors (eg. programmed death-1, PD-1, and cytotoxic lymphocyte associated protein-4, CTLA-4) on effector cells leads to inhibition of activation when these proteins interact with their respective ligands on APCs, preventing further T cell proliferation and cytokine production. Following contraction, the surviving cell population assumes a memory phenotype and is maintained by homeostasis in the absence of re-infection.

Immunological memory, a hallmark of adaptive immunity that has evolved to provide a rapid and robust response to previously encountered pathogens, is achieved by maintenance of a small number of T cell clones. They are typically long-lasting and can persist throughout life. Memory T cells can be broadly categorized as either effector (T_{em}) and central memory (T_{cm}) T cells, according to the expression of the lymphoid homing and adhesion receptors, CD62 ligand (L-selectin) and CCR7. These are expressed on T_{cm} , enabling them to migrate to secondary lymphoid organs where they await recall stimuli, while T_{em} recirculate, acting as sentinels for previously encountered antigens. T_{em} retain some effector functions and are therefore poised to respond to invading pathogens. In contrast T_{cm} have superior proliferative capacity and can replenish the T_{em} population.

The Humoral Immune Response to Viral Infections

B Cells are the Drivers of Humoral Immunity

B lymphocytes develop in the bone marrow from hematopoietic stem cells via the common lymphoid progenitor that also gives rise to T cells. However, unlike T cells they continue their maturation in the bone marrow. Starting out as pro-B cells expressing the pro-BCR complex they quickly upregulate recombination-activating genes and undergo somatic rearrangement of gene segments in the variable region of the Ig locus, a process known as VDJ rearrangement. The resulting pre-B cells undergo further gene rearrangement leading to the expression of either kappa (κ) or lambda (λ) light chains, which are components of a functional B cell receptor (BCR). At this stage, the BCR is a membrane-bound IgM receptor. Immature B cells expressing IgM are tested for affinity to bone marrow-derived antigens in a similar manner to the T cells in the thymus, i.e., those that do not express a functional receptor die, while those responding with high affinity to self-derived antigens are either deleted or undergo receptor editing. Following this selection process, B cells leave the bone marrow as transitional immature B cell and enter the circulation. Through the process of alternative splicing, the IgM receptor is edited to produce IgD + naïve B cells. As for T cells, activation of naïve B cells occurs upon detection of specific antigen, either in a soluble form or presented by APCs, by the BCR. The B cell activation threshold is determined by the affinity of the BCR and dose of antigen as well as innate signals present at the time of activation. The activated B cell then mounts a T cell-dependent (TD) or T cell-independent (TI) immune response depending on the molecular composition of the activating antigen. Polysaccharides and lipid-based antigens typically induce a TI B cell response, with the production of rapid but short-lived antibodies. Structurally, they are composed of repeating antigens, which enables cross-linking of BCRs. Marginal zone B cells and B-1 cells, which are a subset of B lymphocyte expressing a more limited BCR

repertoire, all participate in TI responses. While more typically associated with antibody responses to bacteria, they are also elicited by highly glycosylated glycoproteins, such as the viral envelope of HIV-1. Protein antigens, by contrast, activate B cells to mount a TD antibody response, which leads to the production of long-lived class switched antibodies that arise much later than TI antibodies.

A TD B cell response is elicited by the BCR-mediated capture of antigen by follicular B cells or B-2 cells from follicular dendritic cells in the germinal centers of lymphoid tissue. Follicular B cells function as professional APCs by processing and presenting the antigen in the context of MHC class II to Tfh or Th2 cells. Activated follicular B cells upregulate the T cell zone homing receptor, CCR7, while downregulating CXCR5 to enable migration to the T cell: B cell border, where they interact with pre-Tfh cells. The cognate engagement between the TCR on Tfh cells and peptide-MHC II on follicular B cells, together with CD40 with CD40L and inducible T cell co-stimulator (ICOS) with ICOS ligand (ICOSL), induces signals that promote their survival and further differentiation to Tfh and GC B cells respectively. Activated B and Tfh cells are then able to migrate to follicles in a CXCL-13 dependent gradient where they form GCs to induce GC reactions. Here, B cells clonally expand and undergo affinity maturation of immunoglobulin receptors through the process of somatic hypermutation (SHM). SHM involves editing of the Ig V domain by AID and generates Igs with a higher affinity for their cognate antigen. Secreted Igs, or antibodies, consist of the two heavy chains and two light chains of the BCR, with the V domains (also known as the variable fragment, Fv) forming the antigen-binding region, and constant domains (Fc), that determine the antibody effector function through interaction with innate immune cells bearing Fc receptors. The Fv domains contain hypervariable regions, or CDRs, that contact the antigen surface. As the two Fv domains are identical, this increases antibody avidity to its target antigen. SHM is followed by class switch recombination (CSR) to generate other classes of antibodies (IgG, IgA, and IgE) with a wider range of effector function and tissue localization. The factors critical for initiating CSR include BCR signaling, AID expression, antigenic dose as well as cytokines like IL-4, TGF- β and in mouse models, IFN- γ . Protein antigens, including those expressed by viruses, typically generate IgG (specifically IgG1, 2a and 3 sub-classes) and IgA and can also elicit IgM. The terminal stage of the GC reaction involves the differentiation of GC B cells into long-lived plasma cells that mediate antigen-specific effector responses via the secretion of antibodies, and memory B cells which differentiate rapidly into plasmablasts upon re-exposure to antigen (Fig. 2).

Extrafollicular B cells have also been described that are activated in either a TD or TI manner. These cells migrate to an extrafollicular location where they form a focus of cells secreting robust but short-lived antibodies. They have been described in a number of viral infections including SIV and HIV-1.

Antibodies Play an Essential Role in Antiviral Humoral Immunity

There are three main effector functions of antibodies that play an important role in curtailing viral infections.

Neutralization

Antibodies that block viral entry or reduce viral replication are described as having neutralizing activity. This effector function is mediated by the binding of the Fv region of antibodies to viral antigens. Neutralizing activity is considered to be a crucial component of adaptive immunity as it is strongly correlated with protection against re-infection and with vaccine-induced immune protection. While it typically involves inhibition of viral attachment or fusion with the host cell receptor, it is not defined by a particular mechanism but is an *in vitro* measure of the interaction between a virus and its host cell. The discovery of post-fusion neutralization in the context of influenza virus infection illustrates that the underlying mechanisms may be more diverse than had been appreciated. The targets of neutralizing antibodies are typically exterior proteins that make up the viral envelope or capsid. Neutralizing activity can be demonstrated in sera from individuals who have successfully controlled viral infections. However, they are also found in chronic infections such as HIV-1 and HCV. In this context, the failure of neutralizing antibodies to contain the infection is a consequence of multiple immune evasion mechanisms, including the slow development of antibodies that recognize diverse antigenic determinants in the viral envelope (broadly neutralizing antibodies or bNAbs) in the face of rapid viral evolution.

Opsonization

Secreted antibodies can bind to viral targets either on infected cells or free virus particles, coating these targets in a process known as opsonization, which mediates antibody-dependent cellular phagocytosis (ADCP) by macrophages and neutrophils, or activation of the classical complement pathway. Binding of the first component of the complement system triggers a cascade of protease activity, resulting in the formation and deposition of a membrane attack complex on the virus or viral protein, causing its destruction. Complement-mediated neutralization has been described in a wide range of viral infections. A notable example is Ebola virus disease (EVD), in which antibodies from survivors exhibited IgG and IgA-mediated innate immune effector functions *in vitro*, including ADCP. Opsonization may also play a role in pathogenesis, for example by enhancing viral infection of DCs via complement receptors, as has been described in HSV-2.

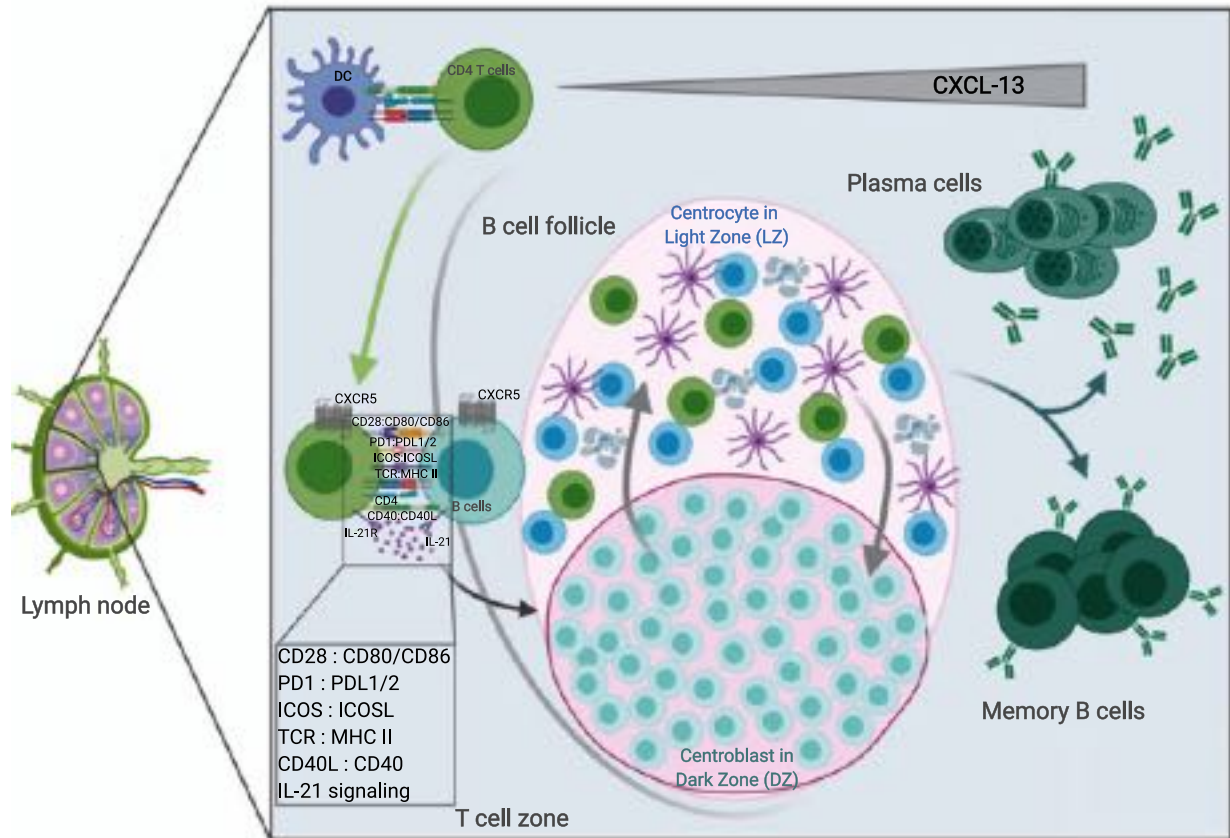


Fig. 2 The germinal center reaction. Antigen-primed CD4 T cells destined to become Tfh cells migrate to the T:B cell border where they interact with primed cognate B cells. Tfh and B cells then migrate to form the GC composed of two distinct microanatomical sites – LZ and DZ. Within the DZ, B cells proliferate and undergo SHM before cycling into the LZ, where high affinity B cell clones get survival signals from Tfh cells and FDCs, and undergo CSR. Cycling between the DZ and LZ continues, leading to the eventual selection of high affinity centrocytes that exit the GC as antibody-secreting plasma cells or memory B cells. GC, germinal center; DZ, Dark zone; LZ, Light zone; FDC, Follicular dendritic cells; CSR, Class switch recombination; SHM, Somatic hypermutation.

Antibody-Dependent Cellular Cytotoxicity (ADCC)

Following binding of the antibody Fab region to viral antigens, binding of the exposed Fc region by Fc receptor (FcR)-expressing cells can initiate direct cell lysis in a process known as ADCC. FcR are found on all innate immune cells and are defined according to their specificity for different antibody subclasses (FcγR – IgA, FcεR – IgE, and FcγR – IgG). On NK cells, the crosslinking of FcγRIIIA (CD16) mediates ADCC. This effector function was shown to be important in the immune response to influenza over 40 years ago. Recently, it has emerged that HIV-1-specific bNAbs may be capable of inducing ADCC, suggesting a greater breadth of effector function for these antibodies, that develop in a minority of individuals during chronic infection. Furthermore, vaccine-induced ADCC has been associated with the modest protection found in the sole HIV-1 vaccine trial that has shown any protective efficacy to date (the Rv144 trial).

Induction of Immunity to Viruses by Vaccination

Vaccination has been used for over 200 years to protect from infectious diseases. The introduction of vaccination against common childhood viral and bacterial infections during the course of the 20th century led to dramatic declines in mortality in children under 5 years of age. The eradication of smallpox (caused by Variola virus) using the live VACV vaccine (from which the intervention takes its name) is one of the most significant achievements in the history of medicine. For decades, many successful vaccines were developed empirically, using whole organisms that were attenuated in cell culture or inactivated by heat or chemical treatment to render them safe for human use (Table 2). The mechanisms of immune protection, and the pathogen components necessary to achieve this, were incompletely elucidated in many cases. However, protection could be demonstrated in animal models by passive transfer of blood components containing antibodies or cells to naïve animals, which were then subjected to pathogenic challenge. The advent of recombinant DNA technology led to the production of the first subunit vaccine against a

Table 2 Successful vaccines and their inferred mechanism of action in mediating protection

<i>Vaccination platform</i>	<i>Vaccine</i>	<i>Licensure year</i>	<i>Inferred mechanism of action</i>
Live-attenuated vaccine	Yellow Fever 17D vaccine	1932	Neutralizing antibody
	Smallpox ^a	1940	Neutralizing antibody
	Influenza vaccine (H1N1, H3N2) ^b	1945 with updates in 2003, 2010, 2012	Strain-specific neutralizing and binding antibodies and CTL responses
	Polio vaccine	1955	Neutralizing antibody
	Measles vaccine ^c	1963	Induction of binding antibodies
	Mumps ^c	1967	Neutralizing antibody
	Rubella ^c	1969	Induction of binding antibodies
	Varicella vaccine	1996	Induction of binding antibodies
	Rotavirus vaccine	1998	Ab response
Inactivated whole organism	Zoster vaccine	2006	Induction of binding antibodies, CD4 T cell responses, Lymphoproliferation
	Japanese encephalitis virus vaccine	1954	Neutralizing antibody
	Rabies vaccine	HDCV 1980	Neutralizing antibody
	Hepatitis A vaccine	1996	Induction of binding antibodies
	Polio vaccine	1955	Neutralizing antibody
	Influenza vaccine (H1N1, H3N2) ^b	1945 with updates in 2003, 2010, 2012	Neutralizing antibody
Purified protein	Hepatitis B virus	1981	Induction of binding antibodies
	Haemophilus influenza type B vaccine	1985	Neutralizing antibody
Virus-like particle	Human Papilloma virus vaccine	2006 updated in 2014 ^d , 2009 ^e	Neutralizing antibody

^aEradicated in 1980.^bRequired yearly modification of vaccine.^cCombined into MMR (measles, mumps and rubella) vaccine in 1971.^dGardasil HPV vaccine.^eCervarix HPV vaccine.

Abbreviation: HDCV, Human diploid cell vaccine.

virus: recombinant HBV surface antigen, a component of the viral envelope, is one of the most successful vaccines to date. In parallel, advances in molecular and cellular technologies have enormously enhanced our understanding of the mechanisms involved in vaccine-induced and natural immunity. The insights gained have been translated into new and improved vaccines.

Over the past decade, high-throughput ‘omics’ approaches have been applied to the study of the human immune response to viruses, using vaccinated individuals as a model of effective immunity. These ‘systems biology’-based studies of the response to vaccines against yellow fever virus (YF-17D) and seasonal influenza highlighted the importance of engaging the innate immune system, via PAMP-PRR interactions, in order to achieve an effective adaptive immune response. Furthermore, induction of potent CD8 + T cell responses appears to be dependent on an acute viral infection, which is achieved with live attenuated viral vaccines such as YF-17D and measles/mumps/rubella. Strong CD8 + T cell responses are also induced with vaccine candidates based on replicating or replication-deficient viral vectors that have limited host range or have undergone targeted gene deletion. This strategy mimics an acute viral infection, with activation of PRRs by PAMPs in the vector following a single round of infection in target cells, and the desired antigen-specific response is induced by expression of the transgene in these cells. The factors that contribute to the induction of strong and durable antibody responses are less clear. Smallpox vaccinees have stable antibody levels for up to 75 years after vaccination, together with long-lived VACV-specific memory B cells. YF-17D also induces virus-specific neutralizing antibodies that can last for several decades. These observations suggest that the ‘acute viral infection’ engendered by live attenuated vaccines is important for induction of durable antibodies. However, human papillomavirus (HPV) virus-like particle (VLP) vaccines, which have proven highly effective against infection with oncogenic and non-oncogenic HPV types, comprise multimerized capsid proteins only, yet they induce neutralizing antibody titers that are several orders of magnitude higher than naturally-induced antibodies and persist for at least 10 years. This is achieved in part by administration with potent adjuvants, such as the TLR4 agonist monophosphoryl lipid A, and also by the structure of the VLPs, which contain epitopic repeats that are a typical feature of PAMPs.

There are still significant challenges in the development of vaccines for many viruses (Table 3). These relate to immune evasion properties such as antigenic variation, the capacity to establish latency or to co-opt host tolerance mechanisms, and the lack of a defined correlate of protection against natural infection. In addition, some vaccine strategies have unexpectedly increased the risk of disease through mechanisms such as antibody-dependent enhancement of infection. In the case of emerging pandemic viruses such as the novel coronavirus, SARS CoV-2, the protective effect of natural immune responses is not yet known, and it is currently

Table 3 Challenges in vaccine development and possible solutions under investigation

<i>Challenges in vaccine development</i>	<i>Possible solutions under investigation</i>
Antigenic diversity ● Rapid mutation as a result of error-prone reverse transcriptase of viruses, leading to within-population and intra-subject antigenic diversity. Mutations also enable escape from host cytotoxic T cell and antibody-mediated responses	● Use of consensus or mosaic immunogen sequences to improve breadth of viral strain coverage and immunogenicity of vaccines ● Modified vaccination strategies including sequential vaccination to mimic virus/antibody co-evolution leading to induction of bNAb development by vaccination and/or the design of immunogens that activate germline precursors of bNAbs to initiate antibody affinity maturation ● Improved mapping of bNAb epitopes on conserved regions of envelope protein for vaccine development to induce bNAb by vaccination ● Development of a universal vaccine capable of broad cross reactivity by increasing breadth of neutralizing antibodies by targeting HA stalk and conserved regions in HA head domain for Influenza vaccines ● Optimized vaccination strategies that induce and maintain CTLs at portal of viral entry prior to virus dissemination
Structural complexity of envelope protein, limiting epitope availability and antibody recognition ● Some viruses like HIV show extensive glycosylation on Env protein. This masks conserved epitopes shielding them from antibody binding ● Conformational changes in envelope protein leading to transiently exposed epitopes or complex conformation ● Structural flexibility of envelope proteins also hampers the mapping of these epitopes	● Use of stabilised Env trimers for vaccination to mimic antigenic surfaces of native trimer ● Focus vaccine development efforts to target conserved non-structural (NS) proteins which induce greater breadth in T cell responses
Poor immunogenicity of certain proteins and vaccination platforms ● Hypervariable regions of envelope protein of viruses like HIV are highly immunogenic while conserved regions are poorly immunogenic. Highly immunogenic regions create a diversionary effect (decoy) from immune response against conserved regions. ● Safety issues associated with live-attenuated vaccination platform ● Whole killed viruses and subunit vaccines not immunogenic	● Use of viral vector vaccination platforms to improve induction of CTLs ● Optimized vaccination strategy to improve immunogenicity and efficacy of various vaccination platforms and regimen inducing broad and potent T cell responses and/or antibody mediated protection e.g. targeting NP and M1 in influenza virus
Lack of appropriate models for vaccine development and testing ● Existing in vitro culture models and/or small animal models do not recapitulate human infection by HIV, HCV and ZIKV ● Absence or poorly defined correlates of protection in natural infection, vaccine trials and virulence factors ● Sporadic and unpredictable epidemics and declining incidence rate causing significant impact on evaluation of vaccine efficacy in clinical trials	● Establishment of human challenge infection models (CHIM) to test efficacy of developed vaccines although this poses ethical issues including the possible development of complications like Guillain-Barre syndrome in ZIKV infection ● Use of humanized small animal models, where possible
Host immunoregulation of bNAb development ● HIV bNAbs are elicited rarely and only in chronic infection, due to requirement for high degree of somatic hypermutation and long HCDR3 regions; these are normally constrained by host tolerance mechanisms	● Transient modulation of immune checkpoint receptors during vaccination to enable bNAb generation
Integration into cellular DNA during primary infection and persistence in latent state in long-lived CD4⁺ T cells	● Therapeutic vaccination strategies involving a 'kick' to reactivate latent HIV virus reservoir followed by a kill of virus infected cells
Antibody-dependent enhancement of infection or disease (ADE) ● Antibodies show cross-reactivity with other flaviviruses facilitating infection (also poses diagnostic challenges) in locations where several other flaviviruses are endemic	● Optimization of the design of vaccine platforms that abrogate ADE e.g. unrelated viral vector (ChAdOx1) for delivery of ZIKV antigens used in mouse studies of ZIKV vaccination

Abbreviations: HA, Haemagglutinin, NHP, Non-human primates.

unclear whether prior infection with related coronaviruses induces cross-protection. Addressing these questions may be crucial to developing an effective vaccine.

Dysregulation of Immune Response to Viruses

Immune dysregulation can manifest as loss of immune effector function, which leads to disease resulting from loss of control of viral replication, or over-activity resulting in damage to healthy tissues.

The importance of type 1 IFN (IFN-I) in orchestrating innate and adaptive immune responses to virus was discussed earlier. However, in chronic infections persistent expression of IFN-I and ISGs may be a double-edged sword: observational studies showed an association between upregulation of ISGs and faster progression of HIV-1 disease or poor control of viral replication. Consistent with this, comparative studies of SIV infection in sooty mangabeys, which do not develop sustained CD4⁺ cell loss despite acute viremia, and Rhesus macaques, which develop acquired immunodeficiency syndrome, showed that the former rapidly attenuate IFN-I expression after acute infection.

Viruses that are able to persist by establishing latency, with periodic gene expression and virion production, can cause chronic immune activation through persistent antigenic stimulation. This is accompanied by progressive loss of functional capabilities of virus-specific CD4⁺ and CD8⁺ effector T cells, including reduction in cytokine secretion and cytotoxic capability, the upregulation of immune co-inhibitory (checkpoint) receptor expression such as PD-1, CTLA-4, T cell immunoglobulin mucin-3 (TIM-3), lymphocyte-activation gene 3 (LAG-3), on the surface of cells, a dysregulated response to homeostatic cytokines IL-7 and IL-15, metabolic dysregulation and epigenetic changes resulting in the expression of transcription factors associated with dysfunction such as eomesodermin, basic leucine zipper transcription factor, ATF-like (BATF), T-bet and nuclear factor of activated T cells (NFAT). This phenomenon is well described in chronic HBV, HCV, and HIV-1 infections. It can be reversed to some degree by treatment with specific immune checkpoint receptor antagonists and/or with antiviral agents that block viral replication, thereby removing the antigenic stimulus.

Given the importance of CD4⁺ T cell help in coordinating and optimizing CD8⁺ T cell and B cell responses to viruses, infections that directly compromise CD4⁺ T cell function lead to profound dysregulation of immune responses, HIV-1 being a

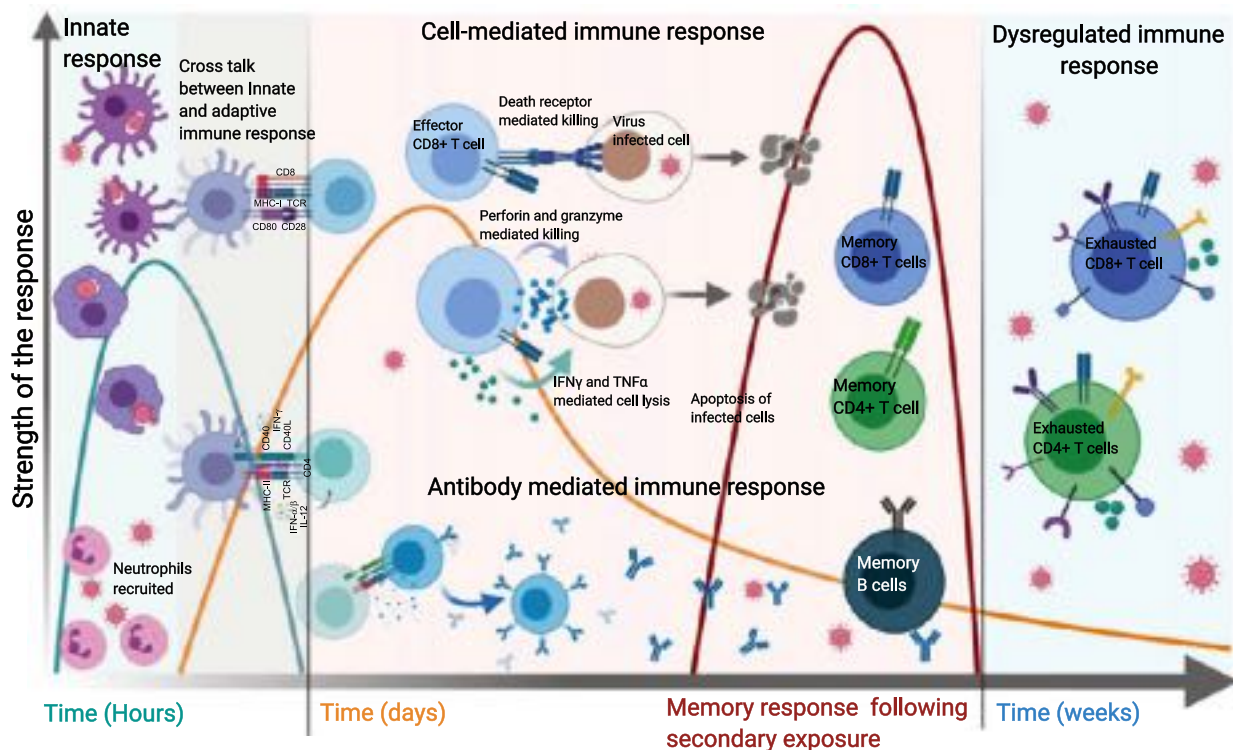


Fig. 3 Immune response to a virus. Invading viruses first encounter phagocytic cells (neutrophils and macrophages) and dendritic cells, triggering non-specific innate immune responses that aim to contain the infection. Dendritic cells capture viruses or virally infected cells and process antigens for presentation to CD4⁺ T cells, CD8⁺ T cells and B cells, priming adaptive immune responses. CD4⁺ T cells provide help to B cells to make antibodies and to CD8⁺ T cells to mature into fully functional effectors. A successful immune response ends with the clearance or control of the viral infection, contraction of the T and B cell responses and development of immunological memory. Failure to contain the infection may result in persistent viral antigen production, leading to immune exhaustion whereby T cells lose their ability to respond effectively to viruses. IFN γ , Interferon- γ ; TNF- α , Tumor necrosis factor- α ; CD, Cluster of differentiation; MHC, Major histocompatibility complex.

prime example. In chronic HIV-1 infection, loss of T_{fh} help severely impairs GC responses, resulting in reduced plasma cell numbers, a failure to class switch antibodies and slow development of bNAbs. In addition, cytotoxic CD4 + T cells can impair antibody responses by deleting marginal zone B cells and macrophages. Conversely, a reduction in Treg frequencies in chronic infection was associated with the emergence of bNAbs, which are normally under host tolerance constraints regulated by Tregs due to their resemblance to self-antibodies. Increased numbers and functional activity of Tregs have also been reported to play a role in the impairment of virus-specific CD4 + and CD8 + T effector responses in chronic HCV infection.

Conclusions

Here, we have discussed the multi-layered nature of the immune response to viruses primarily from a human perspective. However, the immune system has evolved over hundreds of millions of years, from primitive phagocytic cells that are found in invertebrates to the emergence of antigen-specific lymphocyte receptors in vertebrates to the complex cellular and molecular networks described in this article.

The key components of antiviral immunity are: (1) a physical, chemical and microbial barrier that limits viral entry; (2) innate sensing of viruses that breach this barrier by phagocytic cells possessing an array of polyspecific receptors, which initiates an immediate signaling cascade that serves to restrict viral replication and to recruit and activate NK cells; (3) priming of adaptive immune responses through the interactions of professional APCs with CD4 + T cells, CD8 + T cells and B cells; (4) maturation of antigen-specific responses by these effector population; (5) survival of a subset of antigen-specific T cells and B cells to provide immunological memory and thus protect against re-infection (Fig. 3).

Since many viruses have co-evolved with their hosts, elucidation of the mechanisms they use to overcome each layer of immune defense has led to deeper insight into the complexity of the innate and adaptive immune systems. These discoveries continue to provide opportunities to develop new and better therapeutics and vaccines, while also revealing the significant challenges that have yet to be overcome. At the time of writing, a pandemic caused by the novel coronavirus, SARS-CoV-2 has affected nearly 3 million people, with over two hundred thousand losing their lives. While its close genetic relationship to an earlier outbreak pathogen, SARS-CoV, enabled rapid determination of the basic aspects of its biology and life cycle, the exponential spread of SARS-CoV-2 demands that the mechanisms underlying both the pathogenesis and the development of natural immunity be deciphered in real time. Critical questions will be to determine how long any immune response may last, whether it is sufficient to prevent re-infection and whether a vaccine can be developed to provide durable protection.

Further Reading

- Ahmed, R., Burton, D., 2014. Viral vaccines: Past successes and future challenges. *Current Opinions in Virology* 3 (3), 307–308.
- Akira, S., Takeuchi, O., 2009. Innate immunity to virus infection. *Immunology Reviews* 227 (1), 75–86.
- Amanna, I.J., Slifka, M.K., 2018. Successful vaccines. *Current Topics in Microbiology and Immunology*. 1–30.
- Annunziato, F., Romagnani, C., Romagnani, S., 2015. The 3 major innate and adaptive cell-mediated effector immunity. *Journal of Allergy and Clinical Immunology* 135 (3), 626–635.
- Bailey, J.R., Barnes, E., Cox, A.L., 2019. Approaches, progress and challenges to hepatitis c vaccine development. *Gastroenterology* 156, 418–430.
- Graham, B.S., 2013. Advances in antiviral vaccine development. *Immunological Reviews* 255 (1), 230–242.
- Plotkin, S., 2014. History of vaccination. *Proceedings of the National Academy of Sciences of the United States of America* 111 (34), 12283–12287.
- Pulendran, B., 2014. Systems vaccinology: Probing humanity's diverse immune systems with vaccines. *Proceedings of the National Academy of Sciences of the United States of America* 111 (34), 12300–12306.
- Streeck, H., Suscovich, T., Alter, G., 2017. Immune responses to viral infection. In: Richman, D., Whitley, R., Hayden, F. (Eds.), *Clinical Virology*, fourth ed. Washington, DC: ASM Press, pp. 321–350.
- Tan, X., Sun, L., Chen, J., Chen, Z., 2018. Detection of microbial infections through innate immune sensing of nucleic acids. *Annual Reviews* 72, 447–478.
- Wang, H., Mo, Q., Yang, Z., 2015. HIV vaccine research: The challenge and the way forward. *Journal of Immunology Research* 2015 (503978).

Antigenicity and Antigenic Variation

Kuan-Ying A Huang, Chang Gung Memorial Hospital, Taoyuan, Taiwan

Xiaorui Chen and Che Ma, Genomics Research Center, Academia Sinica, Taipei, Taiwan

Dayna Cheng and Jen-Ren Wang, National Cheng Kung University, Tainan, Taiwan

Wan-Chun Lai, Chang Gung Memorial Hospital, Taoyuan, Taiwan

© 2021 Published by Elsevier Ltd.

Introduction

Antigenicity reflects the pattern of antibody responses which the host develops against a virus. For each virus, the host develops an array of distinct antibodies derived from an array of germline segments and usually refined by somatic mutation to include some with high affinity for the antigen. The particular structural regions of proteins recognized by these antibodies are designated as antigenic sites. Upon re-exposure to the viral antigen, the host immune memory reacts promptly and efficiently to produce high-affinity antibodies, which contributes to antibody-mediated immunity and a reduced risk of severe infections in the host. Under the principle of natural evolution, the virus evolves to adapt to survive pressure from the environment and host through the generation of genetic diversity. Since many viruses have a very high mutation rate this generally occurs quickly and the accumulation of genetic mutations potentially results in the modification of viral structural proteins. Antibody recognition occurs over a reasonably extensive area of the protein surface (the antibody footprint) but within this region changes in individual key residues might completely prevent recognition, for instance a larger side chain may sterically block antibody engagement, this would be an escape mutation. By improving the survival fitness of the virus, such structural changes located at or near the antigenic site will lead to the emergence of these antigenic variants in the host and environment.

The outer structural proteins often play a key role in the engagement with cellular receptors and the initiation of the entry into the host cell. Based on their outer structure, viruses can be classified into enveloped and non-enveloped viruses. An enveloped virus is constructed with a lipid bilayer membrane and the presence of surface-exposed glycoproteins. Non-enveloped viruses comprise a capsid built of assembled protein subunits. The exposed glycoprotein of enveloped viruses and the capsid of non-enveloped viruses are the principal targets of host antibodies and thus comprise the primary antigenic sites of viruses.

In most cases, viral antigenicity is analyzed based on post-infection animal sera and monoclonal antibodies. The analysis of escape variants with polyclonal antisera and monoclonal antibodies helps to define neutralizing antigenic sites and trace antigenic variation in the circulating strains. Recent advances in identifying and producing human monoclonal antibodies greatly facilitate the identification of antigenic sites and antigenic variants of human pathogenic viruses, which is of great value in virological surveillance and vaccine development.

The following sections will further explain and exemplify viral antigenicity and antigenic variation in several enveloped and non-enveloped human pathogenic viruses.

Influenza A Virus Hemagglutinin

Influenza A virus is an enveloped RNA virus that infects 5%–10% of the world population each year and kills about half a million people. Antigenicity of the influenza A virus resides in its surface membrane-embedded glycoproteins, mostly hemagglutinin (HA) and, to a lesser extent, neuraminidase (NA). At the moment two major antigenic types of influenza A virus, H1N1 and H3N2 (named after the type of HA and NA they use), are circulating that are clinically relevant to humans. Other antigenic types and their combinations are mainly found in the natural viral reservoir in animals and can occasionally undergo reassortment during a co-infection, which is facilitated by the segmented nature of the viral genome. Such genetic reassortment can potentially lead to the emergence of a radically different and novel antigenic variant, this is called antigenic shift, and it can lead to a severe pandemic, as was caused in 2009 by swine influenza H1N1.

A less dramatic but profoundly important phenomenon observed with influenza A virus is its progressive, unidirectional evolution in which gradual genetic changes accumulate in the HA antigen, a process called antigenic drift. Once mutated, the new antigenic protein is incorporated into new virus particles to be selected for their fitness in terms of survival, infectivity and the capacity to escape antibody recognition. While virus-specific antibodies produced by the immune system can prevent re-infection by a particular virus strain, these antibodies also generate an additional selection pressure to favor escape mutations in the antigen proteins for the virus to survive. Surviving viruses then become a new strain, circulate back in the population and infect individuals without adequate protection. This cycle iterates year after year, driving the direction of antigenic drift and resulting in an influenza-unique evolutionary pattern, in which new antigenic variants are constantly evolving and the old ones never seem to return.

On influenza A viruses, major antigenic sites are found on the surface HA which is a homotrimer with a moderate-to-heavy N-glycosylation profile. Each of its monomers consist of two chains, HA1 and HA2, which are linked by two disulfide bridges (**Fig. 1(a)**). Structurally, the membrane-distal globular head of HA1 binds to cellular surface sialic acid glycans for viral attachment and entry, and a cylinder-like HA2 stem near the membrane refolds into an extended helical bundle to mediate the fusion between

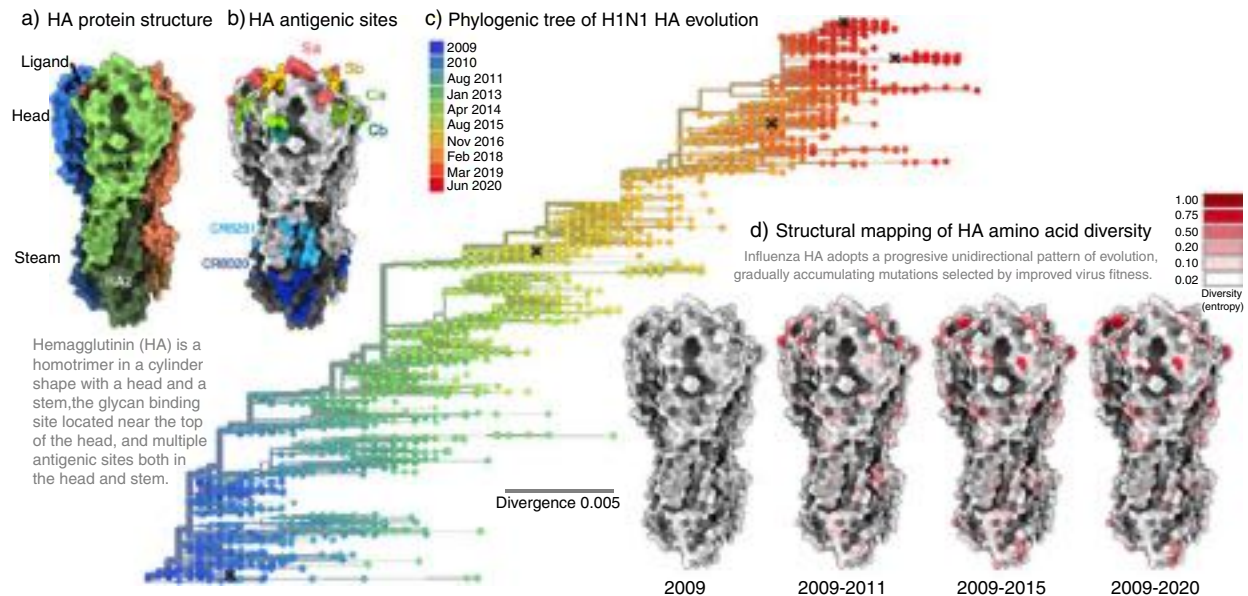


Fig. 1 Antigenic variation of influenza A H1N1 hemagglutinin. (a) Protein structure of influenza A H1 HA (PDB code 3UBJ) is shown as a homotrimer, with each monomer colored differently. Each monomer consists of two chains, HA1 and HA2, the latter in darker colors. Its glycan-binding site is near the top of the head region, and the ligand (LSTa) is shown as spheres colored by atom (gray for carbon and red for oxygen). (b) Classical antigenic sites of HA in the head region are shown in pink (Sa), yellow (Sb), green (Ca) and teal (Cb). Sites in the stem region are also recognized by recently developed neutralizing antibodies, CR6261 (light blue) and CR8020 (dark blue) as examples. (c) A phylogenetic tree of A/H1N1 HA evolution provided by nextstrain.org using data from GISAID shows a progressive pattern from 2009 to now, colored by dates, with each dot denoting a specific strain. (d) Mapping amino acid diversity on influenza A H1 HA structure using entropy data from GISAID, with darker red showing highly variable amino acids and white highly conserved amino acids. Comparison between different time courses reveals an overall consistent pattern of amino acid diversity over different lengths of time, with a few hotspot sites appearing at some time and remaining diversified. All HA structures were drawn using surface representation using ChimeraX, with identical viewing angles.

viral and endosomal membranes for the release of virus genome into the cytoplasm. There are four classical antigenic sites in the HA1 head domain, defined as Sa, Sb, Ca1/Ca2, and Cb (Fig. 1(b)), which have been characterized using escape mutants and monoclonal antibodies. Antibody-binding sites are also found in the HA2 stem region, spanning both N-terminal regions of HA1 and HA2 (Fig. 1(b)), and antibodies binding to these sites inhibit viral fusion resulting in nonreproductive viruses which are eventually cleared by host cells. In the case of influenza A H1N1 virus, progressive accumulations of mutations have occurred in the HA protein throughout the years, generating numerous new strains from the 2009 pandemic to now (Fig. 1(c)). It is noted that some mutation hotspots on HA overlap with or locate close by known antigenic sites (Fig. 1(d)), suggesting an escape mechanism whereby pre-existing antibodies will no longer bind as strongly to the new antigenic variants. At the same time, mutations are also observed in the stem region which was originally thought as highly conserved, and although less-likely it is possible that the efficacy of stem-binding neutralizing antibodies may be reduced as the protein continues evolving in the future.

Dengue Virus Envelope

Dengue virus (DENV) an enveloped RNA virus, is an arthropod-borne human viral pathogen that causes dengue fever epidemics in tropical and subtropical areas every year. DENV is categorized into four distinct antigenic serotypes (DENV1–4), which are further divided into different genotypes. Within each DENV antigenic type, genetic and antigenic diversity exists which results in multiple distinct genotypes.

Primary infection with one DENV serotype has been shown to provide life-long immunity against a similar serotype and short-term immunity against other serotypes. A longitudinal survey of DENV clinical strains over a 20-year time span from Bangkok, where all four DENV serotypes have co-circulated suggests the extinction of one clade and the replacement with another. These shifts have been shown to be linked to the epidemic cycle in the region.

Anti-envelope human antibodies are primarily directed against two glycoproteins on the surface of the viral envelope, the precursor membrane (pre-M) and envelope (E) proteins. Genetic diversity of pre-M and E proteins can significantly modulate antibody neutralization activity against DENV2 strains. Antigenic variation in the E protein has also been shown to impact the neutralization efficiency of monoclonal antibodies, antisera from DENV-infected patients, and immune sera from DENV-vaccinated individuals. Additionally, DENV serotypes contain antigenic heterogeneity. Mapping of E protein amino acid differences from different DENV antigenic serotypes revealed four distinct clusters whereby each serotype was clustered closely together on the antigenic map. However,

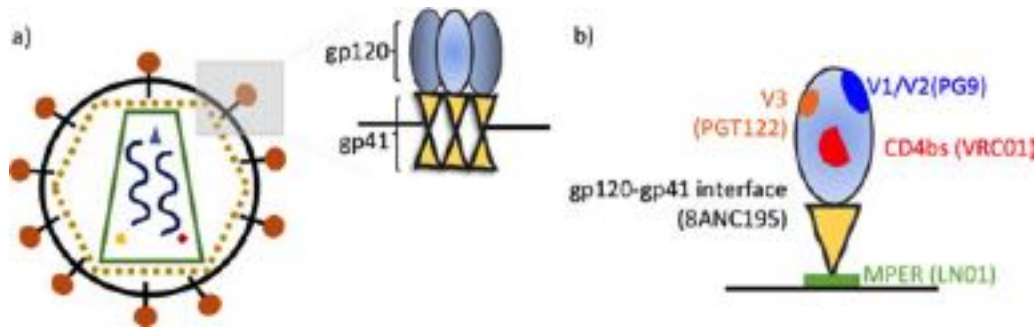


Fig. 2 Antigenicity of HIV-1 envelope glycoprotein gp160. (a) The HIV-1 envelope glycoprotein trimer, composed of gp120 and gp41 proteins. The figure demonstrates the closed conformation of the envelope protein. (b) Conserved neutralizing antigenic sites, including the first and second variable loop (V1/V2) of gp120, the third variable loop (V3), the CD4 binding site (CD4bs), the gp120-gp41 interface and the membrane proximal external region (MPER) of gp41, on the envelope glycoprotein. Examples of broadly neutralizing antibodies are presented in parentheses.

cross-reactivity of antisera with different serotypes can lead to discrepancies in the DENV antigenic cluster and certain DENV strains have shown more antigenic variance to inter-typic viruses than intra-typic viruses.

Antigenic characterization of DENV would greatly aid in DENV vaccine development. In order to promptly identify DENV antigenic variants, an in-silico model for DENV has been devised based on possible antigenicity-dominant positions of the E protein providing a convenient way to calculate the difference in viral antigenicity.

Human Immunodeficiency Virus Type 1 Envelope Glycoprotein

Human immunodeficiency virus type 1 (HIV-1) is an enveloped RNA virus and causes millions of infections around the world. HIV-1 binds and infects the host cell via the envelope glycoprotein gp160 anchored at the viral surface. The gp160 protein is a trimer of heterodimers containing glycoprotein 120 (gp120) and transmembrane glycoprotein 41 (gp41) (Fig. 2(a)). The gp120 protein is responsible for binding to CD4 cellular receptor and another co-receptor (CCR5 on macrophages and CXCR4 on T cells) and conformational changes in the gp41 protein are critical for membrane fusion and cell entry. The gp160 protein is the primary target of human neutralizing antibodies.

After infection with HIV-1, an autologous neutralizing antibody response is directed towards the envelope glycoprotein of the transmitted/founder virus. Antibodies to the envelope commonly recognize the exposed loops of gp120 and this drives the high evolutionary rate of the envelope protein. The rapid development of escape variants allows the virus to evade autologous neutralizing antibodies and the virus ends up developing into a variety of strains, also known as the “quasispecies”. Therefore, an HIV-1-infected person usually fails to establish sterile immunity and finally develops a chronic infection.

There are several mechanisms that allow antigenic variation to occur with high frequency within the HIV-1 virus. It has been noted that a single amino acid substitution, or sequence insertion or deletion, at the antigenic site of envelope glycoprotein can result in neutralization escape. Additionally, the abundance of glycans (N-linked glycosylation) on the outer surface of the gp120 protein can also shield the antigenic sites from antibody.

While autologous strain-specific neutralizing antibody response to viral envelope glycoprotein is generally described and the escape of HIV-1 is characterized by a marked genetic and antigenic divergence within strains, broadly reactive neutralizing antibody can be occasionally detected in a subset of infected persons. Many efforts have been made to delineate antigenic sites recognized by broadly neutralizing human antibodies and several conserved antigenic sites, including the first and second variable loop (V1/V2) of gp120, the third variable loop (V3), the CD4 binding site (CD4bs), the gp120-gp41 interface and the partially lipid membrane embedded membrane proximal external region (MPER) of gp41 have been reported (Fig. 2(b)).

Vaccine design has focused on eliciting neutralizing antibodies to the monomeric gp120, but this vaccine fails to demonstrate efficacy in clinical trials. Recently, the identification of conserved antigenic sites had led to the development of novel vaccine antigen for inducing broadly neutralizing antibodies. For example, a new candidate vaccine containing the soluble stabilized SOSIP trimer that mimics the native envelope has been developed and broadly neutralizing antibodies can be successfully induced in immunized animals. This epitope-focused vaccine design provides a new prospect for a preventive HIV-1 vaccine.

Enterovirus A71 Capsid

Enterovirus A71 (EV-A71) is a non-enveloped RNA virus and one major cause of outbreaks of hand-foot-mouth disease in children. The viral capsid of EV-A71 is composed of VP1, VP2, VP3 and VP4 proteins. VP4 is buried inside the capsid. Antigenic sites have been identified on the VP1, VP2 and VP3 proteins based on murine and human antibodies.

EV-A71 is classified into three genogroups (A-C), which can be further divided into 11 genotypes (A, B1-B5, C1-C5) based on the sequence variations of the VP1 gene. Human antisera show varied neutralization titers against different genotypes of viruses. Antigenic cartography showed antigenic diversity and distinct clustering among the genotypes. Genotypes B1 and B4 are clustered closely together, genotypes C2 and C4 form another cluster, while genotype B5 has been shown to form a distinct cluster of its own. This suggests a difference in EV-A71 antigenic properties and antigenic diversity within strains. Antisera from EV-A71-infected rabbits, mice, guinea pigs, and patients also demonstrate various neutralization titers against different EV-A71 genotypes, supporting the finding of antigenic variation among these viruses. The change in viral antigenicity has been linked to antigenic sites within the VP1 region. Amino acid residues VP1-98, -145, and -164 have been reported as antigenic sites for genotypes B4 and B5 EV-A71. Antigenic changes did not result from single amino acid mutations alone but instead were the result of all three mutations working in a cooperative manner.

Human antisera have varied neutralization titers against reassortant viruses containing different combinations of mutations at VP1-98K, -145Q and -164E, respectively.

Although human antisera may exhibit varied neutralizing activities with different genotypes of viruses, in most cases, cross-neutralization is detected in human sera.

Several antigenic sites on the EV-A71 capsid recognized by a panel of neutralizing human monoclonal antibodies have been identified and characterized in detail. These antigenic sites are found to be conserved among genotypes B and C EV-A71 viruses isolated in the recent decade.

Conclusion

Antigenic variation is a common strategy used by a broad range of viral pathogens to avoid host immune responses. Nevertheless, the rate of genomic mutation and the diversity of antigenic variants vary markedly among different viral species. Viral antigenicity is tightly linked with the capacity of a virus to be bound by specific antibodies, especially neutralizing antibodies. The understanding of antigenicity and associated antigenic change in the virus aids the design of protective vaccines. However, vaccines may have varying efficacy largely resulting from hard-to-predict antigenic variation, as seen for example in influenza A virus and human immunodeficiency virus type 1.

Further Reading

- Adams, B., Holmes, E.C., Zhang, C., *et al.*, 2006. Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in Bangkok. *Proceedings of the National Academy of Sciences of the United States of America* 103 (38), 14234–14239.
- Air, G.M., Laver, W.G., 1986. The molecular basis of antigenic variation in influenza virus. *Advances in Virus Research* 31, 53–102.
- Corti, D., Voss, J., Gambelin, S.J., *et al.*, 2011. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* 333 (6044), 850–856.
- Gerhard, W., Yewdell, J., Frankel, M., *et al.*, 1981. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. *Nature* 290, 713–717.
- Hadfield, J., Megill, C., Bell, S.M., *et al.*, 2018. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* 34 (23), 4121–4123.
- Huang, S.W., Cheng, D., Wang, J.R., 2019. Enterovirus A71: Virulence, antigenicity, and genetic evolution over the years. *Journal of Biomedical Science* 26 (1), 81.
- Huang, K.A., Zhou, D., Fry, E.E., *et al.*, 2020. Structural and functional analysis of protective antibodies targeting the threefold plateau of enterovirus 71. *Nature Communications* 11 (1), 5253.
- Katzelnick, L.C., Fonville, J.M., Gromowski, G.D., *et al.*, 2015. Dengue viruses cluster antigenically but not as discrete serotypes. *Science* 349 (6254), 1338–1343.
- Martinez, D.R., Yount, B., Nivarthi, U., *et al.*, 2020. Antigenic variation of the dengue virus 2 Genotypes Impacts the Neutralization Activity of Human Antibodies in Vaccinees. *Cell Reports* 33 (1), 108226.
- Murphy, B.R., Whitehead, S.S., 2011. Immune response to dengue virus and prospects for a vaccine. *Annual Review of Immunology* 29 (1), 587–619.
- Walker, L.M., Burton, D.R., 2010. Rational antibody-based HIV-1 vaccine design: Current approaches and future directions. *Current Opinion in Immunology* 22 (3), 358–366.
- Wei, X., Decker, J.M., Wang, S., *et al.*, 2003. Antibody neutralization and escape by HIV-1. *Nature* 422 (6929), 307–312.

Antigen Presentation

Andrew J McMichael, University of Oxford, Oxford, United Kingdom

© 2021 Published by Elsevier Ltd.

This is an update of E.I. Zuniga, D.B. McGavern, M.B.A. Oldstone, Antigen Presentation, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00581-1.

Introduction

The immune system is responsible for the tremendous task of fighting a wide range of pathogens to which we are constantly exposed. This system can be broadly subdivided into innate and adaptive components. The innate immune system exists in both vertebrate and invertebrate organisms and represents the first barrier against microbial invasion. This arm of the immune system rapidly eliminates the vast majority of microorganisms that we daily encounter and is responsible for limiting early pathogen replication. The adaptive response is a more sophisticated feature of vertebrate animals involving a broad repertoire of genetically rearranged receptors that specifically recognize microbial antigens (antigen is a generic term for any substance that can be recognized by the adaptive immune system). The hallmark of the adaptive response is the generation of a potent and long-lasting defense specifically directed against the invading pathogen.

B and T lymphocytes represent the effector players of adaptive immunity and carry on their surface antigen-specific receptors, B-cell receptors (BCRs) and T-cell receptors (TCRs), respectively. There are two major classes of T lymphocytes: CD8 cytotoxic and CD4 helper T cells. Upon antigen encounter, lymphocytes undergo clonal expansion and differentiation of their unique functional features. B cells differentiate into plasma cells and secrete antibodies that specifically bind the corresponding antigen. CD8 T cells directly kill infected cells or release cytokines that interfere with viral replication, while CD4 T cells activate other cells such as B cells and macrophages. Unlike B cells, which can directly bind native free antigen, T cells only recognize antigen-derived peptides displayed on cell surfaces in the context of major histocompatibility complex (MHC) class I (MHC-I, CD8 T cells) or class II (MHC-II, CD4 T cells) molecules.

Different pathogens preferentially replicate in distinct cellular compartments. While viruses and intracellular bacteria replicate in the cytosol, microbes such as mycobacterium and protozoan parasites are intravesicular and colonize the endosomal and/or lysosomal compartments. In addition, extracellular bacteria release antigens, such as toxins, that are engulfed by antigen-presenting cells (APCs) to also reach the endosomal pathway. Antigenic peptides derived from these sources are exhibited on cell surfaces by MHC molecules. This process, which represents the major focus of this article, is named 'antigen presentation' and is a fundamental pillar of antimicrobial host defense.

Antigen-Presenting Cells

For initiation of an immune response, naive T cells need to be activated or 'primed'. For that, they require both the recognition of the specific MHC-peptide complex (signal 1) and simultaneous costimulation (signal 2). Although all nucleated cells express MHC-I and can potentially display MHC-I-microbial peptide complexes after infection, only a specialized group of leukocytes, named APCs, express both MHC-I and MHC-II as well as costimulatory molecules. The best-characterized costimulatory molecules are CD80 and CD86, which bind to the CD28 molecule on the T-cell surface. In addition, T cells express CD40 ligand, which interacts with CD40 on APCs further enhancing costimulation and enabling T-cell response. Finally, there is another group of adhesion molecules, such as lymphocyte function-associated antigen-1 (LFA-1) on APCs which binds to ICAM-1 on T cells that seal the APC-T-cell interface. During APC-T-cell interactions, all these molecules cluster together forming a highly organized supramolecular adhesion complex (SMAC), enabling the intimate contact between the two cells that is referred to as the immunological synapse (Fig. 1).

APCs are composed of macrophages, B cells, and dendritic cells (DCs). They differ in location, antigen uptake, and expression of antigen-presenting and costimulatory molecules. Macrophages are localized in connective tissues, body cavities, and lymphoid tissues. Within the secondary lymphoid tissues, macrophages are mainly distributed in the marginal sinus and medullary cords. They specialize in phagocytosis and engulf particulate antigens through scavenger germline receptors such as the mannose receptor. On the other hand, B cells form follicular structures within secondary lymphoid organs and recirculate through the blood stream and lymph seeking their specific antigen. B cells recognize antigens specifically through a rearranged BCR. DCs are the most professional and robust of the APCs. They are widely distributed through the body at an 'immature' stage of development, acting as sentinels in peripheral tissues. They continuously sample the antigenic environment by both phagocytosis and macropinocytosis, which is the engulfment of large volume of surrounding liquid. Within the secondary lymphoid organs, some DCs strategically localize within T-cell areas where they can optimally encounter circulating naive T lymphocytes that actively scan the DC network.

APCs are able to detect components of invading pathogens which trigger their activation/maturation. Specifically, pathogen-associated molecular patterns (PAMPs), as these components are termed, range from lipoproteins to proteins to nucleic acids

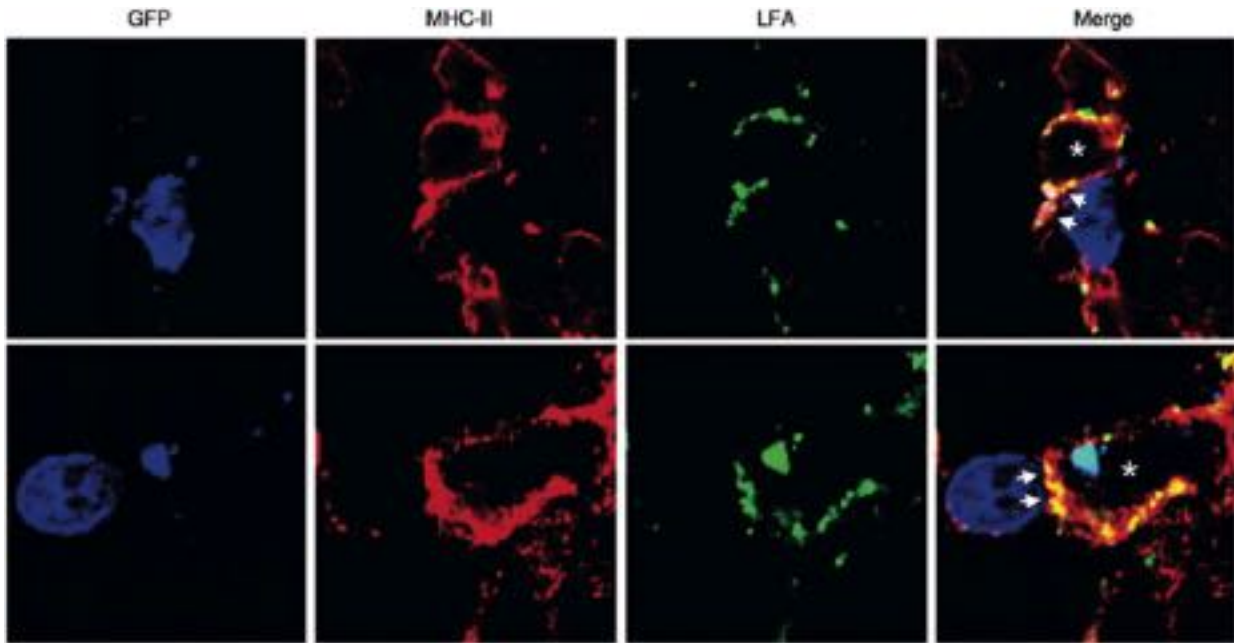


Fig. 1 Interactions between virus-specific T cells and APCs. Three-color confocal microscopy was used to demonstrate immunological synapse formation between lymphocytic choriomeningitis virus-specific T cells (blue) and MHC-II^b APCs (red) in the central nervous system. Immunological synapses were indicated by the polarization of the adhesion molecule LFA-1 (green) between the CTL and APC. Asterisks denote the engaged APC, and arrows denote the contact point between the two cells. LFA-1 is expressed on both CTLs and APCs, but note that all of the CTL-associated LFA-1 is focused toward a contact point at the CTL-APC interface. Reproduced from Lauterbach, H., Zuniga, E.I., Truong, P., Oldstone, M.B.A., McGavern, D.B., 2006. Adoptive immunotherapy induces CNS dendritic cell recruitment and antigen presentation during clearance of a persistent viral infection. *Journal of Experimental Medicine* 203 (8), 1963–1975, with permission from Rockefeller University Press.

carried by potential invaders. These PAMPs are recognized by evolutionarily conserved ‘pattern recognition receptors’ (PRRs) on APCs. Among PRRs, the Toll-like receptors (TLRs) have emerged as critical players in determining APC imprinting on the ensuing immune response. TLR triggering has pleiotropic effects on APCs, promoting survival, chemokine secretion, expression of chemokine receptors, migration, cytoskeletal and shape changes, and/or endocytic remodeling. After interacting with these pathogen signatures, the microbial antigens are processed and presented as peptides associated with MHC molecules and activated APCs upregulate both antigen-presenting and costimulatory molecules initiating a ‘maturation’ process. As part of this process, APCs in peripheral tissues change their chemokine receptors and initiate migration to secondary lymphoid organs where the adaptive immune response is initiated.

The strategic migration and location of DCs into T-cell areas of the secondary lymphoid organs coupled to their superior antigen-presenting capacity make them the most powerful APCs. Indeed, DCs are about 1000 times more efficient than B cells or macrophages in stimulating naive T cells. This has been shown by several experiments in which elimination of DCs prevented the initiation of antigen-specific T-cell responses. Interestingly, DCs are a heterogeneous cell population composed of different subtypes which present unique and overlapping functions. As many as six different subsets of DCs occupy the lymph nodes. Three major defined populations of DCs have been recognized in mouse spleen and humans: CD8⁺ conventional DCs (cDCs), CD11b⁺ cDCs, and plasmacytoid DCs. These subpopulations differ not only in surface phenotype but also in functional potential and localization. In this regard, cDCs are potent activators of naive T lymphocytes, as CD8^b DCs are believed to be specialized in cross-presentation of exogenous antigens. A recent study suggests that CD8⁺ and CD11b⁺ cDCs differ from each other in their intrinsic antigen-processing capacity being specialized in MHC-I and MHC-II antigen presentation, respectively. In contrast, plasmacytoid DCs are poorer activators of T cells, even after stimulation *in vitro*. They likely play a more protagonist role during innate immunity by secreting specific cytokines and chemokines, such as type 1 interferons (IFNs, an important antiviral mediator), and activation of a broad range of effector cells, such as natural killer (NK) cells. Thus, the heterogeneity inherent to DC populations significantly influences the varieties of immune responses to different pathogens, which are subsequently amplified by cross talk between the various subsets.

Major Pathways of Antigen Presentation

Although in all healthy individual’s MHC molecules play the same crucial role of antigen presentation, these molecules are highly polymorphic. There are hundreds of different alleles encoding the MHC molecules in the whole population and each individual

exhibits only a few of them. The major allelic variants of MHC are found in key amino acids forming the peptide-binding cleft. Thus, although a given MHC-I molecule can bind several different peptides; particular amino acids are preferred in certain positions of the peptide resulting in differential peptide sets for particular MHC variants. Importantly, T-cell specificity involves co-recognition of a particular antigenic peptide together with a particular MHC variant, a feature known as T-cell MHC restriction.

Like other polypeptide chains of proteins destined to arrive at the cell surface, MHC molecules are translocated to the lumen of the endoplasmic reticulum (ER) during synthesis. In this compartment, the subunits of MHC molecules are assembled together and the peptide binding groove or cleft is formed. However, MHC molecules are unstable in the absence of bound peptide. In the following sections, we will consider how MHC molecules are folded and generated peptides are bound to MHC-I or MHC-II molecules. After binding, MHC-peptide complexes travel to the cell surface where they are recognized by antigen-specific T cells. Although not discussed in this article, it should be noted that other MHC-like molecules (CD1 and MR1) also display lipid antigens and molecules from riboflavin biosynthesis, contributing to antigen presentation to T cells, especially during mycobacteria and other bacterial infections.

MHC-I Antigen Presentation

MHC-I molecules are expressed in most if not all nucleated cells. MHC-I molecules are heterodimers of a highly polymorphic α -chain (43 kDa) that binds noncovalently to β 2-microglobulin (12 kDa), which is nonpolymorphic. The α -chain contains three domains. The α 3 domain crosses the plasma membrane while the α 1 and α 2 domains constitute the antigen-binding site. The peptides that bind the MHC-I molecule are usually 8–10 amino acids long and contain key amino acids at two or three positions that anchor the peptide to the MHC pocket and are called anchor residues.

As mentioned above, the peptide-binding site of MHC molecules is formed in the ER. However, all proteins, including viral-derived antigens, are synthesized in the cytosol. Numerous studies in the last years outlined the molecular events connecting the antigen generation in the cytosol with the peptide binding to the MHC-I molecule in the ER. A highly conserved multi-catalytic proteasome complex is in part responsible for cytosolic protein degradation into small peptides. The proteasome contains 28 subunits forming a cylindrical structure composed of four rings, each of seven units. Under normal conditions, the proteasome complex exists in a constitutive form. During viral infections, IFNs released by cells of the innate immune system induce the synthesis of three different proteasome subunits, which replace their constitutive counterparts to form the immunoproteasome. This inflammatory form of the proteasome favors the production of peptides with a higher chance of MHC binding. Moreover, IFNs can also enhance the rate of proteasome peptide degradation increasing the availability of peptides and reducing their excessive cleavage. It is important to highlight that other cytosolic proteases also contribute to MHC-I peptide generation and further cleavage can occur within the ER before MHC binding. The source of peptides for MHC-I complexes still holds its secrets. Proteasome substrates may encompass de novo synthesized, mature stable, and/or defective proteins. It is believed that defective ribosomal products (DRiPs), which are proteins targeted for degradation due to premature termination or misfolding, constitute an important source of MHC-I peptides.

Peptides available in the cytosol are transported into the lumen of the ER by ATP-dependent transporters-associated antigen-processing 1 and 2 (TAP-1 and TAP-2) proteins. TAP proteins are localized in the ER membrane forming a channel through which peptides can pass. Within the ER, the newly synthesized MHC-I α -chain binds to a chaperone molecule called calnexin, which retains the incomplete MHC molecule in the ER. After binding to the β 2-microglobulin, calnexin is displaced and the emerging MHC molecule binds to a loading complex composed of the chaperone protein calreticulin, TAP, the thiol oxidoreductase Erp57, and tapasin, which bridges MHC-I molecule and TAP. The TAP molecule holds the MHC class I open and peptide receptive for competing peptides delivered by TAP. After peptide binding, the fully folded MHC-I molecule and its bound peptide are released from the complex and transported to the cell membrane. Importantly, under steady-state conditions, the MHC-I molecules in ER are in excess with respect to peptides allowing the rapid appearance of microbial peptides onto the cell surface during infection. However, since MHC-I molecules are unstable without bound peptide, they also present self-antigens under normal conditions. Because of the absence of microbial signatures, antigen presentation of self-peptides by inactivated/immature APCs leads to T-cell tolerance rather than activation. This is one of the important ways anti-self or autoimmune responses are controlled.

For several years, intracellular peptides were believed to be the only source of MHC-I molecules. However, it is now clear that exogenous proteins also have access to the cytosolic compartment and bind MHC-I in the ER. Also priming of CD8⁺ T cells by dendritic cells involves uptake of foreign antigen by the latter from cells undergoing apoptosis or necroptosis. This mechanism known as cross-presentation is important for initiating T cell responses by specialized dendritic cells. The molecular mechanism by which MHC-I molecules access exogenous peptides is of considerable interest. Different nonexclusive possibilities have been proposed, including sampling of phagosome generated peptides by MHC-I molecules, transference of ER molecules (including MHC-I and its loading complex) into phagosomes, re-entry of plasma membrane MHC-I into recycling endosomes with the subsequent peptide exchange, and finally the acquisition of peptides from other cells through GAP junctions.

MHC-II Antigen Presentation

The MHC-II molecule is composed by two noncovalently bound transmembrane glycoprotein chains, α (34 kDa) and β (29 kDa). Each chain has two domains and altogether they form a four-domain heterodimer similar to the MHC-I molecule. α 1 and β 1

domains form the peptide binding cleft resulting in a groove which is open at the ends, which is different from the MHC-I groove in which the extremes of the peptide are buried at the ends. Peptides that bind to MHC-II are larger than those that bind to MHC-I molecules, being 13–17 amino acids long or even much longer.

Since MHC-II is a surface protein, its biosynthesis is initiated in the ER. To prevent newly synthesized MHC-II molecules from binding cytosolic peptides that are abundant in the ER, its peptide-binding cleft is covered by a protein known as MHC-II-associated invariant chain (Ii) or CD74. Through a targeting sequence in its cytoplasmic domain, the Ii also directs MHC-II molecules to acidified late endosomal compartments, where Ii is cleaved leaving only the Ii pseudopeptide (CLIP) covering the peptide-binding groove. MHC-II molecules bound to CLIP cannot spontaneously bind other peptides, indicating that CLIP must be dissociated or displaced before antigenic peptide can bind.

Priming of CD4⁺ T cells by MHC-II occurs primarily in dendritic cells and B lymphocytes. Proteins that enter dendritic cells through endocytosis, are captured by Fc receptors as immune complexes or are derived from pathogens that replicate in vesicles and are degraded by endosome proteases. B cells bind antigen through their surface antibodies and then internalize it for antigen processing. These proteases become activated as the endosome pH progressively decreases. The final set of peptides available in the endosomal compartment is a result of antigen processing by several acid proteases that exist in endosomes and lysosomes. For instance, the cathepsin S is a very predominant acid protease and mice deficient in this enzyme have a compromised antigen-processing capacity. Vesicles carrying peptides fuse with the vesicles carrying MHC-II molecules, CLIP is then dissociated by the action of non-polymorphic MHC class II molecules, HLA-DM in humans and H-2 M in mice, allowing incorporation of antigenic peptides into MHC-II molecules. HLA-DM/H-2 M thus contributes to 'peptide editing' by removing weakly bound peptides and assuring that the emerging MHC-II-peptide complexes are stable enough to be scanned by CD4 T cells.

MHC-II molecules seem to be in excess and are rapidly degraded unless microbial peptides become available to fill the groove. This excess is important to permit MHC-II availability upon infection. However, during infection, APCs are exposed to both self and microbial peptides. How APCs discriminate between self and non-self represents a fundamental question in immune biology. Recent evidence suggests that the efficiency of presenting antigens from phagocytosed cargo is dependent on the presence of a TLR ligand within the cargo. Thus, TLR signaling would mark a particular phagosome for an inducible mode of maturation dictating the fate of the cargo-derived peptides and favoring their presentation by MHC-II molecules in a phagosome autonomous fashion.

Because they travel through the endocytic pathway, which can be considered a topological continuation of the extracellular space, MHC-II molecules were believed to be specialized in the presentation of exogenous antigens. However, the analysis of the MHC-II peptidome revealed many peptides of cytosolic or even nuclear origin. Autophagy or 'self-eating' explains MHC-II access to cytosolic peptides. This highly conserved pathway could be accomplished by several mechanisms including microphagy (when lysosomal invagination sequesters cytosolic components), macrophagy (when a double membrane structure encloses and isolates cytoplasmic components and eventually fuses with lysosome), and chaperone-mediated autophagy (when cytosolic proteases generate peptides that are transported into lysosomes).

Viral Subversion of Antigen Presentation

Considering the crucial role of antigen presentation for host defense, it is not surprising that many viruses have evolved maneuvers to evade or divert this process. Particularly, the essential role played by APCs in host defense to pathogens makes them an ideal target for viruses to suppress the immune response, thereby maximizing their chances of survival, replication, and transmission. Indeed, many viruses that cause major health problems are able to interfere with the ability of APCs to prime an efficient and effective antiviral immune response. In fact, many viruses have developed different mechanisms to subvert each stage of APC biology. Furthermore, with the greater understanding of antigen presentation pathways comes the discovery of novel viral immune-evasion strategies. In this section, we illustrate selective viral strategies to subvert antigen presentation by describing particular cases.

An interesting example of virus blockade of antigen presentation from very initial steps is the ability of the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV), to dramatically block DC development from early hematopoietic progenitors. Fms-like tyrosine kinase 3 ligand (Flt3L) is known to induce the expansion of undifferentiated progenitors into DCs within the spleen and bone marrow (approximate 20-fold increase), both in mice and humans. In contrast, LCMV-clone (CL) – 13 that suppresses the immune response and causes a persistent infection in mice is associated with DC early progenitors that become refractory to the stimulatory effects of Flt3L.

TLRs function in APCs as an early sensor against pathogens; therefore, impairment in TLR signaling confers another selective advantage to certain infectious agents. As an example, vaccinia virus (VV) blocks TLR signaling and the subsequent maturation of APCs. Specifically, two proteins of VV suppress intracellular signaling of interleukin-1 (a potent pro-inflammatory host factor) and TLR-4.

Migration of DCs is a crucial step in initiating the adaptive immune response. Examples of viruses that have developed mechanisms to prevent migration of infected DCs to lymphoid organs are herpes simplex virus (HSV) 1 and human cytomegalovirus (HCMV). In both cases, there is an inhibition of complete DC maturation and subsequent expression of chemokine homing receptors. In addition, HCMV inhibits DC migration one step further by preventing APCs from arriving at a site of infection by producing homologs of chemokines that interfere with host pro-inflammatory chemokine gradients.

Another effective immune-evasion strategy used by viruses to disrupt APCs is the prevention of or interference with antigen-specific T-cell activation. The ability to disrupt MHC-peptide binding has evolved in many different virus species including adenovirus and human immunodeficiency virus (HIV). Herpesviruses have also evolved to block host cell antigen presentation. Some mechanisms utilized by herpesviruses to disrupt the antigen-presentation pathway include blocking peptide transport to the ER through interference with TAP proteins (HSV ICP47, HCMV US6), transport of particular MHC-I heavy chains from the ER to the cytosol where they are destroyed (HCMV US11, HCMV US2), retention of specific MHC-I heavy chains in the ER (HCMV US3, murine CMV-MCMV-m152), and disruption of T-cell recognition of MHC-I on the cell surface (MCMV m04). That viruses have independently evolved numerous mechanisms to disrupt MHC-peptide presentation indicates the effectiveness and importance of this strategy to the survival of viruses with different infectious life cycles.

Maturation of APCs results in upregulation of costimulatory molecules and expression of cytokines that enable them to stimulate naive T cells. Viruses that can impair T-cell stimulation by preventing the upregulation of costimulatory molecules include Ebola virus, Lassa fever virus (LFV), HSV-1, and HIV. Additionally, a number of viruses (hepatitis C virus (HCV), HIV, measles virus (MV), and dengue virus (DV)) are also able to inhibit interleukin (IL) – 12 production, which is often required for effective T-cell response. HCV does this through the action of its core and nonstructural protein 3 (NS3), which induces production of IL-10. DV, on the other hand, is able to inhibit IL-12 production through an IL-10 independent mechanism. In addition, compelling evidence showed that in vivo persistent infection of mice with LCMV, as well as persistent HCV infection in humans, induces IL-10 production by APCs resulting in the blunting of the CD8 T-cell response and chronic viral infection. Remarkably, antibodies blocking IL-10/IL-10R interactions correct T-cell exhaustion by restoring T-cell function, which results in purging of virus from mice persistently infected with LCMV.

Finally, a novel immunosuppressive molecule, programmed cell death-1 (PD-1), is upregulated in nonfunctional CD8 T cells during chronic infections (LCMV, HIV, HCV). Interaction of PD-1 with PD-ligands on APCs (or parenchymal cells) inhibits lymphocyte activation. It should also be noted that many cancer cells upregulate the PD-1 ligand PDL1 and inhibit T cell responses; recent immunotherapies can block this interaction with some success. Similarly, antibodies interfering with PD-1/PDL interactions also promote viral clearance from a persistently infected host.

The fact that not all viruses are able to block APC maturation does not necessarily represent a failure of the pathogen or a success for the host. A good example of this is observed following MV infection that exploits the ability of DCs to mature and migrate to lymphoid organs in response to infection. MV benefits greatly by having infected DCs home to lymphoid compartments where the infected cells are able to actively suppress T-cell proliferation (mediated through T-cell contact with surface viral glycoproteins) and also facilitate virus spread to more lymphoid cells. Therefore, the full understanding of the virus–host relationship requires not only studying the active mechanisms that viruses use to disable the immune system, but by also asking how a virus benefits by not altering a particular immune function.

Concluding Remarks

Co-evolution of certain hosts and pathogens for millions of years has resulted in a fine-tuned equilibrium that enables survival of both. Antigen presentation is one of the critical elements in this balance. While antigen presentation is an essential process for long-term effective host defense, targeting APCs represents a common maneuver of many viruses to avoid host surveillance and establish a chronic or persistent infection. A major challenge in biomedical research is to thwart microbial APC subversion to promote eradication of the pathogen. A better understanding of the mechanisms used by APC to display microbial antigens as well as the virus strategies to subvert APC functions during immune responses will provide new tools for designing novel vaccination approaches and immunotherapeutic treatments for human infectious diseases.

Further Reading

- Bevan, M.J., 2006. Cross-priming. *Nature Immunology* 7, 363–365.
- Blander, J.M., Medzhitov, R., 2006. On regulation of phagosome maturation and antigen presentation. *Nature Immunology* 7, 1029–1035.
- Blees, A., Janulienė, D., Hofmann, T., *et al.*, 2017. Structure of the human MHC-I peptide-loading complex. *Nature* 551, 525–528.
- Dudziak, D., Kamphorst, A.O., Heidkamp, G.F., *et al.*, 2007. Differential antigen processing by dendritic cell subsets in vivo. *Science* 315, 107–111.
- Itano, A.A., Jenkins, M.K., 2003. Antigen presentation to naive CD4 T cells in the lymph node. *Nature Immunology* 4, 733–739.
- Janeway, C.A., Travers, P., Walport, M., Shlomchik, M.J., 2005. *Immunobiology: The Immune System in Health and Disease*. New York: Garland Science Publishing.
- Lauterbach, H., Zuniga, E.I., Truong, P., Oldstone, M.B., McGavern, D.B., 2006. Adoptive immunotherapy induces CNS dendritic cell recruitment and antigen presentation during clearance of a persistent viral infection. *Journal of Experimental Medicine* 203 (8), 1963–1975.
- Menendez-Benito, V., Neefjes, J., 2007. Autophagy in MHC class II presentation: Sampling from within. *Immunity* 26, 1–3.
- Norbury, C.C., Tewalt, E.F., 2006. Upstream toward the 'DRiP'-ing source of the MHC class I pathway. *Immunity* 24, 503–506.
- Oldstone, M.B., 2007. A suspenseful game of 'hide and seek' between virus and host. *Nature Immunology* 8, 325–327.
- Reis e Sousa, C., 2006. Dendritic cells in a mature age. *Nature Reviews Immunology* 6, 476–483.
- Shen, L., Rock, K.L., 2006. Priming of T cells by exogenous antigen cross-presented on MHC class I molecules. *Current Opinion in Immunology* 18, 85–91.
- Shortman, K., Liu, Y.J., 2002. Mouse and human dendritic cell subtypes. *Nature Reviews Immunology* 2, 151–161.
- Strawbridge, A.B., Blum, J.S., 2007. Autophagy in MHC class II antigen processing. *Current Opinion in Immunology* 19, 87–92.
- Yewdell, J.W., Nicchitta, C.V., 2006. The DRiP hypothesis decennial: Support, controversy, refinement and extension. *Trends in Immunology* 27, 368–373.

Defense Against Viruses and Other Genetic Parasites in Prokaryotes

Kira S Makarova, Yuri I Wolf, and Eugene V Koonin, National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, United States

© 2021 Published by Elsevier Ltd.

Introduction

Parasitism, including genetic parasitism, is a pervasive and, arguably, fundamental feature of life. The constant competitive coevolution between parasites and their hosts is the classical case of the Red Queen evolutionary dynamics and is believed to have profound consequences for the history of life on Earth. It is not surprising, therefore, that genomes of all life forms carry the deep imprint of this constant arms race. Genomes of hosts harbor numerous active and passive defense systems and mechanisms, while those of parasites are replete with systems adapted to overcome these defenses.

A genetic parasite can be defined as a non-cellular DNA or RNA-based life form, which relies on the host resources and cellular machinery to propagate. Genetic parasites include viruses, plasmids and transposable elements. These elements are typically deleterious to the host, but some of them could be only mildly so, neutral, or even beneficial to the host in some conditions. The focus of this article is to review organization and evolution of major types of defense systems that prokaryotes (bacteria and archaea) deploy against genetic parasites. However, it is important to note that many components and principles of organization of these systems are similar to those molecular systems that are involved in conflicts between cellular life forms. The latter include numerous offense systems such as polymorphic toxins and virulence factors and the defense or immunity systems counteracting these toxins.

Bacterial, archaeal and eukaryotic defense systems and mechanisms share similar organization principles ([Table 1](#)). Moreover, in full agreement with the arms race paradigm, the parasites also have an impressive arsenal of both active and passive mechanisms to overcome host defenses. Passive defense prevents the entry of a genetic parasite through the cellular membrane. Typically, viruses and plasmids use specific surface proteins (receptors) and/or cellular DNA uptake machinery to enter the cell. Therefore, these receptors or outer components of DNA uptake machinery could be either modified, preventing their recognition, or physically blocked. There are multiple ways to modify these components, including minor (mutations) and major (rearrangements, insertions, deletions) changes of the respective protein sequences and enzymatic modifications of proteins, such as (partial) proteolysis and glycosylation. Receptor blocking includes protein-protein interactions, glycosylation or occlusion of the receptor by glycan capsule or extracellular matrix. In addition, some bacteria deploy decoy outer membrane vesicles that contain receptors and therefore mislead phages into binding to them instead of the receptors on the cell surface. In eukaryotes, analogous mechanisms of passive protection are also known and some of them involve activity of domesticated viruses which express proteins, blocking cognate receptor and preventing its binding by other viruses.

Active immunity mechanisms typically involve intracellular response to foreign nucleic acid or general stress induced by parasite activity inside the cell. These mechanisms can be roughly classified into specific immunity and non-specific response causing cell suicide (programmed cell death, PCD) or dormancy ([Table 1](#)). The latter mechanisms include numerous toxin-antitoxin (TA) and abortive infection (ABI) systems in prokaryotes discussed below and elaborate apoptosis mechanisms centered around caspase family proteases in eukaryotes. These systems could be also activated by other causes, such as DNA or protein damage. Specific immunity systems are classified into innate immunity and adaptive immunity and depend on the ability to distinguish self versus nonself ([Table 1](#)). Adaptive immunity systems differ from innate immunity in that they develop a highly specific immune response during organism exposure to a parasite; typically, adaptive immunity is more efficient than innate immunity. In prokaryotes, both innate immunity and adaptive immunity operate on the level of nucleic acids recognition and targeting. In eukaryotes, much of innate immunity and all known adaptive immunity mechanisms are based on protein-protein recognition. The only known defense system that is common between prokaryotes and eukaryotes is centered on PIWI domain (nuclease of RNase H fold) containing proteins known as Argonats (AGO). RNA-guided AGOs play key roles in the multiple pathways of RNA interference which protect eukaryotes against RNA viruses; separate branches of the RNA interference network evolved to regulate translation of eukaryotic mRNAs. In contrast, prokaryotic AGOs are either RNA- or ssDNA-guided and so far have been shown to target, primarily, foreign DNA. Several other prominent components of defense systems are also shared between prokaryotes and eukaryotes, but their roles in defense in either prokaryotes or eukaryotes are not well understood. For example, the bacteria and archaea encode numerous immunoglobulin (IG) domains and leucine-rich repeats (LRR) as well as caspase family proteases that can be tentatively predicted to function in defense, but their actual functions remain to be studied experimentally. Conversely, the roles of many eukaryotic nucleases derived from bacterial TA and ABI system remain unclear or are unrelated to defense. Thus, further research might identify more common traits in the organization of defense pathways in all three domains of life.

Table 1 Classification of defense systems and comparison between prokaryotes and eukaryotes

Active response			Passive resistance
Self-nonsel self differentiating			
Adaptive immunity	Innate Immunity	Programmed cell death or dormancy	Avoidance
Bacteria and archaea			
CRISPR-Cas	RM, DND, DPD, PGL/BREX, DISARM, <u>PIWI</u>	TA, ABI	Receptor mutations, Receptor blocking, Decoy OMV
Eukaryotes			
IG immunity, LRR immunity	<u>PIWI</u> , Interferon	Caspase pathways	Receptor mutations, Receptor blocking

Abbreviations: ABI, abortive infection systems; BREX, bacteriophage exclusion system; Cas, CRISPR-associated proteins; CRISPR, clustered regularly interspaced short palindromic repeats; DISARM, defense island system associated with restriction modification; DND, DNA phosphorotiation and degradation system; DPD, 7-deazapurine in DNA system; IG, immunoglobulin; LRR, leucine rich repeats; OMV, outer membrane vesicles; PIWI, argonaute or related RNA or DNA interference systems; RM, restriction modification systems; TA, toxin-antitoxin systems.

Note: The only defense system shared between prokaryotes and eukaryotes, centered around the Argonaute (PIWI) proteins, is underlined.

Classification and Diversity of Defense Systems

Toxins-Antitoxins: PCD or Dormancy Systems

All PCD systems harbor toxic components that either kill the host cell or cause major disruption in house-keeping pathways, triggering reversible cell dormancy. These toxins might be either inactivated or suppressed by a dedicated antitoxin molecule and/or repressed/de-repressed by a dedicated transcriptional regulator. Although it is common to refer to two PCD systems in bacteria and archaea, TA and ABI, the distinction between these is fuzzy. TA modules are most abundant defense/stress response systems in archaea and bacteria. Currently, TA systems are classified into 6 types depending on the nature of the toxin and antitoxin, and on the mode of their interaction. Type I systems consist of protein toxin, typically, a small pore-forming membrane protein, and an RNA antitoxin. The antitoxin RNA hybridizes with the toxin mRNA and prevents its translation. In type II TA, both components are proteins. The antitoxin in these systems often has a dual function. In addition to directly interacting with the toxin and blocking its activity, the antitoxin often contain a DNA-binding domain and doubles as a transcriptional regulator. Most known type II toxins are RNases of 6 distinct structural folds, namely, RelE/Txe (RelE-like fold), MazF/Kid/PemK (SH3-like fold), PIN (PIN-like fold), VapC (ferredoxin-like fold), HicA (dsRNA-binding-like) and HEPN (four-helical up-and-down bundle, named for “higher eukaryotes and prokaryotes nucleotide-binding domain”), which all target host RNAs. A minority of toxins have other toxicity mechanisms. New mechanisms of toxicity continue to be discovered and include most recently characterized acetyltransferases TacT, AtaT, ItaT and related toxins that acetylate amino groups of tRNAs, DarT which ADP-ribosylates single-stranded DNA, and RES domain toxins that either degrade NAD or act as mono-ADP-ribosyl-transferases modifying phosphoribosyl pyrophosphate synthetase, an essential enzyme of nucleotide biosynthesis. In type III systems, such as ToxIN and CptIN, RNA antitoxins interact directly with MazF-like RNase toxins. The ToxIN system is involved in abortive infection mechanism of phage exclusion and was previously discovered and described as ABI system AbiQ in *Lactococcus lactis* plasmid pSRQ900. In the type IV CbtA/CbeA system, toxin CbtA inhibits cell division by targeting FtsZ and MreB whereas antitoxin CbeA protects these proteins from the interaction with the toxin but, in contrast to most of the TA systems, does not bind the toxin. In type V system, GhoT is a membrane protein whose mRNA might be cleaved by the antitoxin GhoS, an endoribonuclease of the ferredoxin fold. Type VI systems consist of the toxin SocB which blocks replication by binding DnaN protein and antitoxin SocA that promotes degradation of the toxin by ClpXP proteases. Recently, the first all-RNA toxin-antitoxin system SdsR/RyeA has been reported. The mechanism behind the toxicity is repression of *yhcB* gene coding for an uncharacterized membrane protein of DUF1043 family. The simple organization of TA systems and several distinct features of each type provide for prediction of new TA variants using comparative genomic and sequence analysis methods. Many new TA systems have been so predicted but not yet explored experimentally, and undoubtedly, many more remain to be discovered.

So far there have been no consistent attempt to classify ABI system because of their enormous diversity, and therefore, these systems are currently treated as one loose group based on the phage exclusion effect of their expression. Moreover, most of them were identified in a single bacterial group, lactococci, so the potential diversity of these systems clearly is far from being exhausted. Many ABI systems consist of two components which is compatible with their classification as TA systems, whereas most of the rest consist of a single gene. Three TA/ABI systems have been characterized experimentally, namely, the AbiQ family of type III TA systems, AbiE family of type IV TA systems that consists of the nucleotidyltransferase toxin AbiEii and the non-interacting antitoxin AbiEi which represses expression of both genes. For the overwhelming majority of the ABI systems, the molecular mechanism is not yet known.

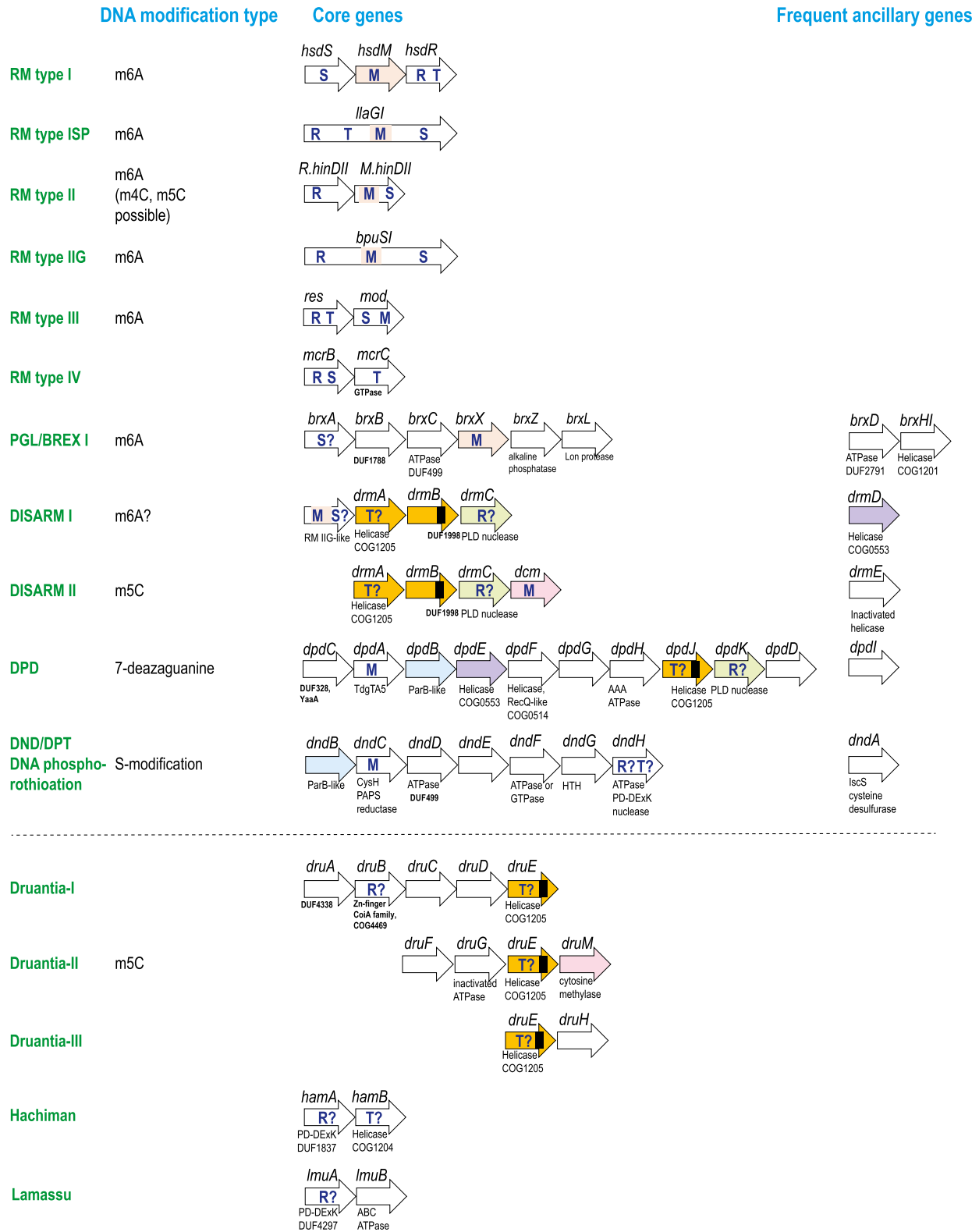
Self Versus Nonself Recognition Innate Immunity Systems

The simplest innate immunity system is the prokaryotic Argonaute (pAGO). pAGOs can use short RNA or ssDNA guides to target foreign DNA in a sequence-specific manner. However, some pAGOs can cleave DNA in a guide-independent mode. The latter activity might provide ssDNA guides for the guide-dependent DNA cleavage. The mechanism of self versus nonself discrimination during target loading is not understood yet. Sequence comparison, in agreement with phylogenetic analysis, divides pAGOs into two major groups that are known as long and short pAGOs depending on the domain organization. Long pAGOs have the same domain composition as eukaryotic AGOs and are comprised of four distinct regions: N-terminal region involved in target-guide duplex positioning, PAZ (PIWI-Argonaute-Zwille) domain responsible for binding 3' guide-RNA binding, MID (middle) domain involved in 5'-guide binding and catalytic PIWI domain of RNase H fold. Short pAGOs contain only the MID and PIWI domains. All short pAGOs and several long pAGOs are predicted to be inactivated because of substitutions in catalytic sites in the PIWI domain. Inactivated long pAGOs are typically associated with different nucleases which might function as effectors, possibly, interacting with pAGO and cleaving DNA. The key feature of short pAGO systems is the presence of the yet uncharacterized APAZ ("analog of PAZ" which, however, displays similarity to the N-terminal domain of AGO) domain-containing protein encoded in the same operon. Putative effector proteins, such as SIR2 (sirtuin) deacetylases or nucleotide-binding TIR (Toll/interleukin-1 receptor) domains, are also encoded in the same operon or fused to the APAZ domain. Their exact functions remain unclear. Another bacterial protein family (PF13032 RNaseH_pPIWI_RE, formerly known as DUF3893) distantly related to pAGOs is linked to the helicase DinG and a nuclease of the PD-DEXK family, and is proposed to function as an RNA-dependent restriction system. So far none of the RNaseH_pPIWI_RE proteins has been studied experimentally.

The innate immunity systems we describe next depend on the DNA modification status for self versus nonself discrimination (Fig. 1). Restriction Modification (RM) systems are by far the best-studied because, from the time of their discovery, they are extensively being used in molecular biology as an essential tool for sequence-specific DNA cleavage. The RM systems are currently classified into four types based on differences in their gene organization, target recognition and location of cleavage sites. Type I includes classic RM systems that consist of three genes with four distinct functions, namely, HsdR (R), endonuclease fused to a helicase or DNA translocase domain (T), HsdM (M), methyltransferase, and HsdS (S), the specificity subunit containing at least one target recognition domain (TRD) that determines sequence specificity (Fig. 1). These proteins form hetero-oligomeric complexes that cleave DNA away from the recognition site. The key difference between type I and type II is the absence of the helicase (translocase) domain in the latter. Many of the type I enzymes are homodimeric or monomeric with all the functionalities (R, M, S) encoded within the same polypeptide. The absence of DNA translocases suggests that the DNA cleavage occurs within the recognition site or close to it. The extremely numerous and abundant Type II RM consist of two genes, SR and SM (Fig. 1). However, domain architectures of these systems are variable. The RM systems of Type IIP (where P stands for "palindromes"), such as EcoRI, contain both S and R domains within the same polypeptide and cleave palindromic sites which makes them particularly well-suited for recombinant DNA technology. Thus, several thousand of these enzymes have been characterized experimentally, many of them in detail. Type III RM systems, such as, for example, EcoP1, typically have all four functionalities combined in two proteins: Mod, containing M and S domains, and Res, containing R and T domains. Type IV is the only variety of RM systems that lack the M component because these enzymes cleave DNA modified by phage methylases. McrBC from *Escherichia coli* K12 is the best studied representative of this RM type.

Two additional microbial defense systems rely on DNA methylation for self versus nonself discrimination; these systems appear to be more complex than the typical RM, and their investigation has started only relatively recently. One of these is BREX (Bacteriophage Exclusion), which methylates bacterial DNA, providing for discrimination between the host and phage genomes, and blocks phage replication. BREX shares several key components with the PGL (phage growth limitation) system. PGL had been originally suggested to function as "reverse RM" because in *Streptomyces coelicolor* this system provides resistance only during the subsequent rounds of infection by the same phages, suggesting that phage DNA is tagged for recognition during the first round of replication making the phage susceptible to PGL later. However, at least some of the BREX systems act in a "direct" way analogously to RM systems and provide phage protection during the first round of infection. There are three key, essential components in these systems: PglX, a methylase homologous to HsdM component of type I RM systems, PglZ, an alkaline phosphatase domain containing protein, and BrxC/PglY, an ATPase domain containing protein (Fig. 1). No nuclease domain was identified in these proteins which is in line with the absence of experimental evidence of DNA cleavage. Thus, the phage restriction mechanism remains to be clarified. Based on PglZ family phylogeny and gene neighborhood analysis, BREX system is divided into 6 subtypes, one of which corresponds to PGL. Apart from the conserved core, each of these subtypes have additional genes. BrxA, a protein structurally similar to the recognition domains of RM systems, BrxB, an uncharacterized protein of DUF1788 family, BrxL, a Lon family protease, BrxW, a serine-threonine protein kinase, and BrxH, a helicase, are among the most common ancillary genes (Fig. 1). However, their role in BREX function remains enigmatic.

The recently identified DISARM (defense island system associated with restriction modification) system has been shown to methylate host DNA and restrict invading dsDNA phages. There are two types of DISARM systems that share three proteins: DrmA, a COG1205 helicase, DrmB, a DUF1998 domain containing protein previously noticed to be fused to a COG1205 helicase, and DrmC, a phospholipase D (PLD) family nuclease. Two common associated genes differ between the two types of DISARM. DISARM I loci encode an additional helicase DrmD (COG0553) and an adenine methylase DrmMI that appears to be a derivative of type IIG RM systems, with the nuclease domain likely inactivated. DISARM II loci encode the cytosine methylase DrmMII and an uncharacterized protein DrmE, an inactivated helicase containing C-terminal helix-turn-helix DNA-binding domain. The



DISARM II system has been studied experimentally and cytosine methylation by DrmMII has been shown. However, it has been also demonstrated that some phages that lack methylation motifs are nevertheless restricted by DISARM II system, suggesting that the protection mechanism could be more complex. This is in line with the observation that the PLD family nuclease is not essential for the protection.

Currently, two innate immunity systems are known that employ DNA modifications other than methylation for self versus nonself discrimination. The DPD (for 7-deazapurine in DNA) system inserts 7-deazaguanine derivatives and restricts unmodified DNA. Typically, the DPD system is encoded in two gene modules that include, respectively, the enzymes of the 7-deazapurine biosynthesis pathway and proteins involved in the recognition and restriction of unmodified DNA. The latter module typically encodes 10 proteins (DpdABCDEFHIJK) but in some bacteria shows more variability, so it is presently unclear which genes are essential for the system functions and which are ancillary. The key gene in the cluster is DpdA/TgtA5, the queuine/archaeosine tRNA-ribosyltransferase family enzyme, which exchanges guanine for the modified base in DNA. In addition, the *dpd* genes encode three distinct helicases, an AAA ATPase, a DNA binding protein containing ParB family NTPase domain, a PLD nuclease and a few uncharacterized proteins (Fig. 1). The PLD nuclease is strongly linked to the COG1205 helicase which is fused to a DUF1998 domain containing protein. The same genes are present in DISARM systems (Fig. 1). The details of the restriction mechanism by the DPD system are not understood yet.

The second system, DND (for DNA degradation), also often referred as DPT (DNA phosphorotioation) system, replaces a nonbridging oxygen in the DNA backbone with sulfur and cleaves unmodified DNA. The DND system consists of two modules: DndABCDE proteins that are essential for DNA modification, and DndFGH proteins that are required for restriction. DndA/IscS, cysteine desulfurase and DndC/CysH, ATP pyrophosphatase related to phosphoadenosine phosphosulfate reductase, appear to be the enzymes that catalyze oxygen replacement with sulfur. Other components, namely, DndB and DndE, two distinct DNA-binding proteins, and DndD, an ATPase of the Mre11/Rad50 family, are likely responsible for specific site targeting and providing energy for sulfur incorporation. It has been demonstrated that DndACDE proteins form a complex and DndB is involved in transcriptional regulation. Although the restriction mechanism remains obscure, sequence analysis suggests that DndH, which contains a PD-DExK endonuclease domain, cleaves unmodified DNA. Other components, DndF, an ATPase and DndG, another DNA binding protein, are also essential for restriction. Interestingly, only approximately half of the genomes that encode DNA phosphorotioation module, also have the DndHGF restriction module, suggesting that alternative restriction mechanisms might be coupled with the phosphorotioation module. Indeed, it has been recently shown that HNH nuclease associated with type IV RM systems cleaves both methylated and phosphorotioated DNA.

CRISPR-Cas – The Adaptive Immunity System

CRISPR-Cas (CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats; Cas – CRISPR associated proteins) is the only known adaptive immunity system in bacteria and archaea. During a (failed) virus infection, short fragments of virus DNA are inserted into a CRISPR array becoming unique spacers that separate the repeats. The array is transcribed, and the primary transcript is cleaved by a dedicated complex of Cas proteins, giving rise to crRNAs (CRISPR RNAs) which serve as guides for the effector machinery of the system. The memory of virus encounters is stored in the genome and is inherited by the progeny of the “vaccinated” bacterium or archaeon, so that the CRISPR-Cas systems embody Lamarckian-type evolution. The typical length of a spacer is about 40 nt, so that CRISPR-cas can be programmed to target unique sites even within a large eukaryotic genome. This property quickly made the CRISPR-Cas system the main tool in the new generation of genome editing techniques and other biotechnological applications, and triggered extensive research on all aspects of the molecular mechanisms, genetic organization, structure and evolution of this system. Numerous up to date reviews discuss different topics of CRISPR-Cas biology and genetic engineering applications (see Further Reading list), so here we briefly review only general features of CRISPR-Cas systems organization and their current classification.

Due to the remarkable diversity of the CRISPR-Cas systems, their classification is hierarchical and is based on a multipronged approach. Such features as gene composition, phylogeny of CRISPR-Cas signature gene *cas1*, sequence and structural similarity of

Fig. 1 Genetic organization of self vs. nonself discriminating innate immunity systems. Abbreviated classification and the type of the system are indicated on the left in green font. Abbreviation of the systems are the same as in the Table 1. DNA modification type abbreviations: m6A – Methylation of the adenosine base at the nitrogen-6 position; m5C – Methylation of the C-5 carbon of cytosine. Arrows schematically indicate the most frequent genes associated with each system. Above an arrow the gene name is indicated if known. Protein family and/or brief enzyme description are indicated below the arrow for less characterized systems. Four typical functionalities are indicated inside the arrows as follows: R – Endonuclease domain, M – Modification domain, S – Recognition or specificity domain, T – DNA translocation domain (typically a helicase). Homologous domain or proteins are shown by the same color and indicate a protein family sharing phenomenon between different systems. Black rectangle indicates DUF1998 (small metal-binding domain), which is typically a C-terminal domain of a larger protein of unknown function often fused to COG1205 helicase. Question mark indicates uncertainty of functional assignment in the absence of experimental data. Dashed line separates a few recently identified defense systems which have some components shared with known self-nonself recognitions systems, but their mechanism is yet to be characterized. COG (clusters of orthologous groups) and DUF (domain of unknown function) correspond to protein families from COG and PFAM databases respectively. Other abbreviations: PAPS, phosphoadenosine phosphosulfate; HTH, helix turn helix domain; PLD (phospholipase D), PD-(D/E)xK – Nucleases from respective superfamilies.

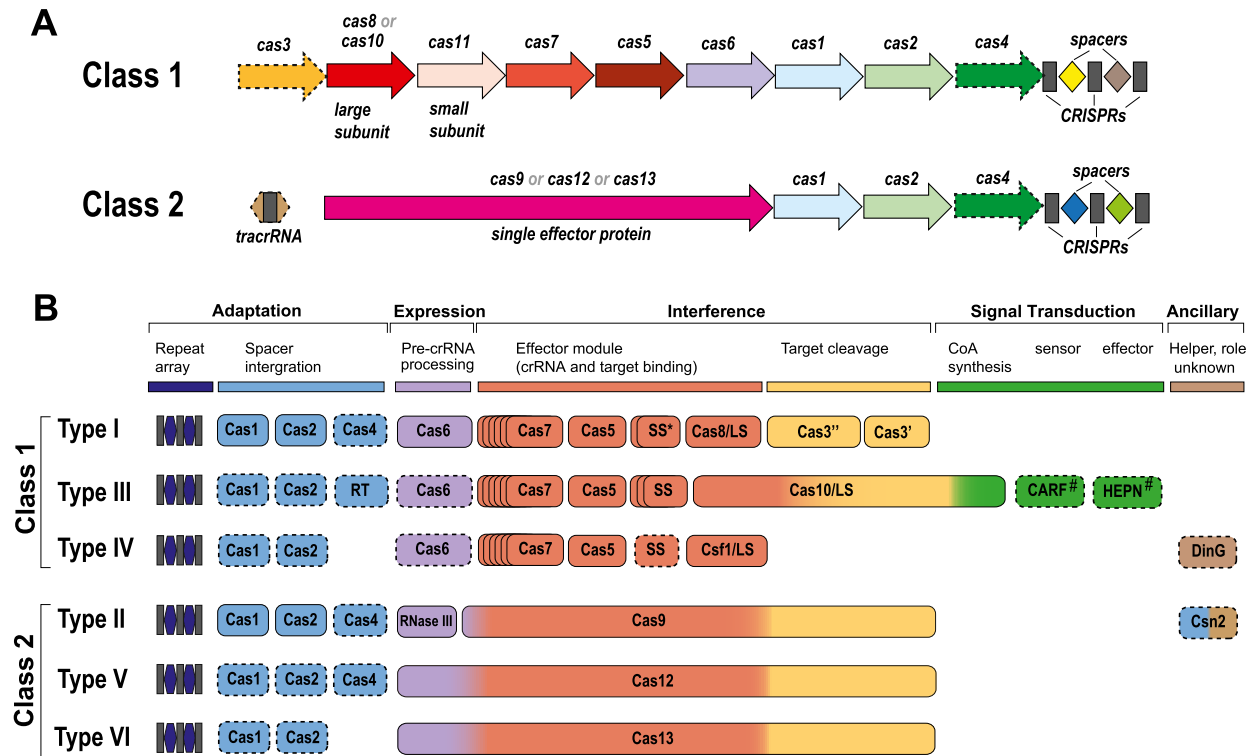


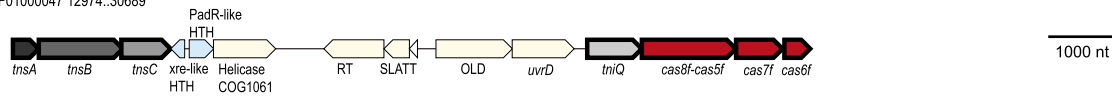
Fig. 2 Core *cas* genes and modular organization of CRISPR-Cas systems. (A). Class 1 and Class 2 CRISPR-Cas systems and 13 core gene families. General architectures of class 1 (multiprotein effector complexes) and class 2 (single-protein effectors) CRISPR-Cas systems is schematically shown. Genes are shown as arrows. Homologous genes are shown by the same color. Gene names follow the current nomenclature and classification. Abbreviation: *tracrRNA*, transactivating crRNA (CRISPR RNA). (B). The middle panel shows the principal building blocks of CRISPR-Cas system types. An asterisk indicates the putative small subunit that might be fused to the large subunit in several Type I subtypes. The dash next to the CARF and HEPN domains indicates that other unknown sensor and effector domains can be involved in the signaling pathway. Dispensable genes are indicated by the dashed outline. The figure is modified from Koonin, E.V., Makarova, K.S., 2019. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B* 374 (1772), 20180087.

effector proteins and complexes are all taken into account. The CRISPR-Cas systems are divided into two classes (Fig. 2(A)). Class 1 includes the variants with multi-subunit effector complexes whereas systems with a single-protein effector (SPE) belong to Class 2. Both Class 1 and Class 2 system effector genes (effector modules) are linked to a CRISPR array containing a variable number of repeats and spacers, and an adaptation module typically consisting of two genes, *cas1* and *cas2* (Fig. 2(B)). Several adaptation modules include an additional gene, *cas4*. *Cas4* protein interacts with *Cas1* and *Cas2* complex and defines several features of spacers that are integrated into CRISPR loci. Class 1 and Class 2 systems are further classified into 6 types and more than 20 subtypes (Fig. 2(B)). A signature protein family is assigned to each subtype. These signature families are *Cas3* (helicase-nuclease), *Cas9* (SPE), *Cas10* (large subunit of effector complex), *Csf1* (large subunit of effector complex), *Cas12* (SPE), and *Cas13* (SPE) for types I, II, III, IV, V and VI, respectively. Despite the extremely low or undetectable sequence similarity between protein subunits, type I, III and IV systems share common principles of effector complex organization and are apparently homologous. Typically, effector complexes of these systems contain one large subunit (*Cas8*, *Cas10* and *Csf1* respectively), several small subunits of the *Cas11* group (which sometimes are fused to large subunits), one *Cas5* family protein and several *Cas7* family proteins (Fig. 2(B)). The *Cas5* subunit binds the 5' handle of the crRNA and interacts with the large subunit. Small subunits interact with the crRNA backbone bound to *Cas7*. The *Cas6* protein is an RNase which is required for crRNA maturation. Type I systems target double-stranded DNA, and the *Cas3* helicase that is typically fused to an HD family nuclease is responsible for the cleavage. Type III systems cleave both DNA, via an HD nuclease that is typically fused to the *Cas10* protein, and RNA via the *Cas7* RNase. Type IV systems are poorly characterized but are predicted not to be active in interference, and accordingly, to function via alternative mechanisms. Multiple additional genes that either help with some steps of CRISPR-Cas function or are involved in regulation and signal transduction constitute the ancillary module and signal transduction modules (Fig. 2(B)). So far only type III systems are known to employ an essential signal transduction pathway in which *Cas10* catalyzes the synthesis of a cyclic oligoadenylate, a signaling molecule that activates the HEPN domain of an ancillary *Cas* protein, a non-specific RNase. This RNase can both degrade viral transcripts and cause PCD or dormancy. To prevent the latter, there are dedicated nucleases that cleave the signaling molecule.

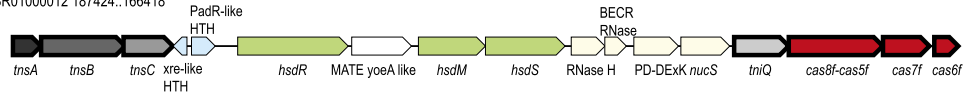
Unlike Class 1, all three distinct types of Class 2 have independent origins although type II and type V share a common domain of RNase H fold (RuvC-like) which is involved in ssDNA nicking. The second domain, HNH family nuclease, in type II SPE (*Cas9*) cleaves the second DNA strand, whereas in type V systems the RuvC like domain cleaves both strands. The crRNA processing in

A

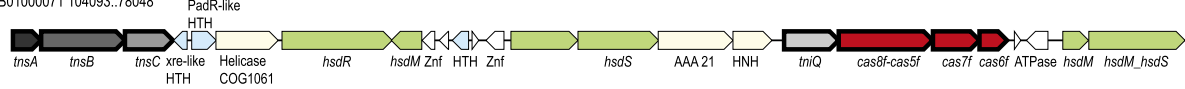
Vibrio campbellii CAIM 519 NBRC 15631
BAOF01000047 12974..30689



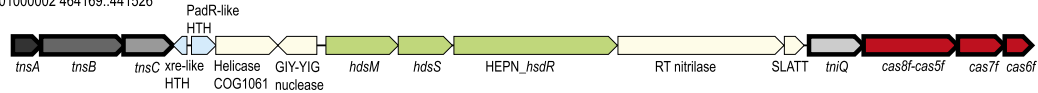
Vibrio cholerae VC35
AMBR01000012 187424..166418



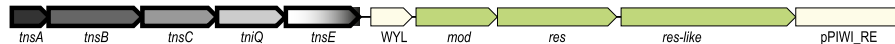
Vibrio crassostreae J5 20
CCKB01000071 104093..78048



Vibrio harveyi ATCC 35084
LRIF01000002 464169..441526

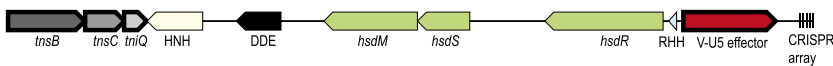


Vibrio parahaemolyticus S023
AWMM01000001 43439..23602

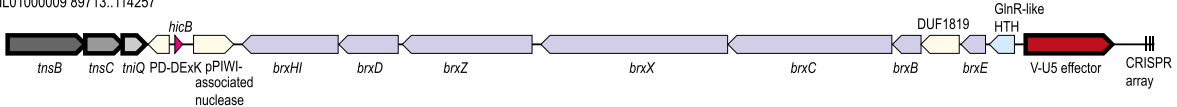


B

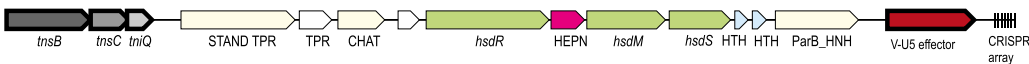
Cyanobacterium PCC 8801
CP001287 3075387..3093284



Pseudanabaena frigida
QBML01000009 89713..114257



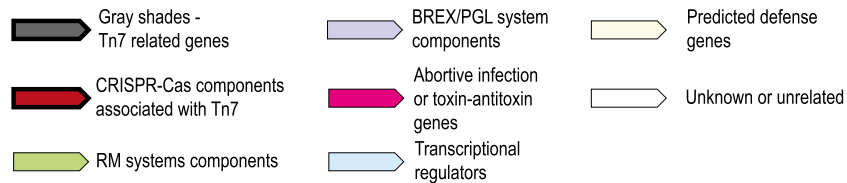
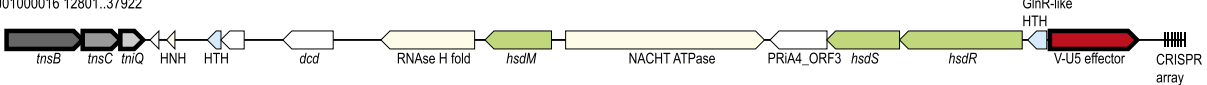
Anabaena cylindrica PCC 7122
CP003659 717365..739767



Cyanobacterium aponinum IPPAS B-1201
PEBC01000041 20555..7088



Nostoc linckia z18
LAHJ01000016 12801..37922



type II and several type V subtypes requires additional host-encoded RNase III and an additional RNA molecule, known as tracrRNA (trans-activating crRNA), that is encoded in the vicinity of the CRISPR-cas loci and apparently evolved by duplication and divergence of a repeat. The tracrRNA remains bound to the corresponding SPE. Several other type V SPEs do not require tracrRNA but contain a dedicated domain involved in pre-crRNA processing. Type V currently includes the largest number of experimentally characterized and predicted subtypes, in particular, small SPEs that are thought to be evolutionary intermediates between non-autonomous transposons encoding RuvC-like domain containing TnpB proteins, but also large, elaborate SPEs such as Cas12a,b,c. Type VI is an RNA targeting system. Cas13, the SPEs of type VI, contains two HEPN domains, involved in RNA cleavage and a domain required for pre-crRNA processing. Once activated by the RNA target recognition, Cas13a becomes a non-specific RNase and might cause dormancy or cell death.

Unclassified Defense Systems

For a number of defense systems too little is known to confidently classify them into any of the categories described above (Table 1). Some of these systems have been experimentally shown to contribute to defense whereas other are only predicted *in silico*. In particular, a dedicated effort on prediction and validation of defense mechanisms has led to the identification of 9 new defense systems. Some of these might be novel self versus nonself recognition systems because they share some components with known defense systems of this type (Fig. 1). For example, all three types of Druantia system (named after a deity from Gallic mythology) contain a COG1205 helicase fused to DUF1998, which is similar to the helicases present in DISARM and DPD systems. Druantia type II system has a DNA methylase of the DCM family similarly to the DISARM II system. Two-component systems consisting of a nuclease and helicase or ABC ATPase, the Hachiman (after Japanese syncretic deity of archery and war) and Lamassu (after Assyrian protective deity) respectively, could be analogs of type IV RM systems (Fig. 1), but currently their mechanisms remain unclear.

Many other defense systems were described by comparative genomics and sequence analysis but remain uncharacterized experimentally. For instance, ATPases of the MORC family are often encoded in the same gene neighborhoods with helicases, nucleases and methyltransferases and might be core proteins of a new class of self versus nonself recognition systems. A distinct subfamily of UvrD-like helicases is linked to Dcm methylases and Cas4-like nucleases and also could be predicted to represent a distinct defense machinery that involves self versus nonself discrimination. Many putative ABI systems were predicted by analysis of defense islands and gene neighborhood of HEPN RNases. Experimental study of these and other predicted defense systems should reveal many new molecular mechanisms of defense.

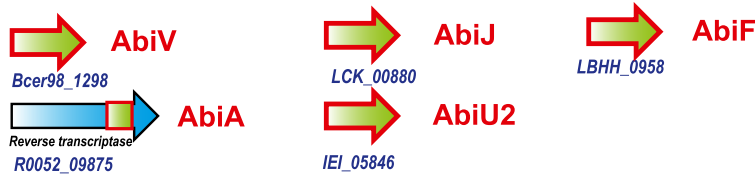
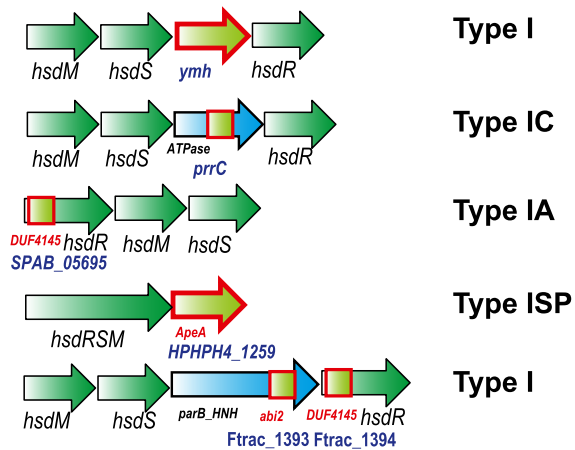
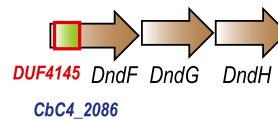
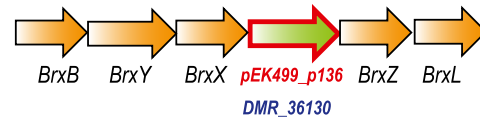
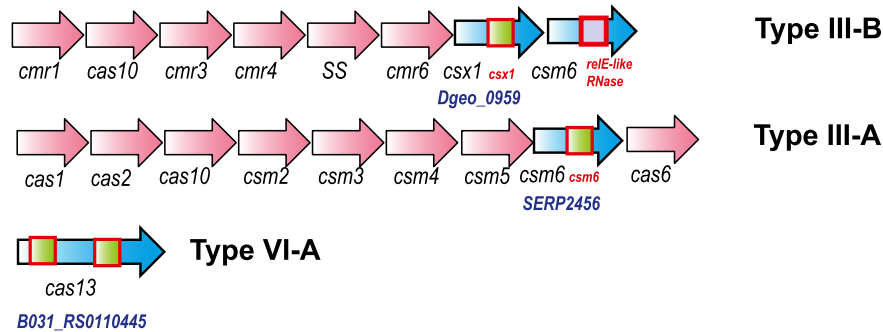
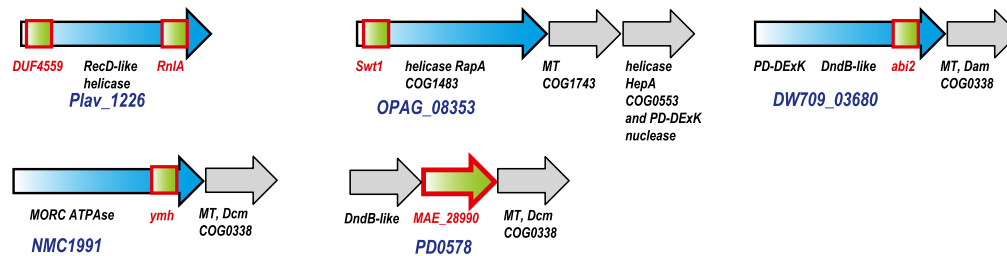
Genomic Organization and Evolution of Defense Systems

Defense Islands and a Tight Link between Defense and Mobilome Genes

Defense genes and typical mobilome components, such as viral, plasmid and transposon genes, show statistically significant clustering in genomic islands. Such islands can be roughly defined as genomic regions flanked by housekeeping genes and containing at least one known defense gene present. Other defense and mobilome genes and uncharacterized genes, if present in the respective genome region, are included in the island. This observation led to the prediction of many new defense genes and systems that are over-represented in such islands. Furthermore, the so-called cargo of many mobile elements includes diverse defense genes that could counter host defense, contribute to competition between mobile elements or serve as addition modules. Defense genes are especially common in some groups of mobile elements, such as Tn7-like transposons and integrative and conjugative elements (ICEs) but, surprisingly, are relatively rare in plasmids and viruses (Fig. 3). It can be argued that all defense genes belong to regions of genome plasticity and, at some point in their evolution, could have been associated with MGE. Not surprisingly, defense genes are more prone to horizontal gene transfer (HGT) than any other functional class of microbial genes except for the mobilome itself, and multiple examples of recent acquisition of defense islands can be observed when closely related genomes are compared.

Remarkably, some of the defense systems were apparently recruited by MGEs to facilitate transposition. The most prominent example is the association of Tn7-like transposons with CRISPR-Cas systems that are competent for precrRNA

Fig. 3 Tn7 cargo genes and recruited CRISPR-Cas systems. (A). Related Tn7 loci linked to CRISPR-Cas system I-F variant. For comparison a related Tn7 locus from *Vibrio parahaemolyticus* S023 without I-F system is shown. (B). Related Tn7 loci linked to CRISPR-Cas system V-U5 type. For each locus, species name, genome accession number and nucleotide coordinates are indicated. The genes of a representative locus are shown by blocked arrows with the scale roughly proportional to gene length and the direction of gene transcription. Genes are color-coded according to the functional groups explained in the inset below. Gene product names are indicated if known, otherwise abbreviated protein family description is provided below or above the respective arrows. see Fig. 1 for innate immunity systems gene names. Abbreviations: dcd, dCTP deaminatase; RT, reverse transcriptase; SLATT, membrane protein frequently associated with immunity systems; TPR, tetratricopeptide repeats containing protein; CHAT, caspase family protease; HTH, helix turn helix; WYL, transcriptional regulator of respective family; Znf, zinc finger domain containing protein; DDE, transposase of respective superfamily; OLD, ATP-dependent DNase of respective family; GIY-YIG, HNH, and PD-D/ExK are distinct DNases families; BECR and HEPN are PCD RNase families.

ABI systems**TA systems****RM systems****DND system****BREX/PGL system****CRISPR-Cas systems****Other putative defense systems**

processing and target binding but not for DNA cleavage that is required for interference. Such CRISPR-Cas systems are hypothesized to guide Tn7 transposition according to the spacer matches which differ from the typical attachment sites recognized by the standard Tn7 transposition machinery. Fig. 3 shows gene organization of closely related Tn7-like elements associated with either CRISPR-Cas I-F (Fig. 3(A)) or V-U5 (Fig. 3(B)) systems but carrying different defense genes as a cargo. Some MGE mobilization genes have been also recruited by some defense systems. For instance, *cas1*, the signature gene of CRISPR-Cas system, likely originated from a distinct group of self-replicating MGEs called casposons and RAG1, the key enzyme of V(D)J recombination of the Ig-based adaptive immunity system in vertebrates, is derived from a eukaryotic transposon of the Transib family.

Gene and Domain Shuffling and Sharing

In addition to being HGT attractors, defense systems and islands are hot spots for non-homologous recombination. This phenomenon is well-studied for the type I RM systems and is often referred to as phase variation. The XerC-like recombinase IvrR is responsible for shuffling between the target recognition domains of the *hdsS* genes, often an *ivrR* gene, and either inverted or direct repeats. Another common mechanism mediating phase variation is DNA polymerase slippage that occurs in simple sequence repeat regions in the TRD region of the Res subunit of type III RM systems. Most likely, these or mechanistically similar processes are also involved in shuffling of domains, individual genes and gene modules in other defense systems. The described examples include restriction endonuclease domain shuffling in RM systems, switching between toxin and antitoxin in toxin-antitoxin systems and effector module shuffling in CRISPR-Cas systems.

Gene sharing is a less thoroughly understood phenomenon. It has been noticed that some superfamily II helicases, especially, COG1205 and COG0553, and ParB family proteins are abundant in defense islands but seemingly do not belong to a single conserved gene neighborhood. Fig. 1 shows that COG1205 helicase and the accompanying DUF1998 domain containing protein belong to core genes sets in DISARM, DPD and Druantia systems. The PLD domain containing nuclease, often encoded next to the COG1205 helicase and the DUF1998 protein, is present in both DISARM and DPD. Despite this link between COG1205 helicase and PLD nuclease in DISARM system, the former is essential for the protection from the phage infection, whereas the latter is not. The COG1205 helicase is also essential in type I Druantia system along with other four genes. Thus, this helicase appears to be an integral component of some defense systems and an ancillary gene in others. Further research is needed to determine whether these genes are essential in other systems and to decipher their specific roles.

Association of Defense Systems With Programmed Cell Death Components

Sequence and gene neighborhood analyzes have identified many domains and genes homologous to RNase toxins of TA and ABI systems in the neighborhoods of RM, CRISPR and other defense systems (Fig. 4). Many of these RNases belong to the HEPN domain superfamily. Only for a few of them, the role in defense is at least partially characterized. The best-studied example is the association of *prnC* anticodon nuclease gene with the type IC RM system. The PrrC proteins contains a HEPN domain and contributes to the phage T4 exclusion mechanism. Inactive PrrC is a subunit of the HsdSMR complex, but it can be allosterically activated by increased levels of dTTP or by the small anti-restriction peptide encoded by the T4 phage. Activated PrrC cleaves the anticodon of tRNA^{lys}. Phage T4 encodes an RNA repair system that consists of the Pnl and Rnl1 proteins and can reverse the damage inflicted by PrrC.

The second example of the defense-PCD association involves another HEPN domain containing protein that is associated with many type III CRISPR-Cas systems. In these proteins, the HEPN domain is fused to a CARF (CRISPR-associated Rossmann fold) domain and is allosterically activated through the binding of a cyclic oligoadenylate that synthesized by the Cas10 protein by the CARF domain. Once activated, HEPN cleaves RNA in a non-specific manner. This activation is reversible because cyclic oligoadenylate messengers are hydrolyzed by dedicated nucleases. Furthermore, as mentioned above, effector proteins of type VI CRISPR-Cas systems become non-specific RNases upon binding the crRNA and the cognate RNA target.

There are two hypotheses explaining the coupling of defense and PCD systems: first, dormancy induction allows the host to “buy time” for the activation of other defense systems, in particular, adaptive immunity; second, dormancy or ‘altruistic’ suicide can be the ‘last resort’ measure to prevent viral spread if other immunity mechanisms fail.

Fig. 4 HEPN domain containing PCD/dormancy RNases linked to various defense systems. The genes are shown as colored arrows. Gene names (if known) are indicated below respective arrows. Colors correspond to distinct defense system groups, which are circles by dashed lines. Multidomain proteins are colored blue, known domains are indicated below the arrows. The HEPN domain is shown by a light green shape with a red outline. Distinct families of HEPN superfamily are indicated on the right for abortive infection systems and below the respective shapes in red font for other systems according to Anantharaman *et al.* (see Further Reading list). For each HEPN domain containing protein locus tag of respective protein is provided. The only other PCD related RNase of RelE family is shown by purple shape with red outline. Abbreviations: HEPN, higher eukaryotes and prokaryotes nucleotide-binding domain, predicted endoribonuclease; MNT, minimal nucleotidyltransferase; HNH, PD-(D/E)xK, DEDD are nucleases from respective superfamilies. RM and CRISPR-Cas gene names and types follow current nomenclature and classification.

Concluding Remarks

Despite the notable progress during the last decades, many aspects of the function, distribution and evolution of prokaryotic defense systems remain poorly understood. One of the deepest mysteries is the enormous variety of defense systems, both across the prokaryotic world and within most individual genomes. At least for self versus nonself recognizing systems, the need to recognize diverse signals could be a plausible explanation (case in point: RM systems) although the existence of separate recognition modules might mitigate this demand. The enormous abundance of PCD/dormancy systems (up to hundreds of TA and ABI modules in a single genome) suggests important roles of processes other than selection for the benefit of the host in the evolutionary dynamics of these systems. Carrying such a large number and variety of PCD/dormancy modules poses a problem for the host cell, namely, the need for regulation that is robust to adversarial tampering but has a low enough spontaneous trigger rate to preclude excessive burdening for the host. Currently, there is scant information about the regulation of TA systems except for the notion that they are upregulated by general stress and DNA damage pathways. It appears likely that the enormous abundance and wide spread of defense systems including those that mediate PCD in the prokaryotic world is, in large part, due to the fact that they themselves behave as mobile elements. The strong link between many innate immunity systems and PCD, especially, through the HEPN superfamily of RNases, clearly points to functional coupling. The mechanisms of such coupling are a major area of future research.

A notable feature of microbial defense systems is domain sharing. In particular, helicase domains are commonly shared and often essential components of these systems. An interesting possibility is that these DNA-dependent ATPases act as translocases rather than helicases, facilitating genome scanning and transport of the defense module along the DNA. This hypothesis remains to be tested experimentally.

Finally, it is important to note that a large part of the toolbox of modern molecular biology and biotechnology consists of components of microbial defense systems, in particular, those that discriminate between self and nonself, namely, RM and CRISPR-Cas. Their ability to recognize specific signals in the genome and to deliver the effector machinery to the targeted loci proved invaluable for genome editing methods. The future research in microbial defense can be expected to bring forth new discoveries and, with them, new tools for practical use.

Further Reading

- Anantharaman, V., Makarova, K.S., Burroughs, A.M., Koonin, E.V., Aravind, L., 2013. Comprehensive analysis of the HEPN superfamily: Identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biology Direct* 8, 15.
- Doron, S., Melamed, S., Ofir, G., *et al.*, 2018. Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359 (6379).
- Faure, G., Shmakov, S., Yan, W.X., *et al.*, 2019. CRISPR-Cas in mobile genetic elements: Counter-defense and beyond. *Nature Reviews Microbiology* 17 (8), 513–525.
- Harms, A., Brodersen, D.E., Mitarai, N., Gerdes, K., 2018. Toxins, targets, and triggers: An overview of toxin-antitoxin biology. *Molecular Cell* 70 (5), 768–784.
- Jangam, D., Feschotte, C., Betran, E., 2017. Transposable element domestication as an adaptation to evolutionary conflicts. *Trends in Genetics* 33 (11), 817–831.
- Knott, G.J., Doudna, J.A., 2018. CRISPR-Cas guides the future of genetic engineering. *Science* 361 (6405), 866–869.
- Koonin, E.V., Makarova, K.S., 2019. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B* 374 (1772), 20180087.
- Kwon, M.J., Oggioni, M.R., De Ste Croix, M., Bentley, S.D., Croucher, N.J., 2018. Excision-reintegration at a pneumococcal phase-variable restriction-modification locus drives within- and between-strain epigenetic differentiation and inhibits gene acquisition. *Nucleic Acids Research* 46 (21), 11438–11453.
- Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2013. Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Research* 41 (8), 4360–4377.
- Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2018. Classification and nomenclature of CRISPR-Cas systems: Where from here? *The CRISPR Journal* 1 (5), 325–336.
- Roberts, R.J., Vincze, T., Posfai, J., Macelis, D., 2007. REBASE—enzymes and genes for DNA restriction and modification. *Nucleic Acids Research* 35 (Database issue), D269–D270.
- Rostol, J.T., Marraffini, L., 2019. (Ph)ighting phages: How bacteria resist their parasites. *Cell Host Microbe* 25 (2), 184–194.
- Swarts, D.C., Makarova, K., Wang, Y., *et al.*, 2014. The evolutionary journey of Argonaute proteins. *Nature Structural & Molecular Biology* 21 (9), 743–753.
- Tong, T., Chen, S., Wang, L., *et al.*, 2018. Occurrence, evolution, and functions of DNA phosphorothioate epigenetics in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 115 (13), E2988–E2996.
- van Valen, L., 1973. A new evolutionary law. *Evolutionary Theory* 1 (1), 1–30.

Relevant Websites

- <https://ncbi.nlm.nih.gov/COG/>
COG.
- <https://crispr.i2bc.paris-saclay.fr/>
CRISPR.
- <https://pfam.xfam.org/>
PFAM.
- <http://rebase.neb.com/rebase/rebase.html>
ReBase – NEB.

Defective-Interfering Viruses

L Roux, University of Geneva Medical School, Geneva, Switzerland

© 2008 Elsevier Ltd. All rights reserved

This is a reproduction of L. Roux, Defective-Interfering Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00379-4.

History

In 1943, Henle and Henle reported the decreased infectivity for mice of influenza virus stocks obtained after a series of undiluted passages in embryonated chicken eggs. In the early 1950s, von Magnus showed that such undiluted passages generate incomplete virus particles capable of limiting the growth of infectious virus (hence exhibiting interference). This first characterization was soon followed by similar reports by Mims, on the one hand, and Cooper and Bellet, on the other, dealing respectively with Rift Valley fever virus and vesicular stomatitis virus (VSV). In the late 1950s, Cooper and Bellet went so far as to assign interference to sedimentable particles, but failed to identify them as antigenically related to VSV. From the mid-1960s on, the characterization of other positive- and negative-stranded RNA virus defective particles continued. In 1970, a review by A. Huang and D. Baltimore set the basic definition of defective interfering (DI) particles and emphasized their widespread occurrence. Since then, DI particles have been described for almost all the known DNA and RNA viruses, including plant and even fungal viruses.

Structure

DI particles have the same protein composition as their homologous nondefective 'parents', often called St. However, they differ from the St particles in the primary structure of their genome. As emphasized later, DI genomes lack part of the genetic information. They may or may not serve as coding sequences. However, they always conserve the *cis*-sequences needed for replication initiation (origins of replication), sometimes present in more than one copy, and sequences involved in encapsidation. Foreign sequences can also be inserted. DI particles can sometimes be separated from St particles on the basis of size, when the size of the particle closely corresponds to the size of the genome (for instance rhabdovirus), or on the basis of particle density differences (changes in nucleic acid to protein ratios). Often, however, only viral stocks enriched in DI particles are available owing to the size heterogeneity of the virus particles.

Generation of DI Genomes

DI DNAs very likely arise from various recombinational events not necessarily linked to genome replication, and which result in deletion, tandem duplication, insertion of host DNA, and polymerization of small monomer sequences. DI RNAs have been proposed to arise almost exclusively during genome replication by a mechanism of 'leaping polymerase' consisting of polymerase stop/fall-off or slippage/reinitiation events (Fig. 1). In this model the replicase complex moves with the nascent RNA still attached to it. Depending on where reinitiation takes place, and on the number and the direction of the leaps, the resulting molecules can be of the copy-back type, with more or less intramolecular inverted complementary sequences (a), of the internal deletion type (b), and of the duplication type (c). Multisteps (b) or (c) and combinations of steps (b) and (c) can, moreover, lead to various mosaic types. Insertion of host RNA is also observed, especially in plant DI RNA. The frequency at which the polymerase leaps and resumes its synthesis is unknown. The probability for this exercise to be successful in producing a viable DI genome has been estimated for VSV to range in the order of 10^{-7} – 10^{-8} per genome replication.

Defectiveness

The DI genomes contain interrupted or rearranged open reading frames. They partly or completely lack the full coding capacity of the viral genome. They are therefore defective, and depend for their replication and for their propagation (formation of virus particles) on the functions provided by the homologous standard virus (helper). Co-infection of cells with DI and standard particles is therefore essential for DI particle multiplication. Consequently, low-multiplicity infections, and particularly plaque purification, represent conditions which decrease, and potentially eliminate, DI particles from a viral preparation.

Interference

As stated earlier, the generation of a defective genome is likely to represent a rare event. This event would never be seen unless it was successfully amplified. During this amplification step the defective genome is preferentially replicated over the nondefective genome. This ability to replicate efficiently at the expense of the nondefective genome is called interference. The mechanisms of

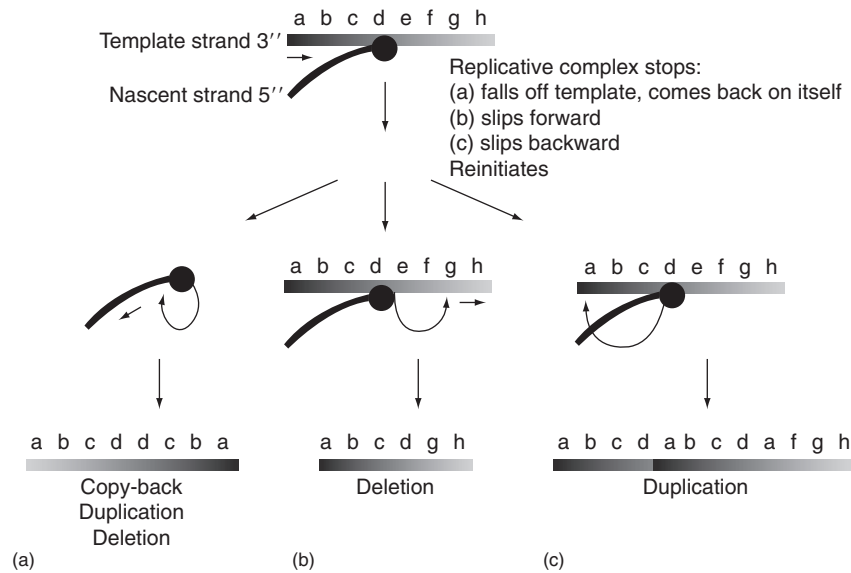


Fig. 1 Defective RNA genome generation.

interference are not completely understood. They obviously change depending on the specificity of the viruses, and appear to be also affected by the host cell types. In general, interference involves an early step in genome replication, and can be pictured as a competition for limiting replication 'factors' (viral replicase, encapsidation proteins, host cell factors). Reiterated origins of replication or encapsidation sites on DI DNAs, presence of higher affinity sites for the replicase or for the encapsidation on both positive and negative polarity DI RNAs, and shorter length of the replicating units, higher availability for replication of molecules not involved in transcription, have been shown or postulated as taking part in the interference mechanism.

Defective Interfering versus Defective Viruses

Based on the outcome of experimental co-infections of defective with nondefective viruses, a distinction has been made between defective interfering or defective noninterfering particles, according to the ability of the defective viruses to selectively restrict nondefective virus replication. This distinction may not apply during the first events following the generation of a defective genome. As this is bound to be a rare event, an interference mechanism has to be invoked any time this defective genome is amplified to the point it can be detected, or become predominant.

Cyclic Variations of Defective Interference

The dependence of DI genome replication on functions provided by the nondefective genome on the one hand, and the interference exhibited by DI genome on the other hand, result in out-of-phase cyclic variations of both DI and St genome replication. As illustrated in **Fig. 2**, efficient St genome replication must precede extensive DI genome replication. This in turn establishes conditions of high interference which results in inhibition of St genome replication. Decrease of helper function availability leads to DI genome replication dampening, and therefore to release from interference, allowing efficient St genome replication to resume. These cyclic variations have been observed in serial passages of St and DI viruses in cell culture, as well as in persistent infections. The periodicity of a complete cycle is generally a matter of days or of a few serial passages.

Assay for DI Particles

DI particles can be detected by physical separation on velocity or density gradients when applicable (see the section titled 'Structure'). The presence of subgenomic nucleic acids in viral stocks or in infected cells (distinct from viral messengers) can also be diagnostic. The ability to decrease the infectivity of a viral stock or to protect infected cells from the lytic infectious virus (see the section titled, 'Biological effects') are used in various biological assays to estimate quantitatively and qualitatively the DI particle composition of a viral stock. These assays, although appropriate to characterize DI particle preparations, are generally not sensitive enough to exclude, when negative, the presence of DI particles in a viral preparation. The test that still remains the most dependable to assess presence or

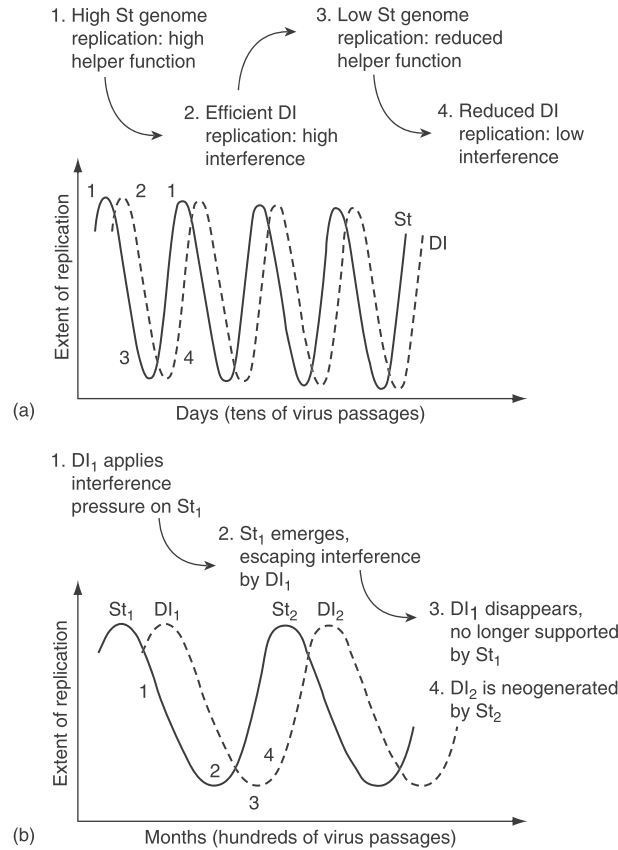


Fig. 2 Cyclic variation of DI replication in (a) days, and (b) months.

absence of DI particles consists of multiple independent serial undiluted passages. It is based on the observation that a viral stock contaminated with an undetectable amount of DI particles will, on subsequent independent serial passages, promote in each series amplification of the same contaminating DI particles. A DI particle-free stock, on the other hand, will yield in each series different DI particles or different sets of DI particles.

Biological Effects

DI particles have been shown to modulate the course of an infection. In cell culture, attenuation of the cytopathic effects is the most frequently described, and DI particles can promote cell survival and establishment of persistent infections. As far as negative-stranded RNA viruses are concerned, copy-back DI RNAs appear to prevent induction of apoptosis through a mechanism which has still to be unravelled, in which a certain category of small leader RNAs may participate. A possible role of the matrix M protein, the concentration of which is decreased to the point where viral assembly and budding at the cell surface is highly diminished in the presence of DI genomes, has also been considered. DI ability selectively allows emergence of St virus (St₂) which escapes interference ([Fig. 2\(b\)](#)). St₂, resistant to interference, is selectively amplified over St₁, still sensitive to interference, and soon becomes predominant. It loses its ability to support DI₁, which therefore disappears. St₂ will generate its own DI₂, which in turn will favor emergence of a new St variant. Thus, DI viruses serve as mutational drivers favoring virus evolution, through cycles of high and low interference whose periodicity is this time measured in months or in hundreds of viral passages (compare [Fig. 2\(a\)](#) and [\(b\)](#)).

DI Particles in Experimental Animals

DI particles are generated and amplified in the whole animal, as well as in cell culture. They can change the symptoms of viral infection from rapid death to slow, progressing paralysis. They can sometimes fully protect the animals from an otherwise lethal infection. Interference is likely to be involved in this modulation of symptoms, but other phenomena like increased interferon induction and immune response modulation are reported.

DI Particles in Natural Infections

Involvement of DI particles in natural infections is still poorly documented. This is partly because the experimental results supporting the strong potential for infection modulation of DI particles have not been fully recognized, as detection of DI particles in natural infections is not straightforward. Unpredictable cyclic variations in DI replication, efficiency of DI replication changing with the types of infected tissues, and potent interfering ability associated with poor DI particle replication are all factors which undoubtedly make DI particle detection difficult *in vivo*. Last but not least, virus isolation, which is used to characterize the infectious agent associated with a disease, often represents conditions (low multiplicity of infection) known to impair DI particle replication drastically.

Nevertheless, association of a chicken influenza virus strain, efficiently producing DI particles, with an epidemic of low morbidity and low mortality, and conversely, a high-mortality epidemic associated with a strain free of DI particles, have been reported. Murine and feline leukemia virus strains causing immune deficiency syndromes are shown to contain predominantly replication-defective viral genomes before onset and during the development of the disease. The pathogenicity of some bovine and swine pestiviruses has clearly been associated with presence of DI RNAs in the animals. For the bovine viral diarrhea virus (BVDV), a pestivirus of the same family as hepatitis C virus, the presence of a particular DI RNA can turn noncytopathic virus into a fatal infectious agent. In plants, at least three examples of DI RNAs are described to be involved in infection modulation. Interestingly, depending on the types of viruses, DI RNAs can either attenuate or exacerbate the symptoms.

DI RNAs are identified in stool and blood samples of humans suffering from hepatitis A virus, an infection known to be rather moderate and prolonged. DI particles are identified in measles virus-attenuated vaccine preparations which have been, and are being, widely and successfully used (raising the question of DI particle participation in vaccine attenuation). Measles viruses defective in viral assembly are currently found associated with human subacute sclerosing panencephalitis (SSPE). The brain cells of SSPE patients were, moreover, shown to harbor many species of measles virus copy-back DI RNAs. Direct amplification of a portion of the HIV *tat* gene from infected patients demonstrates that about a third of the sequences correspond to defective *tat* function. Moreover, human immunodeficiency virus (HIV)-1 sequences isolated from a cohort of six blood or blood product recipients infected with one donor all contained a similar deletion in the *nef* gene. Remarkably, all the patients harboring this deleted viral genome remained free of HIV-related diseases 12–16 years after infection, suggesting that this defective species of HIV genome is responsible for this decreased pathology. Epstein–Barr virus (EBV) replicative infections developing in human epithelial lesions involve a deleted rearranged form of EBV DNA (het DNA). This het DNA is associated, in experimental infections, with disruption of latency and persistent productive infections. Specific identification of viral hepatitis B (HBV) genomes containing an interrupted precore antigen (HBcAg) coding sequence in patients dying from fulminant hepatitis suggests that such defective genomes may be responsible for the exacerbation of the disease. This contrasts with more recent data reporting experimental evidence for the existence of DI-like viruses in HBV human chronic carriers; fluctuations between these naturally occurring core internal deletion variants and helper HBV in three chronic carriers were reminiscent of the cycling phenomenon in other DI viral systems.

Future Perspectives

Defective interfering particles are ubiquitous in the realm of animal and plant viruses. In experimental conditions they appear as necessary companions of their nondefective homologs. Capable of affecting the extent of viral growth, the course of viral infections, and serving as selective pressure to drive mutational changes, they can be seen as natural regulators of virus evolution. The demonstration of their participation in natural infections, and of their ability to affect the course of diseases, constitutes a challenge for the years to come. As pointed out by the few examples listed earlier, their direct detection in infected tissues will certainly be needed to assert their involvement in natural infections. The availability of sensitive detection techniques (like polymerase chain reactions), allowing direct observation of viral genomes without the possible distortion of virus isolation, bears great hope. More than giving increased insights into the physiopathology of viral infections, in times where the modifications of the viral genomes represent an imperative step in the generation of viral recombinant vaccines or of appropriate vectors for gene targeting, DI viral genomes represent natural versions of defective genomes that can serve as model tools for creation of more adapted vectors.

See also: Orthobunyaviruses (*Peribunyaviridae*)

Further Reading

- Barrett, A.D.T., Dimmock, N.J., 1986. Defective interfering viruses and infection of animals. *Current Topics in Microbiology and Immunology* 128, 55.
- Perrault, J., 1981. Origin and replication of defective interfering particles. *Current Topics in Microbiology and Immunology* 93, 151.
- Roux, L., Simon, A.E., Holland, J.J., 1991. Effects of defective interfering viruses on virus replication and pathogenesis. In: Shatkin, A. (Ed.), *Advances in Virus Research*, vol. 40. New York: Academic Press, p. 181.

Ecology and Global Impacts of Viruses

Joanne B Emerson, University of California, Davis, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Auxiliary metabolic genes (AMGs) Genes encoded within viral genomes that potentially function in host metabolic processes.

Lysis Bursting of host cells upon the release of progeny virions after viral infection and replication.

Lysogeny Maintenance of the viral genome within the host, typically either as an inserted “prophage” within the host genome or as an extrachromosomal element similar to a plasmid.

Metagenomics Sequencing and analysis of community DNA, including all genes and all biota from which DNA is extracted from a given sample.

Mycoviruses Viruses that infect fungi.

Temperate Viruses capable of lysogeny or lysis.

Virocell A cell infected by an actively replicating virus; the physiology of a virocell is distinct from an uninfected cell, as metabolism shifts to support viral replication.

Viromics Metagenomics (community DNA sequencing) of the purified viral size-fraction.

Introduction

As members of Earth’s microbiome, viruses exist in consortia that include diverse organisms and other viruses. Even disease-causing viruses of humans, animals, and plants are surrounded by other biota. The focus of this article is on introducing the reader to viruses as members of microbial ecosystems, including the terminology and methods used to study diverse viral communities, the hosts that these viruses can infect, the replication strategies of viruses that infect microorganisms, and the impacts of viruses on microorganisms and the environment. Together, these topics comprise the field of viral ecology, which is the study of viruses as members of ecosystems, including viral diversity and viral interactions with other organisms and the environment.

Depending on the environment, a suite of hosts may exist for the viruses that are present. For example, in soil, hosts could include bacteria, archaea, fungi, protists, plants, insects, and more. In the open ocean, most hosts are likely to be bacteria and other microorganisms, but host-associated viruses are also present in fish, whales, and other creatures. In the human body, there may sometimes be viruses that infect us, like the influenza virus (*Orthomyxoviridae*) or rhinoviruses (*Picornaviridae*) that cause the common cold, but there are also numerous viruses that infect the bacteria in and on our bodies, particularly in our guts. In each of these environments, many viruses co-exist with each other and with a suite of diverse organisms, some of which may be potential hosts. Understanding the different roles that viruses play in these complex ecosystems is one of the main goals of the field of viral ecology.

In addition to the diversity of potential hosts in a given environment, viruses themselves are incredibly genetically and biochemically diverse, which poses a challenge to studying all of the viruses present in a given ecosystem at the same time. Viruses can have DNA or RNA genomes that can be single-stranded or double-stranded, and their genomes can span a wide range of sizes, from just over 1000 nucleotides for some of the smallest circoviruses to 2.5 million base-pairs for some members of the Megavirales group of giant viruses (nucleocytoplasmic large DNA viruses, NCLDV) that infect amoebae. Some viral genomes are spread across multiple segments, while others are contained on a single segment of DNA or RNA. Viruses can be encapsidated within a single protein coat (capsid), or they can be multipartite, with different parts of the genome in multiple capsids. Some viruses are also surrounded by a lipid envelope. Though most viruses can exist as virions in a free particle state in the open environment, some viruses, for example, most mycoviruses (viruses of fungi), are thought to be obligately intracellular and only transmitted vertically from parent organisms to progeny and/or through cell-cell fusion. With all of these different types of viruses, hosts, genome types, replication strategies, and physicochemical forms, it is currently impossible to study all of the viruses in a given ecosystem with a single method. Still, methods for studying viral ecology have vastly improved over the last two decades, and we are getting closer to achieving relatively complete characterizations of viral diversity across environments.

As a field, viral ecology shares some terminology and statistical methodologies with other branches of ecology, particularly with microbial ecology, but there are some important differences. For example, viral population ecology is the study of the diversity, interactions, and processes within a viral population (e.g., growth rates, burst sizes – the number of progeny viruses per infected cell – and controls on these processes), and the study of viral populations is inherently tied to considerations of host populations, as hosts are required for replication. For cellular organisms, populations are typically defined by species boundaries, but there is no simple, universal definition of a viral “species”, nor are there conserved genetic markers common to all viruses, so identifying a meaningful unit of measure to represent a viral population is difficult. And yet, we need to be able to categorize and count viruses within defined groups (populations) in order to take advantage of statistical ecological approaches to compare viral communities. Therefore, a workable definition of a viral operational taxonomic unit (vOTU), or population, is emerging, at least for bacteriophages: genomes and recoverable genome fragments belonging to the same vOTU share an average nucleotide identity (ANI) $\geq 95\%$ along their lengths.

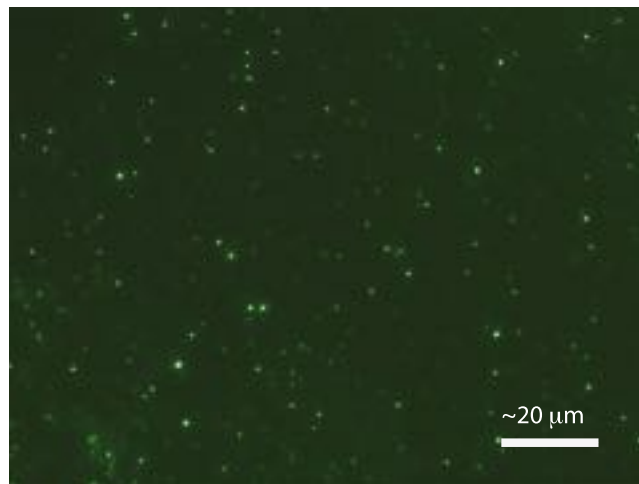


Fig. 1 Epifluorescence micrograph (EFM) of purified viral particles from agricultural soils. 1 ml of post-0.2 μm filtered soil slurry in potassium citrate buffer was placed onto a Whatman 0.02 μm Anodisc filter via vacuum filtration and stained with SYBR Gold. Bright dots are virus-like particles. Image courtesy of Rose Bolle, color-adjusted by JBE for print clarity.

Viral community ecology is a term typically used to mean the study of all viruses recoverable by a given method in a given ecosystem, which is counter to the standard ecological definition of a community, which would include all organisms in a given environment or ecosystem. For simplicity, the viral fraction of an ecosystem is often called a viral community or the virosphere, and, in accordance with field jargon, a viral community in this article includes all of the vOTUs recovered in a given ecosystem by a given method. Similarly, viral community ecology refers to the study of these viruses and their interactions with other organisms and/or the environment. This article will predominantly focus on the emerging field of viral community ecology, which builds on previous studies of single-virus and viral population ecology.

Methods for Studying Viral Community Ecology

Viral communities can be studied with varying levels of detail, depending on the methods used. For many years, continuing today, various microscopic techniques have been used to count the numbers and/or morphological types of viruses in a given community. Transmission electron microscopy (TEM) can resolve viral shapes and can therefore be a useful measure of the numbers of different morphological types of viruses in a given environment. Epifluorescence microscopy (EFM) generally requires adding a fluorescent nucleic acid stain that binds to viral genomes so that the viruses light up under the microscope (Fig. 1). EFM can be useful for counting viruses (or, more accurately, virus-like particles), which appear as bright pinprick dots under the microscope, due to excitation of the nucleic acid stain that emits light. Both of these microscopic approaches can be used to estimate the number of viruses in a sample, which can be extrapolated to estimate the number of viruses in a community or ecosystem. Based on these approaches, the total number of viruses on Earth has been estimated to be approximately 10^{31} , or more than one billion times the number of stars in the Milky Way Galaxy. There are approximately 10^7 viral particles per milliliter of ocean water, and there are approximately 10^7 – 10^{10} viral particles per gram of soil. These counts are based on EFM, which is generally not an effective method for counting viruses with RNA genomes because RNA viruses tend to have smaller genomes that do not stain well, so these approximations, largely based on DNA viruses, may underestimate true, comprehensive viral abundances.

Although microscopy is useful for learning how many viruses are in a sample and for comparing total numbers of viruses among samples, genetic and genomic differences among viruses cannot be determined through microscopy. Some resolution into genomic differences among viral populations has historically been attempted through pulsed-field gel electrophoresis (PFGE), which can distinguish among groups of viruses based on differences in their genome sizes. In this way, a researcher can tell whether a community of viruses from one sample is different from a community of viruses from another sample, based on a “fingerprint” of the genome sizes in each sample. However, many genetically distinct viruses have similar genome sizes and will therefore appear to be the same on a PFGE gel image. So, although PFGE provides slightly more resolution into viral community composition than microscopy, it is still not possible to distinguish among the hundreds or thousands of different viral populations in a given sample through PFGE.

As a further step towards increasing resolution of genetically distinct viral populations, amplicon-based community sequencing can be used to identify specific groups of viruses that share a conserved gene (for example, the gene encoding the g23 major capsid protein of T4-like bacteriophages). This allows for very high resolution into the specific viral populations that contain a specific gene of interest in a given environment. Similar to approaches for characterizing bacterial and archaeal community composition through 16S rRNA gene amplicon sequencing, these specific viral genes can be amplified directly from DNA extracted from the

environment through polymerase chain reaction (PCR). The amplified PCR products contain DNA sequences from each viral population that has this gene in a given sample, thus these PCR products can be sequenced to determine viral population diversity and the relative abundances of each viral population in each sample. Comparing all viral populations that have the gene of interest across samples is also possible via this approach, so this amplicon sequencing technique can provide some insights into viral community ecology, but only for a specific viral group with the gene of interest. For these specific viruses, this may be an excellent approach. However, there is no universal marker gene for viruses, so it is impossible to capture all of the viruses in a given ecosystem through amplicon-based sequencing.

Shotgun metagenomics, or sequencing all of the DNA in a sample, is an approach that comes close to identifying most of the viruses with DNA genomes in a given community. Typically, viral particles are purified from the environment based on their relatively small size through filtration (e.g., by passing a liquid sample through a 0.2 μm filter that removes most bacteria and larger organisms). Filtration is relatively straightforward for aqueous samples, but it generally requires the addition of a buffer and some pre-processing to remove solids from soil or sediment samples. Once the viral particles have been purified from other organisms, the viral community DNA can be extracted and sequenced, and the recovered viral genomes can be reconstructed on the computer via bioinformatics. This approach – purifying viral particles and sequencing and analyzing the viral metagenome – is called viromics. To date, viromic approaches have been used to recover hundreds of thousands of ocean viral populations and thousands of soil viral populations, and the number of viral populations that we are able to recover via viromics continues to increase, such that the actual viral diversity of these ecosystems remains to be seen.

Even with viromics, we cannot study all viruses in a given ecosystem because not all viruses have DNA genomes, and viromics only recovers viruses with DNA genomes. In addition, by purifying viral particles prior to metagenomics, we are only considering viruses that have (and, at the time of sampling, are in) a free viral particle phase, so we will miss viruses that are: integrated in host genomes (see lysogeny in the glossary and later in this article), undergoing active infection and replication inside a host cell, or never have a free viral particle phase in the open environment. Attempts to recover these other viral genome types and replication stages are being made via metatranscriptomics (recovering all of the RNA in a sample) and via bioinformatic mining of viral genomic signatures from total metagenomes. For example, instead of purifying viral particles, shotgun metagenomics or metatranscriptomics can be performed directly on an environmental sample, such that most sequences will be recovered from the dominant organisms in the sample (generally bacteria for marine and soil samples). In the remaining sequences, we can predict genes and proteins to look for viral hallmarks, like capsid proteins, to identify viruses that may not have been recovered in purified viral metagenomes (viromes). Together with long-read sequencing approaches that are beginning to recover complete and near-complete viral genomes in a single sequence, our ever-improving laboratory and bioinformatics approaches to viral ecology are allowing us to access more and more populations and replication stages within each viral and microbial community.

Virus-Host Linkages and Ecology

Most of the subsequent sections in this article will focus on viruses of bacteria (bacteriophages) because our existing viromic methods are best suited to recovering bacteriophages with double-stranded DNA genomes, so we have the most information about bacteriophage ecology. However, as methods continue to improve, we are beginning to uncover diverse viruses in nature that infect other organisms, including eukaryotes. The vast majority of plant and fungal viruses have RNA genomes, so our metagenomic and viromic approaches will not recover these viruses, but metatranscriptomic (shotgun RNA sequencing) approaches have shown that diverse RNA viruses are present in the environment. For example, diverse RNA viruses have been recovered from marine systems and soil, including many putative soil mycoviruses that have vastly expanded the known genomic diversity of mycoviruses. It remains to be seen whether these putative mycoviruses do in fact infect fungi; linking viruses to eukaryotic hosts is currently difficult and is based on shared sequence homology between viruses recovered in a metatranscriptome and known viruses (i.e., viruses with known eukaryotic hosts) in public databases. The more similar the unknown viral sequence in a given metatranscriptome is to a virus in a public database, the more likely it is that they share a host.

Although existing methods are imperfectly suited to identifying divergent viruses with RNA genomes and their (mostly) eukaryotic hosts, bioinformatic methods for linking DNA viruses to host bacteria are much better and continue to improve. For example, some bacteria have clustered regularly interspaced short palindromic repeat (CRISPR) regions that provide adaptive immunity to mobile genetic elements, including viruses (Fig. 2). These regions of some host bacterial and most archaeal genomes contain pieces of DNA from the genomes of viruses that previously infected that host population, allowing the host to recognize and mitigate a new infection from another member of the same viral population based on sequence homology. The host uses the DNA from previous viral infections, contained in its own CRISPR region, as a homing device to identify the same viral sequence in an invading virus. This fascinating mechanism has been adapted as the mainstay of cutting-edge genetic engineering approaches, but for purposes of viral ecology, we can take advantage of the host CRISPR region and viral genomes to link viruses to host bacteria on the computer. Simply comparing the sequences from the host's CRISPR region to viral sequences and looking for exact matches provides a strong indication that a matching viral population has infected that host population in the recent past. With these virus-host linkages, we are beginning to be able to unravel virus-host population dynamics in nature. For example, ratios of virus-host abundance in thawing permafrost soils were shown to vary by host lineage, potentially indicating host lineage-specific differences in viral infection dynamics in response to changing environmental conditions.

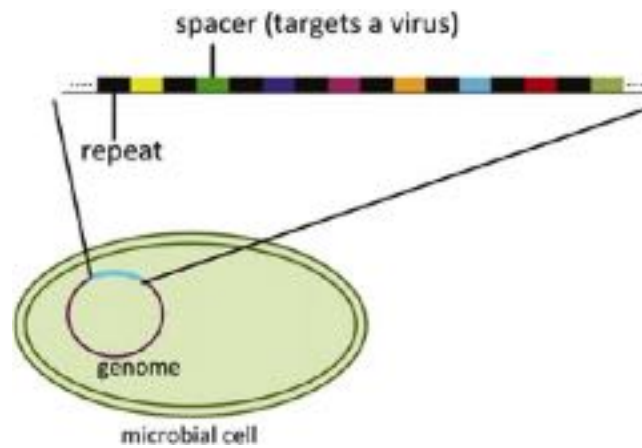


Fig. 2 CRISPR regions of microbial genomes, used to link viruses to bacterial and archaeal hosts through shared sequence homology.

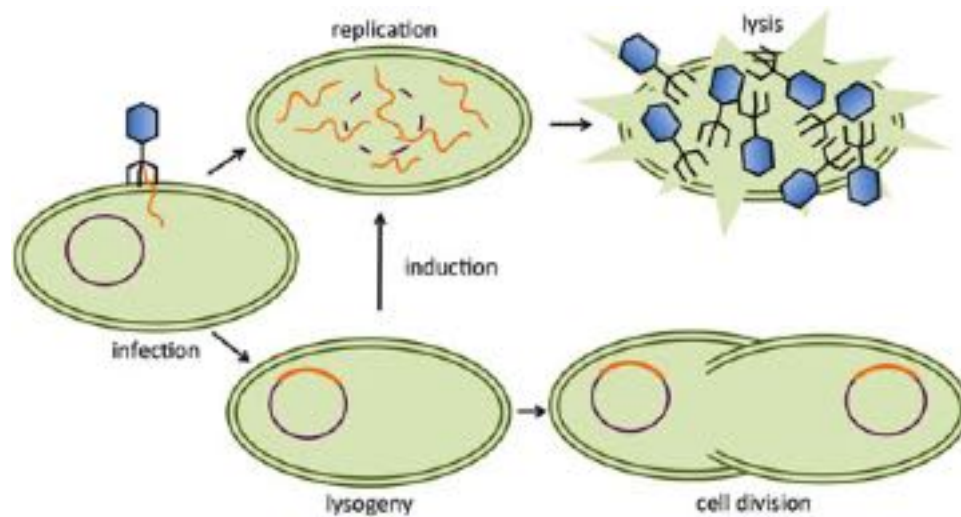


Fig. 3 Schematic overview of common replication strategies for viruses of microorganisms. Orange represents the viral genome, and purple represents the host genome.

Viral Impacts on Carbon, Nutrient, and Biogeochemical Cycling

Viruses of bacteria and archaea typically undergo one of two replication cycles: the lytic or the lysogenic cycle (**Fig. 3**). In the lytic cycle, viral infection leads to near-immediate viral replication inside the host cell and results in cell lysis (bursting) upon the release of progeny virions. In the lysogenic cycle, viral DNA is typically inserted into the host chromosome or maintained extrachromosomally and can therefore be replicated passively through host cellular division unless and until the virus is signaled to undergo the lytic cycle. Temperate phages can undergo the lysogenic or lytic cycles, depending on host and environmental cues, while lytic or virulent phages exclusively undergo the lytic cycle. There are many known chemical and environmental inducing agents that can drive a shift from the lysogenic to the lytic cycle, and these inducing agents generally signal that the current host may soon no longer be viable or that new hosts are likely to be available nearby. For example, known inducing agents include mitomycin C (an antibiotic), ultraviolet (UV) light, and bacterial quorum-sensing molecules, such as *N*-acyl-homoserine lactones (AHLs), all of which have been used in laboratory experiments to induce lysogens to undergo the lytic cycle for recovery in the free viral fraction. Additions of inducing agents to bacterial communities have allowed for bulk estimates of lysogenic viral abundances in soil, sediments, and seawater and have demonstrated seasonal shifts in temperate phage communities from predominantly lysogenic to predominantly lytic in polar marine environments.

These viral replication strategies can have important implications for microbial ecology and biogeochemistry. For example, lysis can contribute to carbon and nutrient cycling by releasing host cellular contents into the environment when the host cell bursts, and an estimated ~20%–40% of marine microbial cells undergo viral lysis daily in the global oceans. In marine and freshwater

systems, lytic events and viral infection in general can mitigate harmful cyanobacterial and algal blooms, alleviating the ecological impacts of eutrophication (excessive nutrient richness, resulting in removal of oxygen) and fish toxin production caused by these microorganisms. In terms of microbial community structure and function, lysis results in host mortality, impacting metabolic functions performed by infected host populations, while lysogeny may increase the fitness of specific host populations by contributing new genes and, therefore, new functional potential. Generally, lysis is presumed to be the dominant viral replication mode in most marine systems, though only a relatively limited number of community-scale evaluations of lysis and lysogeny have been attempted on seawater communities to date. Conversely, it has been hypothesized that temperate phages dominate in soil environments, which could mean that soil viral ecology is fundamentally different from lysis-dominated marine ecology, but this hypothesis has yet to be thoroughly tested.

Given that viruses rely on host cell machinery for replication, it is well known that viruses hijack the host cell throughout the infection cycle, altering host gene expression and metabolism. An infected host cell that contains an actively replicating virus is known as a “virocell”, and the physiology of a virocell is substantially shifted towards metabolic pathways required for viral replication, such as nucleotide biosynthesis. Interestingly, despite by definition not possessing their own metabolism or means of energy conservation, some viruses can also express their own genes to contribute to host metabolism during an active infection. For example, virus-encoded photosynthesis genes can be expressed during cyanophage infection of Cyanobacteria to plug a host metabolic bottleneck, keeping the Cyanobacteria just healthy enough (photosynthesizing just enough) to complete the viral replication cycle. By expressing such virus-encoded auxiliary metabolic genes (AMGs) during infection, viruses are thought to directly impact carbon, nitrogen, sulfur, phosphorous, and other biogeochemical cycles. Although the strongest evidence in support of direct viral AMG impacts on biogeochemistry at ecologically meaningful magnitudes is from cyanophage photosynthesis genes (in some cases, over 50% of the *pbsA* transcripts recovered in marine metatranscriptomes were of viral, as opposed to cyanobacterial, origin), many other host-like metabolic genes have been found in viral genomes, highlighting the intriguing possibility that viruses may impact a suite of biogeochemical cycles by expressing their own versions of metabolic genes during infection. For example, reverse dissimilatory sulfite reductase (*rdsr*) and *sox* genes with presumed roles in sulfur oxidation have been recovered in marine phage genomes, as have genes encoding a wide range of proteins involved in central carbon metabolism, along with the *P-II* gene that encodes a protein regulator of multiple nitrogen-cycling metabolic processes. AMGs that encode ammonia oxidation proteins (e.g., ammonia monooxygenase, *amoC*) have been recovered from marine archaeal viruses, and genes that encode proteins for complex carbon degradation to simple sugars have been recovered from soil viral genomes.

Drivers of Viral Community Composition

Now that methods are available for identifying and counting diverse viral populations (vOTUs) based on sequences recovered from viromes and total metagenomes, the field of viral ecology has begun to leverage established statistical ecological approaches to compare viral community ecology among samples and across biogeochemical conditions. The most substantial advances in this area have come from marine systems, from which hundreds of thousands of viral populations have been recovered and their relative abundances counted and compared. Considering pairwise comparisons of geographically adjacent ocean viromes, with adjacency considered both in terms of surface coordinates and depth within the water column, evidence suggests that viruses travel along ocean currents and tend to sink, perhaps contributing to the sinking aggregates that comprise marine snow and move surface carbon to the deep sea. Locally, marine viral community composition tends to correlate with environmental conditions known to structure host microbial communities, such as oceanic region (Longhurst province), biome, latitude, temperature, and oxygen concentrations, suggesting, perhaps not surprisingly, that viruses tend to be found where their hosts are.

Changes in viral population dynamics and community composition over time have also been explored in aqueous systems. In hypersaline lakes, viral populations were generally found to be highly dynamic over a three-year period, with some populations detected in only a handful of samples, suggesting that viruses are moving in and out of the ecosystem (or at least substantially changing in abundance) over relatively short time scales. In marine systems, seasonal and diel viral dynamics have been observed, suggesting that viruses respond to seasonal changes in temperature, nutrients, and/or host abundances, along with diurnal changes in host dynamics. The predominant bacterial hosts in the surface oceans are photosynthetic and therefore change their expression in response to light, and it appears that, at least to some extent, viral abundances track these patterns of host dynamics. Presumably, most viral responses to changing conditions are indirect responses to those conditions by way of more direct host responses, but there is also some evidence to suggest that viral infection success and population dynamics are directly tied to environmental drivers of viral decay and adsorption dynamics, though to what extent the environment impacts viral infection at an ecologically relevant scale remains to be determined.

Limited information is available for viral community ecology in soil, though the soil viral ecology field is projected to move at a rapid pace, now that laboratory methods for purifying viral particles from the soil matrix have caught up to cutting-edge bioinformatics approaches for reconstructing viral population genomes. Soil viral communities have been shown to contain thousands of diverse viral populations, and these viral communities have been shown to differ by habitat along a permafrost thaw gradient. Differences in viral community composition in thawing permafrost correlated with bacterial and archaeal community composition and known drivers of microbial community composition, including pH and soil moisture content. At this coarse community scale, it would appear that viruses simply follow their hosts, but at the level of individual microbial lineages, evidence suggests that viral abundances are not solely driven by the abundances of their hosts and may be impacted by environmental conditions.

Further Reading

- Brum, J.R., Hurwitz, B.L., Schofield, O., Ducklow, H.W., Sullivan, M.B., 2017. Seasonal time bombs: Dominant temperate viruses affect Southern Ocean microbial dynamics. *The ISME Journal* 11, 588.
- Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., *et al.*, 2015. Patterns and ecological drivers of ocean viral communities. *Science* 348, 1261498.
- Brum, J.R., Sullivan, M.B., 2015. Rising to the challenge: Accelerated pace of discovery transforms marine virology. *Nature Reviews Microbiology* 13, 147–159.
- Emerson, J.B., Roux, S., Brum, J.R., *et al.*, 2018. Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology* 3, 870–880.
- Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., Sullivan, M.B., 2017. Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *The ISME Journal* 11, 1511.
- Paez-Espino, D., Elie-Fadrosh, E.A., Pavlopoulos, G.A., *et al.*, 2016. Uncovering Earth's virome. *Nature* 536, 425–430.
- Pratama, A.A., van Elsas, J.D., 2018. The "neglected" soil virome: Potential role and impact. *Trends in Microbiology* 26, 649–662.
- Schoelz, J.E., Stewart, L.R., 2018. The role of viruses in the phytobiome. *Annual Review of Virology* 5, 93–111.
- Sieradzki, E.T., Ignacio-Espinoza, J.C., Needham, D.M., Fichot, E.B., Fuhrman, J.A., 2019. Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes. *Nature Communications* 10, 1169.
- Williamson, K.E., Fuhrmann, J.J., Wommack, K.E., Radosevich, M., 2017. Viruses in soil ecosystems: An unknown quantity within an unexplored territory. *Annual Review of Virology* 4, 201–219.

The Role of Retroviruses in Cellular Evolution

Andrea Kirmaier and Welkin E Johnson, Boston College, Chestnut Hill, MA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Endogenous retrovirus A locus in an organismal genome comprising a complete or partial retroviral provirus; opposite of 'exogenous'.

Exaptation A special form of adaptation involving cooption of an existing trait for a new function.

Exogenous retrovirus A retrovirus that spreads by horizontal transmission from one host organism to another.

Host dependency factor (HDF) Any cellularly encoded protein a retrovirus depends on for the completion of its replication cycle.

Indel Combined abbreviation for 'insertion/deletion'; mutations resulting from the insertion or deletion of segments of DNA.

Provirus Integrated DNA copy of a retrovirus' RNA genome.

Restriction factor (RF) A cellularly encoded protein with direct anti-viral activity.

Introduction

Retroviruses may have originated more than 400 million years ago, close to the dawn of vertebrate life on earth. It is therefore reasonable to expect that the evolution of vertebrate organisms, as hosts, has been continuously shaped and reshaped by retroviral agents over such vast expanses of evolutionary time. Like all viruses, retroviruses are obligate intracellular parasites and depend on host cells for replication and the generation of progeny virions. What distinguishes retroviruses from most other vertebrate viruses, and adds to their role as selective agents, is their particular mode of replication: all retroviruses replicate by reverse-transcribing the viral RNA genome and integrating the resulting DNA copy permanently into the chromosomal DNA of an infected cell. The resulting insertion is called a provirus, which comprises the minimum set of viral regulatory elements and viral protein-coding genes necessary for expression and propagation of the next generation of progeny virions. The provirus remains as a novel genetic insertion in the host chromosome for as long as the cell itself persists. In the case of dividing cells, the provirus is copied as part of the normal process of chromosomal DNA replication and a copy of the provirus is inherited by each of the daughter cells.

To the extent that any virus causes disease or pathogenesis and affects host fitness, it can act as an agent of natural selection driving changes in alleles of genes that affect susceptibility to infection (or its consequences). Retroviruses are no exception. However, as a consequence of integration, retrovirus infections are often persistent and can result in chronic diseases. Because of integration, retroviruses are also potentially mutagenic, causing disruption or dysregulation of genes and activation of oncogenes (Fig. 1). If integration occurs in germline tissue, producing a heritable provirus known as an endogenous retrovirus (ERV), any phenotypic consequences of insertion are also potentially heritable (passed on to the host's offspring). In such cases, the ERV insertion itself can be subject to selection. Importantly, as with any other alterations to an organism's germline DNA, an ERV may be deleterious, beneficial, or neutral with respect to host fitness.

Molecular Consequences of Retrovirus Integration

The genetic effect of proviral integration into the host genome can be determined in part by the site of integration. Different kinds of retroviruses display preferences for different genomic contexts. Many preferentially integrate into functional regions of the host genome. For example, retroviruses in the genus *Lentivirus*, such as HIV-1, tend to integrate in actively transcribed host genes. Disruption of a host gene's coding region can result in a loss of protein expression from that locus, and a loss of the protein's function in the cell.

Altered gene regulation is another potential consequence of proviral integration. Host gene dysregulation can stem from a loss or disruption of a gene's native regulatory elements (promoter, enhancers, transcription factor binding sites) caused by the integrating provirus, and/or it can stem from the proviral long-terminal repeats (LTRs) acting on nearby genes. LTRs contain the viral promoters and transcription factor binding sites, which can interact with regulatory factors provided by the cell to influence regulation of nearby genes, imposing the virus' regulatory program. Since viral promoters are generally fairly strong, increased and/or aberrant transcription of neighboring host genes can result. Alteration of host gene expression is often seen in infections with viruses in the genus *Gammaretrovirus*, such as Murine Leukemia Virus (MLV), which prefer to integrate near transcriptional control elements such as promoters, transcriptional start sites and enhancers. Moreover, if provirus-mediated dys-regulation happens to genes involved in cell cycling, cells can start to proliferate uncontrollably and ultimately give rise to cancers.

A third way in which retroviral proviruses can genetically interact with the host is through the accidental generation of novel protein products originating from alternative mRNA splicing. Proviruses contain multiple splice donor and splice acceptor sites for

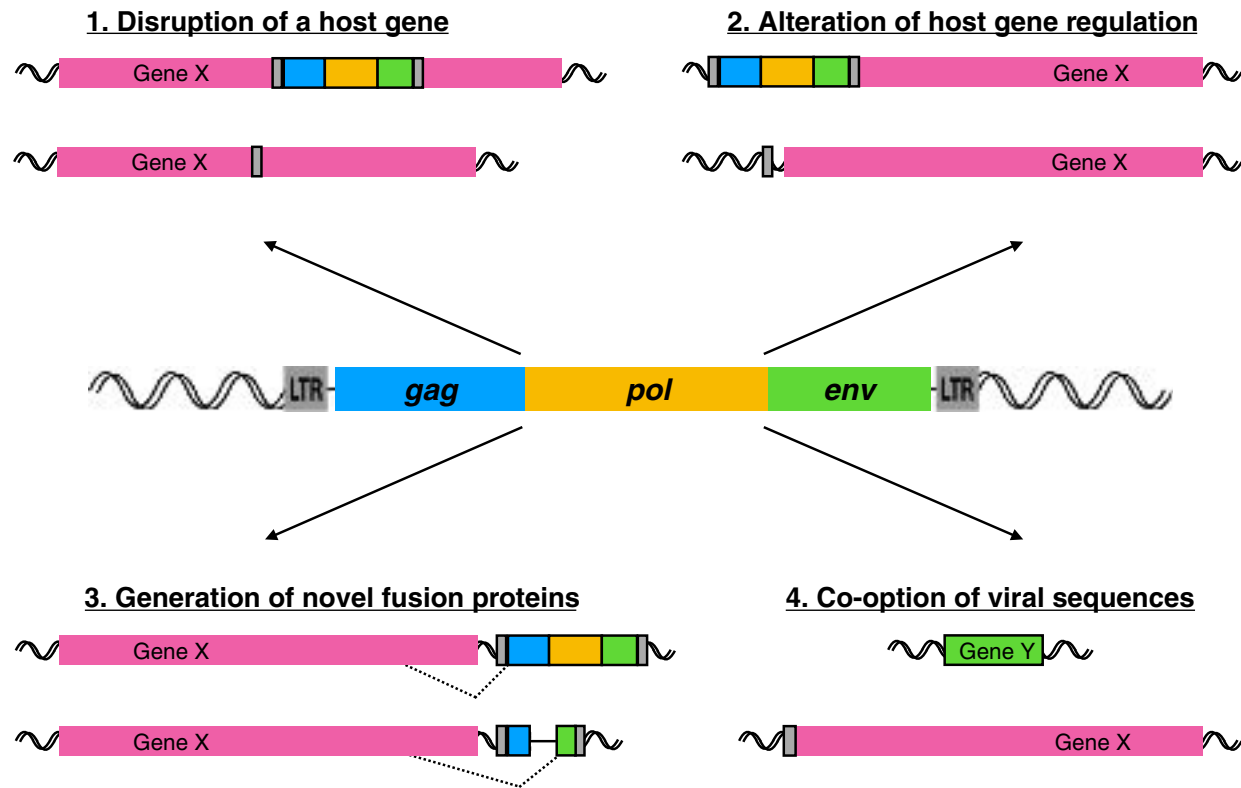


Fig. 1 Genetic interactions of a retroviral provirus with the host genome. The center of the figure depicts an integrated provirus (gray and colored rectangles) flanked by host chromosomal DNA (wavy lines). The provirus comprises LTRs, which contain most of the cis-acting regulatory elements of the virus, and the viral genes (in the illustration, only the canonical *gag*, *pol* and *env* genes are shown). The provirus directs expression of viral genes and the production of progeny virions (not shown). Proviruses can exist in multiple forms; in addition to fully intact proviruses with all regulatory and coding sequences, proviruses can be partially or completely defective for expression due to mutations and indels (exemplified in sub-figure 3, bottom). The provirus can have a variety of effects on host genes: (1) The presence of a provirus or a solo-LTR can disrupt the coding sequence of host gene X, which may result in a non-functional host protein X. (2) A provirus or a solo-LTR near the 5' end of a host gene can cause aberrant transcription of the host gene due to disruption of the host gene regulatory regions or due to the proviral LTR serving as the new regulatory element. Changes in both quantity (i.e., how much) or quality (i.e., when/where) of transcription of gene X are therefore possible. (3) Provirus sequences can give rise to novel splice variants of host genes, for example, if the host gene X transcripts are fused with viral transcripts by altered RNA splicing. This can occur both at the 5' and the 3' end of host gene X (only 3' end is shown; dotted lines = splice junctions). (4) Proviral sequences can be co-opted by the host (a unique form of adaptation known as exaptation). For example, if the expression of a viral protein or altered regulation of a host gene confers a selective advantage to the organism, there is a chance of the novel genetic variant being favored by natural selection. While scenarios (1), (2) and (3) can happen as a consequence of both somatic and germline integration, scenario (4) is exclusive to insertions into the germline (i.e., ERVs).

the generation of sub-genomic mRNA transcripts. Should a provirus integrate within the coding region of a host gene, such that a read-through mRNA is generated by DNA Pol II, the viral splice donors and acceptors can pair with host splicing elements, and thus produce chimeric transcripts. This mechanism can give rise to alternative isoforms of the host mRNA, and even produce novel proteins that are part host-derived, part virus-derived.

Finally, segments of integrated proviruses found in the genomes of many vertebrate organisms, known as endogenous retroviruses (ERVs), can be co-opted by evolution to provide useful cellular functions. This evolutionarily important phenomenon, sometimes referred to as *exaptation*, is discussed further in a later section.

Retroviruses Can Drive Evolution of Host Genes

Retroviruses cause persistent infections, and collectively have been associated with a spectrum of chronic diseases, including various malignancies and immunodeficiencies. To the extent that chronic, long-term infection and its pathogenic consequences reduce the reproductive fitness of host organisms, retroviruses can act on a host population as agents of natural selection, altering allele frequencies of genes that reduce or increase susceptibility to infection and/or severity of disease. Hypothetically, this can encompass a variety of types of genes, including: (1) genes that encode "host dependency factors (HDFs)", which are factors hijacked by viruses for optimal replication and spread; (2) restriction factors (RFs), which are cellular factors that have evolved to

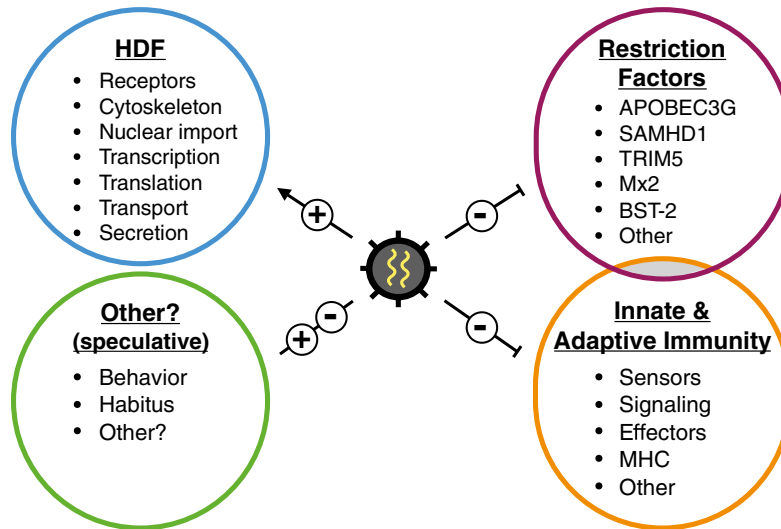


Fig. 2 Retroviruses as selective agents. Any type of host traits or factors that positively or negatively contribute to a retrovirus' ability to infect, replicate and disseminate within and between hosts, may be subject to selection by the virus. There are two general types of host traits and factors relating to retroviruses: (i) traits/factors required for or helpful to the generation and/or spread of progeny virions (designated by "+", left side), and traits/factors impeding the generation and/or spread of progeny virions (designated by "-", right side). Due to retroviral pathogenicity, variants of restriction factors and other innate and adaptive immune effectors with better anti-viral function will be positively selected, as they will result in higher resistance to retroviral infection (purple and orange circles). On the other hand, a retrovirus will convey a selective advantage to individuals encoding alleles of host dependency factors (HDF) that are less suited to interact with the virus, and it will conversely negatively select alleles of HDFs that better support interactions with the virus (blue circle). Other, more complex traits, such as behavior or habitus can also affect the efficiency of retroviral replication and spread, and will be selected accordingly (green circle). For example, higher aggression might lead to more fighting, which can increase the chance of exposure to a virus through direct contact with blood. In this case, the virus would indirectly select for more docile individuals, as they would have a lower chance of contracting an infection through contact. Prolonged, i.e., multi-generational exposure of a population to a retrovirus will result in changes of host allele frequencies based on selection by the virus.

block or suppress viral replication at the level of the cell; (3) adaptive and innate immune factors, including pattern-recognition receptors, effectors, and signaling molecules; and (4) other, more complex traits (Fig. 2).

Host dependency factors (HDFs)

As with all viruses, retroviruses hijack the molecular machinery of the host cell for the purpose of replication and production of progeny genomes and virions. This can involve numerous molecular interactions between viral macromolecules and host factors involved in nucleic acid synthesis, transcription, translation, post-translational modification, nuclear import/export, sorting and trafficking, etc. There is some evidence that retroviruses may select for changes in host dependency factors, the simplest hypothesis being selection should favor changes that eliminate the interaction (and thereby reduce the host's susceptibility to infection or disease). Here, selection may involve a trade-off between reduced cellular function and increased resistance. The classic, non-retroviral example of such a trade-off involves the parasite that causes malaria and the sickle-cell trait that is common in areas with endemic malaria. Despite the deleterious effects of the sickle-cell trait, the increased resistance to malaria favors heterozygous individuals (with one copy of the mutant allele) and thereby maintains the sickle cell mutation at higher than expected frequency. Currently, the best evidence for such trade-offs involving retroviruses comes from adaptations in various cell-surface proteins exploited by retroviruses as entry-receptors. However, it is reasonable to expect that selection has involved trade-offs to reduce or alter other kinds of virus-host molecular interactions.

Restriction factors (RFs)

Currently, the best evidence for selection driven specifically by retroviruses comes from comparative analysis of several well-known restriction factors, including APOBEC3G, TRIM5alpha, BST2/tetherin, SAMHD1 and Mx2. Interspecies comparisons of such genes often reveal signatures of positive selection (selection for changes in amino-acid sequence) affecting subdomains known to interact directly with viral targets or proteins. In some cases, cell-culture experiments and animal models have corroborated the hypothesis that retroviruses may have been the agents of positive selection, primarily by showing that the putatively selected amino-acid changes disrupt or diminish interactions with known retroviral targets. While this kind of evidence is indirect, such patterns suggest that the changes were selected by virtue of conferring resistance to one or more retroviruses in the past.

For example, the cytoplasmic TRIM5alpha protein recognizes and coats the hexameric lattice of retroviral capsid shells in the cytoplasm, thereby blocking subsequent stages of the replication cycle. A comparison of TRIM5alpha orthologs from humans and

other primate species reveals a large number of variable residues in the subdomain involved in capsid recognition, and indeed, the differential ability of primate TRIM5alpha homologs to restrict various retroviruses maps precisely to these same variable sites. This supports the hypothesis that divergent evolution of TRIM5alpha sequence between primates may have been driven by selection in the form of retroviral pathogens. Similar patterns have been reported for several other restriction factors, including APOBEC3G, BST2/tetherin and SAMHD1, among others. In fact, evidence for positive selection has become a commonly applied criterion for discovering putative restriction factors.

Host immunity

Selection on adaptive immune factors, such as antibody genes and MHC, is also likely but more difficult to establish, as these loci likely adapt to a broad array of pathogens (including but not limited to many different kinds of viruses, bacteria and eukaryotic parasites). Studies of HIV-1-positive patient cohorts have clearly identified different MHC alleles associated with positive and negative outcomes – thus, it is possible that an epidemic (such as the HIV-1/AIDS pandemic), extended across multiple generations of a population, could result in selection for increased frequency of protective alleles and reduced frequency of alleles that enhance infection or disease.

Host genes with indirect effects on retrovirus infection, spread or disease

In addition, it is possible that adaptations in any gene or gene network could reflect selection by retroviruses – for example, changes that manifest as alterations in behavior could be so selected, if the result is reduced susceptibility to infection or disease. Presently, there is very little documentation to support this speculative hypothesis, and the idea remains to be investigated.

Germline Insertions (Endogenous Retroviruses) Can Generate Genetic Novelty

Endogenous retroviruses (ERV) are germline-integrated proviruses, and, in contrast to exogenous retroviruses, are part of the heritable genetic information of the host. An organism that originates from a germline cell containing a provirus will have a copy of this provirus in every nucleated cell. Over time, the ERV locus can even become fixed in a population, such that most or all members of a diploid population carry two copies of the ERV locus (one on each chromosome) (**Fig. 3**). The genomes of many vertebrate species contain thousands of ERV loci, ranging in age from thousands to tens of millions of years old. For example, approximately 8% of the human genome, equivalent to hundreds of thousands of loci, are of retroviral origin (for comparison, less than 2% of the human genome comprises protein-coding genes).

As with other genomic loci, ERV insertions will accumulate mutations over generational time and will diverge from the original proviral sequence. ERVs can therefore exist in several states: intact ERVs (retaining both LTRs and most or all coding regions) are usually less than 1 million years old, while more degraded ERVs tend to be older. Point mutations or short indels can inactivate open reading frames and regulatory sequences, so the viral genes are no longer transcribed. Very old ERVs (tens of millions of years and up) typically have accumulated larger deletions and insertions as well as point mutations. Homologous recombination frequently occurs between the 5' and 3' LTRs, which removes the intervening portion of the provirus and leaves behind a solo-LTR. Notably, solo-LTRs often outnumber other forms of ERV by an order of magnitude or more, and are the most common form of ERV element found in vertebrate genomes.

At the time of germline integration, an ERV effectively becomes just another genetic element of the host genome, and as with the rest of the genome, it will be subject to random genetic drift and selection based on whether or not it contributes to host fitness. In fact, there are many documented examples of ERV regulatory elements and ERV protein-coding reading frames retaining (or acquiring) functions beneficial to the host organism. This process is commonly referred to as “co-option” or “exaptation”.

Exaptation of ERVs

There are two general ways by which an ERV might be co-opted by host evolution: (i) exaptation of viral protein-coding genes, when one or more of the retroviral proteins provides a new cellular function, and (ii) exaptation of viral regulatory elements, such as those found in the LTRs, to alter the regulation or expression of cellular gene(s) in a manner that is selectively advantageous to the organism.

Exaptation of ERV-encoded envelope glycoproteins

The best-documented examples of co-option of retroviral proteins involve retrovirus entry glycoproteins (Env). The Env proteins are initially expressed on the surface of the infected cell, and incorporated into progeny virions as they bud from the cell. The primary viral functions of Env proteins are (1) binding to a cognate cell-surface protein that functions as the viral entry receptor, and (2) providing the mechanical energy to fuse the virion and cellular membranes and permit the viral genetic material to enter the cell. In some cases, expression of the Env protein also has a third function (3) blocking receptor availability on the infected cell, which can block reinfection. This phenomenon is known as superinfection interference. In cases where ERV-derived Env glycoproteins have been coopted, the cellular function usually capitalizes on one or more of these three activities.

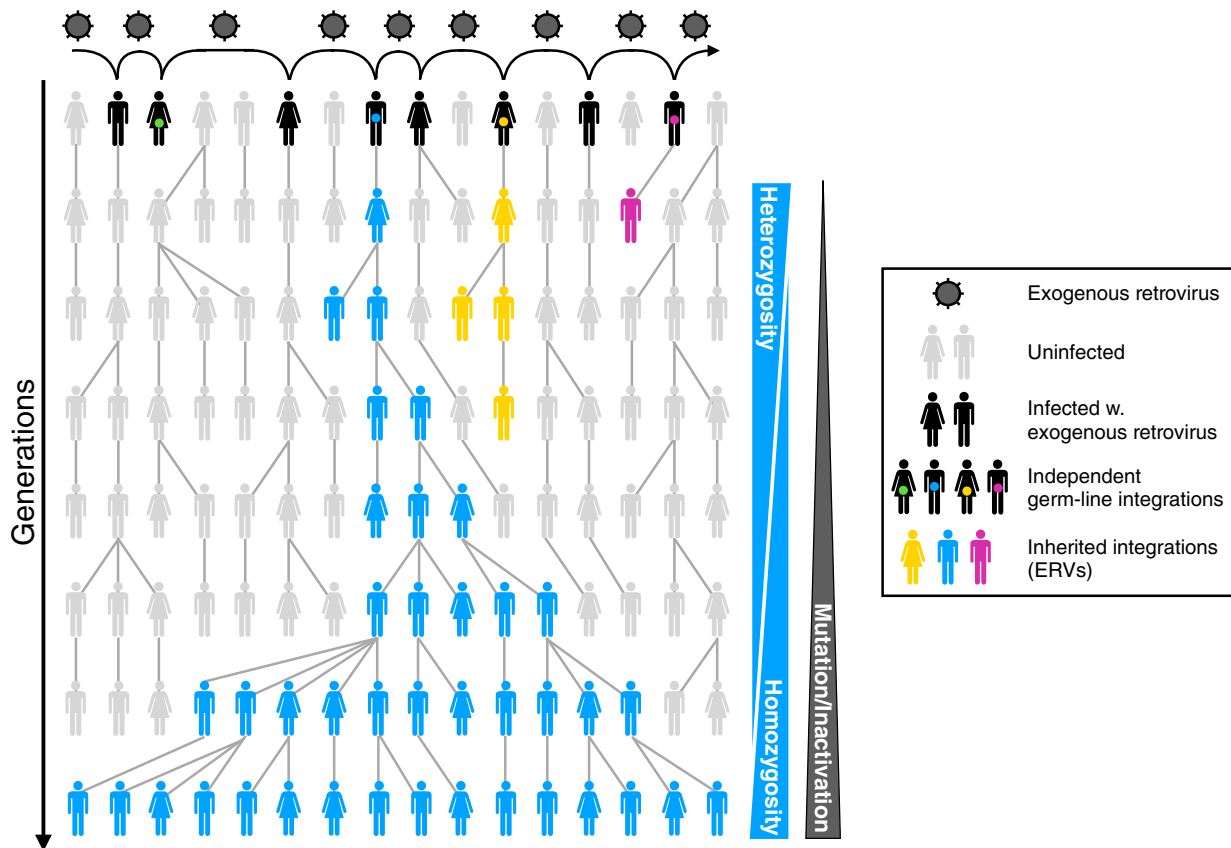


Fig. 3 Origin of ERVs. An exogenous retrovirus can infect somatic or germline cells. In somatic infections, the integrated provirus will persist for the lifetime of the infected cell but it will perish when that cell dies, and the provirus is not heritable. In germline infections, the integrated provirus becomes part of the heritable information of the host (green, magenta, yellow, blue). If an organism with a germ-line integrated retrovirus has offspring, each nucleated cell in the offspring's body will contain a copy of the provirus, including germline tissue that produce gametes – thus, the offspring can pass on the ERV locus to its own offspring, and so on. Over time, an ERV locus can be lost from a population (green, magenta, yellow) or it can increase in frequency and eventually become fixed, i.e., every organism of the population carries the locus (blue). Early on, any individuals that carry the insertion have one copy, i.e., they are heterozygous. As the frequency of the insertion increases, individuals are increasingly likely to carry copies on both chromosomes, i.e., they are homozygous. The ERV locus will also usually accumulate mutations, including mutations disrupting the coding and/or regulatory functions (dark gray triangle). As a consequence of endogenization, ERV sequences can persist long after the original exogenous retrovirus has gone extinct.

The mammalian Syncytins are the best documented example of exapted ERV-encoded Env proteins. These proteins are expressed in trophoblast cells during embryonic development, and function to drive fusion of trophoblast cells to form the syncytiotrophoblast layer. This involved evolutionary repurposing of both the viral receptor-binding and membrane fusion functions of Env. Syncytins and syncytin-like ERV elements have been found in all mammalian clades, including marsupials, as well as in some live-bearing reptiles, indicating that retroviruses have been repeatedly involved in the evolution of live birth.

The ability to bind their cognate receptor also selected for the use of ERV Env proteins as antiviral defenses. Because ERV-encoded Env may retain the ability to bind and block the virus's cognate cell-surface receptor (analogous to superinfection interference), the expression of the ERV Env can inhibit infection by exogenous viruses using that same receptor, thereby providing resistance to infection. In such cases, the former viral Env glycoprotein has in essence become a restriction factor. Examples of this process include the mouse *Fv4* gene, which provides resistance to infection to murine leukemia virus (MLV) by downregulating the viral receptor, and the feline *Refrex* genes, which block infection by feline leukemia virus (FeLV) by a similar mechanism. There is evidence that an ERV in the human genome, known as HERV-T, may have had similar activity against a retrovirus that afflicted our ancestors.

Exaptation of ERV-derived non-coding elements

The provirus contains all of the essential regulatory and cis-acting elements required for expression of the viral genes. For example, the LTRs contain the primary viral promoter, transcription factor binding sites, and signals for mRNA termination and polyadenylation. ERVs in organismal genomes may retain some or all of these elements, and there are now multiple reported examples of co-option of ERV LTRs as components of the regulatory pathways of vertebrate genes and gene networks. For example, the

human globin locus is under the control of an ERV9 LTR, while ERV LTRs upstream of the human *AMY1* genes drive tissue-specific expression of amylase in salivary glands. A particularly striking example of ERV LTR exaptation comes from the discovery that gene regulatory networks involved in innate immunity in mice and humans are regulated, at least in part, via ERV LTRs that have been coopted as cis-acting regulatory elements. Such discoveries lead to the hypothesis that ERV may serve an important evolutionary function by facilitating the formation or rewiring of gene regulatory networks.

Exaptation of other ERV-derived proteins

There is a small but growing list of other ERV-derived retroviral genes that exhibit cellular functions. There is at least one example of a host gene resembling a retrovirus-related protease, known as SASPase. Perhaps the best studied example is the rodent *Fv1* gene, which functions as an ERV-derived restriction factor that can block infection by a wide variety of retroviruses. *Fv1* encodes the remnants of a retroviral *gag* gene, and its expression restricts exogenous retroviruses at an early, post entry stage in the viral replication cycle, possibly via a direct interaction with the capsid shell of the incoming virus as it penetrates the cytoplasm. Another example of exaptation of a retroviral *gag* gene is found in sheep, and confers resistance specifically to the oncogenic Jaagsiekte sheep retrovirus (JSRV). Although such examples are limited, the presence of intact ERV-encoded open-reading frames in the genomes of many vertebrates, including humans, raises the possibility that many more examples will be discovered. Finding candidate sequences is relatively straightforward; the difficulty lies in establishing function, particularly in non-model organisms.

Summary

Darwinian evolution is the combined result of variation and natural selection. The sheer abundance of ancient ERV found in vertebrate genomes, and the widespread presence of extant retroviruses found in nature, together indicate the potential for retroviruses to have had a significant impact on the origins and evolution of vertebrate species. This impact is two-fold: retroviruses can exert selective pressure on host populations, and ERV insertions in the germline can contribute to genetic variation within organismal populations. Going forward, comparative genomics and experimental models will continue to uncover additional examples and help document the actual extent to which retroviruses (and the related LTR-retrotransposons) have contributed to the evolution of cellular organisms.

Further Reading

- Blanco-Melo, D., Gifford, R.J., Bieniasz, P.D., 2017. Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. *eLife* 6. doi:10.7554/eLife.22519.
- Chuong, E.B., Elde, N.C., Feschotte, C., 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351 (6277), 1083–1087. doi:10.1126/science.aad5497.
- Coffin, J.M., 2013. Virions at the gates: Receptors and the host–virus arms race. *PLoS Biology* 11 (5), e1001574. doi:10.1371/journal.pbio.1001574.
- Greenwood, A.D., Ishida, Y., O'Brien, S.P., Roca, A.L., Eiden, M.V., 2017. Transmission, evolution, and endogenization: Lessons learned from recent retroviral invasions. *Microbiology and Molecular Biology Reviews* 82 (1), doi:10.1128/MMBR.00044-17.
- Jern, P., Coffin, J.M., 2008. Effects of retroviruses on host genome function. *Annual Review of Genetics* 42 (1), 709–732.
- Johnson, W.E., 2019. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nature Reviews Microbiology* 17, 355–370.
- Kirmaier, A., Wu, F., Newman, R.M., *et al.*, 2010. TRIM5 suppresses cross-species transmission of a primate immunodeficiency virus and selects for emergence of resistant variants in the new species. *PLoS Biology* 8 (8), e1000462.
- Lavialle, C., Cornelis, G., Dupressoir, A., *et al.*, 2013. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1626).
- Lynch, V.J., 2016. GENETICS. A copy-and-paste gene regulatory network. *Science* 351 (6277), 1029–1030. doi:10.1126/science.aaf2977.
- Meisler, M.H., Ting, C.-N., 1993. The remarkable evolutionary history of the human amylase genes. *Critical Reviews in Oral Biology and Medicine* 4 (3–4), 503–509.
- Meyerson, N.R., Sawyer, S.L., 2011. Two-stepping through time: Mammals and viruses. *Trends Microbiology* 19 (6), 286–294.
- Neil, S., Bieniasz, P., 2009. Human immunodeficiency virus, restriction factors, and interferon. *Journal of Interferon & Cytokine Research* 29 (9), 569–580.
- Vogt, P.K., 2012. Retroviral oncogenes: A historical primer. *Nature Reviews Cancer* 12 (9), 639–648. doi:10.1038/nrc3320.

Relevant Websites

- <https://www.genecards.org>
GeneCards – Human Genes.
- <https://www.malacards.org>
MalaCards – Human disease database.
- https://viralzone.expasy.org/71?outline=all_by_species
Retroviridae.
- <https://talk.ictvonline.org/taxonomy/>
Taxonomy.

The Role of Bacteriophages in Bacterial Evolution

Chris M Rands, University of Geneva Medical School and Swiss Institute of Bioinformatics, Geneva, Switzerland

Harald Brüssow, Laboratory of Gene Technology, Department of Biosystems, KU Leuven, Leuven, Belgium

© 2021 Elsevier Ltd. All rights reserved.

Antagonistic Phage-Bacterium Relationships

Seminal publications influence the scientific discussion well beyond the data described within that specific report. This was certainly the case for the first report describing the discovery of bacterial viruses. Already the naming of bacterial viruses as “bacteriophages” (bacterial eaters) by Félix d’Hérelle in 1917 set the stage for bacteriophages as natural enemies of bacteria. Based on this concept of a predator-prey relationship, phages (short for bacteriophages) were used in the Soviet Union as antimicrobial agents (“phage therapy”) well before the development of antibiotics. This concept of phages as killers of bacteria received strong support in a 1989 study in which Norwegian marine microbiologists reported high viral titers in the ocean and freshwater systems exceeding that of bacteria by a factor of ten. When other microbial ecosystems were explored (soil, plant roots, human gut) the viral to bacterial ratio differed according to the particular ecological situation. However, a clear trend emerged: phages represented the major causes of bacterial mortality in the ecosphere, together with starvation and grazing by protists. Under such a strong selection pressure, one could anticipate a fierce evolutionary arms race between bacteria and phages. Indeed, another seminal paper from 1943 by Luria and Delbrück reported overnight resistance development in bacteria exposed to phages, followed by the development of mutant phages overcoming this bacterial resistance, followed by bacteria becoming resistant to these mutant phages. This so-called Red Queen paradox revealed phages are major drivers of bacterial evolution.

The then young field of molecular biology concentrated their efforts on the study of *Escherichia coli* phages, describing their sophisticated hijacking of the bacterial host metabolic and genetic apparatus. However, bacterial countermeasures against phage attack in the form of restriction-modification (R-M) enzymes were also soon described. When genomics emerged from genetics, the important role of phage-bacterial competition in their coevolution could be read from their genomes. Comparative *E. coli* genomics revealed R-M genes and genes synthesizing lipopolysaccharides (a frequently used primary receptor for phages) as two of the most hypervariable genomic regions in *E. coli* demonstrating that they are under strong selective pressure. Comparative phage genomics showed that phage tail fiber tips, which mediate the first contact with the bacterial receptors, are the principal hypervariable phage genome regions. As such, this genome analysis reveals the foot prints of the arms race between phages and their bacterial hosts.

When molecular biology shifted to more complex model systems, applied scientists dealing with the problem of phage control in the dairy industry maintained phage research and discovered numerous anti-phage defense mechanisms, mostly encoded on plasmids of *Lactococcus lactis*, the major cheese starter bacterium. When these abortive infection (*abi*) genes were introduced into industrial starter bacteria, mutant phages overcoming these resistance mechanisms were frequently documented. Surprisingly, the starter bacterium used in yogurt fermentation, *Streptococcus thermophilus*, suffered a lot from phage attack (Fig. 1), but lacked plasmids and *abi* genes. This paradox led to the discovery of genetic loci on the *S. thermophilus* genome that contained short DNA copies of various dairy phages. These “Clustered Regularly Interspaced Short Palindromic Repeats” (CRISPR) together with “CRISPR-associated” (*Cas*) genes, which encode enzymes that use CRISPR sequences as a guide to recognize and cleave matching DNA in phages, turned out to be the major anti-phage system in *S. thermophilus*. Follow-up research revealed that CRISPR-Cas systems are very diverse and widespread in bacteria and nearly omnipresent in archaea, reflecting the strong impact of phages on bacterial evolution.

While this research identified phages as foes of bacteria locked with their hosts in an age-old and ongoing arms race, ecological theory predicts advantages of predation for other potential prey at the population and ecosystem level. In this area, marine microbiologists took the lead. One concept was that phages, when lysing their host, release nutrients that feed bystander bacteria. Phages thus maintain nutrients in the microbial loop and prevent their loss in the food chain and sedimentation to the bottom of the ocean. As such, phages are now revealed as important drivers of geochemical cycles. Another beneficial role of phages was identified by marine microbiologists based on the threshold concept. Laboratory experiments with coliphages had demonstrated that phages do not infect low abundance bacteria and oceanographers developed the “killing the winning” concept, in which phages efficiently infect outgrowing bacteria in “blooms”. This concept represents a powerful mechanism maintaining bacterial diversity in an ecosystem. Support for this concept has also come from medical microbiologists, demonstrating that *Vibrio cholerae* epidemics show rhythmic alterations with the outgrowth of *V. cholerae*, followed by that of vibriophages in the environment.

Marine microbiologists have opened an important chapter in phage-bacterium coevolution by investigating their interaction in natural ecosystems outside of the laboratory. Historically, phage research has been dominated by work focusing on a set of just seven *E. coli* phages (T1 to T7). While this reductionist approach became a powerful motor for the development of molecular biology, it introduced a bias: all selected phages belonged to the group of virulent phages. Virulent phages are maintained in nature by an uninterrupted chain of bacterial infection cycles. Furthermore, phage infection was investigated under optimized *E. coli* growth in planktonic culture. We know now that slow and unproductive phage replication dominate natural systems, where phages have to deal with bacteria which spent most of their time under limited nutrient conditions. In addition, most bacteria in natural ecosystems grow in biofilms while all model type phages were grown under planktonic growth conditions. Only recently,

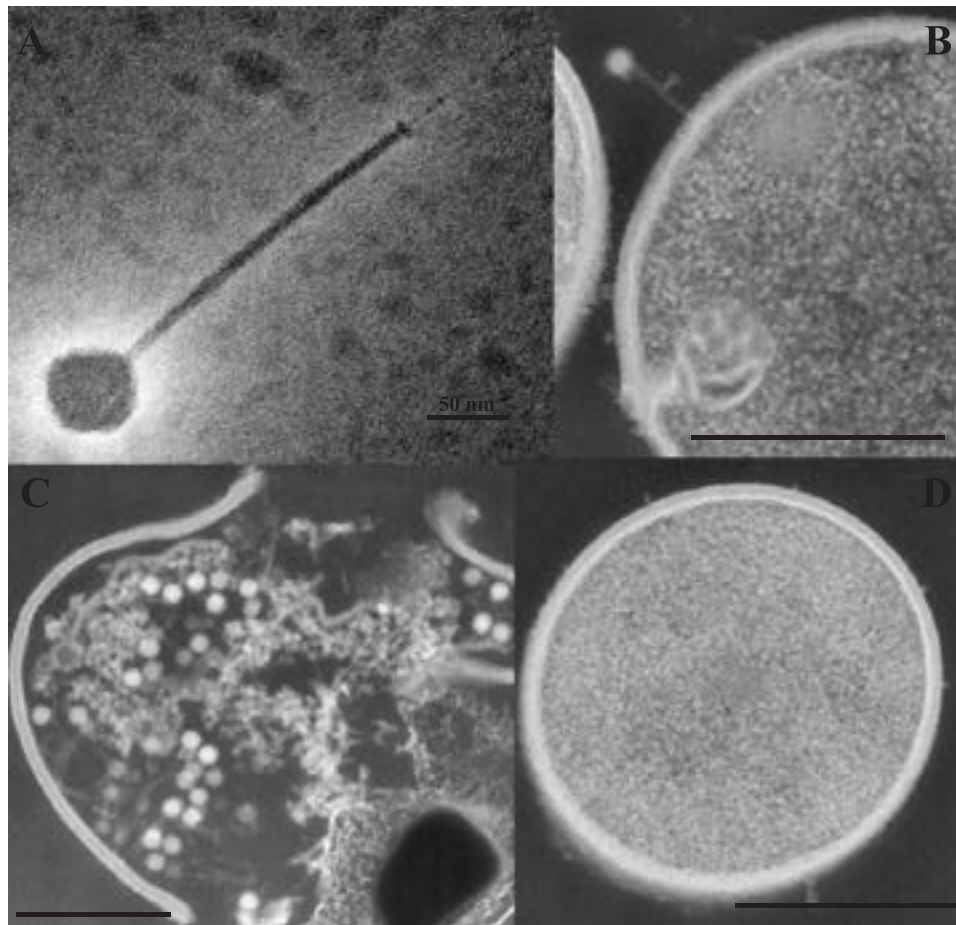


Fig. 1 *Streptococcus thermophilus* phage Sfi21: lytic infection versus lysogeny. (A) Phage Sfi21, a Moineauvirus, is a lambda-like tailed phage of the Siphoviridae family. Negative staining electron microscopy (EM). Size bar: 50 nm. (B) An EM thin section shows phage Sfi21 adsorbed to its bacterial host cell. Size bars in B–D: 200 nm. (C) Lytic infection results in the production of progeny phage in the cytoplasm of the infected cell followed by the lysis of the cell. (D) Alternatively, phage Sfi21 integrates its DNA as a prophage into the bacterial chromosome leading to a lysogenic cell with normal morphology and growth properties. The prophage co-exists with the lysogenic cell until induction results into a new lytic phage replication cycle. *Source:* Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., Brüssow, H., 2003. Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* 6 (4), 417–424.

medical microbiologists have started to explore phages infecting biofilms and discovered that filamentous *Pseudomonas* phages actively structured biofilms in the lung of cystic fibrosis patients. Marine microbiologists discovered that phages directly isolated from the ocean have broader host ranges than phages adapted to laboratory growth, changing an important life history parameter of phage-host interaction for ecological and evolutionary considerations.

Synergistic Phage-Bacterium Relationships

Different aspects of phage-bacterium evolution were revealed with the introduction of *E. coli* phage lambda into phage research in the 1950s. Phage lambda is the prototype of temperate phages. Temperate phages, when infecting a cell, have a choice between two life styles. Either they go into a productive replication cycle like any virulent phage, producing progeny phages and killing the host bacterium, or they integrate their phage DNA as a prophage into the bacterial chromosome, creating a lysogenic cell. The prophage remains, except for the phage repressor, transcriptionally largely silent. Prophages then replicate passively with the bacterial DNA until they receive signals from the host bacterium or the environment that induce the prophage into another lytic phage replication cycle, leading to the death of the bacterial host. This peculiar lifestyle has interesting evolutionary consequences: when phage DNA becomes integrated into the bacterial DNA, the prophage DNA is submitted to the selective forces working on bacterial fitness. One might suspect that prophages are a burden to the bacterial cell since they represent a time bomb that can cause the cell to explode at any time. In addition, when phage lambda integrates into the *E. coli* chromosome, it increases the bacterial genome size by about 1% causing supplementary metabolic costs during DNA replication. One percent more DNA to synthesize might not seem like a large metabolic load, but in some pathogenic *E. coli* strains up to 18 prophages are integrated into

the chromosome, representing 18% of the total bacterial DNA. However, this metabolic cost has not been demonstrated. In fact, in competition experiments, lysogens frequently won over isogenic non-lysogens suggesting that the possession of phages represents a selective advantage for the bacterial cell.

Bacteria, which contain prophage(s) encoding fitness factors useful for the host, will be positively selected, and some prophage genes are of obvious use for a lysogenic bacterium. The phage repressor, which is constitutively transcribed from the lysogenic state in order to prevent prophage induction, also prevents related phages from superinfecting the lysogenic cell (homo-immunity). In addition, many prophages express superinfection immunity genes active against other classes of phages, which prevent superinfection by heterologous phages (hetero-immunity). At the same time, these prophage functions protect the lysogen from phage infection and lysis, not a small selective advantage in a phage-infested environment. However, the advantage of being a lysogen goes beyond this protection against phage infection. Data from the 1970s showed that *E. coli* lysogenic with phage lambda displayed a marked growth advantage over non-lysogenic isogenic cells when grown *in vitro* under carbon-limited conditions. This physiological advantage was attributed to the lambda *rex* gene, which is involved in phage superinfection control by decreasing macromolecule synthesis. This growth advantage was also observed in *E. coli* lysogenic with temperate phages P1, P2 and Mu. Lambda lysogens also showed growth advantages over non-lysogenic *E. coli* in animal infection models and the effects were attributed to the lambda *bor* gene, which encodes a membrane protein that confers increased serum resistance to lysogenic *E. coli*.

Phage lambda is in that respect certainly not an exception, as similar experimental data was obtained for *Pseudomonas aeruginosa* and *Salmonella enterica* prophages suggesting that these prophages confer a selective advantage in animal infection models for lysogens over their isogenic non-lysogen counterparts (Fig. 2). Comparative prophage genomics provided additional evidence: a notable observation in phage genomics – both for temperate and virulent phages – is the possession of many hypothetical genes that are non-essential for *in vitro* phage replication. One might speculate that these viral genes encode accessory functions, assisting phage replication under limiting growth conditions. In the case of a prophage these (under *in vitro* conditions non-essential) genes might encode useful functions for the lysogen under certain physiological growth conditions. Selective forces will maintain these “non-essential” genes since they will increase both prophage and lysogen frequency in the bacterial population under stress conditions, which bacteria experience in their natural niche most of the time.

Bacteria profit from phages using them as efficient gene transfer particles for shuttling bacterial genes. Upon induction, the integrated prophage DNA excises from the bacterial chromosome. Occasionally, a faulty excision process leads to the incorporation of bacterial genes positioned adjacent to the attachment sites. These bacterial genes then become incorporated into the phage particle and are subsequently transferred to the next infected bacterial cell. For phages that integrate into specific bacterial chromosomal sites, only few bacterial genes are transferred (specialized transduction, such as for phage lambda). Transfer of bacterial genes is increased by Mu-like temperate phages that integrate randomly into the bacterial chromosome and package random bacterial genes into phage particles. Phages with a headful *pac*-site packaging mechanism can erroneously incorporate bacterial instead of phage DNA into phage capsids, thus transferring larger pieces of bacterial DNA to the next host (generalized transduction).

Recently, a very efficient form of bacterial gene transfer by phages was demonstrated in *S. aureus*. Several *pac*-site *S. aureus* prophages were induced into phage DNA replication while still being integrated in the bacterial genome, i.e., before excision of the prophage DNA. This led not only to an amplification of bacterial DNA surrounding the integrated prophage, but also to a sequential packaging of headful pieces of this amplified DNA into empty phage particles mediated by the phage terminase. In this way bacterial DNA placed up to seven genome equivalents away from the integrated prophage corresponding to several hundred kb of bacterial DNA could be efficiently transferred by transduction into recipient cells. The researchers called this prophage-initiated process lateral transduction, which was 1000-fold more efficient than generalized transduction achieved by infecting the cells with the same phages. Since *S. aureus* is frequently polylysogenic with several prophages scattered around the bacterial chromosome, lateral transfer could be a very efficient way for horizontal gene transfer of bacterial DNA from many chromosomal locations including *S. aureus* pathogenicity islands (SaPIs). Lateral transduction when confirmed with *pac*-site prophages in other bacterial systems could represent a highly efficient motor for horizontal gene transfer.

Phages are ideal gene transfer particles: the tightly knit capsid protects the incorporated DNA from physicochemical damage in the environment and the phage tail structures assure the recognition of target cells and the efficient injection of the DNA into the new host. In *S. aureus*, intricate interactions between pathogenicity islands and helper phages mediate virulence gene transfer. As long as the recognition process is limited to the same bacterial host species, phages mobilize bacterial DNA only within a given bacterial species. Under individually rare, but cumulatively sufficiently frequent, conditions, phages can infect across the bacterial species and genus borders and can also mediate cross-species or wider gene transfer. Phage-mediated gene transfer from *S. aureus* to *Listeria* was also experimentally demonstrated. The potential breadth of gene transfer across bacterial lineages is also suggested by the distinct GC-content of some extra phage genes – suggesting transfer from bacteria with different GC content – while the remainder of the phage genome shares the GC content of the bacterial host reflecting codon adaptation. This extra or more DNA in phages was dubbed “moron”, and it is found at specific places of phage genomes and frequently contains its own promoter and terminator to allow transcription during the lysogenic state. In this way, the lysogenic cell can acquire a new genetic trait (phenotype) compared to the parental non-lysogenic cell, leading to lysogenic gene conversion (LGC) after the incorporation of a new prophage.

Prophages can also lead to phenotypic changes of the lysogen by integrating into a host gene and inactivating it. Phages like Mu integrate randomly into the bacterial chromosome and phage Mu replication proceeds by further random copy-paste events into the bacterial chromosome mediated by the phage transposase. By this process, many bacterial genes can be inactivated, and mutant bacteria are created – in fact, the phage Mu derives its name from this mutational activity. Gene inactivation can be an adaptive process: in the medical literature cases are documented where the loss of specific bacterial genes is important for the

Prophage genes that affect *Salmonella* / *E. coli* virulence in the mouse model

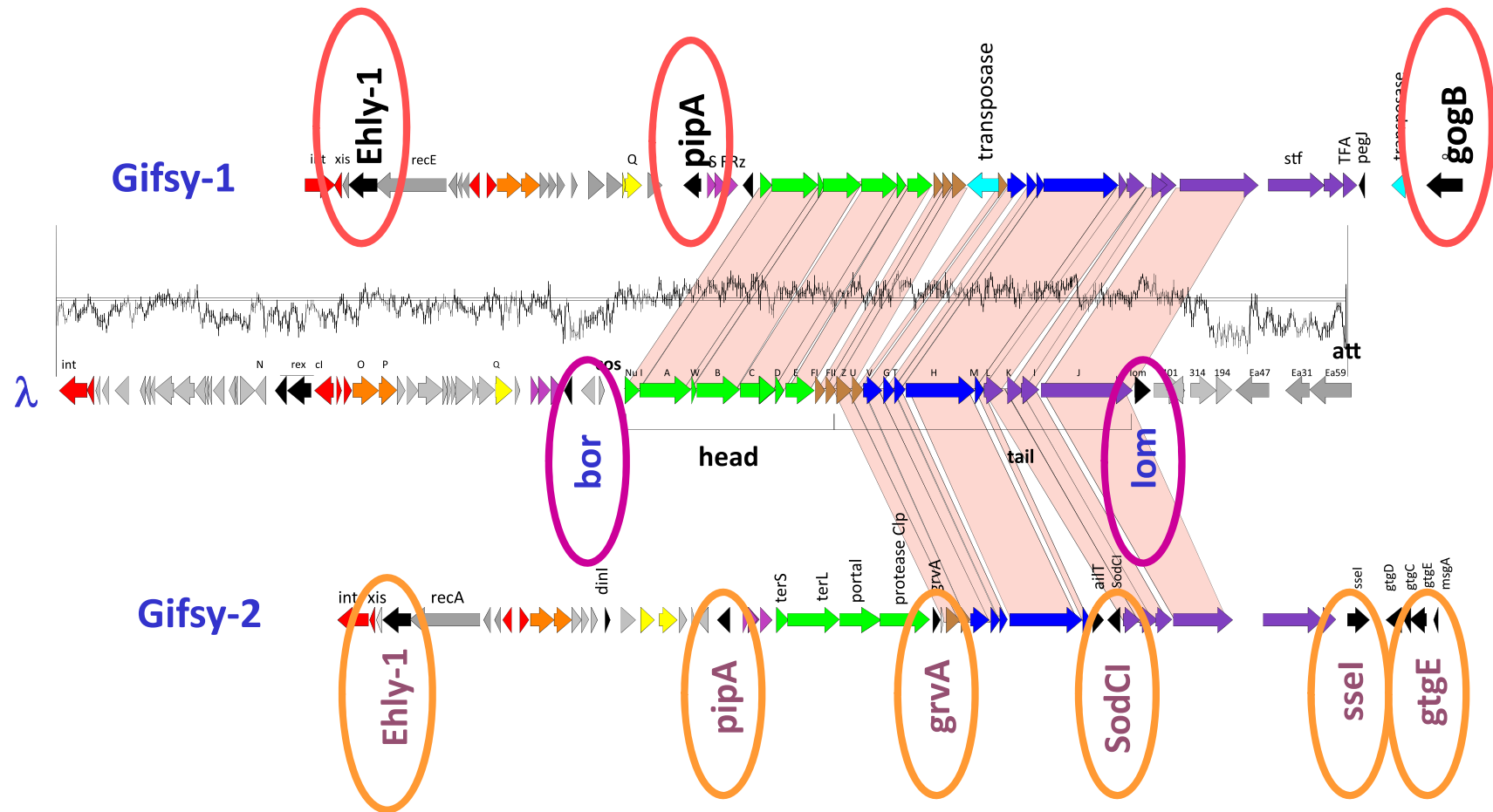


Fig. 2 *Salmonella* / *Escherichia coli*: Prophage genes that affect the virulence of the lysogen in mouse models. Gene maps of *Salmonella enterica* serovar Typhimurium prophages Gifsy-1 and Gifsy-2 were aligned with *E. coli* phage lambda. Sequence-related genes were linked by pink shading. The GC scan for phage lambda suggests distinct origins for the left and right arm of the lambda chromosome. The phage modules are color-coded: Lysogeny, red; DNA replication, orange; regulation, yellow; lysis, pink; head morphogenesis, green; head-tail-joining, brown; tail, blue; tail fiber, mauve; virulence, black (circled and annotated). Source: Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., Brüssow, H., 2003. Phage as agents of lateral gene transfer. *Current Opinion in Microbiology* 6 (4), 417–424.

development of a pathogen (e.g., *Shigella*). Phage Mu also encodes an invertase that inverts the orientation of a phage tail fiber gene to extend the host range of Mu. Such an invertase might also be of use for inverting adjacent bacterial genes. Inversions of large bacterial chromosomal segments were conspicuously flanked by prophage sequences (*Streptococcus pyogenes*, *Xylella fastidiosa*, *E. coli*, *Bartonella*). Therefore, prophage sequences might here have served as cross-over sites for phage or cellular recombination enzymes. In *Bartonella*, chromosome inversions might permit temperature adaptations to the human-flea host change.

Interest in phage-mediated gene transfer grew substantially when bacterial genomics and phylogenetic analyses revealed strong elements of horizontal evolution overlaying and frequently obscuring the vertical evolution of bacteria along a classical phylogenetic tree. Phages belong to mobile DNA elements that accelerate bacterial evolution by allowing rampant horizontal gene transfer. Comparative bacterial genomics showed that beyond the bacterial core genome (shared by most bacteria within a given bacterial species), the accessory bacterial genes (found only in some bacterial isolates from this species) comprised a large proportion of mobile DNA, and particularly of prophage DNA. Phage DNA transfer might thus represent a motor for bacterial evolution only comparable, if not superior to, the role of sexual reproduction in the evolution of higher organisms (Fig. 3).

It seems that the transfer of lysogenic conversion genes (LCGs) by prophages has played a particularly prominent role in the evolution of bacterial pathogens. One can even argue that prophages made certain bacteria pathogens (e.g., *Streptococcus pyogenes* (Figs. 4–6), *Staphylococcus aureus*, *Vibrio cholerae*, *Clostridium botulinum*).

This predilection of phages for pathogens might be explained by their evolutionary history. In early evolution, prokaryotes experienced predation only from prokaryotic viruses. When protists evolved, many eukaryotic cells and small metazoa made a living by grazing on bacteria. Bacteria were thus trapped between two predators. In a stone-scissor-paper game some bacteria might have teamed up with phages that had carried toxin genes. Toxin-producing lysogenic bacteria could thus kill protists that engulfed them. The killed predator released a lot of nutrients, which then fed the former prey, i.e., the lysogenic bacterium, furthering their abundance in the population. When protists evolved into animals, existing lower animals like sponges repurposed bacteria-eating amebocytes as defense systems to protect them from microbial invasion. This primitive cellular immune system evolved into phagocytes patrolling the bodies of higher animals to engulf and eliminate bacterial invaders. The coalition between bacteria and temperate phages could once again pay off when bacteria explored the life style of exploiting animals as a food source evolving either as commensals or pathogens. This hypothesis could explain why a frequent scheme of bacterial pathogenesis is the survival of bacterial pathogens in phagocytes (the functional followers of sponge's amebocytes in animals), why a significant proportion of bacterial pathogens have prophage-encoded virulence genes, and why some virulence factors from animal pathogens are still active against unicellular eukaryotes. In the stone-scissor-paper game other alliances are possible and they are apparently realized. There is increasing evidence that phages have a major impact on the composition and dynamics of the gut microbiota and since the latter are thought to exert a major impact on the health of the animal host, phages might also indirectly impact on human health. Indeed, phages were proposed as an infectious layer in the mucus overlaying the gut epithelia protecting the human body together with secretory immunoglobulins against pathogen invasion. Notably, some phages display immunoglobulin-like folds in their capsid proteins, which might be responsible for interaction of phages with the mammalian immune system and mucus glycoproteins.

Phage-bacterium interactions have many other evolutionary facets. For example, virulent phages from cyanobacteria encode important genes for the photosynthetic apparatus. Their adaptive value might be providing extra energy to the infected cell by replacing readily bleached components of the photosystem in order to maintain a high ATP level that favors phage replication. The nonsystematic phylogenetic distribution of photosynthetic capacity in bacteria suggests a role for horizontal gene transfer in the evolution of bacterial photosynthesis, possibly mediated by phages. If confirmed, phages could have contributed to the spread of one of the major inventions in the evolution of life on earth.

Lysis-Lysogeny Decision: An Ecophysiological Perspective

The classical phage lambda attracted researchers for its lytic-lysogenic lifestyle decision. Spontaneous prophage induction had traditionally been explained as a stochastic process under the influence of biochemical noise. The dramatic cell-to-cell variation in phage burst size demonstrated in the classical fluctuation test was used as an argument for such an interpretation. Instead of the non-deterministic approach to this life style decision, microbiologists, who exploited state of the art single cell analytical methods, now started to unravel the hidden variables driving this decision outcome. This analysis demonstrated large cell-to-cell variation in the transcription and protein expression of the lambda CI repressor that correlated with the switch to lysis decision. Phage infection is now also morphologically investigated at single cell resolution, demonstrating new layers of complexity, such as the location and migration of the phage on the cell surface and the interaction of the injected phage DNA with the chromosomal bacterial DNA.

An interesting recent model, dubbed the “Piggyback-the-Winner” model, proposed that lysogeny predominates at high microbial abundance and high bacterial growth rate. This conclusion was based on metagenomic virome data analyses from coral ecosystems and a meta-analysis of published viral and microbial densities in various ecosystems. It provides an opposing view to the “Killing-the-Winner” model of phage infection and challenges the classical observations where cell starvation – indicating low nutrient availability and thus low chances for lytic propagation – predicts lysogeny to become the preferred strategy. Current data indicates that the number of infecting phages, the volume of the infected cell and the position of the phage on the infected cell surface, influence the infection outcome with phage lambda. For a *Bacillus* phage, a phage-encoded system resembling quorum sensing mechanisms was described and called “arbitrium” system. It consists of a 6 amino acid peptide, an intracellular peptide

receptor and a negative regulator of lysogeny, all encoded by the *Bacillus* phage. This system allows phage-to-phage communication facilitating the best replication strategy. Such observations will pave the way to a sociobiology of phages.

There is also some evidence that lysogeny could become “autocatalytic” within a given cell population. *Salmonella enterica* frequently contains multiple and replication-competent prophages. If such lysogenic cells are mixed with non-lysogenic *S. enterica* cells, spontaneously induced prophages will kill a few lysogens and infect and kill susceptible non-lysogenic *S. enterica* until some become also a lysogen and thus resist further lytic infection. In this way, lysogenic cells can become the dominant bacterial population, at least under *in vitro* conditions. The prophage behaves here as a typical selfish genetic element. Prophages have also recruited genetic elements that stabilize them against deletion. A typical case is the toxin-antitoxin gene cassette. It expresses a longer-lived lethal cell toxin that is neutralized by a shorter-lived antitoxin. Should the cell delete the prophage, it runs out of the antitoxin and will be killed by the persisting toxin. The prophage is thus protected against its loss. However, *in vivo*, the situation is frequently more complicated, and a prophage can become a burden. For example, in the nose *Streptococcus pneumoniae* produces H₂O₂ as a chemical weapon against competing bacteria. This peroxide induces prophages in *Staphylococcus aureus*, killing not only the lysogen but, via the induced prophage, also killing susceptible non-lysogenic *S. aureus*. In nature the phage-bacterium interaction is therefore much more complicated due to a multitude of phage-bacterium, bacterium-bacterium, phage-phage and last, but not least, bacterium-host interactions, which we are just beginning to study in the human gut microbiome and the phyllosphere (the microbiome around plants is even more complex in the soil environment). In these data sets, we see the impact of phages on bacteria through the magnifying glass of ecophysiology.

Coevolution: Polylysogeny

The impact of phages on bacterial evolution can be studied by analyzing the footprints of evolution and selection left in bacterial genome sequences. The survey of 58 *E. coli* and 27 *S. enterica* genomes revealed a mean of 8.2 and 5.6 prophages per bacterial genome, respectively. Polylysogeny, the possession of multiple prophages, documents the possibility of the coexistence of different phages despite the phage-encoded superinfection exclusion mechanisms. This creates complex interactions between prophages within a lysogen: one prophage can for example inhibit the excision of another prophage as seen in *Enterococcus*. Polylysogeny also has interesting consequences for the bacterial cell. In this way, different combinations of lysogenic conversion genes provided by distinct prophages can be combined allowing combinatorial associations of different phage-encoded fitness factors, thus creating a much greater genotypic and phenotypic diversity in poly-lysogens. This permutation principle is actually exploited by several bacterial pathogens. For example, in *S. aureus* variable prophage combinations in polylysogens allow the exploration of different combinations of phage-encoded superantigens to flexibly distract the immune system of the host. In *S. pyogenes*, various combinations of prophages contribute to an astonishingly wide spectrum of clinical disease manifestations. Polylysogeny in *S. pyogenes* thus facilitates the rapid evolution of new pathogenic traits within a decade, as demonstrated by “flesh eating” infections (*S. pyogenes* fasciitis), or the changing epidemiology of scarlet fever depending on the prevalence of the toxin-encoding prophage in the bacterial population. In *Vibrio cholerae*, the cholera toxin is prophage-encoded, and the pathogenicity of this bacterium is the result of a complex physical interaction between three phages.

Coevolution: Prophage Remnants

In the case of the 58 *E. coli* and 27 *Salmonella* genomes most prophages could be attributed to lambdoid phages (59%) followed by P2-like phages (13%). However, this is not a general pattern: *Pseudomonas aeruginosa*, a more distant relative of these bacteria, showed a substantial contribution of additional temperate phages like Mu-like phages and filamentous phages. Filamentous phages cause chronic non-lytic infections in *E. coli*, but do not integrate into the bacterial chromosome because they lack an integrase. In *V. cholerae* filamentous phages have recruited bacterial enzymes for chromosomal integration, while in *Pseudomonas* filamentous phages encode a integrase/recombinase in the small phage genome, possibly explaining the high prevalence of lysogeny with these phages in *P. aeruginosa*.

In the 58 sequenced *E. coli* strains, the genome size of lambdoid prophages was bimodal with a major peak at 40–50 kb genome size corresponding to that of infectious extracellular lambdoid phages and a second peak at about 10 kb. The latter was interpreted as defective lambdoid prophages. The large infectious prophages probably represent recently integrated prophage DNA, while the small

Fig. 3 *Escherichia coli*: Prominent prophage content. The prophages (red rectangles) in the genomes of four different *E. coli* strains. Inner to outer circles represent the food pathogens enterohemorrhagic *E. coli* O157:H7 Sakai and O157:H7 EDL933 strains, the laboratory *E. coli* strain K-12, and the uropathogenic *E. coli* strain CFT073. The prophages of *E. coli* O157:H7 strain Sakai are annotated as lambda, Mu, P2 or P4 like phages (a) A dot-plot for DNA sequence sharing between the lambdoid phages of the Sakai strain (b) Genome maps for four lambda-like Sakai prophages with the integrase gene (red) to the left. Virulence genes are colored black and are annotated. The region around the *lys* gene is boxed in grey (c) *Source*: Modified from Canchaya, C., Proux, C., Fournous, G., Bruttin, A., Brüssow, H., 2003. Prophage genomics. *Microbiology and Molecular Biology Reviews* 67 (2), 238–276. Boyd, E.F., Brüssow, H., 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends in Microbiology* 10 (11), 521–529.

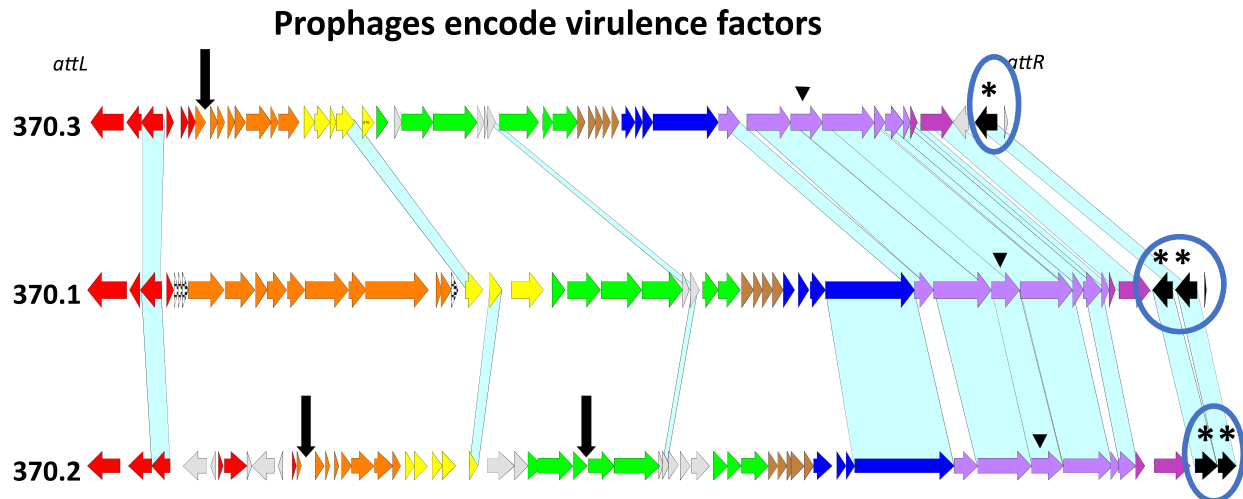


Fig. 4 *Streptococcus pyogenes* prophages: Pathogenicity by the permutation principle. Genome maps of prophages 370.1 to 3 from *S. pyogenes* strain SF370 are displayed. The left and right attachment sites *attL* and *attR* mark the transition from the prophage into the bacterial chromosomal DNA. The phage modules are color-coded. Red: lysogeny; orange: DNA replication; yellow: transcriptional regulation; green: DNA packaging and head; brown: head-to-tail; blue: tail; mauve: tail fiber; pink: lysis; black: lysogenic conversion genes—superantigens and mitogenic factors (*) contributing to the virulence of the lysogenic bacterium. The phage hyaluronidase (triangle) is likely used by both the phage and the pathogenic host. Putative prophage inactivating mutations are marked by arrows. *Source:* Brüssow, H., Hendrix, R.W., 2002. Phage genomics: small is beautiful. *Cell* 108 (1), 13–16.

Prophage DNA determines strain specificity

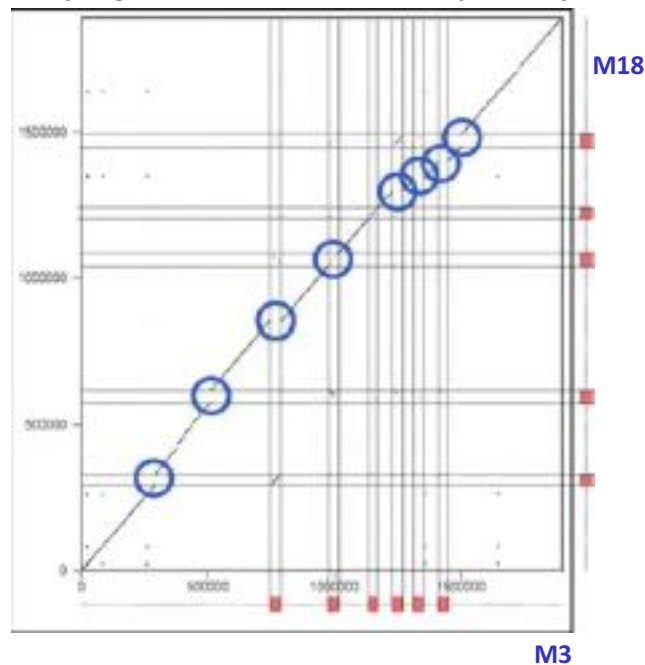


Fig. 5 *S. pyogenes*: Prophage DNA determines strain specificity. Genome alignment of two *S. pyogenes* strains belonging to different M types. Gaps in the alignment (blue circles) correspond to prophage DNA (red rectangles) that are either missing in one strain or the other or are found at a different genome position. Modified from Boyd, E.F., Brüssow, H., 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. *Trends in Microbiology* 10 (11), 521–529.

prophages were interpreted as remnants from initially complete prophage genomes created by a rapid viral DNA deletion, suggestive of a prophage domestication process. This prophage reduction process was interpreted as a consequence of the trend to maintain a small given genome size in bacteria. If such prophage reduction mechanisms did not exist, bacterial genomes would over evolutionary time periods be overrun by selfish DNA invasion (something what seems to have happened in higher plants and animals). However, this deletion process does not seem to be random. Some directionality seems to apply (integrases and repressor genes appear to be

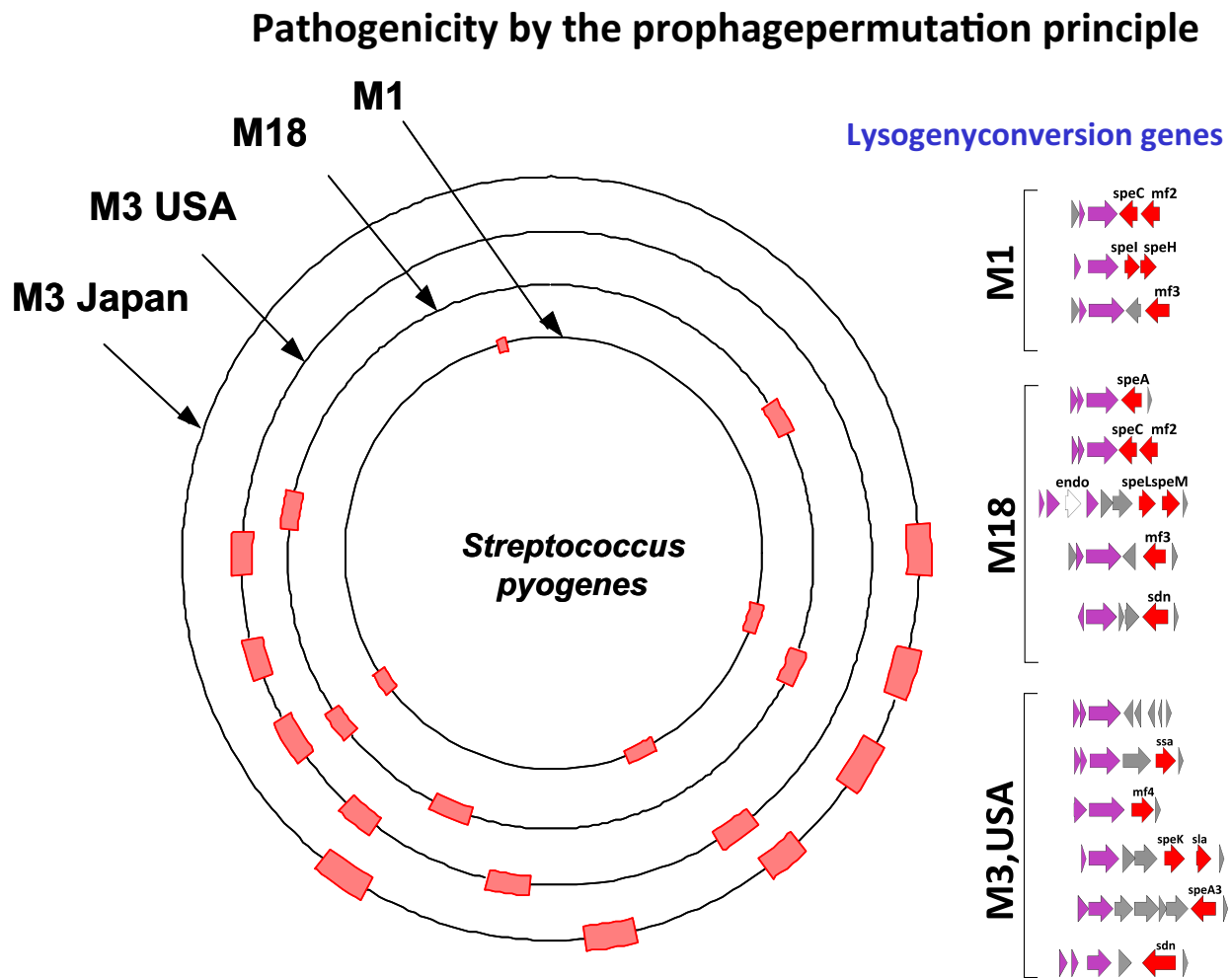


Fig. 6 *S. pyogenes*: Prophage DNA determines bacterial virulence profile. Prophages from *S. pyogenes* encode many potential virulence factors. Left: Prophages are visualized as red boxes on the circular genome maps of four sequenced *S. pyogenes* strains representing three different M types. Right: Prophage genes near the prophage *attR*. Genes of the lysis module are colored in mauve; candidate lysogenic conversion genes are colored in red and are annotated: mf, mitogenic factor; sdn, streptodornase; sla, streptococcal phospholipase A₂; spe, streptococcal pyogenic exotoxins, ssa, streptococcal superantigen. All red genes represent possible virulence factors. Gray arrows represent genes of undetermined function. Source: Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L., Brüssow, H., 2003. Phage as agents of lateral gene transfer. Current Opinion in Microbiology 6 (4), 417–424.

enriched in prophage remnants) and phage genes of selective value to the bacterial host are likely to be conserved. When lysogens and non-lysogens were competing, the lysogen frequently outpaced the non-lysogen, suggesting a strong selective advantage, and this observation also applies to prophage remnants. Indeed, some prophage remnants of *E. coli* K12 have a long evolutionary history and it was experimentally demonstrated that they confer help in stress situations like acid or oxygen exposure, or help in coping with the stress of antibiotic exposure or in biofilm growth. In line with the selective value of these prophage remnants, the purge of non-synonymous mutations in these elements suggests that evolution tries to maintain the gene function in these remaining prophage genes. Genes from defective prophages thus contribute to the fitness of the bacterial hosts well beyond the known phenomena where isolated phage tail structures serve as bacteriocins, or phage capsids evolved into gene transfer agents.

Coevolution: Integration Sites

The coevolution of bacteria and their phages is also documented by the integration sites of temperate coliphages. Prophages integrate into conserved hotspots of the *E. coli* chromosome and frequently target intergenic regions to avoid interference with bacterial gene expression. Prophages are also inserting non-randomly with respect to the large-scale bacterial chromosome organization: when plotting prophage integration sites on the circular *E. coli* chromosome, a clearly increasing origin-to-terminus gradient of integration sites was observed. Prophage integration into the highly expressed genes around the origin of replication

would lower bacterial fitness and are thus counterselected. A possible solution to this interference problem is provided by temperate phages that maintain their genome as an extrachromosomal DNA and assure their maintenance by a toxin-antitoxin gene system like coliphage P1. This extrachromosomal phage lifestyle was considered to be rare, but recent data from *S. aureus* showed that extrachromosomal phage DNA in latently infected cells is more prevalent than initially thought. However, prophages do not always avoid integrating into actively transcribed regions, resulting in gene inactivation with deleterious effects if important genes are disrupted. Recently, it was proposed that such inactivating prophage integration can also be used as a regulatory switch where transcription resumes when the prophage is excised into an episomal stage without leading to cell lysis. This “active lysogeny” might come into play when commensal or pathogenic bacterial cells receive signaling molecules within animal cells (phagocytosis) or in animal hosts (gut). Prophages might also sometimes lead to large scale chromosomal rearrangements when either cellular RecA-like or phage-encoded Rad52-like enzymes mediate recombination across two prophages sharing some sequence identity, but located in distinct chromosomal sites.

Coevolution: In the Patient

Coevolution between phages and bacteria is not only seen from comparative genomics, highly dynamic phage-bacterium relationship changes were also documented over much shorter time periods. In microcosm laboratory experiments *Pseudomonas fluorescens* evolving under phage pressure showed an increased mutation rate and evolved greater genome wide divergence particularly over the LPS gene cluster, the receptor of many phages. The evolved bacteria showed also greater fitness when tested in the absence of phages against the original strain. In *S. pyogenes* a phage integrates into the DNA mismatch repair genes *mutS-mutL*, increasing the mutation rate of the bacterium 200-fold. The cell uses this phage integration as a molecular switch, since the prophage is excised at low cell density restoring normal mutation rates. Short term phage-induced evolution was also seen in cholera patients that showed, in addition to *V. cholerae*, also a vibriophage in the stool. Bacterial colonies isolated from individual cholera patients were practically all resistant to this vibriophage and isogenic to the infecting strain except for mutations in an outer membrane porin and a ToxR signaling protein, which activates virulence gene expression under host environmental stimuli. These ToxR mutants were attenuated in mouse infection model. Apparently, *V. cholerae* evolves under phage pressure to lesser virulence in the later course of human cholera infection. Notably, evolution of the cholera pathogenicity was also described over the course of an individual cholera outbreak and also linked to an increase in phage prevalence. The development of significant environmental phage titer might actually end a cholera season.

Other instructive examples are provided by comparing *S. aureus* colonizing the nose of healthy subjects or the lungs of cystic fibrosis patients. The frequency of genome alterations was significantly higher in the patients than in the controls. In nearly half of the patients, the genome alterations were linked to prophage mobilization, mostly by integration into a single bacterial gene. Phage translocation in the patients leads to a splitting of the bacterial population into various subtypes differing in virulence gene composition, each of which might have different selective advantages for the pathogen in the patient. Phage mobilization seems to be induced by the frequent antibiotic treatment in the patients.

Coevolution: Defense Mechanisms

Coevolution of phages and bacteria is perhaps best documented by the analysis of bacterial defense mechanisms against phages and the countermeasures of phages against this control. This genetic arms race is apparently a major driver of both bacterial and phage evolution. Bacteria use several anti-phage strategies (receptor modification preventing phage adsorption; restriction-modification systems destroying invading phage DNA; abortive infection mechanisms, where a variety of seemingly altruistic cell death processes block phage replication). The latest addition in discovery, but not in evolutionary antiquity is the bacterial CRISPR-Cas system. All these anti-phage systems are characterized by a stunning genetic diversity on the bacterial side (LPS, R-M, *abi*, CRISPR-Cas diversity), illustrating the difficulty to solve the challenge posed by phages. The reason is of course the mutability of the phages, which evolved counter-mechanisms against any of the bacterial defense mechanisms. The details of this arms race can be astonishingly multi-layered. To illustrate this with two examples: phages can escape the control of the CRISPR-Cas system by acquiring point mutations in regions of their genomes that match spacer sequences. Bacteria can then “update” the CRISPR-Cas system by incorporating modified spacers. Molecularly more complex are anti-CRISPR genes located in a specific genome region of Mu-like *Pseudomonas* phages encoding a large variety of small phage proteins inhibiting this bacterial anti-phage system. The diversity of mechanistic action of these anti-anti-phage proteins might be a mirror image of the stunning molecular diversity of CRISPR-Cas systems in some bacteria. A bacterial anti-anti-phage device is suspected in this system but was already demonstrated in another anti-phage system. In order to escape R-M system control phages were selected that either avoid restriction sites or modify their bases: in coliphage T4 cytosine gets a hydroxymethyl addition or a further glycosyl derivation, preventing their recognition by restriction enzymes. *E. coli* evolved restriction enzymes, e.g., GmrSD, that can cleave these modified bases. T4 evolved the internal capsid protein ISI*, which is co-injected with phage DNA, binds and inhibits GmrSD. *E. coli* evolved a GmrSD variant that resists ISI* binding. Furthermore, mutant T4 exist that escape this GmrSD variant. This tit-for-tat aspect seems to be a characteristic of the bacterial-phage coevolution in a never-ending process.

Coevolution: War and Peace

Antagonistic scenarios were observed with a single phage and bacterium in chemostats and depending on the specific coliphage used, different outcomes were described, but phages often lost the race. However, this *in vitro* context excludes external factors (competition from other bacteria, cost of resistance development for bacteria, genetic support from other phages), which will strongly impact the outcome of such a race in nature. In fact, one might even speculate that the balance of selective forces for this battle favors not the supremacy of one or the other partner of this coevolution game, but both work as drivers of their coevolution. Phages are challengers that keep bacteria under pressure to evolve, but phages have no selective interest in winning this race against bacteria. Since phages depend on the bacterial cytoplasm for their replication, this would be a Pyrrhic victory culminating in the phage's extinction. A truce is not possible for a phage as an infectious agent as it must lyse a cell to be retained in evolution. Lysogeny is in some way an escape lane for peaceful coexistence, but it has its limits. In *E. coli* so far, an upper limit of about 20 prophages was described in isolates with a moderately expanded genome size. Even in *Pseudomonas*, which sports with 6 Mb, an unusual large bacterial genome, relaxing control via CRISPR-Cas and R-M systems was only associated with a 300 kb increase in genome size. There are apparently mechanisms that maintain small genome sizes in bacteria, preventing the accumulation of viral-like DNA sequences in the host genome, in contrast to higher animals and plants with retrovirus-like elements where genome sizes are wide-ranging and apparently uncorrelated with organismal complexity as described by the classical C-value paradox. While bacteria could theoretically win the arms race with phages, there is apparently no incentive to eliminate phages. Literature data consistently demonstrates that prophages are of selective value for bacteria, driving phages to extinction in a given bacterial group might compromise their competitiveness in their ecological niche. This argument can be easily extended to lytic phages as a motor for accelerated evolution of bacteria allowing lateral gene transfer by generalized phage transduction and phage-derived gene transfer agents. It thus seems that the paradoxical inefficiency of both bacterial phage-defense systems and phage anti-defense systems have an inborn limitation allowing some, but deliberately not total control of the other player in this coevolution game. The role of phages in bacterial evolution is thus likely to be enormous and will be complemented by other mobile genetic elements to fuel lateral gene transfer in bacteria. Since phages in contrast to plasmids, transposons, insertion elements and the like kill the bacterial cell, phages might play a unique role in bacterial evolution. There are differences with respect to what fitness factors are contributed to bacteria by mobile DNA. While plasmids and transposons are privileged carriers of antibiotic resistance genes (ARG), phages are privileged carriers of virulence genes. There is conflicting data on how frequently phages carry ARG: metagenome data demonstrated a substantial load of ARG in the virome fraction particularly after antibiotic treatment of the host, while the genomes of cultivated phage rarely show ARG. The data are not necessarily mutually exclusive, but phage genomes only rarely contain biologically proven ARG. Whatever the outcome of this controversy, phages that were in the past commonly depicted as parasites of bacteria have with recent research transmogrified into, albeit difficult, partners of bacteria, and necessary elements for diversity in bacterial ecology and evolution.

Further Reading

- Barr, J.J., Auro, R., Furlan, M., *et al.*, 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proceedings of the National Academy of Sciences of the United States of America* 110, 10771–10776.
- Bobay, L.M., Touchon, M., Rocha, E.P., 2014. Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12127–12132.
- Brüssow, H., 2007. Bacteria between protists and phages: From antipredation strategies to the evolution of pathogenicity. *Molecular Microbiology* 65, 583–589.
- Brüssow, H., Canchaya, C., Hardt, W.-D., 2004. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews* 68, 560–602. (table of contents).
- Chen, J., Quiles-Puchalt, N., Chiang, Y.N., *et al.*, 2018. Genome hypermobility by lateral transduction. *Science* 362, 207–212.
- Gutierrez, R., Markus, B., Carstens Marques De Sousa, K., *et al.*, 2018. Prophage-driven genomic structural changes promote *Bartonella* vertical evolution. *Genome Biology and Evolution* 10, 3089–3103.
- Kang, H.S., Mcnair, K., Cuevas, D., *et al.*, 2017. Prophage genomics reveals patterns in phage genome organization and replication. *bioRxiv*. Available at: <https://www.biorxiv.org/content/10.1101/114819v1>.
- Lindell, D., Jaffe, J.D., Coleman, M.L., *et al.*, 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449, 83–86.
- Novick, R.P., Ram, G., 2017. Staphylococcal pathogenicity islands-movers and shakers in the genomic firmament. *Current Opinion in Microbiology* 38, 197–204.
- Rands, C.M., Starikova, E.V., Brüssow, H., *et al.*, 2018. ACI-1 beta-lactamase is widespread across human gut microbiomes in Negativicutes due to transposons harboured by tailed prophages. *Environmental Microbiology* 20, 2288–2300.
- Roux, S., Enault, F., Hurwitz, B.L., Sullivan, M.B., 2015. VirSorter: Mining viral signal from microbial genomic data. *PeerJ* 3, e985.
- Samson, J.E., Magadan, A.H., Sabri, M., Moineau, S., 2013. Revenge of the phages: Defeating bacterial defences. *Nature Reviews Microbiology* 11, 675–687.
- Seed, K.D., Yen, M., Shapiro, B.J., *et al.*, 2014. Evolutionary consequences of intra-patient phage predation on microbial populations. *Elife* 3, e03497.
- Touchon, M., Bernheim, A., Rocha, E.P., 2016. Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal* 10, 2744–2754.

Viruses and Their Potential for Bioterrorism

Dana L. Vanlandingham, College of Veterinary Medicine, Kansas State University, Manhattan, KS, United States

Stephen Higgs, Biosecurity Research Institute, Kansas State University, Manhattan, KS, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S.A. Morse, R.F. Meyer, Viruses and Bioterrorism, In Reference Module in Life Sciences, Elsevier Inc, 2017, doi:[10.1016/B978-0-12-809633-8.11007-6](https://doi.org/10.1016/B978-0-12-809633-8.11007-6).

Glossary

Agro-terrorism A subset of bioterrorism, defined as the deliberate introduction of animal or plant pests (e.g., bacteria, viruses, fungi) with the goal of generating fear, causing economic damage, and/or undermining social stability (Keremidis *et al.*, 2013).

Bio-crime The threat or use of a biological agent for murder, extortion, or revenge (Carus, 2002).

Biological warfare (synonymous with biowarfare)

A specialized type of warfare involving the use of biological

agents conducted by a government against a target (human, agriculture, or infrastructure) (Carus, 2002).

Bioterrorism The threat or use of a biological agent (or toxin) against humans, animals, or plants by individuals or groups motivated by political, religious, ecological, or other ideological objectives (Carus, 2002).

Weapons of mass destruction Weapons that can destroy entire cities, regions, etc. (Merriam-Webster dictionary).

Introduction

The idea of using non-traditional weapons for offensive purposes is not new. Using chemical and biological agents as well as toxins to win wars has a long history. These agents and toxins have also been used by various groups such as military personnel, and governments or non-governmental organizations against civilians and individuals to incapacitate or cause terror in a population. In addition to the use of these weapons against people, another target of these agents and toxins is agriculture, which can trigger severe economic and societal repercussions. Not all utilization of these agents and toxins was to cause destruction; these agents have also been used with the intent of benefitting society, agriculture, or the environment by biologically controlling pests. Whatever the target, intention, and mode of introduction of these agents and toxins, the effects can be long lasting and devastating, as can be shown by several instances of intentional and natural introductions of these agents.

Unconventional Bioweapons

Venomous animals, poisonous plants, and toxins have been used over the centuries as offensive weapons. A common goal of these attacks is often to inhibit or reduce the enemy's ability and capacity to fight. These objectives could be met by killing enemy combatants, incapacitating them through illness, or denying them access to needed resources such as a water supply. The association between illness and fatalities from toxins, primarily through the consumption of poisonous plants, or from the bite or sting of venomous animals or insects must have been obvious long ago. The ability to use these methods to kill on an individual or large scale, such as an army, was used very early in our history. For example, poison could be used to enhance a weapon's immediate effectiveness by adding it to the tip of arrows. Poison is also useful for killing an enemy from a safe distance, in terms of both space and time. Poisoning water supplies, such as wells, has been an effective offensive tactic since as early as the 6th century BC when the fungus, ergot, which grows on rye and other grasses, was used by Assyrians to poison wells of their enemies (Bradley and Sustaita, 2013). Another example of this took place in 600 BCE when roots of hellebore, a poisonous flowering plant in the buttercup family (Romano Jr. 2007), were added to the water supplies to cause diarrhea in the enemies of Solon of Athens. Later, the Romans used the carcasses of dead people and animals to poison water sources used by their enemies (Florus, 1852; Lesho, 1998). In 190 BCE, Hannibal developed another unconventional bioweapon which consisted of pots filled with venomous snakes (Campbell, 2014). These pots were successfully used by launching them onto enemy ships during a naval battle causing the sailors to fight on two "fronts", the snakes and the invading sailors (Nepos, 2015).

Use of Diseases in Early History

Humans began using weapons for what we now consider biological warfare long before they knew what the toxins were or the etiologic agents for any of the diseases. Even without the detailed information that we now have on various disease-causing agents, people were able to use these agents, which spread from one person to another, for offensive goals. Bacterial diseases such as plague and viral diseases such as smallpox were notable for their potential use against an enemy because of their rapid spread and severity.

Widespread and devastating outbreaks of plague occurred unintentionally during warfare as a result of troop movements. This may have been the catalyst for the idea of using plague as a bioweapon. In much the same way as the spear-head dipped in poison enhanced the arrow, the use of the catapult combined with disease causing agents, was used to enhance siege weapons. The development of relatively long-range siege weapons, such as the catapult, enabled armies to use the catapult and plague infected bodies to improve the devastating outcome of their weapons. Bodies of victims who died of plague were placed on the catapult and flung into enemy strongholds and cities where, it was hoped, an outbreak would occur. This effective tactic was first used by the Mongols during their conquest of Asia and Europe in the 14th Century (Wheelis, 2002) and was perhaps last used by Peter the Great in the 1700s (Riedel, 2004). This tactic was also used to spread smallpox into cities under siege.

Spread of smallpox has been used throughout the centuries to kill enemies and civilians, even though it has been only since the end of the 19th Century that we recognized viruses as disease causing agents (MacLachlan, 2017). The highly infectious disease, smallpox, is caused by the virus *Variola major*. It is an ideal bioweapon and is perhaps the viral equivalent of plague. The virus is easily spread from an infected person to a susceptible person via aerosols. Once the person becomes infected, they are highly likely to become sick and suffer a greater than 30% chance of dying from the disease. In addition to the high morbidity and mortality rates, the disease causes fearful symptoms which add to the population's terror. This virus had a devastating effect on the indigenous population when the Spanish conquistadors came to the New World in the 16th century. The dissemination of smallpox by the arrival of approximately 300 conquistadors caused the deaths of about 90% of the indigenous population within a relatively short period of time.

The intentional use of smallpox against an enemy occurred about 150 years later when one of the commanders of British forces fighting in "Pontiac's Conspiracy", General Sir Jeffrey Amherst, suggested the distribution of smallpox-contaminated blankets to an unsuspecting coalition of Indians in 1763 (Eitzen Jr and Takafuji, 1997; Patterson and Runge, 2002). Because of the obvious advantages of using smallpox against an enemy, smallpox was again used as a bioweapon to cause the spread of disease during the American Revolutionary War (Hopkins, 1983; Meyer and Morse, 2016).

20th Century State Sponsored Biological Weapons Programs

Although the use of gas or poisonous smoke was first developed in China in approximately 400 BCE (Mayor, 2008) it was not until the First World War that gas was used on a large scale to clear enemy trenches (Fitzgerald, 2008). From 1914 to the end of the war in 1918 there was a substantial increase in the use of unconventional weapons, with primarily chemical weapons being the unconventional weapons of choice. These chemical weapons included tear gas, chlorine, phosgene, and mustard gas. There are limited reports of the use of biological weapons during this time period; however, one example cited is Germany's operation to use biological weapons to spread the bacteria which causes glanders amongst Allied Army horses and mules (Eitzen Jr and Takafuji, 1997; Kettle and Nicoletti, 2014).

Horror from the use of these weapons was realized after the end of the First World War and in 1925, the Geneva Protocol was signed to prohibit the use in war of "asphyxiating, poisonous or other gases and of bacterial methods of warfare" (UNODA, 2019). In addition to not addressing the use of viruses as weapons, another major drawback of this protocol was that it did not forbid the "production, storage, testing and transfer" of these weapons, nor did it prohibit the use of these weapons for retaliation if a country were to be attacked using chemical or bacterial weapons. Even though the Geneva Protocol had obvious shortcomings, it was an important first step by various nations to address the need to control these weapons.

The 1925 Geneva Protocol did not stop the use of state sponsored biological weapons development in every country. From 1932–1945, around the time of the Second World War, Japan developed a sophisticated biological weapons program that included bacterial and viral pathogens including: plague, anthrax, smallpox, and yellow fever (Riedel, 2004). Some of these weapons were tested on thousands of people in Japanese-occupied China (Eitzen Jr and Takafuji, 1997). Although viruses were used for these experiments, the methods needed to propagate viruses on a relatively large scale and the ability to preserve infectivity for weaponization was not well developed at this time. Therefore, bacterial agents were predominately used as bioweapons through the early 20th Century.

Many nations began to study biological weapons in order to defend against them and to use them in retaliation, if needed. During the time around the Second World War and into the Cold War there were at least 11 countries that developed state-supported offensive biological weapons programs. The countries with offensive weapons programs included: Canada, France, Germany, Hungary, Iraq, Japan, Poland, South Africa, Soviet Union, United Kingdom, and the United States (Millett and Snyder-Beattie, 2017).

The United States and the United Kingdom both participated in the study of potential biological agents, and in some cases, they worked on the development of agents that could be used for biological warfare. A U.S. program was considered in 1941 by Secretary for War, H.L. Stimson, headquartered at Camp Detrick, Fredrick, Maryland in 1943. The program was expanded to a production facility in Terre Haute, Indiana and a test center at Dugway, Utah; both established in 1944. In the 1950s, a bioweapons program in the United States primarily focused on toxins and bacterial agents. However, the U.S. was interested in the study of viruses as potential bioweapons. There were a number of different viral agents studied by the U.S. including: Argentine and Bolivian hemorrhagic fever viruses, Hantavirus, Lassa fever, smallpox, yellow fever, and Western equine encephalitis virus. Of these viruses studied, only Venezuelan equine encephalitis virus was thought to be useful for weaponization.

The UK Biological weapons program, began in 1916 and was primarily based at Porton Down (Hammond and Carter, 2002) with Gruinard Island used as a testing site (Roffey, 2002). Due to the contamination on Gruinard Island, it was not declared to be free of anthrax and de-requisitioned until 1990 (Hammond and Carter, 2002). Beginning in 1948, the British also conducted sea

trials in various locations to study pathogen dispersal and viability. Operation Ozone, conducted in the West Indies in 1954, was the first trial to study a virus: Venezuelan equine encephalitis (VEEV), which was obtained from the U.S. Camp Detrick facility. The virus was again used in the 1945–55 Operation Negation, together with Vaccinia virus which was used as a surrogate for smallpox. As part of defensive programs, inhalation experiments with animals at Porton Down, evaluated VEEV and Vaccinia, and also other viruses such as Ebola, influenza, Marburg, Newcastle disease, rabbit pox, Semliki forest, Variola, and Western equine encephalitis.

A much more extensive state funded bioweapons program was developed in the Soviet Union beginning in the 1920s. This program was initially focused on bacteria such as the ones which cause plague, anthrax, tularemia, and glanders (Leitenberg, 2001; Roffey, 2002). In 1999, it was reported that the Soviet program included work on Machupo, Bolivian hemorrhagic fever, Lassa fever, Marburg, Monkey pox, Rift Valley fever, Russian-Spring Summer encephalitis, Venezuelan equine encephalitis, and Variola major viruses (Alibek and Handelman, 1999). Later sources reported that the Soviet Union was studying at least seven viruses including Junin, Ebola, rinderpest, Japanese encephalitis virus, tick-borne encephalitis virus, and African swine fever virus (Meyer and Morse, 2016). Of these seven viruses, the Soviet Union reportedly weaponized at least three of them: Marburg, Venezuelan equine encephalomyelitis, and Variola major.

In 1972, the Biological Weapons Convention was signed and went into effect in 1975. This treaty was signed by many countries including the U.S., the U.K., and the Soviet Union, and restricted countries from developing, producing, stockpiling, or acquiring biological agents, weapons, and equipment outside of peaceful purposes. The U.S. offensive bioweapons program had already been terminated by the time of the signing of this Convention in 1969 (Riedel, 2004). As the U.S. and U.K. began to destroy their stockpiles and ending their programs, the Soviet Union secretly expanded their program by building new facilities and employing large numbers of workers and scientists (Tucker, 1999) including 9000 key scientists and engineers (Alibek and Handelman, 1999; Federation of American Scientists (FAS), 2019). About 10,000 people worked on agricultural pathogens in perhaps six so-called “agricultural research centers”. These facilities had a large production capacity with the estimated production for anthrax being approximately 4500 metric tons per year (Foster, 2003).

In the last years of the 20th century, there were bioweapons programs started by other countries. For example, Iraq began a biological weapons program focused on bacteria, specifically anthrax, and toxins (Atlas, 1999; Eitzen Jr and Takafuji, 1997). There has been a continued proliferation of these agents for offensive purposes by many other countries. The U.S. Congress Office of Technology Assessment concluded in a 1995 report that biological weapons were probably possessed by Bulgaria, China, Cuba, Egypt, India, Iraq, Iran, Israel, Laos, Libya, North Korea, Russia, South Africa, South Korea, Syria, Taiwan, and Vietnam.

Viruses as Bioweapons

In general, there are several characteristics of viruses that enable them to become effective bioweapons. These characteristics include: high morbidity and in some cases mortality, relatively low infective dose, potential for person-to-person transmission, and with appropriate processing, infectivity by aerosol. In some respects, high morbidity is preferable to high mortality, because caring for the sick places a greater demand on resources, infrastructure and support personnel as compared to the demand on resources associated with fatal cases. In contrast with development of what we now call weapons of mass destruction, including conventional explosives, nuclear, and chemical materials, biological agents have several advantageous characteristics. With modern methodologies, they are relatively easy to obtain, inexpensive to produce and relatively easy to deliver. Until recently, they were difficult to detect and identify, where detection relied on the presence and identification of unusual cases by clinicians. Under suitable climatic conditions, viruses can disseminate over large areas and cause large numbers of casualties. The delay from delivery to the onset of infections, would allow the perpetrator to leave the target site to avoid detection, if necessary. It is also possible to vaccinate the aggressor’s troops so that they can occupy the target site even as the agent begins to be transmitted from initial victims to new ones. Genetic engineering could provide a means by which some of the biological agents can be enhanced with respect to infectivity, lethality, and resistance to countermeasures.

With the characteristics of an effective bioweapon in mind, it is easy to see why Variola major virus, the virus that causes smallpox, has remained a significant consideration for use as a bioweapon. It is highly contagious, requires a low inhalation dose, can be spread from person-to-person, has very high morbidity, and a high mortality rate of 30%–40%. The virus can be stored by lyophilization for long periods of time. As demonstrated by the successful eradication campaign, vaccines are highly effective; however, with eradication accomplished, vaccine availability is now limited. Even with stockpiles, genetic engineering could potentially modify a virus intended for bioterrorism to be resistant to the current vaccine. With an incubation period of up to 17 days and lack of clinical expertise, simply because physicians today have never seen a case, a release of the virus could spread quite widely before detection and activation of countermeasures.

The number of viruses that have been considered as bioweapons to be used on a human population is extensive and includes viruses from a number of different virus families, including: *Bunyaviridae* (Rift Valley fever virus and Hantavirus); *Togaviridae* (Venezuelan, Eastern, and Western equine encephalomyelitis viruses and chikungunya virus); *Flaviviridae* (Japanese encephalitis virus, yellow fever virus, dengue virus and tick-borne encephalitis virus); *Arenaviridae* (Lassa fever, Machupo, and Junin viruses); *Filoviridae* (Ebola and Marburg viruses); *Orthomyxoviridae* (Influenza A); and *Poxviridae* (Variola major) (Meyer and Morse, 2016). Although the viruses listed above have been studied for their potential use as bioweapons, few of them have been used as such. Many of the viruses listed above require significant expertise and some require the highest level of biocontainment (biosafety level four (BSL-4)) in order

to work with them safely. This high degree of expertise and biological containment is prohibitive for non-governmental groups to attempt weaponization.

According to calculations by the World Health Organization (WHO) completed in the 1970s, these viruses could cause devastating results if purposely or accidentally released. The WHO examined the possible effects of Rift Valley fever virus (RVFV) and tick-borne encephalitis virus (TBEV) along a 2 km path, upwind of a city with a population of 500,000 and concluded the release of either agent could result in up to 35,000 infections with RVFV causing 400 deaths and TBEV causing 9,500 deaths. These estimates were based on a release of 50 kg of the virus, which at first may seem to be a substantial amount of virus; however, the capacity to produce Marburg virus was estimated at 250 tons and smallpox virus was 100 tons, so perhaps the 50 kg model for RVFV cited above is not as unrealistic as it initially seems. For comparison, a bacterial release of 50 kg of anthrax would disseminate far more than 20 km from the release line and infect 125,000 people with 95,000 deaths. With extensive use of bioreactors, the Soviet Union bioweapons program was able to produce anthrax in very large quantities. Compared with viruses, bacterial spores are environmentally very stable and can remain in the soil for more than 40 years.

Interestingly, a virus that caused an estimated 20–50 million deaths during an epidemic which started in 1918 was not weaponized. Although influenza virus was studied, it was not weaponized as part of the Soviet biological weapons program. This deadly virus causes countless naturally occurring cases where it infects and kills between 290,000 and 650,000 ([World Health Organization, 2020](#)) people and unknown numbers of animals each year. Although influenza virus may never have been developed as a bio-weapon, it did raise alarm when two teams of investigators conducted experiments which lead to much public debate and caused U.S. officials to impose a “pause” on some influenza research in 2014. The research of concern involved conducting studies for what is termed as “gain-of-function”. In these studies, virus characteristics, for example, change in the susceptible host range, could produce a more dangerous virus through the use of molecular biological methods ([Lipsitch and Ingelsby, 2019](#)).

Entomological Warfare

Since many important pathogens of humans are transmitted by arthropods, it is not surprising that biological weapons programs consider the use of arthropods as a means to deliver and spread certain pathogens. The Japanese successfully used infected fleas to deliver *Yersinia pestis* to cause plague in China ([Wikipedia, 2019](#)). Unsubstantiated reports have suggested that the U.S. considered the use of mosquitoes to deliver agents such as yellow fever virus. Interestingly, in 1939, a Japanese scientist tried to obtain YFV from the U.S. Rockefeller Institute Virology Laboratory presumably for inclusion in the Japanese Biological warfare program ([Regis, 1999](#)). The request was denied since resolutions by the League of Nations and a Congress of Tropical Medicine banned importation of the virus into Asia due to concerns that it might become established in the region. Accusations have also been made that in 1952 during the Korean War, the U.S. dispersed insects and spiders contaminated with anthrax and even rodents possibly infected with plague ([Regis, 1999](#)).

Agro-Warfare

From the above discussion, it is clear that for bio-warfare with humans as the target species, relatively few viruses have been considered since bacteria are more readily weaponized, stored, and delivered. Actual agricultural bio-warfare events have also predominantly focused on diseases caused by bacterial agents such as anthrax (*Bacillus anthracis*) and glanders (*Burkholderia mallei*). As mentioned above, glanders was first reported to be used by Germany to attack horses and mules that were a critical component of the transport and supply infrastructure on the battlefields of World War One ([Wheelis, 1998](#)). Nonetheless, when compared with agents targeted towards humans, those targeted towards livestock include a greater variety of viruses.

As demonstrated by numerous natural outbreaks, the economic impact of viral infections in livestock can be enormous. This is not only due to animal deaths, but also due to trade restrictions that can be imposed immediately upon detection of the pathogen and can last for years until a country is declared free of the pathogen. Viruses that are not zoonotic may be particularly attractive as bioweapons because they can be handled without the risk of the weapon developer becoming infected. Foot-and-mouth disease virus (FMDV) is perhaps the best example of a virus that has been weaponized for agro-warfare. This weaponization took place at the Russian Pokrov Biologics Plant, which is 50 miles east of Moscow ([Warrick, 2002](#)). The campus of this plant included five underground bunkers and had two FMD production lines that could produce 12 metric tons of the virus per production run.

A naturally occurring outbreak of FMD in the United Kingdom in 2001 is a good example of the devastation to agriculture that can be caused by a release of a virus in a new area. The U.K. FMD outbreak lasted for seven months and because at the time, vaccination was not allowed due to a European Union directive, the outbreak resulted in the slaughter of 6.6 million animals. The economic loss was estimated to have exceeded \$20 billion. A subsequent outbreak in 2007 had much less severe consequences lasting for four months with 2160 animals killed and the economic impact was \$200 million. Ironically, the 2007 outbreak was caused by virus released from the Pirbright laboratory site where a Government research facility and commercial vaccine research/production facility were located.

In addition to foot-and-mouth disease virus, there are other viruses that would be a threat to livestock including viruses in the families of: *Asfarviridae* (African swine fever virus) and *Paramyxoviridae* (rinderpest, New Castle disease virus).

The recent spread of African swine fever in Europe and China has caused much concern in agricultural communities. The virus has been introduced and eventually eradicated from several countries outside of Africa during the 20th Century, but in 2007 was identified in the country of Georgia. When it spread to Russia in 2012, Russia accused Georgia of swine fever sabotage, although there was no evidence to suggest an intentional introduction. Human activities, namely transportation of asymptomatic infected pigs and movement of infected feral pigs/wild boar, are potentially the cause of the spread of this virus over these large distances. The virus is now established in Armenia, Azerbaijan, Belarus, Czech Republic, Estonia, Georgia, Hungary, Latvia, Lithuania, Moldova, Poland, Romania, Russia, and Ukraine, and has recently been reported in Belgium. In August of 2018, the virus was reported in China and by March of 2019, infections had occurred in 28 provinces and nearly 1 M pigs were culled in efforts to control the virus. In addition to the devastation to the Chinese pork market due to culling of infected or exposed swine, an additional consequence of this outbreak is the unwillingness of customers to consume pork during the outbreak. This unconsumed pork could be the motivation behind one of the largest seizures of an agricultural product by U.S. customs. In 2019, U.S. Customs found 1 million pounds of smuggled pork from China (Hirtzer, 2019). An unintended consequence of unregulated trade in pork could be the introduction of a virus to a new area. Unfortunately, it is known that following the outbreak in China, there were other outbreaks occurring in Mongolia and South Korea. The question of intentional or naturally occurring outbreaks is not new; for example, an unexplained 1971 outbreak of African swine fever occurred in Cuba. This outbreak led to the slaughter of 500,000 pigs on the island and was blamed on the U.S. Central Intelligence Agency although the evidence to support the accusation has never been revealed (Chronicle, 1977).

Biological Agents for Good or Evil

The use of biological agents to control pests in the environment may be considered bioterrorism or a bio-crime by some people or as biological control by others. For example, in 1997, rabbit hemorrhagic disease virus was used by a small group of farmers to kill rabbits in New Zealand. For over 20 years, the lead perpetrator had been advocating with the New Zealand Agriculture Department for measures to control rabbits, which were causing millions of dollars of losses to farmers every year. Unsatisfied with the results of these discussions, the perpetrators illegally introduced the virus into several areas. This non-State sanctioned use of a virus could certainly be regarded as an act of agro-terrorism, albeit intended to control a wild introduced pest species (the European rabbit) with the purpose of reducing environmental damage, and the economic consequences to crop production and grazing for livestock. Since the 1997 release, the government has determined they should control the release of a variant of the rabbit hemorrhagic disease virus (RHDV1-K5) as a government sponsored project to control the rabbit population.

The use of viral hemorrhagic disease for rabbit control is very similar to the state sponsored use of myxoma virus, which causes myxomatosis, released in Australia on a large scale in 1950. This release is estimated to have reduced the rabbit population from 600 million to 100 million within just two years. The European rabbit, *Oryctolagus cuniculus*, that had been introduced into Australia by settlers, was regarded as a significant agricultural pest as early as 1887, when a £25,000 reward was offered to anyone that developed an effective control measure. The disease had first been reported in 1897 in a stock of European rabbits maintained at a hospital in Montevideo, Uruguay (Sanarelli, 1908). The potential use of the virus to control rabbits in Australia was proposed in 1927 (Aragao, 1927).

The success of the Australian release of myxoma virus probably inspired the release of the virus in France in 1952. Not unexpectedly, the virus quickly spread beyond the initial release site with devastating effects on the European rabbit population, not just in France, but subsequently in the Netherlands, Italy, Spain, and other European countries. Whilst people involved in agriculture were supportive of the effects of the release, hunters were outraged. In 1955, the person responsible for the initial release was prosecuted and fined 5,000 francs for his actions. Ironically in 1956, he was awarded a medal of achievement by the French Department of Rivers and Forests. This prosecution and reward for these actions indicate how controversial such releases of bioweapons or biological control agents are.

The first cases of myxomatosis in the U.K. appeared in the autumn of 1953 in the southern counties of Sussex and Kent. The virus spread rapidly and resulted in reductions in the rabbit populations similar to those experienced in Europe. Interestingly, the virus did not become established in New Zealand even though it was apparently introduced in the 1950s. An interesting aspect of myxomatosis in the U.K., is that it was tested on Gruinard Island where anthrax was also tested. Under the direction of Sir Charles Martin, rabbits on Skokholm Island were inoculated with the virulent strain B of the virus in 1936. Remarkably, there was no effect on the estimated 10,000 rabbit population, with similar failures experienced after follow-up releases in 1937 and 1938. The cause of this mysterious failure was not resolved until 1955. By this time, studies of the transmission cycle of Myxoma virus, had determined a key role of the rabbit flea *Spilopsyllus cuniculi* as a vector of the virus (Lockley, 1954). The rabbit population of Skokholm Island, was perhaps unique in that they were free of this ectoparasite (Lockley, 1955), demonstrating the importance of understanding everything possible about a biological agent in order to optimize its use against a target species. Although effective use of the virus to control rabbits was achieved in most places without this understanding, the knowledge would have been invaluable during the British plans to test the virus on an isolated population.

As a further point of interest, in Australia it was determined that a species of flea, *Echidrophaga myrmecobii*, which parasitizes marsupials, could transmit the virus; but importantly, a type of arthropod, namely mosquitoes, that are not a likely vector of the virus in its Native South America, could be an efficient mechanical vector. The role of mosquitoes as mechanical vectors in other countries does not seem to be well understood, but their role in Australia illustrates that one cannot assume that what is known

about a pathogen in the endemic region, will apply when introduced into a new area. Establishment of West Nile virus in the Americas since 1999, has also demonstrated this principle.

Clearly, these examples illustrate, how the perpetrator and intent can determine the subtleties of how an event is classified. What in some circumstances will define a release of an agent as legitimate and beneficial biological control, can in other circumstances be classified as terrorism, with consequences of fines and perhaps imprisonment.

Effects of Viral Introductions

Often the resultant effects from the natural spread of a virus can be used to model the potential effects of a bioterrorist attack using viruses. Three good examples are the effects from the introduction of West Nile to the Americas, porcine epidemic diarrhea, and highly pathogenic avian influenza viruses. West Nile virus (WNV) was introduced into the U.S. in 1999, although the route(s) of the initial introduction remains unknown, it is believed to not be a bioterrorism related event. Despite West Nile virus being a well-studied-virus with a record of frequent introductions from Africa into other countries, when the virus first came to the U.S. experts assumed that it was St. Louis encephalitis virus. This assumption was probably made because St. Louis encephalitis virus was a virus that occurred in the U.S. However, an astute veterinarian noticed the atypical fatal infections in birds and was insistent that St. Louis encephalitis virus was not the causative agent of the outbreak. In addition to revealing inadequacies of surveillance and initial reporting of a new viral introduction, this also showed inadequacies in the ability to predict viral spread following an introduction. Although initial predictions indicated the virus would be contained, it is now well documented that the virus continued to spread throughout the U.S. each year until it is now permanently established in the U.S., and in some other countries of the Americas, and continues to cause many severe illnesses and some fatalities in people, and in other species, notably horses. Although species of mosquito, for example in the *Culex pipiens* complex, that are principal vectors of WNV in endemic regions, are also present in the U.S., WNV has been detected in far more species in the U.S. than ever seen before. This demonstrates once again that one should be cautious in assuming that the characteristics of a pathogen in a new area will be the same as previously observed. Had the introduction of WNV been a deliberate bioterrorism attack, it would be regarded as a success. From probably a single introduction, its characteristics of being mosquito-borne, infectious to multiple species of vertebrates, and zoonotic with the ability to cause fatal infections, enable the virus not just to cause an outbreak but to become established permanently at a continental level and therefore continue to cause suffering and a financial burden. As a result of the introduction, there were costs to train people, conduct research, develop nationwide surveillance, and develop a blood screening capability and a veterinary vaccine. It is estimated that hospitalization and lost-productivity associated with each encephalitis WNV case can cost up to \$324,000, with a total cost for all hospitalized cases between 1999 and 2012 being \$778 M (Staples, 2014). One virus regarded as a potential bioterrorism agent, RVFV shares several characteristics with WNV. It is mosquito-borne, zoonotic, potentially fatal for humans, and dissemination would likely follow a very similar path as WNV, with the additional potential for large scale livestock losses.

In 2013, porcine epidemic diarrhea virus was introduced into the U.S., probably from imported animal feed products. The virus is now in 34 states and costs an estimated \$900 million to the swine industry with approximately 8 million pigs being culled. In 2015, an outbreak of highly pathogenic avian influenza resulted in the deaths of 50 million poultry with a total economic impact of \$3.3B to the U.S. (Blue Ribbon Study Panel on Biodefense, 2017).

Another viral disease of poultry, Newcastle disease (named because of a 1927 outbreak in Newcastle-upon-Tyne, U.K.), can infect over 250 different bird species with mortality ranging between 70 to 100%. A U.S. outbreak in 2002-03 began in backyard birds with cases in Arizona, Nevada, and Texas. Outbreaks in California commercial poultry industry lasted for 11 months, resulting in trade restrictions, with an estimated cost of over \$160 million, with 3.16 million birds being killed. Again, in California, outbreaks in commercial poultry operations were reported in December of 2018, and are ongoing at the time of writing.

Above, we have discussed threats to agriculture with a focus on pathogens of animals, including zoonotic viruses that also infect humans. In the context of agro-terrorism, plant targets could have an even more devastating impact on food security and the economy, albeit perhaps without the fear factor associated with potentially fatal human diseases. The USDA plant protection and quarantine's select agent and toxins list only includes bacterial and fungal agents. One of these, *Rathayibacter toxicus*, although a pathogen of ryegrass has the unusual property of producing a corynetoxin that at relatively low doses is lethal to vertebrates. Indeed, thousands of sheep and cattle die each year as a result of eating infected grasses. Parker listed banana bunchy top virus as a potential agricultural biowarfare or bioterrorism agent but Wheat dwarf geminivirus and Barley yellow dwarf virus were listed as having potential bioweapon applications (Parker, 2002). Other viruses of plants that could be threats include: Rice tungro virus, Wheat streak mosaic virus, Maize streak virus, Cassava mosaic virus, Tomato spotted wilt virus, Pepino mosaic virus, and the PVYNTN strain of Potato virus Y (Stack, 2019). The Russian biological weapons program supposedly weaponized Wheat stem rust and Rice blast and a fungus-based anti-crop bomb. Another example, is wheat blast which was supposedly developed and put in production by the U.S. in 1951 during the Korean War.

Conclusion

It is clear that there are many biological agents that have the potential to be used in a bioterrorism attack. Many of these agents are viruses that infect, humans, and/or economically important animals and plants that are part of important agricultural practices to provide food. Unlike chemical or radiological-based weapons, many biological agents can be relatively easily obtained at low

costs, since they occur in nature, often over wide geographical areas and relatively frequently. Simple internet searches provide real time information on where and when these agents are present. Having evolved to survive via various transmission cycles, viruses often have the capacity to spread quickly as demonstrated by large scale epidemics, and also, after introduction, to become permanently established in new geographic locations. Some of the viruses and bacteria can be infectious via aerosols, and some can survive for long periods of time, either in the environment or in asymptomatic but potentially infectious hosts. Although development as biological agents for bio-warfare purposes, necessitates large-scale production and processing capabilities, with regards to use for bioterrorism, preparation of an agent may be accomplished using relatively simple equipment and procedures. The availability of vaccines and treatments for several of the viruses of concern, or limitations of host range, for example with viruses that infect swine or plants but not people, working with these viruses could be accomplished by people without fear of infection. New technologies may provide the ability to genetically manipulate agents with increasing ease at low cost without specialist knowledge or equipment. For these reasons, biological agents may be considered as more practical and cost effective than chemical or radiological weapons. As such, we must be increasingly concerned and vigilant with regards to the use of biological agents to attack us or our agricultural systems upon which we depend. As their use becomes easier, we must assume that the probability of their use becomes greater.

References

- Alibek, K., Handelman, S., 1999. *Biohazard: The Chilling True Story of the Largest Covert Biological Weapons Program in the World, Told From the Inside by the Man Who Ran it*, vol. xi. New York: Random House, p. 319.
- Aragao, H.D.B., 1927. Myxoma of rabbits. *Memórias do Instituto Oswaldo Cruz* 20 (2), 225–247.
- Atlas, R.M., 1999. Combating the threat of biowarfare and bioterrorism. *Bioscience* 49 (6), 465–477. doi:10.2307/1313554.
- Blue Ribbon Study Panel on Biodefense, 2017. Special Focus: Defense of Animal Agriculture. Blue Ribbon Study Panel on Biodefense: Washington, DC.
- Bradley, M., Sustaita, G., 2013. Applying scientific evidence to prosecuting perpetrators of well contamination as bio-warfare strategy. *Journal of Biosecurity, Biosafety and Biodefense Law* 4 (1), 5–18. doi:10.1515/jbbbl-2013-0002.
- Campbell, G.L., 2014. *The Oxford Handbook of Animals in Classical Thought and Life*, xix. Oxford, United Kingdom; New York, NY, United State of America: Oxford University Press, p. 633.
- Carus, W.S., 2002. *Bioterrorism and Biocrimes: The Illicit Use of Biological Agents Since 1900*. Amsterdam: Fredonia Books.
- Chronicle, S.F., 1977. CIA Link to Cuban Pig Virus Reported 1977 [updated January 10, 1977; cited 2019 March 19, 2019]. Available from: <http://www.maebrussell.com/Health/CIA%20Pig%20Virus.html>.
- Eitzen Jr., E.M., Takafuji, E.T., 1997. Medical aspects of chemical and biological warfare. In: Sidell, F.R., Takafuji, E.T., Franz, D.R. (Eds.), *Medical Aspects of Chemical and Biological Warfare*. Washington, D.C.: Borden Institute, Office of the Surgeon General, United States Army Medical Department Center and School, United States Army Medical Research and Materiel Command, Uniformed Services University of the Health Sciences, pp. 415–424.
- Federation of American Scientists (FAS), 2019. Biopreparat 2019 [updated October 20, 1998; cited 2019 March 11]. Available from: <https://fas.org/nuke/guide/russia/agency/bw.htm>.
- Fitzgerald, G.J., 2008. Chemical warfare and medical response during world war I. *American Journal of Public Health* 98 (4), 611–625. doi:10.2105/AJPH.2007.11930.
- Florus, L.A., 1852. *Epitome of Roman History*. Harvard University Press. pp. 287–424.
- Foster, D., 2003. The Message in the Anthrax UCLA Department of Edidemiology. [updated September 15, 2003; cited 2019 March 12]. Available from: <https://www.ph.ucla.edu/epi/bioter/messageanthrax.html>.
- Hammond, P.M., Carter, G.B., 2002. *From Biological Warfare to Healthcare, Porton Down, 1940–2000*. Palgrave.
- Hirtzer, M., 2019. U.S. Seizes 1 Million Pounds of Smuggled Chinese Pork: Bloomberg. [updated March 15, 2019; cited 2019 March 19]. Available from: <https://www.bloomberg.com/news/articles/2019-03-15/u-s-seizes-1-million-pounds-of-smuggled-chinese-pork>.
- Hopkins, D.R., 1983. *Princes and Peasants: Smallpox in History*. vol. xx. Chicago: University of Chicago Press, p. 380.
- Keremidis, H., Appel, B., Menrath, A., et al., 2013. Historical perspective on agroterrorism: Lessons learned from 1945 to 2012. In: *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science*, vol. 11 (Suppl 1), S17–S24. doi:10.1089/bsp.2012.0080. PMID: 23971803.
- Kettle, A.N.B., Nicoletti, P.L., 2014. Chapter 36 – Glanders. In: Sellon, D.C., Long, M.T. (Eds.), *Equine Infectious Diseases*. Saunders, pp. 333–336.
- Leitenberg, M., 2001. Biological weapons in the twentieth century: A review and analysis. *Critical Reviews in Microbiology* 27 (4), 267–320. doi:10.1080/20014091096774.
- Lesho, M.E., Dorsey, M.D., Bunner, D., 1998. Feces, dead horses, and fleas. Evolution of the hostile use of biological agents. *Western Journal of Medicine* 168 (6), 512–516.
- Lipsitch, M., Ingelsby, T., 2019. The U.S. is Funding Dangerous Experiments it doesn't Want You to Know About The Washington Post. [updated February 27, 2019; cited 2019 March 5]. Available from: https://www.washingtonpost.com/opinions/the-us-is-funding-dangerous-experiments-it-doesnt-want-you-to-know-about/2019/02/27/5f60e934-38ae-11e9-a2cd-307b06d0257b_story.html?noredirect=on&utm_term=.95cb5f9a298d#click=https://t.co/tSdtPgWk6o.
- Lockley, R.M., 1954. The European rabbit flea, *Spilopsyllus cuniculi*, as a vector of myxomatosis in Britain. *The Veterinary Record* 66.
- Lockley, R.M., 1955. Failure of myxomatosis on Skokholm Island. *Nature* 175 (4464), 906–907.
- MacLachlan, N.J., Dubovi, E.J., Barthold, S.W., Swayne, D.E., Winton, J.R., 2017. *Fenner's Veterinary Virology*, vol. xix. Amsterdam: Elsevier/AP, Academic Press is an imprint of Elsevier, p. 581.
- Mayor, A., 2008. *Greek Fire, Poison Arrows, & Scorpion Bombs: Biological & Chemical Warfare in the Ancient World*, first ed. Harry N. Abrams.
- Meyer, R.F., Morse, S.A., 2016. *Viruses and bioterrorism*. In: *Reference Module in Life Sciences*. Elsevier, pp. 406–411.
- Millett, P., Snyder-Beattie, A., 2017. Human agency and global catastrophic biorisks. *Health Security* 15 (4), 335–336. doi:10.1089/hs.2017.0044.
- Nepos, C., 2015. Chapter 10-Life of Hannibal. In: *Excellentium Imperatorum Vitae* Open Book Publisher.
- Parker, H.S., 2002. *Agricultural Bioterrorism: A Federal Strategy to Meet the Threat*. vol. xii. Washington, D.C.: Institute for National Strategic Studies, National Defense University, (103).
- Patterson, K.B., Runge, T., 2002. Smallpox and the Native American. *The American Journal of the Medical Sciences* 323 (4), 216–222.
- Regis, E., 1999. *The Biology of Doom: The History of America's Secret Germ Warfare Project*, vol. 259. New York: Henry Holt.
- Riedel, S., 2004. Biological warfare and bioterrorism: A historical review. *Baylor University Medical Center Proceedings* 17 (4), 400–406.
- Roffey, R., Tegnelli, A., Elgh, F., 2002. Biological warfare in a historical perspective. *Clinical Microbiology and Infection* 8 (8), 450–454.
- Romano Jr., J.A., Lukey, B.J., Salem, H., 2007. *Chemical Warfare Agents: Chemistry, Pharmacology, Toxicology, and Therapeutics*. second ed. CRC Press.
- Sanarelli, G., 1908. Das myxomatogene virus. *Centralblatt für Bakteriologie* 23 (20).
- Stack J., Higgs, S., 2019. Potential plant viruses for biowarfare.
- Staples, J.E., Shankar, M.B., Sejvar, J.J., Meltzer, M.I., Fischer, M., 2014. Initial and long-term costs of patients hospitalized with West Nile virus disease. *American Society of Tropical Medicine and Hygiene* 90 (3), 402–409. doi:10.4269/ajtmh.13-0206.

- Tucker, J.B., Mahan, E.R., 1972. President Nixon's Decision to Renounce the U.S. Offensive Biological Weapons Program. Available at: http://ndupress.ndu.edu/Portals/68/Documents/casestudies/CSWMD_CaseStudy-1.pdf.
- United Nations Office for Disarmament Affairs, 1925. Geneva Protocol 1919 [cited 2019 March 4]. Available from: <https://www.un.org/disarmament/wmd/bio/1925-geneva-protocol/>.
- Warrick J. Russia's Poorly Guarded Past: The Washington Post; 2002 [updated June 17; cited 2019 March 22]. Available from: https://www.washingtonpost.com/archive/politics/2002/06/17/russias-poorly-guarded-past/603c3de9-22f5-493f-be5b-fc31eb938a9d/?Utm_term=.8032fa23cceb.
- Wheelis, M., 1998. First shots fired in biological warfare. *Nature* 395 (213), 6699. doi:10.1038/26089.
- Wheelis, M., 2002. Biological warfare at the 1346 Siege of Caffa. *Emerging Infectious Diseases* 8 (9), 971–975. doi:10.3201/eid0809.010536.
- Wikipedia, 2019. Operation Big Buzz [updated February 2, 2018; cited 2019 March 12]. Available from: https://en.wikipedia.org/wiki/Operation_Big_Buzz.
- World Health Organization. 2018. Influenza (Seasonal). Available from: [https://www.who.int/news-room/fact-sheets/detail/influenza-\(seasonal\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)).

Further Reading

- Burnette, R., 2013. Biosecurity: Understanding, Assessing, and Preventing the Threat. Hoboken, NJ: John Wiley & Sons.
- Franz, D.R., Takafuji, E.T., Parrott, C.D., 1997. Chapter 19 – The U.S. biological warfare and biological defense programs. In: *Medical Aspects of Chemical and Biological Warfare*. United States. Department of the Army. Office of the Surgeon General. (Available at: <http://www.au.af.mil/au/awc/awcgate/medaspec/Ch-19electrv699.pdf>).
- Frischknecht, F., 2003. The history of biological warfare. Human experimentation, modern nightmares and lone madmen in the twentieth century. *EMBO Reports*. 47–52.
- Keith, J., 1999. *Biowarfare in America*. Atlanta, GA: IllumiNet Press.
- Leitenberg, M., Zilinskas, R.A., Kuhn, J.H., 2012. *The Soviet Biological Weapons Program: A History*. Cambridge, MA: Harvard University Press.
- Lockley, R.M., 1964. *The Private Life of the Rabbit: An Account of the life history and Social Behaviour of the wild rabbit*. MacMillan Publishing Company.
- Lockwood, J.A., 2012. Insects as weapons of war, terror, and torture. *Annual Review of Entomology* 57, 205–227. doi:10.1146/annurev-ento-120710-100618.
- Mangold, T., Goldberg, J., 2000. *Plague wars: A True Story of Biological Warfare*. New York: St. Martin's Press.
- National Research Council (U.S.). Committee on Biological Threats to Agricultural Plants and Animals., National Research Council (U.S.). Board on Agriculture and Natural Resources. and National Research Council (U.S.). Board on Life Sciences, 2003. *Countering Agricultural Bioterrorism*. Washington, D.C.: National Academies Press.
- Smart, J.K., 1997. Chapter 2 – History of chemical and biological warfare: An American perspective. In: Franz, D.R., *et al.* (Eds.), *Medical Aspects of Chemical and Biological Warfare*. United States. Department of the Army. Office of the Surgeon General, pp. 9–86. (Available at: <http://www.au.af.mil/au/awc/awcgate/medaspec/Ch-2electrv699.pdf>).
- Whitby, S.M., 2002. *Biological Warfare Against Crops*. Houndmills, Basingstoke, Hampshire; New York: Palgrave.

The Use of Viral Promoters in Expression Vectors

Ian M Jones, University of Reading, Reading, United Kingdom and London School of Hygiene and Tropical Medicine, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

An Introduction to Promoters

The origins of promoters and their manipulation are an integral part of the history of molecular biology. As DNA sequencing became commonplace following the development of the canonical Sanger and Maxam–Gilbert methodologies the possibility to align sequences and look for common elements arose. A number of sequences, often termed “boxes” in the original alignment studies, were identified upstream of the methionine encoding ATG marking the beginning of a translatable sequence. Such elements were found upstream of both bacterial and eukaryotic open reading frames although the sequences concerned were different. Combined with earlier molecular genetics studies which had defined *operators* (sequences that contained both promoter and repressor elements) it became apparent that such sequence elements were responsible for the control of gene expression. They include the promoter itself, often a control element and sequences around the start codon which serve to stabilize the transcript when it binds to the ribosome (Fig. 1).

Early molecular genetics experiments had shown, using gene fusions for example, that combining the promoter elements of one gene with the coding sequence of another led to a pattern of expression characteristic of the promoter, not the product and the first studies of heterologous gene expression, by the Boyer and Cohen labs, similarly demonstrated that combining bacterial promoters with eukaryotic coding sequences allowed eukaryotic protein expression in *E.coli*. These studies formally demonstrated the plug-and-play possibilities of expression vector design and many promoters were subsequently studied for use therein. Some of these, the *lac* promoter for example, are still used in expression vectors today.

Viral Promoters

The widespread use of viral promoters to control the expression of heterologous genes derives from the fundamental requirements of viruses to (1) initiate replication in multiple cell types and, (2) produce large amounts of the structural proteins necessary for virus assembly. This gives two classes of promoter, relatively *host independent* and *strong* respectively. Viruses typically exhibit tropisms, that is a preference for one cell or tissue type over another, a property that can be exploited when targeting particular cells by virus vectors, in gene delivery applications for example. Tropism can act at any stage of the replication cycle but the predominant restriction operates at the level of virus entry, the presence or not of the viral receptor on the host cell surface. Once

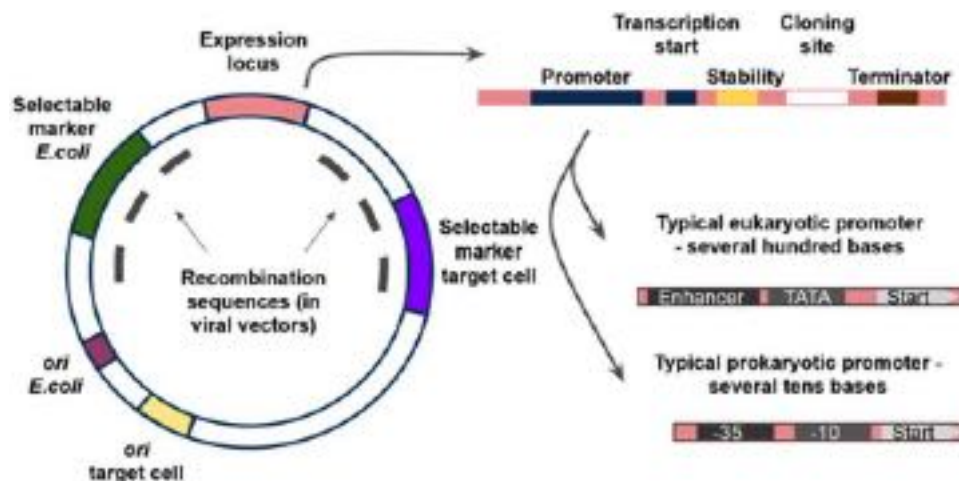


Fig. 1 A typical expression vector and promoter features. Left. All vectors will be usually equipped with a marker and origin of replication for growth in *E.coli*. They may, additionally, have sequences for particular uses, for example recombination sequences for integration into a virus vector or another form of carrier. One or more loci will be engineered to have cloning sites positioned at the optimum distance downstream of a defined promoter for transcription to initiate efficiently. A selectable marker and origin of replication for the target host may also be present. Right. A generalized promoter will bind RNA polymerase and initiate transcription at a defined nucleotide downstream. The 5' untranslated region is of variable length but sequences immediately 5' to the translation start may include Shine-Dalgarno (bacteria) or Kozak (eukaryotic) sequences that stabilize engagement with the ribosome. A transcription terminator will disengage RNA polymerase. Eukaryotic promoters may have an enhancer, necessary for high level or tissue dependent expression and are in consequence usually much larger than prokaryotic promoters. Canonical “boxes” of each are indicated.

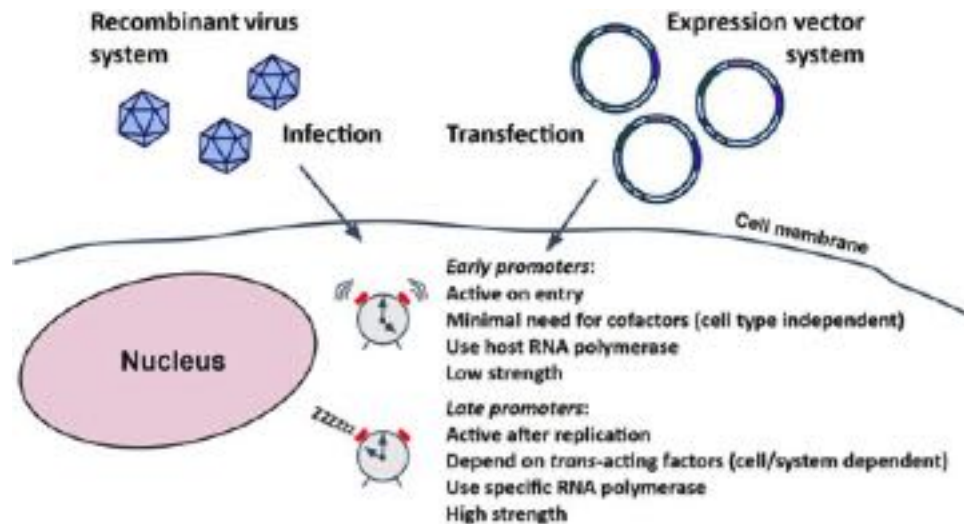


Fig. 2 Viral promoter classes. Expression from viral promoters can result from infection by recombinant viruses or from transfection of DNA or RNA. The former is usually very efficient but may eventually kill the cells while the latter is usually less efficient but generally non-toxic. Once delivered promoters function as either early or late with the features indicated.

inside a cell it is in the best interests of the virus to initiate the replication cycle as soon as possible as failure to do so may result in the virus being degraded and the opportunity for infection lost. Accordingly, with some exceptions, early promoters, defined as those that are used immediately after virus entry are recognized by the host RNA polymerase and initiate RNA transcription without the need for particular co-factors. They are, in essence, universal “go” signals although they may be of limited strength as many early viral components are transcription factors or enzymes that are not needed in abundance. By contrast, late viral promoters are very strong transcription initiators dedicated to the production of copious amounts of message encoding the structural proteins needed late in the replication cycle to permit the assembly of progeny virions (Fig. 2). They require specific *trans* activators and often use a virally encoded polymerase making them much more restricted in terms of permissive cell type. This *early-weak cell independent, late-strong cell dependent* distinction occurs across the field of virology and is the fundamental property that has led to virus promoters being taken up as widely as they are, in viral expression systems themselves and as part of stand-alone expression vectors. There are exceptions, some RNA viruses may forgo the use of promoters, as is the case for some positive strand RNA viruses that translate their genome directly on virus entry, and negative sense and double stranded RNA viruses have promoters that are recognized only by the virus encoded polymerase. They have, nonetheless, been used for expression purposes within particular virus models and some have seen broad take-up in particular fields. The alphaviruses for example, which have a single strong internal promoter recognized by their own RNA polymerase, are in use today as part of the growing field of RNA vaccines although these uses are specialist cases and the promoters concerned are not in widespread use elsewhere.

Viral Promoter Examples

The use of virus promoters and viral systems in biotechnology applications is very widespread but some well-known examples serve to illustrate the general benefits of the early/late promoter choice. These include the use of the cytomegalovirus immediate early promoter in general eukaryotic expression vectors designed for transfection into mammalian cells, the use of the phage T7 late promoter in bacteria and the use of both early and late promoters in the baculovirus insect cell system.

Herpes viruses have gained popularity as vaccine vectors in recent years as they induce robust immunity at the sites where they establish latency, a positive aspect of the virus tropism noted earlier. Encouraging data have been obtained in macaque models for diseases such as HIV and TB that had previously responded only poorly to vectored vaccine approaches. In the recombinant herpes viruses used, up to 6 heterologous genes are expressed from a single virus backbone using, among others, the macaque cytomegalovirus immediate early (MCMV IE) promoter. As noted, immediate early promoters are those that are recognized by host RNA polymerase immediately on cell entry (IE genes are operationally defined as those which are active within the first 6 h of infection). The purpose of these vectors is to expose infected cells to the vaccine antigen in the early, inflammatory phase of virus expansion, leading to the induction of a robust immune response that is then boosted by the alternate cycles of latency and activation that are a herpes virus characteristic. But the CMV IE promoter is also used much more widely than within CMV based vaccine vectors. Along with the promoters of Simian Virus 40 (SV40) and Rous Sarcoma Virus, CMV IE1 was one of the earliest defined eukaryotic promoters and familiarity with its sequence, particularly the mapping of its boundaries, led to its widespread adoption in early plasmid expression vectors. In the virus genome CMV IE contains an active enhancer region several hundred bases upstream of the TATA box (cf. Fig. 1) which serves to bind several host transcription factors leading to the assembly of highly

competent transcription complexes and high levels of transcription. In a typical expression vector in use today, pcDNA3 for example, the enhancer-promoter sequences are found in a ~500 base pair fragment that is part of a selectable vector with other common features (cf. Fig. 2). Several of those features, the promoter for the neomycin resistance gene used for selection of eukaryotic cell transformants and the transcription termination signals are also derived from a virus, SV40, for the same underlying reasons already discussed. Further analysis of CMV IE has determined that the enhancer-promoter is modular and that the enhancer continues to function if appended to promoters from other genes, leading the construction of hybrid enhancer-promoters with even higher levels of transcription. In the pCAG vectors the CMV enhancer is abutted to the chicken actin promoter which also includes a non-translated exon and intron in addition to the promoter itself. Introns have been found empirically to improve the function of several eukaryotic expression vectors. They oblige the primary transcript to be processed by the splicing machinery which leads to a more stable message and higher levels of encoded product. These vectors also contain an origin of replication, again from SV40, that functions in eukaryotic cells expressing large T antigen, meaning that they replicate and pass to daughter cells at cell division so transfected cells are not diluted out in a growing culture.

An example of a virus late promoter in another expression host, *E. coli*, is provided by the T7 system developed over many years by Studier and colleagues. Its efficiency is such that it is the predominant bacterial expression technology in use today. Many additions to the system have been described including options to regulate the levels of expression and the use of particular *E. coli* strains to maximize productivity and the solubility of the translated product. But at its heart the system exploits the strong promoter found upstream of the late genes of bacteriophage T7. In keeping with the early/late divide already described, the T7 early genes are transcribed by *E. coli* RNA polymerase following the entry of the phage into the cell. A cascade of temporal gene expression results in replication of the T7 virus genome and a switch to late gene expression. This is achieved by synthesis of a dedicated RNA polymerase which is not only more processive than *E. coli* RNA polymerase but is also specific for T7 late promoters which are unique and do not share the common promoter elements described in Fig. 1. As a result, anything cloned downstream of a T7 promoter in a plasmid vector is transcribed selectively and abundantly when T7 RNA polymerase is present. If that vector is also multicopy a gene dosage effect also comes into play and very high levels of synthesis are possible. Today's systems match engineered *E. coli* strains, which have the gene for T7 RNA polymerase integrated into the chromosome under control of a *lac* promoter, with a variety of vectors with the T7 promoter. Induction of expression is indirect with activation of the *lac* promoter by IPTG, or manipulation of the growth medium, to induce expression of the T7 polymerase which, in turn, transcribes a vector resident target gene. The specificity of the T7 polymerase for the T7 promoter is a restriction when gene expression from other promoters is required and other bacterial virus promoters, recognized by *E. coli* RNA polymerase are also used, from phages lambda and phage T5 for example. But the strong-late features of the T7 system have made it a default for bacterial expression experiments and variations of it have also been developed for eukaryotic cells although they are not in widespread use. The specificity of polymerase and promoter has also allowed the development of transcription vectors and systems designed for the production of RNA *in vitro*. Phage T7 and the closely related phages T3 and SP6 are all used for these systems as they all share related late promoters that are transcribed efficiently by their cognate RNA polymerases.

A system that has made use of both early and late viral promoters is the baculovirus insect cell system. Its earliest descriptions for heterologous gene expression, and its origins as an expression system, described the use of the very late polyhedrin promoter to transcribe an inserted gene. Polyhedra are crystalline arrays of a single viral protein which trap virus particles to afford them a high degree of environmental protection, from desiccation or UV irradiation, for example, both inherent risks for virus transmission on plant surfaces. Two features follow from this description, very late promoter function is uncommonly strong, to generate the amount of polyhedrin protein required to coat multiple virus particles, and its need for survival in the environment means that it plays no role in tissue culture. Accordingly, the gene encoding polyhedrin can be replaced with a gene of interest so that, late in infection, very high levels of RNA encoding the heterologous protein are produced (Fig. 3). The polyhedrin promoter, and several others similarly classified as very late, conform to the dependency previously noted as their transcription is achieved with a virus encoded RNA polymerase, a eukaryotic equivalent of the T7 system, and so function is restricted to virus infected cells. Despite its strength, the polyhedrin promoter is transcriptionally silent in the absence of virus infection so its general use, as described for CMV IE for example, is not possible. In some recombinant virus systems, the exchange of a viral for recombinant gene results in an attenuated phenotype that requires a *helper* virus or cell line to be overcome. But in recombinant baculoviruses this is not the case and recombinant viruses grow to the same titer as unmodified viruses, facilitating large scale expression, including at industrial scale. As replication is not impaired a gene dosage effect also applies, a further parallel with the T7 system. Variations in the system since its inception in the early 1980s have provided toolboxes for high level expression of single and multiple genes as well as the means to generate the recombinant viruses easily. In fact, the use of early promoters was also investigated, based on the idea that, productive as it is, polyhedrin promoter driven expression occurs at the end of the replication cycle and so recombinant protein expression is limited by the fact the cell is dying (baculovirus infections are lytic). An early promoter driven heterologous gene might allow expression throughout the replication cycle and so accumulate to higher levels overall. This turns out not to be true but the study of early promoters has given rise to a transfection based system for insect cells that mirrors the transfection of mammalian cells with the CMV IE based vectors already described. IE promoters in baculoviruses, including in the default virus used, *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV), are low strength, the default early-weak situation already encountered. However, when combined with homologous regions found in the baculovirus genome, *hrr* sequences, transcription from them improves dramatically. In effect *hrr*s act like the CMV enhancer already described although they do not bind transcription factors but rather act as origins of replication or chromatin opening sequences. Subsequent improvements have systematically optimized the sequence of each component of such IE promoter based transient expression vectors so that their

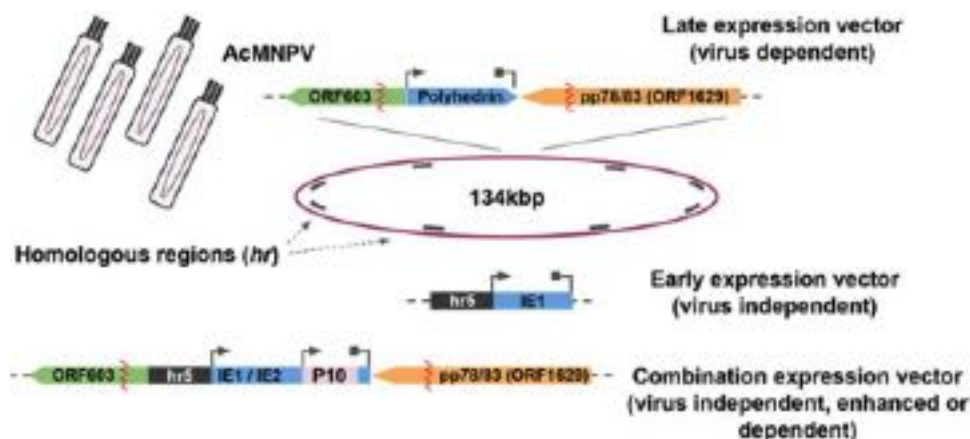


Fig. 3 The baculovirus expression system in which both early and late viral promoter choices are available. The most exploited virus used is *Autographa californica* multiple nuclear polyhedrosis virus. In the canonical system the very late polyhedrin gene is exchanged for a genetic cargo by integration *via* sequences homologous to the flanking genes encoding ORFs 603 and 1629. The approximate extent of sequence used in common vectors is indicated by the red lines. In early promoter vectors, an early promoter such as IE1 is combined with a h sequence, usually hr5, to enable transcription directly upon transfection. In more recently optimized vectors both early and late promoters are present with cloning downstream of both allowing either to operate. Transfection of insect cells results in direct protein expression from either IE1 or IE2 which can be boosted by superinfection by AcMNPV itself which provides the *trans* acting factors to allow p10 to function. True recombinant virus formation can occur *via* the same technology as the canonical example. Irrespective of promoter use the default site of integration is the polyhedrin locus although other sites in the genome have been used. Bent arrows – promoters. Bent blocks – terminators.

productivity following insect cell transfection rivals many mammalian transient expression systems (Fig. 3). The vectors also contain sequences that allow recombination into the baculovirus genome meaning that screening can be performed by transient expression and the most promising construct can be used directly to generate a recombinant virus suitable for scale up.

These three examples, which vary promoter use and the phyla in which they operate, demonstrate the properties of viral promoters that have led to them being so exploited. Indeed, triple viral promoter vectors, comprising the T7 promoter, part of the CMV promoter and a baculovirus promoter are in common use, allowing a screening of expression in various hosts with one plasmid construct.

Further Viral Promoter Use

Viral promoters also feature in less generally used expression systems such as the growing area of expression *in planta*. The attractions of expression in plants are obvious, huge production and lack of adventitious agents, and they remain attractive today even if the vision of edible pharmaceuticals, the “*vaccine in a banana*” has faded somewhat. Virus vectors are an integral part of these developments and a plant virus promoter, the 35S promoter from Cauliflower Mosaic Virus (CaMV), dominates the expression field. 35S is one of two promoters found in the virus and is both strong and relatively cell type independent. Unlike the very late promoters of baculoviruses the 35S promoter is not used as part of recombinant virus system, the plant cell wall limiting how easily virus infections can be performed. Instead 35S is used alone, in various optimized sequence forms, as part of a genetic construct that is delivered to the plant by the invasive bacterium *Agrobacterium tumefaciens* which transfers a section of its DNA, T-DNA, which itself can be modified to carry other sequence, into the plant cell. The combination of plant cell delivery by *Agrobacterium* with the use of the 35S promoter was the predominant methodology of the rapid expansion in transgenic crop experiments from the mid 1980s onwards until this research area was curtailed by concerns over the release of genetically modified organisms. The almost unparalleled performance of the 35S promoter however has continued in the more recent transient plant expression technologies which, effectively, mass transfect plants with vectors for a limited period before harvesting the expressed product. These vectors are based on several types of plant viruses but are used predominantly in a *deconstructed* form in which plasmid based subgenomic versions of the virus genome are used to derive the nucleic acid that will ultimately be delivered into the plant cell (Fig. 4). One example among many is the use of vectors derived from Cowpea Mosaic Virus (CPMV) which have had their non-translated sequences, the sequences which flank the inserted gene, optimized for transcription and translation rather than replication. Such vectors do not replicate, so efficient delivery *via Agrobacterium* is essential, and at scale this is done by vacuum infiltration of the entire plant rather than localized syringe based delivery to individual leaves, but the core expression cassette remains an optimized CaMV 35S promoter flanked by the optimized CPMV sequences (Fig. 4).

A last example of the disparate use of viral promoters is in the field of RNA vaccines, thrown into focus recently by the need to rapidly develop and deploy vaccines for Covid-19. The benefits of an RNA only approach to a vaccine can be found elsewhere but the technology concerned falls into one or other of two types, direct translation of a vaccine candidate antigen by

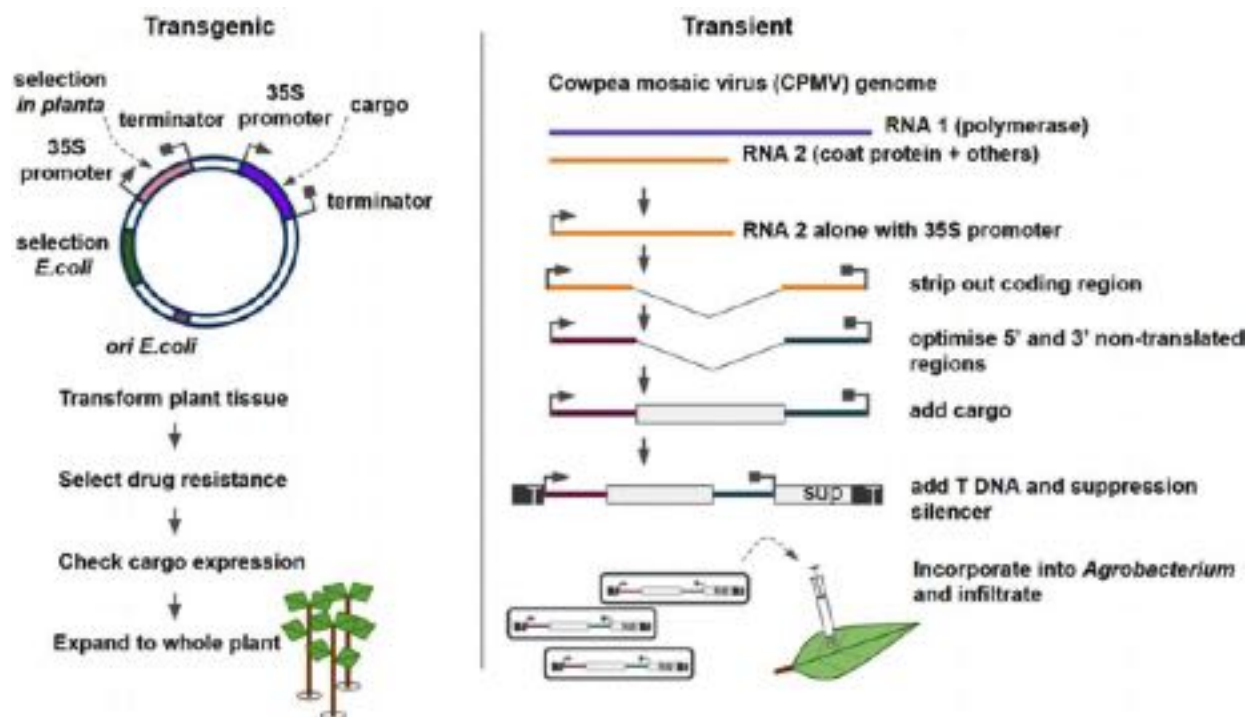


Fig. 4 Viral promoter use in typical plant expression technologies. Left. Transgenic plants are selected following delivery of a vector to plant cells. The Cauliflower mosaic virus 35S promoter is used for expression of both the selectable marker, such as hygromycin resistance, and the transgene. The entire plant is regenerated from the transformed tissue and may be fertile, leading to widespread propagation. Right. Transient plant expression makes use of the RNA2 of Cowpea mosaic virus which normally produces the virus structural proteins. The 35S promoter is used to transcribe the RNA which, as a result of the optimization of the remaining RNA2 sequences is abundantly translated, even in the absence of replication. The system is destructive and cannot be propagated. Bent arrows – promoters. Bent blocks – terminators.

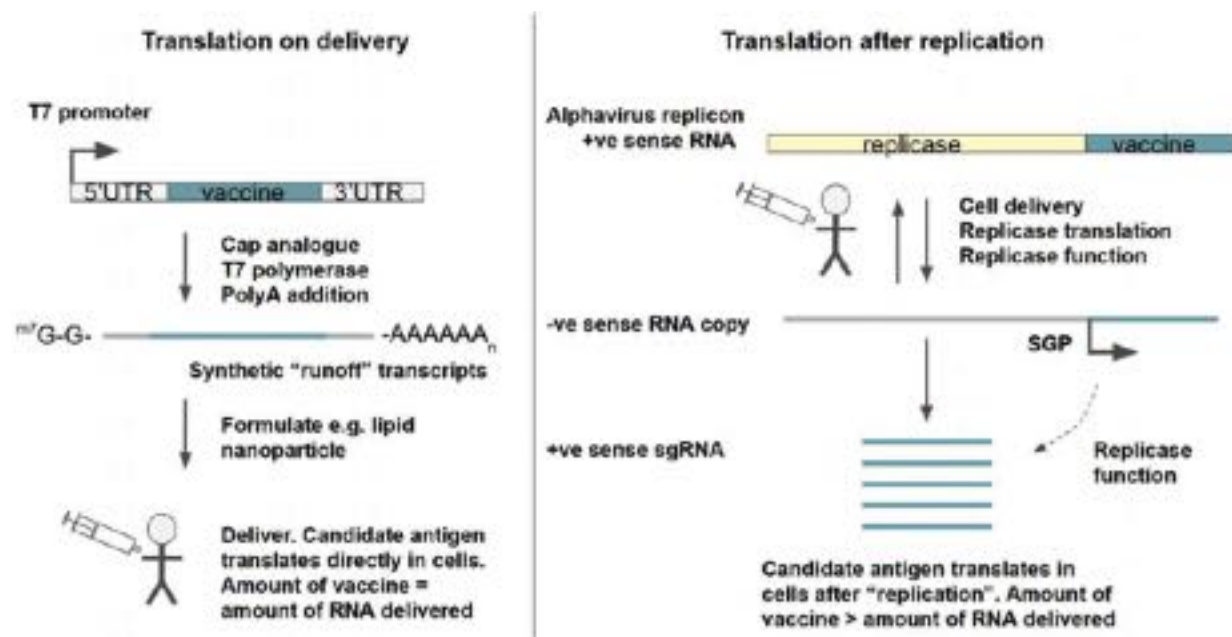


Fig. 5 Viral promoter use in RNA vaccines. Left. Technologies which mimic messenger RNA use efficient *in vitro* transcription with T7 promoter and polymerase to produce copious amounts of a synthetic RNA with optimized 5' and 3' UTRs. The message mimic is achieved by using a cap analog to initiate transcription and including a polyA tail which can be added co- or post-transcription. The transcript may also include modified bases to minimize degradation. The amount of vaccine antigen is directly related to the amount of RNA delivered. Right. Technologies that use self-amplifying RNA make use of the natural replication cycle of alphaviruses but substitute the structural genes for a candidate vaccine. Vaccine antigen is not immediately produced on vector delivery. Rather RNA replication occurs to reveal the SPG promoter on the negative strand and further transcription from it produces the sgRNA messages that encode the vaccine antigen. The system is multiplicative with many more copies of the message produced than numbers of the original RNAs delivered. Bent arrows - promoters.

a synthetic messenger RNA delivered to cells by transfection or a self-replicating RNA which, as a consequence of its design, also produces a translatable message. Both technologies use strong viral promoters, the former for the enzymatic synthesis of the RNA *in vitro* that is to be used, using the phage polymerases discussed for T7 and its relatives, and the latter a specific viral promoter active following RNA replication. The advantage of the self-replicating technology is that lower amounts of RNA need to be delivered to cells to produce the same therapeutic outcome. The self-replicating facility, and the promoter used to drive expression of the vaccine antigen, exploit the alphaviruses noted briefly at the beginning of this article. Alphaviruses have message-like positive sense RNA as their genome but only translate two thirds of it on virus entry into the cell. This produces the viral replicase which sets about replicating the genome to produce a complete negative sense copy of the genome. The negative sense RNA acts as template for more positive sense RNAs including, at this stage, large amounts of a subgenomic RNA (sgRNA) which arises from an internal promoter (SGP) only active in the negative sense strand. In the virus the sgRNA encodes the structural proteins and the use of the incoming template or the later sgRNA effectively mimics the early/late switch of promoter already described for viruses generally. If the virus structural proteins are replaced by sequence encoding a candidate vaccine antigen the hybrid provides a self-amplifying RNA system in which a single incoming transcript generates multiple messages encoding the vaccine antigen and the process continues whilst the replicase remains functional (Fig. 5). Like the polyhedrin or T7 examples the use of SGP is virus specific as it is recognized only by the alphavirus replicase, it has no value for use in generalized expression vectors.

Viruses are the chancers of the natural world and their ability to try and replicate in any cell they encounter underpins their pathology and zoonotic potential. Natural barriers to infection prevent this occurring on a wide scale, from the selective use of receptors, through induction of immunity to specific requirements for virus assembly and release. However, for many viruses bypassing these blocks, by delivery of virus nucleic acid directly into cells, has demonstrated their ability to complete at least part of their replication cycle in many cell types. This ability derives from early viral promoters being able to function in many cellular environments and, if the required *trans* acting factors are present, the ability of the late promoters to drive synthesis of the virus structural proteins. This mix of promoter types, fundamental to virus biology, has been widely exploited in the expression vectors in use today. Understanding virus promoter function, defining the sequences required and their boundaries, has allowed their optimization and, increasingly, their combination with control elements has permitted their development as part of very finely tuned and productive expression applications. A promoter promotes, and in the case of viral promoters that promotion has been of an innovative field of biotechnology as much as of the transcript itself.

Further Reading

- Barrow, N.S., Alderton, D., Sainsbury, S., *et al.*, 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research* 35 (6), e45. doi:10.1093/nar/gkm047.
- Bleckmann, M., Schürig, M., Chen, F.F., *et al.*, 2016. Identification of essential genetic baculoviral elements for recombinant protein expression by transactivation in Sf21 insect cells. *PLoS One* 11 (3), e0149424. doi:10.1371/journal.pone.0149424.
- Chambers, A.C., Aksular, M., Graves, L.P., *et al.*, 2018. Overview of the baculovirus expression system. *Current Protocols in Protein Science* 91, 5.4.1–5.4.6. doi:10.1002/cpps.47.
- Chang, A.C., Nunberg, J.H., Kaufman, R.J., *et al.*, 1978. Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. *Nature* 275 (5681), 617–624. doi:10.1038/275617a0.
- Feldman, R.A., Fuhr, R., Smolenov, I., *et al.*, 2019. mRNA vaccines against H10N8 and H7N9 influenza viruses of pandemic potential are immunogenic and well tolerated in healthy adults in phase 1 randomized clinical trials. *Vaccine* 37 (25), 3326–3334. doi:10.1016/j.vaccine.2019.04.074.
- Hefferon, K., 2017. Plant virus expression vectors: A powerhouse for global health. *Biomedicine* 5 (3), 44. doi:10.3390/biomedicine5030044.
- Itakura, K., Hirose, T., Crea, R., *et al.*, 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 198 (4321), 1056–1063. doi:10.1126/science.412251.
- Sainsbury, F., Cañizares, M.C., Lomonosoff, G.P., 2010. Cowpea mosaic virus: The plant virus-based biotechnology workhorse. *Annual Review of Phytopathology* 48, 437–455. doi:10.1146/annurev-phyto-073009-114242.
- Sawicki, J.A., Morris, R.J., Monks, B., Sakai, K., Miyazaki, J., 1998. A composite CMV-IE enhancer/beta-actin promoter is ubiquitously expressed in mouse cutaneous epithelium. *Experimental Cell Research* 244 (1), 367–369. doi:10.1006/excr.1998.4175.
- Studier, F.W., 2018. T7 expression systems for inducible production of proteins from cloned genes in *E. coli*. *Current Protocols in Molecular Biology* 124 (1), e63. doi:10.1002/cpmb.63.
- Vogel, A.B., Lambert, L., Kinnear, E., *et al.*, 2018. Self-amplifying RNA vaccines give equivalent protection against influenza to mRNA vaccines but at much lower doses. *Molecular Therapy* 26 (2), 446–455. doi:10.1016/j.ymthe.2017.11.017.
- Wang, W., Li, Y., Wang, Y., *et al.*, 2018. Bacteriophage T7 transcription system: an enabling tool in synthetic biology. *Biotechnology Advances* 36 (8), 2129–2137. doi:10.1016/j.biotechadv.2018.10.001.

Oncolytic Viruses

Laura Burga and Mihnea Bostina, University of Otago, Dunedin, New Zealand

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Ad	Adenovirus	ICI	Immune checkpoint inhibitors
ANTXR1	Anthrax toxin receptor 1	ICP	Infected cell protein
ATP	Adenosine triphosphate	IFN	Interferon
CD	Cluster of differentiation	OV	Oncolytic virus
CRT	Calreticulin	PAMP	Pathogen-associated molecular pattern
CTLA4	Cytotoxic T-lymphocyte-associated protein 4	PD-1	Programmed death receptor 1
DAF	Decay-accelerating factor	PD-L1	Programmed death receptor ligand 1
DAMPs	Danger associated molecular patterns	RB	Retinoblastoma
GM-CSF	Granulocyte-macrophage colony stimulating factor	TAA	Tumor-associated antigens
HMGB1	High-mobility group protein box 1	TEM8	Tumor endothelial marker 8
HSV	Herpes simplex virus	TIM3	T cell immunoglobulin and mucin domain 3
ICAM1	Intercellular adhesion molecule 1	TK	Thymidine kinase
ICD	Immunogenic cell death	T-VEC	Talimogene Laherparepvec
		VEGF	Vascular endothelial growth factor
		VSV	Vesicular stomatitis virus

Glossary

Cancer vaccine	A method of treatment that boosts the immune system to fight cancer cells.	Immunotherapy	A method to treat cancer using components of the immune system.
Cold/hot tumor	Tumors that do not allow infiltration of T-cells are characterized by a cold environment; on the contrary tumors infiltrated by immune cells are displaying a hot environment.	Oncolytic virus	A virus that preferentially infects and kills cancer cells.
Combination therapy	A method of treatment that combines two or more therapeutic agents.	Pattern-recognition receptors	Cellular receptors specialized in recognizing antigens found in pathogens; they have an essential role in innate immunity.
Immune checkpoint inhibitors	Therapeutics that block immune checkpoint pathways.	Targeted therapy	A method of treatment that identifies and attacks a certain type of cell.
Immunogenic cell death	A type of cell death that triggers an immune response.	Tumor-associated antigen	Antigens preferentially expressed in tumor cells.

Introduction

Oncolytic viruses (OVs) are replication competent viruses that infect and destroy cancer cells without harming adjacent normal tissue. Evidence for the use of viruses for cancer treatment can be traced to more than a century ago. Several case reports of leukemia and lymphoma patients described signs of transient disease regression following viral infections such as influenza, measles, chicken pox, hepatitis and glandular fever, while a 1910 report presented a case of cervical cancer regression after immunization with the live attenuated rabies vaccine. In the second half of the 20th century, major advances in electron microscopy and tissue culture techniques accelerated the understanding and scientific investigation of viruses. The first clinical trial of an OV started in 1949 to assess the effect of hepatitis B virus in a cohort of 22 patients with Hodgkin's disease. Several patients had a positive response to treatment. However, more than half developed hepatitis resulting in several deaths, highlighting the importance of thorough preclinical investigations. The introduction of animal cancer models offered this opportunity and many other naturally occurring human and animal pathogens were evaluated for their anti-cancer potential such as Egypt 101 virus, Coxsackievirus, adenoidal-pharyngeal-conjunctival virus, and mumps virus. Although full of enthusiasm, the early era of oncolytic virotherapy research and clinical studies was also riddled with unethical practices and was followed by a decrease in scientific interest in the 1980s and 1990s.

In the 1990s, major advances in molecular biology opened the possibility to engineer OVs with increased tumor specificity and reduced pathogenicity. The first OV to be adapted for targeted cancer therapy was herpes simplex-1 (HSV-1). Mutations in the gene encoding for thymidine kinase rendered the virus incapable to propagate in non-dividing cells, thus limiting virus replication to actively dividing tumor cells. Extensive investigations of numerous OV platforms culminated with the first clinical approval of adenovirus-derived Oncorine for the treatment of head and neck cancers in China in 2005, and the first global clinical approval of the modified HSV-1 Talimogene Laherparepvec (T-VEC) for the treatment of unresectable melanoma in 2015. These major

milestones are complemented by the numerous clinical trials of OV currently underway that have resulted from steady scientific progress in oncolytic virotherapy over the last decades.

Modes of Action

OVs employ several mechanisms to infect and destroy cancer cells. In a direct mechanism, OV specifically infect and replicate in tumor cells and the tumor-associated microenvironment, leading to oncolysis and the release of new virus particles that continue to infect neighboring malignant cells. Additionally, viral particles entering the circulatory system can travel, to detect, infect and destroy remote tumors. However, the major mode of action for OV is the induction of immunogenic cell death (ICD) and the stimulation of a cytotoxic immune response against tumor antigens. Tumor growth is conditioned on escaping the surveillance of the immune system. This local immunosuppression is termed a “cold tumor”. OV-induced ICD engages different cell death pathways to infiltrate the tumor with immune cells and convert the “cold tumor” into a “hot tumor” and primes the immune system against cancer cells. Viruses employ numerous methods of ICD: apoptosis, necrosis, necroptosis, pyroptosis, and autophagic cell death, with each OV inducing a predominant form of cell death. ICD causes the release of tumor-associated antigens (TAAs) and danger-associated molecular patterns (DAMPs) such as surface-expressed calreticulin (ecto-CRT), nuclear high mobility group box 1 (HMGB1), ATP, and the production of type I interferons (IFNs). TAAs and DAMPs are recognized by innate effector cells such as dendritic cells and macrophages that activate tumor-specific T cells to establish adaptive immunity. For example, studies of melanoma patients treated with T-VEC have shown an increase of CD8⁺ and CD4⁺ T cells that was positively correlated with the outcome of treatment.

PVSRIP0 is a chimera OV of polio and rhinovirus that showed efficacy in a phase I clinical trial against recurrent glioblastoma, currently under further investigation in a phase II clinical trial. PVSRIP0 has been shown to promote IFN-dominant activation of dendritic cells and tumor-specific cytotoxic T cells in a range of solid tumors *in vitro*, and in an animal model of melanoma. Treatment of melanoma patients with CAVATAK (Coxsackievirus A21) showed an increase of CD8⁺ T cell infiltrates within the tumor microenvironment and up-regulation of immune-related response genes. In addition, several OVs have shown a potent tumor-associated anti-angiogenic effect. However, the process of cell infection by OVs also causes the release of pathogen-associated molecular patterns (PAMPs) that bind to pattern-recognition receptors (PRRs) and trigger the innate and adaptive anti-viral host immune response. Therefore, the success of oncolytic virotherapy is conferred by a balance between the viral escape mechanisms and the anti-viral and anti-tumor immune response.

Tumor Targeting by OVs

Tumor targeting is a complex process that takes advantage of differences between cancer and normal cells and the genetic features of OVs. Numerous OVs use specific receptors to infect tumors which are often overexpressed or restricted to cancer cells. For example, poliovirus employs cluster of differentiation 155 (CD155) as the cellular gateway; Coxsackievirus A21 targets cells expressing the intercellular adhesion molecule-1 (ICAM-1) and the decay-accelerating factor (DAF); anthrax toxin receptor I (ANTXR1), also known as tumor endothelial marker 8 (TEM8) is the high affinity cellular receptor for oncolytic Seneca Valley Virus; and laminin is the cellular mediator for Sindbis virus cellular entry.

The transformation of a cell upon virus infection (i.e. “the hallmarks of virus infection”) shares many similarities with the malignant transformation of cells (known as the “hallmarks of cancer”). These changes contribute to the preferential targeting of tumors associated with many viruses and can be employed to design highly selective virus vectors that complement the phenotype of cancer cells. Some OVs such as Seneca Valley virus, Newcastle disease virus, reovirus and vesicular stomatitis virus have a natural tropism for certain types of tumors. On the other hand, many OVs are genetically engineered to target and destroy tumor cells, and to reduce their toxicity to the non-transformed tissue.

A common feature for cancer cells is a defective interferon pathway employed to escape immune surveillance and apoptosis. At the same time, the loss of interferon defense renders cancer cells more susceptible to virus infection. For example, pelareorep, a wild-type variant of reovirus, is a weak pathogen common to humans that displays oncolytic properties in a variety of cancer types exhibiting activated Ras signaling and a deficient interferon pathway. Recombinant OVs also capitalize on the deficient interferon defense of cancer cells. Vesicular stomatitis virus (VSV) has natural selectivity for IFN-deficient transformed cells. Interestingly, the addition of a human IFN- β gene preserved the cancer cell toxicity of VSV, and decreased toxicity and enhanced the human IFN β -mediated anti-viral activity of neighboring non-transformed cells. VSV-hIFN β is currently under clinical investigations (see “Relevant Websites section”).

The removal of infected cell protein (ICP) 34.5 has eliminated the neurotoxicity of HSV and rendered the virus sensitive to the interferon response in non-transformed cells, thus limiting the virus replication to cancer cells. Clinically approved T-VEC has improved on this platform through additional modifications: the deletion of viral ICP47 and the addition of two copies of human granulocyte-macrophage colony stimulating factor (GM-CSF). Deletion of ICP47 enhanced intra-tumoral virus replication which translated into an increase of tumor cytotoxicity and strengthened the anti-tumor immune response. The addition of GM-CSF transgene is the most utilized modification in the development of recombinant OVs. Local release of GM-CSF promotes dendritic cell maturation that stimulates the anti-tumor immune response by increasing systemic and local CD8⁺ T cells.

Other OV platforms used the deregulation of the cell cycle as a targeted strategy for cancer therapy. Thymidine kinase (TK) is a cell-cycle dependent enzyme and a tumor marker. Deletion of TK in the viral genome causes virus dependence on the cancer cells' high-thymidine nucleotide pool for replication. First reported in 1991, TK deletion is the basis for the tumor targeting ability of the engineered Vaccinia virus. Two modified strains are currently in phase II/III clinical trials for different types of cancer. Other modifications include the addition of GM-CSF in the JX594 strain and the removal of both copies of vaccinia growth factor in Pexa-Vec.

Human adenovirus (Ad) is one of the most investigated oncolytic platforms due to its efficacy and ease of genome manipulation. However, the wild type Ad strains are not specific to cancer cells and display a high immunogenic profile, raising questions about the safety and efficacy of their use. Therefore, oncolytic Ads must be modified for therapeutic use. The first genetic modification of an adenovirus strain removed the E1B-55k protein and was aimed to target p53-deficient cancer cells. E1B-55k is an inhibitor of p53-dependent apoptosis. It was postulated that in normal cells in the absence of E1B-55k, virus infection would trigger p53-dependent cell death and limit virus replication. Therefore, a successful virus infection and replication is limited to inactivated p53 tumor cells. However, in addition to p53 inactivation, E1B-55k mediates late viral RNA export which subsequently was shown to be the trigger for selective amplification of virus in cancer cells. Nevertheless, the rationale for the virus construct design is still valid. Further deletion of E3B gene in a hybrid Ad2/Ad5 virus was designed to escape viral immune surveillance and was tested in clinical trials under the trade name ONYX-015 (Onyx Pharmaceuticals, California, USA) as single therapy or in combination with chemotherapeutics. Onyx-015 was the basis for the very similar virus construct H101 which became the first oncolytic virus to be clinically approved for the treatment of head and neck cancer in China under the tradename Oncorine (Shanghai Sunway Biotech, Shanghai, China). The next generation of modified Ad is represented by the Delta-24-RGD prototype. Based on Ad5, this mutant capitalizes on the array of viral genes that target cell cycle regulators, including the retinoblastoma (Rb) tumor suppressor pathway. Adenovirus replication is dependent on the release of transcription factor E2F from the Rb-E2F complex. A 24-nucleotide deletion in the Rb-binding domain of E1A gene leads to a defective viral protein that is incapable of binding Rb protein and releasing E2F. This limits viral replication to cancer cells where the absence of Rb protein renders E2F free. Further, the insertion of an RGD (Arg-Gly-Asp) peptide in the viral fiber knob increased tumor cell specificity by targeting integrins often expressed in cancers.

A different approach to select for highly efficient oncolytic virus particles that can escape immune surveillance by bioselection or "directed evolution". This strategy was used for the development of enadenotucirev, also known as ColoAd1. Enadenotucirev is a hybrid Ad3/Ad11p generated by serial passaging in colon cancer cells under stringent conditions of a pool of different serotypes of adenovirus. The chimeric virus proved highly efficient in killing epithelial-derived cancer cells and its safety and efficacy profiles have been investigated in a range of clinical trials. Taking it a step further, enadenotucirev was engineered as a drug delivery platform. The viral platform was shown to support the expression of full anti-VEGF antibodies in tumors, without compromising its replication and oncolytic function.

The ability of OVs to stimulate an immune response against tumor cells is exploited in the quest for developing cancer vaccines. As well as the insertion of GM-CSF, arming OVs with immunogenic transgenes such as cytokines, immune checkpoint inhibitors or co-stimulatory molecules was shown to enhance tumor specific immunity in mouse models and human clinical studies.

Combination Therapies

Human tumors are notoriously complex. Understanding the mechanism of virus oncolysis and cancer biology are essential in devising therapeutic strategies, which often include a combination of therapeutic platforms. Numerous studies of oncolytic viruses have shown synergy in combination with standard cancer treatment procedures such as radiotherapy and chemotherapy. Recently, the development of cancer immunotherapy and the emergence of immune checkpoint inhibitors (ICI) have changed the cancer treatment paradigm. Immune checkpoints are molecular pathways that regulate the response of immune cells essential to recognize normal tissue and to promote self-tolerance. Tumor cells have evolved to manipulate these pathways and become invisible to the immune defense by triggering inhibitory signals through the interaction of tumor cell-expressed receptors and their T cell-expressed ligands. For example, the tumor cell programmed death receptor ligand-1 (PD-L1) binds the programmed death receptor 1 (PD-1) on regulatory T-cells; high-mobility group protein box1 (HMGB-1) binds to T cell immunoglobulin and mucin domain 3 (TIM-3); and B7 binds to the cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) expressed in activated T-cells, triggering inhibitory immune response signals against tumor cells. Immune checkpoint inhibitors are antibodies and small molecule chemicals that work to boost the body's immune response to tumor cells. Several ICIs have been tested as monotherapy. Nevertheless, the activity of ICI is enhanced in combination with OVs due to their complementary mode of action. Thus, OVs create a hot tumor environment that has higher immunogenicity, which is a favorable factor for a successful immunotherapy. Indeed, numerous preclinical and clinical trials of combination therapy of OVs and checkpoint inhibitors show promising results. Taking it a step further, arming oncolytic viruses with ICI was also attained. The Western Reserve Vaccinia virus strain modified by inserting anti-PD-1 antibodies was as successful in controlling tumor growth in a mouse model of fibrosarcoma as the two agents administered in combination, and superior to each agent alone.

Conclusion

The development of novel oncolytic virotherapy platforms is an exciting paradigm in the search of innovative cancer therapies. OVs function as self-amplifying drugs with the ability to specifically infect certain cells, making them highly attractive candidates

for the development of targeted cancer therapies. In addition, the possibility to engineer viruses to increase their specificity, reduce toxicity to healthy tissue, and arm them to deliver anti-cancer treatments to tumors, makes them a very versatile tool. There are numerous clinical studies underway that investigate the efficacy of oncolytic viruses either alone or in combination with other therapeutic platforms. The results of these studies will offer more insight into the optimum conditions for treatment, duration and side effects, and will lead to new clinical options for targeted cancer therapy that improve outcomes for patients.

Further Reading

- Andtbacka, R.H.I., Kaufman, H.L., Collichio, F., *et al.*, 2015. Talimogene laherparepvec improves durable response rate in patients with advanced melanoma. *Journal of Clinical Oncology* 33, 2780–2788.
- Bommareddy, P.K., Shettigar, M., Kaufman, H.L., 2018. Integrating oncolytic viruses in combination cancer immunotherapy. *Nature Reviews Immunology* 18, 498–513.
- Filley, A.C., Dey, M., 2017. Immune system, friend or foe of oncolytic virotherapy? *Frontiers in Oncology* 7, 106.
- Garber, K., 2006. China approves world's first oncolytic virus therapy for cancer treatment. *Journal of the National Cancer Institute* 98, 298–300.
- Gromeier, M., Nair, S.K., 2018. Recombinant poliovirus for cancer immunotherapy. *Annual Review of Medicine* 69, 289–299.
- Gujar, S., Bell, J., Diallo, J.S., 2019. SnapShot:Cancer immunotherapy with oncolytic viruses. *Cell* 176, 1240.
- Guo, Z.S., Liu, Z., Bartlett, D.L., 2014. Oncolytic immunotherapy: Dying the right way is a key to eliciting potent antitumor immunity. *Frontiers in Oncology* 4, 74.
- Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: The next generation. *Cell* 144, 646–674.
- Harrington, K., Freeman, D.J., Kelly, B., Harper, J., Soria, J.C., 2019. Optimizing oncolytic virotherapy in cancer treatment. *Nature Reviews Drug Discovery* 18, 689–706.
- Kaufman, H.L., Kohlhaup, F.J., Zloza, A., 2015. Oncolytic viruses: a new class of immunotherapy drugs. *Nature Reviews Drug Discovery* 14, 642.
- Kelly, E., Russell, S.J., 2007. History of oncolytic viruses: genesis to genetic engineering. *Molecular Therapy* 15, 651–659.
- Lawler, S.E., Speranza, M.C., Cho, C.F., Chiocca, E.A., 2017. Oncolytic Viruses in cancer treatment: A review. *JAMA Oncology* 3, 841–849.
- Martin, N.T., Bell, J.C., 2018. Oncolytic virus combination therapy: Killing one bird with two stones. *Molecular Therapy* 26, 1414–1422.
- Russell, S.J., Peng, K.-W., Bell, J.C., 2012. Oncolytic virotherapy. *Nature Biotechnology* 30, 658.
- Senior, M., 2019. Checkpoint inhibitors go viral. *Nature Biotechnology* 37, 12.
- Sivanandam, V., Larocca, C.J., Chen, N.G., Fong, Y., Warner, S.G., 2019. Oncolytic viruses and immune checkpoint inhibition: The best of both worlds. *Molecular Therapy – Oncolytics* 13, 93–106.
- Twumasi-Boateng, K., Pettigrew, J.L., Kwok, Y.Y.E., Bell, J.C., Nelson, B.H., 2018. Oncolytic viruses as engineering platforms for combination immunotherapy. *Nature Reviews Cancer* 18, 419–432.

Relevant Websites

www.clinicaltrials.gov
Home-ClinicalTrials.gov.

Biotechnology Approaches to Modern Vaccine Design

George P Lomonosoff and Daniel Ponndorf, John Innes Centre, Norwich, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Introduction to Virus-Like Particles

Virus-like particles (VLPs) are protein complexes that are similar to, or even indistinguishable from, native virus particles, but that do not contain the viral genome. Hence, they cannot replicate but can mimic the viral antigenicity without being pathogenic. Depending on the biology of the virus, VLPs can be icosahedral or helical (rod-shaped), consist of one or more capsid proteins in multiple copies and may be enveloped or not. While VLPs can occur naturally (e.g., poliovirus empty capsids outside of the cell), they can also be recombinantly produced by expressing the proteins required for VLP production. While some VLPs self-assemble from a single type of coat protein (e.g., L1 of human papillomavirus), other VLPs require the presence of multiple structural proteins (e.g., bluetongue virus VLPs), or a combination of structural and non-structural proteins (e.g., poliovirus VLPs).

Their repetitive surface structure and size of 20–200 nm make VLPs highly immunogenic. They are capable of inducing a strong humoral and cellular immune response and are therefore attractive vaccine candidates for use in humans and other animals. Furthermore, a range of different VLPs have been shown to efficiently present foreign antigens on their surface and/or be loaded with substances of interest, such as, for example, nucleic acids, small peptides and other pharmaceutical substances. These properties make the use of VLPs an attractive approach for the development of new vaccines and as nano carriers for therapy of cancer and chronic diseases (Fig. 1).

Compared to other vaccination strategies, VLPs have advantages regarding safety and efficacy: they are safer than inactivated and live-attenuated viruses, because they cannot replicate and do not require the cultivation of live viruses. Hence, there is no risk of adverse mutations and accidental release of pathogenic viruses to the environment. This, coupled with their size and repetitive structure, makes VLPs ideal vaccine candidates compared to subunit and DNA vaccines.

As VLPs are empty shells, they can be less stable than the native virus in cases where the viral nucleic acid contributes to particle stability. This can cause problems for their production, storage and antigenicity. However, in some cases, for example poliovirus VLPs, the empty capsids can be stabilized by the incorporation of stabilizing mutations within the capsid proteins. Another strategy to improve VLP stability is the production of pseudo-virions, containing non-infectious nucleic acid. This was shown to be effective for HPV VLPs where the co-expression of L1 and L2 of human papillomavirus (HPV) and a circular dsDNA replicon in plants led to the integration of circular DNA into the VLPs and improved their stability compared to empty particles.

Another possible disadvantage with VLPs can be low yield, which makes the production of some VLPs cost intensive. Furthermore, necessary post-translational modifications, for example glycosylation, must be considered during VLP design to achieve the greatest similarity between VLPs and native viruses. These issues can be addressed by choosing a suitable expression system.

Immunological Properties of Virus-Like Particles

VLPs are able to induce both a humoral and cellular immune response as shown in Fig. 2. The high efficiency of VLPs is mainly based on their size, in the range of about 20–200 nm, their highly organized repetitive surface structure and their ability to stimulate toll-like receptors (TLRs). These features are also exploited when VLPs are used to display foreign antigens, unrelated to the VLP itself, and to deliver other pharmaceuticals, for example in cancer therapy or treatment of chronic diseases.

VLPs and the Induction of the Cellular Immune Response

To induce a cellular immune response and activation of cytotoxic T-cells (CTLs) and CD4⁺ T helper cells, antigens must be efficiently taken up by antigen-presenting cells (APC), such as dendritic cells (DC) or macrophages. The efficiency of antigen uptake into APCs depends on size, shape, receptor interactions and surface properties (charge, hydrophobicity, hydrophilicity) of the antigen. Here VLPs are superior compared to larger particles or complex formations over 200 nm. The size of VLPs allows them to freely drain and enter the lymphatic system and lymph nodes without the need for being transported by skin-resident DCs, enabling a fast and efficient uptake of antigen in conventional DCs and lymph node-localized DC subsets such as CD8⁺ lymphoid DCs and plasmacytoid DCs. In addition to the benefits of their size, the shape and surface properties of VLPs are advantageous compared to soluble subunit vaccines. VLPs consist of one or a few proteins in a highly repetitive and organized structure, which differs from almost all macromolecular structures in the host and can be easily detected as foreign by the immune system. Natural IgM antibodies efficiently bind to these surfaces, resulting in activation and recruitment of different complement components, such as mannan-binding protein (C1q) or pentraxin, which facilitates and promotes the trapping of particles in the lymphoid organs and uptake by APCs. Following uptake by APCs, the VLPs are degraded into peptides and can be presented on MHC class I and MHC class II molecules. MHC-class I presentation leads to the priming of CD8⁺ cells, which then mature into memory cytotoxic T cells (CTLs) or effector CTLs. Effector CTLs secrete different cytotoxins, such as perforin and granzymes, which lead to apoptosis of the target cell. MHC-class II molecules activate CD4⁺ cells resulting in the release of T-cell cytokines, which are essential for B-cell antibody class-switching and activation and growth of effector CTLs.

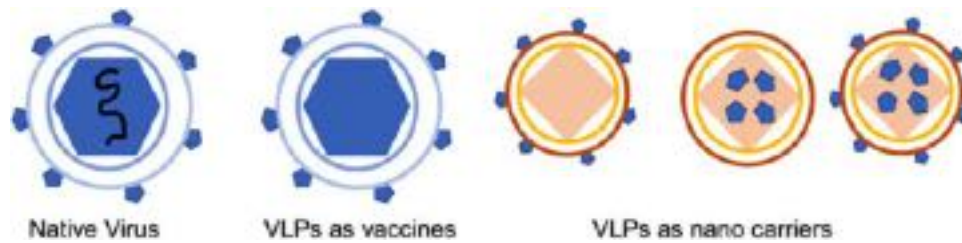


Fig. 1 Biotechnology uses of VLPs: Particles for use directly as vaccines can be produced by expressing the required viral proteins for VLP assembly. These VLPs mimic the antigenic properties of the native virus but do not contain the viral genome. In other approaches, VLPs can be conjugated with one or more antigens and/or loaded with nucleic acid, peptides or other pharmaceutical substances of interest.

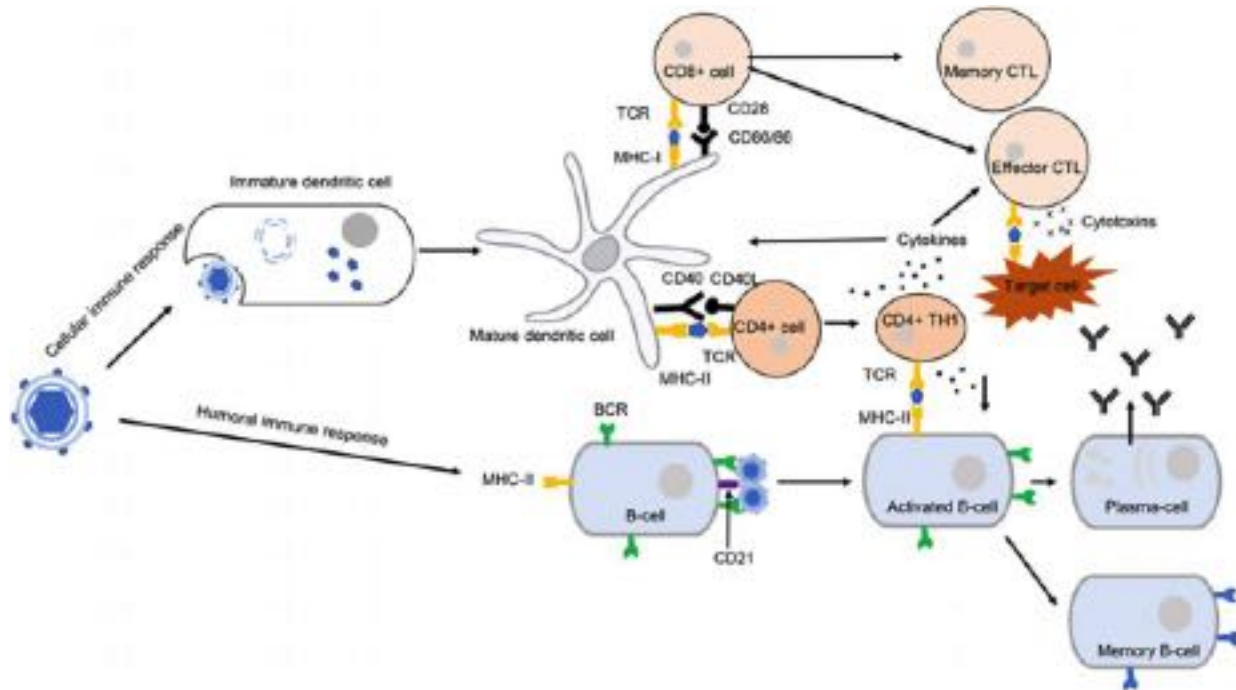


Fig. 2 General induction of humoral and cellular immune response by VLPs: the particles are taken up by antigen-presenting cells, e.g., dendritic cells (DCs) or cross-bind to B-cell receptors (BCR). Antigen processing in DCs leads to presentation of the antigen on MHC-I or MHC-II molecules which leads to the recruitment and activation of $CD8^{+}$ and $CD4^{+}$ cells by binding of the presented antigen to the T-cell receptor (TCR). $CD8^{+}$ cells differentiate into memory and effector CTLs which recognize infected cells and secrete cytotoxins which cause apoptosis of the target cell. $CD4^{+}$ cells differentiate into $CD4^{+}$ helper cells which secrete cytokines, which leads to B-cell antibody class switching and activation and growth of effector CTLs. Activated B cells differentiate into antigen specific plasma cells, which secrete the respective antibody, and antigen-specific memory B cells, the premise for long term immunity.

VLPs and the Induction of the Humoral Immune Response

As described for the cellular immune response, the ability of VLPs to directly enter the lymph nodes has a significant impact on the humoral immune response. By reaching the lymph nodes in the native state, antigens can directly interact with follicular B cells. This is important to directly expose antigens to the B-cells and to develop an antigen depot in the germinal center which maintains B cell activation for several weeks after vaccination. Furthermore, the structure of VLPs enables cross-linking of B cell receptors (BCRs) which is a strong signal for B-cell activation, even in the absence of T-helper cells. The repetitive nature of VLP surfaces activates complement-facilitating B-cell activation through engagement of the CD19-CD20 complex. Cross-linking of this complex and BCRs lowers the required amount of engaged BCRs for B-cell activation and promotes the differentiation of long-lived plasma cells. These plasma cells secrete antigen-specific antibodies which can either bind to the target cell and activate effector T cells or directly neutralize a virus by inhibiting its uptake. Memory B cells carry antigen-specific receptors and enable long-term immunity by recognizing the pathogen and quickly differentiating into pathogen-specific plasma cells.

VLPs and the Stimulation of Toll-Like Receptors

Toll-like receptors (TLRs) are located within the endosome of DCs, B-cells and other activated cells and recognize certain types of nucleic acids including double-stranded RNA (TLR 3), nonmethylated GC-rich DNA (TLR 9) and 5'-triphosphate RNA (TLR 7/8).

These types of nucleic acids are detected by TLRs as an indicator of infection resulting in an enhanced immune response. In line with this, the co-delivery of these TLR ligands during vaccination experiments results in enhanced B- and T-cell response and enhanced IgG response. Hence, administration of TLR ligands, such as CpG-containing oligodeoxynucleotides (CpG-ODNs), a TLR 9 ligand, might result in a stronger and more protective immune response than using empty particles. Because of their previously described size and structural properties, loading of VLPs with TLR ligands enables a more efficient delivery to B- and DC-cells, compared to simply mixing VLPs, and ligands and strongly reduces side effects of “free” TLR ligands.

Expression Systems to Produce Virus-Like Particles

VLPs can be produced in various systems such as bacteria, yeast, mammalian and avian cells, insect cells, plants and plant cell culture as well as cell-free expression systems. Hence, the origin of the VLP is not necessarily important for the choice of the expression system, because VLPs can be produced in an unrelated system (e.g., VLPs from mammalian viruses in plants, VLPs from plant viruses in bacteria). However, the expression system might influence the similarity of VLP to the original virus, especially regarding post-translational modifications (PTMs) and di-sulfide bond-dependent stability. In this section we briefly discuss the different expression systems and the respective advantages and disadvantages for the production of VLPs.

Bacteria

Bacteria are easy to transform, to cultivate and have a high upscaling potential. As a consequence of this, a high yield of VLPs can be achieved in a short time, making them a cost-effective expression system. However, for some VLPs, PTMs and the integration within the host's endoplasmic reticulum (ER) membrane system are necessary to mimic the antigenicity of the original virus and to achieve product stability. Since bacteria are not able to glycosylate recombinant proteins and do not have an ER, they cannot be used to produce these types of VLPs. Bacteria, furthermore, are not a suitable expression system if the VLP stability depends on the formation of the correct disulfide bonds and they can have issues regarding safety because of the presence of endotoxins in the product.

Examples of VLPs successfully expressed in bacteria include HPV, Hepatitis B virus core antigen (HBcAg) and cowpea chlorotic mottle virus (CCMV) VLPs.

Yeast

Yeasts have similar advantages to bacteria in terms of speed but can be used to produce more complex VLPs, including those from enveloped, glycosylated viruses. The pattern of PTMs, however, differs from mammalian cells. Hence, the product similarity to the original virus might be limited. Metabolic engineering to adopt the yeast glycosylation machinery to a mammalian pattern, might solve this issue in the future.

Various VLPs have been successfully produced in yeast, including, for example, dengue virus, poliovirus, hepatitis B virus (HBV) and HPV VLPs. Gardasil®9 (HPV, Merck), Engerix®-B (HBV, GlaxoSmithKline) and Recombivax HB™ (HBV, Merck Sharp and Dohme) are examples for licensed VLP-based vaccines produced in yeast.

Mammalian Cells

The production of VLPs in mammalian cells leads to the closest similarity of a product compared to all other expression systems, if VLPs of mammalian viruses are being produced. However, the high costs of mammalian cell culture and safety concerns, because of possible product contamination with adventitious mammalian pathogens and toxins, remain a challenge for this expression system.

Different mammalian and avian cell lines have been successfully used to produce a broad range of VLPs like for example dengue virus, poliovirus, foot-and-mouth disease virus (FMDV), human respiratory syncytial virus (RSV), Zika virus, HPV, HBV and chikungunya virus VLPs.

Insect Cells

Insect cells can be used in two ways: Infecting cells with baculovirus-based vectors and stable or transient expression in cells without baculovirus infection. Insect cells can be cultivated to high density and large scale and because the PTMs are similar to mammalian cells, they can have an advantage in product similarity compared to the other non-mammalian systems, if mammalian VLPs are produced. The baculovirus design is very time-efficient, but because enveloped baculoviruses are co-produced during product expression, this system has concerns regarding contamination. The latter does not apply to the baculovirus-free system. Examples of VLPs expressed in insect cells include Ebola virus, dengue virus, poliovirus, FMDV, chikungunya virus, human immunodeficiency virus 1 (HIV1), Cervarix® (HPV, GlaxoSmithKline), influenza virus VLPs consisting of haemagglutinin (HA) and matrix protein 1, rabbit hemorrhagic disease virus and porcine parvovirus VLPs.

Plants and Plant Cell Culture

Plants and plant cells can be used to express a broad range of VLPs using two different strategies: stable transformation and transient expression.

In general, plants and plant cell culture are highly scalable but, depending on the method used, the advantages and disadvantages in other areas differ. Stable transformation is time-consuming, and the product yields are usually lower compared to transient expression systems. However, once an elite event is available, whole plants can be the cheapest expression platform, because they only require soil, water and light to grow. By contrast, plant cell culture requires growing media and bioreactors making it more expensive. Another advantage of stable transformed plants and plant cells is that they do not contain endotoxins and are free from mammalian pathogens, a safety benefit of this method. Transient expression leads to higher yields is faster and is essentially a batch process. However, agrobacteria have to be cultured in large quantities and, for large-scale production, plants must be infiltrated using vacuum infiltration systems. This makes it more expensive and, because agrobacteria are used, the endotoxin level of the final product must be monitored. The currently most promising system to produce VLPs in plants is transient expression in *Nicotiana benthamiana*.

The range of plant made VLPs includes, for example, Bluetongue virus (BTV), nervous necrosis virus (NNV), human norovirus, poliovirus, HBcAg VLPs and enveloped influenza VLPs displaying HA (Medicago, Canada). Structural analysis of polio type 3 and NNV revealed that assembly of VLPs in plants lead to high structural similarity compared to the native virus (Fig. 3). Medicago's influenza VLPs represent the first plant-made VLP-based vaccine in Phase III clinical trials and in early 2020 the company announced the successful production of SARS-COV2 VLPs, 20 days after they obtained the necessary genes showing the potential of plants as a rapid VLP production platform.

Cell-Free Expression Systems

Cell-free protein synthesis (CFPS) is based on the use of cell extracts, rather than on living cells. This has the advantage that reaction conditions can be easily controlled and adjusted. CFPS is not sensitive to toxic products and endogenous DNA and mRNA can be removed prior to VLP production, preventing the expression of unwanted proteins. Furthermore, unnatural- or radioactively labeled amino acids can easily be added to the reaction mix for monitoring VLP synthesis. However, upscaling of this system is an issue and is not yet possible at industrial scale.

Cell-free extracts can be based on cell lysates of different organisms and a range of CFPS like *E. coli*, rabbit reticulocytes, wheat germ, insect cells and Yeast *Kluyveromyces* are commercially available. *E. coli* and yeast based CFBS have been successfully used to produced VLPs of different viruses and bacteriophages including M2 bacteriophage, Q β bacteriophage, HBcAg and HPV.

Virus-Like Particles as Nano-Carriers

Besides their direct use as vaccines, VLPs can be used as antigen and/or substance carriers for the development of vaccines against different pathogens and for therapy of chronic infections or cancer. The carrier and therapeutic VLPs do not have to be based on human viruses, but can be derived from bacteriophages, insect-, animal- and plant viruses. Indeed, the use of non-human viruses as nano-carriers has the advantage that there is no pre-existing immune response, such as the presence of neutralizing antibodies against the carrier. Such pre-existing immunity, arising through infection or vaccination, might reduce the delivery efficiency of the VLPs. However, the most important feature for a VLP as a nano-carrier is high stability and structural flexibility to allow antigen display on the particle surface and the loading of the particles. The combination of both peptide display and loading of particles allows the direct targeting to specific cells when a targeting molecule is displayed on the surface of the VLP. This is mainly exploited in cancer therapy research, where VLPs are loaded with anti-cancer drugs and molecules targeting specific cancer cell receptors are presented on their surface. Frequently used and analyzed VLPs include rabbit hemorrhagic disease virus (RHDV), bacteriophage Q β , bacteriophage AP205, bacteriophage MS2, cucumber mosaic virus (CMV), cowpea mosaic virus (CPMV) and HBV (HBcAg or surface antigen (HBsAg) VLPs). Examples for the successful use of antigen display are the malaria vaccines RTS,S/AS01 Mosquirix™ and R21. These vaccines are based on HBsAg VLPs displaying an antigen consisting of 19 NANP repeats from the central repeat region and the C-terminal region of circumsporozoite protein (CSP) of *Plasmodium falciparum*.

Conjugation of antigens to VLPs is possible using different methods such as genetic integration, chemical conjugation and conjugation via the SpyTag/Catcher system, while loading of VLPs can be achieved by in vitro assembly, pH shift and using the natural permeability of the VLPs.

Conjugation of Antigens and Virus-Like Particles

Genetic fusion

The easiest way to conjugate antigens to VLPs is through integration of the DNA sequence of the respective antigen into the viral sequence. The advantage of genetic fusion compared to chemical conjugation or the SpyTag/SpyCatcher system is that only one construct has to be expressed and purified and no in vitro conjugation after purification is necessary. The antigen of interest can be inserted at different locations within the coat protein such as the C-terminus, N-terminus or in exposed loops on the VLP

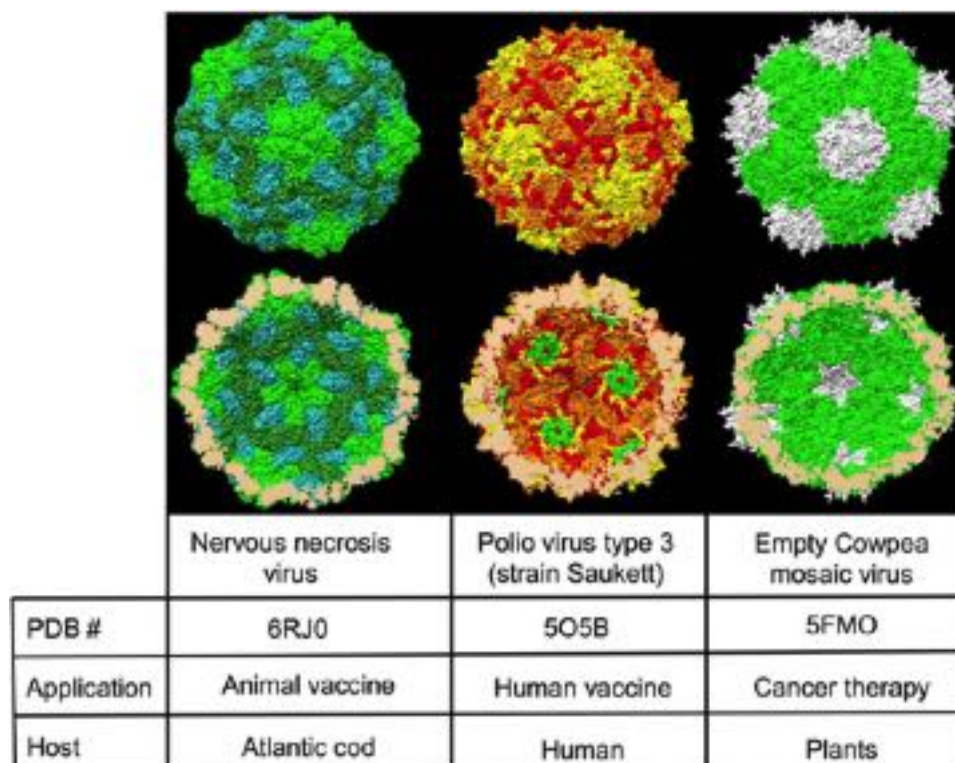


Fig. 3 Examples of plant-made VLPs (not shown to scale). The upper panel shows the external surface and the lower panel the inner surface. Structure determination was carried out by cryo-electron microscopy. Nervous necrosis virus (3.7 Å): The NNV viral asymmetric unit is composed of three copies of the coat protein and three P-domains of the capsid protein interact and form spikes on the particle surface (chain A=light green, chain B=cyan, chain C= dark green). Polio type 3 (3.6 Å): Four structural proteins form the capsid protein: VP1 (yellow), VP2 (orange), VP3 (red) and VP4 (green). Even though VP0 is not cleaved into VP2 and VP4 in plants, the surface structure of assembled VLPs is nearly indistinguishable from wild type virus. Empty cowpea mosaic virus VLPs (2.3 Å): 60 copies of large (green) and small (white) coat protein assemble to T = 3 icosahedral empty VLPs.

surface (e.g., L1 loop on RHDV or the major immunodominant region (MIR) on HBcAg VLPs). For RHDV VLPs it was shown that antigens can be integrated at all three positions at the same time. This allows antigen display on the inside of VLPs, as described for RHDV and BTV. This might be especially of interest for the protection of unstable peptides inside the particles.

However, the genetic fusion strategy has limitations, because insertion of larger antigens or antigens with unfavorable charges can alter the VLP conformation and subsequently the assembly and stability of the particles. Different strategies have been developed to improve the tolerance of VLPs for the presence of larger antigens such as the creation of mosaic particles, the split core and tandem core technology for HBcAg.

To produce mosaic VLPs, the modified viral protein containing the antigen and unmodified viral protein are co-expressed such as in the case of HBcAg/HBsAg proteins. This technology allowed the introduction of up to 213 amino acids of a hantavirus antigen at the C-terminus of a truncated version ($\Delta 144$) of HBcAg. A drawback of mosaic particles is that the ratio of the modified to unmodified protein in the VLPs cannot be easily controlled and thus the VLPs might only contain small amounts of antigen-presenting subunits and some VLPs may not contain any antigen.

Further improvement of antigen integration by genetic fusion was achieved with the split- and tandem core technology. These technologies allow the integration of larger antigens and help to overcome conformational restrictions. The split core technology is based on an artificial separation of the c/e1 loop of HBcAg MIR by the insertion of a stop and a start codon between amino acid 79 and 80, resulting in two open reading frames coreN and coreC without alteration of VLP assembly. The antigen of interest can then be fused to either the C-terminus of the N-terminal half monomer, or the N-terminus of the C-terminal half monomer. The tandem core technology is based on the production of a HBcAg dimer as a single polypeptide chain by fusion of two HBcAg open reading frames. The respective HBcAg proteins are covalently linked by a flexible linker which allows the insertion of large heterologous sequences in only one of the two MIRs in each spike resulting in a higher conformational tolerance of the VLPs after integration of larger antigens or whole proteins such as GFP.

Chemical and affinity conjugation

Chemical conjugation allows the fusion of not only VLPs and larger protein-derived antigens, but also their fusion to non-protein targets, cyclic peptides and whole proteins. This strategy is based on covalent linkage of heterobifunctional conjugation of two

distinct reactive groups that couple two different residues, one on the VLP surface and one on the antigen. Appropriate reactive groups on the carrier VLP are usually the ϵ -amino group of lysine, the carboxyl groups of glutamic and aspartic acids or the sulphhydryl of cysteines. Conjugation of VLP and antigen is then possible via these reactive groups. If the required reaction partner is not naturally present on the VLP surface it can be engineered into the subunits as shown for HBcAg, bacteriophage MS2 and CPMV.

One example of chemical conjugation of a VLP and an antigen is the use of Sulfo-succinimidyl 4-(*N*-maleimidomethyl) cyclohexane-1-carboxylate (Sulfo-SMCC) and a partner that contains a free sulphhydryl group or was treated with *N*-Succinimidyl S-Acetylthioacetate (SATA) to induce the formation of sulphhydryl groups. Treatment of VLPs with Sulfo-SMCC results in maleimide-activated particles which can then be mixed with antigens containing the sulphhydryl group resulting in covalent, irreversible binding of these groups (Fig. 3).

Apart from covalent binding, noncovalent affinity conjugation can be used to fuse VLPs and antigens. This is based on strong interactions between certain partners, such as that between biotin and streptavidin. Another example is the use of a peptide tag that binds to the MIR of HBcAg particles and is based on the natural interaction between HBcAg and HBsAg. However, for the second example a sucrose gradient ultracentrifugation step disrupted the particle-epitope binding indicating instability of this interaction.

Even though chemical conjugation allows the displaying of larger antigens and non-protein conjugation partners, it has drawbacks for downstream processing, because it requires several more steps than using genetic integration thus making it time-consuming and expensive. Chemical conjugation requires the expression and purification of VLPs and conjugation partner in separate steps and after purification one or both partners require chemical activation. The chemical activators have to be removed from the VLP or antigen samples in an additional purification step before the conjugation reaction and finally, uncoupled antigens and VLPs have to be separated.

The SpyTag/SpyCatcher system

Similar to the chemical conjugation approach, the SpyTag/SpyCatcher system is based on the affinity of reactive groups and allows the conjugation of larger antigens to VLPs. In this case, the reactive groups were identified in the CnaB2 domain of the FbaB protein from *Streptococcus pyogenes*. SpyTag is a 13 amino acid long peptide containing aspartic acid as one reactive group, while SpyCatcher is a protein with a size of 12.3 kDa containing a reactive lysine group and a catalytic glutamic acid. In the presence of each other, the carboxyl group of the reactive aspartic acid and the ϵ -amino group of the lysine form a covalent isopeptide bond. Genetic integration of the SpyTag at the N-terminus of a peptide and SpyCatcher on the surface of a VLP (or vice versa) allows the conjugation of the partners. This conjugation can be achieved after their individual expression and purification, followed by an *in vivo* reaction when both components are mixed, or by simply co-expressing the two respective constructs in a single cell. Co-expression might be beneficial, because it is cost saving and simplifies the downstream processing by reducing the number of purification steps required. In a proof of concept study it was shown that co-expression and *in vivo* conjugation of HBcAg VLPs (SpyCatcher) with GFP and P24 respectively (both SpyTag) led to the successful conjugation of VLPs and the respective partners in *Nicotiana benthamiana*. Even though the SpyTag/SpyCatcher system can be very efficient, it is still necessary to distinguish and separate conjugated and non-conjugated partners. This applies for both *in vivo* and *in vitro* conjugation and has to be considered when using this method.

Loading of Virus-Like Particles by Encapsulation

As well as displaying antigens on the outer surface of VLPs, materials can also be incorporated inside the particles. This process is called encapsulating and is not limited to peptides, but also whole proteins, nucleic acids, inorganic compounds and other pharmaceutical substances like cancer drugs can be encapsulated. The main advantages of this delivery system are an improved delivery, the adjuvant effect of the VLPs and the protection of conjugation partners inside of particles against for example proteolysis or thermal denaturation. A broad range of VLPs was so far analyzed for this strategy including VLPs based on bacteriophages (e.g., MS2, Q β , P22), plant viruses (e.g., hibiscus chlorotic ringspot virus (HCRSV), cowpea chlorotic mottle virus (CMV), tobacco mosaic virus), animal viruses (e.g., bluetongue (BTV)) and human viruses (e.g., HBV, SV40).

Two main strategies are available to load particles: statistical- and direct encapsulation. Both strategies can be used *in vitro* and *in vivo*, while for statistical encapsulation the *in vivo* approach is more common.

For statistical encapsulation, VLPs or viral particles are first produced, purified and then disassembled by a pH shift into the viral capsid proteins. If present in the sample, nucleic acid can be removed before cargo and capsid proteins are mixed. The VLPs can reassemble after an additional pH shift and some of the cargo in the sample mix gets encapsulated. For smaller molecules disassembly of VLPs is not obligatory, because they may diffuse through openings in the viral shell, which then can be "closed" by a pH shift resulting in trapping of the cargo inside the particles. However, this strategy requires large amounts of VLP capsid proteins and cargo, and the loading cannot be precisely controlled and might not be reproducible in terms of the amount of encapsulated material. This loading strategy also has requirements regarding to charge and size of the molecule. If the molecule is too small, it may leak out of the VLP shell and for some plant viruses, such as hibiscus chlorotic ringspot virus (HCRSV) and CCMV, it is also necessary that the cargo has a negative charge. Both requirements can be addressed using polyacid association, a process where semi stable complexes between large molecular weight polyacids and the smaller cargo are formed resulting in higher sized aggregates with a negative charge. This enables encapsulation of smaller, positive charged molecules like cancer drugs without altering the deliverability.

Direct encapsulation is similar to surface conjugation and can be based on genetic integration of the sequence of interest within the respective part of the viral genome, the use of specific tags or chemical interactions between amino acid residues of the cargo and the inside of the VLPs. As described for the statistical encapsulation, VLPs or viral particles can be disassembled, mixed with tagged cargo and then reassembled resulting in integration of the tagged material. Furthermore, tagged peptides or proteins can be co-expressed with the capsid proteins or directly genetically fused, saving the *in vitro* steps. This strategy was successfully used to encapsulate proteins and antigens into VLPs of BTV, P22 and RHDV VLPs.

VLP-Based Vaccines for Humans

HPV and HBV

Licensed VLP-based vaccines are currently available against human papilloma- (HPV) and hepatitis B virus.

HPV is a causative agent of cervical-, vaginal-, vulvar, penil-, anal- and oropharyngeal cancer and 200 subtypes of HPV have been isolated. Out of these 200 subtypes, 13 are known as high risk HPV (hrHPV) and 7 of hrHPV (16/18/31/33/45/52/58) are responsible for 98% of all cancers caused by HPV. Low risk HPV subtypes do not cause cancer but induce the formation of genital warts, where subtype 6 and 11 are responsible for 85% of cases. However, the risk of HPV infection can be minimized by vaccination. The current licensed vaccines are based on VLPs which are produced by expressing the L1 coat protein of HPV types in yeast (Gardasil®, Gardasil9®) or insect cells (Cervarix®). Cervarix® is a bivalent vaccine containing VLPs of subtype 16 and 18, while Gardasil and Gardasil9 are quadrivalent (6/11/16/18) or nonavalent (6/11/16/18/31/33/45/52/58) respectively. Immunization with VLPs leads to high antibody titer and long-term protection of vaccinated individuals. While the L1 protein differs between HPV subtypes, the minor capsid protein L2 is highly conserved. However, L2 cannot form VLPs, but L2 display on other VLPs such as Q β - or MS2 is possible and currently pre-clinically tested.

Vaccines against Hepatitis B virus have been available since the 1980s. The first generation of these vaccines were HBsAg particles isolated from the blood of infected individuals. Because of issues like biosafety and tolerability the second generation of vaccines was developed. These vaccines are based on the recombinant expression of HBsAg protein in yeast and the isolation of VLPs. This generation includes licensed vaccines like for example RecombivaxHB or Engerix-B. These vaccines are efficient, safe and have significantly reduced HBV prevalence worldwide. A third generation VLP based vaccine, Sci-B-Vac™ which is produced in Chinese hamster ovary (CHO) cells, is a further improvement. The recombinant HBsAg VLPs mimic three antigenic sites (S, PreS1 and PreS2) instead of only one (S) in the second generation. Sci-B-Vac™ need lower doses compared to the second-generation vaccines and also showed an improved response in obese or older people and in immunocompromised patients, patients with transplants or dialysis and low responders marking a further improvement in HBV vaccination.

Influenza

Besides the licensed examples above, a range of VLP based vaccines is currently in clinical trials. Two examples are Medicago's pandemic influenza vaccine, currently in clinical phase II, and a seasonal influenza vaccine which went through clinical phase III trials and was accepted for scientific review by Health Canada in October 2019 and could be licensed and available in 2020/2021. Both vaccine candidates are produced in plants. The quadrivalent vaccine against seasonal influenza contains VLPs of two strains of influenza A and B respectively, while the pandemic approach uses the respective sequence of the pandemic strain. Influenza VLPs can be produced by the expression of the hemagglutinin (HA) antigen, resulting in assembly of VLPs containing a lipid bilayer envelope. They have different advantages compared to currently licensed products like a short response time and scalability. The response timeline is 3 weeks from obtaining the HA sequence to the first purified VLPs and a production rate of about 1 million doses per month was described. Furthermore, the plant-based recombinant VLP vaccine targets wild-type HA sequences, in contrast to influenza strains grown in eggs or tissue culture that may have mutated during culture.

Polio

Poliovirus is the causative agent of poliomyelitis, a disease which can lead to paralysis of limbs or even death. Introduction of live attenuated (OPV) and inactivated polio vaccines (IPV) led to a 99% decrease in polio cases and wild-type polio type 2 and 3 were declared eradicated by the world health organization (WHO) in 2016 and 2019, respectively. In 2019 about 125 cases of wild type polio 1 infections were recorded, while about 245 cases of vaccine-derived poliomyelitis occurred. These cases are mainly caused by reverse mutation of attenuated polio strains in OPV resulting in regained pathogenicity of the attenuated virus. Thus, while OPV and IPV are efficient and successful vaccines, they are not suitable for the final eradication of poliovirus. Hence, a virus-free polio vaccine is of great interest and VLPs are a promising candidate vaccine.

However, the production of Polio VLPs is not straight forward, because empty polio capsids based on wild-type sequences are antigenically unstable and switch from D to C antigenicity, which makes them unable to induce protective immunity and thus are unsuitable for use as vaccines. This issue was solved by introducing stabilizing mutations into structural proteins resulting in predominant D antigenicity of VLPs. Polio VLPs can be produced by expressing the stabilized structural polyprotein P1 and the viral protease 3CD in various expression systems such as plants, yeast, insect cells, mammalian cells and cell free expression systems.

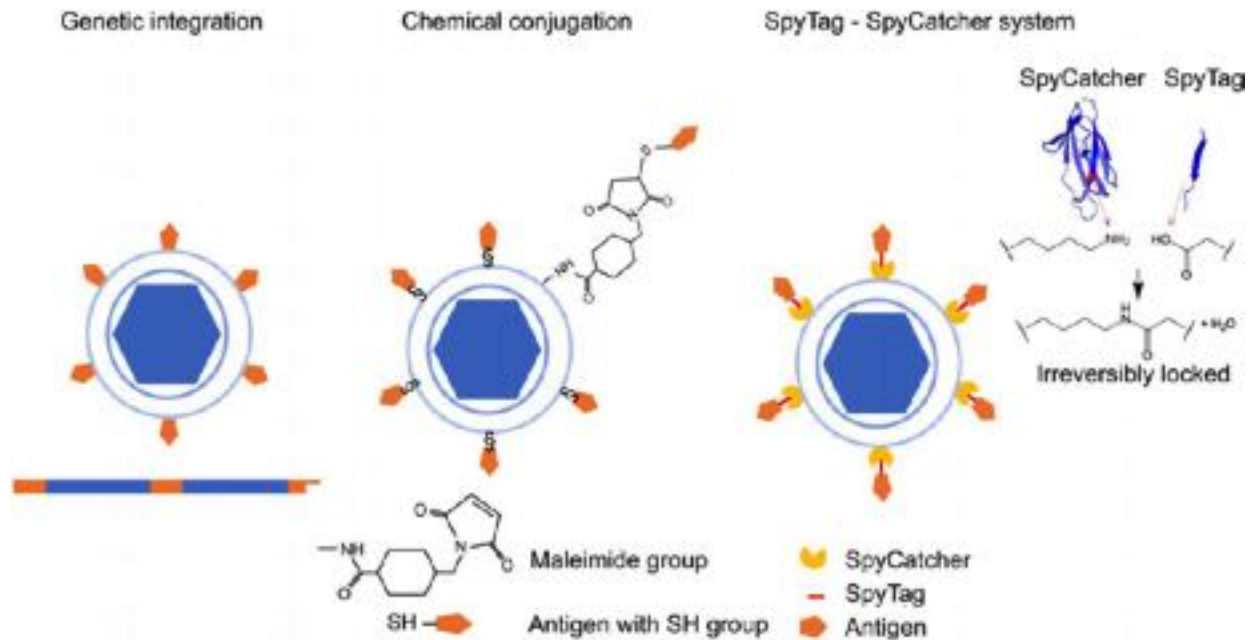


Fig. 4 Methods for antigen display on the surface of VLPs. Genetic integration of the antigen sequence (orange) into the sequence encoding the viral capsid protein is possible at multiple sites, including the N-terminus, the C-terminus or in exposed regions on the VLP surface (e.g., c/e1 loop of HBcAg). Chemical conjugation is based on reactive groups in the two partners. For chemical conjugations the two partners are produced separately and VLPs can be maleimide activated using Sulfo-SMCC which then reacts with the free sulfhydryl group on the antigen. The SpyCatcher/SpyTag system is based on a similar principle. The 13 amino acid long SpyTag carries an aspartic acid residue while the SpyCatcher protein contains a reactive Lysine and a catalytic glutamic acid residue. Once SpyTag and SpyCatcher interact the aspartic acid and lysine irreversibly form an isopeptide bond. SpyTag-Spy catcher reaction with kind permission of PNAS (Zakeri *et al.*, 2012. Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. PNAS 109 (12) 4347–4348).

In plants, efficient processing of the P1 polyprotein into VP1, VP3 and VP0 was confirmed and assembly of the structural proteins resulted in VLPs almost indistinguishable from wild type polio virus (Fig. 4). Preclinical tests in transgenic polio receptor-expressing mice showed that VLPs protected mice from a polio challenge as efficiently as the IPV control. However, this is only true for Polio type 1 and 3, while the production of stable Polio type 2 VLPs remains a challenge in all expression systems.

VLPs as Cancer Vaccines and Therapeutics

VLPs have been intensively investigated for their suitability as a cancer vaccine and as cancer therapeutics. Antigen display and loading of particles, as well as the use of empty CPMV VLPs, have shown great potential in both fields and multiple studies are currently in preclinical, clinical phase I and clinical phase II trials.

For the vaccine approach, the protection induced by VLPs is based on T cell and/or antibody responses. Various VLPs such as RHDV, polyomavirus, cucumber mosaic virus (CMV), CPMV, HBcAg, bacteriophage Q β , bacteriophage MS2 and bacteriophage AP205 have been analyzed for treatment of different cancers including melanoma, breast cancer, pancreatic cancer, cervical cancer and hepatocellular carcinoma. Three different strategies have been mainly analyzed so far: presentation of a single antigen, display of multiple antigens for a certain cancer type and the use of empty VLPs without the presentation of any antigens. Furthermore, these strategies can be combined with the use of different adjuvants, TLRs or checkpoint inhibitors to optimize immunogenicity. The use of multiple antigens seems to be beneficial compared to a single antigen, because of possible outgrowth of antigen-loss tumor cells and higher efficiency. This was shown for RHDV VLPs displaying epitopes from murine topoisomerase IIa and survivin, either in a mono- or multi-target form. Here, the multi-target form significantly prolonged the vaccine-induced remission period compared to the mono therapy approach. Interestingly, empty CPMV and Q β bacteriophage VLPs loaded with A-type CpGs showed promising results in the treatment of different melanoma types without presenting any heterologous antigens. *In situ* vaccination of a lung melanoma with empty CPMV VLPs led to systemic anti-tumour immunity against B16F10 melanoma. This shows that VLPs, even if they are derived from plant viruses, can be highly immunogenic without displaying specific antigens or carrying any adjuvants.

As described before, VLPs can also be loaded with nucleic acids (e.g., mRNAs, micro RNAs, small interfering RNAs), peptides and other pharmaceutical substances as smaller molecules such as cancer therapeutics (e.g., taxol and doxorubicin). Additionally, they can be designed to target specific cancer cells, while leaving healthy cells unaltered. This is possible by conjugation of VLPs with specific targeting ligands such as glycans, receptor-binding ligands (e.g., folate and transferrin), specific antibody fragments

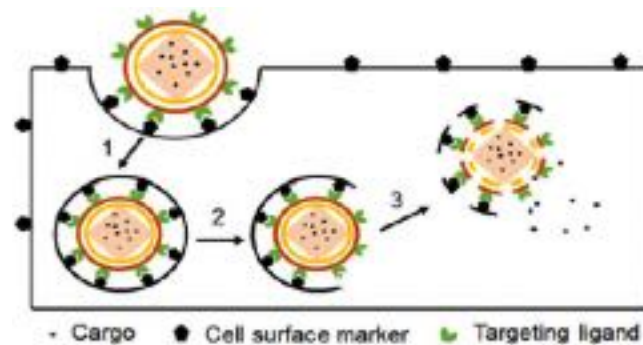


Fig. 5 Cell-specific cargo delivery using VLPs. VLPs are loaded with cargo and display cell-specific target ligands that recognize specific cell surface markers. After binding of VLPs to the markers, endocytosis is triggered (1) and after the VLPs escape from the endosome (2) they get degraded and release the cargo into the cytosol of the target cell (3).

(e.g., anti-HER2 to target melanoma cells or Trastuzumab for breast cancer cells) and RNA- and DNA aptamers (e.g., DNA aptamer with affinity to target tyrosine kinase 7 receptors on Jurkat leukemia T cells).

However, using nano particles as a delivery system has certain challenges. The particles need to remain stable to protect the cargo (e.g., to shield the RNA from RNases) and to avoid nonspecific engulfment by phagocytosis. After endocytosis into the target cell, the carrier needs to be able to escape from the endosome (Fig. 5), get degraded and release the cargo into the cytosol of the host cell. Hereby, the release of the cargo at once is more efficient than a slow release (as described for polymer nano particles). It was shown in animal models and cell culture that a range of VLPs like CCMV, CPMV, Q β , MS2 and P22 meet these criteria and that VLPs are promising nano carriers for cancer treatment using cargo delivery inside of particles.

Further Reading

- Bachmann, M.F., Jennings, G.T., 2010. Vaccine delivery: A matter of size, geometry, kinetics and molecular patterns. *Nature Reviews* 20, 787–796.
- Donaldson, B., Al-Barwani, F., Pelham, S.J., *et al.*, 2017. Multi-target chimeric VLP as a therapeutic vaccine in a model of colorectal cancer. *Journal for Immuno Therapy of Cancer* 5, 69.
- Fuenmayor, J., Godia, F., Cervera, L., 2017. Production of virus-like particles for vaccines. *New Biotechnology* 39, 174–180.
- Khudyakov, Y., Pumpens, P., 2016. *Viral Nanobiotechnology*. Boca Raton FL: CRC Press.
- Lizotte, P.H., Wen, A.M., Sheen, M.R., *et al.*, 2016. In situ vaccination with cowpea mosaic virus nanoparticles suppresses metastatic cancer. *Nature Nanotechnology* 11 (3), 295–303.
- Marsian, J., Fox, H., Bahar, M.W., *et al.*, 2017. Plant-made polio type 3 stabilized VLPs-a candidate synthetic polio vaccine. *Nature Communications* 8, 245.
- Mohsen, M.O., Spenser, D.E., Knuth, A., Bachmann, M.F., 2020. Virus-like particles for vaccination against cancer. *WIREs Nanomedicine and Nanobiotechnology* 12, e1579.
- Mohsen, M.O., Zah, L., Cabral-Miranda, G., Bachmann, M.F., 2017. Major findings and recent advances in virus-like particle (VLP)-based vaccines. *Seminars in Immunology* 34, 123–132.
- Moreno, N., Mena, I., Angulo, I., *et al.*, 1844. Rabbit hemorrhagic disease virus capsid, a versatile platform for foreign B-cell epitope display inducing protective humoral immune responses. *Scientific Reports* 6, 3.
- Rohovie, M.J., Nagasawa, M., Swartz, J.R., 2017. Virus-like particles: Next-generation nanoparticles for targeted therapeutic delivery. *Bioengineering and Translational Medicine* 2, 43–57.
- Rybicki, E., 2019. Plant molecular farming of virus-like nanoparticles as vaccines and reagents. *WIREs Nanomedicine and Nanobiotechnology*. 12. e1587.
- Sarkar, B., Islam, S.S., Zohora, U.S., Ullah, A., 2019. Virus-like particles – A recent advancement in vaccine development. *Korean Journal of Microbiology* 55 (4), 327–343.
- Wege, C., Lomonosoff, G.P., 2018. *Virus-Derived Nanoparticles for Advanced Technologies*. New York: Springer Nature.

Viruses: Impact on Science and Society

Neeraja Sankaran, Utrecht University, Utrecht, The Netherlands

Robin A Weiss, University College London, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Introduction

When we began writing this article in April/May 2020, we were in the middle of the first 'lockdown' in the UK, a quarantine measure adopted in most countries in attempts to contain the exponentially spreading viral disease we call COVID-19. Caused by a novel strain of coronavirus, named SARS-CoV-2, the COVID-19 pandemic, which was first recognized in Wuhan, China in December 2019, had by May 2020 knocked trillions of dollars off international markets and drastically changed the way we lived, at least for the time being. Daily life and livelihoods have been paralyzed by the mandatory, but unfortunately necessary, quarantine measures, separating us from our families, friends and colleagues. Hospitals were overflowing while stores, theaters and restaurants have been closed, many already bankrupt.

Months later, as this article goes to print (November 2020), the death toll of COVID-19 has exceeded one million, and there are no signs that the pandemic is abating, let alone ending. Indeed, some medical historians contend that epidemic endings are a form of collective amnesia, transmuting the disease that remains into merely someone else's problem. The current coronavirus pandemic is neither the first, nor likely the last, viral visitation that humans have experienced on a mass scale. The impact of viruses on our lives predates our knowledge of their existence, as manifested in the way societies have responded to viral epidemics regardless of whether the cause was known at the time.

Often described as entities hovering on the threshold of life, viruses are found in parasitic, commensal or symbiotic association with the entire range of living organisms from single celled Archaea and bacteria to plants, animals and humans. They play a dual role in the laboratory, both as objects of investigation and as experimental tools. They have greatly enhanced our understanding, not only of human disease, but also of the living world more broadly.

In this article we consider the tremendous impact, past and present (and undoubtedly future), of viruses on humans, in both science and society. Recognizing that every interaction is a two-way street, we also remark on the ways that human activities – both deliberate, e.g., in the way we use them in the laboratory, and unknowing, through lifestyle choices – have affected viruses. We also examine how the modern, interconnected world has little resilience in terms of socio-economic consequences of the emergence of a pandemic virus.

Virus: Meanings and Metaphor

Expressions such as "going viral" are so widely used today that it is easy to forget that viruses have a narrower, more precise meaning for scientists than for the public at large. In popular usage, the term is frequently elided with any invisible germ (much to the irritation of virologists), but this imprecision is not a serious matter unless it leads to inappropriate treatment of virus infections by antibiotics that target bacteria.

Etymology and Evolution of the Meaning of Virus

The notion of contagion by invisible agents specific for different diseases was expounded by Girolamo Fracastoro in 1546, and the word 'virus' appears to have been imported into the English language from Latin as a word for poison or venom sometime in the sixteenth century. One of the earliest documented references is in Edward Jenner's classic 1798 report on the prevention of smallpox. Jenner's use of the word virus was not precise, and it was not until the late nineteenth century in the wake of the germ theory that the word virus came to mean a type of disease agent. As the ability of bacteria to cause specific infectious diseases, e.g., tuberculosis, cholera and anthrax, was established, scientists found other infectious diseases from which they could not isolate bacteria. The two earliest examples were mosaic disease of tobacco plants by Dmitri Ivanowski (1892) and foot-and-mouth disease of cattle and swine by Friedrich Löffler and Paul Frosch (1898). They still considered the causative agents to be bacteria, only smaller than known ones. But in 1898, Martinus Beijerinck proposed that the tobacco mosaic agent (what we now call tobacco mosaic virus, TMV) is distinct from bacteria in being somehow dependent for its multiplication on the cells of its host. He described TMV as a *contagium vivum fluidum*, Latin for a 'living soluble contagion'. Without adequate techniques to test it, Beijerinck's idea only gained traction 30 years later, when it helped establish virology as an independent discipline in its own right.

The next couple of decades witnessed many discoveries of viruses and viral diseases, as well as contentious debates over the nature of viruses as living versus non-living organisms. Virological concepts deepened with the development of new instruments and methods of visualizing and cultivating viruses, as well as the emergence of molecular biology, which gave scientists both new

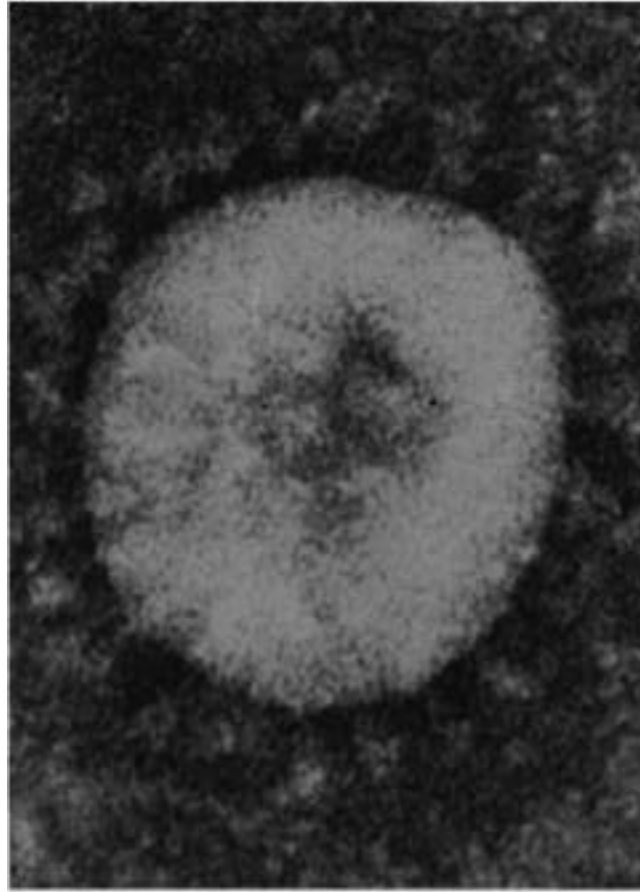


Fig. 1 Electron micrograph of a human common cold coronavirus virus particle (~ 100 nm diameter) of isolate B814 taken by June Almeida in 1966 using the new technique of “negative staining”. It was the ‘characteristic fringe of projections which are rounded or petal shaped, recalling the solar corona’ that led to the adoption of the name Coronavirus for this family in 1968. These projections depict the “S” envelope glycoprotein. Reproduced by permission of the Microbiology Society. From Almeida, J.D., Tyrrell, D.A.H., 1967. The morphology of three previously uncharacterized human respiratory viruses that grow in organ culture. *Journal of General Virology* 1 (2), 175–178.

techniques and a new language. In the 1950s, the French microbiologist, André Lwoff, formulated the modern definition of viruses as ‘obligate intracellular parasites’ composed of a single type of nucleic acid (either DNA or RNA) encased in a protein coat.

Virus as Metaphor

It seems a valid description to call self-propagating malware a “computer virus”, given that it can spread rapidly from one computer to another, reproducing itself and lead to malicious outbreaks. Similarly, descriptions of rumors on social media going “viral” are apt. They allow “pathogenic” entities such as fake news and conspiracy theories to germinate, propagate and become pandemic, such as SARS-CoV-2 escaping from a germ warfare laboratory – or the influence of G5 networks on its spread. Similar conspiracy theories were applied to the emergence of HIV and other viruses in the past.

In times of pestilence, myths of blame or denial often arise. Among the powerful, the blame game for viruses is evident today. As for denial, the former President of South Africa, Thabo Mbeki, became convinced by the notion propagated on the web that HIV did not cause AIDS, and he considered that anti-retroviral drugs were a ruse by Western pharmaceutical companies to profit at Africa’s expense. His ban on providing anti-retrovirals is estimated to have cost 350,000 lives.

Viruses and Science

The relationships between science, viruses and society are complex. Science is the study of the natural world, by which token viruses fall under its ambit. But whereas virology only became a scientific discipline in the twentieth century in the wake of scientific discoveries mentioned earlier (and reviewed by David Rowlands in this volume), viruses themselves have shaped general science for much longer, just as they have influenced society more broadly. Medical specialties such as epidemiology and public health owe much of their

development to infectious diseases we now know to be caused by viruses, as do such medical technologies such as vaccination. The discovery of viruses stimulated a synergistic development between advances in virology and the broader development of science and technology.

Impact of Science and Technology on Virology

The expansion of virology owes a lot to the invention and development of novel instrumentation and laboratory techniques. Many of these tools came from disciplines outside the bounds of biology; from biophysical and engineering sciences and from analytical chemistry. Here too there was synergy, for the viruses spurred improvements of laboratory techniques as well as instrument design.

It was the development of improved bacteriological filters by Charles Chamberland in 1884, originally intended for purifying water, that enabled the discovery of TMV and foot mouth disease virus. The refinement of filtration – more accurately – ultra-filtration, also provided the means for determining the size of virus particles. For example, the British physical chemist, William Elford, developed a series of filters with graded pore sizes to determine the dimensions of a number of different viruses.

Elford also played a pioneering role in harnessing the ultracentrifuge, invented in 1925 by Theodor Svedberg in Sweden, which allowed the separation and concentration of virus particles of different densities. Ultracentrifugation became widely used as both an analytical and preparatory tool in virology, which in turn led to innovations in the design of future generations of the instrument. In 1936, the American physical chemist, Wendell Stanley, used ultracentrifugation to concentrate and purify TMV. He successfully crystallized the purified TMV and used X-ray crystallography to discern its structure thinking it contained only protein, but Frederick Bawden, Norman Pirie and colleagues soon revealed its all-important RNA core. As Bawden later remarked, “before 1936 it was tacitly assumed that *all* viruses were incompressible spheres”. The needle-like crystals of TMV showed otherwise.

The most dramatic change was wrought by the invention and development of the electron microscope, just as the optical microscope had done for bacteria in the seventeenth century which, in Robert Hooke’s words “opened up a whole new invisible World discovered to the Understanding”. The first electron microscope was designed in 1931 by Ernst Ruska, a German electrical engineer and it was his brother, Helmut Ruska, a medical student at that time, who grasped the potential for the new technology in biomedicine. He pursued the visualization of “sub-microscopic agents of disease” and produced the first electron-microscopic images of poxviruses in 1938. Electron microscopy proved a particularly synergistic development with virology; although it stood supreme in the visualization of viruses, it also gained in return, especially in areas of specimen preparation and staining technology (Fig. 1).

While crystallization and electron-microscopy gave scientists a handle on the physical and chemical properties of viruses, they yielded little information about their behavior and biology. The biggest hurdle in this arena was posed by the inability of scientists to isolate and propagate viruses in the laboratory as they could bacteria, because as we now know, viruses are obligate parasites that need a living host cell in order to multiply. With the technique for plating lawns of bacteria in petri dishes already in place, the first viruses to be successfully cultured in the laboratory were the bacteriophages, recognized by the ability to form discrete clear plaques of lysed bacteria. The serial dilution of the “phage” allowed accurate titration and biological cloning, as did titration of TMV on leaves. The cultivation of animal viruses proved more difficult until the propagation of viruses on the chorio-allantoic membrane of embryonated bird’s eggs by Ernest Goodpasture in the United States and Macfarlane Burnet in Australia. Virus propagation in cell cultures was first exploited for polio virus by John Enders, Thomas Weller and Frederick Robbins in 1949.

To illustrate the increasingly rapid pace of technology in virology, let us consider the emergence of novel pathogenic viruses over the past 40 years. Following the appearance of AIDS in 1981, it took two years to identify the virus, HIV-1, and a year more to persuade the world by independent studies that the correct culprit had indeed been found. A further 18 months led to the roll out diagnostic tests based on serology and to the cloning and sequencing of the HIV-1 genome. We had to await the development of the polymerase chain reaction (PCR) technique to monitor viral load in patients. Anti-viral therapy to HIV-1 only became effective in 1996 with the introduction of a combination of drugs targeting more than one step in the replication cycle.

When SARS emerged in 2003, it took less than 2 months to characterize the virus as a member of the coronavirus family, whereas in January 2020, Chinese scientists obtained the full genetic sequence of SARS-CoV-2 from a COVID-19 patient within 2 days. They devised a reliable PCR test to detect infection within a week, although diagnostic tests for SARS-CoV-2 antigens and antibodies have taken a little longer, especially for large scale usage. Pseudotype viruses, based on retrovirus or rhabdovirus vectors incorporating reporter genes that bear the SARS-CoV-2 envelope “S” (spike) glycoprotein, are providing highly sensitive and specific virus neutralization assays without the need for high level containment facilities. These are being employed to study whether recovered patients have protective immunity and as a preliminary test of vaccines based on antibody responses.

DNA sequencing has become ever higher in throughput. In 1987, it was considered quite an achievement to gain the complete DNA sequence of the 150k kilobase genome of human cytomegalovirus. Compare that to today, when a cohort of 1000 COVID patients and healthy controls are each having their total host genome sequenced to identify possible genetic risk factors, and to see if more severely affected people are infected with more virulent variants of SARS-CoV-2.

Virology has benefited from advances in computational sciences, such as bioinformatics, sequence analysis, “big data” and mathematical modeling. The science of tracking and intervening in the course of epidemics using mathematical models began in earnest in 1760, when the Swiss mathematician, Daniel Bernoulli, showed how large-scale variolation against smallpox would

Table 1 Noble Prizes awarded for discoveries relating to viruses

<i>Year</i>	<i>Name</i>	<i>Discovery</i>
1946	Wendell M Stanley ^a	Crystallization of tobacco mosaic virus
1951	Max Theiler	Development of yellow fever virus vaccine
1954	John F Enders Thomas H Weller Frederick C Robbins	Growth of polio virus in various types of cell culture
1958	Joshua Lederberg	Genetic recombination of phage and bacteria
1965	François Jacob André Lwoff Jacques Monod	Genetic control of enzyme and virus synthesis
1966	Peyton Rous	Discovery of tumor-inducing viruses
1969	Max Delbrück Alfred D Hershey Salvador D Luria Renato Dulbecco	Replication and genetic structure of viruses (phage)
1975	David Baltimore Renato Dulbecco Howard M Temin	Reverse transcriptase and tumor virus-host genome interaction
1976	Baruch Blumberg D Carleton Gajdusek	Discovery of hepatitis B virus Transmission of spongiform encephalopathy
1978	Werner Arber Daniel Nathans Hamilton O Smith	DNA restriction enzyme defense against phage infection
1982	Aaron Klug ^a	Crystallographic electron microscopy of viruses
1989	J Michael Bishop Harold E Varmus	Cellular origin of retroviral oncogenes
1992	Richard J Roberts Philip A Sharp	Split genes and RNA splicing
1996	Peter C Doherty Rolf M Zinkernagel	Specificity of cell mediated immunity to viruses
2006	Andrew Z Fire Craig C Mello	RNA interference and gene silencing by double-stranded viral RNA
2008	Harald zur Hausen Françoise Barré-Sinoussi Luc Montagnier	Discovery of human papilloma viruses causing cervical cancer Identification of human immunodeficiency virus
2018	George Smith ^a Gregory Winter ^a	Phage display of peptides and antibodies
2020	Harvey J. Alter Michael Houghton Charles M Rice	Discovery of hepatitis C virus
2020	Emmanuelle Charpentier ^a Jennifer A Doudna ^a	Development of Genome Editing

^aNobel Prize in Chemistry; all others in Physiology or Medicine.

benefit not only the individual but also the population at large, which was the origin of the concept of herd immunity. Had the physicians attending his lecture to the Academie Française heeded his message, Louis XVth and many courtiers at Versailles might not have succumbed to smallpox during the epidemic in May 1774.

Impact of Viruses on Science

Due to their unique characteristics, viruses have wielded a huge influence on advances in biology and medicine, exploited as tiny, precise tools to reveal the inner workings of the cells they infect. One rough measure of the impact of viruses on science can be garnered from the number of Nobel Prizes awarded in Medicine & Physiology and in Chemistry to discoveries made about and through viruses (Table 1).

Leading the way as a scientific tool were the bacteriophages, to which the physicist-turned-biologist, Max Delbrück turned his attention as a means to study genes at the individual and molecular level. Together with Salvador Luria and Alfred Chase, Delbrück created the famed “American Phage Group” which throughout the 1940s laid the foundations of modern molecular biology and set the scene for studies in DNA replication, recombination, and protein studies. Meanwhile André Lwoff and colleagues at the Institut Pasteur in Paris studied another group of bacterial viruses known as the lysogenic or temperate phages, which can persist in latent form within the host bacterium. Lytic and lysogenic phages played a crucial role in defining viruses as

distinct from other replicating entities. Jacques Monod and François Jacob employed phages to study mechanisms of genetic regulation and established much of our basic knowledge of universal rules of replication, transcription and translation.

TMV was aptly dubbed by the virologist Heinz Fraenkel-Conrat as “almost always the first” in virology, being the first, as mentioned, to be crystallized, and a major player in the commercial development of the ultracentrifuge and electron microscope. Animal virology, too, led to important discoveries in other areas such as the inner workings of cells and of genes. Split genes with introns and RNA splicing was revealed in 1977 through studies by Phillip Sharp and by Richard Roberts of transcription of human adenovirus type 2. Enveloped viruses, such as the DNA virus, vaccinia, and RNA viruses, such as influenza virus and alphaviruses (Semliki Forest and Sindbis viruses) have been excellent tools exploited by the Finnish virologist, Ari Helenius, and others to probe endocytosis, membrane fusion, conformational changes, secretory pathways, and post-translational modification of proteins like glycosylation. Paramyxoviruses like Sendai virus, which enter cells at the cell surface rather than through receptor-mediated endocytosis, were used to effect cell-to-cell fusion by inactivating the RNA genome without affecting envelope glycoprotein function. Inactivated Sendai virus was exploited to fuse human cells with murine cells by John Watkins and Henry Harris in pioneering studies of human somatic cell genetics.

It was not until the 1950s that Renato Dulbecco and Marguerite Vogt developed plaque assays for lytic viruses on animal cell monolayers beneath an agar gel overlay. From the mid-1960s to the 1980s oncogenic viruses played the leading role in opening up the molecular biology of cancer, including the discovery of oncogenes, which promote malignancy, and of tumor suppressor genes, which keep it in check. Rous sarcoma virus (RSV), the prototype retrovirus, was discovered by Peyton Rous in 1911, but was largely ignored by cancer experts for many decades, who did not believe that the agent was a virus until, in 1958, Harry Rubin and Howard Temin developed an in vitro cell transformation assay for RSV. Thus, oncogenic viruses began to reveal their secrets, including reverse transcriptase in 1970, the integration of viral genomes into host DNA, and the identification of viral genes with oncogenic potential. Reverse transcription permitted the synthesis of complementary DNA from spliced messenger RNA and hence led to the manufacture in bacteria of eukaryotic proteins and biological pharmaceuticals.

The cellular origin of oncogenes – genes that induce tumors – was first revealed by Harold Varmus, Michael Bishop and colleagues with the transforming *src* gene of RSV, and its gene product became the first known tyrosine kinase. The host homologs of most of the 30 or more oncogenes discovered in various oncogenic animal retroviruses, eg., *abl*, *myc*, *ras*, have important roles in cancer signaling pathways. Several have become successful targets for anti-cancer therapy. Conversely, the discovery of endogenous proviruses in the germ line of chickens and mice led to the realization that a substantial proportion of vertebrate host genomes is represented by ‘fossil’ retroviruses; approximately 8% of human DNA comprises retroviral genomes, some of which have been repurposed by the host for essential functions such as placental development.

Another important milestone with oncogenic viruses was the discovery of the tumor suppressor protein (TSp53) in cells transformed by simian 40 virus (SV40). In 1979, David Lane and Lionel Crawford in London, reported that immunoprecipitation of the large T (tumor) antigen encoded by the SV40 transforming gene also brought down a 53kd host protein. Daniel Linzer and Arnold Levine reported similar data and it became clear that T antigen blocks the natural function of p53 to regulate cell division and to direct cells to an apoptotic pathway. TSp53 became the “canonical” tumor suppressor that Lane dubbed the “guardian of the genome”. Mutations of the p53 gene in human cancers have a similar effect to its sequestration by T-antigen. Many other types of oncogenic DNA viruses encode proteins that block TSp53 function, such as E6 of HPV and the latent nuclear antigen of Kaposi’s sarcoma virus. Knocking out TSp53 function triggers the “S” phase of the cell division cycle which the viruses need to synthesize their own DNA.

Harnessing of Viruses by Humans

In this section, we cite examples of how humans have harnessed or attempted to harness viruses, not only for economic and medical benefits, but also for darker ends.

Commerce and Tulip Mania

Compared to centuries-old industries of fermentation that rely on bacteria and yeasts – think of wine, beer and bread – the exploitation of viruses for commercial purposes is still in its infancy. One exception was the unwitting use in the seventeenth century of plant viruses to develop varieties of tulips with different variegations in their flowers. The tulip mania in The Netherlands led to the sale of single tulip bulbs at prices up to 10-fold higher than the average annual earnings of farmers. In February 1637, the market suddenly collapsed, ruining many businessmen who had invested in tulip “futures”. The most treasured tulips were difficult to breed consistently because the pattern of the color “breaks” in the petals was erratic. We now know that the streaked appearance of the prized tulips is caused by the Potyvirus, Tulip Breaking Virus – discovered in 1928 by Dorothy Caley.

Vaccines

The development of vaccines as a means to prevent viral diseases is one of the best known and oldest examples of humans harnessing viruses. As in the cases of fermentation and horticulture mentioned above, the practice predates actual knowledge about the existence of the effector. The English physician Edward Jenner developed the first safe vaccine against smallpox, which was based on the fact that cowpox causes a mild lesion when introduced into the skin of uninfected persons, but cross-protects them against smallpox itself. It represented a major advance to the previous use of immunization with low doses of variola virus

(the smallpox agent) itself, which had a 2% risk of causing smallpox. Live, attenuated viruses such as Sabin polio vaccine, tend to be efficacious because, while the virus propagates to a lower level than wild-type virus, it is sufficient to elicit immunity without causing disease. Other virulent viruses can elicit protective immune responses after they have been chemically inactivated, e.g., the Salk polio vaccine. Influenza virus vaccines are typically made up of immunogenic components of disassembled viral antigens. Several of the current attempts to design a vaccine against SARS-CoV-2 are targeting the “S” spike glycoprotein.

Despite vaccination having led to immense advances in public health to protect against many infectious diseases, there has been a rise in recent years of antagonism to childhood immunizations such as the combined measles, mumps and rubella vaccine. The fear of parents has been fueled by fake news of vaccine dangers and false associations of vaccines with other diseases. Sadly, this situation has resulted in unnecessary outbreaks of measles and mumps in developed countries.

Phage therapy

When d’Herelle first observed the lytic effect of bacteriophages on bacteria in 1916, he immediately envisaged their use in controlling bacterial diseases, describing his finding as the ‘true microbe of immunity.’ He believed that phages were responsible for natural resistance of populations to diseases such as dysentery and cholera in regions of the world where the diseases were endemic, maintaining an interest throughout his career. Phage therapy attracted wide attention at first, even spurring the imagination of the American novelist Sinclair Lewis who included phages in his Pulitzer Prize-winning novel, *Arrowsmith*. Phage therapy all but disappeared from the West after WWII, owing to the greater efficacy and reliability of antibiotics, but in the pre-antibiotic world in 1930s and early 1940s, various bacteriophage products were commercially available in many countries. It was promoted in the Soviet Union, and continues at the Eliava Institute of Bacteriophage, Microbiology & Virology which d’Herelle helped to found in Tbilisi, Georgia. Recently, the emergence of antibiotic-resistant bacterial pathogens has stimulated a revival of interest in phage therapy, although bacteria can become resistant to phage as well as to antibiotics.

Gene therapy and cancer therapy

Ever since the advent of recombinant DNA technology, medical researchers have been using viruses as vectors or vehicles for delivering specific genes to target host cells as a possible means of therapy for single gene inherited disorders, such as cystic fibrosis, as well as for treating cancer. Vectors such as vaccinia virus and retroviruses may also be designed to carry genes encoding antigens of different viruses or other pathogens for immunization. Recombinant viruses, such as adeno-associated virus, can express high levels of antibody as a means of delivering passive immunotherapy. Lytic viruses can be modified to replicate only in proliferating cells in order to treat cancer. All these approaches are recent developments and are likely to have greater impact on human health in the future.

Bioweapons

The use of viruses for biological warfare is poorly documented. Smallpox may have been spread deliberately by Francisco Pizzaro during the conquest of the Inca empire because its inadvertent effect on the Aztecs 11 years earlier had been noted. One record of the use of viruses in conflict occurred during the Franco-British war for control of North America. Different Native American communities sided with the French or the English. The Redcoat Colonel Henry Bouquet sent his commander, General Lord Jeffrey Amherst, a request “to inoculate the Indians” by sending smallpox-impregnated blankets to them who, under the leadership of Chief Pontiac, were besieging Fort Pitt (Pittsburgh). In a postscript to his reply on 16 July 1763, Amherst approved this request and added that Bouquet should ‘try every method that can serve to extirpate this execrable race’.

We look upon multi-ethnic USA differently 250 years later! In 2016, student activists at Amherst College demanded the deletion of Jeffrey Amherst’s name. The Trustees of Amherst College dismissed the issue of the College’s name by stating that the College is named after the town and not the general. They did, however, remove the College’s “Lord Jeff” mascot and change the name of the “Lord Jeff Inn” on campus.

In the twentieth century, most of the potential germ warfare agents were bacterial such as anthrax and tularemia, but it appears that the Nazi regime also tested highly pathogenic viruses on concentration camp victims. The Friedrich Löffler Institute on Riems island in the Baltic Sea, founded in 1910 by the discoverer of foot and mouth disease virus, became a testing ground for bioweapons during WWII, but is now devoted to vaccine research and development. In the Soviet Union, there was a bioweapons establishment near Novosibirsk in Siberia. Vozrozhdeniya Island, in what used to be the Aral Sea, was a testing ground for bioweapons until the collapse of the Soviet Union and it remains contaminated.

Following the eradication of smallpox as a naturally occurring infection, there was debate whether to destroy the remaining stocks of the virus, officially stored in Russia (at Novosibirsk) and in the USA (at the Centers for Disease Control and Prevention – CDC – in Atlanta). In 1996, on the 200th anniversary of Jenner’s first vaccination of young James Phipps, the World Health Organization voted to destroy all official laboratory stocks of the smallpox virus, but this deliberate extinction of a species has yet to be carried out. The issue remains controversial among virologists, some of whom say that preservation is advisable for future work on vaccines or on any novel pox viruses that might emerge, while others argue that, with full genome sequence data of variola in the open domain, there is no need to preserve the virus itself.

It is not clear whether any terrorist organizations or ‘rogue’ nations possess dangerous viruses such as smallpox or Crimean-Congo Hemorrhagic Fever virus (a frequently fatal orthonairovirus), which is naturally transmitted by animal ticks but can also spread among humans via bodily fluids and aerosols. Unofficial stocks of smallpox virus might still exist, either hidden by those

with malicious intent, or which have simply been overlooked. An example of the latter was the discovery in 2016 of vials of variola virus in a freezer at the US National Institutes of Health in Bethesda, Maryland, which were subsequently shipped to the CDC.

Impact of Viruses on Society

It might be appropriate to call this section “the impact of viruses on *other* aspects of society”, because science is, after all, very much a human activity and thus part of society. Here, we address the impact of viruses beyond the laboratory. Undoubtedly the most recognizable impact of viruses is their ability to cause epidemic disease. While our primary focus here will be on human disease, we also touch on viral pathogens of livestock and crops, before ending with a brief look at the influence of viruses on other spheres of human activity, notably the arts and literature.

Impact on Human Disease

We know little of the precise times and origins of the plagues of old, but they seem to have appeared when dense populations developed on irrigated land. This seems evident from the cultural myths of those regions: the Nile delta, the Fertile Crescent, the Indus Valley and the area between the Yellow and Yangtze rivers. Yet based on both historical records and archeological evidence one can say that the humans have been subject to viral diseases for as long as we have existed.

The success of vaccines and antibiotics in the mid-twentieth century led to hopes that infectious diseases in general may become largely a phenomenon of the past, like smallpox. In their fine book *Natural History of Infectious Disease* (1972), McFarlane Burnet and David White commented: “The most likely forecast about the future of infectious disease is that it will be very dull. There may be some wholly unexpected emergence of a new and dangerous infectious disease, but nothing of the sort has marked the past fifty years”. A half century later, however, we can see that infections have been anything but dull. Ebola broke out in the Congo in 1976, AIDS began to take lives in the early 1980s and at the time of writing we are in the throes of the COVID-19 crisis which began in December 2019.

The origins and historical demographics of viral infections in humans

Some human viruses have always been with us whereas others have transferred from animal hosts. From the “germ’s eye view” it appears that it was the massive changes to human society in the wake of the prehistoric agricultural revolution some 10,000 years ago that provided the opportunity for so many new types of viruses to take up residence in humans. The successful colonization by viruses of foreign origin was a consequence, first of human dispersion out of Africa, second of domestication of livestock and most importantly, the development of large, densely populated communities.

There are many zoonotic viruses, e.g. rabies and H5N1 influenza, that have never become epidemic in humans. An estimated 60% of the 1400 species of infectious microbes known to be pathogenic in humans are transmitted by animals, for which the human usually represents a dead-end host. Viruses that do adapt to human-to-human spread, however, can become self-sustaining providing there is a human population of 250,000 or greater of naïve hosts to infect. Measles, according to recent phylogenetic molecular clock studies, is believed to have diverged from rinderpest of cattle approximately 2400 years ago, and the smallpox strain that was present until eradication, most likely arose in the early Christian era from the Bactrian camel, though burrowing rodents were the ultimate source. These viruses have diverged from their progenitors and only circulate in the human population.

Influenza viruses emerge and re-emerge from reservoirs in waterfowl. Having segmented genomes, they can re-assort with genes of existing strains circulating in humans and other mammals to form new epidemic variants. The “Spanish Flu” pandemic of 1918/1919 was an H1N1 strain that may have arisen in horses of the trenches in World War I. Many modern sources of zoonoses indicate a shift from an origin in livestock or in “companion animals” like dogs and rats, to viruses from exotic species. Deforestation, a taste for bushmeat and other environmental or behavioral factors have increased the opportunities for viruses to reach humans from bats, like Ebola, SARS and Nipah, often via a short-term intermediate host, and from simians, like Zika and other mosquito-borne viruses.

For populations in which epidemic viruses are already resident and widespread, most surviving adults have acquired what is called herd immunity and new infections typically occur in children. However, when the virus reaches a host population that has not previously had experience of that virus, it can infect children and adults alike. The globalization of maritime travel following the explorations of Christopher Columbus and Vasco da Gama at the end of the fifteenth century allowed smallpox, measles and other “Old World” viruses reach new human populations.

The impact on the societies previously unexposed to these viruses proved to be devastating. No other viral disease has affected as many people throughout history as smallpox. Its accumulated mortality was huge, and it has repeatedly ravaged different populations. The introduction of smallpox to the New World during the 1521 conquest of the Aztec empire by Hernán Cortes and his band of Conquistadors, is recounted in vivid detail by Bernal Diaz in his diary, *The Conquest of New Spain*. In the ensuing 100 years, the populous regions of North America from Guatemala to the Mississippi basin lost around 90% of its indigenous peoples from smallpox and measles. This population decline was significantly greater than that in Europe following the Black Death. While it took 200 years for Western Europe to regain its pre-1348 population level, the lot of the survivors improved. For peasants, it became a bull labor market and they fought for and won freedom from serfdom (hereditary indentured labor). The decimation of indigenous peoples in America by smallpox had a greater and more sinister consequence: the need for manpower in the New World plantations engendered the trans-Atlantic slave trade.

This pattern of devastation by viruses was repeated globally as further new worlds were contacted in South America, Australia and Oceania. It continues today when isolated Amazonian tribes are contacted by “civilization”. As observed by Charles Darwin in his *Notes of the Voyage of the Beagle* in 1838: ‘Wherever the European has trod, death seems to pursue the aboriginal ... Most of the diseases have been introduced by ships and what renders this fact remarkable is that there might be no appearance of the disease among the crew which conveyed this destructive importation.’

Viral epidemics in the past century

Perhaps our best weapon against viral infections in the past century and a half has been vaccination; it has effected the eradication of smallpox, and the control of many other viral diseases including rabies, influenza, yellow fever, polio myelitis, measles, mumps and rubella, and most recently the human papilloma viruses associated with cervical cancer.

But we haven't conquered all viruses by any means. Despite the development of vaccines against it, influenza virus continues to wreak havoc, killing an estimated extra 280,000 people as recently as the H1N1 pandemic in 2009. Since it first appeared four decades ago, HIV has accounted for some 39 million AIDS deaths, 66% of them in Africa. In addition, AIDS has had an enormous socio-political impact.

In the early phases of the epidemic in North America, HIV targeted marginalized social groups, notably male homosexuals and intravenous drug-users. On the one hand, this apparent predilection exacerbated the discrimination against homosexuals by fundamentalist religious groups and by political conservatives who opposed certain types of demographic surveys about sexual behavior, which was detrimental for learning about disease epidemiology. On the other hand, the AIDS epidemic mobilized the gay community, especially in the US, to social activism, resulting in widespread appeals and funding for research and public awareness about the disease. Meanwhile, the demographics of AIDS incidence in low income nations, especially Africa, also mobilized efforts to roll out anti-retroviral therapy across Africa, and by public health communities for better sex education and safer sexual practices.

Impact on human capital in the era of COVID-19

Plunged in the COVID-19 pandemic as we are, it is not yet possible to accurately estimate its mortality rate. SARS-CoV-2 has a high mortality rate in hospitalized cases of infection, whereas the mortality in the general population is likely to be much lower, once the prevalence of infection has been ascertained by serological studies. The SARS and MERS coronaviruses are more highly pathogenic than SARS-CoV-2, but fortunately could be contained. That makes them easier to hunt down and box in via contact tracing because, like smallpox virus, there is little onward transmission before symptoms appear.

Four other strains of human coronavirus are highly transmissible and represent around 20% of ‘common colds’. One can speculate whether they caused similar mortality to SARS-CoV-2 when first introduced to humankind. With COVID-19, elderly people are at most risk of severe disease and death, which contrasts with the 1918/19 ‘flu pandemic, when many old people got off “lightly”. The reason why the elderly were relatively protected in 1918/19 was possibly due to an immune memory of a previous outbreak of a related ‘flu virus sometime in the late nineteenth century.

A better social indicator of disease impact than mortality is the disability-adjusted life year (DALY) expressed as the number of years lost due to ill-health, disability or early death. Given that COVID-19 tends to be most severe in older people, if we were to measure DALYs, the cost to society of COVID-19 affecting human capital, the workforce and those still in formative years of education ought to be relatively slight compared to the Spanish ‘flu and HIV. These two 20th century pandemics took out mainly young adults, the very cohorts in whom society had invested maximum social capital.

SARS-CoV-19 infection, nonetheless, appears set to exceed the mortality of any other novel virus since HIV (~39 million over a “flattened” 40 years), and the Spanish ‘flu a century ago (over 50 million deaths within a mere 18 months). The main driver of the profound effects that we are currently witnessing on the global economy and our livelihoods lies in our attempts to ameliorate the spread COVID-19, for which most nations were ill-prepared. It appears that complex, inter-connected modern societies and economies are more fragile in the face of a novel pandemic than 100 year ago.

While we may marvel at the speed of the identification of the new coronavirus and the use of rapid early diagnostic tests by genome amplification, we are humbled by the fact that we still rely on centuries old means of trying to contain the pandemic. We are using quarantine methods (self-isolation) dating from Venice’s approach to the Black Death 670 years ago, and the social distancing imposed on lepers. The advice “wash your hands” is a legacy from Ignaz Semmelweis whose medical colleagues ignored him as they made their way from the morgue to the delivery room. We hope that a COVID-19 vaccine will soon be developed given the number of candidates under preclinical and clinical trial, but an efficacious and long-lasting one may take a little longer to achieve.

Impact of Viruses on Livestock and Crops

Textbooks of veterinary virology and plant virology are as weighty as those about human viruses, owing to the great variety of viruses that threaten animal and plant health. Our 120-year knowledge that foot and mouth disease is caused by a virus has not entirely prevented it, as evidenced by the 2001 outbreak in the UK that resulted in serious economic loss to farmers through obligatory culling of livestock. The epidemic was probably triggered by the illegal addition of kitchen food waste containing imported ham to pig swill, and it then spread rapidly by transport of cattle and sheep traded in markets. Probably the best recognized viral pathogens of animals are strains of influenza, with reservoirs in wild waterfowl but initiating epidemics in domestic birds, horses and pigs. Tick-borne African Swine Fever Virus, endemic in wart hogs and bush pigs, has spread in recent

decades far beyond East Africa and is currently a serious problem in China. The orbivirus causing bluetongue disease in ruminants is broadening its geographic range northwards from Africa to Europe with the spread of its midge vector due to global warming, as well as colonization of new midge species. Tilapia lake virus first emerged in 2014 and is affecting Tilapia pisciculture in South America, Africa and south-east Asia where the fish is an important source of protein in the human diet.

The greatest success story in veterinary virology was the eradication of rinderpest, a disease of cattle caused by a morbillivirus. Regular outbreaks of rinderpest previously had a profound impact on herding communities in East Africa. This disease, together with the trypanosome parasite causing sleeping sickness, rendered large swathes of Africa inhospitable for imported domestic cattle. After a \$5 billion campaign that began in 1995, the United Nations Food and Agricultural Organization declared the global disappearance of rinderpest in June 2011, the second virus after smallpox to be eradicated. In June 2019, the Pirbright Institute in the UK destroyed its repository of different rinderpest virus strains, having documented their genetic sequences.

The best studied viral pathogen of crops is undoubtedly TMV which played such an important role in virology discussed earlier. With hindsight, we know that tobacco itself is a scourge upon humankind, but through its introduction to the Old World in the sixteenth century Colombian Exchange, it became an important commodity. Cauliflower mosaic virus (no relation to TMV) is notable both as a pathogen of cruciferous plants and because it replicates via reverse transcriptase like retroviruses. One of the most important plant viruses that is currently affecting food security is African cassava mosaic virus, a member of the Geminivirus family.

Viruses in Art and Literature

As with almost anything that we humans have encountered, viruses have stirred our imaginations in different ways and have made their way into our creative pursuits. The functional beauty of virus particles has caught the attention of artists, designers and architects. Conversely, buildings have inspired virologists to deduce virus structures more precisely. In 1962, Donald Caspar and Aaron Klug proposed a model of icosahedral (20-sided) particles drawn from the architect-designer Buckminster Fuller's geodesic domes who himself became enchanted with this analogy. He had solved a means to erect stable 'spherical' buildings composed of triangulated, pentagonal and hexagonal sided domes and it became apparent that viruses had solved the problem before humans did. Viruses with small genomes require many copies of one or two proteins to form the capsid surrounding the RNA or DNA genome. Bacteriophage MS2 has this type of structure and so do human adenoviruses. The recombinant human papilloma virus vaccine is self-assembled from one viral protein, L1, to form empty (genome free) "virus like particles" that possess this structural integrity. Moreover, the particulate nature of the vaccine renders it highly immunogenic.

Epidemic disease in particular has been the inspiration and subject of many works of literature. One of the most enduring examples is in Thucydides's *History of the Peloponnesian War*, where he provided a first-hand account of the 430 BCE "Plague of Athens", which may have been a typhus or viral outbreak.

No other disease in the twentieth century has generated as much as the AIDS pandemic in so many different spheres of the literary, performing and visual arts: The plays *A Normal Heart* by Larry Kramer (1985) and *Angels in America* by Tony Kushner (1992), were both explicit in their criticism of the social and medical response to the plight of the gay community in the early years of the AIDS outbreak; Randy Shilts's *And the Band Played On* (1987) provided a chronicle of the new disease and the unfolding of scientific breakthroughs; Thom Gunn produced an anthology of poetry *The Man With the Night Sweats* (1992) while AIDS ravaged San Francisco; and Abraham Verghese provided a doctor's perspective of treating the disease in the deeply conservative rural Tennessee in his moving *My Own Country* (1994). AIDS also gave a new guise to Giacomo Puccini's *La Boheme* – itself based on tuberculosis – when Jonathan Larson's award-winning rock opera *Rent* hit Broadway in the mid-1990s. Last but not least, there is one of the largest and possibly longest-running piece of community folk art in the world: the NAMES Project Memorial Quilt, better known as the AIDS Quilt. Conceived in 1985 by the activist Cleve Jones to celebrate the lives of those lost due to AIDS, it is currently estimated to weigh about 54 tons, and from 2020 will be on permanent display in San Francisco where it originated.

We end with an allusion to two modern classics: Chinua Achebe's novel *Things fall apart* (1958) which describes the strain and collapse when a traditional West African society becomes 'infected' by a foreign culture of missionaries and colonial officers. Although not itself about viral epidemics, the title alludes to the famous lines in WB Yeats' poem *The Second Coming*, written in 1919 at the height of the Spanish 'flu pandemic:

Things fall apart; the center cannot hold;

Mere anarchy is loosed upon the world,

The blood-dimmed tide is loosed, and everywhere

The ceremony of innocence is drowned.

Further Reading

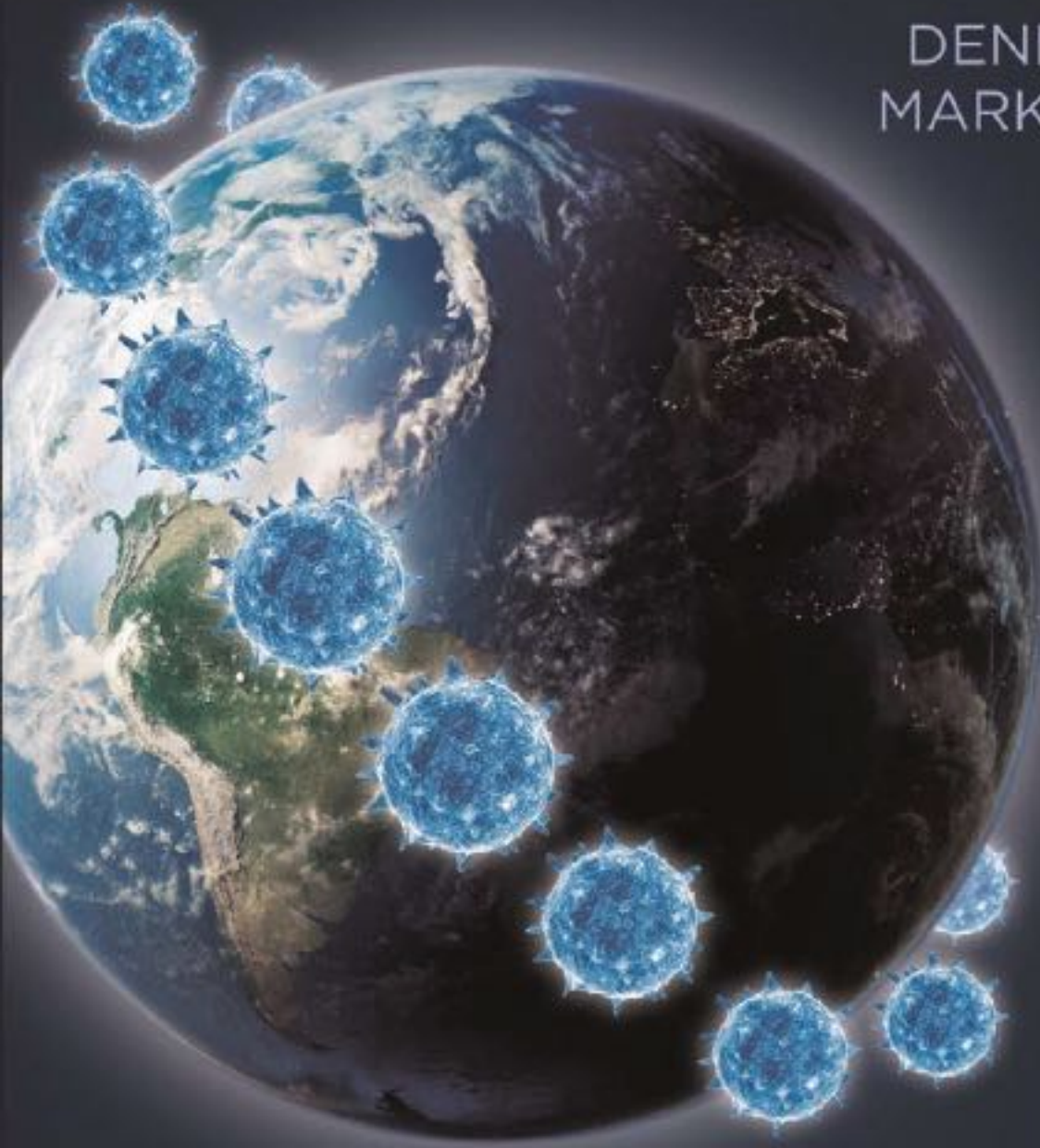
- Crawford, D.H., 2007. *Deadly Companions: How Microbes Shaped our History*. Oxford: Oxford University Press.
- Creager, A., 2002. *The Life of a Virus: Tobacco Mosaic Virus as an Experimental Model, 1930–1965*. Chicago: Univ. of Chicago Press.
- Draulans, D., 2020. 'Finally, a virus got me.' Scientist who fought Ebola and HIV reflects on facing death from COVID-19. *Science*. doi:10.1126/science.abc7042.
- Diamond, J., 1998. *Guns, Germs and Steel*. London: Vintage.
- Garret, L., 1994. *The Coming Plague*. New York: Farrar, Straus & Giroux.
- Helenius, A., 2018. Virus entry: Looking back and moving forward. *Journal of Molecular Biology* 430, 1853–1862.
- Lane, D.P., Crawford, L.V., 1979. T antigen is bound to a host protein in SV40-transformed cells. *Nature* 278, 261–263.
- Lwoff, A., 1957. The concept of virus. *Journal of General Microbiology* 17, 239–253.
- McLean, A.R., May, R.M., Pattison, J., Weiss, R.A. (Eds.), 2005. *SARS: A Case Study in Emerging Infections*. Oxford University Press.
- Sankaran, N., 2021. *A Tale of Two Viruses: Parallels in the Research Trajectories of Tumor and Bacterial Viruses*. Pittsburgh: University of Pittsburgh Press, Pittsburgh.
Available at: <https://upittpress.org/books/9780822946304/>.
- Sharp, P.A., 2005. The discovery of split genes and RNA splicing. *Trends in Biochemical Sciences* 30, 279–281.
- Spinney, L., 2017. *Pale Rider: The Spanish Flu of 1918 and How It Changed the World*. London: Jonathan Cape.
- Weiss, R.A., McMichael, A.J., 2004. Social and environmental risk factors in the emergence of infectious diseases. *Nature Medicine* 10, S70–S76.

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Editors in Chief

DENNIS BAMFORD
MARK ZUCKERMAN



ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Volume 2

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

EDITORS IN CHIEF

Dennis H. Bamford

*Molecular and Integrative Biosciences Research Programme
Faculty of Biological and Environmental Sciences
University of Helsinki, Helsinki, Finland*

Mark Zuckerman

*South London Specialist Virology Centre
King's College Hospital NHS Foundation Trust
London, United Kingdom*

and

*Department of Infectious Diseases
School of Immunology and Microbial Sciences, King's College London Medical School
London, United Kingdom*

Volume 2



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2021 Elsevier Ltd. unless otherwise stated. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-12-814515-9

For information on all publications visit our
website at <http://store.elsevier.com>

		Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org		

Publisher: Oliver Walter

Acquisitions Editor: Priscilla Braglia

Content Project Manager: Katarzyna Miklaszewska

Associate Content Project Manager: Gayathri S.

Designer: Matthew Limbert

EDITORS IN CHIEF



Dennis H. Bamford, PhD, is Professor Emeritus of Virology at the Faculty of Biological and Environmental Sciences, University of Helsinki, Finland. He obtained his PhD in 1980 from the Department of Genetics, University of Helsinki. During 1981–1982 he was an EMBO postdoctoral fellow at the Public Health Research Institute of the City of New York, United States, and during 1983–1992 he worked as a Senior Scientist at the Academy of Finland. In 1993 he was appointed Professor of General Microbiology at the University of Helsinki. He was awarded the esteemed Academy Professorship twice, in 2002–2007 and 2012–2016, and he also served twice as the Director of the Finnish Center of Excellence (in Structural Virology, 2000–2005, and in Virus Research, 2006–2011). Prof. Bamford has had continuous external research funding (e.g., from several European Union, Academy of Finland, TEKES and Jusélius Foundation funds, as well as the Human Frontier Science Program). He is an EMBO member and has held several positions of trust in scientific and administrative organizations.

Prof. Bamford has published approx. 400 articles in international peer-reviewed journals in virology, microbiology, biochemistry, and molecular biology (36 of them in high impact journals). About half of the primary articles have been published with international collaborators showing high international integration. He has also been invited to give 56 keynote and plenary presentations in major international meetings. Prof. Bamford has supervised over 35 Master's and

over 40 PhD theses. Seven of his graduate students or post docs have obtained a professorship and a similar number have a principal investigator status. Prof. Bamford has studied virus evolution from a structure-centered perspective, showing that seemingly unrelated viruses, such as bacteriophage PRD1 and human adenovirus have similar virion architecture. When the corona virion architecture was gradually revealed, it was observed that its structural elements were close to those seen in RNA bacteriophage phi6 so that phi6 has been actively used as surrogate for pathogenic viruses - quite a surprise!



Dr. Mark Zuckerman is Head of Virology, Consultant Medical Virologist, and Honorary Senior Lecturer at South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust and King's College London Medical School, Department of Infectious Diseases, School of Immunology and Microbial Sciences in London, United Kingdom. His interests include the clinical interface between developing molecular diagnostic tests relevant to the local population of patients, respiratory virus infections, herpesvirus infections in immunocompromised patients and blood-borne virus transmission incidents in the healthcare setting. He has chaired the UK Clinical Virology Network, Royal College of Pathologists Virology Specialty Advisory Committee and Virology Examiners Panel and is a member of the Specialty Advisory Committee on Transfusion Transmitted Viruses. He is a co-author on four editions of the "Mims' Medical Microbiology" textbook, has written chapters in a number of other textbooks and has over 100 publications in international peer-reviewed journals and is an associate editor for two journals.

EDITORIAL BOARD

Editors in Chief

Dennis H. Bamford

Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Mark Zuckerman

South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust, London, United Kingdom and Department of Infectious Diseases, School of Immunology and Microbial Sciences, King's College London Medical School, London, United Kingdom

Section Editors

Claude M. Fauquet

St Louis, MO, United States

Michael Feiss

Department of Microbiology and Immunology, Carver College of Medicine, University of Iowa, Iowa City, IA, United States

Elizabeth E. Fry

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Said A. Ghabrial[†]

Department of Plant Pathology, University of Kentucky, Lexington, KY, United States

Eric Hunter

Department of Pathology and Laboratory Medicine, Emory University School of Medicine and Emory Vaccine Center, Emory University, Atlanta, GA, United States

Ilkka Julkunen

Institute of Biomedicine, University of Turku, Turku, Finland

Peter J. Krell

Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Mart Krupovic

Archaeal Virology Unit, Institut Pasteur, Paris, France

Maija Lappalainen

HUS Diagnostic Center, HUSLAB, Clinical Microbiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Hubert G.M. Niesters

Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Massimo Palmarini

MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

David Prangishvili

Institut Pasteur, Paris, France and Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

David I. Stuart

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom and Diamond Light Source, Didcot, United Kingdom

Nobuhiro Suzuki

Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

[†]Deceased.

SECTION EDITORS



Claude Fauquet received his PhD in biochemistry from University Louis Pasteur in Strasbourg, France in 1974. Dr. Fauquet joined the Institut de Recherche pour le Développement (IRD) and worked there as a plant virologist for 28 years, and served in Ivory Coast, West Africa for 14 years. In 1991, he founded the International Laboratory for Tropical Agricultural Biotechnology (ILTAB) at The Scripps Research Institute, CA, United States. ILTAB was then hosted by the Donald Danforth Plant Science Center, St. Louis, MO, from 1999 to 2012. In 2003, he co-founded the Global Cassava Partnership for the 21st Century (GCP21), which he directed until 2019 and which goal is to improve the cassava crop worldwide.

Dr. Fauquet is an international leader in plant virology including taxonomy, epidemiology, molecular virology, and in gene-silencing as an antiviral strategy. He was Secretary of the International Committee on Taxonomy of Viruses (ICTV) for 18 years and the editor of several ICTV Reports including the VIIIth ICTV Report in 2005.

He has published more than 300 research papers in reviewed journals and books. He is a fellow of the American Association for the Advancement of Science, of the American Phytopathological Society and a member of the St. Louis Academy of Sciences. In 2007, Dr. Fauquet was knighted “Chevalier de l’Ordre des Palmes Académiques” by the French Minister of High Education and Research.



Dr. Michael Feiss is Professor Emeritus in the Department of Microbiology and Immunology of the Carver College of Medicine at the University of Iowa, IA, United States. Dr. Feiss received his PhD in Genetics at the University of Washington followed by a postdoctoral traineeship in the laboratory of Dr. Allan Campbell at Stanford. Dr. Feiss is a microbial geneticist who studies virus assembly with an emphasis on how a DNA virus, bacteriophage lambda, packages viral DNA into the empty prohead shell. The lab investigates how sites in the viral DNA orchestrate the initiation and termination of the DNA packaging process. This work includes comprehensive examination of the DNA recognition sites. A related interest is study of terminase, the viral DNA packaging enzyme, including the functional domains for protein–DNA and protein–protein interactions. A second focus has been the roles of the bacterial host’s IHF and DnaJ proteins in the lytic life cycle of the virus. More recent work has involved a genetic dissection of the role of terminase’s ATPase center that powers translocation of viral DNA into the prohead. This interest in the ATP hydrolysis-driven packaging motor involves a multidisciplinary collaboration examining the kinetics of DNA packaging during individual packaging events. Finally, recent studies have also looked at how the packaging process has diverged among several lambda-like phages, including phages 21, N15, and Gifsy-1.



Elizabeth E. Fry is a senior postdoctoral scientist in structural biology at the University of Oxford, Oxford, United Kingdom, where she received her DPhil. for studies relating to the structure determination of Foot-and-Mouth Disease Virus. Specializing in structural virology, Dr. Fry has studied many virus/viral protein structures but her primary focus is on picornavirus structure and function, in particular receptor interactions and virus uncoating. She is particularly interested in rationally designing virus-like-particles as next generation vaccines to reduce the inherent risks in handling live viruses.



Said A. Ghabrial[†] received his BSc in 1959 from Cairo University, Cairo, Egypt, and his PhD from Louisiana State University, Baton Rouge, LA, United States, in 1965. Dr. Ghabrial did postdoctoral research at the University of California, Davis, CA, United States, before returning to Cairo, where he served as a plant virologist in the Ministry of Agriculture. He returned to the United States in 1970 to do postdoctoral research at Purdue University, West Lafayette, IN. In 1972, he joined the Plant Pathology Department at the University of Kentucky, Lexington, KY, United States, where he rose to the rank of professor in 1986 and worked until 2013.

Dr. Ghabrial has served as an associate and senior editor of *Phytopathology*. He served on the editorial boards of the *Encyclopedia of Virology*, 3rd edition and *Encyclopedia of Plant Pathology*, and edited a thematic issue of *Advances in Virus Research* on "Mycoviruses". He was a member of the American Phytopathological Society (APS) and the American Society for Virology (ASV); in July 2002 he was elected as a Fellow of the American Phytopathological Society. He also acted as Chair of the ICTV Subcommittee on Fungal Viruses in 1987–1993 and 2011–2014.

His long professional career allowed him to make many scientific achievements in phytopathology and virology. Among them are molecular dissection of a legume-infecting RNA virus, bean pod mottle virus (BPMV), development of BPMV-based vectors, discovery of a transmissible debilitation disease of the phytopathogenic ascomycete, *Helminthosporium victoriae* (*Cochliobolus victoriae*), establishment of a viral etiology of the *H. victoriae* disease, and advancement of structural biology of diverse fungal viruses.



Eric Hunter, PhD, is Professor of Pathology and Laboratory Medicine at Emory University, Atlanta, GA, United States. He serves as Co-Director of the Emory Center for AIDS Research and is a Georgia Research Alliance Eminent Scholar.

Dr. Hunter's research focus has been the molecular virology and pathogenesis of retroviruses, including human immunodeficiency virus. He has made significant contributions to the understanding of the role of retroviral glycoprotein structural features during viral entry and providing unique insights into the assembly and replication of this virus family. In recent years the emphasis of his research has been on HIV transmission and pathogenesis, defining the extreme genetic bottleneck and selection of viruses with unique traits during HIV heterosexual transmission. He has described the selection of fitter viruses at the target mucosa, a gender difference in the extent of selection bias, and a role for genital inflammation in reducing selection. His research has defined the impact of HIV adaptation to the cellular immune response on immune recognition and control of HIV after

transmission, as well as on virus replicative fitness in vitro and in vivo. Recent work highlights the roles that virus replicative fitness and sex of the host play in defining disease progression in a newly infected individual. His bibliography includes over 300 peer-reviewed articles, reviews, and book chapters. He has also been the recipient of four NIH merit awards for his work on retrovirus and HIV molecular biology.

Dr. Hunter served as the Editor in Chief of the journal *AIDS Research and Human Retroviruses* for 10 years. He was Chair of the AIDS Vaccine Research Subcommittee which is charged with providing advice and consultation on AIDS vaccine research to the National Institute of Allergy and Infectious Diseases and continues to serve on editorial boards for several academic journals and on external advisory committees for several government, academic, and commercial institutions.



Ilkka Julkunen graduated as an MD/PhD in 1984 from the Department of Virology, University of Helsinki, Helsinki, Finland. He worked as a postdoctoral research fellow at Memorial Sloan-Kettering Cancer Center in New York, United States, in 1986–1989, followed by positions as a senior scientist, group leader and research professor at Finnish Institute for Health and Welfare in 1989–2013. In 2013 he became a Professor of Virology at the University of Turku, Turku, Finland. The research interests of Dr. Julkunen have concentrated on innate and adaptive humoral immunity in viral and microbial infections. He has studied intracellular signaling and RIG-I and TLR-mediated activation of interferon system in human macrophages and dendritic cells and stable cell lines in response to human and avian influenza, Sendai, Zika and coronavirus infections. In addition, he has analyzed the downregulation of innate immunity by viral regulatory proteins from influenza, HCV, flavi-, filo- and coronaviruses. He has expertise in vaccinology, biotechnology and development of methods to analyze antiviral immunity, he has also been actively involved in research training and collaborations with biotechnological industry.

[†]Deceased.



Peter Krell started his career in virology early as a summer high school student working for the Canadian Forestry Service studying the resistance of nuclear polyhedrosis viruses (now called baculoviruses) to environmental exposure with Dr. Fred T. Bird at the Insect Pathology Research Institute in Sault Ste. Marie, ON, Canada. He received his BSc and MSc in biology from Carleton University studying the iridovirus *Tipula Iridescent Virus* with Dr Peter Lee, in Ottawa, the Canadian capital. For his PhD he headed east to Dalhousie University in Halifax, Nova Scotia on the Atlantic coast. In addition to enjoying the salt sea air, fresh cod, lobster and mussels, he studied the molecular biology of polydnaviruses under the guidance of Dr Don Stoltz. Heading south to Texas A&M University in College Station, TX, United States, as a Postdoctoral Fellow he worked with Dr. Max Summer, of baculovirus fame, and Dr. Brad Vinson continuing to study polydnaviruses, but also became steeped in the early days of molecular baculovirology. He then accepted a faculty position in the Department of Microbiology and Immunology at the University of Guelph in Guelph, ON, Canada. There he switched to baculovirus research, which was more tractable, due in part to available cell cultures and focused on viral DNA replication and functional genomics,

particularly on chitinase, cathepsin and ME53. In collaboration with Dr. Eva Nagy he studied molecular biology of different animal viruses, notably Fowl Avian adenoviruses and their development as vaccine vectors, but also on the birnavirus infectious pancreatic necrosis virus, the coronavirus porcine endemic diarrhea virus, fowlpox virus and the paramyxovirus Newcastle disease virus. He has been involved extensively with virus taxonomy, being active in the International Committee on Taxonomy of Viruses (ICTV) as member of the *Polydnaviridae* and *Baculoviridae* study groups, national representative of Canada on the ICTV, member of the Executive Committee for the ICTV and Chair of the ICTV Invertebrate Virus Subcommittee. In terms of governance, Peter Krell was President of the Canadian Society of Microbiology, Secretary and later President of the Society for Invertebrate Pathology, as well as being on the Editorial Boards of the *Canadian Journal of Microbiology* and the *ASM Journal of Virology*. While at the University of Guelph, he rose through the ranks to Professor and is currently University Professor Emeritus.



Mart Krupovic is the Head of the Archaeal Virology Unit in the Department of Microbiology at the Institut Pasteur of Paris, France. He received his MSc in Biochemistry in 2005 from the Vilnius University, Vilnius, Lithuania and PhD in 2010 in general microbiology from the University of Helsinki, Helsinki, Finland. His current research focuses on the diversity, origin, and evolution of viruses, as well as molecular mechanisms of virus–host interactions in archaea. He has published over 170 journal articles and serves as an editor or on the editorial boards of *Biology Direct*, *Research in Microbiology*, *Scientific Reports*, *Virology*, and *Virus Evolution*. He is also a member of the Executive Committee of the International Committee on Taxonomy of Viruses (ICTV) and chairs the Archaeal Viruses Subcommittee of the ICTV.



Maija Lappalainen, MD, PhD, Associate Professor of Clinical Microbiology, is the Head of Clinical Microbiology in the HUS Diagnostic Center, HUSLAB, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. In her thesis during the years 1987–1992 she studied the incidence and diagnostics of congenital toxoplasmosis. After PhD, her research interest has been in diagnostic clinical virology, viral hepatitis, respiratory infections, viral infections in the immunocompromised patients and viral infections during pregnancy.



Hubert G.M. Niesters (1958) studied biology and chemistry in Nijmegen, the Netherlands. After obtaining his PhD in Utrecht (Prof. dr. M. Horzinek and Prof. dr. B. van der Zeijst, 1987) on the molecular epidemiology of infectious bronchitis virus, he worked as a post-doctoral fellow with Prof. dr. Jim Strauss at the California Institute of Technology (Pasadena, United States) on the replication of Alphaviruses. He received a Niels Stensen fellowship (The Netherlands) and an E.S. Gosney fellowship (Caltech) during this period.

After returning to the Netherlands (1989), he became a research associate in medical microbiology at the Diagnostic Medical Center (Delft) but moved back to clinical virology as a senior research associate in 1991 at the Erasmus University Medical Center Rotterdam (Head Prof. dr. Ab Osterhaus). From 1993 to 2007, he was responsible for the molecular diagnostics unit. During this period, he was involved in the discovery and characterization of several new viruses and variants. In 2007, he became full professor and director of the Laboratory of Clinical Virology within the Department of Medical Microbiology at the University Medical Center Groningen and University of Groningen.

He has been actively involved in the implementation and development of new technologies like real-time amplification and automation within clinical virology. He has been focusing on molecular diagnostics and its use and the clinical value in a transplant setting, as well as in monitoring treatment of hepatitis viruses. Recently, his interest focuses on rapid regional epidemiology, automation including MiddleWare solutions for molecular diagnostics, as well as the cost-benefit of rapid point-of-impact molecular testing. Special interest is focused on raising awareness for the detection of enteroviruses (enterovirus D68) and its relationship with acute flaccid myelitis (AFM).

Since 2017, he is the Chair of the executive board of QCMD (Quality Control of Molecular Diagnostics, Glasgow). He is an auditor and team leader for the Dutch Council of Accreditation and Co-Editor in Chief of the *Journal of Clinical Virology*. He is an (co)-author of more than 250 peer-reviewed papers, chapters and reviews including emerging viruses, such as enterovirus D68 and hepatitis E virus (H-index 80).

For his entire work, he received in 2016 the “Ed Nowakowski Senior Memorial Clinical Virology Award” from the Pan American Society for Clinical Virology.



Massimo Palmarini is the Director of the MRC-University of Glasgow Centre for Virus Research and Chair of Virology at the University of Glasgow, Glasgow, United Kingdom. A veterinarian by training, his research programs focus on the biology, evolution and pathogenesis of arboviruses and the mechanisms of virus cross-species transmission. His work is funded by the MRC and the Wellcome Trust. Massimo Palmarini has been elected Fellow of the Academy of Medical Sciences, of the Royal Society of Edinburgh and of the Royal Society of Biology and he was a Wolfson-Royal Society Research Merit Awardee. He is a Wellcome Trust Investigator.



David Prangishvili, PhD, Honorary Professor at the Institut Pasteur, Paris, France, and Professor at Tbilisi State University, Tbilisi, Georgia, is one of the pioneers in studies on the biology of Archaea and their viruses. His scientific career spans ex-USSR (Institute of Molecular Biology, Moscow; 1970–1976), Georgia (Georgian National Academy of Sciences, Tbilisi; 1976–1991), Germany (Max-Planck Institute for Biochemistry, Munich; University of Regensburg; 1991–2004) and France (Institut Pasteur, Paris, 2004–2020). In the research groups headed by him, several dozens of new species and eight new families of archaeal viruses have been discovered and characterized, which display remarkable diversity of unique morphotypes and exceptional genome contents. The results of his research contribute to the knowledge on viral diversity on our planet and change the field of prokaryotic virology, leading to the notion that viruses of

hyperthermophilic Archaea form a particular group in the viral world, distinctive from viruses of Bacteria and Eukarya, and to the recognition of the virosphere of Archaea as one of the distinct features of this Domain of Life. David Prangishvili is a member of the Academia Europaea, the European Academy of Microbiology, and the Georgian National Academy of Sciences.



David I. Stuart is MRC Professor of Structural Biology in the Nuffield Department of Medicine, Oxford University, Oxford, United Kingdom, Life Science Director at Diamond Light Source and Director of Instruct-ERIC (pan-European organisation providing shared access to infrastructure and methods for structural biology). He has diverse interests in structural virology from picornaviruses, double-stranded RNA viruses and enveloped RNA viruses. His drive to develop structural techniques led to the determination of the structure of Bluetongue virus (1995) and then the first membrane containing virus, PRD1. More recently, he has been at the fore-front of bringing Cryo-EM technology to bear on virus structure determination and its future role in visualizing virus function in cellulose. In addition to basic science he has a strong commitment to structural vaccinology and the development of antiviral drugs.



Dr. Nobuhiro Suzuki, PhD, received his MSc (1985) in phytopathology and PhD (1989) in virology from Tohoku University in Sendai, Japan. Dr. Suzuki currently serves as a full Professor of the Institute of Plant Stress and Resources, formerly Institute of Plant Sciences and Bioresources at Okayama University and as an Editor of *Virus Research*, *Frontiers in Virology*, *Journal of General Plant Pathology*, *Virology Journal*, and *Biology*. He has also been Guest Editor to *PLoS Pathogens*, *PNAS*, and *mBio*, and an Editorial Board member of *Virology* and *Journal of Virology*.

Suzuki Laboratory focuses on characterization of diverse viruses infecting phytopathogenic fungi and exploration of their interplays. Recent achievements include the discovery of a neo-virus lifestyle exhibited by a (+)ssRNA virus and an unrelated dsRNA virus in a plant pathogenic fungus and of multilayer antiviral defense in fungi involving Dicer. Prior to coming to Kurashiki, Okayama Prefecture, he was a visiting fellow of the Center for Agricultural Biotechnology at the

University of Maryland Biotechnology Institute (UMBI), College Park, MA, United States, for 4 years (1997–2001) to study molecular biology of hypoviruses in the laboratory of Professor Donald L. Nuss. Before visiting UMBI, he served as an assistant professor and a lecturer of the Biotechnology Institute at the Akita Prefectural College of Agriculture, Japan, for 11 years (1988–1998) where he was engaged in a project on molecular characterization of rice dwarf phytoecovirus, a member of the family *Reoviridae*. He received awards from the Japanese Phytopathological Society of Japan and Japanese Society for Virology for his outstanding achievements in plant and fungal virology.

FOREWORD

I am delighted to write the foreword to this wonderful Fourth Edition of the *Encyclopedia of Virology*. The Third Edition was published in 2008, how the world has changed in the intervening years. The release of the updated fourth edition could not be more timely or more prescient. It is superb and a huge tribute to the authors, Elsevier the publisher, and to the brilliant editors, Dennis Bamford and Mark Zuckerman.

SARS-CoV-2 has dominated the world since it emerged in 2019 and affected every continent and every aspect of life. A reminder, if it were needed, of the impact of infectious diseases, the importance of virology and the vulnerability and interconnectivity of our world.

There is no doubt that with rapidly changing ecology, urbanization, climate change, increased travel, and fragile public health systems, epidemics and pandemics will become more frequent, more complex and harder to prevent and contain. Most of these epidemics will be caused by viruses, those we know about and maybe able to predict and some we do not know of that will emerge from animals, plants or the environment. Our changing climate will change the epidemiology of viruses, their vectors and the infections they cause, hence the critical importance of this totally revised Fourth Edition of the *Encyclopedia of Virology* which brings together research and an understanding of viruses in animals, plants, bacteria and fungi, the environment, and among humans. Never has a holistic, one-health understanding been more important.

That starts with an understanding of the fundamentals of virology, a field of science that has been transformed in the years since the Third Edition. An understanding transformed by embracing traditional fields of molecular and structural biology, genomics, and influenced by immunology, genetics, pharmacology and increasingly by epidemiology and mathematics. Events of 2020 and 2021 also show why it is so important to integrate within traditional virology an understanding of the animal and human health and behavior, of climate change and its impact on the ecology of viruses, plant sciences and vectors. And why we must understand the viruses we think we know well, and those viruses less extensively studied. Research is critical to this, research that pushes the boundaries of what we know, has the humility to seek answers to things we do not understand and shares that knowledge with the widest possible community. That research will be most exciting at the interface between disciplines, most impactful when dynamic, open, inclusive, global, and collaborative.

This is what the Fourth Edition of the *Encyclopedia of Virology*, the largest reference source of research in virology sets out to achieve. It is a wonderful contribution to a critical field of knowledge. It contains new chapters, every chapter revised and updated by a dedicated global community who have come together to provide what is a brilliant and inspiring reference. It is an honor to contribute in a very small way to the timely release of the Fourth Edition of the *Encyclopedia of Virology*.

Jeremy Farrar

PREFACE

The fourth edition of the Encyclopedia of Virology is encyclopedic, but we wanted to move away from an alphabetical list, apart from where it was more logical, to a vision that encompassed a different structure. Articles describing novel trends as well as original discoveries in specific subfields of virology have been distributed into a set of five volumes, namely Fundamentals of Virology, Human and Animal Viruses, Plant Viruses, Bacterial, Archaeal, Fungal, Algal and Invertebrate Viruses, and Diagnosis, Treatment and Prevention of Virus Infections.

We had hoped that the new edition would 'go viral' but it was ironic that the time to publication 12 years after the previous edition had been made a bit longer due to a virus infection. The world encountered a devastating global pandemic, COVID-19, caused by a new type of a coronavirus, SARS-CoV-2. Scientists in many disciplines all over the world started immediate efforts to discover solutions as to how to mitigate and stop the spread of the pandemic. Virology moved from being a highly specialized subject to one in which everyone became a virologist, proving just how significant the different aspects of virology are in terms of understanding the nature of viral infection. Since the previous edition, the growth in the field of general virology has been enormous, including huge advances in basic science, identification of novel viruses, diagnostic methods, treatment and prevention. Taking this into account, the introduction of the articles within the Encyclopedia are very timely and crucial for providing a wealth of knowledge of the latest findings in the field of virology to a vast range of people, whether school students, undergraduates, postgraduates, teachers, scientists, researchers, journalists and others interested in infections and the conflict between the host and the pathogen.

Pandemic viruses have become a serious public concern in the changing world. We can ask ourselves whether we have reached the point in which nature can no longer cope with the consequences of increased population density and human activities that are harmful to the environment. Although several pandemics have threatened mankind before, this COVID-19 pandemic has highlighted the massive adverse economic consequences towards the wellbeing of society and the importance of research in virology.

We aimed to produce a Major Reference Work that differs in approach to others and binds all the virology disciplines together. Chapters have been included on origin, evolution and emergence of viruses, environmental virology and ecology, epidemiology, techniques for studying viruses, viral life cycles, structure, entry, genome and replication, assembly and packaging and taxonomy and viral–host interactions. Information has been included on all known species of viruses infecting bacteria, fungi, plants, vertebrates and invertebrates. Additional topics include antiviral classification and examples of their use in management of infection, diagnostic assays and vaccines, as well as the economic importance of viral diseases of crops and their control.

This edition used viral classification according to the 9th Report of the International Committee on Taxonomy of Viruses published in 2012. Updating it to the 10th Report in 2020 was affected by the pandemic and can be found online at <http://ictv.global/report/>.

We wish to acknowledge the hard work, interest, flexibility and patience, during such difficult times both socially and professionally, of everybody involved in the process of writing this edition of the Encyclopedia of Virology, especially Katarzyna Miklaszewska, Priscilla Braglia, Sam Crowe and colleagues at Elsevier. We sincerely thank all the authors and section editors for their excellent contributions to this edition.

Book Cover Image: Viruses are obligate parasites and all cells have their own viruses increasing the total number of viruses to the estimated astronomical number of 10^{31} that extends the number of stars in the universe. The viral string illustrates how pandemic viruses surround the globe. The original picture was created by Dr. Nina Atanasova (Finnish Meteorological Institute and University of Helsinki) and amended by Matthew Limbert at Elsevier.

*Dennis H. Bamford
Mark Zuckerman*

HOW TO USE THE ENCYCLOPEDIA

Structure of the Encyclopedia

All articles in the encyclopedia are arranged thematically as a series of entries within subjects/sections, apart from volume 2 where there it was more logical to have articles arranged alphabetically.

There are three features to help you easily find the topic you are interested in: a thematic contents list, a full subject index, and contributors.

1. Thematic contents list: The alphabetical contents list, which appears at the front of each volume, lists the entries in the order that they appear in the encyclopedia.
2. Index: The index appears at the end of volume 5 and includes page numbers for quick reference to the information you are looking for. The index entries differentiate between references to a whole entry, a part of an entry, and a table or figure.
3. Contributors: At the start of each volume there is a list of the authors who contributed to all volumes.

LIST OF CONTRIBUTORS

Stephen T. Abedon

*The Ohio State University, Mansfield, OH,
United States*

Peter Abrahamian

*Agricultural Research Service, US Department of
Agriculture, Beltsville, MD, United States*

Jônatas S. Abrahão

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Florence Abravanel

*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Nicola G.A. Abrescia

*Center for Cooperative Research in Biosciences, Basque
Research and Technology Alliance, Derio, Spain;
Ikerbasque, Basque Foundation for Science, Bilbao,
Spain; and Center for Biomedical Research in the Liver
and Digestive Diseases Network, Carlos III Health
Institute, Madrid, Spain*

Gian Paolo Accotto

*Institute for Sustainable Plant Protection, National
Research Council of Italy, Torino, Italy*

Elisabeth Adderson

*St. Jude Children's Research Hospital, Memphis, TN,
United States and University of Tennessee Health
Sciences Center, Memphis, TN, United States*

Mustafa Adhab

University of Baghdad, Baghdad, Iraq

Alexey A. Agranovsky

Lomonosov Moscow State University, Moscow, Russia

Nasim Ahmed

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Maher Al Rwahnih

University of California, Davis, CA, United States

Olufemi J. Alabi

*Texas A&M AgriLife Research and Extension Center,
Weslaco, TX, United States*

Aurélien A. Albertini

*Institute for Integrative Biology of the Cell (I2BC),
French Alternative Energies and Atomic Energy
Commission, French National Center for Scientific
Research, Paris-Sud University, University of
Paris-Saclay, Gif-sur-Yvette, France*

Aleksandra Alimova

*The City University of New York (CUNY), School of
Medicine, The City College of New York, New York,
NY, United States*

Juan C. Alonso

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Imran Amin

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Stephanie E. Ander

*University of Colorado School of Medicine, Aurora, CO,
United States*

Danielle E. Anderson

Duke-NUS Medical School, Singapore, Singapore

Ida Bagus Andika

Qingdao Agricultural University, Qingdao, China

Ana C.d.S.P. Andrade

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Juana Angel

Pontifical Javeriana University, Bogota, Colombia

Vanesa Anton-Vazquez

King's College Hospital, London, United Kingdom

Guido Antonelli

Sapienza University of Rome, Rome, Italy

Josefa Antón

University of Alicante, Alicante, Spain

Nanako Aoki

*Tokyo University of Agriculture and Technology, Fuchu,
Japan*

Timothy D. Appleby

King's College Hospital, London, United Kingdom

Miguel Arenas

*Department of Biochemistry, Genetics and Immunology,
University of Vigo, Vigo, Spain and CINBIO
(Biomedical Research Center), University of Vigo, Vigo,
Spain*

Basil Arif

*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Vicente Arnau

*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain*

Gaurav Arya

Duke University, Durham, NC, United States

Leyla Asadi

University of Alberta, Edmonton, AB, Canada

Sassan Asgari

The University of Queensland, Brisbane, QLD, Australia

Nina S. Atanasova

*Finnish Meteorological Institute, Helsinki, Finland and
University of Helsinki, Helsinki, Finland*

Houssam Attoui

*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Silvia Ayora

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Walid Azab

Free University of Berlin, Berlin, Germany

Sasha R. Azar

*The University of Texas Medical Branch, Galveston, TX,
United States*

Fengwei Bai

*The University of Southern Mississippi, Hattiesburg, MS,
United States*

Dalan Bailey

The Pirbright Institute, Pirbright, United Kingdom

S.C. Baker

*Loyola University of Chicago, Maywood, IL,
United States*

Fausto Baldanti

*University of Pavia, Pavia, Italy and Scientific Institute
for Research, Hospitalization and Healthcare, San
Matteo Polyclinic Foundation, Pavia, Italy*

Logan Banadyga

*Public Health Agency of Canada, Winnipeg, MB,
Canada*

Ashley C. Banyard

*Animal and Plant Health Agency, Addlestone,
United Kingdom; University of West Sussex, Falmer,
United Kingdom; and St. George's Medical School,
University of London, London, United Kingdom*

Xiaoyong Bao

*The University of Texas Medical Branch, Galveston, TX,
United States*

Yiming Bao

*Beijing Institute of Genomics, Chinese Academy of
Sciences, University of Chinese Academy of Sciences,
Beijing, China*

Alan D.T. Barrett

*The University of Texas Medical Branch, Galveston, TX,
United States*

Diana P. Baquero

*Archaeal Virology Unit, Institut Pasteur, Paris, France
and Sorbonne University, Paris, France*

Moshe Bar-Joseph

*Agricultural Research Organization, Volcani Center, Bet
Dagan, Israel*

Rachael S. Barr

*Bristol Royal Hospital for Children, Bristol,
United Kingdom*

Ralf Bartenschlager

Heidelberg University, Heidelberg, Germany

David L.V. Bauer

Francis Crick Institute, London, United Kingdom

Oliver W. Bayfield

University of York, York, United Kingdom

Sally A. Baylis

Paul-Ehrlich-Institute, Langen, Germany

Philippa M. Beard

*The Pirbright Institute, Pirbright, United Kingdom and The
Roslin Institute, University of Edinburgh, United Kingdom*

Paul Becher

University of Veterinary Medicine, Hannover, Germany

Björn Becker

Saarland University, Saarbrücken, Germany

Karen L. Beemon

Johns Hopkins University, Baltimore, MD, United States

Martin Beer

Friedrich-Loeffler-Institute, Insel Riems, Germany

Jose Miguel Benito

*Health Research Institute of the Jiménez Díaz
Foundation, Autonomous University of Madrid and Rey
Juan Carlos University Hospital, Móstoles, Spain*

Mária Benkő

*Institute for Veterinary Medical Research, Center for
Agricultural Research, Budapest, Hungary*

Max Bergoin
National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada

Sabrina Bertin
Council for Agricultural Research and Economics,
Research Center for Plant Protection and Certification,
Rome, Italy

Shweta Bhatt
University of Copenhagen, Copenhagen, Denmark

Dennis K. Bideshi
California Baptist University, Riverside, CA,
United States and University of California, Riverside,
CA, United States

Yves Bigot
INRAE – French National Research Institute for
Agriculture, Food and Environment, Nouzilly, France

Richard J. Bingham
University of York, York, United Kingdom

Vera Bischoff
Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany

Kate N. Bishop
Francis Crick Institute, London, United Kingdom

Lindsay W. Black
The University of Maryland School of Medicine,
Baltimore, MD, United States

Romain Blanc-Mathieu
Institute for Chemical Research, Kyoto University, Kyoto,
Japan

Soile Blomqvist
National Institute for Health and Welfare, Helsinki,
Finland

Bryony C. Bonning
University of Florida, Gainesville, FL, United States

Lisa M. Bono
Rutgers, The State University of New Jersey,
New Brunswick, NJ, United States

Alexia Bordigoni
Aix-Marseille University, CNRS, IRD, Mediterranean
Institute of Oceanography, Marseille, France and
Aix-Marseille University, IRD257, Assistance-Publique
des Hôpitaux de Marseille, UMR Microbes, Evolution,
Phylogeny and Infections (MEPHI), IHU Méditerranée
Infection, Marseille, France

Mihnea Bostina
University of Otago, Dunedin, New Zealand

Maxime Boutier
University of Liège, Liège, Belgium

P.R. Bowser
Cornell University, Ithaca, NY, United States

Daniel Bradshaw
Public Health England, London, United Kingdom

Claude Bragard
University of Louvain, Louvain-la-Neuve, Belgium

Aaron C. Brault
Centers for Disease Control and Prevention, Fort Collins,
CO, United States

Nicolas Bravo-Vasquez
St. Jude Children's Research Hospital, Memphis, TN,
United States

Rob W. Briddon
University of Agriculture, Faisalabad, Pakistan

Thomas Briebe
Columbia University, New York, NY, United States

Paul Britton
The Pirbright Institute, Pirbright, United Kingdom

Thomas J. Brouwers
Athena Institute, VU Amsterdam, Amsterdam, The
Netherlands

Kevin E. Brown
Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom

Corina P.D. Brussaard
NIOZ Royal Netherlands Institute for Sea Research,
Den Burg, Texel, The Netherlands and Utrecht
University, Utrecht, The Netherlands

Harald Brüssow
Laboratory of Gene Technology, Department of
Biosystems, KU Leuven, Leuven, Belgium

Joachim J. Bugert
Bundeswehr Institute of Microbiology, Munich, Germany

Jozef J. Bujarski
Northern Illinois University, DeKalb, IL, United States
and Polish Academy of Sciences, Poznan, Poland

Laura Burga
University of Otago, Dunedin, New Zealand

Sara H. Burkhard
University Hospital of Zurich, Zurich, Switzerland

Cara C. Burns
Centers for Disease Control and Prevention, Atlanta,
GA, United States

Felicity Burt
University of the Free State, Bloemfontein, South Africa

Kerry S. Burton
Leamington Spa, United Kingdom

Sarah J. Butcher
University of Helsinki, Helsinki, Finland

Mathias Büttner
Leipzig University, Leipzig, Germany

Jesse Cahill
Sandia National Labs, Albuquerque, NM, United States

Marianna Calabretto
Sapienza University of Rome, Rome, Italy

Thierry Candresse
The National Research Institute for Agriculture, Food and the Environment, University of Bordeaux, Villenave d'Ornon, France

Alan J. Cann
University of Leicester, Leicester, United Kingdom

Lorenzo Capucci
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Irene Carlon-Andres
University of Oxford, Oxford, United Kingdom

José M. Casasnovas
National Center for Biotechnology, Spanish National Research Council (CSIC), Madrid, Spain

J.W. Casey
Cornell University, Ithaca, NY, United States

R.N. Casey
Cornell University, Ithaca, NY, United States

Sherwood R. Casjens
University of Utah, Salt Lake City, UT, United States

Antonella Casola
The University of Texas Medical Branch, Galveston, TX, United States

José R. Castón
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Carlos E. Catalano
University of Colorado Anschutz Medical Campus, Skaggs School of Pharmacy and Pharmaceutical Sciences, Aurora, CO, United States

Roberto Cattaneo
Mayo Clinic, Rochester, MN, United States

Patrizia Cavadini
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Supranee Chaiwatpongsakorn
Nationwide Children's Hospital, Columbus, OH, United States

Supriya Chakraborty
Jawaharlal Nehru University, New Delhi, India

Yu-Chan Chao
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Tyler P. Chavers
Centers for Disease Control and Prevention, Atlanta, GA, United States

Keping Chen
Jiangsu University, Zhenjiang, China

Xiaorui Chen
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Yanping Chen
Bee Research Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

Dayna Cheng
National Cheng Kung University, Tainan, Taiwan

Quentin Chesnais
University of Strasbourg, Colmar, France

Sotaro Chiba
Nagoya University, Nagoya, Japan

Wah Chiu
Stanford University, Stanford, CA, United States

David Chmielewski
Stanford University, Stanford, CA, United States

Irma E. Cisneros
The University of Texas Medical Branch, Galveston, TX, United States

Lark L. Coffey
University of California, Davis, CA, United States

Alanna B. Cohen
Rutgers University, New Brunswick, NJ, United States

Jeffrey I. Cohen
Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Seth Coleman
Rice University, Houston, TX, United States

Miquel Coll
*Institute for Research in Biomedicine, Barcelona, Spain
 and Institute for Molecular Biology of Barcelona,
 Barcelona, Spain*

John Collinge
*UCL Institute of Prion Diseases, London,
 United Kingdom*

Carina Conceicao
The Pirbright Institute, Pirbright, United Kingdom

Gabriela N. Condezo
*National Center for Biotechnology, Spanish National
 Research Council, Madrid, Spain*

Michaela J. Conley
*MRC-University of Glasgow Centre for Virus Research,
 Glasgow, United Kingdom*

Charles A Coomer
University of Oxford, Oxford, United Kingdom

Anne K. Cordes
*Hannover Medical School, Institute of Virology, Hannover,
 Germany*

Mauricio Cortes Jr.
*Department of Chemistry, College of Arts and Sciences,
 Fort Wayne, IN, United States*

Robert H.A. Coutts
University of Hertfordshire, Hatfield, United Kingdom

Jeff A. Cowley
CSIRO Livestock Industries, Brisbane, QLD, Australia

Robert W. Cross
*The University of Texas Medical Branch, Galveston, TX,
 United States*

Henryk Czosnek
The Hebrew University of Jerusalem, Rehovot, Israel

Håkon Dahle
*Department of Biological Sciences, University of Bergen,
 Bergen, Norway*

Janet M. Daly
*University of Nottingham, Sutton Bonington,
 United Kingdom*

Subha Das
Okayama University, Kurashiki, Japan

Indranil Dasgupta
University of Delhi, New Delhi, India

Sibnarayan Datta
*Defence Research Laboratory, Defence Research and
 Development Organisation (DRDO), Tezpur, Assam,
 India*

Amy Davis
*St Jude Children's Research Hospital, Memphis, TN,
 United States*

William O. Dawson
*Citrus Research and Education Center, Lake Alfred, FL,
 United States and University of Florida, Lake Alfred, FL,
 United States*

Erik De Clercq
*Rega Institute for Medical Research, KU Leuven,
 Leuven, Belgium*

Raoul J. de Groot
Utrecht University, Utrecht, The Netherlands

Juan C. de la Torre
*The Scripps Research Institute, La Jolla, CA,
 United States*

Marcelo De las Heras
University of Zaragoza, Zaragoza, Spain

Juliana Gabriela Silva de Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Athos S. de Oliveira
University of Brasília, Brasília, Brazil

Nicole T. de Stefano
University of Florida, Gainesville, FL, United States

Greg Deakin
NIAB-EMR, East Malling, United Kingdom

Philippe Delfosse
University of Luxembourg, Esch-sur-Alzette, Luxembourg

Natacha Delrez
University of Liège, Liège, Belgium

Tatiana A. Demina
*Molecular and Integrative Biosciences Research Program,
 Faculty of Biological and Environmental Sciences,
 University of Helsinki, Helsinki, Finland*

Ismail Demir
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

Zihni Demirbağ
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

X. Deng
*Loyola University of Chicago, Maywood, IL,
 United States*

Cécile Desbiez
*Plant Pathology Unit, INRAE – French National
 Research Institute for Agriculture, Food and
 Environment, Montfavet, France*

Christelle Desnues

Aix-Marseille University, CNRS, IRD, Mediterranean Institute of Oceanography, Marseille, France and Aix-Marseille University, IRD 257, Assistance-Publique des Hôpitaux de Marseille, UMR Microbes, Evolution, Phylogeny and Infections (MEPHI), IHU Méditerranée Infection, Marseille, France

Samantha J. DeWerff

University of Illinois at Urbana-Champaign, Urbana, IL, United States

Daniele Di Carlo

Sapienza University of Rome, Rome, Italy

Arturo Diaz

La Sierra University, Riverside, CA, United States

Alfredo Diaz-Lara

University of California, Davis, CA, United States

Ralf G. Dietzgen

The University of Queensland, St. Lucia, QLD, Australia

Michele Digiario

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Michael Dills

Montana State University, Bozeman, MT, United States

Wayne Dimech

National Serology Reference Laboratory, Fitzroy, VIC, Australia

Savithramma P. Dinesh-Kumar

University of California, Davis, CA, United States

Linda K. Dixon

The Pirbright Institute, Pirbright, United Kingdom

Valerian V. Dolja

Oregon State University, Corvallis, OR, United States

Aušra Domanska

University of Helsinki, Helsinki, Finland

Leslie L. Domier

Agricultural Research Service, US Department of Agriculture, Urbana, IL, United States

Pilar Domingo-Calap

Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, Valencia, Spain

Tatiana Domitrovic

Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Sarah M. Doore

Michigan State University, East Lansing, MI, United States

Lucy Dorrell

University of Oxford, Oxford, United Kingdom

Rosemary A. Dorrington

Rhodes University, Grahamstown, South Africa

Andor Doszpoly

Hungarian Academy of Sciences, Budapest, Hungary

Andreas Dotzauer

University of Bremen, Bremen, Germany

Simon B. Drysdale

St George's University Hospitals NHS Foundation Trust, London, United Kingdom and St George's, University of London, London, United Kingdom

Robert L. Duda

University of Pittsburgh, Pittsburgh, PA, United States

Carol Duffy

University of Alabama, Tuscaloosa, AL, United States

Siobain Duffy

Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

David D. Dunigan

University of Nebraska–Lincoln, Lincoln, NE, United States

Stéphane Duquerroy

University of Paris-Saclay, Orsay, France and Institut Pasteur, Paris, France

Bas E. Dutilh

Utrecht University, Utrecht, The Netherlands and Radboud University Medical Center, Nijmegen, The Netherlands

Michael Edelstein

Faculty of Medicine, Bar Ilan University, Ramat Gan, Israel

Herman K. Edskes

National Institutes of Health, Bethesda, MD, United States

Rosina Ehmann

Bundeswehr Institute of Microbiology, Munich, Germany

Toufic Elbeaino

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Joanne B. Emerson

University of California, Davis, CA, United States

Ann Emery

University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Christine E. Engeland
*University Hospital Heidelberg and German Cancer
 Research Center, Heidelberg, Germany and
 Witten/Herdecke University, Witten, Germany*

Luis Enjuanes
*National Center for Biotechnology – Spanish National
 Research Council (CNB-CSIC), Madrid, Spain*

Katri Eskelin
University of Helsinki, Helsinki, Finland

Rosa Esteban
*Institute of Biology and Functional Genomics,
 CSIC/University of Salamanca, Salamanca, Spain*

Mary K. Estes
Baylor College of Medicine, Houston, TX, United States

Cassia F. Estofolete
*São José do Rio Preto School of Medicine, São José do
 Rio Preto, Brazil*

Alyssa B. Evans
*National Institutes of Health, Hamilton, MT,
 United States*

Øystein Evensen
Norwegian University of Life Sciences, Oslo, Norway

Alex Evilevitch
*Department of Experimental Medical Science, Lund
 University, Lund, Sweden*

Montserrat Fàbrega-Ferrer
*Institute for Research in Biomedicine, Barcelona, Spain and
 Institute for Molecular Biology of Barcelona, Barcelona,
 Spain*

Francesco Faggioli
*Council for Agricultural Research and Economics,
 Research Center for Plant Protection and Certification,
 Rome, Italy*

Bentley A. Fane
University of Arizona, Tucson, AZ, United States

Brian A. Federici
University of California, Riverside, CA, United States

F. Fenner
Australian National University, Canberra, ACT, Australia

Isabel Fernández de Castro
*Cell Structure Laboratory, National Center for
 Biotechnology – Spanish National Research Council
 (CNB-CSIC), Madrid, Spain*

Giovanni Ferrara
University of Alberta, Edmonton, AB, Canada

Elvira Fiallo-Olivé
*Institute for Mediterranean and Subtropical Horticulture
 “La Mayora”–Spanish National Research Council–University
 of Malaga, Algarrobo-Costa, Málaga, Spain*

Andrew E. Firth
University of Cambridge, Cambridge, United Kingdom

Roland A. Fleck
King’s College London, London, United Kingdom

Ricardo Flores
*Polytechnic University of Valencia, Higher Council of
 Scientific Research, Valencia, Spain*

Ervin Fodor
University of Oxford, Oxford, United Kingdom

Anthony R. Fooks
*Animal and Plant Health Agency, Addlestone,
 United Kingdom; University of Liverpool, Liverpool,
 United Kingdom; and St. George’s Medical School,
 University of London, London, United Kingdom*

Patrick Forterre
*Archeal Virology Unit, Institut Pasteur, Paris, France
 and French National Center for Scientific Research,
 Institute of Integrative Biology of the Cell, University
 of Paris-Saclay, Gif sur Yvette, France*

Rennos Fragkoudis
*University of Nottingham, Sutton Bonington,
 United Kingdom and University of Edinburgh,
 Edinburgh, United Kingdom*

Manuel A. Franco
Pontifical Javeriana University, Bogota, Colombia

Giovanni Franzo
*Department of Animal Medicine, Production and Health
 (MAPS), Padua University, Padua, Italy*

Graham L. Freimanis
The Pirbright Institute, Pirbright, United Kingdom

Juliana Freitas-Astúa
*Brazilian Agricultural Research Corporation (Embrapa)
 Cassava and Fruits, Cruz das Almas, Brazil*

Elizabeth E. Fry
*Department of Structural Biology, Nuffield Department
 of Medicine, University of Oxford, Oxford,
 United Kingdom*

Marc Fuchs
Cornell University, Geneva, NY, United States

Tsutomu Fujimura
*Institute of Biology and Functional Genomics,
 CSIC/University of Salamanca, Salamanca, Spain*

- Kuko Fuke
Tokyo University of Agriculture and Technology, Fuchu, Japan
- Toshiyuki Fukuhara
Tokyo University of Agriculture and Technology, Fuchu, Japan
- To S. Fung
South China Agricultural University, Guangzhou, China
- Yahya Z.A. Gaafar
Julius Kuehn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany
- Toni Gabaldon
Barcelona Supercomputing Center-National Center for Supercomputing, Institute of Research in Biomedicine, and Catalan Institution for Research and Advanced Studies, Barcelona, Spain
- Morgan Gaïa
University of Paris-Saclay, Evry, France
- José Gallardo
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain
- Hernan Garcia-Ruiz
University of Nebraska–Lincoln, Lincoln, NE, United States
- Juan A. García
National Center for Biotechnology-Spanish National Research Council, Madrid, Spain
- Matteo P. Garofalo
The University of Texas Medical Branch, Galveston, TX, United States
- Yves Gaudin
Institute for Integrative Biology of the Cell (I2BC), French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, Paris-Sud University, University of Paris-Saclay, Gif-sur-Yvette, France
- Andrew D.W. Geering
The University of Queensland, St. Lucia, QLD, Australia
- Thomas W. Geisbert
The University of Texas Medical Branch, Galveston, TX, United States
- Andrea Gentili
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy
- Volker Gerdts
University of Saskatchewan, Saskatoon, SK, Canada
- Said A. Ghabrial[†]
Department of Plant Pathology, University of Kentucky, Lexington, KY, United States
- Clément Gilbert
Evolution, Genomes, Behavior and Ecology Laboratory, CNRS University of Paris-Sud UMR 9191, IRD UMR 247, Gif-sur-Yvette, France
- Robert L. Gilbertson
University of California, Davis, CA, United States
- Efstathios S. Giotis
Imperial College London, London, United Kingdom and University of Essex, Colchester, United Kingdom
- Laurent Glais
French Federation of Seed Potato Growers/Research, Development, Promotion of Seed Potato, Paris, France and Institute for Genetics, Environment and Plant Protection, Agrocampus West, French National Institute for Agriculture, Food and Environment, University of Rennes 1, Le Rheu, France
- Miroslav Glasa
Biomedical Research Center, Slovak Academy of Sciences, Bratislava, Slovakia
- Ido Golding
University of Illinois at Urbana-Champaign, Urbana, IL, United States
- Esperanza Gomez-Lucia
Complutense University of Madrid, Madrid, Spain
- Zheng Gong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
- Andrea González-González
University of Florida, Gainesville, FL, United States
- Michael M. Goodin
University of Kentucky, Lexington, KY, United States
- Alexander E. Gorbalenya
Leiden University Medical Center, Leiden, The Netherlands
- Paul Gottlieb
The City University of New York (CUNY), School of Medicine, The City College of New York, New York, NY, United States
- M.-A. Grandbastien
INRAE – French National Research Institute for Agriculture, Food and Environment, Versailles, France

[†]Deceased.

Meritxell Granell
*National Center for Biotechnology, Madrid, Spain and
 Institute of Chemical Research of Catalonia (ICIQ),
 Tarragona, Spain*

Patrick L. Green
*The Ohio State University, Columbus, OH,
 United States*

Sandra J. Greive
University of York, York, United Kingdom

Diane E. Griffin
*Johns Hopkins Bloomberg School of Public Health,
 Baltimore, MD, United States*

Jonathan M. Grimes
University of Oxford, Oxford, United Kingdom

Nigel Grimsley
*Integrative Biology of Marine Organisms Laboratory,
 Banuyls-sur-Mer, France and Sorbonne University,
 Banuyls-sur-Mer, France*

Bruno Gronenborn
*Institute for Integrative Biology of the Cell, CNRS,
 University of Paris-Sud, CEA, Gif sur Yvette, France*

Julianne H. Grose
Brigham Young University, Provo, UT, United States

Scott Grytdal
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

Duane J. Gubler
Duke-NUS Medical School, Singapore, Singapore

Peixuan Guo
*College of Pharmacy, The Ohio State University,
 Columbus, OH, United States*

Tongkun Guo
*Beijing Institute of Genomics, Chinese Academy of
 Sciences, Beijing, China*

Anne-Lise Haenni
*Institut Jacques Monod, French National Center for
 Scientific Research, Paris Diderot University, Paris,
 France*

Susan L. Hafenstein
*Pennsylvania State University, Hershey, PA,
 United States*

Ahmed Hafez
*Biotechvana, Valencia, Spain; Pompeu Fabra University,
 Barcelona, Spain; and Minia University, Minya, Egypt*

Marie Hagbom
Linköping University, Linköping, Sweden

Sébastien Halary
*National Museum of Natural History, UMR 7245
 CNRS/MNHN Molécule de Communication et
 Adaptation des Micro-organismes, Paris, France*

Aron J. Hall
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

John Hammond
*Floral and Nursery Plants Research, Agricultural
 Research Service, US Department of Agriculture,
 Beltsville, MD, United States*

Rosemarie W. Hammond
*Agricultural Research Service, US Department of
 Agriculture, Beltsville, MD, United States*

Virginia Hargest
*St Jude Children's Research Hospital, Memphis, TN,
 United States and University of Tennessee Health
 Science Center, Memphis, TN, United States*

Scott J. Harper
Washington State University, Prosser, WA, United States

Balázs Harrach
*Institute for Veterinary Medical Research, Center for
 Agricultural Research, Budapest, Hungary*

Masayoshi Hashimoto
The University of Tokyo, Tokyo, Japan

Muhammad Hassan
University of Agriculture, Faisalabad, Pakistan

Asma Hatoum-Aslan
University of Alabama, Tuscaloosa, AL, United States

Philippa C. Hawes
The Pirbright Institute, Pirbright, United Kingdom

Janelle A. Hayes
*University of Massachusetts Medical School, Worcester,
 MA, United States*

Guijuan He
Virginia Tech, Blacksburg, VA, United States

Klaus Hedman
*University of Helsinki, Helsinki, Finland and Helsinki
 University Hospital, Helsinki, Finland*

Albert Heim
Hannover Medical School, Hanover, Germany

Gary L. Hein
*University of Nebraska-Lincoln, Lincoln, NE,
 United States*

Manfred Heinlein
*IBMP-CNRS, University of Strasbourg, Strasbourg,
 France*

Mercedes Hernando-Pérez
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Carmen Hernández
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Etienne Herrbach
University of Strasbourg, Colmar, France

Stephen Higgs
*Biosecurity Research Institute, Kansas State University,
Manhattan, KS, United States*

Bradley I. Hillman
Rutgers University, New Brunswick, NJ, United States

Deborah M. Hinton
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Judith Hirsch
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Jody Hobson-Peters
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Natalie M. Holste
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Jin S. Hong
Kangwon National University, Chunchon, South Korea

Margaret J. Hosie
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Olivia G. Howell
University of Alabama, Tuscaloosa, AL, United States

Liya Hu
Baylor College of Medicine, Houston, TX, United States

Zhaoyang Hu
Jiangsu University, Zhenjiang, China

Kuan-Ying A. Huang
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Yu Huang
Peking University, Beijing, China

Natalia B. Hubbs
Hanover College, Hanover, IN, United States

Elisabeth Huguet
*Research Institute on Insect Biology, French National
Center for Scientific Research, University of Tours,
Tours, France*

Roger Hull
John Innes Centre, Norwich, United Kingdom

Kiwamu Hyodo
Okayama University, Kurashiki, Japan

Eugénie Hébrard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

Martin Hölzer
University of Jena, Jena, Germany

Tetsuro Ikegami
*The University of Texas Medical Branch at Galveston,
Galveston, TX, United States*

Niina Ikonen
*Finnish Institute for Health and Welfare, Helsinki,
Finland*

Cihan İnan
*Department of Molecular Biology and Genetics,
Karadeniz Technical University, Trabzon, Turkey*

İkbal Agah İnce
*Department of Medical Microbiology, Acibadem
University School of Medicine, Istanbul, Turkey*

Katsuaki Inoue
Diamond Light Source, Didcot, United Kingdom

Toru Iwanami
Tokyo University of Agriculture, Tokyo, Japan

Jacques Izopet
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Fauziah Mohd Jaafar
*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Andrew O. Jackson
China Agricultural University, Beijing, China

Daral J. Jackwood
*The Ohio State University/OARDC, Wooster, OH,
United States*

Jean-Rock Jacques
*Cellular and Molecular Epigenetics (GIGA), Liège,
 Belgium and Molecular Biology (TERRA), Gembloux,
 Belgium*

Tiffany Jenkins
*Nationwide Children's Hospital, Columbus, OH,
 United States and The Ohio State University, Columbus,
 OH, United States*

Jeffrey D. Jensen
Arizona State University, Tempe, AZ, United States

Daohong Jiang
Huazhong Agricultural University, Wuhan, China

Zhihao Jiang
China Agricultural University, Beijing, China

Allison R. Jilbert
The University of Adelaide, Adelaide, SA, Australia

Peng Jing
*Department of Chemistry, College of Arts and Sciences,
 Fort Wayne, IN, United States*

Xixi Jing
Central South University, Changsha, China

Meesbah Jiwaji
Rhodes University, Grahamstown, South Africa

Kyle L. Johnson
*The University of Texas at El Paso, El Paso, TX,
 United States and CQuentia, Memphis, TN, United
 States*

Welkin E. Johnson
Boston College, Chestnut Hill, MA, United States

Ian M. Jones
*University of Reading, Reading, United Kingdom and
 London School of Hygiene and Tropical Medicine,
 London, United Kingdom*

Ramon Jordan
*Agricultural Research Service, US Department of
 Agriculture, Beltsville, MD, United States*

Thomas Joris
*Cellular and Molecular Epigenetics (GIGA), Liège,
 Belgium and Molecular Biology (TERRA), Gembloux,
 Belgium*

Ilkka Julkunen
*Institute of Biomedicine, University of Turku, Turku,
 Finland*

Sandra Junglen
Charité - University Medicine Berlin, Berlin, Germany

Masanori Kaido
Kyoto University, Kyoto, Japan

Laura Kakkola
University of Turku, Turku, Finland

Hannimari Kallio-Kokko
*University of Helsinki and Helsinki University Hospital,
 Helsinki, Finland*

Nassim Kamar
*Toulouse University Hospital, Toulouse, France and
 Toulouse University Paul Sabatier, Toulouse, France*

Phyllis J. Kanki
*Harvard T.H. Chan School of Public Health, Boston,
 MA, United States*

Peter Karayiannis
University of Nicosia, Nicosia, Cyprus

Henry M. Kariithi
*Kenya Agricultural and Livestock Research Organization,
 Nairobi, Kenya*

Brian A. Kelch
*University of Massachusetts Medical School, Worcester,
 MA, United States*

Karen E. Keller
*Horticultural Crops Research Unit, Agricultural Research
 Service, US Department of Agriculture, Corvallis, OR,
 United States*

Japhette E. Kembou-Ringert
University of Tel Aviv, Tel Aviv, Israel

Peter J. Kerr
*University of Sydney, Sydney, NSW, Australia and
 CSIRO Health and Biosecurity, Black Mountain
 Laboratories, Canberra, ACT, Australia*

Tiffany King
*Nationwide Children's Hospital, Columbus, OH,
 United States and The Ohio State University College
 of Medicine, Columbus, OH, United States*

Andrea Kirmaier
Boston College, Chestnut Hill, MA, United States

Thomas Klose
Purdue University, West Lafayette, IN, United States

Barbara G. Klupp
*Friedrich-Loeffler-Institute, Greifswald-Insel Riems,
 Germany*

David M. Knipe
Harvard Medical School, Boston, MA, United States

Nick J. Knowles
The Pirbright Institute, Pirbright, United Kingdom

Guus Koch
*Wageningen Bioveterinary Research, Lelystad,
 The Netherlands*

Renate Koenig
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

Susanne E. Kohalmi
*The University of Western Ontario, London, ON,
Canada*

Hideki Kondo
Okayama University, Kurashiki, Japan

Jennifer L. Konopka-Anstadt
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Eugene V. Koonin
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States and
National Institutes of Health, Bethesda, MD, United States*

Marion P.G. Koopmans
Erasmus Medical Center, Rotterdam, The Netherlands

Richard Kormelink
*Wageningen University and Research, Wageningen,
The Netherlands*

Ioly Kotta-Loizou
Imperial College London, London, United Kingdom

Peter J. Krell
*Department of Molecular and Cellular Biology,
University of Guelph, Guelph, ON, Canada*

Mart Krupovic
Archaeal Virology Unit, Institut Pasteur, Paris, France

Andreas Kuhn
University of Hohenheim, Stuttgart, Germany

Jens H. Kuhn
*National Institutes of Health, Frederick, MD,
United States*

Richard J. Kuhn
Purdue University, West Lafayette, IN, United States

Suvi Kuivanen
University of Helsinki, Helsinki, Finland

Ranjababu Kulasegaram
*Guy's and St Thomas' NHS Foundation Trust, London,
United Kingdom*

Raghavendran Kulasegaran-Shylini
*Department of Pathogen Infection, Faculty of Infectious
and Tropical Diseases, London School of Hygiene and
Tropical Medicine, London, United Kingdom*

Gaurav Kumar
University of Delhi, New Delhi, India

Manish Kumar
Jawaharlal Nehru University, New Delhi, India

Gael Kurath
*US Geological Survey, Western Fisheries Research
Center, Seattle, WA, United States*

Satu Kurkela
*University of Helsinki and Helsinki University Hospital,
Helsinki, Finland*

Wan-Chun Lai
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Kevin Lamkiewicz
University of Jena, Jena, Germany

Rebecca K. Lane
*University of Texas Health Science Center at San
Antonio, San Antonio, TX, United States*

Andrew S. Lang
*Memorial University of Newfoundland, St. John's, NL,
Canada*

Daniel Carlos Ferreira Lanza
Federal University of Rio Grande do Norte, Natal, Brazil

Maija Lappalainen
*HUS Diagnostic Center, HUSLAB, Clinical
Microbiology, University of Helsinki and Helsinki
University Hospital, Helsinki, Finland*

Katherine LaTourrette
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Chris Lauber
*TWINCORE – Center for Experimental and Clinical
Infection Research, Hannover, Germany*

Antonio Lavazza
*The Lombardy and Emilia Romagna Experimental
Zootechnic Institute, Brescia, Italy*

C. Martin Lawrence
Montana State University, Bozeman, MT, United States

Hervé Lecoq
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Young-Min Lee
Utah State University, Logan, UT, United States

Kristen N. LeGault
University of California, Berkeley, CA, United States

James Legg
*International Institute of Tropical Agriculture, Dar es
Salaam, Tanzania*

Anne Legreve
University of Louvain, Louvain-la-Neuve, Belgium

Petr G. Leiman
The University of Texas Medical Branch, Galveston, TX, United States

Stanley M. Lemon
Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Sebastian Leptihn
Zhejiang University-Edinburgh University Institute, Zhejiang University, Haining, China

Dennis J. Lewandowski
University of Florida, Lake Alfred, FL, United States

Sébastien Lhomme
Toulouse University Hospital, Toulouse, France and Toulouse University Paul Sabatier, Toulouse, France

Dawei Li
China Agricultural University, Beijing, China

Guangdi Li
Central South University, Changsha, China

Guoqing Li
Huazhong Agricultural University, Wuhan, China

Yi Li
Peking University, Beijing, China

Zhefeng Li
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Zhenghe Li
Zhejiang University, Hangzhou, China

Jia Q. Liang
South China Agricultural University, Guangzhou, China

Sebastian Liebe
Institute of Sugar Beet Research, Göttingen, Germany

João Paulo Matos Santos Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Bruno Lina
HCL Department of Virology, National Reference Center for Respiratory Viruses, Institute of Infectious Agents, Croix-Rousse Hospital, Lyon, France and Virpath Laboratory, International Center of Research in Infectiology (CIRI), INSERM U1111, CNRS—UMR 5308, École Normale Supérieure de Lyon, University Claude Bernard Lyon, Lyon University, Lyon, France

Walter Ian Lipkin
Columbia University, New York, NY, United States

Jan G. Lisby
Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

Ding X. Liu
South China Agricultural University, Guangzhou, China

Qiang Liu
University of Saskatchewan, Saskatoon, SK, Canada

Sijun Liu
Iowa State University, Ames, IA, United States

Carlos Llorens
Biotechvana, Scientific Park University of Valencia, Valencia, Spain

L. Sue Loesch-Fries
Purdue University, West Lafayette, IN, United States

George P. Lomonosoff
John Innes Centre, Norwich, United Kingdom

L. Letti Lopez
The University of Texas at Austin, Austin, TX, United States

Alan T. Loynachan
University of Kentucky, Lexington, KY, United States

Garry A. Luke
University of St. Andrews, St. Andrews, United Kingdom

M. Luo
University of Alabama at Birmingham, Birmingham, AL, United States

Juan J. López-Moya
Center for Research in Agricultural Genomics and Spanish National Research Council, Barcelona, Spain

Che Ma
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Stuart A. MacFarlane
The James Hutton Institute, Invergowrie, United Kingdom

Saichetana Macherla
J. Craig Venter Institute, La Jolla, CA, United States

Kensaku Maejima
The University of Tokyo, Tokyo, Japan

Fabrizio Maggi
University of Pisa, Pisa, Italy and University of Insubria, Varese, Italy

Melissa S. Maginnis
The University of Maine, Orono, ME, United States

Edgar Maiss
Leibniz University Hannover, Hannover, Germany

Kira S. Makarova
National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, United States

Ariana Manglli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Annette Mankertz
Robert Koch-Institute, Berlin, Germany

Pilar Manrique
The Ohio State University, Wexner Medical Center, Columbus, OH, United States

Shahid Mansoor
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Marco Marklewitz
Institute of Virology, Charité – University Medicine Berlin, Berlin, Germany

Giovanni P. Martelli[†]
University of Bari Aldo Moro, Bari, Italy

Darren P. Martin
University of Cape Town, Cape Town, South Africa

Robert R. Martin
Horticultural Crops Research Unit, Agricultural Research Service, US Department of Agriculture, Corvallis, OR, United States

Manuel Martinez-Garcia
University of Alicante, Alicante, Spain

Francisco Martinez-Hernandez
University of Alicante, Alicante, Spain

Natalia Martín-González
Autonomous University of Madrid, Madrid, Spain

Joaquín Martínez Martínez
Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Manja Marz
University of Jena, Jena, Germany

Andrea Marzi
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, United States

Hema Masarapu
Sri Venkateswara University, Tirupati, India

Chikara Masuta
Hokkaido University, Sapporo, Japan

Carlos P. Mata
University of Leeds, Leeds, United Kingdom

Jelle Matthijnsens
Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

Claire P. Mattison
Centers for Disease Control and Prevention, Atlanta, GA, United States and Cherokee Nation Assurance, Arlington, VA, United States

William McAllister
Rowan University School of Osteopathic Medicine, Stratford, NJ, United States

Alison A. McBride
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Michael McChesney
University of California, Davis, CA, United States

Elaine McCulloch
Quality Control for Molecular Diagnostics (QCMD), Glasgow, United Kingdom

Andrew J. McMichael
University of Oxford, Oxford, United Kingdom

Alexander McPherson
University of California, Irvine, CA, United States

Irene K. Meki
French National Center for Scientific Research, Montpellier, France

Ulrich Melcher
Oklahoma State University, Stillwater, OK, United States

Tomas A. Melgarejo
University of California, Davis, CA, United States

Michael J. Melzer
Department of Plant and Environmental Protection Sciences, University of Hawaii, Honolulu, HI, United States

Luiza Mendonça
University of Oxford, Oxford, United Kingdom

Xiang-Jin Meng
Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Peter P.C. Mertens
University of Nottingham, Sutton Bonington, United Kingdom

[†]Deceased.

Thomas C. Mettenleiter
Friedrich-Loeffler-Institute, Greifswald-Insel Riems, Germany

Philip D. Minor
St Albans, United Kingdom

Ali Mirazimi
National Veterinary Institute, Uppsala, Sweden and Karolinska Hospital University, Huddinge, Sweden

Nischay Mishra
Columbia University, New York, NY, United States

Edward S. Mocarski
Emory University School of Medicine, Atlanta, GA, United States

Florian Mock
University of Jena, Jena, Germany

Volker Moennig
University of Veterinary Medicine, Hannover, Germany

Ian J. Molineux
The University of Texas at Austin, Austin, TX, United States

Aderito L. Monjane
Norwegian Veterinary Institute, Oslo, Finland

Jacen S. Moore
University of Tennessee Health Science Center, Memphis, TN, United States

Marc C. Morais
The University of Texas Medical Branch, Galveston, TX, United States

Cristina Moraru
Institute for Chemistry and Biology of the Marine Environment, Oldenburg, Germany

Hiromitsu Moriyama
Tokyo University of Agriculture and Technology, Tokyo, Japan

Sergey Y. Morozov
Lomonosov Moscow State University, Moscow, Russia

Thomas E. Morrison
University of Colorado School of Medicine, Aurora, CO, United States

Léa Morvan
University of Liège, Liège, Belgium

Benoît Moury
Plant Pathology Unit, INRAE – French National Institute for Agriculture, Food and Environment, Montfavet, France

Muhammad Mubin
University of Agriculture, Faisalabad, Pakistan

Nicolas J. Mueller
University Hospital of Zurich, Zurich, Switzerland

Emmanuelle Muller
The French Agricultural Research Center for International Development, Joint Research Units–Biology and Genetics of Plant-Pathogen Interactions, Montpellier, France and Biology and Genetics of Plant-Pathogen Interactions, University of Montpellier, The French Agricultural Research Center for International Development, French National Institute for Agricultural Research, Montpellier SupAgro, Montpellier, France

John S. Munday
Massey University, Palmerston North, New Zealand

Jacob H. Munson-McGee
Montana State University, Bozeman, MT, United States and Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Hacer Muratoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Kenan C. Murphy
University of Massachusetts Medical School, Worcester, MA, United States

Ugrappa Nagalakshmi
University of California, Davis, CA, United States

Keizo Nagasaki
Kochi University, Nankoku, Japan

Nazia Nahid
GC University, Faisalabad, Pakistan and University of Agriculture, Faisalabad, Pakistan

Venugopal Nair
The Pirbright Institute, Pirbright, United Kingdom

Remziye Nalçacıoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Shigetou Namba
The University of Tokyo, Tokyo, Japan

Rubab Z. Naqvi
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Rachel Nash
The Pirbright Institute, Surrey, United Kingdom

C.K. Navaratnarajah
Purdue University, West Lafayette, IN, United States

Maria A. Navarrete-Muñoz
Biotechvana, Madrid, Spain; Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid; and Rey Juan Carlos University Hospital, Móstoles, Spain

Jesús Navas-Castillo
Institute for Mediterranean and Subtropical Horticulture "La Mayora"—Spanish National Research Council—University of Malaga, Algarrobo-Costa, Málaga, Spain

Muhammad S. Nawaz-ul-Rehman
University of Agriculture, Faisalabad, Pakistan

Christopher L. Netherton
The Pirbright Institute, Pirbright, United Kingdom

Thu V.P. Nguyen
Baylor College of Medicine, Houston, TX, United States

Annette Niehl
Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

Hubert G.M. Niesters
Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Jozef I. Nissimov
University of Waterloo, Waterloo, ON, Canada

Norman Noah
London School of Hygiene and Tropical Medicine, London, United Kingdom

Mauricio L. Nogueira
São José do Rio Preto School of Medicine, São José do Rio Preto, São Paulo, Brazil

Johan Nordgren
Linköping University, Linköping, Sweden

C. Micha Nübling
Paul-Ehrlich-Institute, Langen, Germany

Visa Nurmi
University of Helsinki, Helsinki, Finland

Donald L. Nuss
University of Maryland, Rockville, MD, United States

M. Steven Oberste
Centers for Disease Control and Prevention, Atlanta, GA, United States

Hiroyuki Ogata
Institute for Chemical Research, Kyoto University, Kyoto, Japan

Ane Ogbe
University of Oxford, Oxford, United Kingdom

Hanna M. Oksanen
Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Graziele P. Oliveira
Federal University of Minas Gerais, Belo Horizonte, Brazil

Francesco Origi
University of Bern, Bern, Switzerland

Nikolaus Osterrieder
Free University of Berlin, Berlin, Germany

Robert A. Owens
Beltsville Agricultural Research Center, Beltsville, MD 20705, United States

Emine Özsahin
University of Guelph, Guelph, ON, Canada

Sergi Padilla-Parra
University of Oxford, Oxford, United Kingdom; Department of Infectious Diseases, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom; and Randall Division of Cell and Molecular Biophysics, King's College London, London, United Kingdom

Joshua Pajak
Duke University, Durham, NC, United States

Massimo Palmarini
MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Amanda R. Panfil
The Ohio State University, Columbus, OH, United States

Marcus Panning
Institute of Virology, Freiburg University Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Vitantonio Pantaleo
National Research Council, Research Unit of Bari, Bari, Italy

Anna Papa
Aristotle University of Thessaloniki, Thessaloniki, Greece

Nikolaos Pappas
Utrecht University, Utrecht, The Netherlands

Hanu R. Pappu
Washington State University, Pullman, WA, United States

Kristin N. Parent
*Michigan State University, East Lansing, MI,
United States*

Colin R. Parrish
Cornell University, Ithaca, NY, United States

A. Lorena Passarelli
Kansas State University, Manhattan, KS, United States

Basavaprabhu L. Patil
*ICAR–Indian Institute of Horticultural Research,
Bengaluru, India*

Jade Pattyn
University of Antwerp, Antwerp, Belgium

T.A. Paul
Cornell University, Ithaca, NY, United States

Lillian Pavlik
*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Susan L. Payne
Texas A&M University, College Station, TX, United States

Michael N. Pearson
The University of Auckland, Auckland, New Zealand

Mark E. Peeples
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University College of
Medicine, Columbus, OH, United States*

Ben Peeters
*Wageningen Bioveterinary Research, Lelystad,
The Netherlands*

Joseph S.M. Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Malik Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Judit J. Péntzes
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Miryam Pérez-Cañamás
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Marta Pérez-Illana
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Jaume Pérez-Sánchez
*Institute of Aquaculture Torre de la Sal, Spanish
National Research Council, Castellon, Spain*

Jean-Marie Peron
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Karin E. Peterson
National Institutes of Health, Hamilton, MT, United States

Karel Petrzik
*Biology Center CAS, Institute of Plant Molecular
Biology, České Budějovice, Czech Republic*

Mahtab Peyambari
*Pennsylvania State University, State College, PA,
United States*

Sujal Phadke
J. Craig Venter Institute, La Jolla, CA, United States

Hanh T. Pham
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Mauro Pistello
University of Pisa, Pisa, Italy

Daniel Ponndorf
John Innes Centre, Norwich, United Kingdom

Leo L.M. Poon
The University of Hong Kong, Pok Fu Lam, Hong Kong

Welkin H. Pope
University of Pittsburgh, Pittsburgh, PA, United States

Minna M. Poranen
University of Helsinki, Helsinki, Finland

Claudine Porta
*The Pirbright Institute, Pirbright, United Kingdom and
University of Oxford, Oxford, United Kingdom*

Samuel S. Porter
*National Institute of Allergy and Infectious Diseases,
National Institutes of Health, Bethesda, MD,
United States and University of Maryland, College Park,
MD, United States*

Frank A. Post
*King's College Hospital NHS Foundation Trust, London,
United Kingdom*

Nils Poulicard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

David Prangishvili
*Institut Pasteur, Paris, France and Ivane Javakhishvili
Tbilisi State University, Tbilisi, Georgia*

B. V. Venkataram Prasad
Baylor College of Medicine, Houston, TX, United States

Lalita Priyamvada
Centers for Disease Control and Prevention, Atlanta, GA, United States

Simone Prospero
Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

Elisabeth Puchhammer-Stöckl
Medical University of Vienna, Vienna, Austria

Jianming Qiu
University of Kansas Medical Center, Kansas City, KS, United States

S.L. Quackenbush
Colorado State University, Fort Collins, CO, United States

Killian J. Quinn
King's College Hospital, London, United Kingdom

Diego F. Quito-Avila
Department of Life Sciences, ESPOL Polytechnic University, Guayaquil, Ecuador

Frank Rabenstein
Julius Kühn Institute, Quedlinburg, Germany

Sheli R. Radoshitzky
United States Army Medical Research Institute of Infectious Diseases, Frederick, MD, United States

Saleem U. Rahman
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Mbolarinosy Rakotomalala
FOFIFA, Antananarivo, Madagascar

Norma Rallon
Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid and Rey Juan Carlos University Hospital, Móstoles, Spain

Robert P. Rambo
Diamond Light Source, Didcot, United Kingdom

Bertha Cecilia Ramirez
The Institute for Integrative Biology of the Cell, The French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

María D. Ramos-Barbero
University of Alicante, Alicante, Spain

Chris M. Rands
University of Geneva Medical School and Swiss Institute of Bioinformatics, Geneva, Switzerland

Venigalla B. Rao
The Catholic University of America, Washington, DC, United States

Janne J. Ravantti
University of Helsinki, Helsinki, Finland

Mandy Ravensbergen
Wageningen University and Research, Wageningen, The Netherlands

Georget Y. Reaiche-Miller
The University of Adelaide, Adelaide, SA, Australia

D.V.R. Reddy
International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

Vishwanatha R.A.P. Reddy
The Pirbright Institute, Pirbright, United Kingdom

Juan Reguera
Aix-Marseille University, French National Center for Scientific Research, Marseille, France and French National Institute of Health and Medical Research, Marseille, France

William K. Reisen
University of California, Davis, CA, United States

Jingshan Ren
University of Oxford, Oxford, United Kingdom

Renato O. Resende
University of Brasilia, Brasilia, Brazil

Peter A. Revill
The Peter Doherty Institute of Infection and Immunity, Royal Melbourne Hospital, Melbourne, VIC, Australia

Félix A. Rey
Institut Pasteur, Paris, France

Lara Rheinemann
University of Utah, Salt Lake City, UT, United States

Simone G. Ribeiro
Embrapa Genetic Resources and Biotechnology, Brasília, Brazil

Daniel Rigling
Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Cristina Risco
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Efraín E. Rivera-Serrano

Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Cécile Robin

INRAE – French National Research Institute for Agriculture, Food and Environment, UMR BIOGECO, Cestas, France

Rodrigo A.L. Rodrigues

Federal University of Minas Gerais, Belo Horizonte, Brazil

Elina Roine

University of Helsinki, Helsinki, Finland

Maria R. Rojas

University of California, Davis, CA, United States

Marilyn J. Roossinck

Pennsylvania State University, State College, PA, United States

Vera I.D. Ros

Wageningen University and Research, Wageningen, The Netherlands

Cristina Rosa

Pennsylvania State University, University Park, PA, United States

Hanna Rose

Leibniz University Hannover, Hannover, Germany

David A. Rosenbaum

University of Florida, Gainesville, FL, United States

Shannan L. Rossi

The University of Texas Medical Branch, Galveston, TX, United States

Michael G. Rossmann[†]

Purdue University, West Lafayette, IN, United States

L. Roux

University of Geneva Medical School, Geneva, Switzerland

Simon Roux

US Department of Energy Joint Genome Institute, Walnut Creek, CA, United States

J. Rovnak

Colorado State University, Fort Collins, CO, United States

David J. Rowlands

University of Leeds, Leeds, United Kingdom

Polly Roy

Department of Pathogen Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom and University of Reading, Reading, United Kingdom

Aaron P. Roznowski

The University of Texas at Austin, Austin, TX, United States and University of Arizona, Tucson, AZ, United States

Luisa Rubino

Institute for Sustainable Plant Protection, National Research Council, Bari, Italy

Olli Ruuskanen

Turku University Hospital, Turku, Finland

Eugene V. Ryabov

USDA, Agricultural Research Service, Beltsville, MD, United States

Martin D. Ryan

University of St. Andrews, St. Andrews, United Kingdom

Ki H. Ryu

Seoul Women's University, Seoul, South Korea

Hanns-Joachim Rziha

Eberhard Karls University of Tübingen, Tübingen, Germany

Sead Sabanadzovic

Mississippi State University, Starkville, MS, United States

Roghayeh Safari

Cellular and Molecular Epigenetics (GIGA), Liège, Belgium and Molecular Biology (TERRA), Gembloux, Belgium

Azeez Sait Sahul Hameed

C. Abdul Hakeem College, Melvisharam, India

Nicole Samies

University of Alabama at Birmingham, Birmingham, AL, United States

Carmen San Martín

National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Ruth-Anne Sandaa

Department of Biological Sciences, University of Bergen, Bergen, Norway

Hélène Sanfaçon

Agriculture and Agri-Food Canada, Summerland, BC, Canada

[†]Deceased.

Rafael Sanjuán
*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia-CSIC, Valencia, Spain*

Neeraja Sankaran
Utrecht University, Utrecht, The Netherlands

Fernando Santos
University of Alicante, Alicante, Spain

Cecilia Sarmiento
Tallinn University of Technology, Tallinn, Estonia

Takahide Sasaya
*National Agriculture and Food Research Organization,
Fukuyama, Japan*

Preethi Sathanantham
Virginia Tech, Blacksburg, VA, United States

Panayampalli S. Satheshkumar
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Yukiyo Sato
Okayama University, Kurashiki, Japan

Andreas Sauerbrei
Jena University Hospital, Jena, Germany

Eugene I. Savenkov
*Swedish University of Agricultural Sciences, Uppsala,
Sweden and Linnean Center for Plant Biology, Uppsala,
Sweden*

Carita Savolainen-Kopra
*National Institute for Health and Welfare, Helsinki,
Finland*

Kay Scheets
*Oklahoma State University, Stillwater, OK,
United States*

Uffe V. Schenider
*Copenhagen University Hospital Hvidovre, Hvidovre,
Denmark*

Richard H. Scheuermann
*J. Craig Venter Institute, La Jolla, CA, United States;
University of California, San Diego, CA, United States;
La Jolla Institute for Immunology, La Jolla, CA,
United States; and Global Virus Network, Baltimore,
MD, United States*

Manfred J. Schmitt
Saarland University, Saarbrücken, Germany

James E. Schoelz
University of Missouri, Columbia, MO, United States

Jason R. Schrad
*Michigan State University, East Lansing, MI,
United States*

Declan C. Schroeder
*University of Reading, Reading, United Kingdom and
University of Minnesota, St. Paul, MN, United States*

Stacey Schultz-Cherry
*St. Jude Children's Research Hospital, Memphis, TN,
United States*

Thomas F. Schulz
*Hannover Medical School, Institute of Virology,
Hannover, Germany and German Center for Infection
Research, Hannover-Braunschweig Site, Braunschweig,
Germany*

Catherine A. Scougall
The University of Adelaide, Adelaide, SA, Australia

Kimberley D. Seed
University of California, Berkeley, CA, United States

Joaquim Segalés
*Departament of Animal Health and Anatomy, Faculty of
Veterinary Medicine, Autonomous University of
Barcelona, Barcelona, Spain; Animal Health Research
Center (CReSA) – Institute of Agrifood Research and
Technology (IRTA), Campus UAB, Barcelona, Spain;
and OIE Collaborating Center for the Research and
Control of Emerging and Re-emerging Swine Diseases in
Europe (IRTA-CReSA), Barcelona, Spain*

Mateo Seoane-Blanco
National Center for Biotechnology, Madrid, Spain

Madhumati Sevvana
Purdue University, West Lafayette, IN, United States

Kazım Sezen
*Department of Biology, Karadeniz Technical University,
Trabzon, Turkey*

Arvind Sharma
Institut Pasteur, Paris, France

Sumit Sharma
Linköping University, Linköping, Sweden

James M. Sharp
*University of Zaragoza, Zaragoza, Spain and Edinburgh,
United Kingdom*

Qunxin She
Shandong University, Qingdao, China

Keith E. Shearwin
The University of Adelaide, Adelaide, SA, Australia

Hanako Shimura
Hokkaido University, Sapporo, Japan

Reina S. Sikkema
Erasmus Medical Center, Rotterdam, The Netherlands

Aaron Simkovich
Agriculture and Agri-Food Canada, London, ON,
Canada and The University of Western Ontario,
London, ON, Canada

Peter Simmonds
University of Oxford, Oxford, United Kingdom

Tarja Sironen
University of Helsinki, Helsinki, Finland

Susanna Sissonen
Finnish Institute for Health and Welfare, Helsinki,
Finland

Michael A. Skinner
Imperial College London, London, United Kingdom

Douglas E. Smith
University of California, San Diego, La Jolla, CA,
United States

Melvyn Smith
Viapath Analytics, Specialist Virology Centre, King's
College NHS Foundation Trust, London, United
Kingdom

Thomas J. Smith
The University of Texas Medical Branch, Galveston, TX,
United States

Teemu Smura
Helsinki University Hospital and University of Helsinki,
Helsinki, Finland

Eric J. Snijder
Leiden University Medical Center, Leiden,
The Netherlands

Gisela Soboll Hussey
Michigan State University, East Lansing, MI,
United States

Maria Söderlund-Venemo
University of Helsinki, Helsinki, Finland

Merike Sõmera
Tallinn University of Technology, Tallinn, Estonia

Eun G. Song
Seoul Women's University, Seoul, South Korea

Milan J. Sonneveld
Erasmus University Medical Center, Rotterdam,
The Netherlands

Beatriz Soriano
Biotechvana, Scientific Park University of Valencia and
Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain

Thomas E. Spencer
University of Missouri, Columbia, MO, United States

Pothur Sreenivasulu
Sri Venkateswara University, Tirupati, India

Ashley L. St. John
Duke-NUS Medical School, Singapore, Singapore

David K. Stammers
University of Oxford, Oxford, United Kingdom

John Stanley
John Innes Centre, Colney, United Kingdom

Glyn Stanway
University of Essex, Colchester, United Kingdom

Thilo Stehle
University of Tuebingen, Tuebingen, Germany and
Vanderbilt University School of Medicine, Nashville, TN,
United States

Gregory W. Stevenson
Iowa State University, Ames, IA, United States

Lucy Rae Stewart
Agricultural Research Service, US Department of
Agriculture, Wooster, OH, United States

C.C.M.M. Stijger
Wageningen University and Research Center, Bleiswijk,
The Netherlands

Peter G. Stockley
University of Leeds, Leeds, United Kingdom

David Stone
Weymouth Laboratory, Centre for Environment, Fisheries
and Aquaculture Science, Weymouth, United Kingdom

Ashley E Strother
The University of Texas Medical Branch, Galveston, TX,
United States

Sundharraman Subramanian
Michigan State University, East Lansing, MI,
United States

William C. Summers
Yale University, New Haven, CT, United States

Liyang Sun
Northwest A&F University, Yangling, China

Wesley I. Sundquist
University of Utah, Salt Lake City, UT, United States

Petri Susi
University of Turku, Turku, Finland

Curtis A. Suttle
University of British Columbia, Vancouver, BC, Canada

Nobuhiro Suzuki
Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

Lennart Svensson
Linköping University, Linköping, Sweden and Karolinska Institute, Stockholm, Sweden

Ronald Swanstrom
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Daniele M. Swetnam
University of California, Davis, CA, United states

Moriah L. Szpara
Pennsylvania State University, University Park, PA, United States

Keisuke Tabata
Heidelberg University, Heidelberg, Germany

Anna Taglienti
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Naoki Takeshita
Tokyo University of Agriculture and Technology, Fuchu, Japan

Kana Takeshita Urayama
Tokyo University of Agriculture and Technology, Fuchu, Japan

Michael E. Taliansky
The James Hutton Institute, Dundee, United Kingdom

Pan Tao
The Catholic University of America, Washington, DC, United States

Jacqueline E. Tate
Centers for Disease Control and Prevention, Atlanta, GA, United States

Satyanarayana Tatineni
Agricultural Research Service, US Department of Agriculture, Lincoln, NE, United States and University of Nebraska–Lincoln, Lincoln, NE, United States

Sisko Tauriainen
University of Turku, Turku, Finland

Norbert Tautz
University of Luebeck, Luebeck, Germany

Paulo Tavares
Institute for Integrative Biology of the Cell, CEA, CNRS, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

Nicholas M.I. Taylor
University of Copenhagen, Copenhagen, Denmark

Xu Tengzhi
University of California, Davis, CA, United States

Raquel Tenorio
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Robert B. Tesh
The University of Texas Medical Branch, Galveston, TX, United States

Vaskar Thapa
Pennsylvania State University, State College, PA, United States

John E. Thomas
The University of Queensland, Brisbane, QLD, Australia

Julie A. Thomas
Rochester Institute of Technology, Rochester, NY, United States

Lynn C. Thomason
Frederick National Laboratory for Cancer Research, Frederick, MD, United States

Elizabeth Ashley Thompson
The University of Southern Mississippi, Hattiesburg, MS, United States

Jeremy R. Thompson
Cornell University, Ithaca, NY, United States

Antonio Tiberini
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Peter Tijssen
National Institute of Scientific Research – Armand-Frappier Health Research Centre, Microbiology and Immunology, Laval, QC, Canada

Yuji Tomaru
Japan Fisheries Research and Education Agency, Kanagawa, Japan

Laura Tomassoli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Ruben Torres
National Biotechnology Center–Spanish National Research Council, Madrid, Spain

Jia Q. Truong
The University of Adelaide, Adelaide, SA, Australia

Erkki Truve[†]
Tallinn University of Technology, Tallinn, Estonia

Chih-Hsuan Tsai
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Roman Tuma
University of Leeds, Leeds, United Kingdom and
University of South Bohemia, České Budějovice, Czech Republic

Topi Turunen
Infectious Disease Unit, Espoo, Finland and Finnish
Institute for Health and Welfare, Helsinki, Finland

Reidun Twarock
University of York, York, United Kingdom

Ioannis E. Tzanetakis
University of Arkansas, Fayetteville, United States

Antti Vaheri
University of Helsinki, Helsinki, Finland

Eeva J. Vainio
Natural Resources Institute Finland (Luke), Helsinki, Finland

Anna M. Vaira
Institute for Sustainable Plant Protection, National
Research Council of Italy, Torino, Italy

Steven M. Valles
Center for Medical, Agricultural and Veterinary
Entomology, Agricultural Research Service, US
Department of Agriculture, Gainesville, FL,
United States

Adrián Valli
National Center for Biotechnology-Spanish National
Research Council, Madrid, Spain

Rodrigo A. Valverde
Louisiana State University Agricultural Center, Baton
Rouge, United States

Pierre Van Damme
University of Antwerp, Antwerp, Belgium

Rene A.A. van der Vlugt
Wageningen University and Research Center,
Wageningen, The Netherlands

Bernard A.M. Van der Zeijst
Leiden University Medical Center, Leiden,
The Netherlands

Koenraad Van Doorslaer
University of Arizona, Tucson, AZ, United States

James L. Van Etten
University of Nebraska–Lincoln, Lincoln, NE,
United States

Suzanne van Meer
University Medical Center Utrecht, Utrecht,
The Netherlands

Monique M. van Oers
Wageningen University and Research, Wageningen,
The Netherlands

Mark J. van Raaij
National Center for Biotechnology, Madrid, Spain

Marc H.V. Van Regenmortel
University of Strasbourg, Strasbourg, France

Piet A. van Rijn
Wageningen Bioveterinary Research, Lelystad, The
Netherlands and North-West University, Potchefstroom,
South Africa

Alain Vanderplasschen
University of Liège, Liège, Belgium

Dana L. Vanlandingham
College of Veterinary Medicine, Kansas State University,
Manhattan, KS, United States

Olli Vapalahti
Helsinki University Hospital and University of Helsinki,
Helsinki, Finland

Mark Varrelmann
Institute of Sugar Beet Research, Göttingen, Germany

Nikos Vasilakis
The University of Texas Medical Branch, Galveston, TX,
United States

Michael Veit
Free University of Berlin, Berlin, Germany

Česlovas Venclovas
Vilnius University, Vilnius, Lithuania

H. Josef Vetten
Julius Kühn Institute, Braunschweig, Germany

Marli Vlok
University of British Columbia, Vancouver, BC, Canada

Anne-Nathalie Volkoff
Diversity, Genomes and Insects-Microorganisms
Interactions, National Institute of Agricultural Research,
University of Montpellier, Montpellier, France

Ian E.H. Voorhees
Cornell University, Ithaca, NY, United States

Alex Vorsters
University of Antwerp, Antwerp, Belgium

[†]Deceased.

Jonathan D.F. Wadsworth
UCL Institute of Prion Diseases, London, United Kingdom

Peter J. Walker
The University of Queensland, St. Lucia, QLD, Australia

Paul Wallace
*Quality Control for Molecular Diagnostics (QCMD),
Glasgow, United Kingdom*

Aiming Wang
*Agriculture and Agri-Food Canada, London, ON,
Canada*

Jen-Ren Wang
National Cheng Kung University, Tainan, Taiwan

Lin-Fa Wang
Duke-NUS Medical School, Singapore, Singapore

Nan Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiangxi Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiaofeng Wang
Virginia Tech, Blacksburg, VA, United States

Katherine N. Ward
University College London, London, United Kingdom

Matti Waris
University of Turku, Turku, Finland

Ranjit Warrier
Purdue University, West Lafayette, IN, United States

Daniel Watterson
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Marta L. Wayne
University of Florida, Gainesville, FL, United States

Friedemann Weber
*FB 10 – Institute for Virology, Justus Liebig University
Giessen, Giessen, Germany*

Sung-Chan Wei
*Institute of Molecular Biology, Academia Sinica, Taipei,
Taiwan*

Robin A. Weiss
University College London, London, United Kingdom

Tao Weitao
*Southwest Baptist University, Bolivar, MO,
United States*

Kerstin Wernike
Friedrich-Loeffler-Institute, Insel Riems, Germany

Rachel J. Whitaker
*University of Illinois at Urbana-Champaign, Urbana, IL,
United States*

K. Andrew White
York University, Toronto, ON, Canada

Anna E. Whitfield
*North Carolina State University, Raleigh, NC,
United States*

Richard Whitley
*University of Alabama at Birmingham, Birmingham,
AL, United States*

Reed B. Wickner
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Luc Willems
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Brian J. Willett
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Alexis Williams
*The University of Texas Medical Branch, Galveston, TX,
United States*

Stephen A. Winchester
*Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom*

Clayton W. Winkler
*National Institutes of Health, Hamilton, MT,
United States*

Stephan Winter
*Leibniz Institute – DSMZ – German Collection of
Microorganisms and Cell Cultures GmbH, Braunschweig,
Germany*

William M. Wintermantel
*Agricultural Research Service, US Department of
Agriculture, Salinas, CA, United States*

Jennifer Wirth
Montana State University, Bozeman, MT, United States

Yuri I. Wolf
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States*

Thorsten Wolff
Robert Koch Institute, Berlin, Germany

Blaide Woodburn
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Michael E. Woodson
The University of Texas Medical Branch, Galveston, TX, United States

Courtney Woolsey
The University of Texas Medical Branch, Galveston, TX, United States

Chien-Fu Wu
Tokyo University of Agriculture and Technology, Fuchu, Japan

Mingde Wu
Huazhong Agricultural University, Wuhan, China

Songsong Wu
Huazhong Agricultural University, Wuhan, China

Yan Xiang
University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

Jiatao Xie
Huazhong Agricultural University, Wuhan, China

Zhuang Xiong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Hajime. Yaegashi
Institute of Fruit Tree and Tea Science, NARO, Morioka, Japan

Mehtap Yakupoğlu
Trabzon University, Trabzon, Turkey

Yasuyuki Yamaji
The University of Tokyo, Tokyo, Japan

Meng Yang
China Agricultural University, Beijing, China

Teng-Chieh Yang
Scarsdale, NY, United States

Qin Yao
Jiangsu University, Zhenjiang, China

Tianyou Yao
Baylor College of Medicine, Houston, TX, United States

Nobuyuki Yoshikawa
Iwate University, Morioka, Japan

George R. Young
Francis Crick Institute, London, United Kingdom

Lawrence S. Young
University of Warwick, Coventry, United Kingdom

Mark J. Young
Montana State University, Bozeman, MT, United States

Ry Young
Texas A&M University, College Station, TX, United States

Isaac T. Younker
University of Alabama, Tuscaloosa, AL, United States

Qian Yu
School of Life Sciences, Jiangsu University, Zhenjiang, China

Sang-Im Yun
Utah State University, Logan, UT, United States

Fauzia Zarreen
University of Delhi, New Delhi, India

Francisco M. Zerbini
Federal University of Viçosa, Viçosa, Brazil

Dong-Xiu Zhang
University of Maryland, Rockville, MD, United States

Jianqiang Zhang
Iowa State University, Ames, IA, United States

Junjie Zhang
Texas A&M University, College Station, TX, United States

Long Zhang
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Pan Zhang
Central South University, Changsha, China

Peijun Zhang
University of Oxford, Oxford, United Kingdom and Electron Bio-Imaging Centre, Diamond Light Source, Didcot, United Kingdom

Tao Zhang
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Yongliang Zhang
China Agricultural University, Beijing, China

Zhenlu Zhang
Shandong Agricultural University, Tai'an, China

Lixia Zhou
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Ling Zhu
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Heiko Ziebell
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

John Ziebuhr
*The Queen's University of Belfast, Belfast,
United Kingdom*

Jeffrey J. Zimmerman
Iowa State University, Ames, IA, United States

Falk Zucker
*Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany*

CONTENT OF ALL VOLUMES

Editors in Chief	v
Editorial Board	vii
Section Editors	ix
Foreword	xv
Preface	xvii
Guide to Use	xix
List of Contributors	xxi

VOLUME 1

The Virus as a Concept – Fundamentals of Virology

A Brief History of Virology <i>David J Rowlands</i>	3
The Origin of Viruses <i>Patrick Forterre and Morgan Gaïa</i>	14
The Virocell Concept <i>Patrick Forterre</i>	23
Virus Taxonomy <i>Jens H Kuhn</i>	28
The Greater Virus World and Its Evolution <i>Eugene V Koonin and Valerian V Dolja</i>	38
The Virus Species Concept <i>Peter Simmonds</i>	47
Genetic Diversity and Evolution of Viral Populations <i>Rafael Sanjuán and Pilar Domingo-Calap</i>	53
Mechanisms of RNA Virus Evolution <i>Lisa M Bono and Siobain Duffy</i>	62
Mechanisms of DNA Virus Evolution <i>Moriah L Szpara and Koenraad Van Doorslaer</i>	71
Paleovirology <i>Clément Gilbert</i>	79
Evolution Steered by Structure <i>Nicola GA Abrescia</i>	87

Pairwise Sequence Comparison in Virology <i>Tao Zhang, Zheng Gong, Tongkun Guo, Zhuang Xiong, and Yiming Bao</i>	100
Computational Analysis of Recombination in Viral Nucleotide Sequences <i>Miguel Arenas</i>	108
Phylogeny of Viruses <i>Alexander E Gorbalenya and Chris Lauber</i>	116
Virus Bioinformatics <i>Nikolaos Pappas, Simon Roux, Martin Hölzer, Kevin Lamkiewicz, Florian Mock, Manja Marz, and Bas E Dutilh</i>	124
Metagenomics in Virology <i>Simon Roux, Jelle Matthijnsens, and Bas E Dutilh</i>	133
Database and Analytical Resources for Viral Research Community <i>Sujal Phadke, Saichetana Macherla, and Richard H Scheuermann</i>	141
Classification of the Viral World Based on Atomic Level Structures <i>Janne J Ravantti and Nicola GA Abrescia</i>	153
Isolating, Culturing, and Purifying Viruses With a Focus on Bacterial and Archaeal Viruses <i>Katri Eskelin and Hanna M Oksanen</i>	162
High Throughput Sequencing and Virology <i>Graham L Freimanis and Nick J Knowles</i>	175
Single-Virus Genomics: Studying Uncultured Viruses, One at a Time <i>Manuel Martinez-Garcia, Francisco Martinez-Hernandez, and Joaquín Martínez Martínez</i>	184
Biophysical Characterizations in the Solution State <i>Robert P Rambo and Katsuaki Inoue</i>	191
Virus Crystallography <i>Jonathan M Grimes</i>	199
Advanced Light and Correlative Microscopy in Virology <i>Sergi Padilla-Parra, Charles A Coomer, and Irene Carlon-Andres</i>	208
Atomic Force Microscopy (AFM) Investigation of Viruses <i>Alexander McPherson</i>	218
Cryo-Electron Microscopy (CEM) Structures of Viruses <i>David Chmielewski and Wah Chiu</i>	233
Analysis of Viruses in the Cellular Context by Electron Tomography <i>Peijun Zhang and Luiza Mendonça</i>	242
Mathematical Modeling of Virus Architecture <i>Reidun Twarock</i>	248
Principles of Virus Structure <i>Madhumati Sevvana, Thomas Klose, and Michael G Rossmann[†]</i>	257
Structures of Small Icosahedral Viruses <i>Elizabeth E Fry, Jingshan Ren, and Claudine Porta</i>	278
Structural Principles of the Flavivirus Particle Organization and of Its Conformational Changes <i>Stéphane Duquerroy, Arvind Sharma, and Félix A Rey</i>	290
Reoviruses (Reoviridae) and Their Structural Relatives <i>Liya Hu, Mary K Estes, and B V Venkataram Prasad</i>	303

[†]Deceased.

Structures of Tailed Phages and Herpesviruses (<i>Herpesviridae</i>) <i>Montserrat Fàbrega-Ferrer and Miquel Coll</i>	318
Adenoviruses (<i>Adenoviridae</i>) and Their Structural Relatives <i>Gabriela N Condezo, Natalia Martín-González, Marta Pérez-Illana, Mercedes Hernando-Pérez, José Gallardo, and Carmen San Martín</i>	329
Negative Single-Stranded RNA Viruses (<i>Mononegavirales</i>): A Structural View <i>Juan Reguera</i>	345
Structure of Retrovirus Particles (<i>Retroviridae</i>) <i>David K Stammers and Jingshan Ren</i>	352
Structure of Helical Viruses <i>C Martin Lawrence</i>	362
Giant Viruses and Their Virophage Parasites <i>Rodrigo AL Rodrigues, Ana CdSP Andrade, Grazielle P Oliveira, and Jônatas S Abrahão</i>	372
Viral Replication Cycle <i>AJ Cann</i>	382
Viral Receptors <i>José M Casasnovas and Thilo Stehle</i>	388
Bacterial and Archeal Virus Entry <i>Minna M Poranen and Aušra Domanska</i>	402
Nonenveloped Eukaryotic Virus Entry <i>Ian M Jones and Polly Roy</i>	409
Enveloped Virus Membrane Fusion <i>Aurélien A Albertini and Yves Gaudin</i>	417
Genome Replication of Bacterial and Archaeal Viruses <i>Česlovas Venclovas</i>	429
Viral Transcription <i>David LV Bauer and Ervin Fodor</i>	439
Translation of Viral Proteins <i>Martin D Ryan and Garry A Luke</i>	444
Recombination <i>Jozef J Bujarski</i>	460
Assembly of Viruses: Enveloped Particles <i>CK Navaratnarajah, R Warriar, and RJ Kuhn</i>	468
Assembly of Viruses: Nonenveloped Particles <i>M Luo</i>	475
Virion Assembly: From Small Picornaviruses (<i>Picornaviridae</i>) to Large Herpesviruses (<i>Herpesviridae</i>) <i>Ling Zhu, Nan Wang, and Xiangxi Wang</i>	480
Genome Packaging <i>Richard J Bingham, Reidun Twarock, Carlos P Mata, and Peter G Stockley</i>	488
Virus Factories <i>Isabel Fernández de Castro, Raquel Tenorio, and Cristina Risco</i>	495
Release of Phages From Prokaryotic Cells <i>Jesse Cahill and Ry Young</i>	501
Virus Budding <i>Lara Rheinemann and Wesley I Sundquist</i>	519

Vesicle-Mediated Transcytosis and Export of Viruses <i>Efraín E Rivera-Serrano and Stanley M Lemon</i>	529
Vector Transmission of Animal Viruses <i>Houssam Attoui, Fauziah Mohd Jaafar, Rennos Fragkoudis, and Peter PC Mertens</i>	542
The Human Virome <i>Alexia Bordigoni, Sébastien Halary, and Christelle Desnues</i>	552
Epidemiology of Human and Animal Viral Diseases <i>Michael Edelstein</i>	559
Zoonosis, Emerging and Re-Emerging Viral Diseases <i>Janet M Daly</i>	569
Antiviral Innate Immunity: Introduction <i>Friedemann Weber</i>	577
Humoral and T Cell-Mediated Immunity to Viruses <i>Ane Ogbe and Lucy Dorrell</i>	584
Antigenicity and Antigenic Variation <i>Kuan-Ying A Huang, Xiaorui Chen, Che Ma, Dayna Cheng, Jen-Ren Wang, and Wan-Chun Lai</i>	597
Antigen Presentation <i>Andrew J McMichael</i>	601
Defense Against Viruses and Other Genetic Parasites in Prokaryotes <i>Kira S Makarova, Yuri I Wolf, and Eugene V Koonin</i>	606
Defective-Interfering Viruses <i>L Roux</i>	617
Ecology and Global Impacts of Viruses <i>Joanne B Emerson</i>	621
The Role of Retroviruses in Cellular Evolution <i>Andrea Kirmaier and Welkin E Johnson</i>	627
The Role of Bacteriophages in Bacterial Evolution <i>Chris M Rands and Harald Brüssow</i>	633
Viruses and Their Potential for Bioterrorism <i>Dana L Vanlandingham and Stephen Higgs</i>	644
The Use of Viral Promoters in Expression Vectors <i>Ian M Jones</i>	652
Oncolytic Viruses <i>Laura Burga and Mihnea Bostina</i>	658
Biotechnology Approaches to Modern Vaccine Design <i>George P Lomonosoff and Daniel Ponndorf</i>	662
Viruses: Impact on Science and Society <i>Neeraja Sankaran and Robin A Weiss</i>	671

VOLUME 2

Viruses as Infectious Agents: Human and Animal Viruses

Adenoviruses (<i>Adenoviridae</i>) <i>Balázs Harrach and Mária Benkő</i>	3
---	---

African Horse Sickness Virus (<i>Reoviridae</i>) <i>Piet A van Rijn</i>	17
African Swine Fever Virus (<i>Asfarviridae</i>) <i>Linda K Dixon, Rachel Nash, Philippa C Hawes, and Christopher L Netherton</i>	22
Akabane Virus and Schmallenberg Virus (<i>Peribunyaviridae</i>) <i>Martin Beer and Kerstin Wernike</i>	34
Alphaviruses Causing Encephalitis (<i>Togaviridae</i>) <i>Diane E Griffin</i>	40
Anelloviruses (<i>Anelloviridae</i>) <i>Fabrizio Maggi and Mauro Pistello</i>	48
Animal Lentiviruses (<i>Retroviridae</i>) <i>Esperanza Gomez-Lucia</i>	56
Animal Morbilliviruses (<i>Paramyxoviridae</i>) <i>Carina Conceicao and Dalan Bailey</i>	68
Animal Papillomaviruses (<i>Papillomaviridae</i>) <i>John S Munday</i>	79
Astroviruses (<i>Astroviridae</i>) <i>Virginia Hargest, Amy Davis, and Stacey Schultz-Cherry</i>	92
Avian Hepadnaviruses (<i>Hepadnaviridae</i>) <i>Allison R Jilbert, Georget Y Reaiche-Miller, and Catherine A Scougall</i>	100
Avian Herpesviruses (<i>Herpesviridae</i>) <i>Vishwanatha RAP Reddy and Venugopal Nair</i>	112
Avian Influenza Viruses (<i>Orthomyxoviridae</i>) <i>Nicolas Bravo-Vasquez and Stacey Schultz-Cherry</i>	117
Avian Leukosis and Sarcoma Viruses (<i>Retroviridae</i>) <i>Karen L Beemon</i>	122
Bluetongue Virus (<i>Reoviridae</i>) <i>Raghavendran Kulasegaran-Shylini and Polly Roy</i>	127
Borna Disease Virus and Related Bornaviruses (<i>Bornaviridae</i>) <i>Susan L Payne</i>	137
Bovine Leukemia Virus (<i>Retroviridae</i>) <i>Thomas Joris, Roghaiyeh Safari, Jean-Rock Jacques, and Luc Willems</i>	144
Bovine Viral Diarrhea, Border Disease, and Classical Swine Fever Viruses (<i>Flaviviridae</i>) <i>Paul Becher, Volker Moennig, and Norbert Tautz</i>	153
Capripoxviruses, Parapoxviruses, and Other Poxviruses of Ruminants (<i>Poxviridae</i>) <i>Philippa M Beard</i>	165
Chikungunya Virus (<i>Togaviridae</i>) <i>Thomas E Morrison and Stephanie E Ander</i>	173
Circoviruses (<i>Circoviridae</i>) <i>Giovanni Franzo and Joaquim Segalés</i>	182
Coronaviruses: General Features (<i>Coronaviridae</i>) <i>Paul Britton</i>	193
Coronaviruses: Molecular Biology (<i>Coronaviridae</i>) <i>X Deng and SC Baker</i>	198

Crimean-Congo Hemorrhagic Fever Virus and Nairoviruses of Medical Importance (<i>Nairoviridae</i>) <i>Ali Mirazimi, Felicity Burt, and Anna Papa</i>	208
Dengue Viruses (<i>Flaviviridae</i>) <i>Ashley L. St. John and Duane J. Gubler</i>	218
Ebola Virus (<i>Filoviridae</i>) <i>Andrea Marzi and Logan Banadyga</i>	232
Enteroviruses (<i>Picornaviridae</i>) <i>Carita Savolainen-Kopra, Soile Blomqvist, and Petri Susi</i>	245
Enveloped, Positive-Strand RNA Viruses (<i>Nidovirales</i>) <i>L. Enjuanes, AE Gorbalenya, RJ de Groot, JA Cowley, J Ziebuhr, and EJ Snijder</i>	256
Epstein-Barr Virus (<i>Herpesviridae</i>) <i>Lawrence S. Young</i>	267
Equine Herpesviruses (<i>Herpesviridae</i>) <i>Gisela Soboll-Hussey, Nikolaus Osterrieder, and Walid Azab</i>	278
Equine, Canine, and Swine Influenza (<i>Orthomyxoviridae</i>) <i>Janet M. Daly and Japhette E. Kembou-Ringert</i>	287
Feline Calicivirus (<i>Caliciviridae</i>) <i>Margaret J. Hosie and Michaela J. Conley</i>	294
Feline Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>Brian J. Willett and Margaret J. Hosie</i>	300
Fish and Amphibian Alloherpesviruses (<i>Herpesviridae</i>) <i>Maxime Boutier, Léa Morvan, Natacha Delrez, Francesco Origi, Andor Doszpoly, and Alain Vanderplasschen</i>	306
Fish Retroviruses (<i>Retroviridae</i>) <i>TA Paul, RN Casey, PR Bowser, JW Casey, J Rovnak, and SL Quackenbush</i>	316
Fish Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Gael Kurath and David Stone</i>	324
Foot-and-Mouth Disease Viruses (<i>Picornaviridae</i>) <i>David J. Rowlands</i>	332
Fowlpox Virus and Other Avipoxviruses (<i>Poxviridae</i>) <i>Efstathios S. Giotis and Michael A. Skinner</i>	343
Hantaviruses (<i>Hantaviridae</i>) <i>Tarja Sironen and Antti Vaheri</i>	349
Henipaviruses (<i>Paramyxoviridae</i>) <i>Lin-Fa Wang and Danielle E. Anderson</i>	355
Hepatitis A Virus (<i>Picornaviridae</i>) <i>Andreas Dotzauer</i>	362
Hepatitis B Virus (<i>Hepadnaviridae</i>) <i>Peter Karayiannis</i>	373
Hepatitis C Virus (<i>Flaviviridae</i>) <i>Ralf Bartenschlager and Keisuke Tabata</i>	386
Hepeviruses (<i>Hepeviridae</i>) <i>Xiang-Jin Meng</i>	397
Herpes Simplex Virus 1 and 2 (<i>Herpesviridae</i>) <i>David M. Knipe and Richard Whitley</i>	404

History of Virology: Vertebrate Viruses <i>F Fenner</i>	414
Human Boca- and Protoparvoviruses (<i>Parvoviridae</i>) <i>Maria Söderlund-Venermo and Jianming Qiu</i>	419
Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (<i>Coronaviridae</i>) <i>Ding X Liu, Jia Q Liang, and To S Fung</i>	428
Human Cytomegalovirus (<i>Herpesviridae</i>) <i>Edward S Mocarski</i>	441
Human Immunodeficiency Virus (<i>Retroviridae</i>) <i>Blaide Woodburn, Ann Emery, and Ronald Swanstrom</i>	460
Human Metapneumovirus (<i>Pneumoviridae</i>) <i>Antonella Casola, Matteo P Garofalo, and Xiaoyong Bao</i>	475
Human Norovirus and Sapovirus (<i>Caliciviridae</i>) <i>Sumit Sharma, Marie Hagbom, Lennart Svensson, and Johan Nordgren</i>	483
Human Papillomaviruses (<i>Papillomaviridae</i>) <i>Alison A McBride and Samuel S Porter</i>	493
Human Parainfluenza Viruses (<i>Paramyxoviridae</i>) <i>Elisabeth Adderson</i>	502
Human Pathogenic Arenaviruses (<i>Arenaviridae</i>) <i>Sheli R Radoshitzky and Juan C de la Torre</i>	507
Human Polyomaviruses (<i>Papillomaviridae</i>) <i>Melissa S Maginnis</i>	518
Human T-Cell Leukemia Virus-1 and -2 (<i>Retroviridae</i>) <i>Amanda R Panfil and Patrick L Green</i>	528
Infectious Bursal Disease Virus (<i>Birnaviridae</i>) <i>Daral J Jackwood</i>	540
Infectious Pancreatic Necrosis Virus (<i>Birnaviridae</i>) <i>Øystein Evensen</i>	544
Influenza A Viruses (<i>Orthomyxoviridae</i>) <i>Laura Kakkola, Niina Ikonen, and Ilkka Julkunen</i>	551
Influenza B, C and D Viruses (<i>Orthomyxoviridae</i>) <i>Thorsten Wolff and Michael Veit</i>	561
Jaagsiekte Sheep Retrovirus (<i>Retroviridae</i>) <i>James M Sharp, Marcelo De las Heras, Massimo Palmarini, and Thomas E Spencer</i>	575
Japanese Encephalitis Virus (<i>Flaviviridae</i>) <i>Sang-Im Yun and Young-Min Lee</i>	583
Kaposi's Sarcoma-Associated Herpesvirus (<i>Herpesviridae</i>) <i>Anne K Cordes and Thomas F Schulz</i>	598
Marburg and Ravn Viruses (<i>Filoviridae</i>) <i>Courtney Woolsey, Thomas W Geisbert, and Robert W Cross</i>	608
Measles Virus (<i>Paramyxoviridae</i>) <i>Roberto Cattaneo and Michael McChesney</i>	619
Molluscum Contagiosum Virus (<i>Poxviridae</i>) <i>Joachim J Bugert and Rosina Ehmann</i>	629

Mumps Virus (<i>Paramyxoviridae</i>) <i>Stephen A Winchester and Kevin E Brown</i>	634
Murine Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>George R Young and Kate N Bishop</i>	643
Newcastle Disease Virus (<i>Paramyxoviridae</i>) <i>Ben Peeters and Guus Koch</i>	648
Orthobunyaviruses (<i>Peribunyaviridae</i>) <i>Alyssa B Evans, Clayton W Winkler, and Karin E Peterson</i>	654
Parapoxviruses (<i>Poxviridae</i>) <i>Hanns-Joachim Rziha and Mathias Büttner</i>	666
Parechoviruses (<i>Picornaviridae</i>) <i>Sisko Tauriainen and Glyn Stanway</i>	675
Parvoviruses of Carnivores, and the Emergence of Canine Parvovirus (<i>Parvoviridae</i>) <i>Colin R Parrish, Ian EH Voorhees, and Susan L Hafenstein</i>	683
Polioviruses (<i>Picornaviridae</i>) <i>Philip D Minor</i>	688
Porcine Reproductive and Respiratory Syndrome Virus and Equine Arteritis Virus (<i>Arteriviridae</i>) <i>Jianqiang Zhang, Alan T Loynachan, Gregory W Stevenson, and Jeffrey J Zimmerman</i>	697
Prions of Vertebrates <i>Jonathan DF Wadsworth and John Collinge</i>	707
Pseudorabies Virus (<i>Herpesviridae</i>) <i>Thomas C Mettenleiter and Barbara G Klupp</i>	714
Rabbit Hemorrhagic Disease Virus and European Brown Hare Syndrome Virus (<i>Caliciviridae</i>) <i>Lorenzo Capucci, Patrizia Cavadini, and Antonio Lavazza</i>	724
Rabbit Myxoma Virus and the Fibroma Viruses (<i>Poxviridae</i>) <i>Peter J Kerr</i>	730
Rabies and Other Lyssaviruses (<i>Rhabdoviridae</i>) <i>Ashley C Banyard and Anthony R Fooks</i>	738
Respiratory Syncytial Virus (<i>Pneumoviridae</i>) <i>Tiffany King, Tiffany Jenkins, Supranee Chaiwatpongsakorn, and Mark E Peebles</i>	747
Rhinoviruses (<i>Picornaviridae</i>) <i>Matti Waris and Olli Ruuskanen</i>	757
Rift Valley Fever Virus and Other Phleboviruses (<i>Phenuiviridae</i>) <i>Tetsuro Ikegami</i>	765
Roseoloviruses: Human Herpesviruses 6A, 6B and 7 (<i>Herpesviridae</i>) <i>Katherine N Ward</i>	778
Rotaviruses (<i>Reoviridae</i>) <i>Juana Angel and Manuel A Franco</i>	789
Rubella Virus (<i>Picornaviridae</i>) <i>Annette Mankertz</i>	797
Saint Louis Encephalitis Virus (<i>Flaviviridae</i>) <i>William K Reisen, Lark L Coffey, Daniele M Swetnam, and Aaron C Brault</i>	805
Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (<i>Coronaviridae</i>) <i>Joseph SM Peiris and Leo LM Poon</i>	814

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (<i>Coronaviridae</i>) <i>Malik Peiris</i>	825
Simian Immunodeficiency Virus (SIV) and HIV-2 (<i>Retroviridae</i>) <i>Phyllis J Kanki</i>	827
Sindbis Virus (<i>Togaviridae</i>) <i>Satu Kurkela</i>	837
Tick-Borne Encephalitis Virus (<i>Flaviviridae</i>) <i>Teemu Smura, Suvi Kuivanen, and Olli Vapalahti</i>	843
Transmissible Gastroenteritis Virus of Pigs and Porcine Epidemic Diarrhea Virus (<i>Coronaviridae</i>) <i>Qiang Liu and Volker Gerds</i>	850
Vaccinia Virus (<i>Poxviridae</i>) <i>Yan Xiang and Rebecca K Lane</i>	854
Varicella-Zoster Virus (<i>Herpesviridae</i>) <i>Jeffrey I Cohen</i>	860
Variola and Monkeypox Viruses (<i>Poxviridae</i>) <i>Lalita Priyamvada and Panayampalli S Satheshkumar</i>	868
Vesicular Stomatitis Virus and Bovine Ephemeral Fever Virus (<i>Rhabdoviridae</i>) <i>Peter J Walker and Robert B Tesh</i>	875
West Nile Virus (<i>Flaviviridae</i>) <i>Fengwei Bai and Elizabeth Ashley Thompson</i>	884
Yellow Fever Virus (<i>Flaviviridae</i>) <i>Ashley E Strother and Alan DT Barrett</i>	891
Zika Virus (<i>Flaviviridae</i>) <i>Nikos Vasilakis, Shannan L Rossi, Sasha R Azar, Irma E Cisneros, Cassia F Estofolete, and Mauricio L Nogueira</i>	899

VOLUME 3

Viruses as Infectious Agents: Plant Viruses

An Introduction to Plant Viruses <i>Roger Hull</i>	3
Emerging and Re-Emerging Plant Viruses <i>Sabrina Bertin, Francesco Faggioli, Andrea Gentili, Ariana Manglli, Anna Taglienti, Antonio Tiberini, and Laura Tomassoli</i>	8
Emerging Geminiviruses (<i>Geminiviridae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, and Muhammad Mubin</i>	21
Movement of Viruses in Plants <i>Manfred Heinlein</i>	32
Plant Antiviral Defense: Gene-Silencing Pathways <i>Vitantonio Pantaleo, Chikara Masuta, and Hanako Shimura</i>	43
Plant Resistance to Viruses: Engineered Resistance <i>Marc Fuchs</i>	52
Plant Resistance to Viruses: Natural Resistance Associated With Dominant Genes <i>Mandy Ravensbergen and Richard Kormelink</i>	60

Plant Resistance to Viruses: Natural Resistance Associated With Recessive Genes <i>Masayoshi Hashimoto, Kensaku Maejima, Yasuyuki Yamaji, and Shigetou Namba</i>	69
Plant Viral Diseases: Economic Implications <i>Basavaprabhu L Patil</i>	81
Retrotransposons of Plants <i>M-A Grandbastien</i>	98
Vector Transmission of Plant Viruses <i>Etienne Herrbach and Quentin Chesnais</i>	106
Viral Suppressors of Gene Silencing <i>Hernan Garcia-Ruiz</i>	116
Virus-Induced Gene Silencing (VIGS) <i>Xu Tengzhi, Ugrappa Nagalakshmi, and Savithramma P Dinesh-Kumar</i>	123
Alfalfa Mosaic Virus (<i>Bromoviridae</i>) <i>L Sue Loesch-Fries</i>	132
Alphaflexiviruses (<i>Alphaflexiviridae</i>) <i>Sergey Y Morozov and Alexey A Agranovsky</i>	140
Alphasatellites (<i>Alphasatellitidae</i>) <i>Rob W Briddon and Muhammad S Nawaz-ul-Rehman</i>	149
Amalgaviruses (<i>Amalgaviridae</i>) <i>Ioannis E Tzanetakis, Sead Sabanadzovic, and Rodrigo A Valverde</i>	154
Badnaviruses (<i>Caulimoviridae</i>) <i>Andrew DW Geering</i>	158
Banana Bunchy Top Virus (<i>Nanoviridae</i>) <i>John E Thomas</i>	169
Barley Yellow Dwarf Viruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	176
Bean Common Mosaic Virus and Bean Common Mosaic Necrosis Virus (<i>Potyviridae</i>) <i>Ramon Jordan and John Hammond</i>	184
Bean Golden Mosaic Virus and Bean Golden Yellow Mosaic Virus (<i>Geminiviridae</i>) <i>Francisco M Zerbini and Simone G Ribeiro</i>	192
Beet Curly Top Virus (<i>Geminiviridae</i>) <i>Robert L Gilbertson, Tomas A Melgarejo, Maria R Rojas, William M Wintermantel, and John Stanley</i>	200
Beet Necrotic Yellow Vein Virus (<i>Benyviridae</i>) <i>Sebastian Liebe, Annette Niehl, Renate Koenig, and Mark Varrelmann</i>	213
Benyviruses (<i>Benyviridae</i>) <i>Annette Niehl, Sebastian Liebe, Mark Varrelmann, and Renate Koenig</i>	219
Betaflexiviruses (<i>Betaflexiviridae</i>) <i>Nobuyuki Yoshikawa and Hajime Yaegashi</i>	229
Betasatellites and Deltasatellites (<i>Toleusatellitidae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, Muhammad Hassan, and Muhammad Mubin</i>	239
Bluner-, Cile-, and Higreviruses (<i>Kitaviridae</i>) <i>Diego F Quito-Avila, Juliana Freitas-Astúa, and Michael J Melzer</i>	247
Brome Mosaic Virus (<i>Bromoviridae</i>) <i>Guijuan He, Zhenlu Zhang, Preethi Sathanantham, Arturo Diaz, and Xiaofeng Wang</i>	252

Bromoviruses (<i>Bromoviridae</i>) <i>Jozef J Bujarski</i>	260
Bymoviruses (<i>Potyviridae</i>) <i>Annette Niehl and Frank Rabenstein</i>	268
Cacao Swollen Shoot Virus (<i>Caulimoviridae</i>) <i>Emmanuelle Muller</i>	274
Carmo-Like Viruses (<i>Tombusviridae</i>) <i>Miryam Pérez-Cañamás and Carmen Hernández</i>	285
Cassava Brown Streak Viruses (<i>Potyviridae</i>) <i>Basavaprabhu L Patil</i>	293
Cassava Mosaic Viruses (<i>Geminiviridae</i>) <i>James Legg and Stephan Winter</i>	301
Caulimoviruses (<i>Caulimoviridae</i>) <i>James E Schoelz and Mustafa Adhab</i>	313
Cheraviruses, Sadwaviruses and Torradoviruses (<i>Secoviridae</i>) <i>Toru Iwanami and René AA van der Vlugt</i>	322
Citrus Tristeza Virus (<i>Closteroviridae</i>) <i>Moshe Bar-Joseph, Scott J Harper, and William O Dawson</i>	327
Closteroviruses (<i>Closteroviridae</i>) <i>Marc Fuchs</i>	336
Comoviruses and Fabaviruses (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	348
Cotton Leaf Curl Disease (<i>Geminiviridae</i>) <i>Nasim Ahmed, Imran Amin, and Shahid Mansoor</i>	355
Cowpea Mosaic Virus (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	364
Cucumber Mosaic Virus (<i>Bromoviridae</i>) <i>Judith Hirsch and Benoît Moury</i>	371
Dianthovirus (<i>Tombusviridae</i>) <i>Kiwamu Hyodo and Masanori Kaido</i>	383
Endornaviruses (<i>Endornaviridae</i>) <i>Toshiyuki Fukuhara</i>	388
Fimoviruses (<i>Fimoviridae</i>) <i>Toufic Elbeaino and Michele Digiaro</i>	396
Furoviruses (<i>Virgaviridae</i>) <i>Annette Niehl and Renate Koenig</i>	405
Geminiviruses (<i>Geminiviridae</i>) <i>Jesús Navas-Castillo and Elvira Fiallo-Olivé</i>	411
Hordeiviruses (<i>Virgaviridae</i>) <i>Zhihao Jiang, Meng Yang, Yongliang Zhang, Andrew O Jackson, and Dawei Li</i>	420
Idaeoviruses (<i>Mayoviridae</i>) <i>Robert R Martin and Karen E Keller</i>	430
Illaviruses (<i>Bromoviridae</i>) <i>Aaron Simkovich, Susanne E Kohalmi, and Aiming Wang</i>	439

Luteoviruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	447
Machlomovirus and Panicoviruses (<i>Tombusviridae</i>) <i>Kay Scheets</i>	456
Maize Streak Virus (<i>Geminiviridae</i>) <i>Darren P Martin and Aderito L Monjane</i>	461
Nanoviruses (<i>Nanoviridae</i>) <i>Bruno Gronenborn and H Josef Vetten</i>	470
Necro-Like Viruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Giovanni P Martelli[†]</i>	481
Nepoviruses (<i>Secoviridae</i>) <i>Hélène Sanfaçon</i>	486
Ophioviruses (<i>Aspiviridae</i>) <i>Anna M Vaira and John Hammond</i>	495
Orthospoviruses (<i>Tospoviridae</i>) <i>Renato O Resende and Hanu R Pappu</i>	507
Ourmiaviruses (<i>Botourmiaviridae</i>) <i>Gian Paolo Accotto and Cristina Rosa</i>	516
Papaya Ringspot Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	520
Pecluviruses (<i>Virgaviridae</i>) <i>Hema Masarapu, Pothur Sreenivasulu, Philippe Delfosse, Claude Bragard, Anne Legreve, and DVR Reddy</i>	528
Pepino Mosaic Virus (<i>Alphaflexiviridae</i>) <i>Rene AA van der Vlugt and CCMM Stijger</i>	539
Plant Reoviruses (<i>Reoviridae</i>) <i>Yu Huang and Yi Li</i>	545
Plant Resistance to Geminiviruses <i>Basavaprabhu L Patil, Supriya Chakraborty, Henryk Czosnek, Elvira Fiallo-Olivé, Robert L Gilbertson, James Legg, Shahid Mansoor, Jesús Navas-Castillo, Rubab Z Naqvi, Saleem U Rahman, and Francisco M Zerbini</i>	554
Plant Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Ralf G Dietzgen, Michael M Goodin, and Zhenghe Li</i>	567
Plant Satellite Viruses (<i>Albetovirus, Aumaivirus, Papanivirus, Virtovirus</i>) <i>Mart Krupovic</i>	581
Plum Pox Virus (<i>Potyviridae</i>) <i>Miroslav Glasa and Thierry Candresse</i>	586
Poleroviruses (<i>Luteoviridae</i>) <i>Hernan Garcia-Ruiz, Natalie M Holste, and Katherine LaTourrette</i>	594
Pomoviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	603
Potato Virus Y (<i>Potyviridae</i>) <i>Laurent Glais and Benoît Moury</i>	612
Potexviruses (<i>Alphaflexiviridae</i>) <i>Ki H Ryu, Eun G Song, and Jin S Hong</i>	623

[†]Deceased.

Potyviruses (<i>Potyviridae</i>) <i>Adrián Valli, Juan A García, and Juan J López-Moya</i>	631
Quinviruses (<i>Betaflexiviridae</i>) <i>Ki H Ryu and Eun G Song</i>	642
Reverse-Transcribing Viruses (<i>Belpaoviridae, Metaviridae, and Pseudoviridae</i>) <i>Carlos Llorens, Beatriz Soriano, Maria A Navarrete-Muñoz, Ahmed Hafez, Vicente Arnau, Jose Miguel Benito, Toni Gabaldon, Norma Rallon, Jaume Pérez-Sánchez, and Mart Krupovic</i>	653
Rice Tungro Disease (<i>Secoviridae, Caulimoviridae</i>) <i>Gaurav Kumar, Fauzia Zarreen, and Indranil Dasgupta</i>	667
Rice Yellow Mottle Virus (<i>Solemoviridae</i>) <i>Eugénie Hébrard, Nils Poulicard, and Mbolarinosy Rakotomalala</i>	675
Satellite Nucleic Acids and Viruses <i>Olufemi J Alabi, Alfredo Diaz-Lara, and Maher Al Rwahnih</i>	681
Secoviruses (<i>Secoviridae</i>) <i>Jeremy R Thompson</i>	692
Sequiviruses and Waikaviruses (<i>Secoviridae</i>) <i>Lucy Rae Stewart</i>	703
Solemoviruses (<i>Solemoviridae</i>) <i>Cecilia Sarmiento, Merike Sõmera, and Erkki Truve[†]</i>	712
Tenuiviruses (<i>Phenuiviridae</i>) <i>Bertha Cecilia Ramirez and Anne-Lise Haenni</i>	719
Tobacco Mosaic Virus (<i>Virgaviridae</i>) <i>Marc HV Van Regenmortel</i>	727
Tobamoviruses (<i>Virgaviridae</i>) <i>Ulrich Melcher, Dennis J Lewandowski, and William O Dawson</i>	734
Tobraviruses (<i>Virgaviridae</i>) <i>Stuart A MacFarlane</i>	743
Tomato Leaf Curl New Delhi Virus (<i>Geminiviridae</i>) <i>Supriya Chakraborty and Manish Kumar</i>	749
Tomato Spotted Wilt Virus (<i>Tospoviridae</i>) <i>Hanu R Pappu, Anna E Whitfield, and Athos S de Oliveira</i>	761
Tomato Yellow Leaf Curl Viruses (<i>Geminiviridae</i>) <i>Henryk Czosnek</i>	768
Tombusvirus-Like Viruses (<i>Tombusviridae</i>) <i>K Andrew White</i>	778
Tombusviruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Kay Scheets</i>	788
Tritimoviruses and Rymoviruses (<i>Potyviridae</i>) <i>Satyanarayana Tatineni and Gary L Hein</i>	797
Triviruses (<i>Betaflexiviridae</i>) <i>Yahya ZA Gaafar and Heiko Ziebell</i>	805
Tymoviruses (<i>Tymoviridae</i>) <i>Rosemarie W Hammond and Peter Abrahamian</i>	818

[†]Deceased.

Umbraviruses (<i>Tombusviridae</i>) <i>Eugene V Ryabov and Michael E Taliansky</i>	827
Varicosaviruses (<i>Rhabdoviridae</i>) <i>Takahide Sasaya</i>	833
Virgaviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	839
Viroids (<i>Pospiviroidae</i> and <i>Avsunviroidae</i>) <i>Ricardo Flores and Robert A Owens</i>	852
Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	862

VOLUME 4

Viruses as Infectious Agents: Bacterial, Archaeal, Fungal, Algal, and Invertebrate Viruses

Bacterial Viruses

History of Virology: Bacteriophages <i>William C Summers</i>	3
Icosahedral Phages – Single-Stranded DNA (ϕ X174) <i>Bentley A Fane and Aaron P Roznowski</i>	10
Single-Stranded RNA Bacterial Viruses <i>Peter G Stockley and Junjie Zhang</i>	21
Enveloped Icosahedral Phages – Double-Stranded RNA (ϕ 6) <i>Paul Gottlieb and Aleksandra Alimova</i>	26
Membrane-Containing Icosahedral DNA Bacteriophages <i>Roman Tuma, Sarah J Butcher, and Hanna M Oksanen</i>	36
Tailed Double-Stranded DNA Phages <i>Robert L Duda</i>	45
Helical and Filamentous Phages <i>Andreas Kuhn and Sebastian Leptihn</i>	53
Replication of <i>Bacillus</i> Double-Stranded DNA Bacteriophages <i>Silvia Ayora, Paulo Tavares, Ruben Torres, and Juan C Alonso</i>	61
Lytic Transcription <i>William McAllister and Deborah M Hinton</i>	69
Lysogeny <i>Keith E Shearwin and Jia Q Truong</i>	77
Decision Making by Temperate Phages <i>Ido Golding, Seth Coleman, Thu VP Nguyen, and Tianyou Yao</i>	88
Mobilization of Phage Satellites <i>Kristen N LeGault and Kimberley D Seed</i>	98
Portal Vertex <i>Peng Jing and Mauricio Cortes Jr.</i>	105

Prohead, the Head Shell Pre-Cursor <i>Marc C Morais and Michael E Woodson</i>	115
Enzymology of Viral DNA Packaging Machines <i>Carlos E Catalano</i>	124
DNA Packaging: DNA Recognition <i>Sandra J Greive and Oliver W Bayfield</i>	136
DNA Packaging: The Translocation Motor <i>Janelle A Hayes and Brian A Kelch</i>	148
Biophysics of DNA Packaging <i>Joshua Pajak, Gaurav Arya, and Douglas E Smith</i>	160
Energetics of the DNA-Filled Head <i>Alex Evilevitch</i>	167
Bacteriophage Receptor Proteins of Gram-Negative Bacteria <i>Sarah M Doore, Kristin N Parent, Sundharraman Subramanian, Jason R Schrad, and Natalia B Hubbs</i>	175
Tail Structure and Dynamics <i>Shweta Bhatt, Petr G Leiman, and Nicholas MI Taylor</i>	186
Bacteriophage Tail Fibres, Tailspikes, and Bacterial Receptor Interaction <i>Mateo Seoane-Blanco, Mark J van Raaij, and Meritxell Granell</i>	194
Phage Genome and Protein Ejection In Vivo <i>Ian J Molineux, L Letti Lopez, and Aaron P Roznowski</i>	206
Dealing With the Whole Head: Diversity and Function of Capsid Ejection Proteins in Tailed Phages <i>Lindsay W Black and Julie A Thomas</i>	219
Jumbo Phages <i>Isaac T Younker and Carol Duffy</i>	229
CRISPR-Cas Systems and Anti-CRISPR Proteins: Adaptive Defense and Counter-Defense in Prokaryotes and Their Viruses <i>Asma Hatoum-Aslan and Olivia G Howell</i>	242
Bacteriophage: Therapeutics and Diagnostics Development <i>Teng-Chieh Yang</i>	252
Bacteriophage Vaccines <i>Pan Tao and Venigalla B Rao</i>	259
Bacteriophage Diversity <i>Julianne H Grose and Sherwood R Casjens</i>	265
Genetic Mosaicism in the Tailed Double-Stranded DNA Phages <i>Welkin H Pope</i>	276
Bacteriophages of the Human Microbiome <i>Pilar Manrique, Michael Dills, and Mark J Young</i>	283
Bacteriophage: Red Recombination System and the Development of Recombineering Technologies <i>Lynn C Thomason and Kenan C Murphy</i>	291
Nanotechnology Application of Bacteriophage DNA Packaging Nanomotors <i>Tao Weitao, Lixia Zhou, Zhefeng Li, Long Zhang, and Peixuan Guo</i>	302
General Ecology of Bacteriophages <i>Stephen T Abedon</i>	314
Marine Bacteriophages <i>Vera Bischoff, Falk Zucker, and Cristina Moraru</i>	322

Ecology of Phages in Extreme Environments <i>Tatiana A Demina and Nina S Atanasova</i>	342
---	-----

Archaeal Viruses

Diversity of Hyperthermophilic Archaeal Viruses <i>David Prangishvili, Mart Krupovic, and Diana P Baquero</i>	359
Euryarchaeal Viruses <i>Tatiana A Demina and Hanna M Oksanen</i>	368
Vesicle-Like Archaeal Viruses <i>Elina Roine and Nina S Atanasova</i>	380
Virus–Host Interactions in Archaea <i>Diana P Baquero, David Prangishvili, and Mart Krupovic</i>	387
Antiviral Defense Mechanisms in Archaea <i>Qunxin She</i>	400
Discovery of Archaeal Viruses in Hot Spring Environments Using Viral Metagenomics <i>Jennifer Wirth, Jacob H Munson-McGee, and Mark J Young</i>	407
Metagenomes of Archaeal Viruses in Hypersaline Environments <i>Fernando Santos, María D Ramos-Barbero, and Josefa Antón</i>	414
Extreme Environments as a Model System to Study How Virus–Host Interactions Evolve Along the Symbiosis Continuum <i>Samantha J DeWerff and Rachel J Whitaker</i>	419

Fungal Viruses

An Introduction to Fungal Viruses <i>Nobuhiro Suzuki</i>	431
Cross-Kingdom Virus Infection <i>Liyang Sun, Hideki Kondo, and Ida Bagus Andika</i>	443
Diversity of Mycoviruses in Aspergilli <i>Ioly Kotta-Loizou</i>	450
Evolution of Mycoviruses <i>Mahtab Peyambari, Vaskar Thapa, and Marilyn J Roossinck</i>	457
Mixed Infections of Mycoviruses in Phytopathogenic Fungus <i>Sclerotinia sclerotiorum</i> <i>Jiatao Xie and Daohong Jiang</i>	461
Mycovirus-Mediated Biological Control <i>Daniel Rigling, Cécile Robin, and Simone Prospero</i>	468
Mycoviruses With Filamentous Particles <i>Michael N Pearson</i>	478
Prions of Yeast and Fungi <i>Reed B Wickner and Herman K Edskes</i>	487
Single-Stranded DNA Mycoviruses <i>Daohong Jiang</i>	493
Structure of Double-Stranded RNA Mycoviruses <i>José R Castón, Nobuhiro Suzuki, and Said A Ghabrial[†]</i>	504

[†]Deceased.

<i>Ustilago maydis</i> Viruses and Their Killer Toxins <i>Alexis Williams and Thomas J Smith</i>	513
Vegetative Incompatibility in Filamentous Fungi <i>Songsong Wu, Daohong Jiang, and Jiatao Xie</i>	520
Viral Diseases of <i>Agaricus bisporus</i> , the Button Mushroom <i>Kerry S Burton and Greg Deakin</i>	528
Viral Killer Toxins <i>Manfred J Schmitt and Björn Becker</i>	534
Alternaviruses (Unassigned) <i>Hiromitsu Moriyama, Nanako Aoki, Kuko Fuke, Kana Takeshita Urayama, Naoki Takeshita, and Chien-Fu Wu</i>	544
Barnaviruses (Barnaviridae) <i>Peter A Revill</i>	549
Botybirnaviruses (Botybirnavirus) <i>Mingde Wu, Guoqing Li, Daohong Jiang, and Jiatao Xie</i>	552
Chrysoviruses (Chrysoviridae) - General Features and Chrysovirus-Related Viruses <i>Ioly Kotta-Loizou, Robert HA Coutts, José R Castón, Hiromitsu Moriyama, and Said A Ghabrial[†]</i>	557
Fungal Partitiviruses (Partitiviridae) <i>Eeva J Vainio</i>	568
Fusariviruses (Unassigned) <i>Sotaro Chiba</i>	577
Giardiavirus (Totiviridae) <i>Juliana Gabriela Silva de Lima, João Paulo Matos Santos Lima, and Daniel Carlos Ferreira Lanza</i>	582
Hypoviruses (Hypoviridae) <i>Dong-Xiu Zhang and Donald L Nuss</i>	589
Megabirnaviruses (Megabirnaviridae) <i>Yukiyo Sato and Nobuhiro Suzuki</i>	594
Mitoviruses (Mitoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	601
Mycoreoviruses (Reoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	607
Mymonaviruses (Mymonaviridae) <i>Daohong Jiang</i>	615
Narnaviruses (Narnaviridae) <i>Rosa Esteban and Tsutomu Fujimura</i>	621
Phlegiviruses (Unassigned) <i>Karel Petrzik</i>	627
Plant and Protozoal Partitiviruses (Partitiviridae) <i>Hanna Rose and Edgar Maiss</i>	632
Quadriviruses (Quadriviridae) <i>Hideki Kondo, José R Castón, and Nobuhiro Suzuki</i>	642
Totiviruses (Totiviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	648

[†]Deceased.

Yado-kari Virus 1 and Yado-nushi Virus 1 (Unassigned) <i>Subha Das and Nobuhiro Suzuki</i>	658
Yeast L-A Virus (<i>Totiviridae</i>) <i>Reed B Wickner, Tsutomu Fujimura, and Rosa Esteban</i>	664
Algal Viruses	
Algal Marnaviruses (<i>Marnaviridae</i>) <i>Marli Vlok, Curtis A Suttle, and Andrew S Lang</i>	671
Algal Mimiviruses (<i>Mimiviridae</i>) <i>Ruth-Anne Sandaa, Håkon Dahle, Corina PD Brussaard, Hiroyuki Ogata, and Romain Blanc-Mathieu</i>	677
Miscellaneous Algal Viruses (<i>Alvernaviridae, Bacilladnaviridae, Dinodnavirus, Reoviridae</i>) <i>Keizo Nagasaki, Yuji Tomaru, and Corina PD Brussaard</i>	684
Phycodnaviruses (<i>Phycodnaviridae</i>) <i>James L Van Etten, David D Dunigan, Keizo Nagasaki, Declan C Schroeder, Nigel Grimsley, Corina PD Brussaard, and Jozef I Nissimov</i>	687
Invertebrate Viruses	
An Introduction to Viruses of Invertebrates <i>Peter Krell</i>	699
Ascoviruses (<i>Ascoviridae</i>) <i>Sassan Asgari, Dennis K Bideshi, Yves Bigot, and Brian A Federici</i>	724
Baculovirus–Host Interactions: Repurposing Host-Acquired Genes (<i>Baculoviridae</i>) <i>A Lorena Passarelli</i>	732
Baculoviruses: General Features (<i>Baculoviridae</i>) <i>Vera ID Ros</i>	739
Baculoviruses: Molecular Biology and Replication (<i>Baculoviridae</i>) <i>Monique M van Oers</i>	747
Bidensoviruses (<i>Bidnaviridae</i>) <i>Qin Yao, Zhaoyang Hu, and Keping Chen</i>	759
Bunyaviruses of Arthropods (<i>Myppoviridae, Nairoviridae, Peribunyaviridae, Phasmaviridae, Phunuiviridae, Wupedeviridae</i>) <i>Sandra Junglen</i>	764
Dicistroviruses (<i>Dicistroviridae</i>) <i>Yanping Chen and Steven M Valles</i>	768
Entomobirnaviruses (<i>Birnaviridae</i>) <i>Marco Marklewitz</i>	776
Hytrosaviruses (<i>Hytrosaviridae</i>) <i>Henry M Kariithi and Irene K Meki</i>	780
Iflaviruses (<i>Iflaviridae</i>) <i>Bryony C Bonning and Sijun Liu</i>	792
Iridoviruses of Invertebrates (<i>Iridoviridae</i>) <i>İkbal Agah İnce</i>	797
Mesoniviruses (<i>Mesoniviridae</i>) <i>Jody Hobson-Peters and Daniel Watterson</i>	804

Nimaviruses (<i>Nimaviridae</i>) <i>Peter Krell and Emine Ozsahin</i>	808
Nodaviruses of Invertebrates and Fish (<i>Nodaviridae</i>) <i>Kyle L Johnson and Jacen S Moore</i>	819
Nudiviruses (<i>Nudiviridae</i>) <i>Yu-Chan Chao, Chih-Hsuan Tsai, and Sung-Chan Wei</i>	827
Parvoviruses of Invertebrates (<i>Parvoviridae</i>) <i>Judit J Péntzes, Hanh T Pham, Qian Yu, Max Bergoin, and Peter Tijssen</i>	835
Polydnaviruses (<i>Polydnaviridae</i>) <i>Anne-Nathalie Volkoff and Elisabeth Huguet</i>	849
Poxviruses of Insects (<i>Poxviridae</i>) <i>Basil Arif, Lillian Pavlik, Remziye Nalçacıoğlu, Hacer Muratoğlu, Cihan İnan, Mehtap Yakupoğlu, Emine Özşahin, İsmail Demir, Kazım Sezen, and Zihni Demirbağ</i>	858
Reoviruses of Invertebrates (<i>Reoviridae</i>) <i>Peter Krell</i>	867
Rhabdoviruses of Insects (<i>Rhabdoviridae</i>) <i>Andrea González-González, Nicole T de Stefano, David A Rosenbaum, and Marta L Wayne</i>	883
Sarothroviruses (<i>Sarothroviridae</i>) <i>Azeez Sait Sahul Hameed</i>	888
Solinviviruses (<i>Solinviviridae</i>) <i>Steven M Valles and Andrew E Firth</i>	892
Tetraviruses (<i>Alphatetraviridae, Carmotetraviridae, Permutotetraviridae</i>) <i>Rosemary A Dorrington, Tatiana Domitrovic, and Meesbah Jiwaji</i>	897

VOLUME 5

Diagnosis, Treatment and Prevention of Virus Infections

Diagnosis

Introduction to Virus Diagnosis and Treatment <i>Maija Lappalainen and Hubert GM Niesters</i>	3
Electron Microscopy for Viral Diagnosis <i>Roland A Fleck</i>	5
Serological Approaches for Viral Diagnosis <i>Klaus Hedman and Visa Nurmi</i>	15
A Brief History of the Development of Diagnostic Molecular-Based Assays <i>Hubert GM Niesters</i>	22
Sequencing Strategies <i>Sibnarayan Datta</i>	27
Validating Real-Time Polymerase Chain Reaction (PCR) Assays <i>Melvyn Smith</i>	35
Rapid Point-of-Care Assays <i>Jan G Lisby and Uffe V Schenider</i>	45

Standardization of Diagnostic Assays <i>Sally A Baylis, C Micha Nübling, and Wayne Dimech</i>	52
Quality Assurance in the Clinical Virology Laboratory <i>Paul Wallace and Elaine McCulloch</i>	64
Biosafety and Biosecurity in Diagnostic Laboratories <i>Hannimari Kallio-Kokko and Susanna Sissonen</i>	82
Screening for Viral Infections <i>Walter Ian Lipkin, Nischay Mishra, and Thomas Briebe</i>	91
Clinical Diagnostic Virology <i>Marcus Panning</i>	98
Virus Diagnosis in Immunosuppressed Individuals <i>Elisabeth Puchhammer-Stöckl and Fausto Baldanti</i>	105
Diagnosis; Future Prospects on Direct Diagnosis <i>Marianna Calabretto, Daniele Di Carlo, Fabrizio Maggi, and Guido Antonelli</i>	112

Treatment

Antiviral Classification <i>Guangdi Li, Xixi Jing, Pan Zhang, and Erik De Clercq</i>	121
Antiretroviral Therapy – Nucleoside/Nucleotide and Non-Nucleoside Reverse Transcriptase Inhibitors <i>Timothy D Appleby and Killian J Quinn</i>	131
Protease Inhibitors <i>Vanesa Anton-Vazquez and Frank A Post</i>	139
HIV Integrase Inhibitors and Entry Inhibitors <i>Daniel Bradshaw and Ranjababu Kulasegaram</i>	145
Management of Respiratory Syncytial Virus Infections (<i>Pneumoviridae</i>) <i>Rachael S Barr and Simon B Drysdale</i>	155
Management of Influenza Virus Infections (<i>Orthomyxoviridae</i>) <i>Bruno Lina</i>	160
Management of Herpes Simplex Virus Infections (<i>Herpesviridae</i>) <i>Nicole Samies and Richard Whitley</i>	175
Management of Varicella-Zoster Virus Infections (<i>Herpesviridae</i>) <i>Andreas Sauerbrei</i>	181
Treatment and Prevention of Herpesvirus Infections in the Immunocompromised Host <i>Sara H Burkhard and Nicolas J Mueller</i>	190
Management of Adenovirus Infections (<i>Adenoviridae</i>) <i>Albert Heim</i>	197
Management of Hepatitis A and E Virus Infection <i>Sébastien Lhomme, Florence Abravanel, Jean-Marie Peron, Nassim Kamar, and Jacques Izopet</i>	206
Management of Patients With Chronic Hepatitis B (<i>Hepadnaviridae</i>) and Chronic Hepatitis D Infection (<i>Deltavirus</i>) <i>Milan J Sonneveld and Suzanne van Meer</i>	217
Studying Population Genetic Processes in Viruses: From Drug-Resistance Evolution to Patient Infection Dynamics <i>Jeffrey D Jensen</i>	227

Virus-Based Cancer Therapeutics <i>Roberto Cattaneo and Christine E Engeland</i>	233
Prevention	
Surveillance of Infectious Diseases <i>Norman Noah</i>	247
Preparing for Emerging Zoonotic Viruses <i>Reina S Sikkema and Marion PG Koopmans</i>	256
Use of Immunoglobulins in the Prevention of Viral Infections <i>Leyla Asadi and Giovanni Ferrara</i>	267
Vaccine Production, Safety, and Efficacy <i>Thomas J Brouwers and Bernard AM Van der Zeijst</i>	281
Vaccines Against Viral Gastroenteritis <i>Scott Grytdal, Tyler P Chavers, Claire P Mattison, Jacqueline E Tate, and Aron J Hall</i>	289
Human Papillomavirus (HPV) Vaccines and Their Impact <i>Jade Pattyn, Pierre Van Damme, and Alex Vorsters</i>	295
Influenza Vaccination <i>Topi Turunen</i>	300
Polio Eradication <i>M Steven Oberste, Cara C Burns, and Jennifer L Konopka-Anstadt</i>	310
Subject Index	315

HUMAN AND ANIMAL VIRUSES

Adenoviruses (*Adenoviridae*)

Balázs Harrach and Mária Benkő, Institute for Veterinary Medical Research, Center for Agricultural Research, Budapest, Hungary

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AdV Adenovirus

bp Base pair

HAdV Human adenovirus

kbp Kilo base pair, 1000 bp

SAdV Simian adenovirus

Glossary

CAR A cellular receptor characterized by its property that Coxsackie virus and many adenovirus types attach to it.

Co-evolution The continuous parallel evolution of a host and its virus, consequently showing similar phylogeny (evolutionary tree topology).

Fiber Antenna like projections from the adenovirus virions, which is used to attach the virus to the cells.

PCR Polymerase chain reaction is a molecular biological method to make many copies of a specific DNA segment.

Phylogenetics The study of the evolutionary history and relationships among groups of organisms.

Squamata The largest order of reptiles, comprising lizards, snakes and amphisbaenians (worm lizards) collectively called squamates or scaled reptiles.

Adenoviruses (AdVs) are middle-sized, nonenveloped, icosahedral, double-stranded DNA viruses of vertebrates. The prefix adeno comes from the Greek word *ádēn* (gland), reflecting the first isolation of a virus of this type from human adenoid tissue 67 years ago. AdVs have since been isolated from many hosts, including representatives of every major vertebrate class from fish to mammals. Thanks to the polymerase chain reaction (PCR), a rapidly increasing number of novel AdVs is being detected in wild life, but isolation and in vitro propagation of such viruses are usually hampered by the lack of appropriate permissive cell cultures. Certain AdVs can cause disease or even death in humans or animals; nonetheless they are mostly harmless in non-immuno-compromised, healthy individuals. AdVs have been used as model organisms in molecular biology, and important findings of general relevance have emerged from such studies, including the phenomenon of splicing in eukaryotes. AdVs have become one of the most popular vector systems for virus-based gene therapy and vaccination and have potential as antitumor tools. Wide prevalence in diverse host species and a substantially conserved genome organization make AdVs an ideal model for studying virus evolution.

Classification

Adenoviruses (AdVs) belong to the family *Adenoviridae*. Very recently, higher taxonomical levels have been introduced: realm *Varidnaviria*, kingdom *Bamfordvirae*, phylum *Preplasmiviricota*, class *Tectiliviricetes*, order *Rowavirales*. This is based on similarities to other viruses and consequently on their anticipated evolutionary history. Thus certain bacterial viruses, namely the 'bacteriophages' in *Tectiviridae*, and also a virus of archaea living in hot springs (Sulfolobus turreted icosahedral virus, *Turriviridae*), and the green alga virus *Paramecium bursaria Chlorella virus 1* (*Phycodnaviridae*) were described as having common evolutionary roots with AdVs. *Adenoviridae*, *Tectiviridae*, *Turriviridae* with bacteriophages classified into *Corticoviridae* are grouped into a common class, and with *Phycodnaviridae* and additional families into a common kingdom based mainly on their double jelly-roll major capsid protein.

There are five official genera and one pending genus in the family. Four genera (*Mastadenovirus*, *Aviadenovirus*, *Ichadenovirus*, and *Testadenovirus*) comprise AdVs that are hypothesized to have likely co-evolved with the host, namely mammals, birds, fish, and turtles (testudines), respectively. The remaining two genera (*Atadenovirus* and *Siadenovirus*) have broader host range. *Atadenoviruses* were named after the bias (toward high A + T) in the genome of their initial representatives, found in various ruminant and avian hosts, as well as in a marsupial-associated sample. However, every known AdV detected in squamate hosts (lizards, snakes, and worm lizards) also belongs to this genus. The base composition of the DNA of the squamate-reptilian AdVs is equilibrated. *Siadenoviruses* were isolated from, or detected by PCR in, birds, a frog, and a few tortoise species. This genus was named after the presence of a gene encoding a putative sialidase; however novel *siadenoviruses*, detected in penguins recently, were found devoid of this gene.

Within each genus, the viruses are grouped into species, which are labeled with the name of the host and the genus, supplemented with letters of the alphabet, e.g., *Human mastadenovirus D* (Table 1). Host origin is only one of several criteria that are used to demarcate the species. Phylogenetic distance is the most significant criterion, with species defined as being separated by more than 10%–15% amino acid sequence divergence of the DNA polymerase by maximum likelihood analysis. Genome organization characteristics is another important criterion, yet additional traits such as the G + C content, cross-neutralization values, biological characteristics including pathogenicity, hemagglutination properties and the ability to recombine, are considered, too. Ideally, all these ancillary data are consistent with the results of the phylogenetic calculations.

Table 1 The taxonomy of family Adenoviridae^a

Mastadenovirus			
Species	Type (abbreviated name)	Strain	Full genome ^a
<i>Bat mastadenovirus A</i>	BtAdV – 3	TJM	+
<i>Bat mastadenovirus B</i>	BtAdV – 2	PPV/1	+
<i>Bat mastadenovirus C</i>	BtAdV – 4	WIV9	+
<i>Bat mastadenovirus D</i>	BtAdV – 7	WIV12	+
<i>Bat mastadenovirus E</i>	BtAdV – 8	WIV13	+
<i>Bat mastadenovirus F</i>	BtAdV – 9	WIV17	+
<i>Bat mastadenovirus G</i>	BtAdV – 11	250-A	+
<i>Bat mastadenovirus H</i>	straw-colored fruit bat AdV	06–106	+
<i>Bat mastadenovirus I</i>	Egyptian fruit bat AdV	3085	+
<i>Bat mastadenovirus J</i>	Asian particolored bat AdV	Vs9	+
unassigned bat AdVs	dozens of bat AdVs		–
<i>Bovine mastadenovirus A</i>	BAdV – 1		+
<i>Bovine mastadenovirus B</i>	BAdV – 3	WBR – 1	+
<i>Bovine mastadenovirus C</i>	BAdV – 10	Ma268	–
<i>Canine mastadenovirus A</i>	CAdV – 1, 2	RI261, Toronto A26/61	+
<i>Deer mastadenovirus B</i>	Odokoileus AdV – 2 (OdAdV – 2)	1319	+
<i>Dolphin mastadenovirus A</i>	bottlenose AdV – 2 (BDAdV – 2)	M1	+
<i>Dolphin mastadenovirus B</i>	bottlenose AdV – 1 (BDAdV – 1)	Tt11018	+
<i>Equine mastadenovirus A</i>	EAdV – 1	M1	+
<i>Equine mastadenovirus B</i>	EAdV – 2	385/75.9	+
<i>Human mastadenovirus A</i>	HAdV – 12, 18, 31, chimpanzee strain ch1		+
<i>Human mastadenovirus B</i>	HAdV – 3, 7, 11, 14, 16, 21, 34, 35, 50, simian AdV – 21 (SAdV – 21ch), 27–28ch/go, 29ch, 32–33ch, 35bo/ch, 41go, 46–47go		+
<i>Human mastadenovirus C</i>	BAdV – 9, HAdV – 1, 2, 5, 6, SAdV – 31ch, 34ch, 40ch, 42bo, 43go, 44bo, 45go		all HAdVs/ SAdVs
<i>Human mastadenovirus D</i>	HAdV – 8–10, 13, 15, 17, 19, 20, 22–30, 32, 33, 36–39, 42–49, 51, + 23 genotypes (mainly recombinants)		+
<i>Human mastadenovirus E</i>	HAdV – 4, SAdV – 22–26ch, 30ch, 36ch, 37bo/ch, 38–39ch		+
<i>Human mastadenovirus F</i>	HAdV – 40, 41		+
<i>Human mastadenovirus G</i>	HAdV – 52, SAdV – 1cy, 2rh, 7rh, 11–12rh, 15rh, 51–53rh		52, SAdV – 1, 2, 7, 11, 51–53
<i>Murine mastadenovirus A</i>	MAdV – 1		+
<i>Murine mastadenovirus B</i>	MAdV – 2	K87	+
<i>Murine mastadenovirus C</i>	MAdV – 3		+
<i>Ovine mastadenovirus A</i>	BAdV – 2, OAdV – 2–5		BAdV – 2
<i>Ovine mastadenovirus B</i>	goat AdV – 2 (GAdV – 2), OAdV – 1		–
<i>Ovine mastadenovirus C</i>	OAdV – 6, 8	8: 7508	8
<i>Platyrrhini mastadenovirus A</i>	titi monkey AdV – 1 (TMAdV – 1)	ECC – 2011	+
unassigned New World monkey AdVs	squirrel monkey AdV – 1, > 19 NWM AdV strains		squirrel monkey – 1
<i>Polar bear mastadenovirus A</i>	PBAdV – 1	BK35	+
<i>Porcine mastadenovirus A</i>	PAAdV – 1–3		3
<i>Porcine mastadenovirus B</i>	PAAdV – 4		+
<i>Porcine mastadenovirus C</i>	PAAdV – 5		+
<i>Sea lion mastadenovirus A</i>	California sea lion AdV – 1 (CSLAdV – 1)	Zc11–030	+
<i>Simian mastadenovirus A</i>	SAdV – 3–4rh, 6rh, 9–10 m, 14rh, 48cy		3, 6, 48
<i>Simian mastadenovirus B</i>	SAdV – 5rh, 8cy, 49–50cy, + 10 untyped rh/cy strains, baboon AdV – 1 (BaAdV – 1)		8, 49, 50, + 10 strains, BaAdV – 1
<i>Simian mastadenovirus C</i>	baboon AdV – 2, 3 (olive baboon, BaAdV), SAdV – 19yb		+
<i>Simian mastadenovirus D</i>	SAdV – 13m	P – 9	+
<i>Simian mastadenovirus E</i>	SAdV – 16gr	C – 8	+
<i>Simian mastadenovirus F</i>	SAdV – 17–18gr	17: B – 105,	+
<i>Simian mastadenovirus G</i>	SAdV – 20gr	ATCC VR – 541	+
<i>Simian mastadenovirus H</i>	SAdV – 54rh	23336	+
<i>Simian mastadenovirus I</i>	SAdV – 55gs	WIV19	+
unassigned simian AdVs	> 14 Old World monkey AdVs		–
<i>Skunk mastadenovirus A</i>	SkAdV – 1	PB1	+

Table 1 Continued

Mastadenovirus			
<i>Species</i>	<i>Type (abbreviated name)</i>	<i>Strain</i>	<i>Full genome^a</i>
<i>Squirrel mastadenovirus A</i>	SqAdV – 1	DE/2013/Sciurus vulgaris/ 2013Pa405–00252	+
<i>Tree shrew mastadenovirus A</i>	TsAdV – 1		+
<i>Guinea pig mastadenovirus A</i> (pending)	GPAAdV – 1	AUS96	+
unassigned prosimian AdVs	> 11 prosimian AdV strains		–
unassigned mastadenoviruses	from alpaca, Asian house shrew, cat, Chinese striped hamster, Daurian ground squirrel, deer mouse, harbor porpoise, rabbit, rat, seal, South West China vole		–
<i>Aviadenovirus</i>			
<i>Duck aviadenovirus B</i>	DAdV – 2	GR	+
<i>Falcon aviadenovirus A</i>	FaAdV – 1		–
<i>Fowl aviadenovirus A</i>	FAdV – 1	CELO	+
<i>Fowl aviadenovirus B</i>	FAdV – 5	340	+
<i>Fowl aviadenovirus C</i>	FAdV – 4, 10	ON1, C – 2B	+
<i>Fowl aviadenovirus D</i>	FAdV – 2, 3, 9, 11	P7-A, 75, A2-A, 380	+
<i>Fowl aviadenovirus E</i>	FAdV – 6, 7, 8a, 8b	CR119, YR36, TR59, 764	+
<i>Goose aviadenovirus A</i>	GoAdV – 1–4	P29	4
<i>Pigeon aviadenovirus A</i>	PiAdV – 1	IDA	+
<i>Pigeon aviadenovirus B</i>	PiAdV – 2	YPDS-Y-V1. A19.11–2013	+
<i>Psittacine aviadenovirus B</i>	PsAdV – 4 (red-bellied parrot)	CS15–4016	+
<i>Psittacine aviadenovirus C</i>	PsAdV – 1 (Senegal parrot)	18VIR149_ITA_2018	+
<i>Turkey aviadenovirus B</i>	TAdV – 1	D90/2	+
<i>Turkey aviadenovirus C</i>	TAdV – 4	CS15–4016	+
<i>Turkey aviadenovirus D</i>	TAdV – 5	1277BT, D1648	+
unassigned turkey adenovirus	TAdV – 2	TAV – 2	–
unassigned aviadenoviruses	from cranes, goldfinch, European greenfinch, great tit, gull, Meyer's parrot, neotropic cormorant, Pacific black duck, smooth-billed ani, tropical screech owl, vitelline masked weaver, white-eyed parakeet		–
<i>Atadenovirus</i>			
<i>Bovine atadenovirus D</i>	BAdV – 4, 5, 8	4: THT/62	4
<i>Bovine atadenovirus E</i> (pending)	BAdV – 6		+
<i>Bovine atadenovirus F</i> (candidate)	BAdV – 7		–
<i>Deer atadenovirus A</i>	Odocoileus AdV – 1 (deer AdV – 1, OdAdV – 1)	CA_AdV_Ohc 98–6943	+
<i>Duck atadenovirus A</i>	DAdV – 1		+
<i>Lizard atadenovirus A</i>	LiAdV – 1, 2	gila monster, Mexican beaded lizard	2
<i>Lizard atadenovirus B</i> (pending)	bearded dragon AdV – 1 (BDAdV – 1)	BD5H2	+
<i>Ovine atadenovirus D</i>	goat AdV – 1 (GAdV – 1), OAdV – 7		7
<i>Possum atadenovirus A</i>	PoAdV – 1		–
<i>Psittacine atadenovirus A</i>	PsAdV – 3 (southern mealy amazon)	HKU/Parrot19	+
<i>Snake atadenovirus A</i>	SnAdV – 1	corn snake, python	+
unassigned snake AdVs	SnAdV – 2, SnAdV – 3	California kingsnake, asp viper, gopher snake	–
unassigned avian atadenoviruses	chimney swift, common chaffinch, eastern spinebill, Eurasian bullfinch, Eurasian siskin, European greenfinch, European robin, long-billed corella, song thrush, tropical screech owl, vitelline masked weaver, white-eyed parakeet, white-plumed honeyeater	eastern spinebill ('passerine AdV-1'): AU2787	eastern spinebill – 1
unassigned reptilian atadenoviruses	from anole, blue-tongued skink, chameleon, East African spiny-tailed lizard, emerald monitor, fat-tailed gecko, Greek tortoise, Japalura tree dragon, long-tailed grass lizard, tokay gecko, tree dragon, white-throated monitor, worm lizard		–
<i>Siadenovirus</i>			
<i>Frog siadenovirus A</i>	FrAdV – 1		+
<i>Great tit siadenovirus A</i>	GTAdV – 1		–
<i>Raptor siadenovirus A</i>	RAAdV – 1 (Harris hawk, eagle owls)		+
<i>Skua siadenovirus A</i>	South Polar skua AdV – 1 (SPSAdV – 1)		+

(Continued)

Table 1 Continued

Mastadenovirus			
<i>Species</i>	<i>Type (abbreviated name)</i>	<i>Strain</i>	<i>Full genome^a</i>
<i>Turkey siadenovirus A</i>	TAdV – 3		+
<i>Penguin siadenovirus A</i>	chinstrap penguin AdV – 2, gentoo penguin AdV – 4		+
Psittacine siadenovirus D (pending)	PsAdV – 5, 6	Pacific parrotlet, budgerigar strain BrdKdnyDNA	+
Psittacine siadenovirus E (pending)	PsAdV – 7	little corella	+
unassigned psittacine siadenoviruses	PsAdV – 2 (plum headed parakeet, cockatoo, eastern rosella, scarlet chested parrot, orange-bellied parrot, etc.)		–
unassigned siadenoviruses	from double-barred finch, Eurasian blackcap, Gouldian finch, pigeon, Sulawesi tortoise, zebra finch		–
<i>Ichtadenovirus</i>			
<i>Sturgeon ichtadenovirus A</i>	white sturgeon AdV – 1 (WSAdV – 1)	WSAdV1/1996	+
Testadenovirus (pending)			
Pond slider testadenovirus A (pending)	red-eared slider AdV – 1 (RESAdV – 1)	2010Z01	–
Box turtle testadenovirus A (cand)	box turtle AdV – 1, eastern box turtle AdV – 1 (EBTAdV – 1)	EBTAdV – 1: R08–207	–
Pancake tortoise testadenovirus A (cand)	pancake tortoise AdV – 1	R08–227	–
unassigned testadenovirus	red-footed tortoise AdV – 1	R13–677	–
Unassigned Viruses in the Family			
	crocodile adenovirus		–

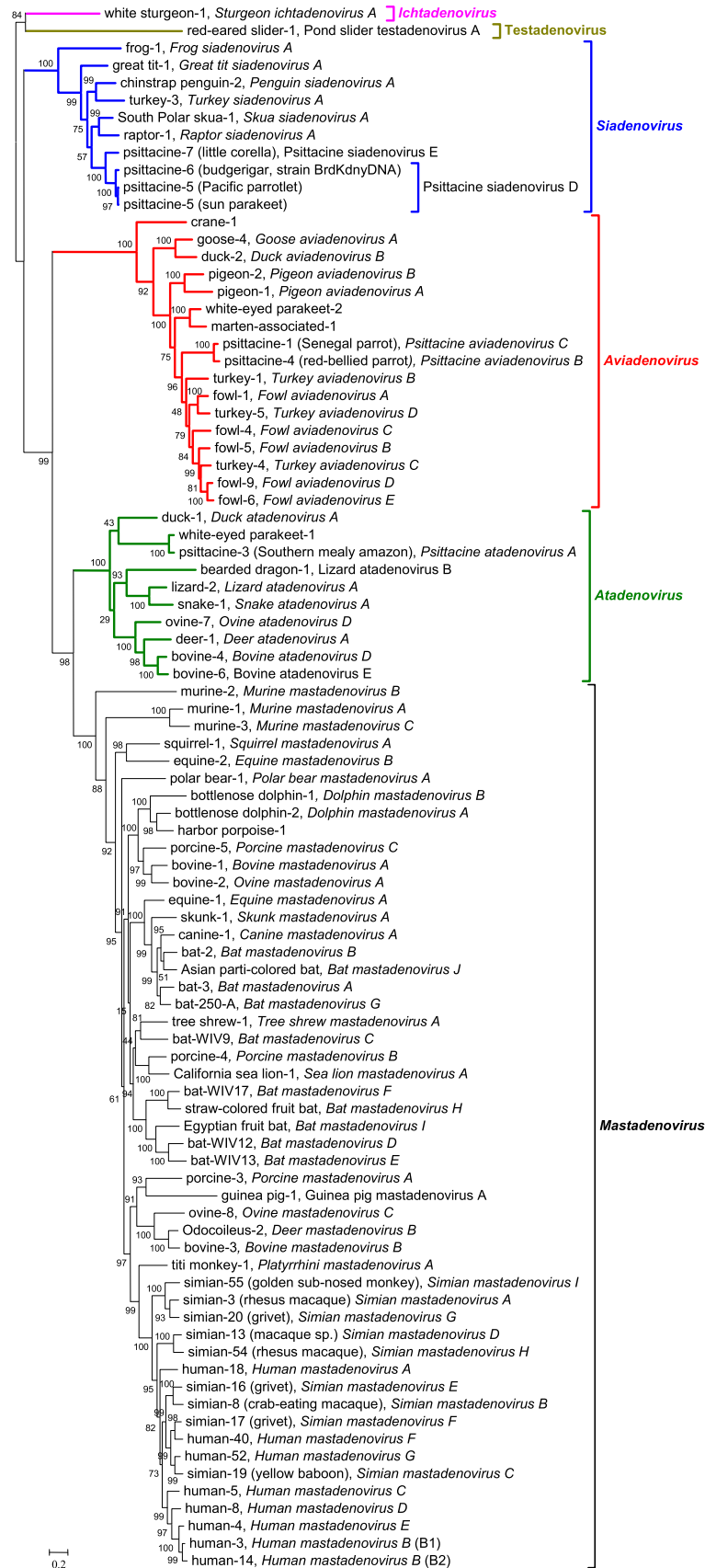
^aAvailable full genome sequences are noted by + or by the type number of the sequenced serotype(s) if those listed are not all sequenced and available. In case of the common name of 'simian AdV', two letter abbreviations show the original hosts: bo, bonobo; ch, chimpanzee; cy, cynomolgus monkey (crab-eating macaque); go, gorilla; gr, grivet; gs, golden snub-nosed monkey; m, macaque sp.; rh, rhesus macaque; yb, yellow baboon.

Along these rules, e.g., chimpanzee AdVs are classified into different human AdV species. Human AdVs have been studied the most extensively and are grouped into seven species (*Human mastadenovirus A* to *Human mastadenovirus G*; abbreviated informally to HAdV–A to HAdV–G). The species are further subdivided to numbered (sero)types, abbreviated in a hyphenated form such as e.g., HAdV–5.

The AdV taxonomy, i.e., clustering similar AdVs together, while separating those that are 'adequately' different, and especially correct naming of the taxa, is prone to human errors and discussions partly because of the lack of knowledge available at a certain time point. As we discover novel AdVs, we find that several AdV types are capable of crossing species barriers of different levels. Thus, some names referring to hosts will not be correct or the best description in the future. Nonetheless, to avoid confusion, it is advisable to keep the historical names based on the first findings. Type classification had been based on serology. In the past decade, however, new type numbers are being assigned to novel AdVs, the full genome sequence of which seems to warrant this. Consequently, HAdVs with type numbers above 52 do not necessarily represent distinct serotypes.

The emerging large quantity of data causes problems also by clearly demonstrating that we will find a very large number of instances of homologous recombination, especially in the case of the best studied human (or in general primate) AdVs. These recombinants have obviously a mixture of biological and genetic characteristics of the 'original' types.

The six clusters of AdVs, defined by gene organization and/or phylogenetic calculations, correspond to the five approved (*Mastadenovirus*, *Aviadenovirus*, *Atadenovirus*, *Siadenovirus*, *Ichtadenovirus*) and the pending sixth genus (*Testadenovirus*) (Fig. 1). The evolutionary distances among the AdVs seem to be generally proportional to those among the main lineages of their vertebrate hosts (in evolutionary order: fishes, turtles, squamates, birds, mammals) supporting the theory on the co-evolution of virus and host. However, there are a few exceptions when very distantly related viruses infect the same host, for example, bovine mastadenoviruses and atadenoviruses infect cattle, whereas the Sulawesi tortoise siadenovirus and testadenoviruses infect testudines (turtles). Obviously, besides co-evolution, several host switches also occurred. According to this hypothesis, the mastadenoviruses, aviadenoviruses, and testadenoviruses represent continuous co-evolving lineages, while squamate AdVs (atadenoviruses) have supposedly switched to ruminants, birds, and common tortoise. The host origin of the lineage of siadenoviruses is presently unknown. Representatives of this genus were found to infect birds, tortoises, and a frog. Similarly, the provenance of the white sturgeon AdV, the single member of the genus *Ichtadenovirus* is not clear either, as no other fish AdV was ever confirmed by targeted examinations or during the aquatic environmental metagenomics studies to date. By phylogenetic analyses, more recent host switches can also be recognized within the genera, for example from Old World monkeys and apes to humans, or from bats to dog-like carnivores, horse, and skunk.



Virion Structure

Adenovirus virions are non-enveloped, icosahedral particles 90 nm in diameter. The capsid consists of 240 nonvertex capsomers (called hexons), each 8–10 nm in diameter, and 12 vertex capsomers (penton bases), each with one or more (2–3) protruding antenna-like structures called fiber(s) 9–77.5 nm in length. The members of genus *Aviadenovirus* usually have two fibers per vertex. The number of fiber structures is not directly connected to the number of fiber genes. Fowl adenovirus 1 (FAdV-1), FAdV-4, and FAdV-10 even have two, tandem fiber genes of different lengths, resulting in two fibers of different sizes at each vertex. Members of species *Human mastadenovirus F* also have two fiber genes of different lengths, but the fiber structures are displayed separately on alternate vertices. Lizard adenovirus 2 (from Mexican bearded lizard) also has two fiber genes of different size. Three long fibers are displayed in some vertices while a single short fiber in the others. The major capsomers (hexons) are formed by the interaction of three identical polypeptides, designated as polypeptide II (the Roman numeral label is related to the relative mobility of structural proteins under reducing conditions in sodium dodecyl sulfate-polyacrylamide gel electrophoresis). Each hexon has two characteristic parts: a triangular top with three ‘towers’, and a pseudo-hexagonal base with a central cavity. Hexons, or more exactly their basal part, are packed tightly to form a protein shell that protects the inner components of the virion.

In members of the genus *Mastadenovirus*, polypeptide IX is located on the outer part of the capsid, keeping together the hexons in each facet (12 copies of protein IX per facet) (Fig. 2). In HAdVs, polypeptide IX also links two facets together across the icosahedral edge. However, polypeptide IX is not present in the members of any other genus. Two monomers of IIIa are located underneath the vertex region, and multiple copies of polypeptide VI are underneath the hexons. Penton bases are formed at the vertices by the interaction of five copies of polypeptide III, and are tightly associated with one, two or three fibers. Fibers are homotrimers of polypeptide IV, and have three structural domains, tail, shaft and the distal knob. There are six monomers of polypeptide VIII per facet: three of them wedged between polypeptide IIIa and the peripentonal hexons, and the other three around the icosahedral 3-fold symmetry axis, contributing with polypeptide IX to the stabilization of the facet. The core consists of the DNA genome complexed with four polypeptides: V, VII, X, and a virus-coded terminal protein (TP). Polypeptide V is found only in mastadenoviruses. Atadenoviruses have two unique structural proteins: p32K and LH3. The major capsid protein, LH3 forms prominent ‘knobs’ on the surface of the atadenoviruses and is located in the same relative position among the hexon subunits as protein IX in the mastadenovirus but sits on top of the central hexon trimers (probably holding the outer capsid together). A virus-coded protease is also packaged bound to the genome.

Adenoviruses are stable on storage in the frozen state. They are insensitive to lipid solvents. Heat sensitivity varies among the genera.

Genome

The adenovirus genome is a single linear molecule of double-stranded DNA (24,630–48,395 bp) containing inverted terminal repeats (ITR) of 26–721 bp at its termini. The 5′ ends of each DNA strand are linked covalently to the TP. The range of nucleotide composition is 33.24%–66.92% G + C. The genetic organization of the central part of the genome is conserved throughout the family, whereas the distal parts show large variations in length and gene content (Fig. 3).

Splicing was first discovered in AdVs, and is a common means of expressing mRNAs in this virus family. In the conserved region, most of late genes are expressed by splicing from the rightward-oriented major late promoter, located in overlap with the DNA-dependent DNA polymerase (pol) gene. The early genes encoding pol, the precursor of TP (pTP), and the DNA-binding protein (DBP) are expressed by splicing from leftward-oriented promoters.

Life Cycle

Replication of various human adenoviruses has been studied in detail, in particular with HAdV-2. Virus entry takes place via interactions of the fiber knob with specific receptors (most often the coxsackievirus and adenovirus receptor or CAR) on the surface of a susceptible cell followed by internalization via interactions between the penton base and cellular α_v integrins. Protein VI mediates the release of virions from the endosomes, allowing dynein-mediated transport on microtubules to nuclear pores. After uncoating, the virus core is delivered to the nucleus, i.e., the site of virus transcription, DNA replication and assembly. Virus infection mediates the shutdown of host DNA synthesis and later RNA and protein synthesis. Transcription of the AdV genome by host RNA polymerase II involves both DNA strands of the genome and initiates (in HAdV-2) from five early (E1A, E1B, E2, E3, and E4), two intermediate (IX and IVa2), the major late (L) and the U exon protein (UXP) promoters. All primary transcripts are capped and polyadenylated, with complex splicing patterns producing families of mRNAs. In primate adenoviruses, there are one

Fig. 1 Phylogenetic tree of adenoviruses based on maximum likelihood analysis of DNA-dependent DNA polymerase (pol) amino acid sequences from adenoviruses one from each species and pending species, or distinct strain with full pol sequence known. Multiple alignment: MultAlin; model selection: ProtTest 2.4; maximum likelihood calculation: PhyML 3.1 with model LG + I + G and Shimodaira-Hasegawa-like branch test for statistical test for branch support on the platform of Galaxy/Pasteur. Unrooted calculation; white sturgeon AdV-1 and slider AdV-1 are selected as outgroup for visualization. From the name of the adenovirus types the word ‘adenovirus/AdV’ was removed for clarity. In case of the common names of ‘psittacine AdV’ and ‘simian AdV’, the original hosts are shown in brackets. Adenovirus species are indicated. The bar indicates 20% difference between two neighboring sequences. SH (Shimodaira-Hasegawa) branch support values shown at the nodes.

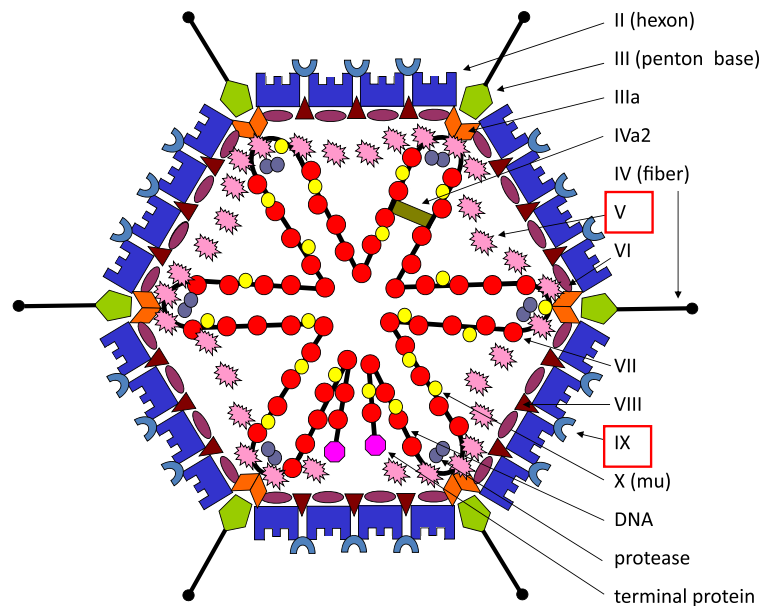


Fig. 2 Stylized section of a mastadenovirus particle. Capsid proteins (II, III, IIIa, IV, VI, VIII and IX) and core proteins (V, VII, X and terminal protein /TP/) are labeled by Roman numbers related to the relative mobility of structural proteins under reducing conditions in sodium dodecyl sulfate-polyacrylamide gel electrophoresis). Protein V and IX (in red squares) occur only in mastadenoviruses. Instead of protein IX, a structural protein called LH3 can be found in almost all adenoviruses. As the structure of the nucleoprotein core has not been established, the polypeptides associated with the DNA are shown in hypothetical locations. Adapted from Harrach, B., Benkő, M., Both, G.W., *et al.*, 2011. Family *Adenoviridae*, In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Classification and Nomenclature of Viruses*. Ninth Report of the International Committee on Taxonomy of Viruses. San Diego, CA: Elsevier, pp. 125–141.

or two virus-associated (VA) RNA genes. These are transcribed by cellular RNA polymerase III and facilitate translation of late mRNAs and block the cellular interferon response. In non-primate mastadenoviruses, no VA RNA genes could be identified. A non-homologous VA RNA gene has been mapped in some aviadenoviruses near the right end of the genome. The replication of aviadenoviruses has been shown to involve significantly different pathways from those characterized in HAdVs. This is not unexpected, given the considerable differences in gene layout between the non-conserved regions of the genome.

About 40 different polypeptides (the largest number being in aviadenoviruses and ichtadenovirus, the smallest in siadenoviruses) are produced. Almost a third of these compose the virion, including a virus-encoded cysteine protease, which processes a number of precursor proteins (these are prefixed with p). With the exception of polypeptides V and IX, the other structural proteins found in mastadenoviruses are conserved in every genus. Products of the four early regions of mastadenoviruses modulate the host cell's transcriptional machinery (E1 and E4), assemble the virus DNA replication complex (E2), and subvert host defense mechanisms (E3). The E2 region (encoding DBP, pol, and pTP) is well conserved throughout the family, while the E3 and E4 regions show great variation in length and gene content even among the mastadenoviruses. E1A, E1B 19K, and the E3 and E4 regions (with the exception of 34K, called ORF6 in HAdVs) occur only in mastadenoviruses. Genes encoding proteins related to 34K are also present in adenoviruses, usually in duplicate. Intermediate (IX, only in mastadenoviruses, and IVa2) and late gene products (in many mastadenoviruses expressed from five transcription units, L1–L5) are concerned with assembly and maturation of the virion. Late proteins include 52K (scaffolding protein) and pIIIa (L1); III (penton base), pVII (major core protein), V (minor core protein, only in mastadenoviruses) and pX (L2); pVI, hexon (II), and protease (L3); 100K, 33K (and its unspliced version, 22K), and pVIII (L4); and fiber (IV) (L5). Seemingly, there are no lipids in adenovirus particles. However, the fiber proteins and some of the nonstructural proteins are glycosylated.

Epidemiology

Adenoviruses are abundant and widespread and theoretically every reptilian, avian and mammalian species could host at least one AdV type. Endemic or enzootic infections that are frequently unapparent seem to be very common. In contrast, epidemic or epizootic occurrence is relatively rare but normally has much more significant consequences. The non-enveloped virions, shed by infected individuals, are rather stable in the environment and can preserve infectivity for long periods. The virions are not only able to tolerate adverse environmental effects, such as drought and moderate temperature or pH changes, but are also resistant to lipid solvents and simple disinfectants. There is evidence that HAdVs can persist and perhaps even retain infectivity in natural or communal water sources. Moreover, freshwater and marine bivalves have also been found to accumulate AdVs. Because of their

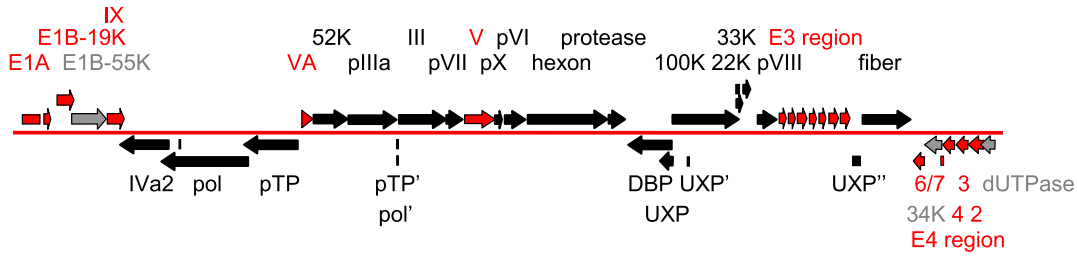
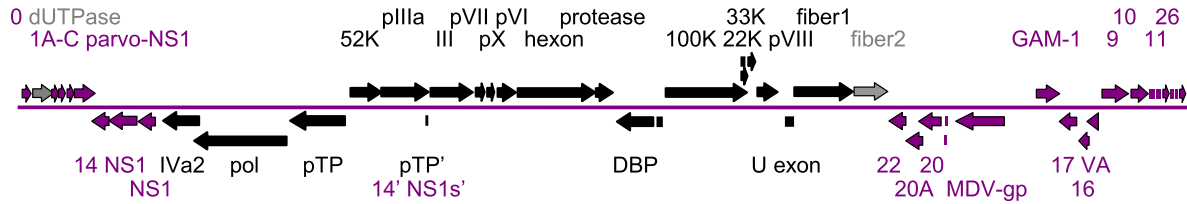
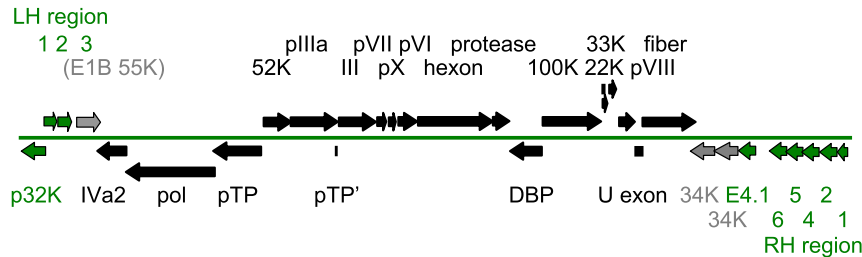
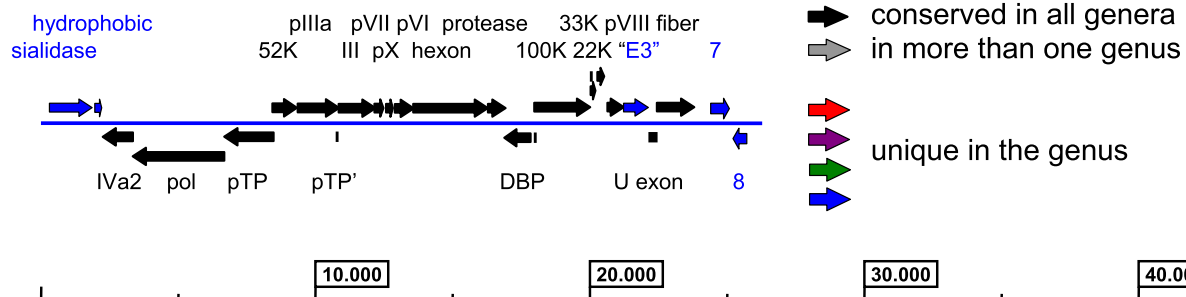
Mastadenovirus (human adenovirus 2)**Aviadenovirus** (fowl adenovirus 1)**Atadenovirus** (ovine adenovirus 7)**Siadenovirus** (frog adenovirus 1)

Fig. 3 Schematic illustration of the different genome organizations found in representative members of the four thoroughly analyzed adenovirus genera. Black arrows depict genes conserved in every genus, gray arrows show genes present in more than one genus, and colored arrows show genus-specific genes. Adapted from Harrach, B., Benkő, M., Both, G.W., *et al.*, 2011. Family *Adenoviridae*, In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Classification and Nomenclature of Viruses*. Ninth Report of the International Committee on Taxonomy of Viruses. San Diego, CA: Elsevier, pp. 125–141. with permission.

stability, AdVs have become a major indicator of viral pollution in the environment. Vertical transmission has also been documented in several host species, including birds and cattle.

Although different AdV types have different affinities for various organs and tissues, primary virus shedding is likely through the intestines. AdVs are often transmitted by person-to-person contact, particularly among young children where fecal–oral spread is also common. Aerosol transmission is probably not rare in crowded populations. Adenovirus DNA may be found in tonsillar tissue, peripheral blood lymphocytes, and lung epithelial cells long after clinical disease has abated. Swimming-pool-related outbreaks, particularly by strains causing keratoconjunctivitis or pharyngitis, are not uncommon. In poultry and other farm animals, transportation, crowding, and mixing of different populations can lead to mass infections accompanied by disease outbreaks.

Clinical Features

The tropism and pathogenicity of HAdVs are largely species- and type-dependent. Respiratory pathogens are common, although infections of a large variety of different organs have been described. The fiber protein mediates primary tissue tropism, and some of the specific determinants, epitopes, and receptors have been mapped. The most common clinical manifestations connected to members of each human AdV species are listed below. Members of *Human mastadenovirus B* are divided into two groups informally termed 'subspecies' HAdV-B1 and HAdV-B2 since they represent two phylogenetic lineages (Fig. 1).

HAdV-A

HAdV-12, HAdV-18, and HAdV-31 can replicate efficiently in the intestines, and, based on serological surveys, are common in the population, especially in children with gastrointestinal disease. However, their role in the etiology of infant diarrhea is yet to be determined. Serotypes belonging to this species are generally difficult to isolate and culture.

HAdV-B1

Respiratory pathogens HAdV-3, HAdV-7, HAdV-16, and HAdV-21 belong here (and HAdV-50). Seasonal outbreaks of febrile respiratory disease, mainly during winter, can be caused in infants by HAdV-7 in most parts of the world. Similar diseases, usually with a less severe outcome, can be seen among school children. In some countries, HAdV-3 is the dominant serotype, whereas HAdV-7 seems to have a number of different genotypes that shift occasionally in certain geographic areas. In addition to HAdV-7 and HAdV-21, HAdV-4 (species HAdV-E), and HAdV-14 (HAdV-B2) are most often implicated in the etiology of acute respiratory outbreaks among freshly enlisted military recruits in the USA.

HAdV-B2

HAdV-11, HAdV-34, and HAdV-35 cause persistent interstitial infection in the kidney and hemorrhagic cystitis. These HAdV types, along with HAdV-50, are most often shed in the feces or urine of acquired immune deficiency syndrome (AIDS) patients and organ or tissue transplantation recipients. Re-emergence of a genomic variant of HAdV-14 causing severe, sometimes fatal respiratory infections was reported first from the USA in 2006, later also detected in Europe and China.

HAdV-C

The low serotype designations (HAdV-1, HAdV-2, HAdV-5, and HAdV-6) of viruses in this species reflect the relative ubiquity of these viruses. These HAdVs can be isolated easily and, indeed, comprise approximately half of all HAdV serotypes reported to the World Health Organization. HAdV-1, HAdV-2, and HAdV-5 are known to maintain endemic infections, and most teenagers will have had infections with more than one serotype. The site of persistence is lymphoid tissue, and shedding can last for a couple of years after primary infection.

HAdV-D

This species contains 32 serotypes, and therefore encompasses well over one half of all known HAdVs. HAdV-8, HAdV-19, and HAdV-37 cause epidemic keratoconjunctivitis, especially in dry climates or densely populated areas. Other serotypes are rarely isolated except from immunocompromised patients. Virus shedding by asymptomatic healthy individuals was found to be common in Africa. Recently, by whole genome analysis, a large number of double or multiple HAdV-D intertypic recombinants, involved in epidemic keratoconjunctivitis, have been recognized. Several recombinant strains were assigned to novel type number (genotypes) without testing or confirming their serological distinctness.

HAdV-E

This species contains a single human serotype, namely HAdV-4, although different variants ('genotypes') of this virus have been described. HAdV-4 strains have been most often implicated in respiratory diseases among military recruits in the USA. The species encompass a number of novel isolates originating from chimpanzees and bonobos. In Japan, HAdV-4 is the second most important cause of adenovirus-associated eye disease after HAdV-8.

HAdV-F

The so-called enteric or fastidious HAdV-40 and HAdV-41 are classified here. The genetic distance between these two serotypes comes close to meriting separation into different species, but, for practical reasons, including indistinguishable pathology, they remain in a single species. These viruses are a major cause of infantile diarrhea all over the world. In Europe, a shift in dominance

from HAdV-40 in the 1970s to HAdV-41 after 1992 has been observed. Because of a deletion in the E1 region near the left end of the genome, HAdV-40 and HAdV-41 can be isolated and cultured only in complementing, transformed cell lines such as 293 or PER.C6.

HAdV-G

The most recently described HAdV serotype, HAdV-52, was recovered from a patient with diarrhea. No further cases or serologic evidence on the circulation of this virus have been reported yet. Interestingly, this species encompasses numerous monkey AdVs besides HAdV-52.

Epizootiology

Simian Adenoviruses

A series of simian viruses (termed SVs) were isolated from various tissue cultures prepared from apes and monkeys. These were numbered serially, irrespective of the virus family to which they belong, since identification and allocation to a family were performed later only. A considerable number of the SV isolates have been identified as adenoviruses. In many cases, unfortunately, the original SV numbers have been retained as the numbers of the AdV type, and thus a somewhat confusing system of SAdV numbering can still be encountered in publications, as well as in the records of the American Type Culture Collection. To date, there are > 55 simian adenovirus types (SAdV-1 through SAdV-55) and dozens of detected but not well characterized SAdVs. SAdV-1 to SAdV-20 originated from Old World monkeys, whereas SAdV-21 to SAdV-47, and further strains were isolated from apes (chimpanzees, bonobos, and gorillas). SAdV-48 to SAdV-55, and further variants are again from Old World monkeys. In the past years, novel adenoviruses have been described from New World monkeys. In one case, the possible transmission of the virus from diseased animals to the caretakers was also observed. Very recently, a dozen different prosimian AdVs were detected by PCR screening of samples directly from Madagascar and from zoos. Data concerning prevalence or pathogenicity of AdVs in their natural hosts are increasing continuously. An interesting mixing and grouping can be observed among the primate AdVs, suggesting that they can sometimes be less host specific and supposedly able to infect individuals of closely related host species.

Canine and Other Carnivoran Adenoviruses

For a long period, only canine AdVs had been known to infect carnivores. CAdV-1 is the causative agent of infectious canine hepatitis (Rubarth's disease), a life-threatening disease of puppies. Regular vaccination worldwide has decreased the number of clinical cases. Viruses, serologically indistinguishable from CAdV-1, have been found to cause encephalitis in foxes as well as infections in numerous other carnivorous animals including wolves, coyotes, jackals, otters, raccoons, and even bears. The disease caused by a genetically closely related virus, CAdV-2, is called kennel cough and is common among dog breeder stocks. There exist a couple of publications reporting cases of inclusion-body-hepatitis-like condition in felids. Yet, focused attempts to find a distinct feline adenovirus yielded only a single case of cat AdV amplification and partial sequencing from paraffin embedded sample. It is likely, however, that cats living in close proximity with humans can harbor HAdV-1 or HAdV-5 from species HAdV-C. More recently, specific adenoviruses have been described from hosts, namely seals and several sea lions, belonging to the superfamily of pinnipeds. Similarly, a distinct AdV had been found in polar bears died in zoos. Phylogeny inference and genome analyses imply that the AdVs isolated from, or detected in, some carnivorous animals (CAdV-1, CAdV-2, skunk AdV-1) have a close common ancestor with the AdVs discovered in different bats recently. This would explain the elevated pathogenicity of these AdVs in the newly invaded hosts. Indeed, skunk AdV was discovered from independent deadly diseases of pet African hedgehogs in Japan and the USA, in a marmoset in a Hungarian zoo and in North American porcupines.

Porcine Adenoviruses

At least five porcine adenovirus types (PAdV-1 to PAdV-5) are recognized. These represent three different species. Species *Porcine mastadenovirus A* contains three serotypes (PAdV-1 to PAdV-3) that are fairly similar to each other and cause no specific diseases. PAdV-4 (in species PAdV-B) has been described as associated with neurological disease, whereas PAdV-5 (in species PAdV-C) is more distantly related to PAdV-A. The pathogenic roles of porcine adenoviruses need further investigation.

Equine Adenoviruses

Two equine adenovirus serotypes (EAdV-1 and EAdV-2) have been isolated to date. EAdV-1 seems to be more prevalent and causes clinical respiratory diseases. It seems to be genetically close to certain bat AdVs so that a host switch has been suspected. On

the other hand, EAdV-2 appears to be the AdV co-evolved with horses. It usually causes gastrointestinal infection with mild or self-limiting disease in foals. Certain Arabian horse lineages carry a genetic defect that, when present in the homozygous state, results in severe combined immunodeficiency disease. Affected animals are incapable of mounting an immune response, and foals often die of pneumonia due to equine adenovirus infection, which would be generally harmless in immunocompetent individuals.

Bovine, Ovine, Caprine, and Other Ruminant Adenoviruses

Ruminant animals can harbor rather diverse AdVs. Several isolates originating from cattle, sheep, goat and deer species belong to various species in the genus *Mastadenovirus*. Unusual BAdVs, substantially differing from mastadenoviruses, as judged for example by the absence of the genus-specific antigen, were described four decades ago. These BAdV serotypes are now classified in a separate genus, *Atadenovirus*. One ovine (OAdV-7), one goat (GAdV-1), and one deer adenovirus (Odocoileus AdV-1) also belong here.

Bovine mastadenoviruses were recognized early on and were found occasionally to cause enzootic bronchopneumonia or calf pneumo-enteritis. However, experimental infection of young calves with the isolated adenovirus seldom, if at all, resulted in reproduction of disease. The early isolates of BAdV-10 originated from diseased or dead animals. In phylogeny reconstructions, BAdV-10 represents an independent lineage. It also has a very simple E3 region and a variable fiber gene. Therefore, it has been proposed to be in the process of switching from one host to another. Crossing the host barrier has also been suggested for BAdV-2, with two subtypes, each infecting cattle and sheep, respectively. Based on evolutionary relationships, BAdV-2 is presently classified into species *Ovine mastadenovirus A* together with serotypes OAdV-2 to OAdV-5. The newest isolated OAdV type is mastadenovirus OAdV-8.

The atadenovirus *Odocoileus AdV-1* caused hemorrhagic epizooty killing thousands of mule deer in California.

Avian Adenoviruses

Birds are known to harbor AdVs from three different genera.

In addition to the host origin, members of the genus *Aviadenovirus* can be distinguished from mastadenoviruses by the lack of the genus-common complement-fixing antigen. Twelve aviadenovirus serotypes have been described from chicken. A specific condition, referred to as hydropericardium syndrome, has been clearly associated with fowl adenovirus 4 (FAdV-4) from species *Fowl aviadenovirus C*. It occurs mainly in Asia causing great economical losses. Another specific disease the so-called gizzard erosion is caused by FAdV-1 (species FAdV-A). Infection with at least ten additional FAdV types may cause inclusion body hepatitis in chicken.

Novel aviadenoviruses have been isolated from, or detected by PCR and sequencing in, numerous host species including falcons, finches, gulls, parrots, pigeon, neotropic cormorant, Pacific black duck, and in unpublished cases from many other wild birds (Table 1).

The genus *Atadenovirus* contains duck adenovirus 1 (DAdV-1), the causative agent of the so-called egg drop syndrome (EDS) of chickens. The disease, causing dramatic decrease in the egg production of laying flocks, was first experienced in Europe in 1976 then soon became known worldwide. Retrospective serological studies confirmed the presence of hemagglutination inhibitory antibodies to EDS virus in archive serum samples originating from a large number of wild and domestic bird species, with predominance in waterfowl, which are now considered as the main reservoir. The presence of novel atadenoviruses has been reported in parrots, songbirds, chimney swift, finches, robin and tropical screech owl (Table 1). Additional yet unpublished atadenoviruses have been detected in various wild birds.

Another poultry pathogen, namely turkey adenovirus 3 (TAdV-3), is classified into the genus *Siadenovirus*. The distinctness of this virus, also known as turkey hemorrhagic enteritis virus, has been recognized early based on the lack of common cross-reacting antigen with aviadenoviruses or DAdV-1. Serologically indistinguishable strains have been associated with various pathological entities in turkey, pheasant, and chicken. Besides TAdV-3, the genus *Siadenovirus* encompasses the single known adenovirus from an amphibian host, frog adenovirus 1 (FrAdV-1). Recently, numerous novel siadenoviruses have been demonstrated by PCR and sequencing in different diseased or dead birds, including finches, parrots, pigeons, owls and a hawk, as well as penguins and skuas from the Antarctica. Psittacine AdV-2 seems to have high pathogenicity and is capable of crossing the species barrier. It has been detected in representatives of more than seven different species belonging to psittaciformes.

Reptilian Adenoviruses

Adenovirus infections in various snake and lizard species are reported frequently, but the number of virus isolates is still limited. In Germany, captive boid snakes, from different collections and breeders, seem to be infected by the same type of AdV, and also by a parvovirus (snake adeno-associated virus). Bearded dragons, a reptile pet of growing popularity, have a specific adenoviral infection and disease often manifesting in neurological signs, the so-called 'star-gazing'. The AdVs found in squamate reptiles, belong to genus *Atadenovirus*, and at least 25 distinct atadenovirus types, detected by PCR and sequencing, have been reported. An unusual virion morphology with triple fiber projection per vertices was discovered during the characterization of an AdV isolate

from Mexican bearded lizard. This virus was also found to have two fiber genes, a trait that usually broadens the host-cell range during infection.

In a dead specimen of the common (Greek) tortoise an adenovirus was detected supposedly as a result of host switching.

Most recently, novel AdVs representing a thus far unknown lineage, proposed as an additional genus (Testadenovirus), have been reported from testudines, including red-eared slider, yellow-bellied slider, box turtle, eastern box turtle, pancake tortoise, and red-footed tortoise.

An interesting case of siadenovirus infection in confiscated Sulawesi tortoises has been described. Significant fatalities and transmission of this virus to other turtles has also been reported. The pathogenic role of the adenoviral infections in reptiles needs further studies nonetheless the introduction of the practice of quarantining is highly recommended.

Frog and Fish Adenoviruses

Data concerning adenovirus infection of aquatic vertebrates are scarce. In fact, only a single AdV has been isolated from each of the two host classes, amphibians and fish. Interestingly, the frog AdV was isolated from a renal tumor of a leopard frog on a reptilian cell line (TH1) prepared from turtle heart tissue. FrAdV-1 has a genome organization similar to that of TAdV-3, and the common evolutionary origin of these two viruses in genus *Siadenovirus* is supported by phylogenetic calculations. The only adenovirus isolate available from fish was obtained from a healthy adult specimen of an ancient Chondrostei species, the white sturgeon. Experimental transmission of the virus failed to produce any clinical signs. Genomic characterization of the sturgeon adenovirus implied the need for creating a new genus, *Ichtadenovirus*. Nuclear inclusion bodies typical of adenoviruses and adenovirus-like particles have been observed by light and electron microscopy, respectively, in damaged tissues of a couple of sea fish species, but molecular confirmation of AdV infection has not been successful. Some of these suspected viruses turned out to have genes similar to those in polyomavirus, but having larger and different genomes thus being candidate for members of a new family, still to be proposed officially (Adomaviridae).

Pathogenesis

Adenovirus infections of man generally occur in childhood, and the outcome varies in severity from asymptomatic to explosive outbreaks of upper or lower respiratory tract manifestations. Less commonly, AdVs cause gastrointestinal, ophthalmic, urinary, and neurological diseases. The vast majority of adenovirus-caused diseases are self-limiting. However, immunocompromised patients, above all organ transplant recipients, individuals infected with human immunodeficiency virus and developing AIDS, and those receiving radiation and chemotherapy against tumors, represent special populations that are prone to experience grave, frequently fatal consequences of AdV infection. In numerous cases, the organ to be transplanted itself proves to be the source of invasive AdVs. Sporadic fatal infections may occasionally occur in healthy, immunocompetent individuals. In such cases, the presence of certain predisposing immunogenetic factors cannot be excluded. Cellular immune responses are also important for the recovery from acute adenovirus infection. Peripheral blood mononuclear cells have been found to exhibit proliferative responses to HAdV-2 antigen. This function is mediated by CD4⁺ T cells, which seem to recognize conserved antigens among different HAdV serotypes.

The incubation period from infection to clinical signs is estimated to be 1–7 days and may be dose dependent. Clinical symptoms during the initial viremia are dominated by fever and general malaise. Recovery from infection is associated with the development of serotype-specific neutralizing antibodies that protect against disease or reinfection with the same serotype.

Pathology by AdVs is partially the consequence of viral replication and cell lysis. Correspondingly, in various tissues and organs, such as bronchial epithelium, liver, kidney, and spleen, disseminated necrotic foci can be observed upon necropsy and histopathological examination. Characteristic intranuclear inclusion bodies and so-called smudge cells contain large amounts of adenovirus capsid proteins. Besides the lesions caused by virus replication, direct toxic effects of high doses of structural proteins, as well as the host's inflammatory response, may contribute to the aggravation of pathology.

Experimental infection of animals with various AdV types seldom results in pathology similar to that experienced with the same virus under natural circumstances. One of the few exceptions is turkey hemorrhagic enteritis, the pathogenesis of which has been studied in detail. TAdV-3 causes intestinal hemorrhages and immunosuppression. By in situ DNA hybridization and PCR, the presence of TAdV-3-specific DNA as evidence for virus replication has been demonstrated in the immunoglobulin M-bearing B lymphocytes and macrophage-like cells, but not in CD4⁺ or CD8⁺ T lymphocytes. Interestingly, fewer virions were present in the intestines, which are the principal site of pathology, than in the neighboring lymphoid organs including spleen and cecal tonsils. This finding strongly suggests that the intestinal lesions induced by TAdV-3 are mediated by the immune system. Systemic or intestinal hemorrhagic disease of ruminants seems to be related to virus replication in endothelial cells.

Diagnosis

With the development of modern techniques, especially PCR and direct DNA sequencing, the number of AdVs detected in various organ samples of human or animal origin has increased rapidly. However, there is no official agreement or convention on the

criteria that would be prerequisites for the approval of new AdV types. The conventional methods, including virus isolation in tissue culture, raising antisera, and performing a large set of cross-neutralization tests, are cumbersome, and the majority of medical and veterinary diagnostic laboratories do not possess appropriate prototype strain and serum collections. Full genomic sequences should validate new types even in the absence of isolated virus strains. However, the value of short sequences from PCR fragments is still a topic of debate. In principle, PCR and sequencing should be able to replace serotyping if appropriate targets are identified.

Human medical laboratories use commercially available tests, such as complement fixation and enzyme immunoassay, to detect adenovirus-specific antibodies that cross-react with all serotypes. Nearly all adults have serologic evidence of past infection with one or more AdV types. For the detection of viral DNA by PCR, the most common target was initially the gene encoding the major capsid protein, the hexon. With subsequent restriction enzyme digestion, typing systems for HAdVs and FAdVs have also been elaborated. Recently, a nested PCR method, targeting the most conserved part of the adenovirus-coded DNA-dependent DNA polymerase gene, became very popular. The highly degenerate consensus primers seem to be capable of facilitating amplification of DNA from every adenovirus, irrespective of its genus affiliation.

Treatment

No specific anti-adenovirus therapy has yet been established. Recent advances in understanding the pathophysiology of fulminant adenovirus diseases in immunocompromised patients have prompted the consideration of applying donor lymphocyte infusions after transplantation.

Cidofovir is a monophosphate nucleotide analog that, after undergoing cellular phosphorylation, competitively inhibits incorporation of dCTP into virus DNA by the viral DNA polymerase. Incorporation of the compound disrupts further chain elongation. Efforts to improve its effect are in progress. There are a growing number of positive experiences with using this drug to combat adenovirus infections.

Successful disinfection of poultry houses and stables can be performed with the use of chlorine-releasing agents, iodophors or quaternary ammonium compounds.

Prevention

In the USA, orally administrable, live, enteric-coated vaccines against HAdV-4, HAdV-7, and HAdV-21 were used in military units for a couple of decades. The vaccine went out of production in 1996 and the supplies were exhausted by 1999. Soon, re-emergence of HAdV-7 and especially HAdV-4 was verified and about 12% of all recruits were affected by adenovirus disease. The vaccination program, restored at the military in October 2011, caused a prompt 75% reduction in the incidence of febrile respiratory illness among the new trainees.

In the veterinary practice, dog vaccination schedules all over the world invariably include a live or killed CAdV-2 component against infectious canine hepatitis. Inactivated vaccine for horses against equine adenoviruses has been prepared in Australia. In farm animals, inactivated bivalent vaccines (containing one mastadenovirus and one atadenovirus BAdV type) have been in use in several countries for controlling enzootic calf pneumonia or pneumo-enteritis. In poultry practice, commercially available or experimental vaccines for the prevention of EDS or turkey hemorrhagic enteritis are applied occasionally. There are several attempts ongoing for the production of recombinant subunit vaccines, which should be safer than vaccines derived from tissue culture or the organs of infected birds.

Adenoviruses as Vectors

HAdV-5 has been engineered and used most extensively as a gene delivery vector with a view to applications in gene therapy, immunization and antitumor medicine. Other human adenoviruses such as HAdV-3, -4, -26, -35, -41, -48, -49, etc. (either replication-competent or replication-deficient) or even nonhuman serotypes (OAdV-7, CAdV-2, chimpanzee and rhesus monkey AdVs), are being tested as a means of overcoming the problem posed by preexisting neutralizing antibodies in the human population, and also to achieve targeting to specific organs and tissues. Bovine, porcine, canine, and fowl AdV types have also been tested as novel antigen delivery vectors for immunizing animals of the respective species.

Further Reading

- Alonso-Padilla, J., Papp, T., Kaján, G.L., *et al.*, 2016. Development of novel adenoviral vectors to overcome challenges observed with HAdV-5 based constructs. *Molecular Therapy* 24, 6–16.
- Condezo, G.N., San Martín, C., 2017. Localization of adenovirus morphogenesis players, together with visualization of assembly intermediates and failed products, favor a model where assembly and packaging occur concurrently at the periphery of the replication center. *PLoS Pathogens* 13, e1006320.

- Crenshaw, B.J., Jones, L.B., Bell, C.R., Kumar, S., Matthews, Q.L., 2019. Perspective on adenoviruses: epidemiology, pathogenicity, and gene therapy. *Biomedicines* 7, e61.
- Doszpoly, A., Harrach, B., LaPatra, S., Benkő, M., 2019. Unconventional gene arrangement and content revealed by full genome analysis of the white sturgeon adenovirus, the single member of the genus *Ichtadenovirus*. *Infection, Genetics and Evolution* 75, 103976.
- Gilson, T., Blanchette, P., Ballmann, M.Z., *et al.*, 2016. Using the E4orf6-based E3 ubiquitin ligase as a tool to analyze the evolution of adenoviruses. *Journal of Virology* 90, 7350–7367.
- Greber, U.F., Flatt, J.W., 2019. Adenovirus entry: From infection to immunity. *Annual Review of Virology* 6, 177–197.
- Hiwarkar, P., Kosulin, K., Cesaro, S., *et al.*, 2018. Management of adenovirus infection in patients after haematopoietic stem cell transplantation: state-of-the-art and real-life current approach. A position statement on behalf of the infectious diseases working party of the European Society of Blood and Marrow Transplantation. *Reviews in Medical Virology* 28, e1980.
- Hoeben, R.C., Uil, T.G., 2013. Adenovirus DNA replication. *Cold Spring Harbor Perspectives in Biology* 5, a013003.
- Lenman, A., Liaci, A.M., Liu, Y., *et al.*, 2018. Polysialic acid is a cellular receptor for human adenovirus 52. *Proceedings of the National Academy of Sciences of the United States of America* 115, E4264–E4273.
- Mennechet, F.J.D., Paris, O., Ouoba, A.R., *et al.*, 2019. A review of 65 years of human adenovirus seroprevalence. *Expert Review of Vaccines* 18, 597–613.
- Needle, D.B., Selig, M.K., Jackson, K.A., *et al.*, 2019. Fatal bronchopneumonia caused by skunk adenovirus 1 in an African pygmy hedgehog. *Journal of Veterinary Diagnostic Investigation* 31, 103–106.
- Pénzes, J.J., Menéndez-Conejero, R., Condezo, G.N., *et al.*, 2014. Molecular characterization of a lizard adenovirus reveals the first atadenovirus with two fiber genes and the first adenovirus with either one short or three long fibers per penton. *Journal of Virology* 88, 11304–11314.
- Podgorski, I.I., Pantó, L., Földes, K., *et al.*, 2018. Adenoviruses of the most ancient primate lineages support the theory on virus – host co-evolution. *Acta Veterinaria Hungarica* 66, 474–487.
- San Martín, C., 2012. Latest insights on adenovirus structure and assembly. *Viruses* 4, 847–877.
- Singh, A.K., Nguyen, T.H., Vidovszky, M.Z., *et al.*, 2017. Structure and *N*-acetyl glucosamine binding of the distal domain of murine adenovirus 2 fibre. *Journal of General Virology* 99, 1494–1508.

Relevant Websites

- <https://sites.google.com/site/adenoseq>
Adenovirus sequences.
Google Sites.
- <https://www.visual-science.com/projects/adenovirus>
Human adenovirus.
Scientific illustration.
Visual Science.
- https://talk.ictvonline.org/files/proposals/animal_dna_viruses_and_retroviruses
Pending Proposals.
International Committee on Taxonomy of Viruses (ICTV).
- <https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=10508>
Adenoviridae.
NCBI.
NIH.
- <https://talk.ictvonline.org/taxonomy>
Virus Taxonomy.
International Committee on Taxonomy of Viruses (ICTV).

African Horse Sickness Virus (*Reoviridae*)

Piet A van Rijn, Wageningen Bioveterinary Research, Lelystad, The Netherlands and North-West University, Potchefstroom, South Africa

© 2021 Elsevier Ltd. All rights reserved.

Classification

African horse sickness virus (AHSV) is a virus species within the genus *Orbivirus* belonging to the subfamily Sedoreovirinae within the *Reoviridae* family. Other orbivirus species are equine encephalosis virus, also infecting horses, and bluetongue virus (BTV) and epizootic hemorrhagic disease virus (EHDV) both infecting instead ruminants. AHSV causes African Horse Sickness (AHS), which is a notifiable disease according to the World Organization of Animal Health (OIE). BTV is the prototype orbivirus and shares morphology and many molecular and virological properties with AHSV. Nine serotypes of AHSV have been recognized.

Virion Structure and Genome

AHSV is a non-enveloped virus of 70–80 nm in diameter with an icosahedral structure. The viral genome of double-stranded (ds) RNA is about 19 kb in total and fragmented in ten genome segments (Seg-1 to Seg-10) varying in length from 0.8 to 3.9 kb. The segmented genome encodes seven structural proteins VP1–VP7 and at least four non-structural proteins NS1, NS2, NS3/NS3a, and NS4. Structural proteins VP2, 3, 5 and 7 form a three-layered virus particle harboring enzymatic proteins VP1, VP4 and VP6 in addition to one copy of each dsRNA genome segment. The inner protein layer consists of 120 VP3 molecules. VP3 displays two conformations in dimers and are clustered around a five-fold axis and form a replication complex together with the enzymatic proteins while VP3 decamers forms a subcore particle in an orbicular configuration. The subcore forms the core particle by a layer of 780 copies of VP7. A third protein layer, which is the outer shell, consists of VP2 and VP5 and determines the serotype of the infectious virus particle. Serotypes have been recognized by specific neutralizing antibodies and reflect genetic differences observed in Seg-2, which encode serotype dominant VP2. Serotypes 1 and 2, 3 and 7, and 5 and 8 induce some cross-neutralizing antibodies and display a higher percentage of genetic homology between each other compared to other serotypes. Non-structural proteins are not part of the virus particle but are expressed in the infected cells and support virus replication.

Life Cycle

AHSV infects cells of the insect vector as well as of the mammalian host. After entering the cell, the outer shell is removed and the uncoated virus, the core particle, is released in the cytoplasm. Messenger RNAs (mRNAs) are synthesized inside the transcriptionally active core particle from all ten dsRNA genome segments and secreted into the cytoplasm. In this way, the viral genome is not exposed to cellular defense mechanisms triggered by dsRNA. Translation of viral mRNAs is specifically enhanced by the NS1 protein. Later after infection, mRNAs are recruited from the cytoplasm into replication complexes in virus inclusion bodies (VIBs) formed by the NS2 protein. Pre-subcore particles are formed within VIB, and initiate the replication of recruited single-stranded RNAs into dsRNA genome segments. The mechanism by which exactly one copy of each RNA segment is recruited into one virus particle is not completely understood but likely occurs according to a “follow-the-leader” process initiated by specific intersegmental RNA-RNA interactions and macromolecule interactions. After formation of core particles and assembly of the outer shell in a late stage, the progeny virus is released from the infected cell through an NS3/NS3a mediated process using cellular transport mechanisms and/or by budding on the cell membrane causing a cytopathogenic effect (CPE). Virus release from culicoides cells is dependent on NS3/NS3a protein and indeed infected culicoides cells show no CPE.

In order to complete its life cycle, replication and transmission of AHSV between equines and midges are crucial. Midges are infected by feeding on viremic equines. Saliva, gut components, viral proteins, virus dose, and likely the gut microbiome influence AHSV infection of gut cells. Progeny virus is released into the hemolymph and, after dissemination throughout the midge, several organs are infected. Infection of salivary glands leads to virus release into saliva and transmission of AHSV to susceptible equines by every next blood meal. The interplay between virus and midge is very specific, since only a very few culicoides species are competent midges able to transmit AHSV to susceptible animals. Saliva, as well as trauma by bites, presumably enhance the infection of the equine host. In conclusion, the life cycle of AHSV requires replication in both the competent midge and susceptible equine hosts and at least two subsequent blood feedings by competent midges on equines.

Epidemiology

AHS is an arthropod-borne noncontagious disease of *Equidae*. AHS is endemic in tropical sub-Saharan Africa where one or more serotypes at any given time are circulating asymptotically in zebra and the African donkey. In subtropical regions,

AHS displays a seasonal occurrence with the first cases appearing in late summer while the disease declines or even disappears in the autumn due to lower temperatures and the reduced midge population. Occasionally, the affected areas can expand by changes of climate and environmental conditions. In the past, AHS outbreaks have also been reported outside Africa in the South-East Asia and the Iberian Peninsula. Local spread of AHS is always associated with the presence of competent infected biting midges spreading by wind and active flight. The most relevant competent midges for AHSV in Africa are *Culicoides imicola* and *Culicoides bolitinos*. Usually, midges become persistently infected and are lifelong virus source for onward transmission. In contrast, susceptible equine hosts are thought to be infectious only temporarily. African wildlife is an important AHSV reservoir. For example, infected zebra can be viremic for up to 40 days as opposed to domestic horses, if they survive the infection, are usually viremic for 2–3 weeks. Movement of viremic equines over long distances can lead to new epidemics but these depend on the presence of local competent midges for onward transmission. For instance, the last recorded European AHS epidemic occurred in Spain (in 1987–90) and was caused by the transmission of AHSV to local horses by indigenous *Culicoides imicola* which fed on imported viremic zebra. In the 21st century, the distribution of several midge-borne orbiviruses and their vectors has been dramatically expanding both to the south and north in different parts of the world. Presumably, climate change and globalization will further expand habitats of potential competent midges. *Culicoides sonorensis* in the Americas and *Culicoides brevitarsis* in Australia are known BTV transmitters but their vector competence for AHSV is unknown. The largest recorded Bluetongue outbreak in Europe in 2006–2009 was unexpectedly spread by indigenous midge species belonging to the *Culicoides obsoletus* group. Accordingly, the epidemiology of AHS could drastically change in the future and might include introduction and expansion of the disease to historically AHS-free regions with a moderate climate.

Clinical Features

The incubation period in horses after natural AHSV infection is 5–9 days, whereas after experimental infection the incubation period is in general 3 days but depends on the dose and route of virus infection. The prognosis for infected horses is extremely poor as the mortality rate is 70%–95%. The fatality rate for African donkeys and zebra is very low (5% to 10%), whereas approximately 50% of infected mules will survive AHSV infection. There are four classical clinical forms of AHS in equines. The peracute, pulmonary form occurs in fully susceptible animals and has a short course. Infected horses with this form will rarely recover. The rise in body temperature starts after 3–5 days and can reach 40–41°C. This disease form is marked by rapidly progressive respiratory failure leading to a higher breathing frequency and further supported by abdominal heave lines. The horse appears exhausted and can stand with the forelimbs spread with an extended head and neck, profuse sweating, and dilated nostrils. Uncontrollable coughing with frothy, serofibrinous fluid exuding from the nostrils is common in the terminal stage of the disease. The respiratory distress usually starts suddenly a few hours before death. The subacute, cardiac, edematous form starts with rising in body temperature to 39–41°C about 7 days after infection and lasts in general for a couple of days. The fatality rate of this form is about 50%. The head may show severe swelling, including eyelids, lips, cheeks, tongue, intermandibular space, and laryngeal region. A characteristic symptom of this form is the swelling of the supraorbital fossae that may include conjunctival swelling with petechiae. The swellings may extend to the neck down to the chest. Remarkably, ventral edema and edema of the lower limbs are not seen. Paralysis of the esophagus may result in aspiration pneumonia and sublingual hemorrhages. The horse becomes very depressed and may lie down frequently for short periods. In the end, the horse dies by heart failure in about two weeks after onset of clinical signs. Horses that recover show gradually reducing swellings. The most commonly observed mixed, acute form displays signs of both the pulmonary and edematous forms with mortality rates around 70%. Mostly, signs are observed at post-mortem examination of fatal cases. Mild pulmonary signs seem to be not progressive, whereas edematous swellings and fluid in the body cavities seem to result in heart failure and death. Still, other typical signs of the pulmonary form are also observed in cases of the mixed form. The fourth form is horse sickness fever, which is the mildest form of the disease and is often missed. The incubation period is long (9 days) and the rise in body temperature can slowly reach 40°C. Clinical signs are not obvious. This mild form is usually seen in resistant or partially protected equines such as zebra, donkeys, or previously vaccinated horses. Horses that recover from natural AHSV infection develop lifelong immunity for the respective serotype and partial protection against closest related serotypes. Foals of immune mares show passive immunity for 3–6 months depending on the amount of AHSV directed antibodies received by uptake of colostrum.

Pathology

The pulmonary form shows the most characteristic post-mortem changes such as edema of the lungs resulting in foam production in the air pipe and bronchia and hydrothorax (Fig. 1).

Extensive alveolar edema and particolored hyperemia of the lungs are seen in peracute cases, whereas extensive interstitial edema is observed for more prolonged courses. The lungs may appear reasonably normal, but hydrothorax is common and the thoracic cavity may contain several liters of fluid. Less frequent are periaortic and peritracheal edematous infiltration. Cardiac lesions are not obvious, but epicardial and endocardial petechial hemorrhages are sometimes observed. Particularly, the lymph nodes in the thoracic and abdominal cavities are edematous and enlarged. Regarding the gastrointestinal tract, hyperemia of the

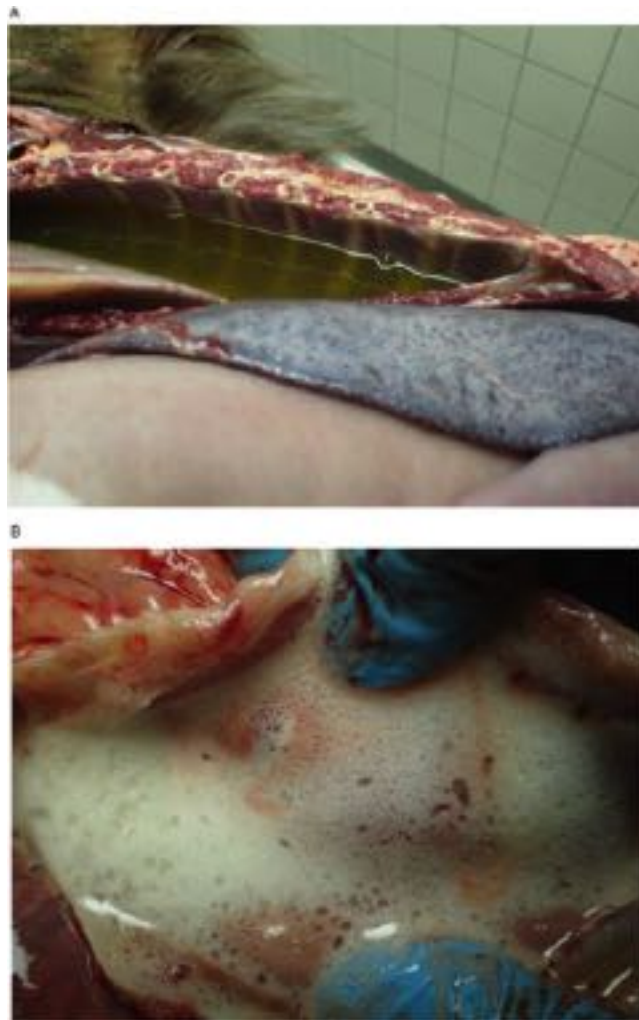


Fig. 1 Severe characteristic post-mortem observations in a horse which died as a result of AHSV infection. (A). Hydrothorax. (B). Formation of foam in the air pipe and bronchia.

glandular fundus of the stomach, hyperemia and petechial hemorrhages in the mucosa and serosa of the small and large intestines, subcapsular hemorrhages in the spleen, and congestion in the renal cortex are reported. The cardiac form typically shows yellow, gelatinous infiltration in the subcutaneous and intermuscular fascia, primarily of the head, neck, and shoulders. Infrequently, lesions may be observed in the chest, ventral abdomen, and rump. Hydropericardium is common, with extensive petechial and ecchymotic hemorrhages on the epicardium and endocardium. In contrast to the pulmonary form, lung edema is not present or very mild, and hydrothorax is rare. In addition to those found for the pulmonary form, submucosal edema of the cecum, large colon, and rectum is more pronounced. The mixed, acute form of the disease shows a combination of those found in the pulmonary and cardiac forms.

Diagnosis

Clinical diagnosis is not very specific especially in the early phase of AHS. However, additional factors support a presumptive diagnosis, such as other AHS suspicions or lethal confirmed AHS cases in the vicinity of the suspected horse. In a later phase of the disease, once characteristic clinical signs typical for AHS develop, differentiation from other equine diseases also improves clinical diagnosis. Specific findings by post-mortem examination further increase the differential diagnosis. AHS can spread very rapidly and causes huge economic and enormous socio-emotional impact by its severity. Therefore, AHS is a notifiable disease according to the World Organization of Animal Health (OIE). Any suspicion of AHS must be promptly reported to national veterinary authorities. The suspected horses should be correctly diagnosed with laboratory diagnostics as soon as possible in order to quickly install effective, feasible, and acceptable control measures. Further, the suspected horse must be protected against bites of midges to prevent further spread of disease by housing and use of horse blankets, repellents, and/or insecticides. The veterinary authority

will advise in this and further arrange and submit blood samples to an authorized laboratory, often the national reference laboratory for AHS. Preferably, all sick and healthy equines on the premises should be blood sampled. In the case of deceased horses, samples of spleen, lung lymph nodes and samples of any other organs with obvious pathology should be collected too, refrigerated and transported on ice for diagnostic purposes. Many highly sensitive and specific laboratory diagnostic tools for AHS have been developed and are operational in authorized laboratories to quickly confirm or exclude AHSV infection. Real-time reverse transcriptase polymerase chain reaction (PCR) assays for AHSV have been developed and are operational in diagnostic laboratories as first-line diagnostic tool. Several genome segments have been targeted which also opens the possibility for confirmatory PCR assays. Confirmation could be also performed with Seg-2 based PCR assays in order to identify the virus serotype. The latter can be supportive to find the original virus source, and is important in case vaccination against the spreading of a specific AHSV serotype is considered. In addition, virus isolation is performed to further characterize and investigate the causing virus, including serotyping with virus neutralization tests using reference sera against each of the nine serotypes. However, virus isolation is not routinely used for diagnostic purposes because this method is less sensitive, laborious and time-consuming. At one week after AHSV infection or later, antibodies against AHSV are detected by ELISA specific for VP7 antibodies and confirms AHSV infection. High throughput ELISAs are routinely used and are highly sensitive and specific detecting all nine AHSV serotypes. Neutralizing antibodies are detected later after infection with serum neutralization tests using nine reference strains of AHSV serotypes. Neutralizing antibodies are in majority serotype specific and directed to VP2.

Treatment, Control, and Prevention

There is no adequate treatment for AHS, and appropriate supportive therapy can be provided based on the observed clinical signs. AHSV infections must be detected and controlled as soon as possible. Importantly, control measures should prevent the spread of AHS to disease-free areas, countries or horse populations. Enforcing quarantine periods, mandatory protocols for clinical inspections and test procedures in combination with insect control measures prior to allowing import of equines are described for international equine transports over long distances. However, local expansion of an affected area is hard to stop due to local transport, movement of animals and, more importantly, spread by competent midges.

National authorities in AHS endemic, as well as in AHS-free countries, together with national reference laboratories, horse industries, owner organizations, and veterinary services have developed contingency plans describing control of AHS according to recommendations of OIE and WHO. In general terms, the suspected horse is temporarily contained with respect to access of midges, and official authorities must be informed. After confirmation by laboratory diagnosis, control measures as described in contingency plans are instantly implemented. Follow-up control measures regarding confirmed cases and (suspected) horses on the same premise are under debate and vary between national contingency plans. For example, infected and/or suspected horses are euthanized or receive supportive therapy while insect control strategies are undertaken. All contingency plans describe the installation of different zones: the affected zone around the affected holding (3 km diameter), surrounded by the protective zone (100 km diameter) and the surveillance zone of an additional 50 km radius around the protective zone. Actions are different for these zones and vary in different national contingency plans. Movement of equines from the affected zone is prohibited and is restricted or allowed in combination with negative test results, quarantine period and vector control measures for the other zones. The temporary or permanent housing of equines, removal of midge breeding places, and use of repellents, insecticides, and horse blankets are also described in contingency plans. Clinical inspections and surveillance are intensified in all zones. These control measures are adopted and specified to the local or national situation with regard to feasibility, such as the capacity of accommodation/stables, possibilities for insect control, the behavior of competent midges (if known), and the number of equine wildlife and domestic horses. Furthermore, contingency plans describe the possibility of vaccination in order to prevent disease and eventually eradicate AHS. In most cases, the decision by authorities to start vaccination campaigns in the affected area or country depends on the success of already undertaken actions, the feasibility, and of course, the availability of suitable AHS vaccine.

Vaccination is the most effective way to protect horses and to eradicate AHS outbreaks. Still, success by vaccination strongly depends on the profile (efficacy, safety, costs, acceptance) of the used vaccine. Currently, conventionally live-attenuated vaccines (LAV) are available for AHS in several African countries. Subsequent vaccinations with affordable cocktail vaccines for multiple serotypes protect horses against all nine serotypes. LAVs can, however, induce viremia and can spread to, and by, midges. Indeed, virulent variants (partially) derived from LAVs by reassortment events and/or by reversion to virulence have been reported. The Spanish AHS outbreak has been eradicated, first by use of LAV, and later by inactivated AHS vaccine which is more expensive but safer. Both vaccine types do not enable DIVA (differentiating infected from vaccinated individuals) and complicate eradication programs, and safe trade of equines is strongly hampered by this shortcoming.

Experimental vector- and subunit vaccines are completely safe and DIVA vaccines in combination with VP7 based ELISAs. Protection is mainly based on immune responses against outer shell proteins. Further, several virus-based vaccine platforms have been developed by genetic modification and applied for all nine serotypes by exchange of outer shell proteins. Consequently, virus-based vaccines induce full-blown immune responses against almost all viral proteins. Disabled infectious single cycle/cell (DISC) vaccine lacks essential protein VP6 and non-essential NS4. In effect, DISC vaccine is produced by *in trans* complementation on cells constitutively expressing VP6 protein. DISC vaccination leads to an abortive infection, since no progeny infectious virus can be produced in equines. AHSV without non-essential NS4 or NS3/NS3a are attenuated, protective and similarly produced as LAVs. AHSV without NS3/NS3a does not cause viremia and does not propagate in midges and is therefore named disabled

infectious single animal (DISA) vaccine. Furthermore, the DISA vaccine platform is DIVA compatible with the experimental NS3/NS3a based ELISA and a DIVA PCR test. All these experimental (vector) vaccine platforms, including easy-to-produce deletion mutants of AHSV, are very promising but are classified as genetically modified organisms and regulatory hurdles hamper their use at the moment. In conclusion, there is a desperate need for improved AHS vaccines in order to reduce the economic and socio-emotional impact caused by this devastating viral disease of horses.

Further Reading

- Alberca, B., Bachanek-Bankowska, K., Cabana, M., *et al.*, 2014. Vaccination of horses with a recombinant modified vaccinia Ankara virus (MVA) expressing African horse sickness (AHS) virus major capsid protein VP2 provides complete clinical protection against challenge. *Vaccine* 32, 3670–3674.
- Erasmus, B.J., 1973. The pathogenesis of African Horsesickness. In: *Proceedings of the 3rd International Conference on Equine Infectious Diseases*. Basel, Paris: Karger. pp. 1–11.
- Erasmus, B.J., 1976. A new approach to polyvalent immunisation against African horse sickness. In: *Proceedings of the Fourth International Conference on Equine Infectious Diseases*, Lyon, France, September 1976. Princeton, N.J.: Veterinary Publications, pp. 401–403.
- Guthrie, A.J., Quan, M., 2009. African horse sickness. In: Mair, T.S., Hutchinson, R.E. (Eds.), *Infectious Diseases of the Horse*. Fordham: Equine Veterinary Journal Ltd, pp. 72–82.
- Lulla, V., Losada, A., Lecollinet, S., *et al.*, 2017. Protective efficacy of multivalent replication-abortive vaccine strains in horses against African horse sickness virus challenge. *Vaccine* 35, 4262–4269.
- Lulla, V., Lulla, A., Wernike, K., *et al.*, 2016. Assembly of replication-incompetent African horse sickness virus particles: Rational design of vaccines for all serotypes. *Journal of Virology* 90, 7405–7414.
- Mertens, P.P.C., Maan, S., Samuel, A., Attoui, H., 2005. Orbiviruses, Reoviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier/Academic Press, pp. 466–483.
- OIE (World Organisation for Animal Health), 2010. African horse sickness. In: Fernandez, P.J., White, W.R. (Eds.), *Atlas of Transboundary Animal Diseases*. Paris: OIE, pp. 12–18.
- Sanchez-Vizcaino, J.M., 2004. Control and eradication of African horse sickness with vaccine. In: Schudel, A., Lombard, M. (Eds.), *Control of Infectious Diseases by Vaccination, Developments in biologicals*, vol. 119. Basel: S. Karger AG, pp. 255–258.
- van de Water, S.G., van Gennip, R.G., Potgieter, C.A., Wright, I.M., van Rijn, P.A., 2015. VP2 exchange and NS3/NS3a deletion in African horse sickness virus (AHSV) in development of Disabled Infectious Single Animal vaccine candidates for AHSV. *Journal of Virology* 89, 8764–8772.
- van Rijn, P.A., Maris-Veldhuis, M.A., Boonstra, J., van Gennip, R.G.P., 2018a. Diagnostic DIVA tests accompanying the Disabled Infectious Single Animal (DISA) vaccine platform for African horse sickness. *Vaccine* 36, 3584–3592.
- van Rijn, P.A., Maris-Veldhuis, M.A., Potgieter, C.A., van Gennip, R.G.P., 2018b. African horse sickness virus (AHSV) with a deletion of 77 amino acids in NS3/NS3a protein is not virulent and a safe promising AHS Disabled Infectious Single Animal (DISA) vaccine platform. *Vaccine* 36, 1925–1933.
- Weyer, C.T., Grewar, J.D., Burger, P., *et al.*, 2016. African horse sickness caused by genome reassortment and reversion to virulence of live, attenuated vaccine viruses, South Africa, 2004–2014. *Emerging Infectious Diseases* 22, 2087–2096.

African Swine Fever Virus (*Asfarviridae*)

Linda K Dixon, Rachel Nash, Philippa C Hawes, and Christopher L Netherton, The Pirbright Institute, Pirbright, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ASFV African swine fever virus
ATF4 Activating transcription factor 4
BER Base excision repair
Bcl-2 B-cell lymphoma 2
CHOP C/EBP homologous protein or GADD153
CSF Classical swine fever
CSFV Classical swine fever virus
dpi days post-infection
dTTP deoxy thymidine triphosphate
ELISA Enzyme-linked immunosorbent assay
ER Endoplasmic reticulum
EU European Union
GADD34 Growth arrest and DNA damage-inducible protein 34
g/cm³ grams per cubic centimetre
IAP Inhibitor of apoptosis protein
ICP34.5 Infected cell protein 34.5
IFN Interferon
IL1 Interleukin 1

kbp kilobase pairs
kDa kilodaltons
MGF Multigene family
ml millilitre
NF-κB Nuclear factor kappa B
NFAT Nuclear factor of activated-T cells
mg milligram
nm nanometre
NTPase Nucleoside triphosphatase
OIE World Organisation for Animal Health
ORF Open reading frame
PCR Polymerase chain reaction
RNAP RNA polymerase
ROS Reactive oxygen species
SUMO Small Ubiquitin-like Modifier
TBP Tata box binding protein
TFIIB Transcription factor IIB
TFS Transcription factor S
VLTf Vaccinia virus late transcription factor
VETfL Vaccinia virus early transcription factor

Glossary

Apoptosis A highly regulated form of programmed cell death.
Capsomers Capsomers are subunits of the capsid, an outer covering of protein that protects the virus genetic material.
Clathrin-mediated endocytosis A vesicular transport event that facilitates the internalization of molecules including receptors. This is involved in various processes including signal transduction, receptor recycling, virus entry. The process involves clathrin, a tri-skeleton shaped scaffold protein.
C-type lectin A type of carbohydrate-binding protein domain known as a lectin. The C-type lectin require calcium for binding. C-type lectin have diverse functions including regulation of cell adhesion, immune responses and apoptosis.
Entry fusion complex The entry fusion complex is a multi-component protein complex required for fusion of the virus envelope with cellular membranes to enable virus entry.
ERCC4-like nuclease An enzyme involved in DNA recombination or repair.
Genotypes This term is used to describe genetic variants of a given microorganism, in this case African swine fever virus.
Filopodia Slender cytoplasmic projections that contain actin filaments cross-linked into bundles by actin-binding proteins.

Guanyl transferase Three enzymes (RNA triphosphatase, guanyl transferase and methyl transferase) which are involved in the addition of the methylated 5' cap to mRNA.
Immunogold electron microscopy Immunogold labelling is a technique which involves binding of antibodies attached to gold particles to proteins. These complexes can then be detected by electron microscopy to localise specific proteins.
Macropinocytosis Is a form of uptake of materials including solutes from the extracellular fluid. The uptake mechanism is non-specific and taken advantage of by many pathogens. Once inside cells macropinosomes mature through endosomes.
Multigene families Multigene families are groups of genes from the same organism that encode proteins with similar sequences over all or part of the sequence.
Myristoylation Involves covalent attachment of a myristoyl group by an amide bond to an alpha amino group of an N-terminal glycine residue. Commonly this can act as a membrane anchor for proteins.
Sylvatic cycle The sylvatic cycle is the fraction of the pathogen population's lifespan spent cycling between wild animals and vectors.

African Swine Fever Virus

Classification

The *Asfarviridae* family consists of a single genus *Asfivirus* which contains one member: African swine fever virus (ASFV). The virus contains a double-stranded DNA genome varying in length between 170 and 193 kbp with terminal cross-links and inverted terminal

repeats. The predominantly cytoplasmic replication cycle and genome structure resembles that of the *Poxviridae* although the icosahedral capsid structure differs. African swine fever virus encodes enzymes and factors required for replication and transcription in the cytoplasm, in common with other large DNA virus families that have a replication stage in the cytoplasm. Genome sequencing has shown that ASFV is only distantly related to other viruses but shares the highest number of conserved genes with several viruses that infect amoeba, including Faustovirus, kaumoebavirus and Pacmanvirus. These giant viruses have genome sizes of greater than 400 kbp and share approximately 30 genes with ASFV including genes involved in replication, transcription and virus morphogenesis.

Virion Structure

ASFV has a multi-layered structure consisting of a nucleoprotein core, 70–100 nm in diameter, inside a protein core shell which is surrounded by an internal lipid layer (Fig. 1). The icosahedral capsid is assembled on this internal envelope. These intracellular mature particles are 170–190 nm in diameter and gain an additional membrane as they bud from the plasma membrane of the infected cell to form the extracellular enveloped virus that is 175–215 nm in diameter. Both the intracellular mature and extracellular enveloped forms of the virus are infectious. The capsid show icosahedral symmetry (T- 189–217) corresponding to 1892–2172 capsomers each observed as a hexagonal prism with a central hole. The distance between capsomers is 7.4–8.1 nm.

Mass spectrometry analysis of purified virus particles has identified 68 virus encoded proteins of which more than 20 have no known function. About 20 additional proteins in the virus particle were identified as host proteins. ASFV encodes for two polyproteins, pp220 (CP2475L) and pp62 (CP530R) which are proteolytically cleaved, by a virus-encoded cysteine protease (S273R). Mature peptide products of molecular weight 150, 37, 34, 14, and 5 kDa from pp20 and 35, 15, and 8 kDa from pp62 are incorporated into the core shell. The main virus capsid protein, p72 (B646L), represents about 10% of the mass of the virus particle. The p49 (B438L) protein is required for forming vertices of the capsid and is also present in the virus particle. Proteins present in the nucleoprotein core include a histone-like DNA binding protein A104R, components of the multi-subunit RNA polymerase and other enzymes and factors required for early transcription for example the mRNA capping and polyadenylation enzymes, early transcription factors and three helicases. Enzymes involved in the virus encoded DNA repair system are also packaged in the virus particle presumably in the core. Only one viral protein has conclusively been demonstrated to be in the external envelope, the type I membrane protein encoded by EP402R which contains an extracellular domain similar to host CD2

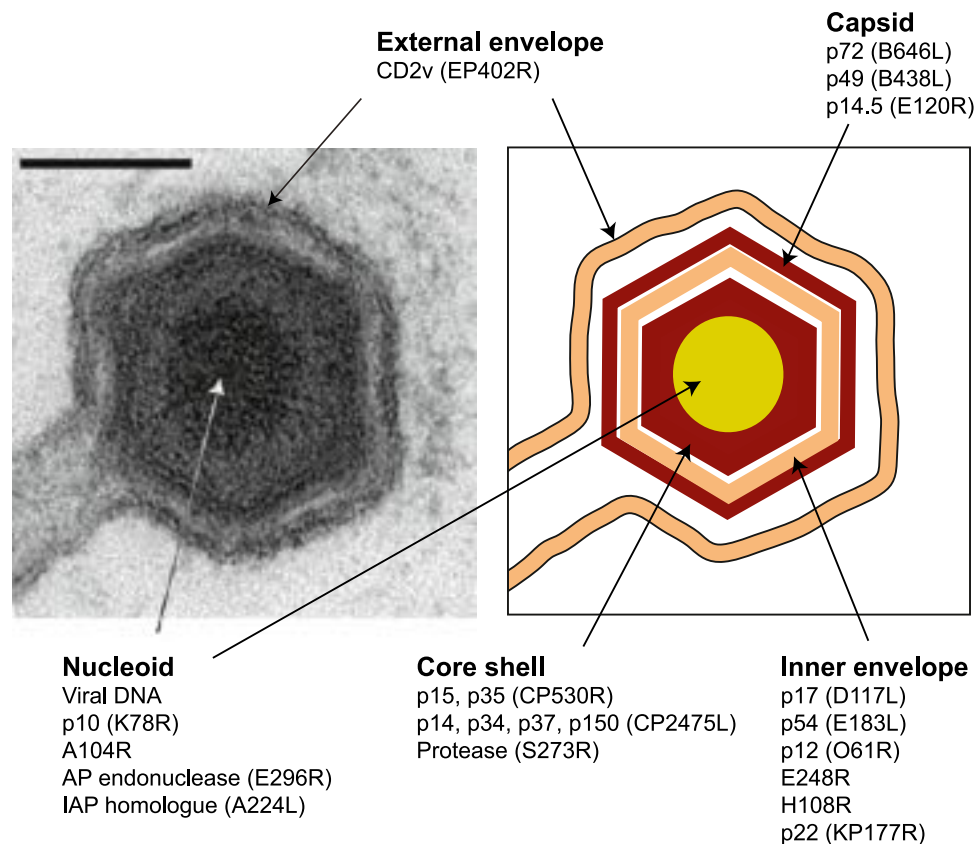


Fig. 1 African swine fever virus structure. Electron micrograph showing the Ba71v isolate of ASFV budding from a Vero cell with a cartoon highlighting the different structural elements. Known viral contents of the nucleoid, core shell, internal envelope, capsid and external envelope are indicated. The scale bar is 100 nm.

protein. The virus CD2 like protein (CD2v) is required for the binding of red blood cells to virus infected cells and extracellular virions. A number of membrane proteins are found on the internal virus envelope including p54 (pE183L), p22 (pKP177R), p17 (pD117L) and p12 (pO61R). The latter was previously shown to be involved in virus attachment. Several ASFV proteins have similarities to components of the vaccinia virus entry fusion complex and it is possible that a similar ASFV complex is localised to the internal envelope.

Virus particles band at 1.095 g/cm^{-3} in Percoll and $1.19\text{--}1.24 \text{ g/cm}^{-3}$ in CsCl. The S_{20W} is about 35,00S. Virus infectivity is stable over a wide pH range between 4 and 13 but is inactivated by treatment at 60°C for 30 min or by UV irradiation. Chemical treatment with some disinfectants, for example 1% formaldehyde, 2% NaOH or paraphenylphenolic compounds effectively inactivate virus.

Genome Organisation

The ASFV genome consists of linear double stranded DNA varying between 170 and 193 kbp depending on the isolate. The genome termini are covalently closed and consist of two flip-flop forms that are inverted and complementary to each other. Adjacent to the termini are tandem repeat arrays. Genome sequences present in sequence databases generally do not include all of the terminal repeat sequences or terminal crosslinks.

The ASFV genome consists of about 150–167 open reading frames (ORFs) which are closely spaced and read from both DNA strands as shown on Fig. 2. In some genome regions a number of consecutive ORFs are on the same DNA strand. Some of these genes are part of multigene families (MGFs) that have evolved on the virus genome by a process of gene duplication and in some examples transposition to the other end of the genome. The number of ORFs is based on the definition that the minimum size for an encoded protein is 60 amino acids and the ORFs do not overlap extensively with other major ORFs. Thus, some smaller ORFs may encode functional proteins but may have been missed from the current sequence annotations of the ASFV genome. Information from the transcript map of the ASFV genome may help in identification of any additional expressed small ORFs. The

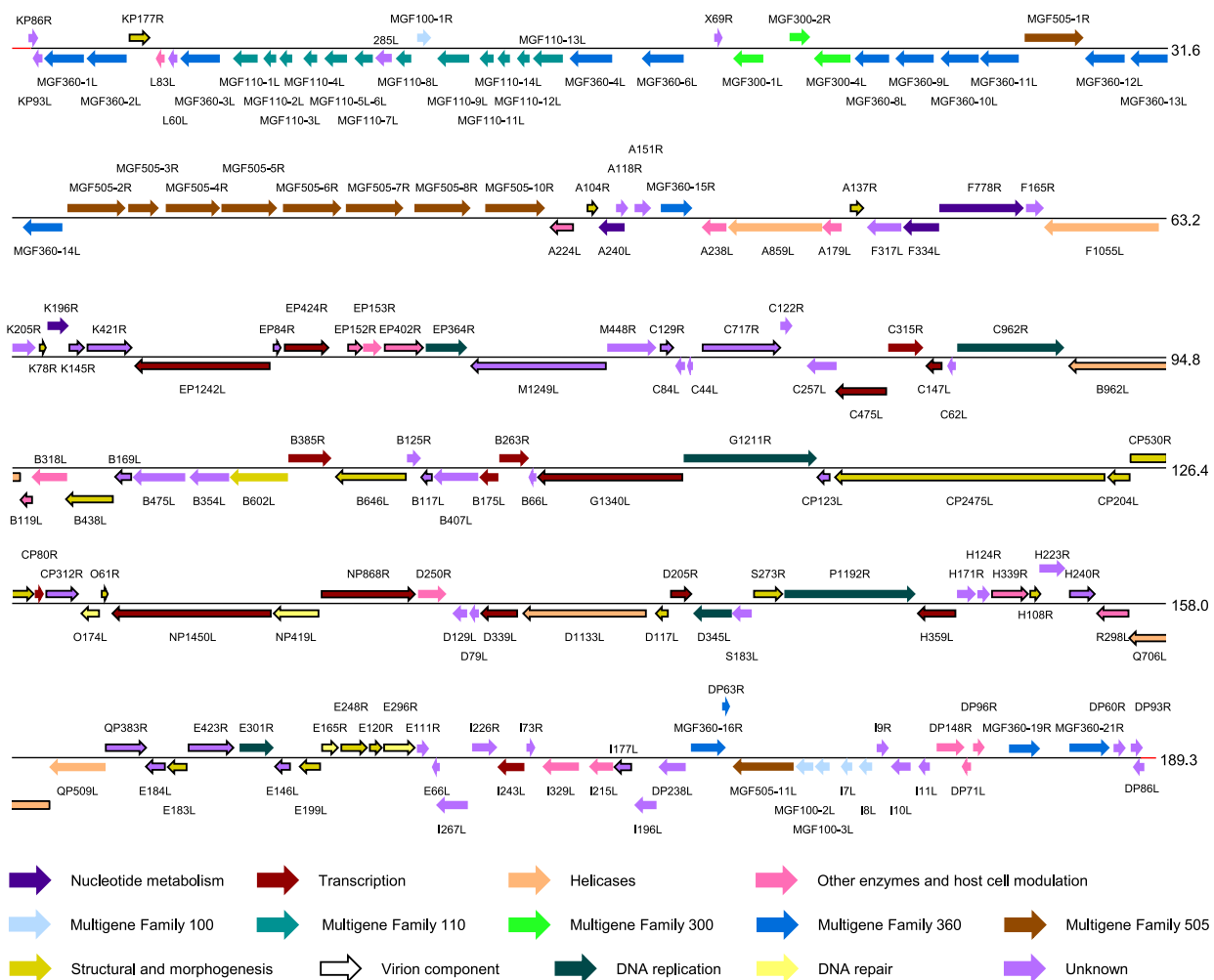


Fig. 2 Genome of African swine fever virus. Schematic representation of the African swine fever virus Georgia 2007/1 isolate. Open reading frames are indicated by arrows indicating the direction that these are read.

intergenic region between ORFs is generally short and contains a short promoter sequence immediately upstream of the ORF. The virus encoded RNA polymerase and transcription factors bind to initiate transcription from the downstream ORF. These promoter sequences are generally AT rich and consensus promoter sequences for replication stage specific transcripts, (immediate early, early, intermediate, late) have not been identified.

Partial sequencing of the gene for the major capsid protein, p72/B646L, has been used to assign ASFV isolates to 24 different genotypes falling in three main clades. These were all found in East Africa, where the ancient sylvatic cycle is present. Genotypes that have spread more widely in domestic pigs include the genotype I in central and West Africa, genotype IX in Kenya and Uganda and genotype II in the Caucasus, Russian Federation, Europe and China. Gain or loss of members of multigene families (MGF) is the main mechanism for genome variability. Loss of some members of MGFs 360 and 505/530 (see below) have been associated with reduction in virulence for domestic pigs. In addition frame shift mutations have been identified in some genes, notably the genes coding for pEP402R/CD2v and the C-type lectin p153R in some low virulence virus isolates. The expression of pEP402R/CD2v protein is required to cause the binding of red blood cells to infected cells, a process referred to as haemadsorption. Thus virus isolates that do not express this protein do not cause haemadsorption.

The different functional classes of ASFV encoded proteins are indicated on the genome map in [Fig. 2](#) and in [Table 1](#) and summarised below.

Multigene families

A striking feature of the ASFV genome is the large number of genes that belong to one of several multigene families, named MGF 100, MGF 110, MGF 300, MGF 360, MGF 505/530 based on the average number of amino acids encoded by family members. The numbers of copies of these genes varies between virus isolates. This varies between 5 and 13 for MGF 110, 3–4 for MGF 300, 11–16 for MGF 360 and 8–10 for MGF 530/505. The MGF 100 contains up to 5 members. In addition the gene for the p22/KP177R virion transmembrane protein has one or two related copies at the opposite genome end in most of the field isolates although these are missing from the genome of the tissue-culture adapted isolate BA71V. These gene families have evolved on the ASFV genome by a process of gene duplication and transposition from one genome end to the other. The functions are largely unknown although some members of MGF 360 and MGF 505/530 play a role in suppressing the type I interferon (IFN) response. MGF110 encode proteins with signal peptides and conserved cysteine motifs and some members contain the amino acid sequence KDEL which are involved in retrieval of proteins from the endoplasmic reticulum (ER).

Polymerases

ASFV transcription does not require the functional host RNA polymerase II since it proceeds in the presence of inhibitors of the host enzyme. From sequence comparisons with eukaryotic RNA polymerase II, ASFV encoded subunits of a multisubunit DNA-dependent RNA polymerase (RNAP) have been identified. These include subunits 1, 2, 3, 5, 6, 7, 10. This is fewer than the 12 subunits coded for by the eukaryotic RNAP suggesting some mechanistic differences. The RNAP 3 subunit has a C-terminal extension dissimilar to the eukaryotic homologue which may represent a fusion with RNAP subunit 11. In addition ASFV codes for homologues of the eukaryotic transcription factors TFIIB and TFS and a putative TATA Box Binding protein (TBP). In eukaryotes TBP binds to specific promoter sequences, initiator factor TFIIB binds TBP and recruits the RNAP. The elongation factor TFS helps the polymerase move past arrest sites. ASFV codes for proteins with sequence similarity to vaccinia virus encoded transcription factors including the early factor VETFL and late factors VLTF2 and VLTF3.

Transcript modification enzymes have also been identified including a poly A polymerase and 5' capping enzyme with guanylyl-transferase capability. Functional characterisation of the ASFV transcription machinery has however not yet been completed.

DNA replication

Sequence comparisons have identified a processive DNA polymerase family B, involved in genome replication and a putative DNA primase involved in initiating DNA replication. Other proteins involved in DNA replication include a DNA topoisomerase class II involved in unwinding of the duplex DNA during replication and a PCNA like protein, potentially involved in clamping the DNA polymerase to the genome. Other proteins potentially involved in replication include those with motifs of helicases and nucleoside triphosphatases (NTPases) and proteins with similarity to ERCC4-like nucleases and a lambda like exonuclease which may be involved in resolution of head to head concatameric replicating genomes into single units. The proteins involved in DNA replication are not packaged in the virus particle but are expressed early during infection.

Helicases

Six helicases are encoded and presumed to have roles in unwinding/winding DNA or RNA to allow access to the virus transcription or replication apparatus. Of these helicases, three are packaged in the virus particle and likely to have roles in enabling early gene transcription.

Nucleotide metabolism

ASFV encodes genes for several enzymes involved in nucleotide metabolism in order to modulate the pool of dNTPs available for DNA replication. These include ribonucleotide reductase, thymidine kinase, thymidylate kinase, deoxyuridine triphosphatase (dUTPase). There is essentially no synthesis of thymidylate in macrophages and both the thymidine kinase and dUTPase genes are required for replication in porcine macrophages, but not the Vero cell culture line that supports replication of some isolates of

Table 1 ASFV genes and their functions. List of ASFV genes and their functions, as well as uncharacterised proteins that have been identified in purified virions by mass spectrometry. Genes are ordered by function and genes which do not code for virion components and no specific function has been described (such as the multigene family proteins) are not included

<i>Gene Name</i>	<i>Function</i>	<i>Virion Component</i>
Nucleotide metabolism		
F778R	Ribonucleotide reductase large subunit	N
F334L	Ribonucleotide reductase small subunit	N
K196R	Thymidine kinase ^a	N
A240L	Thymidylate kinase	N
Transcription		
EP424R	FtsJ-like RNA methyltransferase	Y
NP868R	mRNA-capping enzyme ^a	Y
C475L	Poly A polymerase large subunit	Y
B263R	Putative TATA binding protein	N
NP1450L	RNA polymerase subunit 1	Y
EP1242L	RNA polymerase subunit 2	Y
H359L	RNA polymerase subunits 3 & 11	Y
D205R	RNA polymerase subunit 5	Y
C147L	RNA polymerase subunit 6	Y
D339L	RNA polymerase subunit 7	Y
CP80R	RNA polymerase subunit 10	N
C315R	TFIIB like transcription factor	N
I243L	Transcription factor SII (TFIIS)	N
B175L	Vaccinia A1-like transcription factor	N
B385R	Vaccinia A2-like transcription factor	N
G1340L	Vaccinia A7-like transcription factor	Y
Helicases		
F1055L	Helicase superfamily II	N
QP509L	Helicase superfamily II.	N
Q706L	Helicase superfamily II. Vaccine virus D11 like	Y
A859L	Helicase superfamily II. Vaccinia A18R orthologue	N
D1133L	Helicase superfamily II. Vaccinia D6-like	Y
B962L	Helicase superfamily II. Vaccinia I8 like	Y
DNA replication		
G1211R	DNA polymerase (Group of α -like)	N
P1192R	DNA topoisomerase II ^a	N
EP364R	ERCC4 nuclease domain	N
D345L	Lambda-like exonuclease	N
C962R	Putative DNA primase	N
E301R	Similar to proliferating cell nuclear	N
DNA repair		
E296R	Apurinic/apyrimidinic endonuclease class II	Y
NP419L	DNA ligase ^a	Y
O174L	DNA polymerase X ^a	Y
E165R	dUTPase ^a	Y
Structural proteins and those involved in morphogenesis		
CP2475L	220 kDa polyprotein precursor of structural proteins p5, p150, p37, p14 and p34 ^a	Y
CP530R	62 kDa polyprotein precursor of structural proteins p35, p15 and p8 ^a	Y
E120R	Capsid protein p14.5. Required for virion egress ^a	Y
B438L	Capsid protein p49. Required for vertice formation	Y
EP402R	CD2 homologue ^a	Y
O61R	Inner envelope protein p12. Involved in attachment ^a	Y
D117L	Inner envelope protein p17 ^a	Y
KP177R	Inner envelope protein p22 ^a	Y
E183L	Inner envelope protein p54. Binds to LC8 chain of dynein ^a	Y
H108R	Inner envelope protein ^a	Y
E248R	Inner envelope protein. Involved in post-entry fusion ^a	Y
B646L	Major capsid protein p72 ^a	Y
K78R	Nucleocapsid protein p10 ^a	Y
A104R	Nucleocapsid protein pA104R ^a	Y

Table 1 Continued

<i>Gene Name</i>	<i>Function</i>	<i>Virion Component</i>
CP204L	P32/P30 early phosphoprotein	Y
E199L	Structural protein, transmembrane region ^a	Y
S273R	SUMO-1 protease ^a	Y
A137R	Structural protein p11.5. Protected pigs have antibody response to this protein.	Y
Uncharacterised virion components		
K145R	Uncharacterised late protein	Y
C129R	Uncharacterised protein	Y
C717R	Uncharacterised protein	Y
E184L	Uncharacterised protein	Y
E423R	Uncharacterised protein	Y
F317L	Uncharacterised protein	Y
H124R	Uncharacterised protein	Y
H171R	Uncharacterised protein	Y
H240R	Uncharacterised protein	Y
I73R	Uncharacterised early protein	Y
K421R	Uncharacterised protein	Y
M1249L	Uncharacterised protein	Y
QP383R	Uncharacterised protein	Y
M448R	Uncharacterised protein. Possible RNA ligase	Y
CP312R	Uncharacterised protein. Protected pigs have antibody response to this protein	Y
C105R	Uncharacterised protein. Was C122R in original Ba71v sequence	Y
B117L	Uncharacterised transmembrane protein	Y
B169L	Uncharacterised transmembrane protein	Y
C257L	Uncharacterised transmembrane protein	Y
CP123L	Uncharacterised transmembrane protein	Y
E146L	Uncharacterised transmembrane protein	Y
EP84R	Uncharacterised transmembrane protein	Y
I177L	Uncharacterised transmembrane protein	Y
Other enzymes and host cell modulation		
A179L	Bcl-2 family apoptosis inhibitor ^a	N
L83L	Binds IL-1 ^a	N
EP153R	C-type lectin-like	N
A224L	IAP family apoptosis inhibitor ^a	Y
DP148R	Inhibits interferon signalling ^a	N
DP96R	Inhibits TBK1 and IKK β ^a	N
EP152R	Interacts with BAG6. Essential ^a	Y
A238L	I κ B-like protein. Inhibits host gene transcription ^a	N
DP71L	MyD116, GADD34, ICP34.5 homologue. Translation modulator	N
D250R	Nudix hydrolase/decapping mRNA ^a	N
H339R	α -NAC binding protein ^a	Y
R298L	Serine/threonine protein kinase ^a	Y
B119L	Similar to <i>S.cerevisiae</i> ERV1 and vaccinia E10R ^a	Y
I329L	TLR3 agonist. Inhibits IFN ^a	N
B318L	Trans-geranylgeranyl-diphosphate synthase ^a	N
I215L	Ubiquitin-conjugating enzyme ^a	N

^aIndicates a protein for which there is experimental evidence for its function.

ASFV. It is likely that early expression of these genes increase the dTTP/dUTP ratio to facilitate efficient replication in the macrophage natural target cells for ASFV. These enzymes are not packaged into virus particles.

DNA repair

ASFV codes for components of a base excision DNA repair (BER) system which could be an adaptation required for replication in the highly oxidising environment of the macrophage cytoplasm. The high concentrations of reactive oxygen species (ROS) in macrophages can result in DNA lesions which could either introduce mutations into the viral genome or interrupt reading of the genome by the viral DNA and RNA polymerases. Sequence analysis has identified a DNA polymerase X, an ATP-dependent DNA ligase and a class II apurinic/apyrimidinic endonuclease that could constitute a BER system. A DNA glycosylase is required to initiate the repair process and the apparent lack of an ASFV gene for this enzyme suggests either that the enzyme is a very distant homologue or that the virus recruits a host DNA glycosylase. However, the known components of the BER system are packaged into the virion which implies that the as yet unidentified DNA glycosylase enzyme is also likely to be virally encoded. The BER

pathway is required for replication in the macrophage as deletion mutants lacking the gene for the class II endonuclease are viable in Vero cells, but not macrophages.

Virion morphogenesis and structural proteins

The genes identified as coding for virus structural proteins and those involved in other aspects of virus assembly are shown on [Table 1](#) and on [Fig. 1](#). Initial studies identified virus structural proteins biochemically and mapped the genes for some of these to the virus genome. More recently a proteomic approach using mass spectrometry was used to more fully characterise the proteins present in purified virus particles. The results from these studies generally agree well and are complemented by immunogold electron microscopy to localise proteins in the virus particle. The recent proteomic analysis identified 68 virus proteins in the purified extracellular virions. From these 44 had not previously been characterised and many of these are of unknown function. The major structural proteins include the VP72 (B646L) major capsid protein, the products of polyproteins pp220 and pp62 which are cleaved by a virus encoded SUMO-like cysteine protease into products of molecular weights 150, 37, 34, 14 and 5 kDa or 35, 15 and 8 kDa respectively. These protein products are incorporated into the shell of the nucleoprotein core. Fifteen of the 26 virion proteins containing transmembrane domains were identified in the virus particle and all except one presumed to be incorporated into the inner envelope of the virus particle since only the virus encoded CD2v/EP402R protein has conclusively been shown to be in the outer envelope. Proteins in the inner envelope include several with similarity to the vaccinia virus entry fusion complex and others that have been implicated in virus binding and entry to cells. The vaccinia virus entry fusion complex has 11 proteins each containing a transmembrane domain at either the N or C-terminus but no predicted glycosylation sites or signal peptides. These are features shared with a number of the ASFV virion proteins ([Table 1](#)). Enzymes and factors packaged in the virus also include the RNA polymerase and other components of the early transcription machinery, components of the virus-encoded DNA repair system, others such as two DNA binding proteins, a serine protein kinase and inhibitor of apoptosis protein (see [Table 1](#)).

Proteins that inhibit host defences

By comparison with vaccinia virus, which shares a similar replication cycle to ASFV, it is estimated that more than one third of ASFV encoded proteins may be non-essential for replication in cells but have roles including evasion of host defences or modulation of host cell function.

Inhibitors of apoptosis

ASFV encodes several inhibitors of apoptosis targeting different pathways. These include a member of the Bcl-2 protein family, A179L that binds to a broad range of pro-apoptotic Bcl-2 family members to inhibit their activity. This protein also binds to Beclin-1 and may also inhibit autophagy. The A224L protein belongs to the IAP inhibitor of apoptosis family and has been shown to bind directly to caspase 3 inhibiting its activity. The DP71L protein shares a conserved domain with host protein GADD34 and the Herpes simplex virus-encoded ICP34.5 protein. These proteins up regulate global protein synthesis by targeting host protein phosphatase 1 to dephosphorylate translation initiation factor eIF-2alpha. This activity also prevents translation of a class of mRNAs with short upstream open reading frames that are normally only translated under conditions of cellular stress. These include mRNAs for transcription factors ATF4 and CHOP which activate transcription of genes involved in apoptosis induction. Thus expression of the DP71L protein can this inhibit stress-induced induction of apoptosis. Additional ASFV proteins that inhibit this pathway are predicted since in cells infected with deletion mutants of the DP71L gene stress-induced phosphorylation of eIF-2 alpha is also inhibited.

Inhibitors of type I interferon and inflammatory responses

ASFV encodes a number of inhibitors of the type I interferon response, the main early host response to inhibit virus replication. These were identified either by ectopic expression of ASFV genes identified by sequence comparison or by random screening of ASFV expressed proteins. Alternatively, comparison of type I IFN induction in cells infected with virulent compared to gene deleted attenuated viruses showed that deletion of some genes belonging to MGFs 360 or 505/530 deletion resulted in increased type I IFN induction indicating they function to suppress IFN induction. Other IFN inhibitory proteins include the I329L protein which inhibits the Toll-like receptor 3 activated induction of type I IFN. Other proteins including DP96R inhibit the type I IFN response. Deletion of DP96R, DP148R and MGF 360 and 505/530 genes results in virus attenuation. The A238L protein inhibits transcriptional activation induced by several factors including NF-kB and NFAT mediated by the p300 transcriptional co-activator and its deletion from the virus genome increases several pro-inflammatory responses including TNF-alpha expression, inducible nitric oxide synthesis. The L83L protein has been shown to bind to IL-1 but deletion of the gene did not reduce the virulence of the highly virulent Georgia 2007/1 strain.

Other enzymes

The virus encoded protein kinase, ubiquitin conjugating enzyme and geranyl pyrophosphate synthase may modulate host protein function although their roles are unknown. The protein kinase is packaged in virus particles suggesting it has a role early during infection.

Life Cycle

A diagram of the replication cycle is shown in **Fig. 3**.

Entry

ASFV enters swine macrophages by clathrin-mediated endocytosis as well as macropinocytosis. ASFV can also enter tick midgut phagocytic digestive cells attached to red blood cells. The receptor or receptors on the surface of swine macrophages that ASFV binds to are presently unknown, however it is unlikely to be CD163 as gene-edited pigs lacking this gene are fully susceptible to

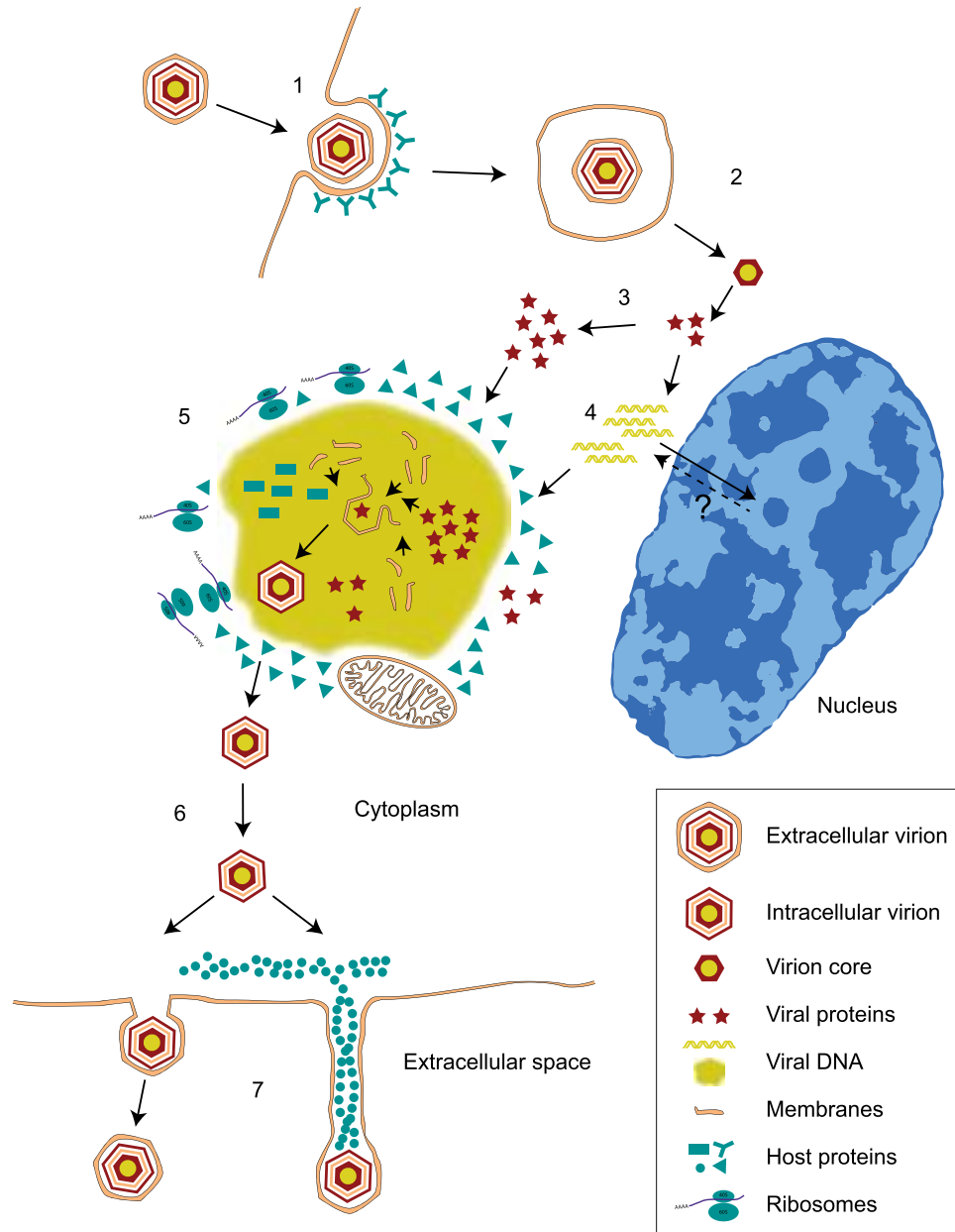


Fig. 3 Replication cycle of African swine fever virus. (1) African swine fever virions enter by receptor mediated endocytosis and enter the endosomal pathway. Virions lose their external envelope and the capsid disassembles in a pH dependent manner in multivesicular bodies. (2) The internal envelope fuses with the endosomal membrane and the core enters the cytoplasm. (3) Expression of early genes begins and (4) DNA replication initiates, possibly involving a nuclear step. (5) Virus factories begin to coalesce in a perinuclear regions of the cytoplasm and exclude host proteins except those required for replication, such as the translational apparatus. Viral proteins and DNA accumulate in factories and striking hexagonal assembly intermediates are seen forming on viral membranes. (6) Fully formed intracellular virions egress along microtubules from the factory to the plasma membrane, from which they bud (7) into the extracellular space gaining an external envelope. Some virions are seen on the tips of cellular projections which label for actin.

ASFV. After entry, the outer envelope and capsid is lost in a pH dependent process within endosomal compartments before the inner envelope fuses with late endosomal membranes delivering the core to cytoplasm. Both the intracellular mature form and extracellular enveloped forms of ASFV are infectious. The precise role of individual viral proteins in entry is unclear. P12/pO63R has long been considered an attachment protein, however this protein has now been shown to localise to the inner envelope rather than the outer and therefore may play a role at a later stage of the entry process. Likewise, antibodies that recognise p54/pE184L, p30/pCP204L and p72/pB646L are capable of neutralising infectivity, but these are components of the inner envelope or capsid. Infection of cells with an ASFV conditional lethal mutant with the E248R gene under control of an inducible promoter showed that expression of the pE248R protein is required for fusion of the virion inner envelope with endosomal membranes. This suggests that this protein may be part of an entry fusion complex similar to that identified in vaccinia virus. Bioinformatics analysis showed similarity of some components of the vaccinia virus entry fusion complex with ASFV virion proteins (see [Table 1](#)). The CD2v/pEP402R protein is the only virus protein that has been localised to the outer envelope of virions although it is not essential for virus replication. This suggests that interaction of the virus outer envelope with cells may involve a non-specific interaction although alternative explanations are possible.

Transcription

ASFV gene expression is tightly regulated with immediate early, early, intermediate and late classes of transcripts described. Early gene transcription begins immediately following entry of virus cores into the cytoplasm. RNA polymerase and other factors packaged in the virus core carry out early gene transcription in partial uncoated cores immediately following entry into the cytoplasm. Shut off of immediate early transcripts is dependent on protein synthesis, whereas initiation of intermediate and late gene expression, as well as the shut off of some early transcripts is dependent on viral DNA synthesis. Whole genome transcription data is currently lacking and therefore the temporal expression patterns are restricted to a few known genes. Late genes are predominately structural proteins involved in the assembly and morphogenesis of progeny virions, whereas early genes tend to be involved in modulating the host immune response and replication, although this is not an absolute rule.

DNA replication

DNA replication takes place predominantly in perinuclear cytoplasmic factory areas at a later stage of replication. The similarity between the genome termini of ASFV and poxviruses suggests they share similarities in DNA replication. This is proposed to involve introduction of single strand nicks in the duplex DNA close to genome termini, providing a template for DNA polymerase to synthesise a strand complementary to some of the terminal inverted repeats and terminal loop structure. Subsequently, these strands fold back such that replication can continue along the genome creating duplexes as predominantly head to head concatemers. Resolution of concatameric full length genomes is proposed to involve extrusion or Holliday- like junctions and resolution by virus encoded enzymes. In support of this model head to head genomic fragments have been detected in ASFV infected cells as well as concatameric genomes of double unit length. An early phase of DNA replication takes place in the nucleus of infected cells and ASFV DNA has been observed juxtaposed to the inner nuclear envelope. Low molecular weight fragments of DNA are synthesised in the nucleus and cytoplasm early during the replication cycle with larger intermediate and full length fragments appearing exclusively in the cytoplasm. It is tempting to speculate that the sub-genomic fragments synthesised in the nucleus are transported to the cytoplasm where they act to prime full length genome synthesis. However evidence for this model is currently lacking since radioactive label incorporated into the shorter fragments could not be chased into the full length genome.

Assembly and morphogenesis

Progeny virions are assembled in discrete perinuclear regions of the cytoplasm called virus factories. Formation of these replication sites coincides with the initiation of cytoplasmic DNA synthesis, however the morphology of virus factories change as replication progresses and serve to concentrate and organise translation, transcription, DNA synthesis and assembly within one cellular compartment. Virions are assembled on the inner envelope with the capsid assembled on the convex surface and the core on the concave. Intermediate forms resembling half formed hexagons are readily visible within virus factories. The inner envelope consists of a single lipid bilayer which by comparison to vaccinia is likely stabilised in part by the major capsid protein, p72. The endoplasmic reticulum has been suggested as the source of the membrane used for the inner envelope, however this was in part based on the detection of viral and cellular endoplasmic reticulum proteins in purified virions by immunoblot which were not present when virions were analysed by mass spectrometry. Membrane targeting of virus proteins is crucial for ASFV morphogenesis. Assembly of the p72 on the convex surface is dependent on a virally encoded chaperone pB602L and correct processing of pp220 into its mature proteins is dependent on myristoylation, as well as the viral protease pS273R mentioned above. Formation of the nucleocapsid and DNA packaging is poorly understood, however is dependent on the prior formation of the core as deletion mutants lacking pp220 and pp62 form 'empty' particles that lack the electron dense nucleocapsid. Once complete, pE120R mediates the transport of virions along microtubules by kinesin motor proteins to the plasma membrane where they gain their external envelope as they bud out of the cell. ASFV can also induce filopodia like actin projections at the cell surface which may aid cell-to-cell spread.

Epidemiology

African swine fever virus (ASFV) was first identified in Kenya in the early 20th century as the cause of a highly lethal disease of domestic pigs that was initially called East African swine fever and later African swine fever (ASF). Wild African suids, in particular the warthog were considered the most likely source of infection. Warthogs did not show clinical signs but spread disease to domestic pigs which they came into contact with. Subsequently, other African wild suids, including the bushpig, red river hog and giant forest hog were also shown to be infected without clinical signs. Soft ticks of the *Ornithodoros* species which inhabit warthog burrows were later shown to be biological vectors of ASFV. The virus can persistently infect ticks over long periods and be transmitted between them transstadially and transexually. Transovarial transmission has been demonstrated for *O. moubata*, the tick vector present in warthog burrows. The virus is very well adapted to replication in hosts in the ancient sylvatic cycle and presumed to have been present over a very long time period. This is supported by the higher genetic diversity of isolates from the regions where this cycle is present.

ASF was subsequently identified in some other East African countries where the ancient sylvatic cycle is present. Since then the virus spread through central and West Africa and is now endemic or causes sporadic outbreaks in almost all sub-Saharan Africa. However, the ancient sylvatic cycle has not been described apart from in East Africa. The virus was contained in sub-Saharan Africa until the late 1950s when there were outbreaks in Spain and Portugal, this was followed with further outbreaks across Western Europe, the Caribbean and Brazil until the 1990s when disease was finally eradicated, although ASFV remains endemic in Sardinia. In 2007 ASFV spread into the Caucasus and from there to the Russian Federation. From 2013 further spread to Eastern Europe was observed. More recently there has been an increase in outbreaks across Europe including: Moldova in September 2016, Czech Republic in June 2017, then Romania in July 2017, followed by Hungary in April 2018, and Bulgaria in August 2018. ASFV was also identified in wild boar in Belgium in September 2018. In addition to European outbreaks ASFV has also been identified for the first time in China in August 2018, with the initial report in the Liaoning region in North East China and has since been reported throughout the country, including Beijing. The virus also appeared in Mongolia in January 2019. The current spread of ASFV presents an enormous threat to the global pig industry.

ASFV can be transmitted through multiple routes including direct contact with infected animals, tick bites, fomites, and contaminated pig products. ASFV is transmitted by direct contact with the virus and in sub-Saharan Africa it is also transmitted by the soft tick *Ornithodoros spp.*, acting as a viral vector. ASFV can be transmitted through direct contact with infected blood which contains up to 10^9 infectious units of virus per ml, or saliva, urine or faeces each of which can contain up to 10^5 infectious units per ml/mg. ASFV has also been shown to be spread by fomites such as feed, equipment and vehicles. Contaminated pig products are another source of infection, with ASFV persisting for months in pork products. This also highlights the risks associated with swill feeding in transmission of ASFV. There is also potential of secretions from infected wild boar being a risk to transmission, by contaminating the environment that is potentially shared with domestic pigs. Wild boar pose another route of transmission to domestic pigs, as in Europe wild boar are equally susceptible to ASFV, and there has been observations in recent outbreaks of dead wild boar from ASFV in the vicinity of ASF-affected farms, as well close to national borders, with some reports suggesting that pig and wild boar interactions with carcass sites play a role in ASFV transmission.

Clinical Features

The clinical signs caused by ASFV infection vary depending on the virus isolate and host. The virus is very well adapted to its ancient wildlife hosts in eastern and southern Africa. These include several wild suid species such as warthogs, bushpigs and red river hogs and soft ticks of the *Ornithodoros* species. In these hosts few clinical signs are observed but virus can persist in tissues, particularly spleen and lymph nodes. In warthogs and bushpigs experimental infection with highly virulent ASFV was observed to cause a transient low fever coincident with a peak of replication in blood. In contrast ASFV infection in domestic pigs and wild boar generally results in an acute haemorrhagic fever with a very high case fatality approaching 100%. Different courses of ASF disease have been described in the field and in experimental infections. Highly virulent isolates can cause a peracute disease which may result in death within 4 days post-infection (dpi) without evident lesions and an acute form of disease which can result in death of almost all infected pigs within 4–21 dpi. In subacute disease forms, caused by moderately virulent isolates, incubation periods are longer, clinical signs tend to be less marked and mortality rates are lower varying between 30% and 70%. A typical disease course results in early clinical signs of fever, reduced eating or loss of appetite and increasing inactivity and lethargy or morbidity. External signs of haemorrhage including cyanosis or skin reddening around the tips of ears, and other extremities, conjunctivitis, bloody diarrhoea can be observed but animals may die without these signs. Clinical signs may typically be observed starting from 3 to 5 dpi but this may vary depending on the dose, route of infection or virus isolate involved. Lower virulence isolates can cause chronic forms of disease characterised by low mortality but signs such as delayed growth, emaciation, joint swelling, skin ulcers and lesions. Some pigs may remain persistently infected with virus in lymph tissues over a prolonged period. Infections of wild boar are clinically very similar to that in domestic pigs and result in similar fatality rates.

Pathogenesis

The main target cells for ASFV replication are those of the myeloid lineage. Monocytes and macrophages at an intermediate to late stage of differentiation are most susceptible to productive infection. Cell markers determining susceptibility to infection have not

been clearly defined. There are reports indicating that dendritic cells may also be susceptible to infection. Vascular endothelial cells have been shown to be infected in cell culture. At later stages of acute disease in animals several other cell types, including lymphocytes and platelets, have been observed to contain virus particles or proteins although virus replication does not occur.

The oral-nasal route is the main route for ASFV infection in suids. In areas where the *Ornithodoros spp* soft tick plays a role infection by tick bites may occur. Following infection the virus replicates mainly in mononuclear phagocytic cells in the tonsils, sub-mandibular or other lymph nodes and spreads through lymph and blood to secondary organs of replication. The virus encodes a number of proteins that inhibit apoptosis but at later stages of the replication cycle apoptosis of the infected cells is observed. Spread of virus in apoptotic bodies may provide a mechanism for virus spread avoiding inflammatory responses induced by necrotic cell death. Virus replication in macrophages has been reported to result in activation of both infected and bystander uninfected macrophages as detected by immunohistochemistry and measured by increased phagocytosis and cytokine secretion. A feature of acute ASF disease is the massive apoptosis observed in both T and B lymphocytes in infected tissues and blood. This apoptosis is observed in uninfected bystander lymphocytes in infected tissues, particularly surrounding infected macrophages. This suggests that factors on the surface of or secreted from infected macrophages are involved in inducing the apoptosis of lymphocytes. Little is known of the mechanism by which apoptosis is induced although TNF-alpha has been identified as a possible contributing factor. In acute and sub-acute forms of disease, ASFV replicates rapidly at primary and secondary sites in lymph tissues including spleen and spreads systemically in blood. This indicates a failure of the host's early innate responses to control virus replication, presumed mainly to result from the many and diverse mechanisms the virus has to inhibit these responses. High levels of virus replication in blood result in typical signs of haemorrhagic pathology including destruction of vascular endothelial cells, leakiness of blood vessels, thrombo and lymphocytopenia and induction of disseminated intravascular coagulation. Typically very high levels of virus replication are observed in blood, up to 10^8 or 10^9 per ml. In some infected animals with moderately virulent isolates, levels of virus replication in blood are lower (up to 10^6 to 10^7 per ml) and some individuals may recover from clinical signs. In pigs infected with low virulence isolates transient low levels of virus in blood may be observed and low levels of virus replication in tissues are detected. Animals that recover may be persistently infected over a period of at least several months. In experimental studies virus was shown to be cleared from persistently infected animals.

Immune Responses

African swine fever virus typically causes a lethal haemorrhagic fever in affected swine, however survivors discovered in the field were found to have robust immunity against homologous strains of the virus. Isolates of reduced virulence obtained from both the field and through experimental passage through tissue culture permitted a more thorough analysis of the immune response to African swine fever and mechanisms that contribute to protection.

Pigs infected with virulent strains of the virus typically exhibit titres of 10^7 – 10^9 infectious units of virus per millilitre of blood. This in combination with high serum levels of pro-inflammatory cytokines such as interferon, tumour necrosis factor alpha and interleukins -1 β , -6 and -8 likely contribute to the lymphopenia and other pathological signs seen in animals suffering acute forms of the disease. A combination of rapid viral replication, disease progression and immunosuppression are likely responsible for the failure to generate ASFV-specific adaptive responses in naïve animals infected with virulent strains of the virus.

Animals infected with low virulence strains typically do not display elevated serum levels of cytokines, but develop both humoral and cellular immune responses that protect animals from acute disease induced by related virulent strains of the virus. Antibodies of the IgM class specific to ASFV first appear between 7 and 11 days post infection and those of the IgG class appear from 14 days onwards. Serum from recovered animals can neutralize ASFV infectivity in vitro as well as mediate lysis of infected cells. The importance of the antibody response in protecting animals from acute disease caused by virulent ASFV strains has been demonstrated by experiments where serum was passively transferred to naïve pigs. Proliferation of ASF specific lymphocytes can be seen from 10 days post infection with attenuated strains of the virus. Pigs recovered from attenuated ASF infection have circulating T-cells that secrete interleukin-2 and interferon gamma after re-exposure to virus, as well as T-cells that have ASFV-specific cytolytic activity. Antibody depletion studies with anti-CD8 monoclonals have shown that the cellular response is essential for the protection afforded by low virulence strains of ASFV. Current research is focussed on identifying precise correlates of protection and identifying which of the 150 or more open reading frames encode for protective antigens.

Diagnosis and Control

Diagnosis of ASF requires laboratory confirmation since clinical signs are similar to several other diseases including classical swine fever (CSF). Detection of virus genome by polymerase chain reaction (PCR) is the usual method used for diagnosis. Detection of infectious virus by replication in primary porcine macrophage cultures is also used, in order to confirm the presence of, and characterise, infectious virus. The haemadsorption of red blood cells to infected monocytes/macrophages or immunofluorescence using antibodies against one of the virion proteins is used to detect infection in cell culture. Kits are commercially available to detect virus antigen by ELISA or using lateral flow devices but the sensitivity of these tests is lower than PCR or virus isolation and they should not be used as the only detection method.

Detection of ASFV specific antibodies using ELISA assay can be used in some situations to identify animals that have been infected. Pigs or wild boar infected with highly virulent isolates usually die before an antibody response is detected. Therefore this method alone is not recommended as the only diagnostic test to detect infection. Detection of ASFV specific antibodies can indicate that isolates of reduced virulence are circulating. To obtain a more complete assessment of the epidemiological situation both ASFV antigen and specific antibodies is recommended.

African swine fever is a notifiable disease to the OIE (World Organisation for Animal Health) as well as national and other international veterinary services. Following confirmation of an ASF outbreak in pigs, veterinary authorities in most countries will cull animals on infected premises and disinfect the premises. In addition movement of pigs is banned from a surrounding protected zone and surveillance for ASF carried out in a wider zone. Infection results in a ban of export trade from the infected area. Varying strategies are applied in different countries to allow continued trade internal to the country or region (for example the EU). If infected wild boar are detected control is more difficult to implement. Following a single point of introduction of ASF to wild boar in the Czech Republic in 2017, disease was eventually eradicated by fencing an extended infected zone and shooting all wild boar in this zone. Reducing contact between wild suids and domestic pigs can also help control ASFV spread particularly if combined with improved on farm biosecurity. In S. Africa double fencing of farms in areas where ASFV infected wildlife are present has proven effective in preventing outbreaks on farms.

The availability of a vaccine would increase options for control. It was established in the 1950s and 1960s that pigs that recover from infection with less virulent strains of ASFV are protected against challenge with related virulent viruses, indicating that vaccine development should be successful. However inactivated virus does not induce protection against challenge even when delivered with recent adjuvants. Moreover it has proven difficult to develop a subunit vaccine since the complexity of the genome makes it difficult to identify protective antigens. Live attenuated vaccines are the most promising approach for vaccine development in the shorter term since high levels of protection can be achieved in the absence of significant replication of the challenge virus. Live attenuated strains isolated from the field have provided excellent models for understanding mechanism of protective immunity but safety issues have precluded their commercial development. Rational attenuation of virus by gene deletion has provided a number of very promising candidates for further evaluation. Deletion of genes that control innate immune responses, in particular the type I interferon response, has been an effective mechanism to attenuate virulent ASFV and induce a protective response. Delivery of live attenuated vaccines in baits to wild boar was successfully used in the eradication of CSFV from Europe and should also be effective for ASFV.

Further Reading

- Afonso, C.L., Piccone, M.E., Zaffuto, K.M., *et al.*, 2004. African swine fever virus multigene family 360 and 530 genes affect host interferon response. *Journal of Virology* 78, 1858–1864.
- Alejo, A., Matamoros, T., Guerra, M., Andres, G., 2018. A proteomic atlas of the African swine fever virus particle. *Journal of Virology* 92, e01293. (18).
- Alonso, C., Borca, M., Dixon, L., *et al.*, 2018. ICTV virus taxonomy profile: Asfarviridae. *Journal of General Virology* 99, 613–614.
- Andreati, J., Khalil, J.Y.B., Sevana, M., *et al.*, 2017. Pacmanvirus, a new giant icosahedral virus at the crossroads between asfarviridae and faustoviruses. *Journal of Virology* 91.
- Blome, S., Gabriel, C., Beer, M., 2013. Pathogenesis of African swine fever in domestic pigs and European wild boar. *Virus Research* 173, 122–130.
- Dixon, L.K., Chapman, D.A.G., Netherton, C.L., Upton, C., 2013. African swine fever virus replication and genomics. *Virus Research* 173, 3–14.
- Dixon, L.K., Sanchez-Cordon, P.J., Galindo, I., Alonso, C., 2017. Investigations of pro- and anti-apoptotic factors affecting African swine fever virus replication and pathogenesis. *Viruses* 9.
- Gonzalez, A., Talavera, A., Almendral, J.M., Vinuela, E., 1986. Hairpin loop structure of African swine fever virus-DNA. *Nucleic Acids Research* 14, 6835–6844.
- Granja, A.G., Perkins, N.D., Revilla, Y., 2008. A238L inhibits NF-ATc2, NF-kappa B, and c-Jun activation through a novel mechanism involving protein kinase C-theta-mediated up-regulation of the amino-terminal transactivation domain of p300. *Journal of Immunology* 180, 2429–2442.
- Hawes, P.C., Netherton, C.L., Wileman, T.E., Monaghan, P., 2008. The envelope of intracellular African swine fever virus is composed of a single lipid bilayer. *Journal of Virology* 82, 7905–7912.
- Hernaez, B., Guerra, M., Salas, M.L., Andres, G., 2016. African Swine fever virus undergoes outer envelope disruption, capsid disassembly and inner envelope fusion before core release from multivesicular endosomes. *Plos Pathogens* 12.
- Jori, F., Vial, L., Penrith, M.L., *et al.*, 2013. Review of the sylvatic cycle of African swine fever in sub-Saharan Africa and the Indian ocean. *Virus Research* 173, 212–227.
- McCullough, K.C., Basta, S., Knotig, S., *et al.*, 1999. Intermediate stages in monocyte-macrophage differentiation modulate phenotype and susceptibility to virus infection. *Immunology* 98, 203–212.
- Netherton, C.L., Wileman, T.E., 2013. African swine fever virus organelle rearrangements. *Virus Research* 173, 76–86.
- Penrith, M.-L., 2013. History of 'swine fever' in Southern Africa. *Journal of the South African Veterinary Association-Tydskrif Van Die Suid-Afrikaanse Veterinere Vereniging* 84.
- Rodriguez, J.M., Salas, M.L., 2013. African swine fever virus transcription. *Virus Research* 173, 15–28.
- Salas, M.L., Andres, G., 2013. African swine fever virus morphogenesis. *Virus Research* 173, 29–41.
- Sanchez-Cordon, P.J., Montoya, M., Reis, A.L., Dixon, L.K., 2018. African swine fever: A re-emerging viral disease threatening the global pig industry. *Veterinary Journal* 233, 41–48.
- Sanchez-Vizcaino, J.M., Mur, L., Gomez-Villamandos, J.C., Carrasco, L., 2015. An update on the epidemiology and pathology of African swine fever. *Journal of Comparative Pathology* 152, 9–21.
- Takamatsu, H.-H., Denyer, M.S., Lacasta, A., *et al.*, 2013. Cellular immunity in ASFV responses. *Virus Research* 173, 110–121.

Akabane Virus and Schmallenberg Virus (*Peribunyaviridae*)

Martin Beer and Kerstin Wernike, Friedrich-Loeffler-Institute, Insel Riems, Germany

© 2021 Published by Elsevier Ltd.

Glossary

Ankyloses Abnormal stiffening and immobility of a joint due to fusion of the bones.

Arthrogryposis Rigid fixation of the joints.

Brachygnathia Shortness or recession of the mandible.

Dystocia Abnormal labor or parturition.

Encephalomyelitis Inflammation of the brain and the spinal cord.

Hydranencephaly Replacement of the brain's cerebral hemispheres by sacs filled with cerebrospinal fluid.

Kyphosis Abnormally excessive convex curvature of the spine.

Lordosis Excessive inward curvature of the spine.

Micromyelia Abnormal smallness or shortness of the spinal cord.

Nystagmus Involuntary, rhythmic oscillating motions of the eyes.

Porencephaly Small cavities in the brain substance.

Scoliosis Lateral deviation of the spine.

Strabismus Condition in which the eyes do not point in the same direction.

Torticollis Twisting of the neck causing the head to rotate or tilt.

Introduction

The genus *Orthobunyavirus* consists of more than 170 viruses and comprise several members of public health or veterinary importance. Among the viruses that cause diseases in animals are Akabane virus and Schmallenberg virus, which infect predominantly ruminants and are able to cross the placental barrier and induce fetal malformation. Both Akabane virus and Schmallenberg virus, are transmitted between their mammalian hosts by blood-sucking insect vectors, specifically biting midges of the genus *Culicoides*.

While Akabane virus was initially described in the 1950th, Schmallenberg virus has been discovered for the first time in 2011. Both viruses are named according to the place of origin of the samples, in which the viruses were detected for the first time. Akabane virus was initially detected in mosquitoes caught in a village in Japan, and Schmallenberg virus was isolated from the blood of acutely infected cattle kept near the German city of Schmallenberg.

Classification

Akabane virus and Schmallenberg virus are members of the genus *Orthobunyavirus* in the family *Peribunyaviridae* (order *Bunyavirales*). Orthobunyaviruses are grouped into 18 serogroups on the basis of serological relatedness and both, Akabane virus and Schmallenberg virus, belong to the Simbu serogroup. Further Simbu serogroup viruses are for example the eponymous Simbu virus, Aino virus, Douglas virus, Peaton virus, Sathuperi virus, Shamonda virus, and Shuni virus. All of them infect animals. The only known zoonotic member of this serogroup is Oropouche virus, which is present in South America.

Virion Structure

Akabane and Schmallenberg virions are spherical, about 100 nm in diameter and relatively simple in their composition, possessing only four structural proteins: the nucleocapsid protein, the viral RNA-dependent RNA polymerase and the surface glycoproteins Gn and Gc. The nucleocapsid protein encapsidates the three RNA segments to form ribonucleoprotein complexes, which associate with the viral RNA-dependent RNA polymerase and are contained within the envelope of the particle (Fig. 1). The viral envelope consists of a lipid layer, which is derived from the Golgi complex of the host cell. On the surface of the virions, locally ordered lattice of spikes protrude from the membrane. These spikes consist of the glycoproteins Gn and Gc, which form heterodimers via disulfide bonds and are anchored in the lipid envelope (Fig. 1).

Genome

The genome of Akabane virus and Schmallenberg virus consists of three segments of negative-stranded RNA, which are named according to their size as large (L), medium (M), and small (S) segment. The L-segment encodes the viral RNA-dependent RNA polymerase. The M-segment encodes the surface glycoproteins Gn and Gc, as well as a non-structural protein (NSm) in the coding

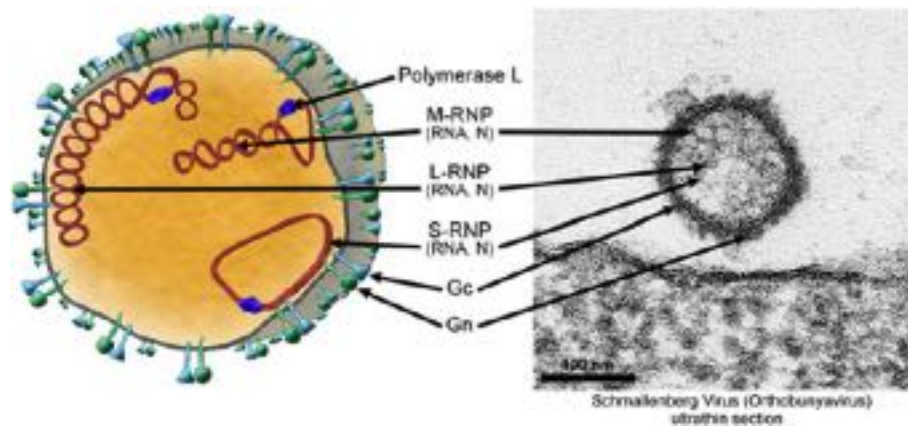


Fig. 1 Virion morphology and genome organization of Schmallenberg virus. The genome consists of three segments of negative-stranded RNA, which are referred to as S, M, and L. The genomic segments have complementary terminal sequences that may base-pair and form circular or panhandle structures. The three RNA segments are encapsidated by the nucleocapsid protein (N) to form ribonucleoprotein complexes, which associate with the viral RNA-dependent RNA polymerase and are contained within the lipid envelope of the particle. The glycoproteins Gn and Gc are inserted in the membrane. Reprinted from Wernike, K., *et al.*, 2014. Schmallenberg virus-two years of experiences. *Preventive Veterinary Medicine* 116 (4), 423-434, Copyright (2014), with permission from Elsevier.

order Gn-NSm-Gc. The S-segment encodes the nucleocapsid protein and the non-structural protein NSs in alternative overlapping reading frames. The three RNA segments have complementary terminal sequences and, as a consequence, the RNAs may base-pair and form circular or panhandle structures (Fig. 1).

Viral Replication

The basic features of the replication cycle are similar to those of other Bunyaviruses, although insect-transmitted viruses such as Akabane virus and Schmallenberg virus have to multiply in two different hosts, i.e., the insect vectors and the mammalian hosts.

Cell entry, which occurs by clathrin-mediated endocytosis, involves interactions between the viral glycoproteins Gn and/or Gc and cell surface receptors, which have not been yet identified. Uncoating of the viral particles likely occurs when the endosomes become acidified, thereby initiating fusion of the viral membrane and endosomal membrane. After the release of the ribonucleoprotein complexes into the cytoplasm, the genomic RNA is transcribed into messenger RNAs (mRNAs) by the virion-associated RNA polymerase. This process is termed primary transcription. The viral RNA polymerase exhibits, in addition to its polymerase activity, an endonuclease activity and this is used to recruit primers for RNA transcription from the host cell mRNA (so-called cap-snatching mechanism). During translation of the mRNA, the virion-associated RNA polymerase switches from primary transcription to replication. Following translation of the primary transcripts into viral proteins, the genome is replicated via a complementary full-length positive-strand RNA, the so-called antigenome, and then further mRNA synthesis ensues, which is termed secondary transcription.

Maturation occurs at the membranes of the Golgi complex. The ribonucleoproteins are transported to the Golgi apparatus that has been modified by the insertion of the viral glycoproteins Gn and Gc. After budding into Golgi membrane-derived vesicles, the viral particles are trafficked to the cell surface via the exocytic pathway involving actin filaments. Subsequently, the fusion of the vesicular membranes with the plasma membrane leads to the release of infectious virions.

Epidemiology

Akabane virus, Schmallenberg virus and other orthobunyaviruses persist in nature by alternate cycling of the virus between mammalian hosts and midges after biting of the mammalian host by the insects. The geographical distribution of the viruses depends on the distribution, seasonal activity, and abundance of the insect vectors responsible for their transmission.

Most of Africa, the Middle East, Asia, and Australia may be regarded as being endemic for Akabane virus. Schmallenberg virus initially emerged in the German/Dutch border region and, thereafter, spread throughout the European continent reaching the Scandinavian countries and the British Isles at the North, the Mediterranean region at the South and Eastern European countries such as Poland, Lithuania and Russia. Within endemically infected areas, Akabane virus and Schmallenberg virus established patterns of cyclic circulation, meaning that for one or perhaps two seasons the viruses are detected only sporadically, while in the intermediate years the viruses re-appear to a greater extent. The differences in the level of virus circulation are presumably related to the overall herd immunity rate in the ruminant population and to the abundance of the vector insects, which is influenced by climate.

The principal vectors of Akabane virus, Schmallenberg virus and further Simbu serogroup members are small biting midges of the genus *Culicoides*. Multiple *Culicoides* species are often present in the same place at the same time. From an epidemiological perspective, important considerations are the infection rates of *Culicoides* with Akabane or Schmallenberg virus and the vector competence, i.e., the physiological ability of a vector organism to acquire, maintain and transmit an infectious agent. The Simbu serogroup viruses are transmitted from the insect vectors to the mammalian hosts with a high level of efficiency, partly because of the relatively high virus infection rates of the *Culicoides* midges and the large number of individual *Culicoides* that may attack an animal. Akabane virus has been detected in the *Culicoides* species *C. brevitarsis*, *C. oxystoma*, *C. imicola*, *C. milnei* and *C. wadai*, while the following *Culicoides* species were implicated in the transmission of Schmallenberg virus: *C. obsoletus* sensu stricto, *C. scoticus*, *C. chiopterus*, *C. dewulfi*, and *C. nubeculosus*. Experimental infection studies with laboratory reared lines of *Culicoides* provided evidence of successful dissemination in *C. sonorensis*. Although Akabane virus has been initially isolated from mosquitoes, they are not considered to be true competent vectors of neither this virus nor of Schmallenberg virus.

In regions with a temperate climate, there is a distinct seasonal pattern of virus transmission, which coincides with warm, moist summer and autumn months and is a consequence of the abundance of the insect vectors. Insect numbers start to increase in the late spring and early summer months, usually peak in early autumn and virus transmission ceases with the onset of very low temperatures and the first frosts. Even in subtropical and tropical regions, there is a tendency towards seasonal transmission, with the highest infection rates in summer and declining transmission rates in the periods of lower rainfall.

Clinical Features

Akabane virus and Schmallenberg virus infect predominantly ruminants. Clinical symptoms have been observed only in domestic animals so far. Direct and indirect detections in association with clinical signs have occurred in adult cattle, sheep, and goats or in their fetuses or newborn offspring.

In addition, Schmallenberg virus genome has been detected in several wild and exotic ruminant species such as alpacas, llama, mouflon, water buffalo or elk.

Furthermore, antibodies against Akabane virus have been found in horses, pigs, camel, deer, and a wide selection of African wildlife, while no infections have been detected in marsupials in Australia. Antibodies against Schmallenberg virus have been detected in various domestic and wild ruminants, free-ranging wild boar, and a range of ungulates kept in European zoos, but clinical signs related to Akabane or Schmallenberg virus infections have not been described in any wild or exotic species.

An infection of adult domestic ruminants with Akabane virus or Schmallenberg virus is either asymptomatic or associated with mild, short-lasting, non-specific clinical signs such as fever, decreased milk production, or diarrhea. Some strains of Akabane virus may induce encephalitis in newborn calves in rare cases and the Iriki strain of Akabane virus, which is present in Japan and Korea, has been occasionally associated with cases of encephalitis in adult cattle.

In contrast to the mild symptoms observed in adult animals, both viruses have significant impact on developing fetuses. When naïve dams are infected during a critical phase of gestation, the viruses can cause stillbirth or premature birth, the birth of mummified fetuses or severe congenital malformations known as “arthrogryposis-hydranencephaly syndrome” (AHS). The range of defects and the percentage of affected newborns varies depending on the herd management and the time period of virus transmission. In herds with a year-round mating system, the full range of congenital defects may be seen, while in herds with a very restricted, seasonal mating period, only a few types of abnormality may be noticed.

An infection early in pregnancy may be associated with embryonic or fetal death and/or abortion, given that an increased number of repeated estrus cycles, which suggest early embryonic or fetal loss, and an increasing rate of abortions have been reported from affected flocks. When fetuses are infected in early to mid-gestation, the viruses may induce congenital lesions. The most common lesions affect the skeletal muscle, the central nervous system and/or the axial skeleton; lesions can affect either one system or can occur in combination. Affected animals may show arthrogryposis ([Fig. 2](#)) associated with histological evidence of muscular hypoplasia, characterized by a reduction in number and diameter of the myofibrils, which may be associated with a loss of cross-striation in myofibrils and fatty replacement. In the central nervous system, the most common lesions are hydranencephaly, porencephaly, cerebellar hypoplasia and micromyelia. These lesions are sometimes associated with nonsuppurative encephalomyelitis and neuronal degeneration and necrosis. The animals are either stillborn or, in cases of less severe malformation, born alive. The following clinical signs or behavioral abnormalities suggestive for lesions in the central nervous system have been described in alive newborns: hypertonicity, hyperreflexia, depression, blindness, ventrolateral strabismus, nystagmus, deafness, dullness, inability to stand, inability to suckle, or aimless wandering. The animals may also display vertebral malformations, including lordosis, kyphosis, scoliosis, torticollis, ankyloses or brachygnathia inferior. These lesions can also appear in isolation with no other clinical sign. In sheep and goats infected with Akabane virus, there may also be pulmonary hypoplasia, which has not been observed after Schmallenberg virus infections. The described defects are usually life-threatening and most affected animals die after birth or are euthanized. The nature of the musculoskeletal and nervous lesions is similar in cattle and sheep. However, only a proportion of newborns whose mothers were infected during gestation show clinical signs or pathomorphological alterations. Even the birth of one malformed and one healthy offspring within the same litter has been described. In cows or ewes delivering severely malformed offspring, dystocia is common and, as a consequence, many of those animals require embryotomy or Caesarian section.

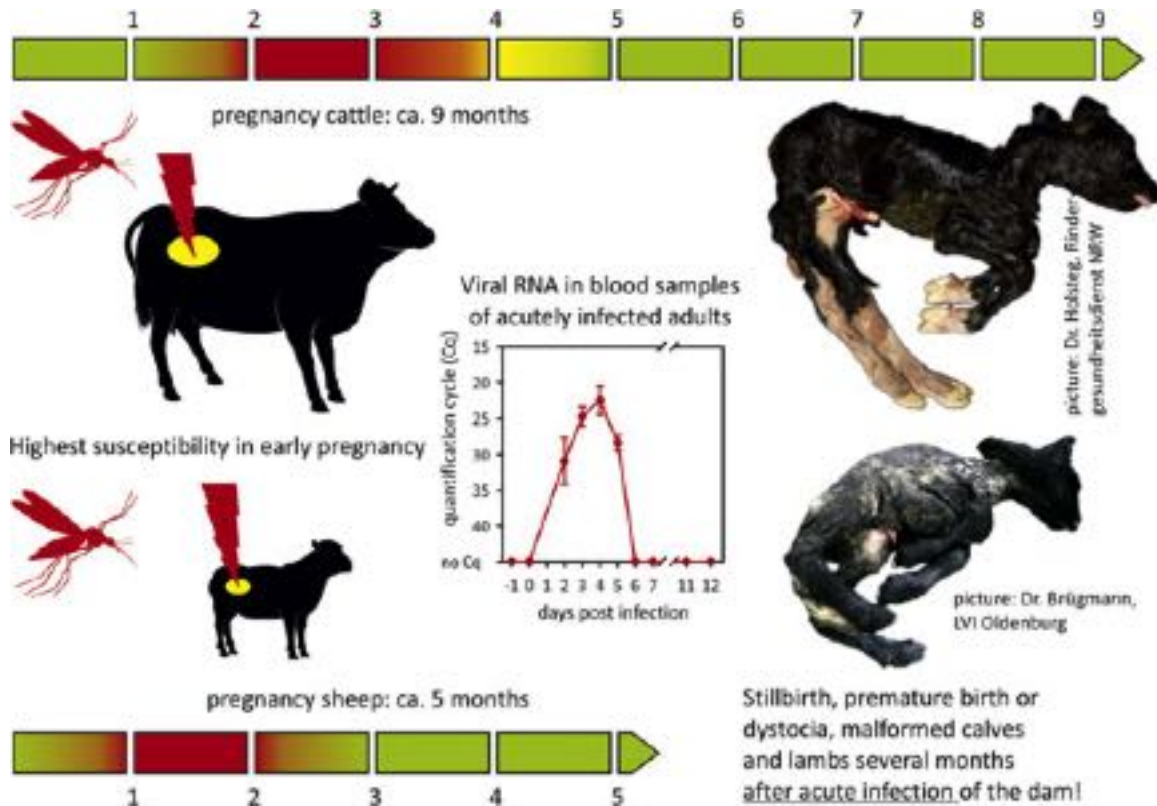


Fig. 2 Genesis of fetal malformations in calves and lambs after infection of their mothers with Akabane virus or Schmallenberg virus during a critical period of pregnancy. When naïve dams are infected, they show a short-lived viremia during which the virus may be transmitted to the developing fetus. Dependent on the stage of pregnancy, fetal infections can lead to stillbirth, premature birth, dystocia or varying degrees of malformation. Modified from Höper, D., *et al.*, 2012. Schmallenberg-Virus, ein neues Virus in Europa. Deutsches Tierärzteblatt 4 (2012), 500-505, with permission.

Pathogenesis

The onset of the two to six days lasting viremia usually occurs one to six days after infection and, in most cases, is not accompanied by clinical signs. The viruses are not associated to any blood cell population, but are circulating freely in the serum, which might assist in their timely elimination from the blood stream.

Similarly to other orthobunyaviruses, the NSs protein of Akabane virus and Schmallenberg virus represents an important virulence factor. In infected mammalian cells, the NSs protein counteracts the shutoff of host cell protein synthesis and the induction of interferon. In addition, NSs is involved in other functions like the regulation of translation, apoptosis, and viral polymerase activity. Viruses, in which the NSs protein has been deleted by reverse genetics, do not induce detectable virus replication in interferon-competent hosts.

As mentioned above, when naïve pregnant animals are infected, the viruses may replicate in the placenta and cross the placental barrier resulting in either abortion, premature birth, still birth or varying degrees of fetal malformation depending on the time of gestation when infection occurred. The susceptibility of the growing embryo or fetus to an infection depends on the maturity of the placentomes and fetal target organs and on the development of the fetal immune system. If the infection of naïve pregnant female takes place before the first placentomes are established (between days 30 and 45 after conception in ruminants), the embryo may be protected from viral invasion. The second key point is related to the maturation of the fetal immune system, since the critical timeframe for the induction of malformations is most probably limited to the time before the fetus becomes immunocompetent. Bovine fetuses develop first antibodies between days 60 and 150 of gestation while in lambs, immunocompetence develops from 65 to 70 days of pregnancy. Consequently, the timeframe during which an infection might lead to malformation ranges from about 30 to 150 days after conception in cattle and from about 30 to 60 days in small ruminants (Fig. 2). Target cells of the viruses within the developing fetuses are neurons in the cerebral cortex, brainstem and spinal cord. An infection of these cells during ontogenesis leads to malformation of the central nervous system such as hydranencephaly, porencephaly, cerebellar hypoplasia or micromyelia. The effects on skeletal muscle are mainly secondary, although a primary viral myositis sometimes occurs. Arthrogryptic newborns show restrictions to movements of the limb joints, which is a result of changes of the soft tissues. The joint surfaces are usually normally developed and the bones are unaffected, while the muscles are

usually reduced in size and paler in color than normal, which is a result of both, neurogenic muscular atrophy and primary infection of the muscles.

Most of the births of dead and/or malformed lambs and calves induced by Akabane virus or Schmallenberg virus infections of their mothers were reported after full-term pregnancy, leading to a time gap of several months between infection and birth (Fig. 2). However, not every infection during the critical period of pregnancy leads to the transplacental transmission of the virus to the fetus. Furthermore, the crossing of the placental barrier does not inevitably result in abortion or fetal malformation. The incidence of affected fetuses or newborns varies depending on the virus, and in the case of Akabane virus, on the strain, but it is usually low. For Schmallenberg virus, an average malformation rate of about 0.5%–3% was estimated for cattle, while on average of 8% of lambs born in infected herds showed typical congenital malformation. However, in some sheep holdings, more than 30% of the pregnant ewes of a lambing season aborted or gave birth to malformed offspring. In cattle naturally infected by Akabane virus, the incidence of defects may be, dependent on the virus strain, as high as 25%–50%, if the animals are infected at the most critical stages of gestation with a very virulent strain. In sheep infected at the most susceptible stages of gestation, the incidence of fetal infection can range from 15% to 80%.

When fetuses are infected once they are immunocompetent or become able to develop specific antibodies during an ongoing infection, the antibodies produced by the fetus itself might assist in the clearance of the virus from the fetus. Due to the occasional elimination of the virus between infection of the fetus and birth, some newborns suspected of Akabane virus or Schmallenberg virus infection test negative for virus or viral RNA. However, the presence of antibodies in the blood of the newborns before the intake of the colostrum of their mothers might be used for diagnostic purposes.

In adult animals, first neutralizing antibodies are commonly detectable between two and three weeks after infection and they prevent from reinfection with the same virus type. Akabane virus-specific antibodies are measurable for at least two years and antibodies against Schmallenberg virus usually persist for many years, if not even lifelong.

Maternally-derived antibodies, which are transferred from seropositive dams to their offspring by colostrum ingestion during the first hours of life, decay over time and are not measurable any more after five to six months.

Diagnosis

The most widely used method for the direct virus detection is the analysis of viral RNA by reverse transcription polymerase chain reaction (RT-PCR) or real-time RT-PCR. Different RT-PCR assays targeting either the L-, M-, or S-segment of either Akabane virus or Schmallenberg virus have been developed. Furthermore, group-specific real-time RT-PCR assays, which allow for the detection of all members of the Simbu serogroup, have been established. In addition to the detection of viral RNA, the viruses can be isolated on insect cell lines (e.g., *Culicoides variipennis* larvae (KC) cells) or mammalian cell lines (e.g., baby hamster kidney (BHK) or African green monkey (Vero) cells). Test systems for antibody detection include micro-neutralization, indirect immunofluorescence tests and various in-house or commercially available ELISAs.

The choice of the sample material is critical for both, virus or viral genome and antibody detection. In adult animals, the preferred material for detecting acute infections and for antibodies is serum, with the detection of virus or viral genome being time-restricted by the short viremia of only a few days. Suitable sample materials for virus or viral genome detection in malformed fetuses or newborns include different parts of the brain, placenta, meconium, and hair swabs (= amniotic fluid collected by a swab from the hair of the fetus or newborn). However, not all calves or lambs by clinical presentation suspected of Akabane virus or Schmallenberg virus infection test positive by RT-PCR. The most accepted explanation for this phenomenon is that the virus, which has caused the initial lesions weeks or months before abortion, stillbirth or birth, is no longer present in the newborn. In addition to the direct virus detection, the detection of antibodies in fetuses or newborns is also a valuable tool to confirm a congenital infection, when serum samples are taken prior to colostrum intake or when fetal heart blood is used as sample matrix from aborted or stillborn fetuses.

For differential diagnosis in cases of newborns showing congenital defects, other teratogenic pathogens should be taken into consideration. Examples include further Simbu serogroup viruses such as Shuni virus or Aino virus, other bunyaviruses (e.g., Cache Valley virus, Rift Valley fever virus) or further viral agents including bovine viral diarrhea, border disease, bluetongue, Chuzan or Wesselsbron viruses. Furthermore, a range of non-infectious causes of fetal malformation such as maternal intoxication (e.g., by ingestion of toxic plants) or an inherited defect should also be considered.

Prevention

Treatment options are not available, but there are two main approaches for prophylaxis: management strategies and vaccination. As Akabane virus and Schmallenberg virus are transmitted by insect vectors, insect-proof nets or mesh to shield animal housing or the use of insecticides or repellents could be taken into consideration to prevent potentially infected vectors from biting susceptible animals. However, case-control studies provided either no or only little evidence for protective effects of those treatments.

An option to prevent from transplacental transmission of the virus from the dam to the developing fetus could be an intelligent management system. At least in regions with temperate climate, the mating period could be adjusted in order to avoid that susceptible animals are in the critical phase of gestation during the season of the highest activity of the insect vectors, i.e., in

summer and autumn. When grazing is used, another option is to adapt the management system and keep young animals outside during the major vector season. In this way, young animals are exposed to insect vectors, which might potentially lead to an infection. The immunity acquired in the course of an infection early in life may prevent fetal infection during a later pregnancy. However, this management concept requires a high number of infected insect organisms every year and a very high transmission rate of the virus from the vector to the animal host to ensure that every young animal is bitten and infected. Therefore, a much more reliable method of prophylaxis is vaccination.

Akabane vaccines have been produced in Japan and Australia. In both countries, inactivated vaccines have been developed either as monovalent Akabane vaccine or as multivalent vaccines to prevent infections with Akabane virus, Aino virus and the likewise teratogenic reovirus Chuzan virus. The vaccination scheme involves two immunizations administered four weeks apart from each other and annual revaccination is recommended. In experimental infection trials, the vaccines were safe when used in pregnant animals. In addition to inactivated vaccines, a live attenuated vaccine has been produced in Japan by serial passage of Akabane virus in cell culture. This vaccine likewise prevents from infection. However, in pregnant ewes this vaccine may cause intrauterine infection of the fetus and, therefore, it is not recommended for use in sheep.

Immunization with inactivated Akabane vaccines do not lead to an *in vivo* cross-protection against Schmallenberg virus infections. Against infections with Schmallenberg virus, virus-specific monovalent inactivated vaccines have been licensed for the European market. These vaccines confer a protective immunity in cattle after two immunizations administered three weeks apart, while sheep have to be vaccinated only once. Annual revaccination is recommended. Apart from inactivated vaccines, live-attenuated, subunit, DNA-mediated and live-vectored vaccines have been developed, but none of them is currently commercially available.

Further Reading

- Elliott, R.M., 2014. Orthobunyaviruses: Recent genetic and structural insights. *Nature Reviews Microbiology* 12, 673–685.
- European Commission: Report on the outcome of Technical and Scientific Studies as described in Annex I of Commission Implementing Decision 2012/349.
- Plyusnin, A., Elliott, R.M., 2011. *The Bunyaviridae: Molecular and Cellular Biology*. Norfolk: Caister Academic Press, (VI, ISBN: 978-1-904455-90-5).
- Saeed, M.F., Li, L., Wang, H., Weaver, S.C., Barrett, A.D., 2001. Phylogeny of the Simbu serogroup of the genus *Bunyavirus*. *Journal of General Virology* 82, 2173–2181.
- World Organisation for Animal Health, 2018. Chapter 2.9.1. Bunyaviral diseases of animals (excluding Rift Valley fever and Crimean-Congo haemorrhagic fever). In: *Manual of Diagnostic Tests and Vaccines for Terrestrial Animals*. World Organisation for Animal Health.
- Zientara, S., Verwoerd, D., Pastoret, P.-P., 2015a. New developments in major vector-borne diseases. Part I: An overview. *Scientific and Technical Review* 34 (1).
- Zientara, S., Verwoerd, D., Pastoret, P.-P., 2015b. New developments in major vector-borne diseases. Part II: Important diseases for veterinarians. *Scientific and Technical Review* 34 (2).

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_9th_report/negative-sense-rna-viruses-2011/w/negrna_viruses/205/bunyaviridae
Bunyaviridae.
- https://talk.ictvonline.org/ictv-reports/ictv_9th_report/negative-sense-rna-viruses-2011/w/negrna_viruses/206/bunyaviridae-figures
Bunyaviridae-Figures.
- <https://www.fli.de/en/institutes/institute-of-diagnostic-virology-ivd/reference-laboratories/nrl-for-sbv/>
NRL for SBV: Friedrich-Loeffler-Institut.
- <http://www.oie.int/en/scientific-expertise/specific-information-and-recommendations/schmallenberg-virus/>
Schmallenberg virus: OIE-World Organisation for Animal Health.

Alphaviruses Causing Encephalitis (*Togaviridae*)

Diane E Griffin, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Arbovirus Virus maintained in an infection cycle involving hematophagous arthropods and vertebrate hosts.

Encephalitis Inflammation of the brain associated with cognitive changes.

History

Nomenclature of the encephalitic alphaviruses recognizes the characteristic geographic distribution of each and the initial identification of the viruses as causes of encephalitis in horses. The first clear record of epidemic equine encephalitis is from 1831, when an outbreak in Massachusetts, USA, resulted in the deaths of 75 horses. Over the next 100 years several outbreaks of encephalitis in horses occurred along the Atlantic seaboard. The virus causing eastern equine encephalitis (EEEV) was first isolated from the brain of an affected horse in 1933 by Tenbroek and Merrill. However, two years earlier Meyer, Haring and Howitt had isolated a virus from the central nervous system (CNS) tissues of two horses involved in an epidemic of equine encephalitis in the San Joaquin Valley of California. EEEV and the Western equine encephalitis virus (WEEV) were antigenically distinct. Both diseases occurred in summertime epidemics, and in 1933, Kelsor showed that mosquitoes could transmit WEEV and shortly thereafter mosquito transmission of EEEV was demonstrated. In 1938 both EEEV and WEEV were isolated from humans with encephalitis occurring in the same regions as the equine cases. In 1936 equine encephalitis occurred in the Goajira region of Venezuela and the virus causing this epizootic, isolated independently by Beck and Wyckoff and Kubes and Rios, was not neutralized by sera from animals immunized with EEEV or WEEV and was designated Venezuelan equine encephalitis virus (VEEV).

Taxonomy and Classification

In 1954 three serologic groups (A, B and C) of arthropod-borne viruses were distinguished by Casals and Brown on the basis of cross-reactivity in hemagglutination inhibition (HI) tests. EEEV, WEEV and VEEV constituted the original group A arboviruses. A second cross-reacting set, including dengue, St. Louis encephalitis and yellow fever viruses, constituted the group B arboviruses and the other viruses, mostly from Brazil, were designated group C viruses. As viruses became classified on the fundamental properties of the virion and genome, the group A viruses became the alphaviruses (genus *Alphavirus*) within the *Togaviridae* family of enveloped plus-strand RNA viruses. HI classifies alphaviruses into six broad antigenic complexes and EEEV, WEEV, and VEEV form three of these complexes. Viruses within each complex have been subtyped using reactivity with monoclonal antibodies, kinetic HI or neutralization assays. The EEE complex includes Madariaga virus (MADV) as well as EEEV. The WEE complex includes Aura, Fort Morgan, Highlands J, Sindbis, Trocara and Whataroa viruses in addition to WEEV. The VEE complex includes Cabassou, Everglades, Mosso das Pedras, Mucambo, Pixuna, Rio Negro and Tonate in addition to VEEV. WEEV and VEEV can also be subdivided into epizootic and enzootic strains. Epizootic strains of VEEV are equine-virulent and are further subdivided antigenically into subtypes IAB and IC (epizootic strains) and ID, IE, and IF (enzootic strains). Sequence information, has confirmed the validity of most of these serologic distinctions.

Virion and Genome Structure

Properties of the Virion

Alphavirus virions are 60–65 nm in diameter. The RNA is enclosed in a capsid formed by a single protein arranged as an icosahedron. The nucleocapsid is enclosed in a lipid envelope derived from the host cell plasma membrane that contains the viral-encoded glycoproteins, E1 and E2. These proteins form heterodimers that are grouped as trimers to form knobs on the virion surface arranged with a $T = 4$ symmetry. Glycoproteins are arranged with 240 copies of each protein interacting with 240 copies of capsid protein (Fig. 1). The virion is sensitive to ether and detergent. Infectivity can be inactivated by heat or acid. The viruses are stable at -70°C for long periods.

Properties of the Genome

The 49S genome is a nonsegmented, capped and polyadenylated message-sense RNA that is infectious. Complete sequence information is available for representative members of all antigenic complexes. The genomes contain approximately 11,700

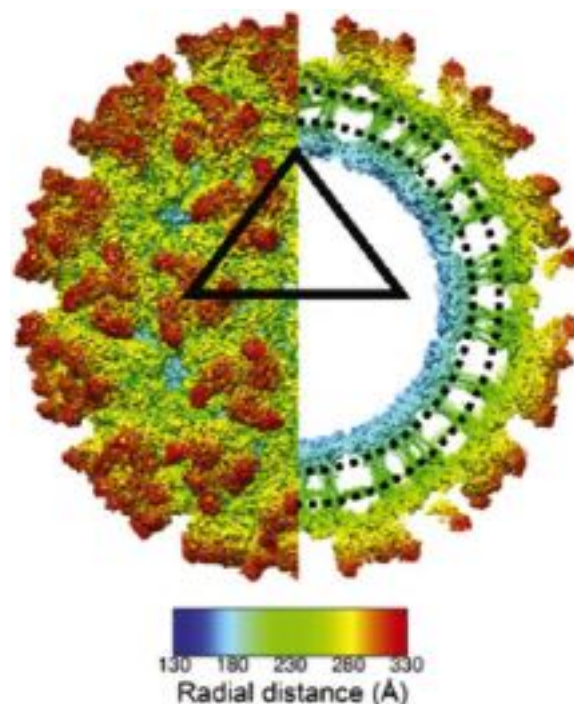


Fig. 1 Cryo-EM structure of eastern equine encephalitis virus showing radially colored surface representation. The right half shows a section of the map with the lipid bilayer indicated between the black dotted lines. From Hasan, S.S., Sun, C., Kim, A.S., *et al.*, 2018. Cryo-EM structures of eastern equine encephalitis virus reveal mechanisms of virus disassembly and antibody neutralization. *Cell Reports* 25, 3136–3147.

nucleotides and have two open reading frames. The nonstructural proteins (nsPs) are encoded at the 5'-end and the structural proteins at the 3'-end. The 3' untranslated region (UTR) is important for negative-strand synthesis, the 5' UTR for genome synthesis and 21 nucleotides at the junction between the nonstructural and structural genes, for synthesis of the subgenomic RNA. Analysis of codon usage shows an underutilization of the dinucleotide CpG.

Properties of the Viral Proteins

Nonstructural proteins

The 5' end of the genome encodes four nsPs that function in replication of viral RNA and in production of the genomic and subgenomic RNAs. The nsPs are translated as a polyprotein from the genomic RNA and sequentially processed to individual proteins by the papain-like protease in the C-terminal portion of nsP2. nsP1 has methyl transferase and guanylyl transferase activities, is palmitoylated and binds the replication complex to membranes. The N-terminal domain of nsP2 has helicase, ATPase, GTPase, methyl transferase and 5'-triphosphatase activity. nsP3 is a phosphoprotein that induces membrane remodeling for the formation of cytoplasmic vacuoles and regulates RNA synthesis in a cell-type-dependent manner. nsP3 has a highly conserved N-terminal macrodomain, a middle zinc-binding region and a poorly conserved, unstructured, acidic and highly phosphorylated C-terminal hypervariable domain that interacts with many cellular proteins to regulate viral replication and assembly of stress granules. nsP4 is the RNA-dependent-RNA polymerase (RdRp).

Structural proteins

Five structural proteins (C, E3, E2, 6K and E1) are translated as a polyprotein from the subgenomic RNA and a sixth small transframe (TF) protein is produced as the result of ribosomal frameshifting. The capsid protein is 259 (EEEV, WEEV) to 275 (VEEV) amino acids long. The N-terminal portion is conserved, basic and binds the genomic RNA and the C-terminal portion interacts with the cytoplasmic tail of E2 and with other capsid proteins to form the nucleocapsid. Capsid also inhibits host gene expression. The E3 protein is 59 (VEEV), 60 (WEEV) or 63 (EEEV) amino acids long, serves as a signal sequence for E2 and, *in vitro*, is shed into the supernatant fluid after cleavage by a furin-like protease in the trans Golgi network. The E2 glycoprotein is a transmembrane protein that is 420 (EEEV) to 423 (WEEV, VEEV) amino acids long and has two (EEEV) or three (WEEV, VEEV) N-linked glycosylation sites. The intracytoplasmic portion has a second stretch of hydrophobic amino acids. The 6K protein is 55 (WEEV, VEEV) or 56 (EEEV) amino acids long, serves as a signal peptide for E1, can form ion channels, remains with interior membranes, interacts with viral glycoproteins and facilitates virus budding. The E1 protein is 439 (WEEV), 441 (EEEV) or 442 (VEEV) amino acids long and has one or two N-linked glycosylation sites. E1 has a short (one or two residue) intracytoplasmic tail and a positionally conserved internal hydrophobic stretch of amino acids in the N-terminal portion of the protein that serves as

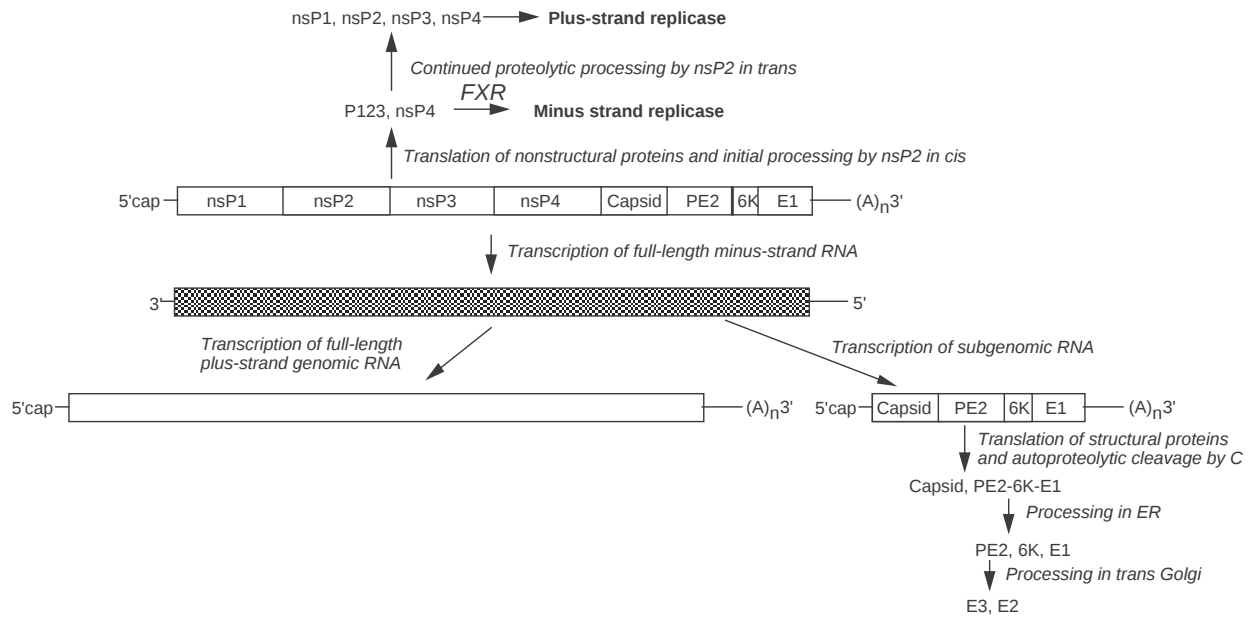


Fig. 2 Replication of new world alphaviruses. Genome is message from which nsPs are translated followed by synthesis of full-length negative sense genomic RNA by P123 + nsP4 (early replicase) and formation of spherules with dsRNA, a step dependent on host FXR. P123 processed to individual nsPs forms the stable plus-strand late replicase and synthesis of genomic and subgenomic RNA. Structural proteins are translated from subgenomic RNA.

the fusion peptide for virion entry into the cell. TF is 70 (VEEV), 72 (WEEV) or 92 (EEEV) amino acids long, palmitoylated, trafficks to the plasma membrane and small amounts are incorporated into virions. Ribosomal frameshifting that leads to TF protein synthesis (<30% of the time) eliminates E1 synthesis.

Replication

Initial attachment of virus to the cell likely involves interaction of the E1 glycoprotein, which has hemagglutination activity, with phospholipids on the cell surface. Efficient subsequent entry involves binding of E2 to a variety of cellular proteins followed by receptor-mediated endocytosis. Within the acidified endosomal compartment the glycoprotein spikes undergo a conformational change that results in fusion with the endosomal membrane and release of the nucleocapsid into the cytoplasm. Acidified endosomes may not be essential for infection of mosquito cells. Once released from the nucleocapsid by interaction with ribosomes, the virion genomic RNA serves as mRNA and translation of the nsPs as 2 polyproteins (P123 and P1234) is initiated (**Fig. 2**). Sequential polyprotein processing regulates the synthesis of minus strand full-length template RNA and plus strand genomic and subgenomic RNAs. nsP4 is first cleaved from P1234 and the resulting P123 polyprotein and nsP4 initiate formation of short-lived early replication complexes within spherules at the plasma membrane that synthesize a genome-length minus-strand template viral RNA. Formation of replication complexes requires interaction of the hypervariable domain of nsP3 with host proteins including members of the fragile X syndrome family of RNA-binding proteins. As P123 is processed to the individual nsPs, these complexes convert to stable replication complexes that synthesize plus-strand 49S genomic and 26S subgenomic RNAs within cytopathic vacuoles.

The 26S RNA represents the 3' portion of the genome from which the structural proteins are translated as a polyprotein despite cessation of host protein synthesis. After translation of the capsid protein is complete, it is autocatalytically cleaved from the nascent chain. pE2 (precursor: E3 plus E2) and E1 are transported into the endoplasmic reticulum, cleaved by cellular signal peptidases and processed through the Golgi with a final processing of pE2 into E2 and E3 by furin. At the plasma membrane nucleocapsids align with regions containing the E1-E2 heterodimers and bud from the cell surface. The process may be different in mosquito cells, where intracytoplasmic compartments containing mature virions are seen *in vitro*. However, these structures have not been seen during ultrastructural studies of alphavirus infection in mosquitoes.

Geographic and Seasonal Distribution

The encephalitic alphaviruses are transmitted primarily by mosquitoes during the warm months of the year and geographically restricted. EEEV is endemic in eastern Canada and from New Hampshire along the Atlantic seaboard and Gulf Coast to Texas in the

United States and in the Caribbean. Inland foci exist around the Great Lakes. Recently EEEV has been detected as far north as Nova Scotia suggesting a northward expansion. In the northern part of the range, cases occur between July and October, while in the southern region cases can occur throughout the year. It is likely that EEEV is periodically reintroduced into northern regions from areas of year-round transmission by migratory birds or wind-borne mosquitoes. Three antigenic subtypes of MADV (formerly South American EEEV) are enzootic in the Amazon River basin, the coasts of South and Central America and Brazil with the cycle maintained in moist forests.

Viruses of the WEE complex are widely distributed throughout the Americas as three closely related but antigenically distinct viruses (WEEV, HJV and Fort Morgan virus). HJV is endemic on the East Coast of the United States in the same areas as EEEV. WEEV is widely distributed in the western plains and valleys of the United States and Canada and is found in South America.

Enzootic VEEVs are perennially active in subtropical and tropical areas of the Americas (subtype II in Florida; ID and IE in Central America; IF and III-VI in South America). Epizootics have appeared in Venezuela, Colombia, Peru and Ecuador at approximately 10-year intervals.

Host Range and Virus Propagation

The life cycle of the encephalitic alphaviruses involves replication in invertebrate vectors, primarily mosquitoes, and in reservoir vertebrate hosts, primarily birds (EEEV and WEEV) or small mammals (VEEV). In epidemic periods additional hosts, such as horses and humans, are infected. EEEV causes encephalitis in humans, horses, emus, pigeons and pheasants. Other birds are susceptible to infection but remain asymptomatic, despite prolonged viremia. The amplifying species for EEEV in North America are wading birds and migratory songbirds. Infection has also been reported in turtles and snakes. In South America, forest-dwelling rodents and marsupials are likely reservoirs for MADV. In invertebrates, EEEV is most consistently recovered from the *Culiseta melanura* in North America and *Culex taeniopus* in the Caribbean with MADV recovered from *Cx. (Melanconion)* spp. in South America. Outbreaks of EEE are initiated when the virus spreads from the enzootic cycle involving ornithophilic mosquitoes (*e.g.*, *Cs. melanura*) into mosquito populations that feed on a wider variety of hosts. Mosquitoes of many species can serve as bridge vectors to humans and equids.

WEEV causes encephalitis in horses and humans, but disease is now rare. The enzootic cycle in North America involves domestic and wild birds and *Cx. tarsalis* for WEEV in the western United States and songbirds and *Culiseta melanura* for HJV in the eastern United States. Serosurveys and virus isolations have demonstrated evidence of natural infection in chickens and other domestic birds, passerine birds, pheasants, rodents, rabbits, ungulates and snakes.

Epizootic strains of VEEV can cause disease in horses and in humans. Infection of horses with enzootic strains of VEEV is asymptomatic and may immunize horses for protection against epizootic strains. Enzootic strains produce mild disease in humans. A variety of wild birds are susceptible to infection, but small mammals serve as an important reservoir, with efficient transmission of infection by *Cx. (Melanconion)* mosquitoes.

In addition to the native hosts, a number of laboratory animals are susceptible to infection. All three viruses cause encephalitis in monkeys, mice, rats, guinea pigs and hamsters. Disease is generally age-dependent, so that young animals more often develop fatal infections than do adult animals. Because of this susceptibility primary isolates of these viruses have often been made in newborn mice.

In vitro, the viruses are routinely propagated in cultures of chick embryo fibroblasts, BHK-21 cells or Vero cells. Most strains will form plaques in these cells and plaque assays provide the usual basis for virus quantification. Mosquito cell lines support replication, often without cytopathic effect. Many other cell lines (*e.g.*, L cells, HeLa cells) support replication as well.

Genetics

Alphaviruses show genetic changes by accumulation of point mutations in the genomic RNA, but this occurs at a rate that is slower than that predicted for other RNA viruses. Recombination is infrequent, but can be demonstrated *in vitro* and occurs at least occasionally in nature because WEEV is the result of recombination between EEEV and Sindbis-like viruses.

Evolution

The alphaviruses derive from a single unknown protoalphavirus and viruses in the alphavirus superfamily have a similar genetic organization and include many RNA plant viruses. In general, amino acids important in secondary structure (*e.g.*, cysteine) have been conserved for the structural glycoproteins E1 and E2 consistent with the similar three-dimensional structure of all alphaviruses. Sequence analyses suggest that the encephalitic alphaviruses evolve slowly in nature. The capsid protein and E1 glycoprotein are the most conserved of the structural proteins, whereas the E2 glycoprotein is more divergent.

EEEV has evolved independently in North and South America over the last 1000 years and there is currently one group in North America and the Caribbean and three groups in South America. Based on sequence divergence, as well as ecologic and pathogenesis distinctions the three South American lineages were recently classified as a different alphavirus species MADV. North

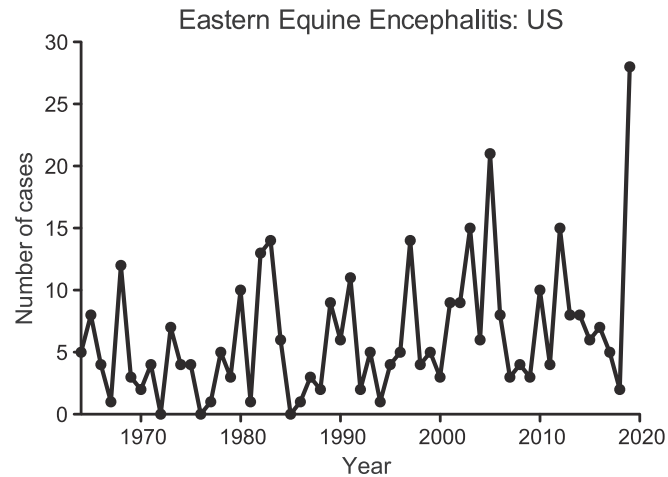


Fig. 3 Numbers of human cases of Eastern equine encephalitis reported in the United States from 1964 to 2019. Data are from the US Centers for Disease Control and Prevention.

American EEEV isolates are highly conserved, varying by less than 3% over 100 years consistent with birds as the reservoir hosts. MADV isolates are evolving more rapidly, consistent with small mammals as the reservoir hosts. Rates of divergence of WEEV and HJV of 0.1%–0.2% per year have been estimated. It is hypothesized that short transmission seasons and limited host mobility constrain genetic diversity in a geographic region.

Epidemiology

EEEV causes localized outbreaks of equine, emu, pheasant and human encephalitis in the summers. Cases of equine encephalitis are usually the first indicators of an outbreak. In North America the primary enzootic cycle is maintained in shaded freshwater swamps where the vector is the ornithophilic swamp mosquito *Culiseta melanura* and the reservoir hosts are migratory passerine songbirds. Young birds are probably most important for virus amplification because they are more susceptible to infection, have a prolonged viremia and are less defensive towards mosquitoes. Human and equine cases usually occur within 5 miles of the swamp, with virus being transmitted by epizootic vectors such as *Coquillettidia perturbans*, *Aedes sollicitans*, and, potentially, *Aedes albopictus*. The enzootic vector in the Caribbean is probably *Culex taeniopus* and for MADV in South America is probably *Culex (Melanoconion)* spp. Epizootics appear approximately every 5–10 years and are usually associated with heavy rainfall that increases the populations of enzootic and epizootic mosquito vectors. There is no evidence for overwintering in mosquitoes so the mechanism for maintenance in northern areas is not known and may require reintroduction from southern areas of transmission. Human infections are unusual, with an average of 5–10 cases per year in the United States (Fig. 3). Serological surveys in the northeastern United States suggest that there are approximately 23 inapparent infections for every case of encephalitis, but children are more susceptible and this declines to only 8:1 for children under 4 years of age.

WEEV is maintained in an endemic cycle involving domestic and passerine birds and *Culex tarsalis*, a mosquito particularly adapted to irrigated agricultural areas. Interseasonal persistence occurs in saltwater marshes, where vertical transmission of WEEV in *Cx. tarsalis* has been demonstrated. Occasional isolations have been made from *Cx. stigmatosoma*, *Aedes melanion* and *Ae. dorsalis*, also competent vectors. Transmission from this enzootic cycle has resulted in small numbers of cases of encephalitis in humans. In the past, there have been widespread summer epizootics of equine encephalitis in North America with significant extension into humans. The estimated case to infection ratio is 1:58 in children under five years and 1:1150 in adults. Recently, for reasons that are not completely clear, cases of equine and human encephalitis have declined with the last human case in 1998 (Fig. 4). In the eastern United States, HJV has been isolated from ornithophilic mosquitoes (*Culiseta melanura*, the enzootic vector for EEEV) in freshwater swamp habitats along the Atlantic coast, as well as from birds. HJV can occasionally cause encephalitis in horses and is a recognized pathogen for turkeys, pheasants, emus and whooping cranes. The lack of significant human disease during equine outbreaks of WEE in South America may be related to the feeding habits of the vector or to a difference in virulence for humans of the South American strains.

VEEV is maintained in enzootic cycles by *Culex (Melanoconion)* spp. mosquitoes that live in tropical and subtropical swamps throughout the Americas and breed near aquatic plants. They feed at dawn and dusk on a wide variety of rodents, birds and other vertebrates. Humans living in these areas have a high prevalence of antibody, but little recognized disease. Epizootic strains of VEEV arise by mutation from enzootic strains and are isolated only during outbreaks. Epizootics have occurred primarily in Latin America in cattle ranching areas during the rainy season. Formalin-inactivated vaccines containing residual live virus are suspected to be responsible for initiating the 1969–72 outbreak in Central America that extended to Texas. During epizootics equids are important amplifying species. Virus has been isolated from several species of mosquitoes including *Aedes taeniorhynchus*, *Ae.*

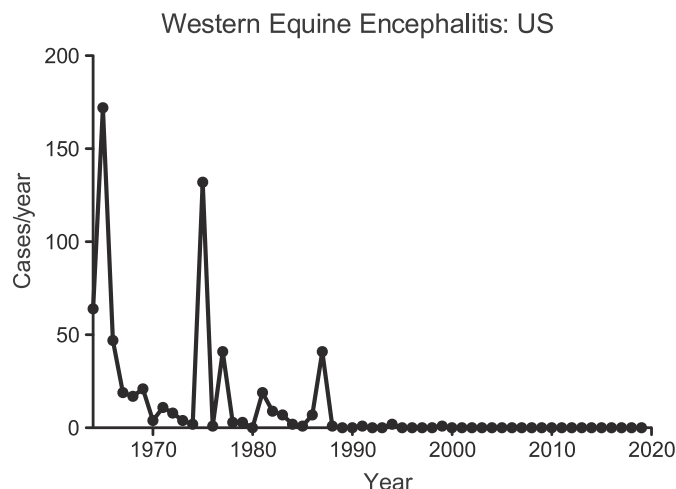


Fig. 4 Numbers of human cases of Western equine encephalitis reported in the United States from 1964 to 2019. Data are from the US Centers for Disease Control and Prevention.

aegypti, *Mansonia dubitans* and *Psorophora confinnis*. The incidence of encephalitis in clinically ill humans is generally less than 5% and the overall mortality less than 1%.

Transmission and Tissue Tropism

The primary mode of alphavirus transmission to birds and mammals is through the bite of an infected mosquito that inoculates virus extravascularly. Mosquitoes become infected by feeding on a viremic host, are able to transmit the virus 4–10 days later (extrinsic incubation) and remain persistently infected. Maintenance of this cycle requires an amplifying host that develops a viremia of sufficient magnitude to infect mosquitoes feeding on it. Other modes of transmission are of occasional importance. EEEV persists in the feather follicles of infected pheasants and secondary transmission among penned pheasants can occur through feather picking and cannibalism. VEEV can be transmitted by the respiratory route between infected horses and to humans in the laboratory. WEEV and VEEV can be transmitted transplacentally.

In mammals, EEEV replicates primarily in muscle and in neurons and glial cells of the CNS, with occasional involvement of liver and lymphatic tissue. Skeletal and myocardial muscle and the CNS of infected birds contain virus. In mosquitoes there is infection of the midgut, muscle and salivary glands without involving the nervous system. WEEV replicates primarily in skeletal and cardiac muscle, brown fat and the choroid plexus, ependyma and neurons in the CNS of mammals. Little is known of the tissue tropism of WEEV in birds. Epizootic strains of VEEV infect the upper respiratory tract, lymphatic and myeloid tissue, pancreas, liver and CNS to varying degrees in different mammals. Mosquito infection is initiated in the midgut epithelial cells and spreads through the hemolymph to the salivary glands and flight muscles.

Pathogenicity

EEEV is the most virulent of the alphaviruses, causing severe encephalitis in humans, horses, dogs, pigs, pigeons, emus, quail and pheasants. The case-fatality rate in humans is 30%–50%, up to 90% in horses and 50%–70% in pheasants. Laboratory studies indicate a similar neurovirulence of EEEV for monkeys, mice and hamsters. Hamsters also develop hepatitis and lymphatic organ infection. At 3–4 weeks of age mice become relatively resistant to peripheral, but not intracranial, inoculation. Birds vary in their susceptibility, with birds of some species developing disease, while birds of many other species show no morbidity or mortality, despite a prolonged viremia. Pheasant deaths are caused by encephalitis while young chickens develop myocarditis. MADV is much less virulent, but infection with lineage III viruses has been associated with cases of fever and encephalitis. Mosquitoes develop persistent infection of the fat body and salivary glands.

WEEV in the western United States caused epidemics of encephalitis in humans, horses and emus, but the case-fatality rate of 10% for humans, 20%–40% for horses and 10% for emus is lower than for EEEV in the eastern U.S. HJV has been rarely reported to cause fatal disease in horses and turkeys. South American strains of WEEV cause fatal encephalitis in horses, but little disease in humans. Epizootic strains are neurovirulent and neuroinvasive in adult mice, whereas enzootic strains are not. HJV is intermediate in virulence between North and South American strains of WEEV. With increasing age mice become relatively resistant to fatal HJV infection, whereas hamsters remain susceptible.

Infection of horses with epizootic strains of VEEV frequently causes fatal disease associated with leukopenia and a high-titered, prolonged viremia. In contrast to EEE and WEE, encephalitis is not always apparent and virus is shed in nasal, eye and mouth

secretions as well as in urine and milk. Experimental infection of laboratory animals produces a variety of disease patterns. In guinea pigs and rabbits, VEEV causes necrosis in lymph nodes, spleen, thymus, intestinal and conjunctival lymphoid tissue, liver and bone marrow. Hamsters develop encephalitis and pancreatitis in addition to widespread infection of myeloid and lymphoid tissues. In rats and mice there is more limited destruction of lymphoid and myeloid tissue and death is usually due to encephalitis. Virus enters the CNS through the olfactory tract and causes neuronal infection and apoptosis. Neonatal mice show extensive replication in many tissues, including brain, myocardium and pancreas. Comparative studies of the virulent Trinidad donkey and avirulent TC-83 strains of IA serotype VEEV and construction of recombinant viruses have led to identification of the 5' noncoding region and the E2 envelope glycoprotein as important determinants of virulence for mice, whereas changes in nsP3 and E2 determine virulence for horses.

Clinical Features of Infection

EEEV is the most virulent of the encephalitic alphaviruses, causing a high mortality due to encephalitis. Prodromal symptoms of fever, headache and myalgias are common. The onset of encephalitis tends to be fulminant and is associated with continued fever, increased headache, meningismus, obtundation and seizures. The overall case-fatality rate is 30% in recent studies, with higher rates in children and the elderly. Recovery is more likely in individuals that have a long (5–7 day) prodrome and do not develop coma. Sequelae are common, with more than 80% of survivors having significant neurological residua, including paralysis, seizures and cognitive deficits. The diagnosis is usually made by detection of antibody in serum or cerebrospinal fluid.

WEEV can cause encephalitis with signs and symptoms similar to those caused by EEEV. There is a 3–5 day prodrome of fever and headache that may progress to irritability, nuchal rigidity, photophobia and altered mental status. Severe disease, seizures, fatal encephalitis and significant sequelae are more likely to occur in infants and in young children.

Infection with epizootic strains of VEEV usually causes relatively mild disease in humans. Illness in adults typically is manifested by fever, headache, myalgias and pharyngitis 2–5 days after exposure. Severe disease, including fulminant reticuloendothelial infection and encephalitis may occur in young children. Children recovering from encephalitis may be left with neurological deficits. Fetal abnormalities and spontaneous abortions may occur with infection during pregnancy.

The cerebrospinal fluid during alphavirus encephalitis usually shows a moderate (up to 4000 cells/mm³) pleocytosis with predominance of either polymorphonuclear or mononuclear cells. Protein is usually elevated and the glucose is normal. The electroencephalogram and magnetic resonance imaging scans are often abnormal, but computed tomographic scans may be normal or indicative only of edema.

Pathology and Histopathology

Initial CNS infection with EEEV in experimental animals is of the capillary endothelial or choroid plexus epithelial cells and spread within the CNS can be cell to cell or through cerebrospinal fluid. The targeted cell within the CNS is the neuron and damage to this cell may be severe and irreversible. Histopathology demonstrates a diffuse meningoencephalitis with widespread neuronal destruction, neuronophagia, gliosis and perivascular inflammation, with polymorphonuclear and mononuclear leukocytes early, and vasculitis and vessel occlusion late. Hamsters also exhibit necrosis of hepatocytes and lymphatic tissue accompanied by local infiltration of mononuclear leukocytes. Initial infection of mosquitoes is of midgut epithelial cells. Infection is facilitated when virus in the serum is concentrated next to these cells as the infected blood meal clots. Infected midgut epithelial cells subsequently degenerate and slough and this process may facilitate penetration of the virus into the hemocoel and rapid dissemination of the infection.

Pathological examination of brains from fatal human cases of WEE demonstrates early perivascular extravasation of blood followed by endothelial hyperplasia, perivascular mononuclear and polymorphonuclear inflammation and parenchymal necrosis. Areas of neuronal degeneration, glial nodules and demyelination are found. Neonatal mice develop acute inflammation and necrosis in skeletal and smooth muscle, cartilage and bone marrow. In animals with encephalitis, the brain shows multifocal areas of necrosis and widespread lymphocytic infiltration of the leptomeninges and perivascular regions of the brain parenchyma. The heart shows a necrotizing, inflammatory myocarditis. Infiltration of mononuclear leukocytes into areas of lung, liver and brown fat also occurs.

The pathology of VEE in horses includes cellular depletion of bone marrow, spleen and lymph nodes, pancreatic necrosis and, in cases with encephalitis, swelling of vascular endothelial cells, edema and mononuclear cell cuffing of cerebral vessels and meningitis followed by the later appearance of demyelinating lesions. Small mammals with widespread involvement of the reticuloendothelial system may develop ileal necrosis.

Both innate and adaptive immune responses are induced by infection. Interferon α/β is produced quickly and virus-specific IgM is often detectable by enzyme immunoassay or neutralization assay very early after onset of the disease. IgM in either serum or cerebrospinal fluid provides a means for rapid diagnosis of infection. IgG antibody appears 10–14 days after disease onset and can be measured by enzyme immunoassay, HI or neutralization. Many lines of evidence suggest that recovery from infection is primarily dependent on the antibody response. Extensive experimental studies to define the antibody specificity and the mechanisms of recovery and protection have been done using infection of mice. Neutralizing and nonneutralizing antibodies against multiple epitopes on the E1, E2 and E3 glycoproteins can protect against challenge and promote recovery.

Cellular immunity has received more limited study, but virus-specific lymphoproliferative and cytotoxic responses have been documented and a mononuclear inflammatory response is common. The importance of cellular immune responses for recovery or for contribution to fatal disease has not been established but interferon- γ can contribute to virus clearance from the CNS. Prior infection with VEEV increases subsequent antibody responses to unrelated antigens. Anti-thymocyte globulin extends time to death in mice infected with VEEV, suggesting a T cell-mediated immunopathogenic component to fatal disease.

Prevention, Treatment and Control

Prevention of infection relies on efforts to control mosquito populations by spraying and reduction of breeding places. Individual use of protective measures such as mosquito repellents and protective clothing are important. Vaccines against EEEV, WEEV and VEEV are available for horses and against EEEV for birds. Experimental human vaccines against EEEV, WEEV and VEEV also are available for laboratory workers exposed to these agents. Most of these vaccines consist of formalin-inactivated virus, but TC-83 is a successful live attenuated vaccine against VEEV for use in horses. TC-83 is also used as a vaccine for humans, although side effects are common. No antiviral agents are of proven usefulness, but supportive treatment of these infections is important.

Future

Construction of full-length cDNA alphavirus clones that can be transcribed into infectious RNA has provided important tools for understanding the functions of various genes and their importance for replication and virulence in the multiple hosts necessary for maintenance of these viruses in their natural cycles. The structures of each of the viral proteins both as individual proteins and polyproteins are needed for interpretation of much of the sequence and virulence data that has been acquired. Further information on the roles of host cell proteins and cell-type dependent differences in replication are likely to provide the next level of understanding of virus-host relationships. There is an ongoing need to develop effective vaccines and therapies.

Further Reading

- Carossino, M., Thiry, E., de la Grandiere, A., Barrandeguy, M.E., 2014. Novel vaccination approaches against equine alphavirus encephalitis. *Vaccine* 32, 311–319.
- Forrester, N.L., Coffey, L.L., Weaver, S.C., 2014. Arboviral bottlenecks and challenges to maintaining diversity and fitness during mosquito transmission. *Viruses* 6, 3991–4004.
- Griffin, D.E., Weaver, S.C., 2019. Alphaviruses. In: Knipe, D.M., Howley, P.M., Whelan, S. (Eds.), *Field's Virology*, seventh ed. Wolters Kluwer.
- Kumar, B., Manuja, A., Gulati, B.R., Virmani, N., Tripathi, B.N., 2018. Zoonotic viral diseases of equines and their impact on human and animal health. *The Open Virology Journal* 12, 80–98.
- Lundberg, L., Carey, B., Kehn-Hall, K., 2017. Venezuelan equine encephalitis virus capsid—the clever caper. *Viruses* 9, 279.
- Ronca, S.E., Dineley, K.T., Paessler, S., 2016. Neurological sequelae resulting from encephalitic alphavirus infection. *Frontiers in Microbiology* 7, 959.
- Sharma, A., Knollmann-Ritschel, B., 2019. Current understanding of the molecular basis of Venezuelan equine encephalitis virus pathogenesis and vaccine development. *Viruses* 11, 164.

Anelloviruses (*Anelloviridae*)

Fabrizio Maggi, University of Pisa, Pisa, Italy and University of Insubria, Varese, Italy

Mauro Pistello, University of Pisa, Pisa, Italy

© 2021 Elsevier Ltd. All rights reserved.

This is an update of P. Biagini, P. de Micco, Anellovirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00694-4](https://doi.org/10.1016/B978-012374410-4.00694-4).

Glossary

Human virome The total collection of viral nucleic acids, both RNA and DNA, that make up the viral community in and on the human body.

Immunosuppression Partial or complete suppression of the immune response of an individual.

MicroRNA A small non-coding RNA molecule that regulates gene expression by sequestering the messenger RNA (mRNA) and interfering with translation of the mRNA into proteins.

Representational difference analysis Molecular technique used to identify and characterize differences between two complex genomes.

Viral species A monophyletic group of viruses the properties of which are unique and can be distinguished from those of other species by multiple criteria.

Viremia The presence of viruses in the blood.

Introduction

Anelloviruses (AV) are a large group of agents discovered in recent years that infect humans and various animal species. The first indication of their existence was in 1997 when a Japanese team applied representational difference analysis to search for novel hepatotropic viruses in the blood of patients with cryptogenetic post-transfusion hepatitis. They found a novel virus with a particularly small genome formed by a circular single-stranded DNA of negative polarity. This new virus was initially believed a novel hepatitis virus and named TT virus (TTV) after the initials of the first patient in whom it was identified. The significance of TTV was then changed in 2004 when it became the acronym of Torquetenovirus (from the Latin words *torques* and *tenuis*, meaning necklace and thin, respectively) to comply with the International Committee on Taxonomy of Viruses (ICTV) rule stating that no official virus designation shall be derived from people's names. The discovery of TTV also opened a Pandora's box of new and novel roles for viruses. Soon after its identification, it became clear that TTV was neither associated to hepatitis nor to any known disease, as the virus was found to be widespread in healthy and sick people. In the 2000s, the discovery of TTV was followed by identification of numerous related, previously unrecognized viruses, detected in diseased and healthy individuals and with genome properties resembling, yet sometime quite divergent from, those of TTV. Furthermore, polymerase chain reaction (PCR) testing of some blood donors yielded amplicons much shorter than expected for a *bona fide* TTV. Further analyses demonstrated the existence of additional viruses clearly related to but with a smaller genome than TTV. Those groups of viruses were named Torquetenominivirus (TTMV), and Torquetenomidivirus (TTMVD). Since 2009, all of these viruses are classified in the newly established family Anelloviridae (from *anellos*, Latin for ring, to indicate the circular genome) (Fig. 1).

Although knowledge on TTV and related AV has accumulated rapidly, many fundamental aspects of their infection remain unresolved. AV have been found to be amazingly widespread, with abundant viral DNA detectable in the plasma of 80% or more of worldwide general population, but their significance for human health is still unknown. It has been proposed that AV should be considered completely nonpathogenic and recent evidence showing that they represent the most abundant members of the human virome speak in favor of this hypothesis. However, at this time, it seems wiser to consider AV a "disease orphan", similar to other viruses that were found to produce significant pathologies many years after their discovery.

Classification

Initially, AV were classified within the Circoviridae family. This grouping was based on the high genome relatedness between AV and the chicken anemia virus (CAV), a prototype circovirus that causes severe damage to the poultry industry, and from the assumption that all known vertebrate viruses with small circular, single-stranded DNA genomes were circoviruses. Subsequent deeper sequence analysis of AV genomes showed that, although they share many genetic features with the circoviruses, they differed considerably from other viruses of this family (e.g., porcine circovirus, beak and feather disease virus of parrots) and had no major sequence homology. Recognition of these differences and other unique attributes have led to group AV in the Anelloviridae, a family coined *ad hoc* by the ICTV.

The progressive discovery of AV human and animal strains exhibiting genomes highly divergent and with size ranging from ~2 kb to ~4 kb, makes reliable phylogenetic and taxonomic analyses of full-length sequences complicated. From this standpoint, nucleotide sequencing and genetic analysis of the entire open reading frame-1 (ORF1) is the most convenient approach for

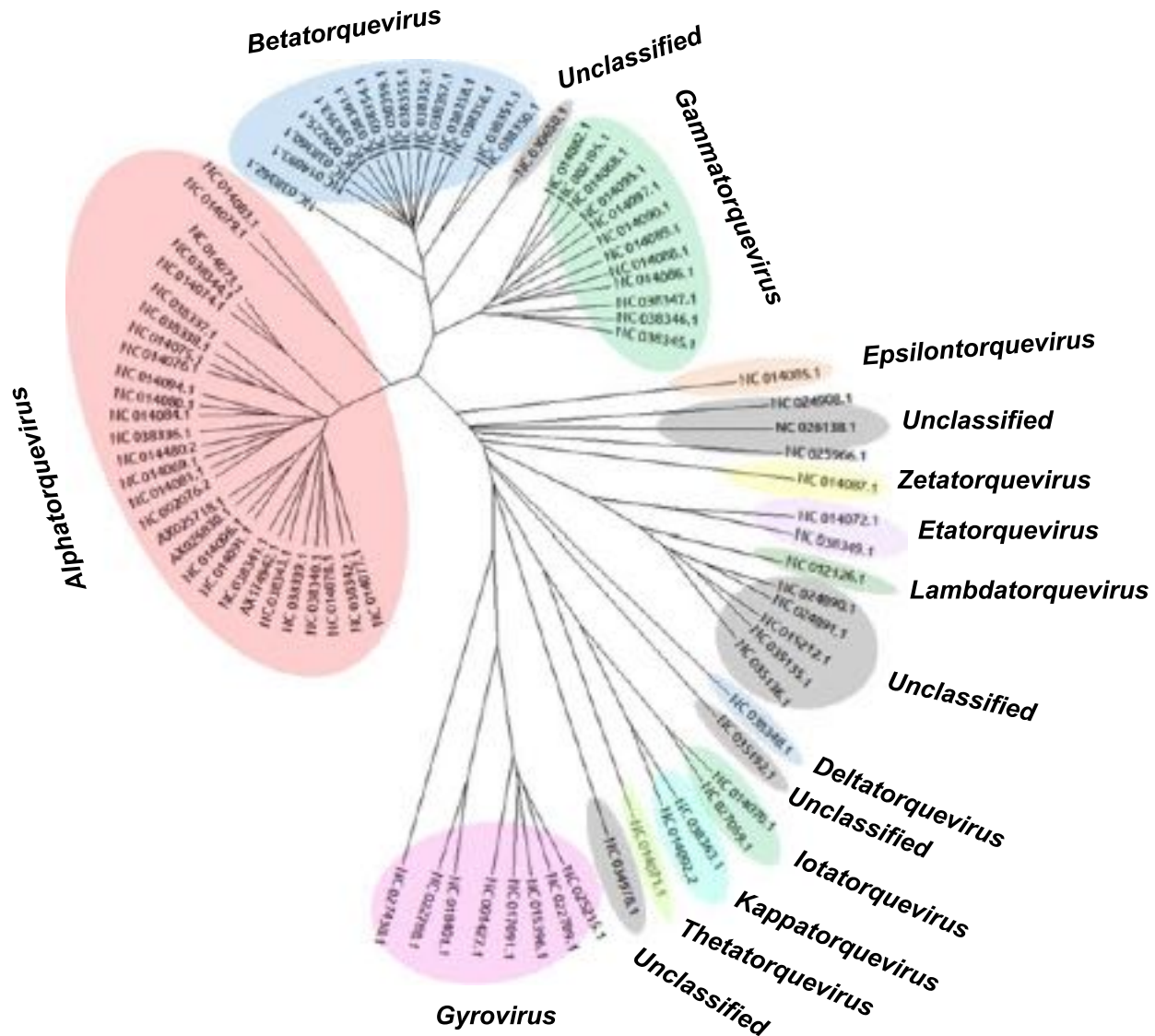


Fig. 1 Phylogenetic tree based on full-length nucleotide sequences of viral species of Anelloviridae.

classifying AV, which are grouped in genera, species and isolates to cover their wide range of genetic variability. Based on the currently available data, AV are divided into 12 genera, each of which is designated by a letter of the Greek alphabet with the only exception of the genus *Gyrovirus*, that has been recently moved from Circoviridae to Anelloviridae (Table 1). Within a given genus, the ORF1 DNA genomes of all members have more than 56% divergence with AV of other genera. A species is designated for those AV within a given genus that share 45%–80% identity, while a viral isolate within a species has more than 80% identity with other isolates within the species. By using these criteria, human AV are clustered among 3 of the genera: *Alphatorquevirus* (identifying the TTV sequences) *Betatorquevirus* (TTMV sequences), and *Gammatorquevirus* (TTMDV sequences); the remaining 9 genera include animal AV only. Human TTV sequences are genetically classified in at least 29 major species, each of which consists of numerous isolates. TTMV and TTMDV has been investigated in less detail, and existing data indicate that they may be more homogeneous compared to TTV, but the recognition of much more diversity is expected as the number of available sequences increases.

Virion Structure and Genome

Knowledge of physical-chemical properties of AV is incomplete and has been obtained by studying plasma and, less commonly, fecal extracts of TTV infected individuals. As determined by PCR analysis of infected serum passed through polycarbonate filters of decreasing pore size, the TTV particle is approximately spherical and has a diameter of between 30 and 32 nm (below 30 nm for TTMV). Isopycnic density was found to be 1.32–1.33 g/ml in cesium chloride (1.27–1.28 g/cm³ for TTMV) and 1.26 g/ml in sucrose gradients. Because

Table 1 List of genera and species in the family of Anelloviridae

Family	Genus	No. of species included	Representative species
Anelloviridae	<i>Alphatorquevirus</i>	29	Human torquetenovirus
	<i>Betatorquevirus</i>	12	Human torquetenominivirus
	<i>Deltatorquevirus</i>	1	Torque teno tupaia virus
	<i>Epsilontorquevirus</i>	1	Torque teno tamarin virus
	<i>Etatorquevirus</i>	2	Torque teno felis virus
	<i>Gammatorquevirus</i>	15	Human torquetenomidivirus
	<i>Gyrovirus</i>	1	Chicken anemia virus
	<i>Iotatorquevirus</i>	2	Torque teno sus virus 1
	<i>Kappatorquevirus</i>	2	Torque teno sus virus k2
	<i>Lambdatorquevirus</i>	1	Torque teno zalophus virus
	<i>Thetatorquevirus</i>	1	Torque teno canis virus
	<i>Zetatorquevirus</i>	1	Torque teno douroucoulis virus
	<i>Unclassified</i>	At least 28	Torque teno rodent virus

density does not change following Tween 80 and is poorly affected by solvent/detergent treatments, TTV has been deduced lack an external lipid envelope and to be as stable to inactivation by chemical and physical agents as parvoviruses and circoviruses.

The component of AV currently best understood is the genome. Early characterization showed that nucleic acids extracted from TTV were sensitive to DNase I and mung bean nuclease but resistant to RNase A and specific restriction enzymes, indicating that the viral genome is single-stranded DNA. Initially, when the prototype Japanese isolate TA278 was sequenced to about 90%, it was proposed that the genome was linear. When sequencing of TA278 genome was further extended, it was then recognized that the two extremities were connected by a guanine and cytosine (GC)-rich stretch of about 100 nucleotides (nt), that was hard to amplify and sequence in earlier studies and that completes the covalently closed circular molecule. Hybridization/nuclease protection assays also showed that virions encapsidate the minus strand.

The genomes of many of the human and animal AV have been sequenced in their entirety, and show similar genomic organization. However, although all consist of a circular molecule of single-stranded DNA, AV genomes differ in size; TTV is 3.5–3.9 kilobases (Kb), TTMDV around 3.2 Kb, and TTMV 2.7–2.9 Kb. All are negative sense, contain an untranslated region (UTR) of about 1.2 kb, and a potential coding region of approximately 2.6 kb. The UTR appears to control virus replication and expression through a number of important regulatory sequences and secondary structures (those formed by the 100 nt high GC region are particularly important), and it possess a short nucleotide sequence highly conserved among all AV isolates. The coding region consists of two major potential protein-coding genes, named ORF1 and ORF2. These ORFs are of different size, are present in the plus strand complementary to the genomic DNA, partially overlap and are in different reading frames. In the prototype TA278 sequence, ORF1 spans nt 589–2898 and ORF2 spans nt 107–712, which correspond to a coding capacity of 770 and 150 amino acid (aa), respectively. The length of the two ORFs may vary in individual isolates. ORF1 is believed to encode the putative capsid protein, which also contains specific motifs: (1) an arginine-rich hydrophobic region at the N-terminus that is believed to mediate binding of the genome to the capsid and its transportation to the cell nucleus; (2) a hypervariable region (HVR) located centrally that encodes potential glycosylation sites that vary in number and location in individual AV strains and that might affect several biological properties of the resulting protein, including antigenic specificity; (3) short aa motifs typical of replication-associated proteins (Rep) of viral DNAs replicating using the rolling circle mechanism. The smaller ORF2 encodes for a putative nonstructural protein of approximately 100–120 aa that is thought to possess phosphatase activity and be involved in viral replication. Sequence analysis of AV isolates predict additional putative ORFs that might encode for proteins which function is poorly characterized. Among these, there are two phosphorylated proteins, one resembles the gene NS5A of hepatitis C virus, the other, named TTV-derived apoptosis-inducing protein (TAIP), has been shown to induce p53-independent apoptosis in various human hepato-cellular carcinoma cell lines at levels comparable with that of CAV apoptin.

Life Cycle

Although the discovery of TTV was over twenty years ago, the AV life cycle is poorly defined. Lack of *in vitro* systems to efficiently cultivate and propagate viral isolates has greatly limited studies of this family of viruses. This includes the screening of cell types supporting replication of AV, the identification of the receptor(s) on the cell surface used for viral entry, the dissection of the biosynthetic machinery exploited to replicate the virus, identification of intracellular signaling pathways perturbed, and visualization of cytopathic effects on infected cells. Our current understanding of AV replication mechanisms is thus mostly based on inference from what is known about other single stranded DNA viruses, even though these have not been studied in great detail either. What is certain about is that AV are found in many organs and infect different cell types. It is presumed, therefore, that the cellular receptor(s) used for viral entry is/are widespread in tissues and exposed by a wide array of cell types. A precise definition of

cell types supporting virus replication is not yet available. Since related single-stranded DNA viruses require actively multiplying cells for productive replication, it is conceivable that replication of most, if not all, AV depends on cellular proteins expressed during the S-phase of cell cycle. It has also been shown that the same cell can be infected with more than one AV, thus explaining why recombination is a frequent occurrence among these viruses.

Transcription analysis of prototype TTV strains have led to the identification of at least three RNA transcripts of different lengths (3.0, 1.2, and 1.0 Kb). Such transcripts share common 5' and 3' termini and are encoded by the positive DNA strand through alternative splicing. However, the transcriptional strategies of different AV vary considerably. The mechanism of DNA replication is not known with certainty but there are several indications pointing to a rolling circle mechanism. Such a mechanism, particularly well-suited for single-stranded DNA viruses, is indicated by the presence in ORF1 of conserved Rep protein motifs typical of the many viruses that replicate in this manner. The absence of motifs typical of known DNA polymerases in the encoded products suggests that the genome is replicated by cellular enzymes. Due to the great genetic diversity of AV, it has also been hypothesized that the viral DNA is replicated by polymerases with poor or no proof-reading activity. Indeed, rates of nucleotide change found in the HVR region of TTV are at least ten-fold higher compared other DNA viruses and are similar to that observed for RNA viruses. Using a combined computational and synthetic approach, microRNA-coding regions have been identified in diverse human TTVs and likely produce TTV microRNA *in vivo*. To this aim, TTV utilizes the host microRNA biogenesis machinery to produce biologically active microRNAs. Although their role in viral replication is unclear, it has been hypothesized that they promote viral persistence and immune evasion.

Modes of virion assembly and release from producer cells are not known. It is likely that virus particles are assembled in the cytoplasm and, in analogy with what observed with other nonenveloped viruses, released by cell lysis. However, recent findings demonstrating that TTV particles are sequestered within circulating extracellular vesicles suggests other routes of egress of progeny virions. Available data do not support the idea that TTV can persist as an episome or is integrated into host cell DNA.

Epidemiology

AV are among the most successful viruses of humans. Indeed, the best documented aspect of AV epidemiology is the remarkably high diffusion of infection in general population and throughout the world. Epidemiological surveys employing PCR assays of high sensitivity and searching UTR or the most conserved segments of encoding regions in the blood, have shown that TTV infects nearly two third of the general population and that prevalence is not related to geographical, socio-economical, and age- and gender-related factors. Interestingly, prevalence is high in geographically segregated populations that have limited contact with the rest of the world. TTMV and TTMDV have a prevalence of active infections similar to TTV, thus individuals who carry multiple AV in their blood and/or other tissues are the rule rather than an exception. Metagenomic analyses provided evidence for the presence of as many as 47 different species of AV in the gut of one infant at 12 months of age. Systematic studies performed using standardized and consistent protocols will allow the proportion of individuals carrying multiple species of TTV to be defined and meaningful comparisons of the study cohorts to be performed.

Although in some surveys viremia rates increase with age and peak in adulthood, AV infection is frequent already in the early months of life. This implies that AV infection is highly contagious and uses various routes of spread. That AV can be transmitted via blood and blood products has been repeatedly documented. Most screenings performed early after TTV discovery found the highest detection rates amongst polytransfused, thalassemic, and long-term hemodialysis patients, as well as hemophiliacs, and intravenous drug abusers, clearly indicating that TTV is a blood-borne virus. These observations, again to comply with ITCV rules, prompted a change of the original TTV acronym from Japanese patient initials to "Transfusion Transmitted Virus". As mentioned above, this denomination was replaced by the current definition as soon as it became evident that TTV viremia rates in general population were far too high to be explained by blood-borne transmission. Indeed, TTV was later found in feces, respiratory secretions, and various biological fluids. The common occurrence of TTV in feces of viremic patients suggests that oral-fecal route is a significant mode of transmission. This finding combined with the high prevalence of virus carriers and a high resistance to physicochemical agents, leading to the presence of the virus in food, wastewater, drinking water, and river water, likely contributes to a particularly high dissemination of TTV in human environment. Attempts are underway to evaluate whether TTV might be useful as a marker of anthropic pollution instead of or in addition to enteric bacteria. Mother-to-fetus transmission of TTV is also very common. TTV DNA sequences similar to those in the corresponding maternal blood were detected in newborns and cord bloods in amounts as high or higher than the respective peripheral blood. Alternative routes of AV spread have been poorly investigated, although TTV and related AV have been detected in nasopharyngeal and genital secretions, suggesting that, at least in some instances, the virus can be transmitted via respiratory and sexual routes. AV DNA sequences are also found in saliva, tears, skin, and hairs implying that household contact transmission is also possible. Whether non-human AV or AV-like viruses, reported to be widespread in chickens, pigs, cats, dogs and other farm and companion animals, contribute to AV epidemiology in humans is unclear. More detailed analysis of these viruses is needed to establish whether animal AV add to the already enormous reservoir of AV genomes existing in nature and contribute to human AV evolution and diversification.

Few studies have investigated the geographical distributions of TTV species. Some species have been found to be very common worldwide, others seem to be less frequent or clustered in specific geographic regions. It is also uncertain whether these data reflect

real differences in species distribution or are biased by preferential amplification of selected species by the molecular assay used. Whether species frequency varies with age of infected subjects and/or route of transmission also remains to be determined. Extensive investigations with next generation sequencing (NGS) or other standardized tools are warranted to obtain a wide and precise epidemiological picture.

Clinical Features

To date, the medical significance of TTV and AV in general is unclear. Undoubtedly, the circumstances in which TTV was discovered greatly influenced initial studies in the attempts to link the virus to a clinical disease. The temporal association between TTV viremia and the elevation of transaminases observed in transfused patients in whom TTV was demonstrated, as well as the apparent ability of TTV to replicate in the liver, led to the suggestion that the virus was a possible cause of acute and chronic liver disease, particularly those that have a putative viral cause but are unrelated to known viruses. Later findings disputed this correlation and, to date, although it cannot be excluded that TTV infection could be occasionally associated with occasional liver injury of varying severity, the virus is no longer considered an important cause of clinically overt liver disease.

Attempts to evaluate the possible role of TTV and other AV in the genesis of extrahepatic illnesses of unknown origin have, therefore, been explored. Given the chronic nature of most if not all TTV infections, most studies have been focused on chronic diseases such as diabetes, cryoglobulinemia, psoriasis, rheumatoid arthritis, systemic lupus erythematosus, Kawasaki syndrome, multiple sclerosis and other neurological diseases for which a viral etiology has long been suspected. In all cases, a direct role of TTV in disease etiology has been excluded or proved inconclusive. Because TTV infection is acquired early in life, other studies were performed to correlate TTV infection with acute pathologies of presumable viral origin occurring in infants. Here, it has been shown that TTV might behave as a respiratory virus and be involved in inducing or aggravating some acute pediatric respiratory diseases in which no other etiological agents were detected. Additional diseases for which a role of TTV has been sought in children are asthma and bronchiectasis. In fact, a correlation between TTV load in nasal fluid and severity of the associated respiratory deficit was observed in both these diseases. Other studies also implicated TTV or TTMV in a few renal ailments and in certain forms of anemia, neutropenia, and thrombocytopenia. However, until final proof of these or other findings, TTV and other AV should be considered “viruses waiting for a disease” or “disease orphans” in the same way that other viruses have been linked to a disease years after their discovery (*e.g.*, echoviruses, reoviruses, Epstein-Barr virus, *etc.*). So far, AV behave exactly as formerly orphan viruses and we are still seeking occasional infections sufficiently aggressive to trigger significant clinical disease. Additionally, it cannot be excluded that only some types of AV are pathogenic but their virulence is blurred by the plethora of non-pathogenic types. An unresolved question is whether AV modulate the host’s adaptive immune response. It has been demonstrated that TTV interacts systematically with many cellular sensors, including those modulating immune and inflammatory responses. Because of the small number of focused studies performed and the limitations of *in vitro* methods, detailed information on the mechanisms involved in AV-cell interactions and pathways involved is lacking. A few studies have showed that TTV ORF2 protein suppressed NFκB activity, which is crucial for expression of many genes connected to inflammation. It has also been shown that the TTV genome, as well as its replicative intermediate in infected cells, may influence the synthesis of pro-inflammatory cytokines evoked following stimulation of Toll-like receptors. Again, different TTV species have been shown to encode microRNAs, which are often sequestered within plasma exosomes of infected individuals. Interestingly, at least one TTV species encodes a microRNA *in vivo* that targets IFN signaling and, therefore, possibly modulates immune evasion by antagonizing the host antiviral response. Overall, these findings support the idea that TTV influences various biological processes and this leads to an elevated inflammatory status that is either generalized or in specific body sites. Further, perturbation of biological processes may be proportional to the type(s) and amount of TTV(s) circulating in the infected host.

From another perspective, put forward following discovery of TTV and its surprisingly high prevalence, the fact that no disease is associated with active infection and that the virus undergoes sustained and continuous replication in healthy individuals has led to the hypothesis that AV is devoid of pathogenic potential. This idea strengthened since initial observations to the point that TTV is now considered integral part of the normal human microflora. Further, TTV can be considered a stepping stone for reconsidering some key aspects of virology. Once thought to be only present in the host during disease, viruses have recently been demonstrated to be numerous in various body sites of healthy subjects, and the term virome has been coined to describe the collection of viral species present in a human organ. This sort of viral “flora” is an ensemble of “innocuous” entities such as bacteriophages, endogenous retroviruses, eukaryotic viruses not associated with diseases, and noxious viruses causing acute, chronic or latent illnesses. Thanks to NGS technology, the human virome has been studied in various compartments, such as respiratory tract, gut, and skin in healthy and sick conditions. To date it is known that some components of human virome are present only in a few areas and few individuals, others are disseminated in almost all body sites of a very high percentage of people. AV, and particularly TTV, are the paradigmatic example of the latter. They are the most representative and abundant viruses of the human virome and their load in plasma is proposed as a simple and general measure of immune system function of infected hosts. In keeping with this notion, it has been shown that TTV viremia increases after autologous hematopoietic stem cell transplantation (HSCT) but returns to pre-HSCT levels once the immune system regains its functionality. Thus, monitoring TTV viremia might prove useful in estimating the time interval between the chemotherapy courses required to minimize the occurrence of opportunistic infections. Again, the immunological status of solid organ transplant recipients and the shape and composition of the virome are closely connected. In transplanted patients, marked expansion of viral species of Anelloviridae family and high titer of virus indicate

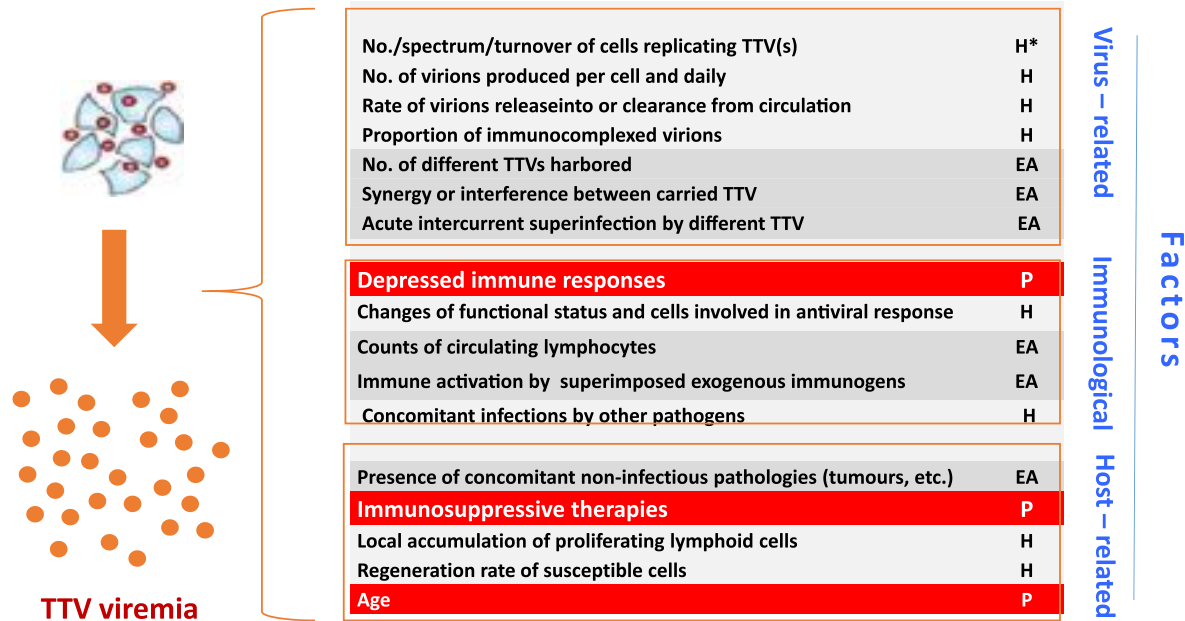


Fig. 2 Factors impacting the size of TTV viremia. P, proven; EA, some direct evidence available; H, hypothetical.

Table 2 Dynamic of chronic TTV viremia, relative to other viruses

<i>Virus</i>	<i>Mean virion half-life in plasma</i>	<i>New virions entering plasma per day</i>
Hepatitis B	28.8 h	$> 1.7 \times 10^{13}$
Hepatitis C	2.7 h	$> 1.3 \times 10^{12}$
HIV-1	< 1 h	$> 9.8 \times 10^9$
TTV	4.3 h	$> 2.0 \times 10^{10}$

strong immunosuppression and risks of opportunistic infection. Conversely, a lower-than-average burden of TTV is indicative of insufficient immunosuppression and risk of graft rejection.

Pathogenesis

A well-known feature of AV is the ability to establish, in many if not all exposed individuals, chronic productive infections characterized by continuous and abundant presence of the virus in the bloodstream. The course of infection is mostly unknown and the small amount of information available almost exclusively concerns TTV. Body sites and cell types where AV undergoes primary amplification following body entry as well as tissues sustaining its continuous replication and shedding in blood circulation are unknown. In children hospitalized for acute respiratory infections, higher TTV loads were measured in the nasal secretions compared to blood and, more interestingly and in some individuals, the same viral isolate was detected earlier in nares than bloodstream, suggesting that the respiratory tract is a site of primary TTV amplification. After its entry, TTV becomes detectable in the peripheral blood within one or a few weeks, indicating that dissemination of infection throughout the body is rapid. The most frequent outcome of TTV infection is a prolonged, likely lifelong TTV viremia sustained by genuine or, in a minority of cases, by a mix of viruses acquired through multiple exposures. Cases of self-limited infection have been also reported in which viremia spontaneously disappears a few weeks to several years post-exposure. It is unclear whether this is the result of an eradication of infection, decreased virus shedding into blood and/or increased virus clearance, or an infection entered in a latent inactive state. In chronically infected persons, the levels of plasma TTV vary extensively, ranging between 10^1 to over 10^9 DNA copies per ml. In many individuals, TTV viremia remains essentially stable over time, in others it undergoes wide fluctuations over time under the influence of various factors that may directly or indirectly impact on viremia size (Fig. 2). During replication, TTV is produced in large amounts: in kinetic studies of TTV viremia in patients treated with interferon for concomitant hepatitis C, it has been calculated that 10^{10} TTV virions are generated per day with nearly all virus entirely cleared and replenished in plasma on a daily basis (Table 2).

Remaining aspects of TTV interactions with the infected host are far less well characterized. For instance, the target tissues and organs from which TTV is continuously released into the bloodstream and the fate of cells sustaining viral replication are unknown. Almost all anatomical body sites have been found to harbor variably abundant TTV. These include liver, lung, pancreas,

spleen, kidney, lymph nodes, skeletal muscles, thyroid gland, bone marrow and circulating leukocytes. Only the cerebrospinal fluid is usually virus negative provided that the blood-brain barrier is intact. The presence of double-stranded viral DNA and viral mRNA, *i.e.*, two important cellular markers of TTV replication, has been reported in liver, bone marrow, and circulating peripheral blood mononuclear cells, but the methods to detecting such replicative intermediates have many pitfalls, thus leaving ample room for questioning the findings. Most findings support the idea that proliferating hematopoietic cells are an important source of circulating TTV. The idea is strengthened by the fact that baseline TTV viremia decreased to undetectable levels in patients under myelosuppressive treatments (*e.g.*, cyclophosphamide and total body irradiation) for bone marrow transplantation. Since neither viral receptor nor *in vivo* tropism are known, it cannot be ruled out that other cells sustain TTV replication and that different species of TTV and AV replicate in different cell types, tissues and body compartments. How rapidly and effectively the host immune system restrains an invading AV is also unknown. While nothing is known about anti-TTV cell mediated immune responses, current knowledge about humoral responses suggests that TTV elicits antiviral antibodies that fail to eradicate the virus, at least in the great majority of cases. The frequent identification of people infected by multiple and heterologous TTV species highlights the inability of anti-TTV specific antibodies to clear the virus or protect against subsequent infections. Sequential examination of patients accidentally infected through contaminated blood and of chimpanzees experimentally infected with human AV has shown that the antibody response mounts rather slowly. TTV-specific IgM become detectable in serum 10–21 weeks post-infection, 2–7 weeks after the first detection of TTV DNA in the blood. IgM declined in titer later on to disappear within 5–11 weeks, although longer persistence was reported. IgG antibodies generally developed with a lag of several weeks, increased in titer concomitant with IgM decline, and last with little variation for a prolonged period of time or even indefinitely.

Most if not all the TTV detected by molecular methods in the serum of chronically infected subjects is immunoprecipitated with anti-human IgG sera, indicating that TTV circulate in the blood bound to Ig. Conversely, in individuals with putative acute infections, free TTV particles in blood are abundant and often overwhelm the IgG complexed ones.

Diagnosis

Laboratory diagnosis of AV infection is still at infancy. This is due to the following reasons: (1) there are no reliable and sensitive tissue culture systems for AV isolation and propagation; (2) there are no easy-to-use immunological methods to detect and titrate AV induced antibodies, and this in part accounts for their unclear significance in the course of viral infection; (3) detection of viral antigens in plasma is also unavailable. Diagnosis of infections relies, therefore, on the detection of TTV genome in blood or other specimens. Here too there are some caveats: most PCR assays are incapable of detecting the entire spectrum of AV species, standardization is far from being achieved because each laboratory sets up its own test independently, there have been few inter-laboratory exchanges of methods and molecular constructs and there are no international standards. As a matter of fact, validated, *in-vitro* diagnostic labeled assays to detect and quantitate TTV DNA are badly needed but uncertainties about the clinical relevance of AV infection has discouraged most commercial companies from investing in this field.

Because of great genetic variability of AV, identification of the viral DNA segment to be targeted for amplification dictates the sensitivity and spectrum of viral species detection. This is true in general for all viruses and crucial for AV, where variability is an order of magnitude higher compared to other viruses. Historically, the first PCR protocols were designed to target the TTV ORF1 gene. Soon after their development, however, it became clear that they greatly underestimated the prevalence rates of the infection since PCR primers annealed to poorly conserved sequences. As reported above, sequence homology of the UTR is higher than ORF1 across the different TTV isolates, particularly for a segment that is shorter than 100 nucleotides, making it suitable for developing PCR assays capable of detecting all or nearly all the 29 species currently identified. Indeed, when compared to ORF1-PCRs, UTR-PCRs greatly increased the rates of samples scored positive for TTV DNA. It has however been pointed out that even UTR-PCR assays can be biased towards some species, unless primers are carefully designed. UTR-PCR assays capable of simultaneous detection of all TTMV and TTMDV species have not been developed at the moment.

It is evident that correlation of AV to a specific disease and/or to use the viruses as potential biomarkers of immune function is difficult without precise quantification of the virus. Recently developed real-time PCRs and droplet digital PCRs targeting TTV UTR and measuring the viral content in plasma and other biological samples hold promises to reduce the gap. TTV viremia as determined with UTR-PCRs is 10- to 100-fold higher compared to ORF1-PCRs, reflecting the higher sensitivity of former assays as well as detection of TTV species poorly or not amplified by latter. The lower limits of detection of quantitative PCR assays are on average 10–100 DNA copies per ml of blood, a threshold that is likely sufficient, based on available information and because of the high levels of viremia established for most diagnostic purposes.

TTV genotyping exploits genomic sequences informative for classification and, for this purpose, ORF1 gene is an apt target as it permits grouping of all AV by either sequencing of the whole gene or using genotype-specific primers in standard PCR assays. The concomitant presence of multiple AV, a frequent occurrence in most subjects, is a confounding factor that impairs reliability and applicability of genotyping assays. NGS technology is a valid alternative for precise genotyping.

In the quest for establishing routes of transmission and disease association it is crucial to extend examination to other clinical specimens. TTV DNA has been detected in feces, saliva, nose swabs, throat swabs, and breast milk of viremic subjects. Of note, some biological fluids scored positive also in persons with no detectable viremia.

Treatment and Prevention

As long as their clinical implications remain undefined, there will be little attention paid to the prevention and control of AV infection. It is also clear that, based on knowledge accrued so far, there is little doubt that prevention of infection is extremely difficult in the absence of specific polyvalent vaccines. There are no drugs of proven efficacy for AV infection, and the current lack of information on AV pathogenicity does not justify specific clinical trials. In patients treated with IFN α alone or in association with ribavirin for HBV or HCV infections TTV viremia did not change or decreased only transiently. Antiretroviral drugs have no anti-TTV activity. Specific antivirals are likely to be developed should one day AV infection be considered clinically relevant and worth treating.

Further Reading

- De Vlaminc, I., Khush, K.K., Strehl, C., *et al.*, 2013. Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* 155, 1178–1187.
- Fernández-Ruiz, M., Albert, E., Giménez, E., *et al.*, 2019. Monitoring of alphatorquevirus DNA levels for the prediction of immunosuppression – Related complications after kidney transplantation. *American Journal of Transplantation* 19, 1139–1149.
- Jaksch, P., Kundi, M., Görzer, I., *et al.*, 2018. Torque teno virus as a novel biomarker targeting the efficacy of immunosuppression after lung transplantation. *Journal of Infectious Diseases* 218, 1922–1928.
- Lim, E.S., Zhou, Y., Zhao, G., *et al.*, 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nature Medicine* 21, 1228–1234.
- Maggi, F., Bendinelli, M., 2009. Immunobiology of the torque teno viruses and other anelloviruses. *Current Topics of Microbiology and Immunology* 331, 65–90.
- Maggi, F., Focosi, D., Statzu, M., *et al.*, 2018. Early post-transplant torquetenovirus viremia predicts cytomegalovirus reactivations in solid organ transplant recipients. *Scientific Reports* 8, 15490.
- Ninomiya, M., Nishizawa, T., Takahashi, M., *et al.*, 2007. Identification and genomic characterization of a novel human torque teno virus of 3.2 kb. *Journal of General Virology* 88, 1939–1944.
- Nishizawa, T., Okamoto, H., Konishi, K., *et al.*, 1997. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochemical and Biophysical Research Communications* 241, 92–97.
- Okamoto, H., Nishizawa, T., Kato, N., *et al.*, 1998. Molecular cloning and characterization of a novel DNA virus (TTV) associated with posttransfusion hepatitis of unknown etiology. *Hepatology Research* 10, 1–16.
- Spandole, S., Cimponeriu, D., Berca, L.M., Mihăescu, G., 2015. Human anelloviruses: An update of molecular, epidemiological and clinical aspects. *Archives of Virology* 160, 893–908.
- Strassl, R., Schiemann, M., Doberer, K., *et al.*, 2018. Quantification of torque teno virus viremia as a prospective biomarker for infectious disease in kidney allograft recipients. *Journal of Infectious Diseases* 218, 1191–1199.
- Takahashi, K., Iwasa, Y., Hijikata, M., Mishiro, S., 2000. Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. *Archives of Virology* 145, 979–993.

Animal Lentiviruses (*Retroviridae*)

Esperanza Gomez-Lucia, Complutense University of Madrid, Madrid, Spain

© 2021 Elsevier Ltd. All rights reserved.

Glossary

CAP Methylated guanosine linked to the first nucleotide of the RNA, essential to initiate protein synthesis, protect RNA from the exonuclease cleavage, and recruit protein factors for several functions.

CD4⁺ T-lymphocyte (or T-cell) T-cell identified by the CD4 marker associated to the TCR (T-cell receptor for the antigen), which enables the cell to recognize antigen processed by antigen processing cells (macrophages, monocytes, dendritic cells). After recognition, the T-cell becomes activated and can collaborate with other cells of the immune system, such as B-cells (for antibody production) and other T-cells (for cell-mediated immune responses).

Chemokine Chemotactic molecule which stimulates inflammation. Chemokines are usually produced by the injured tissues, and attract neutrophils, monocytes, macrophages, lymphocytes, dendritic cells, eosinophils, and basophils.

Endogenous lentivirus Lentivirus which infected sexual cells (sperm or ovum) and became part of the germline, being passed from generation to generation throughout thousands or millions of years. Usually endogenous lentiviruses have acquired mutations which avoid them from productive replication (replication deficient).

Enhancer Sequences in the 5' non-coding region of the genome which are recognized by cellular factors and trigger RNA transcription.

Exogenous lentivirus Lentivirus which is transmitted in a contagious manner, either after budding and being recognized by receptors, or by cell-to-cell contact. Usually it is replication competent.

Immunodeficiency State in which the immune system is unable completely or in part to fight infectious disease and cancer.

Lipid rafts Microdomains in the plasma membrane less fluid than the surrounding lipid bilayer, enriched with cholesterol, saturated phospholipids and membrane proteins.

Macrophage Cell derived from blood monocytes, with high phagocytic ability, which resides in tissues. It participates both in the innate and adaptive immunities.

Promoter Region at the 5' end of genes which triggers RNA transcription for protein synthesis.

Provirus DNA genome of retroviruses, after the viral single-stranded RNA is reverse transcribed into double-stranded DNA and integrates in a host chromosome.

Introduction

Lentiviruses are grouped in a genus within the family *Retroviridae*. The name derives from the Latin word *lenti* which means slow, alluding to the long incubation period of the disease, which in most cases lasts years. During this period the virus seems to be controlled by the immune system but steadily replicates and increases in numbers. When the viral load is too high and the damage to the different systems, mainly the immune system, grows unendurable, the disease becomes apparent and the infection turns fatal. The discovery that acquired immune deficiency syndrome (AIDS) was caused by a lentivirus, human immune deficiency virus (HIV), prompted the study of lentiviruses in other species. Up-to-date lentiviruses have been identified to infect humans and other primates, cats, cattle, sheep, goats and horses. Recently lentiviruses have been shown in rabbits and ferrets. Most animal lentiviral infections are notifiable diseases to the World Organization of Animal Health (OIE). All lentiviruses have similar biochemical, morphological and antigenic properties. They share common features regarding virion particle structure, genetics, infection kinetics, tropism, immune response elicited, transmission, etc. In general, lentiviruses infect immune cells (mainly macrophages and T-lymphocytes) and induce several types of immune dysfunctions (macrophage subset differentiation, T-cell anergy, etc.), inflammation, and organ/tissue impairments. As in other retroviral infections, infected cells carry the retroviral genome (provirus) throughout their life and pass the infection to their progeny. Unlike other viruses, lentiviruses may efficiently infect non-dividing cells.

Classification

Genus *Lentivirus* is one of the six genera of the subfamily *Orthoretrovirinae* within the family *Retroviridae*. The genus, that traditionally contained seven members, has recently expanded to 10 after the addition in the 2018 release of the International Committee on Taxonomy of Viruses (ICTV) of Puma lentivirus (PLV), Jembrana disease virus (JDV), and HIV-2. All lentiviruses are genetically related and must have derived from a single ancestor, which has co-evolved with the hosts that originated from the initial one. Thus, lentiviruses may be classified according to the hosts they infect. Primate lentiviruses include human immune deficiency virus (HIV) type 1 (HIV-1) and type 2 (HIV-2), as well as simian immune deficiency virus (SIV), which has been identified in many different non-human primates with genomic differences between them. Small ruminant lentiviruses (SRLV) include visna-maedi virus (VMV) of sheep and

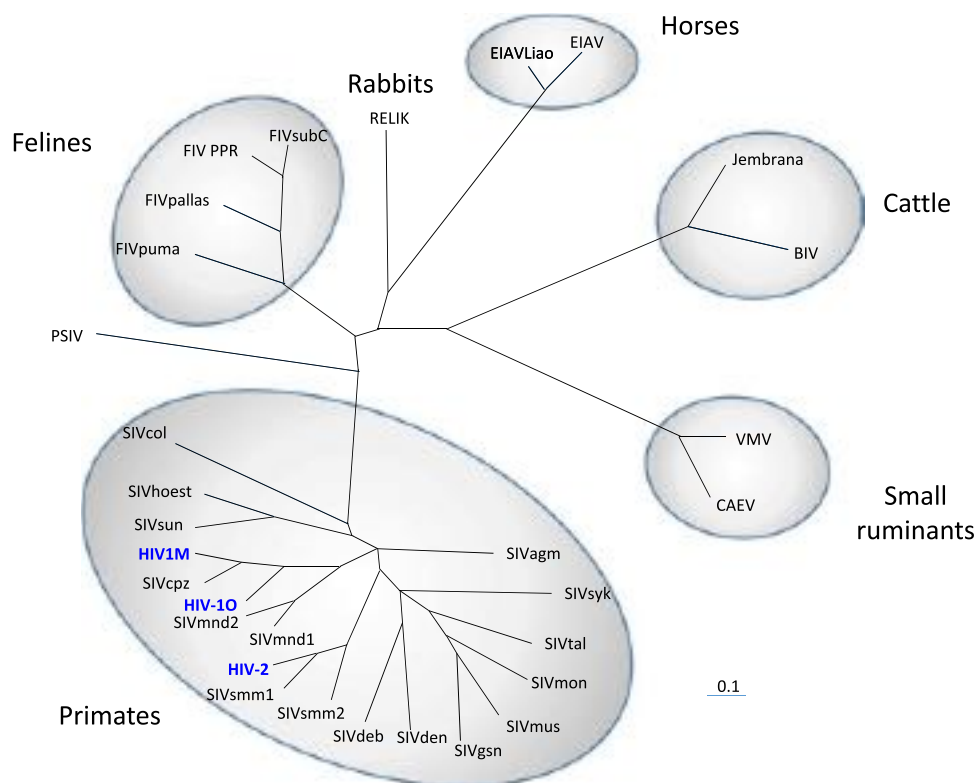


Fig. 1 Phylogenetic tree of the genus *Lentivirus* after alignment. Names of the viruses are shown in [Table 1](#). HIV-1 and HIV-2 are in blue to more easily locate their phylogenetic relationships. RELIK is an endogenous lentivirus found in the rabbit genome.

caprine arthritis encephalitis virus (CAEV) of goats. Cattle lentiviruses had a single representative, bovine immunodeficiency virus (BIV), but as mentioned above, now also the closely related JDV is considered. As of now, lentiviruses of carnivores have been shown to infect felids exclusively and include feline immunodeficiency virus (FIV) and PLV. Lastly, the lentivirus of horses is named equine infectious anemia virus (EIAV) ([Fig. 1](#)). Within the primate lentiviruses, for example, HIV-1 and HIV-2 differ in that the sequences only present no more than 50% homology between them, and that HIV-1 has *vpx* while HIV-2, *vpu*. This reflects that HIV-1 and HIV-2 originated from different ancestors, HIV-1 from chimpanzees and gorillas, and HIV-2 from sooty mangabey. As lentiviruses are genetically related, they have certain cross-reactivity between the Gag proteins, but cross-reactivity is non-existent between proteins codified by the gene *env*. For a long time lentiviruses had been considered to be only exogenous or contagious, but recently the germlines of leporids (rabbits and hares) and ferrets have been shown to carry endogenous lentiviruses, remnants of ancient infections which have lost the ability to produce infectious virions (see below).

Virion Structure

As in the rest of the retroviruses, lentiviral particles are rounded. The size is slightly larger than in other retroviruses, 100–130 nm in diameter; SIV particles can reach even 150 nm in diameter. They are enveloped, with short projections distributed evenly on the envelope, composed of two proteins: surface proteins (SU), the outermost, and transmembrane (TM), which anchors SU to the envelope. Inside the virion electron micrographs reveal a core shaped like a truncated cone or rod, composed of capsid proteins (CA). The matrix proteins (MA) fill the space between the envelope and the core to give consistency to the viral particle. The capsid contains two identical helical nucleocapsids, each consisting of a molecule of RNA, surrounded by the nucleocapsid proteins (NC) and associated with the proteins that are involved in the early stages of viral replication: the protease (PR), reverse transcriptase (RT) and integrase (IN), as well as Nef, Vif and Vpr in the lentiviruses that have them ([Fig. 2](#)).

Genome

Both molecules of single-stranded RNA are identical. They are 7700–11,000 nucleotides (nt) long ([Table 1](#)) and are linked by hydrogen bonds. They have a CAP sequence at the 5' end and a polyadenylated tail around 200 nt long in the 3' end. Regardless of these characteristics, they do not act as mRNA. As in other retroviruses, lentiviruses possess three main genes, *gag* (which encodes the inner proteins), *pol* (which encodes the proteins for replication, protease, reverse transcriptase and integrase) and *env* (which

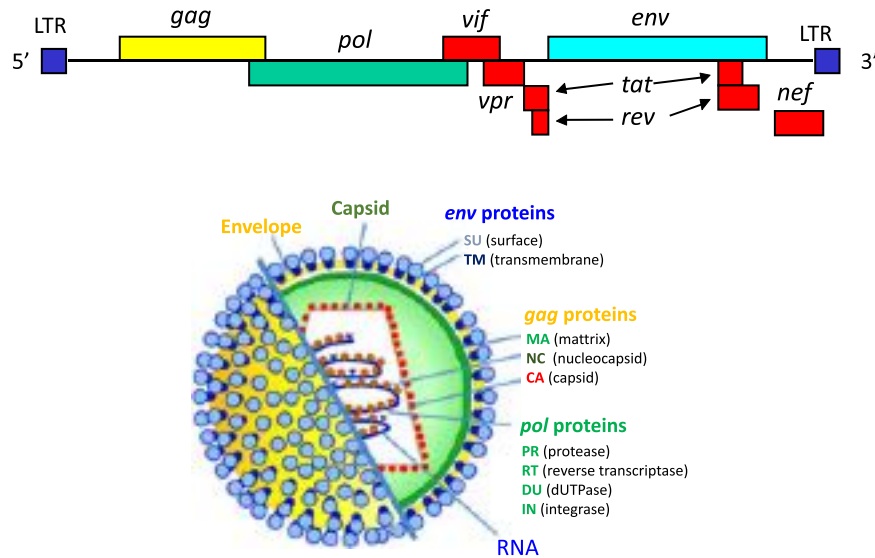


Fig. 2 Schematic representation of the HIV-1 genome and a particle of lentivirus.

Table 1 Species of lentiviruses recognized by the International Committee on Taxonomy of Viruses (ICTV) in the 2018 release (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/reverse-transcribing-dna-and-rna-viruses-2011/w/rt_viruses/161/retroviridae). GenBank Accession Numbers of the prototype strains or that corresponds to the complete sequence of the virus, total length of the genome (expressed in nucleotides, nt) and accessory genes present in each species are shown

Virus	Complete name	Abbreviation	GenBank	nt	Accessory genes
Carnivores	Feline immunodeficiency virus	FIV	EU117992	9891	<i>rev response element, vif, orfA</i>
	Puma lentivirus	PLV	NC_038669	9100	<i>vif</i>
Cattle	Bovine immunodeficiency virus	BIV	M32690	8482	<i>tat, vif, rev, vpw, vpy, tmx</i>
	Jembrana disease virus	JDV	NC_001654	7732	<i>tat, rev, vif, tmx, Jdvgp5</i>
Small ruminants	Visna-Maedi Virus	VMV	NC_001452	9202	<i>vif, rev, vpr-like</i>
	Caprine arthritis and encephalitis virus	CAEV	NC_001463	9189	<i>vif, rev, vpr-like</i>
Primates	Human immunodeficiency virus type 1	HIV-1	K02007	9737	<i>tat, rev, vif, vpr, vpu, nef</i>
	Human immunodeficiency virus type 2	HIV-2	NC_001722	10,359	<i>tat, rev, vif, vpr, vpx, nef</i>
	Simian immunodeficiency virus	SIV	KU892415	10,276	<i>tat, rev, vif, vpr, vpx, nef</i>
Horses	Equine infectious anemia virus	EIAV	M16575	8407	<i>tat, rev, S2</i>

encodes the external glycoproteins). Besides these genes, lentiviruses have genes the products of which regulate the synthesis and processing of the viral RNA (*tat* and *rev*), and other accessory genes (*vif*, *vpr*, *vpu*, *nef*, *tmx*, etc.) that play an important role in the viral replication, in counteracting the restriction factors of the host and in pathogenesis (Fig. 3). HIV and SIV are the lentiviruses with the highest number of accessory genes, followed by BIV (Table 1). Several of these accessory genes may have similar functions to other genes with different names in other lentiviral species, so the list may be misleadingly long. For example, BIV has *tat*, *rev*, *vif*, *vpw*, *vpy*, and *tmx* genes, the last three with functions similar to HIV-1 *vpr*, *vpu* and *nef*, but with limited sequence similarity to the human virus counterparts. The role in pathogenesis of several of these genes has not been ascertained unequivocally or they may have several functions carried out by single genes in other lentiviruses. This is the case of FIV *orfA*, which acts by increasing the net translation of viral gene products similar to HIV Tat, but it does not increase transcription or use a Tar element in the same way as Tat. This gene also shares similar functions to *vpu*, *vpr* and *nef* of HIV-1. Animal lentiviruses, except for HIV and primate lentiviruses, also encode a protein involved in regulating cellular dNTP ratios, the viral dUTPase, although BIV encodes a dUTPase-related gene without enzymatic activity.

At each end of the viral RNA there are the same ~ 100 nucleotides, a sequence known as repeat region (R). In addition, at the 5' end there is a unique region about 70 nucleotide long (U5), and at the 3' one of about 450 nucleotides long (U3). These regions are essential to start reverse transcription, integration and post-integration transcription. During the process of reverse transcription these unique regions are duplicated, forming two identical structures that are known as long terminal repeats (LTR), with a structure U3-R-U5 (Fig. 4). In the stage of provirus synthesis (when the lentiviral genome is double-stranded DNA) LTRs flank the genome at both ends. The 5' end of the R region in the 5' LTR is where transcription begins (+1 transcription). The U3 region of the 5' LTR is extremely important because it effectively controls transcription. In this region there are two main areas: a promoter area and an



Fig. 3 Proviral genomic structure of most common animal lentiviruses, showing genes *gag*, *pol*, and *env* and accessory genes. Several accessory genes are encoded in non-consecutive sequences and are expressed following splicing of the RNA.

enhancer area. The promoter area is closer to the beginning of transcription, and exhibits a TATA box 20–50 nt upstream of the cap site, +1 of transcription. The TATA box is present in the promoter region of many eukaryotic genes. Some lentiviruses also have a CCAAT box, another promoter sequence. The enhancer region is further upstream from the beginning of transcription and contains numerous sequences that are recognized by cell signaling factors. Upon activation of these enhancers, the promoter area triggers the beginning of the synthesis of RNA and DNA (Fig. 4). These sequences which bind cell signaling factors are collectively known as transcription factor binding sites or TBS, and include AP-1, AP-4, ATF, and many others. In some cases there is a TBS for NF- κ B or for LBP-1. In different lentiviruses, a variety of other sites exist, these include a gamma-activated site or GAS (FIV, HIV and CAEV), an interferon stimulated response element (ISRE, in FIV or HIV), interferon response factor-1 (IRF-1, CAEV), and hormone response element (HRE, FIV and HIV). The significance of these TBS is that viruses are silent in non-activated cells, but upon activation, the signaling process culminates in the entrance of molecules to the nucleus that can stimulate lentiviral replication and production of viruses. The other two regions of the 5' LTR, R and U5, play a primary role during the reverse transcription process. The LTRs have different sizes depending on the lentiviral species, ranging from about 311 nt (PLV) to 855 nt (HIV-2).

The region between the 5' LTR and the beginning of *gag* is defined as a non-coding region. It contains the primer binding site (PBS, 18 nt long complementary to the tRNA that functions as primer in the reverse transcription process), the packaging signal (ψ), and the leader region (L), which precedes the first coding sequence. This non-coding region serves as a donor for splicing processes that give rise to different subgenomic mRNAs. All retroviruses contain a sequence rich in purine, or polypurine tract (PPT) immediately before the U3 region of the 3' end. The PPT contains the site for initiating the positive strand of the viral DNA synthesis.

Encoded Proteins

Proteins encoded by essential genes are similar in all lentiviruses. External or surface proteins are usually glycosylated (at least SU), while inner proteins usually are not. Both external and inner proteins are very immunogenic, but the antibodies formed are not

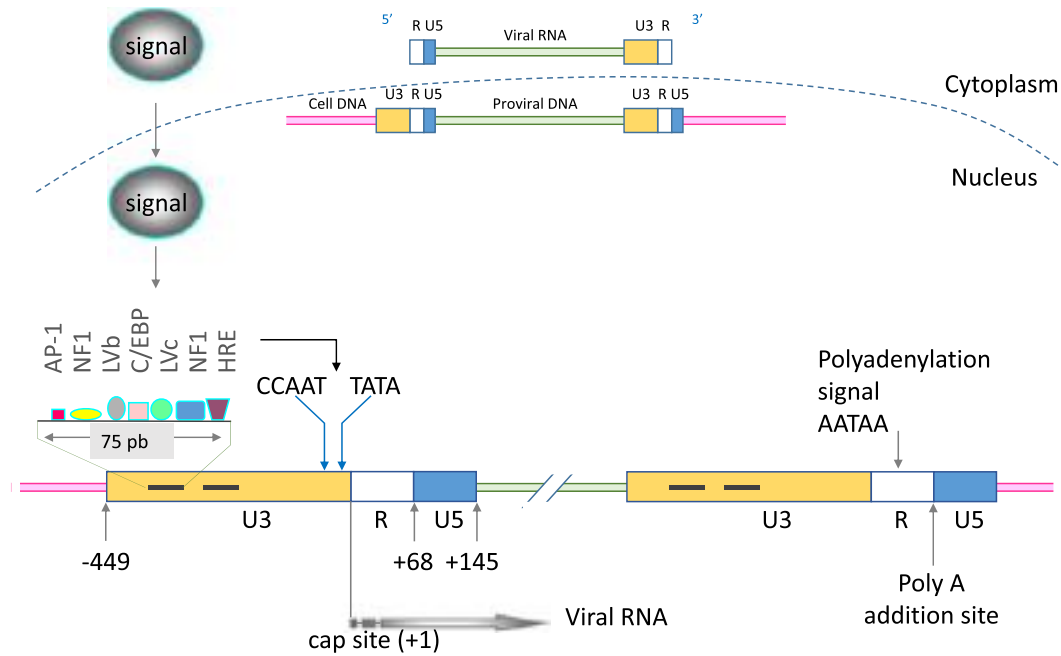


Fig. 4 Schematic representation of viral RNA, proviral DNA and LTRs. The viral RNA contains identical R sequences at both ends, and a unique U5 in the 5' end, and U3 in the 3' end. This U5 and U3 become duplicated during the reverse transcription to give the LTRs, formed by identical U3-R-U5 at both ends. This viral DNA, known as provirus, integrates in the cell genome. There, it remains silent till cell signals cross the nuclear membrane and react with specific sequences in the U3 of the 5' LTR known as transcription factor binding sites or TBS (some of them are shown). This activates the promoter region of the LTR, leading to transcription from the 5' end of R or cap site.

able to eliminate the infection. The inner proteins are relatively conserved and can exhibit cross-reactions between different lentiviruses. They are named according to their weight (**Table 2**). Mature proteins are cleaved from polyprotein precursors. Accessory proteins and their functions are listed in **Table 3**.

Replication Cycle

The replication of retroviruses took time to be elucidated in part because the concept that RNA could be transcribed into DNA went against the accepted dogma that genetic information flows from the DNA to RNA, and from there to proteins. Unlike other retroviruses, which replicate in dividing cells, lentiviruses are preferentially expressed in differentiated cells. Some receptors and co-receptors have been identified, but many other receptors have not been identified yet. For example, despite the similar pathogenesis of HIV and FIV, unlike HIV the latter does not employ CD4 as a receptor, but CD134, a member of the family of receptors for TNF- α and nerve growth factors. In the case of SRLV it is likely that mannosylated residues in the Env protein are sensed by the mannose receptor (CD206), a C-type lectin endocytic receptor expressed on macrophages and dendritic cells. While the receptors for BIV or JDV have not yet been identified, there is evidence that BIV may use the C-C chemokine receptor 5 (CCR5), the main co-receptor used by the majority of strains of HIV-1 for infection of cells.

After the recognition between SU protein and its receptor, there are conformational changes that motivate the exposure of a hidden domain that in the case of HIV allows interaction with the co-receptors, CCR5 or CXCR4. This causes further changes that expose the fusion domain of TM, with the resulting fusion of the virus envelope with the cell membrane or of the endosome, allowing the entry of the viral particle (**Fig. 5**). Another entry mechanism can be in cells with receptors for the Fc domain of immunoglobulins (FcR) that trap antibody-coated particles (e.g., VMV). The penetration process concludes with the release of the core into the cytoplasm. The host cell has mechanisms to prevent this process, including TRIM5 α , which the virus has to circumvent. Once in the cytoplasm, the RT, likely in the context of a partially disrupted capsid, begins to transcribe the retroviral genome. First, thanks to the RNA dependent DNA polymerase activity (RdDp), it transcribes a single-stranded (minus sense) DNA molecule. As the RNA-DNA hybrid moves through the RT its RNase H activity degrades almost simultaneously the original viral RNA which served as template and through the DNA dependent DNA polymerase (DdDp) activity, an additional molecule of RT transcribes the complementary strand of nascent DNA, synthesizing a molecule of double-stranded DNA. In a complex process that involves duplication of regions from the 5' and 3' ends of the viral genome, LTRs that flank the provirus (DNA retroviral genome) are generated. Vif and Nef are important at this early stage, a reason for being present in the viral particle. Nef increases the effectiveness of RT, so that larger amounts of reverse transcribed DNA are produced, and Vif blocks the cell cytidine deaminase APOBEC3 that would increase the rate of error of the RT.

Table 2 Gene products of the lentiviruses, some of their functions and presence in some lentiviral species. In lentiviruses, the number that accompanies the name of the proteins represents its approximate weight in kDa. -, not present. Blank cells, no specific name. Names of the viruses are shown in [Table 1](#)

Gene	Protein	Function	HIV-1	FIV	BIV	VMV	EIAV
<i>gag-pol</i> <i>gag</i>	Gag-Pol	Precursor of the viral structural proteins and enzymes	Pr160		Pr170	Pr175	
	Gag	Precursor of the structural proteins of the virus	Pr55	Pr50	Pr53	Pr55	
	MA	Matrix. Inner membrane layer. Confers rigidity and stability to the virion. Myristilated	p17	p15	p16	p16	p15
	CA	Capsid. Encapsulates the nucleocapsids. Intervenes in packaging. Synthesized in large quantities and its detection is used for diagnosis in some lentiviral infections	p24	p24	p26	p25	p26
	NC P6	Nucleocapsid. Surrounds and protects RNA Protein for viral encapsidation and budding of assembled particle	p7 p6	p10 -	p7 -	p14 -	p11 p9
<i>pol</i>	PR	Protease: Proteolytic cleavage of Gag and Gag-Pol to produce structural proteins and mature virion	p12	p14		p10	
	RT	Reverse transcriptase. Triple function: RNA dependent DNA polymerase (RdDp), DNA dependent DNA polymerase (DdDp) and RNase H to give the end result of double-stranded DNA from single-stranded RNA	p66/p51	p65		p68	
	dUTPase IN	Integrase. Integration of proviral DNA in the host genome	- p32	p14 p31	-	p14 p35	yes
<i>env</i>	PrEnv	Precursor of the viral envelope proteins	gp160		gpp145	pr185	
	SU	Surface protein on the viral envelope located in the outside of the virion. Interaction with receptors. Main domain for receptor binding. Attachment of virus to target cell. Determinant of tropism. Immunogenic, stimulates the synthesis of specific antibodies. It interacts with the extracellular domain of the TM. Always glycosylated	gp120	gp95	gp100	gp135	gp90
	TM	Transmembrane protein. With a cytoplasmic domain and a hydrophobic domain within the cell membrane. The interaction of TM with the lipid bilayer allows the anchoring of the SU-TM complex to the viral envelope. Mediates fusion between viral envelope and cell membrane. Induces syncytia. Immunosuppressant. Frequently glycosylated	gp41	gp40	gp45	gp46	gp4

Table 3 Regulatory and accessory genes, proteins encoded, their function and viruses in which they are present. CD4, molecule on surface of T-helper cells. MHC-I, major histocompatibility complex type I. Full names of the viruses are shown in [Table 1](#)

Gene	Protein	Function	Virus
<i>tat</i>	Tat	Potent transactivator that may greatly stimulate viral expression. Phosphoprotein located in the nucleus and nucleolus of the infected cells. Elongation of mRNA and transcription	BIV, CAEV, EIAV, HIV-1, HIV-2, JDV, SIV, VMV
<i>rev</i>	Rev	Regulates transport, processing and translation of viral mRNAs. Phosphoprotein located in the nucleus and nucleolus of infected cells.	BIV, CAEV, EIAV, HIV-1, HIV-2, JDV, VMV
<i>nef</i>	Nef	Endocytosis of CD4 and MHC-I at the cell membrane. Increase in the efficacy of RT, generating higher amounts of retrotranscribed DNA. Increase in the viral infectivity. Cell activation. Induction of apoptosis. Downregulation of E2 ubiquitin conjugating enzyme	HIV-1, HIV-2, SIV
<i>vif</i>	Vif	Enhancement of viral infectivity. Stabilization of the pre-integration complex. Blockage of the cell restriction APOBEC3G (cytidine deaminase) which increases the error rate of RT	BIV, CAEV, FIV, HIV-1, HIV-2, JDV, SIV, VMV
<i>vpr</i>	Vpr	Arrest of the cell cycle at G2. Transport of the pre-integration complex to the cell nucleus	HIV-1, HIV-2, SIV
<i>vpu</i>	Vpu	Enhancement of virion release. Degradation of CD4 and MHC-I	HIV-1
<i>vpv</i>			HIV-2, SIV
<i>vpw</i>			BIV
<i>vpy</i>			BIV
<i>tmx</i>	Tmx	Similar function to Nef	BIV, JDV
<i>orfA</i>		Similar functions to Tat, Vpu, Vpr and Nef. Arrest of the cell cycle at G2. Downregulation of E2 ubiquitin conjugating enzymes	FIV
<i>Jdvdp5</i>		Unknown	BIV
<i>S2</i>		Similar function to Nef	EIAV

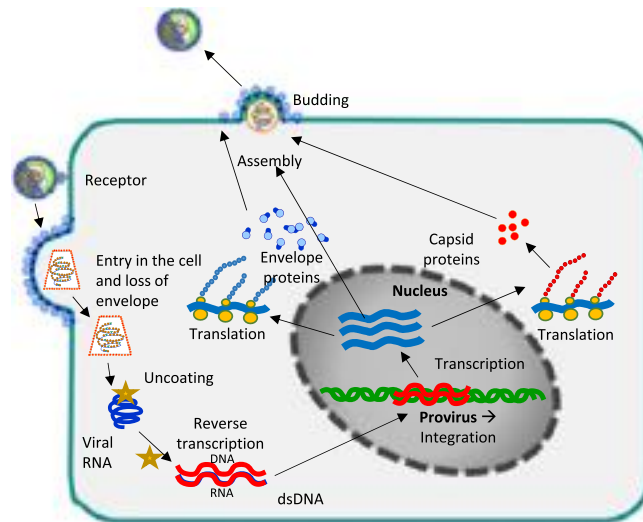


Fig. 5 Schematic representation of the replication cycle of lentiviruses. Reverse transcriptase (RT) is shown in the shape of a star.

Proviral DNA, in the form of a pre-integration complex (PIC), crosses the nuclear membrane and, using the viral enzyme integrase present in the PIC, it is integrated into a host cell's chromosome. Vpr (also present in the virus particle) plays an important role in this process, directing the nuclear localization of the pre-integration complex and interfering with the cell cycle, stopping it in the G2 phase to facilitate integrase-mediated integration of the proviral DNA into the cell genome.

Once integrated, proviral DNA behaves as a Mendelian gene: when the cell divides, the progeny cells will contain the provirus. For this reason, lentiviral infections are lifelong, especially taking into account the cells that they infect, most of them indispensable for the immune response. Although integration is random, HIV-1 prefers to integrate within genes, while other retroviruses prefer exons or non-coding sequences. Integration appears to be a prerequisite for the infectivity, and although there is extra-chromosomal proviral DNA, it is not expressed. The integrated provirus can remain without being expressed for months or years, until the appropriate signaling molecules cross the nuclear membrane. They recognize specific short sequences of the U3 region of the 5' LTR, the TBS mentioned above, triggering the transcription of the lentiviral genome by the cellular RNA polymerase II, as if it were a common cellular gene. Therefore, lentiviruses are not expressed in resting cells and have to wait till these are activated, remaining latent till this occurs. This allows them to wait for adequate conditions to continue their replication cycle.

Integrated DNA serves as a template for the cellular RNA polymerase II to transcribe both spliced subgenomic mRNA, which will be translated to the characteristic polyproteins, and full-length viral RNA, which will be packaged into the viral particle. Accessory genes are expressed first, especially *tat*, *rev*, and *nef*. When the products of these genes reach appropriate levels, Rev facilitates the export of unspliced and partially spliced transcripts to the cytoplasm and translated in the form of polyproteins.

Of the proteins encoded by *env*, at least SU (and usually also TM) undergoes N-linked glycosylation. Both SU and TM are inserted into discrete microdomains in the plasma membrane related to lipid rafts. The internal proteins of the virus undergo limited modification (fatty acylation at N-terminus of MA) and accumulate at the plasma membrane and associate with the proteins SU and TM, in order to initiate the process of budding and assembly. Once concluded, the immature particles are released by budding. The PR cleaves the polyproteins inside the virions, giving rise to the mature virus particles. During these final stages, proteins such as cell MHC-I or CD4 are redirected to the degradation route, so they are not expressed on the cell surface, where they could interfere with virus release.

Reverse Transcription

The process of retrotranscription copies the information of the viral genome, contained in the molecule of single-stranded RNA, into a molecule of double-stranded DNA. During this process, the RT performs three enzymatic functions: DNA-dependent RNA-polymerase (RdDp), Ribonuclease H (RNase H), and DNA-dependent DNA-polymerase (DdDp). The RT activity lacks exonuclease activity (repairing) and, in addition, it has a high frequency of error when incorporating nucleotides, which makes that the final error rate of the RT ranges between 3×10^{-3} and 3×10^{-5} , depending on the retrovirus. The error rate of RT is higher with RNA templates than with DNA templates and varies according to the context of the reading. A characteristic of the activity of the RT is that throughout the process it switches from one strand of nucleic acid to the other, and may make mistakes at this time, allowing recombination.

Reverse transcription begins with the hybridization of the tRNA, which acts as the primer, with the PBS close to the 5' end of the viral genome (Fig. 6). From here, it extends the strand in the direction 5' until it reaches the end of R. RNA that forms part of the newly formed RNA-DNA hybrid is digested by the RNase H activity of the RT. The 3' end of the newly synthesized DNA and, now as a single strand, hybridizes with the R region of the 3' end of the viral RNA. This process is called the first strand transfer. RdDp

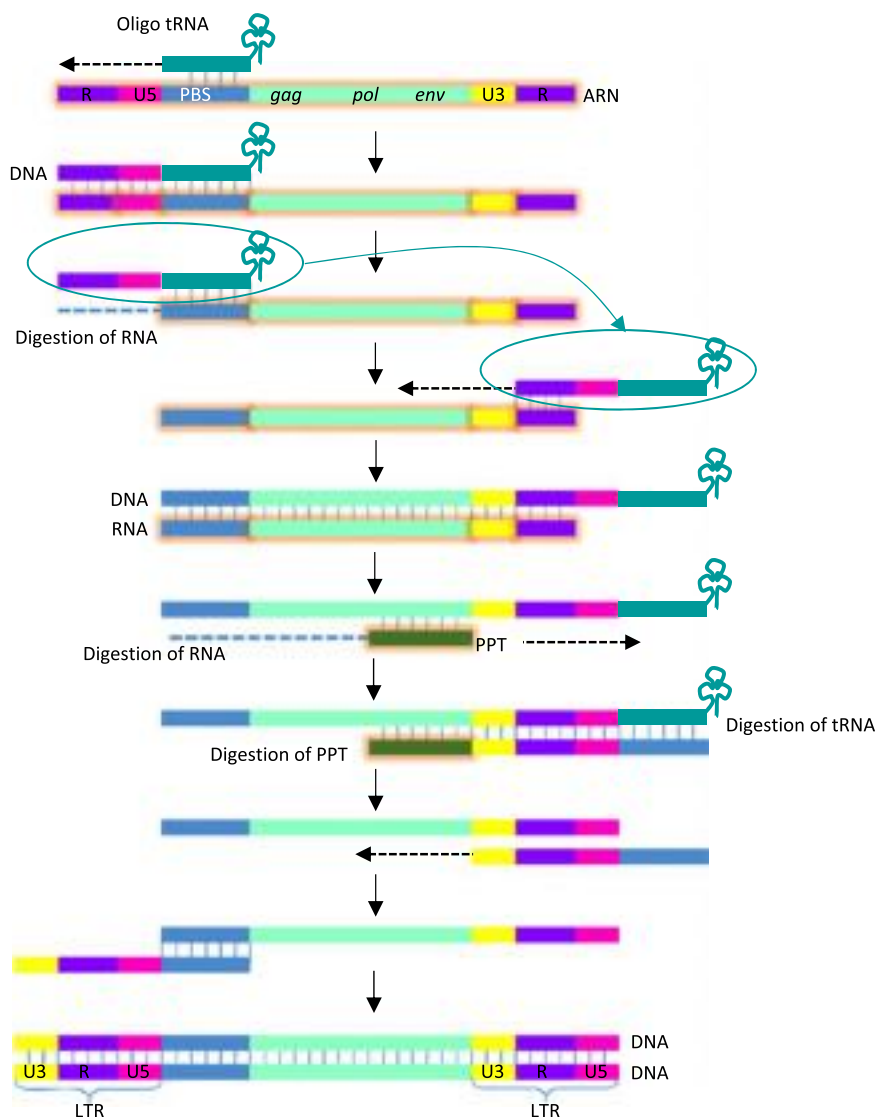


Fig. 6 Process of reverse transcription. The segments delimited by an orange outline represent RNA.

activity continues to extend the DNA molecule all the way towards the 5' end. RNase H activity degrades viral RNA 5' → 3', but is unable to degrade the polypurine tract (PPT) adjacent to the U3 region of the template RNA. This PPT functions as the primer to start DNA synthesis using DNA as template, in a 3' direction. This synthesis is extended by the primer, by copying a new PBS. RNase H activity removes the last vestiges of RNA, both in the area of PPT as in the primer. This creates two complementary PBS areas in each chain of DNA, allowing the second strand transfer to occur. As a new site is formed by where the DNA can be transcribed, RT, through its DdDp activity, synthesizes a strand of DNA complementary to the already synthesized. DNA synthesis continues until RT completes the complementary DNA, concluding the synthesis of the 5' LTR.

Biological Properties

Variability

Lentiviruses can achieve diversity both within individual hosts and within populations through spontaneous mutation rates ranging from 10^{-3} – 10^{-6} substitutions per site per year due to low polymerase fidelity and lack of proofreading capacity of RT. In addition to this remarkable mutation rate, lentiviruses can also create diversity by recombination, because RT needs to switch template from one strand of nucleic acid molecule to the other during the process of transcription (see above). These switches can match the nascent strand with non-corresponding sequences, causing recombination. This occurs in all lentiviruses, but the frequency of crossovers and factors promoting recombination can be variable. From the perspective of the lentivirus, variability has several advantages. Both mutations and recombination increase viral genetic diversity in fluctuating environments; the most

obvious implication of this is the development of escape mutants that allow the viral population to avoid the selective pressure applied by the host immune response or antivirals. Recombination is a mechanism to prevent the accumulation of deleterious mutations in the population, which would decrease its fitness. The extraordinary variability of some of the lentiviruses poses a major obstacle to vaccine development and drug usage. This, together with the persistent presence of lentiviruses in the infected cell, makes it very difficult to fight against these viruses.

Natural Hosts

As mentioned above, lentiviruses derive from a common ancestor that infected its host millions of years ago, and co-evolved with it as the progeny of this host diverged genetically. For this reason, related viruses may infect closely related species. Thus, it is possible that sheep are infected by CAEV and VMV by goats. It has also been shown that HIV-1 and HIV-2 come from SIV. SIV underwent several inter-species jumps; in the case of HIV-1, from chimpanzees and gorillas to humans, and in the case of HIV-2, from the sooty mangabey to humans. However, the mechanisms by which viruses from other animal species are able to infect a new one after these inter-species jumps remain unknown. In the case of FIV, although it usually is not passed between cat species due to the elusive and territorial nature of these animals, variants of the virus in different wild species have been isolated, including pumas, Asian leopard cats, ocelots, bobcats, jaguarundis, Pallas cats, cheetahs, leopards, lions, snow leopards and spotted hyenas. VMV can also infect wild ruminant species, such as moufflons. SIV infects a very high number of simian species, and it has been determined that up to 40 of the 69 species of African non-human primates act as natural hosts of SIV, including African green monkeys and mandrills.

Transmission

As other retroviruses, the lentiviral virions are sensitive to many physical and chemical agents including desiccation, temperature, acidic or basic pH values, etc. For these reasons they are usually transmitted in “short distances” by direct methods. Lentiviruses are usually exogenous viruses. This means that they are transmitted horizontally or in a contagious way, through cell-to-cell contact, as well as to the progeny of the infected cell. However, in the rare occasions in which a germ cell, a spermatozoid or an ovum, is infected, the provirus becomes integrated into the genome and is transmitted vertically from generation to generation of the animal, according to the patterns of Mendelian inheritance. These retroviruses that remain in the genome over millions of years are called endogenous retroviruses. Till recently, it was thought that lentiviruses had no capacity for germline integration. However, endogenous lentivirus-like viruses have been described recently in rabbits and lemurs, called RELIK and PSIV, respectively. This supports the concept that lentiviruses are more widespread than previously thought, can be incorporated into the host germline, and are ancient, existing prior to 12 million years ago. The phylogenetic relationships of RELIK (which is also found in hares) are shown in [Fig. 1](#).

For some lentiviruses, the most common horizontal transmission routes are the sexual route, the direct inoculation in tissues or bloodstream, through syringes with contaminated blood and bites (FIV and SIV), or during childbirth when the virus enters the body of the neonate through mucosal surfaces (HIV). In some cases lentiviruses can be transmitted by colostrum or milk, although not in SIV, possibly due to the low levels of CCR5 in the newborn. Some lentiviruses (i.e., BIV, HIV) can cross the placenta and infect in utero. EIAV is exceptional amongst not only lentiviruses but also retroviruses in that the transmission is performed by arthropod vectors (horseflies and deerflies). In JDV arthropods are also partly responsible for transmission.

Tropism

In many cases, target cells are those involved in the immune response. Some of the lentiviral species are lympho-tropic (infecting lymphocytes), others are macrophage-tropic (infecting cells of the monocyte/macrophage lineage, including in some cases, microglia, such as VMV, CAEV or EIAV), while others have a mixed tropism (infecting both lymphocytes and macrophages, such as SIV, HIV and FIV). Within the lympho-tropic species and strains, some infect T-helper cells ($CD4^+$), cytotoxic T-cells ($CD8^+$), $T\gamma\delta$ cells (common in cattle), null cells or B lymphocytes (e.g., BIV and JDV). Infected T-lymphocytes present an abnormal pattern of cytokine secretion, with increased expression of IFN and certain interleukins, such as IL-12 and IL-18. Some lentiviruses can also infect dendritic cells, which are key in immune responses. CAEV and VMV infect monocytes, but are not expressed until the cells mature to macrophages in the target organs.

Pathogenic Processes

Lentiviral infections usually follow a similar course, with an initial acute transitory stage, a latency stage in which the immune system is able to control the infection, and a final stage consisting of a new phase of fast replication, which determines the typical syndrome. However, many lentiviral infections pass unnoticed in animals because the virus does not produce any clinical sign. Asymptomatic infected animals can transmit the infection (carriers); the infection and the disease become apparent when the virus infects a different population. Typically, the acute stage, which develops 4–6 weeks after infection, is characterized by fever, lymphadenopathy, leukopenia (due frequently to neutropenia) and weight loss. This acute stage can last for a few weeks up to several months. This is followed by an asymptomatic stage, in which the animal appears to be healthy or with minimal clinical signs, which may last several years, during which the virus replicates at low rate, and the immune response is able to control it, but

producing a progressive impoverishment of the affected cell population. Finally, the immune response is unable to control the virus and the disease becomes apparent, with a significant increase of the viral load, an important deterioration of the immune system, immune deficiency with presence of respiratory, digestive, and possibly neurological clinical signs, as well as weight loss and hyperthermia, which lead to the death of the animal.

Immune Deficiency

Immune deficiency is one of the most common pathologies produced by lentiviruses. Several species include in their name this term, such as HIV-1, HIV-2, SIV, FIV and BIV, although in the latter the direct role of the virus in the clinical disease in infected cattle under natural conditions has not been shown clearly, and its true economic impact is hypothetical. Immune deficiency develops because many lentiviruses infect lymphocytes, especially T-cells. In the course of infection, CD4⁺ T-cells (T-helper cells) decrease, either by the direct action of the virus or due to the antiviral defense mechanisms. This decrease in CD4⁺ T-cells is exploited by opportunistic infections that become established and impair the clinical status of the animal, being responsible for its death in the majority of cases.

Immune deficiency is often accompanied by hematological abnormalities, mainly anemia, characteristic of FIV and EIAV infections. It is produced because viral surface glycoproteins coat red blood cells or erythrocytes, triggering complement-mediated destruction of these cells through a mechanism of type II hypersensitivity. Another consequence of immune deficiency is the development of tumors, which can be produced by an uncontrolled growth of cells (lymphoma).

Some of the alterations that appear as a consequence of immune deficiency are located on the skin and oral cavity (gingivitis), respiratory and gastro-intestinal tracts. Frequently the liver, kidneys and central nervous system are impaired.

Inflammation

VMV and CAEV produce mainly inflammation; processes associated with BIV and JDV also seem to be more related to inflammation than with immune deficiency. This could be due to the earlier sacrifice of cattle than of the other animal species mentioned in the paragraph of immune deficiency. The inflammation directly attributed to infection by CAEV or VMV is located in the mammary gland, joints, lungs or central nervous system (meningoencephalitis), prevailing one over the other depending on the infecting virus. As in the other lentiviruses, CAEV and VMV can remain latent for many months with little or no expression. The myriad of mechanisms that regulate viral replication and lead to increased viral loads and more serious outcomes of disease remain unclear. However, it is recognized that specific cell transcription factors activate viral gene expression when interacting with specific proviral DNA sequences included in the LTR, the transcription factor binding sites (TBS) mentioned above.

SRLV constant replication in tissue and the subsequent immune response produce chronic inflammation causing pathological changes observed in the target organs of animals infected by the virus. As a result, the multisystem inflammatory disease that characterizes SRLV infection has an immunopathogenic origin. The main change in the affected tissues is infiltration by mononuclear cells (lymphocytes, macrophages and plasma cells). Cells accumulate and organize in lymphoid follicles. The final phase of the pathogenesis begins when the clinical disease appears. Depending on the clinical form, the animal can die rapidly or remain chronically infected.

The inflammatory processes triggered by lentiviruses are associated with the production of cytokines, which can regulate viral transcription through the cellular signaling pathways mediated by the TBS in the proviral LTR as explained above. For example, cytokines IFN- γ and TNF- α , present in arthritic joints of animals infected by CAEV, have shown to activate viral transcription via recognition of the corresponding gamma activation site (GAS) and TNF activation site (TAS) present in the U3 of CAEV.

Cell Defenses and Viral Mechanisms to Evade Them

When attacked by viruses, cells display innate response mechanisms to avoid their productive infection. Some of the cellular mechanisms to fight against lentiviruses are TRIM5 α , APOBEC3 or tetherin. TRIM5 α induces a premature uncoating of the viral core, by recognizing the incoming viral capsid, inducing its proteasome-dependent degradation. This impairs integration and significantly reduces proviral load and viral production. APOBEC3 is a cytidine-deaminase that causes uracil accumulation in negative-strand nascent DNA, leading to detrimental G-to-A mutations in the proviral DNA. This molecule may be incorporated into progeny virions in producer cells (passenger APOBEC3) and inhibit lentiviral replication in the following replication cycle in target cells. Lastly, tetherin is able to trap virions on the surface of infected cells, preventing virus release, and therefore, virus transmission between cells. These mechanisms are very specific against particular lentiviruses, and this innate immune response is responsible for preserving the species barrier.

In response, lentiviruses have developed mechanisms to counteract these restriction factors. For example, the high genetic heterogeneity found within the capsid region is likely a way to escape TRIM5 α . In HIV, FIV, BIV, SRLV and EIAV infections, the viral protein Vif has the main function of avoiding APOBEC3 incorporation into viral particles, restricting the probability of incorporating wrong nucleotides. Vif also triggers the degradation of most APOBEC3 proteins. Vpu is particularly dedicated to bind tetherin and degrade it via proteasome. This so-called "evolutionary arms race" has led to the selection of somatic genes encoding virus-interacting proteins. So far, these innate mechanisms have been partially unveiled in the simian or human counterparts, and equine, ovine, bovine, and feline lentiviruses are being actively studied to decipher the important restriction mechanisms in these species.

Animal Lentiviral Infections

Feline Immunodeficiency and Puma Lentiviral Infections

Feline immunodeficiency virus was initially isolated from cats (*Felis catus*) in Petaluma (CA, USA) in 1986. As mentioned above, besides cats, other felids, such as cheetah, lions, leopards or hyenas may be infected by the virus. The most recent release of the ICTV (2018) recognizes two different viral species, highly related to FIV, identified in mountain lions (*Puma concolor*) and bobcats (*Lynx rufus*), which are known as puma lentivirus A (PLVA) and B (PLVB). However, the list of FIV-related lentiviruses in wild felids is increasing quickly. FIV is considered a good animal model for the study of HIV, as its biochemical and morphological properties are very similar, though antigenically they are distinct. Also, though both HIV and FIV infect CD4⁺ T-cells, the latter enters cells through CD134, a member of the TNF- α and growth factor receptors. CD134 is highly expressed in active lymphocytes, monocytes/macrophages and dendritic cells. FIV is mostly transmitted through direct inoculation of the virus through bites, the reason why it is most prevalent in adult male cats. After an initial period of 4–6 weeks with no signs, cats develop an acute phase with hyperthermia, lymphadenopathy, leukopenia (mostly neutropenia), and weight loss. This is followed by a long asymptomatic phase which may last many years. In fact, 12% of the healthy cat populations are considered to be unapparent carriers. Infected cats may die in old age of etiologies unrelated to FIV-infection. However, some develop clinical immune deficiency with increased viral loads due to the immune impairment, accompanied with respiratory and intestinal clinical signs, gingiva-stomatitis, hyperthermia, etc., which constitute the so called feline acquired immune deficiency syndrome or FAIDS. FIV, rather than opportunistic bacteria, may be directly responsible for the intestinal dysfunction, as it may cause blunting of the villi and dilation of the crypts. In contrast to FIV, lentiviruses of wild felids have not historically been associated with significant pathogenicity. However, several captive African lions have manifested clinical signs or hematologic abnormalities associated with lentiviral infection.

A commercial vaccine is available to prevent FIV infection in cats, but its use remains controversial, including whether or not the vaccine effectively induces cross-protective immunity against the various naturally circulating virus strains.

Bovine Immunodeficiency (BI) and Jembrana Disease (JD)

BIV was first isolated in 1969 in the USA from a progressively deteriorating 8-year old pregnant Holstein dairy cow (*Bos taurus*), identified as R-29, with wasting syndrome, lymphadenopathy, leukocytosis, lymphocytosis, fatigue, diminished milk yield and neurologic disorders. Serology has shown that BIV is present worldwide; however, the direct role of BIV in clinical disease among cattle infected under natural conditions is not clearly demonstrated and the true economic impact of BIV-infection remains conjectural. However, an acute disease in Bali cattle (*Bos javanicus*), Jembrana disease, endemic to parts of Indonesia, was described in the 1990s and is thought to have arisen from BIV via an unknown transmission event. The clinical signs observed in BIV-infection are more related to the inflammatory events characteristic of VM or CAE (see below) than immune deficiency itself. BIV targets monocytes and macrophages, inducing some immune dysfunction such as decreased superoxide anion production, phagocytic activity and chemotaxis. On the other hand, JDV has a high morbidity and mortality of up to 20%. JD is characterized by an acute stage 5–12 days post infection, with fever, lymphadenomegaly, lethargy, anorexia, panleukopenia, splenomegaly, hemorrhages and very high viral loads. There is no recurrence of disease in recovered animals but viral loads remain high during at least 2 years. JDV is often described as an atypical lentivirus, because of the low level of viral variation in vivo, the delayed antibody response, and the ensuing immunological control which prevents subsequent heterologous infection. There is an inactivated tissue-derived vaccine against JDV to control the disease in Indonesia.

Visna-maedi (VM) and Caprine Arthritis and Encephalitis (CAE)

VM was first described in Iceland in the 1930s. VMV was isolated from cases of interstitial pneumonitis (*maedi* = shortness of breath in Icelandic) and demyelinating leukoencephalomyelitis (*visna* = wasting in Icelandic). CAEV was first isolated in 1980. They are collectively called Small Ruminant Lentiviruses (SRLV). Transmission is related to body fluids, mainly respiratory exudates and milk or colostrum. Both VM and CAE are prevalent worldwide, although the prevalence is much higher in developed countries, which seems to be related to the custom of feeding lambs or kids with milk pooled from all the mothers, which favors transmission, and to the intensive rearing system. Both viruses can infect sheep and goats and cross-infections have been observed both experimentally and naturally. Nevertheless, the infection by VMV is more frequent in sheep and the infection by CAEV in goats. SRLV have in vivo tropism for monocytes, macrophages and dendritic cells, but may also infect cells of the central nervous system and others. As with other lentiviral infections, there is an initial short phase of viremia. Infected monocytes and macrophages spread the infection. The replication in monocytes and macrophages does not take place until these cells mature in the target organs. The infection remains latent for a period of time which depends on the viral strain and mainly on the individual susceptibility, and the animal can die in a short time or remain chronically infected. The continual replication of SRLV in tissues and the ensuing immune response causes a chronic inflammation that originates the pathological changes observed in target organs of SRLV-infected animals. Consequently, as in other lentiviral infections, the multisystem inflammatory disease that characterizes SRLV infections is immune-pathogenic in nature. The final stage of the pathogenesis begins when the clinical disease starts. Clinical forms of both VM and CAE are respiratory, nervous, mammary and joint disease. Although normally the process is

subclinical, a small percentage of the animals can display one or more of these forms or even all four of them. Sheep usually develop respiratory and mammary clinical signs, while goats display nervous and joint forms.

Equine Infectious Anemia (EIA)

EIA is considered a worldwide disease that occurs only in members of the family Equidae. It was first identified as an infectious disease of horses by veterinarians in France in 1843, and shown to be produced by a filterable agent in 1904. The disease is different from other retroviral diseases in that it is transmitted mostly by insect vectors (especially horseflies and deerflies), but also by hypodermic needles and other veterinary instruments. Sexual transmission of EIAV has not been demonstrated to date. The cell target during persistent EIAV infections appears to be exclusively cells of the monocyte/macrophage lineage, especially macrophages resident in tissues, such as liver, kidney, spleen. Viremia observed during episodes of chronic EIA result primarily from the production and release of EIAV from infected tissue macrophages, rather than an extensive infection of blood monocytes.

Field isolates and laboratory strains of EIAV differ markedly in their pathogenicity; processes can be asymptomatic to lethal. Characteristically, the course of the infection is initially acute (first month), followed by chronic disease (first year) to progress into an asymptomatic lifelong stage thereafter. Horses suffer acute and chronic disease, while donkeys and mules typically experience asymptomatic infections. Acute disease is often associated with the first exposure to the virus, with fever, thrombocytopenia and hemorrhages evident from 7 to 30 days after exposure, though anemia or edema is not seen at this point. After this first stage, horses become clinically healthy but recurring episodes of the disease appear at irregular intervals, ranging from weeks to several months. Stress or immunosuppression trigger these viremia and reactivation episodes. Typically, horses will develop weight loss, anemia, diarrhea, and edema. Most of these signs are thought to be the combined effect of immune-mediated lysis of virus-infected cells and an immune complex-mediated inflammatory response. During the first year after infection, clinical episodes in horses with chronic EIA are more severe and frequent, and they usually disappear thereafter, and the horse becomes a lifelong inapparent carrier.

Currently there is no effective vaccine for the prevention of EIAV infection and disease, mostly due to the large antigenic diversity of the virus.

Further Reading

- Bhatia, S., Patil, S.S., Sood, R., 2013. Bovine immunodeficiency virus: A lentiviral infection. *Indian Journal of Virology* 24, 332–341. doi:10.1007/s13337-013-0165-9.
- Burrell, C.J., Howard, C.R., Murphy, F.A., 2017. Retroviruses. In: Fenner and White's Medical Virology. Elsevier, pp. 317–344. doi:10.1016/B978-0-12-375156-0.00023-0.
- Hosie, M.J., Addie, D., Belák, S., *et al.*, 2009. Feline immunodeficiency. ABCD guidelines on prevention and management. *Journal of Feline Medicine and Surgery* 11, 575–584. doi:10.1016/j.jfms.2009.05.006.
- ICTV, 2018. International Committee on Taxonomy of Viruses, Available at: <https://talk.ictvonline.org/taxonomy/>.
- Murphy, B., 2017. Retroviridae. In: Fenner's Veterinary Virology. Elsevier, 269–297. doi:10.1016/B978-0-12-800946-8.00014-3.
- Payne, S., 2017. Family retroviridae. *Viruses*. 287–301. doi:10.1016/B978-0-12-803109-4.00036-2.
- Ryu, W.-S., 2016. Molecular Virology of Human Pathogenic Viruses, first ed. Elsevier Science. Available at: <https://www.elsevier.com/books/molecular-virology-of-human-pathogenic-viruses/ryu/978-0-12-800838-6> (accessed 8.03.19).
- Stover, M.G., Watson, R.R., 2015. Animal lentiviruses. In: Health of HIV Infected People. Elsevier, pp. 349–365. doi:10.1016/B978-0-12-800767-9.00020-0.

Animal Morbilliviruses (*Paramyxoviridae*)

Carina Conceicao and Dalan Bailey, The Pirbright Institute, Pirbright, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Classification

The Morbillivirus genus belongs to the *Paramyxoviridae* family in the order *Mononegavirales*. Within the genus there are now seven recognized species (summarised in Table 1). For the purpose of clarity and concordance with the published literature, we will refer to individual viruses using their commonly used abbreviations (Table 1). The morbilliviruses are genetically closely related, with the most distant, formally classified, virus being feline morbillivirus (FmoPV) (Table 1 and Fig. 1). At the genomic level nucleotide identities vary between 53% (FmoPV and measles virus, MeV) and 76% (canine distemper virus, CDV and phocine distemper virus, PDV). Current theories on the evolution of morbilliviruses are summarised excellently in Nambulli *et al.* (2016) (see "Further Reading"). Of particular interest, rinderpest virus, RPV (or its ancestor) is thought to have emerged following the domestication of cattle. This ancestral virus is also thought to be the source of MeV, following zoonotic transmission to humans. It is well known that MeV spread to the North and South Americas in the 1500s causing devastating epidemics in indigenous populations; however, it is less widely appreciated that CDV is thought to have originated in the Americas, and then spread to Europe following colonisation. PDV may in turn have originated from spillover of CDV into aquatic mammals, perhaps as recently as the twentieth century. The recent identification of a wealth of morbillivirus sequences in bats and rodents now indicates that the established morbilliviruses may have actually emerged from related small mammal viruses. The close molecular relationship of morbilliviruses is also evident in their related antigenicity, with various viruses being able to provide cross-protective immunity.

Virion Structure

Pleomorphism is a feature of morbilliviruses, which translates to the production of particles ranging from 150 to 500 nm in diameter. These viruses are surrounded by a lipid bilayer derived from the host cell decorated with the viral glycoproteins haemagglutinin (H) and fusion protein (F), which can be seen as extruding spikes on the surface of morbillivirus virions (Fig. 2(A)). Inside the viral envelope, the matrix protein M forms a highly ordered lattice that interacts with the lipid bilayer, supporting the virion and packaging the viral genome within. The morbillivirus genome consists of a non-segmented, single-stranded, negative sense RNA molecule encapsidated by the nucleoprotein N in a helical nucleocapsid (Fig. 2(A)). The RNA-dependent RNA polymerase (RdRp) termed the large (L) protein and its co-factor, the phosphoprotein (P), associate with the N-RNA complex, forming the ribonucleoprotein complex (RNP). These RNP complexes assemble into a helical

Table 1 Host range, epidemiology and phylogeny of morbillivirus species identified to date

Species (2018-onwards)	Former species name (pre 2018)	Common abbreviation (For isolates)	Typical host	Sub-species classification	Geographic location	Genome length; % identity to MeV	Example sequence accession number
Canine morbillivirus	Canine distemper virus	CDV	Domestic dogs/ Carnivores	10 lineages (see Section Epidemiology and Host Range)	Global	15,690 nt; 61%	NC_001921.1
Cetacean morbillivirus		CMV	Cetaceans (whales, dolphins)	Previously defined as two species: Porpoise and dolphin morbilliviruses	Uncharacterised (likely global)	15,702 nt; 64%	NC_005283.1
Feline morbillivirus		FmoPV/ FeMV	Domestic cats	Not defined to date	Uncharacterised (likely global)	16,050 nt; 53%	JQ411014.1
Measles morbillivirus	Measles virus	MeV/MV	Humans	19 genotypes identified to date, only 6 currently circulating ^a	Global	15,894 nt	NC_001498.1
Phocine morbillivirus	Phocine distemper virus	PDV	Pinnipeds (seals)	Not defined to date ^b	Uncharacterised (likely global)	15,696 nt; 60%	NC_028249.1
Rinderpest morbillivirus	Rinderpest virus	RPV	Cattle	Lineages 1–3 (see Section Epidemiology and Host Range)	Extinct	15,882 nt; 69%	NC_006296.2
Small ruminant morbillivirus	Peste-des-petits-ruminants virus	PPRV	Sheep and goats	Lineages 1–4 (see Section Epidemiology and Host Range)	Northern and sub-Saharan Africa, Middle East, Asia	15,948 nt; 65%	NC_006383.2

^aCirculating genotypes are B3, D4, D8, D9, G3 and H1 (CDC, <https://www.cdc.gov/measles/lab-tools/genetic-analysis.html>).

^bThe 1998 and 2002 outbreaks may form two distinct lineages.

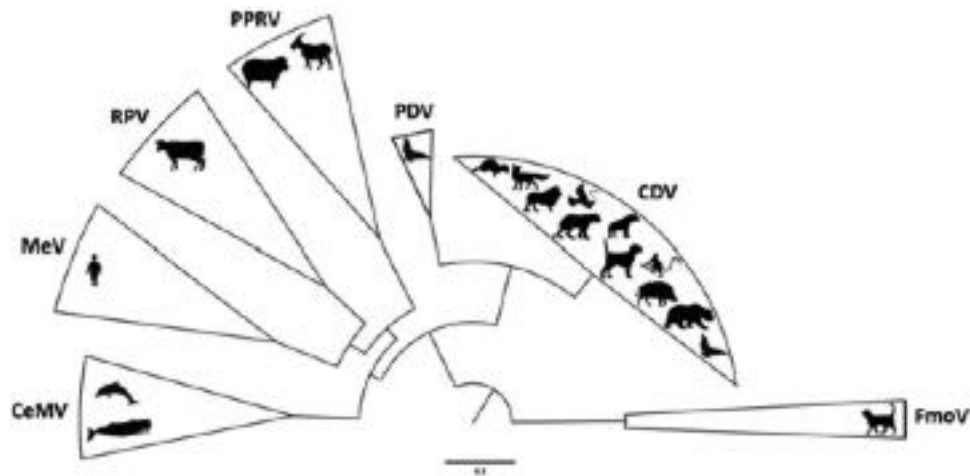


Fig. 1 Genetic diversity and ecology of the morbilliviruses. Rooted phylogenetic tree of the morbilliviruses, constructed using a neighbour-joining algorithm from full-length H sequences with at least 4 representative sequences of each species. Each branch corresponds to the diversity found within the sequences analysed with the genetic distance between branches indicated (0.2). The host silhouettes indicate the naturally infected hosts where disease has been recorded.

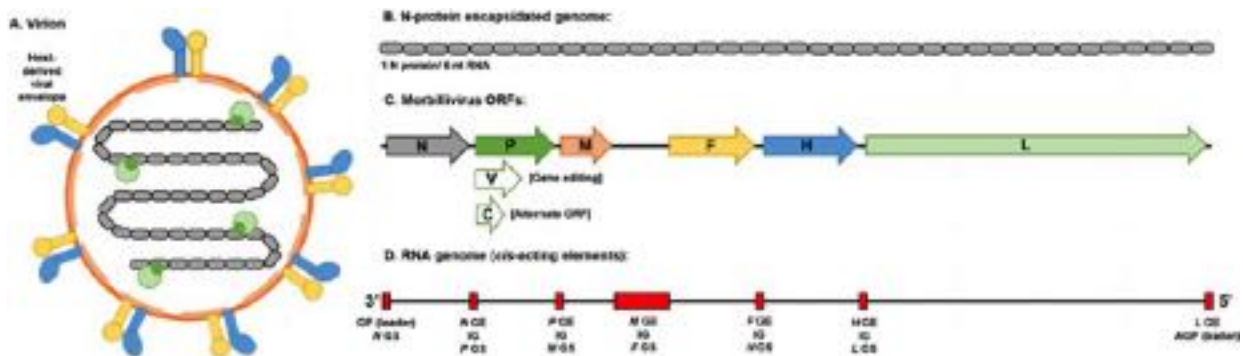


Fig. 2 Virion structure and genome composition of the morbilliviruses. (A) Schematic representation of morbillivirus virions illustrating the virally-encoded glycoproteins embedded in the host-derived envelope. Inside the M protein supports the viral envelope and interacts with the ribonucleoprotein complex (RNP), composed of the RNA genome encapsidated by the N protein, the viral polymerase L and its co-factor, the P protein. (B) Morbillivirus genomes strictly adhere to the “rule of six”, with one monomer of N protein encapsidating 6 nucleotides (nt) of RNA. (C) The morbillivirus genome comprises six genes that encode for six structural proteins (3′ - N, P, M, F, H and L - 5′) as well as two non-structural proteins V and C, expressed from the P gene, by gene editing and through use of an alternative open reading frame (ORF), respectively. The ORFs and intergenic regions are drawn to scale. (D) The RNA genome of morbilliviruses is composed of seven cis-acting elements that define the gene start (GS), IG and gene end (GE) for each of the six genes. At the 3′ and 5′ ends are conserved and complementary sequences that constitute the genome promoter (GP) for the production of viral mRNA and the antigenome promoter (AGP) for the production of full-length RNA genome, respectively.

configuration forming a characteristic herringbone-like structure, visible by electron microscopy. The RNP is the minimum required unit to initiate transcription and replication of viral RNA *in vitro*. Infectious morbillivirus virions possess at least one RNP; however, it has been shown for MeV that more than one RNP can be incorporated without affecting viral fitness.

Genome

Morbillivirus show variable genome lengths ranging from 15.5 to just over 16 kb (Table 1). Genome length conforms to the ‘rule of six’, since each N monomer encapsidates six nucleotides of RNA. The rule of six ensures the whole viral genome is encapsidated by the nucleoprotein, as this would otherwise negatively affect transcription and replication activity by the RdRp (Fig. 2(B)). Interestingly, PPRV was shown to not strictly conform to this ‘rule of six’ as two nucleotide insertions, or one nucleotide deletion at the 5′ end of a minigenome, did not severely affect the RdRp activity. However, the hexameric genome length of morbillivirus is prominently evident in nature, as demonstrated by a novel variant of PPRV that possessed a 6-nucleotide insertion in its genome,

when compared with other strains of PRRV. All morbillivirus genomes are composed of six genes (3' *N, P, M, F, H* and *L* 5') which encode for eight proteins (**Fig. 2(C)**). The two non-structural proteins C and V are encoded within the P open reading frame (ORF), by alternative translation initiation and RNA editing, respectively. All genes are separated by untranslated regions (UTRs), which contain in the following order, gene-stop signals (including polyadenylation signals) at the end of the upstream gene, a non-transcribed trinucleotide-intergenic region (IG) and lastly, gene-start signals for the expression of downstream genes (**Fig. 2(D)**). In morbilliviruses the IG is always 3 nt; however, the whole UTR region between genes can be up to 200 nucleotides in length, except for the UTR between the M and F ORFs, which can be over 1000 nucleotides. This long UTR is unique to morbilliviruses and is thought to play a role in mRNA translation, glycoprotein expression and virulence, as reduction in its size resulted in CDV attenuation *in vivo*. There are also conserved and complementary sequences at the 3' and 5' end of the genome, used as promoters by the RdRp (**Fig. 2(D)**). Preceding the N ORF at the 3' end is a conserved 52-nucleotide long leader sequence, followed by an IG and the 3' UTR of the *N* gene. This 109-nucleotide long region constitutes the genome promoter (GP) that directs the production of viral messenger RNA (mRNA) or full-length antigenome (positive sense RNA), which subsequently serves as the template for the production of nascent full-length genomic RNA (negative sense). Similarly, at the 5' end of the genome there is a conserved 37-nucleotides long trailer sequence that together with the 5' UTR of the *L* gene comprises the antigenome promoter (AGP), which directs production of full-length genomic RNA from the antigenome. Of note, the recently identified FmoPV possesses a 400-nucleotide long trailer sequence, giving it the largest morbillivirus genome identified to date (16,050 nucleotides). Although relatively well conserved, these GP and AGP sequences are thought to be species-specific; however, weak interactions can occur between the RNP of one virus and the GP and AGP of another, as was shown for RPV and PPRV using surrogate minigenome reporter assays. Even minor changes in sequence can have significant consequences on viral replication *in vivo*, e.g., exchanging the GP and AGP between a vaccine and virulent strain of RPV significantly altered pathogenesis in infected cattle.

Proteins

The nucleoprotein is the most abundant structural protein in morbillivirus particles, since significant amounts are required to encapsidate the full length RNA genome during the formation of the RNP. PPRV and RPV encode a 525-amino acid (aa) long protein whereas the N protein of CDV and aquatic morbilliviruses (cetacean morbillivirus, CeMV and PDV) is 523-aa long. All paramyxovirus N proteins can be divided into two main regions: the N_{core} located at the N-terminus and the N_{tail} located at the C-terminus. The N_{core} is a conserved structural domain containing the regions for self-assembly, RNA binding and P binding, while the disordered N_{tail} is involved in mediating RdRp activity *via* interactions with the N_{core} and the RdRp co-factor P. The importance of N_{tail} was illustrated by reports that truncated forms reduced the pathogenesis of recombinant CDV viruses *in vivo*. The predisposition of this protein to create nucleocapsid-like structures can also be observed when the protein is expressed in isolation in different systems. A closer look at these structures revealed that N_{core} is the main domain responsible for formation of nucleocapsid structures, as mutations impaired N-N assembly and RNA packaging.

The functional RNP complex is also composed of the phosphoprotein P, which both acts as a chaperone for the N protein by interacting with N_{core} to maintain solubility in the cytoplasm but also as cofactor for the RdRp L polymerase. While PPRV encodes a 509-aa long P protein, the respective RPV and PDV proteins are 507-aa long, with CeMV encoding the smallest P protein (506 aa). The P gene also encodes two other gene products: the non-structural proteins C and V. The C protein is 186-aa long and produced by leaky scanning from an alternative start codon downstream of the P protein ATG. By contrast, the non-structural V protein is produced by viral RdRp-mediated RNA editing. At a specific motif of three G residues within the P gene sequence a G residue is inserted during transcription, resulting in a subset of V viral mRNAs that share the same N-terminal sequence as P but have a frame-shifted C termini. The C protein has been shown to regulate transcription and replication of genomic RNA in RPV and inhibit type I interferon (IFN) responses in PPRV. Although the C protein is broadly thought to act as an IFN antagonist, Messling *et al.* suggested that this protein is mostly involved in viral spread within the host, while the V protein is the main functional effector of IFN inhibition. In fact, the V proteins of RPV, PPRV and CDV have all been shown to inhibit type I and type II IFN responses. Mechanistically, V protein blocks type II IFN responses through interactions with STAT1, preventing translocation to the nucleus and activation of interferon-stimulated genes (ISGs).

The M protein plays a pivotal role in viral structure and assembly of progeny virions, and is the minimum requirement for the release of virus-like particles (VLPs). During morbillivirus infection, the M protein associates with the RNP *via* interactions with the N-terminus of the N protein, not only mediating viral RNA encapsidation into progeny virions, but also particle synthesis. In addition, the matrix protein is involved in recruiting the glycoproteins to the inner surface of the cell membrane to enable virus budding. The majority of the M protein is found within the cytoplasm; however, it can also be found in the nucleus of infected cells, where for MeV it was reported to inhibit host transcriptional machinery.

The fusion protein F is one of two viral glycoproteins present on the surface of the virion, orchestrating virus-cell and cell-cell fusion during morbillivirus infection. F is produced as an inactive precursor F_0 , which is proteolytically cleaved and folded into the metastable prefusion state consisting of F_1 and F_2 subunits, linked by a disulfide bond. The F_1 subunit is known to contain an N-terminal hydrophobic fusion peptide, two heptad regions (HR1 at the N-terminus and HR2 at the C-terminus), a transmembrane domain and a C-terminal cytoplasmic tail. A third heptad repeat termed HR3 is located in the F_2 subunit. Together, these heptad regions mediate membrane fusion upon F triggering *via* conformational changes that enable the fusion peptide (located in HR1) to dock with the target membrane. For the functionally trimeric F protein to be activated and promote virus-cell membrane

fusion, the viral particle first needs to bind its cellular receptor, an interaction mediated by the viral attachment protein H. The tetrameric H protein interacts with the F trimer primarily through an extended stalk domain. Each H monomer is composed of an N-terminal cytoplasmic tail, a transmembrane domain, stalk and a β -propeller with six β -sheets arranged as a globular head at the C-terminus. Structural studies of the H protein bound to its cellular receptors have revealed that $\beta 4$, $\beta 5$ and $\beta 6$ constitute the receptor binding domain (RBD). The H protein is the main target of the host's humoral immune response and many of neutralizing antibodies (nAbs) raised upon morbillivirus infection recognize the RBD, presumably serving to block attachment. Since this protein binds to two distinct cellular receptors, the ability to alter of the RBD sequence to drift away from nAb recognition is limited, which may explain the mono-serotypic nature of morbillivirus.

Lastly, the largest protein encoded within the morbillivirus genome is the RdRp polymerase. This protein is responsible for transcription of genomic RNA into mRNA as well as replication of nascent full length RNA genomes *via* a full length positive-sense intermediate, the antigenome. Since the entire morbillivirus replication cycle occurs in the cytoplasm of the cell, the L protein also carries out capping, methylation and polyadenylation of nascent viral mRNAs to allow efficient translation by the host cell machinery.

Life Cycle

Morbillivirus infection begins with the attachment of virions to the surface of the host cell *via* a proteinaceous cellular receptor. CD46 was the first receptor identified, however it was later shown that this is only recognized by laboratory-adapted MeV strains and not by wild type (WT) strains. Later, the signalling lymphocyte activation molecule (SLAMF1) was identified as the immune cell receptor for both WT and vaccine strains. This receptor is mainly expressed on activated B and T cells, macrophages and dendritic cells. In addition to being lymphotropic, morbilliviruses are also epitheliotropic; with the cellular adhesion molecule Nectin-4 being the receptor on epithelial cells. Experiments for many of the established morbilliviruses indicate that SLAMF1 and Nectin-4 are universal morbillivirus receptors. Genome entry occurs following F protein-mediated fusion of viral and cellular membranes; however, recently it was shown that SLAMF1 can also mediate endocytosis of the virions in a macropinocytosis-like manner.

The H and F proteins exist as a hetero-oligomeric complex on the surface of the virion. Binding of H to its cellular receptor brings the viral and cellular membranes into close proximity to enable fusion. To date, five models of morbillivirus fusion triggering have been described. The generally accepted model, summarised in Fig. 3, is that morbillivirus fusion is triggered following receptor attachment by H, which results in conformational changes in the head domain of the H protein and partial unfolding of the central stalk region. Unfolding of this region, which is known to interact with F, causes dissociation of the highly energetic and unstable prefusion form of this trimer. Receptor binding by H lowers the energy barrier required for F triggering and this, coupled with the release of the prefusion form, enables the protein to undergo a series of irreversible conformational changes until it stabilizes in its post-fusion state. This includes structural rearrangement of the heptad regions, resulting in docking of the fusion peptide in the target membrane and 'pulling' of the viral and cellular membranes into close proximity. Although fusion with the cellular membrane typically occurs at neutral pH, this is a highly energetic process, thus more than one F trimer might be triggered to overcome this energy barrier, as seen with other paramyxovirus. Finally, an F-mediated membrane pore is formed between the viral and cellular membranes, which expands in an actin cytoskeleton-dependent manner to enable the release of the viral RNP into the cytoplasm of the host cell.

Following F and H-mediated entry, the RNP is released into the cytoplasm of the host cell. Here, the incoming RdRp initiates the process of transcription by first binding to the GP sequence at the 3' end of the genomic RNA. For morbilliviruses, transcription continues *via* a 'stop-start' mechanism, with the RdRp only moving to the next gene once the mRNA from the previous gene has been synthesized and released. Importantly, during this gene-transition process at the IG the RdRp may fall from the genomic

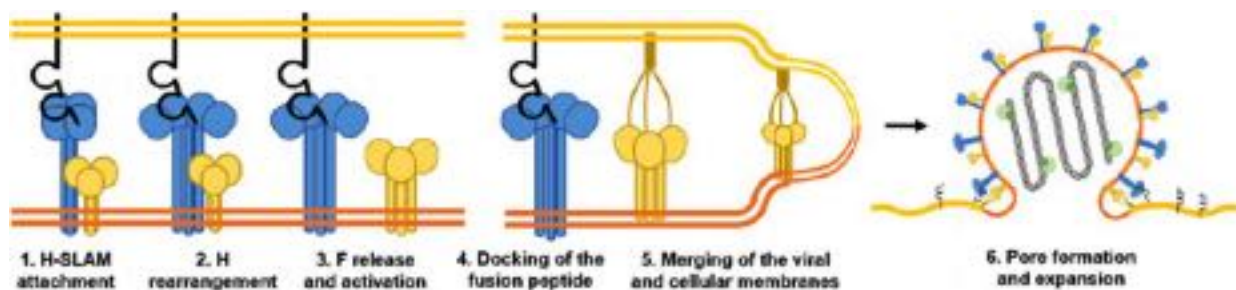


Fig. 3 Step-wise model for membrane fusion in the morbilliviruses. The H protein binds to the SLAM receptor (1), resulting in conformational changes in H (2). These changes promote detachment of the H and F proteins and activation of the latter, triggering structural changes in F (3). The heptad repeats of F form a three-helix bundle wherein the fusion peptide docks with the plasma membrane (4), pulling the two membranes into close proximity and promoting merging of the viral and cellular membranes (5). Lastly, F in its stable post-fusion state enables pore formation and expansion with consequent release of the RNP into the cytoplasm (6).

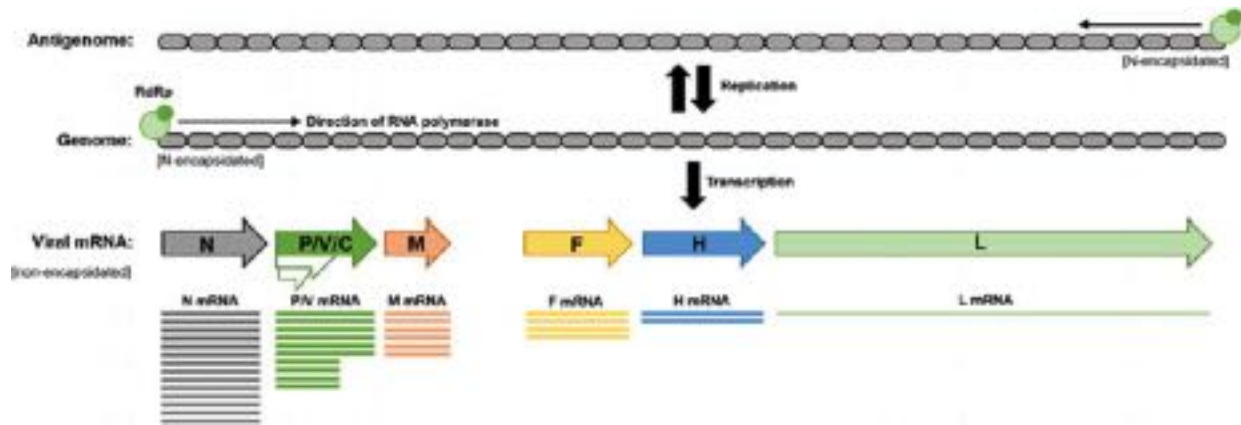


Fig. 4 Transcription and replication of the morbillivirus RNA genome. The RNA genome is transcribed by the viral RdRp, which can only commence transcriptional activity at the 3' end of this molecule. Following transcription of each gene, beginning with *N*, the RdRp either falls off the template at the intergenic region or continues transcription of the next gene, *e.g.*, *P*. This results in a transcriptional gradient whereby the genes closest to the GP are transcribed to higher levels than the promoter distal genes, *i.e.*, *L* is transcribed with the least frequency. The RdRp also carries out genome replication from the antigenome promoter (AGP at the 5' end of the genome), a process which occurs following the production of an encapsidated, full-length intermediate called the antigenome. This serves as the template for production of full-length, N-encapsidated RNA genomes which can be incorporated into progeny virions.

template. Since re-attachment of the polymerase and further transcription can only commence from the 3' GP, a transcriptional gradient results, whereby the genes closest to the GP are transcribed at higher levels than promoter distal genes (Fig. 4). This gene expression strategy reflects the relative requirement for individual proteins in viral replication, with the order of the genes being conserved in all morbilliviruses. Of note, the F and H glycoproteins are synthesized in the ribosomes of the rough endoplasmic reticulum and trafficked through the Golgi complex where post-translational modifications occur, including proteolytic cleavage of F₀ and glycosylation of both proteins.

The RdRp also carries out replication of nascent full-length RNA genomes, initially through production of a positive-sense intermediate; the antigenome, which is also N-encapsidated (Fig. 4). Once this positive-sense intermediate is produced, the RdRp binds to the AGP at the 3' end of the antigenome and synthesizes nascent, full-length, negative-sense RNA genomes, also N-encapsidated. The switch from transcription to replication is poorly understood; however, the most widely accepted model is that increasing levels of soluble N protein mediate the interchange from transcription to replication. Anderson *et al.* showed that deletion of the proximal region of the 3' UTR between the M and F ORFs of CDV influenced replication initiation but the exact mechanism for this remains poorly characterised. It has also been suggested that the presence of two variable-state RdRps, specific for replication or transcription, mediate this switch.

Finally, the synthesized viral components assemble at the plasma membrane to be incorporated into progeny virions that are released from the host cell. The exact mechanism by which morbilliviruses assemble and bud remains largely uncharacterized, therefore we will couple studies on other paramyxoviruses to address this step of the life cycle. It is widely accepted that the M protein plays a major role in assembly of the viral proteins and RNPs as well as budding of nascent virions. Budding of morbilliviruses from polarised epithelial cells occurs at the apical surface, requiring directional movement of the viral components to this surface. Apical budding is largely dependent on specific sorting of the glycoproteins, as otherwise these proteins accumulate at the basolateral membrane. This sorting is accomplished *via* interactions between the M protein and the cytoplasmic tails of the glycoproteins, as demonstrated for MeV, Sendai virus and para-influenza virus 5. Furthermore, budding at the apical surface also requires direct trafficking of the RNP complex to this site. Nakatsu *et al.* recently showed that apical budding of MeV in polarised epithelial cells is dependent on trafficking of the RNP complex to the cell membrane *via* the recycling endosome pathway (REP), employing Rab11A-positive vesicles. In contrast, budding from nonpolarized cells occurs in a REP-independent manner. Another cellular component required for efficient release of morbilliviruses is the actin cytoskeleton, as its specific disruption with pharmacological inhibitors severely impairs the trafficking of the M protein and RNPs to the surface of the cell. It has been suggested that once the viral proteins reach the cell surface they accumulate at lipid raft domains; CDV virions were reported to contain cholesterol while MeV RNPs and M proteins were shown to localize to lipid rafts. In contrast to viral assembly, particle budding requires actin cytoskeleton destabilization, allowing the mature bud to form. Bringolf *et al.* used a structure-guided mutagenesis approach to show that M protein oligomerization plays a key role in promoting virus budding, identifying that accumulation of M dimers at the cell surface induces membrane curvature and formation of the viral bud leading to virus egress, which Salditt *et al.* showed to be ESCRT-independent. Once the viral bud is formed, nascent virions 'pinch off' from the surface of the cell *via* a mechanism that remains largely uncharacterised. Lastly, it has been shown that morbilliviruses downregulate their cellular receptors on the surface of infected cells, a process which could facilitate release of cell-free virions. The life cycle of morbilliviruses is summarised in Fig. 5.

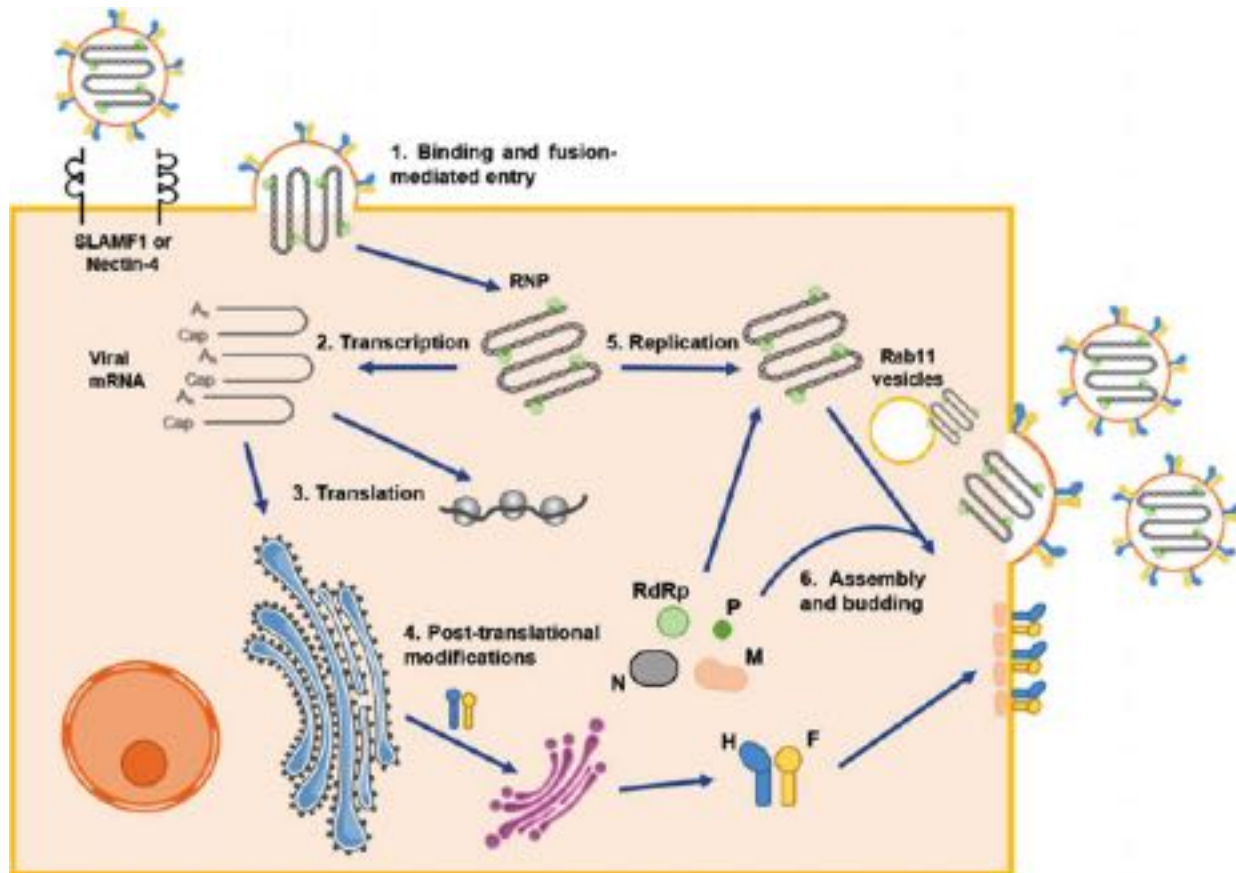


Fig. 5 Morbillivirus life cycle in the cell. A schematic overview of the morbillivirus replication cycle with viral proteins coloured, as in Fig. 2. (1) The virion attaches to the cell receptor through its H protein, which activates the F protein and triggers fusion of the viral and cellular membranes, with release of the RNP into the cytoplasm of the host cell; (2) the viral genome is transcribed, capped and polyadenylated by the RdRp to be translated by the host cell ribosomes (3); (4) Unlike the other viral proteins, the F and H proteins are translated by ribosomes in the rough endoplasmic reticulum and further processed in the Golgi complex; (5) At a later stage, the RdRp switches from genome transcription to genome replication to produce nascent viral RNA that will be incorporated into progeny virions; (6) The RNP as well as the M, F and H proteins are sorted to the surface of the cell membrane where nascent particles are generated by budding.

Epidemiology and Host Range

RPV

Historically recognized as significant diseases of humans and animals, RPV and MeV were properly described during the 16th and 17th centuries; however, the first evidence for rinderpest disease dates back to the 4th century AD. RPV is thought to have originated in Asia, followed by emergence in Europe in the 18th century, where the virus was described as the causative agent of cattle plague. By the end of the 19th century, RPV had reached the African continent, spreading to Australia and South America by the middle of the 20th century. Discovery of the virus, coupled with sequencing technologies, allowed RPV isolates to be classified into three lineages based on sequence variability (up to 11%) in the *F* gene. These lineages accurately represented the geographic distribution of RPV during the 20th century: lineage I included East, West and North African isolates, lineage II represented isolates from East and West African countries, while lineage III included Southern Asian isolates of RPV. The high prevalence of RPV across three continents and its enormous economic impact on agriculture, especially in twentieth century Africa, resulted in the implementation of a coordinated mass vaccination campaign to eradicate the disease. Following years of effort RPV was declared eradicated in 2011 (see below).

PPRV

In the middle of the 20th century, PPRV was identified as the causative agent of '*peste des petits ruminants*' (plague of small ruminants) in West Africa. Although initially thought to be a variant of RPV the origin of this genetically distinct virus remains uncertain. In fact PPRV is thought to have circulated in sheep and goats for a long time before being officially identified.

Nevertheless PPRV spread throughout Africa and reached Asia in the second half of the 20th century. Currently, PPRV circulates in over 70 African and Asian countries and poses a major threat to over 80% of the global sheep and goat populations, a distribution that has made it also a target for eradication. To date, four lineages of PPRV have been classified, based on sequencing of the C-terminal domain of the N gene. As with RPV, this reflects the geographical distribution of this virus across Africa, the Middle East and Asian countries. Lineages I, II and III are endemic in West and East African countries, respectively, while lineage IV is endemic in the Middle East and South Asian countries, although its distribution is increasing. Lineages II and III are thought to have originated independently and at the same time, while lineage IV is endemic in Asia. Worryingly, PPRV has shown evidence of a broadening host range, with lineage IV PPRV recently being detected in camels and saiga antelope, and lineage II in deer. Additionally, evidence of PPRV infection has also been detected in cattle and gazelles.

CDV

CDV is believed to have originated in the Americas before spreading into Europe, Africa and Asia following colonisation. It was initially thought that CDV was restricted to the *Canidae*; however, in the 1990s, CDV outbreaks caused fatalities amongst wildlife species including lions, tigers and leopards. A sudden CDV outbreak in the Serengeti National Park in Tanzania, suspected to have originated in African wild dogs, caused unexpectedly high morbidity and mortality in the African lion and spotted hyena populations. Indeed, CDV outbreaks pose a great threat to endangered species. For example a severe outbreak in the black-footed ferret population, with mortality rates of more than 70%, rendered this species extinct in the wild in the 1980s. More recently, CDV was reported in endangered Ethiopian wolves and it also now poses a major threat to Nicaragua's jaguar population, following transmission from CDV-infected hunting dogs. To date, CDV has been reported to naturally or experimentally infect almost all families of the order *Carnivora*. Significantly, CDV outbreaks have also been reported in monkeys in China and Japan, raising concerns that this virus can cross-over the species barrier to infect primates. Experimental adaptation *in vitro* has shown that CDV can rapidly adapt to use both epithelial and immune human cell receptors.

The worldwide prevalence of CDV has resulted in isolates being categorized into various phylogenetic lineages. CDV lineages are categorized based on H protein variability, with more than 5% amino acid diversity being the cut-off for delineation. Over the last 40 years, 10 lineages of CDV have been described based on this criterion, with the lineages also reflecting geographical distribution: two North American, three European, two South American, two Asian and one South African lineage(s). North America lineage 1 includes the vaccine strains of CDV while North America 2 viruses are known to be currently circulating in the USA; however, recently a third lineage has also been reported in domestic dogs. In South America, Europe lineage 1 strains were predominant until the identification of CDV strains in wildlife in Argentina, which led to the classification of South America lineage 2. The same authors suggested renaming Europe 1 to Europe/South America 1. Furthermore, Espinal *et al.* have also reported the presence of a third lineage in South America in Colombian domestic dogs. In Europe, CDV strains from Europe lineage 3 are currently circulating in Italy, affecting wolf populations in particular, and contributing to a spill-over event in badgers. The first CDV infection in anteaters has recently been reported in Brazil, with these strains clustering in the Europe/South America 1 lineage. Lastly, phylogenetic analysis of Serengeti isolates has revealed that these constitute a different lineage from the currently circulating South African lineage.

Aquatic Mammal Morbilliviruses

PDV was one of the first aquatic morbilliviruses to be identified and, to date, has caused two large outbreaks in European harbor seals. These outbreaks were originally thought to be CDV-related, as an epizootic event was taking place at the same time following contact of infected terrestrial mammals with seals; however, the harbor seals were later shown to be infected with a genetically and antigenically distinct virus that showed similar clinical manifestations. The second aquatic morbillivirus, CeMV was first isolated in 1991 with a host range largely restricted to dolphins, whales and porpoises. To date the exact origin of CeMV remains unknown, however, phylogenetic analysis suggests that CeMV shares a common ancestor with RPV and PPRV, suggesting that its origin might be a spill-over event. CeMV is presumed to be endemic in cetacean populations with an outbreak in pilot whales recently reported in the Canary Islands.

Emerging Morbilliviruses

Recently, a morbillivirus was detected in domestic house cats, firstly in China, with later reports from Japan, the USA and South Africa, suggesting the presence of a previously unidentified morbillivirus, largely restricted to these animals. Phylogenetic analysis based on the six structural proteins revealed that FmoPV clusters together with other morbilliviruses; however, FmoPV is somewhat distinct and less closely related to the other recognized members of the *Morbillivirus* genus. Interestingly, a recent study by Drexler *et al.* showed that bats and rodents may serve as a reservoir for a large number of previously uncharacterized morbilliviruses. The authors isolated multiple viral sequences which upon analysis were closely related to morbilliviruses, yet distinct enough to create their own phylogenetic clusters. This study also identified two bat isolates that were closely related to CDV and PDV, raising concerns of possible zoonotic transmission.

Clinical Features

The following is a summary of the classical clinical signs and symptoms seen in animal morbillivirus infections. A typical acute infection in animals can be classified into five distinct stages: Incubation, Prodromal, Mucosal, Diarrhoeic and Convalescent – a categorisation that corresponds well with the ‘in-tissue’ spread of infection described in the Pathogenesis section below. For RPV infections in cattle, which were often severe, the development of clinical disease was often summarised more succinctly as diarrhoea, dehydration and death. The initial incubation period is followed by a transient pyrexia. Around this period oculo-nasal discharges (associated with virus excretion and consistent with conjunctivitis and rhinitis) begin to develop which can become mucopurulent over time. Similarly, mucosal necrosis develops with erosions and ulcerations in the oral cavity, digestive and/or respiratory tracts. This worsens over time as shedding of necrotic material leads to severe erosions, bloody diarrhoea and a general worsening condition in the animal. Emaciation, dehydration and an associated drop in temperature to sub-normal levels can follow, all contributing to lethality. Other more non-specific signs of infection include restlessness, depression and anorexia as well as shallow and rapid breathing. In pregnant animals infection of the foetus and/or frequent abortion can occur; however, there is little evidence for teratogenic effects or the development of persistently infected ‘carrier states’ in new-borns. The mortality rate in morbillivirus infections can vary greatly, dependent on a number of factors. In immune-naïve herds mortality rates of 100% have been recorded, especially for the ruminant morbilliviruses. However, for other morbilliviruses a significant percentage of infections may be sub-clinical in nature, *e.g.*, CDV infections in domestic dogs. It is known that disease severity varies based on the strain of host as well as its immune status (vaccinated, immunosuppressed or waning maternally-derived immunity). Other complex factors, such as herd immunity levels and geographical endemicity may also contribute to severity in the wild. Variations within the viral sequence strain might also contribute to the development of disease, perhaps best exemplified by the low virulence RPV strains that were circulating in East Africa prior to eradication. In animals that survive an acute infection, one of the hallmarks of morbillivirus infection is a prolonged period of immunosuppression, frequently associated with secondary bacterial infections or indeed viral co-infections. Antigen and viral RNA can persist long after the resolution of acute infection; however, the significance of this to both viral epidemiology and immunosuppression is not well understood.

The progression of disease in morbillivirus-infected animals can also vary between hosts and there are a range of specific clinical signs that appear more specific to one virus than others; however, whether these are the result of virus- or host-specific differences is not well understood. Noticeable differences include a more frequent neural pathology in certain hosts. For CDV and CeMV encephalitis can be prolonged and chronic, with lesions forming in the CNS and brain. This is associated with abnormal behaviour, poor coordination, seizures and possible paralysis and can result in extensive demyelination. CDV-infected dogs can develop old dog encephalitis (ODE) which is reminiscent of the chronic MeV infection in humans, sub-sclerosing pan-encephalitis (SSPE) and many of these neurological symptoms develop long after the resolution of acute infection. Hyperkeratosis is most dominantly seen in CDV-infected domestic dogs; however, keratosis of the flippers, head, trunk, and tail have been reported for aquatic mammal morbilliviruses. In morbillivirus-infected ruminants, digestive pathology is more severe with the Peyer’s patches being particularly badly affected. In the large intestines of ruminants submucosal capillaries running along the crests of the mucosal folds become congested giving rise to zebra-stripes – a pathological hallmark of disease. There are, however, specific differences between the ruminant viruses as well, *e.g.*, PPRV is more likely to cause a primary viral pneumonia with frequent secondary microbial infections involving *Pasteurella* spp. and *Mycoplasma* spp. Significant respiratory infections are also evident in aquatic mammals, where bronchopneumonia is evident as well as necrosis of various tissues accompanied by atelectasis, congestion, edema and emphysema; the presence of parasitic and/or mycotic secondary infections in these animals has also been identified. FeMV infections perhaps represent the greatest outlier and currently there is no clear evidence that the virus utilises the biphasic life cycle described for other morbilliviruses. FeMV infections appear chronic and possibly sub-clinical in domestic cats with the kidneys and urine being associated with virus persistence and transmission, respectively.

Pathogenesis

As discussed previously, morbilliviruses are immuno and epithelio-tropic viruses; tropisms defined by their use of specific proteinaceous receptors. Accordingly, infection of cell-types expressing the two established receptors, SLAMF1 (aka CD150) and Nectin-4, determines the route and progression of disease within the mammalian host. SLAMF1, the immune cell receptor, is found on activated lymphocytes (B and T cells) as well as macrophages and dendritic cell subsets. The physiological role of SLAMF1 is predominately immune cell signalling; however, a role in the innate immune response to Gram-negative bacteria was also recently uncovered. Nectin-4, in contrast, is found on epithelial cells, mainly at mucosal surfaces in the respiratory and digestive tract. Physiologically, Nectin-4 is critical for cellular adhesion, forming an essential component of the adherens junction between polarised epithelial cells.

Morbilliviruses are highly infectious, and despite their enveloped particle, are thought to be capable of persisting in aerosolised droplets or on contaminated fomites or bedding for extended periods of time. For the majority of animal morbilliviruses the primary source of infection is thought to be inhalation of contaminated respiratory droplets. The prevalent model for virus dissemination through the host (summarised in [Fig. 6](#)) is built on various pathogenesis and receptor studies, including animal models of MeV infection. Initially, resident SLAMF1-expressing cells, such as dendritic cells or macrophages in the respiratory tract

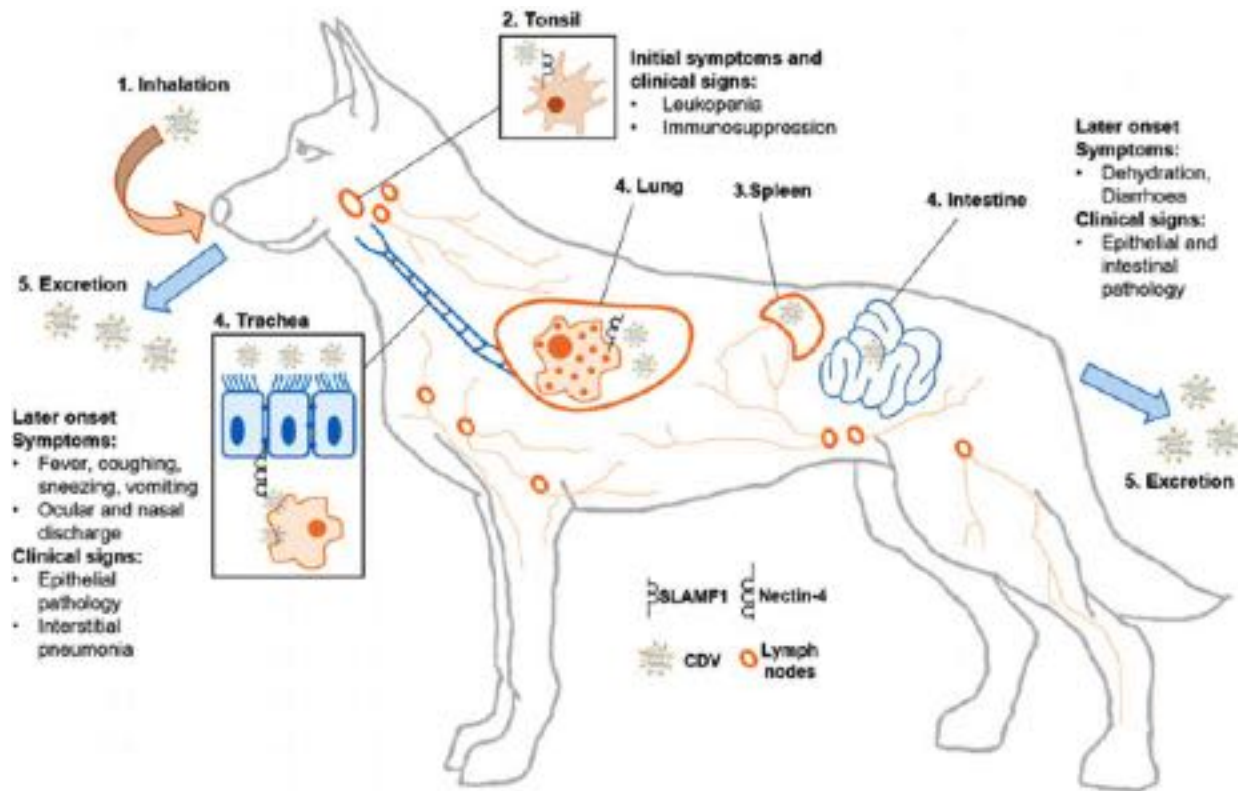


Fig. 6 Model of animal morbillivirus spread in an infected host. Virus particles enter *via* the respiratory route (1), with the virus first infecting SLAMF1-positive immune cells in tissues in the upper respiratory tract, perhaps the tonsil (2). Infected immune cells drain to local lymph nodes seeding subsequent rounds of immune cell infection and enabling the virus to reach more remote lymphatic tissues and immune-related organs such as the spleen (3). Immune cell-associated viraemia subsequently allows viral dissemination to other organs, such as the intestine, lungs and trachea (4). In these tissue infection spreads into the epithelial cells of these organs, *via* Nectin-4 mediated entry; a receptor found on the basolateral surface of polarised cells. This second stage of the infection plays a key role in the development of many clinical signs observed in animal morbillivirus infections. Lastly, nascent virions are released into the lumen of the digestive or respiratory tract, allowing spread to naïve animals (5). Of note, the neurological infections seen in some CDV-infected dogs are not illustrated here.

are infected. However, despite high levels of Nectin-4 expression in the oral cavity and upper respiratory tract this receptor is not immediately accessible because adherens junctions are buried on the basolateral surface of polarised cells. Indeed, infected immune cells first migrate to local lymph nodes, leading to a burst of viral replication in resident B and T cells. Subsequently, infected cells migrate from lymphatic tissues into the blood generating a cell-associated viraemia. It is only then, upon contact between morbillivirus-infected immune cells and the basolateral surface of mucosal epithelia that the virus spreads to Nectin-4-positive cells. Replication within epithelial tissues leads to preferential release of nascent virus into the lumen of the respiratory or digestive tract favouring transmission to new hosts. This biphasic life cycle has led to SLAMF1 and Nectin-4 being referred to as the entry and exit receptors, respectively.

This widely accepted model for morbillivirus infection is supported by pathological observations following autopsies of naturally or experimentally infected animals. Morbilliviruses can be detected in the bronchus-associated lymphoid tissue (BALT) and trachea-bronchial lymph nodes within a few days of infection. In addition, a significant incubation period is observed (without epithelial pathology), associated with an acute leucopenia – indicative of localised replication in lymphatic tissues. Marked necrosis and depletion of lymphocytes can be seen in a range of immune tissues such as the spleen and thymus as well as lymph nodes surrounding both the respiratory and digestive tract, where germinal centre B cells are frequently found to be infected. Subsequent virus spread to the epithelium is associated with immune cell infiltration into the lamina propria and degeneration of epithelial cells. Fused epithelial cell syncytia can often be observed as well as intracytoplasmic and intranuclear inclusion bodies.

Diagnosis and Prevention

The control and prevention of morbillivirus disease is of global relevance to human health, agriculture and wildlife. In fact, in many ways, morbillivirus diseases primed many of the modern approaches to veterinary vaccination and epidemic control (see Rinderpest case study below). The varied approaches are summarised in this section.

Certain animal morbilliviruses, such as RPV or PPRV, are notifiable diseases – enabling a co-ordinated response to infection, especially in non-endemic settings. For these viruses in particular there are a multitude of available tests and tools for diagnosing viral infection, as well as advisory information on which samples and tissues to harvest (managed by the World Organisation for Animal Health, OIE). This advice is likely applicable to all morbillivirus infections; however, the availability of validated diagnostic assays may vary. The recommended sites for sampling are swabs from the eye, nose and any obvious mucosal erosions or lesions, as well as uncoagulated blood. As with many infections the aim is to sample and isolate live virus, viral RNA, viral antigen and/or host-specific antibodies. Virus isolation can be made on a broad range of cell-lines, *e.g.*, Veros, where the efficiency of recovery can be boosted by over-expression of one of the cognate receptors, SLAMF1 or Nectin-4. There are various established RT-PCR or RT-qPCR tests to detect viral RNA, whilst viral antigen can be detected by agar gel immunodiffusion, immunocapture ELISA or penside test assays. The isolation of virus-specific antibodies is important for serological and epidemiological monitoring, as well as the assessment of vaccination responses. Of note, all morbilliviruses appear to be mono-serotypic in nature enabling sero-surveillance. Various established tests are available including virus neutralisation tests (VNTs), indirect ELISAs and competitive ELISAs, all of which target immune responses to the F, H and/or N protein. Sequencing of viral isolates, often targeting the *N*, *F* or *H* genes, has proved beneficial for phylogenetic characterisation and tracking of outbreaks.

Basic approaches to limit the spread of disease are effective at controlling morbillivirus outbreaks. In non-endemic settings the segregation of animals and culling of infected and contact individuals has proven effective, along with adequate disposal of infected carcasses. Often this is not practical in non-endemic or wildlife settings, where vaccination is the only realistic option. The best animal vaccines must have the following properties: They must be efficacious and safe in all animals and breeds, immunity must be easily testable *in vitro*, they must not be immunosuppressive, and finally they should provide lifelong or near-lifelong immunity. Fortunately, many of the licenced morbillivirus vaccines share these properties, the most widely used of which are live attenuated variants. The two commonly used CDV vaccine strains were derived from adaptation to ferrets, chicken embryos and later chicken cell lines (Onderstepoort) or canine kidney cell lines (Rockburn). Whilst these are widely, and successfully, used in dogs their efficacy in wildlife is less clear, with examples of vaccine-associated virulent disease. The difficulties associated with testing and using vaccines in wildlife also means that few vaccine trials have been performed in aquatic mammals, although there is evidence of CDV vaccines providing cross-protective immunity. The other widely used morbillivirus vaccines are the live attenuated PPRV vaccines, generated following serial passage of field strains in tissue culture (Veros). Recently, researchers have demonstrated that the two most widely used vaccines are efficacious against all lineages of circulating PPRV, supporting their use in future control and eradication campaigns. Despite the availability of efficacious vaccines there is a requirement for next generation tools. For example, a vaccine that would allow epidemiologists and veterinarians to distinguish infected from vaccinated animals (DIVA) would prove beneficial in a post-outbreak or eradication setting. Researchers have investigated this by modifying viral epitopes, using alternative viral vectors or recombinant protein platforms to express immunogens or indeed making recombinant morbilliviruses and there is great promise that one such vaccine will soon come to the market, likely for PPRV. Lastly, it is worth noting that recombinant morbilliviruses are also used as vectors themselves to deliver and express antigens from other important viruses, *e.g.*, a CDV expressing Rabies G protein.

Treatment

In general supportive care revolves around maintaining adequate hydration and the avoidance of secondary bacterial infections. Although there are no licenced antivirals for treating sick animals, there are a number of candidate antiviral compounds that have shown efficacy in the lab either by blocking fusion or inhibiting the polymerase.

Rinderpest Eradication: A Success Story

Although RPV was declared eradicated by the FAO and OIE in 2011 the history of control of this disease goes back hundreds of years. Historically, outbreaks of RPV in Europe and Asia are thought to have originated in the Caspian basin (presumably brought in by or with marauding armies). To an extent, the severity and frequency of these outbreaks is thought to have precluded the development of sustainable bovine agriculture. However, in 1711, following RPV disease in the papal herds, Pope Clement XI ordered a physician, Dr. Lancisi, to investigate the issue. He implemented some of the first detailed epidemic control strategies in animals including movement restriction, slaughter and burial of carcasses in lime. The success of this strategy was quickly adopted by other countries in Europe to control RPV infections; many of these approaches are still in use today. Training in the implementation of these strategies led to the development of the first veterinary school (Lyon, France 1761) as well as state veterinary services, with other countries in Europe soon following suit. Rinderpest was also instrumental in establishment of the OIE, while RPV was one of the first issues tackled by FAO in inaugural meetings. The impact of RPV on wildlife should also not be overlooked, as this virus caused devastating epidemics in wild ungulates, especially in Africa. Interestingly, wildlife were initially thought to be the source of infections; however, following the successful removal of RPV from cattle populations in the Serengeti ecosystem, the buffalo and wildebeest populations soared. The history of RPV vaccines tracks a shift from treatment to prevention and finally eradication. Initial vaccinations involved immunization with sera from convalescent animals. This was replaced by killed vaccines and then ultimately live attenuated viruses adapted to animals: goats, rabbits and eggs but later tissue culture

– primary calf kidney cultures. The history and politics of RPV eradication is a fascinating area summarised excellently elsewhere (see further reading); however, the key factors that led to these live attenuated vaccines being successfully used to eradicate RPV are as follows: There was a limited geographical distribution of the disease and no latency or persistence. In addition, infected animals had a short infectious period and transmission was reliant on direct or close indirect contact. This, combined with a safe universal vaccine, usable in all animals and providing lifelong immunity contributed to the success of a global approach led by international organisations with a clear mandate and unified strategy. Many of these criteria apply to PPRV and the FAO and OIE have recently launched a global strategy for PPRV eradication. Whilst eradication is also a meaningful target for the human morbillivirus MeV it seems unlikely that CDV or any of the aquatic mammal morbilliviruses are realistic targets, given their global distribution and more complicated host-range.

Further Reading

- Beineke, A., Baumgartner, W., Wohlsein, P., 2015. Cross-species transmission of canine distemper virus-an update. *One Health* 1, 49–59.
- Chang, A., Dutch, R.E., 2012. Paramyxovirus fusion and entry: Multiple paths to a common end. *Viruses* 4 (4), 613–636.
- de Vries, R.D., Ludlow, M., de Jong, A., *et al.*, 2017. Delineating morbillivirus entry, dissemination and airborne transmission by studying in vivo competition of multicolor canine distemper viruses in ferrets. *PLOS Pathogens* 13 (5), e1006371.
- Drexler, J.F., Corman, V.M., Muller, M.A., *et al.*, 2012. Bats host major mammalian paramyxoviruses. *Nature Communications* 3, 796.
- Duignan, P.J., Van Bresse, M.F., Baker, J.D., *et al.*, 2014. Phocine distemper virus: Current knowledge and future directions. *Viruses* 6 (12), 5093–5134.
- Nambulli, S., Sharp, C.R., Acciaro, A.S., Drexler, J.F., Duprex, W.P., 2016. Mapping the evolutionary trajectories of morbilliviruses: What, where and whither. *Current Opinion in Virology* 16, 95–105.
- Pastoret, P.-P., Yamanouchi, K., Mueller-Doblies, U., *et al.*, 2006. Chapter 5: Rinderpest – An old and worldwide story: History to c. 1902. In: Barrett, T., Pastoret, P.-P., Taylor, W.P. (Eds.), *Rinderpest and Peste des Petits Ruminants*. Oxford: Academic Press. (86–VI).
- Van Bresse, M.F., Duignan, P.J., Banyard, A., *et al.*, 2014. Cetacean morbillivirus: Current knowledge and future directions. *Viruses* 6 (12), 5145–5181.

Animal Papillomaviruses (*Papillomaviridae*)

John S Munday, Massey University, Palmerston North, New Zealand

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Cell mediated immunity The part of the immune response that is directly mediated by cells such as cytotoxic T-lymphocytes.

Humoral immunity The part of the immune response that is mediated by antibodies.

Immunohistochemistry Using antibodies to identify proteins within histological sections.

In situ hybridization Using antisense probes to identify DNA or RNA within histological sections.

Oncoproteins Proteins that promote cell growth and division.

Open reading frame A section of nucleotide sequences between a start and a stop codon.

Papilloma While it is acknowledged that papillomas can either be papillomavirus-induced hyperplastic lesions or non-papillomavirus-induced benign neoplasms of the epithelium, in the text the term 'papilloma' is used to indicate an area of focal epithelial thickening caused by papillomavirus infection. The term papilloma is therefore used to describe a 'wart' in human medicine.

Retinoblastoma protein A protein that controls cell division. Loss of retinoblastoma protein will favor cell division.

Viral capsid The protein shell of a virus.

Classification

The *Papillomaviridae* family includes two subfamilies, the *Firstpapillomavirinae* and the *Secondpapillomavirinae*. The *Firstpapillomavirinae* currently contains 51 genera while only one genus is currently classified within the *Secondpapillomavirinae* subfamily. Genera in the *Firstpapillomavirinae* are named using the Greek alphabet. As there are now more papillomavirus genera than letters in the Greek alphabet, the prefixes dyo- and treis- have been used. For example the deltapapillomavirus genus was the fourth genus established and this was prior to the dyodeltapapillomavirus genus which was established prior to the treisdeltapapillomavirus genus. Within the *Secondpapillomavirinae* subfamily, genera are named according to the Semitic abjads (an abjad is similar to an alphabet, except only consonants are represented). Alefpapillomavirus is only genus currently within the *Secondpapillomavirinae* subfamily.

Knowledge of the genera of papillomaviruses is useful as papillomaviruses within each genus often have closely related hosts. Additionally papillomaviruses within the same genera often have similar biological properties and therefore usually cause similar diseases. For example papillomaviruses in the Lambdapapillomavirus genus all infect carnivores and cause proliferative papillomas, most frequently in the oral cavity.

Within genera, the papillomaviruses are further subdivided into numerically-named species. Currently there are 133 papillomavirus species within the 51 papillomavirus genera. Papillomaviruses within the different species can be expected to have the same, or closely related, hosts and similar biological behaviors. For example, within the deltapapillomaviruses, the species 4 deltapapillomaviruses all cause fibropapillomas in domestic cattle while the species 1 lambdapapillomaviruses all cause oral papillomas in felids including domestic cats, snow leopards, bobcats, and lions.

Within the species, papillomaviruses are subclassified into individual papillomavirus types. Each papillomavirus type is named sequentially using the scientific name of the host species. For example, the sixth fully classified papillomavirus of domestic cattle is *Bos taurus* papillomavirus type 6, which is a species 1 xipapillomavirus within the xipapillomavirus genus. With the notable exception of the bovine species 4 deltapapillomaviruses, papillomaviruses are highly species-specific. Around 500 individual papillomavirus types have been fully classified, including around 175 from the non-human species. Papillomaviruses have been identified in virtually every species that has been intensively studied including cattle, horses, dogs, cats, sheep, deer, rabbits, bears, sea lions, dolphins, primates, birds, a variety of laboratory and free-ranging rodents, bats, snakes, and turtles. Most species are likely to be infected by multiple papillomavirus types from multiple papillomavirus genera and further research will undoubtedly identify many more papillomavirus types in animals.

Classification of papillomaviruses has traditionally been using the genetic sequence of the L1 open reading frame (ORF) (Fig. 1(a)). Viruses of different subfamilies have less than 45% similarity in their L1 ORF while papillomaviruses in different genera have less than 60% similarity. Papillomaviruses with less than 70% similarity are considered to be in different species and less than 90% similarity is present between papillomaviruses of different types. While this scheme is simple, many viruses will show overlap between multiple different genera and species, making classification difficult. To overcome these areas of overlap, papillomaviruses are additionally classified based on their host species and the type of lesions that they cause.

As papillomaviruses tend to infect either mucosal or cutaneous epithelium, human papillomaviruses are also classified into mucosal or cutaneous types. While the same is likely to be true in animals, this classification is not widely used in veterinary medicine. In humans, papillomaviruses are also classified based on the likelihood that they will cause cancer. Therefore the 'low-risk' alphapapillomaviruses cause warts that spontaneously resolve and rarely, if ever, progress to cancer. In contrast, the 'high-risk'

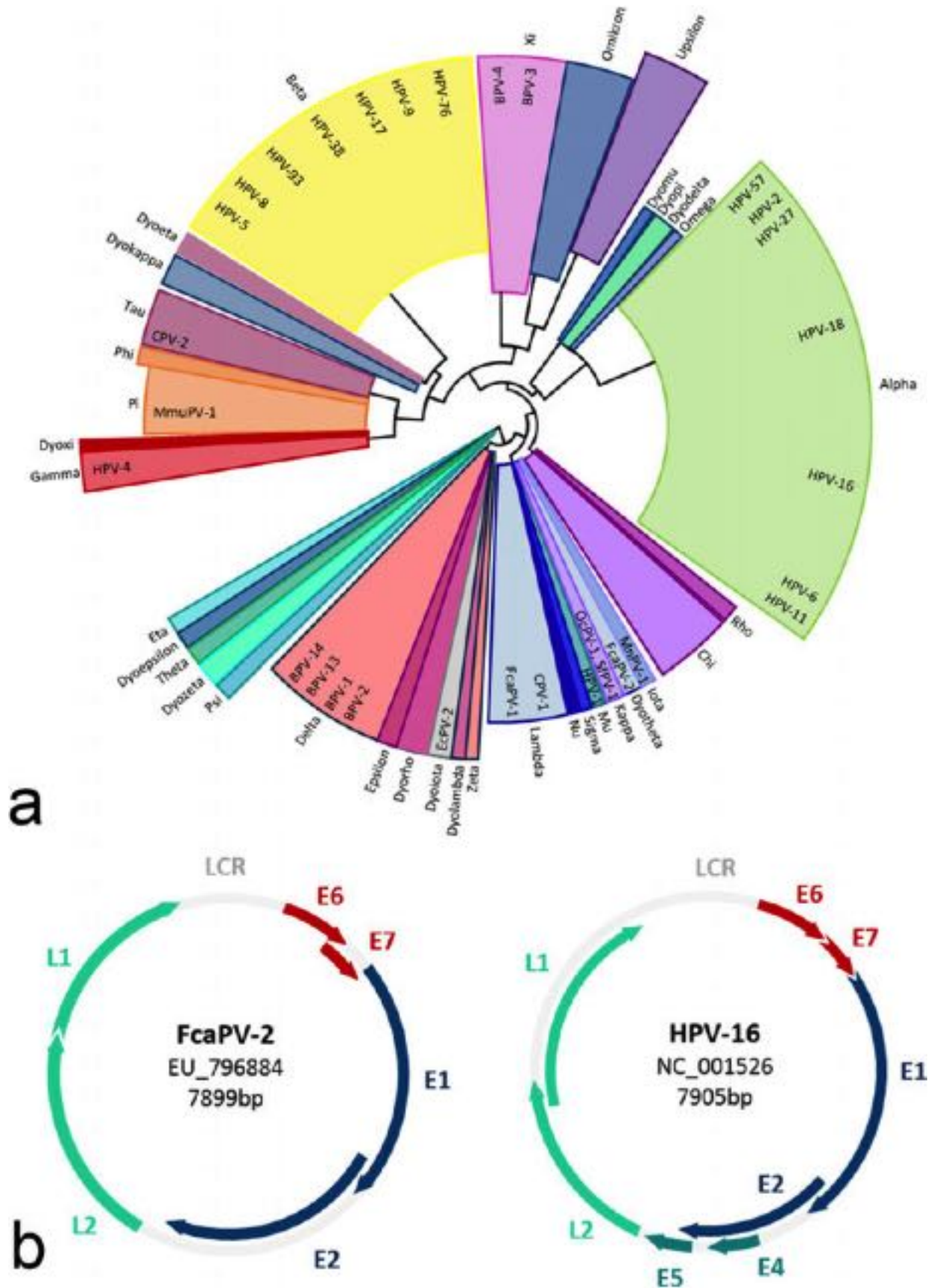


Fig. 1 (a). Schematic representation of select papillomavirus genera (courtesy Dr. Neroli Thomson, Massey University, New Zealand). (b). Genetic organization of *Felis catus* papillomavirus type 2 and human papillomavirus type 16 showing open reading frames within the viral genome and the long control region (LCR; courtesy of Dr. Neroli Thomson, Massey University, New Zealand).

Table 1 Summary of papillomaviruses that have been detected in five domestic animal species. It should be noted that not all papillomavirus types within a genus will cause all the associated lesions listed. For example, *Canis familiaris* papillomavirus (CPV)-17 is the only canine Taupapillomavirus type associated with oral SCCs and only CPV-2 and -7 have been associated with cutaneous papillomas. Additionally, while papillomaviruses are listed as being associated with lesions, for many the role of the papillomavirus in lesion development still remains unproven. SCC, squamous cell carcinoma; BISC, Bowenoid *in situ* carcinoma; BPV, *Bos taurus* papillomavirus; OaPV, *Ovis aries* papillomavirus; EcPV, *Equus caballus* papillomavirus; CPV, *Canis familiaris* papillomavirus; FcaPV, *Felis catus* papillomavirus

Species	Papillomavirus genus	Papillomavirus types	Predominant associated lesions
Cattle	<i>Delta</i>	BPV-1, -2, -13, -14	Cutaneous and esophageal fibropapillomas Bladder neoplasia
	<i>Xi</i>	BPV-3, -4, -6, -9, -10, -11, -12, -15, -17, -20, -23, -24	Cutaneous and upper alimentary papillomas Upper alimentary SCC
	<i>Epsilon</i>	BPV-5, -8	Cutaneous papilloma
	<i>Dyoxi</i>	BPV-7	Cutaneous papilloma
	<i>Dyokappa</i>	BPV-16, -18, -22	Cutaneous papilloma
	Unclassified	BPV-19, -21	Cutaneous papilloma
Sheep	<i>Delta</i>	OaPV-1, -2, -4	Cutaneous fibropapilloma
	<i>Dyolambda</i>	OaPV-3	Cutaneous SCC
Horses	<i>Zeta</i>	EcPV-1	Cutaneous papilloma
	<i>Dyoiota</i>	EcPV-2, -4, -5	Penile papilloma Penile SCC Aural plaque
	<i>Dyorho</i>	EcPV-3, 6, 7	Aural plaque
	Unclassified	EcPV-8	Cutaneous papilloma
	<i>Delta</i>	BPV-1, -2, 13	Equine sarcoid
Dogs	<i>Lambda</i>	CPV-1, -6	Oral papillomas Cutaneous papillomas
	<i>Tau</i>	CPV-2, -7, -13, -17, -19	Cutaneous papillomas Oral SCC
	<i>Chi</i>	CPV-3, -4, -5, -8, -9, -10, -11, -12, -14, -15, -16, -18, -20	Viral pigmented plaques Cutaneous SCC
Cats	<i>Lambda</i>	FcaPV-1	Oral papillomas
	<i>Dyotheta</i>	FcaPV-2	Viral plaques/BISC Cutaneous SCC
	Unclassified	FcaPV-3, -4, -5	Viral plaques/BISC
	<i>Delta</i>	BPV-14	Feline sarcoid

alphapapillomaviruses cause around 5% of all human cancers. While papillomaviruses are not commonly subdivided into 'low' and 'high' risk categories in veterinary medicine, there is increasing evidence that a subset of papillomavirus types cause cancer in animals.

The papillomavirus types that are currently recognized from the major domestic species are contained in [Table 1](#).

Virion Structure

Papillomavirus virions are non-enveloped and around 55 nm in diameter. The viral capsid consists of 360 copies of the L1 protein which are arranged in 72 pentamers (five-sided arrays) and around 12 copies of the L2 protein. Each capsid contains one copy of the papillomavirus DNA although both "empty" and "full" virus particles can be seen by electron microscopy. Papillomaviruses are resistant to diverse environmental insults and infectivity survives lipid solvents and detergents, low pH, and high temperatures.

Genome

The genome consists of a single molecule of circular double-stranded DNA that is typically around 7500 bp in length although the genome of individual papillomavirus types ranges from 5748 bp to 8607 bp. Several different translational strategies are utilized to enhance the limited coding capacity of the papillomavirus genome.

The genome encodes 6–9 proteins depending on the individual papillomavirus type. These proteins are subdivided into the early (E) and late (L) proteins ([Fig. 1\(b\)](#)). The E proteins are non-structural proteins and have important regulatory and replicative

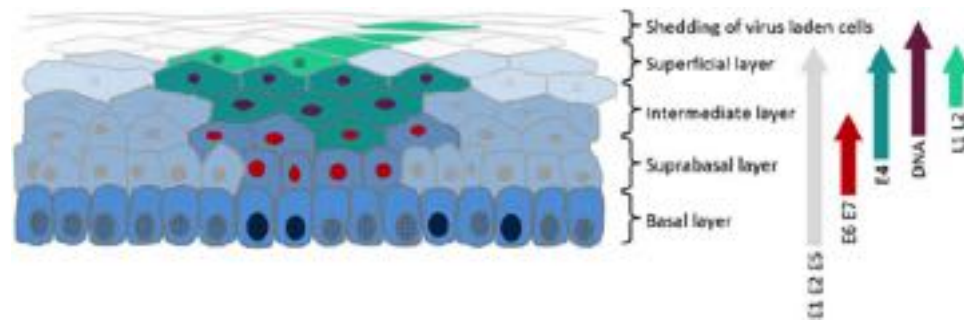


Fig. 2 Schematic representation of papillomaviral replication with stratified epithelium. Arrows indicate where the predominant layers of the epithelium at which each papillomaviral protein is produced and where papillomaviral DNA replication occurs (courtesy Dr. Neroli Thomson, Massey University, New Zealand).

functions. The E1 protein is a viral DNA helicase and is responsible for unwinding the double-stranded DNA. The main function of the E2 protein is to control the papillomavirus replication cycle. As will be discussed, replication of papillomaviruses is closely coordinated with epithelial cell replication and differentiation and the E5, E6 and E7 proteins alter the normal regulation of epithelial cells to maximize papillomavirus replication. As these proteins can promote cell division and therefore potentially cause neoplastic transformation of a cell, they are often referred to as 'viral oncoproteins'. In the majority of papillomavirus types, E7 is considered to be the protein that most significantly influences epithelial cell regulation. The E7 protein influences cell division by binding, and degrading, the retinoblastoma protein, an important cell cycle checkpoint. The E7 of most papillomavirus types has a retinoblastoma binding site (LXCXE) although binding to retinoblastoma protein can occur even without this site, possibly due to poorly-defined alternative pRb binding sites within the E7 carboxyl-terminal domain. In contrast, promotion of cell regulation by bovine deltapapillomaviruses appears to be primarily through interactions between the E5 protein and the platelet-derived growth factor- β receptor. The two L proteins form the viral capsid and these are only expressed close to the completion of the viral replication cycle.

Life Cycle

Infection by most papillomavirus types is currently believed to be restricted to stratified squamous epithelium. This is because papillomavirus replication is dependent on, and intimately linked to, the growth and differentiation of these epithelial cells (Fig. 2). Infection by a papillomavirus occurs when micro-trauma allows the papillomavirus to interact with the basement membrane, which then facilitates viral entry into a basal cell. The papillomavirus genome and some viral E proteins then enter the nucleus resulting in the production of 10–200 episomal copies of viral genomic DNA. As basal cells replicate, the episomal DNA spreads through the basal cell population and infection is maintained indefinitely within the infected epithelium. During this phase of replication, the papillomavirus does not significantly alter normal epithelial cell regulation and such infections are asymptomatic. While the infection is maintained within the host, viral replication does not occur and the infection at this stage is not infective.

For a papillomavirus to complete its life cycle, it has to be present within a basal cell that undergoes terminal differentiation. However, once a basal cell has terminally differentiated, it loses the ability to divide, resulting in degradation of the nucleus. As replication of the papillomavirus DNA is dependent on the host cell nuclear machinery, a key feature of papillomaviruses is their ability to prevent cells from leaving the S-phase of the mitotic cycle. By ensuring the infected keratinocytes keep dividing, not only does the papillomavirus retain the nuclear machinery within the cell, but also amplifies the infection by ensuring that each dividing keratinocyte contains replicating papillomavirus. As the infected keratinocytes approach the surface of the epithelium viral capsids are produced and virions are assembled. An infected cell may produce 10,000–100,000 new virus particles. Papillomaviruses do not cause cell lysis and virions are only released after the epithelial cell has been sloughed from the epithelial surface and degraded.

Replication of the papillomaviral DNA begins when an initiation complex binds to a single unique origin of replication resulting in unwinding of the DNA. A single DNA strand is synthesized continuously in the direction of unwinding with an additional DNA strand synthesized discontinuously in the opposite direction. As replication proceeds, the torsional strain created by the unwinding of the parental strands of DNA is released by the E1 protein. Bidirectional replication proceeds around the full genomic DNA circle resulting in the formation of two new strands of circular DNA.

Infection by papillomaviruses can result in a variety of clinical manifestations that include asymptomatic infection, papillomas (warts), viral plaques, and neoplasia. The manifestation of infection is largely dependent on whether or not the papillomavirus is replicating and the ability of the viral oncoproteins to promote epithelial cell division. Most papillomavirus infections are asymptomatic. Such asymptomatic infections may be due to a papillomavirus infection remaining latent and not replicating.

Alternatively, papillomavirus infections may remain asymptomatic when replication by the papillomavirus occurs only at a slow rate. Slow viral replication only mildly increases epithelial cell proliferation and visible changes in the epithelium will not develop. Papillomavirus infections that are characterized by slow viral replication generally stimulate no, or a weak, immune reaction. Therefore such infections may persist, resulting in the production of small numbers of infectious virions throughout the life of the host. In contrast, if the papillomavirus is able to stimulate marked epithelial proliferation, the thickened epithelium will become folded resulting in a visible papilloma. By stimulating such marked epithelial proliferation large numbers of infective virions are produced. However, the changes to the epithelium will be detected by the host resulting in an immune response, regression of the papilloma, and prevention of further viral replication. Viral plaques tend to be caused by papillomaviruses types that are normally asymptomatic, but an inability of the host to limit the rate of replication enables them to stimulate greater epithelial proliferation. Papillomaviruses can also predispose to neoplasia. The likelihood that a papillomavirus will cause neoplasia is not dependent on the degree of epithelial proliferation that is able to be induced by the papillomavirus. Indeed the human papillomaviruses that cause warts rarely cause cancer.

Epidemiology

There are currently few studies investigating the epidemiology of papillomavirus infections in animals. However, it appears likely that there are significant differences in the epidemiology of infection between papillomaviruses that are generally asymptomatic and papillomaviruses that cause papillomas.

The overwhelming majority of papillomaviruses are well adapted to their hosts and infections by these papillomavirus types typically remain asymptomatic. These papillomaviruses infect a high proportion of individuals with infection being from the dam, either during the birth process or due to the close contact between the mother and newborn shortly after birth. While direct contact is the most accepted mechanism of transfer, there is some evidence that transplacental transfer of papillomaviruses could be possible. Examples of papillomaviruses that are transmitted this way include the human beta-papillomaviruses that ubiquitously infect the skin of people and *Felis catus* papillomavirus type 2 which can be detected in almost all cats from very early in life.

Papillomaviruses that cause papillomas appear to have a different epidemiology as evidence suggests that papilloma development occurs when an animal is first infected by the papillomavirus. This infection could be the result of direct contact or indirect contact from a fomite. The initial infection results in a period of rapid virus replication that results in the development of a visible papilloma. Due to the large number of virions produced in a papilloma, an animal is likely to be highly contagious at this time. However, after a variable period of time, the body mounts a cell-mediated immune response against the papilloma, inhibiting viral replication and causing papilloma regression. Due to the ability of the infection to be maintained within the basal cells, it is likely that the infection is never completely resolved although it appears that few, if any, infectious virions are produced. This method of infection and disease development is illustrated by the 'outbreaks' of papillomas that have been rarely reported in animals.

Clinical Features

Infection by papillomaviruses can result in four distinct clinical manifestations that will be described separately. Due to differences between the species, these manifestations of papillomaviral disease will also be described separately for ruminants, horses, dogs, and cats. It should be noted, many other non-human species develop lesions due to papillomavirus infection and the list of species herein is not intended to be comprehensive.

Papillomas

Papillomas are the result of papillomavirus-induced thickening and folding of the epithelium. They mostly develop in young animals, presumably at the time of first infection by the causative papillomavirus type. Humoral antibodies are formed that protect the animal from additional infection by the papillomavirus type and, after a variable amount of time, the body mounts a cell-mediated immune response that causes spontaneous regression of the papilloma. The infection is probably not completely cleared and later immunosuppression may allow the papillomavirus to cause further papillomas. As papillomas typically regress and do not contain mutations within the cell DNA, they are not considered neoplasms. Papillomavirus-induced papillomas are commonly referred to as 'warts' in humans.

Ruminants

Papillomas in cattle are common with most animals developing these at some time in their lives. Papillomaviruses can be spread by direct contact or by fomites such as contaminated milking equipment, rubbing posts or wire fences. Sexual transmission of papillomavirus-induced genital papillomas is likely.

Papillomas in cattle can be subdivided into squamous papillomas that consist solely of proliferating epithelium and fibropapillomas in which there is proliferation of both the overlying epithelium and the underlying dermal fibroblasts. Cattle develop mucocutaneous papillomas and papillomas of the upper alimentary tract.

Mucocutaneous papillomas of cattle are most common around the head, but also commonly develop on the penis, vulva, and teats (**Fig. 3(a)**). As recent studies have revealed that most papillomas contain numerous BPV types, whether or not papillomas in particular locations are more likely to be caused by a specific BPV type is unknown. However, it appears likely that fibropapillomas are caused by infection with the deltapapillomavirus types, predominantly BPV-1 and -2. Rarely florid persistent mucocutaneous papillomas can be present and these can cause morbidity or mortality by interfering with vision or predisposing to secondary infection. It is currently unknown why some cattle appear unable to make an effective cell-mediated immune response against the papillomas. As these animals are invariably slaughtered, it is also unknown whether these warts would eventually have spontaneously resolved or would have progressed to neoplasia.

Upper alimentary papillomas develop from the caudal aspect of the oral cavity to the rumen. They are subdivided into squamous papillomas and fibropapillomas that are thought to be caused by BPV-4 and BPV-2 respectively. These papillomas typically remain small and self-resolve and clinical disease due to these papillomas is extremely rare. However, in animals that are exposed to the immunosuppressive and carcinogenic chemicals in bracken fern, upper alimentary squamous papillomas can become persistent and these papillomas can undergo neoplastic transformation.

Sheep develop cutaneous squamous papillomas and fibropapillomas most commonly around the muzzle, feet, and mammary gland. As with other species, the papillomas spontaneously resolve. Sheep are infected by three deltapapillomavirus types, although it is currently unknown which of these cause fibropapillomas. Fibropapillomas of the mammary glands have also been reported in goats.

Horses

Equine papillomas can be subdivided into those of the genitals and those that develop in other mucocutaneous areas. All are squamous papillomas and fibropapillomas are not recognized in horses.

Genital papillomas in horses are thought to be caused by *Equus caballus* papillomavirus type 2 (EcPV-2). Papillomas are more common in males and usually develop on the free part of the penis (**Fig. 3(b)**). Unlike papillomas in other species, equine genital papillomas most often develop in middle-aged or older animals. It is currently unknown how EcPV-2 is spread between horses,



Fig. 3 (a). Teat papillomas on an Angus cow. Multiple filiform squamous papillomas are visible. (b). Penile papillomas on a horse. Numerous papillomas are present and these have extended to involve the majority of the free part of the penis (courtesy Dr. Richard Malik, Centre for Veterinary Education, University of Sydney, Australia). (c). Oral and cutaneous papillomas on a dog. Despite their large size and number, such papillomas rarely cause discomfort to the dog. (courtesy Dr. Stephen White, University of California at Davis, California, USA). (d). Pigmented plaques on a dog. Multiple sessile darkly pigmented plaques are visible scattered over the legs of this dog (courtesy Dr. Mark Turnwald, Belmont Veterinary Clinic, North Shore City, New Zealand).

but as papillomas develop in castrated horses, sexual transmission is unlikely. The biological behavior of equine genital papillomas is currently unresolved with some spontaneously resolving, others becoming numerous and persistent, and a proportion progressing to penile squamous cell carcinoma (SCC).

Non-genital warts in horses are common and are thought to be predominantly caused by EcPV-1. They are most common in younger horses on the muzzle, lips, distal limbs, and eyelids. Papillomas are typically multiple and individual papillomas can coalesce into a single large mass. Non-genital papillomas in horses typically regress within 3 months and persistent infections or progression to neoplasia appears to be extremely uncommon.

Dogs

Papillomas are common in dogs and are subdivided into those of the oral cavity and those of the skin. The majority of oral papillomas are believed to be caused by *Canis familiaris* papillomavirus type 1 (CPV-1, formerly canine oral papillomavirus, COPV). Oral papillomas typically develop in younger dogs and the exophytic vegetative lesions can be multiple and extensive (Fig. 3(c)). However, despite the spectacular clinical appearance, it is rare for papillomas to interfere with eating or breathing by the animal. Most cases will spontaneously resolve within 3 months, although around 30% of papillomas will persist for longer and complete resolution can take up to a year in some cases. There are rare reports of persistent canine oral papillomas and these probably develop due to an ineffective papillomavirus-specific cell-mediated immune response. Such papillomas may be pre-disposed to progression to oral SCC, although this appears to be an extremely rare event.

Canine cutaneous papillomas have been associated with CPV-1, -2, -6, and -7. They are most common in young dogs and affected dogs often develop multiple papillomas. Papillomas develop most frequently in areas that are more likely to have skin abrasions such as the face, ears, and extremities. Canine cutaneous papillomas are further subdivided into exophytic papillomas in which the folded epidermis protrudes above the surface of the skin and inverted papillomas in which the folded epithelium is contained within a central depressed cup-shaped structure. Both exophytic and inverted papillomas are expected to spontaneously regress and it is currently uncertain whether these two lesions are caused by different papillomavirus types. Progression of a cutaneous papilloma to a SCC appears to be extremely rare in dogs.

Cats

Domestic cats are unusual because both cutaneous and oral papillomas are extremely rarely reported. Feline oral papillomas appear as clusters of small pale flattened plaques on the underside of the tongue. These papillomas are thought to be caused by FcPV-1. They are not associated with clinical signs of disease and probably spontaneously regress.

Viral Plaques

Viral plaques are much less common than papillomas. They probably develop when the host is unable to suppress infection by a normally asymptomatic papillomavirus type. The failure to inhibit papillomavirus replication results in moderate epithelial thickening and the development of a plaque. As the lesions do not necessarily develop when the animal is first infected, plaques tend to develop on older animals than papillomas which typically develop in younger animals. While spontaneous resolution of plaques can occur, they can persist over a long period of time and some may progress to SCCs.

Horses

Horses develop precancerous plaques of the penis and aural plaques within the concave surface of the ear pinna. Penile plaques are thought to be caused by EcPV-2 and probably represent an intermediate lesion between a papilloma and a SCC. Aural plaques have been associated with EcPV-3, -4, -5 and -6 with the papillomaviruses thought to be spread between horses by species of black flies (*Simulium* spp.). The prevalence of aural plaques is not known and it is possible many horses have these lesions without their owners seeking veterinary attention. Aural plaques can develop on horses of any age and consist of multiple discrete grey-white, roughened papules that can coalesce to form plaques up to several cm in diameter. Aural plaques may be unilateral or bilateral and are typically only a cosmetic concern. Aural plaques do not regress spontaneously, but have not been reported to progress to SCC.

Dogs

Canine pigmented plaques are thought to be caused by a number of closely-related Chipapillomavirus types. Some breeds of dogs, in particular Pugs and Vislas appear to be predisposed to pigmented plaques suggesting a breed-specific failure in normal suppression of replication by these papillomavirus types. Dogs receiving immunosuppressive therapy are also at increased risk for plaque development, although plaques often develop on dogs with no identifiable immunosuppression. Plaques are typically dark, multiple, and 1–10mm in diameter. They are most common on the ventrum and medial aspects of the limbs (Fig. 3(d)). Canine pigmented plaques are considered cosmetically undesirable, but usually do not cause significant morbidity. There are rare reports of progression to SCC although it is currently unknown whether or not plaques caused by specific papillomavirus types are more likely to progress to neoplasia.



Fig. 4 (a). Pigmented viral plaques/Bowenoid in situ carcinomas on the face of a domestic short haired cat. Two lesions are visible on the face of this cat (arrows; courtesy Dr. Richard Malik, Centre for Veterinary Education, University of Sydney, Australia). (b). Papillomas and squamous cell carcinomas of the penis of a horse. The penis of this horse is covered by papillomatous growths, some of which have progressed to a squamous cell carcinoma. (c). Viral plaques/Bowenoid in situ carcinomas and squamous cell carcinomas on a Devon Rex cat. Histology was consistent with the squamous carcinomas developing from pre-existing viral plaque/Bowenoid in situ carcinomas. Some lesions were covered by a thick layer of keratin (arrow; from Munday, J.S., Benfell, M.W., French, A., Orbell, G.M., Thomson, N., 2016. Bowenoid in situ carcinomas in two Devon Rex cats: Evidence of unusually aggressive neoplasm behavior in this breed and detection of papillomaviral gene expression in primary and metastatic lesions. *Veterinary Dermatology* 27, 215–e55, with permission). (d). Equine sarcoid. This sarcoid developed close to the eye of this 3-year-old horse (arrow; courtesy Dr. Luca Panizzi, Massey University, New Zealand).

Cats

Papillomavirus-induced cutaneous plaques in cats have historically been subdivided into viral plaques and Bowenoid *in situ* carcinomas (BISCs). However, as transitional lesions between the two lesions have been reported, they are best considered different severities of the same disease process. Feline viral plaques/BISCs are most often caused by FcPV-2. Most cats are infected with FcPV-2 shortly after birth and disease development appears to be primarily determined by, as yet poorly understood, host factors. Viral plaques/BISCs are uncommon in cats and usually develop in middle-aged or older animals. Lesions are most common on the head and neck of cats and can be single or multiple (**Fig. 4(a)**). Lesions are typically pigmented, mildly raised, and hairless. More advanced lesions can be ulcerated or covered by thick scaling. Lesions can spontaneously resolve, be persistent without progressing, or slowly increase in size and number. In addition, viral plaques/BISCs are pre-neoplastic and all viral plaques/BISCs should be carefully monitored for progression to a SCC. Viral plaques/BISCs tend to develop in Devon Rex and Sphinx cats at an earlier age and these lesions are predisposed to rapid progression to highly metastatic SCCs in these breeds.

Squamous Cell Carcinomas

In humans, papillomaviruses cause around 5% of all cancers including almost all SCCs of the cervix and 20%–30% of oral SCCs. The papillomaviruses that cause these cancers are a select group of sexually-transmitted ‘high-risk’ alpha papillomavirus types. A key feature of these papillomaviruses is the ability of their E6 proteins to prevent accurate replication of epithelial cell DNA and to inhibit cell apoptosis. This predisposes the proliferating epithelial cells to developing genetic mutations which can result in the development of neoplasia. While papillomaviruses in the domestic species have been associated with neoplasia development, the role of the papillomavirus in cancer development and the mechanisms by which the papillomaviruses cause neoplastic transformation are much less well understood.

Ruminants

Vulval papillomas have been reported to progress to SCCs in cattle that have been exposed to high levels of ultraviolet (UV) light. However, it is currently unknown whether or not the vulval papillomas are caused by papillomavirus infection and there is currently little evidence that papillomaviruses cause skin cancer in cattle.

Upper alimentary tract SCCs are extremely common in cattle that ingest bracken fern. As these SCCs can develop within a pre-existing papilloma, it is possible that BPV-4 could influence the development of these cancers. However, the precise role of the papillomavirus is currently uncertain and PV DNA is not detectable in the SCCs. Furthermore, bracken fern has directly carcinogenic properties suggesting that ingestion of bracken fern could be a more significant factor in the development of cancer than the presence of the papillomavirus. The critical role of bracken fern is supported by the rarity of upper alimentary tract SCCs in cattle that are not exposed to bracken fern.

Cutaneous squamous cell carcinomas in sheep were first proposed to be influenced by papillomavirus infection in 1982 when it was observed that some SCCs developed as a progression from a papilloma and that some SCCs contained PV particles. More recently, *Ovis aries* papillomavirus type 3 (OaPV-3) DNA and gene expression was found more frequently in a series of ovine cutaneous SCCs than in non-neoplastic samples of skin. Current evidence suggests that OaPV-3 could be a significant co-factor with UV light in the development of cutaneous SCCs in sheep. However, as SCCs rarely develop in sheep in the absence of high levels of solar radiation, it is difficult to determine the relative importance of papillomavirus infection and UV light in the development of these lesions. It is also impossible to exclude that higher rates of OaPV-3 are present within the SCCs simply because these cancers provide a more permissive environment for papillomavirus infection.

Horses

Penile SCCs are associated with EcPV-2 infection in horses. Currently the mechanisms by which EcPV-2 causes neoplasia are poorly understood. However, it is hypothesized that a high proportion of penile SCCs develop due to progression of a virally-induced papilloma to a precancerous plaque and subsequently to an invasive cancer (Fig. 4(b)). The degree to which EcPV-2 promotes the progression from papilloma to cancer is uncertain and it is possible that additional cofactors are required for neoplasm development. While both smegma and exposure to sunlight have been suggested as potential co-factors, there is currently little evidence supporting either as significant in the development of penile cancer in horses.

In addition to penile SCCs, there is some evidence that EcPV-2 may also cause a proportion of vulval and oropharyngeal cancers in horses, although currently too few cases have been evaluated to draw conclusions about the role of the papillomavirus in these neoplasms.

Dogs

Cutaneous viral plaques and papillomas have occasionally been reported to progress to SCCs in dogs, although the overwhelming majority of cutaneous SCCs in this species do not appear to be caused by papillomavirus infection.

There are just two reports of oral SCCs being influenced by papillomavirus infection in dogs. In one case, CPV-17 was found within multiple oral plaques and SCCs in a dog. These lesions contained histological evidence of viral replication suggesting that the papillomavirus was the cause of the lesions. Interestingly, CPV-17 has not subsequently been detected in any canine lesion. In the second report a dog had oral papillomas for over 18 months before one progressed to an invasive SCC. Overall, as with cutaneous SCCs, evidence currently suggests that papillomaviruses are not a significant cause of oral SCCs in dogs.

Cats

Papillomaviruses were first associated with feline cutaneous SCCs in 2008 when FcaPV-2 DNA was amplified significantly more frequently from SCCs than from non-neoplastic skin samples. Subsequent evidence has suggested that papillomaviruses could influence the development of 33%–45% of feline cutaneous SCCs. A papillomavirus etiology of a SCC may be most likely for a cancer that develops in skin that is protected from sunlight and it is possible that these SCCs could develop from a pre-existing viral plaque/BISC (Fig. 4(c)). The role of papillomaviruses in SCCs that develop in sun exposed skin is less clear. In people there is good evidence that papillomaviruses and sunlight act as co-factors in the development of SCCs and the same could be true in cats. However, as papillomaviruses commonly asymptotically infect skin, it is currently not possible to definitively exclude the papillomaviruses being present in the SCCs simply because they provide a more permissive environment.

While PV DNA is detectable in a small proportion of feline oral SCCs, there is no evidence that PVs are a significant cause of these neoplasms.

Rabbits

Infection of cottontail rabbits with *Sylvilagus floridanus* papillomavirus type 1 (formerly Shope's papillomavirus, cottontail rabbit papillomavirus, CRPV) causes multiple papillomas which progress to SCCs in around a quarter of rabbits. The consistent development of papillomas and cancer due to papillomavirus inoculation enabled rabbits to be used as an animal model during the development of HPV vaccines.

Sarcoids

Sarcoids are proliferations of fibroblasts that are induced by cross-species infection by bovine deltapapillomaviruses. Sarcoids do not metastasize, but tend to be invasive making complete excision difficult.

Horses

Sarcoids have been reported in horses, donkeys, mules and zebras and are the most common skin tumor of equids worldwide. Sarcoids can be single or multiple and may occur anywhere on horses but are more common on the head, neck, and limbs (**Fig. 4(d)**). They often develop at sites of previous surgery or trauma. Sarcoids are most common in young to middle-aged horses and breed predispositions in Quarter horses, Appaloosas and Arabians have been described.

Sarcoids are thought to be caused by infection with BPV-1, -2 or -13, although sarcoids cannot be induced experimentally by inoculation with these papillomaviruses. It is currently not clear whether BPVs are transmitted from infected cattle, from infected horses, or from both. Horses have historically been considered dead-end hosts of BPV infection, with early studies detecting only BPV DNA in dermal fibroblasts, but never infectious BPV virions. However, expression of the BPV L1 gene has been detected in sarcoids, suggesting the possibility of viral replication within the lesion. Additionally, full-length BPV genomes complexed with L1 capsid proteins, possibly representing virions, have been detected in two equine sarcoids.

Cats

Feline sarcoids appear to be caused by BPV-14. As cattle are the definitive host of the presumptive causative papillomavirus, it is unsurprisingly that feline sarcoids are restricted to cats that have contact with cattle and are only seen regularly in 'barn cats' in the Midwest of the United States of America. Feline sarcoids most frequently develop on the nasal philtrum, although they can develop anywhere on the body or in the mouth. The mechanism of spread is poorly understood, but the lesion distribution suggests that biting flies or cat-fight wounds may be important for BPV-14 infection. BPV-14 is a deltapapillomavirus that is closely related to the BPVs that cause equine sarcoids. However, it is interesting that BPV-14 has not been detected in an equine sarcoid and neither BPV-1 nor -2 have been detected in a feline sarcoid. This suggests that, although the deltapapillomaviruses can cause cross-species infection, the range of species able to be infected may be limited.

Other Neoplasms

Deltapapillomavirus DNA and viral proteins, including those of types -2, -13, and -14, have been found more frequently in bladder neoplasms of cattle and water buffalo than in non-neoplastic bladder samples. While this suggests a possible carcinogenic role of the BPVs, the development of bladder neoplasia in cattle is dependent on the ingestion of bracken fern. Therefore, it cannot be excluded that the high rates of BPV detection in the bladder tumors is due to the immunosuppressive action of bracken fern, rather than because the papillomaviruses directly contribute to cancer development.

It is interesting to note that deltapapillomavirus infections of the bladder of cattle are productive and this appears to be a rare example of papillomaviruses being able to replicate in a location other than squamous epithelium. How bladder infection occurs is currently unknown. However, as papillomavirus DNA is detectable in the blood, hematogenous infection of the bladder mucosa may be most likely. Interestingly a wide range of bladder tumors are associated with papillomavirus infection including urothelial and mesenchymal types.

Pathogenesis

Papillomaviruses cause disease by stimulating increased epithelial cell replication. If there is a marked increase in replication then the infection can result in a papilloma while a more moderate increase in replication could result in a viral plaque or an asymptomatic infection. Due to their ability to interfere with the normal regulation of cell division, some papillomaviruses may influence the development of neoplasia.

Because papillomaviruses do not cause cell lysis and only produce viral antigens in the superficial layers of the epidermis, infection may elicit no immune response. If an immune response is made, it consists of the production of antibodies that protect against subsequent infection by this papillomavirus type and a cell-mediated response against cells that are infected by the papillomavirus. The initiation of a cell-mediated response is responsible for the resolution of a developed papilloma.

Diagnosis

A papillomavirus etiology of a lesion can be diagnosed by clinical examination, histology of the lesion, immunohistochemistry, or molecular techniques. Not all methods of diagnosis are appropriate for each of the types of papillomavirus-induced lesion.

Histology of a papillomavirus-induced lesion generally reveals thickened epithelium (**Fig. 5(a)**). If the papillomavirus is replicating within the lesion, papillomavirus-induced cell changes may be visible. These changes include the presence of enlarged cells with increased quantities of blue to grey slightly granular cytoplasm, cells that have shrunken nuclei that are surrounded by a clear halo (koilocytes), cells with basophilic or eosinophilic cytoplasmic bodies, cells with enlarged vesicular nuclei, or cells with eosinophilic intranuclear viral inclusion bodies (**Fig. 5(b)**). Clumping of keratohyalin granules can also be visible within the superficial layers of the epidermis.

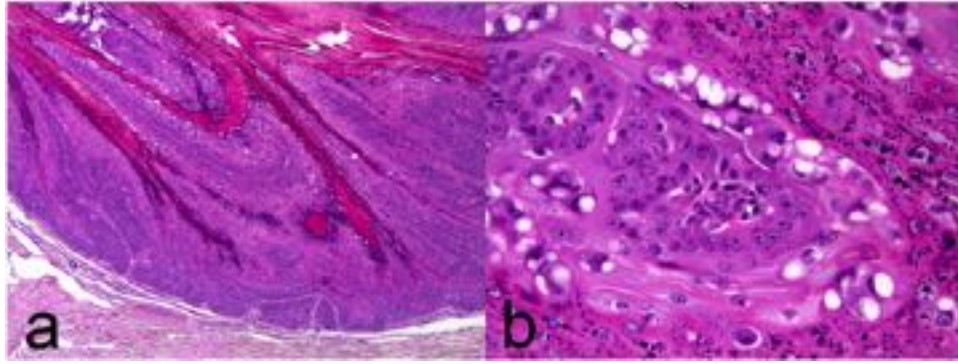


Fig. 5 (a). Canine papilloma. The papilloma consists of marked epithelial proliferation. The papilloma is clearly demarcated from underlying dermis and the thickened epidermis is covered by increased keratin. This papilloma was caused by *Canis familiaris* papillomavirus type 1. (b). Canine papilloma. Higher magnification of the papilloma reveals papillomavirus-induced cell changes. These consist of enlarged cells with clear cytoplasm, cells with vesicular nuclei, and prominent clumping of keratohyalin granules.

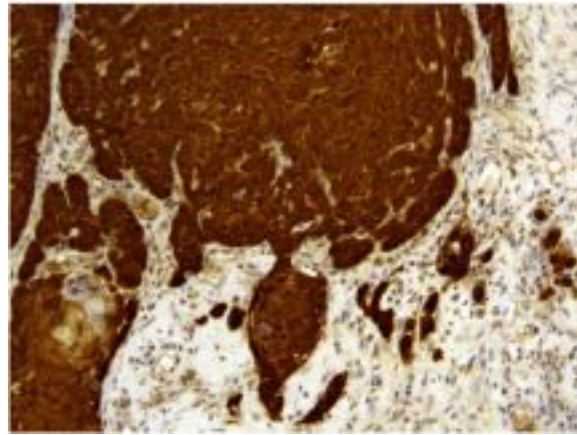


Fig. 6 Papillomavirus-associated cutaneous squamous cell carcinoma from a cat. Note the intense cytoplasmic and nuclear immunostaining using antibodies against p16^{CDKN2A} protein (p16). This immunostaining pattern suggests that the behavior of the neoplastic cells was altered by papillomavirus infection.

There are two strategies that can be used to detect a papillomavirus etiology of a lesion using immunohistochemistry. Firstly, anti-papillomaviral antibodies can be used to detect the papillomavirus L1 protein. A significant disadvantage of this strategy is that L1 protein (and therefore immunostaining) will only be present in lesions that contain active viral replication and it is rare for immunostaining to be present in a lesion that does not contain papillomavirus-induced cell changes. In addition, as there are no commercially-available antibodies to most papillomavirus types of domestic species, the cross-reactivity of the antibodies is often unknown. The second strategy involves using immunohistochemistry to detect p16^{CDKN2A} protein (p16). Papillomaviruses consistently alter cell regulation by mechanisms that increase cell p16. Therefore the detection of increased p16 in a cell is consistent with the behavior of that cell having been altered by papillomavirus infection (Fig. 6). An advantage of this strategy is that p16 immunostaining will be increased regardless of the replication status of the papillomavirus in the lesion. Additionally, a commercially available anti-p16 antibody (clone G175-405) has been shown to cross react with the canine and feline p16 protein, although this antibody does not appear to cross react with the equine p16 protein. As the bovine delta-papillomaviruses cause cell proliferation by alternative pathways, increased p16 is less likely to be present in lesions caused by these papillomaviruses.

Molecular techniques available include PCR and *in situ* hybridization. PCR has the advantage of being extremely sensitive and, unlike immunohistochemistry, able to determine the papillomavirus type present within the lesion. However, as papillomaviruses often incidentally infect the skin of animals, simply detecting papillomavirus DNA within a lesion is not conclusive evidence that the papillomavirus caused the lesion. As *in situ* hybridization is able to localize the DNA within a lesion, this technique provides greater evidence that the papillomavirus was the cause of the lesion. However, as with PCR, the possibility of incidental infection cannot be excluded. The sensitivity of *in situ* hybridization is lower than PCR.

Papillomas

Papillomas are usually diagnosed by clinical examination with the presence of multiple exophytic lesions in the mouth or over the skin of a younger animal consistent with this diagnosis. Histology is rarely needed but reveals thickened folded epithelium. As papillomas are formed due to florid viral replication, papillomas are expected to contain prominent papillomavirus-induced cell changes and immunostaining using anti-papillomaviral antibodies.

Viral Plaques

Viral plaques can be suspected on clinical examination due to their appearance and location. However, canine pigmented plaques and feline viral plaques/BISCs will often require histology for definitive diagnosis. Histology of a viral plaque reveals a well-demarcated focus of mild to moderate epidermal hyperplasia that often has a scalloped appearance. The hyperplasia tends to remain superficial with little involvement of adjacent hair follicles. Cell atypia and crowding can be present, especially in more advanced feline viral plaques/BISCs. Feline and canine plaques can contain increased melanin while hypomelanosis is described as a feature of equine aural plaques. Papillomavirus-induced cell changes and immunostaining using anti-papillomavirus antibodies are variably present. Feline viral plaques/BISCs contain intense immunostaining with anti-p16 antibodies. It is currently unknown whether similar immunohistochemical changes are visible in canine or equine viral plaques.

Squamous Cell Carcinomas and Other Neoplasms

Squamous cell carcinomas are characterized by the presence of invasion of dysplastic cells through the basement membrane of the epithelium. Significant viral replication is not typically present within either SCCs or bovine bladder neoplasms and papillomavirus-associated neoplasms are not expected to contain papillomavirus-induced cell changes or immunostaining using anti-papillomavirus antibodies. There is good correlation between the presence of papillomavirus DNA and p16 immunostaining in feline cutaneous squamous cell carcinomas (SCCs), suggesting that, like human oral SCCs, p16 immunostaining may be a good indicator of a papillomaviral etiology of a feline cutaneous SCC. Additionally, p16 immunostaining was observed in a canine oral SCC that was associated with papillomavirus infection. The papillomavirus-associated cancers are expected to contain viral nucleic acids and molecular tests can be used to localize these within the lesions.

Sarcoids

Histology reveals a proliferation of well-differentiated dermal fibroblast-like cells underlying a thickened epidermis that has characteristic broad interlacing rete pegs. As these lesions support little viral replication, neither papillomavirus-induced cell changes nor immunostaining using anti-papillomavirus antibodies are present. Molecular tests can detect papillomavirus DNA, mainly within proliferating fibroblasts. In cats, as BPV-14 does not appear to infect the skin of cats asymptotically, the detection of BPV-14 from a skin tumor provides strong evidence to support a diagnosis of feline sarcoid. While the detection of BPV DNA from an equine skin mass is supportive of a diagnosis of equine sarcoid, caution has to be used as BPV DNA has been detected in skin from horses without sarcoids.

Treatment

Papillomas

Papillomas in all species typically cause little or no morbidity and are expected to spontaneously regress after a variable length of time. For the overwhelming majority of cases, benign neglect is the treatment of choice. Surgical excision, laser treatment, or cryotherapy can be used for lesions that develop in locations that cause pain or interfere with function. Papillomas can recur after surgical excision, although such warts are expected to subsequently regress once the body mounts an appropriate cell-mediated immune response. Numerous medical therapies have been proposed to hasten the resolution of papillomas. However, none has sufficient evidence of efficacy from controlled studies to be confident of any benefit. Crushing of papillomas or injecting excised papillomas back into animals (autologous vaccination) have been historically used, but is not supported by controlled studies. Likewise the use of papillomavirus vaccines has not been shown to influence papilloma regression in either humans or animals.

Viral Plaques

Viral plaques appear to only rarely spontaneously regress. As plaques are mainly a cosmetic concern in dogs and horses, they can often be left without further treatment. In cats, viral plaques/BISCs should be carefully observed due to the potential of these lesions to progress to a SCC. Treatment options for viral plaques include surgical excision, laser therapy, or cryotherapy. Imiquimod cream has been used to treat plaques in cats and horses. Imiquimod is not specific for papillomaviruses, but works by stimulating a local immune response that can cause the resolution of epidermal lesions.

Squamous Cell Carcinomas and Sarcoids

There are numerous techniques for removing these neoplasms that are out of the scope of this discussion. In cats it is possible to subdivide cutaneous SCCs into those that are associated with papillomavirus infection and those that are not. Evidence suggests that papillomavirus-associated SCCs may have a more favorable prognosis than those not caused by papillomavirus infection.

Prevention

Vaccines that contain the L1 capsid protein prevent papillomavirus infection and subsequent papillomavirus-induced neoplasia in people. This suggests that vaccination could also be used to prevent papillomavirus-induced disease in animals. However, there are some key considerations regarding the use of papillomavirus vaccines in animals.

Firstly, there appears to be little cross-protection between papillomavirus types. Therefore, vaccines would have to be created that protect against each individual papillomavirus type.

Secondly, a vaccination has to be given prior to first infection. All the papillomavirus types that are included in the human vaccines are sexually transmitted. This is critical as it can be guaranteed that people will not be infected by these papillomaviruses prior to vaccination. In the domestic species, the papillomaviruses that cause papillomas appear to most frequently infect their hosts as young adults possibly providing an opportunity to prevent infection by vaccination. In contrast, FcaPV-2, EcPV-2, and the viruses that cause canine pigmented plaques probably infect their hosts soon after birth. If so, novel strategies to allow vaccination prior to first infection will have to be developed. Additional research is required to determine when horses are infected by the BPVs that cause sarcoids.

Lastly, vaccines are likely to be expensive and it seems unlikely that a vaccine against self-regressing papillomas will be a commercial success. If papillomaviruses are found to be a key factor in the development of esophageal and bladder cancer in cattle and water buffalo then vaccination against these economically-important diseases could be viable. However, if bracken fern is found to be the major factor causing these cancers, preventing papillomavirus infection may not result in significant protection against cancer.

Further Reading

- Akgul, B., Cooke, J.C., Storey, A., 2006. HPV-associated skin disease. *Journal of Pathology* 208, 165–175.
- Bergvall, M., Melendy, T., Archambault, J., 2013. The E1 proteins. *Virology* 445, 35–56.
- Borzacchiello, G., Roperto, F., 2008. Bovine papillomaviruses, papillomas and cancer in cattle. *Veterinary Research* 39, 45.
- Doorbar, J., Quint, W., Banks, L., *et al.*, 2012. The biology and life-cycle of human papillomaviruses. *Vaccine* 30 (S5), F55–F70.
- Egawa, N., Doorbar, J., 2017. The low-risk papillomaviruses. *Virus Research* 231, 119–127.
- Gil da Costa, R.M., Peleteiro, M.C., Pires, M.A., DiMaio, D., 2017. An update on canine, feline and bovine papillomaviruses. *Transboundary and Emerging Diseases* 64, 1371–1379.
- Lange, C.E., Favrot, C., 2011. Canine papillomaviruses. *The Veterinary Clinics of North America Small Animal Practice* 41, 1183–1195.
- Munday, J.S., 2014a. Bovine and human papillomaviruses: A comparative review. *Veterinary Pathology* 51, 1063–1075.
- Munday, J.S., 2014b. Papillomaviruses in felids. *Veterinary Journal* 199, 340–347.
- Munday, J.S., Kiupel, M., 2010. Papillomavirus-associated cutaneous neoplasia in mammals. *Veterinary Pathology* 47, 254–264.
- Munday, J.S., Thomson, N.A., Luff, J.A., 2017. Papillomaviruses in dogs and cats. *Veterinary Journal* 225, 23–31.
- Van Doorslaer, K., Chen, Z., Bernard, H.U., *et al.*, 2018. ICTV virus taxonomy profile: Papillomaviridae. *The Journal of General Virology* 99, 989–990.
- Vande Pol, S.B., Klingelutz, A.J., 2013. Papillomavirus E6 oncoproteins. *Virology* 445, 115–137.
- zur Hausen, H., 2009. Papillomaviruses in the causation of human cancers – A brief historical account. *Virology* 384, 260–265.

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsDNA-viruses/w/papillomaviridae
Papillomaviridae – Papillomaviridae – dsDNA Viruses – International.
- <https://pave.niaid.nih.gov/>
PaVE: Papilloma virus genome database.

Astroviruses (*Astroviridae*)

Virginia Hargest, St. Jude Children's Research Hospital, Memphis, TN, United States and University of Tennessee Health Science Center, Memphis, TN, United States

Amy Davis and Stacey Schultz-Cherry, St. Jude Children's Research Hospital, Memphis, TN, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Capsid The protein coat that surrounds the genetic material of a virus.

Fomite An object that can harbor pathogenic organisms and therefore may be involved in the route of transmission of the pathogen.

Subgenomic RNA (sgRNA) Viral RNA which can be directly translated into the desired viral proteins without needing to translate the entire genome. Usually translated in excess compared to genomic RNA.

T=3 Icosahedron In relation to the symmetry of virion structure, a virus with a triangulation number of 3 is composed of 12 pentameric and 20 hexameric capsid subunits for a total of 180 capsid proteins.

Transepithelial Electrical Resistance (TER) Method used to assess the barrier function of epithelial cells by measuring current passing through the cells on both transcellular and paracellular paths.

Uncoating A condition when the protein capsid of the virus is dismantled revealing the viral genome inside the host organism.

Classification

Astrovirus, a non-enveloped positive-sense single-stranded RNA virus, was first discovered in 1975 in the stool of infants suffering from gastroenteritis. The virus obtained its name after Madeley and Cosgrove noted a star-like morphology of a portion of the virions under electron microscopy. In 1993, complete sequencing of the viral genome prompted its classification to a new family to be called *Astroviridae*. The *Astroviridae* family has since been divided into two genera: *Avastrovirus*, those viruses that infect avian species, and *Mamastrovirus*, those viruses that infect mammalian species (Fig. 1). *Avastrovirus* and *Mamastrovirus* are also split into two genogroups based on of their genetic relatedness within the hypervariable capsid protein.

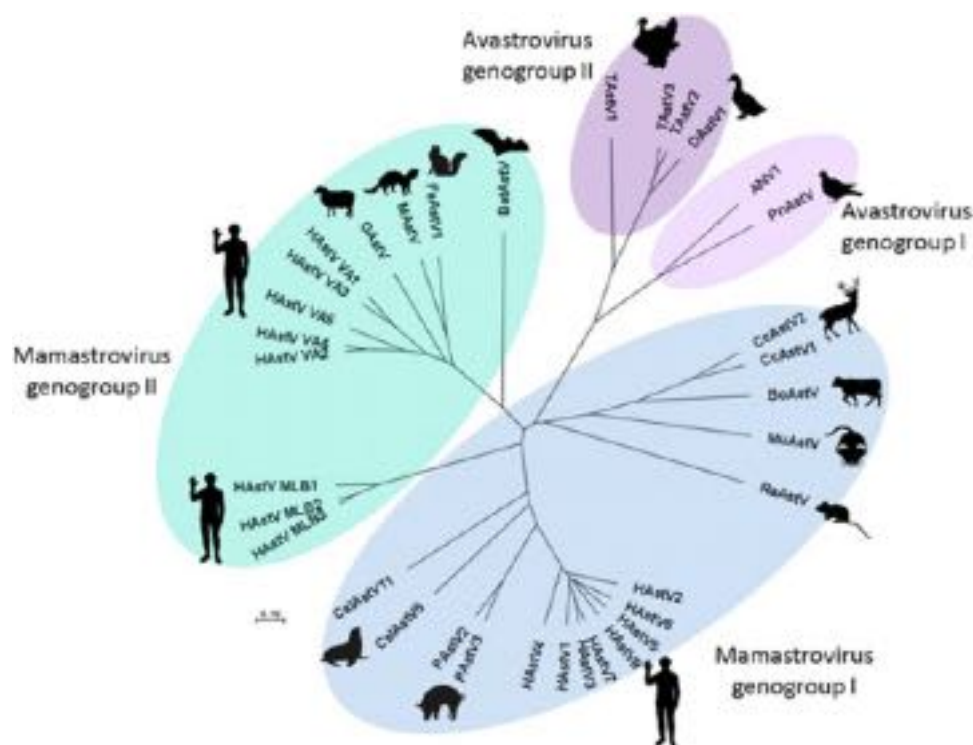


Fig. 1 Phylogenetic classification of *Astroviridae*. Tree generated using capsid protein amino acid sequences. Based on Ninth Report of the International Committee on Taxonomy of Viruses, 2012.

Prior to 2008, humans were thought to be infected by only eight closely related genotypes of astrovirus known as human astrovirus type 1 through 8 (HAstV1–8). These eight genotypes are now known as the classical HAstV genotypes. With the advent of unbiased full-genome sequencing and new pathogen discovery techniques, eight more novel HAstV genotypes were identified, receiving names after the location of their discovery: Melbourne (MLB 1–3) and Virginia (VA 1–5). In the same year the VA genotypes were discovered, the same viruses were discovered in samples from Nigeria, Pakistan, and Nepal. These viruses are sometimes termed HAstV-HMO due to their relatedness to human, mink and ovine astroviruses, however, for the purposes of this review will be termed HAstV-VA. Phylogenetic analysis revealed that the newly identified MLB and VA genotypes are more closely related to other mammalian astroviruses than the classical genotypes creating three genetically distinct clades of HAstV within Mamastrovirus.

Virion Structure

The star-like appearance of astrovirus noted by Madeley and Cosgrove is due to spikes studded on the virion surface. Within the past decade, extensive progress has been made in our knowledge of the capsid's structure using methods including electron and cryo-electron microscopy, X-ray crystallography, and density ultracentrifugation. The crystal structures for the HAstV8 spike, HAstV1 full capsid, HAstV2 spike, and turkey astrovirus type 2 spike have all been described. These studies demonstrate that astrovirus particles are approximately 43 nm in diameter with $T = 3$ icosahedral symmetry (Fig. 2A) and closely resemble the structure of the human hepatitis E virus virion.

The capsid protein is translated as an inactive VP90 precursor that undergoes cleavage before assembly into an infectious virion. First, intracellular caspases cleave VP90 to generate VP70, of which 180 copies are needed to form an immature virion. Two separate studies reported that this caspase cleavage is necessary for viral particles to be released from the cell. Once released from the cell, extracellular proteases further cleave VP70 of the immature virions into VP34, VP27, and VP25 to form mature virions. VP34 makes up the continuous inner core while VP27/VP25 form the spike domains (Fig. 2A). The immature virion has an estimated 90 spikes which is reduced to just 30 spikes on an infectious particle following this final cleavage. Through antigenic studies, these spikes are thought to contain the binding site to the yet to be identified host cell receptor.

The astrovirus capsid appears to have other functional roles although its full role in the replication cycle remains an area of active research. The capsid has been reported to bind both complement 1q and mannose-binding lectin to inhibit the classical and lectin pathways of complement. This could be a clue as to how astrovirus is able to infect without causing noticeable histological changes. The astrovirus capsid has also been shown to act as a novel enterotoxin in vitro and in vivo.

Genome

The astrovirus genome is a single-stranded positive-sense RNA segment of approximately 6.2–7.7 kb. The genetic material contains both 5' and 3' untranslated regions and three open reading frames (ORFs) flanked by 5' viral genome-linked protein (VPg) and a 3' poly(A) tail.

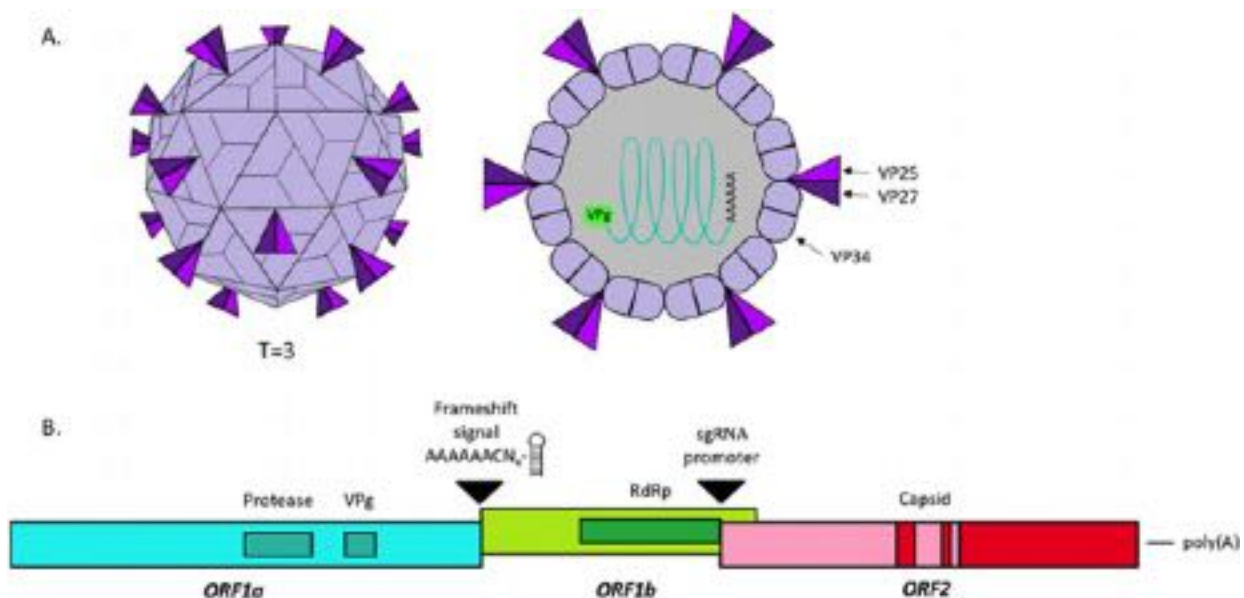


Fig. 2 Astrovirus Genome and Virion Structure. (A) $T = 3$ icosahedral structure of the astrovirus capsid composed by an inner core of VP34 and VP27/VP25 in the spike domains. (B) The three open reading frames (ORF) of astrovirus flanked by a 5' viral genome linked protein (VPg) and a 3' poly(A) tail with known protein coding regions designated by inset boxes. Darker regions of ORF2 indicate hypervariable regions. Modified from SIB Swiss Institute of Bioinformatics, ViralZone Astroviridae.

poly(A) tail (Fig. 2B). ORF1a and ORF1b, located at the 5' end of the genome, both encode the astrovirus nonstructural proteins. While astrovirus could have as many as five nonstructural proteins, the three characterized nonstructural proteins include a serine protease, the VPg, and an RNA-dependent RNA polymerase (RdRp). A ribosomal frameshift signal is present between ORF1a and ORF1b and is critical for the correct translation of the RdRp. ORF2, at the 5' end of the genome, encodes the astrovirus structural protein and is expressed from a subgenomic RNA (sgRNA). This sgRNA was confirmed when two species of viral RNA were detected in astrovirus infected cells, both full-length genomic RNA and a roughly 2.8 kb sgRNA. This allows for an excess of astrovirus capsid protein to be translated and ready for assembly during the viral replication cycle.

Life Cycle

Binding and Entry

Despite dedicated work, much of the astrovirus replication cycle remains unknown. Our understanding comes mainly from studying a few of the classical HAsTV genotypes, and inference from the replication cycles of closely related viral families like *Togaviridae* and *Caliciviridae*. The cellular receptor or receptors for astroviruses are still unknown, and given that different cell types have different susceptibilities for various HAsTV genotypes it is likely a variety of binding factors may be needed for cell entry. For example, while human colorectal adenocarcinoma (Caco-2) cells can support the replication of all HAsTV1–8 genotypes, baby hamster kidney (BHK-21) cells can only be infected by HAsTV2, and human colorectal carcinoma epithelial (HTC-15) cells can support only HAsTV1 replication. Crystallization of the capsid and spike proteins revealed conserved polysaccharide binding motifs meaning these could be involved in viral binding. Nevertheless, pretreatment of Caco-2 cells with heparinase was unable to inhibit viral binding, while pretreatment with chondroitinase was only able to reduce infectivity, showing further research is needed to understand how these sites could be involved in binding.

After attachment of the virus to the host cell, entry appears to be facilitated by clathrin-mediated endocytosis. Electron microscopy has displayed astrovirus particles localized in coated pits and vesicles and blocking clathrin assembly has been shown to inhibit HAsTV8 replication. Once endocytosed, vesicles containing astrovirus particles must reach late endosomes. Blocking Rab7 was shown to block replication of HAsTV8 suggesting the acidic environment found in late endosomes is required for proper uncoating and release of the viral genome.

Uncoating and Replication

While the exact mechanism for viral uncoating is unknown, studies show the process is membrane associated and mediated by an endoplasmic reticulum signal motif located in the C-terminus of VP90. Analysis of recombinant viral-like particles (VLPs) showed that divalent cations in the capsid protein provide stability to the viral particles. Therefore, the acidic environment of late endosomes and typically low cation concentration in the cytoplasm should both promote virus uncoating.

Once the viral genome has been released, the first proteins translated after viral uncoating are the nonstructural proteins encoded by ORF1a and ORF1b. The nonstructural polyprotein, nsp1a, is cleaved by the viral serine protease, conceivably along with unidentified cellular proteases into smaller proteins some whose function is still unknown. One of the nsp1a proteins, nsp1a/4, is a VPg that has been shown to localize with viral RNA and is thought to facilitate viral replication and possibly transcription from the negative-sense RNA template. Although direct interaction between the VPg and viral genome has not been confirmed, pretreatment of viral RNA with proteinase K before transfection into permissive cells significantly reduced the infectious virus produced, indicating a protein moiety is required for successful replication. The production of the capsid protein involves both genomic and sgRNA, with an excess of sgRNA at 12 h post-infection (hpi) coinciding with higher levels of capsid protein production. This is characteristic common among many positive-sense single-stranded RNA viruses as a means of evading repression of translation by the host. While caliciviruses possess a VPg on their sgRNA segments to initiate its translation, it is unclear if astroviruses utilize a similar mechanism.

The precise role host proteins play during astrovirus replication has yet to be fully addressed, although one study has implicated host proteins involved in fatty acid and cholesterol synthesis, phosphatidylinositol and inositol metabolism, and RNA helicase activity, all being necessary viral replication. Our lab has demonstrated that the increase in phosphorylated ERK1/2 protein that occurs within 15 min after viral binding is required for successful viral replication. Recently, others have reported the ubiquitin proteasome system is also needed replication of the viral genome and sgRNA.

Assembly and Release

The location for viral assembly appears to be genotype dependent with HAsTV4 and HAsTV8 accumulating in double membrane vacuoles near the nucleus, while ovine astrovirus has been found in lysosomes and autophagic vesicles. In these membraned sites, viral capsids can self-assemble into full virions or VLPs, and viral release occurs via an unknown, nonlytic mechanism without significant cell death. It has been hypothesized that astrovirus release could occur through similar nonlytic mechanisms used by rotavirus or poliovirus, or through a form of cell membrane destabilization. Overall, many questions remain involving the replication cycle of astrovirus and more research is needed to elucidate key aspects of every step in the cycle including what

receptors and binding factors are used, the mechanism of viral release without inducing cell death, and how cellular factors are involved in this process.

Epidemiology

Astrovirus is spread via the fecal-oral route and can be transmitted through contaminated food and water, human-to-human, or through nosocomial infection. The fecal-oral route of transmission was confirmed through two volunteer studies, where after ingesting astrovirus subjects developed symptoms and detectable serological responses.

Classical HAsTV

Classical HAsTV infections are predominantly reported in pediatric population, but the elderly and immunocompromised are also susceptible to infection. Although less common, healthy adults have been reported with HAsTV infections. The age range of typical infection is newborn to 5 years, with up to 90% of children ages 5 years and up having detectable serum antibodies. Although there tends to be a seasonal pattern with higher incidence in winter, possibly due to enhanced stability in colder temperatures, astrovirus circulates year-round with infections reported in summer months as well.

Classical HAsTVs are thought to be responsible for between 2% and 9% of all acute, non-bacterial diarrhea in children. Reported in 2013, the mean incidence of HAsTV infection was 11% worldwide, with 7% incidence in rural areas and 23% in urban areas. The incidence of infection tends to be higher in developing countries, most likely due to poor hygiene and water treatment practices. Astrovirus infection statistics are greatly hindered by underreporting and asymptomatic infections. However, the development of more sensitive real-time RT-PCR have enabled significantly higher rates of HAsTV infections in asymptomatic children to be detected. With these new screening methods, the rate of HAsTV positive samples from asymptomatic children is thought to be roughly 4%. Of the classical genotypes, HAsTV1 is the most frequently isolated worldwide, with the exception of Mexico where HAsTV2 dominates. The incidence of HAsTV infection has been decreasing over the past three decades from 22% in the 1980s to just 5% in the 2000s. This could be due to novel HAsTVs outcompeting the classical genotypes, since reports have shown their prevalence is rising.

Novel HAsTV

The association between the MLB and VA clades of HAsTV and clinical disease is still somewhat ambiguous compared to the classical genotypes, which makes accurate reporting difficult. Published reports have exhibited this in the large variation in prevalence from one geographic location to another. One review of novel HAsTV prevalence, describes an overall positivity rate from stool samples 1.5% lower than that observed for classic HAsTVs. However, one study from Japan in 2016 found more MLB-HAsTV positive samples, 10.6%, than classic HAsTVs, 5.1%. Although most case reports exhibit a generally low prevalence of novel HAsTVs, studies have revealed the seroprevalence of HAsTV-MLB1 to be 86% and HAsTV-VA1 to be 65% indicating a large portion of the population have seen these viruses previously. In a study of immunocompromised pediatric oncology patients, half of the samples were positive for HAsTV1, however HAsTV-VA2 and HAsTV-MLB1 were also present in 21% and 13% of samples, respectively. The authors of this study did comment that the PCR screening method used was unable to detect other members of the HAsTV-VA and HAsTV-MLB clades, so these genotypes may have been more prevalent than reported.

Clinical Features

Human Disease

The typical symptom of HAsTV infection is watery diarrhea that lasts 2–4 days and can be less commonly accompanied by fever, headaches, abdominal pain and anorexia. Not all HAsTV infections present with symptoms, since many infections in healthy individuals tend to be asymptomatic, which could be due to previous exposure to the virus conferring immunity.

In the immunocompromised population, HAsTV infections are of increased interest due to recent reports of severe symptoms and extra-gastrointestinal dissemination. First reported in 2010, a case report involving a 15-year old boy with X-linked agammaglobulinaemia suggested HAsTV to be the causative agent in encephalitis and meningitis. While virus specific primers were unable to detect the virus, next-generation sequencing (NGS) identified the presence of HAsTV in the patient's biopsy samples. Since then, eight additional cases have linked encephalitis and meningitis to HAsTV infection in predominately immunocompromised patients, though one case was in an otherwise healthy individual. In eight out of the nine cases, a non-classical genotype was identified in patient samples. These studies suggest that astroviruses have the potential to cause systemic infection, highlighting the need to research astrovirus pathogenesis with both classical and non-classical genotypes. Especially when epidemiological studies have shown infections caused by non-canonical viruses are becoming less rare.

Animal Disease

As in humans, diarrhea is the main symptom of astrovirus infection in many animal species including turkeys, chicken, cattle, lambs, piglets, and deer, among others, although cases of asymptomatic infections in animals have been reported. Chicken astrovirus infection has been associated with two different conditions in chicken hatchlings, runting-stunting syndrome and “white chick” syndrome. These syndromes are characterized by decreased hatch-rate, poor weight gain, increased chick mortality and weakness. It can lead to pale plumage in chickens where “white chick” syndrome gets its name. Likewise, astrovirus infections in turkeys also causes stunted growth associated with poult enteritis mortality syndrome (PEMS). PEMS affects newly hatched turkey poults and is characterized by diarrhea, dehydration, and increased mortality. Along with diarrhea associated disease in animals, there have been cases of astrovirus-associated encephalitis and meningitis in mink and cows. Recently, porcine astrovirus type 3 was detected in central nervous system tissues from pigs suffering a neurologic disorder characterized by hind limb weakness to quadriplegia and occasionally convulsions. With improved screening, detection and sequencing techniques, we are now able to detect the variety of symptoms associated with astrovirus infection, however further research is still needed to discover how the virus is causing disease.

Pathogenesis

Obstacles

Several obstacles must be overcome in order to uncover the mechanisms of astrovirus pathogenesis. To date, researchers make do without proper cell culture methods for all astrovirus genotypes or a well-characterized or clinically relevant small animal model. Caco-2 cells are the most common cell culture model for *in vitro* studies on classical HAsV genotypes. However, this is an immortalized cancerous cell line that may not completely mimic the same mechanisms occurring *in vivo*. Furthermore, cell culture systems for the MLB and VA genotypes are limited or non-existent, which greatly limits our ability to study these novel genotypes.

Modeling astrovirus disease *in vivo* has also been challenging. The turkey poult model has historically been the most widely used animal model in astrovirus research. Although poults develop clinical symptoms upon infection with turkey astrovirus type-2 (TAsV2), this virus is an Avastrovirus and genetically distinct from Mamastroviruses. It is still unclear if characteristics of disease resulting from Avastrovirus infection translate to classical or non-classical HAsV infection. Studies using turkey poults are also burdened by the lack of a cell culture system in which to grow virus; researchers are restricted to generating the virus in embryonated turkey eggs. Furthermore, planning experiments is difficult due to a low hatch rate, approximately 50%, decreasing overall sample sizes, and the highly transmissible nature of astrovirus requiring uninfected controls to be housed separately to avoid cross-contamination. Most importantly, turkey-specific antibodies and reagents are rarely commercially available, which greatly reduces which technical methods can be utilized.

In 2012, four murine astroviruses (MuAsV, STL1 – STL4) were identified in mice housed at Washington University in St. Louis, MO. This discovery offered the potential to develop a traditional small animal model to study astrovirus. Unlike TAsV-infected turkey poults, MuAsV-infected mice do not exhibit diarrhea yet exhibit low level persistence and viral shedding for long periods of time. MuAsV has also been identified in mice purchased from multiple vendors, making obtaining mice without underlying astrovirus infection challenging. The characterization of the MuAsV infection is ongoing, but researchers are hopeful this model holds the key many future discoveries. Nevertheless, using the resources available, researchers are expanding the body of knowledge on astrovirus pathogenesis (summarized in [Fig. 3](#)).

Mechanism of Disease

To date, most pathogenesis studies have focused on understanding the mechanism by which astrovirus induces diarrhea. Reports have shown astrovirus infection in turkey poults results in relocalization of the sodium transport proteins NHE3 and SGLT-1 from the cell membrane into the cytoplasm of intestinal epithelial cells, supporting a hypothesis that astrovirus disrupts ionic transport causing sodium malabsorption and osmotic diarrhea. Other intestinal pathogens disrupt the intestinal epithelium by causing inflammation and cell death. However, astrovirus-induced diarrhea occurs without triggering cell death or inflammation, although there is one report of HAsV8 inducing apoptosis in Caco-2 cells. Instead HAsV causes barrier permeability by disrupting intestinal cellular junctions. Tight junctions are highly regulated cell-cell associations that maintain cell polarity and regulate the passage of nutrients, other solutes and microorganisms from the intestinal lumen across the epithelium. Cellular junctions are comprised of the tight junction complex, which includes proteins such as occludin, claudin, ZO-1, and cadherins. Loss of tight junctions can result in increasing fluid in the lumen of the intestine and diarrhea.

Multiple studies report that AsV infection alters permeability of the intestinal barrier caused by redistribution of junctional proteins. *In vitro*, HAsV infection caused a drop in transepithelial resistance (TER) in Caco-2 monolayers and allowed for greater flux of FITC–dextran across the cell monolayer. The decrease in TER and increase in flux began between 16 and 20 hpi, with maximum permeability achieved by 36 and 48 hpi. HAsV1 has been shown to increase intestinal barrier permeability *in vitro*, possibly due to a reduction in actin fibers, reduction of occludin from the junctional complex and loss of E-cadherin by 24 hpi. Studies in the turkey poult yielded comparable results, where infected poults had increased luminal-to-serosal flux and disruption

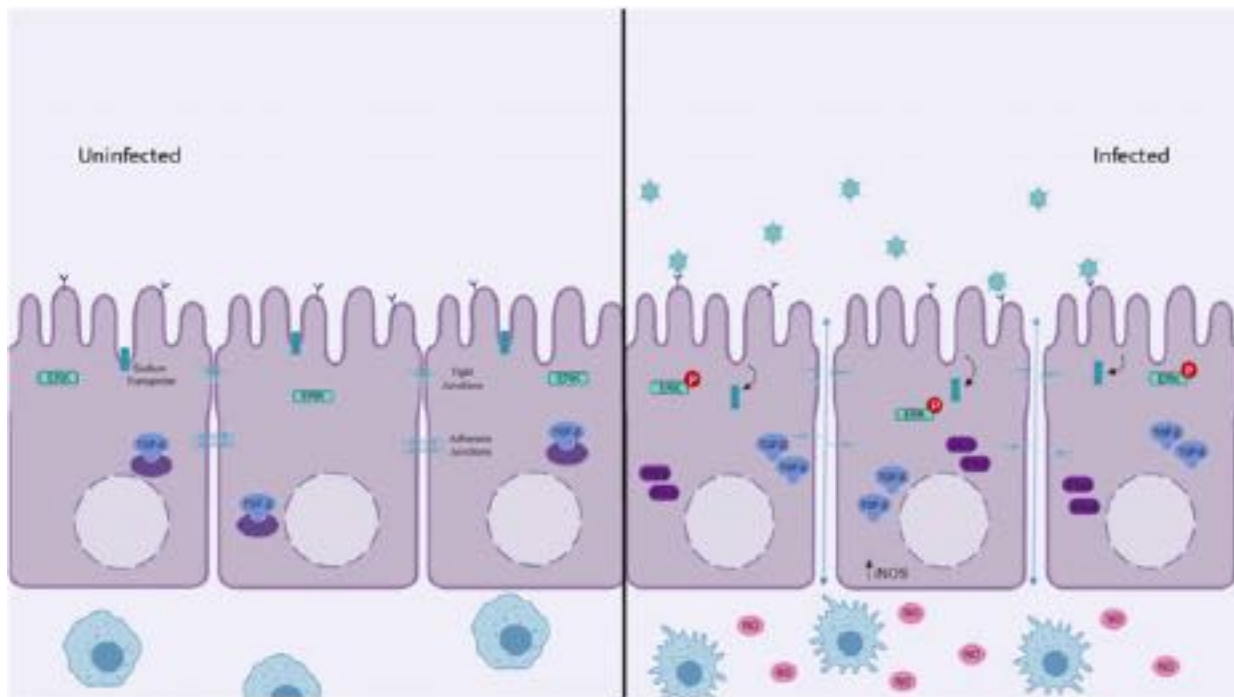


Fig. 3 Astrovirus increases barrier permeability in the gut epithelium. (A) Normal intestinal epithelial cells maintain tightly regulated barrier permeability. (B) In infected cells, intestinal epithelial cells upregulate iNOS inducing macrophages to secrete NO, and TGF- β within the epithelial cells is activated. IFN- β is induced in an effort to control infection. Tight junction proteins and the sodium transporters are relocalized, leading to the breakdown of barrier permeability allowing exchange of ions and solutes across the intestinal epithelium.

of F-actin. There is evidence that MuAstV infection also increased intestinal permeability, as indicated by orally administered FITC-dextran crossing the intestinal epithelium reaching the blood.

Remarkably, astrovirus-induced disruption of barrier integrity was found to be independent of active replication. Reports by Moser *et al.* found both UV-inactivated HAsTV and recombinant capsid protein could decrease barrier permeability in Caco-2 cells as well as activate ERK1/2. In vivo, administration of the recombinant TAsTV2 capsid protein induced acute diarrhea, and capsid inoculated animals experienced relocalization of SGLT-1 from the plasma membrane, comparable to those infected with virus.

Immune Response

The immune response to astrovirus is not well-characterized, but studies suggest that infection does not produce overt inflammation. It is assumed that the innate immune system plays a vital role in controlling the virus because the course of infection with astrovirus is relatively short. One essential arm of the innate immune system is the interferon response. Astrovirus replication was found to be impacted by the type I interferon (IFN) response. Two reports have shown IFN- β levels increase after HAsTV infection and the addition of exogenous IFN- β can protect epithelial barrier integrity. Contrasting the induction of barrier permeability, productive replication is required to induce IFN- β in vitro. In vivo, IFNAR knockout mice, which lack the IFN- α receptor making them unable to respond to type I IFNs, could not clear an astrovirus infection unlike wild-type mice.

Along with inhibiting complement factors, astrovirus may be modulating the immune response by activating the immunosuppressive cytokine, TGF- β . Infected turkey poults exhibited increased levels of active TGF- β to up 12 days after infection. Astrovirus has been shown to not only infect intestinal epithelial cells but also intestinal macrophages, although infection in macrophages led to abortive replication. While astrovirus infection results in decreased macrophage function and viability, infected macrophages produce more nitrogen oxide (NO) because of increased inducible NO synthase (iNOS) produced by infected epithelial cells. The NO produced by macrophages has been reported to then inhibit astrovirus replication in turkey poults.

The innate immune response controls the initial spread of astrovirus and enables the host to mount an effective humoral and cell-mediated adaptive response. Nearly 70% of healthy adults have detectable antibodies against astroviruses. Studies in human volunteers have shown a direct link between having anti-astrovirus antibodies and disease, where subjects without antibodies to the virus had more severe symptoms. Further reports have shown individuals can develop anti-astrovirus antibodies without the presence of symptoms as is the case with poultry abattoir workers with antibodies against TAsTV. This finding was contrary to the widely held belief that astrovirus infections were species-specific, and provided evidence supporting the possibility for zoonotic transmission. However, further work is needed in this area of investigation.

Recently, antibodies that neutralized viral infectivity have been discovered. Bogdanoff *et al.* have mapped the neutralizing epitopes of HAstV2 demonstrating only antibodies to spike protein, not the core, can prevent binding to Caco-2 cells. Additional analysis of several genotype specific neutralizing antibodies determined a quaternary epitope on each side of the dimeric capsid spike necessary for neutralization. These studies could help determine components for future vaccines to induce the production of cross-protective neutralizing antibodies against astrovirus.

In terms of cell-mediated adaptive response, astrovirus specific T cells are also thought to play a significant role in viral clearance. A challenge of biopsy specimens from the small intestine of healthy adults with inactivated HAstV demonstrated HAstV-specific CD4⁺ and CD8⁺ T cells reside in these tissues and could prevent reinfection. Recent work in the mouse model has shown Rag1 knockout mice, which lack B and T cells, shed more virus in their feces than wild-type mice. This study also detected higher viral replication in intestinal tissues, mesenteric lymph nodes, spleen, liver, and kidneys of Rag1 knockout mice, indicating B and T cells are necessary in preventing systemic spread of astrovirus.

Diagnosis

Initially, HAstV was detected via electron microscopy. However, using electron microscopy is time consuming and is impractical for high-throughput screening purposes. In 1995, the first conventional RT-PCR and enzyme immunoassays were developed to determine the antigenic groups of HAstV. Several real-time RT-PCR methods have now been developed for quick screening of HAstV, including many that are multiplexed to detect additional enteric viruses. However, these methods do not detect non-classical HAstV VA or MLB genotypes. To date, there is only one real-time RT-PCR multiplex method accurately detects classical HAstV1–8, MLB1, and VA2. Developing more comprehensive real-time RT-PCR methods to detect the 16 known genotypes has been hindered by limited sequence information available through public databases. Adding to the sequences available and validating new detection methods is essential for researchers to better delineate astrovirus epidemiology and relationship to disease.

Treatment

Currently, there are no vaccine or medical treatment options for HAstV. This is in part due to the short duration and relatively mild symptoms that HAstV infection causes. The low interest in developing a HAstV vaccine could be because of its believed low clinical impact or the need for a multivalent vaccine to cover multiple genotypes. Although most infections resolve without intervention, if treatment is necessary, it generally involves alleviating symptoms. HAstV-induced diarrhea has the potential to cause dehydration in infected individuals and therefore may require oral or intravenous fluid replacement. Since immunocompromised individuals could be at risk of astrovirus spreading systemically, it has been suggested intravenous IgG be used as treatment, although more studies are needed to determine the efficacy of such treatments.

Prevention

The best form of prevention is to stop the spread of the virus. In order to fully disinfect surfaces and contaminated fomites most healthcare facilities regularly use 90% ethanol, however this treatment was shown to be ineffective against astrovirus. Astrovirus is a remarkably hearty virus that can remain stable at low pH, is resistant to many detergents, and even lipid and chlorine solvents are incapable of inactivating the virus. However, reports have shown that formaldehyde and peroxymonosulfate can fully inactivate the virus. Temperature modulation also has very little effect since astrovirus is also very stable at low temperatures and can remain infectious for up to ten years.

Astrovirus has been detected in environmental water samples and shown to persist in drinking water, proving the importance of water treatment in preventing the spread of the virus. Typical water treatment involves chlorine disinfection with residual free chlorine in the range of 0.2–1 mg/L, and a maximum residual disinfectant level of 4mg/L allowed by the Environmental Protection Agency. Astrovirus infectivity has been shown to be decreased by 3 logs at the height of the typical range, 1mg/L free chlorine, for two hours. Most studies to date on disinfection and inactivation have focused on classical HAstV genotypes. While bleach use is recommended for the inactivation of novel HAstV genotypes, more research is necessary to determine exact protocols for their inactivation.

Further Reading

- Arias, C.F., DuBois, R.M., 2017. The astrovirus capsid: A review. *Viruses* 9 (1), 15.
Bogdanoff, W.A., Perez, E.L., López, T., Arias, C.F., DuBois, R.M., 2017. Structural basis for escape of human astrovirus from antibody neutralization: Broad implications for rational vaccine design. *Journal of Virology* 92 (1).
Bosch, A., Pintó, R.M., Guix, S., 2014. Human astroviruses. *Clinical Microbiology Reviews* 27 (4), 1048–1074.

- Cortez, V., Freiden, P., Gu, Z., *et al.*, 2017. Persistent infections with diverse co-circulating astroviruses in pediatric oncology patients, Memphis, Tennessee, USA. *Emerging Infectious Diseases* 23 (2), 288–290.
- Cortez, V., Meliopoulos, V.A., Karlsson, E.A., *et al.*, 2017. Astrovirus biology and pathogenesis. *Annual Review of Virology* 4 (1), 327–348.
- Donato, C., Vijaykrishna, D., 2017. The broad host range and genetic diversity of mammalian and avian astroviruses. *Viruses* 9 (5).
- Espinosa, R., López, T., Bogdanoff, W.A., *et al.*, 2019. Isolation of neutralizing monoclonal antibodies to human astrovirus and characterization of virus variants that escape neutralization. *Journal of Virology* 93 (2), doi:10.1128/JVI.01465-18.
- Gu, Z., Zhu, H., Rodríguez, A., *et al.*, 2015. Comparative evaluation of broad-panel PCR assays for the detection of gastrointestinal pathogens in pediatric oncology patients. *The Journal of Molecular Diagnostics* 17 (6), 715–721.
- Johnson, C., Hargest, V., Cortez, V., Meliopoulos, V.A., Schultz-Cherry, S., 2017. Astrovirus pathogenesis. *Viruses* 9 (1), 22.
- Karlsson, E.A., Small, C.T., Freiden, P., *et al.*, 2015. Non-human primates harbor diverse mammalian and avian astroviruses including those associated with human infections. *PLOS Pathogens* 11 (11), e1005225.
- Meliopoulos, V.A., Marvin, S.A., Freiden, P., *et al.*, 2016. Oral administration of astrovirus capsid protein is sufficient to induce acute diarrhea in vivo. *mBio* 7 (6).
- Mendez, E., Arias, C., 2013. Astrovirus. In: Fields, B., Knipe, D., Howley, P. (Eds.), *Fields Virology*, sixth ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Moser, L.A., Carter, M., Schultz-Cherry, S., 2007. Astrovirus increases epithelial barrier permeability independently of viral replication. *Journal of Virology* 81 (21), 11937–11945.
- Schultz-Cherry, S., 2013. *Astrovirus Research*. New York: Springer.
- Vu, D.L., Bosch, A., Pintó, R.M., Guix, S., 2017. Epidemiology of classic and novel human astrovirus: Gastroenteritis and beyond. *Viruses* 9 (2).

Avian Hepadnaviruses (*Hepadnaviridae*)

Allison R Jilbert, Georget Y Reaiche-Miller, and Catherine A Scougall, The University of Adelaide, Adelaide, SA, Australia

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A.R. Jilbert, W.S. Mason, Hepadnaviruses of Birds, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00414-3.

Nomenclature

nt nucleotide(s)

Glossary

anti-DHBc antibodies Antibodies to DHBcAg
anti-DHBs antibodies Antibodies to DHBsAg
cccDNA Covalently closed circular DNA
CTL Cytotoxic T lymphocyte(s)
DHBcAg Duck hepatitis B virus core Ag
DHBcAg Duck hepatitis B virus e Ag
DHBsAg Duck hepatitis B virus surface Ag

DHBV Duck hepatitis B virus
dsDNA Double stranded-linear hepadnavirus DNA
HBV Human hepatitis B virus
ORF Open reading frame
rcDNA Relaxed circular hepadnavirus DNA
RT Reverse transcriptase
TP Terminal protein
TR Terminal redundancy

Classification

Avian hepadnaviruses belong to the avihepadnavirus genus of the hepadnaviridae family and are related phylogenetically to human hepatitis B virus (HBV) and other viruses of the orthohepadnavirus genus, which infect mammals. Avian hepadnaviruses are related phylogenetically to each other through similarities in genome sequence and organization of open reading frames (ORF). Only two species have been assigned within the avihepadnavirus genus: duck hepatitis B virus (DHBV), first isolated from Pekin ducks (*Anas domestica*), and heron hepatitis B virus (HHBV), isolated from grey herons (*Ardea cinerea*). HHBV was assigned as a species based both on genome divergence and a host range difference from DHBV. DHBV isolates have been found not only in domesticated ducks, but also in the puna teal (*Anas puna*), the Chiloe wigeon (*Anas sibilatrix*) and, in the wild, the mallard, the species from which Pekin ducks and most other domesticated ducks (except Muscovy) were derived.

Avian hepadnaviruses less closely related to DHBV but currently unassigned as distinct species have also been isolated. Viruses isolated from geese include the Ross' goose hepatitis virus (RGHV) from Ross' geese (*Anser rossii*) and Mandarin ducks (*Aix galericulata*), the snow goose hepatitis B virus (SGHBV) from snow geese (*Anser caerulescens*) and the sheldgoose hepatitis B virus from the ashy-headed sheldgoose (ASHBV) (*Chloephaga poliocephala*) and Orinoco sheldgoose (OSHBV) (*Neochen jubata*). The crane hepatitis B virus (CCHBV) has been isolated from demoiselle cranes (*Anthropoides virgo*) and grey crowned cranes (*Balearica regulorum*). Parrot hepatitis B virus (PHBV) has been isolated from ring-necked parakeets (*Psittacula krameri*), an African grey parrot (*Psittacus erithacus*), an Alexandrine Parakeet (*Psittacula eupatria*) and a crimson rosella (*Platycercus elegans*). The stork hepatitis B virus (STHBV) has been isolated from white storks (*Ciconia ciconia*) and the elegant-crested tinamou hepatitis B virus (ETHBV) has been isolated from elegant-crested tinamou (*Eudromia elegans*) (Fig. 1(A); Table 1A).

Virus Structure

Avian hepadnaviruses are enveloped viruses possessing an icosahedral nucleocapsid and ~3000 nt relaxed circular double-stranded DNA (rcDNA) genome (Fig. 2). DHBV, the prototypic avian hepadnavirus, has a diameter of 40–45 nm. The envelope of DHBV is composed of a lipid bilayer containing 2 transmembrane DHBV surface antigen (DHBsAg) proteins, PreS/S and S (36 and 17 kDa respectively). These two proteins are synthesized from the same ORF, with initiation of PreS/S occurring upstream of S and share a common carboxy-terminus. An additional ~28 kDa DHBsAg protein has been detected in DHBV-infected liver but it is unclear if this is a degradation product of PreS/S or is translated from an AUG codon mapping between the start sites of PreS/S and S. The envelope proteins, PreS/S and S, participate in virion formation and also self-assemble into non-infectious DHBsAg particles. DHBsAg assembles in the endoplasmic reticulum and is released from infected hepatocytes in a 500–1000-fold excess over infectious virions. DHBsAg particles do not contain a virus nucleocapsid or viral nucleic acids and are pleomorphic and spherical, with a diameter of ~35–60 nm. Many are almost indistinguishable from DHBV virions by electron microscopy (EM).

The nucleocapsid of DHBV is icosahedral, as observed by cryo-EM, has a diameter of ~38 nm and is composed of 90 or 120 subunits arranged in T = 3 or T = 4 symmetry. The subunits are dimers of the 30 kDa non-glycosylated DHBV core antigen (DHBcAg) protein, or C protein. The rcDNA genome, virus encoded polymerase and cellular chaperones are contained in the nucleocapsid. The DHBV polymerase is a 90 kDa protein that functions as an RNA-dependent DNA polymerase and a DNA-

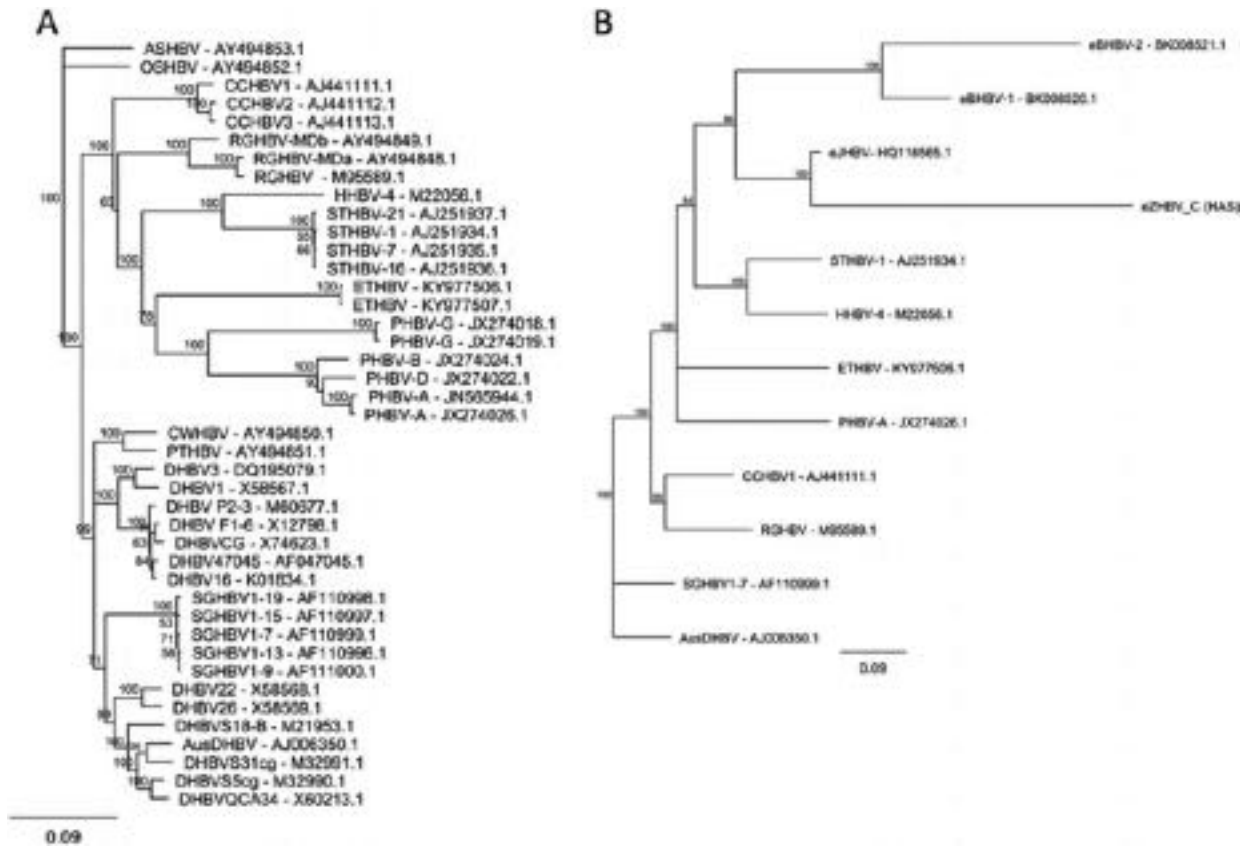


Fig. 1 Phylogenetic relationships of exogenous avian hepadnaviruses and endogenous avian hepadnavirus elements. (A). Phylogenetic analysis of full-length exogenous avian hepadnavirus genomes (Table 1A) was performed using ClustalW sequence alignment with the ClustalW cost matrix, the Neighbor-Joining method with the HKY nucleotide model and 1000 bootstraps and is shown as an unrooted tree. Scale bar = 0.09 nucleotide substitutions per site. (B). Phylogenetic analysis of partial polymerase amino acid (AA) sequences of the endogenous avian hepadnavirus elements, eZHBV_C (HAS) (Hypothetical Ancestral Sequence described by Suh, A., Brosius, J., Schmitz, J., Kriegs, J.O., 2013. The genome of a mesozoic paleovirus reveals the evolution of hepatitis B viruses. *Nature Communications* 4, 1791; 778 AA), eHBV-1-BK008520.1 (679 AA), eHBV-2-BK008521.1 (592 AA) and eHBV-HQ116565.1 (195 AA) (Table 1B), compared to full-length polymerase AA sequences (785–793 AA) of selected exogenous avian hepadnaviruses (Table 1A). Analysis was performed using ClustalW sequence alignment with the BLOSUM cost matrix, Neighbor-Joining method with the Jukes-Cantor model and 1000 bootstrap replicates and is shown as an unrooted tree. Scale bar = 0.09 AA substitutions per site. The percentage of bootstrap replicates that supported each branch is shown in both A and B.

dependent DNA polymerase. The polymerase protein has 4 separate domains listed from the N-terminal end, including a terminal protein (TP), 'spacer', reverse transcriptase (RT) and RNase H domain.

Also released from hepatocytes and circulating in the bloodstream of infected ducks is DHBV e antigen (DHBeAg), which includes proteolytically processed forms of the PreC/C protein ranging in size from a non-glycosylated 27 kDa to ~30–33 kDa glycosylated forms. DHBeAg is thought to be similar in function to the better described HBV e antigen: based upon studies in HBV transgenic mice, HBV e antigen appears to suppress the host immune response to HBV. Thus, it is possible that DHBeAg performs a similar role, facilitating the development of chronic DHBV infections.

Genome

The avian hepadnaviruses have rcDNA genomes with similar genome organization to the orthohepadnaviruses, including the presence of overlapping ORFs all present on the negative strand (Fig. 2). The genomes of DHBV isolates range in size from 3021 to 3027 nt (Table 1A). The negative strand of rcDNA is nicked, with a 9 nt terminal redundancy (r) and with the virus polymerase covalently attached to its 5' end. Similarly, the positive strand of the DHBV genome has an 18 nt RNA primer attached to its 5' end and is incomplete in rcDNA genomes, with a minimum gap of 12 nt (Fig. 2); as illustrated below (Fig. 4), this gap is likely due to the inability of the polymerase to displace the positive strand RNA primer from DR2.

The negative strand of rcDNA contains a P ORF, PreS/S ORF and PreC/C ORF, encoding the polymerase, PreS/S and S proteins, and DHBeAg and DHBeAg proteins, respectively. The P ORF, read in a +1 frame, completely overlaps the PreS/S ORF, read in a

Table 1A Exogenous Avian Hepadnaviruses

Clone name	Accession ^a	Alt. Name ^b	Genome Size (nt)	% ^c	Host	Species	Location ^d
<i>'Chinese' DHBV:</i>							
AusDHBV	AJ006350.1		3027	100.0	Duck, Pekin	<i>Anas domesticus</i>	Australia
DHBVS31cg	M32991.1	HPUS31CG	3027	95.9	Duck, white	<i>Anas domesticus</i>	Shanghai, China
DHBVS5cg	M32990.1	HPUS5CG	3027	94.8	Duck, brown	<i>Anas domesticus</i>	Shanghai, China
DHBVQCA34	X60213.1	DHVBCG	3027	94.7	Duck, domestic	<i>Anas domesticus</i>	Shanghai, China
DHBVS18-B	M21953.1	HPUGA	3024	93.7	Duck, domestic	<i>Anas domesticus</i>	Shanghai, China
DHBV22	X58568.1		3024	91.9	Duck, domestic	<i>Anas domesticus</i>	Chi-tung County, China
DHBV26	X58569.1		3024	91.4	Duck, domestic	<i>Anas domesticus</i>	Shanghai, China
<i>'Western country' DHBV:</i>							
DHBV1	X58567.1		3021	90.0	Goose, domestic	<i>Anas domesticus</i>	Germany
DHBV3	DQ195079.1		3021	90.6	Duck, Pekin	<i>Anas domesticus</i>	Germany
DHBV P2-3	M60677.1	HPUCGE	3021	89.7	Duck, Pekin	<i>Anas domesticus</i>	USA
DHBV16	K01834.1	HPUCGD	3021	89.5	Duck, Pekin	<i>Anas domesticus</i>	USA
DHBV47045	AF047045.1		3021	89.5	Duck	<i>Anas domesticus</i>	Canada
DHBV F1-6	X12798.1		3021	89.4	Duck	<i>Anas domesticus</i>	Germany
DHBVCG	X74623.1	IDHBV	3021	89.0	Duck	<i>Anas domesticus</i>	India
SGHBV1-13	AF110996.1		3024	88.9	Snow Goose	<i>Anser caerulescens</i>	Germany
SGHBV1-15	AF110997.1		3024	89.0	Snow Goose	<i>Anser caerulescens</i>	Germany
SGHBV1-19	AF110998.1		3024	88.9	Snow Goose	<i>Anser caerulescens</i>	Germany
SGHBV1-7	AF110999.1		3024	88.9	Snow Goose	<i>Anser caerulescens</i>	Germany
SGHBV1-9	AF111000.1		3024	89.0	Snow Goose	<i>Anser caerulescens</i>	Germany
PTHBV	AY494851.1		3024	88.9	Puna Teal	<i>Anas puna</i>	USA
CWHBV	AY494850.1		3024	88.9	Chiloe Wigeon	<i>Anas sibilatrix</i>	USA
OSHBV	AY494852.1		3018	86.0	Orinco Sheldgoose	<i>Neochen jubata</i>	USA
ASHBV	AY494853.1		3051	85.1	Ashy-Headed Sheldgoose	<i>Chloephaga poliocephala</i>	USA
CCHBV1	AJ441111.1		3021	83.3	Grey Crowned Crane	<i>Balearica regulorum</i>	Germany
CCHBV2	AJ441112.1		3018	83.3	Grey Crowned Crane	<i>Balearica regulorum</i>	Germany
CCHBV3	AJ441113.1		3018	83.4	Grey Crowned Crane	<i>Balearica regulorum</i>	Germany
RGHBV-MDa	AY494848.1	MDHBVa	3021	81.8	Mandarin Duck	<i>Aix galericulata</i>	USA
RGHBV-MDb	AY494849.1	MDHBVb	3012	82.7	Mandarin Duck	<i>Aix galericulata</i>	USA
RGHBV	M95589.1	HPUGENM	3018	81.7	Ross' Goose	<i>Anser rossii</i>	USA
STHBV-1	AJ251934.1	SHE251934	3033	76.6	White Stork	<i>Ciconia ciconia</i>	Germany
STHBV-7	AJ251935.1	SHE251935	3033	76.6	White Stork	<i>Ciconia ciconia</i>	Germany
STHBV-16	AJ251936.1	SHE251936	3033	76.5	White Stork	<i>Ciconia ciconia</i>	Germany
STHBV-21	AJ251937.1	SHE251937	3033	76.6	White Stork	<i>Ciconia ciconia</i>	Germany
HHBV-4	M22056.1	HPUCG	3027	76.3	Grey Heron	<i>Ardea cinerea</i>	Germany
ETHBV	KY977506.1	160014	3024	75.6	Elegant-Crested Tinamou	<i>Eudromia elegans</i>	Germany
ETHBV	KY977507.1	160050	3024	75.6	Elegant-Crested Tinamou	<i>Eudromia elegans</i>	Germany
PHBV-A	JN565944.1	PL	3048	74.5	Ring-necked Parakeet	<i>Psittacula krameri</i>	Poland
PHBV-A	JX274026.1	P337	3042	74.9	Ring-necked Parakeet	<i>Psittacula krameri</i>	Poland
PHBV-B	JX274024.1	P410	3042	75.3	African Grey Parrot	<i>Psittacula erithacus</i>	Poland
PHBV-D	JX274022.1	P830	3039	74.5	Alexandrine Parakeet	<i>Psittacula eupatria</i>	Poland
PHBV-G	JX274018.1	P902	3036	73.3	Crimson Rosella	<i>Platycercus elegans</i>	Poland
PHBV-G	JX274019.1	P1233	3036	73.3	Ring-necked Parakeet	<i>Psittacula krameri</i>	Poland

^aGenBank accession number.^bAlternate name.^cPercentage sequence identity to AusDHBV-AJ006350.1.^dLocation of the bird/collection where the virus is found.

+ 2 frame, leading to out of frame overlap between the 'spacer' and RT domains of the polymerase and the PreS and S regions of the surface proteins respectively. The P ORF also overlaps out of frame with the downstream region of the C ORF, read in a + 3 frame, leading to overlap between the TP domain of the polymerase and the C-terminal domain of DHBcAg (Fig. 2). These overlaps place evolutionary constraints on the genome with single mutations potentially leading to changes in more than one ORF. The positive strand of rcDNA does not encode any functional ORFs.

The DHBV genome was originally thought to contain only 3 ORFs, and to lack an ORF encoding a protein analogous to the orthohepadnavirus X protein. This is surprising because the HBV X protein is required for transcription of HBV DNA, which is otherwise suppressed by a host-complex targeting episomal DNAs. It was subsequently discovered that DHBV has a fourth ORF, the X ORF, which lacks a conventional start codon but can be read in + 2 frame to direct synthesis of a candidate DHBV X protein. The significance of this

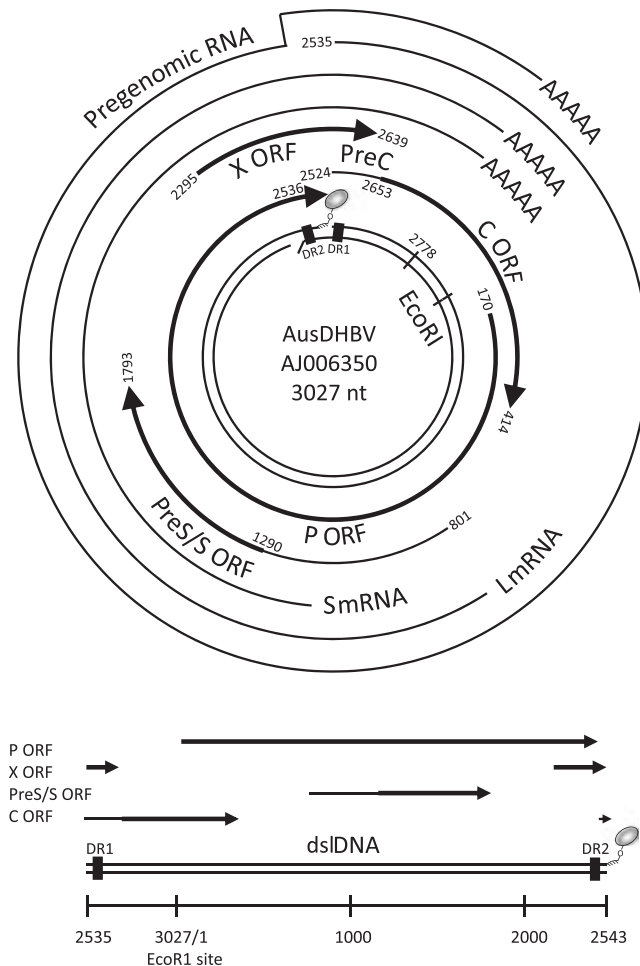


Fig. 2 DHBV genome organization. The rcDNA genome of AusDHBV (GenBank AJ006350) is shown at the top, with the virus polymerase covalently attached to the 5' end of the negative strand and an 18 nt RNA primer attached to the 5' end of the positive strand. The genome is numbered clockwise from a conserved EcoRI site at nt 3027/1 and contains 2 direct repeat (DR) sequences (DR1 and DR2, nt 2541–2552 and 2483–2494 respectively). The location of the P, PreS/S, PreC/C and X ORFs, pregenomic, L and S mRNAs and common polyadenylation site (nt 2778–2783) are also shown. None of the major mRNAs are spliced. Thus, each mRNA has its own promoter. Pregenomic RNA has a terminal redundancy (TR) of 243 nt and is the mRNA for both the DHBcAg and polymerase proteins from the C ORF and P ORF respectively. Pregenomic RNA lacks the AUG of the PreC/C ORF, and DHBcAg is synthesised instead from an mRNA a few nt longer at the 5' end than pregenomic RNA (not shown). The L and S mRNAs encode the DHBsAg proteins, PreS/S and S, respectively. The X ORF, which lacks a conventional start codon, is used to produce X mRNA (not shown) allowing synthesis of a candidate X protein (as described in the text). The more rarely synthesized double stranded-linear DNA (dsDNA) is shown at the bottom.

observation is still unclear. Knockout of the X gene of DHBV did not alter the time course of DHBV infection: the X gene deletion mutant spread rapidly through the liver of newly-hatched ducks and caused persistent infection. Interestingly, other avian hepadnaviruses including HHBV, STHBV, RGHV, ASHBV and OSHBV, all have an AUG codon near the beginning of the X ORF.

The difference in size between the genomes of the avian (e.g., DHBV, 3021–3027 nt) and the orthohepadnaviruses (e.g., HBV, 3200–3300 nt), is due in part to the absence of a 150 nt stretch in the S ORF of the avian virus. This region in the HBV genome encodes the 'a' determinant, a highly immuno-dominant region present in HBV surface antigen. Most neutralizing anti-HBV surface antigen antibodies are directed towards the 'a' determinant. Indeed, HBV strains with mutations in the 'a' determinant can establish infection and replicate in vaccinated hosts as vaccine escape mutants. The effect of the absence of an 'a' determinant in DHBV is unknown but both DHBV PreS/S and S proteins can induce high-titer neutralizing antibodies.

Life Cycle

Although the avian and orthohepadnaviruses share only 40% genome sequence homology, their life cycles and replication strategies are similar. In brief, their rcDNA genome is converted in the nucleus to covalently closed circular DNA (cccDNA), which is used as the transcriptional template for all the viral mRNAs, including the pregenome (Fig. 2). Pregenomic RNA, in addition to functioning as the mRNA for DHBcAg and polymerase, is encapsidated by DHBcAg and reverse transcribed, in the cytoplasm, into rcDNA and double stranded-linear DNA (dsDNA) (Fig. 3).

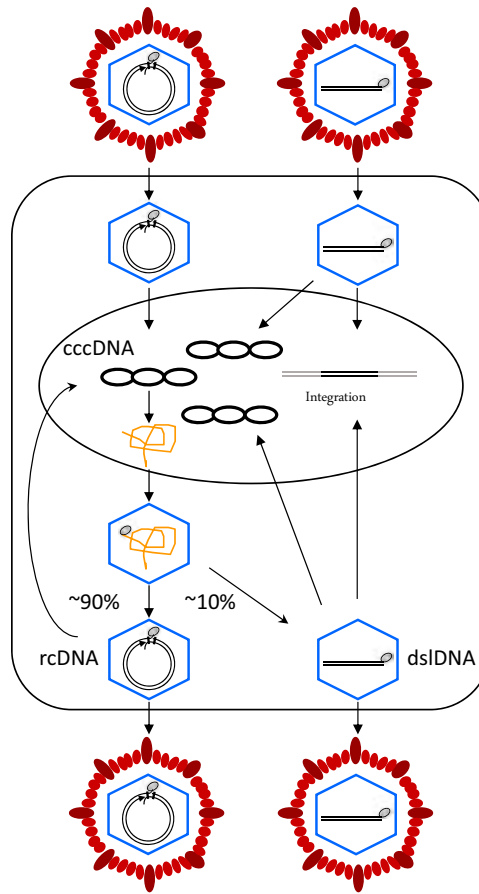


Fig. 3 DHBV infection of hepatocytes. Virions, shown at the top may have either an rcDNA or, more rarely, a dsDNA genome. Upon infection, virus nucleocapsids lose their envelope and move to the nucleus. Virus DNA is released from nucleocapsids and probably enters the nucleus at nuclear pores. Both rcDNA and dsDNA give rise to cccDNA, the virus transcriptional template. cccDNA is used by host cell RNA polymerase II to make all viral mRNAs including the pregenomic RNA that is reverse transcribed into rcDNA or dsDNA. The dsDNA form also has a propensity to integrate into host cell DNA. Early in infection, cccDNA copy number increases to 5–50 copies/cell by intracellular amplification, in which newly made rcDNA and dsDNA migrate to the nucleus rather than being exported as progeny virus; later, nucleocapsids containing rcDNA and dsDNA bud into the ER and are released as progeny virus.

Hepatocytes, the major parenchymal cell of the liver, are the primary site of DHBV infection. Hepatocyte specificity is determined by a cell surface receptor for virus binding and entry and by a viral preference for liver-enriched transcription factors. Receptor binding occurs via the PreS region of DHBsAg and is followed by receptor-mediated endocytosis. The cellular protein carboxypeptidase D (180 kDa) binds DHBsAg particles and DHBV with high affinity and is found on both internal and external membranes of the cell. An additional DHBV binding protein, glycine decarboxylase (120 kDa), has also been identified in liver, kidney and pancreas, all sites of DHBV infection. These 2 proteins are potential components of a DHBV receptor complex but definitive proof that they mediate virus infection is still lacking. In contrast, the receptor used by HBV was recently identified as the human sodium taurocholate co-transporting polypeptide (hNTCP). hNTCP, a multiple transmembrane transporter predominantly expressed in the liver, binds the N-terminal region of the HBV large surface protein, leading to viral entry via receptor-mediated endocytosis.

The immediate sequel to DHBV uptake is transport of rcDNA to the nucleus (**Fig. 3**). Following transport, the covalently linked polymerase protein, the RNA primer and the 9 nt terminal redundancy in the negative strand are removed, the positive strand is completed, and ligation of the ends of each DNA strand takes place to form cccDNA. Once formed, cccDNA is transcribed by host RNA polymerase II to produce pregenomic RNA which is translated allowing assembly of nucleocapsids from DHBcAg and packaging of pregenomic RNA and polymerase.

Pregenomic RNA is greater than genomic length due to a terminal redundancy (TR) of 243 nt (**Figs. 2 and 4**). Its packaging into nucleocapsids is facilitated by the presence of an encapsidation signal, epsilon (Σ), located near its 5' end (at nt 2566–2622) and again at the same position (2566–2622) within the 3' TR (**Fig. 4(A)**). Pregenomic RNA also contains DR1, a 12 nt sequence located 6 nt from the 5' end of the pregenome at nt 2541–2552 and at the same position (nt 2541–2552) in the 3' TR, where it is referred to as DR1*. The same 12 nt sequence is located upstream of DR1 at nt 2483–2494, where it is named DR2.

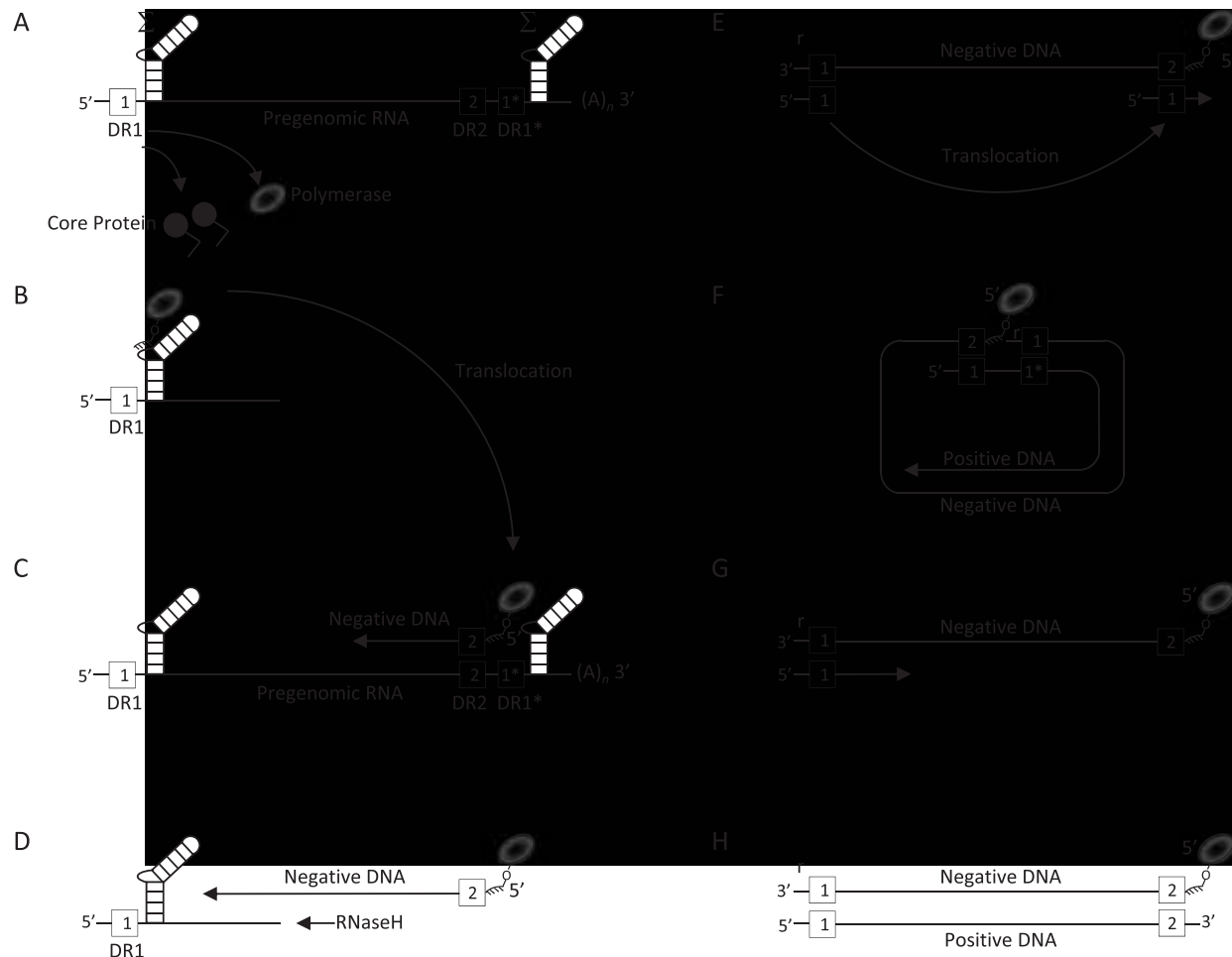


Fig. 4 DHBV DNA synthesis. Pregenomic RNA, which has a terminal redundancy (TR) of 243 nt plus a poly-A tail, has 2 potential encapsidation signals, epsilon (Σ), located at the 5' and 3' ends, but only the 5' copy is functional for pregenome encapsidation. Pregenomic RNA is packaged into nucleocapsids following binding of virus polymerase to the 5' Σ sequence. DNA synthesis then begins with reverse transcription of 4 nt of the bulge in the stem loop structure Σ , 5' UUAC 3' and leads to synthesis of the first 4 nt of negative strand DNA, 5' GTAA 3' as shown in B. Polymerase with the nascent transcript then translocates to the right hand copy of DR1, DR1*, and reverse transcription of the negative strand and degradation of the pregenomic RNA template by RNase H proceeds (C, D). Following completion of the negative strand, the 5' 18 nt of the pregenome, including the 5' cap and DR1, are typically translocated to DR2 and positive strand synthesis then initiates from this RNA primer. In this case, rcDNA formation is of necessity an early step in positive strand elongation (E, F). About 10% of the time, positive strand synthesis initiates without translocation of the primer (G), leading to the formation of dsIDNA (H).

Once the polymerase is translated, it may bind to its own message, pregenomic RNA, via the 5' Σ sequence, blocking its further translation, and facilitating packaging into nucleocapsids. DNA synthesis begins with reverse transcription of 4 nt (5' UUAC 3') in the bulge in the stem loop structure of Σ leading to synthesis of 4 nt of DNA (5' GTAA 3') (Fig. 4(B)). A tyrosine residue in the TP domain of polymerase serves as the primer of reverse transcription and the polymerase remains covalently attached to the 5' end of the newly formed 4 nt DNA sequence.

Following synthesis of the first 4 nt, the DNA-polymerase complex is translocated to DR1* (Fig. 4(C)), where the 4 nt can base pair due to sequence homology. Reverse transcription re-initiates and continues to the 5' end of the pregenome, to produce a full-length negative strand DNA with a 9 nt terminal redundancy. Most of the pregenomic RNA is degraded during negative strand elongation by the RNase H activity of polymerase (Fig. 4(D)). However, the 5' ~18 nt of the RNA pregenome, including the cap and all of the 5' copy of DR1, escapes RNase H degradation and serves as the primer for synthesis of the positive strand.

To facilitate positive strand synthesis, the ~18 nt RNA primer is first translocated to DR2, near the 5' end of the negative strand, where it hybridizes due to the sequence identity of DR1 and DR2 (Fig. 4(E)). Positive strand synthesis then occurs in a 5' to 3' direction to the 5' end of the negative strand template. Circularization then occurs to allow continuation of positive strand synthesis to produce mature rcDNA (Fig. 4(F)). This circularization is facilitated by the 9 nt TR on the negative strand template and the covalent attachment of the TP domain of the polymerase to the 5' end of the negative strand.

In ~10% of cases, translocation of the ~18 nt RNA primer from DR1 to DR2 does not occur and positive strand synthesis occurs by *in situ* priming, to produce dslDNA (Fig. 3, Fig. 4(G, H)). The dslDNA molecule has the 18 nt RNA primer at the 5' end of the positive strand and the polymerase covalently attached to the 5' end of the negative strand and is slightly longer than genome length.

Interestingly, the virus polymerase and the ~18 nt RNA primer remain associated with virus rcDNA and dslDNA genomes throughout virus assembly. Each nucleocapsid contains only a single copy of polymerase, supporting the model that a single protein molecule is simultaneously the TP primer of DNA synthesis and the polymerase for both RNA- and DNA-dependent DNA synthesis.

Nucleocapsids containing newly made rcDNA and dslDNA are enveloped by budding into the endoplasmic reticulum and exported as progeny virus via the Golgi, or are transported to the nucleus to make additional copies of cccDNA, typically found at 5–50 copies/hepatocyte (Fig. 3). DHBV cccDNA molecules exist within the hepatocyte nucleus as a population of virus minichromosomes that bind up to 20 nucleosomes. cccDNA does not undergo semi-conservative replication. New copies are therefore formed solely from rcDNA and dslDNA synthesized in the cytoplasm. However, formation of cccDNA from dslDNA involves illegitimate recombination and typically involves loss of sequences, so that the cccDNA formed from dslDNA is generally defective.

Nuclear transport of nucleocapsids containing rcDNA and dslDNA is negatively regulated by the virus envelope proteins, which direct nucleocapsids into the pathway of virus assembly. Negative regulation is essential because excessive accumulation of cccDNA will kill the host cell. Negative regulation may also be important because of the stability of cccDNA, and new synthesis of cccDNA may only be necessary to restore cccDNA levels in progeny cells following mitosis. Negative regulation of cccDNA synthesis also occurs during HBV infection, but it is not known if the HBV envelope proteins have an essential role in this process.

Nuclear transport of nucleocapsids containing dslDNA and, to a lesser extent, rcDNA, can also result in the integration of viral DNA at random sites in the cell chromosome via illegitimate recombination to create unique virus-cell DNA junctions (Fig. 3). Integrated forms of rcDNA and dslDNA are replication incompetent as they cannot be used to transcribe pregenomic RNA, which is greater than genomic length, and because virus sequences may be lost during the integration process, particularly from the ends of the dslDNA molecule. Integrated DNA has been detected in ~0.01%–0.1% of hepatocytes during transient HBV infections and higher levels accumulate during chronic HBV infections. Because integration occurs at random sites in host DNA, virus-cell DNA junctions can be used as molecular markers to identify and track the fate of individual hepatocyte lineages and possibly to gain insights into the development of primary hepatocellular carcinoma (HCC).

Epidemiology

DHBV is naturally transmitted *in ovo* from an infected female duck to the egg, with virus replication occurring in the yolk sac and liver of the developing embryo, leading to immune tolerance and congenital DHBV infection. Congenital infection is likely further facilitated because DHBV infection is not cytopathic. DHBV infections are therefore maintained within flocks and it is not unusual in some commercial flocks to find the presence of DHBV infection in 100% of birds.

DHBV infection can also be transmitted horizontally through parenteral exposure to infected blood and, since infection of adult ducks usually results in transient infection, duck flocks may contain a mixture of congenitally DHBV-infected and DHBV immune ducks. Female ducks that are immune to DHBV will have anti-DHBs and anti-DHBc antibodies circulating in their bloodstream as well as in the egg yolk which they can pass to their ducklings providing the newly hatched ducks with temporary resistance to DHBV infection.

Pathogenesis and Clinical Features

One major difference between DHBV and HBV are the clinical features of infection that result from differences in their mode of transmission and maintenance. DHBV is maintained primarily by *in ovo* transmission to the developing embryo, while HBV is maintained by horizontal transmission via infected body fluids from mother to child at birth or, during the first year of life, by transmission to the child from infected siblings or adults. Thus, both viruses cause persistent infection, but the degree of immune tolerance to DHBV is much greater following *in ovo* transmission than following horizontal transmission of HBV, which may be why immune responses to HBV-infected cells can lead to severe clinical outcomes of chronic HBV infection.

Humans with chronic HBV infection, if exposed to the virus at less than a year of age, often develop a “healthy carrier” state with high levels of virus replication and little liver disease. This healthy carrier state can last for 10–20 years but can progress to flares of immune-mediated liver disease with fluctuating levels of HBV DNA in the liver and bloodstream and chronic liver disease. If HBV infection is acquired in adulthood and chronic infection develops the accompanying immune-mediated liver disease is similar and can include inflammation, liver failure, fibrosis, cirrhosis, or HCC.

Following *in ovo* transmission congenitally DHBV infected ducks support virus replication in >95% of hepatocytes and have high levels of virus DNA and DHBsAg circulating in the bloodstream but this widespread infection generally results in an absence of clinical features and only mild, if any, hepatitis, similar to the “healthy carrier” state in humans. Interestingly, although >95% of hepatocytes remain infected, levels of DHBV DNA and DHBsAg in the bloodstream gradually decrease over 800 days. Whilst the cause of this gradual decline is unknown, anti-DHBc antibodies, which are usually detected within 7–14 days in experimentally infected adult ducks, are not detected until ~90 days post-hatch, consistent with a high degree of immune tolerance and the delayed activation of immune responses.

In summary, despite a high and persistent level of DHBV infection and replication, congenitally DHBV-infected ducks are immune tolerant to DHBV and progression to immune-mediated liver disease, cirrhosis and HCC have not been reported. Although HCC was detected in ducks in a Chinese province, aflatoxin exposure may have been common and the key contributory factor in the development of HCC. Amyloid disease of the liver is common in domestic ducks but does not appear to be linked to DHBV infection.

DHBV can also be transmitted experimentally to result in persistent or transient infection with clinical outcomes determined by the age of the ducks at the time of infection and the dose of virus inoculated. Experimental inoculation of newly-hatched ducks with DHBV invariably results in persistent infection with an absence of immune responses capable of clearing infection or producing chronic liver disease. In experiments where DHBV was titrated in newly-hatched (1-3-day-old) ducks, persistent infection was established by intravenous inoculation of pooled DHBV-positive serum containing the equivalent of ~ 1 virus genome. Virus spread throughout the liver to infect $>95\%$ hepatocytes within 28 days, showing not only that newly-hatched ducks are highly susceptible to DHBV infection but that the DHBV inoculum, which was derived from pooled sera from congenitally DHBV-infected ducks, had a high specific infectivity with a one-to-one ratio of infectious virus particles to DNA genomes. In related experiments in 3-day-old ducks inoculated with 1500 infectious doses of DHBV, infection was first detected in widely-scattered individual hepatocytes, then in pairs and small groups of cells and resulted in infection of $>95\%$ hepatocytes within 13–15 days. Spread of virus occurred both to adjacent hepatocytes presumably by spread within the space of Disse (cell-to-cell spread) and via the blood by release of virus into the liver sinusoids.

Susceptibility to persistent DHBV infection declines with age. As ducks increase in age, larger doses are required to establish persistent infection, consistent with the idea that larger doses of virus are required to overwhelm the developing immune system: inoculation of 2-week-old ducks with 4×10^4 infectious doses of DHBV resulted in infection of randomly-scattered hepatocytes with subsequent virus clearance. Increasing the dose to 1×10^6 resulted in virus spread to $>95\%$ of hepatocytes and persistent infection. As the age of the duck further increases, larger doses are required to establish persistent infection: inoculation of 6-week-old ducks with $1.5\text{--}4.5 \times 10^{10}$ infectious doses of DHBV resulted in transient infection with successful clearance of infected hepatocytes and accompanying increases in activated Kupffer cells and apoptotic hepatocytes in 11 out of 16 ducks, while the remaining 5 ducks developed persistent infection. Persistent infection was accompanied by mild mononuclear cell infiltration of portal tracts, but no evidence of lobular hepatitis or extensive liver damage was seen within a 1-month timeframe.

While showing that susceptibility to persistent infection decreases with age, these studies also raised the question of how the immune system can clear an infection in which virtually all hepatocytes are infected. This is a problem because under most circumstances, including DHBV clearance, the hepatocyte population is mostly self-renewing. This has been partially resolved by subsequent experiments in DHBV-infected ducks, woodchuck hepatitis virus-infected woodchucks and HBV-infected chimpanzees.

Current understanding is that the “clearance phase” of transient hepadnavirus infection involves 3 steps: (1). A strong CD8-positive CTL response that kills a large fraction of infected hepatocytes, with direct loss of replicative intermediates and cccDNA; (2). A concurrent CTL-induced interferon-gamma response that eliminates nucleocapsids containing replicating DNA from surviving hepatocytes; and (3). Since hepatocytes are a largely self-renewing population, proliferation of surviving hepatocytes to restore liver cell mass. In this way hepatocytes that survive give rise to the uninfected hepatocytes that populate the liver at recovery. What has remained unclear is the fate during this “clearance phase” of the 5–50 copies of cccDNA present in the nucleus of infected hepatocytes; in particular, those hepatocytes not killed by anti-viral CTL. On one hand there is evidence to suggest that cccDNA elimination requires the turnover of the infected hepatocytes (modeling studies have shown that if cytokines eliminate nucleocapsids containing replicative DNA, and cccDNA survives mitosis, cccDNA-free hepatocytes will be generated by dilution of cccDNA through 2–3 turnovers of the entire hepatocyte population, while if cccDNA does not survive mitosis, generation of cccDNA-free hepatocytes would require <1 turnover). On the other hand, published evidence also suggests that cytokine-mediated effects might lead to loss of cccDNA; whether this is a significant effect *in vivo* and sufficient to explain cccDNA loss during virus clearance remains unclear.

Although hepatocytes are the main cell type infected by DHBV, infection also occurs in extrahepatic tissues. In ducks infected *in ovo*, with transmission of DHBV from the bloodstream of the female duck to the egg, virus replication first occurs in the endodermal yolk sac, then in the developing liver and pancreas of the embryo (Fig. 5(A–C)). In the liver, hepatocytes and bile duct cells are infected with DHBV and express high levels of DHBcAg and DHBsAg. Virus replication also occurs in scattered acinar cells and in alpha and beta endocrine islet cells in the pancreas, and in glomeruli and tubular epithelial cells in the kidney of developing embryos. Similar patterns of extrahepatic DHBV infection also occur following experimental transmission in ducklings (Fig. 5(C–E)). Viral antigen accumulation in kidney glomeruli many represent bound immune complexes. Cells located in the red pulp and germinal centers of the spleen have been shown to contain DHBV DNA. Mononuclear cells in the red pulp support DHBV replication early after infection while germinal centers contain non-replicating DHBV DNA associated with follicular dendritic cells. Overall, the significance of viral replication in extrahepatic tissues to the pathogenesis of DHBV infection remains unclear.

Diagnosis

In ducks with congenital and persistent DHBV infection, where $>95\%$ of hepatocytes are infected, DHBV virions are released into the bloodstream resulting in titers of up to 1×10^{10} virions/ml. 35–60 nm DHBsAg particles that lack a virus nucleocapsid are also released from infected hepatocytes, in 500–1000-fold excess over virions, resulting in $\sim 5\text{--}50$ ug/ml of DHBsAg. Serum samples can be tested for DHBV DNA by PCR and for DHBsAg using enzyme-linked immunosorbent assay (ELISA). DHBV

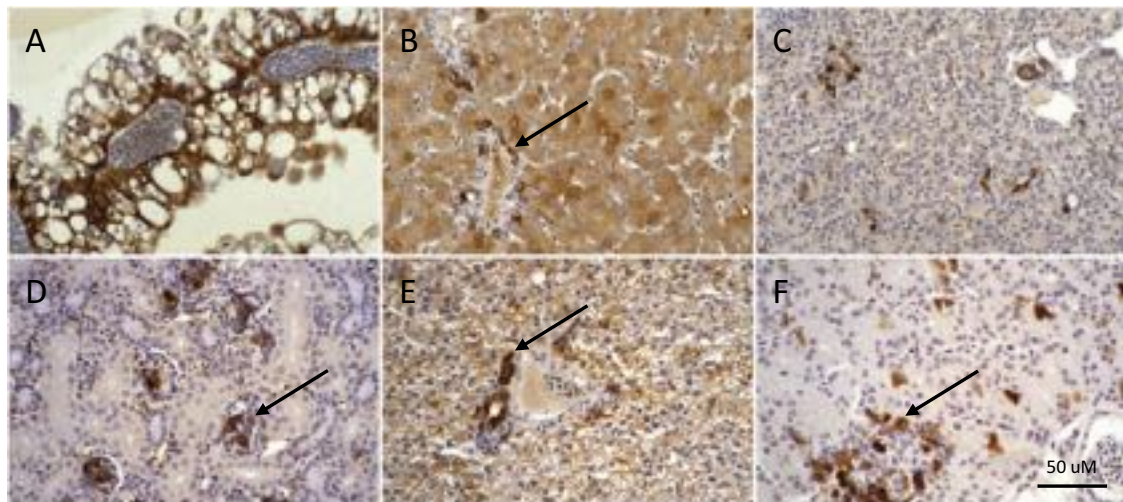


Fig. 5 Hepatic and extrahepatic DHBV infection. Immunoperoxidase staining of DHBcAg in yolk sac (A), liver (B, E), pancreas (C, F) and kidney (D). DHBV is transmitted *in ovo* from the bloodstream of a DHBV-infected female duck to the egg resulting in DHBV infection in (A) yolk sac cells, (B) liver hepatocytes and bile duct cells (arrow) and (C) pancreatic acinar cells in a 14-day fertilized duck embryo. DHBV infection was also detected in (D) kidney glomerular cells (arrow), (E) liver hepatocytes and bile duct cells (arrow) and (F) pancreatic acinar and islets cells (arrow) in a 2-week-old congenitally DHBV-infected duck. Ethanol:acetic acid-fixed tissues were immunostained with polyclonal anti-DHBc antibodies and counterstained with haematoxylin. Magnification $\times 226$. Bar = 50 μ m.

infection can also be detected by immunoperoxidase or immunofluorescence staining of ethanol:acetic acid-fixed and frozen liver tissue using monoclonal or polyclonal anti-DHBs or anti-DHBc antibodies or by using PCR techniques to detect DHBV DNA in extracts from liver and extrahepatic tissues. Detection of cccDNA is performed by selective extraction of protein-free DNA followed by detection by Southern blot hybridization or by a specific PCR assay for cccDNA with primers spanning the incomplete region of the rcDNA genome to reduce detection of rcDNA and replicative intermediates.

DHBsAg and DHBcAg proteins are highly immunogenic. Anti-DHBc antibodies can be detected by ELISA in the serum of ducks during transient and persistent DHBV infections. Anti-DHBs antibodies are produced during the resolution phase of transient infection and are generally not detected in ducks with persistent infection except in complexes with circulating DHBsAg. Anti-DHBs antibodies bind virus particles and are able to block DHBV infection of cells. For this reason, ducks that have recovered from transient infection are immune to challenge with DHBV.

Cellular immune responses that occur during DHBV infections are more difficult to study due to a lack of reagents for studying the immune system in the duck; however, *in vitro* stimulation assays for detection of viral antigen specific T-cell responses have been developed and anti-duck CD4 and CD8 monoclonal antibodies are available for further studies. The recent cloning of cDNA for the duck interferon- α and γ , duck T-cell markers including CD3, CD4, CD8, MHC I and II and duck TLRs now increase the feasibility of studies of the immune response during DHBV infection.

Treatment

Because of its reproducible kinetics and predictable outcomes of infection, the DHBV model has been widely used for the evaluation of anti-viral therapies. Similarities in the replication strategy and structure and activity of the HBV and DHBV polymerase enzymes have allowed anti-viral drugs designed for chronic HBV to be evaluated, either as monotherapy or in combination, in DHBV-infected ducks. A range of nucleoside and nucleotide analogs (NA), that inhibit replication by acting as chain terminators, have been tested including ganciclovir, penciclovir, adefovir, entecavir (ETV), tenofovir disoproxil fumarate (TDF), emtricitabine and clevudine. Treatment of DHBV-infected ducks with NA blocks reverse transcription and inhibits the synthesis of rcDNA and dsDNA in the cytoplasm of infected cells. However, cccDNA, which is present at 5–50 copies in the nucleus of each infected cell at the outset of treatment, is often reduced (presumably by hepatocyte turnover) but not eliminated, and since cccDNA is the template for transcription of viral RNA, viral antigen expression also persists, and virus replication can rebound if treatment is discontinued. A further issue is the high error rate of the viral polymerase, which can introduce mutations into the polymerase RT domain, allowing the emergence of drug-resistant DHBV strains. Some NA (*e.g.*, ETV and TDF) can be used in monotherapy in treatment naïve patients, but long-term anti-viral therapy for chronic HBV infection more commonly uses combinations of different NA to lower the risk of emergence of multidrug-resistant HBV strains. Newer direct-acting anti-viral agents as well as immunomodulators are also in development for chronic HBV infection.

Studies of anti-viral therapy in the DHBV model have demonstrated the stability and long *in vivo* half-life of cccDNA. It is not surprising therefore that cccDNA persists in trace amounts following the resolution of transient HBV infection and that infection

can rebound in situations of immune depletion; for example, during immune suppression following transplantation with donor liver containing residual HBV cccDNA. Studies in the DHBV model have shown that levels of cccDNA present in the liver following the resolution of transient infection are unaffected by treatment with ETV, indicating that residual cccDNA is highly stable and is present in a cell population with a low rate of turnover as compared to typical hepatocytes. The stability and persistence of cccDNA has also been studied in the liver of newly-hatched ducks in which high levels of hepatocyte mitosis occur due to liver growth. Although these results suggested that cccDNA could survive hepatocyte mitosis, there was large duck-to-duck variation in cccDNA levels suggesting that despite therapy with ETV, cccDNA may be maintained to some extent by low levels of viral DNA replication in the growing liver.

DHBV-infected ducks have also provided a model system for anti-viral drug therapy in combination with DNA vaccines and have been used to assess nucleic acid polymers developed by REPLICor Inc (which function as broad spectrum anti-viral agents by inhibiting the release of HBsAg from hepatocytes) to protect against and as therapy for persistent DHBV infection.

The identification and cloning of the cellular receptor used by HBV, hNTCP, and the ability to express the hNTCP in mammalian cells, has conferred susceptibility to HBV infection on a whole range of transformed cell lines and will enable studies of anti-viral drugs, virus replication and cccDNA to be performed *in vitro* in cell lines infected with HBV. These systems are likely to decrease reliance on the DHBV model, but are unlikely to replace the model, which allows studies of the complex interplay between host and virus factors during *in vivo* infection.

Endogenous Avian Hepadnavirus Elements

Recent advances in computational genomics have led to the discovery of endogenous avian hepadnavirus elements in the germline of many bird species, representing most of the major bird orders. Endogenous avian hepadnavirus elements were first identified in the zebra finch (*Taeniopygia guttata*, family Estrildidae) and designated as endogenous zebra finch HBV (eZHBV). eZHBVs have been detected on 10 different chromosomes as site-specific integrated viral fragments that cover ~70% of the avian hepadnavirus genome. eZHBVs have also been found in three other estrildid finches: the black throated finch (*Poephila cincta*), the scaly breasted munia (*Lonchura punctulata*) and the Gouldian finch (*Erythrura Gouldiae*) and in the rough-legged hawk (*Buteo lagopus*) and the black-headed gull (*Chroicocephalus ridibundus*). Endogenous avian hepadnavirus elements have also been found in the dark-eyed junco (*Junco hyemalis*, family Emberizidae) and designated as endogenous Junco HBV (eJHBV). The first full-length endogenous avian hepadnavirus elements were found in the genome of the budgerigar (*Melopsittacus undulatus*, family Psittaculidae) and, along with additional partial genome fragments, were named endogenous budgerigar HBV (eBHBV) (Table 1B). Partial and full-length eBHBVs do not share site-specific orthologous integration points with the identified eZHBVs and, like the individual eZHBVs, represent separate integration events.

All of the eZHBV, eJHBV and eBHBV sequences described to date have lost their AUG start codons or contain stop codons or missense mutations that would prevent them from encoding functional proteins. However, analysis of the putative polymerase AA sequences of eZHBV, eJHBV and eBHBV shows that they are related to, but distinct from the exogenous avian hepadnaviruses, AusDHBV, SGHBV1-7, CCHBV1, RGHV, ETHBV, PHBV-A, STHBV-1 and HHBV-4 (Fig. 1(B)).

Phylogenetic analysis suggests that endogenous avian hepadnavirus elements were created through multiple integrations of rcDNA or dsDNA into germline DNA earlier in evolutionary history when the species, or an ancestor, supported exogenous avian hepadnavirus infection. Integration of hepadnavirus DNA into the host genome occurs at a random site and, once fixed in the host germline, the site-specific integration is passed on to progeny by long-term vertical inheritance and can be used as a genetic marker of ancestry. For example, eZHBVs, which were the first endogenous viral elements to be described from a reverse transcribing DNA virus, revealed a deep timescale of the association between hepadnaviruses and birds. The detection of eZHBV sequences in

Table 1B Endogenous Avian Hepadnavirus Elements

Clone name	Accession ^a	Length (nt) ^b	Host	Species	Location ^c
eZHBV	HQ116569.1	1076	Zebra Finch	<i>Taeniopygia guttata</i>	Fitzroy Crossing, Australia
eZHBV	HQ116568.1	1044	Black Throated Finch	<i>Poephila cincta</i>	Chillagoe, QLD, Australia
eZHBV	HQ116566.1	803	Scaly Breasted Munia	<i>Lonchura punctulata</i>	Pasir Ris, Singapore
eZHBV	HQ116567.1	1034	Gouldian Finch	<i>Erythrura gouldiae</i>	Seattle, USA
eZHBV_C	KC750096.1 ^d	4295	Rough-legged Hawk	<i>Buteo lagopus</i>	Germany
eZHBV_C	KC750099.1 ^e	4291	Black-headed Gull	<i>Chroicocephalus ridibundus</i>	Germany
eBHBV-1	BK008520.1/JQ978775.1	4865/4859	Budgerigar	<i>Melopsittacus undulatus</i>	Shanghai, China
eBHBV-2	BK008521.1/JQ978784.1	3856/3856	Budgerigar	<i>Melopsittacus undulatus</i>	Shanghai, China
eJHBV	HQ116565.1/HQ116576.1	1096/325	Dark-Eyed Junco	<i>Junco hyemalis</i>	Arlington, TX, USA

^aGenBank accession number.

^bLength of GenBank sequence containing partial or full-length endogenous avian hepadnavirus elements + / – fragments of germline DNA.

^cLocation of the bird/collection where the virus is found.

^dShares 93.8% identity with eZHBV_C (HAS) described by Suh *et al.*, 2013.

^eShares 91.8% identity with eZHBV_C (HAS) described by Suh *et al.*, 2013.

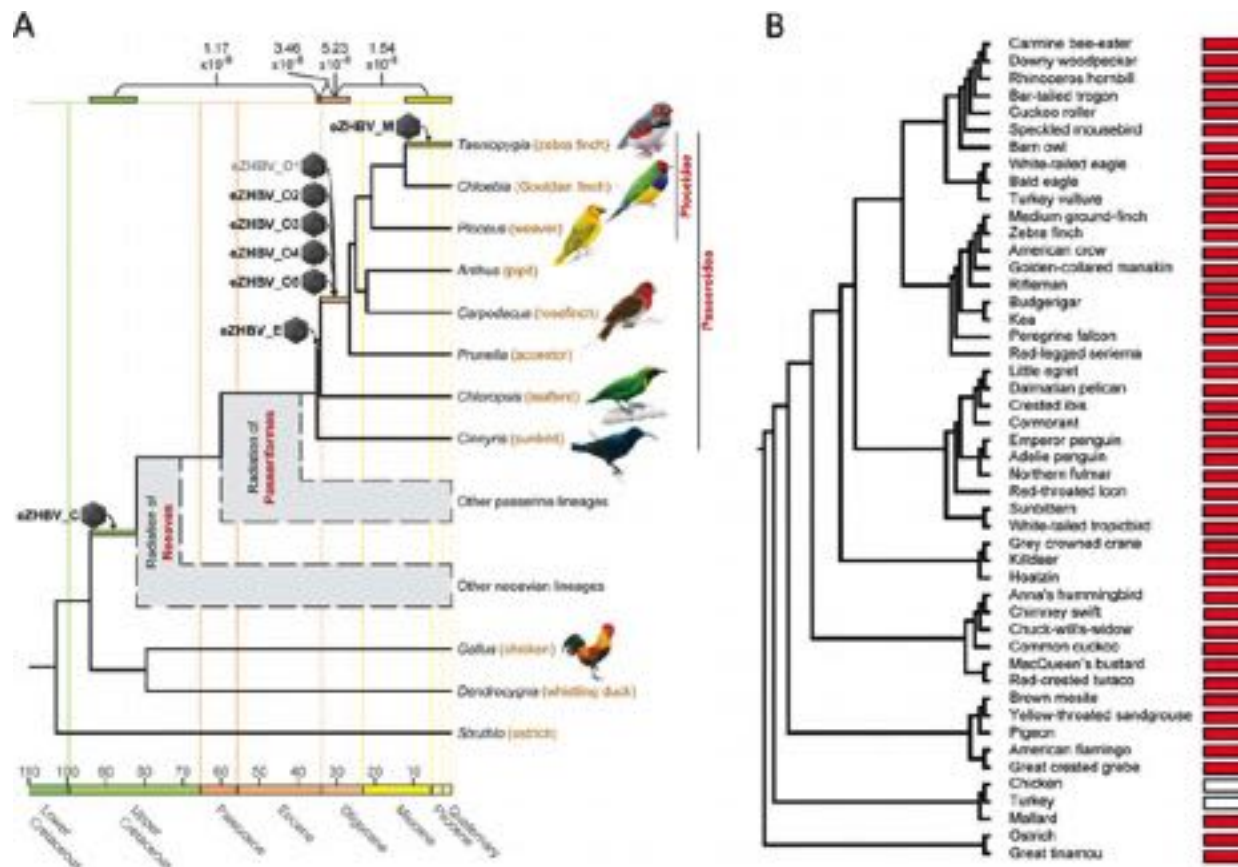


Fig. 6 Evolution of endogenous avian hepadnavirus elements. (A). Endogenous avian hepadnavirus elements reveal a long association between hepadnaviruses and birds. Presence/absence analyses of eZHBV insertion loci revealed eZHBV endogenization events in the ancestor of Neoaves, during early passeroid evolution and during recent ploceid evolution. The youngest eZHBV (eZHBV_M) was found in the zebra finch but not in the Gouldian finch, weaver, pipit, rosefinch, dunnock, leafbird or sunbird, placing its endogenization after the speciation of the zebra finch in the Miocene, 12.1 million years ago (MYA). Five distinct site-specific orthologous eZHBVs integrants (eZHBV_O1-eZHBV_O5) have been identified during the Oligocene (26.6–34.2 MYA) based on their presence in the finch, but not in the leafbird or sunbird, and one (eZHBV_E) during the Eocene (34.2–35 MYA) which was found in finches and the leafbird but not in the sunbird. The oldest eZHBV integrant (eZHBV_C) is believed to have integrated into a zebra finch ancestor between 82 and 94 MYA, during the Cretaceous Period. The avian tree of life is shown as a simplified chronogram based on molecular dates of interordinal avian relationships, as well as relationships within Passeriformes and Ploceidae. Nucleotide substitution rates (substitutions/site/year) based on HAS (hypothetical ancestral sequences) were calculated for the respective parts of eZHBV evolution and denoted above the tree. Endogenization events are depicted as icosahedrons and temporal ranges of insertion events are shown as colored rectangles. Colored scale at the bottom bar represents millions of years. Reproduced from Suh, A., Brosius, J., Schmitz, J., Kriegs, J.O., 2013. The genome of a mesozoic paleovirus reveals the evolution of hepatitis B viruses. *Nature Communications* 4, 1791, Fig. 1, with permission. (B). Distribution of endogenous avian hepadnavirus elements across the avian phylogeny. Chromosome and whole DNA genome shotgun assemblies of 48 avian species were screened in silico using tBLASTn with a library of hepadnavirus protein sequences. 46 out of 48 host species generated high-identity matches to hepadnavirus proteins and are marked in red on the species tree, while 2 DNA samples were negative and are marked in white. The phylogeny is based on the results of a phylogenomics consortium whole DNA genome analyses across all the species shown. Reproduced with modifications from Cui, J., Zhao, W., Huang, Z., et al., 2014. Low frequency of paleoviral infiltration across the avian phylogeny. *Genome Biology* 15 (12), 539, Fig. 1, with permission.

different avian species is illustrated in Fig. 6(A) (reproduced with permission Suh et al., 2013). The youngest eZHBV integrant to be described (eZHBV_M; corresponding to 15% of the avian hepadnavirus genome) is thought to have integrated into the germline genome of a zebra finch ancestor in the Miocene, 12.1 MYA. Five additional integration events have been identified for eZHBVs (eZHBV_O1-eZHBV_O5; 3.3%–9.0% of the avian hepadnavirus genome) during the Oligocene (26.6–34.2 MYA), and one (eZHBV_E; 14.8% of the avian hepadnavirus genome) during the Eocene (34.2–35 MYA). The oldest eZHBV integrant to be dated so far (eZHBV_C; 99% of the avian hepadnavirus genome) is believed to have integrated between 82 and 94 MYA, during the Cretaceous Period. The detection of eZHBV_C in all major taxa of Neoaves (which includes most modern birds) and its absence in non-Neoaves (e.g., the chicken and the ostrich) places the time of integration after the separation of Neoaves and non-Neoaves but before the diversification of Neoaves.

Mining of whole genome sequence data recently made available for 48 bird species (representing thirty-two of thirty-five proposed bird orders and including all Neoavian orders) has revealed the widespread presence of endogenous avian hepadnavirus

elements. As can be seen in **Fig. 6(B)** (adapted and reproduced from Cui *et al.*, 2014), endogenous avian hepadnavirus elements were detected in 46 out of 48 species including the zebra finch and budgerigar as previously discussed, but also detected in species that support exogenous avian hepadnavirus infection including the Grey crowned crane and the mallard, the species from which most domesticated ducks are derived. The diversity of avian species containing endogenous avian hepadnavirus elements suggests that multiple integrations have occurred over the evolutionary history of birds.

Acknowledgments

The authors wish to acknowledge Dr William Mason for critical reading of the manuscript and helpful discussions, and Ms Tahlia Perry and Prof Frank Grutzner for assistance with the phylogenetic analysis.

References

- Cui, J., Zhao, W., Huang, Z., *et al.*, 2014. Low frequency of paleoviral infiltration across the avian phylogeny. *Genome Biology* 15 (12), 539.
- Suh, A., Brosius, J., Schmitz, J., Kriegs, J.O., 2013. The genome of a mesozoic paleovirus reveals the evolution of hepatitis B viruses. *Nature Communications* 4, 1791.
- Cui, J., Holmes, E.C., 2012. Endogenous hepadnaviruses in the genome of the budgerigar (*Melopsittacus undulatus*) and the evolution of avian hepadnaviruses. *Journal of Virology* 86 (14), 7688–7691.
- Foster, W.K., Miller, D.S., Scougall, C.A., *et al.*, 2005. Effect of antiviral treatment with entecavir on age- and dose-related outcomes of duck hepatitis B virus infection. *Journal of Virology* 79 (9), 5819–5832.
- Gilbert, C., Feschotte, C., 2010. Genomic fossils calibrate the long-term evolution of hepadnaviruses. *PLOS Biology* 8 (9).
- Guo, H., Mason, W.S., Aldrich, C.E., *et al.*, 2005. Identification and characterization of avihepadnaviruses isolated from exotic anseriformes maintained in captivity. *Journal of Virology* 79 (5), 2729–2742.
- Jilbert, A.R., Miller, D.S., Scougall, C.A., Turnbull, H., Burrell, C.J., 1996. Kinetics of duck hepatitis B virus infection following low dose virus inoculation: One virus DNA genome is infectious in neonatal ducks. *Virology* 226 (2), 338–345.
- Jilbert, A.R., Wu, T.T., England, J.M., *et al.*, 1992. Rapid resolution of duck hepatitis B virus infections occurs after massive hepatocellular involvement. *Journal of Virology* 66 (3), 1377–1388.
- Lauber, C., Seitz, S., Mattei, S., *et al.*, 2017. Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. *Cell Host Microbe* 22 (3), 387–399.
- Le Mire, M.F., Miller, D.S., Foster, W.K., Burrell, C.J., Jilbert, A.R., 2005. Covalently closed circular DNA is the predominant form of duck hepatitis B virus DNA that persists following transient infection. *Journal of Virology* 79 (19), 12242–12252.
- Mason, W.S., 2015. Animal models and the molecular biology of hepadnavirus infection. *Cold Spring Harbor Perspectives in Medicine* 5 (4).
- Miller, D.S., Bertram, E.M., Scougall, C.A., Kotlarski, I., Jilbert, A.R., 2004. Studying host immune responses against duck hepatitis B virus infection. *Methods in Molecular Medicine* 96, 3–25.
- Piasecki, T., Harkins, G.W., Chrzastek, K., *et al.*, 2013. Avihepadnavirus diversity in parrots is comparable to that found amongst all other avian species. *Virology* 438 (2), 98–105.
- Reaiche-Miller, G.Y., Thorpe, M., Low, H.C., *et al.*, 2013. Duck hepatitis B virus covalently closed circular DNA appears to survive hepatocyte mitosis in the growing liver. *Virology* 446 (1–2), 357–364.
- Summers, J., Jilbert, A.R., Yang, W., *et al.*, 2003. Hepatocyte turnover during resolution of a transient hepadnaviral infection. *Proceedings of the National Academy of Sciences of the United States of America* 100 (20), 11652–11659.

Avian Herpesviruses (*Herpesviridae*)

Vishwanatha RAP Reddy and Venugopal Nair, The Pirbright Institute, Pirbright, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Classification

Avian herpesviruses (AHV) are widespread among most variety of domestic, wild and captive bird species. Similar to mammalian herpesviruses, most AHV have narrow host range in their natural hosts, although some of these viruses infect different bird species belonging to the same or to unrelated families or even orders. Based on the International Committee on Taxonomy of Viruses (ICTV), AHV are classified as alphaherpesviruses, under subfamily *Alphaherpesvirinae*, family *Herpesviridae* and order *Herpesvirales*. These viruses include Gallid herpesvirus type 1 (GaHV1) and Psittacid herpesvirus type 1 (PsHV1) that belong to the genus *Iltovirus*, and Gallid herpesvirus type 2 and 3 (GaHV2 and 3), Meleagrid herpesvirus 1 (MeHV1), Anatid herpesvirus 1 (AnHV1) and Columbidae herpesvirus 1 (CoHV1) that belong to the genus *Mardivirus*.

Several taxonomically unassigned AHV were also identified in a wide variety of birds, such as falcons (Falconid herpesvirus 1, FaHV1), cranes (Gruid herpesvirus 1, GrHV1), cormorant (Phalacrocoracid herpesvirus 1, PhHV1), eagles (Accipitrid herpesvirus 1, AchHV1), storks (Ciconiidae herpesvirus 1, CiHV1), owls (Strigid herpesvirus 1, StHV1), quails (Perdidae herpesvirus 1, PdHV1), penguins (Spheniscidae herpesvirus 1, SpHV1) and loons (Gaviidae herpesvirus 1, GavHV1) (Fig. 1). Additional AHV infections have also been reported in several wild- and free-range bird species such as the prairie falcon, red-headed falcon, little pied cormorant, demoiselle cranes, Sudan crowned cranes, paradise cranes, Japanese cranes, hooded cranes, black storks, American white storks, black-footed penguin, satyr tragopan, toucans, and many passerine birds.

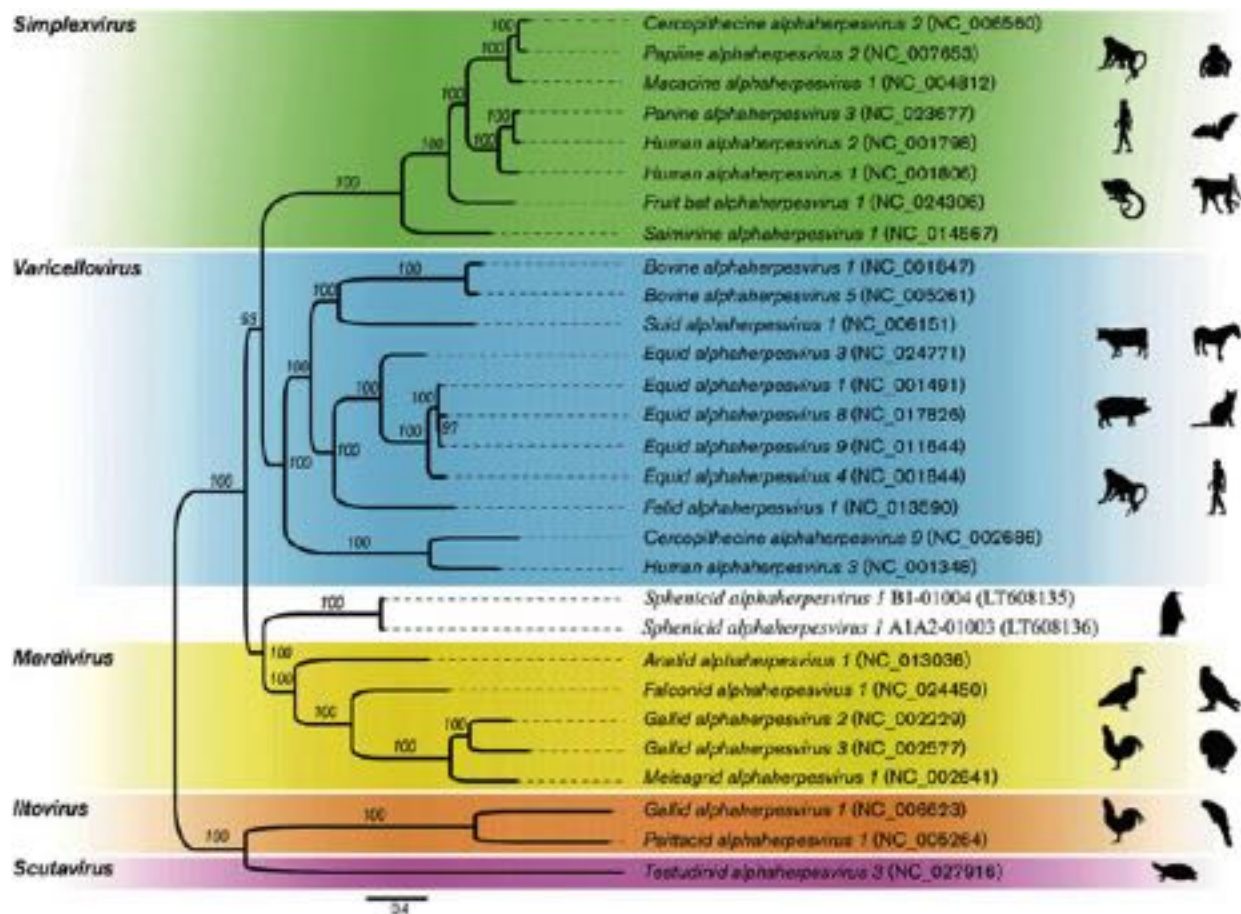


Fig. 1. Genetic relationship of avian herpesviruses with other alphaherpesviruses based on 44 592 amino acid sequences of 52 core genes of the viruses is shown. Gallid herpesvirus 1 (GaHV1) and Psittacid herpesvirus 1 (PsHV1) are closely clustered under *Iltovirus* genus. Gallid herpesvirus 2 and 3 (GaHV2 and 3), Meleagrid herpesvirus 1 (MeHV1), Anatid herpesvirus 1 (AnHV1) and (Falconid herpesvirus 1, FaHV1) are closely grouped under *Mardivirus* genus. Above figure adapted from Pfaff, F., Schulze, C., König, P., et al., 2017. A novel alphaherpesvirus associated with fatal diseases in banded Penguins. *Journal of General Virology* 98, 89–95. doi:10.1099/jgv.0.000698.

Most AHV of wild or free-range birds have been named according to the species of the avian hosts. However, a number of others in wild- and free-range birds still remain as nameless orphan AHV. In some cases, AHV are named after the pathological characteristics associated with the diseases or in honor of scientists who first reported it. Examples include infectious laryngotracheitis virus (ILTV or GaHV1), duck enteritis virus (DEV or AnHV1), hepatosplenitis infectiosa strigum virus (HSIS or StHV1), inclusion body disease virus (FaHV1, GrHV1 and AchV1), herpesvirus of turkeys (HVT or MeHV1), Pacheco's parrot disease virus (PPDV or PsHV1), Marek's disease virus (MDV or GaHV2), and Smadel's disease virus (CoHV1). This article mainly will focus on ILTV, PPDV, MDV and DEV. For information on other AHV, please refer to further reading details.

ILTV is a highly contagious respiratory virus of chickens that is transmitted by inhalation, direct bird-to-bird contact and indirectly through fomites and people. ILTV actively replicates in the epithelium of larynx, trachea and conjunctiva, leading to severe epithelial damage, hemorrhages and laryngotracheitis (hence the term 'infectious laryngotracheitis'). After an acute phase of infection, characterized by clinical signs such as nasal discharge, conjunctivitis and drop in egg production, ILTV establishes a lifelong latency in the trigeminal ganglion. During severe forms of the disease, additional signs such as gasping, expectoration of bloody mucus, dyspnea and death due to asphyxia are also observed. Major economic impact of ILT comes from production losses, high morbidity and mortality and from the costs of vaccination. Live attenuated and recombinant viral vectored vaccines are available to protect against ILT. Most outbreaks are associated with reversions of the live vaccines strains to a virulent strains.

Pacheco's parrot disease virus (PPDV) or Psittacine herpesvirus 1 (PsHV1) is a highly contagious, and potentially fatal respiratory virus of psittacine birds, characterized by acute necrotizing lesions of the crop, intestines, liver and pancreas. Pacheco's parrot disease (PPD) has been reported in psittacine birds in most geographical locations of the world, and is a major concern in the captive parrots of exotic collections. Amazon parrots, cockatiels, Roselle, Galah and macaws are some of the worldwide affected bird species of the PPDV. PPDV outbreaks commonly occur in the quarantine stations and markets, where parrots are held in close proximity. Thus, worldwide exotic pet breeders and companion bird markets are of great concern about the PPD. Other types of psittacine herpesviruses distinct from PsHV1 such as psittacid herpesvirus 2 and 3 have been identified, serotyped and phylogenetically characterized.

Marek's disease virus (MDV-1) or Gallid herpesvirus 2 (GaHV2) is a contagious oncogenic virus, causes neoplastic and neuropathic disease in chickens, commonly known as Marek's disease (MD). Only, MDV-1 causes clinical oncogenic disease in chickens. Gallid herpesvirus 3 (GaHV3) and herpesvirus of turkeys [HVT or Meleagrid herpesvirus 1 (MeHV1)] are non-oncogenic apathogenic herpesviruses of chickens and turkeys, respectively. GaHV3 and HVT are usually used as the live vaccines against pathogenic MD.

Duck enteritis virus (DEV) or Anatid herpesvirus 1 (AnHV1) is a highly contagious virus of waterfowl species such as ducks, geese and swans, which causes acute disease called duck viral enteritis (DVE), sometimes also referred to as duck plague. Waterfowls of all ages are susceptible to DEV and the disease is characterized by vascular damage, tissue hemorrhages, lymphoid organ lesions, digestive mucosal eruptions, degenerative changes in parenchymatous organs and death. DVE causes significant economic losses in domestic and wild water fowls due to mortality, reduced egg production and meat condemnations. Total morbidity and mortality may range from 5% to 100%.

Host Range

Many AHV appear to be group-specific in a given bird species and exhibit a very narrow host range. Chickens are the primary natural hosts of ILTV, although pheasants, pheasant-chicken crosses and peafowls are also infected with this virus. ILTV is commonly isolated and propagated in the embryonated chicken eggs, where it induces opaque plaques on the chorioallantoic membrane (CAM) resulting from necrosis and proliferation of the infected cells. Plaques on CAM can appear as early as two days post-inoculation (PI) and embryo deaths occur between 2 and 8 days PI. ILTV can also be propagated in a variety of primary avian cell cultures such as chicken embryo kidney (CEK), chicken embryo liver (CEL), chicken embryo lung and chicken kidney cells (CKC), as well as in the LMH (leghorn male hepatoma) cell line. PsHV1 grows in primary chicken embryo fibroblasts (CEF) cultures where it readily shows syncytial plaques. MDV-1 is routinely propagated in CEF, embryonic CKC, and/or duck embryo fibroblasts (DEF). Several lymphoblastoid cell lines that grow continuously as suspension cultures have been developed from MD lymphomas. These cell lines are routinely used to understand the pathobiology of MD lymphomas.

Virion Structure

Broadly AHV have hexagonal nucleocapsids with icosahedral symmetry, with typical morphological features of herpesvirus particles. The 80–125 nm in diameter icosahedral nucleocapsids is made of elongated hollow 162 capsomers-150 hexamers and 12 pentamers. Overall diameter of the enveloped virus can vary. ILTV with irregular envelope has a diameter of 195–250 nm, while the enveloped MDV particles are around 150–160 nm. In general, the morphologies of members of the *Mardivirus* are similar, although HVT capsids appear to show a unique crossed appearance.

Genome

AHV isolates have linear double stranded DNA genomes of approximately 150–200 kilobase pairs (kbp). Based on the genome organization, herpesviruses are classified in to six different groups A to F including class D (ILTV and PsHV1) or E (MDV-1, GaHV3

and HVT) genomes. Class D herpesvirus genomes consist of long and short unique regions (U_L and U_S) interspersed with inverted repeat sequences, such as internal repeat (IR) and terminal repeat (TR) sequences. Class E herpesvirus genomes, consist of unique long (U_L) and unique short sequence (U_S), flanked by a sets of inverted repeat sequences: the terminal repeat long (TR_L), internal repeat long (IR_L), internal repeat short (IR_S), and terminal repeat short (TR_S), respectively.

The 153-kbp ILTV genome, predicted to have 79 open reading frames (ORFs), has a 47.9% G+C content, while the 163-kbp PPDV genome, with 73 predicted ORFs, has a 60.9% G+C content. The majority of ILTV and PPDV ORFs have homologous ORFs in other alphaherpesviruses, such as HSV1. However, many unique ORFs also do exist in the ILTV and PPDV genomes. For example, ORFs A, B, C, D, E, F, U_L0 and U_L1 are largely specific of ILTV and PPDV. A large internal inversion of U_L22 to U_L44 genes within the U_L segment has been also reported to be specific to ILTV and PPDV genomes. U_L47 , a conserved homolog among alphaherpesviruses, is absent at the corresponding positions of ILTV and PPDV genomes, but a homolog with weak identity (18%) is found to be inserted between the U_S3 and U_S4 genes of U_S region. ILTV and PPDV genomes also have differences in gene organization, for example, a U_L16 homolog that is conserved throughout all herpesviruses, is absent in ILTV but present in PPDV. In contrast, ILTV contains U_L48 , U_L0 ORFs and U_S9 , neither of which are present in PPDV. Further, the ILTV inverted and terminal repeats are 2535 bp (18.5%) shorter than the PPDV, and each repeat consists of three genes: ICP4, U_S10 and a homolog of the MDV sORF4/3. Conversely, only ICP4 is present in inverted and terminal repeats of PPDV, U_S10 and sORF4/3 are present in U_S region.

The full genome sequences of MDV, GaHV3 and HVT are very similar of approximately 160–180 kb with a varying G+C content of ~43.9%, ~53.6% and ~47.6%, respectively. A total of 103 ORFs in MDV genome, 102 in GaHV3 and 99 in HVT were clearly identified, in addition to a number of microRNAs. The majority of U_L and U_S genes of *Maridiviruses* are homologous to the genes encoded by the alphaherpesviruses. Many virus-specific genes have been identified, and are located predominantly in the IR_L and TR_L regions. Meq (MDV EcoRI Q), vIL8 (viral interleukin 8), vLIP (viral lipase), pp38/pp24 and vTR (RNA telomerase subunit) genes have been identified that are unique for MDV. The Meq gene encodes a 339-amino acid oncoprotein, which contains a basic leucine zipper (bZIP) transcriptional regulator domain at the N terminal and the proline rich repeat transactivator domain at the C terminal. The Meq protein is expressed consistently in the nucleus of lymphoma cells and tumor cell lines. The vIL8 gene is located in the long repeat region, and originally was identified as a spliced Meq variant. The vIL8 functions as a chemoattractant for chicken mononuclear cells, and may be important for the switch of infection from B to T lymphocytes. The vLIP gene, required for efficient lytic replication, encodes a soluble N glycosylated protein similar to α/β beta fold of pancreatic lipases. MDV-1 genome encodes the pp38/pp24 complex within the TR_L/IR_L and the U_L region. MDV-encoded vTR gene, with nearly 88% sequence identity to the chicken telomerase RNA (cTR) indicating its transduction from the host genome, is present in the IR_L/TR_L region of the MDV genome. HVT encodes a two copies of vNr13 protein, a Bcl2 homolog, not seen in the other alphaherpesviruses. The vNr13 is strikingly similar to chicken Nr13 protein (63.7% amino acid identity), an anti-apoptotic member of the Bcl2 like proteins. Most DEV strains have approximately of 160-kbp long genome with a G+C content of 44.89%. With around 78 ORFs, many are homologs seen in other alphaherpesviruses. Five DEV genes were unique among the AHV, while three appear to be specific to DEV.

Replication Cycle

The replication cycle of ILTV is similar to other alphaherpesviruses, which usually starts with the initial attachment to cell surface, which is enhanced by viral glycoproteins gB, gD and gH-L complex. After initial attachment, the nucleocapsid is released into the cytoplasm and transported to the nucleus. Upon injection of the viral genome in the nucleus via the nucleopore, viral DNA replication and transcription occur. In general, the alphaherpesvirus genome transcription strictly follows a regulated cascade at different time points grouped into immediate early (IE), early (E) and late (L) genes. Commonly, IE and E genes, and their products regulate transcription of L genes. First several enzymes and DNA binding proteins are expressed, which are required for DNA replication, e.g., DNA polymerase and thymidine kinase. This is followed by the expression of viral structural proteins such as the capsid, tegument and envelope glycoproteins. The newly synthesized DNA concatemers cleaves into monomeric units and assemble with capsids to form nucleocapsids in the nucleus. The assembled nucleocapsids exploit an envelopment/de-envelopment pathway for transport to the cytosol. The nucleocapsids acquire an envelope by budding through the inner leaflet of the nuclear membrane. The primary enveloped nucleocapsids, undergo a fusion with the outer nuclear membrane, leading to the release of de-enveloped/naked nucleocapsids into the cytoplasm. In the cytoplasm, main tegument proteins are added to the nucleocapsids, before obtaining their final envelope by budding into the Golgi-derived vesicles. Enveloped mature virions released into the extracellular space by exocytosis/cell lysis and membrane vacuolar fusion.

The replication cycle of *Maridiviruses* are similar to other cell-associated herpesviruses such as varicella zoster virus (VZV). The precise virus entry, attachment and replication pathways of MDV remains still unsolved because of strict cell-associated nature of the virus, and only few steps of the replication cycle are available. US3-kinase has been reported to be involved in the morphogenesis as well as cell-to-cell spread of virus through the effect on stress fiber breakdown and polymerization of actin. The glycoproteins gE, gI and gM play a role in the transfer of virus from infected to uninfected cells.

Epidemiology

Majority of AHV are highly contagious especially in intensive production system where the viruses are transmitted very efficiently between birds. Many AHV appear to be group-specific in a given bird species and exhibit a very narrow host range, suggesting their co-

evolution over millions of years. Chickens are the primary natural hosts of ILTV, although pheasants, pheasant-chicken crosses and peafowls are also infected with this virus. Phylogenetic analysis have suggested that AHV species such as ILTV and PsHV-1 have separated from mammalian alphaherpesviruses around 200 million years ago, and have been co-evolving since then with their specific hosts. PsHV1-infected birds shed the virus from the oral or respiratory secretions and droppings. As in the case of ILTV infection, PsHV1 is transmitted by direct bird to bird contact or indirectly by contact with contaminated fomites or environment. Latent virus persists in the flock over generations, and act as source of new infection following reactivation during stress with factors such as the onset of egg laying and mixing of chickens. ILTV has been also reported in extra respiratory tissues like liver, spleen, kidneys, heart, intestines, bursa Fabricius and Harderian glands. Epidemiology of ILTV is also affected by their ability for potential recombination. Water appears to be the natural means of DEV transmission among infected water fowls. DEV establishes latency in the TG and lymphoid tissues including peripheral blood lymphocytes, and recovered waterfowls become carriers and shed the virus intermittently.

Viruses such as MDV mainly infects chickens, although quails, turkeys and pheasants are also susceptible to the disease. Because of their long-term survival in the infected environment in the poultry house dust, coupled with continuous shedding from dander and dust in the latently-infected hosts, diseases such as MD are difficult to be eradicated from the infected flocks. While vaccination still continues to be the most effective approach for control of MD, the leakiness of the current vaccines due to their inability to prevent virus shedding has been thought to be one factor driving virulence of the MDV strains.

Clinical Features

Clinical features of AHV infections vary between viruses and tissues targeted by different viruses. ILTV-associated diseases range from mild to severe forms. The mild form of the disease is characterized by nasal discharge, mild conjunctivitis, mild respiratory rales, reduced egg production and no mortality, while the severe form is characterized by open mouth breathing with expectoration of bloody mucoid material, severe conjunctivitis, high morbidity, and moderate to severe mortality. The clinical signs are reported in chickens of all ages but the most characteristic are observed in adult birds. The important gross pathological changes of ILT include hemorrhages and diphtheritic changes in larynx and trachea, with mucoid plugs/casts in the trachea and presence of eosinophilic inclusion bodies in mucosal cells.

Clinical signs of PPD include anorexia, depression, nasal discharge, diarrhea, tremors and instability. The majority of infected birds die very rapidly before the onset of clinical signs. Gross pathological lesions include abnormal changes in the respiratory, muscular and circulatory system, bone marrow, thyroid and adrenal glands, liver, spleen, kidneys, the urogenital and gastrointestinal tract, and the nervous system. Histopathological changes include necrotizing lesions in many organs, hemorrhages and congestion of the liver, spleen and kidneys, and the presence of intranuclear inclusion bodies.

Clinical manifestations of MD vary according to the specific syndromes. In general, the classical form of the disease is characterized by unilateral paralysis of the legs and/or wings. In acute forms of disease with visceral lymphomas, affected birds may appear clinically normal but have extensive neoplastic involvement visible on post-mortem examination. Non-specific signs such as weight loss, paleness, anorexia, and diarrhea may also be seen in the chickens affected with lymphomas. The incidence of MD is quite variable in commercial flocks and in general are low since the introduction of vaccination. Severely affected peripheral nerves may show loss of cross-striations, yellow discoloration and edematous appearance. Lymphomas may occur in a variety of visceral organs. Lymphomas can be found in the lung, spleen, bursa, thymus, heart, adrenal gland, kidney, liver, mesentery, proventriculus, intestine, iris, skeletal muscle, skin and gonads. Microscopically two main types of lymphoproliferative lesions are recognized in peripheral nerves. Type A lesions are considered as neoplastic mainly consisting of CD30⁺ CD4⁺ T cells. Demyelination and Schwann cell proliferation of peripheral nerves are also observed in type A lesions. Type B lesions are mostly inflammatory, characterized by diffuse, light to moderate infiltration of lymphocytes and plasma cells. MD lymphomatous lesions in MD are more proliferative than those in nerves. Appearance is similar to type A lesions, mainly consist of T lymphocytes and lymphoblasts, B cells, macrophages and natural killer cells (NK cells). Majority of transformed T cells express CD4 and MHC II.

In mature breeder duck flocks acutely infected with DEV, sudden, high and persistent mortality with drop in egg production have been reported. As infection progresses, other signs such as loss of appetite, ataxia, photophobia, associated pasted eye-lids, droopiness, ruffled feathers, nasal discharge and watery diarrhea are observed. Characteristic lesions of DVE are intravascular coagulopathy and necrotic degenerative changes in mucosa and submucosa of gastrointestinal tract in lymphoid and parenchymatous organs. Petechial or larger extravasations of blood may be found or in the myocardium and other visceral organs. Liver, pancreas, intestine, lungs, and kidney surfaces may be covered with petechial hemorrhages. Mucosal lesions are observed in the oral cavity, esophagus, ceca and cloaca. Microscopically, initial lesions occur in the walls of blood vessels. The endothelial lining is disrupted and connective tissue of the wall becomes less compact. Eosinophilic intranuclear and cytoplasmic inclusions are observed in epithelial cells of digestive tract.

Pathogenesis

Pathogenesis of AHV can differ based on their tissue tropism and diseases. ILTV enters its host through upper respiratory and ocular routes, and then primarily replicates in the epithelium of larynx, tracheal and conjunctival mucosae, respiratory sinuses, air sacs and lungs. During the first 7 days of infection, ILTV replication in the larynx and trachea, leads to epithelial damage and hemorrhages. Several authors have reported that active replication of ILTV slows down after 6 days post inoculation (dpi), and virus may remain at very low levels up to 10 dpi. ILTV infection spreads to trigeminal ganglion (TG) in 4–7 days, where the virus establishes latency.

Transmission of MDV occurs through the inhalation of infectious dander via the respiratory route, where the virus replication is thought to be initiated in macrophages. After 24 h of infection, cytolitic replication takes place in the lymphoid organs such as spleen, bursa and thymus, where productive cytolitic infection is initiated in the B cells, followed by the T cells. After the initial cytolitic infection associated with immunosuppression, MDV enters a latency phase in T-cells with potential integration in the telomeric regions of the host chromosomes. Finally, some of the latently-infected CD4⁺ T cells are transformed leading to the development of T-cell lymphomas in multiple organs. MDV cytolitic replication continues in the feather follicle epithelium (FFE), the only known site from where cell-free infectious virus is assembled and released into the environment.

Diagnosis

Monitoring and surveillance of AHV infections are important to recognize the distribution and occurrence conditions in the flocks. Diagnosis of AHV infections in poultry flocks are made using pathological, virological and molecular methods. Pathological diagnosis using gross pathological lesions or histopathological changes are often sufficient if the clinical symptoms and lesions are typical of the diseases associated with the AHV infections. Virological methods of diagnosis include virus isolation using cell culture, chicken embryos and animal models, as well as by the detection of viral antigens in the suspected samples. While virus isolation may not be suitable for all AHV isolates, especially as some of the virus isolates may not easily replicate in the culture systems as further adaptation is required. Serological detection of viral antigen-specific antibodies is also valuable for diagnosis of AHV infections, particularly for detection of infections at the flock levels. PCR-based molecular methods of detection of AHV-specific nucleic acids are increasingly used in the diagnosis due to their inherent speed, sensitivity, reproducibility, and because of the potential to be used on clinical samples. Real-time PCR capable of multiplexing for simultaneous detection of multiple AHV has been developed, which give quantitative data on the copy numbers of AHV isolates in clinical samples. While all of these diagnostic tools are valuable in detecting AHV in the samples, it has to be borne in mind that simple detection of viruses, nucleic acids or proteins alone in the samples does not conclusively show that the clinical disease is indeed caused by these pathogens, especially AHV species such as MDV are ubiquitous present even in clinically normal birds.

Prevention

Control strategies for diseases caused by different AHV can vary based on epidemiological features and disease characteristics although some of the basic principles are common. ILTV is controlled by coordinated efforts of rapid diagnosis, vaccination protocol, and biosecurity measures to prevent further spread. Currently, ILT control is achieved through the use of commercially available live attenuated or genetically-engineered recombinant herpesvirus of turkeys (HVT) or Fowl pox virus (FPV) vectored vaccines. Live attenuated vaccines such as the chicken embryo origin (CEO) vaccines are effective in inducing strong immunity but possess considerable residual virulence, and can cause disease, particularly when birds are stressed. Such viruses can then spread by bird-to-bird passages to initiate new outbreaks as well as causing emergence of virulent viruses. Thus, the use of attenuated vaccines has been restricted to ILTV-enzootic areas. Genetically engineered recombinant vaccines are effective to limit signs of disease, but have only limited effect on virus shedding. Thus further innovations in the design and development of new vaccines are required for more effective control of future ILTV outbreaks.

Inactivated vaccines are available against PsHV1, and are developed by using common inactivating agents. There are three main serotypes of the PsHV1 and infection of one serotype does not protect against with another serotype, suggesting the potential value of autogenous vaccines. However, development of autogenous vaccines requires time for virus isolation and vaccine production. Strict quarantine measures along with diagnostic approaches are also critical, especially for exotic pet bird breeders and companion bird keepers.

For MD, vaccination is the principal strategy for the prevention and control. MD vaccines are administered to 18-day old embryos *in ovo* or 1-day old chicks at hatch. Currently used MD vaccines include the GaHV2 strain CVI988 (Rispen), and naturally apathogenic GaHV3 strain SB-1 and HVT strain Fc126. HVT-based vaccines are also extensively used as recombinant viral vectors for protecting against multiple avian diseases. Recent studies have suggested that MD vaccines such as HVT have the potential to drive virulence because of their limited effects on MDV replication and shedding. Inactivated and attenuated vaccines available against DEV are effective in control especially when combined with quarantine measures during import and export of domestic and wild waterfowl species.

Further Reading

- Garcia, M., Spatz, S., Guy, J.S., 2013. Laryngotracheitis. In: Swayne, D.E., Glison, J.R., McDougald, L.R., *et al.* (Eds.), *Diseases of Poultry*, thirteen ed. pp. 161–179.
- Kalet, E.F., Docherty, D.E., 2007. Avian herpesviruses. In: Thomas, N.J., Hunter, D.B., Atkinson, C.T., *et al.* (Eds.), *Infectious Disease of Wild Birds*. Ames: Blackwell Publishing, pp. 63–86.
- Metwally, S.A., 2013. Duck virus enteritis (duck plague). In: Swayne, D.E., Glison, J.R., McDougald, L.R., *et al.* (Eds.), *Diseases of Poultry*, thirteen ed. pp. 431–440.
- Schat, K.A., Nair, V., 2013. Marek's disease. In: Swayne, D.E., Glison, J.R., McDougald, L.R., *et al.* (Eds.), *Diseases of Poultry*, thirteen ed. pp. 161–179.
- Thureen, D.R., Keeler Jr., C.L., 2006. Psittacid herpesvirus 1 and infectious laryngotracheitis virus: Comparative genome sequence analysis of two avian alphaherpesviruses. *Journal of Virology* 80, 7863–7872.

Avian Influenza Viruses (*Orthomyxoviridae*)

Nicolas Bravo-Vasquez and Stacey Schultz-Cherry, St. Jude Children's Research Hospital, Memphis, TN, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

"DIVA" (differentiating infected from vaccinated animals) Method used to differentiate influenza virus vaccinated from infected animals.

Genetic clade A taxonomic group that comprises a single common ancestor and all descendants of that ancestor.

Highly pathogenic avian influenza virus (HPAI) Designation given to select H5 and H7 avian influenza viruses that kill chickens.

Intravenous pathogenicity index (IVPI) Means of testing the level of pathogenicity of an influenza virus isolate by

observing clinical signs in infected birds over a ten-day period. Strains are considered highly pathogenic if they cause more than 75% mortality within 8 days.

Low pathogenic avian influenza virus (LPAI) Designation given to avian influenza viruses that do not kill chickens.

Reservoir A reservoir host or carrier that harbors pathogenic organisms, without injury to itself and serves as a source from which other individuals can be infected.

Spillover Term describing the spread of influenza virus into another host. Can refer to spread from wild birds into poultry.

Introduction

Avian influenza viruses (AIV) are an ongoing risk to domestic, pet, exotic (zoo) and wild birds worldwide. Disease outbreaks can threaten biodiversity by causing high morbidity and high mortality in endangered species and can also have a significant impact on a country's economy by halting trade among countries, forcing market closures and halting exportations of avian products and sub products from affected countries. While members of the class *Aves* are most susceptible to disease, a broad variety of animal species including humans can be infected with AIV leading to significant public health concerns. Indeed, at least three human influenza pandemics of the 20th century (1918, 1957 and 1968) were caused by avian-human reassorted viruses. More recently, AIV subtypes H5, H7, H9, H10, and H6 have been associated with zoonotic infections in humans resulting in a call for increased global AIV surveillance activities by the World Health Organization (WHO), Food and Agriculture Organization of the United Nations (FAO) and the World Organization for Animal Health (OIE).

Taxonomy

AIV belongs to the *Orthomyxoviridae* family, which is comprised of the following genera: *Alphainfluenzavirus* (including influenza A virus), *Betainfluenzavirus* (influenza B virus), *Gammainfluenzavirus* (influenza C virus), *Deltainfluenza virus* (influenza D virus in cattle and swine), *Thogotovirus* (a tick-borne virus), *Isavirus* and *Quarjanvirus*. As mentioned above, AIV are solely found in the *Alphainfluenzavirus*. Like all influenza viruses, they can be further classified into subtypes based on the genetic and antigenic properties of their surface glycoproteins, HA and neuraminidase (NA). Combinations of all 16 HA subtypes (H1-H16) and 9 subtypes of NA (N1-N9) have been isolated from birds. The endemic circulation of H5 viruses in many parts of the world has led to further subgrouping of the HA into unique genetic clades. A clade is a taxonomic group that comprises a single common ancestor and all descendants of that ancestor. There are geographical differences in the circulation of some HPAI H5 clades.

Virion Structure and Genome

AIV virion is pleomorphic with a size that may vary between 80 and 120 nm and may be either spherical or filamentous. The genome is composed of eight single strand negative-sense RNA segments that code for over 10 proteins including the nucleoprotein (NP), nuclear export protein (NEP) and nonstructural protein (NS), a polymerase heterotrimeric complex (polymerase basic protein 1 (PB1); polymerase basic protein 2 (PB2) and polymerase acid protein (PA)), HA, NA and matrix protein 2 (M2, ion transport channel), and the matrix protein 1 (M1) that maintains the virion shape. The virus is surrounded by a lipid envelope derived from the host plasma membrane. The trimeric HA protein has receptor (cell surface expressed sialic acid) binding activity mediating cell entry. It is the predominant surface glycoprotein being up to 80% of the total proteins present in the viral envelope. The second most predominant surface glycoprotein, the tetrameric NA, has neuraminidase activity that cleaves sialic acid molecules allowing viral egress. HA and NA must work in concert to support productive viral replication. More recently, additional accessory proteins have been identified including PB1-F2, PB2-F2, PB1-N40, PA-X and PA-N155 and PA-N182 that play crucial roles in the viral life cycle.

Life Cycle

Avian influenza viruses follow the same replication cycle as mammalian influenza viruses; attachment, entry, fusion, vRNA transport into the nucleus, transcription, replication, translation, assembly and export. If the life cycle is the same, why don't avian viruses replicate in mammalian cells and vice versa? There are several species-specific restriction factors throughout the replication cycle that block this from occurring. First, HA on the virion attaches to host cell surface expressed sialic acids (SA). The avian influenza viruses have tropism for cellular receptors with terminal $\alpha 2,3$ sialic acid (SA $\alpha 2,3$ Gal), which can be found throughout the respiratory and intestinal tract of birds and the respiratory tracts of mammals including swine, horses and humans (deep lung). Mammalian viruses bind preferentially to $\alpha 2-6$ linked SA.

Second, after attachment, AIV virions enter the cells via receptor-mediated endocytosis where the pH change within the endosome induces a conformation change in HA exposing the HA2 protein resulting in membrane fusion. While the precise pH required for the HA conformational change differs by viral strain and host species, overall HA stability has been linked to pandemic potential and ability to cross-species.

Third, after fusion, the viral ribonucleoproteins (vRNPs) are released into the cytoplasm where they co-opt host factors for transportation into the nucleus for replication initiation. Genome replication represents one of the major host range barriers between avian and mammalian influenza viruses. There are several species-specific host restriction factors that block AIV replication in mammalian cells and vice versa. These include cytoplasmic pattern recognition factors and interferon-stimulated genes (MxA, IFITM3, DDX21 as examples) and specific importin- α nuclear import proteins. Once the vRNPs have entered the nucleus, the genome is transcribed followed by replication. The AIV polymerases perform genome replication poorly in mammalian cells. This is likely due to a combination of species-specific host factors like ANP32A and DDX17 in addition to specific mutations in all three viral polymerase genes as well as in NP, NEP with PB2 E627K being the best understood. Difference in temperature at which polymerase enzymatic activity is regulated also contributes to species-specific blocks. Finally, host responses that limit virus replication may differ for avian and influenza viruses including evasion of autophagy, disruption of MAVS signaling and regulating interferon induction. Understanding the cellular and host mechanisms underlying cross-species infections will increase our ability to predicate viruses with pandemic potential.

AIV are typically propagated in the allantoic cavity of embryonated chicken eggs or in tissue culture cells including Madin-Darby canine kidney (MDCK) cells or primary chicken embryo fibroblasts (CEF) but embryonated duck and turkey eggs can also be used. While embryonated chicken eggs are the preferred option since they can support the growth of a broad range of AIV strains, many field isolates may not rapidly grow, can be contaminated with Newcastle disease virus, and HPAI AIV may be difficult to propagate to high titers in embryonated eggs given the rapid death of the embryo.

Epidemiology

AIV is highly contagious. It is spread by respiratory aerosols, bodily fluids (feces) and feathers from infected animals, contaminated fomites and environment including water. Birds play a fundamental role in the epidemiology of influenza virus. Wild bird, especially wild waterfowl (including Anseriformes) and shorebirds (including Charadriiformes), are considered the natural reservoirs for influenza virus. More than 80 different HA and NA combinations have been identified in wild waterfowl to date; much greater diversity than any other host species. These birds have the potential to spread virus between countries and even continents during migration. AIV spillover events from wild-birds to poultry can result in large outbreaks if not properly contained.

According to the FAO database, ~18,924 confirmed AIV outbreaks were reported during the current decade. The majority occurring in Asia (63%), followed by Africa (19%), Europe (15%) and the Americas (3%) (Fig. 1), primarily due to the endemic spread of HPAI H5 viruses. Since 2004, more than 18,800 HPAI H5 outbreaks have been reported in 61 countries. According to the sanitary code for terrestrial animals of the OIE, H5 and H7 AIV, regardless of pathogenicity, requires immediate notification and control measures.

In many areas of the world, live bird markets (LBM) are commonplace and part of the culture. These are also main risk factors for the spread of AIV within the market, to other areas through the movement of poultry, and for human infections. Most of the birds coming into the market are maintained in subsistence-oriented small-scale production systems commonly known as backyard production systems (BPS). BPS systems are the most common form of animal production worldwide, with domestic birds being the species that are preferably bred in this type of production system, while the production of pigs can be considered as a complementary activity. BPS animals and their products are typically consumed by their owners, transported and sold in local markets or used as gifts. Sick birds may be slaughtered, sold and consumed. Most BPS do not apply basic hygiene and biosecurity measures and the birds are often in contact with wild birds, swine and other mammals including humans. Live bird market closures have been shown to be an effective way to stop human AIV infections.

Clinical Features and Pathogenesis

Unlike humans and other mammals, influenza disease in birds is caused solely by the influenza virus A genera of the *Orthomyxoviridae* family. Disease ranges from asymptomatic to lethal depending on the viral strain and bird species. AIV may be designated as low or highly pathogenic. Historically, this was based solely on the ability of the virus to kill experimentally infected domestic chickens (intravenous pathogenicity index, IVPI). Today, this designation can be initially based on specific genetic

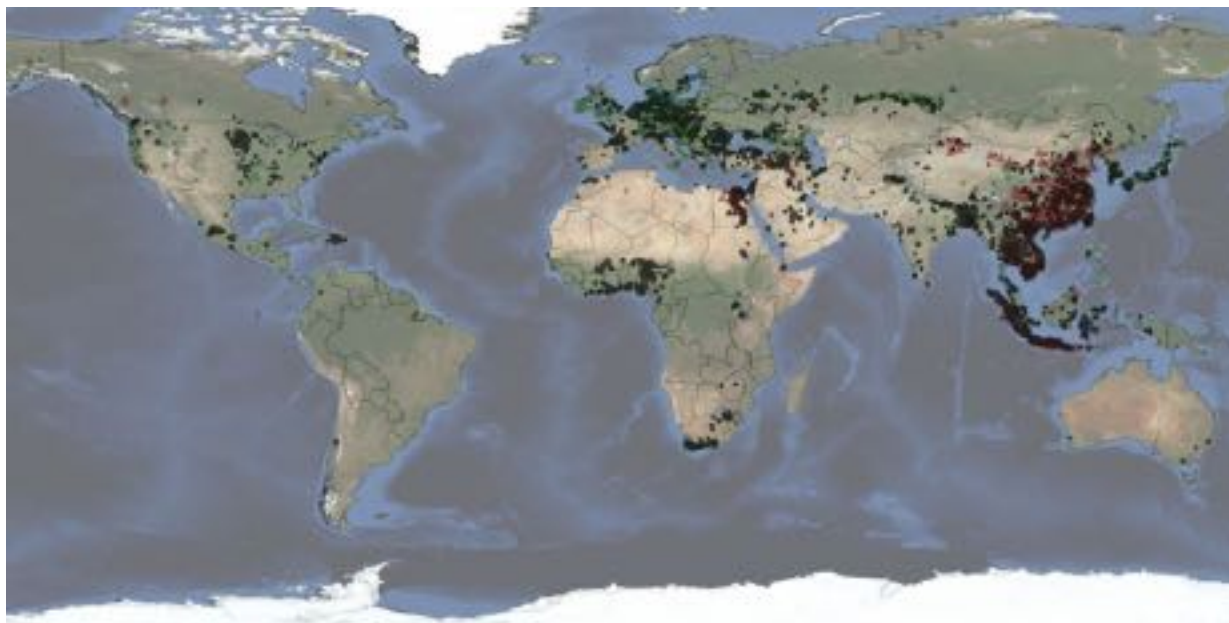


Fig. 1 Global avian influenza virus outbreaks reported to the FAO between 2010 and October 2018. Blue symbol = AIV outbreaks in captive animals; black circles = AIV outbreaks in domestic animals; green squares = AIV outbreaks in wildbirds; and red triangle = AIV outbreaks in humans.

signatures within the hemagglutinin (HA) protein. It is important to note that low and high pathogenicity does not refer to the disease symptoms or infectivity of the virus in humans, mammals or other species of birds.

The majority of AIV infections in wild birds are asymptomatic with the virus replicating primarily in the intestines and can be associated with shedding of high viral concentrations in the feces. Viral replication can also be found in the respiratory tract.

In domestic birds and poultry, disease can also range from a mild drop in egg production to highly pathogenic. Most avian influenza viruses cause a low pathogenic avian influenza virus infection (LPAI), where it is possible to notice respiratory symptoms (coughing, sneezing, sinus swelling and respiratory discharge), sinusitis, and respiratory tract lesions. While less severe than highly pathogenic infections (HPAI), LPAI viral infections can cause considerable losses, especially in turkeys, due to anorexia, depression and decreased egg production.

In contrast, highly pathogenic avian influenza (HPAI) virus infections are associated with severe systemic disease with variable clinical presentations, including respiratory signs, ocular and nasal discharge, cough and dyspnea, swelling of the breasts and head and necrotic combs (Fig. 2). Mortality can reach 100%. Should a bird survive a peracute infection, they may experience central nervous system involvement resulting in lack coordination, wing droop or even paralysis. Microscopic lesions may be present, though their location and severity will vary. HPAI viruses have hitherto been restricted to select H5 and H7 AIV subtypes; however, the majority of H5 and H7 AIVs are low pathogenic. HPAI variants can emerge from LPAI H5 and H7 viruses circulating in domestic poultry populations. It was thought that the primary difference between a HPAI and LPAI virus was solely based on the presence of a polybasic cleavage site in the HA allowing proteolytic processing by ubiquitous host proteases. More recent studies have shown that pathogenicity may be more complicated and involve multiple genetic changes beyond the HA.

In humans, the first example of AIV causing severe and even fatal infections in humans occurred in 1997 when 18 (6 fatal) H5N1 infections were reported in Hong Kong. AIV infections in humans can range from mild to severe and include conjunctivitis, influenza-like illness (fever, cough, sore throat, muscle aches) that can be accompanied by nausea, abdominal pain, diarrhea, and vomiting. Severe respiratory illness including shortness of breath, difficulty breathing, pneumonia, acute respiratory distress, and respiratory failure, neurologic changes and the involvement of other organ systems have also been reported. While the Asian lineage H7N9 and HPAI Asian lineage H5N1 viruses have been responsible for most human illness worldwide to date, including the most serious illnesses and highest mortality, there have been reports of humans infected with H5Nx, H9, H10, and H6 AIV subtypes. In most cases, the infected people had direct contact with infected poultry or objects contaminated by their feces and human-to-human transmission has been inefficient. The possibility that the virus could mutate or reassort and acquire the ability to transmit easily between humans is a continual concern.

Prevention

HPAI outbreaks represent a main threat to the poultry industry. As detailed in the OIE *Terrestrial Animal Health Code*, the OIE must be notified of all cases of HPAI found in any domestic or wild bird by a country's veterinary authorities. LPAI H5 and H7 AIV strains in poultry are also notifiable given their potential to mutate into highly pathogenic viruses or to infect other species.

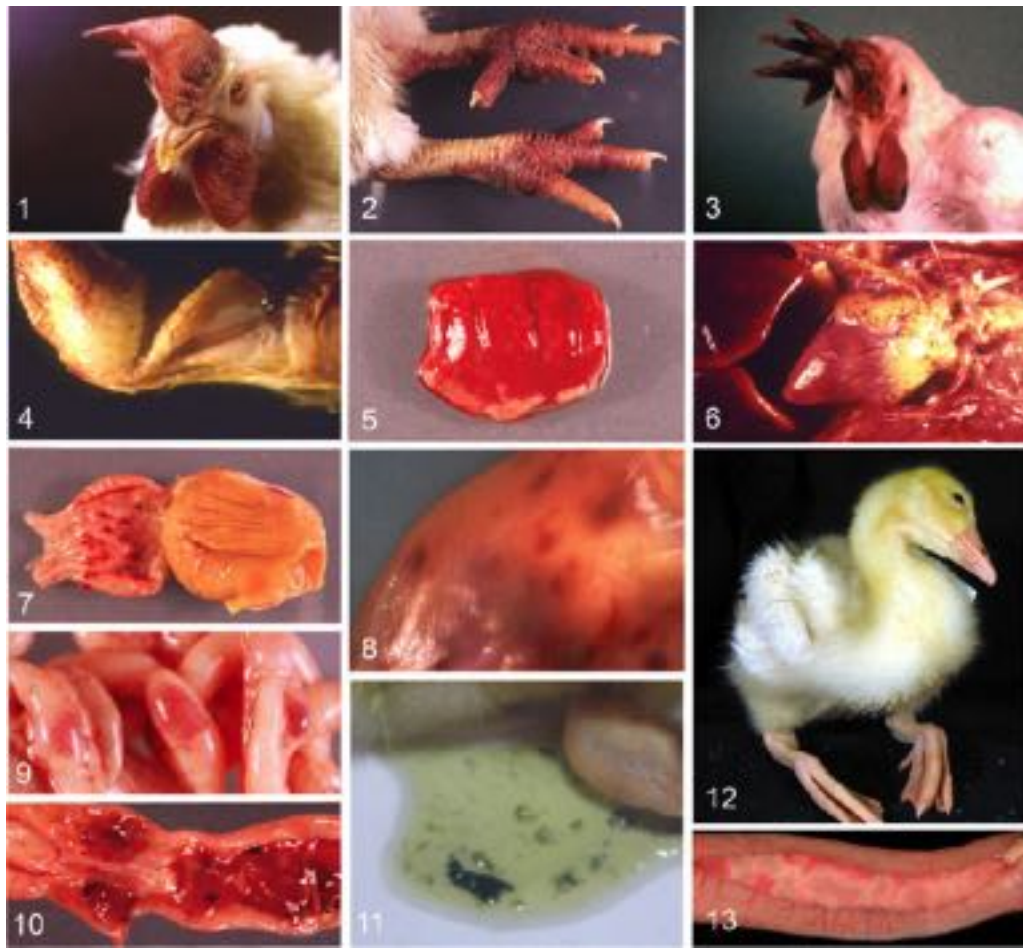


Fig. 2 Gross lesions in chickens, ducks and geese following experimental infection with HPAI viruses. Reproduced from the OIE Scientific and Technical Review (www.oie.int) – Pantin-Jackwood, M.J., Swayne, D.E., 2009. Pathogenesis and pathobiology of avian influenza virus infection in birds. In: Mettenleiter, T., (Ed.), Avian influenza, Revue scientifique et technique (International Office of Epizootics) 28 (1), 113–136. doi:10.20506/rst.28.1.1869. (a) Oedema of comb, wattles, and periorbital tissues, A/chicken/Puebla/1994 H5N2 HPAI virus; (b) Severe subcutaneous haemorrhage and oedema of feet and leg shanks, A/Hong Kong/156/97 (H5N1); (c) Severe oedema, necrosis and haemorrhage of comb and wattles, highly pathogenic embryo derivative A/chicken/NJ/12508/86 (H5N2); (d) Thickened dermis from oedema of distal leg, A/chicken/Queretaro/14589-660/94 (H5N2); (e) Severe pulmonary oedema and haemorrhage in the lung, A/Hong Kong/156/97 (H5N1); (f) Petechial haemorrhages in epicardial fat, A/chicken/NJ/12508/86 (H5N2); (g) Submucosal haemorrhage surrounding ducts of glands in proventriculus, A/chicken/Hong Kong/156/97 (H5N1); (h) Multifocal haemorrhage in the fascial plane of the gastrocnemius muscle (pars intermedia, A/chicken/Hong Kong/220/1997 (H5N1) virus; (i) Haemorrhage in lymphoid tissue of Peyer's patches and Meckel's diverticulum of the jejunum, A/Hong Kong/220/97 (H5N1); (j) Haemorrhage in caecal tonsils and rectum, A/Hong Kong/483/97; (k) Bile-stained loose droppings from a 2-week-old Pekin duck, A/Egret/HK/757.2/02 (H5N1); (l) Two-week-old Embden goose with torticollis, A/chicken/Hong Kong/220/1997 (H5N1); (m) Mottling of the pancreas in a 2-week-old Embden goose, A/chicken/Hong Kong/220/1997 (H5N1).

Veterinary authorities within countries must establish programs to ensure early detection of AIV, response capacities and ongoing implementation of biosecurity measures to protect farms and commercial poultry. If infection is detected, a policy of culling infected and contact animals is typically followed to rapidly contain, control and eradicate the disease. The requirements for containment include:

- Humane destruction of all infected and exposed animals (following OIE animal welfare standards);
- Appropriate disposal of carcasses, litter and all animal products;
- Surveillance and tracing of potentially infected or exposed poultry;
- Strict quarantine and controls on movement of poultry and any potentially contaminated vehicles and personnel;
- Thorough cleaning and decontamination of infected premises. Due to the biochemical characteristics of AIV they are susceptible to a broad variety of disinfectants, among them sodium hypochlorite, quaternary ammonias, 60%–95% ethanol, povidone and other disinfectants. They also can be easily inactivated by medium temperatures (56–60°C for a minimum of 60 min), ionizing radiation and extreme pH either acid or basic (pH 1–3 or pH 10–14);
- A period of at least 21 days before restocking.

This may be applied at the level of the infected farm only or include farms within a short radius of the infected premise. All of this is done in conjunction with active surveillance.

Vaccination can be used to protect susceptible populations from infection. Vaccine strategies can be an effective emergency measure in an outbreak or as a routine measure in endemic areas. A decision to use vaccination must include an exit strategy detailing the conditions that must be met for vaccination to be stopped. The goal of an AIV vaccine for poultry is to reduce virus excretion if they are subsequently infected thus preventing virus circulation. There are a variety of vaccine types available for poultry including monovalent and bivalent inactivated vaccines, subunit killed vaccines, live recombinant vaccines, and DNA-based vaccines as examples. Consideration must be given to routes of administration and timing of administration, differences in host species immune responses, cost, and methods to differentiate vaccinated from infected birds. "DIVA" (differentiating infected from vaccinated animals) was one system that successfully enabled detection of field exposure in a vaccinated population resulting in eradication of LPAI H7 viruses in Italy. The basis of DIVA was use of a vaccine that contained a virus possessing the same HA, but a different NA, as the field virus allowing detection of antibodies to the NA of the field virus. For example, a vaccine containing an H7N3 virus can be used against an H7N1 field virus.

Further Reading

- van Dijk, J.G., Verhagen, J.H., Wille, M., Waldenström, J., 2018. Host and virus ecology as determinants of influenza A virus transmission in wild birds. *Current Opinion in Virology* 28, 26–36. (PMID: 29121508).
- Forment, M., Holmes, E.C., 2015. Avian influenza virus exhibits distinct evolutionary dynamics in wild birds and poultry. *BMC Evolutionary Biology* 15, 120. (PMCID: PMC4481119).
- Global Consortium for H5N8 and Related Influenza Viruses, 2016. Role for migratory wild birds in the global spread of avian influenza H5N8. *Science* 354 (6309), 213–217. (PMCID: PMC5972003).
- Long, J.S., Mistry, B., Haslam, S.M., Barclay, W.S., 2018. Host and viral determinants of influenza A virus species specificity. *Nature Reviews Microbiology* 28. (PMID: 30487536).
- Pantin-Jackwood, M., Swayne, D.E., Smith, D., Shepherd, E., 2013. Effect of species, breed and route of virus inoculation on the pathogenicity of H5N1 highly pathogenic influenza (HPAI) viruses in domestic ducks. *Veterinary Research* 22 (44), 62. (PMID: 23876184).
- Pantin-Jackwood, M.J., Swayne, D.E., 2009. Pathogenesis and pathobiology of avian influenza virus infection in birds. In: Mettenleiter, T., (Ed.), *Avian influenza, Revue scientifique et technique* (International Office of Epizootics) 28 (1), 113–136. doi:10.20506/rst.28.1.1869.
- Ramey, A.M., Deliberto, T.J., Berhane, Y., Swayne, D.E., Stallknecht, D.E., 2018. Lessons learned from research and surveillance directed at highly pathogenic influenza A viruses in wild birds inhabiting North America. *Virology* 518, 55–63. (PMID: 29453059).
- Russell, C.J., Hu, M., Okda, F.A., 2018. Influenza hemagglutinin protein stability, activation, and pandemic risk. *Trends in Microbiology* 26 (10), 841–853. (PMID: 29681430).
- Shaw, M.L., Palese, P., 2013. *Orthomyxoviridae*. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 1151–1186.
- Swayne, D.E., Spackman, E., Pantin-Jackwood, M., 2014. Success factors for avian influenza vaccine use in poultry and potential impact at the wild bird-agricultural interface. *Ecohealth* 11 (1), 94–108. (PMID: 24026475).
- Wright, P.F., Neumann, G., Kawaoka, Y., 2013. *Orthomyxoviruses*. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 1187–1244.
- Yoon, S.W., Webby, R.J., Webster, R.G., 2014. Evolution and ecology of influenza A viruses. *Current Topics in Microbiology and Immunology* 385, 359–375. (PMID: 24990620).

Relevant Websites

- <http://www.offlu.net/fileadmin/home/en/publications/pdf/OFFLUsurveillance.pdf>
OFFLU. Strategy document for surveillance and monitoring of influenzas in animals.
- <http://www.oie.int/en/animal-health-in-the-world/avian-influenza-portal/press-releases/>
OIE. Avian Influenza Portal.
- <http://www.oie.int/en/standard-setting/terrestrial-code/access-online/>
OIE. Terrestrial Animal Health Code.
- <http://www.fao.org/avianflu/en/strategies.html>
Strategy and Policy – FAO.
- <http://www.who.int/influenza/about/en/>
WHO. About the Global Influenza Programme.

Avian Leukosis and Sarcoma Viruses (*Retroviridae*)

Karen L Beemon, Johns Hopkins University, Baltimore, MD, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Avian leukosis virus (ALV) First avian retrovirus, identified in 1908 by Ellerman and Bang, causes lymphomas by insertional activation of host proto-oncogenes.

Avian sarcoma/leukosis viruses (ASLV) Includes ALV and RSV and other avian retroviruses.

Bic Precursor to microRNA 155, discovered by ALV insertional mutagenesis in 1989; over-expressed in many human tumors.

Endogenous retroviruses (ERVs) Retroviruses that have infected germ cells and are transmitted genetically.

Insertional mutagenesis Mechanism of oncogenesis by ALV, which activates a nearby host oncogene after integration into the host genome. Myc expression is commonly induced.

LTR Long terminal repeat formed at both ends of the provirus during reverse transcription and contains U3, R, and U5 sequences.

MC29 Defective ALV containing the myc oncogene.

Myc A transcription factor that is oncogenic when its expression is dysregulated by ALV transduction or insertional mutagenesis.

Oncogene A gene that causes cancer.

Rous sarcoma virus (RSV) First transforming retrovirus, discovered in 1911 by Peyton Rous, causes rapid sarcomas. RSV carries an oncogene *src*, - see below.

Src The first oncogene to be discovered, as part of Rous Sarcoma Virus, and also the first protein tyrosine kinase to be identified.

History

Retroviruses of birds have been studied for over 110 years. Avian leukosis virus (ALV) was discovered in 1908 by Ellermann and Bang in Denmark, who transmitted erythroleukemia between chickens by cell-free filtrates. Rous sarcoma virus (RSV) was isolated by Peyton Rous in 1911 in New York and shown to transmit sarcomas through cell-free extracts from a chicken sarcoma. Rous was awarded the Nobel Prize in 1966 for showing that cancer can be transmitted by a virus; this was the beginning of tumor virology.

Numerous isolates of ALV and of related transforming viruses, which cause sarcomas and a variety of hematopoietic neoplasms, were reported in the decades that followed. Progress in understanding their nature was slow until the development of cell culture assays in the late 1950s and the use of genetic and cell biological approaches to study replication and transformation. The discovery of reverse transcriptase (RT) in RSV in 1970 by Howard Temin and of the origin and mechanism of action of viral oncogenes in the decade following led to an explosion of research activity. In addition to the genetic information of ALV needed for replication, RSV has a transduced oncogene, *src*, that enables it to induce sarcomas rapidly in vivo and to transform cells in culture. *Src* was the first retroviral oncogene to be characterized and the first tyrosine kinase to be identified. Bishop and Varmus received the Nobel Prize in 1989 for showing *src* is derived from its host. ALV-J was isolated in 1988 in England and is currently an economic problem in chicken flocks, especially in Asia.

Classification

The avian leukosis viruses comprise a single genus Alpharetrovirus, of the family *Retroviridae* and the subfamily *Orthoretrovirinae*. Although they share many structural and biological characteristics with the mammalian C-type gammaretroviruses (such as murine leukemia virus), these two groups are not closely related. All ALVs are closely related to one another, sharing considerable sequence and antigenic identity. Isolates are differentiated by subgroup (i.e., based on receptor utilization and determined by the *env* gene), as well as the presence or absence of oncogenes. Subgroups A–E and the more recently isolated subgroups J and K infect chickens. ALV-J was derived by recombination, acquiring a new *env* gene from an endogenous retroelement, EAV-HP. Subgroups F–I refer to endogenous viruses found in ring-necked pheasant (subgroup F), golden pheasant (subgroup G), Hungarian partridge (subgroup H), and Gambel's quail (subgroup I).

ALVs are endemic in flocks of domestic chickens (*Gallus gallus*) worldwide and will replicate efficiently in species closely related to the chicken such as quail, turkeys, and pheasants. Some RSV subgroups can transform mammalian cells and induce tumors in mammals, but with greatly reduced efficiency. The restriction of the virus to avian species is due to a lack of suitable receptors for most subgroups as well as to blocks to viral gene expression. The rare transformants that arise in RSV-infected mammalian cells often display rearrangements in proviral DNA that relieve this block.

A variety of cell types from gallinaceous birds (including chickens, turkeys, and quail) can be used to propagate avian retroviruses. Primary and secondary fibroblast cultures or cell lines (DF-1) are most commonly used, as well as lines derived from quail tumors (QT6). To avoid problems associated with frequent recombination, it is advisable to use cells that do not contain

related endogenous proviruses, such as cells from species other than chickens or from chickens bred to lack endogenous proviruses closely related to ALVs.

Virion Structure

Like all retroviruses, ALVs are transmitted as enveloped virions of about 100 nm diameter, derived by budding from the host cell membrane. Lipids derived from the host plasma membrane are present in the viral envelope and make up 35% of the virion weight. Within the retrovirus family, they are defined as having a C-type morphology. Small, dispersed spikes project from the surface of the virion; these are trimers of the two *env*-encoded proteins, SU (surface) and TM (transmembrane). The internal core of the virion appears in electron micrographs as a centrally located, roughly spherical structure about 30 nm in diameter. Immature virions seen during or shortly after budding have a more open, spherical core structure, substantially larger in diameter than the mature one. The core comprises about 1500 copies each of the five *gag*-encoded proteins (MA, p10, CA, NC, and protease) and about 100 copies each of *pol*-encoded reverse transcriptase and integrase. The avian retroviruses are unique in encoding the viral aspartyl protease as part of the *gag* gene; in other types of retroviruses, it is part of *pol*.

Virions of ALV have an equilibrium density in sucrose solutions of about 1.16–1.18 g/ml and a sedimentation coefficient of about 600S. They are quite labile and are readily inactivated by extremes of pH, as well as by heat or mild detergent treatments. They are somewhat radiation resistant, perhaps reflecting the recombinational repair capability provided by the dimeric genome.

Genome

The ALV genome consists of a homodimer of positive-sense, single-stranded (ss) RNA about 7500 nt in length. The genomic complexity of retroviruses was first determined for RSV by oligonucleotide fingerprinting and found to be the size of one 35S RNA subunit. Subsequently, all retroviruses were found to be diploid. The RNA of RSV was found to be about 2 kb larger than that of ALV due to the presence of the *src* oncogene. Transforming viruses, in which an oncogene has been inserted, have genomes varying in length from about 3200 nt (for UR2 virus) to about 9300 nt (for nondefective RSV). In most of these viruses, the oncogene has replaced some of the viral genes, leading to a genetically defective virus, which requires co-infection with helper ALV for replication. Oncogenes transduced by ALV from chickens include *src*, *myc*, *myb*, *erbA*, *erbB/EGFR*, *fps*, *ros*, and *jun*.

The avian retroviral genome is modified and processed by cell machinery. It contains a 5' m⁷GpppGm cap and a 3' poly (A) sequence, as well as internal m⁶A residues. The gene order of retroviruses was first established with the avian retroviruses.

ALV: 5'-*gag-pol-env*-3'

RSV: 5'-*gag-pol-env-src*-3'

The RSV primary transcript is incompletely spliced to yield an unspliced mRNA (which also serves as the viral genomic RNA) and spliced mRNAs for *env* and *src*. The negative regulator of splicing (NRS) within the *gag* gene, as well as inefficient 3' splice sites, inhibit complete splicing.

Important noncoding regions found near the end of the genome are necessary to provide signals for virus replication. These include an 18–21 base sequence (R) repeated at each end as well as unique sequences U3 (c. 250 nt) near the 3' end and U5 (c. 80 nt) near the 5' end which are duplicated in the long terminal repeat (LTR) during reverse transcription. The LTR contains sequences controlling transcription initiation (enhancers and promoters) and polyadenylation. Adjacent to these are the sites for initiation of reverse transcription: the primer binding (PB) sequence next to U5. Between PB and the beginning of *gag* is an approximately 300 nt leader region, which contains signals important for the dimerization and packaging of the genome into virions. There are also three short open reading frames in this region upstream of *gag*.

The direct repeat (DR) sequences flanking the *src* gene in RSV are necessary for cytoplasmic accumulation of full-length viral RNA. ALV has one copy of the DR sequence, which is involved in export of the RNA from the nucleus to the cytoplasm, with the help of cellular proteins NXF1 and Dbp5. Further, an RNA stability element (RSE) downstream of the Gag termination codon inhibits nonsense-mediated decay of the ALV genome. The RSE binds PTBP1 and inhibits binding of UPF1 to the terminating ribosome complex.

The ALV genome encodes nine proteins, the products of three genes. The Gag proteins constitute the major structural components and are sufficient to form virus-like particles. The Gag-Pro precursor is processed during release of virus into four Gag proteins: MA (matrix, 19 kDa) which interacts with the cell membrane; p10, a 10 kDa protein; CA (capsid, 27 kDa) which forms the core shell structure; and NC (nucleocapsid, 12 kDa), an RNA-binding protein necessary for specific encapsidation of genomic RNA. The Gag-Pro precursor also contains the 15 kDa PR (protease) peptide necessary for processing all internal virion proteins. The Pol reading frame is expressed as a fusion protein with Gag and Pro, processed to yield RT (usually present as a heterodimer of 98 and 66 kDa reflecting partial processing by PR), and integrase (IN, about 32 kDa); these are the two enzymatic components necessary for synthesis and integration of the DNA provirus. The *env* gene encodes the Env precursor (Pr95) which is processed as a membrane protein and cleaved by host cell proteases to yield the SU and TM glycoproteins. SU has an apparent molecular weight of about 85 kDa, about half of which is due to the presence of approximately 14 N-linked carbohydrate side chains. The TM

glycoprotein has an apparent molecular weight of about 37 kDa. The SU and TM products remain as a disulfide-bonded heterodimer with SU containing the activity necessary for receptor binding and TM mediating fusion with the cellular membrane.

Life Cycle

Replication of avian retroviruses requires reverse transcription of the RNA genome to DNA and its integration into the host genome. This group of viruses provided some of the important early models for studying these processes. Howard Temin proposed and demonstrated reverse transcription of RSV RNA by a virus-encoded enzyme and received the Nobel Prize in 1975 for this work. The viral integrase protein is necessary for cleavage of the host genome and insertion of the provirus. The cellular FACT protein complex facilitates this process for the avian viruses. The ALV proviruses show fairly random integration sites with a slight preference for transcription start sites in tissue culture. However, in tumors there is a strong selection for integration near oncogenes, and the ALV regulatory sequences drive expression of the cellular oncogenes.

Entry of the virion into the cell follows interaction with a specific receptor on the cell surface. Viral subgroups (A–J) have been identified on the basis of distinct receptor recognition. The presence of receptors for specific subgroups is polymorphic among birds. Four unlinked genetic loci (Tv-a, Tv-b, Tv-c, and TV-j) for ALV receptors have been genetically identified in chickens. The dominance of susceptibility over resistance alleles at each of these loci implies that they encode the receptor directly. The Tv-b locus has several alleles, controlling susceptibility to subgroups B, D, and E. Receptors for ASLV subgroups A, B, C, D, E, and J have been identified. The Tv-a receptor resembles a portion of the receptor for low-density lipoprotein and is unrelated to other known retroviral receptors. The receptors for B, D, and E are all in the tumor necrosis factor receptor family. The ALV-C receptor is in the butyrophilin family and consists of two immunoglobulin-like domains. ALV-J uses the Na⁺/H⁺ exchanger 1 as its receptor. The newly isolated ALV-K shares the Tv-a receptor with subgroup A viruses. Entry of the virion core into the cell is by fusion of viral and cellular membranes, which is mediated by the viral glycoproteins.

Once within the cytoplasm of the infected cell, reverse transcription within the permeabilized core structure copies the ssRNA genome into double-stranded DNA. This process – which varies little from that of other retroviruses – includes a series of ‘jumps’ from one end of the template to the other. The DNA product is longer than the genomic RNA template and has generated a long terminal repeat (LTR) at both ends. The LTR contains sequences necessary for DNA integration and for synthesis and processing of viral RNA. Like other retroviruses, ALVs exhibit very high rates of homologous recombination – a consequence of the diploid genome and the ‘jumping’ mechanism of reverse transcription. The latter also permits relatively high rates of nonhomologous recombination, leading to frequent rearrangements of the genome as well as the occasional acquisition of foreign sequences such as oncogenes.

Integration of viral DNA into fairly random sites in the cell genome is accomplished by the IN protein which has entered the cell with the virion and remains associated with the DNA. The cellular FACT protein complex interacts with ALV IN and promotes integration. The process of integration leads to the insertion of the linear viral DNA into cell DNA. Integrated ALV DNA is characterized by the loss of two bases from each end of the viral LTR sequence and the duplication of six bases of the cell DNA target sequence at the integration site.

RNA transcription of the provirus is mediated by cellular RNA polymerase II, directed to the correct initiation site by promoter and enhancer sequences in the LTR. The strength of the enhancer elements is a major factor distinguishing pathogenic from nonpathogenic (RAV-0) ALV isolates. Unlike the more complex retroviruses, there is no apparent role of virus-encoded proteins in regulating the transcription process. Processing of the viral transcripts includes addition of poly(A) following a canonical signal (AAUAAA) in the RNA derived from the 3′ LTR and splicing of the fraction of the transcripts destined to become mRNA for the *env* and *src* genes. Splicing removes most of the *gag*, *pro*, and *pol* sequences. Importantly, a fraction of the primary unspliced RNA transcript serves as genomic RNA and as the mRNA for Gag and Pol polyproteins.

Translation of the full-length mRNA leads to two products: The Gag precursor of 76 kDa and the Gag-Pol precursor of 180 kDa. Synthesis of the latter polyprotein is made possible by a –1 translational frameshift about 5% of the time, reading through the termination codon at the end of Gag. Assembly of the precursors appears to be at the cell surface and is coincident with budding. Release of the immature particle (characterized by a hollow, symmetrical core which almost fills the virion) is rapidly followed by cleavage of the Gag and Gag-Pol precursors. This cleavage is accompanied by condensation of the core into its mature form. Since the PR protein embedded in the Gag precursor contains only one-half of the active site, dimerization of this domain is necessary for cleavage to occur. This requirement probably helps to delay cleavage until the appropriate time after budding.

Once infected, the host cell is usually not killed by virus replication. A strong superinfection resistance due to blockage or loss of viral receptors develops soon after infection and prevents accumulation of proviruses by reinfection. In some cases, weak or slow development of superinfection resistance is associated with a cytopathic interaction of the virus with its host cell.

Epidemiology

ALV's have been a major cause of death in poultry, mainly by induction of leukoses in hematopoietic cells. ALV spreads by both vertical and horizontal transmission and is controlled mainly by eradication of breeding stock containing ALV. ALV's subgroup A and B were the predominant subgroups in the 20th century and shown to cause B cell lymphomas, originating in the bursa of Fabricius and metastasizing to other organs. They were largely brought under control in the late 20th century.

However, a new subgroup of ALV (subgroup J), causing myeloid tumors, was isolated from meat-type (broiler) chickens in 1988 in England. Similar viruses were isolated from broiler chickens in the US and China. By 2004 ALV-J infection had spread to layer chickens in China, causing hemangiomas and other diseases, as well as reduced egg laying. It has since become a major economic problem in Asia. Their pathogenicity is evolving rapidly with a change in host range and tumor types. The layer chicken isolates have incurred genetic changes in the 3' UTR and *env* gene.

ALVs are endemic in flocks of domestic chickens (*Gallus gallus*) worldwide and will replicate efficiently in species closely related to the chicken such as quail, turkeys, and pheasants. Some RSV subgroups can transform mammalian cells and induce tumors in mammals, but with greatly reduced efficiency. The restriction of the virus to avian species is due to a lack of suitable receptors for most subgroups as well as to blocks to viral gene expression. The rare transformants that arise in RSV-infected mammalian cells often display rearrangements in proviral DNA that relieve this block.

A variety of cell types from gallinaceous birds (including chickens, turkeys, and quail) can be used to propagate avian retroviruses. Primary and secondary fibroblast cultures or cell lines (DF-1) are most commonly used, as well as lines derived from quail tumors (QT6). To avoid problems associated with frequent recombination, it is advisable to use cells that do not contain related endogenous proviruses, such as cells from species other than chickens or from chickens bred to lack endogenous proviruses closely related to ALVs.

Transmission of virus is principally vertical by infection of the offspring through virus secreted into the egg. Indeed, high titers of ALV are often detectable in commercial hen's eggs. Horizontal spread of virus is naturally much rarer, requiring close contact, but virus can be readily spread from infected birds via contaminated needles during vaccination or through vaccines prepared from infected eggs or cell cultures. All isolates of ALV replicate efficiently in fibroblast cultures and in the bursa. Tropism for other tissues varies among isolates and is determined by both *env* and LTR sequences.

Clinical Features

ALV's of subgroup A and B cause B-cell lymphomas and other neoplasms. The B cell lymphomas originate in the bursa of Fabricius and metastasize to liver, spleen, and kidneys. They cause death in a few months. In contrast, ALV-J causes predominantly myeloid tumors and hemangiomas. Egg production was drastically reduced. Immunosuppression and weight loss also occurred.

ALV's with transduced oncogenes cause a variety of more rapidly occurring (acute) cancers, including sarcoma, myeloblastosis, and erythroblastosis by over-expression of mutated oncogenes from the viral LTRs.

Pathogenesis

ALVs induce a wide spectrum of disease in naturally or experimentally infected animals. The prototypic disease induced by ALV is a B-cell lymphoma arising in the bursa of Fabricius starting a few months after infection and spreading to the liver and other organs during its course. Other malignancies, including myeloid tumors, erythroleukemia, sarcoma, and others, are not uncommon depending on the strain of virus and bird and the time and route of inoculation. The malignant diseases induced by viruses, which do not contain oncogenes, are the consequence of insertional activation of cellular proto-oncogenes (such as *myc* in the case of lymphoma, *erb-B* in erythroleukemia, etc.). This process is termed promoter insertion or insertional mutagenesis and was first observed by Bill Hayward with ALV integrations in *myc*. Additional integrations into the noncoding *bic* gene were observed in 1989. *Bic* was subsequently found to be the precursor of miR-155. When viruses infect 10–14-day embryos, more rapid lymphoma induction is observed with frequent integrations into *myb* and an antisense long noncoding RNA in the TERT promoter. In addition to malignancies, these viruses also induce hemangiomas, osteopetrosis, and wasting diseases. In some cases, an immune response against infected cells may be important; in others, cytopathic effects may play a significant role.

A unique characteristic of ALV and a few other retroviruses is their ability to incorporate host sequences into their genome and to alter the function of proto-oncogenes to generate oncogenes. The presence of an oncogene renders the virus capable of inducing malignant transformation of cells in culture and one of a variety of malignant and rapidly fatal diseases in birds. At least 20 distinct cell sequences have been incorporated by ALV into a number of distinct isolates. RSV, which contains *src*, is the prototype oncogene-containing virus. Other notable oncogene-containing ALV variants include avian myeloblastosis virus (AMV; containing *myb*); avian myelocytomatosis virus-29 (MC-29; *myc*); avian erythroblastosis virus (AEV; *erb-A* and *erb-B*); Fujinami sarcoma virus (FSV; *fps*); and University of Rochester sarcoma virus-2 (RU-2; *ros*). Study of the genetic alterations that distinguish these oncogenes from proto-oncogenes, and the enzymatic and physiological function of the proteins they encode has been an important part of modern cancer research. Incorporation of oncogenes into the virus genome is usually at the expense of some viral genes and co-infection of a cell with a wild-type (helper) ALV is necessary to provide viral proteins for replication of oncogene-containing viruses. RSV, which exists as a replication-competent transforming virus, is the exception. It should be noted that the oncogene-containing viruses are not efficiently transmitted from one animal to another due to their rapid pathogenicity. In most cases, they have probably arisen in the animal from which they were isolated and would have died out if not brought into the laboratory. Not all members of this group are highly pathogenic. RAV-0 (an endogenous virus) can infect susceptible chickens and induce viremia, but disease is rare and occurs only after a long latent period. The reduced virulence is probably an important feature of viruses inherited in the germline.

In infected birds, the only significant immune response is the appearance of type-specific neutralizing antibodies, which apparently recognize the regions of Env involved in receptor recognition. Group specific responses against Env or other proteins are not usually observed in infected chickens, although inoculation of virus into mammals induces antibodies capable of recognizing all virion proteins in the absence of subgroup-specific reactivity. The limited immune response observed in infected chickens has been attributed to the presence of endogenous proviruses whose expression (even at a low level) can induce tolerance to antigens in common with infecting virus. Indeed, it has been suggested that induction of tolerance might be a desirable feature for the animal, since it could prevent or limit immunopathological sequelae of infection. Postinfection immune response seems to be of little consequence in preventing subsequent malignant disease, since the cells, which will eventually form the tumor, are probably infected quite soon after infection, and the long latency reflects the necessity for subsequent rare events (such as mutations in other genes) rather than a continuing period of virus replication.

Diagnosis

Avian retroviruses can be detected by ELISA assays or PCR assays.

Treatment

Control of ALV infection is generally by detection and culling of infected individuals. No useful vaccination strategy has been developed. In principle, it should be possible to virtually eliminate the disease by breeding appropriate receptor alleles into commercial strains.

Lymphoid leukosis caused by ALV (subgroups A and B) was eradicated in primary breeders in the US in the 1980s and 1990s, reducing the disease incidence in commercial layer hens. However, commercial broiler and layer chickens are still struggling with ALV-J and ALV-K infection in many countries, especially in Asia.

Endogenous Retroviruses

ALV's can become established in the germline and inherited stably as endogenous retroviruses (ERVs). Naturally occurring endogenous alpha proviruses form a distinct lineage of ALVs, showing a specific host range (subgroup E) for which many domestic chickens lack receptors (a phenomenon known as xenotropism), and a reduced replication capacity and pathogenicity relative to exogenous viruses. Endogenous viruses are usually expressed at a very low level and are usually defective in sequence.

Analysis of the sequenced chicken genome revealed many endogenous retroviruses related to the gamma and beta retroviruses, with fewer alpha retroviruses. There were also proviral intermediates between alpha and beta genera. About 20% of these ERVs were transcribed and associated with ribosomes. In addition, there were thousands of solo LTRs.

Further Reading

- Bolisetty, M., Blomberg, J., Benachou, F., Sperber, G., Beemon, K., 2012. Unexpected diversity and expression of avian endogenous retroviruses. *mBio* 3 (5), e00344-12.
- Coffin, J.M., Hughes, S.H., Varmus, H.E., 1997. *Retroviruses*. New York: Cold Spring Harbor Laboratory Press.
- Hayward, W.S., Neel, B.G., Astrin, S.M., 1981. Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukosis. *Nature* 290, 475–480.
- Rous, P., 1911. A sarcoma of the fowl transmissible by an agent separable from the tumor cells. *Journal of Experimental Medicine* 13, 397–411.
- Sefton, B.M., Hunter, T., Beemon, K., Eckhart, W., 1980. Evidence that the phosphorylation of tyrosine is essential for cellular transformation by Rous sarcoma virus. *Cell* 20, 807–816.
- Stehelin, D., Varmus, H.E., Bishop, J.M., Vogt, P.K., 1976. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* 260, 170–173.
- Temin, H.M., Mizutani, S., 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226, 1211–1213.
- Venugopal, K., 1999. Avian leukosis virus subgroup J: A rapidly evolving group of oncogenic retroviruses. *Research in Veterinary Science* 67, 113–119.
- Weiss, R.A., Vogt, P.K., 2011. 100 years of Rous sarcoma virus. *Journal of Experimental Medicine* 208, 2351–2355.
- Winans, S., Larue, R.C., Abraham, C.M., *et al.*, 2017. The FACT complex promotes avian leukosis virus DNA integration. *Journal of Virology* 91, e00082-17.

Bluetongue Virus (*Reoviridae*)

Raghavendran Kulasegaran-Shylini, Department of Pathogen Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom

Polly Roy, Department of Pathogen Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom and University of Reading, Reading, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Bluetongue virus (BTV) infects a wide range of ruminants, including both domestic (sheep, cattle, goat, water buffaloes, camels etc.) and wild animals (white-tailed deer, elk, blesbuck etc.) and is common throughout the world, periodically causing serious outbreaks, particularly in sheep and cattle. BTV is transmitted by blood sucking biting midges (gnats) of the genus *Culicoides*.

Bluetongue disease in sheep and cattle was first described in the late 18th century, but was recorded only in the late 19th century by Hutcheon, the Chief Veterinary officer of Cape Colony in South Africa. The clinical features were officially reported in 1902 as ‘malarial catarrhal fever’ of animals. However, since the diseased sheep generally exhibited distinctive lesions in their mouths and on the tongue, with the latter often turning blue, Spreull suggested the use of the common name ‘bluetongue’. In addition to this, Spreull also discovered that the causative agent was filterable and transmissible to cattle and goats causing subclinical infection.

The disease was confined to Africa until 1943 when the first recognized outbreak outside of Africa occurred in Cyprus and soon after, in surrounding countries. Similar outbreaks were also observed in the USA around the same period and virus was isolated from one such outbreak in California in 1952. The disease then spread into southwest Europe and the Indian subcontinent causing severe outbreaks. Over the years, many serotypes of the virus have been isolated from all over the world including Australia, Central and South America and the Caribbean. In recent years, BTV has caused infections in most countries in central and southern Europe and has spread as far north as the United Kingdom and Denmark.

Due to its serious economic importance, BTV has been studied extensively and substantial progress has been made in characterizing the virus at the atomic, molecular, and genetic levels. BTV classification, three-dimensional structure, genetics and the virus replication cycle are summarized here along with epidemiology, pathogenesis and diagnosis, as well as vaccination strategies.

Classification

BTV is the prototype of the orbivirus genus, the largest of the genera within the *Reoviridae* family. Currently, 22 distinct virus species have been classified under the orbivirus genus. Orbiviruses are insect-borne viruses capable of infecting a wide range of hosts, including humans, animals, plants, and insects. Despite basic similarities, orbiviruses differ greatly in their structure, physicochemical properties, replication cycle, pathogenesis, and epidemiology. BTV includes 28 distinct serotypes identified worldwide, including novel isolates of atypical BTV, all of which can be genetically traced to a singular geographic origin.

Virion Structure and Genome

BTV is the most extensively studied virus among the Orbiviruses. The particle has a diameter of 86 nm, and consists of two capsids of concentric protein layers, comprising seven structural proteins and a genome of 10 dsRNA segments. The outer capsid is made up of two proteins, VP2 and VP5, which surround the inner capsid, termed “core”. The core is composed of VP7 and VP3 in two concentric layers, encapsidating the dsRNA genome and the viral replicase complex of three minor proteins, VP1, VP4, and VP6.

The complete genome of BTV-10 was the first sequenced in its entirety in 1989 and was the largest reported complete genome sequence of an RNA virus at the time. The total length of BTV-10 genome was determined to be 19,218 base pairs, with each segment varying in length, from 3954 bp to 822 bp. Based on the size differences, the RNA segments were numbered as S1-S10 and grouped into three size classes, large (S1-S3), medium (S4-S6) and small (S7-S10). All 10 RNA segments share unique 5' and 3' complementary ends and lack a 3' poly-A tail. Segments 1–8 each have only a single open reading frame, encoding six of the seven structural proteins (VP1-VP5 and VP7) as well as two non-structural proteins (NS1 and NS2). The two smallest segments, S9 and S10, have overlapping ORFs, S9 encodes the smallest BTV structural protein, VP6, and a small non-structural protein, NS4 encoded in the +1-reading frame, while S10, encodes the two isoforms of the non-structural protein NS3, namely NS3 and NS3A, with the latter lacking the N-terminal 13 amino acid residues (Table 1). A recent report suggests that S10 may also encode an additional NS protein, NS5, with unknown function.

Mature BTV particles are relatively fragile, unlike for example, reovirus and rotavirus particles, and therefore pose a challenge to purify intact virion particles from infected cells. The initial transmission electron micrographs (TEM) of BTV suggested a double-shelled arrangement of the virion particle. A low resolution (5 nm) cryo-electron microscopy (cryoEM) reconstruction of the virion particles and recombinant virus-like particles (VLPs) formed by the co-expression of four major capsid proteins (VP2, VP3, VP5 and VP7) revealed that the particles have distinct icosahedral morphology. The analysis also showed the specific arrangement of

Table 1 Proteins encoded by BTV segments and their functions

Genome segment	No. of bp ^a	Proteins	No. of aa	Predicted size (Dalton)	Estimated no. of molecules	Location in virion	Function
S1	3954	VP1	1302	149,588	12	Inner core	RNA dependent RNA polymerase
S2	2926	VP2	956	111,112	180	Outer capsid	Receptor binding, virus entry, hemagglutinin, type specific neutralization
S3	2772	VP3	901	103,344	120	Core	Scaffold for VP7 trimers
S4	2011	VP4	654	76,433	~24	Inner core	Capping enzyme with guanylyltransferase, methyltransferases 1 and 2, RNA 5' triphosphatase, inorganic pyrophosphatase NTPase activity
S5	1639	VP5	526	59,163	360	Outer capsid	Membrane penetration, virus entry
S6	1770	NS1	552	64,445	NA	Non Structural	Upregulates viral protein translation, forms tubules
S7	1156	VP7	349	38,548	780	Core surface layer	Viral core protein, group-specific antigenic determinant
S8	1123	NS2	357	40,999	NA	Non Structural	Phosphorylated, recruits viral ssRNA and core components, forms cytoplasmic VIBs, site of viral core assembly
S9	1046	VP6	328	35,750	60–72	Inner core	Involved in viral genomic RNA packaging, has hexameric configuration, helicase
S10	822	NS4	~79	~10,000	NA	Non Structural	Host-virus interaction, regulates interferon response
		NS3	229	25,572	NA/Low level	Non Structural	Glycoprotein, responsible for virus trafficking and egress
		NS3A	216	24,020			

^aBased on BTV 10 genome sequence.

the two outer capsid proteins (VP2 and VP5) for the first time, one with 120 globular densities and the other with 60 spike-like triskelion structures. It was proposed that trimers of VP2 form the triskelion motifs while the globular configurations were the VP5 trimers. With further advancements in cryoEM data acquisition, density reconstitution, and image reconstruction, the structure of the virion particle was further refined to ~24 Å and subsequently to 7 Å. These studies confirmed that the sail-like spikes of triskelion motifs are VP2 (111 kDa) trimers and the globular densities are VP5 (~59 kDa) trimers. The surface of the core forms a middle layer with 260 VP7 (~38 kDa) trimers and an inner layer formed by 60 VP3 (~103 kDa) dimers (detailed below). A recent 3.5 Å structure of BTV further revealed the secondary structural features of VP2 and VP5, and the specific sites of interaction between these two proteins and the scaffolding inner core, at atomic resolution (Fig. 1). Structurally each VP2 monomer (120 in total) is divided into four distinct domains, a hub, a hairpin, a pyramid-shaped body, and a highly flexible external tip. The exposed nature of this domain is consistent with its role as a key determinant of the host antibody response. The hub domain contains a lectin-like β -barrel flanked by helices that drive the trimerization of three monomers such that they sit above the VP7 layer of the core surface. A typical zinc-finger motif, CCCH, with tetrahedral coordination is located at the interface of the hub and the body domains. Each of the 120 VP5 trimers forms a compact globular structure. Each monomer (360 in total) forms a rectangular shaped structure with three distinct domains, an N-terminal flexible dagger sequestered in the canyons underlying the core surface, a helix-rich unfurling domain with a central coiled-coil motif that facilitates trimerization (analogous to a viral fusion protein), and an anchoring domain in the C-terminal region. This domain possesses four antiparallel β -strands with a potential membrane-binding motif. The VP5 trimers surround the VP2 trimers, underlying the gaps in the triskelion arms of VP2 such that the hairpin and hub domains of VP2 are in contact with the unfurling and membrane associating motifs of VP5.

The core (~470 S), unlike the virion particle, is highly robust and was the first particle of BTV analyzed by cryoEM in 1992. The 75 nm structure revealed a T = 13 icosahedral lattice and the presence of concentric layers of VP7, forming the core surface and VP3, forming the inner layer (subcore). The X-ray crystal structure of recombinant VP7 revealed an unusual organization of VP7 trimers, which provided a paradigm in the field and enabled the initial characterization of the assembly of the VP7 layer onto the VP3 layer at 23 Å resolution by cryoEM. These analyses further facilitated solving the subsequent X-ray crystal structure of the core at ~3.5 Å resolution (Fig. 1). The surface of the core is made up of 780 copies of the VP7 monomer arranged as 260 trimers with a T = 13 icosahedral lattice arranged around 132 channels as six-membered rings, with five-member rings at the vertices. In the T13 lattice, VP7 trimers occur as pairs of five quasi-equivalent trimers with each trimer containing two distinct domains, an upper β -rich domain, exposed in the core surface, making extensive contacts with the outer capsid and a helical lower domain that interacts with the VP3 layer below. The underlying subcore layer is composed of 60 VP3 dimers arranged as 12 decamers on a T = 2 symmetry. The VP3 dimers are made up of two nearly identical isoforms (A & B) with a unique plate-like structure resembling a triangular wedge consisting of apical, carapace and dimerization domains. Each VP3 decamer consists of five 'A' molecules arranged around a five-fold apex with a central pore and five 'B' molecules arranged around 'A'. In addition to the pores at the five-fold axes, there are pores at the three-fold axes that are blocked by VP7 trimers in the intact core. VP7 trimers in the core additionally make contact with the tip and hub domains of the VP2 spikes and extensively with all three VP5 domains with the VP5 domains bridging the channel formed by six VP7 trimers in the core. The VP3 layer has been suggested to be in contact with the enclosed viral dsRNA genome.

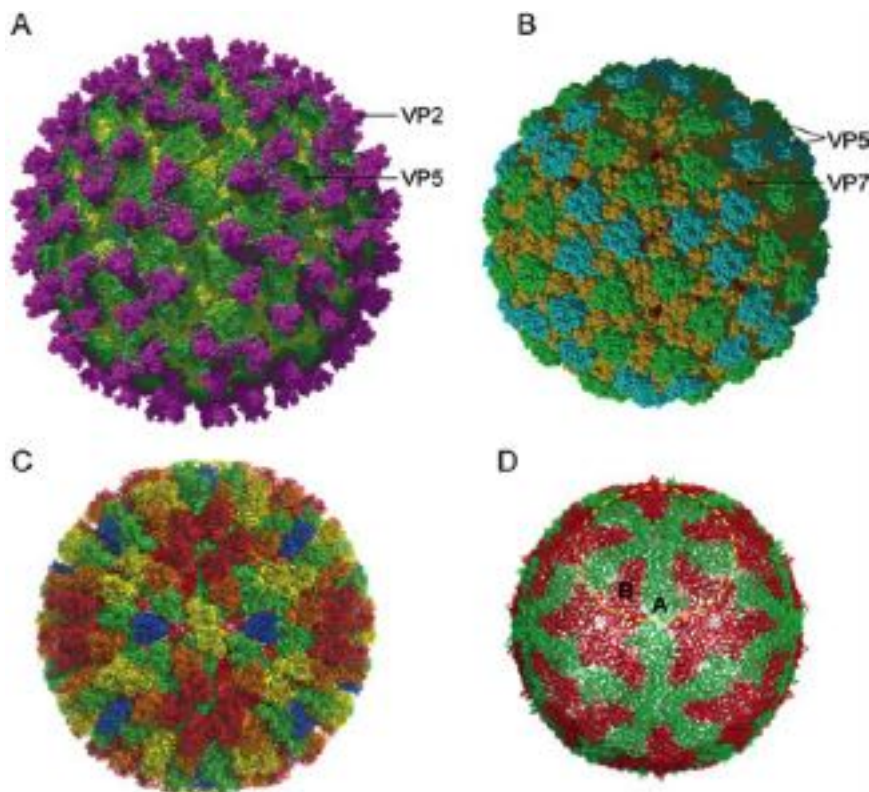


Fig. 1 Structure of BTV virion. (A) and (B) 3.5 Å cryo-EM density map of BTV virion, shown as radially colored surface representation (A) Outer layer showing VP7 triskelions (magenta), VP5 trimers (green) and VP7 trimers (yellow) (B) VP5 layer showing the two different VP5 trimer conformers (green and cyan) and VP7 trimers (yellow). (C) and (D) X-ray crystal structure of BTV core particles showing (C) 260 VP7 trimers arranged as five quasi-equivalent trimers in $T = 13$ icosahedral symmetry (colored in red, orange, green, yellow and blue), (D) VP3 scaffold showing the unique A (green) and B (red) isoforms and highlighting VP3 decamer (dashed ellipse). Adapted from: (A) and (B) Zhang, X., Patel, A., Celma, C.C.P., *et al.*, 2016. Atomic model of a nonenveloped virus reveals pH sensors for a coordinated process of cell entry. *Nature Structure Molecular Biology* 23, 74–80. (C) and (D) Grimes, J.M., Burroughs, J.N., Gouet, P., *et al.*, 1998. The atomic structure of the bluetongue virus core. *Nature* 395 (6701), 470–478.

While these studies provided detailed structural descriptions of the VP7 and VP3 layers, the exact positioning of the three minor proteins and the dsRNA genome within the core remains unclear. However, X-ray diffraction patterns of the BTV cores suggest that most of the genomic dsRNA is packed as well-ordered layers surrounding putative transcription complexes at the apices of the core particle. Further, cryoEM reconstruction of recombinant core-like particles (CLPs) of VP3 and VP7 along with two minor proteins, VP1 (polymerase) and VP4 (capping), revealed a flower-shaped density at the 5-fold axes underneath the VP3 layer.

Life Cycle

In both mammalian and insect cells, BTV follows the typical steps of virus replication, which begin with virus entry and release of intact viral cores, followed by synthesis of viral ssRNA, translation of the viral proteins, amplification of the viral genome, assembly of nascent virus particles and, finally, egress (Fig. 2).

Virus Attachment and Entry Into Cells

In mammalian cells, BTV infection is initiated by attachment of the virus to the cell surface by the outer capsid protein, VP2, primarily via the tip domain that protrudes from the surface of the particle. While the identity of the receptor and the exact mechanism of cell entry have not been elucidated, it has been reported that sialoglycoproteins and sialic acids play roles in the virus entry process. BTV primarily uses the endocytic pathway and is transported to the cytoplasm via clathrin-coated vesicles. Once internalized, the clathrin coat is released and the endocytic vesicle fuses with an early endosome. The CCCH zinc-finger motif in VP2 acts as a pH sensor in the early endosome, triggering a conformational change, and the eventual disengagement of VP2 from the virus particle. The particle lacking VP2 is then transported to the late endosomal compartment, where a histidine cluster – located at the interface of the anchoring domain and the helices of the unfurling domain of VP5 – senses the late-endosomal pH. This triggers the release of

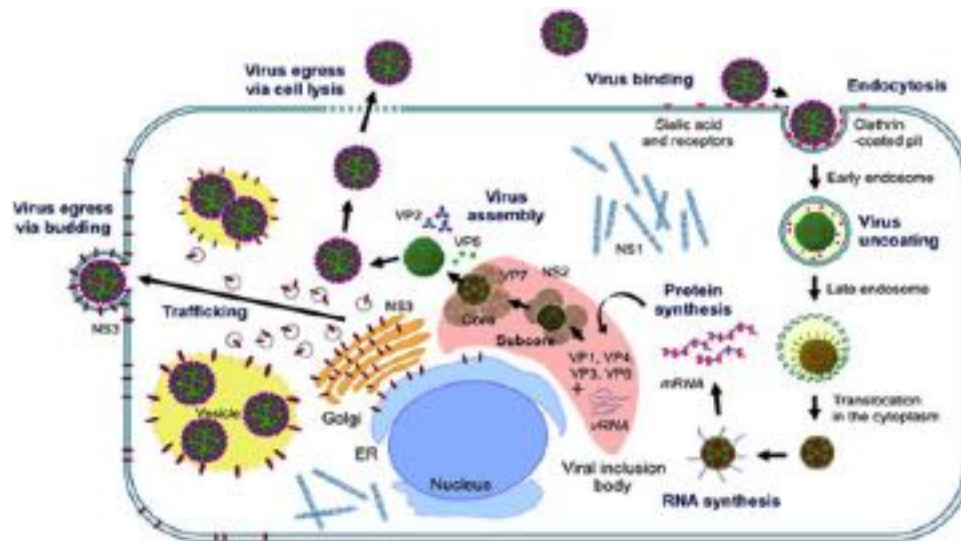


Fig. 2 Overview of BTV replication cycle. Virus entry occurs via sialic acid attachment by VP2 followed by clathrin-mediated endocytosis. VP2 senses the pH in the early endosome leading to its detachment. The core particle with exposed VP5 layer then moves to late endosome where VP5 senses the acidic pH leading to a conformational change and membrane permeabilization. Core is then released into the cytosol. Transcription and translation of viral proteins occurs leading to formation of tubules (NS1) and viral inclusion body (NS2) where components for core assembly are concentrated. Assembled core particles leave the VIB, leading to maturation by addition of the outer capsid proteins VP5 and VP2. The particles are trafficked using exocytic vesicles mediated by NS3 interaction with host factors leading to egress via budding or via host cell lysis.

the putative membrane-interaction element leading to the dramatic refolding of the dagger domain and the helical unfurling domain to form a new filamentous barb-like structure that protrudes out while the anchoring domain remains attached to the VP7 layer below. It is likely that the filamentous structures interact with the endosomal membrane and initiate permeabilization and destabilization of the endosomal vesicles, facilitating the release of the highly stable core, containing the dsRNA genome and the viral replication complex, into the host cytoplasm.

Transcription and Replication

Once released into the host cytoplasm the transcriptionally active core remains intact and initiates replication of the viral genome. Single-stranded RNAs (ssRNAs) from the 10 genomic segments are repeatedly transcribed by the enzymes within the core and released into the host cell cytoplasm through the channels that connect the interior of the core with the exterior. The released transcripts (mRNAs) are capped at the 5'-end but lack the canonical 3'-poly-A, and can serve as templates for both translation and for negative-strand RNA synthesis to form new genomic dsRNA segments.

The advent of a BTV reverse genetics (RG) system has enabled the dissection of genome replication within mammalian cells, which was found to occur in two stages. In the first, newly synthesized transcripts initiate the primary replication cycle by generating the primary replicase complex (PRC) consisting of the enzymatic minor proteins, VP1, VP4, and VP6, the inner core protein VP3, as well as two non-structural proteins, NS1 and NS2. NS1 is required for viral protein synthesis, whilst NS2 plays an essential role in selectively recruiting the four subcore proteins and the ssRNAs during particle formation (discussed below). In the second step, 10 ssRNAs segments are recruited to the PRC, forming the subcore. The viral replicase complex, composed of VP1, VP4, and VP6, then replicates the viral ssRNA to form dsRNA.

VP1 (150 kDa) is the largest of the BTV proteins and has shown RNA dependent RNA polymerase activity in vitro. Recombinant VP1 is capable of *de novo* dsRNA synthesis when supplied with BTV or Rotavirus mRNAs. However, it cannot synthesize dsRNAs from non-viral RNAs unless they are fused with BTV specific 5' and 3' sequences, suggesting that the recognition of a specific sequence or secondary structure element is essential for the initiation of RNA replication. No high-resolution structures are available for VP1. However, molecular modeling of the sequence based on the known Reovirus λ 3 structure, suggests that VP1 has the typical RdRp organization with an N-terminal domain, a central polymerase domain, and a C-terminal domain.

VP1 lacks the ability to cap the dsRNAs produced. This activity is instead performed by VP4. VP4 (76 kDa) functions as a single capping protein and contains all the necessary domains to catalyze the reactions required to generate type-1-like cap structures on ssRNAs in vitro, similar to those found on BTV mRNAs. VP4 was the first capping enzyme within the *Reoviridae* family to have its structure solved by X-ray crystallography. This revealed a unique multi-domain structure with an hourglass morphology with defined domains arranged in a linear layout (Fig. 3). While the atomic structure of VP4 did not account for its RNA-triphosphatase function, it provided the molecular basis for its guanylyltransferase and two methyltransferase activities (guanine-N7-methyltransferase, N7MTase and nucleoside-2-O-methyltransferase, 2'OMTase) activities characterized by in vitro experiments previously.

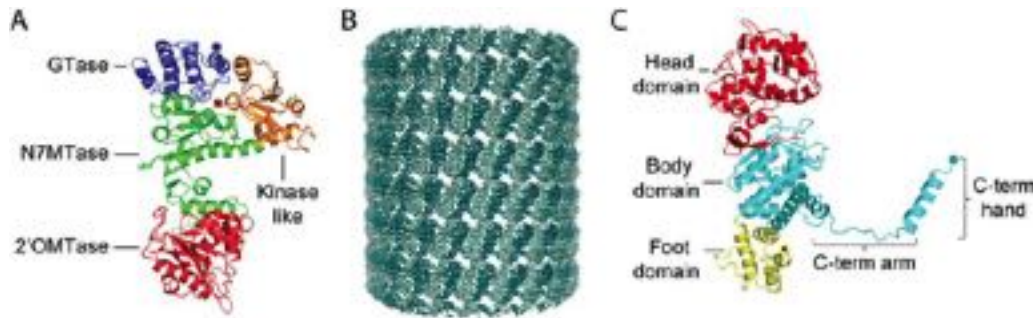


Fig. 3 Structure of BTV VP4 and NS1. (A) X-ray crystal structure of VP4 showing its domain organization. (B) Cryo-EM density reconstruction of NS1 tubule structure. (C) Atomic model of NS1 monomer showing the different domains.

VP6 (36 kDa) is the smallest of the proteins in the replication complex. *In vitro* analysis of recombinant VP6 suggests that it possesses helicase and nucleotide triphosphatase (NTase) activities and has high affinity for both ssRNAs and dsRNAs. Recent studies have shown that VP6 is essential for viral RNA replication and genome packaging. This attribute of VP6 has been utilized in the development of new replication defective vaccine strains, named 'ECRA' (Entry Competent Replication Abortive) vaccines (discussed below). Sedimentation analysis, size exclusion chromatography and EM analysis suggest that the VP6 forms hexamers with ring-like structures similar to those observed for other DNA and RNA helicases. While the VP6 structure is yet to be fully characterized, Nuclear Magnetic Resonance (NMR) analysis in 2014 suggested the presence of two large dynamic loops.

In tissue culture, the first BTV-specific proteins are detectable 2 h post infection and the rate of protein synthesis increases rapidly until 12–13 h post infection, after which it slows but continues until cell death. The two NS proteins, NS1 (64 kDa) and NS2 (41 kDa), involved in primary replication cycle, are synthesized abundantly early in BTV infection forming two virus-specific intracellular structures, tubules, and viral inclusion bodies (VIBs), respectively. These two structures are the hallmarks of BTV infection. By contrast, synthesis of NS3 and NS3A (26 and 25 kDa) varies from barely detectable to highly expressed, depending on the host cells.

NS1 has been suggested to positively influence the expression of BTV proteins and initial cryoEM analysis suggested that its tubules are made up of helically coiled NS1 dimers. Near-atomic resolution structure of two NS1 tubular forms has recently been obtained by cryoEM and the structure revealed that the same NS1 monomer gives rise to tubules with different helical assemblies and varying diameter (500 Å to 580 Å). Each NS1 monomer structure consists of an N-terminal foot domain, a head domain, and a body domain connected to an extended C-terminal arm (Fig. 3). The C-terminal arm of an NS1 subunit wraps above the head domain of another NS1 subunit through hydrophobic interactions in the tubular form and deletion of this C-terminal arm perturbs tubule formation. It does not, however, inhibit virus recovery and NS1 is still functional, indicating that NS1 tubules are not necessary for mRNA translation. The structure also revealed the presence of two metal-bound zinc-finger motifs in the NS1 monomers. Biological analysis confirmed that while both are needed for tubule assembly, only one is involved in the regulatory function of NS1 during replication.

NS2 is a phosphoprotein and is synthesized abundantly throughout infection. Initially distributed throughout the cytoplasm of infected cells, NS2 rapidly accumulates to form dense globular structures, known as virus assembly factories or virus inclusion bodies (VIBs). NS2 has a strong affinity for ssRNAs and viral ssRNAs accumulate within the VIBs soon after their formation. NS2 is phosphorylated at serine 249 and serine 259 by casein kinase II (CKII) and mutations of either one or both the serine residues influence VIB morphology. More recently, it has been shown that protein phosphatase 2A dephosphorylates NS2 and along with CKII modulates the phosphorylation status of NS2 in infected cells thereby regulating VIB morphology and virus replication. The N-terminal region of NS2 (160aa) was shown to have a β -sandwich motif in its monomeric form and forms homo-multimers composed of 10–11 monomers. The VIBs are the site of virus assembly and act as a scaffold for subcore formation. They contain viral RNAs in abundance along with the other viral proteins required for genome packaging, replication, and core assembly. Upon formation of the subcore in the VIBs, VP7 is added onto the VP3 layer to form the stable mature core particle.

The viral core is then released into the cytoplasm, where the outer capsid proteins (VP2 and VP5) are layered on to the core surface. The mature progeny virions are then released from the host cell. NS3/NS3A, which is the only glycosylated transmembrane protein of BTV, traffics through the endoplasmic reticulum and Golgi apparatus, and is involved in virus budding at the plasma membrane (discussed below), while the smallest protein, NS4 (77–79 residues), acts as an interferon antagonist and appears to be a virulence factor in sheep.

Core Assembly

NS2 plays a major role in recruiting each of the core components including the ssRNA transcript to facilitate the entire core assembly process within the VIBs. However, since core-like and subcore-like particles can be assembled *in vitro* by expressing the recombinant structural proteins, it has been possible to define the assembly pathway of VP3, from dimer to VP3 layer via the decamer intermediates and to show how 260 pre-formed VP7 trimers are layered onto the VP3 subcore. Recombinant core assembly data has further been supported by an *in vitro* cell-free assembly (CFA) system, which provided additional insights into

the entire core assembly process. In this assay, in vitro translated VP1, VP4, and VP6 were initially incubated with the ten ssRNA transcripts, followed by sequential addition of VP3 and VP7 to reconstitute a subcore and core structures with all ten ssRNA segments. The subcores were sensitive to RNase treatment prior to addition of VP7, suggesting that VP7 stabilizes the core. Furthermore, the assay also revealed an essential role for VP6 in genome packaging. In the presence of triphosphates the packaged ssRNA could serve as template for dsRNAs synthesis. Most importantly, in vitro reconstituted core particles were infectious in BTV permissive cells. Thus, NS2 was dispensable for core assembly in vitro suggesting its role in vivo is passive, acting as a condenser of the necessary components required for assembly.

Assembly of Genomic RNA and Packaging

The mechanism by which BTV and related viruses with multipartite genomes package their genomes into capsids has proved elusive until very recently. The ten BTV RNA segments (3.95–0.8 kb) vary in sequence but share common complementary short 5' and 3' untranslated regions (UTRs) of variable length, which include highly conserved hexanucleotides at both ends. Recent evidence has shown that the sorting of BTV RNA segments starts by the recruitment of genomic ssRNAs via these UTRs. Further, intersegment interactions leading to complex RNA networks are the driving force for sorting and packaging of a correct set of genomic segments. There is a specific order of segment recruitment, the smallest segment, S10, triggers RNA-RNA interaction with other three small segments and forms the initial complex. The other six larger ssRNA segments are subsequently recruited in a size dependent manner to form the final complex. Highly specific segment assorted signals (SAS) dispersed across the ssRNA segments are responsible for driving the specific interactions between RNA segments and thereby promoting RNA complex formation. Perturbation of these SAS by mutagenesis or by the introduction of specific competitive oligoribonucleotides, disrupts the RNA assembly process and affects RNA packaging in vivo. These studies indicate how selective packaging is achieved for BTV and suggest that similar mechanisms may apply to other members of the *Reoviridae* family.

Virion Maturation and Egress

Upon core assembly and dsRNA synthesis, nascent cores dissociate from the VIBs and the outer capsid proteins VP2 and VP5 are rapidly recruited to the core generating the mature virus particles. The precise mechanism behind the recruitment of VP2 and VP5 is not yet known, but it is likely that host factors are involved, VP5 with lipid raft and VP2 with vimentin and, in addition, both proteins have been found to concentrate on the surface of intracellular vesicles in preparation for their recruitment.

The release of the mature virus particles can be either lytic or non-lytic, with the host cell influencing the mechanism of egress. Infection of mammalian cells results in extensive cytopathic effect, such that the egress in these cells is predominantly by cell lysis although non-lytic budding has been also documented. In contrast, infected insect cells show little or no cytopathic effect and infected cells are characterized by continual non-lytic budding of nascent virus.

Non-lytic egress is primarily mediated by the viral glycoprotein NS3/3A, which is predominately found anchored to the plasma membranes and associated with the intracellular vesicles together with VP2 and VP5. It has been predicted that this protein forms homo-oligomers in infected cells and possesses a viroporin activity. The NS3/3A has two transmembrane domains with a short extracellular domain between them and the N and C termini project into the cytoplasm. The short C-terminal domain interacts with nascent viral particles via the outermost structural protein VP2, while the N-terminal domain of NS3/3A is capable of recruiting several host components of trafficking and exocytic pathways as well as the endosomal sorting complex required for transport (ESCRT) pathway. At the plasma membrane, NS3 directly binds to the membrane and facilitates the budding process. It has been shown that the NS3 protein is present in the envelope surrounding budding particles, and in cholesterol enriched membrane patches in close association with VP5 protein that also possessed an autonomous membrane-targeting signal. Finally, NS3 recruits TSG101, another ESCRT component, which eventually promotes scission of the budding viruses. Together these cellular factors are thought to traffic NS3/3A, along with the bound virus, to the plasma membrane. Thus, NS3/3A can be considered a bridge linking the virus particle to essential cellular trafficking machinery.

The importance of NS3/3A to non-lytic egress was highlighted in studies using an NS3/3A defective virus. These viruses remained viable in mammalian cells, although with less efficiency, however, they were dramatically attenuated in insect cells. Altogether, these findings describe a striking role for the NS3 protein in mediating viral egress, using mechanisms similar to those described for enveloped viruses.

Epidemiology

Bluetongue virus alternates its infectious cycle between the insect *Culicoides* vector and ruminant mammals. In many countries, the prevalence of BTV infection is high in ruminants although clinical disease is often not recorded. Due to the presence of *Culicoides* midges, the disease is prevalent throughout the tropics and subtropics, within separate ecosystems. The chance of infection depends in part on the genotype of the midge, the strain of virus, the level of viremia, and environmental factors. The epidemiology of bluetongue disease can be considered in the context of three major zones of infection: an endemic zone (generally subclinical infection), an epidemic zone (disease occurs at regular intervals), and an incursive zone (disease occurs usually only at extended intervals, but when it does occur, the epidemic is often extensive). Seasonal incursions of the virus into more temperate latitudes, sometimes accompanied by

disease, may occur under favorable climatic conditions at certain locations. Interestingly, cases of major epidemics in regions where disease occurs only sporadically, have often been traced to windborne transport of infected *Culicoides* from distant areas.

Many BTV strains have been isolated from different sources since its first characterization in the early 1900s. Initial cross-immunity studies of multiple BTV strains and subsequent serum-virus neutralization tests confirmed the existence of multiple serotypes of BTV. The complete virion particle was purified in 1969, and subsequent gel electrophoresis analysis showed that the BTV genome contained 10 discrete dsRNA segments. The genetic diversity and variation in sequences of different genome segments within a single BTV serotype was identified initially by RNA oligonucleotide fingerprint analyses of BTV field samples, which demonstrated that both genetic drift and genetic shift contribute to BTV evolution. Nevertheless, some segments were clearly less variable within a single serotype, but more variable between types, than others (e.g., segments encoding outer capsid proteins). Further, field isolates from both animals and insect vectors consisted of mixtures of RNA segments from populations of circulating virus strains. High-frequency of genome segment reassortment was shown to occur readily between different BTV serotypes and within other orbivirus serogroups but not between viruses belonging to different serogroups.

Early molecular hybridization studies demonstrated that genome segments 2 and 5 (encoding the two outer capsid proteins) exhibited little or no cross-hybridization between serotypes, whereas the other eight RNA segments gave consistently higher hybridization signals. Thus, genome segment reassortment may be an important factor in the generation of genetic diversity in orbivirus populations in nature.

These results were initially confirmed by sequence comparisons of segments encoding VP5 and VP2 from six different serotypes in the 1980s. This was further affirmed by recent extensive phylogenetic analyses of more than 200 different isolates of different serotypes. The accumulating data suggested that while VP5 showed up to 70% sequence identity, VP2 sequences varied from 22% to 73% between serotypes. While BTV isolates can be separated into twenty eight distinct serotypes on the basis of genome segment 2, localized circulation of the virus in different regions throughout the world has led to evolution of distinct geographical variants broadly divided into Eastern (endemic in Asia, Indonesia and Australia) and Western lineages (endemic in Africa, the Caribbean and the Americas). Serotypes have additionally been classified into distinct nucleotypes based on phylogenetic analysis of segment-2 (VP2) and segment 5 (VP5). VP2 was classified into 12 nucleotypes: A-L that correlates with BTV serotypes while VP5 was classified into 10 nucleotypes: A-J and have partial correlation with BTV serotypes.

The global distribution of different BTV serotypes has undergone drastic alteration over the recent years. Changes in climate and increased travel and trade have been implicated in the global distribution of BTV, and of *Culicoides* species susceptible to BTV infection. Molecular epidemiology studies indicate that individual strains can move large distances due to the windborne movements of infected insects and that a single bite from a fully infected adult *Culicoides* is sufficient to reliably transmit the virus to a susceptible mammalian host.

Many serotypes of BTV previously not present, have now been detected in Europe, North America, Australia, Korea, India, the Middle East, and South America, with more than one BTV serotype identified in many regions. Particularly in Europe, the epidemiology of BTV has changed dramatically since 1998, when six different BTV serotypes emerged across southern European countries. Most of these serotypes have persisted and spread both westwards and northwards, providing evidence for an epidemiological 'step-change'.

By 2002, different BTV serotypes were identified in Greece, Italy, Bulgaria, Kosovo, Serbia, Montenegro, Macedonia, Croatia, Albania, and Bosnia and Herzegovina. Several serotypes gradually spread north and in particular, a BTV-8 epidemic broke out in northern Europe in 2006 with the first case being reported in the Netherlands. Outbreaks of BTV-8 were then reported in Belgium, Germany, and Northern France although the exact mode and mechanism of these incursions remains unclear. BTV-8 subsequently spread across southern European countries and further north into Luxembourg, the Czech Republic, Switzerland, Denmark, and the UK. New outbreaks were also reported in Hungary, Austria, Sweden, and Norway. In 2008, BTV-6 was identified in Germany and the Netherlands and BTV-11 was identified in Belgium. The high number of cases reported for the BTV-8 outbreak in Belgium, Germany, Netherlands, France, and Luxembourg, led to the implementation of a vaccination strategy against the serotype across several countries in Europe, and the outbreak was finally controlled in 2009.

Pathogenesis and Clinical Features

BTV is transmitted to vertebrate hosts through blood-feeding midges of the *Culicoides* genus. BTV infection has an incubation period of 4–12 days and produces a spectrum of conditions from sub-clinical infection to severe and fatal disease, depending on virus strain and host species. Clinical signs of infection are generally more severe in sheep, resulting in severe illness and high mortality (as high as 70%). The insect vectors carry the virus in their salivary glands and ducts and infect the host during a blood meal with virus inoculation at the site of biting. The regional lymphatic system underlying the infected skin laceration mediates the dissemination of the virus to the endothelial cells, and peripheral blood mononuclear phagocytes (PBMs). The virus proliferates in the PBMs and lymph nodes and is further disseminated via the blood stream to sites of secondary infection. In the final phase of infection, the virus begins circulating in the bloodstream. Viral RNA can persist for several months. BTV binds to glycoporphins on the surface of bovine and ovine erythrocytes where it may remain in an infectious state as invaginations of the erythrocyte cell membrane. This minimizes contact with antibody and T cells and thus provides multiple opportunities for transmission by infection of blood-sucking midges.

In sheep, viremia is usually detectable around 3–5 days post infection. During the initial stages the virus associates with both platelets and erythrocytes, however, in the late stages of infection, the virus is associated exclusively with erythrocytes, facilitating

the prolonged infection of ruminants and increasing transmission to the insect vectors that feed on them. BTV infection of mammalian cell lines induces cell death, either by triggering multiple apoptotic pathways or via necrosis. BTV replication in endothelial cells causes cell injury and infected ruminants develop lesions due to mononuclear cell infiltration. The accumulation of large acidophilic intra-cytoplasmic masses which, results in degeneration and necrosis. Infected ruminants suffer severe systemic damage, characterized by widespread edema and hemorrhage. In infected sheep with severe disease, swelling of the tongue and lack of oxygen make the tongue and mucous membranes appear blue (hence the name of the disease). Infection of pregnant cattle and sheep with BTV can result in maternal death, abortion, and fetal death, or congenital anomalies. BTV infection in most wild ruminant species is typically asymptomatic or subclinical, although in some cases severe disease in cattle and camelids has been described. Infected cattle and goats sporadically display clinical signs including nasal discharge, swelling, and, ulceration of the mouth.

Diagnosis

Since bluetongue is a notifiable disease, animals diagnosed with clinical signs of bluetongue disease and pathology are required to be reported to local health officials. BTV infection is confirmed by virus isolation, detection, and characterization of the infecting serotype. Earlier methods relied primarily on serological tests to detect BTV antibodies in infected animals. Complement fixation, agar gel immunodiffusion (AGID) and haemagglutination-inhibition relied on virus/serum neutralization methods that enabled virus serotype identification based on the interaction of the outer capsid protein VP2 with serotype-specific antibodies. Among the various serogroup-specific assays for antibody detection, AGID and competitive enzyme-linked immunosorbent assay (cELISA) are the most widely used. However, although the AGID test is simple to perform and rapid, it is not highly sensitive or quantitative and has limitations in its specificity. In contrast, the cELISA, in which a group-specific monoclonal antibody based on BTV core protein VP7 is used, has proved to be the most sensitive and specific assay for detection of antibodies to most, if not all, serotypes and strains of BTV. Following extensive national and international validation, the cELISA became a universal test to certify ruminants for trade purposes and to diagnose BT infection in domestic and wild animals. The test is highly specific and sensitive detecting antibodies to most, if not all, serotypes and strains of BTV. For the detection of serotype-specific antibodies in animal sera, the cell culture-based microtiter serum neutralization (MTSN) is the most commonly used assay. MTSN is often used to monitor animal population for specific serotypes of BTV in epidemiological investigations.

Recent advances in molecular and high-throughput sequencing technologies have enabled rapid identification of newly isolated strains by reverse transcription PCR (RT-PCR) assay, real-time RT-PCR, or sequencing specific segments that are highly variable across the different isolates. RT-PCR assays employing serotype-specific primers have provided the most rapid and specific information regarding isolate serotype. Selective sequencing of virus segments or whole genome sequencing provides serotype as well as other unique sequence information of isolates. These new approaches have enabled rapid and reliable confirmation of BTV serotype, and are potential methods that should be applied at the onset of an outbreak to allow for the early selection of vaccine.

Vaccination and Control

Vaccination strategies in different countries have been developed according to their individual policies, the geographic distribution of the circulating BTV serotypes and the availability of appropriate vaccines. In South Africa, the country where bluetongue was first characterized, live attenuated virus vaccines have been used for over 50 years and are known to induce an effective and lasting immunity. Due to the high number of circulating serotypes however, live modified vaccines often have to be administered as a multivalent vaccine. Despite their success in endemic areas, these attenuated virus vaccines have exhibited some drawbacks, including abortion/embryonic death and teratogenesis in fetuses of pregnant females. In addition, the strength of the protective antibody response to attenuated vaccines correlates directly with their ability to replicate in vaccinated animals and when responses are sub-optimal viremia is observed that is sufficient for the vaccine strain to be transmitted to biting midges. Furthermore, there is a likelihood of reversion to virulence and/or reassortment of the vaccine strain with wild type viruses.

In recent years, chemically inactivated BTV vaccines were used for mass vaccination of sheep and cattle in European countries during the BTV-8 and BTV-1 outbreaks. While these inactivated virus vaccines have been demonstrated to be safe and immunogenic in animals, they are only available for a limited number of serotypes. Additionally, in some cases adverse reactions have been associated with some of these inactivated virus preparations. Inactivated vaccines generally induce relatively transient immunity in comparison to live, attenuated vaccines and require booster immunizations. The high cost of vaccine production due to the need for large amounts of antigen required for each vaccination is another major drawback.

The biggest limitation to the use of both the live attenuated and inactivated vaccines is the significant economic impact they have on agricultural trade. These vaccines do not allow the differentiation of naturally infected versus vaccinated animals (DIVA) as these vaccine strains are not free of viral protein components that interfere with serological tests. Thus, neither of these vaccines is DIVA compliant and restrictions are imposed on animal trade and movement if an animal is diagnosed positive, even if that positivity is the result of vaccination. Further, due to the presence of genomic RNA after vaccination and short-term BTV viremia in animals in the case of the live attenuated vaccine, a genetic DIVA is not always feasible.

A number of recombinant protein technologies have been employed in order to provide DIVA compliant vaccine candidates and to reduce the potential adverse effects associated with inactivated and attenuated virus vaccines. Recombinant proteins have

predominantly been assessed in sheep in experimental vaccination challenge trials and shown to afford protection, but have yet to be commercialized. Recombinant outer capsid protein VP2, the serotype determining protein, could indeed elicit protective neutralizing antibody in vaccinated animals and later studies, which combined both outer capsid proteins (VP2 + VP5) resulted in better protection than VP2 alone in vaccinated sheep. In addition, VP2-based vaccines have been developed using other live virus vector systems (i.e., poxviruses, herpesvirus, and adenovirus), which have been designed to express either a single or multiple proteins in animals following vaccination.

Enhanced protective immunity afforded by recombinant VP2 + VP5 led to the development of recombinant virus-like particles (VLPs) containing the VP2 + VP5 and the two core proteins (VP3 and VP7), mimicking the virus particles without the viral genome or enzymatic proteins. The VLPs had the characteristic icosahedral structure with the four recombinant proteins in the same position and ratios as the native proteins. Multiple vaccination studies with homologous and heterologous VLPs (~10 µg) resulted in complete protection against virulent virus challenges in sheep. In recent years, extensive vaccination trials with VLPs of European serotypes had been undertaken in European breeds of sheep. These trials have also assessed VLPs delivered singly or in cocktail, as well as the evolutionary lineage of the virus strains, in order to assess their potential for commercialization.

The influence of evolutionary divergence (Eastern and Western topotypes) on vaccine development has clearly highlighted that the development of serotype specific neutralizing antibodies was more critical than the topotype of the circulating strain. This suggested that isolate variation or evolutionary distance of the strain is not critical for BTV vaccines. Furthermore, the delivery of VLPs as a cocktail of different serotypes did not limit the ability of each serotype to raise a serotype specific neutralizing antibody response or to interfere with the protective efficacy of the vaccine when animals were challenged. These findings have indicated that multiple serotype vaccines could be developed using known serotypes rather than generating new vaccines for each BTV outbreak. Most importantly, it is possible to distinguish between VLP-vaccinated animals and those that are infected with the virus, thus addressing one of the major problems with current vaccines. Further, VLPs are entirely protein-based and there is no genetic component avoiding all issues of reversion or recombination.

The use of the BTV reverse genetics system has further allowed direct introduction of mutations/deletions in the viral genome leading to the recovery of replication-defective virus strains using specific complementary cell lines. Entry Competent Replication Abortive (ECRA) vaccines, formerly known as Disabled Infectious Single-Cycle (DISC) vaccines were designed based on this principle. The ECRA vaccines are modified particles that contain only 9 of the 10 genome segments, they lack the essential catalytic gene VP6, and are hence replication incompetent. The defective virus can enter susceptible cells where they undergo a single cycle of replication but fail to make progeny virus due to the inability to package the dsRNA genome. Administration of ECRA vaccine results in expression of the viral proteins at the site of infection but does not cause infection in animals and hence combines the efficacy of live-attenuated vaccines with the safety of inactivated vaccines. Recent trials in cattle and sheep have shown promising results. A single dose of a DISC virus strain was sufficient to confer complete protection against a challenge in sheep 42 days after vaccination while a cocktail of six different vaccines protected animals as early as 21 days against any challenge with virulent strains. The animals were completely protected from challenge with individual virulent serotypes, even after 5 months of challenge without any clinical disease. ECRA vaccines are more cost-effective and since they are replication incompetent, the ECRA RNA segments are present in the bloodstream for a very short period thus minimizing the risk of transmission via insect vector, reassortment, or reversion to virulence.

A second RG based candidate vaccine strain has been made against BTV known as Disabled Infectious Single-Animal (DISA), in which segment 10 is deleted, encoding NS3/NS3A. Unlike the ECRA strains, the DISA strain still replicates in normal cells and in sheep although it is significantly attenuated. The DISA vaccine strain is also DIVA compliant and is another promising candidate vaccine although rigorous animal studies are still lacking.

While a completely new range of different vaccine candidates, showing promising levels of efficacy and safety, have been developed recently, none of them are near commercial production. The voluntary vaccination policy adopted in Europe and the acceptance of endemic disease in many parts of the world has resulted in a dearth of commercial organizations interested in producing these new vaccines. Given the likelihood of new BTV outbreaks across Europe, there is however, a need to develop suitable vaccines that can be produced efficiently as vaccination is the best way to control outbreaks.

Future Perspectives

The advances in understanding BTV, at all levels, has led to it being one of the most well characterized RNA viruses, including the details related to its replication in both mammalian and insect hosts. Physically the resolution of the virion structure has increased substantially and the atomic detail of nearly all of the particle is now known. This level of understanding has provided insights on how the particle disrobes on entry and crosses the membrane to enter the cytoplasm. We also have a greater understanding of how transcription and replication proceed. Importantly, key insight has been obtained on how multiple segmented dsRNA genomes are incorporated into assembling virions, while cellular messenger RNAs are excluded. The role of cellular factors and virus encoded proteins in leading the virus out of the cell has never been better described.

Yet despite this phenomenal progress, much remains to be learned. The structure of the particle, while instructive, does not represent the dynamic aspects of the capsid, that is less understood. The level of “breathing and the triggers” for channel opening to allow the initiation of transcription of the assembled particles are still unclear. How VP5 leads to membrane destabilization is also not well understood. BTV is a large particle and it is difficult to envisage it crossing a membrane barrier with ease.

The dynamic nature of VIB is another area of uncertainty and while its description as a virus factory is true, it does not address how exactly all BTV components coalesce there or how, after assembly, the newly formed virions leave this amorphous mass in an orderly fashion. What are the triggers and timings of these processes and do they pertain to all cell types or only to some? Virions acquire their outer layer as it migrates the cell to the exterior. A role for an “enveloped” BTV has been proposed, but detail is sketchy. Not least, the astonishing difference in CPE observed in mammalian and insect cells is something that requires a much more work.

Behind all experimental approaches, molecular, structural or immunological, is the desire to see new findings translated into improved options for disease control. Here, recent progress has been promising with some clever approaches to the generation of candidate vaccines that include DIVA compliance. Better yielding and genetically tighter variants will certainly be produced in the future opening the possibility that such intricately engineered vaccines may become the default for future outbreak control. Bluetongue is a disease with economic repercussions in both developed and developing countries. Hence, better control measures have the opportunity to improve also the quality of life for the livestock of subsistence farmers, which in turn can have knock-on benefits, to the communities that depend upon animal productivity.

Further Reading

- Belbis, G., Zientara, S., Bréard, E., *et al.*, 2017. Bluetongue virus: From BTV-1 to BTV-27. In: Beer, M., Höper, D. (Eds.), *Advances in Virus Research* 99. Academic Press, pp. 161–197.
- Bhattacharya, B., Roy, P., 2010. Role of lipids on entry and exit of bluetongue virus, a complex non-enveloped virus. *Viruses* 2, 1218–1235.
- Boyce, M., Celma, C.C.P., Roy, P., 2008. Development of reverse genetics systems for bluetongue virus: recovery of infectious virus from synthetic RNA transcripts. *Journal of Virology* 82, 8339–8348.
- Celma, C.C.P., Stewart, M., Wernike, K., *et al.*, 2016. Replication-deficient particles: New insights into the next generation of bluetongue virus vaccines. *Journal of Virology* Dec 91 (1), e01892–16.
- Coetzee, P., Van Vuuren, M., Stokstad, M., Myrmel, M., Venter, E.H., 2012. Bluetongue virus genetic and phenotypic diversity: Towards identifying the molecular determinants that influence virulence and transmission potential. *Veterinary Microbiology* 161, 1–12.
- Huisman, H., 1979. Protein synthesis in bluetongue virus-infected cells. *Virology* 92, 385–396.
- Janowicz, A., Caporale, M., Shaw, A., *et al.*, 2015. Multiple genome segments determine virulence of bluetongue virus serotype 8. *Journal of Virology* 89 (10), 5238–5249.
- Lourenco, S., Roy, P., 2011. In vitro reconstitution of bluetongue virus infectious cores. *Proceedings of the National Academy of Sciences* 108, 13746–13751.
- Maan, S., Maan, N.S., Samuel, A.R., *et al.*, 2007. Analysis and phylogenetic comparisons of full-length VP2 genes of the 24 bluetongue virus serotypes. *Journal of General Virology* 88, 621–630.
- MacLachlan, N.J., Drew, C.P., Darpel, K.E., Worwa, G., 2009. The pathology and pathogenesis of bluetongue. *Journal of Comparative Pathology* 141, 1–16.
- Matsuo, E., Roy, P., 2013. Minimum requirements for bluetongue virus primary replication in vivo. *Journal of Virology* 87, 882–889.
- Nomikou, K., Hughes, J., Wash, R., *et al.*, 2015. Widespread reassortment shapes the evolution and epidemiology of bluetongue virus following European invasion. *PLOS Pathogens* 11 (8), e1005056.
- Roy, P., 2017. Bluetongue virus structure and assembly. *Current Opinion in Virology* 24, 115–123.
- Shaw, A.E., Hughes, J., Gu, Q., *et al.*, 2017. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PloS Biol* 15, e2004086.
- Sung, P.Y., Roy, P., 2014. Sequential packaging of RNA genomic segments during the assembly of Bluetongue virus. *Nucleic Acids Research* 42, 13824–13838.
- Zhang, X., Patel, A., Celma, C.C.P., *et al.*, 2016. Atomic model of a nonenveloped virus reveals pH sensors for a coordinated process of cell entry. *Nature Structure Molecular Biology* 23, 74–80.

Borna Disease Virus and Related Bornaviruses (*Bornaviridae*)

Susan L Payne, Texas A&M University, College Station, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

Introduction

The name bornavirus derives from Borna disease, a rare neurologic disease that mainly effects horses and sheep. The disease presents as an array of behavioral changes including ataxia, head tilt, muscle fasciculation, hind limb paresis, localized hypo- or hyperesthesia, disturbances in chewing and swallowing, and aggression. Borna disease is endemic in a fairly small geographic area encompassing central and southern Germany and neighboring countries. The disease name is linked to two major outbreaks that occurred from 1894 to 1896, among cavalry horses near the town of Borna, Germany. The infectious nature of Borna disease was not confirmed until the early decades of the 20th century and it would take approximately another 50 years before persistently infected cell cultures were established. However, even then the infectious agent remained uncharacterized because persistently infected cultures did not provide sufficient cell-free virus for convincing biochemical or structural studies. Finally, in the early 1990s molecular studies revealed the similarities of Borna disease virus (BoDV-1) to other unsegmented, negative strand RNA viruses.

Classification

Until 2008 BoDV-1 was the sole member of the family *Bornaviridae*. In 2008 novel bornaviruses were identified as the etiologic agents of proventricular dilatation disease (PDD) of birds. Identification of these avian bornaviruses was accomplished by using nucleic acid hybridization and unbiased high throughput sequencing techniques. Since 2008, polymerase chain reaction (PCR) and unbiased nucleic acid sequencing methods have been used to identify additional bornaviruses. Currently the International Committee on the Taxonomy of Viruses Master Species List contains 10 species of bornavirus. Two species have been isolated from mammals (Mammalian 1 orthobornavirus and Mammalian 2 orthobornavirus), 5 from birds (Passeriform 1 orthobornavirus, Passeriform 2 orthobornavirus, Psittaciform 1 orthobornavirus, Psittaciform 2 orthobornavirus and Waterbird 1 orthobornavirus) and 3 from snakes (Elapid 1 orthobornavirus, Queensland carbovirus, and Southwest carbovirus).

Virus Structure

Bornavirus virions are enveloped and are assumed to have a helical nucleocapsid; their size is broadly estimated to be 80–130 nm in diameter. However, unlike many viruses, detailed ultrastructural analyses of bornavirus virions have not been performed, due to the highly cell-associated nature of the virus. The illustration of a bornavirus particle shown in [Fig. 1](#) is theoretical, based on relatively low resolution electron micrographs of virions and the known structures of other unsegmented negative strand RNA viruses.

Genome

The genome of bornaviruses is a single negative stranded molecule of approximately 8.9 kb in length. The genome organization of mammalian and avian bornaviruses is highly conserved ([Fig. 2](#)) and similar to other members of the order *Mononegavirales* ([Fig. 2](#)). Bornaviruses contain 6 open reading frames and short non-coding regions are found at the 5' and 3' ends of the genome. Bornavirus genomes lack the discrete intergenic sequences characteristic of the paramyxoviruses and rhabdoviruses. Instead, transcription termination sequences for upstream transcripts are found just downstream of initiation sequences for the next transcript.

A promoter near the 3' end of the BoDV-1 genome initiates the N transcript, the most abundant viral transcript. The N transcript is used to produce two isoforms of N, p40 and p38 by use of alternative start codons. The X/P transcript produces two proteins, X and the phosphoprotein P. X and P are produced from overlapping reading frames. P is a 24 kDa phosphoprotein that serves as a co-factor for L, the catalytic subunit of the transcription/replication complex. The 10 kDa X protein interacts with P and has role in RNA replication. X also associates with mitochondria and prevents induction of apoptosis in infected cells. A third transcription start site is used to produce the M, G and L transcripts; both G and L are spliced. M (p16) is found in both the cytoplasm and nucleus of infected cells. M co-localizes with N, P, and X in punctate structures in the nucleus. As shown in [Fig. 2](#), there are short overlaps of the M, G and L reading frames. The glycoprotein, G, is post-translationally modified by N-glycosylation to yield a 93-94-kDa primary product. G contains a proteolytic cleavage site and a portion of G is cleaved to produce amino terminal (G-N) and carboxyl terminal (G-C) products. Virus neutralization studies have demonstrated the importance of G in virus entry.

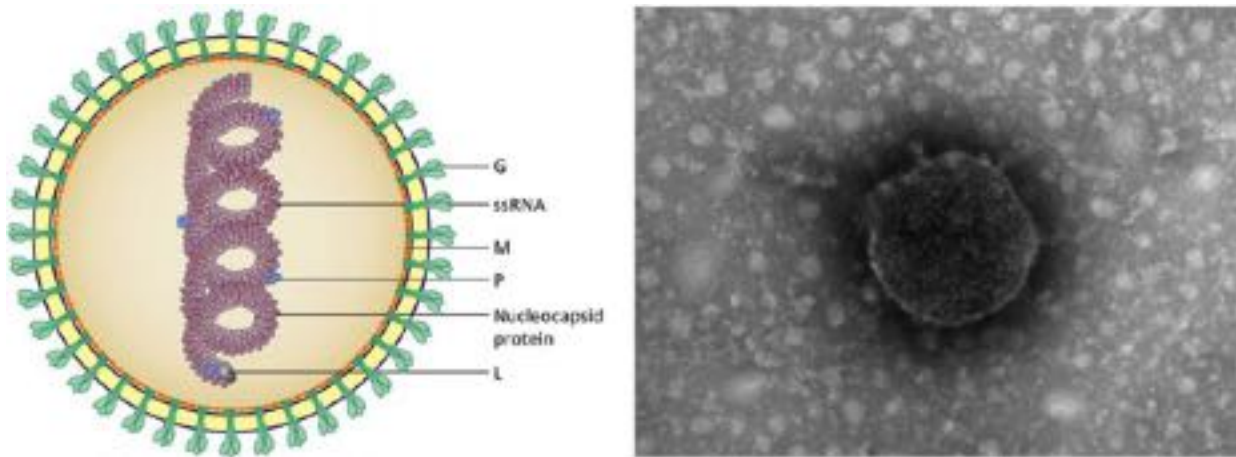


Fig. 1 Virion structure (left). Virus-like particle (83 nm in diameter) from the eye fluid of an Eclectus parrot with confirmed PDD and ABVO infection (right). From: Payne, S., 2017. Chapter 22 – Family Bornaviridae. In *Viruses* (edited by S. Payne), Academic Press, pp. 191-196, <https://doi.org/10.1016/B978-0-12-803109-4.00022-2>.

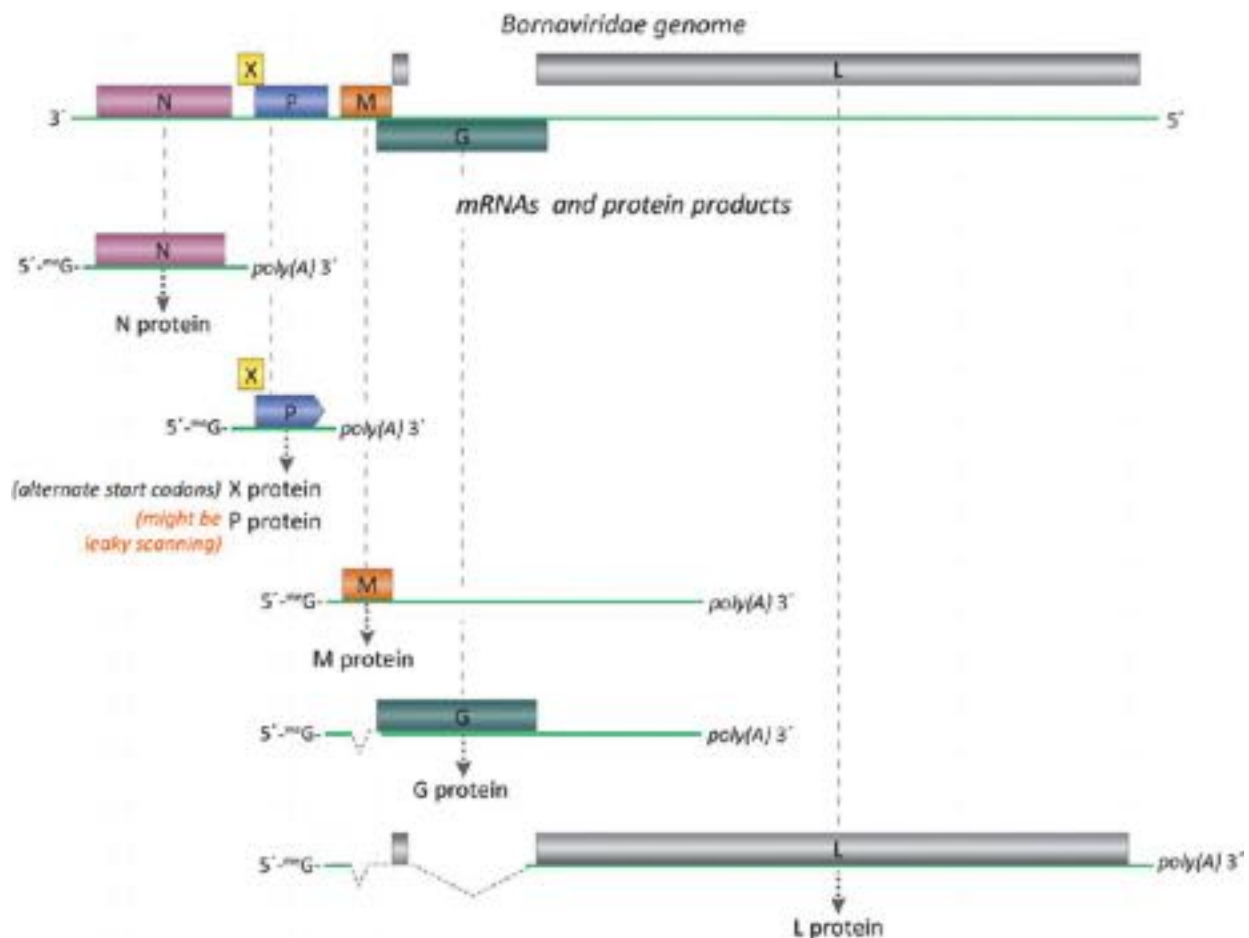


Fig. 2 Bornavirus genome organization and transcription pattern. From: Payne, S., 2017. Chapter 22 – Family Bornaviridae. In *Viruses* (edited by S. Payne), Academic Press, pp. 191-196, <https://doi.org/10.1016/B978-0-12-803109-4.00022-2>.

Life Cycle

Most molecular studies of bornaviruses replication have been carried out using BoDV-1, but it is likely that mechanisms of transcription and genome replication elucidated for BoDV-1 will hold true for the avian bornaviruses as well. Bornaviruses are highly cell-associated, however some infectious particles are released from infected cells. In the case of cell-free virus, bornavirus G is the candidate for receptor attachment and subsequent fusion, but host receptors have not been determined. Although the molecular details are not yet known, it is clear that the nucleocapsids traffic to the nucleus where transcription initiates. L catalyzes RNA synthesis and likely serves to cap and polyadenylate viral mRNAs. Translation of N, X, P, M and L are cytosolic, while G is processed in the endoplasmic reticulum. All viral proteins with the exception of G are transported back into the nucleus where genome replication and assembly of ribonucleoprotein particles (RNP) occur. Immunostaining of persistently infected cells shows large punctate structures in the nucleus. These so-called vSPOTS are virus replication factories with a size of ~66 µm in diameter. vSPOTS are cage-like assemblies of the N protein with 59–180 nm pores. Viral genomes and the proteins X, P, M and L are also found in vSPOTS, therefore these structures contain all of the components necessary to be sites of transcription and/or genome replication.

Filamentous viral RNPs, containing genomic RNA, N, P, X and L, associate with chromatin by using core histones as a docking platform. High mobility group box protein 1 (HMGB1), a chromatin-remodeling DNA architectural protein binds to viral P, stabilizes the RNP on chromosomes and appears to be required for efficient transcription. In mitotically active cells, the tight association of the RNP with chromatin facilitates transmission of RNPs to daughter cells. Within infected hosts, virus (in the form of RNPs) spreads long distances along neural networks to access the CNS. RNPs appear to move along axons in signaling endosomes. There is little to no free virus in the infected animal. In a cockatiel model of avian bornavirus infection, virus injected into pectoral muscle reached the spinal cord at about 20–25 days post-inoculation. From there virus spread quickly to the brain. From the CNS, virus spread to ganglia in the gastrointestinal tract, adrenal gland, heart, and kidneys. At late points of infection (80–100 days) viral antigen was also seen in smooth muscle and/or scattered epithelial cells of numerous tissues.

Epidemiology and Clinical Features

Borna Disease

The white-toothed shrew (*Crocidura leucodon*) is the animal reservoir for BoDV-1 in Borna disease endemic areas. The virus replicates in numerous tissues in the shrew and is excreted into the environment. In shrews the virus causes persistent infection without signs of disease. Borna disease, caused by infection with BoDV-1, is quite rare and predominantly affects horses and sheep in Central Germany. There is no evidence that the virus is naturally transmitted among infected livestock, therefore horses and sheep are 'accidental' and dead-end hosts.

Disease is often fatal. Signs include ataxia, head tilt, muscle fasciculation, hind limb paresis, localized hypo- or hyperesthesia, disturbances in chewing and swallowing, and aggression. Grazing animals, in particular horses and sheep, likely inhale excreted virus and are infected by the olfactory route. The incubation period last from weeks to months as virus moves from the initial site of infection to the central nervous system.

Staggering Disease in Cats

Beginning in the late-1970s, a neurologic disease characterized by stiffness, leading to paralysis and death was observed in cats in Central Europe and Scandinavia. Histopathology of affected cats revealed lymphohistiocytic inflammation of the CNS. In the 1990s BoDV-1 nucleic acids were isolated from some of cats with staggering disease, and challenge of pathogen-free cats with this virus induced neurologic disease. Staggering disease has been reported in cats world-wide and serologic surveys of healthy cats reveal that many are also positive for bornaviral antigens. Studies in Japan show approximately 25% of tested cats are positive for bornaviral antigens by Western blot assays. This is similar to findings in birds (see below) and underscores the fact that many BoDV-1 infections are asymptomatic. Epidemiology of bornavirus infection in cats in Japan points to vertical transmission, while reports from Europe suggest cats with outdoor access are at higher risk for developing staggering disease. For these cats, infection likely results from predation upon birds, rodents and/or other small mammals. In one European study, approximately 80% of cats with staggering disease tested positive for bornavirus-specific antibodies by immunofluorescence assays, compared to 16% of a healthy comparison group.

Bornavirus Associated Encephalitis in Humans

Beginning in the 1980s there were efforts to link human bornavirus infection and psychiatric disorders. Many of the early studies claiming positive links relied on poorly described serologic techniques and many other studies failed to show links. However, while the relationship between bornavirus infection and human psychiatric disease remains unclear, three recent case reports provide striking evidence that mammalian bornaviruses can cause severe, fatal neurologic disease in humans. In all, 6 human deaths have been strongly associated with bornaviral infection. The first case report of fatal human bornavirus infection involved

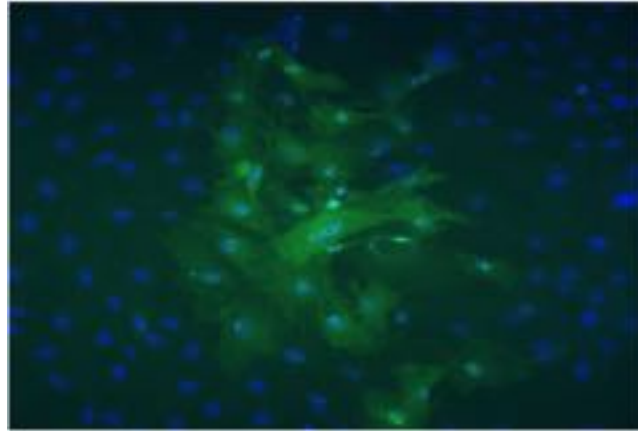


Fig. 3 Fluorescent antibody labeling of duck embryo fibroblasts infected with an avian bornavirus. A focus of infected cells (green) sits amidst uninfected cells. The uninfected cells are visualized by DAPI staining (blue) of their nuclei. From: Payne, S., 2017. Chapter 22 – Family Bornaviridae. In *Viruses* (edited by S. Payne), Academic Press, pp. 191-196, <https://doi.org/10.1016/B978-0-12-803109-4.00022-2>.

three breeders of squirrels in Germany that presented with encephalitis between 2011 and 2013. All three died within 2–4 months after onset of symptoms. After eliminating suspected causes of encephalitis, a genomic approach was used as a means of identifying novel pathogens. This approach revealed common bornaviral sequences in the brains of the three victims. Exchange of breeding squirrels linked the three human victims and an identical viral sequence was recovered from the brain of a variegated squirrel (*Sciurus variegatoides*). The squirrel bornavirus, (named variegated squirrel 1 bornavirus (VSBV-1) is phylogenetically distinct from BoDV-1 and is a member of the species Mammalian 2 orthobornavirus).

Another case report from 2018 underscores the danger of handling squirrels. This report describes a fatal case of limbic encephalitis in a zoo caretaker in Germany. Molecular assays and immunohistochemistry revealed a limbic distribution of viral antigen in brain tissue. Sequencing and phylogenetic analysis of the brain-associated virus demonstrated a likely spillover infection from a Prevost's squirrel (*Callosciurus prevostii*). Prompted by the human cases of VSBV-1, studies of squirrels have been carried out in order to assess the risk of infection. Small numbers of VSBV-1 infected squirrels have been detected among small breeders and zoos in Germany, the Netherlands and Croatia. The variegated squirrel is a native of southern Mexico and Central America while the Prevost's squirrel is a native of Asia, thus the origins and natural reservoirs of VSBV-1 are unknown.

Organ transplantation is another source of human bornavirus infection. A 2018 report presents strong evidence of BoDV-1 infection transmitted via solid organ transplant. As described in the case report, the organ donor was a 70 year old man from the Bavarian region of Germany who had no signs or symptoms of neurologic disease. Two recipients of kidneys developed neurologic disease that progressed to irreversible coma. A third patient, who received a liver graft, developed transient neurologic deficits but went into remission from the disease. A metagenomics analysis of brain tissues from the two deceased organ recipients was performed, revealing BoDV-1 sequences. Immunohistochemical analysis also revealed bornaviral antigens in brain tissues. Antibodies to BoDV-1 were detected in the third organ recipient. It is interesting to note that to date, these well-documented cases of bornavirus-associated encephalitis in humans have occurred in Central Europe, in Borna disease endemic areas. It remains to be seen if additional cases of fatal human encephalitis will be linked to bornavirus infection in the future.

Proventricular Dilatation Disease of Birds

PDD is a fatal neurologic condition first recognized in the mid-1970s among captive parrots. A viral etiology was suspected early on, but the agent was not identified until 2008 when the genomes of novel bornaviruses were identified in the brains of affected birds. Infectious viruses were quickly recovered and used to produce PDD in experimental infections. While large parrots and macaws are the captive birds most likely to be diagnosed with PDD, the disease has also been described for other pet birds such as canary birds and finches. Clinical neurologic signs of PDD include weakness, ataxia, proprioceptive deficits, seizures, and blindness. Gastrointestinal malfunctions are also very common and result in severe weight loss, passage of undigested food, regurgitation and delayed crop emptying. Gastrointestinal malfunction results in the most common gross pathologic lesions associated with PDD – dilation and thinning of the walls of the proventriculus and ventriculus. Histopathologically, PDD is characterized by nonsuppurative inflammation in the central, peripheral, and autonomic nervous systems.

ABV infects primarily the central and enteric nervous systems of birds. In experimental models, virus is inoculated into muscle tissue and onset of disease is highly variable, but is typically from several weeks to months. Although the virus is noncytopathic examination of brain tissue reveals selective loss of glial cells and neurons. Based on the known pathogenesis of Borna disease, it is believed that cell loss is due to T cell cytotoxicity rather than direct cytopathic effects of the virus. Treatment of affected birds with immunosuppressive agents can suppress or delay disease development in experimental models, supporting an immunopathologic basis for the disease.

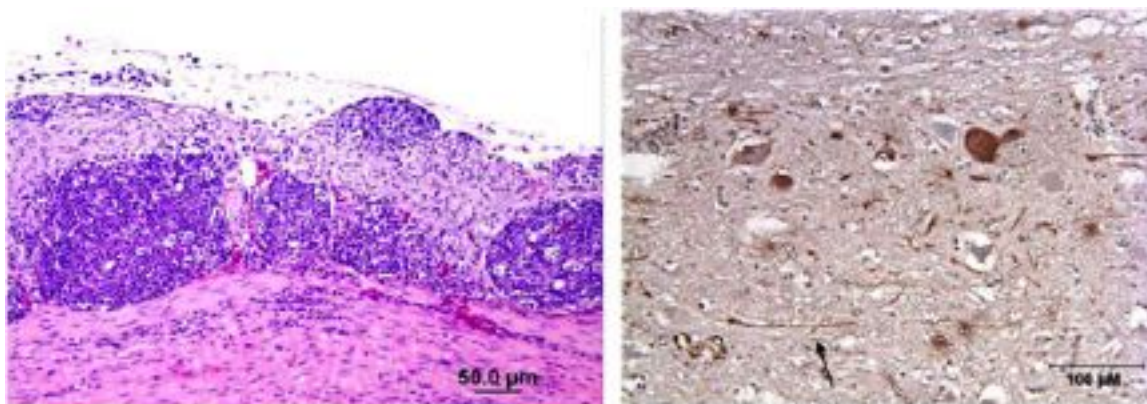


Fig. 4 Microscopic lesions from a birds with avian bornavirus infection and PDD showing lymphocytic infiltration (dark purple cells) (left) and presence of bornavirus antigen (discrete areas of dark brown staining) in neurons (right). From: Payne, S., 2017. Chapter 22 – Family Bornaviridae. In Viruses (edited by S. Payne), Academic Press, pp. 191-196, <https://doi.org/10.1016/B978-0-12-803109-4.00022-2>.

In birds, bornavirus replication is not restricted to neurons. Lymphocytic infiltrates and viral antigens and nucleic acids are abundant in the heart, kidneys, adrenals, gonads and skin (Fig. 4). Bird droppings contain viral nucleic acids and droppings have been used a source of infectious virus. Droppings are a likely vehicle for transmission of avian bornaviruses but it is interesting to note that attempts to experimentally infect birds via the oral or intranasal route have, to date, been unsuccessful.

The first clues to bornavirus infection of waterfowl came from retrospective surveys of Canada geese and trumpeter swans suffering from neurologic disease of unknown origin but with neuropathology consistent with PDD. These studies were carried out at the Ontario Veterinary College, Guelph, Ontario, Canada. Bornaviral sequences and antigens were present in brain tissues of many affected birds. Subsequent surveys of healthy birds (hunter harvested, or culled as nuisance birds at airports and landfills) revealed that avian bornaviruses infection is widespread in North America and Europe. In North America, bornaviruses have been detected in the brains or urofeces of Canada geese (*Branta Canadensis*), snow geese (*Chen caerulescens*), Ross's geese (*Chen rossii*), trumpeter swans (*Cygnus buccinator*), mute swans (*Cygnus olor*) and a variety of ducks and gulls. The majority of waterfowl infections are associated with the species Waterbird 1 orthbornavirus. The virus has also been detected in raptors suggesting that predation may be a source of infection. The overall impact of bornavirus infection of wild bird populations remains unknown.

Endogenous Bornavirus-Like Elements

In recent years, whole genome sequencing has revealed insertions of many non-retroviral genome fragments into the genomes of diverse animal species. Among the viral genome fragments discovered in animal genomes are those of bornaviruses. Fragments of bornavirus genomes have been detected in fish, snakes, turtles, bats, rodents, squirrels, elephants, nonhuman-primates, and humans. Phylogenetic analyzes show that the establishment of endogenous Borna-like (EBL) elements took place took place as multiple independent events, as long as 65 million years ago. Bornavirus N, M, G and L genes are present in eukaryotic genomes. EBLs are derived from viral mRNAs and evidence supports a role for LINE-1 retrotransposition.

Some EBLs comprise complete open reading frames and indeed, some are expressed. In humans there are seven EBL N elements (EBLNs) and all are expressed in at least one tissue. Robust protein expression of Homo sapiens EBLN-1 (hEBLN-1) occurs in various cancer-derived cell lines and it functions in regulation of genome stability and cell cycle. Knockdown of hEBLN-1 RNA results in abnormal cell growth, arrest of cell cycle at G2/M and increased apoptosis. In an experimental system, the thirteen-lined ground squirrel (*Ictidomys tridecemlineatus*) EBLN inhibits the replication of BoDV-1 in cell cultures by suppression of polymerase activity. Thus, some EBLs may play roles in intrinsic immunity to bornaviruses.

Pathogenesis

Non-cytopathic infection and persistence are hallmarks of bornavirus infection, both in cultured cells and within the infected host. Relatively few infectious virions are produced as bornaviruses are highly cell-associated. Immunostaining of infected cell cultures reveals slow 'spread' of the virus as infected cells divide (Fig. 3). Infected cells undergoing mitosis reveal chromatin associated 'particles', suggestive of RNPs, being distributed to daughter cells.

In the infected host, suppression of innate immune responses may favor establishment of infection and viral persistence. Bornaviruses have redundant mechanisms by which they interfere with innate responses.

BDV N interferes with the IKK/NF-κB pathway by inhibiting the cleavage of NF-κB1 p105. N also interferes with IRF7 activation (prevents nuclear localization) and interferon (IFN) type 1 expression. P inhibits the activation of type I interferon (IFN) as well,

through the TBK1-IRF3 pathway. BDV X interferes with IFN signaling and also localizes to mitochondria, blocking apoptosis mediated by the mitochondrial signaling protein MAVS.

Other factors that favor persistence in the infected host include regulation of transcription and genome replication. An interesting and unique feature of bornavirus genome replication is truncation of genome strands. Many genomes cloned from persistently infected cells are truncated and lack sequences required for genome replication. It is speculated that this process leads to down-regulation of replication and or transcription. Genome truncation would also remove 5' terminal triphosphates from RNA, leaving 5' monophosphate ends, providing a mechanism for avoiding recognition by RIG-I.

In horses and sheep, BoDV-1 is exclusively associated with neurons. Virus replication is non-cytopathic but in response to infection, large numbers of T-lymphocytes infiltrate the CNS and are found at sites of virus replication. Thus, there is an immune mediated component to BoDV-1 pathogenesis.

Laboratory rodents are used as models for BoDV-1 infection (although rodents vary greatly in their susceptibility to BoDV-1 infection and different model systems employ specific rodent and virus strains). In the rodent model, virus replication is restricted to neurons and virus spreads through neural networks, similar to the case of rabies virus. BoDV-1 mainly targets hippocampal neurons, but also infects astroglial and oligodendroglial cells in the brain. The outcome of infection depends on the genetic background, age of infection and immune status of the animal. Infection of adult, immunocompetent rats results in an encephalitis similar to that seen in natural infections. It has been clearly demonstrated that encephalitis is immune-mediated, with cytotoxic T-cells being the major players. The pathology of the immunocompetent rat model is identical to the natural infection of horses and sheep. In contrast, experimental infection of neonatal rats (not yet immunocompetent) or immune-suppressed rats, leads to persistent infection, in the absence of acute disease. However, infection is not completely without consequence as persistently infected rats show behavioral, emotional and cognitive impairments.

Neurological disease mediated by BoDV-1 comes in two forms: immune-mediated pathology and alterations in cell function as a direct result of viral proteins. Studies carried out in cultured cells and laboratory rodents reveal myriad effects of BoDV-1 infection at the cellular level. BoDV-1 proteins have various effects on cell signaling, apoptosis, and neurogenesis. Specific effects vary with virus isolate and cell types tested, however it is clear that BoDV-1 proteins cause significant changes in cell physiology and function. Some examples include: Up or down regulation of important signaling proteins such as PKC, Raf, MEK and ERK, changes in gene expression of (TGF)- β family members, and inhibition of histone acetylation. BoDV-1 infection of neuronal progenitor cells impairs neurogenesis and causes diminished neuritic outgrowth and synapse formation. Many of these cellular perturbations are linked to BoDV-1 P and it has been shown that expression of P in transgenic mice induces measurable behavioral abnormalities.

Diagnosis

Clinical presentation alone is insufficient for a diagnosis of bornaviral disease in animals as differential diagnosis for an animal with encephalitis will include any number of viral agents (which will vary by species). Currently there is no single assay recommended for diagnosing bornavirus infection, however a number of serological assays for detection of antiviral antibodies have been described. These include immunofluorescence assays (targeting bornavirus-infected cells) and Western blot and ELISA (targeting recombinant N or P antigens). Results of serologic assays must be interpreted carefully as a negative antibody test does not rule out bornavirus infection, as some infected animals will be antibody negative. Conversely, antibody positive animals may be completely healthy and positive serology is not a predictor of disease among healthy animals. In birds with suspected PDD, serologic assays are highly problematic as antibody production in experimentally infected birds is highly variable.

Detection of bornaviral RNA by reverse transcriptase (RT) PCR is a more reliable indicator of infection. Most PCR assays target the N or P genes. In cats, recommended samples include blood, serum, urine, conjunctival, oral, nasal and/or anal swabs. In birds, avian bornaviral RNAs can be detected in cloacal swabs and the urofeces but virus shedding is intermittent. While positive PCR results are indicative of bornavirus infection, negative results indicate only a current lack of viral shedding. Currently, PCR remains the best indicator test for avian bornavirus infection.

Bornavirus induced encephalitis is usually fatal, thus many diagnoses are made post-mortem. Typical findings include non-suppurative lymphocytic infiltrates in the brain or spinal cord. Perivascular cuffing is a common finding upon histological examination of brain tissues. Immunohistochemistry using N or P antisera can be used to stain bornavirus infected neurons to achieve a definitive diagnosis. RT-PCR of RNA extracted from brain tissue may also be used to confirm bornavirus infection.

Control and Prevention

There are currently no bornaviral vaccines approved for veterinary use in either mammals or birds. Other than among pet birds, symptomatic bornavirus infections of animals are quite rare, even in confirmed endemic areas of Central Europe. Infections in this region are likely the result of contact with urine or feces of natural hosts, such as the white-toothed shrew. Currently the highest risk of human bornavirus infection appears to be contact with certain squirrels (*S. variegatoides* and *C. prevostii*). Four human deaths from Mammalian 2 orthobornaviruses have been reported in Germany, and all occurred among persons having contact with either *Sciurus* or *Callosciurus* squirrels which are exotic to Europe. The original source of the squirrel infections is unclear but

keeping exotic squirrels as pets may be inadvisable. In 2018 two human deaths due to BoDV-1 infection were reported in Germany. These were associated with organ transplantation from a BoDV-1 infected donor. It is currently not clear if the risk of human bornavirus encephalitis is higher in Germany than in other locals, or if these cases were detected due to heightened awareness of the potential for bornavirus infection in this area. Unfortunately serologic assays will not detect 100% of bornavirus infections and even the most sensitive PCR assays will not detect virus if replication is restricted to the central nervous system. Controlling the spread of avian bornaviruses among pet birds might be achieved by testing birds using PCR assays of urofeces.

Future Prospects

Since the advent of metagenomics analyzes and high-throughput sequencing, the list of known bornaviruses has grown from 1 to 5 recognized species, geographic range has expanded from an area of Central Germany to world-wide, and hosts have expanded to include birds and reptiles. Much of our understanding of bornavirus replication and pathogenesis comes from studies of BoDV-1 and it is likely that use of well-characterized laboratory rodent models will continue to add to our knowledge base. In recent years there has been debate over a role for bornavirus involvement in human psychiatric disease. The recognition of novel bornaviruses and the development of rigorous and validated methods for assessing the infection status of patients should allow clarification of this important issue. Indeed, recent reports of bornavirus-associated fatal encephalitis in humans clearly demonstrate the ability of humans serve as spillover hosts. We currently have only a rudimentary understanding of the natural reservoirs of bornaviruses and continued surveys are needed to gain a more complete understanding of the full host range and genetic diversity of these fascinating viruses. On a very different note, the study of EBLs should be pursued as a tractable model for understanding the ancient history of the association between eukaryotes and their viruses.

Further Reading

- Formisano, S., Hornig, M., Yaddanapudi, K., *et al.*, 2017. Central nervous system infection with borna disease virus causes Kynurenine pathway dysregulation and neurotoxic quinolinic acid production. *Journal of Virology* 91, e00673.
- Hoffmann, B., Tappe, D., Höper, D., *et al.*, 2015. A variegated squirrel bornavirus associated with fatal human encephalitis. *The New England Journal of Medicine* 373, 154.
- Horie, M., Kobayashi, Y., Suzuki, Y., Tomonaga, K., 2013. Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. *Philosophical Transactions of the Royal Society London B Biological Sciences* 368, 20120499.
- Hornig, M., Mervis, R., Hoffman, K., Lipkin, W.I., 2002. Infectious and immune factors in neurodevelopmental damage. *Molecular Psychiatry* 7, S34.
- Myers, K.N., Barone, G., Ganesh, A., *et al.*, 2016. The bornavirus-derived human protein EBLN1 promotes efficient cell cycle transit, microtubule organisation and genome stability. *Scientific Reports* 6, 35548.
- Tomonaga, K., Kobayashi, T., Ikuta, K., 2002. Molecular and cellular biology of borna disease virus infection. *Microbes and Infection* 4, 491.

Bovine Leukemia Virus (*Retroviridae*)

Thomas Joris, Roghaiyeh Safari, Jean-Rock Jacques, and Luc Willems, Cellular and Molecular Epigenetics (GIGA), Liège, Belgium and Molecular Biology (TERRA), Gembloux, Belgium

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AGID Agar Gel Immunodiffusion test

BLV Bovine Leukemia Virus

CAT1/SLC7A1 Cationic Amino acid Transporter 1/Solute Carrier family 7 member 1

EBL Enzootic Bovine Leukemia

ELISA Enzyme-Linked Immunosorbent Assay

FPPS Farnesyl Pyrophosphate Synthetase

HTLV-1 Human T-cell Leukemia Virus type 1

ICTV International Committee on Taxonomy of Viruses

LTR Long Terminal Repeat

PBMCs Peripheral Blood Mononuclear Cells

PCR Polymerase Chain Reaction

PL Persistent Lymphocytosis

PVL Proviral Load

VPA Valproic Acid

Glossary

Bidirectional transcription The long terminal repeats located at the 5' and 3' ends of the provirus contain promoter sequences that direct bi-directional transcription.

Clonal expansion As a retrovirus, BLV replicates upon expression of an integrated provirus, budding of the viral particle and infection of a new cell (the infectious cycle). When integrated into the host chromosome, BLV replication

also occurs by cell mitosis through a mechanism of clonal expansion.

MicroRNA Besides genomic and subgenomic mRNAs, BLV also transcribes 10 viral microRNAs that are exported into the plasma of infected animals.

Provirus BLV is a retrovirus whose RNA genome is reverse transcribed into DNA, the provirus, that integrates into the host chromosome.

Classification

According to the International Committee on Taxonomy of Viruses (ICTV, 2019 release), BLV is classified in the *Ortervirales* order, *Retroviridae* family, *Orthoretrovirinae* subfamily and *Deltaretrovirus* genus.

Epidemiology and Clinical Features

The bovine leukemia virus (BLV) is an exogenous, oncogenic retrovirus that naturally infects B-cells of cattle (*Bos taurus*), water buffalo (*Bubalus bubalis*), yak (*Bos grunniens*) and zebu (*Bos taurus indicus*) (Meas *et al.*, 2000; Burny *et al.*, 1988). BLV-associated disease was initially observed in 1871 in an enlarged spleen of a cow (Gillet *et al.*, 2007). BLV is the etiological agent of enzootic bovine leukosis (EBL) characterized by lymphoma and lymphosarcoma (extranodular lymphoma in different organs such as eyes, heart, liver, lung, and skin) (Burny *et al.*, 1985). While most are asymptomatic carriers, a minority of infected animals develop EBL several years after infection (Barez *et al.*, 2015). The frequency of tumors and clinical latency depend on herd prevalence: a typical picture is 10% death after three years in a herd having 50% BLV prevalence.

In approximately one third of BLV-infected cows, a benign condition called persistent lymphocytosis (PL) is characterized by an excessive accumulation of B-lymphocytes in the peripheral blood (Aida *et al.*, 2013). In PL, the number of B cells can exceed that of T lymphocytes. Animals with PL have a higher risk of EBL although tumor development also occurs without persistent lymphocytosis. BLV infection is associated with immunosuppression that favors opportunistic infections (e.g., mastitis) (Suzuki *et al.*, 2013).

Diagnosis

BLV infection initiates a strong humoral and cytotoxic immune response that tightly controls viral replication (Meas *et al.*, 2000; Kabeya *et al.*, 1999; Pyeon *et al.*, 1996). Since assays based on cytotoxic T-cell responses are not robust, diagnosis is primarily based on the presence of anti-BLV antibodies in the plasma or serum. The historical agar gel immunodiffusion test (AGID) is nowadays being replaced by the enzyme-linked immunosorbent assay (ELISA) (LaDronka *et al.*, 2018; Gutiérrez *et al.*, 2009; Portetelle *et al.*, 1991). The viral antigens that elicit the best response belong to the capsid (p24) and envelope (gp51) proteins. Commercially available ELISAs have an excellent performance in terms of specificity and sensibility. The issue with these tests is their inability to discriminate between viral infection and maternally derived anti-BLV antibodies. Approximately 1–4 weeks after infection, the humoral immune response indeed produces anti-p24 and anti-gp51 antibodies that cannot be dissociated from passive immunity acquired through the colostrum

(Ferrer and Piper, 1981; Kuczewski *et al.*, 2019). This issue is solved by different PCR methods for detecting proviral sequences (e.g., end-point, real-time, nested, filter, CoCoMo and isothermal) (Jimba *et al.*, 2010; Jaworski *et al.*, 2018; Ruggiero *et al.*, 2018).

Prevention and Treatment

BLV is a major animal health problem worldwide causing important economic losses (Norby *et al.*, 2016). A series of attempts were developed to reduce prevalence, chiefly by eradication of infected cattle or segregation of BLV-free animals (Rodríguez *et al.*, 2011). The “test and eliminate” strategy requires identification of BLV-positive animals either by genomic or serological methods, removal of positive cases from the herd and slaughtering. Although having been instrumental in regions such as in Europe, this strategy was unsuccessful elsewhere mainly due to excessive costs (EFSA AHAW Panel, 2015). Since efficient transmission correlates with high proviral loads, an alternative is to identify and discard only the virus shedders. Although the impact on genetic and reproductive potential is reduced in the herd, BLV infection persists over longer periods (Ruggiero and Bartlett, 2019).

The “test and segregate” approach detects and separates BLV-infected and healthy cattle in different areas. The need of separated operational structures duplicates housing facilities and equipment, thereby excessively increasing the costs.

“Test and manage” BLV-infected and healthy animals in the same housing facilities requires permanent surveillance of the herd. This strategy is cost-effective, requires only minimal investment on facilities and does not need replacement of culled BLV-infected cattle (Bartlett *et al.*, 2014). The “Test and manage” approach is nevertheless intensively laborious, requires strict adherence to the rigorous implemented measures and needs long-term commitment to the program. Management includes the single-use of disposable material (needles, syringes and obstetrical sleeves), the disinfection of reusable surgical instruments (dehorning, tattooing, castration or ear-tagging) and elimination of hematophagous insects (Rodríguez *et al.*, 2011).

The strategies of testing and slaughtering/segregating/managing could be implemented with a treatment that limits the proviral loads. For example, valproic acid (VPA), an HDAC inhibitor, has been successfully used to reduce the PVL in leukemic sheep (Achachi *et al.*, 2005). Although VPA treatment does not clear infection, reduced PVL correlates with decreased transmission and pathogenesis.

BLV transmission can also be prevented by feeding the calves with colostrum from BLV-infected cows (Ferrer and Piper, 1981; Kuczewski *et al.*, 2019). Although anti-BLV antibodies from the colostrum efficiently prevent infection, there is a risk of transmission when plasma titers decrease.

Vaccination against BLV has been quite difficult to achieve probably because of its ability to stably integrate into the host genome, undergo long-term latency in a proportion of infected cells and thereby escape immune response. Since clearance of the virus is impossible once infection is established, the primary goal was to achieve sterilizing immunity. Besides efficacy, safety is the major issue since vaccination has been associated with increased infection or reversion to pathogenicity. After a series of failures due to the fast decline of protective antibody titers and poor stimulation of cytotoxic response, an efficient and safe vaccine is now available (Gutierrez *et al.*, 2014; Abdala *et al.*, 2019). The concept of this vaccine is to establish a permanent infection with an attenuated strain that impedes wild-type challenge by activating the anti-viral immune response. Since the vaccine strain replicates at very low levels, is not pathogenic and does not transmit to uninfected sentinels or from cow-to-calf. Moreover, the low level of replication also prevents a significant impact on the host immunocompetence, thereby avoiding immunosuppression and opportunistic infections.

Life Cycle

The infectious cycle of BLV replication initiates with the interaction of the viral particle with the cell membrane (Fig. 1). The BLV envelope glycoprotein gp51 binds to the cationic amino acid transporter 1 (CAT1)/solute carrier family 7 member 1 (SLC7A1) receptor (Bai *et al.*, 2019). The transmembrane protein gp30 then promotes fusion between the lipid bilayers of the viral and cell membranes (Willems *et al.*, 1995). Upon uncoating of the capsid, the BLV genomic RNA is reverse transcribed into a double stranded DNA provirus that is transported into the nucleus after circularization (Gillet *et al.*, 2007). Integrase-mediated proviral insertion into the host cell genome is random albeit BLV proviral integration significantly favors transcribed regions of the genome (Gillet *et al.*, 2013). Negative selection then eliminates 97% of the clones detected at seroconversion and disfavors BLV-infected cells carrying a provirus located close to a promoter or a gene. Nevertheless, among the surviving proviruses, clone abundance positively correlates with proximity of the provirus to a transcribed region. The integrated provirus can remain silent or be transcribed under the control of various stimuli (e.g., PKC, calmodulin) (Kerkhofs *et al.*, 1996). Translation of viral proteins, encapsidation of genomic RNA and budding at the cell membrane release a new infectious virion.

Besides this infectious cycle, BLV also replicates by mitosis of its host cell (i.e., clonal expansion) (Gillet *et al.*, 2013). The relative importance of these two cycles in viral replication varies during infection. The majority of infected clones are created early before the onset of an efficient immune response. Later on, the main replication route is mitotic expansion of pre-existing infected clones. The mechanism that stimulate preferred mitosis of BLV-infected cells is still incompletely understood but involves oncogenic viral proteins such as Tax and G4 (Willems *et al.*, 1990; Kerkhofs *et al.*, 1998).

Virion Structure

The BLV RNA genome is packaged in a capsid surrounded by an external envelope (Fig. 2). The p12 nucleocapsid (NC) protein is a proline-rich polypeptide that is tightly bound to the packaged genomic RNA. The p24 protein is the major constituent of the capsid

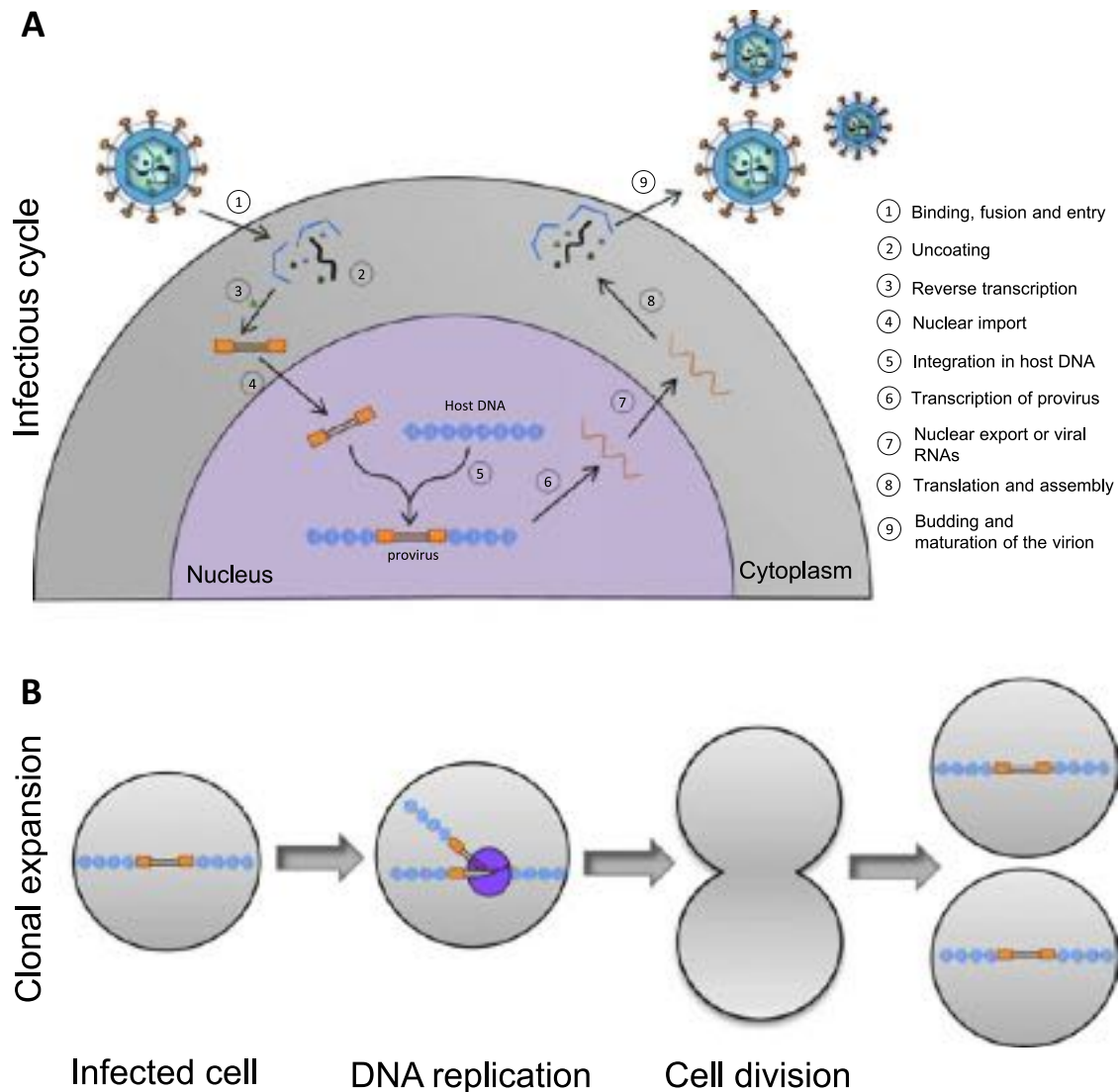


Fig. 1 Schematic representation of the two modes of BLV replication. (A) In the infectious cycle, the virion enters the cell by envelope-mediated membrane fusion. After reverse transcription of the genomic RNA, the provirus integrates into the host chromosome. Upon transcription of the provirus and translation, viral proteins and genomic RNA form a virion that buds out of the cell (B) An infected cell harboring an integrated provirus undergoes mitosis. Further proliferation generates a population of cell clones characterized by a single proviral integration site.

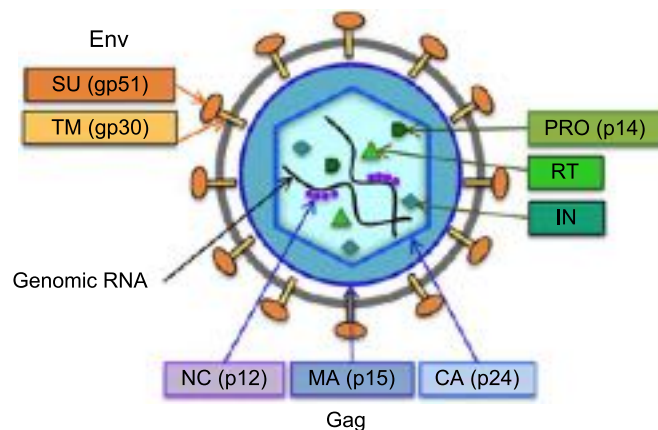


Fig. 2 Structure of the BLV virion. Env, Gag and Pro-Pol proteins are indicated by orange, blue and green arrows. Gag = group-specific antigen; Env = envelope; MA = matrix; NC = nucleocapsid; CA = capsid; PRO = protease; RT = reverse transcriptase; IN = integrase; SU = extracellular envelope protein; TM = transmembrane protein.

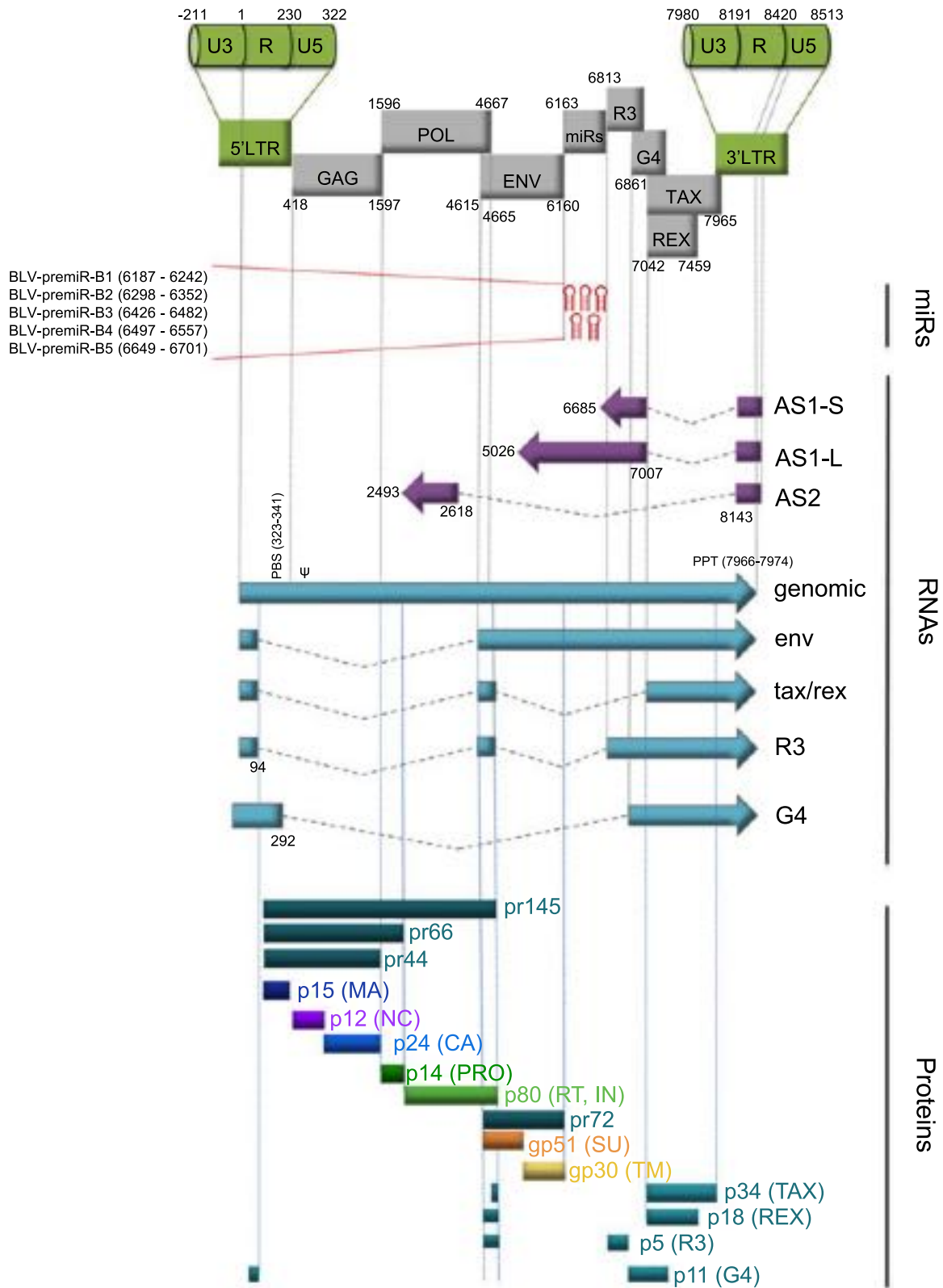


Fig. 3 Schematic representation of an integrated BLV provirus, viral RNAs and proteins. The long terminal repeats (LTR in green) flank the viral genes (in gray). The provirus is integrated into chromatin. Sense and antisense transcripts are colored in blue and purple, respectively. GAG = group-specific antigen; ENV = envelope; miR = viral microRNAs; premiR = precursor microRNA; U3 = unique 3' end; R = redundant region; U5 = unique 5' end; PPT = polypurine tract; PBS = primer binding site; ψ = packaging Psi sequence; AS1-S = antisense 1 short; AS1-L = antisense 1 long; AS2 = antisense 2; pr = precursor; p = protein; gp = glycoprotein; MA = matrix; NC = nucleocapsid; CA = capsid; PRO = protease; RT = reverse transcriptase; IN = integrase; SU = extracellular envelope protein; TM = transmembrane protein. Numbers are coordinates according to the GenBank AF033818.1 reference genome (1 is the transcription initiation site).

(CA) of BLV virions and a major target for the host immune response. The matrix (MA) protein p15, which corresponds to the NH₂-terminal end of the gag precursor, is a myristylated and phosphorylated polypeptide. MA proteins bind the genomic viral RNA but also interact with the lipid bilayer of the viral membrane. The Gag proteins (MA, CA and NC) are derived from the proteolytic cleavage of the Pr44gag precursor carried out by the viral protease p14 (PRO). The BLV virion also contains two enzymes required to reverse transcribe the RNA genome into proviral cDNA (RT) and for its integration into the chromosome (IN). Finally, the capsid is surrounded by an envelope of cellular and viral origin. The extracellular gp51 (SU) and the transmembrane gp30 (TM) proteins are glycosylated polypeptides that associate through disulfide bonds. TM is embedded through a lipid bilayer acquired upon budding from the cell membrane. Insertion, in an oblique orientation, of the N-terminal fusion peptide of TM into the lipid bilayer of a target cell promotes cell fusion and viral entry. SU is involved in receptor binding and cell fusion.

It should be emphasized that, although BLV virions are spontaneously expressed in ex vivo cultures from peripheral blood mononuclear cells (PBMCs), there is no viremia in infected animals, likely due to a very efficient immune response.

Genome Organization and Expression

The BLV provirus integrated into chromatin is schematized in **Fig. 3**. The genome contains the classical gag-pro-pol-env structural and enzymatic genes required to produce the viral particle. As a delta-retrovirus, BLV also harbors 4 genes (Tax, Rex, R3 and G4) involved in different regulatory mechanisms (see the paragraph on pathogenesis). A series of non-coding RNAs are transcribed from a microRNA cluster (miRs) and in an antisense orientation (AS1-S, AS1-L, AS2). The provirus is flanked by two identical long terminal repeats (5'LTR and 3'LTR) that direct initiation and termination of transcription.

The U3-R boundary of the 5' LTR identifies the transcription initiation site of the sense RNAs (**Fig. 3**). Transcription terminates by polyadenylation at the R-U5 junction of the 3'LTR. Sense transcription is initiated by a TATA box (GATAAT) and controlled by a series of enhancers. Among these, three 21 bp-long Tax-responsive elements (TxREs) contain imperfectly conserved cyclic-AMP responsive elements (AGACGTCA; TGACGGCA and TGACCTCA). These enhancers mediate activation of transcription by the viral Tax protein upon interaction with CREB/ATF (*Calomme et al., 2004; Adam et al., 1994*) (**Fig. 4**). Other regulators of sense transcription include binding sites for NF- κ B, PU.1/Spi-B, glucocorticoid responsive element (GRE), E box and interferon regulatory factor (IRF) (see review (*Gillet et al., 2007*) for further details).

The genomic RNA contains the Psi site (ψ) required for its packaging inside the virion and the tRNA primer-binding site (PBS) that is involved in initiation of reverse transcription. During reverse transcription, an RNA polypurine tract (PPT) resists digestion by RT and primes plus-strand DNA synthesis. The genomic RNA is capped with 7-methyl guanosine and is spliced into 4 transcripts (*env*, *tax/rex*, *R3* and *G4*).

The RT and IN proteins are translated from a 145-kDa polypeptide via a frameshift mechanism of the gag-protease precursor (pr66gagprt). Different Gag proteins (MA, CA and NC) are derived from the proteolytic cleavage of the Pr44Gag precursor. This post-translational maturation is carried out by the viral protease PRO (p14) that is encoded by a region located between the *gag* and the *pol* genes. p14 is synthesized from a Gag-protease precursor (pr66GagPrt) via a frameshift suppression of the gag termination codon by a lysine-specific tRNA.

The sequences coding for the envelope partially overlap in a different frame the 3' end of the *pol* gene by 51 nucleotides (**Fig. 3**). A single-spliced RNA encodes the pr72Env precursor that is cleaved into gp51 (SU) and gp30 (TM) by the subtilisin/kesin-like convertase furin. The SU and TM glycosylated proteins associate through disulfide bonds and interact with the cell membrane.

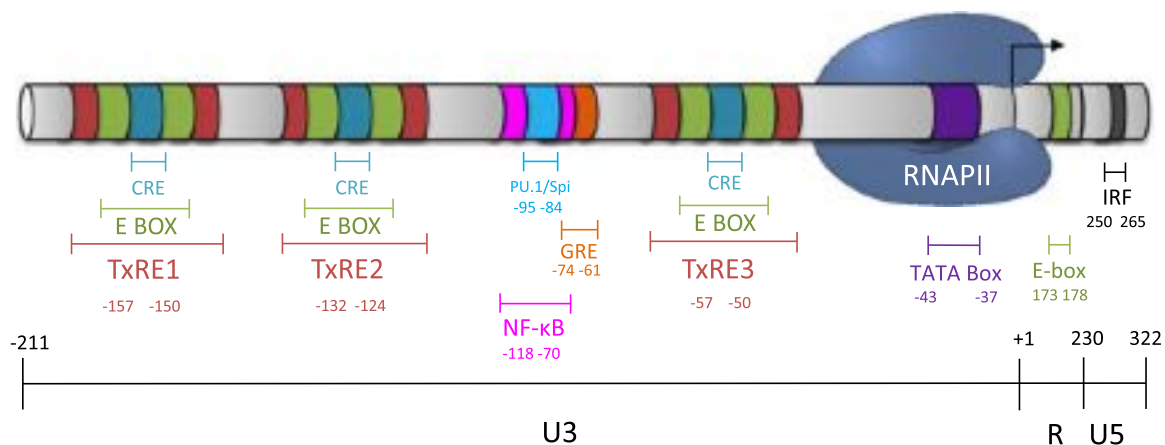
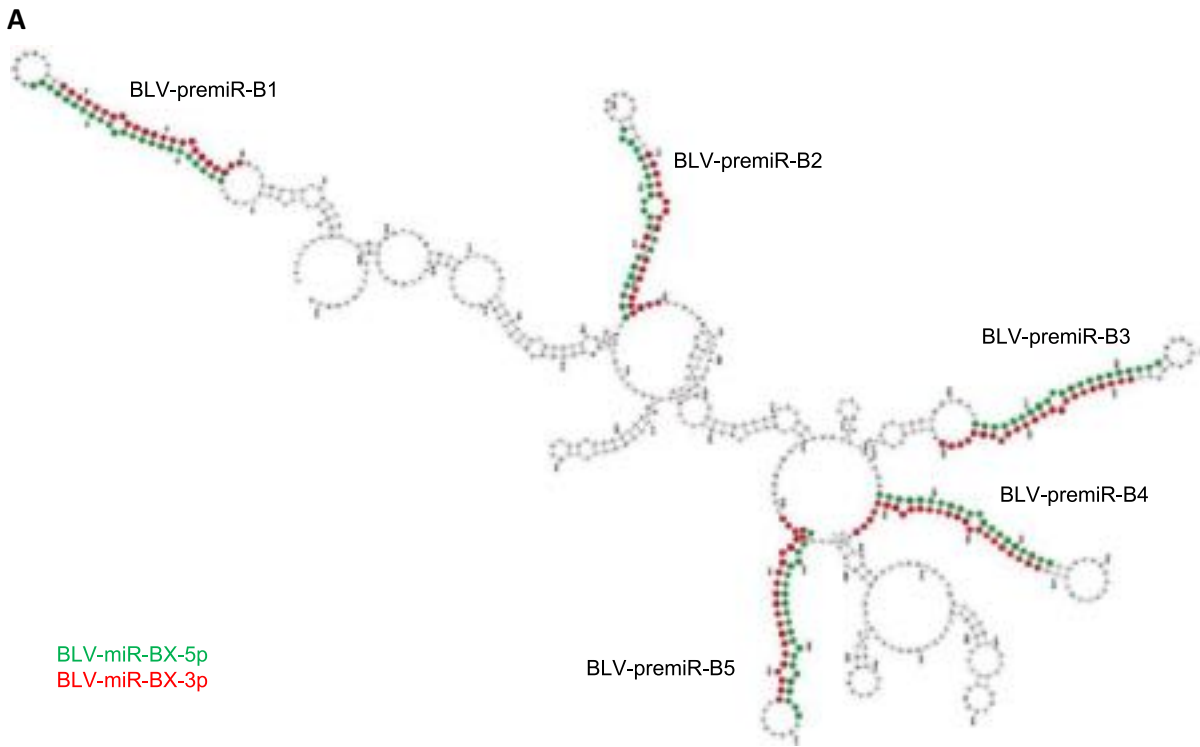


Fig. 4 Regulatory elements of the LTR promoter. Transcription of the 5'-LTR is initiated by RNAPII at position +1. TxRE = Tax responsive element; CRE = cyclic-AMP response element; GRE = glucocorticoid responsive element; E-box = enhancer box; NF- κ B = nuclear factor κ B; TATA = TATA box; IRF = interferon regulatory factor; U3 = unique 3' end; R = redundant region; U5 = unique 5' end. All coordinates are indicated according to the GenBank: AF033818.1 reference genome.



B

Name	Sequence	Coordinates	miRBase (22.1) accession
BLV-miR-B1-5p	AGGCGUGGGUGGUGCACUGGCUU	6187 - 6209	MIMAT0027346
BLV-miR-B1-3p	UCAGUGUACCAUCAAGCCUCU	6220 - 6242	MIMAT0025859
BLV-miR-B2-5p	AUGACUGAGUGUAGCGCAGAGA	6298 - 6319	MIMAT0025860
BLV-miR-B2-3p	UGCGUGUCGUCAGUCAUUUU	6331 - 6352	MIMAT0025861
BLV-miR-B3-5p	AUCCCCUGCCAGCGUUGGUC	6426 - 6246	MIMAT0025862
BLV-miR-B3-3p	UAACGCUGACGGGGGCGAUUUUCU	6460 - 6482	MIMAT0025863
BLV-miR-B4-5p	GCGGGAGGCUCUGGUGCUGG	6497 - 6516	MIMAT0027347
BLV-miR-B4-3p	UAGCACACAGUCUCUGCGCCUUU	6534 - 6557	MIMAT0025864
BLV-miR-B5-5p	AGGAAGGUUGGUCUCAGAGGU	6649 - 6670	MIMAT0025865
BLV-miR-B5-3p	CUCGAGCCGCAACCCUUUCU	6679 - 6701	MIMAT0025866
BLV-premiR-B1	AGGCGUGGGUGGUGCACUGGCUU-AGUGGAGUAG- UCAGUGUACCAUCAAGCCUCU	6187 - 6242	MI0022561
BLV-premiR-B2	AUGACUGAGUGUAGCGCAGAGA-GAUUGUCGUUC- UGCGUGUCGUCAGUCAUUUU	6298 - 6352	MI0022562
BLV-premiR-B3	AUCCCCUGCCAGCGUUGGUC-UAGUGGAAAGAAC- UAACGCUGACGGGGGCGAUUUUCU	6426 - 6482	MI0022563
BLV-premiR-B4	GCGGGAGGCUCUGGUGCUGG-GGAUAAGAUUGGCCCU- UAGCACACAGUCUCUGCGCCUUU	6497 - 6557	MI0022564
BLV-premiR-B5	AGGAAGGUUGGUCUCAGAGGU-UAAAAUAG- CUCGAGCCGCAACCCUUUCU	6649 - 6701	MI0022565

Fig. 5 Sequences of the BLV miRNAs. (A) Structural prediction of the transcribed BLV miRNA cluster, predicted from sequence 6163–6813 by the RNAfold WebServer of the Vienna RNA Websuite. Nucleotide 1 is at position 6163 of the GenBank AF033818.1 reference genome. 5p- and 3p-arms are highlighted in green and in red, respectively (B) Sequences, coordinates and references of the 10 BLV miRNAs.

Among regulatory proteins, Tax is a phosphoprotein that activates directed transcription via CREB/ATF and TxRE elements located in U3 (Willems *et al.*, 1987, 1998) (Fig. 4). Tax is oncogenic in immortalization assays in primary cells and in mouse models (Willems *et al.*, 1990), a property that is shared with G4 (Kerkhofs *et al.*, 1998). Localized in the mitochondria and the nucleus, the G4 protein interacts with farnesyl pyrophosphate synthetase (FPPS) involved in the mevalonate/squalene pathway and prenylation of Ras (Lefebvre *et al.*, 2002). The role of nuclear R3 protein is still unclear but could pertain to regulation of Rex activity. The 18 kDa Rex nuclear phosphoprotein interacts with the RxRE responsive RNA element located between the AATAAA polyadenylation signal and the polyadenylation site (8156–8420) of viral transcripts (Derse, 1988). p18 mediates splicing and is required for the accumulation of viral genomic and env transcripts. Tax and Rex are essential genes while R3 and G4 are dispensable for infectivity (Willems *et al.*, 1994; Florins *et al.*, 2007b).

BLV also transcribes two antisense RNAs (AS1 and AS2) initiating at two transcriptional start sites of an active RNAPII promoter located in the 3' LTR (Van Driessche *et al.*, 2016; Durkin *et al.*, 2016). The AS1 RNA is differentially polyadenylated, generating AS1-L (2200 bp) and, more abundantly, AS1-S (600 bp) (Fig. 3). The AS1-L isoform overlaps the microRNA cluster and is therefore cleaved in the cytoplasm by the RISC complex. The antisense transcripts are predominantly located in the nucleus and are predicted to be untranslated. Their possible role as long non-coding RNAs mediating mechanisms such as epigenetics remains to be determined.

Finally, five precursor miRNAs are transcribed from a cluster located 3' of the env gene on the BLV genome (Rosewick *et al.*, 2013; Kincaid *et al.*, 2012) (Fig. 3). The pre-miRNAs are directly transcribed from RNA polymerase III (RNAPIII) promoters and are further processed in the cytoplasm into 10 mature miRNAs by Dicer (Fig. 5). The pre-miRNAs are therefore not processed by the Drosha enzyme, but still largely depend on the DUSP11 phosphatase and Dicer (Burke *et al.*, 2016). In contrast to genomic and subgenomic RNAs, miRNAs are abundantly expressed in infected animals (i.e., up to 40% of all detected cellular miRNAs) (Rosewick *et al.*, 2013; Gillet *et al.*, 2016). Mechanistically, the BLV miRNAs target genes involved in proliferation, oncogenesis, cell signaling, apoptosis and immunity (e.g., GZMA, FOS, PPT1, ANXA1, MAP2K1 and PIK3CG) (Gillet *et al.*, 2016). Deletion of the miRNA cluster reduces viral replication in the natural (bovine) and experimental (ovine) models. Ablation of miRNAs mainly affects proliferation and abolishes pathogenesis in BLV-infected sheep (Safari *et al.*, 2020).

It thus appears that BLV proviral transcription is characterized by a complex profile of miRNAs, sense and antisense RNAs driven by RNAPII and RNAPIII, respectively.

Pathogenesis

Although the mechanisms associated with BLV pathogenesis still need to be further investigated, a series of experimental results generates a hypothetical but consistent model (Safari *et al.*, 2017; Gazon *et al.*, 2018; Florins *et al.*, 2007a). The hypothesis is based on a tight equilibrium between a virus attempting to proliferate and an infected host that develops an efficient immunity to control viral replication. Evidence includes the following:

- (1) Despite a strong anti-viral immune response, BLV persists indefinitely throughout life;
- (2) BLV encodes oncogenes promoting cell proliferation and transformation (Kerkhofs *et al.*, 1998; Twizere *et al.*, 2003);
- (3) Expression of structural (*gag*, *env*), enzymatic (*pol*) and regulatory (*tax*, *rex*, *R3* and *G4*) genes is barely detectable in vivo (Lagarias and Radke, 1989);
- (4) Viral miRNAs expressed abundantly in cells and in plasma are required to maintain high proviral loads (Rosewick *et al.*, 2013; Gillet *et al.*, 2016);
- (5) Non-coding antisense transcripts are silent to the immune response and may affect epigenetics (Durkin *et al.*, 2016);
- (6) BLV integrates randomly but favors transcriptionally active sites nearby cancer drivers (Gillet *et al.*, 2013; Rosewick *et al.*, 2017);
- (7) Forced proliferation of infected cell clones generates replication-associated mutations that are predicted to drive oncogenicity (Rosewick *et al.*, 2017);
- (8) Persistent infection induces inflammation and immunosuppression (Pyeon *et al.*, 1996; Konnai *et al.*, 2013)
- (9) The host's genetic background and viral determinants affect proviral loads (de Brogniez *et al.*, 2015; Willems *et al.*, 2000; Hayashi *et al.*, 2017; Tajima and Aida, 2000; Takeshima *et al.*, 2017)
- (10) Host immunity tightly controls viral replication (Florins *et al.*, 2006; Florins *et al.*, 2009; Florins *et al.*, 2011; Frie *et al.*, 2018)

Regular switches between 5' and 3' transcription may allow transient Tax expression and fast silencing of viral expression. Associated with viral miRNAs, this mechanism would allow Tax-driven cell proliferation and synthesis of viral particles in presence of the host immunity. This model thus illustrates the dynamic equilibrium between a virus attempting to proliferate under a tight control exerted by the immune response.

Acknowledgments

This work was supported by the "Fonds National de la Recherche Scientifique" (FNRS, grant PDR T.0261.20), the Télévie, "Crédits Sectoriels de Recherche en Sciences de la Santé" of the university of Liège and the Fondation Léon Fredericq (FLF). LW is research director of the FNRS.

References

- Abdala, A., Alvarez, I., Brossel, H., *et al.*, 2019. BLV: Lessons on vaccine development. *Retrovirology* 16, 1–6.
- Achachi, A., Florins, A., Gillet, N., *et al.*, 2005. Valproate activates bovine leukemia virus gene expression, triggers apoptosis, and induces leukemia/lymphoma regression in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 102 (29), 10309–10314.
- Adam, E., Kerkhofs, P., Mammerickx, M., *et al.*, 1994. Involvement of the cyclic AMP-responsive element binding protein in bovine leukemia virus expression in vivo. *Journal of Virology* 68 (9), 5845–5853.
- Aida, Y., Murakami, H., Takahashi, M., Takeshima, S.-N., 2013. Mechanisms of pathogenesis induced by bovine leukemia virus as a model for human T-cell leukemia virus. *Frontiers in Microbiology* 4, 328.
- Bai, L., Sato, H., Kubo, Y., Wada, S., Aida, Y., 2019. CAT1/SLC7A1 acts as a cellular receptor for bovine leukemia virus infection. *FASEB Journal* 33, 14516–14527.
- Barez, P.Y., de Brogniez, A., Carpentier, A., *et al.*, 2015. Recent advances in BLV research. *Viruses* 7, 6080–6088.
- Bartlett, P.C., Sordillo, L.M., Byrem, T.M., *et al.*, 2014. Options for the control of bovine leukemia virus in dairy cattle. *Journal of the American Veterinary Medical Association* 244, 914–922.
- Burke, J.M., Kincaid, R.P., Nottingham, R.M., Lambowitz, A.M., Sullivan, C.S., 2016. DUSP11 activity on triphosphorylated transcripts promotes Argonaute association with noncanonical viral microRNAs and regulates steady-state levels of cellular noncoding RNAs. *Genes Development* 30 (18), 2076–2092.
- Burny, A., Bruck, C., Cleuter, V., *et al.*, 1985. Bovine Leukemia Virus, a versatile agent with various pathogenic effects in various animal species. *Cancer Research* 45 (suppl), 4578–4583.
- Burny, A., Cleuter, Y., Kettmann, R., *et al.*, 1988. Bovine leukaemia: Facts and hypotheses derived from the study of an infectious cancer. *Veterinary Microbiology* 17 (3), 197–218.
- Calomme, C., Dekoninck, A., Nizet, S., *et al.*, 2004. Overlapping CRE and E box motifs in the enhancer sequences of the bovine leukemia virus 5' long terminal repeat are critical for basal and acetylation-dependent transcriptional activity of the viral promoter: implications for viral latency. *Journal of Virology* 78 (24), 13848–13864.
- de Brogniez, A., Bouzar, A.B., Jacques, J.-R., *et al.*, 2015. Mutation of a single envelope N-linked glycosylation site enhances the pathogenicity of bovine leukemia virus. *Journal of Virology* 89 (17), 8945–8956.
- Derse, D., 1988. Trans-acting regulation of bovine leukemia virus mRNA processing. *Journal of Virology* 62 (4), 1115–1119.
- Durkin, K., Rosewick, N., Artesi, M., *et al.*, 2016. Identification and characterization of novel Bovine Leukemia Virus (BLV) antisense transcripts reveals their constitutive expression in leukemic and pre-leukemic clones. *bioRxiv*. 1–31.
- EFSA AHAW Panel, 2015. Scientific opinion on enzootic bovine leukosis. *European Food Safety Authority. EFSA Journal* 13, 4188.
- Ferrer, J.F., Piper, C.E., 1981. Role of colostrum and milk in the natural transmission of the bovine leukemia virus. *Cancer Research* 41 (12), 4406–4409.
- Florins, A., de Brogniez, A., Elemans, M., *et al.*, 2011. Viral expression directs the fate of B cells in bovine leukemia virus-infected sheep. *Journal of Virology* 86 (1), 621–624.
- Florins, A., Gillet, N., Asquith, B., *et al.*, 2007a. Cell dynamics and immune response to BLV infection: A unifying model. *Frontiers in Bioscience* 12, 1520–1531.
- Florins, A., Gillet, N., Boxus, M., *et al.*, 2007b. Even attenuated bovine leukemia virus proviruses can be pathogenic in sheep. *J Virol* 81 (18), 10195–10200.
- Florins, A., Gillet, N., Asquith, B., *et al.*, 2006. Spleen-dependent turnover of CD11b peripheral blood B lymphocytes in BLV-infected sheep. *Journal of Virology* 80 (24), 11998–12008.
- Florins, A., Reichert, M., Asquith, B., *et al.*, 2009. Earlier onset of δ -retrovirus-induced leukemia after splenectomy. *PLoS One* 4 (9), 1–7.
- Frie, M.C., Droscha, C.J., Greenlick, A.E., Coussens, P.M., 2018. MicroRNAs encoded by bovine leukemia virus (BLV) are associated with reduced expression of B cell transcriptional regulators in dairy cattle naturally infected with BLV. *Frontiers in Veterinary Science* 4 (January), 1–16.
- Gazon, H., Chauhan, P., Hamaidia, M., *et al.*, 2018. How does HTLV-1 undergo oncogene-dependent replication despite a strong immune response? *Frontiers in Microbiology* 8 (January), 1–6.
- Gillet, N., Florins, A., Boxus, M., *et al.*, 2007. Mechanisms of leukemogenesis induced by bovine leukemia virus: Prospects for novel anti-retroviral therapies in human. *Retrovirology* 4, 18.
- Gillet, N.A., Gutierrez, G., Rodriguez, S.M., *et al.*, 2013. Massive depletion of bovine leukemia virus proviral clones located in genomic transcriptionally active sites during primary infection. *PLoS Pathogens* 9 (10), e1003687.
- Gillet, N.A., Hamaidia, M., de Brogniez, A., *et al.*, 2016. Bovine leukemia virus small noncoding RNAs are functional elements that regulate replication and contribute to oncogenesis in vivo. *PLoS Pathogens* 12 (4), e1005588.
- Gutiérrez, G., Alvarez, I., Fondevila, N., *et al.*, 2009. Detection of bovine leukemia virus specific antibodies using recombinant p24-ELISA. *Veterinary Microbiology* 137 (3–4), 224–234.
- Gutiérrez, G., Rodríguez, S.M., De Brogniez, A., *et al.*, 2014. Vaccination against δ -retroviruses: The bovine leukemia virus paradigm. *Viruses* 6 (6), 2416–2427.
- Hayashi, T., Mekata, H., Sekiguchil, S., *et al.*, 2017. Cattle with the BoLA class II DRB3*0902 allele have significantly lower bovine leukemia proviral loads. *The Journal of Veterinary Medical Science* 79 (9), 1552–1555.
- Jaworski, J.P., Pluta, A., Rola-Luszczak, M., *et al.*, 2018. Interlaboratory comparison of six real-time PCR assays for detection of bovine leukemia virus proviral DNA. *Journal of Clinical Microbiology* 56, 1–11.
- Jimba, M., Takeshima, S., nosuke, Matoba, K., Endoh, D., Aida, Y., 2010. BLV-CoCoMo-qPCR: Quantitation of bovine leukemia virus proviral load using the CoCoMo algorithm. *Retrovirology* 7, 91.
- Kabeya, H., Ohashi, K., Sugimoto, C., Onuma, M., 1999. Characterization of immune responses caused by bovine leukemia virus envelope peptides in sheep. *The Journal of Veterinary Medical Science* 61 (5), 475–480.
- Kerkhofs, P., Adam, E., Droogmans, L., *et al.*, 1996. Cellular pathways involved in the ex vivo expression of bovine leukemia virus. *Journal of Virology* 70 (4), 2170–2177.
- Kerkhofs, P., Heremans, H., Burny, A., Kettmann, R., Willems, L., 1998. In vitro and in vivo oncogenic potential of bovine leukemia virus G4 protein. *Journal of Virology* 72 (3), 2554–2559.
- Kincaid, R.P., Burke, J.M., Sullivan, C.S., 2012. RNA virus microRNA that mimics a B-cell oncomiR. *Proceedings of the National Academy of Sciences of the United States of America* 109 (8), 3077–3082.
- Konnai, S., Suzuki, S., Shirai, T., *et al.*, 2013. Enhanced expression of LAG-3 on lymphocyte subpopulations from persistently lymphocytotic cattle infected with bovine leukemia virus. *Comparative Immunology, Microbiology and Infectious Diseases* 36 (1), 63–69.
- Kuczewski, A., Hogeveen, H., Orsel, K., *et al.*, 2019. Economic evaluation of 4 bovine leukemia virus control strategies for Alberta dairy farms. *Journal of Dairy Science* 102 (3), 2578–2592.
- LaDronka, R.M., Ainsworth, S., Wilkins, M.J., *et al.*, 2018. Prevalence of bovine leukemia virus antibodies in US dairy cattle. *Veterinary Medicine International* 2018, 5831278.
- Lagarias, D.M., Radke, K., 1989. Transcriptional activation of bovine leukemia virus in blood cells from experimentally infected, asymptomatic sheep with latent infections. *Journal of Virology* 63 (5), 2099–2107.
- Lefebvre, L., Vanderplasschen, A., Ciminale, V., *et al.*, 2002. Oncoviral bovine leukemia virus G4 and human T-cell leukemia virus type 1 p13l accessory proteins interact with farnesyl pyrophosphate synthetase. *Journal of Virology* 76 (3), 1400–1414.
- Meas, S., Ohashi, K., Tum, S., *et al.*, 2000. Seroprevalence of bovine immunodeficiency virus and bovine leukemia virus in draught animals in Cambodia. *The Journal of Veterinary Medical Science* 62 (7), 779–781.

- Norby, B., Bartlett, P.C., Byrem, T.M., Erskine, R.J., 2016. Effect of infection with bovine leukemia virus on milk production in Michigan dairy cows. *Journal of Dairy Science* 99, 2043–2052.
- Portetelle, D., Limbach, K., Burny, A., *et al.*, 1991. Recombinant vaccinia virus expression of the bovine leukaemia virus envelope gene and protection of immunized sheep against infection. *Vaccine* 9 (3), 194–200.
- Pyeon, D., O'Reilly, K.L., Splitter, G.A., 1996. Increased interleukin-10 mRNA expression in tumor-bearing or persistently lymphocytotic animals infected with bovine leukemia virus. *Journal of Virology* 70, 5706–5710.
- Rodríguez, S.M., Florins, A., Gillet, N., *et al.*, 2011. Preventive and therapeutic strategies for bovine leukemia virus: lessons for HTLV. *Viruses* 3 (7), 1210–1248.
- Rosewick, N., Durkin, K., Artesi, M., *et al.*, 2017. Cis-perturbation of cancer drivers by the HTLV-1/BLV proviruses is an early determinant of leukemogenesis. *Nature Communications* 8, 15264.
- Rosewick, N., Momont, M., Durkin, K., *et al.*, 2013. Deep sequencing reveals abundant noncanonical retroviral microRNAs in B-cell leukemia/lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* 110, 2306–2311.
- Ruggiero, V.J., Bartlett, P.C., 2019. Control of bovine leukemia virus in three US dairy herds by culling ELISA-positive cows. *Veterinary Medicine International* 2019.
- Ruggiero, V.J., Benitez, O.J., Tsai, Y.L., *et al.*, 2018. On-site detection of bovine leukemia virus by a field-deployable automatic nucleic extraction plus insulated isothermal polymerase chain reaction system. *Journal of Virological Methods* 259, 116–121.
- Safari, R., Hamaidia, M., de Brogniez, A., Gillet, N., Willems, L., 2017. Cis-drivers and trans-drivers of bovine leukemia virus oncogenesis. *Current Opinion in Virology* 26, 15–19.
- Safari, R., Jacques, J., Brostaux, Y., Willems, L., 2020. Ablation of non-coding RNAs affects bovine leukemia virus B lymphocyte proliferation and abrogates oncogenesis. *PLoS Pathogens* 16 (5), e1008502.
- Suzuki, S., Konnai, S., Okagawa, T., *et al.*, 2013. Expression analysis of Foxp3 in T cells from bovine leukemia virus infected cattle. *Microbiology and Immunology* 57 (8), 600–604.
- Tajima, S., Aida, Y., 2000. The region between amino acids 245 and 265 of the bovine leukemia virus (BLV) tax protein restricts transactivation not only via the BLV enhancer but also via other retrovirus enhancers. *Journal of Virology* 74 (23), 10939–10949.
- Takeshima, S.N., Sasaki, S., Meripet, P., Sugimoto, Y., Aida, Y., 2017. Single nucleotide polymorphisms in the bovine MHC region of Japanese Black cattle are associated with bovine leukemia virus proviral load. *Retrovirology* 14, 1–7.
- Twizere, J.-C., Krus, V., Lefebvre, L., *et al.*, 2003. Interaction of retroviral tax oncoproteins with tristetraprolin and regulation of tumor necrosis factor- α expression. *Journal of the National Cancer Institute* 95 (24), 1846–1859.
- Van Driessche, B., Rodari, A., Delacourt, N., *et al.*, 2016. Characterization of new RNA polymerase III and RNA polymerase II transcriptional promoters in the Bovine Leukemia Virus genome. *Scientific Reports* 6, 31125.
- Willems, L., Burny, A., Collete, D., *et al.*, 2000. Genetic determinants of bovine leukemia virus pathogenesis. *AIDS Research and Human Retroviruses* 16 (16), 1787–1795.
- Willems, L., Gatot, J.S., Mammerickx, M., *et al.*, 1995. The YXXL signalling motifs of the bovine leukemia virus transmembrane protein are required for in vivo infection and maintenance of high viral loads. *Journal of Virology* 69 (7), 4137–4141.
- Willems, L., Gégonne, A., Chen, G., *et al.*, 1987. The bovine leukemia virus p34 is a transactivator protein. *EMBO Journal* 6 (11), 3385–3389.
- Willems, L., Grimonpont, C., Kerkhofs, P., *et al.*, 1998. Phosphorylation of bovine leukemia virus Tax protein is required for in vitro transformation but not for transactivation. *Oncogene* 16 (17), 2165–2176.
- Willems, L., Heremans, H., Chen, G., *et al.*, 1990. Cooperation between bovine leukaemia virus transactivator protein and Ha-ras oncogene product in cellular transformation. *The EMBO Journal* 9 (5), 1577–1581.
- Willems, L., Kerkhofs, P., Dequiedt, F., *et al.*, 1994. Attenuation of bovine leukemia virus by deletion of R3 and G4 open reading frames. *Proceedings of the National Academy of Sciences of the United States of America* 91 (24), 11532–11536.

Bovine Viral Diarrhea, Border Disease, and Classical Swine Fever Viruses (*Flaviviridae*)

Paul Becher and Volker Moennig, University of Veterinary Medicine, Hannover, Germany
Norbert Tautz, University of Luebeck, Luebeck, Germany

© 2021 Elsevier Ltd. All rights reserved.

Classification

Bovine viral diarrhea virus 1 (BVDV-1), BVDV-2, Border disease virus (BDV) and Classical swine fever virus (previously called *Hog cholera virus*) belong to the genus *Pestivirus* within the family *Flaviviridae*. In addition, a growing number of additional pestiviruses has been identified from various hosts of the order *Artiodactyla* including pig, cattle, sheep, goat, wild boar, giraffe, Pronghorn antelope, and various other wild ruminant species. More recently, distinct pestivirus sequences have been detected in samples from several rodent and bat species. These viral sequences significantly differ from artiodactylous pestiviruses and form two separate clusters. The recently identified atypical porcine pestivirus (APPV) which causes congenital tremor in newborn piglets is most closely related to pestiviruses from bats. Taken into account the growing number and the expanded host range of genetically highly variable pestivirus species, the *Flaviviridae* study group of the International Committee on Taxonomy of Viruses (ICTV) has proposed a revision to the taxonomy of the genus *Pestivirus*. This included the modification of the criteria for pestivirus species demarcation, the creation of seven additional new pestivirus species and naming species in a host-independent manner using the format *Pestivirus X*. This revision is consistent with the recently revised format used for the genera *Hepacivirus* and *Pegivirus* of the *Flaviviridae* family. BVDV-1, BVDV-2, CSFV, and BDV are termed *Pestivirus A*, *Pestivirus B*, *Pestivirus C*, and *Pestivirus D*, respectively. The seven additional pestivirus species (with reference viruses or previously used designations provided in parentheses) are *Pestivirus E* (pronghorn pestivirus), *Pestivirus F* (Bungowannah virus), *Pestivirus G* (giraffe pestivirus), *Pestivirus H* (Hobi-like pestivirus), *Pestivirus I* (Aydin-like pestivirus), *Pestivirus J* (rat pestivirus) and *Pestivirus K* (atypical porcine pestivirus) (Fig. 1). The proposed changes only relate to the nomenclature of virus species names, while virus isolates and established pestiviruses are still referred to their original names. Pestiviruses identified from pigs (Linda virus), sheep (MG770617, MK618726), sheep and goat (Tunisian sheep pestivirus, not included in Fig. 1) as well as from bats, rodents and whales (not included in Fig. 1) may represent additional species (Fig. 1). Demarcation of pestivirus species is based on sequence relationships and antigenic relationships, and is frequently associated with differences in host range and diseases caused by these viruses. For the bat and rodent pestiviruses, virus isolates are not available and therefore antigenic relationships have not been studied.

Phylogenetic analyses of complete and partial genomic sequences revealed that CSFV, BVDV-1, BVDV-2, and BDV can be further segregated into defined genotypes and subgenotypes. The genetic variations identified in individual virus isolates may be implicated in disease control. It has been reported that vaccines and diagnostics that work well against homologous virus strains can be less efficacious for genetically different viruses. Moreover, genetic typing of CSFV and BVDV isolates can improve the understanding of epidemiological relations between outbreaks and may assist in tracing the source of disease outbreaks.

CSFV can be divided into at least three major genotypes (1, 2 and 3), each comprising three to seven subgenotypes (1.1–1.7; 2.1–2.3; 3.1–3.4). Recent studies suggested that the genetic diversity of CSFV might be broader than previously reported and proposed up to seven subgenotypes within genotype 2. For assignment of newly identified CSFV isolates to a (sub)genotype, either full-length genomic sequences or partial sequences have to be determined and subsequently compared to the available sequences of other known CSFV isolates. In addition to full-length genomic sequences, complete E2 coding sequences provide a solid basis for phylogenetic analysis and assignment to a defined subgenotype. More than 500 complete E2 coding sequences are available for CSFV. Compared to BVDV-1 and BDV, the genetic variability of CSFV is lower. During long lasting epidemics, only a very limited number of nucleotide changes have been observed.

BVDV-1 can be segregated into at least twenty-one subgenotypes (BVDV-1a-1u), while four subgenotypes have been described for BVDV-2. Epidemiological studies have demonstrated that various BVDV subgenotypes predominate in different countries and geographic regions. For BDV at least eight distinct and highly variable genotypes have been reported. Future investigations of BVDV-1, BVDV-2, BDV, and CSFV isolates may result in the identification of additional genotypes and subgenotypes.

Genome Structure and Polyprotein Processing

The pestiviral genome is a single-stranded RNA with a length of about 12.3 kb but naturally occurring virus mutants with genomes up to 16.5 kb have been described. The single large open reading frame (ORF) is flanked on both ends by untranslated regions (UTRs). The 5'UTR folds into a complex structure which is critical for RNA replication and the formation of an internal ribosome entry site (IRES) mediating internal initiation of translation. In the 3'UTR, which also contains structural RNA elements of functional importance for viral RNA replication, highly conserved binding sites for the small cellular RNAs miR-17 and let-7 have been identified. Binding of these RNAs increases genome stability and stimulates translation.

The polyprotein consists of approximately 3900 amino acids and is co- and posttranslationally processed by host- and virus-encoded proteases into 12 mature proteins. However, also cleavage intermediates play a crucial role in the pestiviral life cycle. The organization of the polyprotein is NH₂–N^{pro}/C/E^{ms}/E1/E2/p7/NS2/NS3/NS4A/NS4B/NS5A/NS5B–COOH (Fig. 2). The first protein encoded in the ORF is the N-terminal protease (N^{pro}) which cleaves off itself from the remainder of the polyprotein

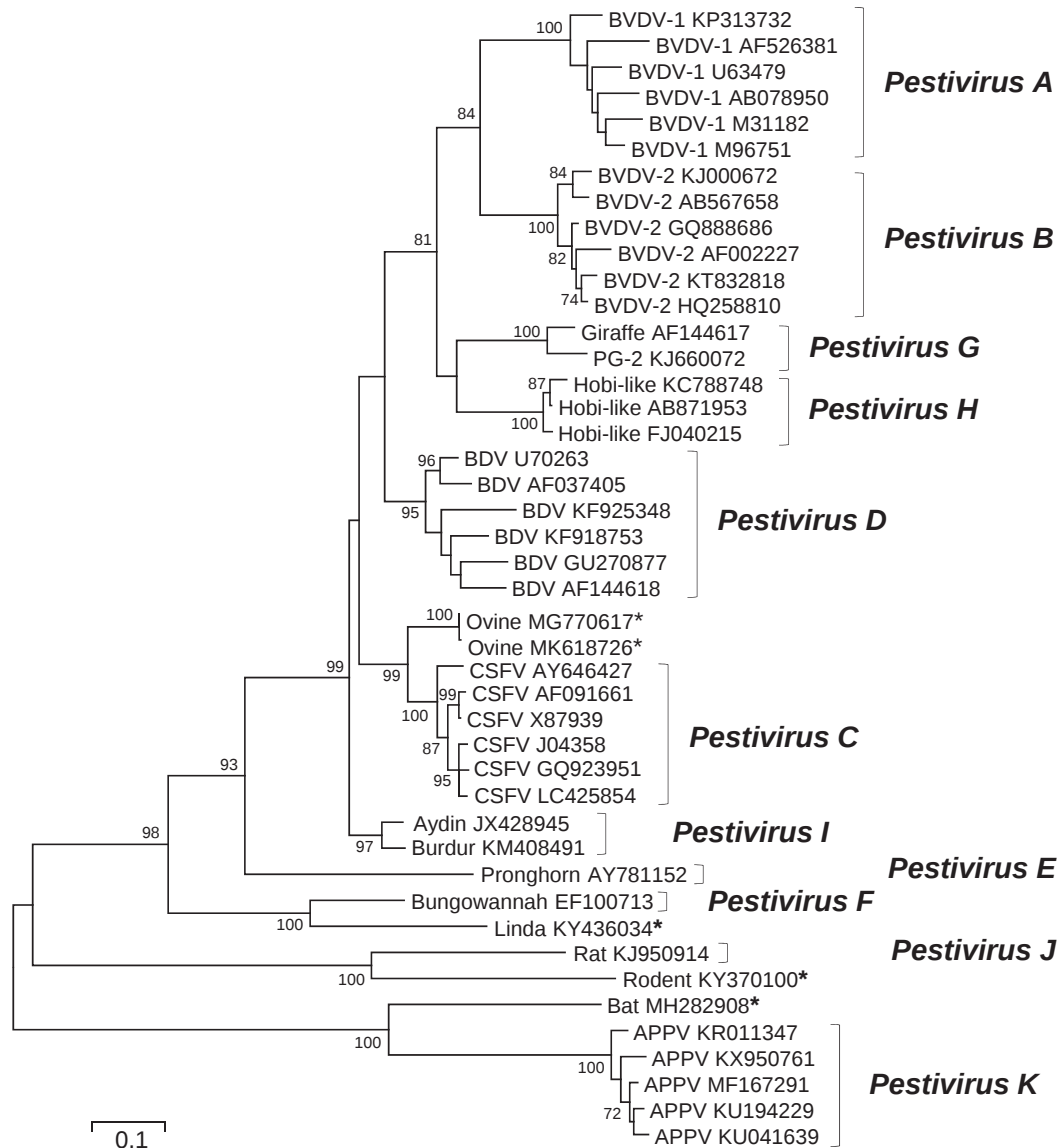


Fig. 1 Phylogenetic relationship of pestiviruses. Phylogenetic trees were constructed using MEGA 6 and based upon distances between amino acid sequences for amino acid positions 3312–3899 (NS5B region) by maximum likelihood using a JTT + G model as recommended by the Flaviviridae Study Group of the International Committee on Taxonomy of Viruses (ICTV). Accession numbers of the individual sequences allocated to the eleven established pestivirus species (*Pestivirus A-K*) or representing tentative pestivirus species (Ovine MG770617 and Ovine MK618726, Linda KY436034, Bat MH282908 and Rodent KY370100; highlighted by astrisks) are indicated. Up to 6 sequences were used for each species. Numbers indicate branches supported by >70% of bootstrap replicates. The authors thank Gökçe Nur Cagatay, University of Veterinary Medicine Hannover, Germany, for providing the phylogenetic tree.

thereby generating the N terminus of the core protein (C). The cleavages required for the release of the structural proteins C, E^{rns}, E1 and E2 as well as p7 are mediated by the cellular signal peptidase residing in the ER. The C terminus of C is further processed by cellular signal peptide peptidase. NS2 acts as an autoprotease generating its own C terminus and thereby the N terminus of NS3. NS3 is multifunctional with proven protease, helicase and NTPase activities. NS3 is believed to generate autocatalytically its own C terminus. For full protease activity NS3 forms a complex with its cofactor NS4A which intercalates with one beta-strand into the protease domain of NS3. The NS3/4A complex catalyzes all cleavages further downstream in the polyprotein.

Virion Structure

The virions of pestiviruses are enveloped and have a size of about 50 nm (**Fig. 3**). Early immuno-EM studies using monoclonal antibodies against E^{rns} and E2 have shown for CSFV and BVDV the presence of both proteins at the surface of the virion. Already in

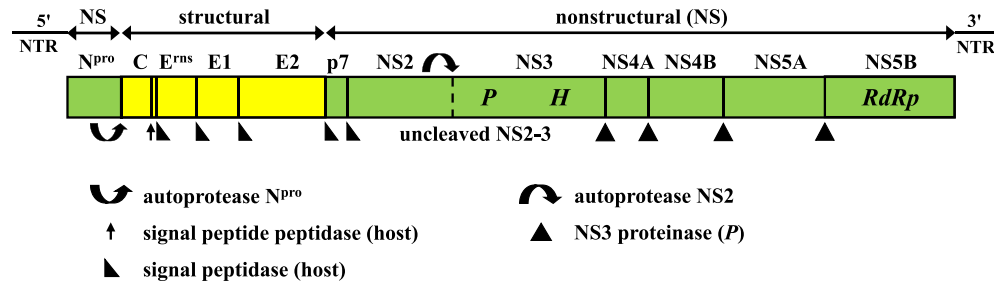


Fig. 2 Pestivirus genome organization. The genomic RNA contains one open reading frame flanked by 5' and 3' nontranslated regions (NTR). The proteinases involved in polyprotein processing are indicated. *H*, helicase; *RdRp*, RNA-dependent RNA polymerase. Modified from Becher, P., Thiel, H.J., 2011. Pestivirus. In: Tidona, C., Darai, G. (Eds.), The Springer Index of Viruses. Springer Science + Business Media, LLC. doi:10.1007/978-0-387-95919-1.



Fig. 3 Extracellular virions on MDBK cells infected with giraffe pestivirus (Pestivirus G), 15 h p.i., bar: 100 nm. Reproduced with permission from Schmeiser, S., Mast, J., Thiel, H.J., König, M., 2014. Morphogenesis of pestiviruses: New insights from ultrastructural studies of strain Giraffe-1. Journal of Virology 88, 2717–2724.

those studies E^{ns} was preferentially detected. A more recent, detailed structural and biochemical analysis of purified BVDV virions revealed a generally low envelope protein content with E^{ns} being more abundant than the other envelope proteins. The averaged image of the virions did not show glycoprotein spikes even after acidification of the sample, indicating that these proteins may not form a regular structure at the virion. Alternatively, those structures are not visualized due to their low abundance. The envelope proteins detected in the virions were covalent E^{ns} homodimers as well as E1/E2 heterodimers but hardly any monomers or E2 dimers. Even though the virions bud from the endoplasmatic reticulum (ER), their lipid composition differed from the one of the ER membrane indicating lipid-sorting during virion morphogenesis. In contrast to hepatitis C virus (HCV), the virions of pestiviruses showed no apolipoprotein association.

While E2 is the receptor binding protein and most neutralizing antibodies are directed against this component of the virion, a limited degree of neutralization was also observed for antibodies against E^{ns}. Structural information for E2 of BVDV has been obtained in two studies. These structures revealed no similarity of E2 to the class II fusion proteins of the alphaviruses or members of the genus *Flavivirus*. Therefore, it remains to be finally determined whether E2 or E1, the least well investigated glycoprotein of pestiviruses, act as fusion protein in the process of cell entry.

A duplication of the core protein as well as fusions of core with YFP (yellow fluorescent protein) did not lead to abrogation of virion morphogenesis. Surprisingly, even the deletion of C could be partially compensated by single mutations located in the C-terminal region of NS3. Taken together, the enormous flexibility of C with respect to formation of infectious virions may indicate that C does not form a rigid icosahedral structure in the virion but acts as an RNA binding protein with a potential RNA chaperone activity. This assumption is further corroborated by the fact that pestiviral genomes with massively increased genome size can still be efficiently packaged into infectious viral particles. For a final clarification of the architecture of the pestiviral capsids, a high resolution structure will be required.

E^{ns} is unique to pestiviruses and its most unusual feature is its RNase activity. Structural analysis further verified the postulated similarity to T2 RNases. The RNase activity is neither required for the infectivity of the virion nor for viral replication in cell culture. However, inactivation of the RNase led to an attenuation of BVDV in its natural host. Besides being part of the infectious virion, E^{ns} is also present in significant amounts in the serum of infected animals. Detailed analyses have shown that an unusual in plane membrane anchor is critical for the shedding of E^{ns} from infected cells. This alpha-helix also seems to play a distinct role in the formation of infectious viral particles. Furthermore, E^{ns} was shown to act as an antagonist of the innate host response to infection. In cell culture experiments E^{ns} was able to suppress the induction of interferon- β and interferon-induced protein expression after stimulation by extracellular dsRNA or ssRNA. The RNase activity of E^{ns} is required for its immunosuppressive effect. Secreted E^{ns} can bind to cells, be taken up by Clathrin-dependent endocytosis and interfere with IFN induction by extracellular dsRNA. Viral RNA shedded from pestivirus infected cells, most likely via exosomes, was shown to induce IFN secretion by plasmacytoid

dendritic cells (PDCs). This stimulation of PDCs could be antagonized by E^{ms} via an RNase-dependent mechanism. For the *in vivo* function homodimerisation of E^{ms} via a disulfide bridge seems to be of importance since a strong positive selection for the cysteine critical for dimerization has been observed. The cleavage of ss or dsRNA by E^{ms} does, however, not depend on its dimerization.

Pestiviral Nonstructural Proteins and the Regulation of RNA Replication and Virion Morphogenesis

N^{pro}

N^{pro} is the only viral protein dispensable for RNA replication and virion morphogenesis. However, it is one of the major repressors of Interferon-1 (IFN-1) induction in pestivirus infected cells. Recently the crystal structures for N^{pro} of a HoBi-like BVDV-3 strain and of CSFV strain Alfort 187 were reported. These studies did finally prove that the protease activity is based on a catalytic diad composed of His49 and Cys69. Moreover, the structural information depicted that the C terminus of N^{pro} is locked into the catalytic site after autocleavage, thereby disabling any trans-cleavage activity of this enzyme. In line with these findings, the protease activity of N^{pro} is not required for the suppression of IFN-1 response. N^{pro} interacts directly with interferon regulatory factor 3 (IRF3), one of the major transcription factors responsible for the transcription of IFN- α/β mRNA. For this interaction a structurally important TRASH motif coordinating a Zn ion, is essential. N^{pro}/IRF3 interactions finally lead to an ubiquitin-dependent proteasomal degradation of IRF3 which strongly interferes with the induction of an IFN- α/β response in pestivirus infected cells.

p7

The small hydrophobic protein p7 in part remains uncleaved at the C terminus of E2 while most p7 gets released by signal peptidase cleavage. p7 is so far not detected in virions and thus classified as nonstructural protein. The E2-p7 precursor is not essential for the viral life cycle, at least in cell culture. Due to its sequence characteristics a viroporin function was postulated for p7 and finally demonstrated. Together with results obtained by studying p7 of HCV the ion channel activity of p7 may be required to protect immature intracellular virions from inactivation by acidification. In addition to its viroporin function, p7 may be involved in critical protein/protein interactions during virion morphogenesis as observed in the HCV system.

NS2

NS2 has a length of about 450 amino acids and thus about twice the size of NS2 of HCV. NS2 is composed of a hydrophobic N-terminal half anchoring the protein to membranes and a C-terminal protease domain with a putative structural Zn-binding site residing in the cytoplasm. However, so far neither the membrane topology nor the structure of the protein has been clarified. The activation of the cysteine protease domain requires an interaction with the cellular chaperone DNAJC14 to become active thereby releasing itself from NS2-3. The finding that NS2 becomes only active for cleavage *in trans* after C-terminal truncation suggests that the C terminus of the protein remains after autocleavage in the active site rendering the cleavage product enzymatically inactive. Actually, the crystal structure of the HCV NS2 autoprotease revealed such an autoblockade, which is also described above for N^{pro}. At the moment no function of the released NS2 has been identified. However, in the context of NS2-3 the NS2 part plays an essential role in virion morphogenesis (see below).

NS3

NS3 is an essential component of the viral replicase which can't be functionally complemented by uncleaved NS2-3. It is a multifunctional protein with an N-terminal serine protease domain and a helicase/NTPase domain connected by a flexible linker. The crystal structure of a single chain NS3-4A protease of CSFV verified that, analogous to findings from the HCV system, NS4A intercalates with its central peptide into the beta-barrel domain of the NS3 protease. Mutations based on the closed conformation structure of CSFV NS3/4A revealed a critical role for the interaction between helicase and protease domain in RNA replication. Only in complex with NS4A the NS3 protease gains its full protease activity and can process e.g., the NS5A/B site. Interestingly, for NS3/4A a surface interaction between NS3 and the part of NS4A located downstream of the central peptide could be visualized in the crystal structure. Mutagenesis studies demonstrated that inactivation of protease, helicase or NTPase function of NS3 interferes with viral RNA replication. For the most C-terminal part of NS3, namely helicase subdomain 3 a surprising compensatory role in CSFV variants with a core deletion was described. Single amino acid exchanges were able to partially rescue the core deletion variants with respect to virion morphogenesis. However, those viruses were attenuated in their natural host. These findings raise the question whether this domain may also be involved in virion morphogenesis in wild type pestiviruses, especially since in this virus system all nonstructural proteins, besides N^{pro}, are required for this process. In this context it is remarkable that the members of the genus *Pegivirus* do not encode a core protein, but still produce infectious progeny. Moreover, also for other members of the *Flaviviridae* family mutations located in the helicase were able to rescue virion morphogenesis deficiencies.

NS2-3

Cleavage of NS2-3 is mediated by the autoprotease in NS2. Since NS3 but not NS2-3 is an essential component of the viral replicase the release of NS3 is of regulatory importance for viral RNA replication efficiency. The NS2 autoprotease depends in its activity on the stable interaction with DNAJC14. This chaperone is available only in low amounts in part due to the existence of an upstream ORF in the mRNA. Accordingly, NS2-3 produced at early times after infection will form a complex with DNAJC14 leading to activation of the NS2 protease and NS3 release. However, when translation of the viral polyproteins proceeds, NS2-3

translated at later time points will not be able to recruit DNAJC14 from the depleted intracellular pool and therefore will not undergo cleavage. Since NS3 is an essential component of the viral replicase this mechanism leads to a temporal downregulation of viral RNA replication which has been shown to be required for the noncp biotype of pestiviruses which is in turn a prerequisite for viral persistence in the animal. A shRNA-mediated knock down of the DNAJC14 mRNA in MDBK cells severely reduced the titers of BVDV-1. However, with this approach no conclusive data were obtained for other pestivirus species. By generating bovine and porcine DNAJC14 knock-out cells it was finally possible to prove that the replication of 6 distinct noncp pestivirus species, namely species A-D, F and G strictly depends on DNAJC14. Furthermore, this study allowed the conclusion that all cytopathogenic pestiviruses which are often products of RNA recombination replicate independent of DNAJC14. Those viruses are characterized by an overexpression of NS3 which is correlated with upregulation of viral RNA synthesis and cytopathogenicity.

NS2-3 is a cleavage intermediate in polyprotein processing. A product-precursor relationship between NS3 and NS2-3 was observed for the noncp CSFV strain Alfort Tübingen. Incomplete cleavage results from DNAJC14 deprivation in the infected cells (see above). In all cells infected with noncp pestiviruses significant amounts of uncleaved NS2-3 are detected. This fact is in line with the observation that uncleaved NS2-3 is essential for the process of virion morphogenesis. This function of NS2-3 became apparent when ubiquitin was inserted between NS2 und NS3 in the context of an infectious cDNA clone of BVDV-1 strain NADL. In such a polyprotein cellular ubiquitin specific C-terminal hydrolases process the C terminus of ubiquitin resulting in complete cleavage of NS2-3 which correlated with a defect in virion morphogenesis. An analogous result was obtained for a virus mutant with an IRES insertion between the sequences coding for NS2 and NS3. These findings could be reproduced for CSFV. The latter study revealed that virion morphogenesis of these mutants could be rescued by authentic NS2-3/4A complexes supplied *in trans*, in which, however, none of the enzymatic activities of NS2 or NS3 was required.

When a cell culture system for HCV became available similar studies led to the surprising result that virion morphogenesis of HCV does not depend on uncleaved NS2-3. This finding inspired research on pestiviral gain of function mutants capable of virion production in the absence of uncleaved NS2-3 which mimics the situation in the HCV system. These studies led to the identification of two gain of function mutations, one in NS2 and one in NS3 which were able to compensate the loss of virion morphogenesis due to an ubiquitin-insertion between NS2 and NS3 in BVDV. These mutants and the structural information obtained by the crystal structure of a NS3-4A single chain protease were used to address the question how NS3/4A complexes are functionalized for either RNA replication or virion morphogenesis. Reverse genetics and biochemical studies revealed that a surface interaction between NS3 and the C-terminal part of NS4A (kink region) is decisive whether a NS3/4A complex functions as a basic building block for viral replicase assembly or in the pathway of virion morphogenesis. Further studies are needed to understand those pathways in more detail.

NS4A

NS4A is a small protein which acts in the NS3/4A serine protease complex as protease cofactor. The central domain as well as the kink domain are required for the stimulation of the NS3 protease activity. The N-terminal transmembrane domain of NS4A serves as membrane anchor for the NS3/4A complex. *Trans*-complementation studies demonstrated that NS4A is also involved in virion morphogenesis. Surface interactions between NS4A and NS3 further functionalize the protein (see above).

NS4B

NS4B is a hydrophobic protein of about 35 kDa which is essential for RNA replication. However, structure and function of NS4B are largely unknown. Recently, a unique ER-like localization was described for NS4B of CSFV. In the N-terminal region of NS4B an amphipathic α -helix structure was determined which serves as membrane anchor. Mutations in this protein region could be correlated with RNA replication efficiency and virulence of CSFV. For mutations in a sequence motif indicative of an NTPase no conclusive *in vivo* data were obtained with respect to CSFV replication and virulence.

NS5A

NS5A is a phosphoprotein with a size of about 58 kDa. It is anchored to membranes via an N-terminal amphipathic α -helix. A three domain structure with intervening low complexity sequences (LCS) was postulated for NS5A of BVDV. Domain I is essential for RNA replication and contains a structural Zn-binding motif. The downstream LCS I and domain II were more tolerant to deletions/insertions leading to the establishment of a reporter replicon with a mCherry-insertion in this part of the protein. Using those replicons BVDV NS5A was found to localize to lipid droplets, an analogy to the HCV system. NS5A is so far the only replicase component that can be complemented *in trans*. For NS5A a direct interaction with and a regulatory effect on NS5B has been described and relevant amino acids have been identified. Interestingly, those residues were found to be conserved between pestiviruses and HCV.

NS5B

NS5B has RNA-dependent RNA polymerase (RdRp) activity. However, for replication of the viral RNA *in vivo* at least viral proteins NS3 to NS5B are required. The RdRp activity of NS5B can be stimulated by GTP and is capable of starting *de novo* RNA synthesis. Structural information has been obtained for NS5B of BVDV and more recently also for CSFV. The crystalized CSFV NS5B structure included a novel N-terminal domain. NS5B of BVDV has also been shown to be involved in virion morphogenesis.

Life Cycle

Cell culture studies have shown that pestiviruses attach via E^{tns} to heparan sulfate and use pH-triggered clathrin-mediated endocytosis for cell entry. Since pestiviruses are acid resistant, an additional activation step was postulated. CD46 was identified as receptor for BVDV-1 with E2 as viral interaction partner. However, expression of bovine CD46 did not confer susceptibility to non-susceptible cell lines indicating the existence of critical co-receptors. Remarkably, direct cell to cell transmission was described for BVDV even in the presence of neutralizing antibodies directed against E2. CD46 is dispensable for this process, which still employs clathrin-mediated endocytosis. Cell to cell spread may also be involved in the transmission of noncp pestiviruses over the placenta, which is required for establishing viral persistence in the fetus. For CSFV a Caveolar-dependent uptake into Alveolar macrophages has been reported involving Rab5, Rab7, and Rab11. Thus different entry pathways may be used depending on the cell type. Replication and virion formation are believed to take place at ER-derived membranes. Co-localization of the viral core protein and the envelope glycoprotein E2 with the ER marker protein disulfide isomerase (PDI) was demonstrated by immunogold labeling of cryosections while NS5A was found by *in vivo* imaging at lipid droplets which also reside in close proximity to the ER. The efficiency of RNA replication is regulated via the extent of NS2-3 cleavage (see above) and correlates with the two biotypes observed in cell culture. While the noncp biotype is capable of persistence, the cp biotype produces higher amounts of intracellular RNA which correlates with viral cytopathogenicity, rendering this biotype incompetent for establishing persistent infections. Intracellular virions were found during exocytosis in transport vesicles and the Golgi apparatus. Significant titers of infectious particles are released at about 12 h post infection. The detection of core and double-stranded RNA in multi vesicular bodies may be an indication for an involvement of the endosomal sorting complexes required for transport (ESCRT) machinery in virion budding and secretion.

Epidemiology

Classical Swine Fever

The natural hosts of CSFV are domestic pigs and wild boar (*Sus scrofa ferus*). In endemic situations CSFV can be transmitted between domestic pigs and wild boar and vice versa. Other members of the *Suidae* family like common warthogs (*Phacochoerus africanus*) and bushpigs (*Potamochoerus larvatus*) are also susceptible to infection with CSFV. For a long time, it was commonly accepted that cattle can be infected with CSFV only under experimental conditions, while a recent report suggested the occurrence of natural CSFV infections of cattle in China. The significance of CSFV infection in cattle and, possibly, other ruminants remains to be elucidated.

CSF is still endemic or sporadically occurs in large parts of the world. According to recent official reports to the OIE (since 2014), CSF is still present in China, Russia, South-East Asia, the Caribbean as well as several countries of South-America. While the African continent has no official OIE status for CSF, recent reports suggest the presence of the disease in at least some regions (e.g., Madagascar). A complete list of currently CSF-free countries is published by the OIE. Other regions of the world including Australia, New Zealand, United States of America, and Canada have been CSF-free for at least forty years. In the European Union (EU) there were only a few sporadic outbreaks of CSF in the Eastern parts of the EU during the past ten years. The last CSF outbreaks in the EU occurred in Lithuania (2009, 2011) and in Latvia (2012–2015).

A detailed knowledge of the different routes of transmission is of particular importance to understand the epidemiology of CSF and to implement suitable and efficient control measures for eradication of the disease, surveillance and maintenance of a CSF-free status. CSFV can be transmitted by horizontal and vertical routes. Infected pigs shed high titers of infectious virus via saliva, lacrimal secretions, feces, urine, and semen. Meat, blood and other products of infected pigs can also contain high viral loads. After introduction into a pig farm CSFV spreads by direct contact. These horizontal infections mainly occur through the oronasal route. Spread of CSFV within a herd or within a wild boar population progresses rather slowly. Consequently, the disease may remain unrecognized for long time periods after introduction of CSFV into a holding. The occurrence of less virulent CSFV strains during the past decades is another important factor contributing to longer high risk periods. In addition to direct contact between infected and susceptible animals, CSFV can be efficiently transmitted via contaminated materials (e.g., cloth, shoes, vehicles, agricultural and veterinary equipment) and contaminated food and waste. While swill feeding is prohibited in most countries with industrial pig production, it still represents a common practice for feeding of backyard pigs in large parts of the world. The (illegal) import of virus-contaminated food represents a major threat for introduction of the disease in CSF-free countries. The recent re-occurrence of CSFV in Japan (2018) after 26 years of absence provides an illustrative example how easy CSFV can be introduced even in a country with high biosecurity measures and subsequently can spread among domestic pigs and wild boar populations.

Vertical transmission of CSFV from pregnant sows to fetuses can occur throughout all stages of gestation and can result in the birth of persistently infected offspring. These persistently infected pigs shed the virus continuously or intermittently for the rest of their life and represent a permanent source for transmission of CSFV to susceptible animals of the same herd or to other pig holdings. The occurrence of persistently infected offspring is also known for wild boar and together with the decreased virulence of CSFV observed during the past decades may be implicated in the establishment of endemic situations in wild boar populations representing a constant risk of virus transmission to domestic pigs.

Bovine Viral Diarrhea and Border Disease

The host range of BVDV-1, BVDV-2, and BDV comprises cattle, sheep, goats, domestic pigs, wild boar as well as many other animal species of the order *Artiodactyla*. BVDV-1 and BVDV-2 mainly infect cattle, while BDV has been detected most frequently in sheep. In some countries infections of sheep and goat with BVDV-1 or BVDV-2 are as frequent as infections with BDV. All these ruminant pestiviruses can cause Border Disease in sheep. Infections of cattle with BDV represent very rare events. Transmission of ruminant pestiviruses, in particular BDV, to pigs can interfere with CSF diagnosis, mainly due to the induction of cross-reacting antibodies.

Bovine viral diarrhea is endemic and widely distributed in all parts of the world. In contrast to CSF, only a few countries in Europe have implemented effective BVD control and eradication programmes resulting in a BVD-free status (Sweden, Norway, Finland, Denmark) or in a significantly reduced prevalence of BVDV in cattle (Switzerland, Germany, parts of Austria and Northern Italy).

As described above for CSFV, ruminant pestiviruses like BVDV-1, BVDV-2 and BDV can be transmitted horizontally and vertically. Horizontal infections can occur via direct and indirect contact. The viruses are excreted via nasal and oral fluids, lacrimal secretions, urine, feces, and semen. The main route of infection is oronasally.

Vertical transmission of BVDV-1, BVDV-2, BDV, and other ruminant pestiviruses from pregnant cattle (days 40–125 of gestation) and sheep (days 16–80 of gestation) to the fetuses can result in a specific acquired immunotolerance against the infecting virus leading to persistent infection of the offspring. These persistently infected (PI) animals are unable to mount an effective humoral and cellular immune response against the persisting virus. PI animals either develop severe disease (see below) or can survive without apparent clinical signs of disease for many years. They shed constantly large amounts of viruses until death and represent the main source of transmission of the virus to susceptible animals within a herd. Accordingly, these PI animals are significantly implicated in the spread of these viruses to other herds, other susceptible hosts, and to other countries.

Clinical Features and Pathogenesis

Classical Swine Fever

Infection with CSFV can result in acute, chronic, and *late onset* forms of CSF, which are characterized by a wide range of clinical symptoms. All forms of CSFV can be observed in both, domestic pigs and wild boar. CSFV infection usually results in fever, thrombocytopenia, leucopenia, and immunosuppression favoring secondary infections. The severity of disease mainly depends on age and immune status of the host animal and virulence of the virus. The disease is more severe and often fatal in young animals, while older breeding animals usually show milder symptoms. Highly, moderately and low virulent viruses can be distinguished. The incubation period of CSF in individual animals is three to ten days. After introduction of low and moderately virulent CSFV in large holdings the disease may remain unrecognized for several weeks. Slow transmission from animal to animal, absence of severe disease and lack of awareness represent significant factors implicated in such high risk periods and can contribute to unrecognized spread of CSF to other farms.

Acute CSF

After infection with moderately and highly virulent CSFV strains piglets and young animals usually develop high fever ($>40^{\circ}\text{C}$) and various clinical signs. In the initial stage of the disease animals may show anorexia, lethargy, conjunctivitis, respiratory and gastrointestinal symptoms. In addition, incoordination of movement, weakness of hind legs, convulsions and other neurological symptoms can be frequently observed. In typical cases of CSF, hemorrhages of the skin complete the clinical picture of the disease. Hemorrhages can be observed during the second week and later stages after infection and most frequently occur on the ears, abdomen, tail, and distal parts of the limbs. In piglets, fatality is usually high and can be up to 100%. In fattening and breeding pigs clinical disease is less severe and less specific. Such older animals can mount an effective immune response resulting in overcoming the virus and recovery from the disease.

Chronic CSF

A minority of infected animals is not able to develop an effective immune response and remain chronically infected. Such chronically infected animals constantly shed infectious virus. While virus-specific antibodies are produced by these animals, they fail to eliminate the virus. The animals show symptoms similar to those observed in acute CSF, but the disease in its initial stage is milder. Later, the disease is characterized by rather unspecific clinical signs including chronic enteritis, wasting, and intermitting fever. Secondary infections with other pathogens frequently occur and can contribute to the clinical signs of the disease. Chronic CSF is always fatal and the animals usually die after two to four months of infection.

Late Onset CSF

As a common feature of pestiviruses, CSFV is able to cross the placenta and infect the fetus. The consequences of diaplacental infections mainly depend on the time of gestation and on viral virulence. Infections during early pregnancy result in abortion, stillbirth, malformations and mummification. Intrauterine infections between days 50–70 of pregnancy with low and moderately virulent CSFV strains which do not kill the pregnant sows can result in the birth of persistently infected offspring and late onset CSF. The persistently infected animals are characterized by an acquired immune tolerance against the infecting virus and shed high

amounts of infectious CSFV for the rest of their life. Animals can appear clinically healthy for longer time periods before they develop a chronic disease condition. Infected animals show poor growth, chronic wasting or occasionally congenital tremor and usually die within the first year of life. Late onset CSF is always fatal. Moreover, it has been reported that early post-natal infections can also result in CSFV persistence.

Bovine Viral Diarrhea and Mucosal Disease

BVDV infection can result in a wide range of clinical consequences. Acute infections of immunocompetent non-pregnant animals usually lead to no or only mild enteric or respiratory disease and frequently remain unnoticed. However, a minority of highly virulent BVDV-1 and BVDV-2 strains can cause a hemorrhagic syndrome with high mortality. BVDV infections of pregnant animals result in transmission to the fetus and a number of severe consequences including embryonic death, abortion, stillbirth as well as malformations and persistent infection of the offspring (see below). Accordingly, the outcome of BVDV infections in cattle is mainly determined by the route of infection (horizontal versus vertical), the immunological status of the host animal, and the virulence of the virus.

Acute BVDV infection

In immunocompetent cattle, acute BVDV infections result in a transient viremia accompanied by fever, transient leukopenia, lymphopenia and immunosuppression. Viremia starts at 3 days after infection and usually lasts up to two weeks until a specific immunity leads to elimination of the virus. In rare cases viremia can last up to three weeks. Clinical presentations of acute BVDV infections can include mild diarrhea or respiratory disease, but in many cases such infections remain unnoticed. BVDV induced immunosuppression decreases the capacity of the immune system to control other viral and bacterial pathogens. Consequently, infections with other infectious agents can result in more severe disease of the respiratory or gastrointestinal tract. In rare cases, highly virulent BVDV strains (mainly BVDV-2) can cause severe thrombocytopenia with hemorrhages and acute fatal disease.

Field BVDV isolates can be categorized into noncp and cp biotypes according to their effects in cell culture. In contrast to cp BVDV, noncp BVDV isolates interfere with the host innate immune response by inhibition of interferon type 1 mediated antiviral defense. Furthermore, it has been reported that cp BVDV strains are faster and more efficiently eliminated by the host compared to noncp BVDV strains. Consequently, cytopathogenicity of BVDV is not associated with increased virulence in acutely infected immunocompetent animals.

Fetal infection with BVDV

Horizontal transmission of BVDV from pregnant cattle to the fetus can result in different severe consequences depending on the age of the fetus at the time point of infection. During the first weeks of pregnancy, prior to the development of cotyledons, BVDV cannot infect the embryo as the virus is not able to penetrate the zona pellucida. After day 30 of gestation infection of the embryo can result in embryonic death which may represent an important cause for reduced conception rates of dams. Fetal death is most frequently observed after infection of the dam during early gestation, but can occur at all stages of gestation. Another important consequence after infection of the fetus with noncp BVDV between 40 and 125 days of gestation is the establishment of persistent infection of the fetus resulting in the birth of persistently infected (PI) calves. After infection of the dam between 80 and 180 days of gestation, the virus is transmitted to the fetus during the period of organogenesis and can cause cerebellar hypoplasia, ocular degeneration, brachygnathism, and skeletal malformations. After infection during the last trimester of gestation the fetus can mount an effective immune response including neutralizing antibodies capable of elimination of the virus. Such infections usually do not cause disease and the respective animals are apparently healthy at birth.

Persistent BVDV infection

Infection of the fetus with noncp BVDV at a stage prior to immunocompetence of the fetus can result in a highly specific immunotolerance against the infecting virus and the birth of PI calves. As a consequence, persistently infected animals are not able to produce virus-specific antibodies and to eliminate the virus. The ability of noncp BVDV to inhibit the induction of type 1 interferon in the fetus is a crucial prerequisite for the establishment of lifelong persistent infection. PI animals are efficient virus producers, shed large amounts of viruses in all secretions and excretions and thus are significantly implicated in the spread of BVDV. PI animals can either appear clinically healthy or more frequently present as weak calves with various clinical signs due to increased susceptibility to secondary infections with other infectious agents. Another consequence is the onset of mucosal disease.

Mucosal disease

Only persistently infected animals may develop mucosal disease. Affected animals show acute severe diarrhea, dehydration, as well as erosion and ulceration of the mucosa of the oral cavity, esophagus, and gastrointestinal tract. Mucosal disease frequently occurs within the first two years of life and is always fatal. Affected animals die within a few days or a few weeks after onset of disease. In mucosal disease affected animals, a cp BVDV variant is always present in addition to the persisting noncp BVDV. Accordingly, the emergence of cp BVDV in animals persistently infected with noncp BVDV is crucially implicated in the pathogenesis of mucosal disease. In most cases the cp viruses emerge in the PI animal by RNA recombination including deletions and duplications of viral sequences as well as insertions of cellular RNA sequences. In rare cases point mutations in the NS2 protein coding region can result

in the emergence of cp BVDV. Alternatively, superinfection with an antigenically highly similar cp BVDV strain can directly cause mucosal disease within a few weeks after superinfection. A so called late onset mucosal disease results from RNA recombination between a superinfecting antigenically less similar cp BVDV strain with the persisting noncp BVDV strain and usually occurs about 3–4 months after infection with a cp BVDV field strain or after vaccination with a live attenuated cp BVDV vaccine.

Border Disease

Border disease (BD), also known as *hairy shaker disease*, was first described in the late 1950s. Again, intrauterine infections of pregnant sheep and goat are by far more important than acute infections of immunocompetent animals. Intrauterine infection during early pregnancy can cause fetal death, mummification, abortion, and stillbirth. Intrauterine infections during the 17th and 80th day of gestation can result in the birth of lambs with tremor, ataxia, brain malformations, hairy fleece, and poor growth. Alternatively, apparently healthy lambs can be born. Both clinically affected and apparently healthy lambs can be persistently infected. As described above for CSF and BVD, persistently infected animals constantly shed large amounts of infectious virus and therefore play an important role in virus transmission. In rare cases, PI lambs which recovered from disease can develop a syndrome similar to mucosal disease. Infections of immunocompetent sheep and goat mostly lead to either no or mild transient clinical signs.

Prevention and Control

In order to prevent economic losses caused by pestiviral infections and to prevent spread of the disease in domestic or wild animals control measures are often applied. Vaccination, eradication or a combination of both are common methods. There are several ways to prevent pestiviral infections and different strategies have been developed. Reliable laboratory diagnostic tests and efficacious vaccines are prerequisites of efficient preventive control measures.

Diagnosis

Acute infection with pestiviruses results in transient viremia. Since these viruses grow readily in cell cultures of their species of origin, classical diagnostic methods were based on the isolation of viruses from blood or organs of infected animals. Isolated cp BVDV can be detected using plaque assays, while noncp viruses can be visualized either by immunofluorescence or enzyme immuno techniques.

Since tissue culture is time consuming and expensive, rapid screening tests have been developed. Immunohistochemistry using mono- or polyclonal antibodies can be applied in post-mortem samples of infected animals.

Antigen-capture enzyme-linked immunosorbent assays (AgC-ELISA) are being used to diagnose cattle persistently-infected with BVDV. Since monoclonal antibodies are used in these test systems they are highly specific. Either blood or ear-notch samples are suitable for these tests. The most common method for the rapid diagnosis for ruminant and porcine pestiviruses is RT-PCR, which is applied using either individual or pooled samples of blood, milk, other body fluids, ear notches of cattle or organ suspensions.

In order to achieve a more precise characterization, viruses are often sequenced. Either parts or the whole genome are used, respectively. By comparing the sequence data with published sequences in data bases the information can be used to obtain detailed taxonomic or epidemiological insights.

As the result of acute infection or vaccination animals seroconvert. Classical methods for detecting virus-specific antibodies are virus neutralisation assays. These assays are highly specific and sensitive and they allow serological differentiation of viral strains. However, since they are based on the use of cell culture, they are expensive, time consuming and labor-intensive. Today these assays are still used for reference purposes. For the rapid and large scale screening of samples a number of alternative tests, in particular ELISAs for the detection of antibodies against structural and/or non-structural viral proteins are being used. Variants of these tests are direct, indirect or competition or double antigen ELISAs. Recombinant viral proteins and monoclonal antibodies have been used to improve sensitivity and specificity of these assays. All these tests have in common that they are relatively inexpensive and they are suitable for mass screening because they can be automated. Antibody detection can be performed in individual or pooled serum samples as well as in bulk milk samples of ruminants. Despite good characteristics of these ELISAs with respect to sensitivity and specificity, positive results are sometimes required to be confirmed by virus neutralization test.

Vaccines

Vaccines have been used as one of the most important control tools against pestiviral infections. Today there are a number of vaccines available, in particular against BVD and CSF.

Vaccines against classical swine fever

Currently modified live vaccines (MLV) are the vaccines of choice for the prevention of CSF. Conventional MLVs, e.g., the lapinised C-(China) strain of CSFV, are safe, highly efficacious and relatively inexpensive. However, when using these vaccines a serological distinction between vaccinated and infected animals is not possible. This might impede an organized effort for the control of the infection. Therefore, numerous efforts were made to develop DIVA (Differentiating Infected from Vaccinated Animals) vaccines, which would allow the distinction of naturally infected and vaccinated pigs. First DIVA vaccines were based on the immunogenic

E2 envelope protein of the virus, which had been expressed in baculoviruses. These vaccines were on the market for a few years, however, they have been abandoned because their efficacy was inferior to MLV. Nevertheless, at least one E2 subunit vaccine has recently been licensed in China. Moreover, considerable efforts have been undertaken to develop live DIVA vaccines. Current live DIVA vaccines have been developed using recombinant techniques, e.g., by constructing a chimeric virus using a BVDV backbone and the immunogenic E2 envelope protein of CSFV. Other vaccines are based on the use of adenovirus vectors. In parallel serological tests were developed to discriminate infected from vaccinated animals.

Vaccines against bovine viral diarrhea

First MLVs against BVD were already developed in the 1960ties. They were highly efficacious, however, they could not be used in pregnant cattle, because transplacental transmission of vaccine virus often produced damage comparable to natural infection with BVDV. In response to these safety issues killed vaccines were developed, which in general had a very good safety record and animals of all ages could be vaccinated including pregnant animals. However, killed vaccines had to be administered repeatedly and the immune response was inferior to that of MLVs. In consequence fetal protection of pregnant animals was often insufficient. BVDV infection of vaccinated pregnant cattle often interfered with pregnancy and the whole spectrum of BVDV-related malformations was observed including the birth of PI calves. A rare phenomenon associated with one particular inactivated vaccine was observed in 2007 and subsequent years in Europe. A new disease “bovine neonatal pancytopenia (BNP)” leading to the death of newborn calves emerged. Epidemiological studies indicated a connection between this new syndrome and vaccination with an inactivated vaccine against BVD. In serum and colostrum of dams which had given birth to calves with BNP, alloantibodies reacting with blood leukocytes of calves were detected. Apparently impurities derived from bovine cells or fetal calf serum that was used for the production of the vaccine virus were presented by a powerful adjuvant in the vaccine thereby eliciting the production of alloantibodies. The vaccine was withdrawn from the European market. Today about 20 predominantly killed vaccines are licensed in the EU. All but one product are directed against BVDV-1 only. In North America close to 200 BVD vaccines, killed and live and sometimes in combination with other antigens, are on the market. Some of the vaccines are directed against BVDV-1 and BVDV-2. Since a few years, a novel MLV is on the market that is based on a combination of deletion mutants of noncp BVDV-1 and BVDV-2 both lacking N^{pro} and the ribonuclease activity of E^{ms}. This MLV is safe to administer to all age classes of cattle including pregnant animals and has been shown not to establish an infection in foetuses.

The main purpose of vaccination against BVD is to achieve a strong fetal protection in order to prevent the production of new PI animals. Vaccines with this potency are well suited to also prevent all other consequences of acute BVDV infection, i.e. transient immunosuppression and subsequent clinical or subclinical symptoms.

Control of Classical Swine Fever in Domestic Pigs

CSFV is arguably the most important pathogen of pigs. As mentioned above, domestic and feral pigs, as well as wild boar are equally susceptible. Suitable methods for the control of CSF in domestic pigs are stamping out of the infection, i.e., killing and destruction of infected and suspected contact animals in combination with movement restrictions and other strict biosecurity measures. A ban of swill feeding is a mandatory and very effective prophylactic control measure. Usually farmers are compensated for their losses. Stamping out is generally used in countries that are free from CSF and where vaccination is prohibited.

Prophylactic vaccination can also be used as a control tool. It is applied, when new outbreaks can no longer be controlled by stamping out alone. In countries with endemic CSF vaccination is used in order to reduce economic losses. When used systematically, i.e., when the vast majority of a pig population is covered, the number of susceptible animals will decrease to an extent that the spread of the infection is no longer supported. The basic reproductive number of infection (R_0) drops below 1. One infected animal infects less than one susceptible animal. As a result the infection fades out and CSF will eventually be eradicated. This strategy has been used successfully in domestic pigs in the 1970ties in some regions of The Netherlands and from 2007 to 2010 in the Romanian backyard pig population.

Control of Classical Swine Fever in Wild Boar

CSFV can also infect wild boar. The epidemiology of the infection depends on a number of factors, in particular virulence of the virus and density of the animal population. From early experience it was thought that CSF outbreaks in wild boar are self-limiting. However, more recent outbreaks, e.g., in the 1990s in Germany became endemic. Possible reasons were high population densities and moderate virulence of the virus. CSF in wild boar is a severe threat to domestic pig holdings in the affected area. Therefore effective control measures are crucial.

Key parameters for the successful control of CSF in wild boar are the number of susceptible animals, R_0 (basic reproduction number of infection) and viral virulence. Fresh outbreaks result in a large number of deaths in younger age classes, while most older animals survive and will become seropositive. The latter are immune for the rest of their lives. This scenario may lead in low density populations to a drop of the basic reproduction number below the critical threshold ($R_0 < 1$), and outbreaks are self-limiting. In contrast, when population density is high there will always be enough susceptible animals due to intense reproduction to maintain the infection ($R_0 > 1$). CSF may become endemic for many years. Viral virulence influences these situations by the proportion of killed vs. surviving animals.

The prime purpose of any control measure is the reduction of susceptible animals to a threshold that R_0 drops below one. In sparsely populated areas this might be the case after the initial phase of the outbreak when a part of the population has succumbed to the infection and the remaining animals are immune. In these situations hunting is often banned in order not to disturb the population and keep the infection in the affected area. Natural boundaries, e.g., major rivers or mountains are helpful in containing the infection. In dense populations the number of susceptible animals has to be reduced by external intervention:

- Hunting activities should focus on the young age classes of wild boar, since older animals in an infected area are most likely to be immune. However, due to the high reproduction rate of wild boar experience has shown that hunting alone as a control tool is insufficient and does not result in eradication of the infection.
- Trapping might be considered a variant of hunting. Positive experiences using this control tool have been made, e.g., in Bulgaria. This strategy does not disturb the population and larger numbers of susceptible animals can be removed from the population.
- Oral immunization with baits containing MLV has been shown to be highly effective, and when applied over a period of a few years may significantly contribute to the extinction of the infection in endemic areas. A proven strategy is to vaccinate twice per year, i.e., in spring and autumn. Each vaccination is repeated once after four weeks. Baits are distributed in small restricted areas. Although very young age classes are not well reached by oral immunization, this strategy can reduce the number of susceptible animals sufficiently to reach $R_0 < 1$.

Sometimes a combination of the above strategies leads to success.

Control of Bovine Virus Diarrhea

Systematic control of BVD including the removal of PI animals was long discussed controversially, because of the ubiquitous occurrence of the infection and the stealthy nature of the infection. In addition the economic impact of BVD was underestimated for a long time. Today the economic dimension of BVD is fully appreciated. Transient immunosuppression and fetal infections produce disease and economic losses can be dramatic. Meanwhile BVD is considered to be the most important viral disease of cattle. The infection is spread by direct contact of acutely infected animals and by fomites. However, the most important source of BVDV spread are persistently infected (PI) cattle. They are the virus reservoir in cattle populations and consequently they play a pivotal role in any control programme. The life span of PI animals ranges from a few hours after birth up to seven or eight years. During their lifetime they shed large amounts of virus at any time. This is unique and the number of new infections transmitted by PI cattle may be unlimited, depending on frequency of animal contacts.

For more than fifty years vaccines have been used to prevent clinical BVD. Once the economic consequences of intrauterine infection were recognized, an additional requirement for licensing of vaccines was reliable fetal protection. However, despite widespread use of vaccination BVD prevalence did not change over the decades. The main reason for this failure was due to the remaining of PI animals as a powerful reservoir in cattle populations. Poor efficacy of some vaccines might have been an additional factor. In contrast to many other animal infections, e.g., CSF or rinderpest, BVD will never be eradicated using vaccination alone.

With the advent of inexpensive laboratory screening methods, it was possible to identify and remove PI cattle. Scandinavian countries were the first in the early 1990s to implement systematic control programmes by searching with serological methods for herds with active BVD infections. PI animals were identified in seropositive herds and removed. After their removal herds were monitored in regular intervals and, when seronegative, they were declared BVD-free. Strict biosecurity measures and movement restrictions for non-negative cattle were accompanying measures of these compulsory programmes. A few years of systematic control resulted in a BVD-free status of the Scandinavian countries. Vaccination was banned throughout the programme. Economic analyses showed that the cost benefit ratio of BVD control was highly positive. Austria followed a few years later with a similar control programme. In 2008 Switzerland chose a slightly different approach: The whole national cattle population was tested for PI animals. Thereafter all newborn calves were tested for BVD. The final phase of the programme is the constant serological surveillance of the cattle population. Within 3 years PI prevalence was reduced from 1.3% to 0.02%. Germany started a systematic effort to reduce PI prevalence in 2011 by testing newborn calves for BVDV. Vaccination as an additional control tool was allowed. PI prevalence was reduced from 0.48% (2011) to 0.01% (2017). Similar control programmes were initiated in Belgium, Scotland and Ireland.

In countries or regions without a compulsory programme for systematic BVD control it is possible to prevent economic damage on a herd basis by removing PI animals, regular vaccination in order to prevent reinfection with BVDV and strict biosecurity.

Control of Border Disease of Sheep

There are a few vaccines against BD available. Since the epidemiology of BD is very similar to BVD analogous strategies for its control can be used.

Further Reading

- Becher, P., Tautz, N., 2011. RNA recombination in pestiviruses: Cellular RNA sequences in viral genomes highlight the role of host factors for viral persistence and lethal disease. *RNA Biology* 8, 216–224.
- Blome, S., Moss, C., Reimann, I., Koenig, P., Beer, M., 2017. Classical swine fever vaccines-state-of-the-art. *Veterinary Microbiology* 206, 10–20.

- Lindenbach, B.D., Thiel, H.J., Rice, C.M., 2013. *Flaviviridae*. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed., vol. 1. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 712–746.
- Moennig, V., Becher, P., 2015. Pestivirus control programs: How far have we come and where are we going? *Animal Health Research Reviews* 16, 83–87.
- Moennig, V., Becher, P., 2018. Control of bovine viral diarrhea. *Pathogens* 7 (1).
- Nettleton, P.F., Gilray, J.A., Russo, P., Dlissi, E., 1998. Border disease of sheep and goats. *Veterinary Research* 29, 327–340.
- Postel, A., Austermann-Busch, S., Petrov, A., Moennig, V., Becher, P., 2018. Epidemiology, diagnosis and control of classical swine fever: Recent developments and future challenges. *Transboundary and Emerging Diseases* 65 (Suppl 1), 248–261.
- Tautz, N., Tews, B.A., Meyers, G., 2015. The molecular biology of pestiviruses. *Advances in Virus Research* 93, 47–160.
- Yeşilbağ, K., Alpaya, G., Becher, P., 2017. Variability and global distribution of subgenotypes of bovine viral diarrhea virus. *Viruses* 9 (6).

Capripoxviruses, Parapoxviruses, and Other Poxviruses of Ruminants (*Poxviridae*)

Philippa M Beard, The Pirbright Institute, Pirbright, United Kingdom and The Roslin Institute, University of Edinburgh, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Poxviruses cause disease in farmed and free-living ruminants, ranging from mild, self-limiting lesions to systemic, fatal disease. The species of ruminant poxvirus covered in this article are listed in [Table 1](#). The most common is Orf virus (ORFV) which is found worldwide and causes mild, transient, proliferative lesions on young sheep and goats. In contrast, the most severe poxvirus diseases of ruminants are caused by Sheeppox virus (SPPV), Goatpox virus (GTPV) and Lumpy skin disease virus (LSDV). These three viruses are recognised as causing high consequence transboundary diseases with a substantial morbidity and mortality burden.

Classification

The family Poxviridae contains two subfamilies – Entomopoxvirinae and Chordopoxvirinae. Poxviruses that affect ruminants are found within the Chordopoxvirinae subfamily and include species within the genera Capripoxvirus, Orthopoxvirus and Parapoxvirus.

Virion Structure

Poxviruses have a complex virion structure. The viral genome and transcription factors are believed to be contained in a folded tubular “nucleocapsid” structure. This is contained within the very recognisable biconcave core structure which has a lateral body in each concavity. The core structure is then enclosed within one to three lipid membranes.

Orthopoxviruses produce four virion types that are distinguished by the number of membranes surrounding the core and their cellular location – the intracellular mature virion (IMV), the intracellular enveloped virion (IEV), the cell-associated enveloped virion (CEV) and extracellular enveloped virion (EEV). In some recent papers IMVs are referred to as mature virions, IEVs as wrapped virions, and EEVs as extracellular virions.

Electron microscopy studies have shown parapoxviruses (PPV) produce IMVs, IEVs, CEVs and EEVs analogous to orthopoxviruses. These studies have also revealed a unique feature of the PPV virion – a raised surface tubule protein that is wrapped around the virion core to produce a “ball of wool” effect. This protein is encoded by the ORF104 gene; mutated viruses with this ORF disrupted produced morphologically aberrant particles and failed to create EEVs.

Very little work has been reported on the virion structure of capripoxviruses (CPPV). Examination of LSDV, SPPV and GTPV purified virions using electron microscopy have identified CPPV virion size as length 293–299 nm and width 262–273 nm, consistent with other poxviruses, and the absence of any surface tubular proteins. Electron microscopy images from LSDV-infected tissue have revealed IMV and IEV-like structures ([Fig. 1](#)).

Genome

The CPPV genome structure is similar to other poxviruses. It consists of a double stranded linear DNA genome that is 25% GC-rich, approximately 150 kbp in length, and encodes around 156 ORFs. Each end of the linear genome encodes for a

Table 1 Poxviruses of ruminants

<i>Genus</i>	<i>Species</i>	<i>Abbreviation</i>
Capripoxvirus (CPPV)	Sheeppox virus	SPPV
	Goatpox virus	GTPV
	Lumpy skin disease virus	LSDV
Parapoxvirus (PPV)	Orf virus	ORFV
	Bovine papular stomatitis virus	BPSV
	Pseudocowpox virus	PCPV
	Parapoxvirus of red deer in New Zealand	PVNZ
Orthopoxvirus (OPXV)	Vaccinia virus	VACV
Unclassified	Deerpox virus	
	Mule deerpox virus	

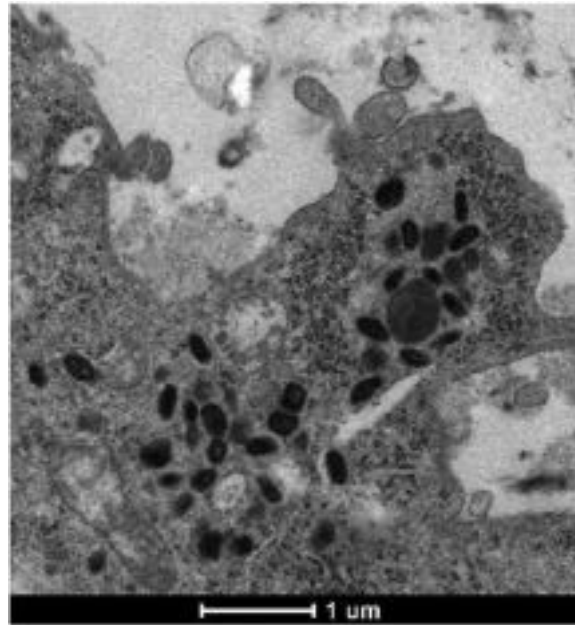


Fig. 1 Ultrastructural detail of LSDV morphogenesis. Tissue was collected from a cutaneous nodule on a calf experimentally inoculated with LSDV, then fixed and processed for electron microscopy using standard techniques. Acknowledgment: Beatriz Sanz-Bernardo, Philippa Hawes and Jennifer Simpson.

2200–2300 bp inverted terminal repeat sequence, and the linear ends of the genome are joined with a hairpin loop. The central region of the CPPV genome contains ORFs predicted to encode proteins required for virus replication and morphogenesis and exhibit a high degree of similarity with genomes of other mammalian poxviruses. In the outer regions of the CPPV genomes this similarity is lower. These ORFs contain genes encoding proteins likely to be involved in viral virulence and host range determinants. The CPPV isolates sequenced to date are genetically very similar, exhibiting 96% nucleotide identity. However phylogenetic analysis groups the isolates into the three species – SPPV, GTPV and LSDV.

The genome of the PPVs is one of the smallest amongst the poxviruses at 134–139 kbp. It is also unusual in its high GC content, around 64%. The only other poxvirus with a similarly high GC content is the human pathogen molluscum contagiosum virus. The PPV genome encodes 130–132 ORFs and exhibits the standard poxvirus genome structure with inverted terminal repeats and hairpin loops. Genes encoding essential functions are found centrally and ORFs encoding proteins involved in virus pathogenesis towards the periphery. PPV species exhibit greater genetic variation than CPPV, for example the amino acid identity of PCPV compared to BPSV is only 72%.

Life Cycle

It is widely presumed that the cellular life cycle of CPPV and PPV mirrors that of the well-studied orthopoxvirus genus. This is supported by microscopic and ultrastructural findings indicating cellular entry is followed by formation of cytoplasmic viral factories, and morphogenesis from immature virions to IMV, IEV, and CEV followed by virion release.

Sheeppox and Goatpox Virus

Epidemiology

Sheeppox and goatpox occur in Africa, the Middle East, and Asia including India and China. There are regular outbreaks on the periphery of these endemic zones, as for example Vietnam (2005 and 2008), Taiwan (2010), and Greece (2018). Participatory studies using farmer interviews have indicated these diseases are widespread in low and middle income countries. For example, 80% of farmers in Sudan had experienced an outbreak of sheeppox at some time in the past, while 14% of sheep farmers in Algeria reported current cases of sheeppox in their flock. Serological studies to estimate the prevalence of SPPV and GTPV are few but support the theory that the diseases are widespread. 17% of sheep and 14% of goats tested in west Amhara region of Ethiopia were seropositive, while 17% of goats were seropositive in the Layyah Punjab region of Pakistan. Morbidity, mortality and case fatality rates (CFR) of sheeppox and goatpox vary between outbreaks. In an epidemic of sheeppox in Israel morbidity was 20.5% and CFR 19.3%. In an epidemic of sheeppox in Mongolia morbidity was 14.5% and CFR 6%. In imported, unvaccinated sheep in a feedlot

in Jordan an outbreak of sheeppox caused 21.5% morbidity and CFR 41%. In more endemic settings in Sudan morbidity due to sheeppox varied between flocks from 0.45%–15.9% and CFR 0%–33%.

SPPV and GTPV are widely accepted to cause substantial economic hardship with particularly severe impacts on subsistence livestock owners in low and middle income countries. However the economic impact of SPPV and GTPV is poorly studied. A study from 2000 in India estimated it would take 6 years for a farm to recover from an outbreak of SPPV or GTPV, with average annual losses in income of 30%–43%, depending on flock type and the owner's actions.

SPPV and GTPV spread from animal to animal via direct contact or insect vectors such as biting flies (*Stomoxys calcitrans*). Aerosol infection of sheep has been achieved experimentally by spraying SPPV or GTPV into the close environment or directly into the face of sheep, although this method of transmission was found to be less effective than prolonged direct contact with a diseased animal. Under experimental conditions, 100% of animals inoculated intradermally with a virulent strain develop disease.

The host tropism of SPPV and GTPV strains varies. Some SPPV and GTPV viruses are highly host specific and cause disease in the field only in sheep or goats respectively. Other strains cause disease in both sheep and goats. The host specificity of SPPV and GTPV has been tested experimentally. Two GTPV strains, one from Vietnam and one from Yemen, were inoculated into both goats and sheep. Both isolates caused more severe disease in goats, the Yemen isolate was also moderately pathogenic in sheep. However the Vietnam isolate caused only very mild disease in sheep. This supports evidence from the field that different SPPV and GTPV strains exhibit variable host preference. The genetic determinants of SPPV and GTPV strain virulence and host preference are unknown.

Clinical features and pathology

Experimentally the incubation period of sheeppox and goatpox is 4–6 days following intradermal or subcutaneous inoculation. Both diseases are characterised by the appearance of round to oval, slightly raised, well demarcated lesions up to 2 cm in diameter on the skin. The appearance of skin lesions coincides with other clinical signs including severe pyrexia, purulent ocular and nasal discharge, anorexia, dyspnoea, lethargy, diarrhoea and superficial lymphadenopathy. The cutaneous lesions of sheeppox and goatpox are similar to the classic orthopoxvirus “pocks” and develop from macule to papule to pustule and then scab. Ulceration of the cutaneous lesions is a feature of some outbreaks. Lesions are often more numerous on sparsely haired areas of the body. Post mortem examination of cases of sheeppox and goatpox may reveal lesions in internal organs, described as firm red to white nodules in the lungs and intestines, liver and kidneys.

Sheeppox and goatpox can occur in the field as a mixed infection concurrent with bluetongue virus, peste-des-petits-ruminants virus, orf virus or *Mycoplasma capricolum* subsp. *capripneumoniae*. These mixed infections often result in mortality over 50%.

Histopathology

The dominant histopathological features seen in sheep and goats infected with SPPV and GTPV are keratinocyte hyperplasia and necrosis accompanied by vesicle formation in the epidermis, with large basophilic intracytoplasmic inclusion bodies often present in the degenerate and necrotic keratinocytes and in the epithelium of adnexal structures. Within the dermis oedema and haemorrhage is accompanied by infiltrating large histiocytic cells with a vacuolated nucleus and marginated chromatin. These cells are often found in perivascular areas, and can contain prominent large basophilic intracytoplasmic inclusion bodies. A necrotising vasculitis with histiocytic cells infiltrating the walls of blood vessels has been reported within the dermis and subcutis. In the lungs, a severe necrotising bronchiointerstitial pneumonia is present, with large histiocytes and other mononuclear cells infiltrating the interstitial tissues, particularly in peribronchiolar and perivascular areas. Large histiocytic cells accompanied by necrosis and vasculitis have also been noted in the liver, kidney, lymph nodes and gastrointestinal tract. Immunolabeling of tissues from affected animals have identified CPPV antigen in histiocytic cells, keratinocytes, and bronchiolar epithelial cells.

Pathogenesis

Most of the work on sheeppox and goatpox pathogenesis comes from time course studies following experimental infection. After intradermal or subcutaneous inoculation, virus disseminates from the inoculation site to distant cutaneous sites and other organs, possibly via a monocyte/macrophage associated viraemia. Virus is detected in the skin distant from the inoculation site 3–4 days later, and virus load increases up to day 7 or 8 and remains high until day 14 when it begins to gradually decrease. The skin and lung appear to be the major target organs for these viruses.

Diagnosis

SPPV and GTPV are often strongly suspected based on the clinical signs displayed by the affected animals. Differential diagnoses include orf, dermatophilosis, dermatophytosis, bacterial pyoderma, and mange. Confirmation of the diagnosis of sheeppox or goatpox is usually made on the basis of detection of SPPV or GTPV DNA within lesions or blood. A PCR based on the CPPV-encoded G-protein-coupled chemokine receptor (GPCR) gene can be used to differentiate between the three CPPV species. A CPPV neutralisation test can be used to detect neutralising antibodies in immune animals. Histopathology and electron microscopy are occasionally used for confirmation of the disease but are time consuming and expensive.

Treatment

Due to their high consequence outbreaks of SP and GP are often treated by euthanasia of affected and in-contact animals ("stamping out"). In circumstances where treatment is appropriate and economically viable sheeppox and goatpox may be treated with supportive therapy and antibiotic coverage to reduce secondary bacterial infections.

Prevention

Live attenuated strains of CPPV are widely used as vaccines to prevent sheeppox and goatpox. Similar to orthopoxviruses CPPVs are cross-protective within genus therefore one CPPV live attenuated vaccine can be used to protect against all three CPPV species.

Lumpy Skin Disease Virus

Epidemiology

Lumpy skin disease affects cattle and water buffalo and occurs in Africa, the Middle East, southeast Europe, the Caucasus, Russia and Kazakhstan. LSD was first described in southern Africa in the 1920s, from where it gradually spread northwards throughout Africa and entered the Middle East in the 1980s. Outbreaks occurred sporadically in this region until 2012 when the virus began to spread rapidly over long distances into new areas including Iran, Iraq, Jordan, Turkey, Russia and, in 2015, into mainland Europe (Greece). The disease has since been reported in many Balkan and Caucasian countries. The reasons behind the rapid and sudden spread of LSDV through the Middle East, Europe and western Asia since 2012 are unclear. Hypotheses include climatic conditions conducive to LSDV spread, or civil conflict hampering veterinary services and disrupting livestock movement patterns.

LSD is highly host specific with extensive serological surveys of wildlife in southern Africa finding no evidence for LSDV infection of any species other than cattle and water buffalo.

The mode of transmission of LSDV differs from SPPV and GTPV. Experimental studies found no evidence for direct transmission of the virus between animals, instead transmission is strongly believed to be via arthropod vectors. It is unclear if this vector borne transmission is biological or mechanical. LSD has been successfully transmitted experimentally via *Aedes aegypti* mosquitoes. LSDV was retained by the *Ae. aegypti* mosquitoes and transmitted to recipient cattle up to 6d after feeding on donor cattle. This work represents the only unequivocal demonstration of experimental transmission of LSD via vectors. In contrast two other species of mosquito (*Culex quinquefasciatus* and *Anopheles stephensi*) retained LSDV for only 24 h and did not transmit LSDV to recipients. Similarly, LSDV was detected in *Stomoxys calcitrans* (biting flies) and midges (*Culicoides nubeculosus*) after feeding the insects on a LSD-affected animal but transmission of the virus or disease to a naïve recipient animal was unsuccessful. Experimental transmission of LSDV has been demonstrated by transferring a *Rhipicephalus appendiculatus* ticks from donor to recipient animals mid-feed although the latter did not show clinical disease.

LSDV has shown a propensity for the male reproductive tract, with virus isolated from necrotic lesions in the testes and epididymis, and from the semen of LSD-affected bulls. Clinical LSD accompanied by a severe necrohaemorrhagic vulvovaginitis and metritis was demonstrated in a group of heifers artificially inseminated with semen spiked with virulent LSDV, suggesting this could be a route of LSD transmission.

A number of detailed epidemiological studies on LSD have been published. Communal grazing, proximity to water sources, temperature and rainfall have all been associated with outbreaks of LSD and are consistent with vector-borne transmission of virus. A well-documented LSD outbreak in a large dairy herd in Israel in 2006 was analysed and transmission by indirect or direct contact modelled. Indirect transmission was the only parameter that could explain the outbreak dynamics with an overall effect over 5 times larger than direct contact. Outbreaks in south-east Europe were strongly seasonal, occurring in spring and summer with very few, if any, outbreaks over winter. The median LSD spread rate in these outbreaks was calculated to be 7.3 km/week, consistent with vector-borne spread.

Morbidity and mortality in LSD outbreaks are lower than sheeppox and goatpox. Recent reports from epidemic areas in the Middle East and Europe have indicated morbidity of 9%–26%, and mortality from 0.5% to 2%. In comparison, data from 1161 LSD outbreaks over 15 years in an endemic area in Uganda reported a lower disease impact with morbidity 4.77% and mortality 0.03%.

Clinical features

LSD has an incubation period of approximately 7 days after experimental inoculation. The first sign of disease is pyrexia, followed by the appearance of distinctive and numerous raised cutaneous lesions from 0.5 to 5 cm in diameter which develop over 3–4 days from macule to papule to nodule. The cutaneous lesions are often accompanied by oral, nasal and ocular discharge, lethargy, anorexia, and in lactating animals a rapid drop in milk production. Brisket oedema and superficial lymphadenopathy are also reported. After 1–2 weeks the skin nodules become necrotic and the centre eventually sloughs, leaving a defect in the skin which can predispose the animal to secondary bacterial infection or myiasis.

Pathology

Gross pathology of LSD is characterised by multifocal, acute to chronic nodular dermatitis, with necrosis and ulceration. Histologically there is mild acanthosis, ballooning degeneration of keratinocytes with intracytoplasmic intracellular inclusion bodies, histiocytic inflammation of the dermis and fibrinonecrotic vasculitis (Fig. 2).

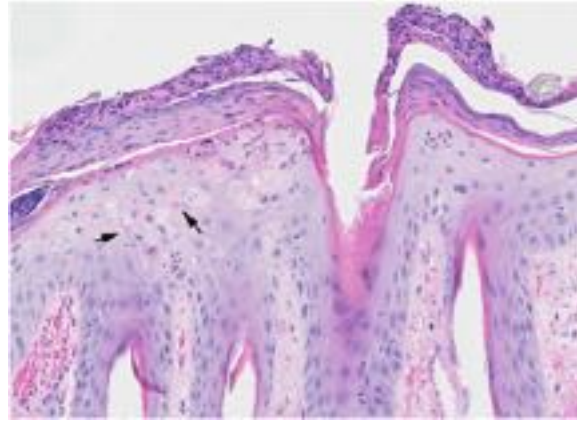


Fig. 2 Microscopic image of the histopathological changes in the epidermis of a nodule from a calf experimentally inoculated with LSDV. The tissue was fixed in buffered formal saline, processed to a wax block, and stained with haematoxylin and eosin using routine methods. The arrows point to eosinophilic intracytoplasmic inclusion bodies present in the keratinocytes.

Pathogenesis

Few details of the pathogenesis of LSD have been determined, and our understanding of how the virus causes disease is extrapolated from the more meticulously studied poxviruses such as the orthopoxviruses. It is believed that arthropod vectors inoculate LSDV into the skin of the animal, the virus then circulates in the blood stream via a low level viraemia and disseminates into distant cutaneous sites as well as the reproductive tract and other internal organs. Experimental models of LSD have consistently found that only around 50% of cattle inoculated with LSDV develop clinical disease; the host factors which control this variable susceptibility are unclear.

Diagnosis

LSDV is often strongly suspected based on the characteristic clinical signs. PCR is used to detect viral DNA in fluids or tissues to confirm disease. Virus isolation and electron microscopy can be used for diagnosis but are relatively expensive and time-consuming techniques. A virus neutralisation test can be used to detect neutralising antibodies in immune cattle. ELISA-based serological tests to detect antibodies to LSDV have proved challenging to develop due to the complex and variable antibody response to poxvirus infection. Differential diagnoses for LSDV are bovine herpesvirus 2 infection (sometimes known as pseudolumpy skin disease), dermatophilosis, demodicosis, besnoitiosis, Hypoderma bovis infection, and cutaneous tuberculosis.

Treatment

In epidemic situations treatment of LSD is rarely attempted, with stamping out of affected and in-contact cattle often mandated. In circumstances where treatment is appropriate and economically viable LSD may be treated with supportive therapy and antibiotic coverage to reduce secondary bacterial infections of the skin lesions.

Prevention

LSD can be prevented by vaccination with a live-attenuated CPPV strain. The most commonly used vaccine for LSD is the "Neethling" attenuated LSDV strain. Vaccination can be accompanied by movement restrictions to prevent the spread of the disease. Detailed analyses of the recent outbreaks in Europe and neighbouring countries have found vaccination to be a highly effective method of disease prevention, while questioning the effectiveness of a comprehensive stamping out policy.

Orf Virus

Epidemiology

Orf, also known as scabby mouth or contagious ecthyma, is caused by infection with orf virus (ORFV) and found worldwide wherever small ruminants are reared. The disease can be readily recognised from its characteristic scabby lesions around the mouth and nares of lambs and kids. The morbidity can be high but mortality is rare, and the disease is usually self-limiting. ORFV is very robust and can survive months in dry conditions, likely promoting recurrence of the disease in the spring lambing season. Like many other PPVs, ORFV is zoonotic with regular reports of transmission of the virus to people where it causes a focal proliferative dermatitis. Cases in humans have been associated with high risk professions such as veterinarians and farm workers, and with cultural celebrations which involve animal slaughter.

Molecular epidemiological studies of PPVs have been carried out, often targeting ORF011 (the B2L gene), which is the PPV orthologue of the vaccinia virus (VACV) Copenhagen strain gene F13L. A recent comprehensive study compared 23 isolates using ORF011 and ORF032 to enable categorisation of the PPV isolates to the correct genus.

Clinical features and gross pathology

Orf is characterised by proliferative, focal to regionally extensive scabby lesions most commonly around the mouth of lambs and kids. The lesions develop from an initial erythema, to macules, papules, vesicles, pustules and proliferative scabs. The disease has no systemic phase and is usually self-limiting however it can cause reduced growth rates in affected animals, most likely by inhibiting feeding and/or suckling. Less commonly the infection can spread to the udder of the ewe or doe, causing a dermatitis of the teat. There are occasional reports of ORFV causing more severe outbreaks of disease.

Reinfection can occur but the disease is milder and resolves more rapidly.

Histopathology

The histopathology of orf is focal to regionally extensive moderate to marked proliferative dermatitis or cheilitis, with marked dermal vascular proliferation and dilation. Early lesions are characterised by severe acanthosis and ballooning degeneration of keratinocytes, with elongated rete pegs often a key feature. Dermal changes include a moderate inflammatory cell infiltrate (mainly lymphocytes and histiocytes) with oedema occasionally causing subepidermal vesicles.

Pathogenesis

Orf virus (ORFV) infects broken or scarified skin and is therefore more often reported in animals grazing thistle, gorse or stubble. The virus initially replicates in regenerating keratinocytes immediately below the stratum corneum, before spreading laterally into adjacent epithelium. The inflammatory response consists of neutrophils, basophils and mast cells followed by abundant MHC class II+ histiocytic cells, T and B lymphocytes.

A number of virulence factors encoded by ORFV have been studied and their mechanism of function elucidated. These include GIF and VEGF-E.

GM-CSF and IL-2 inhibitory factor (GIF) is encoded by ORF117 of ORFV. This protein appears to be unique to PPVs with no orthologue found in any other Chordopoxvirus species. GIF is expressed as an immediate-late protein and binds sheep, but not human, GM-CSF and IL-2 and antagonises the bioactivity of both. More specifically GIF interferes with the myeloproliferative and chemoattractant properties of GM-CSF and the activation of lymphocytes and NK cells by IL-2. Recently, a structural biology approach was taken to identify how GIF targets these two different chemokines. This work revealed that GIF forms a dimer and binds two molecules of the target cytokine. GIF achieves this without sharing any structural similarity with either cognate cytokine receptor. GIF binds both GM-CSF and IL-2 with high affinity, but uses mutually exclusive interaction sites for each cytokine. GIF is the only such dual-purpose viral decoy receptor characterised to date.

Orf virus encodes a fifth member of the vascular endothelial growth factor family, known as VEGF-E. This protein binds to VEGF-receptor 2 (VEGFR2) and neuropilin-1, but not VEGFR1 or VEGFR3. VEGF-E exhibits biological activity consistent with its receptor binding, causing increased mitogenic activity of endothelial cells and increase vascular permeability. An ORFV strain lacking VEGF-E has been constructed and characterised *in vivo*. This mutated strain exhibited reduced gross pathology with particularly striking loss of the characteristic erythema seen in early orf lesions. Histologically the strain lacking VEGF-E produced less inflammatory cell influx, reduced hyperplasia of the epithelium, smaller pustules which resolved more rapidly, and reduced neovascularisation in the dermis. VEGF-E is therefore a key factor driving the pathological changes associated with ORFV infection, with a strong angiogenic role contributing to the characteristic neovascularisation and capillary dilation. The biological activities of VEGF-E have been used in a mouse model to promote wound healing. Expression of recombinant VEGF-E in normal skin caused epidermal thickening and increased numbers of endothelial cells and blood vessels in the dermis. In wounded skin, VEGF-E enhanced wound re-epithelialisation and increased thickness of neovascularisation, suggesting a therapeutic application for VEGF-E in promotion of wound healing.

Diagnosis

Clinical signs are often used to diagnose orf. Polymerase chain reaction (PCR) can be used for confirmation if needed. The main differential diagnoses for orf are sheeppox, goatpox, foot and mouth disease, bluetongue, peste-des-petits-ruminants, and mange.

Treatment

Most cases of orf are self-limiting, however support therapy is indicated when the clinical signs are more severe.

Prevention

Vaccination with a live attenuated strain of ORFV is the recommended method for reducing the disease. Vaccination is required on an annual basis.

Bovine Papular Stomatitis Virus

Bovine papular stomatitis virus (BPSV) causes mild disease in calves, characterised by flat to slightly raised red to brown macules and papules, sometimes progressing to ulcers, in the mucus membranes of the oral cavity. Diagnosis, treatment and prevention measures are rarely required, as severe disease is very rare. In common with other PPVs, BPSV is zoonotic with transmission to animal handlers reported. Older publications suggest differentiation of BPSV and PCPV can be made based on lesion location,

however cases of BPSV affecting the teats of cattle have been reported. Hence, molecular techniques are required to correctly diagnose these diseases.

Pseudocowpoxvirus

Pseudocowpoxvirus (PCPV) causes vesicular, pustular and scabby lesions most commonly on the teats and udders of milking cows. The clinical signs are often described as an incomplete ring of scabs in the shape of a horseshoe. The disease is mild and rarely of economic importance. PCPV should not be confused with the orthopoxvirus Cowpox virus which causes localised to systemic disease in cats, humans, rats, and other species.

Parapoxvirus of Red Deer in NZ

Parapoxvirus of red deer in NZ (PVNZ) is recognised as a fourth species of PPV after being isolated from lesions on the skin of growing antlers (velvet) of farmed red deer in New Zealand. The disease was originally described on eight farms, and characterised by scabby lesions on the muzzle, lips, face, ears, neck and velvet of farmed red deer. Viruses with strong genetic similarities to PVNZ have been isolated from healthy deer in Bavaria and from proliferative lesions around the mouth of red deer in Italy.

Vaccinia Virus

Vaccinia virus (VACV) is known to have a wide host range but rarely causes severe disease in an immunocompetent host. Sporadic cases of mild VACV infection of cattle (bovine vaccinia) have been reported in the literature for many years, however since the late 1990s bovine vaccinia in Brazil has increased to endemic levels in some regions. Bovine vaccinia presents as papules, vesicles and pustules, leading to ulcers and scabs, on the teat and udders of milking cows and oral cavity of suckling calves. There have been numerous reports of zoonotic transmission of the virus to animal handlers, mainly milking staff who develop vesicopustular and ulcerative lesions on their hands and arms, accompanied by fever, lymphadenopathy and lethargy. The disease can be severe enough to lead to hospitalisation. Human to human transmission of bovine vaccinia appears rare but has been documented.

The pathogenesis of bovine vaccinia is unclear. Experimental intradermal inoculation of cattle with VACV in the skin of the teat showed evidence of local disease and systemic spread, with detection of virus antigen in gastrointestinal-associated lymphoid tissue, and excretion of virus in faeces.

Molecular studies have shown at least two genetically distinct populations of VACV are responsible for the recent epidemic of bovine vaccinia in Brazil. The origins of the two VACV strains are unknown, hypotheses include a derivative of a VACV strain used during the smallpox eradication vaccination campaign in Brazil, or the emergence of a novel orthopoxvirus strain. Surveys of wild and domestic animals in Brazil have found serological and/or virological evidence of orthopoxvirus infection in a wide range of species including horses, rodents and marsupials, dogs, cats and wild-living primates. The role of any of these species in either the re-emergence or persistence of bovine vaccinia is unknown.

A strain of VACV known as buffalopox virus causes sporadic outbreaks of disease in buffalo and, less commonly, in cows in India and other Asian countries. The disease is usually mild, causing typical pox lesions on the teats, udder, and extremities of the head of affected animals.

Cervidpoxvirus

Cervidpoxvirus is a proposed poxvirus genera containing species provisionally named Deerpox virus or Mule deerpox virus. These viruses are associated with mucocutaneous ulcerative lesions in deer that are clinically dissimilar to proliferative PPV lesions. The full genome sequence of an isolate from skin lesions of a free-ranging mule deer has been reported and shown to be genetically distinct from other poxvirus genera. Experimental inoculation of black-tailed deer with an isolate of Deerpox virus resulted in mild clinical disease with gross and microscopic pathology consistent with poxvirus infection (focal to multifocal erythema, papules, pustules, ulcers and scabs), and confirmed by virus isolation. More extensive cutaneous and gastrointestinal lesions caused by Deerpox virus have been reported in captive and farmed white-tailed deer.

Further Reading

- Babiuk, S., Bowden, T.R., Parkyn, G., *et al.*, 2009. Yemen and Vietnam capripoxviruses demonstrate a distinct host preference for goats compared with sheep. *Journal of General Virology* 90 (Pt 1), 105–114.
- Beard, P.M., 2016. Lumpy skin disease: A direct threat to Europe. *Veterinary Record* 178 (22), 557–558. doi:10.1136/vr.i2800.
- Carn, V.M., Kitching, R.P., 1995. The clinical response of cattle experimentally infected with lumpy skin disease (Neethling) virus. *Archives of Virology* 140 (3), 503–513.
- Deane, D., McInnes, C.J., Percival, A., *et al.*, 2000. Orf virus encodes a novel secreted protein inhibitor of granulocyte-macrophage colony-stimulating factor and interleukin-2. *Journal Virology* 74 (3), 1313–1320.

- Felix, J., Kandiah, E., De Munck, S., *et al.*, 2016. Structural basis of GM-CSF and IL-2 sequestration by the viral decoy receptor GIF. *Nature Communications* 7, 13228.
- Matos, A.C.D., Rehfeld, I.S., Guedes, M., Lobato, Z.I.P., 2018. Bovine vaccinia: Insights into the disease in cattle. *Viruses* 10 (3).
- Mercier, A., Arsevska, E., Bournez, L., *et al.*, 2018. Spread rate of lumpy skin disease in the Balkans, 2015–2016. *Transboundary and Emerging Diseases* 65 (1), 240–243.
- Meyer, M., Clauss, M., Lepple-Wienhues, A., *et al.*, 1999. A novel vascular endothelial growth factor encoded by Orf virus, VEGF-E, mediates angiogenesis via signalling through VEGFR-2 (KDR) but not VEGFR-1 (Flt-1) receptor tyrosine kinases. *EMBO Journal* 18 (2), 363–374.
- Savory, L.J., Stacker, S.A., Fleming, S.B., Niven, B.E., Mercer, A.A., 2000. Viral vascular endothelial growth factor plays a critical role in orf virus infection. *Journal of Virology* 74 (22), 10699–10706.
- Wise, L.M., Veikkola, T., Mercer, A.A., *et al.*, 1999. Vascular endothelial growth factor (VEGF)-like protein from orf virus NZ2 binds to VEGFR2 and neuropilin-1. *Proceedings of the National Academy of Sciences of the United States of America* 96 (6), 3071–3076.

Chikungunya Virus (*Togaviridae*)

Thomas E Morrison and Stephanie E Ander, University of Colorado School of Medicine, Aurora, CO, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Arbovirus Viruses that are maintained in nature in transmission cycles between susceptible vertebrate hosts and blood-feeding arthropod vectors.

Arthralgia Pain in a joint.

Autochthonous transmission Spread of a disease within a local population by local arthropod vectors.

Conserved sequence elements (CSEs) Highly conserved genomic elements important for viral replication.

Cytopathic vacuoles Collection of rearranged host membranes and spherules that are a site of replication for alphaviruses.

Spherules Individual plasma membrane protrusions that serve as a site of alphaviral replication.

Sylvatic cycle of CHIKV transmission Transmission between nonhuman primates and *Aedes* species mosquitoes.

Synovitis Inflammation of synovial membranes in the joints.

Urban cycle of CHIKV transmission Transmission within a human population, where humans serve as the vertebrate reservoir and mosquitoes transmit the virus from human to human.

Viremia High presence of virus in the bloodstream; required for the transmission of arboviruses from vertebrate to invertebrate hosts.

Virus-like particles Non-infectious particles composed of viral structural proteins but containing no viral genome.

Classification

Chikungunya virus (CHIKV) is an enveloped virus with a positive-sense, single-stranded RNA genome. CHIKV is classified within the *Alphavirus* genus of the family *Togaviridae* based on its antigenicity, genome sequence and organization, and replication strategy. The *Alphavirus* genus is divided into 7 antigenic complexes; determined by antibody cross-reactivity to the virion glycoproteins E1 and E2, epitopes of which share a high degree of conservation between genetically similar viruses. Along with Bebaru, Getah, Mayaro, o'nyong 'nyong, Ross River, Sagiyama, Semliki Forest (SFV), and Una viruses, CHIKV is a member of the Semliki Forest virus antigenic complex.

Virion Structure

CHIKV particles are 60–70 nm in diameter and composed of an icosahedral nucleocapsid enclosed within a lipid envelope. The capsid is composed of 240 capsid monomers arranged in a $T = 4$ symmetry, and it contains a single copy of the viral genome. The viral envelope is derived from the host cell plasma membrane. Embedded within the viral envelope are 80 trimeric spikes composed of viral glycoproteins E2 and E1. A single spike consists of a trimer of E2/E1 heterodimers, which facilitate viral attachment and entry. Also associated with the E2/E1 complex is E3. During glycoprotein trafficking through the ER-Golgi complex, E3 is covalently bound to the immature E2/E1 dimer, where it aids the folding and association of the dimer; after furin cleavage of E3 in the Golgi, E3 remains non-covalently associated to protect against inappropriate activation of the E1 fusion loop. In the related SFV, E3 dissociates from the E2/E1 complex at neutral pH, however, cryo-EM analysis of infectious CHIKV particles indicates that some E3 is retained in the virion. Also incorporated into CHIKV particles at substoichiometric levels is the structural protein TF, as determined by mass spectrometry of purified virus particles. As TF is produced by ribosome slippage during synthesis of the viral protein 6K and shares the same N-terminus, 6K may also be present at low levels in CHIKV particles.

Genome Organization and Features

The CHIKV genome is a positive-sense, single-stranded RNA of approximately 12 kb in length. Like cellular mRNAs, the CHIKV genomic RNA (gRNA) and subgenomic RNA (sgRNA) possess both a 5' cap and 3' poly(A) tail. The capping and polyadenylation of alphavirus RNAs are mediated by viral proteins and the cap type 0 is different from eukaryotic mRNAs. The first two nucleotides of the genome (AU) are highly conserved within the *Alphavirus* genus, and substitution of these nucleotides inhibits gRNA replication.

The CHIKV genome is organized in a similar fashion to other alphaviruses, consisting of two sequential open reading frames (ORFs), encoding the nonstructural (ns) and structural proteins, flanked by untranslated regions (UTRs) and an internal non-coding region between the two ORFs (Fig. 1). The 5' ORF encodes the ns polyprotein, termed P1234, which is processed by the viral nsP2 protease into 4 mature ns proteins (nsP1 to nsP4). These ns proteins constitute the replicase that synthesizes new viral RNAs. The 3' ORF encodes the structural polyprotein composed of capsid, PE2 (E3 and E2), 6K/TF, and E1. As described in detail

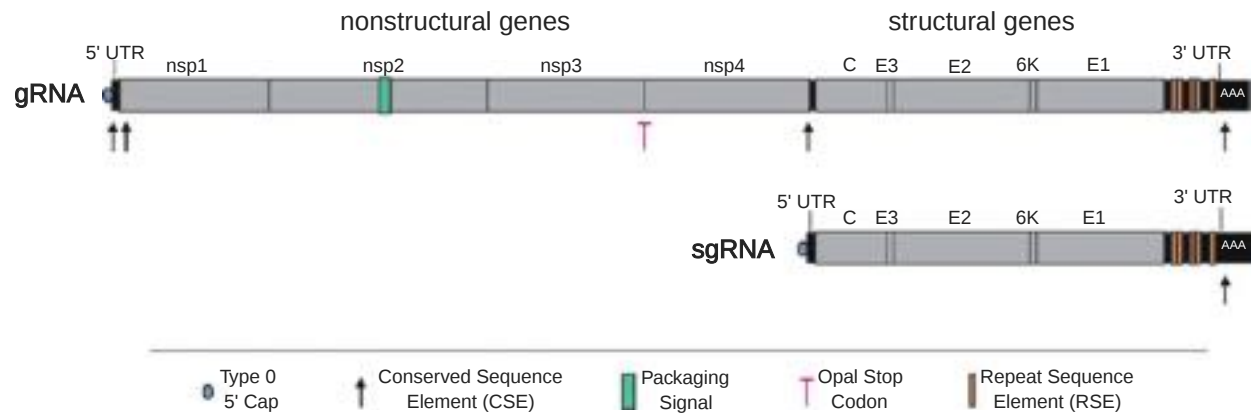


Fig. 1 CHIKV gRNA and sgRNA organization. The CHIKV gRNA serves as a transcript that is translated into the nonstructural polyprotein and a template for negative-strand RNA synthesis. The sgRNA is transcribed from the negative-strand RNA and is translated into the viral structural polyprotein. In addition to the protein-coding regions (indicated in gray), the gRNA and sgRNA encode *cis*-acting elements important for the viral life cycle. To promote translation, both RNAs have a 5' cap and a poly(A) tail. Conserved sequence elements (CSEs) have roles in the evasion of host defense mechanisms, transcriptional enhancement, and transcription of viral RNAs. The packaging signal located within the nsP2 gene promotes encapsidation of gRNAs within the capsid of progeny virions. NsP4 is expressed by a readthrough mechanism of the opal stop codon. Repeat sequence elements (RSEs) at the 3' UTR vary across the CHIKV genotypes in number and length and are important for viral replication.

below, this structural polyprotein is processed by capsid autocleavage and cellular proteases into individual structural proteins that assemble into CHIKV virions.

Protein-Coding Regions

The ns polyprotein is translated directly from the gRNA. A leaky opal stop codon near the 3' end of the nsP3 coding region results in the synthesis of two protein products, P123 and P1234. The structural polyprotein is translated from the sgRNA, which itself is transcribed from the subgenomic promoter active in the negative-strand viral RNA. As described in detail in the viral life cycle section, viral protein expression is temporally regulated by both the continual processing of the ns polyprotein and subsequent synthesis of viral RNA templates.

Non-Coding Elements

In addition to the protein-coding functions of the genome, specific non-coding elements of the alphavirus RNA genome have important roles in genome stability, packaging, viral RNA replication, evasion of host defenses, and host range determination. These activities are mediated by *cis*-acting elements, such as conserved sequence elements (CSEs), packaging signals, and read-through and frameshift elements. The function of these elements is dependent on the secondary RNA structure of the element, the nucleotide sequence, or a combination of both. These *cis*-acting elements are mostly localized to the 5' and 3' UTRs, but some also exist within the protein-coding regions.

There are four CSEs in the CHIKV genome: the 5' CSE, the 51-nt CSE, the sgRNA promoter, and the 3' CSE. These CSEs are highly conserved amongst alphaviruses in both nucleotide sequence and genome location. (1) The 5' CSE of CHIKV and other alphaviruses consists of 60 nucleotides that fold into a stem-loop structure that effectively occludes binding by the mammalian anti-viral protein IFIT1, a key player in the type I interferon-induced antiviral state. Based on studies with Sindbis virus, this stem-loop structure is also critical for RNA synthesis. (2) The 51-nt CSE in the nsP1 coding-region is another *cis*-acting element present in the 5' region of alphavirus genomes, and it forms two stem-loop structures. Both the structure and nucleotide sequence are important for this CSE to act as a transcriptional enhancer. (3) The promoter for the sgRNA is active in the negative-strand RNA, and it overlaps with the nsP4 coding sequence to encompass ~19 nucleotides upstream and 2–5 nucleotides downstream of the sgRNA transcription start-site. Synonymous substitutions at this site inactivate the sgRNA promoter. (4) The 3' CSE consists of 19 nucleotides and is located just upstream of the polyA tail at the 3' end of the CHIKV genome. This element functions as the promoter for negative strand RNA synthesis. Specifically, the last 5 nucleotides of the 3' CSE are required for initiation of negative strand RNA synthesis, where the last nucleotide of the 3' CSE (a cytosine) is the site of transcription initiation. Furthermore, the function of the 3' CSE is complemented by the 5' UTR; as both are required for negative strand synthesis, it is hypothesized that viral RNA replication is assisted by genome circularization.

In addition to the CSEs, the CHIKV genome also contains sequence regions important for genome packaging and viral mRNA translation. A packaging signal located within the CHIKV nsP2 coding region is required for genome encapsidation, and it permits the capsid protein to distinguish gRNA from sgRNA and cellular RNAs. Other functionally important regions of the CHIKV genome dictate nsP4 and TF gene expression. Translation of the gRNA predominately results in the synthesis of P123 instead of

P1234, however, low-frequency read-through of the opal stop codon near the junction of the nsP3–nsP4 coding regions results in the synthesis of P1234. A large stem-loop structure is predicted to be present immediately downstream of the opal stop codon, where it is positioned to promote readthrough events at the nsP3–nsP4 junction. A mutation of the opal stop codon to form an arginine readthrough codon decreased CHIKV-induced disease in a mouse model. The -1 frameshift site (U UUU UUA) within the alphavirus 6K gene is responsible for the expression of the structural protein TF. Most alphavirus sequences downstream of this frameshift site form either an RNA pseudoknot or a hairpin that acts to further stimulate the -1 frameshift. However, the -1 frameshift in CHIKV and other members of the SFV complex appears to be independent of an RNA structural element. Instead, the frameshift is proposed to be stimulated by either the downstream RNA sequence or an unidentified *trans*-factor.

CHIKV 3' UTR

The CHIKV 3' UTR is 498–723 nucleotides long, depending on the strain genotype (CHIKV genotypes are detailed in the Section “Epidemiology”), and is amongst the longest of the alphaviruses. Much of the length of the CHIKV 3' UTR is composed of repeat sequence elements (RSEs), and these vary in both number and length amongst the CHIKV genotypes. In mammalian cells, the third CHIKV RSE serves as a docking site for the cellular HuR protein which protects the viral RNA from degradation. However, most RSEs appear to be insignificant for viral replication in mammalian cells while they are required for optimal viral replication in mosquito cells. In the mosquito cell, the CHIKV 3' UTR RSEs are proposed to contain binding sites for host factors that regulate viral replication in a host-specific manner.

Life Cycle

In general, alphaviruses have a similar life cycle, and much of what is understood of the CHIKV life cycle has been inferred from other alphaviruses. The alphaviral life cycle is generally composed of viral attachment and entry, ns protein expression, viral RNA replication, structural protein expression, and virion assembly (Fig. 2).

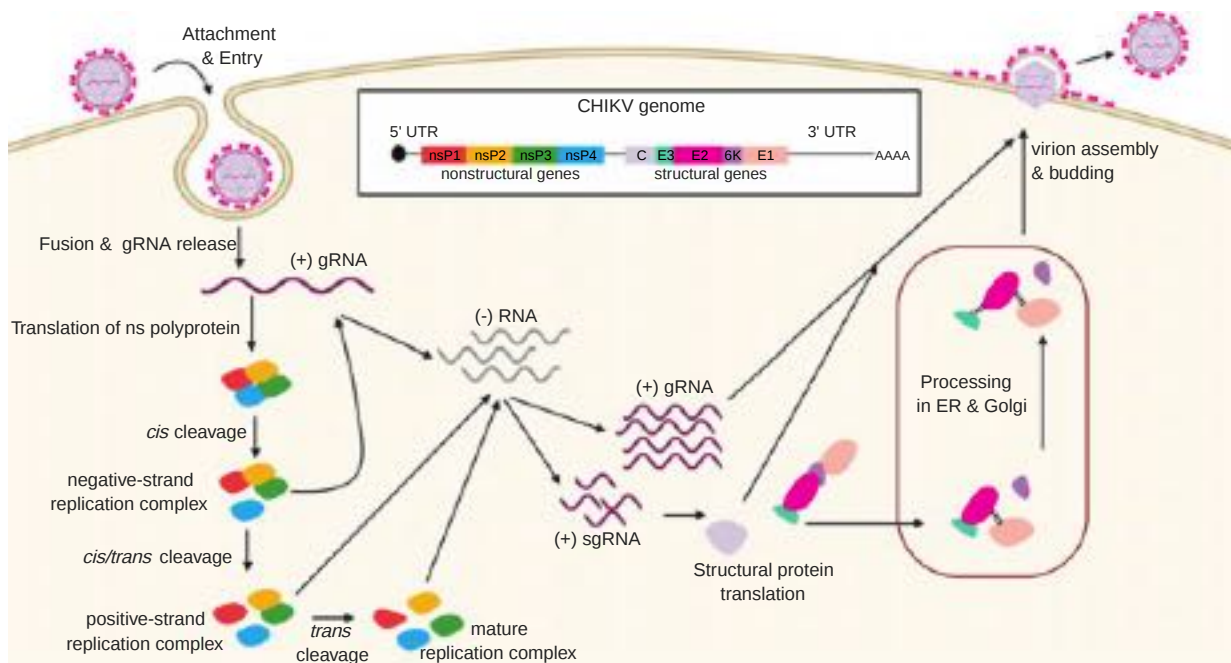


Fig. 2 CHIKV life cycle. Upon attachment and endocytosis of CHIKV virions, the viral envelope fuses with the endosomal membrane and releases the gRNA into the host cell cytoplasm. The gRNA is translated to produce the nonstructural polyprotein P1234. *Cis* cleavage of the polyprotein by the nsP2 protease produces a negative-strand replication complex (P123 plus nsP4); leading to the synthesis of negative-strand RNA. The nsP2 protease then further processes P123 to yield a positive-strand replication complex (nsP1 plus P23 and nsP4), which transcribes the positive-strand gRNA and sgRNA. Complete processing of P23 produces the mature replication complex (nsP1, nsP2, nsP3, nsP4), capable of synthesizing positive-strand gRNA and sgRNA. The structural proteins are translated from the sgRNA as a polypeptide. However, capsid self-cleavage during translation yields the capsid and E3–E2–6K–E1 polypeptide. An N-terminal signal peptide in E3 directs the E3–E2–6K–E1 polypeptide into the ER. In the ER, E3–E2–6K–E1 is processed by host signal peptidases to yield PE2 (E3–E2), 6K, and E1. E1 and PE2 associate to form heterodimers. The cleavage of E3 occurs during trafficking in the Golgi complex, but E3 remains non-covalently associated with the E2/E1 heterodimer until their arrival at the plasma membrane. Meanwhile, nucleocapsid cores are formed in the cytoplasm and are transported to the plasma membrane. During budding, progeny virions acquire their E2/E1-studded lipid envelope as they egress from the host cell.

Attachment and Entry

Initial CHIKV attachment to cells can be mediated by virion binding to cellular glycosaminoglycans (GAGs), a family of large, complex polysaccharides (e.g., heparan sulfate) that are present on the cell surface of most mammalian cells. Furthermore, the acquisition of enhanced GAG binding is a common cell culture adaptation of CHIKV and other alphaviruses. Virion-GAG interactions are mediated by residues in the domains A and B of the CHIKV E2 glycoprotein. For example, a mutation of CHIKV E2 domain A amino acid residues 79 and 82 to positively-charged residues enhances virion-GAG interactions, and results in an increased infectivity in cultured mammalian cells and decreased virulence in a mouse model of human disease. CHIKV attachment to cells also can be mediated by phosphatidylserine present in the viral envelope. Cellular phosphatidylserine receptors, such as T cell immunoglobulin and mucin domain-1 (TIM-1), can enhance the entry of CHIKV pseudovirions and virus-like particles in mammalian cells.

Mxra8 was identified as a CHIKV entry receptor in a genome-wide CRISPR-Cas9 screen and the observation was confirmed in both cell culture and mouse models. Mxra8 is a cell adhesion molecule that also serves as a cell entry receptor for several other arthritogenic alphaviruses including Mayaro, Ross River, and o'nyong nyong viruses. It possesses two immunoglobulin-like domains that form three contact points with E2/E1 heterodimers. Another putative mammalian cell entry factor for CHIKV is prohibitin, a multifunctional protein conserved in eukaryotes. Pretreatment of susceptible cells with anti-prohibitin-1 antibodies reduces CHIKV infection, and prohibitin-E2 interactions were shown with colocalization and immunoprecipitation studies. In mosquito cells, ATP synthase and heat shock protein 70 are the only putative CHIKV entry receptors identified to date.

Following viral attachment and entry, cell-bound virus is internalized to release the gRNA into the cytoplasm. The particular details of CHIKV internalization appear to differ based upon cell type and the species from which the cell is derived. However, there are some common features regarding CHIKV internalization, including initial virion uptake by the endocytosis pathways of the host cell and the pH-dependent fusion of the viral envelope with the endosomal membrane. Regarding endocytosis of CHIKV particles, both clathrin-mediated endocytosis (CME) and clathrin-independent mechanisms may contribute. CME is a constitutively active cellular process that involves the formation of clathrin-coated pits at the cellular surface which pinch-off into the cell to form clathrin-coated vesicles containing cargo such as transferrin and low-density lipoprotein. Microscopy studies examining CHIKV particle interactions identified viral interactions with both clathrin and clathrin-coated pits in human and mosquito cells, and selective targeting of endocytosis pathways can differentially affect CHIKV infection depending upon the cell type. Additional cellular factors including Eps15, dynamin, and FUZ can promote CHIKV internalization. Eps15 is required for the formation of clathrin-coated pits and dynamin mediates the pinching-off of endocytic vesicles from the plasma membrane. Each of these proteins can participate in both CME and clathrin-independent entry pathways. Finally, macropinocytosis also may mediate CHIKV internalization in some cells. Given that no study to date has been able to completely block CHIKV infection by targeting a single pathway, it may be that CHIKV particles exploit several cell-mediated internalization pathways to gain entry into host cells.

Following internalization, the virion envelope fuses with the endosomal membrane to release the gRNA into the host cell. Upon endocytosis, cargo is typically delivered first to the early endosome then trafficked to late endosomes and lysosomes. In mammalian cells, CHIKV fusion occurs within Rab5⁺ early endosomes. Viral fusion is a rapid process and can be completed within 40 s of virus particle delivery to the early endosome in mammalian cells. In mosquito cells, CHIKV particles associate with both Rab5⁺ early endosomes and Rab7⁺ late endosomes. Furthermore, knockdown of Rab7 reduces CHIKV infection of C6/36 *Aedes (Ae.) albopictus* cells, implying that viral fusion occurs in late endosomes in mosquito cells. Fusion is driven by the low pH of the endosome, and the differences between fusion in mammalian and mosquito cells may be due to species variation in endosomal pH. The mammalian early endosome has a pH of 6.8–5.5, whereas the pH threshold for CHIKV is pH 6.2–5.9 depending upon the specific virus strain.

In the mammalian early endosome, low pH induces destabilization of the E2/E1 heterodimer and a conformational change in E2 that exposes the buried hydrophobic fusion loop of E1, which mediates membrane fusion. In brief, low pH-induced movement of the alphavirus E2 B domain uncovers the E1 fusion loop, which extends and inserts into the endosomal membrane. This allows E1 to trimerize via interactions between domains I and II of three E1 molecules in the virion spike. Afterwards, E1 domain III refolds, forming a hairpin trimer that forces the viral and host membranes together. A fusion pore forms and releases the nucleocapsid and its packaged genome into the cytoplasm where ns protein expression and genome replication can start. The endosomal membrane fusion is further promoted by the presence of cholesterol and sphingomyelin in the endosomal membrane. In addition, host factor TSPAN9 (tetraspanin 9) can indirectly promote CHIKV and SFV fusion in the early endosome, likely either by recruiting a proviral factor or sequestering an antiviral factor.

Non-Structural Protein Expression and Viral Replication

Alphavirus RNA synthesis is associated with protrusions of the host cell plasma membrane. These rearranged membranes, called spherules, function as replication organelles and are the sites of RNA replication. Each spherule contains a single replication complex, and its formation is dependent on expression of viral nsPs and cleaved nsP4, in particular. nsP3 is hypothesized to contribute to spherule formation through association with host membrane curvature proteins. Later during the infection, individual spherules are endocytosed in bulk to form type-1 cytopathic vacuoles. In SFV, this assembly of cytopathic vacuoles is dependent on an endocytic process requiring PI3K and shuttling along actin and microtubule networks.

Table 1 General CHIKV viral protein functions

Viral protein	Enzymatic activity	Function in CHIKV replication
nsP1	Methyltransferase, guanylyltransferase	5' capping of gRNA and sgRNA; anchors replication complex to membranes
nsP2	Cysteine protease, helicase, NTPase/RTPase	Processing of ns polyprotein; RNA replication; 5' capping of gRNA and sgRNA
nsP3	ADP ribosylhydrolase	Spherule formation; association with host factor FHL1 required for productive infection
nsP4	RNA-dependent RNA polymerase (RdRp), adenylyltransferase	RNA replication; transcription of sgRNA
Capsid	Serine protease	Binds full-length gRNA; oligomerizes to form nucleocapsid
E3	–	Contains signal peptide for translocation of structural polypeptide into the ER; stabilizes E2-E1 heterodimers <i>en route</i> to surface expression
E2	–	Viral attachment protein; virion assembly
6K/TF	–	Stability and infectivity of virions; putative role in mediating membrane permeability, viral budding, and ion channel formation
E1	–	Viral fusion protein

The first genes expressed during CHIKV infection are the ns polyproteins P123 and P1234. The latter is synthesized when the ribosome translating the gRNA reads through the opal codon near the nsP3-nsP4 junction (see above). Read-through to produce P1234 occurs at a low-frequency in alphaviruses, yielding lower quantities of P1234 than P123. During infection, the ns proteins are continuously processed by the nsP2 protease, and the timing and order of this processing regulates the progression of viral RNA synthesis. Upon translation, P1234 is cleaved *in cis* by nsP2 to form P123 and nsP4, which functions as a negative-strand replication complex. The newly synthesized negative-strand RNA is used as a template for both gRNA and sgRNA synthesis by the positive-strand replication complex. The positive-strand complex is produced when P123 is further processed by nsP2 *in cis* and *trans* to yield nsP1, P23, and nsP4. Subsequent cleavage of P23 *in trans* produces the mature replication complex. The mature replication complex is present within hours of infection and is markedly incapable of synthesizing negative-strand RNA. However, complete processing of the nonstructural polyprotein is not required for viral RNA replication.

The individual ns proteins have distinct functions during viral replication (Table 1). The nsP1 protein encodes methyltransferase and guanylyl transferase activities that mediate 5' capping of the positive-sense gRNA and sgRNA. nsP1 also functions to associate the replication complex to host membranes and also may promote CHIKV particle release by downregulating the expression of tetherin, a host antiviral factor that traps virus particles at the cell surface during egress.

As discussed above, nsP2 encodes a protease domain at the C-terminus that is required for the processing of the alphavirus nonstructural proteins. nsP2 also encodes an S-adenosyl-L-methionine (SAM) dependent RNA methyltransferase-like domain at the C-terminus and helicase and NTPase domains at the N-terminus. Both the helicase and NTPase domains play important functions in viral transcription and replication; the former recognizes CSEs important for gRNA and sgRNA synthesis, while the latter modifies the viral RNAs to have a 5' cap. Lastly, some CHIKV nsP2 protein also translocates to the host cell nucleus where it mediates transcriptional shutoff by promoting the polyubiquitination, and consequent degradation, of the RNA polymerase II catalytic subunit RPB1. This effect is mediated by a small, highly variable loop in the nsP2 SAM MTase-like domain.

CHIKV nsP3 has three structural domains: a macrodomain, an alphavirus unique domain (AUD), and a hypervariable domain (HVD). The macrodomain in CHIKV has ADP-ribose hydrolase activity that promotes CHIKV replication and virulence. The AUD is capable of binding zinc, but possible other functions of this domain are presently not known. In contrast to the macrodomain and AUD, the sequence and length of the HVD, as the name implies, is highly variable among alphaviruses. This disordered region serves as a docking site for various host proteins in a virus-specific manner to promote viral replication complex assembly and function. Due to its disordered nature, this region permits large protein insertions, including fluorescent proteins and luciferase. During an infection, nsP3 associates with small, replication complexes at the plasma and endosomal membranes and large complexes associated with the cellular cytoskeleton. The HVD of CHIKV nsP3 also mediates binding with host factor FHL1; not only is FHL1 required for negative strand synthesis, but its expression determines cellular permissiveness for CHIKV. CHIKV nsP3 also sequesters cellular pools of stress granule proteins G3BP1 and G3BP2 and the mosquito homolog, Rasputin, which is important for viral replication complex formation. Other host factors known to interact with the CHIKV HVD of nsP3 include BIN1, CD2AP, MYBBP1A, NAP1L1, NAP1L4, and SH3KBP1.

CHIKV nsP4 is expressed at much lower levels than the other ns proteins due to low frequency readthrough events (discussed above) and N-terminus rule degradation. nsP4 functions as an RNA-dependent RNA polymerase (RdRp) required for the synthesis of viral RNAs. Poor *in vitro* expression of nsP4 has prevented structural studies, but sequence analysis predicts two functional domains. The predicted N-terminal domain is highly disordered and varies between alphaviruses. It is sensitive to point mutations and thought to mediate interactions with other viral and host proteins. The structure of the C-terminal domain is predicted to be similar to other RdRps, and this domain in SINV possesses an adenylyl-transferase activity, suggesting a mechanism for the presence of the poly(A) tail critical for viral RNA stability and function.

Structural Protein Expression and Virion Assembly

The alphavirus structural proteins are expressed from the sgRNA and are initially translated as a polyprotein that is processed to produce five individual proteins (Table 1). The first cleavage of the polyprotein occurs at the capsid-E3 junction, mediated by the protease activity of the capsid protein, and it occurs before translation of the polyprotein is completed. Upon capsid self-cleavage and completion of translation, two structural polyproteins are produced due to a low frequency ribosome slippage causing – 1 frameshift: E3–E2–6K–E1 and E3–E2–TF. E3 contains an N-terminal signal peptide that directs the polyprotein into the endoplasmic reticulum where complete processing will occur by host signal peptidase cleavage at both the N- and C-termini of 6K to yield PE2 (E3–E2), 6K/TF, and E1. E1 and PE2 form noncovalent heterodimers in the endoplasmic reticulum and are post-translationally modified in the Golgi complex. Further processing of PE2 by furin in the Golgi complex produces mature E2 and E3. Following this cleavage, E3 remains noncovalently associated with E2 and prevents premature fusion of the E2/E1 dimers in the low pH of the secretory pathway. Upon arrival at the plasma membrane, E3 can dissociate from the E2/E1 spikes or remain non-covalently associated. Nucleocapsid cores, formed by the oligomerization of capsid protein around a single gRNA, are transported to the plasma membrane. Viral egress occurs via budding, whereby the nucleocapsid cores acquire a lipid envelope decorated with E2/E1 spikes.

Epidemiology

CHIKV is an arbovirus that replicates in mosquitoes and mammals. The primary mosquito vectors that transmit CHIKV to humans are *Ae. aegypti* and *Ae. albopictus*. Adaptation of CHIKV to *Ae. albopictus* is a relatively recent event that arose due to the acquisition of an A226V point mutation in the E1 glycoprotein, and consequently, increases the risk of viral transmission in temperate climates. Transmission cycles of CHIKV can be sylvatic or urban. In the sylvatic cycle, the virus is primarily transmitted between nonhuman primates and arboreal *Aedes* species mosquitoes, and incidental spillovers can cause disease in humans. In the urban cycle, humans serve as the vertebrate reservoir, and mosquitoes transmit the virus from human to human.

Historically, CHIKV was associated with localized outbreaks in Africa and Asia, but due to a dramatic re-emergence in 2004, CHIKV is now found in tropical and sub-tropical areas throughout the world. Phylogenetic analysis of extant strains suggest that CHIKV originated in Africa. The first confirmed epidemic of CHIKV infection in humans occurred in modern-day Tanzania in 1953, and the second recorded epidemic occurred in 1958 in Thailand. In 2004–2005, an outbreak in coastal Kenya spread to islands in the Indian Ocean region that had not previously been associated with CHIKV activity, resulting in hundreds of thousands of infections. In late 2005, CHIKV activity exploded in India, and viremic travelers continued to spread the virus to temperate regions, including Europe. As the virus strain implicated in this epidemic acquired mutations in the E1 and E2 glycoproteins capable of expanding the pool of mosquito vector species, autochthonous transmission was established in the temperate climates of Italy and France. In 2013, CHIKV arrived in the Caribbean region, marking the introduction of the virus to the western hemisphere. CHIKV spread rapidly across the Caribbean region and the Americas, resulting in millions of confirmed and suspected cases. In addition, locally transmitted cases occurred in Florida in 2014.

Prior to 2004, three distinct genotypes of CHIKV had been described: Asian, East/Central/South African (ECSA), and West African lineages (Fig. 3). However, the 2005–2006 outbreak in the Indian Ocean was associated with a new CHIKV sublineage. This Indian Ocean Lineage (IOL) arose from the ECSA genotype, and it is notably characterized by the presence of the E1 A226V mutation in many isolates that increases CHIKV infectivity for *Ae. albopictus* mosquitoes. This point mutation occurred independently in different geographical regions and within different ECSA genetic backgrounds. During the 2005–2006 outbreak, the IOL genotype was globally disseminated by viremic travelers, including its introduction into Southeast Asia. However, the endemic Asian CHIKV strains were not displaced; instead these strains continued to be associated with CHIKV outbreaks in Asia and the Pacific region at the same time. Despite the explosive spread of the IOL/ECSA genotypes, the CHIKV strains that were first identified in the Americas were of the Asian genotype rather than of the IOL/ECSA lineage. Today, most of the CHIKV strains circulating in the Americas phylogenetically form a separate clade within the Asian genotype. However, in 2014 an ECSA CHIKV strain was introduced into Brazil. Given this genotype's previously expanded vector range, it has elicited heightened concern over the risk of CHIKV spread to more temperate areas in the western hemisphere, including the United States.

Clinical Features

CHIKV has an incubation period of 1–12 days, and in most cases, symptoms persist 1–2 weeks (Fig. 4). Unlike most other arbovirus infections, a high percentage (75%–95%) of CHIKV infections are symptomatic. For example, a case-control study during the 2009 epidemic in Thailand found 91% of laboratory-confirmed cases to be symptomatic. The acute disease presents with a sudden high fever, excruciating musculoskeletal pain (including polyarthralgia) in the periphery, headache, and rash; the afflicted joints are typically inflamed. Patients can have a very high viremia (up to 10^9 viral genome copies/mL) and typically remain viremic for about 1 week. In some cases, CHIKV disease is associated with atypical manifestations such as neurological and cardiovascular symptoms and even death in neonates, the elderly, and patients with pre-existing medical conditions.



Fig. 3 *Geography of CHIKV genotypes.* There are three genotypes of CHIKV strains: Asian, East/Central/South African (ECSA), and West African lineages. The geographic distribution of the ECSA and Asian genotypes has expanded dramatically since 2004, likely due to increased intercontinental travel, expansion of the CHIKV mosquito host range, and other factors.

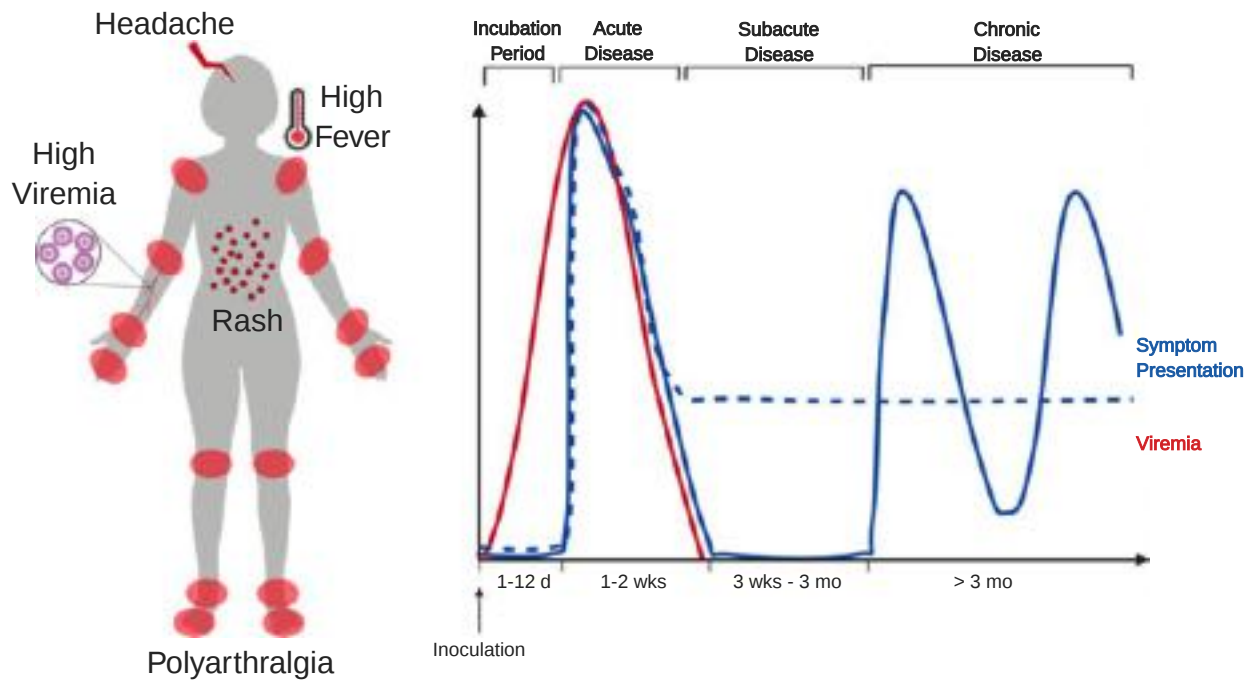


Fig. 4 *Clinical features of CHIKV infection.* Unlike many other arbovirus infections, a high proportion of CHIKV infections are symptomatic. Many infected individuals develop a high titer viremia. The most common signs and symptoms include headache, fever, rash, and polyarthralgia. Symptoms typically appear 1–12 days after infection, and the acute disease can last for 1–2 weeks. Some patients can completely recover after the acute phase of the disease, however, many patients suffer from a subacute and chronic form of the disease. A subacute disease is characterized by the persistence of arthralgia for 3 weeks to 3 months after the initial infection. The chronic form of the disease can include a constant arthralgia over 3 months post-infection or a sudden relapse in clinical disease.

A hallmark of CHIKV infection is the relapse of symptoms or continued symptoms well after the acute infection period has passed. As many as half of all CHIKV patients will relapse or continue to experience arthralgia for weeks, months, or even years after the initial infection. This chronic form of CHIKV disease can be classified into two major types, a musculoskeletal disease and a chronic inflammatory rheumatic syndrome. During the relapse, pain is typically experienced in the same joints that were affected

during the acute phase of the infection, at an equal or a more intense level than previously experienced. The chronic disease form is associated with long-term sequelae, including synovitis and tenosynovitis in the wrists and ankles, joint effusion, and bone erosion/lesions. Analysis of human patients have identified some risk factors for the development of chronic CHIKV disease, which include increased age, severity and duration of initial joint pain, female sex, and pre-existing rheumatologic disease.

Pathogenesis

In vitro, CHIKV is capable of productively infecting various mammalian cells, including those of human, primate, and rodent origin. A study examining CHIKV tropism for human cell types found productive infection of various adherent cells, including epithelial and endothelial cell lines, primary fibroblasts, and primary monocytes and macrophages; but no infection of T and B lymphocytes or monocyte-derived dendritic cells was seen. Productive infection is associated with a cytopathic effect and typically leads to apoptosis of virus-infected cells. Viral replication is completed within 8–16 h post-infection and results in titers of 10^5 – 10^8 PFU/mL depending on the cell type.

In vivo, CHIKV replicates at the site of inoculation in the skin where it predominately targets fibroblasts, but it may also infect resident dendritic cells or macrophages. The virus then disseminates to lymph nodes, spleen, muscles, liver, and peripheral joints where it replicates in connective tissue fibroblasts, skeletal muscle cells, osteoblasts, and possibly in other cell types. During this acute phase of the disease, CHIKV potently induces type I interferon and other proinflammatory cytokines (including IL-1 β , IL-6, IP-10, and RANTES), of which elevated levels have been correlated with increased disease severity in patients and experimentally-infected animals. Virus replication results in musculoskeletal tissue damage and inflammation (including synovitis, tenosynovitis, and myositis) and immune cell infiltration leading to edema and degeneration of cellular and tissue architecture. Various leukocyte populations have been implicated as mediators of CHIKV-induced musculoskeletal tissue pathology in mouse models. Monocyte and neutrophil accumulation in musculoskeletal tissues is associated with CHIKV infection, and disrupting their recruitment to these sites can prevent more severe forms of pathology including the loss of bone. In addition, inoculation of CHIKV into the foot of mice causes a biphasic pattern of swelling with a small peak occurring 2–3 days after the infection and a larger peak occurring at days 6–8 postinfection. This second peak is associated with the influx of CD4⁺ T cells into the foot tissue, and mice lacking CD4⁺ T cells have reduced foot swelling. Thus, although CD4⁺ T cells are required to develop an effective B cell response against CHIKV infection, these cells also exacerbate the joint disease.

Post-acute and chronic CHIKV disease are typically defined as the persistence of disease signs and symptoms from three weeks to three months or for more than three months, respectively, after the acute disease onset. The underlying causes of chronic CHIKV disease are not well understood. The induction of autoimmunity and/or viral persistence in joint tissues have been hypothesized as contributing factors. However, chronic CHIKV disease also could be due to unresolved injury in joint tissue either caused or exacerbated by the acute infection. Animal models have provided evidence of viral persistence associated with chronic synovitis, perhaps as a result of prolonged immune activation. In mice, infectious virus can be recovered from aged mice at 60–90 days post infection, and CHIKV RNA was found in joint-associated tissues and feet up to 16 weeks after the initial infection. A recent mouse study demonstrated that muscle and dermal cells maintain CHIKV infection. In addition, CHIKV RNA can persist in these cells with no detectable surface expression of viral proteins. In nonhuman primates, infectious virus and viral RNA can be recovered weeks post inoculation from joints, muscles, lymphoid organs, and the liver. In aged macaques, RNA can be detected in the spleen as long as 5 weeks post infection. Moreover, CHIKV antigen has been detected in perivascular synovial macrophages and muscle satellite cells in tissue biopsies collected from patients diagnosed with chronic CHIKV disease.

Diagnosis, Treatment, and Prevention

CHIKV infection can be misdiagnosed as dengue due to similarities in patient symptoms and geographic distribution of these two viruses, resulting in an under-diagnosis of CHIKV infection. Key clinical indicators for a CHIKV diagnosis are fever and arthralgia, and studies examining human cases have correlated these symptoms with a 74%–79% positive predictive value and with 73%–84% sensitivity. However, RT-PCR or CHIKV-specific serology is required to confirm a suspected CHIKV infection.

Although studies are underway to identify CHIKV-specific antivirals, there is currently no specific treatment for CHIKV infection. Instead, patients receive general care in the form of anti-inflammatories and analgesics to reduce fever and to alleviate pain and other symptoms. Biological drugs have shown some promise; monoclonal antibodies against CHIKV have been shown to be protective against the disease in nonhuman primates and mice, even when administered at late stages of the infection.

As with treatment, there are no specific prevention measures available against CHIKV. Mosquito control and methods to reduce risk of mosquito bites are the only preventative measures currently available. There is no approved vaccine available, but there are several in clinical trials. The first live, attenuated CHIKV vaccine strain (vaccine TSI-GSD-218; CHIKV strain 181/25) was developed by a serial passage of a clinical isolate in cell culture. The resulting 181/25 strain is immunogenic and protective against CHIKV challenge in animal models. While phase II clinical trials found it to be highly immunogenic, safety concerns regarding bulk production and reversion and instability of attenuation led to its discontinuation. Currently, there are two other CHIKV vaccines in phase II clinical trials: VRC-CHKVLP059-00-VP, a virus-like particle vaccine employing only the CHIKV structural proteins; and MV-CHIKV, another virus-like particle strategy using CHIKV pseudotyped, live-attenuated measles virus vaccine.

Further Reading

- Basore, K., Kim, A.S., Nelson, C.A., *et al.*, 2019. Cryo-EM structure of Chikungunya virus in complex with the Mxra8 receptor. *Cell* 177, 1725–1737.
- Chen, R., Puri, V., Fedorova, N., *et al.*, 2016. Comprehensive genome scale phylogenetic study provides new insights on the global expansion of Chikungunya virus. *Journal of Virology* 90, 10600–10611.
- Chen, R., Wang, E., Tsetsarkin, K.A., Weaver, S.C., 2013. Chikungunya virus 3' untranslated region: Adaptation to mosquitoes and a population bottleneck as major evolutionary forces. *PLoS Pathogens* 9, e1003591.
- Gardner, J., Anraku, I., Le, T.T., *et al.*, 2010. Chikungunya virus arthritis in adult wild-type mice. *Journal of Virology* 84, 8021–8032.
- Hawman, D.W., Fox, J.M., Ashbrook, A.W., *et al.*, 2016. Pathogenic Chikungunya virus evades B cell responses to establish persistence. *Cell Reports* 16, 1326–1338.
- Kendall, C., Khalid, H., Marietta, M., *et al.*, 2019. Structural and phenotypic analysis of Chikungunya virus RNA replication elements. *Nucleic Acids Research* 47, 9296–9312.
- McCarthy, M.K., Davenport, B.J.J., Morrison, T.E., 2018. Chronic chikungunya virus disease. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer, pp. 1–26.
- Meertens, L., Hafirassou, M.L., Couderc, T., *et al.*, 2019. FHL1 is a major host factor for Chikungunya virus infection. *Nature* 574, 259–263.
- Ooi, Y.S., Stiles, K.M., Liu, C.Y., Taylor, G.M., Kielian, M., 2013. Genome-wide RNAi screen identifies novel host proteins required for alphavirus entry. *PLoS Pathogens* 9, e1003835.
- Teo, T.H., Lum, F.M., Claser, C., *et al.*, 2013. A pathogenic role for CD4⁺ T cells during Chikungunya virus infection in mice. *Journal of Immunology* 190, 259–269.
- Thiberville, S.D., Moya, N., Dupuis-Maguiraga, L., *et al.*, 2013. Chikungunya fever: Epidemiology, clinical syndrome, pathogenesis and therapy. *Antiviral Research* 99, 345–370.
- Tsetsarkin, K.A., Vanlandingham, D.L., McGee, C.E., Higgs, S., 2007. A single mutation in Chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathogens* 3, e201.
- Voss, J.E., Vaney, M.C., Duquerroy, S., *et al.*, 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468, 709–712.
- Young, A.R., Locke, M.C., Cook, L.E., *et al.*, 2019. Dermal and muscle fibroblasts and skeletal myofibers survive Chikungunya virus infection and harbor persistent RNA. *PLoS Pathogens* 15, e1007993.
- Zhang, R., Kim, A.S., Fox, J.M., *et al.*, 2018. Mxra8 is a receptor for multiple arthritogenic alphaviruses. *Nature* 557, 570–574.

Circoviruses (*Circoviridae*)

Giovanni Franzo, Department of Animal Medicine, Production and Health (MAPS), Padua University, Padua, Italy

Joaquim Segalés, Department of Animal Health and Anatomy, Faculty of Veterinary Medicine, Autonomous University of Barcelona, Barcelona, Spain; Animal Health Research Center (CRESA) – Institute of Agrifood Research and Technology (IRTA), Campus UAB, Barcelona, Spain; and OIE Collaborating Center for the Research and Control of Emerging and Re-emerging Swine Diseases in Europe (IRTA-CReSA), Barcelona, Spain

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A. Mankertz, Circoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00702-0](https://doi.org/10.1016/B978-012374410-4.00702-0).

Nomenclature

CSB Chondroitin sulfate B

cryo-EM cryo-electron microscopy

ELISA Enzyme-linked immunosorbent assay

HS Heparan sulfate

IHC Immunohistochemistry

ISH In situ hybridization

ORF Open Reading Frame

PBMC Peripheral blood mononuclear cell

PCR Polymerase Chain Reaction

RCR Rolling circle replication

RER Rough endoplasmic reticulum

Glossary

Apoptosis Programmed self-induced cell death.

Botryoid Grape-like appearance.

Bursa of Fabricius Specialized organ in birds, that is necessary for B-cell development.

Cryogenic electron microscopy (cryo-EM) Cryogenic electron microscopy is an electron microscopy (EM) technique applied on samples cooled to cryogenic temperatures and embedded in an environment of vitreous water. It is an alternative to X-ray crystallography or NMR spectroscopy for

macromolecular structure determination without the need for crystallization.

Open reading frame (ORF) Open reading frame is the part of a reading frame that has the ability to be translated.

Phylogenetic tree Phylogenetic trees depict the evolutionary relationships among species that are believed to have a common ancestor.

Rolling circle replication (RCR) Rolling circle replication is a process of unidirectional nucleic acid replication that can rapidly synthesize multiple copies of circular molecules of DNA or RNA.

Classification

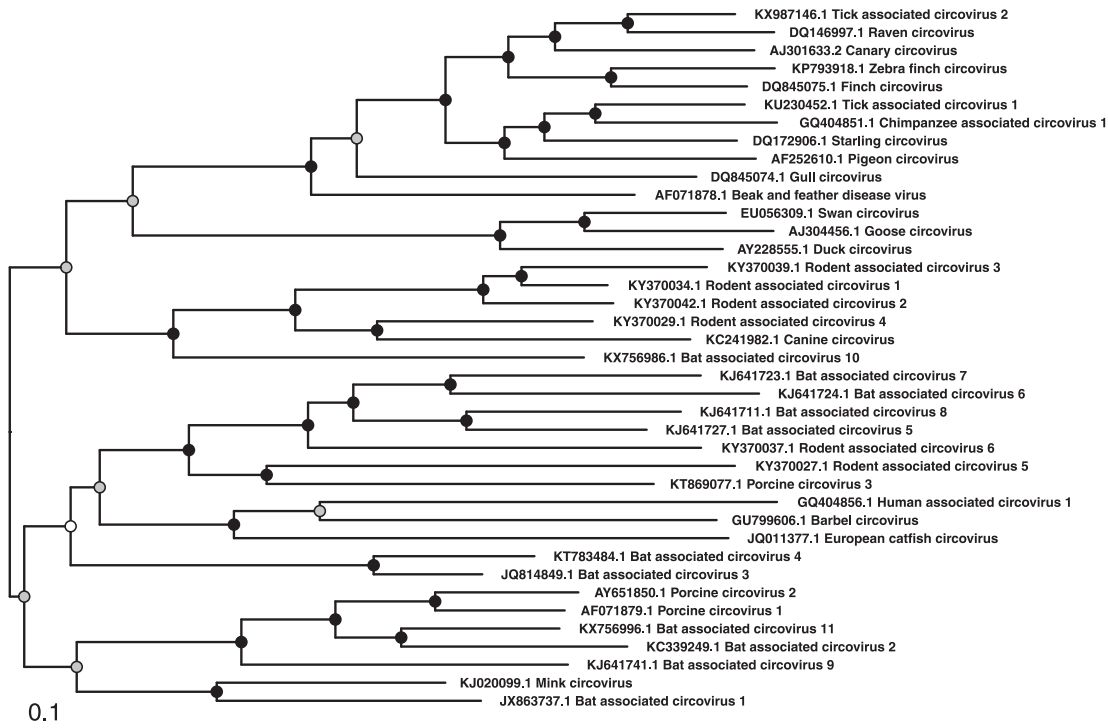
The family *Circoviridae* was initially established in the mid-1990s and included animal viruses with circular, single-stranded DNA (ssDNA) genomes, distinct from other eukaryotic ssDNA viruses classified at the time. While originally only the genus *Circovirus* was defined, the variability in ssDNA genome sequences and structures prompted the creation of 2 genera (Virus Taxonomy: 2018b Release) within the *Circoviridae* family: *Circovirus* and *Cyclovirus*.

Until 2010, only viral species responsible for relevant disease of avian species and swine were classified within the genus *Circovirus*, with the exception of a non-pathogenic porcine circovirus (PCV-1). However, metagenomic-based studies and degenerate PCR for circoviruses in unconventional hosts have since identified the presence of circovirus genomes in several other species including various other mammals and avian species, freshwater fishes and even ticks. However, the actual host has not been confirmed for some of these newly-detected circoviruses. A caution criterion should be followed, especially when viral sequences were obtained from samples associated to the digestive tract (e.g., feces or pharyngeal/rectal swabs) or from hematophagous parasites (i.e., tick-associated circoviruses) since the detection of viral DNA could be due to an infected meal.

Members of *Cyclovirus* have been described since 2010 in both vertebrates and invertebrates. However, similarly to several circoviruses, a definitive host has been difficult to assign for most, if not all, cycloviruses. In fact, all of them have been identified through metagenomic analysis and degenerate PCR primers. Although they share several genomic features with circoviruses, the putative origin of replication (*ori*), which is marked by the nonanucleotide motif, is located on the Cap-encoding strand in cycloviruses while it is on the Rep-encoding strand in *Circovirus* genus members. Moreover, introns have been identified within Open Reading Frames (ORFs) of several cyclovirus genomes, but not in circoviruses.

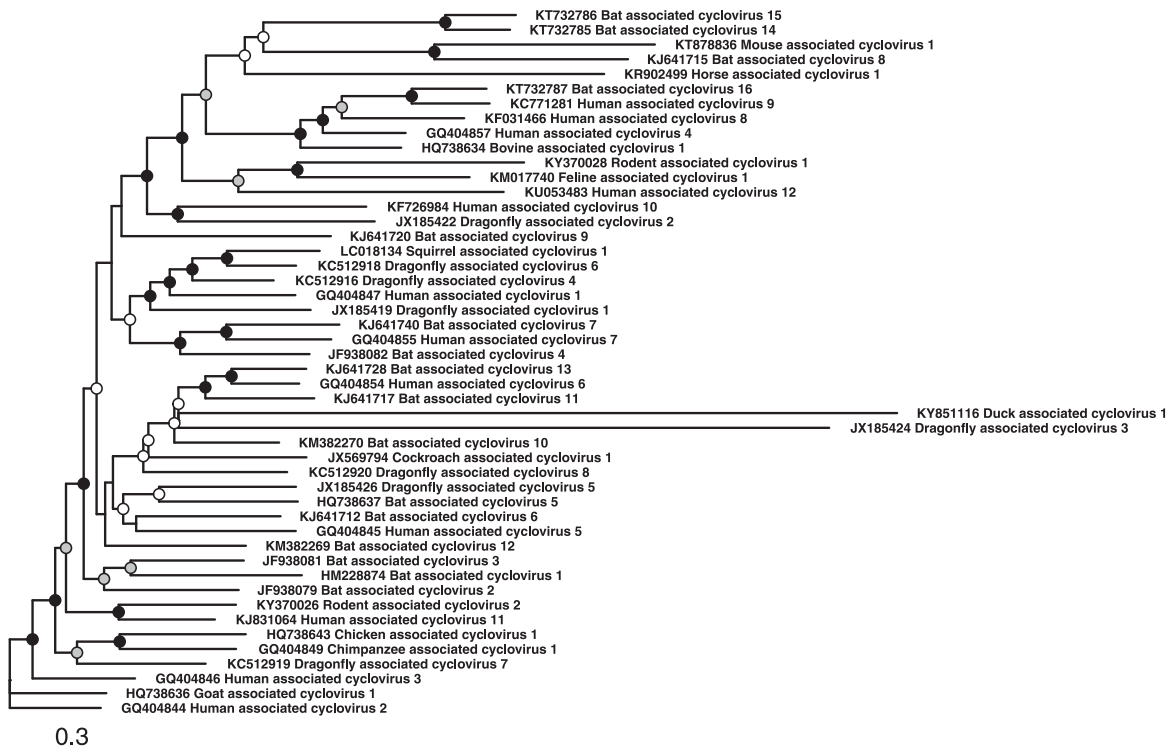
Based on genome-wide pairwise identities of Circoviruses and Cycloviruses, a species demarcation threshold of 80% genomic identity, well supported phylogenetically, has been proposed and accepted. Currently, 39 and 48 species have been classified within the *Circovirus* and *Cyclovirus* genera, respectively. ([Fig. 1](#)).

Circovirus tree



Bootstrap Percentage(BP) ● BP ≥ 90 ○ 70 ≤ BP < 90 ◻ BP < 70

Cyclovirus tree



Bootstrap Percentage(BP) ● BP ≥ 90 ○ 70 ≤ BP < 90 ◻ BP < 70

Fig. 1 Maximum likelihood (ML) phylogenetic trees based on the complete genome of members of the viral species within the genus *Circovirus* (top) and *Cyclovirus* (bottom). The ML trees were inferred using Iq-Tree software with the GTR + G + I model of substitution after aligning complete genome sequences using the MAFFT algorithm. Branch support, estimated by 10,000 ultrafast bootstrap replicates, has been categorized and color-coded. The phylogenetic trees were midpoint-rooted.

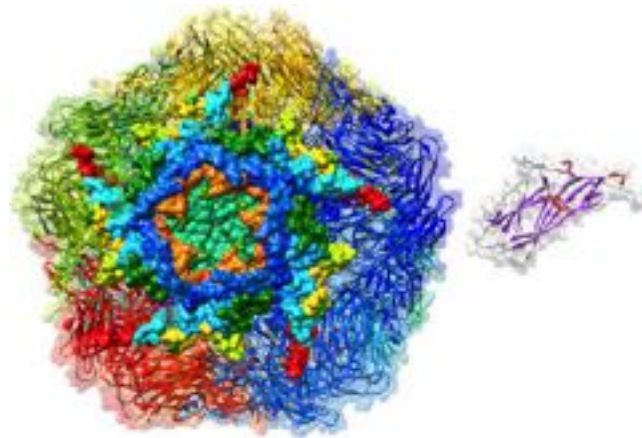


Fig. 2 Structure of the PCV-2 capsid estimated from cryo-EM of virus-like particles (PDB: 5ZBO). The surface of different capsomers has been colored with different nuances. In transparency, the ribbon structure has been depicted. The surfaces of the subunits constituting one of the pentamer clusters (center of the figure) are color-coded based on the corresponding amino-acid position within the protein. In the right insert, a single capsomer is displayed with the secondary structure highlighted in different colors (i.e., alpha helix and beta-sheets are displayed in red and purple, respectively).

Virion Structure

The virion structure has been accurately characterized for some member of the genus *Circovirus* only, while no morphological data is available for the *Cyclovirus* genus members.

Based on the available evidences, circoviruses are non-enveloped viruses with icosahedral $T = 1$ symmetry. Virion sizes change depending on the viral species, although *Porcine circovirus 1* and 2 (PCV-1 and PCV-2) and *Beak and feather disease virus* (BFDV) virions range from 15 to 25 nm in diameter. A similar viral structure can be expected based on the similarities in the capsid (Cap) protein among members of the same genus. In fact, the crystallographic three-dimensional reconstruction of PCV-2 and BFDV, despite their relative genetic distance (higher than approximately 55% and 65% at nucleotide and protein level, respectively), revealed a common organization, with 60 capsid protein subunits arranged in 12 pentameric clusters (Fig. 2).

The PCV-2 capsid protein is responsible for viral attachment through binding to heparan sulfate (HS) and chondroitin sulfate B (CSB) glycosamino-glycans on the cell surface. A recent structural study based on cryo-electron microscopy (cryo-EM) suggests that each subunit possesses up to five binding sites for heparin but only one can be occupied at a time. Therefore, each capsid might be composed of a unique combination of subunits ligated to heparin. Capsid protein also displays several basic amino acids in the N-terminal region, likely involved in DNA interaction and packaging.

Genome

Recognized members of the family *Circoviridae* are featured by a monopartite ambisense single stranded circular DNA genome ranging from approximately 1.7 to 2.1 kb. Two main ORFs (ORF1 and ORF2), located on different strands in the double-stranded DNA (dsDNA) replicative form, are present in all circoviruses and encode the Rep proteins (encoded by ORF1), fundamental for genome duplication, and Cap protein (encoded by ORF2), the only constituent of the viral capsid.

Although less characterized, other ORFs have been predicted *in silico* in several circoviruses. Unfortunately, experimental confirmation of their actual expression and function are often absent, with the most remarkable exception of BFDV and PCV-2. In the latter, up to 4 additional ORFs have been characterized and associated with several functions, including regulation of cell cycle, apoptosis and immune responses.

Circoviridae genomes display a nonanucleotide motif marking the origin of replication (*ori*). The Rep proteins introduce a nick in the virion-sense strand between positions 7 and 8 of this motif, initiating circovirus genome replication through rolling circle replication (RCR).

The two genera display some differences in the overall genome organization. *Circovirus* display two intergenic regions (IR) between ORF1 and ORF2. Also *Cyclovirus* genomes contain an IR between the 5' ends of Rep and Cap coding ORFs; however, the IR between the 3' ends of these ORFs is absent or significantly smaller than that observed in *Circovirus*. Additionally, as previously mentioned, the putative *ori* is located on the Rep-encoding strand of *Circovirus*, while in *Cyclovirus* is located on the opposite strand.

Life Cycle

Knowledge of the circovirus biological cycle and cell tropism is limited, and essentially obtained through studies on PCV-2.

PCV-2 has been proven to display a wide cell tropism *in vitro* (i.e., PK-15, PEDSV.15, aortic endothelial cells, fetal cardiomyocytes, gut epithelial cells, fibrocytes, monocytes, and a few others) and it is detected in several cells and tissues *in vivo*. However, the capability of each cell to sustain a productive infection is limited and subject of debate.

Likely, epithelial, lymphoblast and monocyte derived cells are the cell-types involved in the PCV-2 life cycle, although with different roles. In fact, while PCV-2 can be abundantly present in monocytes, macrophages and dendritic cells, its replication in these cells, if occurring, is limited. On the contrary, PCV-2 is able to actively replicate in certain subpopulations of lymphocytic cells, and mainly in epithelial cells both *in vitro* and *in vivo*.

Viral binding is mediated by capsid interaction with sulfate-rich glycosaminoglycans like HS and CSB. The internalization process follows different paths depending on the cell line involved, at least in *in vitro* experiments. In the monocytic cell line 3D4/31, the process is clathrin mediated, while in epithelial cells the virus may follow two internalization pathways: a clathrin-mediated one and a dynamin- and cholesterol-independent, but actin- and small GTPase-dependent one. Remarkably, only the latter leads to a full infection. In monocytes, macrophages and monocyte-derived and bone marrow-derived dendritic cells, PCV-2 enters the cell through a mechanism that is not fully characterized. Similarly, virus disassembly and DNA release in the cytoplasm depend on the involved cell line. In the monocytic cell line 3D4/31, it is released at acid pH, while in epithelial cells the process seems to occur at neutral pH. In monocytes, macrophages and monocyte-derived and bone marrow-derived dendritic cells, disassembly appears to not occur. Therefore, the virus can be stored for a long time without replicating in those cells, which likely participate in spreading the virus through the body. After uncoating, viral DNA is transferred to the nucleus, where replication occurs. Since circoviruses do not possess a DNA polymerase, they depend completely on the cell machinery to replicate. PCV-2 DNA replication begins during the first S-phase and depends on both S and G2/M phases of the cell cycle. Viral RNA transcription is mediated by cellular enzymes after ssDNA is converted to dsDNA. The transcription strategy is complex and several RNAs have been detected, although the role of most is unknown. Two Rep proteins (Rep and Rep') are translated from PCV-2 ORF1 RNA. While they share the same 5' end, they have different 3' (Rep' is obtained through alternative splicing). While both proteins display domains involved in RCR and are fundamental for genome replication, some of their other functions are likely different. Additionally, the replicase proteins form homo- and hetero-dimers and the ratio of Rep versus Rep' changes over the course of the infection, suggesting that different protein complexes could have a role in specific reactions. In contrast, only one Rep protein has been identified for other circoviruses.

DNA encapsidation occurs in the nucleus and virions are released into the cytoplasm by budding. Viral factories observed in the cytoplasm are bounded by multiple membranes, suggesting that the rough endoplasmic reticulum (RER) or Golgi apparatus may play a role in virion assembly. Clusters of viral particles, potentially moving from organelle membrane to the cytoplasm in the final replication stages, are then released from the plasma membrane by exocytosis or by budding from the cell membrane. Finally, an increase in cell death during the last hours of infection coupled with the presence of large clusters of virions forming paracrystalline arrays suggests that cell lysis could be another path for virus release.

Knowledge of other circovirus life cycles is essentially absent; however, the genome and protein resemblance, coupled with some similarity in the involved tissues and induced lesion (when present), could suggest a similar cycle.

No consistent data are currently available on the biological cycle of cycloviruses. Since the putative cyclovirus *ori* and predicted Rep share conserved features with members of the genus *Circovirus*, cycloviruses are also thought to replicate through RCR.

Epidemiology

The first *Circovirus* (later renamed PCV-1) was recognized in the 70s as a contaminant of porcine kidney cell lines. Thereafter, other circoviruses (i.e., BFDV, *Pigeon circovirus* (PiCV) and *Goose circovirus* (GoCV)) associated with clinical syndromes were reported in avian species in the 80s. However, despite overt clinical signs, the economic impact of these diseases remained limited. In the middle 90s, a dramatic change occurred with the emergence and discovery of PCV-2 as one of the most devastating viral diseases of the swine industry. This event drew attention towards this group of viruses. Since then, many *Circovirus* members have been reported in other avian and mammalian species (Fig. 1). Noteworthy, in most instances, no association with clinical signs could be demonstrated. *Canine circovirus* (CanineCV) and *Porcine circovirus 3* (PCV-3) are two potential exceptions since some evidence of their potential role in overt clinical syndromes has been reported.

Similar conclusions can be drawn for *Cyclovirus* members. Although some of these viruses were initially recognized in diseased subjects, including human beings, no causal association could be consistently established. Moreover, the actual host tropism is poorly known for most of them due to the lack of viral isolation and experimental models. In addition, several *Cyclovirus* DNAs have been detected and characterized from feces or digestive tract samples, which hinders the differentiation between contamination and actual infection.

Paleovirology studies, investigating the presence of sequences related to circoviruses integrated in the genome of vertebrate species, suggest that the origin of this virus family may exceed 40–50 million years. Therefore, the very recent discovery of new species is likely attributable to the improved identification and diagnostic capabilities. However, at least in some instances, other factors likely played a role in the emergence and expansion of these viruses. PCV-2 is a particularly illustrative example.

Retrospective studies demonstrated its presence in archived samples since 1962 and molecular clock-based studies dated it back at least at the beginning of the 20th century. Consequently, the switch from an apparently subclinical infection to a clinical one was likely due to the modern intensive farming conditions, which provided several cofactors (i.e., high animal density, stressing factors, management and husbandry related factors, co-infections, etc.) that enhanced viral replication and lead to signs and lesion appearance. A differential susceptibility of particular genetic lines of commercial pigs also seems to be involved. Finally, the evolution and adaptation of the virus is likely, as suggested by the emergence of new strains and genotypes over time, characterized by a partially different virulence and immunological features.

The multifactorial nature of PCV-2 disease is supported by the frequent subclinical course of the infection and by the evidence that PCV-2 alone is typically not sufficient to cause disease under experimental conditions. A comparable scenario could be true for PCV-3, a virus that likely circulated in the swine population for decades and is now frequently detected in both symptomatic and asymptomatic animals. Even in non-intensively raised species like dogs and psittacine birds, CanineCV and BFDV infection can be detected both in severely ill and healthy subjects. The reasons behind this behavior are not fully understood. If a similar picture can be expected for other *Circoviridae* remains to be established.

Unfortunately, most *Circovirus* and *Cylovirus* species have been detected only one or a few times. Therefore, data on their prevalence and distribution are largely lacking. The species for which adequate epidemiological and molecular epidemiology data are available, like BFDV, PCV-1, PCV-2, PCV-3, Duck circovirus (DuCV), GoCV and CanineCV, display a worldwide distribution with an absent-to-limited geographic clustering. However, these viruses affect domestic animals and their spreading patterns appear deeply connected to animal trade and movements. A different pattern could characterize circoviruses infecting wild species only. Significantly, some human cycloviruses initially considered to be limited to certain areas have demonstrated a wider geographic distribution than originally thought.

Circoviruses were traditionally thought to have a narrow host range. Nevertheless, even if a “preferred” host can be identified, sporadic detection in other species is described with increasing frequency. PCV-2 has been detected in wild boars, feral pigs, peccaries, but also in mice (used as experimental models) and rats (especially captured on commercial swine farms), ruminants and in fur animals (minks, foxes and raccoon dogs). Similarly, PCV-3 was detected in cows, dogs and laboratory mice, as well as in wild animals including wild boar, chamois, roe deer and related ticks. The actual clinical impact and epidemiological role of those hosts remains unclear, with the only exception being wild boars whose high PCV-2 and PCV-3 infection frequency suggests they could be a PCV reservoir in nature. The close genetic similarity between strains detected in wild boar and domestic pigs testifies to a frequent strain exchange between these populations, although the directionality of the viral flux is still unknown.

Detection of CanineCV in significant numbers of wild carnivores, including foxes, wolves and badgers, has been described, sometimes even with the presence of clinical signs.

As a group, ssDNA viruses are considered to be rapidly evolving organisms, characterized by a substitution rate comparable to that of RNA viruses. Although few studies have been performed on circovirus evolution, they seem to make no exception. A certain heterogeneity in evolutionary rate characterize different species. For example, PCV-2 substitution rates approach 10^{-3} substitutions/site/year, while that of PCV-3 is about two orders of magnitude lower. A similar slower rate evolution was also estimated for PCV-1. Detailed studies explaining such differences are still lacking, but a combination of higher population size and stronger selective pressures imposed by the host could play a pivotal role. Within-host viral variability has been investigated for PCV-2 only, revealing the emergence of several subpopulations over the course of infection. Significantly, the higher within-host variability was observed in the capsid coding region, suggesting the immune response as one of the major drivers of viral evolution. A comparable tendency has been described at a population level, where the capsid coding gene is the most variable and the one under stronger selective pressure. Importantly, evidence of vaccine-driven immune-selection have been provided and the emergence of multiple genotypes (i.e., cluster of strains with an independent evolution) alternating over time is partially due to an escape from the herd immunity developed against previously circulating genotypes.

In addition to mutation, frequent recombination has been reported within *Circovirus* member species, occurring especially in the intergenic regions. Epidemiological studies performed on PCV-2 have confirmed that recombination is a common phenomenon both within and between genotypes and demonstrated that recombinant viruses can have an evolutionary success comparable to parental ones.

Viral persistence, shedding and infection routes have been investigated for a limited subset of circoviruses. Oronasal is the primary route of PCV-2 transmission, which was detected in most animal secretions and excretions, including saliva, nasal, tonsillar, bronchiolar and ocular secretions, milk, colostrum, semen, feces and urine. Vertical, transplacental infection can also occur. Even though experimental studies have not been performed for PCV-3, virus detection in the same excretions, secretions and in aborted fetuses suggests similar transmission modes. Orofecal transmission is likely for CanineCV. Horizontal transmission occurs for BFDV, PiCV, DuCV and GoCV, even though vertical transmission cannot be excluded; in fact, vertical transmission has been suggested for PiCV.

Clinical Features

Circoviruses have been detected in the presence of several clinical syndromes in different species. However, a consistent association between infection and emergence of overt clinical signs is often lacking. Moreover, even when a causal nexus has been inferred, circovirus diseases are typically multi-factorial, infection being necessary but not sufficient to induce disease.

PCV-2 is the most studied pathogen belonging to this family and represents an illustrative example. PCV-2 seroprevalence suggests that PCV2 infection is ubiquitous all over the world and most pigs come in contact with the virus during their life. Nevertheless, morbidity is much lower and variable among countries, farms and over time. Even if most of infections remain subclinical (PCV-2-SI), their relevance cannot be ignored since even asymptomatic cases have been associated to a decrease in productive performance.

A number of clinical syndromes have been linked to PCV-2 infection, collectively named porcine circovirus diseases (PCVD). These can be broadly divided in systemic diseases and organ specific ones, although the differentiation can be only effectively made at a histological rather than clinical level. PCV-2 systemic disease (PCV-2-SD) is characterized by wasting or weight loss, pallor of the skin, respiratory distress, diarrhea, low grade fever and occasionally icterus. Enlarged subcutaneous lymph nodes are also a common finding. The onset of disease may be acute, leading to death within a few days in some pigs while others display a more chronic course and fail to gain weight or thrive. Morbidity in affected farms can reach 4%–30% (with a sporadic higher peak) and mortality ranges from 4% to 20%. PCV-2 lung disease (PCV-2-LD) and PCV-2 enteric disease (PCV-2-ED), postulated once as independent entities, are now believed to be part of PCV-2-SD. The frequent presence of co-infections and the role of PCV-2 in porcine respiratory disease complex (PRDC) further complicates the clinical scenario. PCV-2 is also involved in reproductive disease (PCV-2-RD), leading to mummifications, late term abortions and stillbirths. Return-to-estrus might also be linked to early embryo infection. However, the relevance of PCV-2-RD is apparently low under field conditions, likely because of the high seroprevalence and presumably immune coverage in adult sows. Actually, PCV-2-RD affected herds are typically start-up herds with high number of gilts or PCV-2 seronegative herds. Porcine dermatitis and nephropathy syndrome (PDNS) has also been linked to PCV-2 infection, although the role of PCV-2 in pathogenesis is debated. The most evident sign is the presence of irregular, red-to-purple macules and papules in the skin, primarily on the hind limbs and perineal area, but sometimes with a broader distribution, which over time evolve to dark crusts. The disease prevalence is low (less than 1%), but the lethality can reach 50%–100%. Pigs typically die within a few days after the onset of clinical signs due to acute renal failure while the survivors recover in 7–10 days.

Similar clinical symptoms have been described in PCV-3 infected pigs, including PDNS, reproductive disorders, enteric and respiratory disease. Nevertheless, PCV-3 is often detected in healthy animals. To date, high amounts of PCV-3 have been found in few cases of reproductive disease and nursery pigs with periarteritis and myocarditis.

CanineCV was initially identified in dogs with vasculitis and/or hemorrhagic gastroenteritis in the United States. Over time, it has been reported in several countries in presence of enteric signs, such as vomiting and diarrhea with hematochezia, and lymphadenitis. CanineCV has also been identified in histological lesions of the cerebrum from foxes with neurological disorders.

Among avian species, psittacine BFDV is probably the most relevant. Like in PCVDs, mortality and clinical signs in birds infected with BFDV are variable and dependent on the species of bird, age at the time of infection, presence or absence of concurrent infections and possibly the viral strain involved. Typically, the disease is detected in young birds of less than 3 years and is clinically characterized by a chronic, progressive loss of feathers, and, in some species, deformities in beak and claw. The infection can last for years; however, more often the animals die after some months due to secondary infections, since the condition causes an immunosuppressive state.

Circoviral infection in pigeons is associated with high morbidity and low mortality in most disease outbreaks. A broad range of signs and severity is observed in infected pigeons, including poor performance, diarrhea, and ill thrift.

Circovirus particles were detected in a canary affected by a condition known as “black spot”, which is associated with abdominal enlargement, gall bladder congestion, failure to thrive, dullness, anorexia, lethargy, and feather disorder. Growth retardation, feathering disorders and increased losses have been reported in circovirus-infected geese. Clinical signs of circovirus infection in Mallard ducks include feather dystrophy along the dorsum (hemorrhagic shafts) and poor body condition with low weight gain. In one report, circovirus infected finches displayed nasal discharge, dyspnea, anorexia, and depression.

A brief summary of most relevant clinical signs associated to circovirus infection are reported in [Table 1](#).

The clinical role of circoviruses is essentially unknown. Of note, CyCVs have been detected in cases of respiratory, enteric and neurological infections in humans and the presence of the virus has been found in serum, stool or cerebrospinal fluid (CSF) of patients with acute flaccid paralysis, paraplegia or suspected central nervous system infections. However, a causal association has not been proven.

Pathogenesis

Most circoviruses are considered able to cause disease in their corresponding hosts under certain conditions (not always well determined), although the majority of infections end up with subclinical scenarios. Circoviruses mainly replicate in epithelial and/or endothelial cells and, to a limited extent, in immune cells such as macrophages and lymphocytes. These animals developing clinical disease tend to suffer from immunosuppression and damage in the lympho-reticular tissues.

Among avian circoviruses, pathogenesis is mainly inferred from the knowledge generated on BFDV. This virus is considered to be epitheliotropic, targeting the basal epithelial layer of the feather and feather follicles. In fact, it is possible to detect circovirus-specific intranuclear inclusions in these cells, which are actively dividing. BFDV inclusions are also present in the cytoplasm of macrophages; although it has been suggested that most of the accumulated virions in the inclusions are due to the phagocytosing

Table 1 Summary of the main criteria to diagnose the well-recognized circoviral diseases in different animal species

Disease	Diagnostic criteria ^a		
	Clinical signs and gross pathology	Histopathology	Viral detection
PCV-2-systemic disease (PCV-2-SD)	Wasting Lymphadenopathy Interstitial pneumonia	Moderate to severe lymphocyte depletion with granulomatous inflammation of lymphoid tissues	Moderate to high amount of PCV-2 in lymphoid tissues by IHC/ISH ^b
PCV-2-reproductive disease (PCV-2-RD)	Abortion Mummified fetuses Weak-born piglets Myocardial necrosis	Necrotizing and fibrosing myocarditis	Moderate to high amount of PCV-2 in myocardium by IHC/ISH
PCV-2-subclinical infection (PCV-2-SI)	No clinical signs Lesser average daily weight gain	None or mild lymphocyte depletion with or without mild granulomatous inflammation of lymphoid tissues	Low amount of PCV-2 in tissues by IHC/ISH (or lack of PCV-2 in tissues by IHC/ISH but positive PCR)
Porcine dermatitis and nephropathy syndrome (PDNS)	Skin red papules and macules Kidney generalized petechiae Nephromegaly	Fibrinous-necrotizing glomerulonephritis Systemic necrotizing vasculitis	PCV-2 definitive causality on PDNS has not been demonstrated, so, its detection is not yet part of the diagnostic criteria
PCV-3 associated disease ^c	Abortion Mummified fetuses Weak-born piglets Congenital malformation	Non-suppurative systemic perivascularitis Non-suppurative myocarditis	Moderate to high amount of PCV-3 by ISH or low Ct values for qPCR ^d
Canine circovirus infection	Bloody diarrhea Vomiting Hemorrhagic enteritis	Systemic vasculitis Intestinal crypt necrosis Lymphocyte depletion of lymphoid tissues	Moderate to high amount of canine circovirus in tissues
Mink circovirus infection	Diarrhea Loss of feathers	Gastro-enteritis Epidermal cell necrosis, epidermal hyperplasia and hyperkeratosis	Amount of virus not defined to establish a diagnostic criterion
Beak and feather disease (BFD)	Deformity of the beak Deformity of the claw	Macrophage intracytoplasmic inclusions in affected epidermis and feather pulp	Moderate to high amount of BFDV by ISH or qPCR in tissues
Pigeon circovirus infection	Increased rearing losses Atrophy of thymus and bursa of Fabricius	Follicular hyperplasia and mild lymphocyte necrosis to lymphocyte depletion and histiocytosis Macrophage intracytoplasmic inclusion bodies	Moderate to high amount of PiCV by ISH or qPCR in tissues
Duck circovirus infection	Feathering disorders Growth retardation and low body weight	Lymphocytic depletion, necrosis, and degeneration of lymphoid tissues	Moderate to high amount of DuCV in tissues by qPCR
Finch, goose, gull and canary circovirus infections	Poorly described, but mainly growth retardation and low body weight in all these species	Lymphocytic depletion and necrosis of the bursa of Fabricius Macrophage intracytoplasmic inclusion bodies	Not well established criterion; presumed, moderate to high amount of these viruses

^aclinical signs and gross lesions are indicative; the final diagnosis is established by means of histopathology and virus detection.^bimmunohistochemistry/*in situ* hybridization.^calthough not yet clarified, PDNS has also been linked to PCV-3 infection.^dquantitative PCR.

activity of these cells rather than from endogenously replicated virus. The mechanism by which lymphoid organs such as the thymus and bursa of Fabricius become depleted in clinically affected birds is not entirely clear; it has been speculated that an indirect mechanism involving a cytokine-mediated reduction in lymphocyte differentiation might happen. Lymphocyte depletion is probably transient in subclinically infected animals, but persistent in diseased birds. Therefore, BFD-affected birds are susceptible to opportunistic secondary bacterial, chlamydial, and fungal infections.

Pathogenesis of circovirus infection in mammals has been documented more thoroughly than that of avian viral species. Specifically, the vast majority of data derives from PCV-2, the most studied *Circoviridae* family member, which has been associated with a number of economically significant diseases (PCVD). The greatest amount of PCV-2 is found in the cytoplasm of monocyte/macrophage lineage cells, both in PCV-2 diseased and subclinically infected pigs. *In vitro* studies have shown that PCV-2 is able to infect these cells, persisting for extended periods of time with apparently no or minimal active replication. Therefore, monocytic

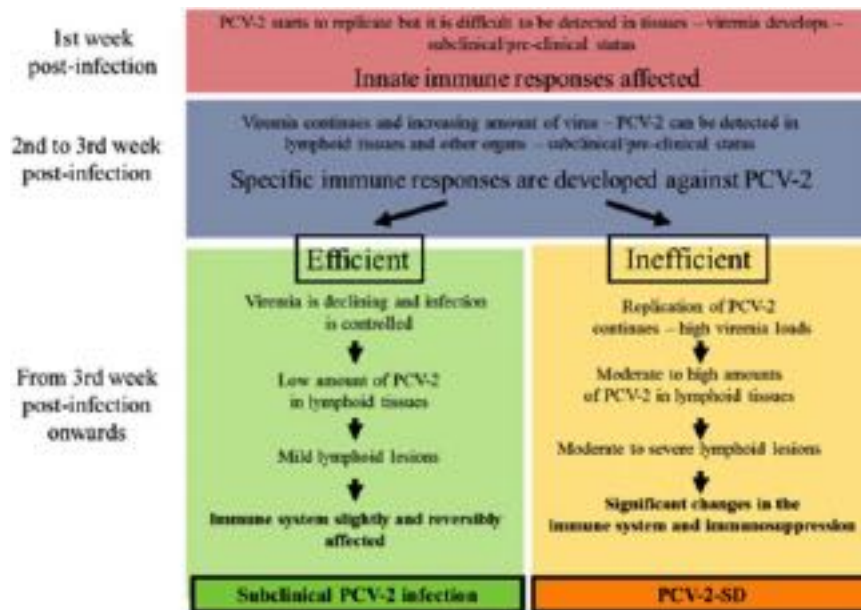


Fig. 3 Schematic representation of the pathogenesis of PCV-2 infection in post-natal swine. Adapted from: Porcine circovirus type 2: the virus, the disease and the vaccine, from Segalés, J. Grupo Asís Biomedica S.L., 2017.

cells seem not be the primary target for PCV2 replication and may represent a dissemination mechanism for PCV-2 throughout the host. Moreover, PCV-2 causes impairment of dendritic cell (DC) functionality *in vitro*, which may alter innate and virus-specific immune responses. However, *in vivo* information about these effects is minimal.

Cell types that supports PCV-2 replication *in vivo* are not completely elucidated. Initial work suggested that macrophages and lymphocytes did not play a significant role in PCV-2 replication, but subsequent research indicated that those cells types (mainly macrophages) may support replication to a certain degree. Ultrastructural studies in cell culture and lymph nodes from PCV-2-SD affected pigs have suggested that mitochondria may play a role in PCV-2 replication. Further characterization of PCV2-infected leukocyte subpopulations from peripheral blood mononuclear cells (PBMCs) indicated that mainly circulating T ($CD4^+$ and $CD8^+$) and potentially B lymphocytes, support PCV-2 replication; PBMC-derived monocytes apparently do not. PCV-2 also replicates in fetal myocardiocytes in experimentally infected porcine fetuses and in zona pellucida-free morulae and blastocysts.

Fig. 3 displays a proposed scheme for the post-natal pathogenesis of PCV-2 infections. PCV-2 viremia is first detectable around 7 days post-inoculation (PI), and viral load increases and reaches a peak between days 14 and 21 PI. Around 10–14 days PI, seroconversion against PCV2 is usually detected, especially if pigs remain subclinically infected. Under field conditions, the most likely age of infection is at the late nursery period and sero-conversion consequently occurs around 7–12 weeks of age; antibodies may last at least until 28 weeks of age. However, PCV-2 infection can occur at any age, and so evidence of infection and/or sero-conversion must be monitored by means of viral detection (usually PCR or quantitative PCR) and/or serological (mainly ELISAs) tests, respectively. Besides its detection in serum samples, PCV-2 can be detected in several organs, with lymphoid tissues containing the highest viral loads. PCV-2 amounts measured by quantitative PCR methods are usually 2–4 logs higher in lymphoid tissues compared to serum. The virus can be detected in virtually all tissues of a viremic pig. Immunohistochemistry and *in situ* hybridization techniques have shown presence of PCV-2 in epithelial cells from the kidney and respiratory tract, endothelial cells, lymphocytes, enterocytes, hepatocytes, smooth muscle cells, and pancreatic acinar and ductular cells.

The capacity of a pig to mount an adaptive immune response is key to prevent the progression of PCV-2 infection towards PCV-2-SD (**Fig. 3**). The onset of PCV-2 antibodies is followed by a decrease of viral loads and, eventually, resolution of viraemia. Long-lasting PCV2 viraemia and/or detection of PCV-2 in tissues, however, has been described in experimentally subclinically infected pigs despite the presence of high PCV-2 antibody titers. Inefficient humoral responses and, in particular, a poor PCV-2-neutralizing antibody response, are linked to increased viral replication, which results in severe lymphoid lesions and PCV-2-SD (**Fig. 3**). Under field conditions, from lactation to the growing-finishing period, the proportion of PCV-2 infected pigs and viral loads increase gradually, coincidentally with the waning of maternal-derived immunity.

Porcine embryos are susceptible to PCV-2 infection and their susceptibility varies with developmental stage. Virological and clinical outcomes of direct intra-fetal inoculation of PCV-2 at 57, 75 and 92 days of gestation have been described. At 21 days post-infection it was found that PCV-2 replicated in all inoculated fetuses but virus replication was significantly higher in fetuses inoculated at 57 days of gestation. In such fetuses, the heart displayed myocarditis-like lesions and harbored the highest virus load and number of infected cells. Fetuses inoculated at 75 days of gestation were stillborn at birth, while fetuses inoculated at 92 days of gestation did not show any lesion at birth. Fetuses inoculated at 75 or 92 days of gestation, but not the fetuses inoculated at

57 days, had antibodies against PCV-2 at delivery. PCV-2-infected cells identified in fetuses inoculated at different stages of gestation include cardiomyocytes, hepatocytes and cells of the monocyte/macrophage lineage.

Intranasal inoculation of pregnant sows or artificial insemination with purposely contaminated semen have also caused PCV-2 infection of fetuses/newborn piglets and reproductive failure. In contrast, sows artificially inseminated with semen of boars experimentally infected with PCV2 did not cause reproductive problems or infection of dams. Therefore, it is believed that the amount of PCV-2 naturally shed in boar semen is not sufficient to infect sows or their fetuses.

Little is known about the pathogenesis of PDNS, but it has been suggested that excessive PCV-2 serum antibody titers may be linked with the triggering of the condition. This hypothesis still awaits experimental confirmation: to date, studies on tissue sections from pigs with PDNS have failed to consistently demonstrate PCV-2 antigen or nucleic acid associated with PDNS lesions. Therefore, PCV-2 causality of PDNS is mainly based on circumstantial epidemiological and pathological evidences, but there is no experimental demonstration of such causation.

No data is currently available regarding the pathogenesis of PCV-3 infection. The first lesion descriptions associated to PCV-3 resembled those traditionally described for PCV-2-SD and PDNS, including as well myocarditis and systemic inflammation. The latest reports have emphasized the presence of PCV-3 genome mainly in inflammatory infiltrates in animals displaying perivascular inflammation at a systemic level, affecting especially kidney, liver, lymph nodes, lung and brain. Moreover, viral nucleic acid has also been found in follicular areas of lymph nodes, but usually in a low amount, similar to the location and distribution observed in pigs subclinically infected by PCV-2. Detection of the virus in tissues of stillborns and pigs with congenital anomalies points to a potential role in reproductive disease and, therefore, vertical transmission as an important route of dissemination.

Little is also known about the pathogenic role of CanineCV. On postmortem examination, some infected dogs showed histological lesions of necrotizing vasculitis and hemorrhage throughout the gastrointestinal tract, as well as granulomatous lymphadenitis of the lymph nodes. By *in situ hybridization* (ISH) analysis, abundant cytoplasmic viral nucleic acid has been detected in macrophages and monocytes in lymphoid tissues, including germinal centers and subcapsular and medullary sinuses of lymph nodes and ileal Peyer's patches. However, CanineCV has been detected in feces of healthy dogs and dogs with hemorrhagic diarrhea with a similar frequency. Therefore, CanineCV does not seem to be involved as a primary agent in dogs, and other co-factors might be necessary to elicit disease, similarly to other circoviruses.

The pathogenesis of cyclovirus infections is virtually unknown. All viruses of this genus have been detected in a number of samples by means of metagenomics and degenerate PCR methodologies, without an apparent association to a particular disease or condition. In fact, such detection has usually focused on the virome in the vertebrate and invertebrate species where they were detected. So far, the meaning of cyclovirus detection under disease scenarios is purely speculative. Importantly, techniques able to detect cycloviruses *in situ* in tissues or their putative humoral and cellular immune responses are still lacking.

Diagnosis

Diagnosis is the investigation or analysis of the cause of a condition, situation, or problem. Regarding *Circoviridae* family members, the terminology diagnosis should be applied at two major levels: identification of a disease problem (causal association of the agent with the condition) and identification of an agent, its components or antibodies against it. Taking into account the ubiquitous distribution of circoviruses and cycloviruses, disease causality is difficult to be established since in most of the cases a multifactorial origin is involved. Moreover, for just a limited number of circoviruses (but not of cycloviruses), a relatively consistent association with disease has been demonstrated (Table 1). The consistency of the association between infection and clinical disorder has been established using three main criteria: presence of compatible clinical signs, characteristic gross and histopathologic lesions, and presence of the virus (ideally by immunohistochemistry and/or *in situ* hybridization methods) within the lesions.

The most studied diseases have been by far PCVDs caused by PCV-2 and BFD caused by BFDV. In both cases, a significant clinical problem was first detected in swine and psittacine birds and the agent was subsequently discovered. Relevant gross and microscopic lesions were described in both diseases, being lymphocyte depletion and granulomatous inflammation of lymphoid tissues together with the presence of intracytoplasmic botryoid inclusion bodies common and very characteristic hallmarks. These histopathological lesions were subsequently confirmed in a number of circoviral infections in avian species (i.e., pigeon and geese). Importantly, avian circoviral diseases have been also associated to chronic, progressive loss of feathers, and, in some species, deformities of the beak and claws. The key diagnostic aspects of circoviral diseases are summarized in Table 1.

Techniques able to detect the viral antigen or nucleic acid within tissues displaying histopathological lesions are key to establish a sound diagnosis of disease. Immunohistochemistry (IHC) and ISH methodologies have been successfully developed to detect a number of circoviruses. Most of the techniques have been used to investigate the viral tropism for pathogenesis research purposes. Nevertheless, a routine diagnostic activity using IHC or ISH is performed only on a limited number of circovirus infections of clinical relevance such as PCVDs and BFD.

The study of *Circoviridae* family members has benefited enormously from the development of molecular techniques. As previously mentioned, *Cycloviruses* and most of *Circovirus* species have not been isolated yet. Therefore, the identification of their genome represents the only proof of their existence. In consequence, care must be taken in terms of interpretation of a given result.

A positive result indicates the presence of nucleic acid of the virus in a sample but does not provide any clue about the effects of the agent on the vertebrate/invertebrate animal. Quantification of the genome implies a further step, since the amount of genome for a number of circoviruses has been correlated with disease occurrence. However, the number of genome equivalent copies is also difficult to interpret for these agents that have never been associated with a certain disease. Taking into account that their infection dynamics are unknown and a particular amount of nucleic acid for clinical infections versus being part of the normal virome has not been established, the significance of their presence is still to be established.

A number of laboratory techniques have been developed to detect antibodies against circoviruses (not yet against cycloviruses). Some techniques such as the immunofluorescent antibody test and the immunoperoxidase monolayer assay have been used mainly in research, but implied the availability of a viral isolate, which limited their use almost exclusively for PCV-2. The development of ELISA tests using recombinant proteins widened the possibility of detecting antibodies against different circoviruses, such as PCV-2, BFDV, PiCV, GoCV, DuCV and mink circovirus (MiCV). However, only PCV-2 ELISA commercial kits are currently available for routine diagnosis. No serological tests have so far been developed for the detection of antibodies against cycloviruses.

Disease Control

Control of circoviral diseases have been mainly focused on counteracting the specific clinical signs of pet animals by means of supportive treatments. Importantly, these signs can vary greatly among individuals due to the immunosuppressive nature of the disease in clinically affected animals.

The initial control strategies for livestock species, mainly against PCV-2-SD, were developed to counteract the different infectious and non-infectious risks or triggering factors for the development of clinical disease. Noteworthy, most of the implemented measures were based on empirical results and not on contrasted and fully demonstrated efficacy. Overall, efforts were mainly directed to diminish/control the concomitant microbiological burden, decrease viral transmission and potentiate resistance against disease. Among these measures, the Madec's 20-point plan (a list of general zootechnical actions to lower the impact of disease) was the most practiced one, since they significantly decreased the percentage of mortality in severely affected farms. These measures were designed to reduce the infection pressure of PCV-2 as well as other infections, improve hygiene and reduce stress at different production stages.

The advent of commercial PCV-2 vaccines radically changed the clinical situation regarding PCV-2-SD and allowed the discovery of the so-called PCV-2-SI. The excellent efficacy of PCV-2 commercial vaccines surpassed the most optimistic expectations, to the point that nowadays it is the most used swine vaccine in the worldwide market. PCV-2 vaccines allowed a decrease in the impact of clinical signs and mortality, and more homogeneous animals in terms of body weight and improved average daily weight gain. Importantly, several commercial vaccines have appeared in the market and the different meta-analyses published on the subject indicated no significant differences in their efficiency to control PCV-2-SD and a significant impact for all the products when compared to non-vaccinated animals.

The most widely used vaccination scheme includes the use of the commercial products in piglets around weaning, using one (most often) or two doses. Also, several vaccines have been licensed to be applied in sows, which have been used to prevent PCV-2-SD in piglets (sows/gilts are vaccinated at the end of gestation) or to potentially prevent PCV-2-RD (sows/gilts are vaccinated pre-mating).

So far, no commercial vaccines are available for other circoviral diseases. Although the lack of virus isolates may potentially jeopardize initial testing of potential vaccine prototypes, most of the experimental work is performed with recombinant proteins derived from the sequences of the Cap gene from circoviruses. Using this approach several vaccine prototypes have been developed against some avian circoviruses and tested under experimental conditions. Such developments include prototypes against BFDV, DuCV and PiCV.

No vaccine products have been developed against the rest of circoviruses, mainly because they do not represent a significant threat to the corresponding infected animal populations. Also, it is not expected that vaccines will be developed against cycloviruses since their association with disease outcome is still to be demonstrated.

Conclusions

The number of family *Circoviridae* members has been significantly expanded in the last 10 years, and the novel taxonomy implied the removal of the genus *Gyrovirus* and the addition of the genus *Cyclovirus* together with the genus *Circovirus*. A number of circovirus species are unequivocally linked to clinical disease in animals, mainly in pigs and birds, although the associated disorders are of multifactorial nature. A common denominator of these conditions is the affection of the immune system of the animals, leading to immunosuppression in severely affected ones. Curiously, a large number of circoviruses and all known cycloviruses to date have never been isolated, and the only evidence of their existence is genome detection. Contrasting with the simplicity of these viral genomes is the complex and still poorly understood pathogenesis. Progress in molecular biology techniques and phylogenetic studies will surely facilitate elucidating evolutionary relationships among classified as well as newly identified members of the family *Circoviridae* in the future.

Further Reading

- Franzo, G., Cortey, M., Segalés, J., Hughes, J., Drigo, M., 2016. Phylodynamic analysis of porcine circovirus type 2 reveals global waves of emerging genotypes and the circulation of recombinant forms. *Molecular Phylogenetics and Evolution* 100, 269–280.
- Klaumann, F., Correa-Fiz, F., Franzo, G., *et al.*, 2018. Current knowledge on Porcine circovirus 3 (PCV-3): A novel virus with a yet unknown impact on the swine industry. *Frontiers in Veterinary Science* 5, 1–13.
- Ly, Q., Guo, K., Zhang, Y., 2014. Current understanding of genomic DNA of porcine circovirus type 2. *Virus Genes* 49, 1–10.
- Meng, X.J., 2013. Porcine Circovirus Type 2 (PCV2): Pathogenesis and Interaction with the Immune System. *Annual Review of Animal Biosciences* 1, 43–64.
- Meng, X.J., 2012. Spread like a wildfire-The omnipresence of porcine circovirus type 2 (PCV2) and its ever-expanding association with diseases in pigs. *Virus Research* 164, 1–3.
- Nauwynck, H.J., Sanchez, R., Meerts, P., *et al.*, 2012. Cell tropism and entry of porcine circovirus 2. *Virus Research* 164, 43–45.
- Rosario, K., Breitbart, M., Harrach, B., *et al.*, 2017. Revisiting the taxonomy of the family *Circoviridae*: Establishment of the genus *Cyclovirus* and removal of the genus *Gyrovirus*. *Archives of Virology* 162, 1447–1463.
- Segalés, J., 2012. Porcine circovirus type 2 (PCV2) infections: Clinical signs, pathology and laboratory diagnosis. *Virus Research* 164, 10–19.
- Steiner, E., Balmelli, C., Herrmann, B., Summerfield, A., McCullough, K., 2008. Porcine circovirus type 2 displays pluripotency in cell targeting. *Virology* 378, 311–322.
- Todd, D., 2004. Avian circovirus diseases: Lessons for the study of PMWS. *Veterinary Microbiology* 98 (2), 169–174.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/circoviridae
Circoviridae.

Coronaviruses: General Features (*Coronaviridae*)

Paul Britton, The Pirbright Institute, Pirbright, United Kingdom

© 2021 Published by Elsevier Ltd.

This is an update of D. Cavanagh, P. Britton, Coronaviruses: General Features, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00370-8](https://doi.org/10.1016/B978-012374410-4.00370-8).

Glossary

Infectious clone A full-length DNA copy of an RNA virus genome from which full-length viral RNA can be generated, leading to production of infectious virus.

Nidovirales (nidoviruses) An order comprising positive-sense RNA coronaviruses, toroviruses, arteriviruses, and roniviruses that have a common genome organization and expression, similar replication/transcription strategies, and

form a nested set of 3' co-terminal subgenomic mRNAs (*nidus*, Latin for nest).

Ribosomal frameshifting Movement (shift) backward by one nucleotide of a ribosome that is on an RNA, caused by particular RNA structures and sequences. Subsequent continuation of the progress of the ribosome is in a different open reading frame.

Introduction

Coronaviruses are known to cause disease in humans, other mammals, and birds. They cause major economic loss, sometimes associated with high mortality, in neonates of some domestic species (e.g., chickens, pigs). In humans, they are responsible for respiratory and enteric diseases. Coronaviruses do not necessarily observe species barriers, as illustrated most graphically by the recent spread of three zoonotic viruses in humans, severe acute respiratory syndrome (SARS) coronavirus, Middle East respiratory syndrome (MERS) coronavirus and in 2019 SARS-CoV-2 responsible for the disease Covid-19 causing a global Pandemic infection. All of which appear to have originated from bats via an intermediate wild animal species before infecting humans, with lethal consequences. As a group, coronaviruses are not limited to particular organs; target tissues include the nervous system, immune system, kidney, and reproductive tract in addition to many parts of the respiratory and enteric systems. A great advance in recent years has been the development of systems ('infectious clones') for modifying the genomes of coronaviruses to study all aspects of coronavirus replication, and for the development of new vaccines.

Taxonomy and Classification

Coronaviruses are part of the Order *Nidovirales*, which is divided into five Suborders; the *Arnidovirineae*, which contains the Family *Arteriviridae*; the *Cornidovirineae*, which contains the Family *Coronaviridae*; the *Tornidovirineae*, which contains the Family *Tobamiviridae*; the *Mesonidovirineae*, which contains the Family *Mesonviridae*; and the *Ronidovirineae*, which contains the Family *Roniviridae*. The viruses generically known as coronaviruses fall within the *Coronaviridae* Family which are divided into four genera, the *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*. Members of the *Nidovirales* Order have a similar genome organization and produce a nested set of subgenomic mRNAs and contain related specialized enzymes that are involved in replication of the RNA. Coronaviruses have been placed into different genera, initially on the basis of serological relationships which has subsequently been refined by gene sequencing.

Virion Properties

Virions have a buoyant density of approximately 1.18 g ml^{-1} in sucrose. Being enveloped viruses ([Fig. 1\(a\)](#)), they are destroyed by organic solvents such as ether and chloroform.

Virion Structure and Composition

All coronaviruses have four structural proteins in common ([Fig. 1\(b\)](#)): a large surface glycoprotein (S; c. 1150–1450 amino acids); a small envelope protein (E; c. 100 amino acids, present in very small amounts in virions); integral membrane glycoprotein (M; c. 250 amino acids); and a phosphorylated nucleocapsid protein (N; c. 500 amino acids). Some *Betacoronaviruses* have an additional structural glycoprotein, the hemagglutinin-esterase protein (HE; c. 425 amino acids). This is not essential for replication in vitro and may affect tropism in vivo.

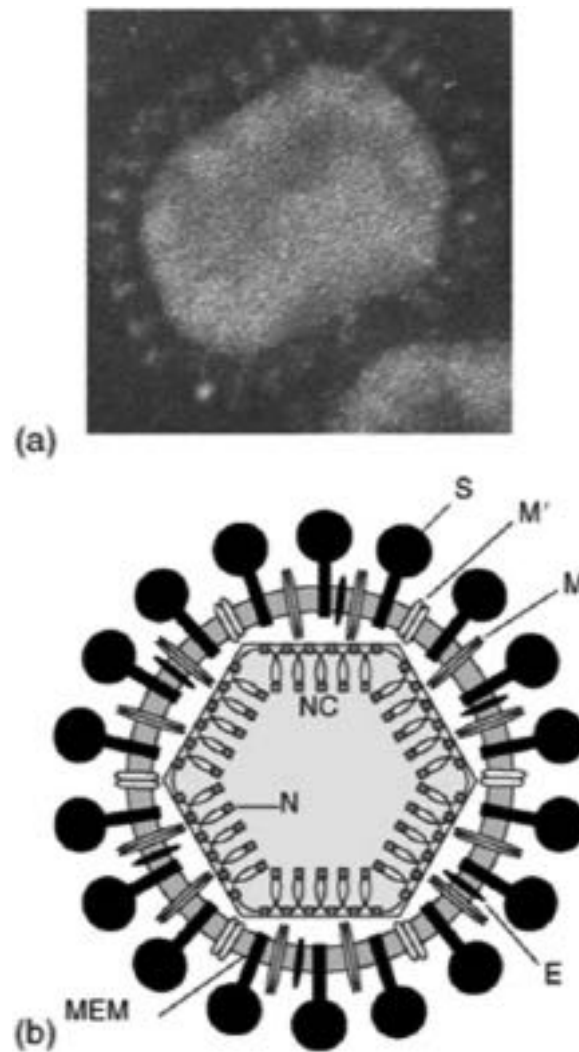


Fig. 1 (a) Electron micrograph of an IBV virion, showing the bulbous S protein. (b) Diagrammatic representation of the composition and structure of a coronavirus virion: S, spike glycoprotein; M, M', integral membrane glycoprotein; E, small envelope protein; N, nucleocapsid protein; NC, nucleocapsid (nucleoprotein) comprising the RNA genome and N protein. Cryoelectron microscopy of TGEV has indicated a core structure comprising the NC and the M protein. Two forms of M protein (M, M') have been observed for TGEV (see main text). The coronavirus membrane proteins, S, E, M, and M', are inserted into a lipid bilayer (MEM) derived from internal cell membranes. (b) Reproduced from González, J.M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A.E., Enjuanes, L., 2003. A comparative sequence analysis to revise the current taxonomy of the family. *Coronaviridae*. *Archives of Virology* 148, 2207–2235, with permission from Springer-Verlag.

Virions are *c.* 120 nm in diameter, although they can be up to twice that size, and the ring of S protein spikes is approximately 20 nm deep. When present, the HE protein forms a layer 5–10 nm deep. In some species, the S protein is cleaved into two subunits, the N-terminal S1 fragment being slightly smaller than the C-terminal S2 sequence. The S protein is anchored in the envelope by a transmembrane region near the C-terminus of S2. The functional S protein is highly glycosylated and exists as a trimer. The bulbous outer part of the mature S protein is formed largely by S1 while the stalk is formed largely by S2, having a coiled-coil structure. S1 is the most variable part of the S protein; some serotypes of the avian coronavirus, infectious bronchitis virus (IBV) differ from one another by 40% of S1 amino acids. S1 is the major inducer of protective immune responses. Variation in the S1 protein enables one strain of virus to avoid immunity induced by another strain of the same species.

The M glycoprotein is the most abundant protein in virions. In most cases, only a small part (~20 amino acids) at the N-terminus protrudes at the surface of the virus. There are three membrane-spanning segments and the C-terminal half of the M protein is within the lumen of the virus. In transmissible gastroenteritis virus (TGEV), a proportion of M molecules have four membrane-spanning segments, resulting in the C-terminus also being exposed on the outer surface of the virus (M' in Fig. 1(b)). The E protein is anchored in the membrane by a sequence near its N-terminus.

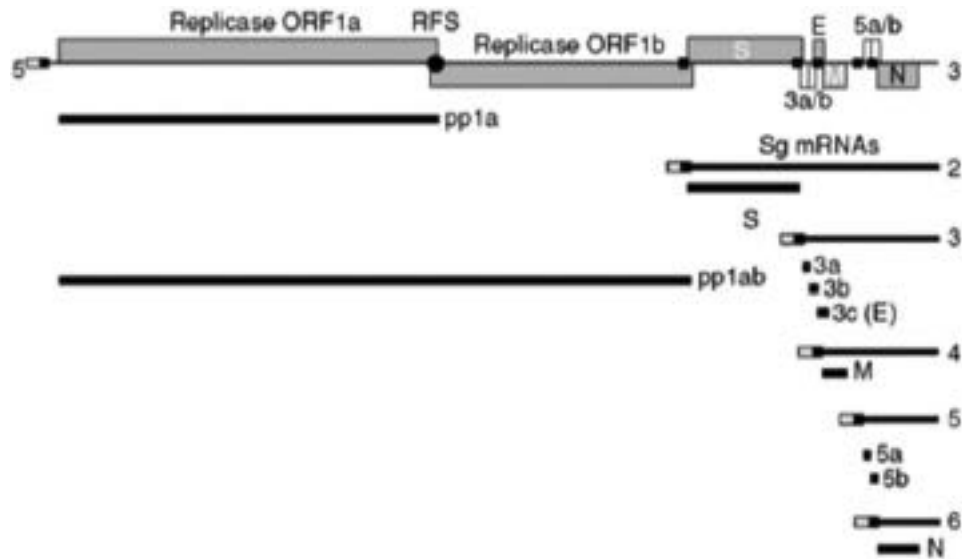


Fig. 2 Schematic diagram representing the genomic expression of the avian coronavirus IBV. The upper part of the diagram shows the IBV genomic RNA, with the various genes highlighted as boxed regions. The black boxes represent the transcription regulatory sequences (TRSs) found upstream of each gene and direct the synthesis, via negative-sense counterparts, of the sg mRNAs (2–6 for IBV). The leader sequence, represented by a gray box, is at the 5' end of the genomic RNA and at the 5' ends of the sg mRNAs. The genomic RNA is translated to produce two polyproteins, pp1a and pp1ab, that are cleaved by virus-encoded proteases to produce the replicase proteins. The structural proteins, S, E, M, and E, and the accessory proteins, 3a, 3b, 5a, and 5b, produced from IBV genes 3 and 5, respectively, are translated from the sg mRNAs. The proteins produced by the sg mRNAs are represented by lines below the corresponding sg mRNA. All of the sg mRNAs, except the smallest species, are polycistronic but only produce a protein from the 5'-most gene. The ribosome frameshift (RFS) region, denoted as a black circle on the genomic RNA, directs the -1 frameshift event for the synthesis of pp1ab. Translation of the genomic RNA results in the production of pp1a. However, the translating ribosomes undergo the -1 frameshift about 30% of the time resulting in pp1ab. The 5' and 3' UTR sequences are represented as single lines downstream of the leader and N gene sequences, respectively.

Genome Organization and Expression

Coronaviruses have the largest known RNA genomes, which comprise 28–32 kb of positive sense, single-stranded RNA. The overall genome organization being 5' UTR–polymerase gene–structural protein genes–3' UTR, where the UTRs are untranslated regions (Fig. 2). The first 60–90 nucleotides at the 5' end form a leader sequence. The structural protein genes are in the same order in all coronaviruses: (HE)–S–E–M–N. Interspersed among these genes are one or more other genes, often referred to as accessory genes as they have been demonstrated to be non-essential for replication in several coronaviruses using reverse genetic systems. The accessory genes encode small proteins of mainly unknown function. Some of these genes encode two or three proteins. In some cases (e.g., gene 3 of IBV and gene 5 of murine hepatitis virus (MHV)), translation of the third and second open reading frame (ORF), respectively, is affected by the preceding ORFs acting as internal ribosome entry sites. The proteins encoded by these small ORFs are mostly not required for replication *in vitro*; some of them might function as antagonists of innate immune responses.

Following entry into a cell and the release of the virus ribonucleoprotein (genome surrounded by the N protein) into the cytoplasm, ribosomes translate gene 1, which is approximately 20 kb, into two polyproteins (pp1a and pp1ab). These are cleaved by gene 1-encoded proteases, to generate 15 or 16 proteins (Fig. 3). Translation of gene 1 results in two polyproteins, translation of the ORF 1a region results in pp1a and translation of the ORF 1b region results in pp1ab, the latter involves ribosomal frameshifting, which has two elements, a slippery site followed by an RNA pseudoknot. At the slippery site (UUUAAAC in IBV), the ribosome slips one nucleotide backward and then moves forward, this time in a -1 frame compared with translation ORF 1a, resulting in the synthesis polyprotein 1ab, in which the proteins encoded by ORF 1b are in effect fused with the proteins from ORF 1a.

Proteins, including the RNA-dependent RNA polymerase (RdRp), from gene 1 associate to form the replicase complex, which is membrane associated. Coronavirus subgenomic mRNAs are generated by a discontinuous process. At the beginning of each gene is a common sequence (CUUAACAA in the case of IBV) called a transcription regulatory sequence (TRS). When the polymerase producing the nascent negative sense RNA, reaches a TRS, RNA synthesis is attenuated, followed by continuation at the 5' end of genomic RNA. This results in the addition of a negative copy of the leader sequence to the negative-sense RNA, resulting in a negative-sense copy of an sg mRNA. Of course, progress of the polymerase is not always halted at a TRS. Rather, it sometimes continues, producing a nested set of negative-sense sg mRNAs. These negative-sense copies of the sg mRNAs are the templates for the generation of the positive-sense sg mRNAs (Fig. 2). The amount of each sg mRNA does not necessarily decrease in a linear fashion; the efficiency of termination by a TRS is dependent on adjacent sequences, which are different for each gene. The leader sequence is found at the very 5' end of the genomic RNA and at the 5' ends of each sg mRNA.

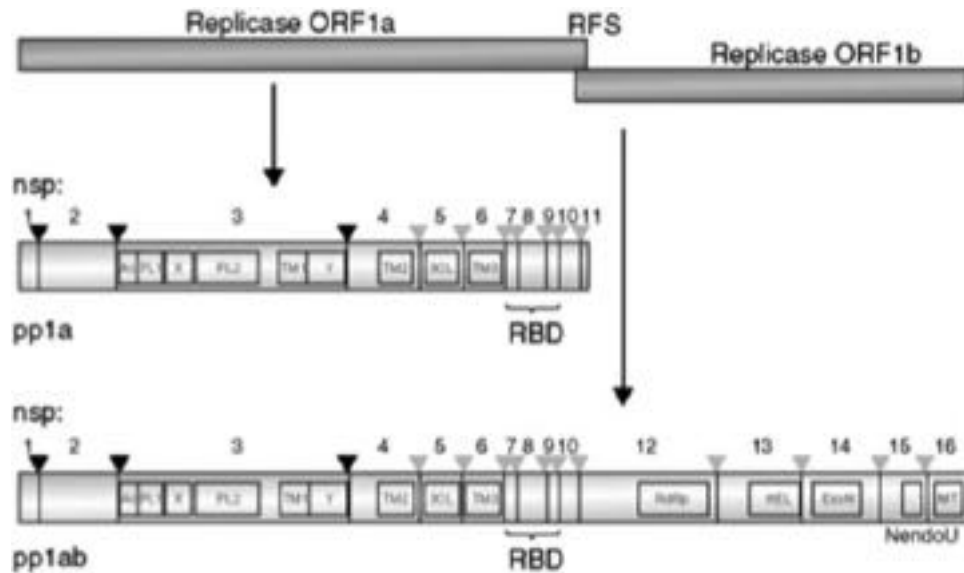


Fig. 3 Organization of the coronavirus replicase gene products. Translation of the coronavirus replicase ORF 1a and ORF 1b sequences results in pp1a and pp1ab; the latter is a C-terminal extension of pp1a, following a programmed -1 frameshift event (see legend to Fig. 2). The two polyproteins are proteolytically cleaved into 10 (pp1a; nsp1–11) and 16 (pp1ab; nsp1–16) products by the papain-like proteinases (PL1^{pro} and PL2^{pro}) and the 3C-like (3CL^{pro}) proteinase. The PL^{pro} proteinases cleave at the sites indicated with a black triangle and the 3CL^{pro} proteinase cleaves at the sites indicated with a gray triangle. The nsp11 product of pp1a is produced as a result of the ribosomes terminating at the ORF 1a translational termination codon, a -1 frameshift results in the generation of nsp12, part of the pp1ab replicase gene product. Various domains have been identified within some of the replicase products: Ac is a conserved acidic domain; X = ADP-ribose 1'-phosphatase (ADRP) domain; PL1 and PL2 the two papain-like proteinases; Y is a conserved domain; TM1, TM2, and TM3 are conserved putative transmembrane domains; 3CL = 3CL^{pro} domain; RdRp, RNA-dependent RNA polymerase domain; HEL, helicase domain; ExoN, exonuclease domain; NendoU, uridylate-specific endoribonuclease domain; MT, 2'-O-ribose methyltransferase domain. nsp's 7–9 contain RNA-binding domains (RBDs).

Replication Cycle

The N-terminal (S1) part of the S protein mediates attachment to cells. It is a determinant of host species specificity and, in some cases, pathogenicity, by determining susceptible cell range (tissue tropism) within a host. The C-terminal S2 part triggers fusion of the virus envelope with cell membranes (plasma membrane or endosomal membranes), which can occur at neutral or slightly acidic pH, depending on species or even strain. The virus glycoproteins (S, M, and HE, when present) are synthesized at the endoplasmic reticulum. Both subunits of the S protein are multiply glycosylated, while the M protein has one or two glycans close to its N-terminus. Interestingly, glycosylation of the M protein can be either N- or O-linked, depending on the type of coronavirus, although experiments using reverse genetics showed that conversion of an O-linked glycosylated M protein to an N-linked version had no effect on virus growth.

Early and late in infection, formation of virus particles can occur in the endoplasmic reticulum–Golgi intermediate compartment (ERGIC) and endoplasmic reticulum, but most assembly occurs in the Golgi membranes. The M protein is not transported to the plasma membrane; its location at internal membranes determines the sites of virus particle formation. It interacts with the N protein (as part of the RNP) and C-terminal part of the S protein, retaining some, though not all, of the S protein at internal membranes. The E protein is essential for virus particle formation, though it is not known how it functions. It has a sequence that determines its accumulation at internal membranes, and its interaction with the M protein. The latter's interaction with the N protein enables the formation of virus particles with spikes.

Genome Replication and Recombination

Following infection of a susceptible cell, the coronavirus genomic RNA is released from the virion into the cytoplasm and immediately recognized as an mRNA for the translation of the replicase pp1a and pp1ab proteins. These proteins are cleaved by ORF1a-encoded proteases, after which they become part of replicase complexes for the synthesis of either complete negative-sense copies of the genomic RNA or negative-sense copies of the sg mRNAs. The negative-sense RNAs are used as templates for the synthesis of genomic RNA and sg mRNAs (Fig. 2). RNA synthesis appears to take place in replication complexes associated with rearranged host membranes generated by several pp1a derived proteins. Following synthesis of the sg mRNAs, the structural proteins are produced for the assembly and encapsidation of the *de novo*-synthesized genomic RNA, resulting in the release of new infectious coronavirus virions. The release of new virions starts 4–7 h after the initial infection. As indicated above, the synthesis of the sg mRNAs is the result of a discontinuous process in which the synthesis of a negative-sense copy of an sg mRNA is completed

by the addition of the negative-sense leader sequence by a recombination mechanism. If a cell is infected with two related coronaviruses, the polymerase may swap between two RNA templates, in a similar way to addition of the leader sequence. This 'copy-choice' mechanism of genetic recombination results in a chimeric RNA. Such RNAs may give rise to new viruses with modified genomes with a capacity to infect a different cell and, in some cases, new host species.

Diseases and Host Range

Probably all coronaviruses replicate in epithelial cells of the respiratory and/or enteric tracts, though not necessarily producing clinical damage at those sites. The avian coronavirus (a *Gammacoronavirus*), IBV, not only causes respiratory disease but can also damage gonads in both females and males, and causes serious kidney disease (dependent on the strain of virus, and to some extent on the breed of chicken). IBV is able to replicate at virtually every epithelial surface in the host. Some coronaviruses have their most profound effect in the alimentary tract (e.g., the porcine coronavirus transmissible gastroenteritis virus (TGEV) causes $\geq 90\%$ mortality in neonatal pigs). Human coronaviruses are known to be associated with enteric and respiratory diseases (e.g., diarrhea), in addition to respiratory disease. SARS-CoV was also associated with diarrhea in humans, in addition to serious lung disease. Other coronaviruses, for example, MHV and porcine HEV, spread to cells of the central nervous system, producing disease, for example, acute or chronic demyelination in the case of MHV.

Coronavirus replication and disease are not necessarily restricted to a single host species. Canine enteric CoV and feline CoV can replicate and cause disease in pigs; these two viruses have proteins with very high amino acid identity to those of porcine TGEV. Canine respiratory CoV has proteins, including the S protein (which is the attachment protein and a determinant of host range), with very high amino acid identity ($\geq 95\%$) to other group 2 viruses Hu CoV-OC43 and BCoV. This raises the possibility of co-infection in these hosts. There is evidence that pheasant CoV can infect chickens, and IBV infect teal (a duck), though without causing disease. The most dramatic demonstration that coronaviruses can have a wide host range was provided by SARS-CoV. This appears to have its origin in bats, was transferred to various other species (e.g., civet cat) that were captured for trade, and then caused lethal disease in humans. A similar zoonotic pathway has been hypothesized for the infection of humans by SARS-CoV-2, again with an origin in bats but via, as yet unknown, intermediary host or hosts, though generally accepted via a wild animal in a live animal market.

Persistent infections *in vivo* are well known for MHV, and less well known for other coronaviruses (e.g., IBV). Following infection of very young chickens, IBV is re-excreted when hens start to lay eggs. The trigger for release is probably the stress of coming into lay.

The S protein is a determinant of both tissue tropism within a host and host range. This has been elegantly demonstrated by genetic manipulation of the genome of MHV, which is unable to attach to feline cells. Replacement of the MHV S protein gene with that of CoV from feline coronavirus resulted in a recombinant virus that was able to attach, and subsequently replicate in, feline cells. However, other proteins can also affect pathogenicity. Research with genetically modified coronaviruses, using targeted recombination or 'infectious clones', has shown that modifications to proteins encoded in ORF1 and the small genes interspersed among the structural protein genes, result in attenuation of pathogenicity. Although the roles of these 'accessory proteins' are not known, this may offer a route to the development of a new generation of live vaccines. Currently, the most widely used prophylactics for control of IBV in chickens include killed vaccines and live vaccines attenuated by passage in embryonated eggs. However, disease control is complicated by extensive variation in the S1 protein which is the inducer of protective immunity.

See also: Coronaviruses: Molecular Biology (*Coronaviridae*). Enveloped, Positive-Strand RNA Viruses (*Nidovirales*). Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (*Coronaviridae*). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (*Coronaviridae*)

Further Reading

- Britton, P., Cavanagh, D., 2007. Nidovirus genome organization and expression mechanisms. In: Perlman, S., Gallagher, T., Snijder, E.J. (Eds.), *Nidoviruses*. Washington, DC: ASM Press, pp. 29–46.
- Britton, P., Cavanagh, D., 2007. Avian coronavirus diseases and infectious bronchitis vaccine development. In: Thiel, V. (Ed.), *Coronaviruses: Molecular and Cellular Biology*. Norfolk, UK: Caister Academic Press, pp. 161–181.
- Cavanagh, D., 2003. SARS vaccine development: Experiences of vaccination against avian infectious bronchitis coronavirus. *Avian Pathology* 32, 567–582.
- Cavanagh, D., 2005. *Coronaviridae: A review of coronaviruses and toroviruses*. In: Schmidt, A., Wolff, M.H. (Eds.), *Coronaviruses with Special Emphasis on First Insights Concerning SARS*. Basel: Birkhäuser, pp. 1–54.
- Cavanagh, D., 2005. Coronaviruses in poultry and other birds. *Avian Pathology* 34, 439–448.
- Enjuanes, L., Almazán, F., Sola, I., Zúñiga, S., 2006. Biochemical aspects of coronavirus replication: A virus–host interaction. *Annual Reviews in Microbiology* 60, 211–230.
- González, J.M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A.E., Enjuanes, L., 2003. A comparative sequence analysis to revise the current taxonomy of the family *Coronaviridae*. *Archives of Virology* 148, 2207–2235.
- Masters, P.S., 2006. The molecular biology of coronaviruses. *Advances in Virus Research* 66, 193–292.
- Siddell, S., Ziebuhr, J., Snijder, E., 2005. Coronaviruses, toroviruses and arteriviruses. In: Mahy, B.W.J., ter Meulen, V. (Eds.), *Topley and Wilson's Microbiology and Microbial Infections, Virology*. London: Hodder Arnold, pp. 823–856.

Coronaviruses: Molecular Biology (*Coronaviridae*)

X Deng and SC Baker, Loyola University of Chicago, Maywood, IL, United States

© 2021 Elsevier Ltd. All rights reserved.

This is a reproduction of X. Deng, S.C. Baker, Coronaviruses: Molecular Biology, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:10.1016/B978-0-12-801238-3.02550-2.

Glossary

Cell tropism Process that determines which cells can be infected by a virus. Factors such as receptor express can influence the cell type that can be infected.

Discontinuous transcription Process by which the coronavirus leader sequence and body sequence are joined to generate subgenomic RNAs.

Double membrane vesicles (DMVs) Vesicles that are generated during coronavirus replication when viral replicase proteins sequester host cell membranes. These vesicles are the site of coronavirus RNA synthesis.

Transcriptional regulatory sequences (TRSs) Sequences that are recognized by the coronavirus transcription complex to generate leader-containing subgenomic RNAs.

Introduction

Coronaviruses (CoVs) were first identified during the 1960s by using electron microscopy to visualize the distinctive spike glycoprotein projections on the surface of enveloped virus particles. It was quickly recognized that CoV infections are quite common, and that they are responsible for seasonal or local epidemics of respiratory and gastrointestinal disease in a variety of animals. CoVs have been named according to the species from which they were isolated and the disease associated with the viral infection. Avian infectious bronchitis virus (IBV) infects chickens, causing respiratory infection, decreased egg production, and mortality in young birds. Bovine coronavirus (BCoV) causes respiratory and gastrointestinal disease in cattle. Porcine transmissible gastroenteritis virus (TGEV) and porcine epidemic diarrhea virus (PEDV) cause gastroenteritis in pigs. These CoV infections can be fatal in young animals. Feline infectious peritonitis virus (FIPV) and canine coronavirus (CCoV) can cause severe disease in cats and dogs. Depending on the strain of the virus and the site of infection, the murine CoV mouse hepatitis virus (MHV) can cause hepatitis or a demyelinating disease similar to multiple sclerosis. CoVs also infect humans. Human coronaviruses (HCoVs) 229e and OC43 are detected worldwide and are estimated to be responsible for 5–30% of common colds and mild gastroenteritis. Interestingly, HCoV-OC43 and BCoV share considerable sequence similarity, indicating a likely transmission across species (either from cows to humans or vice versa) and then adaptation of the virus to its host. In contrast to the relatively mild infections caused by HCoV-229e and HCoV-OC43, the CoV responsible for severe acute respiratory syndrome (SARS-CoV) causes atypical pneumonia with a 10% mortality rate. Two additional HCoVs, HCoV-NL63 and HCoV-HKU1, were identified using molecular methods and are associated with upper and lower respiratory tract infections in children, and elderly and immunosuppressed patients. Most recently, a coronavirus likely endemic in the camel population has been associated with Middle East Respiratory Syndrome (MERS) in humans. MERS-CoV can cause a lethal respiratory disease and renal syndrome, and mortality is highest in older individuals with co-morbidities such as diabetes or immunosuppression. CoVs are grouped according to sequence similarity as alpha-, beta, gamma or deltacoronaviruses (see www.viprbrc.org for a complete list of reference genomes for coronaviruses).

To date, the most infamous example of zoonotic transmission of a CoV is the outbreak of SARS in 2002–03. We now know that the outbreak started with cases of atypical pneumonia in the Guangdong Province in southern China in the fall of 2002. The infection was spread to tourists visiting Hong Kong in February, 2003, resulting in the dissemination of the outbreak to Hong Kong, Vietnam, Singapore, and Toronto, Canada. After attempting to treat cases of atypical pneumonia in Vietnam and acquiring the infection himself, Dr. Carlo Urbani alerted the World Health Organization (WHO) that this disease of unknown origin may be a threat to public health. The WHO rapidly organized an international effort to identify the cause of the outbreak, and within months a novel CoV was isolated from SARS patients and identified as the causative agent. Sequence analysis revealed that the virus was related to, but distinct from, all known CoVs. This led to an intensive search for an animal reservoir for this novel CoV. Initially, the masked palm civet and raccoon dog were implicated in the chain of transmission, since a SARS-CoV-like virus could be isolated from some animals found in wild animal markets in China. However, SARS-CoV-like viruses were never detected if these animals were captured from the wild, indicating that the civets may have only served as an intermediate host in the chain of transmission. Further investigation revealed that the likely reservoir for SARS-CoV is the Chinese horseshoe bat (*Rhinolophus* spp.), which is endemically infected with a virus, named bat-SARS-CoV, that is closely related to SARS-CoV (Fig. 1(a)). The bringing together of CoV-infected bats and susceptible civet cats and humans in live animal markets in China likely contributed to the emergence of SARS-CoV in 2002–2003.

In 2012, a novel coronavirus eventually termed Middle East Respiratory Syndrome Coronavirus (MERS-CoV) was isolated from a patient suffering from pneumonia in Saudi Arabia. MERS-CoV was then linked to multiple outbreaks of severe respiratory disease in the Middle East. Sequence analysis revealed that MERS-CoV was similar to but distinct from CoVs previously identified in bats in Europe and Asia. Further epidemiologic studies revealed that MERS-CoV could be isolated from the respiratory and

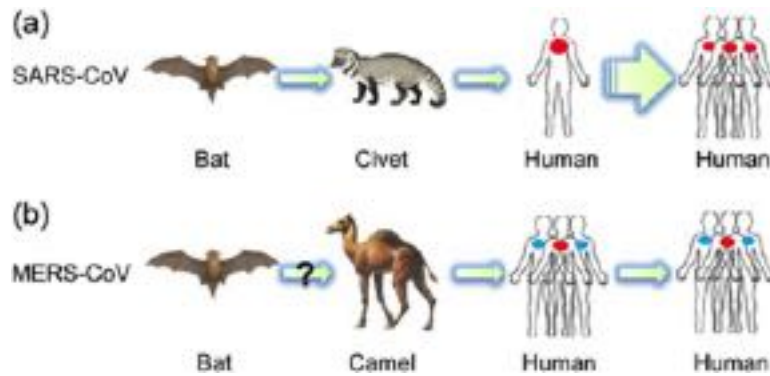


Fig. 1 Zoonotic transmission of emerging coronaviruses. (a) SARS-like coronaviruses are endemic in bats in Southern China. Virus transmission events may occur when animals are brought together in live animal markets. Replication of the virus in an intermediate host allows for the acquisition of adaptive mutations that allow for an expanded host range. SARS-CoV adapted to use the ACE-2 receptor in humans and can be spread by respiratory droplets. Infection with SARS-CoV generally results in symptomatic infection (red dots) characterized by fever and pneumonia. Super-spreader events (one person infecting many others) followed by global travel led to the pandemic of 2003. The quarantine of infected individuals and their contacts led to the cessation of the outbreak. (b) MERS-like viruses have been identified in bats and may be the progenitor of the virus now recognized to be endemic in camels in parts of the Middle East and Africa. MERS-CoV infection of camels seems to cause a respiratory disease with no significant mortality in young or adult camels. MERS-CoV infection in humans, proposed to be acquired from contact with camels or raw camel milk or meat, may cause pneumonia (red dot) or mild or asymptomatic illness (blue dot). Currently, there seems to be limited human-to-human transmission of MERS-CoV.

gastrointestinal tract of young camels and that the majority of adult camels tested had neutralizing antibodies to MERS-CoV, consistent with the idea that this virus was endemic in the camel population in the Middle East and in some regions of Africa. Currently, the majority of MERS-CoV infections are proposed to be acquired from either contact with camels or camel products such as raw milk or meat. There is some evidence supporting human to human transmission of MERS-CoV in hospital and home settings, however to date the virus has not evolved to be transmitted easily from human to human ([Fig. 1\(b\)](#)). Estimates of mortality from MERS-CoV are in the range of 30–40%, so there is significant concern about the pandemic potential of this emerging coronavirus should it evolve to transmit efficiently between humans. The existence of animal reservoirs for SARS-CoV and MERS-CoV presents the possibility for sporadic re-emergence of these significant human pathogens. By improving our understanding of the molecular aspects of CoV replication and pathogenesis, we may facilitate development of appropriate antiviral agents and vaccines to control and prevent diseases caused by known and potentially emerging CoVs.

Molecular Features of CoVs

CoV virions ([Fig. 2\(a\)](#)) are composed of a large RNA genome, which combines with the viral nucleocapsid protein (N) to form a helical nucleocapsid, and a host cell-derived lipid envelope which is studded with virus-specific proteins including the membrane (M) glycoprotein, the envelope (E) protein, and the spike (S) glycoprotein. CoV particles vary somewhat in size, but average about 100 nm in diameter. The genomic RNA (gRNA) inside the virion, which ranges in size from 27 to 32 kb for different CoVs, is the largest viral RNA identified to date. CoV gRNAs have a broadly conserved structure which is illustrated by the SARS-CoV genome shown in [Fig. 2\(b\)](#). The gRNA is capped at the 5' end, with a short leader sequence followed by two long open reading frames (ORFs) encoding the replicase polypeptide. The remaining part of the genome encodes the viral structural and so-called accessory proteins. The structural protein genes are always found in the order S–E–M–N, but accessory protein genes may be interspersed at various sites between the structural genes. SARS-CoV has the most complex genome yet identified, with eight ORFs encoding accessory proteins. The expression of these ORFs is not required for viral replication, but they may play a role in the pathogenesis of SARS and MERS. In addition, the products of accessory genes may be incorporated into the virus particle, potentially altering the tropism or enhancing infectivity. For SARS-CoV, the proteins encoded in ORFs 3a, 6, 7a, and 7b have been shown to be incorporated in virus particles, but the exact role of these proteins in enhancing virulence is not yet clear.

The features of CoV structural proteins are shown in [Fig. 3](#). For each structural protein, a schematic diagram of the predicted structure of the protein is shown on the left and a linear display of the features is shown on the right. The CoV spike glycoprotein is essential for attachment of the virus to the host cell receptor and fusion of the virus envelope with the host cell membrane. CoV spike glycoproteins assemble as trimers with a short cytoplasmic tail and hydrophobic transmembrane domain anchoring the protein into the membrane. The spike glycoprotein is divided into the S1 and S2 regions, which are sometimes cleaved into separate proteins by cellular proteases during the maturation and assembly of virus particles. S1 contains the receptor-binding domain (RBD) and has been shown to provide the specificity of attachment for CoV particles. The cellular receptors and corresponding RBDs in S1 have been identified for several CoVs. MHV binds to murine carcinoembryonic antigen-related cell adhesion molecules (mCEACAM1 and mCEACAM2); TGEV, FIPV, and HCoV-229e bind to species-specific versions of

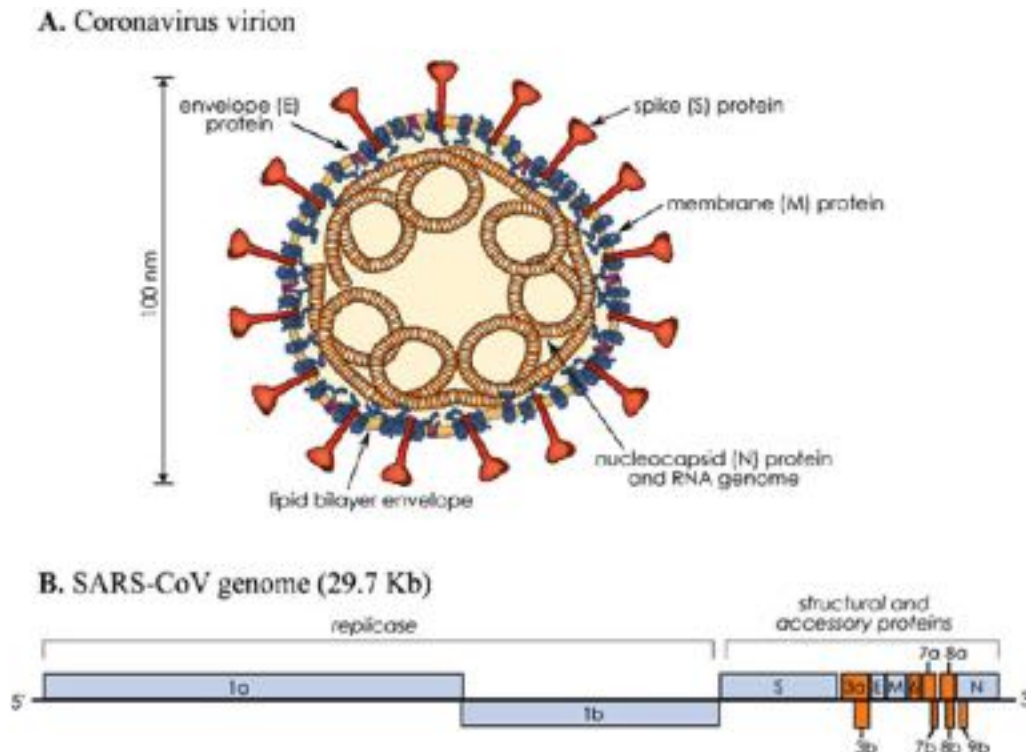


Fig. 2 CoV virion and the genome of SARS-CoV. (a) Schematic diagram of a CoV virion with the minimal set of four structural proteins required for efficient assembly of the infectious virus particles: S, spike glycoprotein; M, membrane glycoprotein; E, envelope protein; and N, nucleocapsid phosphoprotein which encapsidates the positive-strand RNA genome. (b) Schematic diagram of the gRNA of SARS-CoV. Translation of the first two open reading frames (ORF1a and ORF1b) generates the replicase polyprotein. ORFs encoding viral structural and accessory (orange) ORFs are indicated at the 3' end of the genome. (a) Reprinted from Masters PS (2006) *The molecular biology of coronaviruses. Advances in Virus Research* 66: 193–292, with permission from Elsevier.

aminopeptidase N. Interestingly, both HCoV-NL63 and SARS-CoV have been shown to bind to human angiotensin-converting enzyme 2 (ACE2). MERS-CoV binds to dipeptidyl peptidase (DPP4 or CD26). ACE2 and DPP4 are expressed in both the respiratory and gastrointestinal tracts, consistent with virus replication at both of these sites.

Once the S1 portion of the spike has engaged the host cell receptor, the protein undergoes a dramatic conformational change to promote fusion with the host cell membrane. Depending on the virus strain, this can occur at the plasma membrane on the surface of the cell, or in acidified endosomes after receptor-mediated endocytosis. The critical elements in the conformational change are the heptad repeats, HR1 and HR2, and the fusion peptide, F. After engaging the receptor, there is a dissociation of S1 which likely triggers the rearrangement of S2 so that HR1 and HR2 are brought together to form an antiparallel, six-helix bundle. This new conformation brings together the viral and host cell membranes and promotes the fusion of the lipid bilayers and introduction of the nucleocapsid into the cytoplasm. During infection, the spike glycoprotein is also present on the surface of the infected cell where it may (depending on the virus strain) promote fusion with neighboring cells and syncytia formation. The spike glycoprotein is also the major antigen to which neutralizing antibodies develop. The spike protein is a target for development of therapeutics for treatment of CoV infections. Monoclonal antibodies directed against the spike neutralize the virus by blocking binding to the receptor; synthetic peptides that block HR1-HR2 bundle formation have also been shown to block CoV infection.

The membrane (M) and envelope (E) proteins are essential for the efficient assembly of CoV particles. M is a triple-membrane-spanning protein that is the most abundant viral structural protein in the CoV virion. The ectodomain of M is generally glycosylated, and is followed by three transmembrane domains and an endodomain which is important for interaction with the nucleocapsid protein and packaging of the viral genome. The E protein is present in low copy numbers in the virion, but is important for efficient assembly. In the absence of E protein, few or no infectious virus particles are produced. The exact role of the E protein in the assembly of virus particles is still unknown, but recent studies suggest that E may act as an ion channel. The nucleocapsid protein (N) is an RNA-binding protein and associates with the CoV gRNA to assemble ribonucleoprotein complexes. The N protein is phosphorylated, predominantly at serine residues, but the role of phosphorylation is currently unknown. The N protein has three conserved domains, each separated by highly variable spacer elements. Domains 1 and 2 are rich in arginine and lysine residues, which is typical of many RNA-binding proteins. Domain 3 is essential for interaction with the M protein and assembly of infectious virus particles. The N protein has been shown to be an important cofactor in CoV RNA synthesis and is proposed to act as an RNA chaperone to promote template switching, as described below.

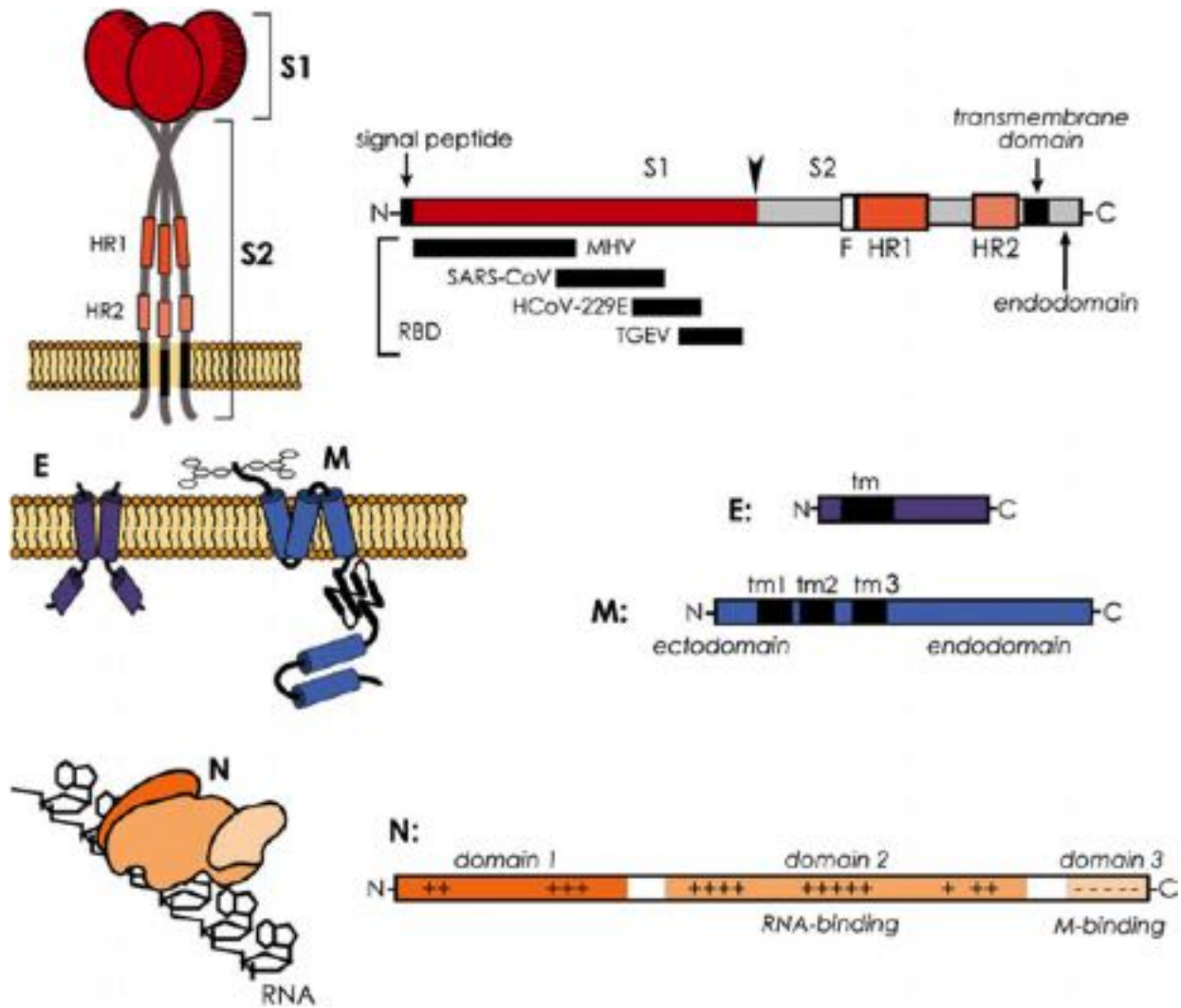


Fig. 3 Diagrammatic representation of the spike trimer assembled on membranes, with the S1 receptor binding domain (RBD) and S2 fusion domain indicated. The linear map of spike indicates the location of the RBDs for three CoVs, and the relative location of the heptad repeat domains 1 and 2 (HR1 and HR2) which mediate the conformational changes required to present the fusion peptide (F) to cellular membranes. The membrane (M), envelope (E), and nucleocapsid (N) proteins represented in association with membranes or viral RNA. The linear map of each protein highlights the transmembrane domains of M and E and the RNA-binding and M protein-binding domains of N. Domains 1 and 2 of N are rich in arginine and lysine (indicated by +). Reprinted from Masters PS (2006) The molecular biology of coronaviruses. *Advances in Virus Research* 66: 193–292, with permission from Elsevier.

Replication and Transcription of CoV RNA

The replication and transcription of CoV RNA takes place in the cytoplasm of infected cells (Fig. 4). The CoV virion attaches to the host cell receptor via the spike glycoprotein and, depending on the virus strain, the spike mediates fusion directly with the plasma membrane or the virus undergoes receptor-mediated endocytosis and spike-mediated fusion with endosomal membranes to release the viral gRNA into the cytoplasm. Once the positive-strand RNA genome is released, it acts as a messenger RNA (mRNA) and the 5' end (ORF1a and ORF1b) is translated by ribosomes to generate the viral RNA-dependent RNA polymerase polypeptide, termed the viral replicase. Translation of ORF1b is dependent on ribosomal frameshifting, which is facilitated by a slippery sequence and RNA pseudoknot structure present in all CoV gRNAs. The replicase polypeptide is processed by replicase-encoded proteases (papain-like proteases and a poliovirus 3C-like protease) to generate 16 mature replicase products. These viral replicase proteins sequester host cell membranes to generate distinctive double-membrane vesicles (DMVs) that have been shown to be the site of CoV RNA synthesis. The replicase complex on the DMVs then mediates the replication of the positive-strand RNA genome to generate full-length and subgenomic negative-strand RNAs, and the subsequent production of positive-strand gRNAs and sgmRNAs. The sgmRNAs are translated to generate viral structural and accessory proteins, and virus particles assemble with positive-strand gRNA in the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) and bud into vesicles, with

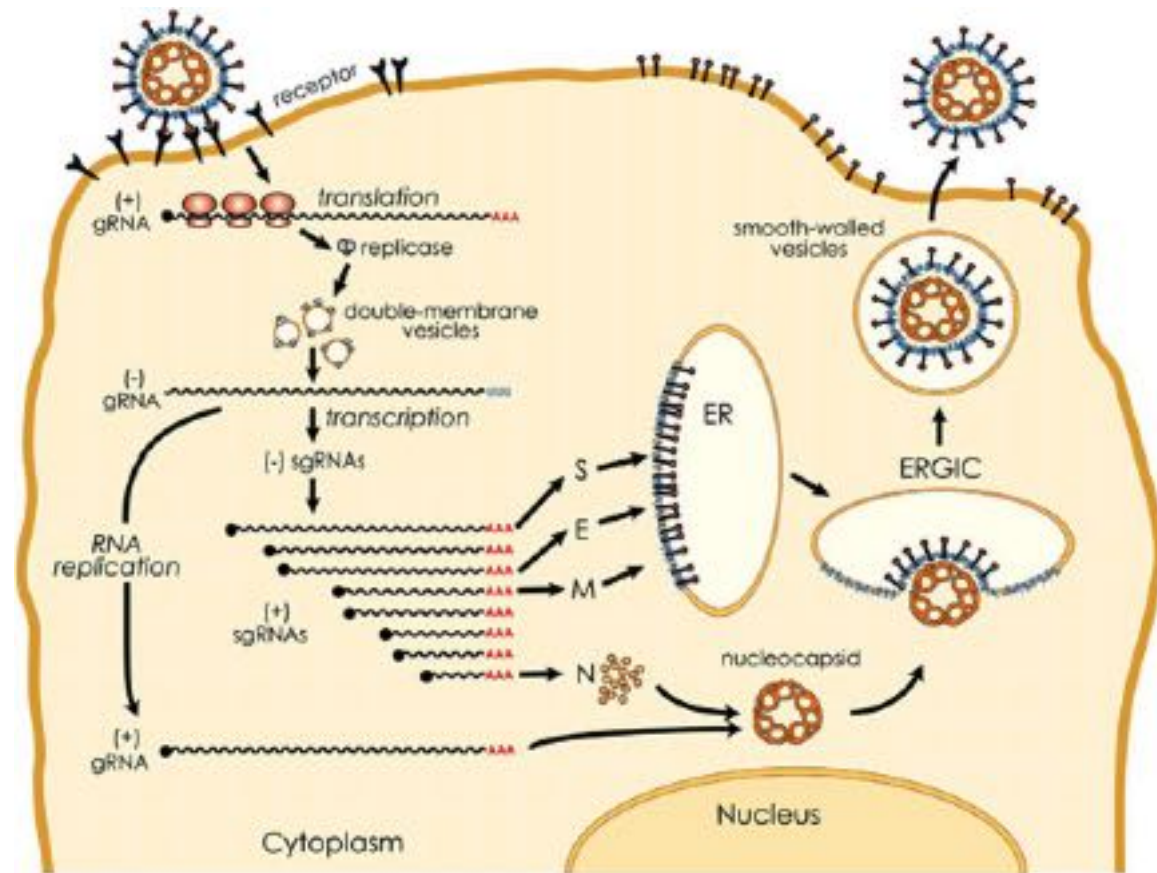


Fig. 4 Replication cycle of CoVs. The spike glycoprotein on the virus particle interacts with host cell receptors to mediate fusion of the virus and host cell membranes and release of the positive-strand RNA genome into the cytoplasm. The 5'-proximal open reading frames (ORF1a and ORF1b) are translated from the gRNA to generate the replicase polyprotein. The replicase polyprotein is processed by viral proteases into 16 nonstructural proteins which assemble with membranes to generate double-membrane vesicles (DMVs) where RNA synthesis takes place. A nested set of 3' co-terminal subgenomic (sg) RNAs is generated by a discontinuous transcription process. The sgRNAs are translated to generate the viral structural and accessory proteins. Viral gRNA is replicated and associates with nucleocapsid protein and viral structural proteins in the endoplasmic reticulum-Golgi intermediate compartment (ERGIC), where virus particles bud into vesicles before transport and release from the cell. Reprinted from Masters PS (2006) *The molecular biology of coronaviruses. Advances in Virus Research* 66: 193–292, with permission from Elsevier.

subsequent release from the cell. Depending on the virus strain, this replication can be robust and result in destruction of the host cell or a low-level, persistent infection that can be maintained in cultured cells or infected animals.

A hallmark of CoV transcription is the generation of a nested set of mRNAs, with each mRNA having the identical 'leader' sequence of approximately 65–90 nt at the 5' end (Fig. 5(a)). The leader sequence is encoded only once at the 5' end of the gRNA. Each subgenomic mRNA (sgmRNA) has the identical leader sequence fused to the 5' end of the body sequence. How are the leader-containing mRNAs generated during CoV transcription? Current evidence supports a model of discontinuous transcription, whereby the replicase complex switches templates during the synthesis of negative-strand RNA (Fig. 5(b)). The key sequence element in this process is the transcriptional regulatory sequence (TRS). The TRS is a sequence of approximately 6–9 nt (5'-ACGAAC-3' for SARS-CoV) which is found at the end of the leader sequence and at each intergenic region (the sites between the open reading frames encoding the viral structural and accessory proteins). Site-directed mutagenesis and deletion analysis has revealed the critical role of the TRS in mediating transcription of sgmRNAs. Deletion of any intergenic TRS results in loss of production of the corresponding sgmRNA. In addition, the CoV leader TRS and the intergenic TRS sequences must be identical for optimal production of the sgmRNAs. A three-step working model for template switching during negative-strand RNA synthesis has been proposed to describe the process for the generation of CoV leader-containing sgmRNAs (Fig. 5(b)). In this process, the 5' end and 3' end of the gRNA form a complex with host cell factors and the viral replication complex. The 3' end of the positive strand is used as the template for the initiation of transcription of negative-strand RNA. Negative-strand RNA synthesis continues up to the point of the TRS. At each TRS, the viral replicase may either read through the sequence to generate a longer template, or switch templates to copy the leader sequence. The template switch allows the generation of a leader-containing sgmRNA. In this model, alignment of the leader TRS, the newly synthesized negative-strand RNA, and the genomic TRS is critical for the template switching to occur. Disruption of the complex, or loss of base-pairing within the complex, will result in the loss of production of that

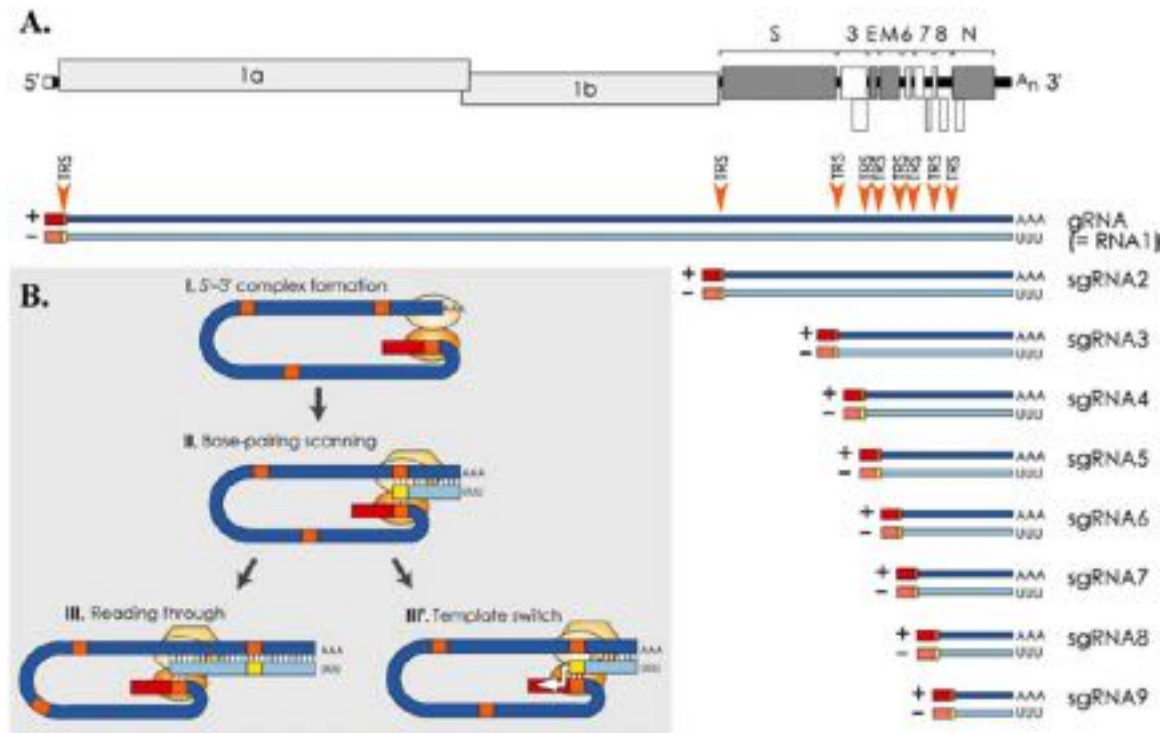


Fig. 5 Model of SARS-CoV gRNA and sgRNAs, and a working model of discontinuous transcription. (a) Diagram of gRNA and the nested set of sgRNAs of SARS-CoV. The 5' leader sequence, the transcriptional regulatory sequences (TRSs), and the positive- and negative-sense sgRNAs are indicated. (b) A working model of CoV discontinuous transcription. I. 5'-3' complex formation. Binding of viral and cellular proteins to the 5' and 3' ends of the CoV gRNA is represented by ellipsoids. The leader sequence is indicated in red, the TRS sites are in orange. II. Base-pair scanning step. Minus-strand RNA (light blue) is synthesized from the positive-strand template by the viral transcription complex (hexagon). At the TRS site, base-pairing may occur between the template, the nascent negative-strand RNA, and the leader TRS sequence (dotted lines). III. The synthesis of negative-strand RNA can continue to make a longer sgRNA III, or a template switch can take place III' to generate a leader-containing subgenomic negative-strand RNA, which could then serve as the template for leader-containing positive-strand sgRNAs. Modified from Enjuanes L, Almazán F, Sola I, and Zunia S (2006) Biochemical aspects of coronavirus replication: A virus–host interaction. *Annual Reviews in Microbiology* 60: 211–230.

sgmRNA. Further studies of the CoV replication complex may yield new insights into the role of the viral helicase and endoribonuclease in the generation of the leader-containing CoV RNAs.

Another hallmark of CoV replication is high-frequency RNA recombination. RNA recombination occurs when a partially synthesized viral RNA dissociates from one template and hybridizes to similar sequences present in a second template. Viral RNA synthesis continues and generates a progeny virus with sequences from two different parental genomes. This RNA recombination event is termed copy-choice recombination. Copy-choice RNA recombination can be demonstrated experimentally when two closely related CoV strains (such as MHV-JHM and MHV-A59) are used to coinfect cells. Recombinant viruses with cross-over sites throughout the genome can be isolated, although sequences within the spike glycoprotein may be a 'hot spot' for recombination due to the presence of RNA secondary structures that may promote dissociation and reassociation of RNA. It has been proposed that copy-choice recombination is also the mechanism by which many CoVs have acquired accessory genes, and it has been exploited experimentally for the deletion or insertion of specific sequences in CoV genomes to assess their role in virus replication and pathogenesis.

CoV Accessory Proteins

Sequence analysis of CoVs isolated from species ranging from birds to humans has revealed that all CoVs encode a core canonical set of genes, replicase (rep), spike (S), envelope (E), membrane (M), and nucleocapsid (N), and additional, so-called accessory genes (Table 1). The canonical genes are always found in the same order in the genome: rep-S-E-M-N. Reverse genetic studies (see below) have shown that this is the minimal set of genes required for efficient replication and assembly of infectious CoV particles. However, the genomes of all CoVs sequenced to date encode from one to eight additional ORFs, which code for accessory proteins. As the name implies, these accessory proteins are not required for CoV replication in tissue culture cell lines, but they may play important roles in tropism and pathogenesis *in vivo*. How were these additional genes acquired? Current evidence indicates that these additional sequences may have been acquired by RNA recombination events between co-infecting viruses. For example, the hemagglutinin-esterase (HE) glycoprotein present in four different CoVs (MHV, BCoV, HCoV-OC43, and HCoV-HKU-1) was

Table 1 Coronavirus **canonical** and *accessory* proteins

<i>Virus</i>	<i>Proteins: canonical (rep-S-E-M-N) and accessory</i>
<i>Alphacoronavirus</i>	
TGEV	rep-S-3a,3b-E-M-N-7
FIPV	rep-S-3a,3b,3c-E-M-N-7a,7b
HCoV-229E	rep-S-4a,4b-E-M-N
PEDV	rep-S-3-E-M-N
HCoV-NL63	rep-S-3-E-M-N
BatCoV-HKU8	rep-S-3-E-M-N-7
BatCoV-HKU2	rep-S-3-E-M-N-7
<i>Betacoronavirus</i>	
MHV	rep-2a, HE-S-4-5a,E-M-N,7b
BCoV	rep-2a, HE-S-4a,4b-5,E-M-N,7b
HCoV-OC43	rep-2a, HE-S-5,E-M-N,7b
HCoV-HKU1	rep-HE-S-4-E-M-N,7b
SARS-CoV	rep-S-3a,3b-E-M-6-7a,7b-8a,8b-N,9b
Bat-SARS-CoV	rep-S-3-E-M-6-7a,7b-8-N,9b
MERS-CoV	rep-S-3-4a,4b-5-E-M-N-8b
BatCoV-HKU5	rep-S-3a,3b,3c,3d-E-M-N
BatCoV-HKU4	rep-S-3a,3b,3c,3d-E-M-N
<i>Gammacoronavirus</i>	
Avian IBV	rep-S-3a,3b,3c-E-M-5a,5b-N
Beluga whale CoV	rep-S-E-M-5a,5b,5c-6-7-8-9-10-N
<i>Deltacoronavirus</i>	
Munia CoV	rep-S-E-M-5-N-7a,7b,7c
Thrush CoV	rep-S-E-M-5-N-7a,7b,7c

likely acquired by recombination of an ancestral CoV with the HE glycoprotein gene of influenza C. Interestingly, the expression of the HE gene has no effect on replication of the virus in cultured cell lines, but has been shown to enhance virulence in infected animals. Other CoV accessory genes may have been acquired through recombination with host cell mRNA or other viral mRNAs. The specific role of the accessory proteins in CoV replication and pathogenesis is under investigation. For SARS-CoV, accessory protein 6 has been implicated as an important factor in viral pathogenesis. Researchers have shown that mice infected with murine CoV expressing SARS-CoV protein 6 rapidly succumb to the infection, indicating that the protein 6 enhances virulence. In addition, studies suggest that SARS-CoV accessory proteins may play a role in blocking host cell innate immune responses, which may enhance viral replication and virulence. Other accessory proteins, such as SARS-CoV 3a and 7a, have been shown to be packaged into virus particles, where they may enhance infectivity or alter cell tropism. Future studies will be aimed at elucidating how CoV accessory proteins modulate the innate immune response to coronavirus replication.

Manipulating CoV Genomes Using RNA Recombination and Reverse Genetics

Genetic manipulation of CoV sequences is challenging because of the large size (27–32 kbp) of the RNA genomes. However, two approaches have been developed to allow researchers to introduce mutations, deletions, and reporter genes into CoV genomes. These approaches are (1) targeted RNA recombination and (2) reverse genetics using infectious cDNA constructs of CoV. The first approach exploits high-frequency copy-choice recombination to introduce mutations of interest into the 3' end of the CoV gRNA. In the first step of targeted RNA recombination, a cDNA clone encoding the region from the spike glycoprotein to the 3' end of the RNA is generated. These sequences can be easily manipulated in the laboratory to introduce mutations or deletions, or for the insertion of reporter or accessory genes, into the plasmid DNA. Next, RNA is transcribed from the plasmid DNA and the RNA is transfected into cells coinfecting with the CoV of interest. RNA recombination occurs between the replicating CoV and the transfected substrate RNA, and viruses with the 3' end sequences derived from the transfected substrate RNA will be generated. The recombinant viruses are generated by high-frequency copy-choice recombination, but the challenge is to sort or select for the recombinant virus of interest from the background of wild-type virus. To facilitate selection of recombinant viruses, Masters and Rottier introduced the idea of host range-based selection. They devised a clever plan to use a mouse hepatitis virus (MHV) that encodes the spike glycoprotein from a feline CoV as the target for their recombination experiments. This feline-MHV, termed fMHV, will infect only feline cell lines. Substrate RNAs that encode the MHV spike and mutations of interest in the 3' region of the genome can be transfected into feline cells infected with fMHV, and progeny virus can be collected from the supernatant and subsequently selected for the ability to infect murine cell lines. Recombinant CoVs that have incorporated the MHV spike gene sequence (and the downstream substrate RNA with mutations of interest) can be selected for growth on murine cells, thus allowing for the rapid isolation of the recombinant virus of interest. This host range-based selection step is now widely used by virologists to generate recombinant viruses with specific alterations in the 3' end of the CoV genome.

The second approach for manipulating CoV sequences, generating infectious cDNA constructs of CoV, has been developed in several laboratories. Full-length CoV sequences have been cloned and expressed using bacterial artificial chromosomes (BACs), vaccinia virus vectors, and from an assembled set of cDNA clones representing the entire CoV genome. The generation of a full-length cDNA and subsequent generation of a full-length CoV gRNA allows for reverse genetic analysis of CoV sequences. Successful reverse genetics systems are now in place to study the replication and pathogenesis of SARS-CoV, MERS-CoV, MHV, HCoV-229e, and IBV. These reverse genetics systems have allowed researchers to introduce mutations into the replicase gene and identify sites that are critical for enzymatic activities of many replicase products such as the helicase, endoribonuclease, and the papain-like proteases. Reverse genetic approaches are also being used to investigate the role of the TRSs in controlling the synthesis of CoV mRNAs. Interestingly, the SARS-CoV genome can be 're-wired' using a novel, noncanonical TRS sequence, which must be present at both the ends of the leader sequence and at each intergenic junction. This 're-wired' SARS-CoV may be useful for generating a live-attenuated SARS-CoV vaccine. An important feature of this 're-wired' virus is that it would be nonviable if it recombined with wild-type virus, since the leader TRS and downstream TRS would no longer match in a recombinant virus. The development of reverse genetics systems for CoVs has opened the door to investigate how replicase gene products function in the complex mechanism of CoV discontinuous transcription, and provides new opportunities to generate novel CoVs as potential live-attenuated or killed virus vaccines to reduce or prevent CoV infections in humans and animals.

Vaccines and Antiviral Drug Development

Because of the economic importance of CoV infection to livestock and domestic animals, a variety of live-attenuated and killed CoV vaccines have been tested in animals. Vaccines have been developed against IBV, TGEV, CCoV, and FIPV. However, these vaccines do not seem to provide complete protection from wild-type virus infection. In some cases, the wild-type CoV rapidly evolves to escape neutralization by vaccine-induced antibodies. In studies of vaccinated chickens, a live-attenuated IBV vaccine has been shown to undergo RNA recombination with wild-type virus to generate vaccine escape mutants. Killed virus vaccines may also be problematic for some CoV infections. Vaccination of cats with a killed FIPV vaccine has been shown to exacerbate disease when cats are challenged with wild-type virus. Therefore, extensive studies will be required to carefully evaluate candidate vaccines for SARS-CoV or MERS-CoV. A variety of approaches are currently under investigation for developing CoV vaccines, including analysis of killed virus vaccines, live-attenuated virus vaccines, and viral vector vaccines (such as modified vaccine virus Ankara, canarypox, alphavirus, and adenovirus vectors). Studies have shown that removing the envelope (E) protein from either SARS-CoV or MERS-CoV is an effective approach for generating a live attenuated virus vaccine. In the absence of this structural protein, the virus replicates to lower titers in the lungs of mice, but still induces an immune response that is protective from challenge with wild type virus. The development of improved animal models for human coronaviruses will be essential for evaluating any candidate vaccines. Transgenic mice expressing human ACE-2 (for SARS-CoV) can be used as a model system for SARS, however the pathogenesis in ACE-2 expressing transgenic mice does not fully mimic the pathogenesis seen during SARS-CoV infection in humans. CoVs can be adapted for replication in small animal models. Passaging SARS-CoV in mice led to the generation of a mouse-adapted strain, SARS CoV-MA15, which induces a lethal respiratory tract infection of mice. CoV vaccine studies will benefit from an improved understanding of conserved viral epitopes that can be targeted for vaccine development.

The use of neutralizing monoclonal antibodies directed against the SARS-CoV or MERS-CoV spike glycoprotein is another approach that may provide protection from severe disease. The success in the development and use of humanized monoclonal antibodies against respiratory syncytial virus (family *Paramyxoviridae*) to protect infants from severe disease indicates that this approach is certainly worth investigating. Studies have shown that patient convalescent serum and monoclonal antibodies directed against the SARS-CoV spike glycoprotein efficiently neutralize infectious virus. Further studies are essential to evaluate any concerns about potential antibody-mediated enhancement of disease and to determine if neutralization escape mutants arise rapidly after challenge with infectious virus. Studies evaluating monoclonal antibodies directed against a variety of structural proteins, and monoclonal antibodies directed against conserved sites in the spike glycoprotein will provide important information on the efficacy of passive immunity to protect against emerging coronaviruses.

Currently, there are no antiviral drugs approved for use against any human CoV infection. With the potential for the emergence or re-emergence of pathogenic CoV from animal reservoirs, there is considerable interest in identifying potential therapeutic targets and developing antiviral drugs that will block viral replication and reduce the severity of CoV infections in humans. Two promising targets for antiviral drug development are the SARS-CoV protease domains, the papain-like protease (PL^{Pro}) and the 3C-like protease (3CL^{Pro}, also termed the main protease, M^{Pro}) (Fig. 6). These two protease domains are encoded within the replicase polyprotein gene and protease activity is required to generate the 16 replicase nonstructural proteins (nsp1–nsp16) that assemble to generate the viral replication complex. The crystal structure of the 3CL^{Pro} was determined first from TGEV and then from SARS-CoV. Rational drug design, much of which was based on our knowledge of inhibitors directed against the rhinovirus 3C protease, has provided promising lead compounds for 3CL^{Pro} antiviral drug development. Interestingly, these candidate antivirals have been shown to inhibit the replication of SARS-CoV and other group 2 CoVs such as MHV, and the less related group 1 CoV, HCoV-229e. This indicates that the active site of 3CL^{Pro} is highly conserved among CoVs and that antiviral drugs developed against SARS-CoV 3CL^{Pro} may also be useful for inhibiting the replication of more common human CoVs such as HCoV-229e, HCoV-OC43, HCoV-NL63, and HCoV-HKU1. Further studies are needed to determine if these inhibitors can be developed into clinically useful antiviral agents.

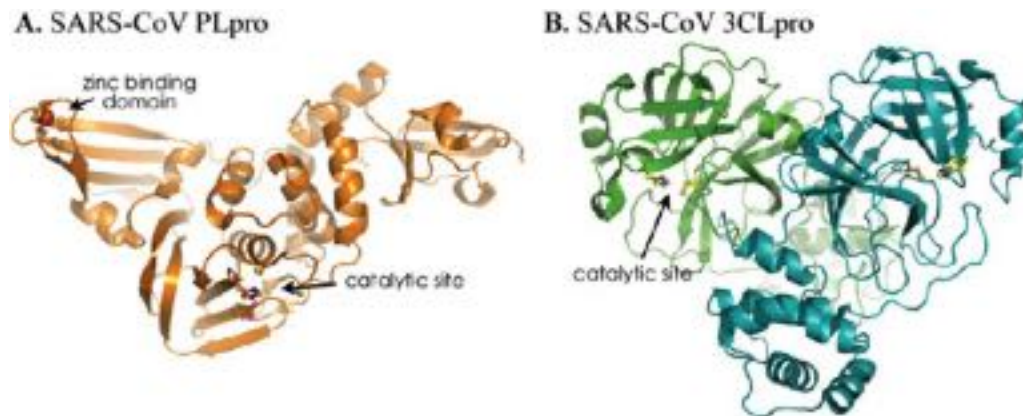


Fig. 6 CoV proteases are targets for antiviral drug development; X-ray structures of the two SARS-CoV protease domains encoded in the replicase polyprotein. (a) The SARS-CoV papain-like protease (PL^{pro}) with catalytic triad cysteine, histidine, and aspartic acid residues, and zinc-binding domain indicated. (b) The 3C-like protease (3CL^{pro}, also termed main protease, M^{pro}) dimer with catalytic cysteine and histidine residues indicated.

Analysis of SARS-CoV papain-like protease led to the surprising discovery that this protease is also a viral deubiquitinating (DUB) enzyme. The SARS-CoV PL^{pro} was shown to be required for processing the amino-terminal end of the replicase polyprotein and to recognize conserved cleavage site (-LXGG). The LXGG cleavage site is also the site recognized by cellular DUBs to remove poly-ubiquitin chains from proteins targeted for degradation by proteasomes. Analysis of the X-ray structure of the SARS-CoV PL^{pro} has revealed that it has structural similarity to known cellular DUBs. These studies suggest that CoV papain-like proteases have evolved to have both proteolytic processing and DUB activity. The DUB activity may be important in preventing ubiquitin-mediated degradation of viral proteins, or the DUB activity may be important in subverting host cell pathways to enhance viral replication. PL^{pro} inhibitors are now being developed using structural information and by performing high-throughput screening of small molecule libraries to identify lead compounds. The recent development of a chimeric virus system which uses Sindbis virus to express the SARS-CoV or MERS-CoV papain-like protease, facilitates the study of protease activity and the efficacy of protease inhibitors. The chimeric virus system can be used in a biosafety level 2 environment, compared to the biosafety level 3 containment required for the study of SARS-CoV or MERS-CoV. Additional CoV replicase proteins, particularly the RNA-dependent RNA polymerase, helicase, and endoribonuclease, are also being targeted for antiviral drug development.

Future Perspectives

The development of targeted RNA recombination and reverse genetics systems for CoVs has provided new opportunities to address important questions concerning the mechanisms of CoV replication and virulence, and to design novel CoV vaccines. In the future, improved small animal models for testing vaccines and antivirals, and the availability of additional X-ray crystallographic structure information for rational drug design will be critical for further progress toward development of effective vaccines and antiviral drugs that can prevent or reduce diseases caused by CoVs.

See also: Coronaviruses: General Features (*Coronaviridae*). Enveloped, Positive-Strand RNA Viruses (*Nidovirales*). Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (*Coronaviridae*). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (*Coronaviridae*)

Further Reading

- Almazán, F., DeDiego, M.L., Sola, I., *et al.*, 2013. Engineering a replication-competent, propagation-defective Middle East respiratory syndrome coronavirus as a vaccine candidate. *MBio* 4, e00650. 13.
- Deng, X., Agnihortham, S., Mielech, A.M., *et al.*, 2014. A chimeric virus-mouse model system for evaluating the function and inhibition of papain-like proteases of emerging coronaviruses. *Journal of Virology* 88 (20), 11825–11833.
- Enjuanes, L., Almazán, F., Sola, I., Zuñiga, S., 2006. Biochemical aspects of coronavirus replication: A virus–host interaction. *Annual Reviews in Microbiology* 60, 211–230.
- Ge, X.Y., Li, J.L., Yang, X.L., *et al.*, 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503, 535–538.
- Graham, R.L., Donaldson, E.F., Baric, R.S., 2013. A decade after SARS: Strategies for controlling emerging coronaviruses. *Nature Reviews. Microbiology* 11, 836–848.
- Haagmans, B.L., Al Dhahiry, S.H.S., Reusken, C.B.E.M., *et al.*, 2014. Middle east respiratory syndrome coronavirus in dromedary camels: An outbreak investigation. *The Lancet Infectious Diseases* 14, 140–145.
- Jiang, L., Wang, N., Zuo, T., 2014. Potent neutralization of MERS-CoV by human neutralizing monoclonal antibodies to the viral spike glycoprotein. *Science Translational Medicine* 6, 234ra59.
- Lau, Y.L., Peiris, J.S., 2005. Pathogenesis of severe acute respiratory syndrome. *Current Opinion in Immunology* 17, 404–410.

- Li, W., Wong, S.K., Li, F., *et al.*, 2006. Animal origins of the severe acute respiratory syndrome coronavirus: Insights from ACE2-S-protein interactions. *Journal of Virology* 80, 4211–4219.
- Lu, G., Hu, Y., Qi, J., *et al.*, 2013. Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 500, 227–232.
- Masters, P.S., Rottier, P.J.M., 2005. Coronavirus reverse genetics by targeted RNA recombination. *Current Topics in Microbiology and Immunology* 287, 133–160.
- Perlman, S., Netland, J., 2009. Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews. Microbiology* 7, 439–450.
- Raj, V.S., *et al.*, 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 495, 251–254.
- Ratia, K., Kilianski, A., Baez-Santos, Y.M., *et al.*, 2014. Structural basis for the ubiquitin-linkage specificity and deISGylating activity of SARS-CoV papain-like protease. *PLoS Pathogens*. <http://dx.doi.org/10.1371/journal.ppat.1004113>.
- Ratia, K., Pegan, S., Takayama, J., *et al.*, 2008. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proceedings of the National Academy of Sciences, USA* 105, 16119–16124.
- Roberts, A., *et al.*, 2007. A mouse-adapted SARS-coronavirus causes disease and mortality in BALB/c mice. *PLoS Pathogens* 3, e5.
- Scobey, T., *et al.*, 2013. Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *Proceedings of the National Academy of Sciences, USA* 110, 16157–16162.
- Thiel, V., Siddell, S., 2005. Reverse genetics of coronaviruses using vaccinia virus vectors. *Current Topics in Microbiology and Immunology* 287, 199–228.
- Woo, P.C., *et al.*, 2012. Discovery of seven novel mammalian and avian coronaviruses in the genus *Deltacoronavirus* supports bat coronaviruses as the gene source of *Alphacoronavirus* and *Betacoronavirus* and avian coronaviruses as the gene source of *Gammacoronavirus* and *Deltacoronavirus*. *Journal of Virology* 86, 3995–4008.
- Yang, H., Xie, W., Xue, X., *et al.*, 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biology* 3, 1742–1751.
- Yount, B., Roberts, R.S., Lindesmith, L., Baric, R.S., 2006. Rewiring the severe acute respiratory syndrome coronavirus (SARS-CoV) transcription circuit: Engineering a recombination-resistant genome. *Proceedings of the National Academy of Sciences, USA* 103, 12546–12551.
- Zaki, A.M., van Boheemen, S., Bestebroer, T.M., *et al.*, 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine* 367, 1814–1820.

Relevant Websites

<http://www.cdc.gov/coronavirus>

Severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) resource, Centers for Disease Control and Prevention, USA.

<http://www.viprbrc.org>

viral genome bioinformatics resource.

<http://www.who.int/csr/disease/coronavirus>

World Health Organization Global Alert and Response for emerging coronaviruses.

Crimean-Congo Hemorrhagic Fever Virus and Nairoviruses of Medical Importance (*Nairoviridae*)

Ali Mirazimi, National Veterinary Institute, Uppsala, Sweden and Karolinska Hospital University, Huddinge, Sweden

Felicity Burt, University of the Free State, Bloemfontein, South Africa

Anna Papa, Aristotle University of Thessaloniki, Thessaloniki, Greece

© 2021 Elsevier Ltd. All rights reserved.

This is an update of C.A. Whitehouse, Crimean–Congo Hemorrhagic Fever Virus and Other Nairoviruses, in Encyclopedia of Virology (Third Edition), Edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00620-8.

Introduction

CCHFV is a tick-borne viral zoonosis found in Africa, Asia, eastern Europe and the Balkans. The geographic distribution of the virus correlates with that of ticks belonging to the genus *Hyalomma*, which are considered to be the principal vectors of the virus. Crimean-Congo hemorrhagic fever orthonairovirus (CCHFV) is a member of the *Orthonairovirus* genus of the family *Nairoviridae* within the *Bunyavirales* order of viruses. Species within the genus were grouped originally on the basis of antigenic relationships and morphological and genetic similarities. Recent taxonomic changes based on genomic characterizations, resulted in renaming of seven species within the genus and the addition of five new species to the genus. Taxonomic changes included revising the grouping of Hazara virus, historically considered to be the closest relative to CCHFV, and now placed in a separate species with Tofla virus. Neither of these viruses, Hazara and Tofla, are known to cause disease. Within the genus, CCHFV, the etiologic agent of mild to severe disease with fatalities, is considered to be the most significant public health concern. Nairobi sheep disease virus (NSDV), an important veterinary pathogen, causes severe disease in domestic sheep and goats in Africa and Ganjam virus, considered an isolate of NSDV, circulates in the Indian subcontinent. Lesser known human pathogenic orthonairoviruses include Dugbe virus and Kasokera virus, associated with febrile illness in Africa; Erve virus proposed to be the cause of thunderclap headache in Europe; and Issyk-Kul virus associated with febrile disease in central Asia. The remaining viruses within the *Orthonairovirus* genus have mostly been isolated from ticks, and the medical and/or veterinary significance of them is unknown.

A hemorrhagic disease with symptoms suggestive of CCHF infection was described in eastern Europe and Asia as far back as the 12th century. The disease given the name Crimean hemorrhagic fever (CHF) was first reported in an outbreak involving 200 soldiers and peasants on the Crimean peninsula in 1944. The following year the disease was shown to be caused by a filterable agent, but was only isolated in mice in 1967. In 1956 a virus was isolated from blood samples collected from a febrile child in Stanleyville in the Belgian Congo (now referred to as Kisangani in the Democratic Republic of the Congo) and given the name Congo virus. The availability of a laboratory host made it possible to further characterise the two viruses and in 1969 it was reported that the virus from Europe and Asian Republics of USSR and Bulgaria were antigenically identical to Congo virus from Africa. The names were combined and the virus is now referred to as Crimean-Congo hemorrhagic fever virus.

Due to the high fatality rate, the propensity to cause nosocomial infections and a lack of specific anti-viral treatment or vaccine, the virus is classified among biosafety level 4 pathogens. This dictates that the virus can only be handled and cultured within the confines of a maximum containment facility. This requirement for a biosafety level 4 laboratory limits the number of facilities that culture the virus and has likely contributed to our incomplete understanding on the pathogenesis of disease and the correlates of protection. Development of vaccines was initially thwarted by the lack of suitable animal models however the identification of murine models, and more recently a non-human primate model, has contributed to recent development of candidate vaccines.

The emergence and re-emergence of CCHFV in Balkan countries and south western regions of the Russian Federation, and more recently in southern regions of Europe, highlights the potential for this pathogen to spread to non-endemic regions where ticks of the *Hyalomma* genus are present.

Virion Structure, Genome and Life Cycle

The CCHF virion is spherical and approximately 80–100 nm in diameter with a bilayered lipid envelope. The viral glycoproteins G_N and G_C protruding from the lipid envelop are likely associated with receptor recognition and are involved in the early steps of viral replication. The CCHFV contain the negative-sense small (S), medium (M) and large (L) genome segments, encapsidated by the nucleoprotein (N), and the RNA dependent RNA polymerase (RdRp), which are responsible to initiate transcription and genome replication in the host cell. The terminal complementary sequences 5'-UCUCAAGA and 3'-AGAGUUUCU are conserved in all nairoviridae members.

The S Segment

The S segment encodes the N, which is the most abundant viral protein in the CCHFV particles, playing a key role in the encapsidation of viral genomic. The CCHFV N structure organize a large globular domain, to which N- and C-terminal portions of the polypeptide

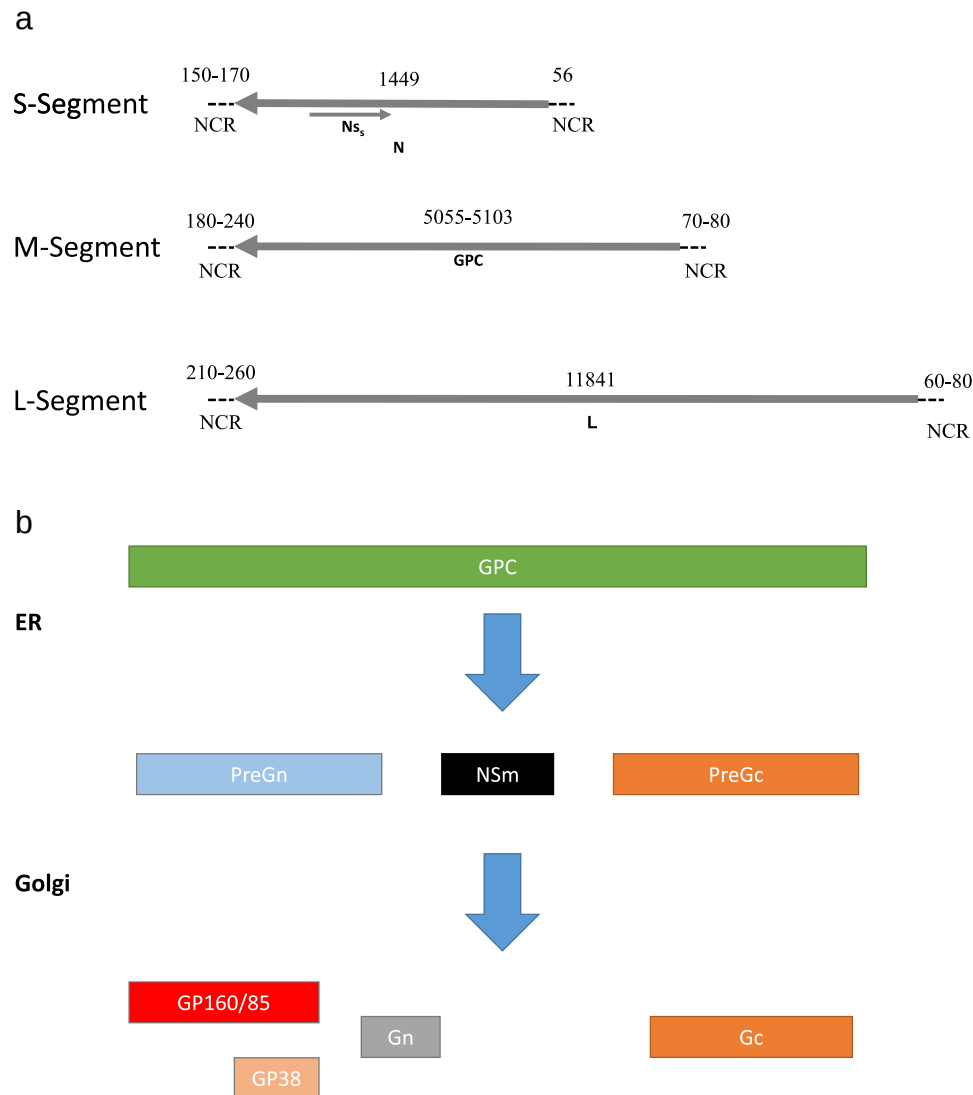


Fig. 1 CCHF virus genome. CCHF virus consist of tri-segmented negative sense RNA genome. The small (~1.6 kb), medium (~5.4 kb) and large (~12.1 kb) segments, code for the N, the GPC and the L protein, respectively. The small segment also codes for a non-structural S protein (NSS) in opposite sense. The coding regions are flanked by non-coding regions (NCRs).

contribute, plus a protruding “arm” containing a conserved caspase-3 cleavage site on the top of the arm. Most probably the globular region is responsible for RNA binding. The S segment can also codes for a non-structural protein in ambisense direction (**Fig. 1**).

The M Segment

The M segment encodes a single polyprotein, the glycoprotein precursor (GPC) (**Fig. 1**). The maturation of CCHFV GPC is very complex and does not share much similarity with glycoprotein processing in other bunyaviruses. GPC maturation yields to two type I transmembrane glycoproteins, G_N and G_C which are spiked on the virus envelope. The GPC processing in the endoplasmic reticulum (ER) and Golgi results in two structural glycoproteins, G_N and G_C , and nonstructural proteins designated GP160, GP85, and GP38 [and the non-structural M protein (NSM). The GPC is heavily glycosylated and subsequently cleaved by host proteases including the proprotein convertases.

The L Segment

Nairoviridae have an unusually large RNA dependent RNA polymerase protein (nearly twice the size) compared to those of other members (**Fig. 1**). This segment has a single open reading frame, and encodes an approximately 4000-amino acid polyprotein. The

amino acids 1–609 contain domains with the ovarian tumor (OTU) cysteine protease function. Sequences located between the OTU domain and the RdRp conserved motifs contain a potential leucine zipper and a C2H2 zinc finger motif.

Virus Life Cycle

The CCHFV GN and/or GC are involved in early step of replication cycle of CCHFV. It has been suggested that GC is responsible for binding to the cellular receptors and mediates fusion in endosomes (Fig. 2). The processes of internalization, virion assembly

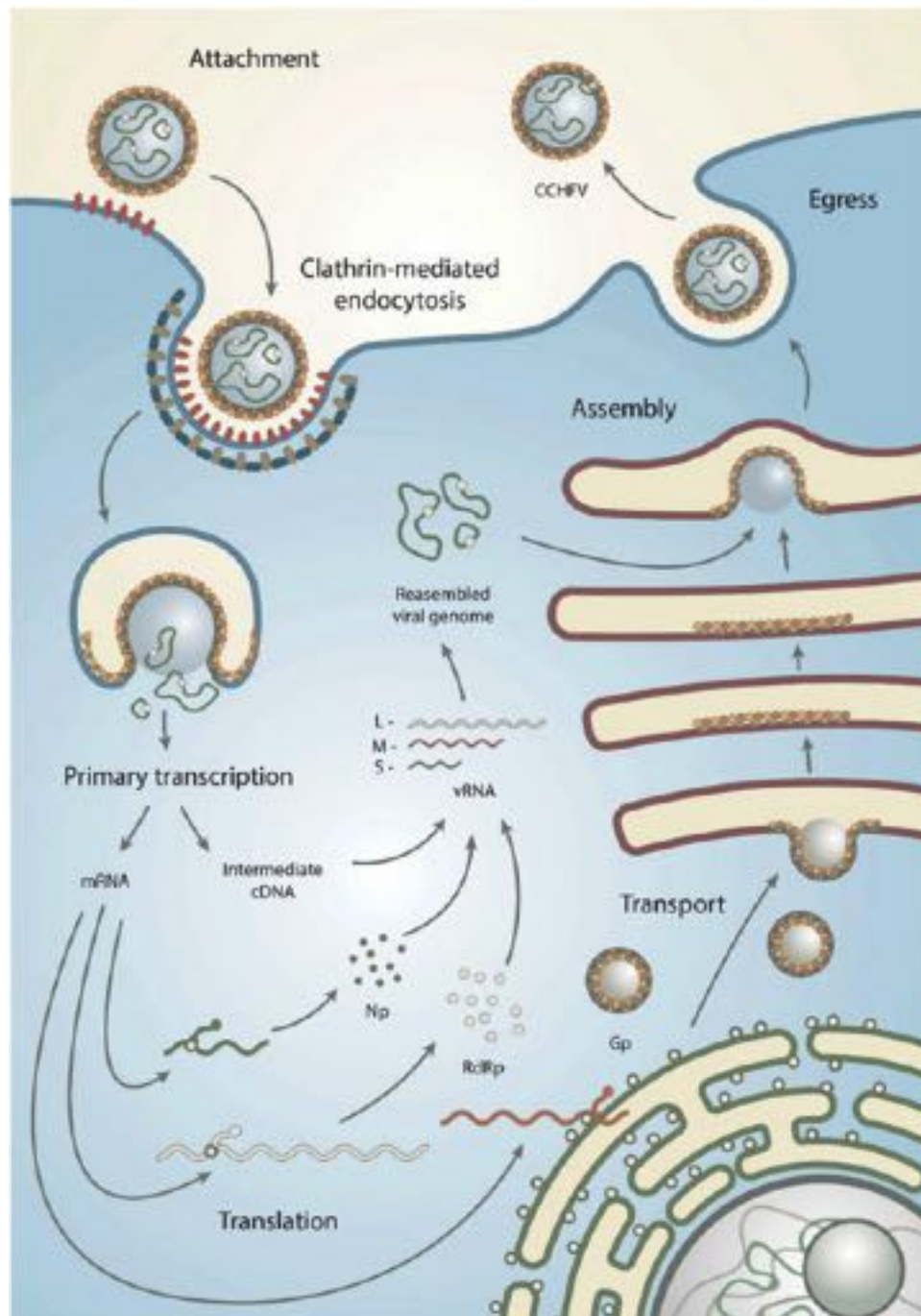


Fig. 2 CCHFV replication cycle. CCHFV binds to an unknown cellular receptor and penetrate into the cells in a clathrin-dependent manner. After fusion, the viral genome is released to the cytoplasm. The viral mRNAs are translated into the NP and L proteins at cytoplasm, and the glycoprotein precursor (GPC) translated into endoplasmic reticulum (ER). The NP and L protein helping to replicate the genomic RNA. The GPC is processed and matured to Gn and Gc in the ER and the Golgi. The new CCHFV particles most probably assemble in the Golgi followed by virus release in Golgi-derived vesicles. The figure illustrated by Dr. J Vermeulen.

and egress are dependent on host-cell skeleton. The cellular receptors for CCHFV entry have not been identified. A functional interaction has been suggested between CCHFV GC and cell surface nucleolin. Following adsorption of CCHFV at plasma membrane, the CCHFV is endocytosed through a clathrin-mediated endocytosis mechanism. CCHFV entry has also been suggested to be dependent on cholesterol and a low pH. Following endocytosis, CCHFV particles are transported to early endosomes and subsequently to multivesicular bodies (MVB), where the CCHFV membrane is fused to the endosomes membrane. Following this process the genomic RNPs are released into the cytosol and transcription and replication process will be initiated. The replication of the RNPs minimally require the L protein and N. Since CCHFV is a negative strand RNA virus, genomic RNPs are used as template to synthesize capped mRNA. These process are followed by transcription and viral protein processing and subsequent maturation. The assembly and egress process for CCHFV is most probably like other bunyaviruses.

Epidemiology

CCHF is the most geographically extensive viral hemorrhagic fever and the most widespread medically significant tick-borne viral disease. It is estimated that 3 billion people in Africa, Asia and Europe are at risk, and 10,000–15,000 infections and 500 fatalities occur annually. Ticks of the *Hyalomma* genus (mainly *H. marginatum*, *H. rufipes*, *H. anatolicum*, and *H. asiaticum*) are the principal competent vectors and reservoirs of CCHFV. The occurrence of the disease coincides with the geographic distribution of *Hyalomma* spp. ticks, with a 50° north latitude limit, which, however, is predicted to increase to northern areas due to warming climate. Although CCHFV has been detected or isolated from ticks in several other genera, their role in CCHF epidemiology remains unclear.

CCHFV circulates unnoticed in nature between ixodid ticks and vertebrate hosts. The virus is maintained in ticks transovarially, transovarially, venereally and by co-feeding on infected or uninfected animals. Ticks remain persistently infected. The immature *Hyalomma* spp. ticks (larvae and nymphs) feed mainly on ground-feeding birds and small mammals and remain attached for up to 26 days. Adult ticks actively seek larger mammals, including hares, wild and domesticated ungulates, or humans. Overwintering infected ticks maintain the virus in nature. Transmission to humans occurs by bite or crushing of an infected tick or by direct contact (percutaneous and permucosal) with blood or tissues from viremic patients or livestock. Therefore, groups at higher risk are agricultural and abattoir workers, individuals involved in backyard slaughtering in endemic areas and health-care personnel. Nosocomial infections have been reported when exposure to the virus was substantial; the risk is higher when the patient presents hemorrhagic manifestations. Isolation of the patient and barrier nursing should be applied. Epidemiological and behavioral factors that play a role in acquisition of the CCHFV infection differ among countries; e.g., while tick bite is the main mode of transmission in Turkey, the contact with the blood of infected livestock, seems to be the main route of CCHFV transmission to humans in Iran.

Cases are seen from spring to autumn and they are usually sporadic. Non-nosocomial epidemics or outbreaks occur when environmental (e.g., mild winters, dry summers) and anthropogenic conditions (e.g., land fragmentation, socioeconomic changes) result in high tick populations leading to increased human exposure to ticks. Several biotic and abiotic drivers contribute to shape a favorable ecosystem. Given that CCHFV can be spread in long distances by migratory birds infested with infected ticks, or transported through international livestock trade, virus introduction in a suitable new habitat can lead to disease emergence and establishment. This was seen in Turkey, India and Spain where the disease emerged in 2002, 2011 and 2016, respectively. Based on environmental suitability, predictive risk maps have been constructed which can direct surveillance studies.

The epidemiology of CCHF is changing, as the disease emerged in new places, while the incidence differs between years. Currently, the most CCHF-affected country is Turkey; since 2002, when CCHF emerged in the country, more than 10,000 cases have been registered, with the majority of them being observed in 2008 and 2009 (more than 1000 cases annually). The most endemic areas are in the Middle Anatolia-Black Sea region. Sporadic CCHF human cases and outbreaks have been reported in more than 30 countries in three continents. In Asia, CCHF has been reported from Afghanistan, China, India, Iran, Iraq, Kazakhstan, Kuwait, Oman, Pakistan, Saudi Arabia, Tajikistan, Turkey, Turkmenistan, United Arab Emirates, and Uzbekistan. In Europe, CCHF has been reported in Albania, Bulgaria, Georgia, Kosovo, Russia, Ukraine and Spain, while one single case has been reported in Greece. Among African countries, CCHF has been reported in Burkina Faso, Democratic Republic of Congo, Egypt, Kenya, Mauritania, Nigeria, Senegal, Somalia, Republic of South Africa, Namibia, Sudan, Tanzania and Uganda. Furthermore, detection of the virus in ticks or detection of viral antibodies in humans and/or livestock, in the absence of clinical case, has been reported in several countries. Wild and domestic animals infected with CCHFV develop only a transient viremia without presenting symptoms of the disease. Seroprevalence studies in humans, livestock and wild animals enable the identification of CCHFV endemic foci.

CCHFV presents the most extensive genetic diversity among arboviruses. The highest level of diversity is seen in the M RNA segment, and especially in the mucin-like region. This phenomenon can be explained by the fact that the virus has to adapt to a variety of host (tick and vertebrate) cells, prompting for genetic changes and complex virus evolution. Based on S RNA segment sequences, seven genetic clades (three from Africa, two from Asia and two from Europe) can be identified which present a geographic correlation (Fig. 3).

Phylogenetic studies show that the Greek strain AP92 (clade Europe 2) and West African viruses (clade Africa 1) are the oldest CCHFV lineages. The AP92 strain was initially isolated from *Rhipicephalus bursa* tick in 1975 in Greece, and only a limited number of human cases has been associated with this lineage. Several studies showed that this lineage is strongly associated with

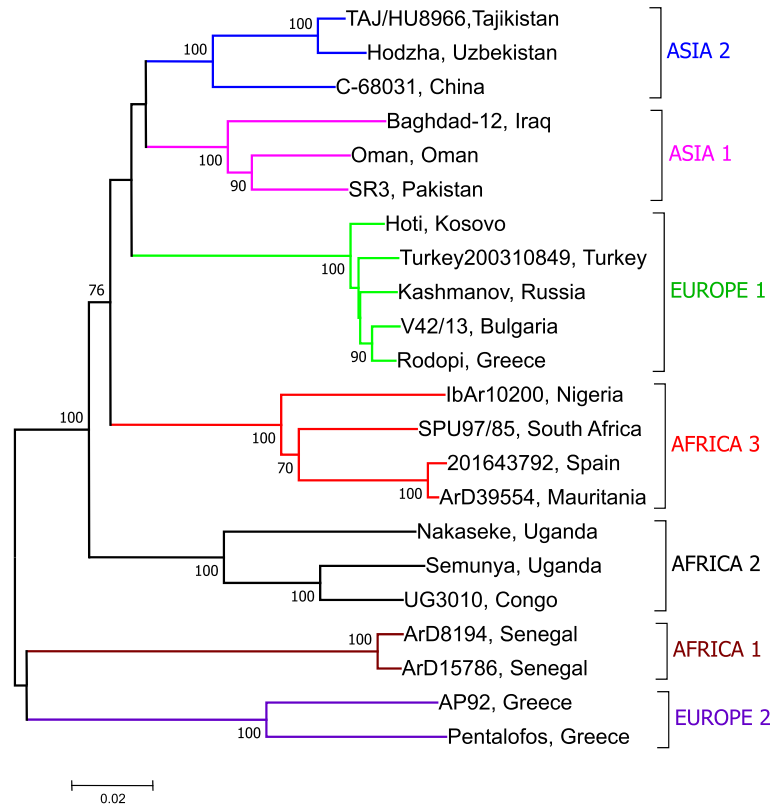


Fig. 3 Neighbor-Joining phylogenetic tree based on whole open reading frame of CCHFV S RNA segment (1446 nucleotides) computed using the Kimura 2-parameter method. Evolutionary analysis was conducted in MEGA software version 7.0.

Rhipicephalus spp. ticks. Genetic recombination is rare, but reassortment events are common in CCHFV, especially in the M RNA segment which encodes the envelope glycoproteins, Gn and Gc, involved in the attachment of the virus to the host cells. Detection of genetic clades in unrelated geographic regions is common, and can be explained by the passive transportation of the virus to distant regions across continents through migratory birds and livestock trade. The increasing number of whole genome CCHFV sequences is expected to elucidate further the molecular epidemiology of CCHF and increase the knowledge on the evolutionary history of the virus.

Clinical Features

Humans are the only vertebrates which present symptoms after CCHFV infection (Fig. 4),

CCHFV-infected humans may present asymptomatic, mild, severe, or even fatal disease. It is estimated that more than 80% of CCHFV infections are asymptomatic or subclinical and remain unnoticed. The case fatality rate (CRF) in hospitalized patients is 10%–40%. The course and outcome of the disease depend on the inoculation viral dose, the virus strain and the immune status and response of the host. In severe and fatal cases the production of antibodies is usually low, delayed or absent. CCHF resembles Ebola virus disease in clinical symptoms, underlying mechanisms and cytokine production.

Following initial virus amplification at the site of inoculation (dermal dendritic and Langerhans cells), the virus is spread through the lymphatic system to the target organs, mainly liver and spleen, and then is spread to additional organs. Endothelial damage leading to increased permeability and impaired immune response seem to play an important role in the clinical course and severity of the disease.

Typically, the clinical course follows four distinct phases: incubation, pre-hemorrhagic, hemorrhagic and convalescence. The incubation period is 1–14 days and varies depending on the viral dose and the route of exposure (it is usually shorter after a bloodstream infection, e.g., tick bite or needle-stick injury). The pre-hemorrhagic phase (1–5 days) is characterized by abrupt onset of fever lasting for 4–5 days, chills, severe headache, photophobia, conjunctivitis, fatigue, dizziness, myalgia, flushed face and rash, as well as gastrointestinal symptoms, such as nausea, vomiting, abdominal pain, and diarrhea. Hepatomegaly, jaundice, splenomegaly, pleural effusions and cardiovascular symptoms, such as bradycardia or tachycardia and hypotension, can be also seen. Some patients present changes in mood, confusion and aggression, followed by depression and somnolence. Patients with severe disease enter the hemorrhagic phase (3–4 days) and develop hemorrhagic manifestations on the mucous membranes and skin ranging from petechiae, epistaxis, bleeding at venipuncture and injection sites, ecchymosis, hematomas, and gingival hemorrhage



Fig. 4 Large hematomas in a Crimean-Congo hemorrhagic fever patient (day 10 of illness), Greece, 5 June 2018.

to severe hemorrhages from various systems (melena, hemoptysis, hematemesis, hematuria, nose and vaginal bleeding, and cerebral hemorrhage). Hemophagocytosis is often detected in the bone marrow biopsy. Necropsy findings in a fatal case in Spain revealed massive hepatocyte necrosis, complete apoptosis of colonic enterocytes, hemorrhages in bone marrow without inflammatory infiltrates in any of the organs examined providing support for the view that unlike other viral hemorrhagic fevers, there is a primary cytopathic pathogenic effect in CCHF. Death occurs 5–14 days after symptom onset. The survivors show improvement signs on the 10th day onwards, but recovery may be delayed by one month or even longer. Some persisting symptoms include weakness, lethargy, dizziness, hearing loss, and amnesia. The hospitalization time is approximately 10 days. The clinical course of the disease is usually milder in children.

The most frequently reported abnormal laboratory findings include thrombocytopenia, prolonged coagulation (prothrombin and activated partial thromboplastin) times, decreased fibrinogen, leucopenia, elevated levels of aminotransaminases, creatine phosphokinase, and lactate dehydrogenase. Collecting, handling and processing clinical samples require trained personnel and suitably equipped laboratories.

Differential diagnosis of CCHF includes other viral hemorrhagic fevers, rickettsiosis and other tick-borne bacterial infections, leptospirosis, viral hepatitis, brucellosis and bacterial sepsis, while non-infectious diseases may also cause similar symptomatology. Additional diseases have to be included when symptoms from respiratory, gastrointestinal or central nervous system are present. The list of diseases depends highly on the geographic location of the case and the travel history of the patient. The differential diagnosis is difficult when the patient presents only a mild flu-like illness.

Disease severity scoring systems have been developed. High viral load, severe thrombocytopenia, elevated liver enzymes and prolonged bleeding times are the most important predictors of severity and fatality in CCHF. Viral loads higher than 1×10^9 RNA copies/mL of plasma were strongly associated with fatality in CCHF patients. Specific HLA alleles and genetic polymorphisms in Toll-like receptors, as well as elevation of specific cytokines (e.g., IFN-inducible protein 10 and monocyte chemoattractant protein-1) and soluble adhesion molecules, have been also associated with severity and outcome of the disease.

Pathogenesis

CCHFV infection in human can result in a severe hemorrhagic fever. Interestingly, serological evidence indicates that CCHFV can infect numerous vertebrate species. However, even within humans, CCHFV infection may result in a mild or sub-clinical infection.

The factors contributing to development of a severe disease are unknown. The incubation period ranges from a few days up to 1 week, the length most probably depending on the transmission route and the amount of inoculum. The severity of disease has been demonstrated to correlate to the amount of virus in the blood. Furthermore, it has also been shown that antibody responses have a highly significant inverse correlation with viral loads. The pathogenesis of CCHF is only poorly characterized due to several reasons such as (1) infections occur sporadically and in areas where facilities are limited for performing complete autopsies, (2) virus handling requires biosafety level 4 (BSL-4) containment laboratories, and (3) a lack of available animal models of disease. The limited knowledge about CCHF pathogenesis is mostly derived from blood analyzes and liver biopsies of patients, using materials e.g., from outbreaks in Turkey and South Africa. One of the most comprehensive study involving 50 CCHF patients from South Africa, which describe that cerebral hemorrhage, severe anemia, severe dehydration, and shock associated with prolonged diarrhea, lung edema, and pleural effusion are the factors causing fatal outcome. Almost all patients who died developed multiple organ failure. In the fatal cases, platelet counts can be extremely low from an early stage of illness. Increases in aspartate and alanine aminotransferase (AST and ALT) levels in the serum, prolongation of prothrombin and partial thromboplastin times have also been observed.

A common suggested feature of agents causing hemorrhagic fever is the deregulation of host immune responses by combating and attacking cells involved in initiation of antiviral responses. The primary pathophysiological events appear to be leakage of erythrocytes and plasma through the vasculature into tissues. Endothelial damage can contribute to coagulopathy by deregulated stimulation of platelet aggregation, which in turn activates the intrinsic coagulation cascade, ultimately leading to clotting factor deficiency causing hemorrhages. For CCHFV, vascular leakage may be caused either by destruction of endothelial cells or by a disruption of the tight junctions which constitute the endothelial barrier between cells. Moreover, it is unclear whether these events are a direct consequence of infection or whether virus-induced host factors cause the endothelial dysfunction. However, in an epithelial cell line model CCHFV neither caused disruption of tight junctions nor necrosis or apoptosis of cells. This could suggest that the hemorrhages and coagulation disturbances may be caused indirectly, possibly by high levels of proinflammatory cytokines. Previous studies demonstrated higher level of IL-6 and TNF- α in CCHFV patients with fatal outcome compared to the non-fatal cases [38].

The ability of CCHFV to cause severe or lethal disease in mice deficient in the type I interferon system but not wild-type (WT) mice suggests that innate immune responses in vertebrate hosts play a substantial role in limiting CCHFV pathogenesis. Understanding the role of the adaptive immune response in the control of primary CCHFV infection has been limited by the lack of suitable animal models. However, a newly described humanized mouse model of CCHFV demonstrated that T cells were activated following CCHFV infection and may have a role in controlling the disease. In a recent study using mice treated with an interferon blockade antibody, demonstrating that adaptive immune responses can control CCHFV in mice. Recently a cynomolgus macaque model of CCHF for CCHFV has been described, in which cynomolgus macaques infected with a human clinical isolate of CCHFV, strain Hoti. These infected macaques develop a spectrum of disease outcome similar to that of human CCHF cases from asymptomatic to severe or lethal infections. Histological analysis showed that CCHFV infection resulted mainly in pathological changes in the liver and spleen.

In vitro studies have suggested retinoic acid-inducible gene I (RIG-I) as an innate immune sensor of CCHFV. However, it is most likely that more innate immune sensors contribute to sensing of CCHFV. The L segment of CCHFV encodes an ovarian tumor-like de-ubiquitinase (OTU) domain that recently has been shown to suppress innate immune responses. Several studies indicated that host apoptosis process may also be involved in the pathogenesis and controlling of CCHFV infection. There are a few studies in humans have identified correlates between polymorphisms in Toll-like receptors (TLRs), nuclear factor-kappa B and severity of disease. CCHFV also antagonizes innate immune signaling.

Diagnosis

Confirmation of infection early in the course of disease is important for isolation of patients to protect healthcare workers and for implementing appropriate supportive therapy. Although human cases can be identified using clinical criteria, laboratory confirmation is considered essential to distinguish the infection from other conditions that share similar manifestations.

Laboratory confirmation during the acute phase is achieved by isolation of the virus, amplification of viral nucleic acid using reverse transcriptase polymerase chain reaction (RT-PCR) or detection of viral antigen. The virus is classified among biosafety level 4 pathogens and hence culturing the virus is restricted to high containment facilities of which there are limited numbers worldwide. In the absence of level four facilities, clinical samples can be inactivated prior to testing.

Virus is most frequently isolated from clinical serum samples collected during the acute phase of illness from days 1–6 after onset of symptoms, although isolation has been reported from samples collected up to day 12. Although the virus replicates in a variety of mammalian primary and line-cell cultures, Vero cells are the most frequently used for routine isolation. The virus is poorly cytopathic and immunofluorescent assays are required to confirm infection of cells. Virus can also be isolated by intracerebral inoculation of 24 h old mice. Isolation of the virus in mice is usually more sensitive than cell culture but it requires 7–10 days for mice to succumb to infections whereas isolation in cell cultures can be confirmed within 3–6 days. Molecular techniques have distinct advantages over traditional virus isolation methods in that a result can be obtained far more rapidly and the assays can be performed using inactivated samples without the requirement for biosafety level four

facilities. The first diagnostic molecular assay was a conventional nested RT-PCR based on alignment of the NP gene of seven geographically distinct isolates of CCHFV and identification of consensus primer pairs. Since the development of the first in house RT-PCR, the availability of nucleotide sequence data for complete and partial genomes from various geographic regions and identification of genetic relationships, defining genotypes and their distribution, has contributed to further refinement of molecular assays. The availability of data allows design of molecular assays to accommodate nucleotide diversity and detect genetically different lineages. Real time RT-PCR platforms have been developed using simple intercalating dye and a melt curve analysis to detect amplicons; sequence specific probes with multiple primer pairs and probes to detect all lineages; low density macroarrays and padlock probes with colourimetric readout. Isothermal assays such as loop-mediated isothermal amplification (LAMP) and an isothermal recombinase polymerase assay have been described with potential for diagnostic application. Commercial molecular assays based on primer and probe technology are now available. Real time assays with probes allow quantification of viral load which has been used as a prognostic indicator and a high viral load can be considered to be a predictor of a fatal outcome.

Detection of viral antigen using ELISA has been described and although it may have a role in low resource laboratories it lacks the sensitivity required for use as a diagnostic tool. With the availability of molecular assays, including isothermal RT-RPA that can be performed without the need for thermal cyclers, viral antigen ELISA are unlikely to be important in a diagnostic facility. However detection of viral antigen in paraffin embedded tissues using immunohistochemistry is a useful tool for confirming infection in fatal cases although molecular assays can also be used to detect viral nucleic RNA in post mortem samples.

During the convalescent phase of illness, serological assays, such as in house or commercial ELISA and indirect immunofluorescent assays (IFA), are used for detection of IgG and IgM antibody. Human infections can be confirmed by detection of a specific IgM response, or alternatively, demonstration of seroconversion or at least a four-fold increase in IgG antibody activity in paired serum samples. An antibody response is most commonly detectable from day 5 after onset of illness in non-fatal cases, whereas in fatal infections patients may not develop demonstrable antibody responses.

Although assays such as targeted amplification and next generation sequencing based identification, Taqman cards and multiplexed Luminex based immunoassays have been investigated as diagnostic tools, the usefulness of these techniques will likely be limited by accessibility to suitable equipment and cost effectiveness in routine application. The need for rapid sensitive and specific point of care assays remains a requirement for increasing diagnostic capacity.

In summary, diagnosis of human infections is usually achieved using a combination of assays including isolation of virus, detection of CCHF viral RNA and demonstration of antibody response. Interpretation of results for exclusion or confirmation of infection usually requires an accurate history of illness and consideration of kinetics of viremia and antibody responses.

Prevention

The CCHFV constitutes a public health risk in endemic countries as well as neighboring areas. To date, WHO put CCHFV in their blue print and highlight the importance on developing vaccines and treatment for this virus. A vaccine against CCHFV will prevent infection in human populations at risk, and reduce the number of cases of CCHF. To date, there is a licensed vaccine against CCHFV available in Bulgaria. This vaccine has been developed in 1970 in Russia and is based on CCHFV cultivated in suckling mouse brain and subsequently inactivated by chloroform, heated at 58°C, and absorbed on Al(OH)₃.

Data from the Bulgarian Ministry of Health suggest a fourfold reduction in reported CCHF cases since they used this vaccine in Bulgaria. However, this observation can be due a several factors which have decreased tick exposure in the region per se. The first detailed analysis of the cellular and humoral immune response in healthy individuals following vaccination with the Bulgarian vaccine demonstrated that vaccinated individuals developed anti-CCHFV immunity however, responses were low and required three booster vaccinations to improve immune responses. It should also be highlighted that the use of mouse brain derived vaccine unlikely gain widespread international regulatory approval due to several reasons such as possibility to autoimmune responses induced by myelin basic protein and etc.

The most obvious problem for development of vaccine candidates against CCHFV, has been the lack of a good animal models. Whilst studies on the immunogenicity of vaccine candidates can increase our knowledge on immune response to these candidates, the outcome are difficult to interpret as no defined immune correlate of protection is available. Nevertheless, to date there are two mice model available which make it possible to assess the protection studies.

The data on vaccine candidates for CCHFV are presented in (Table 1), and the vaccine candidate which has demonstrated efficacy data are discussed in more detail below.

DNA Vaccines

Recently, A DNA vaccine coding for CCHFV Gc, Gn and NP demonstrated induction of both humoral and cell-mediated immune responses which elicited complete protection from lethal disease in a mouse model. A DNA vaccine expressing only the M-segment glycoprotein precursor gene of CCHFV elicited strong antigen specific humoral immune responses with neutralizing titers after three vaccinations, demonstrated 75% protection and highlight that a DNA vaccine expressing the glycoprotein genes of CCHFV elicits protective immunity against CCHFV.

Table 1 Approaches for human vaccines against CCHFV

Vaccine candidate	Antigen	Immune response		Protection	References
		Antibody	T cell		
Inactivated virus (mice brain)	Whole Virus	Positive results	Positive results	Not tested	Mousavi-Jazi, M., et al. Vaccine 2012;
Inactivated virus (cell culture)	Whole Virus	Positive results	Not tested	80%	Canakoglu, N., et al. PLoS Negl Trop Dis 2015;
DNA vaccine	M-seg	Positive results	Not tested	Not tested	Spik, K., et al. Vaccine 2006;
	M-seg	Positive results	Positive results	75%	Garrison, A.R., et al. PLoS Negl Trop Dis. 2017
	GC, GN and N	Positive results	Positive results	100%	Hinkula, J., et al. J Virol 2017.
Transgenic Plant Protein	GpC	Positive results	Not tested	Not tested	Ghiasi, S.M., et al. Vaccine Immunol: CVI 2011;18:
	GN	Positive results	Not tested	Negative results	Kortekaas, J., et al. Vector Borne Zoonotic Dis 2015;
	GC	Positive results	Not tested	Negative results	Kortekaas, J., et al. Vector Borne, Zoonotic Dis 2015;
VLP	GC, GN	Postive results	Positive results	40%	Hinkula, J., et al. J Virol 2017.
Adenovirus	M-seg	Postive results	Postive results	Negative results	Sahib, M.M., et al. 2010
	N	Postive results	Not tested	75%	Zivcec, M., et al. PLoS Negl Trop Dis. 2018
Modified Vaccinia Ankara (MVA)	M-seg	Postive results	Postive results	100%	Buttigieg, K.R., et al. PloS One 2014;
	N	Postive results	Postive results	Negative results	Dowall, S.D., et al. Hum Vaccines Immunother 2016;

Inactivated Vaccines

To date there is a report on a vaccine candidate based on a purified, formalin inactivated of CCHFV derived from cell culture which avoid the complications of a mouse brain derived vaccine (the Bulgarian licensed vaccine). Vaccination with this inactivated virus required the use of adjuvant. The efficacy in the IFNAR^{-/-} mouse model of CCHF showed 80% protection using this vaccine candidate,. While inactivated virus vaccines have proven effective for protecting against many viral diseases, those based on highly pathogenic viruses such as CCHFV are difficult to manufacture in bulk within a high containment setting.

Modified Vaccinia Virus Ankara (MVA)

There is a report describing a MVA poxvirus vector containing the glycoprotein encoding M segment ORF (MVA-GP). This vaccine candidate demonstrated 100% efficacy in animal studies. This MVA-GP vaccine induced antibody and cellular responses across a range of CCHFV epitopes. MVA platform has a proven safety record, having been used in the latter stages of the smallpox eradication campaign, and it is thermostable avoiding the requirement to maintain a cold chain and conferring an extra advantage for regions where such a vaccine might be used.

Adenovirus Vaccine

A human serotype 5 adenovirus (AdHu5) vector carrying the M segment of CCHFV has been investigated as a vaccine candidate using IFNAR^{-/-} mice. Despite the induction of both cellular and humoral responses, the vaccine failed to confer protection in a murine model. Recently, using a construct of human adenovirus type 5 expressing the CCHFV N protein, it was demonstrated that mice immunized with Ad-N developed an anti-N humoral immune response. It was shown that a prime-boost of this vaccine candidate regimen provided 78% protection.

Treatment

The current medical management is largely supportive and on the treatment of symptoms. To date, there are no licensed specific antiviral compounds available for CCHF. Ribavirin, a broadspectrum antiviral drug, has been used as a common treatment approach. Several reports have suggested clinical benefits of using ribavirin, especially if administered early in infection. However, recently several studies have reported no effect on mortality rates. Recent studies have shown that T-705 (favipiravir) has a higher efficacy than ribavirin in animal model. It should be also highlighted that there are other reports demonstrating antiviral activity for several compounds against CCHFV, however, these studies are in vitro. The lack of efficient treatment and vaccines options, emphasizes the urgency for developing CCHF therapeutics.

Further Reading

- Bente, D.A., Forrester, N.L., Watts, D.M., *et al.*, 2013. Crimean-Congo hemorrhagic fever: History, epidemiology, pathogenesis, clinical syndrome and genetic diversity. *Antiviral Research* 100 (1), 159–189.
- Hawman, D.W., Feldmann, H., 2018. Recent advances in understanding Crimean-Congo hemorrhagic fever virus. *F1000 Research* 7.
- Hoogstraal, H., 1979. The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa. *Journal of Medical Entomology* 15 (4), 307–417.
- World Health Organization. Crimean-Congo haemorrhagic fever (CCHF). Available from: <https://www.who.int/emergencies/diseases/crimean-congo-haemorrhagic-fever/en/>.
- Zivcec, M., Scholte, F.E.M., Spiropoulou, C.F., Spengler, J.R., Bergeron, É., 2016. Molecular Insights into Crimean-Congo Hemorrhagic. *Viruses* 8, 106. doi:10.3390/v8040106.

Dengue Viruses (*Flaviviridae*)

Ashley L. St. John and Duane J. Gubler, Duke-NUS Medical School, Singapore, Singapore

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Arbovirus A virus transmitted to vertebrates by hematophagous (blood-feeding) arthropods.

Endemic A disease constantly present in a specific geographic region.

Extrinsic incubation time The time between infection and becoming infectious.

Hyperendemicity The co-circulation of multiple dengue virus serotypes in the same population.

Vertical transmission Transmission of a virus from a female arthropod to its progeny.

History

Dengue fever is a very old disease; the earliest record of a dengue-like illness found to date is in a Chinese encyclopedia of disease symptoms and remedies, first published during the Chin Dynasty (AD 265–420) and formally edited in AD 610 (Tang Dynasty) and again in AD 992 during the Northern Song Dynasty. There are reports of epidemics of dengue-like illnesses in the French West Indies in 1635 and in Panama in 1699. By the late 1700s, a disease clinically compatible with dengue had spread worldwide, with epidemics occurring in 1779 in Batavia (Jakarta), Indonesia and Cairo, Egypt, and in 1780 in Philadelphia, Pennsylvania, USA. From the late 1700s to World War II, repeated epidemics of dengue-like illness occurred in most tropical and subtropical regions of the world at 10- to 40-year intervals. There is no documentation, however, that dengue viruses were responsible for all of these epidemics because the diagnosis was based only on clinical reports. Clinical descriptions of some early epidemics were also compatible with chikungunya virus infection, which has a transmission cycle similar to that of the dengue viruses. It is likely that epidemic chikungunya did occur, but recent data show that the dengue viruses, not chikungunya virus, were responsible for the majority of epidemics in the past 60 years.

The virus etiology of dengue fever was not documented until 1943–44, when Japanese and American scientists simultaneously isolated the viruses from soldiers in the Pacific and Asian regions during World War II. Albert Sabin isolated dengue viruses from soldiers who became ill in Calcutta (India), New Guinea, and Hawaii. The viruses from India, Hawaii, and one strain from New Guinea were antigenically similar, whereas three others from New Guinea were different. These viruses were called dengue 1 (DENV-1) and dengue 2 (DENV-2) and designated as prototype viruses (DENV-1, Hawaii, and DENV-2, New Guinea C). The Japanese virus, isolated by Susumu Hotta, was later shown to be DENV-1. Two more serotypes, called dengue 3 (DENV-3) and dengue 4 (DENV-4), were subsequently isolated by William McD. Hammon and his colleagues from children with hemorrhagic disease during an epidemic in Manila, the Philippines, in 1956. There have been viruses isolated that were suspected to represent new DENV serotypes, but characterization remains ongoing and identification of additional DENV serotypes has not been confirmed.

Many early workers suspected that dengue viruses were transmitted by mosquitoes, but actual transmission was first documented by H. Graham in 1903. In 1906, T. L. Bancroft demonstrated transmission by *Aedes aegypti*, later known to be the principal urban mosquito vector of dengue viruses. Subsequent studies in the Philippines, Indonesia, Japan, and the Pacific showed that *Aedes albopictus*, *Aedes polynesiensis* and other members of the *Aedes* (*Stegomyia*) *scutellaris* complex were also efficient secondary vectors for dengue viruses.

During and following World War II, *A. aegypti* greatly expanded its distribution in Asia, becoming the dominant day-biting mosquito in most Asian cities. Multiple dengue virus serotypes were also disseminated widely at that time. At the end of the war, both the vector (*A. aegypti*) and the viruses (all four serotypes) were widespread in Asia. A dramatic increase in human population growth and urbanization in the postwar years, especially in southeast Asia, created ideal conditions for increased transmission of urban mosquito-borne diseases. As a result, epidemics of dengue became widespread in the region. The first epidemics where dengue hemorrhagic fever (DHF) was reported occurred in the Philippines and Thailand in the mid 1950s. In the 1960s, but accelerating in the 1970s–80s, jet air travel increased dramatically. Population growth and urbanization combined with increased movement of people within and among countries of the region, resulted in increased spread of dengue viruses among population centers, increased frequency of epidemic activity, the development of hyperendemicity (co-circulation of multiple dengue serotypes), and the emergence of epidemic dengue hemorrhagic fever/dengue shock syndrome (DHF/DSS) in the countries of southeast Asia and India during the 1960s. During the 1970s, epidemic DHF/DSS expanded geographically with an accelerated rate in Asia, to the Pacific Islands after an absence of 25 years, and to the Americas. *A. aegypti* had been eradicated from many countries in tropical America in the 1950s and 1960s as part of a program to control urban yellow fever. The program was terminated in the early 1970s and *A. aegypti* began reinfesting the countries where it had been eliminated. Increased epidemic dengue fever followed as the region was reinfested by *A. aegypti* and new viruses were introduced during the 1970s, 1980s, and 1990s.

With the development of global hyperendemicity in the tropics, severe dengue has emerged as one of the world's most important global public health problems in the 21st century. It is a leading cause of hospitalization and death among children in tropical developing countries and causes a high economic and social toll where the disease is endemic. Epidemics occur frequently.

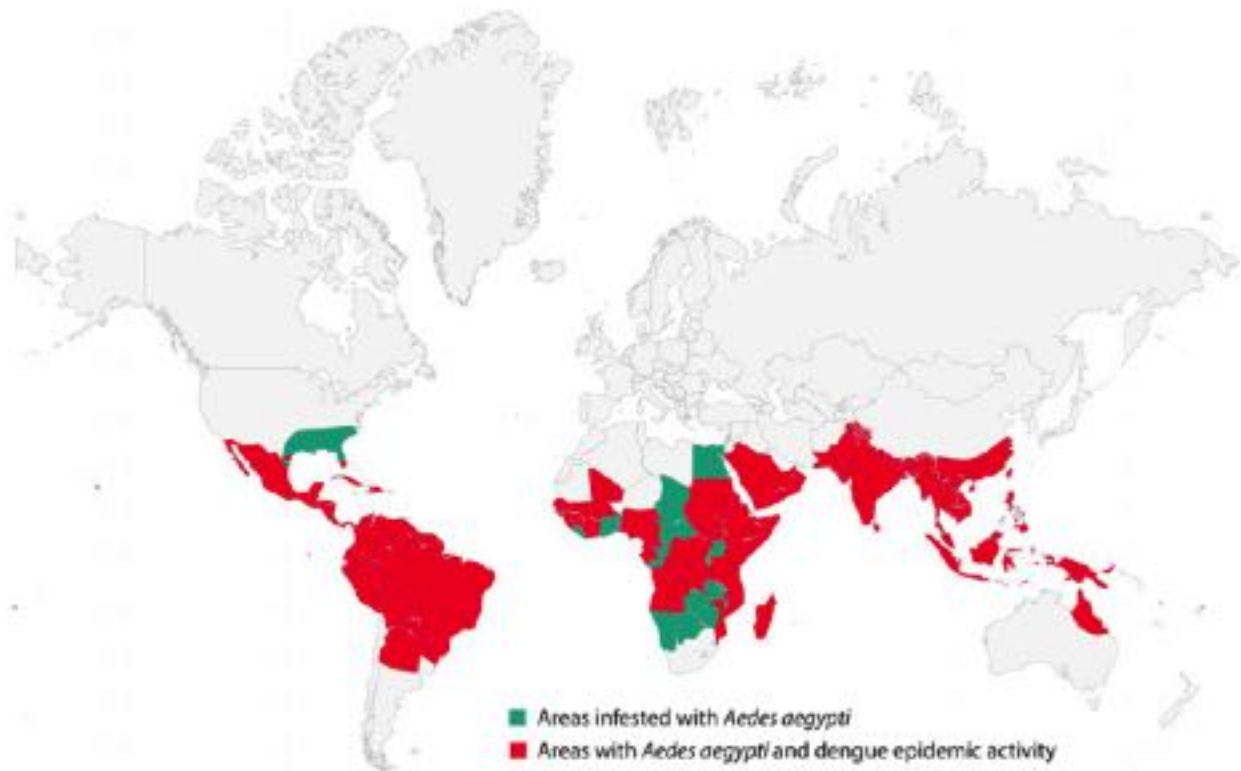


Fig. 1 Global distribution of dengue fever and its principal epidemic mosquito vector, *Aedes aegypti*, 2019.

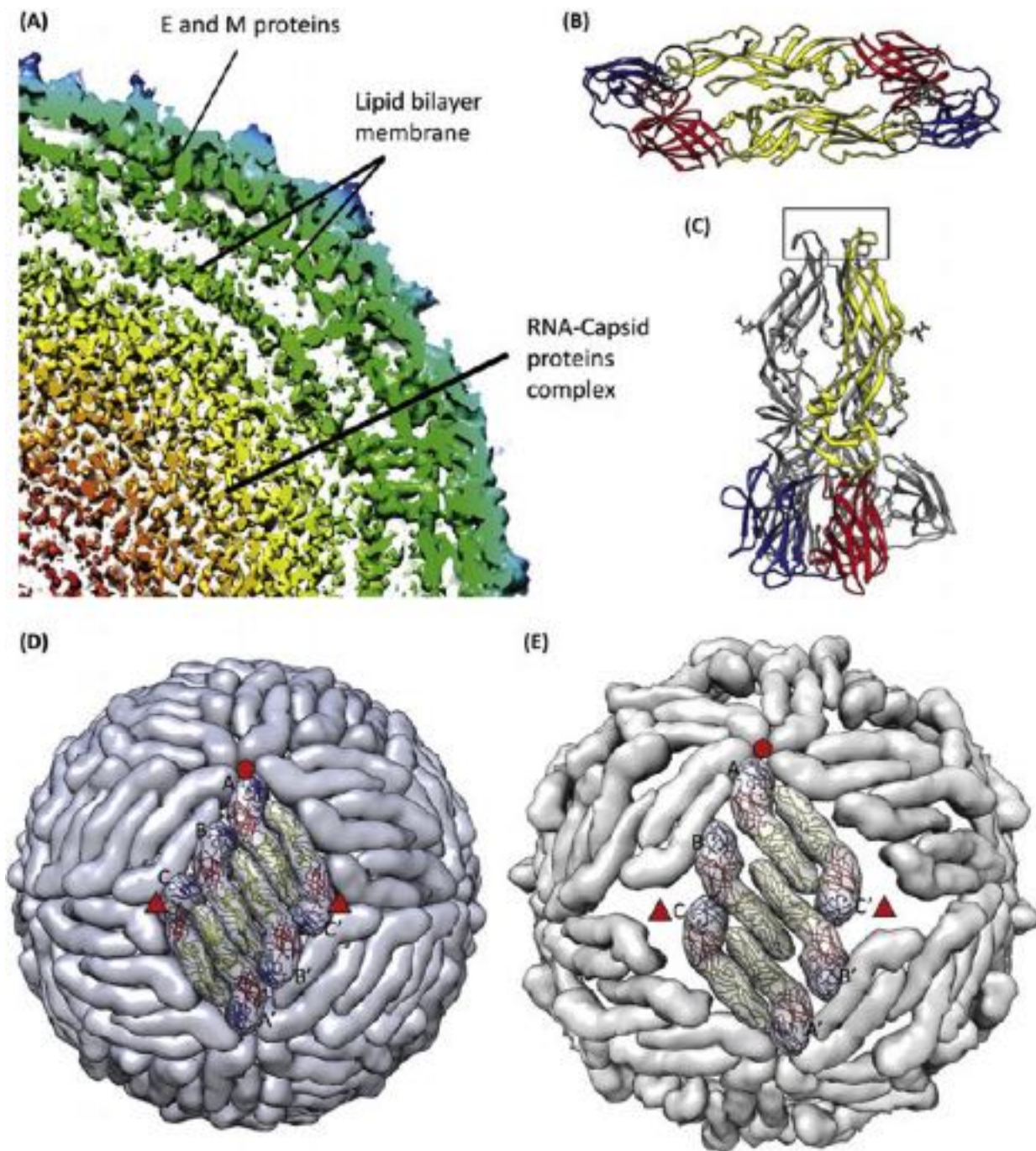
In 2020, dengue fever is the most important arboviral disease of humans, with more than half of the world's population (an estimated 3.6 billion people) living in areas at risk for dengue in tropical and subtropical areas of the world (Fig. 1). An estimated 400 million dengue infections occur each year. The average case–fatality rate for patients with severe dengue is variable, depending on the country, but averages about 5%.

Taxonomy and Classification

Dengue viruses belong to the family *Flaviviridae*, genus *Flavivirus*. There are four serotypes: DENV-1, DENV-2, DENV-3, and DENV-4. Dengue viruses belong to a larger, heterogeneous group of viruses called arboviruses. This is an ecological classification that requires transmission between vertebrate hosts, including humans, and hematophagous arthropod vectors. A discussion of the epidemiologic, evolutionary, biologic, and immunologic relationships of these viruses with each other and with other flaviviruses is beyond the scope of this article.

As are other flaviviruses, dengue viruses are comprised of a single-stranded positive sense RNA genome surrounded by an icosahedral nucleocapsid. The latter is covered by a lipid envelope, which is derived from the host cell membrane from which the virus buds. The mature virion is approximately 50 nm in diameter (Fig. 2). The virus genome encodes three structural and seven nonstructural proteins in a single open reading frame that are expressed as one polyprotein. The open reading frame is flanked by two non-coding regions. The three structural proteins include the nucleocapsid core (C), a membrane associated protein (M), and an envelope (E) glycoprotein. The seven non-structural proteins include the NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (Fig. 3). The functional domains responsible for virus neutralization, hemagglutination, fusion, and interaction with virus receptors are associated with the E protein. Epitope mapping has demonstrated three major domains on the E protein. Immunogenic epitopes have been mapped to all three domains, but domain II appears to be most important for dengue virus neutralization.

There are 53 virus species in the genus *Flavivirus* recognized by the International Committee on Taxonomy of Viruses. The flaviviruses are divided into four subgroups based on the mode of transmission: (1) insect-specific viruses that have only been isolated from various mosquito species; (2) viruses that have no known arthropod vector, and which have been isolated only from rodents and bats; (3) mosquito-borne viruses; and (4) tick-borne viruses. The mosquito-borne and tick-borne groups are the largest, but many insect-specific viruses have been identified in recent years. The dengue viruses belong to the mosquito-borne subgroup, and are considered as one virus species with four serotypes instead of four distinct viruses according to the current classification. Other members of the mosquito-borne group that are of public health importance include the genus prototype yellow fever virus, and Japanese encephalitis, Murray Valley encephalitis, St. Louis encephalitis, West Nile, Zika and others.



Trends in Microbiology

Fig. 2 Mature Dengue Virus (DENV) Structures. (A) A quarter of the central cross-section of a cryoEM map of DENV. The map is colored radially: red (1–130 Å), yellow (131–200 Å), green (201–234 Å), cyan (235–249 Å), and blue (>250 Å). The part of the DENV structure colored in red and yellow indicates the approximate radius that contains the RNA–capsid proteins complex. In green, the lipid bilayer membrane is shown with the E and M transmembrane regions; cyan indicates the E and M ectodomains, and blue indicates the glycosylation sites of E and M proteins. (B) Crystal structure of the prefusion dimeric E protein [Protein data bank (PDB) 1OAN]. DI, DII, and DIII are colored in red, yellow, and blue, respectively. The fusion loop on each E protein is circled in black. (C) Crystal structure of the postfusion trimeric E proteins (PDB 1OK8). One E protein is colored as indicated in (B), the other two are shown in gray. Fusion loops are assembled at one end (enclosed in the black box) and likely interact with the endosomal membrane. (D) Organization of E proteins on the compact mature DENV at 28°C (PDB 3J27). Ribbon representation of an E protein raft is shown. One asymmetric unit contains three individual E proteins: mols A, B, and C. The neighboring symmetry-related E proteins within a raft are labeled as A', B', and C'. All E proteins on the virus surface are also shown as gray surfaces. The 5- and 3-fold vertices are indicated by red pentagon and triangle shapes, respectively. (E) The cryoEM structure of the DENV2 class III particles at 37°C (PDB 3ZKO). All of the E proteins have moved to higher radius. The E protein dimer near the 5-fold vertex (mols A–C') has undergone rotation. The B and B' dimer dissociate from each other. *Source:* Lok, S.-M., 2016. The interplay of dengue virus morphological diversity and human antibodies. *Trends in Microbiology* 24 (4), 284–293.

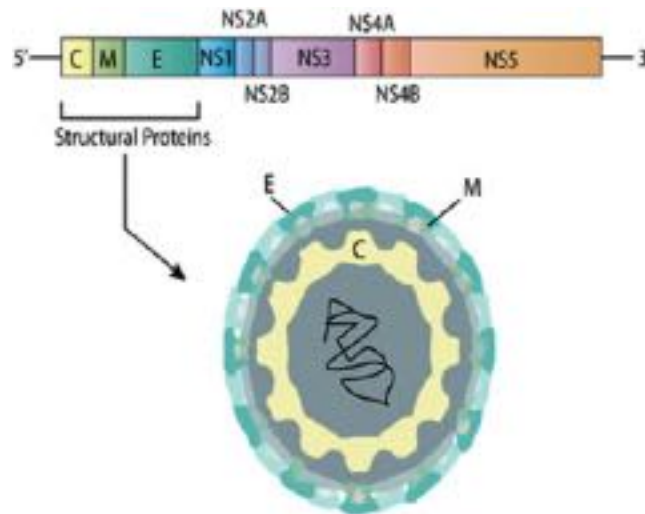


Fig. 3 Organization of the dengue virus genome.

Antigenically, the dengue viruses make up a unique complex within the mosquito-borne group of *flaviviruses*. Although the four dengue serotypes are antigenically distinct, there is evidence that serologic subcomplexes may exist within the group. For example, a close genetic relationship has been demonstrated between DENV-1 and DENV-3 and between DENV-2 and DENV-4. The sizes of the genomic open reading frames of DENV-1, DENV-2, DENV-3, and DENV-4 are 3392, 3391, 3390, and 3387 amino acids, respectively, the shortest among the mosquito-borne flaviviruses. An amino acid sequence positional homology of 63%–68% is observed among the DENV serotypes compared to 44%–51% between DENVs and other flaviviruses such as yellow fever and West Nile.

Geographic and Seasonal Distribution

Dengue viruses have a worldwide distribution in tropical and subtropical areas (**Fig. 1**). The viruses are endemic in most urban centers of the tropics, with transmission occurring year-round, and epidemics occurring every 3–5 years. It is well documented, however, that dengue viruses are maintained for the most part by silent transmission during interepidemic periods in most tropical urban areas and, although the risk of infection is lower than during epidemic periods, there is still substantial risk for infection.

Peak seasonal transmission of dengue viruses during both epidemic and interepidemic periods is usually associated with periods of high rainfall and moderated temperature in most endemic countries. Factors influencing seasonal transmission patterns of dengue viruses are not well understood, but obviously includes mosquito density, which may increase during the rainy season, especially in those areas where water level in larval habitats is dependent on rainfall. In areas where water in storage containers is not influenced by rainfall, however, other factors such as higher humidity and moderate ambient temperatures associated with the rainy season increases survival of infected mosquitoes, thus increasing the probability of secondary transmission. Virus strain, serotype, herd immunity, mosquito species, mosquito behavior and mosquito vector competence also influence the transmission dynamics.

Host Range and Virus Propagation

There are only three known natural hosts for dengue viruses: *Aedes* mosquitoes, humans, and non-human primates. Viremia in humans may last 2–12 days (average, 4–5 days) with titers of infectious virus ranging from undetectable to more than 10^8 mosquito infectious doses 50 (MID₅₀) per ml. Experimental evidence shows that several species of primates (chimpanzees, gibbons, and macaques) become infected and develop viremia titers high enough to infect mosquitoes, but do not develop illness. Viremia levels in lower primates are more transient, often lasting only 1–2 days if detectable, with titers seldom reaching 10^6 MID₅₀ per ml. However, recent studies in Cambodia have suggested that humans with undetectable viremias by PCR were still capable of infecting mosquitoes. This has important implications for virus maintenance during interepidemic periods.

Dengue viruses are known to cause clinical illness and disease only in humans and some non-human primates. Baby mice, which are used for the isolation and assay of many other arboviruses, may show no signs of illness after intracerebral inoculation with most unpassaged strains of dengue viruses. Experimentally, however, some strains can be adapted to produce illness and death in baby mice by serial passage. SCID, Rag^{2-1-8c-1} and AG129 mice have been used for pathogenesis studies since these immunocompromised strains have defects either in innate or adaptive immune responses, allowing robust viral replication. Some strains of dengue, primarily low passage clinical isolates, are also capable of replicating transiently in immunocompetent mice.

Mosquito species of the genus *Aedes* appear to be the natural hosts for dengue viruses. Species of the subgenus *Stegomyia* are the most important vectors in terms of human transmission, and include *A. (S.) aegypti*, the principal urban vector worldwide, *A. (S.) albopictus* (Asia, the Pacific, Americas, Africa, and Europe), *A. (S.) scutellaris*, *A. (S.) polynesiensis*, *A. (S.) hensilli*, *A. (S.) pseudoscutellaris* (Pacific), *A. (S.) africanus*, *A. (S.) taylori* and *A. (S.) luteocephalus* (Africa). It is uncertain what role *A. albopictus* plays in transmission in areas where it has been recently introduced. Species of the subgenera *Finlaya* (Asia) and *Diceromyia* (Africa) appear to be important mosquito hosts involved in forest maintenance cycles of these viruses. Two other species, *Ochlerotatus* (= *Aedes*) (*Gymnotetop*) *mediovittatus* (Caribbean) and *Oc. (=Aedes)* (*Protomacleaya*) *triseriatus* (North America), have been shown to be excellent experimental hosts of dengue viruses.

Low passage or unpassaged dengue viruses can efficiently be propagated only in laboratory-reared mosquitoes and in mosquito cell lines. Mosquito species most commonly used for in vivo propagation include *A. aegypti*, *A. albopictus*, and *Toxorhynchites* spp., all of which can be reared with ease in the laboratory. Only three mosquito cell lines show high susceptibility to dengue viruses: C6/36 from *A. albopictus*, AP-61 from *A. pseudoscutellaris*, and TRA-284 from *T. amboinensis*.

Dengue viruses can also be propagated in baby mice (see above) and in several vertebrate cell lines. These infection models have lower susceptibility to infection than the mosquito cells. However, dengue viruses must be adapted to each system by serial passage before consistent results can be obtained. Mammalian cell lines commonly used include LLC-MK2 and VERO (monkey kidney), BHK-21 (baby hamster kidney), FRhL (fetal rhesus lung), and PDK (primary dog kidney) cells. Propagation of dengue viruses in mosquito cell lines is associated with genetic stability due to slower accumulation of mutations than propagation in mammalian cell lines.

Genetics

Laboratory and epidemiologic studies have demonstrated that genetic variants of dengue viruses occur in nature. This was first observed with DENV-3 isolated during the epidemics in Puerto Rico in 1963 and 1977, and in Tahiti in 1965 and 1969. The viruses in these epidemics were shown to be antigenically and biologically very similar to each other, but very different biologically from Asian strains of the same serotype. Antigenic differences were observed between Caribbean and Asian strains of DENV-4 as well.

Oligonucleotide fingerprinting, restriction enzymes, primer extension, sequencing, and nucleotide sequence comparisons have been used to study the genetic variation among dengue viruses. The number of genetic subtypes identified in each serotype varies with the method used, but based on sequencing of the envelope protein, there are five distinct genotypes of DENV-1, five of DENV-2, four of DENV-3, and four of DENV-4. In general, viruses circulating in the same geographic region during the same time periods have shown genetic homogeneity, while differing from viruses of the same serotype from other regions. However, with increased transmission and spread of dengue viruses worldwide via airplane, they have increased in diversity, influencing both the virulence and epidemic potential. Since there is not an ideal animal model for dengue, it is not well understood how genetic changes or variation influences phenotypic expression of the viruses. That said, however, epidemiologic, clinical and virologic evidence is accumulating showing the importance of the virus strain in determining disease severity and epidemic potential.

Genetic evidence of the diversity of dengue viruses influencing transmission dynamics comes from clinical and epidemiological studies in a series of epidemics in Asia, the Pacific and Caribbean Islands from the 1970s to the present time. In the earlier studies, viruses were isolated and stored unpassaged at -70°C for years until technological advances allowed their study. Full-length genome sequencing was conducted on viruses isolated before, during and after these epidemics where DENV-2, -3 and -4 were predominant. Epidemiological and sequence data from three geographic regions (Asia, Pacific and Americas) and all three serotypes (DENV-2, -3 & -4) show limited genetic changes in the viruses (two-three amino acid changes mainly in the non-structural genes) correlated with phenotypic changes, including clinical severity, viremia levels in humans, virus replication in mosquitoes and cell culture, and epidemic potential. More recent studies have defined the molecular basis for phenotype changes of DENV-1 and -2. Collectively, the data suggest that dengue virus evolution via genetic drift and positive selection may result in the emergence of new virus clades or subtypes, often of the same genotype, that have greater epidemic potential, thus playing an important role in the epidemiology and pathogenesis of severe dengue. Epidemic potential, disease severity and pathogenesis of dengue are determined by a complex interaction of viral, host and environmental factors.

Although there have been reports on intraserotypic recombination events among dengue viruses, the data suggest that this has not been an important factor in their evolution. However, with increased occurrence of the cocirculation of multiple serotypes in an area (hyperendemicity), there have been increased reports of concurrent infections with two serotypes increasing the probability that interserotypic recombination may occur.

Evolution

The origin of dengue viruses is not known. It is not clear which of the four flavivirus groups (insect specific, vertebrate, mosquito-borne, tick-borne) is the oldest. Given the large number of insect-specific viruses identified in mosquitoes in recent years, which have tentatively been placed in the genus *Flavivirus*, however, it seems plausible that an ancestral flavivirus was a mosquito or a tick virus that diverged by adapting to a variety of vertebrate hosts, including rodents, birds, bats, and nonhuman primates. From these early introductions, the no-known vector, tick-borne encephalitis, and mosquito-borne (yellow fever, dengue, and Japanese encephalitis) subgroups arose. Speculation that tick-borne and mosquito-borne viruses had a common ancestor is supported by

the fact that several mosquito-borne flaviviruses (Koutango, Saboya, West Nile, St. Louis encephalitis and yellow fever) have all been isolated from ticks. Also, it has been reported that some tick-borne viruses replicate in mosquitoes or mosquito cell cultures. On the other hand, a comprehensive replication study of 66 flaviviruses in mosquito, tick, and vertebrate cell cultures showed conclusively that there is a strict host range specificity for viruses in the four subgroups, and that the conventional classification based on epidemiologic and phylogenetic relationships is correct.

Regardless of the geographic and evolutionary origin, the data collectively suggests that the dengue viruses most likely evolved in mosquitoes before becoming adapted to lower primates and then to humans, an estimated 1500–2000 years ago. Biologically, dengue viruses are well adapted to their mosquito hosts, being maintained by vertical transmission (from female mosquito to its offspring) in mosquito species responsible for sylvatic cycles, with periodic amplification in lower primates.

As noted above, forest cycles have been documented in South-East Asia and Africa, and possibly in Sri Lanka, India, Vietnam and China. These cycles involve several species of lower primates and three subgenera (*Stegomyia*, *Finlaya*, and *Diceromyia*) of canopy-dwelling mosquito species of the genus *Aedes*. At some point in the past, probably with the clearing of the forests and development of human settlements in Asia, these viruses moved out of the jungle and into a rural environment where they were, and still are, transmitted to humans by peridomestic mosquitoes such as *A. albopictus*. Ultimately, the migration of people moved the viruses into the cities of the tropics where they became “urbanized” and transmitted by the highly domesticated, urban *A. aegypti* mosquito, which had been spread around the world from Africa via sailing ships and increased commerce.

Because of the rather slow rate of change (genetic drift) of the dengue virus genome, viruses isolated over long periods of time in the same geographic region still show striking homogeneity. The greatest genetic differences between dengue virus strains were observed between DENV-2 and DENV-3 isolated from forest mosquitoes in Africa and Asia, respectively, and viruses of the same serotype isolated from humans or mosquitoes in nearby urban areas. This suggests that there is limited gene flow between the forest and urban cycles.

Serologic Relationships and Variability

Dengue viruses share a common morphology, genomic structure, and antigenic determinants with 52 other flaviviruses. Serologic tests most frequently used to determine antigenic relationships have included the hemagglutination-inhibition (HI), complement fixation (CF), and the plaque reduction neutralization (PRNT) tests. Because all flaviviruses share common antigenic determinants, identification of individual family members using these serologic tests is difficult. The dengue viruses make up one antigenic complex within the genus *Flavivirus*. They share complex-specific antigenic determinants on both structural and nonstructural proteins. Serotypes within the dengue virus complex are most accurately and easily identified with an indirect immunofluorescent antibody (IFA) assay using serotype-specific monoclonal antibodies which react with epitopes on the structural proteins, or by a polymerase chain reaction (PCR) using serotype specific primers.

Both antigenic and biological variation among dengue viruses has been documented. As mentioned above, DENV-3 viruses isolated in the Caribbean and the South Pacific in the 1960s were found to be antigenically distinct from the prototype and Asian strains of DENV-3 using PRNT. They were also biologically unique in that they did not grow as well in baby mice and mosquitoes as did the Asian strains. DENV-4 viruses isolated in the Caribbean after the introduction of this serotype into that region in 1981 were antigenically distinct from DENV-4 viruses from Asia.

Field and epidemiological evidence for natural strain variation among dengue viruses is more circumstantial. When DENV-2 was introduced into the South and Central Pacific islands in 1971 after being absent for more than 25 years, epidemics occurred on numerous islands. Marked variation was observed in disease severity, viremia levels, and epidemic potential in epidemics on different islands. This variation was observed with both DENV-1 and DENV-2 in the Pacific and with DENV-3 in Sri Lanka and Indonesia. Some DENV strains appeared naturally attenuated, causing mild illness with low viremia levels of short duration, whereas others caused rapidly spreading epidemics with severe hemorrhagic disease and high viremia levels. Factors that could influence epidemic transmission and disease severity, other than differences in the virus strain, were ruled out as a cause of this variation. Recent studies in Sri Lanka and Puerto Rico have shown that genetic mutations in strains of DENV-3, DENV-2 and DENV-4 have also influenced the epidemic transmission potential of the virus and the disease severity.

Epidemiology

Dengue viruses occur in nature in three basic maintenance cycles. The primitive forest cycle involves canopy-dwelling mosquitoes and lower primates. A rural cycle, primarily in Asia and the Pacific, involves peridomestic mosquitoes (*A. albopictus* and *A. scutellaris* spp.) and humans. The urban cycle, which is the most important epidemiologically and in regard to public health and economic impact, involves the highly domesticated *A. aegypti* mosquito and humans. The viruses have fully adapted to humans and are maintained in most large urban centers of the tropics, with epidemics occurring at periodic intervals of 3–5 years.

A combination of increased urbanization in the tropics, changing life styles, and the lack of effective mosquito control has made most tropical cities highly permissive for transmission of dengue viruses by *A. aegypti*. Increased air travel by humans provides an ideal mechanism to transport dengue viruses between population centers. As a result, in the past 50 years there has been a dramatic global increase in the geographic spread of dengue viruses within and among regions, resulting in increased

epidemic activity, development of hyperendemicity, and increased incidence of severe and fatal forms of the disease, especially DHF/DSS. After being observed only in Southeast Asia, epidemic DHF/DSS has spread to west Asia, the Peoples Republic of China, Taiwan, the Pacific islands, the Americas, Africa and the Middle East in the past 50 years.

Factors responsible for the emergence and spread of the severe form of the disease, DHF/DSS, are not fully understood. The changing disease pattern described above provides support for the principal hypotheses regarding the pathogenesis of DHF/DSS, a secondary infection, and an increased virulence of the virus. Thus, increased transmission in urban areas and the development of hyperendemicity increases the probability of secondary infections and of genetic changes in the virus which may result in more severe disease due to antibody dependent enhancement (ADE) or to an epidemic virus strain with greater virulence. And increased spread of viruses between population centers via modern transportation increases the probability of introducing a virus strain with increased epidemic potential and virulence into new geographic areas.

Increased transmission of multiple dengue serotypes thus increases the probability that severe disease will occur, regardless of whether the underlying cause is due to increased virulence, ADE, or, more likely, a combination of both.

Dengue is primarily an urban disease. Most major epidemics of DHF/DSS occur in tropical urban cities where large and crowded human populations live in intimate contact with the principal mosquito vector, *A. aegypti*. This mosquito is a highly domesticated, day-biting species that lives and breeds in and around people's homes. High mosquito densities often occur in tropical cities because of water-storage practices and the accumulation of domestic waste. Primary larval habitats for *A. aegypti* include a variety of domestic water-storage containers such as clay jars and pots, metal drums, cement cisterns, and many other artificial containers found in the domestic environment that collect and hold rain water. The latter include, but are not limited to flower vases and pots, used automobile tires, buckets, bottles, cans, old machinery, etc.

Transmission and Tissue Tropism

Dengue virus transmission occurs mostly by bites of infective mosquito vectors. Any of the four serotypes may cause high levels of viremia in humans ($\geq 10^8$ MID₅₀ per ml) that lasts on an average of 4–5 days (range, 2–12 days). If a competent mosquito vector takes a blood meal from a person during a viremic phase, virus is ingested with the blood meal and infects the cells of the mosquito mesenteron. After 8–12 days, depending on ambient temperature and the strain of mosquito, the virus will disseminate outside the midgut and infect other tissues, including the mosquito ovaries, nerve ganglia, brain and salivary glands. When the mosquito takes a subsequent blood meal, virus is injected into the person along with the salivary fluid. Dengue virus infection has no apparent effect on the mosquito, which is infected for life.

A. aegypti is a highly competent epidemic vector of dengue viruses. It lives in close association with humans because of its preference to lay eggs in artificial water-holding containers in the domestic environment, and to rest inside houses and feed on humans rather than other vertebrates. It has a nearly undetectable bite and it is very restless in that the slightest movement will interrupt its feeding and make it fly away. It is not uncommon, therefore, for a single mosquito to bite several persons in the same room over a short period of time. If the mosquito is infective, every person bitten may become infected.

In addition to transmitting the virus to humans or non-human primates, the female mosquito may also transmit the virus vertically through her eggs to her offspring. Although the implications of vertical transmission are not fully understood, it is thought to be an important mechanism in the natural maintenance cycles of dengue viruses, especially in rural and forest settings.

The primary site of replication of dengue viruses after injection into humans by the feeding mosquito is not well understood. Infective mosquitoes will often probe the skin multiple times when looking for a blood vessel on which to engorge. Each time it probes, an infective mosquito injects salivary fluid infected with dengue viruses. Dengue viruses deposited in the epidermal and dermal layers likely infect a variety of cells, including dendritic cells, macrophages (Fig. 4). Additional responding cells such as mast cells, T cells, NK cells and others may be activated (Fig. 4). It is uncertain how systemic infection evolves, but it is likely that cells infected in the skin are recruited into the lymphatic system, allowing the infection to spread to the draining lymph nodes and beyond. Human tissues from which dengue viruses have been isolated or have had documented infection include liver, lungs, kidneys, lymph nodes, stomach, intestine, spleen and brain, but it is not known to what extent the virus replicates in these tissues. Infected cell types that have been identified include Langerhans cells, dermal dendritic cells, endothelial cells and macrophages and monocytes. Virus replication has also been detected in hepatocytes and Kupffer cells in the liver. Pathological changes similar to those observed in yellow fever, with focal central necrosis, have been observed in the liver of some patients who have died of dengue virus infection. There is some evidence that the viruses also replicate in endothelial cells and possibly in bone marrow cells. Encephalopathy has been documented in dengue infection but whether dengue viruses cross the blood–brain barrier and replicate in the central nervous system is still an open question. Dengue viruses have been transmitted by blood transfusion and organ transplantation.

Pathogenesis

Dengue disease expression in humans is the result of the complex interaction between an individual with specific genetic and immunological characteristics and the strain of dengue virus with specific virulence and biological characteristics. The result may range from asymptomatic infection to a severe disease and death. Unfortunately, animal models do not fully replicate the symptoms and kinetics of human infections, which vary greatly. Most of what we know or think we know is based on in vitro

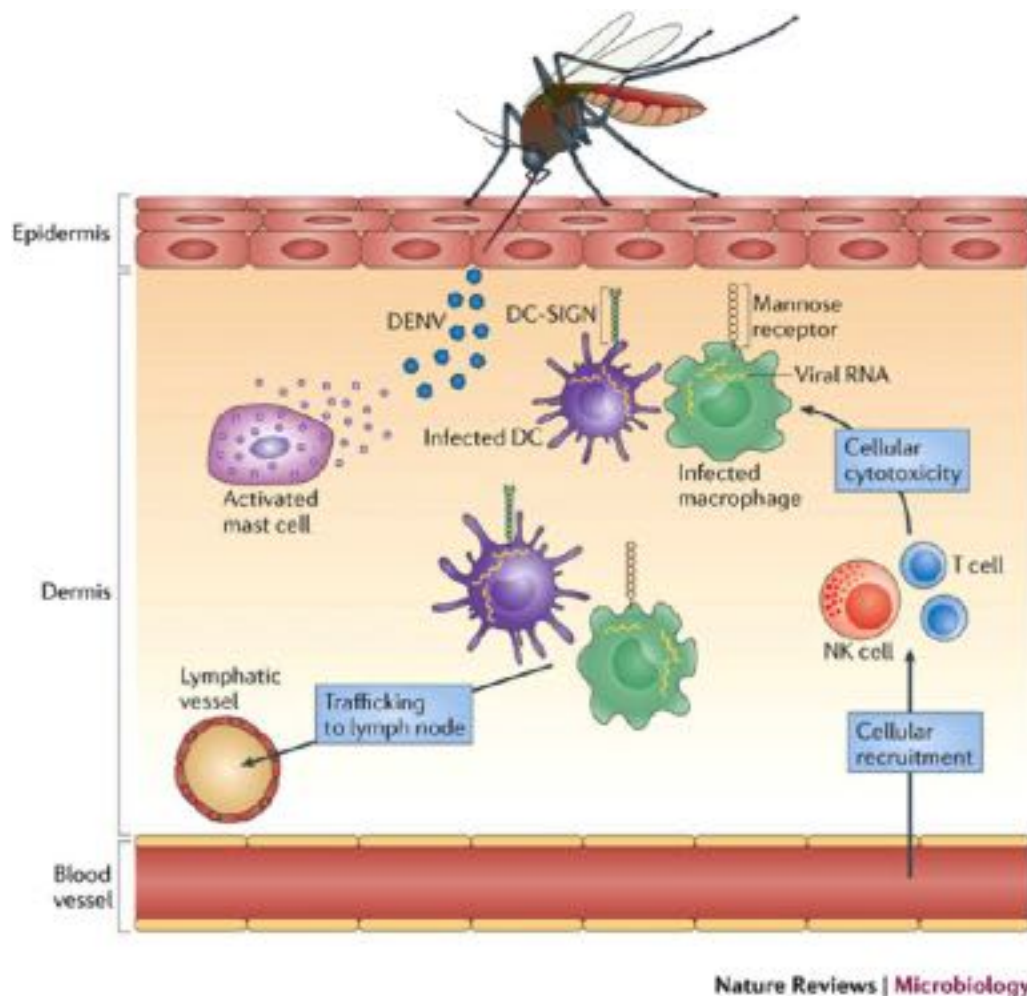


Fig. 4 Host responses to cutaneous dengue virus. Most infections by dengue virus (DENV) occur after subcutaneous injection of the virus into the skin. Released viral particles may infect nearby cells such as monocytes or dendritic cells (DCs) or activate resident immune cells like mast cells. A local inflammatory response to DENV in the skin induces the recruitment of leukocytes from the vasculature, including natural killer (NK) cells and T cells, which promote the killing of virus-infected cells at the site of injection. DENV is thought to then travel to draining lymph nodes via lymphatic vessels to establish systemic infection. These localized inflammatory responses occur days before there are any signs of severe infection. *Source:* St. John, A.L., Abraham, S.N., Gubler, D.J., 2013. Barriers to pre-clinical investigations of anti-dengue immunity and immune-pathology. *Nature Reviews Microbiology* 11 (6), 420–426.

studies using highly passaged (adapted) viruses and artificial host systems, although low-passage clinical isolates are increasingly used for experimental studies. Therefore, there is still considerable controversy about the pathogenesis of severe dengue disease, including DHF/DSS.

Disseminating virus is first detected in the draining and remote lymph nodes, leading to a systemic infection and viremia, which may be detectable 24–48 h prior to clinical symptoms. Leukopenia, neutropenia and thrombocytopenia are common, with altered bone marrow cellularity. Evidence suggests that at least two pathogenetic mechanisms are associated with severe dengue infection. Classical DHF/DSS is characterized by a vascular leak syndrome which, if not corrected, may rapidly lead to hypovolemia, shock, and death (Fig. 5). Multiple theories of dengue pathogenesis exist and these primarily focus on the role of immune-mediated pathology in the initiation of vascular leak syndrome. In secondary infections, the mechanisms implicating non-neutralizing antibodies, or weakly cross-reactive T cells in generating high viral titers have been proposed to contribute to severe disease. The underlying pathogenetic mechanism for this syndrome is thought to be an immune enhancement phenomenon in which the infecting virus complexes with non-neutralizing dengue antibodies, thus enhancing infection of mononuclear phagocytes. Phagocytes produce vasoactive mediators, which may result in vascular permeability and leakage. Serum anti-dengue virus antibody levels may also be a factor in antibody-mediated pathologies. Moderate concentrations of cross-reactive antibodies are thought to promote enhanced disease severity, whereas low concentrations may be insufficient to induce the phenomenon and high concentrations could offer some cross-protection. Poorly neutralizing cross-reactive antibodies are also able to induce enhanced activation of mast cells, which release vasoactive mediators. Another hypothesis emphasizes the

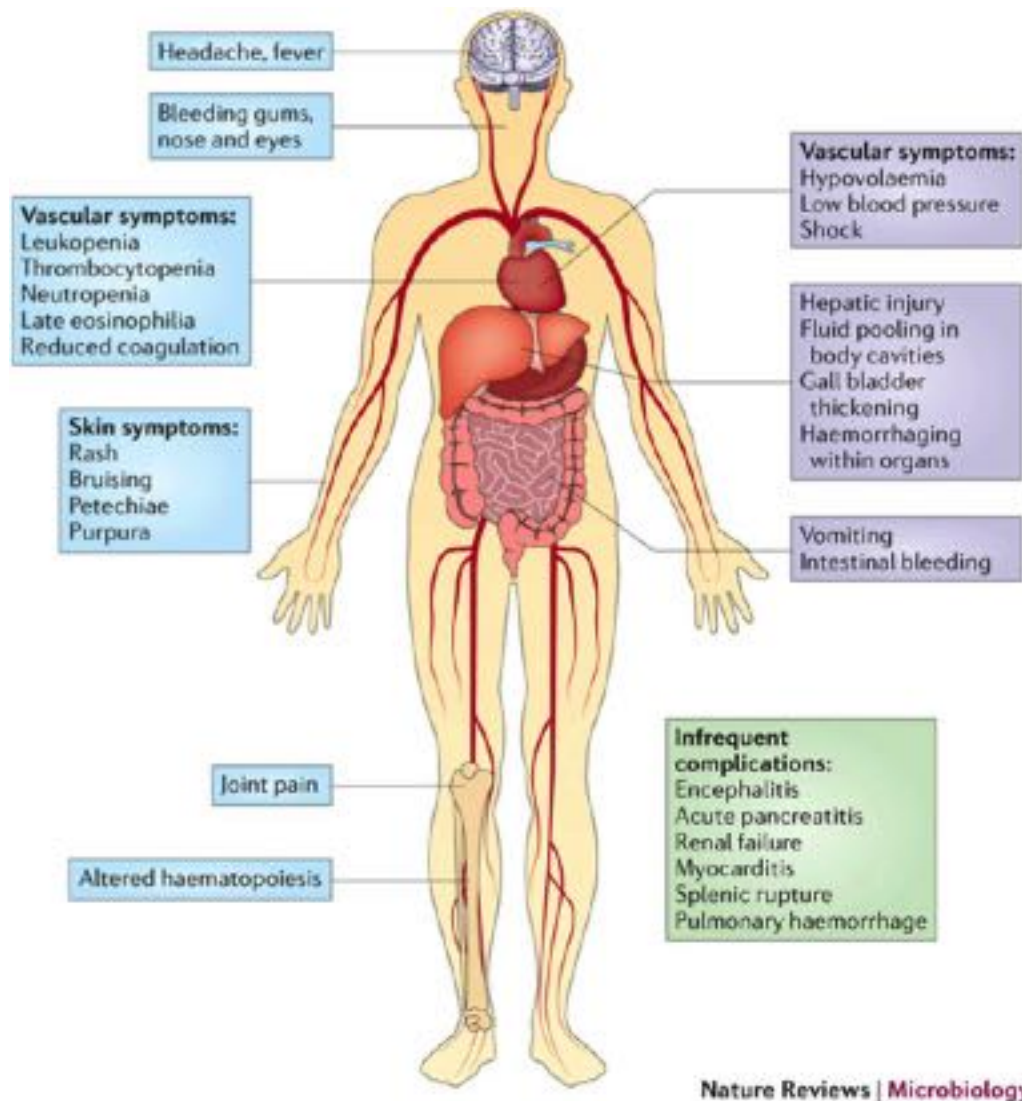


Fig. 5 Clinical symptoms and pathogenesis of dengue in humans, across the spectrum of mild to severe disease. *Source:* St. John, A.L., Abraham, S.N., Gubler, D.J., 2013. Barriers to pre-clinical investigations of anti-dengue immunity and immune-pathology. *Nature Reviews Microbiology* 11 (6), 420–426.

importance of T cells in the clearance of dengue-infected cells. During secondary infection, memory recall responses by T cells that are only weakly cross-reactive for the secondary strain may exhaust immune resources and outcompete naïve T cells that would be more specific for the strain causing secondary infection. These activated T cells may also produce cytokines, in addition to infected cells, that could promote vascular leakage. Several vasoactive cytokines and host inflammatory factors have been identified in different human studies as being associated with severe dengue syndrome, including TNF, VEGF, and mast cell derived proteases. Loss of plasma from the vascular compartment may range from mild and transient to severe and prolonged, the latter often resulting in life threatening irreversible shock and death. Although there is a slightly higher risk (~10%) of developing classical DHF/DSS in persons experiencing secondary dengue infections, DHF/DSS may also occur in primary infections. This suggests that additional mechanisms of vascular pathogenesis exist that are not dependent on heterologous adaptive immune responses. Interestingly, DHF/DSS rarely or never occurs in the 3rd or 4th infection.

In vitro studies have shown that not all dengue viruses can be enhanced and that not all antibodies can enhance dengue infection. That, in addition to epidemiologic and virological studies showing that dengue viruses in nature vary greatly in their epidemic potential, and virulence (see Section “Genetics” above), raises the question as to whether dengue virus strains vary in their ability to stimulate the production of enhancing antibodies, whether this is associated with virulence, and, if so, how this relates to the immune enhancement hypothesis. Unfortunately, animal models cannot explain the disease mechanisms of different dengue strains in humans. However, accumulating sequence and field data suggest the dengue viruses, like most other animal

viruses, vary in their virulence and in their epidemic potential. As already noted, both host immunity and virus virulence are likely playing a role in the pathogenesis of severe dengue disease.

Patients infected with dengue viruses may also experience other forms of severe disease, including severe and uncontrolled bleeding, usually from the upper gastrointestinal (GI) tract. This severe hemorrhagic disease may be associated with multiple organ failure, and it is more difficult to manage than classical vascular leak syndrome. The underlying pathogenetic mechanism for this type of bleeding is not fully understood.

A third type of severe and fatal dengue infection, which may or may not involve overt hemorrhagic disease, is neurological symptoms. Although many patients with this syndrome present clinically as viral encephalitis, conclusive evidence that dengue viruses commonly infect the central nervous system has not yet been obtained. Available data suggests that neurological symptoms may be secondary to cerebral hemorrhage, edema, or other indirect effects of dengue virus infection.

Clinical Features of Infection

Dengue virus causes a spectrum of illness in humans ranging from subclinical to severe and fatal hemorrhagic disease with the latter representing only a small proportion of all infections. The incubation period may be as short as 3 days and as long as 14 days, but most often it is 4–7 days (Fig. 6). The majority of patients present with a mild, nonspecific febrile illness, or with classical dengue fever. The latter is generally observed in older children and adults, and is characterized by a sudden onset of fever, frontal headache, retro-ocular pain, and myalgias. Rash, joint pains, nausea and vomiting, lymphadenopathy and hemorrhagic manifestations such as nose and gum bleeding, purpura and petechiae are common (Fig. 5). The acute illness, which lasts for 3–7 days, is usually benign and self-limiting, but it can be very debilitating. Convalescence may be prolonged for several weeks.

As noted, severe dengue may present in several forms. The most common form of severe dengue is a vascular leak syndrome, often observed in children under 15 years of age, although this syndrome may also occur in adults in areas of lower endemicity. Severe dengue is characterized by acute onset of fever and a variety of nonspecific signs and symptoms that may last 2–7 days (Fig. 6). During this stage of illness, dengue is difficult to distinguish from many other viral, bacterial, and protozoal infections. In children, upper respiratory tract symptoms caused by concurrent infection with other viruses or bacteria are not uncommon. The differential diagnosis should include other hemorrhagic fevers, hepatitis, leptospirosis, typhoid, malaria, measles, influenza and other diseases causing fever.

The critical stage in severe dengue occurs when fever subsides to or below normal. This usually occurs 5–7 days after onset of fever (Fig. 6). Most patients recover without complications, but the patient's condition may deteriorate rapidly with signs of circulatory failure, neurological manifestations, shock and death, if proper management is not available. The main pathophysiological features of severe dengue involve immune system activation, increased microvascular permeability and plasma leakage into the extravascular spaces. There may be altered hemostasis involving thrombocytopenia, impaired coagulation and vascular dysfunction. Skin hemorrhages such as petechiae, easy bruising, bleeding at the sites of venipuncture, and purpura/ecchymosis are the most common hemorrhagic manifestations; GI hemorrhage may occur, usually after, but in some cases before, the onset of shock.

The World Health Organization (WHO) has recently redefined the definition of severe dengue. Previously limited to only DHF/DSS, other forms of severe disease not associated with vascular leakage are now included (Fig. 7). The strict criteria to define DHF remain unchanged with four major clinical manifestations: fever or recent history of fever, hemorrhagic manifestations, vascular leakage, and thrombocytopenia. Patients with DSS have DHF with objective signs of circulatory failure. Thrombocytopenia and hemoconcentration are constant features of DHF/DSS. As noted above, other severe forms of dengue include severe hemorrhage, myocardiopathy and neurological disease, which have been added to the new WHO guidelines. Patients may present with severe and uncontrollable upper GI bleeding with shock and death in the absence of hemoconcentration or other evidence of the vascular leak syndrome. Also, they may present with a variety of neurological and psychiatric disorders, including headache, dizziness, hysteria, and depression. In addition,

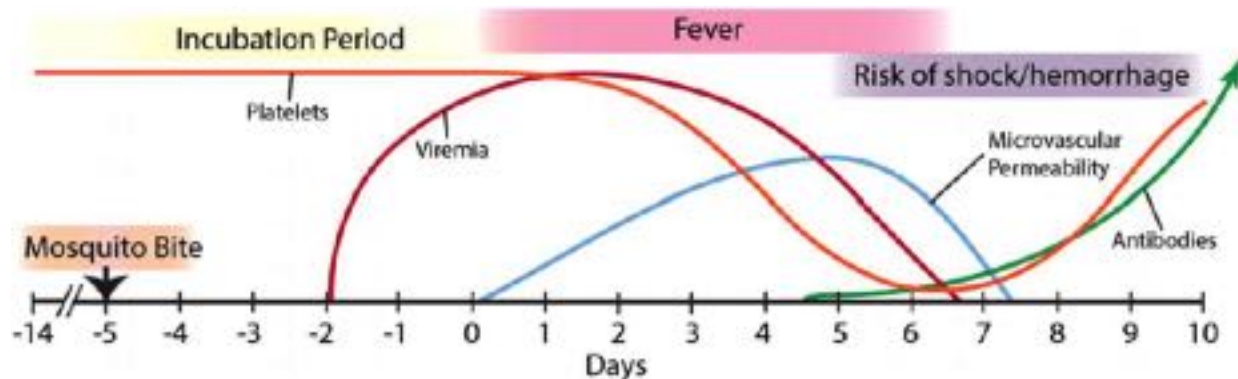


Fig. 6 A time course of dengue clinical symptoms during infection.

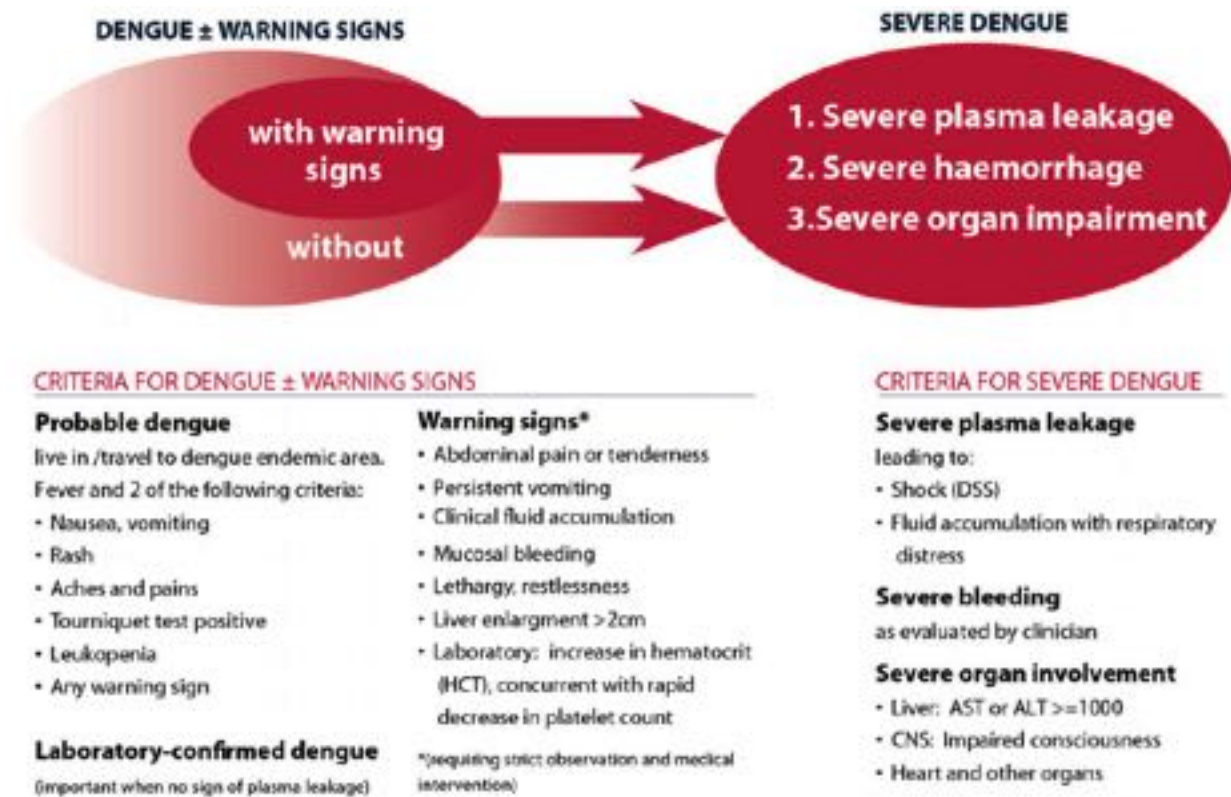


Fig. 7 Suggested dengue case classification and levels of severity; adapted from the 2009 WHO Dengue guidelines for diagnosis, treatment, prevention and control.

some patients present with clinical symptoms of viral encephalitis and Guillain-Barré syndrome, but as noted above, there is no conclusive evidence that CNS infection occurs.

Treatment of dengue infection is symptomatic since there are no specific antiviral drugs. Management of vascular leakage syndrome, however, is very effective with a corrective fluid replacement therapy. The prognosis of patients with DHF/DSS depends on early recognition, monitoring, and management of vascular leakage. Good nursing care is critical. A definitive diagnosis of dengue virus infection can only be made in the laboratory by serologic and/or virologic methods.

Pathology and Histopathology

The pathology of dengue virus infection is not well understood because systematic postmortem studies have not been done on patients representing all types and stages of severe dengue disease. The pathology may also be complicated by patient age, and co-morbidities. In general, the pathology of dengue is non-specific and difficult to distinguish from yellow fever, other viral hemorrhagic fevers, leptospirosis, and other diseases. Petechiae, serous effusion and organ hypoperfusion are common. Liver pathology is similar to that seen in yellow fever with centrilobular and mid-zonal necrosis, apoptosis and steatosis. There is often an increased proportion of immune cells seen in blood and in the spleen. Studies have not revealed destructive inflammatory vascular lesions, but some swelling and occasional necrosis has been observed in endothelial cells. Changes in the kidneys are suggestive of an immune complex type of glomerulonephritis. Complement deposition has also occasionally been identified in tissues. There is impaired bone marrow function, which improves when the patient becomes afebrile. Biopsy studies of the skin rash have demonstrated perivascular edema with infiltration of lymphocytes and monocytes.

Immune Response

Persons infected with dengue viruses produce IgM and IgG antibodies, both of which appear 5–7 days after onset of illness in primary infections. The highest titers of IgM antibodies are detected in primary dengue infections, but an IgM response is also detected in secondary and tertiary infections. The IgM response is transient and generally disappears 30–90 days after onset of illness in primary infections and after shorter periods of time in secondary and tertiary infections. IgG antibodies, in contrast, may persist for at least 60 years and probably throughout the life span of the patient. In persons experiencing their first dengue or

flavivirus infection, peak IgG levels are detected 14–21 days after onset of illness and seldom exceed titer values of 640–1280 as measured by hemagglutination-inhibition or neutralization tests. There are exceptions, however, with some infections inducing antibody titers typical of secondary or tertiary infections during primary infection. In secondary infections, there is a rapid anamnestic IgG immune response to dengue complex-and/or other flavivirus-specific antigenic determinants. Both IgM and IgG antibodies neutralize dengue viruses, and infection provides life-long immunity to the infecting virus serotype.

Both IgM and IgG antibodies to dengue viruses cross-react with other flavivirus antigens, including those of yellow fever, Japanese encephalitis, West Nile, Zika, and St. Louis encephalitis viruses. Those flavivirus cross-reactive antibodies are generally non-neutralizing. Cross-reactivity with viruses in the dengue complex makes the interpretation of serologic results difficult. In geographic areas where several flaviviruses are endemic, definitive laboratory diagnosis can only be made by virus isolation or nucleic acid detection, and in patients with a primary infection, by PRNT. Normally, a combination of laboratory (serologic and virologic), clinical, and epidemiologic data is used to confirm a diagnosis of dengue or other flavivirus infections.

Because dengue virus-specific IgG antibodies persist for many years or decades, their presence in a single serum sample is not diagnostic of current or recent infection. Lower IgG titers simply indicate that the person has had a previous infection at some time in the past. Paired serum samples are required to confirm a current infection by demonstrating a fourfold or greater rise in IgG antibody levels. The presence of detectable IgM antibody in a single serum sample is considered to be diagnostic for a recent infection because this isotype does not persist for long periods of time. Increasingly, NS1 protein is used as a diagnostic marker of acute dengue infection since this viral protein is secreted from infected cells and can be detected in patient serum. NS1, like RNA detection by PCR, is a surrogate marker for the presence of infectious virus.

Extensive work done in recent years has demonstrated that immunopathogenetic mechanisms play a major role in the pathophysiology of vascular leakage syndrome. Generally, CD4⁺ and CD8⁺ T-cell responses are directed against multiple viral proteins, including both structural and non-structural proteins. After primary infection, memory T cells proliferate in response to multiple dengue serotypes providing both specific and cross-reactive responses. After a secondary infection, T cell proliferation is of low-affinity and may be more reactive to the first infecting virus serotype than to the second virus strain, a phenomenon analogous to original antigenic sin seen in the antibody response. Whether T cell activation is involved in immune pathology is controversial and results from different human studies vary considerably. T cells have increasingly been suggested to contribute to infection control, both through the development of cytotoxic responses and through promoting high affinity antibody production. NK cells have also been identified to be activated during dengue infection; however, it is not yet clear whether they are primarily protective or potentially contribute to immune pathology.

The vascular leakage syndrome is commonly associated with a cascade of immunopathologic events in which increased infection of dendritic cells, macrophages and other cell types lead to a cytokine storm that is thought to increase vascular permeability. The literature supporting this hypothesis is extensive, but highly variable with many studies lacking controls or performed in immune compromised animals, where there is a potential of enhancing the role of virus replication in immune pathology or promoting infection in cell types that are normally resistant to infection. There is strong evidence, however, that pro- and anti-inflammatory cytokines are elevated during the acute febrile phase of the disease. Aside from cytokines, additional vasoactive inflammatory mediators and inhibitors of coagulation are released during an acute infection, including products such as mast cell proteases, heparin, and matrix metalloproteinases.

There is accumulating evidence that cell-mediated immunity also plays a role in terminating dengue infections. Dendritic cells (DCs) are likely the initial site of virus replication. Dengue virus infection stimulates DC maturation and activation, and the production of TNF- α , IFN- γ and IFN- α/β that are known to inhibit virus infection. DCs migrate to T-cell-rich lymphoid organs where T cells are activated, stimulating memory responses and releasing cytokines and chemokines. Transcriptional activation of DCs has been associated with asymptomatic infections, supporting the concept that T cell responses can be protective. During secondary infections, T cells may be specific for a previous infecting serotype of dengue and cross-reactive for the secondary serotype. In that situation, T-cell responses are likely to be dominated by a subset of memory T cells, which produce IFN- γ and CD40L, resulting in a better DC activation, T cell stimulatory capacity, IL-12 release, increased secretion of TNF- α and IFN- γ , as well as potentially altered cytokine responses. When viremia is cleared, the cascade of events initiated by a Th1 cytokine response in the absence of effective viral control may contribute to the pathogenesis of DHF. Cross-reactive CD8 T cells that have low affinity to the second infecting virus serotype (see above) may be ineffective in clearing viremia, thus permitting a higher virus load and a more severe disease. However, cross-reactive CD4 T cells may contribute to cross-protection by promoting germinal center responses and enhancing the kinetics of antibodies that can neutralize the secondary serotype.

Prevention and Control of Dengue

The options available for prevention and control of dengue have been limited, and dependent on controlling the mosquito vector, which has failed in most countries during the past 50 years. A highly effective vaccine against dengue is not available despite more than 70 years of research. However, recent progress in developing new mosquito control tools, vaccines and antiviral therapeutics has been promising, but it is unlikely that any single method or approach will be successful in reversing the trend of increased epidemic dengue. Most likely, a combination of vector control and vaccination will be needed to effectively control this disease.

Vaccines

Effective vaccination to prevent dengue will likely require a live attenuated vaccine since this strategy ensures induction of both T and B cell memory responses. One vaccine produced by Sanofi Pasteur under the name Dengvaxia, has been licensed for use in people 9–45 years of age in some countries. In clinical efficacy trials Dengvaxia has been shown to be effective in preventing severe disease by 93% and decreasing hospitalization by 81%. The vaccine is most effective in individuals who have previously been exposed to the virus. However, 6 year follow up of vaccinees showed that there is a small increased risk of severe disease and hospitalization in vaccinees who were dengue-naïve at the time of vaccination, leading WHO to recommend that it not be used in seronegative individuals. The WHO recommended pre-vaccination screening for serological evidence of previous dengue infection, but also suggested that Dengvaxia could be used in highly endemic areas with a seroprevalence of 80% or higher. Currently, it is uncertain whether or how Dengvaxia will be used in dengue control programs.

The dengue vaccine pipeline currently includes 2 candidate vaccines in phase 3 clinical trials, one in phase 2 and two in phase 1 (Table 1). In addition, there are several third generation vaccines under development using new molecular technology. The two most promising live attenuated candidate vaccines (CDC/Takeda and NIH/Merck) are expected to be licensed in the next 3–5 years. It is not known whether a balanced tetravalent protective response will be obtained. From a public health perspective, however, that may not be necessary. Current thinking is that a vaccine that protects against three or even two of the virus serotypes may be effective in protecting against a severe disease since 3rd and 4th infections are rarely if ever associated with a severe disease.

A very promising issue has also been the rapid progress in developing antiviral drugs and therapeutic antibodies that can be used in the treatment of dengue infection, and perhaps, even in prevention and control programs. Research is currently focused on developing a safe and inexpensive drug that inhibits dengue virus replication that can be used in a manner similar to influenza antivirals. While most effort has been directed towards developing a small molecule drug, many studies are now screening drugs that were developed against other viruses and that have proven to be safe in humans.

Monoclonal antibodies are also being developed for dengue therapeutics. Antibodies are able to limit virus infection through several known mechanisms including by aggregating virus particles, destabilizing the virus structure, preventing virus attachment to target cells, or by preventing fusion of the virus with the target cell membrane. Some monoclonal antibodies that are most effective in neutralizing DENV bind to the virus surface, locking the virus structure and preventing the dynamic movements required for virus attachment and fusion with the cell membrane. A DENV-1-specific therapeutic monoclonal antibody has shown promising results in early clinical trials and a DENV-2-specific antibody is being developed. These antibodies could be used for both treatment and for prophylaxis.

Currently, the most widely used way to prevent dengue infection is to control the mosquito vector that transmits the virus. Unfortunately, with the exception of Singapore, programs to control *A. aegypti* mosquitoes have not been successful. For more than 50 years, the recommended method to control mosquitoes has been spraying ultralow volume (ULV) or thermal fog application of insecticides to kill adult mosquitoes. Field trials in Puerto Rico, Jamaica, and Venezuela, however, showed that this method was not effective in significantly reducing natural mosquito populations for any length of time. This supports epidemiologic observations that space spraying has little or no impact on epidemic transmission of dengue viruses. Unfortunately, however, the method is still the most widely recommended.

Recent research has led to the development of a number of new and innovative chemical, genetic and biologic mosquito control tools. The insecticides include new classes of chemicals that will provide residual activity for 6 months to a year, a characteristic that is critical to control adult mosquitoes emerging from cryptic or hidden larval habitats. Other new insecticides can be formulated as spatial repellants to prevent mosquitoes from entering specific areas or to treat both natural and artificial oviposition sites. Both genetic and biological approaches to mosquito control are equally promising.

The most promising genetic approach is the release of sterile male mosquitoes into a natural population to mate with wild female mosquitoes. Male sterilization can be achieved by genetic modification, irradiation or biologically by Wolbachia infection. Sterile males can be produced by the millions and released as eggs, pupae or adults into natural populations. If the sterile males are released in high enough numbers, egg hatch rate will decrease and the wild mosquito population will be suppressed. Field trials have been conducted in Malaysia, Cayman Islands, Panama, Brazil, Singapore, the United States and China, all with promising results.

Table 1 The dengue vaccine pipeline by stage of clinical development

Company	Vaccine strategy	Development phase
Sanofi Pasteur	Chimeric, YF17-D backbone, DENV1–4	Licensed
NIH/Merck	LAV, DENV-1,-3 & -4, Δ30/31 + Chimeric, 2/4	Phase 3
CDC/Takeda	LAV, DENV2 PDK53; + Chimeric DENV1/2, 3/2 & 4/2	Phase 3
WRAIR/GSK	Inactivated, DENV1–4	Phase 2
NMRC	DNA, DENV1–4	Phase 1
Merck/Hawaii Biotech	Subunit, DENV1–4	Phase 1 (on hold)

Source: YF17-D, Yellow fever vaccine strain 17-D; LAV, live attenuated virus; GSK, Glaxo-Smith Kline; NIH, National Institutes of Health; CDC, Centers for Disease Control and Prevention; NMRC, Navy Medical Research Center; PDK53, a live-attenuated DENV2 vaccine developed at Mahidol University, Thailand.

A promising biological control tool is the infection of *A. aegypti* mosquitoes with a common bacterium, Wolbachia, that naturally infects over 70% of all insects in nature, but for reasons that are unknown, not *A. aegypti*. Several strains of Wolbachia have been adapted to infect *A. aegypti*, causing their decreased survival and decreased susceptibility to dengue, yellow fever, Zika and chikungunya viruses. Field trials have been carried out in Australia showing effective spread of Wolbachia into the natural *A. aegypti* populations and effective control of dengue in one city. Other trials in Vietnam, Indonesia, Brazil and Colombia are in progress.

An effective method of controlling *A. aegypti* is larval control, ie, to eliminate or control the larval habitats where the mosquitoes lay their eggs. Unfortunately, that is not possible in old cities that have expanded over the centuries and have many hidden or cryptic larval habitats. Fortunately, a number of these new mosquito control tools may be useful in reaching these sites, most of which are found in the domestic environment, where most transmission occurs. To have sustainability of prevention and control programs, some responsibility for mosquito control should be transferred from governments to citizens. For long-term sustainability, mosquito control programs must be regional, but operational control must be local, community-based and integrated. Persons living in *A. aegypti* infested communities have to be educated to take responsibility for their own health and help government agencies control vector mosquitoes, and thus prevent epidemic dengue. Effective sustainable prevention and control programs must involve a partnership between government agencies and the community.

Countries with endemic dengue should develop active, laboratory-based surveillance systems that can provide some degree of early warning for epidemic transmission. Finally, prevention of excess mortality associated with DHF/DSS can be achieved by educating physicians in endemic areas on clinical diagnosis and management of severe dengue. As demonstrated in countries such as Thailand, early recognition and proper clinical management are the key factors reducing DHF/DSS case-fatality rates.

Future

Continued population growth and urbanization of the tropical and subtropical areas, changing lifestyles, globalization and lack of effective mosquito control have been the most important factors responsible for the dramatically increased incidence and geographic expansion of dengue in the past 50 years. Epidemic dengue has become a global public health problem in the tropics and it is anticipated that this trend will continue unless more effective prevention and control strategies are developed and implemented worldwide. Fortunately, the dengue control pipeline includes a number of new and innovative tools, including vaccines, antiviral drugs, therapeutic antibodies, insecticides and other mosquito control tools, which if used in an integrated and synergistic way, could allow effective control of dengue in the near future.

Further Reading

- Bhatt, S., Gething, P.W., Brady, O.J., *et al.*, 2013. The global distribution and burden of dengue. *Nature* 496, 504–507. doi:10.1038/nature12060.
- Gubler, D.J., Kuno, G., Markoff, L., 2007. Flaviviruses. In: Knipe, D., Howley, P. (Eds.), *Fields Virology*, fifth ed. Philadelphia: Lippincott Williams and Wilkins, pp. 1153–1252.
- Gubler, D.J., Ooi, E.E., Vasudevan, S., Farrar, J. (Eds.), 2014. *Dengue and Dengue Hemorrhagic Fever*, second ed. Wallingford: CAB International.
- Gubler, D.J., 1989. *Aedes aegypti* and *Aedes aegypti*-borne disease control in the 1990's: Top down or bottom up. *American Journal of Tropical Medicine and Hygiene* 40, 571.
- Gubler, D.J., 2011. Dengue, urbanization and globalization: The unholy trinity of the 21(st) century. *Tropical Medicine and Health* 39, 3–11.
- Rico-Hesse, R., 2003. Microevolution and virulence of dengue viruses. *Advances in Virus Research* 59, 315–341.
- Simmons, C.P., Farrar, J.J., Nguyen, V.V., Wills, B., 2012. Dengue. *New England Journal of Medicine* 366, 1423–1432.
- St. John, A.L., Abraham, S.N., Gubler, D.J., 2013. Barriers to preclinical investigations of anti-dengue immunity and dengue pathogenesis. *Nature Reviews Microbiology* 11 (6), 420–426.
- St. John, A.L., Rathore, A.P.S., 2019. Adaptive immune responses to primary and secondary dengue virus infections. *Nature Reviews Immunology*. doi:10.1038/s41577-019-0123-x.
- World Health Organization, 1997. *Dengue Haemorrhagic Fever: Diagnosis, Treatment and Control*. Geneva: World Health Organization, p. 58.
- World Health Organization (WHO), 2009. *World Health Organization (WHO) Dengue 2009: Guidelines for Diagnosis, Treatment, Prevention and Control*. Geneva: World Health Organization.

Ebola Virus (*Filoviridae*)

Andrea Marzi, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, United States
Logan Banadyga, Public Health Agency of Canada, Winnipeg, MB, Canada

© 2021 Elsevier Ltd. All rights reserved.

This is an update of K.S. Brown, A. Silaghi, H. Feldmann, Ebolavirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00662-2](https://doi.org/10.1016/B978-012374410-4.00662-2).

Classification

Ebola virus (EBOV) is the best known member of a relatively small but growing family of viruses known as *Filoviridae*, which itself belongs to the much larger order of negative-sense, single-stranded RNA viruses called *Mononegavirales* (Table 1). Within the family *Filoviridae* are five recognized genera—*Ebolavirus*, *Marburgvirus*, *Cuevavirus*, *Striavirus*, and *Thamnovirus*—each of which contains one or more species represented by specific type viruses. The genus *Ebolavirus* is composed of five species each with a single virus: *Zaire ebolavirus* (type virus: EBOV), *Sudan ebolavirus* (type virus: Sudan virus; SUDV), *Bundibugyo ebolavirus* (type virus: Bundibugyo virus; BDBV), *Tai Forest ebolavirus* (type virus: Tai Forest virus; TAFV), and *Reston ebolavirus* (type virus: Reston virus; RESTV). A sixth species of *Ebolavirus*, tentatively termed “*Bombali ebolavirus*” (type virus: Bombali virus; BOMV), and a sixth genus of *Filoviridae*, tentatively termed “*Dianlovirus*” (containing Mënglà virus; MLAV), have been proposed but have yet to be ratified by the International Committee on Taxonomy of Viruses. Informally, all viruses belonging to the *Filoviridae* family are referred to as filoviruses, while viruses belonging to the *Ebolavirus* genus are referred to as ebolaviruses.

Only ebolaviruses and marburgviruses are known to cause human disease, which is generally referred to as EBOV disease (EVD) or Marburg virus disease (MVD), replacing the outdated terms Ebola and Marburg hemorrhagic fever (EHF and MHF). Considering the severity of the diseases they cause, and the threat they pose to global public health, filoviruses, including EBOV, are considered Category A Bioterrorism Agents and Tier 1 Select Agents by the US Centers for Disease Control and Prevention and the Department of Health and Human Services, respectively. All work with filoviruses is therefore restricted to maximum containment (biosafety level 4; BSL-4) laboratories.

Epidemiology

Ebolaviruses were first identified in 1976 following two near-simultaneous outbreaks that occurred around Yambuku, Zaire (now the Democratic Republic of the Congo) and Nzara, Sudan (now South Sudan). Although these two outbreaks were originally attributed to a single virus—EBOV—years later it was determined that the etiological agent of the Sudanese outbreak was the highly related SUDV. Since then, four additional ebolaviruses have been discovered and dozens of outbreaks have occurred, mostly in Central Africa (Table 2). TAFV was discovered in 1994 following a single non-fatal infection in the Tai Forest reserve of Côte d'Ivoire, and BDBV was discovered in 2007 following an outbreak in the Bundibugyo region of Uganda. RESTV was first identified in 1989 in a group of fatally infected nonhuman primates (NHPs) that had been imported from the Philippines to Reston, Pennsylvania, and it appears to be apathogenic in humans. Recently, full genome sequences for a putative sixth ebolavirus, BOMV, have been identified in free-tailed bats from Sierra Leone, Guinea and Kenya, although, to date, infectious virus has not been isolated and no known human cases have been reported.

By far the largest and deadliest outbreak of any filovirus was that caused by EBOV in West Africa from 2013 to 2016. From a single index case in the village of Méliandou, Guinea, EBOV eventually spread throughout Western Africa to cause tens of thousands of cases. Guinea and the neighboring countries of Sierra Leone and Liberia bore the brunt of the epidemic, but several cases were exported to nearby African countries and numerous other countries around the world, including some European countries and the United States. Altogether the outbreak lasted nearly two and a half years and resulted in at least 28,616 cases and 11,310 deaths—although these numbers likely underestimate the true toll of the epidemic. In addition to the unprecedented number of human casualties, the West African EBOV epidemic also incited fear and anxiety around the world while at the same time devastating the public health infrastructure and economies of Guinea, Sierra Leone, and Liberia. Currently, the world's second largest filovirus outbreak—also caused by EBOV—is occurring in the North Kivu and Ituri provinces of the Democratic Republic of the Congo (DRC). At the time of writing, over 2000 deaths and 3000 cases have been recorded. Although these are obviously not the only outbreaks that have been caused by ebolaviruses, they are the worst, and they serve as a reminder of the ongoing threat posed by these viruses to global public health and biosecurity.

Ebolaviruses are considered to be zoonotic agents introduced into human populations either directly from the reservoir host or via intermediate hosts. The identity of this reservoir (or reservoirs) remains unknown; however, considerable circumstantial evidence suggests that this role may be fulfilled by bats. Antibodies specific to EBOV antigen, as well as EBOV-specific RNA sequences, have been identified in multiple species of bats (including *Hypsignathus monstrosus*, *Epomops franqueti*, and *Myonycteris torquata*), and epidemiological connections have been made between bats and index cases for certain outbreaks. Moreover, the complete genome of BOMV was isolated from bats (*Chaerephon pumilus* and *Mops condylurus*), and extensive lines of evidence

Table 1 *Filoviridae* family classification

Order	Family	Genus	Species	Viruses
Mononegavirales	<i>Filoviridae</i>	<i>Marburgvirus</i>	<i>Marburg marburgvirus</i>	Marburg Virus (MARV) Ravn Virus (RAVV)
			<i>Zaire ebolavirus</i>	Ebola Virus (EBOV)
		<i>Ebolavirus</i>	<i>Sudan ebolavirus</i>	Sudan Virus (SUDV)
			<i>Reston ebolavirus</i>	Reston Virus (RESTV)
			<i>Tai Forest ebolavirus</i>	Tai Forest Virus (TAFV)
			<i>Bundibugyo ebolavirus</i>	Bundibugyo Virus (BDBV)
			<i>"Bomali ebolavirus"</i>	Bomali Virus (BOMV)
		<i>Cuevavirus</i>	<i>Lloviu cuevavirus</i>	Lloviu Virus (LLOV)
			<i>"Dianlovirus"</i>	Mengla Virus (MLAV)
		<i>Striavirus</i>	<i>Xi'lang striavirus</i>	Xi'lang virus (XILV)
		<i>Thamnovirus</i>	<i>Huangjiao thamnovirus</i>	Huangjiao virus (HUJV)

Note: Quotation marks are used to indicate classifications that have not yet been ratified by the International Committee on Taxonomy of Viruses (ICTV).

Table 2 Ebolavirus outbreaks

Year	Location	Virus	Total Cases	CFR
1976	Yambuku, DRC	EBOV	318	88%
1976	Nzara, Maridi, Tembura, & Juba, South Sudan	SUDV	284	53%
1977	Tandala, DRC	EBOV	1	100%
1979	Nzara & Yambio, South Sudan	SUDV	34	65%
1989	Philippines	RESTV	3	0%
1989	Reston, Virginia ex Philippines	RESTV	0	0%
1990	Reston, Virginia ex Philippines	RESTV	4	0%
1992	Siena, Italy ex Philippines	RESTV	0	0%
1994	Ogooué-Inwindo Province, Gabon	EBOV	52	60%
1994	Tai Forest, Ivory Coast	TAFV	1	0%
1995	Kikwit, DRC	EBOV	315	81%
1996	Mayibout, Gabon	EBOV	37	57%
1996–1997	Booué, Gabon; Johannesburg, South Africa	EBOV	62	74%
1996	Alice, Texas ex Philippines	RESTV	0	0%
2000–2001	Gulu district, Mbarrara, & Masindi, Uganda	SUDV	425	53%
2001–2002	Ogooué-Inwindo province, Gabon; Cuvette region, RC	EBOV	124	78%
2002–2003	Ogooué-Inwindo province, Gabon; Cuvette region, RC	EBOV	143	90%
2003	Mbomo & Mbandza, RC	EBOV	35	83%
2004	Yambio county, South Sudan	SUDV	17	41%
2005	Etoumbi & Mbomo, RC	EBOV	12	83%
2007	Kasai Occidental province, DRC	EBOV	264	71%
2007–2008	Bundibugyo district, Uganda	BDBV	149	25%
2008	Philippines	RESTV	6	0%
2008–2009	Kasai Occidental province, DRC	EBOV	32	47%
2011	Nakisimata, Uganda	SUDV	1	100%
2012	Kibaale, Uganda	SUDV	24	71%
2012–2013	Luwero, Uganda	SUDV	6 ^a	50% ^a
2012	Isiro, DRC	BDBV	36 ^a	36% ^a
2013–2016	Guinea, Liberia, & Sierra Leone	EBOV	28,646 ^b	40% ^b
2014	Équateur province, DRC	EBOV	69	71%
2017	Bas Uélé province, DRC	EBOV	8	50%
2018	Équateur province, DRC	EBOV	54	61%
2018–present	North Kivu & Ituri provinces, DRC	EBOV	3346 ^c	66% ^c

^aLaboratory-confirmed cases only.

^bIncludes cases exported to Italy (1), Mali (8), Nigeria (20), Senegal (1), Spain (1), the United Kingdom (1), and the United States (4).

^cAccurate as of 19 December 2019.

Abbreviations: CFR, Case fatality ratio; DRC, Democratic Republic of the Congo; RC, Republic of the Congo.

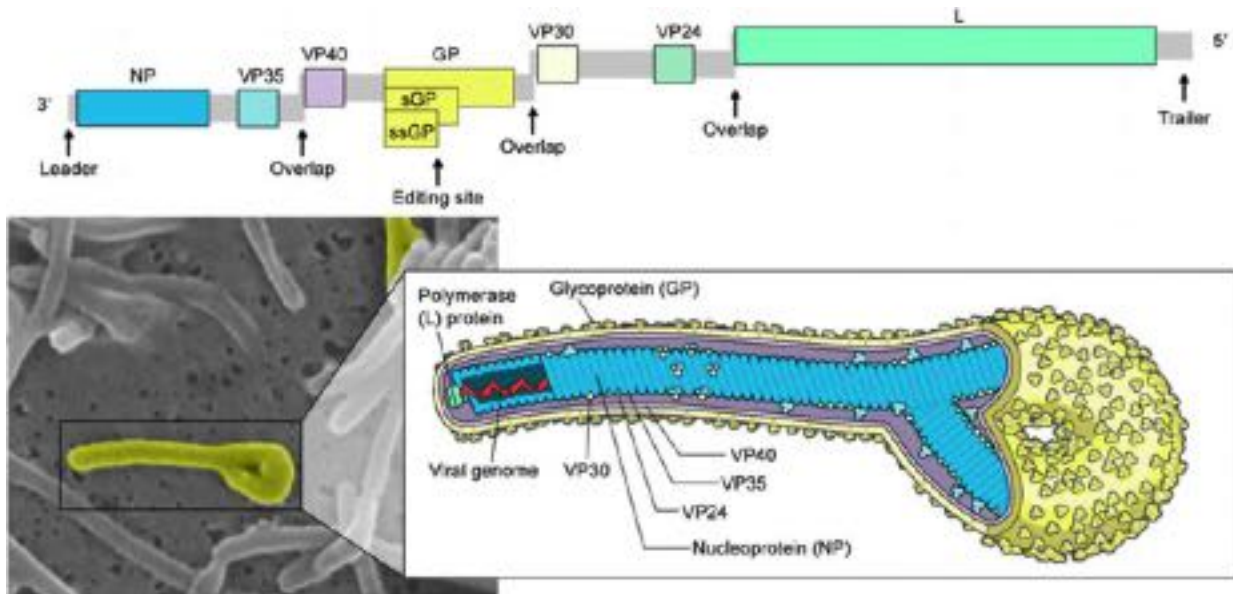


Fig. 1 Ebola Virus Genome Structure and Morphology. The Ebola virus (EBOV) genome is a single-stranded, negative sense RNA molecule, approximately 19 kb in length. The genome consists of seven genes, each indicated by a coloured box: nucleoprotein (NP), virion protein 35 (VP35), VP40, glycoprotein (GP), VP30, and the RNA-dependent RNA polymerase (L). Transcriptional editing of the GP gene at a conserved editing site (indicated by an arrow) produces two additional transcripts: soluble GP (sGP) and small soluble GP (ssGP). The grey regions of the diagram represent the 3' and 5' leader and trailer regions, respectively, as well as the untranslated and intergenic regions. Gene overlaps are indicated by arrows. A scanning electron microscope image depicts an EBOV virion, highlighted in yellow. A schematic depicts the structural proteins (NP, VP35, VP40, GP, VP30, VP24, and L), as well as the viral genome, that comprise the distinct, filamentous morphology of the virion.

suggest that the Egyptian rousette bat (*Rousettus aegypticus*) serves as a definitive reservoir for the related marburgviruses. Serological evidence of SUDV, TAFV, BDBV, and RESTV has also been reported in bats, although it is possible that cross-reactivity may complicate the specificity of these results. NHPs, such as gorillas and chimpanzees, along with other animals, such as duikers, are likely important dead end hosts. NHPs, in particular, are highly susceptible to ebolavirus infection and are common sources of bushmeat, the butchering and consumption of which has been linked to virus spillover into human populations.

The majority of ebolavirus outbreaks appear to cluster in distinct environments depending on the viral species, suggesting that different ebolaviruses occupy distinct ecological niches, a hypothesis that is supported by mathematical modeling. Moreover, the relatively low genetic diversity among different ebolaviruses provides further evidence suggesting strong ecological adaptation. Interestingly, although all human-pathogenic ebolaviruses (EBOV, SUDV, TAFV, BDBV) seem to exist near equatorial Africa, RESTV is found in the Philippines, suggesting that ebolaviruses may be more widely distributed than is currently appreciated. Indeed, serological studies have found evidence of ebolaviruses in Bangladesh and China, Lloviu virus (the sole member of the *Cuevavirus* genus) was identified in bats from Spain, and Mënglà virus (genus "*Dianlovirus*") was discovered in bats from Southern China. Ultimately, however, since the geographic distribution of ebolaviruses is directly related to the reservoir species, a complete understanding of ebolavirus epidemiology awaits the definitive identification of the reservoir host or hosts.

Genome

The EBOV genome, like that of all ebolaviruses, comprises a single, linear strand of negative sense RNA approximately 19 kb in length (Fig. 1). The genome contains seven genes, each of which encodes one of seven structural proteins: the nucleoprotein (NP), the 35-kDa virion protein (VP35), VP40, the glycoprotein (GP), VP30, VP24, and the RNA-dependent RNA polymerase "Large" protein (L) (Table 3). As a result of transcriptional editing, the GP gene also encodes two non-structural proteins: soluble GP (sGP) and small soluble GP (ssGP). A small 40 amino acid peptide, known as Δ -peptide is also generated by a post-translational cleavage of sGP (Table 3). Each gene is flanked at its 3' and 5' end by highly conserved untranslated regions (UTRs), which contain transcription start (3'-CUnCnUnUAAUU-5') and stop (3'-UAAUUCUUUUU-5') signals, respectively, and incorporate a conserved pentamer (UAAUU). For all ebolaviruses except RESTV, three intergenic regions of variable length exist between NP/VP35, VP40/GP, and VP30/VP24, while three small gene overlaps (where the start site of one gene occurs prior to the stop site of the preceding gene) exist between VP35/VP40, GP/VP30, and VP24/L. Conversely, RESTV lacks the gene overlap between GP and VP30. The genome itself is flanked by a conserved 3' leader sequence, approximately 60 base pairs in length, and 5' trailer sequence, which varies in length depending on the virus, ranging from as little as 25 bp in RESTV to as long as 676 bp in EBOV. The leader and trailer sequences, which possess a high degree of complementarity and are thought to form secondary structures, contain

Table 3 Major defined roles of ebolavirus proteins

	<i>Structural</i>	<i>Transcription/Replication</i>	<i>Immunomodulatory</i>
NP	Major component of the nucleocapsid; binds and protects viral genome RNA	Critical component of the viral polymerase complex	Not known
VP35	Critical component of the nucleocapsid	Polymerase cofactor; critical component of the viral polymerase complex	Binds and sequesters dsRNA; inhibits type I IFN gene expression
VP40	Major matrix protein; critical for virion assembly and budding	Negative regulator of transcription/replication	Not known
GP _{1,2}	Critical component of virion; required for virus attachment and fusion		Inhibits tetherin activity
sGP			Antibody decoy
Δ-peptide			Not known
ssGP			Not known
VP30	Component of the nucleocapsid	Critical transcription factor	Not known
VP24	Critical component of the nucleocapsid	Negative regulator of transcription/replication	Inhibits IFN gene expression and signaling
L	Component of the nucleocapsid	RNA-dependent RNA polymerase	Not known

conserved signals critical for genome replication and transcription. Importantly, because the genome of ebolaviruses—like all filoviruses, and, indeed, all mononegaviruses—is negative sense and therefore complementary to the mRNA sequences, it is not infectious. The viral polymerase complex is thus required to initiate replication and production of the (positive sense) antigenome prior to mRNA transcription.

Overall, the genomes of viruses from different species within the *Ebolavirus* genus differ from each other, on the nucleotide level, by approximately 23%–36% while the genomes of ebolaviruses differ from those of other filoviruses by greater than 50%. For the most part, these differences are recapitulated at the level of the GP gene, and highlight the diversity within the *Filoviridae* family (Fig. 2). Intriguingly, however, within a given species of *Ebolavirus*, different variants are highly conserved, typically displaying less than 5% divergence.

Virion Structure

Filoviruses derive their name from the Latin word *filum*, meaning “thread”, which accurately describes the overall morphology of the virion (Fig. 1). Ebolavirus virions, like those of all filoviruses, are long, filamentous particles that are often linear but can also bend into U-shaped, comma-shaped, or toroidal structures, indicating a high-degree of structural flexibility. Despite this flexibility, the EBOV virion maintains a uniform diameter of 96 nm, as measured from opposing outer edges of the viral envelope, and the length of a particle containing a single genome is approximately 980 nm long. However, approximately half of EBOV virions are polyploid, in which a single virion contains more than one copy of the viral genome, each arranged end-to-end continuously or separated from one another by a small gap. The result of this polyploidy is a population of exceptionally long virus particles of varying length, with some particles observed to contain as many as 22 genomes and measure over 21 μm in length. Empty virions containing no viral genome are also observed, possessing much smaller diameters (48–52 nm) and random lengths. Branched filaments are frequently observed in tissue culture supernatants, although the branches are typically empty.

Virion morphology is ultimately determined by the seven ebolavirus structural proteins, most of which compose some part of the helical nucleocapsid (Fig. 1). Indeed, the EBOV nucleocapsid is among the most complex of all mononegaviruses and consists of the virus genome in complex with five viral proteins: NP, VP35, VP24, VP30, and L. NP binds directly to genomic RNA in a sequence-independent manner and oligomerizes to form a left-handed helical, loosely-coiled structure that is then more tightly condensed via interactions with VP24 and VP35, which are thought to form a heterodimeric “bridge” along the outer edge of the helix. Although NP appears to form vertical contacts along the inner edge of the helix with neighboring NP monomers on adjacent helical layers, no such connections are apparent along the outer edge of the helix, thus permitting the nucleocapsid’s characteristic flexibility. Overall, the nucleocapsid has an inner-edge diameter of ~22 nm, an outer-edge diameter of ~41 nm, and a pitch of ~7 nm with 10.8 NP units per helical turn. Analysis of virion protein contents suggests that NP, VP35, VP24, and VP30 are all present at equimolar ratios, which is consistent with their involvement in forming repeating units of a highly ordered nucleocapsid. The role that VP30 plays in this structure remains unclear, however, since its presence or absence does not appear to affect nucleocapsid morphology. Likewise, although L is also considered to be a part of the nucleocapsid, it is present at such low copy numbers that it is unlikely to play a major structural role. VP40, the membrane-associated matrix protein, is not a part of the viral nucleocapsid, but instead serves to simultaneously drive the budding of nascent virions from the plasma membrane of the host cell and recruit nucleocapsids to the site of budding. In the virion, a layer of VP40 approximately 9 nm thick is thought to underlie the viral envelope, and VP40 extensions from this layer are thought to bridge the gap between the envelope and the nucleocapsid. In this way, VP40 keeps the nucleocapsid centered in the virion, with a spacing of 7–8 nm between the nucleocapsid and the

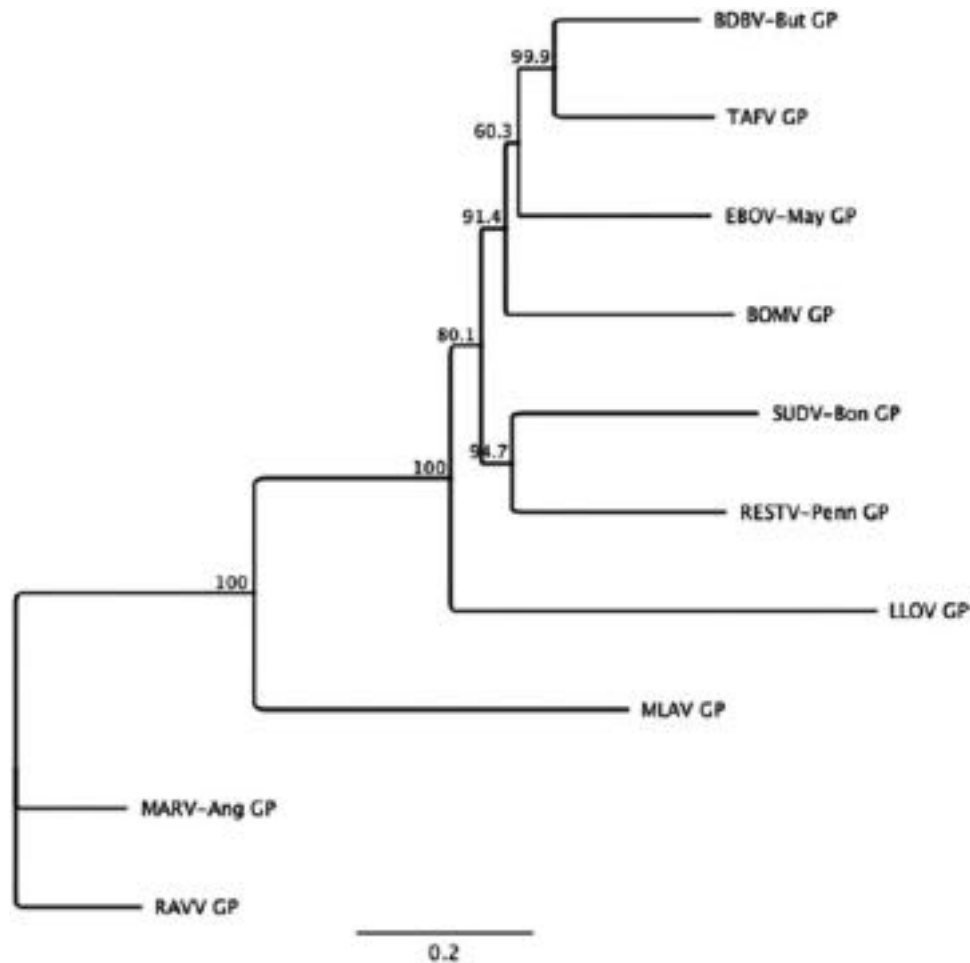


Fig. 2 *Representative Filovirus Phylogeny.* GP nucleotide sequences were aligned using MUSCLE, and a Neighbor-Joining phylogenetic tree was generated using Geneious Tree Builder. Percent consensus support is indicated at each branch, which was bootstrapped 1000 times. Genbank Accession Numbers: BDBV variant Butalya, KR063673.1; TAFV, KU182910.1; EBOV variant Mayinga, AF086833.2; BOMV, MF319185.1; SUDV variant Boniface, MH121162.1; RESTV variant Pennsylvania, AF522874.1; LLOV, NC_016144.1; MLAV, KX371887.2; MARV variant Angola, KU978782.1; RAVV, KU179482.1.

envelope. The outer surface of the viral envelope is studded with the heavily glycosylated viral glycoprotein (GP_{1,2}), the virus's only transmembrane surface protein and the sole structure responsible for receptor-mediated entry into host cells. GP_{1,2}, a class I fusion protein ~13 nm in length, is composed of a trimer of three identical disulfide-bound dimers consisting of proteins GP1 and GP2, which mediate receptor binding and membrane fusion, respectively.

Replication Cycle

Virus Attachment and Fusion

The virus replication cycle begins with virion attachment to a host cell (Fig. 3). Although ebolaviruses are thought to preferentially infect dendritic cells and macrophages early following infection, they have a broad cell and tissue tropism overall. Accordingly, with a notable exception of lymphocytes, ebolaviruses are capable of infecting many different cell types, including fibroblasts, epithelial and endothelial cells, and hepatocytes. Attachment to a host cell is mediated by at least two distinct mechanisms: (1) carbohydrate-binding receptors on the cell surface interacting with the heavily glycosylated GP_{1,2} and (2) phosphatidylserine (PS) receptors interacting with PS in the viral envelope. It is thought that the N- and O-linked glycans on GP_{1,2} interact promiscuously and non-specifically with a variety of carbohydrate-binding receptors, including C-type lectin receptors (such as DC-SIGN, L-SIGN, and LSECtin) and glycosaminoglycans, but that additional interactions between PS and PS receptors, such as those in the TIM (T-cell, immunoglobulin, mucin domain receptor) and TAM (Tyro3, Axl, Mer) families, are necessary for virus internalization. The broad distribution and redundancy of these attachment factors likely helps to explain the equally broad cell and tissue tropism

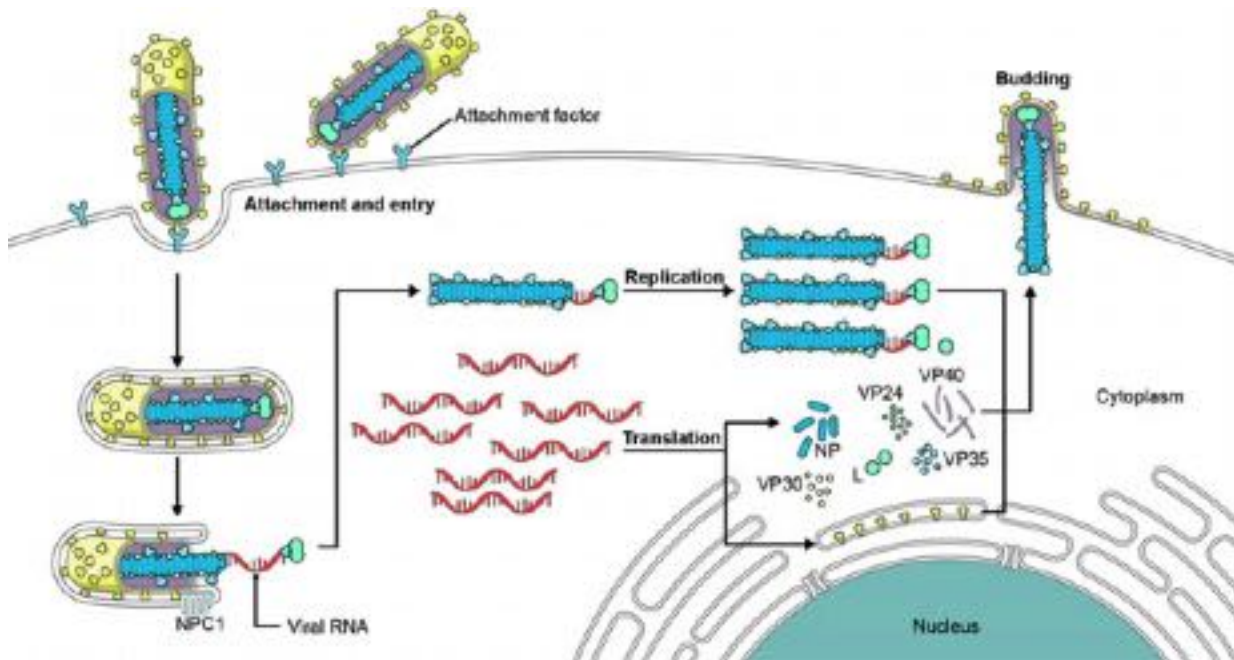


Fig. 3 The Ebola Virus Replication Cycle. Virion attachment to the host cell is mediated in part by non-specific interactions between EBOV GP and a number of different carbohydrate-binding factors, although interactions between phosphatidylserine in the virion envelope and phosphatidylserine receptors in the cell membrane also facilitate attachment. Following attachment, the virion is internalized by micropinocytosis, after which fusion between the viral and cellular envelopes is mediated by GP and Niemann-Pick C1 (NPC1), the cellular entry receptor for the virus. Primary transcription occurs following release of the NP-encapsidated viral genomic RNA into the cytoplasm and is mediated by the viral ribonucleoprotein complex (RNP) consisting of L, VP35, VP30. Host cell machinery translates the mRNA into protein. NP-encapsidated genome serves as a template for genome replication, carried out by L and VP35, via a positive-sense antigenomic intermediate. NP-encapsidated genomes are further encapsidated by VP35, VP24, VP30, and L, to generate highly ordered helical nucleocapsids that are trafficked to the cell membrane, where VP40 drives virus budding.

exhibited by ebolaviruses, and it also explains why a single surface-expressed cellular protein has yet to be identified as indispensable for virus entry.

Following attachment, virions are thought to be internalized primarily via macropinocytosis, although other cellular uptake pathways may play a role, depending on cell type and the attachment factors engaged (Fig. 3). The internalized virion is then trafficked through the progressively-acidic endocytic pathway where cleavage by the cysteine proteases cathepsin L and B removes the GP_{1,2} mucin-like domain and glycan cap, exposes the receptor binding site, and prepares the molecule for fusion-dependent conformational changes.

The cellular entry receptor for ebolaviruses was identified to be Niemann-Pick C1 (NPC1), a large, highly-conserved protein involved in cholesterol transport that is ubiquitously expressed and located on the interior membrane of late endosomes. NPC1 is critical for ebolavirus entry, and cells lacking this molecule are resistant to infection. Likewise, transgenic mice lacking the *Npc1* gene are resistant to infection. Proteolytically cleaved GP_{1,2} binds to domain C of NPC1 via the exposed receptor binding site, and through a process that is still incompletely defined, a series of conformational changes are thought to occur in GP_{1,2} that release the internal fusion loop and lead to fusion between the virion and endosomal membranes. In addition to the requirement for NPC1, the acidic environment of the late endosome/early lysosome is also required for fusion, and additional but unidentified endosomal factors may also be required. Ultimately, fusion of the virion and endosomal membranes releases the viral nucleocapsid into the cytoplasm, from where primary transcription commences.

Primary Transcription & Translation

Because the viral genome is in the negative sense and cannot be directly translated into protein, positive sense viral mRNAs must first be transcribed from the genome before the remainder of the replication cycle can take place (Fig. 3). The viral polymerase L binds to a conserved promoter in the 3' (Leader) end of the NP-encapsidated genome and moves successively towards the 5' end, initiating and terminating transcription at each gene's transcriptional start and stop sites, respectively. Newly transcribed viral mRNAs are capped by the viral polymerase and polyadenylated by a polymerase stuttering mechanism at the transcription start site. mRNAs are monocistronic and, unlike the virus genome, are not encapsidated by NP. Because a certain proportion of the polymerase falls off the genome at each transcription stop site, the probability of the polymerase transcribing a given gene decreases with the gene's distance from the 3' end of the genome. This results in a gradient of mRNA concentrations, with transcripts for NP being the most abundant and transcripts for L being the least abundant. Importantly, VP35 is an essential—but non-enzymatic—cofactor for the viral

polymerase that is thought to connect L to the NP-encapsidated genome by interacting with both proteins. VP30 is also required for transcription, particularly for activation of transcription from the leader sequence and re-initiation from each gene's transcription start sites, and its function depends heavily on its phosphorylation state. Along with L, NP, and the virus genome, these components form the ribonucleoprotein complex (RNP), and they are necessary and sufficient for ebolavirus transcription. Subsequent translation of viral mRNAs is completed by host cellular machinery and is thought to occur at distinct sites within the host cell.

One key feature of ebolavirus transcription that distinguishes it from that of the marburgviruses, is the use of co-transcriptional editing to produce multiple products from the *GP* gene. The primary product of the *GP* gene is the soluble glycoprotein (sGP), which is produced when the translational machinery encounters a stop codon roughly in the middle of the gene. Post-translational cleavage of sGP at its C-terminus produces the Δ -peptide, and both products are secreted from the infected cell. Thanks to a stretch of uridine residues before the stop codon, however, the viral polymerase occasionally stutters during transcription, resulting in the addition of a non-templated adenosine residue added to the mRNA. This additional residue results in a frameshift that allows translation to continue along the second half of the *GP* transcript, producing the pre-cursor protein to GP_{1,2}. Post-translational cleavage of this protein by furin-like proteases produces the GP1 and GP2 subunits, which then dimerize and trimerize to form the functional GP_{1,2}. Interestingly, the addition of two adenosine residues to the *GP* mRNA results in a different frameshift that creates a third stop codon ahead of the first two, thus producing the small soluble glycoprotein (ssGP). Like sGP and Δ -peptide, ssGP is also secreted from the cell. The function of these three products has yet to be clearly elucidated.

Replication

As transcription and translation continue, increasing levels of viral protein—including NP and VP30—are thought to trigger the switch to replication, although the details remain unclear. While dephosphorylated VP30 is required for the transcriptional activity of the RNP complex, phosphorylation of VP30 causes its dissociation from the RNP complex and shifts RNA synthesis towards replication. NP-encapsidated genome also serves as a template for viral RNA replication, which requires L and VP35. The polymerase complex binds to a bipartite replication promoter in the leader sequence of the genome and synthesizes a complementary positive-sense genome, referred to as the antigenome, which is immediately encapsidated by NP (**Fig. 3**). The antigenome then serves as a replication intermediate for the synthesis of additional copies of the viral genome. EBOV replication has been demonstrated to occur within peri-nuclear viral inclusion bodies, which appear to be dynamic and complex aggregates of viral proteins and RNA. Some evidence exists for the involvement of cellular proteins in replication/transcription, but further research is required.

Assembly & Budding

Although the structure of the mature viral nucleocapsid has been well studied, the dynamics of assembly are incompletely understood. Following transcription and replication, the NP-encapsidated viral genomes are further encapsidated by VP35, VP24, VP30, and L, which together generate the highly ordered, helical nucleocapsid described above (**Fig. 3**). Based on the ability of VP24 to inhibit transcription and replication in *in vitro* experimental systems, it has been proposed that the addition of VP24 to the nascent nucleocapsid represents the final step in nucleocapsid assembly, serving to condense the structure and restrict access to the genome by the viral replication machinery.

Mature nucleocapsids must then travel from the perinuclear region in which they are assembled to the plasma membrane, from where virus budding occurs. VP40, the matrix protein, is thought to be transported through the retrograde late endosomal pathway or in association with COPII vesicles to the underside of the plasma membrane, where it exists as a higher-order oligomer. Similarly, GP_{1,2} is transported to the plasma membrane via the ER-Golgi secretory pathway and collects in areas rich in VP40, likely due to interactions between the GP_{1,2} cytoplasmic domain and VP40. Actin-dependent transport mechanisms are thought to propel EBOV nucleocapsids along their long axis towards these sites on the plasma membrane, where interactions between NP and VP40 facilitate the envelopment of the nucleocapsid. Budding of the nascent virion from the cell is thought to require, at least in part, components of the endosomal sorting complex required for transport (ESCRT), which interact with two overlapping late domains present in VP40. Additional mechanisms may also be involved in virus budding, and the potential for direct cell-to-cell transmission remains a relatively under-investigated possibility.

Immune Evasion

The successful completion of the virus replication cycle, and the subsequent dissemination of infection, relies on the ability of the virus to evade the host's immune defences. Two of the most critical ebolavirus immunomodulatory proteins are VP35 and VP24. In addition to its roles in virion structure and transcription/replication, VP35 also functions to inhibit the induction of a type I interferon (IFN) response by binding to and sequestering double-stranded RNA (dsRNA) from intracellular nucleic acid receptors, such as RIG-I. Moreover, direct interactions between VP35 and components of the IFN-signaling cascade, including IKK ϵ , TBK1 and IRF7, also serve to limit the induction of the IFN response. VP24, on the other hand, binds to members of the NFI-1 subfamily of karyopherin- α proteins and prevents their translocation of STAT1 into the nucleus. Since STAT1 nuclear translocation is a critical step in the induction of IFN-stimulated genes (ISGs), VP24 effectively blunts the antiviral response triggered by type I IFN signaling. VP24 has also been shown to inhibit RIG-I-dependent IFN gene expression. GP_{1,2} also plays a role at the plasma

membrane in preventing the activity of tetherin, a viral restriction factor that interferes with virus budding, and sGP, may serve as a decoy molecule for the antibody response. Given the multi-functional nature of most ebolavirus proteins, additional immunomodulatory activities of other proteins likely exist, although further investigation is required.

Clinical Features

Prior to the West African epidemic, there was limited data available characterizing EVD, and, as a result, much of what we understood about this disease was obtained from animal models, usually in the context of EBOV infection. Different animal species, such as mice, guinea pigs, hamsters, ferrets, and NHPs have been used to model EVD and evaluate countermeasures—and each animal has its own set of advantages and disadvantages. Although inexpensive and widely accessible, immune-competent rodents are inherently resistant to EBOV infection, and virus must therefore be adapted through serial passaging in these animals to produce a lethal disease. Conversely, ferrets and NHPs are susceptible to wild type EBOV and exhibit disease remarkably similar to what is observed in humans. Rhesus and cynomolgus macaques, in particular, display a majority of the clinical signs of human disease, and they are considered the gold-standard model for research. Indeed, much of what is understood about EVD in humans was originally described in the NHP infection model. Our current understanding of human EVD is based predominantly on what we have learned from the West African EBOV epidemic, which infected more people than all previous EBOV outbreaks combined and produced a survivor cohort of thousands of people.

Early Disease

Like many other infectious diseases, the early stages of EVD include an asymptomatic incubation period followed by the onset of nonspecific signs of disease. These symptoms include fatigue, malaise and myalgia often coinciding with the onset of fever. Because these symptoms are not specific for EVD, the disease can be mistaken for other, more prevalent diseases in Africa, such as malaria, yellow fever, and Dengue. Although diagnosis of EVD in this early stage may be possible, it remains realistic, since the early symptoms precede the detectable presence of virus and viral RNA in the blood.

Peak Disease

Within the first week of illness, EVD patients become progressively weaker and develop nausea, vomiting, and diarrhea. Some cases also develop a maculo-papular rash and oozing from venipuncture sites or mucosal membranes. During this time, patients develop disseminated intravascular coagulation (DIC) along with thrombocytopenia. Patients are also extremely infectious, and the virus is present in all body fluids. If patients are not properly rehydrated at this point, loss of fluid from diarrhea and vascular leakage can result in severe dehydration and hypovolemic shock, ultimately leading to organ failure and death. Since EBOV replicates systemically during peak illness, almost all organs are affected. Liver damage is documented by an increase in liver enzyme levels, and an increase in creatinine levels is indicative of an acute kidney injury and subsequent renal failure. Viral antigens can be found systemically but is most abundant in the liver and spleen. In many EVD cases, central nervous system (CNS) dysfunction and life-threatening meningoencephalitis are observed.

Recovery

The survival of EVD is associated with a decline in viremia and the development of a life-saving adaptive immune response towards the end of peak disease. However, EVD survivors can develop sequelae, as documented in limited reports from previous outbreaks. The EVD epidemic in West Africa resulted in the largest cohort of EVD survivors to date, and many have been followed to document late-onset sequelae. The most commonly reported sequela is uveitis, which occurs in ~20% of EVD survivors and results in light sensitivity as well as blurred vision or temporary loss of vision. Other sequelae include hepatitis, myelitis, tinnitus, difficulty hearing and temporary hearing loss, cognitive difficulties, fatigue, muscle pain, and general weakness. These symptoms have been described to persist for one year or longer.

Virus Persistence

Until the epidemic in West Africa, EBOV persistence had only been anecdotally described from the outbreak in Kikwit, DRC in 1995. Now, there is increasing evidence that EBOV persists in immune-protected sites and can cause late manifestations of disease months and years after initial infection. Indeed, infectious EBOV has been isolated from the eyes, CNS, and male reproductive organs (mainly the testes) from EVD survivors. During the epidemic, there were a couple examples of sexual transmission of EBOV originating from male EVD survivors. These events were discovered and documented based on EBOV sequence data. There is limited data available on what triggers the reactivation of EBOV from immune-protected sites and causes a disease; however, this topic is the focus of ongoing research, and several animal models are being used to model and analyze this phenomenon.

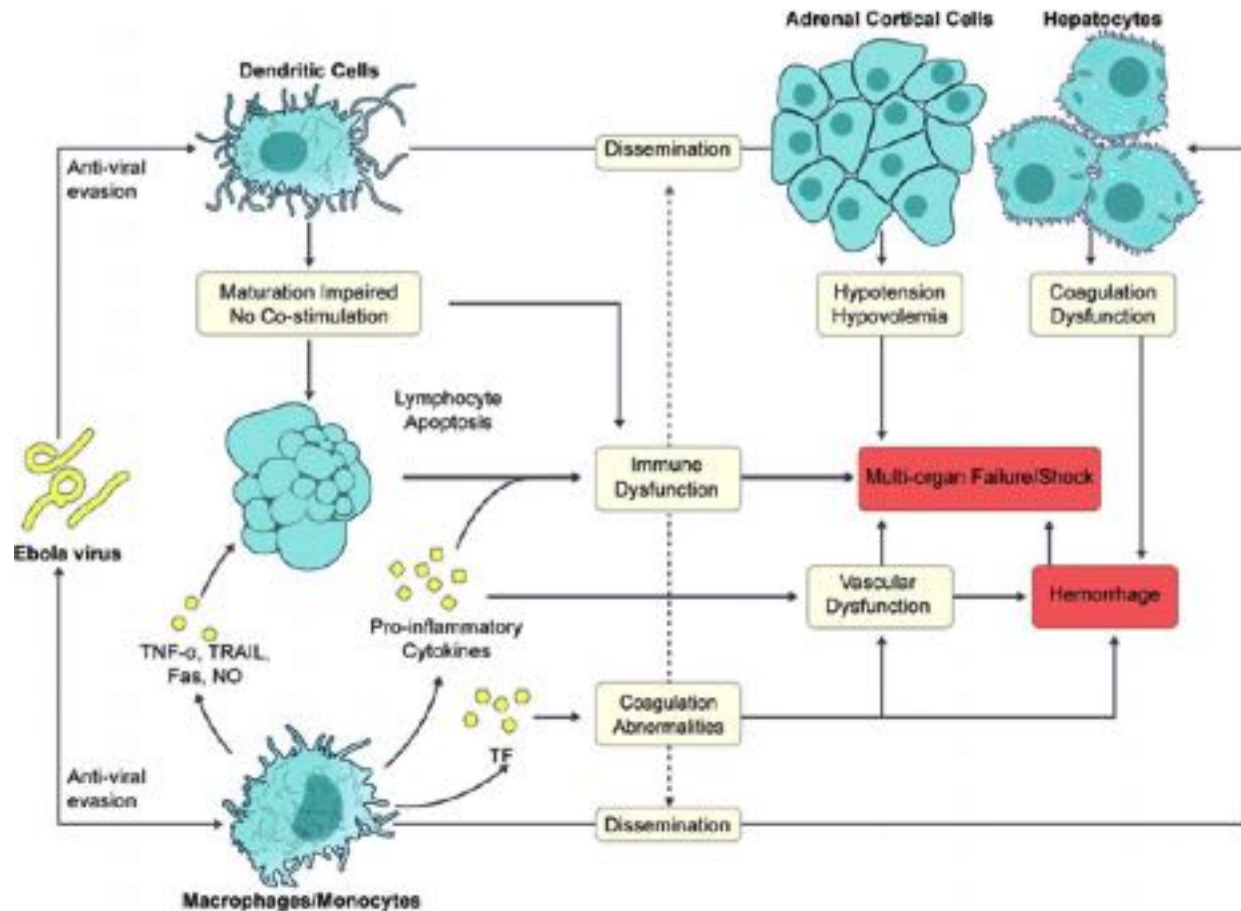


Fig. 4 *Ebola Virus Pathogenesis.* The primary target cells of Ebola virus infection are thought to be monocytes, macrophages, and dendritic cells. Infection of these cells not only permits virus dissemination, but also affects the immune response to infection, resulting in the production of inflammatory chemokines and cytokines as well as the impairment of maturation and lymphocyte co-stimulation. Immune dysfunction exacerbates virus dissemination and, along with rampant virus replication, contributes to organ damage. Damage to the adrenal cortex contributes to hypotension and hypovolemia, while damage to the liver contributes to coagulation dysfunction that, along with vascular dysfunction, can produce hemorrhage. Ultimately, death is precipitated by multi-organ failure and hypovolemic shock.

Pathogenesis

Most of the information available regarding EBOV pathogenicity is derived from infections in animal models, as well as human cases from the West African epidemic. Fatal EVD cases appear to be the result of systemic shock due to vascular dysfunction and leakage, which is caused by a complex interaction of the immune system with vascular physiology. There are three major processes that are thought to lead to hypovolemic shock and death: (1) increased vascular permeability; (2) DIC; and (3) impaired protective immune responses. The first two processes result from a series of events starting with the infection of primary target cells, namely monocytes, macrophages and dendritic cells (Fig. 4).

Infected macrophages are strongly activated—secreting pro-inflammatory molecules such as interleukin 6 (IL-6), tumor necrosis factor alpha (TNF)- α , IL-1 β , nitric oxide (NO)—and serve to spread the infection systemically. In contrast, infection of dendritic cells results in their impaired activation, with no upregulation of co-stimulatory molecules, such as MHC class I and II molecules, CD80, CD86, and CD40. Activation of endothelial cells by inflammatory mediators such as TNF- α and NO is believed to be the main cause for the decrease in their barrier function. TNF- α , NO, and pro-inflammatory cytokines increase vascular permeability and the expression of adhesion molecules on the surface of endothelial cells, which are necessary in extravasation of immune cells in inflamed tissue. Although these pro-inflammatory cytokines are an integral part of a normal, localized immune response by attracting and activating immune cells to the site of infection, in the context of a systemic and extensive response, they have a negative effect by inducing shock. Endothelial cells also serve as target cells for EBOV, and infection results in their activation—as indicated by the upregulation of adhesion molecule expression followed by cytolysis. However, the destruction of endothelial cells in vivo can only be observed at the end stage of the disease.

Infection of macrophages also results in the expression of tissue factor (TF), which can impair the anticoagulant protein-C pathway by downmodulating thrombomodulin. TF is believed to be an important protein in the development of DIC, since

inhibition of TF delays death in NHPs and offers limited protection. TNF- α can also induce TF expression in endothelial cells, thus amplifying coagulopathy and DIC. Moreover, coagulopathy is known to enhance inflammation, leading to a vicious cycle where inflammation enhances coagulopathy, which in turn amplifies inflammation. Rapid systemic viral replication is also an integral part of EVD pathogenesis. EBOV infection in humans and NHPs results in rapid massive viremia and high titers in many organs, especially in the spleen and liver. This correlates with an impaired innate and adaptive immune response.

Interestingly, high viremia and lethality are only seen in mice if a host-adapted virus or immune-deficient animals are used. These observations suggest that evasion of immune responses plays a pivotal role in pathogenesis. Indeed, mice lacking the interferon (IFN) α/β receptor or STAT1, the main signaling molecule in response to type I IFN, are extremely susceptible to EBOV infection and develop a lethal disease. Moreover, mouse-adapted (MA)-EBOV is less sensitive to IFN- α treatment in murine macrophages compared to wild type EBOV, and analysis of genomic mutations in MA-EBOV indicated that mutations in NP and VP24 are sufficient for a lethal phenotype. Although VP24 inhibits the interferon response in vitro, it remains unclear whether the mutations in MA-EBOV are required for type I IFN evasion or other functions. Nevertheless, while treatment of NHPs with IFN- α does not protect the animals, it does delay death, suggesting that the type I IFN response may be central to EBOV pathogenesis. EBOV is also able to inhibit IFN production through VP35, and deletion/mutation of the region in VP35 involved in its IFN antagonism results in a highly attenuated virus that does not cause disease in NHPs. Furthermore, comparative in vitro gene microarray analysis demonstrated a correlation between cytotoxicity and IFN antagonism, which was strongest with EBOV and weakest with RESTV.

Adaptive immune responses are also impaired during infection. As mentioned above, upregulation of co-stimulatory molecules is inhibited in dendritic cells, which reduces activation of T cells. There is also a dramatic reduction in the number of T and natural killer cells, despite the fact that lymphocytes do not support EBOV replication. It is believed that lymphopenia is caused by “bystander” apoptosis, although the mechanism is not well understood.

Diagnosis

Diagnosing EBOV infections in humans is difficult early on as the disease presents with symptoms common to many other infectious diseases. EBOV particles (viral antigens and viral RNA) and IgM antibodies specific to EBOV proteins are among the first parameters that can be detected in blood and serum samples generally at the same time as symptom onset. In the early outbreaks in Zaire (now DRC), Sudan and Gabon samples were shipped to Europe and the USA where antibodies were detected by ELISA and immunofluorescence; in addition, virus isolation was attempted. During the Kikwit, DRC outbreak in 1995, antigen detection was performed, and EBOV RNA was identified via conventional reverse transcription polymerase chain reaction (RT-PCR). The US CDC used both assays afterwards with good success in the field laboratories in rural areas. In the early 2000s, real-time RT-PCR assays using fluorescent probes were designed for the detection of EBOV from patient samples. Eventually, the real-time RT-PCR tests replaced all other tests since the application is manageable in the field, with a turn-around time of ~2–3 h from the sample receipt to a test result. Several laboratory-developed and commercial tests are currently being used, but only limited data is available regarding their specificity and sensitivity. These tests have also been developed to work on automated diagnostic systems for high-throughput capacity.

Other tests with a fast turn-around time include antibody-based lateral flow immunoassays designed for the detection of EBOV VP40 antigen or a combination of VP40, NP and GP antigens. Limited data is available regarding the sensitivity and specificity of these assays, and more testing is needed.

Treatment

Monoclonal Antibody Therapy

Shortly after the discovery of EBOV and SUDV, scientists started to develop potential treatment strategies. Convalescent serum therapy was among the first to be tested in animal models, but it demonstrated limited efficacy and was not developed further. However, during the EBOV outbreak in Kikwit, DRC in 1995 the treatment was used with a questionable success. Similarly, EBOV-specific equine IgG treatment was evaluated in NHPs with limited success. Despite these early setbacks, the scientific community continued to pursue antibody-based therapeutics. The human-derived monoclonal antibody KZ52 was tested in rodents with good efficacy but failed to protect NHPs from a lethal disease.

In 2012 it was demonstrated that purified EBOV-specific IgG from NHPs could protect naïve NHPs from a lethal disease when multiple doses were administered even after infection. Based on this report significant efforts were made to develop poly- and monoclonal antibody therapies. Two antibody cocktails, MB-003 and ZMAb, each consisting of three monoclonal antibodies, were developed approximately the same time. MB-003 used plant-produced human-mouse chimeric antibodies, whereas ZMAb comprised a mixture of mouse monoclonal antibodies. Three doses of each cocktail protected NHPs from the lethal disease depending on antibody amount and time of administration. Notably, a cocktail composed of two ZMAb antibodies and one MB-003 antibody resulted in what is now known as ZMapp and provided superior efficacy in NHPs. In 2014, during the West African epidemic, ZMapp was administered to two patients repatriated to the USA after they fell ill from EVD in Liberia, and it was

subsequently used on at least four additional patients evacuated from West Africa. ZMapp was also tested in phase 3 clinical trials towards the end of the West African outbreak, where it showed benefit compared to the standard care alone, although the results were not significant. More recently, a next-generation cocktail of three human antibodies, known as REGN-EB3, and a single human-derived monoclonal antibody, known as mAb114, have shown promising protective efficacy in NHPs. All three monoclonal antibody therapies are currently being used in clinical trials in the ongoing outbreak in the DRC, with an initial report showing superior efficacy and higher survival rates following a single intravenous dose of mAb114 or REGN-EB3 compared to three intravenous doses of ZMapp.

Besides antibody therapeutics, other approaches have also been tried to counter EVD, including small interfering RNAs (siRNAs), phosphorodiamidate morpholino oligomers (PMOs), small-molecule compounds, immunomodulatory approaches, and post-exposure vaccination. The most successful ones are explained in detail below.

Small Interfering RNAs (siRNAs)

The treatment potential of siRNAs against EBOV infection in NHPs was first described in 2010 when four intravenous doses of a cocktail of three siRNAs targeting VP24, VP35 and L packaged in a lipid nanoparticle protected animal from a lethal disease. The cocktail was further modified to contain siRNAs targeting L and VP35 and was termed TKM-Ebola. TKM-Ebola was administered to EVD patients evacuated from West Africa in 2014; however, because these patients received other care in addition to the siRNA, it was unclear whether the treatment was beneficial. The siRNA sequences were subsequently modified to match the outbreak strain (EBOV variant Makona), and this new product, now termed TKM-130803, was uniformly protective when seven doses were administered intravenously to EBOV-Makona-infected NHPs. In 2015, a small phase 2 clinical trial of TKM-130803 was conducted in Sierra Leone. Twelve patients received treatment, but only three individuals received the full seven doses and survived. Disease was too far advanced in the other nine patients, all of whom died before the remaining doses could be administered.

Small-Molecule Compounds

Favipiravir, also known as T-705, has broad antiviral activity against a number of RNA viruses. It has shown partial protective efficacy in preclinical NHP studies against EBOV when administered intravenously twice daily for 14 days starting two days before a lethal EBOV infection. Treatment success was dependent on the dose, with the highest dose resulting in 60% survival. The drug inhibited virus replication with the virus mutating over time. Even though this data was published after the EBOV epidemic in West Africa, clinical trials of favipiravir were still conducted during the epidemic, since this drug was the only one available with a fairly well-known safety profile in humans, oral bioavailability, and demonstrated efficacy in EBOV-infected mice. The drug was also used in patients evacuated to Europe and the USA. Unfortunately, despite the clinical trial data and data obtained from repatriated cases, a beneficial effect of favipiravir in EVD patients could not be demonstrated.

Remdesivir, also known as GS-5734, has been shown to prevent lethal EVD in NHPs when a relatively high intravenous dose of 10 mg/kg was initiated three days after infection and continued once daily for 12 days. The drug was used in a relapsed EBOV patient who had been evacuated to Europe and treated with other experimental drugs during the epidemic. Remdesivir is currently being used with good success in the ongoing outbreak in the DRC. A first report from the outbreak demonstrates that patients receiving an intravenous dose (200 mg) on day 1 followed by daily treatment with 100 mg until days 9–13, depending on virus load, had a ~50% survival rate. Interestingly, ~85% of people with a high viral load at the start of treatment survived, compared to 29% survival in patients with a low viral load. Although the efficacy of remdesivir is lower than those of mAb114 and REGN-EB3, remdesivir has the potential for combination therapy. No preclinical data is yet available for combined approaches.

Prevention

Treatment strategies are the best option during an outbreak; however, when an entire population of an endemic area needs to be protected from a disease, vaccination is the best preparedness strategy. For EBOV, vaccine development started shortly after the virus was discovered. Early vaccine approaches included inactivated virus and vaccinia virus-based vaccines. While these vaccines protected rodents from disease, NHPs were not protected. With additional funding becoming available in the early 2000s for countermeasure development against potential bioterrorism pathogens, efforts to develop an EBOV vaccine increased. Several vaccine platforms were developed most of which used EBOV GP_{1,2} as the antigen, including DNA-, protein-, and virus-like particle (VLP)-based vaccines, as well as virus-vectored vaccines based on alphavirus replicons, adenoviral vectors, modified vaccinia Ankara (MVA), human parainfluenza virus 3, rabies virus, and vesicular stomatitis virus (VSV). Although each of these vaccine approaches showed preclinical efficacy in NHP studies, there was limited interest from pharmaceutical companies to bring any of them to licensure. The situation changed in 2014, however, during the unprecedented West African epidemic, when most vaccination strategies with preclinical efficacy and an industrial partner for production were accelerated into human clinical trials worldwide. While these clinical trials involved several different vaccination approaches with demonstrated immunogenicity and safety in humans, this article will focus on the two most promising candidates that are currently being used in the ongoing

outbreak in the DRC. Notably, while China and Russia both licensed EBOV vaccines in 2018 (without human efficacy data), these vaccines are currently not being used in the DRC.

Adenovirus 26 (Ad26) & Modified Vaccinia Ankara (MVA)

The deletion of genes essential for replication in the adenovirus genome produces a replication-deficient particle that can be used as a vector for an antigen of choice, such as EBOV GP. Indeed, human adenovirus 5 expressing EBOV GP_{1,2} was first used to generate an EBOV vaccine, but it was shown to have limited efficacy due to pre-existing adenovirus immunity in the human population. In order to circumvent this problem, several other adenoviruses, including human adenovirus 26 and 35, as well as chimpanzee adenovirus 3, were developed expressing EBOV GP_{1,2}. Vaccination with an adenovirus 26-based vaccine (known as Ad26-EBOV) resulted in strong humoral immune responses and was able to protect NHPs from a lethal EBOV challenge. However, at least two doses were needed for protection and there were concerns with the longevity of the protective immune response. In order to address these concerns, a prime/boost strategy was implemented using Ad26-EBOV as a prime vaccination followed by a boost with the replication-incompetent MVA encoding GP_{1,2} (known as MVA-EBOV). This vaccination approach offered 100% protection in NHPs and has since been evaluated in several phase 1 human clinical trials, starting during the West African epidemic. The trials demonstrated a good safety profile and immune responses specific to EBOV GP. The Ad26-EBOV/MVA-EBOV vaccine approach is currently being used in the DRC in areas with no active EBOV cases to help increase the number of protected people with the hope to end the large ongoing outbreak in that country.

Vesicular Stomatitis Virus

In contrast to Ad26 and MVA, the VSV-based EBOV vaccine is replication competent. The VSV glycoprotein was removed in this vector and replaced with EBOV GP_{1,2}. Preclinical efficacy in NHPs was already demonstrated in 2005, yet no industry partner was found to produce human-grade vaccine until 2013. Further preclinical studies with NHPs demonstrated that a single dose of the VSV-EBOV vaccine is fast acting, protecting the animals within one week of vaccination; however, the longevity of the protective immune response remains unclear. During the West African epidemic, several phase 1–3 human clinical trials were conducted world wide, including a phase 3 trial in Guinea that demonstrated that a single dose of the vaccine was efficacious and prevented new infections. As this is a replication-competent vaccine, mild to moderate adverse effects after vaccination with a single dose of VSV-EBOV were reported. This vaccine has been used since the beginning of the ongoing outbreak in the DRC with remarkable success.

Remarkably, the VSV-EBOV vaccine was licensed (under its commercial name ERVEBO) in November 2019 by the European Medicines Agency and a month later by the US Food and Drug Administration. With the long-awaited clinical licensure of VSV-EBOV, as well as the licensing of separate vaccines in Russia and China and the anticipated licensure of Ad26-EBOV/MVA-EBOV, there is increasing hope that future EBOV outbreaks will be rapidly contained and no longer pose a threat to global public health.

Acknowledgments

We thank Ryan Kissinger (NIAID) for assistance with figure production. Filovirus research in the Immunobiology & Molecular Virology Unit, Laboratory of Virology is funded by the DIR, NIAID, NIH. This work was supported, in part, by the Public Health Agency of Canada.

Further Reading

- Agua-Agum, J., Allegranzi, B., Ariyaratne, A., *et al.*, 2016. After Ebola in West Africa-unpredictable risks, preventable epidemics. *The New England Journal of Medicine* 375 (6), 587–596. doi:10.1056/NEJMSr1513109.
- Banadyga, L., Dolan, M.A., Ebihara, H., 2016. Rodent-adapted filoviruses and the molecular basis of pathogenesis. *Journal of Molecular Biology* 428 (17), 3449–3466. doi:10.1016/j.jmb.2016.05.008.
- Banadyga, L., Wong, G., Qiu, X., 2018. Small animal models for evaluating filovirus countermeasures. *ACS Infectious Diseases* 4 (5), 673–685. doi:10.1021/acsinfecdis.7b00266.
- Baseler, L., Chertow, D.S., Johnson, K.M., Feldmann, H., Morens, D.M., 2017. The pathogenesis of Ebola virus disease. *Annual Review of Pathology: Mechanisms of Disease* 12, 387–418. doi:10.1146/annurev-pathol-052016-100506.
- Broadhurst, M.J., Brooks, T.J.G., Pollock, N.R., 2016. Diagnosis of Ebola virus disease: Past, present and future. *Clinical Microbiology Reviews*. 773–793. doi:10.1128/CMR.00003-16.
- Cross, R.W., Mire, C.E., Feldmann, H., Geisbert, T.W., 2018. Post-exposure treatments for Ebola and Marburg virus infections. *Nature Reviews Drug Discovery* 17, 413–434. doi:10.1038/nrd.2018.73.
- Feldmann, H., Feldmann, F., Marzi, A., 2018. Ebola: Lessons on vaccine development. *Annual Review of Microbiology* 72, 423–446. doi:10.1146/annurev-micro-090817-062414.

- Furuyama, W., Marzi, A., 2019. Ebola virus: Pathogenesis and countermeasure development. *Annual Review of Virology* 6 (1), 435–458. doi:10.1146/annurev-virology-092818-015708.
- Hoenen, T., Brandt, J., Cai, Y., Kuhn, J.H., Finch, C., 2017. Reverse genetics of filoviruses. *Current Topics in Microbiology and Immunology* 411, 421–445. doi:10.1007/82_2017_55.
- Messaoudi, I., Amarasinghe, G.K., Basler, C.F., 2015. Filovirus pathogenesis and immune evasion: Insights from Ebola virus and Marburg virus. *Nature Reviews Microbiology* 13 (11), 663–676. doi:10.1038/nrmicro3524.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/filoviridae
Filoviridae.

Enteroviruses (*Picornaviridae*)

Carita Savolainen-Kopra and Soile Blomqvist, National Institute for Health and Welfare, Helsinki, Finland
Petri Susi, University of Turku, Turku, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AFM Acute flaccid myelitis
AFP Acute flaccid paralysis
CNS Central nervous system
CSF Cerebrospinal fluid
CV Coxsackievirus
E Echovirus
EIA Enzyme immunoassay
EV Enterovirus
HFMD Hand-foot-and-mouth disease
IgG Immunoglobulin G
IgM Immunoglobulin M

IPV Inactivated polio vaccine
IRES Internal ribosome entry site
mRNA Messenger RNA
NCR Non-coding region
OPV Oral poliovirus vaccine
ORF Open reading frame
PCR Polymerase chain reaction
PV Poliovirus
RNA Ribonucleic acid
VP Viral protein
WHO World Health Organization

Glossary

Acute flaccid paralysis Any sudden weakness or loss of muscle tone.
Onychomadesis Shedding of the nails.

Recombination Process by which genetic material is broken and joined to other genetic material, genetic rearrangement.

Classification (Compact)

Enteroviruses form one of the largest group of viruses, which infects both humans and other species. Currently, there are more than 150 enterovirus types in 15 species. Most types infect humans but enteroviruses have also been identified in many mammals including bovines, pigs, monkeys, dromedary camels and rodents. Within species different types share greater than 70% amino acid (aa) identity in the polyprotein, greater than 60% identity in the capsid protein coding region (P1), and greater than 70% aa identity in the non-structural proteins 2C + 3CD. They also have a limited range of host cell receptors, a limited natural host range, a genome base composition within a 2.5% variation range as well as significant compatibility in proteolytic processing, replication, encapsidation and genetic recombination. Current EV typing system is based on genetic relationships of VP1 capsid protein.

Human enteroviruses are currently classified based on their genome organization, sequence similarity and biological properties into four species: Enterovirus A (EV-A), EV-B, EV-C and EV-D (**Fig. 1**), and include representative members such as polioviruses, coxsackie A and B viruses and echoviruses. The most recently identified types have been named according to numbering and species convention, such as enterovirus B-100. Each human enterovirus species consists of types; 25 EV-A types, 63 EV-B types, 20 EV-C types including 3 poliovirus types and 5 EV-D types (**Fig. 1**). Types can be distinguished by nucleotide sequence identity in the VP1 protein coding region. Identity of at least 75% (85% aa identity) between a prototype strain and an isolated strain suggests serotypical identity given that the next highest identity with an other prototype strain is less than 70%. Recent evolutionary studies indicate, however, that for e.g., CVA13, CVA20, E1, E13 and E30 nucleotide sequence identity within type may even be 73%, and thus amino acid identity should also be used as the basis of type determination. Within types enterovirus can further be classified into subtypes or genotypes with demarcation cut-offs varying between 9%–20% in the VP1 protein coding region. Genomic regions outside the capsid do not correlate with serotype identity due to frequent recombination in the non-structural parts of the genome.

The sequence relatedness in the P1 region coding for the capsid proteins correlates reliably with the traditional definition of a serotype. Antigenic groupings known as serotypes share similarities in immunologic response of the human host, protection from the disease, receptor usage and to some extent symptoms of clinical disease. Within a serotype isolates can be distinguished based on antigenicity measured with antisera raised in animals. Serotype still remains an important physical and immunologic property distinguishing different enteroviruses.

Most of the prototype strains of enteroviruses were originally isolated in the 1940–1950s. Development of molecular methods has facilitated detection of new enterovirus types during the past twenty years. In addition, molecular epidemiological studies have indicated the need for re-classification of some types, e.g., human rhinovirus 87 as EV-D68 and closely related rhinoviruses as three different species under the genus *Enterovirus*.

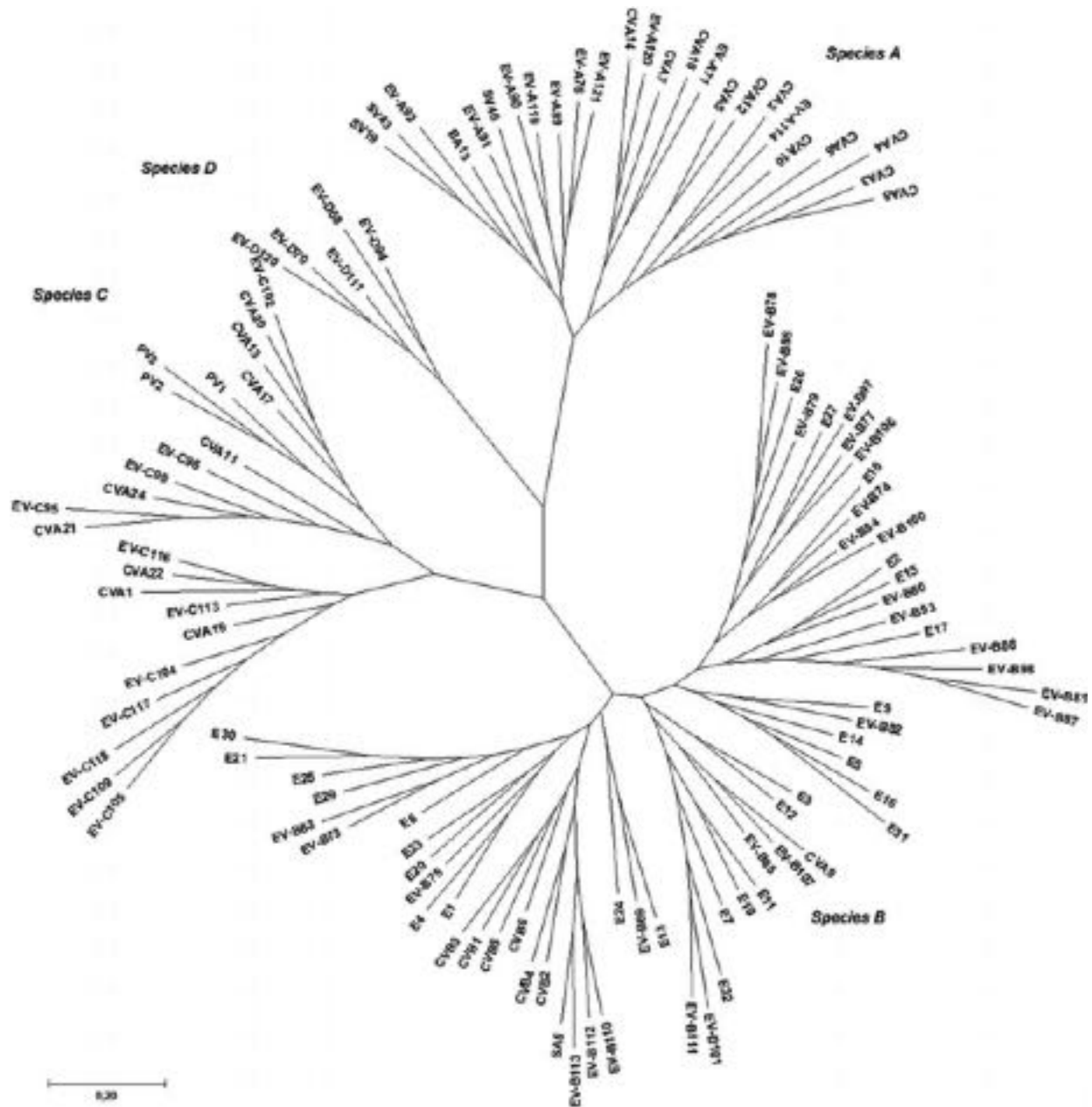


Fig. 1 Phylogenetic tree depicting genetic relationships of enterovirus types in *Enterovirus* species A to D. The evolutionary history was inferred by using the Maximum Likelihood method and Tamura-Nei model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 116 nucleotide sequences coding for the complete VP1 capsid protein region. Evolutionary analyzes were conducted using MEGA vs. 10.0.5.

Enteroviruses evolve through mutations and recombination events. Substitution rates within types are in the range of 0.41×10^{-2} to 3.07×10^{-2} substitutions per site per year. According to recent evolutionary analyzes times of the most recent common ancestors of individual enterovirus types vary between 55 and 200 years ago. Both inter- and intratype recombination events are common. Recombination occurs more frequently in the non-structural protein coding region and is restricted to viruses belonging to the same species. Interspecies recombination has been detected in the 5' NCR and VP4 junction.

Virion Structure

Enteroviruses were the first animal viruses for which structures with atomic detail were determined. Since 1985, when the X-ray structures of poliovirus and rhinovirus B-14 were determined, numerous high-resolution enterovirus particle structures with

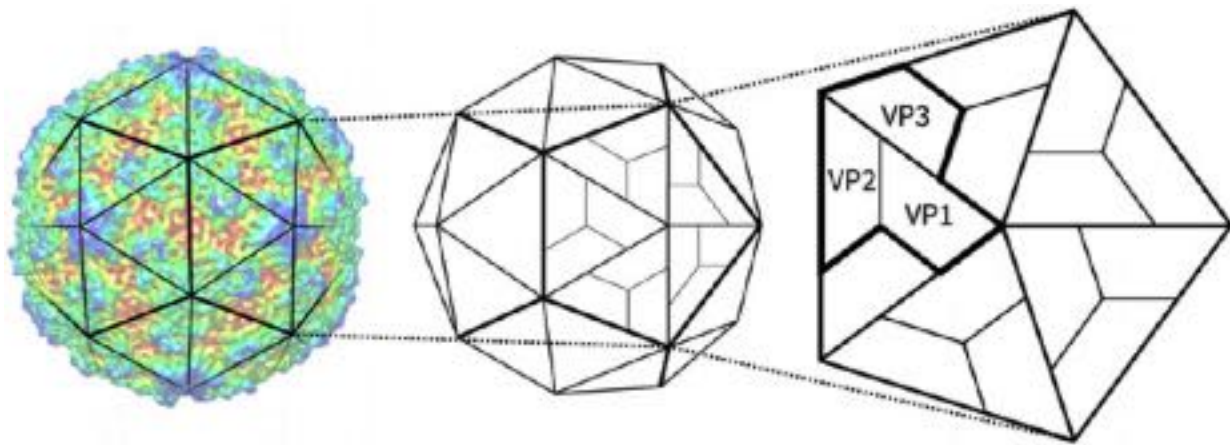


Fig. 2 Enterovirus structure. The capsid consists of 60 protomers, each consisting of four polypeptides that comprise the structural proteins VP1, VP2, VP3, and VP4 within capsid. VP4 faces the interior and is not visible on capsid surface. Cryo-EM model of coxsackievirus A7 (left). Reproduced from Seitsonen, J.J., Shakeel, S., Susi, P., *et al.*, 2012. Structural analysis of coxsackievirus A7 reveals conformational changes associated with uncoating. *Journal of Virology* 86, 7207–7215.

similar particle architecture have been resolved. Enterovirus particles are spherical with a diameter of about 28 nm and a density close to 1.34 g/ml, which is less than in most picornaviruses. Virus particles lack lipid membrane, which makes them insensitive to organic solvents and sturdy enough to ensure survival in harsh conditions in the gut where they replicate. Most enteroviruses are acid stable and retain infectivity at pH values lower than 3.0. However, virus particles are also flexible and able to undergo structural rearrangement during the entry and uncoating processes to release viral RNA.

Core particle features include small and smooth icosahedral structure and modularity of viral subunits with conserved Swiss roll domains. Enterovirus particles are constructed of 60 repeated protomer units, which are composed of four structural proteins, VP1 to VP4, encoded by proteolytic processing of long viral polyprotein (**Fig. 2**). The protomers form an icosahedral shell with a pseudo- $T = 3$ arrangement that encapsidates the viral RNA genome. External VP1 proteins are located around the five-fold axes, while VP2 and VP3 alternate around the 2- and 3-fold axes. These proteins have no sequence homology, but they share similar protein topology; each protein adopts eight-stranded, antiparallel β -barrel fold (Swiss roll) with C-termini located on the surface of the virion and N-termini facing the interior, which allows the formation of a dense and rigid protein shell. The main structural differences between VP1, VP2 and VP3 can be found in the loops that connect the β -strands and in the extended N- and C-termini. The amino acids within these domains define the morphology and antigenicity of the enterovirus types. VP4 is a short myristoylated capsid protein located in the interior of the virion. Viral RNA is tightly packaged within the capsid and in direct contact particularly with N-terminal amino acid residues of VP4. The N-termini of capsid proteins together with the myristate of VP4 have a major role in virus assembly and stability of the particle.

Most enteroviruses have a deep surface depression – also known as “canyon” – encircling the five-fold axes of symmetry that often serves as the cellular receptor binding site. The surface area of receptor binding in the canyon has been proposed to allow neutralization by neutralizing antibodies, which block viral infectivity, but also immune evasion of enteroviruses via mutations. Since VP1 is located around the five-fold axes and accounts most of the accessible surface, it is the most important target for neutralizing antibodies. VP1 is also the most variable and immunogenic particle protein. However, not all enteroviral receptors bind to the canyon structure suggesting that other mechanisms of neutralization (and immune evasion) exist. Indeed, several mechanisms for antibody neutralization besides interference with receptor binding has been proposed; virus stability, uncoating or virus aggregation may be affected by neutralizing antibodies.

The floor of the canyon within the VP1 β -barrel of all enteroviruses harbors a small, hydrophobic tunnel or pocket, which is in some enteroviral structures filled with a lipid moiety (for example, sphingosine, lauric acid or myristic acid) and likely to be involved in regulating particle stability. This pocket has been extensively studied for the development of novel antiviral substances. Several clinical trials have been performed with antiviral capsid binders, known as the WIN compounds (produced by Sterling-Winthrop and Janssen Pharmaceuticals). These molecules occupy the hydrophobic pocket and displace the pocket factor, thereby increasing particle stability and preventing receptor binding and/or genome release. However, antiviral effects *per* enterovirus type and clinical benefits have not been sufficient for clinical approval and use.

As of January 2019, there are 147 structure depositions for enterovirus structures in the VIPERdb (see “Relevant Websites section”) including native structures and structures of mutant viruses, viruses complexed with receptors, antiviral compounds, peptides and antibody fragments determined by X-ray crystallography or cryo-electron microscopy. Progress in high-resolution cryo-electron microscopy techniques is likely to increase the number of high-resolution structures in the near future and to enhance our understanding of the structural diversity within the family *Picornaviridae*. Expedited structure determination is also likely to facilitate research on antiviral capsid binders.

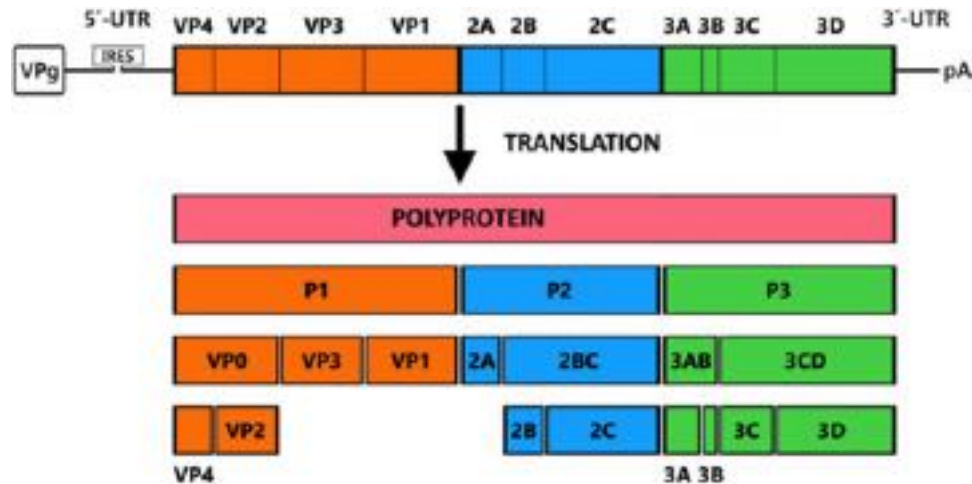


Fig. 3 Enterovirus genome organization and polyprotein cleavage patterns.

Genome

Enterovirus genome is a single positive-stranded RNA molecule with a length of approximately 7500 nucleotides (**Fig. 3**). Enteroviral RNA can function as a messenger RNA (mRNA) and it is infectious. At the 5' end of the genome there is a covalently linked VPg protein (virion protein, genome linked). VPg is not required for infectivity, but it is suggested to act as a primer for RNA synthesis.

The structure of the genome (**Fig. 3**) is common among many picornaviruses. The 5'-non-coding region (NCR) with an approximate length of 700 nt controls genome replication and translation. It harbors the internal ribosome entry site (IRES) that directs translation of the mRNA by internal ribosome binding. In the 3' end of the genome there is a short, less than 100 nt, 3'NCR. It has also a secondary structure, a pseudoknot, that controls viral RNA synthesis. The genome ends with a poly-A tail in both virion and mRNA. The poly-A tail has an effect on infectivity; the genome is non-infectious if the poly-A tail is removed.

Between the NCRs there is a single long open reading frame (ORF). The ORF is processed to form individual viral proteins. The gene expression is also similar in all picornaviruses including EVs. The polyprotein cleavage occurs during translation by virus-encoded proteinases. The polyprotein is divided in three regions: P1-3 (**Fig. 3**). P1 encodes viral capsid proteins, P2 and P3 encode non-structural or functional proteins. 2A^{pro}, 3C^{pro}, 3CD^{pro} are involved in protein processing and 2B, 2C, 3AB, 3B^{VPg}, 3CD^{pro}, 3D^{pol} in genome replication.

Life Cycle

To complete a successful infectious cycle, an enterovirus must proceed via several steps including attachment, internalization/entry, endocytosis, replication, particle assembly and exit (**Fig. 4**). Recognition of cell surface receptor(s) by virus particle determines the host range and tropism of cell and tissue infection and is the first step in the virus replicative cycle. Currently, there are more than ten different receptors known to mediate enterovirus infection, which may explain the numerous diseases that are caused by enteroviruses. These include e.g., CAR (coxsackievirus-adenovirus receptor), CD54 (intercellular adhesion molecule 1; ICAM-1), CD55 (decay-accelerating factor, DAF), CD155 (poliovirus receptor, PVR), integrin (e.g., $\alpha 2\beta 1$, $\alpha V\beta 1$, $\alpha V\beta 3$, $\alpha V\beta 6$), heparan sulfate and sialic acids. Cellular receptors are classified as true receptors (a cell surface protein or sugar moiety) that mediate internalization and/or genome release and co-receptor(s)/attachment factors (e.g., heparan sulfate) that act in the attachment of a virus to cell surface. For some enteroviruses, a single cellular receptor functions in both binding and entry (e.g., PVR) while other enteroviruses use several receptors in a coordinated fashion where binding to the "first" receptor leads to signaling cascade that allows the virus to bind to the true entry receptor (e.g., DAF-CAR).

Virus binding to the cell surface receptor may initiate a cascade of events leading to extensive irreversible structural changes (externalization of the N-termini of VP1 and the release of VP4) or have no effect on virus structure. Irrespective of this, most enteroviruses are internalized via receptor-mediated endocytosis mechanism (receptor itself is not known to be endocytosed) and transported to the cytosolic replication site in endosomal vesicles for viral RNA release instead of direct release of viral RNA into cell cytosol upon contact with the receptor. Nevertheless, the receptor usage is highly dependent on cell type, virus strain or even virus passage number (due to adaptation), and therefore, the actual role of any of these molecules in clinical infection needs further studies.

While the endocytic process has been analyzed in detail for only a few enteroviruses, it is evident that they can use several endocytic mechanisms. Previously, endocytosis was divided into clear-cut mechanisms, such as clathrin- or caveolin-mediated endocytosis, but more recently it has become evident that there are more exceptions than rules to enterovirus entry mechanisms. That is, the virus entry is defined by pathway-specific entry components. For example, poliovirus (PV) is internalized using actin

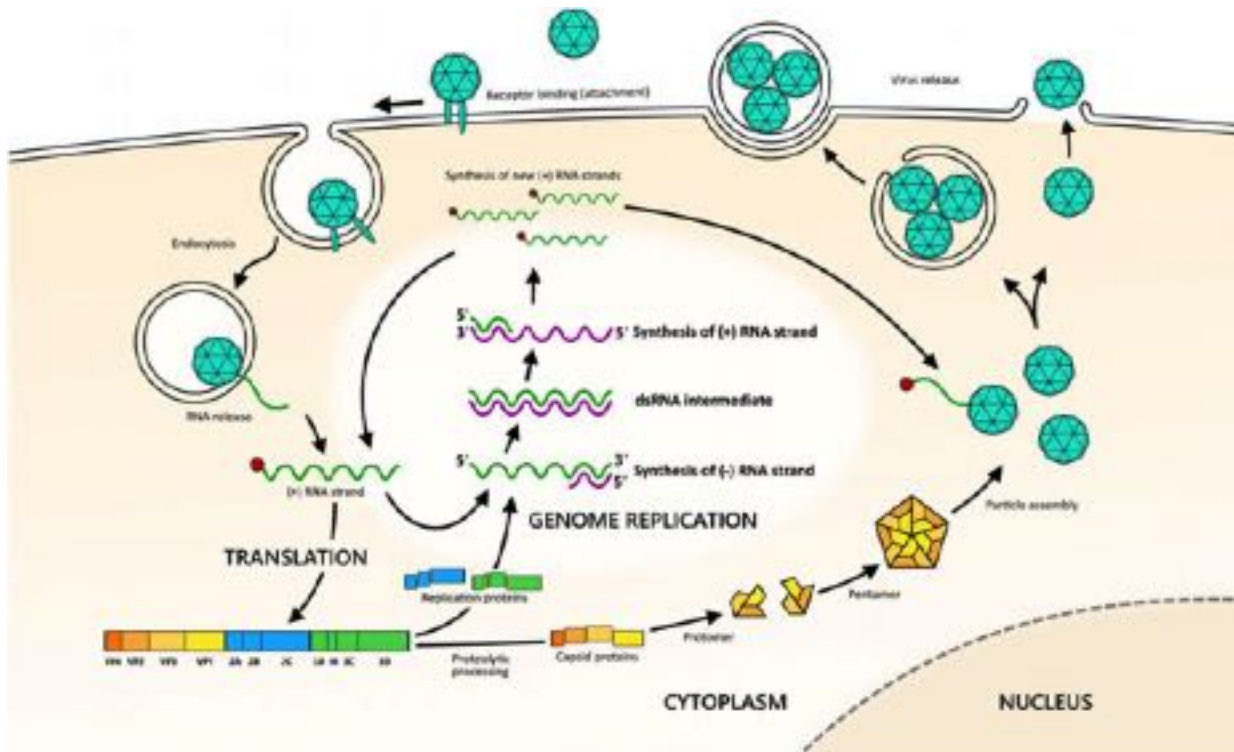


Fig. 4 Enterovirus life cycle. Following the binding to cell surface receptor(s), virus is endocytosed into the cell cytoplasm, where its positive-strand (+) RNA genome is released from viral capsid and used in translation and genome replication. Viral protein VPg (3B), which is used as primer in replication, is covalently linked to viral genome. Genome translation yields a single polyprotein that is proteolytically cleaved into replication proteins (2A–2C and 3A–3D) and capsid proteins (VP0, VP1 and VP3), which are used to make new viral capsids. Genome replication is mediated by viral RNA-dependent RNA polymerase (3D), which synthesizes negative-strand (–) RNAs that serve as templates for synthesis of new (+) RNA molecules. New (+) RNA molecules can either enter a new round of replication or be packaged into progeny virions. Capsid proteins self-organize into protomers and pentamers and assemble into provirions that are converted into infectious, mature virions by genome-induced cleavage of VP0 into VP4 and VP2. Mature virions exit the cell via non-lytic release of extracellular vesicles or via cell lysis.

and tyrosine kinases but not clathrin, caveolin or microtubules with rapid release of viral RNA in the periphery of HeLa cells. In contrast, in endothelial cells PV entry occurs slowly and is dependent on caveolin-1 and dynamin. There are also no rules for receptor-specificity and endocytosis. Echovirus 7 (E7) uses DAF as the primary receptor, but it is endocytosed via a clathrin-mediated pathway, while coxsackievirus B3 (CVB3) entry via DAF is caveolin-1-dependent. In addition, the entry process of CVB3 requires occludin but not dynamin. Echovirus 1 (E1) uses specifically integrin $\alpha 2\beta 1$ receptor in a process involving components of macropinocytosis, such as PKC, Pak1 and Rac1, and ESCRT proteins. Recent studies within *Enterovirus B* species have revealed remarkable similarities between virus types, which use a variety of cellular entry receptors, and it is apparent that many factors contributing to endocytosis are associated with macropinocytic uptake. These examples imply that enterovirus entry pathways are subject to virus type (and strain), cellular receptor and cell line used in the study.

Viral RNA release or uncoating follow the endocytosis to the site of replication. This may result in conformational changes in the virus particles, which are detected by sedimentation of virus particles with different molecular weights (160S for native particles versus 135S) in a differential ultracentrifugation run. According to some studies, the resulting 135S particles have exposed N-terminus of VP1 and lack VP4 but contain viral RNA. It has been hypothesized that hydrophobicity of VP1 N-termini increase the affinity of 135S forms to vesicular membranes, and as a result a pore through which viral RNA is released to cytosol is formed. The directionality of enteroviral RNA release during uncoating is unknown but it has been shown for a related rhinovirus (RV-A2) that the 3' end of the genome exits the virion before the 5' end. The site of RNA release from the virions is under debate and subject of virus type; the release may occur both at the two-fold, near the quasi-three-fold and five-fold axes depending on the cues used to trigger the uncoating process (such as heat, acid or soluble receptor molecules).

The mechanism of RNA replication is well conserved among all enteroviruses; it is mediated by an internal ribosome entry site (IRES), and requires several IRES trans-acting factors and a non-canonical replication machinery assembled on virus-induced ER-derived membrane structures enriched with phosphatidylinositol-4-phosphate and cholesterol lipids. These lipids regulate the activity of the replication machinery and its core enzyme, viral-encoded RNA-dependent RNA polymerase (3Dpol), and thus viral RNA synthesis. After delivery to the cytosol, viral RNA is directly translated by ribosomes into a single large polyprotein, which is proteolytically processed by viral proteinases, 2A^{pro}, 3C^{pro} and 3CD^{pro}, into functional proteins that mediate both genome replication and virus assembly. Genome replication by 3D^{pol} is initiated with the synthesis of negative-stranded RNAs of the viral genomic RNAs to generate double-stranded RNA

Table 1 Clinical disease entities associated to enterovirus types

<i>Disease</i>	<i>EV types most often detected</i>	<i>Comments</i>
<i>AFP/AFM</i>	PV, EV-D68, EV-A71	PV cases almost ceased. Other types associated include CVA7, CVA9, CVB1-B6, echoviruses, EV-D70
<i>AHC</i>	EV-D70, CVA24 variant	
<i>Encephalitis</i>	EV-A71	All EV types
<i>HFMD/herpangina</i>	CVA6, CVA10, CVA16, EV-A71	CVA1-A5, CVA8, CVA22 CVB types and echoviruses associated more rarely
<i>Meningitis</i>	Echoviruses, CVA and CVB viruses, EV-A71	All EV types
<i>Myocarditis</i>	CVB1–6	
<i>Neonatal sepsis</i>	CVB2-B5, E11	E5, E6, E16, E19 have also been associated to severe sepsis
<i>Pleurodynia</i>	CVB1–6	CVA types and echoviruses rarely
<i>Respiratory disease</i>	EV-D68, CVA21, CVA24	All EV types

replication intermediates. After strand separation, the negative strands serve as templates for the synthesis of new positive-stranded viral RNAs, which subsequently serve as templates for new rounds of replication or IRES-guided translation. Capsid proteins assemble into protomers and pentamers, which together with viral RNAs form mature virions. Thus, the particle assembly is directly coupled with the viral RNA replication. Besides cleaving the viral polyprotein, viral proteinases 2A^{Pro} and 3C^{Pro} cleave several host proteins to optimize virus translation, replication and virus spread and to suppress antiviral cellular responses. This leads to morphological alterations and ultimately virus particle release by disruption of cellular membranes. However, enteroviruses can also egress non-lytically in phosphatidylserine enriched, non-canonical autophagosome-like vesicles. This exit strategy may provide a simultaneous (masked) route of transmission of multiple genomes to uninfected cells, which enables the virus to maintain the viral genetic integrity and evade the immune system.

Epidemiology

Enterovirus infections are very common. Age, socioeconomic status and sex have an effect on the frequency and outcome of the infection. Enteroviruses are seen in people of all ages, but children are most prone due to lack of protecting antibodies from previous infections. Also the severity of infection, variety of clinical symptoms and prognosis varies between different age groups with children more likely to present with significant clinical symptoms. Encephalitis and aseptic meningitis are most frequently seen in children 5–14 years of age. Encountering the first enterovirus infection at an older age increases the severity of symptoms. Enterovirus infections are more prevalent among people with low socioeconomic status as well as in those living in urban areas. Males, in general, are more prone to enteroviral disease and severe infection than females.

Enterovirus infections typically peak during summer and early autumn in temperate climates. In tropical environments enterovirus infections occur all year round or they are associated with rainy seasons. Enteroviruses are isolated from both lower and upper gastrointestinal tract and are readily excreted to feces. They survive for extended periods, i.e., months, in the environment, e.g., sewage, in favorable conditions. Neutral pH, moisture, low temperature and presence of organic matter protect enteroviruses from inactivation.

The main form of transmission of enteroviruses is via fecal-oral route; from respiratory/upper gastrointestinal tract to the gut and from there to the brain. Also, respiratory transmission occurs for certain types, e.g., EV-D68, and to a lesser extent, mother-to-infant transmission. Hand contact of upper respiratory mucosa (mouth, nose, conjunctiva) may facilitate transmission. Transmission occurs readily within households, neighborhoods and communities. In addition, nosocomial transmission has also been reported, especially among newborns. Humans are thought to be the main reservoir of enteroviruses. Simian enteroviruses are genetically closely related and human enteroviruses have been detected in free-living non-human primates. In addition, swine vesicular disease virus is a porcine variant of CVB5.

Based on seroprevalence studies neutralizing antibodies are readily found in different populations indicating high incidences of past infections. Some enterovirus types are considered to remain endemic in certain areas, e.g., EV-B type E11 in Europe, whereas other viruses cause outbreaks with 5–10 year intervals, e.g., E30. Nucleic acid sequencing and molecular epidemiology have been widely used as a tool for the detection of genetic relationships of enteroviruses and identification of epidemiological patterns. Molecular epidemiology has provided information on geographic and temporal origin of the strains to strengthen epidemiologic links, to observe similarities among isolates in an outbreak and differences among isolates in different geographic regions.

Clinical Features

Most enterovirus infections are clinically unnoticeable. When symptoms occur, the prevailing clinical presentations are common cold, non-specific febrile illness and/or rash, which are indistinguishable from diseases caused by other viruses. However,

enteroviruses can cause a variety of more specific clinical outcomes, e.g., hand-foot-and-mouth disease (HFMD), meningitis, encephalitis, acute flaccid paralysis/myelitis, conjunctivitis, sepsis-like disease in newborns and myocarditis (Table 1). Different enterovirus types can cause similar diseases during the same geographically restricted outbreak and, vice versa, one type can cause a variety of clinical symptoms.

Hand-foot-and-mouth disease is the most significant form of clinical enterovirus infection globally. The WHO defines HFMD as a febrile illness with papulovesicular rash on palms and soles with or without vesicles/ulcers in the mouth. In herpangina, the patient has febrile illness with multiple oral ulcers on the posterior parts of the oral cavity. Causative agents are species A enteroviruses and usually multiple different types co-circulate in HFMD outbreaks. The incidence of CVA16 is high world-wide. CVA6 emerged in 2000s and alternates with CVA16 as the most common enterovirus in outbreaks. EV-A71 has caused extensive HFMD outbreaks especially in the Asia-Pacific region during last decades.

In most cases, HFMD and herpangina are self-limiting diseases, recovery occurs within 1–2 weeks after onset, and skin blisters heal without scarring. Onychomadesis is a recently reported specific consequence of CVA6 infection. Severe HFMD complications are rare, but occasionally central nervous system (CNS) is involved, particularly in children under 5 years of age, and when the causative agent is EV-A71. During recent EV-A71 outbreaks in Asia, 10%–30% of hospitalized HFMD patients developed CNS complications such as meningitis, brainstem encephalitis, encephalomyelitis and/or acute flaccid paralysis. The most severe EV-A71 cases have acute refractory myocardial dysfunction and fulminant cardiorespiratory failure with relatively high mortality rate. The survivors are in high risk for long-term sequelae, which range from limb weakness to ventilator dependence and presence of multiple physical and psychobehavioral disabilities. The severity of the disease depends on multiple factors, including age of the patient, pre-existing immunity, co-infections and the type and (sub-) genotype of the outbreak virus strain. The susceptibility to a severe EV-A71 infection may be associated with certain HLA types of the host as well.

Enteroviruses are the most important causes of meningitis, an inflammation of the tissue covering brain and spinal cord. Individuals of all ages can contract meningitis, but most prone are children younger than 5 years of age and people having impaired immune system. Typical symptoms are headache, fever, muscle and stomach aches, stiffness in the neck, photophobia and phonophobia. The prognosis is usually good and the infection is cleared in one to two weeks. The prominent causative enterovirus types are echoviruses and coxsackie B viruses. Large outbreaks may occur in the summer-fall season in temperate climates. Encephalitis is a more severe presentation of CNS infection. The direct inflammation of brain can be caused by many enterovirus types, but it is mostly associated with EV-A71.

All enteroviruses can cause respiratory symptoms. The disease spectrum includes acute upper respiratory infection, pneumonia, bronchiolitis, wheezing, respiratory distress and asthma exacerbation. Enteroviruses have also been detected in the middle ear fluid in young children at the time of acute otitis media. EV-D68 has long been associated with respiratory diseases and, at least some of the strains, share biological properties, like acid-sensitivity, with human rhinoviruses. Clinical isolates of EV-D68 were rare up to 2005. Since then, EV-D68 has caused wide outbreaks with considerable proportion of severe respiratory infections. Especially young children may require hospitalization and intensive care due to bronchiolitis or pneumonia. In rare cases, and mostly in young children, EV-D68 infection has been lethal.

Acute flaccid paralysis (AFP), is a case definition used in poliovirus surveillance, meaning any sudden weakness or loss of muscle tone. Acute flaccid myelitis (AFM), is a more specific definition for a paralytic case with lesions in the brainstem and spinal cord gray matter, which are detectable by magnetic resonance imaging. CSF findings in enterovirus AFM are non-specific, lymphocytic pleocytosis is mild to moderate, glucose is normal and protein levels are normal or elevated. Virus is not usually detected in CSF. In the past, the most important AFP/AFM causing enteroviruses were polioviruses. In the 2010s, when poliovirus eradication is almost complete, most of the AFP/AFM cases are caused by other, non-polio enteroviruses. During EV-D68 outbreak in the US and Canada in 2014, the incidence of AFM followed that of EV-D68 respiratory disease. EV-D68 was detected in CSF in only few cases, but more often in the respiratory specimens on the onset of AFM. The full recovery occurred in only a small percentage of the EV-D68 AFM patients, and the majority had residual weakness still one year after the onset. In EV-A71 HFMD cases, the onset of paralysis may occur in 2–6 days after the first symptoms.

Newborns are susceptible to severe enterovirus diseases, although the majority of enterovirus infections are asymptomatic or mild in neonates as well. The most severe presentation of neonatal enterovirus disease is septic shock with multiorgan failure. The disease can include specific manifestations like pneumonia, hepatitis, pancreatitis, myocarditis, meningitis or encephalitis. Mortality is highest in neonates with both hepatitis and myocarditis. Neonates with symptomatic enterovirus infection usually have mother or other contact with recent symptomatic illness. Perinatal transmission is also possible. Severe disease has been associated especially with CVB2 – CVB5 and E11.

Acute hemorrhagic conjunctivitis (AHC), first detected in 1969, is a highly contagious infection of eyes. EV-D70 and a variant of CVA24 are the leading causative agents of AHC and have caused pandemics as well as smaller outbreaks world-wide. In addition to general symptoms, like fever, sore throat and headache, the infection results in eye irritation, photophobia, discharge and subconjunctival hemorrhage, which ranges from discrete petechiae to large patches of hemorrhage covering the bulbar conjunctiva. Recovery is usually complete within less than 10 days without ocular sequelae. During EV-D70 pandemics, neurological symptoms were reported in about 1 out of 10 000 AHC patients.

Viral myocarditis, pericarditis and myopericarditis are rare consequences of EV infections, usually caused by coxsackie B viruses. Bornholm disease or epidemic pleurodynia, also known as epidemic myalgia, is caused by coxsackie B viruses. The typical symptom of Bornholm disease is acute, severe pain in the chest, often only on one side. Other symptoms include headache, fever, sore throat and general malaise.

There is also evidence that some enteroviruses may trigger onset of chronic diseases. Some coxsackie B viruses and echoviruses are associated to the onset of type 1 diabetes, but the exact mechanisms are not yet fully understood.

Pathogenesis

Polioviruses were historically the model of enterovirus pathogenesis research. With the success of the on-going poliovirus eradication campaign (see “Relevant Websites Section”) pathogenesis research of other enteroviruses has gained increasing interest. The most current research on enterovirus pathogenesis is focused on coxsackie B viruses, particularly CVB3, and enterovirus A71 (EV-A71) in Asia. The prevalence, modes of transmission, genetic variability (quasispecies), spectrum of tropism, specific features in cellular life cycle and ability to cause both lytic and persistent infections in the human host make it challenging to not only recognize but also link prior enteroviral infections and subsequent chronic syndromes.

Primary infection occurs via respiratory or gastrointestinal tract mucosa. Viremia infection occurs within a day after contact with the virus enabling virus dissemination to multicellular organs, and lasts for 7–14 days until it is cleared by virus-activated adaptive immune response. Enteroviruses also replicate in actively proliferating cells, which are evident during organ development in the young. In addition, infectivity in the young may be related to tropism, i.e., expression of virus-specific cellular receptor in target cells, and more specifically to age-related expression of virus receptors, such as CAR in case of coxsackie B viruses. Enteroviruses have been shown to infect muscle, heart, pancreas and CNS causing respective diseases (see Section Clinical Features), some of which can be fatal or result in long-lasting organ dysfunction. CVB3 has tropism to heart and pancreatic tissues and is thus causative agent of virus-induced myocarditis and pancreatitis while EV-A71 is considered to be a serious emerging CNS pathogen.

Since EVs are transmitted via fecal–oral route of infection, a detailed understanding of how the virus travels from the gut to the brain is essential in order to understand the pathogenesis. It is possible that during viremia virus passively passes blood-brain-barrier (BBB) independent of receptor interactions. However, it is more likely that virus dissemination between organs is at least partially defined by tissue tropism. The current model favors the idea that virus infects CNS using “Trojan horse” tactic by infecting myeloid stem cells, which mask the virus from the immune system and enable long-distance dissemination from the gut to CNS.

At cellular level the most prominent feature of EV infection is cytolysis, i.e., ability to cause direct cell lysis following virus replication. In cells, EVs shut-down host protein synthesis and inhibit host cell protein secretion by virus-specific 2A and 3C proteases and 2B and 3A proteins, respectively. In addition, enteroviruses benefit from a cellular degradation process known as autophagy, which is known to play an important role particularly in preventing cellular damage in neurons. EVs use autophagosomal membranes as scaffolds for viral replication, which eventually leads to cytopathic cell lysis. It has also been suggested that EV proteins possess pro-apoptotic properties promoting cell death.

Enteroviruses are known to persist (i.e., to cause infection in which virus is not cleared but remains silent in specific cells) up to 90 days with occasional activation of virus with disease symptoms. The genetic mechanism by which enteroviruses persist is apparently related to the RNA genome; enteroviral genomes with the first 49 nucleotides deleted have been detected in samples from human heart and in murine model of chronic myocarditis. This results in virus variants that replicate slowly and express very low levels of genomic RNA. As a consequence, enteroviruses that do not induce cytolysis are generated in barely detectable amounts. During the persistent infection viral 2A protease is expressed. Viral 2A protein has been shown to cleave dystrophin, which is a vital heart protein that supports muscle fiber strength. Cleavage of dystrophin is thought to contribute to dilated cardiomyopathy. It is not known whether autophagy contributes to viral persistence. CVB3 infection in CNS tissue may be linked to its ability to establish persistent infection only in such target stem cells, which support sporadic reactivation whenever they become activated to generate downstream progenitor cells. Together with cellular effects, this data may explain why enterovirus is the most important causative agent of aseptic meningitis.

Despite of vast number of cellular studies and epidemiological studies with phylogenetic analyzes, relatively little is known about the determinants of enteroviral virulence. In general, there are not many publications where a single amino acid mutation is linked to changes in disease symptoms. This is greatly related to quasispecies nature of enteroviruses, which makes it difficult to identify a uniform virus (population) with a disease-related point mutation. Nevertheless, differences may lie in either NCRs, capsid or functional regions. For example, CVB4-VP1 gene has been linked to CVB4-induced pancreatitis; VP1 mutation in a virulent strain increases both virus amounts and immunogenicity enhancing disease symptoms. Studies with chimeric viruses (generated by the exchange of fragments between the attenuated and the virulent strains) have suggested that the 5′-NCR plays a major role in poliomyelitis and myocarditis, latter of which is directly related to enterovirus persistence with 5′-terminal deletions. However, it is evident that many genomic regions contribute not only to symptom development in myocarditis but also to other disease phenotypes.

Diagnosis

Enterovirus diagnosis is based on the detection of enterovirus or its genome from appropriate clinical specimens. Clinical diagnosis is rarely adequate enough because enterovirus symptoms are indistinguishable from other viral or bacterial diseases. Only in well-described community outbreaks, where circulation of a particular enterovirus type is confirmed, the diagnosis may rely solely on clinical and epidemiological findings.

Stool specimens and rectal swabs are the most sensitive for enterovirus detection. Viruses are excreted in stools in large concentrations, and this may last few weeks or even months after the recovery. In AFP/AFM, two stool specimens should be collected and analyzed for poliovirus. Respiratory specimens are a good choice especially in the beginning of the disease and they are obligatory for detection of e.g., EV-D68. Challenging in the non-sterile respiratory and intestinal tract specimens is that the etiological connection to the disease cannot be assured, as a significant proportion of healthy individuals are positive for enteroviruses as well. Specimens collected according to the symptoms are recommended for confirmation of the diagnosis. CSF is analyzed when CNS is involved and, with the exception of EV-D68, EV-A71 and polioviruses, enteroviruses can be detected in most meningitis cases, but rarely in encephalitis. From newborns with a systemic sepsis-like disease, blood is used as a supplementary specimen for enterovirus detection. In AHC, enterovirus is most frequently detected in conjunctival and oropharyngeal specimens. Vesicle swabs contain a high viral load in cases of HFMD and herpangina. Biopsy samples, including nails in onychomadesis, can be analyzed as well.

The specimens should be collected as soon as possible after the symptoms have appeared, but at least within four days of the disease onset is recommended. The virus can be detected e.g., in feces for weeks, but etiological connectivity to the disease decreases in time.

The detection of enteroviruses is performed almost solely by molecular methods, and particularly by real-time PCRs. The PCR exceeds the traditional virus isolation both in speed and sensitivity. For enterovirus genome detection, the most sensitive assays rely on highly conserved sequence stretches in the 5' NCR of the genome. There are numerous published in-house protocols guiding the set-up of sensitive and specific assays. Commercial assays are available for either detection of enterovirus only or detection of enterovirus together with other major respiratory, neurotropic or gastrointestinal viruses by multiplexed PCR assays. The 5' NCR assays are highly sensitive, but allow the detection only in the pan-EV level, not in the type level. It is also noteworthy that some of the 5' NCR protocols do not separate enteroviruses from human rhinoviruses, which is a challenge when respiratory samples are analyzed.

Enterovirus type-specific diagnosis is based on the sequences of the capsid coding parts of the genome. The VP1 coding region is the best one for EV typing as it has the most prominent antigenic sites, which are the basis for the classification of enteroviruses. Type-specific real-time PCRs have been designed for example HFMD causing strains CVA6, CVA10 and EV-A71 and these both confirm the presence of enterovirus and type the virus in one PCR reaction. Specific PCR assays are especially useful in large outbreaks, when rapid screening of large amounts of clinical specimens is needed.

Sequencing of the VP1 coding region or its part is a common practice in many clinical and national laboratories. This can be done directly from the clinical specimens, but the assay is usually less sensitive than the PCRs targeting the 5' NCR and it is not recommended for the first choice in diagnostics. The appr. 350 nucleotide long sequence from the amino-terminal region of the VP1 gene is sufficient for reliable typing and molecular epidemiological studies. The complete VP1 sequence is needed for the classification of a new enterovirus type. In addition to VP1, occasionally coding regions for proteins VP4 and VP2 are used for typing. Interpretation of the results is, however, complicated due to possible recombination events between the VP4 and VP1 genomic parts. Next generation sequencing is used for sequencing the complete enterovirus genomes. In most cases, the complete genomic sequence is not relevant in diagnosing individual patients, but it gives important information e.g., on the emergence of new outbreak strains, recombinant types and evolutionary history.

Although largely replaced by molecular assays in primary diagnostics, the golden standard in enterovirus diagnostics – isolation in cell culture – is used in national enterovirus/poliovirus laboratories globally. Most enteroviruses grow rapidly in continuous monkey and human cell lines. The typical cytopathic effect (CPE) appears in 2 to 3 days after inoculation, negative samples are followed for at least 10 days. The CPE is not specific for enteroviruses and the virus isolate should be confirmed as an enterovirus by a real-time PCR assay, sequencing or neutralization with type-specific antisera or pools made of them.

Enterovirus-specific IgM and IgG antibodies are measured by commercial or in-house EIA assays. IgM antibodies are indicative of a recent symptomatic or asymptomatic infection. The IgG antibodies are determined from acute and convalescent sera, and a significant increase is an indication of an acute or very recent infection. The antigen in the IgM and IgG assays is mostly a mixture of VP1 antigens from multiple enterovirus types and gives group-specific, not type-specific results. In addition to IgM- and IgG-EIA, neutralizing antibodies can be measured against a specific virus type in a cell culture assay. The positivity is a reliable indication of a past infection – two serum specimens are needed for the detection of a rise or decline in titers to predict a recent infection. Although not practical in rapid diagnostics, both antibody assays are used widely in seroprevalence studies where large populations are screened for their infection history.

Treatment

In general, enterovirus infections are cleared by host immune system in few days or at least in weeks. Anti-enteroviral drugs have not yet been approved for clinical use. The treatment of enterovirus disease is symptomatic and aims to prevent complications. Immune-modulating treatments, such as intravenous immunoglobulin (IVIG), have been used in severe neonatal cases, hypogammaglobulinemia patients and disease outbreaks, but the evidence of benefits is not conclusive. However, there is supporting evidence that early administered IVIG may reduce the risk of autonomic nervous system dysregulation and death in severe EV-A71 infections, and the treatment is recommended by the WHO. The other passive immunotherapy options are high-dose corticosteroids, plasmapheresis and convalescent serum.

The need for broad-reactive, efficient and safe anti-enteroviral drugs is obvious and has led to an active development during the past decades. Some drugs have been tested in clinical trials but either the lack of efficiency or the emergence of side effects has prevented their licensing for clinical use.

Any step of the enterovirus life cycle may be a target for antiviral therapy. The antivirals can affect the functions of either virus proteins or host factors essential to viral infection. The side effects are supposed to be less prominent if the goal is the virus itself, but targeting the virus may soon result in the emerge of drug-resistant virus strains. The drug candidates are screened from compound libraries and optimized chemically or synthesized *de novo*. As a new approach, drug repurposing is performed to screen for compounds that have already been proved to be safe and that are used in treatment of other diseases.

So far the most promising antivirals have been capsid binders, which occupy the hydrophobic pocket in the capsid and interfere with receptor binding and uncoating by stabilizing the virus particle. The capsid binders can be either pan-enteroviral with broad specificity or targeted to a single type, e.g., EV-A71. In clinical trials, for example Pleconaril, Pirodavidir, Vapendavidir and Pocapavidir have shown efficiency in inhibiting some enteroviruses, but not all types. The fast emergence of resistant viruses was also noticed.

Promising targets have also been viral proteases 2A and 3C, which are involved in viral polyprotein processing. Rupintrivir targets the active site of 3C and inhibits rhinovirus and EV-A71 infection. Viral replication can be affected by inhibiting the 3D polymerase. This can be done either by using nucleoside analogs such as ribavirin and gemcitabine or non-nucleoside inhibitors like amiloride. 2C inhibitors have also been described, one of them being drug fluoxetine.

Virus replication needs essential host factors, which may also be the targets for antivirals. For example, enviroxime, which is an inhibitor of phosphatidylinositol 4-kinase- β , was effective against all enterovirus species in clinical trials, but the drug caused adverse side effects.

Prevention

The most effective prevention of enterovirus infection is achieved by good hygiene, which interrupts the transmission of the virus. Enteroviruses are spread in droplets by coughing and sneezing and they are excreted in feces. In addition, depending on the clinical picture, enteroviruses may be present in eye secretions and blisters. The guidelines emphasize the need for hand washing with soap and water as alcohol-based disinfectants do not have an effect on enteroviruses. Contaminated surfaces should be washed with chlorine or soap and water.

Enterovirus diseases could be prevented efficiently with vaccines, as shown by almost completed eradication of polioviruses. Both inactivated poliovirus vaccine (IPV) and live, attenuated oral vaccine (OPV) have been safe and efficient in preventing poliomyelitis and diminishing virus circulation. The experience on the poliovirus vaccines is of great value when developing new vaccines against other enteroviruses. The major disadvantage of IPV is the weak prevention of transmission, which is better obtained by OPV, which induces better mucosal immunity. The disadvantage of OPV is the reversion of attenuated poliovirus strains to neurovirulent mutants, which may occasionally cause vaccine associated poliomyelitis or, in certain epidemiological situations, circulation of vaccine-derived polioviruses.

Designing a broad-spectrum vaccine against many or all enterovirus types has been too challenging so far. However, attempts have been made to develop vaccines against prevalent enterovirus types such as EV-A71. The first formalin-inactivated whole-virus vaccine against EV-A71 was approved by the Food and Drug Administration of China in 2015. The second vaccine was licensed in 2016, and further products are in clinical trials at the moment. The EV-A71 strains used in the first two vaccines are from different origin, but they both belong to sub-genotype C4a, which has been the most important virus type in Chinese outbreaks. Phase III clinical trials have involved over 30,000 infants and showed that the vaccines are safe and can prevent over 90% of EV-A71-associated HFMD or herpangina.

Besides inactivated and live-virus-vaccines, attenuated whole-virus vaccines are under development against some other EV types as well. Other potential vaccine constructs include virus-like particles, recombinant proteins and DNA vaccines.

Further Reading

- Baggen, *et al.*, 2018. The life cycle of non-polio enteroviruses and how to target it. *Nature Reviews* 16, 368–381.
- Bauer, *et al.*, 2017. Direct acting antivirals and host-targeting strategies to combat enterovirus infections. *Current Opinion in Virology* 24, 1–8.
- Bitnun, A., Yeh, E.A., 2018. Acute flaccid paralysis and enteroviral infections. *Current infectious disease reports* 20, 34.
- CDC & WHO/EURO, 2015. Enterovirus Surveillance Guidelines. Guidelines for Enterovirus Surveillance in Support of the Polio Eradication Initiative (2015). WHO. (ISBN 9789289050814).
- Chen, *et al.*, 2015. Phosphatidylserine vesicles enable efficient en bloc transmission of enteroviruses. *Cell* 60, 619–630. doi:10.1016/j.cell.2015.01.032.
- Harik, *et al.*, 2018. Neonatal nonpolio enterovirus and parechovirus infections. *Seminars in Perinatology* 42, 191–197.
- Harvala, *et al.*, 2018. Recommendations for enterovirus diagnostics and characterization within and beyond Europe. *Journal of Clinical Virology* 101, 11–17.
- Jones, *et al.*, 2018. Outcomes following severe hand foot and mouth disease: A systematic review and meta-analysis. *European Journal of Paediatric Neurology* 22, 763–773.
- Lukashev, *et al.*, 2018. Molecular epidemiology and phylogenetics of human enteroviruses: Is there a forest behind the trees? *Reviews in Medical Virology* 28 (6), e2002. doi:10.1002/rmv.2002.
- Marjomäki, *et al.*, 2015. Infectious entry pathway of Enterovirus B species. *Viruses* 7 (12), 6387–6399. doi:10.3390/v7122945.
- Rhoades, *et al.*, 2011. Enterovirus infections of the central nervous system. *Virology* 411, 288–305. doi:10.1016/j.virol.2010.12.014.

- Seitsonen, J.J., Shakeel, S., Susi, P., *et al.*, 2012. Structural analysis of coxsackievirus A7 reveals conformational changes associated with uncoating. *Journal of Virology* 86, 7207–7215.
- World Health Organization. Regional Office for the Western Pacific, 2011. A Guide to Clinical Management and Public Health Response for Hand, Foot and Mouth Disease (HFMD). WHO.
- Yi, *et al.*, 2017. Enterovirus 71 infection and vaccines. *Clinical and Experimental Vaccine Research* 6, 4–14.

Relevant Websites

- <https://ecdc.europa.eu/en/enteroviruses>
European Centre for Disease Prevention and Control.
- <http://polioeradication.org/>
Global Polio Eradication Initiative: GPEI.
- <https://www.cdc.gov/non-polio-enterovirus/index.html>
Non-Polio Enterovirus.
- <http://www.picornaviridae.com/>
Picornavirus Home.
- <http://www.picornastudygroup.com/>
Picornaviridae Study Group.
- <http://viperd.b.scripps.edu>
Welcome to VIPERdb.
Scripps Research.

Enveloped, Positive-Strand RNA Viruses (*Nidovirales*)

Luis Enjuanes, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

AE Gorbalenya, Leiden University Medical Center, Leiden, The Netherlands

Raoul J de Groot, Utrecht University, Utrecht, The Netherlands

Jeff A Cowley, CSIRO Livestock Industries, Brisbane, QLD, Australia

John Ziebuhr, The Queen's University of Belfast, Belfast, United Kingdom

Eric J Snijder, Leiden University Medical Center, Leiden, The Netherlands

© 2008 Elsevier Ltd. All rights reserved.

This is a reproduction of L. Enjuanes, A.E. Gorbalenya, R.J. de Groot, *et al.*, Nidovirales, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-0-12374410-4.00775-5.

Glossary

3CL^{pro} or M^{pro} 3C-like proteinase, or main proteinase.

ADRP ADP-ribose-1''-phosphatase.

CS TRS Core sequence.

ExoN 3' to 5' exoribonuclease.

NendoU Nidovirus endoribonuclease.

O-MT Ribose-2'-O-methyltransferase.

PL^{pro} Papain-like cysteine proteinase.

TRS Transcription-regulating sequence.

Taxonomy and Phylogeny

The order *Nidovirales* includes the families *Coronaviridae*, *Roniviridae*, and *Arteriviridae* (Fig. 1). The *Coronaviridae* comprises two well-established genera, *Coronavirus* and *Torovirus*, and a tentative new genus, *Bafinivirus*. The *Arteriviridae* and *Roniviridae* include only one genus each, *Arterivirus* and *Okavirus*, respectively. All nidoviruses have single-stranded RNA genomes of positive polarity that, in the case of the *Corona*- and *Roniviridae* (26–32 kbp), are the largest presently known RNA virus genomes. In contrast, members of the *Arteriviridae* have a smaller genome ranging from about 13 to 16 kbp. The data available from phylogenetic analysis of the highly conserved RNA-dependent RNA polymerase (RdRp) domain of these viruses, and the collinearity of the array of functional domains in nidovirus replicase polyproteins, were the basis for clustering coronaviruses and toroviruses (Fig. 2). The more distantly related roniviruses also group with corona- and toroviruses, thus forming a kind of supercluster of nidoviruses with large genomes. By contrast, arteriviruses must have diverged earlier during nidovirus evolution. The current taxonomic position of coronaviruses and toroviruses as two genera of the family *Coronaviridae* is currently being revised by elevating these virus groups to the taxonomic rank of either subfamily or family.

A comparative sequence analysis of coronaviruses reveals three phylogenetically compact clusters: groups 1, 2, and 3. Within group 1, two subsets can be distinguished: subgroup 1a that includes transmissible gastroenteritis virus (TGEV), canine coronavirus (CCoV), and feline coronavirus (FCoV), and subgroup 1b that includes the human coronaviruses (HCoV) 229E and NL63, porcine epidemic diarrhoea virus (PEDV), and bat coronavirus (BtCoV) 512 which was isolated in 2005. Within group 2 coronaviruses, two subsets have been recognized: subgroup 2a, including mouse hepatitis virus (MHV), bovine coronavirus (BCoV), HCoV-OC43, and HCoV-HKU1; and subgroup 2b, including severe acute respiratory syndrome coronavirus (SARS-CoV) and its closest circulating bat coronavirus relative, BtCoV-HKU3. A growing number of other bat viruses has been recently identified in groups 1 and 2. It is currently being debated whether some of these viruses (e.g., BtCoV-HKU5, BtCoV-133 (isolated in 2005), and BtCoV-HKU9) may in fact represent novel subgroups or groups. Avian infectious bronchitis virus (IBV) is the prototype of coronavirus group 3, which also includes several other bird coronaviruses. In arteriviruses, there are four comparably distant genetic clusters, the prototypes of which are equine arteritis virus (EAV), lactate dehydrogenase-elevating virus (LDV) of mice, simian hemorrhagic fever virus (SHFV) infecting monkeys, and porcine reproductive and respiratory syndrome virus (PRRSV) which infects pigs and includes European and North American genotypes.

Roniviruses are the only members of the order *Nidovirales* that are known to infect invertebrates. The family *Roniviridae* includes the penaeid shrimp virus, gill-associated virus (GAV), and the closely related yellow head virus (YHV).

More than 100 full-length coronavirus genome sequences and around 30 arterivirus genome sequences have been documented so far, whereas only very few sequences have been reported for toroviruses, bafiniviruses, and roniviruses. Therefore, information on the genetic variability of these nidovirus *taxa* is limited.

Diseases Associated with Nidoviruses

Coronavirus infections are mainly associated with respiratory, enteric, hepatic, and central nervous system diseases. In humans and fowl, coronaviruses primarily cause upper respiratory tract infections, while porcine and bovine coronaviruses establish enteric

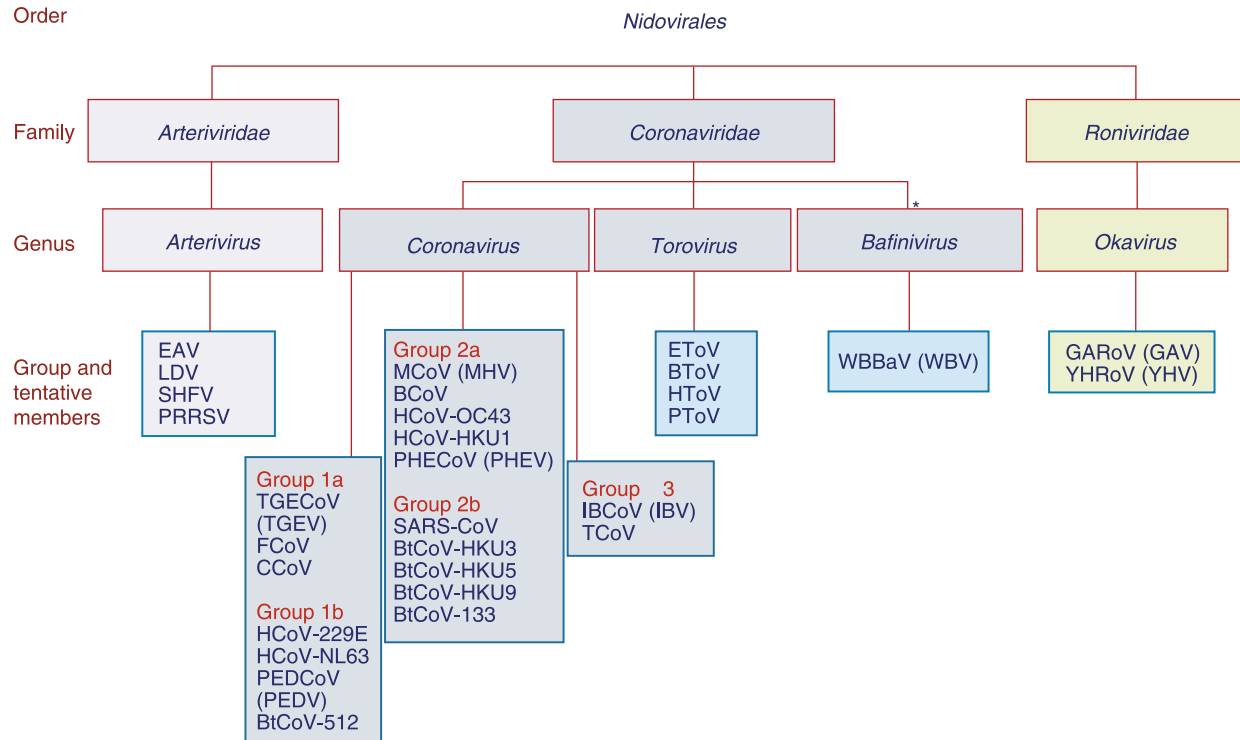


Fig. 1 Nidovirus classification and prototype members. The order *Nidovirales* containing the families *Coronaviridae* (including the established genera *Coronavirus* and *Torovirus*, and a new tentative genus *Bafinivirus*), *Arteriviridae*, and *Roniviridae*. Phylogenetic analysis (see Fig. 2) has confirmed the division of coronaviruses into three groups. In arteriviruses, four comparably distant genetic clusters have been differentiated. To facilitate the taxonomy of the different virus isolates, the types Co, To, Ba, Ro, standing for coronavirus, torovirus, bafinivirus, or ronivirus, respectively, have been included. The following CoVs are shown: human coronaviruses (HCoV) 229E, HKU1, OC43 and NL63, transmissible gastroenteritis virus (TGEV), feline coronavirus (FCoV), porcine epidemic diarrhoea virus (PEDV), mouse hepatitis virus (MHV), bovine coronavirus (BCoV), bat coronaviruses (BtCoV) HKU3, HKU5, HKU9, 133 and 512 (the last two isolated in 2005), porcine hemagglutinating encephalomyelitis virus (PHEV), avian infectious bronchitis virus (IBV), and severe acute respiratory syndrome coronavirus (SARS-CoV); ToV: equine torovirus (EToV), bovine torovirus (BTov), human torovirus (HToV), and porcine torovirus (PToV); BaV: white bream virus (WBV); Arterivirus: equine arteritis virus (EAV), simian haemorrhagic fever virus (SHFV), lactate dehydrogenase-elevating virus (LDV), and three (Euro, HB1, and MLV) porcine reproductive and respiratory syndrome viruses (PRRSV); RoV: gill-associated virus (GAV) and yellow head virus (YHV). Human viruses are highlighted in red. Some nodes are formed by a pair of very closely related viruses (e.g., SARS-CoV and BtCoV-HKU3). Asterisk indicates tentative genus.

infections, often resulting in severe economic losses. In 2002, a previously unknown coronavirus that probably has its natural reservoir in bats crossed the species barrier and caused a major outbreak of SARS, which led to more than 800 deaths worldwide.

Toroviruses cause gastroenteritis in mammals, including humans, and possibly also respiratory infections in older cattle. Bafiniviruses have been isolated from white bream fish but there is currently no information on the pathogenesis associated with this virus infection. Roniviruses usually exist as asymptomatic infections but can cause severe disease outbreaks in farmed black tiger shrimp (*Penaeus monodon*) and white pacific shrimp (*Penaeus vannamei*), which in the case of YHV can result in complete crop losses within a few days after the first signs of disease in a pond. Infections by arteriviruses can cause acute or persistent asymptomatic infections, or respiratory disease and abortion (EAV and PRRSV), fatal age-dependent poliomyelitis (LDV), or fatal hemorrhagic fever (SHFV). Arteriviruses, particularly PRRSV in swine populations, cause important economic losses.

Virus Structure

In addition to the significant variations in genome size among the three nidovirus families mentioned above, there are also major differences in virion morphology (Fig. 3) and host range. Nidoviruses have a lipid envelope which protects the internal nucleocapsid structure and contains a number of viral surface proteins (Fig. 3). Whereas coronaviruses and the significantly smaller arteriviruses have spherical particle structures, elongated rod-shaped structures are observed in toro-, bafini-, and ronivirus-infected cells. The virus particles of the *Corona*- and *Roniviridae* family members carry large surface projections that protrude from the viral envelope (peplomers), whereas arterivirus particles possess only relatively small projections on their surface. Coronaviruses have an internal core shell that is formed by a nucleocapsid featuring a helical symmetry. The nucleocapsid (N) protein interacts with the carboxy-terminus of the envelope membrane (M) protein. The intracellular forms of torovirus, bafinivirus, and ronivirus

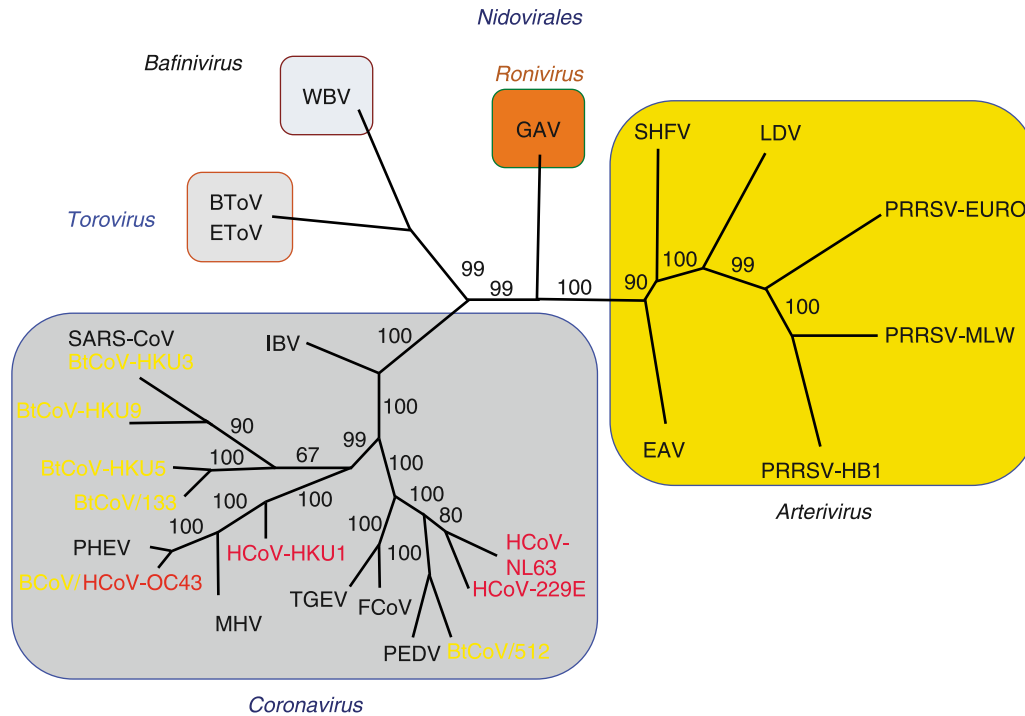


Fig. 2 Nidovirus phylogeny. Tree depicting the evolutionary relationships between the five major groups of nidoviruses as shown in Fig. 1 (Coronavirus, Torovirus, Bafinivirus, Ronivirus, and Arterivirus). This unrooted maximum parsimonious tree was inferred using multiple nucleotide alignments of the RdRp-HEL region of a representative set of nidoviruses with the help of the PAUP*v.4.0b10 software (AEG, unpublished). Support for all bifurcations from 100 bootstraps performed is indicated. The phylogenetic distances shown are approximate. For acronyms, see Fig. 1.

nucleocapsids have extended rod-shaped (helical) morphology. By contrast, mature (extracellular) toroviruses (but not bafini- and roniviruses) feature a remarkable structural flexibility, which allows them to adopt crescent- and toroid-shaped structures also. Unlike other nidoviruses, arteriviruses have an isometric core shell. In all nidoviruses, the nucleocapsid is formed by only a single N protein that interacts with the genomic RNA.

Both the number and properties of structural proteins vary between viruses of the three families of the *Nidovirales* and may even vary among viruses of the same family. Nidoviruses usually encode at least three structural proteins: a spike (S) or major surface glycoprotein, a trans-membrane (M) or matrix protein, and the N protein (Fig. 3). Ronivirus particles are unique in that they possess two envelope glycoproteins, gp116 (S1) and gp64 (S2), but no M protein. Coronavirus and arterivirus particles possess another envelope protein called E that is not conserved in toroviruses, bafini-, and roniviruses. Toroviruses and subgroup 2a coronaviruses, such as MHV, have a hemagglutinin esterase (HE) as an additional structural protein, whereas the SARS-CoV has at least four additional proteins that are present in the viral envelope (encoded by ORFs 3a, 6, 7a, and 7b). The proteins may promote virus growth in cell culture or *in vivo*, but they are dispensable for virus replication.

The major envelope proteins are the S and M proteins in coronaviruses and toroviruses, the GP₅ and M proteins in arteriviruses, and the S1 and S2 proteins in roniviruses. Among these, only the corona-, toro-, and bafinivirus S and M proteins share limited sequence similarities, possibly indicating a common origin. Whereas S proteins can differ in size, they share an exposed globular head domain and, with the exception of roniviruses, a stem portion containing heptad repeats organized in a coiled-coil structure. The S proteins of corona- and toroviruses (and most likely those of bafiniviruses) form trimers that bind the cell surface receptor whereas receptor binding in roniviruses is probably mediated by gp116 (S1). The arterivirus envelope proteins form two higher-order complexes: one is a disulfide-linked heterodimer of GP₅ and the M protein; and the other is a heterotrimer of the minor structural glycoproteins GP₂, GP₃, and GP₄. Except for the E and M proteins, all arterivirus structural proteins are glycosylated. By contrast, the M proteins of corona- and toroviruses (and, most likely, bafiniviruses) are glycosylated, and they share a triple-spanning membrane topology with the amino-terminus exposed on the outside of the virions and the carboxy-terminus facing the nucleocapsid. In TGEV, a proportion of the M proteins has a tetra-spanning membrane topology leading to the exposure of both termini on the virion surface.

In the virion, the coronavirus E protein has a low copy number (around 20) and deletion of the E protein gene from the genome of the group 1 coronavirus TGEV blocks virus maturation, preventing virus release and spread. In the group 2 coronaviruses MHV and SARS-CoV, deletion of the E protein results in a dramatic reduction, of up to 100 000-fold, of virus infectivity. The coronavirus E and SARS-CoV 3a proteins are viroporins, that is, they belong to a group of proteins that modify membrane permeability by forming ion channels in the virion envelope.

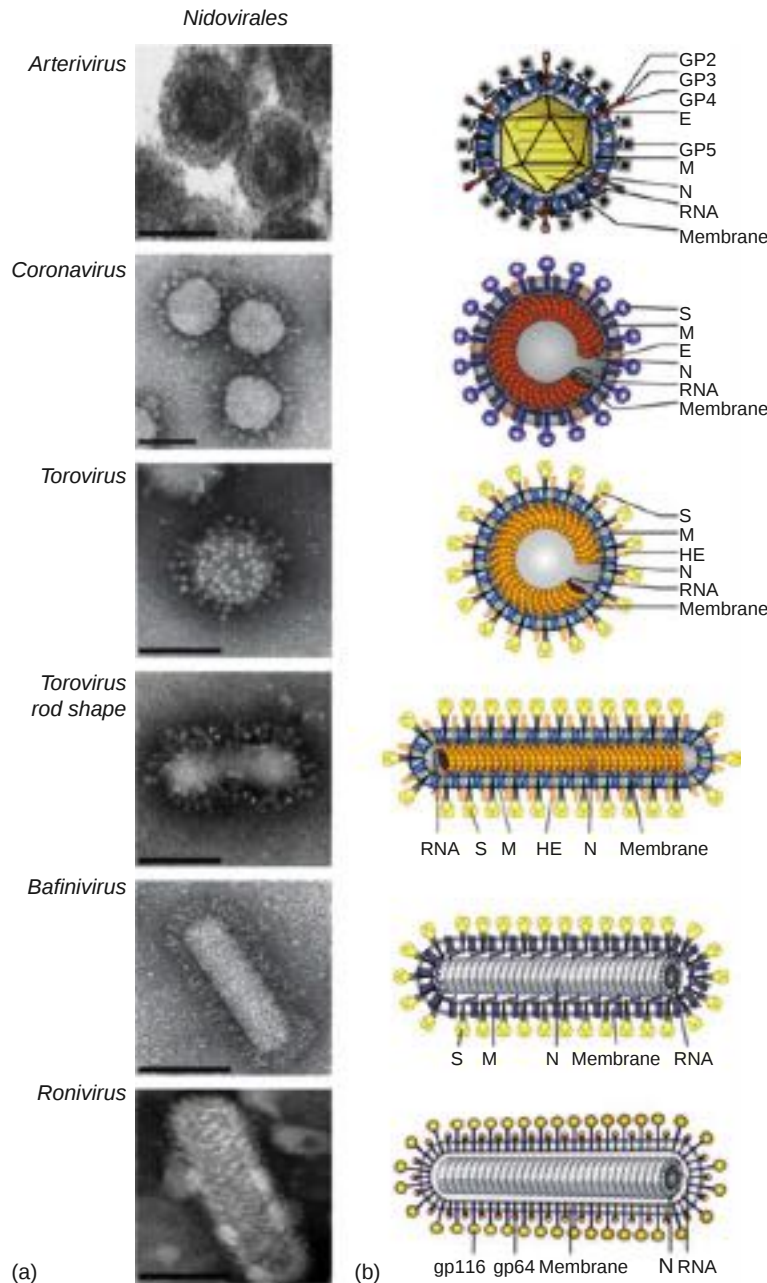


Fig. 3 Nidovirus structure. Architecture of particles of members of the order *Nidovirales*: electron micrographs (a) and schematic representations (b) are shown. N, nucleocapsid protein; S, spike protein; M, membrane protein; E, envelope protein; HE, hemagglutinin-esterase. Coronavirus M protein interacts with the N protein. In arterivirus, GP₅ and M are major envelope proteins, while GP₂, GP₃, GP₄, and E are minor envelope proteins. Toro-, bafini-, and roniviruses lack the E protein present in corona- and arteriviruses. Proteins gp116 and gp64, ronivirus envelope proteins. Different images were reproduced with permission from different authors: arterivirus, E. Snijder (Leiden, The Netherlands); ronivirus, P. J. Walker (CSIRO, Australia); bafinivirus, J. Ziebuhr (Queen's University, Belfast) torovirus, D. Rodriguez (CNB, Spain); coronavirus, L. Enjuanes (CNB, Spain).

Genome Organization

Nidovirus genomes contain variable numbers of genes, but in all cases the 5' terminal two-thirds to three-quarters of the genome is dedicated to encoding the key replicative proteins, whereas the 3' proximal genome regions generally encode the structural and, in some cases, accessory (group- and virus-specific) proteins (Fig. 4). Nidovirus genome expression is controlled at the translational and post-translational levels. Thus, for example, ribosomal frameshifting is required for the expression of ORF1b, and the two replicase polyproteins (pp1a and pp1ab) are proteolytically processed by viral proteases. The proteolytic processing occurs in a

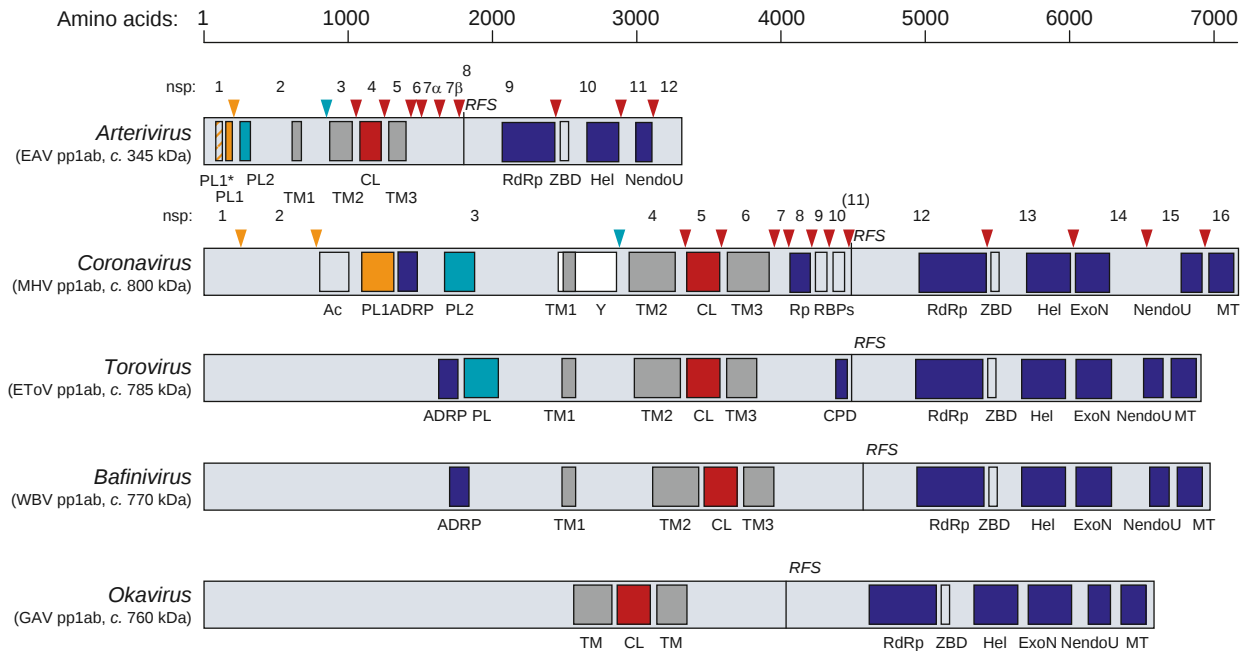


Fig. 5 Nidovirus replicase genes. Polyprotein (pp) 1ab domain organizations are shown for representative viruses from the five nidovirus genera. Acronyms as in Fig. 1. Arterivirus and coronavirus pp1ab processing pathways have been characterized in considerable detail and are illustrated here for EAV and MHV. N-proximal polyprotein regions are cleaved at two or three sites by viral papain-like proteases 1 (PL1) and 2 (PL2), whereas the central and C-terminal polyprotein regions are processed by the main protease, M^{pro}. PL1 domains are indicated by orange boxes and cognate cleavage sites are indicated by orange arrowheads. PL2 domains and PL2-mediated cleavages are shown in green and CL domains and CL-mediated cleavages are shown in red. Note that EAV encodes a second, but proteolytically inactive PL1 domain (PL1*; orange-striped box). For the genera *Torovirus*, *Bafinivirus*, and *Okavirus*, the available information on pp1ab proteolytic processing is limited and not shown here. Other predicted or proven enzymatic activities are shown in blue: ADRP, ADP-ribose 1'-phosphatase; Rps, noncanonical RNA polymerase ('primase') activity; RdRp, RNA-dependent RNA polymerase; HEL, NTPase/RNA helicase and RNA 5'-triphosphatase; ExoN, 3'-to-5' exoribonuclease; NendoU, nidoviral uridylyte-specific endoribonuclease; MT, ribose-2'-O methyltransferase; CPD, cyclic nucleotide phosphodiesterase. Regions with predicted transmembrane (TM) domains are indicated by gray boxes. Other functional domains are shown as white boxes: Ac, acidic domain; Y, Y domain containing putative transmembrane and zinc-binding regions; ZBD, helicase-associated zinc-binding domain; RBPs, RNA-binding proteins. Expression of the C-terminal part of pp1ab requires a ribosomal frameshift, which occurs just upstream of the ORF1a translation stop codon. The ribosomal frameshift site (RFS) is indicated.

to develop selective protease inhibitors that block viral replication, suggesting that nidovirus main proteases may be attractive targets for antiviral drug design.

In arteri-, corona-, and toroviruses, the M^{pro} is assisted by 1–4 papain-like ('accessory') proteases (PL^{pro}) that process the less well-conserved N-proximal region of the replicase polyproteins (Fig. 5). Nidovirus PL^{pro} domains may include zinc ribbon structures and some of them have deubiquitinating activities, suggesting that these proteases might also have functions other than polyprotein processing. Bafini- and roniviruses have not been studied in great detail and it is not yet clear if these viruses employ papain-like proteases to process their N-terminal pp1a/pp1ab regions (Fig. 5).

The replicase polyproteins of 'large' nidoviruses with genome sizes of more than 26 kb (i.e., corona-, toro-, bafini-, and roniviruses) include 3'-5' exoribonuclease (ExoN) and ribose-2'-O-methyltransferase (MT) activities that are essential for coronavirus RNA synthesis but are not conserved in the much smaller arteriviruses (Fig. 5). The precise biological function of ExoN has not been established for any nidovirus but the relationship with cellular DEDD superfamily exonucleases and recently published data suggest that ExoN may have functions in the replication cycle of large nidoviruses that, like in the DEDD homologs, are related to proofreading, repair, and recombination mechanisms.

NendoU is a nidovirus-wide conserved domain that has no counterparts in other RNA viruses. It is therefore considered a genetic marker of the *Nidovirales*. The endonuclease has uridylyte specificity and forms hexameric structures with six independent catalytic sites. Cellular homologs of NendoU have been implicated in small nucleolar RNA processing whereas the role of NendoU in viral replication is less clear. Reverse genetics data indicate that NendoU has a critical role in the viral replication cycle.

Two other RNA-processing domains, ADP-ribose-1''-phosphatase (ADRP) and nucleotide cyclic phosphodiesterase (CPD), are conserved in overlapping subsets of nidoviruses (Fig. 5). Except for arteri- and roniviruses, all nidoviruses encode an ADRP domain that is part of a large replicase subunit (nsp3 in the case of coronaviruses). The coronavirus ADRP homolog has been shown to have ADP-ribose-1'-phosphatase and poly(ADP-ribose)-binding activities. Although the highly specific phosphatase activity is not essential for viral replication *in vitro*, the strict conservation in all genera of the *Coronaviridae* suggests an important (though currently unclear) function of this protein in the viral replication cycle. This may be linked to host cell functions and,

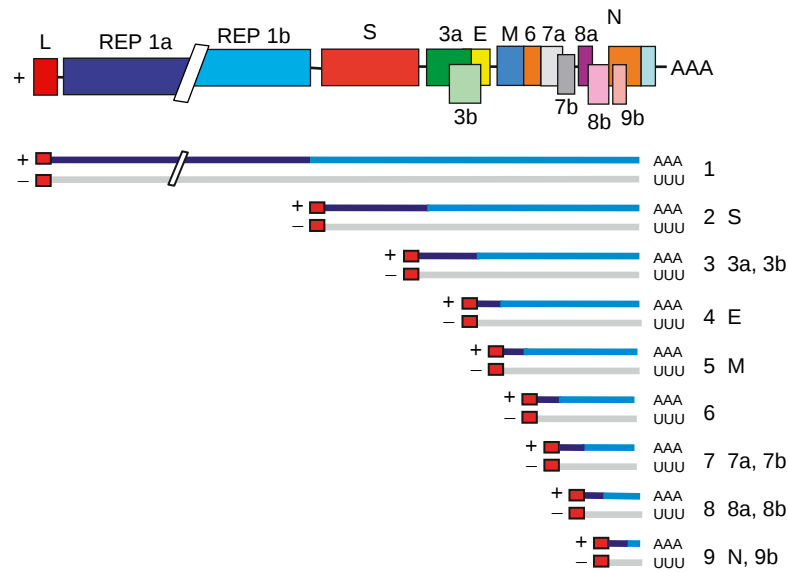


Fig. 6 SARS-CoV genome organization and sg mRNA expression. A diagram of CoV structure using SARS-CoV as a prototype. Below the top bar a set of positive- and negative-sense mRNA species synthesized in infected cells is shown. Dark and light blue lines (+), mRNA sequences translated and nontranslated into viral proteins, respectively. Light gray lines (–), RNAs complementary to the different mRNAs. L, leader sequence. Poly(A) and Poly(U) tails are indicated by AAA or UUU, respectively. Rep 1a and Rep 1b, replicase genes. Other acronyms above and below the top bar indicate structural and nonstructural proteins. Numbers and letters to the right of the thin bars indicate the sg mRNAs.

particularly, to the activities of cellular homologs called ‘macro’ domains which are thought to be involved in the metabolism of ADP-ribose and its derivatives.

The CPD domain is only encoded by toroviruses and group 2a coronaviruses. In toroviruses, the CPD domain is encoded by the 3′ end of replicase ORF1a (Fig. 5), whereas in group 2a coronaviruses, the enzyme is expressed from a separate subgenomic RNA. The enzyme’s biological function is not clear. Coronavirus CPD mutants are attenuated in the natural host whereas replication in cell culture is normal, suggesting some function *in vivo*. The available information suggests that nidovirus replicase polyproteins (particularly, those of large nidoviruses) have evolved to include a number of nonessential functions that may provide a selective advantage in the host.

ORF1a of all nidoviruses encodes a number of (putative) transmembrane proteins, like the coronavirus nsps 3, 4, and 6 and the arterivirus nsps 2, 3, and 5. These have been shown or postulated to trigger the modification of cytoplasmic membranes, including the formation of unusual double-membrane vesicles (DMVs). Tethering of the replication–transcription complex to these virus-induced membrane structures might provide a scaffold or subcellular compartment for viral RNA synthesis, possibly allowing it to proceed under conditions that prevent or impair detection by cellular defense mechanisms, which are usually induced by the double-stranded RNA intermediates of viral replication.

Finally, recent structural and biochemical studies have yielded novel insights into the function of a set of small nsps encoded in the 3′-terminal part of the coronavirus ORF1a. For example, nsp7 and nsp8 were shown to form a hexadecameric supercomplex that is capable of encircling dsRNA. The coronavirus nsp8 was also shown to have RNA polymerase (primase) activity that may produce the primers required by the primer-dependent RdRp residing in nsp12. For nsp9 and nsp10, RNA-binding activities have been demonstrated and crystal structures have been reported for both proteins. Nsp10 is a zinc-binding protein that contains two zinc-finger-binding domains and has been implicated in negative-strand RNA synthesis.

Structural and Accessory Protein Genes

In contrast to the large genome of *Coronaviridae*, which can accommodate genes encoding accessory proteins (i.e., proteins called ‘nonessential’ for being dispensable for replication in cell culture (Fig. 6)), the smaller genomes of arteriviruses only encode essential proteins (Fig. 4). Coronaviruses encode a variable number of accessory proteins (2–8), while the torovirus genome contains a single accessory gene encoding a hemagglutinin-esterase (HE). Coronavirus accessory genes may occupy any intergenic position in the conserved array of the four genes encoding the major structural proteins (5′-S-E-M-N-3′), or they may reside upstream or downstream of this gene array. Roniviruses are unique among the presently known nidoviruses in that the gene encoding the N protein is located upstream rather than downstream of the gene encoding the glycoproteins. Several members of the coronavirus group 1a are exceptional in that they contain genes downstream of the N protein gene, which has not been reported for other coronaviruses. The ronivirus glycoprotein gene is also unique in that it encodes a precursor polyprotein with two

internal signal peptidase cleavage sites used to generate the envelope glycoproteins S1 and S2 as well as an amino-terminal protein with an unknown function.

The accessory genes are specific for either a single virus species or a few viruses that form a compact phylogenetic cluster. Many proteins encoded by accessory genes may function in infected cells or *in vivo* to counteract host defenses and, when removed, may lead to attenuated virus phenotypes. Group 1 coronaviruses may have 2–3 accessory genes located between the S and E genes and up to two other genes downstream of N gene. Viruses of group 2 form the most diverse coronavirus cluster, and they may have between three and eight accessory genes. In this cluster, MHV, HCoV-OC43, and BCoV form the phylogenetically compact subgroup, 2a, that is characterized by the presence of (1) two accessory genes located between ORF1b and the S gene encoding proteins with CPD and HE functions, (2) two accessory genes located between the S and E protein genes, and (3) an accessory gene, I, that is located within the N protein gene. Of this set of five accessory proteins, only three homologs are encoded by the recently identified HCoV-HKU1, which is the closest known relative of the cluster formed by MHV, HCoV-OC43, and BCoV. In contrast, the most distant group 2 member, SARS-CoV, has seven or eight unique accessory genes, two between the S and E protein genes, four to five between the M and N protein genes, and ORF9b which entirely overlaps with the N protein gene in an alternative reading frame. In group 3 avian coronaviruses, of which IBV is the prototype, several accessory genes, which are expressed from functionally tri- or bicistronic mRNAs, have been identified in the region between the S and E protein genes (gene 3) and between the M and N protein genes (gene 5).

Some functionally dispensable ORF1a-encoded replicase domains may also be considered as accessory protein functions. For instance, MHV and SARS-CoV nsp2 turned out to be nonessential for replication in cell culture.

Replication

Like in all other positive-stranded RNA viruses, nidovirus genome replication is mediated through the synthesis of a full-length, negative-strand RNA which, in turn, is the template for the synthesis of progeny virus genomes. This process is mediated by the viral replication complex that includes all or most of the 14–16 nsps derived from the proteolytic processing of the pp1a and pp1ab replicase polyproteins of arteriviruses and coronaviruses. The replication complex, which is likely to include also cellular proteins, is associated with modified intracellular membranes, which may be important to create a microenvironment suitable for viral RNA synthesis as well as for recruitment of host factors. Electron microscopy studies of cells infected with arteriviruses (EAV) and coronaviruses (MHV and SARS-CoV) have shown that RNA synthesis is associated with virus-induced, DMVs. The origin of DMVs is under debate and different intracellular compartments including the Golgi, late endosomal membranes, autophagosomes, and the endoplasmic reticulum have been implicated in their formation.

Studies of *cis*-acting sequences required for nidovirus replication have mainly relied on coronavirus defective-interfering (DI) RNAs replicated by helper virus. Genome regions harboring minimal *cis*-acting sequences have been mapped to around 1 kb domains of the genomic 5' and 3' ends. Studies with MHV DI RNAs have indicated that both genome ends are necessary for positive-strand synthesis, whereas only the last 55 nt and the poly (A) tail at the genomic 3' end are required for negative-strand synthesis. It has been postulated that the 5' and 3' ends of the genome may interact directly during RNA replication, as predicted by computer-aided simulations of MHV and TGEV genomic RNA interactions in protein-free media. There is, however, some experimental evidence supporting protein-mediated cross-talk between both genome ends in the form of RNA–protein and protein–protein interactions.

Several experimental approaches have implicated, in addition to the nsps encoded by the replicase gene, the N protein in coronavirus RNA synthesis. Early in infection, the coronavirus N protein colocalizes with the site of viral RNA synthesis. In addition, the N protein can enhance the rescue of various coronaviruses from synthetic full-length RNA, transcribed *in vitro* or from cDNA clones. In contrast, arterivirus RNA synthesis does not require the N protein.

Host factors that may participate in nidovirus RNA synthesis have been identified mainly from studies of coronaviruses and arteriviruses. In coronaviruses (MHV and TGEV), heterogeneous nuclear ribonucleoprotein (hnRNP) A1 has been identified as a major protein binding to genomic RNA sequences complementary to those in the negative-strand RNA that bind another cellular protein, polypyrimidine tract-binding protein (PTB). hnRNP A1 and PTB bind to the complementary strands at the 5' end of coronavirus RNA and could mediate the formation of an RNP replication complex involving the 5' and 3' ends of coronavirus genomic RNA. The functional relevance of hnRNP A1 in coronavirus replication was supported by experiments showing that its overexpression promotes MHV replication, whereas replication was reduced in cells expressing a dominant-negative mutant of hnRNP A1. There is also experimental evidence to suggest that the poly(A)-binding protein (PABP) specifically interacts with the 3' poly(A) tail of coronavirus genomes, and that this interaction may affect their replication. Other cellular proteins found to bind to coronavirus genomic RNA, such as aconitase and the heat shock proteins HS40 and HS70, might be involved in modulating coronavirus replication. Similarly, interactions of cellular proteins such as transcription cofactor p100 with the EAV nsp1, or of PTB or fructose biphosphate aldolase A with SHFV genomic RNA, suggest that, in arterivirus replication also, a number of cellular proteins may be involved.

Transcription

RNA-dependent RNA transcription in some members of the *Nidovirales* (coronaviruses, bafiniviruses, and arteriviruses), but not in others (roniviruses), includes a discontinuous RNA-synthesis step. This process occurs during the production of subgenome-length

negative-strand RNAs that serve as templates for transcription and involves the fusion of a copy of the genomic 5'-terminal leader sequence to the 3' end of each of the nascent RNAs complementary to the coding (body) sequences (Fig. 6). The resulting chimeric sg RNAs of negative polarity are transcribed to yield sg mRNAs that share both 5'- and 3'- terminal sequences with the genome RNA. Genes expressed through sg mRNAs are preceded by conserved 'transcription-regulatory sequences' (TRSs) that presumably act as attenuation or termination signals during the production of the subgenome-length negative-strand RNAs. In arteriviruses and coronaviruses, the TRSs preceding each ORF are presumed to direct attenuation of negative-strand RNA synthesis, leading to the 'jumping' of the nascent negative-strand RNA to the leader TRS (TRS-L). This process is guided by a base-pairing interaction between complementary sequences (leader TRS and body TRS complement) and it has been proposed that template switching only occurs if the free energy (ΔG) for the formation of this duplex reaches a minimum threshold. This process is named 'discontinuous extension of minus strands' and can be considered a variant of similarity-assisted template-switching that operates during viral RNA recombination. The genome and sg mRNAs share a 5'-leader sequence of 55–92 nt in coronaviruses and 170–210 nt in arteriviruses.

Toroviruses are remarkable in that they employ a mixed transcription strategy to produce their mRNAs. Of their four sg mRNA species, the smaller three (mRNAs 3 through 5) lack a 5' common leader and are produced via nondiscontinuous RNA synthesis. In contrast, sg mRNA2 has a leader sequence that matches the 5'-terminal 18 nt of the genomic RNA and its production requires a discontinuous RNA-synthesis step reminiscent of, but not identical, to that seen in arteri- and coronaviruses.

Synthesis of torovirus mRNAs 3 through 5, and possibly of the two mRNAs in roniviruses, is thought to require the premature termination of negative-strand RNA synthesis at conserved, intergenic, TRS-like sequences to generate subgenome-length negative-strand RNAs that can be used directly as templates for sg mRNA synthesis. In the case of torovirus mRNA2, a TRS is lacking. Fusion of noncontiguous sequences seems to be controlled by a sequence element consisting of a hairpin structure and 3' flanking stretch of 23 residues with sequence identity to a region at the 5' end of the genome. It is thought that during negative-strand synthesis, the hairpin structure may cause the transcriptase complex to detach, prompting a template switch similar to that seen in arteri- and coronaviruses.

In addition to regulatory RNA sequences, viral and host components involved in protein–RNA and protein–protein recognition are likely to be important in transcription. For example, the arterivirus nsp1 protein has been identified as a factor that is dispensable for genome replication but absolutely required for sg RNA synthesis. The identification of host factors participating in nidovirus transcription is a field under development and specific binding assays have recently identified a limited number of cellular proteins that associate with *cis*-acting RNA regulatory sequences. For example, differences in affinity of such factors for body TRSs might regulate transcription in nidoviruses by a mechanism similar to that of the DNA-dependent RNA-polymerase I termination system, in which specific proteins bind to termination sequences.

Origin of Nidoviruses

The complex genetic plan and the replicase gene of nidoviruses must have evolved from simpler ones. Using this natural assumption, a speculative scenario of major events in nidovirus evolution has been proposed. It has been speculated that the most recent common ancestor of the *Nidovirales* had a genome size close to that of the current arteriviruses. This ancestor may have evolved from a smaller RNA virus by acquiring the two nidovirus genetic marker domains represented by the helicase-associated zinc-binding domain (ZBD) and the NendoU function. These two domains may have been used to improve the low fidelity of RdRp-mediated RNA replication, thus generating viruses capable of efficiently replicating genomes of about 14 kbp. The subsequent evolution of much larger nidovirus genomes may have been accompanied by the acquisition of the ExoN domain. This domain may have further improved the fidelity of RNA replication through its 3'–5' exonuclease activity, which might operate in proofreading mechanisms similar to those employed by DNA-based life forms. It has been suggested that the ORF1b-encoded HEL, ExoN, NendoU, and O-MT domains may provide RNA specificity, whereas the relatively abundantly expressed CPD and ADRP might control the pace of a common pathway that could be part of a hypothetical oligonucleotide-directed repair mechanism used in the present coronaviruses and roniviruses. The expansion of the replicase gene may have been associated with an increase in replicase fidelity, thus also supporting the further expansion of the 3'-proximal genome region to encode the structural proteins required to form complex enveloped virions.

Effect of Nidovirus Infection on the Host Cell

Compared to other viruses, the interactions of nidoviruses with their hosts have not been studied in great detail. In many cases, information is based on relatively few studies performed on a limited number of viruses from the families *Coronaviridae* and *Arteriviridae*. Also, most studies have been performed with viruses that have been adapted to cell culture and therefore may have properties that differ from those of field strains. Coronaviruses and arteriviruses are clearly the best-studied members of the *Nidovirales* in terms of their interactions with the host.

Coronavirus infection affects cellular gene expression at the level of both transcription and translation. Upon infection, host cell translation is significantly suppressed but not shut off, as is the case in several other positive-RNA viruses. The underlying mechanisms have not been characterized in detail, but data obtained for MHV and BCoV suggest that they may involve the 5'-leader sequences present on coronavirus mRNAs. The viral N protein was reported to bind to the 5'-common leader sequence and

it has been speculated that this might promote translation initiation, leading to a preferential translation of viral mRNAs. Furthermore, host mRNAs were reported to be specifically degraded in MHV- and SARS-CoV-infected cells, further reducing the synthesis of cellular proteins. Another mechanism affecting host cell protein synthesis may be based on specific cleavage of the 28S rRNA subunit, which was observed in MHV-infected cells.

Studies on cellular gene expression following nidovirus infections have mainly focused on the coronaviruses MHV and SARS-CoV. For example, SARS-CoV infection was reported to disrupt cellular transcription to a larger extent than does HCoV-229E. Differences in cellular gene expression have been proposed to be linked to differences in the pathogenesis caused by these two human coronaviruses. Apart from the downregulation of genes involved in translation and cytoskeleton maintenance, genes involved in stress response, proapoptotic, proinflammatory, and procoagulating pathways were significantly upregulated. Both MHV and SARS-CoV induce mitogen-activated phosphate kinases (MAPKs), especially p38 MAPK. In addition, activation of AP-1, nuclear factor kappa B (NF- κ B), and a weak induction of Akt signaling pathways occur after SARS-CoV infection and the N and nsp1 proteins were suggested to be directly involved in inducing these signaling pathways.

Nidoviruses have also been reported to interfere with cell cycle control. Infection by the coronaviruses TGEV, MHV, SARS-CoV, and IBV was reported to cause a cell cycle arrest in the G0/G1 phase and a number of cellular proteins (e.g., cyclin D3 and hypophosphorylated retinoblastoma protein) and viral proteins (MHV nsp1, SARS-CoV 3b 7a, and N proteins) have been proposed to be involved in the cell cycle arrest in G0/G1.

Many viruses encode proteins that modulate apoptosis and, more generally, cell death, which allows for highly efficient viral replication or the establishment of persistent infections. Infection by coronaviruses (e.g., TGEV, MHV, and SARS-CoV) and arteriviruses (e.g., PRRSV and EAV) have been reported to induce apoptosis in certain cell types. Apoptosis has also been reported in shrimp infected with the ronavirus YHV and is thought to be involved in pathogenesis. Both apoptotic and antiapoptotic molecules have been found to be upregulated, suggesting that a delicate counterbalance of pro- and antiapoptotic molecules is required to ensure cell survival during the early phase of infection, and rapid virus multiplication before cell lysis occurs. Coronavirus-induced apoptosis appears to occur in a tissue-specific manner, which obviously has important implications for viral pathogenesis. For instance, SARS-CoV was shown to infect epithelial cells of the intestinal tract and induce an antiapoptotic response that may counteract a rapid destruction of infected enterocytes. These findings are consistent with clinical observations of a relatively normal endoscopic and microscopic appearance of the intestine in SARS patients. Furthermore, SARS-CoV causes lymphopenia which involves the depletion of T cells, probably by apoptotic mechanisms that are triggered by direct interactions of the SARS-CoV E protein with the antiapoptotic factor Bcl-xL. Also the MHV E protein has been reported to induce apoptosis. The SARS-CoV 7a protein was found to induce apoptosis in cell lines derived from lung, kidney, and liver, by a caspase-dependent pathway. Apoptosis has also been associated with arterivirus infection but information on underlying mechanisms and functional implications is limited.

Coronavirus and arterivirus infections trigger proinflammatory responses that often are associated with the clinical outcome of the infection. Thus, for example, there seems to be a direct link between the IL-8 plasma levels of SARS patients and disease severity, similar to what has been described for pulmonary infections caused by respiratory syncytial virus. In contrast, despite the upregulation of IL-8 in intestinal epithelial cells, biopsy specimens taken from the colon and terminal ileum of SARS patients failed to demonstrate any inflammatory infiltrates, which may be the consequence of a virus-induced suppression of specific cytokines and chemokines, including IL-18, in the intestinal environment.

Innate immunity is essential to control vertebrate nidovirus infection *in vivo*. The induction of type I IFN (IFN- α/β) varies among different coronaviruses and arteriviruses. Whereas some coronaviruses such as TGEV are potent inducers of type I IFN, other coronaviruses (MHV and SARS-CoV) or arteriviruses (PRRSV) do not stimulate its production, thus facilitating virus escape from innate immune defenses. Type I interferon is a key player in innate immunity and in the activation of effective adaptive immune responses. Upon viral invasion, IFN- α/β is synthesized and secreted. IFN- α/β molecules signal through the type I interferon receptor (IFNR), inducing the transcription of several antiviral mediators, including IFN- γ , PKR, and Mx. IFN- γ is critical in resolving coronavirus (MHV and SARS-CoV), and also arterivirus (EAV, LDV, and PRRSV) infections. Like many other viruses, coronaviruses have developed strategies to escape IFN responses. For example, it has been shown that the SARS-CoV 3b, 6, and N proteins antagonize interferon by different mechanisms, even though all these proteins inhibit the expression of IFN by interfering with the function of IRF-3.

In arteriviruses such as PRRSV, IFN- γ is produced soon after infection to promote Th1 responses. However, PRRSV infections or vaccination with attenuated-live PRRSV vaccines cause only limited IL-1, TNF- α , and IFN- α/β responses. This then leads to IFN- γ and Th1 levels that fail to elicit strong cellular immune responses.

See also: Coronaviruses: General Features (*Coronaviridae*). Coronaviruses: Molecular Biology (*Coronaviridae*). Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (*Coronaviridae*)

Further Reading

- de Groot, R.J., 2007. Molecular biology and evolution of toroviruses. In: Snijder, E.J., Gallagher, T., Perlman, S. (Eds.), *The Nidoviruses*. Washington, D.C.: A.S.M. Press, pp. 133–146.
- Enjuanes, L. (Ed.), 2005. *Current Topics in Microbiology and Immunology*, vol. 287: Coronavirus Replication and Reverse Genetics. Berlin: Springer.
- Enjuanes, L., Almazan, F., Sola, I., Zuniga, S., 2006. Biochemical aspects of coronavirus replication and virus–host interaction. *Annual Review of Microbiology* 60, 211–230.
- Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. *Nidovirales*: Evolving the largest RNA virus genome. *Virus Research* 117, 17–37.

- Masters, P.S., 2006. The molecular biology of coronaviruses. *Advances in Virus Research* 66, 193–292.
- Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. *Journal of Virology* 81, 20–29.
- Siddell, S.G., Ziebuhr, J., Snijder, E.J., 2005. Coronaviruses, toroviruses, and arteriviruses. In: Mahy, B.W.J., ter-Meulen, V. (Eds.), *Virology*, 10th edn., vol. 1. London: Hodder-Arnold, pp. 823–856.
- Snijder, E.J., Siddell, S.G., Gorbatenya, A.E., 2005. The order *Nidovirales*. In: Mahy, B.W.J., ter-Meulen, V. (Eds.), *Virology*, 10th edn., vol. 1. London: Hodder-Arnold, pp. 390–404.
- Snijder, E.J., Spaan, W.J.M., 2007. Arteriviruses. In: Knipe, D.M., Howley, P.M., Griffin, D.E., *et al.* (Eds.), *Fields Virology*, vol. 1. Philadelphia: Lippincott Williams and Wilkins, pp. 1205–1220.
- Spaan, W.J.M., Cavanagh, D., de Groot, R.J., *et al.*, 2005. Nidovirales. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press, pp. 937–945.
- van Vliet, A.L.W., Smits, S.L., Rottier, P.J.M., de Groot, R.J., 2002. Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO Journal* 21 (23), 6571–6580.
- Walker, P.J., Bonami, J.R., Boonsaeng, V., *et al.*, 2005. Roniviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press, pp. 975–979.
- Ziebuhr, J., Snijder, E.J., 2007. The coronavirus replicase: Special enzymes for special viruses. In: Thiel, V. (Ed.), *Molecular and Cellular Biology: Coronaviruses*. Norfolk, UK: Caister Academic Press, pp. 31–61.
- Zuñiga, S., Sola, I., Alonso, S., Enjuanes, L., 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *Journal of Virology* 78, 980–994.

Epstein–Barr Virus (*Herpesviridae*)

Lawrence S Young, University of Warwick, Coventry, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Classification

Epstein-Barr virus (EBV) is a member of the genus *Lymphocryptovirus* (LCV) which belongs to the lymphtropic *Gammaherpesvirinae* subfamily of the family *Herpesviridae*. EBV is closely related to the LCVs present in hominids (e.g., chimpanzees) and other Old World non-human primates (e.g., baboons and macaques). These viruses share homologous sequences and genetic organization, and infect the B lymphocytes of their host species resulting in the establishment of persistent infection in vivo, B cell lymphomagenesis and the ability to immortalize B cells from their own natural host in vitro. LCVs have been isolated from common marmosets which are New World monkeys but these viruses show considerable divergence from the genomes of EBV and Old World primate LCVs suggesting that these viruses represent a more primitive predecessor of the LCVs infecting higher-order primates. LCV-specific genes can be classified into three groups based on evolutionary analysis: (1) universal herpesvirus genes common to all herpesviruses including genes important for virus replication and structure, (2) ancestral genes specific to the LCV genus responsible for intrinsic properties of these viruses such as B cell tropism, B cell transformation and virus reactivation, and (3) acquired LCV-specific genes that evolved later in the LCV genus and therefore are present in EBV and Old World LCVs but not New World LCVs, these include genes associated with the evasion of host immune responses.

Virion and Genome Structure

EBV is about 120–180 nm in diameter and has a similar structure to other herpesviruses with a core composed of linear dsDNA surrounded by a protein nucleocapsid comprising 162 capsomeres. There is a protein tegument which encompasses the nucleocapsid and is surrounded by a cell membrane-derived envelope containing both lipids and glycoprotein spikes which mediate virus infection (Fig. 1).

The EBV genome is composed of linear double-stranded DNA, approximately 172 kilobase pairs (kb) in length. EBV has a series of 0.5 kb terminal direct repeats (TRs) and internal repeat sequences (IRs) that divide the genome into short and long, largely unique sequence domains (Fig. 2). EBV was the first herpesvirus to have its genome completely cloned and sequenced. Since the EBV genome was sequenced from an EBV DNA *Bam*HI fragment cloned library, open reading frames (ORFs), genes and sites for transcription or RNA processing are frequently referenced to specific *Bam*HI fragments, from A to Z, in descending order of fragment size (Fig. 2). The virus has the coding potential for around 85 proteins, not all of which have been identified or characterized.

History

In 1958 Dennis Burkitt described a lymphoma that represented the most common tumor affecting children in certain parts of East Africa. The geographical distribution of this malignancy suggested that the development of this tumor, named Burkitt lymphoma (BL), might be due to an infectious agent possibly related to malaria. In 1964 the successful establishment of cell lines from explants of BL enabled Tony Epstein and Yvonne Barr to identify herpesvirus-type particles by electron microscopy within a subpopulation of tumor cells in vitro (Fig. 1). Werner and Gertrude Henle subsequently demonstrated that BL-derived cell lines expressed antigens that were recognized not only by sera from patients with BL but also by sera from patients with infectious mononucleosis (IM) and, at lower levels, from normal healthy individuals. Similar seroepidemiological studies also suggested a link between infection with this virus (now called Epstein-Barr virus after its discoverers) and undifferentiated nasopharyngeal carcinoma (NPC) leading to the subsequent direct demonstration of EBV DNA in the tumor cells of NPC. The ability of EBV to efficiently immortalize B lymphocytes in vitro and to induce tumors in non-human primates established this virus as a putative oncogenic agent in humans. Over the last 50 years EBV has been implicated in a variety of other lymphoid and epithelial malignancies.

Epidemiology and Natural History of Infection

EBV is the most common and persistent virus infection in humans with approximately 95% of the world's population sustaining a life-long asymptomatic infection. In underdeveloped countries, primary infection with EBV usually occurs during the first several months to few years of life and is predominantly asymptomatic. However, in developed populations, primary infection is more frequently delayed until adolescence or adulthood, in many cases producing the characteristic clinical features of infectious mononucleosis (IM) otherwise known as glandular fever.

EBV is orally transmitted and infectious virus can be detected in oropharyngeal secretions from IM patients, from immunosuppressed patients and at lower levels from healthy EBV seropositive individuals. The precise site of EBV replication in healthy individuals remains unknown but there is evidence that the virus replicates in differentiating epithelial cells in the oropharynx, in

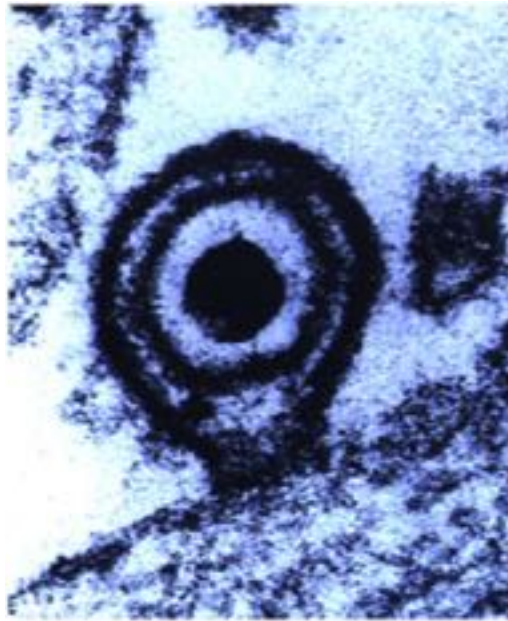


Fig. 1 Electron micrograph of the Epstein-Barr virus virion.

salivary glands and in gingival cells. The nasopharynx and oropharynx are associated with a ringed arrangement of lymphoid tissue (so-called Waldeyer's ring) which may contribute to EBV shedding from virus-infected B cells being reactivated within the local mucosal environment and/or B cell-derived virus infecting and then replicating in differentiating epithelial cells.

EBV can be completely eradicated by irradiation in bone marrow transplant recipients suggesting that B lymphocytes are the main site of EBV persistence. This is supported by the B lymphotropism of EBV which is mediated by the binding of the major viral envelope glycoprotein gp350 to the CR2 receptor on the surface of B cells. Virion penetration of the B cell membrane requires further interactions between the EBV glycoprotein gp42 (which forms a ternary complex with gH and gL viral glycoproteins) and HLA class II molecules. It appears that the presence or absence of HLA class II in virus producing cells influences the tropism of EBV for B cells or epithelial cells by affecting the availability of gp42. Other CR2-independent pathways are responsible for EBV infection of epithelial cells, including secretory component-mediated IgA transport, integrin interactions with polarized epithelium and direct cell-to-cell contact, but these are relatively inefficient and of unknown relevance to EBV infection *in vivo*.

The B cell tropism of EBV and its ability to establish a latent infection in these cells both *in vitro* and *in vivo* further supports the contention that B lymphocytes are the reservoir of lifelong persistent infection. EBV can efficiently transform B cells *in vitro* converting approximately 3%–10% of all target B cells into permanently growing lymphoblastoid cell lines (LCLs) in which every cell carries multiple episomal copies of the EBV genome and expresses all latent genes (referred to as latency III or the "growth" program), including six Epstein-Barr nuclear antigens (EBNAs 1, 2, 3A, 3B and 3C) and EBNA leader protein (EBNA-LP), latent membrane proteins LMP1 and LMP2 (which encodes two isoforms, LMP2A and LMP2B) and the non-coding EBV-encoded RNAs (*EBER1* and *EBER2*) and viral microRNAs (miRNA) (Fig. 2). When peripheral blood lymphocytes from healthy EBV seropositives are placed in culture, the few EBV-infected B lymphocytes that are present regularly give rise to spontaneous outgrowth of EBV-transformed LCLs provided that immune T lymphocytes are either removed or inhibited by the addition of cyclosporin A to the culture. LCLs can also be generated by direct infection of resting B lymphocytes with EBV derived from the throat washings of seropositive individuals or from producer B cell lines. LCLs have provided an invaluable, albeit incomplete, model of the lymphomagenic potential of EBV.

Detailed examination of EBV infection *in vivo* has shown that the virus persists in the IgD⁺CD27⁺ memory B cell subset and that these cells have down-regulated the expression of most, if not all, viral genes. The precise route of entry of EBV-infected B cells into the memory compartment remains a subject of much debate. This reservoir of infected cells is stably maintained thereafter, apparently subject to the same physiologic controls as the general mucosa-associated memory B cell pool. EBV persistence within this B cell population brings with it the possibility of fortuitous antigen-driven recruitment of infected cells into germinal centers, leading to progeny that either re-enter the circulating memory pool or differentiate to plasma cells that may migrate to mucosal sites (Fig. 3). The different forms of latency that are manifest in virus-associated malignancies (discussed below) may represent latency programs that have evolved to accommodate such changes in host cell physiology (Figs. 3 and 4). Thus germinal center transit appears to activate a latency program where only the genome maintenance protein EBNA1 is expressed (latency 0), while exit from germinal centers is possibly linked to the transient expression of the EBV-encoded latent membrane proteins, LMP1 and LMP2 (latency II). The ability of these proteins to mimic the key signals required for B cells to undergo a germinal center reaction, namely T cell help via the CD40 pathway (constitutively provided by EBV-encoded LMP1) and activation of the cognate B cell receptor (augmented by EBV-encoded LMP2A), supports this strategy whereby EBV exploits the physiological process of B cell

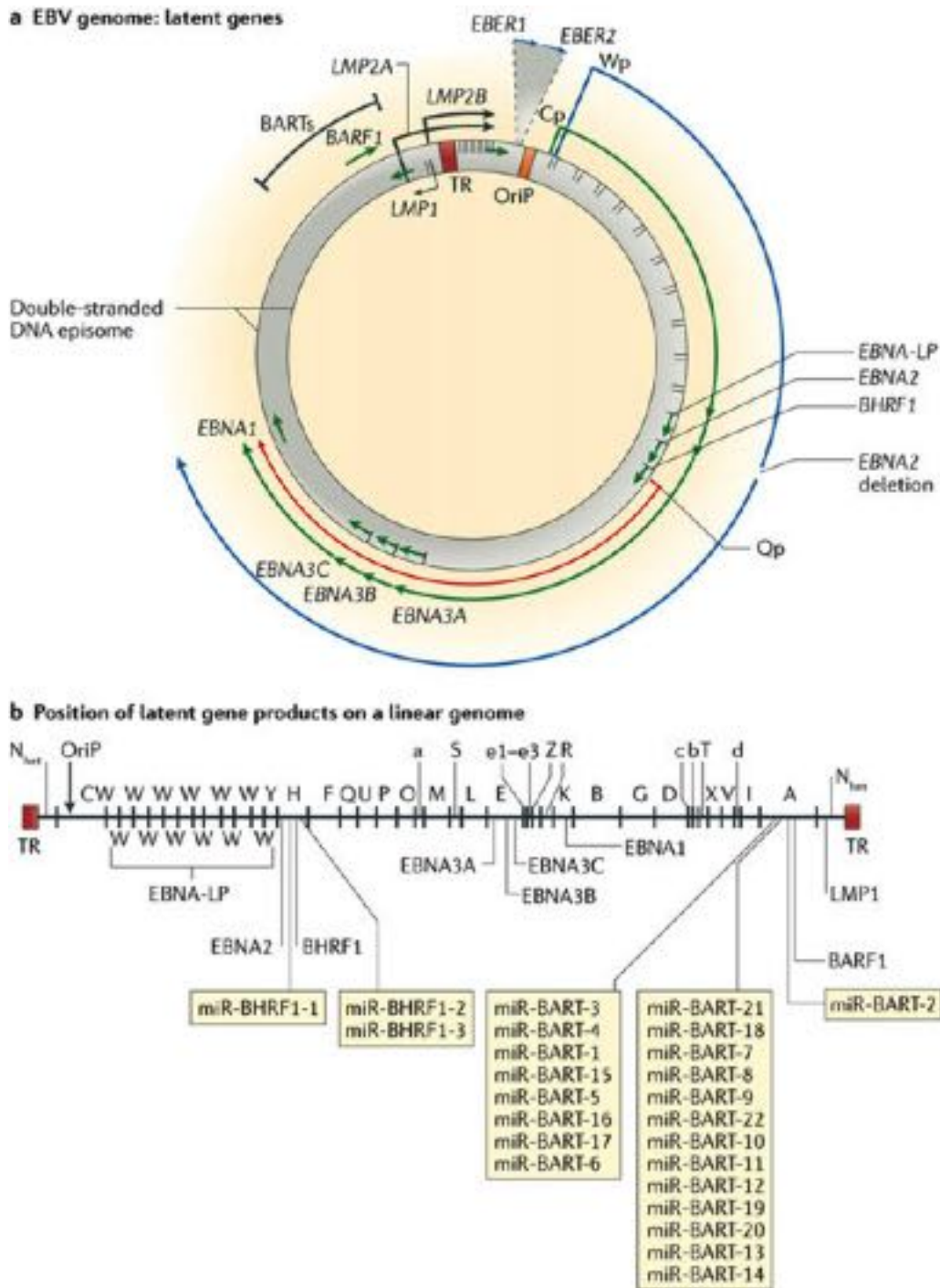


Fig. 2 EBV and its latent genes. (a) Location and transcription of Epstein-Barr virus latent genes on the double-stranded viral DNA episome. The origin of plasmid replication (OriP) is shown in orange. The short green thick arrows represent exons encoding latent proteins: six nuclear antigens (EBNAs 1, 2, 3A, 3B and 3C, and EBNA-LP), latent membrane proteins (LMPs 1, 2A and 2B), BamHI fragment H rightward open reading frame (BHRF1) and BamHI-A fragment rightward reading frame (BARF1). The short blue arrows at the top represent the highly transcribed non-polyadenylated RNAs, *EBER1* and *EBER2*. The middle long green arrow represents EBV transcription during latency III, in which all the EBNAs are transcribed from either the Cp or Wp promoter; the different EBNAs are encoded by individual mRNAs that are generated by differential splicing of the same long primary transcript. The inner red arrow represents the EBNA1 transcript, which originates from the Qp promoter during latency I and latency II. Latency II is characterized by expression of EBNA1 together with the latent membrane proteins. Wp restricted latency is initiated from the Wp promoter and there is expression of all the EBNAs, except EBNA2, which is deleted in this form of latency (outer long blue arrow). (b) Location of open reading frames for EBV latent proteins on the BamHI restriction map of the prototype EBV B95-8 genome and including the viral microRNAs in sequence order. The BamHI fragments are named according to size, with A being the largest. Lowercase letters indicate the smallest fragments. TR refers to terminal repeats at each end of the genome. This region was often referred to as Nhet to indicate heterogeneity in this region according to the numbers of TRs within different virus isolates. BART, BamHI-A rightward transcript. Reproduced from Young, L.S., Yap, L.-F., Murray, P.G., 2016. Epstein-Barr virus: Over 50 years and still providing surprises. *Nature Reviews Cancer* 16, 789–802.

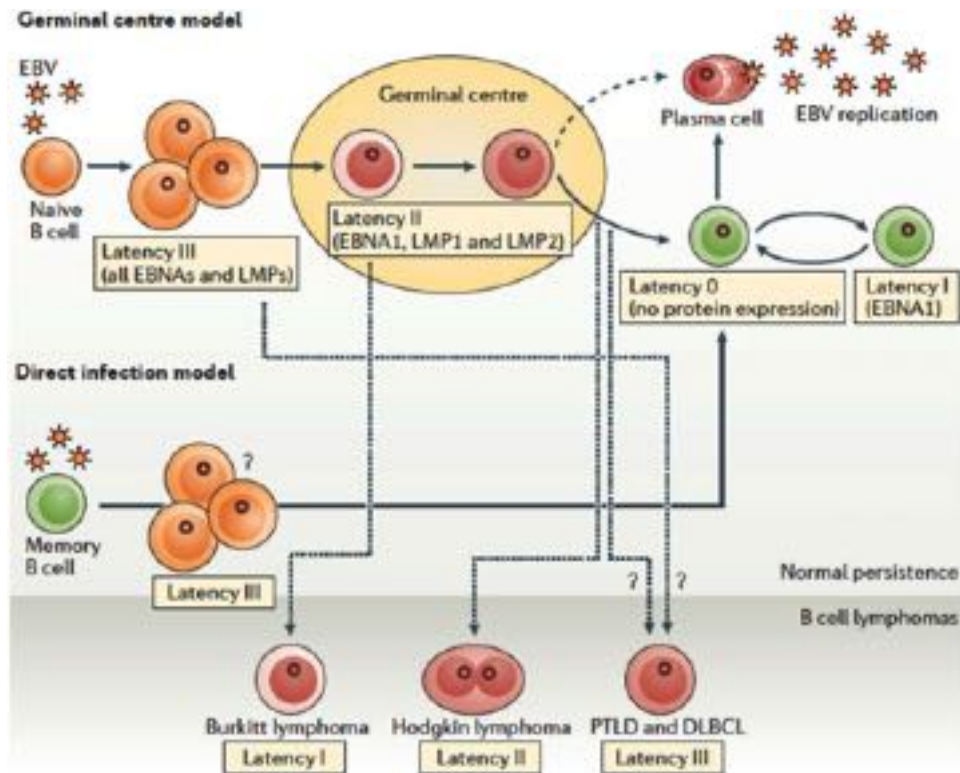


Fig. 3 Virus persistence in the B cells of the human host and the origin of EBV-associated B cell lymphomas. Upper panel: normal virus persistence. EBV resides in memory B cells of asymptomatic hosts. There are two models to explain how the virus enters memory B cells. In the germinal center model, EBV-infection of naïve B cells leads to a latency III ‘growth program’ in which the proliferation and expansion of the infected B cell pool is driven by expression of all EBV latent genes. Cells then enter the germinal center and express latency II (‘default program’), characterized by expression of EBNA1, LMP1 and LMP2. LMP1 provides a CD40-like signal and LMP2 a surrogate B cell receptor like signal; mimicking the same signals provided to the antigen-specific B cell in the normal germinal center. EBV-infected germinal center B cells leave the germinal center and enter the memory B cell pool. Here EBV protein expression is silenced (Latency 0); these cells are replenished by the same signals that normally induce the proliferation of memory B cells. Proliferating memory B cells require EBNA1 expression (Latency I) for viral episome segregation. Memory B cells can terminally differentiate into plasma cells (solid arrow) which triggers virus replication. EBV-infected germinal center B cells might also differentiate directly into plasma cells. (dashed arrow). In the ‘direct infection’ model (shown below), EBV accesses memory B cells following the direct infection of these cells which may involve a latency III intermediary. Lower panel: The origin of the EBV-associated B cell lymphomas. There is uncertainty about the exact stages of differentiation from which the EBV-positive B cell lymphomas arise (as indicated by the black dotted lines), as it cannot be assumed that the pattern of latency observed in the progenitor cell is recapitulated in the corresponding tumor. The figure illustrates the presumed (question marks indicate this uncertainty) cell of origin based on current evidence for post-transplant lymphoproliferative disease (PTLD), diffuse large B cell lymphoma (DLBCL), Hodgkin’s lymphoma and Burkitt’s lymphoma. Reproduced from Young, L.S., Yap, L.-F., Murray, P.G., 2016. Epstein-Barr virus: Over 50 years and still providing surprises. *Nature Reviews Cancer* 16, 789–802.

differentiation to access and maintain persistent infection in the memory B cell pool (Fig. 4). The possible commitment of these cells to plasmacytoid differentiation may trigger entry into lytic virus replication thereby providing a source of low-level virus shedding into the oropharynx. There may also be circumstances in which infected cells in the reservoir can reactivate back to proliferative infections similar to those of an EBV-transformed LCL.

EBV Strain Variation

The B cell-derived B95.8 strain of EBV was fully sequenced in 1984 and was, at that time, the largest DNA sequence (172 kb) ever determined. While variations in repeat regions of the EBV genome are observed amongst different EBV isolates, the genomes of viruses from different regions of the world or from patients with different virus-associated diseases are very similar. Strain variation over the EBNA2-encoding (BamHI WYH) region of the EBV genome permits all virus isolates to be classified as either “type 1” (EBV-1, B95.8-like) or “type 2” (EBV-2, Jijoye-like). This genomic variation results in the production of two antigenically distinct forms of the EBNA2 protein which share only 50% amino acid homology. Similar allelic polymorphisms (with 50%–80% sequence homology depending on the locus) related to the EBV type occur in a subset of latent genes, namely those encoding EBNA-LP, EBNA3A, EBNA3B and EBNA3C. These differences have functional consequences as EBV-2 isolates are less efficient in *in vitro* B lymphocyte transformation assays compared with EBV-1 isolates. A combination of virus isolation and sero-

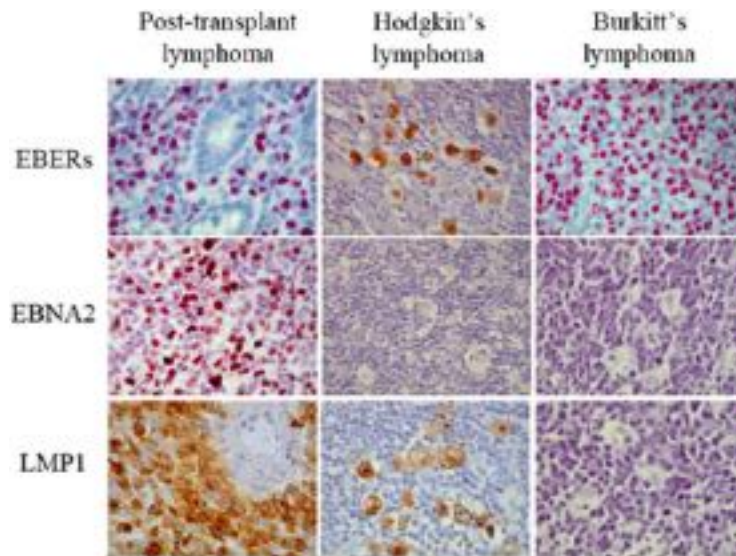


Fig. 4 EBV-associated lymphomas display different patterns of virus latent gene expression. While the post-transplant lymphomas associated with EBV infection expression a pattern of latent gene expression similar to that observed in virus-transformed B cells *in vitro* (latency III), Hodgkin's lymphoma (latency II) and Burkitt's lymphoma (latency I) express more restricted forms of EBV latency. Upper panels show *in situ* hybridization for the abundant non-polyadenylated EBER transcripts which are expressed in all forms of EBV latent infection. The middle panels depict immunohistochemical staining for the EBNA2 protein which is only expressed in post-transplant lymphoma and not either Hodgkin's lymphoma or Burkitt's lymphoma. The bottom panels show immunohistochemical staining for the LMP1 protein which is expressed in post-transplant lymphoma and HL but not in BL.

epidemiological studies suggest that type 1 virus isolates are predominant (but not exclusively so) in many Western countries, whereas both types are widespread in equatorial Africa, New Guinea and certain other regions. The evolutionary relationship between the type 1 and type 2 strains of EBV remain obscure. They may have evolved from a common progenitor virus or through recombination of the ancestral LCVs infecting Old World primates. The pronounced (but not exclusive) segregation of EBV-2 isolates within equatorial regions suggests that environmental factors (including immunological competence) may have influenced EBV evolution and may still be responsible for the effective competition between EBV-2 isolates and the ubiquitous EBV-1 family.

In addition to this broad distinction between EBV types 1 and 2, there is also minor heterogeneity within each virus type. Individual strains have been identified on the basis of changes, compared with B95.8, ranging from single base mutations to extensive deletions. While infection with multiple strains of EBV was originally thought to be confined to immunologically compromised patients, more recent studies demonstrate that normal healthy seropositive individuals can be infected with multiple EBV isolates and that their relative abundance and presence may vary over time.

The possible contribution of EBV strain variation to virus-associated tumors such as BL and NPC which is common in China and South-East Asia, remains unknown. A number of studies have failed to establish an epidemiological association between EBV strain variation and disease concluding that the specific EBV gene polymorphisms detected in virus-associated tumors occur with similar frequencies in EBV isolates from healthy virus carriers from the same geographic region. However, these studies focussed on specific regions of the EBV genome rather than comparing the entire viral DNA sequence. More recent work using hybrid capture technologies and next generation sequencing (NGS) have facilitated routine sequencing from primary material and have confirmed that, while there is a high level of overall similarity of tumor-derived EBV strains with the prototypical EBV genome, variation exists in viral genes that might result in functional differences that influence tumor development. A recent case-control study of EBV strain variation in NPC has found distinct polymorphisms in the virus genome associated with the tumor thereby defining a high risk EBV variant that could be used to screen at risk populations and develop more precise diagnostic and prognostic tests.

Immune Response

EBV elicits both humoral and cell-mediated immune responses in infected hosts. Primary infection with EBV is associated with the rapid appearance of antibodies to replicative antigens such as VCA, EA and MA (gp350/220) with a later serological response to the EBNA proteins. In IM these responses are exaggerated and are accompanied by autoantibodies such as rheumatoid factor as well as a heterophile antibody response directed against antigens on the surface of sheep erythrocytes. These autoantibodies are the result of EBV-induced polyclonal B cell activation. In the chronic asymptomatic virus carrier, antibodies to VCA, MA and EBNA are found, the titers of which remain remarkably stable over time. Of these antibodies those against MA are particularly important

as they have virus neutralizing ability and can also mediate antibody-dependent cellular cytotoxicity. The levels of these EBV-specific antibodies are elevated in different EBV-associated diseases.

As with other persistent viruses, cell-mediated immunity plays an important role in controlling EBV infection. Primary EBV infection elicits a robust cellular immune response and the lymphocytosis observed in IM is a consequence of the hyperexpansion of cytotoxic CD8⁺ T cells with reactivities against both latent and lytic viral antigens. These reactivities are subsequently maintained at high level (up to 5% of the total circulating pool) in the CD8⁺ memory T cell pool. An EBV-specific CD4⁺ T cell response also contributes towards the control of EBV infection and, along with the CD8⁺ response, appears to be important in preventing the uncontrolled proliferation of EBV-infected B cells. Thus, impairment of the T cell response, either by immunosuppressive drug therapy or HIV infection, is responsible for the development of polyclonal lymphoproliferations that can progress to frank monoclonal non-Hodgkin's lymphomas (see below). These lesions can be controlled by adoptive therapy with EBV-specific T cells. The growth and survival of BL, NPC and Hodgkin's lymphoma (HL) in immunocompetent individuals implies that the tumor cells can evade EBV-specific T cell surveillance. This may be achieved by restricting EBV latent gene expression to those viral proteins not efficiently recognized by the T cell responses (i.e., away from expression of the highly immunogenic EBNA3 family), and/or by downregulation of target cell molecules required for immune recognition such as MHC class I/II and the antigen processing machinery which is mediated by the production of immunosuppressive EBV-encoded proteins such as those expressed during the EBV lytic cycle that inhibit antigen processing (e.g., BNLF2a), suppress macrophage function (vIL-10) and disrupt MHC class II presentation (e.g., BGLF5).

EBV-Associated Diseases

Infectious Mononucleosis

Primary infection with EBV in childhood is usually asymptomatic but when delayed until adolescence or early adulthood can manifest clinically as IM, a self-limiting lymphoproliferative disease characterized by a quartet of symptoms – sore throat, cervical lymphadenopathy, fever and fatigue. IM is associated with the massive expansion of activated CD8⁺ T cells that are reactive predominantly to EBV lytic cycle antigens with some limited responses against latent antigens. Immunodominant CD8⁺ T cell responses in IM are focussed on immediate early proteins (e.g., BZLF1 and BRLF1) and early proteins (e.g., BALF2, BMRF1, BORF2) and these reactivities are subsequently maintained in the CD8⁺ T cell memory pool at high levels. Activated CD4⁺ T cell responses specific for EBV epitopes in EBV lytic and latent antigens are also elevated in IM patients but fall rapidly as the disease resolves. IM appears to result in the global down-regulation of the alpha chain of the IL-15 receptor on T cells and NK cells, an effect which lasts for years after acute infection and may influence other disease risks. The incidence of IM is low in developing countries where asymptomatic primary infection predominantly occurs in childhood. In certain poorly defined situations IM-like symptoms can persist resulting in chronic active EBV infection (CAEBV) associated with elevated antibody titers to virus lytic antigens but low titers to the EBV-encoded nuclear antigens (EBNAs) along with elevated EBV infection in circulating T cells and NK cells. CAEBV appears to originate from an EBV-infected lymphoid progenitor and is associated with intragenic deletions in the EBV genome.

Multiple Sclerosis

Multiple Sclerosis (MS) is a chronic inflammatory demyelinating disease of the central nervous system (CNS), causing severe progressive disability. Genetic and environmental factors contribute to the development of MS and many studies implicate EBV including the universality of EBV infection in MS patients, elevated levels of antibodies against EBV in serum (particularly against the EBNA1 protein) during the course of MS and often years prior to the onset of the disease. Impairment of the EBV-specific CD8⁺ T-cell response, increased shedding of EBV from saliva and accumulation of EBV-infected B cells and plasma cells in the brain are also found in MS patients. It has been hypothesized that a defect in the CD8⁺ T cell control of EBV predisposes to MS by allowing EBV-infected autoreactive B cells and plasma cells to accumulate in the CNS. A recent phase I clinical trial using autologous EBV-specific T cell therapy demonstrated some clinical improvement in MS patients.

Lymphomas in Immunosuppressed Individuals

Patients with primary immunodeficiency diseases such as X-linked lymphoproliferative syndrome (XLP) and Wiscott-Aldrich syndrome are at increased risk of developing EBV-associated lymphomas. Because these tumors are extremely rare little is known about the precise contribution of EBV and the associated pattern of viral gene expression in these lymphomas. Mortality from XLP is high with around 50% of patients developing fatal IM after primary infection with EBV and an additional 30% of patients developing malignant lymphomas. The defect responsible for XLP is mutation of an adaptor molecule, SAP, which mediates signaling in a wide range of immune cells and is involved, via its interaction with the SLAM family of receptors, in both innate and adaptive immune reactions.

Allograft recipients receiving immunosuppressive therapy and patients with AIDS are also at increased risk for development of EBV-associated post-transplant lymphoproliferative disease (PTLD) and lymphomas. The incidence of non-Hodgkin's lymphoma (NHL) in AIDS patients, prior to the advent of highly active antiretroviral therapy, was increased approximately 60-fold compared

to the normal population. Around 60% of these tumors were large B cell lymphomas like those found in allograft recipients, 20% were primary brain lymphomas and 20% were of the Burkitt's lymphoma (BL) type. EBV infection was present in approximately 50% of AIDS-related NHL, nearly all the primary brain and Hodgkin's lymphomas, and in around 40% of the BL tumors. These NHLs range from polyclonal EBV-driven lymphoproliferations, much like that observed in vitro in virus-transformed LCLs (latency III, Figs. 2–4), to aggressive monoclonal non-Hodgkin's B cell lymphomas (NHLs) in which additional cellular genetic changes (e.g., mutation of p53) are present. The incidence of both PTLD and NHL B cell lymphomas in allograft recipients varies with the type of organ transplanted and with the type of immunosuppressive regimen used. PTLD in solid organ recipients are derived predominantly from recipient B cells whereas PTLD following hematopoietic stem cell transplantation are mainly of donor origin. Allogeneic bone marrow or solid organ transplantation into EBV seronegative children is a particular risk factor for the development of these lesions. A proportion of these lesions in post-transplant patients resolve in response to a reduction in immunosuppression or to targeted therapies such as anti-CD20 monoclonal antibody (rituximab) or adoptive EBV-specific T cells.

Diffuse Large B-Cell Lymphomas

EBV-positive diffuse large B-cell lymphomas (DLBCL) are monoclonal B-cell tumors that occur in patients over the age of 50 with no evidence of immunodeficiency or history of lymphoma. These lymphomas are thought to arise as a consequence of immunosenescence, present as two different morphologic variants (monomorphic and polymorphic) and are associated with extranodal involvement with the gastrointestinal tract, skin and bone marrow being the most commonly affected sites. EBV-positive DLBCL is more common in Asia than the West, representing around 8%–10% of DLBCL in Asian countries and 2%–5% of DLBCL in Western countries. The EBV-positive DLBCLs have a distinctive molecular profile with increased activity of certain cell signaling pathways (e.g., NF- κ B, STAT3, MAP kinase), probably as a consequence of EBV latent gene expression which is either of the latency II (EBNA1, LMP1, LMP2A) or latency III type. Intragenic deletions of various regions of the EBV genome have been found in a proportion of DLBCL.

EBV-positive DLBCL is an aggressive tumor with worse prognosis than the EBV-negative disease, particularly in Asian patients. EBV-positive DLBCL has also been found in younger patients but here the disease is very similar to EBV-negative DLBCL in terms of clinical outcome.

Burkitt's Lymphoma

The endemic form of BL which is found in areas of equatorial Africa and New Guinea represents the most common childhood cancer (peak age 7–9 years) in these regions with an incidence of up to 10 cases per 100,000 people per year. This high incidence of BL is associated with holoendemic malaria thus accounting for the climatic variation in tumor incidence first recognized by Dennis Burkitt. More than 95% of these endemic BL tumors (eBL) are EBV-positive compared with 20% of the low incidence, sporadic form of BL (sBL) which occurs worldwide. In areas of intermediate BL incidence, such as Algeria and Malaysia, the increased number of cases correlates with an increased proportion of EBV-positive tumors. Malaria is thought to contribute to endemic BL by inducing an intense polyclonal B cell activation, thereby increasing the target population for EBV infection, along with transient impairment of T cell immunity.

Seroepidemiological studies have demonstrated elevated antibody titers to EBV capsid antigen (VCA) and early antigens (EA) in BL patients compared to children without the tumor. These elevated antibody titers have been found to precede the development of BL and can therefore be used to screen "at risk" individuals. While sBL in high-income countries responds well to intensive chemotherapy, such regimens are not feasible or safe in low-income countries where late presentation is common and severe treatment-associated toxicities are difficult to manage. Recently modified protocols using cyclophosphamide and doxorubicin in eBL have resulted in much improved disease-free survival. Nevertheless, treatment failure is common and often associated with incomplete treatment, treatment-related mortality, relapse or progression of disease.

The pattern of EBV gene expression in BL is generally restricted to EBNA1 and the non-polyadenylated EBER transcripts (latency I) although broader virus gene expression involving the EBNA3 proteins and the viral Bcl-2 homolog BHRF1 has been observed in around 15% of tumors (Figs. 2–4). These so-called Wp-restricted BLs (named after the alternative virus Wp promoter that is used in these cells to drive EBV gene expression) carry an EBNA2-deleted EBV genome and, as a consequence of BHRF1 expression, are more resistant in vitro to apoptosis induced by cytotoxic drugs.

A consistent feature of BL tumors, irrespective of geographical location or EBV status, is chromosome translocations involving the long arm of chromosome 8 (8q24) in the region of the c-myc proto-oncogene and either chromosome 14 in the region of the immunoglobulin (Ig) heavy-chain gene or, less frequently, chromosomes 2 or 22 in the region of the Ig light-chain genes. BL tumors display a phenotype reminiscent of germinal center centroblasts with lower mutational frequencies in cancer driver genes such as *MYC*, *ID3*, *TCF3*, *CCND3* and *TP53* as compared to sBL. In contrast, a greater overall mutational load including higher frequency of mutations in *ARID1A* and loss of the regulatory role of PTEN have been found in eBL versus sBL. Some of these effects can be attributed to the influence of EBV-encoded genes in driving proliferation and survival in eBL without the need for mutations in other genes such as *TCF3* and *ID3*. Endemic BLs infected with EBV-2 appear to have a higher proportion of somatic mutations (similar to the levels observed in sBL) than those containing EBV-1. Thus, there are clear genetic and molecular distinctions between EBV-positive and EBV-negative BL that reinforce the crucial role of EBV in eBL pathogenesis.

Hodgkin's Lymphoma

Epidemiological studies originally suggested a possible role for EBV in the etiology of Hodgkin's lymphoma (HL). Thus, elevated antibody titers to EBV antigens have been detected in HL patients and these are present before the diagnosis of the disease. Furthermore, there is an increased risk of HL following IM. EBV has been demonstrated in around 40% of HL cases with both viral DNA, RNA and virus-encoded latent proteins (EBNA1, LMP1, LMP2 – latency II) localized to the malignant component of HL, the so-called Hodgkin's and Reed-Sternberg (HRS) cells (**Fig. 4**). These malignant cells are of germinal center origin but lack a functional B cell receptor, implicating EBV infection in the rescue of these so-called crippled cells from their usual apoptotic fate. Genetic mutations in components of various cell signaling pathways (e.g., NF- κ B, JAK/STAT, PI3K/Akt) are consistently found in EBV-negative HL as compared to the EBV-positive form of the disease. Thus, in EBV-positive HRS cells, the virus-encoded LMP1 oncoprotein is apparently sufficient to drive aberrant activation of these pathways in the absence of cellular mutations. Thus, EBV and cellular mutations probably provide mutually exclusive means to the same pathogenic end-point. The association of HL with EBV is age-related; pediatric and older adult cases are usually EBV-associated whereas HL in young adults is less frequently virus-positive. The proportion of EBV-positive HL in developing countries is high consistent with a greater incidence of HL in children and more frequent prevalence of the mixed cellularity histiotype; the histological subtype of HL in which EBV infection is most frequently detected. Although the incidence of HL is relatively low (1–3/100,000 per year), this tumor is not geographically restricted making its association with EBV significant in world health terms.

EBV-Associated T Cell and NK Cell Lymphomas

EBV-positive monoclonal lymphomas of either CD4+ or CD8+ T cell origin are more frequently found in Southeast Asian populations arising as a consequence of either virus-associated haemophagocytic syndrome (VAHS) or in the setting of CAEBV. There is also a distinct spectrum of EBV-positive T/NK lymphoproliferative disorders ranging from cutaneous lymphoid proliferations to aggressive extranodal T/NK lymphomas such as lethal midline granuloma, an erosive lesion of the nasal cavity. These lesions often develop in individuals with CAEBV, exhibit a latency II pattern of EBV latency with some evidence of early virus gene expression and frequently carry EBV genomes with intragenic deletions particularly over the BART miRNA region of the virus genome.

Nasopharyngeal Carcinoma

The association of EBV with undifferentiated NPC (WHO types II and III) was first suggested by serological evidence and then confirmed by the demonstration of EBV DNA in NPC biopsy material. NPC is particularly common in areas of China and South-East Asia, reaching a peak age-standardized incidence of around 25 cases per 100,000 per year. Incidence rates are particularly high in Cantonese males highlighting an important genetic predisposition as well as a role for environmental cofactors such as dietary components (e.g., salted fish). NPC tumors are characterized by a prominent stroma and the interaction between these stromal cells (activated lymphocytes, tumor-associated macrophages, cancer-associated fibroblasts etc.) and adjacent carcinoma cells appears to be crucial for the continued growth of the malignant component. EBV latent gene expression in NPC is restricted to EBNA1, the LMP2A/B proteins, the EBER, BARF1 and the BART miRNAs with a variable proportion of tumors also expressing the oncogenic LMP1 protein (latency II) (**Fig. 2**). While latent EBV infection appears to be a key driver in the development and progression of NPC, more detailed analysis of virus gene expression by RNA sequencing reveals a more complex pattern with expression of lytic genes associated with EBV replication. The precise levels and consistency of such EBV lytic gene expression in NPC remains unknown as does whether this has any impact on NPC pathogenesis.

Extensive serological screening for EBV-specific antibody titers in high incidence areas, in particular IgA antibodies to VCA and EA, have proved useful in diagnosis and in monitoring the effectiveness of therapy. The quantitative analysis of cell-free tumor-derived EBV DNA in the plasma of patients with NPC using real-time PCR (Q-PCR) is of both diagnostic and prognostic utility. A recent study has demonstrated that Q-PCR for plasma EBV DNA levels can also be used as a primary screening test, detecting early stage NPC which is more responsive to treatment.

The precise contribution of EBV infection to the development of NPC remains unclear. The presence of monoclonal EBV episomes in NPC tumor cells indicates that virus infection precedes the clonal expansion of the malignant cell population. However, the lack of epithelial EBV infection in normal nasopharyngeal biopsies from individuals at high risk of developing NPC suggests that epithelial infection may not be the initiating event in virus-associated carcinogenesis. EBV infection as detected by *in situ* hybridization to the EBER RNAs is found in high grade (severe dysplastic and carcinoma *in situ*) preinvasive lesions in the nasopharynx but not in low grade disease or histologically normal nasopharyngeal epithelium. Both the high grade and carcinoma *in situ* lesions carry monoclonal EBV genomes. A scheme has been proposed whereby loss of heterozygosity occurs early in the pathogenesis of NPC possibly as a result of exposure to environmental co-factors such as dietary components (i.e., salted fish) creating low grade preinvasive lesions that after additional genetic and epigenetic events become susceptible to EBV infection (**Fig. 5**). Once infected, EBV latent genes provide growth and survival benefits resulting in the development of NPC. Additional genetic and epigenetic changes occur after EBV infection and contribute to metastatic disease. Analysis of the genomic landscape in NPC has confirmed a role for chromatin modification in the carcinogenic process and identified additional somatic events including mutations in the ERBB-PI3K signaling pathway. A crucial role for the NF- κ B pathway in NPC pathogenesis has been recently identified and this appears to be either a consequence of LMP1 expression or of genomic aberrations in negative regulators of the NF- κ B signaling.

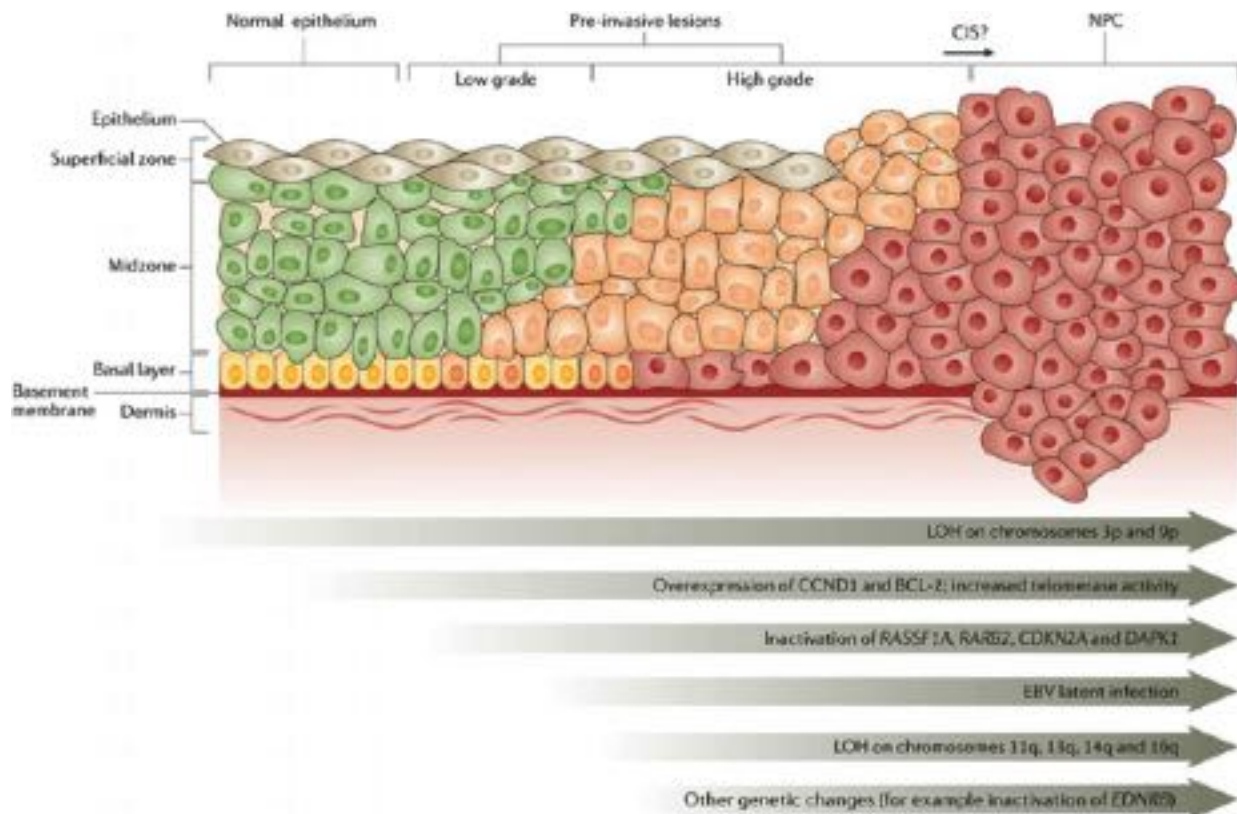


Fig. 5 Schematic representation of the pathogenesis of NPC. In nasopharyngeal carcinoma (NPC), the tumor cells carry monoclonal viral genomes indicating that EBV infection must have occurred prior to expansion of the malignant cell clone: EBV infection is detected in high-grade pre-invasive lesions in both NPC but not in low-grade disease. Multiple genetic changes have been found in NPC, with frequent chromosomal deletions and promoter hypermethylation of specific genes on chromosomes 3p, 9p and 11q. Deletions in both 3p and 9p have been identified in low-grade EBV-negative dysplastic lesions and in normal nasopharyngeal epithelium from individuals who are at high risk of developing NPC. *In vitro* studies demonstrate that CCND1 over-expression (a consequence of cyclin-dependent kinase inhibitor 2A [*CDKN2A* which encodes p16^{INK4A}] deletion on chromosome 9p and amplification of the *CCND1* locus on chromosome 11q) facilitates persistent EBV infection of immortalized nasopharyngeal epithelial cells. Increased expression of the anti-apoptotic protein BCL-2 and elevated telomerase activity in low grade dysplastic lesions may also promote the establishment of latent EBV infection. A scheme has been proposed whereby loss of heterozygosity (LOH) occurs early in the pathogenesis of NPC possibly as a result of exposure to environmental co-factors such as dietary components (i.e., salted fish) creating low grade preinvasive lesions that after additional genetic and epigenetic events become susceptible to stable EBV infection. Once infected, EBV latent genes provide growth and survival benefits resulting in the development of NPC. Additional genetic (e.g., LOH on chromosomes 11q, 13q, 14q and 16q) and epigenetic changes occur after EBV infection and contribute to metastatic disease. The precise timing of epigenetic events in the pathway of NPC development is unknown although some changes such as RASSF1A inactivation may occur early while others (e.g., inactivation of RARβ2, p14ARF and DAPK) may be influenced by the ability of EBV to enhance the activity of the methylation machinery. Recent analysis of the genomic landscape in NPC has confirmed a role for chromatin modification in the carcinogenic process and identified additional somatic events including mutations in the ERBB-PI3K signaling pathway. The pathogenesis of EBV-associated GC is similar to NPC with pre-malignant lesions occurring before EBV infection and a distinctive hypermethylation phenotype. CIS, carcinoma *in situ*; EDNRB, endothelin receptor B. Reproduced from Young, L.S., Yap, L.-F., Murray, P.G., 2016. Epstein-Barr virus: Over 50 years and still providing surprises. *Nature Reviews Cancer* 16, 789–802.

EBV infection is also associated with the more differentiated keratinizing form of NPC (WHO type I) particularly in those geographical regions with a high incidence of the type III tumor Carcinomas with similar features to undifferentiated NPC have been described at other sites (e.g., tonsils, uterine cervix) but, other than stomach tumors, these so-called undifferentiated carcinomas of nasopharyngeal type (UCNT) are not consistently EBV positive. EBV has been demonstrated in thymic epithelial tumors from Chinese but not Western patients. Salivary gland UCNTs are EBV-associated in Greenland Eskimos and Chinese but not in Caucasian patients.

Gastric Carcinoma and Other Epithelial Tumors

EBV infection is also present in around 10% of typical gastric adenocarcinomas, accounting for up to 90,000 cases per year worldwide. These EBV-GC tumors resemble NPC in carrying monoclonal EBV genomes, having a restricted pattern of EBV gene expression (EBERs, EBNA1, LMP2A, BARTs and BARF1) and in the appearance of virus infection as a relatively late event in the

carcinogenic process. EBV-GCs have distinct phenotypic and clinical characteristics compared to EBV-negative GC tumors including loss of p16 expression, p73 promoter methylation, wild type p53, a different pattern of allelic loss and improved patient survival. Comprehensive genomic and molecular characterization of GC tumors has confirmed that EBV-GCs are a distinctive entity with a high CpG island methylator (CIMP) phenotype and apparent driver mutations in the PIK3CA and ARID1A genes.

Control and Prevention of EBV Infection

While a number of nucleoside (e.g., acyclovir, ganciclovir, famciclovir) and nucleotide (e.g., cidofovir, foscavir) analogs are approved for the treatment of herpes simplex virus (HSV-1 and HSV-2), varicella-zoster virus and human cytomegalovirus, none are used routinely in the treatment of acute EBV infection. This may reflect difficulties in diagnosing IM, its long incubation period and, more significantly, the immunopathology of IM which results in symptoms that are related to the immune response to EBV-infected B lymphocytes rather than virus replication. Nevertheless, antivirals such as acyclovir and ganciclovir have been used in patients with CAEBV and in immunosuppressed patients where such prophylaxis reduces the incidence of PTLT.

The conventional approaches of chemotherapy and radiotherapy are used to treat primary EBV-associated tumors. It is only in the context of relapsed, refractory or metastatic cancer that novel therapeutic strategies specifically targeting EBV proteins or exploiting the presence of the virus in malignant cells are being developed. Demethylating agents such as 5'-azacytidine or histone deacetylase (HDAC) inhibitors such as SAHA are able to induce the EBV lytic cycle including virus-encoded kinases (EBV thymidine kinase and BGLF4, a protein kinase) that can phosphorylate the nucleoside analog ganciclovir to produce its active cytotoxic form. These agents, which are continuing to be assessed in clinical trials, also induce expression of the more immunogenic EBV latent genes in NPC and HL as well as creating a tumor immune microenvironment that supports enhanced anti-tumor immune responses. Other experimental approaches are based on targeting individual EBV latent proteins either indirectly by inhibiting signaling pathways activated, for instance, by LMP1 (e.g., NF- κ B, PI3-kinase) or more directly using specific inhibitors such as siRNA or dominant negative molecules. More recently specific inhibitors of the EBNA1 protein have been developed and these are showing therapeutic promise in pre-clinical models.

The use of adoptive EBV-specific T cell therapy for the treatment of existing PTLT and in the prophylactic setting has been extremely successful. This approach is also showing signs of some clinical efficacy in NPC when T cells are enriched for reactivities to subdominant targets (e.g., LMP2A and EBNA1). More direct vaccine approaches to treat patients with EBV-associated tumors or to prevent disease development are being examined including: (1) whole EBV latent antigens delivered by a virus vector or in autologous dendritic cells; (2) peptide or polytope vaccination and (3) prophylactic vaccination against MA (gp350/220) on the virion surface or virus-like particles to generate a neutralizing antibody response.

Future Perspectives

EBV was discovered over 50 years ago and its DNA was fully sequenced in 1984. It remains the most common persistent virus infection in humans – testimony to the intimate interaction between EBV and the immune host. This relationship relies on the ability of EBV to persist in the memory B cell pool of normal healthy individuals and corruption of this process leads to the development of virus-associated lymphomas. The role of EBV infection in NPC is less clear and may be a consequence of the aberrant establishment of *virus* latency in epithelial cells that have already undergone pre-malignant genetic changes.

The development of more efficient *in vitro* systems for studying EBV infection and replication in different cell types along with the use of recombinant forms of EBV is shedding light on the complex interplay between the virus and host. Improvements in virus genome sequencing technologies as well as the ability to clone wild-type EBV strains for biological characterization will facilitate a better understanding of the natural history of EBV infection and how the virus contributes to the oncogenic process. These systems, along with detailed analysis of EBV-associated tumors, are already providing insights into the interaction of the virus with the tumor microenvironment and how this influences the local cytokine milieu to promote tumor growth and immune evasion. There are many interesting aspects of EBV biology that remain to be understood including the role of virus-encoded microRNAs and the possible contribution of lytic cycle antigens to virus persistence and oncogenesis. The challenge will be to exploit these new mechanistic insights both to gain a better understanding of EBV infection *in vivo* and to develop novel therapies for treating virus-associated disease.

Further Reading

- Dawson, C.W., Port, R.J., Young, L.S., 2012. The role of the EBV-encoded latent membrane proteins LMP1 and LMP2 in the pathogenesis of nasopharyngeal carcinoma (NPC). *Seminars in Cancer Biology* 22, 144–153.
- Farrell, P.J., 2015. Epstein-Barr virus strain variation. *Current Topics in Microbiology and Immunology* 390, 45–69.
- Feederle, R., Klinke, O., Kutikhin, A., *et al.*, 2015. Epstein-Barr virus: From the detection of sequence polymorphisms to the recognition of viral types. *Current Topics in Microbiology and Immunology* 390, 119–148.
- Fitzsimmons, L., Kelly, G.M., 2017. EBV and apoptosis: The viral master regulator of cell fate? *Viruses* 9 (11), 339.
- Kimura, H., Cohen, J.I., 2017. Chronic active Epstein-Barr virus disease. *Frontiers in Immunology* 8, 1867.
- Lam, W.K.J., Chan, K.C.A., Lo, Y.M.D., 2019. Plasma Epstein-Barr virus DNA as an archetypal circulating tumour DNA marker. *The Journal of Pathology* 247, 641–649.

- Longnecker, R.M., Kieff, E., Cohen, J.I., 2013. Epstein-Barr virus. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia: Lippincott, Williams & Wilkins, pp. 1898–1959.
- Shannon-Lowe, C., Rickinson, A.B., Bell, A.I., 2017. Epstein-Barr virus-associated lymphomas. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 372, 1732.
- Taylor, G.S., Long, H.M., Brooks, J.M., Rickinson, A.B., Hislop, A.D., 2015. The immunology of Epstein-Barr virus-induced disease. *Annual Review of Immunology* 33, 787–821.
- Thorley-Lawson, D.A., Allday, M.J., 2008. The curious case of the tumour virus: 50 years of Burkitt's lymphoma. *Nature Reviews Microbiology* 6, 913–924.
- Thorley-Lawson, D.A., Hawkins, J.B., Tracy, S.I., Shapiro, M., 2013. The pathogenesis of Epstein-Barr virus persistent infection. *Current Opinion in Virology* 3, 227–232.
- Tsao, S.W., Tsang, C.M., To, K.F., Lo, K.W., 2015. The role of Epstein-Barr virus in epithelial malignancies. *The Journal of Pathology* 235, 323–333.
- Vrzalikova, K., Sunmonu, T., Reynolds, G., Murray, P., 2018. Contribution of Epstein-Barr virus latent proteins to the pathogenesis of classical Hodgkin lymphoma. *Pathogens* 7 (3), 59.

Equine Herpesviruses (*Herpesviridae*)

Gisela Soboll Hussey, Michigan State University, East Lansing, MI, United States
Nikolaus Osterrieder and Walid Azab, Free University of Berlin, Berlin, Germany

© 2021 Elsevier Ltd. All rights reserved.

Introduction

The first disease attributed to a herpesvirus of horses, the agent that is now referred to as equid herpesvirus type 1 (EHV-1; equine abortion virus), was documented at the University of Kentucky Agricultural Experiment Station in Lexington, KY. EHV-1 was first shown to be associated with spontaneous abortions in pregnant mares in 1932. In 1941 and 1949, equine abortions and/or paresis were found to be associated with mild respiratory disease with clinical signs similar to those associated with equine influenza virus infections. EHV-1 was also shown to be the etiological agent of epizootic respiratory disease in young horses. Based on these investigations, the disease was termed viral rhinopneumonitis, and the agent was called equine rhinopneumonitis virus. In the 1980s, however, the two disease manifestations, viral abortion and viral rhinopneumonitis, were shown to be caused by two closely related but clearly distinct viruses that are now classified as EHV-1 (equine abortion virus) and EHV-4 (equine rhinopneumonitis virus). EHV-1 is the cause of respiratory disease in young animals, as well as equine abortions, highly fatal neurological disease, fulminating neonatal pneumonitis, ocular chorioretinal lesions and very rarely, an exanthematous condition involving the external genitalia of the mare.

Equid herpesvirus type 2 (EHV-2) was first isolated from horses in 1963. The cytopathology caused by this virus closely resembled that of cytomegalovirus infections which were first described in 1921 in humans. EHV-2 is a ubiquitous slow-growing virus that infects horses at a very young age (1–2 years) and establishes a life-long chronic infection such that the virus is continually shed. To date, no major disease has been attributed to EHV-2. An association of EHV-2 with chronic throat infections (the lumpy bumpies) or recurrent eye disease (keratoconjunctivitis) has been established; however, causal relationships have not been unequivocally proven. In addition, EHV-2 may be a cofactor in EHV-1 and/or EHV-4 infections in that it may be able to modulate EHV-1 and EHV-4 replication by an immunosuppression causing general malaise or by modulation of EHV-1 gene expression through EHV-2-specific transcriptional transactivators.

Equid herpesvirus type 3 (EHV-3; equine coital exanthema virus, ECE virus) was first isolated in 1968, concurrently in Canada, Australia and the US. EHV-3 is the etiological agent of equine coital exanthema, a generally mild genital infection of both mares and stallions that is transmitted venereally. Clinical signs include the formation of papules, vesicles, pustules, and ulcers on genital organs. Although mild, EHV-3 infection has negative impact on equine breeding and can cause temporary disruption of the mating season, artificial insemination, and embryo transfer. Sporadic cases of EHV-3 infection with periodic virus shedding and virus isolation are still reported.

Equid herpesvirus type 4 (EHV-4; equine rhinopneumonitis virus), is associated mainly with respiratory disease, but has also been associated occasionally with equine abortions. EHV-4 is most commonly detected in weaned foals and yearlings. Adult horses are also susceptible to EHV-4 infection, however, they show milder clinical signs or infections remain subclinical.

Equid herpesvirus type 5 (EHV-5) is closely related to EHV-2 and is frequently detected in (healthy) horse populations worldwide. In 1970, EHV-5 was first isolated from the nasal cavity of two horses in quarantine in Australia. It has been thought that the virus is harmless and cause no clinical disease but there is mounting evidence that EHV-5 is associated with equine multinodular pulmonary fibrosis (EMPF). More definite proof for direct causation of EHV-5 and EMPF and a clearer understanding of the pathogenesis is needed, however, before a solid conclusion can be reached.

Equid herpesvirus type 9 (EHV-9), previously referred to as gazelle herpesvirus 1, was first isolated from a captive Thomson's gazelle suffering neurological signs in 1993 in Japan. It has been shown to be closely related, but clearly distinct, from EHV-1. The natural host and the true host range of EHV-9 is still unknown but the virus has a relatively wide host range and infects equid and non-equid species in the wild and in captivity. EHV-9 was used for experimental infection of a number of animal species, including dogs, cats, mice, cattle, hamsters, goats, pigs, macaques and marmosets, in all of which it causes a highly lethal neurological disorders.

Classification

EHV-1, EHV-2, EHV-3, EHV-4, EHV-5, and EHV-9 are all members of the *Herpesviridae* but belong to two different subfamilies (*Alpha-* and *Gammaherpesvirinae*) in this large virus family. EHV-6, 7 and 8 are also referred to as asinine herpesviruses 1, 2, and 3, respectively (AsHV-1 to AsHV-3), and will not be discussed in this article. The morphology of all six members is typical of the herpesviruses in that they are enveloped, contain an icosahedral capsid, and have a proteinaceous layer, the so-called tegument, which surrounds the nucleocapsid. Equine herpesviruses are composed of six distinct species: (1) EHV-1 is the major equine pathogen causing abortions, respiratory illness, neurological disease and ocular disease; (2) EHV-2 establishes asymptomatic long-term persistent infection and might be associated with keratoconjunctivitis; (3) EHV-3 is the causative agent of mild pro genital

Table 1 The equine herpesviruses

<i>Properties of the viral genome</i>				
<i>Member</i>	<i>Subfamily</i>	<i>Isomers</i>	<i>G + C ratio (%)</i>	<i>Clinical manifestations</i>
EHV-1	<i>α-herpesvirinae</i>	2	57	abortion, respiratory infection, paralysis
EHV-2	<i>γ-herpesvirinae</i>	1	58	chronic throat infection; keratoconjunctivitis
EHV-3	<i>α-herpesvirinae</i>	2	68.1	equine coital exanthema
EHV-4	<i>α-herpesvirinae</i>	2	50.5	respiratory infections
EHV-5	<i>γ-herpesvirinae</i>	1	55	equine multinodular pulmonary fibrosis
EHV-9	<i>α-herpesvirinae</i>	2	56	encephalitis in non-definitive hosts

exanthema; (4) EHV-4 is a major respiratory pathogen that differs significantly from EHV-1 at the DNA level, and is associated occasionally with equine abortions; (5) EHV-5 may be associated with equine multinodular pulmonary fibrosis; (6) EHV-9 can cause subclinical encephalitis in horses under experimental conditions. Experimental and accidental natural infections of non-equid species are mostly lethal due to severe neurological disorders. In this latter respect, EHV-9 possesses biological properties very akin to the alphaherpesvirus of pigs, pseudorabies virus.

EHV-1, 3, 4, and 9 are *Alphaherpesvirinae* members and belong to the genus *Varicellovirus*. EHV-2 and EHV-5 are members of the *Gammaherpesvirinae* and are classified into the genus *Percavirus*.

Virion Structure

Infectious EHV-1 and EHV-4 particles contain an envelope that is mainly composed of glycoproteins. To date, twelve glycoproteins have been identified and an association of gB, gC, gD, gE, gH, gL, gM, gp2, gI, gK, ORF10 (UL49.5) with purified virions has been demonstrated. The glycoproteins were shown in other alphaherpesviruses to be involved in virus attachment and binding, penetration, egress and cell-to-cell spread of infection as well as immune evasion. Less is known for proteins that make up the third component of the mature virion, the tegument. In all cases, EHV-1 and EHV-4 encode homologs to most of the HSV-1 tegument proteins that have been surmised to exert the same function.

Three different EHV-1 capsid species were identified and probably correspond to the capsid forms found in HSV-1, which are designated type A, B, and C capsids. The EHV-1 capsid species were designated: (1) L capsids, which appear to be empty capsids; (2) I capsids, which possess an electron lucent, immature core structure in the shape of a cross; and (3) H capsids, which contain an electron dense, mature core. All three capsids appear at approximately 6 h postinfection. I capsids are believed to be a major precursor in the formation of mature capsids. The major capsid protein has an apparent size of 148 kD, while other structural proteins have also been identified.

Genomes

All six equine herpesviruses contain a linear, double-stranded DNA genome ranging between 140 and 184 kbp. The reported approximate molecular masses of the genomes are: EHV-1 (Ab4 strain), 150 kbp; EHV-2, 184 kbp; EHV-3, 152 kbp; EHV-4, 145 kbp; EHV-5, 182 kbp; and EHV-9, 148 kbp. The genomes of EHV-1, EHV-3, EHV-4, and EHV-9, exist in two isomeric forms, since the short region (S) can invert relative to the fixed orientation of the unique long region (UL). The S region is comprised of a central segment of unique sequences (US) bracketed by a pair of inverted sequences (IR). In contrast, the genomes of EHV-2 and EHV-5 exist as one isomer and are comprised of a large central segment of unique sequences that is bracketed by a pair of direct repeat sequences. The EHV-2 genome contains direct terminal repeats (TR) of 17,553 and 18,332 bp, respectively. While the TR of EHV-5 is much smaller (10 bp) in each terminal. Other characteristics of the genomes of EHV-1 and EHV-9 are shown in [Table 1](#) and [Fig. 1](#).

There are varying degrees of homology at the DNA level among the six equine herpesviruses. The sequences shared by the EHV-1, EHV-3, EHV-4, and EHV-9 appear to be arranged colinearly. EHV-1 and EHV-9 exhibit the highest degree of nucleotide identity, ranging from 86% to 95%. EHV-1 and EHV-4 have 55%–84% sequence identity at the DNA level. EHV-1, EHV-4, and EHV-9 are antigenically very closely related and antibodies can cross-react. EHV-3 is the most divergent among the equid alphaherpesviruses and only shares an overall identity of approximately 60% with EHV-1, EHV-4 and EHV-9 ([Fig. 2](#)). EHV-2 and EHV-5 show approximately 62% identity on the DNA level.

The EHV-1 and EHV-4 genomes have been cloned as bacterial artificial chromosomes (BAC). These infectious clones provide a basis for rapid and efficient mutagenesis in prokaryotic cells. A variety of recombinant viruses that lack non-essential genes or portions of an essential gene or with swapped genes have been generated and used in experiments to elucidate the functions of these genes in virus replication and/or pathogenesis.

The complete EHV-1 (Ab4 strain) genome is 150,223 base pairs in size, while that of EHV-4 is 145,597. The EHV-1 genome has at least 78 unique corresponding genes [63 ORFs in the UL region, 6 ORFs in each inverted repeat (IR and TR), and 9 ORFs in the US segment]; however, the actual number of open reading frames (ORFs) in the genome is 84 due to duplication of the six genes

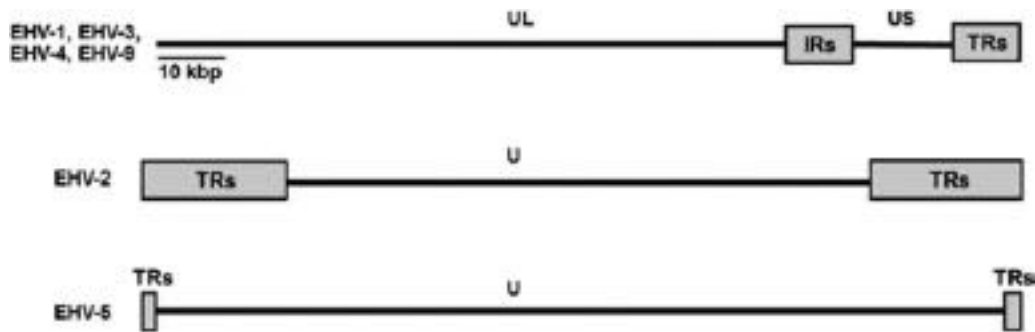


Fig. 1 Schematic presentation of the genomic organization of equine herpesvirus genomes.

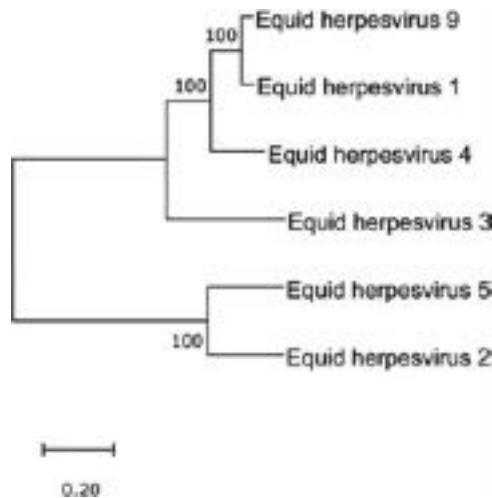


Fig. 2 Phylogeny of equid herpesviruses. Genetic relationships of four equid *Alphaherpesviruses* and two equid *Gammapherpesviruses* are shown.

in the inverted repeat, and, based on recent discoveries in other herpesviruses, this number likely is a large underestimation. The EHV-4 genome harbors at least 76 unique genes and 79 ORFs due to duplication of three genes of the inverted repeats. The 63 open reading frames of the UL region are arranged colinearly between EHV-1 and EHV-4 and with those in the genomes of herpes simplex virus and varicella zoster virus. Several genes mapping within IR and US region differ in arrangement from those of other alphaherpesviruses. In addition, EHV-1 and EHV-4 contain a limited number of unique genes that are present neither in HSV-1 nor in VZV which might represent the viruses' gene repertoire determining host specificity.

A comparative genome sequencing study of three EHV-1 strains (Ab4: neuropathogenic strain; RacL11: abortogenic strain; KyA: non-pathogenic strain after serial passages in mouse fibroblast 1-M cells) revealed that KyA genome is the shortest with 141,350 bp when compared to the RacL11 genome, 147,469 bp, and the Ab4 genome, 150,223 bp. Sequence analysis showed that ORF1 and ORF2 genes are deleted in both KyA and RacL11 strains, while other deletions of ORFs 17, 73, 74, and 75 were only detected in the KyA strain. Identical in-frame deletions in ORFs 14, 63, and 68 were observed in both RacL11 and KyA, while extra in-frame deletions in ORFs 61 and 71 were observed only in KyA strain. On the other hand, EHV-4 viruses appeared to be more conserved with minimal changes in their genomes, even when analyzing the genomes of EHV-4 that was suspected to be the cause of abortion in mares.

Six genes and the origin (ORI) of replication have been mapped to the IR sequences of EHV-1, while the remainder of the genes locates in the unique regions of the genomes. Unique genes (ORFs 2, 3, 34, 58, 67, 75, and IR2 and IR3) of EHV-1 and some of EHV-4 with no homologs in HSV-1 have been identified and mapped to the UL, US, and repeat regions. In addition, another interesting gene (ORF71, gp2) was mapped in the US segment. This gene encodes a highly O-glycosylated protein referred to as gp2. Some size variations of gp2 were documented for EHV-1 strains KyA and wild-type strains Ab4 or RacL11. A gp2-null mutant was apathogenic in a murine model of EHV-1 infection. Replacement of the truncated gp2 gene of the apathogenic KyA virus with the gp2 gene of the pathogenic RacL11, which encodes for the full-size gp2 of 791 amino acids, resulted in a "transfer" of pathogenic properties, such that full-length gp2 is a major determinant of virulence.

The EHV-9 genome is shorter, 148,371 bp, than the genome of EHV-1 (Ab4 neuropathogenic strain). The genome has in total 80 open reading frames that also present in EHV-1; however, no information about the presence of IR2 or IR3 is available. The

Table 2 Regulatory genes of EHV-1

Gene	Temporal class	Gene product	Function in replication
IE	Immediate early	1,487aa	Trans-activates early genes and activates some late genes
IR4	Early	293aa	Enhances IE protein DNA binding and binds TBP
UL5	Early	470aa	Binds the IE protein, TFIIB, and TBP
EICP0	Early		Promiscuous trans-activator, which Antagonizes IE protein function; binds IE protein, TFIIB, TBP
IR2	Early	1,165aa	Dominant negative regulator, which blocks IE protein binding to promoters
IR3	Early	0.9kb RNA	Antisense to IE mRNA; precursor to microRNA?
ETIF	Late	479aa	Trans-activates the IE promoter; essential for virus egress

EHV-9 genome has a high degree of similarity with EHV-1, with ORFs 42, 52, and 53 showed the highest degrees of identity (95%) and ORF71 the lowest (86%). One frameshift in ORF68 was detected when compared to EHV-1.

The genomic sequences of EHV-2, EHV-3, and EHV-5 have now been determined, but intensive research on gene functions has not yet been performed.

Life Cycle

The initial attachment of EHV-1 and EHV-4 virions to cells occurs through the interaction of gC with cellular heparan sulfate, while gD is essential for virus entry after binding to specific cellular receptors. EHV-1 gD was shown to utilize unique and wide range of entry receptors that enables the virus to enter and infect cell types of different origins. Equine major histocompatibility (MHC) class I molecules were determined as the gD-binding receptors for both EHV-1 and EHV-4. Yet, these two viruses can still infect other non-equine cells independently of equine MHC-I. The interaction between EHV-1 gH and cellular $\alpha 4\beta 1$ -integrins have been shown to induce release of Ca^{2+} from intracellular stores and facilitate fusion of the viral envelope with the plasma membrane.

Transcription and translation of EHV-1 genes have been studied best for the KyA strain. The genes of the KyA virus are regulated at the transcriptional and translational levels in a temporal fashion, and three kinetic classes of genes designated immediate-early (IE), early (E) and late (L) have been described. The sole IE gene (IR1) maps in both IR segments and gives rise to a spliced 6.0 kbp mRNA. Multiple IE polypeptide species have been observed, and the major IE protein (IE1, 203 kD) is a nuclear-localized phosphoprotein that is capable of *trans*-activating other viral genes and auto-regulating its own transcription. Previous studies revealed that the IE protein harbors domains for binding general transcription factors, such as TFIIB and TATA-binding protein (TBP), and thus serves to promote the formation of pre-initiation complexes that mediate viral transcription. Following IE polypeptide synthesis, approximately 45 early transcripts can be detected. Four of these early proteins serve as regulatory proteins and are designated IR4 (EICP22), UL5 (EICP27), EICP0 (UL63) and IR2 (Table 2). The IR4 protein physically interacts with the IE protein and serves to enhance the DNA-binding of the IE protein to its target sequence (ATCGT) present within the promoters of EHV-1 genes characterized to date. IR4p also binds to TBP and is present at early viral promoters in association with the IE protein and TBP. These interactions explain the synergistic effect on the *trans*-activation of viral genes mediated by the IE and IR4 proteins. The EHV-1 homolog of ICP27 of HSV-1 is essential for virus replication in cell culture. It acts synergistically with either the IE protein or the EICP0 early regulatory protein to activate expression of both early and late viral gene expression. The third early regulatory protein is EICP0p, which is a powerful and promiscuous *trans*-activator that can independently activate expression of viral genes of all three temporal classes. Deletion of the *EICP0* gene greatly impairs virus replication and severely retards late gene expression, suggesting that this regulatory protein is important in the switch from the early to late phase of viral gene programming. The fourth early regulatory protein is the IR2 protein that is a truncated form of the IE protein that lacks its essential *trans*-activation and serine-rich domains. The IR2 protein serves a negative regulatory role as it down-regulates viral gene expression by acting as a dominant negative protein that blocks IE protein binding to viral promoters and/or by squelching the limited supply of TFIIB and TBP. In addition to these four early auxiliary regulatory proteins, the EHV-1 unique IR3 gene contributes a regulatory role as it encodes a small transcript that is antisense to a portion of the IE transcript. In transient transfection assays, the IR3 transcript down-regulates IE gene expression and is only minimally expressed to a protein. Initial studies suggest that the IR3 transcript is processed to a microRNA, and thus the IR3 gene may use novel mechanisms to down-regulate IE gene expression at late times of infection. Early gene expression is followed by viral DNA replication and the production of approximately 29 late transcripts. Although these 75 transcripts have been positioned on the viral genome, only a small number of protein products have been identified and characterized (see above).

Viral DNA replication initiates at approximately 4 h postinfection and requires the virus-encoded DNA polymerase. DNA replication is thought to occur by the rolling circle mechanism whereby long concatemers of the viral genome are generated, cleaved and then packaged into the maturing virions. The UL15 homolog of HSV-1, one of the two spliced genes of EHV-1 known to date, appears to be essentially involved in the generation of unit-length genomes and their packaging into mature capsids.

The start of viral DNA synthesis initiates the late phase of viral gene expression and the synthesis of the late regulatory protein ETIF, a counterpart to the *alpha trans* inducing factor of HSV-1. This 60 kDa protein is multifunctional and plays at least three roles in EHV-1 replication. ETIF in several molecular sizes is present in the tegument and thus contributes to the composition of this

virion component and overall virion structure. Secondly, after virus entry and uncoating, ETIF serves to *trans*-activate the *IE* promoter by binding to cellular factors that mediate its association with the TAATGARATT sequence at nt -630 to -620 in the *IE* promoter. This *trans*-activation function to initiate viral gene programming is important but is not essential as EHV-1 DNA is infectious and virus progeny are produced in cells transfected with plasmids carrying the EHV-1 genome. The activation of a viral promoter by ETIF is specific for the *IE* promoter as ETIF is not able to *trans*-activate any early or late promoter tested to date. ETIF-deleted virus (Δ ETIF) and an ETIF complementing cell line revealed that ETIF plays a key role in secondary envelopment. As reported for other herpesviruses, EHV-1 maturation occurs by interaction of mature nucleocapsids with the inner portion of the nuclear membrane resulting in the formation of enveloped particles. Studies with Δ ETIF and Δ US3 EHV-1 support the model of the sequential envelopment/de-envelopment/re-envelopment for the egress of EHV-1. Nuclear egressed-nucleocapsids are transported to the TGN, where additional tegument proteins are recruited and the envelope with glycoproteins is acquired. The mature virions are released extracellularly through exocytosis.

Latency

Like other members of the *Alphaherpesvirinae*, equid herpesviruses establish latency in different cell types in the horse after primary infection. Under stress or following immunosuppression, periodic reactivation from latency usually happens and leads to viral shedding and horizontal transmission. EHV-1 has been shown to establish latent state in leukocytes, lymphoid tissues, and neuronal cells of the trigeminal ganglion. EHV-4 can establish latency in trigeminal ganglia and lymphoid tissues. EHV-2 is thought to establish latency in B-lymphocytes, macrophages, lymphoid tissues, peripheral and central nervous systems, and possibly Langerhans cells. Potential sites of EHV-3 and EHV-5 latency are not yet identified; however, mononuclear cells are surmised to be the major site of EHV-5 latency. The location of EHV-3 latency is presumed to be the sciatic or sacral ganglia. Interestingly, latent state of EHV-1 and EHV-9 in the sensory ganglia of zebras has been identified.

Epidemiology

In general, herpesviruses are species-specific and it has long been thought that these viruses have co-evolved with their hosts over millions of years; that is how equine herpesviruses acquired their nomenclature. Advances in sequencing technology and screening a wide range of species might now take the lead to change this general concept (i.e., herpesviruses are strictly species-specific) as there is clear evidence of frequent virus transfer beyond the definitive host.

The horse is the natural host of EHV-1, -2, -3, -4, and -5, while the definitive host of EHV-9 is still unknown. Wild equids (zebras and onagers) and domestic donkeys can also serve as a host of EHV-1. Further, EHV-1 has a broad host range and the virus can infect several non-equid species including captive gazelles, antelopes, cattle, alpacas, llamas, black bears, polar bears, black and Indian rhinoceros, and Indian tapir. EHV-9 infections have been also documented in non-equid species, such as gazelles, giraffes and polar bears. A recombinant virus of EHV-1 and EHV-9 was detected in lung and brain tissues of an Indian rhinoceros and polar bears. Serological studies indicated the presence of EHV-1 antibodies in African white and black rhinoceroses, while EHV-9 antibodies were widespread high-titered in zebras and even more so in white and black rhinoceroses. It is generally thought, however, that both EHV-1 and EHV-9 will not replicate to high levels in most non-definitive hosts and that no interindividual spread occurs.

Epidemiological studies concerned with determining the host range of other equid herpesviruses are limited. The currently available reports show that the host range of EHV-2, EHV-3, EHV-4, and EHV-5 is more restricted than that of EHV-1. Domestic donkeys have been shown to be infected with EHV-2, EHV-4, and EHV-5.

In addition to equine influenza virus, EHV-1 and 4 are considered to be the most relevant clinical equine viral respiratory pathogens. However, of the equine herpesviruses, subclinical detection of virus in respiratory secretion or blood samples is most common for EHV-5 and EHV-2 followed by EHV-4 and then EHV-1. For all of the equine herpesviruses the main epidemiologically relevant features include: (1) High incidence of respiratory infection early in life; (2) Establishment of latency in a high percentage of infected horses; (3) Frequent reactivation of latent virus with subsequent shedding, resulting in transmission to naïve hosts. Currently, it is estimated that as many as 80% of all horses are latently infected with EHV-1 and seroprevalence for EHV-4 is typically even higher. Prevalence may vary depending on geographic region, age and population. Furthermore, testing technology and tissues sampled for detection of latency affect sensitivity of testing and actual numbers of latently infected horses may be even higher.

Rhinopneumonitis caused by EHV-1 and EHV-4 is spread by direct or indirect contact (ingestion and inhalation). The viruses are shed in nasal droplets for up to three weeks and are present in large amounts in aborted fetuses and the placenta. EHV-1 infection can also result in spontaneous abortions in pregnant mares between the eighth and eleventh months of pregnancy. The peak incidence is in the ninth and tenth months at which time approximately 70% of abortions occur. Presence of EHV-1 and histological lesions following infection with EHV-1 have also been reported in testis and semen of infected stallions but a venereal transmission has not been reported. Duration of shedding following recrudescence varies, but because prolonged shedding has been detected particularly in cases of equine herpesvirus myeloencephalopathy (EHM) outbreaks, current recommendations are to wait 28 days before lifting quarantine measures.

EHV-2 and EHV-5 are endemic worldwide with high seroprevalence, and co-infections are common. The viruses have been isolated from the respiratory tract, conjunctiva, blood leukocytes, kidneys, spleen, testicles, genital tract and rectum, however

isolation is primarily reported from the respiratory tract, the conjunctiva and blood cells, which are presumed sites of primary infection and latency. Once infected, the horse is a life-long carrier and excreter of the virus. The exact mode of spread for EHV-2 and EHV-5 is unknown but horizontal transmission is the likely mode of transmission under field conditions.

EHV-3 causes a mild coital exanthema that is spread by genital contact and – rarely – the respiratory route or via contact with contaminated objects. Information on seroprevalence is similarly limited and reported rates range from 6% to 48%. EHV-3 infection is usually cleared after 14 days and is not associated with equine abortions.

Pathogenesis and Immune Response

Although EHV-1 and EHV-4 respiratory infections are clinically indistinguishable, their pathogenesis is quite different. Primary infection with EHV-1 and EHV-4 occurs via the respiratory tract by inhalation of infectious virus, nose-to-nose contact, or contact with contaminated tissues or fomites. Following infection and replication in the respiratory airway epithelium, EHV-1 can be detected by 24–48 h in the local lymph nodes of the respiratory tract and a cell-associated viremia is established between days 4 and 10 post infection. In contrast, a cell-associated viremia is not commonly observed with EHV-4, which stays restricted to the respiratory tract. For EHV-1, the cell-associated viremia transports the virus to sites of secondary infection where contact between infected leukocytes and the vascular endothelium leads to endothelial cell infection, inflammation, thrombosis and tissue necrosis and secondary disease manifestations directly following viremia. Secondary disease manifestations for EHV-1 include equine herpesvirus myeloencephalopathy (EHM), EHV-1 abortions, neonatal foal death and chorioretinopathies. Because a cell associated viremia is the pre-requisite for transport of the virus to sites of secondary infection, including EHM and late term abortions, EHV-4 is only occasionally associated with abortion and extremely rarely with neurological disorders. While a positive correlation between the duration and magnitude of viremia and incidence of EHM has been identified, and EHM is unlikely to occur in the absence of viremia, only a small percentage (~10%) of viremic horses subsequently develop EHM. A combination of host and viral factors have at this point be identified to be associated with incidence of EHM including age, breed, gender, season as well as a single nucleotide polymorphism in the viral polymerase gene that results in a coding change (D752 *vs.* N752) and is strongly associated with neuropathogenicity (D752) and high levels of viremia. However, strains containing each polymerase variant (D752 or N752) have the potential to cause EHM, and strict control measures should be taken regardless of the variant identified.

Infection with EHV-3 occurs through direct skin-to-skin contact during coitus or via secretions containing live virus (i.e., contaminated materials during artificial insemination, hands or even lips and nose of horses nuzzling or sniffing each other). No disruption of the skin barrier is necessary to establish infection and viral replication is typically limited to the stratified epithelium, leading to a localized inflammatory response associated with the classical cutaneous lesions. Spread of the virus to the underlying tissues and dissemination via the blood stream does not occur preventing systemic dissemination. Primary infection with EHV-3 often leads to secondary infection with *Streptococcus equi zooepidemicus* affecting the severity and duration of clinical disease. Most horses clear the infection on their own within 2–3 weeks of infection but a high percentage of horses establish latency, and recrudescence of the virus and clinical signs can be observed in horses in consecutive breeding seasons.

The exact pathogenesis of EHV-2 and EHV-5 is not fully understood. Infection with both viruses is likely acquired horizontally early in life usually by inhalation. The frequent isolation of EHV-2 from pharyngeal lymphoid tissue suggest that this might be a primary replication site from which the virus is then spread by peripheral blood leukocytes (PBL) to other organs. EHV-5 has recently been associated with EMPF. The pathogenesis is thought to mirror that of a murine model of lung fibrosis caused by MHV-68, marked by the progressive deposition of interstitial collagen, increased TGF- β and Th2 cytokine expression, and hyperplasia of type-2 pneumocytes.

The six equine herpesviruses are antigenically distincts although they share certain antigens. EHV-1, EHV-4, and EHV-9 are closely related antigenically, such that cross-neutralizing antibodies are generated. Also, EHV-2 and EHV-5 are closely related. Almost no data are available on the relationship of EHV-3 to other members of the equine herpesviruses, however, all of the EHV-3 are believed to share complement-fixing antigens.

Neutralizing antibodies are detected in the serum soon after an EHV-1 or EHV-4 infection. However, it has become clear that: (1) virus-neutralizing antibodies can reduce viral nasal shedding but do not correlate with protection from EHM or abortions, (2) protection from EHV-1 depends critically on induction of cytotoxic T-cell (CTL) responses and prevention of viremia, and (3) EHV-1 uses multiple mechanisms to evade the induction of protective immunity, providing a unique challenge for the development of preventatives. More recently there has been an increased interest in innate immunity to EHV-1 because early immunity is not only critically important for immediate protection but likely shapes subsequent adaptive immune responses. On the flip side, over the past years multiple immune evasion mechanisms employed by the virus have been identified. These include NK-cell lysis, alteration of cytokines regulating B- and T-cell responses, alteration of the chemokine network, interference with antigen presentation, inhibition of antibody dependent cytotoxicity, induction of T-regulatory cells and alteration of CTL responses. A number of EHV-1 candidate immunomodulatory genes have so far been identified. These include the UL49.5 and UL56/UL43 genes that interfere with MHC-I antigen presentation and induction of CTL responses, as well as glycoprotein G, which functions as a viral chemokine-binding protein that selectively inhibits IL-8 and CCL-3 mediated chemotaxis of phagocytes (neutrophils and macrophages). Other viral modulatory genes are likely present and may be identified in the future.

Immunity to EHV-2, EHV-3, EHV-5, and EHV-9 is poorly understood; however, virtually all horses have antibodies to EHV-2 and EHV-5 confirming the general apathogenicity of these viruses.

Clinical Features and Pathology

Clinical manifestations of EHV-1 and EHV-4 include primary respiratory disease, late term abortions, neonatal foal death, EHM and chorioretinopathy. EHV-4 infections are typically restricted to the respiratory tract but secondary disease has been reported occasionally. EHV-1 and EHV-4 respiratory disease can be mild or asymptomatic in older or previously exposed horses. In contrast, the respiratory disease observed in young immunologically naïve horses is often severe and lasts for 2–3 weeks. Infection is characterized by a short incubation time (<1 day), fever (39–41°C), which can last between 2 and 5 days, sometimes with a second spike approximately one week after the primary pyrexia and coinciding with the onset of viremia. Animals also suffer from serous nasal discharge and congestion of the nasal mucosa and conjunctiva. Less frequently, one can detect a transitory period of anorexia, enlargement of the submandibular lymph nodes and edematous swelling of the lower parts of the body and the extremities. An initial leukopenia is followed by leukocytosis before the temperature falls. Coughing is reported in cases with severe respiratory disease and occasionally in milder cases depending on the environment. Lower respiratory disease associated with secondary bacterial infection, tachypnea, anorexia and depression can be observed in young foals. Clinically EHV-1 and EHV-4 respiratory disease is difficult to distinguish, unless the virus is identified diagnostically.

EHV-1 is also the cause of late-term abortions and premature delivery of foals that die soon after birth. Mares infected with EHV-1 commonly do not display respiratory disease and abort 2 weeks to several months after infection or re-activation of the virus. Abortion typically occurs in the last trimester of pregnancy but has been reported as early as the second trimester. Abortions happen without warning signs and the placenta is often found together with the fetus, which has died from asphyxia or dies shortly after birth. Sporadic abortions in individual mares are most common, but EHV-1 outbreaks with attack rates >50% “abortion storms” have been reported and depend on herd management, immune status and viral factors. Mares typically recover and deliver a healthy foal in the next breeding season. Occasionally apparently healthy foals are delivered that become ill within 2 days of delivery and show respiratory distress, fever, failure to nurse, weakness, diarrhea and leukopenia and do not respond well to treatment.

Clinical signs of neurological disease caused by EHV-1 are highly variable and include head pressing, ataxia, and paralysis with complete recumbency. Both, EHV-1 and EHV-4 have been isolated from the central nervous system of infected horses, but neurological disease due to EHV-4 is extremely rare. EHV-1 does not invade the extravascular nervous tissue and is not neurotropic per se. Rather, the virus spreads from the respiratory tract to the CNS via infected leukocytes and infects endothelial cells of the blood vessels supplying the spinal cord. While the neurological form of the disease is usually lethal, some horses recover fully from neurological disease with no permanent neurological sequelae.

EHV-1 infection can also lead to chorioretinopathy causing permanent “shotgun” lesions in a substantial proportion of infected horses. A recent study has shown that >50% of yearling horses exhibit classical shotgun lesions following experimental infection with EHV-1, but that lesions are not visible *in vivo* until 1–3 months after primary infection. Lesions can be focal, multifocal or in rare occasions diffuse affecting the whole eye. Clinically, only diffuse lesions have a significant impact and cause loss of vision.

EHV-3 replication is restricted to the stratified epithelium of the skin or muco-cutaneous margins and lesions are restricted to the superficial skin of the external genitalia in both mares and stallions. An incubation period as long as 10 days can be observed following natural infection. The initial lesions are small (1–2 mm), raised, and reddened papules. The lesions then progress rapidly to the pustular form, and there is a general reddening of the vaginal mucosa in the mare. The number of lesions increases in the first few days, and by day 6, many of the lesions form ulcers up to 20 mm in diameter and 5 mm deep. Lesions can also be seen on the vulva, perineal skin, as well as the penis and prepuce of the stallion. The disease is usually mild such that the temperature, pulse, appetite and respiration remain close to normal. Occasionally, severely affected horses may be febrile and show depression and anorexia. Stallions may show a loss of libido and refusal to mount and mares may show frequent urination, arching of the back and vulvar discharge. The severity of the disease can be increased by secondary bacterial infections; however, uneventful cases are usually cleared within 2 weeks. EHV-3 is not abortigenic and does not lead to infertility.

The role of EHV-2 and EHV-5 in clinical disease is less understood. EHV-5 has been implicated as the cause of EMPF in recent years. EMPF is typified by low-grade fever, weight loss, and progressive exercise intolerance, along with radiographic evidence of nodular pulmonary interstitial fibrosis and isolation of the virus from affected lungs. EHV-2 has been associated occasionally with chronic throat infections. However, pharyngitis as well as conjunctivitis, coughing and nasal discharge has so far only been experimentally induced in young immuno-suppressed animals.

Infections with EHV-9 lead to encephalitis in horses under experimental conditions, but, more severely, in other species as well.

Respiratory disease caused by EHV-1 and EHV-4 results in inflammation, congestion and sometimes necrosis of the upper respiratory tract. Extensive swelling of the nasal mucosa may occur, and in later stages, the lungs may become involved. One can find typical herpesvirus inclusion bodies in the nuclei of the respiratory epithelium. The respiratory infection can become more serious if followed by a secondary bacterial infection, which may lead to bacterial pneumonia.

Fetuses that are aborted as a consequence of EHV-1 infection present with widespread hemorrhages and edema, as well as a yellowish discoloration of the fetal conjunctiva and splenomegaly – if virus is transmitted from mother to animal. Abortions without infection of the fetus also occur and the pathogenesis in the uterine vasculature largely resembles that of the vasculitis observed in the neurological form of the disease. The classic histopathology includes the presence of eosinophilic intranuclear inclusion bodies in various organs of the aborted fetuses. However, gross pathological and histopathological alterations are sometimes not as obvious and only sensitive methods (virus isolation, PCR, qPCR) are able to confirm the EHV-1 abortion. Classically, features associated with EHV-1 induced abortions differ in those fetuses aborted during the first 6 months of gestation

as compared with those aborted after 6 months. Those before 6 months present with widespread cell necrosis and inclusion bodies in the liver and lung. Those after 6 months exhibit jaundice, subcutaneous edema, excessive pleural fluid, pulmonary edema, splenomegaly and necrosis of the liver.

Similar to histopathological findings in the CNS and the pregnant uterus, vasculitis is common in the choroid of the eye, as are mild or moderate perivascular and interstitial infiltrates of lymphocytes and histiocytes around the chorioretinal vasculature.

Finally, vasculitis with perivascular to scattered interstitial lymphohistiocytic infiltrates can also be observed in the testis of intact males following experimental infection. In addition, occasionally, discrete lymphofollicular structures within the interstitium of the testes, tubular degeneration and arteritis in the fascia surrounding the testes or fibrous interstitium may be observed.

Tissues affected following an infection with EHV-3 include the vaginal and vestibular mucosa, penis, prepuce and the skin of the perineal region. One of the characteristics of an infection with EHV-3 is the sloughing of the surface epithelial cells. On occasion, the skin of the lips and mucus membranes of the respiratory tract may become involved, but the exanthema is usually mild.

EHV-2 and also EHV-5 infection becomes widespread throughout the body, and the viruses have been isolated from a variety of tissues. Horses affected by EMPF show coalescing or discrete nodules of fibrosis and an enlargement of bronchial lymph nodes. Histologically the lesions are centered on the alveolar parenchyma and are characterized by an expansion of collagen accompanied by infiltration of inflammatory cells into the interstitium.

Control

Standard diagnostic methods for equid herpesviruses are well established, including virus isolation, serological assays, particularly serum neutralization tests (SNT) and type-specific ELISAs, as well as molecular assays that are now almost exclusively based on quantitative PCR protocols.

A number of vaccines are available to combat EHV-1 infections, among them modified live vaccines (e.g., Rhinomune™ or Prevaccinol™), inactivated vaccines (e.g., Pneumabort K™, Equip™ EHV1,4, Prestige™, Prodigy™), inactivated combination vaccines, which – among others – also contain EHV-4 (e.g., Innovator™, Vetera™). Unfortunately, many EHV vaccines do not afford acceptable levels of protection, especially when inactivated combination vaccines are considered. All vaccines are given repeatedly to pregnant mares usually in the third, fifth, seventh and ninth months of pregnancy. To protect against viral rhinopneumonitis outbreaks, the vaccine is usually given to all horses every 3–6 months. There is considerable ongoing discussion as to proper vaccination against the neurological disease, which has become more prevalent during past years. Recent studies seem to suggest that modified live virus vaccines may be superior to especially inactivated combination vaccines with respect to duration of fever and virus excretion from the nasal mucosa; at this point there is however little evidence that vaccination can prevent EHM, and conflicting information is available on the use of vaccination for prevention of abortion. Clinical management often involves the use of antibiotics to prevent severe bacterial complications following the viral rhinopneumonitis. Because of the vaccine limitations, control of EHV-1 and EHV-4 infections involves isolation and quarantine of infected horses (up to 28 days) and sound hygiene for prevention of viral infection, since the viruses are highly contagious. Quarantine procedures are required with EHV-1 infections, since EHV-1 can lead to more serious disease of the CNS and the viruses can be spread easily in contaminated feed and water. In addition, minimizing stress and close contact of large groups of horses can prevent the spread of disease. It is important to recognize that EHV-1 infection is a reportable disease in some states of the U.S., and in these states, quarantine requirements will be determined by state officials.

No vaccines are available for preventing EHV-3 infection at present and unless secondary bacterial infection develops, genital lesions caused by EHV-3 infection usually heal without therapeutic intervention.

Future Perspectives

Considerable progress in unraveling the genetic makeup of EHV-1, EHV-2, EHV-3, EHV-4, EHV-5, and EHV-9 has been made during the last years, and the genomes have been sequenced in their entirety. With this information in hand, it will be possible to perform studies on gene expression and on proteins that are involved in virulence of each of the viruses. These studies will in turn open the possibility for a rational design of anti-EHV vaccines, especially against the most important pathogens, EHV-1 and EHV-4. These goals may be achieved by the use of viral deletion mutants that carry targeted gene deletions, which is greatly facilitated by the advent of a number of infectious DNA clones (bacterial artificial chromosome – BAC – clones) during the past years. Using novel molecular approaches, a better understanding of the viruses' biology will likely be possible in the near future and allow a finer definition of diseases caused by the different EHV.

Further Reading

- Abdelgawad, A., Damiani, A., Ho, S.Y., *et al.*, 2016. Zebra alphaherpesviruses (EHV-1 and EHV-9): Genetic diversity, latency and co-infections. *Viruses* 8 (9), doi:10.3390/v8090262.
- Azab, W., Kato, K., Arii, J., *et al.*, 2009. Cloning of the genome of equine herpesvirus 4 strain TH20p as an infectious bacterial artificial chromosome. *Archives of Virology* 154 (5), 833–842.

- Azab, W., Osterrieder, K., 2017. Initial contact: The first steps in herpesvirus entry. *Advances in Anatomy, Embryology and Cell Biology* 223, 1–27. doi:10.1007/978-3-319-53168-7_1.
- Barrandeguy, M., Thiry, E., 2012. Equine coital exanthema and its potential economic implications for the equine industry. *The Veterinary Journal* 191 (1), 35–40. doi:10.1016/j.tvjl.2011.01.016.
- Fortier, G., van Erck, E., Pronost, S., Lekeux, P., Thiry, E., 2010. Equine gammaherpesviruses Pathogenesis, epidemiology and diagnosis. *The Veterinary Journal* 186 (2), 148–156. doi:10.1016/j.tvjl.2009.08.017.
- Hussey, G.S., Goehring, L.S., Lunn, D.P., *et al.*, 2013. Experimental infection with equine herpesvirus type 1 (EHV-1) induces chorioretinal lesions. *Veterinary Research* 44, 118doi:10.1186/1297-9716-44-118.
- Ma, G., Azab, W., Osterrieder, N., 2013. Equine herpesviruses type 1 (EHV-1) and 4 (EHV-4) – masters of co-evolution and a constant threat to equids and beyond. *Veterinary Microbiology* 167 (1–2), 123–134. doi:10.1016/j.vetmic.2013.06.018.
- Paillot, R., Ellis, S.A., Daly, J.M., *et al.*, 2006. Characterisation of CTL and IFN-gamma synthesis in ponies following vaccination with a NYVAC-based construct coding for EHV-1 immediate early gene, followed by challenge infection. *Vaccine* 24 (10), 1490–1500. doi:10.1016/j.vaccine.2005.10.019.
- Rudolph, J., O'Callaghan, D.J., Osterrieder, N., 2002. Cloning of the genomes of equine herpesvirus type 1 (EHV-1) strains KyA and racL11 as bacterial artificial chromosomes (BAC). *Journal of Veterinary Medicine B, Infectious Diseases and Veterinary Public Health* 49 (1), 31–36.
- Shakya, A.K., O'Callaghan, D.J., Kim, S.K., 2017. Comparative genomic sequencing and pathogenic properties of equine herpesvirus 1 KyA and RacL11. *Frontiers in Veterinary Science* 4, 211doi:10.3389/fvets.2017.00211.
- Soboll Hussey, G., Ashton, L.V., Quintana, A.M., *et al.*, 2014. Equine herpesvirus type 1 pUL56 modulates innate responses of airway epithelial cells. *Virology* 464–465, 76–86. doi:10.1016/j.virol.2014.05.023.
- Soboll, G., Whalley, J.M., Koen, M.T., *et al.*, 2003. Identification of equine herpesvirus-1 antigens recognized by cytotoxic T lymphocytes. *Journal of General Virology* 84, 2625–2634. doi:10.1099/vir.0.19268-0.
- Van de Walle, G.R., Sakamoto, K., Osterrieder, N., 2008. CCL3 and viral chemokine-binding protein gg modulate pulmonary inflammation and virus replication during equine herpesvirus 1 infection. *Journal of Virology* 82 (4), 1714–1722.
- Wilkie, G.S., Kerr, K., Stewart, J.P., Studdert, M.J., Davison, A.J., 2015. Genome sequences of equid herpesviruses 2 and 5. *Genome Announcements* 3 (2), doi:10.1128/genomeA.00119-15.
- Williams, K.J., Maes, R., Del Piero, F., *et al.*, 2007. Equine multinodular pulmonary fibrosis: A newly recognized herpesvirus-associated fibrotic lung disease. *Veterinary Pathology* 44 (6), 849–862. doi:10.1354/vp.44-6-849.

Equine, Canine, and Swine Influenza (*Orthomyxoviridae*)

Janet M Daly, University of Nottingham, Sutton Bonington, United Kingdom

Japhette E Kembou-Ringert, University of Tel Aviv, Tel Aviv, Israel

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

(H)/HA Hemagglutinin (subtype)
(N)/NA Neuraminidase (subtype)
HA0 Hemagglutinin polypeptide
HA1, HA2 Hemagglutinin subunits
HI Hemagglutination inhibition test
IAV Influenza A virus

M1/M2 Matrix protein 1/2
mRNA Messenger ribonucleic acid
NP Nucleoprotein
PA Polymerase acidic protein
PB1 Polymerase basic protein 1
PB2 Polymerase basic protein 2
vRNA Viral ribonucleic acid

Glossary

Antigenic drift The gradual accumulation of amino acid changes in the surface glycoproteins of influenza viruses.
Antigenic shift The sudden appearance of a novel combination of surface glycoproteins in a host as a result of reassortment.
Pandemic Outbreak of disease affecting a large region.

Reassortment The exchange of genome segments between distinct influenza A viruses (typically viruses of different subtypes).

Subtype Classification of influenza A viruses based on the antigenicity of the hemagglutinin and neuraminidase proteins.

Classification (Compact)

Most equine, canine and swine influenza viral infections are caused by influenza A viruses (IAV) in the *Influenzavirus A* genus of the family *Orthomyxoviridae*. Classification of viruses in this family is based on, but not limited to, the negative-sense and segmented nature of their RNA genome. Influenza A viruses (IAV) are further classified based on the antigenic reactivity of their surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). Until recently, 16 HA (H1 to H16) and 9 NA (NA1 to NA9) subtypes of IAV viruses had been identified in different host species, with aquatic birds and primarily ducks being regarded as the reservoir hosts. Even though up to 144 possible subtype combinations of H1 to H16 and N1 to N9 are possible, only a few combinations have consistently been isolated from pigs, horses and dogs. Two additional subtypes of HA (H17 and H18) and NA (N10 and N11) have recently been identified in bats.

Consistent isolation of H1N1, H1N3 and H3N2 IAV from pigs suggests that these subtypes are associated with endemic infections. In addition, H2, H4, H5, H7, and H9 subtypes have also been reported to cause sporadic infections. There are only two subtypes of IAV known to be endemic in horses, H3N8 and H7N7, although the latter is now believed to be extinct. An equine-origin H3N8 virus and avian-origin H3N2 virus are currently circulating in dogs.

Although most cases of influenza virus infection reported in pigs, horses and dogs involve IAV, it appears that pigs at least can also be infected by other types of influenza viruses. Influenza C viruses (ICV), which are primarily human pathogens, have occasionally been isolated from pigs and the more recently discovered influenza D virus (IDV) was first isolated from pigs, although cattle are thought to be the major host.

As a general rule, the nomenclature of influenza viruses includes: the antigenic type of the virus (A, B, C or D); the host of origin for non-human isolates (e.g., swine, equine, canine); the geographical location of origin of the isolate; the viral strain or isolate number and the year of isolation; and the HA and NA subtype in brackets. For example, an IAV isolate obtained from pigs in Hong Kong in 2010 with the laboratory reference number 35 will be named: A/swine/Hong Kong/35/10 (H5N1).

Virion Structure

Mammalian IAV share the same morphological features. Virions are enveloped (with the viral envelope deriving from the host cell) and often pleomorphic. The diameter of spherical particles is around 80 and 120 nm, whereas filamentous forms of the virus often found in clinical isolates can reach up to 20 µm in length. The HA and the NA glycoproteins are inserted in the viral envelope in a ratio of approximately four to one (Fig. 1). The HA is a trimeric glycoprotein that binds to the cellular receptor during the entry stage of the virus. The NA protein, on the other hand, is a globular tetrameric glycoprotein. Additionally, another smaller integral matrix protein 2 (or M2 ion channel) is embedded in the envelope in a ratio of one M2 channel per 10–100 HA molecules. The lipid membrane is lined by another viral matrix protein (M1), which gives strength and rigidity to the viral envelope, which surrounds the viral genome.

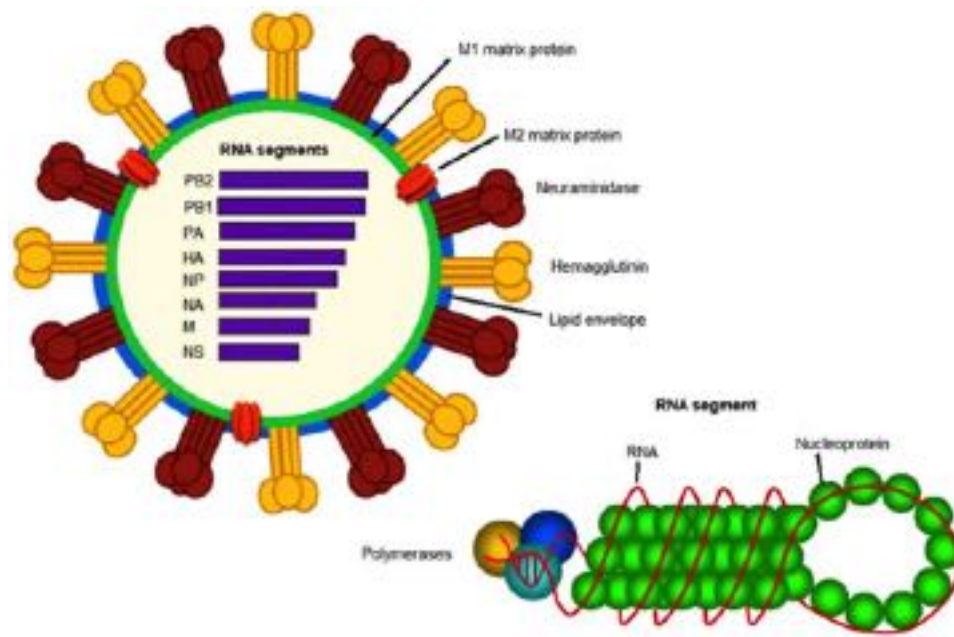


Fig. 1 Influenza A virus structure of a virus particle and an RNA segment.

Genome

The IAV genome consists of eight negative-sense, single-stranded viral RNA (vRNA) segments encoding for around 12–14 viral proteins depending on the viral strain. In addition to coding regions, each viral segment also contains non-coding regions of various length at both the 5' and 3' ends. The distal 10–12 nucleotides of these non-coding regions appear to be highly conserved among the genome segments. These conserved regions allow the formation of “panhandle” structures (helical hairpins) by partial inverted complementarity between the 5' and 3' ends; this unique feature serves as a promotor for viral mRNA transcription and plays a critical role during genome replication and packaging. Additionally, the 5' non-coding regions include the messenger RNA (mRNA) polyadenylation signal, a uridine-rich region that serves as a template on which the viral polymerase stutters to generate the poly-A tail at the 3' end of each viral mRNA.

Viral ribonucleoproteins (vRNPs) consist of viral RNA segments covered with numerous copies of the nucleoprotein (NP) and a heterotrimeric RNA-dependent RNA polymerase complex composed of two “polymerase basic” (PB1 and PB2) proteins and one “polymerase acidic” (PA) protein. Within the virions, the vRNPs interact with each other and the M1 to form a “7 + 1” configuration with seven segments of different lengths surrounding a central segment, which is thought to play an important mechanistic role during genome packaging and virus budding.

Life Cycle

The virus life cycle can be broken down into six stages: (1) attachment and internalization; (2) fusion and uncoating; (3) mRNA synthesis and viral RNA replication; (4) translation; (5) assembly and packaging and (6) budding (**Fig. 2**).

Binding of IAV to the target host cell mainly occurs via the HA protein, which recognizes and binds to sialic acid (N-acetylneuraminic acid, Neu5Ac) residues that are attached to an underlying galactose sugar chain by α 2,3 or α -2,6 linkages on the surface of the host cell. After IAV particles are brought into the vicinity of the host cell surface, the HA and NA act co-operatively; HA binds cell surface sialic acid residues while NA locally degrades and removes unnecessary HA-sialic acid interactions. This creates iterative associations and dissociations of the HA with cell surface sialic acid residues, allowing the virus to ‘crawl’ across the cell surface. This probably ensures that the virus is directed to the right entry location on the cell surface, where HA-receptor mediated endocytosis will be triggered. Although it is still not clear whether binding of HA to sialic acid residues alone induces IAV internalization, induction of host cell signaling is required to direct IAV to its specific entry route and this certainly requires specific transmembrane receptors that mediate HA-receptor endocytosis. Additional host factors other than sialic acid residues have been identified as receptor determinants including C-type lectins and annexin V (A5). Virus particles are internalized via receptor-mediated endocytosis. Most virions are internalized via a clathrin-mediated endocytic process during which Epsin-1 is recruited to the binding sites of influenza viruses in synchrony with the assembly of clathrin-coated vesicles. Dynamin protein is then required for membrane fission, to pinch off the vesicles from the plasma membrane. Various studies have established that the virus can also infect cells via a clathrin- and caveolin-independent endocytic pathway showing all the characteristics of macropinocytosis. Recent

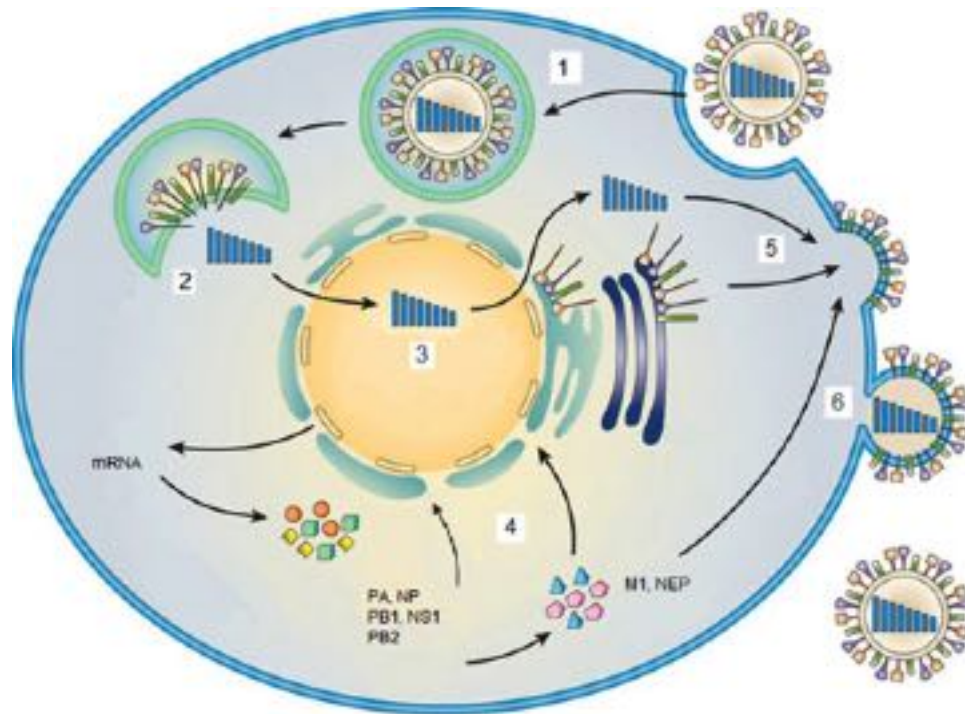


Fig. 2 Influenza A virus life cycle. 1 = Attachment and internalization; 2 = fusion and uncoating; 3 = mRNA synthesis and viral RNA replication; 4 = translation; 5 = assembly and packaging; and 6 = budding. Adapted from Neumann, G., Noda, T., Kawaoka, Y., 2009. *Nature* 459, 931–939.

approaches based on single-virus tracking and imaging established that both clathrin-mediated and clathrin-independent pathways lead to efficient viral fusion.

Following internalization of the virus, virions are initially transported from the cell periphery to early endosomes and then transferred to late endosomes where a late acidification step occurs. This low endosomal pH (≈ 5) triggers a conformational change in the HA protein of the virus, which exposes the 23 residues near the N-terminus of HA2, also known as the fusion peptide. The HA is synthesized as a precursor protein (HA0) and must be cleaved into HA1 and HA2 by host proteases in order to be infectious. At neutral pH, the fusion peptide is hidden in a hydrophobic pocket in the HA2 trimeric interface. The exposed fusion peptide inserts into the endosomal membrane on the N-terminus of HA2 while its C-terminus remains anchored to the viral membrane. HA2 trimers then bend back on themselves in a hairpin-like conformation, bringing both membranes close to each other to merge the two outermost leaflets of both membranes (often referred to as hemifusion). Subsequently, the lipid stalk ruptures to form a fusion pore, which in turn activates the M2 ion channel. Hydrogen ions from the endosome are then pumped into the viral core via the M2 ion channel, creating an internal acidification of the virions, which disrupts protein–protein interactions that hold the viral core together and causes the release of the vRNPs from the M1 into the cytoplasm via the fusion pore.

Replication of IAV takes place in the cell nucleus. After vRNPs are released into the cytoplasm, nuclear localization sequences (NLS) exposed on NP molecules recruit the cellular adapter protein importin- α . Importin- α then binds to the importin- β transport receptor and the entire complex directs the vRNPs to the nuclear pore complex, via which vRNPs are released into the nucleus. Replication and transcription are carried out by the viral RNA-dependent RNA polymerase complex made up of the PB1, PB2 and PA proteins. Because this complex is an integral part of each vRNP imported into the nucleus, each vRNP becomes an independent functional unit of replication and transcription. Given the negative sense nature of the genome, vRNA replication begins with an initial step of transcription, where vRNPs are transcribed into a complementary RNA (cRNA), which is then used as template for the synthesis of new vRNAs. As opposed to cRNA synthesis, influenza viral mRNA transcription is a unique primer-dependent transcription process during which the viral polymerase complex, recruited to the promoter-associated Pol II, binds the 5' cap of cellular mRNAs via its PB2 sub-unit. Subsequently, the PA sub-unit, containing the endonuclease activity, cleaves cellular mRNAs approximately 10 to 13 nucleotides from their 5' end. The PB2 cap-binding domain then rotates to position the stolen cellular capped primer sequences into the catalytic site of PB1, where it is then used as a primer for elongation of the mRNA, using vRNA as a template. Viral transcripts are then polyadenylated by a stuttering process of the polymerase on a short poly-U sequence located at the 5' end of the viral RNA. The coding capacity of IAV is enhanced by the presence of donor-acceptor splice sites. As a result, during mRNA transcription, these viral transcripts are capable of recruiting the cellular spliceosome machinery to produce additional spliced transcripts such as the M2 and the NEP. A third transcript of the NS gene segment (NS3) in which the donor/acceptor splice site exhibits a novel GC/GUA motif was recently identified.

The newly synthesized vRNPs are selectively exported out of the nucleus. It is thought that the viral nuclear export protein (NEP), via its leucine-rich nuclear export signals, plays a central role in targeting vRNP complexes to the chromosomal maintenance 1 (CRM1) nuclear export pathway, but nuclear export signals also exist in M1 protein and NP. Nuclear export of capped IAV mRNAs is believed to occur in an analogous manner to the nuclear export of cellular mRNAs. It is believed that following capping, the 5' caps of nascent viral transcripts bind to the cap-binding complex. The viral NS1 protein recruits proteins involved in nuclear export of cellular mRNAs but also forms inhibitory complexes with some essential export factors (such as Nxt1) to block the export of endogenous cellular mRNAs, as a way to ensure the preferential nuclear export of viral transcripts.

Viral enveloped proteins (HA, NA and M2 for IAV) are synthesized on membrane-bound ribosomes into the endoplasmic reticulum where they are folded and trafficked to the Golgi apparatus for post-translational modifications, whereas PB1, PB2, PA, NP, NS1, NS2 and M1 synthesis is carried out by cytosolic ribosomes. NP, PB1, PB2 and PA proteins are synthesized during early infection. The NLS sequences on these proteins target them back into the nucleus by the recruitment of importin- α and - β to assist in viral mRNA transcription and viral replication. In addition to generating new transcripts by alternative splicing, IAV also generates so-called accessory proteins by either alternative translation initiation sites or ribosomal frameshift. Segment 2, for instance, additionally encodes for PB1-F2 and PB1-N40 (depending on the viral strain) by using alternative translation initiation sites. Likewise, segment 3 can also encode for PA-X (a fusion protein that results from a +1 ribosomal frameshift), while PA-N155 and PA-N182 are N-terminally truncated forms of PA.

Viral RNPs are trafficked toward the apical plasma membrane in polarized epithelial cells. The M1 protein has also been identified as a key component in the assembly process of virions via protein-protein interactions. M1 is thought to interact with itself, the vRNPs, and the HA and NA surface glycoproteins to both facilitate the concentration of viral components at the budding site and allow the discrimination of viral proteins from host proteins. By interacting with the cytoplasmic tail and the trans-membrane domain of envelope glycoproteins HA and NA, M1 functions as a bridge between the viral envelope and vRNPs.

It is believed that budding of IAV occurs at distinct apical plasma membrane regions known as rafts. It has been suggested that IAV promotes budding by inducing membrane curvature. The expression and abundance of HA and NA on one side of the membrane are sufficient to induce budding, and the presence of M1 greatly increases budding efficiency. It is thought that M1 oligomerizes upon reaching the cellular membrane where it is modeled to form curved structures. This property could explain its role in the formation of either spherical or filamentous virion forms. To facilitate the release of virions from the plasma membrane, NA catalyzes the hydrolysis of glycosidic linkages that attach sialic acids to underlying sugar molecules. The sialidase catalytic site is located in a highly conserved deep pocket in the center of each monomer of the NA tetramer.

The segmented nature of the genome means that IAV viruses can undergo reassortment, i.e., when more than one virus infects a single host cell a virus with a different combination of genome segments can arise. Antigenic shift occurs if the reassortant virus generated has a novel combination of HA and NA proteins. The lack of proofreading enzymes associated with RNA replication results in a high frequency of mutation. Together with high rates of replication, this enables IAV to evolve rapidly. The process of antigenic drift results from the accumulation of amino acid changes in the HA and NA proteins that enable the virus to escape neutralization by pre-existing antibody.

Epidemiology

Three major subtypes of IAV circulate in the pig population worldwide. The H1N1, H3N2 and H1N2 subtypes are often associated with acute respiratory disease in pigs and, in some cases, with more chronic, multifactorial respiratory disease problems in combination with other viruses or bacteria. Although the same subtypes circulate in Europe and the USA, their genetic evolution is significantly different between continents. In Europe, H1N1 subtypes were introduced into the pig population from wild ducks. In contrast, two lineages of H1N1 circulate in the USA; the 'classical' H1N1 (cH1N1) lineage was first identified in 1930 is also found in Asia and the 'reassortant' H1N1 (rH1N1) viruses with HA and NA of cH1N1 and the internal proteins of a recently emerged H3N2 subtype virus.

In European swine, circulating H3N2 viruses are reassortants of the classical H1N1 (avian-like) virus with the human H3N2 (avian-like HA and NA with internal genes from human H3N2 virus). In a separate reassortment event, the swine (avian-like) H1N1 viruses also gained the human N2 gene, generating a novel H1N2 reassortant, now endemic to Europe. Although different lineages of human H3N2 have also been detected in pig populations in Asia, these lineages seem to undergo less antigenic drift compared to their ancestral human lineages. In contrast to Europe, the H3N2 virus circulating in North American swine is a triple reassortant containing: HA, NA and PB1 of human origin; NP, M and NS genes originating from the classical swine H1N1; and PA and PB2 genes of north American avian virus origin. These triple reassortants are believed to have further acquired HA genes of diverse origins, leading to the appearance of the triple reassortant H1N2 virus (acquiring its HA from the classical swine H1N1). Other reassortant swine viruses have been reported in Europe (H1N7), Asia (H3N1) and the USA (H2N3) and some that only transiently affect pigs in other parts of the world (H9N2, H3N6 and H5N2). Novel reassortants of swine viruses with equine IAV have also been reported, highlighting the role swine play in the generation of new influenza viruses. The fact that most swine IAV are reassortants, usually of avian and human IAV, led to the hypothesis that swine serve as 'mixing vessels' for IAV through which novel viruses can be transmitted to man. The 2009 influenza pandemic caused by an H1N1 virus that originated in pigs and was a reassortant with genome segments from several viruses circulating in swine. Since the 2009 pandemic, further reassortant viruses of the 2009 H1N1 pandemic virus and other IAV strains circulating in pigs have been detected in many different countries.

Two subtypes of IAV have circulated endemically in equids: H7N7 and H3N8. The prototype equine H7N7 virus was isolated from horses in Prague in 1956 when it was termed 'equine subtype 1'. There is little evidence that H7N7 viruses are still circulating in the equine population. Virus of the H3N8 subtype was first isolated from horses in 1963. The virus is thought to have emerged in South America, spreading northwards and subsequently almost worldwide. The equine H3N8 viruses were initially a single lineage, which later evolved into two lineages (American and Eurasian). The Eurasian lineage, composed of viruses isolated from horses in Europe and Asia, continues to form a single lineage, whereas the American lineage has further diverged into three sublineages: 'South American', 'Kentucky', and 'Florida'. Viruses in the Florida sublineage further evolved into two clades. Florida clade 1 (FC1) viruses predominate on the American continent but have also caused large outbreaks in Africa, Australia, Europe, Asia and South America. Florida clade 2 (FC2) viruses, on the other hand, predominate in Europe, but viruses from this clade have also been isolated in Asia and North America. Overall, viruses in the Florida sublineage have dominated since its emergence. Equine IAV of the H3N8 subtype is endemic in many American, European and Asian countries and occasionally causes epidemics. The most recent noteworthy epidemic of equine IAV was in Australia in 2007/2008. Australia had previously been free from equine IAV and succeeded, at great expense, in eradicating it within 6 months of its emergence. Sporadic infection of equids with other IAV subtypes occurs rarely. A highly pathogenic avian influenza of the H5N1 subtype was isolated from donkeys living in close contact with diseased birds. It was thought that transmission of IAV from horses to other species was unlikely until transmission of the equine H3N8 virus to dogs occurred. Here is some evidence that the equine H3N8 subtype may have occasionally caused human infection in people but experimental infection of human volunteers with equine IAV was largely unsuccessful.

Historically, there were occasional reports of sporadic transmission and sub-clinical infection of dogs with IAV of human origin with no evidence of onward transmission. More recently, two subtypes of IAV (H3N8 and H3N2) have become established in dogs. The canine H3N8 virus is believed to have originated from horses. The virus was first isolated in 2004, but retrospective serological investigation suggested it had been circulating among greyhounds for some years prior to its identification. The canine H3N8 virus spread to other dog populations throughout the USA, mainly affecting dogs housed in large groups, e.g., dogs in animal shelters, boarding kennels and day-care centers. However, by 2018, the canine H3N8 virus seemed to have mostly died out, being limited to one large rescue center. A closely related virus was identified as the cause of an outbreak of respiratory disease among a group of foxhounds in the UK in 2002 and some dogs became infected during the large outbreak of equine H3N8 IAV in Australia in 2007/2008. However, the virus did not become established in dogs in Australia or the UK.

The second canine IAV subtype originated from birds. The virus is reported to contain gene segments from several different avian subtypes. Infections with this subtype were first reported in South Korea, where dogs are farmed for meat, and spread to other Asian countries including China and Thailand. The virus was introduced into the USA in 2015, presumed to be due to importation of dogs rescued from the Far East. It was first identified in an animal shelter in Chicago and spread to other states within the USA far more rapidly than the H3N8 subtype had done.

Dogs are sporadically infected with other types of IAV but these are not maintained in dog populations. An Asian lineage of a highly pathogenic avian influenza (HPAI) H5N1 virus was detected in a dog that had scavenged carcasses of chickens infected with the virus in Thailand. Antibodies were detected in other dogs, but no transmission of disease to in-contact dogs was observed. Other rare reports of avian subtypes of IAV affecting dogs include an HPAI H5N2 virus and H9N2 viruses. In a worrying development, an H3N1 virus was recently isolated from a dog with respiratory disease in Korea. This virus appeared to be a reassortant between an endemic canine H3N2 virus and the human 'pandemic' H1N1 subtype.

Pathogenesis

Influenza A virus infection in pigs, horses and dogs is a disease of the upper respiratory tract. There is usually a short incubation period (2–3 days). The virus replicates in epithelial cells, leading to disruption of the mucociliary elevator, which is one of the factors that predisposes to secondary bacterial infection. Both host and virus factors (particularly the HA, polymerase proteins and NS1) contribute to IAV pathogenesis. As described above, proteolysis is required to cleave the inactive HA0 protein into the HA1 and HA2 subunits. In the endemic porcine, equine and canine viruses, the HA0 is cleaved by trypsin-like proteases found in the gut and respiratory tract that recognize a single arginine in the cleavage site (a monobasic cleavage site). The acquisition of a multi-basic cleavage site in avian H5 and H7 subtype IAV during infection of chickens means that these viruses can be cleaved by ubiquitously expressed subtilisin-like proteases. This is one reason why HPAI viruses can cause severe and often fatal systemic infection in various species. Mutations in the polymerase complex proteins can contribute to IAV pathogenicity by enhancing transcription and replication of the viral genome. The NS1 protein is a highly multifunctional protein that is important for efficient virus replication and also modulates the host's innate immune response.

The type of sialic acid linkage in the host cell receptor of IAV has been shown to affect the HA receptor binding affinity in a host-specific manner leading to the suggestion that species-specific differences in the distribution of linkages on the respiratory epithelial cells influences the ability of IAV to transmit between species. The detection of both α 2,3 and α -2,6 linkages in the pig respiratory tract was thought to be central to the role of the pig as a bridging host by enabling the co-infection of pigs with avian and human IAV leading to the generation of reassortant viruses. However, several studies have demonstrated that HA receptor-binding preference for a specific sialic acid linkage is not essential for infection but is more critical for virus transmission. Mutations in the HA can alter the receptor binding affinity and therefore adaptation of IAV to a new host. Indeed, a tryptophan (W) to leucine (L) mutation of residue 222 of the HA was associated with adaptation of both the equine H3N8 and avian H3N2

viruses to dogs. However, interspecies transmission is not solely due to adaptive mutations in the HA and NA proteins. The virus must also overcome intrinsic and innate immune restrictions to replicate in a novel host. For example, interspecies transmission of avian IAV to mammalian hosts requires adaptation of the viral polymerase complex to a different importin- α isoform. Several recent studies have shown that differences in the pH stability of the HA protein reflect adaptation to different host species. A more pH-stable HA protein allows the virus to survive longer in mildly acidic environments and during infection, membrane fusion occurs in late rather than early endosomes. Avian IAV typically have a pH of activation of >5.5 whereas viruses that replicate in the upper respiratory tract of mammals generally have a pH of activation of ≤ 5.5 . Swine IAV have a broad range of pH (5.3–6.0), suggesting that pH of activation of the HA is another factor contributing to the role of pigs as a bridging host for adaptation of avian viruses to human hosts. Furthermore, porcine respiratory epithelial cells have a lower interferon response to IAV infection than equivalent human cells, which may also contribute to the lower barrier to infection of pigs with IAV from other species.

Clinical Features

The clinical signs of IAV infection are similar among pigs, dogs and horses. Typical signs include high fever, cough, nasal discharge, anorexia and apathy. Morbidity rates are usually high, but animals typically recover after 2–6 days. Although the mortality rate is typically low (less than 1%), the disease can be exacerbated by the development of secondary bacterial pneumonia.

Diagnosis

Although a rapidly spreading febrile respiratory disease characterized by sudden onset of repeated coughing is very suggestive of IAV infection in pigs, horses and dogs, the clinical signs are not pathognomonic and are often suppressed in animals that become infected despite being vaccinated. Laboratory confirmation is therefore necessary. Detection of virus in clinical specimens (nasopharyngeal swab or nasal or tracheal wash samples) is usually achieved by RT-PCR, typically using primers for the more conserved matrix or nucleoprotein genes. This approach is more rapid and sensitive than virus isolation, but isolation of viruses remains important to fully characterize circulating viruses. Viral shedding is typically limited to day 2–4 post infection, declining rapidly during the clinical phase (Fig. 3). Therefore, to optimize the likelihood of obtaining a positive result, samples for virus detection or isolation should be obtained within 1–2 days of onset of clinical signs and/or from other animals in contact with the index case. Serological assays are important to confirm infection both as a diagnostic aid for ongoing cases and for surveillance purposes. The World Organization for Animal Health (OIE) recommends the hemagglutination inhibition (HI) test. The assay is based on binding of antibodies to a standardized dose of virus inhibiting its ability to agglutinate red blood cells. A variety of factors can affect the sensitivity and specificity of the test. For canine, equine and porcine influenza, receptor-destroying enzyme and periodate and heat treatment are used to inactivate non-specific inhibitors in serum and the test typically uses 0.5 or 1% chicken or turkey red blood cells. As there may be pre-existing antibodies (from previous infection or vaccination), paired serum samples should be taken, with a 4-fold increase of titer between an acute phase sample and a convalescent sample taken 10–21 days later usually taken to indicate recent infection (Fig. 3). Competitive ELISA for detection of antibodies to nucleoprotein are also increasingly used to diagnose influenza infection.

Treatment

Although three classes of influenza antivirals (inhibitors of the M2 matrix protein, neuraminidase and PA polymerase) are licensed for human use, treatment for veterinary species is largely supportive. Treatments include fluids to help maintain hydration or

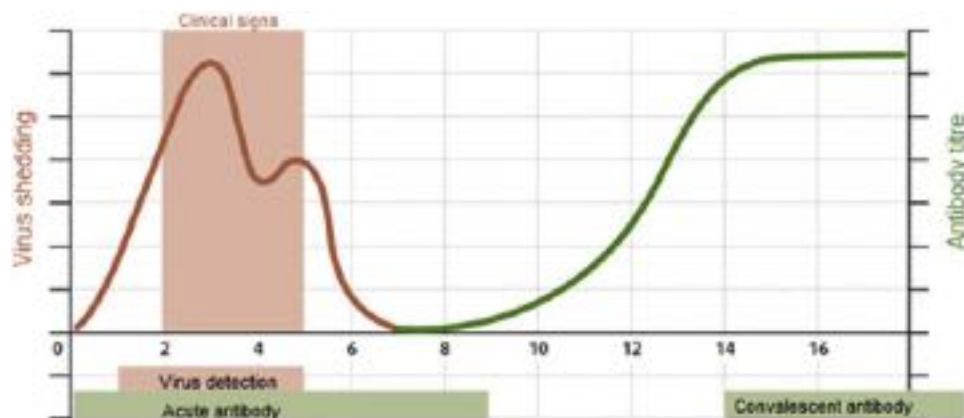


Fig. 3 Optimal sampling times for diagnosis of influenza A virus infection.

correct dehydration, non-steroidal anti-inflammatory drugs to reduce fever and inflammation and antimicrobials for known or suspected secondary bacterial infections.

Prevention

Inactivated virus vaccines are commercially available for use in pigs. In the US, it is recommended that autogenous vaccines are considered if the strains in commercially-available vaccines are not a close match for circulating strains, but autogenous vaccines are only permitted for use on the farms for which the vaccine was generated.

Inactivated virus vaccines against equine IAV first became available in the late 1960s, shortly after the first human IAV vaccines were developed. Subsequently, subunit ('split') vaccines were introduced. Today, a range of different equine IAV vaccines is available; whole inactivated virus, subunit immune-stimulatory complex (ISCOM/ISCOMatrix), live attenuated and recombinant live vector. The live canarypox-vectored vaccine expresses only the HA protein of IAV HA enabling a differentiation of infected from vaccinated animals (DIVA) strategy to be adopted in combination with an ELISA to detect antibodies to NP, which is not present in the vaccine. Demonstration that vaccine effectiveness is compromised if vaccine strains are not updated periodically led to the formation of an OIE (World Organization for Animal Health) Expert Surveillance Panel on Equine Influenza Vaccine Composition, which meets annually.

The first vaccine to protect dogs against the H3N8 IAV strain was launched in 2009 in the USA. Vaccines were developed against the H3N2 IAV soon after it emerged and there is currently a bivalent vaccine available containing both H3N8 and H3N2 strains in the USA.

Concluding Remarks

Swine, equine and canine IAV have been referred to as 'neglected influenza viruses'. It is unlikely that the characteristics of the next pandemic IAV, or where it may emerge from, can be accurately predicted therefore, surveillance of IAV strains circulating in these neglected species is of paramount importance.

Further Reading

- Barba, M., Daly, J.M., 2016. The influenza NS1 protein: What do we know in equine influenza virus pathogenesis? *Pathogens* 5, 57.
- Dou, D., Revol, R., Østbye, H., Daniels, R., 2018. Influenza A virus cell entry, replication, virion assembly and movement. *Frontiers in Immunology* 9, 1581.
- Mostafa, A., Abdelwhab, E.M., Mettenleiter, T.C., Pleschka, S., 2018. Zoonotic potential of influenza A viruses: A comprehensive overview. *Viruses* 10, 497.
- Smith, G.J., Vijaykrishna, D., Bahl, J., *et al.*, 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459, 1122.

Feline Calicivirus (*Caliciviridae*)

Margaret J Hosie and Michaela J Conley, MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Classification

Feline calicivirus (FCV) is a non-enveloped, single stranded, positive sense RNA virus classified within the *Vesivirus* genus of the *Caliciviridae*. Vesiviruses have been shown to infect and cause varying clinical signs among many animal species. FCV was originally isolated in 1957 in New Zealand from cells in culture and has been shown to cause respiratory illness among felids. FCV was originally classified as a picornavirus (known as feline picornavirus for some time) prior to amendment and inclusion within the vesivirus genus of the *Caliciviridae* family. The *Caliciviridae* family was first classified in the third report of the International Committee on Taxonomy of Viruses (ICTV) in 1979. Calicivirus is derived from the latin word for cup, *calix*. The classification was due to the use of a single capsid protein which assembled to form icosahedral particles of 30–40 nm in diameter with 32 cup-shaped depressions on their surface.

Virus Characteristics

In the 1970s, small individual particles of 35 nm in diameter were observed in FCV infected cells by electron microscopy. FCV particles were also seen in association with smooth membranes within cells.

The major capsid protein of FCV, VP1, assembles into capsomeres that form arch-like structures on the surface of the virus, giving rise to the characteristic cup-shaped depressions on the virus particle. The VP1 dimer conformations differ slightly, forming A, B and C isoforms. A/B (bent) and C/C (flat) capsomeres assemble at different positions in the viral particle which results in $t=3$ icosahedral symmetry with 180 subunits (Fig. 1). A/B capsomeres are assembled around the five-fold symmetry axes while C/C capsomeres are located at two-fold symmetry axes (resulting in the alternation of A/B and C/C capsomeres around the three-fold symmetry axes).

The major capsid protein VP1 comprises 3 domains: the shell (S) domain forms the floor of the capsid, P1 makes up part of the viral spikes and P2 (which is an insertion into P1) forms the outer surface of the capsid protrusions and has a role in binding the cellular receptor molecule as well as neutralising antibodies (Fig. 2). The S domain of VP1 contains an eight-stranded β -barrel structure, with strands designated B to I. These strands form two, four-stranded β -sheets (composed of strands BIDG and CHEF) as well as two alpha helices between strands C–D and E–F. The P domain is joined to the S domain via a flexible hinge region, allowing slightly different conformations to be adopted by the A, B and C VP1 isoforms. The P2 domain contains a six stranded β -barrel (A' to F') of varying lengths and positions, as well as a hyper-variable region that is responsible for the variation observed between strains. These hypervariable regions form neutralising epitopes which serve as targets for the host's immune system.

All members of the *Caliciviridae* family are morphologically similar to FCV, although the P2 domains of VP1 can differ, as shown in Fig. 3. For example, some capsomeres form rounded structures e.g., rabbit haemorrhagic disease virus and norovirus GII.10, while other caliciviruses, including FCV, form rhomboid structures.

Genome Organisation and Viral Proteins

FCV encodes a 7.7 kb single stranded, positive sense RNA genome with a VPg protein covalently linked to the 5' end and a 3' polyA tail. An additional RNA species of 2.4 kb (known as subgenomic mRNA) is also produced during infection which

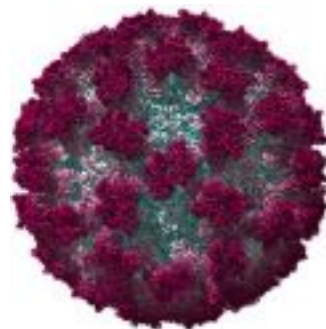


Fig. 1 Structure of FCV strain F9. Three-dimensional structure of FCV solved by cryo-electron microscopy and three-dimensional image reconstruction at 3.0 Å resolution. Domains of the major capsid protein VP1 are coloured: S domain (cyan), P1 domain (pink) and P2 domain (magenta). Image provided by Dr. Michaela J. Conley and Dr. David Bhella (MRC-University of Glasgow, Centre for Virus Research).

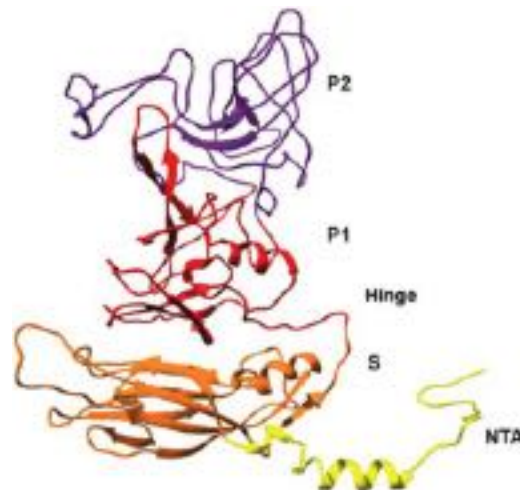


Fig. 2 Structure of major capsid protein VP1. Ribbon diagram of FCV strain F9 major capsid protein VP1, comprising the N-terminal arm (NTA, yellow), the shell domain (S, orange), P1 domain (red) and P2 domain (purple). The hinge region between the S and P1 domains is also shown. Image provided by Dr. Michaela J. Conley and Dr. David Bhella (MRC-University of Glasgow, Centre for Virus Research).

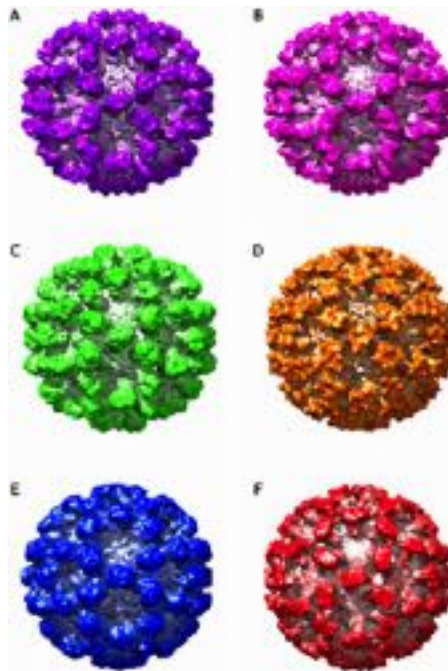


Fig. 3 Comparison of three-dimensional structures of six caliciviruses. Vesivirus 2117 (A), a chimeric sapovirus (B), rabbit haemorrhagic disease virus (C), feline calicivirus F9 strain (D), norovirus genotype II.10 (E) and Norwalk virus (F). Reproduced with permission from Conley, M., Emmott, E., Orton, R., *et al.*, 2017. Vesivirus 2117 capsids more closely resemble sapovirus and lagovirus particles than other known vesivirus structures. *Journal of General Virology* 98 (1), 68–76.

encompasses the coding region of the structural proteins of the virus. FCV encodes three open reading frames (ORFs), ORF1 encodes the non-structural proteins, ORF2 encodes the major capsid protein, VP1, while the minor capsid protein, VP2, is encoded on the third ORF (**Fig. 4**).

The major capsid protein, VP1, forms the majority of the viral capsid (180 copies per particle) with the minor capsid protein, VP2, incorporated to a lesser extent.

The subgenomic RNA forms the template for translation of VP1 which forms a 76 kDa precursor that is subsequently cleaved to yield the mature capsid protein of 62 kDa. This cleavage of VP1 is mediated by the viral protease and is essential for the life cycle of FCV. The leader of the capsid protein (LC) is cleaved post-translationally and has been shown to cause the cytopathic effect

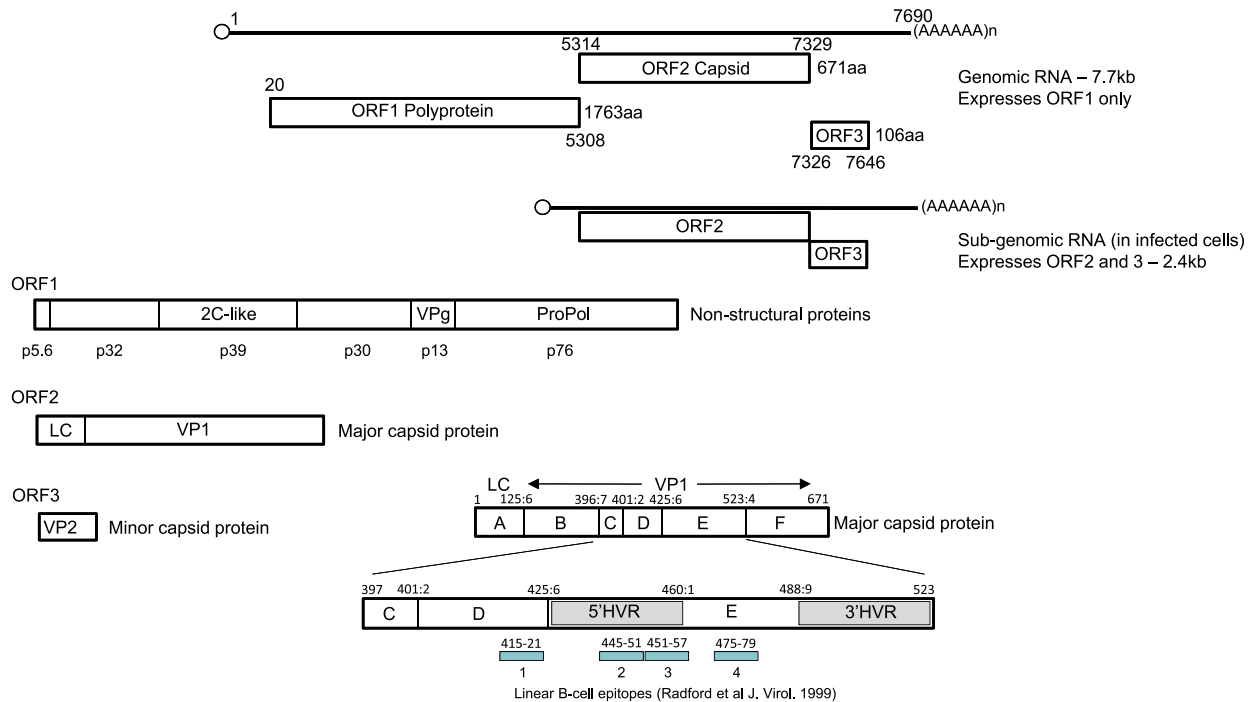


Fig. 4 Genome organisation of FCV. Diagrammatic representation of the open reading frames (ORF) and proteins encoded by the FCV genome and the two RNA species produced during infection.

observed in infected cells. Many viral products form protein-protein interactions. VP1 has been shown to interact with VPg, the protease-polymerase (NS6–7) as well as the minor capsid protein, VP2.

The minor capsid protein, VP2, is crucial for viral replication and release of progeny viral particles. Each virion is predicted to incorporate one or two copies of VP2, which is involved in the self-assembly of the capsid.

Non-structural proteins

Almost all of the non-structural proteins of FCV associate with membranous replication complexes in infected cells, where translation and viral genome replication are co-ordinated. These proteins include NS2, NS3 (helicase) and NS6–7 (protease-RNA dependent RNA polymerase).

Virion protein g (VPg/NS5) is a small protein that is not required for FCV infection, although it is essential for translation of the viral RNA (unless replaced with a cap structure). Polypyrimidine tract binding protein (PTB) is redistributed by FCV to the cytoplasm (from the nucleus), where it interacts with both the genomic and subgenomic viral RNAs via their 5' ends. PTB has been shown to partly co-localise with FCV replication complexes and provides a switch from translation to viral genome replication. VPg is also associated with NS4 in replication complexes which suggests a role in anchoring the small VPg protein near the complexes for close proximity during viral RNA replication.

NS2 interacts with NS3, NS4 and NS6–7 in the membranous replication complexes. NS2, NS3 and NS4 have been proposed as possible transmembrane proteins which localise to the endoplasmic reticulum and may result in membranous rearrangements which are characteristic of FCV infection. NS3 has the ability to prevent the activation of IRF3 (interferon response factor-3) and therefore plays a role in the suppression of the innate immune response during FCV infection.

The FCV protease/polymerase (NS6–7) forms oligomers and interacts with VPg as well as ORF2 (encoding VP1). A weak interaction with VP2 has also been described. The viral polyprotein is cleavage in *cis* by the FCV protease to yield the individual non-structural proteins, however, the protease-polymerase (non-cleaved protein) is thought to be the active form of the RdRp. Nucleolin is essential for FCV replication and interacts with both viral RNA and the protease-polymerase. NS6, appears to reduce cellular translation through the cleavage of poly A binding protein (PABP) and has recently been shown to also impair the formation of stress granules within infected cells.

Life Cycle

FCV infection is restricted to cells of feline origin although this can be overcome by transfection of viral RNA into cells, indicating that the restriction occurs at the stage of binding and entry rather than replication. Attachment of FCV to cells is most efficient at pH 6.5 and occurs as quickly as 15 min post cell exposure. The entry of FCV is pH dependent; the low pH generated by the

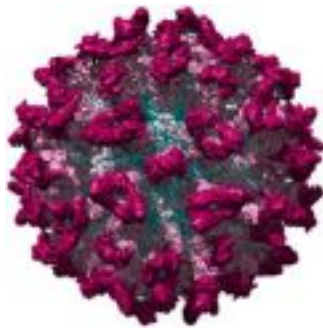


Fig. 5 Three-dimensional reconstruction of FCV decorated with fJAM-A. Cryo-electron microscopy and three-dimensional reconstruction were used to solve the structure of FCV (F9 strain; cyan and pink) decorated with the soluble ectodomain of fJAM-A (magenta) at 3.5 Å resolution. Image provided by Dr. Michaela J. Conley and Dr. David Bhella (MRC-University of Glasgow, Centre for Virus Research).

acidification of endosomes is essential for the uncoating of FCV and the release of the viral RNA into the host cell before replication can ensue. The entry of FCV is dependent upon clathrin mediated endocytosis and the release of the viral genome is likely mediated by the permeabilisation of membranes during FCV entry.

Many viruses utilise carbohydrate moieties (e.g., sialic acid) as attachment factors or as parts of their receptors. Neuraminidase treatment of feline cells (to cleave terminal sialic acid residues) reduced the binding of FCV to cells by approximately 23%, suggesting a potential role of sialic acid in FCV attachment and entry. FCV binds to α 2,6-linked sialic acid (but not α 2,3-linked), which is thought to be present on an N-linked glycoprotein on the surface of susceptible cells. This sialic acid linkage preference could contribute to tissue specificity and cell tropism.

The functional receptor for FCV is feline junctional adhesion molecule A (fJAM-A), which is found at tight junctions between epithelial and endothelial cells. JAM-A can also be found on platelets and leucocytes, as described for both the murine and human JAM-A proteins. FCV infection results in the re-localisation of fJAM-A from tight junctions into the cytosol of the cell. Antibodies targeting JAM-A have been shown to reduce FCV binding to Crandell-Reese feline kidney cells, while the expression of fJAM-A in non-permissive cells rendered them susceptible to infection with FCV. JAM-A contains a short cytoplasmic tail, a transmembrane region and an ectodomain (composed of sub-domains D1 and D2). D1 of fJAM-A is sufficient for FCV attachment and feline JAM-A shares over 75% amino acid identity with its human counterpart. Feline JAM-A contains a predicted N-glycosylation site at position 185, based on this sequence identity.

All fJAM-A domains are necessary for infection with FCV although the PDZ-binding motif (within the cytoplasmic domain of fJAM-A) is not required for virus entry. FCV is able to interact with both monomeric and dimeric forms of fJAM-A with residues D42, K43 and S97 (in D1 of the fJAM-A ectodomain) being important for virus binding. FCV binds both monomeric and dimeric fJAM-A, since the virus binding site and fJAM-A dimerisation site are located at distinct regions/faces of the fJAM-A ectodomain.

The structure of FCV bound to the fJAM-A ectodomain has been determined by cryo-electron microscopy and three-dimensional reconstruction (**Fig. 5**). Binding of fJAM-A to FCV appears to trigger a conformational change in the capsid resulting in the anti-clockwise rotation (15°) of the VP1 capsomeres. Two fJAM-A proteins bind to two VP1 proteins (a single VP1 capsomere) in a head-to-tail manner (i.e., the D1 domain of the first fJAM-A protein lies adjacent to D2 domain of the second fJAM-A protein). The fJAM-A proteins also appear to bind to FCV in their monomeric form. The conformational change observed upon fJAM-A binding likely acts as a priming step for virus entry into the host cell via clathrin-mediated endocytosis.

Epidemiology

FCV is widespread in the domestic cat population, it has been isolated from up to 24% of healthy cats at cat shows and 10% of cats attending veterinary surgeries. The prevalence increases as the number of cats in a household increases, with the highest prevalence observed in large groups of cats housed together in multi-cat households. As many as 25%–40% of cats kept in colonies and shelters are infected, whereas the prevalence in small groups of cats is approximately 10%. The prevalence amongst colonies can range from 50% to 90%.

FCV is host-specific; there are no known reservoirs or alternative hosts and the virus does not infect humans. FCV-like viruses have been isolated from dogs, although it is unlikely that such viruses play a role in the epidemiology of FCV infection. Following infection of cats, FCV is shed in the oral and nasal secretions. The majority of cats shed virus for more than 30 days, although shedding can last for weeks or even several years after the acute clinical signs have resolved.

Pathogenesis

Cats become infected via the nasal, oral or conjunctival routes and the primary site of viral replication is the oropharynx. A transient viraemia occurs 3–4 days following infection and during this period the virus can be detected in many other tissues.

Following acute infection, the virus induces necrosis of epithelial cells and vesicles are observed, typically on the margin of the tongue. The vesicles subsequently develop into ulcers and it has been demonstrated that the dermis becomes infiltrated with neutrophils in the affected regions before the mouth ulcers resolve 2–3 weeks later. Occasionally other tissues are infected, such as the lungs and joints. Focal alveolitis can occur, leading to acute exudative pneumonia and proliferative interstitial pneumonia and lameness has also been observed, with acute synovitis and thickening of the synovial membrane.

The recent emergence of highly virulent systemic strains of FCV (VS-FCV) has been reported in the US, Europe and the UK. Unlike the majority of FCV infections, VS-FCV infected cats develop severe, diverse lesions, including cutaneous oedema and ulcerative lesions on the skin and paws, hepatic necrosis, pancreatitis, pulmonary oedema and disseminated intravascular coagulation. Outbreaks usually occur in multi-cat environments and have been characterised by rapid onset and high mortality. It has been reported that VS-FCV strains emerge independently from genetically distinct FCV strains but attempts to identify genetic patterns within the viral genome that define the highly virulent FCV biotype have been inconclusive so far. The pathogenesis of virulent systemic FCV infection might also include immune mediated components and/or environmental and management factors.

Following recovery from acute disease, the majority of cats clear FCV infection within approximately 30 days, although some cats can shed virus for longer. In healthy carrier cats, FCV is localised within the tonsillar epithelium, but, since tonsillectomy does not terminate the carrier state, it is likely that the virus also persists in other sites. It has been postulated that amino acid mutations in the viral capsid protein might lead to the appearance of FCV variants that escape the host immune response and persist within infected cats.

Control and Prevention

Disease Control

FCV particles can survive for up to 2 weeks on solid surfaces and are stable between pH values ranging from 4 to 8.5 and the virus is resistant to many common disinfectants. Effective disinfectants include sodium hypochlorite (5% bleach diluted at 1:32), potassium peroxy-monosulfate, chlorine dioxide and commercial products that have been approved for their virucidal activity. FCV infections often arise in cat shelters, but virus transmission can be limited, or even prevented, by appropriate shelter design and by instigating management measures to avoid cross infection of cats. For example, cats should be housed individually, only cats originating from the same household should be allowed to mix and flea control should be implemented to minimise the risk of transmission of FCV and other diseases. New cats arriving in shelters that are healthy should be vaccinated as soon as possible, preferably with a modified live virus vaccine to induce protection more rapidly.

FCV can be a major problem in breeding catteries, characterised by upper respiratory disease in kittens as young as 4–8 weeks of age, coinciding with the waning of maternally-derived antibodies. FCV infection can be severe in young kittens, often affecting all kittens in affected litters and some kittens may develop pneumonia and die. Breeding queens should be vaccinated so that kittens acquire high levels of maternally-derived antibody via colostrum and milk, which will protect them from FCV infection for several weeks. It is recommended that breeding queens should be vaccinated prior to mating rather than during pregnancy. Modified live virus vaccines are not licensed for use in pregnant cats and only inactivated vaccines should be used. The earliest age for which FCV vaccines are licensed for kittens is six weeks, but vaccination could be considered in younger kittens if they are at risk of infection. Alternatively, early weaning from four weeks of age will protect kittens against infection from infected breeding queens.

Vaccination

Both modified live and inactivated parenteral FCV vaccines are available. Modified live intranasal vaccines can still be used in US but are not available in Europe. It is thought that FCV vaccines provide protection mainly via the induction of virus neutralising antibodies, although cellular immunity has not been extensively examined following either FCV vaccination or infection. As FCV rapidly mutates and gives rise to viral variants, it has been suggested that field isolates could evolve and become resistant to vaccines based on strains that have been used for several decades; indeed, vaccine manufacturers are seeking to identify and introduce newer vaccine strains that might provide wider cross protection. At present, the most commonly used vaccine strains are F9 (the oldest strain that was isolated in the 1950s), FCV 255 and two more recent strains, G1 and 431.

Several *in vitro* studies have provided evidence that older vaccine strains are less effective at cross-neutralising current circulating field strains isolated from sick cats, but these findings are controversial and further independent studies will be required to investigate vaccine efficacy against different FCV strains. Nevertheless, it is generally recommended that all healthy cats should be vaccinated against FCV, since vaccination provides good protection against acute oral and upper respiratory tract disease in the majority of cats. However, sterilising immunity is not induced, and so vaccinated cats can still become infected and subsequently shed FCV. It is important to consider that current vaccines are not able to protect cats equally well against all the currently circulating field strains of FCV.

Further Reading

- Conley, M.J., McElwee, M., Azmi, L., *et al.*, 2019. Calicivirus VP2 forms a portal-like assembly following receptor engagement. *Nature* 565 (7739), 377–381.
 Coutts, A.J., Dawson, S., Willoughby, K., Gaskell, R.M., 1994. Isolation of feline respiratory viruses from clinically healthy cats at UK cat shows. *Veterinary Record* 135, 555–556.

- Hansman, S.G., Jiang, X.J., Green, K.Y. (Eds.), 2010. Feline Calicivirus by Christine Luttermann and Gregor Meyers in *Caliciviruses: Molecular and Cellular Virology* 2010. Caister Academic Press.
- Pedersen, N.C., Elliott, J.B., Glasgow, A., Poland, A., Keel, K., 2000. An isolated epizootic of hemorrhagic-like fever in cats caused by a novel and highly virulent strain of feline calicivirus. *Veterinary Microbiology* 73, 281–300.

Relevant Websites

- <http://www.abcdcatsvets.org/feline-calicivirus-infection-2012-edition/>
Feline Calicivirus infection.
- <https://icatcare.org/advice/cat-health/feline-calicivirus-fcv-infection-0>
Feline calicivirus (FCV) infection – International Cat Care.

Feline Leukemia and Sarcoma Viruses (*Retroviridae*)

Brian J Willett and Margaret J Hosie, MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J.C. Neil, Feline Leukemia and Sarcoma Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00397-6.

History

In the early 1960s, Harry Pfaff, a veterinary practitioner in the Glasgow area, observed a cluster of lymphoma cases in a local multi-cat household. Suspecting a possible infectious etiology, he contacted Professor William Jarrett, a veterinary pathologist at the University of Glasgow. Using electron microscopy, Professor Jarrett and his colleagues visualized virus-like particles in tumor material from the cats, and demonstrated that these particles bore a striking resemblance to the virus associated with murine leukemia, reporting their findings in *Nature* in 1964. Subsequent studies demonstrated that these feline leukemia viruses (FeLVs) could be transmitted to cats where they induced lymphosarcomas and a range of degenerative diseases, including anemias and thymic atrophy.

Following these studies, Snyder and Theilen isolated a retrovirus from a feline fibrosarcoma that rapidly reproduced this tumor on inoculation into experimental cats. It is now recognized that feline sarcoma viruses (FeSVs) arise by recombination between FeLV and cellular proto-oncogenes and that, in contrast to FeLV, these viruses are not transmitted from cat to cat.

FeLV remains an important virus clinically, affecting both domestic cats and wild felids. Several highly efficacious FeLV vaccines are available and are used widely. This, in conjunction with reliable testing and control programmes, has led to a significant decline in the prevalence of FeLV infection (<2%) across Europe and North America. Prevalence remains high in some areas, for example N.W. China (~11%) and Thailand (~25%).

As a naturally occurring disease in an outbred host, FeLV has served as a paradigm for the natural history and molecular pathogenesis of the gammaretrovirus subfamily. It also played a foundational role in cancer genetics as a transducing agent which led to the discovery of novel cellular transforming genes.

Taxonomy and Classification

FeLVs are RNA viruses and belong to the family *Retroviridae*. They are further classified in the genus *Gammaretroviruses*.

Virion and Genome Structure

The virion particles are around 100 nm in diameter and consist of an outer membrane derived from the host cell surrounding a spherical core particle. The core encapsidates the viral genome which, as in other members of this viral family, is present as two linear single-stranded RNA molecules linked as a dimer. The virion RNA is positive-stranded and resembles cellular RNA having a 5' cap and a 3' poly(A) tract. As deduced from sequencing of proviral forms, the FeLV genome is around 8 kb long with a 67 base terminal redundancy and the gene order *gag-pol-env*.

The particles have surface spikes composed of multimers of the two *env*-coded proteins, the gp70 surface glycoprotein (SU), and the p15E transmembrane anchor protein (TM). Inside the envelope are the structural *gag* gene products which form a spherical core particle composed of p27, the major capsid protein (CA), with an outer layer formed by the p15 matrix protein (MA). Another *gag* product, the p10 nucleocapsid (NC), is associated with the virion RNA. Other minor virion proteins encoded by the *pol* gene comprise the protease (PR), reverse transcriptase (RT), and integrase (IN) enzymes.

Replication

Following attachment to specific host cell-surface receptors and fusion of the viral envelope with the cellular membrane, the viral core is internalized into the cytoplasm of the cell. Uncoating is followed by reverse transcription in which the virion RNA is converted into a double-stranded DNA form by the virion RT which uses a proline tRNA primer and carries a ribonuclease H function that degrades the virion RNA. After nuclear translocation, viral integration is catalyzed by the IN protein and involves the creation of a staggered cut in cellular DNA with a consequent 4 bp duplication of host DNA at the insertion site.

The proviral form of FeLV is flanked by long terminal repeats (LTRs) of 480–560 bp. These are generated during reverse transcription by duplication of unique sequences at the 3' (U3) and 5' (U5) ends of the RNA genome. The LTRs contain promoters and enhancers that drive transcription of viral RNA and also processing signals for cleavage and polyadenylation of the RNA transcripts. The 5' LTR functions to initiate transcription while the 3' LTR acts primarily as an RNA processing signal.

The virion RNA can function as a messenger RNA for Gag and Pol products, while a spliced subgenomic mRNA of around 3 kb encodes the Env products. Most full-length RNA translation products terminate at the 3' end of *gag* to produce the Pr65gag precursor, while a small percentage read through into *pol* by misreading of a UAG termination codon, generating the Pr180Gag-pol precursor. During and after the budding process, the virion aspartyl PR catalyzes the cleavage of both precursor proteins to their mature forms. The envelope gene products are synthesized as a Pr80env precursor and processed by cellular PRs to the mature, disulfide-linked gp70 and p15E envelope proteins. The *gag* gene is also abundantly expressed in an alternative, glycosylated form *via* an upstream AUG codon. This product is expressed on the cell surface and shed after cleavage by cellular PRs. It is dispensable for *in vitro* virus replication but is highly conserved and may play a role *in vivo*, for example in counteracting intrinsic immune activity as has been described for the equivalent gene product in murine leukemia virus.

Assembly of virus particles occurs at the cell surface by extrusion of cores which form at the budding site, concomitantly acquiring a host cell-derived outer membrane with virus-coded surface spikes. Virus replication and release is often noncytopathic.

Geographic Distribution

FeLV occurs worldwide in domestic cat populations, although prevalence varies significantly and has declined in pet cats in areas where reliable testing, vaccination and control have been instigated. FeLV has also been isolated from several wild felids, including the European wild cat (*Felis silvestris*), Florida panther (*Puma concolor coryi*), Iberian lynx (*Lynx pardinus*) and Jaguar (*Panthera onca*). Endogenous retroviral sequences closely matched to FeLV are found in the sand cat (*Felis margarita*) and the jungle cat (*Felis chaus*). Although not a direct source of disease, these endogenous sequences may potentially participate in recombination with exogenous FeLV to generate variant viruses with altered host range.

Epidemiology

The outcomes of FeLV infection fall into several categories, which vary in likelihood according to the age and immune status of the exposed host. The majority of cats undergo a transient infection lasting up to 3 months, during which they are viremic and shed virus. They then develop neutralizing antibody and concomitantly clear infectious virus and, a little later, virus antigen from the blood stream. Some cats appear to clear virus infection successfully, while others may harbor latent virus in the bone marrow for some years. A recent study based on analysis of virus loads by polymerase chain reaction (PCR) suggests that animals that control viremia may be further subdivided into either abortive, or regressive, sometimes known as latent, infection. Even where latent infection persists, reactivation is not a frequent event and the vast majority of these cats do not develop an FeLV-related disease. However, at this stage virus can be reactivated by immunosuppression. The final group (persistent or progressive infection) remains actively infected, shedding virus from epithelial surfaces and displaying high titer plasma viremia. Such cats may remain apparently healthy for 2–3 years before succumbing to an FeLV-related disease.

The proportion of cases falling into these groups differs between multi-cat households and those households containing one or two free-ranging cats. In the former case, the introduction of an FeLV carrier results in repeated exposure of susceptible cats, often at a young age so that up to 30%–40% become persistently viremic and at risk of disease. Approximately 50% of free-ranging urban and suburban cats have serological evidence of exposure to FeLV but only 1%–5% of these cats are actively infected and the disease incidence is correspondingly lower.

Dual infection with FeLV and feline immunodeficiency virus (FIV) occurs and is associated with rapid disease onset, particularly if cats with preexisting FeLV infection encounter FIV. Rapid death of dually infected animals may reduce the apparent overlap of these agents in the field, but the populations at risk of infection also differ. FIV infection rates increase directly with age while persistent FeLV infection has a peak incidence in young cats.

FeLV Subgroups and Host Range

FeLV isolates were initially classified into three subgroups; A, B, C according to their viral interference properties in feline fibroblast cells *in vitro*. Viruses of a given subgroup prevent superinfection (interfere) with other viruses of the same subgroup (Table 1). This property is based on the use of three different host cell-surface receptors by FeLV-A, B, and C, and the blockade and/or down-modulation of these receptors in persistently infected cells by viral Env glycoproteins. Primary receptors for subgroups A, B, and C have been identified and shown to be transmembrane transporter molecules which the virus has subverted to gain entry to the host cell (Table 1). Subsequent studies have identified additional, less frequent viral variants with distinct interference properties and receptor usages, although the contributions of such viral variants to the pathogenesis of disease in the wider cat population remains unclear. The generation of novel variants occurs by either mutation or recombination. During the process of mutation, viruses may emerge with an intermediate phenotype, for example isolate FY981 can use either the FeLV-A or C receptors for infection. Notable variants include FeLV-T, a “lymotropic” variant with complex entry requirements. FeLV-T does not replicate well in fibroblast cells and appears to use the FeLV-A receptor in conjunction with an auxiliary mechanism in which a truncated *env* gene product encoded by endogenous FeLV sequences (FeLIX) is used as a co-receptor through binding to the FeLV-B receptor.

Table 1 Properties of feline leukemia virus subgroups

Subgroup	Origin	Receptor	Function	Pathogenesis
A	Exogenous	feTHTR1	Thiamine transport	Minimally pathogenic to acute immunosuppression
B	Recombination FeLV-A × endogenous FeLV	Pit-1 (Pit-2)	Phosphate transport	More common in leukemic cats Some isolate-specific diseases, <i>e.g.</i> , FeLV-GM1 myeloid leukemia
C	Mutation of FeLV-A (Env vrA)	FLVCR1 (FLVCR2)	Heme export	Erythroid hypoplasia
D	Recombination FeLV-A × endogenous retrovirus ERV-DC	Unknown	Unknown	Unknown
T	Mutation of FeLV-A (outside RBD)	FeLV-A receptor + FeLIX, Pit-1		Acute immunosuppression
TG35-2	Mutation of FeLV-A (Env vrA)	feRFC	Reduced folate transport	Unknown

FeLV-D is formed by recombination between exogenous FeLV and the endogenous cat retrovirus ERV-DC. FeLV-D appears to use a distinct receptor to the A, B and C subgroups. FeLV-TG35-2 is a virus isolated from a single cat in Japan and uses the feline reduced folate carrier as an entry receptor.

Natural isolates contain either subgroup A alone, or mixtures of subgroups A + B, A + C, or A + B + C. FeLV isolates of subgroup A have a limited tropism in cell culture due to their use of the receptor THTR1 (*e.g.*, FEA and HEK293 cells), whereas subgroup B and C isolates have a greatly expanded host range due to their use of Pit-1 and FLVCR1 respectively, infecting cell lines from diverse species, including cat, human, mink, and canine cells. FeLV infection is generally noncytopathic and persistent and the virus is commonly propagated in long-term cultures of embryo-derived fibroblasts. Some strains such as FeLV-T or FeLV-C are cytopathic or induce apoptosis in cells of lymphoid origin *in vitro*, reflecting their *in vivo* pathogenic properties.

Clinical Features and Pathology

Of those cats that become persistently viremic following FeLV exposure, over 80% die within 3.5 years. Most young cats infected with FeLV die from degenerative diseases rather than from tumors. Profound immunosuppression associated with thymic atrophy is a common finding in kittens. Other diseases seen in FeLV-infected cats include enteritis, immune complex glomerulonephritis, pancytopenia, and hemolytic anemia. Erythroid hypoplasia, an acute disease involving failure of red cell development past the burst-forming unit (BFU) stage, is specifically associated with FeLV subgroup C.

The most common neoplasm induced by FeLV is lymphosarcoma of T-cell origin, usually restricted to the thymus or sometimes occurring as a multicentric tumor in lymph nodes. The tumors often develop between 1 and 3 years after infection and the first signs may be chronic wasting and anemia. At presentation, the normal architecture of the lymphoid organ has usually been destroyed by a monomorphic infiltrate of lymphoblastic cells. The thymic tumors frequently display a rearrangement of the T-cell antigen receptor β -chain gene and may also express the co-receptor molecules CD4 and/or CD8. FeLV is also commonly associated with myeloid leukemias and a myelodysplastic syndrome-like disease, as well as with other forms of hematopoietic malignancy. Multicentric fibrosarcoma is a rare sequel to FeLV infection but these tumors are often associated with the *de novo* generation of a FeSV. FeLV is also found in association with 35% of alimentary tumors, primarily of B cell origin, but the virus is not always clonally integrated in these tumors and the role of the virus is, therefore, unclear. FIV can also increase the frequency of malignant diseases in cats, and some rare cases arise on a background of dual infection.

Envelope Gene Variation and Pathogenicity

The common infectious form of FeLV is FeLV-A which is a remarkably highly conserved virus as shown by sequence analysis of several strains and serotypic analysis of a much larger number. FeLV variants frequently arise from FeLV-A by mutation and recombination, and such variants are often implicated in the acute diseases that develop in persistently infected cats. The variant viruses generated from FeLV-A are generally dependent on the continued presence of the prototype for their propagation *in vivo*. However, as the variants are transmitted less efficiently and are in some cases rapidly fatal, they tend to die out with the host while the prototype FeLV-A continues to colonize new hosts (Fig. 1).

The most commonly isolated FeLV recombinants are subgroup B viruses (Table 1). These are derived by recombination between FeLV-A and endogenous FeLV-related proviruses which are found in the genome of the domestic cat and related small feline species. Although the endogenous FeLV-related proviruses all appear to be replication defective, their envelope genes can be rescued by the recombination process leading to the generation of FeLV-B viruses. FeLV-B can infect cells refractory to, or already containing, FeLV-A by virtue of their distinctive receptor specificity.

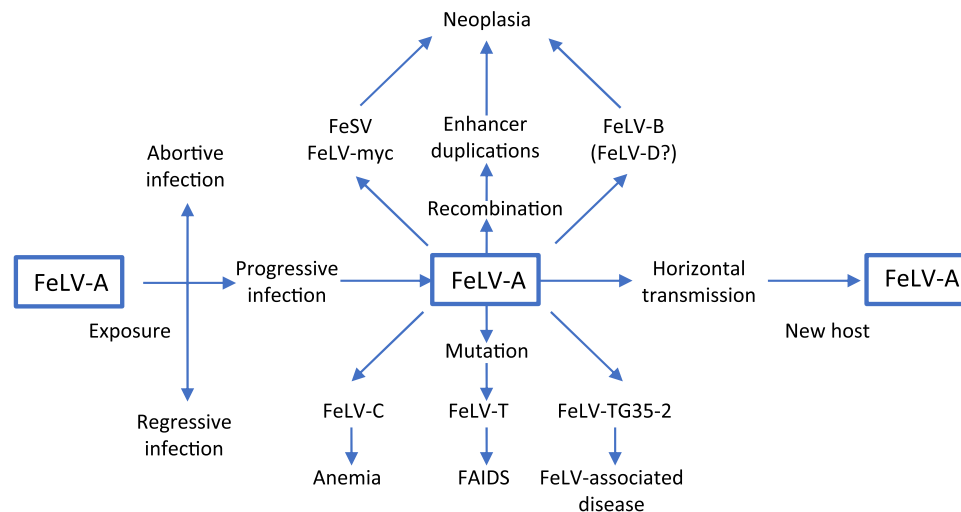


Fig. 1 Life cycle of FeLV in its natural host, the domestic cat. FeLV-A is the most readily transmitted form and is found in 100% of isolates. In cats which become persistently viremic (progressive infection), the virus may evolve by recombination or point mutations to generate pathogenic variants that can lead to the rapid demise of the infected host animal. Multiple variants can arise in a single host. Few of these variants show any capacity for horizontal transmission to new hosts with the exception of FeLV-B. In this way, FeLV inflicts a substantial disease burden without significantly reducing host numbers.

Table 2a FeLV gene transduction in neoplasia

Gene	Normal function of host gene product	Associated tumor	Examples ^a
<i>abl</i>	Plasmamembrane protein kinase	Fibrosarcoma	FeSV-HZ2
<i>fes</i>	Plasmamembrane protein kinase	Fibrosarcoma	FeSV-GA, -HZ1, -ST
<i>fgr</i>	Plasmamembrane protein kinase	Fibrosarcoma	FeLV-GR, -TP1
<i>fms</i>	Receptor protein kinase (CSF-1 receptor)	Fibrosarcoma	FeSV-SM, -HZ5
<i>kit</i>	Receptor protein kinase (SCF receptor)	Fibrosarcoma	FeSV-HZ4
<i>myc</i>	Transcription factor	T-cell lymphoma	FeLV (T3, T17, FTT)
<i>Notch2</i>	Transmembrane receptor	T-cell lymphoma	(Inoculum FeLV-61E) ^a
<i>sis</i>	Growth factor (B chain PDGF)	Fibrosarcoma	FeSV-PI
<i>trc</i>	T-cell antigen receptor (β -chain)	T-cell lymphoma	FeLV-T17

^aIsolated from naturally occurring tumors apart from the indicated exception.

The anemia-inducing FeLV-C isolates are rarer and appear to be derived from FeLV-A by mutation of a single variable domain (VRA) of the *env* gene. The acute disease properties of these variants appear to be due to compromised viability of erythroid progenitor cells due to perturbation of the FeLV-C receptor, a vital heme exporter.

Minor *env* mutations also appear to give rise to the acutely immunosuppressive FeLV-FAIDS variants. The prevalence of acutely immunosuppressive viruses in nature is unknown, but immunosuppressive disease is a common manifestation of FeLV infection.

The relationship of subgroup variation to oncogenesis is complex. FeLV-B recombinants are more common in tumor-bearing than in healthy infected cats. This higher frequency might reflect merely longer-standing infection, but some FeLV-B-containing isolates have an altered spectrum of neoplastic disease. For example, FeLV-GM1, which contains a replication-defective FeLV-B component, induces mainly myeloid leukemia.

FeLV Oncogenesis: Virus Evolution and Mutagenesis of Cellular Oncogenes

Two modes of virus-induced host gene mutation have been described in FeLV-associated cancers. The first is 'transduction', where recombination leads to the generation of an acutely oncogenic variant in which viral gene sequences are replaced by a host-derived insert. Such viruses are replication defective and are found in nature in association with a replication-competent FeLV helper virus.

Multicentric fibrosarcomas of young cats are relatively rare, but are generally FeLV positive and frequently involve a novel sarcoma virus. Similarly, transduction of *c-myc* has been observed in up to 20% of naturally occurring thymic lymphosarcomas in FeLV positive cats. In all, nine different host cell genes have been shown to be transduced by FeLV (Table 2a). The transducing viruses induce tumors with short latency in cats and in the case of FeSVs may transform cells in tissue culture.

Table 2b Insertional mutagenesis and FeLV oncogenesis

<i>Common integration site</i>	<i>Gene function</i>	<i>Tumor</i>
<i>fit-1</i>	Transcription factor (<i>c-myc</i>)	T-cell lymphoma
<i>flvi-1</i>	Unknown	non-T, non-B lymphoma
<i>flvi-2 (bmi-1)</i>	Transcription factor	T-cell lymphoma
<i>c-myc</i>	Transcription factor	T-cell lymphoma
<i>pim-1</i>	Protein kinase	T-cell lymphoma
<i>flit-1</i>	Growth factor receptor?	T cell lymphoma

Alternatively, host genes can be affected by proviral 'insertional mutagenesis' (*cis*-activation). Four known oncogenes and two uncharacterized novel integration loci have been identified as common tumor-specific insertion sites for FeLV in thymic lymphosarcomas (Table 2b). In this respect FeLV oncogenesis appears remarkably similar to that of the murine gammaretroviruses.

Changes within the LTR are also a feature of tumor-associated FeLV. In thymic lymphosarcomas, sequence duplications of the core enhancer domain are found frequently and have been shown to arise *de novo* from infection with molecularly cloned FeLV isolates lacking such features. By analogy with the murine oncoretroviruses, the duplications are likely to increase the oncogenicity of the virus and reduce the latent period for tumor formation, possibly by increasing the potency of viral enhancer activity on nearby cellular promoters. There is evidence that these adaptive changes to the LTR operate tissue-specifically and proviruses carrying different duplications of LTR regions 5' and 3' to the core-enhancer region have been identified in myeloid leukemias and non-T, non-B splenic lymphomas, respectively. Chimeric murine retroviruses carrying the FeLV enhancer region have been generated, confirming that the tissue specificity is carried by this structure.

Immune Response

Unlike infection with the lentiviruses HIV and FIV, FeLV infection of cats can lead to recovery. There is evidence that both neutralizing antibodies and virus-specific cytotoxic T cells (CTLs) play a role in resistance as either can be used to prevent or limit infection by passive transfer. In natural infection, CTLs to virus structural protein epitopes can be detected as early as 1 week after infection. In contrast, neutralizing antibodies may not be detected until approximately 6 weeks post infection.

In the early literature on FeLV, a distinction was made between antiviral immunity and antitumor immunity. Cats with antitumor (FOCMA) antibody were thought to be protected from tumor development. This antibody response is now believed to be directed to endogenous FeLV *env* proteins and its role in modulating tumor development is uncertain.

Transmission

Cats persistently infected with FeLV shed virus in their saliva, urine, and feces but, as the virus is fragile, close contact is required for transmission. The most frequent routes involve saliva and transplacental spread. Kittens infected *in utero* become persistently infected, but the consequences of infection in older cats depend on a number of factors. There is an age-related resistance to infection such that cats up to 12 weeks of age are highly susceptible, but above 16 weeks they are difficult to infect either naturally or experimentally.

FeLV subgroup A is always found in field isolates while approximately half also contain FeLV-B, whereas FeLV-C is present in only 1%–2% of isolates. Although FeLV-B can arise *de novo* by recombination, it can also be transmitted between cats. This occurrence is dependent on pseudotype formation in which the genome of the B virus becomes enclosed in an envelope containing glycoproteins of the A subgroup.

Prevention and Control

Successful control measures can be adopted in multi-cat households by removing or isolating persistently infected animals. Productively infected cats are identified routinely by the detection of soluble FeLV antigen (p27) by ELISA or immunomigration test, proviral DNA PCR on whole blood, RT-PCR for viral RNA in blood or saliva, or less frequently by direct virus isolation in culture. A few cats remain persistently antigenemic but nonviremic. These cats do not usually transmit the virus unless they are shedding virus in the milk or saliva.

Numerous vaccine strategies have been shown to confer protection against FeLV under laboratory conditions and FeLV was the first retrovirus for which commercial vaccines were developed. Vaccines in current use include whole inactivated virus preparations, subunit vaccines comprising recombinant viral Env protein expressed in bacterial cells, and a canarypox recombinant virus

expressing Gag and Env. These vaccines offer significant protection against experimental challenge and are under evaluation for longer-term efficacy in the field. These vaccines do not generate sterilizing immunity but appear to prime the immune system to favor clearance of virus infection instead of persistent viremia.

Future Perspectives

There is continuing interest in control of FeLV infection due to its importance in veterinary medicine. In the future we can look forward to improvements in vaccine efficacy and further dissection of the host responses that confer protection. While the focus of attention of cancer genetics has moved on to more easily manipulated models, FeLV remains as a useful touchstone for our understanding of retroviral pathogenesis in an outbred, naturally infected host.

Further Reading

- Anai, Y., Ochi, H., Watanabe, S., *et al.*, 2012. Infectious endogenous retroviruses in cats and emergence of recombinant viruses. *Journal of Virology* 86, 8634–8644.
- Hanlon, L., Barr, N.I., Blyth, K., *et al.*, 2003. Long-range effects of retroviral activation on c-myc over-expression may be obscured by silencing during tumor growth *in vitro*. *Journal of Virology* 77, 1059–1068.
- Lutz, H., Addie, D., Belák, S., *et al.*, 2009. Feline leukaemia. ABCD guidelines on prevention and management. *Journal of Feline Medicine and Surgery* 11, 565–574.
- Mendoza, R., Anderson, M.M., Overbaugh, J., 2006. A putative thiamine transport protein is a receptor for feline leukemia virus subgroup A. *Journal of Virology* 80, 3378–3385.
- Miyazawa, T., 2002. Infections of feline leukemia virus and feline immunodeficiency virus. *Frontiers in Bioscience* 7, D504–D518.
- Roca, A.L., Nash, W.G., Menninger, J.C., Murphy, W.J., O'Brien, S.J., 2005. Insertional polymorphisms of endogenous feline leukemia viruses. *Journal of Virology* 79, 3979–3986.
- Studer, N., Lutz, H., Saegerman, C., *et al.*, 2019. Pan-European study on the prevalence of the feline leukaemia virus infection – Reported by the European advisory board on cat diseases (ABCD Europe). *Viruses* 11, 993–1020.
- Torres, A.N., Mathiason, C.K., Hoover, E.A., 2005. Re-examination of feline leukemia virus: Host relationships using real-time PCR. *Virology* 332, 272–283.
- Tsatsanis, C., Fulton, R., Nishigaki, K., *et al.*, 1994. Genetic determinants of feline leukemia virus-induced lymphoid tumors: patterns of proviral insertion and gene rearrangement. *Journal of Virology* 68, 8294–8303.
- Willett, B.J., Hosie, M.J., 2013. Feline leukaemia virus: Half a century since its discovery. *The Veterinary Journal* 195, 16–23.

Fish and Amphibian Alloherpesviruses (*Herpesviridae*)

Maxime Boutier, Léa Morvan, and Natacha Delrez, University of Liège, Liège, Belgium

Francesco Origgi, University of Bern, Bern, Switzerland

Andor Doszpoly, Hungarian Academy of Sciences, Budapest, Hungary

Alain Vanderplasschen, University of Liège, Liège, Belgium

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AcHV Acipenserid herpesvirus
AngHV Anguillid herpesvirus
BAC bacterial artificial chromosome
BfHV Bufonid herpesvirus
CyHV Cyprinid herpesvirus
dsDNA double-stranded DNA
EsHV Esocid herpesvirus
GaHV Gadid herpesvirus
IchV Ictalurid herpesvirus

ICTV International Committee on Taxonomy of Viruses
kbp kilobase pair
ORF open reading frame
PeHV Percid herpesvirus
RaHV Ranid herpesvirus
SalHV Salmonid herpesvirus
SiHV Silurid herpesvirus
TK thymidine kinase
TR terminal repeat

Glossary

Alloherpesvirus Herpesvirus classified in the family *Alloherpesviridae*, order *Herpesvirales*, grouping all herpesviruses infecting fish and amphibians described until now.

Bacterial artificial chromosome Large DNA vector that can allow the stable maintenance and efficient mutagenesis of herpesvirus genome in *Escherichia coli*, followed by the reconstitution of progeny virions after its transfection into permissive eukaryotic cells.

Behavioral fever Phenomenon by which ectotherms increase their body temperature by moving to warmer places in response to an infection.

Ectotherm Organism unable to internally self-regulate its body temperature, therefore relying on external sources of heat.

Hyperplastic lesion Increase of the size of a tissue or organ as a result of cell proliferation.

Latency Maintenance of viral genome as a non-integrated episome associated with minimal viral protein expression and absence of infectious particle production.

Viral reactivation Phenomenon by which the state of viral infection switches back from latency to productive infection.

Introduction and Classification

Herpesviruses form a diverse group of viruses initially described by a conserved virion structure containing a large, linear double-stranded DNA (dsDNA) genome. Indeed, the unifying property of herpesviruses is virion morphology rather than genetic content. Herpesviruses are now classified in the order *Herpesvirales* and share only one convincing conserved gene, i.e., the ATPase subunit of the terminase. Phylogenetic analysis of this gene has led to the current classification of herpesviruses by the International Committee on Taxonomy of Viruses (ICTV). The order *Herpesvirales* is now composed of three viral families, namely the families *Herpesviridae*, *Alloherpesviridae* and *Malacoherpesviridae*. Until now, herpesviruses of amniotes (reptiles, birds and mammals) have been grouped in the family *Herpesviridae*. Herpesviruses of molluscs are encompassed within the family *Malacoherpesviridae*. Finally, herpesviruses of fish and amphibians are classified in the family *Alloherpesviridae*.

The family *Alloherpesviridae* contains four genera with 13 virus species currently accepted by the ICTV (Table 1). The alloherpesviruses of amphibians (*Ranid herpesvirus* 1; 2 and 3 [RaHV]) belong to the genus *Batrachovirus*, while fish alloherpesviruses have been clustered into three genera. The genus *Cyprinivirus* comprises the alloherpesviruses of cyprinids (*Cyprinid herpesvirus* 1; 2 and 3 [CyHV]) and that of eel (*Anguillid herpesvirus* 1 [AngHV]). The genus *Ictalurivirus* includes the alloherpesviruses isolated from catfish (*Ictalurid herpesvirus* 1 and 2 [IchV]) and sturgeon species (*Acipenserid herpesvirus* 2 [AcHV]), while the genus *Salmonivirus* consists of herpesviruses of salmonids (*Salmonid herpesvirus* 1; 2 and 3 [SalHV]) (Fig. 1(A)). Interestingly, viruses of the genera *Cyprinivirus* and *Ictalurivirus* are able to infect fish from two different superorders (Elopomorpha, Ostariophysi) or subclasses (Chondrostei, Neopterygii), respectively. This observation supports a more recent evolution of the family *Alloherpesviridae*, associated with a lesser degree of coevolution of some alloherpesviruses with their hosts. In addition to the accepted viral species, more than 10 putative alloherpesviruses were detected by PCR from a wide range of fish species (Table 1). Unfortunately, only partial genome sequences are available. The official taxonomical classification of these viruses is still pending. However, according to phylogenetic analyses, these viruses undoubtedly belong to the family *Alloherpesviridae*. Some of them clearly cluster into already existing genera (Fig. 1(B)), e.g., SalHV-4 and SalHV-5 are likely going to expand the genus *Salmonivirus*. Other viruses (e.g., AcHV-1; EsHV-1) cannot be grouped with viruses of any existing genera (Fig. 1(B)). Further genomic data are needed from these virus species in order to establish their evolution and taxonomic classification. The

Table 1 Classification of fish and amphibian alloherpesviruses according to the ICTV (Order *Herpesvirales*, Family *Alloherpesviridae*)

Genus	Viral species (ICTV list)	Viral name	Viral acronym	Alternative viral name (s)	Susceptible host (s)
<i>Batrachovirus</i>	<i>Ranid herpesvirus 1</i> ^a	Ranid herpesvirus 1	RaHV-1	Lucké tumor herpesvirus	Northern leopard frog (<i>Lithobates pipiens</i>), green frog (<i>L. clamitans</i>), pickerel frog (<i>L. palustris</i>)
	<i>Ranid herpesvirus 2</i>	Ranid herpesvirus 2	RaHV-2	Frog virus 4	Northern leopard frog (<i>Lithobates pipiens</i>)
	<i>Ranid herpesvirus 3</i>	Ranid herpesvirus 3	RaHV-3		Common frog (<i>Rana temporaria</i>)
<i>Cyprinivirus</i>	<i>Anguillid herpesvirus 1</i>	Anguillid herpesvirus 1	AngHV-1	European eel herpesvirus	Japanese and European eel (<i>Anguilla japonica</i> and <i>A. anguilla</i>)
	<i>Cyprinid herpesvirus 1</i>	Cyprinid herpesvirus 1	CyHV-1	Carp pox herpesvirus	Common carp (<i>Cyprinus carpio</i>)
	<i>Cyprinid herpesvirus 2</i>	Cyprinid herpesvirus 2	CyHV-2	Goldfish haematopoietic necrosis virus	Goldfish (<i>Carassius auratus</i>), gibel carp (<i>Carassius gibelio</i>)
	<i>Cyprinid herpesvirus 3</i> ^a	Cyprinid herpesvirus 3	CyHV-3	Koi herpesvirus	Common carp (<i>Cyprinus carpio</i>)
<i>Ictalurivirus</i>	<i>Acipenserid herpesvirus 2</i>	Acipenserid herpesvirus 2	AcHV-2	White sturgeon herpesvirus 2	Sturgeon (<i>Acipenser</i> spp.)
	<i>Ictalurid herpesvirus 1</i> ^a	Ictalurid herpesvirus 1	IchHV-1	Channel catfish virus	Channel catfish (<i>Ictalurus punctatus</i>), striped catfish (<i>Pangasius hypophthalmus</i>)
	<i>Ictalurid herpesvirus 2</i>	Ictalurid herpesvirus 2	IchHV-2	Ictalurus melas herpesvirus	Black bullhead catfish (<i>Ameiurus melas</i>)
<i>Salmonivirus</i>	<i>Salmonid herpesvirus 1</i> ^a	Salmonid herpesvirus 1	SalHV-1	Herpesvirus salmonis	Rainbow trout (<i>Oncorhynchus mykiss</i>)
	<i>Salmonid herpesvirus 2</i>	Salmonid herpesvirus 2	SalHV-2	<i>Oncorhynchus masou</i> herpesvirus	Salmon and trout (<i>Oncorhynchus</i> spp.)
	<i>Salmonid herpesvirus 3</i>	Salmonid herpesvirus 3	SalHV-3	Epizootic epitheliotropic disease virus	Lake trout (<i>Salvelinus namaycush</i>)
<i>Unclassified</i>	N/A	Acipenserid herpesvirus 1	AcHV-1	White sturgeon herpesvirus 1	White sturgeon (<i>Acipenser transmontanus</i>)
	N/A	Bufoid herpesvirus 1	BfHV-1		Common toad (<i>Bufo bufo</i>)
	N/A	Cyprinid herpesvirus 4	CyHV-4	Sichel herpesvirus	Sichel (<i>Pelecus cultratus</i>)
	N/A	Esocid herpesvirus 1	EsHV-1	Blue spot disease virus	Northern pike (<i>Esox lucius</i>)
	N/A	Gadid herpesvirus 1	GaHV-1	Atlantic cod herpesvirus	Atlantic cod (<i>Gadus morhua</i>)
	N/A	Percid herpesvirus 2	PeHV-2	European perch herpesvirus	European perch (<i>Perca fluviatilis</i>)
	N/A	Salmonid herpesvirus 4	SalHV-4	Atlantic salmon papillomatosis virus	Atlantic salmon (<i>Salmo salar</i>)
	N/A	Salmonid herpesvirus 5	SalHV-5	Namaycush herpesvirus	Lake trout (<i>Salvelinus namaycush</i>)
	N/A	Silurid herpesvirus 1	SiHV-1		Glass catfish (<i>Kryptopterus bicirrhus</i>)

^aDesignates the type species of each genus (ICTV, 2018).

complexity and precision of the current classification are likely to rapidly evolve in the future based on the description of new alloherpesvirus members. Finally, many fish species have experienced integration of alloherpesvirus sequences fused with transposons into their genomes.

Virion Structure

In terms of virion structure, alloherpesviruses are no exception to the remarkably conserved and morphologically distinct architecture found throughout the *Herpesvirales* order. It was demonstrated, for example, that the nucleocapsid of IchHV-1 was strikingly similar to that of human herpesvirus 1 (HHV-1 also known as herpesvirus simplex 1). The structure of herpesviruses is, from the center outwards, as follows: (i) a densely packed genome consisting of a single copy of linear dsDNA encapsulated in an icosahedral nucleocapsid ($T = 16$) with a diameter of 100–130 nm; (ii) an amorphous proteinaceous tegument layer; and (iii) a lipid bilayer envelope acquired from the host and bearing various viral glycoproteins.

Genome

Until now, 11 complete alloherpesvirus genomes have been sequenced and partial genome sequences are available from 11 more viruses. The full-length genomes of SiHV-1 (potential ictalurivirus) and BfHV-1 (potential batrachovirus) were released recently.

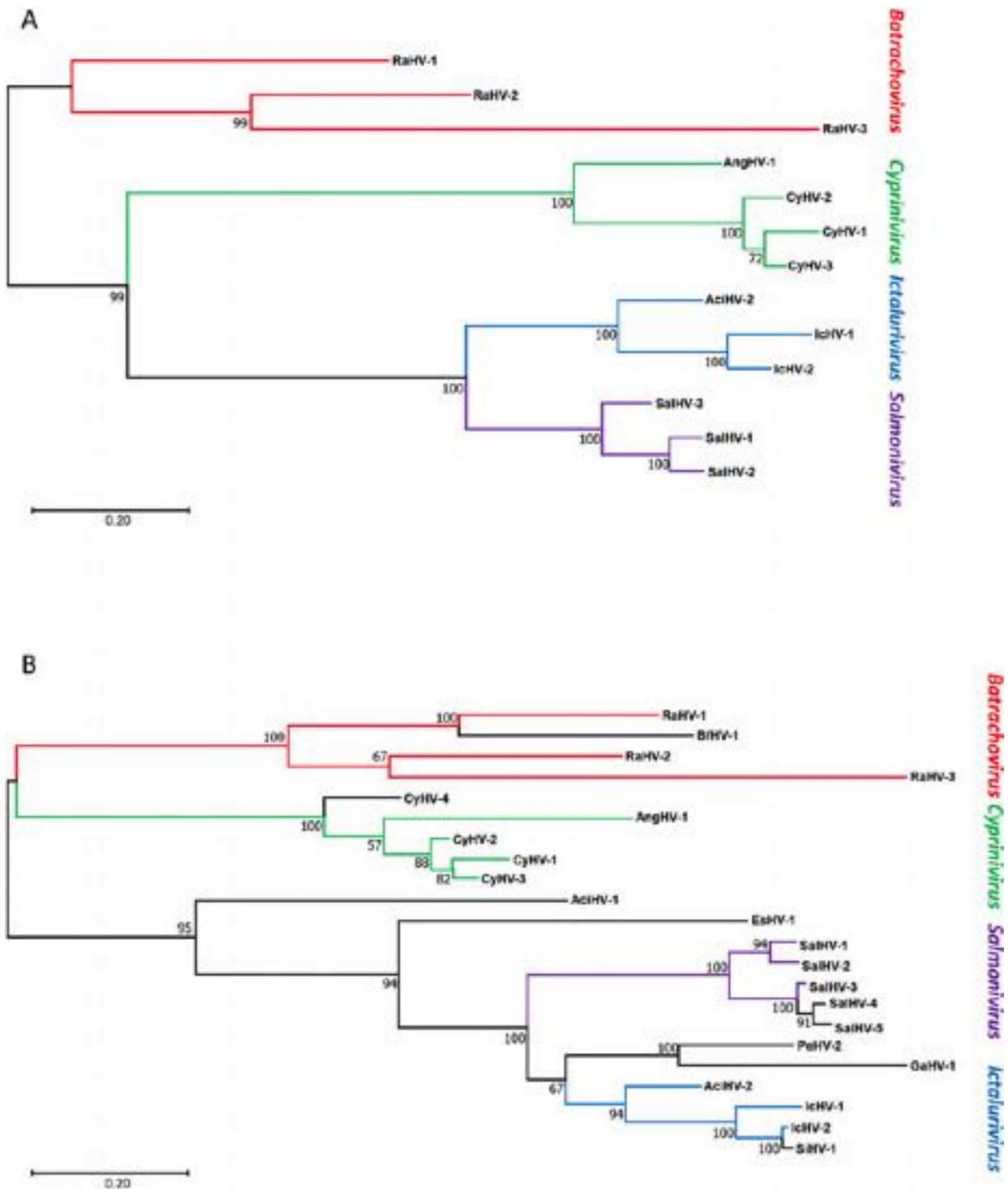


Fig. 1 Phylogenetic tree of the family *Alloherpesviridae*. The analyses were based on the Bayesian analysis (WAG amino acid model) of the deduced amino acid sequences of DNA polymerase genes (134 and 113 amino acid residues, respectively for panel A and B). High statistical values confirm the topology of the trees. The four main genera within the family *Alloherpesviridae* are designated by different colored lines on the trees. Panel A: phylogenetic tree of the classified viral species in the *Alloherpesviridae* family. Panel B: phylogenetic tree of the classified and unclassified potential members of the *Alloherpesviridae* family. AciHV: acipenserid herpesvirus; AngHV: anguillid herpesvirus; BfHV: bufonid herpesvirus; CyHV: cyprinid herpesvirus; EsHV: esocid herpesvirus; GaHV: gadid herpesvirus; IchiHV: ictalurid herpesvirus; PeHV: percid herpesvirus; RaHV: ranid herpesvirus; SalHV: salmonid herpesvirus; SiHV: silurid herpesvirus. GenBank and RefSeq accession numbers: AciHV-1: EF685903; AciHV-2: FJ815289; AngHV-1: NC_013668; BfHV-1: MF143550; CyHV-1: NC_019491; CyHV-2: NC_019495; CyHV-3: NC_009127; CyHV-4: KM357278; EsHV-1: KX198667; GaHV-1: HQ857783; IchiHV-1: NC_001493; IchiHV-2: NC_036579; PeHV-2: MG570129; RaHV-1: NC_008211; RaHV-2: NC_008210; RaHV-3: NC_034618; SalHV-1: EU349273; SalHV-2: FJ641908; SalHV-3: EU349277; SalHV-4: JX886029; SalHV-5: KP686090; SiHV-1: MH048901.

Table 2 Data on complete genome sequences of fish and amphibian alloherpesviruses

Viral name	Genome size (kbp)	Genome structure	Structure sizes (kbp)	GC%	ORFs (No.) ^a	Genbank accession number
Anguillid herpesvirus 1	248.53	TR-U-TR	U: 226 TR: 11	53.0	134	NC_013668.3 ^b
Cyprinid herpesvirus 1	291.14	TR-U-TR	U: 224 TR: 33	51.3	143	NC_019491.1 ^b
Cyprinid herpesvirus 2	290.3	TR-U-TR	U: 260 TR: 15	51.7	154	NC_019495.1 ^b
Cyprinid herpesvirus 3	295.15	TR-U-TR	U: 250 TR: 22	59.2	163	NC_009127.1 ^b
Ictalurid herpesvirus 1	134.23	TR-U-TR	U: 97 TR: 18	56.2	90	NC_001493.2 ^b
Ictalurid herpesvirus 2	142.92	TR-U-TR	U: 101 TR: 20	53.8	91	NC_036579.1 ^b
Silurid herpesvirus 1	149.34	TR-U-TR	U: 100 TR: 25	53.7	94	MH048901.1
Ranid herpesvirus 1	220.86	TR-U-TR	U: 219 TR: 0.6	54.6	132	NC_008211.1 ^b
Ranid herpesvirus 2	231.8	TR-U-TR	U: 230 TR: 1	52.8	147	NC_008210.1 ^b
Ranid herpesvirus 3	207.91	U	U: 208	41.8	186	NC_034618.1 ^b
Bufonid herpesvirus 1	158.25	U	U: 158	40.6	152	MF143550.1

^aPredicted to encode functional proteins. Includes ORFs duplicated in TR.^bDesignates Reference Sequences (RefSeq) in the GenBank.

Note: U: Unique region; TR: Terminal repeat; kbp: kilobase pair.

The genome size of alloherpesviruses range between 134 and 295 kilobase pairs (kbp) containing 90–186 Open Reading Frames (ORFs). There are significant differences in the genome size of viruses belonging to different genera. The (potential) ictalviruses have the smallest genomes ranging from 134 to 149 kbp, while the estimated genome size of salmoniviruses is around 170 kbp. The genomes of the (potential) batrachoviruses are comprised within the range 158–232 kbp, and the largest genomes are presented by cypriniviruses ranging from 245 to 295 kbp (Table 2). The genome of CyHV-3 is the largest of all herpesviruses sequenced so far. The structure of alloherpesvirus genomes is simpler than that of most of the *Herpesviridae*. Most of their genomes contain only one unique region which is flanked by direct terminal repeats (TR). On the contrary, several of the members of the *Herpesviridae* family are more complex and composed of one long region associated with one or sometimes two short unique regions flanked by internal and terminal repeats. Fish alloherpesviruses have long TR, while frog herpesviruses have very short or even no detectable TR (Table 2). Alloherpesviruses differ from the members of the *Herpesviridae* family in the number of conserved genes. The latter have a subset of about 40 convincingly conserved genes in members of the family *Herpesviridae*. While alloherpesvirus genomes share 40–60 homologous genes within a genus, there are only 12 genes which have homologs in all completely sequenced alloherpesvirus genomes. These encode major structural proteins and enzymes involved in replication and packaging of viral DNA into virions.

Life Cycle

Infection begins with viral entry, where surface glycoproteins attach to cellular receptors, leading to cell penetration. Two non-exclusive mechanisms of viral penetration have been described in the *Herpesviridae* family, i.e., direct fusion with the plasma membrane, or endocytosis followed by fusion of the viral envelope with the endosome membrane, both resulting in the release of the nucleocapsid, coated by tegument proteins, into the cytoplasm (Fig. 2). Nevertheless, the entry mechanisms of alloherpesviruses are still poorly known. Retrograde transport ensues along microtubules to the nuclear pores, where the viral linear genome is delivered into the nucleus. After circularization of the genome, cellular transcription machinery is used for expression of viral genes as a cascade, starting with immediate early genes, followed by early and late categories. As a result of early gene expression, replication of virus DNA is mediated by viral DNA polymerase, leading to the synthesis of branching concatemers. Finally, late gene expression results in the generation of structural proteins that assemble in the nucleus as capsids. Concatemers are then cleaved before the viral genomes are loaded into capsids. Nuclear egress begins with primary envelopment whereby budding at the inner nuclear membrane occurs, with release into the perinuclear space. From there, nucleocapsids exit the nucleus by fusion of the primary envelope with the outer nuclear membrane (de-envelopment). Naked capsids then get to sites of virion assembly and bud into vesicles belonging to the *trans*-Golgi network, where tegument addition and secondary envelopment with acquisition of the glycoprotein-studded envelope arise. Lastly, enveloped virions undergo egress via exocytosis.

Increasing evidence points to alloherpesviruses displaying the capacity, as found in the *Herpesviridae* family, to establish latent infections. The balance between productive and latent infection is intricately linked to host biology, as latency may be interrupted by reactivation that reverts to a productive infection. In alloherpesviruses, water temperature could be a key factor in switching between productive and latent infections.

Epidemiology

Fish Alloherpesviruses

Alloherpesviruses share the general tendency of herpesviruses to possess a narrow host range, although limited infection can sometimes occur in non-natural host species. For example, CyHV-1 is usually restricted to common carp but was detected in

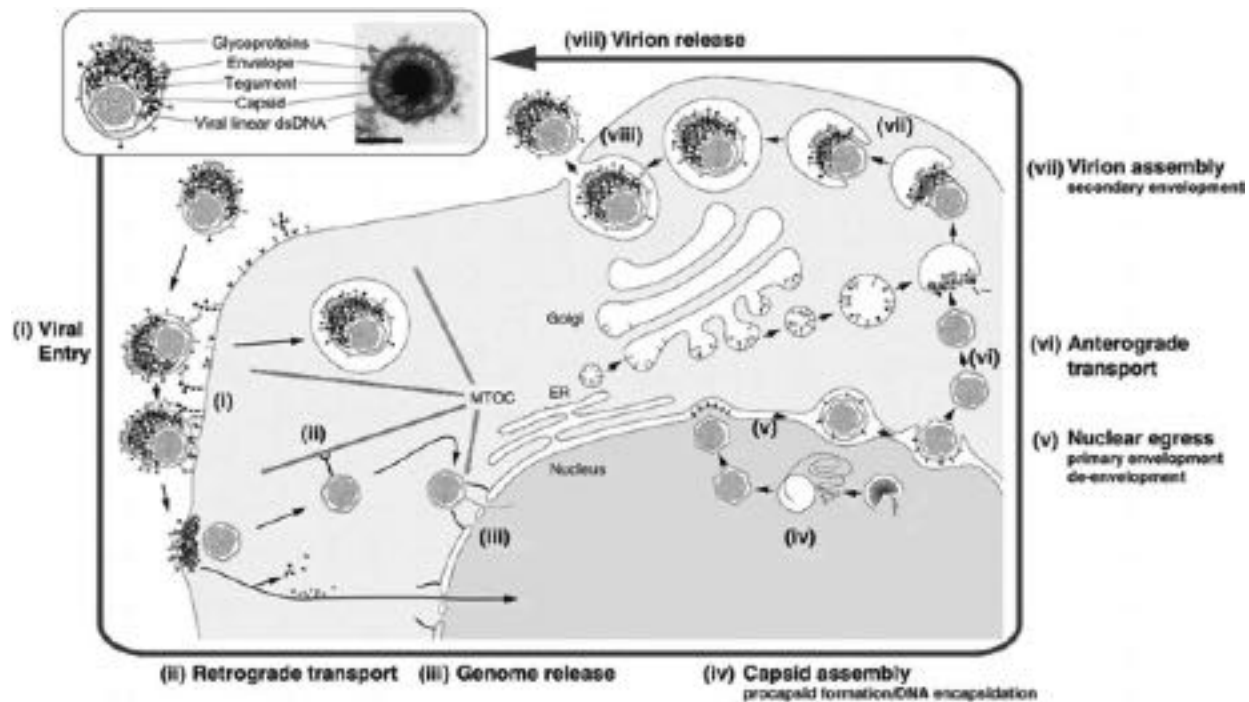


Fig. 2 Schematic representation of herpesvirus productive life cycle. Schematic representation and electron microscopy examination of CyHV-3 virion (upper left panel, bar represents 100 nm). Typical replication cycle of herpesviruses. Diagrammatic representation of the herpesvirus replication cycle, including virus entry and dissociation of the tegument, transport of incoming capsids to the nuclear pore, and release of viral DNA into the nucleus. Capsid assembly, DNA packaging, primary, secondary envelopment and egress are also illustrated. MTOC: Microtubule Organizing Center; ER: Endoplasmic Reticulum. Modified from Zeev-Ben-Mordehai, T., *et al.*, 2014. *Current Opinion in Virology* 5 (100), 42–49. Electron microscopy picture adapted with permission from Mettenleiter, T.C., *et al.*, 2009. *Virus Research* 143 (2), 222–234 Copyright Elsevier.

golden ide (*Leuciscus idus*) imported from Germany into North America. For CyHV-3, multiple vertebrate and invertebrate species have been identified as bearing viral DNA and being a potential epidemiological risk factor of viral spreading.

The geographical distribution of alloherpesviruses is variable: for instance, whereas members of the *Cyprinivirus* genus are found worldwide, others have only been associated with restricted geographical areas and contexts. Such is the case for IchV-2 that caused mass mortalities in black bullhead catfish farms in Italy in the late 1990s. In the case of CyHV-3, the first outbreaks in 1997 in Germany and its rapid worldwide distribution were attributed to international trade of common and koi carp and notably the methods used in koi carp competitions before awareness of this pathogen was commonplace. It can therefore be stated that the epidemiology of some alloherpesviruses is tightly interwoven with the precise use and trade of the domesticated host species.

Deadly outbreaks of CyHV-3 have occurred in many natural habitats such as in some North American lakes, but have not caused any long-term decrease of common carp populations. Recently, the high virulence of CyHV-3 associated with its believed absence in Australia has led to the hypothesis of using this virus to control common carp populations which are considered as an invasive species in this country. However, the predicted safety and efficacy of this method have recently been under important scientific debate.

Amphibian Alloherpesviruses

The real causes of alloherpesvirus emergence and spread in free-ranging amphibians are mostly unknown. However, the impact of human activities on their habitat and the climate may play a relevant role. Of the known batrachoviruses, RaHV-1 has been detected in free-ranging leopard frogs (*Lithobates pipiens*) in specific areas of the North American continent where, to the best of our knowledge, it has remained confined. The prevalence of RaHV-1 associated tumors ranged between 1% and 9% according to the collection site. A drastic reduction of the infected populations of leopard frogs was recorded in the 70's. Currently, no information is available concerning possible reemergence of the virus. Differently, RaHV-3 and BfHV-1 have been recorded in Europe only, in the common frog (*Rana temporaria*) and common toad (*Bufo bufo*), respectively. Their actual and current significance in the context of amphibian disease ecology is currently unknown.

Pathogenesis

Fish Alloherpesviruses

Fish alloherpesviruses are characterized by their temperature dependency, a consequence of the ectothermic nature of their hosts. The optimal temperature for viral replication and disease is variable among alloherpesviruses. Whereas AngHV-1 infection is

promoted by high temperatures (25°C), SalHV-3 outbreaks are reported when temperatures are low (8–10°C). In natural conditions, if the environmental water temperature is suboptimal for virus replication, the induced disease will be less severe or even asymptomatic, explaining the seasonal occurrence of alloherpesvirus diseases. Interestingly, expression of behavioral fever was described in common carp infected by CyHV-3.

Studies of CyHV-3 pathogenesis mimicking natural conditions demonstrated that the skin is the major portal of entry of the virus after inoculation of carp by immersion, whereas the virus would penetrate through the pharyngeal periodontal mucosa following oral contamination. Viral spreading within the infected host is probably promoted by infected white blood cells allowing rapid spread in almost all tissues soon after inoculation.

Horizontal transmission of fish alloherpesviruses can occur either directly between fish or indirectly, through several potential vectors (e.g., water, fish droppings). Salmoniviruses were detected in ovarian fluids, suggesting a possible vertical transmission. Vertical transmission was demonstrated for IchV-1 after detecting the virus in offspring from IchV-1 positive broodstocks.

The hallmark of herpesviruses is their ability to establish latency in their host after initial productive infection. Some evidence of latency has been described for CyHV-1, CyHV-3, AngHV-1, SalHV-2 and IchV-1 fish alloherpesviruses. The latency of alloherpesviruses in fish surviving the primary infection allows viral transmission to naïve fish upon disease reactivation. This aspect of the virus biology results in a certain spatial and temporal flexibility for viral maintenance in the host population.

Amphibian Alloherpesviruses

Pathogenesis of the *Batrachovirus* associated disease is virtually unknown. RaHV-1 has been shown to be the etiologic agent of Lucké's adenocarcinoma in leopard frogs. However, it is not known how the virus induces tumor development although environmental temperature may play a role. The disease could be reproduced by only infecting frogs during their larval stage with tumor extract. Viral replication appears to be dependent on low environmental temperatures, whereas tumor growth and metastasis would be promoted at higher temperatures. Regarding temperature and seasonality, the proliferative dermatitis associated with RaHV-3 and BfHV-1 are observed right after hibernation, during the mating season, between the end of winter and beginning of spring, when the environmental temperature is low. No virus-associated lesion has been reported during the later spring and summer. The molecular bases of the associated diseases are not known and Koch's postulates have not been fulfilled for these two viruses.

Clinical Features

Fish Alloherpesviruses

Clinical aspects induced by fish alloherpesviruses can vary substantially among outbreaks or individuals. Indeed, variations are observed in the severity of clinical signs according to diverse factors including, but not restricted to, those related to the environment (temperature, density of the population) and host (age of the individuals, genetics, immune status). Primary infection involves productive viral replication that may result in viremia and acute disease associated with high morbidity and mortality. In general, fish exhibit non-specific behavioral changes including lethargy, hyperexcitability, erratic swimming or gasping at the water surface. Diseases associated with infection of naïve individuals by alloherpesviruses can include epithelial lesions, diffuse multisystemic hemorrhages, ascites, kidney and liver necrosis (Fig. 3(A), Table 3).

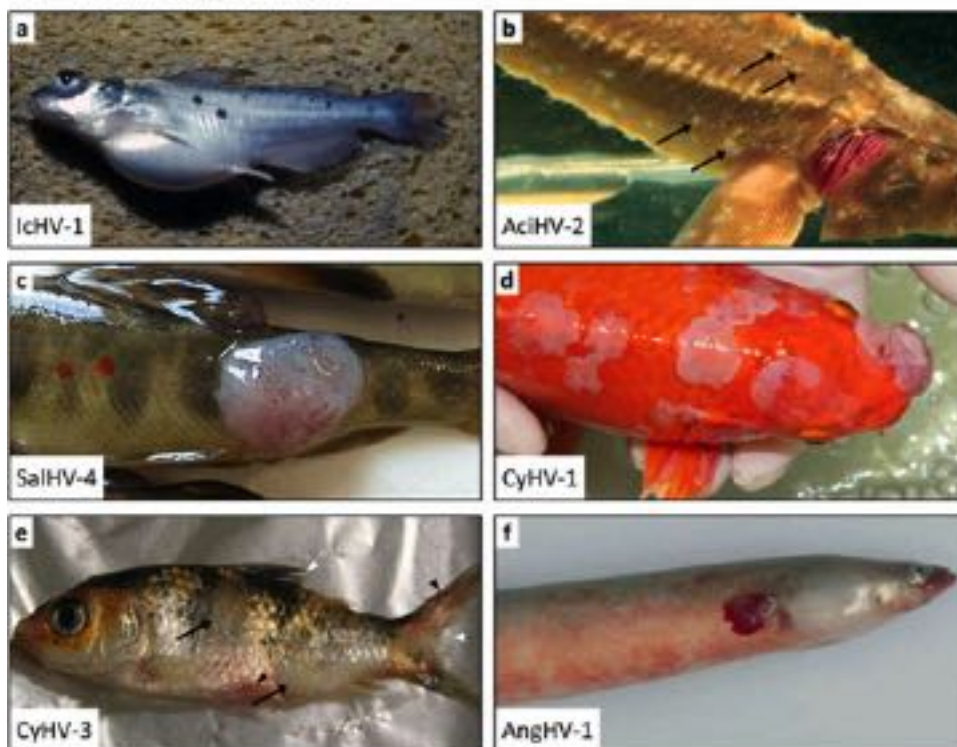
Contrarily to their *Herpesviridae* family counterparts, alloherpesviruses usually induce severe diseases in their hosts. This may result from a milder degree of coevolution with their respective hosts. Alloherpesvirus infections might nevertheless be artificially enhanced by aquaculture methods. Among the most important fish alloherpesviruses, CyHV-2, CyHV-3 and IchV-1 cause a systemic and lethal disease (Fig. 3(A); Table 3) characterized by mass outbreaks associated with high mortalities. During the summer, when temperature is optimal for the occurrence of IchV-1 disease, 90% of fry and fingerlings may die in less than two weeks. During CyHV-3 infection, infected individuals may begin to die at 6–8 days post infection. Mortality rates are in general up to 70%–80% of infected individuals but they can reach 100% in some cases. Mass mortalities within 24 h were also recorded in CyHV-2 infected gibel carp. CyHV-1, IchV-2, AciHV-2, SalHV-2 and SalHV-3 primary infections can also induce lethal diseases.

Development of a benign epithelial hyperplasia, also called herpetic lesion or papilloma is another clinical pattern largely observed in alloherpesvirus infections. This clinical feature can be observed during primary productive infection with AciHV-2, SalHV-3 and SalHV-4 (Fig. 3(B)). In addition, this lesion can arise during recrudescence of a previous infection. Indeed, while CyHV-1 and SalHV-2 primary infection is typically characterized by an acute systemic disease, the recrudescence of the disease in survivors is associated with epithelial tumors. Papillomas associated with CyHV-1 infection appear 5–6 months after primary exposure in most survivors and are more commonly observed on the skin (body, fins, mouth) and at the site of intraperitoneal inoculation. Similarly, epitheliomas in SalHV-2 disease are observed around the mouth, at the caudal fin, operculum, and the body surface. These lesions may persist for one year. Experimental challenges show that tumor occurrence varies between species (12%, 40%–60%, and 100% in rainbow trout, chum salmon and masu salmon, respectively).

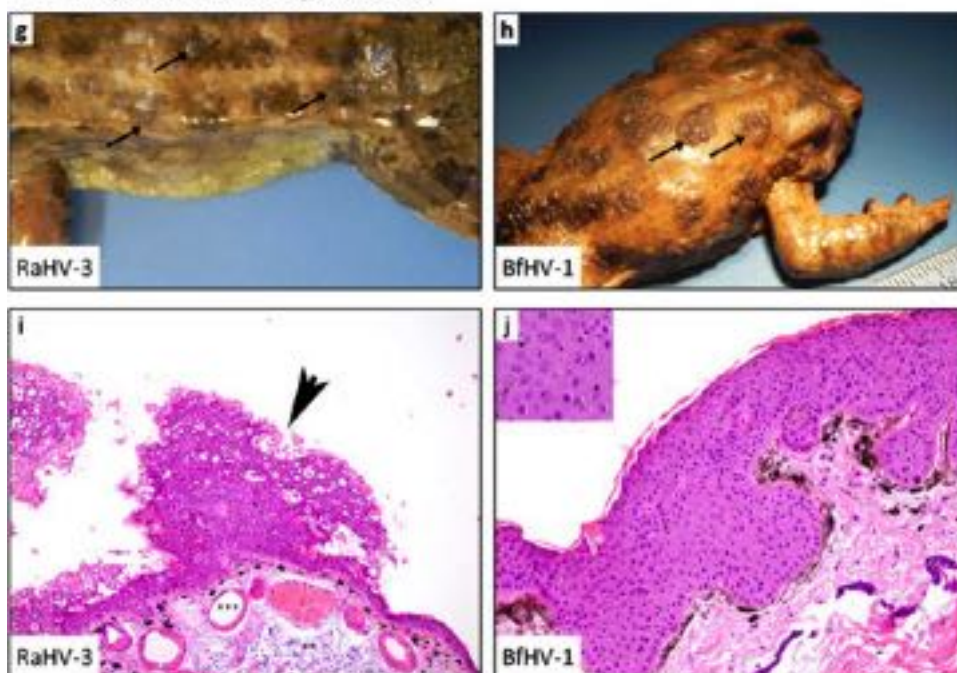
Amphibian Alloherpesviruses

RaHV-1 causes adenocarcinomas in leopard frogs (Lucké's adenocarcinoma). The tumor can grow extensively and metastasize, involving and invading other organs. Papillary projections are often observed in the neoplastic mass. Differently, the only reported

A – Fish alloherpesviruses



B – Amphibian alloherpesviruses



lesion associated with RaHV-2 is the occurrence of severe edema in frogs infected at larval stages. RaHV-3-associated disease is characterized by multifocal to coalescent gray skin patches scattered throughout the integument of affected frogs, which correspond to areas of prominent epidermal hyperplasia during microscopic examination (Fig. 3(B)). Large numbers of intranuclear eosinophilic inclusions are seen in the thickened epidermis. Toads infected with BfHV-1 show similar but distinct skin lesions. Relatively flat to slightly elevated, dark brown, cauliflower-like skin proliferations are observed grossly. Histologically, the lesions correspond to areas of epidermal hyperplasia, frequently associated with hyperkeratosis. Intranuclear inclusions are seen in infected toads, but less numerous and prominent than in frogs infected with RaHV-3. Interestingly, minimal to no associated inflammation is observed in both infections.

Diagnosis

Diagnosis of alloherpesvirus infection can rely on various methods. First, diagnosis can be oriented by analysis of the clinical signs, anatomopathological and histopathological lesions. However, the latter are only indicative of the alloherpesvirus infection due to the absence of known pathognomonic signs and the high frequency of multipathogen co-infection. Still, hyperplastic epidermal lesions are typical and very indicative of some alloherpesvirus infections (SalHV-4, CyHV-1, RaHV-3, BfHV-1).

Specific diagnosis involves the use of viral isolation using specific cell lines, viral detection using molecular methods or serological methods. Viral isolation of most alloherpesviruses has been achieved using cell lines. Nevertheless, viral isolation is still restricted by the absence of permissive cell lines: for example, among batrachoviruses, only RaHV-2 has been successfully isolated in cell culture. Specific molecular methods are now available for detection of most alloherpesviruses (PCR, qPCR, monoclonal and polyclonal antibodies). These methods can be tentatively used to diagnose productive, but also potentially latent infection (PCR, qPCR). Regarding the latter, the specific sites of alloherpesvirus latency are frequently unknown. Viral DNA can be detected in the brain and B lymphocytes of fish previously exposed to CyHV-3. Note that the sensitivity of the molecular tools developed could be decreased by the limited knowledge about genetic coalescence within each alloherpesvirus species. Detection of the antibody response against alloherpesviruses (ELISA, seroneutralization) can help to diagnose a previous (latent) infection (e.g., screening of broodstocks). Nevertheless, interpretation should be cautious, as specific antibody half-life is rather short in fish. Specific antibodies against CyHV-3 can barely be detected from sera of fish 280 days post infection. Finally, on-site diagnostic methods including lateral flow devices and loop-mediated isothermal amplification have been developed for some alloherpesviruses.

Management and Treatment

Alloherpesviruses are not equally threatening the aquaculture sector. Some alloherpesvirus infections, like those induced by SalHV-1, were found to have limited impact on aquaculture. Other alloherpesviruses (e.g., IchV-1), can be controlled by management procedures (screening of broodstocks, disinfection of facility and incoming water, microbiological independence of produced fish batches), improving rearing conditions (fish density, water quality, stress reduction) or by regulating the water temperature.

As explained above, alloherpesvirus replication and disease expression correlates strongly with environmental temperature. This peculiarity has been exploited for the treatment of alloherpesvirus infection by changing environmental temperature to levels suboptimal or non-permissive for viral replication. For example, fish infected by CyHV-3 can be transferred to non-permissive

Fig. 3 Illustration of typical clinical features observed during some alloherpesvirus infection. Panel A. Illustration of clinical features induced by fish alloherpesviruses. (a) channel catfish (*Ictalurus punctatus*) infected by IchV-1, typical exophthalmia and distended abdomen caused by ascite. (b) Multiple plaques of epidermal proliferation (arrows) in a fingerling of Siberian sturgeon (*Acipenser baeri*) infected AciHV-2. (c) Severe papillomas caused by SalHV-4 in wild young Atlantic salmon (*Salmo salar*) captured at 10°C. (d) Proliferative cutaneous lesions in adult koi carp (*Cyprinus carpio*) infected by CyHV-1, multifocal to coalescing, slightly raised to nodular, white cutaneous proliferations along the dorsum, mouth, and pectoral fins. (e) Koi carp (*Cyprinus carpio*) infected by CyHV-3, extensive circular necrosis of the skin covering the body (black arrows), fin erosion (white arrowheads), and hemorrhages on the skin and fins (black arrowheads). (f) Diseased cultured European eel (*Anguilla anguilla*) with AngHV-1 infection. The eel shows a patchy pattern of hemorrhages in the skin, and severe hemorrhagic fins. Panel B. Illustration of clinical features induced by amphibian alloherpesviruses. (g) Multifocal to coalescent light tan to gray patches of skin are present, more frequently on the dorsum and along the flanks of a common frog (*Rana temporaria*) (black arrows). (h) Bufonid herpesvirus 1 infected common toad (*Bufo bufo*) associated gross skin lesions. Multifocal raised, dark brown, patchy areas of skin are scattered over the dorsum of the toad (black arrows). (i) Skin, Common frog (*Rana temporaria*) naturally infected with Ranid herpesvirus 3. The epidermis is focally, severely elevated (hyperplastic), forming a papillary projection (arrowhead) characterized by several clear spaces (epidermal vacuoles) partially filled with degenerating and necrotic keratinocytes. The underlying dermis shows loosely arranged collagen fibers (edema-two asterisks) and variably ectatic mucous glands (three asterisks). (j) Skin, Common toad (*Bufo bufo*) naturally infected with Bufonid herpesvirus 1. The epidermis is diffusely thickened (hyperplasia) up to 5 times as normal. Among the several nuclei are a number of intranuclear eosinophilic inclusions (inset, asterisks). The underlying dermis is characterized by loose collagen fibers (edema). (a) Courtesy of J. A. Plumb. (b) Adapted with permission from Shchelkunov, *et al.*, 2009. Diseases of Aquatic Organisms 86 (3), 193–203. Copyright Inter-Research. (c) Adapted with permission from Doszpoly, *et al.*, 2013. Diseases of Aquatic Organisms 107 (2), 121–127. Copyright Inter-Research. (d) Adapted with permission from Crossland, *et al.*, 2018. Journal of Aquatic Animal Health 30 (3), 185–190. Copyright John Wiley and Sons. (e) Adapted from Michel *et al.*, 2010. Emerging Infectious Diseases 16 (12), 1835–1843. (f) Adapted from Haenen, O.L.M., *et al.*, 2002. Bulletin of the European Association of Fish Pathologists 22 (4), 247–257. (g, i, j) Courtesy of F.C. Origgi. (h) Adapted from Origgi, *et al.*, 2018. Scientific Reports 8 (1), 14737.

Table 3 Clinical features caused by some alloherpesviruses

<i>Viral name</i>	<i>Susceptible host</i>	<i>Clinical signs (non-exhaustive)</i>	<i>Mortality</i>
Batrachovirus			
Ranid herpesvirus 1	Northern leopard frog (<i>Lithobates pipiens</i>)	Renal adenocarcinoma	Unknown
Ranid herpesvirus 2	Northern leopard frog (<i>Lithobates pipiens</i>)	Unclear, severe edema when infected at larval stage	Unknown
Ranid herpesvirus 3	Common frog (<i>Rana temporaria</i>)	Proliferative dermatitis	Unknown
Cyprinivirus			
Cyprinid herpesvirus 1	Common carp (<i>Cyprinus carpio</i>)	Systemic disease in fry, hyperplastic cutaneous proliferation (papillomas, pox-like lesions) in surviving/older fish	Up to 87.5% (fry)
Cyprinid herpesvirus 2	Goldfish (<i>Carassius auratus</i>), gibel carp (<i>Carassius gibelio</i>)	Lethargy, anorexia, pale patches and hypersecretion of mucus in the gills, hemorrhages on the body and gills	Up to 100% (all ages)
Cyprinid herpesvirus 3	Common carp (<i>Cyprinus carpio</i>)	Lethargy, anorexia, hypersecretion of mucus on the skin, hyperemia of the fins, skin fin & gill necrosis, fin erosion, hemorrhages on the skin, neurological symptoms	Up to 100% (all ages)
Anguillid herpesvirus 1	Japanese and European eel (<i>Anguilla japonica</i> and <i>A. anguilla</i>)	Decreased growth rates, hemorrhages and ulcerative lesions on the skin (head, beak and caudal abdominal surface) fin and gills	Up to 10%
Ictalurivirus			
Ictalurid herpesvirus 1	Channel catfish (<i>Ictalurus punctatus</i>), striped catfish (<i>Pangasius hypophthalmus</i>)	Reduced growth rates, erratic swimming, distended abdomen, hemorrhages, exophthalmia	>90% (fry and fingerling channel catfish), 30%–40% (striped catfish)
Ictalurid herpesvirus 2	Black bullhead catfish (<i>Ameiurus melas</i>)	Erratic swimming, hemorrhages	High mortality (fry and juveniles)
Acipenserid herpesvirus 2	Sturgeon species (<i>Acipenser</i> spp.)	Erratic swimming, lethargy, hyperplastic skin lesions, hyperemia of the ventral scutes, mouth and anus	Up to 80% (juvenile white sturgeon), up to 100% (juvenile Siberian sturgeon)
Salmonivirus			
Salmonid herpesvirus 1	Rainbow trout (<i>Oncorhynchus mykiss</i>)	Darkened pigmentation, exophthalmia, abdominal distention, haemorrhages in the fins, pale gills	Low
Salmonid herpesvirus 2	Salmon and trout (<i>Oncorhynchus</i> spp.)	Lethargy, anorexia, darkened appearance, skin ulcers, focal epithelial proliferation on the mouth and body, exophthalmia Epithelial tumors (papilloma), preferably on the mouth, in surviving fish	Up to 100%. High losses in juveniles
Salmonid herpesvirus 3	Lake trout (<i>Salvelinus namaycush</i>)	Lethargy, swim near the surface, behavioral modification, mucoid patches on the skin, hemorrhages on eyes and jaw	Up to 100%. High losses in juveniles
Unclassified			
Acipenserid herpesvirus 1	White sturgeon (<i>Acipenser transmontanus</i>)	Few external clinical signs associated with a diffuse hyperplastic dermatitis	Up to 35% (juvenile white sturgeons)
Bufoid herpesvirus 1	Common toad (<i>Bufo bufo</i>)	Proliferative dermatitis	Unknown
Esocid herpesvirus 1	Northern pike (<i>Esox lucius</i>)	Whitish circular plaques on the fins and dorsal parts	NA
Salmonid herpesvirus 4	Atlantic salmon (<i>Salmo salar</i>)	White plaques of epithelial tumors (papilloma) on skin, or caudal peduncle	>50% when secondary infections involved
Salmonid herpesvirus 5	Lake trout (<i>Salvelinus namaycush</i>)	Abnormal swimming, discolored musculature	

water temperatures (>30°C), allowing fish to clear viral infection and develop an adaptive immune response. This method was also used as a natural immunization method following iatrogenic CyHV-3 contamination. Unfortunately, fish infected with these methods may become latently infected by CyHV-3. Interestingly, fish infected by CyHV-3 have been shown to naturally seek for warmer environments (behavioral fever), thereby surviving viral infection; this feature could be exploited in aquaculture. The specific temperature range allowing alloherpesvirus replication and disease is variable among viral species. For instance, eel farms experiencing AngHV-1 infection usually decrease environmental temperature to decrease clinical signs and mortality.

Antiviral treatments were tested against alloherpesviruses, but to the best of our knowledge, no commercial treatment is currently available. Extracts of the *Clinacanthus nutans* (Burm. f.) Lindau medicinal plant used as a prophylactic or therapeutic treatment against CyHV-3 infection showed positive effects on common carp survival. Nucleoside analogues such as acyclovir are able to limit IchV-1, CyHV-3 or SalHV-2 replication in cell culture, and to be effective against SalHV-2 *in vivo*.

Prevention

Facing the lack of efficient therapeutic methods, safe and efficacious vaccines represent the most promising control approach against alloherpesviruses. Vaccination of fish should involve delivery methods adapted to mass vaccination of the target species, minimizing the cost for vaccination and maximizing the duration of immunity with an early in life and limited number of vaccine administrations.

Multiple vaccine candidates against alloherpesviruses have been studied but very few are commercially available nowadays. Inactivated vaccines consist of the entire killed viral particle, thus containing all structural viral antigens. Inactivated vaccine candidates have been described for SalHV-2, CyHV-2 and CyHV-3. Injection of inactivated CyHV-2 vaccines to gibel carp induced partial protection. A formalin-inactivated CyHV-3 vaccine, trapped within a liposomal compartment and used for oral administration, induced partial protection. The high amount of viral antigen required, the vaccine delivery method, and a generally low efficacy are the main concerns and restrictions for inactivated vaccine development.

DNA vaccines consist of genetically engineered DNA plasmids encoding specific antigens from pathogens. DNA vaccines were tested against ICHV-1 and CyHV-3. Injectable DNA vaccines consisting of plasmids encoding CyHV-3 ORF25 envelope glycoprotein were shown to be partially efficacious against immersion challenge, but not against a cohabitation challenge. Attempts to deliver the DNA plasmids encoding ORF25 orally failed to induce an efficient protection. DNA vaccines against ICHV-1 showed equivocal results, even when cocktails of plasmid DNA encoding multiple ORFs were used. Studies on DNA vaccines against alloherpesviruses highlighted two main challenges for future research, namely selecting the best antigens for immunization and improving vaccine delivery methods.

Attenuated vaccines consist of live viral strains with a virulence that has been reduced or suppressed by serial passages in cell culture (conventional attenuated vaccine) or by targeted viral genome editing (recombinant attenuated vaccine). Attenuated vaccines have the advantages of being compatible with mass vaccination and simulating a natural viral infection of the host. However, they raise safety concerns including residual virulence, reversion to virulence, and spread from vaccinated to naïve subjects. Conventional and recombinant attenuated vaccines were developed against ICHV-1 and CyHV-3. A conventional anti-CyHV-3 attenuated vaccine has been developed by serial passages in cell culture and UV irradiation. This vaccine is commercialized in Israel for the vaccination of koi and common carp. The determinism of the attenuation is unknown, and consequently, reversions to a pathogenic phenotype cannot be ruled out. The attenuated strain has residual virulence for fish weighing less than 50 g.

Due to scientific advances in molecular biology and molecular virology, the development of attenuated vaccines is evolving from empirical to rational design. A viral genome can be edited to delete genes encoding virulence factors in such a way that reversion to virulence cannot occur. This approach has been tested for CyHV-3 by targeting different genes such as the thymidine kinase (TK) and deoxyuridine triphosphatase. The previous cloning of ICHV-1 and CyHV-3 as infectious bacterial artificial chromosomes (BAC) allows viral mutagenesis. Recently, a vaccine candidate based on the double deletion of ORF56 and ORF57 was produced using BAC mutagenesis. This strain is safe for vaccinated fish, is limitedly transmitted to naïve fish and induces full protection against a lethal cohabitation challenge. ORF57 was found to be the crucial virulence factor and is conserved in the genus *Cyprinivirus*. A recombinant attenuated vaccine candidate was also developed for ICHV-1 by deletion of the TK gene and was shown to be effective. Attenuated recombinant vaccines, associated with the tools provided by BAC mutagenesis, allow efficient development of vector vaccine strategies. Fish immunized with an ICHV-1 TK deleted recombinant attenuated strain expressing a reporter gene developed an antibody response against the transgene product.

Further Reading

- Adamek, M., Steinhagen, D., Imazarow, I., *et al.*, 2014. Biology and host response to Cyprinid herpesvirus 3 infection in common carp. *Developmental and Comparative Immunology* 43, 151–159.
- Boutier, M., Ronsmans, M., Rakus, K., *et al.*, 2015. Cyprinid herpesvirus 3: An archetype of fish alloherpesviruses. In: Margaret Kielian, K.M., Thomas, C.M. (Eds.), *Advances in Virus Research*. Academic Press, pp. 161–256.
- Costes, B., Raj, V.S., Michel, B., *et al.*, 2009. The major portal of entry of koi herpesvirus in *Cyprinus carpio* is the skin. *Journal of Virology* 83, 2819–2830.
- Hanson, L., Dishon, A., Kotler, M., 2011. Herpesviruses that infect fish. *Viruses* 3, 2160–2191.
- Hanson, L., Doszpoly, A., van Beurden, S.J., de Oliveira Viadanna, P.H., Waltzek, T., 2016. Chapter 9 – Alloherpesviruses of fish. In: Kibenge, F.S.B., Godoy, M.G. (Eds.), *Aquaculture Virology*. San Diego: Academic Press, pp. 153–172.
- Ilouze, M., Dishon, A., Kotler, M., 2006. Characterization of a novel virus causing a lethal disease in carp and koi. *Microbiology and Molecular Biology Reviews* 70, 147–156.
- Marshall, J., Davison, A.J., Kopf, R.K., *et al.*, 2018. Biocontrol of invasive carp: Risks abound. *Science* 359, 877.
- McColl, K.A., Cooke, B.D., Sunarto, A., 2014. Viral biocontrol of invasive vertebrates: Lessons from the past applied to cyprinid herpesvirus-3 and carp (*Cyprinus carpio*) control in Australia. *Biological Control* 72, 109–117.
- Origgi, F.C., Schmidt, B.R., Lohmann, P., *et al.*, 2018. Bufonid herpesvirus 1 (BfHV1) associated dermatitis and mortality in free ranging common toads (*Bufo bufo*) in Switzerland. *Scientific Reports* 8, 14737.
- Rakus, K., Ronsmans, M., Forlenza, M., *et al.*, 2017. Conserved fever pathways across vertebrates: A herpesvirus expressed decoy TNF- α receptor delays behavioral fever in fish. *Cell Host and Microbe* 21, 244–253.
- Van Beurden, S., Engelsma, M., 2012. Herpesviruses of fish, amphibians and invertebrates. In: Magel, G.D., Tying, S. (Eds.), *Herpesviridae – A Look into this Unique Family of Viruses*. Rijeka: InTech, pp. 217–242.

Fish Retroviruses (*Retroviridae*)

TA Paul, RN Casey, PR Bowser, and JW Casey, Cornell University, Ithaca, NY, United States
J Rovnak and SL Quackenbush, Colorado State University, Fort Collins, CO, United States

© 2021 Elsevier Ltd. All rights reserved.

This is a reproduction of T.A. Paul, R.N. Casey, P.R. Bowser, *et al.*, Fish Retroviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00670-1](https://doi.org/10.1016/B978-012374410-4.00670-1).

Glossary

Cytopathic effect Morphologic changes in cells resulting from virus replication.

Dysplasia An abnormality in the appearance of cells and indicative of a preneoplastic change.

Hyperplasia An increase in the number of cells of an organ or tissue.

Leiomyosarcoma Neoplasia of smooth muscle cells.

Malpighian cells Cells of the deepest layer of the epidermis.

Neoplasia An abnormal growth of cells.

Oncogene Cellular or viral gene whose products are capable of inducing a neoplastic phenotype.

Proto-oncogene A normal cellular gene which when inappropriately expressed or mutated becomes an oncogene.

Introduction

Retroviruses have been documented in a wide range of vertebrate species including lower vertebrates such as frogs, snakes, sharks, fish, birds, and turtles. To date, however, the majority of these reported sequences represent only partial fragments of the retrovirus. In recent years, the complete genomic sequence of exogenous fish retroviruses from Atlantic salmon (*Salmo salar*), walleye (*Sander vitreus*), and snakehead (*Ophicephalus striatus*), and an endogenous retrovirus of zebrafish (*Danio rerio*) have been determined. Analysis of these viral genomes indicates a high degree of diversity among the fish retroviruses as well as several unique features compared to mammalian and avian retroviruses.

The etiological relationship between retroviruses and cancer is well established in mammalian and avian systems. Simple retroviruses can promote cell proliferation by inducing the ectopic expression of captured cellular oncogenes (viral transduction for acutely transforming viruses) in which an oncogene recombines into the viral genome, or by integration near or within the coding sequence of cellular proto-oncogenes (nonacute transforming viruses). Transcription factors encoded by some complex retroviruses, like human T-cell leukemia virus-1 (HTLV-1), also have nonacute oncogenic potential by deregulating cellular pathways controlling cell proliferation.

Like their mammalian and avian counterparts, many fish neoplastic/proliferative diseases are suspected of having a retroviral etiology. Retroviruses are reported to be associated with 13 spontaneous proliferative diseases of fish based on the observation of retrovirus-like particles and, in some cases, reverse transcriptase activity in neoplastic lesions. Seven of these diseases with putative viral etiologies display a seasonal cycle, that is, they develop and regress annually. Retrovirus-associated tumors of the skin have received the most attention, since they are most visible and easily sampled. Skin tumors with suspected viral etiologies have been found in white sucker (*Castostomus commersoni*), walleye, yellow perch (*Perca flavescens*), and European smelt (*Osmerus eperlanus*). Retroviruses have also been linked to lymphoma in northern pike (*Esox lucius*), leukemia in chinook salmon (*Oncorhynchus tshawytscha*), and leiomyosarcoma of the swimbladder in Atlantic salmon. One noteworthy piscine retrovirus, the snakehead fish retrovirus (SnRV), has not been associated with tumor induction. In most of these systems, the etiologic relationship of retroviral infections and neoplasms rests on circumstantial evidence such as observations of viral particles by electron microscopy and the presence of reverse transcriptase activity. However, sequencing of fish retroviruses has facilitated use of molecular-based diagnostic reagents and, in some cases, shed light on potential modes of pathogenesis. In particular, the walleye retroviruses, walleye dermal sarcoma virus (WDSV) and walleye epidermal hyperplasia virus type 1 (WEHV-1) and type 2 (WEHV-2), express a subset of accessory genes exclusively during oncogenesis, specifically implicating the encoded proteins in the process of tumorigenesis.

Retroviruses and their Life Cycle

Retrovirus particles are composed of viral structural and enzymatic proteins and two copies of the positive-sense single-stranded RNA genome. The proteins are surrounded by a lipid bilayer, acquired from the host cell, in which viral envelope glycoproteins are embedded. Retroviral infection is initiated when the surface glycoprotein binds to its cognate receptor on the surface of a susceptible cell. Upon fusion of the viral and cellular membranes, the viral core enters the cytoplasm where the virally encoded enzyme, reverse transcriptase, copies the RNA genome into DNA. The viral integrase protein binds reverse transcribed DNA and mediates its integration into the genomic DNA of the host. In rare cases, retroviral infection of host germ-line cells and subsequent integration into the chromosomes can establish retroviral sequences as heritable genetic elements, known as endogenous retroviruses.

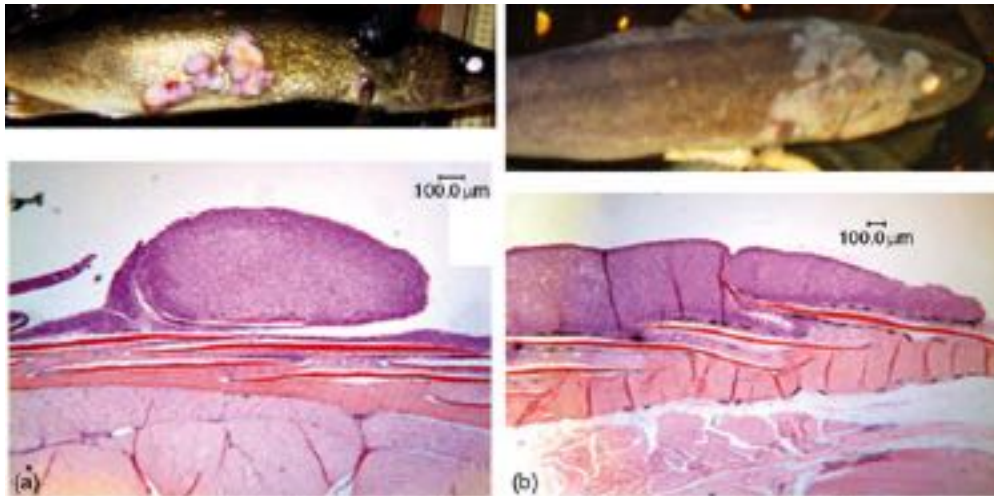


Fig. 1 (a) Gross and histological images of walleye dermal sarcoma and (b) walleye epidermal hyperplasia.

Based on the complexity of the viral genome, retroviruses can be divided into two broad categories: simple and complex. Simple retroviruses, including members of the *Gammaretroviruses* and *Alpharetroviruses*, contain three genes: *gag*, *pol*, and *env*. Expression of a full-length unspliced messenger RNA (mRNA) and a spliced mRNA serve as the templates for translation of the retroviral structural and enzymatic proteins. These proteins are first translated as polyproteins from *gag* and *pol*, respectively, and are processed by a virally encoded protease. The viral envelope glycoprotein is encoded from a singly spliced viral transcript containing the *env* gene.

Complex retroviruses are distinguished from simple retroviruses on the basis of their additional coding capacity and pattern of viral gene expression. This category of retroviruses includes *Lentiviruses* such as human immunodeficiency virus type 1 (HIV-1), *Spumaviruses*, and the *Deltaretroviruses*, human T-cell leukemia virus (HTLV-1), and bovine leukemia virus (BLV). Complex retroviruses produce multiply spliced transcripts that encode two classes of accessory/regulatory proteins. One class includes transcriptional regulatory proteins such as HTLV-1 Tax and HIV-1 Tat, which act *in trans* to directly regulate the activity of the viral promoter. A second class, including HTLV-1 Rex and HIV-1 Rev, act post transcriptionally, to facilitate transport of unspliced and singly spliced viral transcripts to the cytoplasm. The combined action of these proteins divides the replication cycle into two temporal phases: an early, regulatory phase, and a later, structural phase. During the regulatory phase, low levels of fully spliced transcripts encoding *trans*-activators increase levels of transcription leading to the accumulation of post-transcriptional regulatory proteins. These proteins allow production of unspliced and singly spliced mRNAs that encode the viral structural proteins.

Both simple and complex retroviruses have been identified in fish species. Retroviruses isolated from walleye and snakehead fish resemble complex retroviruses based upon their genome structure and transcriptional profile, while simple retroviruses have been identified in Atlantic salmon and zebrafish.

Skin Tumors in Walleye and Their Associated Retroviruses

Walleye dermal sarcoma (WDS) and walleye discrete epidermal hyperplasia (WEH) are neoplastic and hyperplastic skin lesions in walleye that have been etiologically associated with infection by three distinct retroviruses (Fig. 1). WDS and WEH were first reported in 1969 on fish from Oneida Lake in New York. Since then, WDS and WEH have been reported on walleyes throughout North America. These diseases have a seasonal cycle; tumor incidence is highest in the late fall and early spring months at frequencies of 27% and 5% for WDS and WEH, respectively. Lower tumor prevalence in summer months is associated with regression on individual adult fish.

WDS appears as a cutaneous mesenchymal neoplasm arising multicentrically within the superficial dermis and overlaid with epidermis. Fall tumors appear highly vascularized due to a network of capillaries, while spring tumors are frequently white with ulcerated surfaces. Invasive tumors are rarely observed in feral fish. WEH lesions are benign mucoid-like plaques with distinct boundaries. They can range from 2 to 3 mm in diameter to large lesions with irregular borders that may be as large as 50 mm across. Histologically, this disease appears as an epidermal proliferation consisting primarily of Malpighian cells.

Experimental transmission of WDS and WEH has been achieved with cell-free filtrates. Interestingly, only cell-free filtrates prepared from the spring, regressing WDS, not fall, developing WDS, are able to transmit disease. Infection of walleye fingerlings less than 9 weeks of age results in the frequent development of invasive tumors.

Electron microscopic observation of retrovirus particles and cell-free transmission of both diseases suggest a retroviral etiology for WDS and WEH. WDSV was isolated from WDS tissue, and two independent retroviruses, WEHV1, WEHV2, were isolated from

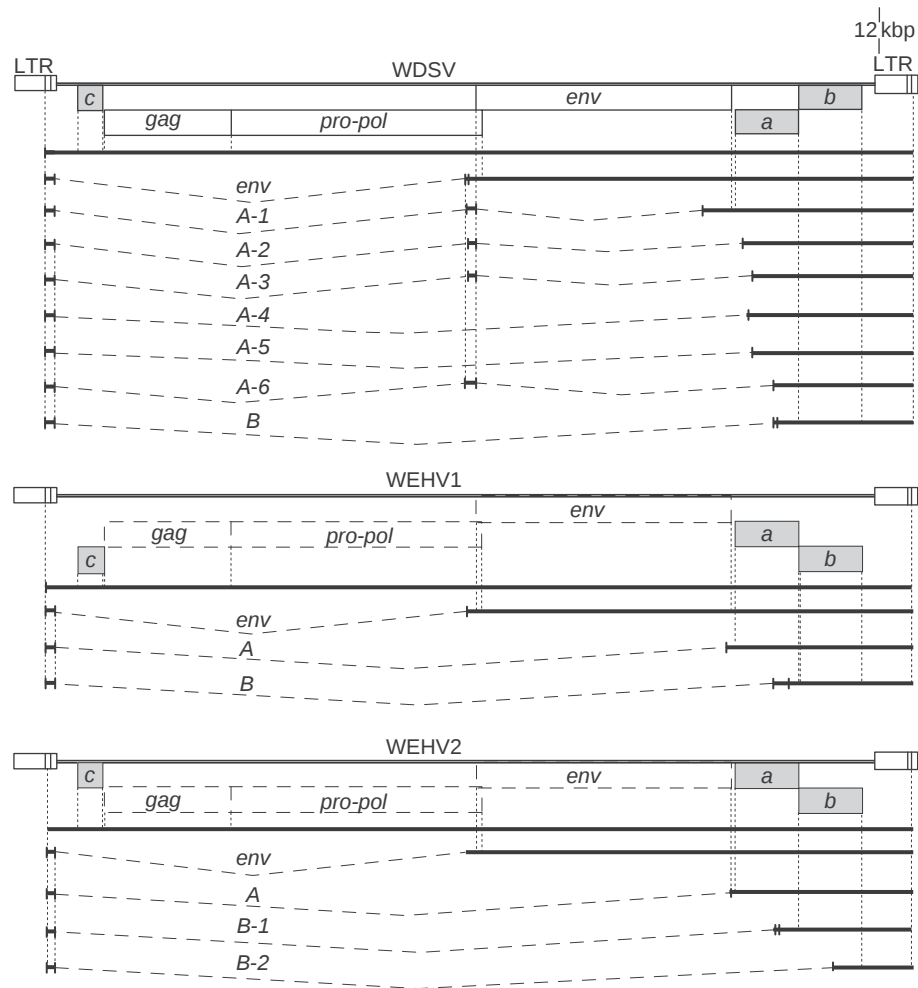


Fig. 2 Genomic and transcriptional maps of WDSV, WEHV1, and WEHV2. Genomic organization of viral DNA with retroviral genes and accessory genes *orfa*, *orfb*, and *orfc* is depicted under each virus heading. The envelope-spliced transcript (*env*) and additional spliced transcripts (*A* and *B*), which are capable of expressing accessory genes, are illustrated. The vertical dashed lines indicate the boundaries of the exons, introns, and open reading frames.

WEH lesions. Similar hyperplastic lesions on yellow perch in Oneida Lake are associated with two new retroviruses, perch epidermal hyperplasia virus types 1 and 2 (PEHV1, PEHV2).

Isolation and Sequencing of Retroviruses from WDS and WEH Lesions

WDSV, WEHV1, and WEHV2 have been molecularly cloned and sequenced (Fig. 2). The genome structures of WDSV, WEHV1, and WEHV2 indicate large and complex viral genomes (12.7, 12.9, and 13.1 kbp, respectively). All three viruses have intact open reading frames (ORFs) capable of encoding the structural and enzymatic genes *gag*, *pol*, and *env*. *gag* (viral capsid) and *pol* (reverse transcriptase) are in the same reading frame and synthesized as a polyprotein through a termination suppression mechanism. *pro* (viral protease), responsible for cleaving the polyprotein, is located in the same reading frame as *pol*. A unique feature of all three viruses is the presence of additional orfs that have tentatively been called *orfa*, *orfb*, and *orfc*. The *orfa* and *orfb* genes are located between *env* and the 3' long terminal repeat (LTR), while *orfc* is located between the 5' LTR and the start of *gag*. Database searches with OrfA protein amino acid sequences suggest limited homology with D-type cellular cyclins within the cyclin-box motif, and are referred to as rv-cyclins. WEHV1 rv-cyclin has a 20% amino acid identity and 35% similarity with human cyclin D-3, whereas WEHV2 and WDSV rv-cyclin are most similar to human cyclin D-1 (22%/35% and 19%/29% amino acid identity/similarity). Similarities with known walleye and other piscine cyclins are not significantly greater. The WEHV cyclin homologs are 37% identical within the cyclin box motif and 21–28% identical with the WDSV cyclin. Sequence homologies between rv-cyclin and OrfB in the walleye retroviruses suggest they arose by a gene duplication event.

Roles of the WDSV Accessory Proteins in Pathogenesis

Transcriptional mapping by RT-PCR and Northern blot analyses of developing and regressing WDS and WEH have demonstrated temporal gene expression profiles and complex splicing patterns analogous to those seen in the complex retroviruses of mammals. In developing WDS tumors, only low levels of multiply spliced subgenomic transcripts are detected. These transcripts predominantly contain the coding sequences for rv-cyclin and OrfB. Regressing spring tumors contain high levels of full-length and subgenomic transcripts as well as unintegrated viral DNA and transmissible virus. OrfC is likely encoded by the full-length transcript coincident with tumor regression.

A detailed transcriptional analysis of WDSV showed an alternative splicing pattern of the *orfa* transcript (Fig. 2). In developing tumors, the *orfa* transcript contains the coding sequence for a full-length rv-cyclin protein that localizes to the nucleus of mammalian and walleye cells. Alternatively spliced forms of this transcript encode amino-terminal truncated forms of the rv-cyclin protein, which localize in the cytoplasm of cells. The different forms of rv-cyclin may play functionally different roles in developing and regressing tumors. A single, spliced transcript encodes the OrfB protein, which is predominantly localized in the cytoplasm in tumor explant cells, but is capable of shuttling into and out of the nucleus when expressed in piscine and mammalian cells.

The OrfC protein from WDSV has been shown to localize to mitochondria and to disrupt mitochondrial function, which results in apoptosis. Only the full-length viral transcript found in regressing tumors is capable of encoding OrfC. Apoptotic cells are present in regressing, but not developing, WDS tumor sections. Therefore, the OrfC protein may be responsible for a direct viral mechanism of tumor regression.

WDSV *orfa* and *orfb* are the only virus transcripts found in developing tumors, suggesting direct roles for rv-cyclin and OrfB in the process of tumorigenesis, and their oncogenic potential has been demonstrated by several experimental approaches. WDSV *orfa* supported growth of a yeast (*Saccharomyces cerevisiae*) strain conditionally deficient for the synthesis of the G1 to S cyclins that are necessary for cell cycle progression. WEHV *orfa* did not support yeast growth in this model. Transgenic expression of WDSV *orfa* in mice from a skin-specific promoter caused a moderate to severe squamous epithelial hyperplasia and dysplasia dependant on skin injury. This suggests that rv-cyclin is not oncogenic as a result of a 'single-hit' mechanism, but rather, that secondary genetic or epigenetic events (possibly wound repair) are necessary for tumor development and progression. A similar phenotype has been described in transgenic mice expressing *v-jun* oncogene from the H-2K^k major histocompatibility complex (MHC) class I antigen gene promoter. When expressed in mammalian or piscine cell culture, WDSV rv-cyclin localized in the nucleus and was found by co-immunoprecipitation to be associated with cyclin dependent kinase 8 and cyclin C, general transcription initiation factors, and RNA polymerase II transcription complexes. In piscine cells, rv-cyclin inhibited transcription from the WDSV promoter independent of *cis*-acting DNA sequences. However, rv-cyclin can activate other viral promoters in fish cells, and can activate the WDSV promoter in select mammalian cells; thus rv-cyclin activation and inhibition of transcription are dependent on both the promoter and the cell type. WDSV rv-cyclin contains a defined transcription activation domain in the carboxy end of the protein, and the isolated activation domain is capable of interacting with co-activators of transcription. WEHV rv-cyclins do not have a corresponding carboxy-region activation domain. WDSV rv-cyclin may exhibit oncogenic potential by differential regulation of host gene expression such as proto-oncogenes or tumor suppressors.

Less is known about the capabilities of the OrfB protein, but initial studies indicate its direct association with the regulation of signal transduction pathways. It is concentrated, in tumor explant cells, at focal adhesions, and along actin stress fibers. Established OrfB-expressing lines are resistant to the chemical induction of apoptosis, suggesting a role in oncogenesis.

The origins of the accessory genes of complex retroviruses are unclear. This includes the origins of the WDSV accessory genes encoding rv-cyclin, OrfB, and OrfC. In the case of rv-cyclin and OrfB, their extreme divergence from host cyclin sequences indicates that any transduction event was ancient and led, ultimately, to an exclusive, complex viral species. The OrfC protein, like the accessory proteins of other complex retroviruses, has no clear homology to host proteins.

Control of the Seasonal Cycle of Disease

An interesting aspect of WDS is the control of the seasonal switch in viral gene expression and associated tumor regression. Potential modulators of viral gene expression include host immunity and endocrine activity (accompanying spawning) and environmental factors such as water temperature, physical trauma, and sunlight. A number of *cis*-acting elements important for transcription activation have been identified in the WDSV promoter. There is differential binding of proteins from developing and regressing tumor nuclear extracts to a 15 bp repeat region in the WDSV promoter. This element may be critical to the induction of high levels of virus expression. The WDSV rv-cyclin protein negatively regulates the WDSV promoter in tissue culture cells. Presumably, it is advantageous for the virus to have lower levels of gene expression during tumor development to avoid immune surveillance as well as the cytopathic effects associated with virus production. While rv-cyclin may function in the repression of virus expression during tumor growth, the host, environmental, or viral signals that switch on full virus expression are yet to be determined.

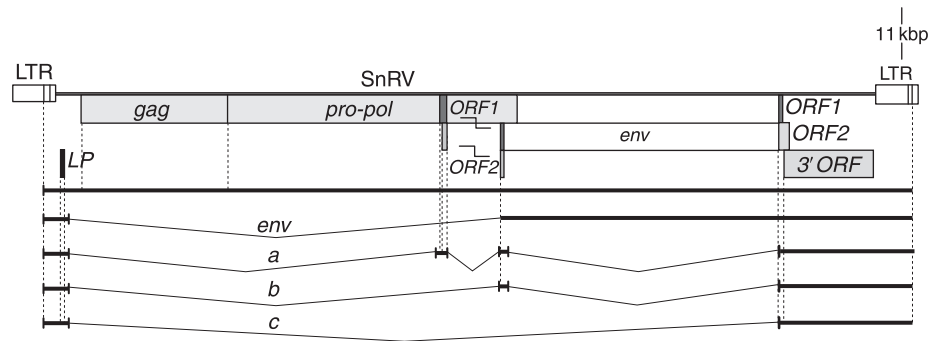


Fig. 3 Genomic and transcriptional map of SnRV. Predicted coding regions are shown as open or shaded boxes. The envelope-spliced transcript (*env*) and additional spliced transcripts (*a*, *b*, *c*), which are capable of expressing accessory genes, are illustrated. LP indicates the location of a predicted leader peptide. The vertical dashed lines indicate the boundaries of the exons, introns, and open reading frames.

Snakehead Retrovirus

The SnRV was isolated from a productively infected cell line derived from a Southeast Asian striped snakehead fish. Cell culture supernatant from these infected cells demonstrated high levels of RT activity and the presence of type C-like retrovirus particles. Additionally, supernatant from the infected cell line induced cytopathic effects in cultures of a bluegill cell line (BF-2).

Molecular approaches were utilized to identify and sequence the SnRV from infected cells. The large 11.2 kbp genome of SnRV contains intact coding regions for *gag*, *pol*, and *env* (Fig. 3). An arginine tRNA primer binding site used for reverse transcription initiation distinguishes SnRV from other retroviruses. The structure and transcriptional profile of SnRV suggests a complex expression pattern capable of encoding an ORF located between *env* and the 3' LTR and two very small ORFs termed ORF1 and ORF2. Within the leader sequence of SnRV resides a start codon located just upstream of the major splice donor site that could potentially encode a 14 amino acid peptide (LP). Expression of Env from a singly spliced transcript is predicted to utilize the LP initiation codon and fuses this 14 amino acid peptide in frame to downstream *env* sequences. A transcript with four exons and two initiator codons would encode ORF1 and ORF2 proteins. The 3' ORF would be expressed from a transcript containing three exons and would encode a protein of 24 kDa. Finally, the fourth spliced transcript may encode 3' ORF or possibly a LP-Env cytoplasmic domain (CD) fusion protein. The 3' ORF contains an N terminal acidic domain, cysteine residues, and a basic region, motifs commonly found in transcriptional activators. These small ORFs have no significant homology to any known proteins in the databases, and their role in the viral life cycle is unknown. No endogenous copies of SnRV have been identified in the snakehead genome and an uninfected cell line has been established.

Although the presence of SnRV in wild populations has not been thoroughly examined, the virus has been independently isolated from two separate snakehead cell lines derived from whole fry tissue and from caudal peduncle tissue from juvenile fish. In all cases, fish from which these cell lines were derived appeared healthy.

Salmon Swimbladder Sarcoma-Associated Retrovirus

An outbreak of neoplastic disease of the swimbladder of Atlantic salmon was first reported at a Scottish commercial marine fish farm in 1975. Affected salmon were sluggish and in poor condition. A viral etiology was suspected, and electron microscopic evaluation of tumors revealed the presence of budding viral particles with retroviral morphology. A second outbreak occurred between 1995 and 1997 in brood stock salmon held at the North Attleboro National Fish Hatchery in Massachusetts. The fish exhibited skin discoloration and hemorrhages on the fins and body and showed a general debilitation and lack of vigor. In May of 1997, significant mortality was noted at the facility such that 35% of the population was affected by late spring of 1998. All of the affected fish displayed swollen abdomens due to multinodular masses on internal and external surfaces of the swimbladder. In several cases, these multinodular masses occupied the entire swimbladder (Fig. 4). Histologic examination revealed that the tumors were composed of well-differentiated fibroblastic cells that were arranged in interlacing bundles, which were classified as leiomyosarcomas.

Isolation of a Retrovirus from Salmon Swimbladder Sarcoma

An exogenous retrovirus, termed Atlantic salmon swimbladder sarcoma virus (SSSV), was initially identified in tumors by degenerate RT-PCR of tumor RNA, and the entire viral sequence completed by DNA sequencing of proviral DNA. In contrast to the complex walleye retroviruses, SSSV is a simple retrovirus. The viral genome contains ORFs capable of expressing *gag*, *pol*, and *env* (Fig. 5). Additionally, a short 25 amino acid leader peptide of unknown function is located upstream of the Gag-Pol polyprotein

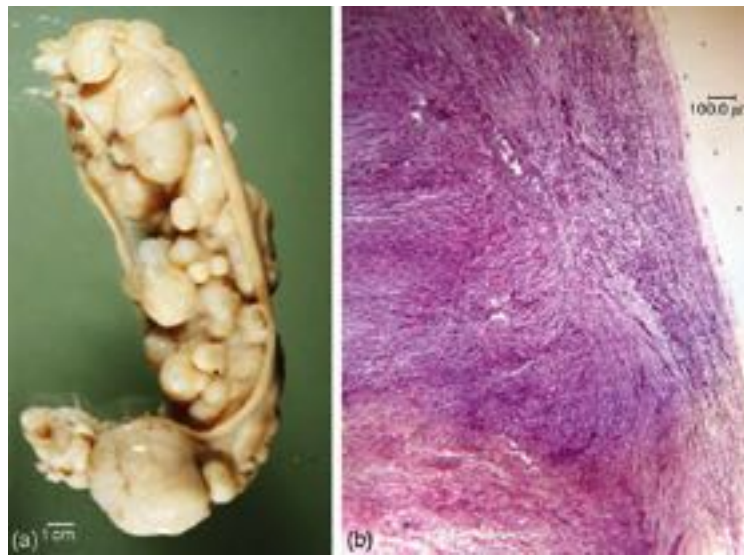


Fig. 4 (a) Gross and (b) histological images of Atlantic salmon swimbladder sarcoma.

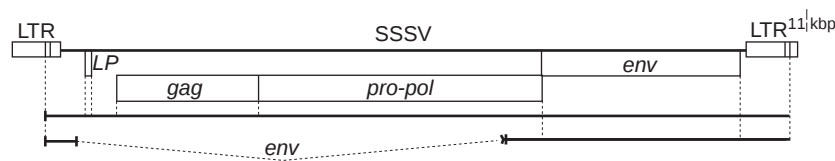


Fig. 5 Genomic and transcriptional map of SSSV. Predicted coding regions are shown as open boxes. The envelope-spliced transcript (*env*) is illustrated. LP indicates the location of a predicted leader peptide. The vertical dashed lines indicate the boundaries of the exons, introns, and open reading frame.

ORF. SSSV differs from other simple retroviruses by not having related endogenous sequences in the host genome, and SSSV is the only retrovirus to use a methionine-tRNA as a plus-strand primer. Additionally, sequences in *pol* display homology to central polypurine tract regions identified in complex retroviruses. Central polypurine tracts facilitate the formation of a single-stranded DNA region called a central DNA flap as a result of reverse transcription priming from this internal site. In HIV-1, this feature plays a major role in complex retroviral replication and allows efficient infection of nondividing cells. It is intriguing that SSSV displays a high proviral copy number (greater than 30 copies per cell) with a polyclonal integration pattern in swimbladder tumors. SSSV must be capable of initiating multiple rounds of infection within the same cell. The specific mechanisms leading to the high copy number and its implications in the pathogenesis of disease are of significant interest for future research.

Sequence Comparisons of SSSV with Other Retroviruses

A comparison of the sequence of SSSV viral proteins with other retroviruses indicates large regions of homology with mammalian type C and type D viruses in *Pol* and *Env*. Additionally, a 179 amino acid region in the C terminus of *Gag* and a 1064 amino acid region of *Pro-Pol* displays 23% and 33% identity to the related WDSV proteins, respectively. SSSV has striking homology to the sequence of the zebrafish endogenous retrovirus (ZFERV). BLAST analysis of the *Gag*, *Pol*, and *Env* ORFs of SSSV reveals a 25% identity with ZFERV over 533 amino acids within *Gag*, a 40% identity over 533 amino acids of *Pol*, and a 39% identity over 429 amino acids of *Env*.

Prevalence, Seasonality, and Transmission of SSSV

A PCR diagnostic assay was developed using sequences within the *pol* gene of SSSV. Prevalence at the North Attleboro fish facility has been found to be 52% and 5% of a natural stock of Pleasant River salmon were found to harbor SSSV.

Like other retroviral diseases of fish, observations suggest that salmon swimbladder sarcoma is seasonal. First, of the 34% fish mortality at the North Attleboro facility between 1997 and 1998, 57% occurred in June. Second, in the following year, the surviving salmon had an SSSV incidence (measured by PCR) that cycled with a peak in late summer to early winter and then diminished in late winter to early spring. The period of highest SSSV incidence correlated with salmon spawning runs in late fall.

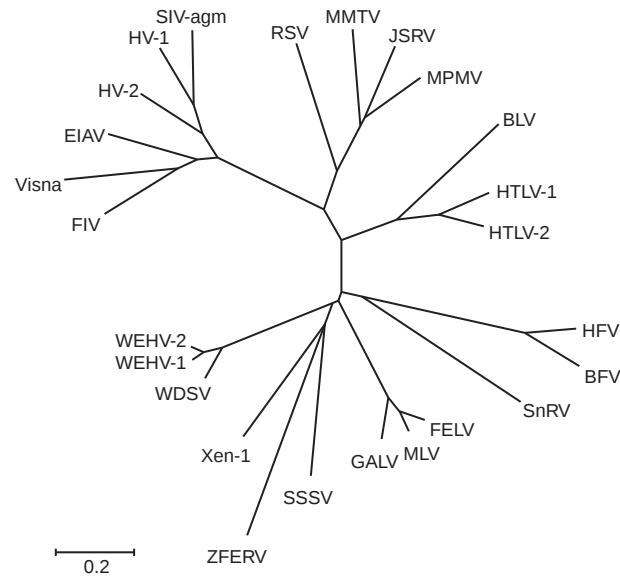


Fig. 6 Unrooted phylogenetic tree of representative retroviruses based on an amino acid alignment of seven conserved domains in reverse transcriptase. Retroviruses are designated as follows: MPMV (Mason–Pfizer monkey virus), JSRV (Jaagsiekte sheep retrovirus), MMTV (mouse mammary tumor virus), RSV (Rous sarcoma virus), EIAV (equine infectious anemia virus), FIV (feline immunodeficiency virus), Visna (visna virus), HIV-2 (human immunodeficiency virus-2), HIV-1 (human immunodeficiency virus-1), SIV-agm (simian immunodeficiency virus-agm), HTLV-1 (human T-cell leukemia virus type 1), HTLV-2 (human T-cell leukemia virus 2), BLV (bovine leukemia virus), SSSV (salmon swimbladder sarcoma virus), SnRV (snakehead retrovirus), BFV (bovine foamy virus), HFV (human foamy virus), FeLV (feline leukemia virus), MLV (murine leukemia virus), GALV (gibbon ape leukemia virus), WDSV (walleye dermal sarcoma virus), WEHV-1 (walleye epidermal hyperplasia virus type 1), WEHV-2 (walleye epidermal hyperplasia virus type 2), ZFERV (zebrafish endogenous retrovirus), Xen-1 (Xenopus endogenous retrovirus-1).

Interestingly, WDSV prevalence, which peaks during late spring, also correlates with the time of spawning. This suggests that endocrine changes during spawning runs may be a critical factor in the observed seasonal variation in disease incidence.

Pathogenesis of SSSV

The existence of SSSV sequences in association with an outbreak of the swimbladder sarcoma suggests a role for the retrovirus in pathogenesis of the disease. As a simple retrovirus with no transduced cellular oncogenes, one possible mechanism of tumorigenesis is the insertional activation of a cellular proto-oncogene. The high copy number of proviruses in tumors has made the analysis of insertional activation events difficult, but multiple insertions increase the likelihood of such a mechanism. It is also feasible that SSSV may express an oncogenic viral gene product. This mechanism of oncogenesis has been proposed for the Jaagsiekte sheep retrovirus (JSRV) in the induction of ovine pulmonary adenocarcinoma and in Friend spleen focus-forming virus (SFFV)-associated erythroid hyperplasia in mice. Additionally, the presence of common integration sites within JSRV-associated tumors suggests that insertional mutagenesis may act in concert with the envelope in tumor development. Thus, a multifactorial mechanism for the development of salmon swimbladder sarcoma must also be considered.

In addition to sarcoma, a number of other distinct pathologies are associated with SSSV infection. These diseases are frequently debilitating and are, therefore, of significance to the aquaculture industry. Atlantic salmon at the North Attleboro hatchery, during the SSSV outbreak, presented with multifocal hemorrhages, sloughing of the epidermis, lethargy, wasting, and failure to mature sexually in addition to swimbladder tumors.

Zebrafish Endogenous Retrovirus

Endogenous retroviruses have been identified in almost all vertebrate genomes, and most are defective due to mutations and deletions. An endogenous retrovirus, ZFERV with a genome of 11.2 kbp has been identified in the Tubingen stock of zebrafish, and contains intact coding regions for *gag*, *pol*, and *env*. The *gag* and *pol* genes are in the same reading frame. While the majority of endogenous retroviruses are transcriptionally silent because of mutation or methylation, ZFERV remains transcriptionally active. In addition to genomic transcripts that encode Gag and Pol proteins, an unusual multiply spliced *env* transcript is produced. Expression appears highest in the larval and adult zebrafish thymus, and no expression was detected in 2-day old embryos, suggesting that ZFERV expression may be tied to thymic development.

Phylogeny of Fish Retroviruses

Retroviruses have been classified into seven genera based largely on highly conserved amino acid sequences in the retroviral reverse transcriptase gene. While the majority of the viral sequences employed in this classification represent mammalian and avian retroviruses, a new genus termed *Epsilonretroviruses*, representing the fish retroviruses WDSV, WEHV-1, WEHV-2, has been added to the most recent classification. As more retroviral sequences from lower vertebrates have been identified, it has become apparent that this classification scheme may be inadequate to represent the apparent diversity.

Based on the unique characteristics of SnRV, including its genomic organization, tRNA primer, and complex transcriptional profile, the virus is yet to be definitively placed in the current classification. The large size of the genome, genetic organization, and presence of additional ORFs suggest that SnRV is closest to the spumaviruses and walleye retroviruses, but its limited sequence homology suggests SnRV is divergent from these groups.

Phylogenetic analysis indicates that, while the walleye retroviruses cluster in a group representing the *Epsilonretroviruses*, SSSV and ZFERV appear to represent a new branch of piscine retroviruses between the walleye retroviruses and the *Gammaretroviruses*, a genera that includes the murine leukemia virus (MLV)-related retroviruses (Fig. 6). The SnRV appears quite divergent from the other fish retroviruses by its placement in a distinct branch near the *Spumaviruses*. This would suggest that SnRV is quite divergent from the genus *Epsilonretrovirus* and may represent yet another group of retroviruses. Interestingly, a more encompassing phylogenetic analysis using all known retroviral sequences from lower vertebrates, including partial endogenous retroviral *pol* fragments from Stickleback (*Gasterosteus aculeatus*), brook trout (*Salvelinus fontinalis*), Brown trout (*Salmo trutta*), freshwater whiting (*Coregonus lavaretus*), and puffer fish (*Fugu rubripes*), indicates that the majority of the fish viruses, excluding SnRV, cluster together with MLV-related viruses in a group separate from most non-MLV related mammalian retroviruses. This raises the possibility that some retroviral groups maybe restricted to particular vertebrate classes. However, it is evident from the diversity among the fish retroviruses, that there is a high degree of heterogeneity within this group.

Acknowledgment

This research was supported in part by USDA grants 99–35204–7485 and 02–35204–12777 to J.W.C., National Oceanic and Atmospheric Administration award no. NA86RG0056 to the Research Foundation of State University of New York for New York Sea Grant to P.R.B., American Cancer Society grant RPG-00313–01-MBC to S.L.Q., and National Institutes of Health grant CA095056 to S.L.Q. T.A.P. was supported by National Institutes of Health training grant 5T32CA09682.

See also: Feline Leukemia and Sarcoma Viruses (*Retroviridae*). HIV Integrase Inhibitors and Entry Inhibitors. Simian Immunodeficiency Virus (SIV) and HIV-2 (*Retroviridae*)

Further Reading

- Bowser, P.R., Wolfe, M.J., Forney, J.L., Wooster, G.A., 1988. Seasonal prevalence of skin tumors from walleye (*Stizostedion vitreum*) from Oneida Lake, New York. *Journal of Wildlife Diseases* 24, 292–298.
- Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), 1997. *Retroviruses*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Hart, D., Frerichs, G.N., Rambaut, A., Onions, D.E., 1996. Complete nucleotide sequence and transcriptional analysis of the snakehead fish retrovirus. *Journal of Virology* 70 (6), 3606–3616.
- Hernious, E., Martin, J., Miller, K., Cook, J., Wilkinson, M., Tristem, M., 1998. Retroviral diversity and distribution in vertebrates. *Journal of Virology* 72 (7), 5955–5966.
- Holzschu, D.L., Fodor, S.K., Quackenbush, S.L., et al., 1995. Nucleotide sequence and protein analysis of a complex piscine retrovirus, walleye dermal sarcoma virus. *Journal of Virology* 69 (9), 5320–5331.
- Lairmore, M.D., Stanley, J.R., Weber, S.A., Holzschu, D.L., 2000. Squamous epithelial proliferation induced by walleye dermal sarcoma retrovirus cyclin in transgenic mice. *Proceedings of the National Academy of Sciences, USA* 97 (11), 6114–6119.
- LaPierre, L.A., Casey, J.W., Holzschu, D.L., 1998. Walleye retroviruses associated with skin tumors and hyperplasias encode cyclin D homologs. *Journal of Virology* 72, 8765–8771.
- LaPierre, L.A., Holzschu, D.L., Bowser, P.R., Casey, J.W., 1999. Sequence and transcriptional analyses of the fish retroviruses walleye epidermal hyperplasia virus types 1 and 2: Evidence for a gene duplication. *Journal of Virology* 73 (11), 9393–9403.
- Paul, T.A., Quackenbush, S.L., Sutton, C., Casey, R.N., Bowser, P.R., Casey, J.W., 2006. Identification and characterization of an exogenous retrovirus from Atlantic salmon swimbladder sarcomas. *Journal of Virology* 80 (6), 2941–2948.
- Poulet, F.M., Bowser, P.R., Casey, J.W., 1994. Retroviruses of fish, reptiles, and molluscs. In: Levy, J.A. (Ed.), *The Retroviridae*, vol. 3. New York: Plenum, pp. 1–38.
- Quackenbush, S.L., Holzschu, D.L., Bowser, P.R., Casey, J.W., 1997. Transcriptional analysis of walleye dermal sarcoma virus (WDSV). *Virology* 237, 107–112.
- Rovnak, J., Hronek, B.W., Ryan, S.O., Cai, S., Quackenbush, S.L., 2005. An activation domain within the walleye dermal sarcoma virus retroviral cyclin protein is essential for inhibition of the viral promoter. *Virology* 342 (2), 240–251.
- Rovnak, J., Quackenbush, S.L., 2002. Walleye dermal sarcoma virus cyclin interacts with components of the mediator complex and the RNA polymerase II holoenzyme. *Journal of Virology* 76, 8031–8039.
- Shen, C.H., Steiner, L.A., 2004. Genome structure and thymic expression of an endogenous retrovirus in zebrafish. *Journal of Virology* 78 (2), 899–911.

Fish Rhabdoviruses (*Rhabdoviridae*)

Gael Kurath, US Geological Survey, Western Fisheries Research Center, Seattle, WA, United States

David Stone, Weymouth Laboratory, Centre for Environment, Fisheries and Aquaculture Science, Weymouth, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

This is an update of G. Kurath, J. Winton, Fish Rhabdoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00493-3.

Glossary

Exophthalmia Abnormal protrusion or bulging of the eyeball

Finfish Fish with fins, as opposed to shellfish

Salmonid Fish in the family *Salmonidae*

Classification

The family *Rhabdoviridae* currently has 18 genera accepted by the International Committee for Virus Taxonomy (ICTV), and three of those genera contain fish rhabdoviruses. In the genera *Novirhabdovirus*, *Sprivirus*, and *Perhabdovirus* all viruses infect fish hosts, and there are no fish viruses in any of the other 15 rhabdovirus genera. In the overall phylogeny of the *Rhabdoviridae* the three fish virus genera are well separated from each other, and the novirhabdovirus genus occupies a position basal to all other genera (Fig. 1). The novirhabdovirus genus was established in 1998 and contains four viral species, as listed in Table 1. Two of these species contain infectious hematopoietic necrosis viruses (IHNV) and viral hemorrhagic septicemia viruses (VHSV), which are globally important fish pathogens that are well characterized in terms of molecular biology and genetic diversity, with many known genetic subgroups and strains within each species. Hiram rhabdoviruses (HIRRV) are also important pathogens but with a more restricted geographic and host range, and snakehead rhabdoviruses (SHRV) have not been extensively characterized.

The sprivirus and perhabdovirus genera were established more recently, in 2013 and 2012, respectively. There are two recognized species in the sprivirus genus (Table 1). The species *Carp sprivirus* contains a single virus, spring viraemia of carp virus (SVCV); another of the important fish pathogens that has multiple genetic subgroups distributed across the globe. The species *Pike fry sprivirus* is comprised

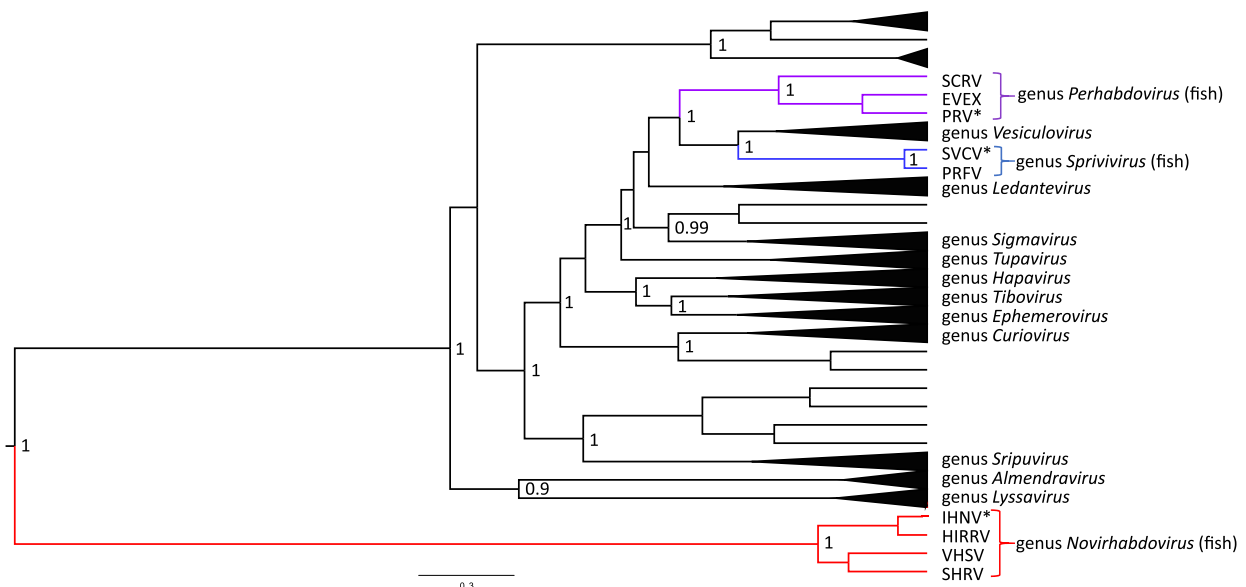


Fig. 1 Phylogenetic analysis of 105 animal rhabdovirus sequences illustrating relationships among 14 established genera, two proposed genera, and several unassigned viruses. Viral species in the three fish rhabdovirus genera (see Table 1) are shown as individual taxa, while taxa within other genera are collapsed into triangles to conserve space. The three fish rhabdovirus genera are shown as colored branches, and an asterisk denotes the type species in each genus. Unlabeled branches and triangles represent individual viruses and proposed genera, respectively, that have not yet been accepted by the International Committee for Virus Taxonomy. The four genera of plant rhabdoviruses are not included here. The tree results from full length polymerase (L) amino acid sequences aligned and analyzed with gaps (2246 characters) under a Bayesian implementation of the Yule speciation model, using the LG substitution model and constant population size prior. Posterior probability (pp) values above 0.7 are shown for critical nodes, and taxa within each established genus clustered with pp of 1.0. Scale bar is substitutions per site. Tree provided by Dr. Rachel Breyta, University of Washington.

Table 1 Classification of fish rhabdoviruses into 3 genera and several unassigned viruses^a

Genus	Species	Virus(es)	General host range ^b	Geography	Genome
<i>Novirhabdovirus</i>					
	<i>Salmonid novirhabdovirus</i>	Infectious hematopoietic necrosis virus (IHNV)	salmonid fish including rainbow trout	North America, Asia, Europe	GQ413939.1
	<i>Piscine novirhabdovirus</i>	Viral hemorrhagic septicemia virus (VHSV)	diverse range of marine and freshwater species, rainbow trout	Europe, Asia, North America	KC778774.1
	<i>Hirame novirhabdovirus</i>	Hirame virus (HIRRV)	olive flounder, ayu	Japan, Korea	AF104985
	<i>Snakehead novirhabdovirus</i>	Snakehead virus (SHRV)	snakehead	Southeast Asia	AF147498
<i>Sprivirus</i>					
	<i>Carp sprivirus</i>	Spring viremia of carp virus (SVCV)	wide range of cyprinid species	Europe, Asia, North America	U18101
	<i>Pike fry sprivirus</i>	Pike fry rhabdovirus (PFR) Grass carp rhabdovirus (GrcRV) Tench rhabdovirus (TenRV)	wide range of cyprinid species, pike and brown trout	Europe	JF872827 NC025376.1 KC113517.1
<i>Perhabdovirus</i>					
	<i>Perch perhabdovirus</i>	Perch rhabdovirus (PRV)	perch	Europe	JX679246
	<i>Sea trout perhabdovirus</i>	Lake trout rhabdovirus (LTRV)	brown trout varieties	Europe	AF434992
	<i>Anguillid perhabdovirus</i>	Eel virus European X (EVEX) Eel virus American (EVA)	eel	Europe, Asia	X827265
Unassigned					
		Starry flounder rhabdovirus	starry flounder	North American Pacific coast	AY450644
		<i>Siniperca chuatsi</i> rhabdovirus	mandarin fish	China	NC_008514.1
		Chinese rice-field eel rhabdovirus	Asian swamp eel	China	MH319839.1
		<i>Scophthalmus maximus</i> rhabdovirus	turbot	China	HQ003891.1

^avirus listed under the 3 genera have been established as species by the ICTV (2018 taxonomy). Viruses listed as unassigned are examples of those that have been described in the literature but are not currently recognized by the ICTV.

^bformal classification for fish hosts: salmonid, *Oncorhynchus* and *Salmo* sp.; rainbow trout, *Oncorhynchus mykiss*; olive flounder, *Paralichthys olivaceus*; ayu, *Plecoglossus altivelis*; snakehead murrel, *Channa striata*; starry flounder, *Platichthys stellatus*; cyprinid, *Cyprinidae* sp.; pike, *Esox lucius*; brown trout, *Salmo trutta*; perch, *Perca fluviatilis*; eel, *Anguilla* sp.; mandarin fish, *Siniperca chuatsi*; Asian swamp eel, *Monopterus albus*; turbot, *Scophthalmus maximus*.

of three recognized viruses; pike fry rhabdovirus (PFRV), grass carp rhabdovirus (GrcRV) and tench rhabdovirus (TenRV). These viruses are confined in their geographical distribution and are generally considered to have less of a disease impact when compared to SVCV.

The perhabdovirus genus has three virus species: *Perch perhabdovirus*, the type species which consists of a single virus member, perch rhabdovirus (PRV); *Anguillid perhabdovirus* which consists of two recognized viruses, eel virus European X (EVEX) and eel virus American (EVA); and *Sea trout perhabdovirus* which consists of two recognized viruses, the lake trout rhabdovirus (LTRV) and the Swedish sea trout virus (SSTV). These viruses are also confined in their geographical distribution, and the evidence that they have a significant disease impact is limited. Beyond the three established genera, several viruses were observed historically in a wide range of fish species and tentatively identified as rhabdoviruses based on the characteristic virion morphology. However, they remained unclassified due to the absence of genetic data, and most of the sample material in these cases has been lost so it is not possible to establish how they relate to the current taxonomy. Other more recently isolated fish rhabdoviruses are currently unassigned because they have not yet been sufficiently well characterized (Table 1).

Virus Structure

Fish rhabdovirus virions have the bullet-shaped morphology that is characteristic of rhabdoviruses in general. Virus particles have a host-derived lipid envelope, and they are sensitive to inactivation by desiccation, as well as heat or UV exposure. Novirhabdovirus virions are approximately 150–190 nm in length and 65–75 nm in diameter (Fig. 2(A)). Sprivirus virions are approximately 120–180 nm in length and 60–90 nm in diameter, and the perhabdovirus virions have a more compact morphology and measure approximately 115–130 nm in length and 85–95 nm in diameter. The unassigned starry flounder rhabdovirus is even more compact, with bullet-shaped virions measuring approximately 100 × 50 nm (Fig. 2(B)). Virions are comprised of five viral proteins. The single glycoprotein (G) on the enveloped surface of the virus particles has been shown to be the major antigenic protein and is by itself capable of eliciting a protective immune response. The matrix protein (M) is an inner component of the virions that binds both the cytoplasmic domain of the G protein

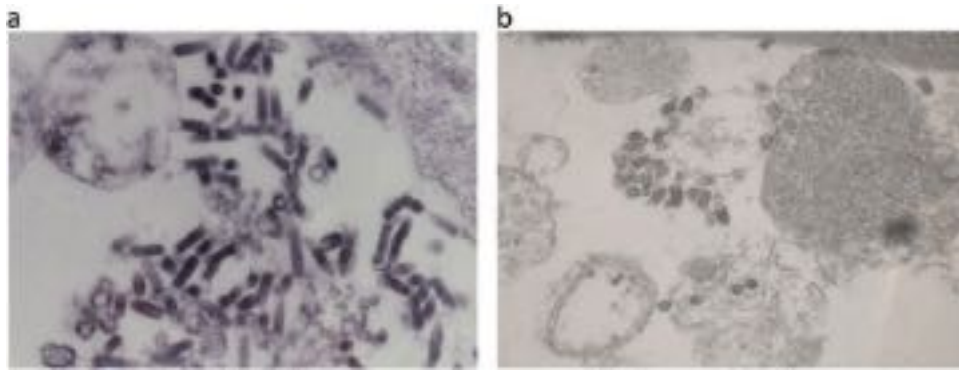


Fig. 2 Electron micrograph showing virions of fish rhabdoviruses. (a) Viral hemorrhagic septicemia virus (VHSV), in the genus *Novirhabdovirus*, (b) Starry flounder rhabdovirus (SFRV), an unassigned rhabdovirus of fish. Photos provided by J.R. Winton.

and the viral nucleocapsid, and it has been shown for IHNV to down-regulate host transcription. The major structural component of the nucleocapsid is the viral nucleocapsid (N) protein that associates with full length viral genomic RNA and facilitates its interaction with the viral polymerase. The large polymerase protein (L) is a complex protein that carries out the enzymatic functions of genome replication and transcription, and the phosphoprotein (P) is an essential co-factor that is required for polymerase function. These five virion proteins are found in all known rhabdoviruses and to date it appears that their roles are conserved across genera. In addition to the five familiar rhabdovirus proteins, the genus-specific hallmark of the novirhabdoviruses is the presence of a sixth protein, the nonvirion (NV) protein, which is not found in virus particles, but it is synthesized, in trace amounts, within infected cells. It has recently been demonstrated that the viral NV protein plays a role in abrogating host innate immune responses and suppressing apoptosis. Through the use of reverse genetics it has been shown that the NV gene and its encoded protein are not essential for replication in cultured cells or in fish, but in most cases it greatly enhances replication and pathogenicity. The absence of an NV gene in the genomes of both the spriviviruses and perhabdoviruses indicates that whatever the role of the NV protein is in novirhabdoviruses, it is not essential for rhabdovirus replication in fish.

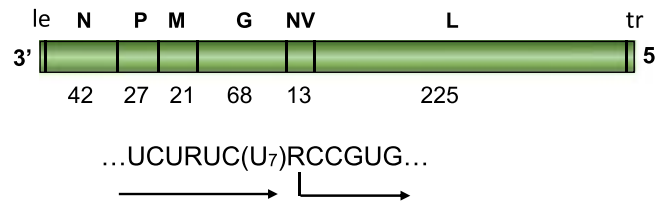
Genome

Fish rhabdovirus genomes are single stranded minus-sense RNA molecules of approximately 11,000–11,100 nt for novirhabdoviruses and spriviviruses, or 11,400–11,800 nt for perhabdoviruses. All contain the five canonical rhabdovirus genes in the order 3' N-P-M-G-L 5', and novirhabdovirus genomes have the additional NV gene located between the G and L genes, as illustrated in **Fig. 3**. Genomic termini have leader and trailer regions ranging from 50 to 100 nt for 3' leaders, and 19–115 nt for 5' trailers, and these sequences have inverse complementarity near the genome ends. Untranscribed intergenic regions are generally single nt for novirhabdoviruses, 2 nt for spriviviruses, and more variable for perhabdoviruses. Putative transcription initiation and termination/polyadenylation signals (**Fig. 3**) are conserved across genes within each virus and among viruses within a genus. These signals are identical for spriviviruses and perhabdoviruses but they differ from those of novirhabdoviruses and other rhabdovirus genera. Within genes the untranslated regions upstream and downstream of the open reading frames are variable in length. For example, the putative transcript encoding the G gene of SVCV viruses differs by up to 85 nt due to variation in the 3' untranslated regions of the gene.

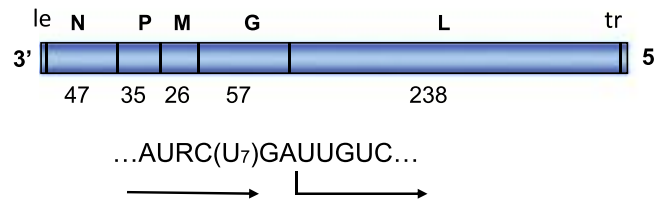
Genetic Diversity

Molecular epidemiology and phylogenies have been described at local, regional, and global levels for all major fish rhabdoviruses. Hundreds or in some cases thousands of IHNV, VHSV, and SVCV field isolates have been characterized by genetic sequencing, providing a thorough understanding of the genetic diversity and evolutionary history of these viruses. Genetic typing of IHNV field isolates from North America, Europe, Russia and Asia has shown surprisingly low diversity, with no more than 9% nucleotide divergence globally in a variable region of the G gene. Despite this low diversity, phylogenetic analyses clearly resolve IHNV into five major global genogroups that vary in geographic origin and host specificity. In North America the IHNV genogroups U, L, and M occur in partially sympatric ranges along the Pacific coast, and they show host-specific adaptation and virulence for sockeye salmon (*Oncorhynchus nerka*), Chinook salmon (*O. tshawytscha*), and rainbow trout (*O. mykiss*), respectively. The IHNV J and E genogroups arose in rainbow trout aquaculture in Asia and Europe, respectively, after intercontinental introduction from North America. Genetic typing of global collections of VHSV isolates has revealed greater diversity, with up to 18% nucleotide divergence in full length G gene sequences. Global VHSV phylogenies have 4 major phylogenetic groups designated genotypes I–IV. Genotypes I, II, and III occur in marine waters around Europe, mostly in wild fish, but within genotype I there are major subgroups that contain VHSV isolates from European rainbow trout farms, where they cause a severe disease burden. VHSV isolates in genotype IV are found in North America and Asia, where they cause disease outbreaks in several wild fish species and cultured flounder, but they are not virulent for rainbow trout.

a. Novirhabdovirus



b. Sprivivirus



c. Perhabdovirus

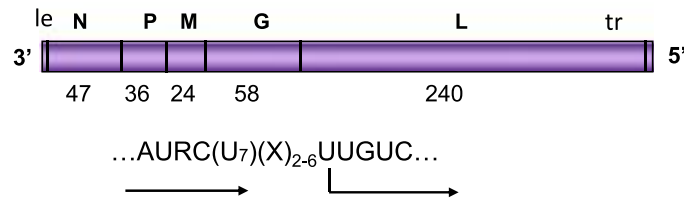


Fig. 3 Genomes of representatives of the three genera of fish rhabdoviruses. All have minus-sense single-stranded RNA genomes of approximately 11,000–11,800 nucleotides (nt) containing nucleocapsid (N), phosphoprotein (P), matrix (M), glycoprotein (G), and polymerase (L) genes as shown. Species in the genus *Novirhabdovirus* have an additional gene encoding a nonvirion (NV) protein at the G-L junction. Approximate sizes (in kDa) of encoded proteins are shown below each gene. Conserved gene junction sequences for each virus are shown in minus-sense (viral RNA) orientation below the genome with arrows indicating putative transcription stop (ending with seven U's) and start signals.

Spriviruses, particularly the SVCV isolates, are well characterized, and partial G gene sequences revealed up to 17.3% nucleotide divergence. Four SVCV genogroups (a-d) have been identified, and the clustering of the genogroups appears closely associated to geographical location, suggesting that the virus has evolved independently in different geographical regions. Genogroup a contains isolates from Asia, and more recently North America and the UK; genogroups b and c contain isolates from Eastern Europe; and genogroup d contains isolates from multiple European countries. To date, the pike fry spriviruses have only been isolated in Europe, where they show relatively high diversity, with up to 28.5% divergence based on partial G gene sequences. The genetic diversity of Perhabdoviruses is less well defined, however, the limited studies have shown that nt diversity within species ranges from 7.9% for *Sea trout perhabdovirus* to 9.3% for *Anguillid perhabdovirus* and 14.2% for *Perch perhabdovirus*.

Global phylogenies of fish rhabdoviruses show clear evidence of diversification into multiple lineages associated with geographic translocations, host jumps and adaptation to different fish hosts, and emergence events. Most recognized host jumps are into cultured rainbow trout, which has occurred multiple times in the histories of both IHNV and VHSV. Viral emergence events often occur after introduction of fish rhabdoviruses into a new geographic range. A classic example is the emergence of SVCV genogroup a in the USA, Canada and the United Kingdom following its potential introduction with infected carp from Asia in the early 1990s. Another example is introduction of VHSV into the Great Lakes region of North America in approximately 2003, resulting in emergence of a new VHSV genotype IVb, that caused large-scale die-offs in many wild fish species for several years. On a practical level molecular epidemiology of fish rhabdoviruses has been useful for fish health management decisions, tracking viral movements, defining viral transmission routes, and supporting risk assessment. Resulting viral phylogenies show clear evidence of anthropogenic influence on directions of virus evolution, with major impacts on fish health and aquaculture world-wide.

Life Cycle

At the cellular level the replication cycle of fish rhabdoviruses is similar to that of rhabdoviruses in other vertebrate hosts. The viral glycoprotein (G) forms trimeric spikes on the virion surface that attach to receptors on fish host cells, facilitating entry by receptor-mediated endocytosis. Fibronectin has been identified as a primary cellular receptor for novirhabdoviruses but the receptor(s) used by most fish rhabdoviruses have not been identified. Once inside the cell the virus is uncoated, releasing the viral nucleocapsid, and the replication cycle occurs in the cytoplasm. Primary transcription of the negative-sense viral genome to produce mRNAs is

accomplished by viral polymerase complexes that contain the viral L, N, and P proteins and are carried in the virus particles. Viral mRNAs are produced in order from the N gene at the 3' end of the genome through the L gene at the 5' end, and translation by host cell machinery generates viral proteins. Based on knowledge from mammalian rhabdovirus systems it is assumed that viral proteins form more polymerase complexes and secondary mRNA transcription produces more viral proteins that accumulate and signal for the polymerase to switch from transcription to replication. Replication creates full-length anti-genomes that serve as templates for progeny RNA genomes. Viral proteins and genomes traffic to the cell surface and assemble, budding through the cell membrane to acquire the viral envelope as new virions are released.

Epidemiology

IHNV, VHSV and SVCV are each currently distributed over several continents in the northern hemisphere, while many other fish rhabdoviruses are distributed more regionally, either in Europe or Asia ([Table 1](#)). IHNV was originally endemic to Western North America before spreading to Asia and Europe in the 1960s and 1980s, respectively. IHNV has a relatively narrow host range, naturally infecting various species of salmon and trout (*Oncorhynchus* sp., *Salmo* sp., *Salvelinus* sp.) and showing variation in host-specific virulence for individual salmonid species. HIRRV also appears to have a narrow host range, causing disease primarily in cultured and wild olive flounder (*Paralichthys olivaceus*) in several Asian countries, but there are rare reports of disease in other species and its host range has not been tested extensively. In contrast, VHSV has an extremely broad host range, naturally infecting over 80 species of fish from diverse taxonomic families. Oceanographic surveys have revealed that VHSV is broadly distributed in numerous species of wild marine fish in the Northern Atlantic and Northern Pacific Oceans, serving as a reservoir that occasionally spills over into cultured fish populations. For VHSV the concept of host specificity generally refers to whether virus strains are of low or high virulence for rainbow trout: VHSV strains from marine fish may cause disease in other host species but they have very low virulence in rainbow trout, while VHSV isolates from trout farms are highly virulent for rainbow trout.

Although the host ranges for the spriviruses and perhabdoviruses are not as great as for VHSV, they are found in a range of fish species including salmonids, cyprinids and percids. Naturally occurring SVCV infections have been recorded predominantly in cyprinid fish and non-cyprinid hosts include sheatfish (*Silurus glanis*), pike (*Esox lucius*), Siberian sturgeon (*Acipenser baerii*), largemouth bass (*Micropterus salmoides*) and bluegill sunfish (*Lepomis macrochirus*), and possibly rainbow trout (*Oncorhynchus mykiss*). The initial identifications were based on serological tests, but subsequent molecular characterization has shown that many of the original carp sprivirus isolates may belong within the pike fry sprivirus rather than carp sprivirus. PFRV was initially reported in pike, sheatfish and grass carp (*Ctenopharyngodon idella*), but given the recent findings, further genetic characterization is required to establish the full host range of the pike fry spriviruses.

PRV, the type species of the perhabdoviruses was first isolated from cultured perch (*Perca fluviatilis*) following mortality events in France and Norway, and antigenically similar viruses were then recovered from diseased pike perch (*Stizostedion stizostedion*), grayling (*Thymallus thymallus*) and largemouth bass (*Micropterus salmoides*) in France and pike in Denmark. SSTV was first discovered in brown trout (*Salmo trutta m. lacustris*) in Finland and sea trout (*Salmo trutta trutta*) from Sweden, but more recently, viruses assigned to the species *Sea trout perhabdovirus* have been isolated from perch, European eel (*Anguilla anguilla*) and brown trout (*Salmo trutta*) in the republic of Ireland. EVA was isolated from American elvers (*Anguilla nostrata*) and EVEX was isolated from European eels.

Both field observations and controlled experimental studies clearly show that transmission of fish rhabdoviruses occurs horizontally by waterborne virus that is shed from infected fish and subsequently infects naive fish without the need for fish contact or vectors. Transmission can also occur vertically from parent to progeny fish by virus associated with the surface of fish eggs or milt. In aquaculture settings disinfection of fish eggs with commercial iodophores is generally effective in preventing vertical transmission. Common routes of virus introduction into fish culture facilities include import of infected fish or eggs, or virus in the water supply due to the presence of wild or feral infected fish in source waters. Within a fish farm or hatchery viruses spread with water flow, feed, or inadequate disinfection of equipment. On a regional scale virus can be spread due to general connectivity and flow of water in aquatic environments, interactions among fish culture facilities, or movement of wild fish populations. On a global scale the history of fish rhabdoviruses includes several examples of intercontinental spread of viruses such as IHNV and SVCV during the mid-1900s, due to movement of contaminated fish eggs or juvenile fish for aquaculture. Among fish that survive virus infection the existence of asymptomatic long-term carriers and reservoirs has been demonstrated in some cases and the biological role of viral persistence is an active area of research.

Clinical Features

The majority of fish rhabdovirus disease outbreaks occur in cultured fish populations, including land-based fish farms, freshwater and marine netpen farms, and conservation hatcheries that rear fish for release into natural environments. In these settings fish infected with rhabdoviruses may exhibit petechial hemorrhages, accumulation of ascites, darkening of body color, or exophthalmia ([Fig. 4](#)), although many fish die without visible clinical signs. Mortality is frequently highest in younger fish and can be explosive in aquaculture settings, resulting in losses approaching 100% in some cases. The impacts of fish diseases include direct loss of fish due to mortality in disease outbreaks, culling of infected fish to avoid spread of disease, and costs of biosecurity

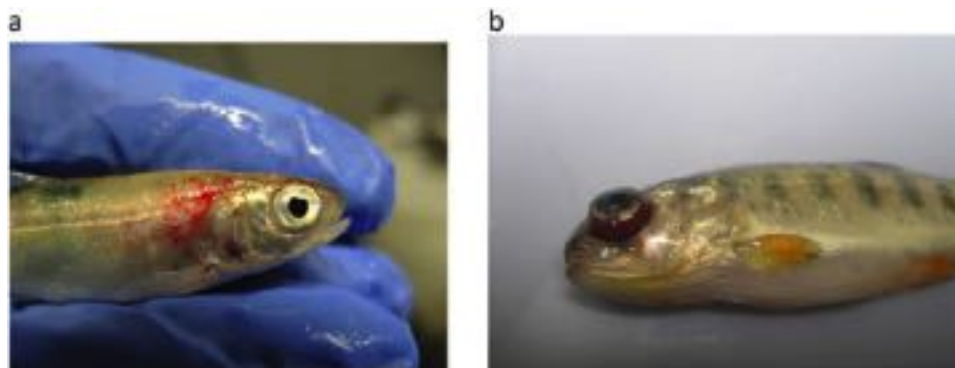


Fig. 4 Clinical signs of disease caused by fish rhabdoviruses. (a) skin hemorrhage in sockeye salmon (*Oncorhynchus nerka*) infected with IHNV (U genogroup strain), (b) exophthalmia (popeye) with eye hemorrhage in rainbow trout (*O. mykiss*) infected with IHNV (M genogroup strain). Photos by G Kurath.

and surveillance programs. Outbreaks of disease due to IHNV, VHSV, and SVCV have also been documented in wild and feral fish populations, with potential impacts on ecosystem health. The significance of fish rhabdovirus disease is evident in the OIE Aquatic Animal Health Code, which includes IHNV, SVCV, and VHSV as three of the 10 globally reportable pathogens of fish. Other fish rhabdoviruses, such as HIRRV, are regionally important pathogens.

Due to the poikilothermic (cold-blooded) nature of their hosts, a unique aspect of fish rhabdoviruses is the strong influence of temperature on virus replication and severity of disease. Fish rhabdoviruses replicate with varying temperature optima typically between 10 and 25°C, and they are inactivated at temperatures well below 37°C. In SVCV infections, clinical disease is usually observed at water temperature between 5°C and 18°C, and is most severe at temperatures below 10°C when it is believed the host's protective immune response is most likely suppressed or delayed.

Pathogenesis and Immune Response

Fish rhabdoviruses gain entry into hosts through the gills and skin. A study using bioluminescence imaging of live fish infected with fluorescent-labeled IHNV indicated the skin at fin bases as a major portal of entry. Inside the host the viruses spread to multiple organs, with hematopoietic tissues of the kidney and spleen as common targets where virus accumulates to high levels. Virus is thought to shed from infected fish in urine, feces, and sloughing of mucus, but this process has not been characterized in detail. The diseases caused by fish rhabdoviruses are generally characterized as acute, hemorrhagic septicemias affecting multiple organs. Death is usually due to organ failure and subsequent loss of osmoregulation. Internally, necrosis of multiple organs is evident upon histological examination.

The immune system of finfish has innate and adaptive immune responses typical of all vertebrates. The response to acute infection with fish rhabdoviruses is a strong interferon-based (IFN) innate response that is rapid and non-specific, resulting in transient cross-protection among different fish rhabdoviruses. It has been hypothesized that fish novirhabdovirus G proteins function as pathogen associated molecular patterns (PAMPs) to trigger signaling that activates the innate IFN response. In many cases this strong innate response is successful in controlling viral infection, leading to survival and clearance of virus. However, in other cases virulent strains of IHNV or VHSV are able to continue replicating despite the innate response, leading to disease and mortality, or viral persistence in some fish. Innate immune evasion mechanisms have been identified in novirhabdoviruses, including host cell shutoff by the viral M protein, and suppression of both the IFN-stimulated gene expression and apoptosis by the NV protein.

Fish that survive rhabdovirus infection typically develop long-term protective immunity. Adaptive immunity in fish includes both humoral and cellular responses. It has been clearly shown that neutralizing antibodies produced in response to the rhabdoviral G protein are necessary and sufficient to provide long-term immunity against fish rhabdoviruses. These antibodies become detectable in fish sera several weeks after infection. Timing of the antibody response is dependent on environmental temperature, generally occurring more rapidly at higher temperatures. There is also evidence of a role for specific cellular immunity, such as cell mediated cytotoxicity, in response to fish rhabdoviruses. Recently new reagents have become available for distinguishing among T cell subsets, which will improve the ability to characterize these cellular immune responses.

Diagnosis

Diagnostic methods for detection and identification of fish rhabdoviruses are well established. Definitive diagnosis of listed pathogens, VHSV, IHNV and SVCV is by isolation of the virus in cell culture followed with confirmation using serological and/or molecular methods. The World Organization for Animal Health (OIE), in its *Manual of Diagnostic Tests for Aquatic Animals*, includes

comprehensive methodology to diagnose and identify all three pathogens see “Relevant Websites section”. Many cell lines from finfish were developed beginning in the 1960s, and they support the replication of a broad range of rhabdoviruses from aquatic animals. Infection of these cell lines produces typical cytopathic effects (CPE) consisting of cell rounding and necrosis following the release of mature virions by budding. The rate of the appearance of CPE is related to the incubation temperature, which is typically set to mirror the temperature optimum of the fish host (e.g., 10–25°C). Cell culture-based assays are also used routinely to detect the presence of infections or to quantify infectious virus. Serological methods using polyclonal or monoclonal antibody reagents have been commonly used for both diagnostics and research, but more recently molecular techniques such as real-time RT-qPCR are often used for surveillance and outbreak tracking, with conventional RT-PCR and sequencing for confirmatory diagnosis or identification of the infectious agent. In many regions of North America, Europe, and Asia cultured fish stocks are routinely surveyed for viral pathogens, including rhabdoviruses, by fish health professionals. For most fish rhabdoviruses surveillance and genotyping data are available as internet-accessible public sequence databases.

Prevention

Control of fish rhabdovirus diseases in aquaculture is based primarily on avoiding infection through biosecurity programs designed to prevent introduction and spread of virus in fish culture facilities. Biosecurity includes strict hygiene practices, use of pathogen-free water sources (such as well water) for early life-stage rearing, application of water treatment technologies for some facilities using water from open sources, and frequent fish health inspections that make use of current technologies. In fish culture settings, the ability to break the cycle of vertical transmission between generations by the disinfection of eggs with an iodine solution has been important in controlling fish rhabdoviruses. Treatment options for fish showing signs of disease are limited to modifying husbandry conditions, typically by reducing fish density, or taking fish off feed to reduce stress. Due to the inadvertent intercontinental spread of several fish rhabdoviruses in the mid-1900s both national and international standards have been established that require fish health inspections for certain fish species that are involved in international or interstate transport. Several fish rhabdoviruses are on the list of pathogens for which inspections are indicated, and the global trade in aquaculture species is reasonably well regulated. However, the huge global trade in ornamental fish is essentially unregulated and is a documented source of virus movement.

Research on vaccines and immunostimulants to protect fish from rhabdovirus infections has been conducted for more than 40 years. Early work focused on traditional vaccination strategies using inactivated viruses or live-modified vaccines, and immunostimulants such as interferon-inducer double-stranded RNA and bacterial lipopolysaccharide. Later, recombinant DNA technology produced experimental vaccines using bacterial or baculovirus expression systems as well as peptide immunogens. However, while some of these traditional and molecular vaccines were reported to be efficacious in laboratory and field trials, prohibitive costs, inconsistent efficacy, licensing barriers, safety concerns, and lack of mass delivery methods for juvenile fish prevented their use on a commercial scale. Since the 1990s, DNA vaccines expressing the viral G gene have been engineered for protecting fish against the novirhabdoviruses, IHNV, VHSV, and HIRRV. These vaccines have been studied extensively and show exceptionally high efficacy against severe viral challenges under a wide range of conditions. Although commercialization of these vaccines faces challenges similar to those noted above, a DNA vaccine against IHNV was licensed in Canada in 2005 for use in Atlantic salmon marine netpen aquaculture. This was one of the first DNA vaccines licensed for use in veterinary medicine and there is currently an industry requirement for 100% vaccination, so the IHNV DNA vaccine is used for millions of fish annually. Interestingly, DNA vaccines for SVCV also provide significant protection, but their efficacy is less than novirhabdovirus DNA vaccines for unknown reasons. In the absence of licensed vaccines for VHSV or SVCV, the national or international legislative controls and sound biosecurity strategies remain the most effective way of preventing the spread of these diseases. Also in the absence of commercially available vaccines some private aquaculture companies may use autogenous vaccines against fish rhabdoviruses, but the extent of this practice is not well documented.

Conclusions

Due to their economic importance, the rhabdoviruses of finfish are among the best studied of all aquatic animal viruses, providing models of how aquatic animal pathogens function, spread, and persist in their environments. Extensive oceanographic surveys have revealed the role of wild fish as important reservoirs of virus for infection of fish in freshwater and marine aquaculture. Conversely, the potential for aquaculture fish to amplify pathogens that may impact wild fish is also a current concern. Field studies are providing insights into the ecology and evolution of fish viral disease, identifying anthropogenic impacts on fish health in altered environments, and supporting development of landscape transmission models for some fish rhabdoviruses.

In laboratory studies fish rhabdoviruses also serve as useful components of model systems to study vertebrate virus disease, epidemiology, and immunology. The availability of a variety of established fish cell lines, the creation of new-generation reagents and tools, and the well-established laboratory challenge models that can use statistically robust numbers of relevant fish hosts make these model systems particularly attractive and powerful. New high throughput sequencing technologies have made the discovery, characterization and classification of viruses a simpler and more rapid process, and consequently the number of recognized rhabdovirus species is likely increase dramatically over the next decade. Experimental infections *in vivo* are increasingly

important for investigating viral pathogenesis, virulence, host-specificity of new emerging viral diseases and assessing the impact of environmental factors, or efficacy of control options such as vaccines. Reverse genetics systems for IHNV, VHSV and SHRV have been used to investigate the roles of individual viral proteins and the genetic determinants of virulence, as well as the use of recombinant viruses as gene vectors or attenuated vaccines. In the area of immunology, quantitative real-time PCR assays, microarray analyzes, and transcriptomics have been used to profile the fish immune gene response following either viral infection or vaccination. Collectively these studies provide essential information for managing fish rhabdoviruses in fish health and contribute to the basic science of virology.

Further Reading

- Ahne, W., Bjorklund, H.V., Essbauer, S., *et al.*, 2002. Spring viremia of carp (SVC). *Diseases of Aquatic Organisms* 52, 261–272.
- Bootland, L.M., Leong, J.C., 2011. Infectious hematopoietic necrosis virus. In: Woo, P.T.K., Bruno, D.W. (Eds.), *Fish Diseases and Disorders*, vol. 3. Wallingford: CAB International, pp. 66–109.
- Einer-Jensen, K., Ahrens, P., Forsberg, R., Lorenzen, N., 2004. Evolution of the fish rhabdovirus viral haemorrhagic septicaemia virus. *Journal of General Virology* 85, 1167–1179.
- Harmache, A., LeBerre, M., Droineau, S., Giovannini, M., Bremont, M., 2006. Bioluminescence imaging of live infected salmonids reveals that the fin bases are the major portal of entry for Novirhabdovirus. *Journal of Virology* 80, 3655–3659.
- Kurath, G., 2008. Biotechnology and DNA vaccines for aquatic animals. *Review Science Technology* 27, 175–196.
- Kurath, G., Winton, J., 2011. Complex dynamics at the interface between wild and domestic viruses of finfish. *Current Opinion in Virology* 1, 73–80.
- Kurath, G., Garver, K.A., Troyer, R.M., *et al.*, 2003. Phylogeography of infectious hematopoietic necrosis virus in North America. *Journal of General Virology* 84, 803–814.
- Lorenzen, N., LaPatra, S.E., 1999. Immunity to rhabdoviruses in rainbow trout: The antibody response. *Fish and Shellfish Immunology* 9, 345–360.
- Mork, C., Hershberger, P., Kocan, R., Batts, W., Winton, J., 2004. Isolation and characterization of a rhabdovirus from starry flounder (*Platichthys stellatus*) collected from the northern portion of Puget Sound, Washington, USA. *Journal of General Virology* 85, 495–505.
- Purcell, M.K., Laing, K.J., Winton, J.R., 2012. Immunity to fish rhabdoviruses. *Viruses* 4, 140–166.
- Salonius, K., Simard, N., Harland, R., Ulmer, J.B., 2007. The road to licensure of a DNA vaccine. *Current Opinion in Investigational Drugs* 8, 635–641.
- Smail, D.A., 1999. Viral haemorrhagic septicaemia. In: Woo, P.T.K., Bruno, D.W. (Eds.), *Fish Diseases and Disorders*, Volume 3: Viral, Bacterial and Fungal Infections. New York: CAB International, pp. 123–147.
- Stone, D.M., Ahne, W., Denham, K.L., *et al.*, 2003. Nucleotide sequence analysis of the glycoprotein gene of putative spring viraemia of carp viruses and pike fry rhabdovirus isolates reveals four distinct piscine vesiculovirus genogroups. *Diseases of Aquatic Organisms* 53, 203–210.
- Stone, D.M., Kerr, R.C., Hughes, M., Radford, A.D., Darby, A.C., 2013. Characterisation of the genomes of four putative vesiculoviruses: tench rhabdovirus, grass carp rhabdovirus, perch rhabdovirus, and eel rhabdovirus European X. *Archives of Virology* 158, 2371–2377.
- Vakharia, V.N., Li, J., McKenney, D.G., Kurath, G., 2019. The nucleoprotein and phosphoprotein are major determinants of the virulence of viral hemorrhagic septicemia virus in rainbow trout. *Journal of Virology*. in press.
- Walker, P.J., Blasdel, K.R., Calisher, C.H., *et al.*, 2018. ICTV virus taxonomy profile: *Rhabdoviridae*. *Journal of General Virology* 99, 447–448.
- Wolf, K., 1988. *Fish Viruses and Fish Viral Diseases*. Ithaca: Cornell University Press.

Relevant Websites

- www.fishpathogens.eu
Fish Pathogen Database.
- bioinfo/ihb.ac.cn
Fish-Associated Virus Database (FVD).
- <https://talk.ictvonline.org/taxonomy>
ICTV Virus Taxonomy 2018 Release.
- gis.nacse.org/ihnv
Molecular Epidemiology of Aquatic Pathogens (MEAP)-IHNV.
- <http://www.oie.int/en/standard-setting/aquatic-manual/access-online/>
OIE Manual of Diagnostic Tests for Aquatic Animals.

Foot-and-Mouth Disease Viruses (*Picornaviridae*)

David J Rowlands, University of Leeds, Leeds, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

History

The earliest account that clearly describes foot and mouth disease was made by Fracastorius in 1546. Today, FMD is still one of the most important diseases of domestic livestock.

It was the first animal disease shown to be caused by a filterable agent in 1897 by Loeffler and Frosch, who also demonstrated the presence of neutralizing antibody in serum. It is also the first virus for which serotype differences were recognized. Further milestones were the demonstration by Waldmann and Pope in 1921 that guinea-pigs could be infected by the virus and by Skinner in 1951 that it caused a lethal infection in suckling mice, thus providing small animal models for FMD research. Subsequently, cultivation of FMDV in tissue culture cells has enabled studies of viral structure and replication and also the large-scale production of vaccines. The combination of strict control measures and the use of vaccines has led to elimination of the disease from some parts of the world but many regions remain endemic for FMD.

Classification

FMDVs comprise a species within the *Aphthovirus* genus of the *Picornaviridae*. The nature and organization of the genome, mode of replication and structure of the virion are, in general, similar to other viruses in the family. The original subdivision of the *Picornaviridae* into the four genera, *Enteroviruses*, *Rhinoviruses*, *Cardioviruses* and *Aphthoviruses* was based on physicochemical properties such as susceptibility to acid inactivation, buoyant density of CsCl solution and the nucleotide composition. Analysis of evolutionary relationships by nucleotide sequence comparisons have partly endorsed the original classifications; however, with further virus discovery the family has now expanded greatly and currently includes 110 species which are classified into 47 genera.

The genus aphthovirus originally comprised the seven serotypes of FMDV (Fig. 1) With the discovery of more related viruses the genus was expanded to include 4 species; FMDV, bovine rhinoviruses A and B and equine rhinovirus A (Fig. 2). A combination of properties which distinguish the FMDVs include: (1) extreme sensitivity of the virion to acid inactivation (< pH 6.8 in low ionic strength buffer); (2) high buoyant density in CsCl (1.43–1.50 g cm⁻³); (3) an exceptionally long multi-domain untranslated region (UTR) at the 5' end of the genomic RNA, including a large poly C tract (a property shared with some cardioviruses); (4) three separately encoded VPg proteins; (5) the use of two alternative in-frame protein translation initiation sites; (6) a leader protease protein located in the N-terminal region of the polyprotein; (7) interruption of translation (ribosome skipping) to facilitate polyprotein processing and; (8) very rapid replication.

Virion Structure

As for other picornaviruses, FMDV particles are non-enveloped icosahedrons comprising 60 copies each of four structural proteins, VP1–4, encapsidating a single copy of the single-stranded positive-sense genomic RNA. Determination of the structure of the virus by X ray crystallography and cryo electron microscopy (Fig. 3) has shown that the three larger proteins (VP1–3) have the eight-stranded antiparallel β -barrel folding motif seen in other picornaviruses and some plant viruses. VP4 is located on the inner surface of the particle. In common with other picornaviruses, VP1 molecules are disposed around the axis of five-fold symmetry whereas VP2 and VP3 alternate around the two- and three-fold symmetry axes. In contrast to many other picornaviruses, such as enteroviruses, in which the virion remains intact as an empty particle following thermally induced release of the RNA, heat or acid degradation of FMDV particles results in dissociation into pentameric subunits, consisting of five copies each of VP1–3, with release of the RNA and VP4. Several features of the structure are peculiar to FMDV. The protein shell is generally thinner and the external surface is smoother than in other picornaviruses. This is a result of the smaller sizes of VP1–3 (VP1 213, VP2 218, VP3 220 amino acids for virus O1) compared to other picornaviruses. The truncations are in the loop regions linking the core elements of the β barrels, which in other picornaviruses form prominent features at the outer surface. The deep grooves or pits encircling the five-fold axes of many picornaviruses are not present in FMDV and the position equivalent to these invaginations is occupied by the C-terminal portion of VP1. An important exception to the generally smooth contours of the surface of FMDV is provided by the G-H loop of VP1 (Fig. 3). In serotype O viruses this large loop extends from residue c.130 to 160 of which c. 135–158 are too disordered to be visible in electron density maps. The length of the VP1 G-H loop differs between different serotypes of the virus. In serotype O1 viruses the disorder of the VP1 G-H loop is induced by a disulfide bond between the cystine residues VP2 130 and VP1 134. Under reducing conditions this bond is broken and the G-H loop collapses onto the surface of the virus in an ordered configuration. In all serotypes the VP1 G-H loop includes an immunodominant antigenic site to which a high proportion of virus neutralizing antibodies are directed. Synthetic peptides representing sequences from this region are immunogenic and can induce

Genetic Relationships Between the Seven Serotypes of
Foot-and-Mouth Disease Virus

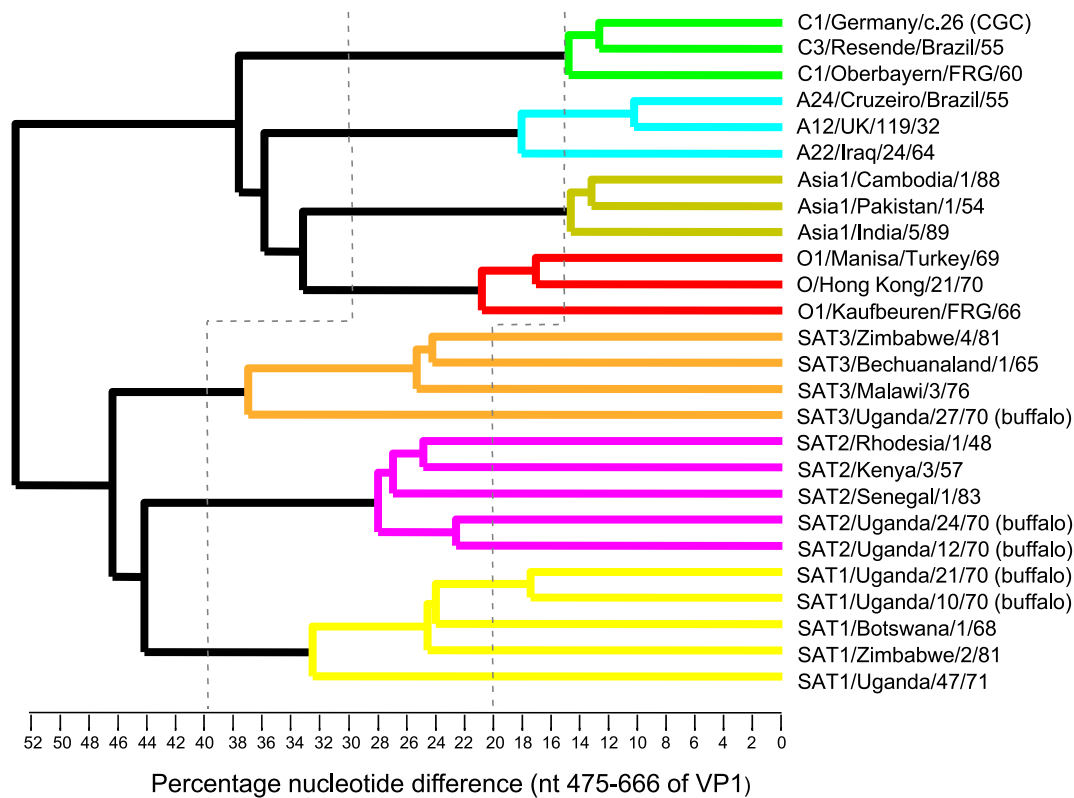


Fig. 1 Genetic relatedness of FMDV strains and serotypes based on sequences within the VP1 gene (courtesy of N. Knowles).

Genetic Relationships within the genus Aphthovirus

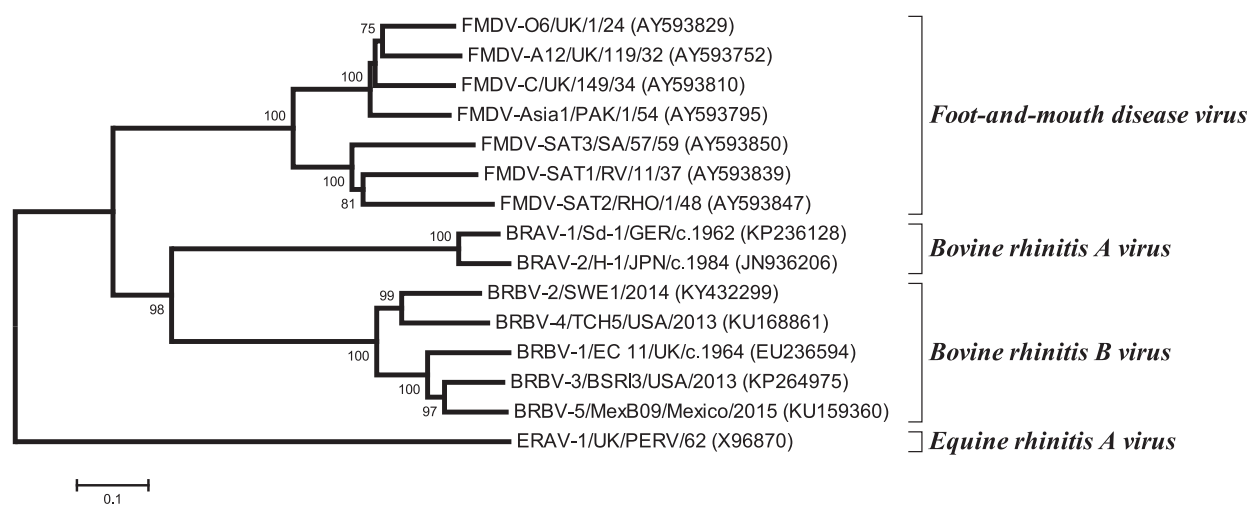


Fig. 2 Genetic relationships within the genus Aphthovirus. (Courtesy of N. Knowles).

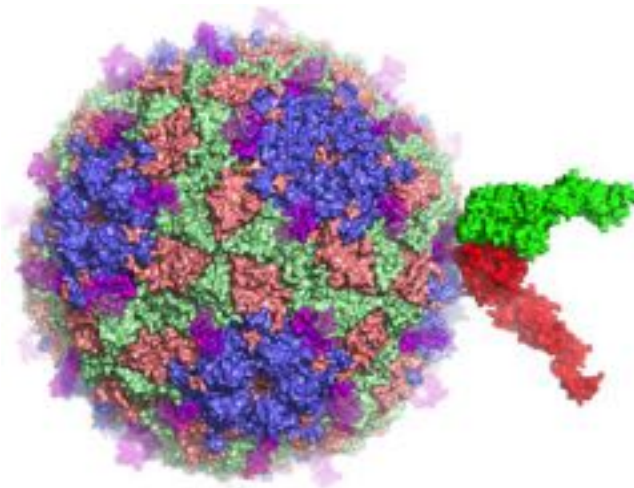


Fig. 3 The FMDV O1M/ $\alpha_v\beta_6$ receptor complex, both the O1M virus and integrin receptor are depicted using a surface representation. The capsid proteins are colored with VP1 in blue, VP2 green and VP3 salmon. The VP1 GH loop is drawn using semi-transparent magenta spheres in the “up” orientation to engage with receptor. For the integrin, the alpha subunit is green and the beta subunit red. (Courtesy of E. Fry & J. Ren).

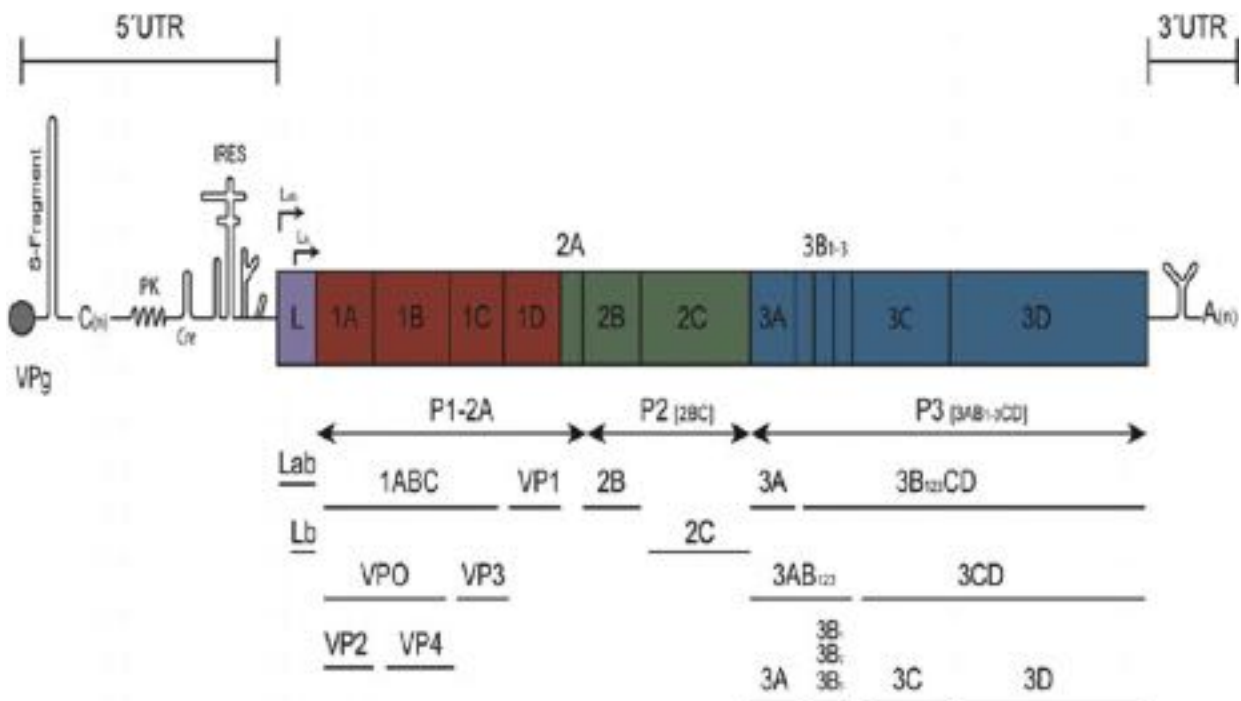


Fig. 4 Genome map of FMDV. Boxed region is the polyprotein translation product with processing sites arrowed. (Courtesy of J. Ward).

protective immunity. The sequence of this region is variable both in composition and length between different viruses with the exception of a highly conserved receptor-binding motif which includes the triplet, Arg, Gly, Asp.

Genome

The genome consists of a single molecule of single-stranded positive-sense RNA, which is infectious if transfected directly into cells in the absence of the protein components of the virion. The genome is longer than for most picornaviruses, comprising approximately 8500 nucleotides. The order of the gene products on the genome is basically similar to other picornaviruses but there are some unique/unusual features (Fig. 4). The genomic RNA terminates at the 5' untranslated end with a small protein, VPg

Structure of the 5' end of FMDV RNA

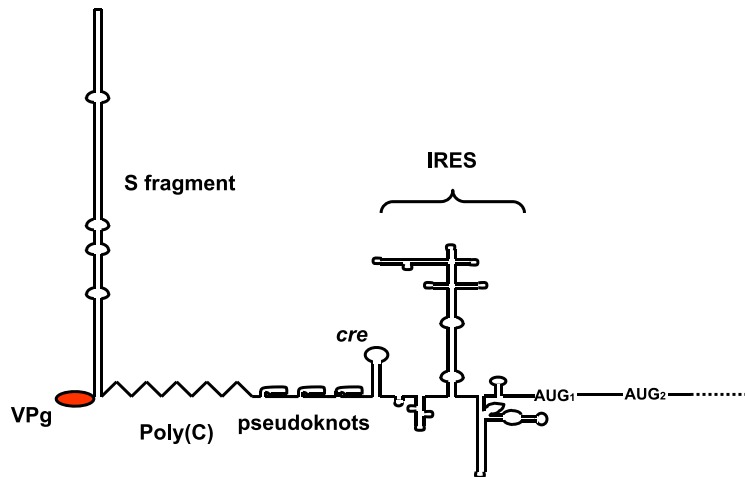


Fig. 5 Secondary structure prediction for the 5'UTR of FMDV RNA with major functional motifs individually labeled.

(or 3B), linked by a phosphodiester bond through a Tyr at position 3 to the terminal uridine residue of the RNA genome. There is a variable length of poly(A) tract at the 3' end. Uniquely, the FMDV genome encodes three distinct VPgs, which are used with equal efficiency. This is a consistent feature of all FMDVs isolated to date (except for one which has only two copies of VPg) but the rationale for this gene triplication is unknown. In fact, virus derived from infectious cDNA clones from which one or two of the VPg copies are deleted is still infectious, although RNA synthesis is slower and viruses with reduced numbers of copies of VPg may have a competitive growth disadvantage. The 5' untranslated region (UTR) is exceptionally long (c. 1300 nucleotides) even by picornavirus standards and includes at least five distinct structural motifs (Fig. 5).

The 5' end of the genome has a ~400 nucleotide sequence, known as the S fragment, which is predicted to fold into an almost complete hairpin. Its function is unknown but by analogy to the highly structured 5' sequence of poliovirus RNA it is likely to be involved in the control of RNA replication. Deletion experiments have shown that viable virus can be recovered in cultured cells (BHK-21) from genomes in which a large proportion of the predicted hairpin structure has been removed. However, truncation of the S fragment alters the host susceptibility profile and there is evidence that it interacts with cellular proteins in a species-specific manner to down-regulate innate immune response mechanisms.

The S fragment is followed by a poly(C) tract of 100–200 residues (exceptionally in excess of 400), depending on virus isolate. This tract is composed almost entirely of cytidine residues, occasionally interrupted with uridines. The function of the poly(C) tract is unknown and its presence in other aphthoviruses is equivocal. Its importance can be implied from its presence in all virus isolates and the observation that truncated poly(C) tracts increase in length following serial cell passage in cells. There is some evidence from FMDV and Mengovirus (a cardiovirus) that the length of poly(C) is related to pathogenicity. For example, deletion of the poly(C) tract from the highly virulent Mengovirus renders it completely non-pathogenic in mice.

To the 3' side of the poly(C) tract there is a variable number of repeat domains, which are predicted to fold as pseudoknot structures. The number of pseudoknots varies between different virus isolates but none has been isolated with fewer than two copies. The number of nucleotides involved in these deletions/insertions corresponds closely to the number of nucleotides required to form the predicted pseudoknot structures, providing circumstantial evidence for their structural integrity. Although the role of the pseudoknots is unclear their evolutionary retention suggests that they provide the virus with a competitive advantage.

To the 3' side of the pseudoknots there is a stem-loop structure known as the *cre* element (cis active replication element). This acts as a template for the uridylation of the VPg protein by 3D^{pol}, the viral RNA dependent RNA polymerase, in the replication complex. The uridylated VPg then functions as a primer for RNA replication. Although *cre* elements are found generally in picornaviruses, they are frequently present within the protein coding sequence and location within the 5' UTR is peculiar to FMDV.

The 3' 435 nucleotides of the 5' UTR fold into a series of stem loop structures similar to those present in the equivalent region of cardiovirus 5' UTRs and function as an internal ribosome entry site (IRES) to initiate protein synthesis. There is a short pyrimidine-rich tract between the IRES and the first of two in-frame alternative translation initiation codons.

The 3' UTR is approximately 100 nucleotides in length and includes two stem-loop structures (SL1 and SL2) which make important interactions with features in the 5' UTR to facilitate translation and RNA replication. It terminates in a poly(A) tract of variable length. Bioinformatic analysis of FMDV genome sequences has revealed an underlying feature termed GORS (genome ordered RNA structure) not seen in other picornaviruses, such as enteroviruses. It is speculated that this inherent structural feature results in the production of short double stranded RNA products following genome degradation, which may be inhibitory to innate immune mechanisms. The high GORS score may contribute to the ability of FMDV to establish persistent infections and it is also a feature of hepatitis C virus which regularly induces chronic infections.

Protein Products

The protein-coding region is a continuous open reading frame of 6999 or 6915 nucleotides for FMDV serotype A¹⁰, depending on which of two functional in-frame initiation codons is used. The order of the gene products is shown in [Fig. 4](#).

Leader protein

The leader protein L^{pro}, which precedes the structural proteins, is a papain-like cysteine protease which cleaves the polyprotein at the L-P1 junction. It also cleaves a number of host proteins with profound consequences for protein translation (see "Section Translation") and innate immune functions. L^{pro} invariably occurs as two forms, L_{ab} and L_b, which are translated from different in-frame initiation sites separated by 84 nucleotides. Both forms of the protein are equally active proteolytically and the reasons for the conservation of two initiation sites and hence the expression of two forms of the protein are still unclear. The proportion of L_b produced exceeds that of L_{ab} due to a less efficient L_{ab} translation initiation site and the presence of low abundance codons in the L_{ab} – L_b region. The L^{pro} proteins are not essential for virus viability per se since virus derived from an infectious clone with deleted L^{pro} domain can replicate. However, it has an attenuated phenotype and has been proposed as a candidate live vaccine.

P1 region

The P1 region consists of the structural proteins. P1A, B, C and D are equivalent to VP4, 2, 3 and 1 respectively. The precursor protein P1 is cleaved by the viral protease 3C^{pro} into VP0 (VP4 + VP2), VP3 and VP1 which together comprise the monomeric structural subunit. Five monomers assemble to form a pentameric subunit, which is the building block for icosahedral virion assembly. Cleavage of VP0 into VP4 + VP2 is a final assembly maturation step and the mechanism is unclear.

P2 region

The P2A protein is vestigial in size compared to other picornaviruses, being only 18 amino acids long, but it enables the nascent separation of the polyprotein at the P2A-2B junction. The amino acid sequence of 2A facilitates termination of translation of the polyprotein at this point without dissociation of the translation complex from the ribosome. Translation then continues but without the formation of a peptide bond between 2A and 2B. Cellular release factors determine the efficiency of this process and it is proposed that changes in the levels of these during infection can alter the relative amounts of structural and non-structural proteins produced. There is some direct evidence for this in Theiler's virus – a cardiovirus – which adopts a similar mechanism of interrupted translation. The rationale for such a mechanism is that it provides for the differential production of structural and non-structural proteins as requirements change during the replication cycle. 2B multimerizes to form an ion channel (viroporin) and can alter the distribution of ions across cellular membranes. It antagonizes components of the cellular innate immune mechanisms. The precursor protein 2 BC influences membrane trafficking in infected cells, inhibiting transport from the ER to the Golgi complex. P2C has a nucleoside triphosphate binding motif and has ATPase activity. It appears to have a role in RNA replication since mutations in this protein can relieve the inhibition of RNA synthesis seen in the presence of guanidine hydrochloride. It forms hexameric complexes and is thought to function as an RNA helicase during viral replication.

P3 region

The role of P3A in FMDV is unclear but it does not seem to alter cellular membrane architecture like 3A protein of poliovirus. The N terminal portion of the protein is conserved and includes both hydrophilic and hydrophobic domains through which it associates with membranes, where it functions as an important component of the RNA replication complex. The C terminal portion is variable between virus isolates and has an important influence on host specificity. Viruses with a significant deletion in 3A are highly pathogenic for pigs but are much attenuated in cattle. P3B or VPg occurs as three tandem copies of 23, 24 and 24 amino acids. Although differing in sequence each is rich in Pro, Arg and Lys residues and has a single Tyr at position three. Each of the individual VPg molecules is highly conserved between the A, O and C serotype viruses. The VPg molecules are post translationally modified to function as primers in RNA synthesis. All encapsidated RNA molecules terminate with a VPg molecule and each of the three forms is found in equal abundance in encapsidated viral RNA. Actively translating viral RNA lacks VPg, and cell extracts contain an enzyme that cleaves the phosphodiester bond to produce RNA terminating with a 5'-monophosphate. For poliovirus the cellular enzyme responsible for the removal of VPg has been identified as 5' tyrosyl-DNA phosphodiesterase 2 (TDP2) and this is likely to be also the case for FMDV. The observation that all virion associated RNA terminates in VPg suggests that it has a role in the selection of molecules for encapsidation.

P3C^{pro} is the protease responsible for the majority of the processing cleavages. In common with other picornaviruses, sequence analysis suggests that the catalytic site of 3C^{pro} is related to trypsin, a serine protease, but with the replacement of the nucleophilic serine residue with cysteine.

P3D^{pol} is the RNA-dependent RNA polymerase responsible for RNA replication and VPg uridylation. The crystal structure has been solved in the apo form and in complex with primer and template. As has been demonstrated for poliovirus, FMDV 3D^{pol} can form fibrillar structure in vitro but the significance of these in RNA replication is unknown.

P3CD precursor has a distinct function as a catalyst for the 3D^{pol} mediated uridylation of VPg. A range of precursor proteins derived from the P3 region is present in infected cells resulting from the dynamics and ordering of the proteolytic processing cascade. It is likely that some of these intermediate products have distinct roles in the replication process.

The molecular structures of virus particles, L^{pro}, 3C^{pro} and 3D^{pol} have all been determined by X ray crystallography.

Life Cycle

Cell Attachment and Entry

The large mobile G-H loop of VP1 located at the surface of FMDV particles mediates their binding to susceptible cells. The sequence of the loop contains a highly conserved motif, Arg, Gly, Asp, which is a hallmark of ligands for a number of heterodimeric cell surface molecules called integrins. Synthetic peptides including this sequence can compete with virus for cell attachment and treatment of the virus with proteolytic enzymes such as trypsin, which cleave within the G-H loop, also prevents virus binding. Although the virus can bind to a number of integrins there is good evidence that $\alpha_v\beta_6$ is the preferred receptor *in vivo*. The structure of a complex of $\alpha_v\beta_6$ bound to FMDV has been determined by cryo electron microscopy (Fig. 3).

Following attachment, the virus is internalized by endocytosis and release of the RNA is triggered by acidification in early endosomes. Reduction of the pH below c. 6.8 seems to be all that is required to initiate the infection process and the mechanism by which the genome is delivered to the cytoplasm is unknown. It is unlikely that cell entry requires complete dissociation of virions into pentameric subunits and genomic RNA triggered by acidification of the endosome since this would expose the RNA to the potentially hostile environment within the lumen. Transient empty particles are produced from aphthovirus particles under some conditions and these may represent intermediate structures that protect the genome during entry into the cytoplasm. Many tissue culture adapted strains of the virus have evolved *in vitro* to use heparin sulfate at cell surface as an alternative receptor.

RNA Replication

RNA of infecting virus functions as a template for the synthesis of a negative-sense complementary strand, which serves as template for the synthesis of positive-sense strands, identical to the original infecting molecule. Positive-strands are synthesized in a complex structure [replicative intermediate (RI)] consisting of a single negative-strand template and several (c.6) nascent positive-strands. RNA synthesis is asymmetrical in favor of positive-strands. A proportion of the negative-strand templates occur as full-length double-stranded hybrids [replicative form (RF)] and appear to take no further part in RNA synthesis. RF molecules accumulate in the cell during viral replication. Each RNA replication event is initiated by priming with an uridylated VPg molecule.

A single molecule of viral RNA is sufficient to initiate infection, which implies that it can function sequentially as a template for translation, to produce the polymerase enzyme(s), and as a template for RNA replication. Single-stranded viral RNA is infectious in the presence of inhibitors of host cell DNA-dependent RNA polymerase. Double-stranded viral RNA is also infectious but not in the presence of inhibitors of the cellular polymerases.

As for other positive sense RNA viruses, the RdRp of FMDV is error prone due to the absence of a proof-reading capability. Consequently, point mutations occur at a high rate and the viral population exists as a quasi-species – a swarm of variant molecules the sequence of which represents a consensus. In addition to a high replication error rate, recombination occurs readily. Both of these features contribute to the genetic flexibility of FMDV and enable it to respond rapidly to environmental selective pressures.

Translation

FMDV RNA is translated efficiently in a cell-free system (e.g., rabbit reticulocyte lysate) to produce protein products similar to those found in infected cells. The 435 nucleotides upstream of the first AUG initiation codon is folded into a complex structure similar to that of the equivalent region of cardiovirus RNAs. This sequence acts as an IRES, allowing initiation of translation in the absence of the host cell CAP-binding protein, eIF4E. In addition to canonical translation initiation factors, IRESs recruit a number of cellular IRES trans-activating factors (ITAFs) which appear to act as chaperones to facilitate optimal conformational folding of the IRES sequence. The rate of total protein synthesis in virus-infected cells does not change until its decline towards the end of the growth cycle, when cytopathic effects (CPE) are apparent. There is, however, a marked change in the profile of proteins produced. When viral replication is maximal, virtually no host cell proteins are produced. This “swap over” of translation from host to viral products is similar to the situation in cardiovirus-infected cells and differs from the kinetics of translation following infection of cells with enteroviruses. In the latter, infection results in a rapid shutdown of host cell protein translation, which is followed later by a resumption of protein synthesis due to the increasing production of viral proteins. The shutdown induced by enteroviruses is largely, if not entirely, due to virus-induced cleavage of a host protein, p220 or eIF4G, an important component of the CAP-binding complex required for the initiation of translation of host mRNAs. In these viruses, eIF4G is cleaved by P2A^{Pro}. Cardioviruses do not induce eIF4G cleavage but downregulate host mRNA translation via P2A induced hypophosphorylation of initiation factor 4E-BP1 and so out-compete host mRNAs for utilization of the translation machinery. Although the kinetics of protein translation in FMDV-infected cells resemble those of cardiovirus-infected cells, eIF4G is cleaved. In contrast to enteroviruses, in FMDV eIF4G is cleaved by L^{Pro} and not by P2A^{Pro}. 3C^{Pro} can also cleave eIF4G but at a different site from L^{Pro} and later in the infection cycle. Inhibition of host cell protein synthesis is likely to be of advantage to the virus both by removing competition for access to the translation machinery and also by inhibiting the expression of innate immunity response genes in infected cells. FMDV from which L^{Pro} is deleted has a reduced ability to inhibit the interferon system.

Post-Translation Processing

The polyprotein translation product of FMDV RNA is proteolytically processed by three of four virus-encoded enzyme activities (Fig. 2). Three of the cleavages occur nascently on the growing polypeptide chain. The first separates L^{pro} from P1, the structural protein precursor, and is carried out by L^{pro} itself. The cleavage is within a Lys-Gly dipeptide and probably occurs normally in *cis* but can also occur in *trans*. L-P1 cleavage is the only processing step which is inhibited by the tripeptide D-Val-Phe-Lys-CH₂Cl. Both Lb and Lab are proteolytically active.

The second primary “cleavage” occurs at the junction of P2A and P2B and is facilitated by the 18-amino acid P2A sequence. There is good evidence that the separation of 2A from 2B and the remainder of the polyprotein does not occur by proteolytic hydrolysis of a peptide bond but rather involves a novel process of interrupted translation, referred to as ribosomal skipping.

The third primary cleavage is between P2C and P3A and is catalyzed by 3C^{pro}. This protease is responsible for all other processing cleavages, apart from that which generates VP2 (P1B) and VP4 (P1A) from the precursor VP0 (P1AB). Many cleavages catalyzed by 3C^{pro} occur at Glu-Gly junctions but other dipeptides are recognized and 3C^{pro} of FMDV is the most promiscuous of the picornavirus proteases. The cleavage event to produce VP2 and 4 from the precursor, VP0, occurs in the final stages of virus maturation. The mechanism of this cleavage is not known.

The production of viral proteins via cleavage from a polyprotein allows for alternative cleavage pathways and provides a range of intermediate precursor proteins. It is thought that some of the intermediate products have roles additional to those of the mature cleavage products, thereby expanding the functional capacity of a small viral genome.

Virus Assembly and Release

A variety of assembly intermediates containing equimolar amounts of VP1, 2 and 0 are detected in infected cells. These correspond to monomer and pentamer subunits of the icosahedral capsid and 75S empty particles, which lack viral RNA but possess antigenic properties similar to mature viral particles. Following pulse-labeling experiments, empty particles can be “chased” into viral particles, but it has not been shown that they are on the direct morphogenetic pathway. All encapsidated RNA terminates with VPg, suggesting that this plays a role in selection for packaging and/or that only nascent RNA is encapsidated before VPg is cleaved off by TDP2. There is also evidence for the presence of multiple short RNA sequence motifs (packaging signals) distributed through the genome, which facilitate virion assembly.

Paracrystalline arrays of virus are visible in infected cells and are released by lysis of the cell. In addition there is evidence that some viral particles are secreted prior to cell disruption or are released within exosomes.

Epidemiology

FMD occurs widely and is endemic in many countries, especially in tropical regions. Many regions, including Europe, North America, Australia and New Zealand are free of the disease and maintain this status by rigorous application of import controls and quarantine. Mass vaccination campaigns have eliminated the virus from some areas, e.g., Europe, but have been less effective in others, due largely to logistical problems of vaccine distribution and the techniques of animal husbandry employed. The global distribution of the serotypes is shown in Figs. 6 and 7. There have been no reports of isolation of viruses belonging to serotype C since 2004 and these appear to now be extinct.

In regions endemic for FMD, the virus is most likely maintained in persistently infected animals. It has been shown experimentally that infected bovines can secrete virus for long periods after the initial episode of disease. In some areas the wild animal population may act as a reservoir for infection (e.g., Cape buffalo in Africa). In non-endemic areas infection may be introduced from a variety of sources such as the importation of infected livestock, contaminated animal products such as carcasses containing bone (in contrast to meat, the post-mortem acidification of bone marrow and other non-muscular tissues is insufficient to inactivate the virus), or contaminated materials. More locally, transmission of infection is by direct transport of contaminated animals or materials or by wind-borne carriage of infectious aerosols. It is also suspected that the virus can be passively transmitted by migrating birds. High-throughput deep sequencing methodologies now allow virus transmission to be traced forensically at high resolution.

Host Range

The virus typically infects cloven-hoofed species with domestic cattle being the most susceptible. Domestic pigs are also important hosts and are particularly effective in propagating the disease, since they secrete large quantities of virus in the form of aerosols. In sheep and goats, the clinical manifestations of infection are usually less severe than those seen in cattle and pigs making initial symptomatic diagnosis in these species more difficult. Natural infection of Indian elephants and of camels has been reported. Many wild species of deer and antelope are susceptible to infection and in African Cape buffalo infection is asymptomatic. Persistent infection with prolonged shedding of virus for months or years has been reported in wild and domestic bovine species. Although direct transmission of infection from carrier animals to naive animals has proven difficult under laboratory conditions

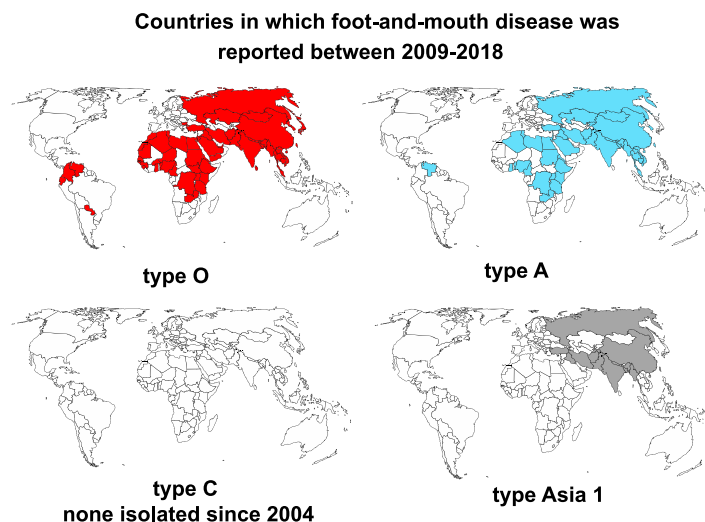


Fig. 6 Global distribution of FMDV serotypes O, A, C and Asia 1 between 2009 and 2018. (the FAO World Reference Laboratory for FMD and the OIE/FAO Reference Laboratory Network).

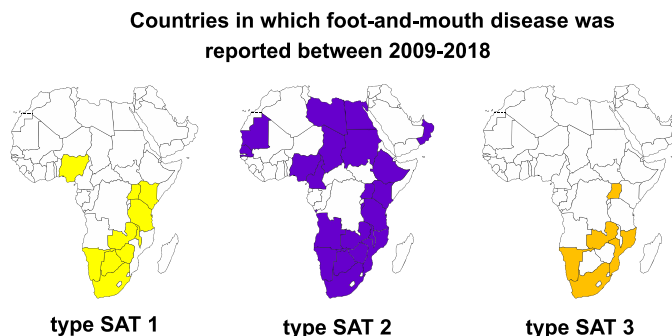


Fig. 7 Distribution of FMDV serotypes SAT 1, SAT 2 and SAT 3 between 2009 and 2018. (the FAO World Reference Laboratory for FMD and the OIE/FAO Reference Laboratory Network).

there is epidemiological and some experimental data which supports the importance of persistent infection in maintenance and transmission of disease in the wild. A wide range of animals, including Australian marsupials and even birds, have been infected under laboratory conditions and, very rarely, infection of man has been demonstrated.

The most important small animals for laboratory investigations are the guinea-pig and the suckling mouse. In the former, injection of virus intradermally into plantar pads results in the formation of vesicular lesions both at the site of injection and in the mouth and the remaining feet, and so resembles the lesion distribution in naturally infected susceptible species. Intra-peritoneal infection of suckling mice results in rapid death. The viruses can be propagated in primary cells and cell lines of bovine or porcine origin. Cells derived from the BHK-21 line are most widely used for research or vaccine production purposes. The virus can be titrated by plaque assay in cultured cell monolayers or by cytopathic end-point dilution assay in microtiter plates. Primary bovine cells (e.g., bovine thyroid cells) are used for field virus isolation, as they are more susceptible to non-culture adapted virus strains.

Genetics and Evolution

In common with other RNA viruses, the mutation rate of FMDV is extremely high and virus populations exist as quasispecies in which each individual genome is likely to differ from every other. Antigenic sites on the viral particle are tolerant of sequence variation, and antigenic diversity is a significant property of the virus.

In addition to evolution by the accumulation of point mutation, genomic recombination occurs at a high rate in vitro and comparative sequencing of field isolates suggests that it is also a frequent event in vivo. The frequency and genomic location of recombinatorial events mirrors the genetic relatedness of the parental viruses.

Serological Relationships and Variability

Seven serotypes of FMDV are recognized, the definition of serotypes being that an animal convalescent from infection by virus of one serotype is fully susceptible to viruses of any of the remaining six. In addition to the major serotype differences, there is considerable antigenic variation between viruses within serotypes. The serological relationships between FMDV isolates are paralleled by genetic relationships as evidenced by RNA sequence analyses (Fig. 1). Major advances in nucleotide sequencing technologies in recent years have resulted in these becoming the most accurate and rapid methods for evolutionary and epidemiological studies.

For epidemiological studies and vaccine strain selection the serological relationships between field virus isolates or laboratory strains are expressed as r values, i.e., the ratio of the neutralizing titers of immune sera against heterologous and homologous viruses. The serological relationships between virus isolates are frequently nonreciprocal, showing that closely related viruses may induce broadly cross-reactive or narrowly specific immune responses.

Pathogenesis and Clinical Features

The principal route of infection appears to be via aerosol impinging on the pharynx and respiratory tract. Aerosols may be produced locally during feeding on contaminated foodstuff or may be transmitted over considerable distances under appropriate meteorological conditions. Pigs secrete particularly high levels of virus-contaminated aerosols. In addition to mucosal secretion, high levels of virus are found in milk. Pasture may be contaminated with virus from urine and feces. Although the virus is labile when purified in laboratory conditions it can survive for long periods in certain environments (e.g., in slurry) at low temperatures.

Vesicular lesions appear in the mouth on the tongue, gums and cheeks and later on interdigital mucosa and coronary bands of the feet. Virus can be isolated from many tissues in the body. The onset of clinical disease is usually very rapid and lesions can develop as early as 1–2 days after infection, depending on the virus strain and level of exposure.

The pathogenicity of FMDV varies according to the virus strain, host species and age. The factors that govern the virulence of FMDVs are poorly understood and as with most viruses are probably multifactorial. Domestic cattle are the most susceptible species and morbidity is usually c. 100% in non-vaccinated animals. Wild bovines, such as African Cape buffalo, may produce no clinical manifestations. Although the disease is rarely fatal in adult domestic animals (< 5%), significant mortality may occur in young animals (c. 50%). Overt symptoms of disease are less obvious in pigs, although this species produces large volumes of infectious aerosol and are very important in disease transmission. Symptoms are much less severe in ovine species and can be unapparent, which can complicate surveillance for the presence of FMD.

Infection of bovines typically produces a rapidly progressing febrile illness and the development of massive vesicular lesions in the mouth and on the feet. Infection primarily involves mucosal epithelia in the oropharynx, which correlates with expression of the principle receptor, integrin $\alpha_v\beta_6$. The lesions rupture with considerable loss of epithelial tissue (Fig. 8). The resulting discomfort discourages feeding until the lesions heal by the infiltration of fibrous tissue. The severity of the disease results in long-term loss of productivity in terms of meat and milk yield, and lameness may be a serious consequence for draft animals. Abortion and chronic subfertility are also common. Other organs infected include mammary glands, pancreas and heart.

Immune Response

Infection elicits a vigorous humoral antibody response and, after recovery, immunity to re-infection with viruses of the same serotype is prolonged. The humoral antibody response is induced by both T-cell independent and T-cell dependent mechanisms. The role of cytotoxic T cells in recovery is unclear.



Fig. 8 Severe symptoms of FMD showing extensive sloughing of lingual epithelium.

Four antigenic sites recognized by antibodies capable of neutralizing the virus have been described. One of these is an immunodominant linear sequence comprising the G-H loop of VP1, and synthetic peptides representing this tract are effective in inducing high levels of virus neutralizing antibody and can protect animals from infection. The remaining antigenic sites are conformational and one spans the interface between protomers in the assembled particles and so is lost on particle dis-assembly.

Antibody responses to several of the non-structural proteins are induced during infection and these are being developed as markers to distinguish vaccinated from infected animals.

Diagnosis

Initial diagnosis in the field is based on clinical observations, which are usually obvious for cattle but less so for other species, especially sheep. The presence of FMDV must then be confirmed using more specific virological and/or serological methods. Cultivation of virus from clinical samples is the gold standard for confirmation but this may be complicated by lack of containment facilities to perform the isolation, difficulties in growing field virus strains in culture cells and poor quality of samples. It is also a lengthy process. Virus antigens can be detected using antibody based assays such as enzyme-linked immunosorbent assay (ELISA) using polyclonal sera or monoclonal antibodies. Genome sequence based methods relying on the reverse transcriptase – polymerase chain reaction (RT PCR) are used increasingly for diagnosis and for high accuracy epidemiological studies. Modern diagnostic techniques are becoming ever more sophisticated through the development of “pen side” rapid lateral flow devices, analogous to those used in pregnancy testing kits, and portable automated RT PCR apparatus.

Prevention and Control

In disease free regions, natural protection afforded by geographical barriers is rigorously reinforced by strict controls on the importation of susceptible animals and potentially contaminated materials. Where outbreaks occur sporadically, due to occasional introduction from external sources, embargoes on animal movement and slaughter of infected herds have been successful in maintaining a disease-free national herd. In endemic areas, control is by mass vaccination. Ring or barrier vaccination is also used to limit the spread of infection during outbreaks.

Dramatic demonstrations of the vulnerability of naive livestock to large-scale epidemics of FMD were provided by the outbreaks of serotype O virus in the UK and of serotype A virus in South America in 2001. The outbreak in the UK resulted in the slaughter of over 10 million animals (most of which were uninfected) to control spread of the disease and the cost to the country was many billions of pounds. In South America the outbreak was controlled by the resumption of blanket vaccination of all livestock.

The first vaccines against FMD were produced by formalin inactivation of lymph drawn from lesions on the tongues of infected cattle. This source of immunizing virus was replaced by the Frenkel method of culture in fragments of epithelium stripped from the tongues of slaughtered cattle. Most vaccine in use today is produced by growing the virus in suspensions of BHK-21 cells in fermentation vessels of up to 10,000 liters capacity. Approximately 2×10^9 monovalent doses of vaccine are administered annually. Aziridines have largely replaced formalin as the inactivant, since the inactivation kinetics of the latter are non-linear and residual live virus in vaccines has occasionally been the source of outbreaks of disease. The serotype and strain composition of vaccines have to be tailored for local requirements. Inactivated virus is usually adjuvanted by adsorption onto aluminum hydroxide gel, and saponin may be added to enhance potency. In pigs, such vaccines elicit only immunoglobulin M responses and for this species vaccines are formulated with oil adjuvants. Oil adjuvanted vaccines are widely used for all species in S. America and China.

Solid protection requires high levels of neutralizing antibodies and to achieve this with inactivated vaccines, immunization is repeated two to three times a year. The development of live attenuated vaccines has not progressed due mainly to the complexity of antigenic diversity and the fear of reversion to a virulent phenotype. However, greater understanding of the structure and function of the virus provides the basis for development of rationally designed live vaccines which may overcome the disadvantages inherent in earlier attempt to derive attenuated vaccine strains.

Future Perspectives

There is great scope for developments in FMD vaccines to improve safety, stability, cross-protective efficacy and duration of immunity. Experimental peptide vaccines were demonstrated to elicit protective levels of neutralizing antibodies more than 30 years ago but poor responses in target species as compared to small animal models curtailed their development. However, research continues in this field. Production of virus-like particles (VLPs) via the co-expression of the structural precursor protein, P1, together with the protease, 3C^{pro} in recombinant expression systems, such as baculovirus transduced insect cells, is developing rapidly. The inherent instability of FMDV VLPs is being addressed by the identification and inclusion of stabilizing mutations. Adenovirus vectored P1/3C^{pro} constructs are being investigated as live recombinant delivery systems and naked DNA constructs capable of expressing the same proteins have also been shown to elicit anti FMDV responses. In addition, despite the failure

through reversion to virulence of attenuated vaccines developed many decades ago, increased understanding of the structure and function of the viral genome at the molecular scale is leading to the development of rationally designed attenuated vaccine strains (virtually) incapable of reverting to a virulent phenotype. Massive outbreaks of FMD in the UK and in South America in 2001 stimulated interest in the development of drugs to halt the rapid spread of infection. Improved rapid diagnostic methods are being explored as is the ability to accurately distinguish infected from vaccinated animals.

As to the molecular properties of the virus, both the determination of the near atomic structure of the particle and several of the non-structural proteins and the cloning and manipulation of full-length infectious cDNA molecules are important steps towards understanding the unique features of the virus.

Further Reading

- Bartling, S.J., Vreewijk, J., 1991. Developments in foot and mouth disease vaccines. *Vaccine* 9, 75.
- Brooksby, J.B., 1982. Portraits of viruses: foot and mouth disease virus. *Intervirology* 18, 1–23.
- Diaz-San Segundo, F., Medina, G.N., Stenfeldt, C., Arzt, A., de los, Santos, T. Foot-and-mouth disease vaccines, 2017. *Veterinary Microbiology* 206, 102–112.
- Fry, E., Logan, D., Fox, G., *et al.*, 1990. Architecture and topography of an aphthovirus. *Semin. Virol.* 1, 439–451.
- Foot-and-mouth disease virus. Mahy, B.J.W. (Ed.), Heidelberg: Springer.
- Foot-and-mouth disease virus. In: Rowlands, D.J. (Ed.), *Virus Research* 91. Elsevier. [Special Issue].
- Foot-and-mouth disease – Current perspectives. Sobrino, F., Domingo, E. (Eds.), Great Britain: Horizon Bioscience.
- Foot-and-mouth Disease Virus – Current Research and Emerging Trends. Sobrino, F., Domingo, E. (Eds.), Norfolk, UK: Caister Academic Press.
- Yuan, Gao, Sun, Shi-Qi, Guo, Hui-Chen, 2016. Biological function of Foot-and-mouth disease virus non-structural proteins and non-coding elements. *Virology Journal*. [13/107].

Relevant Website

<http://www.picornaviridae.com/>
The Pirbright Institute.

Fowlpox Virus and Other Avipoxviruses (*Poxviridae*)

Efstathios S Giotis, Imperial College London, London, United Kingdom and University of Essex, Colchester, United Kingdom

Michael A Skinner, Imperial College London, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

APV Avipox viruses

CNPV Canarypox virus

CPE Cytopathogenic effect

dsDNA Double-stranded DNA

FWPV Fowlpox virus

GC-content Guanosine cytosine content

ICTV International Committee on Taxonomy of Viruses

PCR Polymerase chain reaction

PEPV Penguinpox virus

PGPV Pigeonpox virus

PV Poxviruses

REV Reticuloendotheliosis virus

TKPV Turkeypox virus

VACV Vaccinia virus

Glossary

Borell bodies Virus-containing granules that cluster to form “Bollinger bodies” and are found in cells infected with FWPV.

Inclusion bodies (Or elementary bodies) are nuclear or cytoplasmic aggregates of stable substances, usually proteins.

Passerine A perching bird of the order Passeriformes.

Photolyase Enzymes that repair DNA damage caused by exposure to ultraviolet light.

Psittacine A bird of the parrot family.

Scab Skin lesion.

Viroplasm (Or virus factories) cytoplasmic inclusion body where viral replication and assembly occurs.

Introduction

Avipoxviruses (APV), of which fowlpox virus (FWPV) is the prototypic member, have received considerable attention due to their devastating effects to populations of domesticated, pet and wild birds, and due to their narrow host range, which allows them to be used as safe, non-replicating vectors for vaccination of birds and mammals. Despite their impact and increased frequency in wild birds, our understanding of the APV molecular and biological characteristics is restricted to FWPV and canarypox virus. Only few APV genomes are currently available and the viruses’ global distribution, phylogenetics and epidemiology are not as well understood as those of *Orthopoxviruses*.

Classification

Avian pox viruses or avipox viruses (APV) form one genus of the Chordopoxvirinae subfamily. APV infect and cause diseases specifically in birds and have been isolated from more than 230 avian species worldwide. Traditionally, avipoxviruses have been assigned to species on the basis of original host, growth and morphological characteristics in the chorioallantoic membranes of embryonated eggs or cell cultures, and/or on the basis of clinical manifestations. Moreover, restriction enzyme and single-gene PCR (e.g., *P4b* gene) analyses have been used to detect and differentiate APV. Genome DNA sequencing has confirmed most provisional species demarcations, although some strains are now regarded as variants of other APV. According to the ICTV, ten APV species have so far been identified and classified. These include *Fowlpox virus* (FWPV), *Canarypox virus* (CNPV), *Juncopox virus* (JNPV), *Mynahpox virus* (MYPV), *Pigeonpox virus* (PGPV), *Psittacinepox virus* (PGPV), *Quailpox virus* (QUPV), *Sparrowpox virus* (SRPV), *Starlingpox virus* (SLPV), and *Turkeypox virus* (TKPV). Other viruses which may be members of the genus but have not yet been approved a separate species include *Crowpox virus* (CRPV), *Peacockpox virus* (PLPV) and *Penguinpox virus* (PEPV). APV fall into three major clades of phylogenetically related viruses. The first one (Clade A, 7 subclades) is related to FWPV which is the prototypic member of the genus, and the causal agent of a widespread, enzootic disease of domestic chickens and other gallinaceous birds. The second one (Clade B, 3 subclades) is related to CNPV which causes high mortality rates to songbirds such as canaries and magpies and the third clade (Clade C) comprises the psittacine PV.

Virion Structure

Since the recognition of APV and orthopoxviruses as the causative agents of disease in the early decades of the nineteenth century, scientists have striven to elucidate their structure. Nevertheless, it has only recently become possible to acquire data on the precise location of poxvirus structural proteins at subviral resolution by using super-resolution microscopy and computational tools. Most

Table 1 List of complete *Avipoxviruses* genomes

Strain	Organism	Host	Isolation source	GenBank Sequence Accession	Size (Kbp)	Open reading frames
ATCC VR – 111	<i>Canarypox virus</i>	<i>Serinus canaria domestica</i> (Canary)	USA, 1948	NC_005309	359	328
FGPV FGPVKD09	<i>Flamingopox virus</i>	<i>Phoenicoparrus minor</i> (Lesser flamingo)	South Africa, 2008	NC_036582	293	256
HP1 – 438 FP9	<i>Fowlpox virus</i>	isolate HP – 438/Munich, passage 438, clone FP9	UK, 2003	AJ581527	266	241
FPVUS	<i>Fowlpox virus</i>	Iowa strain	USA, 2000	NC_002188	288	261
PEPVSPan92	<i>Penguinpox virus</i>	<i>Spheniscus demersus</i> (African penguin)	South Africa, 1992	NC_024446	306	242
PGPV FeP2	<i>Pigeonpox virus</i>	<i>Columba livia</i> (Feral pigeon)	South Africa, 2014	NC_024447	282	224
SWPV – 1	<i>Shearwaterpox virus</i>	<i>Ardenna carneipes</i> (Flesh-footed Shearwater)	Australia, 2016	KX857216	326	310
SWPV – 2	<i>Shearwaterpox virus</i>	<i>Ardenna pacificus</i> (Wedge-tailed Shearwater)	Australia, 2015	KX857215	351	312
TKPV HU1124	<i>Turkeypox virus</i>	<i>Meleagris gallopavo</i> (Wild turkey)	Hungary, 2011	NC_028238	188	171

structural studies in poxviruses (PV) have been focused on the archaetypal poxvirus vaccinia virus (VACV), which is the vaccine strain that eradicated smallpox (variola virus).

In common with the other PV, APV are large viruses (about $330 \times 280 \times 200$ nm) that can just be seen by light microscopy. Electron microscopy revealed that APV display the characteristic brick or ovoid morphology with a biconcave viral core. The viral core contains the linear dsDNA genome and enzymes for the virus uncoating and genome replication. It is enclosed by a lipoprotein envelope and flanked by two protein structures called the lateral bodies. All these structures are contained within an outer coat composed of randomly arranged surface tubules. The role of lateral bodies is not entirely clear and until recently it was proposed to be merely structural. Recent studies in VACV (and FWPV; unpublished) suggest that lateral bodies pack immune modulatory proteins, which can be delivered (upon virus uncoating) into the cytoplasm of target cells. They can therefore serve as facilitators of an early innate immune evasion strategy.

Genome

APV have a single, unsegmented, linear molecule of double-stranded DNA (260–365 kbp) and encode on average more than 230 proteins. Their genome has a low GC-content (30%–40%) compared to >60% for parapoxviruses and molluscipoxviruses and 20% for entomopoxviruses. Similar to all PV, the central region of their genome is flanked by two identical inverted terminal repeats covalently linked by hairpin loops. The central region contains nearly 100 relatively conserved (among *Chordopoxviruses*) genes, which encode viral DNA replication, transcription and RNA modification, and virion formation. The terminally located genes are more prone to mutation and recombination, are thus more variable and encode proteins typically involved in host range restriction, immunomodulation and pathogenesis.

Although APV have evolved to infect a wide range of avian species, to date only few complete APV genomes are available (Table 1); a pathogenic strain of *Fowlpox virus* (FPVUS), an attenuated strain of *Fowlpox virus* (FP9), a virulent *Canarypox virus* (VR-111), a pathogenic South African strain of *Pigeonpox virus* (FeP2), a *Penguinpox virus* (PSan92), a pathogenic strain of *Turkeypox virus* (HU1124), a *Flamingopox virus* (FGPVKD09) and two novel avipoxviruses from Pacific shearwaters (*Ardenna* spp; SWPV-1 and SWPV-2).

A comparison of the available APV genomes revealed overall synteny in their genome arrangement. There are however significant differences (i.e., gene duplications, translocations etc) both in the terminal and the internal regions in contrast to the high levels of conservation of central genomic regions in orthopoxviruses. Comparison of the plaque-purified tissue culture-adapted attenuated European strain FP9 with the US virulent FWPV strain revealed 118 differences, of which 71 genes were affected by deletion (26 of 1–9334 bp), insertion (15 of 1–108 bp), substitution, termination or frame-shifts. Only 110 (42%) of FWPV genes share significant similarity to those in other PV. The CNPV genome is about 80–100 Kbp larger than FWPV, with 39 genes lacking FWPV homologs and approximately 47% amino-acid divergence primarily in the terminal variable regions. The TKPV genome is considerably smaller than the other APV genomes. FeP2 and PEPV are more closely related to FWPV but the presence of whole or disrupted genes that are absent in FWPV but similar to CNPV sequences, suggests that they originate from a common ancestor of CNPV and FWPV. The DNA sequences of SWPV-1 and SWPV-2 are significantly different from each other but nevertheless have good similarity with CNPV (67% and 98% respectively). Genomic profiles of mynahpox, and quailpox viruses show marked differences from FWPV. Compared to mammalian PV, APV show considerable genome reorganization. The most closely related mammalian PV are the molluscipoxviruses and parapoxviruses with which APV share key aspects of gene organization as well as several aspects of immunomodulation and pathogenesis.

Many field FWPV strains carry an active copy of the retrovirus reticuloendotheliosis virus (REV). REVs are closely related to mammalian retroviruses and have also been reported in the gallid herpesvirus 2 (GHV-2) that causes Marek's disease in chickens

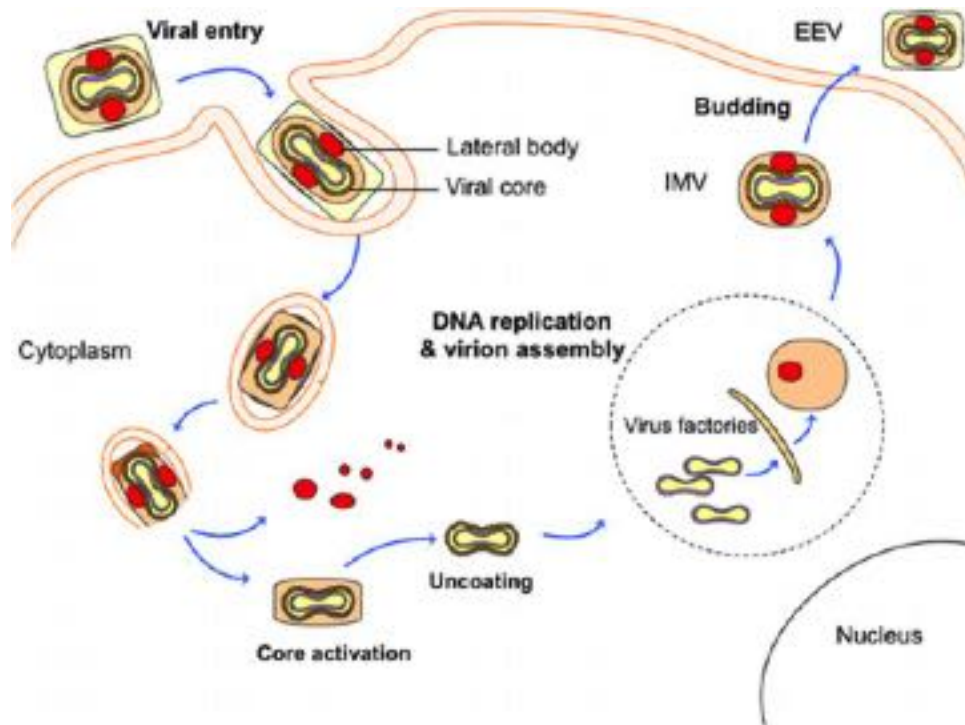


Fig. 1 The replication cycle of APV. Certain elements of the schematic are from Motifolio.

and other galliform birds. The origin of REV insertion into the APV genome is a point of conjecture. Current theories propose that REV endogenization is either due to an ancestral insertion event following natural infection or that it has instead a iatrogenic origin. The REV insertion may account for the strains' increased virulence and fitness. Certainly, REV has been considered one of the most important vaccine contaminants as it can interfere with vaccine-induced immunity. Several vaccine strains (i.e., FPV-S) have already been withdrawn from the market over fears that they are contaminated with REV. Nevertheless, most passaged laboratory and commercial vaccine strains appear to have lost most of the provirus, sometimes leaving only a single long terminal repeat sequence. In these cases the inserted REV sequences cannot give origin to a replication competent virus and therefore should not pose risks for the vaccinated animals.

Life Cycle

The life cycle of PV, illustrated in [Fig. 1](#), is a highly conserved (among *Chordopoxviruses*) and complex sequence of events that commences with the binding of PV onto the cell surface and ends with the release of virions from the cells. APV replicate exclusively in the host cell cytoplasm and have two distinct infectious virus particle types: the unenveloped mature virus (MV) or intracellular MV (IMV) and the extracellular enveloped virus (EEV). The MV and EEV virions differ in their surface glycoproteins and number of surrounding membranes and can both initiate infection.

Viral entry involves either macropinocytosis followed by fusion of the viral envelope and the endocytic vesicle membrane or direct fusion of the virion with the plasma membrane. Upon fusion, the viral cores containing the viral DNA, viral transcription factors, and RNA polymerases pre-bound to early promoters, are liberated into the host cell cytoplasm and undergo a morphological modification termed activation. Core activation includes the switching of the virion shape from biconcave to oval, the uncoupling of the lateral bodies and the initiation of early viral gene expression under the control of viral early promoters. This step is followed by the rupture of the core structure, which releases the viral DNA into the cytoplasm. DNA replication is initiated immediately and is followed by two waves of transcription (intermediate and late), which are mediated by host-derived transcription factors. The newly synthesized viral DNA serves as a template for the subsequent cycles of genome replication and for the transcription of viral genes. Replication and virion assembly take place in cytoplasmic structures called virus factories or viroosomes.

APV generally undergo the various morphogenic stages similar to VACV but with some notable differences. Electron microscopy studies in FWPV showed that the first viral structures are crescent-shaped structures, formed in the virus factories, that begin to incorporate viral DNA. These structures develop into spherical immature virions (IV), which subsequently mature into the characteristic infectious brick- or ovoid-shaped IMVs. The two-lipid bilayer membranes of IMVs derive from the compartment between the endoplasmic reticulum and the *cis*-Golgi cisternae. In VACV, IMVs can either remain in the cells until cell lysis, bud through the plasma membrane or acquire an additional bilayer membrane by a process called "wrapping". The VACV intracellular enveloped virions (IEV), formed after wrapping, migrate by actin tails to the cell surface where the external membrane of the virus

fuses with the plasma membrane, releasing EEV. In contrast to VACV, electron microscopy and proteomic studies have demonstrated that FWPV, CNPV and PGPV lack some of the key genes involved in the wrapping step and the formation of actin tails. FWPV and CNPV appear to exit cells predominantly by budding of the IMV through the plasma membrane to form EEV.

Epidemiology

The epidemiology of the APV is still unclear, due to a lack of studies, the scarcity of genomic data and consequent inability to accurately assign viral isolates into species. APV appear fairly host restricted, but the picture is complicated because many atypical species-species transmissions can occur e.g., prey to predator transmissions, close proximity of diverse animals in zoos and animal facilities etc. APV are considered a threat to both passerine and non-passerine birds including endangered species. Dramatic losses of passerine birds caused by CNPV-like viruses have been reported worldwide in wildlife ecosystems such as in Hawaii, the Galapagos, Australia and New Zealand. Hosts included Andean condors, Australian magpie, Currawongs, Hawaiian crows and geese, palila, apapane and silvereyes. Since the 1950s, APV have been reported in a number of bird species including blackbirds, carrion crows, chaffinches, doves, dunnocks, goldfinches, greenfinches, starlings, jackdaws, pigeons, and sparrows. Since 2005, there has been a marked increase in reports of APV in great tits, with the disease confirmed in several central and northern European countries including the UK. Blue and coal tit populations have been less affected. APV outbreaks have been associated with wildlife translocations or human interventions. However, without long-term data on the (sero)prevalence of APV in wild birds, it is difficult to trace the infection sources and assess the associated risks for wild life.

APV are transmitted via biting insects (e.g., mosquitos, mites, flies) and/or aerosols and invade birds through broken skin. The transmission potential of APV has been suggested to be correlated to the range of insect vector populations. In the face of climate change, seasonal and geographical shifts of insect populations may have broad repercussions in future outbreaks of APV. FWPV (and most likely all APV) is resistant and can survive in the environment for extended periods of time in desiccated scabs shed from infected hosts. Resistance is due to embedment of viruses within large cytoplasmic inclusions and production of photolyase which protect virions following cell lysis and during their environmental exposure.

Pathogenesis

Studies in domestic poultry with FWPV and in canaries with CNPV have been responsible for most of the information currently available in APV pathogenesis. There is evidence that the effects of APV can vary in different avian species. The generalized pathogenicity model is that APVs replicate at the site of inoculation and cause dermal hyperplasia (pock lesions) and leukocyte infiltration. APV infection is typically found in one of two forms: the cutaneous and the diphtheritic form. Both forms can occur at the same time. A third rare systemic form (generalized) can also occur occasionally in canaries, causing primarily depression and anorexia. The cutaneous form of the disease ('dry' pox) is relatively mild and usually characterized by nodular, proliferative growths that are most commonly restricted to the eyes, beak or unfeathered skin of the body. Atypical forms of dry pox can be found in the feathered skin of the birds i.e., feather folliculitis. Infection is slow-developing and can become persistent and extensive although most often is self-limited. With few exceptions, dry pox only causes a drop in productive performance of birds. The nodular lesions are rarely fatal but may restrict the birds' vision and ability to hunt/forage thus making them more susceptible to predation and secondary infections. In canaries the cutaneous lesions can be accompanied by often severe inflammation mainly in the periocular space that may be complicated by secondary infections.

Inhalation or ingestion of aerosols from the dried scabs can lead to more severe infection the so-called "diphtheritic infections" ("wet pox"). This is characterized by fibronecrotic lesions on the mucous membranes of the respiratory and digestive tracts. These lesions resemble those of the avian herpesvirus infectious laryngotracheitis virus. Symptoms of wet pox also resemble those of other avian viruses that cause upper respiratory disease such as avian influenza, or Newcastle Disease virus, thus complicating initial diagnosis. Wet pox causes up to 15% mortality in chicken flocks by occlusion of the larynx and/or secondary bacterial/fungal infections. Upon histological examination of FWPV-infected tissues, variably sized and often large, eosinophilic intracytoplasmic inclusions (Bollinger bodies) can be observed with elementary bodies called Borrel's bodies. Other confirmatory techniques for laboratory diagnosis of FWPV include PCR, electron microscopy and viral culture.

The mortality associated with FWPV disease in poultry is usually low, but in flocks under stress, mortality can reach up to 50%. CNPV is generally associated with higher mortality rates (approaching 100%) often without the characteristic skin lesions.

Host range

FWPV affects mainly chickens and turkeys and has been reported to infect ducks, geese, pheasants, quail, canaries and hawks. Mammals are not susceptible to natural infection with FWPV (or any of the other APV). Nelson (1941) reported mild pathology in mice following intranasal inoculation with FWPV, with no virus replication. Atypical infections have also been reported (i.e., rhinoceros). More recent (albeit limited) studies have shown the presence of infectious viral particles and CPE of FWPV in mammalian cell cultures, such as embryonic bovine tracheal cells and baby hamster kidney cells. These studies challenge the general dogma that FWPV cannot undergo a full replication cycle in mammalian cells. Unlike the FWPV-like, the CNPV-like APV appear able to infect a wider range of avian species. CNPV causes mild infection to its most likely natural hosts, the native songbirds of temperate climes, and more severe infections in non-native canaries.

The target cell determinants of PV binding are thought to be ubiquitously expressed glycosaminoglycans and no specific cell receptors are yet known to be required for virion fusion and cell entry. Host range restriction of PV is attributed to post-entry events associated with antiviral mechanisms such as the type I IFN response or T-cell responses.

Immunomodulation

PV have developed strategies to prevent or evade early host innate responses induced as a consequence of the viruses' binding/fusing with the cell membrane and their progressive migration through diverse cell types and tissues. APV encode many gene homologs with possible host range functions such as chemokine-like molecules (e.g., fpv061, fpv121), putative chemokine receptors (e.g., fpv021, fpv206) as well as inhibitors and homologs of genes involved in apoptosis, autophagy, cell growth and various signaling cascades. Both FWPV and CNPV encode mimics of transforming growth factor (TGF)- β , and CNPV encodes an IL-10-like protein. These proteins are predicted to stimulate regulatory T-cells and suppress the host inflammatory responses. Viral mimicry of cytokines, chemokines and their receptors is well documented in orthopoxviruses. Whether APV use similar mechanisms to evade detection and destruction by the host immune system remains to be confirmed upon availability of chicken reagents.

Our understanding of the strategies deployed by APV to disarm the interferon (IFN) response, the first line of host defense against invading viruses, remains rudimentary and is limited to few FWPV immunomodulatory proteins. FWPV and CNPV (unpublished), unlike other avian viruses (e.g., infectious bursal disease virus) do not induce type I IFN response in chicken embryo fibroblast culture. Moreover, both the wild-type pathogenic FWPV and the attenuated vaccine strain FP9 can block the induction of expression of the chicken interferon β (ChIFN2) promoter normally seen upon transfection with the dsRNA analog, poly(I:C). A broad-scale loss-of-function genetic screen involving a library of 48 FWPV in vitro-generated mutants, each defective in a single, nonessential gene, identified *fpv012* as a modulator of poly(I:C)-mediated ChIFN2 induction. *fpv012* is a member of a family of genes that are highly expanded in the APV (31 in FWPV; 51 in CNPV), encoding proteins containing N-terminal ankyrin repeats (ANKs) and C-terminal F-box-like motifs (ANK-PRANC proteins). Under ectopic expression, the first ANK of *fpv012* is dispensable for inhibitory activity and the CNPV ortholog is also able to inhibit induction of ChIFN2. Conversely, a gain-of-function approach in which 4–8 kbp fragments of the FWPV FP9 strain were introduced into the vaccine strain modified vaccinia Ankara, identified that another ANK-PRANC protein (FPV014) contributes to increased resistance to exogenous recombinant chicken IFN- α .

Control & treatment

FWPV infections in chickens and turkeys are generally well controlled with a standard vaccination program. The timing of vaccinations is dictated by the number of affected birds in a flock (when <20% of the birds have lesions), the onset of disease, the flock size and the type of commercial operation. Modified live FWPV, and PGPV vaccine strains of cell-culture origin are commercially available. The vaccines are usually applied by the wing-web stick method (or thigh-stick methods for turkeys), or injection between the age of 10–14 weeks of age. Birds can be vaccinated at any age if necessary. In some areas where mosquito populations are present throughout the year two vaccinations per year are used. For CNPV infections in pet canary birds and other passerine birds, a live attenuated CNPV vaccine is considered the best preventive measure. Vaccination could also play a role in the conservation management of other endangered avian species that are significantly affected by APV disease. Transmission prevention measures include an insect control program, the improvement of farm hygiene and animal husbandry practices (e.g., dust control, disinfection, ventilation, prevention of fighting of male birds in bird housing, quarantine procedures).

Currently there is no treatment available and diseased birds are treated symptomatically i.e., by washing and sterilizing infected areas, removal of necrotic tissues from mouth and throat of infected birds etc. In the event of an outbreak, disinfecting water lines with a liquid iodine disinfectant, appears to reduce mortality. Broad-spectrum antibiotics and vitamin supplements are beneficial for the birds if there is evidence of secondary bacterial infection.

Use of APV as recombinant vaccine vectors

There is no known zoonotic risk associated with APV. APV cause a productive infection in avian hosts and a non-productive or abortive infection in mammalian hosts. This feature allowed the development of host-restricted APV vectors as safe platforms for vaccines or gene delivery including CNPV (ALVAC) and FWPV (TROVAC). Other desirable features that make APV ideal vaccine vectors include their large genome size which supports a large coding capacity, their ability to readily enter most vertebrate cells to express antigens and their ability to modulate the host innate and adaptive immune responses. Numerous attenuated, live APV-based vaccine strains have been developed since the 1920s against diverse pathogens including avian influenza, Newcastle disease, rabies, cancer, some of which are currently tested or in use in both veterinary and human medicine. A landmark vaccination campaign using for the first time recombinant FWPV was employed successfully in Mexico in the mid-1990s to control high pathogenicity avian influenza H5N2 with almost a billion doses used. Since then, recombinant FWPV have been used in large-scale vaccinations in China, Southeast Asia and elsewhere.

The subject of APV vaccine-vector development is beyond the scope of this article and has been comprehensively reviewed by others. As more information is gathered on APV virulence, immune evasion and host range, it is likely that this knowledge will facilitate the engineering of safer and more efficient vectors for vaccines and gene therapy.

Further Reading

- Bidgood, S.R., Mercer, J., 2015. Cloak and dagger: Alternative immune evasion and modulation strategies of poxviruses. *Viruses* 7, 4800–4825.
- Bolte, A.L., Meurer, J., Kaleta, E.F., 1999. Avian host spectrum of avipoxviruses. *Avian Pathology* 28, 415–432.
- Boulanger, D., Smith, T., Skinner, M.A., 2000. Morphogenesis and release of fowlpox virus. *Journal of General Virology* 81, 675–687.
- Boyle, D.B., 2007. Genus Avipoxvirus. Poxviruses. In: Mercer, A., Schmidt, A., Weber, O. (Eds.), *Birkhäuser Advances in Infectious Diseases*, first ed. Basel: Birkhäuser Verlag, pp. 217–251. [217–251].
- Giotis, E.S., Skinner, M.A., 2019. Spotlight on avian pathology: Fowlpox virus. *Avian Pathology* 48, 87–90.
- Gubser, C., Hué, S., Kellam, P., Smith, G.L., 2004. Poxvirus genomes: A phylogenetic analysis. *Journal of General Virology* 85, 105–117.
- Gyuranecz, M., Foster, J.T., Dán, Á., *et al.*, 2013. Worldwide phylogenetic relationship of avian poxviruses. *Journal of Virology* 87, 4938–4951.
- Herbert, M.H., Squire, C.J., Mercer, A.A., 2015. Poxviral ankyrin proteins. *Viruses* 7, 709–738.
- McFadden, G., 2005. Poxvirus tropism. *Nature Reviews in Microbiology* 3, 201–213.
- Nelson, J.B., 1941. The behavior of pox viruses in the respiratory tract: IV. The nasal instillation of fowl pox virus in chickens and in mice. *Journal of Experimental Medicine* 74, 203–212.
- Niewiadomska, A.M., Gifford, R.J., 2013. The extraordinary evolutionary history of the reticuloendotheliosis viruses. *Plos Biology* 11, e1001642.
- Pattison, M., McMullin, B., Alexander, D., 2008. *Poultry Diseases*, sixth ed. India: Elsevier, pp. 333–339.
- Schat, K.A., Skinner, M.A., 2014. Chapter 16: Avian immunosuppressive diseases and immuno-evasion. In: Schat, Karel A., Kaspers, Bernd, Kaiser, Pete (Eds.), *Avian Immunology*, second ed. London: Academic Press, pp. 275–297.
- Trupkiewicz, J., Garner, M.M., Juan-Sallés, C., 2018. Chapter 33: Passeriformes, caprimulgiformes, coraciiformes, piciformes, bucerotiformes, and apodiformes. In: Terio, K.A., McAloose, D., St. Leger, J. (Eds.), *Pathology of Wildlife and Zoo Animals*, first ed. London: Academic Press, pp. 799–823.
- van Riper, C., Forrester, D.J., 2007. Avian Pox. In: Thomas, N.J., Hunter, D.B., Atkinson, C.T. (Eds.), *Infectious Diseases of Wild Birds*. Oxford: Wiley Blackwell Publishing.
- Wei, S.C., Tryland, M., 2011. Avipoxviruses: infection biology and their use as vaccine vectors. *Virology Journal* 8, 49.

Relevant Websites

- www.avianvirusresearch.org
AvianVirusResearch.
Bringing avian virologists together.
- <http://www.ictvonline.org>
International Committee on Taxonomy of Viruses (ICTV).
- www.msdsvetmanual.com
Veterinary Manual.
- <https://www.viprbrc.org/>
Virus Pathogen Database and Analysis Resource.
- <http://www.thepoultrysite.com>
The Poultry Site.
Your poultry knowledge hub.
- <http://www.violinet.org>
VIOLIN: Vaccine Investigation and Online Information Network.

Hantaviruses (*Hantaviridae*)

Tarja Sironen and Antti Vaheri, University of Helsinki, Helsinki, Finland

© 2021 Published by Elsevier Ltd.

This is an update of A. Vaheri, Hantaviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00411-8.

Glossary

Coevolution Evolution of a virus together with its reservoir host.

Hantavirus A virus in the genus *Orthohantavirus*.

Reassortant A virus having genome segments of two different viruses.

Historical Introduction

Hantavirus infections are not new to humankind. The first description of a hemorrhagic fever with renal syndrome (HFRS)-like disease can be found in a Chinese medical account written in about AD 960 and the earliest definite description of HFRS comes from Far East Russian clinical records dating back to 1913. During World Wars I and II, HFRS became an important military problem; for example, 'field nephritis' in Flanders during World War I may well have been caused by a hantavirus. In Manchuria in the mid-1930s, 12,000 Japanese soldiers caught the disease and military researchers were investigating the cause of the disease, sometimes using prisoners of war in infection experiments. Finnish and German soldiers encountered an HFRS-like epidemic in Finnish Lapland in 1943–1944. During the Korean conflict in 1950–1953, the disease gained again much attention when about 3000 United Nations troops contracted it – since then known as Korean hemorrhagic fever – with a 5%–10% case–fatality rate. The milder form of HFRS, nephropathia epidemica (NE), common in Fennoscandia (Scandinavia and Finland), was first described by Swedish authors in 1934. However, the infecting agent of HFRS remained unknown until 1976, when the Korean researcher Ho Wang Lee discovered that cryostat-sectioned lungs of striped field mice (*Apodemus agrarius*), trapped near the Hantaan river, contained virus-specific antigen reactive with HFRS-patient sera. By a similar approach, the causative agent of NE was demonstrated in 1980 in bank voles, *Myodes glareolus*, trapped in Puumala, Eastern Finland. Another agent, which identified in urban rats in Seoul, Korea, was also found to cause HFRS. In the early 1990s, a further distinct European human-pathogenic hantavirus was isolated from the yellow-necked mouse, *Apodemus flavicollis*, near the village of Dobrava in Slovenia, and a few years later, a related, less pathogenic variant, Saaremaa virus from striped field mice on Saaremaa Island, Estonia. Thottapalayam virus from an insectivore (*Suncus murinus*, a shrew) in India, Prospect Hill virus from a meadow vole (*Microtus pennsylvanicus*) in USA, and Thailand virus from a bandicoot rat (*Bandicota indica*) had already been isolated in the early 1970s to the 1980s – apparently they all are nonpathogenic for humans.

Human disease had not been known to be caused by a hantavirus in the Americas until a cluster of acute respiratory distress syndrome cases with a high (40%) case–fatality rate in the Four Corners area of Southwestern USA (where Arizona, Utah, Colorado, and New Mexico are contiguous) was recognized in May 1993. Subsequent studies led to the discovery of Sin Nombre virus and other hantaviruses that cause hantavirus cardiopulmonary syndrome (HCPS), all transmitted from sigmodontine rodents indigenous to the New World. An increasing number of pathogenic hantaviruses have been reported in South America, including Andes virus, which can be transmitted from person to person, and which possess high case–fatality rates. While only a few thousand HCPS cases have been reported so far, approximately 50,000 HFRS cases are estimated to occur worldwide annually.

In the era of advanced sequencing methods, the rate of hantavirus discovery has accelerated. A great majority of the novel hantaviruses are carried by soricomorpha hosts, i.e., shrews and moles, and few of them by insectivorous bats. These variants appear to be non-pathogenic to humans. In 2016, the new order Bunyavirales was established, elevating the previous genus Hantavirus to the family *Hantaviridae*. Simultaneously, the species demarcation criteria were updated and the family was divided into 4 subfamilies, 7 genera, and 47 species. The hantaviruses causing human infections belong to the subfamily *Mammantavirinae* and genus *Orthohantavirus*.

Virology

Hantaviruses are enveloped viruses with a tripartite negative-stranded RNA genome (Fig. 1). The 6.4 kb L (large) genome segment encodes a ~250 kDa RNA polymerase, the ~3.6 kb M (medium) segment, two glycoproteins 68–76 kDa Gn and 52–58 kDa Gc (formerly known as G1 and G2), and the ~1.7 kb S (small) segment a 50–54 kDa nucleocapsid protein (N). In addition, the S segment of hantaviruses carried by cricetidae rodents has another overlapping (+1) open reading frame (ORF) named NSs, which has been shown to inhibit the interferon response. Viral messenger RNAs of the members of the *Bunyaviridae* are not polyadenylated and are truncated relative to the genomic RNAs at the 3' termini. Messenger RNAs have 5'-methylated caps and 10–18 nontemplated nucleotides derived from host cell mRNAs. The termini of all three segments are highly conserved and complementary to each other, a feature that has assisted in cloning and discovery of new hantaviruses (Fig. 2).

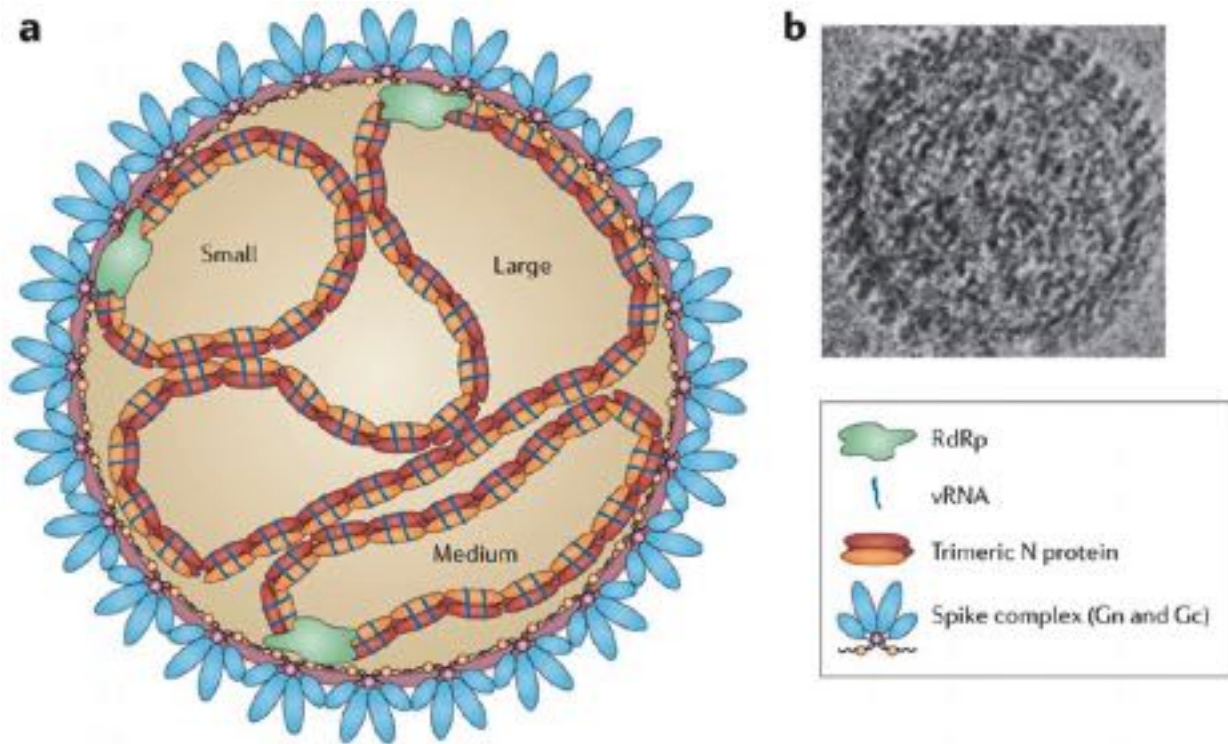


Fig. 1 Structure of an orthohantavirus. (a) The hantavirus particle contains the trisegmented viral RNA (vRNA) genome, comprising the small, medium and large ORFs. These are encapsidated by nucleocapsid (N) protein. The outer part of the virion consists of spikes comprising four units of each glycoprotein, Gn and Gc. The viral genome is replicated and transcribed by RNA-dependent RNA polymerase (RdRp). (b) A hantavirus particle viewed by cryoelectron microscopy. The spike height is invariably 12 nm, and the median diameter of the virion is 135 nm. Part b image is courtesy of P. Laurinmäki and S. Butcher, University of Helsinki, Finland.

Unlike most other bunyaviruses, hantaviruses are not arthropod borne, their reservoirs are rodents, insectivores or bats. Each hantavirus is primarily carried by a distinct reservoir host species, although several host switches seem to have occurred during tens of millions of years of their co-evolution with their carrier hosts. We now know that the genetic diversity of hantaviruses is generated by (1) genetic drift (accumulation of point mutations and insertions/deletions) leading to mixtures of closely related genetic variants, quasispecies; (2) genetic shift (reassortment of genome fragments within a given virus genotype/species); and (3) to some extent, by homologous recombination.

Ecology and Epidemiology

Hantavirus infections are prime examples of emerging and reemerging infections. Like most of these infections, hantaviral diseases are zoonoses. With the exception of the South American Andes virus, hantavirus infections are thought to be transmitted to humans primarily from aerosols of rodent excreta (feces, urine, saliva). Andes virus has also been shown to be transmitted human-to-human. Infections of hantaviruses in the rodents are asymptomatic and persistent, and have only a minor effect on the host. Some hantaviruses cause disease in humans. In Asia, Hantaan virus, carried by *Apodemus* spp. mice, cause severe HFRS, and Seoul virus, carried by rats, causes a milder disease. In Europe, there are two major hantaviral pathogens: Puumala virus carried by bank voles causes NE, while Dobrava virus (DOBV), carried by yellow-necked or striped-field mice, causes severe HFRS. Four different genotypes of DOBV circulate in Europe: the genotypes Dobrava and Sochi cause severe illness, while the genotypes Kurkino and Saaremaa a mild, NE-like disease. There also are reports that Seoul virus (SEOV), carried by rats (*Rattus norvegicus* and *Rattus rattus*), causes HFRS of moderate severity both in Europe and North America. Recent findings suggest that pet rats may also act as reservoirs of SEOV. In addition, European common voles (*Microtus arvalis* and *Microtus rossiaemeridionalis*) carry Tula hantavirus, which can asymptotically infect humans on rare occasions.

Hantavirus infections are quite common in Europe. Puumala virus occurs in Northern Europe, European Russia, and parts of Central and Western Europe. Dobrava genotype of DOBV is found mainly in the Balkans and the neighboring Central European areas. The genotypes Kurkino and Saaremaa have been detected in Eastern and Central Europe, while the genotype Sochi (carried by *Apodemus ponticus*) is circulating in the southern European Russia. Apart from infections of laboratory rats, Seoul virus has been detected in wild rats in France and in several city harbors. Seoul virus infections from pet rats have been increasing lately in Europe, but also in US. It is apparent that many parts of Europe, and most of Africa, remain completely or relatively unstudied with regard to hantaviruses. This suggests either that HFRS is rare or nonexistent in these regions or is not generally recognized and is not diagnosed by local biomedical communities.

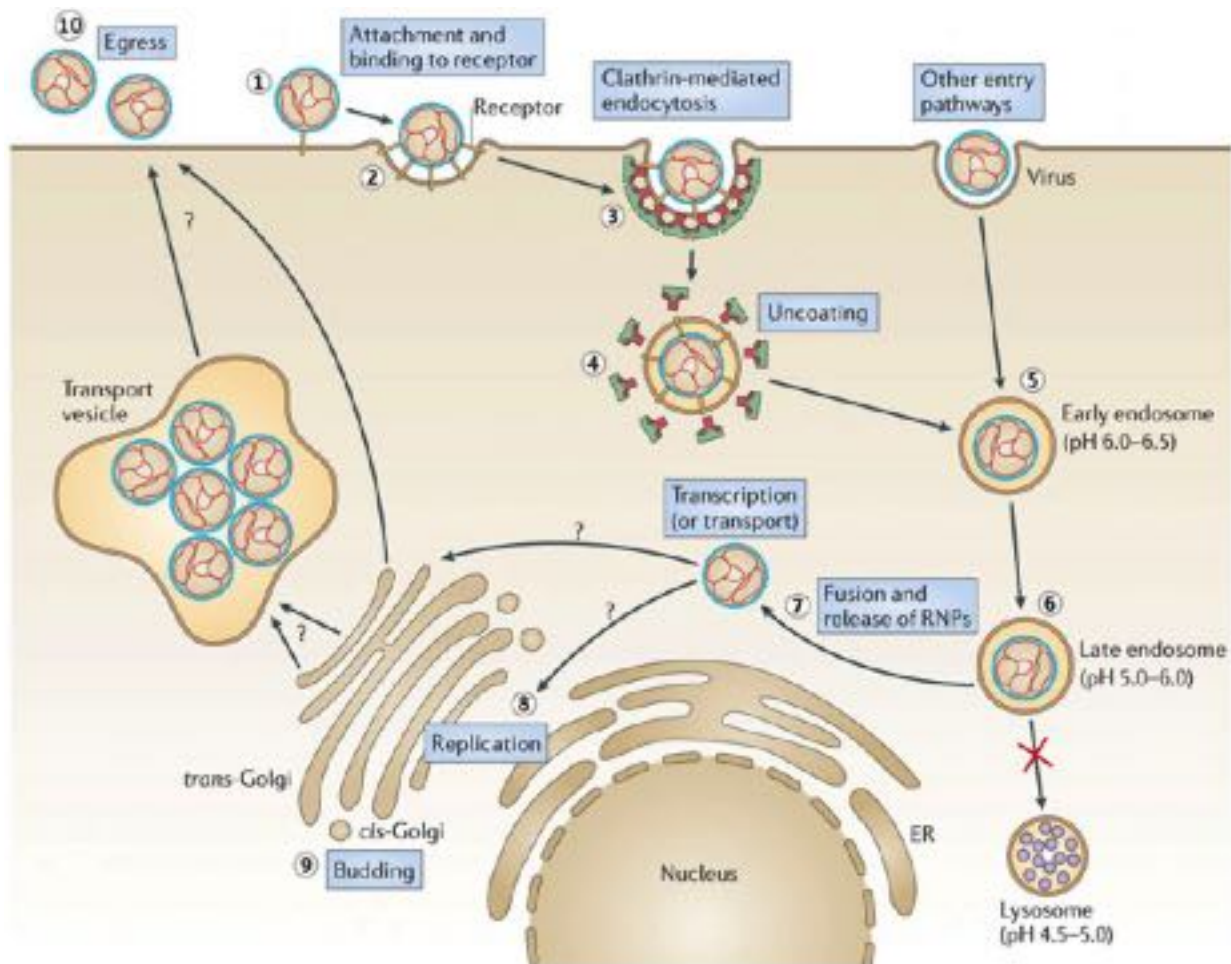


Fig. 2 Life cycle of a hantavirus in an infected cell. The hantavirus virion attaches to a receptor on the cell surface (step 1). This binding event induces endocytosis signaling (step 2), after which the virion enters the cell in clathrin-coated vesicles (step 3). Other entry pathways have also been observed for some hantaviruses. In the case of clathrin-mediated endocytosis, the clathrin coat of the vesicle is disassembled (step 4), and the virion-harboring vesicle enters the early endosome (step 5), which matures into a late endosome (step 6). Fusion between the viral and endosomal membranes is driven by acid-induced conformation changes in the viral fusion protein in the late endosome. This results in release of the viral ribonucleoproteins (RNPs) (step 7). Initial transcription might take place at the site of release; alternatively, the RNPs might be transported to the ER–Golgi intermediate compartment (ERGIC) for transcription. It is also possible that the virus is directly transported to the Golgi complex from the late endosome, either before or after fusion. Viral replication is thought to occur in viral factories that might be located at the ERGIC or the *cis*-Golgi (step 8). The nascent viruses are thought to bud into the *cis*-Golgi (step 9), from where they are transported to the plasma membrane for release, presumably via recycling endosomes. The egress of progeny virions takes place at the plasma membrane (step 10).

In Northern Europe, HFRS as well as the carrier rodents exhibit peaks in 3–4 year cycles, while in Central Europe the HFRS incidence follows the fluctuations of “mast years”, that is, the abundance of beech and oak seeds for the hantavirus-carrying rodents. In Central Europe, HFRS peaks in the summer whereas in Northern Europe most cases occur in late autumn and early winter, from November to January. Risk factors for acquiring hantavirus infections and HFRS include professions such as forestry, farming, and military, or activities such as camping, and the use of summer cottages. Cigarette smokers and males are more likely to be infected than are females. In the Americas, the increased precipitation during El Niño/Southern Oscillation in South and North America has been suggested as the main reason for the peaks in rodent population densities and for the consequent increased number of HCPS cases.

Sigmodontine-borne hantaviruses circulating in North America form three phylogenetically distinct groups: those associated with *Peromyscus* spp. and *Reithrodontomys* spp. rodents, and the third carried by *Sigmodon* spp. and *Oryzomys* spp. rats. The first group carries both human pathogens (Sin Nombre virus) and viruses not thus far associated with human disease (e.g., Blue River virus, Limestone Canyon virus). *Reithrodontomys*-borne viruses have not been shown to cause disease in humans. Hantaviruses discovered in South America are associated with several tribes of Sigmodontinae, mostly Oryzomyini-associated viruses, several of which are important human pathogens, including Andes virus. Results of phylogenetic analyzes of South American hantaviruses suggest that several host-switching events have occurred during coevolution with their rodent hosts. The phylogenetic split

between murine and sigmodontine rodents presumably dates to a divergence that occurred between subfamilies 30 million years ago, when the precursors of the sigmodontine rodents crossed the Bering Strait into the Americas. Notably, sigmodontine rodents are found only in the Americas and thus it is unlikely that HCPS would occur in Eurasia, unless imported.

Clinical Picture and Pathogenesis

Dobrava, Hantaan, Seoul and Puumala viruses all cause HFRS but the infections differ considerably in severity. All are characterized by acute onset, fever, headache, abdominal pains, backache, temporary renal insufficiency (first oliguria, proteinuria, and increase in serum creatinine, and then polyuria), and thrombocytopenia, but the extent of hemorrhages (hematuria, petechiae, internal hemorrhages), requirement for dialysis treatment, hypotension, and case–fatality rates are much higher in HFRS caused by Dobrava, or Hantaan viruses than in NE caused by Puumala or Saaremaa and Kurkino genotype viruses. About a third of NE patients experience temporary visual disturbances (myopia), which is a very characteristic if not pathognomonic sign of the

Table 1 Orthohantavirus types

<i>Virus</i>	<i>Host</i>	<i>Distribution</i>	<i>Disease</i>
MURIDAE-BORNE			
Hantaan (HTNV) ^a	Striped field mouse (<i>Apodemus agrarius</i>)	Asia (Korea)	HFRS
Da Bie Shan	Chinese white-bellied rat (<i>Niviventer confucianus</i>)	Asia (China)	NR
Seoul (SEOV)	Rat (<i>Rattus rattus</i> , <i>R. norvegicus</i>)	Worldwide (Korea)	HFRS
Dobrava-Belgrade (DOBV)	Yellow-necked mouse (<i>A. flavicollis</i>)	Europe (Slovenia)	HFRS
Sangassou (SANGV)	African wood mouse (<i>Hylomyscus simus</i>)	Africa (Guinea)	NR
Thailand (THAIV)	Bandicoot rat (<i>Bandicota indica</i>)	Asia (Thailand)	NR
Tigray (TIGV)	Ethiopian white-footed rat (<i>Stenocephalemys albipes</i>)	Africa (Ethiopia)	NR
CRICETIDAE/ARVICOLINAE-BORNE			
Puumala (PUUV)	Bank vole (<i>Myodes glareolus</i>)	Europe (Finland)	HFRS
Tula (TULV)	European common vole (<i>Microtus arvalis</i>)	Europe (Russia)	NR
Prospect Hill (PHV)	Meadow vole (<i>M. pennsylvanicus</i>)	North America (USA)	NR
Khabarovsk (KHAV)	Reed vole (<i>M. fortis</i>)	Asia (Far East Russia)	NR
Fugong (FUGV)	(<i>Eothenomys eleusis</i>)	Asia (China)	NR
Fusong (FUSV)	Reed vole (<i>Microtus fortis</i>)	Asia (China)	NR
Luxi (LUXV)	Yunnan red-backed vole (<i>Eothenomys miletus</i>)	Asia (China)	NR
CRICETIDAE/SIGMODONTINAE BORNE			
<i>New World mice and rats</i>			
Sin Nombre (SNV)	Deer mouse (<i>Peromyscus maniculatus</i>)	North America (USA)	HCPS
Bayou (BAYV)	Rice rat (<i>Oryzomys palustris</i>)	North America (USA)	HCPS
Black Creek Canal (BCCV)	Hispid cotton rat (<i>Sigmodon hispidus</i>)	North America (USA)	HCPS
Muleshoe (MULV)	Hispid cotton rat (<i>S. hispidus</i>)	North America (USA)	NR
Andes (ANDV)	Long-tailed pygmy rice rat (<i>Oligoryzomys longicaudatus</i>)	South America (Argentina)	HCPS
Choclo (CHOV)	Pygmy rice rat (<i>O. fulvescens</i>)	Central America (Panama)	HCPS
Laguna Negra (LANV)	Vesper mouse (<i>Calomys laucha</i>)	South America (Paraguay)	HCPS
Rio Mamore (RIOMV)	Small-eared pygmy rice rat (<i>O. microtis</i>)	South America (Bolivia)	NR
Caño Delgadito (CADV)	Cane mouse (<i>S. alstoni</i>)	South America (Venezuela)	NR
El Moro Canyon (ELMCV)	Western harvest mouse (<i>Reithrodontomys megalotis</i>)	North America (USA)	NR
Maporal (MAPV)	Delicate pygmy rice rat (<i>Oligoryzomys delicatus</i>)	South America (Venezuela)	HCPS
Montano (MTNV)	Orizaba deer mouse (<i>Peromyscus beatae</i>)	Central America (Mexico)	NR
Necocli (NECV)	<i>Zygodontomys brevicauda</i>	South America (Colombia)	NR
SORICOMORPHA BORNE			
<i>Insectivores</i>			
Asama (ASAV)	Japanese shrew mole (<i>Urotrichus talpoides</i>)	Asia (Japan)	NR
Asikkala (ASIV)	Eurasian Pygmy Shrew (<i>Sorex minutus</i>)	Europe	NR
Bowe (BOWV)	Doucet's musk shrew (<i>Crocidura douceti</i>)	Africa	NR
Bruges virus (BRGV)	European mole (<i>Talpa europaea</i>)	Europe	NR
Cao bang (CBNV)	Chinese mole shrew (<i>Anourosorex squamipes</i>)	Asia (Vietnam)	NR
Jeju (JJUV)	Asian lesser white-toothed shrews (<i>Crocidura shantungensis</i>)	Asia (Korea)	NR
Kenkeme (KKMV)	Flat-skulled shrew (<i>Sorex roboratus</i>)	Asia (Russia)	NR
Oxbow (OCBV)	American shrew mole (<i>Neurotrichus gibbsii</i>)	North America	NR
Seewis (SWSV)	Common shrew (<i>Sorex araneus</i>)	Europe (Switzerland)	NR
Rockport (RKPV)	Eastern mole (<i>Scalopus aquaticus</i>)	North America (USA)	NR
Yakeshi (YKSV)	<i>Sorex isodon</i>	Asia (China)	NR

^aType species.

Abbreviation: HFRS, hemorrhagic fever with renal syndrome; HCPS, hantavirus pulmonary syndrome; NR, disease not recorded.

disease. Notably, the clinical consequences of all of the hantaviral pathogens in humans vary from none to fatal. Severe NE is associated with a certain haplotype, HLA-B8, DR3, DQ2 alleles, severe HCPS with HLA-B35, and mild NE with HLA-B27. Yet, although Puumala virus infection is generally associated with mild HFRS, NE may have significant long-term consequences. A 5-year follow-up study demonstrated that 20% of NE patients had a somewhat increased systolic blood pressure and proteinuria. This is important, since the infection is relatively common in many areas of Europe. In addition, in some patients, Puumala virus infection may infect the pituitary gland and lead to mortality or at least to hypophyseal insufficiency requiring hormone-replacement therapy. Hormonal deficiencies are common in acute NE and 5 years after the infection, 17% of the patients are diagnosed with a chronic, overt pituitary hormone deficiency.

The pathogenesises of HFRS and HPS are not fully understood. However, it is known that β 3 integrins can mediate the entry of pathogenic hantaviruses and that hantaviruses can regulate apoptosis. We and others have studied the pathophysiology of Puumala virus (PUUV) and other hantavirus infections and have come to the conclusion that vascular leakage induced by bradykinin (BK) is a key element in the pathophysiology. Orthohantaviruses predominantly infect microvascular endothelial cells of different organs in humans, and the vascular permeability can be a direct effect of viral infection in these cells or result of an imbalanced immune response. Other important elements of pathophysiology include thrombocytopenia, complement activation, circulating histones, cytotoxic and regulatory T cells (CTLs and T regs), certain cytokines (IL-6 and TNF- α in particular) and neutralizing antibodies. In the novel therapy we have successfully blocked vascular leakage in several very severely ill hantavirus-infected patients by the BK receptor antagonist icatibant. Recent developments of both primate and Syrian hamster models that mimic hantavirus diseases should assist in elucidating the mechanisms of pathogenesis.

Hantavirus disease in the Americas is characterized by pulmonary edema but death often results from a cardiac failure; thus the term hantavirus cardiopulmonary syndrome (HCPS) has been proposed for the disease. HFRS and HCPS, although primarily targeted at kidneys and lungs, respectively, share a number of clinical features, such as capillary leakage, increased serum TNF- α levels, and thrombocytopenia; notably, hemorrhages and alterations in renal function occur also in HPS and pulmonary involvement is not rare in HFRS.

Of the four structural proteins, both in humoral and cellular immunity, the nucleocapsid protein appears to be the principal immunogen. Cytotoxic T-lymphocyte responses are seen in both HFRS and HPS and may be important for both protective immunity and pathogenesis in hantavirus infections.

Diagnostics and Prevention

The diagnosis of an acute hantavirus infection is primarily based on serology. Both immunofluorescence tests and enzyme immunoassays are widely used for the detection of specific IgM or low-avidity IgG antibodies, characteristic for an acute infection. In addition, immunochromatographic 5 min IgM-antibody tests have been developed. Hantaviruses show extensive serological cross-reactivity, especially within each of the three virus subgroups (murid, arvicolid, and sigmodontine borne; [Table 1](#)), but for accurate typing, neutralization tests are needed. For example, in Paraguay, a considerable seroprevalence of hantavirus antibodies, as high as 40%, is noted in people without a history of HCPS. Average seroprevalence in Finland and Sweden suggest that only 10%–25% of Puumala virus infections are diagnosed; thus, most infections are either subclinical or mild or atypical and remain undiagnosed.

Viral RNA can usually be detected in the blood of HCPS and HFRS patients using polymerase chain reaction with samples collected during the first week of illness, which is useful because it also identifies the infecting virus genotype.

Vaccines against hantaviral infections have been used for years in China and Korea, but not in Europe or the Americas. A candidate pan-hantavirus vaccine has been reported to be protective in animal models, but is not yet in clinical trials. No specific antiviral therapy is used in Europe, but both ribavirin and interferon- α have been administered in trials in China. A single dose of icatibant, a bradykinin receptor antagonist, has been successfully used to treat patients with severe HFRS caused by Puumala. A major problem is that at the time the patients are hospitalized, the rate of virus replication is already declining, thus the reduction of virus replication by antiviral drugs is no longer necessary for the patient. Thus, prevention of hantaviral infections continues to rely on reduced contact with excreta from infected rodents.

Further Reading

- Figueiredo, L.T., Souza, W.M., Ferres, M., Enria, D., 2014. Hantaviruses and cardiopulmonary syndrome in South America. *Virus Research* 187. doi:10.1016/j.virusres.2014.01.015.
- Forbes, K.M., Sironen, T., Plyusnin, A., 2018. Hantavirus maintenance and transmission in reservoir host populations. *Current Opinion in Virology* 28, 1–6. doi:10.1016/j.coviro.2017.09.003.
- Hepojoki, J., Strandin, T., Lankinen, H., Vaheri, A., 2012. Hantavirus structure – Molecular interactions behind the scene. *Journal of General Virology* 93, 1631–1644.
- Holmes, E.C., Zhang, Y.Z., 2015. The evolution and emergence of hantaviruses. *Current Opinion in Virology* 10, 27–33. doi:10.1016/j.coviro.2014.12.007.
- Jonsson, C.B., Figueiredo, L.T., Vapalahti, O., 2010. A global perspective on hantavirus ecology, epidemiology, and disease. *Clinical Microbiological Reviews* 23, 412–441.
- Kariwa, H., Yoshimatsu, K., Arikawa, J., 2007. Hantavirus infection in East Asia. *Comparative Immunology, Microbiology and Infectious Diseases* 30, 341–356.
- Laenen, L., Vergote, V., Calisher, C.H., et al., 2019. *Hantaviridae*: Current classification and future perspectives. *Viruses* 11 (9), 788. doi:10.3390/v11090788.
- Mills, J.N., Amman, B.R., Glass, G.E., 2010. Ecology of hantaviruses and their hosts in North America. *Vector Borne and Zoonotic Diseases* 10, 563–574.
- Noack, D., Goeijenbier, M., Reusken, C.B.E.M., Koopmans, M.P.G., Rockx, B.H.G., 2020. Orthohantavirus pathogenesis and cell tropism. *Frontiers in Cellular and Infection Microbiology* 10, 399. doi:10.3389/fcimb.2020.00399.

- Olsson, G.E., Leirs, H., Henttonen, H., 2010. Hantaviruses and their hosts in Europe: Reservoirs here and there, but not everywhere? *Vector Borne and Zoonotic Infections* 10, 549–561.
- Sironen, T., Plyusnin, A., 2011. Genetics and evolution of hantaviruses. In: Plyusnin, A., Elliott, R.M. (Eds.), *Bunyaviridae, Molecular and Cellular Biology*. Norfolk: Caister Academic Press, pp. 61–94.
- Vaheri, A., Vapalahti, O., Plyusnin, A., 2008. How to diagnose hantavirus infections and detect them in rodents and insectivores. *Reviews in Medical Virology* 18, 277–288.
- Vaheri, A., Henttonen, H., Voutilainen, L., *et al.*, 2013a. Hantavirus infections in Europe and their impact on public health. *Reviews in Medical Virology* 23, 35–49.
- Vaheri, A., Strandin, T., Hepojoki, J., *et al.*, 2013b. Uncovering the mysteries of hantavirus infections. *Nature Reviews in Microbiology* 11, 539–550.
- Yanagihara, R., Gu, S.H., Arai, S., Kang, H.J., Song, J.W., 2014. Hantaviruses: Rediscovery and new beginnings. *Virus Research* 187, 6–14. doi:10.1016/j.virusres.2013.12.038.

Henipaviruses (*Paramyxoviridae*)

Lin-Fa Wang and Danielle E Anderson, Duke-NUS Medical School, Singapore, Singapore

© 2021 Elsevier Ltd. All rights reserved.

This is an update of B.T. Eaton, L.-F. Wang, Henipaviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00653-1](https://doi.org/10.1016/B978-012374410-4.00653-1).

Glossary

Biosafety level 4 A biosafety level is a set of biocontainment precautions required to isolate dangerous biological agents in an enclosed laboratory facility. The levels of containment range from the lowest biosafety level 1 (BSL-1) to the highest at level 4 (BSL-4).

Pteropid bats These are among the largest bats in the world taxonomically placed within the genus *Pteropus*. They are commonly known as fruit bats or flying foxes and live in the tropics and subtropics of Asia (including the Indian

subcontinent), Australia, East Africa, and some oceanic islands in the Indian and Pacific Oceans. There are at least 60 extant species in the genus.

Rule-of-six It describes a requirement for paramyxoviruses to have a genome length with a multiple of six nucleotides for efficient virus replication.

Zoonosis A zoonosis is an infectious disease that is transmitted between species from animals other than humans to humans or from humans to other animals.

Introduction

Hendra virus (HeV), the first henipavirus, was isolated in 1994 following an outbreak of acute respiratory disease in horses in a stable in Brisbane, Australia. In a period of less than three weeks, 21 of 30 horses became infected and 14 either died or were euthanized. Two people who worked at the stable also became infected and one died. Since the initial outbreak in 1994, there have been at least 60 recognized occurrences of HeV in Australia, with at least one occurrence per year since 2006.

The second member of the genus, Nipah virus (NiV), emerged in 1998–99 in Perak State, Peninsular Malaysia as the cause of an outbreak of respiratory and encephalitic disease of low morbidity and mortality in pigs. The virus spread to humans causing febrile encephalitis among pig farmers and those who had a direct contact with pigs in Malaysia and Singapore, resulting in more than 100 human casualties.

A highly related variant of NiV emerged in Bangladesh in 2001, and outbreaks of human NiV infections have been detected from that country almost every year since 2001. At least three separate NiV outbreaks have occurred in India, from east Bangladesh-adjacent regions to the Kerala state on the west. In 2015, an outbreak of NiV or closely related virus occurred in the Philippines with lethal infections in both horses and humans.

In addition to the six countries where henipavirus outbreaks have been detected in humans, molecular and serological evidence suggests that henipavirus or henipa-related viruses are present in bats from Australia and Indonesia in the east to Madagascar and west Africa. Three other henipaviruses have been discovered since 2009: Cedar virus was isolated from flying fox in Australia, the full-length genome sequence was determined for the Ghana virus from bats in Ghana and Mojiang virus from rats in China.

Classification

When the full-length genome sequence of HeV was determined, it was clear that it represented a previously unknown member of the family *Paramyxoviridae*, subfamily *Paramyxovirinae*, but one with unique genetic features that precluded classification in the existing three genera *Morbillivirus*, *Respirovirus* or *Rubulavirus* in the subfamily at that time. Complete genome sequencing of NiV revealed a high degree of similarity to HeV and in 2002, the genus *Henipavirus* was created to accommodate these novel paramyxoviruses.

In the latest report from the paramyxovirus study group released in 2019, the genus *Henipavirus* has been expanded to include three new species, Cedar henipavirus (Cedar virus, CedV), Ghanaian bat henipavirus (Ghana virus, GhV) and Mojiang henipavirus (Mòjiāng, MojV). *Henipavirus* is now classified as one of the seven ICTV-approved genera in the family *Paramyxoviridae* (Fig. 1).

Virion Structure

Henipavirus particles are pleomorphic, varying from spherical to filamentous and ranging in size from 40 to 1900 nm. Nucleocapsids, composed of three major structural proteins (nucleocapsid protein, phosphoprotein and a large protein or RNA-dependent RNA polymerase), have a diameter of 18–19 nm with an average pitch of 5 nm. Examination by electron

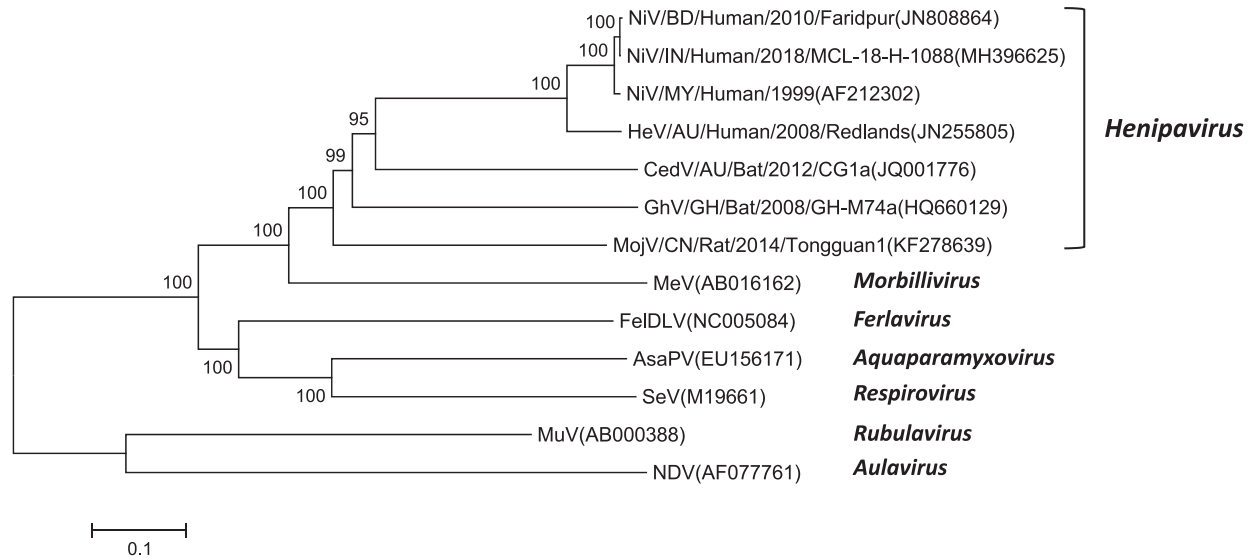


Fig. 1 Phylogenetic tree derived from the RdRP proteins of selected henipaviruses and prototype species of the other genera in the family *Paramyxoviridae*. The tree was constructed by using Neighbor-Joining approach with 1000 bootstrap replicates. Numbers at the nodes represent bootstrap values. Scale bar indicates amino acid substitutions per site. Genbank accession numbers are provided in brackets on the right.

microscopy revealed that HeV has a unique double-fringed appearance, caused by the presence of surface projections 15 ± 1 nm and 8 ± 1 nm in length. In contrast, NiV possesses a single layer of surface projections with an average length of 17 ± 1 nm.

Genome

Several molecular features distinguish henipaviruses from other paramyxoviruses. When the genome length of HeV was revealed at 18,234 nucleotides (nt), it was approximately 15% larger than all known paramyxoviruses at that time. The genomes of the Malaysian (MY, 18,246 nt) and Bangladesh (BD, 18,252 nt) strains of NiV differ by 6 nt because of a 6-nt insertion in the 3' untranslated region of the F gene in NiV-BD. The large genome size of > 18 kb is a conserved feature of all henipaviruses known to date, including the more divergent CedV, GhV, and MojV with genome sizes varying between 18,162 nt (CedV) to ~18,430 nt (GhV).

It is not the additional length per se that is unique to henipaviruses. For example, we know now that the genomes of J virus and Beilong virus in the family *Paramyxoviridae* are over 19,000 nt in length. The novel henipavirus genomic feature is the length of the untranslated region at the 3' end of five of the six genes, which in most cases are from 3 to 13 times longer than their *Paramyxoviridae* counterparts. The function of these regions is unknown. Another unique feature of henipaviruses is the increased size of the P protein, at 707 and 709 aa for HeV and NiV, respectively. This is more than 100 aa larger than most of the known paramyxovirus P proteins.

For all members of the family *Paramyxoviridae*, the genome length can be divided by six, due to the requirement of each N protein in the viral ribonucleoprotein to bind 6 nt residues. This "rule-of-six" is also true for henipaviruses despite their much larger genome sizes. The importance of this has been functionally demonstrated for henipaviruses using minigenome replicon assays.

The genome organization of henipaviruses resembles that of the genera *Respirovirus* and *Morbillivirus*. As for other paramyxoviruses, the first 12 nt of the 3' and 5' genomic terminal sequences of henipaviruses are highly conserved. The first 3 nt of the genome termini are 5'-ACC-3' – a sequence that is fully conserved in the family *Paramyxoviridae* and different from that found in the family *Pneumoviridae*.

There are six coding genes in the henipavirus genome coding for the following major proteins (from 3' to 5' order of the genome): nucleocapsid protein (N), phosphoprotein (P), matrix protein (M), fusion protein (F), receptor binding protein (RBP, previously known as G) and RNA-dependent RNA polymerase (RdRp, also known as L). A comparison of henipavirus genome structure to other well known paramyxoviruses is provided in Fig. 2.

Similar to other paramyxoviruses, the henipavirus P gene undergoes transcriptional editing to produce 5'-co-linear mRNAs which result in the production of V, W and C proteins in addition to the P protein. These additional accessory proteins play a role in antagonizing host innate immune responses. However, the CedV P gene lacks the coding capacity for these additional proteins.

Life Cycle

Henipaviruses can replicate in a variety of mammalian cell lines. The rate of replication, the size of syncytia generated, and the location of nuclei in syncytia vary depending on the cell type and virus species. Infected Vero cell monolayers yield titers in excess

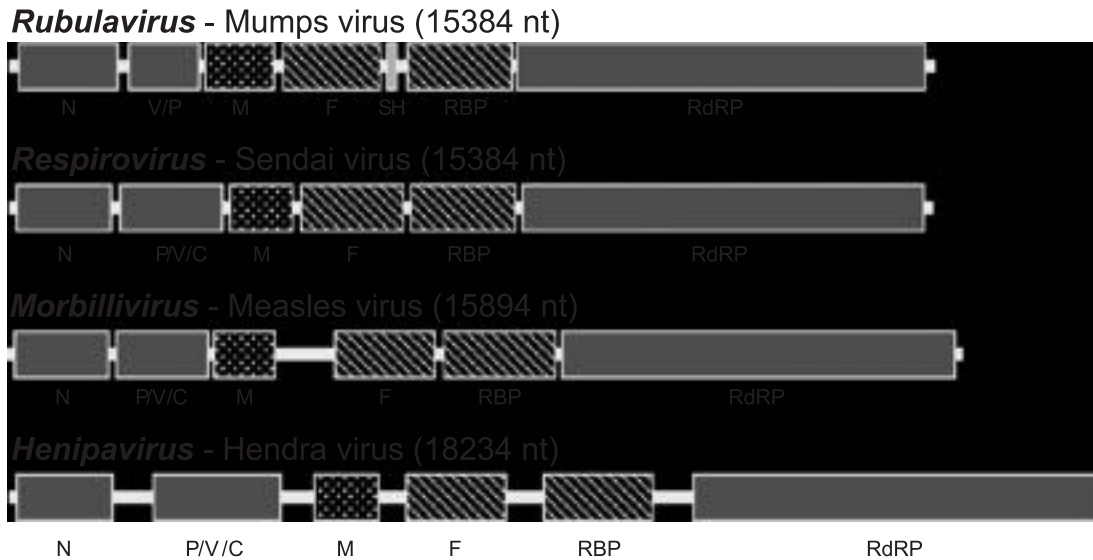


Fig. 2 Genome size and organization of Hendra virus in comparison with the type species of the three other most characterized genera in the family *Paramyxoviridae*. Abbreviations: N = nucleocapsid protein; P = phosphoprotein; M = matrix protein; F = fusion protein; RBP = receptor binding protein (formerly G protein); RdRp = RNA-dependent RNA polymerase.

of 10^8 infectious virions per milliliter. The broad host range of henipaviruses has been demonstrated by the multiple natural susceptible hosts from bats, horses, pigs, humans to cats, dogs and rats. This is further corroborated by experiments with direct infection of cells of different species using live virus or pseudotyped viruses containing the F and G/RBP proteins of henipaviruses, and vaccinia virus-mediated cell surface expression of F and G/RBP glycoproteins. The wide host range of henipaviruses is considered uncommon for paramyxoviruses, which usually adopt a narrow host range.

The ephrin B ligands, functioning in the cell-cell signaling with their cognate EphB receptors, have been identified as the entry receptor for henipaviruses with ephrin-B2 being the main receptor and ephrin-B3 and other homologs as alternative receptor(s). The ephrin proteins are highly conserved among vertebrates, which explains, at least in part, the unusually broad species tropism of henipaviruses. Ephrin-B2 is expressed in arteries, arterioles, and capillaries in multiple organs, including neurons, arterial smooth muscle and bronchiolar epithelial cells. In the brain, ephrin-B3 is expressed at a higher level than ephrin-B2. It should be noted that mice are less susceptible to henipaviruses although the level of ephrin B expression is similar, indicating that other factors, in addition to the receptor molecules, play a role in the overall susceptibility of a mammalian host to henipaviruses.

For susceptible hosts, data obtained from field and laboratory studies indicate an oronasal or oropharyngeal route of entry. The initial site of replication is mostly localized within the respiratory system. Secondary infection can occur in most organs, but with key lesions associated with the respiratory and CNS system, most likely via hematogenous spread of the virus with replication occurring in vascular endothelium. Inflammation of blood vessels (vasculitis) occurs in most organs but is particularly prominent in the brain, lungs, heart, kidneys and spleen.

It is important to note that while human lymphocytes (mainly T and B cells), and monocytes are not permissive for henipavirus infection, it has been shown that they can nevertheless bind and transfer NiV to permissive microvascular endothelial cells. This trans-infection phenotype may play an important role in the hematogenous spread of the virus in addition to the free form that makes up the plasma viremia seen in human patients in the acute phase of the lungs.

Epidemiology

Human henipavirus infections in Australia and the Philippines and in Malaysia occurred as a result of transmission of the viruses from flying foxes via horses and pigs, respectively. In contrast, NiV may have been directly transmitted to humans in Bangladesh (and India, with less direct evidence) where it is believed that date palm juice contaminated with bat secretions constituted a potential transmission route. The mode of transmission from bats to spillover hosts is not known. It has been suggested that horses and pigs may have been infected by HeV and NiV through masticated fruit pulp discarded by flying foxes or by flying fox urine which contaminated pastures or pig sties. Alternatively, the source of virus may have been infected fetal tissues or fluids, a suggestion based on the fact that HeV outbreaks occurred during the birthing period of some species of flying fox, the isolation of HeV from a pregnant flying fox and its fetus, and the transplacental transmission of HeV in experimental infections.

HeV has been transmitted from horses to humans in four occasions. The source of the virus may have been the saliva of infected horses, the nasal discharge commonly found at the terminal stages of the disease, or a wide range of infected tissues made

Table 1 List of countries where infection of Nipah or Nipah-related virus in bats has been detected

<i>Virus isolation</i>	<i>Molecular detection</i>	<i>Serological detection</i>
Bangladesh	Bangladesh	Annobón
Cambodia	Ghana	Bangladesh
Malaysia	India	Bioko
	Thailand	Cambodia
		Cameroon
		China
		Ghana
		India
		Madagascar
		Malaysia
		Malawi
		Príncipe
		Rio Muni
		São Tomé
		Tanzania
		Thailand
		Uganda
		Vietnam
		Zambia

accessible by necropsy. The paucity of virus in the bronchi or bronchioles of infected horses suggests transmission of HeV to either man or horse by aerosol is highly unlikely and experimental infection has confirmed the poor transmissibility of HeV in horses.

Risk factors for human infection by NiV in Malaysia were contact with pigs or fresh pig products and because the virus was present in a wide range of organs, the greatest likelihood arose for those in direct contact with sick or dying pigs on farms during farrowing or slaughtering or in abattoirs. NiV is readily observed in the respiratory epithelium of naturally and experimentally infected pigs, a feature suggesting that the virus probably spread to humans and within the pig population by aerosol or by direct contact with oropharyngeal or nasal secretions. Although NiV was present in the urine and respiratory secretions of patients, human-to-human transmission in Malaysia was extremely rare. In contrast, in Bangladesh, epidemiologic evidence indicates the spread of the virus from one person to another.

Fruit bats (flying foxes) in the genus *Pteropus*, family *Pteropodidae*, suborder Yinpterochiroptera (previously known as Megachiroptera) are the main reservoir hosts of HeV and NiV. In Australia, HeV has been shown to occur in all four flying fox species. In other parts of the world, serologic tests have shown that NiV or NiV-related virus is widely distributed in bats from Indonesia to the west border of Africa. In Australia, HeV has been isolated from at least bats of two different species, *Pteropus alecto* and *Pteropus poliocephalus*. NiV was first isolated from the urine of Island flying foxes and from the saliva on partially eaten fruit and has since been isolated from Lyle's flying foxes (*Pteropus lylei*) in Cambodia and *Pteropus medius* in Bangladesh. In addition to direct isolation, the presence of NiV in different bat populations has also been demonstrated by either molecular or serological detection in the following countries or regions: Annobón, Bangladesh, Bioko, Cambodia, Cameroon, China, Ghana, India, Madagascar, Malaysia, Malawi, Príncipe, Rio Muni, São Tomé, Tanzania, Thailand, Uganda, Vietnam, and Zambia (Table 1).

Genome sequencing revealed that HeV strains isolated from equine and human sources during the outbreak in Brisbane are identical and differ only little from HeV strains isolated from flying foxes 2 years later. Subsequent sequencing of several additional horse and human HeV isolates confirmed the high genetic similarity. Similar observations were made in Malaysia between the human isolates from the outbreak and bat isolates several years later. In Bangladesh, four human isolates obtained in 2004 demonstrate significant genetic heterogeneity, suggesting multiple spillovers of NiV from flying foxes into the human population. The NiV sequences detected from human patients in India from 2001 to 2018 were more related to NiV strains from Bangladesh than the strains isolated in Malaysia.

Clinical Features

Henipaviruses display either predominantly respiratory or neurological symptoms depending on the host. In natural infections of horses and young pigs with HeV and NiV, respectively, and in experimental infections of different animals with either virus, respiratory symptoms predominate. Neurological symptoms were also observed in a proportion of HeV-infected horses. For human infection with henipaviruses, the incubation period ranges from 2 to 45 days, but for 90% of patients, the incubation time is two weeks or shorter. Experimental infection of horses, ferrets and non-human primates is usually fatal, with death or euthanasia occurring 5–10 days post infection. However, observation following natural infection of horses during the initial outbreak indicated that some animals displayed respiratory symptoms but they survived.

HeV-induced respiratory disease in horses may be accompanied by facial swelling and ataxia and in extremis, a copious frothy nasal discharge. The first two patients infected with HeV presented with fever, myalgia, headaches, lethargy, and vertigo. While one recovered, the other patient developed pneumonitis, respiratory failure, renal failure, and arterial thrombosis, and died of cardiac arrest. Findings at autopsy were consistent with a disseminated viral infection in various organs.

Natural infection of pigs with NiV is usually asymptomatic, an outcome also observed after experimental administration of NiV by the ocular and oronasal route. When symptoms are present, they vary according to the age of the pig. Young animals present primarily with respiratory symptoms. Older animals display increased salivation and nasal discharge and occasionally develop neurological signs such as trembling, muscle spasms, and an uncoordinated gait. In Malaysia, pigs displaying symptoms of NiV infection, such as nasal discharge and rapid and labored respiration, had a harsh and nonproductive cough, which gave rise to the name 'barking-pig-disease'.

In contrast, NiV infection in humans is usually associated with a severe acute encephalitis and although a proportion of cases presented with respiratory symptoms, particularly in Bangladesh, the majority of the patients displayed fever, headache, drowsiness, dizziness, myalgia, vomiting, and a reduced level of consciousness. Clinical signs such as the absence of reflexes and the irregular twitching of muscles or parts of muscles and an abnormal doll's eye reflex are indicative of brainstem and upper cervical spinal cord dysfunction. In the Malaysian outbreak, 105 of 256 patients died, a mortality rate of 41%. However, this figure is reduced to 30% when individuals who experienced either a mild or asymptomatic infection are considered. In Bangladesh, a combined case-fatality rate of multiple outbreaks is approximately 70%. For the most recent NiV outbreak in Kerala, India in 2018, the case-fatality rate was 94%.

Both HeV and NiV can cause prolonged infection in humans before manifesting with severe neurological symptoms. In Malaysia, NiV persisted in a proportion of patients (c. 10%) who either recovered from encephalitis or who had experienced an asymptomatic infection. Such cases of relapsed and late-onset encephalitis, respectively, presented from months to years after the initial infection, with the longest reported case 11 years after an initial acute infection. In relapsed cases, although serum anti-NiV IgG antibody levels were elevated, no IgM response or vasculitis was observed, and no live virus was isolated from the throat and nasal secretions. Only four cases of human HeV infection have been recorded. Two displayed influenza-like symptoms and one died. A third, fatal case of encephalitis occurred over a year after a self-limiting episode of meningitis, attributed to HeV infection.

Pathogenesis

As stated above, data from clinical and pathological studies of human cases as well as animal infection and cell tropism experiments all indicate that the initial site of replication is in the respiratory system. The respiratory disease caused by HeV in horses is characterized by pulmonary edema and congestion. Viral antigen is found in endothelial cells in a range of organs including the lungs, lymph nodes, kidneys, spleen, bladder, and meninges, and virus can be recovered from several internal organs, including the lungs, and from saliva and urine. Young pigs infected with NiV present primarily with respiratory symptoms characterized by tracheitis and bronchial and interstitial pneumonia. After experimental infection by the ocular and oronasal routes, virus replication occurs in the oropharynx and spreads to the respiratory tract and lymphoid tissues before appearing in the trigeminal ganglion and neural tissue. Viral antigen is found, particularly in clinically affected animals, in both lungs and meninges and virus is recovered from a range of tissues including tonsil, nasal, and throat swabs and lungs.

The outcome of henipavirus infections can range from high mortality, as seen with HeV in horses and NiV in humans, to low mortality and morbidity, best exemplified by NiV infection in pigs, to asymptomatic infection in flying foxes. Experimental infections reveal that henipaviruses are infectious following either oronasal or parenteral administration. Figures for the minimum lethal dose of henipaviruses suggest that 200–300 plaque-forming units (pfu) is sufficient to produce a productive infection in some animal models.

Transmission is mostly via the oronasal route. Infectious NiV was recovered from urine and the tracheal and nasopharyngeal secretions of infected patients in the early phases of their illness in the original Malaysian outbreak. However, human-to-human transmission in the Malaysian outbreak was extremely limited. In contrast, human-to-human transmission in Bangladesh has been well documented, probably due to more frequent and closer contact with familial caretakers and religious personnel.

As stated above, henipaviruses are uncommon amongst the paramyxoviruses in respect to their broad host range. The ability to replicate, spread, and cause acutely lethal disease amongst multiple species suggest that these viruses, in addition of using a highly conserved molecules as entry receptor, have evolved effective means to antagonize or evade innate antiviral responses of evolutionally distant hosts from bats, horses, pigs to humans. As in other paramyxoviruses, the major anti-interferon (IFN) defences are encoded by the P gene of henipaviruses. It has been shown that the V, W and C proteins produced from the henipavirus P genes all play a role in modulating host immune system by directly interacting with multiple innate immune signaling molecules, including, MDA5, RIG-I, and TRIM25 which regulate IFN gene expression and STATs which are involved in IFN signaling. Interestingly, the NiV M protein has also been shown to inhibit type I IFN induction by indirectly antagonizing the activity of IKK ϵ , one of the kinases responsible for IRF-3 activation. In this context, it is worth noting that CedV lacks the editing site in the P gene and the virus does not produce the V protein and fail to cause any overt disease in ferrets and guinea pigs, in which both HeV and NiV infection can lead to a severe disease.

In susceptible non-pteropid species, henipaviruses elicit strong humoral immune responses. Antibodies to P, N, and M proteins are evident in Western blots and anti-F and anti-G/RBP antibodies are readily detected by various immunological methods. In

contrast to a consistent antibody response in susceptible, nonvolant species, in flying foxes there appears to be less direct correlation between detectable virus replication and the appearance of antiviral antibodies. Antibody production is irregular and of uncertain longevity. Anti-NiV antibodies were present in most patients with clinical NiV encephalitis and IgM antibodies occurred more frequently than IgG antibodies in both serum and cerebrospinal fluid (CSF). The appearance of specific IgM antibodies in serum preceded their appearance in the CSF, a sequence consistent with virus replicating in primary infection site(s) prior to invasion into the central nervous system.

The role played by cellular immunity in henipavirus infection is less known. In the most recent NiV outbreak in Kerala, careful analysis of humoral and cellular immune responses in the 2 survivors revealed that there was a dramatic increase in activated (Ki67⁺) CD8⁺ T cells with an acute effector phenotype (CD38⁺/HLA-DR⁺/Granzyme B⁺/PD1⁺), suggesting that virus-specific naïve T cells were actively proliferating in response to the infection. It is interesting to note that clearance of viremia coincided with the peak of proliferating Ki67⁺ CD8⁺ T cells, which occurred before the peak IgG antibody levels.

Diagnosis

Traditional diagnostic methods such as virus isolation, EM, immuno-EM, immunohistochemistry (IHC), serology, and polymerase chain reaction (PCR) played key roles in the initial discovery of HeV and NiV.

As henipaviruses are BSL4 agents, it is important to note that while initial attempts to culture suspected henipaviruses may be conducted in a BSL3 laboratory, it is essential to continue further characterization inside a BSL4 facility once a cytopathic effect (CPE) is observed and the growth of henipavirus is confirmed by PCR or immune staining. All henipaviruses can replicate in different cell lines – a feature that contributed to the efficiency with which they were isolated during the initial disease outbreak investigations, but Vero cells remains the most commonly used cell line for initial virus isolation trials. More recently, different bat cell lines have been used with some success.

For molecular diagnosis, PCR is still the method of choice. With the discovery of greater genetic diversity of henipaviruses in different parts of the world, and current PCR tests may not work for unknown viruses, next generation sequencing (NGS) is starting to be employed for outbreak investigation with suspected of henipaviruses or henipa-related viruses as well to any previously unknown pathogens. A probe-based NGS containing hybridization baits covering all known henipavirus genomes has recently been developed for more efficient NGS characterization, especially from clinical samples where host genetic materials are expected to give very background signal masking the low level of virus-specific NGS reads.

For henipaviruses, serologic tests are important both during outbreak investigation and for disease surveillance. The virus neutralization test (VNT) is accepted as the standard reference method. Few laboratories, however, can conduct neutralization tests because of the requirement to handle live virus at BSL4. Two surrogate VNT test platforms which do not require BSL4 containment have been developed: (1) Luminex bead-based test: recombinant soluble G/RBP proteins of HeV or NiV are coated on Luminex beads and the binding of a soluble ephrin-B2 protein is used to mimic the virus-receptor interaction. When neutralizing antibodies are present, they will block this G-ephrin-B2 interaction, thus acting as a surrogate VNT; (2) Pseudotype virus: different pseudotype systems, based on Vesicular stomatitis or lentivirus, carrying the henipavirus F and G/RBP proteins have been developed as a surrogate VNT for detecting henipavirus-specific antibodies. Incorporation of reporter genes in these systems have resulted in a greater sensitivity and reproducibility.

Treatment

At the present time, there is no licenced treatment available for henipavirus infections. Ribavirin, which inhibits the replication of HeV *in vitro*, was used during the NiV outbreak in Malaysia in 1998 and HeV outbreak in 2008 in Australia. While the trial in Malaysia demonstrated a 35% reduction ($P = 0.011$) in mortality, it had no detectable therapeutic effect on the two HeV patients in Australia. In the absence of other therapies, ribavirin may be an option for the treatment of henipavirus infections with patient consent, but the efficacy of ribavirin as a therapy or prophylaxis in people remains uncertain at best. In different animal models, ribavirin either delayed but did not prevent the death or had no effect on infection caused by NiV or HeV.

Chloroquine, an antimalarial drug, was shown to block the critical proteolytic processing needed for HeV F maturation and function. Although the drug inhibited NiV and HeV infection *in vitro*, it did not show any apparent clinical benefit *in vivo*.

In contrast to existing small molecule antivirals, a henipavirus-specific therapeutic heptad peptide-based fusion inhibitor, analogous to the HIV-1-specific enfuvirtide (Fuzeon), has shown promising efficacy in animal studies. The α -helical heptad repeat (HR) domain of henipavirus F1 glycoprotein resembles other fusion glycoproteins which play a role the formation of a trimer-of-hairpins structure during the membrane fusion and virus infection process. In a hamster model of NiV infection, the administration of cholesterol tagged HR-derived peptides provided the first evidence of *in vivo* effectiveness of a fusion inhibiting peptide against NiV.

Finally, passive immunotherapy based on recombinant human monoclonal antibodies to the henipavirus G/RBP protein also appear to have a promising therapeutic potential. In particular, the m102.4 mAb is shown to be effective in a non-human primate model against HeV or NiV infection. While human clinical application is yet to be proven, a phase I clinical trial has been completed, demonstrating the safety for human applications and an effective half-life in humans. The mAb has also been used in

humans with high risk exposure to henipaviruses under emergency and compassionate grounds with no severe side effects observed.

Prevention

A number of promising strategies to develop henipavirus vaccines are being explored. NiV F and G/RBP glycoproteins, either delivered via viral vectors or by administration as purified proteins with adjuvant, elicit neutralizing antibodies in many different animal models.

The most studied and tested vaccine is based on the HeV-sG protein produced in CHO cells. The vaccine was licensed by Zoetis™ Inc., (formerly Pfizer Animal Health) and developed as an equine vaccine, under the trade name Equivac® HeV, for use in Australia. This remains to be the first and only licensed vaccine for any BSL4 agent. The vaccine is 100% effective in protecting horses under experimental challenge studies. Since the vaccine release in 2012, laboratory confirmed HeV infections in horses have only occurred in unvaccinated horses.

NiV is one of the first three high priority targets set by the Coalition for Epidemic Preparedness Innovations (CEPI). Currently there are two CEPI-funded NiV vaccine projects ready for phase I clinical trial: the US-led HeV-sG vaccine platform (an extension of Equivac® HeV to NiV infection in humans) and the Japan-led measles viral delivery platform. Human phase I trials will be completed in the next few years and we are confident that a licenced human henipavirus vaccine will be available thereafter.

Further Reading

- Ang, B.S.P., Lim, T.C.C., Wang, L.-F., 2018. Nipah virus infection. *Journal of Clinical Microbiology* 56. doi:10.1128/JCM.01875–17.
- Bean, A., Baker, M., Stewart, C.R., *et al.*, 2013. Studying immunity to zoonotic diseases in the natural host – Keeping it real. *Nature Reviews Immunology* 13, 851–861.
- Broder, C.C., 2012. Henipavirus outbreaks to antivirals: The current status of potential therapeutics. *Current Opinion in Virology* 2, 176–187.
- Broder, C.C., Xu, K., Nikolov, D.B., *et al.*, 2013. A treatment for and vaccine against the deadly Hendra and Nipah viruses. *Antiviral Research* 100, 8–13.
- Eaton, B.T., Broder, C.C., Middleton, D., Wang, L.-F., 2006. Hendra and Nipah viruses: Different and dangerous. *Nature Reviews Microbiology* 4, 23–35.
- Field, H.F., Mackenzie, J.S., Daszak, P., 2007. Henipaviruses: emerging paramyxoviruses associated with fruit bats. *Current Topics in Microbiology and Immunology* 315, 133–160.
- Luby, S.P., Gurley, E.S., 2012. Epidemiology of henipavirus disease in humans. *Current Topics in Microbiology and Immunology* 359, 25–40.
- Mahalingam, S., Herrero, L.J., Playford, G., *et al.*, 2012. Hendra virus: An emerging paramyxovirus in Australia. *Lancet Infectious Diseases* 12, 799–807.
- Pernet, O., Wang, Y.E., Lee, B., 2012. Henipavirus receptor usage and tropism. *Current Topics in Microbiology and Immunology* 359, 59–78.
- Thibault, P.A., Watkinson, R.E., Moreira-Soto, A., Drexler, J.F., Lee, B., 2017. Zoonotic potential of emerging paramyxoviruses: Knowns and unknowns. *Advances in Virus Research* 98, 1–55.
- Wang, L.-F., Anderson, D.E., 2019. Viruses in bats and potential spillover to animals and humans. *Current Opinion in Virology* 34, 79–89.
- Wang, L.-F., Mackenzie, J.S., Broder, C.C., 2013. Henipaviruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed., vol. 2. Philadelphia: Lippincott Williams and Wilkins, pp. 286–313.

Relevant Websites

- <https://cepi.net/>
CEPI.
New Vaccines For A Safer World.
- <http://www.outbreak.gov.au/for-vets-and-scientists/hendra-virus>
Hendra virus.
National pest & disease outbreaks.
- <https://www.who.int/emergencies/diseases/hendra-virus/en/>
Hendra virus infection.
World Health Organization.
- <https://gvn.org/>
Home.
GVN.
- <https://www.who.int/csr/disease/nipah/en/>
Nipah virus infection.
World Health Organization.
- <https://www.business.qld.gov.au/industries/service-industries-professionals/service-industries/veterinary-surgeons/guidelines-hendra/incident-summary>
Summary of Hendra virus incidents in horses.
Business Queensland.

Hepatitis A Virus (*Picornaviridae*)

Andreas Dotzauer, University of Bremen, Bremen, Germany

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Antibody prevalence The percentage of a population with antibodies against a certain disease at a given time.

Aplastic anemia Disease characterized by a decrease in blood cells resulting from underproduction due to bone marrow failure; called also hypoplastic anemia.

COPII Coat protein complex II.

cre Cis-acting replication element.

ERGIG ER-Golgi intermediate compartment.

Exosome Vesicle (50–120 nm) released from cells by fusion of late endosomal multivesicular bodies with the plasma membrane that contain various biologically active proteins and RNAs.

Incidence The rate of occurrence of new cases of a particular disease in a population in a certain period of time.

Leukocytopenia Decrease in the number of white blood cells.

Polyprotein processing Cascade of proteolytic cleavage events resulting in release of biologically active intermediate and finally mature proteins from the polyprotein.

Ribosome stalling Transient translation stop.

Serine-like protease Proteolytic enzyme characterized by a catalytic triad similar to that in serine-type proteases with Ser, His, Asp.

History

Jaundice has been known as an epidemic disease for centuries, but the earliest outbreaks of what was almost certainly hepatitis A were documented by Rayrun "MRWmats_scpger, occurring in Preßburg, now Slovakia, in 1674 and 1697. Besides sporadic occurrence of the disease, characterized by a slow increase and slow decrease of the number of infected persons in the course of months (spread by person-to-person contact), with an overall small number of infections, larger vehement epidemics (spread by contaminated water and food) were reported in later times. The first pandemic HAV wave occurred in the 1860s and the first considerable record of the disease was registered during the American Civil War.

In the second half of the nineteenth century, the disease became known as "icterus catarrhalis" (catarrhal jaundice; inflammation of the biliary tract was supposed) or "icterus epidemicus" (epidemic jaundice) as well as campaign jaundice, as epidemics are common in military medical history.

At the turn of the nineteenth to the twentieth century, the disease was recognized as infectious and transmissible by person-to-person contact. The terms "hepatitis epidemica" (epidemic hepatitis) and "hepatitis infectiosa" (infectious hepatitis) were introduced as synonyms for catarrhal jaundice.

Detailed epidemiologic recordings have been conducted since the beginning of the twentieth century. Two large pandemic waves were observed during the twentieth century, the first one originating during World War I and reaching its summit between 1918 and 1922, and the second one originating in the early 1930s reaching its widest distribution with the beginning of World War II.

During World War II, hepatitis was demonstrated to be caused by at least two separate filterable agents, and the resulting diseases were called hepatitis A (infectious hepatitis) and hepatitis B (serum hepatitis), and the etiological agents hepatitis A virus (HAV) and hepatitis B virus (HBV), respectively. We now know three more viruses that have a specific tropism for the liver: HCV, HDV and HEV. These five viruses share no taxonomic and virological relationship.

After World War II, epidemiologic studies in human volunteers showed that hepatitis A is spread by the fecal–oral route, and provided information on the duration of viremia and shedding of virus in feces. In the late 1960s and early 1970s, it was shown that chimpanzees and several small, non-human primates could serve as animal models for human hepatitis A. Meanwhile, mice genetically deficient for *Ifnar1*, a component of the type 1 interferon receptor, or for the mitochondrial antiviral signaling protein (*Mavs*) or the interferon regulatory factors (*Irf3/Irf7*), which are components of the induction pathway for interferon expression, as well as chimeric mice with human liver cells were found to be infectable by HAV and are also used as models for the disease. The replication of HAV in cell cultures was established between 1979 and 1981.

The etiologic agent was identified through immune electron microscopy by Feinstone *et al.* in 1973.

The molecular cloning of the HAV genome in the 1980s revealed that the genomic organization of HAV is similar to that of picornaviruses, and the first infectious cDNA clone of HAV was reported by Cohen *et al.* in 1987.

The disease manifestations by hepatocellular destruction could be attributed to an immunopathogenic mechanism in the late 1980s by Vallbracht *et al.*

An inactivated HAV vaccine has been available since 1992.

Taxonomy and Evolutionary Origin

HAV is a member of the genus *Hepatovirus* within the family *Picornaviridae*.

All human HAV strains known belong to only one serological group, but nucleotide sequence analysis of the VP1-2A junction region and the VP1 coding region, respectively, revealed that several distinct HAV genotypes (seven in the case of VP1-2A junction analysis, six in the case of VP1 region analysis; more than 15% nucleotide variation), which include several subgenotypes (nucleotide variation of more than 7.5%), can be distinguished. These genotypes correlate with the geographic origin of the virus isolates. Genotypes I and III are the most common types worldwide, particularly genotype IA. The cell culture-adapted viruses most commonly used are variants of the Australian strain HM175 (genotype IB) and the German strain GBM (genotype IA).

Recently, multiple novel species, which have the same unique properties as human HAV, have been identified in bats and other small mammals, seals and even marsupials. Phylogenetic analyses indicate that these viruses have frequently jumped between the host species, and ancestral reconstructions suggest that human HAV descended from rodent viruses, which have originated in small insectivorous mammals. Based on the unique capsid structure of HAV, the HAV-like viruses of ancient insectivores might be direct descendants from a picornavirus-like insect virus.

Morphology and Physicochemical Properties

The infectious spherical virion shed in feces is a nonenveloped particle with a diameter of approx. 27 nm. The pentakisdecagonal capsid, which embodies the viral RNA genome, contains 60 copies of each of the 4 structural proteins, VP1 (also known as 1D), VP2 (1B), VP3 (1C) and VP4 (1A) (see [Table 1](#)). The mature capsid has a different structure than other mammalian picornaviruses, as the canyon surrounding the 5-fold symmetry axes, which is a prominent feature of picornaviruses and represents the attachment site to cellular receptors, is missing, and VP2 extends N-terminally across the 2-fold symmetry axis to the adjacent protomer, a feature seen in picornavirus-like dicistroviruses of insects. HAV forms a single serotype and the antigenic structure of the capsid is conserved among the human and the non-primate viruses. In blood as well as in supernatants of infected cell cultures HAV appears enveloped in exosome-like vesicles (eHAV). These vesicles may enclose up to four infectious virus particles, which still contain uncleaved VP1-2A.

The HAV particle has a buoyant density of approx. 1.33 g cm^{-3} in CsCl (in iodixanol of approx. 1.22 g cm^{-3} ; eHAV particles band at approx. 1.08 g cm^{-3} in iodixanol gradients) and a sedimentation coefficient of approx. 160S. Empty capsids found in feces have a sedimentation constant of about 70S. HAV is extremely resistant to acid (pH 1.0 for 2 h at room temperature) and thermal inactivation (60°C for 1 h). The isoelectric point of HAV is between values of 2.8 and 5.8.

Genome Organization and Expression, Replication, Morphogenesis

The linear, single-stranded, positive-sense RNA genome of HAV is approx. 7500 nucleotides in length and is not capped but covalently linked by a tyrosine- O^4 -phospho-diester bond to the 2.5 kDa viral protein 3B (also known as VPg; see [Table 1](#)). It encompasses a structurally complex 5' nontranslated region (NTR) of approx. 740 bases, followed by a single open reading frame encoding a polyprotein (approx. 250 kDa; approx. 2230 amino acids) and a 3' NTR of approx. 60 bases which terminates with a poly(A) tail of about 60 nucleotides. In the polyprotein, the capsid proteins and those with functions during virion assembly represent the N-terminal third (VP4, VP2, VP3, VP1 (numbered according to molecular mass and also known as P1 region), and 2A) (see [Table 1](#) and [Fig. 1](#)) with the remainder of the polyprotein comprising a series of proteins required for RNA replication (2B and 2C followed by 3A to 3D) (see [Table 1](#) and [Fig. 1](#)).

The viral polyprotein is translated directly from the messenger-sense genomic RNA, which is released into the cellular cytoplasm after uncoating of the virion. An internal ribosomal entry site (IRES) located within the 5' NTR (see [Table 1](#)) is involved in the cap-independent initiation of protein synthesis. The IRES of HAV differs from that of other picornaviruses and forms its own group (type III picornavirus IRES). Translation from the HAV IRES requires all of the initiation proteins, including eIF4E and intact eIF4G, and infection with HAV does not result in cleavage of the translation initiation factor eIF4G to block the cap-dependent host protein synthesis, as featured by other picornaviruses. The HAV IRES-directed initiation of translation, which depends on the entire 5' NTR, is enhanced by sequences of the 5'-terminal coding region and by certain host cell factors, e.g., glyceraldehyd 3-phosphate dehydrogenase, poly(C) binding protein and polypyrimidine tract binding protein. The efficiency of translation is decreased by an extraordinarily high proportion of rare codons (particularly in the P1 coding region). But, as HAV does not induce shutdown of cellular protein synthesis, this not optimal codon usage might avoid strong competition for tRNAs and in addition may support proper protein folding by ribosome stalling. A further characteristic of the genome is a low G/C ratio and the avoidance of CpG dinucleotides.

Proteolytic processing of the primary polyprotein (see [Fig. 1](#)) occurs simultaneously with translation and is largely carried out by the viral 3C protease, which is a serine-like protease in which cysteine replaces the nucleophilic serine in the catalytic triad of the active center. As cellular substrate, PABP (poly(A) binding protein) has been described for mature 3C, obviously resulting in late negative regulation of translation and contributing to the switch to replication. Furthermore, processing intermediates of 3C, in this case 3ABC and possibly 3CD, interfere with the induction of IFN- β transcription by 3C-mediated cleavage of MAVS (mitochondrial antiviral signaling protein), and of TRIF (TIR domain-containing adaptor inducing IFN- β), respectively (see [Fig. 3](#)).

Table 1 HAV regulatory genomic regions and HAV proteins; all numbering refers to HAV strain HM175 (accession no. NC_001489).

<i>Regions/Proteins</i>	<i>Nucleotide position</i>	<i>Length in amino acids</i>	<i>Proposed or known function</i>	<i>Remarks</i>
5' nontranslated region (5' NTR)	1–734	–	5' stem-loop region (nucleotides 1–94): required for RNA replication Polypyrimidine tract pY1 (nucleotides 99–138): unknown function Internal ribosomal entry site (IRES) (nucleotides 152–734): directs cap-independent initiation of translation Involved in regulation of RNA replication	Mechanisms regulating switch from translation to replication are not clear Picornaviral Type III IRES
3' nontranslated region (3' NTR)	7416–7478	–		
1A	735–803	23	Extended by 40–80 adenylates Capsid protein VP4 Morphogenesis Pore-forming activity: may facilitate release of genome from the capsid	Involvement in translation is not clear Not myristoylated
1B	804–1469	222	Capsid protein VP2	During morphogenesis, precursor VP4 - VP2 (VP0) is likely cleaved autocatalytically
1C	1470–2207	246	Capsid protein VP3	
1D	2208–3023/ 3026/3029	272/273/274	Capsid protein VP1	
				Heterogeneous C terminus, depending on HAV strain and host cell
				During morphogenesis, precursor VP1–2A (also known as VP1pX), is likely cleaved by cellular protease So far not identified in infected cells
2A	3024/3027/ 3030–3242	71/72/73	Morphogenesis	
2B	3243–3995	251	Structural rearrangements of intracellular membranes for replication	Peripheral association with membranes Influence on membrane permeability Induces structural alterations of the ERGIC structure Mutations accompany adaptation to growth in cell culture Interferes with induction of beta interferon transcription by an unknown mechanism (affects MAVS and the IRF-3 kinases)
2C	3996–5000	335	Structural rearrangements of intracellular membranes for replication	Integral association with membranes Mutations accompany adaptation to growth in cell culture Interferes with cellular macromolecular synthesis
3A	5001–5222	74	As 3AB VPg precursor(pre-VPg)	Interacts with membranes, may anchor pre-VPg to cellular membranes through a central region of 21 hydrophobic amino acids Induces structural alterations of intracellular membranes Tyr ³ attached to 5' terminus of genomic RNA
3B	5223–5291	23	VPg (virus protein genomic) Supposed protein primer for RNA replication	
3C	5292–5948	219	Sole protease	Serine-like protease, with replacement of Ser by Cys in catalytic triad: Cys ¹⁷² , His ⁴⁴ , Asp ⁸ Involved in switching viral translation to replication by cleavage of PABP Intermediate 3ABC interferes with induction of beta interferon transcription by cleavage of MAVS
3D	5949–7415	489	RNA-dependent RNA polymerase	Interferes with cellular COPII-mediated, classical secretory pathway cre-element consisting of 110 nucleotides with conserved AAACG motif at 5' coding region required for replication: template for VPg (3B) uridylation

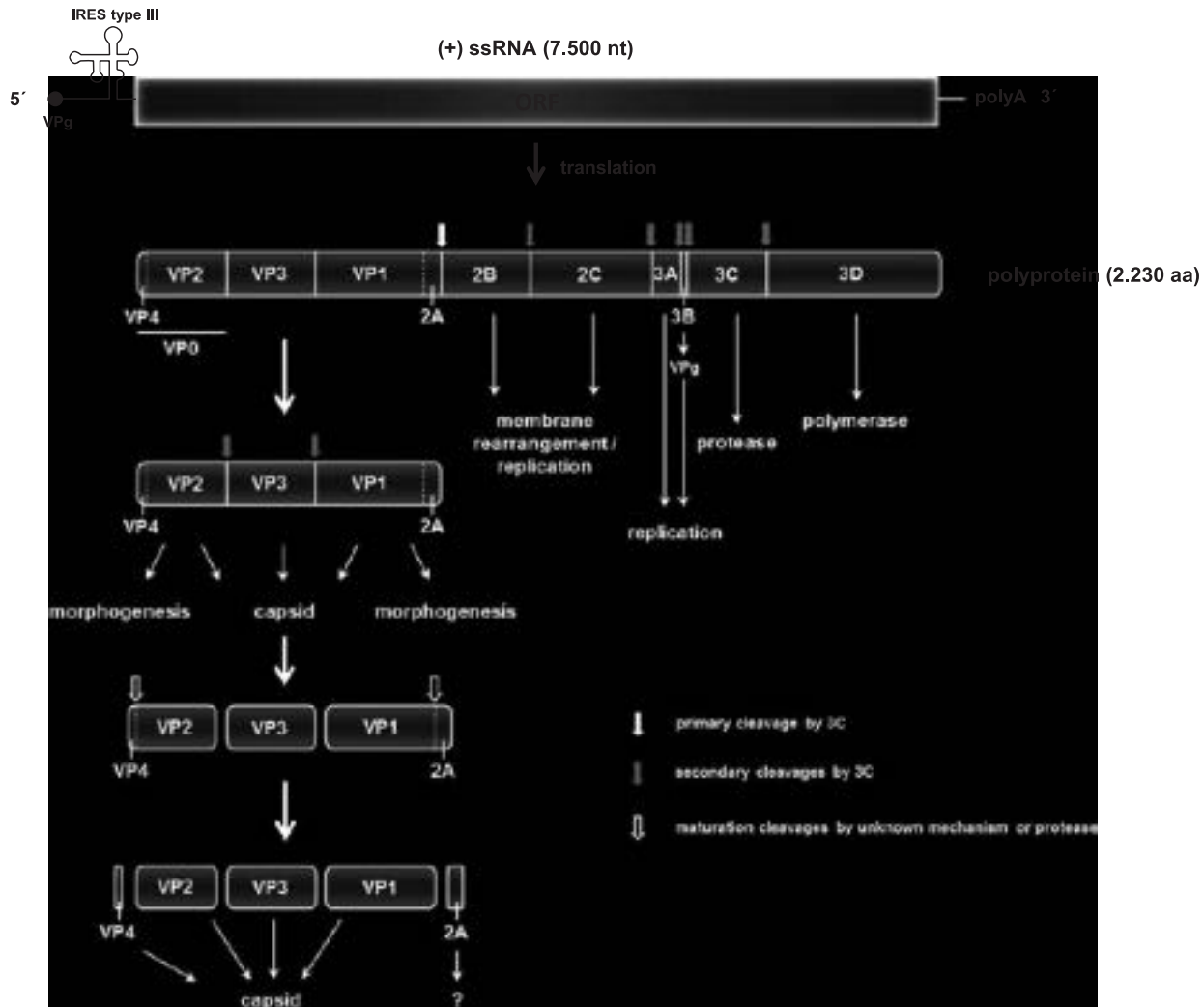


Fig. 1 HAV genomic RNA and HAV polyprotein processing including known or supposed functions of the viral proteins during morphogenesis and replication. The polyprotein that results from IRES-directed translation is cleaved by the viral protease 3C to release certain precursor proteins with biologically relevant functions as well as the mature nonstructural proteins. Primary cleavage results in the release of the structural precursor VP4-VP2-VP3-VP1-2A, which associate to assembly intermediates (pentamers), which are stabilized by 3C cleavage of the VP3 junctions (VP4-VP2 (VP0), VP3, VP1-2A)₅. After assembly of the intermediate protein building blocks and the viral genome to provirions, maturation cleavages occur through an unknown proteolytic activity to release VP4 from VP2 and through a cellular, presumably a lysosomal protease to release 2A from VP1, resulting in the infective virion with 60 copies of each of the main structural proteins VP1 to VP4.

Synthesis of viral RNA by the viral 3D polymerase follows the accumulation of the nonstructural proteins spanning 2B to 3D, which induce the assembly of a macromolecular replication complex on membranes that are recruited from intracellular membranes. It is not clear which cellular membranes and which viral proteins are involved in this process. But it was demonstrated that the proteins 2B, 2C, 3A, 3B and their processing intermediates are able to interact with intracellular membranes and to induce structural membrane rearrangements. RNA transcription is protein-primed, with the uridylylated VPg protein 3B (VPg-pUpU), which is produced in a reaction templated by a cis-acting replication element (*cre*) in the 5'-terminal coding region of 3D, representing the primer for the negative-sense RNA replication intermediate and the subsequent positive-sense genomic RNA synthesis. A participation of the 5'-terminal NTR structures in the switch from translation to replication on the same viral RNA is suggested.

HAV morphogenesis is poorly understood. The primary polyprotein cleavage event occurs at the 2A/2B junction mediated by the 3C protease resulting in the structural precursor P1-2A. The steps resulting in particle formation are not entirely clear, but a model for this process has been suggested (see Fig. 1). The P1-2A structural precursors assemble to a pentameric structure and are further cleaved by the 3C protease to generate the precursors VP4-VP2 (also known as VP0) and VP1-2A (also known as VP1pX), as well as VP3 resulting in a structural building block (VP0,VP3,VP1-2A)₅. The 2A C-terminal extension of VP1 is a critical structural intermediate in virion morphogenesis, probably clamping the pentamer at the fivefold symmetry axis. After assembly of 12 such pentamers with the genomic RNA to provirions (12 × (VP0,VP3,VP1-2A)₅-RNA), VP0 is cleaved into VP4 and VP2 by an unknown mechanism. These immature particles, still containing VP1-2A, are present in lysosome-like vesicles as well as in multivesicular

bodies (MVB). Entry into MVBs is supposed to be mediated by an interaction between VP2 and the cellular endosomal sorting complex required for transport (ESCRT) and from there still immature HAV is secreted non-lytically in exosome-like vesicles (eHAV) from the basolateral site of hepatocytes into blood. How immature particles get access to lysosome-like vesicles is unknown, but the maturation cleavage at the VP1-2A junction leading to the removal of 2A seems to result from the action of a lysosomal protease before mature HAV virions are apically released into bile non-lytically via a secretory lysosomal pathway.

Host Range, Transmission and Tissue Tropism

Until recently, when several HAV species were discovered in diverse mammals, only humans and certain species of non-human primates were considered to be infectable by HAV under natural conditions. Non-human primates, which are also used as animal models, include chimpanzees, marmosets and owl monkeys. As mice with an impaired interferon system become permissive to human HAV infection, the ability of HAV to interfere with the species-specific interferon system (e.g., cleavage of MAVS) seems to contribute to HAV's host tropism.

HAV is transmitted via the fecal-oral route. As the virus is excreted in feces, it is typically acquired by ingestion of feces-contaminated food or water. Direct person-to-person spread occurs under poor hygienic conditions.

The site of HAV replication is the liver. The events that occur during the passage of HAV across the intestinal epithelium into blood, in which the virus reaches the liver via the portal circulation, are not clearly understood. Although HAV antigen could be detected in different organs, such as the kidneys, spleen and gastrointestinal tract, no extrahepatic sites of HAV replication have been clearly identified. The only evidence for replication within the gut comes from studies of orally inoculated owl monkeys, but it was demonstrated that replication of HAV in polarized intestinal cells does not result in basolateral release of the virus progeny and thereby not in penetration of the epithelium.

A functional cellular receptor could not be identified so far. The ubiquitously expressed cell surface T cell immunoglobulin mucin 1 (TIM1) protein, which binds HAV and had been suspected to be a HAV receptor (also known as HAV_{cr1}), is not essential for cellular entry. It can be supposed that the HAV receptor is expressed by different cell types, and conserved through evolution, and it is unlikely that the receptor is the main determinant of the strong hepatotropism. Some studies suggest that the hepatotropism of HAV is supported by fecal immunoglobulin A (IgA)-virus complexes (HAV/IgA), as HAV-specific IgA facilitates transcytosis of virus particles antivectionally through polarized epithelial cells via the polymeric immunoglobulin receptor (pIgR), and as subsequently HAV/IgA uptake via the hepatocellular asialoglycoprotein receptor (ASGPR) results in infection of hepatocytes (IgA-carrier hypothesis).

As for eHAV, which is released into blood but not into the fecal transmission pathway, it is unknown how it enters hepatocytes and that selectively. The entry of eHAV might occur via an endocytotic pathway in which the virus is amenable to HAV-specific neutralizing antibodies, which develop in the course of the infection.

After entry of HAV into susceptible cells, the structural protein VP4 might facilitate the release of the genomic RNA into the cytoplasm by its pore-forming ability (pores of 5–9 nm in diameter) in endosome-like membranes at acidic pH.

The virus progeny produced in the liver is released back into the intestinal tract via bile for transmission.

Cell Culture and Growth Characteristics

In vitro HAV can infect a variety of primate and nonprimate cell lines, including nonhepatic cells. The virus exhibits a protracted replication cycle and normally establishes a noncytolytic, persistent infection with low virus yields, and there is no evidence that HAV notably interferes with the macromolecular synthesis of the host cell. After infection of cultivated cells with the wild-type virus, a minimum of 8 weeks elapses before HAV can be isolated. Although a more rapid replication and higher virus titers are obtained after serial virus passages in cultivated cells resulting in cell culture-adapted viruses, even the replication of these virus variants is not detectable within the first days after infection. Adaptation of HAV to growth in cultivated cells seems to be achieved by varying sets of multiple interacting mutations, with adaptive mutations within the IRES enhancing viral translation in a cell-type-specific fashion, and mutations clustering in the 2B and 2C proteins (see [Table 1](#)) increasing replication regardless of the cell line used.

The virus apparently downregulates its own replication and this may, for example by supporting the ability of HAV to inhibit innate cellular anti-viral defense mechanisms, be important for the establishment of persistent infections. A large proportion of the virus progeny remains cell associated, but extensive release of HAV and eHAV from the cells also occurs.

Several cytopathogenic variants of HAV have been isolated which induce apoptosis resulting in cell death. These variants are highly cell culture adapted and characterized by a rapid replication phenotype and high virus yields. In these variants, both the downregulation of viral replication and the ability to inhibit the innate defense mechanisms are less effective. The molecular mechanisms resulting in apoptosis are not known.

Clinical Features and Pathology

HAV infection may lead to a wide spectrum of manifestations ranging from a silent infection, over icteric courses to a fatal fulminant hepatitis. The acute icteric course of infection varies between common, over prolonged to relapsing hepatitis A. Persistent infections or chronic disease states have not been described.

The clinical presentation of the disease depends on the age of the infected individual. The likelihood of having symptoms and the severity of the disease increases with the age of the patient. Silent infections (asymptomatic or at least anicteric) are normally observed in very young children, under the age of 2. However, a clinically detectable disease can occur even in infancy (aged 2 weeks to 8 months) and it may be characterized by prolonged courses. In children under 5 years of age approximately 3%, in children 5–15 years old approximately 30% and in individuals over the age of 18 years as many as 70% may develop a clinically detectable disease.

The heterogenous outcome of HAV infections seems to depend on individual physiological conditions and multiple independent factors, such as the amount of virus progeny, the strength of IgA and IgG responses or the activity of the cytotoxic T lymphocytes.

Common Course of Infection

The incubation period ranges from 2 to 6 weeks with a mean duration of 4 weeks. The prodromal (preicteric) period of normally 4–6 days (which may vary from 1 day to more than 2 weeks) is characterized by nonspecific symptoms, like anorexia, nausea with vomiting, malaise, abdominal pain, loss of appetite, accelerated pulse, rash, headache and fever (38–39°C) as well as by gastrointestinal symptoms, typically as obstipation, but diarrhea is also observed.

The prodromal symptoms disappear with the onset of jaundice, which is seldom abrupt (in 15% of the cases no obvious prodromal symptoms are observed before the appearance of jaundice). The icteric phase, which ranges from 2 to more than 22 days (mean duration 3 weeks), is marked by jaundice, which may start with scleral icterus, dark-colored urine (conjugated bilirubinuria), clay-colored stools and clearly decelerated pulse.

The reconvalescence period ranges from 3 to 6 weeks, but fatigue, dullness, right upper quadrant tenderness and fast exhaustion may remain for 2–4 months. In almost all cases the liver is enlarged.

The clinical symptoms are accompanied by several abnormal biochemical parameters (see Fig. 2). The elevation of aminotransferase levels in serum (alanine aminotransferase (ALT) and to a lesser degree aspartate aminotransferase (AST)), which reflect hepatocellular damage with the release of liver enzymes into the circulation, roughly correlate with the severity of the disease. The elevation of serum alkaline phosphatase activity and serum bilirubin level relate to intrahepatic cholestasis.

At the onset of symptoms, seroconversion to HAV occurs.

Large amounts of HAV, which are produced in the liver and released into the gastrointestinal tract via bile, are secreted in the feces already during the late incubation period when no clinical symptoms are detectable and are shed for approximately 3 weeks until a few days after the onset of elevated levels of liver enzymes in the serum (see Fig. 2). Fecal shedding of HAV reaches its maximum just before the onset of hepatocellular injury and terminates around the time when the IgG antibody response is detectable.

Viremia occurs a few days before and during the early acute stage of the clinical and biochemical hepatitis, in which it roughly parallels the shedding of virus in feces, but at a lower magnitude (see Fig. 2).

More sensitive methods (especially nucleic acid amplification technologies) demonstrated that low levels of viral RNA may be present in feces and blood for many weeks.

Prolonged Hepatitis A

In 8.5%–15% of the cases, jaundice lasts for up to 17 weeks. In general, the symptoms are initially stronger than during the common course of the infection and the antibody titers reached are higher and sustain for a longer period of time. Virus in feces and blood may be detectable for 2–3 months. The biochemical abnormalities last at least 5 months. Occasionally, prolonged courses are accompanied by exceptionally high serum aminotransferase levels (ALT up to 560 IU dl⁻¹) and by high serum bilirubin levels (up to 7.4 mg dl⁻¹) persisting for months (cholestatic hepatitis A). The cholestatic form of HAV infection is associated with extensive itching of the skin.

The pathogenesis of prolonged hepatitis is not understood and different factors could contribute to the extended course: reinfections at different sites of the liver may occur by enterohepatic cycling of HAV, which may be supported by anti-HAV immunoglobulin A (IgA) (IgA-carrier hypothesis) and by serum eHAV (enveloped HAV), respectively. HAV/IgA is neutralized for infection of hepatocytes via the ASGPR only late in the course of infection by highly avid immunoglobulin G (IgG) and eHAV is resistant to neutralizing antibodies. Membrane enveloping of HAV might also facilitate viral spread in the liver supporting infection for a longer time. Also NF-κB, which is activated by HAV, may contribute to prolonged courses by attenuation of the effects of HAV-specific cytotoxic T cells. Furthermore, infection by different HAV genotypes resulting in an elongated course of the disease has been speculated.

Relapsing Hepatitis A

After initial improvement in symptoms and liver test values (serum aminotransferase levels), one or more relapses of the disease (mostly biphasic) are described for up to 20% of the patients (the average age in adults is 30 years, in children 9 years). These relapses occur between 30 and 90 days after the primary episode, when high titers of neutralizing antibodies are already present. The severity of symptoms, the serum aminotransferase levels and the immunoglobulin M (IgM) response are similar or weaker than observed during the initial phase, with a higher tendency for cholestasis (30% of these patients). The relapses last longer

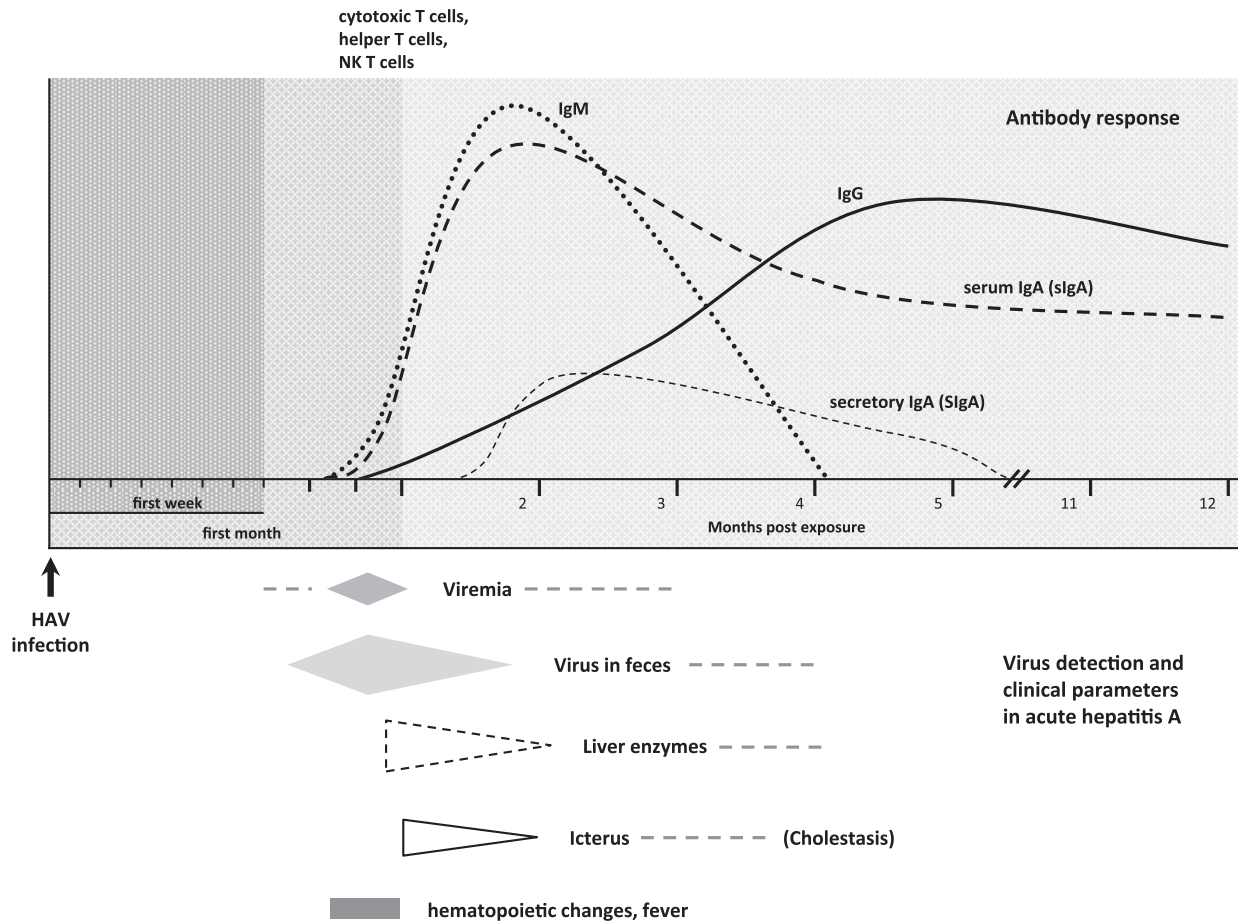


Fig. 2 Course of clinically relevant events in acute hepatitis A. This figure schematically shows the mean duration and intensity of certain parameters. The dotted lines indicate the duration and intensity of the events that may vary. NK T cells: Natural killer T cells.

(mean duration 6 weeks) than the initial phase, HAV is present in blood and the serum bilirubin level is higher. During the relapse the tendency for purpura and itching of the skin is increased.

No specific risk factors could be identified and the pathogenesis of relapsing hepatitis is not understood. The relapses may represent a manifestation of endogenous reinfections by an enterohepatic cycling of HAV/IgA (IgA-carrier hypothesis) until the reinfection cascade is terminated by competitive replacement of the IgA in the HAV/IgA complexes by large amounts of the avidity-matured IgG response.

Fulminant Hepatitis A

In rare cases, acute hepatitis A results in a fatal deterioration in liver function with massive destruction of liver cells. Surprisingly, no vigorous inflammatory response is observed. The fatality rate is below 1.5% of all hospitalized icteric HAV infection cases. This course of the disease is accompanied with fever over 40°C. This outcome is more frequent in adults, especially in patients over 50 years of age, and the risk is increased in patients with underlying chronic liver disease.

Extrahepatic Manifestations

Extrahepatic manifestations of the disease are rare and the etiology is unclear. Besides frequently observed transient suppression of hematopoiesis (a significant leukocytopenia occurs 1–7 days before the appearance of jaundice correlating with the viremic and the fever phases; see Fig. 2), rare cases of aplastic anemia (occurring 1.5–2 months after appearance of jaundice) with a lethality rate of over 90% are described, and it was demonstrated that HAV infects monocytes and inhibits their further differentiation. In some patients interstitial nephritis, pancreatitis, myocarditis or Guillain-Barré syndrom were observed. By statistical evaluation of medical records, it was suggested that HAV exposure may leave a protective effect on the development of asthma and allergic diseases. However, due to contradictory findings in different studies this is still a subject of debate.

Pathology

The pathology of the liver in acute hepatitis A has been studied in humans and some animal models. The pathological lesions, which are caused by an immunopathogenic mechanism, are characterized by hepatocellular necrosis, which is most prominent in periportal regions, accompanied with large inflammatory infiltrates of mononuclear cells.

Innate and Adaptive Immune Response

Innate Immune Response

HAV, which does not interfere with the replication of other viruses, prevents the synthesis of beta interferon (IFN- β). However, the virus is sensitive to alpha and beta interferon (IFN- α/β) exogenously added to persistently infected cells. It can be demonstrated that HAV inhibits the transcription of the IFN- β gene by blocking effectively interferon regulatory factor 3 (IRF-3) activation due to a cooperative interaction of the HAV proteins 2B and 3ABC with the mitochondrial antiviral signaling protein MAVS, which is a component of the retinoic acid-inducible gene-I (RIG-I) and melanoma differentiation-associated gene 5 (MDA-5) signaling pathway (see Fig. 3). While MAVS is targeted for HAV protease 3C-mediated proteolysis by 3ABC, an intermediate of polyprotein

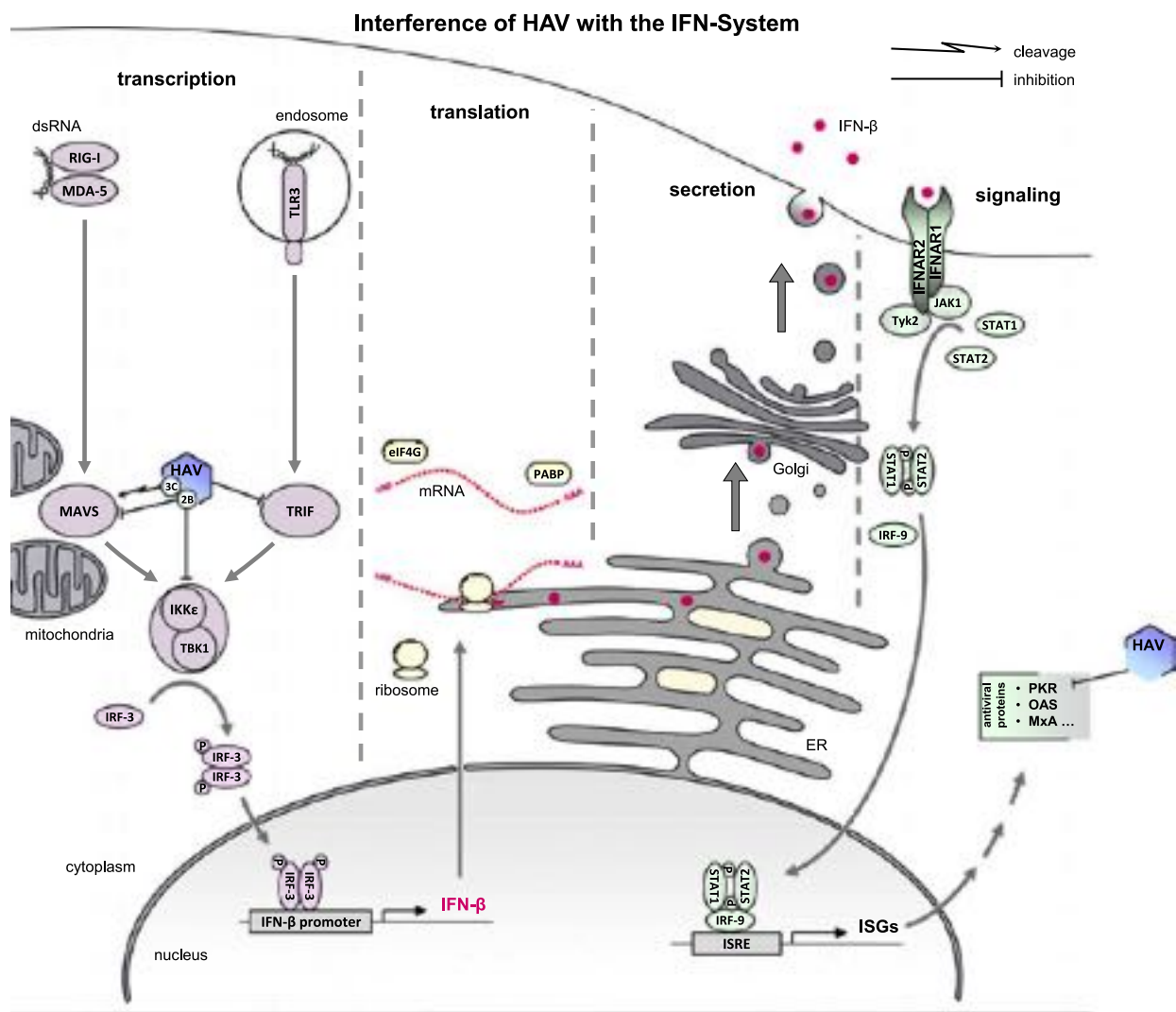


Fig. 3 Interference of HAV with the interferon system. This figure schematically shows the sites and effects of the interference of hepatitis A virus with IFN induction. The figure also displays that, besides inhibiting protein kinase R (PKR) activity, HAV does neither interfere with interferon signaling nor significantly with cellular macromolecular synthesis. dsRNA, Double-stranded RNA; IFN, Interferon; IRF, Interferon regulatory factor; ISGs, Interferon stimulated genes; ISRE, Interferon stimulated response element; MAVS, Mitochondrial antiviral signaling protein; MDA, Melanoma differentiation-associated gene; OAS, Oligoadenylate synthetase; RIG-I, Retinoic acid-inducible gene I; TLR, Toll-like receptor.

processing localized to mitochondria by 3A, 2B indirectly seems to interfere with MAVS as well as with the kinases responsible for IRF-3 phosphorylation (TBK1/IKK ϵ) through interactions with cellular membrane structures. Signaling through the Toll-like receptor 3 (TLR-3) pathway is also impaired by an interaction with TRIF (TIR domain-containing adaptor inducing IFN- β) (see Fig. 3). Gamma interferon (IFN- γ) produced by HAV-specific HLA-dependent cytotoxic T lymphocytes (CTL) may contribute to the elimination of the HAV infection by inducing an antiviral state in the later stages of the infection. The importance of NK cells for the elimination of HAV is controversially discussed.

As CpG containing codons are avoided in the coding sequence, HAV evades recognition of unmethylated viral CpG by Toll-like receptor 9. Furthermore, it could be shown that HAV inhibits the activity of protein kinase R (PKR) by a so far unknown mechanism (see Fig. 3).

Whether HAV induces or inhibits the activation of NF- κ B is controversially discussed.

HAV also has the ability to prevent apoptosis induced by accumulating dsRNA, but the underlying mechanism is not clear.

Adaptive Immune Response

Anti-HAV IgM antibodies are present in almost all patients at the onset of the symptoms (see Fig. 2). These antibodies reach their maximum level 2 months post exposure, have only weak neutralizing activity and disappear in the course of 3 months, but in the case of prolonged infections IgM can be detected up to 1 year after onset of icterus.

Anti-HAV IgA antibodies are also detectable at the onset of the symptoms (see Fig. 2). This response reaches its peak titer 50 days post infection and may last for more than 5 years. The majority of the IgA remains as serum IgA in circulation and is not secreted into the intestinal tract as secretory IgA by the polymeric immunoglobulin receptor (pIgR) pathway. But a significant fraction of this serum IgA is released into the gastrointestinal lumen via bile by liver functions under participation of the hepatocellular IgA-specific asialoglycoprotein receptor (ASGPR). Fecal samples contain IgA from 5 weeks till 3–6 months post infection. Salivary anti-HAV IgA is also detectable in HAV patients, its course parallels that of fecal IgA. The role of secretory IgA antibodies in the protection against HAV infection appears to be limited, as neutralizing activity is barely detectable. Studies suggest that HAV-specific IgA can serve as a carrier molecule for a liver-directed transport of the virus, supporting and enhancing the hepatotropic infection by uptake of HAV/IgA immunocomplexes via the ASGPR (IgA-carrier hypothesis). As IgA-coated HAV is translocated antiverotically from the apical to the basolateral site of epithelial cells via the pIgR, it was assumed that fecal HAV/IgA immunocomplexes, whose stability enables their fecal-oral transmission, are able to support the primary infection utilizing the IgA receptors. Furthermore, an enterohepatic cycling of HAV may be established during infection by HAV/IgA resulting in endogenous reinfections of the liver until large amounts of highly avid IgG replace IgA in HAV/IgA immunocomplexes. Depending on individual factors, this mechanism may play a role in the development of the different courses of hepatitis A. It is not clear by which processes and mechanisms the induction of the anti-HAV IgA response occurs.

Neutralizing anti-HAV immunoglobulin G (IgG) antibodies are also detectable 3 weeks post infection for the first time, but this response develops slowly, reaching its peak titer approximately 4 months post infection (see Fig. 2), a time point late in the convalescence phase. Envelopment of HAV in blood (viremia) by cellular membranes (eHAV) unrecognizable for the immune system may contribute to the slow IgG antibody response, as well as the transient perturbation of hematopoiesis, which may result in an attenuation and retardation of the inflammatory response and of the induction of the adaptive IgG response. Anti-HAV IgG persists lifelong, although the titer may fall to undetectable levels after several decades. Neutralizing antibodies, which are effective in eliminating the virus from the blood, do recognize a conformational epitope clustered into a major, immunodominant antigenic site involving residues contributed by VP3 (Asp₇₀, Gln₇₄) and VP1 (Ser₁₀₂, Val₁₇₁, Ala₁₇₆, Lys₂₂₁). Neutralization of eHAV by antibodies is possible during an assumed endocytotic re-uptake from blood by hepatocytes in an endosomal compartment.

HAV-specific, HLA-restricted cytotoxic CD8⁺ T-lymphocytes (CTL) accumulate in the liver during the acute phase of the infection (after destruction of the infected hepatocytes, approximately 2–3 weeks after onset of the icterus, the CTLs leave the liver back into blood) and they play a prominent role both in eliminating the virus and in causing a liver injury (see Fig. 2). Gamma interferon (IFN- γ), released by these CTLs, may stimulate HLA class I expression on hepatocytes and in the following promote upregulation of the normally low level display of antigen on liver cells. Multiple dominant T-cell epitopes could be identified in the proteins VP1, VP2, VP3, 2B, 2C and mainly 3D (amino acids 2025–2033). This multitude of T-cell epitopes combined with an inhibitory effect of HAV on CTL-suppressing regulatory T cells (by binding of HAV to HAV_{cr1} expressed on T_{regs}) during the acute phase of the disease seems to result in a strong activity of HAV-specific CTLs. Studies in experimentally infected chimpanzees suggest a more complex T cell response with the involvement of anti-viral cytokines producing CD4⁺ helper T cells also. In addition, NK T cells and possibly bystander-activated CD8⁺ T cells seem to be involved in the elimination of HAV and the destruction of hepatocytes. Therefore, multiple immune mechanisms seem to contribute to the acute liver injury in humans (immunopathogenesis), also leading to an efficient elimination of HAV, which might prevent persistence of the virus.

Diagnosis

Since the clinical presentation of hepatitis A cannot be distinguished from hepatitis caused by the other hepatitis viruses, serologic tests or nucleic acid amplification techniques are necessary for a virus-specific diagnosis.

The routine diagnosis of acute hepatitis A is made by detection of anti-HAV IgM in the serum of patients (see Fig. 2). A further option is the detection of virus in the feces.

In order to improve the safety of blood and blood products, blood screening with HAV-specific polymerase chain reaction (PCR) is performed, which reduces the window period of up to 3 weeks post infection during which HAV infection fails to be diagnosed by serologic assays.

Epidemiology

HAV occurs worldwide and accounts for over 1.5 million clinical cases reported annually. The seroprevalence pattern ranges from high endemicity, such as in Africa, Southern Asia and Latin America, where infection normally occurs in childhood, over intermediate endemicity, such as in Eastern Europe and the northern parts of Asia, to low endemicity, such as in Western Europe and North America, where the majority of the population remains susceptible to HAV infection. However, the epidemiology pattern is complex and continuously changing, with considerable heterogeneity among different countries. In general, the anti-HAV antibody prevalence inversely correlates with the quality of the hygienic standards, and the incidence declines in many populations through improvements in public sanitation and living conditions. These improvements result in an increase of the pool of susceptible adults, with a shift in the age of infection to older age groups, in which a more severe disease form is observed, leading to an increased morbidity. HAV has been recognized as a main cause of fulminant hepatic failure in a growing number of countries.

A minor seasonal variation of HAV infections is observed, with a peak occurring during the fall and winter months, mentioned in almost all earlier and contemporary reports, nowadays possibly as a result of exposure during summer vacations spent in endemic countries.

At special risk for acquisition of hepatitis A are international travelers from areas of low endemicity to endemic areas, employees of child-care centers and sewage plants, gully workers, injecting drug users, homosexually active men and persons with an increased risk of developing a fulminant disease, such as persons with chronic hepatitis C virus (HCV) infections.

The high physical stability of HAV provides a good opportunity for common-source transmission. Community-wide outbreaks are reported in association with infections of food handlers, and linked to contaminated food and drink, or uncooked clams from contaminated water. Hepatitis A is most commonly acquired by sharing the household with an infected person.

Prevention and Control

There is no specific antiviral treatment for hepatitis A. As almost all HAV infections are spread by the fecal-oral route, good personal hygiene, high-quality standards for public water supplies, and proper disposal of sanitary waste are important measures to reduce virus transmission.

Until the availability of an active prophylaxis, the disease could be prevented for up to 5 months with a certainty of 80%–90% by passive immunization with pooled IgG of at least 100 IU anti-HAV antibodies, a prophylaxis which is effective within hours. IgG is still used for postexposure prophylaxis. If administered within 2 weeks after exposure, either development of the disease is prevented or the severity of the disease as well as virus shedding is reduced.

Since 1992, inactivated vaccines for active immunization have been available. These vaccines contain purified, attenuated, formalin-inactivated HAV produced in cell culture, which are absorbed to an aluminum hydroxide adjuvant. They are highly immunogenic and protect against both the infection and the disease caused by all strains of HAV with 100% efficacy for at least 10 years, which is consistent with the finding that all human HAV strains belong to one single serotype. HAV vaccines are also effective when given within 2 weeks after exposure.

Live, attenuated HAV vaccines have been developed in China for a single-dose, subcutaneous injection using virus adapted to grow in cell culture, which protect with similar efficacy as the inactivated vaccines. No reversion of the attenuated virus to virulence has been reported so far. However, viral shedding is possible, which must be taken into account in vaccinations in populations with large numbers of susceptible adults (see Epidemiology section).

Although the minimum level of neutralizing antibodies that protect against infection and disease is unknown, an estimate of a minimal protection level is approximately 20 mIU ml⁻¹ blood.

Further Reading

- Bell, B.P., 2002. Global epidemiology of hepatitis A: Implications for control strategies. In: *Proceedings of the 10th International Symposium on Viral Hepatitis and Liver Disease*.
- Debing, Y., Neyts, J., Thibaut, H.J., 2014. Molecular biology and inhibitors of hepatitis A virus. *Medicinal Research Reviews* 34 (5), 895–917.
- Dotzauer, A., Kraemer, L., 2012. Innate and adaptive immune responses against picornaviruses and their counteractions: An overview. *World Journal of Virology* 1 (3), 91–107.
- Gerety, R.J. (Ed.), 1984. *Hepatitis A*. London: Academic Press.
- Gust, I.D., Feinstone, S.M., 1988. *Hepatitis A*. Boca Raton: CRS Press.

- Jacobsen, K.H., 2010. The Global Prevalence of Hepatitis A Virus Infection and Susceptibility: A Systematic Review. Geneva: WHO Press.
- Lanford, R.E., Walker, C.M., Lemon, S.M., 2017. The chimpanzee model of viral hepatitis: Advances in understanding the immune response and treatment of viral hepatitis. *ILAR Journal*, 1–18.
- Lemon, S.M., Ott, J.J., Van Damme, P., Shouval, D., 2018. Type A viral hepatitis: A summary and update on the molecular virology, epidemiology, pathogenesis and prevention. *Journal of Hepatology* 68, 167–184.
- Wasley, A., Fiore, A., Bell, B.P., 2006. Hepatitis A in the era of vaccination. *Epidemiologic Reviews* 28, 101–111.

Hepatitis B Virus (*Hepadnaviridae*)

Peter Karayiannis, University of Nicosia, Nicosia, Cyprus

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

IU/ml International units

Glossary

Immunomodulation Changes in the body's immune system, caused by agents that activate or suppress its function.

Nucleos(t)ide analogue Antiviral medication that is incorporated into viral nucleic acid during synthesis causing chain termination.

Perinatal Of, relating to, or being the period around childbirth, especially the five months before and one month after birth.

Percutaneous Through the skin following injury or puncture with a needle.

Quasispecies A population of closely related variants of a virus isolated from an individual at a given time.

Introduction

Since the discovery and recognition of HBV as a human hepatitis causing pathogen, great strides have been made in our understanding of its epidemiology, virology, life cycle, pathogenesis, treatment and vaccine protection. Discovered by Bloomberg and co-workers just over fifty years ago, it remains a major cause of morbidity and mortality in countries of the world where the virus is still endemic, in spite of effective regimens of antiviral drugs which suppress viral replication and vaccine campaigns to protect those at greater risk. Recent estimates suggest that 291 million people are chronically infected with HBV representing a global prevalence of 3.6%–3.9%. Accurate figures of chronic infection are incomplete from many parts of the world and even from Western countries entry data on death certificates for example may not be accurate. Thus there is a reliance on mathematical models that take into account various confounding factors. These suggest that as chronic infection over many years may lead to cirrhosis and hepatocellular carcinoma, the incidence of these two conditions for all viral hepatitis is 1.4 million annually, of which nearly 700,000 are HBV-related. This represents an increase of over 33% over the period 1990–2013, when viral hepatitis climbed from 10th to the 7th position of leading causes of mortality worldwide.

Although, the prevalence of HBV infection is decreasing in several highly endemic countries due to actions which will be explained later, acute infections still occur primarily in high risk populations. What is more, recent population movements and migration from war-torn regions of the world to low endemicity countries may be changing the prevalence and incidence of HBV in these countries, particularly if such individuals evade surveillance. Vaccination programs and antiviral treatment remain the only means of reducing infections and halting progression of chronic liver disease to cirrhosis and HCC respectively.

The study of the life cycle and molecular biology of HBV, in the absence of a reliable cell culture system for the propagation of the virus has been re-invigorated in recent years following the identification of its hepatocyte receptor. Nevertheless, studies in the early 1980s employing genetic engineering techniques led to the cloning of the viral genome and its sequencing. This allowed expression of its proteins, enabling the study of their function, and more importantly, unraveled the unique and fascinating mechanism of its replication strategy. In this respect, the contribution of animal models has been invaluable. These initially employed chimpanzees, and subsequently ducks, woodchucks, transgenic and more recently chimeric mice. This has allowed the study of many aspects of the life cycle, molecular biology and immuno-pathogenesis of the virus, and facilitated vaccine development and antiviral testing. What follows is what we know to-date about this very important pathogen.

Classification

HBV is the prototype member of the *hepadnaviridae* which comprises a group of hepatotropic DNA viruses that have been steadily growing in numbers. These viruses are species specific and are broadly divided into two genera depending on their natural host. Thus, the mammalian *Orthohepadnavirus* genus includes members that infect rodents such as woodchucks (*Marmota monax*) and ground squirrels (*Spermophilus beecheyi*), bats (various species) and primates (humans, chimpanzees, gorillas, gibbons, baboons, orangutans, macaques and woolly monkeys). The *Avihepadnavirus* genus infects birds such as pekin ducks (*Anas domestica*), herons, storks, geese and parrots with about 80% homology between them. The homology between the two genera is around 40% and the genetic relatedness of sequenced isolates from various species are evident in Fig. 1. Three branches are discernible within the orthohepadnaviruses, with human and primate isolates grouping together, and two other groups representing rodent and bat isolates respectively.

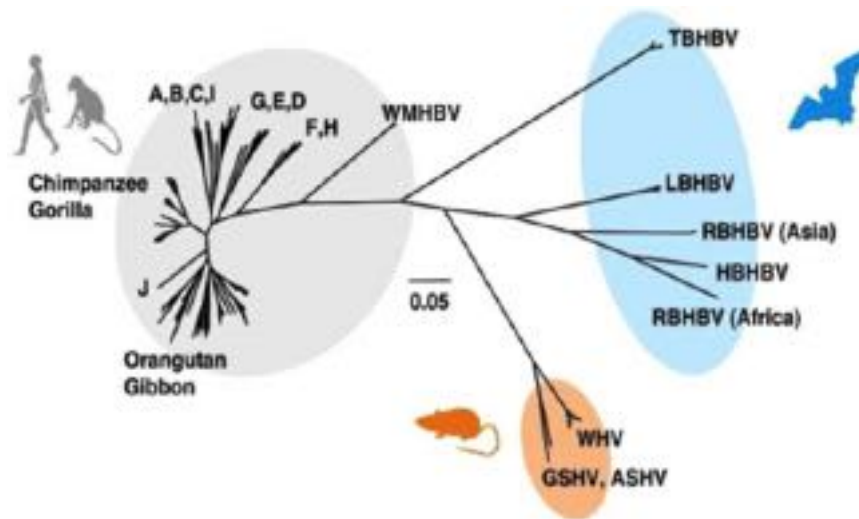


Fig. 1 Phylogenetic tree based on complete genome sequences of mammalian, rodent and bat HBV isolates, showing three separate groupings. Human HBV genotypes A-J group together with primate isolates as shown. Similarly, rodent isolates from woodchuck (WHV), ground and arctic squirrel hepatitis virus (GSHV, ASHV) group together. Bat HBV isolates from different species group together on the third branch (e.g., LBHBV, long-fingered bat HBV). Reprinted by permission from Springer Nature, Rasche, A., Souza, B.F.C.D., Drexler, J.F., 2016. Bat hepadnaviruses and the origins of primate hepatitis B viruses. *Current Opinion in Virology* 16, 86–94.

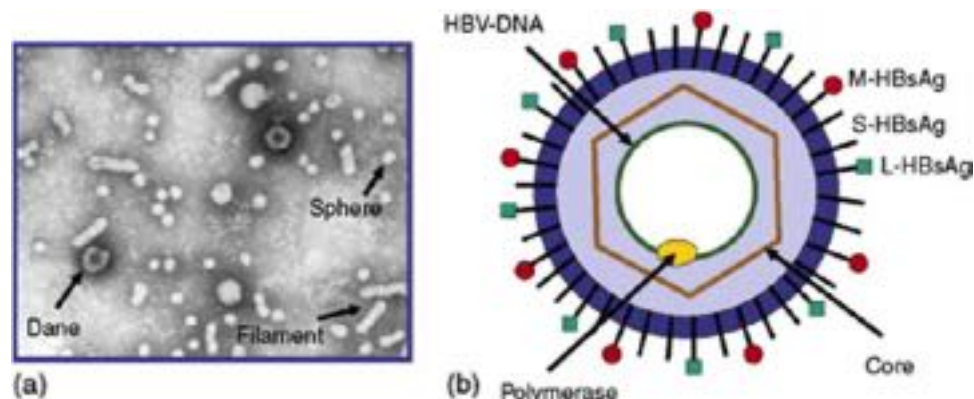


Fig. 2 Structure of the hepatitis B virion. (a) Electron micrograph of HBV purified from plasma showing the infectious Dane particle and the spherical and filamentous subviral particles. (b) Cartoon of the virion and its components.

Virion Structure

HBV has pleomorphic appearance under the electron microscope (**Fig. 2**) and this because other than the infectious virions, often referred to as Dane particles, there is an abundance of subviral particles which are known as filaments and spheres. The Dane particles were named after the scientist who led the team that discovered them and have a diameter of 42–45 nm. The virions have an outer glycolipid envelope made of hepatitis B surface antigen (HBsAg) in a lipid bilayer, which is thought to derive from the endoplasmic reticulum membrane. This in turn encloses the nucleocapsid or core of the virus consisting of hepatitis B core antigen (HBcAg) and measuring 30–32 nm in diameter. The cores contain the viral genome and a copy of its polymerase. Additionally the virus produces a soluble protein known as hepatitis B e antigen (HBeAg) which is not a structural protein, but rather constitutes a marker of active viral replication.

The abundant sub-viral particles circulating in serum consist entirely of HBsAg and are devoid of cores and any viral nucleic acid. In this respect, these particles are not infectious. As mentioned above, these are the 25 nm spheres and the 22 nm diameter filaments, which outnumber infectious virions by 100–10,000-fold. Their production and secretion in such excess may serve to mop up anti-HBs antibodies or cause T cell anergy, both of which may contribute to immune tolerance and thus viral persistence. Similarly, HBeAg is thought to function as a tolerogen.

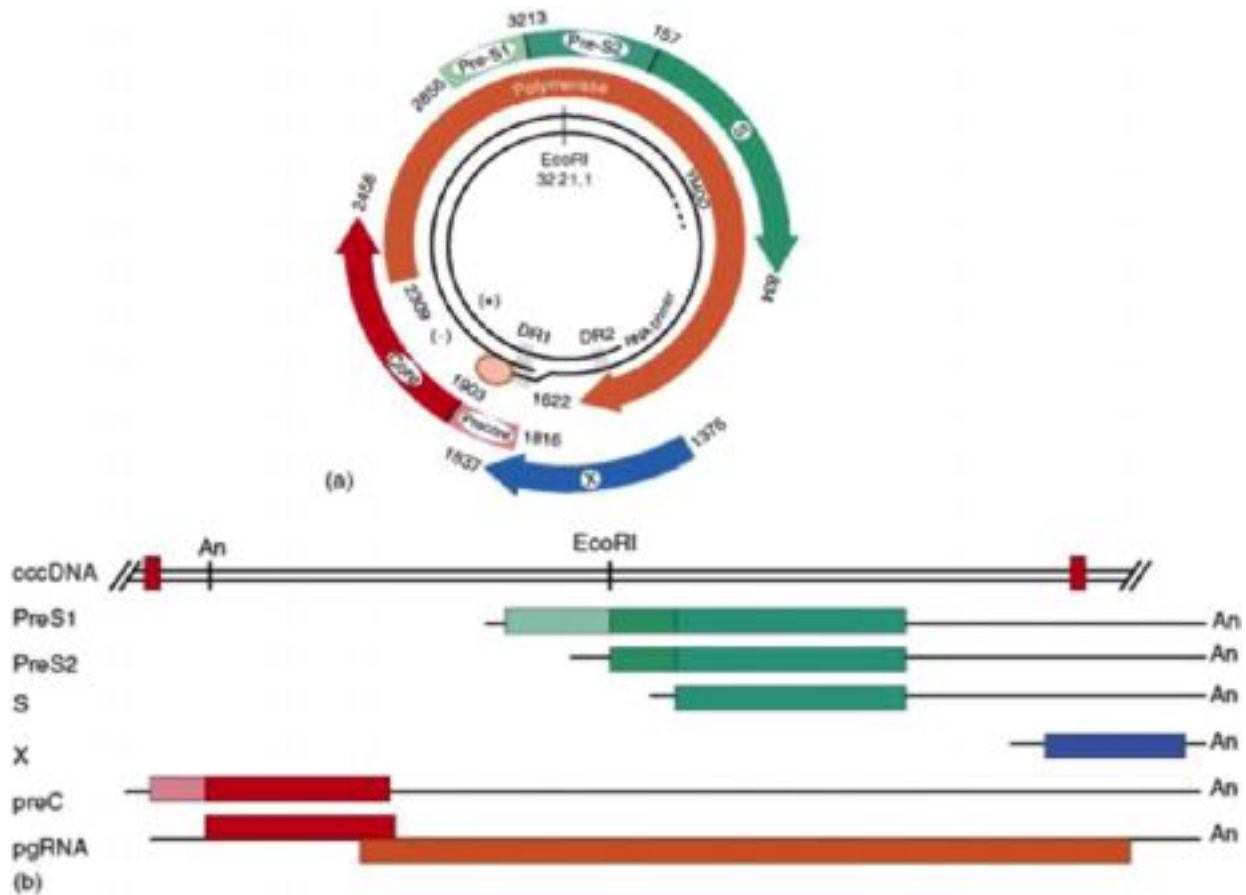


Fig. 3 Genome organization of the virus (a) and the transcripts encoding the various viral proteins (b), synthesized from cccDNA. All of them terminate at the common polyadenylation site. Reprinted (panel a) by permission from Wiley, Hunt, C.M., McGill, J.M., Allen, M.I., Condeary, L.D., 2000. Clinical relevance of hepatitis B viral mutations. *Hepatology* 31, 1037–1044.

Genome

The partially double stranded DNA viral genome is a relaxed circular molecule (rcDNA) 3.2 kb in length, and is the smallest among DNA viruses. The strands overlap as shown in [Fig. 3](#) with the minus (–)-strand being complete and the positive (+)-strand incomplete at its 3' end and of variable length (about 600 nucleotides missing). It is one of the most compact genomes as it contains 4 wholly or partially overlapping open reading frames (ORFs) ([Fig. 3](#)). What is more, every nucleotide of the genome lies within the coding region of one of the 4 ORFs. These ORFs also incorporate in them all of the regulatory elements of the virus, including two enhancers (Enh1, Enh2), the four promoters (core, S1, S2 and X), polyadenylation, encapsidation (ϵ) and replication signals (DR1, DR2). The four ORFs are those of the Pre-S/S, Pre-core/core, X and polymerase and between them they encode for a total of 7 proteins which are translated from 6 co-terminal mRNAs. These are unspliced and capped, ending at a common polyadenylation signal embedded in the core ORF ([Fig. 3](#)). The synthesis of these mRNAs is regulated by the aforementioned enhancers and promoters which bear motifs which are recognized by transcription factors highly enriched in the nuclei of hepatocytes which are the target cells of the virus.

Transcription and Translation

Surface proteins

The Pre-S/S ORF encodes three proteins which are glycosylated and are part of the viral envelope. These are known as the large (L), middle (M) and small (S) HBsAg, produced by differential initiation of translation at each of three in-frame initiation codons. Of these, the S-HBsAg constitutes the majority protein of the viral envelope. The proteins are translated from two mRNA transcripts of 2.4 kb and 2.1 kb in length, the synthesis of which is under the control of the S1 and S2 promoters respectively. The 2.4 kb transcript is utilized for the synthesis of L-HBsAg whilst the 2.1 kb transcript is responsible for the translation of M- and S-HBsAg, the latter through leaky ribosome scanning. The proteins are co-terminal and the sequence of S-HBsAg consisting of 226 amino-acids is shared by the other two and constitutes their C-terminus. The M-HBsAg has an additional 55 amino-acids at its N-terminus

encoded by the Pre-S2 region, whilst the L protein has 107–118 (depending on genotype) further amino-acids encoded by the Pre-S1 region. A stretch of 75 amino-acids at the N-terminus of Pre-S1 appears to be responsible for the interaction of the virus with its hepatocyte receptor. As mentioned previously, the three proteins are glycosylated, although the L- and S-HBsAg may be present in an unglycosylated form in particles also. In addition to glycosylation the L-HBsAg undergoes one other post-translational modification in the form of myristylation, a step that appears necessary for infectivity and envelopment of the core particles. These proteins are synthesized in close proximity to the ER membrane adopting a transmembrane configuration that facilitates budding of the virus into the lumen during virion maturation.

Core and polymerase

The synthesis of two longer than genome length mRNAs (3.5 kb) which differ with respect to the start of their 5' end is under the control of the core promoter. The precore mRNA is longer by a small number of ribonucleotides amongst which the initiation codon for the synthesis of the precursor of the soluble HBeAg often referred to as the precore protein. This is synthesized when the first of two in frame initiation codons that the precore/core ORF contains is utilized. The N-terminus of this protein has a signal peptide that anchors the protein in the ER membrane. A signal peptidase in the lumen of the ER cleaves the first 19 amino-acids of the protein and what remains undergoes furin cleavage for the removal of an arginine rich domain at its C-terminus. The resulting protein constitutes the 15 kD HBeAg, which is secreted in serum. Initiation of translation at the second initiation codon results in the synthesis of the nucleocapsid or core protein (21 kD). The core has the capacity to dimerize and form nucleocapsids consisting of 240 copies (120 dimers) of the protein through self-assembly. The HBeAg and core proteins are translated from two separate transcripts known as the precore mRNA and the pregenomic RNA (pgRNA), respectively.

The pgRNA is bicistronic in nature as it contains in addition the ORF encoding the polymerase (Fig. 3), a 90 kD protein that encompasses almost $\frac{3}{4}$ of the genome. The polymerase has multifunctional enzymic activities that include a reverse transcriptase (rt), DNA polymerase and RNase H activity. In addition, the pgRNA constitutes the template for synthesis of the (–)-DNA strand by reverse transcription during the replication of the viral genome, as explained later.

X protein

Finally, the fourth ORF encodes for the 17 kD HBx protein that is translated from the shortest transcript of 0.7 kb in length. It consists of 154 amino-acids and has been implicated in a number of actions ranging from modulation of host-cell signal transduction pathways, gene transactivation under experimental conditions, activation of transcription factors and epigenetic control of the covalent closed circular DNA (cccDNA) minichromosome that is essential for transcriptional activation. Moreover, cccDNA can persist in hepatocytes indefinitely thus maintaining chronic lifelong infection until such time as the host cell is eliminated through immune lysis.

Life Cycle

Attachment

Other than the requirement of transcription factors enriched in hepatocytes, it is now evident that species specificity and hepatotropism are determined by the expression of the recently described sodium taurocholate co-transporting peptide (NTCP) on human hepatocytes, which is the HBV receptor (Fig. 4). NTCP is a bile acid transporter expressed at the basolateral membrane of hepatocytes. The receptor-virus interaction is effected by binding of the N-terminal end of L-HBsAg as described above. This may not be the only requirement as evidence suggests co-operative binding in the process of attachment and uptake involving in addition, heparan sulfate proteoglycans and glypican.

Penetration and Uncoating

The virion is likely internalized through clathrin mediated endocytosis. The steps from uncoating to transfer of the nucleocapsid to the nuclear pore are not clear, but it seems that transport factors such as importin alpha and beta and component nucleoporin 153 assist in nucleocapsid delivery to the nuclear basket. The disassembly of the nucleocapsid is followed by the release of the rcDNA genome into the nucleoplasm with its covalently attached polymerase.

Transcription/Translation

Within the nucleoplasm the rcDNA is converted into the cccDNA molecule involving at least three stages. The viral polymerase which is covalently attached to the 5' end of the (–)-DNA strand is removed, as is the short RNA oligomer from the 5' end of the (+)-DNA strand which is used to prime (+)-DNA strand synthesis. Next, the variable (+)-DNA strand is completed and the ends of the two now complete strands ligated to form a double stranded circular molecule, referred to as cccDNA. This is quite stable and functions as a minichromosome following association with a number of epigenetic factors that include histones H3 and H4, transcription factors that include CREB, ATF, STAT1 and STAT2, chromatin modifying enzymes, histone acetyltransferases and deacetylases, as well as the HBc and HBx proteins. Transcriptionally active cccDNA constitutes the template for viral transcript

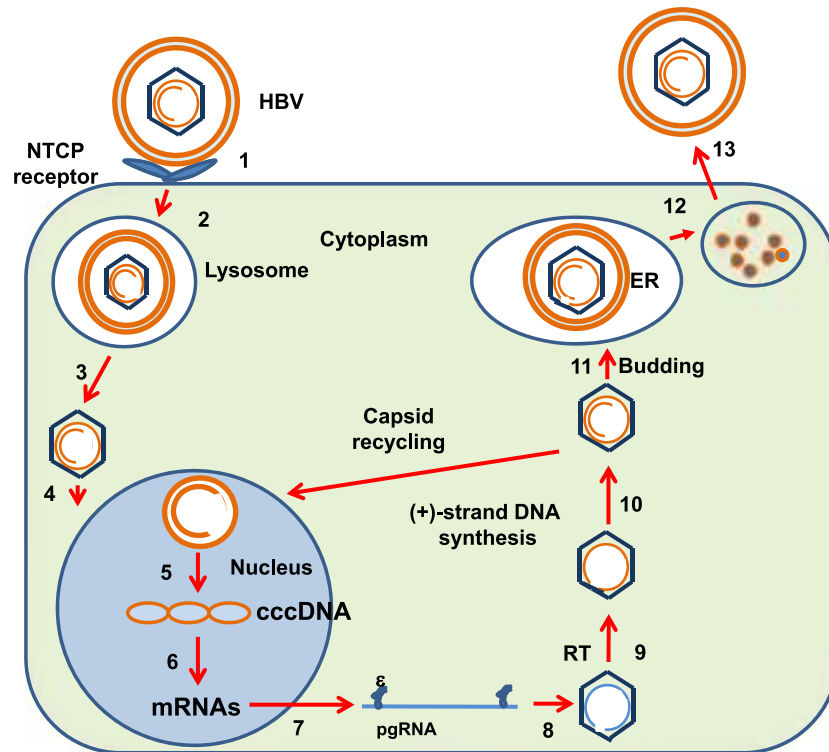


Fig. 4 Diagrammatic representation of the stages of the life cycle of the virus as explained in the text: 1. Attachment; 2. Endocytosis; 3. Capsid release; 4. rcDNA entry into the nucleus; 5. cccDNA synthesis; 6. Transcription; 7. mRNA transfer to the cytoplasm; 8. Encapsidation; 9. (–)-DNA strand synthesis by reverse transcription; 10. (+)-DNA strand synthesis; 11. Budding of virions into the ER lumen; 12. Virus transfer by multivesicular bodies (MVB) to hepatocyte surface; 13. Release. Modified from Karayiannis, P., 2017. Hepatitis B virus: Virology, molecular biology, life cycle and intrahepatic spread. *Hepatology International* 11, 500–508, with permission from Elsevier.

synthesis by the host RNA polymerase II. Newly synthesized viral transcripts are exported to the cytoplasm for viral protein synthesis.

Encapsidation

As mentioned above the bicistronic pgRNA encoding for the core and polymerase proteins, also serves as the template for DNA synthesis by reverse transcription. The pgRNA is longer than genome length by virtue of a terminal redundancy which duplicates the sequence at its 5' end by about 120 nt and terminates at the polyadenylation signal. The host RNA polymerase appears not to recognize this signal during the first pass but does so on the second encounter. Thus, the redundancy contains a second copy each of the direct repeat 1 (DR1) and the encapsidation signal ε , preceding and incorporating the precore nucleotide sequence respectively.

The encapsidation of the pgRNA into the nucleocapsid involves both viral and host factors. The polymerase has four domains which in sequence are the terminal protein involved in DNA priming, a spacer region of unknown function, the reverse transcriptase/polymerase (rt) and the RNase H which are involved in DNA synthesis and degradation of the pgRNA respectively. The polymerase recognizes and binds to the secondary structure of ε located at the 5' end of the pgRNA, an event that triggers encapsidation of the complex by the core protein (Fig. 5). This process is facilitated by the cap at the 5' end of the pgRNA, as well as eIF4E and heat shock proteins, which in addition support the stabilization and activation of the polymerase. The arginine rich C-terminus of the core protein is involved in pgRNA binding thus facilitating encapsidation also.

The steps that follow are performed within the nucleocapsid. The secondary structure that ε assumes consists of a lower stem, an upper stem, a side bulge and an apical loop, which are the result of base pairing of palindromic nucleotide sequences. Part of the sequence of the side bulge of ε serves as a template for the synthesis of only a 4 nucleotide long DNA primer (Fig. 5(A)). A phosphodiester linkage between dTTP and the hydroxyl group of a tyrosine residue in the terminal protein (position 63), results in the covalent attachment of the primer to this domain of the polymerase. This is followed by a translocation event whereby the polymerase-primer complex localizes to the 3' of the pgRNA, where the primer hybridizes with a homologous region within DR1 (Fig. 5(B)). This is assisted by two other elements, ϕ and ω , which interact with ε . Priming of the pgRNA is followed by the synthesis of the (–)-DNA strand initiated by reverse transcription as the complex proceeds towards the 5' end of the pgRNA, having a terminal redundancy of about 10 nucleotides (Fig. 5(C)). The RNase H activity of the polymerase degrades in the process

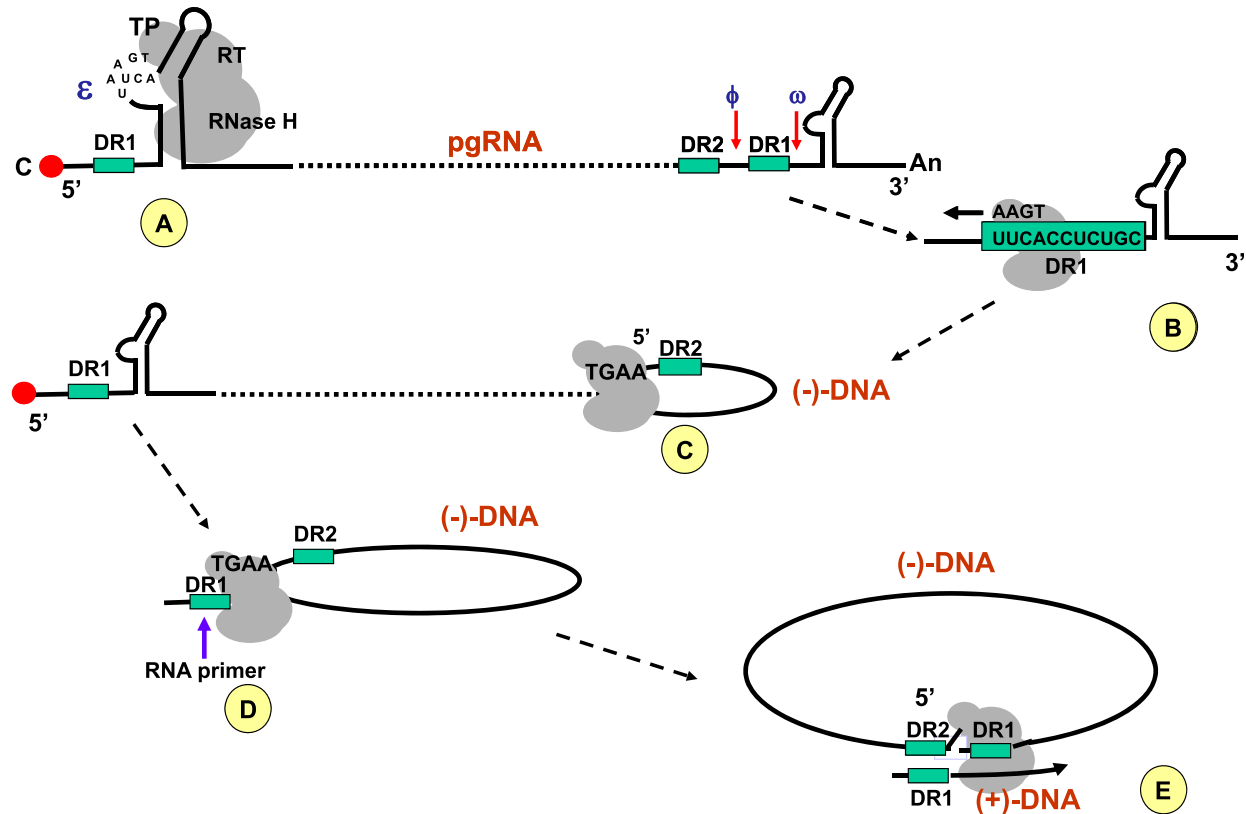


Fig. 5 Replication strategy of the virus. (a) Primer synthesis; (b) translocation and binding to DR1; (c) synthesis of the (–)-DNA strand; (d) RNA primer fragment preserved from the degradation of the pgRNA; (e) (+)-DNA strand synthesis. Reproduced from Karayiannis, P., 2017. Hepatitis B virus: Virology, molecular biology, life cycle and intrahepatic spread. *Hepatology International* 11, 500–508, with permission from Elsevier.

the pgRNA except for the final 11–16 or so ribonucleotides which contain the DR1 region at the 5' of the pgRNA (**Fig. 5(D)**). (+)-DNA strand synthesis is primed by the ribonucleotide fragment which hybridizes with DR2 at the 5' end of the newly synthesized (–)-DNA strand with which it shares homology. This event necessitates a second translocation event which is followed by (+)-DNA strand synthesis that proceeds to the 5' end of the (–)-DNA strand. Circularization facilitated by the short terminal redundancy of the (–)-DNA strand allows template exchange and continuation of (+)-DNA strand synthesis along the 3' end of the (–)-DNA strand (**Fig. 5(E)**). The partially double stranded nature of the HBV genome is the result of the depletion of the nucleotide pool which is not replenished once the maturing nucleocapsids are enveloped by budding through the endoplasmic reticulum membrane. The passage of nucleotides is facilitated through pores within the nucleocapsid at the early stages of replication. As a result the (+)-DNA strand remains incomplete.

Maturation

Early in infection and until sufficient amounts of HBsAg accumulate, nucleocapsids containing rcDNA are shuttled back to the nucleus in order to replenish the cccDNA pool. Hepatocyte nuclei may contain more than 50 copies of cccDNA through this amplification loop (**Fig. 4**). In the final stages, virions bud through the endoplasmic reticulum membrane where HBsAg proteins are already localized, into the lumen, acquiring in the process their outer envelope. This is assisted by the topology and conformational arrangement of L-HBsAg. Studies have shown that about half of the protein produced exposes its N-terminus on the cytosolic side of the ER membrane, facilitating binding of the nucleocapsid and initiation of budding. The rest of L-HBsAg has its N-terminus protruding in the lumen which enables exposure on the surface of the virion and thus accessibility for binding to the NTCP receptor.

Egress

Virions that accumulate in the ER lumen rely on the multivesicular body (MVB) pathway for their export. This pathway depends on the endosomal sorting complex required for transport (ESCRT) system which uses ESCRT-0, -I, -II and -III complexes and associated proteins. SVPs on the other hand bud off independently and are processed via the ER-Golgi pathway.

Epidemiology

In spite of the availability of an effective vaccine against HBV and concerted immunization campaigns in endemic areas, there are still about 291 million chronic carriers of HBV infection worldwide, representing 3.9% of the world's population (estimates of 2016). The prevalence of HBV infection has been dropping in many endemic areas, but still this varies by geographical region. In North Western Europe, North America and Australia the prevalence is between 0.1% and 2%, in the Mediterranean region, Eastern Europe, Middle East, Indian subcontinent, Central and South America it ranges between 2% and 7%, whilst in Africa and the Far East it is between 2% and 14%. About two thirds of the world's chronic carriers reside in Africa and the Far East. The geographical distribution of HBV infection is mirrored by the incidence of HCC in the same regions. In areas of high endemicity, the virus is transmitted perinatally from carrier mothers to their infants, or horizontally from infected siblings and other children in early childhood. In areas of intermediate endemicity apart from perinatal transmission, household and sexual contact, as well as percutaneous exposure, are likely routes of infection. Finally in Western countries transmission is nowadays through sexual contact or intravenous drug use. Following infection, the risk of becoming a chronic carrier is age dependent. Exposure to the virus perinatally and up to one year of age has the highest chronicity rate of 90%, an intermediate risk of 20%–50% for ages up to 5, and 5% for older children and adults.

Genotypes

Whole HBV genome nucleotide sequencing of human isolates from various regions of the world, followed by phylogenetic analysis have identified 10 genotypes designated A–J based on sequence divergence of > 8%. These have characteristic geographical distribution such that Genotypes A and D are frequently found in Africa, Europe, and India, genotypes B and C in Asia, genotype E confined to West Africa, and genotype F in Central and South America. Genotypes G and H have been isolated from Central America and Southern Europe but otherwise their distribution is less clear. Two new genotypes designated I and J have recently been described. Genotype I identified in Vietnam, Laos and Eastern India appears to be an intergenotypic recombinant between genotypes A, C, and G, whilst genotype J was isolated in Borneo, and is a recombinant between genotype C and gibbon HBV. Currently, 44 sub-genotypes also exist and are found in Genotypes A, B, C, D, F and I. These are based on nucleotide divergence of 4%; A1–6, B1–9, C1–16, D1–9, F1–4 and I1–2. Of the other recombinants, B/C and C/D types represent the majority, whilst other intergenotypic recombinants that occur less frequently involve most of the other genotypes.

Likely differences between genotypes in relation to pathogenesis and response to antiviral treatment are beginning to emerge. Genotype C is more frequently associated than B with abnormal liver function tests, lower rates of seroconversion to anti-HBe, higher levels of serum HBV-DNA, cirrhosis and hepatocellular carcinoma. Moreover, there is a better sustained response to interferon treatment in patients with genotype B than those with C, and in patients with genotype A than those with D. Genotype A infection appears to be associated with biochemical remission and clearance of HBV DNA more frequently than genotype D, and has a higher rate of HBsAg clearance compared with genotype D. In contrast to the differences observed in response to interferon therapy, treatment with nucleos(t)ide analogs does not show differential responses between genotypes. Recently, quantitative measurements of HBsAg levels have been associated with potentially favorable responses to interferon treatment in both HBeAg-positive and -negative patients. For example, a reduction in HBsAg levels during treatment, particularly at 12 and 24 weeks, may be predictive of a favorable outcome in the case of Peg-IFN treatment in HBeAg positive patients.

Subtypes

The nucleotide sequence encoding HBsAg is shared by all three envelope proteins. What is more, it contains the main virus neutralizing epitope which is referred to as the “a” determinant region. This epitope is hydrophilic in nature and thus displayed on the surface of the virion. It is conformational, assisted in this by a number of disulfide bridges, and thought to encompass amino-acids 110–160 of the protein. It is shared by all virus isolates. Subtypic specificities also exist and these can be defined either by specific antibodies or by the amino-acids present at positions 122 and 160 of the protein. Lysine (K) or arginine (R) at position 122 confers d or y specificity, respectively, whilst specificities w and r are conferred by the presence of K or R at position 160. Thus each subtype is given a 3 letter code e.g., adw, whereby the a is the group determinant, and d and w based on the amino-acids present at positions 122 and 160 respectively. The w sub-determinant is further divided into w1–w4 specificities. Nine subtypes of the virus are recognized in all (Table 1) depending on the presence of other subtypic determining amino-acids elsewhere in the “a” determinant region.

Variants

Other than the above stable genotypes and subgenotypes, a number of genomic mutations have been described which arise during the natural history of the infection. As described above, the replication step in the HBV life cycle that entails the reverse transcription of the pgRNA replicative intermediate to synthesize the (–)-strand DNA of the virus is prone to errors by the viral

Table 1 Genotype determining amino acid variation over the “a” determinant region of HBsAg, and relationship between genotypes and subtypes.

[illegible]

Source: a = R or K.

Note: G and I share complete amino-acid homology over the "a" determinant. Distinction between the two is possible as R in genotype G is substituted by K at position 24 of S-HBsAg in genotype I. The latter is also distinguished by the presence of H, A, N, V and V at positions 56, 60, 87, 90 and 91 of L-HBsAg respectively and I at position 22 of M-HBsAg.

reverse transcriptase. It is estimated that the HBV genome evolves at a rate of $1.4\text{--}3.2 \times 10^{-5}$ nucleotide substitutions/site/year. As a result, the virus circulates in the serum of infected patients as a population of very closely related genetic variants, referred to as quasispecies. The viability of these variants is dependent on the site of the nucleotide substitutions given the overlapping nature of the ORFs and the localization of the various regulatory elements. Therefore some of these mutations may be deleterious to the virus whilst others may be advantageous, offering either a replication advantage or facilitating immune escape. These are discussed later in the clinical settings in which they have been described.

Clinical Features

HBV Infection may run an asymptomatic, acute icteric (jaundice) or in a minority of patients a fulminant hepatitis course (0.1%–0.5%). As already mentioned the risk of the infection progressing to chronicity is inversely related to age with approximately 5% of adults and 95% of young children becoming chronically infected. Other than age, outcome depends on genetic factors which likely influence the efficiency of the host immune response. These include polymorphisms of the MHC class II glycoproteins involved in viral peptide presentation during induction of the cellular immune response and mannin binding lectins which bind to mannose terminated carbohydrate residues, such as those present on the carboxy-terminus of the pre-S2 region of the middle envelope protein, facilitating phagocytosis.

Acute HBV Infection

The incubation period following exposure is variable ranging from 3 to 6 months and may be dependent on the size of the inoculum and route of exposure. Before onset of symptoms some patients may experience a serum sickness-like syndrome including arthralgia, fever and urticaria. The clinical picture varies from asymptomatic anicteric infection to protracted icterus and, in some patients (< 1%), liver failure (fulminant hepatitis). The acute infection is self-limiting and most patients recover within 1–2 months after the onset of icterus.

Chronic HBV Infection

This is defined as persistence of HBs antigenaemia of longer than 6 months duration and may or may not be accompanied by viraemia and hepatic inflammation. The latter is based on histological examination of liver biopsy material that is followed by assignment of a score for necro-inflammatory activity (out of 18) and stage of fibrosis (out of 6) which are used to assess disease progression and to determine whether a patient needs therapy.

Course of Chronic Infection

Chronic HBV infection is quite variable and is typically characterized by four main phases. These have recently been renamed as the “HBeAg-positive chronic HBV infection” (previously immune tolerant), the “HBeAg-positive chronic hepatitis B” (immune clearance), the “HBeAg-negative chronic HBV infection” (non-replicative), and the “HBeAg-negative chronic hepatitis B” (reactivation phase) which may be seen in some patients, particularly in Southern Europe and the far East (**Fig. 6**). There may be a fifth phase in some patients, referred to as the “HBsAg-negative phase”, which is characterized by absence of serum HBsAg and positivity for anti-HBc with or without detectable anti-HBs. This phase is also known as “occult HBV infection”. During the HBeAg-positive chronic HBV infection phase the patient is HBeAg positive with high levels of HBV-DNA, but with near normal or minimally elevated alanine aminotransferase (ALT) levels and absent or minimal necroinflammation changes in the liver. This phase is more likely in children infected at birth or soon after, and may last 2–3 decades. Patients are highly infectious and HBeAg loss during this phase is very low indeed. Patients from this phase progress to the HBeAg-positive chronic hepatitis B phase which is more commonly seen in those infected in adult life. This phase is characterized by the presence of HBeAg and HBV-DNA the levels of which gradually decline, as well as increased ALT levels associated with necroinflammatory changes in the liver. Loss of HBeAg may be accompanied by an ALT flare, culminating in loss of HBeAg and seroconversion to anti-HBe in a large proportion of patients followed by entry to the HBeAg-negative chronic HBV infection. During this phase plasma HBV-DNA may remain undetectable, ALT levels return to normal whilst annual loss of HBsAg and/or seroconversion may occur spontaneously in 1%–3% of cases. A few patients from this phase and the rest from the previous may proceed to the HBeAg-negative chronic hepatitis B. These patients have raised ALT levels that may fluctuate from time to time, viraemia and continued necroinflammatory activity in the liver that may lead to fibrosis, and faster development of cirrhosis and HCC. Viral DNA integration into the host genome may take place early on in infection and HBsAg expressed from such sequences contributes to the total throughout chronic infection. Moreover, HBV DNA integration may contribute to the development of cancer in the long term. Finally, patients in the fifth HBsAg-negative phase have normal ALT values and usually, but not always, undetectable serum HBV DNA. cccDNA quiescent in the liver may cause viral reactivation following immunosuppression as may be the case for organ transplantation.

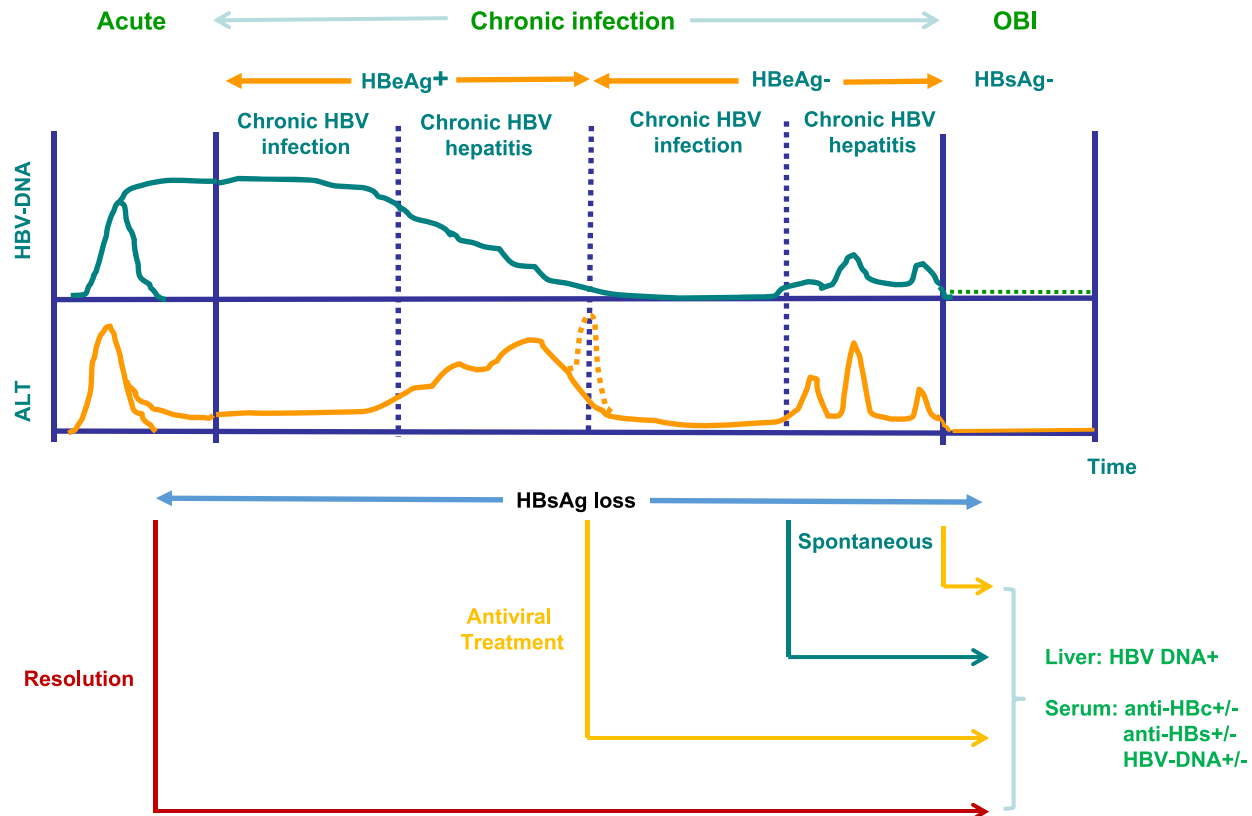


Fig. 6 Diagram of the natural history of HBV infection showing progression from acute to the newly renamed chronic phases of infection. Loss of HBsAg at any time during chronic infection as indicated, may lead to the establishment of occult B infection (OBI).

Mutant Viruses and Chronic Infection

Patients in the HBeAg-negative chronic hepatitis B carry natural mutants of the virus which impact HBeAg production by either reducing levels (core promoter variants) or abrogating completely HBeAg (pre-core variants) production. These variants are selected at the time of, or soon after seroconversion, and become dominant during the reactivation phase. The most common precore mutation is the G1896A substitution, which creates a premature stop codon in the precursor protein from which HBeAg is elaborated. This mutation is located in the stem of the ε encapsidation signal, which it does not destabilize but affords stronger base-pairing with the A1896 change in genotypes with a T at position 1858 of the precore region, such as B, C, D and E. A double mutation in the basal core promoter region (A1762T, G1764A) is suggested to result in decreased transcription of the precore mRNA, with a knock on effect on HBeAg production, whilst pgRNA synthesis remains unaffected. Additional mutations in this region may also contribute to this phenotype.

Pathogenesis

The outcome of infection following exposure to HBV is variable with regards to probability of progression to chronic infection, the likelihood of a favorable outcome after antiviral treatment and the risk of progression to unwelcome chronic sequelae such as cirrhosis and HCC. All of these are dependent on viral, environmental, demographic and clinical factors, but also host genetics and other factors which have been identified such as defective viral mutants (pre-S/S deletion mutants), and more recently spliced mRNAs, miRNAs and exosomal vesicles.

Acute self-limiting infection is characterized by necroinflammation in the liver, consistent with the rise in transaminase levels. It seems that the innate immune response in the initial stages of infection is not robust enough while viraemia decline prior to symptom appearance may indicate non-cytolytic suppression of virus replication, mediated perhaps by $\text{IFN}\gamma$ secreted by NK and NKT cells. Studies in animal models have indicated that an HBV-specific cytotoxic T cell response (CTL) is essential for clearance of infected hepatocytes. Such CTL responses are strong and polyclonal in nature, directed against viral proteins such as HBeAg, and are the prime cause of immunopathology in the liver as the virus is not directly cytopathic. Clearance of free virions and prevention of attachment are dependent on the development of neutralizing anti-HBs.

In chronic HBV infection on the other hand, the virus and high levels of HBsAg or HBeAg may disrupt maturation of antigen presenting cells such as dendritic cells (DCs), thus inducing tolerance. Studies have shown that NK cells and plasmacytoid DCs

(pDCs) are deficient or dysfunctional. Monocytes/macrophages, NK, NKT and T cells (including regulatory T cells) infiltrate the liver and cause chronic inflammation, as a result of induction of pro-inflammatory cytokines, resulting in immunopathology. CTL responses are narrow and monospecific and not sufficient to clear infected hepatocytes. Such cells are deficient in proliferation and production of cytokines such as IFN γ . Intrahepatic CD8 $^{+}$ T cells express programmed cell death protein 1 (PD-1) which suggests that they may become exhausted in the presence of cognate antigen expression.

Long term infection with HBV associated with active viral replication can lead to cirrhosis and HCC. HBV is responsible for 30% of cases with cirrhosis and almost 50% of cases with HCC. The annual risk of HCC in patients with cirrhosis has been reported to be 2%–5%. Recurrent liver necro-inflammation due to the immunopathological events described above, accelerate hepatocyte turnover and deposition of collagen initially but gradually leading to fibrosis, cirrhosis and eventually HCC. This is a long term process, multifactorial in nature and implicating both host and viral factors. These include high HBV DNA levels (>2000 IU/ml), high transaminase levels, prolonged HBeAg positivity, genotypes B, D and F, precore and basal core promoter variants, male sex and older age. Co-infections with HIV and hepatitis D virus, comorbid conditions such as diabetes mellitus and obesity, abuse of alcohol, carcinogen exposure such as smoking and aflatoxin B1, are also contributing factors. Integration of the viral genome into hepatocyte DNA can lead to chromosomal instability and expression of functional preS/S and X proteins which are truncated at their extreme carboxyl end. PreS/S proteins with deletions and mutations in HBx have been associated with increased risk of HCC. HBx is thought to promote viral replication and is a known transactivator *in vitro*, it inhibits TNF- α , Fas-induced apoptosis, whilst it promotes fibrogenesis and remodeling of extracellular matrix. In addition, spliced HBV mRNAs have been implicated in tumorigenesis as have microRNAs (miRNAs) which are involved in regulating gene expression, cell cycle progression, differentiation and apoptosis. Finally, exosomal vesicles are involved in virion export, cell-to-cell spread of the virus, miRNA export etc., which may be additional contributors to viral pathogenesis.

Diagnosis

Acute or chronic infection with HBV is based on serological detection of HBsAg in serum. Its persistence longer than 6 months indicates progression to chronic infection, whilst appearance of anti-HBs (antibody to HBsAg) indicates recovery from infection, or acquired immunity after prophylactic vaccination. Detection of HBeAg indicates active viral replication, as does the detection of serum HBV-DNA by qualitative or quantitative polymerase chain reaction (PCR) tests. The latter is particularly useful in monitoring efficacy of antiviral treatment. Seroconversion to anti-HBe occurs after recovery from acute infection, and less often during the chronic phase, either spontaneously or after therapeutic intervention. In the latter case, this leads to a quiescent phase of disease that can be long-term, but does not necessarily mean that virus replication has stopped entirely (see below). Antibody to core antigen (anti-HBc) of IgM class at high level is a marker of acute infection, whereas total anti-HBc (primarily IgG) is detectable following acute and throughout chronic infection.

Treatment

Antiviral treatment is the only means available to prevent progression of chronic infection to cirrhosis and eventually HCC. The primary aim is the long term suppression of viral replication to undetectable levels. Moreover, in HBeAg positive patients the loss of this marker and preferably the development of the respective antibody (anti-HBe) is a favorable outcome as it denotes immune control of the virus. These responses are accompanied by normalization of biochemical markers. A secondary outcome which is achieved less often is the loss of HBsAg with or without development of anti-HBs. Decision to treat is based on a viral load >2000 IU/ml, an ALT level above the ULN and the presence of moderate necroinflammation and fibrosis. Patients with HBV DNA levels >20000 and ALT levels more than 2xULN should be treated without histological assessment. Two approaches are currently available for antiviral treatment of chronic HBV infection; nucleos(t)ide analogs (NAs) and immunomodulators, such as Pegylated Interferon α (PegIFN α).

NAs are lamivudine (LAM), adefovir dipivoxil (ADV), entecavir (ETV), telbivudine (TBV), tenofovir disoproxil fumarate (TDF) and tenofovir alafenamide (TAF). Of these ETV, TDF and TAF are the drugs of choice as they have a very high barrier to HBV resistance as opposed to the other three which have a low resistance barrier. NAs are taken orally and act as chain terminators at the stage of DNA synthesis. They are chemically modified so that their incorporation into nucleic acids prevents chain elongation and therefore cause premature termination of synthesis. These lend themselves for long-term treatment which effectively leads to suppression of HBV DNA to undetectable levels very quickly which is maintained whilst the patient remains on treatment. They are quite safe and can be used in any HBV infected patient whether HBeAg-positive or -negative, but also various subgroups including those with decompensated cirrhosis, liver transplant patients, acutely infected or those with severe flares during chronic HBV. NAs can also be used in patients with occult HBV to prevent reactivation if they are due to be immunosuppressed.

5 years of ETV or TDF treatment in HBeAg-positive patients results in nearly 99% and 97% cumulative virologic response and near 53% and 49% HBeAg loss respectively. HBsAg loss was achieved in 10% and HBsAg seroconversion in 8%. For HBeAg-negative patients, treatment with ETV and TDF over the same period achieved virological and biochemical responses of 98% and 95% and 99% and 88%, respectively. Very few such patients (1%) achieved HBsAg loss during treatment exceeding 5 years duration. Data on TAF treatment is available for the first two years since its licensing with comparable results to TDF. In HBeAg-

positive patients, the rates of virologic response at week 96 was 75%, whilst HBeAg loss and anti-HBe seroconversion were achieved in 22% and 18%, respectively. Only 1% of patients cleared HBsAg. In HBeAg-negative patients, TAF achieved a virologic response in 90% of them by week 96.

PegIFN α acts by promoting cytotoxic T-cell activity for lysis of infected hepatocytes and by stimulating cytokine production for control of viral replication. Weekly intramuscular injections in HBeAg-positive patients for 1 year can lead to seroconversion to anti-HBe in 32% of patients. This is sustainable in over 95% of patients. In HBeAg-negative patients the intention is to suppress HBV- DNA levels followed by ALT normalization a goal which is achievable in about 44% of them. However, durability is only possible in around 20%. To manage such relapses following initial interferon treatment, nucleos(t)ide therapy should be started and continued long term in both HBeAg positive and negative patients not achieving a sustained response on PegIFN. Rates of HBsAg loss following 12 months of treatment are 3%–7%. Overall, among sustained responders whether HBeAg-positive or -negative 30% will clear HBsAg in the long-term.

Resistance

Treatment with NAs of low genetic barrier to resistance over time can lead to increasing resistance levels. In the case of LAM for example, this is about 24% at year 1 rising to >70% by year 5. Adefovir resistance on the other hand is delayed being 0% at year 1, 3% at year 2 and rising to 29% by year 4. Entecavir resistance has only been seen so far in patients with lamivudine resistant strains. Resistance is associated with the acquisition of amino-acid substitutions in the rt domain of the HBV polymerase which has 6 spatially separated subdomains from A-F and which contribute to nucleic acid synthesis. Most importantly, amino-acid substitutions in subdomain C which contains the characteristic YMDD (tyrosine-methionine-aspartate-aspartate) motif of the catalytic site of the enzyme can cause resistance. Resistance to LAM relates to substitution of the methionine (M) at position 204 to either valine (YVDD, rtM204V) or isoleucine (YIDD, rtM204I). The rtM204V substitution mutation is almost always associated with a second one involving a substitution of leucine with methionine at position 180 (rtL180M) in subdomain B. ADV resistance is conferred by mutations rtN236T in subdomain D and rtA181V in subdomain B, whilst the small number of entecavir resistant cases have in addition to the lamivudine resistant substitutions additional ones such as rtI169T, rtT184G, rtS202I and rtM250V.

Prevention

Protection against HBV infection is afforded by prophylactic vaccination which was first introduced in the early 1980s. Current vaccines consist of recombinantly expressed HBsAg in yeast such as *Saccharomyces cerevisiae*. In adults, the vaccine is administered through intramuscular injection in the deltoid in a three dose schedule at 0, 1 and 6 months. In countries of high and medium sero-prevalence, universal vaccination programs have been instituted, and HBV vaccination is recommended for infants born to carrier mothers preferably within the first 12 h of life, given together with hepatitis B immune globulin (HBIG). Measuring anti-HBs levels 1–4 months after the last dose of the vaccine ensures development of protective immunity defined as a level >10 mIU/ml.

Response rates to the vaccine in healthy individuals range from 90% to 95%, whilst in haemodialysed and hemophiliac patients these are lower (70%). The durability of the response to the vaccine with persistence of anti-HBs levels >10 mIU/ml has extended beyond 12 years in up to 80% of individuals immunized at a young age. Booster immunizations therefore may not be necessary for at least 10 years after vaccination. However, high risk professions such as healthcare personnel and individuals in high risk groups need to monitor antibody levels perhaps every 5 years.

The impact of HBV vaccination on infection prevalence has become apparent from epidemiological studies around the world and in countries of medium to high endemicity. They have recorded a dramatic drop in HBV prevalence. For example, in Taiwan, 15 years after the start of the vaccination program, the prevalence of HBsAg in children under 15 years of age had decreased from 9.8% in 1984 to 0.9% in 1999 and by 2014 to 0.3%. Similarly, the incidence of HCC has been on the decline from 0.7 per 10,000 children between 1981–1986, to 0.57 and 0.36 between 1986–1990 and 1990–1994, respectively.

In spite of vaccination, breakthrough infections in vaccinated subjects have been detected through monitoring in spite of the presence of a satisfactory antibody levels. Such infections may be due to the wild type virus or mutant viruses, the commonest of which involves a G145R substitution in the “a” determinant region of HBsAg. This and a number of other mutations in this region, which can occur in combination also, alter the antigenic structure of this epitope, causing failure in its recognition by neutralizing antibody. Mutant viruses have also been described in the setting of liver transplantation where use of HBIG or monoclonal anti-HBs is recommended in an attempt to prevent infection of the new liver graft.

Further Reading

- Cornberg, M., Wong, V.W., Locarnini, S., *et al.*, 2017. The role of quantitative hepatitis B surface antigen revisited. *Journal of Hepatology* 66, 398–411.
Lampertico, P., Agarwal, K., Berg, T., *et al.*, 2017. EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection. *Journal of Hepatology* 67, 370–398.

- Ganem, D., Prince, A.M., 2004. Hepatitis B virus infection – Natural history and clinical consequences. *New England Journal of Medicine* 350, 1118–1129.
- Ginzberg, D., Wong, R.J., Gish, R., 2018. Global HBV burden: Guesstimates and facts. *Hepatology International* 12, 315–329.
- Hadziyannis, S.J., Papatheodoridis, G.V., 2006. Hepatitis B e antigen-negative chronic hepatitis B: Natural history and treatment. *Seminars in Liver Disease* 26, 130–141.
- Jazayeri, S.M., Alavian, S.M., Dindoost, P., Thomas, H.C., Karayiannis, P., 2014. Molecular variants of hepatitis B surface antigen (HBsAg). In: Thomas, H.C., Lok, A.S.F., Locarnini, S.A., Zuckerman, A.J. (Eds.), *Viral Hepatitis*, fourth ed. London: Wiley Blackwell, pp. 107–126.
- Karayiannis, P., Carman, W.F., Thomas, H.C., 2014. Molecular variants of the precore, core and core promoter regions of hepatitis B virus, and their clinical significance. In: Thomas, H.C., Lok, A.S.F., Locarnini, S.A., Zuckerman, A.J. (Eds.), *Viral Hepatitis*, fourth ed. London: Wiley Blackwell, pp. 127–142.
- Karayiannis, P., 2017. Hepatitis B virus: Virology, molecular biology, life cycle and intrahepatic spread. *Hepatology International* 11, 500–508.
- O'Brien, T.R., Yang, H.I., Groover, S., Jeng, W.J., 2019. Genetic factors that affect spontaneous clearance of Hepatitis C or B virus, response to treatment, and disease progression. *Gastroenterology* 156, 400–417.
- Rehermann, B., Thimme, R., 2019. Insights from antiviral therapy into immune responses to Hepatitis B and C virus infection. *Gastroenterology* 156, 369–383.

Relevant Websites

- <https://www.cdc.gov/hepatitis/hbv/index.htm>
Hepatitis B Information.
- www.hepb.org
Hepatitis B Foundation.
- <https://www.who.int/news-room/fact-sheets/detail/hepatitis-b>
Hepatitis B
World Health Organization.

Hepatitis C Virus (*Flaviviridae*)

Ralf Bartenschlager and Keisuke Tabata, Heidelberg University, Heidelberg, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of R. Bartenschlager, S. Bühler, Hepatitis C Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00416-7](https://doi.org/10.1016/B978-012374410-4.00416-7).

Glossary

Chronic hepatitis C A liver disease with continued presence of HCV RNA in the blood for six months or more after primary infection.

DAAs Direct acting antivirals, i.e., drugs targeting distinct viral proteins and inhibiting their function, thus disrupting virus replication.

HCV viremia Presence of HCV RNA in the blood of an infected individual.

Internal ribosome entry site (IRES) A folded RNA element that allows for cap-independent initiation of RNA translation. The specific type of IRES determines which subset of the canonical translation initiation factors is required for RNA translation.

Polyprotein A precursor polypeptide from which distinct proteins are produced by proteolytic cleavage.

Positive-strand RNA virus A virus with a single-strand RNA genome that has the polarity of a cellular mRNA and therefore, can be used by host cell ribosomes for RNA translation to produce viral proteins. For that reason, the positive-strand RNA genome is infectious, i.e., it is sufficient to produce virus progeny in a permissive host cell.

Replicon A DNA or RNA molecule capable of self-replication in a cell or an adequate in vitro system (e.g., cell lysate).

Sustained virological response (SVR) Continuous absence of virus from the serum, starting at a given time point after cessation of antiviral therapy. For instance, in HCV infection SVR6 means absence of HCV RNA 6 months after termination of therapy.

Classification

HCV has been classified as a member of the genus *Hepacivirus* and is grouped together with the genera *Pestivirus*, *Pegivirus*, and *Flavivirus* in the family *Flaviviridae*. Based on genomic heterogeneity, 8 genotypes having more than 30% nucleotide sequence divergence and more than 80 subtypes differing from each other by 10%–30% at the nucleotide sequence level have been defined. Subtypes are designated by lower case letters following the number of the genotype (e.g., genotype 1 subtype b = 1b). Genotype 1 is the most prevalent worldwide (~45% of all HCV cases), followed by genotype 3 (~30% of all HCV cases). Genotypes 2, 4, and 6 are responsible for ~23% of all HCV infections, whereas infections with genotypes 5 constitute ~1% of all HCV cases worldwide. Currently, there is no information available about the global prevalence of genotype 7.

Infections with genotype 1–3 are found in almost all countries, whereas HCV genotypes 4–7 are to a large extent restricted to distinct geographical regions like Africa and Middle East (genotype 4), South Africa (genotype 5), and Southeast Asia (genotype 6). Genotype 7 has been isolated in Canada and Belgium from patients possibly infected in Central Africa. A novel genotype 8 was recently identified in 4 patients that migrated from Punjab State, India to Canada. Individual genotypes have not been ascribed to particular disease manifestations except for a higher prevalence of hepatosteatosis for patients infected with genotype 3 viruses. However, in the case of interferon (IFN)-based therapy, the infecting genotype is an important predictor of therapy outcome, although with the implementation of IFN-free antiviral drugs, the genotype has become much less relevant (see below).

Virion Structure and Properties

HCV particles are enveloped and spherical with a very pleomorphic structure. They are most likely composed of a single copy of the viral RNA genome that associates with multiple copies of the core protein (Fig. 1). It is unclear whether a regular nucleocapsid is formed or a more unstructured ribonucleoprotein complex. Surrounding the capsid is a lipid envelope, probably derived from the endoplasmic reticulum (ER). Into this envelope, multiple copies of a heterodimeric complex of the two envelope glycoproteins E1 and E2 are embedded. A hallmark of HCV particles is their association with various apolipoproteins of the host cell, most notably apolipoprotein E (ApoE). This association occurs within the cell, perhaps during assembly or virion egress, but also outside of the cell with released virus particles binding to apolipoprotein-containing very low-density lipoproteins (VLDL), LDL and high-density lipoproteins (HDL). This association is mediated most likely by the viral envelope glycoproteins binding to various apolipoproteins. Because of this tight association, infectious HCV particles have an irregular structure, a very low density and therefore, they have been designated as lipoviro-particles (Fig. 1). Moreover, HCV particles have a unique lipid composition containing an unusually high amount of cholesterol esters and a surprisingly low amount of phosphatidylethanolamine, thus resembling the composition of (V)LDL, but being very different from the lipid composition of e.g., HIV-1 particles. Because of the

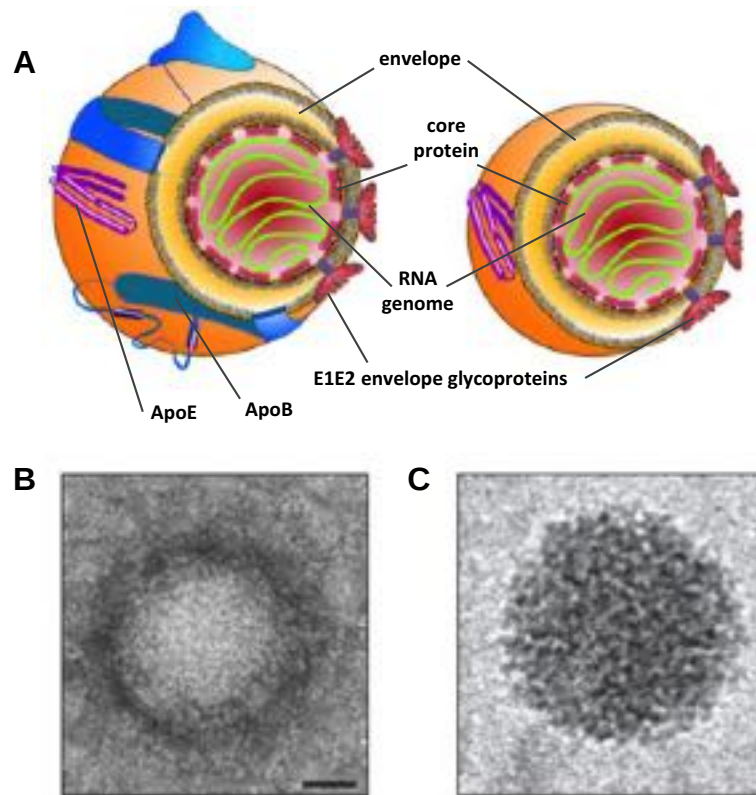


Fig. 1 *Morphology and composition of HCV particles.* (A) Schematic presentation of the HCV particle structure and composition. (B) Electron micrograph of negatively stained HCV particles. Virions were produced in cell culture and captured using electron microscopy grids coated via protein A with anti-HCV antibodies. Scale bar: 20 nm. (C) Cryo-electron microscopy image of an HCV particle at 78,000 $\times$ magnification. Virions are heterogeneous in size ranging from 45 to 86 nm in diameter, with a mean diameter of 68 nm. The illustration is adapted from Bartenschlager, R., Penin, F., Lohmann, V., André, P., 2011. Assembly of infectious hepatitis C virus particles. *Trends in Microbiology* 19, 95–103. Copyright 2010, Elsevier Ltd. Images are reproduced from Catanese, M.T., Uryu, K., Kopp, M., *et al.*, 2013. Ultrastructural analysis of hepatitis C virus particles. *Proceedings of the National Academy of Sciences of the United States of America* 110, 9505–9510. with permission of the authors.

association of HCV particles with apolipoproteins and (V)LDL and HDL components, size, density and composition of the virions differ, depending on the culture conditions. HCV produced in the human hepatoma cell line Huh7, which supports robust virus replication, have a density of 1.1 g/mL whereas virions produced in more authentic cell systems such as Huh7 cells cultured in human serum or primary human hepatocytes can have a much lower density (close to 1.0 g/mL). The association with lipoprotein components reduces sensitivity of the virus to neutralization by antibodies and increases virion infectivity.

Genome Organization and Viral Proteins

The genome of HCV is a single-stranded RNA molecule of positive polarity and has a length of ~ 9600 nucleotides (**Fig. 2(A)**). It is flanked by two highly structured non-translated regions (NTRs). The about 340 nucleotides long 5' NTR contains an internal ribosome entry site (IRES) (**Fig. 2(B)**), mediating cap-independent RNA translation giving rise to an around 3,000 amino acids long polyprotein. The 3' NTR is composed of an about 40 nucleotides long variable region, a polypyrimidine tract (heterogeneous in length) and a 98 nucleotides long highly conserved 3' terminal X-tail (**Fig. 2(B)**). Both the 5' NTR and the 3' NTR contain cis-acting RNA elements (CREs) that are required for viral replication. Within the 3' terminal part of the NS5B gene 3 additional CREs are localized (5BSL3.1 to 5BSL3.3) with 5BSL3.2 being absolutely required for viral RNA replication by forming a long-distance RNA-RNA interaction with the middle stem-loop in the X-tail. Likewise, 4 CREs have been identified in the core protein coding region with two of them (SL47 and SL87) contributing to HCV RNA translation and robust replication in cell culture. The importance of the structural integrity of SL87 for robust HCV replication was confirmed *in vivo* using mice xenografted with primary human hepatocytes and in chimpanzees.

Flanked by the two NTRs is the coding region of the HCV polyprotein that is cleaved proteolytically into 10 products (**Fig. 2(A)**). The structural proteins core, E1 and E2 are located within the N-terminal part of the polyprotein preceding the p7 protein, which appears to be an ion channel and therefore, was grouped into the viroporin protein family. The non-structural (NS) proteins NS2, NS3, NS4A, NS4B, NS5A, and NS5B are encoded in the remaining part of the polyprotein. The NS2 protein, together with the N-terminal protease domain of NS3, is responsible for the autocatalytic cleavage of the NS2-NS3 junction (**Fig. 2(A)**). NS2 is a dimeric cysteine

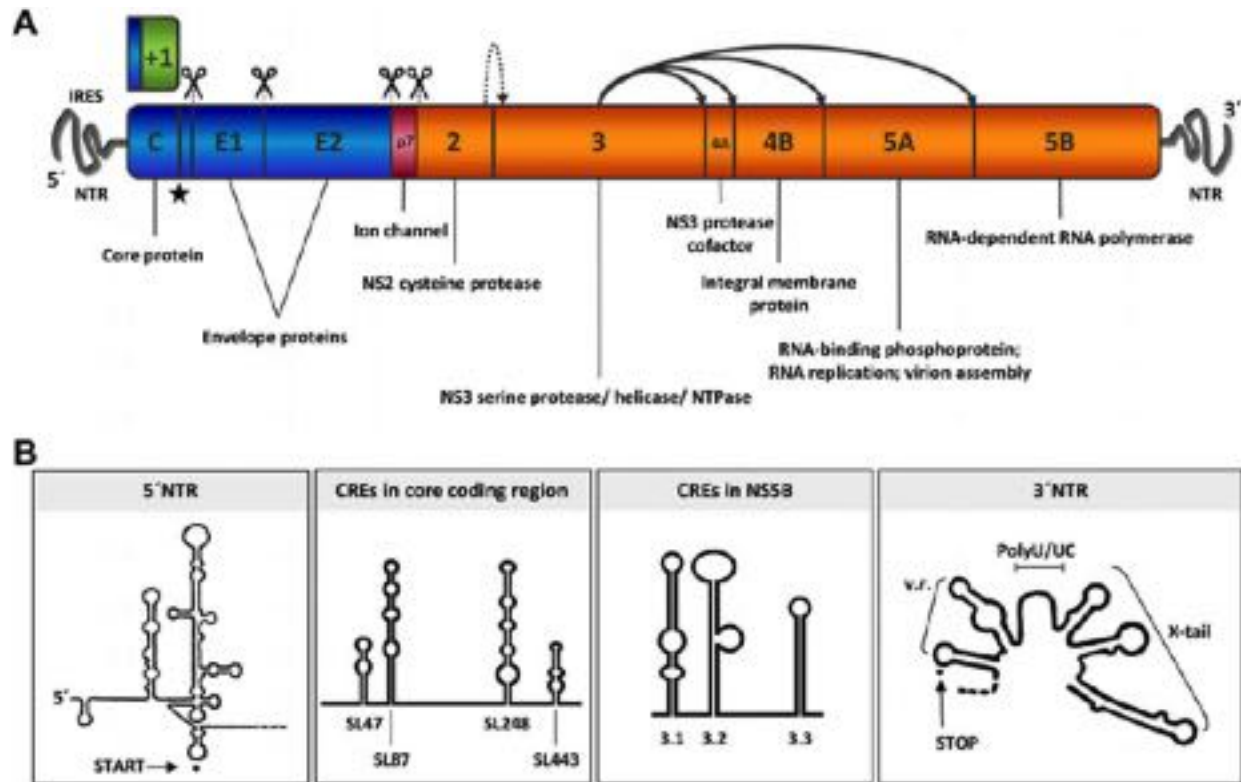


Fig. 2 HCV genome organization, gene products and *cis*-acting RNA elements (CREs). (A) The HCV polyprotein coding region is flanked by two non-translated regions (5' and 3' NTR) represented in dark gray. The 5' NTR contains an internal ribosome entry site (IRES). Polyprotein cleavage by cellular signal peptidases is indicated by scissors, whereas cleavage by the viral NS3 serine protease is indicated by solid arrows. The cleavage at the NS2-NS3 junction by the NS2 cysteine protease is indicated by a dotted arrow. The star indicates the cleavage site for the cellular signal peptide peptidase removing the carboxy-terminal region of the core protein. Functions of the individual cleavage products are specified in the figure. (B) Schematic representation of CREs. The 5' NTR (left panel) contains the IRES as well as structures important for viral RNA replication. The middle left panel depicts the four CREs in the core coding region. Three CREs are located in the 3'-terminal part of the NS5B coding region (middle right panel). Stem-loop 5BSL3.2 forms a long-distance RNA-RNA interaction essential for RNA replication with the loop of the middle stem-loop in the X-tail. The right panel displays the organization of the 3' NTR with the variable region (v.r.), the poly U/UC-tract and the X-tail. Positions of start and stop codons are indicated with dots.

protease with a composite active site. The N-terminal NS3 domain carries a serine-type protease that after association with its cofactor NS4A, cleaves the four junctions between NS3 and NS5B. Two additional enzymatic activities (RNA helicase and nucleoside triphosphatase) reside in the carboxy-terminal two thirds of NS3.

Structural integrity of intracellular membranes is altered by NS4B that has a complex and dynamic membrane topology. NS4B cooperates with NS5A, which is a zinc-binding phosphoprotein, composed of the following domains (from the N- to the C-terminus): an amphipathic α -helix serving as a peripheral membrane anchor, the structured domain I, low complexity sequence I, domain II, low complexity sequence II and domain III. Interestingly, domains II and III are intrinsically unfolded, but they also contain segments of transient secondary structure. The X-ray crystal structure of domain I was resolved and found to form various types of homodimers with one of them forming a basic RNA binding groove. NS5A is phosphorylated by several cellular kinases including the casein kinase I isoform alpha (CK1 α), CKII, polo-like kinase 1 (PLK1) and calmodulin-dependent kinase II gamma (CaMKII γ) and CaMKII δ . The phosphorylation state of NS5A affects its interaction with several host cell proteins such as VAP-A/B, arguing that distinct phosphorylation patterns confer different functions to NS5A, depending on an associating partner molecule. In this respect, NS5A exerts multiple functions; it is essential for viral RNA replication and virion assembly, but it is also involved in counteracting antiviral defenses such as the IFN response. The exact mechanism by which NS5A contributes to RNA replication has not been revealed, but it was found that the sole expression of NS5A can induce the formation of double membrane vesicles (DMVs), which are the presumed sites of HCV RNA replication (Fig. 3, inset). Moreover, drugs targeting NS5A block DMV formation, suggesting that the induction of the HCV replication organelle, designated "membranous web", is one way by which NS5A contributes to HCV RNA replication. In addition, NS5A interacts with the core protein, for which phosphorylation of the carboxy-terminal region of NS5A is required. Thus, phosphorylation of NS5A at this site appears to trigger the assembly of infectious HCV particles, a process that may be supported by the interaction between RNA-loaded NS5A and the core protein.

The 68-kDa protein NS5B is an RNA-dependent RNA polymerase (RdRp). It is a tail-anchored membrane protein with a structural organization similar to that of other polymerases, i.e., composed of palm, finger and thumb subdomains. However,

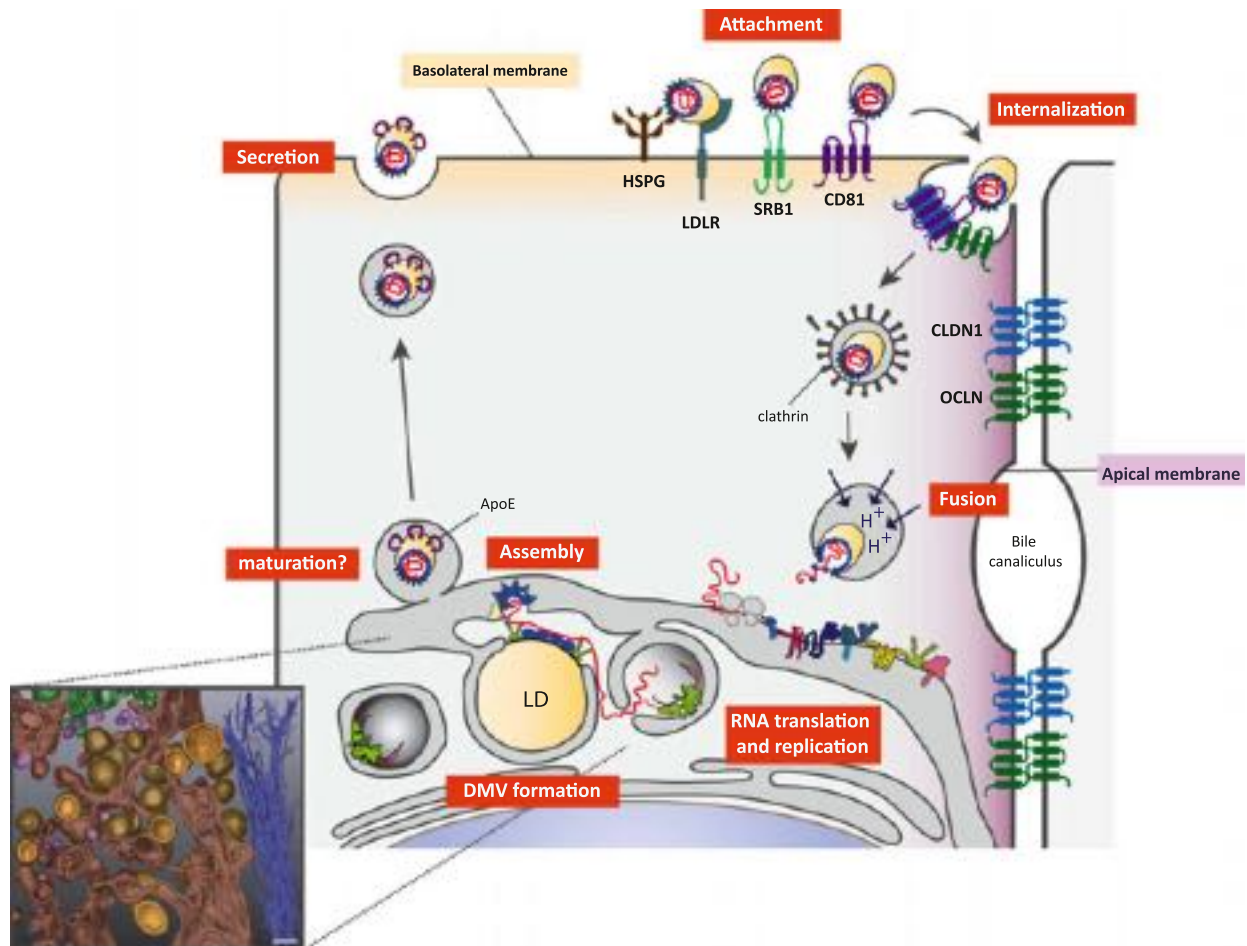


Fig. 3 *HCV replication cycle.* (Attachment) HCV binds to one or several receptors on the surface of the host cell. HSPG, SRB1, CD81 and presumably LDLR are required for or contribute to virus binding and entry. (Internalization) The virus particle enters the cell via clathrin-mediated endocytosis likely at tight junctions where Claudin-1 (CLDN1) and occludin (OCLN) reside. (Fusion) After a low-pH mediated fusion step from within an acidic endosome and uncoating, the HCV genome is released into the cytoplasm of the host cell. (RNA translation and replication) The viral RNA genome (red line) is translated at the rER and a membranous web, presumably originating from ER membranes in close proximity to lipid droplets (LD), is formed (DMV formation). DMV is the site of viral RNA amplification, which occurs via negative-strand RNA intermediates (blue line). Newly synthesized positive-strand RNA is used either for translation, or for replication, or the RNA is packaged into nascent capsids (Assembly). This most likely occurs at the ER where E proteins are retained with virus particles forming via budding into the ER lumen. (Maturation) During or after envelopment virus particles acquire ApoE, which is thought to be required for virion infectivity (indicated with the question mark). The association of HCV particles with ApoE and other apolipoproteins, occurs intra- and extracellularly. (Secretion) Virions are release after fusion of the transport vesicle with the plasma membrane. A three-dimensional reconstruction of the membranous web in HCV-infected cells is shown on the bottom left. The inner DMV membrane is indicated in light brown whereas ER is depicted in dark brown. Single-membrane vesicles, cytoskeletal filaments and Golgi cisternae are shown in violet, blue and green, respectively. Scale bar: 100 nm. The illustration is adapted from Neufeldt, C.J., Cortese, M., Acosta, E.G., Bartenschlager, R., 2018. Rewiring cellular networks by members of the Flaviviridae family. *Nature Reviews Microbiology* 16, 125–142. The inset image is reproduced from Romero-Brey, I., Merz, A., Chiramel, A., *et al.*, 2012. Three-dimensional architecture and biogenesis of membrane structures associated with hepatitis C virus replication. *PLOS Pathogens* 8, e1003056.

NS5B differs from most other RdRps by having a fully encircled active site, which is due to tight interactions between the finger and the thumb subdomain.

Apart from the polyprotein a heterogeneous group of HCV proteins are expressed either by ribosomal frameshifting into the + 1 ORF or by internal translation initiation. The resulting proteins, collectively designated core + 1 (Fig. 2(A)), are not essential for replication and virus production in cell culture and they appear to be dispensable *in vivo*.

Life Cycle

Hepatocytes are the primary target cells for HCV. However, viral RNA was also detected in PBMCs and bone marrow cells, but it is unclear whether a productive infection is possible in these cell types. In cell culture, HCV RNA replication was demonstrated in non-liver

cells like human T and B cell lines, human embryonic kidney cells (HEK293) and even neuroepithelioma-derived cell lines arguing that some of the extrahepatic manifestations observed with chronic hepatitis C patients, such as cryoglobulinemia or central nervous system abnormalities, might be caused by infection of these cell types. Furthermore, certain mouse cell lines can support the replication of engineered HCV replicons demonstrating that the viral replication machinery is also functional in a murine host cell environment.

In hepatocytes, HCV infection starts with virion binding to host cell attachment factors such as heparan sulfate proteoglycans (HSPG) and low-density lipoprotein receptors (LDLR) located at the basolateral membrane (Fig. 3). Subsequently virus particles interact with the scavenger receptor class B member 1 (SRB1) and the tetraspanin CD81, which guide the lateral translocation of virion - receptor complexes towards tight junctions upon activation of the epidermal growth factor receptor (EGFR) and the EGFR/Shc/Grb2/Hras signaling pathway. Following these events, CD81 associates with claudin-1 (CLDN1) and occludin (OCLN) to trigger virus internalization through the clathrin-dependent pathway and a subsequent fusion step from within an acidic endosomal compartment (Fig. 3).

Upon release of the RNA genome into the cytoplasm, the polyprotein is synthesized by IRES-dependent RNA translation occurring at the rough ER (rER) where host cell signal peptidases, signal peptide peptidases and viral proteases catalyze polyprotein cleavage. During or after polyprotein cleavage an ER-derived membranous web is formed where the positive-strand RNA genome is amplified via a negative-strand RNA intermediate (Fig. 3). Newly synthesized positive-strand RNAs are used for the synthesis of multiple polyprotein copies (probably > 1000 protein copies per RNA molecule), or they serve as templates for further RNA synthesis, or they are incorporated into assembling virions. Viral envelope (E1-E2) glycoproteins are retained at rER membranes suggesting that viral envelopes are generated by budding into the lumen of this organelle. The envelopment of virus particles depends on components of the ESCRT (Endosomal Sorting Complex Required for Transport) machinery. The ESCRT-0 complex component HRS (hepatocyte growth factor-regulated tyrosine kinase substrate) interacts with HCV NS5A and NS2 proteins, at least in case of the latter via recognition of polyubiquitinated residues. Other ESCRT proteins may interact with HRS to coordinate HCV assembly, as the depletion of the components of ESCRT-III or VPS4-VTA1 complexes causes assembly defects. Given the strong functional relationship between HCV and lipoproteins, it was speculated that HCV particles exit the cell via the canonical secretory pathway, comparable to VLDL. However, recent evidence suggests that HCV may also use an alternative secretion route bypassing the Golgi apparatus. In either case, virus particles are released after fusion of the transport vesicle with the plasma membrane as cell-free virions, or they are transmitted directly to neighboring cells via a cell-to-cell spread mechanism.

Formation of the HCV Replication Organelle

The most abundant components of the membranous web are double membrane vesicles (DMVs) that have an average diameter of about 200 nm (Fig. 3, inset). Although the sole expression of NS5A suffices to induce DMVs, the efficiency of DMV formation is greatly increased upon the expression of an NS3-5B polyprotein. Morphologically, DMVs correspond to protrusions from the ER membrane with a fraction of DMVs being connected via the outer DMV membrane to the ER membrane. DMVs isolated from HCV-replicating cells contain active viral replicase arguing that DMVs are the site where HCV amplifies its genome. However, only a minority of DMVs has an opening connecting the DMV lumen with the cytoplasm, whereas the majority of DMVs appears to be completely "sealed". Thus, viral RNA produced inside DMVs can either be released into the cytoplasm as long as the DMVs are not completely sealed or DMVs contain specific "pores" spanning both membranes and allowing RNA export out of the DMV interior. Interestingly, similar DMV-type viral replication organelles have been described for other positive-strand RNA viruses, including members of the distantly related *Coronaviridae*, *Arteriviridae*, and *Picornaviridae* families.

At least in the case of HCV, the formation of DMVs may be linked to autophagy as deduced from the impaired RNA replication and reduced DMV number upon depletion of distinct factors of the autophagy machinery. Given the structural similarity of autophagosomes and DMVs, a link to autophagy appears plausible. However, in the case of poliovirus, it was shown that DMVs form via "secondary" invagination of single membrane vesicles, which appear to be the primary membrane structures induced by poliovirus early after infection. In contrast, single membrane vesicles have been observed only rarely in HCV-infected Huh7 cells, but they have been reported as sole structures in hepatocytes of HCV-infected individuals. Thus, it remains to be determined whether DMVs are *bona fide* HCV replication sites formed only in hepatoma cell lines or whether they are also found in vivo.

Although HCV-induced DMVs are derived from the ER membrane, their lipid composition differs substantially from the originating membrane by having much higher levels of cholesterol and sphingolipids. This change in lipid composition is achieved by the NS5A-mediated recruitment and activation of the lipid kinase phosphatidylinositol 4-kinase III α (PI4KA) that is enriched at (NS5A-containing) membranes to produce excessive amounts of PI4P. In addition, NS5A binds to and recruits, probably via VAP-A/B, lipid-transfer proteins such as oxysterol-binding protein that may deliver cholesterol into the DMV membrane in exchange for PI4P. Probably in a similar way, the glucosylceramide transfer protein FAPP2 may be recruited to DMVs to release lipids in exchange for PI4P into the DMV membrane. In this way cholesterol and sphingolipids would be enriched leading to the formation of lipid rafts likely required for HCV replicase activity and possibly the assembly of (lipidated) HCV particles.

Host Cell Factors Contributing to Various Steps of the HCV Replication Cycle

During the last few years, numerous host cell factors affecting the HCV replication cycle have been identified, but only a few can be mentioned here. The first one is microRNA-122 (miR122). The best-known role of miRNAs, which are short single-strand, non-coding

RNA molecules (20–24 nucleotides long) is the downregulation of protein expression by translational repression of the target mRNA or promotion of its cleavage. The HCV genome contains several conserved binding sites for miR-122, the most abundant miRNA in the liver. However, in contrast to its canonical function in the case of HCV, binding of miR-122 to two sites in the 5' UTR increases viral RNA translation and replication primarily via two different mechanisms. First, miR-122 forms an extensive Argonaute-2-containing oligomeric complex at the 5' end of the HCV genome and protects this RNA from nucleolytic degradation by the cellular 5'-to-3' exoribonuclease 1 (Xrn1) and Xrn2. Second, binding of miR-122 suppresses the formation of energetically favorable structures of the 5' UTR interfering with IRES activity, thus promoting RNA translation, which in turn interferes with viral RNA replication.

Another host factor essential for HCV replication is the prolyl-peptidyl *cis-trans* isomerase cyclophilin A (CypA). Binding of CypA to the HCV NS5A protein appears to catalyze the isomerization of peptidyl-prolyl bonds in NS5A domain I and domain II, resulting in conformational changes that stimulate NS5A RNA binding and enhance viral replication. CypA activity also seems to be required for the formation of an HCV replication organelle, as deduced from the observation that CypA antagonists block *de novo* formation of the membranous web.

A third important factor is SEC14L2, a cytosolic lipid-binding protein family member, which emerged as an important HCV host dependency factor. It was shown that ectopic expression of SEC14L2 enhances the replication of patient-derived HCV isolates in diverse human hepatoma cell lines that do not express this factor. This enhancement might be linked to the capacity of SEC14L2 to increase the accumulation of intracellular vitamin E protecting HCV from the detrimental effects of lipid peroxidation.

As mentioned above, apolipoproteins, such as apolipoproteins A-I, B, C-I, and E are components of HCV virions. Of these, ApoE fulfills a dual function in the viral life cycle. First, ApoE interacts with HSPG, LDLR and SRB1 and facilitates virus entry; second, it interacts with NS5A and the envelope glycoproteins and contributes to infectious virus particle production. An association of ApoE with E1-E2 seems to be required for lipidation and maturation of HCV particles at a post-envelopment step (Fig. 3). However, this association can also happen extracellularly with secreted ApoE that is part of lipoprotein particles. Thus, it is possible that ApoE incorporation starts intracellularly during assembly and continues extracellularly after virus particle release.

Host Cell Lipids of Relevance to the HCV Replication Cycle

To accommodate the net increase of intracellular membrane surface and the lipidation of HCV particles, a *de novo* synthesis of membrane lipids is induced in HCV-infected cells, along with an inhibition of lipid secretion. Indeed, lipidomic profiling of HCV-infected cells revealed distinct alterations of the lipid composition of host cells, including elevated levels of phosphatidylcholines and polyunsaturated fatty acids. At least two lipid biosynthesis pathways are activated by HCV: First, the transcriptional induction of lipid biosynthesis-related genes via the sterol regulatory element-binding protein (SREBP) pathway. In this case, HCV infection triggers trafficking of the ER-localized inactive SREBP precursor to the Golgi where it is proteolytically cleaved by site 1 protease (S1P) and S2P. The released N-terminal fragment is transported into the nucleus and it activates the transcription of lipogenic factors such as the fatty acid synthase and 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA), which is a rate-limiting enzyme of the cholesterol biosynthetic mevalonate pathway. Second, the 3' UTR of the HCV RNA genome binds to the cytosolic RNA helicase DEAD box polypeptide 3 X-linked (DDX3X), activating a signaling cascade that promotes SREBP-mediated transcription. In addition, HCV may mobilize fatty acids from lipid droplets.

The strong dependency of HCV on lipids may explain the antiviral activity of inhibitors of distinct lipid biosynthesis pathways. For instance, treatment of cells containing replicating HCV RNA with lovastatin, an inhibitor of 3-hydroxy-3-methylglutaryl CoA reductase, or with an inhibitor of protein geranylgeranyl transferase I induced the disintegration of the HCV replication organelle. Moreover, fatty acids can inhibit or stimulate HCV replication, depending on their degree of saturation. While saturated and monounsaturated fatty acids stimulate viral replication, polyunsaturated fatty acids impair replication but enhance the production of infectious virions.

Activation and Inhibition of Innate and Adaptive Immunity by HCV

Both in cell culture, and in the majority of patients, treatment with IFN- α leads to a rapid and efficient block of HCV replication. This result is somewhat surprising given the high rate of persistence of HCV infections (50%–80%) and the finding that in the infected liver IFN-induced genes (ISGs) are activated. In several studies, it was concluded that HCV proteins such as the core or NS3 proteins interfere with various steps of IFN- α/β induced signaling and that some HCV proteins block individual IFN- α/β induced effectors. However, in primary human hepatocytes infected with HCV, a robust IFN response is induced and the virus is rapidly cleared. Thus, it is still uncertain whether HCV can block the IFN-induced antiviral defense. It also remains to be clarified by which mechanism IFNs block HCV replication.

Much better established is the mechanism by which HCV aggravates the induction of innate antiviral defense. Several studies have shown that the NS3 protease proteolytically cleaves mitochondrial antiviral-signaling molecule (MAVS) relaying the activation of retinoic-acid-inducible gene 1 (RIG-I) to IRF-3 phosphorylation. As a result, IRF-3 dependent genes are not expressed and cells remain sensitive to virus infection. Indeed, it was shown that HCV controls ISG activation in infected cells, including the production of IFN- λ , which is the predominant IFN type in hepatocytes. In contrast, in cells expressing a MAVS variant that cannot be cleaved by HCV ISG expression and IFN- λ secretion are not affected, concomitant with reduced HCV replication. Although these

results clearly show the control of the MAVS signaling pathway by HCV, in infected cells the virus induces a robust IFN response. This may be due to activation of the Toll-like receptor 3 (TLR3) pathway that is less efficiently controlled by HCV. It was found that double-strand HCV RNA, a replication intermediate, is sensed by TLR3 in HCV-infected cells, but released via exosomes from infected cells. Inhibition of exosome release leads to intracellular accumulation of double-strand RNA, enhancing TLR3 activation and thus, induction of IFN gene expression. Apart from that HCV RNA-containing exosomes appear to be sensed by dendritic cells, contributing to the activation of IFN production.

In addition to the innate cytokine response, a rigorous and multi-specific T cell response is critical for the control and elimination of HCV. Thus, a successful antiviral defense generally encompasses multiple MHC class-I and class-II restricted T cell epitopes and an efficient expansion of CD8⁺ and CD4⁺ T cells. In contrast, persistent infections are characterized by oligoclonal T cell responses, a low frequency of HCV-specific T cells and the induction of poorly neutralizing antibodies. The underlying reasons for the weak response in the majority of patients are not clear, but several possibilities are discussed. These include an impaired antigen presentation e.g., due to interference of HCV with dendritic cell function; a CD4⁺ T cell failure due to deletion or anergy; a mutational escape in important T and B cell epitopes; a functional impairment of HCV-specific CD8⁺ T cells. The mechanisms underlying T cell attenuation are unclear, but one attractive possibility is that the defect induced in innate immunity may result in a defect in CD4⁺ T cell help. In fact, HCV-induced loss of T cell help appears to be the key event of immune evasion. In any case, CD8 T cells in chronic hepatitis C patients are exhausted, but this defect is partially restored in patients that have eliminated the virus in the course of antiviral therapy. This observation clearly shows that continuous HCV replication is required to suppress HCV-specific adaptive antiviral response, although an immunological “scar” remains upon virus elimination.

The role of HCV-specific antibodies in controlling viral infection is much less clear. Antibodies are probably less important for viral clearance and in most cases they do not protect from reinfection, neither in experimentally infected chimpanzees nor in humans, even after multiple exposure to HCV. However, there is increasing evidence that the presence of HCV-specific antibodies at least partially attenuates infection. For instance, antibodies neutralizing HCV particles of different genotypes have been detected in sera of chronic hepatitis C patients but the frequency of these antibodies is low. Nevertheless, in-depth studies of selected individuals who spontaneously cleared HCV infection have led to the identification of broadly neutralizing antibodies (bNAbs). In combination with the recently solved crystal structure of the E2 ectodomain, important contact sites between bNAbs and regions in E2, which are of relevance for neutralization across genotypes and subtypes, can now be identified, which should pave the way for the development of a prophylactic HCV vaccine.

Epidemiology

HCV is mainly transmitted by parenteral exposure to blood, blood products and objects contaminated with blood. The development of effective screening tests for blood and blood products and the implementation of viral disinfection procedures have almost excluded this route of transmission in countries where these measures are in place. Thus, the major remaining risk factor for acquiring HCV infection in developed countries is the use of contaminated needles in injection drug use and certain type of risk behavior. In some countries (e.g., Egypt) HCV infection has spread primarily by the use of inadequately sterilized medical instruments. In contrast to HBV infection, sexual transmission, and vertical transmission of HCV are much less frequent.

Modeling of HCV prevalence indicates that 71 million people (62.5–79.4) in 2015 were living with chronic HCV infection, accounting for approximately 1% of the world population. Prevalence rates vary significantly between and within countries. China, Pakistan, India, Egypt, Russia, and the USA have the highest number of cases, corresponding to around 50% of global HCV disease burden. However, only 20% of HCV-infected individuals have been diagnosed. Although the total number of HCV infections is declining since 2007, there are regions in the world where the prevalence is increasing. The reasons for this increase include unsafe health-care-related injections (e.g., Azerbaijan, India, Iraq, Syria, Uzbekistan), i.v. drug use (e.g., Iran, Russia, Latvia, and the USA), or immigration from endemic countries (e.g., Qatar, United Arab Emirates). Of note, in 2015 the number of people dying from HCV-related liver cirrhosis and hepatocellular carcinoma (~400,000) and being cured (~850,000) was exceeded by the number of newly infected persons (~1.75 million). Therefore, the HCV pandemic is continuing unless efficient interventions are scaled-up and implemented globally.

Clinical Features

Acute HCV infections are usually asymptomatic, or in about 30% of cases associated with non-specific symptoms such as abdominal pain, fatigue, weakness, poor appetite, nausea and icterus (Fig. 4). During an incubation period of 15–75 days, HCV RNA becomes detectable in serum by RT-PCR and virus titers usually peak to 10⁵–10⁷ genome equivalents/mL between week 6 and 10, irrespective of disease outcome (Fig. 4(A)). Two to four weeks after onset of viremia, serum ALT (alanine aminotransferase) levels begin to rise indicative of hepatocellular injury. Due to these uncharacteristic symptoms, acute HCV infection remains often unrecognized and therefore, undiagnosed. Only 20%–50% of infected individuals clear the infection, with those having a clinically overt acute infection having a higher chance to clear HCV, arguing that pathogenesis is driven mainly by the immune response.

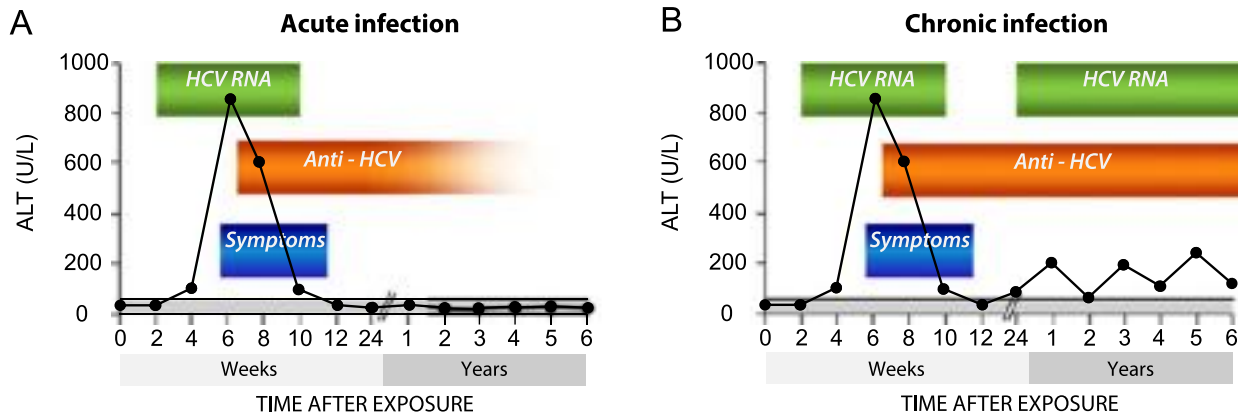


Fig. 4 Course of acute and chronic HCV infection. (A) In acute HCV infection viral RNA is typically detectable between week 2 and 10 after virus exposure. ALT levels rise between week 4 and 10, peaking around week 6. By this time, the first symptoms manifest. HCV-specific antibodies arise late in infection. In self-limiting infection, HCV is cleared and anti-HCV antibodies may wane within years after resolution. (B) In chronic hepatitis C infection the early phase is similar to acute infection but later on, the virus persists. In the chronic phase, ALT levels fluctuate as does viremia. Antibodies against different HCV proteins are constitutively detectable.

While the duration of viremia in acute hepatitis C is unpredictable and can vary between 2 to 4–6 months, some patients even become HCV RNA negative during early convalescence, but later on viremia rebounds. Overall 50%–80% of HCV infections lead to a chronic carrier state (Fig. 4(B)). About 30% of these chronically infected persons progress to liver cirrhosis within 10–30 years after primary infection and hepatocellular carcinoma (HCC) occurs in up to 2.5% of these patients. In contrast to chronic hepatitis B, in persistent HCV infection HCC develops only to those patients, who have developed liver cirrhosis.

Pathogenesis

HCV infection is most likely not cytolytic. Instead, hepatocyte destruction is mediated by the immune response, arguing that HCV, like HBV infection, causes an immune pathology. Since hepatocytes are the primary target of HCV infection, the histological alterations of chronic hepatitis C are hepatocellular injury, portal and parenchymal inflammation and necrosis. Liver damage is typically spotty and focal with accompanying chronic inflammatory cells, including macrophages, and the disease eventually develops to a variable degree of fibrosis. The progression rate to hepatic fibrosis is the major determinant for the outcome of chronic hepatitis C in terms of developing cirrhosis and HCC. Although the liver is the primary target, persistent HCV infection is often associated with extrahepatic symptoms like renal complications, lymphoma and diabetes. A high proportion of patients with chronic hepatitis C develop cryoglobulinemia, which may account for some of these extrahepatic manifestations.

The mechanisms underlying HCV-associated tumorigenesis are poorly understood, but they are thought to arise from a combination of indirect and direct mechanisms. The most important indirect effect is a chronic inflammation, sustained at least in part by the antiviral immune response that attacks and destroys hepatocytes, yet this response is unable to eradicate the virus and eliminate the infection. Amongst the more direct effects are: (1) a possible reduction of the tumor suppressor miR-122 in infected cells due to binding to viral RNA and sequestration of miR-122 from normal cellular target RNAs; (2) a possible oncogenic effect of HCV core protein that has been observed in certain transgenic mouse models; (3) an elevated expression of EGFR, induced by HCV infection and playing an important role in HCC development; (4) the binding of NS5B to the tumor suppressor Rb, resulting in Rb degradation via the ubiquitin-proteasome pathway simultaneously with stimulation of hepatocellular proliferation. (5) persistent changes of the epigenome of infected cells.

Diagnosis

Routine screening tests for detecting HCV infections are based on serological assays measuring HCV specific antibodies (most often by ELISA or similar immunoassay) and nucleic acid-based tests (NAT) such as RT-PCR to determine viral RNA. Current serological assays have a specificity of >99% and they are positive in 99% or more of immunocompetent patients where viral RNA is detectable. Once a patient is tested positive in a serological assay, the sample should be tested by RT-PCR. A seropositive person with a positive PCR test result may have an acute, or, much more likely, a chronic HCV infection. A seropositive person with a negative PCR test result might have had an early or eliminated infection, a false negative test result, or a false positive serologic test result, or a chronic HCV infection with a very low viremia at the time of blood sampling. Thus, repeated testing is required in those cases.

Given the higher sensitivity, the diagnostic window can be reduced by using NAT. Quantitative RNA detection assays have been implemented in many blood banks e.g., in the EU and the US. The risk to acquire a transfusion-associated hepatitis C in such countries has been reduced to less than 1 in a million blood donations. The determination of HCV genotypes, which is an important parameter for IFN-based antiviral therapy, is based on analyzing viral RNA either by hybridization or direct sequence analysis or by using genotype-specific primers for RT-PCR. In addition to NAT, other (direct) detection methods of HCV have been developed, including the quantitative detection of HCV core protein. However, these assays are not widely available.

Treatment

For many years, antiviral therapy of chronic hepatitis C was based on the combination of a polyethylene glycol (PEG) conjugated form of IFN- α with ribavirin resulting in an overall sustained viral response (SVR) rate of ~60% in eligible patients. However, the success rate depended very much on two important parameters. The first one was the genotype of the infecting virus. While up to 85% of genotype 2- and 3-infected patients developed SVR, only around 45% of patients infected with genotype 1 viruses did so. In addition, this therapy had numerous serious side effects, including neuropsychiatric symptoms and hemolytic anemia, and therefore, many patients were either not eligible for this treatment or had to discontinue the therapy. The second parameter determining the success of IFN therapy is 3 distinct SNPs near the IFN- λ 3 and - λ 4 genes. While the causality between HCV infection and these SNPs remains to be determined, they appear to be of predictive value for other diseases including nonalcoholic fatty liver disease or infections with cytomegalovirus or HIV.

The limitations of IFN-based therapy of chronic hepatitis C have been overcome by the implementation of all-oral IFN-free direct acting antiviral drugs (DAAs) in 2014. These drugs allow virus elimination in >95% of individuals within a treatment period of 8–12 weeks; they are well tolerated and can be given to patients even with a decompensated liver cirrhosis. DAAs are targeting three viral proteins (Fig. 5): the NS3 serine protease, the NS5A replicase and assembly factor, and the NS5B RdRp. Numerous drugs against these three targets have been developed and although the drug names are complicated, the drug target is disclosed in the name. Protease inhibitors are ending with “-previr”, NS5A inhibitors with “-asvir”, and NS5B inhibitors with “-buvir”. While most of the first generation DAAs had a rather low resistance barrier, this limitation has been overcome with 3rd generation DAAs that are also rather pan-genotypic. Therefore, the first-line treatment of hepatitis C patients is based on a combination of at least two DAAs, e.g., a protease and a NS5A inhibitor or in addition a NS5B inhibitor (Fig. 5).

Although IFN-free DAA treatment has been in use for just a few years, it is becoming clear that the incidence of end-stage liver disease caused by HCV infection is declining with patients that mount a SVR. However, it is also becoming clear that in patients with advanced cirrhosis, HCV elimination is of limited use to prevent HCC because these patients still have a high risk for liver tumor development. While this observation argues for a “point-of-no-return” beyond which the presence of HCV is no longer relevant for HCC development, at the same time it has become clear that antiviral therapy should be given as early as possible to HCV infected individuals and not be limited to people with advanced liver disease. Fortunately, the price of the DAA therapy is reducing enabling the treatment of a larger proportion of HCV infected individuals.

Prevention

With the advent of highly efficacious DAAs global eradication of HCV has been considered as a reachable goal. While this is theoretically possible, major limitations need to be overcome. These include the limited availability of DAAs in many countries, the still high costs for these drugs, the insufficient health care systems in many high-endemicity countries and the fact that only a small proportion of HCV infections have been diagnosed. Moreover, reinfections with HCV are common, especially in high-risk populations such as intravenous drug users, and the number of new infections is rising, e.g., as a result of the opioid crisis in the USA (see above). For these reasons, global eradication strategies must not ignore the development of a prophylactic HCV vaccine. Although scientifically challenging, effective immunization against HCV appears to be feasible. Both in experimentally infected chimpanzees and in humans a rigorous and multi-specific T cell response can be induced and some individuals develop cross-neutralizing antibodies. What we need is a better understanding of the correlates of immune protection and how to induce a protective immune response with an ideal vaccine formulation. Thus far, only two vaccine candidates have been pursued in clinical trials. The first one is a recombinant envelope glycoprotein prophylactic vaccine that induced sterilizing immunity against experimental challenge of chimpanzees with a homologous HCV strain and reduced chronicity upon challenge with a heterologous strain of the same genotype. The second vaccine candidate is based on an adenovirus vector encoding an HCV NS3–5B polyprotein fragment used to prime the immune response, followed by a boost with a modified vaccinia Ankara (MVA) vector encoding the entire genome of the same HCV type. This combination, aiming to prime and boost CD4 and CD8 T cell responses, was found to be safe and to induce polyfunctional, broad HCV-specific memory CD4 and CD8 T cells. However, in a recent phase 2 clinical trial this vaccine failed to reach clinical significance in reducing the incidence of chronic HCV infection compared to placebo among HCV-uninfected intravenous drug users. Therefore, further efforts are required to identify vaccine candidates inducing a protective immune response. Without such a vaccine, achieving global eradication of chronic HCV infection is questionable.

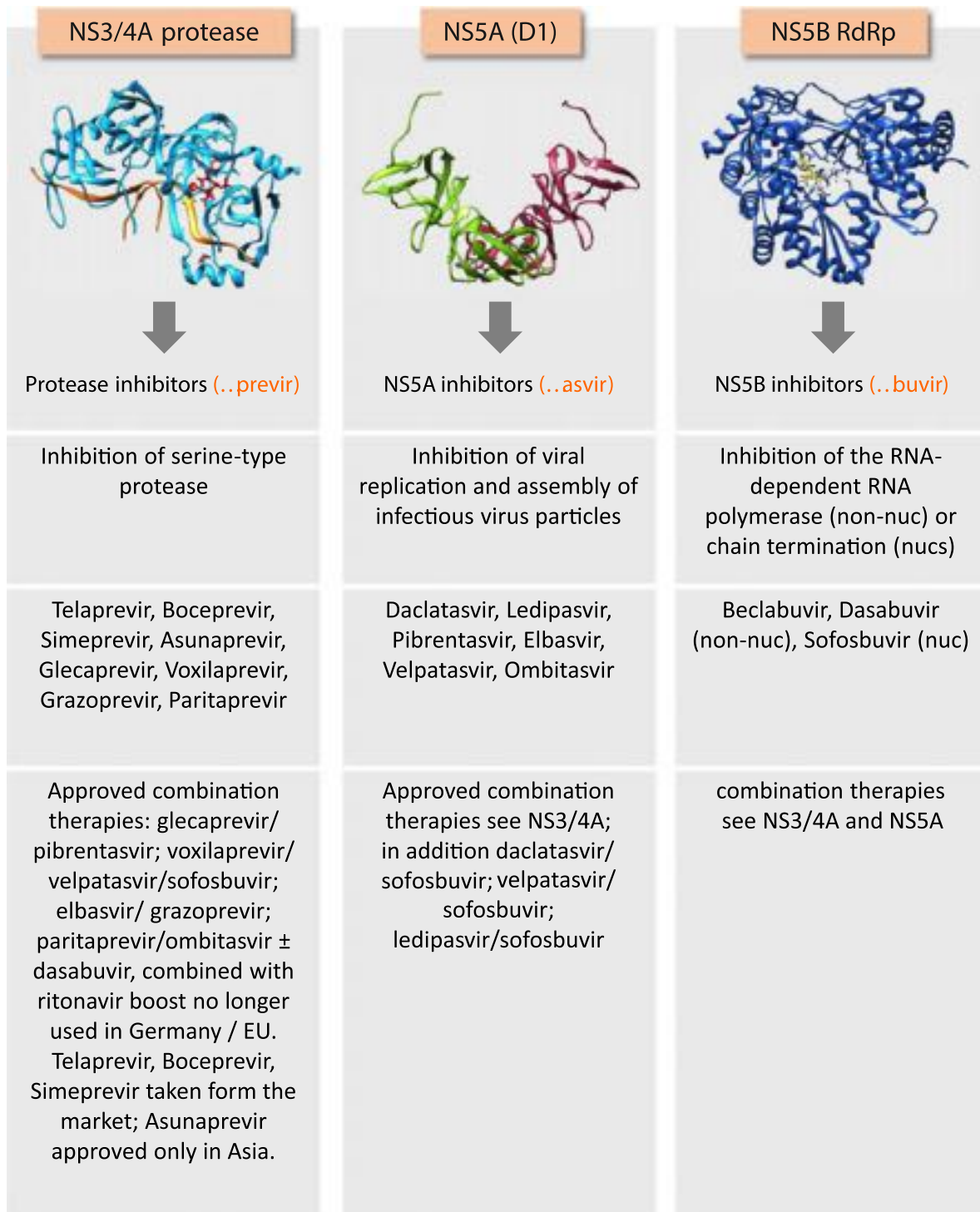


Fig. 5 Drug targets and direct acting antiviral drugs (DAAs) for treatment of HCV infection. Ribbon diagrams of the 3D structures of the NS3/4A protease complex, the homodimer of NS5A domain I and the NS5B RdRp domain are displayed on the top. Shown are the NS3 protease domain (light blue) complexed with the NS4A peptide (orange) and Boceprevir (red); the NS5A domain I dimer with monomers represented in red and green; the NS5B RdRp domain complexed with UTP (yellow). DAAs targeting each of these proteins and clinically used drug combination therapies are given below each viral protein. Protein structures are derived from Protein data bank (PDB) accession 2OC8 for NS3/4A, PDB accession 1ZH1 for NS5A and PDB accession 1GX6 for NS5B.

Acknowledgments

The authors gratefully acknowledge Dr. Eliana G. Acosta (University of Heidelberg) for excellent assistance with the preparation of the manuscript and the figures.

Further Reading

- Bartenschlager, R., Lohmann, V., Penin, F., 2013. The molecular and structural basis of advanced antiviral therapy for hepatitis C virus infection. *Nature Reviews Microbiology* 11, 482–496.
- Bartenschlager, R., Penin, F., Lohmann, V., André, P., 2011. Assembly of infectious hepatitis C virus particles. *Trends in Microbiology* 19, 95–103.
- Catanese, M.T., Uryu, K., Kopp, M., *et al.*, 2013. Ultrastructural analysis of hepatitis C virus particles. *Proceedings of the National Academy of Sciences of the United States of America* 110, 9505–9510.
- Choo, Q.L., Kuo, G., Weiner, A.J., *et al.*, 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244, 359–362.
- Lohmann, V., Körner, F., Koch, J., *et al.*, 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science* 285, 110–113.
- Manns, M.P., Buti, M., Gane, E., *et al.*, 2017. Hepatitis C virus infection. *Nature Reviews Disease Primers* 3, 17006.
- Neufeldt, C.J., Cortese, M., Acosta, E.G., Bartenschlager, R., 2018. Rewiring cellular networks by members of the Flaviviridae family. *Nature Reviews Microbiology* 16, 125–142.
- Ortega-Prieto, A.M., Dorner, M., 2017. Immune evasion strategies during chronic hepatitis B and C virus infection. *Vaccines* 5, E24.(Basel).
- Paul, D., Bartenschlager, R., 2015. Flaviviridae replication organelles: Oh, what a tangled web we weave. *Annual Review of Virology* 2, 289–310.
- Polaris Observatory HCV Collaborators, 2017. Global prevalence and genotype distribution of hepatitis C virus infection in 2015: A modelling study. *Lancet Gastroenterology & Hepatology* 2, 161–176.
- Romero-Brey, I., Merz, A., Chiramel, A., *et al.*, 2012. Three-dimensional architecture and biogenesis of membrane structures associated with hepatitis C virus replication. *PLOS Pathogens* 8, e1003056.
- Ross-Thriepland, D., Harris, M., 2015. Hepatitis C virus NS5A: Enigmatic but still promiscuous 10 years on!. *Journal of General Virology* 96, 727–738.
- Smith, D.B., Bukh, J., Kuiken, C., *et al.*, 2014. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology* 59, 318–327.
- Tu, T., Bühler, S., Bartenschlager, R., 2017. Chronic viral hepatitis and its association with liver cancer. *Biological Chemistry* 26, 817–837.
- Wensch, F., Crouchet, E., Ligat, G., *et al.*, 2018. Hepatitis C virus (HCV)-apolipoprotein interactions and immune evasion and their impact on HCV vaccine design. *Frontiers in Immunology* 21, 1436.

Relevant Websites

- https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification
HCV Classification.
International Committee on Taxonomy of Viruses (ICTV).
- <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>
Hepatitis C.
World Health Organisation (WHO).
- <https://www.cdc.gov/hepatitis/hcv/cfaq.htm>
Hepatitis C Questions and Answers for the Public.
CDC.
- https://clarolineconnect.univ-lyon1.fr/uploads/webresource/189e3e07-df18-48b5-9164-8a6312cb8628.zip/hcv_life_cycle_site_web/index.html
Hepatitis C Virus Life Cycle.
- <https://www.hcvguidelines.org/>
HCV Guidance: Recommendations for Testing, Managing, and Treating Hepatitis C.
- <https://www.niaid.nih.gov/news-events/trial-evaluating-experimental-hepatitis-c-vaccine-concludes>
Trial Evaluating Experimental Hepatitis C Vaccine Concludes.

Hepeviruses (*Hepeviridae*)

Xiang-Jin Meng, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Animal Reservoir A large animal population serving as a source of virus supply for the transmission of viruses to other animals including humans.

Fulminant hepatitis A clinical condition characterized by massive necrosis of liver parenchyma and decrease in liver size that is usually caused by hepatitis virus infections or drug and toxin-induced liver injury.

Quasi-enveloped virus particles Non-enveloped viruses but present as membrane-wrapped quasi-enveloped virions

in the blood of infected hosts. As a non-enveloped virus, hepatitis E virus is present in multivesicular body as membrane-associated quasi-enveloped particles with intraluminal vesicles resembling exosomes in infected cells and blood.

Zoonotic infection Infectious diseases transmissible under natural conditions from vertebrate animals to humans.

Classification

Hepatitis E virus (HEV) is classified in the family of *Hepeviridae*, which comprises two genera. The genus *Orthohepevirus* includes all mammalian and avian hepatitis E virus isolates, whereas the genus *Piscihepevirus* consists of the cutthroat trout virus. Within the genus *Orthohepevirus*, there exist four distinct species: *Orthohepevirus A* consisting of HEV isolates from human, wild and domestic pigs, deer, mongoose, rabbit, and camel, *Orthohepevirus B* comprising avian HEV isolates from birds, *Orthohepevirus C* containing HEV isolates from rat, greater bandicoot, Asian musk shrew, ferret and mink, and *Orthohepevirus D* including HEV isolates from bat (Table 1).

HEV isolates known to infect humans all belong to the species *Orthohepevirus A*, which consists of at least 8 distinct genotypes. Genotype 1 HEV (mostly Asian isolates) and genotype 2 HEV (a Mexican isolate and some African isolates) infect only humans, whereas genotypes 3 (human, pig, rabbit, deer, mongoose), and 4 (human and pig) HEVs infect humans and several other animal species. Genotypes 5 and 6 HEVs infect wild boars with unknown zoonotic potential. Genotype 7 HEV infects dromedary camels and reportedly a human, and the genotype 8 HEV infects Bactrian camel (Table 1).

Virion Structure

HEV is an icosahedral, spherical virus particle of approximately 32–34 nm in diameter. HEV particles secreted in bile and stool are non-enveloped, however, in circulating blood and in supernatant of infected cell culture, HEV appears as membrane-associated quasi-enveloped particles with intraluminal vesicles resembling exosomes that are infectious but resistant to antibody neutralization. The capsid protein encoded by ORF2 is the only known structural protein on the virion, although the majority of ORF2 protein in serum and supernatant of HEV-infected cell culture exists as a secreted form (ORF2^S) that is not associated with virions.

The virion of HEV is a $T=3$ icosahedral capsid lattice consisting of 180 copies of capsid protein with three functional domains: S (shell), M (middle), and P (protruding). The S domain forms the icosahedral shell, while the P domain is the binding site for an unknown cellular receptor. The M domain interacts with S and P domains contributing to virion stability. The virion buoyant density is 1.35–1.40 g/cm³ in CsCl, and 1.29 g/cm³ in potassium tartrate and glycerol. The virion sedimentation coefficient is 183S. The HEV virion is more heat labile than is hepatitis A virus (HAV), another enterically transmitted hepatitis virus. Infectious HEV was still detectable after incubation at 37°C for 21 days, at room temperature for 28 days, and at 4°C for 56 days with a 2.7-log decrease of virus titer. Incubation of infectious virus at 60°C for 1 min resulted in moderate decreases in infectivity, 2- to 3.5-log decreases in infectivity between 65°C and 75°C for 1 min, and complete inactivation at $\geq 80^\circ\text{C}$ for 1 min. The fecal-oral route of transmission indicates that HEV is resistant to inactivation by acidic and mild alkaline conditions in the gastrointestinal tract.

Genome

The genome of HEV, which possesses a 7-methylguanosine cap structure at its 5' end, is a single-stranded, positive-sense RNA molecule of approximately 7.2 kb in length. The genome consists of a short 5' noncoding region (NCR), three open reading frames (ORFs 1, 2, and 3), and a 3' NCR. ORF3 overlaps ORF2, but neither overlaps ORF1 (Fig. 1). The ORF1 encodes nonstructural proteins responsible for viral replication and protein processing. Functional domains within ORF1 have been identified, including methyltransferase (Met), papain-like cysteine protease (PCP), hypervariable region (HVR) or proline rich region (PRR), X (or

Table 1 Host range of hepatitis E virus and zoonotic infection in humans^a

Natural animal host	Classification [genus/species, genotypes (gt)]	Experimental hosts for cross-species infection	Zoonotic infection in humans
<i>Orthohepevirus A</i>			
Human	gt 1, 2, 3, 4	Non-human primates, pigs (gt3, gt4), rabbit (gt1, gt4), lamb (gt1), Wistar rats (gt1)	
Domestic swine	gt 3, 4	Non-human primates, rabbit, Mongolian gerbil (gt4), Balb/C mice (gt3, gt4)	Yes
Wild boar	gt 3, 4, 5, 6		Yes (gt 3, 4), likely (gt 5, 6)
Deer	gt 3		Yes
Rabbit	gt 3	Pig	Yes
Mongoose	gt 3		Likely
Dromedary camel	gt 7	Non-human primates	Yes
Bactrian camel	gt 8	Non-human primates	Unknown
Moose	unknown		Unknown
<i>Orthohepevirus B</i>			
Chicken, wild birds	aHEV gt 1, 2, 3, 4	Turkey	No
<i>Orthohepevirus C</i>			
Rat			Yes
Ferret			Unknown
Greater bandicoot			Unknown
Asian musk shrew			Unknown
Mink			Unknown
Common kestrel			Unknown
Red-footed falcon			Unknown
Bat	<i>Orthohepevirus D</i>		No
Cutthroat trout	<i>Piscihepevirus</i>		No

^aModified from Meng, X.J., 2016. PLoS Pathogens 12(8), e1005695. doi:10.1371/journal.ppat.1005695, which was published under a Creative Commons License.

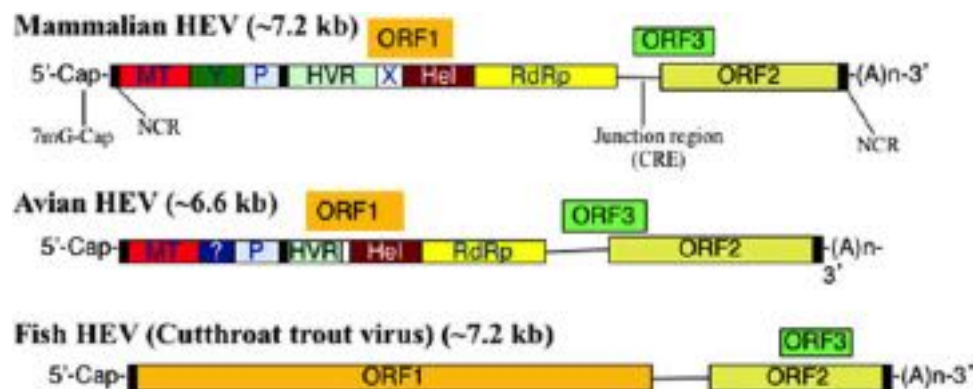


Fig. 1 A schematic diagram of comparative genomic organization of mammalian, avian, and fish HEVs. The three open reading frames (ORFs) are labeled and shown as boxes. ORF2 overlaps ORF3 but neither overlaps ORF1. ORF1 encodes nonstructural proteins with the putative functional domains indicated inside the box. ORF2 encodes the capsid protein, and ORF3 encodes a small multifunctional protein that is involved in virus replication. The HEV genome is capped (m7G Cap) at the 5' end, and contains a poly A tail at the 3' end. There are noncoding regions (NCR) at the 5' and 3' end of the viral genome. There is a junction region between ORF1 and ORF3 for mammalian and avian HEVs, which contains a stem-loop structure and a *cis*-reactive element (CRE). The avian HEV genome is approximately 600 bp smaller than the mammalian and fish HEVs. Hel, helicase; HVR, hypervariable region; MT, methyltransferase; NCR, noncoding region; P, a papain-like cysteine protease; RdRp, RNA-dependent RNA polymerase; X, macro domain. Reproduced from Meng, X.J., 2016. PLoS Pathogens 12(8), e1005695. doi:10.1371/journal.ppat.1005695, which was published under a Creative Commons License.

macro) domain, helicase (Hel), and RNA-dependent RNA polymerase (RdRp) (Fig. 1). An RNA element located in the 5' end region of the ORF1 binds to capsid protein and is thought to be the RNA packaging signal. The viral genomic RNA contains two *cis*-reactive RNA elements (CREs) important for virus replication. CRE1 overlaps the 3' end of ORF2 and continues into the 3' NCR, and forms two stem-loop structures that interact with RdRp. CRE2 locates in the intergenic junction region between ORF1 and ORF3, and forms a stem-loop structure that is part of the promoter for synthesis of the subgenomic RNA. The intragenomic

subgenomic promoter regulating subgenomic RNA synthesis is mapped to the junction region extending to the 3' end of ORF1. Both the sequence and the stem-loop structure in the junction region play important roles in HEV replication.

ORF2 encodes capsid protein that contains a signal peptide sequence, and three glycosylation sites. The capsid protein contains an immunogenic epitope, induces neutralizing antibodies against HEV, and is the target for vaccine development. A secreted form of ORF2 protein (ORF2^S) is translated from an AUG start codon that is 15 amino acids upstream to the N terminus of the capsid protein. The ORF2^S does not block HEV cell entry but inhibits antibody-mediated neutralization. ORF3 encodes a small cytoskeleton-associated phosphoprotein. The N-terminus of ORF3 has a cysteine-rich region, binds to HEV RNA, and enters into a complex with the capsid protein. The C-terminus of ORF3 is a multifunctional domain involved in HEV virion morphogenesis and pathogenesis. The ORF3 is essential for virus infectivity in vivo, although it is not required for virion assembly or infection of hepatocellular carcinoma cells in vitro. Overlapping the macro and Hel domains in genotype 1 HEV, but not other HEV genotypes, is a small frameshifted ORF4 that is controlled by an internal ribosome entry site (IRES)-like RNA structure, which induces ORF4 protein translation under endoplasmic reticulum (ER) stress.

Life Cycle

Attachment and entry: The attachment and entry of HEV into susceptible cells remains largely unknown, due in large part to the lack of an efficient cell culture system to propagate HEV. The capsid protein attaches to susceptible cells via heparin sulfate proteoglycans (HSPGs) to facilitate HEV entry, although a specific receptor for HEV has not yet been identified. The virions are internalized through a clathrin and dynamin-2 based pathway. The entry of quasi-enveloped HEV particles reportedly involves the small GTPases Rab5 and Rab7 and degradation of the quasi-envelope membrane within the lysosome. The uncoating process of the capsid is not understood, although the P domain of the capsid contains polysaccharide binding sites that may be involved in capsid disassembly to facilitate the uncoating process.

Transcription and translation: Upon uncoating and delivering the viral genomic RNA to the cytosol, the ORF1 polyprotein is translated directly from the positive-sense viral genomic RNA by cap-dependent translation, and eukaryotic translation initiation factors 4A, 4G, and 4E are involved in HEV replication. Whether the ORF1 polyprotein requires further processing for functionality remains debatable, since contradictory results have been reported with some studies showing no processing at all while others reporting cleaved products corresponding to the sizes of Met, Hel, and RdRp. The ORF2 and ORF3 proteins are also translated in a cap-dependent fashion from a single bicistronic subgenomic mRNA by a process involving leaky ribosome scanning in which ribosomes bypass the ORF3 initiator AUG to initiate protein synthesis downstream at the ORF2 AUG initiator codon.

Replication of viral genome: The replication of HEV genome involves the transcription of a negative-sense replicative RNA intermediate by viral RdRp. This negative-sense RNA subsequently serves as the template for the production of the positive-sense, progeny viral genomic RNA as well as the bicistronic subgenomic mRNA. In HEV-infected animal tissues, the negative-sense replicative RNA intermediate has been detected by negative strand-specific RT-PCR. A viral replication complex (VRC) forms on ER membranes or on vesicles consisting of ER-derived membranes, although the VRC composition is not well understood. For genotype 1 HEV, the VRC appears to contain the ORF4 protein, Hel, RdRp, and macro domains of ORF1, along with host eukaryotic elongation factor 1 isoform-1 (eEF1 α 1) and tubulin β .

Assembly and release: Little is known about the assembly and release steps in the life cycle of HEV. The capsid protein packages the HEV genome through binding to a putative RNA packaging signal located in the 5' end region of the viral genome. Binding of viral RNA to the N-terminus of capsid leads to the formation of C-terminal/C-terminal dimers that are important for completing viral capsid assembly. It has been shown that ORF3 is involved in viral egress via interaction with host late domain proteins. Quasi-enveloped HEV has been detected within multivesicular bodies (MVBs) associated with trans-Golgi network protein 2 and CD63, suggesting that HEV capsids bud into intracellular vesicles requiring the MVB pathway to release virions. The ORF3 protein is a viroporin forming a functional ion channel at plasma membrane to facilitate the release of infectious virus particles.

Epidemiology

Morbidity and mortality: In developing countries, a high seroprevalence of HEV antibodies was reported. For example, more than 70% of the general populations in Egypt are seropositive for HEV antibodies. Surprisingly, in some industrialized countries where clinical cases of hepatitis E are rare, the seroprevalence of HEV antibodies is much higher than expected, with up to 30% seropositivity in blood donors in the United States. The morbidity for hepatitis E is not very reliable due in large part to the lack of accurate etiology-specific data for acute viral hepatitis. Most sporadic cases of acute viral hepatitis are not routinely tested for HEV and thus are not reported. Historically, the incidence of hepatitis E in large outbreak settings was estimated at 1400–1650 per 100,000 population in India and Nepal. More recently, a large outbreak of 2621 cases of acute hepatitis E was reported in refugee camps in Darfur, Sudan with an attack rate of 3.3% and a case-fatality rate of 1.7%. The highest attack rate of clinical disease is generally in young adults of 20–29 years of age, although the seroprevalence of HEV antibodies is age-dependent and increases with age. HEV-associated mortality in the general population is <1% but it can reach up to 25% in infected pregnant women. According to the World Health Organization, annually there are an estimated 20 million HEV infections worldwide leading to approximately 3.3 million cases of hepatitis E with >44,000 HEV-related deaths.

Animal reservoirs and cross-species infection: Since the discovery of the first animal strain of HEV, swine HEV, from pigs in the United States, genetically-diversified strains of HEV have now been identified from more than a dozen other animal species including chicken, deer, mongoose, rabbit, rat, ferret, moose, shrew, kestrel, egret, bat, goat, camel, and cutthroat trout. The ecology, natural history, zoonotic potential and pathogenicity for most of these animal hepeviruses are still largely unknown, although the genotype 3 HEV infecting pig, rabbit, deer and mongoose, and the genotype 4 HEV infecting pigs are known to be zoonotic and infect humans.

Under natural conditions, zoonotic infections in humans by genotypes 3 and 4 HEVs from various animal species have been well documented. The sporadic and cluster cases of hepatitis E in humans from industrialized countries are caused primarily by the zoonotic genotypes 3 and 4 HEVs of animal origin. Cases of chronic hepatitis E in immunocompromised individuals are almost exclusively caused by the zoonotic genotype 3 HEV of animal origin. Pig caretakers such as farmers and swine veterinarians had a higher risk of zoonotic HEV infection. In addition to pigs, the zoonotic genotype 3 HEV has also been identified in rabbit, deer, and mongoose, suggesting that these animals are also reservoirs. Zoonotic transmissions of hepatitis E from deer to humans and from rabbit to humans have been reported. A genotype 7 camelid HEV also reportedly infected a human liver transplant recipient. More recently, human infections by rat HEV have also been reported.

Under experimental conditions, genotypes 3 and 4 HEVs can infect across species barriers: genotypes 3 and 4 swine HEVs infect rhesus macaques and chimpanzee, and conversely, genotypes 3 and 4 human HEVs infect pigs. Genotype 3 rabbit HEV was also shown to infect pigs and cynomolgus macaques. Balb/c nude mice are reportedly susceptible to experimental infection with a genotype 4 swine HEV, although others failed to infect C57BL/6 mice with genotype 1, 3 and 4 HEVs. Rabbits have been reportedly infected with genotypes 1 and 4 HEV, although independent confirmation is still lacking. Avian HEV infects turkeys, but not rhesus macaques, suggesting that avian HEV is not zoonotic.

Sources and risk factors of HEV infection: For epidemics, the main source of infection is HEV-contaminated drinking water in areas with poor sanitation such as villages in developing countries, and refugee camps. Hepatitis E epidemics are often associated with contamination of well water or leakage of untreated sewage into city water treatment plants, and poor sanitary conditions such as washing hands in a group basin. Person-to-person transmission of hepatitis E is not common during an epidemic. In contrast, the main sources for sporadic or cluster cases of hepatitis E appear to be contaminated animal meats, shellfish, and direct contact with infected animals.

Hepatitis E epidemics are mainly in developing countries of Asia, Africa, and Latin America, particularly in regions with hot climates or poor sanitation conditions. However, sporadic or cluster cases of hepatitis E occur in both developing and industrialized countries. A high incidence of hepatitis E outbreaks is observed in seasons with monsoon rains and flooding. The use of river water for drinking and cooking, washing, and disposal of human excreta carries a higher risk of HEV infection. Pigs are reservoirs for HEV, and therefore pig farmers, swine veterinarians and other pig caretakers in both developing and industrialized countries are at increased risk of HEV infection. In the United States, swine veterinarians were 1.51 times more likely to be seropositive for HEV antibodies than normal blood donors. Individuals from traditionally major swine states are more likely to be seropositive for HEV antibodies than those from traditionally non-swine States. Individuals with occupational exposure to wild animals also have an increased risk of zoonotic HEV infection. Sewage workers in India had a significantly higher seroprevalence of HEV antibodies (56.5%) than controls (19%), although such an association was not found in Switzerland. In a large hepatitis E outbreak in 2004 in refugee camps in Darfur, Sudan, the risk factors included age of 15–45 years (OR=2.13) and drinking chlorinated surface water (OR=2.49).

Transmission: HEV is primarily transmitted by fecal-oral route, although under experimental conditions it is rather difficult to infect non-human primates or pigs via oral route of inoculation. Waterborne transmission via sewage-contaminated drinking water and contaminated well or river water used for washing and drinking purpose is the main source of HEV infection in developing countries. Feces from infected animals contain large amounts of HEV leading to contamination of irrigation or coastal water with concomitant contamination of produce or shellfish. HEV strains of both human and swine origin have been detected in sewage water.

Foodborne transmission is responsible for most sporadic and cluster cases of acute hepatitis E that were linked to the consumption of raw or undercooked animal meats. Sporadic cases of hepatitis E were associated to the consumption of raw or undercooked wild boar meat deer meat, and pork products. HEV RNA was detected in commercial pork products sold in grocery stores, and the contaminating virus in commercial pig livers from grocery stores remains infectious. Raw pig liver sausages (figatelli) were the source for some sporadic and cluster cases of hepatitis E in France. Consumption of raw deer meats was responsible for cluster cases of acute hepatitis E in Japan. Also, HEV infections were linked to the consumption of contaminated shellfish and strawberries. Zoonotic transmission through direct contact with infected animals occurs for genotypes 3, 4 and genotype 7 HEVs within the species *Orthohepevirus A*. Potential transmissions of hepatitis E from a pet cat and a pet pig to human owners were also reported.

Vertical transmissions from mother-to-fetus associated with a high neonatal mortality has been reported. Among the 62 HEV-infected pregnant women, vertical transmission of HEV occurred in 33.3% of the cases. HEV RNA was detected in some cord blood and in colostrum of HEV-infected mothers, although breast-feeding appears to be safe for infants. Vertical transmission has been reported in HEV-infected human patients, but experimental evidence of vertical transmission of HEV in animal models is still lacking. Pregnant sows and pregnant rhesus monkeys experimentally-infected with HEV failed to transmit the virus to offspring. Egg whites from eggs of chickens experimentally infected with a chicken strain of HEV contain infectious virus, but evidence of complete vertical transmission is lacking.

Blood-borne transmission through blood transfusions, although not common, has been reported in patients worldwide. HEV RNA has been detected in a small proportion of blood donors in Japan and some European countries, thus raising a potential

concern of transfusion-transmitted hepatitis E. In Germany, HEV viremia was detected in approximately 0.12% of blood donors at a tertiary care center. In the United States, in one study HEV RNA was not detected in blood donors or recipients, while in a separate study three out of 128,020 blood donors from 27 states tested positive for genotype 3 HEV RNA. The incidence of viremia and viral loads in blood donors, and the total plasma volume the recipients receive are the determining factors for transfusion-associated HEV infection. It remains debatable as to whether donor blood should be screened for acute markers of HEV infection, although several European countries have already implemented HEV RNA screening of blood donors.

Clinical Features

The majority of HEV infections are subclinical, and a proportion of the infected individuals do develop self-limiting acute viral hepatitis, but the acute disease usually does not go into chronicity. However, in immunocompromised individuals, HEV infections can progress into chronic hepatitis E, which has become an emerging and significant clinical problem. Numerous neurological diseases such as Guillain-Barré syndrome and neuralgic amyotrophy have also been linked to HEV infection, although the underlying mechanism of neurological invasion by HEV remains unknown. Fulminant hepatitis E with a high mortality have been reported in HEV-infected pregnant women especially during the third trimester.

Self-limiting acute hepatitis E: Only a proportion of HEV-infected individuals develop self-limiting acute viral hepatitis. The incubation period typically ranges from 2 weeks to up to 2 months. The clinical symptoms of acute hepatitis E, which include Jaundice, anorexia, abdominal pain, nausea, vomiting, and fever, are essentially indistinguishable from those of other types of acute viral hepatitis. Viremia and fecal virus shedding are detectable approximately 3 weeks post-infection and 1 week prior to the onset of clinical disease. A transient appearance of IgM HEV antibodies followed by long-lasting IgG antibodies appear late during the period of viremia and fecal virus shedding. The infected patients typically recover from acute hepatitis E and are protected against subsequent HEV infection.

Chronic hepatitis E: Although HEV generally causes a self-limiting acute viral hepatitis in immunocompetent individuals, the majority of HEV infections in immunocompromised individuals such as solid organ transplant recipients and patients with HIV/AIDS, lymphoma, or leukemia progress into chronic hepatitis E. Organ transplant recipients with chronic hepatitis E have persistently elevated aminotransferase levels, viremia, and hepatic lesions, and also shed HEV in feces for a prolonged period of time. Cases of chronic hepatitis E are almost exclusively caused by the zoonotic genotype 3 HEV, which infects pig, rabbit, deer and mongoose. Therefore, the source of chronic HEV infection is of zoonotic origin presumably through consumption of undercooked HEV-contaminated animal meats.

Under experimental conditions, aimed at mimicking the conditions of immunocompromised organ transplant patients, pigs were treated with an immunosuppressive regimen and experimentally infected with genotype 3 HEV. The infected pigs developed chronic HEV infection with fecal virus shedding lasting for at least 22 weeks post-infection. Active suppression of cell-mediated immune responses under immunocompromised conditions facilitates the establishment of chronic HEV infection. Similarly, cynomolgus monkeys treated with tacrolimus, an immunosuppressive drug, and experimentally infected with a genotype 3 swine HEV also developed chronic HEV infection with pronounced hepatic lesions at 160 days after infection.

Fulminant hepatitis with a high mortality during pregnancy: Accumulating evidence indicates that HEV infection during pregnancy causes fulminant hepatic failure with a high mortality. In India, one study showed that the mortality rate in HEV-infected pregnant women is 65.8% (25/38) due to fulminant hepatic failure. In another study, the mortality rate among the HEV-positive pregnant women was 27%, and approximately two-thirds of the infected pregnant women had preterm deliveries. However, in Egypt, a study of 2428 pregnant women did not link HEV exposure to a history of liver disease. Genotype 1 HEV is usually associated with fulminant hepatic failure in infected pregnant women, although zoonotic genotypes 3 and 4 HEVs are also associated with fulminant hepatic failure.

Under experimental conditions, however, severe and fulminant hepatitis could not be reproduced in pregnant pigs or monkeys: pregnant sows experimentally infected with genotype 3 HEV had no clinical signs of hepatitis or elevation of liver enzymes. Similarly, pregnant rhesus macaques experimentally infected with genotype 1 HEV did not develop more severe hepatitis than non-pregnant monkeys. Recently, however, fulminant hepatitis with high mortality was successfully reproduced in genotype 3 rabbit HEV-infected pregnant rabbits. Fetal mortality rates ranging from 67% to 80% were observed. Miscarriages, high fetal and maternal mortality, and fulminant hepatitis were also reproduced in these genotype 3 rabbit HEV-infected pregnant rabbits. The underlying mechanism of fulminant hepatitis with high mortality during pregnancy is still unknown, although the social economic status of the patients, hormonal changes during pregnancy, immunological status of the patients, and the existence of other co-infecting agents in patients could all play a potential role.

Neurological diseases: Neurological diseases such as Guillain-Barré syndrome, neuralgic amyotrophy, encephalitis, and myelitis have been associated with zoonotic genotype 3 HEV infection and, to lesser extent, genotypes 1 and 4 HEV infections. Guillain-Barré syndrome and neuralgic amyotrophy are associated with HEV infection in approximately 5% and 10% of the cases, respectively. In some patients, HEV RNA was detected in cerebrospinal fluid, indicating an infection of central nerve system. The neurological disorders in humans are mostly associated with infection by genotype 3 HEV that infects pig, deer, rabbit and mongoose, and occur during both acute and chronic HEV infections. Under lab conditions, numerous human neuronal-derived cell lines have been shown to support HEV replication. In animals peripherally infected with HEV, viral RNA and antigen were detected in brain tissues. The available data thus far suggest that HEV infects cells in the central nerve system leading to manifestation of various neurological diseases, which has become an emerging and significant clinical problem.

Pathogenesis

The mechanism of HEV pathogenesis is not well understood. HEV enters the host through the fecal-oral route. It is thought that, after oral ingestion of HEV, the virus first replicates in the gastrointestinal tract before reaching its target organ, the liver, via viremia. In pigs experimentally infected with human and swine HEVs, virus replication has been identified in small intestines, colon, hepatic and mesenteric lymph nodes, although the clinical significance of these extrahepatic sites of virus replication remains unknown. HEV replication in the liver has been detected in non-human primates, pigs and chickens experimentally infected with HEV. It is believed that, after replication in liver, HEV is released to the gallbladder from hepatocytes and is excreted in feces. Fecal virus shedding and viremia proceeds the onset of clinical and biochemical hepatitis, and virus shedding ends when the level of serum liver enzyme returns to baseline. In pigs experimentally infected with a genotype 3 human HEV, mild-to-moderate multifocal lymphoplasmacytic hepatitis and focal hepatocellular necrosis were observed. In chickens experimentally infected with an avian strain of HEV, gross lesions including subcapsular hemorrhages and slightly enlarged right intermediate lobe of the livers, and histological lesions including foci of lymphocytic periphlebitis and phlebitis were also observed.

Diagnosis

HEV serological assays are commercially available in some countries, although considerable variations in the specificity and sensitivity of these assays have been reported. These serological assays primarily use the recombinant ORF2 capsid protein or a mixture of both ORF2 and ORF3 proteins as the antigen. Some serological assays are approved for HEV diagnosis by regulatory agencies in Asia and Europe, however an FDA-approved diagnostic assay in the United States is still lacking. IgM anti-HEV is detectable in up to 90% of the acute HEV infections within 1–4 weeks after onset of the disease. IgG anti-HEV persists in infected individuals lasting from <1 year in some pediatric patients to 2–14 years in some adult patients.

RT-PCR and quantitative RT-PCR assays have also been developed to detect and quantify HEV genomes. However, significant variations in the sensitivity and specificity of these PCR-based assays were reported, largely due to the extensive heterogeneity of HEV. At least five genotypes of HEV, genotypes 1–4 and 7, within the species *Orthohepevirus A* are known to infect humans, thus stressing the need for a standardized universal PCR-based assay for HEV detection.

Treatment

Most HEV infections are self-limiting and acute in nature, and do not warrant antiviral treatment. However, a subpopulation of HEV-infected individuals, such as pregnant women with a high mortality and immunocompromised individuals with a high tendency of progressing into chronicity, do require antiviral treatment options. Currently there is no HEV-specific therapy, however broad-spectrum antivirals such as ribavirin and pegylated-interferon have been used with some success. Ribavirin monotherapy inhibits HEV replication and induces a sustained virus inhibition in patients with chronic HEV infections. Pegylated-interferon- α also induced a sustained antiviral response in HEV-infected patients. However, both ribavirin and interferon- α have severe side effects, and are contraindicated in pregnant women. Therefore, development of effective HEV-specific antivirals is warranted.

Prevention

A commercial vaccine based on the recombinant ORF2 capsid protein of a genotype 1 HEV has been licensed for use only in China but not elsewhere. Other experimental HEV vaccines based on the capsid protein also appear to be very promising. However, significant challenges remain for a more cost-effective modified live-attenuated vaccine due to the lack of an efficient cell culture system to propagate HEV. The practice of good hygiene and avoiding drinking water of unknown purity or consuming raw or undercooked animal meats are important preventive measures.

Further Reading

- Cao, D., Cao, Q.M., Subramaniam, S., *et al.*, 2017. Pig model mimicking chronic hepatitis E virus infection in immunocompromised patients to assess immune correlates during chronicity. *Proceedings of the National Academy of Sciences of the United States of America* 114 (27), 6914–6923.
- Dalton, H.R., Kamar, N., van Eijk, J.J., *et al.*, 2016. Hepatitis E virus and neurological injury. *Nature Reviews Neurology* 12 (2), 77–85.
- Dalton, H.R., Kamar, N., 2016. Treatment of hepatitis E virus. *Current Opinion in Infectious Diseases* 29 (6), 639–644.
- Kamar, N., Selves, J., Mansuy, J.M., *et al.*, 2008. Hepatitis E virus and chronic hepatitis in organ-transplant recipients. *The New England Journal of Medicine* 358 (8), 811–817.
- Kenney, S.P., Meng, X.J., 2018. Hepatitis E virus: Animal models and zoonosis. *Annual Review of Animal Biosciences* 7, 427–448. doi:10.1146/annurev-animal-020518-115117.
- Meng, X.J., Purcell, R.H., Halbur, P.G., *et al.*, 1997. A novel virus in swine is closely related to the human hepatitis E virus. *Proceedings of the National Academy of Sciences of the United States of America* 94, 9860–9865.

- Meng, X.J., 2014. Hepatitis E. In: Kaslow, R.A., Stanberry, L.A., LeDuc, J.W. (Eds.), *Viral Infections of Humans: Epidemiology and Control*, fifth ed. NY: Springer Publishing, pp. 439–454.
- Meng, X.J., 2016. Expanding host range and cross-species infection of hepatitis E virus. *PLoS Pathogens* 12 (8), e1005695.
- Meng, X.J., Shivaprasad, H.L., 2019. Avian hepatitis E virus infections. In: Swayne, D., Boulianne, M., McDougald, L., *et al.* (Eds.), *Diseases of Poultry*, fourteenth ed. Hoboken, NJ: John Wiley & Sons, Inc, pp. 528–534.
- Meng, X.J., Halbur, P.G., Opriessnig, T., 2019. Hepatitis E virus. In: Zimmerman, J.J., Karriker, L.A., Ramirez, A., *et al.* (Eds.), *Diseases of Swine*, eleventh ed. John Wiley & Sons, Inc, pp. 544–549.
- Nimgaonkar, I., Ding, Q., Schwartz, R.E., Ploss, A., 2018. Hepatitis E virus: Advances and challenges. *Nature Reviews Gastroenterology & Hepatology* 15 (2), 96–110.
- Purcell, R.H., Emerson, S.U., 2013. Hepatitis E virus. In: Knipe, D.M., Howley, P.M., Cohen, J.I. (Eds.), *Fields Virology*, sixth ed. Lippincott: Williams and Wilkins, Philadelphia, PA, pp. 2242–2258.
- Purdy, M.A., Harrison, T.J., Jameel, S., *et al.*, 2017. ICTV virus taxonomy profile: Hepeviridae. *Journal of General Virology* 98 (11), 2645–2646.
- Shukla, P., Nguyen, H.T., Torian, U., *et al.*, 2011. Cross-species infections of cultured cells by hepatitis E virus and discovery of an infectious virus-host recombinant. *Proceedings of the National Academy of Sciences of the United States of America* 108 (6), 2438–2443.
- Yin, X., Ying, D., Lhomme, S., *et al.*, 2018. Origin, antigenicity, and function of a secreted form of ORF2 in hepatitis E virus infection. *Proceedings of the National Academy of Sciences of the United States of America* 115 (18), 4773–4778.

Relevant Websites

- <https://ecdc.europa.eu/en/hepatitis-e/facts>
Facts about hepatitis E
European Center for Disease Control and Prevention.
- <https://www.who.int/news-room/fact-sheets/detail/hepatitis-e>
Hepatitis E infection
World Health Organization.
- <https://www.cdc.gov/hepatitis/hev/index.htm>
Hepatitis E information
Division of Viral Hepatitis
CDC.
- <https://discontools.eu/component/desease/desease.html?id=73>
Hepatitis E virus
Discontools.
- <https://www.viprbrc.org/brc/home.spg?decorator=hepe>
Hepeviridae
Virus Pathogen Database and Analysis Resources (ViPR).

Herpes Simplex Virus 1 and 2 (*Herpesviridae*)

David M Knipe, Harvard Medical School, Boston, MA, United States

Richard Whitley, University of Alabama at Birmingham, Birmingham, AL, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Encephalitis Infection of the central nervous system.

Gingivostomatitis Infection of the oral cavity.

Innate immunity Immune responses induced by recognition of pathogen-specific molecular patterns.

Intrinsic resistance Host factors that are constitutively expressed and restrict infection.

Keratitis Infection of the cornea.

Latent infection An infection in which the viral genome can be detected but no infectious virus can be detected, but in which infectious virus can be reactivated upon appropriate stimuli.

Lytic infection An infection that produces progeny virus and leads to death and lysis of the infected cells.

Pro-drug A compound that is metabolized in the body to an active drug molecule.

Classification

Human herpes virus 1 (HHV-1), commonly known as herpes simplex virus 1 (HSV-1), and human herpes virus 2 (HHV-2), commonly known as herpes simplex virus 2 (HSV-2), are two species of viruses in the *Simplexvirus* genus in the *Herpesviridae* family of the *Herpesvirales* order. These two viruses are sometimes called type 1 and type 2, but they are formally species, so the word “type” is incorrect. Many of the genetic and phenotypic properties of HSV-1 and HSV-2 are similar, so we will describe them together as HSV where genes and mechanisms are shared by the two viruses.

Virion Structure

The HSV virion (extracellular virus particle) consists of the core containing the double-stranded DNA genome surrounded by an icosahedral capsid, enclosed by a largely unstructured protein layer (tegument), surrounded by the outer lipid bilayer membrane (envelope) which contains embedded glycoprotein spikes (Fig. 1). The core consists of the DNA wrapped as a toroid in a liquid crystalline state. The DNA has no associated histones, and the phosphate charges on the DNA may be neutralized by polyamines. The capsid is a protein shell comprised of four viral proteins assembled with icosahedral symmetry based on 162 capsomeres or structural units defining the 16 triangular faces on the icosahedron. Twelve of the capsomeres form the vertices, of which 11 are pentons (five-fold symmetry) and one is the portal, through which virion genomic DNA is inserted, while 140 of the capsomeres are hexons (six-fold symmetry) that form the faces of the icosahedral structure. The tegument is comprised of at least 18 viral proteins and is largely unstructured, although there is some evidence for icosahedral organization imposed by interactions with the capsid and some granular and filamentous character. The envelope is a lipid bilayer membrane derived from host cell membranes that has as many as 13 viral proteins embedded in it.

Genome

The HSV genome is a linear double-stranded DNA molecule of approximately 150,000 base pairs (bp) with a one bp 3' extension at each end. The virion DNA contains nicks and gaps and a G + C content of 68% for HSV-1 and 69% for HSV-2 DNA. The HSV-1 and HSV-2 genomes are largely collinear with many of the lytic genes being highly homologous. HSV-1 X HSV-2 recombinants are readily isolated with cross-over events throughout the genomes, further demonstrating the similarity of the HSV-1 and HSV-2 genomes.

The HSV genome can be considered to consist of two covalently linked components, the L (long) and S (short), each of which is bounded by inverted repeats. The termini of the molecules contain a short repeated sequence that is also present in inverted forms at the L-S junction. The inverted repeated sequences provide homology for recombination that results in inversion of the L and S components, relative to each other. As a result, virion and infected cell DNAs contain approximately equimolar populations of four isomers of HSV DNA that result from inversion of one component or the other or both. The reason for the inverted repeated sequences has not been defined, but they may play a role in recombination-mediated replication of the viral DNA genome.

Life Cycle

HSV undergoes a lytic productive infection in epithelial cells and fibroblasts when it is inoculated onto the oral (usually HSV-1) or genital (usually HSV-2 but increasingly HSV-1) mucosa (Fig. 2). The virus spreads and causes the primary herpetic lesion. The virus

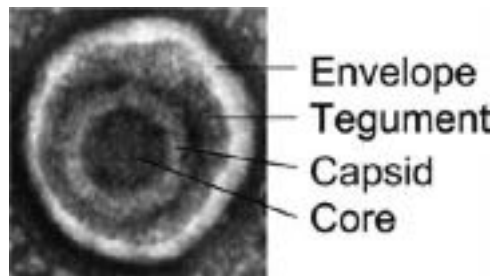


Fig. 1 Structure of the HSV virion. Electron micrograph of a negative-stained HSV-1 virion. The envelope, tegument, capsid, and core are indicated. Copyright, Travis Taylor and David Knipe.

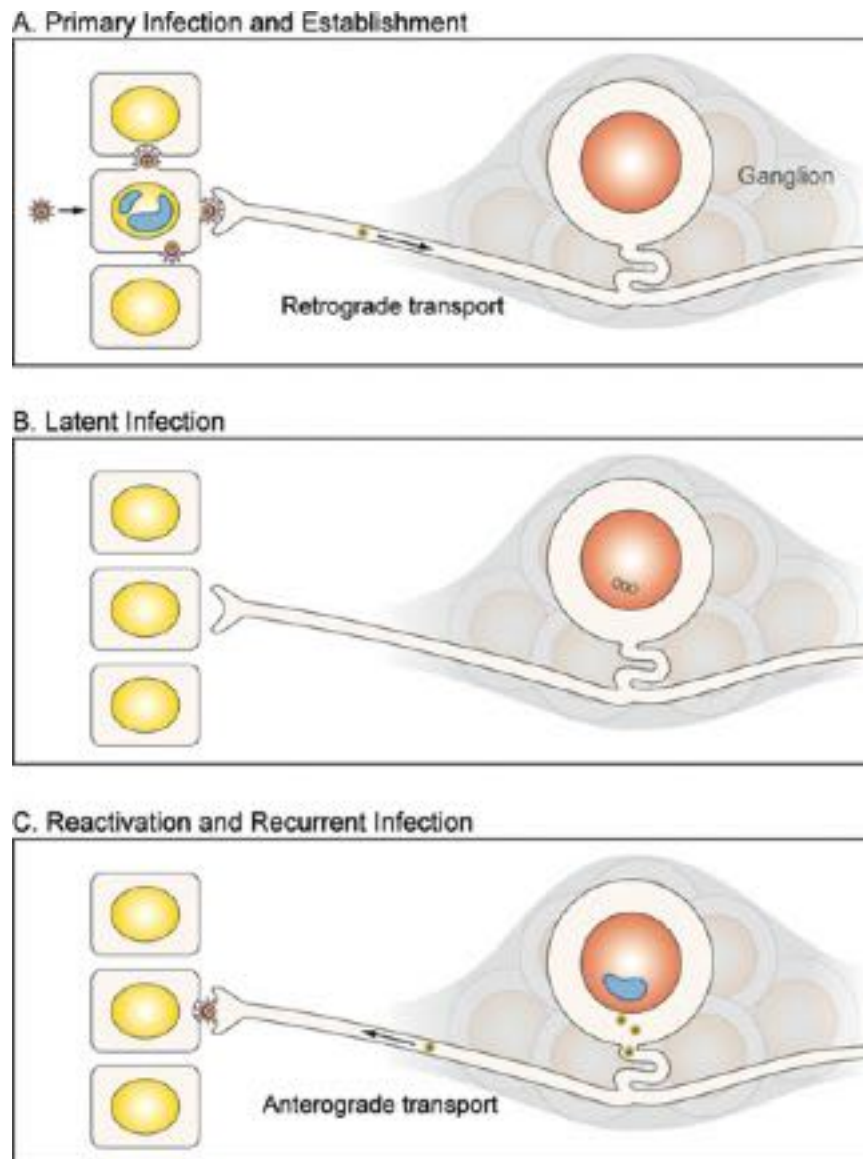


Fig. 2 Stages of herpes simplex virus (HSV) infection of the host. A: HSV is introduced onto a mucosal surface or a break in the skin, and it replicates productively in epithelial cells at the site of inoculation and spreads through the tissue. Virus enters sensory neuron axons and is transported to the cell body in a ganglion. B: HSV establishes a latent infection in the neuronal cell nucleus. Viral DNA is circular and assembled in chromatin. C: Upon neuronal damage or activation, the virus reactivates and undergoes at least a limited productive cycle. Capsids are transported by anterograde transport to the axonal termini, and virions are released. Reactivated virus causes a recurrent infection of the mucosal tissue, causing the shedding of virus. Copyright, Patrick T. Waters and David Knipe.

enters the axons of sensory neurons and establishes a latent infection in the neuronal cell body. Upon stress or damage to the neuron, the virus reactivates and causes the recurrent infection.

Lytic Infection

HSV lytic infection of epithelial cells or fibroblasts is initiated by binding of the virion to the cell surface and entry by fusion at the plasma membrane (Fig. 3, step 1a) or endocytosis followed by fusion in an endosome (Fig. 3, step 1b). The nucleocapsid and tegument are transported along microtubules (Fig. 3, step 2) to the nucleus where the capsid docks at the nuclear pore and releases the virion DNA into the nucleus. The input genome rapidly circularizes (Fig. 3, step 5) and is loaded with histones by host cell mechanisms. Virion protein 16 (VP16) localizes into the nucleus (Fig. 3, step 4) and assembles a complex of host proteins that promote transcription of the immediate-early (IE) mRNAs (Fig. 3, step 6). The IE mRNAs are transported to the cytoplasm where they are translated to yield the five IE proteins. Four of the five proteins localize to the infected cell nucleus where they promote early (E) and late (L) gene transcription. The E mRNAs are transported to the cytoplasm where they encode the E proteins, including seven viral proteins directly involved in viral DNA synthesis and other proteins that increase deoxynucleotide triphosphate levels for viral replication in resting cells. The E viral DNA polymerase and thymidine kinase (TK) are the targets of the nucleoside analogs used for antiviral therapy as described below.

The seven essential viral proteins together with host proteins replicate the input viral genome to yield progeny viral genomes as concatemers (Fig. 3, step 8). Replication of the viral DNA is thought to occur in two stages: (1) Cairns circle replication of the input viral genome, and (2) rolling circle and/or recombination-based replication generating concatemeric progeny molecules of viral DNA. Following initiation of viral DNA synthesis, the replicative intermediates and progeny DNA molecules serve as substrates for

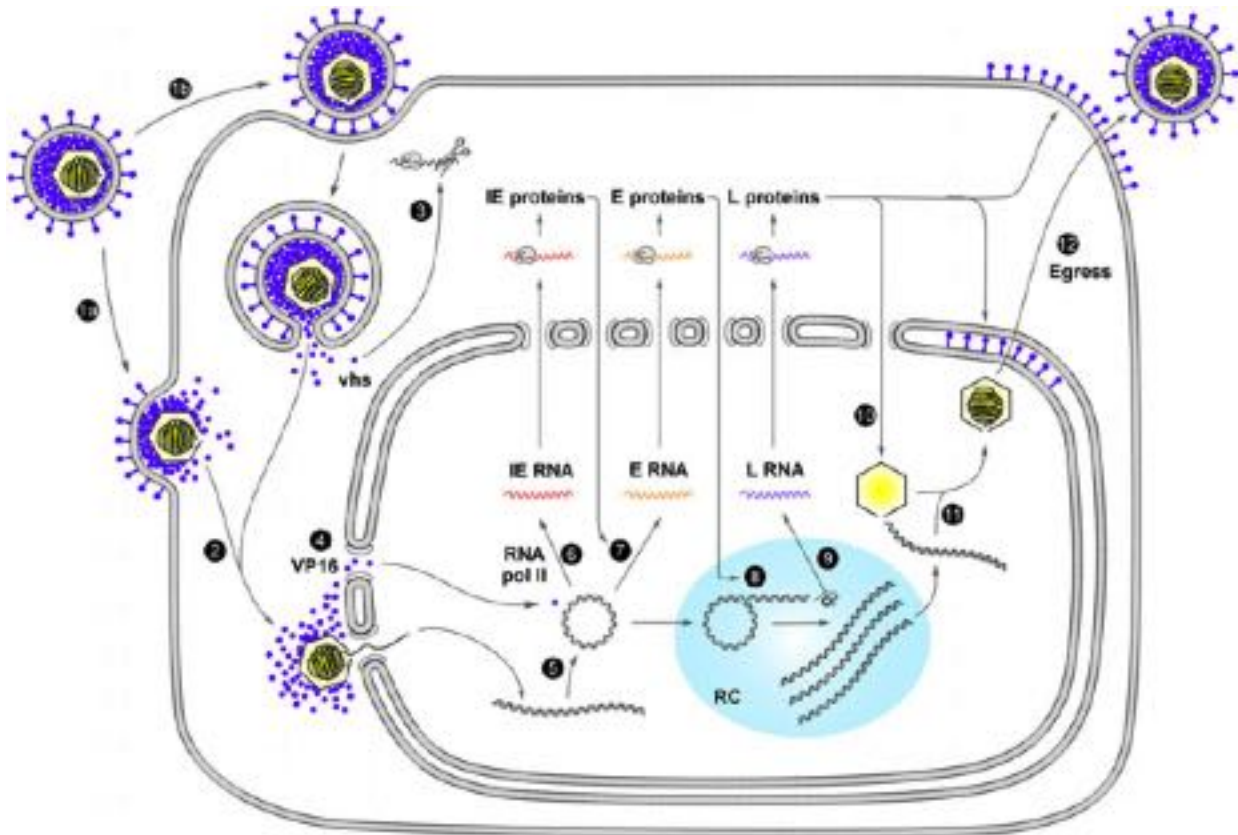


Fig. 3 Diagram of the replication cycle of herpes simplex virus. 1: The virus binds to the cell plasma membrane and the virion envelope fuses with the plasma membrane (1a) or the virus enters by endocytosis (1b), releasing the capsid and tegument. proteins into the cytoplasm. 2: The capsid is transported to the nuclear pore, where the viral DNA is released into the nucleus. 3: The vhs (virion host shutoff) protein causes degradation of host messenger RNAs (mRNAs). 4: VP16 localizes into the nucleus. 5: The viral DNA circularizes. 6: It is then transcribed by host RNA polymerase II to give first the immediate-early (IE) mRNAs. Gene transcription is stimulated by the VP16 tegument protein. Five of the six IE proteins act to regulate viral gene expression in the nucleus. 7: IE proteins transactivate early (E) gene transcription. 8: The E proteins are involved in replicating the viral DNA molecule. 9: Viral DNA synthesis stimulates late (L) gene expression. 10: The L proteins are involved in assembling the capsid in the nucleus and modifying the membranes for virion formation. 11: DNA is encapsidated in the capsid. 12: The filled capsid buds through the inner membrane to form an enveloped virion, and the virion exits from the cell by mechanisms described in the text. Copyright, Patrick T. Waters and David Knipe.

transcription of the L genes (Fig. 3, step 9). The L mRNAs are translated in the cytoplasm to yield structural proteins for assembly of progeny virions. The capsid proteins localize into the infected cell nucleus and assemble empty capsids (Fig. 3, step 10), and viral DNA is inserted into the capsid through the portal on the capsid (Fig. 3, step 11). Once the capsid is filled with approximately a unit-length of viral DNA, the concatemeric DNA is cleaved and the portal is plugged. Some tegument proteins are localized into the nucleus and inner nuclear membrane to promote budding of the nucleocapsid through the inner nuclear membrane. Egress of the virion from the host cell (Fig. 3, step 12) is thought to then involve a complicated pathway involving primary envelopment at the inner nuclear membrane, de-envelopment of the virus particle at the outer nuclear membrane, and then secondary envelopment at membranes of the Golgi apparatus. The enveloped virion then travels out of the cell by the secretory pathway.

Viral DNA replication, late gene transcription, and viral DNA encapsidation take place within virus-specific nuclear structures called replication compartments (Fig. 3, steps 8–10). The concentration of viral host factors in these factories is thought to enhance the rate of the reactions, as well as possibly provide a platform for the viral processes. The accumulation of viral nucleic acid and capsids in the replication compartments gives rise to the nuclear inclusion bodies that show specific staining properties characteristic of herpesviral infection.

The process of lytic infection involves the viral gene products inhibiting host cell metabolism, which eventually results in the death and lysis of the host cell. Shut off of the host cell protein synthesis involves the virion host shutoff (vhs) protein, a ribonuclease that targets to polyribosomes and degrades host cell mRNA (Fig. 3, step 3) and provides an opportunity for translation of host mRNAs. Shut off of host transcription and of host RNA splicing also reduces host gene expression.

Epigenetic Regulation

HSV virion DNA is not associated with histones, but when the viral DNA enters the cell nucleus, host cell mechanisms rapidly load histones with heterochromatic post-translational modifications on the viral DNA so as to silence it. Virion protein 16 (VP16) from the tegument is released into the cytoplasm and localizes into the infected cell nucleus where it organizes a complex of host cell proteins that remove the heterochromatin marks and reduce the histone loading on the viral DNA, so that transcription factors and RNA polymerase II can have access to the viral DNA for its transcription. The IE ICP0 protein then acts to allow E and L gene expression. ICP0 is an E3 ubiquitin ligase that promotes the degradation of a number of host restriction factor proteins that increase heterochromatin on HSV DNA, including PML, IFI16, and ATRX. This leads to euchromatin across the HSV genome, allowing broad transcription of the genome.

Host Factors

Like all viruses, HSV uses numerous host proteins and structures for its replication. Human cells have also evolved numerous mechanisms to restrict replication of HSV. First, the cells detect foreign viral macromolecules, and activate mechanisms by constitutively expressed host proteins that block the activity of the viral molecules, a process called intrinsic resistance. One of the first mechanisms of resistance is the rapid loading of heterochromatin on the incoming viral genome, as described above. Alternatively, recognition of HSV DNA, altered host RNAs, or viral proteins can induce innate responses which result in expression of host proteins that reduce viral replication in the infected cell and surrounding cells. These responses may involve induction of cytokines such as interferons and interleukins. Type 1 interferons such as interferon α and interferon β are produced in infected cells and bind to receptors on that cell and surrounding cells to induce the expression of multiple antiviral genes collectively called interferon-stimulated genes (ISGs). ISGs encode protein products that block or reduce viral gene expression or replication. Interleukins induced by viral infection may induce antiviral mechanisms or recruit immune cells to the sites of viral infection and that reduce viral replication.

HSV has evolved a number of mechanisms to fight back against intrinsic resistance and innate immunity mechanisms. As described above, ICP0 promotes the degradation of host epigenetic restriction factors. ICP0 also blocks certain innate signaling pathways. In addition, the HSV-1 ICP34.5 protein blocks type 1 interferon signaling so the virus is less sensitive to type 1 interferons. The HSV-1 US3 protein kinase also blocks interferon signaling. Other examples can be found in the cited reviews.

Latent Infection

HSV latent infection of sensory neurons involves the binding of the virion to the termini of the neuronal axons and entry, likely using the same entry molecules as for lytic infection described above, into the axon. The capsid moves along the axon by retrograde transport to the neuronal cell body, where the viral genome is released into the neuronal nucleus. The genome circularizes and is loaded slowly with histones due to the small pool of free histones in the resting neuron; nevertheless, lytic gene transcription is limited due to the lack of localization of VP16 to the neuronal nucleus and the cytoplasmic localization of at least one of the host proteins that forms a complex with VP16 to activate IE gene expression. The major viral gene that is transcribed is the gene encoding the latency-associated transcript (LAT), which is a long (greater than 10 kb) non-coding RNA primary transcript. LAT is processed by splicing to yield a stable 2 kb circular intron, and several microRNAs (miRNAs). LAT expression acts to promote the survival of the infected neuron by increasing heterochromatin on the viral genome and reducing viral lytic gene expression and inhibiting apoptosis. LAT increases the levels of the H3K9me3 (histone H3 lysine 3 residue trimethylated) facultative heterochromatin marker, possibly by

recruiting the Polycomb repressive complex 2 histone methyltransferase. Facultative heterochromatin seems ideal for latent HSV chromatin because it is silencing, but it would constitute a form of chromatin that is poised for reactivation.

During maintenance of latent infection, there is minimal but detectable expression of HSV lytic genes. Because the neuron is a resting cell, the viral DNA genome can be maintained stably without replicating or being tethered to host chromosomes. Some have argued that CD8⁺ T cells surrounding the latently infected neuron maintain latent infection, but an alternative explanation is that they are targeting reactivating neurons. In any event, neuronal damage, stress, and/or activation lead to kinase signaling pathways believed to phosphorylate histones in viral chromatin to allow generalized transcription of the latent genome to provide enough VP16 to trigger the lytic transcriptional cascade. This allows at least a limited amount of viral gene expression to form capsids, and capsids or virions move along the axon by anterograde transport to release reactivated at the axonal termini to cause a recurrent infection. It has been debated as to whether the reactivating neuron survives or not, but neuronal survival is likely inversely proportional to the level of reactivated gene expression and not an all-or-none effect.

Epidemiology

HSV infection is transmitted by direct contact with an infected person's lesions, mucosal surfaces, or genital or oral secretions. The epidemiology of infection is best defined by the seroprevalence of HSV-1 and HSV-2. HSV infections remain common despite a recent steady decrease in prevalence (Fig. 4). The seroprevalence of HSV-1 and HSV-2 increases linearly with age and is higher in females. Primary HSV-1 infections usually occur in the young child and are most often asymptomatic. During the period of 1999–2002, HSV-1 seroprevalence in U.S.-born children ages 6–13 years was 31%, progressively increasing with age. In the most recent National Health and Nutrition Examination Survey (NHANES) of 2015–2016, the seroprevalence in people age 14–19 years was 27%, consistent with overall trends of decreasing prevalence of HSV-1 in early childhood over the last two decades. By adulthood, about half of the U.S. population (48%) has experienced HSV-1 infection. HSV-1 prevalence differs by race and ethnicity, with prevalence highest in the Mexican-American population and lowest in non-Hispanic whites, a finding that remains true despite a steady decline in overall prevalence in the U.S. The HSV-1 contribution to genital herpetic disease has increased over the past two decades. The percentage of people who randomly tested positive for HSV-1 only and had a diagnosis of genital herpes increased to 1.8% in NHANES survey 1999–2004 when compared to 0.4% in the NHANES survey 1988–1994. A retrospective evaluation of college students with newly diagnosed genital infection from 1993 to 2001 showed a striking increase of HSV-1 as a cause of symptomatic genital herpes from 30.9% in 1993 to 77.6% in 2001.

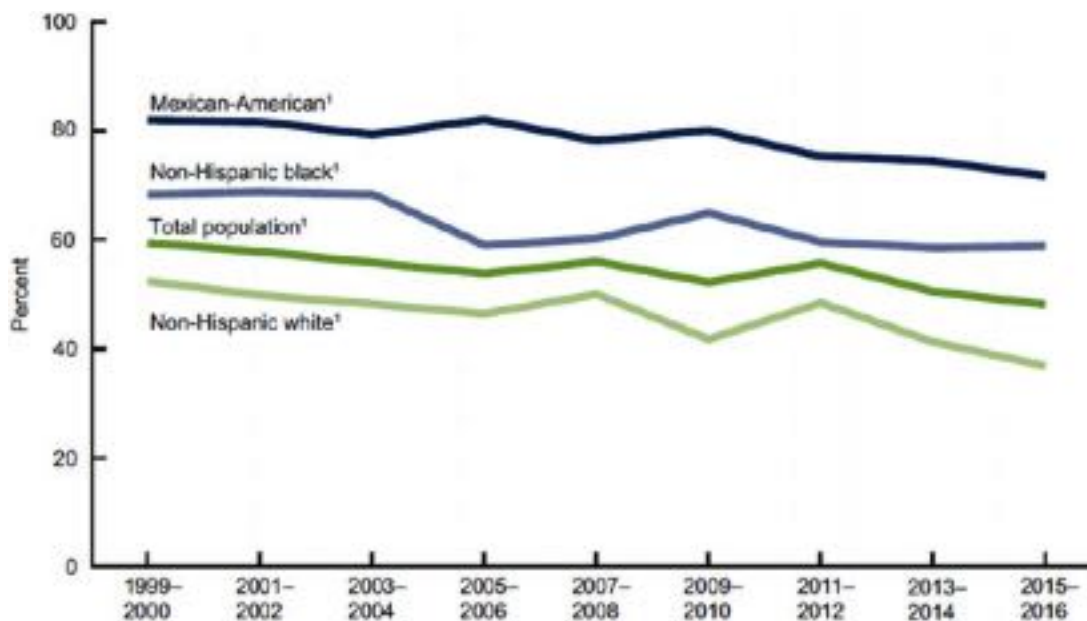


Fig. 4 Trends in age-adjusted prevalence of herpes simplex 1 among persons aged 14–49, for the total population and by race and Hispanic origin: United States, 1999–2000 through 2015–2016. ¹Significant decreasing linear trend over time, $p < 0.05$. **NOTES:** Age adjusted by the direct method to the 2000 U.S. Census population, using age groups 14–19, 20–29, 30–39, and 40–49 years. Total population includes all race and Hispanic-origin groups including those not shown separately. Data for the Asian subpopulation are only available since 2011, so this subpopulation is not shown separately, but included in the total population. Access data table at https://www.cdc.gov/nchs/data/databriefs/db304_table.pdf#2. Source: NCHS, National Health and Nutrition Examination Survey, 1999–2016.

Geographic location, socioeconomic status, and age influence the occurrence of HSV infection, regardless of the mode of assessment. The global estimates of the incidence and prevalence of HSV-1 infections vary by region, being around 67%, and highest in African, Southeast Asian and Western Pacific countries. The prevalence of HSV-1 is higher and occurs earlier in childhood outside of Europe and the Americas, with estimates reaching 90% in Africa by 10 years of age. As noted, most of these infections are asymptomatic.

The annual global incidence of HSV-1 infections is estimated to be 2% in 2012. The contribution of HSV-1 to the incidence of genital infections is less clearly defined in the developing world because prospective studies are lacking. Estimates of HSV-1 genital infections vary by region and are also impacted by both cultural sexual practices and the prevalence of childhood HSV-1 infections. In Africa, for example, the prevalence of genital HSV-1 is estimated to be 0%–0.1% (early childhood HSV-1 prevalence overall of 90%) compared to Eastern Mediterranean region with a 2%–3% prevalence of genital HSV-1 (early childhood HSV-1 prevalence close to 70%).

Acquisition of HSV-2 usually occurs in association with onset of sexual activity and directly correlates with the number of lifetime sexual partners. Overall, seroprevalence of HSV-2 in the U.S. decreased in recent years after a significant increase in the mid-1990s. The most recent studies showed total prevalence of 12% during 2015–2016, which represents a decrease from 17% during 1999–2004. HSV-2 remains more common among women when compared to men and non-Hispanic blacks compared to other race and ethnic groups. HSV-2 prevalence is also affected by factors such as age (linear increase with age), poverty (higher in people below poverty line), social status (higher in those who are divorced, separated or widowed), and number of lifetime sexual partners (3.8% in people who reported one partner compared to 39.9% in people reporting ≥ 50 sexual partners lifetime). Previous studies had indicated that men who have sex with men (MSM) have an increased prevalence of HSV-2. More recently, the NHANES 2001–2006 study did not detect a difference in HSV-2 prevalence overall between homosexual, bisexual or heterosexual men; however, there was a slightly increased prevalence of HSV-2 in homosexual individuals with HIV infection compared to heterosexual men. Worldwide estimates of HSV-2 infection in 2012 indicated an overall prevalence of 11.3%. Estimates vary significantly by gender and geographic area. Prevalence in females is higher, compared to males (14.8% versus 8%) and was highest in Africa with a prevalence of 31.5%. The estimated incidence of HSV-2 infections also differs by gender and geographic areas, being typically higher in the younger age groups. A decline of incidence with age was observed in areas with higher prevalence, which could be attributed to age-related behavioral changes or decrease in susceptible people at older ages. In general, women acquire HSV-2 infection more frequently than do men, irrespective of the number of partners.

For pregnant women, approximately 1% will excrete virus at the time of delivery. Nevertheless, the incidence of neonatal HSV infection is only approximately 1 in 1500 live born infants in the U.S. The incidence of neonatal HSV infection is affected by multiple factors, the most important of which is the maternal infection status. Pregnant women who contract genital HSV for the first time during the third trimester have the highest risk for transmitting infection to the fetus, as 57% of those neonates develop neonatal HSV infection. The rate of exposed neonates developing the disease decreases to 25% when pregnant women contract a first genital HSV infection but have a prior history of HSV infection of the other species. The risk drops significantly to 2% at most, for neonates born to women with established recurrent genital HSV of either species. The decrease in neonatal HSV infection with documented recurrent genital HSV infection is likely due to the protective transplacental antibodies. Cesarean delivery reduces the risk of neonatal HSV infection by avoiding exposure to viral shedding in maternal vaginal secretions. Other risk factors for neonatal HSV infection, particularly in women with established genital infection, include application of a fetal scalp electrode, virus species (higher risk with HSV-1 versus HSV-2), and duration of membrane rupture. Global estimates of annual neonatal HSV incidence is about 14,257 neonates from 2010 to 2015 with HSV-2 being responsible for about two thirds of the total number. The global rate of neonatal HSV is estimated to be around 10.3 per 100,000 live births. The estimated global annual incidence of neonatal HSV disease is significantly lower than that of the U.S. The reason behind this significant difference is not clear yet, and further studies are needed to identify potential explanations. Both HSV-1 and HSV-2 seem to contribute almost equally to neonatal HSV diseases in the U.S. This is not the case in other areas of the world where childhood HSV-2 prevalence is significantly higher, as in Africa where almost all neonatal HSV cases are thought to be due to HSV-2.

Nosocomial HSV infection has been documented both in newborn nurseries as well as in intensive care units.

Clinical Features

Mucocutaneous Infections

Gingivostomatitis

Mucocutaneous infections are the most common clinical manifestations of HSV-1 and HSV-2. Gingivostomatitis is usually caused by HSV-1 and occurs most frequently in children under five years of age. Clinical disease is characterized by fever, sore throat, pharyngeal edema, and erythema, followed by the development of vesicular or ulcerative lesions of the oral or pharyngeal mucosa. Skin lesions are usually grouped vesicles on an erythematous base. Mucous membrane vesicles usually rupture, resulting in a painful ulcerative lesion on an erythematous base, which is prone to bleeding. In children, clinical fussiness occurs and is associated with decreased oral intake that may lead to dehydration. Auto-inoculation of HSV from oral lesions to fingers results in herpetic whitlow, not uncommon in the pediatric population.

Another manifestation of mucocutaneous HSV is eczema herpeticum, which develops in patients with skin eczema and may not be limited to one area, but rather involves multiple patches of the skin affected by poorly controlled eczema. Mucocutaneous HSV lesions usually heal without leaving a scar at the site of infection. Recurrent HSV-1 infections of the oropharynx frequently manifest as herpes simplex labialis (cold sores) and appear on the vermilion border of the lip. Intraoral lesions, as a manifestation of recurrent disease, are uncommon in the normal host but do occur frequently in the immunocompromised host.

Genital herpes

HSV-2 was the major cause of genital herpes in the past, but recent studies indicate that probably over half of new genital herpes cases are caused by HSV-1. Primary infection in women usually involves the vulva, vagina, and cervix. In men, initial infection is most often associated with lesions on the glans penis, prepuce, or penile shaft. In individuals of either sex, primary disease is associated with fever, malaise, anorexia, and bilateral inguinal adenopathy. Women frequently have dysuria and urinary retention due to urethral involvement. As many as 10% of individuals will develop an aseptic meningitis with primary infection. Aseptic meningitis, often presenting as recurrent disease known as Mollaret's meningitis, usually presents with fever, photophobia, headache, and neck stiffness. It is more likely to happen in women with genital HSV-2 infections. Sacral radiculomyelitis may occur in both men and women, resulting in neuralgias, urinary retention, or obstipation. The complete healing of primary infection may take several weeks. The first episode of genital infection is less severe in individuals who have had previous HSV infections at other sites, such as herpes simplex labialis. Primary genital HSV is more likely to be asymptomatic regardless of viral species. Recurrent lesions with viral shedding could also be asymptomatic, explaining the high incidence of new infections.

Recurrent genital infections in either men or women can be particularly distressing. The frequency of recurrence varies significantly from one individual to another. For those with HSV-2 infection, approximately one-third of individuals with genital herpes have virtually no recurrences, one-third have approximately three recurrences per year, and another third have more than three per year, while those individuals with HSV-1 infection are much less likely to suffer recurrences. By applying polymerase chain reaction assays to genital swabs from women with a history of recurrent genital herpes, viral DNA can be detected in the absence of culture proof of infection. Further, HSV DNA can be detected in over 20% of days in women. This finding suggests the chronicity of genital herpes in some women, as opposed to a recurrent infection.

Epidemiological studies have also shown that HSV-2 infection increases the risk of human immunodeficiency virus (HIV) acquisition and transmission. It is postulated that inflammation in the genital tract caused by HSV-2 infection leads to increased activated CD4⁺ T cells that are the host cells for HIV infection, and this would lead to increased acquisition and transmission of HIV. This is of importance because genital herpes is potentially a major risk factor for HIV infection.

Herpetic keratitis

HSV keratitis is usually caused by HSV-1, either primary or recurrent and is accompanied by conjunctivitis in many cases. It is considered among the most common infectious causes of blindness in the U.S. The characteristic lesions of HSV keratoconjunctivitis are dendritic ulcers best detected by fluorescein staining. Deep stromal involvement has also been reported and may result in visual impairment. Acute necrotizing retinitis is a rare complication that leads to painless loss of vision. Chorioretinitis is another rare HSV ophthalmic complication that could develop in neonates or immunocompromised patients with HSV infections.

Other skin manifestations

HSV infections can manifest at any skin site. Common among health-care workers are lesions on abraded skin of the fingers, known as herpetic whitlows. Similarly, because of physical contact, wrestlers may develop disseminated cutaneous lesions known as herpes gladiatorum. HSV infections have also been recognized as a trigger for erythema multiforme.

Neonatal Herpes Simplex Virus Infection

The incidence of neonatal HSV is estimated to be 1 in 1500 live births in the U.S. Most (85%) affected neonates acquire HSV from exposure to infected secretions during birth (intrapartum acquisition), a small number (10%) acquire the virus postnatally, and rarely (~5%) the virus is acquired *in utero*, resulting in congenital HSV infection. Manifestations of neonatal HSV infection can be divided into three categories: (a) skin, eye, and mouth disease (SEM) (45% of all neonatal HSV cases); (b) CNS disease (30%); and (c) disseminated infection (25%). Most HSV infections during the neonatal period present with fever, lethargy, poor feeding, skin lesions, or seizures. In SEM, as the name implies, skin, eye, and mouth disease consists of cutaneous lesions (in >80%) and does not involve other organ systems. Involvement of the central nervous system may occur with encephalitis or disseminated infection and generally results in a diffuse encephalitis and results in seizures and lethargy. Cerebrospinal fluid analysis characteristically reveals an elevated protein and a mononuclear pleocytosis. Disseminated infection involves multiple organ systems and can produce disseminated intravascular coagulation, hemorrhagic pneumonitis, encephalitis, hepatitis, and cutaneous lesions. Diagnosis can be particularly difficult in the absence of skin lesions. The mortality rate for each disease classification varies from zero for skin, eye, and mouth disease to 15% for encephalitis and 60% or higher for neonates with disseminated infection in the absence of therapy. In addition to the high mortality associated with these infections, morbidity is significant in that children with encephalitis or disseminated disease develop normally in only approximately 40% of cases, even with the administration of appropriate antiviral therapy. Congenital HSV (acquired *in utero*, about 5% of neonatal HSV cases) manifests as a triad of skin

involvement (lesions, scarring and change of pigmentation), CNS involvement (calcifications and microcephaly), and ophthalmologic involvement (chorioretinitis and optic atrophy). Affected neonates are usually very ill.

Herpes Simplex Encephalitis

HSV encephalitis is characterized by hemorrhagic necrosis of the inferomedial portion of the temporal lobe. Disease begins unilaterally, then spreads to the contralateral temporal lobe. It is the most common cause of focal, sporadic encephalitis in the U.S. today and occurs in approximately 1 in 250,000 to 500,000 individuals. Most cases are caused by HSV-1. The actual pathogenesis of herpes simplex encephalitis is unknown, although it has been speculated that primary or recurrent virus can reach the temporal lobe by ascending neural pathways, such as the trigeminal tracts or the olfactory nerves.

Clinical manifestations of herpes simplex encephalitis include headache, fever, altered consciousness, and abnormalities of speech and behavior. Focal seizures may also occur. The cerebrospinal fluid formulae for these patients is variable, but usually consists of a pleocytosis of monocytes. The protein concentration is characteristically elevated, and glucose is usually normal. Red blood cells may or may not be present. Historically, a definitive diagnosis could be achieved only by brain biopsy, because other pathogens may produce a clinically similar illness. However, the application of polymerase chain reaction (PCR) for detection of virus DNA has replaced brain biopsy as the standard for diagnosis. The mortality and morbidity are high, even when appropriate antiviral therapy is administered. At present, the mortality rate is approximately 14%–19% even with appropriate intravenous (IV) acyclovir (mortality exceeds 70% without effective therapy). Approximately 50% of survivors will have significant neurologic sequelae.

Herpes Simplex Virus Infections in the Immunocompromised Host

HSV infections in the immunocompromised host are clinically more severe, may be progressive, and require more time for healing. Manifestations of HSV infections in this patient population include pneumonitis, esophagitis, hepatitis, colitis, and disseminated cutaneous disease. Individuals suffering from human immunodeficiency virus infection may have extensive perineal or orofacial ulcerations. HSV infections are also noted to be of increased severity in individuals who are burned.

Pathogenesis

Acquisition of HSV infections requires intimate contact between a person who is shedding virus and a susceptible host. After inoculation of HSV onto the skin or mucous membrane and an incubation period of 4–6 days, clinical disease may occur or, as is more common, the infection will be asymptomatic. HSV replicates in epithelial cells and fibroblasts. As replication continues, cell lysis and local inflammation ensue, resulting in characteristic vesicles on an erythematous base. Regional lymphatics and lymph nodes become involved; therefore, viremia and visceral dissemination may develop, depending upon the immunologic competence of the host. In all hosts, the virus generally ascends peripheral sensory nerves and reaches sensory ganglia. Replication of HSV within neural tissue is followed by anterograde axonal spread of the virus back to other mucosal and skin surfaces via the peripheral sensory nerves. Virus replicates further in the epithelial cells, reproducing the lesions of the initial infection, until infection is contained through both systemic and mucosal immune responses.

The initial host response to HSV infection is derived mainly from the innate immune system including natural killer (NK) cells, neutrophils, B cells, and T cells, as well as production of cytokines and recruitment of adaptive and humoral immune response. HSV-specific antibodies and, subsequently, CD8 and CD4 lymphocytes develop with time. Prevention of latent infection is not possible; thus, latency is inevitable. The production of HSV-specific neutralizing antibodies contributes to the control of infection and is also used in laboratory testing to detect the viral infection.

Latency is established when HSV reaches the dorsal root ganglia after retrograde transmission via sensory nerve pathways. The virus does not actively replicate during latency. Reactivation of latent infection results in one of two outcomes: asymptomatic viral shedding (more likely) or reactivation that results in disease with a clinical presentation similar to, but usually milder than, the original symptomatic infection. Reactivation of a latent infection may be symptomatic despite the asymptomatic nature of the original primary infection.

Rarely, HSV can infect the central nervous system and cause encephalitis. The focality and temporal lobe affinity suggest direct extension of virus along neural tracts. Encephalitis caused by HSV is characterized by necrosis of the inferior medial portion of the temporal lobe, initially unilaterally and then contralaterally. This necrotic process accounts for the high morbidity and mortality of infection. Specific genetic mutations in the innate immune system make affected hosts susceptible to severe or recurrent disease. For example, toll-like receptor 3 (TLR-3) mutations have been reported to increase the risk for HSE in pediatric and young adults from the same family.

Infection of the neonate is usually the consequence of direct contact with infected maternal genital secretions, accounting for approximately 85% of cases of neonatal herpes. The remaining 15% are caused by *in utero* infection (5%), secondary to viremia, or postnatal acquisition whereby the baby comes in contact with infectious virus in the environment.

Diagnosis

The diagnosis of HSV infections is predicated on clinical evaluation of mucocutaneous manifestations. However, confirmation of the diagnosis requires isolation of HSV in appropriate cell culture systems or the detection of viral gene products or, alternatively,

the detection of viral DNA by polymerase chain reaction (PCR). HSV grows readily in tissue culture, producing cytopathic effects within a few days in a wide variety of mammalian cell lines. Routine testing to distinguish HSV-1 from HSV-2 in the isolate is important for genital isolates because HSV-1 is much less likely to recur.

PCR is an important method for diagnosing HSV infections, particularly those involving the central nervous system. The detection of HSV DNA by PCR in the CSF has replaced brain biopsy as a method of diagnosis of central nervous system infections.

Serologic assays that distinguish HSV-1-versus HSV-2-specific antibodies are commercially available. The utilization of immunoblot detection of specific glycoproteins that distinguish HSV-1 from HSV-2, namely, glycoprotein G-1 (gG-1) and gG-2, are available for determining prior exposure to HSV-1 and HSV-2 infections. Serologic testing is also used for population-based studies to identify the prevalence of the disease.

Historically, Tzanck smears have been used to diagnose HSV infections. Tzanck smears are not sensitive enough for routine diagnostic purposes. However, immunofluorescent staining of cell trap preparations from lesions is both sensitive and specific for the diagnosis for HSV infections.

Treatment

Infections due to HSV are amenable to treatment. Acyclovir and its pro-drug (valacyclovir) as well as famciclovir (pro-drug of penciclovir) are the mainstays of treatment. Acyclovir is mono-phosphorylated by the viral TK, and di- and tri-phosphorylates are added by cellular enzymes. Acyclovir tri-phosphate is recognized specifically by the viral DNA polymerase and incorporated into the growing viral DNA chain. Acyclovir causes chain termination and inhibits viral DNA synthesis. Valacyclovir is a pro-drug that metabolizes to acyclovir and results in good serum levels, similar to levels achieved with IV acyclovir. Famciclovir undergoes metabolic conversion to penciclovir which is active against HSV. Dose, route, and duration of therapy vary considerably depending on the site of infection and status of the patient. The use of high-dose acyclovir and addition of suppressive therapy has improved outcomes significantly in neonatal HSV infections.

Primary mucocutaneous infections in the immunocompetent host are treated with oral acyclovir or valacyclovir for 7–10 days; famciclovir can also be used. Valacyclovir is FDA-approved in the pediatric population 12 years of age or older. Recurrent orolabial HSV infection can be treated with any of these three drugs. Suppressive therapy may be considered for patients with very frequent recurrent lesions. Patients with genital HSV infection can be treated with acyclovir, valacyclovir, or famciclovir. IV acyclovir should be considered in the initial phase of treating severe orolabial disease leading to decreased oral intake in a pediatric patient or in an adult with genital herpes associated with aseptic meningitis. Treatment of genital herpes in the HIV-infected patient may be followed by long term suppressive therapy with acyclovir, valacyclovir or famciclovir.

Topical therapy with one of several antiviral ophthalmic preparations is appropriate for HSV keratoconjunctivitis. However, the treatment of choice is Viroptic or trifluorothymidine. Secondary choices include vidarabine ophthalmic or topical idoxuridine. Consultation with an ophthalmologist for confirmation of the diagnosis and management of HSV ocular disease is indicated.

Neonatal HSV infection should always be treated with IV acyclovir. The duration of treatment varies: CNS and disseminated diseases are treated for 21 days and SEM only 14 days. All forms of neonatal HSV should be followed by a minimum of 6 months of suppressive therapy with oral acyclovir. Neonatal disease involving the eyes should also be treated with topical ophthalmic drugs (trifluridine or topical ganciclovir), and an ophthalmologist consulted.

IV acyclovir is the mainstay of HSV encephalitis during adulthood. Encephalitis patients require longer treatment duration of 14–21 days as compared to 7–10 days for aseptic meningitis. Acyclovir resistance can develop. Resistance is usually related to viral TK mutations. The treatment of choice for infections caused by acyclovir-resistant strains is foscarnet. Consultation with specialists is recommended when resistance is known or suspected.

Prevention

At the present, there is no licensed vaccine for the prevention for HSV infections. Vaccines in clinical testing include DNA vaccines, replication-defective viruses, and protein subunits with new adjuvants. Consequently, prevention of HSV infections resides in the most part on knowledge of the mechanisms of transmission, including person to person as well as in the hospital environment. Individuals with known recurrent HSV infections should be counseled on the possibility of transmission of infection while lesions are present. The use of condoms for individuals with recurrent genital herpes is encouraged in that detection of HSV DNA by PCR can occur even in the absence of lesions. Similarly, for individuals who have recurrent herpes labialis, kissing should be discouraged.

There is a risk of nosocomial transmission of HSV within the hospital environment. Because many individuals shed HSV in the absence of clinical symptoms, it is impossible to exclude all workers from the hospital environment who could transmit infection. Thus, many authorities simply recommend strict hand washing and covering of lesions, should they exist.

Finally, no data exist on the prevention of neonatal HSV infection. It has been theorized that anticipatory administration of acyclovir to babies delivered through an infected birth canal may prove of value, particularly for women who have first episode genital herpetic infection. However, no data exist to substantiate this hypothesis. Because over 1% of all women at delivery excrete HSV and the rate of neonatal HSV infection is only 1 in 1500 live born infants as noted earlier, the routine administration of acyclovir to all children born to HSV-positive women is not reasonable. Alternative approaches, namely administration of acyclovir to known HSV-2-infected women is gaining acceptance.

Further Reading

- Baines, J.D., 2011. Herpes simplex virus capsid assembly and DNA packaging: A present and future antiviral drug target. *Trends Microbiology* 19, 606–613.
- Dai, X., Zhou, Z.H., 2018. Structure of the herpes simplex virus 1 capsid with associated tegument protein complexes. *Science* 360, eaao7298.
- Johnston, C., Gottlieb, S.L., Wald, A., 2016. Status of vaccine research and development of vaccines for herpes simplex virus. *Vaccine* 34, 2948–2952.
- Knipe, D.M., Lieberman, P.M., Jung, J.U., *et al.*, 2013. Snapshots: Chromatin control of viral infection. *Virology* 435, 141–156.
- Kurt-Jones, E.A., Orzalli, M.H., Knipe, D.M., 2017. Innate immune mechanisms and herpes simplex virus infection and disease. In: Osterrieder, K. (Ed.), *Cell Biology of Herpes Viruses*. *Advances in Anatomy, Embryology and Cell Biology* 223. Springer, pp. 49–75.
- McQuillan, G., Kruszon-Moran, D., Flagg, E.W., Paulose-Ram, R., 2018. Prevalence of herpes simplex virus type 1 and type 2 in persons aged 14–49: United States, 2015–2016. *NCHS Data Brief* 304, 1–8.
- Orzalli, M.H., Knipe, D.M., 2014. Cellular sensing of viral DNA and viral evasion mechanisms. *Annual Reviews of Microbiology* 68, 477–492.
- Roizman, B., Knipe, D.M., Whitley, R.J., 2013. Herpes simplex virus. In: Knipe, D.M., Howley, P. (Eds.), *Fields Virology*, sixth ed. Philadelphia: Lippincott Williams & Wilkins, pp. 1823–1897.
- Whitley, R.J., Roizman, B., 2017. Herpes simplex viruses. In: Richman, D.D., Whitley, R.J., Hayden, F.G. (Eds.), *Clinical Virology*, fourth ed. Washington DC: ASM Press, pp. 415–445.

Relevant Websites

- <https://www.cdc.gov/std/herpes/default.htm>
Genital Herpes.
- <https://www.who.int/news-room/fact-sheets/detail/herpes-simplex-virus>
Herpes simplex virus
World Health Organization.

History of Virology: Vertebrate Viruses

F Fenner, Australian National University, Canberra, ACT, Australia

© 2008 Elsevier Ltd. All rights reserved.

This is a reproduction of F. Fenner, History of Virology: Vertebrate Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00424-6](https://doi.org/10.1016/B978-012374410-4.00424-6).

Introduction

As in other branches of experimental science, the development of our knowledge of viruses has depended on the techniques available. Initially, viruses were studied by pathologists interested in the causes of the infectious diseases of man and his domesticated animals and plants and these concerns remain the main force advancing the subject. The idea that viruses might be used to probe fundamental problems of biology arose in the early 1940s, with the development of the knowledge of bacterial viruses. These studies helped establish the new field of molecular virology in the period 1950–70, and this has revolutionized the study of viruses and led to an explosion of knowledge about them.

The Word Virus

Since antiquity the term virus had been synonymous with poison, but during the late nineteenth century it became a synonym for microbe (Pasteur's word for an infectious agent). It did not acquire its present connotation until the 1890s, after the bacterial or fungal causes of many infectious diseases had been discovered, using the agar plate, effective staining methods, and efficient microscopes. It then became apparent that there were a number of infectious diseases of animals and plants from which no bacterium or fungus could be isolated or visualized with the microscope. After the introduction in 1884 of the Chamberland filter, which held back bacteria, Loeffler and Frosch demonstrated that the cause of foot-and-mouth disease was a filterable (or ultramicroscopic) virus. The first compendium of all then known viruses was edited by T. M. Rivers of the Rockefeller Institute and published in 1928. Entitled *Filterable Viruses*, this emphasized that viruses required living cells for their multiplication. In the 1930s chemical studies of the particles of tobacco mosaic virus and of bacteriophages showed that they differed from all cells in that at their simplest they consisted of protein and nucleic acid, which was either DNA or RNA. Gradually, the adjectives filterable and ultramicroscopic were dropped and the word viruses developed its present connotation.

Early Investigations

Foot-and-Mouth Disease Virus

In 1898 F. J. Loeffler and P. Frosch described the filterability of an animal virus for the first time, noting that “the filtered material contained a dissolved poison of extraordinary power or that the as yet undiscovered agents of an infectious disease were so small that they were able to pass through the pores of a filter definitely capable of retaining the smallest known bacteria.” Although the causative agent of foot-and-mouth disease passed through a Chamberland-type filter, it did not go through a Kitasato filter which had a finer grain. This led to the conclusion that the causative virus, which was multiplying in the host, was a corpuscular particle. Loeffler and Frosch gave filtration a new emphasis by focussing attention on what passed through the filter rather than what was retained and established an experimental methodology which was widely adopted in the early twentieth century in research on viral diseases.

Yellow Fever Virus

Following the acceptance of the notion of filterable infectious agents, pathologists investigated diseases from which no bacteria could be isolated and several were soon shown to be caused by viruses. One of the most fruitful investigations, in terms of new concepts, was the work of the United States Army Yellow Fever Commission headed by Walter Reed in 1900–01. Using human volunteers, they demonstrated that yellow fever was caused by a filterable virus which was transmitted by mosquitoes and that the principal vector was *Aedes aegypti*. They also showed that infected persons were infectious for mosquitoes only during the first few days of the disease and that mosquitoes were not infectious until 7–10 days after imbibing infectious blood, thus defining the extrinsic incubation period and establishing essentially all of the basic principles of the epidemiology of what came to be called arboviruses (arthropod-borne viruses).

Physical Studies of Viruses

Further advances in understanding the nature of viruses depended on physical and chemical studies. As early as 1907 H. Bechhold in Germany developed filters of graded permeabilities and a method of determining their pore size. Subsequently, W. J. Elford, in London, used such membranes for determining the size of animal virus particles with remarkable accuracy.

In Germany, on the eve of World War II, H. Ruska and his colleagues had produced electron microscopic photographs of the particles of tobacco mosaic virus, bacteriophages, and poxviruses. Technical improvements in instrumentation after the war, and the introduction of negative staining for studying the structure of viruses by Cambridge scientists H. E. Huxley in 1957 and S. Brenner and R. W. Horne in 1959 resulted in photographs of the particles of virtually every known kind of virus. These demonstrated the variety of their size, shape, and structure, and the presence of common features such as the icosahedral symmetry of many viruses of animals.

Following a perceptive paper on the structure of small virus particles by F. H. C. Crick and J. D. Watson in 1956, in 1962 D. L. D. Caspar of Boston and A. Klug of Cambridge, England, produced a general theory of the structure of regular virus particles of helical and icosahedral symmetry. Structure of the virus particle became one of the three major criteria in the system of classification of viruses that was introduced in 1966.

The Chemical Composition of Viruses

If the ultramicroscopic particles found in virus-infected hosts were the pathogens, what did these particles consist of? Following observations on plant and bacterial viruses in 1935, in 1940 C. L. Hoagland and his colleagues at the Rockefeller Institute found that vaccinia virus particles contained DNA but no RNA. Thus evidence was accumulating that viruses differed from bacteria not only in size and their inability to grow in lifeless media, but in that they contained only one kind of nucleic acid, which could be either DNA or RNA.

The development of restriction-endonuclease digestion of DNA, based on studies of phage restriction by W. Arber (Nobel Prize, 1978) of Basel in the 1960s, and then elaborated by biochemist H. O. Smith of Baltimore in 1970, has simplified the mapping of the genomes of DNA viruses, a study initiated by D. Nathans (Nobel Prize, 1978), also of Baltimore, using Simian virus 40. The development of the polymerase chain reaction in 1985 by K. B. Mullis (Nobel Prize, 1993) revolutionized sequencing methods for both DNA and RNA, leading to the availability of the complete genomic sequences of most viruses and the ability to make diagnoses using minute amounts of material.

Investigations of bacterial viruses were motivated by scientific curiosity, but animal virology was developed by pathologists studying the large number of diseases of humans and livestock caused by viruses, and it has retained this practical bias. Over the first two decades of the twentieth century, testing of filtrates of material from a number of infected humans and animals confirmed that they were caused by viruses. Among the most important was the demonstration in 1911 by P. Rous (Nobel Prize, 1966), at the Rockefeller Institute, that a sarcoma of fowls could be transmitted by a bacteria-free filtrate.

The Cultivation of Animal Viruses

The first systematic use of small animals for virus research was Pasteur's use of rabbits, which he inoculated intracerebrally with rabies virus in 1881. It was not until 1930 that mice were used for virus research, with intracerebral inoculations of rabies and yellow fever viruses. By using graded dilutions and large numbers of mice, quantitative animal virology had begun. The next important step, initiated by E. W. Goodpasture in 1931–32, was the use of chick embryos for growing poxviruses. This was followed a few years later by the demonstration by F. M. Burnet that many viruses could be titrated by counting the pocks that they produced on the chorioallantoic membrane, whereas others grew well in the allantoic and/or amniotic cavities of developing chick embryos.

Tissue cultures had first been used for cultivating vaccinia virus in 1928, but it was the discovery by J. F. Enders, T. H. Weller, and F. C. Robbins of Harvard University (Nobel Prize, 1954) in 1949 that poliovirus would grow in non-neural cells that gave a tremendous stimulus to the use of cultured cells in virology. Over the next few years their use led to the cultivation of medically important viruses such as those causing measles (by J. F. Enders and T. C. Peebles in 1954) and rubella (by T. H. Weller and F. C. Neva in 1962). Even more dramatic was the isolation of a wide variety of new viruses, belonging to many different families. Also the different cytopathic effects produced by different viruses in monolayer cell cultures were found to be diagnostic.

The next great advance, which greatly increased the accuracy of quantitative animal virology, occurred in 1952, when the plaque assay method for counting phages was adapted to animal virology by R. Dulbecco (Nobel Prize, 1975), using a monolayer of chick embryo cells in a petridish. In 1958 H. Temin and H. Rubin applied Dulbecco's method to Rous sarcoma virus, initiating quantitative studies of tumor viruses. Biochemical studies of animal virus replication were simplified by using continuous cell lines and by growing the cells in suspension.

Biochemistry

During the 1950s virus particles were thought to be 'inert' packages of nucleic acid and proteins, although in 1942 G. K. Hirst of the Rockefeller Institute had shown that particles of influenza virus contained an enzyme, later identified by A. Gottschalk of Melbourne as a neuraminidase.

Further advances in viral biochemistry depended on methods of purification of virus particles, especially the technique of density gradient centrifugation. In 1967 J. Kates and B. R. McAuslan demonstrated the presence of a DNA-dependent RNA polymerase in purified vaccinia virus virions, a discovery that was followed the next year by the demonstration of a double-stranded (ds) RNA-dependent RNA polymerase in reovirus virions and then of single-stranded (ss) RNA-dependent RNA polymerases in virions of paramyxoviruses and rhabdoviruses. In 1970 came the revolutionary discovery by D. Baltimore of Boston (Nobel Prize, 1975) and H. Temin of Wisconsin (Nobel Prize, 1975), independently, of the RNA-dependent DNA polymerase, or reverse transcriptase, of Rous sarcoma virus. Many other kinds of enzymes were later identified in the larger viruses; for example, no less than 16 enzymes have now been identified in vaccinia virus virions.

Just as investigations with bacterial viruses were of critical importance in the development of molecular biology, studies with animal viruses have led to the discovery of several processes that have proved to be important in the molecular biology of eukaryotic cells, although many of these discoveries are too recent for serious historical appraisal. Thus, in 1968 M. Jacobson and D. Baltimore of Boston showed that the genomic RNA of poliovirus was translated as a very large protein, which was then cleaved by a protease that was subsequently found in the viral replicase gene. RNA splicing was discovered in 1977, independently by P. A. Sharp (Nobel Prize, 1993) and L. T. Chow and their respective colleagues during studies of adenovirus replication. Work with the same virus by Rekosh and colleagues in London in 1977 resulted in the definition of viral and cell factors for initiation of new DNA strands using a novel protein priming mechanism. Capping of mRNA by m⁷G and its role in translation was discovered in 1974, during work with reoviruses by A. J. Shatkin. Other processes first observed with animal viruses and now known to be important in eukaryotic cells were the existence of 3' poly A tracts on mRNAs, the pathway for synthesis of cell surface proteins, and the role of enhancer elements in transcription.

Analysis of viral nucleic acids progressed in parallel with that of the viral proteins. Animal viruses were found with genomes of ssRNA, either as a single molecule, two identical molecules (diploid retroviruses), or segmented; dsRNA; ssDNA, dsDNA, and partially dsDNA. The so-called unconventional viruses or prions of scrapie, kuru, and Creutzfeld-Jacob disease, which have been extensively investigated since 1957 by D. C. Gajdusek of the US National Institutes of Health (Nobel Prize, 1976) and S. B. Prusiner of the University of California at San Francisco (Nobel Prize, 1997), appear to be infectious proteins. Soon after the discovery that the isolated nucleic acid of tobacco mosaic virus was infectious (see below), it was shown that the genomic RNAs of viruses belonging to several families of animal viruses were infectious; namely those with single, positive-sense RNA molecules. Then in 1973 came the discovery of recombinant DNA by P. Berg of Stanford University and his colleagues, using the animal virus SV40 and bacteriophage λ .

Structure of the Virion

Using more sophisticated methods of electron microscopy, many animal virus virions have been shown to be isometric icosahedral structures, or roughly spherical protein complexes surrounded by a lipid-containing shell, the envelope, which contains a number of virus-coded glycoprotein spikes. X-ray crystallography of crystals of purified isometric viruses has revealed the molecular structure of their icosahedra; similar revealing detail has been obtained for the neuraminidase and hemagglutinin spikes of influenza virus.

Tumor Virology

After the discovery of Rous sarcoma virus in 1911, there was a long interval before the second virus to cause tumors, rabbit papilloma virus, was discovered by R. E. Shope in 1933. However, after that viruses were found to cause various neoplasms in mice, and in 1962 J. J. Trentin of Yale University showed that a human adenovirus would produce malignant tumors in hamsters. Since then, direct proof has been obtained that several DNA viruses (but not adenoviruses) and certain retroviruses can cause tumors in humans. Molecular biological studies have shown that oncogenicity is largely caused by proteins they produce that are encoded by viral oncogenes, a concept introduced by R. J. Huebner and G. Todaro of the US National Institutes of Health in 1969 and corrected, refined, and greatly expanded since the mid-1970s by J. M. Bishop, H. Varmus (Nobel Prize, 1989), and R. A. Weinberg of Boston.

Only one group of RNA viruses, the retroviruses, which replicate through an integrated DNA provirus, cause neoplasms, whereas viruses of five groups of DNA viruses are tumorigenic. Study of these oncogenic viruses has shed a great deal of light on the mechanisms of carcinogenesis. Oncogenic DNA viruses contain oncogenes as an essential part of their genome, which when integrated into the host cell DNA may promote cell transformation, whereas the oncogenes of retroviruses are derived from proto-oncogenes of the cell.

Impact on Immunology

Immunology arose as a branch of microbiology and several discoveries with animal viruses were important in the development of important concepts in immunology. It was the discovery in the 1930s by E. Traub in Tübingen of persistent infection of mice with

lymphocytic choriomeningitis virus that led to the development by F. M. Burnet (Nobel Prize, 1960) of the concept of immunological tolerance in 1949. It was work with the same virus that led to the discovery of MHC restriction by P. C. Doherty and R. Zinkernagel in Canberra (Nobel Prize, 1996) in 1974.

Vaccines and Disease Control

From the time of Pasteur's use of rabies vaccine in 1885, a major concern of medical and veterinary virologists has been the development of vaccines. Highlights in this process have been the development of the 17D strain of yellow fever virus by M. Theiler (Nobel Prize, 1951) of the Rockefeller Foundation laboratories in 1937, the introduction of influenza vaccine in 1942, based on Burnet's 1941 discovery that the virus would grow in the allantois, the licensing in 1954 of inactivated poliovaccine developed by J. Salk and in 1961 of the live poliovaccine introduced by A. B. Sabin, both based on the cultivation of the virus in 1949 by Enders, and development of the first genetically engineered human vaccine with the licensing of yeast-grown hepatitis B vaccine in 1986. Finally, 1977 saw the last case of natural smallpox in the world and thus the first example of the global eradication of a major human infectious disease, the result of a 10-year campaign conducted by the World Health Organization with a vaccine directly derived from Jenner's vaccine, first used in 1796.

Recognition of HIV-AIDS

In 1981 a new disease was described in the male homosexual populations of New York, San Francisco, and Los Angeles. It destroyed the immune system, and the disease was called the acquired immune deficiency syndrome (AIDS) and the causal virus, human immunodeficiency virus (HIV), a retrovirus, was first isolated by L. Montagnier, of the Pasteur Institute in Paris, in 1983. A sexually transmitted disease, it has now spread all over the world, with about 40 million cases worldwide, and without very expensive chemotherapy is almost always fatal. It is now clear that it arose from an inapparent but persistent infection of chimpanzees in Africa. In spite of enormous efforts over the past 20 years, it has so far been impossible to devise an effective vaccine.

Arthropod-Borne Viruses of Vertebrates

Arthropods, mainly insects and ticks, were early shown to be important as vectors of virus diseases of vertebrates (the arboviruses) and of plants. Sometimes, as in myxomatosis, carriage was found to be mechanical, but in most cases the virus was found to multiply in the vector as well as the vertebrate concerned. The development of methods of growing insect cells in culture by T. D. C. Grace in Canberra in 1962 opened the way for the molecular biological investigation of the replication of insect viruses and arboviruses in invertebrate cells.

Taxonomy and Nomenclature

After earlier tentative efforts with particular groups of viruses, viral nomenclature took off in 1948, with the production of a system of latinized nomenclature for all viruses by the plant virologist F. O. Holmes. This stimulated others, in particular C. H. Andrewes of London and A. Lwoff of Paris, to actions which resulted in the setting up of an International Committee on Nomenclature (later Taxonomy) of Viruses in 1966. Adopting the kind of viral nucleic acid, the strategy of replication, and the morphology of the virion as its primary criteria, this committee has now achieved acceptance of its decisions by the great majority of virologists. One interesting feature of its activities is that its rules avoid the controversies about priorities that plague taxonomists working with fungi, plants, and animals.

The Future

The future will see an explosive expansion of understanding and knowledge of viruses of vertebrates, with the application of techniques of genetic engineering, nucleic acid sequencing, the polymerase chain reaction, and the use of monoclonal antibodies. All these discoveries are addressed in the appropriate sections of this encyclopedia. Molecular biology, which was initially conceived during studies of bacterial viruses, is now being used to study all the viruses of vertebrates and the pathogenesis and epidemiology of animal viral diseases. The expansion in knowledge of these subjects in the next decade can be expected to exceed that of the previous century.

See also: An Introduction to Plant Viruses. History of Virology: Bacteriophages

Further Reading

- Bos, L., 1995. The embryonic beginning of virology: Unbiased thinking and dogmatic stagnation. *Archives of Virology* 140, 613.
- Cairns, J., Stent, G.S., Watson, J.D. (Eds.), 1962. *Phage and the Origins of Molecular Biology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory of Quantitative Biology.
- Fenner, F., Gibbs, A.J. (Eds.), 1988. *Portraits of Viruses. A History of Virology*. Basel: Karger.
- Horzinek, M.C., 1995. The beginnings of animal virology in Germany. *Archives of Virology* 140, 1157.
- Hughes, S.S., 1977. *The Virus. A History of the Concept*. London: Heinemann Educational Books.
- Joklik, W.K., 1996. Famous institutions in virology. The Department of Microbiology, Australian National University and the Laboratory of Cell Biology, National Institute of Allergy and Infectious Diseases. *Archives of Virology* 141, 969.
- Matthews, R.E.F. (Ed.), 1983. *A Critical Appraisal of Viral Taxonomy*. Boca Raton, FL: CRC Press.
- Porterfield, J.S., 1995. Famous institutions in virology. The National Institute of Medical Research, Mill Hill. *Archives of Virology* 141, 969.
- van Helvoort, T., 1994. History of virus research in the twentieth century: The problem of conceptual continuity. *History of Science* 32, 185.
- Waterson, A.P., Wilkinson, L., 1978. *An Introduction to the History of Virology*. Cambridge: Cambridge University Press.

Human Boca- and Protoparvoviruses (*Parvoviridae*)

Maria Söderlund-Venermo, University of Helsinki, Helsinki, Finland

Jianming Qiu, University of Kansas Medical Center, Kansas City, KS, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AAV adeno-associated virus, in genus *Dependoparvovirus*
AGE acute gastroenteritis
ARTI acute respiratory tract infection
B19V human parvovirus B19, in *Erythroparvovirus*
BocaSR bocavirus-transcribed small RNA
BPV1 bovine parvovirus 1, in *Bocaparvovirus*
BuV bufavirus, in *Protoparvovirus*
CPV canine parvovirus, in *Protoparvovirus*
CSF cerebrospinal fluid
CuV cutavirus, in *Protoparvovirus*
CTCL cutaneous T-cell lymphoma
dsDNA double-stranded DNA
EIA enzyme immunoassay
HAE human airway epithelium
HAE-ALI human airway epithelium cultured at an air-liquid interface
HBoV human bocavirus, in *Bocaparvovirus*
HEK293 human embryonic kidney 293 cells
IgG immunoglobulin G
IgM immunoglobulin M
kDa kilo Dalton
LEH left-end hairpin (of the genome)
MVC minute virus of canines, in *Bocaparvovirus*

MVM minute virus of mice, in *Protoparvovirus*
NCR noncoding region (of genome)
NP1 nuclear phosphoprotein 1
NPS nasopharyngeal sample
NS1–4 non-structural protein 1–4
nt nucleotide
OAS original antigenic sin (an immunological phenomenon)
ORF open reading frame
(pA)d distal polyadenylation site
(pA)p proximal polyadenylation site
PCR polymerase chain reaction
PLA₂ phospholipase A₂
Pre-mRNA precursor mRNA
qPCR quantitative PCR
REH right-end hairpin (of genome)
RT-PCR reverse transcription PCR
ssDNA single-stranded DNA
TuV tusavirus, in *Protoparvovirus*
VP1–3 viral capsid proteins 1–3, with long common C-termini
VP1u a short unique N-terminus of VP1, which is not shared with VP2 or 3

Glossary

Cutaneous T-cell lymphoma (CTCL) A rare form of cancer affecting mature T cells in skin, forming skin lesions. CTCL is a class of non-Hodgkin lymphoma with unknown etiology, and is divided into several subtypes, including mycosis fungoides and Sézary syndrome.

HAE-ALI Human airway epithelium cultured at an air-liquid interface: Human bronchial/tracheal epithelial cells are cultured on microporous membrane inserts in wells. After 3–4 weeks at an air-liquid interface, cells are polarized into apical side that is exposed to the air and basolateral side that is dipped into liquid media.

IgG avidity The avidity, or binding force, of IgG to its antigen increases with the slowly maturing immune response after acute primary infection and can be measured by a protein-denaturing EIA disrupting the weak antigen-antibody bond in acute infection but not the stronger bond in past immunity.

Original antigenic sin (OAS) An immunological phenomenon (long-known from influenza and dengue virus infections) observed when a prior virus infection of an individual fully or partly hinders the induction of antibodies to the currently infecting related virus, and instead the original virus-specific IgG response is enhanced.

Introduction

Novel parvovirus sequences have been obtained after filtration, DNase treatment, and random PCR, followed by large-scale sequencing and bioinformatics. The first human bocavirus, HBoV1, was discovered in 2005 in airway samples of children with acute respiratory tract infection (ARTI). HBoV1 causes mild to life-threatening ARTI in young children worldwide. It is thus the second human-pathogenic parvovirus after parvovirus B19 (B19V). Three other closely related human bocaviruses, HBoV2, 3, and 4, have since been discovered in stool, but their association with acute gastroenteritis (AGE) is disputed. More recently, in 2012–2016, three novel human protoparvoviruses were discovered in stool samples of children with diarrhea and named bufavirus (BuV), tusavirus (TuV) and cutavirus (CuV). BuV has been associated with gastroenteritis and CuV with cutaneous T-cell lymphoma (CTCL), while more studies are needed to reveal if TuV really is a genuine human virus.

Classification

Parvoviruses are small non-enveloped icosahedral ($T = 1$) viruses with linear single-stranded (ss) DNA genomes forming double-stranded (ds) hairpin structures at both ends. The family *Parvoviridae* includes two subfamilies, *Parvovirinae* with 8 genera comprising viruses infecting vertebrates and *Densovirinae* with 5 genera comprising viruses infecting invertebrates, but with a third subfamily, *Hamaparvovirinae*, proposed with viruses of both vertebrates and invertebrates. Human parvoviruses are found in 5 different genera, *Erythroparvovirus*, *Dependoparvovirus*, *Tetraparvovirus*, *Bocaparvovirus* and *Protoparvovirus*, with human bocaviruses (HBoVs) and bufavirus (BuV), tusavirus (TuV), and cutavirus (CuV), belonging to the two latter genera, respectively. Virus species within the genera are monophyletic and their NS1 protein sequences (with some flexibility) are $> 30\%$ identical to each other but $< 30\%$ identical to those of other genera, which are general taxonomic criteria for parvoviruses. Characteristic for members of the *Bocaparvovirus* genus is a unique ~ 25 kDa nuclear protein, NP1, encoded in the middle of the genome. For both boca- and protoparvoviruses, the encapsidated linear ssDNA genomes are heterotetrameric and $\geq 90\%$ of genomes are of negative sense. The founder viruses of *Bocaparvovirus* are bovine parvovirus 1 (BPV1) and minute virus of canines (MVC), from which the genus name was created. *Protoparvovirus* is an older, but recently renamed, well characterized genus, including one of the first known parvoviruses, Kilham rat virus, discovered already in 1959. New viral members are being discovered at a rapid pace, but at the time of writing this review, there are 21 species in *Bocaparvovirus* and 11 in *Protoparvovirus*, with viruses from two species each infecting humans. HBoV1, HBoV3, a gorilla and a chimpanzee bocavirus belong, according to the NS1 amino-acid sequence identities, to the species *Primate bocaparvovirus 1*, and HBoV2 and HBoV4 belong to the species *Primate bocaparvovirus 2*. Similarly, all three BuV genotypes belong to *Primate protoparvovirus 1* and CuV to *Primate protoparvovirus 3*, while a rhesus monkey bufavirus founded the *Primate protoparvovirus 2* species. TuV has been proposed to found the species *Primate protoparvovirus 4*. The current status of parvovirus taxonomy and phylogenetic trees, can be seen at the “Relevant Websites section”.

Virion Structure

Parvoviruses are small (~ 25 nm) non-enveloped viruses with icosahedral $T = 1$ capsid symmetry. The capsid comprises thus 60 copies of the common C-terminal part of the capsid proteins (VPs), the vast majority of which is provided by the shortest VP (VP3 for HBoVs and VP2 for protoparvoviruses), and the minority (5%–10%) by the longer VP1 (and VP2 for HBoV) proteins. Typical to most parvoviruses, the short unique N-terminus of VP1, VP1u, carries phospholipase (PLA_2) activity, which is important in endosome escape. The capsid has multiple roles in parvovirus biology, including host cell recognition by binding to cellular receptors, intra-cellular trafficking by the activity of its PLA_2 motif, nuclear entry by its signaling sequences, genome packaging through its 5-fold cylinder, and host immune recognition of its neutralizing epitopes.

The HBoV major capsid protein is VP3 of ~ 60 kDa, and all three VPs seem to be translated from the same transcript, in a ratio of 1:1:10. However, for use as antigens or for structural studies, virus-like capsids (VLP) can be assembled based on VP3 alone. The major capsid protein in BuV, TuV and CuV is VP2 of the same size. The structures of HBoV1, HBoV3, HBoV4 and BuV1–3 VLPs, obtained by the baculovirus expression system, have been determined to high (~ 3 Å) resolution, by cryo-electron microscopy and 3D image reconstruction. The HBoV and BuV capsid cores are conserved and formed by an eight-stranded antiparallel β -barrel motif and an α -helix, similar to all other parvoviruses. The specific surface of the capsid is shaped by long amino acid loops between these secondary structures, containing ten variable regions (VR) of which VR-III and VR-IX have been identified as potential tropism determinants. Seemingly unique for the bocaparvoviruses, there is another α -helix within the surface loops. The HBoV and BuV capsids follow the typical parvoviral topology, including an open channel at the icosahedral 5-fold axes with a surrounding canyon, a depression at the 2-fold axes, and trimeric protrusions at the 3-fold axes. The HBoV 3-fold protrusions are less prominent than those of AAVs but more pronounced than those of parvovirus B19, and among the HBoVs, HBoV1 has the least prominent topology. Surprisingly, these trimeric 3-fold protrusions in BuV are separated, and therefore more similar to those of HBoV and B19V than those of animal protoparvoviruses, perhaps indicating a host-specific structural evolution. The conserved 5-fold channel is made by five β -strand ribbons forming a cylinder, through which DNA packaging is thought to occur. Unique for bocaparvoviruses is the presence of a density beneath the 5-fold axis that extends into the capsid interior. In both HBoV and BuV capsids, the location of the disordered VP1u and N-terminus of VP2 have not been structurally confirmed, but recent observations support the hypothesis that the N-termini would lie inside the capsid, from where a flexible VP1u could, when associated with a cell, be externalized through the 5-fold channel exposing its PLA_2 moiety. Four HBoV-specific antigenic epitopes have been mapped in the capsid, three being specific for HBoV1 at the 3-fold protrusions and the wall between the 2- and 5-fold axes, and a fourth HBoV1–4 cross-reacting epitope at the 5-fold axis. Structural analyses of TuV and CuV are ongoing.

Genome

All parvovirus virions have linear ssDNA genomes with dsDNA hairpin ends. Among viruses in the genus *Bocaparvovirus*, only the genomes of HBoV1, BPV1 and MVC have been sequenced in full length, including their terminal hairpins. The HBoV1 genome is 5543 nucleotides (nt) in length with nonidentical hairpin structures at the termini (Fig. 1(A)). In bocavirus virions, approximately over 90% of the viral genome are of negative sense. The HBoV1 left-end hairpin (LEH) is 140nt in length and is predicted to be ‘Y’ shaped with short axial “rabbit ears” and mismatches that cause an unpaired “bubble”. The LEHs of both MVC and BPV1 have

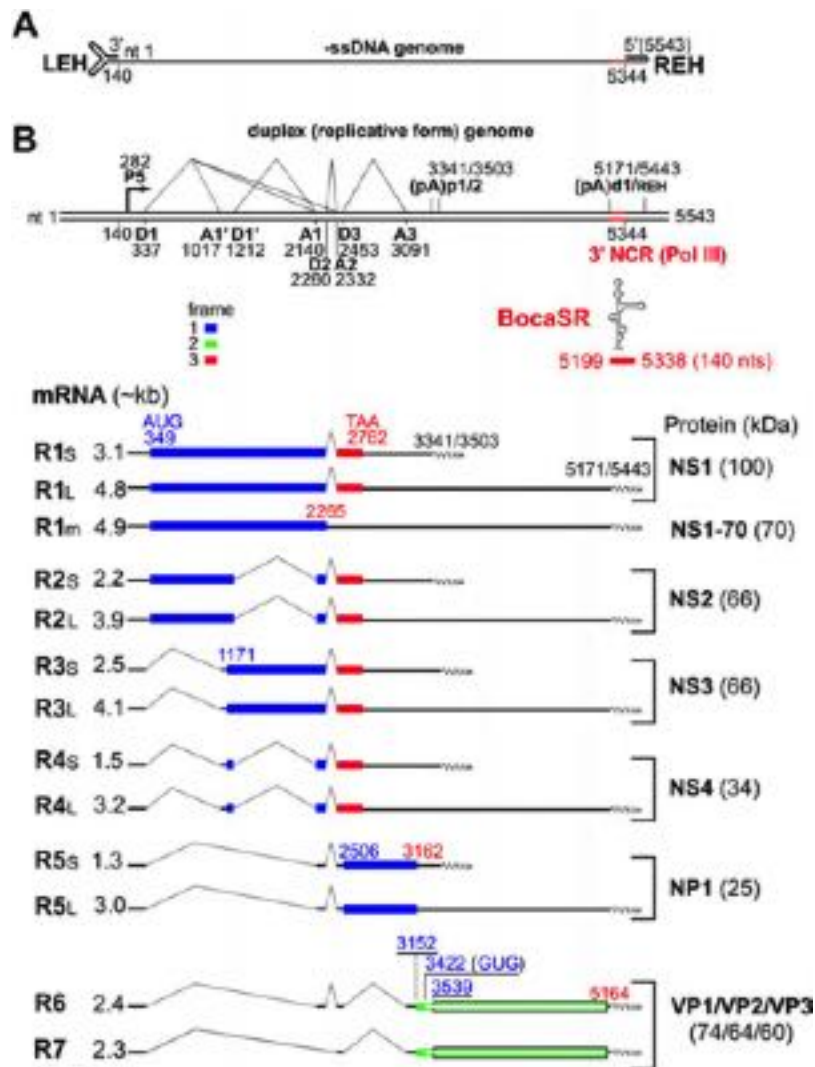


Fig. 1 Gene expression profile of HBoV1. The ssDNA genome of HBoV1 is depicted to scale in negative sense with transcriptional and RNA processing units: the P5 promoter, 5' splice-donor sites (D1, D1', D2, and D3) and 3' splice-acceptor sites (A1, A1', A2, and A3), the proximal and distal polyadenylation sites ((pA)p1/2 and (pA)d1/REH, respectively), and the 3' noncoding region (NCR). The left-end hairpin (LEH) and right-end hairpin (REH) structures of the genome are shown. Seven groups of HBoV1 mRNA transcripts are detected during infection, with either long-form mRNA (R_{XL}) that reads through the (pA)p site or short-form mRNA (R_{XS}) that is polyadenylated at the (pA)p site. R1 mRNA has a minor species (R1m) that is unspliced at the central small intron (D3–A3). Major open reading frames (ORFs) are depicted in colored boxes with the nucleotide number (nt) and amino acid (aa) of the start and stop codons indicated. Proteins expressed from each mRNA are indicated beside their respective mRNAs with molecular weight in kDa. *BocaSR* is diagramed, and is transcribed from the 3' NCR of nt 5199–5388 that bears an intergenic RNA polymerase III promoter (Pol III).

been isolated in two configurations, called flip and flop, suggesting that also the HBoV1 LEH should exist in both configurations. The right-end hairpin (REH) of MVC and BPV1, but not HBoV1, harbors sequences that have the potential to be folded into a cruciform structure near the end tip, although it is thermodynamically less favorable for BPV1. Notably, HBoV1 REH shares an identical 21-nt sequence with the MVC REH at the end, and the HBoV1 LEH shares two identical sequences of 6nt and 8nt, respectively, with the BPV LEH. Both the HBoV1 and MVC REHs are perfect duplex (palindromic) structures without any nucleotide mismatches and, therefore, only one configuration is present.

The intracellular duplex or replicative form of the HBoV1 genome is transcriptionally capable (Fig. 1(B)). The left half of the duplex genome encodes nonstructural, NS, proteins, and the right half encodes capsid, VP, proteins. The HBoV1 genome possesses two unique features among parvoviruses: (1) an additional NP1 protein is encoded from an open reading frame (ORF) lying in the middle of the genome, and (2) a viral noncoding RNA – bocavirus-transcribed small RNA (*BocaSR*), from the noncoding region (NCR) at the right-hand end of the duplex genome, after the VP ORF but before the REH. Parvoviruses utilize cellular DNA and RNA polymerases for replication and transcription, respectively. HBoV1 has only one (RNA polymerase II) promoter, P5, which transcribes a single precursor mRNA (pre-mRNA). The viral pre-mRNA is processed through alternative polyadenylation and

alternative splicing in the nucleus, which generates 7 groups of viral mRNA transcripts during infection (Fig. 1(B)). Notably, there is also an intergenic RNA polymerase III promoter at nt 5199–5338 that transcribes the 140-nt long *BocaSR* during infection.

While NS1 is essential to viral DNA replication, both NP1 and *BocaSR* enhance viral DNA replication. In *ex vivo* human airway epithelium cultured at an air-liquid interface (HAE-ALI), also NS2 is essential to virus replication, whereas NS3 and NS4 are dispensable. NP1 also plays an important role in VP expression by overcoming a blockage of the internal polyadenylation in the middle of the genome, and is essential to the splicing of viral pre-mRNA at the A3 acceptor site that lies in front of the (pA)p sites (Fig. 1(B)). *BocaSR* not only plays a role in viral DNA replication but also in the expression of NS1, NS2, NS3, and NP1 (but not NS4). The VP proteins constitute the capsid and allow packaging of the viral genome. Notably, VP1–3 are likely translated from a single mRNA, R6 or R7, and VP2 utilizes a noncanonical translation initiation site, GUG, at nt 3422 (Fig. 1(B)).

The human protoparvoviruses, BuV, TuV and CuV, have, similar to HBoVs, the same general genome organization as those of the other parvoviruses, with NS1 genes on the left half and VP genes on the right half of the negative-sense genome. In addition, BuV and CuV seem to have a small middle ORF, with unknown function. Contrary to HBoV and B19V, protoparvoviruses have two different promoters, one in the 3' end and the other in the middle of the genome, but only one termination site at the far-right end. The exact transcription maps for BuV, TuV and CuV have not been determined, however, they are most likely similar to those of the other better known animal protoparvoviruses.

Life Cycle

HBoV1 is detected in nasopharyngeal samples (NPS) of patients with acute respiratory tract infections (ARTI), suggesting that HBoV1 naturally infects human airways. Productive HBoV1 infection has been observed only in HAE-ALI. Although HBoV1 does not productively infect monolayer-cultured (proliferating) human cells, including human embryonic kidney 293 (HEK293) cells, HEK293 cells fully support replication of the HBoV1 duplex genome – delivered by transfection of a full-length plasmid clone of HBoV1 (pHBoV1) – and produces progeny virions that are fully infectious in HAE-ALI, but not in monolayer-cultured airway epithelial cell lines, such as A549 and BEAS-2B cells. HAE-ALI cultures generated from primary or repopulating primary human bronchial/tracheal epithelial cells, as well as immortalized human airway epithelial cells (CuFi-7), support productive infection of HBoV1. Recently, it has been reported that Caco-2 cells, derived from human colon cancer, can be infected with HBoV1, but only after polarization/differentiation.

HBoV1 can infect HAE-ALI cultures through both the apical and the basolateral sides, suggesting that the (unknown) virus receptor is expressed on both sides of the airway epithelia (Fig. 2, steps 1–2). As with other parvoviruses (e.g., B19V, CPV, AAV2, and MVM), HBoV1 likely enters the cells through receptor-mediated endocytosis, and intracellularly traffics from early endosome to late endosome (Fig. 2, steps 3). Virions may directly disrupt the nuclear envelope or be transported through the nuclear pore to translocate into the nucleus (Fig. 2, step 4). Within the nucleus, the viral genome is released and likely recognized by the cellular repair machinery (Fig. 2, step 6). Then, the ssDNA viral genome is converted into a duplex form (Fig. 2, steps 7a and 7b). In polarized HAE cells that are mitotically terminated (differentiated), cellular DNA repair DNA polymerases η/κ are involved in HBoV1 DNA replication. This has not been shown for other parvoviruses. It is thus believed that the Y-family polymerases η/κ synthesize the complementary strand, utilizing the 3'-OH hairpin end as primer. After expression from the duplex genome, NS and NP1 proteins are associated with the viral genome for viral DNA replication (Fig. 2, steps 12–16). NS or NP1 proteins are likely required for genome packaging as well (Fig. 2, steps 16–18); however, they are not packaged into the mature virions themselves. The viral DNA replication process involves complex hairpin structure transfers and protein-genome interactions, which produces viral DNA intermediates in various lengths and structures (Fig. 2, steps 12–16). Viral pre-mRNA is transcribed from the duplex form genome and further processed to generate NS-, NP-, and VP-coding mRNAs (Fig. 2, step 8). Capsid proteins VP1, VP2, and VP3, produced in the cytoplasm, are assembled into oligomers before translocating into the nucleus (Fig. 2, steps 10 and 11). Empty capsids are assembled in the nucleus after which viral ssDNA is encapsidated (Fig. 2, steps 17–18). Eventually, the mature virions are released from the nucleus into the cytoplasm and transported out of the infected cells (Fig. 2, steps 19–20), which undergo pyroptotic cell death. BuV, TuV and CuV have not yet been propagated in cell culture.

Epidemiology

Human Bocaviruses

HBoV1, being a respiratory virus, is most likely transmitted by the respiratory route, whereas HBoV2 and 3 are considered enteric and are most likely transmitted by the fecal-oral route. Findings supporting this are mainly that HBoV1 DNA is detected by quantitative (q)PCR in high amounts in airway samples during the acute phase of ARTI, whereas HBoV2–4 seldom are found in such samples but are detected in stool. HBoV1 is most common in children between 0.5 and 5 years of age, with a mean age of infection of 2 years. HBoV1 DNA is detected in 2%–20% of NPS of children with upper or lower ARTI, but due to its persistence in the airways for up to a year, it may be detected in lower DNA loads also in healthy children. The most common HBoV in stool samples is HBoV2 that has been found in 1%–20% of stools of both symptomatic and non-symptomatic subjects, followed by HBoV1, HBoV3 and HBoV4. None of these have, however, indisputably been shown to cause gastroenteritis. Further, HBoV4 is too

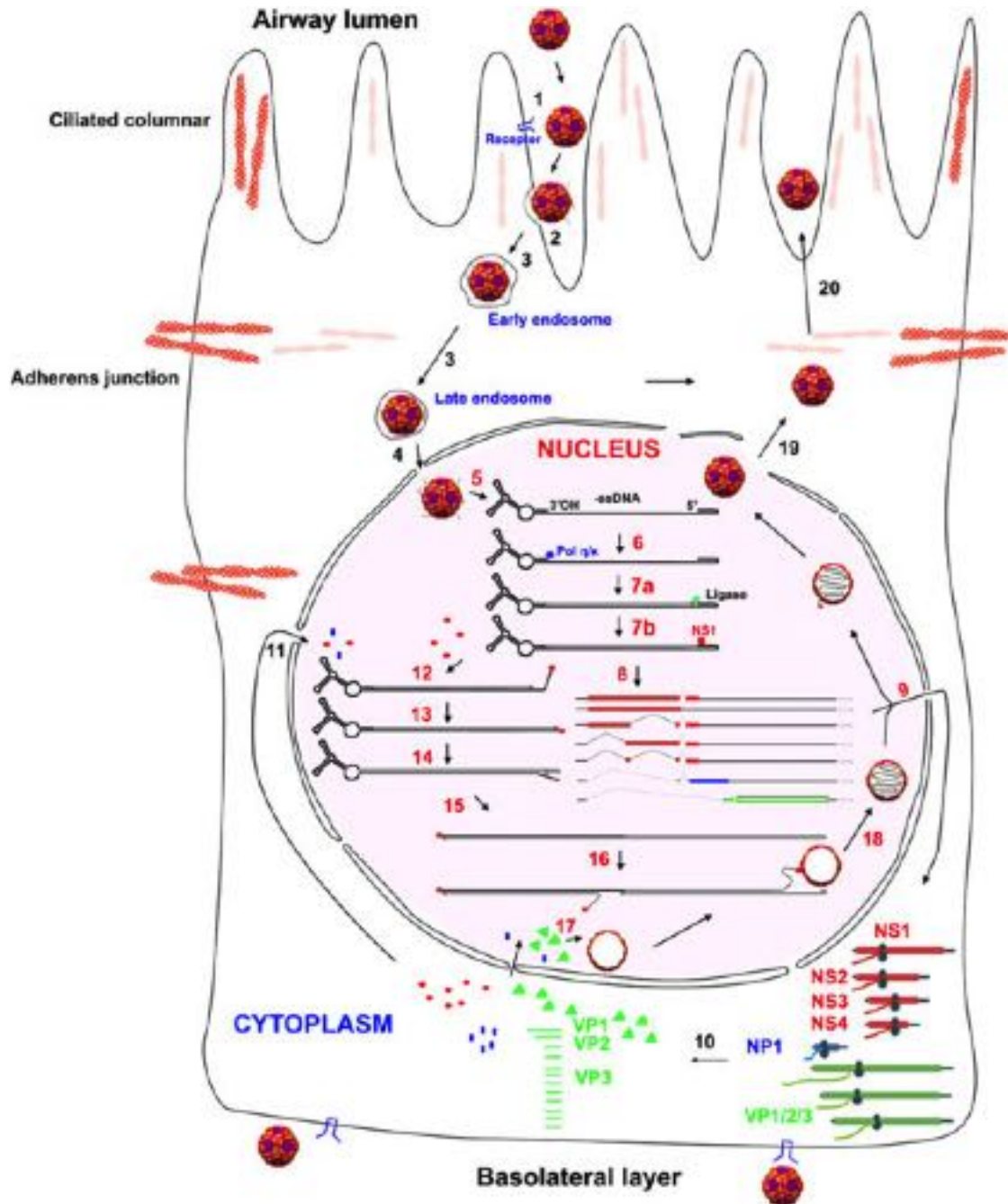


Fig. 2 The infection life cycle of HBoV1. HBoV1 infection of human airway epithelium cultured at an air–liquid interface (HAE-ALI) is illustrated with a polarized airway epithelial cell featured with the apical side towards the airway lumen, basolateral layers, columnar cilia, and adherent junction molecules. Each step of the virus life cycle is explained in the text. Ligase, an unknown cellular DNA ligase; -ssDNA, negative single-stranded DNA; pol η/κ , DNA repair DNA polymerase η and κ .

rare to state anything of its transmission, pathogenicity or tissue tropism. Contrary to B19V, none of the HBoVs seem to be transmitted by blood or blood products, even though a short viremic phase does occur during acute infection and HBoV1 DNA is infrequently detected in donated blood. Besides serum, respiratory and enteric samples, HBoV DNAs have been detected also in saliva, urine, cerebrospinal fluid (CSF), tonsils, and intestinal biopsies, as well as in environmental river water and in sewage. HBoV1 DNA is globally detected throughout the year with no obvious seasonality observed, however, the long persistence of HBoV1 DNA in the airways makes such studies problematic.

HBoVs cause systemic infections, likely leading to immunity. Most HBoV1-specific antibody responses are strong and life-long, but ~30% are weaker and some may be totally nonexistent. Further, the IgG responses against the enteric HBoVs are generally

weaker and even more prone to waning than those against HBoV1. This variation is due to both cross reactivity between the HBoVs and to an immunological phenomenon called “original antigenic sin” (OAS). If a person has been infected by HBoV2 earlier in life, his/her current HBoV1 infection might not induce HBoV1-specific IgG responses, leading to a false negative IgG result and a lower seroprevalence. Despite these challenges, the seroprevalence of HBoV1–4 in Finland already at the age of 6 have been determined to be 80%, 50%, 10% and ~0%, and in China among adults to 67%, 50%, 40% and 1.4%, respectively, indicating that HBoV1 is the most common human bocavirus circulating worldwide. The median age of infection is 2 years for HBoV1 and slightly lower for the enteric HBoVs, while in adults and the elderly, HBoV infections are much less frequent.

Human Protoparvoviruses

BuV, TuV, and CuV were all discovered by metagenomics in stools from children with diarrhea. The BuV DNA prevalence is between 0% and 4% in stools of diarrheic patients from Africa, Europe, Asia, and South America, and viral DNA can be found also in river water. In a study from Finland, also close to 1000 nasal swabs were studied for BuV DNA, only one being positive, indicating that BuV is not a respiratory virus. All BuV sequences typed to-date have been of genotype 1 or 3, while genotype 2 DNA has been identified only in the stool of a single child from Burkina Faso, described in the original discovery report. Likewise, TuV DNA has also been detected in a single pediatric stool sample, obtained from Tunisia, whereas the CuV genoprevalence was 1%–2% among stools from Brazil and Botswana.

The three BuV genotypes, sharing only 65%–73% VP2 amino-acid identity, can be separately identified in EIAs and are thus also considered different serotypes. The global BuV seroprevalence pattern is strikingly different, varying among adults from 2% to 4% in Finland and the USA to 56%–85% in Kenya and the Middle East. The predominant genotype in Kenya was BuV-3 and that in the Middle East BuV-1, whereas BuV-2 was the second-most prevalent genotype on both continents. This finding contrasts the PCR findings, where BuV-2 DNA has been found in only one sample. TuV IgG was not found in the sera from any of the four continents, while CuV IgG was evenly distributed with a seroprevalence of 0%–6%. BuV-2 and CuV VP2s are more similar to each other than those of BuV-1 and -3, and they therefore need blocking by a competition EIA not to show considerable cross reactivity of IgG.

Clinical Features

Human Bocaviruses

HBoV1 is one of the most common respiratory viruses detected in NPS of children with ARTI. However, HBoV1 DNA is frequently (20%–80%) detected together with other respiratory pathogens and it is also detected in NPS of healthy children. In the early studies, researchers were therefore questioning the clinical significance of HBoV1 in ARTI and many considered it a bystander virus. Later, this high prevalence of HBoV1 DNA was shown to be due to virus persistence in the nasopharynx for several weeks to up to a year after acute infection, during which time the child may have several other respiratory tract infections. Consequently, the mere presence of this virus in the airways, determined by sensitive qualitative PCRs, cannot be a reliable marker for a HBoV1 etiology in the current ARTI. Comprehensive clinical studies, applying monodetection and/or more accurate diagnostic markers, including serology, qPCR, reverse transcription (RT)-PCR and antigen detection, have since provided convincing evidence of HBoV1 being a true respiratory pathogen in children.

HBoV1 may cause both upper or lower ARTI, most often in children of 0.5–5 years of age, while smaller infants are mostly protected by passive maternal antibodies. The most frequent clinical features are those associated with the common cold, bronchiolitis, pneumonia, and exacerbations of asthma. The symptoms do not differ from those of other viral ARTIs; cough, fever, rhinitis, and wheezing. In addition, acute otitis media and diarrhea are often reported during HBoV1 infection and viral DNA has been detected in middle-ear fluids and stool. Typical signs of pneumonia, with patchy or interstitial infiltrates in chest radiographs, hyperinflation and atelectasis, as well as hypoxia and tachypnea are frequently seen in HBoV1-caused lower ARTI. HBoV1 has further been the sole finding also in several life-threatening infections with respiratory failure demanding intensive care with mechanical ventilation.

In adults, HBoV1 DNA detections in airway samples are infrequent, possibly due to the high seroprevalence, but usually associated with respiratory symptoms. Summarizing 5 major studies in China with > 12,000 adult ARTI patients, the prevalence of HBoV1 DNA in airway samples was 0.3%. In one study on Finnish elderly adults (> 65 years of age, with a mean age of 83 years), no acute HBoV1–4 infections were recorded among 438 episodes of ARTI, as diagnosed by serology and only 2 (0.5%) had HBoV1 DNA in their respiratory samples. Among these patients, 88% were HBoV1 IgG seropositive. A striking difference between these elderly adults and children was the low prevalence of HBoV1 (0%) and high prevalence of RSV (7%) genome detections in the samples of 283 asymptomatic elderly controls.

The four HBoVs have also been found in up to 25% of diarrheic stool samples, but neither their DNA loads nor prevalence suggests a causal association with AGE. In most studies equally often HBoV DNA is detected in stools of symptomatic as of asymptomatic children. A prolonged excretion of HBoV1 and 2 in stool has furthermore been reported, making disease associations difficult. Nevertheless, one case is reported of a life-threatening hypovolemic shock due to diarrhea associated with a disseminated HBoV1 infection in an immunocompromised 9-year-old boy, with no other pathogens detected.

HBoV1, HBoV2, and HBoV3 have infrequently been detected, as sole microbes, in CSF of patients with severe encephalitis, indicating a possible role in infections of the central nervous system. In a HBoV2 DNA-positive CSF sample also typical parvoviral virions were detected by electron microscopy; the child was further seronegative but viremic. The severity of symptoms in the HBoV-positive patients did not differ from those of patients with other virus findings in CSF (dengue-, echo- or adenovirus). In addition, there are case reports of HBoV-associated myocarditis and hepatitis. A 13-month-old girl had a fatal subacute myocarditis with HBoV2 DNA detected in NPS, plasma, urine, ascites, mesenteric node, skeletal muscle, and lung, and she died of heart failure. A previously healthy 2.5-year-old girl had bronchiolitis and hepatitis with high-load HBoV1 viremia as sole finding, and another 7-month-old boy with an underlying T-cell deficiency, had hepatitis with HBoV1 DNA in consecutive NPA, feces, urine, and serum samples and with diagnostic increases in both HBoV1 IgM and IgG. These cases suggest that the full picture of the clinical impact of HBoVs is still to be revealed.

Human Protoparvoviruses

Very little is known about the clinical pictures of BuV, TuV, and CuV infections. BuV has been shown to associate with diarrhea, since its DNA has been detected by PCR only in diarrheic stools, while non-diarrheic stools all have so far been PCR negative. However, the frequencies as well as viral loads have been low, suggesting a nonsignificant role of BuV in gastroenteritis. No further TuV and CuV PCR studies of stool samples have been reported, leaving their relationship with gastroenteritis open. CuV DNA has, however, in three separate studies been detected in skin lesions of cutaneous T-cell lymphoma patients, but not in skin biopsies of healthy subjects, reaching statistical significance. It cannot be proven at this time, whether or not this association is causal or consequential. Parvoviruses are usually not directly oncogenic, but they may cause chronic inflammation possibly leading to cell transformation. CuV DNA has infrequently been detected also in melanoma lesions and in both healthy and malignant skin biopsies of immunocompromised transplant patients.

Pathogenesis

As animal models of HBoV1 infection are currently not available, the respiratory pathogenesis of HBoV1 is not well understood and thus is limited to HBoV1 infection in the ex vivo HAE-ALI culture model. HBoV1 infection of polarized HAE-ALI cultures is persistent, and can last 45 days with release of progeny virions, or as long as the ALI cultures can be maintained. The infection induces progressive airway epithelial damage with disruption of the tight junction barrier, thinning of the epithelium, loss of cilia, and epithelial cell hypertrophy indicated by enlarged nuclei. HBoV1 infection induces pyroptotic cell death of airway epithelial cells, represented by activation of caspase 1 but not of caspase 3, as well as an increase in the expression of cytokines IL-1 and IL-18. The pyroptosis induced by HBoV1 infection is mediated by the NLRP3 (NLR family pyrin domain containing 3 protein)-ASC (apoptosis-associated speck-like protein containing a CARD)-caspase-1 inflammasome-induced pathway.

Mechanistically, HBoV1 infection upregulates the expression of the anti-apoptotic genes *BIRC5* (Survivin, also called baculoviral inhibitor of apoptosis repeat-containing 5) and *IFI6* (interferon alpha inducible protein 6), which inhibit HAE cell apoptosis, and therefore the infected cells die via pyroptosis. Notably, pyroptosis does not inhibit HBoV1 replication. HBoV1 infection-induced chronic inflammation may serve to promote its persistent infection. HBoV1-infected patients are mostly young children, still with maturing immune systems. If HBoV1-infected cells are not quickly eliminated by effective immune responses, the intensifying inflammation cycle could possibly even lead to lung tissue injury in HBoV1-infected young children.

Pyroptosis is an inflammasome-mediated cell death with release of active IL-1, IL-18, and gasdermin D (GSDMD), which cause the cell to swell and eventually lyse. Therefore, HBoV1 infection-caused cell death is not limited to the infected cells but also spreads to the entire epithelia cultured on the insert, including uninfected cells of the inserts. It is currently unknown which viral components are sensed as pathogen-associated molecular patterns (PAMPs) by any pattern recognition receptors (PRRs) or which damage-associated molecular patterns (DAMPs) are generated into the cytosol during infection. Further studies are warranted to examine the nature of cell death in clinical specimens such as cells in airway washes from patients with monoinfection of HBoV1.

Diagnosis

Diagnosis of HBoV1 ARTI is challenging because of viral persistence in the airways long after infection (Fig. 3). Qualitative PCR of NPS, the predominantly used diagnostics for ARTI, is therefore insufficient. For an accurate diagnosis of acute HBoV1 infection, a combination of many different approaches is needed. One such approach is quantitative PCR, which gives a better estimate of how recently the infection began; a high viral load (typically over 10^4 or 10^6 copies/ml) would therefore suggest an acute HBoV1 infection. However, as verified by serology, about a third of patients with low viral load and even a few symptomatic patients with HBoV1 PCR-negative NPS, still had an acute HBoV1 infection and conversely, some patients with high loads of HBoV1 DNA in NPA, yet do not have acute infection. Obviously, if no other pathogens are detected, it strengthens the probability of an HBoV1 etiology, but again, ~60% of patients with other pathogens co-detected, experience also true acute HBoV1 infections. Another useful diagnostic method for respiratory samples is reverse transcription (RT)-PCR for detection of viral mRNA. Viral transcripts are

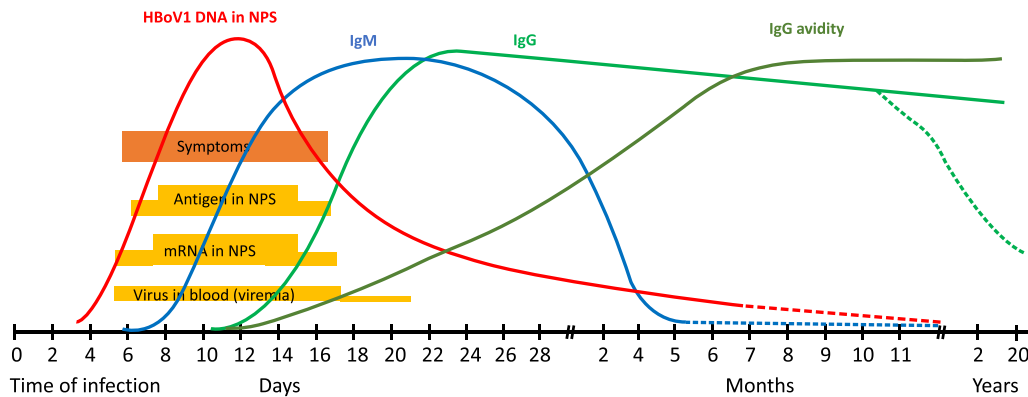


Fig. 3 Clinical course of HBoV1 infection. Appearance of respiratory symptoms, of HBoV1 DNA, mRNA and antigen in the nasopharynx, of HBoV1-specific DNA, IgM and IgG in blood, as well as increase of IgG avidity, after acute primary HBoV1 infection. The exact time of onset of symptoms related to time of infection (i.e., incubation time) is unknown, but has here been estimated based on the antibody kinetics and proportion of the other diagnostic markers in wheezing children with and without seroconversion. NPS, nasopharyngeal samples; IgM, immunoglobulin M; IgG, immunoglobulin G.

expressed mostly in the acute phase of HBoV1 infection and can thus be considered an indirect marker of virus replication and ongoing infection. By a novel commercial antigen assay, also viral capsid protein can be detected in NPS, which would indicate an acute phase of infection. The specificity is good but the sensitivity is lower than for the PCR-based assays.

When an accurate diagnosis of HBoV1 respiratory tract infection is required, also serum (or plasma) is an important sample type. HBoV1 causes a systemic infection resulting in a short-phase viremia, and can therefore, during the very acute phase, be found by PCR in blood or serum. HBoV1 DNA has, however, been shown to persist longer in some individuals, slightly decreasing the clinical specificity. The most compelling evidence perhaps for an acute HBoV1 infection is, however, the detection of HBoV1-specific IgM together with a seroconversion or a $\geq$ four-fold increase of IgG in paired serum samples, taken at the acute phase and at the convalescent phase two weeks later (Fig. 3). A low IgG avidity (even in a single IgG-positive sample) further verifies the acute HBoV1 infection, whereas a high IgG avidity indicates past immunity from a prior infection a few months to several years ago. The IgG-avidity EIA thus further increases the diagnostic value of serology. Nevertheless, two issues hamper the reliability of serology. First, the four HBoVs induce cross-reacting antibodies to common epitopes, showing reactivity in all four HBoV EIAs, despite only one of them is specific reactivity. Such cross-reactivity can, however, be avoided by blocking the HBoV1–4 common antibodies from binding by adding the heterotypic VLP antigens to the serum prior to the EIA, leaving only the antibodies to specific HBoV1 epitopes for measurement. Secondly, the four related HBoVs may occasionally interact immunologically resulting in OAS, a phenomenon described above (Epidemiology), perhaps leading to a slightly decreased negative predictive value. However, this OAS effect seems to occur more often in past immunity than in recent infection.

Taken together, due to virus persistence in the airways, the detection of HBoV1 DNA by a qualitative PCR in NPS is insufficient. The most accurate diagnosis of acute HBoV1 respiratory tract infection is obtained by combining different diagnostic approaches; HBoV1-DNA monodetection, high viral-DNA load, and detection of viral mRNA or antigen in respiratory specimens, viral DNA detection in serum, and detection of HBoV1-specific IgM coupled with seroconversion or ≥ 4 -fold increase in paired sera of HBoV1-specific IgG antibodies with low avidity.

For the human protoparvoviruses, only qPCR and IgG EIA methodology so far exist, but development of IgM EIAs is ongoing for future clinical studies.

Prevention

Since HBoV1 is transmitted by close contact, droplets or aerosols, and HBoV2–4, BuV and CuV possibly by the fecal-oral route, standard hygienic practices like hand washing with soap may reduce spreading the virus. However, alcohol-derived gels are less effective against non-enveloped parvoviruses. Avoiding large crowds in closed areas during the cold seasons, may also diminish the risk of getting respiratory tract infections. Currently there are no vaccines or specific medication available for human parvovirus infections.

Further Reading

- Allander, T., Tammi, M.T., Eriksson, M., *et al.*, 2005. Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proceedings of the National Academy of Sciences of the United States of America* 102, 12891–12896.
- Christensen, A., Kesti, O., Elenius, V., *et al.*, 2019. Human bocaviruses and paediatric infections. *Lancet Child and Adolescent Health* 3, 418–426. doi:10.1016/S2352-4642(19)30057-4.
- Cotmore, S.F., Agbandje-McKenna, M., Canuti, M., *et al.*, 2019. ICTV virus taxonomy profiles: *Parvoviridae*. *Journal of General Virology* 100, 367–368. doi:10.1099/jgv.0.001212.

- Deng, X., Yan, Z., Cheng, F., Engelhardt, J.F., Qiu, J., 2016. Replication of an autonomous human parvovirus in non-dividing human airway epithelium is facilitated through the DNA damage and repair pathways. *PLoS Pathogens* 12, e1005399.
- Deng, X., Zou, W., Xiong, M., *et al.*, 2017. Human parvovirus infection of human airway epithelia induces pyroptotic cell death by inhibiting apoptosis. *Journal of Virology* 91, pii: e01533-17.
- Huang, Q., Deng, X., Yan, Z., *et al.*, 2012. Establishment of a reverse genetics system for studying human bocavirus in human airway epithelia. *PLoS Pathogens* 8, e1002899.
- Jartti, T., Hedman, K., Jartti, L., Ruuskanen, O., Söderlund-Venermo, M., 2012. Human bocavirus – The first 5 years. *Reviews of Medical Virology* 22, 46–64.
- Mietzsch, M., Kailasan, S., Garrison, J., *et al.*, 2017. Structural insights into human bocaproviruses. *Journal of Virology* 12, 91.
- Ong, S.Y., Schuurman, R., Heikens, E., 2016. Human bocavirus in stool: A true pathogen or an innocent bystander? *Journal of Clinical Virology* 74, 45–49.
- Schildgen, O., Qiu, J., Söderlund-Venermo, M., 2012. Genomic analysis of the human bocaviruses. *Future Virology* 7, 31–39.
- Söderlund-Venermo, M., Brown, K., Erdman, D., 2017. Chapter 30. Human parvoviruses. In: Richman, D.D., Whitley, R.J., Hayden, F.G. (Eds.), *Clinical Virology*, fourth ed. ASM Press, pp. 679–699.
- Söderlund-Venermo, M., Lahtinen, A., Jartti, T., *et al.*, 2009. Clinical assessment and improved diagnosis of bocavirus-induced wheezing in children, Finland. *Emerging Infectious Diseases* 15, 1423–1429.
- Qiu, J., Söderlund-Venermo, M., Young, N.S., 2017. Human parvoviruses. *Clinical Microbiology Reviews* 30, 43–113.
- Väisänen, E., Fu, Y., Hedman, K., Söderlund-Venermo, M., 2018. Human protoparvoviruses. *Viruses* 9, 354.
- Väisänen, E., Fu, Y., Koskenmies, S., *et al.*, 2019. Cutavirus DNA in malignant and non-malignant skin of cutaneous T-cell lymphoma and organ transplant patients but not of healthy adults. *Clinical Infectious Diseases* 68, 1904–1910.
- Väisänen, E., Mohanraj, U., Kinnunen, P.M., *et al.*, 2018. Global distribution of human protoparvoviruses. *Emerging Infectious Diseases* 24, 1292–1299.
- Xu, M., Arku, B., Jartti, T., *et al.*, 2017. Comparative diagnosis of human bocavirus 1 respiratory infection with messenger RNA reverse-transcription polymerase chain reaction (PCR), DNA quantitative PCR, and serology. *Journal of Infectious Diseases* 215, 1551–1557.

Relevant Websites

<https://talk.ictvonline.org/taxonomy/>
 International Committee on Taxonomy of Viruses.
www.ictv.global/report/parvoviridae
 Parvoviridae
 ssDNA Viruses
 International.

Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (*Coronaviridae*)

Ding X Liu, Jia Q Liang, and To S Fung, South China Agricultural University, Guangzhou, China

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ACE2 Angiotensin converting enzyme 2
APN Aminopeptidase N
BAX Bcl-2-associated X protein
BST2 Bone marrow stromal antigen 2
CNS Central nervous system
CsA Cyclosporin A
CypA Cyclophilin A
DMV Double-membrane vesicle
DUB Deubiquitinating
EIA Enzyme-linked immunoassays
eIF2 α Eukaryotic initiation factor 2
ER Endoplasmic reticulum
ERGIC ER/Golgi intermediate compartment
ERK Extracellular signal-regulated kinase
ExoN Exoribonuclease
GRP Glucose regulated protein
HCoV Human coronavirus
HR Heptad repeat
IBV Avian infectious bronchitis coronavirus
IC Ion channel
ICTV International Committee on Taxonomy of Viruses
IFITM Interferon inducible transmembrane protein
IFN Interferon
IRE1 Inositol requiring enzyme 1

JNK c-Jun N-terminal kinase
MAP Mitogen-activated protein
MDA5 Melanoma differentiation-associated 5
MERS-CoV Middle East respiratory syndrome coronavirus
MHV Mouse hepatitis virus
Mpro Main protease
NendoU Endoribonuclease
NF- κ B Nuclear factor- κ B
nsp Non-structural protein
ORF Open reading frames
PERK PKR-like ER kinase
PLpro Papain-like proteases
RBD Receptor-binding domain
RdRP RNA-dependent RNA polymerases
RTC Replication-transcription complex
RT-LAMP Reverse transcription loop-mediated isothermal amplification
SARS-CoV Severe acute respiratory syndrome coronavirus
sgRNA Subgenomic RNA
TMPRSS2 Transmembrane serine protease 2
TRS Transcription-regulated sequences
UPR Unfolded protein response
UTR Untranslated region
XBP1 X-box protein 1

Glossary

Apoptosis A form of programmed cell death that occurs in multicellular organisms. Biochemical events lead to characteristic cell changes (morphology) and death.
Autophagy The natural, regulated mechanism of the cell that removes unnecessary or dysfunctional components. It allows the orderly degradation and recycling of cellular components.
ER stress A cellular process that is triggered by a variety of conditions that disturb the folding of proteins in the endoplasmic reticulum.
Innate immunity Refers to nonspecific defense mechanisms that come into play immediately or within hours upon pathogen recognition in the body. These mechanisms include physical barriers such as skin, chemicals in the blood, cytokine production and immune cells that attack foreign cells in the body.

Mitogen-activated protein kinase Serine-threonine protein kinases, important molecules in mediating the signal transduction from cell surface to nucleus, regulating cellular activities such as gene expression, mitosis, differentiation, and cell survival/apoptosis.
Type I interferon A large subgroup of interferon proteins that help regulate the activity of the immune system, and a pleiotropic cytokine with antiviral, antitumor and immunoregulatory functions.
Unfolded protein response A cellular stress response related to the endoplasmic reticulum stress. It has been found to be conserved between all mammalian species, as well as yeast and worm organisms.
Viroporin A group of small hydrophobic viral proteins that tend to oligomerize to form hydrophilic pores or ion channels in cellular membrane.

Background

The first human coronavirus (HCoV), strain B814, was isolated in 1965 from the nasal discharge of a patient with a common cold. Since then, more than 30 additional strains were identified. Among them, the prototypic strain HCoV-229E (named after a student specimen coded 229E) was isolated using standard tissue culture. HCoV-OC43 (Organ Culture 43) was later recovered using tracheal organ culture and found to be serologically distinct from HCoV-229E. These two viruses were the focus of HCoV research

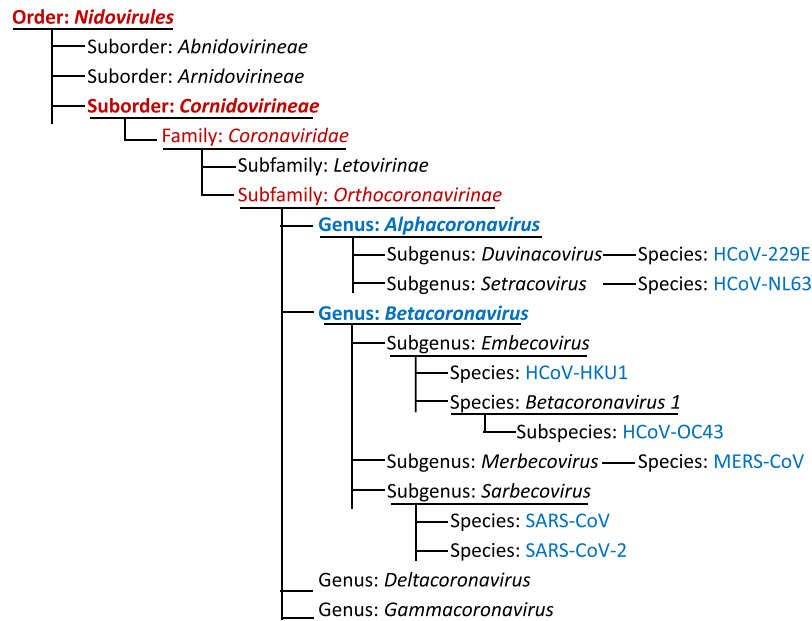


Fig. 1 The updated classification scheme of HCoVs according to the ICTV.

in the following years, until the emergence of the highly pathogenic severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002–2003. In the post SARS era, two more HCoVs were identified. HCoV-NL63 (NetherLand 63) was isolated from the aspirate of a 7-month-old infant with bronchiolitis in 2004, whereas HCoV-HKU1 (Hong Kong University 1) was isolated from a Hong Kong patient with pneumonia in 2005. Since then, two more zoonotic HCoVs have emerged, namely as the Middle East respiratory syndrome coronavirus (MERS-CoV) and the 2019 novel coronavirus (2019-nCoV, a.k.a. SARS-CoV-2). Unlike SARS-CoV, MERS-CoV and SARS-CoV-2 that are associated with severe respiratory disease, the four common HCoVs (229E, OC43, NL63, and HKU1) generally cause mild to moderate upper-respiratory tract illness, presumably contributing to 15%–30% of cases of common colds in human.

Classification

On the basis of the 10th International Committee on Taxonomy of Viruses (ICTV) report, coronaviruses are classified under the order *Nidovirales*, suborder *Cornidovirineae*, family *Coronaviridae*, subfamily *Orthocoronavirinae* (Fig. 1). According to serology studies and genomic analysis, *Orthocoronavirinae* is further divided into four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. Under the genus *Alphacoronavirus*, HCoV-229E and HCoV-NL63 belong to the subgenus *Duvinacovirus* and *Setracovirus*, respectively; under the genus *Betacoronavirus*, both HCoV-OC43 and HCoV-HKU1 belong to the subgenus *Embecovirus*.

Virion Structure

Under the electron microscope, coronavirus virions are spherical or pleomorphic. Coronavirus particles are enveloped, about 80–120 nm in diameter, with club-like projections of the spike (S) protein decorating the surface. In some *betacoronaviruses*, including HCoV-OC43 and HCoV-HKU1, shorter projections of the hemagglutinin-esterase (HE) protein are also observed. The viral envelope is supported by the membrane (M) protein and contains a small amount of the envelope (E) protein. Inside the viral envelope, the genome is bound by the nucleocapsid (N) protein to form a helical symmetric nucleocapsid. The common structural and functional features of HCoV structural proteins are briefly summarized as follows.

The S protein is a type I transmembrane protein, with a molecular weight of 128–160 kDa before glycosylation and 150–200 kDa after N-linked glycosylation. As a class I viral fusion protein, the S protein forms homotrimer and is cleaved by host proteases into a S1 subunit for receptor binding and a S2 subunit for membrane fusion. The ectodomain of the S protein is also modified by disulfide bonds, whereas the very short cytosolic tail is modified by palmitoylation. The S protein is the major determinant of host and tissue tropism, and may also contribute to viral pathogenesis by activating the endoplasmic reticulum (ER) stress response.

The HE protein is also a type I transmembrane protein, about 48 kDa before glycosylation and 67 kDa after N-linked glycosylation. It forms homodimer via disulfide bonds. With its sialic acid-binding hemagglutinin activity, the HE protein may serve as a cofactor of S protein and facilitate virion attachment. Additionally, as it possesses esterase activity that removes acetyl groups from O-acetylated sialic acids, it has been postulated to have a role as a receptor-destroying enzyme that facilitates the

release of progeny virions from nonpermissive host cells, thereby enhancing virion spreading in the extracellular milieu. In fact, the HE protein of HCoV-HKU1 mediated receptor-destroying enzyme activity specific to the *O*-acetylated sialic acids recognized by its own S protein.

The M protein (25–30 kDa) is the most abundant structural protein and possesses three transmembrane domains. The short N-terminal ectodomain of the M protein is modified by *O*-linked glycosylation in HCoV-OC43 and some animal coronaviruses including mouse hepatitis virus (MHV) and bovine coronavirus (BCoV). However, in HCoV-229E, HCoV-NL63, and most other coronaviruses, the ectodomain of M protein is modified by N-linked glycosylation. The M protein forms homodimer and interacts with other viral structural proteins to orchestrate the assembly of the coronavirus particle. This protein may also contribute to viral pathogenesis. For example, retinoic acid-inducible gene 1 (RIG-I)-dependent induction of type I interferon (IFN) is observed in cells overexpressing the M protein of SARS-CoV but not HCoV-HKU1.

The E protein is a small (8–12 kDa) integral membrane protein found in low amounts in the virion. Current evidence strongly suggests that the E protein adopts an N-ecto/C-endo topology with one transmembrane domain. The SARS-CoV E protein is modified by N-linked glycosylation and three cysteine residues in its endodomain are modified by palmitoylation. Additionally, the E protein of SARS-CoV and avian infectious bronchitis coronavirus (IBV) has been shown to form homopentamers with ion channel (IC) activity. The IC activity may modulate the process of virion release and contribute to viral pathogenesis. Although the deletion of the E gene is not lethal for SARS-CoV, the mutant virus is severely defective in virion morphogenesis and attenuated *in vivo* compared with the wild type control.

Underneath the viral envelope, the N protein (43–50 kDa) forms dimer and binds to the genomic RNA in a beads-on-a-string fashion, forming a helically symmetric nucleocapsid. In SARS-CoV and other coronaviruses, the N protein is phosphorylated by cellular kinases such as glycogen synthase kinase 3 (GSK3) and ataxia-telangiectasia mutated and Rad3-related. Other modifications such as SUMOylation, ADP-ribosylation, and proteolytic cleavage by caspases has also been demonstrated in the N protein of some coronaviruses. The N protein facilitates RNA packing and is involved in many other processes, including viral genome replication and evasion of the immune response.

Genome

With a single-stranded, positive-sense RNA genome containing approximately 27 to 32 kilobases (kb), coronavirus has the largest viral RNA genome described so far. The genome sizes are approximately 27.5 kb for HCoV-229E and HCoV-NL63, and more than 30 kb for HCoV-OC43 and HCoV-HKU1 (Fig. 2). Because the genomic RNA harbors a 5'-cap structure and a 3'-polyadenylate tail, it can act directly as a messenger RNA (mRNA) encoding the viral replicase. Additionally, the genome also serves as a template for RNA replication and the genome is packed into progeny virions. There are two untranslated regions (UTRs) flanking the coding region. The 5'-UTR is 292, 210, 286, and 205 nucleotides long in HCoV-229E, -OC43, -NL63, and -HKU1, respectively, and contains a leader sequence (~70 nucleotides long) at its 5'-terminus. At the other end of the genome, the 3'-UTR is 462, 288, 287, and 281 nucleotides long in HCoV-229E, -OC43, -NL63, and -HKU1 respectively, and it contains a highly conserved octameric sequence ~70 nucleotides upstream from the poly(A) tail.

The HCoVs replicase gene comprises the 5'-two thirds of the genome and is made up of two overlapping open reading frames (ORFs)—ORF1a and ORF1b. ORF1a is directly translated from the RNA genome, producing the polyprotein pp1a; whereas the translation of ORF1b requires a programmed ribosomal frameshift near the 3' end of ORF1a, leading to the production of polyprotein pp1ab. The autoproteolytic cleavage of pp1a and pp1ab then gives rise to 16 non-structural proteins (nsp1–16).

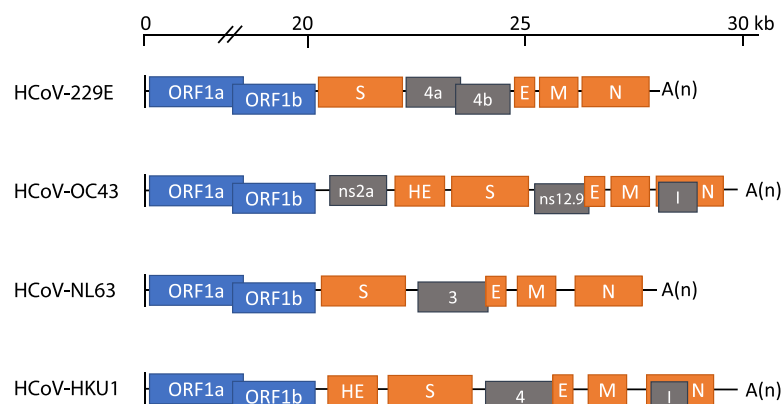


Fig. 2 Genome structure of human coronaviruses (HCoVs) – HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU1. The open reading frame 1a (ORF1a) and ORF1b are represented as shortened navy blue-boxes. The genes encoding structural proteins spike (S), envelope (E), membrane (M), nucleocapsid (N), and hemagglutinin-esterase (HE) are shown as orange boxes. The genes encoding accessory proteins are shown as dark gray boxes.

The remaining one third of the HCoV genome contains ORFs for viral structural proteins in the order of 5'-(HE)-S-E-M-N-3', as well as several accessory proteins that are distinct among HCoVs of different species and genus. Specifically, HCoV-229E encodes two accessory proteins 4a and 4b; HCoV-OC43 encodes three accessory proteins ns2a, ns12.9 (a.k.a. ns5a), and protein I (a.k.a. N2 or N internal ORF protein); HCoV-NL63 encodes a single accessory protein 3; and HCoV-HKU1 encodes two accessory proteins: protein 4 and protein I. Although these accessory proteins are dispensable for viral replication in cell culture, they may be involved in viral pathogenesis and contribute to virulence in vivo. Unlike the viral replicase, structural proteins and accessory proteins are translated from a 3'-nested set of subgenomic RNA (sgRNA) species. The coding sequences of some accessory proteins overlap with those of the structural proteins, but are translated in distinct reading frames. In fact, the accessory gene encoding protein I of HCoV-OC43 and HCoV-HKU1 are encoded internally of the N genes.

Non-Structural Proteins

The HCoV polyproteins pp1a and pp1ab are autocatalytically processed by viral proteases into 16 non-structural proteins. Nsp1, the most N-terminal cleavage product of the polyproteins, has been shown to suppress host protein synthesis and IFN response.

Nsp3 encodes one or two papain-like proteases (PLpro). Nsp3 of the four common HCoVs contains two PLpro domains (PLP1 and PLP2), whereas the nsp3 of SARS-CoV and MERS-CoV contains only one PLpro domain. In general, PLP1 processes at cleavage site 1 to release nsp1, whereas PLP2 is responsible for the processing at both cleavage sites 2 and 3 to release nsp2 and nsp3. The deletion of PLP2 is lethal, whereas the proteolytic activity of PLP1 is dispensable for HCoV-229E replication. Notably, a recent study reported that ectopic expression of HCoV-NL63 PLP2 induces proteasomal degradation of p53, thereby inhibiting p53-dependent production of type I IFN and the innate immune response.

Following the translation of ORF1a and ORF1ab, nsp5 is properly folded within the context of the replicase polyprotein and orchestrates its own autoproteolytic processing. Nsp5 cleaves pp1a/pp1ab at as many as 11 sites to produce a total of 13 mature proteins, and is therefore indispensable for virus replication. Nsp5 is also referred to as the coronavirus main protease (Mpro).

Nsp6 of some coronaviruses, such as IBV, MHV, or SARS-CoV, activates the formation of autophagosomes from the ER via an omegasome intermediate. Nsp3, nsp4, and nsp6 are also responsible for remodeling cellular membranes to form double-membrane vesicles (DMVs) or ER spherules, onto which the coronavirus replication-transcription complex (RTC) is assembled and anchored.

The complex of HCoV-229E nsp7 and nsp8 is capable of synthesizing short RNA strands of ~6 nucleotides. Coronavirus nsp8 protein also has template-dependent RNA polymerase activities resembling those of RNA primases or even canonical RNA-dependent RNA polymerases (RdRP). Current evidence suggests an essential cofactor function of nsp7 and nsp8 for the RNA-dependent RNA polymerase activity of nsp12. In a recent study, it is found that nsp8 of HCoV-229E has a metal ion-dependent RNA 3'-terminal adenylyl transferase (TATase) activity, when a partially double-stranded RNA with a short 5' oligo-U sequence is provided as the template strand. This supports the notion that nsp8 may catalyze the 3'-polyadenylation of HCoV genomes.

Nsp9, a single-stranded DNA/RNA-binding protein which exists as a dimer in physiological situations, has been found to be indispensable for virus replication based on reverse-genetics experiments. Nsp10 is a double-stranded RNA-binding zinc-finger protein. Nsp7, nsp8, nsp9, and nsp10 are all closely associated with the replication complex built around the RNA-dependent RNA polymerase (nsp12).

Nsp13 separates the double-stranded replicative intermediates to provide single-stranded templates for RNA synthesis. HCoV-229E nsp13 contains an N-terminal zinc-binding domain and a C-terminal superfamily 1 helicase domain. It exhibits a variety of enzymatic activities including NTPase, dNTPase, and RNA/DNA helicase activity. Using the NTPase active site, this protein also has an RNA 5'-triphosphatase activity, which may be involved in the capping of viral RNAs.

Nsp14 demonstrates exoribonuclease (ExoN) activities. RNA viruses generally have high mutation rates that allow for rapid viral adaptation in response to selective pressure. Nsp14-ExoN is the first proofreading enzyme identified for an RNA virus, and it functions together with other CoV replicase proteins to perform the crucial role of maintaining CoV replication fidelity.

Nsp15 is an endoribonuclease (NendoU) and a type I interferon antagonist. HCoV nsp15-NendoU can excise both single- and double-stranded RNA and specifically recognize uridylylates to produce 2-3'-cyclic phosphodiester products, thereby preventing the activation of the host innate immune response.

Nsp16 has a 2'-O-methyltransferase activity. The 2'-O-methylation capping protects viral RNA from recognition by melanoma differentiation-associated protein 5 (MDA5) and thus prevents MDA5-dependent production of type I interferon in virus-infected cells.

Accessory Proteins

Although not essential for viral replication in cell culture, coronavirus accessory proteins may play in vivo functions that have not yet been fully elucidated. Most HCoV accessory proteins are genus-specific and show low homology to known proteins. But an accessory gene between the S and E gene is encoded by three HCoVs (3a of SARS-CoV, 4a of HCoV-229E, and ns12.9 of HCoV-OC43), suggesting a conserved role during HCoV infection. Indeed, all three proteins have been shown to serve as viroporins that regulate viral replication. Viroporins are oligomeric hydrophobic viral proteins that form and insert ion channels into the host cell membrane. The HCoV-OC43 ns12.9 protein is a recently identified viroporin that facilitates virion

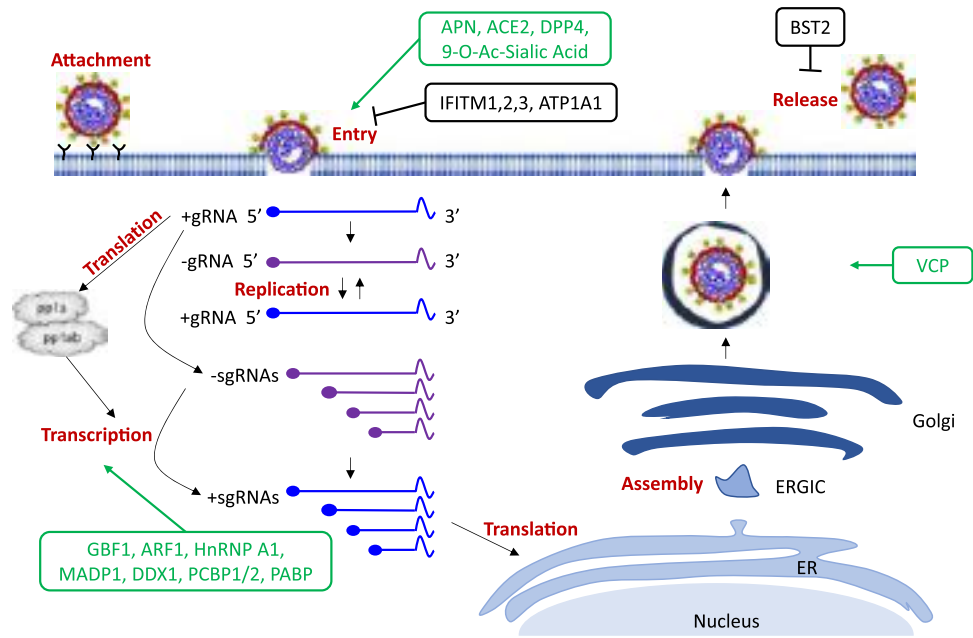


Fig. 3 HCoV replication cycle. HCoV infection is initiated by binding of virions to cellular receptors, which drives the conformational change in the S2 subunit in S, promoting the fusion of the viral and cell plasma membrane. The gRNA serves as the template for translation of polyproteins pp1a and pp1ab, which are cleaved to form nsps. Nsps induce the rearrangement of cellular membrane to form DMVs, where the viral RTCs are anchored. Full-length gRNA is replicated via a negative-sense intermediate, and a nested set of sgRNA species are synthesized by discontinuous transcription. These sgRNAs encode viral structural and accessory proteins. The viral products produced will be assembled in the ERGIC, and bud out as a smooth-wall vesicle to the plasma membrane to egress via exocytosis. Host factors that promote infection and inhibit infection are highlighted in green and black, respectively.

morphogenesis and pathogenesis. The expression of important mediators of innate immune response was downregulated in cells expressing of HCoV-OC43 ns12.9, presumably due to inhibited promoter activity of ISRE, IFN- β , and NF- κ B.

Interestingly, because of a scenario in which HCoV-229E acquired an out-of-frame insertion or deletion, instead of an ORF4a and ORF4b, an intact ORF4 was present in some HCoV-229E clinical isolates. ORF4a of HCoV-229E was expressed in infected cells and localized at the ER/Golgi intermediate compartment (ERGIC). The ORF4a protein formed homo-oligomers through disulfide bridges and possessed ion channel activity in both *Xenopus* oocytes and yeast.

The NS2 protein of HCoV-OC43 has cyclic phosphodiesterase activity, which may modulate cAMP-mediated signaling and important physiological processes such as lipid metabolism and apoptosis. Finally, although the accessory protein encoded by HCoV-NL63 ORF3 appears to be nonessential in cell culture, there were differences in RNA synthesis, protein expression, plaque morphology, and virus growth in cells infected with the ORF3-deleted mutant compared with the control.

Life Cycle

The replication cycle of HCoV can be arbitrarily divided into five steps: attachment to host cells, viral entry and uncoating, expression of the viral replicase and formation of the replication-transcription complex, viral RNA synthesis, and virion assembly and release. Each step will be briefly introduced as follows (Fig. 3).

Attachment

Coronavirus infection is initiated by binding of virions to cellular receptors. The S protein includes two functional domains: S1 (bulb) is the part binding to the receptor(s) and S2 (stalk) is responsible for fusion between virion and cell membranes. The receptor-binding domain (RBD) of S1 varies among different coronaviruses. RBDs of HCoV-229E, HCoV-NL63 and HCoV-HKU1 are located in the C-terminal region but not the N-terminal domains of the respective S1 subunits.

Receptor binding is critical to initiate viral infection. HCoV has been shown to use either cellular proteins or carbohydrates displayed on the plasma membrane as receptors. Interestingly, all known protein receptors for HCoVs are cell surface peptidase, such as aminopeptidase N (APN) for HCoV-229E, dipeptidyl peptidase 4 (DPP4) for MERS-CoV, and angiotensin converting enzyme 2 (ACE2) for HCoV-NL63, SARS-CoV and SARS-CoV-2. On the other hand, HCoV-OC43 and HCoV-HKU1 employ glycan-based receptors carrying 9-O-acetylated sialic acid.

In addition to the receptor binding by the S protein, other HCoV structural proteins may also facilitate the early stage of attachment. For example, the M protein of HCoV-NL63 binds to target cells using heparan sulfate proteoglycans as the initial attachment factors. This is followed by the engagement of the S protein with the ACE2 receptor protein.

Viral Entry and Uncoating

Specific binding between S1 and the cognate receptor leads to a conformational change in the S2 subunit and large-scale rearrangements of the S protein, resulting in the fusion between virus and cell membranes and the release of viral nucleocapsid into the cytoplasm. Many host factors are involved in the entry and uncoating of HCoVs.

The cleavage of S protein into S1 and S2 subunits is mediated by one or more cellular proteases. HCoV-229E can enter host cells via two distinct pathways: one is mediated by surface proteases like type II transmembrane protease serine 2 (TMPRSS2), and another is mediated by endosomal cathepsin L. However, to avoid triggering an innate immune response, it is more likely that HCoV-229E enters cells via the TMPRSS2 pathway, as the endosome is a main site for recognition by toll-like receptors. Similarly, the entry of SARS-CoV depends on the endosomal cysteine protease cathepsin L and another trypsin-like serine protease to activate the S protein.

Coated vesicles and the cytoskeleton are utilized by some HCoV for entry. For example, HCoV-OC43 employs caveolin-1-dependent pathway of endocytosis, and the scission of virus-containing vesicles from the cell surface is dynamin-dependent. Internalization of these vesicles also requires rearrangements of the actin cytoskeleton. The interaction between HCoV-NL63 S protein and the ACE2 receptor molecule triggers the recruitment of clathrin. Subsequent vesicle scission by dynamin results in virus internalization, and the newly formed vesicle passes the actin cortex, which requires active cytoskeleton rearrangement.

Finally, acidification of the endosomal microenvironment is required for successful fusion and release of the viral genome into the cytoplasm. Additionally, virion release of HCoV-229E and IBV from early endosomes is shown to be dependent on the host factors valosin-containing protein (VCP), an AAA ATPase family protein that normally facilitates the export of misfolded proteins from the ER to the cytoplasm.

On the other hand, some host factors could prevent the entry and uncoating of HCoVs. Interferon inducible transmembrane proteins (IFITMs) exhibit broad-spectrum antiviral functions against various RNA viruses. IFITMs restricted the entry of HCoV-229E and HCoV-NL63, as well as SARS-CoV and MERS-CoV. Conversely, IFITM2 or IFITM3 serve as an entry factor to promote the infection of HCoV-OC43.

Formation of the Replication-Transcription Complex

Among RNA viruses, the transcription of coronavirus RNA is unique. First, in order to maintain genetic stability, the large genome size requires unusual enzymatic activities, such as an exoribonuclease and an endoribonuclease activity. Also, the synthesis of a nested set of sgRNAs by discontinuous transcription demands a huge and complicated RTC.

Following the release of viral nucleocapsid into the cytoplasm, the genomic RNA serves as a transcript encoding the viral replicase. The replicase gene includes two ORFs, ORF1a and ORF1b. ORF1a is translated into the polyprotein pp1a (440–500 kDa). Owing to a slippery sequence and an RNA pseudoknot near the end of ORF1a, a programmed –1 ribosomal frameshifting event can occur with a frequency of 25%–30%. This allows a continuous translation into ORF1b, producing a larger polyprotein pp1ab (740–810 kDa). Similar to likely all coronaviruses, the pp1a and pp1ab are autoproteolytically processed into 16 nsps, collectively forming the RTC for viral RNA synthesis. Notably, nsp3, 4, and 6 appear to be responsible for remodeling of cellular membranes to form DMVs or spherules, onto which the HCoV RTC is assembled and anchored. Host factors of the early secretory pathway, such as Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1 (GBF1) and its effector ADP ribosylation factor 1, may also contribute to DMV formation and RTC assembly.

Viral RNA Synthesis

The assembly of RTC sets a foundation for viral RNA synthesis. By using the genomic RNA as a template, new copies of full-length viral genome are synthesized, using the full-length negative sense genomic RNA as the intermediate template. Meanwhile, the polymerase can switch templates at short motifs called transcription-regulated sequences (TRS) during negative sense RNA synthesis, producing a 5'-nested set of negative sense sgRNAs. These are in turn used as templates to synthesize a 3'-nested set of positive sense sgRNAs, which serve to encode the structural and accessory proteins. The core sequence of HCoV TRS is a conserved hexamer CUA AAC in HCoV-229E and HCoV-NL63, and a conserved heptamer UCU AAC in HCoV-OC43 and HCoV-HKU1.

Different numbers of sgRNAs are produced by the four common HCoVs. For instance, seven major viral RNA species are produced during HCoV-229E infection. The full-length genome (mRNA1) encodes the viral replicase, whereas mRNAs 2, 4, 5, 6, and 7 encode the S protein, accessory protein 4, E protein, M protein, and N protein, respectively. Notably, mRNA 3 is considered defective, because it contains a truncated version of the S gene that is not translated.

Although the replication and transcription of viral genome is mainly carried out by the replicase, the involvement of other factors including viral structural protein and host proteins has been implicated. For instance, by serving as an RNA chaperone, the coronavirus N protein can facilitate template switching during sgRNA synthesis. Host proteins such as heterogeneous

nuclear ribonucleoprotein A1, polypyrimidine tract-binding protein, mitochondrial aconitase, and polyadenylate-binding protein have been suggested to participate in coronavirus RNA synthesis, presumably mediated by their RNA binding activity.

Assembly and Release of Virion

Coronavirus structural and accessory proteins that are membrane-associated (such as S, HE, M, and E) are translated by ribosomes in the ER, whereas other viral proteins (such as N) are translated by free ribosomes. Most coronavirus structural proteins are also subjected to posttranslational modifications that modulate their functions. The building parts converge at the assembly site of ERGIC. Assembly is orchestrated by the M protein: homotypic interaction of M protein provides the scaffold for virion morphogenesis, whereas heterotypic interactions of M protein with other structural proteins, such as M-S and M-E, facilitate their recruitment and incorporation. Virion assembly is completed by the condensation of the nucleocapsid with the envelope components, a process mediated by M-N interactions. A small amount of E protein may provide the driving force for envelope morphogenesis by inducing membrane curvature.

After assembly, progeny virions are transported in smooth-wall vesicles, trafficked to the plasma membrane via the secretory pathway, and released by exocytosis. The cytoskeletal system also participates in HCoV assembly and release. For instance, the interaction between tubulin and the cytosolic domain of S protein are required for the assembly and release of infectious virions during HCoV-229E and HCoV-NL63 infection. In addition, bone marrow stromal antigen 2 (BST2, a.k.a. tetherin), an interferon-inducible antiviral protein, blocks the release of various envelope viruses by interfering with the budding step at the plasma membrane. Although budding of HCoV-229E occurs at the ERGIC, it is recently shown that BST2 can trap HCoV-229E virions in the intracellular vesicles, thereby suppressing the release of progeny viruses. Interestingly, the S protein of SARS-CoV downregulates BST2 at the protein level by promoting its lysosomal degradation, thus antagonizing the BST2 tethering of SARS-CoV, HCoV-229E, and HIV-1 virus-like particles.

HCoV-Host Interactions

As intracellular obligate parasites, HCoVs exploit the host cell machinery for their own replication and spread. Since virus–host interactions also form the basis of viral pathogenesis, knowledge about their interplay is of great research interest (Fig. 4).

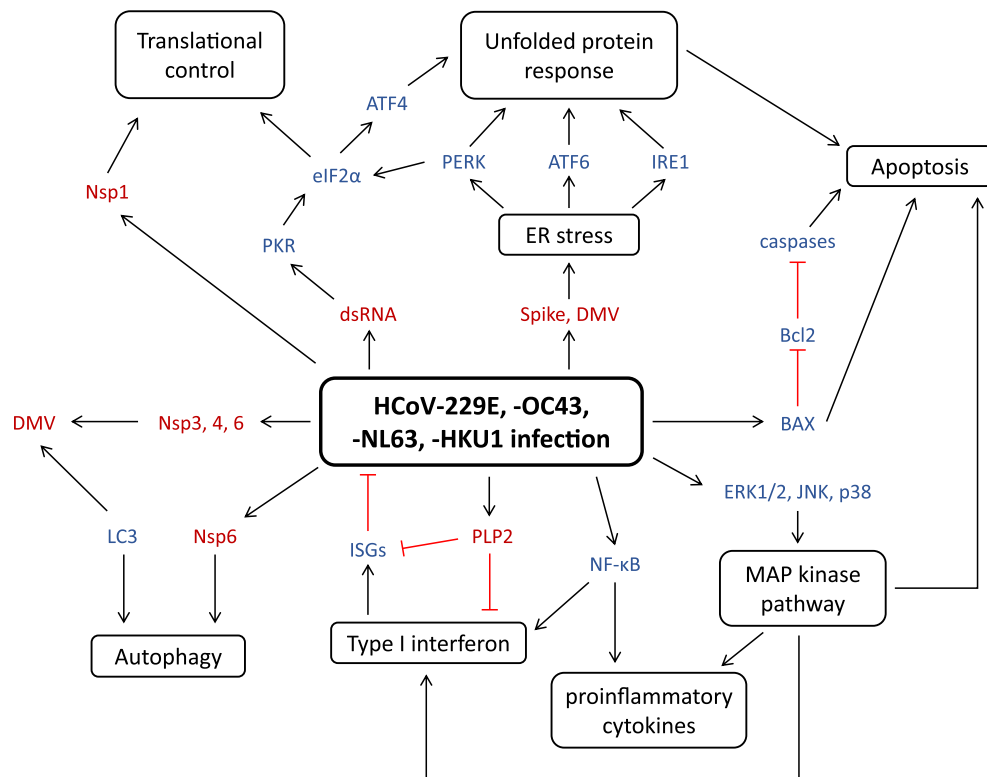


Fig. 4 HCoV-host interaction. Schematic diagram showing the host signaling pathways activated during HCoV infection. Black pointed arrows indicate activation, and red blunt-ended lines indicate inhibition. Viral components modulating the pathway are in red, and host proteins are in blue. See text for detail.

Translational Control

Viruses must utilize the host translation machinery to ensure efficient viral protein translation. In response to acute viral infection, host cell would shut down the protein translation system to cope with the infection stress, which is regarded as an integrated stress response. Integrated stress response is marked by the phosphorylation of the α -subunit of eukaryotic initiation factor 2 (eIF2 α), downregulation of the general cap-dependent protein synthesis, and up-regulation of the expression of certain transcription factors, such as activating transcription factor 4 (ATF4). Due to the translation competition between cellular and viral mRNAs for limiting number of ribosomes and associated factors, coronavirus must hijack the host translational machinery to produce its own proteins.

Some HCoVs, such as SARS-CoV and MERS-CoV, have been shown to induce host translation shutoff in susceptible cells. Specifically, SARS-CoV infection leads to sustained phosphorylation of eIF2 α in 293T/ACE2 cells. Coronavirus nsp1 also suppresses host protein synthesis and IFN response. For example, SARS-CoV nsp1 suppresses the expression of host genes including type I IFN, thereby counteracts the host innate immune response and contributes to virulence. MERS-CoV nsp1 also negatively regulates host gene expression by inhibiting the translation and inducing the degradation of host mRNAs. Nsp1 of HCoV-229E and HCoV-NL63 inhibits the expression of reporter genes, probably by binding to ribosomal protein S6 and blocking the mRNA binding to the 40S ribosomal subunit.

ER Stress Response

The ER is a cellular organelle important for protein synthesis, folding, and post-translational modifications. In normal circumstances, the ER can be loaded with a very high concentration of proteins without perturbing its unique luminal environment. However, when the protein load exceeds the ER folding and processing capacity, misfolded or unfolded proteins will accumulate within the ER, resulting in ER stress. ER stress activates the signaling pathways collectively known as unfolded protein response (UPR). These are initiated by three ER transmembrane sensors: protein kinase R (PKR)-like ER kinase (PERK), inositol requiring enzyme 1 (IRE1) and activating transcriptional factor 6 (ATF6). The activation of UPR restores the ER homeostasis by enhancing protein folding, attenuating protein translation and ER-associated degradation (ERAD).

HCoV infection triggers ER stress and activates UPR in virus-infected cells. Overexpression of the S protein of HCoV-HKU1 and SARS-CoV can activate PERK and the promoters of GRP78 and GRP94. Infection with HCoV-OC43 activates IRE1 and induces X-box protein 1 (XBP1) mRNA splicing, thereby upregulating downstream UPR effector genes. Introduction of two point mutations (H183R and Y241H) into the S protein of HCoV-OC43 induces a higher degree of XBP1 mRNA splicing and results in a more pronounced apoptotic cell death. It is also reported that when the E gene is deleted from SARS-CoV, the mutant virus also induces a higher level of XBP1 mRNA splicing and apoptosis, compared with the wild type control. This suggests that activation of the IRE1-XBP1 pathway may be generally pro-apoptotic during HCoV infection.

MAP Kinase Pathway

The family members of the mitogen-activated protein (MAP) kinases mediate a wide variety of cellular processes in response to extracellular stimuli. Four distinct subgroups within the MAP kinase family have been described: extracellular signal-regulated kinase 1/2 (ERK1/2), ERK5, c-Jun N-terminal kinase (JNK), and the p38 group of protein kinases. Activation of the ERK pathway has been observed in cells infected with a number of HCoVs, including SARS-CoV, MERS-CoV, and HCoV-229E. Phosphorylation of the 90-kDa ribosomal protein S6 kinase (p90RSK), a key substrate of ERK, was also observed in SARS-CoV-infected Vero E6 cells. Activation of p38 and its upstream kinases has been detected in cells infected with HCoV-229E, SARS-CoV, and MERS-CoV. In fact, the replication of HCoV-229E is suppressed by treatment with p38 inhibitor SB203580 or chloroquine, which attenuates p38 activation. As for JNK, phosphorylation of JNK and its upstream kinases was observed in cells infected with SARS-CoV or in cells overexpressing the SARS-CoV S protein. Notably, treatment with JNK inhibitor abolished persistent infection of SARS-CoV. Apart from their involvement in cell survival and apoptosis, MAP kinases also contribute significantly to the induction of pro-inflammatory cytokines during HCoV infection.

Autophagy

Autophagy is a conserved cellular process involving self-eating. Specifically, cells under stress conditions, such as starvation, growth factor deprivation, or infection by pathogens, initiate autophagy in nucleation sites at the ER, where part of the cytoplasm and/or organelles are sequestered in DMVs (autophagosomes) and degraded in response to a fusion with lysosomes. Autophagy is regulated by highly conserved autophagy-related genes.

Coronavirus infection activates the formation of autophagosomes, but inhibition of autophagy does not affect viral replication. The non-structural protein 6 (nsp6) is a transmembrane protein implicated in the formation of DMVs during SARS-CoV infection. Overexpression of nsp6 of IBV, MHV, or SARS-CoV induced the formation of autophagosomes from the ER via an omegasome intermediate. However, autophagosomes induced by IBV infection or overexpression of coronavirus nsp6 had smaller diameters compared with those induced by starvation, indicating that nsp6 may also restrict the expansion of autophagosomes. Notably, recent studies have shown that coronavirus employs the host machinery for COPII-independent vesicular ER export to derive cellular membranes for DMV formation. Although this process requires an autophagy-related gene called LC3, it is independent of host autophagy.

Apoptosis

Apoptosis is a form of programmed cell death that is tightly regulated. When cells undergo apoptosis, they demonstrate specific hallmarks such as cell shrinkage, extensive plasma membrane blebbing, nuclear condensation, and DNA fragmentation. During viral infections, apoptosis is induced as one of the host antiviral responses to limit virus replication and production. Two main mechanisms of apoptosis have been established – the extrinsic and intrinsic pathways. The extrinsic pathway is initiated by the binding of extracellular death ligands to death receptors from the tumor necrosis factor super-family. The intrinsic pathway occurs internally in the cell and involves changes in the mitochondrial outer membrane permeability based on the ratio of pro-apoptotic and anti-apoptotic B-cell lymphoma 2 (Bcl2) family proteins.

As HCoVs are known to infect tissue cultures, they have been associated with apoptosis induction in a wide spectrum of cell types during infection, including intestinal mucosal cells, kidney tubular cells and neuronal cells. Apoptosis in neuronal cells infected with HCoV-OC43 involved mitochondrial translocation of Bcl-2-associated X protein (BAX) but this phenomenon was independent of caspase activation. HCoV-229E infection resulted in massive cytopathic effect (CPE) and cell death in dendritic cells, albeit independent of apoptosis induction. Since dendritic cells are prevalent throughout the human body, it is possible that they are used as a vehicle to facilitate viral spread. Induction of apoptosis during HCoV infection is also regulated by cellular stress response pathways such as the UPR and MAP kinase pathways described above.

Pathogenesis

HCoV-229E, -OC43, -NL63, and -HKU1 are considered as pathogens causing upper respiratory tract disease and responsible for up to 15%–30% of common colds in adults. Unlike SARS-CoV that spreads from the upper airway to cause a severe lower respiratory tract infection, HCoV-229E and HCoV-OC43 replicate principally in the upper respiratory tract epithelial cells, where they produce virus and cause local respiratory symptoms.

There are striking differences in the extent of genetic variability when isolates of HCoV-OC43 and HCoV-229E are compared. HCoV-229E isolated at geographically distinct locations shows little genetic variability, whereas for HCoV-OC43 the opposite is true. The ability of HCoV-OC43 to tolerate mutations probably accounts for its ability to grow in mouse cells and to infect the mouse brain as well. As for HCoV-NL63, it shares homology with HCoV-229E and phylogenetic analyses suggest that HCoV-NL63 and HCoV-229E diverged approximately 1000 years ago.

HCoVs attach to cellular receptors by the S protein on the surface of the virion. Internalization into host cells occurs by direct fusion with the plasma membrane or by endocytosis. Viral receptors, components that actively promote host cell entry, differ greatly from one virus to another, and each with its own distinct physiological functions. Both APN, the receptor for HCoV-229E, and ACE2, the receptor for SARS-CoV and HCoV-NL63, exist as prominent zinc-dependent peptidases on the plasma membrane. However, unlike SARS-CoV, HCoV-NL63 does not require cathepsin L or endosomal acidification to infect ACE2-expressing cells. In addition to the receptor ACE2, the entry of HCoV-NL63 also requires heparan sulfate proteoglycans on the cell surface, which act as attachment factors that increase the virus density and possibly facilitate receptor binding. Proteolytic cleavage of the HCoV S protein is also an important regulatory mechanism. In recent studies, the proteolytic activation of the HCoV-229E S protein is analyzed using trypsin-like serine proteases. It is found that fusion activation is not dependent on the cleavage of the S1/S2 site, but is highly dependent on the cleavage in the S2' region. This is very similar to the fusion activation of the IBV S protein, which requires furin-dependent cleavage at the S2' site.

Compared with SARS-CoV and MERS-CoV, the ability of other HCoVs causing mild infections to suppress the host antiviral response is less well studied. HCoV-229E replication is shown to attenuate the inducible activity of the transcription factor nuclear factor kappa-B (NF- κ B) and to restrict the nuclear accumulation of NF- κ B subunits. Overexpression of structural or accessory proteins of HCoV-OC43 also leads to downregulation of over 30 genes related to innate immune response, including genes encoding MAP kinases, toll-like receptors, interferons, interleukins, and signal transduction proteins. Similar to the PLPro of SARS-CoV and MERS-CoV, the PLP2 of HCoV-NL63 possesses deubiquitinating (DUB) activity. This DUB activity may remove ubiquitins from critical intermediates in the innate immune signaling pathway, thereby suppressing host antiviral response. PLPro of SARS-CoV and MERS-CoV also recognizes and catalyzes the removal of another ubiquitin-like modifier called interferon-stimulated gene 15 (ISG15). Such deISGylating activity has not been fully characterized for other HCoVs.

HCoV infection is not always confined to the upper respiratory tract and can invade the central nervous system (CNS) under circumstances that are presently uncharacterized. Although evidence for a significant correlation between the presence of HCoV-229E and HCoV-OC43 RNA and multiple sclerosis has not been demonstrated, accumulating evidence from cell culture and animal models highlights their neurotropic and neuroinvasive potential. HCoV-229E RNA is detected in about 44% (40 of 90) of human brains tested, with similar frequencies in brains from multiple sclerosis patients and patients who died from other neurologic diseases or normal control subjects. The detection of HCoV RNA in human brain samples clearly demonstrates that these respiratory pathogens are naturally neuroinvasive in humans and suggests that they may establish a persistent infection in the human CNS. Therefore, the close structural and biological relatedness of HCoV to the neurotropic animal coronaviruses has led to speculation of the possible involvement of HCoV in neurological diseases. Research proved that classical apoptosis associated with the BAX protein does not play a significant role in HCoV-OC43-induced neuronal cell death and that RIP1 and MLKL, two cellular proteins that are usually associated with necroptosis, another form of regulated cell death when apoptosis is

not adequately induced. It is likely that in HCoV-OC43-infected neuronal cells, RIP1 and MLKL are activated to induce necroptotic cell death in an attempt to limit viral replication. However, this regulated cell death also leads to neuronal loss in the mouse CNS and accelerates the neuroinflammation process, reflecting the severity of neuropathogenesis.

Epidemiology and Clinical features

HCoV-229E, -OC43, -NL63, and -HKU1 are distributed globally. By spreading via coughing and sneezing, they cause mild upper respiratory tract diseases in adults. However, in infants, young children, elderly, and immunocompromised individuals, they may sometimes cause life-threatening bronchiolitis and pneumonia. Apart from respiratory illnesses, they may also result in enteric and neurological diseases. Natural infection is probably acquired in a fashion similar to that of many other respiratory viruses, with primary infection of ciliated epithelial cells in the nasopharynx. Although HCoVs were found to cause epidemics every 2–3 years with a high probability of reinfections, there is a paucity of evidence on the epidemiology and clinical manifestations of these four HCoVs worldwide. Reinfection of HCoVs demonstrates that infection does not induce long-lasting protective immunity.

HCoV-229E

First isolated in 1966, HCoV-229E is proposed to originate from African hipposiderid bats and adopt camelids as intermediate hosts. The evolutionary history of HCoV-229E likely shares important characteristics with that of the recently emerged highly pathogenic MERS-CoV. HCoV-229E infection is associated with common cold symptoms in healthy adults. But younger children and the elderly are vulnerable to lower respiratory tract infections. In particular, immunocompromised patients have been reported to suffer severe and life-threatening lower respiratory tract infections attributed to HCoV-229E. Moreover, the serological test suggested the possible involvement of HCoV-229E in the development of Kawasaki syndrome.

The incubation period of HCoV-229E is approximately 2–5 days, followed by illness lasting 2–18 days. Symptoms of HCoV-229E infection include headache, nasal discharge, sneezing, sore throat and general malaise. A few patients exhibit fever and cough. The clinical features can be distinguished from respiratory tract infections caused by other pathogens such as influenza A virus. HCoV-229E tends to be epidemic during winter in temperate-climate countries, and tests of a HCoV-229E laboratory strain suggested it is relatively stable in the environment.

HCoV-OC43

First isolated in 1967, HCoV-OC43 has no serological cross-reactivity with HCoV-229E. However, patients infected with HCoV-OC43 or HCoV-229E cannot be distinguished based on clinical symptoms alone; although coryza (inflammation of the mucous membrane of the nose) occurs more often during HCoV-229E infections, while sore throat manifestations occur more often during HCoV-OC43 infections.

HCoV-OC43 is generally associated with mild upper respiratory tract infections, although it has also been shown to have neuroinvasive properties. In vivo studies in mice have shown that HCoV-OC43 can infect neurons and cause encephalitis. The virus has also been shown to cause persistent infections in human neural cell lines. Since the first isolation of HCoV-OC43 in the 1960s, seven genotypes (A–G) have been identified by phylogenetic analysis. HCoV-OC43 is also transmitted primarily during the winter in temperate climates.

HCoV-NL63

HCoV-NL63 was isolated from a 7-month-old girl with coryza, conjunctivitis, fever, and bronchiolitis in the Netherlands in 2004. HCoV-NL63 and HCoV-OC43 cause most of the respiratory infections leading to hospitalization. Although HCoV-NL63 infections often exists in mixed viral infections, the mixed infection does not usually increase the severity of the disease. In about 71% of the cases, patients are co-infected with other respiratory viruses, such as human rhinovirus, enterovirus and parainfluenza viruses. In most patients, HCoV-NL63 infection is associated with relatively mild symptoms like fever, cough, sore throat and rhinitis. Furthermore, HCoV-NL63 is one of the main causes of croup in children.

HCoV-NL63 infects people in all ages, with the highest infection rate occurring before 5 years of age. It is estimated that 1%–10% of the population suffers annually from cold-like symptoms related to infection with HCoV-NL63. HCoV-NL63 epidemic shows a peak during the spring and summer in Hong Kong, indicating that the seasonality of HCoV-NL63 infection may not be restricted to the winter in tropical and subtropical regions. Surprisingly, studies show that HCoV-NL63 virions are exquisitely stable in liquid media and can be stored also without preservatives at ambient temperature for up to 14 days.

HCoV-HKU1

HCoV-HKU1 was isolated from an adult who had a chronic pulmonary disease in Hong Kong. The clinical symptoms of HCoV-HKU1 infection included rhinorrhea, cough, nasal congestion, fever, sputum, sore throat, chills, postnasal discharge, and tonsillar hypertrophy. About 50% of patients infected with HCoV-HKU1 experience febrile seizures, but the symptoms of HCoV-HKU1 infections in the respiratory tract are not easily separable from those caused by other respiratory viruses. HCoV-HKU1 co-circulates with respiratory syncytial virus and an epidemic usually appears prior to influenza season. In addition, HCoVs have been associated with wheezing and exacerbations of asthma.

Similar to other three mild disease causing HCoVs, HCoV-HKU1 is distributed worldwide. HCoV-HKU1 infection is relatively frequent in adults. One recent study has examined the circulation of HCoV in Israel during 2015–2016. In the winter influenza survey, 10.36% of patients were infected with the other three common HCoVs; among them, 43.43% were infected with HCoV-OC43, 44.95% with HCoV-NL63, and 11.62% with HCoV-229E. Although it is absent from the winter survey, 22.6% of the hospitalized patients were positive for HCoV-HKU1, mainly during the spring-summer period.

Diagnosis

In most cases, HCoV infections, other than SARS-CoV or MERS-CoV, are not well identified by clinical diagnosis because they cause mild, upper respiratory tract disease, and no specific therapy is available. At the same time, HCoV detection is carried out in a limited number of virological laboratories, because HCoV infections cannot be easily distinguished clinically from other causes of upper respiratory tract infections. The presence of multiple genotypes and the co-infection with other respiratory viruses, such as human rhinoviruses, enteroviruses and parainfluenza viruses, render the detection of HCoV even more challenging. HCoVs are also sometimes detected in asymptomatic patients/individuals; so that the presence of these viruses may not be etiologically related to the illness.

Electron microscopic examination of the clinical materials, although laborious, has contributed to the identification and characterization of many HCoVs. Electron microscopy has been extensively used in most of the initial studies, also because isolation of HCoVs from infected individuals and their propagation in tissue culture cells have been technically challenging. HCoV-229E has sometimes been isolated in human diploid cell lines. However, HCoV-OC43 initially required cell organ culture systems for isolation, although this virus can now be grown in tissue culture cells. HCoV-NL63 can infect monkey kidney LLC-MK2 cells or Vero cells, whereas HCoV-HKU1 has been grown only in primary human airway epithelial cells. Sometimes, these cell or organ culture techniques are labor intensive, time consuming, and relatively insensitive.

HCoVs can be detected by RT-PCR with greater sensitivity than standard culture techniques. Nasopharyngeal aspirates are the most common respiratory samples for diagnosis. PCR primers can be designed to be broadly reactive or strain specific, based on primer binding sites (usually in the viral replicase and/or the N gene). A multiplex real-time RT-PCR assay capable of detecting all four common HCoVs (HCoV-229E, OC43, NL63 and HKU1) has become the diagnostic method of choice. A simple and sensitive assay for rapid detection of human coronavirus NL63 (HCoV-NL63) has been developed by colorimetric reverse transcription loop-mediated isothermal amplification (RT-LAMP). This method employs six specially designed primers that recognize eight distinct regions of the HCoV-NL63 nucleocapsid protein gene for amplification of target sequences under isothermal conditions at 63°C for 1 h. This RT-LAMP assay can achieve a high sensitivity of 1600 RNA copies per reaction with high specificity, and is suitable for clinical settings without conventional PCR equipment.

Various serologic assays can also be used to detect HCoV infections, including complement fixation assays, neutralization assays, immunofluorescence assays (IFA), enzyme-linked immunoassays (EIA), and hemagglutination inhibition (HI) for viruses with an HE protein. Monoclonal antibodies have also been generated for direct detection of HCoV antigens in nasopharyngeal aspirates samples. Complement-fixing and enzyme-linked immunosorbent assays (ELISA) for HCoV-229E and HCoV-OC43 have been published, but are not yet available in clinical laboratories.

Serologic tests for specific antibody responses against HCoVs are mainly reserved for epidemiologic studies. Initially, virus lysates or inactivated whole viruses were used as antigens for serologic assays. More recently, cloned expressed proteins, synthesized peptides, and pseudoviruses are also used. For example, recombinant N protein of HCoV-HKU1 expressed by *E. coli* and baculovirus has been used for IgG and IgM detection of sera from patients and normal individuals, using Western blot and EIA.

Treatment

At present, there are no specific antiviral drugs available for treating HCoV infections, and therapy is therefore primarily supportive. Studies have shown that, if diagnosed early, HCoV-NL63 replication can be suppressed by treatment with PEG-IFN. The effect of various chemicals to inhibit HCoV replication has been studied in cell culture. Potential targets for anti-HCoV drugs include inhibiting virus entry, viral RdRP, and viral protease such as Mpro.

Similar to SARS-CoV and MERS-CoV, peptide-based membrane fusion inhibitors targeting the heptad repeat 1 (HR1) and HR2 of HCoV-229E S protein have been developed. They effectively inhibit S-mediated membrane fusion and the replication of

HCoV-229E in cell culture. Moreover, when administered intranasally to a mouse model, these inhibitors could widely distribute in the upper and lower respiratory tracts and maintain fusion-inhibitory activity. Some polymers, such as a cationically modified chitosan known as N-(2-hydroxypropyl)-3-trimethylammonium chitosan chloride and its derivatives, have been shown to block HCoV-NL63 entry by inhibiting the receptor binding of S protein. Viral entry can also be inhibited using serine proteases (such as camostat) and cysteine protease inhibitor (such as K11777). These compounds inhibit the critical proteolytic activation of HCoV S protein, thereby blocking the entry step of HCoV infection. Chloroquine, a compound that inhibits acidification of endosomes, can also strongly inhibit the replication of HCoV-OC43 in vitro.

The nucleoside analog ribavirin has been used to treat RSV infection, hepatitis C, and some viral hemorrhagic fevers. However, ribavirin only shows inhibition of HCoV-OC43 replication in cell culture at high concentrations ($IC_{50} \sim 10 \mu M$), which may not be applicable to humans in clinical treatment. Remdesivir (GS-5734) is a novel adenosine analog prodrug. By targeting the viral RdRP, remdesivir exhibits potent antiviral activity against HCoV-OC43 and HCoV-229E with submicromolar IC_{50} values in cell culture.

Targeting host factors required for replication may be another viable therapeutic approach against HCoV infection. A systemic investigation of HCoV-host interactome has identified cyclophilins as targets for pan-coronavirus inhibitors. In fact, knockdown of cyclophilin A (CypA) in Caco-2 cells inhibits the replication of HCoV-NL63, suggesting that CypA is required for virus replication. Notably, when cells are treated with cyclosporinA (CsA) or FK506, potent immunosuppressive drugs that bind to CypA or related immunophilins, the replication of HCoV-NL63, HCoV-229E, and SARS-CoV is inhibited. Importantly, recent studies also show that CsD Alisporivir, NIM811, and other novel non-immunosuppressive derivatives of CsA and FK506 can also strongly inhibit the growth of HCoV-NL63 in cell culture at low micromolar, non-cytotoxic concentrations. Therefore, non-immunosuppressive CypA inhibitors may be promising candidates as antivirals against HCoV infection. Interestingly, the replication of HCoV-229E and MERS-CoV is also suppressed when exogenous arachidonic acid or linoleic acid is added to the culture medium, suggesting that the regulation of host lipid metabolism could also be a common and druggable target for coronavirus infections.

Prevention

Due to the mild diseases associated with the infection by the four common HCoVs, no vaccines are currently available or being developed. Risk of infection can be reduced by good hand hygiene and coughing etiquette and avoiding close contact with infected individuals.

Acknowledgment

This work was partially supported by National Natural Science Foundation of China grant 31972660 and grant 31900135, Natural Science Foundation of Guangdong Province grant 2018A030313472, and Guangdong Province Key Laboratory of Microbial Signals and Disease Control grant MSDC-2017-05 and MSDC-2017-06.

Conflict of Interest

The authors declare no conflict of interest.

Further Reading

- Corman, V.M., Muth, D., Niemeyer, D., Drosten, C., 2018. Hosts and sources of endemic human coronaviruses. *Advances in Virus Research*. 163–188. doi:10.1016/bs.aivir.2018.01.001.
- Dubé, M., Le Coupanec, A., Wong, A.H.M., *et al.*, 2018. Axonal transport enables neuron-to-neuron propagation of human coronavirus OC43. *Journal of Virology* 92. doi:10.1128/JVI.00404–18.
- Fung, T.S., Liu, D.X., 2019. Human coronavirus: Host-pathogen interaction. *Annual Review of Microbiology* 73. doi:10.1146/annurev-micro-020518–115759.
- Milewska, A., Nowak, P., Owczarek, K., *et al.*, 2018. Entry of human coronavirus NL63 into the cell. *Journal of Virology* 92. doi:10.1128/JVI.01933–17.
- Ou, X., Guan, H., Qin, B., *et al.*, 2017. Crystal structure of the receptor binding domain of the spike glycoprotein of human betacoronavirus HKU1. *Nature Communications* 8, 15216. doi:10.1038/ncomms15216.
- Su, S., Wong, G., Shi, W., *et al.*, 2016. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends in Microbiology* 24, 490–502. doi:10.1016/j.tim.2016.03.003.
- Tvarogová, J., Madhugiri, R., Bylapudi, G., *et al.*, 2019. Identification and characterization of a human coronavirus 229E nonstructural protein 8-associated RNA 3'-terminal adenylyltransferase activity. *Journal of Virology* 93. doi:10.1128/JVI.00291–19.
- Woo, P.C.Y., Lau, S.K.P., Yip, C.C.Y., Huang, Y., Yuen, K.-Y., 2009. More and more coronaviruses: Human coronavirus HKU1. *Viruses* 1, 57–71. doi:10.3390/v1010057.
- Zeng, Z.-Q., Chen, D.-H., Tan, W.-P., *et al.*, 2018. Epidemiology and clinical characteristics of human coronaviruses OC43, 229E, NL63, and HKU1: A study of hospitalized children with acute respiratory tract infection in Guangzhou, China. *European Journal of Clinical Microbiology & Infectious Diseases* 37, 363. doi:10.1007/s10096-017-3144-z.
- Zumla, A., Chan, J.F.W., Azhar, E.I., Hui, D.S.C., Yuen, K.-Y., 2016. Coronaviruses – Drug discovery and therapeutic options. *Nature Reviews Drug Discovery* 15, 327–347. doi:10.1038/nrd.2015.37.

Relevant Websites

<https://www.cdc.gov/coronavirus/index.html>
 Coronavirus Disease 2019.
<https://medlineplus.gov/coronavirusinfections.html>
 Coronavirus Disease 2019 (COVID-19)
 MedlinePlus.
<https://www.medicalnewstoday.com/articles/256521.php>
 Coronaviruses
 Medical News Today.
https://www.who.int/csr/disease/coronavirus_infections/faq_dec12/en/
 Frequently asked questions on novel coronavirus
 Update.
<https://www.sciencedirect.com/topics/neuroscience/human-coronavirus-229e>
 Human Coronavirus 229E.
https://en.wikipedia.org/wiki/Human_coronavirus_HKU1
 Human Coronavirus HKU1.
<https://www.sciencedirect.com/topics/medicine-and-dentistry/human-coronavirus-nl63>
 Human Coronavirus NL63.
<https://www.sciencedirect.com/topics/veterinary-science-and-veterinary-medicine/human-coronavirus-oc43>
 Human Coronavirus OC43.
<https://talk.ictvonline.org/p/coronavirus-genomes>
 ICTV.

Human Cytomegalovirus (*Herpesviridae*)

Edward S Mocarski, Emory University School of Medicine, Atlanta, GA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

allogeneic MHC-mismatched cell, tissue or organ.

allograft A tissue or organ from someone other than an identical twin.

alphaherpesvirus A subfamily of herpesviruses characterized by epithelial infection with latency in sensory neurons.

antibody Secreted protein, called immunoglobulin, made by immune system plasma cells that protect by recognizing a specific antigen such as on the surface of viruses.

antigen a specific pathogen component recognized by antibody, B cell receptor or T cell receptor.

betaherpesvirus A subfamily of herpesviruses that includes the cytomegaloviruses and closely related salivary gland viruses.

capsid The protein shell part of a virion into which viral DNA is packaged.

capsomer A subunit of the capsid, also called a protomer.

cleavage/packaging A process of inserting viral DNA into a capsid and cutting viral DNA during nucleocapsid maturation.

concatemer A long repetitive head-to-tail structure, applied to replicating viral DNA prior to cleavage/packaging.

congenital Evident at birth.

core gene A subset of evolutionarily conserved common herpesvirus genes that carry out entry, replication, virion formation and maturation.

cytopathology An alteration in appearance of cells, typically during infection.

dense body A capsidless particle composed of tegument surrounded by an envelope.

differentiation state-dependent A process that only occurs in cells at a particular point in their developmental or somatic differentiation.

DNA A type of nucleic acid that encodes genes and forms the genome of some viruses and all organisms.

early gene A gene expressed prior to the initiation of viral DNA replication.

egress A process during viral maturation where the virion leaves the cell.

encapsidation The process of packaging viral DNA into nucleocapsids.

envelope A lipid bilayer membrane that surrounds the tegument of virions and dense bodies that contains viral glycoproteins responsible for binding to and entering into cells.

envelopment The process of adding an envelope to nucleocapsid.

epigenetic non-heritable.

ERGIC Endoplasmic reticulum-Golgi intermediate compartment, the cellular site of envelopment.

gammaherpesvirus A subfamily of herpesviruses characterized by oral transmission, oncogenicity and latency in lymphocytes.

genome The nucleic acid molecule that encodes all of the genes of a virus or an organism.

helicase-primase A replication enzyme that unwinds nucleic acid and synthesizes short primers that are extended by DNA polymerase.

hematopoietic Blood-forming.

herpesvirus A virus family characterized by a large DNA genome, similar-appearing enveloped virion particle and lifelong latency.

hexon A type of capsomere/protomer forming most of the capsid structure.

homologous gene products Proteins with evolutionary relatedness.

immediate early gene The first set of genes expressed during infection.

immunity Resistance to infection and disease through natural (innate) or acquired (adaptive) host defense mechanisms

immunocompetent Having the capacity for immunity.

immunocompromised Lacking the capacity for immunity.

immunodeficiency An inborn or acquired lack of immunity.

immunosuppression A reduction in immunity.

inapparent Lacking in signs or symptoms.

infectious Capable of person-to-person transmission (transmissible).

late gene The last set of genes expressed during infection.

latency Persistence in the absence of active replication with the potential to reactivate.

leukodepletion Reducing the leukocytes.

major histocompatibility complex (MHC) Tissue antigens that present antigen to T cells that vary in the population such that no two people are alike (except for identical twins).

monoclonal antibody Antibody produced by a single B cell or plasma cell recognizing a specific antigen.

mononucleosis The proliferation of mononuclear leukocytes (lymphocytes).

neurodevelopmental Related to the development of the nervous system.

nucleocapsid A DNA-containing capsid.

opportunistic A pathogen that infects and causes disease in immunocompromised hosts.

opportunistic pathogen Causing disease the immunocompromised host.

pathogen A disease-causing virus, microbe or multicellular organism.

pathogenesis	The process of causing disease.	seronegative	Lacking pathogen-specific antibody.
penton	A type of capsomere that composes capsid vertices.	seropositive	Having pathogen-specific antibody.
persistence	The act of remaining.	seroprevalence	Proportion of seropositive individuals in a population.
plaque forming unit	A single infectious virus as determined by plaque assay.	tegument	The layer of protein between the capsid and envelope in a virion.
portal	An entrance.	transfusion	The transfer of blood from one person to another.
promoter	A segment of DNA that directs transcription of a gene.	transplacental	Crossing the maternal-fetal interface during pregnancy.
reactivation	The process of transitioning from latency to active replication.	transplant	The transfer of a tissue or organ from one individual to another.
reinfection	The act of becoming infected with the same pathogen again.	tropism	The likelihood of a particular tissue or cell to support virus replication.
replication	The process of making more viral DNA and virus.	viral assembly compartment	A cytoplasmic location where virus assembly and release takes place.
restriction factor	A gene product that slows down or blocks a pathogen.	virion	A assembled, infectious virus particle, typically observed outside of a host cell.
roseolavirus	A subset of betaherpesviruses that includes HHV-6A, HHV-6B and HHV-7.		
sequence identity	Having similar nucleic acid (DNA or RNA) or amino acid (protein) order.		

Classification

Cytomegaloviruses (CMVs) are defining members of the betaherpesviruses, one of the major subfamilies of herpesviruses infecting mammalian hosts. Herpesviruses are a biologically diverse, ancient group of viruses that are characterized by large DNA genomes (from just over 100,000 kilobase pairs to nearly 300,000 kilobase pairs), as well as a common virion structure, replication scheme and the biological property of lifelong latency. Betaherpesviruses generally, and CMVs in particular are more highly host restricted than other herpesviruses. CMVs were initially referred to as salivary gland viruses because they are sporadically shed into saliva which, along with urine and breast milk, represent primary sources for transmission. Although restricted to a single host species in nature, all CMVs share a common ancestor, biological features and characteristic cytopathology that includes a nuclear site of replication and a cytoplasmic assembly compartment where final envelopment and egress occur (**Fig. 1**). Despite species specificity, amino acid sequence similarity in proteins provides molecular evidence supporting the biological relatedness of betaherpesviruses that infect various animal species (**Fig. 2**). The cytomegalovirus (CMV) of humans (HCMV), officially classified *Human herpesvirus* (HHV)5 within the genus *Cytomegalovirus*, subfamily *Betaherpesvirinae* of the family *Herpesviridae*, has the largest and most complex genome of any human herpesvirus with a potential to encode hundreds of gene products. The other human betaherpesviruses, roseolaviruses classified as HHV6A, HHV-6B and HHV-7 have smaller genomes and retain homologs of approximately half of the canonical, annotated HCMV genes (shaded genes in **Table 1**). Betaherpesviruses of mice, rats, guinea pigs and monkeys are all species-restricted viruses that are more similar to HCMV than to roseolaviruses. Because they naturally infect laboratory animals, all have been used to model aspects of HCMV infection, immunity, latency and pathogenesis. Mammalian betaherpesviruses replicate slowly in cultured cells which must be from their natural host species, remain cell-associated within the host, and show genetic relatedness to each other more so than they do to either alphaherpesviruses or gammaherpesviruses.

Virion Structure

Virions of HCMV appear similar to other herpesviruses (**Fig. 1**) despite a larger genome and a more complex coding capacity. The ~230 nm enveloped virion includes a linear double-stranded DNA genome tightly packed within a 130 nm icosahedral nucleocapsid surrounded by a tegument layer with ~30 virus-encoded phosphoproteins surrounded by a lipid bilayer envelope containing ~22 viral membrane glycoproteins (**Table 1**). The most abundant tegument constituents are phosphoprotein (pp)65 (ppUL83) and pp150 (ppUL32). Although the most structurally complex of any characterized herpesvirus, the HCMV capsid is composed of herpesvirus-conserved major capsid protein (MCP), the primary constituent of capsomers (both hexons and pentons), as well as minor capsid proteins that form intercapsomeric triplexes, the smallest capsid protein (SCP) that decorates the hexon tips, and the portal protein (PORT) comprising one penton position responsible for conducting viral DNA encapsidation. In addition to a linear molecule of DNA, RNA associates with the origin of lytic DNA replication (*oriLyt*) in virions. The most biologically important HCMV envelope constituents are glycoprotein (g)B and complexes composed of gH:gL:gO (gH/gL trimer), gH:gL:pUL128:pUL130:pUL131A (gH/gL pentamer) and gM:gN which all play crucial roles in replication. Envelope gB and gH:gL complexes are very important targets of host humoral (antibody) immunity. Numerous additional envelope glycoproteins

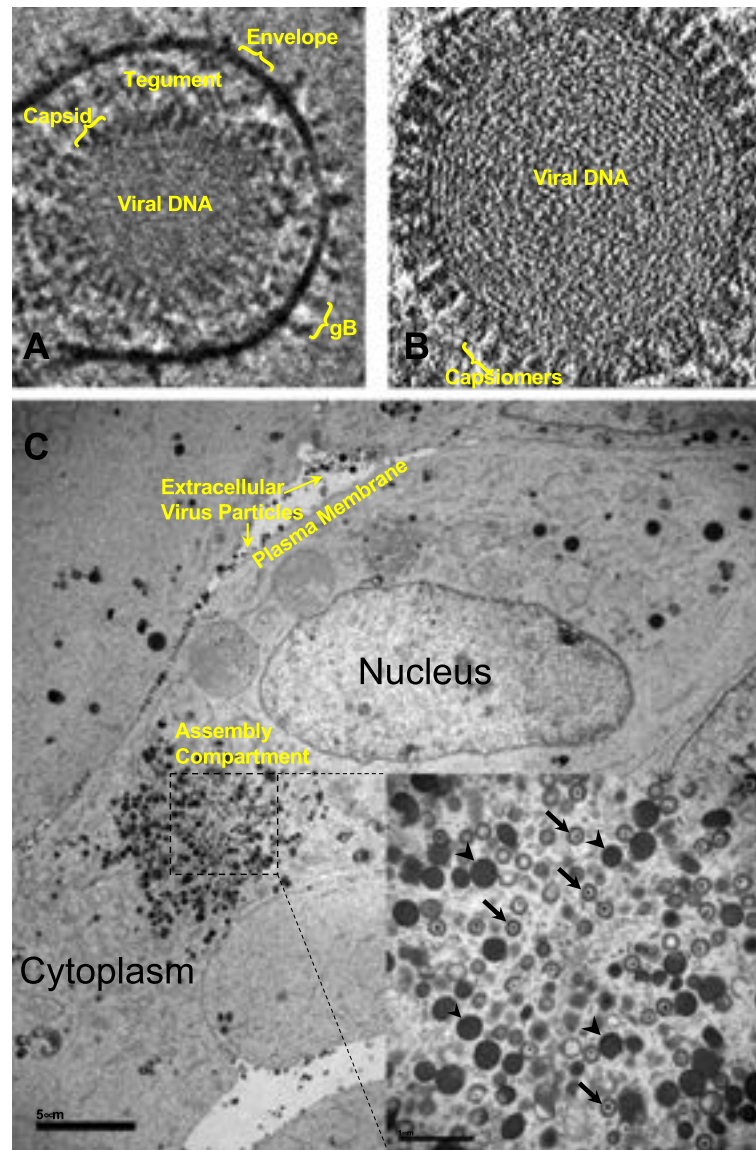


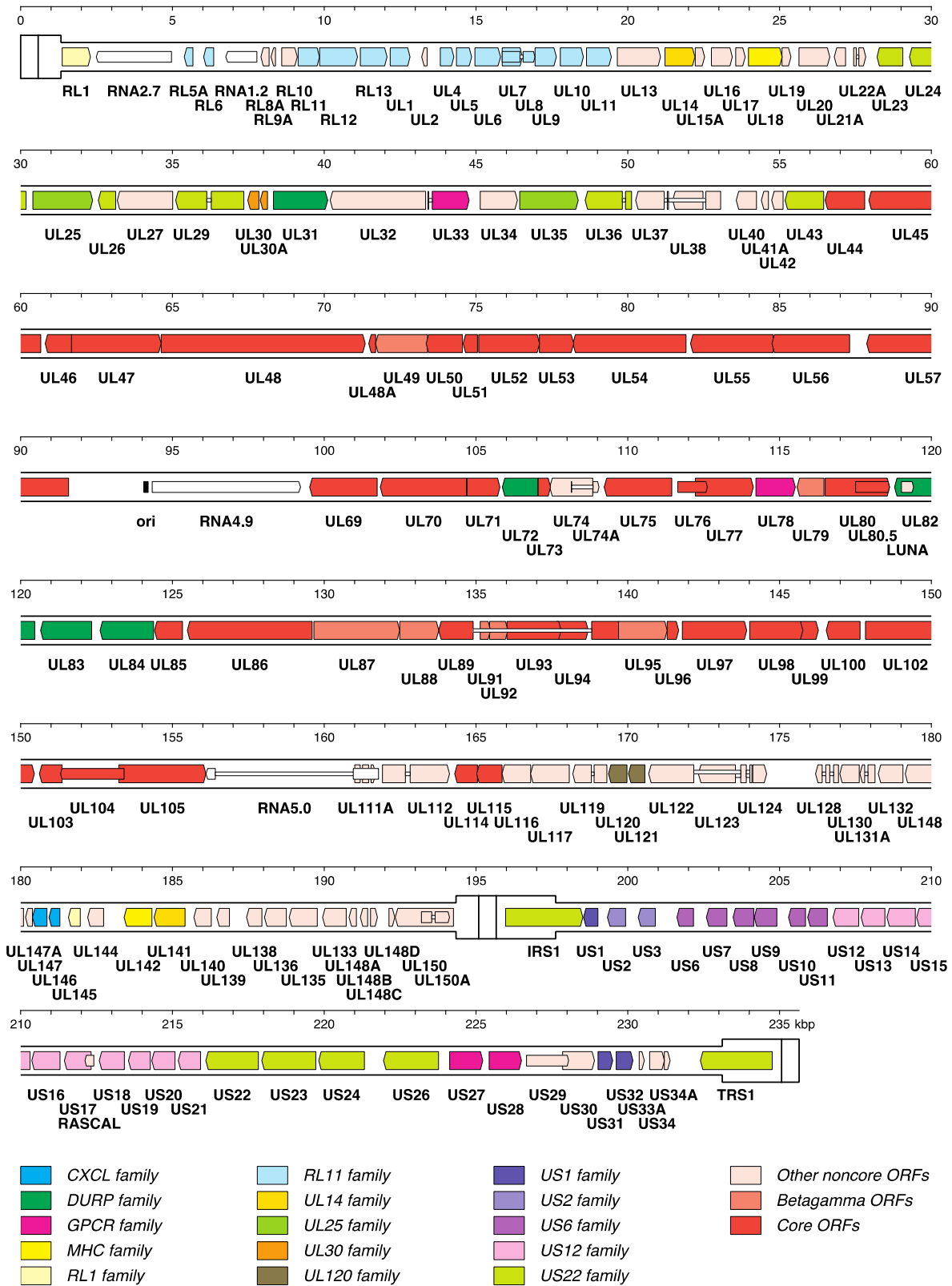
Fig. 1 Cryo and transmission electron micrograph (EM) of HCMV particles and HCMV-infected human fibroblast cells. A. Cryo-EM virion particle structure (strain AD169) showing the nucleocapsid, tegument and envelope with embedded gB (but without gH: gL pentamer) after sedimentation (note distortion of the envelope). B. Cryo-EM nucleocapsid particle from within a virion showing capsomeres and appearance of encapsidated viral DNA. C. Transmission EM of HCMV infection (multiplicity of infection = 3) at day five postinfection showing a cell with nucleus and mature virions and dense bodies within the cytoplasmic assembly compartment where final envelopment precedes release by exocytosis through the plasma membrane to the extracellular space. Released mature extracellular virus particles accumulate between adjacent cells of the monolayer. Inset shows a close-up of mature virions (arrows) and dense bodies (arrow heads) within the assembly compartment. Intracellular virus particles are contained within vesicles. Panels A and B courtesy of Hong Zhou, UCLA.

(Table 1) are dispensable for replication in cultured cells although each likely contributes to virus biology in the host by influencing cell tropism, tempering host response to infection and dictating pathogenesis as well as persistence. Capsidless particles (called dense bodies), composed solely of tegument surrounded by an envelope, are produced in numbers equivalent to virions during infection (Fig. 1). Dense bodies as well as noninfectious capsid-containing particles predominate in all virus preparations, such that particle-to-plaque forming unit (PFU) ratios of HCMV stocks range from a low of 100:1 to over 1000:1.

Genome

Structure

The HCMV genome is a single ~236,000 bp double-stranded DNA molecule composed of unique long (U_L) and unique short (U_S) regions flanked by terminal and internal repeats in an arrangement depicted $ab-U_L-b'a'c'-U_S-ca$ (alternative depiction,



TRL-UL-IRL-IRS-US-TRS). This structure is similar to human herpes simplex viruses but, with the exception of chimpanzee CMV, is not observed with nonhuman CMVs. Because recombination occurs between the inverted repeats of these herpesvirus genomes, virus particles carry one of four different sequence arrangements, or isomers, that are all infectious. The genome depicted in **Fig. 2** includes conventional annotated and canonical protein-coding genes, four long noncoding (lnc) RNA genes and an origin of replicative DNA synthesis (*oriLyt*) located between UL57 and UL69, cleavage/packaging signals (*pac1* and *pac2* within terminal *a* sequences) and the major immediate early (IE) promoter (MIEP), a transcriptional enhancer that controls transcription of IE1 (p72) and IE2 (p86) genes (located between UL124 and UL128). The genome also encodes 23 micro (mi)RNA genes not depicted in **Table 1** or **Fig. 2**. Importantly, HCMV is known to encode many more than the ~200 annotated gene products shown in such conventional depictions. Hundreds of additional candidate viral proteins have been detected experimentally. The genome organization of human roseolaviruses (HHV6A, HHV6B and HHV7) as well as other nonhuman CMVs are co-linear with the HCMV genome, although their smaller genomes (145,000–162,000 base pairs) encode only ~85 proteins where ~70 genes are colinear with HCMV, from UL23 to UL124 (shaded rows in **Table 1**). Despite these evolutionarily conserved gene products, human roseolaviruses and HCMV do not exhibit significant DNA sequence identity. All characterized betaherpesvirus genomes also have a similarly positioned *oriLyt*, MIEP-enhancer and terminal direct repeats with conserved genome-terminal cleavage/packaging signals.

Evolution

Based on the similarity to HCMV gene products and genome organization, CMVs have been estimated to have coevolved with common ancestors of mammals over the past 150 million years. The core genes that mediate the herpesvirus-conserved processes of viral DNA synthesis, capsid assembly and genome cleavage/packaging (nuclear events shown in **Fig. 3**), as well as a set of genes involved in transcription of true-late gene products, including a viral TATA binding protein (**Table 1**), likely have their evolutionary origin in a distant DNA bacteriophage ancestor. These and virion maturation events from nuclear egress through final envelopment and release require the function of herpesvirus core genes (**Table 1** and **Fig. 3**). Because of their long evolution, betaherpesviruses exhibit a greater level of divergence than their respective host species. As such, non-human primate CMVs exhibit diversity in gene content and function evident even when comparing great ape and human CMVs. This extreme diversity in biologically related, but species-specific viruses is much greater than observed in alphaherpesviruses or gammaherpesviruses. The greatest conservation in genome organization is between viruses infecting closely related host species such as HCMV and chimpanzee CMV or murine CMV and rat CMV. There even appears to be a level of diversity generated during lifelong infection with HCMV, above and beyond the presence of multiple strains in a single host. Complete genome sequences have been determined for many circulating HCMV strains, including those that have been detected directly in secretions or infected tissues. An average of >95% DNA sequence identity is observed between HCMV strains. A similar range of viral strains appear to circulate worldwide. The presence of multiple strains and within-host variation as well as the detection of multiple distinct strains within a single individual are all features that may influence disease pathogenesis but most certainly document susceptibility to reinfection. Thus, in every animal species that has its own CMV, the virus enjoys both a widespread distribution and the ability to superinfect.

Life Cycle

HCMV, and related mammalian viruses, cause systemic infection primarily targeting myelomonocytic, epithelial and endothelial cells. Like other herpesviruses, HCMV establishes chronic/persistent infection characterized by lifelong latency. Only a minor proportion of viral genes, mostly betaherpesvirus-common and herpesvirus core proteins, carry out DNA replication and cleavage/packaging as well as virion assembly, maturation and release (**Table 1**). A majority of cytomegalovirus gene products modulate host defense mechanisms, acting within an infected cell (cell autonomous) as well as systemically to alter the course of innate and adaptive immunity, subverting some pathways and enhancing others (see Pathogenesis section). CMVs target conserved host defense pathways, evidence that these master manipulators co-evolved during the radiation of mammalian species as immune mechanisms were acquired.

Host Range

HCMV only infects humans. There are no known susceptible nonhuman hosts, in nature or under experimental conditions. Furthermore, cultured human epithelial, endothelial, neuronal, myeloid and fibroblast cells are all susceptible. Chimpanzee

Fig. 2 Genetic organization and annotated gene content of wild-type HCMV, based on low passage strain Merlin. The 236 kbp HCMV genome is depicted on eight lines with each gene arranged by orientation. Inverted repeats TRL/IRL (*ab-b'a'*) and TRS/IRS (*a'c'-ca*) are shown in a thicker format than the UL₁ and US₁ components. Protein-coding regions (ORFs) are indicated by large colored arrows grouped according to the key shown at the bottom. 15 different gene families, core ORFs conserved across all mammalian and avian herpesviruses, betagamma ORFs conserved only in betaherpesviruses and gammaherpesviruses and other non-core ORFs are each depicted by a separate color. Noncoding RNAs are indicated by medium, white arrows (micro RNAs are not shown). Introns are shown as narrow white bars. Gene nomenclature is shown below the genome depiction. UL72 is both a core ORF and a member of the DURP gene family. Reproduced by permission of the Society for General Microbiology; modified by A. Davison.

Table 1 Annotated cytomegalovirus gene products: Characteristics and functions

HCMV ^{a,b}	Kinetic Class, Gene Family (Characteristics) ^c	Function (Comments) Gene Product Name ^d
RL1	LL <i>RL1</i>	
lncRNA2.7	E (most abundant noncoding RNA)	Binds mitochondrial complex I (↓ cell death) β2.7
RL5A	LL <i>RL11</i> (membrane-like gp)	
RL6	LL <i>RL11</i> (membrane-like gp)	
lncRNA1.2	E (abundant noncoding RNA)	
RL8A	LL (multiply spliced)	
RL9A	LL	
RL10	LL (envelope gp)	
RL11	LL <i>RL11</i> (envelope gp, modulation)	Binds IgG Fc (↓ NK-mediated ADCC) gp34/gpRL11
RL12	LL <i>RL11</i> (membrane gp, modulation)	IgG Fc-binding (antibody decoy) gp95/gpRL12
RL13* [#]	LL <i>RL11</i> (envelope gp, modulation)	NUDT14-binding protein (drives host restriction but allows spread in the presence of HCMV-specific antibody) gpRL13
UL1	TL <i>RL11</i> (envelope gp, modulation)	(Cell tropism determinant) gpUL1
UL2	LL (membrane gp)	
UL3	(modulation)	Disrupts nuclear PML and Cajal bodies (↓ IFN response) pUL3
UL4	E & TL <i>RL11</i> (envelope gp)	(Expression regulated by translational stalling at uORF2) gp48/gpUL4
UL5	E <i>RL11</i> (envelope gp)	gpUL5
UL6	LL <i>RL11</i> (membrane-like gp)	gpUL6
UL7	LL <i>RL11</i> (membrane gp; CD229/SLAM homolog, modulation)	Activates FLT3R (↓ inflammatory cytokine production; ↓ myeloid cell differentiation) gpUL7
UL8	LL <i>RL11</i> (CD229/SLAM homolog, modulation)	(↓ cytokine production, contains UL7-common exon) gpUL8
UL9*	LL <i>RL11</i> (membrane gp)	(↓ replication in fibroblasts)
UL10*	LL <i>RL11</i> (membrane gp)	(↓ replication and cytokine production in fibroblasts)
UL11	LL <i>RL11</i> (membrane gp, modulation)	Activates CD45, driving IL-10 (↓ T cell activation) gpUL11
UL13	IE (secreted gp)	
UL14	E <i>UL14</i> (membrane gp)	
UL15A	E (membrane-like gp)	
UL16*	E (membrane gp, modulation)	Sequesters NKG2D ligands (↓ NK cell killing) gpUL16
UL17	LL (membrane gp)	gpUL17
UL18	LL <i>UL18</i> (MHC1 homolog, modulation)	Binds LIR-1 (↓ NK activation, requires gpUL40) gpUL18
UL19	LL	
UL20	E (TCR γ chain homolog)	
UL21A*	E (regulation)	Destabilizes cyclin A (↓ anaphase promoting complex) pUL21A
UL22A (UL21.5)	LL (virion-associated secreted gp, modulation)	Binds RANTES (↓ CC chemokine response)
UL23* ^e	LL <i>US22</i> (tegument, modulation)	Disrupts Nmi recruitment of STAT1 (↓ IFN _γ response) pUL23
UL24	LL <i>US24</i> (tegument)	(Interacts with capsids via pUL46 and pUL85) pUL24
UL25	LL <i>UL25</i> , (tegument, modulation)	(↑ resistance to IFN, dependent on pUL26) pp85/pUL25
UL26	LL <i>US22</i> (tegument, modulation)	↓ IκB kinase (↑ resistance to IFN; ISGylated) pUL26
UL27	E (regulation)	Inhibits histone acetyltransferase KAT5/Tip60 (↑ p21/Cip1-dependent cell cycle arrest, requires UL97) pUL27
UL29-28	E <i>US22</i> (tegument pp, regulation)	(↓ p53 and NURD complex; stimulate IE expression) ppUL29
UL30	LL <i>UL30</i> (modulation)	(↓ nuclear PML and Cajal bodies) pUL30
UL30A	LL <i>UL30</i>	
UL31	LL <i>DURP</i> (modulation)	(↓ cGAS signaling)
UL32	LL (tegument pp, maturation)	Capsid proximal (↑ nucleocapsid stabilization) pp150
UL33x1-x2	LL <i>GPCR</i> (envelope gp, modulation)	(↓ pUS28, CCR5 and CXCR4 signaling)

Table 1 Continued

UL34	LL (regulation)	AAACACCGTA/T-specific DNA interaction (↓ US3/US9 transcription; enhances viral DNA synthesis) pUL34
UL35	LL, <i>UL25</i> (tegument pp, regulation)	Binds to ppUL82/pp71 (remodels of PML; ↑ expression of IE genes; supports tegumentation) ppUL35
UL36x1-x2	IE, <i>US22</i> (tegument, modulation)	Binds caspase-8 (inhibitor of caspase-8 apoptosis) vICA
UL37x1	IE (modulation)	Binds Bax and Bak (mitochondrial inhibitor of apoptosis; resistance to T cell cytolysis) vMIA/pUL37x1
UL37x1-x3	IE (membrane gp, regulation)	Gene regulation (includes UL37x1 amino terminus) gpUL37
UL38	E (tegument, modulation)	(↓ tuberous sclerosis complex ER stress and USP24-dependent lysosomal stress) pUL38
UL40	E (membrane gp, modulation)	Leader peptide binds HLA-E (↓ NK cell killing; ↑ NKG2C+ NK cell expansion) gpUL40
UL41A	LL (envelope gp)	(also called UL41.5)
UL42	(membrane gp, modulation)	Interacts with Nedd4 family E3 ligase Itch (↓ ubiquitination; ↓ cGAS response) gpUL42
UL43	LL <i>US22</i> (tegument)	
UL44	E & TL (core) (DNA synthesis)	POL complex (processivity subunit of DNA polymerase, PPS; recruits ppUL84 and ppUL112-113) pp50/ppUL44
UL45	E (core) (tegument, modulation)	Complexes with pUL48 (↓ RIPK1-dependent NF-κB activation; enzymatically inactive large subunit ribonucleotide reductase) RR1/pUL45
UL46	LL (core) (capsid assembly)	Capsid triplex complex (structural) ppUL46
UL47	LL (core) (tegument pp, entry and maturation)	pUL48 binding protein (↑ stabilization of capsid during uncoating and assembly) LTPbp/ppUL47
UL48	LL (core) (tegument pp, entry/maturation)	Capsid vertex complex (links capsid triplexes and terminase; ↑ stabilization of capsid) LTP/ppUL48
UL48A	LL (core) (capsid assembly)	Binds to pUL85 at hexon tips (structural, smallest capsid protein, SCP) pUL48A
UL49	E (membrane protein)	(Alternatively spliced) pUL49
UL50	LL (core) (nuclear membrane, maturation)	Nuclear egress complex (controls nucleus to cytoplasm transport of nucleocapsid; suppresses ISGylation) pUL50
UL51	LL (core) (maturation)	Terminase complex (DNA cleavage/packaging into capsids) ppUL51
UL52	LL (core) (maturation)	(Supports terminase function during packaging) pUL52
UL53	LL (core) (nuclear membrane, maturation)	Nuclear egress complex (controls nucleus to cytoplasm transport of nucleocapsid) pUL53
UL54	E (core) (DNA synthesis)	POL complex (catalytic subunit of DNA polymerase) POL/pUL54
UL55	LL (core) (envelope gp, entry)	Membrane fusion (heparan-binding; entry into cells) gB
UL56	LL (core) (maturation)	Terminase complex (binds <i>pac1</i> sequence to direct DNA cleavage/packaging into capsids) ppUL56
UL57	E (core) (DNA synthesis)	Single-stranded DNA-binding protein SSB/ppUL57
oriLyt	(DNA synthesis)	Lytic DNA replication origin (vRNA-1 and vRNA-2 bound)
lncRNA4.9	E (abundant, noncoding RNA)	Potential source of vRNA-1 & vRNA-2 associated with oriLyt; role in latency
UL69	E (core) (tegument, regulation)	Regulates RNA transcription/export (↑ nucleocytoplasmic shuttling; ↑ mRNA export and translation of unspliced mRNA; binds SPT6; SUMO ligase; cell cycle block) pUL69
UL70	E (DNA synthesis)	DNA helicase-primase complex (unwinds dsDNA) pUL70
UL71	E (core) (tegument, maturation)	Forms oligomers; controls secondary envelopment at ERGIG
UL72	E (core) <i>DURP</i> (modulation)	Enzymatically inactive deoxyuridine triphosphatase pUL72
UL73	LL (core) (envelope gp)	(Component of gM:gN complex) gN
UL74	TL (envelope gp, entry)	(Component of gH:gL:gO complex) gO
UL75	TL (core) (envelope gp, entry)	(Component of complexes that bind cellular receptors - gH:gL:gO and gH:gL:pUL128:pUL130:pUL131) gH
UL76	TL (core) (tegument, regulatory)	pUL31 interaction; nucleolar; translational impact on pUL77
UL77	LL (core) (capsid assembly)	Capsid vertex complex (links capsid triplexes and terminase; stabilizes capsid during assembly/uncoating) pUL77
UL78	E <i>GPCR</i> (membrane gp, modulation)	(↓ pUS28, CXCR4 and CCR5 by heterodimerization)

(Continued)

Table 1 Continued

UL79	LL (betagamma) (regulation)	Elongation factor supporting L gene transcription
UL80	LL (core) (capsid assembly)	Maturation protease (PR; N terminus) and capsid assembly scaffold protein (AP; C terminus) precursor pUL80
UL80.5	LL (core) (capsid assembly)	Scaffold protein precursor (capsid assembly) pAP/pUL80.5
UL82	LL <i>DURP</i> (tegument pp, regulation)	Major virion transactivator (localizes to and disrupts PML; degrades Daxx) pp71/ppUL82/upper matrix protein
UL83	LL <i>DURP</i> (tegument, modulation)	Scaffold for optimal tegumentation (↓ IRF3 activation) pp65/ppUL83/lower matrix protein
UL84	E <i>DURP</i> (DNA synthesis, regulation)	Binds IE2, ppUL112-113, ppUL44 and RNA/DNA hybrid (initiation of DNA replication) ppUL84
UL85	LL (core) (capsid assembly)	Forms capsid triplexes (minor capsid protein; pUL85)
UL86	LL (core) (major capsid protein)	Forms capsid (hexons and pentons) MCP/pUL86
UL87	LL (betagamma) (regulation)	TATA-binding protein (supports L gene transcription) pUL87
UL88	LL (core) (tegument, maturation)	(cytoplasmic egress) pUL88
UL89x1-x2	LL (core) (maturation)	Terminase complex (DNA cleavage/packaging into capsids; likely cleaves after <i>pac2</i> site) pUL89
UL90		
UL91	LL (betagamma) (regulation)	(Supports L gene transcription) pUL91
UL92	LL (betagamma) (regulation)	(Supports L gene transcription) pUL92
UL93	LL (core) capsid vertex	Capsid vertex complex (links capsid triplexes and terminase; stabilizes capsid during assembly/uncoating) pUL93
UL94	TL (core) (tegument, maturation)	Secondary envelopment (in association with ppUL99) pUL94
UL95	LL (betagamma) (tegument, maturation)	(Supports L gene transcription) pUL95
UL96	LL (tegument, maturation)	(↑ capsid stabilization during maturation and egress) pUL96
UL97	LL (core) (tegument, regulation)	Viral serine-threonine protein kinase (regulates DNA synthesis, DNA packaging and nuclear egress) VPK/pUL97
UL98	LL (core)	Deoxyribonuclease, NUC pUL98
UL99	TL (core) (myristylated tegument pp, maturation)	Secondary envelopment (in association with ppUL94) pp28/ppUL99
UL100	TL (core) (envelope gp, maturation)	(Component of gM:gN complex) gM
UL102	E (core) (DNA synthesis)	DNA helicase-primase complex (unwinds dsDNA) pUL102
UL103	LL (core) (tegument, maturation)	(↑ vesicle transport of mature cytoplasmic virions) pUL103
UL104	LL (core) (capsid assembly)	Portal for DNA encapsidation (capsid formation) pUL104
UL105	E (core) (DNA synthesis)	DNA helicase-primase complex (unwinds dsDNA) pUL105
IncRNA5.0	IE, (noncoding RNA)	
UL111A	LL (interleukin 10, modulation)	(↓ T cell response) cmvIL-10
UL112-113	E (pp, regulatory, DNA synthesis, multiply spliced)	(Transcriptional activation, initiation of DNA replication) p43 and p84 are essential
UL114	E (core) DNA synthesis accessory	Uracil-DNA glycosylase (Excises uracil from DNA and regulates DNA replication) UNG/pUL114
UL115	TL (core) (envelope gp, entry)	(Component of complexes that bind cellular receptors - gH:gL:gO and gH:gL:pUL128:pUL130:pUL131) gL
UL116	LL (envelope gp)	(Complexes with gH, replacing gL) gpUL116
UL117	LL	↓ MCM complex proteins (↑ viral replication compartment formation; ↓ cell DNA synthesis) pUL117
UL118-119	LL (envelope gp, modulation)	Binds IgG Fc (antibody decoy) gp68
UL120	LL <i>UL120</i> (membrane-like gp)	
UL121	LL <i>UL120</i> (membrane gp)	
UL122	IE (major regulatory protein)	Activates cellular and viral transcription machinery (autorepressor; cell cycle modulation) IE2-p86
UL123	IE (regulation and modulation)	Enhances IE2-dependent activation (binds HDACs; disrupts PML bodies; tethers to chromatin) IE1-p72
ORF94	E	↓ 2',5' oligoadenylate synthetase
UL124	LL (membrane gp)	
UL128[#]	LL (envelope gp, entry)	Endothelial and epithelial cell tropism determinant as a component of gH:gL:pUL128:pUL130:pUL131A complex
UL130[#]	LL (envelope gp, entry)	Endothelial and epithelial cell tropism determinant as a component of gH:gL:pUL128:pUL130:pUL131A complex
UL131A[#]	LL (envelope gp, entry)	Endothelial and epithelial cell tropism determinant as a

Table 1 Continued

		component of gH:gL:pUL128:pUL130:pUL131A complex
UL132	LL (envelope gp)	(↑ endocytosis and assembly of virions) gpUL132
UL148	LL (membrane gp)	ER localized; retains CD58 (↑ unfolded protein/ integrated stress responses via IRE1 and PERK) gpUL148
UL147A	TL (membrane-like gp)	
UL147	TL <i>UL146</i> (secreted gp)	(CXC chemokine-like) gpUL147
UL146	TL <i>UL146</i> (secreted gp)	CXC chemokine (hCXCR1/hCXCR2-specific) vCXCL1
UL145	E <i>RL1</i> (regulation)	Recruits Cullin4 E3 ligase (degrades HLTF, enhancing IE gene expression) pUL145
UL144	LL (membrane gp) HVEM homolog	Triggers coinhibitory BTLA (↓ T cell activation; high sequence variability) gpUL144
UL142	TL <i>UL18</i> (putative membrane gp) MHC-I homolog	Sequesters NKG2D ligands (↓ NK cell killing) gpUL142
UL141	LL <i>UL14</i> (membrane gp)	Sequesters TRAIL-R and restrains DNAM-1 ligands CD155 and CD112, works with US2 (↓ NK cell killing) gpUL141
UL140	LL (membrane gp)	
UL139	LL (membrane gp)	
UL138	E (membrane gp)	Golgi localized; enhances TNFR1 and EGFR levels (impairs replication in CD34+ cells) gpUL138
UL136	E (membrane gp)	Golgi localized; binds to ATP1B1 (post-entry tropism) gpUL136
UL135	E (secreted gp)	Remodels actin and degrades EGFR pUL135
UL133	E (membrane gp)	
UL148A	(membrane gp)	Degrades NKG2D ligand (↓ NK cell killing)
UL148B	(membrane-like gp)	
UL148C	(membrane-like gp)	
UL148D	E (membrane gp)	
UL150	LL (secreted gp)	(multiply spliced)
UL150A		(multiply spliced)
IRS1	IE <i>US22</i> (tegument, modulatory)	Binds dsRNA and PKR (↓ PKR pathway) pIRS1
US1	TL <i>US1</i>	
US2	E <i>US2</i> (membrane gp)	Dislocates and degrades MHC-I (ER to proteasome) gpUS2
US3	IE <i>US2</i> (membrane gp)	Binds MHC-I (↓ ER to PM transport of peptide-MHC-I) gpUS3
US6	E <i>US6</i> (membrane gp)	Binds TAP (↓ peptide transport into ER) gpUS6
US7	LL <i>US6</i> (membrane gp)	gpUS7
US8	E <i>US6</i> (membrane gp)	(Binds MHC-I) gpUS8
US9	E <i>US6</i> (membrane gp)	ER/mitochondria-localized; sequesters NKG2D ligand (↓ NK cell killing; ↓ MAVS/STING/IRF3 signaling) gpUS9
US10	E <i>US6</i> (membrane gp)	Degrades HLA-G (↓ nonclassical MHC-I) gpUS10
US11	E <i>US6</i> (membrane gp)	Dislocates and degrades MHC-I (ER to proteasome) gpUS11
US12	E <i>US12</i> (7TM-like membrane gp)	Sequesters NKG2D ligand (↓ NK cell killing) gpUS12
US13	E <i>US12</i> (7TM-like membrane gp)	Sequesters NKG2D ligand (↓ NK cell killing) gpUS13
US14	E <i>US12</i> (7TM-like membrane gp)	Sequesters PTPRM (↑ tyrosine kinase signaling) gpUS14
US15	E <i>US12</i> (7TM-like membrane gp)	Sequesters IL6ST (↓ inflammatory signaling) gpUS15
US16*	E <i>US12</i> (7TM-like membrane gp)	Promotes entry (↑ gH/gL pentamer levels) gpUS16
US17	E <i>US12</i> (7TM-like membrane gp)	Restricts virion pp65 levels (↑ IRF3 signaling) gpUS17
cORF29	E	Nuclear rim-associated HCMV protein (localizes with the p50/p53 egress complex) RASCAL/pORF29
US18	E <i>US12</i> (7TM-like membrane gp)	Degrades NKG2D and NKp30 ligands (↓ NK cell killing) gpUS18
US19*	E <i>US12</i> (7TM-like membrane-like gp)	(↑ replication in HMVECs) gpUS19
US20	E <i>US12</i> (7TM-like membrane gp)	Promiscuous degradation of NKG2D and NKp30 ligands (↓ NK cell killing; ↑ replication) gpUS20
US21	TL <i>US12</i> (7TM-like membrane gp)	Viroporin (↓ intracellular Ca ⁺⁺ ; cell death suppressor) gpUS21
US22	E <i>US22</i> (tegument)	
US23	E <i>US22</i> (tegument)	
US24	E <i>US24</i> (tegument)	(↑ early infection) pUS24

(Continued)

Table 1 Continued

US26	E US22	(↑ release) gpUS26
US27	LL <i>GPCR</i> (envelope gp)	(↑ release) gpUS27
US28	LL <i>GPCR</i> (envelope gp)	Receptor for diverse chemokines; constitutive signaling (↑ cell activation, migration and cell differentiation) gpUS28
US29	LL (membrane-like gp)	(↑ replication in RPE cells) gpUS29
US30*	LL (membrane-like protein)	
US31	LL <i>US1</i>	Interacts with IκB (↑ NFκB activation) pUS31
US32	LL <i>US1</i>	Localization to and alteration of PML pUS32
US33A		
US34	(secreted-like gp)	
US34A	(membrane-like gp)	
TRS1	IE US22 (tegument pp)	Binds dsRNA and PKR (↓ PKR pathway) pTRS1

^a Commonly annotated ("canonical") ORFs of HCMV. Certain genes ([#]) spontaneously mutate during isolation in human fibroblasts. HCMV mutants made experimentally may grow better than wild type (*), be indistinguishable from wild type (normal type) or exhibit one of two growth deficient phenotypes: failure to replicate (bold black[†]) or reduced replication (bold grey^{*}) when on human fibroblasts.

^b Betaherpesvirus-common genes include 40 herpesvirus core functions shown in shaded fields (labeled "core"). Gene characteristics and functional information are provided (a blank indicates the absence of information).

^c Kinetic information is based on detection of the gene product, either a protein or long noncoding (lnc)RNA.

Abbreviations: phosphoprotein, pp; glycoprotein, gp; g-protein-coupled receptor, GPCR; seven transmembrane-spanning, 7TM; immediate early, IE; early, E; leaky late, LL; true late, TL.

^d Functional information may be speculative and based on homologs in other herpesviruses if no direct evidence has been presented for HCMV.

^e Shaded fields identify the betaherpesvirus-common genes where homologs are found in the human roseolaviruses.

fibroblasts support HCMV replication, but cells from other species do not. Except for certain glioblastoma cell lines, undifferentiated or transformed cells are not susceptible to HCMV. Nonpermissive human cell lines as well as many nonhuman cells support viral entry but do not support replication owing to the presence of host restriction factors. Clinical isolates of HCMV show considerable sequence variability and sometimes represent mixtures of strains arising as a result of multiple infection or reinfection. Clinical strains undergo adaptation to cultured fibroblasts by losing expression of specific viral genes, mediated through point mutation, deletion and inversion, thereby confounding designation as well as experimental evaluation of wild type virus (Table 1).

HCMV naturally initiates infection in mucosal epithelium, disseminates within myelomonocytic leukocytes that carry the virus in the bloodstream to other locations, such as endothelial and epithelial sites, in the host. In immunocompetent individuals, virus characteristically disseminates to and replicates in ductal epithelial cells of the salivary glands and kidney tubules, allowing shedding in saliva and urine that may continue for months to years, particularly in young children. Host innate and adaptive immune mechanisms dictate the level of systemic dissemination, with a much broader tissue distribution and greater levels in infected immunocompromised individuals or newborns compared to adults with fully mature immune systems. Levels of virus infection and extent of tissue involvement dictated by host immune status determine the risk for disease following infection. Tissue samples from immunocompromised individuals with acute HCMV disease affirm that HCMV replicates naturally in epithelial, fibroblast, endothelial, macrophage, neuronal and dendritic cells. Viral antigens are readily detected in infected tissues as well as in peripheral blood neutrophils of immunocompromised individuals with a risk of HCMV disease using monoclonal antibody reagents. One of the most important, yet poorly understood aspects of host cell range, is intrauterine infection and transplacental transmission where virus gains access to the developing fetus and results in congenital disease. HCMV infection during pregnancy starts with systemic dissemination to the maternal side of the placenta and continues with access to the maternal-fetal interface but little is known about the cells involved in this medically significant process. When transmission occurs during primary infection of a pregnant woman, there is a ~40% chance that the fetus will become infected with about 10% of infected newborns showing signs of neurodevelopmental damage and disease at birth. In the case of transmission during recurrent infection, the chance of infecting the fetus has been estimated to be <1% with a similar proportion (10% of infected newborns) showing neurodevelopmental damage at birth. HCMV reactivates and is shed from ductal epithelium in mother's milk, an important route of virus transmission in the human population worldwide.

Entry into cells depends on specific envelope glycoproteins that recognize distinct cell surface receptors and lead to the fusion with plasma membrane and deposition of the HCMV nucleocapsid into the cytoplasm of the cell (Fig. 3). The cellular receptors mediating HCMV entry vary with cell type but include human platelet-derived growth factor receptor (PDGFR), epithelial growth factor receptor (EGFR), neuropilin-2, OR1411 and various integrins. PDGFR binds to trimeric envelope (gH:gL:gO) complex, mediating efficient attachment to and entry into fibroblasts and some other cell types; whereas, neuropilin-2 and OR1411 are cell surface receptors for a pentameric complex (gH:gL:pUL128:pUL130:pUL131A), that mediate attachment to and entry into

Attachment & Penetration

gB (UL55)
 gH:gL (UL75:UL115)
 gH:gL:gO (UL75:UL115:UL74)
 gH:gL:pUL128:pUL130:pUL131

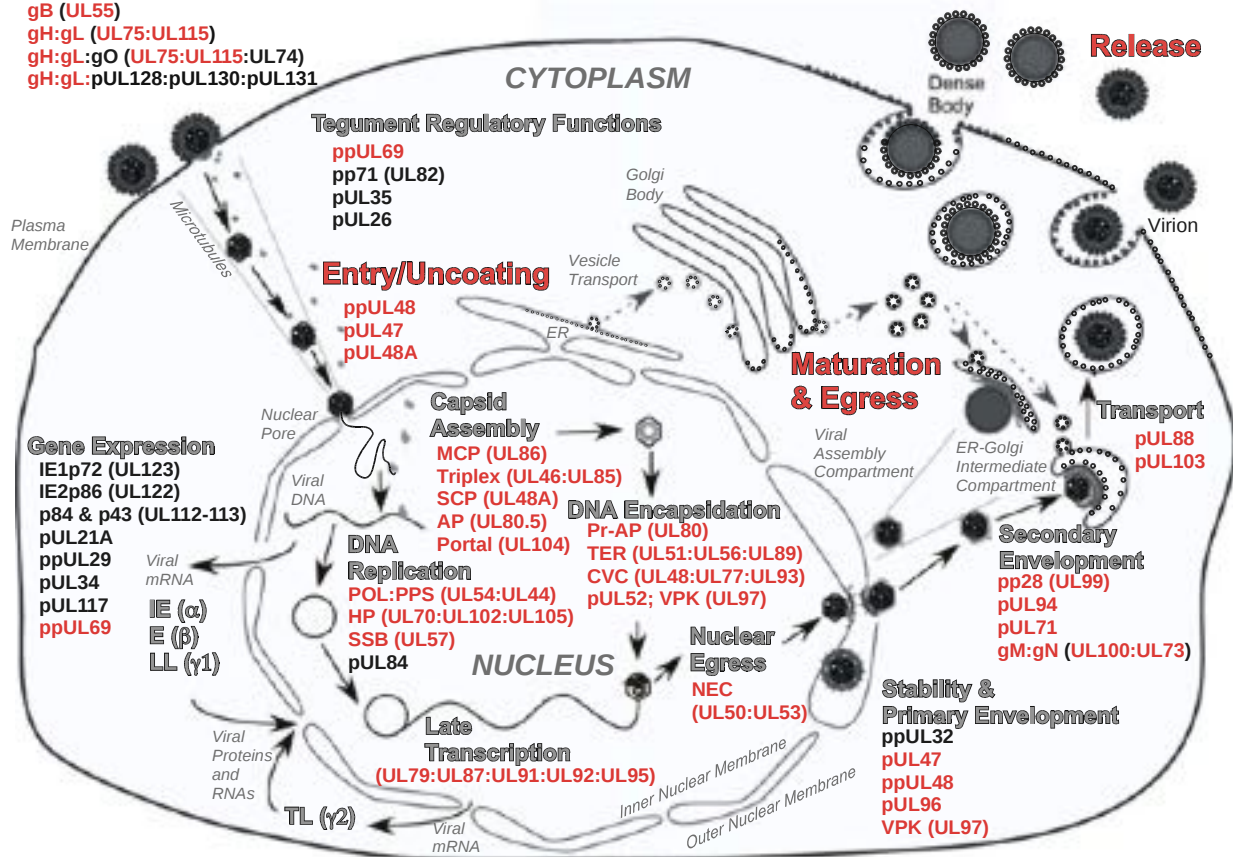


Fig. 3 Summary of the HCMV replication pathway with focus on herpesvirus core (red) and betaherpesvirus-conserved (black) gene products playing known or predicted roles in the viral replication cycle from entry (upper left) to release of progeny virions and dense bodies (upper right). Major steps in the productive replication cycle are indicated in large red outlined font and major activities are indicated in large gray outlined font, with black arrows indicating the procession and individual functions identified by abbreviated names. The entry pathway shown employs direct fusion at the plasma membrane (attachment and penetration) although an endocytic pathway is important in some cell types. Entry requires gB, along with gH:gL:gO and/or gH:gL:pUL129:pUL130:pUL131A, depending on the cell type. Virion tegument proteins (pUL69, pUL82, pUL35 and pUL26) are involved in regulation of the initial steps in replication. Nucleocapsid-associated pUL48, pUL47 and pUL48A are likely to control transport on microtubules, docking at nuclear pores and release of virion DNA from the capsid into the nucleus. Host cell RNA pol II is responsible for transcription of viral and host cell genes. Gene expression is regulated by products of IE genes (IE1-p72, IE2-p86) working together with E genes (UL34, UL112-UL113 and UL69) and LL genes (UL21A, UL29, UL117), together with a late gene transcription complex that includes UL79, UL87, UL91, UL92 and UL95 gene products, all modifying the activity of RNA pol II. Viral DNA replication, which also takes place in the nucleus, depends on core proteins (POL: PPS, HP and SSB) as well as three betaherpesvirus-specific gene products (UL84, UL34 and UL112–113). Capsid assembly incorporates proteins (MCP, triplex and SCP) around an AP scaffold initiated by the portal protein, as occurs in other herpesviruses. Pre-formed capsids translocate to sites of DNA replication where maturational protease Pr-AP works in conjunction with a terminase complex (UL51, UL56 and UL89) and capsid vertex complex (UL48, UL77 and UL93), together with UL97 and UL52 gene products to cleave and package viral DNA. Nuclear egress and primary envelopment at the inner nuclear membrane is under the control of UL50 and UL53 gene products and capsid stability is provided by UL32, UL48, UL47 and UL96 gene products together with UL97 protein kinase. De-envelopment at the outer nuclear membrane, translocation on microtubules to the viral assembly compartment in the cytoplasm is followed by secondary (final) envelopment controlled UL99, UL94 and UL71 gene products with the UL97 protein kinase at specific cytoplasmic membranes. Virion envelope glycoproteins follow an independent vesicle transport pathway from the ER via the Golgi body to site of final envelopment (ER-Golgi intermediate compartment) in the cytoplasm (dashed gray arrows). HCMV modifies the cellular exocytic transport pathway from sites of secondary envelopment to release via the action of the UL88 and UL103 gene products.

epithelial, endothelial and certain myelomonocytic cell types. Envelope gB is responsible for fusing the viral envelope with cellular membranes and delivering the nucleocapsid to the cytoplasm during entry regardless of cell type (Fig. 3).

Although crucial to cell tropism within the host, functions that control attachment and entry do not dictate the restricted host range characteristic of CMVs. For HCMV, this is primarily a consequence of a block to early post-entry step such that nonpermissive cells from distant host species (for example rodent) allow all steps through initial (IE) HCMV gene expression but do not support viral DNA synthesis or later phases necessary for production of progeny virus. Current evidence implicates early activation of cell death together with other host factors acting prior to viral DNA replication in restricting CMV replication within non-natural host species.

Replication Steps

The following categorical steps in replication are common across herpesviruses (depicted in [Fig. 3](#)): (1) virion binding to cell surface receptors through gH:gL complexes; (2) fusion mediated by virion envelope gB targeting the plasma membrane, either directly or after endocytosis into vesicles, and resulting in the release of nucleocapsids into the cytoplasm; (3) nucleocapsid association with microtubules with translocation to nuclear pores; (4) deposition of the viral genome out of the capsid into the nucleus; and, (5) coordinately regulated RNA pol II transcription that produces immediate early (IE), early (E) and late (L) gene products ([Table 1](#)). As with all herpesviruses, different temporal classes of genes are interspersed across the HCMV genome. For example, the 11 IE genes map to diverse locations in the U_L, U_S and IR_S/TR_S regions of the viral genome ([Table 1](#) and [Fig. 2](#)). Replication proceeds over several days before progeny virus are detected within cells or released extracellularly. Host RNA pol II transcription machinery becomes modified by virus-encoded proteins within the cell nucleus and synthesizes messenger (m)RNA as well as three types of noncoding RNAs, long noncoding RNAs, micro RNAs and ori_{Lyt}-associated vRNAs. Translation of virus-encoded mRNAs into proteins is carried out by host cell polyribosomes in the cytoplasm where active translation of several hundred uncharacterized proteins that expand the coding potential of this virus well beyond annotated genes ([Fig. 2](#)). Initially, the HCMV MIEP-enhancer provides regulatory control over the decision whether to replicate, and epigenetic regulation dictates outcome in a cell type- and differentiation state-dependent manner that is influenced by viral gene products as well as the host cell immune environment. The MIEP generally confers very strong constitutive transcriptional activity, even in nonpermissive cells, enabling this transcriptional element to be widely adapted for experimental and commercial production of recombinant proteins. Epigenetic restriction of viral transcription is interrupted by several viral modulatory proteins, both from the input tegument as well as gene products produced during infection. These balance replication, persistence, latency and reactivation in a cell-type-dependent fashion.

Tegument proteins regulate viral gene expression and modulate the host response from earliest post-entry steps through late stages of replication. Input virion tegument proteins control the release of virion DNA into the nucleus, suppress intrinsic host defense mechanisms that would otherwise restrict viral replication and mediate activation of IE gene transcription. Later, tegument proteins modulate the cell-autonomous host response and support key steps in envelopment and egress from the nucleus and cytoplasm. During the initial steps of infection, the MIEP-enhancer is transactivated by virion tegument protein complex, enhancing RNA pol II transcription of IE genes. The IE gene products activate E and L gene transcription, including E and L products from the MIE regions itself, while suppressing the host cell response to infection, innate immune cell death and cytokine-mediated immunity.

Expression of ~52 E genes requires functional IE1 and IE2 proteins. The replication cycle of HCMV is much slower than herpes simplex virus, requiring 24–36 hrs for the progression from E to L phase, 48–72 h until progeny start to accumulate, and maximum levels of virus release continuing from 96 hrs to a week post infection. Viral DNA synthesis initiates through events that depend upon a subset of regulatory IE, and E proteins ([Table 1](#)). Following initiation, the viral DNA polymerase complex assembles at replication forks together with single stranded DNA binding protein and helicase-primase complex that together produce long double-stranded DNA concatemers that are both a template for late transcription and a substrate for genome cleavage/packaging machinery ([Fig. 3](#)). All steps in replication depend on protein phosphorylation by a viral protein kinase together with host cell cycle kinases. At least two categories of L genes are encoded based on patterns of expression and dependence on viral DNA synthesis, known as leaky late (LL) and true late (TL). A late transcription complex that is also employed by gammaherpesviruses modifies the behavior of host RNA pol II to facilitate transcription of TL viral genes, likely in close collaboration with viral DNA replication machinery.

HCMV capsid formation and maturation (capsid assembly, maturation and egress from the nucleus) are largely carried out by herpesvirus core genes ([Table 1](#)). Non-core gene products provide a means to evade and engage the host response to infection for the benefit of the virus. Viral DNA synthesis and capsid assembly proceed independently in the nucleus but converge at the point where concatemeric DNA is cleaved and packaged, or encapsidated, into preformed capsids ([Fig. 3](#)), filling each progeny capsid with a unit-length genome. Seven herpesvirus core encapsidation proteins process a concatemeric DNA template by recognizing sequence-specific cleavage/packaging (*pac*) sites located near genomic termini once a genome-length molecule has been inserted into a capsid. The resulting filled nucleocapsids become stabilized by the addition of capsid-proximal tegument proteins, and then proceed through the inner nuclear membrane under control of a nuclear egress complex. Transfer of the nucleocapsid into the cytoplasm most likely occurs by initial addition of a lipid bilayer envelope at the inner nuclear membrane and removal of this envelope through fusion with the outer nuclear membrane. The tegumented capsid gains a final (second) lipid bilayer envelope at ER-Golgi intermediate compartment (ERGIC) membranes decorated by virus-encoded glycoproteins within modified region of the cytoplasm called the viral assembly compartment (VAC), an area where viral egress also occurs ([Fig. 1](#)). Dense bodies form when tegument becomes enveloped in the cytoplasm and proceeds through cytoplasmic maturation steps along with virions. Thus, mature virions, noninfectious capsid-containing and capsidless dense body particles all accumulate at sites of final envelopment within the VAC prior to release from cells. Vesicle transport (exocytosis) machinery mediates the translocation and release of virus particles. Progeny virions are composed of ~100 different virus-encoded proteins enclosing a single viral DNA molecule. The maturation process results in an excess of noninfectious and defective particles, such that only 0.1%–1% of progeny particles are able to initiate a new infection. Once maturation starts, infected fibroblasts remain viable and continue to release virus for days. A final burst of virus release accompanies the eventual death of infected cells after about five to seven days.

Epidemiology

Patterns of Infection

HCMV is a ubiquitous virus with worldwide prevalence ranging from ~60% to greater than 90%. Overall 95% of the world population becomes infected before old age. Contact with infected bodily secretions drives transmission without involvement of any intermediate host or any seasonal prevalence. Infection is typically subclinical in immunocompetent individuals. Transmission is more efficient in higher density living conditions, with larger family size and in settings of reduced hygiene. Transmission is reduced by simple measures such as avoiding salivary transfer hand washing. On a population level, HCMV is only rarely acquired congenitally (~1%). Breast milk and contact with other bodily secretions account for the majority of transmissions throughout life. Thus, infection posing a significant risk of disease when acquired in utero and is a certain risk when primary infection occurs during pregnancy. Developing areas (Eastern Europe, Africa, Asia and South America) exhibit higher rates of transmission at younger ages, than more developed areas (North America, Australia and Western Europe). Infants and young children infected with HCMV shed persistently in saliva and urine and represent important sources of new infections. In such settings, avoidance of direct contact with infected secretions and good hygiene practices reduces transmission from children to adults as well as between adults. During lifelong latency, virus may reactivate and be shed sporadically in saliva, urine and genital secretions. Blood and tissues from infected, HCMV seropositive individuals transmit virus to recipients of blood transfusions and transplants. Use of prescreened blood products from HCMV seronegative donors and leukodepletion of blood from seropositive donors has prevented transfusion-associated infections in many areas of the world.

In US-specific surveys, overall HCMV seroprevalence has remained constant at ~60% for decades, similar to other developed areas of the world. HCMV is acquired by 40%–50% of the US population by midlife, increasing to 90% by old age. Transmission patterns are influenced by host innate and adaptive immunity. Immunoprophylaxis and, particularly, vaccination would be expected to interrupt person-to-person transmission of this virus. Thus, transplacental transmission occurs much more frequently in pregnant women with primary infection than in HCMV-seropositive women with long-term infection and pre-existing immunity. The importance of adaptive cellular immunity against HCMV in preventing active infection is supported by the clinical experience in immunocompromised transplant recipients as well as AIDS. In solid organ transplantation, the HCMV seronegative recipient of an organ from a donor with long-term infection is most at risk of infection and disease. In this setting, the organ introduces the virus into a recipient. In allogeneic hematopoietic stem cell transplantation (HSCT), performed with either mobilized blood or bone marrow, the HCMV seropositive recipient of hematopoietic stem cells from a HCMV naïve host runs the greatest risk of HCMV disease. In this setting, the transplanted HCMV-naïve immune system cannot suppress virus. The nature of immune control in settings of congenital disease remains poorly understood. Whereas exogenously acquired primary infection is the sole source of virus in a HCMV-naïve woman, reactivation of latent virus, mixed infection with additional strains of virus have been observed in long-term seropositive women. All of these sources may contribute to intrauterine transplacental transmission, congenital infection and consequent disease in progeny of seropositive women.

Serologic assays provide an accurate indication of lifelong infection with HCMV, but do not provide insights into viral shedding levels or susceptibility to reinfection. Levels of antiviral antibody and cellular immunity rarely correlate in seropositive individuals. Although strain differences are recognized using virus antigen-specific monoclonal antibodies that have facilitated characterization of mixed infections, distinct HCMV serotypes are not recognized. Together with various molecular genetic approaches that assess within host viral genome sequence variation, these approaches have provided evidence that different viral strains may predominate within a single individual at different times and that such mixed infections can arise from sequential reinfection.

Virus Propagation

HCMV is most readily isolated from saliva or urine, sites of virus replication in ductal epithelial cells of salivary glands and kidney. Virus may be isolated on cultured susceptible cells from the original host species. Fibroblast cultures are most commonly used for virus isolation. The ability to replicate in epithelial and endothelial cells is lost when virus is propagated in fibroblasts because components of the gH/gL pentamer complex (UL128, UL130 and UL131) accumulate mutations. Primary or secondary cultures of epithelial, endothelial, macrophage and dendritic cells sustain gH/gL pentamer function. Additional mutations that disrupt genes that are either unnecessary or deleterious to virus replication occur. Fully wild type HCMV is very difficult to maintain in a culture. Several genes must be modified by conditional mutation. Variants of common laboratory strains, typically referred to by a single name (e.g., AD169 or Towne) have arisen as these strains have been serially propagated in different laboratories around the world. This makes comparisons between laboratories somewhat challenging even when using a strain that carries a common name. Clonal virus strains maintained as bacmids in bacteria currently predominate for experimental work; however, these strategies also generate unintended mutants and variants. Thus, clonal virus stocks should be prepared for which the full genome sequence has been determined. The difficulty of working with clinical isolates, the species specificity of replication and the rapid generation of variants when HCMV DNA is transfected into cells (or is propagated in cell culture) are well-recognized experimental confounders.

HCMV is easiest to isolate from saliva or urine of infected individuals, virus may also be isolated from semen, cervical secretions, blood cells and breast milk, particularly during primary infection. Infectious virus or viral DNA is only rarely shed in HCMV seropositive adults with clinically silent long term latent/persistent infection. DNA PCR methods have been most successful for the detection of recurrent shedding, as a surrogate for virus in saliva, urine, milk, blood, tissues and other secretions. These

methods have revealed that children with congenital or perinatal infection naturally shed virus for prolonged periods lasting up to years, that HCMV-seropositive adults shed less than 5% of the time and that shedding increases following pregnancy.

Clinical Features

HCMV infection is generally inapparent. Infectious mononucleosis can be a rare consequence but is typically less severe than that caused by Epstein-Barr virus. Women of childbearing age often encounter HCMV primary infection without showing specific symptoms. Such asymptomatic infections are the major way this virus is transmitted to offspring during pregnancy. As a consequence of viral transmission during primary or recurrent infection, HCMV remains the single most important infectious cause of congenital disease in newborns, as significant as major genetic diseases such as Down syndrome and more frequent than other genetic congenital diseases. HCMV is a classic opportunist such that infection in immunocompromised hosts results in systemic HCMV disease affecting one or more organs, referred to as CMV syndrome. HCMV has been known to unveil immunodeficiency in individuals with a previously unrecognized combined or T cell-specific immune deficit, and as such is one of the most prominent AIDS-defining infections.

As a ubiquitous herpesvirus that causes systemic infection, lifelong infection with HCMV correlates with chronic disease, immune defects and frailty, particularly in the elderly. There have been postulated contributions to atherosclerotic vascular disease, autoimmune vasculitis, systemic lupus erythematosus, Sjogren's syndrome, periodontal disease, cancer, immunosenescence and risk of some specific diseases such as Crohn's disease and glioblastoma multiforme. These associations between a common virus and chronic diseases require further scrutiny.

Congenital Disease

Congenital disease represents the hallmark of HCMV pathogenesis. This disease follows transplacental transmission to the fetus during pregnancy. The highest incidence of mother-to-child transmission occurs from the first trimester through the gestational midpoint. Neurodevelopmental disease consequences such as hearing and eyesight compromise are most common. Broader neurological symptoms recognized at birth or during childhood are less common, but important. HCMV congenital disease remains a well-recognized, though medically underappreciated public health concern. HCMV is also a substantial cause of spontaneous abortion and still birth. The at-risk population is difficult to identify because infection in the pregnant woman is asymptomatic. HCMV congenital infection affects between 0.2% and 2% of pregnancies, where transmission remains negatively correlated with socioeconomic conditions. Nevertheless, congenital disease affects roughly 20% of HCMV-infected newborns regardless of socioeconomic status, with about half of babies with disease evident at birth and half developing in the newborn over the first years of life. Left untreated, disease generally worsens. Mild, but developmentally and medically significant congenital disease is often overlooked. Demographic characteristics contribute to transmission patterns, with younger (teenage) females more likely to transmit during pregnancy than women in their twenties or thirties. Severe congenital disease is relatively uncommon but involves extensive neurological symptoms such as retinitis, microcephaly, seizures and neurodevelopmental disabilities as well as damage to the reticuloendothelial system (petechial rash, hepatosplenomegaly, jaundice, extramedullary hematopoiesis) and low birth weight. This severe disease syndrome is known as CMV inclusion disease (CID). Antiviral drugs such as ganciclovir are used to treat CID. In addition, low birthweight, premature infants transfused with blood from an HCMV seropositive donor or breast fed by an actively shedding mother risk developing systemic disease (sepsis, hepatitis, pneumonitis) that is clinically distinct from CID. By the 1970s, the spectrum of congenital CMV disease was well-characterized, although medical awareness has not improved over the intervening decades primarily because of the belief that little could be done except to address the consequent disability. It is now clear that early intervention to improve sensorineural perception improves quality of life outcomes. Except for neurological disease and multiorgan CID, intervention with antiviral drugs, either in pregnancy or in newborns, has not yielded of significant benefit. Furthermore, the substantial contribution of recurrent maternal infection to transmission worldwide and the protean nature of neurodevelopmental damage has resulted in testing for HCMV infection at birth in some countries. Overall, the disease consequences of congenital CMV can be addressed through enhanced testing, increased clinical recognition and appropriate directed therapies.

It is now very clear that, in addition to infection of HCMV-naïve women during pregnancy, long-term maternal infection also makes a substantial contribution to patterns of intrauterine infection, transplacental transmission and congenital disease. Transmission in women known to already be infected with HCMV is recognized to occur worldwide, predominating even in developed countries such as the US. Natural levels of acquired antiviral immunity provide a level of protection that apparently reduces, but that is insufficient to completely prevent transplacental transmission. The recognized benefit of natural maternal HCMV-specific immune memory has been difficult to quantify, although antiviral antibodies and T cell immunity likely impede reinfection, dissemination to the maternal-fetal interface and transplacental transmission to the fetus. Transplacental transmission occurs with a frequency less than 2% in women with long-term infection. In contrast, transmission occurs in upwards of ~40% in women with primary infection during pregnancy. Thus, prior natural exposure to HCMV provides immunity that significantly reduces risk of congenital disease.

While most newborns with congenital CMV infection survive, CID may be lethal and often results in permanent damage even when the infant is treated with antiviral therapy. Given estimated transmission frequencies of ~0.2 to 2% of live births, HCMV causes congenital disease more frequently than any other pathogen. It is estimated that there are >1 million HCMV-infected infants born each year based on a total of 150 million births worldwide. HCMV-infected young children, regardless of whether

infected in utero or postnatally, shed virus in saliva and urine for extended periods. Thus, children represent a silent, but major reservoir of virus transmitted to parents as well as other children. There is still a remarkable need to understand pathogenesis and immune control of congenital disease, including the processes of transplacental transmission and dissemination within the developing fetus.

Premature infants who acquire HCMV infection during birth or post-partum (from breast milk, horizontal transmission or blood transfusion) are susceptible to systemic disease distinct from CID. Risk factors include very low birth weight, exposure to multiple units of blood and birth to a seronegative mother who has been recently infected but has not yet started producing antibodies. Transmission of HCMV from mother's milk to newborn may be prevented by brief heat treatment (Holder Pasteurization) that inactivates infectivity but preserves nutritional benefits. Transfusion-associated HCMV disease in premature newborns is controlled by the use of blood products from seronegative (or leukocyte-depleted) donors.

Immunocompromised Host

HCMV has become one of the most common and medically important opportunistic pathogens as allograft transplantation has become more common and as incidence of HIV infection has increased worldwide. HCMV is a highly significant pathogen in immunocompromised hosts receiving immunosuppressant therapy following solid organ transplantation, during the period of engraftment following allogeneic HSCT, in the course of cancer immunotherapy and anytime cellular immune function becomes compromised. HCMV is a very common complication during HIV/AIDS. Immunodeficient or immunocompromised individuals support higher levels of HCMV replication than immunocompetent hosts, predisposing to disease.

Clinical manifestations of disease in solid organ transplant patients include systemic, febrile illness (CMV syndrome, with malaise, arthralgia and rash; neutropenia, thrombocytopenia and elevated liver enzymes), as well as a direct or indirect impact on specific organs. Disease risk has been reduced but not eliminated by preemptive therapy or prophylaxis with antiviral drugs. Clinical detection of viral antigens or nucleic acids without the need for virus culture increased the opportunity for preemptive intervention to prevent acute HCMV disease in the first few weeks following transplantation. Preemptive therapy has also benefitted allogeneic HSCT where late onset disease and indirect consequences of infection have subsequently emerged. Threshold detection nucleic acids (usually PCR detection of viral DNA in plasma or leukocytes) in blood is used to predict impending disease. Solid organ transplant recipients may develop HCMV pneumonitis, gastrointestinal lesions, hepatitis, retinitis, pancreatitis, myocarditis, and, in rare circumstances, encephalitis or peripheral neuropathy, whereas HSCT recipients typically develop HCMV pneumonitis and gastrointestinal disease. Establishing an etiologic role for HCMV requires detection of virus in affected tissue or bronchoalveolar lavage or quantification of HCMV in blood or plasma. The principle indirect effects of HCMV, particularly during solid organ transplantation, include allograft rejection, vascular disease complications and increased risk of opportunistic fungal and bacterial infections. Primary infection is most likely to lead to disease when a naïve recipient receives an organ from a HCMV seropositive donor, although the particular organ, level of major histocompatibility complex (MHC) match and immunosuppression regimen strongly influence outcome.

HIV-associated destruction of CD4 T lymphocyte during AIDS removes critical antiviral immune control, resulting in retinal, gastrointestinal, pulmonary or multi-organ HCMV disease. Disease may follow primary infection, reactivation of latent virus, and reinfection with additional strains of virus. Disease pathogenesis is complicated because HCMV shedding and viremia are common in the face of impaired cellular immunity. Clinically inapparent HCMV infection may predominate even in the face of antiviral intervention and may only progress to disease as T cell-mediated immune surveillance becomes severely compromised, such as occurs in AIDS patients with declining CD4 T cell counts.

Pathogenesis

HCMV infection and systemic dissemination precedes the development of disease; however, disease severity is not necessarily proportional to levels of virus, viral antigens or viral nucleic acid appearing in blood or tissues. Levels of virus shed in bodily secretions does not correlate with disease risk, whether considering transplacental transmission, congenital infection, transplantation or AIDS. With HCMV, direct viral damage to tissues may result from viral replication or from a combination of replication together with host inflammatory and immune response mechanisms. It is reasonable to expect that the pathogenic signature of HCMV in most settings is due to a combination of both. Direct tissue damage in severely immunocompromised patients is a hallmark of disease and develops as virus levels rise. The degree of immunosuppression matters such that HCMV disease risk increases with T cell depletion in HSCT but may be moderated by adoptive transfer of HCMV-specific T cells, by backing off therapeutic immunosuppressants in solid organ transplantation, or by the control of an immunosuppressive pathogen as seen with the rescue of HCMV immunity following highly active antiretroviral therapy (HAART) in AIDS. In such settings, the levels of HCMV DNA in the bloodstream provide an accurate indication of disease risk and is the basis of monitoring for pre-emptive therapy. Antiviral CD8 T cells are the primary immune mechanism for bringing HCMV under control in transplant settings. Shortcomings in this arm of immunity underlie late onset disease.

In AIDS, antiviral CD4 T cells provide crucial support of overall immunity. HAART suppresses HIV replication in these T cells, reconstitutes or preserves cellular immune function, and thereby eliminates risk of HCMV retinitis, gastrointestinal disease and other problems associated with AIDS progression. This reconstitutes immunity and leads to a risk of uveitis where inflammatory disease is triggered by HCMV antigens in the affected tissues.

It is settings of less dramatic immune compromise, including pregnancy, where transplacental transmission results in congenital disease in offspring, that the host response to infection is suspected of making contributions to pathogenesis. Pregnancy is a state of relative maternal immune compromise, resulting from maternal tolerance of fetal antigens in a complex and changing immune environment. Transmission and intrauterine replication may be enhanced by inflammatory damage. Fetal immune response patterns, though not fully developed, may contribute to virus control as well as risk of sensorineural damage in congenital disease.

Transmission

HCMV transmission requires contact with virus-containing body fluids, as is the case with most human herpesviruses. HCMV is shed, often for prolonged periods of time, in urine, saliva, tears, semen and cervical secretions, as well as in breast milk. HCMV transmission rates are highest following direct contact with body fluids. Two distinct types of exposures are associated with horizontal transmission of HCMV: sexual transmission in adults and contact with urine or saliva from young children or saliva from adults. Infection is often transmitted among children and from young children to parents. One of the highest risk populations for acquisition of this virus is therefore an HCMV naïve pregnant woman who is a parent or caregiver of young children. This setting places the woman at risk of infection herself as well as of transmitting to the developing fetus. Adults may shed virus for weeks after primary infection as well as sporadically throughout life. General hygiene practices, such as regular hand washing, reduce transmission. Although HCMV shedding is common among hospitalized individuals, particularly immunocompromised patients, good hygiene practices by medical care professionals generally prevents HCMV transmission. Transmission via blood transfusion and HSCT has been reduced by donor screening. Transplantation of solid organs and tissues from seropositive donors to HCMV naïve recipients very efficiently transmits HCMV and remains clinically significant.

HCMV is transmitted vertically during pregnancy, during birthing as well as via breast milk, a pattern that is distinct from roseolaviruses, alphaherpesviruses or gammaherpesviruses. In developed countries such as the US, primary infection occurs in a small proportion of pregnant women where there is a high risk for transplacental transmission. Approximately 40% of newborns delivered by women with primary infection show signs of HCMV infection, compared to only 0.2% to 2% of newborns delivered by long-term HCMV-seropositive women. Transmission at birth as well as through mother's milk helps maintain HCMV infection in the population, with between 25% and 50% of infants nursed by seropositive mothers acquiring HCMV by one year of age. Importantly, disease sequelae are not generally associated with intrapartum or post-partum acquisition of HCMV in full-term newborns; however limited acute disease has been observed in rare cases.

Persistence and Latency

HCMV infection is lifelong. HCMV seropositive individuals are at risk of disease due to virus reactivation should they become immunocompromised. Tissues and organs from these individuals can transmit infection to naïve transplant recipients, a clear indication that this virus persists in cell types with a wide distribution. Based on detection of viral DNA, natural latency occurs in CD34 + stem cells and CD14 + lineage-committed granulocyte-macrophage progenitors that give rise to macrophages and dendritic cells, but not in mature granulocytes or lymphocyte populations. In HCMV seropositive individuals, very few (0.004% to 0.01%) bone marrow-derived mononuclear cells carry a few genome copies of viral DNA. Somewhat surprisingly, HCMV gene expression during lifelong quiescent infection is characterized by the same subset of early and late genes that are expressed during acute infection, albeit at lower levels in the very small proportion of bone marrow-derived hematopoietic cells that carry viral DNA. On the one hand, this diversity of viral gene expression contrasts the restricted latent gene expression of other herpesviruses. On the other hand, the question of whether HCMV persists in a chronic active infection rather than as latent virus remains for experimental consideration. It may certainly be the case that the balance of persistence and latency is determined at the level of the individual infected cell. The presence of viral DNA and gene expression are strong indications that myelomonocytic lineage cells are the reservoir of lifelong HCMV infection. Sporadic shedding that is observed in immunocompetent individuals may be due to reseeded ductal cells in the salivary gland by these myeloid cells; however, it remains possible that ongoing replication in ductal epithelium (or elsewhere) may seed the leukocytes and be responsible for the observed patterns of gene expression.

Leaving aside the complexities, quiescent infection is observed naturally in progenitor (stem) cells from the bone marrow (myeloid lineage), characterized by an absence of IE gene expression with repressive chromatin structure metering epigenetic control over virus replication. Despite ambiguities, several viral protein and RNA gene products have been implicated in latency and differentiation-dependent reactivation or in modulation of the host immune control over latency. HCMV infection of CD34 + stem cells drives differentiation towards monocyte-like cells dependent on inflammatory nitric oxide and STAT3 transcription factor under the influence of the virus-encoded US28 chemokine receptor. Thus, even in the absence of immune control, immature myeloid cells carry viral DNA without supporting replication and retain the potential to reactivate upon differentiation. The generation of macrophages or dendritic cells from infected myeloid progenitors, typically achieved by stimulation with proinflammatory cytokines, results in viral IE gene expression and replication. Clinically relevant reactivation in allogeneic transplant settings appears to depend on the elaboration of

proinflammatory cytokines that are themselves components of tissue rejection. Untreated, patients with compromised immunity to HCMV undergo a pattern of uncontrolled viral replication that results in disease. Clinical reactivation occurs in settings such as severe trauma or bacterial sepsis, where inflammation and the consequences of tissue damage suffice to induce replication possibly as a consequence of reperfusion injury. Severely immunocompromised patients, such as terminal AIDS patients, support high levels of HCMV replication with associated disease. Pregnancy in long-term seropositive women supports viral replication at levels that pose a small, but important risk of congenital infection in offspring.

Immunity

HCMV is subject to innate immune control, including via cell death, cytokines and natural killer (NK) cell mechanisms. This virus is controlled by pathogen-specific adaptive immunity mediated by T cells and, to a lesser extent, antibodies. On one hand, virus-specific cellular immunity is crucial for lifelong control over HCMV. NK and T cells collaborate to control of acute infection. Memory T cell surveillance suppresses replication and sustains clinical latency. Immunocompromised hosts with defects in cellular immunity are particularly susceptible to HCMV, as illustrated by increased disease in long-term HCMV seropositive recipients of hematopoietic cell transplants from HCMV naïve donors as compared to allogeneic transplants from MHC-matched long-term HCMV immune donors. In such settings, despite sustained serum antibody levels, host control depends on donor memory T cells that may be enhanced by administration of MHC-matched CMV-specific CD8 T cell therapy. On the other hand, anti-HCMV antibody makes a contribution to host resistance, best illustrated by the long-standing observation that transferred maternal antibody prevents transfusion-acquired HCMV in low birthweight infants. In addition, an envelope gB subunit vaccine reduces rates of infection in HCMV naïve women that is likely due to an antibody-mediated mechanism. Finally, there is growing consensus that intravenous immune globulin with HCMV-specific antibody helps reduce intrauterine infection, transplacental transmission and congenital disease.

HCMV emergence (whether amplification of persistent virus or reactivation from latency) is triggered by inflammatory signals and likely occurs sporadically throughout life even in fully immune individuals. This results in restimulation of HCMV-specific immunity such that memory T cells accumulate to exceptionally high levels with age. This cellular immunity affords protection because, once compromised, virus levels increase and disease consequences follow. The immunocompromised host provides the most convincing setting where HCMV control by T cell immune surveillance is crucial, supporting the widely held view that a cellular immune response, particularly by NK and memory T cells, maintain lifelong infection in immunocompetent individuals.

Modulation of the Host Response to Infection

HCMV commits most of its genome coding capacity to altering the host response to infection within the infected host cell and organism (**Table 1**). All classes of viral gene products (IE, E and L, as well as those associated with latency) contribute to the modulation of the host response to infection. These viral functions (whether protein or RNA) are typically not required replication in cultured cells. There are three major categories of such functions: (1) those acting within the infected cell (cell autonomous), modulating gene regulation, pathogen sensing, cellular metabolism, cell cycle and cell death pathways; (2) those acting within the host organism at the innate immune level, modulating innate cytokine and cellular responses; and, (3) those acting within the host organism at the adaptive immune level, modulating T and B cell immune effector mechanisms and memory. The virus employs modulatory functions to either evade host clearance mechanisms or to exploit host response pathways. Either way, these viral functions serve to enhance viral infection, pathogenesis and persistence. As with other herpesviruses, most of these functions are committed to suppressing host immunity. However, HCMV encodes cytokines and chemokines as well as chemokine receptors that exploit cytokine-dependent signaling within infected cells to benefit systemic dissemination. One recent insight shows that viral chemokine receptor US28 signaling drives differentiation of hematopoietic stem cells to become a dedicated monocytic reservoir of infection. Such fundamental orchestration of cellular development serves to ensure lifelong persistence.

Modulation of cell autonomous response to infection

HCMV gene products deflect intrinsic host cell antiviral responses mediated at the transcriptional level through activation of the cytokine pathways, cell stress, cell cycle, metabolism and cell death (**Table 1**). Protein-coding genes counteract host cell epigenetic downregulation of gene expression as well as interferon (IFN) activation. Similarly, epigenetic regulation involved in switching from latency to reactivation are influenced by viral regulatory proteins. Pathogen recognition receptor for DNA cGAS, NF- κ B activation and IRF3 activation are all targeted by multiple viral proteins. Inflammatory transcriptional activation, IFN response, and IFN-dependent activation of protein kinase R and 2',5' oligoadenylate synthetase are all suppressed. During productive infection as well as latency and reactivation, HCMV enhances signal transduction through TNF receptors and encodes a chemokine receptor to interface with cellular chemokines and chemokine receptors and dramatically enhance cell activation, migration and differentiation, potentially to benefit replication as well as latency and reactivation. HCMV-encoded chemokine likely recruit leukocytes to benefit the virus and host FLT3R becomes activated and enhances viral persistence in stem cell models of latency.

In a manner that has been compared frequently to tumor viruses, HCMV dramatically alters cellular metabolism and stress markers, downregulating ER and lysosomal stress, upregulating the unfolded protein/integrated stress response as well as tyrosine kinase signaling, activating certain cell cycle processes but imposing a striking block to cellular DNA synthesis. HCMV suppressed

various cell death pathways via both protein and RNA gene products engaging mitochondrial targets such as BCL2 family members Bax and Bak as well as caspase-8 to prevent apoptosis and alternate death pathways driven by serine proteases and protein kinases. These functions appear to foster an environment in infected cells that sustains infection, whether the outcome is productive replication or persistence.

Modulation of the innate immune response to infection

HCMV modulates cytokine signaling and encodes cytokine, chemokine, chemokine binding protein and chemokine receptor strategies to respond to host chemokines attract granulocytes and monocytes, apparently orchestrating the inflammatory response. This virus disarms the innate NK cell response through a large number of strategies sequestering ligands, impeding receptor engagement and mimicking MHC, all with the apparent goal of evading NK cell killing.

Modulation of adaptive immune response to infection

Natural HCMV infection is highly immunogenic, resulting in broad humoral antibody as well as cellular T lymphocyte responses that persist for life and maintain control over the virus infection. Immune responses to certain viral antigens are slow to develop, likely due to virus modulation of adaptive antibody and T cell responses. The antibody response to HCMV is targeted by several virus-encoded IgG Fc receptors, some of which also enhance virus spread dependent on the presence of antiviral antibody. T cell immune surveillance provides crucial control preventing amplification of virus throughout life. Some of the first HCMV-encoded immunomodulatory functions discovered downmodulate the ability of T cells to recognize HCMV-infected cells. Thus, MHC class I antigen presentation is reduced, while cellular and virus-encoded cytokines act together to impede T cell activation and effector function to benefit the virus. An important immunomodulatory impact of HCMV infection is that the antiviral T cell response is sustained at very high levels and increases with age of the host, a phenomenon called "memory inflation".

Diagnosis

Diagnosis of HCMV progressively improved from the 1980s and 1990s as virus isolation methods gave way to rapid immunological and, particularly, nucleic acid detection that are more rapid, less costly, quantitative and potentially automatable. HCMV DNA is readily detected in the tissues, plasma, leukocytes and whole blood of individuals at risk of developing HCMV disease. Immunohistochemical detection of HCMV antigens in biopsy samples and blood can confirm infection. Commercial diagnostic tests that rely on quantification of viral DNA in whole blood or plasma are in common clinical use. DNA amplification (predominantly PCR) methods are used to monitor at-risk populations with the intention to preemptively treat when levels of viral DNA reach a threshold value predictive of impending disease. The levels of viral DNA in the bloodstream are predominantly cell-associated but detected proportionally in plasma as well, providing an accurate prediction of disease risk. Detection of HCMV DNA in saliva, urine or dried blood spots of newborns within the first three weeks of life is diagnostic of transplacental transmission during pregnancy; however, levels of shedding in this or any other clinical setting does not predict disease severity in the infected host. Widespread HCMV screening of newborn blood spots facilitates the recognition of newborns at-risk of HCMV congenital disease who should be monitored for progressive hearing loss during early childhood.

Prophylaxis and Treatment

Antiviral prophylaxis has proven to be very effective in preventing HCMV disease in solid organ transplant recipients. However, late onset HCMV disease may follow prophylaxis in as many as 5% of patients. Prophylaxis or preemptive antiviral treatment in hematopoietic cell transplant recipients reduces the incidence of HCMV disease during the first 3–4 months from >80% to around 5%, significantly reducing mortality in adults and, particularly, in children. Late onset disease, usually more than 100 days after transplant, remains a clinical challenge. Long-term HIV antiviral therapy reduces incidence of HCMV disease in AIDS patients, although HCMV remains a significant risk in HIV-positive individuals when CD4 T cell counts drop to low levels, triggering HCMV-specific therapeutic intervention.

The advent of antiviral drugs such as ganciclovir (and valganciclovir), cidofovir, foscarnet, and letermovir, has allowed better clinical management of HCMV disease. Despite therapeutic intervention, this virus remains medically significant. The incidence of HCMV pneumonia in allogeneic HSCT as well as the preponderance of HCMV retinitis AIDS contributed to the development of antiviral drugs ganciclovir, foscarnet and cidofovir as well as to the development of orally administered valganciclovir for therapy. Most recently, letermovir for has been approved for antiviral prophylaxis. Valganciclovir, ganciclovir, foscarnet and cidofovir are approved for treatment of HCMV disease in the USA. All of these target the viral DNA polymerase to block viral DNA replication, with ganciclovir, valganciclovir and cidofovir acting as nucleoside analogs, and foscarnet acting as a pyrophosphate analog. Intravenous ganciclovir and orally administered prodrug valganciclovir are first choices in most settings for prophylaxis, preemptive therapy or therapy of active disease. These antiviral drugs come with significant toxicities. The antiviral, letermovir targets viral genome packaging machinery and is orally administered like valganciclovir but lacks toxicity so enables earlier prophylaxis in allogeneic HSCT. Additional oral drugs such as maribavir, targeting the UL97 protein kinase, have been under development. Prolonged use of any investigational or licensed drug

selects for resistant virus. Importantly, drug resistant mutants may complicate clinical management of the patient; however, fitness compromise of drug resistant mutants is such that they have not been found to circulate in the community.

Prevention

The medical significance and impact of HCMV congenital disease is recognized worldwide, making the development of a vaccine a universal public health priority. Attempts to develop live, attenuated, killed and subunit vaccines over the past 50 years have not yet resulted in a licensed product. The Centers for Disease Control and Prevention currently recommends that parents and caregivers of young children be informed of how HCMV is transmitted and of hygienic measures that reduce transmission, particularly to women who care for young children and may become pregnant.

The significantly lower rate of congenital disease following recurrent compared to primary maternal infection during pregnancy supports the potential benefits of universal vaccination. The limited success of a gB subunit vaccine strategy in reducing new infections by half in women of childbearing age lends a note of optimism that effective prevention may be possible. Additional subunit, mRNA and vectored vaccines incorporating gB along with the gH:gL pentamer plus antigens to stimulate T cell immunity are being evaluated as candidates for universal vaccination to prevent congenital infection or to protect stem cell and solid organ transplant recipients. Replication defective HCMV vaccines are also under clinical evaluation. Some vaccines are focused on recipients of organ and tissue transplants. Vaccines must either reduce transmission or suppress viral replication in tissues and organs depending on the clinical setting.

Future Perspectives

Despite tremendous gains in knowledge of HCMV molecular biology and pathogenesis, this virus continues to be a medically challenging cause of disease especially for immunocompromised hosts and early in fetal development. To date, antiviral chemotherapy has been only partially successful in controlling HCMV infection in transplant and other immunocompromised patients. There remains a continued and pressing need for vaccination to reduce the incidence of congenital HCMV infection. An important goal for future research will be to translate the large and growing body of basic knowledge of HCMV biology into improved treatments and effective vaccines.

Further Reading

- Britt, W.J., 2019. Cytomegalovirus. In: Bennett, J.E., Dolin, R., Blaser, M.J. (Eds.), *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, ninth ed. Oxford: Elsevier, pp. 1857–1872.
- Close, W.L., Anderson, A.N., Pellett, P.E., 2018. Betaherpesvirus virion assembly and egress. *Advances in Experimental Medicine and Biology* 1045, 167–207.
- Jackson, S.E., Sedikides, G.X., Okecha, G., *et al.*, 2017. Latent cytomegalovirus (CMV) infection does not detrimentally alter T cell responses in the healthy old, but increased latent CMV carriage is related to expanded CMV-specific T cells. *Frontiers in Immunology* 8, 733.
- Jackson, S.E., Sedikides, G.X., Okecha, G., Wills, M.R., 2019. Generation, maintenance and tissue distribution of T cell responses to human cytomegalovirus in lytic and latent infection. *Medical Microbiology and Immunology* 208, 375–389.
- Mocarski Jr., E.S., Shenk, T., Griffiths, P., Pass, R.F., 2013. Cytomegalovirus. In: Knipe, D.M., Howley, P.W. (Eds.), *Fields Virology*, sixth ed. Philadelphia: Lippincott, Williams and Wilkins, pp. 1960–2014.
- Patel, M., Vlahava, V.M., Forbes, S.K., *et al.*, 2018. HCMV-encoded NK modulators: Lessons from in vitro and in vivo genetic variation. *Frontiers in Immunology* 9, 2214.
- Patro, A.R.K., 2019. Subversion of immune response by human cytomegalovirus. *Frontiers in Immunology* 10, 1155.
- Schwartz, M., Stern-Ginossar, N., 2019. The transcriptome of latent human cytomegalovirus. *Journal of Virology* 93, e00047.
- Si, Z., Zhang, J., Shivakoti, S., *et al.*, 2018. Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate. *PLoS Pathogens* 14, e1007452.
- Stern, L., Withers, B., Avdic, S., *et al.*, 2019. Human cytomegalovirus latency and reactivation in allogeneic hematopoietic stem cell transplant recipients. *Frontiers in Microbiology* 10, 1186.
- Tandon, R., Mocarski, E.S., Conway, J.F., 2015. The A, B, Cs of herpesvirus capsids. *Viruses* 7, 899–914.
- Zhu, D., Pan, C., Sheng, J., *et al.*, 2018. Human cytomegalovirus reprogrammes haematopoietic progenitor cells into immunosuppressive monocytes to achieve latency. *Nature Microbiology* 3, 503–513.
- Zuhair, M., Smit, G.S.A., Wallis, G., *et al.*, 2019. Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. *Reviews in Medical Virology* 29, e2034.

Relevant Website

<https://www.cdc.gov/cmrv/overview.html>
About Cytomegalovirus and Congenital CMV Infection.
CDC.

Human Immunodeficiency Virus (*Retroviridae*)

Blaide Woodburn, Ann Emery, and Ronald Swanstrom, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Combinatorial antiretroviral therapy (cART) Combination antiretroviral therapy is the current standard of care for those living with HIV-1. These regimens are often described as a “cocktail” of different antiretrovirals that keep the virus from replicating via varying modes of drug action. First line cART usually consists of two nucleoside reverse transcriptase inhibitors and an integrase inhibitor.

Latent Reservoir Refers to the HIV-1 DNA that remains integrated in host cells following suppression of viral replication using cART. These viral genomes are generally not actively being expressed to produce infectious virus but do so stochastically and thus are responsible for viral rebound when antiretroviral therapy is stopped. Much of the latent reservoir is composed of defective viral genomes with intact genomes making up only a fraction of the viral sequences that persist as DNA in cells.

Reverse transcription The process of using single-stranded viral RNA as a template to produce full-length, double-stranded viral DNA.

RNA splicing A form of RNA processing that allows the generation of multiple messenger RNAs from a single large precursor RNA leading to the translation of multiple different proteins. This process is often utilized by viruses that have smaller genomes to increase protein diversity.

Virion A term used to describe a fully budded viral particle.

90–90–90 Initiative An initiative from the World Health Organization that aims to have 90% of all infected individuals aware of their infection status, 90% of those aware of their status in care and taking ART, and 90% of those taking ART to have the drugs fully suppressing viral replication.

Classification

The human immunodeficiency virus type 1 (HIV-1) is an enveloped, plus-stranded RNA virus of the *Retroviridae* family, further classified into the orthoretrovirus subfamily. The name retrovirus implies a special feature of the viral life cycle - the reverse flow of genetic information. Retroviruses reverse the usual flow of genetic information where DNA is used as a template to synthesize RNA (transcription) to now include a step of *reverse* transcription where viral RNA is used as a template to synthesize viral DNA. Many of the drugs used to control an HIV-1 infection target the viral DNA polymerase (called Reverse Transcriptase or RT) to block the step of reverse transcription.

There are a number of genera in the *Retroviridae* family, and HIV-1 is part of the *Lentivirus* genus. Lentiviruses have a worldwide distribution and have been found in cats, sheep, goats, horses, cows, and rabbits. Within the *Lentivirus* genus is also a lineage of primate lentiviruses, with HIV-1 being a member of this lineage.

The Origin of HIV-1

There is an important milestone that puts a time boundary as to when lentiviruses entered the primate lineage. Specifically, about one-half of the approximately 70 primate species in Africa are infected with a primate lentivirus (collectively called simian immunodeficiency viruses, or SIV) while none of the primates in the Americas are naturally infected. Thus we can infer that a lentivirus entered the primate population in Africa after the split of old world and new world primates and has spread through that population. In these naturally occurring infections the virus is rarely pathogenic, in contrast to the infection of humans with HIV-1.

Sequence analysis of viral genomes has become a powerful tool for tracking the history of viruses, and this tool has been especially important for understanding the HIV-1 epidemic. The SIV most similar to HIV-1 is found in chimpanzees ([Fig. 1](#)). Using sequence analysis of the viral genome we can infer that two different SIV lineages (from two other primates: SIVrcm from red-capped mangabeys and SIVgsn from greater spot-nosed monkeys) infected a chimpanzee and recombined to create a virus that could grow and adapt in chimps (SIVcpz). Given the similarity between chimps and humans this meant that to a large extent this adaptation would also improve replication in humans. About 100 years ago a single transmission event between chimps and humans occurred that gave rise to what we know as the HIV-1 epidemic, a lineage of virus in humans known as the main or M group. Chimpanzees have been used as food sources for human populations, with hunting and dressing carcasses being an obvious path for the introduction of these viruses into humans. It is hypothesized that this is how the first transmission event of SIV from chimpanzees to humans might have occurred. In addition, several other closely related SIVcpz strains have been transmitted to humans (possibly with gorillas as an intermediate) but these have not reached epidemic proportions as the M group lineage has.

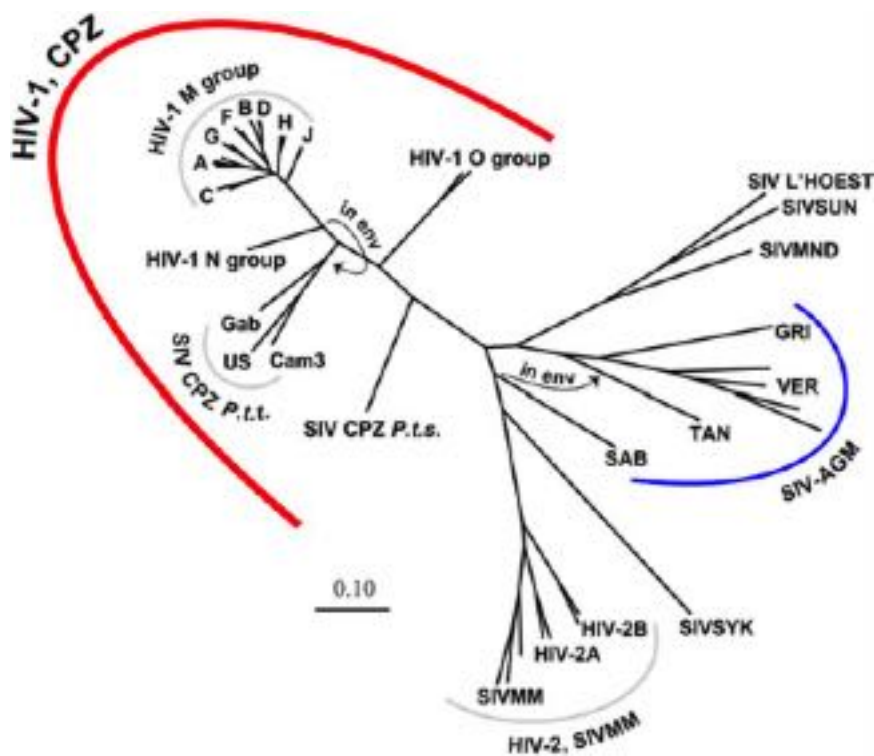


Fig. 1 Phylogenetic tree of the SIV and HIV lineages. Phylogenetic analysis of SIV and HIV sequences shows that each HIV-1 group (M, N, and O) originated from an independent transmission event of chimpanzee SIVcpz to humans followed by diversification into subtypes (e.g., Group M into Subtypes A, B, C, etc.). This analysis also demonstrate how HIV-2 has been transmitted multiple times to humans from its origin as SIVmm in sooty mangabeys. Figure adapted from Kuiken, C.L., Foley, B., Hahn, B.H., *et al.*, 1999. Human Retroviruses and AIDS: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. Theoretical Biology and Biophysics Group. Los Alamos, NM: Los Alamos National Laboratory.

There is a second SIV that has infected humans, known as HIV-2. This virus is most similar to an SIV found in sooty mangabeys (SIVsm). The SIV of sooty mangabeys has been transmitted into humans a number of times giving rise to the smaller epidemic of HIV-2 in west central Africa with a virus that is significantly less pathogenic than HIV-1.

Organization of the Genome

HIV-1 has a single-stranded RNA genome of about 9700 nucleotides that is packaged into the virion as two identical copies of RNA (**Fig. 2(A and B)**). The genomic RNA is plus stranded, meaning it is in the same polarity as its messenger RNAs (i.e., the strand that can be translated into protein). However, for retroviruses the genomic RNA is not translated when it enters the cell but rather copied into double-stranded DNA. Thus the genome exists in two states, genomic RNA that is packaged into the virion and a DNA copy of the genome that is integrated into the host cell genomic DNA. Once integrated into the host genome, transcription of viral DNA into RNA begins. Only one transcript type is made – a full length transcript equivalent to genomic RNA (**Fig. 2(C)**). The transcript is made by the cellular RNA Pol II and processed by cellular machinery, so it looks like a cellular mRNA with a 5' cap and a 3' poly A tail (**Fig. 2(A)**). The RNA and DNA forms of the genome are very similar, retaining the gene order. However, during DNA synthesis portions of the RNA sequence at the ends of the RNA are rearranged and duplicated at the ends of the DNA to give the DNA long terminal repeats (LTR; **Fig. 2(C)**). These duplications are mediated during DNA synthesis by a short direct repeat at the ends of genomic RNA called R. This allows the virus to place in its DNA form the transcription regulatory signals in U3 upstream of the transcription start site, signals that are duplicated near the 3' end of the RNA and thus not lost when they are excluded from the transcribed RNA at the 5' end. In the 5' untranslated region (UTR) the genome folds into extensive secondary structures with important regulatory functions. This region contains a packaging signal, ψ (Psi), that identifies the viral genome as the RNA that should be packaged into newly forming viral particles. Genomic RNA is packaged in dimers and a dimerization signal (DIS) initiates the dimeric folded structure.

Downstream of the 5' UTR are the viral genes. HIV-1 has 10 genes (*gag*, *pro*, *pol*, *vif*, *vpr*, *tat*, *rev*, *env*, and *nef*). Like the 5' end of the genome, the 3' end of the genome also contains sequences that function in the RNA that are important for replication (such as the polyadenylation signal).

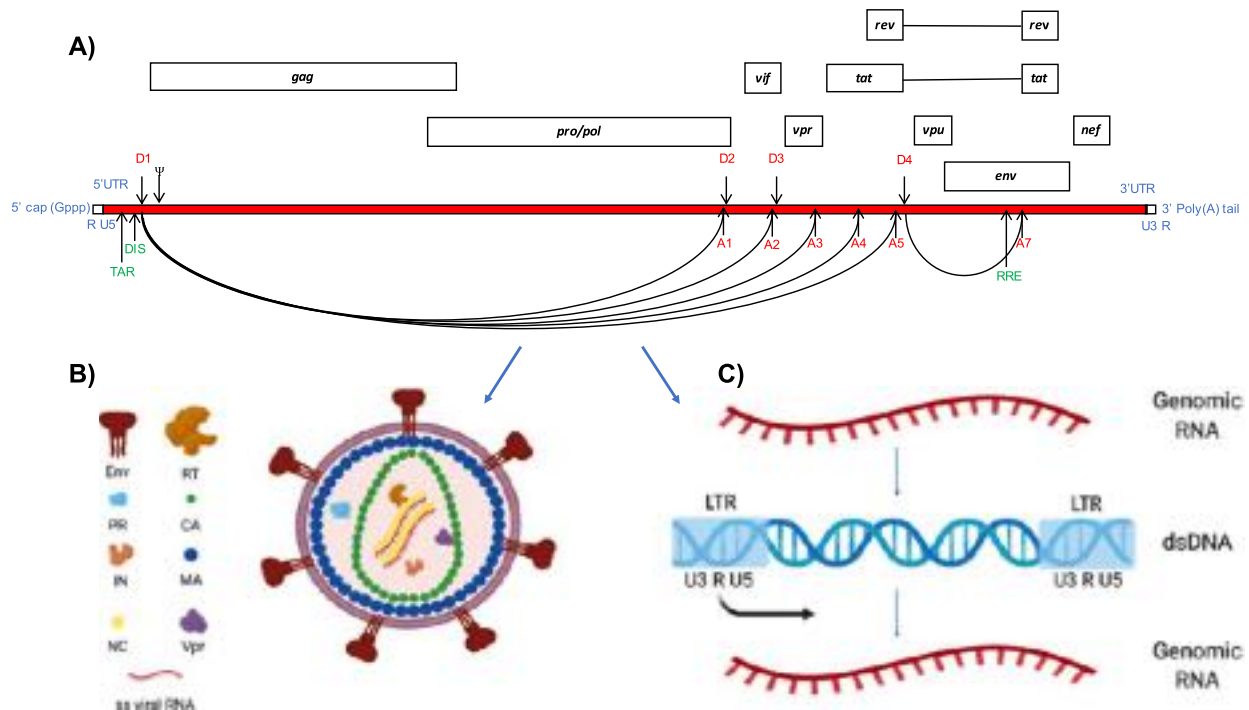


Fig. 2 HIV-1 virion structure and genome. (A) The HIV-1 RNA genome is approximately 9.7 kilobases and consists of 10 genes (shown with boxes) that code for 15 mature proteins. Each end of the RNA genome is comprised of a repeated sequence known as R with sequences inside of R labeled as U5 near the 5' end and U3 near the 3' end. The 5' and 3' untranslated regions (UTR) encode cis-acting replication signals as described in the text. Genomic RNA has the hallmarks of a eukaryotic RNA with a 5' cap and a 3' poly A tail. Also shown is the major splice donor site D1 and the other splice donors D2-D4, as well as the positions of the major splice acceptor sites A1-A5 and A7. (B) The HIV-1 particle is spherical, approximately 100 nm in diameter, and has a modest number of Env protein trimers (maroon) in its membrane envelope that facilitate entry into the host cell. The particle contains several structural proteins that dictate its geometry, including the matrix (MA; blue) and capsid (CA; green) proteins. Additionally, the particle is prepackaged with functional proteins reverse transcriptase (RT; orange) and integrase (IN; pink), and also includes the viral protease (PR; blue) which functioned during assembly/maturation. Two identical plus strands of viral RNA comprise the viral genome, which is contained in the capsid cone of the virion. (C) Through reverse transcription, HIV-1 RNA (red) is copied into double-stranded linear DNA (blue) that is then integrated into the host cell genome. DNA synthesis creates a direct repeat of sequences that in the RNA were discontinuous (U3 R U5, blue box), with this direct repeat called the long terminal repeat, or LTR. Following integration, the viral promoter within the LTR (U3) drives expression of the DNA form of the genome into full length RNA, creating new viral RNA to be packaged into assembling virions or used in the expression of viral proteins (with or without RNA splicing).

Expression of Viral Genes

The full-length 9700 nucleotide-long transcript can be used either as genomic RNA and packaged into virions or as a messenger RNA. In eukaryotes only the 5' most reading frame can be translated on an mRNA. For HIV-1 this means that when the full-length transcript serves as an mRNA it is for the *gag* gene. About 95% of the time, a Gag protein is synthesized and translation stops at the end of the *gag* reading frame. However, embedded in the RNA near the end of the *gag* coding region is a stretch of uridines and a stem-loop structure that occasionally (5% of the time) causes the ribosome to slip back one nucleotide while decoding the RNA (−1 frameshift); this changes the reading frame the ribosome is in with translation now continuing into the *pro* and *pol* genes creating a Gag-Pro-Pol polyprotein precursor. This is one strategy the virus uses to get around the limitation of having translation limited only to the 5' most open reading frame of an RNA.

The other strategy the virus uses to express its downstream genes is RNA splicing. Splicing is used in the expression of most cellular genes to excise introns from a long RNA precursor and piece together the exons that make up the final mRNA. This is what happens to the virus except here the internal regions of the genome serve as either exon or intron, depending on which gene is being expressed. The cellular splicing machinery generates over 40 spliced mRNAs from the full length genomic viral RNA, and a simplified scheme of these splicing events is shown in Fig. 2(A). If the full-length transcript is spliced, it always uses the main splice donor D1 (in the 5' UTR) and splices to one of the downstream acceptors, A1-A7. Splicing excises the section of the genome under the arc, so, for example, a splice to A1 removes the section between D1 and A1 (Fig. 2(A)). This joins the 5' UTR to the translation start site for the *vif* open reading frame (which is just downstream of the A1 splice site) to make a transcript that will be translated to generate Vif protein. Similarly, splicing to the other acceptors moves the translation start sites for each of the downstream genes next to the 5' UTR to allow it to be translated. There are still several twists. First, both the *env* and *vpu* open reading frames are

translated from the same mRNA (bending the rule of only the 5' most reading frame being translated). Second, RNAs that are spliced at D1 may or may not splice again from D4 to A7. The exons for Tat and Rev span this D4-A7 splice site, requiring the D4-A7 splice to form their complete open reading frames but at the expense of the intact *env* reading frame.

The D4-A7 splicing event is regulated to allow the formation of mRNAs for Tat, Rev, and Nef when splicing occurs, and both the full length genomic RNA and mRNA for the Env protein when splicing is blocked. Cells typically require RNAs to be fully spliced in order to leave the nucleus, and the mRNAs for Tat, Rev, and Nef follow this rule (and the normal cellular mRNA export pathway). However, for the unspliced genomic RNA and partially spliced *env* mRNA (retaining the D4-A7 intron) the virus must trick the cell to export these RNAs. The viral Rev protein binds to a region of RNA secondary structure in the *env* intron called the Rev Response Element (RRE) and links the unspliced and partially spliced RNAs to an alternative RNA export pathway by binding to the host protein CRM1 to move these RNAs out of the nucleus and into the cytoplasm.

Virion Proteins

Most of the viral proteins that are found in the virion are initially synthesized as larger polyprotein precursors that must be cleaved by a protease. Thus the *gag* gene encodes the Gag polyprotein precursor that is targeted to the cell plasma membrane where the virus buds out of the cell. During virion formation/maturation the viral protease (see just below) cleaves the Gag protein into its membrane-bound, N-terminal matrix (MA) protein, its capsid (CA) protein that will form the characteristic cone structure within the enveloped virus, its nucleocapsid (NC) protein which binds the viral RNA, and the C-terminal p6 protein which is involved in interacting with a host cell pathway to allow the membrane that surrounds the budding virus to pinch off and leave the cell.

Approximately 5% of the time translation of Gag slips into the -1 reading frame to create the Gag-Pro-Pol polyprotein precursor. When the Pro domain dimerizes this creates the viral protease (PR), which is responsible for cleaving the Gag and Gag-Pro-Pol polyprotein precursors. This happens during the budding/maturation process. The protease is responsible for cleaving itself out of the precursor, for cleaving Gag, and for releasing the viral DNA polymerase reverse transcriptase (RT). The protease cleavage event at the C terminus of RT releases the viral integrase protein (IN). PR and RT function as dimers, while IN functions as an even higher order oligomer. The p6 domain at the C terminus of Gag has a binding site for the viral Vpr protein. There are about 2000 copies of Gag in a virion, 100 copies of the Gag-Pro-Pol precursor, and 75 copies of Vpr.

The other major viral protein is the envelope or Env protein on the surface of the virion. It takes a different path to the assembling virion. As a membrane protein, it is synthesized in the endoplasmic reticulum (ER) in the cell with the N terminus threaded into the lumen of the ER until the appearance of an encoded translocation stop signal about 200 amino acids from the C terminus of Env. The initial protein is called gp160 based on its apparent molecular weight and the fact that it is highly glycosylated. The protein trimerizes and migrates to the cell surface, always embedded in the membrane near its C terminus. In the Golgi, a cellular protease (furin) cleaves the protein into two subunits: an extracellular surface protein gp120 (SU) and the transmembrane protein gp41 (TM), with the gp120 and gp41 subunits maintaining noncovalent interactions. The final Env trimer is composed of three subunits of gp120 and three subunits of gp41. For HIV-1 there are only about 10–20 trimers on a virion, occupying only a small amount of the surface of the enveloped virus. gp120 interacts with the CD4 protein on the surface of helper T cells to determine which cell gets infected, i.e., CD4 is the primary receptor for the virus. After binding CD4, gp120 goes through a series of conformational changes to create a binding site for a coreceptor, usually a cellular surface protein called CCR5. After binding these two cell surface proteins, gp41 then mediates fusion between the viral and host cell membrane to allow entry of the capsid into the cytoplasm of the cell, where viral DNA synthesis occurs. Small amounts of other viral and cellular proteins have been reported to be associated with virions but these are largely of unknown significance.

The Other Viral Proteins Work Intracellularly

All retroviruses encode the main replicative proteins Gag, Pro, Pol, and Env. For some retroviruses these are the only proteins encoded in the viral genome. Other retroviruses encode a variety of additional proteins, often called accessory proteins since they are in addition to the replicative proteins. However, adapter proteins might be a better name since most of them function to bring together proteins in the cell (think an adapter for an electrical plug) in ways that benefit the virus. HIV-1 encodes six such proteins: Tat, Rev, Nef, Vif, Vpr, and Vpu.

Tat and Rev play critical roles in initiating the expression of the viral genome. The newly integrated viral DNA is transcribed by RNA Pol II from the host cell. However, most of the initial transcripts are short, with only a fraction being full length, with these few full length RNAs being spliced to form Tat and Rev mRNAs. The initial Tat protein that is synthesized binds to a feature of RNA secondary structure at the 5' end of the RNA, a structure called the TAR loop. In this way the Tat protein recruits the transcription and elongation factor P-TEFb. P-TEFb includes a protein kinase (CDK9) which phosphorylates the C terminus of RNA Pol II to increase its elongation efficiency allowing the synthesis of more full-length RNA. In parallel the early Rev protein binds to the RRE to allow partially spliced and unspliced viral RNAs to exit the nucleus thus allowing for all of the viral RNAs, including the unspliced genomic RNA, to exit the nucleus (as described above).

The viral Vif, Vpr, Vpu, and Nef proteins counteract cellular antiviral restriction factors and regulate cell processes to suit viral replication. These proteins have multiple targets and effects. However, a common theme is to bind a cellular protein that has a deleterious effect on viral replication and link it to the proteasome where it gets degraded.

A clear example of how this group of viral proteins acts as an adapter is seen with Vif. Vif targets the cellular factor APOBEC3G for proteasomal degradation. In the absence of Vif, APOBEC3G is packaged into virions and causes cytosine deamination in the minus strand of viral DNA during reverse transcription in the next cell, which results in G to A mutations in the positive strand. The minus strand is initially present as single-stranded DNA and this is the target of APOBEC3G. Viruses lacking Vif are unable to grow in cells that produce APOBEC3G. Vif brings APOBEC3G to a cellular ubiquitin ligase where it is ubiquitinated, targeting it for degradation by the proteasome. Vpr also binds to a cellular ubiquitin ligase but its host cell target is still a focus of research. Vpr induces the infected cell to arrest in the G2 phase in the cell cycle, perhaps to induce DNA repair enzymes that ultimately assist in integration. However, such activity would have to be carried out by just the Vpr packaged in the virion since integration happens before viral gene expression. The extent to which the small amount of virion-associated Vpr is the effective form of the protein or if it has functions after expression of larger amounts of protein after integration is unknown. Vpu counteracts the cellular tetherin protein. Tetherin binds to both the cellular membrane and to the membrane of the newly budding virus, thus tethering the virus to the cell surface. Vpu reduces the amount of tetherin on the cell surface. Nef has multiple functions, such as down regulation of CD4 from the surface of the cell to reduce interaction with the newly synthesized Env trimers, down regulation MHC class I from the surface of the cell to reduce immune surveillance, reduction in the amount of SERINC5 which when incorporated into the virion inhibits subsequent fusion to the next cell, and interacting with a cellular kinase to affect cellular activation.

Life Cycle

It is convenient to divide the viral life cycle into two phases: an early phase starting with binding of the virus particle to the target cell through the integration of viral DNA, and a late phase starting with transcription of the integrated DNA through virion budding and maturation.

Early Phase

Attachment and Entry

The first step of the virus life cycle is attachment to the host cell via the CD4 protein on the surface of helper T cells using the gp120/SU portion of the Env protein (**Fig. 3**). The normal form of HIV-1 requires a high density of CD4 on the surface of the cell, marking the CD4⁺ T helper cell as its primary target cell. As noted above, fusion of the viral and host membranes requires the presence of the primary surface receptor, CD4, and one of two secondary surface receptors, usually CCR5 or sometimes CXCR4, two chemokine receptors that are found on a subset of cells. Engagement with the coreceptor exposes the hydrophobic gp41 fusion peptide, which then inserts into the host cell membrane. Insertion of each gp41 in the trimer brings the virus in even closer proximity to the host cell membrane; gp41 then forms a hairpin mediated by two helical domains in the protein, and since gp41 is in a trimer this results in a six helix bundle. This six helical bundle serves as the driving force behind membrane fusion by bringing the membranes close enough to form a fusion pore through which the virus can release its capsid into the cytosol of the host cell where viral DNA synthesis can begin. The need for a high density of CD4 for efficient entry and the use of CCR5 as the coreceptor make the predominant form (and the transmitted form) of the virus appropriately called HIV-1 R5 T cell-tropic.

Uncoating, Reverse Transcription, Nuclear Localization, and Integration

Following release of the viral capsid into the host cell cytoplasm, the capsid core begins to uncoat, although the extent to which the p24 CA capsid protein is released is unclear. The partially uncoated core utilizes the host cytoskeletal network to traffic toward the host cell nucleus. Inside the viral capsid core, reverse transcriptase (RT) begins converting the single-stranded viral RNA into dsDNA, with DNA synthesis being primed initially using a host cell tRNA that was earlier packaged into the virion and placed near the 5' end of the RNA genome (**Fig. 3**). Linear, double-stranded DNA is fully synthesized in the cytoplasm. This replication complex contains integrase, and while in the cytoplasm the integrase protein removes 2 bases from each 3' end of viral DNA in preparation for integration in the nucleus, making the pre-integration complex (PIC).

Once reaching the nucleus the PIC is transported across the nuclear membrane through the nuclear pore complex, in part mediated by interactions with residual CA protein. Inside the nucleus the PIC associates with host factors (such as LEDGF) that target the PIC to transcribed regions of the viral genome. These host factors help direct integration to regions of the host genome where the viral genome will be more likely to be transcribed. After interacting with the host chromatin, the integrase protein cleaves the host DNA and, in a strand transfer reaction, covalently links the viral DNA to the host DNA where it can be transcribed by the host transcription machinery (**Fig. 3**).

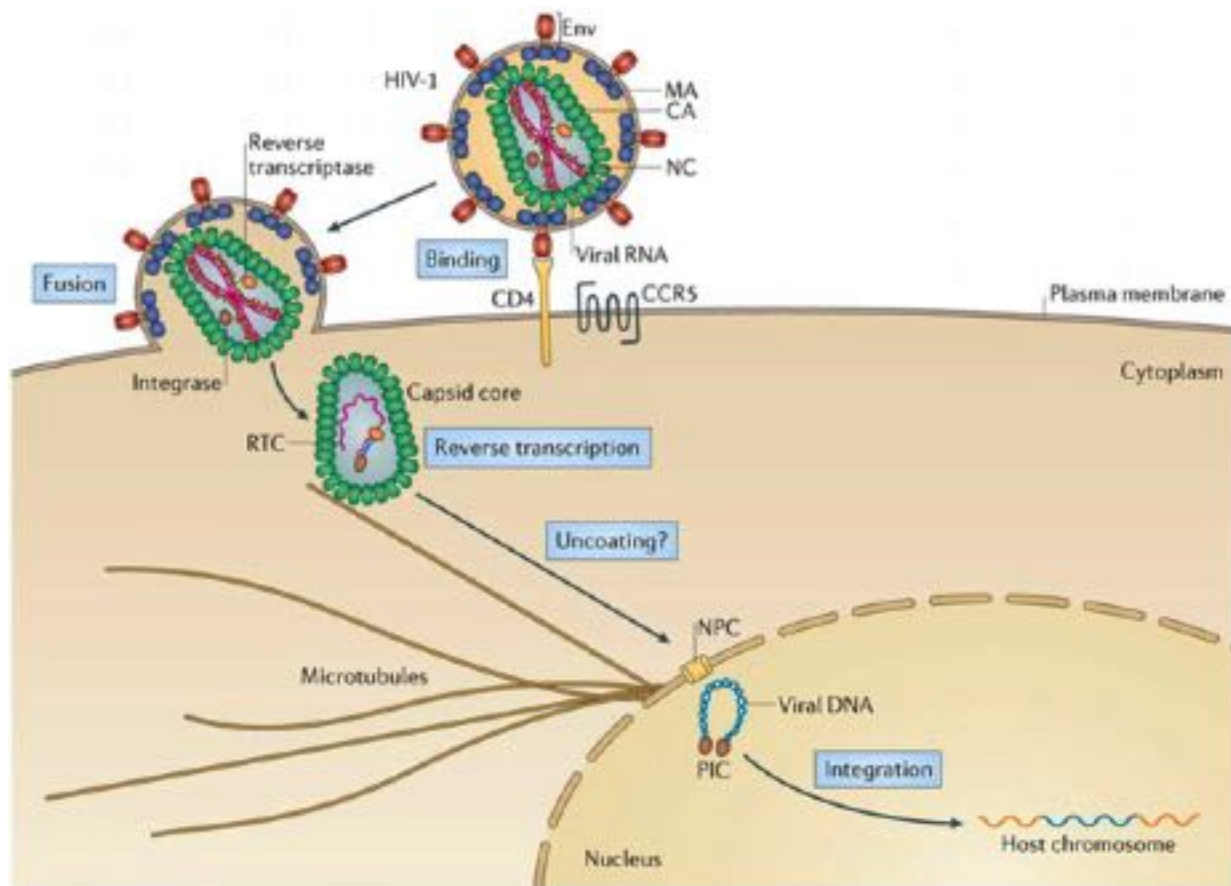


Fig. 3 Early phase of the HIV-1 life cycle. HIV-1 begins its life cycle by binding to its primary receptor CD4 and then its secondary receptor CCR5 (shown) or CXCR4 to facilitate membrane fusion. In the cytoplasm there is some level of “uncoating” to form the Reverse Transcription Complex (RTC) where reverse transcription of the viral RNA genome occurs while making its way to the nucleus using a vast array of microtubule networks. After completion of DNA synthesis the viral DNA is associated with integrase forming the Pre-Integration Complex (PIC), which moves through nuclear pore complex (NPC) to traverse the nuclear membrane. The PIC migrates to a host chromosome where integration occurs. Figure adapted from Campbell, E.M., Hope, T.J., 2015. HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nature Reviews Microbiology* 13, 471–483. with permission.

Late Phase

Gene Expression and Assembly

After integration, sequences in the LTR are recognized by host RNA polymerase II to synthesize viral RNA. However, initially the resulting transcripts are short and prematurely terminated, with only a small fraction being full length genomic RNA. The fraction that is full length is spliced by the host cell machinery initially to give mRNAs for Tat, Rev, and Nef. Tat recruits a host transcription factor complex P-TEFb with Tat binding to an RNA hairpin in the short transcript (TAR). Bringing P-TEFb close to the host RNA polymerase II leads to modification of the polymerase to allow more of the transcripts to be full length (i.e., making the polymerase more processive).

Normally mRNAs have to be fully spliced to leave the nucleus. However, HIV-1 gene expression requires the use of partially spliced or even unspliced RNA (the form of genomic RNA). To accomplish this the viral Rev protein binds to another large RNA hairpin within the D4-A7 intron in the *env* gene. Rev links partially spliced and unspliced viral RNA to a different nuclear export pathway, mediated by the host protein Crm1, such that the RNA avoids the quality control the host applies to RNAs going through the mRNA nuclear export pathway. Over 40 RNA transcripts are generated from the full length viral genomic RNA transcript by alternate splicing patterns, which allow expression of all of the viral proteins and full length RNA for packaging into new virions.

As noted above, some of the viral proteins function to manipulate the host cell (Tat, Rev, Nef, Vif, Vpr, Vpu) while others are used to make new virion particles. The Env protein migrates through the ER/Golgi apparatus of the cell to get to the cell surface/plasma membrane (Fig. 4). The Gag and Gag-Pro-Pol proteins are synthesized in the cytoplasm but are targeted to the inner face of the cell surface/plasma membrane. The Gag protein alone is sufficient to mediate budding of a particle from the cell membrane (with Gag-Pro-Pol being mixed in at a 1–20 ratio). The Env protein interacts with the N-terminal domain of Gag (the MA domain) to increase its incorporation into budding virion particles.

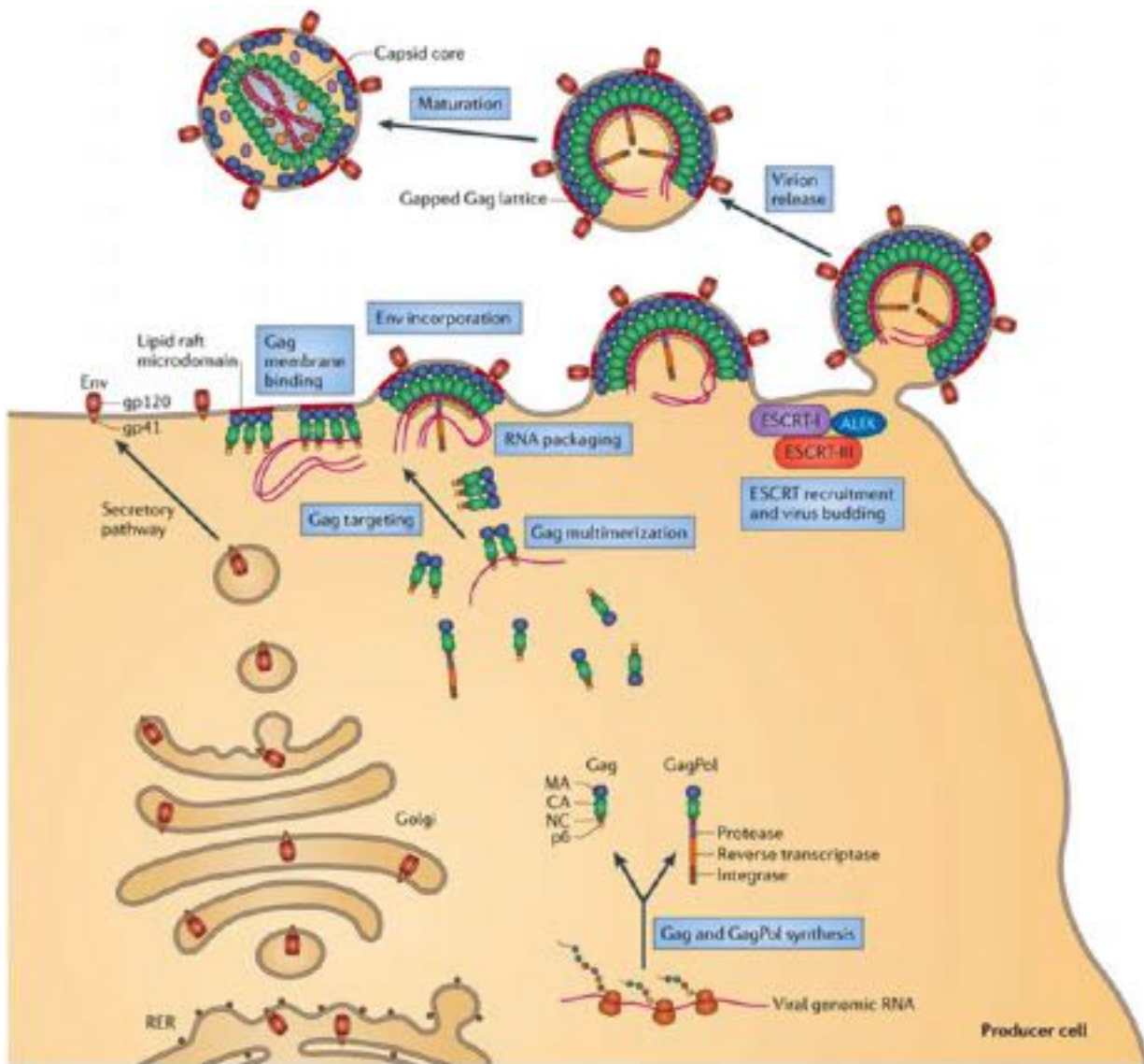


Fig. 4 Late Phase of the HIV-1 life cycle. Following integration, the HIV-1 genome is transcribed into full length RNA followed by complete splicing to produce Tat and Rev, which increase the level of expression and allow for unspliced and partially spliced RNAs needed for virion assembly and to function as mRNA for the structural proteins. HIV-1 Env is processed and trafficked to the host cell membrane via the secretory pathway (ER and Golgi) while Gag, with some bound to full length viral RNA, and Gag-Pro-Pol (here called GagPol) migrate through the cytoplasm toward the host cell membrane to form an immature, budding viral lattice. Membrane fission and release of the newly assembled virion from the host cell is mediated by ESCRT I, ESCRT-III, and ALIX proteins, and maturation of the virion is dependent on viral protease cleavage of Gag and Gag-Pro-Pol into their respective proteins. Figure adapted from Freed, E.O., 2015. HIV-1 assembly, release and maturation. *Nature Reviews Microbiology* 13, 484–496. with permission.

Budding and Maturation

As a Gag lattice forms at the surface of the plasma membrane it creates a budding structure. However, the particle must undergo membrane fission to be released from the host cell in the form of an immature virion. Host proteins in the ESCRT pathway target the C-terminal domain of Gag (p6) to mediate this membrane fission allowing the enveloped virus to detach from the cell (Fig. 4). During this budding process, the viral protease PR dimerizes in the Gag-Pro-Pol precursor to become active, first cleaving itself out then cleaving the Gag and Gag-Pro-Pol precursor polypeptides into their respective proteins to initiate the maturation of the virion into its infectious form with its familiar cone-shaped capsid core.

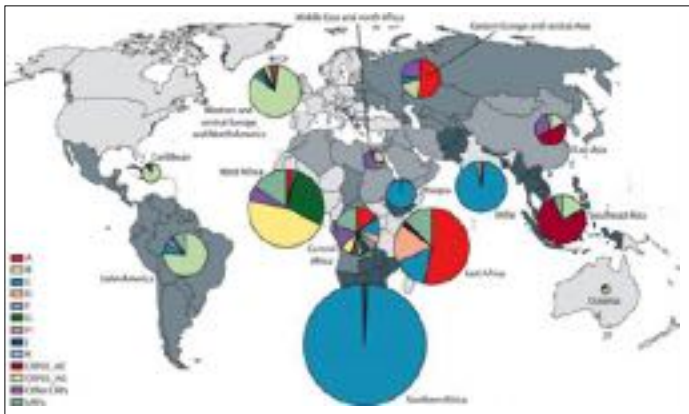


Fig. 5 HIV-1 global subtype distribution and 2019 UNAIDS statistics. Countries forming a region are shaded in the same color on the world map. The surface area of each pi-chart corresponds to the total number of people living with HIV-1 in each region. CRF = circulating recombinant form. URFs = unique recombinant forms. Figure adapted from Hemelaar, J., Elangovan, R., Yun, J., *et al.*, 2018. Global and regional molecular epidemiology of HIV-1, 1990–2015: A systematic review, global survey, and trend analysis. The Lancet Infectious Diseases 19, 143–155. with permission.

2019 UNAIDS HIV Statistics
37.9 million people living with HIV
23.3 million people had access to ART
1.7 million people became newly infected
770,000 died from AIDS-related illness
90-90-90 Initiative
79% living with HIV knew their status
78% of those aware of their status were receiving treatment
86% of those receiving treatment were virally suppressed
At Risk Populations
Men who have sex with men (22 times more likely)
Injection drug users (22 times more likely)
Sex workers (21 times more likely)
Transgender people (12 times more likely)

Epidemiology

Currently we recognize distinct lineages of HIV-1 called subtypes or clades (designated subtypes A, B, C, etc. and also called subgroups; Fig. 5) which are all part of the M group. Virus isolates within a subtype are more similar at the sequence level than they are to virus isolates in different subtypes. We interpret the existence of these subtypes as the result of having a diverse (in sequence composition) population of viruses within the human population within western Africa around 60 years ago. The virus then spread in single transmission events to different areas of the world, creating a sequence or genetic bottleneck that defined the subsequent expansion of that virus in these different places. These are all strains of HIV-1 with either little or no biological differences but we can still see the evidence of the genetic bottleneck coming to North America for example, where we have viruses that are related to each other and called subtype B (or clade B), the same predominant viral lineage that is found in western Europe. We can see subtypes A and D in eastern Africa, and another jump occurred subsequently to create another A subtype cluster from the east African virus in eastern Europe. Subtype C virus entered southern Africa then spread through India and into China. These are broad stroke descriptions of the early spread of the virus into the larger human population. Where two subtypes are circulating in the same human population it is possible for a person to become infected with both subtypes where recombinant genomes between the two viruses can occur and get transmitted, making the inference of its history more difficult. In addition, as people have moved between different regions of the world over the years the different lineages of viruses have become more mixed.

Modes of Transmission and Risk Factors

The modes of HIV-1 transmission are well-defined. HIV-1 can be transmitted through sexual activity, exposure to blood, and from mother to child during pregnancy, birth, and breast feeding. There are six bodily fluids by which HIV-1 transmission can occur, including blood, semen, pre-semen, vaginal secretions, anal secretions, and breast milk. Additionally, there are four common routes of transmission that include the vagina, the tip of the penis, anus, and open sores (especially in the genital tract). In general, exposure of mucosal surfaces to virus-containing fluids creates the risk of transmission. There is no epidemiological data to suggest that HIV-1 transmission can occur via exchange of saliva (kissing, sharing drinks and food) or mosquito bites, mostly because the virus is unable to reach transmissible titers in these fluids.

Unprotected sexual contact is the primary mode of transmission worldwide. HIV-1 has a relatively low rate of transmission, with transmission rates at least 10–100 times less per sexual encounter than other sexually transmitted infections (STIs) such as gonorrhea, chlamydia, and syphilis. However, there are numerous factors that can enhance infectivity of the source partner such as a high viral load, other STIs (in either partner), and blood contact during sex. The probability of transmission increases with the likelihood of microtrauma during sexual contact and direct exposure of blood to infected fluids (Table 1).

Outside of sexual contact, another common mode of transmission is through exposure of HIV-infected blood via injection drug use. Sharing of needles with HIV-contaminated blood has the highest rate of transmission of any exposure, aside from blood transfusion with HIV+ blood which has a transmission rate of essentially 100%. Mother-to-child transmission can take place

Table 1 Estimated per-act risk for acquisition of HIV-1, by exposure route

Exposure route	Risk per 100 exposures to an infected source
Blood transfusion	90
Needle-sharing injection-drug use	0.67
Percutaneous needle stick	0.3
Receptive anal intercourse	0.5
Receptive penile-vaginal intercourse	0.1
Insertive anal intercourse	0.065
Insertive penile-vaginal intercourse	0.05
Receptive oral intercourse	0.01
Insertive oral intercourse	0.005
Mother-to-child transmission (without breast-feeding)	30
Breast-feeding for 18 months	15

Source: Adapted from Smith, D.K., Grohskopf, L.A., Black, R.J., *et al.*, 2005. Antiretroviral postexposure prophylaxis after sexual, injection-drug use, or other nonoccupational exposure to HIV in the United States: Recommendations from the US Department of Health and Human Services. *MMWR Recommendations and Reports* 54, 1–20. Kourtis, A.P., Lee, F.K., Abrams, E.J., Jamieson, D.J., Bulterys, M., 2007. Mother-to-child transmission of HIV-1: Timing and implications for prevention. *The Lancet Infectious Diseases* 6 (11), 726–732.

during pregnancy, during delivery, and during breastfeeding, although this transmission risk is essentially reduced to zero if the mother is on suppressive antiviral therapy.

Prevalence

HIV-1 has been detected in every region of the world. Currently it is estimated that 37.9 million people are living with HIV. In 2019, 1.7 million people became newly infected and 770,000 died from AIDS-related illnesses (**Fig. 5**). Fortunately, 23.3 million individuals had access to antiretroviral therapy (ART), and this number will only continue to rise with global initiatives to increase sexual health education and access to treatment.

The prevalence of HIV-1 varies globally, disproportionately affecting Sub-Saharan Africa. Eastern and Southern Africa contain ~60% of the total global infections, with Asia and the Pacific (15%), West and Central Africa (13%), Latin America (2%), the Caribbean (0.9%), the Middle East and North Africa (0.6%), Eastern Europe and Central Asia (4%), and Western/Central Europe and North America (5%) comprising the other ~40%. Prevalence is not uniform across the population in these regions. Both engaging in acts that can mediate transmission and the level of pre-existing infection within the subgroup where these acts are shared have led to significantly elevated levels of infection among certain subgroups of people, which has led to focused efforts in treatment and prevention within these subgroups.

Diagnosis

A diagnosis of an HIV-1 infection is made based on one of three different measures. Most commonly the diagnosis is made based on the host antibody response to the virus. By examining a group of people with suspected time of exposure it is possible to determine that a person becomes antibody positive in an ELISA (enzyme-linked immunosorbant assay) test around 22 days after transmission. This test is usually confirmed with a Western Blot test where viral proteins transferred to a thin membrane are bound by host antibodies and the presence of the bound antibodies detected. While these tests are usually done using a sample of blood, antibodies are also shed into saliva which has allowed a “rapid test” to be developed based on a saliva sample to test for the presence of HIV-specific antibodies, although this test becomes positive later than the antibody blood test.

The detection of antibodies is a measure of the host response to the replicating virus. The two other tests are designed to measure the presence of the virus itself. Virus is replicating in the body before the antibody response appears. Thus detection of the virus itself offers the possibility for detection of the infection even before the appearance of the host antibody response. The two tests that are used are a sensitive ELISA assay to detect the viral capsid protein (CA/p24), and a PCR-based assay to detect the presence of viral RNA. Most people who get infected are initially infected with a single or a few virus particles that then start replicating in the new host. Very early after infection there is not enough virus to measure with these methods so for about 7–10 days there is a period called the eclipse phase where the viral infection can’t be seen by any of these tests. The detection of viral RNA in the blood is the most sensitive test and the first to be positive after the eclipse phase, followed within a week by the ELISA test for viral capsid protein, followed by one week for the detection of viral antibodies. In the initial absence of an immune response the virus multiplies rapidly, reaching a peak level of viremia in the blood at about 3 weeks. With the onset of the host immune response the virus is partially controlled and brought down to a “set point”, measured as copies of viral RNA per milliliter of blood (usually in the range of 10,000 to 100,000 copies of viral RNA per milliliter). This level of viral RNA is relatively constant or trends upward over time and represents a steady-state level of viral replication and the death of infected cells over time.

Table 2 *AIDS Defining Illnesses.* There are a set of illnesses often associated with HIV-1 infection and the onset of AIDS. These range from bacteria and other viral infections to virus-induced cancers. This is the list of AIDS defining illnesses as designated by the U.S. Center for Disease Control

AIDS-defining illnesses

Bacterial infections (recurrent)	Kaposi sarcoma
Candidiasis of bronchi, trachea, or lungs	Lymphoid interstitial pneumonia or pulmonary lymphoid hyperplasia complex
Candidiasis of esophagus	Lymphoma, Burkitt
Cervical cancer	Lymphoma, immunoblastic, primary brain
Coccidioidomycosis, disseminated or extrapulmonary	Mycobacterium avium complex or Mycobacterium kansasii, disseminated or extrapulmonary
Cryptococcosis, extrapulmonary	Mycobacterium tuberculosis of any site, pulmonary, disseminated, or extrapulmonary
Cryptosporidiosis, chronic intestinal	Mycobacterium, other species or unidentified species, disseminated† or extrapulmonary
Cytomegalovirus disease (other than liver, spleen, or nodes), onset at age > 1 month	Pneumocystis jiroveci pneumonia
Cytomegalovirus retinitis (with loss of vision)	Pneumonia, recurrent
Encephalopathy, HIV related	Progressive multifocal leukoencephalopathy
Herpes simplex: chronic ulcers (> 1 month's duration) or bronchitis, pneumonitis, or esophagitis (onset at age > 1 month)	Salmonella septicemia, recurrent
Histoplasmosis, disseminated or extrapulmonary	Toxoplasmosis of brain, onset at age > 1 month
Isosporiasis, chronic intestinal (> 1 month's duration)	Wasting syndrome attributed to HIV

Source: Schneider, E., Whitmore, S., Glynn, K.M., *et al.*, 2008. Revised surveillance case definitions for HIV infection among adults, adolescents, and children aged < 18 months and for HIV infection and AIDS among children aged 18 months to < 13 years—United States, 2008. MMWR Recommendations and Reports 57 (RR-10), 1–12.

Once a person becomes positive for anti-HIV-1 antibodies they will remain antibody-positive for life. However, combination antiretroviral therapy (see below) can fully suppress viral replication such that the clinical tests for viral RNA and viral capsid protein become negative. For HIV-1, even though these two measures of virus turn negative the virus is still present in a latent state, and for most people the virus will "rebound" in the blood within 2–4 weeks if antiviral therapy is stopped. Thus people with HIV-1 must initiate life-long therapy to stop viral replication (and disease progression) and keep the virus from reappearing in the blood.

Clinical Features

There is an early feature of infection with HIV-1 that occurs shortly after transmission, although its frequency varies greatly among different cohorts. About 2–4 weeks after infection there can be transient flu-like symptoms associated with the initial host immune response. As the viral load is brought down to its set point these symptoms resolve. When symptoms appear at this stage they are referred to as acute infection syndrome.

As these early symptoms resolve, a person transitions into a state of clinical latency where the virus is replicating and damaging the immune system but not yet to a point where symptoms occur. It is the cumulative damage to the immune system that leads to the defining feature of an HIV-1 infection, i.e., immunodeficiency. In the absence of antiviral treatment, there is an ongoing loss of CD4⁺ T cells in the body, a cell type that is central to coordinating the host immune response. As the CD4⁺ "helper" T cells are lost (we start with around 1000 such cells per microliter of blood) the host becomes increasingly susceptible to infections from other viral, bacterial, and fungal agents as well as virus-induced cancers (Table 2). Ultimately it is these other infectious agents or tumors that lead to death. The clinical state of immunodeficiency is defined as the appearance of one of these opportunistic infections or when the CD4⁺ T helper cell count drops below 200 cells per microliter in the blood (which defines a state of enhanced risk of opportunistic infection).

Two features of this process are highly variable. Some people progress rapidly to a state of immunodeficiency (1–2 years) while others progress slowly (15–20 years) or not at all, there being a small group of people (1%) who have a very low set point with no evidence of disease progression. Overall the average time from infection to the onset of immunodeficiency is in the range of 8–10 years.

The other feature that is highly variable is what happens as a person becomes immunodeficient. Some of the infections are acquired from our environment and are agents that we are exposed to regularly but control in an immunocompetent state. Others are agents that we are already infected with and normally control in a latent state but then cannot. A person experiences only a subset of these infections but we have little understanding why the passage into an immunodeficient state can have such varied consequences. However, individually or together these infections present serious challenges for the body. Under certain circumstances HIV-1 itself may become the opportunistic infection. In untreated infection a condition named AIDS-associated dementia ensues that involves replication of the virus within the central nervous system and the adaptation of the virus to that environment.

In this case the virus expands from replication in blood and lymphoid cells into the CNS where it is associated with a distinct clinical outcome.

Pathogenesis

There is one overarching principle in HIV-1 pathogenesis - the virus infects and kills CD4⁺ T cells and these cells are lost over time until a state of immunodeficiency is reached. It is tempting to make viral infection and loss of these cells the sum total of viral pathogenesis. However, there are many complex changes in the body that occur over the course of an untreated infection and it has been difficult to determine whether these ancillary changes contribute to pathogenesis or are the result of the immunodeficient state induced by the cumulative loss of CD4⁺ T cells. For example, it appears that too few CD4⁺ T cells are infected and killed by the virus to account for the magnitude of loss of these cells over time. Thus there have been studies of “standby” effects where other mechanisms contribute indirectly to the loss of CD4⁺ T cells and thus ultimately contribute to the immunodeficient state. Ideas that have been explored include a loss of the ability to produce new CD4⁺ T cells after years of ongoing replacement, loss of normal architecture in lymphoid tissue where most T cells reside, and local effects where cells near an infected cell also die.

Even the mechanism of how an infected cell dies is unclear. The viral protein Vpr induces cell cycle arrest which precludes the cell from dividing. The viral Env protein may continue to bind its CD4 protein receptor in the infected cells inducing aberrant signaling (several viral proteins down regulate CD4, perhaps for this reason). Budding of viral particles from the cell's plasma membrane could cause loss of membrane integrity. Virus replication may induce one of several “cell death” pathways within the cell.

Viral evolution within an infected person is a key feature of viral pathogenesis. The virus replicates every one to two days within a cell and then must jump to a new target cell to continue replicating. At steady-state (i.e., with a constant set point) one new cell is infected for each cell that dies. If the immune system were fully competent we might expect it to win over the virus and clear the virus from the body, as happens with most RNA viruses (such as influenza virus). In contrast, HIV-1 sustains replication in the face of an active immune response. It is possible that the virus blunts the host immune response specifically as the virus replicates most efficiently in activated cells, such as those CD4⁺ T cells responding to HIV-1 itself. Given the impaired host response, the virus is able to maintain a sufficiently large population size that, with its short replication time and relatively error prone replication apparatus, allows sequence diversity to accumulate in the viral population. Out of this diverse population variants that are able to escape the host's targeted immune response grow out, and this process is repeated over the course of the infection as the host response changes and the virus continues to evolve.

Broadly, there are two types of adaptive (specific) immune responses the host mounts: CD8⁺ T cells can target short sequences in viral proteins that are distributed among all of the viral proteins as they are “presented” on the surface of an infected cell (recognized as foreign resulting in the killing of that cell); alternatively antibodies target the surface proteins of the virus particle, for HIV-1 this is the Env protein. Since the Env protein is the target of both CD8⁺ T cells and antibodies this explains why this protein is the most variable in the viral population, undergoing the most evolution to escape both arms of the host immune response.

Viral evolution plays a role in several other features of replicating virus beyond immune escape: expansion of the host range of cell types that can be infected, and the evolution of drug resistance. In expanding its host range HIV-1 goes through distinct evolutionary pathways for its use of both its receptor CD4 and its chemokine coreceptor.

Coreceptor Switch

The predominant form of HIV-1 uses the host chemokine receptor CCR5 as its coreceptor. Late in disease as the CD4⁺ T cell population is lost the probability of encountering a preferred target cell decreases. Under these circumstances there is selective pressure to be able to replicate in a different cell type. There is a subpopulation of CD4⁺ T cells that express CXCR4 but not CCR5. HIV-1 evolves to be able to use CXCR4 as its coreceptor thus expanding the number of cell types it can infect. Such viruses are typically seen late in untreated disease and correlate with more rapid disease progression, although it is not known if this is because of the expanded host range or as an indirect marker of the loss of CD4⁺/CCR5⁺ T cells. The coreceptor switch involves evolution of sequences within the viral Env protein, which is responsible for interacting with the coreceptor during viral entry.

Macrophage-Tropic HIV-1 and Infection of the CNS

One of the important AIDS-defining illnesses that became apparent shortly after the AIDS syndrome was identified was HIV-associated dementia (HAD), which appears in about 25% of people with untreated infections. The appearance of HAD focused attention on a potential role for HIV-1 in infection of the central nervous system (CNS). The brain/CNS is somewhat immune privileged and as a result there are not very many CD4⁺CCR5⁺ T cells to support infection. However, there are two other cell types in the CNS that express CD4 and CCR5: the endogenous myeloid cell microglia and inflammatory macrophages that have trafficked across the blood-brain barrier into the CNS (perivascular macrophages that appear around blood vessels in the brain). However, there is one big and important difference with these cell types compared to CD4⁺ T cells, the density of CD4 on

microglia and macrophages is about 25 times lower than it is on T cells. As a result, the normal form of HIV-1 (R5 T cell-tropic) is very inefficient at infecting these myeloid cells. This leads to the second evolutionary pathway to expand the host range of infected cells - evolution to efficiently use a low density of CD4 for entry, which we associate with macrophage tropism (R5 M-tropic). There are claims that HIV-1 can infect yet another cell type in the brain, astrocytes, although this is hard to reconcile with the fact that astrocytes do not express the viral receptor CD4.

While CD4⁺ T cells are not abundant in the CNS they are always trafficking from the blood at a low level and can be found in cerebral spinal fluid (CSF). Some of these cells are infected thus the brain is constantly being exposed to virus that is equivalent to the virus in the blood (R5 T cell-tropic). It has been hypothesized that monocytes can carry virus into the CNS, where they differentiate into tissue macrophages. However, the low density of CD4 on monocytes in the blood makes these cells poor targets for infection by the R5 T cell-tropic form of the virus found in blood. Thus trafficking CD4⁺ T cells seem like the more likely culprit for introducing virus into the CNS.

While the CNS gets exposed to virus in all people via infected T cell trafficking, it is only in a subset of people where there is detectable replication of HIV-1 in the CNS. There appears to be little movement of virus from the CNS back to the blood (at least that is detectable), thus when virus starts replicating in the CNS it appears as a new genetic lineage of the virus that is distinct from the virus in the blood. This is called a compartmentalized viral population. In the case of the CNS this is usually associated with the ability to use CD4 more efficiently for entry, i.e., the evolution of macrophage tropism. Between 10% and 20% of people have evidence of virus growing in the CNS within the first two years of untreated infection (seen as genetically compartmentalized virus in the CSF compared to the blood). This rises to 40% of people in late stage untreated infection, and includes most of the people who are diagnosed with HAD. An active area of research is determining the extent to which virus can persist in the CNS during suppressive antiretroviral therapy.

HIV-associated dementia represents the end-stage of CNS pathogenesis, however it is now clear there is a broader spectrum of CNS-related disease states that can be measured and are collectively called HIV-associated neurocognitive defects (HAND). HAND is subclassified in the clinic based on symptomatic presentation and severity as either Asymptomatic Neurocognitive Impairment (ANI), Mild Neurocognitive Disorder (MND), or the most severe HIV-associated Dementia (HAD). These states are seen in people with untreated infections but there is concern that time spent untreated either leaves a person in one of these states even after therapy starts or makes it more likely they will progress even while on therapy. For people on antiretroviral therapy, the frequency of HAD has greatly decreased, occurring in only 2%–8% of treated people; however, ANI and MND have a prevalence of 30% and 20%–30%, respectively. As a general rule drugs are not efficient at crossing the blood-brain barrier leaving drug concentrations suboptimal in the CNS. With less potent drugs a phenomenon of CNS escape has been seen where virus is replicating within the CNS, sometimes with associated symptoms, while being suppressed in the blood. This is less of a problem with the recent more potent regimens.

Evolution of Drug Resistance

With a large population size of virus within the host, a rapid rate of replication (300 times per year), and an error prone DNA polymerase (reverse transcriptase, about 1 nucleotide substitution in every three genomes synthesized), the virus generates a lot of sequence diversity. Most nucleotide substitutions (mutations) are deleterious and are lost from the population, but they are constantly reappearing with the error prone polymerase. However, as selective pressures change then minor variants in the population may suddenly become more fit and grow out. This is the case with drug resistance. If drug concentration or potency is too low and allows continued viral replication then resistant variants will be selected causing the drugs to lose potency. The success of highly potent drug regimens is that they quickly shrink the size of the pool of replicating virus which blocks the evolution of drug resistance to multiple drugs at one time. Ironically, the same three drugs that together block viral replication in a typical antiretroviral regimen, if given sequentially would simply select for resistance.

Latency

Even when viral replication is completely blocked by suppressive therapy (see below), HIV-1 can persist in the body, referred to as a latent infection. In this state the virus exists as integrated DNA that is not transcribed. Expression of viral DNA into RNA and virus particle release happens stochastically. When suppressive antiviral therapy is present this transiently expressed virus cannot replicate and is lost. However, if therapy is discontinued this low level of stochastically released virus restarts the systemic infection. Latently infected cells appear to be a biological accident but with significant consequences. The virus has no "program" to become latent, its goal is to replicate as fast as possible. Thus when therapy almost clears all of the infected cells it is only a small number of long-lived cells that got infected but are not expressing the viral genome that accidentally create this reservoir. However, this latent reservoir decays so slowly that it is present for the entire life of the infected person. There is great interest in the research community to find ways of removing this latent reservoir so that a person could stop taking antiretroviral therapy and not experience reappearance of the virus. This has been accomplished several times in the context of an allogeneic transplant of blood cells in the context of cancer therapy where the patient was given new blood cells from a donor who had both copies of their CCR5 gene mutated, making their new transplanted CD4⁺ T cells resistant to infection. While this represents a "proof-of-concept" for an HIV-1 cure, this treatment is too extreme to use in the general population.

Table 3 FDA approved antiretrovirals for the treatment of HIV-1 infection. There are currently six classes of drugs approved for the treatment of HIV-1 infection, including Non-Nucleotide Reverse Transcriptase Inhibitors, Nucleotide Reverse Transcriptase Inhibitors, Protease Inhibitors, Fusion Inhibitors, Entry Inhibitors, and Integrase Inhibitors. Another drug class worthy of noting is Latency Reversing Agents, which aim to activate viral expression in latently infected cells to allow immune-mediated clearance, but these agents are still being investigated in the clinic and have yet to be approved by the FDA for the treatment of HIV-1

Nucleoside reverse transcriptase inhibitors		Protease inhibitors	
Brand name	Generic name	Brand name	Generic name
Combivir	lamivudine and zidovudine	Agenerase	amprenavir, APV (no longer marketed)
Emtriva	emtricitabine, FTC	Aptivus	tipranavir, TPV
Epivir	lamivudine, 3TC	Crixivan	indinavir, IDV,
Epzicom	abacavir and lamivudine	Fortovase	saquinavir (no longer marketed)
Hivid	zalcitabine, dideoxycytidine, ddC	Invirase	saquinavir mesylate, SQV
Retrovir	zidovudine, azidothymidine, AZT, ZDV	Kaletra	lopinavir and ritonavir, LPV/RTV
Trizivir	abacavir, zidovudine, and lamivudine	Lexiva	Fosamprenavir Calcium, FOS-APV
Truvada	tenofovir disoproxil fumarate and emtricitabine	Norvir	ritonavir, RTV
Videx EC	enteric coated didanosine, ddl EC	Prezista	darunavir
Videx	didanosine, dideoxyinosine, ddl	Reyataz	atazanavir sulfate, ATV
Viread	tenofovir disoproxil fumarate, TDF	Viracept	nelfinavir mesylate, NFV
Zerit	Stavudine, d4T	<i>Fusion inhibitors</i>	
Ziagen	abacavir sulfate, ABC	Fuzeon	enfuvirtide, T – 20
<i>Nonnucleoside reverse transcriptase inhibitors</i>		<i>Entry inhibitors</i>	
Edurant	rilpivirine	selzentry	maraviroc
Intelence	etravirine	<i>Integrase inhibitors</i>	
Rescriptor	delavirdine, DLV	Isentress	raltegravir
Sustive	efavirenz, EFV	Tivikay	dolutegravir
Viramune	nevirapine, NVP	Vitekta	elvitegravir

Treatment

The current standard of care for people living with HIV is combination antiretroviral therapy, or ART. Over a 15-year period we went from the identification of HIV-1 as the cause of AIDS (early 1980s) to the identification of the first active drug (AZT, late 1980s), to the discovery of potent three drug regimens capable of suppressing viral replication and stopping the disease course (late 1990s). Over the last 20 years drug potency has continued to increase and newer formulations allow drug delivery to become increasingly convenient, now one pill once a day for most people. Modern regimens allow people to live a near normal life with a full expected life span. However, HIV-1 is maintained as a latent infection even in the face of inhibitors that block all viral replication. This results in the need to take antiretroviral therapy throughout life since stopping typically results in the rebound/reappearance of high levels of replicating virus in less than a month. There are four main classes of antiretrovirals used to treat HIV, each targeting a different stage of the virus life cycle. These include: Entry Inhibitors, Reverse Transcriptase Inhibitors, Protease Inhibitors, and Integrase Inhibitors. A full list of current FDA-approved drugs for the treatment of HIV-1 can be found in [Table 3](#).

Entry Inhibitors

Entry inhibitors block entry of the virus into the cell. One type of entry inhibitor binds to the host cell coreceptor CCR5 preventing its interaction with the viral Env protein (e.g., maraviroc) thus blocking the transition to the membrane fusion function of gp41. Another drug (enfuvirtide) is a peptide mimetic of a portion of the gp41 sequence that binds to block formation of the six helix bundle, thus preventing fusion. As a peptide this particular drug must be injected and thus is used only infrequently in people who have virus that is extensively drug resistant to other regimens.

Reverse Transcriptase Inhibitors

Reverse Transcriptase Inhibitors (RTIs) target the viral DNA polymerase reverse transcriptase, which is packaged in the virus particle and is responsible for reverse transcribing the single-stranded viral RNA genome into double-stranded DNA in the newly infected cell. RTIs can be subdivided into two classes, nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs). NRTIs are nucleoside or nucleotide analogs that are missing the 3'-OH on the ribose. This OH- is used to extend the growing DNA chain as each nucleotide precursor is added. Thus when one of the NRTIs gets incorporated into the growing viral DNA chain it terminates synthesis since the growing point now cannot get extended with the

next nucleotide. NNRTIs are non-nucleoside analogs that bind in a pocket in RT to allosterically prevent enzymatic activity. NRTIs commonly comprise two-thirds of the three-drug cocktail prescribed as ART, with the third drug historically being a boosted protease inhibitor, an NNRTI, or more recently an integrase inhibitor.

Protease Inhibitors

Protease Inhibitors (PIs) bind the active site of the viral protease, which is responsible for cleaving the HIV-1 polyproteins, namely the Gag and Gag-Pro-Pol precursors, into their mature and active protein domains. Protease inhibitors are peptide mimetics of the cleavage sites present in the precursor proteins. Inhibition of the viral protease prevents maturation of the viral proteins, leaving virions in either an immature state or a state of mixed processed and unprocessed proteins which results in a noninfectious virion. An early PI (ritonavir) had the fortuitous property of binding and inhibiting the enzyme responsible for its clearance, cytochrome P450-3A4, and in so doing greatly extended the half-life of the drug in the body. Ritonavir (or another drug with similar properties) is now often given in lower doses as a "booster" to increase the half-life of other drugs in the regimen that are metabolized by P450-3A4.

Integrase Inhibitors

Integrase has a series of conserved acid amino acids that coordinate the binding of a Mg^{2+} ion through an acidic and polar amino acid interface. This interface orients the Mg^{2+} ion allowing it to be involved in a nucleophilic attack of a host DNA phosphodiester bond by the viral 3' hydroxyl, resulting in integration of viral DNA into the host DNA via a strand transfer reaction. Integrase strand transfer inhibitors, or ISTIs, bind to the same Mg^{2+} to get oriented within the active site to inhibit the strand transfer reaction and, consequently, integration. ISTIs have recently emerged as the drug of choice over boosted protease inhibitors and NNRTIs in most ART regimens due to their increased potency, which results in less undesirable side effects and a lower probability that the virus will develop resistance mutations.

Prevention

Of the 37.9 million people currently living with HIV-1, approximately 60% are currently receiving ART. Further, the percentage of people living with HIV-1 and on ART is expected to continually rise as governments and health organizations around the world increase efforts to achieve the '90-90-90 Goal', an initiative from the World Health Organization that aims to have 90% of all infected individuals aware of their status, 90% of those aware of their status administered ART, and 90% of those administered ART to be virally suppressed. While diagnosis, treatment, and adherence are each crucial elements for controlling the HIV-1 epidemic, prevention efforts will also contribute significantly to making the 90-90-90 initiative successful. In this regard, Treatment as Prevention (TasP) has been one of the most effective prevention strategies in that people who are successfully suppressed on therapy have essentially a zero chance of transmitting the virus.

Biomedical Prevention

Biomedical intervention is key to preventing the transmission of HIV-1 and further controlling the epidemic. There are numerous biomedical interventions that have been shown to decrease the rate of transmission. For example, circumcision has been shown to decrease a male's risk of acquiring HIV-1 by 50–60%. HIV-1 prevalence tends to be higher in countries where male circumcision at birth is not standard medical practice. Thus, encouraging voluntary male circumcision could potentially prevent a significant number of new infections, if there was significant uptake. Additionally, female-oriented prevention strategies such as ART-containing vaginal gels as prophylaxis have been shown to decrease transmission rates up to 54% in clinical trials, although equivalent real-world efficacy of these gels has been harder to realize.

For many viral diseases there is a vaccine that allows control of virus spread and is associated with a reduced disease course. The development of a vaccine for HIV-1 has proven to be especially challenging. One large trial showed about 30% efficacy in preventing infection, enough to inspire hope that there may be a path forward but too low to be useful in its current form. The search for an HIV-1 vaccine is currently driving an intense effort to understand how the body interacts with foreign antigens and to define how best to deliver these antigens to the immune system.

What is PrEP?

The idea of Pre-Exposure Prophylaxis, or PrEP, as a prevention strategy grew from the observation that HIV-infected women on ART rarely transmit to their offspring during birth, as long as the women are virally suppressed. PrEP is treatment with anti-retrovirals prior to exposure to decrease the risk of acquiring HIV-1 if exposed. During acute infection, HIV-1 first replicates in

mucosal tissue at the point of entry, such as the genital or rectal tract. The presence of antiretroviral drugs in these tissues prevents the initial replication of the virus leaving the person uninfected. PrEP typically includes one or two antiretroviral drugs rather than the three used for suppressive treatment of an infection. The ongoing development of long acting drug delivery systems (lasting weeks to months) is likely to have a significant impact on the uptake of PrEP in people at increased risk of exposure to HIV-1. Another new line of prevention research is to use broadly neutralizing monoclonal antibodies to neutralize incoming virus at exposure. If such an approach proves efficacious then genes for these antibodies could potentially be introduced using gene therapy vectors so that each person will be able to make these antibodies over an extended period of time.

Social and Educational Prevention

While biomedical interventions are crucial tools for preventing HIV-1 transmission, there will continue to be an important role for social and educational prevention initiatives. It is clear that greater access to information regarding sexual health and safe sex practices, reducing risk associated with sharing injection equipment, and knowledge about basic HIV-1 transmission biology can significantly impact transmission rates.

In countries such as Zimbabwe, Uganda, Zambia, Malawi, Kenya, and Thailand, initiating behavioral interventions has been successful in decreasing transmission rates. Particularly, improved sexual education courses, more thorough and informed HIV-1 testing and counseling, and well-organized, state-run prevention programs such as Thailand's '100% Condom' program, which promotes the use of condoms in sex work, have contributed to a decline in transmission. Unfortunately, many of these programs have yet to fully reach at-risk populations. Stronger political representation and funding opportunities will be crucial for expanding existing programs and creating new programs to aid these important populations. Ultimate success in fully controlling the HIV-1 epidemic will likely require some level of success in using all of these different approaches to prevention.

Further Reading

- Barré-Sinoussi, F., Chermann, J.C., Rey, F., *et al.*, 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220, 868–871.
- Bushman, F., Nabel, G., Swanstrom, R., 2012. HIV: From Biology to Prevention and Treatment. Cold Spring Harbor Laboratory Press.
- Coffin, J., Hughes, S., Varmus, H., 1997. Retroviruses. Cold Spring Harbor Laboratory Press.
- Crawford, D., 2013. Virus Hunt: The Search for the Origin of HIV. Oxford University Press.
- Flint, J., Racaniello, V., Rall, G., Skalka, A., 2015. Principles of Virology: 4th Edition. ASM Books.
- Joseph, S., Swanstrom, R., 2018. The evolution of HIV-1 entry phenotypes as a guide to changing target cells. *Journal of Leukocyte Biology* 103, 421–431.
- Levy, J., 2007. HIV and the Pathogenesis of AIDS. ASM Press.
- Margolis, D., Garcia, V., Hazuda, D., Haynes, B., 2016. Latency reversal and viral clearance to cure HIV-1. *Science* 353, aaf6517.

Relevant Websites

- <https://www.cdc.gov/>
Center for Disease Control and Prevention.
- <https://www.hiv.lanl.gov/content/index>
Los Alamos H.I.V. Database.
- <https://www.unaids.org/en>
unaids.
- <https://www.who.int/>
World Health Organization.

Human Metapneumovirus (*Pneumoviridae*)

Antonella Casola, Matteo P Garofalo, and Xiaoyong Bao, The University of Texas Medical Branch, Galveston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Capping Addition of a guanosine nucleotide to the 5'-end of mRNAs, which is then methylated by a methyl transferase.

Glycosaminoglycan Long, unbranched, highly negatively-charged polysaccharides that consist of repeating disaccharide subunits that are located on the cell surface, or in the extracellular matrix.

Lipid raft Membrane microdomains enriched in cholesterol and sphingolipids that modulate the membrane distribution of receptors and signaling molecules.

PAMP Molecules (lipids, proteins, nucleic acids) derived from invading pathogens that trigger specific receptor-mediated host responses.

RDG Peptide motif responsible for protein interaction with adhesion proteins called integrins.

Ribavirin A nucleoside analog (1-beta-D-ribofuranosyl-1,2,4-triazole-3-carboxamide) that has antiviral activity against a number of RNA viruses, including hMPV.

Syncytium A multinucleated cell, formed by fusion of the plasma membrane of an infected cell with those of its neighbors.

Transcriptional elongation Process through which nucleotides are added to a growing RNA chain during gene transcription.

Classification

Human metapneumovirus (hMPV) is a non-segmented, negative-sense (–), single-stranded RNA (ssRNA) virus, formerly a member of the *Paramyxoviridae* family. In 2016 it was reclassified into the family *Pneumoviridae* of the order *Mononegavirales*. This new family consists of two genera, *Metapneumovirus* and *Orthopneumovirus*, with the former including human and avian metapneumoviruses, and the latter human and bovine respiratory syncytial virus (RSV), as well as murine pneumonia virus (MPV). Based on genetic and phylogenetic analysis, hMPV is divided into two groups, A and B, with each group divided in two subgroups, i.e., A1 and A2; B1 and B2. Subgroup A2 is again subdivided into A2a and A2b. Despite differences in genetic sequences, there is significant antibody cross-reactivity and neutralization capacity and therefore protection against various genotypes. All hMPV strains share similar genome and virion structure and several different hMPV genotypes cocirculate each epidemic season, with the predominant genotypes varying from season to season.

Virion Structure and Virus Proteins

In tissue culture, hMPV virions consist of enveloped pleomorphic, spherical or filamentous particles, ranging from 150 to 600 nm in size, with short protein spike projections of approximately 13–17 nm.

HMPV encodes 9 proteins, classified as either integral membrane proteins, embedded in the virus envelope, or internal proteins, which associate with the virus nucleocapsid beneath the virus envelope, as illustrated in [Fig. 1](#). The transmembrane proteins, which represent the virion spikes, are glycosylated and include the fusion protein F, glycoprotein G, and small hydrophobic protein SH. The F protein, which is well conserved among strains, mediates fusion of viral and cellular membranes during hMPV entry, it induces syncytium formation in infected cells, and determines the cellular host range. In addition, it is a major immunogenic protein. The functions of the G protein include cellular glycosaminoglycans binding (strain-dependent), and host immunity regulation. The SH protein, whose function is still unclear, also plays a role in modulating innate immune responses in airway epithelial cells (AECs) and dendritic cells (DCs).

There are five internal proteins, four of which are part of the nucleocapsid and polymerase complex: the nucleoprotein N, the phosphoprotein P, the transcription regulatory proteins M2-1 and M2-2, and the RNA-dependent RNA polymerase L. The viral genome interacts with N, P, and L to form the virus ribonucleoparticle (RNP). The matrix protein M surrounds the RNP underneath the virus envelope, and, together with the F and G proteins, plays an important role in virus assembly and budding. A summary of hMPV proteins and their function is presented in [Table 1](#). A further description of the various functions of hMPV proteins is presented in subsequent sections of this article.

Genome

The hMPV genome consists of about 13,000 nucleotides, with a size ranging from 13,280 to 13,378 nucleotides depending on the genotype, and it contains 8 genes, which encode 9 proteins, with the order of viral protein expression as the following: 3'-N-P-M-F-M2-1-M2-2-SH-G-L-5' ([Fig. 2](#)). Among those, M2-1 and M2-2 are both encoded by hMPV M2 gene. The M2-1 open reading

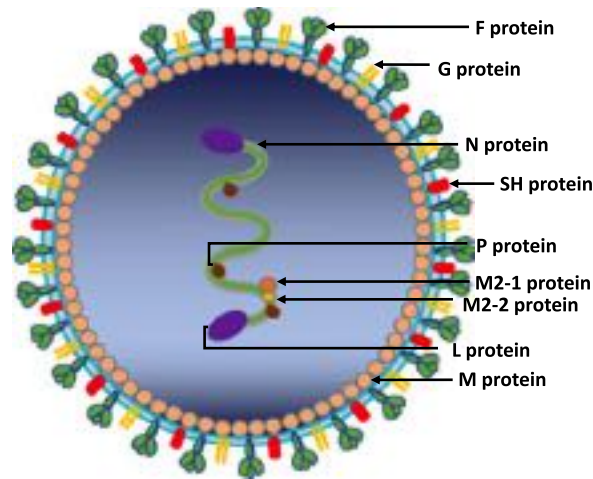


Fig. 1 hMPV virion structure. Schematic representation of name and location of hMPV proteins, either present in the envelope (F, G, SH) or underneath it (M), or associated with viral RNA (N, P, L, M2-1 and M2-2).

Table 1 hMPV proteins and function

Gene	Protein	Function
F	Fusion protein	Cell binding, membrane fusion, virus assembly
G	Attachment glycoprotein	Binding to cellular glycosaminoglycans, virus assembly, innate immune response inhibition
SH	Small hydrophobic protein	Innate immune response inhibition
N	Nucleoprotein	RNA genome encapsidation, virus intercellular spread
P	Phosphoprotein	Polymerase cofactor
M2	M2-1 protein	Essential for virus replication in vivo
	M2-2 protein	Viral RNA synthesis
L	polymerase protein	RNA-dependent RNA polymerase responsible for mRNA generation and genome replication
M	Matrix protein	Viral assembly and budding

frame (ORF) is presumed to start with the first AUG codon of M2, while the M2-2 ORF possibly initiates with the second or third AUG (s) of M2, overlapping the M2-1 ORF by about 50 to 40 nucleotides. HMPV RNA synthesis comprises two independent events: viral replication and viral gene transcription, which are carried out by three essential proteins: the nucleoprotein N, phosphoprotein P, and large protein L. In addition to these three proteins, studies using recombinant hMPV lacking M2-2 demonstrated that the M2-2 protein is also important for viral RNA synthesis. Whether hMPV M2-2 is a key regulatory factor involved in the switch of the viral RNA polymerase from viral gene transcription to viral genome replication, as with the RSV M2-2 protein, is still unknown. Although RSV M2-1 functions as a transcriptional elongation and antiterminator factor, and is essential for viral replication, hMPV M2-1 protein was found to be dispensable for virus viability or replication in tissue culture, although it was very important for replication in vivo, as its deletion resulted in a highly attenuated virus.

Life Cycle

The Entry of hMPV

The major cell target of hMPV infection is the airway epithelial cell, although it can infect other cell types, like macrophages and DCs, leading to abortive replication in those cells. hMPV F is sufficient for mediating virus entry, as hMPV lacking G or SH is still infectious, although hMPV B strain also uses the G protein to bind to glycosaminoglycans to improve viral fitness. HMPV F protein shares conserved functional domains with other pneumovirus F proteins, and requires cleavage from a precursor protein F0 into F1 and F2 as a prerequisite for fusion activity. In cell culture, it also requires the addition of exogenous proteases, different from the orthopneumovirus RSV. HMPV F attaches to cellular receptors, including integrins, via interaction with its Arg-Gly-Asp (RDG) sequence, and catalyzes virus membrane fusion with cellular membranes during virus entry. In addition to plasma membranes, hMPV particles have been shown to be internalized by clathrin-mediated endocytosis and then fuse with endosomal membranes. Engagement of RGD-binding integrins has been shown to be required for endosomal trafficking and full virus membrane fusion with intracellular membranes. Although low pH exposure was initially thought to be necessary for hMPV fusion, inhibition of

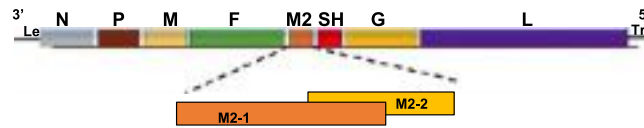


Fig. 2 HMPV genomic organization. Schematic representation of hMPV genome, illustrating gene order from the 3' to the 5' end and the M2-1 and M2-2 open reading frames both encoded by hMPV M2 gene.

endosomal acidification has only a modest effect on fusion for many hMPV viruses, and this event appears to be strain-specific. In addition to clathrin-mediated endocytosis, there is evidence that hMPV enters target cells via lipid rafts. When lipid raft inhibitors, such as cyclodextrin and nystatin were used, hMPV cell entry was inhibited. Virus internalization in immune cells, such as monocyte-derived DCs, seems to occur via macropinocytosis and also via clathrin-mediated endocytosis through the interaction with the C-type lectin receptor DC-SIGN.

Viral RNA Synthesis

Upon entry, viral nucleocapsids are released into the cell cytoplasm, followed by synthesis of viral components. It is generally believed that the viral synthesis of hMPV is similar to the better known pneumovirus RSV. The viral RNA-dependent RNA polymerase (RdRp) complex, formed by the L and P proteins, is presumed to bind the N-encapsidated genome at the leader region, then sequentially transcribes each gene by recognizing the start and stop signals flanking the viral genes, leading to synthesis of mRNAs, which are capped and polyadenylated. Protein translation occurs in the endoplasmic reticulum, with transport of proteins through the Golgi apparatus, for post-translational modifications, mainly glycosylation, to sites of virus assembly. Genome replication presumably starts when enough nucleoprotein is present to encapsidate newly synthesized antigenome and genome copies. The negative-sense genome is replicated into a positive-sense antigenome, which serves as a template for synthesis of many copies of the viral genome. In addition to P and L, the M2-1 and M2-2 protein are also important for viral RNA synthesis, as mentioned above. Recent structural studies on P protein revealed it to be a multifunctional hub which interacts with numerous components of the viral RNA-synthesis machinery, including the RNA-free N, the M2-1, the L polymerase, and the nucleocapsid, demonstrating its importance in viral RNA synthesis. Experimentally, colocalization of the virus -RNA, +RNA, mRNA, as well as, N and P proteins was found to occur in specific cytoplasmic structures called inclusion bodies (IBs), but not in others – such as stress granules, or processing bodies. This strongly suggests that IBs are the sites of active hMPV replication and gene transcription, similar to other negative sense RNA viruses, such as filoviruses and rhabdoviruses.

Viral Particle Assembly and Spread

From IBs, newly synthesized nucleocapsids are transported to assembly sites, likely via actin cytoskeleton, to be incorporated into nascent virions. Similar to human RSV (hRSV) and paramyxoviruses, the M protein plays a fundamental role in hMPV morphogenesis, however, its presence is not sufficient for the formation of virus-like particles, which requires the presence of the F, G and possibly N proteins. The fully assembled virions are then released into the extracellular environment from the apical surface of polarized infected cells. The process of budding and the release of hMPV has not yet been elucidated. Many enveloped viruses, including several paramyxoviruses, use the endosomal sorting complex required for transport (ESCRT) machinery for budding. However, recent studies have shown that hRSV does not utilize this pathway, similar to influenza A virus and alphaviruses, but employs a Rab11-dependent mechanism, which is part of the apical recycling endosomal pathway. Similarly, the ESCRT is not required for budding of the avian metapneumovirus C, which shares about 80% homology with hMPV, suggesting that it is unlikely to be involved in hMPV budding, as well. Viral RNA, and other viral components, can also be transferred to other cells via cell-to-cell syncytia formation or via formation of intercellular extensions, which have been shown to be covered by branching filaments that contain viral proteins and RNA, likely corresponding to filamentous hMPV virions. A summary of the virus life cycle is presented in Fig. 3.

Epidemiology and Clinical Features

hMPV was first discovered in the Netherlands in 2001. Through electron microscopy and real-time reverse transcriptase-polymerase chain reaction (RT-PCR), researchers identified this new viral agent in nasopharyngeal specimens collected from children with a respiratory illness. Since then, hMPV has been isolated in samples derived from a variety of age groups worldwide, with a higher prevalence in children younger than 5 years and adults older than 65 years of age. Although the disease burden for hMPV is lower compared to hRSV, the other important human virus member of the *Pneumoviridae* family, hMPV remains a cause of considerable morbidity and mortality, currently estimated at 33 million cases/year worldwide.

The virus is believed to have circulated for a very long time, as hMPV-neutralizing antibodies have been detected in stored serum samples from the 1990s and as far back as the 1950s. From seroconversion studies, we know that in the first few months of

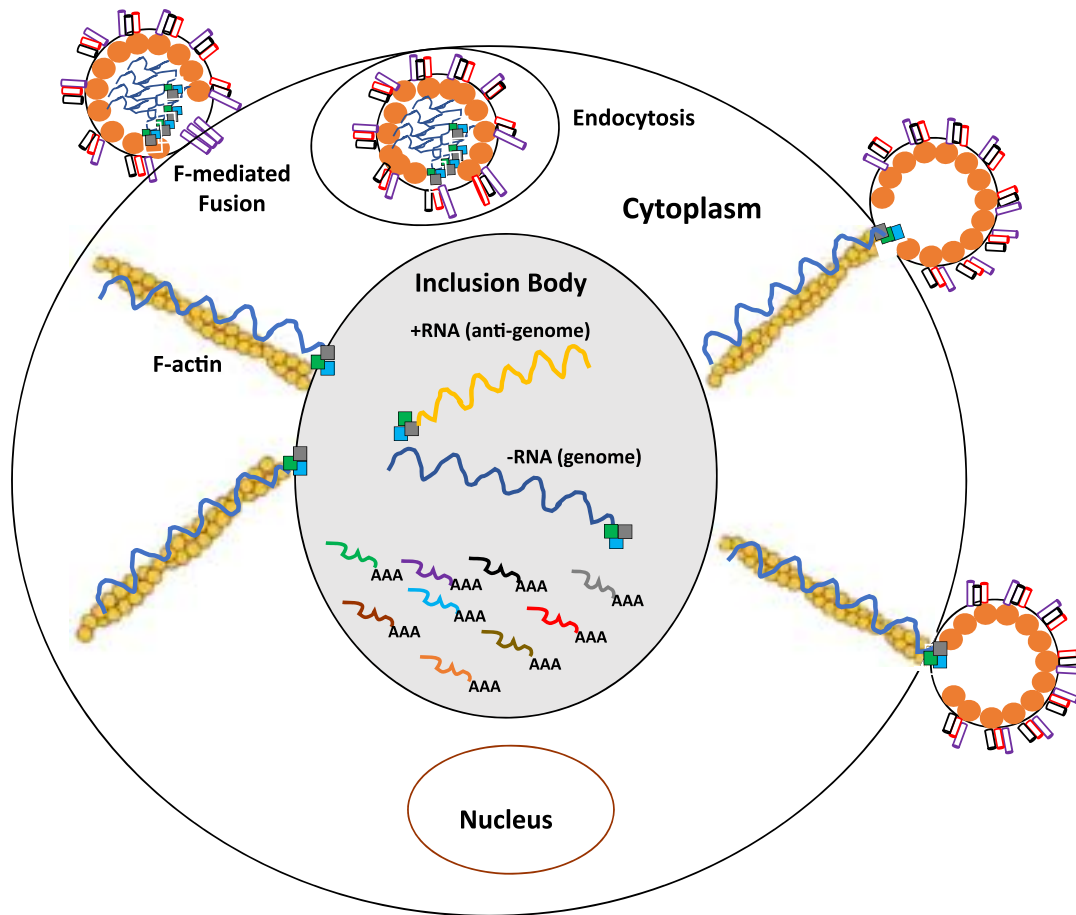


Fig. 3 Schematic diagram of hMPV life cycle. Schematic representation of main steps of hMPV replicative cycle, from virus entry via endocytosis to RNA synthesis in specialized structures known as inclusion bodies.

life most children have maternally derived antibodies, and that primary infection occurs predominantly in the first two years of life, followed by common reinfections during childhood, with positive IgG and IgM detected during an acute respiratory illness.

In a similar fashion to RSV, and most viruses associated with respiratory illnesses, hMPV infections show a seasonal cycle which spans between November and May in temperate climates, with the peak of disease mostly occurring during the months of March and April. In recent surveillance studies conducted in the US from 2008 to 2014, the data confirmed previous observations that hMPV season occurs later than that of RSV, with the RSV season occurring first, followed by influenza, and then hMPV. In addition, the data demonstrated a biannual pattern of early and late seasons, which did not occur for RSV.

Different hMPV genotypes can circulate during the same epidemic year, and severe infection forms may be associated with different lineage viruses throughout the year. Some studies have suggested a correlation of specific genotypes with disease presentation, with hospitalization and diagnosis of pneumonia being more frequent in patients infected with genotype A, and wheezing with genotype B, however, there is no significant correlation with disease severity. The level of viral replication may also affect the clinical picture. Some studies have reported that, in the absence of a preexisting condition or immunosuppression, the disease severity and the duration of illness were associated with high hMPV viral load in the lower respiratory tract, although other studies showed conflicting results.

The incidence of hMPV infection is higher early in life, in fact it is two times more common among children before 2 years of age, compared to those between 2 and 5 years of age. Hospitalization is also most common during the first year of life. Among infants and very young children, prematurity and pre-existing heart and lung conditions have been identified as predisposing factors for a severe disease. The estimated prevalence of hMPV infection in hospitalized children is between 5% and 7%, with some studies reporting up to 10% of all pediatric hospitalizations for respiratory illnesses, with a rate of 1.2 hospitalizations per 1000 children <5 years old per year. In the outpatient population, the impact of hMPV on acute respiratory illnesses in the pediatric population becomes as high as 15% during the peak season.

hMPV infections in otherwise healthy young adults are usually mild and often go undetected, but it is known that reinfection occurs frequently during adult life, and the majority of adults have stable titers of neutralizing antibodies. Severity of disease increases again in the elderly, with higher morbidity and mortality rates, and hospitalization rates as high as 60%.

In addition to children and the elderly, patients with preexisting conditions such as pulmonary neoplasia, lymphoma, chronic obstructive pulmonary diseases or asthma are also at greater risk of developing pneumonia and respiratory failure. The relationship of hMPV to asthma is still under investigation, but current evidence seems to support both higher prevalence of hMPV in patients hospitalized due to acute asthma exacerbations, and a higher risk of developing asthma later in life for those patients who experienced severe disease from hMPV infection during childhood.

Coinfection with other respiratory viruses or bacteria may also occur. Dual infection with RSV has been more commonly reported, and there are conflicting results on whether virus co-infections lead to increased disease severity. Although less common than infection with influenza, superimposed bacterial pneumonia, whether a nosocomial infection or community-acquired, has been reported, which leads to increased mortality rate.

Infection with hMPV is acquired through proximity or contact with infected subjects, mostly via respiratory secretions (droplets) or fomites. After an incubation period of 4–6 days, patients usually develop symptoms associated with upper respiratory tract infection (URTI), which may include fever, sore throat, cough, and development of acute otitis media. Patients who progress with the development of a lower respiratory tract infection (LRTI) show tachypnea, dyspnea, wheezing and hypoxia. High fever, found frequently among hospitalized children, is less frequent in adults, especially compared to infection with RSV. In rare cases, seizures and encephalitis have also been reported in association with hMPV. The virus can be isolated in secretions for up to 14 days after the onset of symptoms, making person-to-person spread within families or communities common.

In the pediatric population, URTI symptoms last on average between 7 and 10 days, while LRTI requiring hospitalization have a longer duration. Bronchiolitis and interstitial pneumonia are usually confirmed by radiologic diagnosis. In systematic evaluations of the radiographic characteristics of acute respiratory illnesses associated with pediatric hMPV infection, peribronchial/perihilar opacities were the most commonly observed abnormality, occurring in 87% of children. Hyperinflation also occurred frequently (69%), with atelectasis (40%) and consolidation (18%) appearing less frequently. Subjects with different clinical diagnoses had different radiologic findings, e.g., children with pneumonia, were more likely to have consolidation and perihilar/peribronchial opacities, while children with asthma and bronchiolitis were more likely to have hyperinflation. When additional laboratory tests are performed, complete blood counts may show lymphopenia with or without neutrophilia, and abnormal chemistries, namely elevated transaminases on liver function tests.

In immunocompromised patients, and the elderly, the presentation of hMPV LRTI is usually pneumonia, which may be severe. Because of increased transmission and severe disease, the mortality rate associated with hMPV infection is higher in elderly patients confined within long-term care institutions.

Pathogenesis

Cellular Responses

The airway epithelium represents the primary infection site of respiratory viruses, including hMPV. After infection, epithelial cells, together with lung resident cells – such as alveolar macrophages, secrete a variety of innate immune molecules, including antiviral substances and proinflammatory and immunoregulatory cytokines and chemokines, which initiates specific hMPV immune responses, both innate and adaptive, that are protective, but also in part, pathogenetic. For example, the depletion of alveolar macrophages have been shown to decrease hMPV replication, lung inflammation and disease severity in a mouse model of infection. Similarly, type I interferon (IFN) secretion, which plays a role in controlling hMPV replication, is also a determinant of a disease, as mice lacking type I IFN receptor have less lung inflammation and airway dysfunction. Increased levels of cytokines and chemokines have been shown to correlate with clinical disease in animal models of infection. For example, the thymic stromal lymphopoietin (TSLP) has been recently found to induce pulmonary inflammation and enhance hMPV replication in mice, and to correlate with wheezing and disease severity in young children with the infection. IL-17 has been shown to suppress regulatory T cell recruitment, increase Th1 and Th2 cells and neutrophil lung influx, and be involved in airway dysfunction in the course of hMPV infection, suggesting that IL-17 skews the immune response in the lung toward an inflammatory profile.

Studies in mice have shown that neutrophils are the predominant cell type recruited to the airways early during hMPV infection. In this animal model, neutrophil depletion is associated with increased pulmonary inflammation and severe clinical disease, without changes in viral clearance, and with increased recruitment of lung $\gamma\delta$ T cells, suggesting that the latter cells contribute to hMPV-induced disease pathogenesis.

T cell responses are required for virus clearance in hMPV-infected mice, together with neutralizing antibody production, however, they also contribute to pathology. Depletion of CD4⁺ T cells reduces lung pathology and airway obstruction, without affecting viral replication, although it is associated with impaired generation of neutralizing antibodies. This impairment does not affect the protection against hMPV reinfection, suggesting that protection can be provided by an intact CD8⁺ T cell compartment. Cytotoxic CD8⁺ T lymphocytes have been shown to be capable of clearing hMPV in Rag^{-/-} mice, underlying their importance in hMPV protection, however, they have also been shown to contribute to pathogenesis.

Immunity to hMPV is incomplete, as reinfections can occur throughout the life. One mechanism utilized by hMPV to evade the adaptive immune response is the upregulation of programmed cell death-1 (PD-1) receptor on T cells and its ligand, PD-L1, in the lung, leading to functional impairment of hMPV-specific CD8⁺ T cells. Inhibition of PD-1 signaling, or lack of the receptor in a

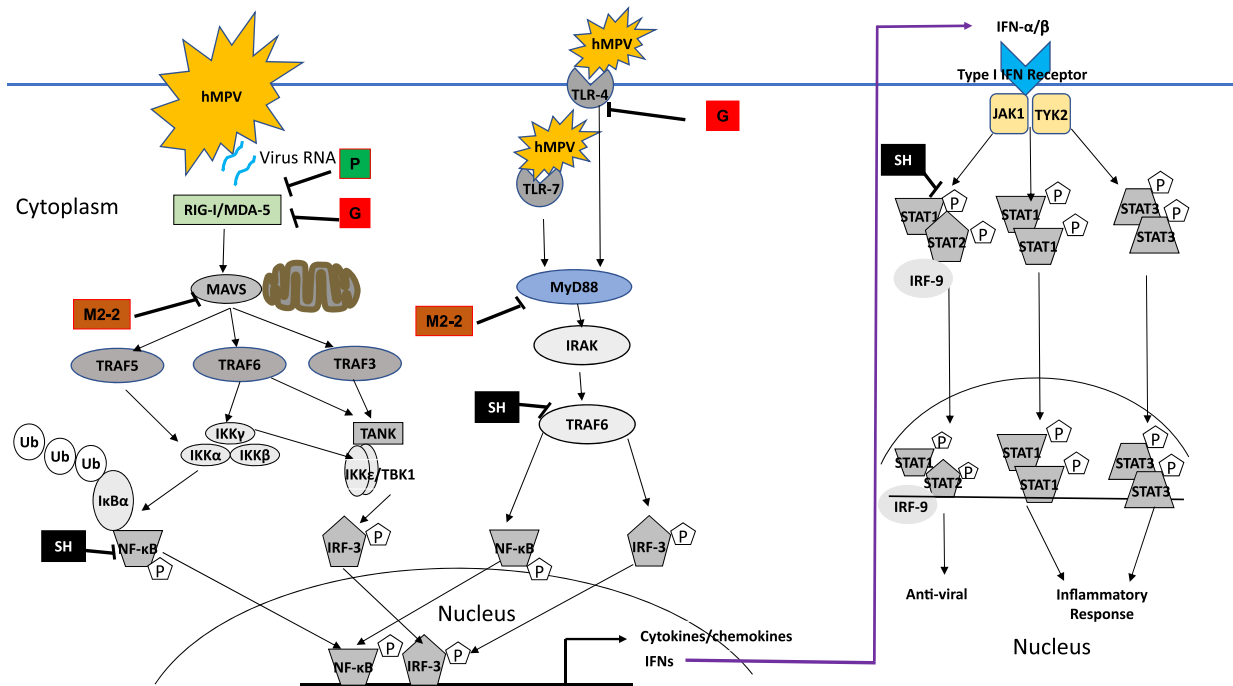


Fig. 4 Schematic diagram of sites of antagonism of cellular signaling and IFN production/responses by hMPV proteins. Major cellular targets of hMPV proteins in the TLR4, RIG-I/MAVS and type I IFN signaling pathways. P indicates phosphorylation.

murine model of infection, result in a higher number of hMPV-specific CD8 + T cells, both during primary and secondary infections, suggesting that this pathway could play an important role in the ability of hMPV to reinfect the host.

Pattern Recognition Receptors (PRRs) and Antagonism of Immune Responses

Activation of the innate responses to infections is initiated by the recognition of specific pathogen-associated molecular patterns (PAMPs) derived from the invading pathogens, through a variety of host PRRs. Multiple PRR families, including RIG-I-like receptors (RLRs), Toll-like receptors (TLRs), and NOD-like receptors (NLRs), and their corresponding signaling adapters, have been shown to detect hMPV and initiate cellular responses to infection.

Retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated protein 5 (MDA5) are the two major RLRs activated by hMPV infection and they share a common adapter, the mitochondrial antiviral-signaling protein (MAVS), for signaling. The RIG-I/MAVS pathway has been shown to be important for the cytokine/chemokine/IFN induction by hMPV in AECs, while MDA5 is essential to induce IFN secretion in human and mouse DCs, as well as, in an experimental mouse model of infection, where it plays a protective role against a severe disease. The role of MAVS in hMPV-induced inflammatory and antiviral mediator secretion has also been investigated in mice, and it was found to be necessary for the generation of protective innate immune responses, through immune mediator production and recruitment and maturation of immune cells, which in turn, modulate activation of T cell responses.

Several TLRs have also been shown to be important in hMPV-induced host responses. In human monocyte-derived DCs, TLR4 plays an important role in the induction of cytokines, chemokines and type I IFNs following hMPV infection. Its absence, in a mouse model of infection, is associated with significantly reduced lung inflammatory response, clinical disease severity, and airway dysfunction. TLR7 also plays a role in type I IFN production in response to hMPV infection, as TLR7-deficient plasmacytoid dendritic cells (pDCs) show reduced IFN- β secretion compared to wild type cells. Although the expression of TLR9 has been shown to be increased in a hMPV murine model of bacterial superinfection, its role in hMPV-induced host responses has not been revealed. MyD88 is a common TLR adapter, and its absence in DCs and in a mouse model of infection is associated with reduced proinflammatory mediators and type I IFN production, which results in amelioration of hMPV-induced pulmonary inflammation and clinical disease.

The elucidation of the role of NLRs in hMPV infection has only recently been initiated, therefore, limited information is available. HMPV has been shown to activate NLRP3 inflammasome in an SH protein-dependent manner both in cells and mice, playing a detrimental role during infection, as inflammasome inhibition ameliorates clinical disease and lung inflammation without affecting viral replication. Furthermore, children experiencing a more severe disease form have significantly higher cellular and secreted levels of the inflammasome-regulated NLRP3 and IL-1 β , respectively.

Like many other viruses, hMPV has developed strategies to subvert cellular signaling and antiviral responses. The P protein of hMPV-B strains, but not the one from A strains, suppresses RIG-I-dependent type I IFN production. The hMPV G protein has also been shown to suppress RIG-I-mediated immune responses in AECs by disrupting the RIG-I-MAVS interaction and to inhibit TLR4-mediated innate antiviral signaling in DCs. The SH protein of hMPV downregulates type I IFN signaling pathway by reducing the expression and phosphorylation of signal transducer and activator of transcription 1 (STAT1). SH also blocks type I IFN induction in pDCs by inhibiting the TLR7/MyD88/TRAF6 pathway, and suppresses the nuclear factor (NF)- κ B transcriptional activity in AECs. The M2-2 protein is also an important immune regulatory protein by targeting the common adapter molecules MAVS and MyD88. A summary of the viral protein antagonism of cellular signaling pathways is presented in [Fig. 4](#).

Diagnosis

As hMPV infection shares clinical signs and symptoms with a variety of other respiratory viruses, it is not possible to make a diagnosis based solely on the patient's clinical presentation, thus, laboratory testing is necessary to confirm the infection. HMPV can be detected by culture, nucleic acid amplification, antigen detection and serology, however, not all methods have clinical utility. Traditional virology methods, such as cell culture, are not very helpful with hMPV in clinical practice because isolation may take over 2 weeks, and the mild cytopathic effect caused by hMPV is not consistent among different genotypes. Furthermore, growing hMPV in cell culture requires exogenous trypsin and is only successful in a few specific cell lines.

Real time PCR, targeting usually either the N or the F protein gene, is the method of choice in most medical facilities for hMPV diagnosis – especially with the current availability of multiplex assays, which allow simultaneous detection of several respiratory viruses, and therefore, establishes the presence of virus coinfections.

Viral antigens can be detected in cells of respiratory secretions by immunofluorescence assay, but this method, despite being of value during seasonal epidemics, is less sensitive than RT-PCR. Enzyme-linked immunoassays (ELISA) can be used for serologic testing, however, as the majority of individuals test seropositive for hMPV after childhood, serology only yields useful information in cases of seroconversion, or when a four-fold increase in subsequent specimens is observed.

Imaging studies, such as chest radiography, and computed tomography without contrast, are clinically useful for the overall evaluation of the patient and clinical decision-making process, but hMPV cannot be differentiated from other respiratory infections associated with bronchiolitis or pneumonia on the basis of imaging alone.

Treatment

The treatment options available to those infected with hMPV are limited to supportive care with intravenous hydration and oxygen supplementation, if necessary. Similar to RSV infection, bronchodilators and corticosteroids have been used in patients with airway obstruction, with no clear benefits, therefore, they are generally not recommended for pediatric or adult patients.

Ribavirin and nonspecific polyclonal intravenous immunoglobulins (IVIG) have been used with mixed results in some specific cases and case series. Though preclinical studies of ribavirin and IVIG for the treatment of hMPV did show promising results, only sporadic case reports have shown positive outcomes when they were used in adults and children mostly immunocompromised due to cancer or solid-organ transplant. Because of conflicting conclusions, and the lack of randomized controlled trials, there is no recommendation regarding the therapeutic use of ribavirin, nor IVIG.

Several experimental products are under research and development for the treatment of hMPV, including specific monoclonal antibodies, fusion inhibitors, and small interfering RNAs. Among therapeutic immunoglobulins, monoclonal antibodies developed against hMPV are currently in the preclinical research pipeline, having shown efficacy *in vitro* and also *in vivo*, by reducing pulmonary viral load and severe disease in small animal models of infection. Additional therapeutic approaches under experimentation are fusion inhibitors and RNA interference, targeting primarily N and P proteins, which have been tested *in vitro* with a certain degree of effect.

Prevention

Currently, the prevention of hMPV infection is mostly achieved through epidemiological countermeasures. Standard droplet isolation infection control practices are recommended in clinical settings where infected patients are being treated, as well as, in long-term care facilities when cases of hMPV infection are suspected.

While a vaccine to prevent hMPV is not currently available, researchers are exploring different strategies to develop a safe and effective vaccine, including the generation of live-attenuated vaccines, vector and chimeric-based vaccines, as well as, subunit vaccines. Live-attenuated virus candidates in preclinical development include, cold-passaged temperature-sensitive strains and recombinant vaccine candidates generated by deletion (M2-2, G and/or SH proteins) or mutation (M2-1 or F proteins) of selected genes, which have shown good immunogenicity, and reduced viral replication in the lungs of various animal models of infection, with no enhanced disease – different from inactivated vaccines, both formalin and heat-inactivated, which increased disease pathogenicity. The recombinant virus lacking SH expression has been recently proved to be immunogenic in healthy adults, and it is now in the research pipeline as the parent virus for vaccine development.

Chimeric, virus-like particle and subunit vaccines have also been developed and tested primarily in rodent models of infection. In general, those vaccines that including the F protein were shown to be immunogenic and protective, however, in the case of the F subunit vaccine, the antibody levels seemed to rapidly wane in nonhuman primates.

Experimental monoclonal antibody preparations have also been studied for their prophylactic use to prevent severe disease. In rodent models, the use of these antibodies has shown promising results. In animals treated prior to hMPV challenge, viral load was low or undetectable in lung specimens and microscopic examination revealed reduced tissue damage compared to untreated or mock-treated animals.

Further Reading

- Biacchesi, S., Pham, Q.N., Skiadopoulos, M.H., *et al.*, 2005. Infection of nonhuman primates with recombinant human metapneumovirus lacking the SH, G, or M2-2 protein categorizes each as a nonessential accessory protein and identifies vaccine candidates. *Journal of Virology* 79, 12608–12613.
- Boivin, G., Abed, Y., Pelletier, G., *et al.*, 2002. Virological features and clinical manifestations associated with human metapneumovirus: A new paramyxovirus responsible for acute respiratory-tract infections in all age groups. *The Journal of Infectious Diseases* 186 (9), 1330–1334.
- Chu, H.Y., Renaud, C., Ficken, E., *et al.*, 2014. Respiratory tract infections due to human metapneumovirus in immunocompromised children. *Journal of the Pediatric Infectious Diseases Society* 3 (4), 286–293.
- Falsey, A.R., Erdman, D., Anderson, L.J., Walsh, E.E., 2003. Human metapneumovirus infections in young and elderly adults. *The Journal of Infectious Diseases* 187, 785–790.
- Jartti, T., van den Hoogen, B., Garofalo, R.P., *et al.*, 2002. Metapneumovirus and acute wheezing in children. *Lancet* 360 (9343), 1393–1394.
- Kolli, D., Bao, X., Casola, A., 2012. Human metapneumovirus antagonisms of innate immune responses. *Viruses* 4, 3551–3571.
- van den Hoogen, B.G., de Jong, J.C., Groen, J., *et al.*, 2001. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nature Medicine* 7, 719–724.
- Wen, S.C., Williams, J.V., 2015. New approaches for immunization and therapy against human metapneumovirus. *Clinical and Vaccine Immunology* 22 (8), 858–866.
- Williams, J.V., Wang, C.K., Yang, C.F., *et al.*, 2006. The role of human metapneumovirus in upper respiratory tract infections in children: A 20-year experience. *The Journal of Infectious Diseases* 193, 387–395.

Human Norovirus and Sapovirus (*Caliciviridae*)

Sumit Sharma and Marie Hagbom, Linköping University, Linköping, Sweden

Lennart Svensson, Linköping University, Linköping, Sweden and Karolinska Institute, Stockholm, Sweden

Johan Nordgren, Linköping University, Linköping, Sweden

© 2021 Elsevier Ltd. All rights reserved.

This is an update of K.Y. Green, Noroviruses and Sapoviruses. In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-0-12374410-4.00555-0](https://doi.org/10.1016/B978-0-12374410-4.00555-0).

Introduction

Acute gastroenteritis (AGE) remains a major cause of morbidity and mortality in people in all age groups. Enteric caliciviruses, including human noroviruses and sapoviruses, are today recognized as important etiologic agents of this illness. However, before 1970, it was assumed that bacteria or parasites caused almost all acute infectious diarrhea. It was not until the introduction of electron microscopy (EM) and negative staining in the early 1970s that several viruses were discovered and found associated with acute diarrhea. In 1972, virus particles measuring 27 nm in diameter, were detected by immuno-electron microscopy in the feces of volunteers who had received filtered fecal suspensions from an outbreak of winter vomiting disease that had originated from a primary school and spread to the local community of Norwalk in the US. The particle was therefore first named 'Norwalk agent'; decades later, it was named norovirus and has a 'star of David' appearance under EM (**Fig. 1**). It has now been established as a causative agent for one-fifth of all acute gastroenteritis, with approximately 200,000 deaths annually worldwide.

In 1976, Madeley and Cosgrove and Flewett and Davies, reported morphologically 'typical human caliciviruses' in the feces of children with acute gastroenteritis. In the years that followed, these viruses were detected by several investigators and were named 'typical human caliciviruses' or 'Sapporo-like viruses' after an outbreak in 1982 in Sapporo, Japan. Today, these viruses are known as sapovirus.

Although human norovirus and sapovirus were first detected nearly five decades ago, several questions related to their pathogenesis and host–virus interactions still remain unanswered, partly due to the lack of a small-animal model and the inability to cultivate these viruses in vitro. However, most recently, human intestinal enteroids (HIEs) originating from stem cells in the small intestine have been established and are being used successfully to propagate human noroviruses. This recent major breakthrough has increased our knowledge of norovirus biology significantly. It is anticipated that the recent successful cultivation of sapovirus in HIEs will also lead to a similar advancement of understanding.

Norovirus

Viral Classification

Norovirus is a genus in the family *Caliciviridae*, which also includes the genera Bavovirus, Lagovirus, Minovirus, Nacovirus, Nebovirus, Recovirus, Salovirus, Sapovirus, Valovirus, and Vesivirus (**Table 1**). The norovirus genus contains a diverse group of viruses that infect several mammalian species including humans, dogs, cats, pigs, mice, and cattle.

The classification within the norovirus genus has changed significantly through the years and was recently updated.

Noroviruses are classified into 10 capsid genogroups (GI to GX) based on amino acid sequence diversity of the complete major capsid gene (VP1). Strains that infect humans belong to GI, GII, and GIV genogroups, with GI and GII causing the vast majority of human disease. Similarly, partial RNA dependent RNA polymerase (RdRp) nucleotide diversity is used to classify norovirus into eight P-groups (GI.P, GII.P, GIII.P, GIV.P, GV.P, GVI.P, GVII.P, and GX.P).

These genogroups are further divided into capsid (VP1) or RdRp genotypes. This dual genotyping system can account for that recombination is a major factor in norovirus diversity, with the major recombination site present in the junction between the RdRp and VP1 genes (open reading frame [ORF]1–ORF2 junction).

Today, VP1 genogroup GI has been divided into nine capsid genotypes (GI.1–GI.9) and GII into 27 capsid genotypes (GII.1–GII.27). The GI.P and GII.P RdRp genogroups are divided into 14 and 37 P-genotypes, respectively (GI.P1–GI.P14 and GII.P1–GII.P8, GII.P11–GII.P13, GII.P15–GII.P18, GII.P20–GII.P41). The division into VP1 capsid genotypes is based on the amino acid diversity of the complete VP1 gene, whereas the P-genotype is currently based on the nucleotide difference of 762 nucleotides at the 3' end of ORF1.

The predominant human genotype GII.4 is further classified into variants based on phylogenetic clustering after becoming epidemic in at least two geographical locations.

Genome

Human norovirus is a single-stranded positive-sense RNA virus [ssRNA(+)]. The genome is ~7.5 kb long and organized in three ORFs. The genome is covalently linked to a virus-encoded protein called VPg at the 5' end, and is polyadenylated at the 3' end

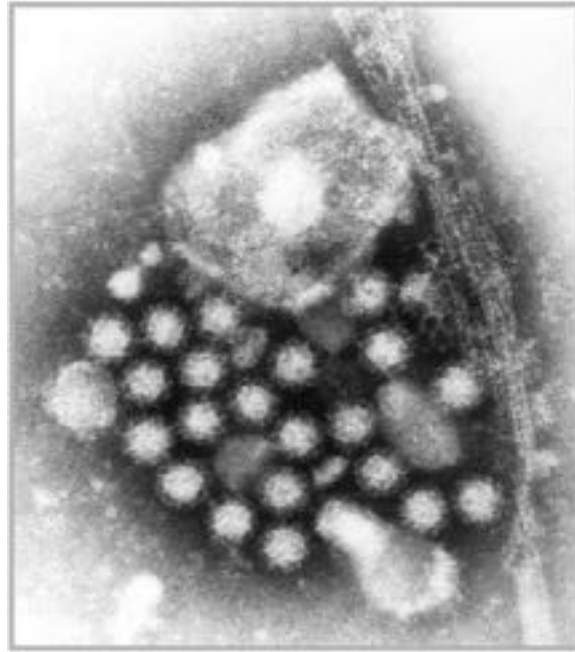


Fig. 1 Norovirus particles visualized by negative staining electron microscopy.

Table 1 Calicivirus classification detailing genus and type species

<i>Genus</i>	<i>Type Species</i>
Norovirus	Norwalk virus (NV)
Sapovirus	Sapporo virus (SV)
Lagovirus	Rabbit hemorrhagic disease virus (RHDV) European brown hare syndrome virus (EBHSV)
Vesivirus	Vesicular exanthema of swine virus (VESV) Feline calicivirus (FCV)
Nebovirus	Newbury – 1 virus (NBV)
Recovirus	Tulane virus
Valovirus	St Valerian calicivirus
Bavovirus	Chicken calicivirus
Nacovirus	Turkey calicivirus
Salovirus	Atlantic salmon calicivirus
Minovirus	Fathead minnow calicivirus

Adapted from Desselberger, U., 2019. Caliciviridae other than noroviruses. *Viruses* 11, 286. Jan Vinjé, J., Estes, M.K., Esteves, P., *et al.*, 2019. ICTV virus taxonomy profile: Caliciviridae. *Journal of General Virology* 100, 1469–1470.

(Fig. 2(A)). The untranslated regions (UTRs) at either end of norovirus contain conserved RNA secondary structures important for viral replication, translation and pathogenesis.

ORF1 encodes a large polypeptide that is cleaved by the virally encoded cysteine proteinase into six mature non-structural proteins (Fig. 2(A)).

ORF2 encodes the major structural protein, VP1, which in turn is organized into internal shell (S) domains and protruded domains (P). The S domain is the most conserved part of the VP1 gene. The P domain can be further divided into the P1 domain and the hypervariable P2 domain, which encodes epitopes that are targets for neutralizing antibodies as well as sites involved in binding to attachment factors/receptor.

ORF3 encodes a minor structural protein called VP2, which has a putative role in assisting capsid assembly and genome encapsidation.

Virion Structure

Early experiments to understand norovirus virion structure using electron cryo-microscopy and image processing techniques involving recombinant Norwalk-like virus (rNV) particles have shown that norovirus exhibits $T = 3$ icosahedral symmetry,

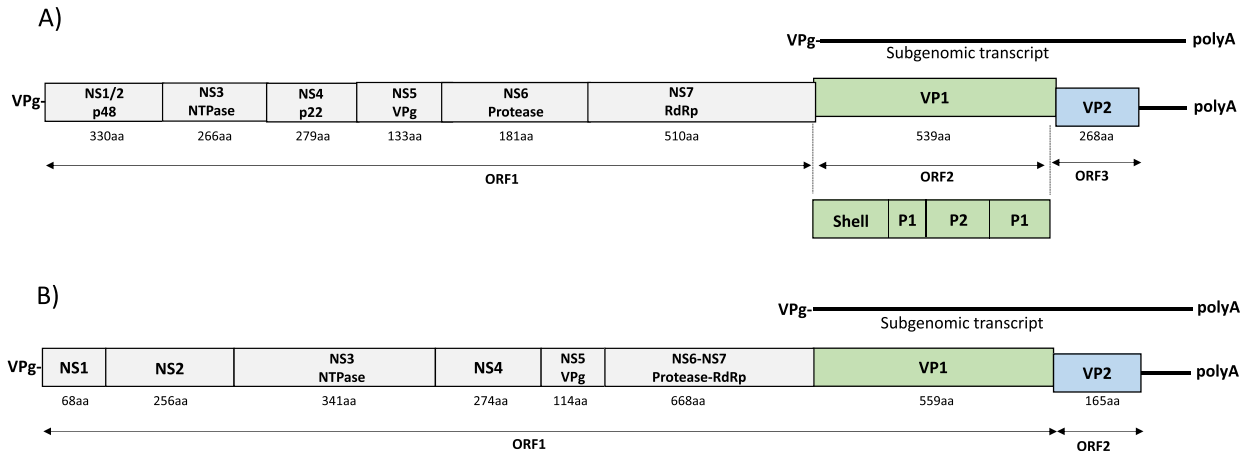


Fig. 2 Norovirus and sapovirus genome organization. (A) The norovirus genome consists of three open readings frames. ORF1 (in grey) encodes the nonstructural proteins, with ORF2 (in green) and ORF3 (in blue) encoding the structural proteins VP1, the major capsid protein, and VP2, a minor structural protein, respectively. The non-structural polyprotein of ORF1 is processed by the cysteine protease (Pro), into six mature proteins: p48, an N-terminal protein proposed to be involved in replication complex formation; NTPase, a nucleoside triphosphatase; p22, a protein of unknown function but proposed to be involved in replication complex formation; VPg, covalently attached to the 5' end of the viral genome and involved in translation and replication; Protease, the cysteine proteinase, and RdRp, which is the viral RNA-dependent RNA polymerase. The major capsid protein is further divided into the shell and protruding domains (P), with the P domain subsequently divided into two different subdomains, P1 and P2. The approximate size of proteins expressed as number of amino acid (aa) are from the GII.4 MD-145 strain (AY032605). (B) Sapovirus genome consists of two open reading frames, ORF1 (non-structural proteins and VP1, in grey and green, respectively) and ORF2 (VP2, in blue). A subgenomic RNA is also produced during replication. A viral protein (VPg) is linked to the 5' end and a poly-A tail to the 3' end of both the genomic and subgenomic RNA. The approximate size of proteins expressed as number of amino acid (aa) are from the GI.1 Manchester strain (X86560).

wherein 180 capsid protein molecules form 90 dimers. X-ray crystallography studies have revealed that the capsid protein structure has S domain and a P domain, which are linked by a flexible hinge. The NH₂ terminal part of the capsid protein (up to 225 aa) is involved in the formation of the icosahedral shell, while the P domain is formed by two subdomains, i.e., P1 (aa 226–278 and aa 406–520) and P2 (aa 279–405).

Life Cycle

Until recently, there were no robust cell culture systems supporting human norovirus propagation, which has hampered the study and characterization of the viral life cycle. Hence, most information on the norovirus life cycle is obtained from surrogate virus systems such as murine norovirus and feline calicivirus, or the use of gnotobiotic pigs that can be infected with human norovirus. Different human norovirus replicon systems have also been used. These models have provided some insight into the life cycle of human noroviruses, but all with specific limitations. Murine norovirus, for example, is a common model for molecular studies due to the availability of cell culture and reverse genetics systems, but differs from human norovirus in terms of host interaction and pathogenesis. The recent development of human intestinal enteroids (HIEs) able to propagate human norovirus *in vivo* have already provided new insights into the human norovirus life cycle, and will be an essential tool for future studies.

Enterocytes in the small intestine are likely the primary sites of norovirus infection. Biopsies from immunocompromised patients with chronic norovirus infections have shown the presence of viral antigens in small-intestine enterocytes. Small-intestine enterocytes were also norovirus antigen-positive in a gnotobiotic pig model of human norovirus infection, and human noroviruses can be propagated in enterocytes present in *in vitro* culture systems of HIEs. Moreover, B-cells can be infected with human norovirus *in vitro*, although evidence of B cell infection *in vivo* is lacking. Patients with severe combined immunodeficiency (SCID) who lack B cells can be infected with norovirus, albeit with a slightly lower viral load. Thus, B cells are not essential for infection and replication but may be important for norovirus pathogenesis, although further studies are needed to determine their exact role.

Much is still unknown of the different steps of the live cycle of norovirus infecting a cell. However, an essential feature is that attachment factors for most human norovirus are cell-associated host glycans such as HBGAs (histo blood group antigens) present in epithelial surfaces of the small intestine, which can facilitate entry into the cell. For example, non-secretors, lacking the expression of α 1,2-linked fucosylated HBGAs in the small intestine are not infected with the predominant GII.4 genotype. Furthermore, GII.4 strains of human norovirus only infect HIEs derived from secretors but not those from non-secretor individuals. The soluble form of HBGAs and bile salts also increase virus infectivity. The subsequent receptor engagement remains unknown for norovirus. Several models have been proposed, such as binding to glycosphingolipids on the cell surface, followed by the invagination and release of norovirus-containing vesicles by dynamin II-mediated scission of the invagination, and viral genome release into the cytosol. The VPg-linked RNA genome of norovirus will subsequently act as an mRNA, where the VPg

protein is important for mediating translation and replication. Translation of the ORF1 polyprotein is followed by co-translational and post-translational processing by a viral protease, resulting in replication complex formation. Translation of the viral proteins VP1 and VP2 occurs mainly from the subgenomic RNA (sgRNA), likely to produce higher levels of the major capsid protein for virus assembly, as the sgRNA is present at higher levels and each icosahedral capsid contains 180 copies of VP1.

The processes behind norovirus assembly, encapsidation, and exit are largely unknown. The ability of VP1 to self-assemble into virus-like particles (VLPs) that are morphologically and antigenically highly similar to native virions suggests that capsid assembly during virus replication may occur without the involvement of cellular proteins.

Non-enveloped viruses, including norovirus, were usually thought to exit infected cells through cellular lysis. Recent findings, however, have shown that cell destruction is not always necessary for norovirus exit. Stools from infected people show norovirus residing in vesicles of exosomal or plasma membrane origin, which indicates that at least some norovirus are shed non-lytically inside exosomes.

Epidemiology

Norovirus causes sporadic and epidemic gastroenteritis in both children and adults and has been described from a variety of geographic regions and settings. The estimated global prevalence of norovirus is approximately 20% of all cases of AGE. Norovirus is also common in asymptomatic infections, with a globally estimated prevalence of 7%. All ages are affected, but the incidence is the highest in children aged <5 years. Norovirus is estimated to cause approximately 200,000 deaths annually worldwide, with 70,000 or more among children in developing countries. In children, norovirus is typically second only to rotavirus in causing acute gastroenteritis, and in several countries with national rotavirus immunization, it has become the most common etiology in children with AGE.

Norovirus is mainly transmitted through the fecal-oral route. Besides sporadic cases, norovirus is also the most common agent in outbreaks of gastroenteritis due to its high stability in the environment and ease of spread through contaminated water and food (Table 2). Waterborne norovirus outbreaks have been linked to potable water, wells, recreational water use such as swimming, and fountains. Norovirus is further estimated to be responsible for almost 50% of foodborne outbreaks globally. Fruits, vegetables, berries, shellfish, cereals, spices, and oysters are common in foodborne norovirus outbreaks, as they are consumed raw and subject to contamination from water sources. Contamination of food can also occur during its preparation.

Norovirus outbreaks occur frequently in settings with close human contact, such as schools, nursing homes, retirement communities, cruises, hotels, hospitals, or day care centers. The spread during an outbreak is predominantly via direct contact, exposure to aerosols or fomites. Many factors contribute to norovirus transmission and its ability to cause outbreaks. Norovirus is an extremely stable particle in the environment, and is highly resistant to disinfectants such as alcohols and to freezing and high temperatures. Furthermore, a low dose, i.e., 10–100 viral particles, is required for symptomatic infection, and norovirus can be shed in the stool at up to $1-10^{10}$ genome equivalents per g feces, and up to 10^6 genome equivalents per ml of vomit. Prolonged asymptomatic shedding further increase the risk of secondary spread (Table 2).

Seasonality

As the term ‘winter vomiting disease’ implies, norovirus infections exhibit a clear winter seasonality in many countries. This winter seasonality, however, is predominant in temperate regions, with peaks in December–February in the Northern Hemisphere or June–August in the Southern Hemisphere. In many tropical regions, such as Sub-Saharan Africa or South America, less data is available. Studies from several tropical countries have reported no or less clear patterns of seasonality. In some tropical countries

Table 2 Characteristics of norovirus facilitating transmission in different outbreak settings and in person-to-person transmission

Characteristics	Observations	Consequence
Stability	High survival in the environment. Resilient to freezing, heating and disinfectants such as alcohols	Difficult to eliminate in water and surfaces, leading to infections from e.g., oysters, recreational water, food irrigated with sewage, and fomites. Increased risk of infections in closed settings such as hospitals
Long-term/Asymptomatic shedding	Patients can shed norovirus up to three weeks after resolution of symptoms. Immunocompromised patients shed the virus even longer	Increased risk of secondary spread, can be particularly problematic concerning food handlers.
Viral diversity	Multiple genetic and antigenic strains exist, using different attachment factors in humans	Re-infections can occur more easily with different strains. Diagnostic and detection methods may not be sensitive for all strains
Low infectious dose High viral load in feces and vomitus	Only ~10–100 virus particles are needed for symptomatic infection	Increases risk of person-to-person transmission, aerosols, secondary spread, food, water and surface contamination
Short-term immunity	Symptomatic re-infection with homotypic strains can occur	Symptomatic re-infections occur. Adults are not protected although infected as children. May hinder development of effective vaccines with long-term protection

Note: Modified from Nordgren, J., 2009. Norovirus epidemiology: Prevalence, transmission, and determinants of disease susceptibility. Linköping: Linköping University Electronic Press. Patel, M.M., Hall, A.J., Vinjé, J., Parashar, U.D., 2009. Noroviruses: A comprehensive review. Journal of Clinical Virology 44, 1–8.

such as Nicaragua, norovirus peaks during the rainy season, and its prevalence has been associated with average rainfall in other countries, whereas some studies have found an association with rainfall only for norovirus outbreaks, and not for sporadic cases. The impact of rainfall on norovirus infections is not understood, but may be linked with environmental spread of the virus.

Molecular epidemiology

Genotype GII.4 is the predominant genotype worldwide for both sporadic infections and outbreaks of norovirus. A global review of norovirus in children (aged <18 years) found GII.4 accounting for 70% of the capsid genotypes and 60% of the polymerase genotypes, followed by the capsid genotype GII.3 (16%). A further systematic review of norovirus in low-income countries found that GII.4 was the predominant genotype in all settings (47%), followed by GII.3 and GII.6. The most prevalent GI strain was GI.3. In some geographic settings, other capsid genotypes such as GII.17 and GII.2, can become temporarily predominant. Nevertheless, >50% of norovirus infections yearly are caused by norovirus GII.4, with pandemics occurring every ~2–7 years. The first known human norovirus pandemic occurred in 1995–1997, with subsequent pandemics in 2002, 2004, 2006, 2009, and 2012. The reason GII.4 is the pandemic-causing genotype is not fully understood but could be related to higher biological fitness, seen as higher levels of viral shedding in infected patients, more diverse attachment receptor specificity and fast mutation frequency leading to antigenic drift and putative change of HBGA attachment factors, both leading to escape of previous herd immunity. The predominant genotype GII.4 has been linked to more severe symptoms and is less common in asymptomatic infections. Some studies have indicated that genotype GII.4 is relatively less common in food and waterborne outbreaks compared to other non-GII.4 or non-GI strains, which might be associated with higher resistance of some genotypes to environmental factors.

Population genetics and norovirus epidemiology

HBGAs, specifically the secretor, Lewis and ABO families, are strongly associated with susceptibility to norovirus in a genotype-dependent manner. Secretor status, mediating the presence of α 1,2 fucose-linked blood group antigens in epithelial surfaces of the small intestine, has been strongly linked to susceptibility to a wide array of norovirus genotypes. As expression of these HBGA glycans are linked to ethnicity and differ widely between populations, this likely has an effect on norovirus epidemiology. Many common norovirus genotypes such as the predominant GII.4, the recently emerging GII.17 and the common GII.6 genotype have all shown strong secretor specificity. Other genotypes, whose clinical importance in many cases is unknown, infect both secretors and non-secretors. As no genotype infects non-secretors exclusively, it is likely that populations with a high prevalence of secretors are more infected with GII.4 as well as other secretor specific or non-specific genotypes, which might lead to differences in disease burden, particularly as genotype GII.4 is associated with more severe symptoms.

Clinical Features

Norovirus cause vomiting and diarrhea, both of which may happen together or independently. The incubation period is short, usually 1–2 days, with symptoms lasting 2–3 days. Besides vomiting and diarrhea, abdominal pain, fever, headache, chills, myalgia, and malaise can also occur. Other unusual symptoms include encephalopathy and convulsions. Convulsions among norovirus-infected children are reported more often than for rotavirus infections. Norovirus disease is more severe among the elderly and people with accompanying medical conditions. Besides old age, very young children are also vulnerable. A study from Japan reported that the duration of illness among children aged <2 years was twice that in children aged 2–4 years. Additionally, hematopoietic stem cell transplant (HSCT) recipients and other immunocompromised individuals can have persistent norovirus infection with a duration of months or even years.

Pathogenesis

Until very recently, the lack of a cell culture system and an efficient animal model restricted our knowledge of norovirus pathogenesis. Our limited understanding has been obtained from histological or biochemical studies among human volunteers and data from porcine, bovine, and murine models. Histological examination of the small intestine of norovirus-infected human volunteers showed villi broadening and blunting, microvilli shortening, increased cytoplasmic vacuolization, and intercellular oedema. Like rotavirus infection, norovirus infection results in delayed gastric emptying.

Besides epithelial cells, human norovirus capsid protein has also been detected in macrophages, T cells and dendritic cells (DCs) of norovirus-infected immunocompromised patients. Recently, in vitro studies have reported human norovirus propagation in differentiated human intestinal organoid/enteroid cultures, and in a few established human B cell lines.

Mild inflammation of the lamina propria has also been reported in volunteer studies and animals. A large amount of understanding on norovirus pathogenesis comes from a murine infection model. Murine strains that cause both acute (MNV-1 and CW3) and persistent (CR6 and MNV-3) infection have been used to better understand norovirus pathogenesis. During the initial phase of MNV-1 infection in mice, the macrophages, DCs, B cells and T cells in the gut-associated lymphoid tissues become infected. It is suggested that murine noroviruses that cause acute infection involve M cells to overcome the epithelial barrier and cause initial infection. On the other hand, persistent murine norovirus strain CR6, which can be continuously shed in feces has been reported to infect tuft cells. Besides immune cells in the gut, MNV-1 infection also leads to extra-intestinal spread in immunocompetent mice with MNV-1 virus reportedly detected in spleen, liver, lungs, and lymph nodes. Furthermore, STAT1 (signal transducer and activator of transcription 1) not only limits MNV-1 virus replication in the intestines but also limits its spread to the peripheral tissues; however, MNV-1 infection in STAT1-deficient mice resulted in weight loss, gastric bloating, and

diarrhea. Differences in disease severity between MNV-1 and MNV-3 norovirus strains have been reported in wild-type mice, wherein MNV-1 results in modest to severe intestinal pathology, whereas MNV-3 infection is attenuated. Another interesting aspect was that no correlation between viral titres and disease severity was observed. Wild-type mice infected with MNV-1 showed intestinal pathology, while high viral load with no intestinal pathology was observed in MNV-3-infected wild-type mice. Although there are some similarities between human and murine noroviruses, such as involvement of carbohydrate attachment factors, and potential for continued viral shedding, there are certain differences as well, most notably is the inability of murine norovirus to produce severe diarrhea in mice.

Recent studies have been addressing the role of the host microbiota in norovirus infections. It has been reported that human norovirus interacts with both commensal and pathogenic bacteria via norovirus capsid and HGBA-like carbohydrates expressed on bacterial surfaces. Antibiotic-administered mice prevented persistent murine norovirus infection. However, it should be noted that infection of HIEs with human norovirus clinical isolates does not require bacteria, warranting further research on the importance of commensal bacteria in human norovirus disease.

It has also been suggested that norovirus might serve as a trigger for intestinal bowel syndrome or intestinal bowel disease. Norovirus infection increases intraepithelial cytotoxic T cells in the duodenum for up to 6 days after the onset of symptoms. Additionally, apoptosis of norovirus-infected epithelial cells has also been suggested.

Immunity

Protection from norovirus can be divided into genetic and immunological components. Briefly, the genetic factors for norovirus susceptibility are associated with expression of HBGAs in the small intestine. Secretor status is defined by the presence of a functional fucosyltransferase 2 (FUT2) enzyme which determines the presence of α 1,2 fucose linked HBGAs (A, B, H-type 1, and Lewis b). Lack of a functional FUT2 enzyme (yielding the so-called secretor-negative phenotype) is a strong restriction factor for several norovirus genotypes, including the predominant genotype GII.4. The overall prevalence of the non-secretor phenotype is approximately 20%, but large differences exist between populations.

Immunologically, humoral responses have an important role in the protection against norovirus infection. Higher serum levels of antibodies blocking norovirus VLP binding to HBGAs correlate with a lower risk of developing illness to some norovirus genotypes. The potential mechanism of action is blocking of the HBGA binding site of the virus capsid hindering the attachment to HBGAs on the cell surfaces of the small intestine. Higher levels of norovirus-specific memory B cells also correlate with lower risks of developing illness. This in combination with that immunodeficient persons lacking B-cells can develop chronic norovirus infections, suggest B cells to be important for preventing norovirus infection. Cross-reactive antibodies can be developed after infection, and immune response following infection of one genotype can provide protection against other genotypes. However, norovirus infections do not elicit long-lasting immunity, and a person can develop symptoms upon reinfection even with the same norovirus strain. The estimates of the degree and duration of homotypic and/or heterotypic protection after infection varies widely.

Diagnosis

Since the initial detection of norovirus using electron microscopy in the early 1970s, various diagnostic approaches have been used to detect noroviruses in clinical and environmental samples. Currently, both antigen and molecular (for detecting viral genome) tests are widely used.

(1) Antigen tests

Enzyme immunoassays

Enzyme immunoassays have been developed for norovirus GI and GII antigens in feces, but these assays have varying sensitivity (32%–92%) and specificity (65%–100%). Although most kits include cross-reactive monoclonal and polyclonal antibodies, variability in norovirus load in clinical specimens, the circulation of antigenically distinct genotypes and variability within the genotype due to antigenic drift affect their performance. However, their biggest advantage is the high throughput and feasibility for use in laboratories with minimum infrastructure.

Besides conventional enzyme immunoassays, Immunochromatographic (ICG) lateral flow assays can be used for rapid detection as they do not require specialized laboratory equipment or highly trained personnel. However, their low sensitivities, warrant further screening of the negative samples.

(2) Molecular tests

Reverse transcription-PCR (RT-PCR)-based assays are today considered the gold standard for detecting norovirus and have been continuously updated over time to ensure high specificity and sensitivity.

Conventional RT-PCR

One of the first RT-PCR assays used primers to the first detected norovirus strain, the Norwalk virus. Subsequent generations of RT-PCR assays used primers from a more conserved region (the viral polymerase gene). Although having higher performance than that of enzyme immunoassays; hybridization or sequencing was required to ensure high specificity and sensitivity. Another

shortcoming of the conventional RT-PCR over real-time RT-PCR is the processing required for visualizing the amplicon after genome amplification.

Real-time RT-PCR and multiplex molecular assays

Real-time RT-PCR assay today is the most common method for norovirus detection. It is usually applied in a multiplex fashion to distinguish GI and GII genogroup noroviruses. Most real-time PCR assays use primers and probes targeting the most conserved part of the norovirus genome, the ORF1-ORF2 junction.

Recently, the use of large multiplex molecular assays (Biofire FilmArray Gastrointestinal Panel, Luminex xTag Gastrointestinal Pathogen Panel, and Taqman Array Card) for simultaneous detection of known gastrointestinal pathogens, including viruses, has increased. This reduces the turnaround time for accurate results and aids in identifying infections and/or co-infections that remained undiagnosed by routine test methods for single pathogens.

Single-unit tests

Besides multiplex assays, there also exist single-unit, easy-to-use norovirus automated molecular detection tests that require very little time, and staff training, and one such assay is the Xpert Norovirus assay (Cepheid, Sunnyvale, CA, USA).

Treatment

In immunocompetent individuals, norovirus gastroenteritis is self-limiting, and the treatment of norovirus gastroenteritis is generally supportive. Administering oral rehydration solutions (ORS), or intravenous rehydration, is the primary treatment strategy for correcting the altered electrolyte and fluid imbalance. Antiemetics, such as ondansetron can also be given to reduce vomiting and facilitate oral rehydration therapies. Loperamide, an anti-motility drug, is often used by travellers; however, its use in children is not recommended in all countries.

Therapeutics for persistent norovirus infection is of clinical importance and studies involving antiviral therapy to norovirus are limited. Nitazoxanide, a broad-spectra antiparasitic and antiviral agent, shortened the duration of norovirus illness, and resolved diarrhea in a norovirus-infected patient with refractory acute myelogenous leukemia and HSCT. Similarly, enteric administration of human immunoglobulin resolved chronic diarrhea in a transplant patient. Besides oral immunoglobulins and nitazoxanide, the anti-viral drug ribavirin has been linked to viral clearance in common variable immunodeficiency (CVID) patients.

Prevention

Norovirus spreads from person-to-person, contaminated food, water and other environmental contaminants. Thus, epidemic outbreaks as well as food and waterborne outbreaks are common, with norovirus being responsible for ~50% of foodborne AGE outbreaks. Various fruits and ready-to-eat vegetables that come into contact with contaminated water are often the source of norovirus-related food borne outbreaks. To date, there is no consensus regarding what viral load should be considered as a cut-off for contamination. However, a recent study using HIE model reported that a ct (threshold cycle) cutoff of 30 in a fecal sample could be considered as infectious. One of the ways of preventing norovirus infection is to ensure that water used for irrigation, washing vegetables or culturing shell fish is not contaminated with norovirus. Another major issue to be considered is that the water used for irrigation be pre-treated and not contaminated with fecal wastewater.

Besides food itself, the norovirus can also be transmitted via restaurant personnel infected with norovirus, who are often the index source of foodborne norovirus outbreaks in restaurants or cruises. Hence, appropriate disinfection measures are needed to prevent this contagious disease. Good hand hygiene remains an important measure for controlling norovirus spread. Cleaning with soap and running water is better for removing norovirus than alcohol-based hand disinfection. Until very recently, the lack of a cell culture system for propagating human norovirus hindered assessment of norovirus inactivation systems. A recent study using HIEs as a model to examine the effect of chlorine and alcohols (ethanol and isopropanol) on inactivating human norovirus infectivity. The authors reported that, irrespective of alcohol treatment duration/concentration, complete norovirus inactivation could not be achieved. On the other hand, chlorine treatment resulted in a complete norovirus inactivation.

In norovirus outbreaks in hotels, restaurants or cruise ships, besides isolating infected individuals and maintaining personal hygiene, the disinfection of contaminated surfaces is equally important. These include tables, chairs, commodes, doorknobs, bathroom, floors, and carpets. Hypochlorite at a minimum concentration of 1000 ppm and quaternary ammonium compounds are preferred for decontaminating surfaces and objects although the latter are less effective. In health care settings, limiting norovirus spread at the earliest instance is of utmost importance. Besides following the above routine disinfection procedures, additional measures such as discarding all unused disposable items present in the patient's room should be considered.

Vaccination

There is no licensed norovirus vaccine available till date; however, there are vaccine candidates at various stages of clinical development. Most of these are VLP-based vaccines administered through an oral, intranasal, or intramuscular route. Recently, a safety and immunogenicity study of a bivalent (GI.1/GII.4) VLP-based vaccine (administered intramuscularly) showed that it was well tolerated and elicited pan-immunoglobulin and HBGA-blocking antibodies. Furthermore, a single-site, randomized, double-blind, placebo-controlled clinical trial of an oral non-replicating adenovirus-based vaccine expressing VP1 gene of the GI.1

norovirus strain with a double-stranded RNA (dsRNA) adjuvant showed that it was well tolerated and generated both systemic and mucosal immune responses.

Sapovirus

Classification

The family *Caliciviridae* consists of 11 genera including sapovirus (Table 1). Human sapoviruses are highly heterogeneous, and genetic classification is usually based on phylogenetic analyses of the sequence of the major capsid gene VP1, the most diverse region of the genome. Today, sapoviruses are divided into 15 genogroups, of which viruses within genogroups GI, GII, GIV and GV infect humans. Viruses in the other genogroups have been detected in swine (GIII and GV–GXI), sea lions (GV), mink (GXII), dogs (GXIII), bats (GXIV, GXVI–GXIX), and rats (GXV). These genogroups are further divided into genotypes, but a standardized genotyping method is still lacking, although this would be important in order to facilitate worldwide comparison of the data. The International Calicivirus Scientific Committee in 2010 proposed that new sapovirus genotypes should only be assigned based on the complete capsid sequence.

Virion Structure

Human sapoviruses are about 30–38 nm in diameter and have an icosahedral structure, with characteristic cup-shaped depressions on the surface, and are morphologically distinguishable from other gastroenteritis viruses (e.g., rotavirus, astrovirus or adenovirus) by the ‘star of David’ appearance. The surface structure is formed from 180 copies of the VP1, arranged in a $T = 3$ icosahedral symmetry. Each VP1 subunits contains two principal domains: the S and P domains, which are connected via a flexible hinge region.

The major component of the complete virus, VP1, is a ~60-kDa protein, while the VP2 protein has not been identified in the virion but is predicted to be a strong basic protein in the interior of norovirus particles. VP1 expression results in the spontaneous assembly of VLPs.

Genome

The human sapovirus genome is a 7.4–7.5 kb ssRNA(+). The genome contains two ORFs (Fig. 2(B)). A sgRNA consisting of VP1 and VP2 genes is also produced during replication. The viral protein VPg is covalently attached to the 5′ end of genomic RNA and sgRNA, and the 3′ end is polyadenylated. ORF1 encodes a polyprotein that is cleaved post-translationally into non-structural proteins (NS1, NS2, NS3, NS4, NS5, and NS6–NS7) and a major capsid protein VP1, whereas ORF2 encodes the minor capsid protein VP2. A third ORF (ORF3) has been predicted in several human and bat sapovirus strains; however, its function is unknown.

Life Cycle

There is yet no cell culture model for the human sapovirus, which has limited our knowledge of its replication cycle. Most of replication cycle studies of the *Caliciviridae* family arise from feline and murine norovirus studies and from porcine sapovirus in porcine cell lines. It is thought that sapovirus has a similar replication cycle to norovirus, with receptor binding, endocytosis, uncoating and release of genomic RNA into the cytoplasm. The endocytosis of porcine sapovirus has been shown to be clathrin- and cholesterol-mediated. Bile acids likely support porcine sapovirus replication via escape from the late endosome into the cytoplasm to initiate virus replication, as in the absence of bile acid, the virus remains in the late endosomes/lysosomes and is destined to be degraded. Recently, it was shown that porcine sapovirus-infection induced an early activation of the PI3K-AKT (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha-AKT) and MEK-ERK (extracellular signal-regulated kinase) pathways that facilitated early to late endosome trafficking as well as the acidification and release of virus from the late endosome. The dsRNA genome is synthesized from the genomic ssRNA(+), and transcribed/replicated, thereby providing viral mRNAs/new ssRNA(+) genomes. Calicivirus genomes do not have a 5′ cap structure; instead, a small virus-encoded protein (VPg) is covalently linked to the 5′ end of the genome. Similar, the porcine sapovirus genome is covalently linked to VPg, and this linkage is required for the translation and infectivity of viral RNA. The VPg protein is bound directly to the 4F subunit of the eukaryotic translation initiation factor (eIF4F) complex in infected cells, which was required for initiating viral translation and infectivity. Lytic or non-lytic mechanisms encapsidate the genomic RNA during virus assembly and release. Moreover, the cyclooxygenase/prostaglandin E2 (COX-PGE2) pathway is induced during porcine sapovirus replication, a mechanism proposed to be used by the virus for efficient viral growth and which could provide a potential future target for the control of sapovirus replication.

Epidemiology

Sapovirus has been detected all over the world, both in outbreaks and sporadic cases and is a public health concern. The virus infects and causes disease in people of all ages but is most common in young children and the elderly. Similar to norovirus, sapovirus shows winter predominance in many countries. A recent 1-year study of out-hospitalized patients in Spain found that

sapovirus was the second major gastroenteritis pathogen, with a prevalence of 15%, and was most prevalent in infants, young children and the elderly. Similarly, in Nicaragua, having introduced rotavirus vaccination, sapovirus was the second most commonly detected enteropathogen among children aged <5 years and was observed in 15% of hospitalizations due to gastroenteritis.

Clinical Features

Sapovirus infection manifests with typical viral gastroenteritis symptoms, characterized by vomiting, watery diarrhea and abdominal pain. Other symptoms, including fatigue, nausea, headache, myalgia and chills can also be present. The clinical picture is very similar to norovirus, but the symptoms tend to be milder than those of norovirus and rotavirus infections. Laboratory diagnosis is essential for virus detection. Similar to norovirus infection, there is usually no or low fever. The incubation period is 1–4 days, but asymptomatic cases have been reported. Although the person does not show symptoms, they are still shedding high amounts of infectious virus and are capable of spreading the virus through the fecal-oral route. In immunocompetent individuals, diarrhea is usually resolved within one week.

Pathogenesis

Knowledge of the pathogenesis and cell tropism of human sapovirus is limited due to a lack of a small-animal model. Pathological lesions have only been studied in pigs infected with porcine sapovirus, and visualization by histopathological staining showed that the virus infects cells of the small intestine. Tissue sections show histological changes with villus atrophy in the duodenum and/or jejunum of inoculated pigs at day 3–7 post infection. These results were confirmed by scanning EM which also revealed villi shortening, blunting, fusion, or absence of villi.

Cell Culture

None of the human sapoviruses had previously been propagated in cell culture. However, porcine enteric calicivirus (PEC) grows in primary porcine kidney cells and in a porcine kidney cell line (LLC-PK), in the presence of porcine intestinal contents or bile acids to permit virus replication. It was thought that the bile acid or intestinal contents support growth in vitro and in vivo by downregulating innate immunity, through cyclic AMP/PKA signaling, which downregulates STAT1 and antiviral interferon (IFN) signaling. However, sapovirus growth in cell lines deficient in their ability to induce or respond to IFNs showed a 100–150-fold increase in infectious virus production, indicating that the primary role of bile acids is not inactivation of the innate immune response. Although the addition of bile acids is an essential co-factor for porcine sapovirus replication in cell culture, the mechanism has yet to be fully elucidated.

Recently, the stem-cell derived HIE culture system was used for successful replication of human norovirus. After this historical success, the non-transformed cell culture system has been used for other human enteric viruses such as enterovirus, rotavirus, adenovirus type 41 and astrovirus. This human intestinal non-transformed cell culture method is being currently explored for cultivation of human sapovirus.

Immunity

Seroprevalence studies of human sapoviruses have demonstrated gradually increasing seroprevalence during the early years, reaching high levels in school-aged children that remain high in adults. Reinfection with the same genogroup is common, but is rare with the same genotype, suggesting that infection generates a genotype-specific immunity from each infection. The fact that sapovirus gastroenteritis occurs more frequently in young children than in older children and adults suggests that natural infection likely provides long term protection against sapovirus and provides hope for success of a possible future sapovirus vaccine.

In contrast to norovirus infection, sapovirus infection has not been associated with host HBGA phenotypes. Neither the fucosyltransferase 2 (Fut2)^{-/-} non-secretor mutation (428A) nor any of the ABO blood groups or Lewis phenotypes protected children against sapovirus-infection.

Diagnosis

Visually, sapoviruses are distinguishable from other gastroenteric viruses, except norovirus, by their ‘Star of David’ morphology in EM, but EM has low sensitivity. Enzyme-linked immunosorbent assay (ELISA) can be used for antigen detection of human sapovirus, but the method is less sensitive than the current commonly and routinely used RT-PCR methods. Due to the high genetic diversity of sapoviruses, primers targeting well-conserved regions, may also amplify other gastroenteritis viruses. Multiple primers are often used for detecting sapoviruses of different genogroups. Targeting the RdRp-VP1 junction region, the overlapping part of the RdRp and VP1 genes, has the highest detection rate and is often the first choice for screening of clinical samples. For genotyping, VP1-targeting primers are preferred, as they allow reliable genotyping of the sequenced PCR products, since the RdRp-VP1 junction region is too short and conserved for sequence analysis.

Prevention

Sapoviruses can be transmitted from person to person via contacts with sapovirus-positive feces, vomitus, or sapovirus-contaminated materials/surfaces, or via contaminated food and drinking water. Sapovirus, like norovirus, is probably transmitted efficiently by aerosol from vomiting. To date, no vaccines or antiviral drugs are available for preventing or treating human sapovirus infections.

Infectivity and stability testing of the porcine sapovirus has shown that the virus is stable at pH 3.0–8.0 at room temperature (RT) for 1 h, sensitive to ethanol treatment (60% and 70%) at RT for 30 s, inactivated by 200 mg/L (or ppm) sodium hypochlorite at RT for 30 min and inactivated by heating at 56°C for 2 h. The stability of human sapovirus needs to be further investigated.

Further Reading

- Ahmed, S.M., Lopman, B.A., Levy, K., 2013. A systematic review and meta-analysis of the global seasonality of norovirus. *PLoS One* 8, e75922.
- Atmar, R.L., Ramani, S., Estes, M.K., 2018. Human noroviruses: Recent advances in a 50-year history. *Current Opinion in Infectious Diseases* 31, 422–432.
- Bartnicki, E., Cunha, J.B., Kolawole, A.O., *et al.*, 2017. Recent advances in understanding noroviruses. *F1000Research* 6, 79.
- Chhabra, P., de Graaf, M., Parra, G.I., *et al.*, 2019. Updated classification of norovirus genogroups and genotypes. *Journal of General Virology* 100, 1393–1406.
- Costantini, V., Morantz, E.K., Browne, H., *et al.*, 2018. Human norovirus replication in human intestinal enteroids as model to evaluate virus inactivation. *Emerging Infectious Diseases* 24, 1453–1464.
- Desselberger, U., 2019. *Caliciviridae* other than noroviruses. *Viruses* 11.
- Estes, M.K., Ettayebi, K., Tenge, V.R., *et al.*, 2019. Human norovirus cultivation in nontransformed stem cell-derived human intestinal enteroid cultures: Success and challenges. *Viruses* 11.
- Ettayebi, K., Crawford, S.E., Murakami, K., *et al.*, 2016. Replication of human noroviruses in stem cell-derived human enteroids. *Science* 353, 1387–1393.
- Nordgren, J., Svensson, L., 2019. Genetic susceptibility to human norovirus infection: An update. *Viruses* 11.
- Oka, T., Wang, Q., Katayama, K., *et al.*, 2015. Comprehensive review of human sapoviruses. *Clinical Microbiology Reviews* 28, 32–53.
- Robilotti, E., Deresinski, S., Pinsky, B.A., 2015. Norovirus. *Clinical Microbiology Reviews* 28, 134–164.

Relevant Website

<https://www.cdc.gov/norovirus/index.html>

Norovirus.

Home.

CDC.

Human Papillomaviruses (*Papillomaviridae*)

Alison A McBride, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States
Samuel S Porter, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States and University of Maryland, College Park, MD, United States

Published by Elsevier Ltd.

This is an update of P.F. Lambert, A. Collins, Papillomaviruses: Molecular Biology of Human Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00458-1. and G. Orth, Papillomaviruses: General Features of Human Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00457-X.

Taxonomy, Classification, and Evolutionary Relationships

Human papillomaviruses (HPVs) belong to the family *Papillomaviridae* that was established by the International Committee on the Taxonomy of Viruses in 2005 (previously, they had been classified with *Polyomaviridae* under the now defunct *Papovaviridae*). *Papillomaviridae* has two sub-families, the *Firstpapillomavirinae* and *Secondpapillomavirinae* (the latter contains only one genus). *Firstpapillomavirinae* are further classified into genera, based on nucleotide sequence homology of the L1 gene. Viruses with 61% or greater L1 nucleotide identity are considered to be within the same genus. *Firstpapillomavirinae* has 52 genera that are named by a Greek letter prefix.

Five genera from the *Firstpapillomavirinae* contain the human papillomaviruses: *Alpha*-, *Beta*-, *Gamma*-, *Mu*-, and *Nupapillomavirus*. Within each genus, viral genomes are further separated into species that share between 71% and 89% nucleotide identity within the L1 gene. For example, within the *Alphapapillomavirus* genus, the species are named α -1, α -2, α -3, etc. At the next level, viral genomes are divided into types and the L1 sequence of each type must differ > 10% from any other type. Within each type, there are usually distinct variant lineages. For example, the oncogenic virus HPV16 can be further separated into genetic lineages A1–4, B1–4, C1–4, and D1–4.

In the HPV community, a virus receives an official designation as a reference genome only if the complete viral genome has been sequenced and a viral genomic DNA clone has been sent to the HPV Reference Center in Sweden for validation. To date, reference genomes have been officially designated up to HPV226. However, an even greater number of additional novel HPV types have been identified and sequenced but have not been submitted to the HPV Reference Center. The sequences of these Non-reference genomes can be found in the Papillomavirus Episteme database (see “Relevant Websites” section). Fig. 1 shows a phylogenetic tree of the named reference HPV types found in the *Alpha*-, *Beta*-, *Gamma*-, *Mu*-, and *Nu*- genera.

Papillomaviruses are an ancient group of viruses that have co-evolved with their hosts for millions of years. When *Homo sapiens* emerged they were likely already infected with multiple ancestral HPV types related to those in the genera we recognize today. The HPVs have continued to slowly evolve, adapting to different anatomical and biological niches. Current thinking is that niche adaptation to susceptible host cells inadvertently led to the emergence of HPV-associated oncogenesis.

Virion Structure

HPVs have non-enveloped icosahedral capsids with $T = 7d$ symmetry and are approximately 55 nm in diameter. The capsid is composed of two proteins, L1, the major structural protein, and L2, the minor protein. Each capsid consists of 360 L1 proteins arranged into 72 pentamers (or capsomeres) and a variable number of L2 monomers (observed to range from 12 to 72 molecules). HPV particles are very stable, and are resistant to degradation by various chemical and enzymatic processes. The L1 proteins can self-assemble into Virus-Like Particles (VLPs), and these VLPs are the main components of the current HPV prophylactic vaccines (See Fig. 2).

The HPV capsid encloses a circular, double-stranded DNA genome of approximately 7–8 kbp in length. A relatively unusual feature of papillomaviruses is that the capsids package the viral DNA in the form of chromatin; host core histones organize the genome into a chromatinized minichromosome.

Genome

The circular dsDNA viral genome is organized into the following three parts: the upstream regulatory region (URR) contains elements that regulate transcription and replication; the early region contains genes expressed in the early and maintenance phases of infection; and the late region encodes the proteins expressed at late stages of the viral life cycle. Fig. 3 shows the genomic arrangement of an oncogenic Alpha-HPV genome.

All viral mRNA is expressed from one strand of the viral genome. In the oncogenic Alpha-HPVs, the early viral genes are transcribed from promoters located in the URR (just upstream from the early region) and terminate at the early polyadenylation site. An intermediate class of transcripts (that encode E1, E2, and E1E4) use the late promoter (located in the E7 open reading frame (ORF) in the oncogenic Alpha-HPVs) and terminate at the same early polyadenylation site. Finally, late transcripts initiate from the late promoter and terminate at the late polyadenylation site at the end of the coding region. Other HPVs have additional

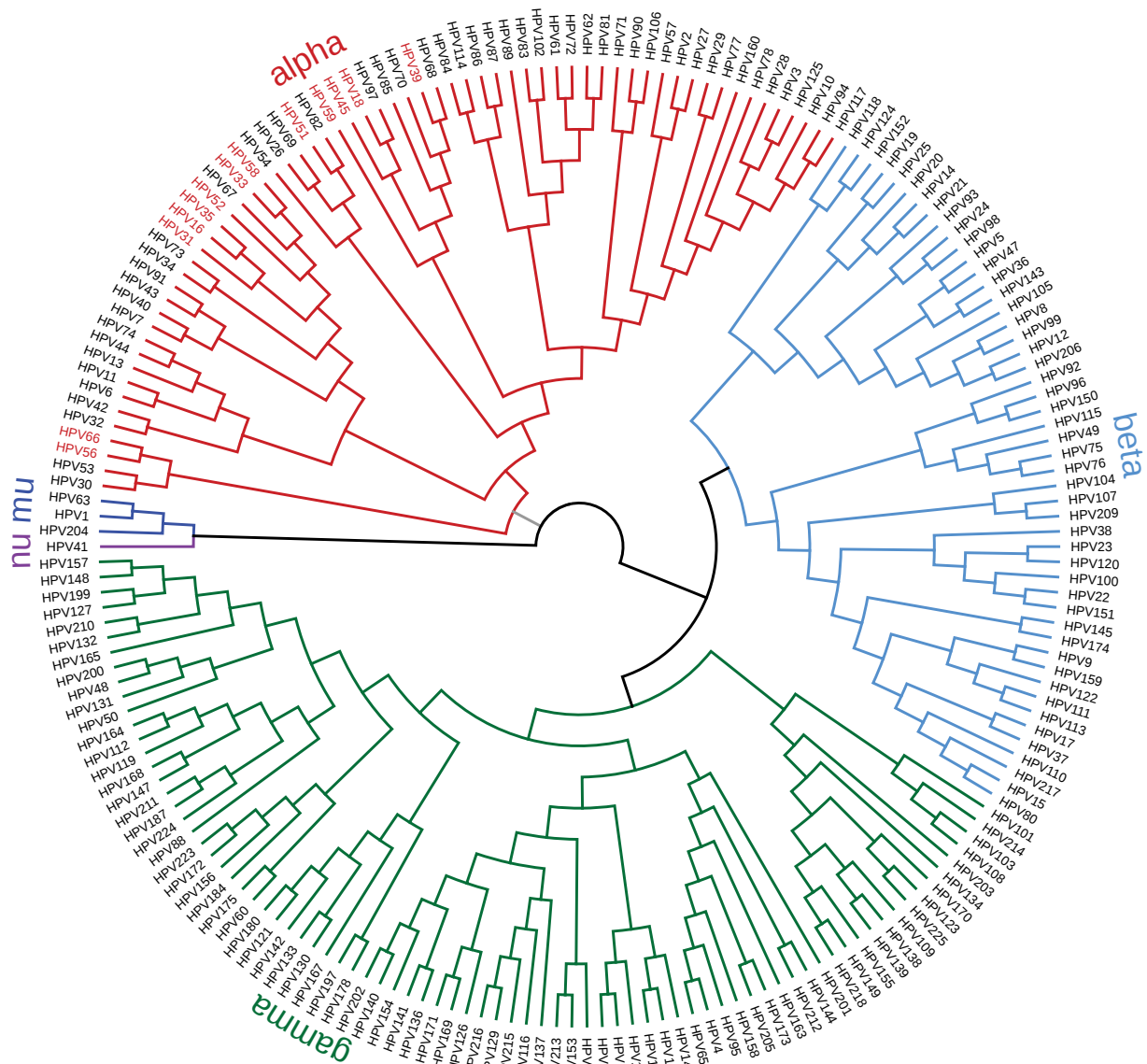


Fig. 1 Phylogenetic tree of human papillomaviruses. Phylogenetic tree of reference HPVs based on the nucleotide sequence of the L1 gene. The five genera of human viruses are shown and the names of oncogenic Alpha-HPVs are shown in red.

promoters in the early coding region of the viral genome. The use of different viral promoters and polyadenylation signals results in a diverse assortment of polycistronic transcripts, many of which are further alternatively spliced, thus allowing efficient use of the limited coding capacity of the small genome.

All papillomaviruses contain four conserved open reading frames that encode two replication proteins, E1 and E2, and two capsid proteins, L1 and L2. All HPVs also express two additional proteins encoded from spliced transcripts: E8E2 is a truncated E2 protein that antagonizes the function of full-length E2 and E1E4 is an intermediate protein encoded primarily from an alternative open reading frame in the E2 gene. Almost all HPVs express E6 and E7 proteins. In addition, the oncogenic Alpha-HPVs (and those that are evolutionarily closely related) express a spliced E6 transcript, E6*. In the Alpha-HPVs, short, hydrophobic, transmembrane proteins (called E5) are encoded at the end of the early region. Sometimes more than one evolutionarily distinct E5-like protein is encoded in this region and they are named E5 α , E5 β , E5 γ , or E5 δ .

Host Range and Tissue Tropism

Papillomaviruses are highly species specific and tissue tropic. HPVs will only infect and replicate in the cutaneous and mucosal epithelial surfaces of *Homo sapiens*. HPVs enter and establish a persistent infection of keratinocytes and rely on the differentiation

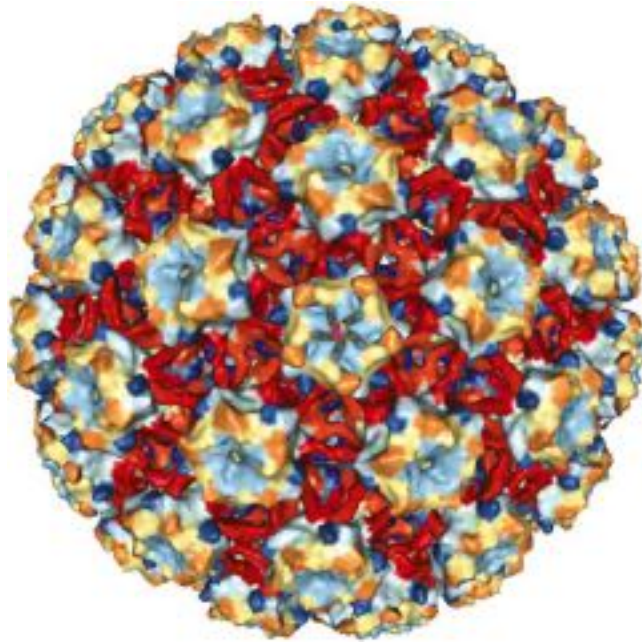


Fig. 2 Structure of an HPV L1 capsid. Electron cryo-microscopy of HPV16 L1 capsid, created from the PDB structure 3J6R using the NGL viewer. The icosahedral capsid is composed of 360 L1 monomers arranged into 72 star-shaped pentamers.

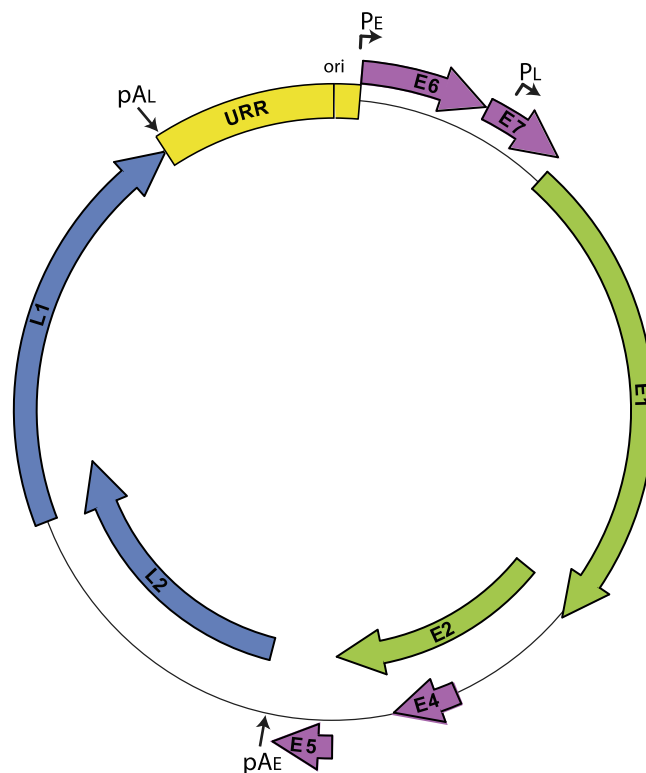


Fig. 3 Viral genome. A genome map of the oncogenic Alpha-papillomavirus HPV18. The core early genes shown in green, and those encoding the late, capsid proteins in blue. The E4, E5, E6, and E7 genes are shown in purple. The URR (upstream regulatory region) contain transcriptional regulatory elements and the origin of replication (ori). The early and late promoters (P_E and P_L) and polyadenylation sites (pA_E and pA_L) are indicated.

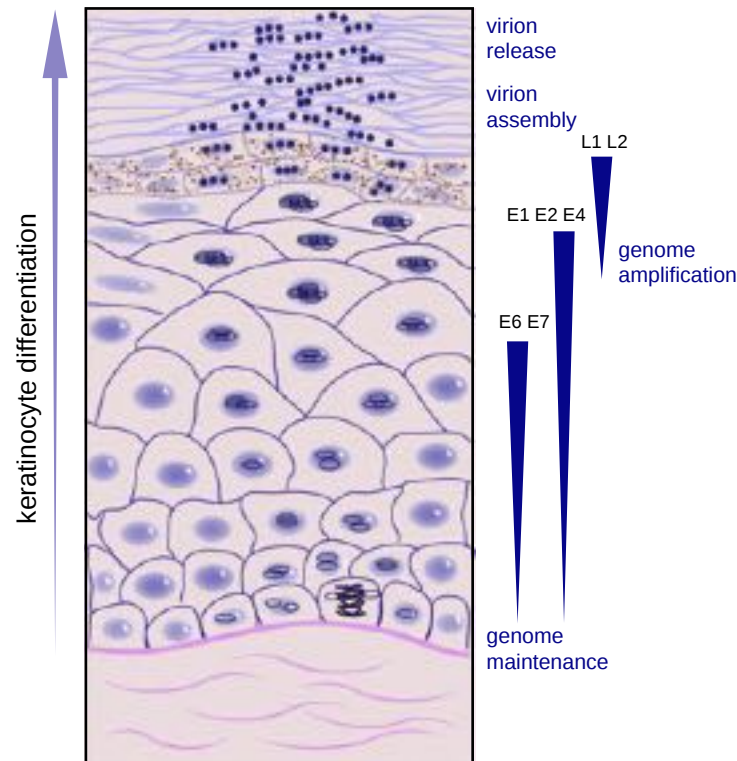


Fig. 4 *Differentiation-dependent Life Cycle of Papillomaviruses.* The stages of the HPV lifecycle in a stratified epithelium. The virus enters the basal cells through a fissure in the epithelium and the genome replicates here at low levels. As the infected cells move up to the surface of the epithelium during the differentiation process, high levels of viral gene expression and genome amplification are induced. Viral particles assemble in the upper layers and are shed from the surface in squames.

program of these cells to support viral production. In addition, each HPV replicates in specific types or anatomical regions of the host epithelium. This tropism is not well understood and does not appear to be due to entry receptors; rather it is thought that each virus relies on specific cellular processes (e.g., transcription, tissue renewal, differentiation) to support the viral life cycle. Likely, the functions of the E5, E6, and E7 proteins of each HPV evolved to take advantage of these niches. Cancers of the cervix often arise at the transformation zone, which is the junction between squamous and columnar epithelial cells. Cells that reside here are especially susceptible to HPV-associated carcinogenesis.

Life Cycle

The HPV life cycle is intricately and inexorably tied to the differentiation program of the host keratinocyte (Fig. 4). In a stratified epithelium, only the basal cells divide. Keratinocyte division can occur symmetrically to produce two basal cells, or asymmetrically to produce one basal cell and a second daughter cell that differentiates and moves slowly towards the surface of the epithelium as part of the terminal differentiation process. Infection of the basal cells ensures that the viral genome is maintained for long periods of time and at the same time will continually yield cells that eventually produce progeny viral particles.

The different stages of viral replication and gene expression are temporarily and spatially regulated at the different stages of keratinocyte differentiation. Viral gene expression and DNA replication are very low in infected basal cells but are activated in suprabasal cells to amplify the viral genome to high levels. In even more differentiated cells, late viral gene expression produces the L1 and L2 proteins to package the viral genome and form infectious progeny virions. Compared to many viruses, the HPV infectious cycle is relatively long, with at least three weeks from initial infection to the first production of progeny virions.

Entry

The HPV life cycle begins when an infectious virion accesses the basal keratinocytes exposed by microabrasions in the host epithelium. The major L1 capsid protein initially interacts with heparin sulfate proteoglycans on the exposed basement membrane. Here, L1 is cleaved by Kallikrein-8 (KLK8), causing a conformational change in the capsid. Subsequent interaction with cyclophilin

Table 1 Functions of the HPV Proteins

<i>Protein</i>	<i>Function</i>
E1	Origin binding replicative helicase
E2	DNA binding protein; E1 helicase loader, transcriptional regulator, tethering and partitioning of viral genome
E8E2	DNA binding protein; repressor of viral transcription and replication
E1E4	Late protein; disrupts keratin filaments; aids virion release
E5	Regulates cell cycle, contributes to viral egress
E6	Immune evasion, manipulates cell pathways to produce a cellular environment conducive to persistent viral replication
E7	Immune evasion, manipulates cell pathways to produce a cellular environment conducive to persistent viral replication
L1	Major structural protein
L2	Minor structural protein, trafficking of viral DNA during entry, packaging viral DNA

B triggers a second conformational change in the virion, exposing a furin cleavage site in the minor capsid protein L2. Furin cleavage of L2 exposes the RIG-1 epitope and facilitates the transfer of the virion to an unknown secondary receptor on the basal keratinocyte. The viral particle is internalized by endocytosis and the low pH in the endosome, and the activity of cyclophilin B, causes L1 to disassemble from the particle, leaving a complex of the L2 protein and viral chromatin. L2 is inserted into the endosomal membrane resulting in virus containing vesicles that move through the trans golgi network. These vesicles enter the nucleus at the onset of mitosis when the nuclear membrane breaks down. They bind to condensed mitotic host chromosomes until cell division is complete.

Early Stages of Infection

After cell division is complete, the L2-viral minichromosome complex is released from the membrane vesicles and localizes to ND10 bodies to begin the early part of the life cycle. Many viruses associate with ND10 bodies early in infection, even though these bodies are thought to be involved in anti-viral defense. However, viruses often disrupt and reorganize ND10 bodies by displacing repressive components and taking advantage of positive factors. Similarly, the HPV L2 protein displaces the host Sp100 protein from ND10 bodies to promote a local nuclear environment conducive to early viral replication and transcription. The viral early promoter drives transcription of early genes, including those of the replication proteins E1 and E2. E1 and E2 bind the replication origin, unwind the viral DNA, and work in concert with the cellular DNA replication machinery to amplify the viral genome to a low copy number (just a few copies per cell). The new viral genomes establish a persistent infection by localizing to beneficial regions of the host nucleus that will support low level gene expression, and evade host defenses.

Maintenance Phase of Infection

During this phase of the life cycle, the late structural genes are not transcribed and no infectious particles are formed, thus it is dubbed the “non-productive” phase. To maintain a long term non-productive infection, the virus must keep a low immunogenic profile and faithfully pass the genome onto successive daughter cells as the basal keratinocytes proliferate. The E2 protein binds to sites in the viral genome and tethers the extrachromosomal viral genome to the host chromosomes. This ensures that the viral genome is retained in the nucleus and distributed to daughter cells in mitosis. In this way, the viral genome persists as a low copy extrachromosomal circular DNA.

Late Stages of Infection

When cells maintaining HPV genomes detach from the basement membrane and begin their journey upward through the epithelium, the virus begins the late (or “productive”) phase of its life cycle. In these terminally differentiating cells, expression of the viral E5, E6, and E7 proteins delay the normal keratinocyte differentiation program and disrupt cell cycle regulation. High levels of the E1 and E2 proteins induce massive amplification of the viral genome in differentiated cells that are in a G2-like phase. HPVs rely on the cellular DNA replication machinery to replicate viral DNA but this is not present in G2 phase or differentiated cells. Therefore, HPVs induce a DNA damage response and repair factors traffic to the viral genome to replicate viral DNA in nuclear foci using what is presumed to be recombination-directed replication mechanisms. Following this, transcription and translation of the late structural proteins L1 and L2 occur; these proteins assemble around chromatinized progeny genomes to generate infectious virions. Further maturation of the virions occurs in the oxidizing environment of the uppermost, dying keratinocytes. These cells, containing arrays of mature virions, are sloughed off the surface of the epithelium by desquamation. Prior activities of the E4 and E5 proteins increase the fragility of these squames, so that they will more easily release viral particles.

Functions of Individual Viral Genes (Table 1)

The E1 Protein

The E1 gene product is a sequence-specific origin binding replicative helicase that is essential for viral DNA replication. Homodimers of the E1 and E2 proteins initially bind cooperatively to their respective binding sites in the viral origin of replication. A conformational shift in the E1 protein displaces E2 and recruits additional E1 proteins to form a double hexamer, which is the functional replicative helicase that unwinds the DNA bidirectionally and allows cellular replication machinery to bind and process the template.

The E2 Protein

The viral E2 protein is multifunctional; it is involved in replication, transcription, and maintenance of the viral genome. E2 consists of an N-terminal transactivation domain linked by a flexible “hinge” sequence to the C-terminal DNA binding/dimerization domain. E2’s main role in initiation of replication is to assist in loading the E1 helicase onto the origin of replication, but it further contributes by displacing nucleosomes and recruiting cellular replication proteins. E2 additionally functions in maintenance replication by partitioning the viral DNA to daughter cells by tethering the genome to host chromatin. E2 is also the primary regulator of HPV transcription, functioning predominantly by recruiting cellular activating or repressing factors to promoters and enhancers in the viral genome through its E2 binding sites (E2BSs).

The E8E2 Protein

The E8E2 protein arises from a spliced transcript that encodes a short peptide from the E8 ORF fused to the C terminal DNA binding and dimerization domain of E2. E8E2 can competitively bind to E2 binding sites and recruits repressive cellular NCoR/SMRT complexes which restricts HPV transcription and replication.

The E1E4 Protein

The E4 protein is expressed as an E1E4 fusion protein from a spliced transcript encoding the first few codons of E1 joined to the E4 ORF. The resulting protein is a small, mostly cytoplasmic protein that interacts with a variety of cellular partners. Despite its location in the early region of the viral genome, high levels of E1E4 expression correlate with the start of productive viral DNA synthesis in the later stages of the life cycle, where it is among the most abundant protein of the infected cell. The exact role of E1E4 in viral genome replication is not fully understood. E1E4 induces G2/M cell cycle arrest by sequestering cyclins and cyclin dependent kinases, and interacts with many other cellular proteins. The major function of E1E4 may be its contribution to the release of progeny HPV virions by associating with and destabilizing cornified cell envelopes and keratin filaments, thus aiding virion egress from infected, desquamated cells.

The E5 Protein

The E5 proteins are small hydrophobic membrane proteins that are encoded only by the human Alpha-papillomaviruses (as well as many animal papillomaviruses). At the sequence level, there are several different types of E5 protein (named alpha, beta, gamma and delta) that are each evolutionarily distinct. For the most part, the oncogenic Alpha-HPVs encode alpha-E5 proteins and these are the best studied.

E5, together with E6 and E7, promotes proliferation of infected cells. Alpha-E5 proteins increase cell growth by enhancing the signaling capacity of the epidermal growth factor receptor (EGFR). Additionally, E5 forms a complex with the 16K subunit of the V-ATPase and this may impair keratinocyte cell-cell communication. E5 also hinders cellular antigen presentation by blocking HLA-I from trafficking to the cell surface and by inhibiting the maturation of MHCII. The alpha-E5 proteins encoded by oncogenic high-risk HPVs inhibit apoptosis and maintain the proliferative capacity of differentiating keratinocytes. Moreover, HPV E5 (along with E6) has been proposed to contribute to viral egress by inducing koilocytic vacuoles in the nucleus of infected cells, thus increasing cell fragility and aiding virion release from desquamated cells.

The E6 Protein

The E6 proteins contain two zinc finger domains linked together by an alpha helix. Together, these domains give E6 proteins the ability to interact with different cellular proteins that contain acidic, helical LxxLL motifs. The Alpha-HPV E6 proteins bind to the cellular E3 ubiquitin ligase, E6 Associated Protein (E6AP). In the oncogenic Alpha-HPVs this complex ubiquitinates p53, leading to p53 degradation by the proteasome and allowing the cell cycle to proceed with G1/S progression. This action by E6 to degrade p53 is required for the long-term maintenance of the viral genome. On the other hand, the Beta- and Mu-HPV E6 proteins bind LxxLL motifs on mastermind-like (MAML) proteins. MAML proteins are transcriptional coactivators for Notch signaling and inhibit the ability of Notch to promote squamous epithelial differentiation. The oncogenic Alpha-HPV E6 proteins contain a

C-terminal region that allows them to bind and degrade PDZ domain proteins involved in cell polarity. These E6 proteins also induce telomerase expression, allowing proliferating cells to escape senescence due to eroding telomeres. In contrast, many Beta-HPV E6 proteins increase tolerance to UV-mediated DNA damage and in this way could act as a cofactor in the development of squamous cell skin cancer. The multifunctional HPV E6 proteins also inhibit innate immune response pathways.

The E7 Protein

Almost all HPVs encode an E7 protein, a small early protein that binds multiple cellular targets. The main functions of E7 are to ensure that the cellular environment is favorable for HPV replication. In the high-risk Alpha-HPVs, the E6 and E7 proteins manipulate the cell in such a way as to promote genetic instability and oncogenic progression of the infected cell are thus considered to be oncogenes. For example, high-risk E7 promotes cell cycle progression by binding and degrading the tumor suppressor retinoblastoma protein (pRb), thus overriding its ability to inhibit the progression of the cell cycle from G1 into S phase and promoting cellular proliferation. This stabilizes p53, which in turn is degraded by high-risk E6. E7 also causes global, epigenetic reprogramming by upregulating Lysine Demethylase 6 (KDM6) A and B. E7 also induces cellular DNA damage signaling to promote viral DNA amplification in differentiated cells. Similar to the E6 proteins, E7 proteins inhibit the host innate immune responses.

The L1 Protein

L1 is the major structural protein of the HPV capsid. The L1 protein forms pentamers called capsomeres, and 72 of these self-assemble into an icosahedral particle with $T = 7$ symmetry. The L1 protein is sufficient to form virus-like particles (VLPs). L1 initiates the infection process by binding to the extracellular matrix (ECM) at the base of a stratified epithelium, and then facilitates entry of the virion into the cell by interaction with a secondary receptor on the basal keratinocyte. While most of the L1 dissociates from the viral genome throughout the trafficking to the nucleus, some L1 remains associated with HPV DNA. In the late stage of the viral life cycle, L1 is expressed in the uppermost layer of the epithelium where, together with L2, it encapsidates viral DNA to produce progeny virions.

The L2 Protein

L2 is the minor capsid protein and it has several important functions during viral entry and trafficking of the viral genome to the nucleus. After endocytosis, L2 becomes transmembranous and redirects the viral genome complex to the nucleus in a membrane vesicle via the retromer pathway. When the infected cell undergoes mitosis, the L2-genome containing membrane vesicle associates with host mitotic chromosomes. After the nuclear envelope reforms, the viral complex localizes to the ND10 nuclear bodies. Here, L2 displaces Sp100, a restriction factor of HPV, to facilitate an environment conducive to viral transcription. At late stages of infection L2 is expressed in the upper layers of the epithelium where it assists in packaging of the viral genome.

Natural History, Transmission and Epidemiology

HPVs are ubiquitous and, in most cases, cause asymptomatic or transient, self-limiting infections. The viral particles are very stable and are released from the desquamating surface of infected epithelia. Transmission of cutaneous types are by direct host to host contact, or through contact with a contaminated surface. Anogenital and oral HPV types are usually transmitted by sexual contact. In both cases, the virus enters the epithelium and accesses the basal epithelial cells through small scratches or microabrasions.

HPV infection is extremely common. Children and adolescents have the highest prevalence of warts of the hands and feet, and a high proportion of individuals become infected with anogenital HPVs when they become sexually active. In most cases, the infections are self-limiting and regress without treatment. However, individuals persistently infected with oncogenic HPV types are at risk of progression to cancer. Widespread vaccination with the current HPV vaccine can prevent infection with the most common oncogenic types.

Clinical features, Pathogenesis and Histopathology of HPV Infection

There are hundreds of different HPV types and they each infect different regions of the cutaneous epithelium (skin), the anogenital mucosal epithelium, and the epithelium of the oral cavity. These infections give rise to common, plantar, flat plane or anogenital warts (papillomas). Lesions and pre-cancers are also found in the oral cavity and anogenital regions. Many HPV types (in particular those from the Beta and Gamma genera) persist asymptotically in the cutaneous epithelium, resulting in lesions only in individuals with immunodeficiencies.

The cell layers of the epithelium are greatly expanded in many common warts resulting in exophytic lesions with rough, cauliflower-like appearances. Plantar warts (verruca) are endophytic and infected cells grow deep into the sole of the foot. Plane warts present as flat papules on cutaneous skin surfaces. These warts often show characteristic histological features of acanthosis

(thickening of the epithelium), parakeratosis (nuclei retained into the cornified layer), and hyperkeratosis (a thickened cornified layer). Lesions often contain koilocytes, which are cells with shrunken hyperchromatic nuclei that are displaced by a large perinuclear vacuole. The presence of koilocytes is usually indicative of HPV infection. Anogenital warts (condyloma acuminata) present as discrete, small papules on the genital regions. Laryngeal papillomatosis (recurrent respiratory papillomatosis) is a rare condition associated with low-risk HPV infection (possibly due to an unidentified immunodeficiency). Here, benign papillomas form in the respiratory tract and tend to recur despite multiple surgeries.

Oncogenic HPVs

A subset of Alpha-papillomaviruses are associated with cancer. The E6 and E7 proteins (oncogenes) of these viruses promote unscheduled cellular proliferation and inactivate cell cycle checkpoints. This leads to genetic instability as infected cells accumulate mutations that promote carcinogenesis. Infection with oncogenic HPVs is extremely common and most infections resolve without treatment. Only long term, persistent infection with oncogenic HPVs is associated with oncogenic progression.

The International Agency for Research on Cancer (IARC) has reported that 13 Alpha-HPVs are oncogenic and that close, evolutionarily related types are “probably” oncogenic. [Fig. 1](#) shows a phylogenetic tree of human papillomaviruses with the oncogenic Alpha-HPVs indicated in red.

In many cancers associated with HPV infection, the viral genome is integrated into the host chromosomes. Integration is not part of the normal viral life cycle and is a dead end for virion production. The only viral proteins consistently expressed in cancers are the E6 and E7 oncogenes. It is thought that the viral genome accidentally integrates into regions of the host genome that promote dysregulated expression of the E6 and E7 oncoproteins. The resulting increase in cellular proliferation, absence of cell cycle checkpoints, and promotion of genomic instability drives infected cells toward oncogenesis.

HPV-associated Cancers

About 5% of human cancers can be attributed to oncogenic Alpha-HPV infection. Persistent HPV infection is the causative agent of almost all cancers of the uterine cervix and the majority of vaginal, vulvar, penile, anal and rectal cancers. A large proportion of oropharyngeal cancers are also due to HPV infection. Beta-HPVs cause asymptomatic infection of the skin and are often found in squamous cell carcinomas. However, they are not essential for the maintenance of the tumors and may play a role in the development of cancer by impairing the repair of UV damage. Conversely, these infections may also enhance CD8⁺ T cells anti-tumor activity and suppress cancer.

Immune Response

HPVs must avoid host immune defenses and the virus replication cycle itself is a strategy of immune evasion. There is no viremia or cell lysis associated with HPV infection and high levels of viral genome replication, transcription and protein synthesis are confined to terminally differentiated cells. Viral particles are shed from the surface of the epithelium and so do not induce an immune response. The virus also suppresses interferon and other cytokine responses and downregulates innate immune sensors.

Most infections are self-limiting and regression is due to a cell-mediated immune response. Infection does not always lead to seroconversion and levels of natural antibodies to HPV antigens are low. This is in contrast to the immune response to HPV vaccines, which induce extremely high levels of anti-L1 neutralizing antibodies. The levels of these antibodies contained in serum exuded at epithelial microabrasions are sufficient to prevent infection.

However, HPV infection is not completely invisible to host immune defenses as evidenced by the great increase in pathogenic HPV lesions and cancers in cases of immunosuppression and immunodeficiency. Individuals with uncontrolled HIV infection, or immunosuppressed organ transplant recipients, often have prolific HPV disease. Some genetic primary immunodeficiencies also give rise to uncontrolled HPV infections. For example, individuals with WHIM syndrome (Warts, Hypogammaglobulinemia, Infections, and Myelokathexis) are very susceptible to HPV warts, but treatment of the underlying immune defect results in resolution of HPV lesions. In the rare genetic disease Epidermodysplasia verruciformis, Beta-HPVs that usually cause asymptomatic infections in healthy individuals give rise to plaques of HPV infection on the cutaneous epithelium with a high propensity to develop into squamous cell carcinomas on UV-exposed skin.

Diagnosis

Dermatologists and gynecologists diagnose common, plantar, plane and genital warts from clinical and histological appearance; evaluation of HPV type is usually unnecessary. However, because cervical HPV infections are common (and can lead to pre-cancerous and cancerous lesions), a routine gynecological exam incorporates a Pap smear (invented by George Papanicolaou). In this test, cells are brushed from the surface of the cervix and analyzed by cytopathologists for signs of HPV infection or other abnormalities. For follow up, cervical (or anal) pre-cancerous lesions are treated in situ with acetic acid so that they can be visualized by colposcopy or anoscopy. Small biopsies of these lesions are further analyzed histologically to determine the stage of cervical or anal squamous intraepithelial lesions (ASIL or CSIL). HPV types that infect the anogenital region can be classified as

“low oncogenic risk” or “high oncogenic risk” and tests are available to determine the HPV DNA type and thus the relative risk of disease progression.

Treatment

HPV infections are very common, and in most healthy individuals are self-limiting and require no treatment. If bothersome, common warts of the hands and feet can be treated with salicylic acid (available as over-the-counter treatments). Topical application of creams containing imidazoquinoline-based IRMs (Immune Response Modifiers) can enhance the host immune response and induce regression of various types of warts. Persistent pre-cancerous lesions of the cervix are removed by ablation techniques such as laser vaporization, cryotherapy or LEEP (loop electrosurgical excisional procedure), while HPV associated cancers are treated by surgery, chemotherapy, and radiotherapy.

Prevention

Vaccination with prophylactic HPV vaccines is the most effective means of prevention. To date, three highly effective vaccines have been developed against the most common HPV types that cause anogenital infection and associated cancers. Each of these vaccines contains the HPV L1 protein, self-assembled into VLPs (Virus-Like Particles) and an adjuvant. Cervarix contains HPV16 and HPV18 VLPs, Gardasil-4 contains HPV6, 11, 16, and 18 VLPs, and Gardasil-9 contains HPV6, 11, 16, 18, 31, 33, 45, 52, and 58. The VLPs are extremely immunogenic and when delivered intramuscularly induce high titers of neutralizing antibodies. The efficacy of these vaccines in preventing warts and cancer precursor lesions in the anogenital region is excellent and there is hope that in countries with high rates of vaccination uptake, HPV-associated cancers could eventually be eliminated.

Acknowledgments

The research of the authors is supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health.

Further Reading

- Cubie, H.A., 2013. Diseases associated with human papillomavirus infection. *Virology* 445, 21–34.
- de Oliveira, C.M., Fregnani, J., Villa, L.L., 2019. HPV vaccine: Updates and highlights. *Acta Cytologica* 63, 159–168.
- Lambert, P.F., McBride, A., Bernard, H.U., 2013. Special Issue: The Papillomavirus Episteme. Academic Press.
- McBride, A.A., Munger, K., 2018. Expert views on HPV infection. *Viruses* 10.
- Schiffman, M., Doorbar, J., Wentzensen, N., *et al.*, 2016. Carcinogenic human papillomavirus infection. *Nature Reviews Disease Primers* 2, 16086.
- Strickley, J.D., Messerschmidt, J.L., Awad, M.E., *et al.*, 2019. Immunity to commensal papillomaviruses protects against skin cancer. *Nature* 575, 519–522.
- van Doorslaer, K., Chen, Z., Bernard, H.U., *et al.*, 2018. ICTV virus taxonomy profile: Papillomaviridae. *Journal of General Virology* 99, 989–990.
- Vande Pol, S., 2015. Papillomavirus E6 oncoproteins take common structural approaches to solve different biological problems. *PLoS Pathogens* 11, e1005138.

Relevant Websites

- <https://hpvcentre.net/>
HPV INFORMATION CENTER.
- <https://www.cdc.gov/hpv/>
HPV, the Vaccine for HPV, and Cancer Caused by HPV.
CDC.
- <https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-and-cancer>
Human Reference clones.
hpvcenter.
- https://www.hpvcenter.se/human_reference_clones/
Human Reference clones.
hpvcenter.
- <https://pave.niaid.nih.gov/>
PaVE: Papilloma virus genome database.
NIH.

Human Parainfluenza Viruses (*Paramyxoviridae*)

Elisabeth Adderson, St. Jude Children's Research Hospital, Memphis, TN, United States and University of Tennessee Health Sciences Center, Memphis, TN, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of E. Adderson, A. Portner, Parainfluenza Viruses of Humans, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00459-3](https://doi.org/10.1016/B978-012374410-4.00459-3).

Classification

The family Paramyxoviridae, in the order Mononegavirales, consists of large negative-sense, single-stranded, enveloped RNA viruses belonging to 4 subfamilies and 14 genera. The 4 human parainfluenza viruses (HPIVs) belong to 2 separate genera within this family. HPIV-1 and HPIV-3, as well as bovine, murine, and porcine parainfluenza viruses, are members of the Respirovirus genus. HPIV-2 and HPIV-4 are members of the Orthorubulavirus genus, which also includes mumps virus and several pathogens of non-human mammalian species. HPIV-4 is further divided into two subtypes, HPIV-4a and HPIV-4b, on the basis of serologic reactivity.

Virion Structure

HPIVs are roughly spherical viral particles 150–250 nm in diameter consisting of a host-cell plasma membrane–derived lipid bilayer surrounding an 18-nm diameter internal helical nucleocapsid containing the negative-sense, non-segmented, single-stranded RNA genome. Projecting from the surface of the virion are 2 transmembrane glycoproteins, hemagglutinin–neuraminidase (HN) and fusion (F) protein. These glycoproteins are anchored to the plasma membrane of the infected cell or the virion envelope by a hydrophobic transmembrane region. A schematic representation and an electron micrograph of a naturally occurring parainfluenza virus (PIV) are shown in [Fig. 1](#).

Genome

The genomes of HPIVs range from 15,400–17,000 nucleotides and encode 6 common structural proteins, linked in tandem in the order 3′-nucleocapsid protein (NP)–phosphoprotein (P)–matrix protein (M)–F–HN–large protein (L)–5′. In general, the genome organization of HPIV is remarkably similar, but differences exist in gene sequences, intergenic regions, and nonstructural proteins expressed from the P gene. Sequence and immunological analyses suggest that human HPIV-1 and murine Sendai virus are closely related, as are human- and bovine-origin PIV-3 [Fig. 2](#).

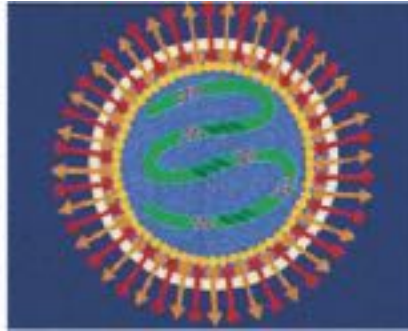
The RNA nucleocapsid, a complex consisting of approximately 2000 molecules of NP bound to the RNA genome, rather than the RNA genome itself, serves as a template for RNA synthesis. The viral RNA-dependent RNA polymerase complex consists of a heterocomplex of about 200 P and 20 L molecules. The P gene is unique in its capacity to express not only P protein but also alternative proteins. This varying expression is determined by either use of internal initiation codons in the same or different reading frames or by insertion of non-templated G residues during the editing of the P gene specific mRNA. These alternative proteins have been designated V, C, and D, depending on the PIV. The approximate molecular weights in Daltons (Da) of the HPIV proteins, as exemplified by HPIV-3, are as follows: NP (58,000 Da); P (68,000 Da); M (40,000 Da); F (63,000 Da); HN (72,000 Da); and L (256,000 Da) [Table 1](#).

Life Cycle

HPIV replication starts in ciliated respiratory epithelial cells, initially in the nose and throat, and may subsequently progress to affect ciliated and alveolar cells of the lower respiratory tract. Infection is initiated by the attachment of viral particles to the host cell, the energy-dependent fusion of the viral envelope and target cell membrane, and the formation of pores through which the viral genome is introduced into the cell. The HN and F glycoproteins cooperatively mediate the direct binding of the virion to sialic acid–containing glycoprotein or glycolipid receptors on upper and lower respiratory tract epithelial cells. The same process is responsible for the hemagglutination of avian and mammalian erythrocytes. HN also causes the enzymatic (neuraminidase) cleavage of sialic acid residues from the carbohydrate moiety of glycoproteins and glycolipids, which functionally serves to prevent the self-aggregation of virus during release and likely aids in the spread of virus from infected cells. HN possesses an N-terminal stalk region of approximately 130 amino acids anchoring a large glycosylated hydrophilic globular head region to the viral envelope. A small uncharged hydrophobic peptide near the N-terminus spans the viral envelope, and a small hydrophilic domain is internal to the membrane. The globular head contains the active site for virus attachment, neuraminidase activity, and antigenic determinants that induce neutralizing antibodies for HPIVs. HN exits on the surface of the virion as disulfide-linked dimers of homodimers [Fig. 3](#).



Fig. 1 HPIV genome. The genome consists of 6 common structural proteins, linked in tandem in the order 3'-nucleocapsid protein (NP) – phosphoprotein (P) – matrix protein (M) – fusion protein (F) – hemagglutinin-neuraminidase (HN) – and large protein (L) – 5'.



Membrane proteins

- **HN**: receptor-binding, neuraminidase, fusion promotion
- **F**: membrane fusion
- **M**: assembly

Nucleocapsid

- **NP**: encapsidate viral RNA
- **P**: forms polymerase complex with L
- **L**: RNA synthesis, capping, polyadenylation

Fig. 2 Parainfluenza virus structure.

Table 1 HPIV-3 proteins

Name	Function	Size (Da)
Nucleocapsid protein (NP)	Structure, RNA binding	58,000
Phosphoprotein (P)	RNA-directed RNA polymerase	68,000
Matrix protein (M)	Structural, virion assembly	40,000
Fusion protein (F)	Fusion of virus membrane with host cell membrane	63,000
Hemagglutinin-neuraminidase (HN)	Viral attachment and entry into host cell, sialidase	72,000
Large protein (L)	RNA-directed RNA polymerase	256,000

Besides being needed for attachment, HN is essential to or enhances the fusion activity of the F protein. F proteins are class I fusogenic viral proteins that are synthesized as inactive precursors, then post-translationally cleaved by a host cell trypsin-like protease to form prefusion homotrimers. Following receptor binding, a conformational change in HN activates the F protein, triggering the insertion of hydrophobic membrane attack domains into the target cell membrane. The cell and viral membranes are then approximated by refolding of F into a hairpin structure. Membrane fusion precedes the formation of a pore through which the viral nucleocapsid is deposited into the host cell cytoplasm.

After the infecting nucleocapsid is introduced into the host cell, the polymerase complex transcribes viral mRNAs and initiates genome replication by generating an intermediary positive-sense RNA antigenome template. HN and F proteins are synthesized, post-translationally modified, and transported to the host cell membrane, which becomes the envelope for progeny virions. Once protein synthesis is underway, N protein encapsulates newly synthesized genomes that, in association with P, form new nucleocapsids that are also transported to the cell plasma membrane. The M protein lines the inner surface of the viral envelope. During infection, M associates with the inner leaflet of the plasma membrane, where it orchestrates the release (budding) of progeny virus into the respiratory tract lumen by interacting with specific sites on the cytoplasmic tail of the viral glycoproteins and the nucleocapsid and by transporting viral components to the budding site. HN facilitates the release of new virus particles by cleaving sialic acid residues on the host cell membrane.

Embryonated hen's eggs can support the growth of some strains of HPIV-1, HPIV-2, and HPIV-3. All 4 HPIV types grow well in primary monkey or human kidney cells, which are also used to isolate virus from clinical samples. LLC-MK2, a rhesus monkey kidney cell line, offers an efficient system for isolating PIVs and an experimental tissue culture system. HPIV-2 and HPIV-1 require trypsin in the medium to cleave the F glycoprotein for cell growth. Viral infection of tissue culture can be detected by hemadsorption with chicken and guinea pig erythrocytes and by the cytopathic effects produced by the viruses.

Epidemiology

The HPIVs are important causes of upper and lower respiratory tract infections worldwide (Table). Between 25%–40% of HPIV detections are in children aged 2 years or younger; and half are in children aged 7 years or younger. By 5 years of age, almost all persons have serological evidence of HPIV-3 infection, and 75% have been exposed to HPIV-1. HPIV-2 infection, but not infection

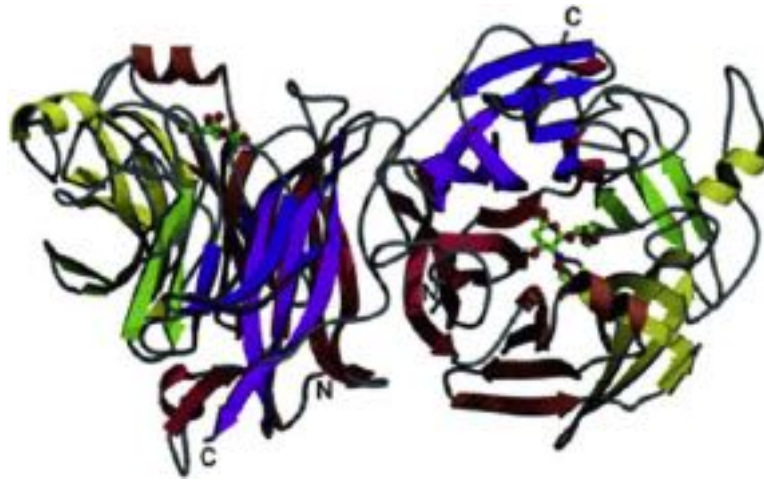


Fig. 3 Structure of an HPIV-3 hemagglutinin–neuraminidase dimer shown in colored ribbon representation, complexed with β -Sialic acid in red and green ball-and-stick representation. Reproduced from Lawrence, M.C., Borg, N.A., Streltsov, V.A., *et al.*, 2004 Structure of the hemagglutinin–neuraminidase from human parainfluenza virus type III. *Journal of Molecular Biology* 335, 1343–1357, used with permission.

Table 2 Epidemiological and Clinical Features of HPIV

	<i>Infections</i>	<i>Seasonality</i>
HPIV-1	Croup Other URTI ^a Pneumonia in young infants, elderly adults	Autumn, biannual in odd-numbered years
HPIV-2	Croup Other URTI ^a Pneumonia in young infants, elderly adults	Autumn, biannual in even-numbered years
HPIV-3	Bronchiolitis, pneumonia in young infants, elderly adults Severe respiratory infection in immunocompromised patients Croup Other URTI ^a	Spring-summer, annual
HPIV-4	Other URTI ^a Pneumonia in young infants	Autumn, annual

^aOther URTI includes rhinitis, pharyngitis, sinusitis, otitis media.

due to other HPIVs, is more common in males. In the United States, HPIV-1 and HPIV2 infections generally peak in the autumn of alternate years (odd- and even-numbered, respectively), whereas HPIV-3 infections are more common in the spring and summer. The circulation of HPIV-4 infections has been less well studied, but detections appear to peak annually in the winter time. Overall, HPIV-3 is the most commonly detected HPIV, followed by HPIV-1, HPIV-2 and, lastly, HPIV-4. The simultaneous detection of more than one HPIV in an individual is unusual, but about one-third of HPIV are co-detected with other respiratory viruses, most commonly with rhino- and enteroviruses and influenza. Reinfections are possible, but generally cause milder disease than primary infection does. Outbreaks in healthcare and long-term care settings have been reported [Table 2](#).

Clinical Features

HPIVs are highly communicable and are transmitted person-to-person by respiratory droplets and by direct contact with e.g., hands and surfaces contaminated by respiratory secretions. The incubation period is 2–6 days; persons are most contagious in the early stages of infection. Virus is typically shed for 4–21 days, longer in children than in adults. Infections can cause rhinitis, pharyngitis, otitis media, sinusitis, infectious laryngotracheobronchitis (croup), bronchitis, bronchiolitis, and pneumonia. Less common manifestations include apnea, febrile seizures, aseptic meningitis, encephalitis, parotitis, pericarditis, myocarditis, disseminated infection, and Guillain-Barré syndrome. Although most infections are mild and self-limited, severe disease may occur in young children, the elderly, and immunocompromised persons. HPIVs are responsible for about 7%–17% of pediatric hospitalizations for febrile or acute respiratory infections. Certain serotypes are associated with specific clinical syndromes: notably

HPIV-1 and HPIV-2, with infectious laryngotracheobronchitis (croup), and HPIV-3, with bronchiolitis and pneumonia in infants. HPIV-4 causes milder disease than other HPIV types do. HPIV, particularly HPIV-3, may cause severe and protracted infections among persons with cellular immunodeficiencies. Rates of progression of upper respiratory tract infection (URTI) to lower respiratory tract infection (LRTI) may be up to 40% in patients with combined immunodeficiency and recipients of hematopoietic stem cell transplants and reported mortality rates range from 10% to 40%. HPIV LRTIs are associated with allograft dysfunction and rejection among lung transplant recipients and cryptogenic organizing pneumonia and persistent airflow obstruction in hematopoietic stem cell recipients, particularly those with graft-versus-host disease. Secondary bacterial and fungal infections may complicate HPIV, and these co-infections are associated with increased morbidity and mortality. HPIV infections may also trigger exacerbations of reactive airways (asthma), chronic obstructive pulmonary disease, and post-infectious olfactory dysfunction.

HPIV-1 and HPIV-2 cause 60%–75% of croup in pediatric outpatients in the United States, resulting in up to 65,000 hospitalizations annually. The illness, which is most common in children aged 6 months to 2 years, is characterized by inflammation of the larynx and trachea, subglottic edema, and airway obstruction. Children typically present with inspiratory stridor and barking cough, generally preceded by a brief history of coryza or other upper respiratory symptoms. Respiratory failure develops in up to 3% of affected children, but mortality is rare.

Pathogenesis

HPIVs bind to specific sialic acid-containing receptors on upper respiratory tract epithelial cells. The virus and host factors contributing to disease pathogenesis are not completely understood. Both viral (F and HN structure) and host features (genetic susceptibility, host cell proteases, immunocompetence, previous infection) are likely to influence the severity of infection. Viral shedding is prolonged, and disease is more severe in persons with T-cell immunodeficiencies, implying that cellular immunity is important in the control of an acute infection. Severe infection occurs following the spread of the infection to the lower respiratory tract. HPIV has direct cytopathic effects, causing ciliary damage, epithelial necrosis, and recruitment of a mononuclear inflammatory response. Recent studies suggest that host inflammatory responses, rather than the direct effects of viral replication, are important determinants of the severity of signs and symptoms of HPIV infection. Alterations in HN receptor binding or activity may influence host inflammatory responses independent of viral replication or the ability to infect epithelial cells. Infection may also trigger long-term bronchial hyperreactivity in genetically susceptible persons.

HPIV infection triggers cellular and innate responses that reduce viral replication and the persistence of virus in the respiratory tract. Protection from reinfection is associated with the development of serum and secretory antibodies against F and HN. Local IgA responses appear to be particularly important for sustained immunity. Various HPIVs can interfere with host immune responses by inhibiting interferon responses, for example by the accessory C protein. HPIVs share certain antigenic determinants, but structural and antigenic variation occurs between serotypes and, to a lesser degree, within strains belonging to the same type. Antigenic diversity is not generally progressive but may contribute to the propensity of HPIV to cause recurrent infections. The duration of protection after acute illness is relatively short, especially in infants, and protection from infection by heterotypic strains is incomplete.

Diagnosis

The gold standard for diagnosis of HPIV infections is culture on primary monkey kidney cells, human embryonic kidney cells, or human lung carcinoma cells, but this process is labor-intensive and slow, requiring incubation for 4–7 days. Nucleic acid amplification tests (NAATs), performed as single assays or as components of multiplexed respiratory panels, are several-fold more sensitive than is culture and are the diagnostic testing modality of choice where available. The sensitivity and specificity of these assays is >95%. Recent study results suggest a potential role for quantitative viral load testing to identify patients at high risk of disease progression or poor outcomes. Antigen testing and enzyme immunoassays have also been developed for HPIV diagnostics; these methods are rapid but less sensitive and specific than NAATs are. Appropriate samples for diagnostic testing include nasal or nasopharyngeal washes or swabs and bronchoalveolar wash fluid. Detecting specific IgM and IgG in paired acute and convalescent sera has limited utility in diagnosing acute infections, but these approaches are useful in research and epidemiological studies.

Treatment and Prevention

Most HPIV infections are mild and self-limited. Treatment is mainly supportive, concentrating on the provision of adequate hydration and supplemental ventilatory support, if needed. Oral or intramuscular corticosteroids reduce airway edema and reduce the duration and severity of symptoms and the risk of respiratory failure in children with croup. Nebulized epinephrine may have similar effects of short duration. Bronchodilators may be useful in patients with wheezing. In immunocompromised patients with severe infections, reductions in immunosuppressive therapy should be considered if feasible. Some experts also recommend using intravenous gammaglobulin in this population, especially those with hypogammaglobulinemia, although its efficacy in this setting is unproven. No licensed antivirals active against HPIV are currently available. Although ribavirin, a synthetic

guanosine analog, has in vitro activity against HPIV, it has not been effective in reducing the progression of URTI to LRTI or in reducing mortality in immunocompromised patients. DAS181 is a recombinant sialidase fusion protein that inhibits HPIV replication by removing cell-surface sialic acids. In early-phase clinical trials of inhaled DAS181 in immunocompromised patients with severe HPIV pneumonia, treatment was well-tolerated and appeared to reduce morbidity; a phase III trial is ongoing. Other antiviral therapies in development include selective inhibitors of HN, fusion, and IMP dehydrogenase and small-molecule inhibitors of RNA polymerase. Adoptive T-cell transfer using donor lymphocytes specific for HPIV is another strategy being investigated for use in severely immunocompromised patients.

Paramyxoviruses are labile and readily inactivated by heat, organic solvents, detergents, light, and low pH. Droplet and contact infection prevention precautions are appropriate in the health care setting. Despite considerable effort, no licensed vaccines to prevent HPIV exist. Early formalin-inactivated vaccines were poorly immunogenic and were scrutinized because of the potential of similarly designed vaccines for respiratory syncytial virus to enhance pulmonary inflammatory responses to infection with wild-type virus. Vaccines currently in early clinical trials include live attenuated HPIV-3 and HPIV-2, bovine PIV-3, human-bovine chimeras, bovine PIV or Sendai virus expressing heterologous HPIV F and HN proteins, and oligomannose-coated liposomal HN. These vaccines have, in general, been safe and immunogenic, although multiple doses are likely to be required, and immunity may be incomplete or of short duration.

Further Reading

- Batista, M.V., El Haddad, L., Chemaly, R.F., 2018. Paramyxovirus infections in hematopoietic stem cell transplant recipients. *Current Opinion in Infectious Diseases* 31, 542–552.
- Branche, A.R., Falsey, A.R., 2016. Parainfluenza virus infection. *Seminars in Respiratory and Critical Care Medicine* 37, 538–554.
- DeGroot, N.P., Haynes, A.K., Taylor, C., *et al.*, 2011–2019. Human parainfluenza virus circulation, United States. *Journal of Clinical Virology* 124, 104261. doi:10.1016/j.jcv.2020.104261.
- Hijano, D.R., Maron, G., Hayden, R.T., 2018. Respiratory viral infections in patients with cancer or undergoing hematopoietic cell transplant. *Frontiers in Microbiology* 9, 3097. doi:10.3389/fmicb.2018.03097.
- Pawelczak, M., Kowalski, M.L., 2017. The role of human parainfluenza virus infections in the immunopathology of the respiratory tract. *Current Allergy and Asthma Reports* 3, 16. doi:10.1007/s11882-017-0685-2.
- Tang, J.W., Lam, T.T., Zaraket, H., *et al.*, 2017. Global epidemiology of non-influenza RNA respiratory viruses: Data gaps and a growing need for surveillance. *Lancet Infectious Disease* 17, e320–e326.

Human Pathogenic Arenaviruses (*Arenaviridae*)

Sheli R Radoshitzky, United States Army Medical Research Institute of Infectious Diseases, Frederick, MD, United States

Juan C de la Torre, The Scripps Research Institute, La Jolla, CA, United States

© 2021 Published by Elsevier Ltd.

History and Classification

The first arenavirus, lymphocytic choriomeningitis virus (LCMV), was isolated in 1933 from a suspected case of St. Louis encephalitis. Shortly after LCMV was found to cause aseptic meningitis in humans, and the house mouse was identified as its natural reservoir. By the late 1960s the *Arenaviridae* family was recognized to include a group of viruses that shared with LCMV common morphology, serology, biochemical features, and a natural history of establishing long-term chronic infections in their natural rodent hosts. The pathogenic potential of arenavirus infections in humans was underscored with the discovery of Junín virus (JUNV) in the late 1950s as the causative agent of Argentinian hemorrhagic fever (AHF), and the subsequent identification of Machupo virus (MACV) and Lassa virus (LASV) in 1962 and 1969 as the etiologic agents of Bolivian hemorrhagic fever (BHF) and Lassa fever (LF), respectively. In the following years, additional pathogenic arenaviruses were discovered, including Guanarito (GTOV), Sabiá (SBAV), and Chapare (CHAPV) viruses in South America, and Lujo virus (LUJV) in Africa. During the last decade the use of the new technologies of deep sequencing and genome discovery has resulted in the discovery of many new arenaviruses, but their contribution to human disease remains to be determined. Notably, arenaviruses have been also discovered in reptiles (snakes) and fish, representing a significant expansion of the host range of arenaviruses.

Current arenavirus classification is based on pairwise sequence comparisons (PASC) of coding-complete genomes. Based on the most current sequence dataset, S segment and L segment nucleotide sequence identities for viruses within the same genus need to be higher than 40% and 35%, respectively. This analysis and additional characteristics, such as genome architecture, host range, modes of transmission and/or sites of replication in the cell, led to the establishment of four genera within the family *Arenaviridae*: *Antennavirus*, *Hartmanivirus*, *Mammarenavirus*, and *Reptarenavirus*. The hosts of mammarenaviruses are rodents, with the exception of Tacaribe virus (TCRV) that has been found in phyllostomid bats and ixodid lone star ticks. Reptarenaviruses and hartmanivirus infect reptilian hosts, and these infections can cause boid inclusion body disease (BIBD). Antennaviruses have a piscine host. Based on antigenic properties, mammarenaviruses have been divided traditionally into two distinct groups. Old World (OW) mammarenaviruses ("Lassa–LCMV serocomplex") include viruses indigenous to Africa and the worldwide distributed LCMV, whereas New World (NW) mammarenaviruses ("Tacaribe serocomplex") include viruses indigenous to the Americas. This classification is largely consistent with phylogenetic data and murid rodent host phylogeny, with OW mammarenaviruses infecting murid rodents primarily in Africa and NW mammarenaviruses infecting cricetid rodents primarily in the Americas.

This article focuses on arenaviruses known to cause disease in humans, all of them within the *Mammarenavirus* genus.

Mammarenavirus Virion Structure

Mammarenavirions are pleomorphic, ranging in size from 40 to more than 200 nm in diameter, with dense lipid envelopes. The virion's surface is decorated with evenly spaced spike projections composed of heterotrimer complexes of the viral glycoproteins GP1 and GP2 and the stable signal peptide (SSP) (Fig. 1(A)). Cryo-EM studies have shown that surface GP complexes are aligned with subagent Z and ribonucleoprotein (RNP) densities, which are packed into a two-dimensional lattice at the inner surface of the viral membrane. Virions contain the L and S RNP complexes organized in circular configurations. The L and S genomic RNAs are not present in equimolar amounts within virions (L:S ratios ~ 1:2), and low numbers of both L and S antigenomic RNAs are also present within virions. In addition, host ribosomes are documented to be incorporated into virions, but the biological significance of this incorporation remains uncertain.

Mammarenavirus Genome Organization and Proteins

Mammarenaviruses have a bi-segmented negative-stranded RNA genome (Fig. 1(B)). Each genome segment, L (large, 7.3 kb) and S (small, 3.5 kb) uses an ambisense coding strategy to direct the synthesis of two proteins from two non-overlapping open reading frames (ORF) of opposite polarities that are separated by non-coding intergenic regions (IGRs). The S RNA encodes the viral NP and the glycoprotein precursor (GPC), whereas the L RNA encodes for the viral RNA dependent RNA polymerase (RdRp) (L protein) and the matrix Z protein.

Mammarenavirus genomes exhibit high degree of sequence conservation at their 3'-termini and similarly to other viruses with segmented negative strand (sNS) RNA genomes, they exhibit 5'- and 3'-inverted complementary sequences at their L and S genome segments that are predicted to form panhandle structures. This prediction is supported by EM data showing the existence of circular RNP complexes within arenavirions. For several arenaviruses, a non-templated G residue has been detected at the 5' end of progeny genomic RNAs during replication. The IGRs are predicted to fold into stable hairpin structures. Transcription of the

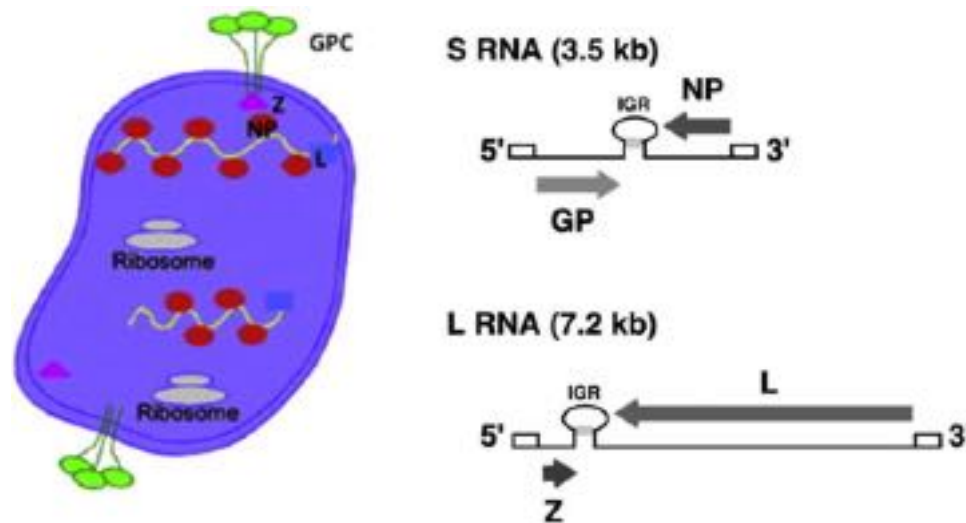


Fig. 1 A. Schematic illustration of an arenavirus particle. Shown is the pleomorphic shape of the enveloped virion particle decorated by spikes composed of heterotrimers of GP1 and GP2 associated with the SSP (not shown in the figure). The surface glycoproteins are aligned with subjacent matrix Z protein that serves also as a bridge between the surface glycoproteins and the S and L vRNP complexes inside the particle. Each vRNP consists of the NP (red) and L polymerase (blue) together with the S or L genome RNA species. Mammarenavirus virions are thought to incorporate ribosome particles. B. Mammarenavirus genome organization. Each genome segment uses an ambisense coding strategy to direct the synthesis of two different proteins in opposite direction and separated by a non-coding intergenic region (IGR) that serves as a bona fide transcription termination signal. The S segment encodes for the virus NP and GPC precursor, whereas the L segment encodes for the L polymerase and the matrix Z protein.

S-derived NP and GP mRNAs was shown to terminate at multiple sites within the predicted distal stem of the IGR, supporting the view that a structural motif rather than a sequence-specific signal promotes the release of the arenavirus polymerase from the template RNA. There are significant differences in sequence and predicted folded structure between the S and L IGRs, but among isolates and strains of the same arenavirus, the S and L IGR sequences are highly conserved. In addition to their role in control of transcription termination, IGRs have been shown to play a critical role in production of infectious particles.

GPC is co-translationally cleaved by the signal peptidase to generate a 58-amino acid stable signal peptide (SSP), and post-translationally processed by the cellular protease subtilisin kexin isozyme-1 (SKI-1)/ site 1 protease (S1P) to generate the mature virion surface glycoproteins GP1 and GP2. GP1, GP2 and SSP form the GPC complexes that decorate the surface of virions and mediate virus receptor recognition and cell entry. GP1 mediates virion interaction with host cell-surface receptors, whereas GP2 directs fusion of virus and host cell membranes. The fusion process depends on a low pH-driven conformational change of GP2 from a metastable prefusion structure to a more stable postfusion six-helix bundle. Notably, the SSP contributes to trafficking and processing of GPC as well as to the GP2-mediated pH-dependent fusion process.

NP is the most abundant viral polypeptide both in infected cells and virions. It is the main structural component of the viral RNP responsible for directing RNA genome replication and gene transcription. The C-terminus of NP exhibits a type I interferon (IFN-I) counteracting activity. Crystallographic studies identified distinct N- and C-terminal domains within NP of LASV, LCMV and JUNV, features that are expected to be present in other mammarenavirus NPs. The N-terminal domain of NP contains an RNA-binding site and plausible cap-binding activity. The C-terminal part of NP contains a functional DEDDH 3'-5' exoribonuclease folding domain similar to the one described for the non-structural 14 protein of severe acute respiratory syndrome (SARS) coronavirus. This 3'–5" exoribonuclease activity of NP plays critical roles in NP's anti-IFN activity and in other additional steps of mammarenavirus multiplication yet to be elucidated.

Mammarenavirus L proteins have a central region that includes conserved motifs characteristically found in the RdRp domains of negative-sense RNA viruses. Residues critical for LASV L function have been shown to be located both within and outside the predicted RdRp domain. Bioinformatic analysis and structural studies revealed that the N-terminus of LASV L has an endonuclease domain of similar structure to the cap-snatching endonuclease domains of influenza A virus polymerase acidic (PA) and La Crosse virus L proteins. In addition, EM characterization of a functional MACV L has revealed a core ring-domain decorated by appendages, which likely reflects a modular organization of the arenavirus polymerase.

Mammarenavirus Z proteins exhibit a modest degree of overall conservation. However, three regions within Z have a higher degree of conservation: (1) an N-terminal myristoylation site marked by a highly conserved G motif that is critical for membrane anchoring and interaction with other viral proteins; (2) a RING domain that binds two zinc ions through three conserved motifs; and (3) a C-terminal region containing proline-rich motifs that serve as functional late budding motifs. Structural studies revealed that the N- and the C-terminal arms flanking the RING domain are disordered, which may enable Z to recruit a variety of cellular partners to modulate virus multiplication. In addition, Z assembles in a dodecameric manner through a head-to-tail dimerization of the RING domain. These findings are consistent with Z playing a key role in virion assembly, which is further supported by EM

Mammarenavirus Life Cycle

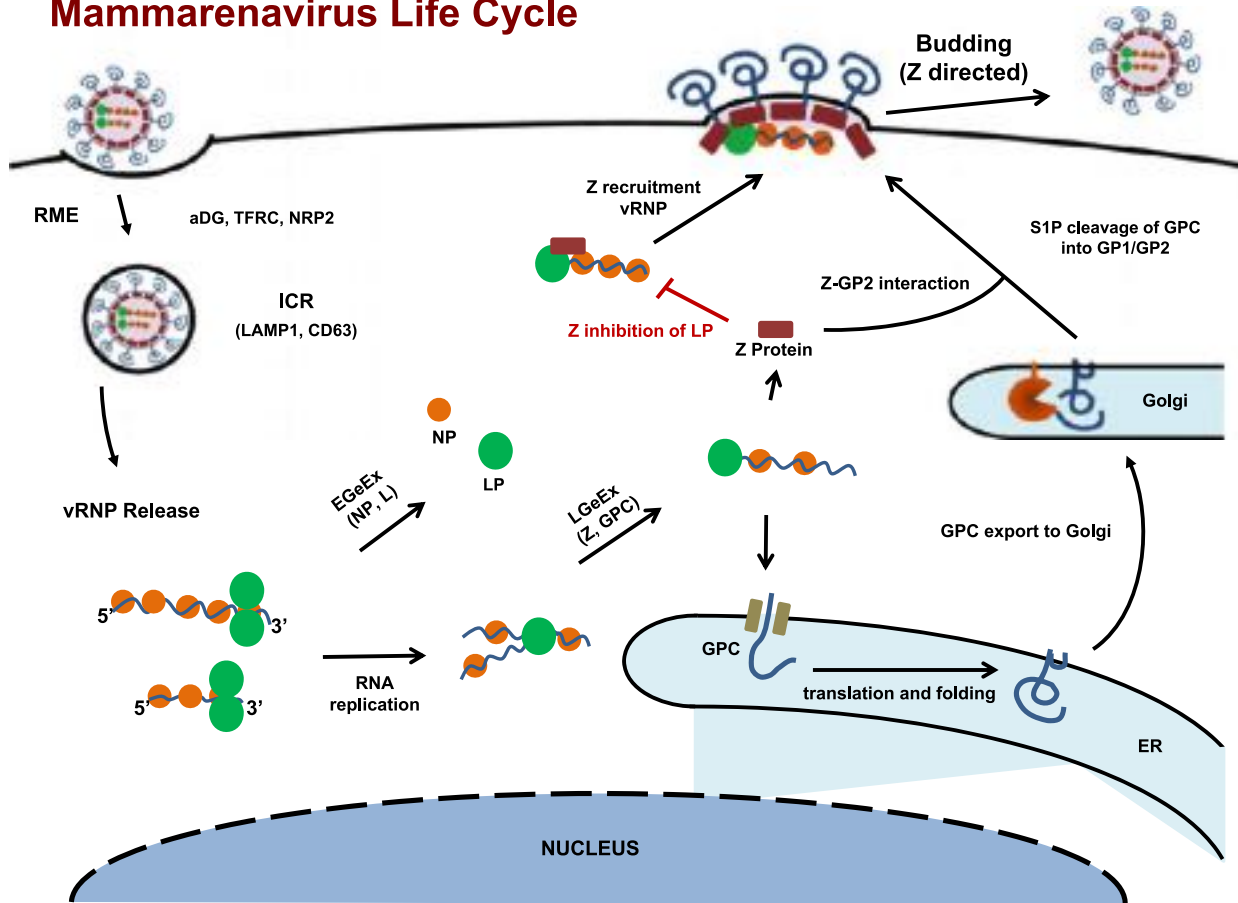


Fig. 2 Mammarenavirus life cycle. Mammarenavirus enter cells via receptor-mediated endocytosis (RME). There are some differences on the specific steps of the pathway used by different mammarenaviruses to reach the late endosome where the acidic environment triggers a pH-dependent fusion event between viral and cell membranes that releases the vRNP into the cytoplasm of the infected cells where viral RNA replication and gene transcription take place. Correct processing of GPC by the cellular signal peptidase and S1P protease is required for the formation of infectious progeny. The matrix Z protein plays critical roles mediating interactions between the vRNP and surface GP complex required for the formation of infectious progeny. Z is also the main driving force of mammarenavirus budding.

studies indicating that Z bridges the viral RNP to the viral envelope proteins. Z is also involved in suppression of the host cell type I IFN response via its interaction with RIG-I-like receptors.

Mammarenavirus Life Cycle

Cell Attachment and Entry

Receptor-mediated endocytosis is the main cell entry pathway used by mammarenaviruses. The acidic environment of the late endosome facilitates a pH-dependent conformational change in the GP complex and subsequent GP2-mediated fusion step between viral and cell membranes. Following fusion, the viral RNP is released into the cytoplasm where it directs both replication and transcription of the viral genome (Fig. 2). The conserved and widely expressed cell-surface receptor for extracellular matrix proteins α -dystroglycan (aDG) is a main receptor for the OW mammarenaviruses LCMV and LASV. Posttranslational glycosylation modification of aDG by the like-acetylglucosaminyltransferase (LARGE) is critical for aDG's function as a mammarenavirus receptor, and specific LARGE alleles have been shown to be positively selected among the Yoruba of Western Africa where LASV is endemic. However, non-pathogenic mammarenaviruses also use aDG-mediated cell entry, indicating that aDG does not play a unique direct role in mammarenavirus pathogenesis. Secondary alternative receptors, including members of the Tyro3/Axl/Mer and T-cell immunoglobulin (TIM) and mucin receptor families may account for LASV and LCMV infection of cells lacking fully glycosylated aDG. Cell entry of the OW hemorrhagic fever (HF) mammarenavirus LUV is mediated by neuropilin (NRP)-2, a cell-surface receptor for semaphorins. NRP-2 is highly expressed in microvascular endothelial cells, which may contribute to LUV-induced coagulopathy. Human transferrin receptor 1 (TfR1) is the main cellular receptor used for cell entry of pathogenic NW mammarenaviruses, including JUNV and MACV. Consistent with the use of TfR1 as a primary receptor, JUNV enters cells through clathrin-mediated endocytosis.

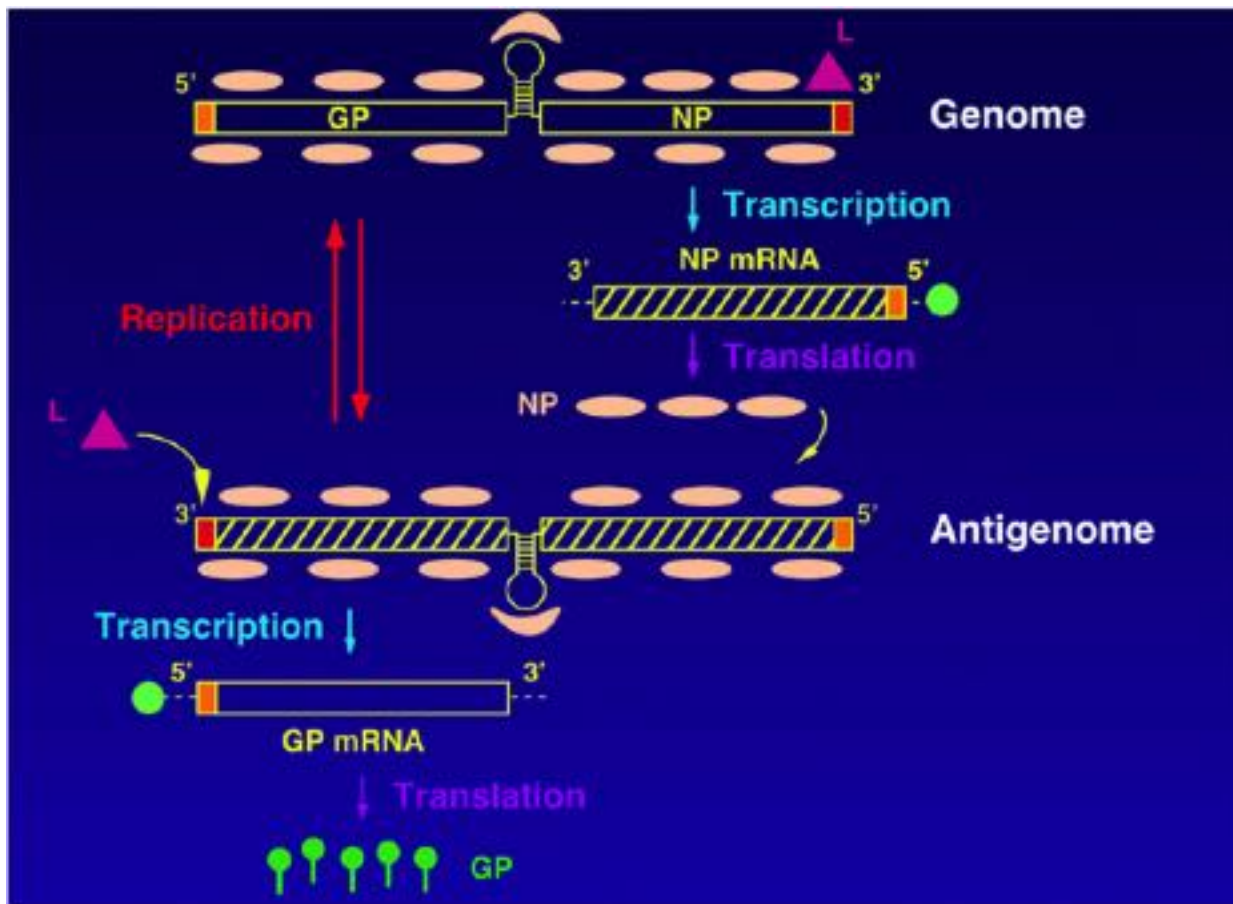


Fig. 3 Mammarenavirus RNA replication and gene transcription. The basic steps of mammarenavirus RNA replication and gene transcription are illustrated for the S segment. Following the pH-dependent fusion event between the viral and cell membranes in the late endosome, the vRNPs are delivered in to the cytoplasm where they initiate transcription from the genome promoter located at the genome 3'-end. Primary transcription results in synthesis of NP and L mRNA from the S and L segments, respectively. Subsequently, the virus polymerase adopts a replicase mode and moves across of the IGR to generate copies of the full length antigenome RNA (agRNA) species that serve as template for the synthesis of the GPC (agS) and Z (agL) mRNAs. The agRNA species serve also as templates for the amplification of the corresponding genome RNA species.

Notably, completion of the cell entry process for LASV and LUJV involves a late endosomal receptor switch mechanism. LASV uses the late endosomal resident proteins LAMP1 and LUJV uses CD36 in these final entry stages.

Expression and Replication of the Viral Genome

NP and L coding regions are transcribed into a genomic complementary mRNA. However, the GPC and Z coding regions are not translated directly from genomic RNA, but rather from genomic sense mRNAs (Fig. 3). These genomic sense mRNAs are transcribed from templates of the corresponding antigenome RNA, which also function as replicative intermediates. Transcription initiation is carried out through a mechanism shared by all negative sense viruses called cap-snatching, in which short fragments of 5'-capped host cell mRNAs are used by virus polymerase complex to prime the synthesis of viral mRNAs. Transcription termination occurs within the distal side of the IGR and generates non-polyadenylated viral mRNAs. Virus replication proceeds in two steps: first, the *de novo* (primer-independent) synthesis of an intermediate RNA with positive polarity (cRNA) and second, using this cRNA as template, the synthesis of progeny genomic vRNA with negative polarity. The 5'-end of arenavirus genome and antigenome RNAs each contain a non-templated G residue that has been proposed to reflect a prime-and-realign mechanism for RNA replication mediated by L.

Using cell-based minigenome (MG) assays, NP and L have been identified as the minimal viral *trans*-acting factors required for efficient RNA synthesis mediated by the virus polymerase. For LCMV, both genetic and biochemical evidence indicate that oligomerization of L is required for polymerase activity. Z is not required for RNA replication and gene transcription mediated by the virus L polymerase, but rather exhibits a dose-dependent inhibitory effect on both RNA biosynthetic processes. Consistent with these findings, studies using in vitro reconstitution of RNA synthesis directed by MACV polymerase have provided evidence that Z, via direct interaction with the polymerase, is able to lock the polymerase in a promoter-bound, catalytically inactive state.

Mutation-function analysis of the genome 5'- and 3'-termini using cell-based MG assays for LCMV, LASV, and MACV indicate that the activity of the arenavirus genomic promoter is dependent on both sequence specificity within the highly conserved 3'-terminal 19 nt of arenavirus genomes and on the integrity of the predicted panhandle structure. This panhandle is formed via sequence complementarity between the 5'- and 3'-termini of viral genome RNAs. Mammarenavirus RNA replication and gene transcription are regulated in a coordinated manner but intracellular levels of NP do not determine the balance between virus RNA replication and transcription. MG-based assays confirmed the IGR role as a *bona fide* transcription termination signal, but synthesis of translation-competent viral mRNAs does not strictly require the presence of the IGR.

Assembly and Budding

Production of infectious mammarenavirus progeny requires both Z and GPC and the correct processing of GPC into GP1 and GP2. Z is a structural component of the virion and cryo-electron microscopy structural studies reveal a location of LCMV Z within virions consistent with its role as a matrix protein. Consistent with its matrix protein function, Z is the main driving force of mammarenavirus budding, a process mediated by the interaction of Z's late domain budding motifs, PTAP or PPPY, or both, with components of the endosomal sorting complexes required for transport (ESCRT) within the vacuolar protein sorting (Vps) pathway. Myristoylation of Z is strictly required for its targeting to the plasma membrane, the location of arenavirus budding (Fig. 2).

Epidemiology of Human Pathogenic Mammarenaviruses

Typically, mammarenaviruses cause persistent, frequently asymptomatic infections in their reservoir hosts, which are characterized by virus multiplication in many different tissues and chronic viremia and viruria. Both, vertical transmission (exposure to infectious virus early in ontogeny) and horizontal transmission (exposure to virus later in life through aggressive or venereal behavior) may contribute to the mammarenavirus chronic carrier state in rodents. Human infections occur via exposure to rodent fomites, ingestion of contaminated food, exposure to broken skin or mucous membranes, or by inhalation of aerosolized virions from contaminated material.

Lassa Fever (LF)

LF is caused by LASV, which is endemic in vast areas of western sub-Saharan Africa. Imported cases of LF have been reported in the United States, Canada, the United Kingdom, Japan, Germany, Netherlands, and Israel. The case-fatality rate (CFR) of LF is about 1%–2% in the endemic areas, with estimated 300,000 infections annually. Most LASV infections in Africa are asymptomatic, mild or subclinical, but the CFR in hospitalized confirmed cases of LF can be as high as 69%. The disease is especially severe late in pregnancy with fetal death, miscarriage, or spontaneous abortion occurring in nearly all cases.

The main reservoir host of LASV is the Natal multimammate mouse (*Mastomys natalensis*). Maintenance of LASV in Natal *mastomys* is thought to occur via vertical transmission but recent studies suggest the existence of an additional horizontal transmission mechanism. Other rodents might also serve as LASV hosts or have roles in LASV transmission. Humans become infected with LASV through direct contact with infected rodent tissues, excreta, or blood or via inhalation of aerosolized virus. Infected meat could contribute to virus transmission in populations that include peridomestic rodents as part of their diet. Person-to-person transmission of LASV can occur in nosocomial settings via direct contact with body fluids from symptomatically infected individuals or corpses. Human infections tend to be more common in the dry season, which may reflect a higher prevalence of Natal *mastomys* within human dwellings. LASV prevalence is focal and varies greatly between geographical regions. For example, in Sierra Leone, seroprevalence ranges from 8% to 52% and in Nigeria from 13% to 37%.

LASV exhibits a high degree of genetic diversity; nucleotide differences can reach up to 32% for the L segment and 25% for the S segments. The genetic diversity among LASV strains correlates with geographic distribution rather than time. This high genetic diversity of LASV might explain the observed variability of LF's clinical presentation and possible regional differences associated with CFRs and disease symptoms.

Lujo Hemorrhagic Fever (LHF)

LHF, caused by LUJV, was identified in 2008 in South Africa. The only known nosocomial outbreak involved five cases of viral HF disease with clinical symptoms remarkably similar to LF, despite LASV and LUJV being distantly related. LHF disease has an abrupt onset and exhibits disseminated intravascular coagulation (DIC), features that distinguish LHF from LF. The overall lack of epidemiological data on the prevalence of LUJV in South Africa and the limited number of verified cases prevents any conclusion on whether the high (80%) CFR associated with the LHF outbreak reflects a more severe disease compared to LF. The natural reservoir of LUJV remains unknown.

Lymphocytic Choriomeningitis (LCM)

The primary host of LCMV is the house mouse. Other rodents, including pet hamsters and domesticated guinea pigs can become infected and transmit the virus to humans. LCMV is transferred vertically from one generation to the next and chronically infected

rodents shed the virus in their urine, saliva, nasal secretions, and droppings throughout life. Mounting evidence indicates that LCMV is a neglected human pathogen of clinical relevance as an under-recognized cause of neurologic disease in the fetus, and adult, and important threat to immune compromised individuals. Humans become infected with LCMV through close contact with infected rodents, solid organ transplantation, or by vertical transmission associated with congenital infections. The seroprevalence of LCMV within house mouse populations and in humans is highly variable, even within the same geographic region, resulting in the focal and uneven spatial distribution of the virus. As with LASV, LCMV exhibits high genetic diversity. Up to 18% and 25% nucleotide divergence was observed within the S and L segments, respectively, of LCMV lineages. The high genetic diversity of LCMV and the lack of clear correlation of virus genetic lineages to particular geographic locations likely reflect the long and complex phylogeographic history of the common house mouse host.

Argentinian Hemorrhagic Fever (AHF)

Junín virus (JUNV), the etiologic agent of AHF, is endemic to the Pampas in Argentina. In the absence of treatment, the CFR rate can reach 20%–30%. Pregnant women have higher CFR and many miscarry, especially if infection occurs during the third trimester. AHF is typically a seasonal disease, with a peak of frequency occurring during the corn-harvesting season, and infected cases are primarily rural male agricultural workers. The drylands laucha (*Calomys musculinus*) is the main reservoir of JUNV but other animals can also become infected. The patchy spatial distribution of the drylands laucha has been suggested to account for the focal distribution of AHF. Annual increase in the number of these rodents coincides with the corn harvesting season providing opportunities for rodent-to-man transmission. Human transmission is thought to occur predominantly by inhaling aerosolized viral particles from contaminated soil and plant litter, which are disturbed during the mechanized harvesting process, or by exposure to primary aerosols of rodent urine or contact with contaminated nesting materials in border habitats. Horizontal transmission via aggressive encounters among adult, male rodents is the primary mode of viral persistence in nature, which is facilitated by the high viral load in saliva of drylands lauchas. Vertical transmission might contribute to the maintenance of JUNV via intra-generational infection by horizontal transmission when population numbers are reduced. Seroprevalence in humans range from 4.7% to 12.3% in AHF-endemic areas, and significantly lower (0.44%) in non-endemic areas. In rodents, prevalence of infection can be as high as 10.9% with highest prevalence observed in current epidemic areas and lowest (0.2%) in non-endemic areas. In contrast to LASV and LCMV, Nucleotide similarity between JUNV strains is high, reaching up to 94.5% and 95.4% for GP and NP, respectively.

Bolivian Hemorrhagic Fever (BHF)

The CFR of BHF, caused by Machupo virus (MACV), is approximately 5%–30% with the highest rates occurring among those under 5 and over 55 years of age. BHF cases are more common at the peak of agricultural activity, during the dry season, with males over 15 years of age more frequently affected. The big laucha (*Calomys callosus*) is the reservoir host of MACV. Both horizontal and vertical transmission have been shown as possible maintenance mechanism of MACV within its reservoir rodent. Person-to-person transmission of MACV is possible but rather uncommon, likely reflecting the infrequent and low virus detection in blood or from throat and oral swabs of infected patients.

Venezuelan Hemorrhagic Fever (VeHF)

Guanarito virus or GTOV emerged in 1989 as the cause of VeHF, a severe hemorrhagic illness with CFR of $\approx 30\%$. The disease has focal distribution in the southern and southwestern portions of Portuguesa state and in adjacent areas of Barinas state. VeHF cases peak during the period of agricultural activity in these regions and involve mainly male agricultural workers. The short-tailed zygodont (*Zygodontomys brevicauda*) is the main reservoir host of GTOV. The virus is also commonly found in Alston's cotton rats (*Sigmodon alstoni*). These hosts have not been reported within houses or farm building. Therefore, infections are assumed to occur outdoors, in rural areas, and persons that have frequent contact with rodent-infested grassland habitats are at higher risk.

Pathogenesis and Pathology

The mechanisms underlying pathogenesis of mammarenaviral disease in humans are not well understood. Pathological findings in patients' autopsies do not account for the severity of symptoms seen in many cases of human mammarenaviral disease. Mammarenaviruses typically enter humans in an aerosolized form and are deposited in the lung, where initial viral replication occurs. Antigen-presenting cells (APCs), DCs and macrophages, are prominent targets in the initial stages of infection, and facilitate virus access to the lymphoid system and subsequent systematic spread to other organs and tissues including liver, kidneys, lungs, adrenal glands, and heart. LASV has also been recovered from placenta, mammary glands, and aborted fetal tissues, and in cases of AHF, virion-like particles were detected in the central nervous system (CNS), ovaries, and testes. LCMV load can reach high levels in meninges, choroid plexus, and ventricular ependymal linings, where the inflammatory response produces the characteristic LCM pathology. LCMV exhibits a strong tropism for the fetal brain, where LCMV congenital infection produces its

most common and severe pathologic effects, including microencephaly, periventricular calcifications, hydrocephalus, cerebellar hypoplasia, focal cerebral destruction, and gyral dysplasia.

The extent of hemorrhagic and coagulopathy manifestations differs between human pathogenic mammarenaviruses. Hemorrhages are common in AHF cases, but are rare and mainly limited to mucosal surfaces in LF cases. Bleeding is usually associated with thrombocytopenia and platelet dysfunction. Fatal cases of LF are associated with lower levels of platelet activating factor (PAF) and PAF-like molecules, as well as hemoglobin breakdown products D-urobilinogen and I-urobilin. Disseminated intravascular coagulation (DIC) or complement activation do not appear to play a role in mammarenavirus pathogenesis.

Impaired vascular endothelium function, including increased permeability, likely plays a central role in LF and AHF pathogenesis. However, only minimal vascular histological lesions are detected in fatal human LF and AHF cases and infected non-human primates (NHPs), and the mechanisms responsible for virus induced increase in vascular permeability remain to be elucidated. In contrast to other HF-causing viruses, LASV and JUNV do not trigger a “cytokine storm” that could interfere with the integrity of the vascular endothelium, but infection of the endothelium might cause changes in endothelial cellular function leading to increased fluid flow and subsequent edema.

Pathological findings in LF patients include both macroscopic and microscopic abnormalities. Gastric mucosal, renal, and subconjunctival hemorrhages, petechiae, and increased vascular permeability are common macroscopic observations, whereas microscopic observations include multifocal hepatocellular necrosis with modest inflammatory cell involvement, splenic necrosis, necrosis of renal tubular cells, adrenocortical cell necrosis, focal renal interstitial lymphocytic infiltrates, mild mononuclear interstitial myocarditis, alveolar edema with capillary congestion, interstitial pneumonitis, and rhabdomyositis.

The most consistent pathological hallmark of LF in humans is multifocal hepatocellular necrosis. High virus titers in liver tissue correlate with severe hepatitis. However, the degree of hepatic tissue damage is insufficient to cause hepatic failure, and there is no correlation between the degree of hepatic necrosis and chemical indicators of liver damage, indicating that LASV-induced hepatitis is unlikely to be the primary cause of death in fatal cases of LF.

The most common macroscopic abnormality observed in severe cases of NW mammarenaviral disease is widespread hemorrhage affecting the skin and mucous membranes, Virchow-Robin space, kidneys, pericardium, spleen, adrenal glands, and lungs. Microscopic lesions include acidophilic bodies, focal liver necrosis, acute tubular and papillary necrosis in the kidneys, and reticular hyperplasia of the spleen and lymph nodes. Pathognomonic symptoms of AHF illness can vary depending on the specific JUNV strain and can be “hemorrhagic”, “neurologic”, “mixed”, and “common”. Infection with “hemorrhagic” strains results in a pronounced bleeding tendency with disseminated cutaneous and mucous membrane hemorrhage. In contrast, infection with “neurologic” strains show little or no hemorrhagic manifestations but result in overt and progressive signs of neurologic dysfunction.

LCM disease correlates with detection of viral antigen in meninges and cortical neurons, as well as mononuclear cell infiltrates in the meninges and around vessels and glial nodules in the deeper structures, suggesting an immune-mediated pathology associated with LCMV infection. An animal model of congenital LCMV infection revealed a very strong tropism of LCMV for neuroblasts and altered neuron migration, findings that may explain the location of periventricular calcifications and the gyral malformations in children with congenital LCMV.

Clinical Features

LF and LHF

LF is mild or asymptomatic in about 80% of infected individuals, but 20% develop acute LF. The incubation period can range from 5 to 21 days but is typically 10–14 days. Within 2–4 days of infection, many patients experience an array of symptoms including headache, myalgia, arthralgia, lower back, abdominal and retrosternal chest pain, dizziness, nausea, tinnitus and sore throat. Cough, vomiting, diarrhea and constipation are also common, whereas lymphadenopathy, oliguria, tachycardia, vertigo, splenomegaly, hepatomegaly, and jaundice have been reported occasionally. Disease progression is often associated with pharyngitis, conjunctivitis, respiratory distress, pleural and pericardial effusions, and facial and neck edema. In most cases of LF recovery begins 8–10 days after disease onset, and 4–11 days later viremia is undetectable. In contrast, in severe cases of LF, worsening of symptoms and hemorrhagic and neurologic manifestations are more common. Hemorrhagic manifestations can affect skin and mucosal surfaces. Neuroglial findings including diffuse encephalopathy, confusion, tremors, coma, and convulsions are common prior to shock and death. Widespread edema can be observed in children with LF, which usually has a fatal outcome.

Common clinical findings in LF cases include proteinuria, albuminuria, and elevated AST levels, whereas moderate leukopenia has been reported only in some patients. Survivors of LF often recover without sequelae. However, long-term unilateral or bilateral sensorineural deafness can affect a significant ($\approx 13\%$ – 30%) number of survivors.

Known cases of LHF disease presented with similar clinical symptoms to those documented for LF cases. Initial symptoms of nonspecific febrile illness increase in severity over 7 days with the development of diarrhea, pharyngitis, and facial and neck edema. Terminal features are shock and multi-organ system failure often with evidence of DIC that resulted in death of four of the five LHF cases.

LCM

In most healthy individuals, LCMV infection course is asymptomatic, or manifested as a mild febrile illness. Following 1–2 weeks of incubation, some infected individuals may develop a flu-like illness that could include anorexia and gastrointestinal

symptoms. Sore throat, cough, pharyngitis, and other symptoms of respiratory tract involvement are less common. Leukopenia, moderate thrombocytopenia, mild elevations of AST, and infiltrates on chest radiographs can be seen in some patients. Additional, but rather infrequent, reported clinical findings are lymphadenopathy, dysesthesia, conjunctivitis, arthralgia, arthritis, rash, testicular or parotid pain, as well as abdominal, back, and chest pain. LCMV invasion of the CNS can occur in some infected individuals, usually resulting in symptoms of classic aseptic meningitis. In some cases, encephalitis, encephalomyelitis, meningoencephalitis, acute hydrocephalus, ascending or bulbar paralysis, or transverse myelitis may develop. CSF pleocytosis is characteristically observed during this CNS phase of illness and LCMV can be isolated from the spinal fluid. Infections are rarely fatal, and most patients recover without sequelae, but some might experience asthenia, headaches, confusion and difficulty in concentration.

LCMV Infection during pregnancy, especially during the first trimester, increases the risk for miscarriage, *in utero* fetal death, fetopathy, and severe neurologic sequelae. Congenital LCMV can result in severe CNS or ocular malformations and can mimic the signs/symptoms of classic TORCH syndrome. In immunosuppressed organ recipients, LCMV infection can be fatal due to multi-organ system failure with LCMV-associated hepatitis as a prominent feature. In most immunosuppressed patients CNS manifestations are also observed. Clinical findings in these patients include elevated transaminases levels, leukopenia, and thrombocytopenia, as well as increased protein levels and leukocyte numbers in CSF.

NW Mammarenaviral HFs

AHF has an incubation period of 1–2 weeks, followed by a prodromal phase characterized by fever and malaise, headache, myalgia, epigastric pain, and anorexia. After 2–4 days, signs become increasingly severe including prostration and GI disturbances. In some cases, dizziness, photophobia, retro-orbital pain, or disorientation may also occur. Initial signs of vascular damage may appear at this early phase of AHF. During the second week of illness, about 20%–30% of patients develop severe hemorrhagic and neurologic manifestations, or secondary bacterial infections. Hemorrhagic manifestations include bleeding from mucous membranes and ecchymosis at needle puncture sites, but only very modest blood loss is observed. Neurologic manifestations include seizures, convulsions, tremor of the hands and tongue, and less frequently, delirium, coma, encephalitis and meningoencephalitis. Following shock, death usually occurs 7–12 days after disease onset. Clinical laboratory findings include leukopenia and thrombocytopenia and non-specific electrocardiogram abnormalities, whereas chest radiography is usually normal in the absence of secondary infections. Patients' clinical improvement during the second week of AHF correlates with the appearance of neutralizing antibodies. Convalescence often lasts several weeks with polyuria, fatigue, alopecia, and dizziness.

BHF and VeHF caused by MACV and GTOV, respectively, as well as NW mammarenaviral HF caused by CHAPV and SBAV, display clinical symptoms similar to those described for AHF. Proteinuria and elevated hematocrit during the peak of hemorrhagic manifestations are characteristic of severe cases of BHF. Early symptoms of VeHF disease are indistinguishable from dengue fever, also common in Venezuela. CNS manifestations of VeHF including encephalitis are associated with poor prognosis. Hearing loss has been reported in convalescence cases of VeHF. The only reported case of a naturally acquired infection with SBAV caused disease with symptoms similar to those of other NW mammarenaviral HF, but also included extensive liver necrosis.

Diagnosis

LF

LASV can be isolated from blood during the febrile phase of LF disease, as well as from autopsy tissue samples. Serum detection of LASV antigen by ELISA is robust, reliable, and can be completed in a short time. LASV-specific antibodies can be detected by immunofluorescence (IF) and ELISA. ELISA IgM titers appear earlier and persist longer than IF IgM titers. Virus-specific IgG ELISA antibody detection persists for long periods, whereas IF antibody appears to wane over time below detectable limits. Reverse transcriptase-PCR (RT-PCR) can detect virus RNA in blood with high sensitivity. Nanopore sequencing using the MiniON device was successfully used to genetically characterize LASV isolates during the 2018 upsurge of LF cases in Nigeria. The ability of this portable sequencing technology to genetically characterize *in situ* RNA viral samples in real-time cases of LASV infection associated with disease symptoms will represent a major breakthrough in the study of LF epidemiology.

LF presents with symptoms indistinguishable from those of other febrile illnesses including malaria, and therefore it is difficult to diagnose LF clinically, but it should be suspected in patients with fever ($\geq 38^{\circ}\text{C}$) not responding to antimalarial and antibiotic drugs. Useful clinical findings for diagnosis of LF are fever, pharyngitis, retrosternal pain, and proteinuria. Viremia levels, fever, sore throat, vomiting, edema, and bleeding are the predictors of poor prognosis. CNS signs, face and neck edema, jaundice, bleeding, hematuria, and proteinuria were shown to be associated with a fatal outcome. Several biomarkers including increased levels of blood urea nitrogen and creatinine (biomarkers for kidney function), serum AST (liver malfunction), and serum electrolyte potassium, have been shown to have a strong correlation with CFR, highlighting the role of liver and renal dysfunction and electrolyte disturbance in the severity of LF.

LCM

LCMV can be isolated from blood during the febrile phase of LCM and during meningitis symptoms, but CSF contains higher viral load, and PCR-based tests using CSF samples have been successfully used. LCMV-specific IgM antibodies can be detected by ELISA and IF in serum and CSF in acute cases of LCM. Neutralizing antibodies appear late after onset of disease and their diagnostic value is limited.

NW Mammarenaviral HF

During the acute febrile phase of disease, virus can be isolated from blood samples by inoculation of newborn hamsters or mice, but cocultivation of patient's peripheral blood mononuclear cells with Vero cells may offer higher level of sensitivity. Virus can also be isolated from autopsy tissues, with the exception of brain. Viral antigen in blood and tissues from patients with JUNV, MACV, SBAV, or GTOV infection can be usually detected by antigen-capture ELISA. Serologic diagnosis of AHF and BHF is usually made by IF and CF, but the limited specificity and sensitivity of these tests pose problems. The ELISA test is the most useful and practical for rapid detection of IgM and IgG antibodies in a clinical setting and sero-epidemiologic surveys. The plaque neutralization test can be valuable for evaluation of convalescent plasma units intended for therapeutic use in AHF.

Prevention and Control

Medical Management

Supportive therapy including hydration and pain relief management represent standard of care in suspected cases of mammarenaviral HF. Platelet transfusions and factor replacement could help to control bleeding. Infusion of crystalloids to counteract the commonly observed moderate degree of vascular permeability should be carefully considered due to a high risk of pulmonary edema. The low cardiac output observed in human cases of AHF support the use of Swan-Ganz catheterization. LCMV ependymal infection and inflammation may cause acute hydrocephalus and a need for surgical shunting. Management of patients that undergo shock is difficult.

Patients with mammarenaviral HF pose in general a low risk of contagion. However, nosocomial outbreaks and infection of multiple contacts have occurred when the index case was severely ill with high viremia and viral load in tissues. Patients may excrete virus in urine or semen for weeks after recovery from disease. Therefore, body fluids should be monitored for infectivity before the patient is released, and counseling should be provided emphasizing protection of sexual partners and the use of disinfectant prior to use of toilets. The highest exposure risk is parenteral and can be minimized through staff training. Respiratory protection against small-particle aerosols should be implemented with caregivers and people in close proximity to patients. Special precautions are indicated when blood and other body fluids are handled in the clinical laboratory.

Antiviral Drugs

The prophylactic and therapeutic value of the nucleoside analog ribavirin (Rib) against human pathogenic mammarenaviruses is well supported by results from both cell culture and animal models of infection. Importantly, Rib reduced both morbidity and mortality in human cases of LF, and evidence suggests that Rib treatment can have also clinical benefits in cases of JUNV, MACV and SABV infections. However, the need of intravenous administration for optimal efficacy, poor penetration into the CSF, and side effects, including anemia and congenital disorders, pose some limitations to the use of Rib. In addition to Rib, several compounds have been reported to have anti-mammarenaviral activity in cultured cells, but their safety and efficacy *in vivo* remain to be determined. The broad-spectrum antiviral favipiravir and the mammarenaviral GP-mediated fusion inhibitor ST-193 have yielded very promising results in different animal models of mammarenaviral disease.

Progress in mammarenavirus molecular genetics has facilitated the development of screens to identify drugs targeting specific steps of the arenavirus lifecycle. These drugs include direct-acting antivirals (DAAs) that target specific viral gene products and functions, and drugs that target host-cell factors (HTAs) required for the completion of virus replication cycle. Combination therapy of DAAs and HTAs should counteract the emergence of drug resistant variants often observed with monotherapy strategies. Moreover, related viruses are likely to rely on the same host machinery, thus providing an opportunity for the development of broad-spectrum antiviral therapeutics.

Limited market opportunities are an obstacle for the development and licensing of new antiviral drugs against human pathogenic mammarenaviruses. It would be therefore beneficial to use the newly developed screening platforms to implement drug repurposing strategies that can significantly reduce the time and resources required to advance a candidate antiviral drug into the clinic. Moreover, drug repurposing can generate new knowledge on virus biology by uncovering previously unexplored pathways and specific host cell factors contributing to virus multiplication, knowledge that can be harnessed to identify new targets and therapeutics.

Antibody Therapy

Treatment of AHF cases with convalescent plasma from survivors of AHF has been shown to be quite effective in reducing AHF-associated mortality. About 10% of patients treated with convalescent plasma may exhibit usually self-limited neurologic signs including fever,

headache, cerebellar tremor, and cranial nerve palsies, which may reflect the prior invasion of the CNS by the virus. Studies in animal models of infection suggest that passive antibody therapy could be also helpful to treat BHF, VeHF and LF. Experimental studies of passive protection by monoclonal antibodies have been successful in LCMV. Notably, LASV GPC-specific recombinant human monoclonal antibodies derived from survivors of LF with strong neutralization profiles in cell-based assays have shown robust therapeutic efficacy in guinea pig and non-human primate (NHP) models of LF, supporting the feasibility of antibody-based therapy against LF.

Vaccines

The safety, immunogenicity, and protective efficacy of the JUNV live-attenuated Candid #1 strain was demonstrated in preclinical studies in both guinea pig and NHP models of AHF. Vaccination campaigns focused on agricultural workers in the JUNV-endemic area showed Candid #1 to be an effective and safe vaccine in humans, and it was licensed in 2006 for use exclusively in Argentina. However, comparison of Candid#1 clonal isolates from blood of vaccinated NHPs revealed a 1000-fold range of virulence among them, raising some concerns about the stability of the Candid #1 vaccine.

LASV is the mammarenavirus that poses the highest concern to human health due to morbidity and lethality associated with LF together with the vast LASV-endemic regions and size of the population at risk in Western Africa. Accordingly, LF has been included on the revised list of priority diseases of the WHO R&D Blueprint for which the development of a vaccine is encouraged. Studies involving LF survivors and animal models of LASV infection indicate that early and robust virus-specific CD4⁺ and CD8⁺ T cell responses, rather than the presence of neutralizing antibodies, which appear late after infection and at low titers, are the best correlates of recovery and protection. Although antibodies do not appear to contribute to viral control and recovery during acute LASV infection, genetically engineered neutralizing monoclonal antibodies can be successfully used for immunotherapy. These data suggest that an ideal LASV vaccine should trigger both long-term robust cell-mediated immunity (CMI) and humoral responses following a single immunization, features frequently displayed by live-attenuated vaccines (LAV). A number of LASV vaccine platforms have shown promising result in animal models of LASV, including LAV candidates based on vaccinia virus, vesicular stomatitis Indiana virus (VSIV), Mopeia virus (MOPV) an OW arenavirus closely related to LASV, yellow fever virus 17D vaccine strain, measles virus (MV)-based vector, alphavirus replicons and the ML29 reassortant carrying the L segment of the non-pathogenic MOPV and the S segment of LASV.

The Coalition for Epidemic Preparedness Innovations (CEPI) was created as non-profit organization to accelerate vaccine development against emerging epidemic infections for which the commercial market is insufficient to justify private investment. LASV vaccine candidates supported by CEPI include recombinant VSV and MV expressing LASV GPC and DNA-based vaccines. A single injection of rVSV/LASV-GPC experimental vaccine, in which VSIV G was replaced with LASV GPC, fully protected guinea pigs and NHPs against LASV strains from the same clade. Despite these promising results, there are still some safety concerns because of reported side effects in individuals vaccinated with an Ebola virus vaccine based on the same platform (rVSVΔG/EBOV-GP), which included post-vaccination arthritis, vector RNAemia, and detection of infectious vaccine virus in the skin of vaccinated individuals. The MV-based vector vaccine platform has shown promising results in human clinical trials providing antibody-based protection. Whether this platform would be effective for development of a vaccine against LASV with predominant T cell-mediated mechanism of protection remains to be determined. Currently there is no human licensed preventive DNA vaccine for any infectious disease, which poses great technical obstacles for developing an DNA-based vaccine against LASV.

The LASV candidate vaccine reassortant ML29 induced a robust cross-reacting and protective, sterilizing cell-mediated immune response against strains from distantly-related LASV lineages in a guinea pig model of LF. In addition, ML29 was shown to be stable and safe in animal models of LF, including immunocompromised NHPs. This finding is relevant because the LASV vaccine target population will likely include individuals with some degree of undiagnosed immune suppression due to the high prevalence of malaria and HIV-1 infections in Western Africa. Progress in arenavirus molecular genetics can also facilitate the implementation of novel strategies for the development of safe and effective LF vaccines based on the use of codon deoptimization and reorganization of the coding and non-coding intergenic regions.

Perspective

Since the discovery of LCMV more than 70 years ago, mammarenaviruses have served as important model systems for the study of host-virus interactions. Specifically, studies with LCMV in mice, the virus's natural reservoir, have uncovered a wide range of principles in the fields of virology and immunology that apply universally to other viral infections, including virus-induced immunopathology, MHC restriction, the contribution of negative immune regulators to viral persistence, and the concept of T cell exhaustion and demonstration that PD-1/PD-L1 blockade can rescue T cell function. In addition to their impact on basic science, mammarenaviruses include several human pathogens that remain a significant public health risk in much of the world. Control of human pathogenic mammarenaviruses faces significant obstacles including current limited diagnostic protocols and sparse resource allocation for vaccine development and other medical countermeasures. However, progress in molecular genetics has facilitated novel approaches for the investigation of mammarenavirus molecular and cell biology, whereas the application of next generation sequencing technologies has brought a major breakthrough to the investigation of mammarenavirus epidemiology including their prevalence and distribution, as well as genetic diversity. This new knowledge has had also great implications for the development of new approaches to antivirals and vaccines to combat human pathogenic mammarenaviruses.

Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the US Department of the Army, the US Department of Defense, the US Department of Health and Human Services, or of the institutions and companies affiliated with the authors.

Further Reading

- Andersen, K.G., Shapiro, B.J., Matranga, C.B., *et al.*, 2015. Clinical sequencing uncovers origins and evolution of Lassa virus. *Cell* 162 (4), 738–750.
- Basler, C.F., 2017. Molecular pathogenesis of viral hemorrhagic fever. *Seminars in Immunopathology* 39 (5), 551–561.
- Bederka, L.H., Bonhomme, C.J., Ling, E.L., Buchmeier, M.J., 2014. Arenavirus stable signal peptide is the keystone subunit for glycoprotein complex organization. *mBio* 5 (6), e02063.
- Boisen, M.L., Hartnett, J.N., Shaffer, J.G., *et al.*, 2018. Field validation of recombinant antigen immunoassays for diagnosis of Lassa fever. *Scientific Reports* 8 (1), 5939.
- Buchmeier, M.J., Peters, C.J., de la Torre, J.C., 2007. Arenaviridae: The virus and their replication. In: Fields, B.N., Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, fifth ed., vol. 2. Philadelphia, USA: Wolters Kluwer Health/Lippincott Williams & Wilkins, pp. 1792–1827.
- Golden, J.W., Hammerbeck, C.D., Mucker, E.M., Brocato, R.L., 2015. Animal models for the study of rodent-borne hemorrhagic fever viruses: Arenaviruses and hantaviruses. *BioMed Research International* 2015, 793257.
- Hastie, K.M., Kimberlin, C.R., Zandonatti, M.A., MacRae, I.J., Saphire, E.O., 2011a. Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3' to 5' exonuclease activity essential for immune suppression. *Proceedings of the National Academy of Sciences of the United States of America* 108, 2396–2401.
- Hastie, K.M., Zandonatti, M.A., Kleinfelter, L.M., *et al.*, 2017. Structural basis for antibody-mediated neutralization of Lassa virus. *Science* 356, 923–928.
- Jae, L.T., Raaben, M., Herbert, A.S., *et al.*, 2014. Virus entry. Lassa virus entry requires a trigger-induced receptor switch. *Science* 344, 1506–1510.
- Kranzusch, P.J., Schenk, A.D., Rahmeh, A.A., *et al.*, 2010. Assembly of a functional Machupo virus polymerase complex. *Proceedings of the National Academy of Sciences of the United States of America* 107, 20069–20074.
- Lukashevich, I.S., 2013. The search for animal models for Lassa fever vaccine development. *Expert Review of Vaccines* 12 (1), 71–86.
- Raaben, M., Jae, L.T., Herbert, A.S., *et al.*, 2017. NRP2 and CD63 are host factors for Lujo virus cell entry. *Cell Host Microbe* 22, 688–696.
- Radoshitzky, S.R., Buchmeier, M.J., Charrel, R.N., *et al.*, 2019. ICTV virus taxonomy profile: Arenaviridae. *Journal of General Virology* 100 (8), 1200–1201. doi:10.1099/jgv.0.001280.
- Sarute, N., Ross, S.R., 2017. New world arenavirus biology. *Annual Review of Virology* 4 (1), 141–158.
- Vela, E., 2012. Animal models, prophylaxis, and therapeutics for arenavirus infections. *Viruses* 4, 1802–1829.
- Watanabe, Y., Raghvani, J., Allen, J.D., *et al.*, 2018. Structure of the Lassa virus glycan shield provides a model for immunological resistance. *Proceedings of the National Academy of Sciences of the United States of America* 115, 7320–7325.

Human Polyomaviruses (*Papillomaviridae*)

Melissa S Maginnis, The University of Maine, Orono, ME, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of M. Safak, Polyomaviruses of Humans, In Reference Module in Biomedical Sciences, Elsevier Inc., 2018, doi:[10.1016/B978-0-12-801238-3.02646-5](https://doi.org/10.1016/B978-0-12-801238-3.02646-5).

Nomenclature

BKPyV	BK polyomavirus	NCCR	Non-coding control region
HIV	Human immunodeficiency virus	NJPyV	New Jersey polyomavirus
HPyV	Human polyomavirus	PML	Progressive multifocal leukoencephalopathy
JCPyV	JC polyomavirus	PyV	Polyomavirus
KIPyV	KI polyomavirus	PyVAN	Polyomavirus-associated nephropathy
LIPyV	Lyon IARC polyomavirus	STAg	Small T (tumor) antigen
LTAg	Large T (tumor) antigen	STLPyV	St. Louis polyomavirus
MCC	Merkel cell carcinoma	TSPyV	Trichodysplasia-spinulosa polyomavirus
MCPyV	Merkel cell polyomavirus	VP	Viral protein
MWPyV	Malawi polyomavirus	WHO	World Health Organization
		WUPyV	Washington University polyomavirus

Glossary

IHC Immunohistochemistry, an experimental method used in diagnosis to test for the presence of antigens, such as viral antigens, in a tissue section through the use of antigen-reactive antibodies.

Receptor A molecule that is expressed in the cell membrane and serves as an attachment site for a ligand such as a virus.

Seroprevalence The frequency of positive occurrence of a particular disease or condition (viral infection) within a population as measured by serological (blood) tests.

Tissue tropism Specific cells and tissues that are able to support the replication of a particular virus.

Introduction

Human polyomaviruses (HPyVs) commonly infect the majority of the population, and serological studies suggest that transmission occurs during childhood but seroprevalence continues to increase into adulthood. Transmission of HPyVs is thought to occur via direct person-to-person contact, contact with fomites, or through the ingestion of contaminated food or water. HPyVs demonstrate a high degree of species-specificity, and thus tractable animal models for studies of HPyVs are lacking, and the majority of the clinical data is from observations of human subjects. In addition, this group of viruses exhibits a narrow tissue tropism, generally targeting specific tissues within the host. HPyVs establish life-long persistent infections in target host cells resulting in asymptomatic infections. However, under conditions of immunosuppression such as HIV infection or use of immunosuppressive agents for immune-mediated diseases, HPyVs can become reactivated and cause a range of serious clinical diseases.

BK and JC polyomaviruses were first identified in the 1970s after isolation from patients with nephropathy and progressive multifocal leukoencephalopathy (PML), respectively, and are named after the patients. Another human polyomavirus was not identified until the 2000s, and since that time, there have been 11 newly-identified human polyomaviruses isolated from various tissues and fluids including respiratory samples, urine, stool, blood, skin, and liver ([Table 1](#)). The increase in HPyV isolation and identification was ignited by improved technologies such as rolling circle amplification (RCA) and high-throughput viral sequencing. Karolinska Institute polyomavirus (KIPyV) and Washington University polyomavirus (WUPyV) were both isolated from nasopharyngeal tissue and named for the research institute credited with the identification. Merkel cell polyomavirus (MCPyV) was isolated from individuals with Merkel cell carcinoma (MCC) and was identified as a causative agent of MCC. Shortly thereafter, HPyV6 and HPyV7 were identified from normal skin tissues, although they are now linked to pathologies of the skin including dermatoses. The eighth polyomavirus identified was Trichodysplasia Spinulosa-associated polyomavirus (TSPyV) named for the TS skin lesions from which it was isolated. HPyV9 was isolated from blood, skin, and urine of a kidney transplant recipient, and HPyV10 or Malawi (MWPyV) was isolated from stool samples and from condylomas from a patient with warts, hypogammaglobulinaemia, infections, and myelokathexis (WHIM) syndrome. St. Louis polyomavirus (STLPyV)/(HPyV11) was identified in stool samples from children in Malawi, The Gambia and St. Louis. HPyV12 was isolated from human liver tissue and stool, and New Jersey polyomavirus (NJPyV)/(HPyV13) was isolated from a muscle biopsy and skin lesions in a pancreas transplant recipient with vasculitis and necrotic plaques. The most recently identified polyomavirus, Lyon IARC polyomavirus

Table 1 Human polyomaviruses and associated diseases

Number Designation	Polyomavirus	Isolation source	Year of Disease association ID
HPyV1	BK polyomavirus (BKPyV)	Urine	1971 Polyomavirus-associated nephropathy, hemorrhagic cystitis
HPyV2	JC polyomavirus (JCPyV)	Brain	1971 Progressive multifocal leukoencephalopathy
HPyV3	Karolinska Institute polyomavirus (KIPyV)	Nasopharyngeal tissue	2007 <i>Potential mild respiratory illness</i>
HPyV4	Washington University polyomavirus (WUPyV)	Nasopharyngeal tissue	2007 <i>Potential mild respiratory illness</i>
HPyV5	Merkel cell polyomavirus (MCPyV)	Tumors, skin	2008 Merkel cell carcinoma
HPyV6	Human polyomavirus 6	Skin	2010 Pruritic and dyskeratotic dermatitis
HPyV7	Human polyomavirus 7	Skin	2010 Pruritic and dyskeratotic dermatitis
HPyV8	Trichodysplasia spinulosa-associated polyomavirus (TSPyV)	Skin lesions	2010 Trichodysplasia spinulosa
HPyV9	Human polyomavirus 9	Blood, skin, urine	2011 ?
HPyV10	Malawi polyomavirus (MWPyV)	Stool, condylomas	2012 ?
HPyV11	St. Louis polyomavirus (STLPyV)	Stool	2012 ?
HPyV12	Human polyomavirus 12	Liver, rectum, stool	2013 ?
HPyV13	New Jersey polyomavirus (NJPyV)	Muscle, skin	2014 ?

(LIPyV), discovered in 2017, was isolated from skin swabs and oral fluids. LIPyV has not yet been formally assigned as a human polyomavirus species, yet it could be the fourteenth HPyV.

HPyVs within the family of *Polyomaviridae* share remarkable similarities in structure and genome composition and organization. In addition to their overall similarities in virion structure and genome organization, there are many parallels in the general schema of the infectious viral cycle including viral attachment, entry, trafficking, and the temporally regulated gene expression. However, the specific cellular factors required for a productive infectious cycle for each virus are disparate. Furthermore, polyomaviruses demonstrate a broad tissue tropism in the human host infecting skin, kidney, and brain cells, yet pathogenic viral disease is generally limited to those with an underlying immunosuppression. Polyomavirus disease ranges from nephritis to a fatal infection of the brain to skin cancer. The commonalities across structure, genome, and life cycles of HPyVs is described herein as well as detailed sections on specific differences in viral life cycles and clinical diseases of select HPyVs, including BK polyomavirus (BKPyV), JC polyomavirus (JCPyV), Merkel cell polyomavirus (MCPyV), and other recently-discovered HyPVs associated with significant human disease manifestations.

Virion Structure

Human polyomaviruses are classified by a nonenveloped protein capsid that is ~40–45 nm in diameter (Fig. 1). The virion capsid is comprised of two to three structural proteins, viral protein 1 (VP1), VP2, and VP3. VP1 is a pentameric protein that associates with either a VP2 or a VP3 in the interior of the virion capsid (Fig. 1). VP1 is the only capsid protein expressed on the exterior of the surface and serves as the viral attachment protein. Polyomavirus capsids are made up 360 molecules of VP1 arranged into 72 pentamers. The pentameric VP1 proteins associate with neighboring VP1 pentamers via C-terminal extensions in which the 72 pentamers are organized as 60 hexamers and 12 pentamers allowing for the formation of a spherical, capsid with a $T=7d$ skewed icosahedral structure. VP1 pentamers from all human polyomaviruses exhibit a β barrel jelly roll structure of VP1 monomers, formed by anti-parallel β strands, and the connecting loops of which serve as the binding site for many HPyV sialic-acid containing receptors. The virion capsid is stabilized by intra- and inter-pentameric disulfide bonds and calcium ions, and assembly occurs entirely in the nucleus where the concentration of calcium ions is highly regulated. The N-terminal region of the VP1 pentamers faces the interior of the virion and binds the minor capsid proteins, VP2 or VP3, which can interact with genomic DNA (Fig. 1).

Genome

Encased within the virion capsid is the circular double-stranded (ds) DNA genome that is approximately 5000 bp in size and packaged with cellular histones. The simple genome is comprised of 5–6 protein-coding genes resulting in a relatively limited protein repertoire, yet the polyomaviruses genome structure and organization are fairly conserved among HPyVs (Fig. 2). Polyomavirus genomes are divided into three regions organized in a bidirectional manner into early and late regions separated by a non-coding control region (NCCR) or regulatory region (RR). The early region of the genome encodes for the early viral proteins, the nonstructural proteins large T antigen (LTAg), small t antigen (STAg), and in some cases middle t antigen (MT), alternative T (ALT), and alternatively spliced T' proteins are encoded from an alternative splicing of a single mRNA transcript. The late viral

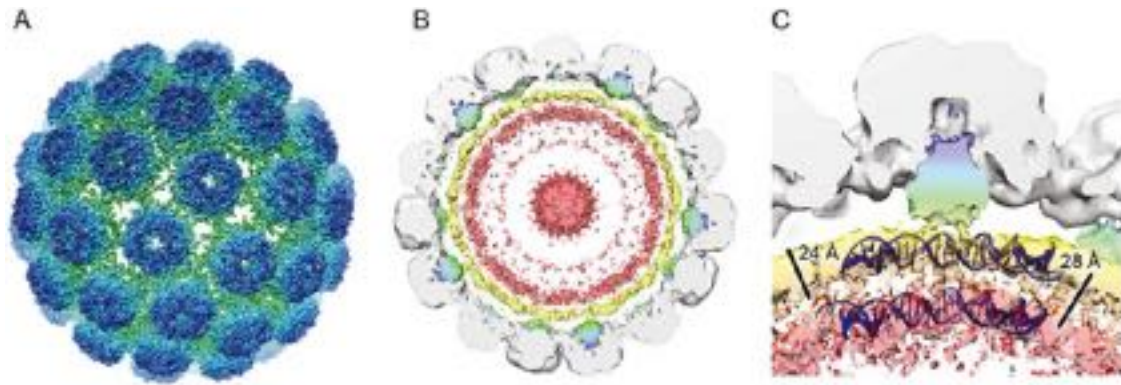


Fig. 1 Icosahedral structure of nonenveloped human polyomavirus virion. (A) High resolution structure of BKPyV resolved by cryo-electron microscopy (EM) at 3.8 Å demonstrating. (B) A view of the BKPyV virion interior with VP1 pentons (gray), VP2 (blue), and VP3 (green) and packaged dsDNA (yellow, pink). (C) Enlargement from structure (in (B)) demonstrates VP2/VP3 interactions with dsDNA, which is wrapped in cellular histones. Modified and reproduced from Hurdiss *et al.*, 2018. Structure. 26 (5), 839–847. and Hurdiss *et al.*, 2016. Structure. 24 (4), 528–536. under Creative Commons licenses.

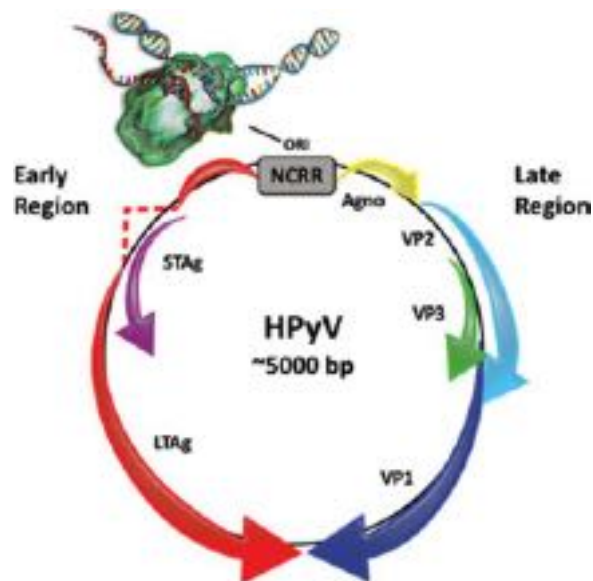


Fig. 2 General schematic of human polyomavirus genome. The dsDNA genome size is ~5000 bp and is organized into early and late regions separated by the non-coding control region (NCCR). The NCCR contains the origin of replication (ORI), and replication is driven through several functions of TAg, including helicase activity to unwind the viral dsDNA genome. The early region encodes for T antigen proteins including large T antigen (LTAg), small T antigen (STAg), and other splice variants of T antigen (indicated by dashed line, varies by HPyV). Not pictured are middle T Ag and truncated T Ag, present in some HPyVs. The late region encodes for structural proteins viral protein 1 (VP1), VP2, and VP3 (lacking in MCPyV). Some HPyVs, including JCPyV and BKPyV also encode for the nonstructural protein agnoprotein. Figure artwork generated by Michael Wilczek.

genes include the viral structural proteins VP1, VP2, and VP3 (except for MCPyV). JCPyV and BKPyV encode for an additional protein, agnoprotein, which is expressed predominantly in the cytoplasm of infected cells and is required for release of infectious progeny into the supernatant. JCPyV Agno protein has been demonstrated to play a role in egress through the nuclear membrane. The NCCR demonstrates genomic diversity among polyomaviruses and contains promoters for early and late genes, the origin of replication (ORI), a binding domain for LTAg, and promoters and enhancers. HPyV NCCRs undergo rearrangements within the host during active viral replication, yielding insertions, deletions, and duplications. The rearranged NCCR of polyomaviruses are generally associated with enhanced viral replication, expansion of tissue tropism, and thus viral pathogenesis. Viral strains with rearranged NCCRs are generally only found during viremia or pathogenic states. A number of HPyVs, including BKPyV, JCPyV, and MCPyV have been demonstrated to also encode for microRNAs (miRNAs) complementary to LTAg. These miRNAs can function by binding to LTAg mRNA and regulating expression of early genes and contribute to viral immune evasion.

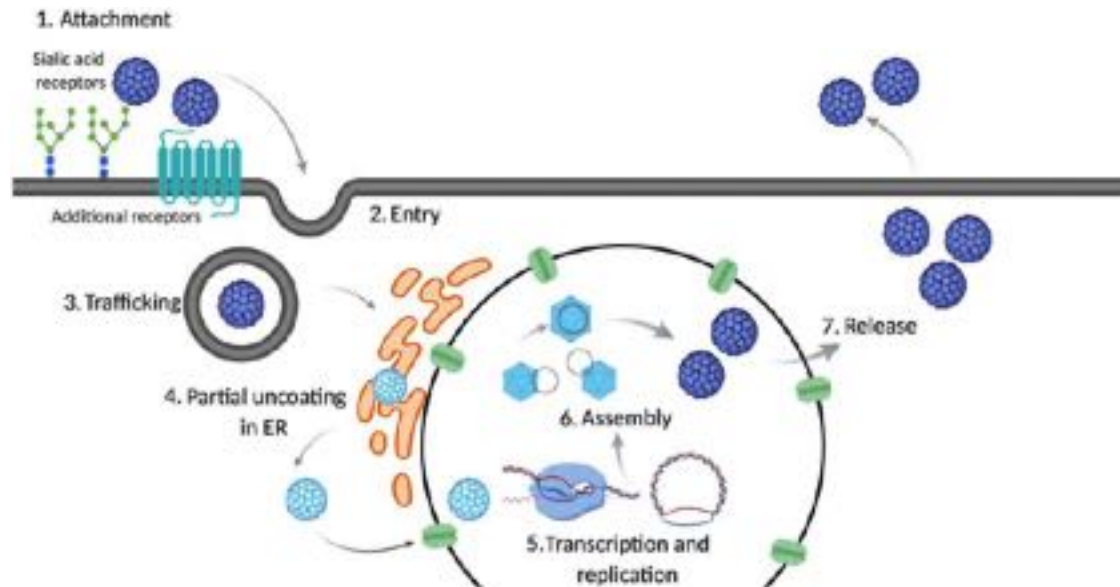


Fig. 3 General schematic of human polyomavirus life cycle. (1) Viral attachment is mediated by VP1 binding to sialic-acid containing receptors or alternate receptors such as glycosaminoglycans or serotonin receptor. (2) Viral entry is mediated via endocytosis (mechanism varies by HPyV). (3) HPyVs traffic through the endocytic compartment to the endoplasmic reticulum (ER), and (4) the virions are partially disassembled in the ER. (5) Transcription begins with viral early genes followed by replication of the dsDNA genome then transcription of viral late genes. (6) Virions are assembled in the nucleus before (7). release, possibly through cell lysis. Figure generated by Jeanne DuShane and created in [BioRender.com](#).

Infectious Cycle

Despite differences in tissue tropism and broad effects on host cells, the current information for human polyomaviruses suggests they follow the same general schema for infection of host cells ([Fig. 3](#)). HPyVs bind to cell surface receptors on host cells, and for most HPyVs this has been demonstrated to be a carbohydrate sialic acid receptor. The nature of the carbohydrate receptor differs among polyomaviruses but can include $\alpha 2,3$ -, $\alpha 2,6$ -, or $\alpha 2,8$ -linked sialic acids or alternate receptor such as glycosaminoglycans. In fact, a number of HPyVs attach to a sialic acid receptor expressed on a ganglioside, a glycosphingolipid comprised of a ceramide in a lipid-rich region of the membrane with an extracellular carbohydrate moiety with sialic acid residues. The utilization of gangliosides results in a lipid-mediated caveola- or lipid raft-mediated internalization process for some HPyVs. However, JCPyV requires a unique sialic acid receptor $\alpha 2,6$ -linked lactoseries tetrasaccharide c (LSTc) and also requires a secondary receptor, the 5-hydroxytryptamine family of receptors (5-HT) in the 2 subfamily and clathrin-mediated endocytosis viral internalization. BKPyV has also been demonstrated to enter specific cell types via a caveolae- and clathrin-independent mechanism.

Regardless of the internalization strategy, polyomaviruses are internalized into the endocytic compartment and traffic to the endoplasmic reticulum (ER). Within the ER, resident chaperones, disulfide isomerases and reductases, and components of the ER-associated protein degradation (ERAD) pathway facilitate partial uncoating of the viral capsid. The partially disassembled virion is then deposited into the cytoplasm and retrotranslocates to the nucleus where transcription and replication ensues in a temporally-regulated process. Transcription of viral early genes is initiated through host cell RNA polymerase II and transcription factors binding to the promoter region of early genes. A precursor mRNA is generated, which is then alternatively spliced to produce large tumor (T) antigen (LTag), small T antigen (STAg) and TAg splice variants. LTag is considered a master regulator of polyomavirus infection, exerting multiple functions necessary for the infectious process. The expression of LTag and STAg drives cell cycle progression from G0 or G1 into S phase by binding to cell cycle regulators pRb, p53, and PP2A. LTag binds pRb family members through the conserved LXCXE motif. Some HPyVs have additional open reading frames (ORFs) that encode for proteins although their functions are not well described. TSPyV and MCPyV encode for additional T antigens, middle T antigen (MTAg) and alternative tumor antigen (ALT), respectively. While the function of MTAg and ALT are not yet well characterized, they are expected to play roles in cellular proliferation and transformation due the presence of Rb binding motifs. Viral replication is carried out through reliance on the cellular replication machinery and proceeds in a bidirectional manner. LTag binds to the ori in the NCCR, unwinds the dsDNA genome through helicase activity, and drives viral replication. Finally, late viral gene transcription ensues on the complementary strand in the opposing direction from the early genes to produce the late viral genes VP1, VP2, and, in most cases, VP3, which then encapsidate the virion in the nucleus. MCPyV lacks a VP3, which in other HPyVs is translated from an internal start site within the VP2 transcript. JCPyV and BKPyV also produce the late viral gene product agnoprotein, which has a number of reported properties including roles in transcription and replication of the viral genome and viral release through viroporin activity. Polyomaviruses are thought to be released from infected cells upon viral lysis of the host cells, such as through the viroporin activity of JCPyV agnoprotein. However, nonlytic release for polyomaviruses, such as BKPyV, has also been reported and thus the process is not well understood.

BKPyV

BKPyV Clinical Features

BKPyV, also classified as HPyV1, was first identified in 1971 when it was isolated from the urine of a renal transplant patient with ureteric stenosis. BKPyV is shed in urine and is usually acquired in childhood, and 50%–90% of individuals are seropositive by adulthood. Transmission is thought to occur via a peroral or respiratory route, as BKPyV can be detected in the respiratory tract and tonsils, but could also occur through semen, transfusion, or organ transplantation (renal allografts). Initial infection results in either an asymptomatic infection or a mild respiratory illness before dissemination to the kidney and urogenital tract, where the virus establishes a persistent infection. BKPyV is shed in the urine of both healthy and immunosuppressed individuals, yet those with immune suppression have higher viral loads, which could be due to viral reactivation. Thus, BKPyV can be diagnosed by qPCR of urine or plasma samples. It is also detected through urine cytology as non-cell associated particles, and virions in infected “decoy cells” via electron microscopy. In immunosuppressed individuals, BKPyV can become reactivated resulting in clinical disease, most commonly polyomavirus-associated nephropathy (PyVAN) and hemorrhagic cystitis (PyVHC).

PyVAN results from BKPyV infection in renal tubular epithelial cells in kidney transplants and causes renal dysfunction. BKPyV reactivation due to immunosuppression is most common in kidney transplant (KT) recipients, occurring in up to 10% of KT recipients, which can result in 50%–70% allograft loss. Due to a high incidence of allograft rejection due to PyVAN, kidney transplant patients are regularly screened for BKPyV by viral load in the urine or plasma, and PyVAN is diagnosed by tissue biopsy. However, other BKPyV-associated clinical diseases including ureteric stenosis, encephalitis, meningoencephalitis, pneumonia, and vacuolopathy have been reported in immunocompromised patients. In individuals with PyVAN, BKPyV can also be detected in the urothelial cells of the bladder. On the other hand, PyVHC can occur in up to 15% of individuals who receive an allogenic hematopoietic stem cell transplantation or those otherwise immunosuppressed due to damage to the bladder. Additionally, the WHO International Agency for Research on Cancer (IARC) Working Group has classified BKPyV as Group 2B for “possibly carcinogenic to humans,” as increasing evidence suggests a role for BKPyV in the development of bladder cancer and other cancers of the urinary epithelia in organ transplant recipients, yet causation has not been established. Ureteric stenosis has been observed in transplant patients, characterized by damage to the ureteric epithelium due to inflammation. Finally, some cases of BKPyV-induced CNS disease have been observed including meningoencephalitis in an individual with AIDS and viral encephalitis in a pediatric heart transplant recipient. A case of BKPyV-induced progressive multifocal leukoencephalopathy (PML) has also been reported, in which samples were positive for BKPyV and not JCpV, which is the etiological agent for PML. Treatments for BKPyV-induced disease varies depending upon the disease pathogenesis and stage of disease. Treatment for PyVHC is supportive and involves palliative care and analgesia. The first line in PyVAN treatment involves altering the dosage or type of immunosuppressive drugs administered to transplant patients. Antiviral drugs are often used in addition to immunosuppressive agents to treat PVAN including nucleoside analog cidofovir or the lipid conjugate brincidofovir.

BKPyV Infectious Cycle

BKPyV attaches to ganglioside receptors containing $\alpha 2,3$ - or $\alpha 2,8$ -linked sialic acid of the b-series gangliosides GD2, GD3, GD1b, and GT1b. Following attachment to sialic acid receptors, BKPyV is internalized into cells through either a caveolae- or lipid-mediated pathway or through a clathrin- and caveolae-independent pathway, suggesting cell-type dependent differences in viral entry mechanisms. BKPyV then traffics through the endocytic compartment along microtubules to eventually arrive in the ER where the capsid is partially disassembled. In addition, autophagic and Rab-18 + vesicles are reported to play a role in trafficking of BKPyV. Similar to other polyomaviruses, BKPyV utilizes ER reductases, disulfide isomerases, and components of the ERAD pathway, including Derlin and heat shock proteins (Hsp), for particle disassembly. Derlin-1, Hsp105, and small glutamine rich tetratricopeptide repeat containing alpha (SGTA) mediate the movement of BKPyV from the ER to the cytosol. BKPyV is capable of transport into the nucleus through the nuclear pore complex via the importin α/β pathway through the nuclear localization signals (NLS) in VP2 and VP3. Like most DNA viruses, once in the nucleus, BKPyV hijacks the cellular transcription and replication machinery. In order to reprogram the cell into a viral factory, BKPyV activates the cellular signaling pathways mitogen activated protein kinase (MAPK) and Akt/ mTOR (mammalian target of rapamycin), which leads to increased cellular proliferation, and thus aids in the production of virus.

Viral transcription begins with production of the early genes, yielding the production of T antigens including LTag, STAg, and truncated TAg (truncTAg). LTag binds to pRb and p53, pushing the cell into S phase, further inducing cell division and preventing apoptosis. In addition, BKPyV activates the DNA damage response signaling pathway by activating the ATM (ataxia-telangiectasia mutated) and ATR (ATM- and Rad3-Related) kinases, which protect the host cells during viral replication. This counteraction of the DNA damage induced by BKPyV infection and activation of ATR is critical for production of viral progeny. Production of TAg and regulation of the cell cycle drive viral replication. The BKPyV archetype strain is isolated from urine samples and is the transmissible form of the virus. The archetype strain has five sequence blocks (O, P, Q, R, and S) with the ORI, TATA boxes, and binding sites for cellular transcription factors necessary for infection including Ap-1, Ets-1, NF1, NFAT, nuclear factor kappa B (NF- κ B), and Sp1. The NCCR of viral isolates from kidney and other sites during disease states exhibit variation in the NCCR with rearrangements of the sequence blocks. The rearranged strains usually contain deletions of P, Q, R, or S blocks and multiplication of the P block. The Gardner strain is the prototype of a rearranged BKPyV strain, and the Dunlop strain represents another rearranged strain commonly used in

BKPyV studies. BKPyV rearranged strains have been isolated from the kidneys of patients with BKPyVAN and from the blood, and those strains are associated with increased viral replication and higher viral loads. The rearrangements of the NCCR likely confers enhanced cellular tropism, allowing for increased replication in additional tissue types within the host and in vitro, as archetype strains replicate poorly in most cell culture models. BKPyV produces the late genes VP1, VP2, VP3, and agnoprotein. BKPyV release from cells following virion assembly has been shown to occur via lytic and non-lytic egress, and again may reflect cell-type dependent differences. Agnoprotein may play a role in virion release from cells, as proposed for JCPyV.

JCPyV

JCPyV Clinical Features

JC polyomavirus (JCPyV), or HPyV 2, was also identified in 1971 when it was isolated from the brain of a patient with PML. JCPyV infects the majority of the population with seroprevalence ranging from 30% to 80%, depending upon geographical location and sampling. JCPyV is shed in the urine and can be found in untreated wastewater, and thus is predicted to spread perorally. Initial infection is thought to occur in the tonsillar tissue but virus has not been detected in the saliva. JCPyV then spreads to secondary sites of replication including the kidneys and B cells of the bone marrow, where it establishes a lifelong persistent infection. While infection of the kidney is asymptomatic, periodic viral shedding occurs as virus can be detected in the urine of healthy individuals. JCPyV remains in the kidney in healthy individuals with full immune function, yet in individuals with severe immunosuppression, JCPyV can become reactivated and spread to the central nervous system (CNS) possibly crossing the blood-brain barrier through B lymphocytes or another mechanism. Another model proposes that JCPyV may persist in the brain and then become reactivated locally in the CNS. Within the brain, JCPyV infects glial cells, astrocytes and oligodendrocytes, which are responsible for production and maintenance of myelin in the brain. JCPyV infection of these cells results in cytolytic destruction resulting in demyelination and the causation of the fatal, neurodegenerative disease PML. Symptoms of PML include impaired cognitive function, motor dysfunction such as paralysis, hemiplegia, and speech and vision disturbances. In addition, JCPyV replication in neurons of the cerebellum has been associated with development of granule cell neuronopathy (GCN), which results in clinical symptoms including ataxia and progressive cerebellar atrophy. GCN can occur in conjunction with PML or independently, and has been associated with mutations in the C-terminus of VP1.

PML is diagnosed by magnetic resonance imaging (MRI) and the presence of JCPyV DNA in the cerebrospinal fluid (CSF). The development of PML is rapid and progressive, leading to demyelination within the white matter of the brain resulting in multifocal lesions. Individuals at risk for development of PML include those with HIV or those receiving immunomodulatory therapies. In the HIV-infected population, PML develops in ~5% of individuals. While the use of highly active antiretroviral therapy (HAART) has lessened the burden on HIV-infected individuals, PML is still considered an AIDS-defining illness with a 50% mortality rate in this population. The second most at-risk population are those receiving the immunomodulatory therapy natalizumab for immune-mediated diseases such as relapsing remitting multiple sclerosis (MS). Natalizumab is a monoclonal antibody specific for very late antigen-4 (VLA-4), $\alpha 4\beta 1$ integrin, which blocks T cell extravasation to the brain, where autoreactive T cells can attack myelin, the hallmark of MS pathogenesis. In natalizumab treatment of a JCPyV-seronegative patient, the incidence of PML development is 1:1000, yet the incidence can increase to as great as 1:100 dependent upon JCPyV serostatus, duration of natalizumab treatment, and prior immunosuppressive use. Further, there have been cases of PML in individuals receiving humanized monoclonal antibody treatments (rituximab, efalizumab, and infliximab) or dimethyl fumarates and other fumaric acid ester-containing drugs for immunomodulatory therapies such as rheumatoid arthritis (RA), Crohn's disease, and psoriasis. However, a particular predilection for PML in MS patients is illuminated by the nearly 800 cases of natalizumab-induced PML in this patient subset since 2006. Thus, natalizumab remains on the market but contains a black box warning of the potentially fatal outcomes due to PML.

PML is fatal within 1–2 years of symptom onset, and can be fatal within 3 months of symptom onset if left completely untreated. Currently the treatment options for JCPyV are very limited. In most cases, PML treatment begins with addressing the underlying immunosuppression. Treatment of the immunosuppression for HIV-infected individuals includes initiation or modifications to HAART regimens, while those on immunosuppressive therapies will undergo cessation of therapy. While these treatments can help to slow PML progression, they do not halt PML development. Further, there is currently no way to reverse demyelination, and thus these individuals often suffer severe debilitation. Moreover, removal of the immunosuppressive therapy can result in immune reconstitution inflammatory syndrome (IRIS), which can cause neurological worsening and result in fatality. Novel treatments for JCPyV are currently under development and have yielded some promising results under “compassionate use” trials. JCPyV VP1 vaccines in combination with immune boosting adjuvants cytokine interleukin 7 and toll-like receptor (TLR) 7/9 agonist led to clinical improvement and slowed PML progression. In addition, monoclonal antibodies capable of neutralizing PML variants with VP1 mutations also demonstrated promising results in the clinic. Use of selective serotonin reuptake inhibitors (SSRIs), such as mirtazapine, in an off-label use, has resulted in mixed clinical results, suggesting that this avenue of treatment requires further investigation.

Additionally, JCPyV infection has been detected in the brain in other non-PML disease states. For instance, JCPyV-associated encephalopathy has been reported, in which cortical pyramidal neurons of the gray matter were infected. Viral meningitis with the presence of JCPyV and absence of other viral infections have also been reported. JCPyV infection in the kidney is generally not associated with clinical disease, yet JCPyV has been reported to cause PyVAN in rare cases. Further, JCPyV has also been designated as Group 2B for “possibly carcinogenic to humans”, due to potential associations with colon cancer, yet definitive causation has not been established.

JCPyV Infectious Cycle

JCPyV infection is mediated by VP1 binding to cellular sialic acid receptor, $\alpha 2,6$ LSTc. The X-ray crystal structure of VP1 in complex with LSTc revealed specific VP1 residues (L54, S266, S268) that make key contacts with LSTc. Mutation of these residues *in vitro* results in reduced infectivity and specifically reduced viral attachment. Following viral attachment to sialic acid receptors, JCPyV enters cells through clathrin-mediated endocytosis (CME) through the 5-hydroxytryptamine (5-HT) 2 subfamily of receptors. The internalization of JCPyV via CME is dependent upon the activation of clathrin scaffolding proteins adapter protein 2 (AP2), β -arrestin, and the GTPase dynamin. Following internalization, JCPyV traffics into the endocytic compartment through the Rab 5-positive endosomes and caveolin 1-positive vesicles and traffics along microtubules and microfilaments. The virus eventually arrives at the ER before retrotranslocation to the nucleus for transcription and replication of the viral genome. JCPyV infection also induces the activation of the cellular signaling pathway, MAPK and activation of the terminal kinase extracellular signal-regulated kinase (ERK), which can reprogram cells to induce proliferation and growth. The activation of the MAPK pathway is necessary for viral transcription and infection in culture. JCPyV relies on the expression and activation of specific cellular transcription factors in order to complete the infectious process. Transcription and production of the JCPyV early gene product, LTag regulates the JCPyV infectious cycle. LTag binds to Rb and p53 to drive the cell into S phase inducing cell cycle progression and inhibiting apoptotic responses, respectively.

Initial infection of the kidney and subsequent transmission of JCPyV in the urine of infected individuals occurs with the archetype (CY) strain. However, within the infected individual, the virus undergoes significant mutations in the NCCR, during high levels of replication, yielding rearranged NCCRs. Isolates from the brain, denoted "PML-types" contain rearranged NCCRs in which the six sequence blocks (*a, b, c, d, e, and f*) have undergone deletions of repressor elements and duplication of enhancer elements. The Mad1 prototype strain is a PML-type strain that was isolated from the brain of a PML patient in Madison, WI. Mad1 contains deletions of *b* and *d* and duplication of an enhancer element (*a, c, e, a, c, e, f*). These rearrangements result in duplication of the TATA box and transcription factor binding sites for transcription factors that have been reported to play a critical role in JCPyV infection including DDX-1, Sp1, Spi-B, NFAT4, NF-1X, NF- κ B, Oct-6, and YB-1/Purx. While the NCCR rearrangements facilitate additional transcription factor binding sites, these alterations are also necessary for neurotropism. Interestingly, mutations in the sialic-acid binding sites in VP1 are also associated with neurotropism in the host, as 90% of viral isolates from individuals with PML contain at least one mutation in one of the sialic acid-binding sites: L54, S266, and/or S268. The mutation of VP1 and NCCR rearrangements suggest that mutations may occur during times of enhanced replication, ultimately resulting in increased neurovirulence.

MCPyV

MCPyV Clinical Features

MCPyV is a common polyomavirus of the skin with ~60%–80% seroprevalence in the population, and the majority of individuals exposed during childhood. While MCPyV persists in the skin of healthy individuals, it also has been identified as a causative agent of Merkel cell carcinoma (MCC). To date, MCPyV is the only polyomavirus with a clear association with human cancer. MCC is a rare, yet aggressive form of skin cancer that occurs in sun-exposed regions of the body including the head, neck, arms, and legs. Elevated risk for MCC is associated with advanced age, exposure to sunlight and UV radiation, and immunosuppression. The link to immunosuppression guided the search for an infectious cause of MCC, especially given that risk of MCC development is 10-fold higher in individuals with HIV/AIDS. Immunosuppression due to autoimmune conditions such as rheumatoid arthritis (RA), solid organ transplants, and other co-morbidities such as chronic lymphocytic leukemia and non-Hodgkin lymphoma also increase the risk of MCC development. MCC is most common in those with advanced age, as the median age of diagnosis is 76 years old (National Cancer Data Base). However, HIV-positive individuals can experience an earlier age of onset and development of MC tumors in areas not exposed to the sun such as the oral or anogenital mucosa. MCC has a 46% rate of disease-associated mortality, and prognosis is especially poor in AIDS patients that can have advanced tumor stage and reduced survival.

Diagnosis of MCC is usually carried out by skin biopsy and pathological examination. Biopsy samples are stained with antibodies such as CM2B4, which detects MCPyV T antigen, and the cytokeratin 20 antibody that is a marker of Merkel cells. In addition, individuals with MCC may undergo further testing such as sentinel lymph node biopsy and/or a CT scan to determine if the cancer has metastasized. Further, serology testing for MCPyV oncoprotein antibody titer has been recommended for determining patient prognosis, as MCPyV-positive tumors usually have a better prognosis in comparison to MCPyV-negative tumors. Thus, tracking antibody titers can help in MCC disease management and risk stratification for treatments. It is recommended that PCR testing for the presence of MCPyV should be used in conjunction with other methods, as PCR results can include false positives due to contamination as well as false negatives due to strain-specific differences.

The treatment options for MCC vary greatly depending upon the stage of cancer, MCPyV status, and underlying immunosuppression. In some cases of early state MCC, surgical excision or Mohs micrographic surgery may be the only course of treatment necessary. Yet, in other cases, additional radiation therapy is warranted. In metastasized MCC, chemotherapy may be the treatment course, and while responsive in the short term, there is usually developed resistance and also immunosuppression, yielding reduced immune reaction to the tumors. Another treatment option for MCC, which has yielded successful results, include immune checkpoint inhibitors (ICIs), programmed death 1 (PD-1) and programmed death-ligand 1 (PD-L1) antibodies. PD-L1 is upregulated in MCPyV-positive MCC tumors, and when PD-L1 is recognized by PD-1 receptors on T lymphocytes, the immune

response to the tumor is down-regulated, resulting in tumor immune evasion. Thus, ICIs targeting either PD-1 or PD-L1 stimulate the anti-tumor immune response so immune cells can recognize and destroy tumor cells. This type of treatment is only possible in individuals who have a stable, robust immune system, and thus is not effective in those who are immunocompromised due to HIV or immunosuppressive therapy for underlying immune-mediated disease. In fact, patients with higher levels of CD8⁺ and CD3⁺ T lymphocytes have a better prognosis, reduced metastases, and improved survival rates.

MCPyV was first detected in 2008 when it was isolated from tumors of individuals with Merkel cell carcinoma. MCPyV was identified using digital transcriptome subtraction. The presence of MCPyV in tumors was confirmed by sequencing and Southern blot analysis. Of the 10 tumors tested, 8 were positive for MCPyV, confirming a strong association (80% MCPyV positive), which led to the WHO IARC classification as a Group 2A, "carcinogenic to humans." The genome of MCPyV was found to be clonally integrated into MCC tumors, making MCPyV the only human polyomavirus that is known to integrate into the host genome. The integration occurs before clonal expansion of the tumor and is hallmarked by expression of small T antigen (ST) and a mutated form of large T antigen (LT), which inhibits viral replication. However, the viral structural proteins VP1 and VP2 are not expressed in tumors suggesting that active viral replication is not occurring in these tissues. The mutations that arise in LT include a truncation, deletion of the DNA binding region (ORI), and deletion of the helicase domains and growth suppressing domains. The LXCXE motif that binds to the tumor suppressor pRb remains intact, and thus when LT binds to Rb the tumor suppressor functions are inactivated contributing to tumor development. Despite the role of LT in tumorigenesis, MCPyV ST is the major oncogenic protein leading to tumor progression in MCC. In fact, expression of ST can lead to cellular transformation *in vitro* and *in vivo* independent of LT expression. ST can increase levels of LT through the stabilizing domains, and ST can also alter the function of FBXXW7 to inhibit Ub ligase activity leading to accumulation of LT and transcription factors that regulate cell cycle progression including c-Myc and cyclin E. The ST protein also induces hyperphosphorylation of 4E-binding protein (4E-BP1) to promote protein translation and increase cellular proliferation and transformation.

MCPyV Infectious Cycle

The mechanism of MCPyV pathogenesis and resultant MCC development is not fully understood, and the cell types infected in the host are not clear. MCC was originally described in 1972 as a trabecular carcinoma of the skin. The name was later changed to MCC named for the Merkel cells, mechanoreceptor cells that aid in perception of touch, within the tumor tissue. However, Merkel cells do not support MCPyV replication. While this might indicate that a lack of replication could facilitate viral integration and the development of MCC tumors, the cells of origin of MCC tumors remains unclear. Epithelial cells and fibroblasts, including human dermal fibroblasts, are infected by MCPyV. Thus, it has been proposed that tumors may be of dermal fibroblast origin or perhaps Merkel cells are infected through a bystander mechanism. Further, expression of particular cell markers of pro/pre-B cells suggest that MCC tumors may originate from B cells. Importantly, tissue tropism of viruses in the host is dictated by the expression of specific host-cell factors necessary for infection and the virus-host cell interactions that ensue. However, a lack of a tractable cell culture model for MCPyV replication has limited the understanding of the MCPyV infectious cycle. Infection is initiated by attachment to cell-surface glycosaminoglycans (GAGs) such as heparan sulfate (HS) proteoglycans. The viral capsid protein VP1 also binds specifically to sialic-acid containing ganglioside receptors with the Neu5Ac motif, such as GT1b, presumably to facilitate a post-attachment cell entry step. The subsequent steps in the infectious cycle remain to be elucidated and may differ greatly in cell types that support active replication in comparison to the integration in pre-tumor cells.

Other HPyVs

Since the beginning of the 2000s, there have been 11 (possibly 12), human polyomaviruses identified. The HPyVs represent broad sites of isolation, tissue tropism, and disease outcomes. Due to the relative novelty of these viruses, researchers are still working to understand more about the infectious cycle and viral pathogenesis associated with infections. In some cases, a tissue culture model or an infectious experimental form of the virus have not been established, and thus alternatives are utilized such as molecular, biochemical, or *in silico* analyses along with utilization of virus like particles (VLPs) or pseudoviruses (PsV). VLPs are generated by the expression of viral capsid proteins, while pseudoviruses are comprised of the viral capsid proteins along with a reporter. Each of these particle types have demonstrated an overall similar size and structure to native polyomavirions and have been useful in experimental studies for the early steps in the infectious cycle. While many HPyV experimental models are still under development, the clinical understanding and seroprevalence of HPyVs has been a focus of study to understand the medical importance of HPyV infections.

Polyomaviruses of the Skin

Of the thirteen HPyV species, seven of the HPyVs have been isolated from skin including MCPyV, TSPyV, HPyV6, HPyV7, HPyV9, MWPyV (HPyV10), and HPyV13, suggesting common circulation of these skin HPyVs in the population. As described above, MCPyV is a causal agent in the development of MCC. The other HPyVs of the skin have been associated with other asymptomatic or non-malignant diseases.

TSPyV has been isolated from individuals with TS, which is characterized by spiny follicular papules and hyperkeratotic spicules usually localized to the face in immunocompromised individuals. The progression of TS can result in thickening of the skin, disfigurement, and facial spines. The underlying immunosuppression in TS patients has generally been the result of immunosuppressive therapies, organ transplants, or leukemia. Approximately 70% of the population is seropositive for TSPyV, and there is a higher seroprevalence in solid organ transplant patients, including renal transplant recipients. Interestingly, samples obtained from skin show low levels of seroprevalence in the general population. TS is diagnosed through physical examination in combination with biopsy and histopathological analysis. Samples can be analyzed by immunohistochemistry (IHC) with VP1-specific antibodies, electron microscopy for viral particles, or through PCR analysis of purified viral DNA. Transmission of TSPyV is thought to occur via peroral or respiratory route, and viral persistence within tonsillar tissue has been suggested. While TSPyV has been mostly linked to the skin pathology of TS, additional studies have demonstrated that TSPyV can be isolated from blood, feces, urinary tract, and respiratory samples. A case of a respiratory sample from an immunocompromised child with respiratory illness without facial papules spines has been described, suggesting a potential for respiratory illness. TS has been treated with a topical cidofovir cream leading to partial resolution of symptomatic disease. Studies using a TSPyV PsV demonstrated that TSPyV binds to sialylated glycans with an α 2,3- and α 2,6-linked sialic acid. While the replication cycle for TSPyV has not been characterized, the viral genome encodes for five ORFs: STAg, LTA_g, VP1, VP2, and VP3. STAg contains the conserved binding motif for PP2A while LTA_g has binding sites for pRb and p53 but a strong interaction of LTA_g with p53 has not been confirmed.

HPyV6 and HPyV7 were identified from skin swabs from healthy individuals, and the seroprevalence of HPyV6 is reported to be ~80% and HPyV7 is ~66%. Given the high rate of seroprevalence, HPyV6 and 7 infections are likely asymptomatic in healthy individuals, yet clinical disease has been noted in individuals with immunosuppression due to immunosuppressive therapy, solid organ transplantation, HIV infection, or uncharacterized immunosuppression. Additionally, HPyV6 has also been isolated from tonsils, CSF from a PML patient, feces, and the respiratory tract. HPyV7 has been identified in tonsils, urine, feces, and the respiratory tract. HPyV6 and 7 are can infect keratinocytes in the epidermis of the skin, and both been associated with pruritic and dyskeratotic dermatitis. Additionally, HPyV6 has been associated with epithelial neoplasm, angiolymphoid hyperplasia, and keratoacanthomas. HPyV7 has been associated with thymic epithelial tumors. Both viruses are diagnosed by PCR and IHC, as TAg and VP1 can be detected in samples with specific antibodies. HPyV6 and 7 express LTA_g, STAg, and VP1, 2, and 3. The receptors for HPyV 6 and 7 have not been identified but structural studies demonstrate that the sialic acid-binding site present in the VP1 attachment protein of many HPyVs is occluded in these viruses, and thus a receptor remains to be identified. LTA_gs of HPyV 6 and 7 have been shown to bind specifically to p53.

HPyV9 was isolated from a kidney transplant patient but has also been found from skin swabs, with a low seroprevalence of ~20%–50%. HPyV10, MWPyV, and Mexico polyomavirus (MXPV) represent the different isolates of the same virus, demonstrating a 95%–99% sequence identity. HPyV10 was identified from condyloma samples from a patient with WHIM syndrome, which is characterized by uncontrolled human papillomavirus (HPV) infections. Additionally, HPyV10, MWPyV, and MXPV have been isolated from stool samples, and seroprevalence is ~75% of the population based on a VP1 ELISA. The MWPyV LTA_g can bind to pRb and p53, and STAg can bind to PP2A. Interestingly, the recently identified LIPyV, which was isolated from skin swabs and oral fluids has a seroprevalence of ~5%, similar to that of NJPyV. LIPyV has been detected in the stool of diarrhetic cats, suggesting a possible feline origin and a host jump mode of transmission.

Polyomaviruses of the Other Tissues

KIPyV and WUPyV were originally isolated from the respiratory samples from children with acute respiratory illness. Seroepidemiology studies suggest that transmission occurs during childhood and either causes a mild respiratory illness or is asymptomatic and remains as a persistent infection, possibly in the tonsils. Seroepidemiological studies suggest that 70% of the population are seropositive for KIPyV and 80% are seropositive for WUPyV, as detected by VP1 antibodies by ELISA. The prevalence of both KI and WU in respiratory secretions is higher in immunosuppressed individuals, such as those with HIV. Additionally, KIPyV has been isolated from respiratory and stool samples of a patient with hematopoietic stem cell transplantation. STLPyV was isolated from stool samples, and can be detected in ~70% of the population. TAg of STLPyV has a unique splice variant, referred to as 229T. Additionally, the STLPyV represents genetic mosaicism with the closely related MWPyV. HPyV12 was isolated from resected liver tissue but is also found in rectum tissues and fecal samples, with a seroprevalence of ~20%. NJPyV was isolated from a pancreatic transplant recipient with retinal blindness and vasculitis myopathy, with a seroprevalence of ~5%. A high degree of sequence identity to NJPyV, suggests a possible transmission of this virus from shrews to humans, yet this remains to be confirmed.

Perspectives

Human polyomaviruses represent a diverse group of human pathogens that generally cause asymptomatic infection in healthy individuals and persist in an asymptomatic state at varying frequencies in the population. However, HPyVs can cause serious disease in immunocompromised individuals ranging from skin cancer to allograft rejection to a fatal neurodegenerative disease.

HPyVs share many commonalities including virion structure, organization of the genome and protein-coding regions, binding sites for particular host-cell transcription factors, conserved mechanisms of viral replication and virion assembly. Less is known about the recently identified polyomaviruses, which also have the propensity to cause serious disease and could represent viral emergence into the population from other species. Thus, additional research on polyomaviruses will continue to provide insights into these disease mechanisms. Lastly, given the species-specificity of polyomaviruses, animal models are lacking. Development of animal models would revolutionize the polyomavirus field and provide tremendous opportunities to explore viral infection and pathogenesis *in vivo*.

Further Reading

- Ambalathingal, G.R., Francis, R.S., Smyth, M.J., *et al.*, 2017. Clinical Microbiology Reviews 30 (20), 503–528.
- Assetta, B., Atwood, W.J., 2017. The biology of JC polyomavirus. Journal of Biological Chemistry 398 (8), 839–855.
- DeCaprio, J.A., 2017. Merkel cell polyomavirus and Merkel cell carcinoma. Philosophical Transactions of the Royal Society B 372, 20160276.
- DeCaprio, J.A., Garcea, R.L., 2013. A cornucopia of human polyomaviruses. Nature Reviews Microbiology 11 (4), 264–276.
- Feltkamp, M.C.W., Kazen, S., van der Meijden, E., *et al.*, 2013. From Stockholm to Malawi: Recent developments in studying human polyomaviruses. Journal of General Virology 94, 482–496.
- Ferency, M.W., Marshall, L.J., Nelson, C.D., *et al.*, 2012. Molecular biology, epidemiology, and pathogenesis of progressive multifocal leukoencephalopathy, the JC virus-inducing demyelinating disease of the human brain. Clinical Microbiology Reviews 25 (3), 471.
- Haley, S.A., Atwood, W.J., 2017. Progressive multifocal leukoencephalopathy: Endemic viruses and lethal brain disease. Annual Review of Virology 4, 439–467.
- Harms, P.W., Harms, K., Moore, P.S., *et al.*, 2018. The biology and treatment of Merkel cell carcinoma: Current understanding and research priorities. Nature Reviews Clinical Oncology 15, 763–776.
- Helle, F., Brochet, E., Handala, L., *et al.*, 2017. Biology of the BKPyV: An update. Viruses 9, 327.
- Kamminga, S., van der Meijden, E., Feltkamp, M.C.W., *et al.*, 2018. Seroprevalence of fourteen human polyomaviruses determined in blood donors. PLoS One 13. e0206273.
- Liu, W., MacDonald, M., You, J., 2016. Merkel cell polyomavirus infection and Merkel cell carcinoma. Current Opinion in Virology 20, 20–27.
- Mayberry, C.L., Nelson, C.D.S., Maginnis, M.S., 2017. JC polyomavirus attachment and entry: Potential sites for PML therapeutics. Current Clinical Microbiology Reports 4 (3), 132–141.
- Moen, U., Krumbholz, A., Ehlers, B., *et al.*, 2017. Biology, evolution, and medical importance of polyomaviruses: An update. Infection, Genetics and Evolution 54, 18–38.
- Rinaldo, C.H., Tylden, G.D., Sharma, B.N., 2013. The human polyomavirus BK (BKPyV): Virological background and clinical implications. APMIS 121, 728–745.
- Sheu, J.C., Tran, J., Rady, P.L., *et al.*, 2019. Polyomaviruses of the skin: Integrating molecular and clinical advances in the emerging class of viruses. British Journal of Dermatology 180, 1302–1311.

Relevant Websites

- <https://merkelcell.org/about-mcc/>.
About Merkel Cell Carcinoma.
Merkel Cell Carcinoma.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsDNA-viruses/w/polyomaviridae.
Polyomaviridae.
Polyomaviridae.
dsDNA Viruses.
ICTV.
- <https://www.ninds.nih.gov/Disorders/All-Disorders/Progressive-Multifocal-Leukoencephalopathy-Information-Page>.
Progressive Multifocal Leukoencephalopathy.
NINDS.
NIH.

Human T-Cell Leukemia Virus-1 and -2 (*Retroviridae*)

Amanda R Panfil and Patrick L Green, The Ohio State University, Columbus, OH, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Clonality The number or abundance of HTLV-infected T-cell clones in an infected individual.

Immortalization Process by which cells evade normal cellular senescence and begin to proliferate or grow indefinitely. An early step in oncogenesis.

Latency Period of restricted or altered viral gene expression.

Leukemia Cancer of the body's blood forming tissues; cancer of the blood or bone marrow.

Lymphoma Cancer of cells in the lymph nodes.

Oncogenesis Development or induction of cancer.

Proliferation An increase in the number of cells caused by a disturbance in the balance between cell division and cell death.

Proviral load Viral DNA burden in an HTLV-infected individual. The number of proviral copies per set number of peripheral blood mononuclear cells.

Provirus Viral genome that is integrated into the host cell DNA or chromosome.

Transformation Change in the growth or reproduction of cells.

Tumorigenesis The formation of tumors or cancer.

Classification

Human T-cell leukemia viruses (HTLVs) are classified within the retroviridae family. To date, four distinct types of HTLVs have been identified: HTLV-1, HTLV-2, HTLV-3, and HTLV-4. HTLV-1 was first detected and isolated in the laboratory of Dr. Robert Gallo. The virus was discovered in a T-cell line established from a cutaneous T-cell lymphoma patient. Around the same time, researchers in Kiyoshi Takatsuki's laboratory in Japan defined a new, distinct form of leukemia with unusual clinical features and cellular morphology. This new leukemia was called adult T-cell leukemia (ATL) and subsequent studies would identify and isolate HTLV-1 from this tumor. Over the past several decades, several epidemiological and molecular studies have shown that HTLV-1 has the ability to disrupt growth regulation and immortalize human T-cells, linking this virus to several types of human disease (to be discussed). HTLV-2 is genetically and immunologically related to HTLV-1, with around 70% nucleotide identity and several overlapping gene functions. HTLV-2 was originally isolated from a patient with a variant form of hairy T-cell leukemia. Like HTLV-1, HTLV-2 can also infect and immortalize human T-cells *in vitro*. However, there is no strong disease association related to this virus. Indeed, a vast majority of HTLV research has been focused on HTLV-1 due to its significant disease connection. Nevertheless, comparative studies between the related HTLV-1 and HTLV-2 have been, and continue to be, extremely informative for identifying specific virus-host interactions associated with the pathogenic process (Table 1). More recently, HTLV-3 and HTLV-4 have been identified in bush meat hunters in Cameroon. While HTLV-3 and HTLV-4 have similar genomic organization to HTLV-1 and HTLV-2, their biology and disease association remain less clear.

Soon after the discovery of HTLV-1, highly related novel viral sequences were found in T-cells of several non-human primates and termed STLVs (simian T-cell leukemia viruses). Like HTLV-1, STLV-1 can cause T-cell leukemia/lymphoma in non-human primates and these tumors display many of the same hallmarks found in ATL patients. According to several different phylogenetic algorithms, it is suspected that many HTLV-1 subtypes arose from zoonotic interspecies transmission between monkeys and humans. Thus, STLV

Table 1 Summary and comparison of various aspects of HTLV-1 and HTLV-2 viruses

	HTLV-1	HTLV-2
<i>Genome</i>	Complex deltaretrovirus	70% nucleotide similarity to HTLV-1
<i>Infection</i>	Cell-to-cell	Cell-to-cell
<i>Viral entry</i>	HSPG and NRP1 for binding, GLUT1 for entry	GLUT1 and NRP1 for both binding and entry
<i>Transcription</i>	Tax-1-mediated LTR activation; antagonized by HBZ	Tax-2-mediated LTR activation; antagonized by APH-2
<i>Transmission</i>	Breastfeeding, contaminated blood products, sexual transmission	Mainly contaminated blood products, some mother-to-child and sexual transmission
<i>Immortalization</i>	CD4+ T-cells	CD8+ T-cells
<i>Transforming proteins</i>	Tax-1, HBZ	Tax-2
<i>Disease</i>	Yes; ATL, HAM/TSP, HTLV-1-associated uveitis, dermatological conditions	No; some neurological abnormality
<i>Prevalence</i>	5–10 million people worldwide; areas of endemic infection	800,000 people worldwide
<i>Immune response</i>	CTL against viral gene products	CTL against viral gene products
<i>Treatment</i>	Chemotherapy, antiviral therapy, targeted therapies, alloHSCT	None reported

is a useful animal model to study infection, seroconversion, and disease pathogenesis. STLV-2 and STLV-3 share close sequence similarity with HTLV-2 and HTLV-3, respectively. However, there is no known simian counterpart to HTLV-4 identified to date.

Virion Structure

HTLVs, like all retroviruses, contain two single-stranded positive-sense copies of RNA as their genetic material. The RNAs are packaged with various viral enzymes (reverse transcriptase (RT), integrase, and protease) and surrounded by capsid proteins forming a polymorphic, roughly spherical viral nucleocapsid. Also contained within the nucleocapsid is a cellular factor called tRNA^{pro} that serves as the primer for viral RT initiation. A lipid bilayer envelope of host-cell origin containing viral envelope proteins surrounds the capsid. HTLV-1 virions measure approximately 105–115 nm in diameter based on electron tomography of naturally infected T-cells and cell lines.

Viral Entry

Like most retroviral envelope proteins (Env), HTLV Env is a type-I transmembrane glycoprotein synthesized as a precursor protein in the endoplasmic reticulum, delivered through the Golgi to the plasma membrane, and subsequently acquired on the surface of virions during budding from infected cells. The Env glycoprotein precursor is cleaved by a cellular protease to produce surface (SU) and transmembrane (TM) subunit proteins. Specific interactions between Env and particular receptors on the surface of target cells are the initial and essential step in HTLV infection. Of note, infection can also occur following internalization via an endocytic pathway. In both instances, entry is initiated by SU binding to a cellular receptor, followed by fusion of viral and cellular membranes directed by the TM subunit. Prior to binding of SU to a receptor, the TM subunit is maintained in close association with the SU subunit in a fusogenic-inactive metastable conformation. This conformation of SU/TM keeps a fusion peptide found on the TM domain buried, thus preventing premature membrane fusion, Env inactivation, and cell toxicity. TM activation occurs when SU binds to a cellular receptor, triggering alterations of the SU/TM interaction and enabling TM to attain its fusogenic state.

The receptors and co-receptors for HTLV virion binding and entry are expressed on a broad range of cell types. Indeed, both HTLV-1 and HTLV-2 have been detected in a variety of hematopoietic cells from virus-infected individuals. The primary HTLV receptor is glucose transporter type 1 (GLUT1). Loss of GLUT1 expression inhibits virion binding and infection in cell culture, while ectopic expression of GLUT1 restores HTLV infectivity. GLUT1 binds SU and promotes Env-mediated virion entry via SU residues D106 and Y114. Two additional factors, neuropilin-1 (NRP1) and heparan sulfate proteoglycans (HSPG), function as co-receptors for HTLV. NRP1 is overexpressed in transformed cells and upregulated in activated T-cells, the preferential cell target of HTLV infection. NRP1 binds SU, and together with GLUT1, form a stable tripartite complex when co-expressed in cells. Interestingly, HTLV-2 SU binding to target cells does not depend on HSPGs. Based on this collective body of work, a potential model for HTLV entry posits interaction of SU with heparan sulfate chains, presumably on cell surface HSPG, initiates virus attachment (at least for HTLV-1). The SU is then able to bind NRP1, which commonly associates with heparan sulfate chains, to allow virus binding. The binding of SU to HSPG/NRP1 allows a conformation change in SU that exposes the GLUT1-binding domain, thus triggering virus/cell fusion and release of the HTLV nucleocapsid into the cytoplasm.

HTLV-1 and HTLV-2 SU proteins share 65% identity at the amino acid level and share many of the same cell receptors (GLUT1, NRP1). However, the *in vivo* tropism of HTLV-1 and HTLV-2 are distinct with HTLV-1 primarily detected in CD4⁺ T-cells and HTLV-2 found primarily in CD8⁺ T-cells. This discrepancy in T-cell tropism has been investigated using HTLV-1/HTLV-2 recombinant viruses and mapped to the Env region in the viral genome. An obvious explanation for this could be differences in the receptors the viruses use to enter T-cells. However, as previously discussed, HTLV-1 and HTLV-2 share many of the same receptors/co-receptors. An alternative explanation could therefore be the expression level of these receptors on the surface of T-cells. Indeed, HSPGs are only essential for HTLV-1 SU binding, not HTLV-2 SU. Interestingly, activated CD4⁺ T-cells express high levels of HSPG and low levels of GLUT1, while activated CD8⁺ T-cells express low levels of HSPG and high levels of GLUT1. Overexpression of HSPG in CD8⁺ T-cells increases HTLV-1 internalization, while overexpression of GLUT1 in CD4⁺ T-cells enhances HTLV-2 internalization. This suggests HTLV-1 is more dependent on HSPG, while HTLV-2 relies heavily on GLUT1 levels. In summary, HTLV-1 Env requires HSPG and NRP1 for binding and GLUT1 for entry, while HTLV-2 Env requires GLUT1 and NRP1 for both binding and entry.

Intriguingly, recent studies found both HTLV-1 and HTLV-2 proviruses were detected as early as 1 week post-infection in both CD4⁺ and CD8⁺ T-cells using an *in vivo* rabbit model of infection. This would argue that the tropism and transformation by HTLV-1: CD4⁺ T-cells and HTLV-2: CD8⁺ T-cells is not solely conferred by the viral envelope at the level of entry. Indeed, this same study showed early proliferation of both CD4⁺ and CD8⁺ T-cells irrespective of HTLV-1 or HTLV-2 virus type using *in vitro* cell culture immortalization assays. A late selection and outgrowth of the preferred T-cell type was observed several weeks after initial viral infection, indicating the predominance of T-cell type occurs during the clonal expansion process. This also highlights the potential role of HTLV Env not only in viral entry, but also during transformation and clonal proliferation.

Genome

HTLV-1 and HTLV-2 are complex deltaretroviruses that contain the standard retroviral structural and enzymatic genes; *gag*, *pol*, and *env* (Fig. 1). In addition, there is a unique region in the 3' end of the genome that encodes several regulatory and accessory genes

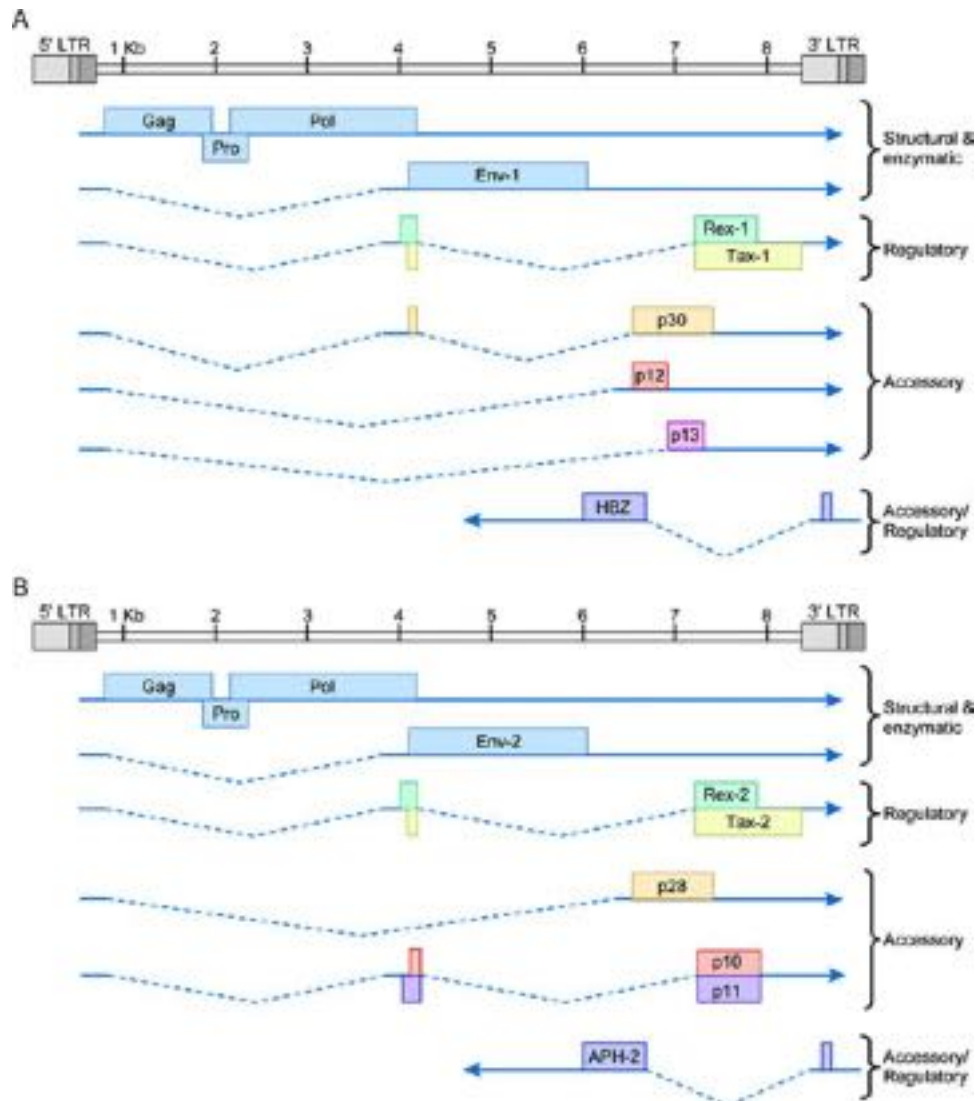


Fig. 1 HTLV genomes. Schematics of (A) HTLV-1 and (B) HTLV-2 proviral genomes. The viral *gag*, *pro*, *pol*, and *env* structural/enzymatic genes are flanked by 5' and 3' LTRs. The pX region at the 3' end of the viral genome contains several regulatory and accessory genes. Drawing is intended to be illustrative and not to exact scale.

on both the sense and antisense genomic strands. Originally this region was termed 'pX'. The HTLV proviral genome is roughly 9 kb in length and is flanked by 5' and 3' long terminal repeats (LTRs). The LTRs are exact duplicates comprised of a U3, R, and U5 region. These regions contain promoter elements, polyadenylation signal sequences, and other regulatory sequences necessary for proper viral transcription. HTLV-1 expresses 8 major mRNA species, while HTLV-2 expresses 7 major mRNA species from the sense strand. Both viruses express 1 major mRNA species from the antisense strand.

Like other retroviruses, the unspliced mRNA serves as both genomic RNA and mRNA that encodes Gag (MA-CA-NC) and Gag-Pol (MA-CA-NC-RT-IN) polyproteins. The HTLV protease (Pro) is generated as part of the fusion protein Gag-Pro or Gag-Pro-Pol, which are dependent on ribosomal frameshifts for their generation. The Env glycoprotein precursor is encoded by a major singly spliced mRNA. Tax and Rex regulatory proteins are encoded by a doubly spliced, or completely spliced, mRNA. They share partial overlapping reading frames and their start codons lie in the 2nd exon. Tax translation is preferentially favored over Rex due to a strong Kozak consensus sequence. Tax is a potent activator of HTLV mRNA transcription through activation of the viral promoter. Rex regulates the export of unspliced and singly spliced or incompletely spliced viral mRNA to the cytoplasm. The major role of HTLV-1 p30 is to sequester Tax/Rex mRNA in the nucleus and thus downregulate their expression. p30 has also been shown to affect the cell cycle and interfere with Tax transcriptional activation, thus favoring a latent cell environment. The HTLV-2 genomic and functional equivalent to p30 is p28. The viral proteins p12, p13, p10, and p11 are not required for viral replication *in vitro*, but are suspected to play a role in viral persistence *in vivo*. For a detailed description and summary of these accessory genes, please see a review by Edwards, *et al.* The antisense encoded HBZ protein (HTLV-1 bZIP transcription factor) functions to

antagonize Tax transcriptional activation through sequestration of cellular transcription factors, thus regulating viral transcription/replication and facilitating the establishment of latency. The HTLV-2 equivalent to HBZ is called APH-2 (anti-sense protein of HTLV-2) and it shares significant amino acid homology and functional similarities (to be discussed).

Life Cycle

The HTLV capsid core contains two copies of viral genomic RNA along with reverse transcriptase, integrase and viral protease. Similar to HIV-1, HTLV is believed to undergo complete reverse transcription of its viral genome after intracellular uncoating of the capsid core. The partially uncoated viral core, which contains the preintegration complex, is then translocated to the nucleus where it can integrate into the host cell chromosome to form the provirus. As with other retroviruses, integration is catalyzed by integrase and facilitated by virus-specific host factors. Recently, an integration host factor for HTLV-1 and HTLV-2 was identified as cellular serine/threonine protein phosphatase 2A (PP2A). HTLV-1 does not appear to integrate into preferred sites of the cellular genome in HTLV-1-infected asymptomatic individuals. However, in patients with HTLV-1-associated diseases such as HAM/TSP, clinical disease diagnosis has been correlated with proviral integration into transcriptionally active regions of the cellular genome. Integration site selectivity studies with HTLV-2 have not been performed to date.

Viral Gene Transcription

From the integrated provirus, viral mRNAs are produced to generate an array of HTLV proteins. Transcription is driven from the 5' LTR which contains the promoter/enhancer, while the polyadenylation signal is found in the 3' LTR. The first mRNA produced after integration is the doubly-spliced mRNA encoding Tax and Rex. The HTLV-encoded transactivating transcriptional regulatory protein, Tax, functions to enhance viral transcription. Thus, generating a positive transcriptional feedback loop. Briefly, Tax recruits CREB to a set of conserved 21-base pair repeats in the U3 region of the LTR known as Tax-responsive elements (TREs). The TREs consist of 3 domains; A, B, and C. The A and C domains are C/G rich, while the B domain is similar to the cAMP response element (CRE). Tax does not directly bind DNA, but instead facilitates the binding of CREB to the TREs. CREB then recruits cofactors p300 and CBP to the promoter. Together, this protein complex promotes transcription of viral gene products. The anti-sense derived protein HBZ inhibits Tax-1-mediated viral transcription. HBZ interacts with CBP and p300 and competes these essential cofactors away from the Tax-1 complex, thereby decreasing Tax-driven viral transcription. The HTLV-2 equivalent to HBZ is termed APH-2. APH-2, unlike HBZ, lacks a bZIP (basic leucine zipper) domain but still interacts with CREB to repress Tax-2-mediated viral transcription.

The other gene product encoded from the doubly-spliced mRNA encoding Tax, is Rex. Rex is involved in post-transcriptional gene regulation. During normal cellular gene transcription, intron-containing mRNAs are retained in the nucleus until they are fully spliced or degraded. HTLV expresses several gene products from unspliced mRNA (genomic and gag/pol) and singly spliced mRNA (env) containing introns. Therefore, the virus must overcome the default cellular splicing pathway to efficiently replicate. Rex exports unspliced or incompletely spliced viral mRNA from the nucleus to the cytoplasm by binding target viral mRNA within the Rex response element (RxRE). Previous studies revealed Rex-1, Rex-2 and their RxREs are structurally similar and functionally interchangeable. The HTLV-1 RxRE is a 205 nt sequence located in the U3 and R regions of the 3' LTR, while the HTLV-2 RxRE is 226 nt in length and is located in the R/U5 region of the 3' LTR. The secondary structure of RxRE contains four stem loops, which are critical for Rex-dependent mRNA export. Also contained within the RxRE is a cis-acting repressive sequence (CRS). Experimental data suggests the CRS retains the unspliced mRNA in the nucleus to presumably ensure sufficient amounts of Tax and accumulation of Rex substrate. Once Rex binds to the RxRE, this overcomes the CRS effects and Rex can then export the mRNA to the cytoplasm. Rex also interacts with the nuclear export factor CRM1. The Rex/CRM1/RxRE containing mRNA can then be shuttled out of the nucleus through the nuclear pore. Once released from the viral mRNAs, Rex binds importin β and is transported back to the nucleus. Several studies have identified phosphorylation sites at serine and threonine residues within Rex. Many of these modified sites are necessary to regulate Rex function.

The viral accessory proteins p30 (HTLV-1) and p28 (HTLV-2) also play a role in post-transcriptional regulation by inhibiting the export of doubly spliced *tax/rex* mRNAs. HTLV-2 p28 is functionally homologous to HTLV-1 p30. p30/p28 bind to and retain *tax/rex* mRNA in the nucleus, thereby decreasing the export and expression of completely spliced *tax/rex* mRNA. Consequently, this RNA redistribution promotes the expression of other viral gene products.

Viral Translation

Translation of accessory and regulatory HTLV gene products proceed in a similar manner to cellular translation. However, the structural and enzymatic gene products follow an alternative translation pathway. For every ten Gag proteins that are translated, a Gag-Pro-Pol fusion protein is produced. This fusion protein is generated via two successive ribosomal frameshifts. A -1 frameshift occurs near the 3' end of Gag open reading frame (ORF) that allows the ribosome to continue into Pro. A second -1 frameshift occurs near the end of Pro ORF that allows the ribosome to continue reading through Pol. This alternative translation helps ensure the correct ratios of gene products are generated.

Assembly, Budding, and Maturation

Like other retroviruses, the unspliced mRNA serves as both genomic RNA and mRNA that encodes Gag (MA-CA-NC) and Gag-Pol (MA-CA-NC-RT-IN) polypeptides. Gag (p55) precursor protein is myristoylated and targeted to the cellular membrane. Gag is cleaved by the viral protease into matrix (MA, p19), capsid (CA, p24), and nucleocapsid (NC, p15) proteins. This cleavage of Gag is essential for the assembly of infectious virions. The MA protein contains amino acid residues necessary for Gag homodimerization, the amino terminus of the CA protein is responsible for particle formation, and the NC protein coats the viral gRNA. After assembly, the virus particles bud from the surface of the host cell and mature into infectious virions via proteolytic cleavage.

Epidemiology

HTLV-1 is estimated to infect 5–10 million people worldwide with areas of endemic infection. This data is based strictly on reliable epidemiologic data of people in HTLV-1 endemic areas. Epidemiologic researchers do not yet have reliable data from some of the more densely populated areas of the world; therefore, this infection rate may be much higher. Regions with prevalent infection include the Southwestern region of the Japanese archipelago, parts of the Caribbean, foci in South America (Columbia, French Guyana, Brazil), areas of intertropical Africa, the middle East (Mashhad region in Iran), clusters in Australo-Melanesia, and Romania. The uneven worldwide distribution of HTLV-1 ranges from infection rates of 1%–2% in high endemic regions, although these rates can reach 20%–40% in people older than 50 in some particular areas. Recently, a hospital and community-based cohort study in Central Australia reported HTLV-1 infection rates of 33.6% in certain communities. The infection rates increased with age, reaching over 48% in men older than 50. This uneven geographical and ethnic divide is not well understood. It is hypothesized to be due to a founder effect in some cases, followed by the persistence of a high viral transmission rate due to various local environmental and cultural customs/traditions. The prevalence of HTLV-1 infection in an endemic area has been linked to age (especially in women), sex, and economic level.

HTLV-1 is genetically stable, which is highly unusual for a retrovirus given the nature of reverse transcriptase. The genetic stability is mostly due to viral amplification that occurs via clonal expansion of infected cells vs. viral replication and subsequent new infections (which is the case for HIV-1). As a result, HTLV-1 has low sequence variation that can be utilized to track the migrations of infected populations worldwide. There are 4 major geographic subtypes, or genotypes, of HTLV-1: Cosmopolitan a-subtype, the Central African b-subtype, the new Central African d-subtype, and the Australo-Melanesian c-subtype. A small number of rare subtypes are also found in Central Africa (e, f, and g). The most widespread subtype is the Cosmopolitan a-subtype and it has very little sequence variation in the envelope gene and the LTR region. There is no strong evidence to suggest specific mutations or genotypes are associated with the development of HTLV-1-associated diseases.

HTLV-2 is estimated to infect 800,000 individuals worldwide based on healthy blood donor rates of infection. The largest region of infection is the United States at 400,000–500,000 infected people, followed by Brazil at 200,000–250,000 infections, and finally 20,000–40,000 people in Europe. High rates of HTLV-2 infection have been reported in Amerindian populations and are widespread in injection drug users (IDU). Overall, the global burden of HTLV-2 infection is 6–12 fold lower than HTLV-1.

Transmission

HTLV infection of new cells is heavily dependent upon cell-to-cell transmission (as opposed to cell-free transmission). Therefore, new infections require the transfer of virus-infected cells and transmission primarily occurs by 3 routes: mother-to-child, sexual intercourse, and contaminated blood products. Mother-to-child transmission via breastfeeding occurs in 10%–25% of breast fed children born to HTLV-1 infected mothers. A high proviral load in the mother's milk or blood cells, high antibody titers in the mother's serum, and duration of breastfeeding (>6 months) are all increased risk factors for mother-to-child transmission. Vertical mother-to-child transmission due to transplacental transmission or during delivery is more rare (<3% risk). Sexual transmission of HTLV is usually more prevalent in male-to-female transmission than female-to-male, but this is not exclusive. On-going sexual transmission of HTLV-1 is prevalent in central Australia, Japan, and South America. The exposure to HTLV-infected blood products by blood transfusion, organ donation, and sharing of blood-contaminated needles can lead to viral infection with variable success (12%–44%) in blood recipients. Although HTLV is endemic in many regions around the world, many countries still do not screen their blood supply or organ transplants for HTLV. HTLV-2 is mainly transmitted via contaminated blood products by IDU, and to a lower degree sexual and mother-to-child transmission.

Clinical Features

HTLV-1 is the causative infectious agent of several human diseases and conditions, including ATL, HAM/TSP, HTLV-1-associated uveitis, and various dermatological conditions. Hereafter, HTLV-1-associated disease will be focused on ATL and HAM/TSP. These diseases develop in roughly 5%–10% of infected individuals and generally only after a lengthy clinical latency period of several decades. Although very similar to HTLV-1 in nucleotide sequence and coding similarities, the clinical data associating HTLV-2 with malignancy have not been conclusive to date.

ATL

HTLV-1 was first identified as the cause of an aggressive non-Hodgkin's peripheral T-cell malignancy called adult T-cell leukemia/lymphoma (ATL). ATL is most commonly a CD4 + T-cell disease; however there have been rare examples of CD8 + disease origin. The clinical features that classify ATL include lymphadenopathy, skin lesions, increased abnormal lymphocytes, frequent blood and bone marrow involvement, hypercalcemia, and lytic bone lesions. The abnormal lymphocytes detected in ATL patients are called 'flower' cells. These cells have multilobular nuclei with homogeneous and condensed chromatin, small or absent nucleoli, and an agranular and basophilic cytoplasm. Prospective studies have indicated HTLV-1 infected patients with a higher proviral load (> 4 copies/100 peripheral blood mononuclear cells) are at a higher risk for developing ATL.

ATL is a heterogeneous disease and is thus divided into four clinical subtypes based on the 1980s Shimoyama classification: acute, lymphoma, smoldering, and chronic. Patients who develop smoldering or chronic ATL present with a skin rash and minimal blood involvement. These two subtypes are generally less aggressive and are collectively known as indolent ATL. The 'aggressive' ATL subtypes, acute and lymphoma are characterized by a large tumor burden, lymph node and blood involvement, and hypercalcemia. Patients who develop aggressive ATL typically display > 5% of their total T-cells as flower cells. These subtypes are also accompanied by multiorgan failure and frequent opportunistic infections due to intrinsic T-cell immunodeficiency. Aggressive ATL subtypes also frequently display multidrug and chemotherapy resistance, leading to their poor prognosis. In a retrospective analysis of 1594 patients diagnosed and treated for ATL in Japan between 2000 and 2009, the median survival rate and four year survival rate (OS4) were as follows: acute ATL, 8 month median survival, 11% OS4; lymphoma ATL, 11 month median survival; 16% OS4; chronic ATL, 32 month median survival; 36% OS4; and smoldering ATL, 55 month median survival; 52% OS4.

HAM/TSP

HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) is a slow progressing, chronic neurodegenerative disease caused by HTLV-1. HAM/TSP is characterized by painful stiffness and weakness of the legs (lower limb spasticity), bowel and bladder disturbances, and slow but steady progression over several years. These symptoms develop due to inflammation or swelling that occurs in the spinal cord. HTLV-1-infected cells infiltrate the central nervous system and cause persistent immune activation against proliferating HTLV-1-infected T-cells. HAM/TSP patients frequently display high proviral loads and high levels of HTLV-1-specific antibodies in the cerebrospinal fluid (CSF). Many symptoms of HAM/TSP are also found in patients with MS. In fact, the clinical course of HAM/TSP and MS are similar, with steady worsening of neurological function without any distinct relapse or period of remission.

HTLV-2-Associated Symptoms

HTLV-2 has the ability to transform T-cells in culture, but clinical data demonstrating HTLV-2 association with malignancy have not been conclusive. There are however several reports throughout the 1990s correlating HTLV-2 infection with a few cases of neurological disease. HTLV-2-associated neurological symptoms include leg weakness, reflex abnormalities, sensory abnormalities, urinary tract abnormalities, sphincter disturbance, spastic paraparesis, and ataxia spastic. Interestingly, these symptoms appear after several decades of infection and closely mimic the neurological symptoms of HTLV-1-associated HAM/TSP. Ultimately, a clear link between HTLV-2 and neurological disease has been difficult to determine, mainly due to lower worldwide prevalence and concomitant infections with HIV-1, hepatitis B and C, and human herpes virus.

Pathogenesis

HTLV-1 and HTLV-2 are transmitted efficiently in a cell-to-cell mediated fashion. Both viruses immortalize T-cells *in vitro* and *in vivo* and induce T-cell proliferation. However, these two viruses have distinct oncogenic properties. A large body of work was initially focused on the Tax proteins and the role they play in triggering transforming activity. Loss of Tax by mutation or deletion prevents HTLV-mediated cellular transformation during *in vitro* cell culture assays. Reciprocally, Tax over-expression, in the absence of other viral factors, can induce cellular transformation in cell culture models. The ability of Tax to contribute to cellular transformation is dependent on the interaction of Tax with several important signaling pathways involved in cell survival, proliferation, and genomic stability. Recently, the anti-sense derived HBZ and APH-2 proteins have garnered more attention. A brief description and summary of the Tax and anti-sense derived proteins for both HTLV-1 and HTLV-2 follows.

Tax

Tax-1 and Tax-2 share approximately 85% amino acid similarity. Tax-1 is 353 amino acids long, while Tax-2 is 356 amino acids in length. Tax-1 contains a CREB-binding domain and nuclear localization signal at its N-terminus. The central portion of Tax-1 contains two leucine zipper-like motifs (LZ) at amino acids 116–145 and 225–232. These motifs are required for DNA interactions and for protein dimerization. The amino terminal LZ is important for Tax activation of the classical NF- κ B pathway, while the carboxyl terminal LZ is important for alternative NF- κ B pathway activation. The C-terminus of Tax-1 contains an ATF/CREB

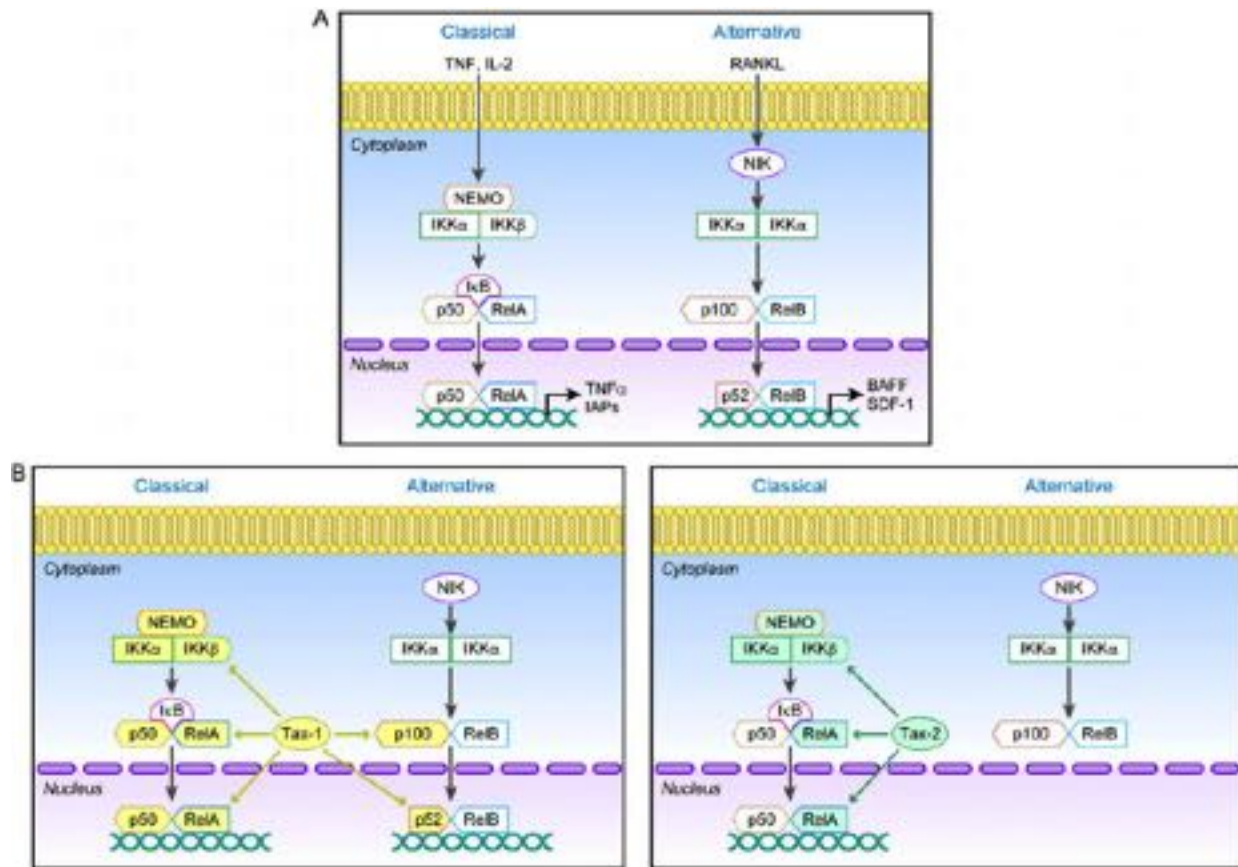


Fig. 2 Tax proteins activate the NF-κB pathways. (A) Brief schematic outlining the classical (left) and alternative (right) NF-κB pathways. (B) Tax-1 (left panel) activates both the classical and alternative NF-κB pathways at various steps within the pathway (shown in yellow). Tax-2 (right panel) only activates the classical NF-κB pathway (shown in green).

activation domain, p300/CBP binding domain, and PDZ binding domain. Tax-2 is less well characterized compared to Tax-1. Like Tax-1, Tax-2 also contains a CREB-binding domain at its N-terminus. However, instead of a NLS, Tax-2 contains a nuclear localization determinant (NLD) within the N-terminus. Also similar to Tax-1, Tax-2 contains an ATF/CREB activating domain at its C-terminus. Unlike Tax-1, Tax-2 lacks a PDZ domain. While both Tax-1 and Tax-2 possess transcriptional transactivating activity and transforming capacity, Tax-2 functions less efficiently than Tax-1.

One of the major cellular pathways affected by Tax is the NF-κB pathway, a pathway of transcription factors involved in the regulation of cell survival and proliferation. Briefly, the NF-κB family of transcription factors are controlled by inhibitory IκB proteins which sequester NF-κB/Rel in the cytoplasm as multiprotein complexes (Fig. 2(A)). In the classical NF-κB pathway, upon activation by extracellular stimuli such as TNF-α and IL-2, the IκB proteins become serine phosphorylated by the IKK complex composed of NEMO, IKKα, and IKKβ. This targets the IκB proteins for ubiquitination and degradation by the proteasome, freeing the p50/RelA heterodimer to translocate to the nucleus where it can activate transcription of NF-κB responsive genes. The classical pathway is quickly activated in response to stimuli, with IκB degradation occurring within minutes of stimulus recognition. However, this pathway is also short lived and is quickly repressed when stimulus is withdrawn or lost. Tax activates the classical NF-κB pathway via its interaction with, and activation of, the IKK complex. Tax expression leads to constitutive activation of the IKK complex and thus continuous degradation of IκB protein. Tax maintains this activation via interaction with NEMO. Activation of the IKK complex is increased by ubiquitination of Tax, which promotes the translocation of Tax to the cytoplasm. Both Tax-1 and Tax-2 have been extensively shown to activate the classical NF-κB signaling pathways through various protein-protein interactions at various stages of the signaling pathway (Fig. 2(B)).

Tax-1 can also activate the alternative NF-κB pathway through its ability to induce the processing of p100 to p52. Briefly, the alternative NF-κB pathway is activated by stimulus, such as RANKL signaling (Fig. 2(A)). Upon stimulation of the cell, an IKKα homodimer is phosphorylated and activated by the NF-κB inducing kinase (NIK). Activation of the IKKα homodimer leads to the phosphorylation of p100. p100 is usually dimerized with RelB. Upon phosphorylation and subsequent ubiquitination, p100 is cleaved into the transcriptionally active p52 protein. This cleavage removes an inhibitory ankyrin repeat and allows the p52/RelB dimer to translocate into the nucleus and drive expression of target genes. The alternative pathway is slowly activated due to the kinetics involved in processing p100 to p52. Complete activation can take up to several hours. However, the alternative pathway is

also long-lived with cellular transcription remaining active after stimulus is lost. Interestingly, only Tax-1, but not Tax-2, is able to efficiently activate the alternative pathway (Fig. 2(B)).

Although NF- κ B activation should lead to cell proliferation, HTLV-infected cells cease to proliferate one cell division cycle after infection. Tax activation of NF- κ B signaling increases the cyclin-dependent kinase inhibitors, p21^{cip1/waf1} and p27^{kip1} and thus become arrested in senescence. Therefore, in order for Tax to promote oncogenesis of the infected cell, the cellular senescence response would have to be circumvented.

In addition to being a potent activator of NF- κ B signaling, Tax-1 has also been shown to inactivate cell cycle checkpoints, alter DNA repair pathways, induce micronuclei formation, and disrupt spindle assembly and cytokinesis. This helps establish an environment of genetic instability that contributes to oncogenesis. Tax-1 also uses multiple mechanisms to promote cellular proliferation, including activation of the Akt pathway. In contrast, Tax-2 lacks micronuclei inductive ability. By comparison, Tax-1 has a more robust ability to modulate cell cycle progression and apoptosis compared to Tax-2.

The Tax-1 protein is essential for cellular transformation and contributes to genomic instability within the infected cell. HTLV-1-associated diseases progress very slowly and are characterized by a long clinical latency period. Thus, with each round of DNA replication, Tax-1 further increases genomic DSBs (double-strand breaks) and accumulation of mutations. Tax-1 is crucial for its ability to drive viral replication and immortalize T-cells. However, Tax-1 is frequently lost from ATL tumor cells through epigenetic silencing, 5' LTR deletion, or abortive protein mutations within the Tax-1 gene. This suggests other HTLV-1 proteins contribute to the transformation process.

Hbz

The other HTLV-1 product that is involved in oncogenic transformation and the pathogenic process is the anti-sense derived *Hbz* gene. Unlike Tax, the *Hbz* transcript is detected in all ATL cells. *Hbz* transcription initiates in the mostly epigenetically unmodified 3' LTR and is regulated by viral CREs and several SP1 binding sites. There are both spliced and unspliced *Hbz* transcript variants and the proteins encoded by these transcripts have nearly identical amino acid sequence and function. However, spliced *Hbz* is more abundant in infected cells and therefore has been well-studied. Subsequent discussion of *Hbz* refers to the major spliced variant.

Hbz mRNA possesses proliferative potential in addition to serving as a protein-coding template. *Hbz* mRNA alone is able to support proliferation of an IL-2-dependent T-cell line when cultured at suboptimal concentrations of IL-2. However, the precise mechanism is unclear. Recently, *Hbz* mRNA was shown to inhibit apoptosis in mouse CD4 + T-cells and increase the transcription of the anti-apoptotic gene *survivin*. Thus, *Hbz* mRNA serves both as a proliferative and anti-apoptotic factor.

HBZ is a 206 amino acid nuclear protein comprised of three functional domains: an N-terminal activation domain, a central basic region, and a C-terminal bZIP domain (Fig. 3(A)). The activation domain contains two well-studied LXXLL-like motifs which enable HBZ to bind the KIX domain of CBP/p300. The LXXLL motifs are also required for HBZ to activate TGF- β signaling. Through interactions with CBP/p300, HBZ is able to sequester these factors away from Tax-1 and repress Tax-1-mediated LTR-activation. The bZIP domain allows HBZ to heterodimerize with cellular bZIP proteins of the AP1 superfamily (CREB2, c-Jun, JunB, JunD, CREB, MafB, and ATF3) and affects their binding to DNA recognition sites. HBZ mutant proviral clones are able to immortalize T-cells *in vitro*, but a rabbit model of infection revealed that HBZ is required for efficient HTLV-1 infection and persistence *in vivo*.

Viral persistence and immune evasion are key determinants for a successful HTLV lifecycle. To ensure survival, the virus employs a period of latency, or restricted viral gene expression. HBZ helps to maintain a persistent latent infection in cells through several mechanisms (Fig. 3(C)). In contrast to Tax-1, HBZ does not affect the alternative NF- κ B pathway and actually inhibits the classical NF- κ B pathway. HBZ interacts with p65 (RelA) and induces its proteasomal degradation. Therefore, HBZ prevents NF- κ B hyperactivation (induced by Tax) and therefore prevents cellular senescence.

Like Tax-1, HBZ also affects genomic integrity through the induction of DSBs. These HBZ-induced DSBs are dependent on the miR17 and miR21 HBZ-inducible miRNAs. miR17 and miR21 target and suppress expression of the gene that encodes hSSB2, a single-stranded DNA-binding protein that prevents genomic instability. HBZ also enhances hTERT expression through its JunD interaction. hTERT, the enzyme responsible for maintenance of telomere length, is often upregulated in HTLV-1-infected cells. Upregulated hTERT translates to increased cellular proliferation and further accumulation of deleterious mutations.

HBZ is able to usurp cellular apoptosis through various pathways. HBZ suppresses expression of Bim, a pro-apoptotic gene by repressing its transactivation. HBZ also activates the mTOR pathway via interactions with GADD34. This pathway regulates cellular metabolism and promotes cellular proliferation in response to environmental stimuli.

HBZ promotes proliferation of infected T-cells in a number of ways. HBZ interacts with ATF3, thus preventing ATF3 activation of p53 tumor-suppressor signaling. HBZ also interacts with JunD and enhances its transcriptional activation. Loss of JunD inhibits HBZ-induced cellular proliferation, highlighting the significant role JunD plays in HTLV-1 biology. HBZ upregulates transcription of the noncanonical Wnt ligand Wnt5a while suppressing the canonical Wnt pathway. Wnt5a enhances proliferation and migration of ATL cells. Finally, IRF-1, a known tumor suppressor, is negatively regulated by HBZ.

HTLV-2 also encodes an anti-sense transcript named APH-2 (anti-sense protein of HTLV-2) (Fig. 3(B)). Like HBZ, the APH-2 mRNA is transcribed from the 3' LTR, spliced, and polyadenylated. APH-2 mRNA is expressed in a majority of HTLV-2-infected patient PBMCs (peripheral blood mononuclear cells). However, loss of APH-2 has no effect on *in vitro* cell immortalization and contrary to HBZ, its loss enhances *in vivo* viral replication in rabbits.

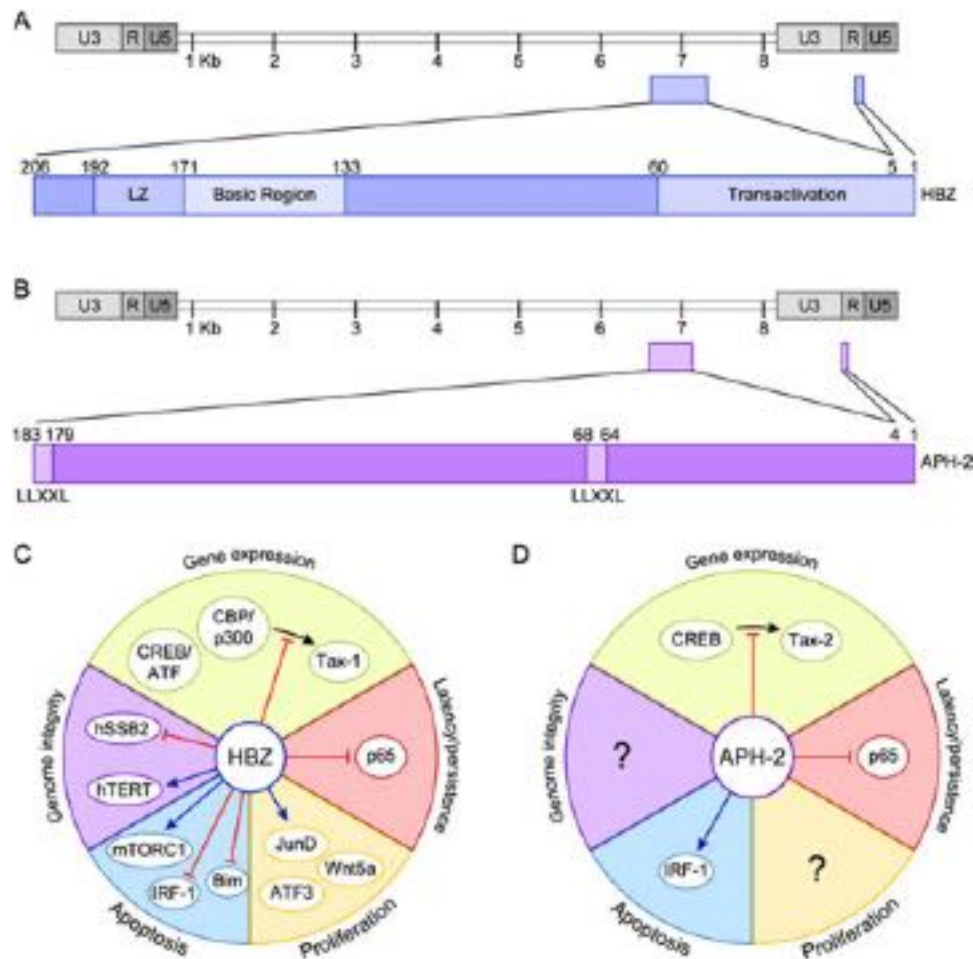


Fig. 3 The anti-sense derived proteins of HTLV. (A) HBZ is encoded from the antisense strand of the HTLV-1 proviral genome. Domain structure of HBZ is depicted. (B) APH-2 is encoded from the antisense strand of the HTLV-2 proviral genome. Domain structure of APH-2 is depicted. Brief summary of the various cellular pathways and processes affected by (C) HBZ and (D) APH-2 proteins.

APH-2 protein shares less than 30% homology to HBZ. However, the two proteins share several functional similarities. Although it lacks a classical bZIP domain, APH-2 still interacts with CREB (through an LXXLL motif and LXXLL-like motif) and thus inhibits Tax-2-mediated viral transcription. APH-2 inhibits p53 (RelA), like HTLV-1 HBZ, but not as strongly. Conversely, unlike HBZ, APH-2 enhances transcriptional activity of IRF-1 – suggesting HTLV-2 cells are more prone to IRF-1-mediated apoptosis. This finding is supported by data that shows APH-2 deficient proviral clones replicate more efficiently compared to WT in a rabbit model of *in vivo* infection (Fig. 3(D)). Taken together, this data suggests APH-2 does not affect proliferation or infectivity to the extent of its counterpart HBZ.

Viral Clonality

DNA analyses of asymptomatic HTLV-1-infected individuals show little variation in HTLV-1 proviral DNA sequence – unlike HIV-1. This suggests HTLV replication occurs not by error-prone RT, but through mitotic cell division. Therefore, HTLV exists predominantly in a latent form where the genome is maintained by mitotic host cell division, followed by occasional re-activation and de novo infection.

Clonality is the number and abundance of HTLV-1-infected T-cell clones in an individual. Initial research estimated that in asymptomatic HTLV-1-infected individuals, there were as many as 100 clones in circulation. However, more recent high throughput techniques have estimated the number of clones in circulation in a single host is > 20,000. Interestingly, 91% of ATL cases have a predominant, malignant T-cell clone with a single proviral genome. Clonality studies of non-malignant HTLV-1-carriers who developed ATL suggest a malignant clone does not necessarily develop from the largest pre-existing infected T-cell clone, but often from a clone of previously low abundance. These recent advances in clonality studies have also shown there are several less abundant clones which underlie the largest malignant clone. This data reaffirms the risk of malignant transformation depends on the longevity of the cell and subsequent number of cell divisions that enable accumulation of DNA damage (Fig. 4). Unlike other retroviruses,

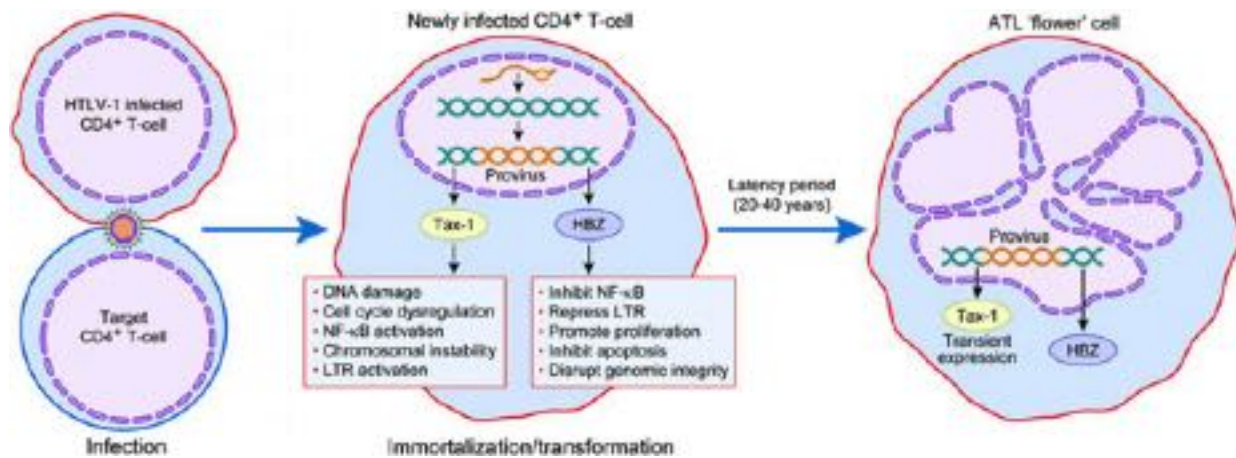


Fig. 4 HTLV-1 immortalization and transformation of infected cells. HTLV-1 infection (Left panel) occurs primarily in a cell-to-cell mediated fashion. The HTLV-1 viral genome is integrated into the host chromosome in newly infected CD4⁺ T-cells (Center panel). Expression of the viral genes Tax-1 and HBZ from the proviral genome affects a variety of important processes within the cell. These cellular changes lead to eventual immortalization and transformation. The development of disease (ATL) occurs after a prolonged clinical latency period of several decades (right panel). ATL cells consistently express HBZ and have variable intermittent expression of Tax-1.

HTLV-1 is a trans-activating retrovirus. The virus carries no dominant viral oncogene and does not cause cis-activation of a cellular oncogene. Instead, viral proteins (Tax-1, HBZ) activate transformation with very slow tumor formation kinetics.

In Vivo Animal Models

Several different animal models have been established and utilized since the discovery of HTLV almost 40 years ago. New Zealand white rabbits have been used to study both early immune response and viral persistence. Inoculation of rabbits with lethally irradiated HTLV producer cells results in seroconversion and T-cell infection, but the absence of disease. Mutational analysis of various viral genes and elements using proviral clones can be utilized with this system. Infection of rabbits with cloned mutant viruses allows investigators to address how specific alterations affect viral replication, proviral load, and alter persistence in the presence of a functional immune system. In addition to rabbits, rats and non-human primates have proved useful for studying HTLV-1 viral transmission, viral infection, neurologic HTLV-1 disease, and vaccine efficacy studies.

The mouse model is the most popular animal model used in HTLV-1 studies. Transgenic mice expressing Tax-1 or HBZ have been used to study the role of these gene products during disease development. NSG mice (NOD.Cg-PrkdcSCIDIL2rgtm1Wjl/SzJ) can be used in a tumor transplant model for HTLV-1. Following transplantation of HTLV-1-transformed cell lines, the mice will develop tumors. Interestingly, only certain HTLV-1 infected cell lines are tumorigenic in NSG mice. Immortalized cell lines generated in culture from a proviral clone are non-tumorigenic compared to cell lines generated in culture from co-cultured ATL patient leukemic cells. This confirms that although HTLV-1 can immortalize T-cells *in vitro*, there are additional molecular events that occur, contributing to the development of tumorigenesis. The NSG tumor transplant model is useful for testing therapeutic reagents and investigating the importance of individual viral genes in the context of the intact viral genome. The recent advancement of a 'humanized' mouse offers a model system for ATL development in a human microenvironment. However, while the reconstituted immune cells are phenotypically similar to human, they lack full functionality.

Immune Response

The host immune response – both antibody and cell-mediated – against HTLV occurs within a few weeks of exposure to the virus. The first two months after infection, Gag antibodies predominate the body. Shortly after this, antibodies to Env appear. In approximately 50% of infected individuals there is a detectable Tax antibody response. The infected host also develops a strong cytotoxic T-cell response. The majority of cytotoxic T lymphocytes (CTLs) are against Tax, with a smaller percentage directed against Gag, Env, and other nonstructural gene products. The immune response helps control, but does not eliminate the virus. Little to no cell-free HTLV virions are found in the plasma. Therefore, proviral load is used to quantify viral infection or viral burden. Proviral load is the number of proviral copies per set number of PBMCs. A high proviral load is associated with both inflammatory and malignant disease caused by HTLV-1. Although antibody titer can vary between infected patients, it directly correlates with the proviral load. The immune response plays an important role in reducing proviral load and thereby protecting the infected person against inflammatory disease. Most of what is currently known concerning antibody responses against HTLVs

have been carried out with HTLV-1. The immune response to HTLV-2 is less well understood. What is known is that high frequencies of CTLs specific for Tax-2 are often found in patients.

Several studies have suggested the immune response can determine the course of HTLV-1-associated disease. The development of ATL is observed more frequently in patients who acquired HTLV-1 early in life (e.g., via breastfeeding). ATL is characterized by a late onset and it is believed the initial reduced anti-viral immune response favors viral persistence. HAM/TSP patients typically acquire their HTLV-1 infection via infected blood products. This elicits a vigorous immune response and these patients have 10–100 fold higher proviral load than asymptomatic patients. The humoral and cell-mediated CD4 + and CD8 + T-cell responses are highest in HAM/TSP compared to ATL and asymptomatic carriers. Indeed, a vigorous expansion of CD8 + T-cells, the presence of Tax specific CTLs in the cerebral spinal fluid (CSF), and high levels of anti-HTLV-1 antibodies in the serum and CSF characterize HAM/TSP. There has been a suggested role of Tax-specific CTLs in the cellular destruction and inflammation in the central nervous system and spinal cord of HAM/TSP patients.

Diagnosis and Prevention

The diagnosis of HTLV infection is based on the detection of antibodies by ELISA, or enzyme-linked immunosorbent assay. However, this type of assay is unable to distinguish HTLV-1 from HTLV-2. A follow-up confirmation test by immunoblot and/or PCR is required. The PCR analysis will also provide the patient's proviral load. This type of test is not convenient for routine screening and makes it difficult to make a more accurate approximation of HTLV prevalence worldwide. Many HTLV-1 infected individuals are asymptomatic for many years and have the capacity to transmit the virus.

There is currently no cure, preventative vaccine, or prophylactic vaccine for HTLVs; therefore infected individuals remain capable of passing infectious virus for their lifetime. However, since HTLVs are blood borne pathogens, prevention strategies that are effective for HIV and Hepatitis B & C also reduce the risk of HTLV transmission. Blood and organ donor testing have been shown to decrease HTLV-1 prevalence and HTLV-1 diseases in endemic regions. In Japan, HTLV-1 prevalence rates decreased from 20% in 1987 to 2.5% in 2017 after the implementation of universal antenatal care and avoidance of breastfeeding in HTLV-1 carriers.

Treatment

HTLV-1 and HTLV-2 are lifelong infections and there is currently no cure or recommended treatment strategy for asymptomatic HTLV-infected patients. The possibility of a vaccine against HTLV-1 exists because the virus displays low antigenic variability and natural immunity occurs in humans. Experimental vaccination with attenuated vaccinia virus-derived candidates encoding Env provided partial protection from HTLV-1 challenge in squirrel monkeys, offering promising potential for vaccine possibilities. Overall, a variety of vaccine approaches have been attempted for the prevention of HTLV-1 infection, including DNA, live vectors, synthetic peptides, or combinations of these approaches. Unfortunately, despite progress in HTLV research, there is currently no licensed vaccine against HTLV-1. Therefore, HTLV-infected individuals should avoid blood and organ donations, practice safe sex, discourage needle sharing, and avoid breastfeeding in efforts to prevent the spread of the virus.

Treatment strategies for HTLV-1-associated diseases vary depending on the severity of symptoms and diagnostic criteria presented. There is no uniform standard of care for ATL. Several treatment options include 'watchful waiting', chemotherapy, antiviral therapy, allogeneic hematopoietic stem cell transplantation, and molecular targeted therapies. Patients diagnosed with the less severe, indolent forms of ATL (chronic and smoldering subtypes) generally display few symptoms. These patients are often advised to utilize 'watchful waiting' until disease progression to a more aggressive phenotype. When following this standard of care, patients with chronic or smoldering types of ATL have a median survival of only four years. When treated with an antiviral therapy regimen of zidovudine (AZT) and interferon-alpha (IFN α), patients with indolent forms of ATL have a better prognosis. AZT/IFN α therapy of more aggressive ATL subtypes has also been shown to improve patient survival. Patients with acute ATL, treated with first-line AZT/IFN α therapy, had an 82% 5-year survival rate. Cell culture studies suggest AZT/IFN α treatment causes HTLV-1-infected cells to undergo apoptosis.

Conventional chemotherapy has been used for the treatment of aggressive ATL subtypes with varying degrees of success. Although chemotherapy regimens continue to be prescribed as a treatment for ATL, overall patient survival has been minimally improved. ATL tumors are frequently chemotherapy resistant and contribute to patient relapse after therapy. Chemotherapy resistance arises from p53 mutations – a common feature of ATL cells – and immune deficiency due to early stages of disease.

Several alternative therapies against ATL have been reported, including allogeneic stem cell transplantation, proteasome inhibitors, HDAC inhibitors, JAK 1/2 inhibitors, arsenic trioxide, monoclonal antibodies directed against ATL/T-cell surface molecules, anti-folates, nucleotide phosphorylase inhibitors, antiangiogenic therapy, and immune-modulatory drugs. Several of these therapies are part of ongoing or proposed clinical trials and will not be described in more detail due to space constraints.

Many of the treatment options available for HAM/TSP patients are directed at the treatment of symptoms associated with the disease. Baclofen or tizanidine is used to treat muscle stiffness, urinary symptoms are treated with medications which reduce the activity of the bladder muscle, constipation is treated with increased fiber and laxatives, non-steroidal anti-inflammatory medication is used for nerve pain and inflammation of the spinal cord can be controlled by corticosteroids. Unfortunately, no therapy has been shown to alter the long-term disability associated with HAM/TSP.

Further Reading

- Ciminale, V., Rende, F., Bertazzoni, U., Romanelli, M.G., 2014. HTLV-1 and HTLV-2: highly similar viruses with distinct oncogenic properties. *Frontiers in Microbiology* 5, 398.
- Doueiri, R., Green, P.L., 2012. Human T-cell leukemia virus type 2 (HTLV-2) biology and pathogenesis. In: Robertson, E.S. (Ed.), *Cancer Associated Viruses*. Springer.
- Enose-Akahata, Y., Vellucci, A., Jacobson, S., 2017. Role of HTLV-1 Tax and HBZ in the pathogenesis of HAM/TSP. *Frontiers in Microbiology* 8, 2563.
- Fochi, S., Mutascio, S., Bertazzoni, U., Zipeto, D., Romanelli, M.G., 2018. HTLV deregulation of the NF-kappaB pathway: An update on tax and antisense proteins role. *Frontiers in Microbiology* 9, 285.
- Fujii, M., Matsuoka, M., 2013. Human T-cell leukemia virus types 1 and 2. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*. sixth Ed., Lippincott Williams & Wilkins.
- Giam, C.-Z., 2012. HTLV-1 and oncogenesis. In: Robertson, E.S. (Ed.), *Cancer Associated Viruses*. Springer.
- Giam, C.Z., Semmes, O.J., 2016. HTLV-1 infection and adult T-cell leukemia/lymphoma – A tale of two proteins: Tax and HBZ. *Viruses* 8.
- Higuchi, M., Fujii, M., 2009. Distinct functions of HTLV-1 Tax1 from HTLV-2 Tax2 contribute key roles to viral pathogenesis. *Retrovirology* 6, 117.
- Ilinskaya, A., Heidecker, G., Jones, K., 2010. Interaction between the HTLV-1 envelope and cellular proteins: Impact on virus infection and restriction. *Future Medicinal Chemistry* 2, 1651–1668.
- Mahieux, R., Gessain, A., 2011. HTLV-3/STLV-3 and HTLV-4 viruses: Discovery, epidemiology, serology and molecular aspects. *Viruses* 3, 1074–1090.
- Matsuoka, M., Green, P.L., 2009. The HBZ gene, a key player in HTLV-1 pathogenesis. *Retrovirology* 6, 71.
- Nasr, R., Marçais, A., Hermine, O., Bazarbachi, A., 2017. Overview of targeted therapies for adult T-cell leukemia/lymphoma. *Methods in Molecular Biology* 1582, 197–216.
- Panfil, A.R., Al-Saleem, J.J., Green, P.L., 2013. Animal models utilized in HTLV-1 research. *Virology: Research and Treatment* 4, 49–59.
- Yamagishi, M., Fujikawa, D., Watanabe, T., Uchamaru, K., 2018. HTLV-1-mediated epigenetic pathway to adult T-cell leukemia-lymphoma. *Frontiers in Microbiology* 9, 1686.

Relevant Websites

- <http://hamtsp-net.com/english/about/index.html>
HAM-net HTLV-1 Associated Myelopathy (HAM) registration.
- <https://www.htlvaware.com>
HTLV AWARE
HTLV Home.
- [https://www.nationalmssociety.org/What-is-MS/Related-Conditions/HTLV-I-Associated-Myelopathy-\(HAM\)](https://www.nationalmssociety.org/What-is-MS/Related-Conditions/HTLV-I-Associated-Myelopathy-(HAM))
HTLV-I Associated Myelopathy (HAM)
National MS Society.
- <https://rarediseases.org/rare-diseases/htlv-type-i-and-type-ii/>
HTLV Type I and Type II
National Organization for Rare.
- <https://htlv.net>
International Retrovirology Association
HTLV.
- <https://www.cdc.gov/mmwr/preview/mmwrhtml/00021234.htm>
Recommendations for Counseling Persons Infected.

Infectious Bursal Disease Virus (*Birnaviridae*)

Daral J Jackwood, The Ohio State University/OARDC, Wooster, OH, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

cvIBDV Clinical virulent pathotype of IBDV
hvVP2 Hypervariable region of the protein VP2

scIBDV Subclinical pathotype of IBDV
VP Viral protein
vvIBDV Very virulent pathotype of IBDV

Glossary

Bursa of Fabricius The organ in a chicken responsible for the maturation of B-lymphocytes.

Icosahedron A polyhedron shape with 20 faces.

Introduction and Classification

Infectious bursal disease virus (IBDV), a member of *Birnaviridae*, genus *Avibirnavirus*, causes infectious bursal disease (IBD) also known as Gumboro disease, as it was first described in Gumboro, Delaware, USA. IBDV causes an immunosuppressive disease, through infection of immature B-lymphocytes in the bursa of Fabricius of chickens. The immune suppression caused by IBDV can lead to secondary infections from opportunistic pathogens that exacerbate the disease. IBD is often not recognized in chickens until these secondary infections cause an immune suppression related disease such as gangrenous dermatitis or respiratory disease. In addition, the immune suppression can cause clinical reactions from live-attenuated vaccines, and reduced efficacy of vaccination programs for other diseases. Economic losses result from morbidity, mortality, poor feed efficiency, slower growth, uneven bird weights within a flock, longer times to market and increased condemnations at the processing plant.

Isolates of IBDV have been given names based on a variety of criteria including clinical observations, the person that first described a given isolate, antigenic differences, alpha-numeric designations and geographic locations. Establishing a nomenclature system based on the antigenic and pathogenic types of IBDV is difficult because standard assays for determining these characteristics are not possible on a worldwide basis. As a result, nucleotide sequence comparisons and phylogenetic analysis of the VP2 gene have been used to categorize the viruses into genogroups. These procedures are universally available and have identified at least seven genogroups to date (Fig. 1). In addition, a new system for naming historic and future IBDV isolates by their genogroup classification has been also proposed.

There are two major serotypes of IBDV. Serotype 1 viruses were the first strains isolated from chickens. Serotype 2 viruses were originally isolated from turkeys but have also been found in other avian species including penguins. No disease has been attributed to the serotype 2 viruses in any avian species. Nucleotide sequencing of the VP2 gene and virus-neutralization assays can be used to distinguish serotype 2 from serotype 1 viruses.

Antigenic drift and genetic recombination have contributed to the development of several antigenic subtypes of serotype 1 viruses. The first IBDV strains identified in Gumboro Delaware are known as classic viruses. The term variant was used to describe those viruses that were able to replicate in the presence of antibodies to the classic strains. Although classic and variant are still used to describe the antigenic diversity among serotype 1 IBDV, there are numerous antigenic variations of these viruses. In some cases, partial cross-neutralization and protection is observed between antigenic strains of IBDV. However, there are other strains that have substantial antigenic drift resulting in little or no cross-neutralization or protection between the strains.

The antigenicity of serotype 1 IBDV is dictated by the amino acids in the hypervariable region of VP2 and the VP2 epitopes are dependent on the proper folding of this protein. Neutralizing monoclonal antibodies bind to the P domain amino acids in VP2 and amino acid substitution mutations in the P_{BC} and P_{DE} domains were found to contribute to the antigenic drift of IBDV strains.

Virion Structure and Genome

Infectious bursal disease virus is a non-enveloped icosahedron (Fig. 2). The genome of IBDV consists of two segments of double-stranded RNA. The RNA-dependent RNA polymerase, VP1 is encoded by the smaller genome segment B. The larger segment A encodes a polyprotein that is self-cleaved by the viral encoded protease VP4. Cleavage of this polyprotein yields two capsid proteins, pVP2 and VP3, and the VP4 protease (Fig. 3). The pVP2 protein is further processed at the COOH terminus to yield the mature capsid protein VP2. VP2 is comprised of base (B), shell (S) and projection (P) domains. The P domain contains four loop structures designated P_{BC}, P_{DE}, P_{FG}, and P_{HI} that make up the surface of the virion. Trimers of VP2 comprise the structural units of

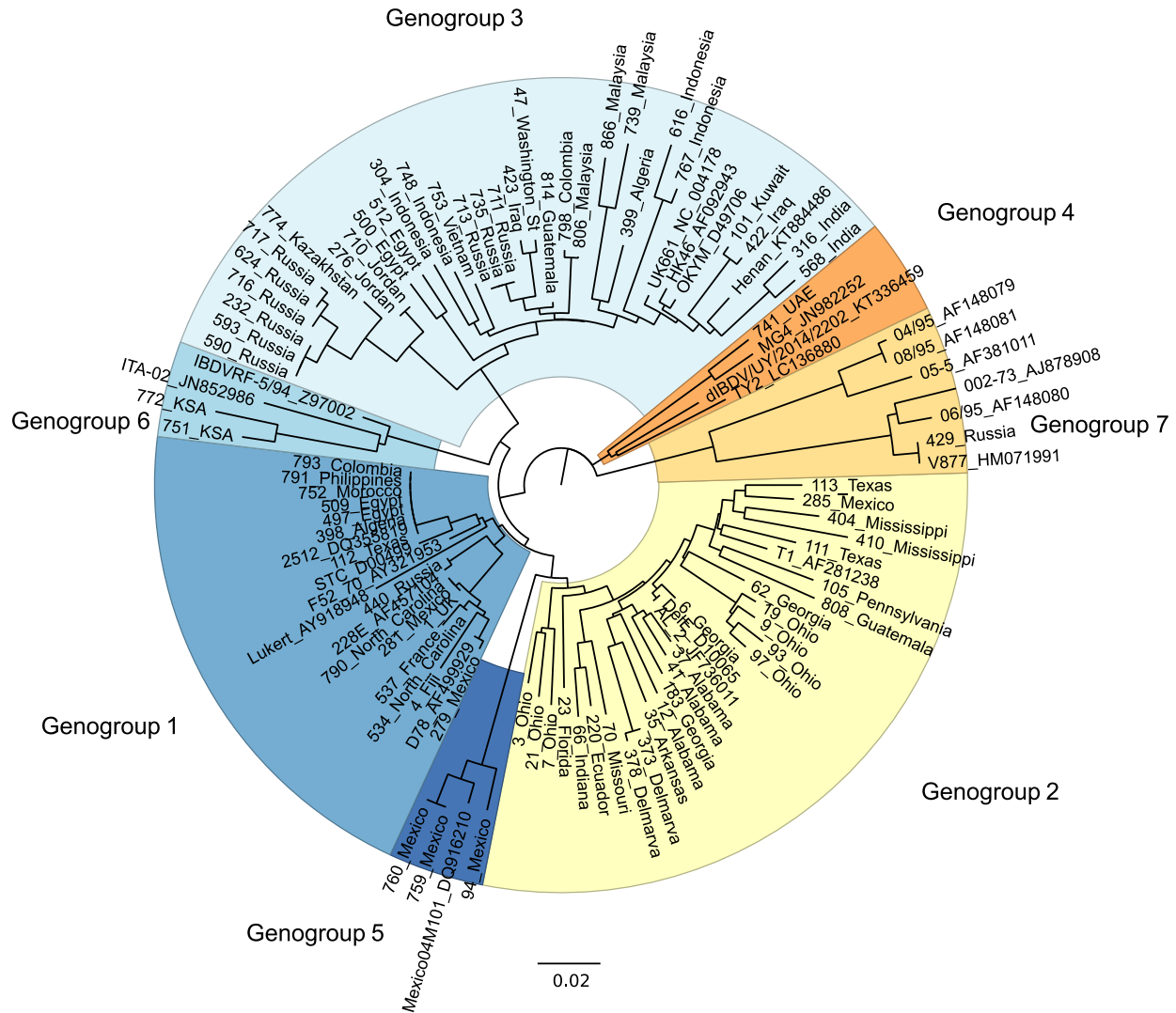


Fig. 1 IBDV genogroups. Phylogenetic analysis of 105 IBDV nucleotide sequences of the hypervariable region of VP2 identifies 7 genogroups. The analysis was conducted using neighbor-joining method with 1000 bootstrap replicates. The evolutionary distances were computed using the maximum composite likelihood method. Reference strains are identified by GenBank accession numbers. From: Michel, L.O., Jackwood, D.J., 2017. Classification of infectious bursal disease virus into genogroups. Archives of Virology. doi:[10.1007/s00705-017-3500-4](https://doi.org/10.1007/s00705-017-3500-4).

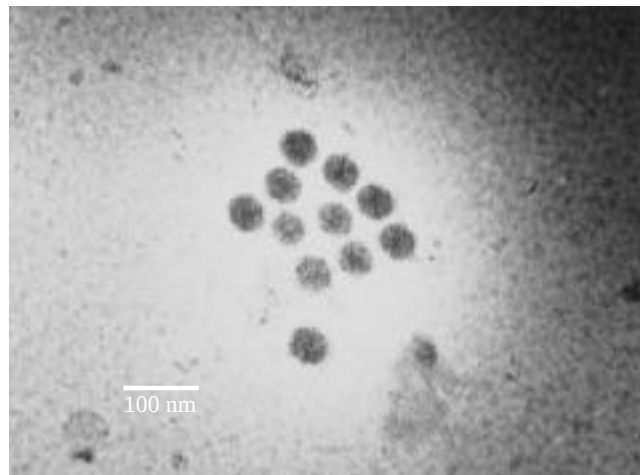


Fig. 2 Electron micrograph of IBDV.

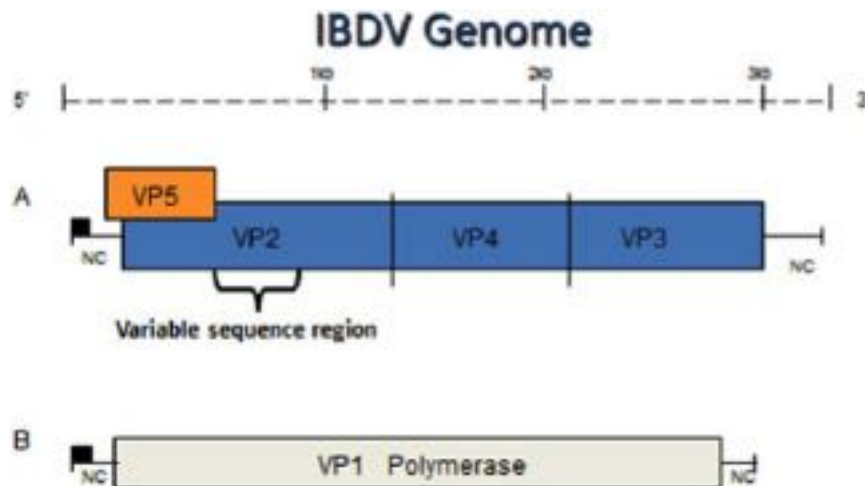


Fig. 3 Schematic presentation of genome segments A and B of IBDV and the proteins they encode.

the T3 viral capsid and the P domain of VP2 is responsible for inducing neutralizing antibodies and a protective immune response in chickens. The larger genome segment also has an overlapping reading frame with the polyprotein that encodes a non-structural protein VP5 which is involved in apoptosis.

Pathogenesis and Clinical Features

Infectious bursal disease is highly contagious. Pathogenicity ranges from attenuated to highly virulent and strains are typically placed in one of three pathogenic groups; subclinical (sc)IBDV, clinical virulent (cv)IBDV and very virulent (vv)IBDV. Immune suppression can be caused by viruses in all three of the pathogenic groups. Infections with the cvIBDV may be subclinical in chicks before 3 weeks of age due to only partial protection by maternal immunity to the infecting IBDV strain. Chickens are most susceptible to clinical disease at 3–6 weeks of age when maternal antibodies have waned. The vvIBDV usually cause clinical disease in young chickens and have caused severe clinical infections in Leghorn chickens up to 18 weeks of age.

Early IBDV infections, even though they may be subclinical, can cause severe, long-lasting immunosuppression. Following an early IBDV infection, immature B-lymphocytes in the bursa of Fabricius are destroyed before they can populate secondary lymphoid organs such as the spleen and Harderian gland. Consequently, chickens immunosuppressed by early IBDV infections do not respond well to vaccination and are predisposed to infections by opportunistic viruses and bacteria. The IBDV induced immune suppression can exacerbate the severity of disease caused by other poultry pathogens such as Newcastle Disease virus, infectious bronchitis virus and avian influenza.

In clinical infections, the onset of clinical signs occurs after an incubation of 3–4 days. Chickens can exhibit severe prostration, incoordination, watery diarrhea, soiled vent feathers, vent picking, and inflammation of the cloaca. The bursa of Fabricius will become edematous, yellowish, and occasionally hemorrhagic. Congestion and hemorrhage of the bursa and skeletal muscles is sometimes seen in vvIBDV infected birds. Chickens that have recovered from IBDV infections have small bursas. Microscopic lesions in the bursa include inflammation, severe lymphocyte necrosis, atrophy of the follicles and follicular depletion of lymphocytes.

Diagnosis and Control

IBDV can cause morbidity, mortality and gross lesions in the bursa are usually indicative of an infection. However, it can also be difficult to diagnose IBD because some strains cause a sub-clinical disease. In these cases, the recognition that IBD has infected a flock often occurs when opportunistic microorganisms cause secondary infections resulting from the IBDV induced immune suppression. Gross lesions in the bursa can signal an IBD outbreak but some strains cause atrophy of the bursa without gross lesions. The diagnosis of IBDV using traditional virus isolation methods can yield variable results. Many of the pathogenic strains do not replicate in cell culture unless they are passed multiple times which can lead to genetic and phenotypic changes. Passage in 9 day old chicken embryos has been used successfully to isolate IBDV but there is the risk that an IBDV vaccine or subpopulation of the field virus will be selected using this method. Most diagnostic laboratories use reverse transcriptase-polymerase chain reaction (RT-PCR) followed by nucleotide sequencing of the VP2 gene to detect and identify IBDV strains. As described above, genogroups of the virus have been defined and are based on the sequence of the hypervariable VP2 sequence region (Fig. 1).

There is no treatment for IBD. The virus is fastidious and thus disinfection of contaminated farms after depopulation has had limited success. The first line of defense is achieving high levels of maternal antibody to IBDV in young chicks. High titers of anti-IBDV antibodies in breeder hens are transferred through the egg yolk to the young chicks. It is important to use IBDV vaccines that are very similar or an antigenic match to the pathogenic IBDV endemic to the region where the chicks will be housed.

Vaccinating broiler chicks or pullets with live-attenuated vaccines is practiced to protect them after maternal immunity wanes. It is important to administer these vaccines at the proper time so residual maternal antibodies to IBDV do not interfere with their replication. To help solve this problem, recombinant herpesvirus of turkeys (HVT) vectors containing the VP2 gene of IBDV (HVT-IBD vaccines) have been produced and are used to vaccinate broilers and pullets *in ovo* or at one-day of age. These live HVT-IBD vector vaccines are not neutralized by maternal antibodies to IBDV and thus can be administered when maternal immunity is high. The immunity induced by HVT-IBD vector vaccines can be slow to develop in young chicks. Therefore, it is important to make sure the maternal immunity adequately protects the birds until active immunity to the IBDV strain is achieved by these vaccines.

Further Reading

- Becht, H., 1980. Infectious bursal disease virus. *Current Topics in Microbiology and Immunology* 90, 107–121.
- Cosgrove, A.S., 1962. An apparently new disease of chickens-avian nephrosis. *Avian Diseases* 6, 385–389.
- Coulibaly, F., Chevalier, C., Delmas, B., Rey, F.A., 2010. Crystal structure of an Aquabirnavirus particle: Insights into antigenic diversity and virulence determinism. *Journal of Virology* 84, 1792–1799.
- Jackwood, D.J., 2017. Recent advances in vaccine research against economically important viral diseases of food animals: Infectious bursal disease virus. *Vaccines Veterinary Microbiology* 206, 121–125. (special issue).
- Jackwood, D.J., 2019. Overview of infectious bursal disease in poultry. In: Line, S., Moses, M.A. (Eds.), *The Merck Veterinary Manual*. Whitehouse Station, NJ: Merck & Co., Inc.
- Michel, L.O., Jackwood, D.J., 2017. Classification of infectious bursal disease virus into genogroups. *Archives of Virology*. doi:10.1007/s00705-017-3500-4.
- van den Berg, T.P., 2000. Acute infectious bursal disease in poultry: A review. *Avian Pathology: Journal of the W.V.P.A* 29, 175–194.
- van den Berg, T.P., Morales, D., Etteradossi, N., *et al.*, 2004. Assessment of genetic, antigenic and pathotypic criteria for the characterization of IBDV strains. *Avian Pathology: Journal of the W.V.P.A* 33, 470–476.

Relevant Websites

- <https://www.oie.int/doc/ged/d9315.pdf>
Infectious Bursal Disease (Gumboro disease) – OIE.
- <https://www.merckvetmanual.com/poultry/infectious-bursal-disease/overview-of-infectious-bursal-disease-in-poultry>
Overview of Infectious Bursal Disease in Poultry.
- <https://www.youtube.com/watch?v=Op4z86V3l8>, <https://www.youtube.com/watch?v=j0iyVM15Z-Q>
Poultry Health Today.

Infectious Pancreatic Necrosis Virus (*Birnaviridae*)

Øystein Evensen, Norwegian University of Life Sciences, Oslo, Norway

© 2021 Elsevier Ltd. All rights reserved.

This is an update of Ø. Evensen, N. Santi, Infectious Pancreatic Necrosis Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00772-X](https://doi.org/10.1016/B978-012374410-4.00772-X).

Glossary

Fingerlings A life-cycle stage when young salmonids are one finger in length.

Fry Newly-spawned fish that has fully absorbed their yolk sack.

Milt The sperm from male fish.

Parr A young salmonid with parr-marks (dark blotches) on the sides, before migration to the sea.

QTL Quantitative trait loci.

Smolts Salmonid fish going through smoltification; physiological changes that will allow fish to change from life in freshwater to life in seawater. The smolt state follows the parr state.

Start-feeding The time when the fry take up feed for the first time.

Introduction

Infectious pancreatic necrosis (IPN) is an acute and contagious disease of juvenile and post smolt salmonids which include the genus *Salmo*, *Oncorhynchus*, and *Salvelinus*, typically occurring at time of start-feeding of fry and after sea transfer in post-smolt Atlantic salmon. Virus characteristics such as variations between virus serotypes and strains, and viral infection loads can also play an important role in determining mortality rates. The clinical manifestation of the disease has changed over the last two decades. While IPN outbreaks were frequently seen in Atlantic salmon farming at start-feeding and after sea transfer, both manifestations of disease have declined with the introduction of IPN-QTL fry in major salmon-producing countries, Norway, UK, and recently also in Chile. When occurring, outbreaks are seen during the first few weeks after start-feeding of fry. Disease outbreaks can also be seen in 10–20 g salmon in freshwater but are less frequent. In the mid-1980s the number of outbreaks in post-smolt Atlantic salmon after seawater transfer emerged as a significant disease problem. Mortality rates in freshwater in genetically susceptible fish vary considerably from negligible to almost 100%. Disease outbreaks in seawater typically result in 5%–10% cumulative mortality but can reach 70% at cage level in genetically susceptible fish. This variation in mortality has been ascribed to factors related to the host species (e.g., the age and/or genetic resistance of fish) and environmental stressors. IPN-QTL (homozygous) fish are in principle resistant to IPN.

Classification

Infectious pancreatic necrosis virus is the type species of the genus *Aquabirnavirus* within the family *Birnaviridae*. IPN in salmonid fish is characterized by necrosis of the exocrine pancreas, giving the name to the disease, and the name of the type species. IPN in Atlantic salmon also causes necrosis of the liver parenchyma. Aquabirnaviruses have been isolated from a number of freshwater and marine fish species as well as from marine invertebrates worldwide and are associated with a wide range of host pathologies. Aquatic birnaviruses are an antigenically diverse group of virus.

Serological classification based on neutralizing antibodies separates IPNV isolates into serogroups A and B. Serogroup A contains nine serotypes (A1–A9) comprising most of the isolates, whereas serogroup B so far consists of only one serotype. Serological classification has not been standardized and serotyping is today replaced by genetic classification. This includes phylogenetic analysis based on sequencing the VP2-coding region of segment A clusters IPNV isolates into six genogroups. Genogrouping based on 310 basepairs (bp) at the VP2/NS junction region yields seven genogroups with an additional genogroup that includes a Japanese aquabirnavirus isolates. (Fig. 1).

There are no agreed species demarcation criteria for the genus *Aquabirnavirus*, and this has led to difficulty in nomenclature. Distantly related strains may cause similar pathologies, whereas closely related strains may range from avirulent to very virulent. Closely related strains are found in both farmed fish and marine species in the same geographical region. Therefore, the species concept does not apply very well to members of the genus aquabirnaviruses. Another aspect contributing to this is the intrinsic variability of RNA viruses like the aquabirnaviruses, derived from the error-prone process of replication. For future characterization of aquabirnavirus isolates, it should be sufficient to classify them according to genus and genogroup.

Virion Structure and Genomic Organization

The IPNV virion is 60 nm in diameter and a single-shelled icosahedral structure with no envelope (Fig. 2). The virus genome consists of two segments of double-stranded RNA. For serotype Jasper, segment A comprises 3097 base pairs (bp) and the smaller segment B comprises 2784 bp.

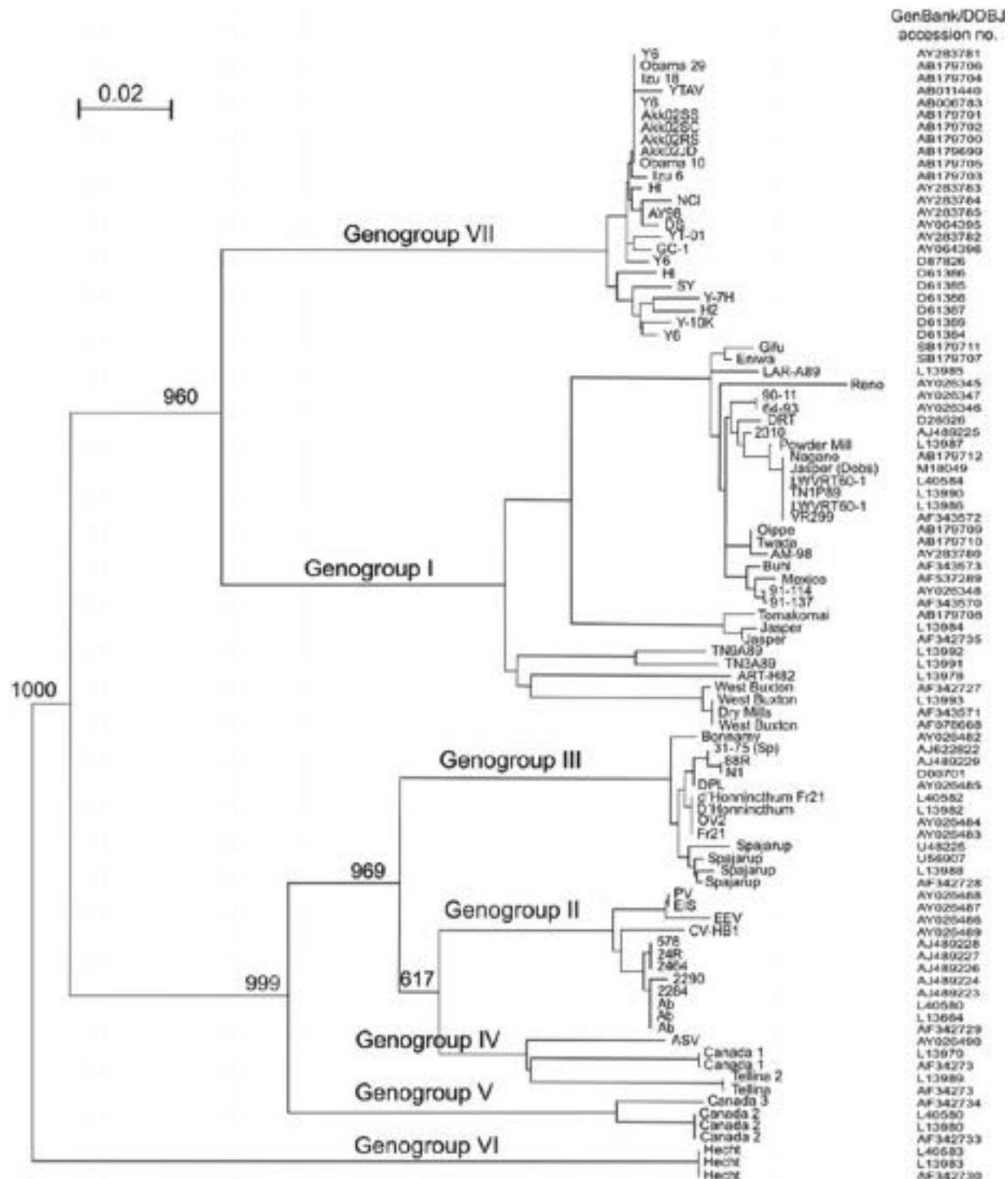


Fig. 1 Molecular phylogenetic tree based on nucleotide sequences of the VP2/NS junction among 93 worldwide isolates of IPNV and other aquabimaviruses (corresponding GenBank/DBJ accession numbers are shown). Bootstrap values from 100 replicates are shown at major nodes. Scale: 0.02 replacement nucleotides per site. Reproduced from Nishizawa, T., Kinoshita, S., Yoshimizu, M., 2005. An approach for genogrouping of Japanese isolates of aquabimaviruses in a new genogroup, VII, based on the VP2/NS junction region. *Journal of General Virology* 86, 1973–1978, with permission from Society for General Microbiology.

Genome segment A contains a large open reading frame (ORF) encoding a 106 kDa polypeptide (NH₂-VP2-protease (VP4)-VP3-COOH), which is co-translationally cleaved to generate pVP2 (the precursor of VP2), VP4 and VP3. The processing of the large polypeptide is controlled by VP4, which is a viral protease with a serine/lysine catalytic dyad. The cleavage sites have been identified at the pVP2-VP4 and VP4-VP3 junctions, and the structure of VP4 has been elucidated by X-ray crystallography. The

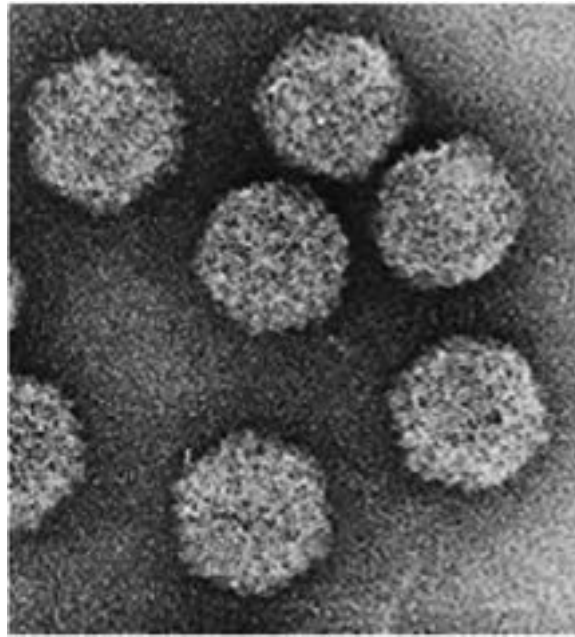


Fig. 2 Negative staining of IPNV serotype Sp grown on Chinook salmon embryonic (CHSE-214) cells and purified by density gradient centrifugation.

Table 1 Point mutations in VP2 result in a marked reduction to almost complete loss of virulence whereby 217 and 221 are key determinants. Substitution of threonine for proline at residue 217 (T217P) reduces the virulence, producing a mortality rate approx. 50% lower than the high-virulent variant (TAT). Similarly, single amino acid substitutions of alanine for threonine at residue 221 render highly virulent strains (TAT) almost avirulent. Passage in culture of high virulent strains is associated with motif changes in position 221 (A to T) for TAT and PAA variants, each of which results in avirulence

Amino acid residues of the VP2 protein			<i>In vivo</i> virulence characteristics
217	221	247	
T	A	T	Highly virulent
T	T	T	Avirulent
P	A	A	Moderately virulent
P	T	A	Avirulent

pVP2 precursor is processed further into the mature VP2, which is the major capsid protein. In addition, three small peptides are generated, that can be detected in viral particles.

VP2 spontaneously self-assembles into particles with a diameter of about 25 nm, which suggests that late maturation avoids premature assembly during particle morphogenesis. Some studies have indicated that elements of VP3 may be exposed on the surface of the virion, serotype specific and neutralizing antibodies are almost exclusively directed against continuous and discontinuous epitopes of VP2. Indeed, VP2 is the type-specific antigen and the neutralizing sites reside within the hypervariable region of VP2 ([Table 1](#)). The region covering amino acid residues 183–335 has been found by several authors to contain the neutralizing epitopes, although it has so far not been possible to identify their exact localization or sequence. This part of the VP2 protein carries several loops that stick out from the surface over this stretch ([Fig. 3](#)).

VP3 is an internal capsid protein associated with the viral genome. Studies using yeast two-hybrid systems in combination with co-immunoprecipitation, showed that VP3 binds to VP1. In addition, VP3 was shown to specifically bind to dsRNA in a sequence independent manner. The binding between VP3 and VP1 is not dependent on the presence of dsRNA.

Genome segment A also contains a small ORF that proceeds and partly overlaps the large ORF encoding the 106 kDa polyprotein. The smaller ORF encodes a highly basic, arginine-rich polypeptide of varying size. The protein has been detected in infected cells but has not, to date, been conclusively demonstrated to be present in purified viral particles. This apparently non-structural protein, designated VP5, is encoded by most but not all IPNV strains examined. In an Asian IPNV strain, an anti-apoptotic effect has been demonstrated for VP5 and this may aid viral replication in the initial phase of the infection. However, the amino acid sequence of VP5 is not well conserved and the anti-apoptotic effect does not occur in all IPNV strains encoding the polypeptide.

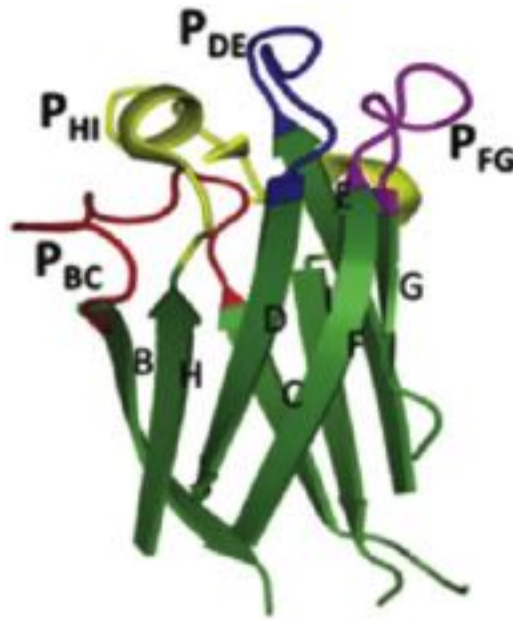


Fig. 3 Structural layout of the VP2 protein showing the P_{BC}, P_{DE}, P_{FG}, and the P_{HI} loops which represents amino acid residues, 210–247, 254–262, 282–288, and 314–334, respectively.

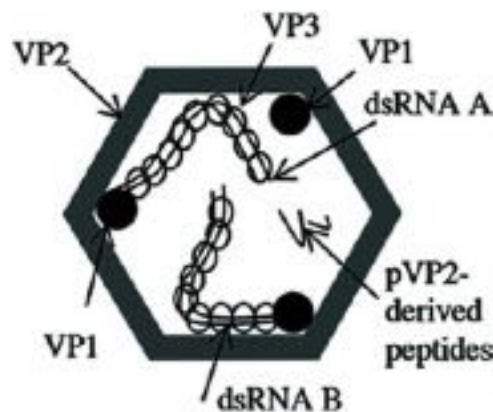


Fig. 4 Schematic presentation of the IPNV virus particle. VP2 is located on the surface, VP3 bound to viral RNA and VP1 bound to the dsRNA or free inside the virus particle.

Segment B encodes a 90 kDa protein (VP1) which functions as an RNA-dependent RNA polymerase (RdRp). This enzyme directs the transcription of non-polyadenylated mRNA from genome segments A and B. Within the viral particle, VP1 occurs as a free polypeptide and covalently linked to the 5' terminus of each genome segment (**Fig. 4**).

Life Cycle

IPNV replicates in the cytoplasm and a single cycle of replication takes 16–20 h. IPNV infects a wide range of cultured fish cells, including CHSE-214, RTG-2 and BF-2 cells. These cell lines are routinely used for laboratory propagation the virus. IPNV replicates well at temperatures ranging from 15–22°C, resulting in a characteristic cytopathic effect. Available information on attachment, penetration and uncoating is limited. Purified virus attaches to CHSE-214 cells by specific and non-specific binding to cell membrane components. After binding, IPNV can be observed in endosomes after 20–30 min, suggesting that the entry of IPNV is attained by endocytosis. The entry does not seem to be dependent on the acid pH of the endosomes, and the virion RdRp VP1 is active without proteolytic treatment of the virus, indicating that uncoating may not be a precondition for virus replication. VP1 can be guanylated *in vitro* whereupon it becomes a primer for *in vitro* RNA synthesis. Unlike other viral VPg polypeptides, VP1 can guanylate itself *in vitro* in a template-independent manner, and the guanylation site has been mapped to serine 163. Segment A

specific mRNA is synthesized in larger amounts (2–3 times) than segment B, reflecting the relative abundance of the viral proteins. The assembly of virions takes place in the cytoplasm of infected cells, and virus release occurs *via* cell destruction.

Apoptosis is induced in CHSE-214 cells following IPNV infection and the so-called “McKnight” cells associated with sloughing of mucosal cells in the intestines of juvenile salmonids suffering from IPN have morphological resemblance to apoptotic cells. Similar apoptotic bodies are also found in the liver of Atlantic salmon post-smolts suffering from IPN.

Epidemiology

It has long been known that survivors of an IPN outbreak become carriers and surviving fish continue to shed virus without any apparent clinical symptoms. In addition, sub-clinical infection is important for horizontal transmission of the virus. The efficiency by which the virus is spread in a population of susceptible individuals varies between strains and there are indications this might correlate with virulence characteristics whereby highly virulent variants are more efficient at spreading. Transmission of the virus to progeny has been shown to occur vertically *via* eggs in brook trout and rainbow trout. The virus is possibly present inside the eggs since there are indications that surface disinfection of the eggs is not fully effective in preventing the introduction of IPNV to hatcheries. One possible route of intra-ovum introduction is *via* the milt, by adsorption of virus to the sperm head.

Surveys for the prevalence of IPNV in susceptible strains of salmon, have concluded that most North Sea salmon farms harbor IPNV carrier fish, making selection for IPNV-free brood fish difficult. Once introduced to a hatchery, the virus is difficult to eliminate, as it is stable in water and relatively resistant to many disinfectants, and this has coined the term house-strains to these hatchery specific IPNV strains. Sea water outbreaks of IPN are the result of persistent infection obtained during the freshwater period.

Pathogenesis and Clinical Features

IPNV causes acute disease of juvenile salmonids and is the major pathogen of concern in brook trout (*Salvelinus fontinalis*), rainbow trout (*Oncorhynchus mykiss*), and Atlantic salmon (*Salmo salar* L.) hatcheries. Today, outbreaks also occur in fingerlings and parr later in the freshwater phase for Atlantic salmon, as well as in post-smolts after transfer to seawater.

Salmonid fish with acute IPNV infection show darkened skin, abnormal swimming behavior (whirling or swimming on the side), and a distended abdomen. Whitish threads of debris are often seen hanging from the anal orifice. At necropsy, common findings are ascites and hemorrhages in perivisceral adipose tissue. The liver is often pale macroscopically. Histopathologically, IPN is characterized by acute multifocal necrosis of the exocrine pancreas and liver necroses. In cases of more protracted disease, evidence of a cellular response can be observed. Multifocal hepatic necroses are typical and acute catarrhal enteritis with necrosis and sloughing of the intestinal epithelium have been reported. Viral antigen can be detected within lesions by immunohistochemistry.

A high proportion of fish surviving an IPN outbreak become persistently infected without apparent signs of disease. These sub-clinically infected individuals are of major importance as sources of horizontal transmission of the virus. The carrier state has no direct negative impact on infected individuals, although there are some indications of suppressed immune response. Recent studies have shown that there are differences in the ability of different strains of IPNV Sp serotype to establish persistent infections in Atlantic salmon fry. Carrier fish shed virus in their feces but titers fluctuate over time and increase during periods of stress. In addition, persistently infected brood fish can transfer IPNV to their progeny. Recently it has been documented that persistently infected fish carry IPNV from the hatchery to the grow-out farm. This was shown by sequence analysis of strains from different stages of the production.

Virus can be isolated from peripheral leukocytes a short time after experimental infection but not during chronic stages. The virus is located in head kidney leukocytes in fish that recover from an acute infection and most likely in macrophages or macrophage-like cells of the head kidney in persistently infected fish. It has also been shown that IPNV multiplies in adherent leukocytes isolated from carriers, although a lytic infection is not seen and, concordantly, virus is not found in the supernatant medium from cultures of isolated head kidney macrophages. Similarly, IPNV is also able to induce persistent, non-lytic infections in cell cultures in which a high proportion of the cells are infected but produce relatively low quantities of virus. Different strains of IPNV appear to vary in their ability to induce persistence in cell culture. Infected cells are resistant to superinfection by IPNV, both homologous or heterologous strains or serotypes, and cultures can be cleared of infection by several passages in the presence of virus neutralizing antiserum. The underlying mechanisms of the establishment and maintenance of IPNV persistence *in vivo* and *in vitro* are largely unknown.

Molecular Determinants of Virulence

The mortality rate during an IPN outbreak can vary, ascribed to host factors and traits related to the virus. Viral factors influencing mortality rates are dose of infection as well as virulence variations between serotype and strains. IPNV stains of different serotypes are highly heterogenic and the Sp serotype is considered to be more virulent than the Ab serotype. Within the Sp serotype different genotypes exhibit different virulence characteristics studied by the use of reverse genetics.

In vitro transcribed positive-sense cRNAs of IPNV are infectious when transfected into permissive cell cultures, allowing the recovery of genetically modified isolates. The reverse genetics technique is a valuable tool for studying the IPNV life cycle and the role

of individual viral proteins in the pathogenesis of infection. Experimental studies using such recombinant IPNV strains have demonstrated that VP5 expression is not required for efficient replication *in vitro* or *in vivo*, or for virulence or persistence of the virus. Reassortant viruses containing segment A from a high-virulence strain and segment B from a low-virulence strain, prepared by either reverse genetics or co-infection techniques, have also been used to demonstrate that segment B had no direct influence on IPNV virulence. For IPNV serotype Sp, it has been shown that certain amino acid sequence motifs in VP2 strongly correlate with virulence. Residues 217 and 221 appear to be critical determinants of virulence, and a genetic signatures involving these amino acids are strongly associated with avirulent, moderately and highly virulent strains, as shown in Table 1. Furthermore, serial passage of the Sp serotype in cell culture leads to the attenuation of virulence as a result of an A → T mutation at amino acid residue 221.

In summary, the neutralizing domain, the virulence signatures and the cell culture adaptation with subsequent loss of *in vivo* virulence, seem all to be located in the VP2 hypervariable region for IPNV. A limited number of serotypes and strains of IPNV have been characterized for their virulence signatures, and the same applies to identifying the immunogenic domain but indications are that the stretch between amino acids 183–337 of VP2 are important.

Control

Good management includes the environment of fish to minimize stress and to reduce the likelihood of transmission of the pathogen, particularly important at the time of first feeding of fry, when smolt are transferred to sea and during the first few months after sea water transfer.

Transfers of fish between farms are considered as one of the highest risks in the transmission of IPNV. Another factor is the movement of eyed eggs and fry between hatcheries, which is common practice in Norway and other salmon producing countries. Vertical transmission has not been conclusively demonstrated in Atlantic salmon but there is evidence from other salmonid species to indicate that vertical transmission is a strategy utilized by IPNV.

Diagnosis of IPN is made by combination of clinical signs, histopathological examination combined with immunohistochemistry for *in situ* detection of virus antigen in exocrine pancreas or liver. Combinations of clinical signs and real-time PCR methods are also frequently used to document IPN virus infection.

A number of studies have addressed the efficacy of IPN vaccines, most of these based on inactivated vaccine preparations delivered by injection (intraperitoneally). In addition, plasmid-based vaccines have also been tested under experimental conditions although with limited efficacy. Currently, the majority of vaccines used are inactivated whole virus vaccines (IWV) while vaccination with an *E. coli*-expressed VP2 recombinant protein is used less frequently. Reliable challenge models for IPN in (genetically) susceptible Atlantic salmon post-smolts have been established and high level of protection can be obtained with high antigen dose experimental vaccines (Fig. 5).

Plasmid vaccines expressing the structural proteins of IPNV and delivered by the oral route to rainbow trout, confer high level of protection against lethal challenge (> 80% relative percent survival). This is in contrast to the level of protection obtained in Atlantic salmon vaccinated with plasmid vaccines. Mode of delivery might play a role but has not been defined.

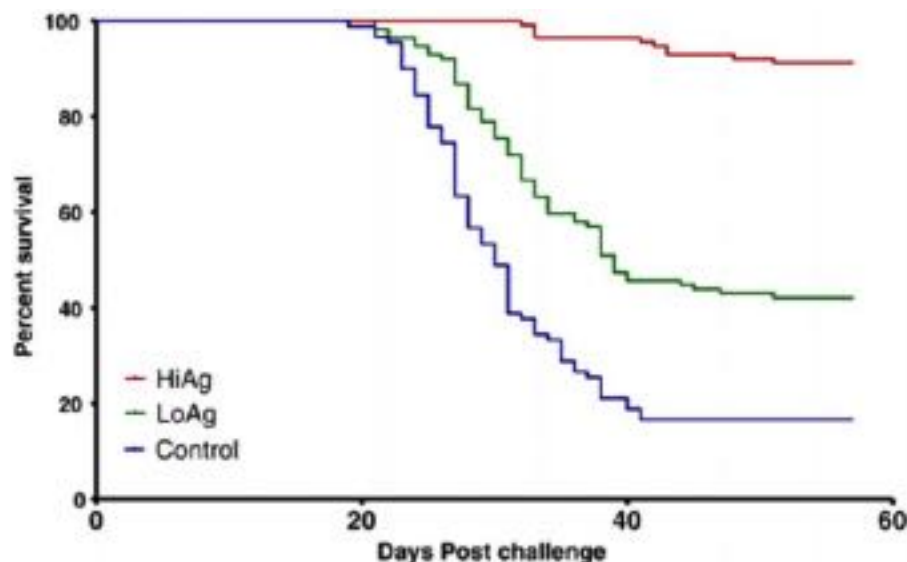


Fig. 5 Susceptible Atlantic salmon parr were vaccinated with two doses of inactivated whole virus vaccines (oil-adjuvanted) and challenged at 450 degree days. Protection against mortality coincided with antigen dose in the vaccine with close to 90% relative protection in the high dose group while less than 50% in the low-antigen dose group. Control mortality at 87%.

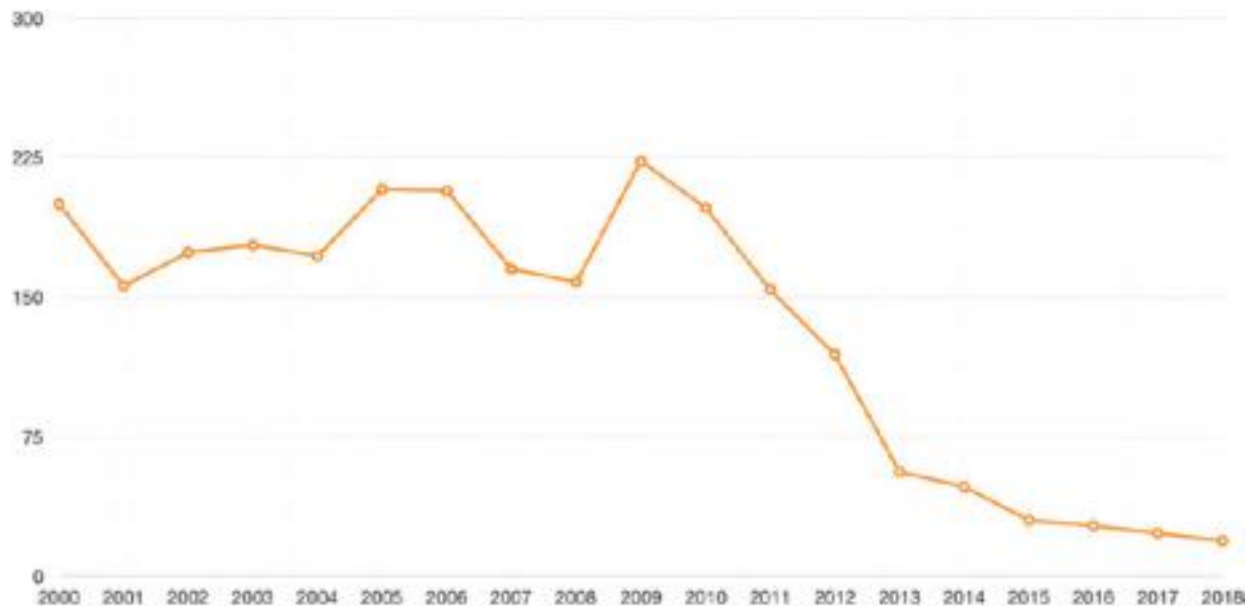


Fig. 6 Number of reported cases of IPN in Atlantic salmon and rainbow trout farming in Norway between 2000 and 2018. IPN-QTL fish were introduced in 2009–2010 and there has been a steady decline in number of reported outbreaks of IPN since.

In addition, a number of studies have been published over the last years related to genetic resistance to IPN virus infection in Atlantic salmon. Currently, only IPN-QTL eggs are available from breeding companies. Recent studies have shown that epithelial cadherin (*cdh1*) determines IPN virus infection in Atlantic salmon, and putatively a single nucleotide polymorphism within the full-length *cdh1* gene was identified as a causal factor but not conclusively determined. Infected salmon carrying the QQ_{QTL} genotype appear with very low levels of virus in target organs, like liver. There are indications that the *cdh1-1* gene product serves a function as receptor or co-receptor for IPNV, based partly on a co-localization of IPNV and *cdh1-1* proteins in infected liver tissue.

Since IPN-QTL eggs were introduced, the number of reported cases of IPN over the last 6–7 years in Norway have declined markedly (**Fig. 6**). In Norway, the number of reported cases peaked in 2009 (223 reported cases) and in 2017, 23 cases were reported. In addition to introduction of IPN-QTL fish, eradication programs in individual hatcheries have also contributed to the reduced number of reported cases.

Further Reading

- Ahmadvand, S., Soltani, M., Behdani, M., *et al.*, 2017. Oral DNA vaccines based on CS-TTP nanoparticles and alginate microparticles confer high protection against infectious pancreatic necrosis virus (IPNV) infection in trout. *Developmental and Comparative Immunology* 74, 178–189.
- Coulibaly, F., Chevalier, C., Gutsche, I., *et al.*, 2005. The birnavirus crystal structure reveals structural relationships among icosahedral viruses. *Cell* 120, 761–772.
- Dobos, P., 1995. The molecular biology of infectious pancreatic necrosis virus (IPNV). *Annual Review of Fish Diseases* 5, 25–54.
- Duncan, R., Nagy, E., Krell, P.J., Dobos, P., 1987. Synthesis of the infectious pancreatic necrosis virus polyprotein, detection of a virus-encoded protease, and fine structure mapping of genome segment A coding regions. *Journal of Virology* 61, 3655–3664.
- Hill, B.J., Way, K., 1995. Serological classification of infectious pancreatic necrosis (IPN) virus and other aquatic birnaviruses. *Annual Review of Fish Diseases* 5, 55–77.
- Hong, J.R., Wu, J.L., 2002. Induction of apoptotic death in cells *via* Bad gene expression by infectious pancreatic necrosis virus infection. *Cell Death and Differentiation* 9, 113–124.
- Houston, R.D., Haley, C.S., Hamilton, A., *et al.*, 2010. The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity* 105, 318–327.
- Kristoffersen, A.B., Devold, M., Aspehaug, V., *et al.*, 2018. Molecular tracing confirms that infection with infectious pancreatic necrosis virus follows the smolt from hatchery to grow-out farm. *Journal of Fish Diseases* 41, 1601–1607.
- Liao, L., Dobos, P., 1995. Mapping of a serotype specific epitope of the major capsid protein VP2 of infectious pancreatic necrosis virus. *Virology* 209, 684–687.
- Moen, T., Torgersen, J., Santi, N., *et al.*, 2015. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics* 200, 1313–1326.
- Mutoloki, S., Jøssund, T.B., Ritchie, G., Mungang'andu, H.M., Evensen, O., 2016. Infectious pancreatic necrosis virus causing and subclinical infections in Atlantic salmon have different genetic fingerprints. *Frontier Microbiol* 7, 1393.
- Nishizawa, T., Kinoshita, S., Yoshimizu, M., 2005. An approach for genogrouping of Japanese isolates of aquabirnaviruses in a new genogroup, VII, based on the VP2/NS junction region. *Journal of General Virology* 86, 1973–1978.
- Santi, N., Sandrø, A., Sindre, H., Song, H., *et al.*, 2005. Infectious pancreatic necrosis virus induces apoptosis *in vitro* and *in vivo* independent of VP5. *Virology* 342, 13–25.
- Santi, N., Song, H., Vakharia, V.N., Evensen, Ø., 2005. Infectious pancreatic necrosis virus VP5 is dispensable for virulence and persistence. *Journal of Virology* 79, 9206–9216.
- Song, H., Santi, N., Evensen, Ø., Vakharia, V.N., 2005. Molecular determinants of infectious pancreatic necrosis virus virulence and cell culture adaptation. *Journal of Virology* 79, 10289–10299.

Influenza A Viruses (*Orthomyxoviridae*)

Laura Kakkola, University of Turku, Turku, Finland

Niina Ikonen, Finnish Institute for Health and Welfare, Helsinki, Finland

Ilkka Julkunen, Institute of Biomedicine, University of Turku, Turku, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ARDS Acute respiratory distress syndrome

bp Base pair

cRNA Complementary ribonucleic acid

cRNP Complementary ribonucleoprotein

GISR Global Influenza Surveillance and Response System

HA Hemagglutinin

IAV Influenza A virus

IFITM Interferon-induced transmembrane protein

IFN Interferon

Ig Immunoglobulin

IL Interleukin

IRF Interferon regulatory factor

ISG Interferon stimulated gene

M1 Matrix protein 1

M2 Matrix protein 2

MDA5 Melanoma differentiation-associated protein 5

mRNA Messenger ribonucleic acid

NA Neuraminidase

NEP Nuclear export protein

Nf- κ B Nuclear factor kappa-light-chain-enhancer of activated B cells

NLRP Nucleotide-binding oligomerization domain, leucine rich repeat and pyrin domain containing

NOD Nucleotide-binding oligomerization domain

NP Nucleoprotein

NPA Nasopharyngeal aspirate

NS1 Non-structural protein 1

PA Polymerase acidic protein

PB1 Polymerase basic protein 1

PB2 Polymerase basic protein 2

PKR Protein kinase R

RdRp RNA-dependent RNA polymerase

RIG-I Retinoic acid-inducible gene 1

RNA Ribonucleic acid

RT-PCR Reverse transcription polymerase chain reaction

STAT Signal transducer and activator of transcription

TLR Toll-like receptor

TNF α Tumor necrosis factor alpha

vRNA Viral ribonucleic acid

vRNP Viral ribonucleoprotein

WHO World Health Organization

Glossary

Antigenic drift Mutations occurring to influenza genes during virus replication triggered by immunological pressure or RNA polymerase errors.

Antigenic shift Reassortment of viral segments when a cell or a host is simultaneously infected with two or more different IAV strains or subtypes.

ARDS Acute respiratory distress syndrome is a life-threatening condition caused by a breakage in endothelial-epithelial barrier due to IAV infection killing the epithelial cells, resulting in the production of inflammatory mediators leading to accumulation of leukocytes and fluid in the lungs leading to an acute respiratory failure.

Classification of Influenza A Viruses

Influenza A viruses (IAV) belong to the family *Orthomyxoviridae*, together with two other human-infecting influenza viruses, influenza B and C viruses. In addition to humans, influenza A viruses infect various animals, including ducks, chickens, pigs, whales, horses, seals, cats, and bats. Influenza A viruses are subtyped based on the hemagglutinin (HA) and neuraminidase (NA) proteins. To date, eighteen different HA proteins (H1-H18) and eleven NA proteins (N1-N11) have been identified.

IAV strains are named according to HA and NA subtype, the location or country of isolation and the year of isolation. For example A/Finland/9/2019 (H1N1) describes IAV strain isolated in 2019 in Finland with a strain number of 9, and the isolated virus is a subtype H1N1.

Virion Structure

Influenza virions are 100–140 nanometers in diameter. Virions are pleomorphic and mainly elliptical in shape. The enveloped IAV virion consists of cell-derived lipid bilayer, with NA and HA proteins sticking out from the membrane and forming trimers (HA) and tetramers (NA). HA is cleaved into two subunits, HA1 and HA2, that are attached by disulfide bridges. Matrix protein 2 (M2) forms ion channels through the lipid bilayer, and matrix protein 1 (M1) forms a lining underneath the lipid bilayer (**Fig. 1**).

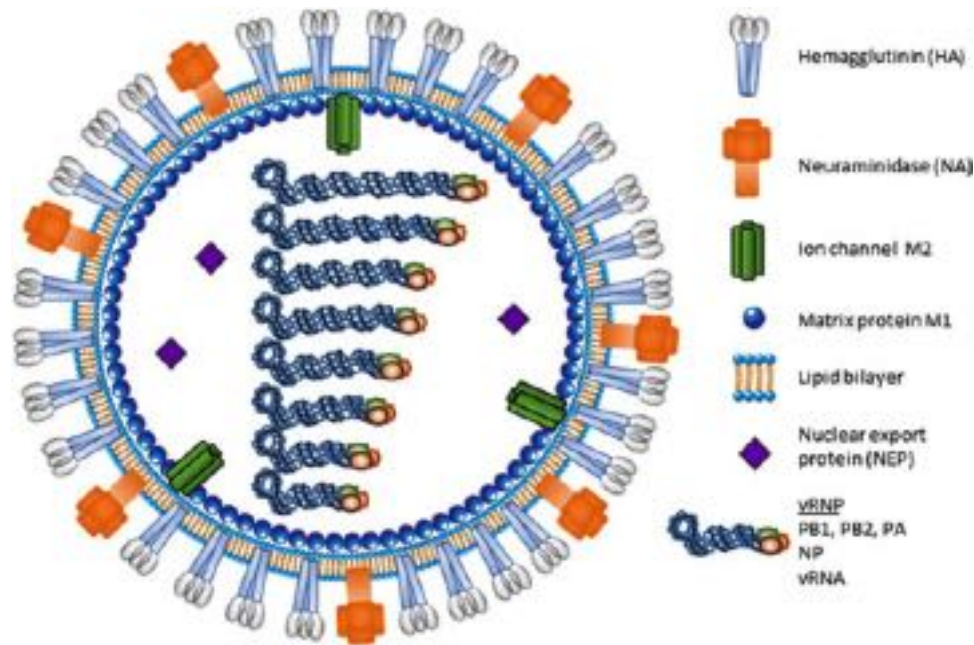


Fig. 1 Structure of influenza A virus. Influenza A virus particles are composed of a viral envelope with trimeric HA and tetrameric NA envelope glycoproteins, envelope-associated matrix protein and eight separate viral ribonucleotide protein complexes (vRNP) that include separate vRNA segments covered by nucleoprotein (NP) and polymerase proteins attached to viral 3' and 5' ends. M2 ion channel protein is embedded into the viral membrane and some nuclear export protein, NEP is also found in virus particles.

Nucleoproteins (NP) wrap eight genomic RNA segments, and the polymerase subunits PA, PB1 and PB2 that form the RNA-dependent RNA polymerase complex (RdRp) are attached to each segment. NP prevents vRNA segments from forming secondary structures and assists in virus replication. Nuclear export protein (NEP) is packaged within the virion.

Genome

Genomic Organization

Influenza genome is approximately 13,500 bp long. IAV genome consists of eight segments of negative sense single-stranded RNA. Segments are numbered from 1 to 8 based on the molecular weight. Each genomic segment encodes proteins (**Fig. 2** and **Table 1**): segment 1 encodes RNA polymerase subunit PB2, segment 2 encodes via different reading frames RNA polymerase subunit PB1 and PB1-F2 protein, segment 3 encodes RNA polymerase subunit PA and by a ribosomal frameshift PA-X protein, segment 4 encodes HA, segment 5 encodes NP, segment 6 encodes NA, segment 7 encodes by different reading frames and splicing matrix proteins M1 and M2, and segment 8 encodes by different reading frames and splicing non-structural proteins NS1 and NEP.

Each RNA segment has non-coding regions at both ends that vary in length, containing poly-adenylation and packaging signals. The ends of the segments are highly conserved and partially complementary. This allows the segments to form helical hairpin structures which serve as promoter areas for RNA replication and transcription, and onto which viral polymerase subunits are bound. Polymerase subunits PB1, PB2 and PA are bound to the ends as a single heterotrimeric complex. The entire viral genomic RNA segments are covered by NP and vRNAs with the associated proteins form viral ribonucleoprotein complexes (vRNPs). NP wrapped around the vRNA segments prevents RNA in forming secondary structures and opens the vRNA for viral polymerase-directed RNA synthesis.

Reassortment and Mutagenesis

Due to a segmented genome, influenza virions are capable of exchanging segments between virus strains. This genomic shuffling occurs when two or more different strains or subtypes infect the same cells in a host. This phenomenon is called antigenic shift.

Due to error prone viral RNA-dependent RNA polymerase (RdRp), IAVs mutate rather fast. The HA and NA, in particular, are under constant immunological pressure and mutations in HA or NA genes can create new escape mutants. The mutation rate has been estimated to be ca. 0.8%–1.0% per year at the amino acid level of the HA molecule. The phenomenon when mutations in virus genes occur during virus replication is called antigenic drift.

Both antigenic drift and antigenic shift can result in new IAV subtypes or strains creating new pandemic or seasonal epidemic-causing viruses.

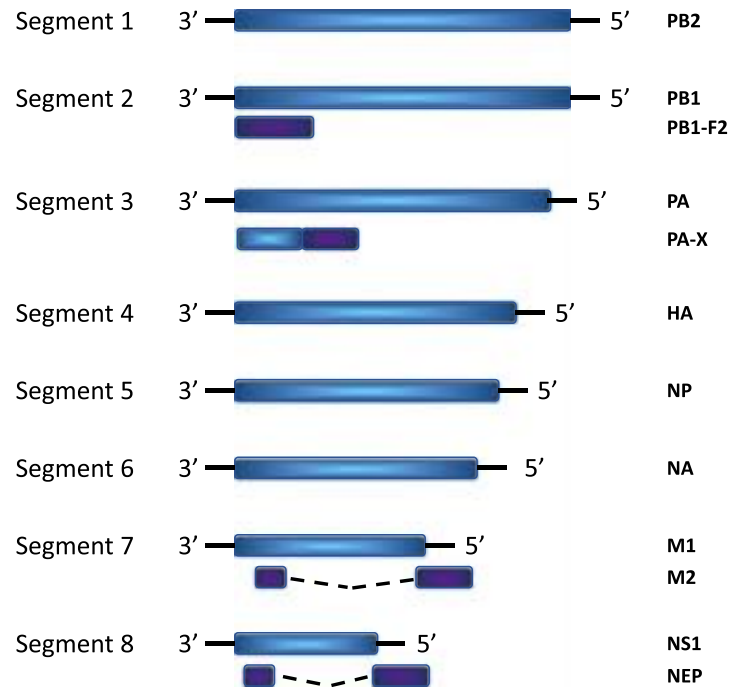


Fig. 2 Proteins encoded by the eight IAV genomic RNA segments.

Table 1 IAV genome segments, their sizes and functions of the IAV encoded proteins in the virus life cycle and some cell regulatory functions

Genomic segment	Protein	Nucleotides, amino acids	Function
1 (2341 nt)	PB2	2280 nt, 759 aa	Polymerase subunit, 5'-cap snatching, transcription and replication initiation, type I IFN induction inhibition
2 (2341 nt)	PB1	2274 nt, 757 aa	Polymerase subunit, RNA-dependent RNA polymerase
	PB1-F2	264 nt, 87 aa	Apoptosis inducer, type I IFN induction inhibition
3 (2233 nt)	PA	2151 nt, 716 aa	Polymerase subunit, RNA synthesis and replication, endonuclease activity
4 (1778 nt)	HA	1698 nt, 550 aa	Surface glycoprotein, binding to receptor, fusion with endosome membrane
5 (1565 nt)	NP	1497 nt, 498 aa	vRNA encapsidation, nuclear import of vRNPs
6 (1413 nt)	NA	1365 nt, 454 aa	Surface glycoprotein, releases HA from sialic acids, cleaves sialic acids on cell surfaces and in mucus, assists in budding
7 (1027 nt)	M1	759 nt, 252 aa	Matrix protein, vRNP interaction, RNA nuclear export regulation, viral budding
	M2	294 nt, 97 aa	Ion channel, virus uncoating and assembly
8 (890 nt)	NS1	693 nt, 230 aa	Interferon antagonist, splicing regulator
	NEP	366 nt, 121 aa	Nuclear export of vRNPs

Life Cycle

Influenza A viruses infect lung cells and immune cells residing in the lungs. New virions are produced locally in the tissues, and coughing or sneezing spreads various size aerosol particles which transmit the viruses to new recipients. Influenza virus is mainly spread via larger aerosol particles within 1.5–2 m from an index cases as well as with direct contact with upper respiratory tract secretions.

Attachment and Entry

A schematic representation of the life cycle of influenza A virus is shown in [Fig. 3](#). IAV attaches to the sialic acid residues of the cell surface glycoproteins and glycolipids. Attachment to these receptors on the target cell surface occurs by the viral HA protein. Seasonal (human) IAVs specifically bind to sialic acids with $\alpha 2,6$ -linkage to galactose in the cell surface glycoproteins that function as viral receptors. Virus is internalized by a receptor-mediated endocytic pathway. At low pH of the endosomes, the M2 ion

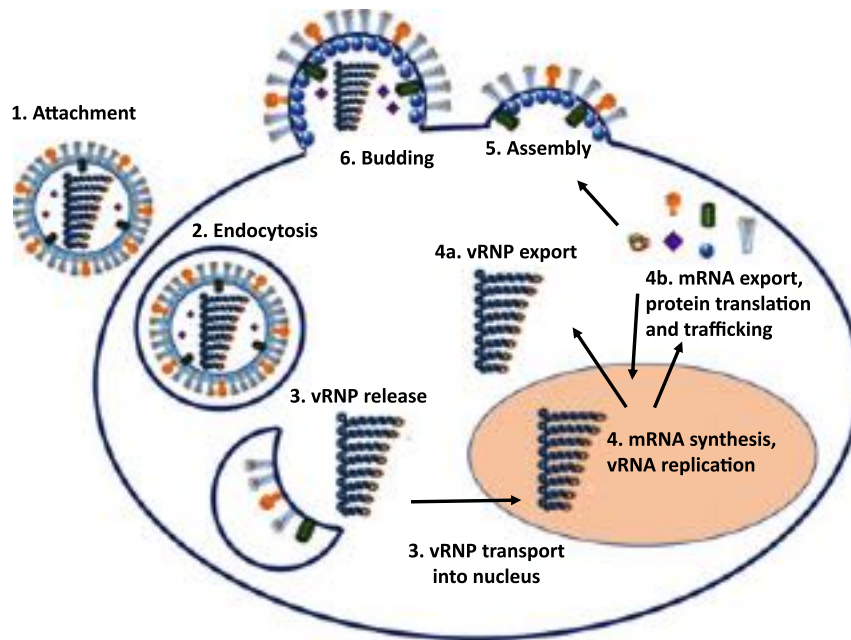


Fig. 3 Life cycle of IAV in infected cell. IAV attaches to terminal sialic acid residues on the host surface via its HA molecule (phase 1). The virus enters the cell via clathrin-mediated endocytosis (2). In endosomes, due to a conformational change in the trimeric HA molecule triggered by low pH, the membranes of the virus and endosomes are fused and vRNPs are released into the cell cytoplasm (3). vRNPs are transported into the nucleus where the primary transcription, replication and secondary transcription are taking place (4). Viral mRNAs are transported into the cytoplasm and viral proteins are synthesized. Newly synthesized Ps, NP, M1, NS1 and NEP proteins are transported into the nucleus where they regulate vRNA replication and packaging of newly synthesized vRNPs that are exported from the nucleus by the aid of M1 and NEP (4a, 4b). vRNPs and viral envelope glycoproteins HA and NA are transported onto the plasma membrane where the assembly (5) and budding of new viral particles takes place (6).

channel opens and proton influx through the M2 ion channel acidifies the interior of the viral capsid. The subsequent acidification of the endosome causes conformational change in the HA protein, resulting in the exposure of a fusion peptide in the N-terminal end of the HA2 subunit. The insertion of the HA2 fusion peptide into the endosomal membrane causes the viral and endosomal membranes to fuse. Low pH inside the virus particle causes M1 proteins to degrade and dissociate. This membrane fusion and virus capsid acidification results in the release of vRNP complexes into cytoplasm. In the cytoplasm, the released vRNPs are transported into the nucleus via cellular importin transport machinery. This transport is facilitated by the nuclear localization signal on NP onto which cellular importin α subtypes bind and with the aid of importin β molecule vRNPs are translocated into the nucleus through nuclear pore complexes. All eight vRNP complexes are transported into the nucleus within an hour after virus attachment. In the nucleus, three forms of viral RNAs are produced: complementary viral RNA (cRNA) that is used as a template for the new viral genomic RNA (vRNA) and for the viral mRNA.

Transcription and Translation

Viral RNA transcription and RNA processing take place in the nucleus. Transcription of viral mRNAs is initiated by the PB2 subunit of the polymerase complex that binds to the host pre-mRNAs. The PA subunit of the polymerase complex is an endonuclease that cleaves 5'-cap structures from cellular pre-mRNAs in a process called cap snatching. The cap-structure is needed later to promote viral mRNA translation by host cell ribosomes. The PB1 subunit ligates the 5'-cap structure to be used as a primer for the synthesis of viral mRNAs. PB1 is also an RNA-dependent RNA polymerase that incorporates nucleotides to the newly synthesized viral mRNAs. Polyadenylation of viral mRNAs takes place at the 5'-end of the vRNA segment by reiterative stuttering of the polymerase complex. All eight viral RNA segments are transcribed to mRNAs. Transcripts for M and NS contain splicing sites, and cellular spliceosome is used to produce mRNAs encoding M2 and nuclear export protein NEP. Processed viral mRNAs are transported to the cytoplasm for translation.

Translation of viral proteins takes place on cytosolic ribosomes (for PB1, PB2, PA, NP, NS1, NS2, and M1) and on endoplasmic reticulum associated ribosomes (for HA, NA and M2). Cellular ribosomes recognize the 5'-cap structures and translate the viral mRNAs into proteins. Of the newly synthesized viral proteins, the polymerase subunits PB1, PB2 and PA, and NP, NEP, M1, NS2 and NS1 are transported into the nucleus where the proteins regulate viral RNA replication, mRNA processing and downregulate innate immunity (NS1). HA, NA and M2 are incorporated into the membranes of endoplasmic reticulum (ER) during translation. HA and NA form trimers (HA) and tetramers (NA) on ER membranes and these molecular complexes are glycosylated at the Golgi

complex. HA is cleaved into HA1 and HA2 by cellular proteases before entering the plasma membrane. This cleavage is a prerequisite for the progeny virions to be infectious and have a functional HA2-associated fusion peptide to initiate the next round of infection.

Genome Replication

In addition to mRNA transcription, the new viral genomic RNA segments are replicated in the nucleus. First, viral cRNA is synthesized by an unprimed process by viral RNA-dependent RNA polymerase using the incoming vRNAs as templates. These newly-synthesized cRNAs are wrapped by newly-produced viral NPs and a polymerase (PB1, PB2 and PA) complex to form a cRNP structure. This cRNP is subsequently used as a template to produce more vRNA by RNA-dependent RNA polymerase. Newly synthesized vRNA segments are wrapped by NPs, newly-produced polymerase subunits are added, and these vRNPs are ready to be exported from the nucleus.

Viral Assembly

Viral ribonucleoprotein complexes (vRNA, NPs and polymerase subunits) are assembled in the nucleus. NEP exports the vRNPs and M1 complexes out from the nucleus. vRNP-M1 complexes created by a nuclear egress process are transported to the plasma membrane by an unknown mechanism that involves cellular protein Rab11. Viral M1 directs the formation of the progeny virions on the plasma membrane and viral surface glycoproteins HA and NA associate with vRNP-M1 complexes.

Budding of New Virions

Progeny virus assembly and virus budding occur on distinct areas on the plasma membrane. HA and NA, along with M2, localize on the plasma membrane to raft regions that are rich in cholesterol and sphingolipids. Viral proteins HA, NA and M1 from the incoming vRNP-M1 complexes induce plasma membrane curvature and new progeny viruses are formed by wrapping of the plasma membrane around the 8 segments of vRNPs. M2 on the plasma membrane is localizing to the budding site and assists in membrane bending and scission. These events result in the budding of the new virions from the plasma membrane. Viral NA protein cleaves sialic acid from the cell surface glycoproteins and glycolipids facilitating progeny virions to be released from the host cell. This sialidase activity of NA also facilitates the virus movement and entry into new host cells by cleaving the sialic acids in the mucus of respiratory track.

Immunological Responses and Immune Evasion

Non-specific structural protective barriers, such as mucosa, ciliated epithelial cells, and protease inhibitors at the airways, restrict influenza viruses to gain access to epithelial cell surfaces. If IAV bypasses these barriers and is able to enter the target cell, various types of immunological responses are initiated in the host.

Innate Immune Responses

The entry of IAV initiates innate immune responses in the target cell leading to the production of interferons and cytokines. The RNA structures of the incoming and replicating virus are recognized by cellular pattern recognition receptors. In influenza infection, the viral structures are recognized by cytoplasmic cellular dsRNA helicase enzymes retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated protein 5 (MDA-5), and by cellular membrane-associated Toll-like receptors (TLR3 and TLR7). The recognition of viral RNAs initiates a signaling cascade in the cell, leading to the activation of transcription factors such as interferon regulatory factor 3 (IRF3), IRF7 and nuclear-factor-kappa-B (NF- κ B). These transcription factors translocate into the nucleus and activate the transcription of type I and type III interferons (IFN- α/β and IFN- λ , respectively) and pro-inflammatory cytokines such as tumor necrosis factor alpha (TNF- α), interleukin 6 (IL-6) and IL-1 β . Another innate immune response mechanism, inflammasome, is activated through NOD-like receptors, such as NOD-like receptor family pyrin domain containing 3 (NLRP3), leading to transcription of pro-IL-1 β , pro-IL-18, and pro-caspase-1, which are subsequently cleaved into active forms of the proteins.

Interferons are secreted from the cell, and they act both as para- and autocrine signaling molecules to activate antiviral and inflammatory responses in infected cells and in neighboring cells. Secreted type I or III IFNs bind to their specific receptors IFNAR or IFNLR, respectively, on the cell surface and a signaling cascade is initiated. Through activation of cellular Janus kinase (JAK) pathways transcription factors called signal transducer and activator of transcription (STAT) are activated. STAT1 and STAT2 form heterodimers, translocate into nucleus where they associate with interferon regulatory factor 9 (IRF9). In the nucleus these activators initiate the transcription of antiviral and proinflammatory genes, that are called interferon-stimulated genes (ISGs). During virus infection, several hundred of ISGs can be induced. The main antiviral proteins produced in IAV infection include

protein kinase R (PKR), Mx-proteins and interferon-induced transmembrane protein family proteins (IFITMs). These antiviral proteins inhibit the early steps of IAV infection.

Interferons and interleukins, i.e., cytokines, regulate immune responses and inflammatory reactions, and act as mediators in initiating the activation of adaptive immune responses.

Adaptive Immune Responses

Adaptive immune responses are specifically targeted to various antigenic structures of the invading pathogen. Adaptive immune responses include antibody-mediated response (B cells) and cell-mediated response (T cells). IAV infection leads to both B cell and T cell-mediated immune responses. On the site of infection, dendritic cells take up IAV particles, mature and migrate to local lymph nodes to present virus-derived antigens to T cells. Activated virus-specific lymphocytes migrate from the lymph nodes to the site of infection where locally secreted cytokines further enhance the development of adaptive immune response. Cell-mediated immunity is important in the recovery from influenza infection. Cytotoxic T cells (CD8⁺) kill virus-infected cells and secrete cytokines to restrict virus replication. T cells also develop to memory T cells that are specific for conserved IAV epitopes found e.g., in NP, M1, and PA proteins. However, the cytotoxicity of the memory T cells fades in time. Helper T cells are important in promoting B cell responses against IAV infection. Viral antigens in lymph nodes are taken up by B cells and with additional B cell stimulatory signals IAV-specific antibody response is activated. B cells produce IgA and IgG antibodies that act on mucosal membranes and in tissues, respectively. The majority of antibodies are targeted against HA, NA and NP proteins. Anti-HA and anti-NA antibodies are protective and neutralizing towards the reinfection with IAV. However, due to relatively fast mutation rate of IAVs, this protection is not always effective against infections with new IAV strains. In addition, the capability of B cells to produce antibodies wanes with age.

Immune Evasion Strategies of IAV

IAVs, as with other pathogenic viruses, have several strategies to evade or delay host cell immune responses. The main player in this is the viral NS1 protein. In infected cells, NS1 is rapidly expressed to high levels and the protein is found mainly in the nucleus, but some protein is also found in the cell cytoplasm. The RNA-binding domain of NS1 interacts with viral double-stranded RNA to protect it and to interfere with the interaction of viral RNA with cellular RNA recognition receptors RIG-I-like receptors and TLRs. In addition, the effector domain of NS1 interacts with host cell mRNA processing. This leads to shut-down of host protein synthesis facilitating virus replication and viral protein translation. NS1 interferes also with immune response pathways with several mechanisms: it interferes with cell antiviral gene transcription in the nucleus, it inhibits RIG-I pathway and interferon-induced pathway, it inhibits the function of antiviral proteins, it affects the signaling pathways, and it inhibits adaptive immune response by affecting the maturation of dendritic cells. Another immune response antagonist, PB1-F2, activates apoptosis and cell death by targeting mitochondrial membranes. It also interferes with the RIG-I pathway by interacting with mitochondrial antiviral signaling (MAVS) protein on the mitochondrial membrane.

Evolution and Epidemiology

Evolution of IAV

Surface glycoproteins, HA and NA, are under constant immunological pressure. These proteins mutate rather fast due to antigenic drift when the RNA-dependent RNA polymerase makes mistakes during RNA replication (**Fig. 4**). IAV attaches to the cell surface receptor with HA protein and the antibodies targeting especially receptor-binding part of HA restrict the binding IAV on the cell surface receptors and reduce the likelihood of infection. However, amino acid changes in the major antigenic sites and changes in glycosylation pattern reduce the ability of anti-HA and anti-NA antibodies to bind to the virus, and return the ability of virions to infect the cells. IAV NA protein acts as a sialidase and cleaves sialic acids enabling the virions to be released from the cell surface after budding from the plasma membrane. The antibodies targeting the catalytic sites of NA prevent the release of progeny virions. However, mutations that retain catalytic activity but hinder antibody binding again assist the infectious cycle of virions. This constant mutation of HA and NA proteins is referred as antigen evolution. Due to this antigen evolution, pre-existing immunity to other IAV strains does not fully protect from new epidemic strains. Hence new influenza epidemics occur yearly, and the influenza strain composition of vaccines has to be constantly evaluated.

IAV Epidemics and Pandemics

Influenza A viruses have caused severe pandemics and yearly epidemics. Of the potential HA/NA combinations, three influenza A subtype viruses (H1N1, H2N2, and H3N2) have caused pandemics or epidemics in humans. In addition, avian influenza A virus subtypes H5N1, H7N7, H9N2, and H7N9 have been able to infect humans. The most variable part of the HA protein has 25%–80% homology between the IAV subtypes (**Fig. 5**). Currently two main subtypes, H1N1 (pdm09) and H3N2, are circulating

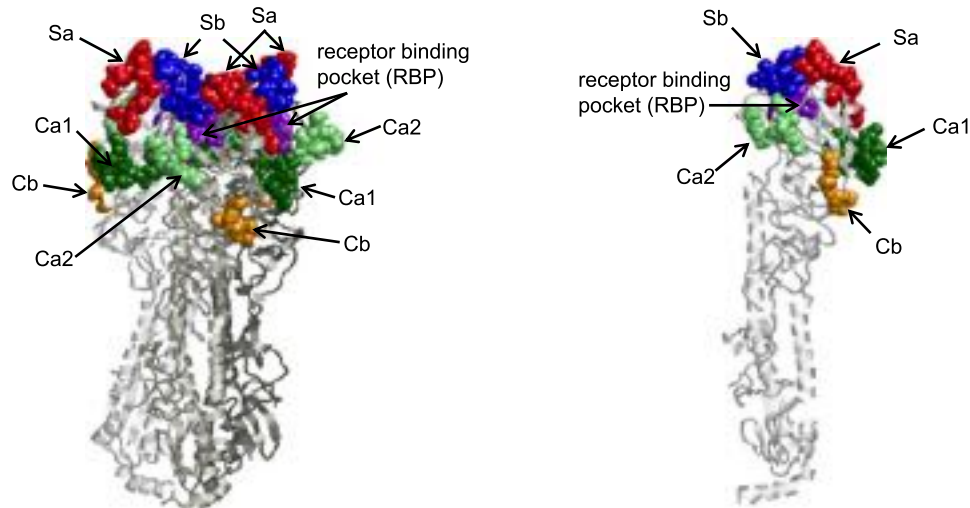


Fig. 4 Major antigenic sites of H1 virus (Ca, Cb, Sa and Sb) shown on a trimeric (on left) and on a monomeric (on right) HA molecule. Different antigenic sites are shown in different colors and the receptor binding pocket is shown in purple.

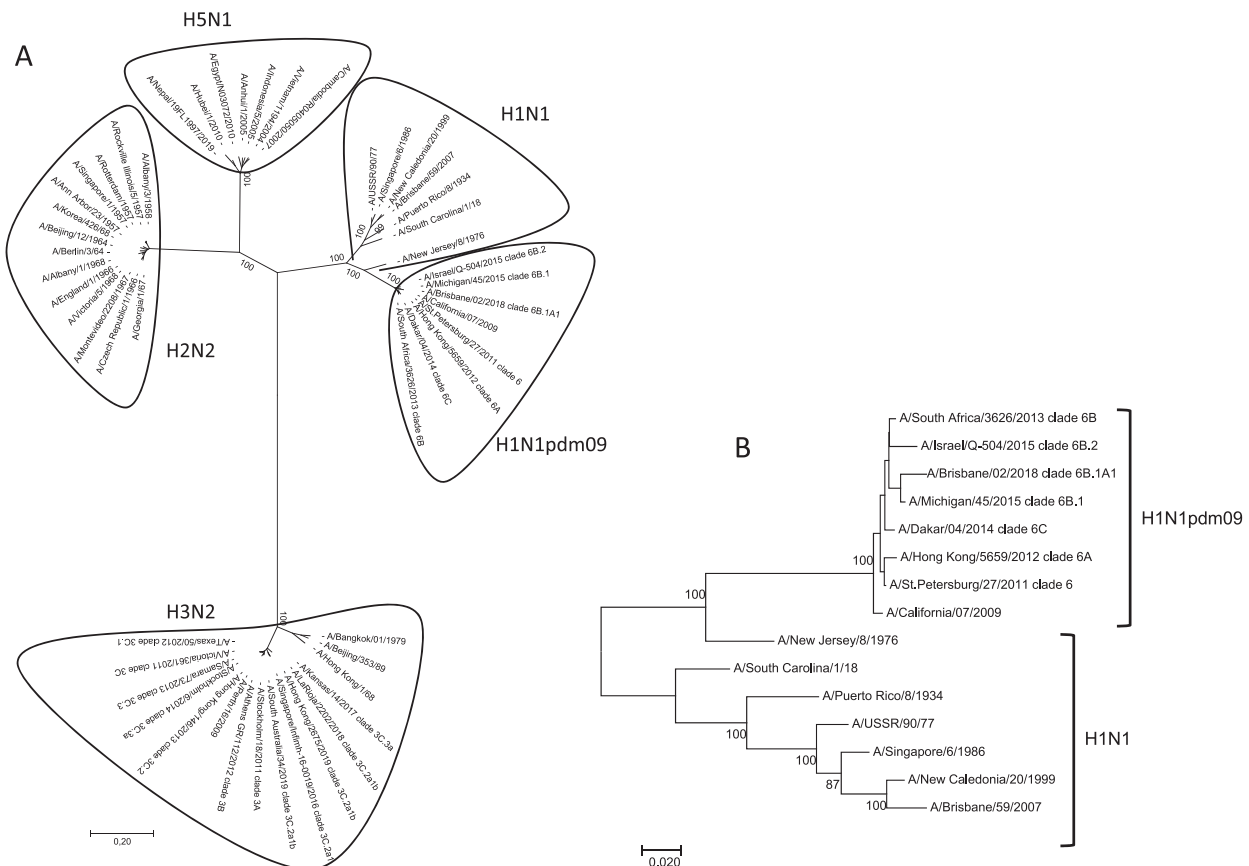


Fig. 5 Phylogenetic trees based on HA sequences of various IAV subtypes. A. The tree of five different subtypes was constructed by the Neighbor-Joining method with Mega software version 7. The length of branches are proportional to the evolutionary distances. B. Phylogenetic tree based on HA sequences of H1N1 subtypes. The tree was constructed by the Neighbor-Joining method. The horizontal lines are proportional to the number of nucleotide changes.

in the human population world wide. Influenza epidemics occur in the Northern hemisphere usually in November-April, and in the Southern hemisphere in June-October. In countries near the equator this seasonality is not that distinct, however, the peak activity of IAV infections is mainly during the rainy season.

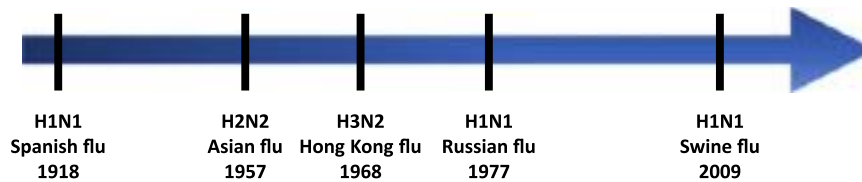


Fig. 6 Timeline of IAV pandemics and global epidemics in the 20th and 21st centuries.

WHO has estimated that seasonal influenza virus infections (both IAV and IBV) cause 3–5 million cases of severe respiratory illness annually and 290,000–650,000 deaths. This accounts for roughly 2% of all annual respiratory illness-related deaths. The intensity of the epidemic varies between countries due to the heterogeneity of the epidemic virus strains and differences in the immune status of a given population.

H3N2 subtypes of IAV have caused seasonal epidemics since 1968 when the H3N2 subtype caused a pandemic. H3N2 subtypes readily infect all age groups, and they are the cause for the most of the IAV related hospitalizations and deaths. H1N1 subtypes caused seasonal epidemics less frequently until 2008. H1N1 epidemics usually affect younger age groups than the elderly ones. In 2009 a new pandemic H1N1 subtype (“swine flu”) caused a world wide pandemic and since then this subtype has been circulating and causing annual epidemics.

IAVs have caused several pandemics (Fig. 6). New pandemics arise when the virus has a new set of reassortant segments (antigenic shift). The first identified pandemic was caused by an H1N1 subtype (Spanish flu) in 1918. It has been estimated that the 1918 virus infected one third of the world population causing at least 50 million deaths. The high mortality among healthy adults was exceptional for this virus. However, the 1st World War, low hygienic conditions and lack of antibiotics and intensive care may have contributed to the poor prognosis of the patients which was often due to secondary bacterial infections. The next two pandemics, in 1957 (Asian flu) and 1968 (Hong Kong flu), were caused by H2N2 and H3N2 subtypes, respectively. The Asian flu strain was a reassortant between human H1N1 and avian H2N2, and this subtype caused annual epidemics for the next ten years, being then replaced by the H3N2 Hong Kong flu subtype. Both pandemics resulted in an estimated 1 million deaths each. In 1977 the reappearance of a previous 1950s-like H1N1 subtype (Russian flu) caused a wide global epidemic, affecting the younger population since older individuals had protective immunity from infections caused by the descendants of the 1918 H1N1 virus. The H1N1 subtype causing the pandemic in 2009 and resulting in roughly 400,000–800,000 deaths, was exceptional in that sense that it originated from North America. The virus had accumulated genomic segments from IAVs of birds, swine and humans. The 2009 pandemic, like the one in 1918, affected the younger generations since the elderly individuals (born before 1950s) had pre-existing immunity due to infections caused by the 1918 descendant H1N1 viruses. The 2009 subtype was named as the H1N1pdm09 virus.

Avian influenza strains have caused limited epidemic clusters or sporadic cases of infection in humans. Avian IAV subtype H5N1 has caused almost 900 human cases with 50% mortality in ca. 30 countries. Subtype H7N7 caused an outbreak in the Netherlands with mild symptoms in humans (though one died), as also H9N2 in Hong Kong and China. Subtype H7N9 has caused nearly 1600 cases with 35% mortality in China and this subtype is causing new human cases annually. These avian influenza strains are transmitted to humans mainly directly from birds, and human-to-human transmission is either absent or extremely rare. As for future, the avian IAVs are posing a threat of starting a new deadly pandemic. To prevent this from happening, global measures have been put in place (instructions on farming the birds and culling of poultry if infection is detected) and the appearance of IAV strains with amino acid changes enhancing the likelihood of human-to-human transmission is intensively monitored.

Clinical Features

Seasonal IAV infections can cause asymptomatic or mild disease, or a more pronounced disease which may even lead to death. Incubation time from the time of exposure until the onset of the disease is usually 1–4 days, but can be up to 7 days. Virus is shed in the respiratory secretions a day before symptoms, and shedding usually continues for up to five days.

Symptomatic disease often begins with a sudden high fever that lasts for several days. Other common symptoms are headache, chills, muscle pain, dry cough, sore throat, fatigue and loss of appetite. Nasal congestion and runny nose follow later. Cough and fatigue can last for a few weeks. In young children another common symptom is diarrhea. Common complications to follow IAV infection in adults are bronchitis and sinusitis, and in children otitis media.

Most infected people recover from the infection but some develop more severe, sometimes even life-threatening conditions. Usually people belonging to high risk groups develop these complications: > 65 years of age, pregnant women, children < 2 years of age, and people with certain medical chronic conditions such as disease of the lungs, diabetes, or cardiovascular diseases. The most severe complication is pneumonia, caused either by the virus or by concurrent bacterial infection. Other severe complications may include myocarditis, encephalitis, myositis, rhabdomyolysis, ARDS or multi-organ failure of, for example, respiratory or kidney organs. Influenza infection can also worsen the underlying disease.

Pathogenesis

Virus initially replicates in the epithelial cells of the trachea and bronchia, and can disseminate into other parts of the respiratory system. Replication of the virus and host inflammatory responses cause cellular damage to the epithelium. The recovery process of the epithelium begins in few days, but can last up to a month.

Viral pneumonia occurs in some patients when the virus replicates in alveolar epithelial cells. This causes the walls of the alveoli and bronchioles to rupture, leading to pneumonia. Often pneumonia is a combined viral-bacterial pneumonia. The patient appears to recover from IAV infection, but then a few days later symptoms of pneumonia develop. This is caused by a secondary infection with bacteria such as *Streptococcus pneumoniae*, *Staphylococcus aureus*, or *Hemophilus influenzae*. This condition can be treated with antibiotics, but can be fatal.

In some patients IAV infection can lead to an acute respiratory distress syndrome (ARDS). This life-threatening condition is caused by a breakage in the endothelial-epithelial barrier due to IAV infection killing the epithelial cells. This results in the production of cytokines that attract leukocytes and activate endothelial cells, leading to infiltration of leukocytes into the lungs. The responses triggered by immune cells then further break the barrier, resulting in the accumulation of fluids into the lungs and a respiratory failure.

Another severe pathological outcome in IAV infection is a cytokine storm. In addition to lung epithelial cells, IAV infects also endothelial cells and lung-resident macrophages. All these infected cell types secrete cytokines leading to the expression of ISGs. This host response is tightly autoregulated, however, in some patients the regulation fails. As a result, there is an excessive inflammatory response in the lungs, leading to permeabilization of lung endothelium and leakage of fluids into the lungs.

Diagnosis

IAV infections are diagnosed based on the presence of virus or viral RNAs or proteins in the sample (acute infection) or by detection of an increase in influenza antibodies in paired serum specimens. IAV can be cultured from nasal swabs or nasopharyngeal aspirates (NPAs). For treatment decisions, fast diagnosis is required. Thus virus cultivation is not an optimal method of choice, although it is important in epidemiological surveillance of circulating strains and in selecting vaccine strains. IAV particles in the samples can be detected by antigen detection methods, such as rapid point-of-care tests that are fast but robust, and give accurate results only in the early phases of the infection when the viral load is the highest. Presently, the most commonly used method is the detection of IAV RNA in a sample by RT-PCR. Different PCR methods are fast, some can quantitate the amount of viral RNA in the sample and can subtype the IAV strain causing the infection.

Treatment and Prevention

Antivirals

More detailed description of antivirals and vaccines is available elsewhere in the Encyclopedia. Briefly, commercially available and widely used drugs against influenza viruses are of three categories. Neuraminidase inhibitors, such as zanamivir and oseltamivir, target viral surface NA proteins. Adamantane compounds, amantadine and rimantadine, block the M2 ion channel protein. Amantadine compounds are effective only for IAV and not for IBV. Baloxavir marboxil, a new antiviral against influenza, is a cap-endonuclease inhibitor. Antivirals are effective when started within 48 h after the onset of infection, making the treatment at the population level challenging. In addition, wide use of antivirals creates viral escape mutants that are resistant to drugs, and in fact, all currently circulating IAV strains are resistant to amantadines. Resistance to NA inhibitors is presently not very common. Treatment with antivirals given in due time or prophylactically can reduce the severity of symptoms and occurrence of secondary infections, reduce the duration of the disease, and prevent the IAV infection.

Vaccines

The most effective countermeasure for IAV infections is vaccination. Yearly updating of influenza vaccines (against influenza A and B) is needed due to constant immunological pressure on influenza viruses that gives rise to new progeny virions with altered protein sequences with ability to escape immune responses elicited by vaccines. Thereby, WHO estimates twice a year the composition of the annual vaccines based on viral strains obtained and characterized by WHO Global Influenza Surveillance and Response System (GISRS).

To manufacture influenza vaccines, the whole viruses are inactivated (split vaccines), viruses are attenuated or vaccines include only the HA and NA proteins of influenza virus (subunit vaccines). Presently, influenza vaccines include four influenza strains: two influenza A (H1N1 and H3N2) and two influenza B (Victoria and Yamagata) strains. In addition to influenza virus components, different types of adjuvants may be included in the vaccine to strengthen the immune response. There are two ways to administer influenza vaccines, by injection (inactivated) or inhalation (live attenuated). Influenza vaccination is recommended for risk groups, i.e. mainly infants, elderly people and those with predisposing underlying conditions, such as lung and cardiovascular

diseases, diabetes, impaired immunity or cancer. The vaccine efficacy varies every year, it depends on the timing of the vaccination campaign, the age and immune status of the vaccinees, the match of vaccine strains to circulating virus strains, and the testing procedures of the vaccine efficacy. Usually the efficacy is 40%–70%. In addition to prevention of infection, vaccination has been shown to reduce the severity of the influenza disease and to prevent secondary bacterial infections.

See also: Avian Influenza Viruses (*Orthomyxoviridae*)

Further Reading

- Chen, X., Liu, S., Goraya, M.U., *et al.*, 2018. Host immune response to influenza A virus infection. *Frontiers in Immunology* 9, 320.
- Dou, D., Revol, R., Østbye, H., Wang, H., Daniels, R., 2018. Influenza A virus cell entry, replication, virion assembly and movement. *Frontiers in Immunology* 9, 1581.
- Ferhadian, D., Contrant, M., Printz-Schweigert, A., *et al.*, 2018. Structural and functional motifs in influenza virus RNAs. *Frontiers in Microbiology* 9, 559.
- Gounder, A.P., Boon, A.C.M., 2019. Influenza pathogenesis: The effect of host factors on severity of disease. *Journal of Immunology* 202 (2), 341–350.
- Hsu, A.C., 2018. Influenza virus: A master tactician in innate immune evasion and novel therapeutic interventions. *Frontiers in Immunology* 9, 743.
- Kash, J.C., Taubenberger, J.K., 2015. The role of viral, host, and secondary bacterial factors in influenza pathogenesis. *American Journal of Pathology* 185 (6), 1528–1536.
- Klemm, C., Boergeling, Y., Ludwig, S., Ehrhardt, C., 2018. Immunomodulatory nonstructural proteins of influenza A viruses. *Trends in Microbiology* 26 (7), 624–636.
- Krammer, F., 2019. The human antibody response to influenza A virus infection and vaccination. *Nature Reviews Immunology* 19 (6), 383–397.
- Krug, R.M., 2015. Functions of the influenza A virus NS1 protein in antiviral defense. *Current Opinion in Virology* 12, 1–6.
- Kuiken, T., Taubeberger, J.K., 2008. Pathology of human influenza revisited. *Vaccine* 26 (Suppl 4), D59. [66].
- Lowen, A.C., 2017. Constraints, drivers, and implications of influenza A virus reassortment. *Annual Review of Virology* 4 (1), 105–121.
- Meineke, R., Rimmelzwaan, G.F., Elbahesh, H., 2019. Influenza virus infections and cellular kinases. *Viruses* 11 (2).
- Padiilla-Quirarte, H.O., Lopez-Guerrero, D.V., Gutierrez-Xicotencatl, L., Esquivel-Guadarrama, F., 2019. Protective antibodies against influenza proteins. *Frontiers in Immunology* 10, 1677.
- Taubenberger, J.K., Morens, D.M., 2008. The pathology of influenza virus infections. *Annual Review of Pathology* 3, 499–522.

Relevant Websites

- <https://www.gisaid.org/>
 GISAID.
 Initiative.
- <https://www.cdc.gov/flu>
 Influenza (Flu).
 CDC.
- <https://www.ecdc.europa.eu/en/seasonal-influenza>
 Seasonal influenza.
 ECDC.
 Europa EU.
- <https://www.who.int/influenza/en/>
 WHO.
 Influenza.
 World Health Organization.

Influenza B, C and D Viruses (*Orthomyxoviridae*)

Thorsten Wolff, Robert Koch Institute, Berlin, Germany

Michael Veit, Free University of Berlin, Berlin, Germany

© 2021 Elsevier Ltd. All rights reserved.

Glossary

EC50 Half maximal effective concentration. EC50 describes the concentration of a substance that gives half maximal response of a biological pathway.

GISRS Global Influenza Surveillance and Response Systems. GISRS is the joint collaboration of WHO with 6 WHO Collaborating Centers, four Essential Regulatory Laboratories and National Influenza Centers in more than 140 countries dedicated to protect people from the threat of influenza. It was established in 1952.

HA Hemagglutinin.

HEF Hemagglutinin-esterase-fusion protein. HEF mediates host cell attachment and fusion of the incoming virions as well as progeny virion release after completion of the replication cycle.

IBV Influenza B virus.

IC50 Inhibitory concentration 50. IC50 describes a concentration of a compound sufficient to limit virus titers by 50% when added to a host or a host cell.

ICV Influenza C virus.

IDV Influenza D virus.

NA Neuraminidase.

NEP Nuclear export protein.

NLS Nuclear localization signal. NLS describes a topogenic amino acid sequence that directs the import of a protein into the nucleus of a eukaryotic cell.

NS1 Non-structural protein.

PKR dsRNA-activated protein kinase.

RIG-I Retinoic acid induced gene I.

SC35 Serine/arginine-rich splicing factor 35.

vRNP Viral ribonucleoprotein. The genomic RNAs of influenza viruses are complexed with the viral nucleoprotein and polymerase into vRNPs that represent the smallest unit for expression of the encoded protein(s).

WHO World Health Organization.

Classification (Compact)

Influenza B virus (IBV) was identified for the first time during an outbreak of an acute respiratory disease in 1940 in Northern America, in which it was found to lack antigenic cross-reactivity to influenza A virus (IAV). IBV is classified as the only member of the genus influenza virus B within the order of the *Orthomyxoviridae*. IBV was estimated to have diverged from IAV around 4000 years ago based upon the calculated rate of amino acid substitutions within HA proteins, but the mechanisms of replication and transcription, as well as functionality of most viral proteins appear to be largely conserved, with some unique differences. IBV has been detected in rare occasions in seals and pigs, but humans are the major host species of IBV, in which it causes typically the same spectrum of symptoms of influenza-like illness as IAV although pathologies outside the respiratory tract have been determined in a minority of IBV infections. A large recent metagenomic study detected an IBV-like virus in a fish species suggesting that the genus may have additional vertebrate host reservoirs outside mammalian species. Sentinel surveillance detects IBV in some seasons as the dominating circulating influenza virus type highlighting its significant socio-economic impact. There are currently two lineages of IBV (B/Yamagata-like and B/Victoria-like viruses) with limited antigenic cross-reactivity. Viruses of both lineages circulate to varying extents and ratios in all parts of the world. The lack of a large animal host reservoir able to foster the appearance of pandemic variants, a lower evolutionary rate and generally considered milder courses of disease in comparison to IAV infection are typical features of IBV.

Influenza C virus (ICV) was first isolated in 1947 from a human patient having mild respiratory symptoms. Since the virus showed no cross-reactivity with antisera against influenza A or B viruses, it was classified as a new genus of the *Orthomyxoviridae* (influenza C virus). The virus has a worldwide distribution and the majority of humans develop antibodies against ICV early in life. Humans are the main reservoir of ICV, but occasionally the virus may also infect dogs and pigs.

In 2011 a new Influenza C-like virus was isolated from clinically ill pigs. Since this isolate is unable to reassort with ICV it is now officially named as influenza D virus (IDV). Subsequent studies showed that IDV can also infect small ruminants and cattle, the latter of which is apparently its main reservoir.

The first part of this article reviews current knowledge on type B influenza virus, whereas the second part addresses important aspects of type C and D viruses.

Influenza B virus

Virion Structure

IBV virions show pleomorphic structures that can have spherical, irregular or filamentous appearances and are surrounded by a membrane, in which the viral membrane proteins HA, NA, NB and BM2 are integrated. Type A and B influenza viruses cannot be distinguished based upon their morphology or size as determined by electron microscopy (EM) (Fig. 1). The average size of spherical IBV virions of the prototypic B/Lee/40 strain grown in embryonated chicken eggs was determined to be 137 ± 27 nm. Apparently, different organizational patterns have been reported for the eight viral negative-strand RNA gene segments, which together with the viral nucleoprotein and polymerase exist as viral RNPs inside IBV virions. Analysis of ultrathin sections by transmission electron microscopy suggested that virions of laboratory and clinical IBV strains package mostly eight RNP structures that are organized in a 7 + 1 configuration with a central long RNP encircled by 7 others. In contrast, 3D reconstructions by cryo-EM tomography were rather indicative for a disordered organization of IBV vRNPs. Studies of purified virions by electron microscopy revealed the presence of 400–500 spike-like projections per spherical virion corresponding to the viral glycoproteins. Quantitative analysis of purified virions by mass spectrometry showed that M1 is the most abundant IBV virion component followed by NP, HA and NA, whereas the level of NB and P proteins were the lowest.

Genome

The genome of IBV is organized into eight single-stranded RNA molecules that are of negative polarity and vary in length from 1.1 to 2.4 kb (Table 1). Each of the segments carries characteristic highly conserved sequences of 10 (AGUAG (A/U) AACA) and 9 nucleotides (UCGUCUUCG) at their 5' and 3'-termini, respectively, which themselves flank short segment-specific non-coding regions (NCR) that encompass the coding sequences and short U-rich elements directing polyadenylation. Each IBV gene segment

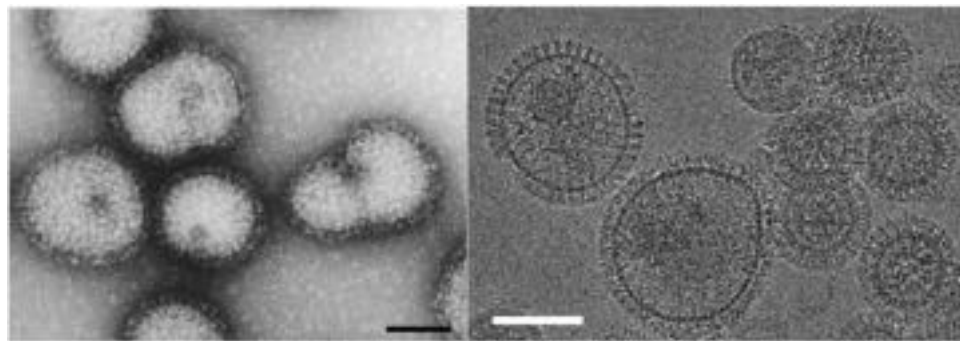


Fig. 1 EM pictures of influenza B and C viruses. Left: Influenza B virus. Negative staining EM. (B/Harbin/7/94, Source: RKI, ZBS4). Right: Influenza C virus. Cryo EM of C/Johannesburg/1/66. Source: Kai Ludwig (Core Facility BioSupraMol of the Free University Berlin). Scale bar = 100 nm.

Table 1 Identities and lengths of genome segments and encoded gene products from prototypic influenza B virus, influenza C virus and influenza D virus strains

Segment	B/Lee/40 ^a vRNA (nts)	B/Lee/40 ^a Protein(s) (aa)	B/Colorado/06/2017 (Vic) ^a Protein(s) (aa)	B/Phuket/3073/2013 (Yam) ^a Protein(s) (aa)	C/Jhb/1/1966 ^a Protein(s) (aa)	D/swine/Ok/1334/2011 ^b Protein(s) aa
1	2396	PB2 (769)	PB2 (770)	PB2 (770)	PB2 (774)	PB2 (772)
2	2368	PB1 (752)	PB1 (752)	PB1 (752)	PB1 (754)	PB1 (753)
3	2304	PA (726)	PA (726)	PA (726)	P3 (709)	P3 (710)
4	1882	HA (584)	HA (583)	HA (584)	HEF (655)	HEF (664)
5	1841	NP (560)	NP (560)	NP (560)	NP (565)	NP (552)
6	1557	NA (466)	NA (466)	NA (466)	M1 (242)	M1 (246)
		NB (100)	NB (100)	NB (100)	CM2 (115)	CM2 (?)
7	1191	M1 (248)	M1 (248)	M1 (248)	NS1 (246)	NS1 (243)
		BM2 (109)	BM2 (109)	BM2 (109)	NS2 (182)	NS2 (184)
8	1096	NS1 (281)	NS1 (282)	NS1 (281)		
		NEP (122)	NEP (123)	NEP (122)		

^asource: Influenza Virus Database, NCBI (<https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database>).

^bsource: Hause, B.M., Collin, E.A., Liu, R., *et al.*, 2014. Characterization of a novel influenza virus in cattle and Swine: Proposal for a new genus in the *Orthomyxoviridae* family. *Mbio* 5, e00031–14.

Abbreviations: nts, nucleotides; aa, aminoacids.

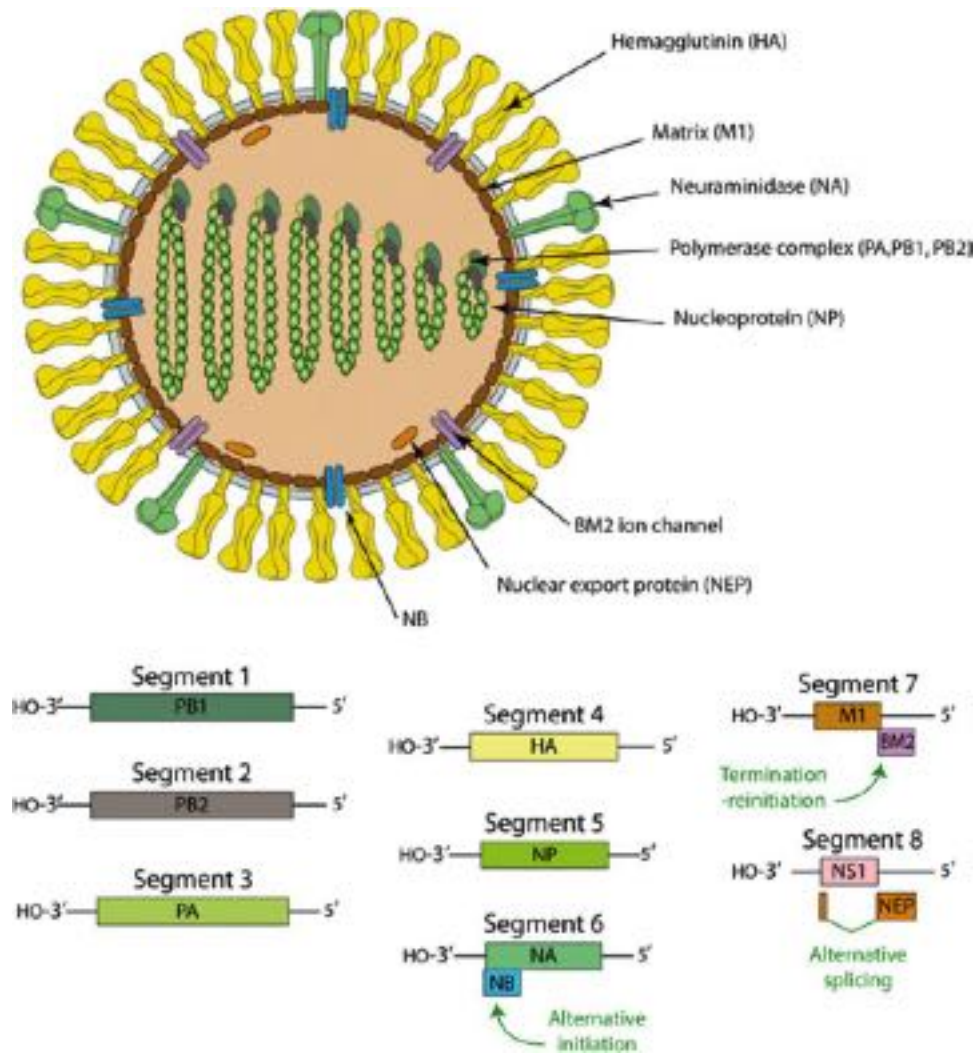


Fig. 2 Scheme of an IBV particle and its genome segments. Adapted from viralzone https://viralzone.expasy.org/80?outline=all_by_species with permission.

encodes one or two of the eleven known viral gene products (PB1, PB2, PA, HA, NP, NA, NB, M1, BM2, NS1, NEP) utilizing distinct strategies to expand viral coding capacity **Fig. 2**.

The nomenclature of the viral gene segments resembles those of IAV, with the largest three encoding the polymerase subunits PB2 (segment 1), PB1 (segment 2) and PA (segment 3), followed by the monocistronic HA (segment 4) and NP genes (segment 5). The residual three segments each encode for two viral proteins, which is achieved by different strategies: Segment 6 differs from its IAV counterpart by encoding not only the viral neuraminidase (NA), but also the fourth viral transmembrane glycoprotein NB in an overlapping reading frame. Its translation is initiated at the first AUG start codon, just four nucleotides upstream of the NA start codon. The NA and NB open reading frames overlap for 292 nucleotides. Translational initiation at the NA start codon is efficient despite being located downstream of the NB start.

Also the expression strategy of segment 7 encoding the viral M1 matrix and the BM2 ion channel protein differs from the one used by IAV, in which the primary mRNA transcript is processed by cellular splicing machinery proteins. In contrast, the IBV M1 and BM2 proteins are both translated from a collinear transcript in which the M1 stop codon and the BM2 start codon are assembled into a UAAUG pentanucleotide, which facilitates ribosomal re-initiation of protein synthesis following completion of M1 translation. Translation of BM2 coding sequence depends on the presence of a conserved *cis* element of 45 nucleotides preceding the stop-start signal. This element contains a nucleotide stretch complementary to helix 26 of 18S rRNA, which has been suggested to facilitate the recruitment of the 40S ribosomal subunit to the mRNA upon termination at the M1 stop codon.

The NS segment (segment 8) of IBV encodes NS1 and NEP proteins, which are expressed either from a collinear mRNA (NS1) or from a spliced transcript, from which an intronic sequence is removed (NEP). NEP exon 1 mRNAs utilizes the AUG start and ten codons of the NS1 reading frame, which is directly spliced to exon 2 that overlaps with the NS1 coding region for 155 nucleotides

in a +1 reading frame. No homologs of accessory protein genes identified in IAV such as PB1-F2 or PA-X have been reported for IBV.

Reassortment, i.e., the exchange of gene segments between two viruses infecting one cell, is a general evolutionary mechanism of segmented RNA viruses, which is also a major driving force for the appearance of novel epidemic IBV strains both from within and in between the Victoria- and Yamagata lineages (see below). Classical coinfection experiments indicated that there can be phenotypic mixing of type A and B viruses generating particles that contain antigenic determinants derived from both parental types. However, IBV has never been observed to produce stable reassortant viruses harboring a gene segment(s) of IAV and vice versa. At least some part of this restriction appears to operate on the level of genome incorporation. Recombinant viral gene segments carrying the IBV HA or NA reading frames could be incorporated together with six other IAV segments into infectious chimeric influenza A/B viruses, when the coding regions were flanked with the packaging signals of the type HA and NA segments. The unique packaging signals of each of the viral gene segments have been mapped for IAV to contain the non-coding regions (NCRs) as well as terminal parts of the coding sequences, but have been less well studied for IBV. The appearance of defective interfering (DI) RNAs derived from viral gene segments containing the conserved terminal promoter sites with large internal deletions of the coding regions, especially in cells infected at high multiplicity, has also been described for IBV. However, the precise mechanism(s) by which the viral polymerase generates viral DI RNA remain to be elucidated.

Life Cycle and Functions of Viral Proteins

Virus Entry and Genome Expression and Replication by Polymerase Proteins

The main targets of IBV are the epithelial cells lining the upper respiratory tract. The available evidence indicates that the replication cycles of IBV and IAV follow very similar or even identical principles, which are, however, executed by divergent and a few type-specific proteins, respectively, resulting in some virus type-specific mechanisms that are described in detail below. Hence, IBV and IAV were shown to enter a permissive host cell by receptor-mediated endocytosis, which is initiated by binding of the viral hemagglutinin (HA) to sialic acid residues that are present either as glycoprotein or glycolipids on the host cell surface. Intracellular release of the viral genome requires both acidification of the interior of the virion through the viral BM2 ion channel, which is followed by HA-mediated fusion of the viral and endosomal membranes in the acidic environment of the late endosome. The pH optimum for fusion activity of the IBV HA has been determined at pH of 5.4–5.6, which is slightly above the range known for HA of human IAV (pH = 5.0–5.4). The viral RNPs are transported from the cytosol into the nucleus, where the viral RNA polymerase, a trimeric enzyme of ca. 270 kDa initiates transcription and replication of the viral gene segments. These events have been less well studied in cells infected with IBV compared to type A viruses. However, the high conservation of structural elements within the three polymerase subunits PB1 (enzymatic core), PB2 (cap binding domain, aa 318–480) and PA (endonuclease site, aa 1–195) as determined by X ray crystallography, as well as functional studies of recombinant polymerases strongly argue for a conserved mode of action.

Within the nucleus each vRNP acts as an independent unit that can either be transcribed into polyadenylated positive-sense viral mRNA using a cap-snatching mechanism, or serves as the template for synthesis of a positive-sense complementary RNA (cRNA) which in turn directs the replication of another virion RNA (vRNA) molecule. The IBV PB2 subunit binds not only to methylated G-capped RNAs like its IAV counterpart, but also to unmethylated GpppG caps indicating a larger flexibility of the IBV PB2 cap binding domain. The viral RNA polymerase complex binds the vRNA or the cRNA at the promoter sites located on the highly conserved stretches of nearly complementary nucleotides at the 3'- and 5'-termini. Successful expression and purification of soluble recombinant IBV polymerase has recently facilitated detailed in vitro studies of binding parameters to viral vRNA and cRNA mimics as well as characterizations of the enzymatic activities of IBV P proteins. It is generally believed that influenza viruses express an error-prone RNA polymerase which contributes to high viral evolutionary rates, but this has not been determined experimentally for the IBV polymerase.

The major component of the vRNPs is the viral nucleoprotein (NP) that is a basic (pI > 9.0) 62 kDa protein and wraps around the viral RNA. It is an essential cofactor for viral transcription and replication and it consists of a head domain, a body domain and a tail loop like its IAV counterpart, to which it has sequence identity below 40%. The tail loop of one NP monomer can insert into a neighboring monomer, which drives the formation of homo-oligomers. A distinct feature of the IBV NP protein is an extended N-terminal region of 70 amino acids that appeared intrinsically disordered in structural analyzes. In a study by NMR and small-angle X ray scattering the N-terminal NP region was bound to nuclear import factor importin- α 7, which together with biochemical and mutational analyzes suggested that the N-terminal region of NP that is enriched in basic amino acids between positions 30 and 71 directs its nuclear targeting.

Intracellular expression of viral mRNA directs the synthesis of eleven known IBV proteins that facilitate amplification and nucleo-cytoplasmic export of the viral genome, the production of progeny virions and/or manipulate innate antiviral defenses. The newly synthesized vRNPs are exported to the cytosol with the support of the viral NEP and, possibly, the M1 protein, which both contain nuclear export signals. The nuclear export of vRNPs is sensitive to inhibition by leptomycin B indicating a dependency on the exportin 1-dependent pathway. IBV encoded NS1 protein suppresses the stimulation of cytosolic pattern recognition receptors such as RIG-I and PKR through newly replicated vRNPs. IBV replication triggers an apoptotic response in infected cells and the activation of both extrinsic and intrinsic induction pathways has been described. The assembly and budding of IBV progeny

virions occurs at the plasma membrane, which is followed by virion release that depends on NA activity. In the following we will discuss the remaining structural and regulatory IBV gene products with a focus on their virus type-characteristics.

Hemagglutinin (HA)

The HA of IBV is synthesized in the ER as a HA₀ precursor of approx. 63 kDa with a predicted leader peptide of 15 amino acids that needs to be proteolytically cleaved by the cellular signal peptidase. HA₀ precursor is cleaved by trypsin-like proteases into HA₁ (approx. 346 amino acids) and HA₂ subunits (approx. 223 amino acids) to become biologically active. HA has an essential role in the viral life cycle as it mediates binding of the virion to host cell surface receptors as well as promoting the release of the viral genome from the late endosome. The HA of IBV is heavily glycosylated containing 10–12 N-linked glycosylation sites and it is assembled into homotrimers. The high number of glycan attachment sites may reflect the prolonged circulation of IBV in humans as glycosylation is believed to be a strategy, by which influenza viruses mask antigenic sites. Cleavage activation of the HA occurs within a motif that is highly conserved among viruses of the Yamagata and Victoria lineages (PAKLLKER↓GFFGAIAGFLE), which liberates the hydrophobic fusion peptide at the N-terminal end of HA₂. Trypsin and the serine proteases TMPRSS2 and TMPRSS11d were shown to execute HA cleavage activation in vitro. TMPRSS2 knock out mice were shown susceptible to IBV infection suggesting the existence of additional proteases that can activate the HA. Several X ray structures have been reported for the neutral pH form of HA expressed by the divergent IBV lineages (Fig. 3). Despite a low degree of sequence identity of approx. 25% to IAV HA these studies revealed an overall conserved fold of the trimeric HA ectodomain composed of a membrane-distal globular head domain (mainly located on HA₁) and an elongated stem (mainly located on HA₂) formed by each monomer. Each head domain carries a receptor binding pocket that is built by the 190 helix (aa 193–202) on the top, the 240 loop (aa 237–242) as the left boundary and the 140 loop (aa 136–143) as the right edge with additional contributions by the conserved Phe 95, Trp 158 and Tyr 202 residues. The principal receptor determinant of IBV HA are sialic acid (SA) conjugated to cellular surfaces. An interesting variation among clinical IBV isolates has been noted in terms of their receptor specificity by glycan array analysis. 19 of 21 Yamagata-like viruses isolated between 2001 and 2006 in Taiwan preferentially bound to α -2,6 linked sialic acid that is also the

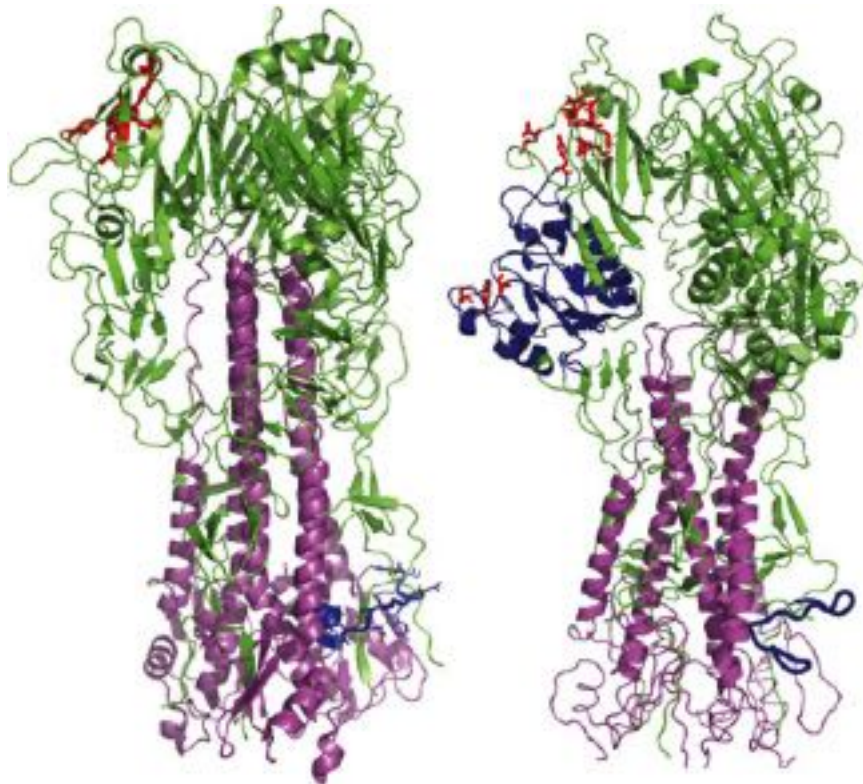


Fig. 3 Crystal structure of an HA (left) and HEF (right) trimers. Polypeptide chains are displayed as green (HA₁, HEF₁) and magenta cartoons (HA₂, HEF₂). HA (left): The amino acids contributing to receptor-binding (Phe-95, Ser-140, Trp-158, His-191, Tyr-202, Pro-238 and Ser-240) are shown as red stick in one HA₁ subunit. The fusion peptide at the N-terminus of the HA₂ subunit is displayed as blue sticks. HEF (right): The esterase domains shown in blue within one HEF₁ subunit. The catalytic triad of the esterase domain (serine 57, aspartate 352, histidine 355) is shown as red sticks. The amino acids contributing to receptor-binding (Tyr 127, Thr 170, Gly 172, Leu 184, Thr 186, Phe 225, Tyr 227, Arg 236, Phe 293) are shown as red stick in one HEF₁ subunit. The fusion peptide at the N-terminus of the HEF₂ subunit is displayed as a blue line. Figures were created with Pymol2.2 (<https://pymol.org/2/>) from pdb file 4M40 showing HA from influenza virus B/Yamanashi/166/1998 and from pdb file 1FLC containing C/Johannesburg/1/66.

major receptor determinant for IAV in the human upper respiratory tract. In contrast, 12 of 23 Victoria-like viruses additionally bound α -2,3 linked sialic acid, which is the receptor for avian IAV and can be found in humans mainly in the lower respiratory tract (LRT). 6 of 23 Victoria-like IBV recognized specifically sulfated glycans. Receptor specificity is considered an important determinant of influenza virus pathogenesis, but it remains to be established whether IBV strains with dual receptor specificity are more prone to infect the LRT which is usually associated with more severe symptoms.

The HA is a main target for antibodies elicited after infection or immunization. Phylogenetic and structural analyzes revealed four major canonical antigenic sites reflecting the strong immune pressures on the IBV HA. Antigenic sites have been mapped to the 120 loop (116–137), the 150 loop (141–150), the 160 loop (162–167), and the 190 helix (194–202) which contain sequence-variable immunodominant epitopes. The IBV HA protein contains a highly conserved cytosolic tail domain of 10 amino acids (RDNVSCSICL) with important regulatory roles in virion assembly and budding. The two cysteine residues are modified by palmitoylation thereby increasing hydrophobicity of the tail. Removal or shifting of palmitoylation sites affected membrane fusion activity of the HA. Studies by reverse genetic analyzes suggested that the cytosolic tail is important for the incorporation of IBV HA into virions and localization of viruses in lipid raft microdomains during entry. A deletion in the HA tail did not affect the fusion activity of HA, but yet it strongly reduced viral propagation and incorporation of HAs into virions. Interestingly, the phenotype of the HA truncation mutant was partially rescued by a compensatory E136K mutation in M1 suggesting that the assembly of HA into budding virions depends on an interaction with M1.

Neuraminidase (NA)

The neuraminidase, the receptor destroying enzyme of influenza A and B viruses, is a type II transmembrane glycoprotein found on the surface of virus-infected cells and in the membrane of the virions. Its main function is thought to be the removal of inhibitory receptor determinants from mucus in the respiratory tract to facilitate virion docking to the host cell, as well as to mediate the release of progeny virions from the host cell, which is the last step of the virus replication cycle. Hence, inhibition of NA by small molecules or antibodies diminishes virus propagation. There are many structural similarities between the NA molecules of type A and B viruses. In both viruses NA matures into a tetrameric structure despite the low degree of sequence identity of 30%. The monomers form a head domain composed of six-bladed β -propellers that are linked to membrane-proximal elongated stalks, which together build a box-like tetramer. The enzymatic and antigenic sites are located in the head domain. The eight NA amino acids that directly bind to sialic acid and are important for the catalytic activity of NA (R118, D151, R152, R224, E276, R292, R371, and Y406; N2 numbering) are highly conserved among type A and B viruses. NA is the main target for current antiviral therapy of human infections by circulating influenza A and B viruses (see below).

NB

NB is an enigmatic second transmembrane glycoprotein encoded by segment 6 with an unknown function. It consists of an N-terminal ectodomain of 18 amino acid, a predicted 22 amino acid transmembrane domain and a cytosolic part of 60 amino acids. It has two glycosylation sites in the ectodomain. Mature NB forms a tetrameric structure and is incorporated into virions, but only in low amounts as determined by quantitative mass spectrometry. NB transport to the cell surface was suggested to depend on palmitoylation of cysteine 49. Most interestingly, a recombinant influenza B/Florida/04/2006 virus engineered to lack NB expression by introduction of an early stop codon not affecting NA expression replicated well in stable cell lines or primary tissue cultures. The NB-deficient mutant virus had no specific phenotype in mice and propagated in a ferret model in virtually identical manner to the wildtype virus with a retained (low) capacity for respiratory droplet transmission. These findings partially contrast with a weakly delayed viral growth for a B/Lee/40 mutant virus in which NB palmitoylation at cysteine 49 was abrogated, which may have affected trafficking of NB to the cell surface. Yet, the likely essential function of the NB protein in IBV replication, pathogenesis or transmission, which is expected by the strong conservation of the NB reading frame, remains to be discovered.

Matrix protein (M1)

The M1 proteins are the most abundant morphogenic virion proteins of IBV and IAV. The IBV M1 protein has a conserved length of 248 amino acid and it shares about 30% sequence identity with its type A virus counterpart. The physicochemical properties, domain structure and regulatory roles of the IAV M1 protein in facilitating nucleocytoplasmic export of vRNPs late in infection have been well studied, but considerably less is known about the characteristics of the IBV M1. However, IBV M1 has been suggested to engage into a binding interaction with the viral BM2 protein and that this interaction is important for incorporation of M1 and vRNPs into virions. Mutational analysis identified within the IBV M1 protein a bipartite nuclear localization signal (amino acids 74–94) and two leucine-rich exportin 1-dependent nuclear export signals at positions 3–14 and 124–133. However, the role of these topogenic sequences in the viral replication cycle remain to be determined. In case of IAV the M1 protein is thought to bridge an interaction between vRNP and the viral nuclear export protein (NEP), which facilitates nuclear export of the viral genome via the exportin 1-pathway. However, the IBV NEP has a reported ability to bind directly to vRNPs suggesting that M1-mediated complex formation would be dispensable or redundant for IBV.

BM2

BM2 is a type III transmembrane protein that provides a proton channel function essential for the intracellular release of vRNPs from late endosomes. BM2 consists of a short 7 amino acid ectodomain, a 19 amino acid transmembrane domain and a cytosolic domain of 83 amino acids, and it exists in the form of a homotetramer. BM2 is a functional homolog with the IAV M2 protein to

which it has little sequence identity apart from a conserved 19-HxxxW-23 proton channel motif. BM2 was recently shown to efficiently complement the multi-cyclic growth of a recombinant M2-deficient influenza A virus, which was otherwise confined to a single cycle infection. BM2 differs from the M2 protein in being insensitive to channel blockers such as amantadine resulting in natural resistance of type B influenza viruses to this first generation class of anti-influenza compounds. The resistance to amantadine was explained by differences in the interior part of the proton channel, which is covered by polar amino acids and not hydrophobic residues as in the case of the IAV M2 protein, as well as by a reduced size of the channel pore. Apart from serving as a proton channel BM2 protein has also important morphogenic activity since recombinant viruses with C-terminal BM2 deletions of as little as six amino acids displayed irregularly formed virions and the virus was strongly attenuated in its replication ability. The large cytosolic BM2 domain adopts a coiled-coil tetramer. Biochemical analyzes showed that the C-terminus of BM2 is important for efficient membrane association of the M1 protein in flotation centrifugation, as well as incorporation of M1 and vRNPs into virion particles. Within the cytosolic BM2 domain, there is an electrostatic interaction with a positive charged N-terminal part (residues 44–71) and a negative C-terminus (72 – 103), the latter of which has been suggested to bind to the IBV M1 protein. From these analyzes, a model emerges in which the BM2 protein is a central factor that recruits viral RNPs and the M1 protein to virion assembly sites as a prerequisite for efficient budding. Whether BM2 also contributes to scission of virus particles from the plasma membrane as described for M2 for IAV remains to be investigated.

Non-structural protein 1 (NS1)

The NS1 protein is abundantly expressed in virus-infected cells and the main antagonistic protein enabling IBV to interfere with the innate type I interferon (IFN) defense and to propagate efficiently in mammalian hosts. Hence, there are strong parallels to the functions of the IAV NS1 protein although they share a sequence identity of less than 25%, and there are also type-specific activities of the two NS1 proteins. The IBV NS1 protein has a conserved length of 281 amino acids with a predicted molecular mass of 32 kDa. It is described as a non-structural protein although low levels of NS1 were detected in purified virions by mass spectrometric analysis. Early biochemical analyzes identified in the N-terminal NS1 domain an RNA binding activity (amino acids 1–93) and the ability to bind to the IFN-induced protein ISG15 (1 – 103), a ubiquitin-like modifier of nascent proteins. Later on, it was found that the C-terminal part of NS1 (94 – 281) can also bind RNA. Reverse genetic analysis confirmed that in epithelial cells the NS1 protein inhibits the accumulation of IFN- β transcripts and activation of IRF3, a central transcriptional regulator of IFN- β expression. For this activity the C-terminal NS1 domain was essential. However, IBV is a stronger activator of type I and III IFN expression early in infection compared to IAV. This depends on the cellular pattern recognition receptor RIG-I and it is possibly caused by a more rapid release of vRNPs from endosomes. Interestingly, the IBV NS1 protein does not bind RIG-I, but rather targets the ubiquitin E3 ligase TRIM25 that activates RIG-I by ubiquitination, whereas the type A NS1 protein binds RIG-I and TRIM25 directly. Phenotypic analyzes of recombinant mutant IBV demonstrated that NS1 facilitates efficient viral propagation in vitro and in vivo by blocking the activation of the central immune kinase PKR that can shut-off protein synthesis in infected cells. Silencing of PKR strongly depended on a binding interaction of the N-terminal NS1 domain with the kinase domain of PKR. Structural information from X-ray crystallography has been reported for the N-terminal domain (15 – 93) of NS1 that adopts a six helical dimeric fold and a large C-terminal fragment (141 – 281). It is uncertain whether the IBV NS1 protein forms higher order structures as was recently shown for the type A NS1. ISG15 has been shown to regulate influenza A and B virus infections. It was therefore not surprising to find out that the IBV NS1 protein not only binds ISG15, but it can also prevent its conjugation to target proteins. Interestingly, the IBV NS1 protein binds strongly to the human ISG15 and only poorly to ISG15 homologs in other mammalian species, which provides an important example of a species-specific viral antagonism of the type I IFN response.

The intracellular localization of the IBV NS1 protein is highly dynamic throughout the infection. At early times of infection, NS1 accumulates in nuclear SC35-positive speckle domains enriched in RNA processing factors and leads to a rounding of their usually irregular appearance. This activity was localized within the N-terminal 90 amino acids which also displays a monopartite NLS at position 46–57 inside the N-terminal RNA binding domain. However, speckle association of NS1 is resolved in late phases of infection when vRNPs are exported to the cytosol, and is accompanied by a cytosolic relocation to downregulate cellular immune receptors such as PKR and RIG-I. It has been speculated that the transient interaction of the IBV NS1 protein with speckles reflects a role in promoting export of viral mRNAs. The IBV NS1 protein lacks two important properties, which are found in IAV NS1 proteins: First, the absence of interference with the cellular polyadenylation machinery, which is mediated by binding to a 30 kDa subunit of cleavage and polyadenylation factor. Second, IBV NS1 has the inability to activate phosphatidylinositol-3-kinase (PI3K) signaling by binding to its p85 β subunit, which has been suggested to prevent premature apoptosis in infected cells.

NEP

The IBV NEP protein has a conserved length of 122–123 amino acids. In spite of its similar size the sequence identity of IBV NEP with IAV NEP is below 25%. However, both NEPs have conserved functions as vRNP export factors and regulators of the viral RNA polymerase. The IBV NEP has been shown to interact with human nucleoporins and exportin 1, and this interaction has been mapped in the hydrophobic N-terminal region at positions 10–19 of NES. In contrast to IAV, the IBV NEP has been reported to bind directly to vRNPs, and may, hence, be able to facilitate nucleocytoplasmic export of the viral genome independently of the M1 protein. An additional activity of NEP in regulating polymerase protein functions was observed in minigenome reporter assays, in which increasing amounts of NEP reduced the accumulation of viral mRNA and enhanced cRNA levels without affecting vRNA levels. Neither the precise target of IBV NEP within the viral RNPs nor the mechanism underlying this regulation have been established.

Epidemiology

IBV is considered to cause seasonal influenza epidemics every 2–4 years, but less comprehensive systematic analyzes on its epidemiology are available compared to IAV. Two IBV lineages derived from the prototypic B/Victoria/2/87- and B/Yamagata/16/88-strains co-circulate with the IAV H1N1 and H3N2 subtypes to cause seasonal influenza. However, the extent of human infections for each of the four viruses is highly variable and can differ from one geographic region to the other even in the same epidemic season. Sentinel surveillance data in Europe since 2001 showed that the detection of IBV among circulating influenza viruses ranged from 1% to 70% through all age groups and that IBV was in five seasons the most prevalent virus type. A long-term global comparison suggested that IBV is the most prevalent influenza virus type affecting older children (5–17 years) and is second behind A(H3N2) in the elderly (> 65 years). Moreover, in Germany the rates for hospitalization attributable to IBV infection were estimated to be 81 per 100,000, which was in between the rates for A(H3N2) and A(H1N1) influenza (99 and 56 per 100,000). Significantly, IBV infections were the main cause of mortality attributable to influenza in the US in four epidemic seasons between 1997 and 2009. IBV was detected in 1222 (73%) of the 1674 fatal influenza cases reported through the German Public Health notification system for infectious diseases in the severe 2017/2018 season. Taken together, studies show that IBV imposes a substantial burden of disease.

The IBV HA genes evolve at a rate of 2.0×10^{-3} substitutions/site/year, which is somewhat slower compared to the seasonal IAV HA subtypes H3N2 (5.5×10^{-3} substitutions/site/year) and H1N1 (4.0×10^{-3} substitutions/site/year). This finding correlates with a slower antigenic drift of IBV and may indicate higher structural and/or functional constraints on its HA. The B-Yamagata and B-Victoria lineages split in the 1970s and the lineages were distinguished based upon genetic and antigenic differences of their HA and NA genes (Fig. 4). During their co-circulation multiple reassortment events occurred either within or in between the lineages. Recent molecular genetic analyzes indicated that only the HA, PB1 and PB2 genes continue to evolve in a lineage-specific manner. Current viruses with the Yamagata-like HA have diverged into clades 2 and 3, whereas most strains with a

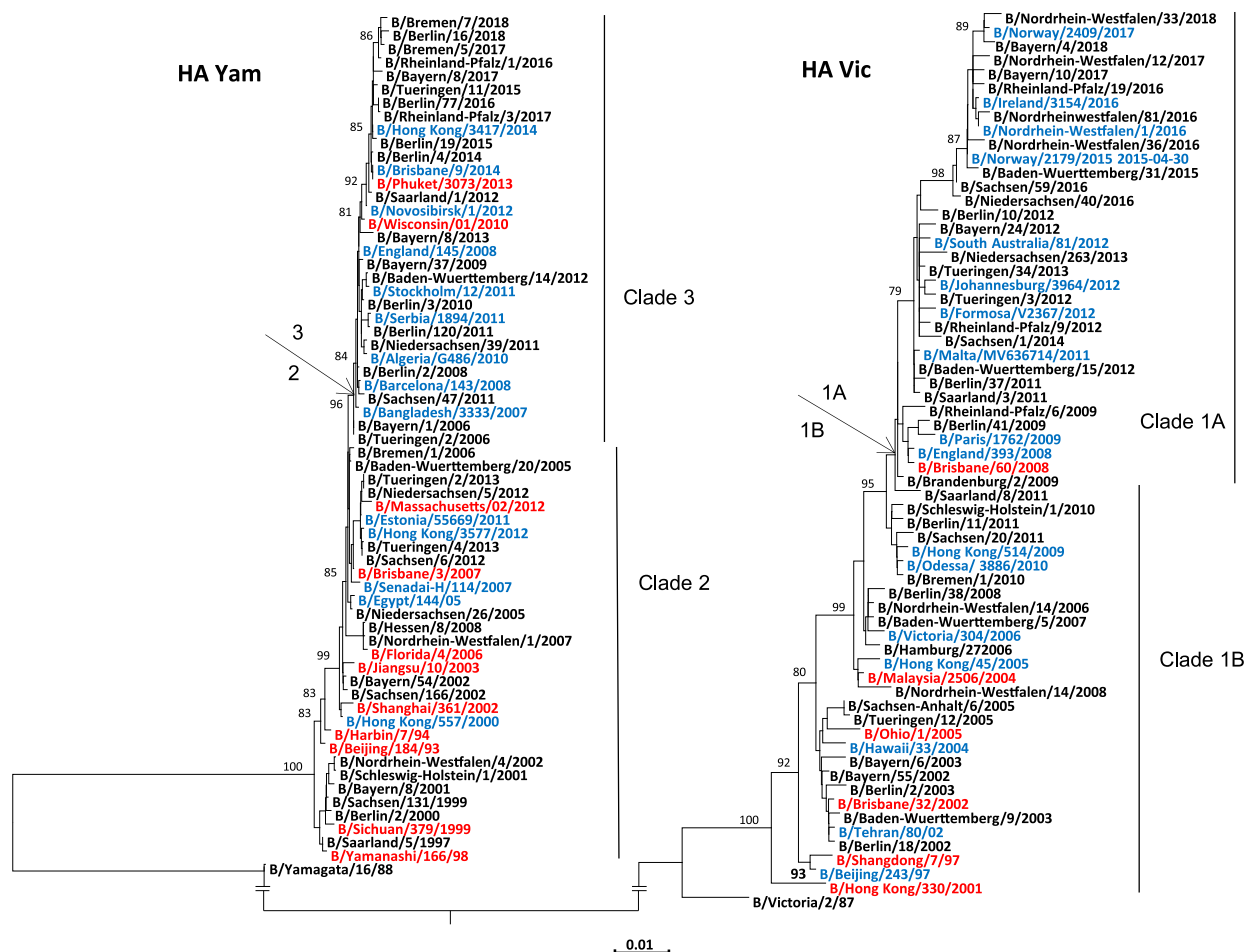


Fig. 4 Phylogenetic tree of IBV HA genes with representatives of B-Yam and B-Vic lineage using the Neighbor-joining method. Vaccine viruses (red), reference viruses (blue) and circulating isolates (black) are shown. The continued evolution of IBV HA genes is presented on <https://nextstrain.org>.

Victoria-like HA belong to clade 1A. There is low antigenic cross-reactivity between the two IBV lineages and often circulating strains and the IBV vaccine component in trivalent vaccines have not matched well.

Clinical Features

Infections by type B or A viruses cause febrile influenza-like illness characterized by a sudden onset of fever accompanied by cough, nasal congestion, headache, myalgia and other types of malaise that appear after a typical incubation period of 1–3 days. There is a large spectrum in the severity of symptoms ranging from an inapparent infection with no virus shedding to a fulminant disease, which may depend on the immune status, underlying risk factors, and age of the patient. Some studies on pediatric populations reported slightly increased frequencies of myalgia, sore throat and hoarseness in IBV infection compared to IAV, but other studies have not confirmed this observation. From outpatient-attended influenza infections or experimental analyses there are no data suggesting that the clinical features of IBV infection are different from the ones caused by IAV. Complications of IBV infection affect the lungs (e.g., bronchiolitis or pneumonia), but have also been observed in other organ systems leading to neurological (encephalitis), muscular (myositis) or cardiologic presentations (myocarditis).

Pathogenesis

Clinical symptoms of seasonal influenza are in general similar for IBV and IAV and no unique pathogenetic mechanism has been ascribed for IBV, when it infects and spreads through the human respiratory tract. One autopsy study of 45 fatal influenza B cases showed that both lineage viruses caused inflammation of the trachea and bronchi with necrotic cells, pulmonary edema and hemorrhages as common findings. Viral antigen was mainly found in the columnar epithelium of the trachea and bronchi as well as in submucosal glands even in the absence of bacterial coinfection. These findings were similar to the ones seen in fatal influenza A virus infection. Yet, it is unclear how representative these findings from post mortem tissue are to the pathogenesis in uncomplicated influenza for which there is only little data available. The pathogenesis of IBV infection in experimental animal models depend on virus adaptation (e.g., mice), the use of specific virus strains (ferrets, guinea pigs) or in some models there is no pathological changes (e.g., Syrian golden hamsters).

Diagnosis

It is not possible to diagnose an influenza B virus infection based on clinical presentation alone as there is a large range in symptom severity and a number of other respiratory pathogens that cause similar or identical illness. Hence, laboratory testing of respiratory patient samples such as nasopharyngeal swabs or broncho-alveolar lavage fluid is required to confirm a suspected IBV infection by the presence of viral RNA or antigen. In principle, reverse transcription-dependent real-time PCR assays, viral culture or rapid point-of-care tests detecting viral antigen or the viral genome are used, which differ in sensitivity and testing time. RT-PCR assays have a high sensitivity close to 100% and can be completed within a few hours, but usually RT-PCR analyzes require the samples to be transported to a diagnostic laboratory operating the required equipment (e.g., thermocyclers). PCR primers and probes often target the M gene to detect IBV and utilize the HA sequence for discrimination between the B-Yamagata and Victoria lineages. Results from rapid antigen detection or isothermal nucleic acid amplification may be available within 30 min, but these assays have on an average slightly lower sensitivity and specificity.

Treatment

There are two classes of compounds licensed for the treatment of acute influenza A and B virus infections, which target either the viral NA or the PA subunit of the viral RNA polymerase. The neuraminidase inhibitors (NAI) include the oral oseltamivir (given as the prodrug oseltamivir phosphate), the inhaled zanamivir and the intravenous peramivir, which all target the substrate binding pocket of NA and reduce enzymatic activity in the nanomolar range to stall the spread of newly synthesized virions to other host cells. It is a consistent finding that IC₅₀ values for IBV NA are up to tenfold higher compared to concentrations needed to inhibit IAV NA. For oseltamivir this observation has been explained by a lower flexibility of the active site residue E276 that impedes high affinity binding to the hydrophobic pocket. The development of resistance towards NAI in IBV has been observed in experimental studies as well as in patients undergoing antiviral treatment. However, the genetic barrier to acquire resistance mutations appears to be high, possibly due to a loss of viral fitness. Global surveillance by the GISRS detected NAI resistance in circulating IBV in five consecutive seasons between 2012–13 and 2016–17 at frequencies between 0.2% and 2.0% using phenotypic assays suggesting that these compounds remain suitable for clinical use. Single amino acid mutations conferring NAI-resistance in clinical isolates have been identified either directly at the catalytic site (R152K), at the NA framework (D198N) or near the sialidase center (G109E, G402S).

Baloxavir marboxil, the prodrug of an oral anti-influenza compound, was licensed for the treatment of seasonal influenza in Japan and the US in 2018. It targets the cap-dependent endonuclease activity located on the PA polymerase subunits of IAV and IBV, and the drug can be given as a single dose. The EC₅₀ of Baloxavir acid for the propagation of IBV in MDCK cells was reported in the range of 5.6–8.5 n mol. Analyses of A(H1N1) and A(H3N2) viruses isolated from patients in the treatment studies showed that single mutations of the conserved residue I38 to T, F or M reduced the sensitivity to Baloxavir by more than 10-fold. No resistant IBV were reported in the clinical studies, but reverse genetics showed that PA I38 mutations increase the EC₅₀ of IBV by 2- to 7-fold. Structural analyses showed that resistance mutations reduced the affinity of the inhibitor to PA. Strong conservation of the PA I38 position in current human influenza B viruses suggests only a low background level of resistance to Baloxavir. Importantly, no cross-resistance has been found for Baloxavir and NA1.

Prevention

Vaccination is the most effective measure to protect against seasonal influenza A and B, for which either inactivated or live-attenuated vaccines are available. Since IAV and IBV strains circulate in unpredictable patterns, killed vaccine formulations contain a mixture of inactivated whole viruses or purified HA and NA proteins from one A(H1N1), one A(H3N2) and one (trivalent influenza vaccine, TIV) or two IBV strains (quadrivalent influenza vaccine, QIV) designed to elicit a protective humoral immune response. The specific strains to be included into the vaccine of the forthcoming season are recommended by WHO twice a year based upon extensive genetic analysis and antigenic profiling of circulating virus strains. Live-attenuated vaccines contain a mixture of three or four attenuated virus strains with the same antigenic profiles. High vaccine effectiveness depends on a good antigenic match between the vaccine strains and the circulating epidemic viruses. The elderly may also have weaker immune responses against the vaccine. Since there is only limited antigenic cross-reactivity between the two IBV lineages that co-circulate since 2001 with unpredictable frequencies, it has turned out a formidable challenge to predict the specific IBV lineage component for TIV. In fact, TIV was in five of ten seasons from 2001 to 2010 reported to confer little or no protection against the circulating IBV strains in the US. Still, recent meta-regression analysis quantified that there was reduced but substantial IBV vaccine effectiveness against laboratory confirmed infection even in seasons with a type B lineage mismatch (67% for matched seasons, 35% for mismatched seasons). The suboptimal performance of TIV led to a first recommendation in 2012/2013 for the inclusion of both IBV lineages into a seasonal QIV, which have been increasingly used by different countries. Modeling studies suggested a good potential that the simultaneous coverage of both IBV lineages by quadrivalent formulations can substantially reduce the burden of influenza B infection if vaccine uptake is sufficiently high.

Influenza C and D viruses

Virion Structure and Genome

Influenza C (ICV) and D virus (IDV) particles exhibit various morphologies: elliptical, spherical with a diameter of 80–120 nm or filamentous with a similar diameter but with lengths in the μm range (**Fig. 1**). A peculiar feature of ICV and IDV particles not observed for IAV virions is the arrangement of its unique spike, the hemagglutinin-esterase-fusion glycoprotein (HEF) in a reticular structure that consist mainly of hexagons. The regular polymeric structure can also be observed in isolated HEF molecules indicating that the hexagonal arrangement is an intrinsic feature of the protein likely involving lateral interaction between its ectodomains.

The interior of the virus particle contains the viral ribonucleoprotein (RNP) complexes, which consists of negative-sense, single-stranded RNA bound to the nucleoprotein NP and to three polymerase proteins (**Fig. 5**). In contrast to IAV and IBV, the genomes of ICV and IDV consist of only seven (not eight) different segments (**Table 1**), but most virus particles nevertheless package eight RNPs arranged in the typical 1 plus 7 pattern in which a central RNP is surrounded by seven circularly arranged RNPs. It is still not known how specific packaging of genome segments is achieved, but the segmented genome provides an evolutionary advantage of allowing the exchange of individual genome segments with those of other strains.

All RNA segments are flanked by non-coding sequences, which are more variable in length than those of influenza A and B viruses. The first twelve nucleotides at each 3' end as well as the first eleven nucleotides at each 5' end are conserved between genome segments and are partially complementary to each other, which enables the single stranded RNAs to form "panhandle" structures. This unique structure serves as the promotor for transcription of mRNAs and plays a pivotal role in genome replication and packaging. A uridine-rich region at the 5' end of each segment serves as the poly A tail template for mRNA transcription.

The longest three gene segments encode the proteins PB2, PB1 and PA/P3 that form the trimeric polymerase complex, which is bound to one end of each RNP. PB1 houses the polymerase active site, whereas PB2 and PA/P3 contain cap-binding and endonuclease activities, respectively, required for transcription initiation by cap-snatching, a process during which a short nucleotide sequence is cleaved from the 5' end of cellular pre-mRNAs. The 3D-structure of polymerase proteins and the active site of the polymerase and endonuclease as determined by X-ray crystallography are conserved between influenza viruses suggesting that the mechanism of transcription and replication is very similar in all influenza virus types. The fourth segment of ICV encodes the glycoprotein HEF, the only spike of the viral membrane, which is discussed in more detail below. The fifth segment encodes

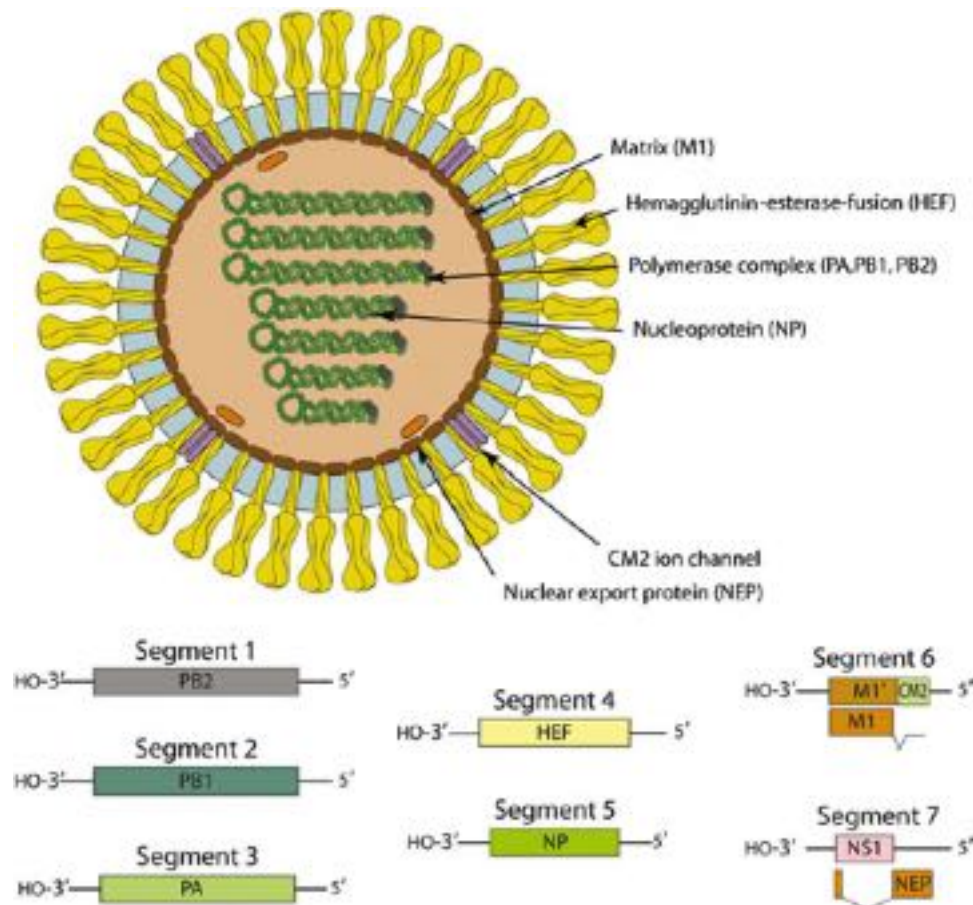


Fig. 5 Scheme of an ICV particle and its genome segments. Adapted from viralzone https://viralzone.expasy.org/81?outline=all_by_species with permission.

the nucleoprotein NP that associates with the viral genome segments in a double-helical conformation in which two NP strands of opposite polarity are associated with each other. The sixth segment encodes two proteins, the matrix protein M1, a peripheral membrane protein that covers the viral envelope on its inside, and CM2, a short transmembrane protein with proton-channel activity required for virus entry. M1 contributes to the morphology of virus particles and also to another unique feature of ICV, the formation of cord-like structures, very long (up to 500 μm) tubules containing HEF and M1, but no RNPs emanating from the surface of virus-infected cells.

In contrast to IAV, the larger M1 protein of ICV and IDV is translated from a spliced RNA, but the mechanisms of splicing differ between ICV and IDV. In ICV the removal of an intron generates a stop codon whereas in IDV splicing generates a second exon encoding a 4-residue peptide that is added to the primary M1 transcript. The unspliced mRNA encoding CM2 translates into a long precursor protein called M1' or p42. P42 contains an internal signal peptide, which targets the protein into the ER. Residues C-terminal to the signal peptide are translocated into the lumen of the ER until translocation is stopped by a second hydrophobic region, the transmembrane region of CM2. The signal peptide is then cleaved by signal peptidase yielding the mature CM2 and the p31 protein, which is rapidly degraded.

The seventh vRNA encodes the two non-structural proteins NS1 (246 amino acids) and NS2 (182 amino acids) that are generated via mRNA splicing: the unspliced mRNA is translated into NS1 and the spliced mRNA translates the shorter NS2 protein. The N-terminal residues of NS1 and NS2 are identical in sequence, splicing then generates a shift in the ORF in such a way that the remaining residues are translated from a different reading frame. Similar to IAV and IBV, it is believed that NS1 counteracts the cellular interferon response and NS2, which is also designated nuclear export protein (NEP), mediates the nuclear export of RNPs. Small amounts of NEP may be incorporated into virus particles.

Functions of the HEF Glycoprotein During the Life Cycle

Research on ICV and IDV has concentrated mainly on its only glycoprotein HEF that combines the function of HA and NA of IAV and IBV, i.e., it possesses receptor-binding, receptor-destroying and membrane fusion activities. HEF (like HA) is a typical type 1

transmembrane protein with a short N-terminal, cleavable signal peptide, a long ectodomain, a transmembrane region and a very short, cytoplasmic tail which contains (in contrast to HA of Flu B) the longer acyl chain stearate. HEF is synthesized as a precursor HEF₀, which is cleaved into the N-terminal HEF₁ subunit and the membrane-anchored HEF₂ subunit. Although there is only 12% amino acid identity between HA and HEF, crystallography of their ectodomains revealed that the overall structures as well as the folding of individual domains are quite similar. Both glycoproteins form a mushroom-shaped trimer consisting of a membrane-near stalk (containing the regions involved in membrane fusion) and a globular head domain carrying the receptor-binding site. HEF contains an additional bulge located at the lower part of the globular domain that carries the receptor-destroying region that is not present in HA (Fig. 3).

Influenza C virus does not use sialic acid as its receptor on the target cell, but an acetylated derivative, namely 9-O-acetyl-N-acetylneuraminic acid. 9-O-Ac-Neu5Ac as the terminal sugar on both glycoproteins and glycolipids can be used as a receptor, regardless of whether it is attached to galactose via an α -2,3 or α -2,6 linkage. There is some evidence that the abundance of 9-O-Ac-Neu5Ac influences the tropism, since MDCK II cells (a standard cell line used for growth IAV and IBV) is resistant to ICV infection due to insufficient number of receptors. Instead, ICV is usually grown and plaque-titrated in another subline of Madin-Darby canine kidney cell, namely MDCK I cells.

The receptor binding site is located near the top of the globular head in a shallow groove surrounded by residues from four secondary structure elements: the 170-loop, 190-loop, 230-helix and 270-loop and thus exhibits a similar structure as the binding site for Neu5Ac in HA (Fig. 3). Interestingly, the crystal structure of HEF of IDV, which is otherwise almost identical to the structure of HEF of ICV, revealed a more open receptor-binding cavity, which may be the basis for the broader cell tropism of IDV. Some coronaviruses, such as the prototype member mouse hepatitis virus and human and bovine coronaviruses, contain a hemagglutinin esterase (HE) protein that also binds 9-O-Ac-Neu5Ac, but the ligand is bound in an opposite orientation.

Intrinsic temperature sensitivity is another peculiar feature of HEF of ICV that distinguishes it from HA. Folding of the protein in the endoplasmic reticulum (and hence cell surface expression) is less efficient at 37°C than at 33°C. Probably as a consequence, ICV grows to higher titers in cells cultured at 33°C compared with 37°C. This temperature sensitivity may be one of the factors that prevents the replication of ICV in the lungs that has a higher temperature compared with the upper respiratory tract. On the contrary, HEF of IDV exhibits an exceptional temperature and acid-stability and hence IDV is the most stable of the four types of influenza viruses, a feature that likely affects the tissue tropism and cross-species transmission of the virus.

Influenza C virus is proposed to enter cells via the endocytic pathway and to subsequently fuse with the membrane of the endosome. This process is also mediated by HEF and requires its proteolytic cleavage into HEF₁ and HEF₂ subunits and subsequent exposure to mildly acidic pH. Although the fusion mechanism of HEF has not been investigated, it is believed to be similar to HA where well-studied conformational changes catalyze membrane fusion. This assumption is based on the very similar structures of HA and HEF at neutral pH and especially the presence of a long central α -helix in HEF₂ and a smaller helix on the outside packed in an antiparallel fashion, which undergo large acid-induced conformational changes that are the main drivers for membrane fusion (Fig. 3). The long α -helix contains a predicted heptad-repeat, a characteristic element of most viral fusion proteins that interact in the low pH structure to form a stable 6-helix coiled-coil domain.

A notable difference between HA and HEF is the fusion peptide, a hydrophobic region at the N-terminus of the small HEF₂ subunit which contains hydrophobic, aromatic, but also a few negatively charged residues. It is fully conserved between ICV and IDV HEF (IFGIDDLI), but very different from the respective, strictly conserved amino acids of HA2 (GLFGAIAAGFIE). In addition, while the fusion peptide of HA is completely buried in the interior of the molecule, the fusion peptide of HEF has a loop-like structure, which is partly exposed at the surface of the virus. HEF may therefore be regarded as having an internal fusion peptide, similar to virus fusion proteins that do not require cleavage activation.

HEF from both ICV and IDV possess a monobasic cleavage site (one arginine at the C-terminus of HEF₁) and it is in this respect similar to the HAs from mammalian and low pathogenic avian IAVs. The replication of these viruses is restricted to the site of virus infection, usually the respiratory tract. Interestingly, the amino acids surrounding the monobasic cleavage site of HEF differ between ICV and IDV. HEF from ICV contains three closely spaced basic residues at the C-terminus of HEF₁ (TVTKPKSR), whereas HEF₁ from IDV contains only two, widely spaced basic residues (RTLTPATR). The molecular features of the cleavage site in viral glycoproteins determine which of the many trypsin-like serine proteases recognize the precursor protein and the expression of this protease determines whether a certain cell type produces infectious virus particles. Thus, differential cleavage activation of HEF may contribute to the differences in host range of ICV and IDV, i.e., human versus cattle.

The last step in the viral replication cycle is the release of virus particles from infected cells, which requires the removal of all putative receptors. In influenza A and B the receptor-destroying activity is carried out by the second viral glycoprotein, the neuraminidase NA, whereas HEF performs this activity for ICV. In accordance with its receptor binding specificity, HEF is not a neuraminidase, but an esterase that cleaves acetyl from the C9 position of terminal 9-O-Ac-Neu5Ac. The esterase activity, which belongs to the class of serine hydrolase, is included in an additional domain of HEF₁. The esterase domain is structurally the most conserved part of HEF; the position of the catalytic triad characteristic for serine hydrolases (serine 57, aspartic acid 356 and histidine 359) is identical in ICV and IDV HEFs, which exhibit 55% amino acid identity and an almost identical 3D-structure. The structure of the esterase site is also very similar between HEF and coronavirus HE and it has been proposed that the HE arose from Influenza C-like HEF by a lateral gene transfer of the middle part of the HEF gene encoding mainly the receptor-binding and esterase domains.

Epidemiology

Although influenza C virus infections occur primarily in a pattern of sporadic cases or in limited outbreaks, serological studies indicated that this virus is widely distributed around the world and that the majority of humans develop antibodies against the virus early in life, which prevent subsequent infections with ICV or mitigate the symptoms. According to a phylogenetic analysis of their HEF genes the existing strains are divided into six genetic and antigenic lineages. The average rate of nucleotide substitution in HEF has been calculated to be 4.9×10^{-4} substitutions/site/year, an order of magnitude slower compared to the corresponding rate for the HA gene of IAV. Only 40 residues ($\sim 5\%$) are not completely conserved through all lineages of HEF, the variable amino acids are mainly located in the globular head domain of HEF₁. Reassortment between viruses of different lineages occurred frequently, and newly emerged reassortants replaced previously circulating viruses and became epidemic strains. The primary host and reservoir of influenza C virus are humans, but antibodies against influenza C virus are present in dogs and especially in pigs. A number of influenza C virus strains were isolated from pigs in China and these strains could be experimentally transmitted from pig to pig.

Infection of humans by IDV has not been reported, but based on serological studies it is assumed that the virus circulates widely in cattle, especially in calves and the virus may spread to pigs and small ruminants (sheep and goat). Based on the comparison of the nucleotide sequence of the HEF gene, two distinct co-circulating lineages are currently described, which frequently reassort with each other, but not with ICV. HEF of IDV is also stable in evolution, but the substitution rate for HEF from IDV is 1.54×10^{-3} per site and year, and thus faster than in the HEF of ICV. The reason for the slow evolutionary rate of HEF from ICV is unknown. It may be either a low error rate of the polymerases and/or that HEF does not tolerate many amino acid changes without compromising its functionality.

Clinical Features and Pathogenesis

Influenza C virus usually causes inflammation of the upper respiratory tract, especially in children from one to six years of age. Clinical symptoms, such as cough, fever and malaise are typically mild. Occasionally the virus can spread to the lower respiratory tract, causing symptoms such as bronchitis and pneumonia, particularly in children younger than two years.

Experimental infection of pigs and cattle with IDV causes only mild symptoms (rhinitis and tracheitis) and virus replication is usually restricted to the upper respiratory tract. This is somewhat in conflict with the observation that IDV is associated with a more pronounced disease in the field and it was thus proposed that the manifestation of clinical symptoms requires coinfection with other respiratory viruses or bacteria that colonize the respiratory tract.

Diagnosis

Due to the usually mild clinical symptoms no rapid diagnostic tests are in place for ICV. These infections are usually identified in respiratory patient samples in retrospective studies using RT-PCR targeting the viral M or NP genes and/or virus isolation in tissue culture. Also, infection by IDV is mainly detected by RT-PCR analyzes of respiratory samples such as nasal and pharyngeal swabs or lung tissue.

Treatment and Prevention

Drugs developed against influenza A virus are not effective against ICV or IDV, since their target molecule is not present (neuraminidase inhibitors) or there are different amino acids at the binding site (M2 inhibitors). Since ICV is not thought to be a significant concern for human health, no vaccines have been developed. Prototype vaccines against IDV are currently under development.

Further Reading

- Burnham, A.J., Baranovich, T., Govorkova, E.A., 2013. Neuraminidase inhibitors for influenza B virus infection: Efficacy and resistance. *Antiviral Research* 100, 520–534.
- Carrat, F., Vergu, E., Ferguson, N.M., *et al.*, 2008. Time lines of infection and disease in human influenza: A review of volunteer challenge studies. *American Journal of Epidemiology* 167, 775–785.
- Hause, B.M., Ducatez, M., Collin, E.A., *et al.*, 2013. Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses. *PLoS Pathogens* 9, e1003176.
- Herrler, G., Klenk, H.D., 1991. Structure and function of the HEF glycoprotein of influenza C virus. *Advances in Virus Research* 40, 213–234.
- Herrler, G., Rott, R., Klenk, H.D., *et al.*, 1985. The receptor-destroying enzyme of influenza C virus is neuraminidase-O-acetyltransferase. *The EMBO Journal* 4, 1503–1506.
- Langat, P., Raghvani, J., Dudas, G., *et al.*, 2017. Genome-wide evolutionary dynamics of influenza B viruses on a global scale. *PLoS Pathogens* 13, e1006749.
- Muraki, Y., Hongo, S., 2010. The molecular virology and reverse genetics of influenza C virus. *Japanese Journal of Infectious Diseases* 63, 157–165.
- Nakatsu, S., Murakami, S., Shindo, K., *et al.*, 2018. Influenza C and D viruses package eight organized ribonucleoprotein complexes. *Journal of Virology*. 92.

- Paul Glezen, W., Schmier, J.K., Kuehn, C.M., Ryan, K.J., Oxford, J., 2013. The burden of influenza B: A structured literature review. *American Journal of Public Health* 103, e43–e51.
- Pflug, A., Lukarska, M., Resa-Infante, P., Reich, S., Cusack, S., 2017. Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Research* 234, 103–117.
- Reich, S., Guilligay, D., Pflug, A., *et al.*, 2014. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 516, 361–366.
- Rosenthal, P.B., Zhang, X., Formanowski, F., *et al.*, 1998. Structure of the haemagglutinin-esterase-fusion glycoprotein of influenza C virus. *Nature* 396, 92–96.
- Su, S., Fu, X., Li, G., Kerlin, F., Veit, M., 2017. Novel influenza D virus: Epidemiology, pathology, evolution and biological characteristics. *Virulence* 8, 1580–1591.
- Wang, J., Pielak, R.M., McClintock, M.A., Chou, J.J., 2009. Solution structure and functional analysis of the influenza B proton channel. *Nature Structural & Molecular Biology* 16, 1267–1271.
- Wang, Q., Tian, X., Chen, X., Ma, J., 2007. Structural basis for receptor specificity of influenza B virus hemagglutinin. *Proceedings of the National Academy of Sciences of the United States of America* 104, 16874–16879.

Relevant Websites

- <https://www.gisaid.org/>
GISAID.
Global Initiative on Sharing All Influenza Data.
- <https://www.fludb.org/brc/home.spg?decorator=influenza>
Influenza Research Database.
- <https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database>
Influenza Virus Database.
- <https://nextstrain.org>
Nextstrain.org.
- <https://reactome.org/PathwayBrowser/#/R-HSA-168254>
Reactome Pathway Database.
- <https://viralzone.expasy.org/>
ViralZone root.
ExPASy.
- <https://www.who.int/influenza/en/>
WHO.
Influenza.
World Health Organization.

Jaagsiekte Sheep Retrovirus (*Retroviridae*)

James M Sharp, University of Zaragoza, Zaragoza, Spain and Edinburgh, United Kingdom

Marcelo De las Heras, University of Zaragoza, Zaragoza, Spain

Massimo Palmarini, MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Thomas E Spencer, University of Missouri, Columbia, MO, United States

© 2021 Elsevier Ltd. All rights reserved.

History

Jaagsiekte sheep retrovirus (JSRV) is the causative agent of a naturally occurring lung adenocarcinoma of sheep known as ovine pulmonary adenocarcinoma (OPA, also known as jaagsiekte or sheep pulmonary adenomatosis). OPA was recognized for the first time in South Africa in the nineteenth century as a cause of dyspnea in herded sheep, hence the origin of the Afrikaans name 'jaagsiekte', meaning driving (=jaagt) sickness (=ziekte). OPA is one of the original 'slow diseases' of sheep (along with scrapie, maedi-visna and paratuberculosis) originally described in the 1930s by the Icelandic physician Björn Sigurdsson. The slow diseases of sheep were of great biological importance as they allowed, for the first time, the recognition that an infectious agent could cause clinical disease many months or years after the initial infection of the host.

Studies on JSRV were hampered for years by the lack of a suitable tissue culture system for the cultivation of JSRV. The isolation of full-length JSRV molecular clones (JSRV₂₁ and JS7) allowed the *in vitro* generation of infectious viral particles by a transient transfection system, which has sparked a variety of studies that have elucidated many aspects of the molecular biology of JSRV. JSRV is a remarkable virus in many respects. It is the only known virus that induces a naturally occurring lung adenocarcinoma. In addition, the JSRV envelope glycoprotein is an oncoprotein. This is a unique example among retroviruses and oncogenic viruses in general in which a structural protein functions also as a dominant oncoprotein.

Classification

JSRV belongs to the genus *Betaretrovirus* within the family of the *Retroviridae*. Retroviruses are divided on the basis of their modality of transmission as "exogenous" and "endogenous" viruses. Exogenous retroviruses are horizontally transmitted between infected and uninfected hosts. Endogenous retroviruses are stably integrated in the genome of the host species from which they derive, are usually defective, and are transmitted vertically like any other Mendelian gene. JSRV is an exogenous retrovirus and it is highly related to enzootic nasal tumor virus (ENTV) of sheep (ENTV-1) and goats (ENTV-2). JSRV and ENTV have common pathogenic characteristics, and both cause low-grade adenocarcinomas of secretory cells in different portions of the respiratory tract of small ruminants. Interestingly, sheep, goats, and other members of the *Caprinae* have several copies of nonpathogenic JSRV-related "endogenous" retroviruses (commonly referred as enJSRVs) stably integrated in their genome. The phylogenetic relationship between JSRV, ENTV, and enJSRVs is indicated in [Fig. 1](#).

Genetic Organization and Virion Proteins

The JSRV virions are enveloped particles of approximately 100 nm in diameter and a density by isopycnic centrifugation in sucrose gradients of 1.15 g ml⁻¹ for particles obtained from cell cultures. Virions purified from lung secretions of OPA-affected sheep have a slightly higher density (1.16–1.18 g ml⁻¹).

JSRV has the typical genomic organization of a simple retrovirus; the genomic RNA of 7455 nt (in the JSRV₂₁ infectious molecular clone) contains the canonical retroviral genes *gag*, *pro*, *pol*, and *env* ([Fig. 2](#)). Apart from Env, few studies have been undertaken to assign functions to the JSRV proteins and these are assumed to be the same as other retroviruses. The *gag* gene encodes the structural proteins of the viral core. Gag is expressed as an immature polypeptide that is cleaved upon exit by the viral protease. The JSRV Gag is cleaved into at least five proteins: MA(p23), p15, CA (p26), NC, and p4. MA is myristoylated and presumably interacts with the cell membrane during viral egress. CA is the major capsid protein, while NC interacts tightly with the genomic RNA although no specific studies have been conducted on the JSRV NC. The *pro* gene overlaps *gag* and encodes most probably a dUTPase (DU, deoxyuridine triphosphatase) and the viral protease (PR). The main function of DU is to avoid misincorporation of uracil into DNA during reverse transcription (see below). As mentioned above, PR cleaves the Gag polypeptide upon exit and it is absolutely required in order to obtain mature infectious viral particles. The *pol* gene overlaps *pro* and is predicted to encode the viral reverse transcriptase (RT) and integrase (IN). Both RT and IN are virion-associated enzymes; RT copies the viral single-stranded RNA genome into double-stranded DNA, while IN serves to join the proviral DNA into the host genome. RNaseH is a subdomain of RT and serves to degrade the viral genomic RNA once it has been copied into DNA.

An additional open reading frame (*orf-x*) overlaps *pol* and has some unusual features, including a codon usage different from other genes within JSRV, and a very hydrophobic predicted amino-acid sequence that shows no strong similarities to any known

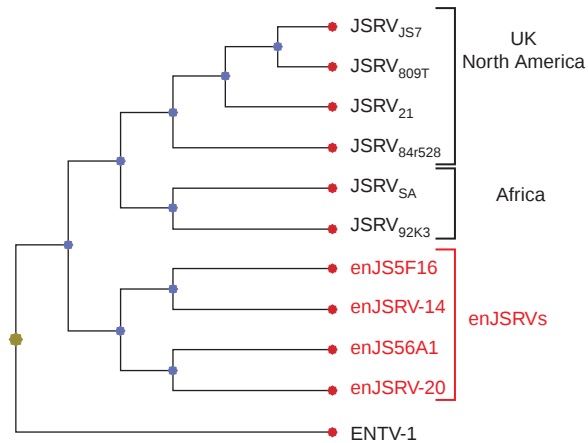


Fig. 1 Phylogenetic analysis of sheep betaretroviruses. A phylogenetic tree of representative small ruminant betaretroviruses was derived by the neighbor-joining method using JSRV, ENTV-1, and enJSRVs *env* aligned using Clustal W. JSRV, enJSRVs, and ENTV-1 cluster in three distinct phylogenetic groups. Note that JSRV isolates from Africa cluster in separate branches from European and North American isolates.

protein and only weak homology to a member of the G protein coupled receptor family. The role of this open reading frame is unknown. Orf-x is conserved among all of the exogenous JSRVs examined to date, but it does not seem to be required for JSRV replication *in vitro* nor for cell transformation either *in vitro* or *in vivo*.

The *env* gene encodes the glycoproteins of the viral envelope (Env). The JSRV Env is formed by two subunits, the surface domain (SU) which interacts with the cellular receptor and mediates viral entry, and the transmembrane domain (TM), which fixes SU to the lipid bilayer. A unique feature of the JSRV *env* is that it functions essentially as a dominant oncogene, and its sole expression is sufficient to induce cell transformation *in vitro* and *in vivo* (see below).

According to some studies, an additional protein, named Rej or JSRV Env signal peptide, is also encoded with in the *env* gene. It moderately enhances nuclear export of unspliced viral RNA and considerably increases viral particle release by favoring a post-transcriptional step of the replication cycle. Noncoding regions are present at the 5' and 3' end of the genome. The R region is repeated at the 5' and 3' end of the genome. U5 is present uniquely at the 5' end, and the U3 is present at the 3' end of the genome. The viral promoter and enhancer regions are present in the U3 (see below).

Genomic sequence variability among JSRV strains is very low. For instance, the infectious molecular clones JSRV₂₁ and JSRV_{JS7} are 99.3% identical along the entire genomic sequence, and they were derived from naturally occurring OPA cases from Scotland collected many years apart. JSRV₂₁ and JSRV_{JS7} Env proteins are 100% identical. JSRV isolates from Africa can be distinguished phylogenetically from the UK and North American isolates, although there is still a high degree of homology (~93% along the entire nucleotide genomic sequence) among the two groups.

Replication Cycle

In general, the replication cycle of JSRV is not thought to be markedly different from other betaretroviruses. The lack of a suitable tissue culture system for the propagation of JSRV has not allowed detailed studies on the replication of this virus.

JSRV interacts with a specific cellular receptor to enter the cell. The use of retroviral pseudotypes identified hyaluronidase 2 (HYAL2) as the cellular receptor for JSRV entry. HYAL2 is a glycosylphosphatidylinositol (GPI) linked membrane protein with a low hyaluronidase activity and is widely expressed on many different cell types. Besides ovine HYAL2, the human ortholog also allows JSRV entry, while mouse and rat Hyal2 do not.

Detailed steps of JSRV entry and post-entry would need more clarification. However, using pseudovirus of Moloney leukemia murine virus (MoMlv) it was shown that JSRV-mediated entry is a relatively slow and non-classical pH-dependent process that requires dynamin-associated endocytosis. Fusion activity of the JSRV Env protein is dependent on low pH and it is modulated by the cytoplasmic tail. Despite this particular aspects of JSRV other steps are likely to be similar to those of other retroviruses including uncoating, reverse transcription of the viral genome into double-stranded DNA, entry into the nucleus of the pre-integration complex, and stable integration of the proviral DNA into the host DNA. During the process of reverse transcription, the noncoding regions at the 5' and 3' ends of the genome (R, U5, and U3) are duplicated and give origin to the viral long terminal repeats (LTRs).

The retroviral LTRs are major determinants of retrovirus tropism, as the 5' LTR of the provirus initiates transcription and the U3 region contains the majority of *cis*-acting sequences interacting with the cellular RNA polymerase II and with cellular transcription factors. The exogenous JSRV LTRs are particularly active in reporter assays using type II pneumocytes/Clara cell lines and interact with lung-specific transcription factors such as forkhead box A2 (FOXA2; alias HNF-3 β). These features may explain the preferential expression of exogenous JSRV in the transformed cells of the lungs, which have phenotypic characteristics of type II pneumocytes and Clara cells.

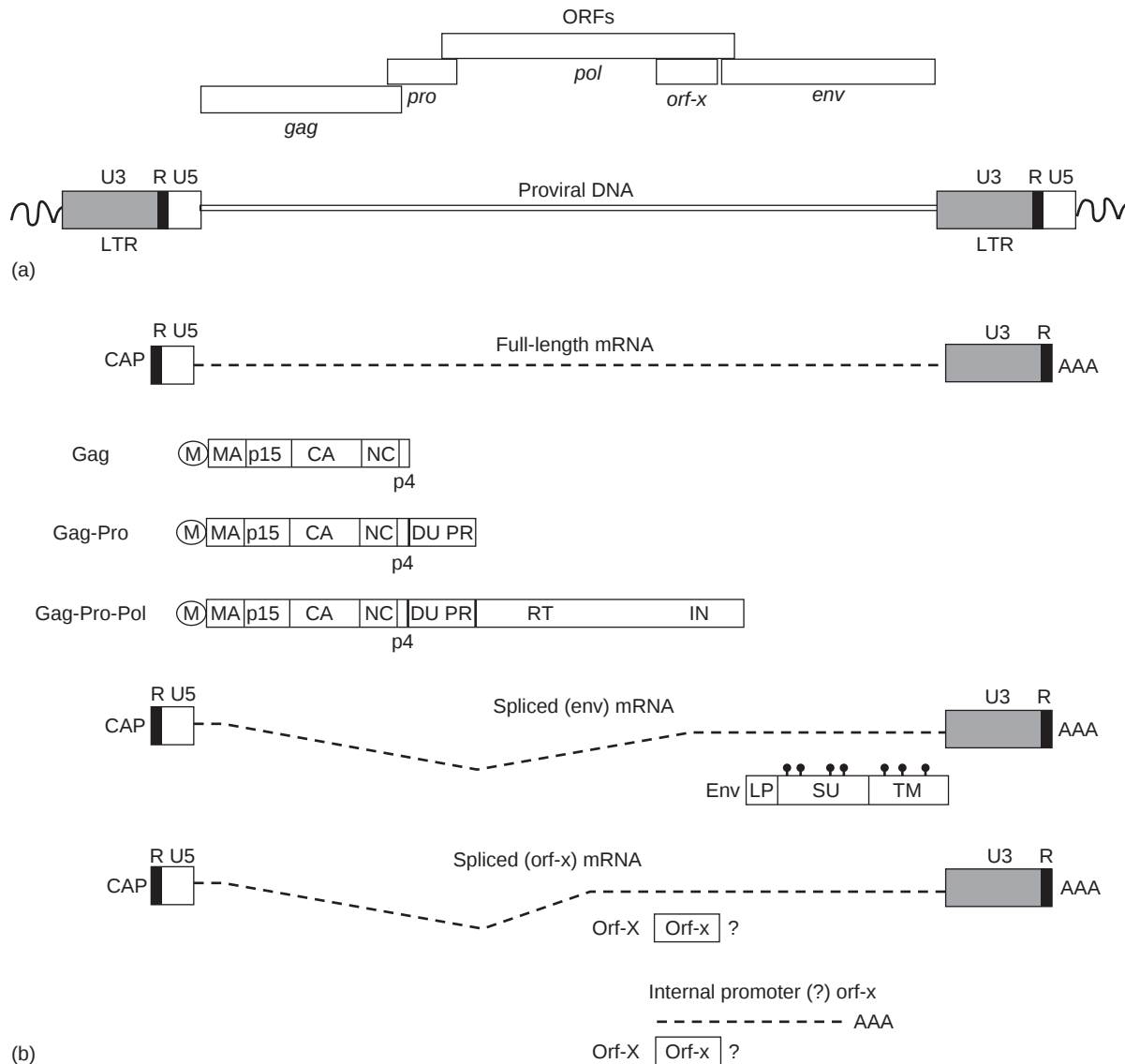


Fig. 2 Genomic organization, mRNAs and viral proteins of JSRV. (a) Schematic organization of the JSRV open reading frames (ORFs) and their relative position in the JSRV provirus. The JSRV LTR, repeated at both ends of the provirus genome, is divided into U3, R and U5. (b) Major JSRV RNAs and proteins. The JSRV provirus transcribes various mRNAs. A full-length mRNA encodes for the Gag, Pro and Pol proteins and serves as genome for the newly synthesized viral particles. A spliced mRNA encodes the viral Env. The Env glycoprotein is glycosylated, and putative glycosylated sites are indicated by full circles. Two additional mRNAs have been detected and encompass the *orf-x* reading frame, but their biological significance is unknown at present. MA, matrix; CA, capsid; NC, nucleocapsid; DU, dUTPase; PR, protease; RT, reverse transcriptase; IN, integrase; LP, leader peptide; SU, surface; TM, transmembrane.

The expression of JSRV, as in all retroviruses, is believed to follow the basic transcriptional events of cellular mRNAs including capping of the 5' end and polyadenylation at the 3' end. JSRV encodes a full-length mRNA that serves as genome of the viral progeny and is also translated into the Gag polyprotein and the proteins encoded by *pro* and *pol*. The latter are expressed as fusion proteins with Gag. As in all betaretroviruses, *pro* and *pol* are in different open reading frames to *gag* and are expressed by ribosomal frameshifting. The viral Env is produced from a mRNA which is spliced using the splice donor in the untranslated *gag* region and a splice acceptor immediately before *env*.

Another spliced mRNA has been found to use the same splice donor of the *env* mRNA and a splice acceptor before the *orf-x* reading frame. This mRNA presumably expresses a protein encoded by *orf-x*, but there are no published data supporting this assumption. A possible mRNA expressing *orf-x* has been found to lack the R-U5 regions (typical of all other mRNAs starting from the U3) and is thought to derive from the expression of an internal promoter that has not been characterized. Other mRNAs deriving from the use of secondary splice acceptors and non-conventional polyadenylated sites have been detected but their biological significance, if any, is unknown.

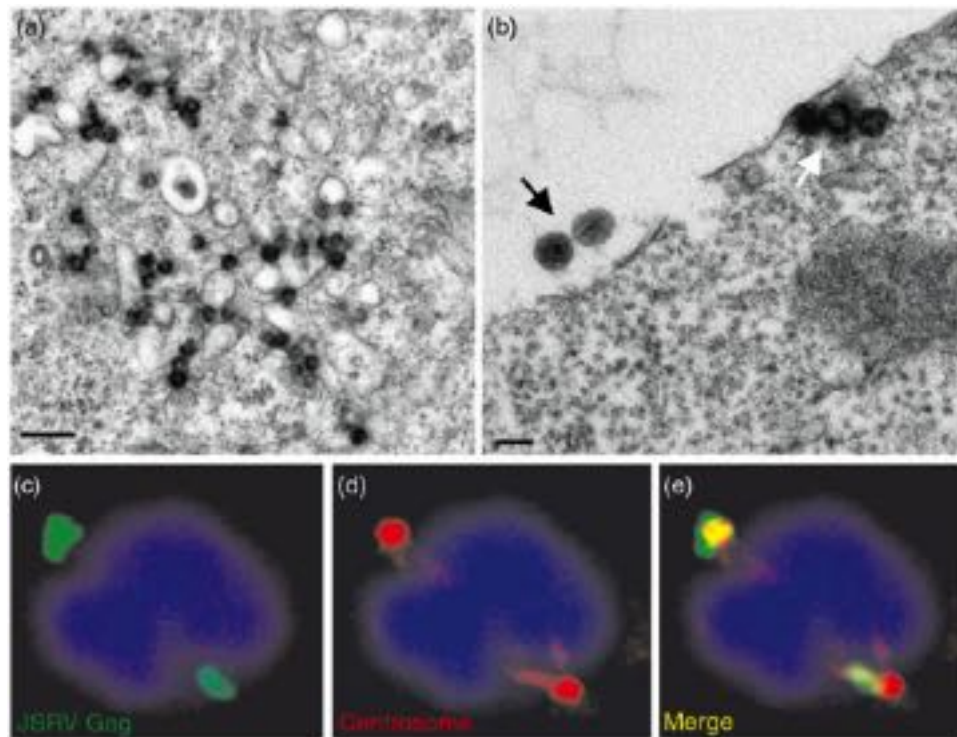


Fig. 3 Electron and confocal microscopy of cells expressing JSRV. Electron microscopy in 293T cells transiently transfected with the JSRV₂₁ infectious molecular clone showing intracytoplasmic particles (a; scale = 200 μ m). In panel b, intracellular particles in the vicinity of the cell membrane (white arrow) and extracellular particles complete with viral envelope (black arrow) are also visible (scale = 100 μ m). (c–e) Confocal microscopy in HeLa cells transiently transfected with JSRV and probed with an antiserum towards the JSRV matrix (c) and γ -tubulin (d). The JSRV Gag concentrates in the pericentrosomal area (e).

JSRV assembles in the cytoplasm (**Fig. 3(a)** and **(b)**) like all betaretroviruses, most likely in the pericentriolar region (**Fig. 3(c)–(e)**). Other retroviruses such as lentiviruses and gammaretroviruses assemble instead mostly at the cell membrane. The JSRV intracellular particles interact with Env at the cell membrane or in a cellular compartment not yet fully identified and egress from the cell. Upon exit, the Gag polyprotein is cleaved by the viral protease into the mature proteins described above. A more detailed description of the replication cycle of retroviruses is discussed elsewhere in this encyclopedia.

enJSRVs

The sheep genome contains at least 27 copies of endogenous retroviruses highly related to JSRV (hence the name enJSRVs). Endogenous retroviruses are believed to derive from integration events of ancestral exogenous retroviruses into the germline of the host. enJSRVs have a high degree of homology with JSRV. For example, JSRV₂₁ and enJS56A1 are 92% identical at the nucleotide level along the entire genome. Major differences are located in the U3 region, in two regions in Gag (termed variable regions 1 and 2), and in the cytoplasmic tail of the transmembrane domain of the Env (termed variable region 3). At least some of these highly divergent regions are the basis of important biological differences between JSRV and enJSRVs.

enJSRV transcripts have been detected in most tissues by sensitive reverse transcriptase polymerase chain reaction (RT-PCR) assays. However, high levels of enJSRVs (both mRNA and proteins) are found specifically in the epithelia of the genital tract of the ewe and, particularly, in the epithelia of the uterus (**Fig. 4**). In the placenta, enJSRVs are expressed in the mononuclear trophoblast cells of the conceptus (embryo/fetus and associated extra-embryonic membranes), but are most abundant in the trophoblast giant binucleate cells (BNCs) and multinucleated syncytial plaques of the placentomes. The temporal expression of enJSRVs envelope (*env*) gene in the trophoblast is coincident with key events in the development of the sheep conceptus. Indeed, using a morpholino antisense oligonucleotide to induce loss-of-function *in utero*, enJSRVs Env knockdown caused a reduction in trophoblast outgrowth and inhibited/prevented trophoblast giant binucleate cell differentiation during blastocyst elongation and formation of the conceptus. Thus, enJSRVs have been proposed to be essential in sheep for *peri*-implantation growth and differentiation.

enJSRVs expression appears to be regulated by progesterone and expression of the progesterone receptor. However, the specific enJSRV loci that are transcriptionally active are not known at present. enJSRVs are also expressed in the sheep fetus, supporting the idea that sheep are tolerized toward JSRV. Indeed, JSRV-infected naïve sheep (with or without lung adenocarcinoma) have no

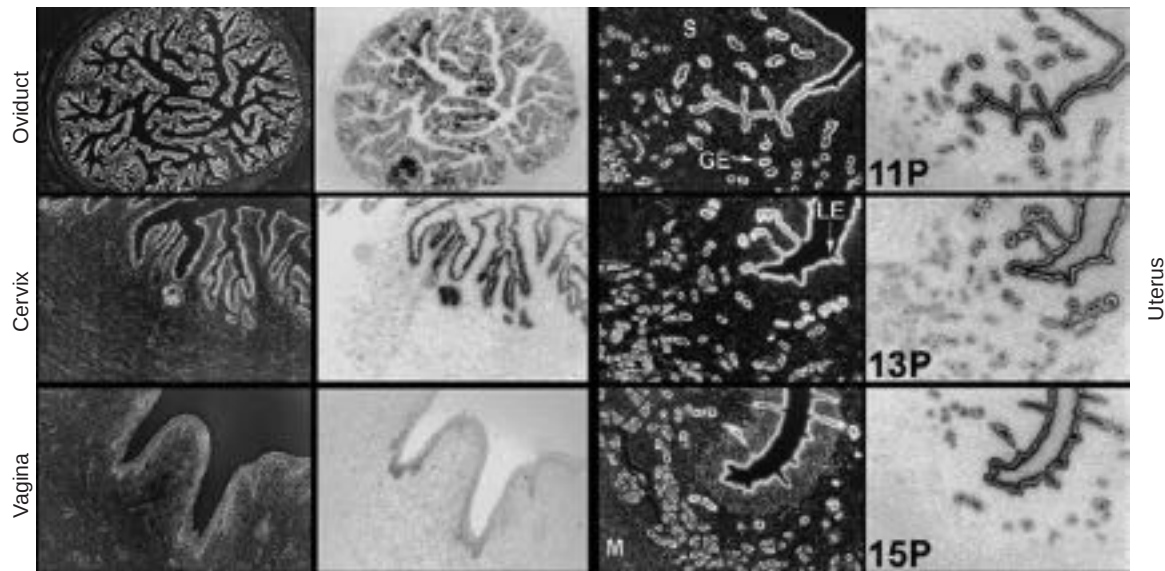


Fig. 4 enJSRVs are highly expressed in epithelia of the genital tract of the ewe. *In situ* hybridization analysis of enJSRVs *env* mRNA in the oviduct, uterus, cervix, and vagina. Note the specific expression of enJSRVs *env* mRNA in the epithelia lining the oviduct, cervix, and vagina of the cycling ewe. During the estrous cycle and pregnancy (P), enJSRVs mRNA is particularly abundant in the luminal and glandular epithelia of the uterus. LE, luminal epithelium; GE, glandular epithelium; M, myometrium; S, stroma. Modified from Palmarini, M., Gray, C.A., Carpenter, K., *et al.*, 2001. Expression of endogenous betaretroviruses in the ovine uterus: Effects of neonatal age, estrous cycle, pregnancy, and progesterone. *Journal of Virology* 75 (23), 11319–11327.

detectable specific humoral or cellular immune responses, although recombinant JSRV CA, in the presence of adjuvants, can induce antibody production and specific T-cell responses in vaccinated sheep.

Three full-length enJSRV loci have been cloned and characterized (enJS56A1, enJS5F16, and enJS59A1). All three loci have deletions or stop codons that make them replication incompetent. However, enJS56A1 and enJS5F16 maintain intact ORFs for *gag* and *env*. By using retroviral vectors pseudotyped by the enJSRVs Env, it was found that they too use HYAL2 as a cellular receptor and interfere by receptor competition with JSRV entry.

One of the enJSRVs loci, enJS56A1, is defective for viral exit when overexpressed in transfected cells, although abundant intracytoplasmic Gag is detected and intracytoplasmic viral particles are visible by electron microscopy. The replication defect of enJS56A1 is determined by its Gag protein, which is *trans*-dominant over the exogenous JSRV Gag if co-expressed in the same cell. A tryptophan residue in position 21 of the enJS56A1 Gag (replacing an arginine in JSRV) is the main determinant for the block induced by enJS56A1. Thus, enJS56A1 exerts a unique mechanism of retroviral interference, which occurs at a late step of the replication cycle. The mechanism and timing of the block induced by enJS56A1 are not yet understood. However, the observation of viral particles by electron microscopy in cells expressing enJS56A1 (or co-expressing enJS56A1 and JSRV) suggests that enJS56A1-induced interference depends on a defect in Gag trafficking. Recent studies suggest that enJS56A1 appears to block JSRV, most likely in *trans*, by hampering the ability of the latter to reach the centrosome, the proposed site of assembly for betaretroviruses.

enJSRVs are not expressed in the differentiated epithelial cells of the lungs but are expressed in the epithelium of the genital tract. As mentioned above, enhancer regions are located in the U3 and indeed enJSRVs transcription is regulated by progesterone *via* the progesterone receptor. Interestingly, the enJSRVs LTRs do not respond to lung transcription factors (such as FOXA2) unlike the exogenous JSRV LTRs. Thus, the LTR appears to be a major contributor to the different tropisms (genital tract *vs.* respiratory tract) shown by enJSRVs and JSRV.

The enJSRVs Env (or at least the Env of those enJSRVs loci cloned so far) is not able to transform cells *in vitro*, unlike the highly related JSRV Env. Indeed, the cytoplasmic tail of the JSRV Env is where major determinants of transformation are located. In particular, an SH2 binding domain is present in the exogenous JSRV Env, but is absent in the enJSRVs Env sequenced to date. It is hypothesized that enJSRVs have been selected in the sheep for their ability to confer resistance to infection of the host by the related exogenous betaretroviruses. This innate resistance could have provided a selection pressure for betaretroviruses with tropism towards the respiratory tract (*i.e.*, the current JSRV) rather than the genital tract.

Ovine Pulmonary Adenocarcinoma

OPA is a naturally occurring lung cancer of sheep that has been reported in most sheep-rearing countries. It is absent from Australia and New Zealand and has been eradicated from Iceland. The disease is characterized by a progressive respiratory condition caused by the growth of the lung adenocarcinoma.

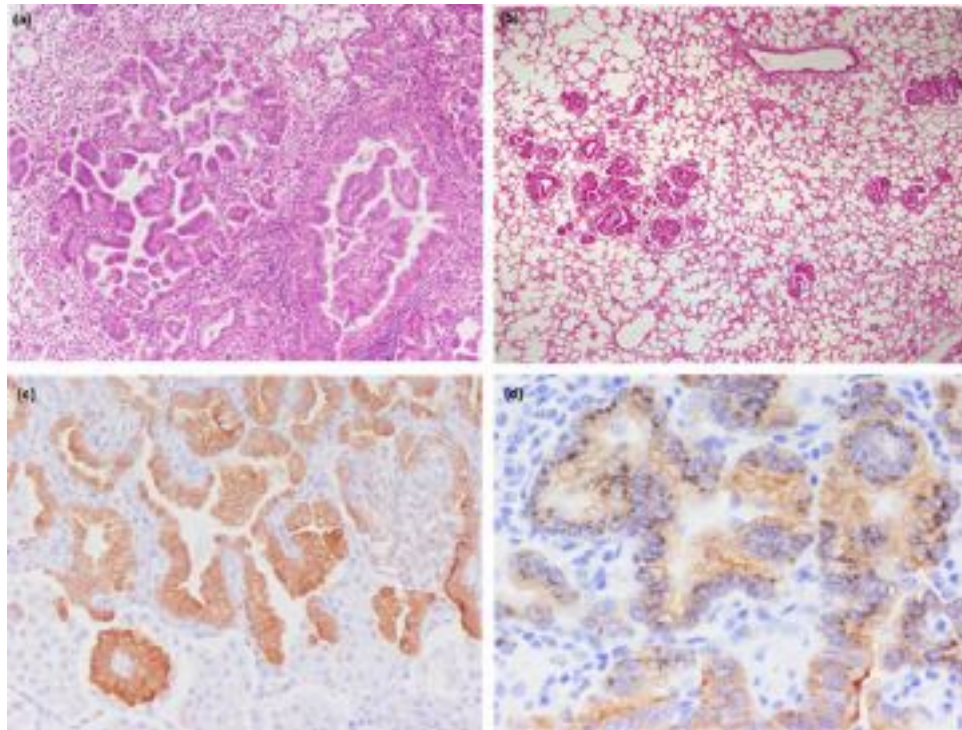


Fig. 5 JSRV-induced tumors *in vivo*. Histology from a lung tumor section of naturally occurring cases of OPA (a) and experimentally reproduced tumor in young lambs (b). Sections were stained with routine hematoxylin and eosin. Note the presence of papillary to acinar neoplastic lesions that replace the normal alveolar and bronchiolar structures of the lungs. Immunohistochemistry (IHC) in lung tumor sections from natural OPA affected sheep showing the expression to the viral Env (characterized by membrane and intracytoplasmic brown color) (c). Double IHC in the lung sections from the same tumor combining surfactant protein C marker (black dots), a marker for alveolar type II cells, and Env JSRV protein marker (light brown) (d). Carazzi's hematoxylin was used as a counterstain.

In vivo JSRV infects differentiated and undifferentiated epithelial cells of the terminal bronchioli and alveoli. Natural tumors are heterogeneous variably containing cells expressing markers of Club and ATII together with stem/progenitor cells of the lung airway epithelia (Fig. 5). In addition, mesenchymal growths with intense expression of vimentin and desmin are also found.

OPA is invariably fatal once the disease is diagnosed, and affected sheep die as a result of compromised respiratory function caused by tumor enlargement or from secondary bacterial infections. The incubation period of the naturally occurring disease can be very long lasting several months to years.

Susceptibility/resistance to JSRV of some breeds has been suggested. A single genetic association study associated CCR5 and MX1 genes with resistance and susceptibility, respectively, in Spanish Latxa diary sheep.

OPA can be transmitted experimentally only with material that contains JSRV, such as lung secretions collected from affected animals or virions obtained by transiently transfecting cells with JSRV infectious molecular clones. The experimental OPA model is highly reproducible. JSRV infection can be induced in most, if not all, inoculated lambs aged 1–6 months at the time of inoculation and a high proportion of them develop clinical signs and OPA lesions. Neonatal lambs are most susceptible and, in contrast to naturally occurring OPA, the incubation period for experimentally induced OPA can be as short as a few weeks.

Although most experimentally inoculated lambs develop clinical signs and OPA lesions, under natural conditions the majority of JSRV-infected sheep do not develop OPA during their commercial lifespan. Aerogenous transmission is well recognized but, recently, the importance of milk and colostral transmission has been highlighted. JSRV in these secretions can reach the intestinal lymphoid tissues of lambs nursed by infected mothers. Interestingly, most infected animals harbor the virus as a disseminated infection of their lymphoid tissues and do not show detectable pulmonary lesions. Although lymphoreticular cells appear to serve as the principal reservoir of virus infection, viral antigens are detected only rarely in this compartment where sensitive PCR assays are necessary to detect viral RNA or proviral DNA.

The short incubation period in young lambs experimentally infected with JSRV may be explained by the combination of high virus infectious doses present in the inoculum and the higher abundance of the permissive target cells (type 2 pneumocytes and Clara cells) in lambs compared to adult animals. In natural infections, the mechanisms involved in converting the stable persistent lymphoid infection in peripheral tissues to a progressive pulmonary epithelial tumor are not clear. In addition, it is not known whether and how efficiently sheep infected by JSRV, but with no neoplastic lesions, are able to transmit the virus to uninfected sheep. In an affected flock, control of OPA is very difficult as no vaccines or effective diagnostic tests are available.

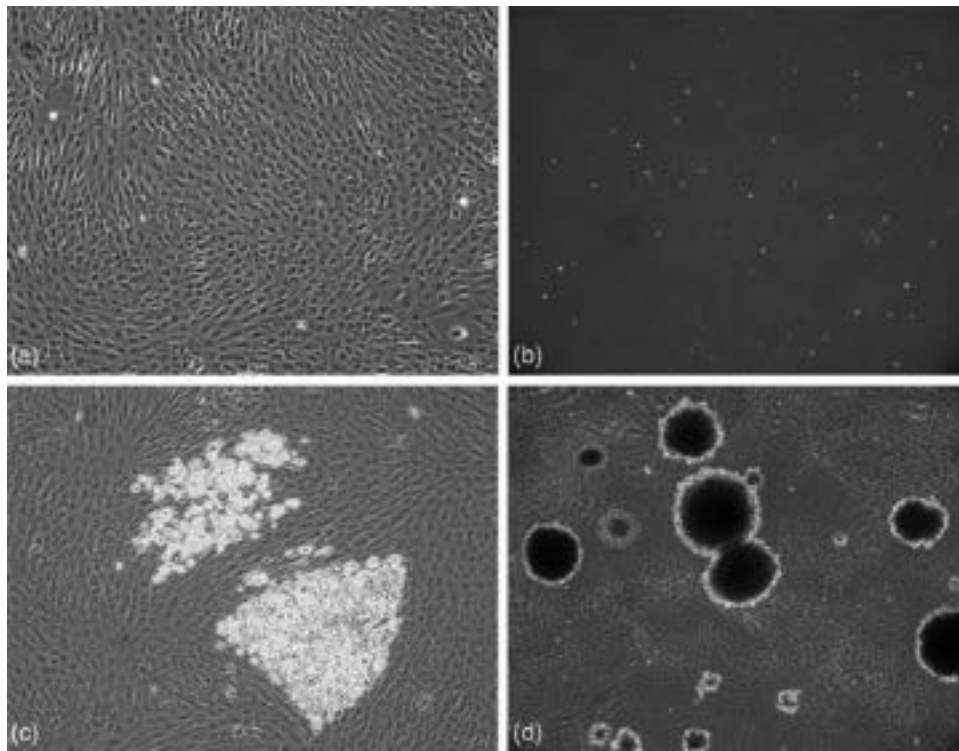


Fig. 6 Expression of the JSRV Env transforms cells *in vitro*. Transformation of the rat fibroblast cell line 208F by the JSRV envelope glycoprotein. 208F cells were mock-transfected (a) or transfected with an expression plasmid for the JSRV Env (c). Mock-transfected 208F cells are morphologically flat, possess a strong contact-inhibition and do not grow in soft agar (b). On the other hand, 208F cells transfected with an expression plasmid for the JSRV Env show the onset of foci of transformed cells (c), which are able to form colonies in soft agar (d).

Mechanisms of Virus-Induced Cell Transformation

The mechanisms used by JSRV to induce cell transformation are different from those followed by the majority of oncogenic retroviruses.

OPA tumors are multiclonal, and common proviral integration sites have been rarely observed. However, no cellular oncogenes have been found to be transduced by JSRV. The oncogenic properties of JSRV are due directly to one of its structural proteins.

Transfection of a variety of fibroblast and epithelial cell lines with expression plasmids for the JSRV Env leads to efficient cell transformation (Fig. 6). Moreover, experimental infection of mice and lambs, with replication incompetent vectors expressing JSRV Env, induced tumors in the inoculated animals. Thus, the JSRV Env is a dominant oncoprotein both *in vitro* and *in vivo* and viral spread is not necessary for tumorigenesis.

The mechanisms involved in JSRV Env-induced transformation have not been fully elucidated; however, signal transduction involving the PI3K-Akt and H/N Ras-MEK-MAPK pathways are important.

Additional pathways including mTOR/p70S6K, EGFR and Wnt signaling and AGR-2-YAPI-AREG axis are active in OPA and may contribute to oncogenesis in this disease.

Conflicting reports are available on the involvement of other cellular oncogenes such as the Stk/Ron tyrosine kinase in the onset of JSRV Env-induced cell transformation.

One of the major determinants of JSRV Env transformation is the transmembrane domain (TM), although other regions may be important. In particular, a putative docking site (Y-X-X-M) for phosphatidylinositol 3-kinase (PI-3K) is critical for JSRV-Env induced cell transformation. Within this motif, Y590 is crucial for JSRV Env-induced cell transformation, although Y590 mutants maintain a reduced ability to induce cell transformation in some cell lines. In summary, the JSRV Env acts as a dominant oncogene *in vitro* and *in vivo* and its expression is sufficient to induce lung adenocarcinoma in the target species.

Enzootic Nasal Tumor Virus

ENTVs of sheep and goats are distinct betaretroviruses highly related to each other and JSRV. ENTV causes a contagious tumor of the mucosal nasal glands, as well as respiratory and olfactory mucosa, in their respective target species known as enzootic nasal adenocarcinoma (ENA). Clinically, ENA is characterized by respiratory distress caused by the enlargement of the tumor, nasal discharge, and skull deformation. Like OPA, ENA has a long incubation period.

The molecular biology of ENTVs (ENTV-1, ENTV2) has not been studied in great detail, but the virus displays many features in common with JSRV. The cellular receptor for ENTVs is also HYAL2 and the ENTVs Env, like its JSRV counterpart, is a dominant oncogene that appears to follow the same mechanisms of cell transformation used by JSRV. The main differences between the ENTVs and JSRV appear to be in LTR, small regions of Gag, Orf-x and the transmembrane region (TM) of Env (concentrated in the U3 region of the viral LTRs). The ENTV-1 LTR, unlike the JSRV LTR, does not bind lung-specific transcription factors such as FOXA2, which is likely the basis of tropism differences between the two viruses.

Future Perspectives

Small ruminant betaretroviruses are a fascinating group of viruses with unique characteristics that are of broad interest through their veterinary, comparative medical and biological importance. The veterinary importance arises from the economic impact in many sheep rearing countries of the diseases induced by JSRV and ENTVs combined with the absence of any effective control tools or mechanisms. Their comparative medical interest stems from the striking similarity of some forms of human lung adenocarcinoma to OPA.

Investigations looking for the JSRV in human lung cancer have not found evidences for its presence in human pulmonary adenocarcinoma but data do not rule out a role for a unknown retrovirus. In addition, the comparison of the transcriptomes of OPA and human lung adenocarcinomas indicate close similarity, but highlights that there remain many differences between the sheep and human diseases.

OPA is considered an excellent outbred large animal model for these tumors with opportunities to investigate issues that are not available from other systems.

Lung cancer is the main cause of death among cancer patients and effective therapeutic strategies are greatly needed to improve patient survival and well-being. OPA is a large animal model that can identify and test the efficacy of new therapeutic interventions in a highly reproducible system.

enJSRVs are an especially active group of endogenous retroviruses and offer insights into several areas of general biological interest, such as viral replication, interference, and reproductive biology. Understanding the mechanisms of the enJS56A1-induced block could inspire the design of novel anti-retroviral strategies and shed light on early events in retroviral assembly and/or trafficking. In particular, this unique viral block provides additional clues on the variety of mechanisms shaping co-evolution of endogenous/exogenous retroviruses and their hosts.

enJSRVs are highly expressed in the genital tract of the ewe and are intimately involved in early placental development in this animal species. The sheep/enJSRVs model can be useful to experimentally address the hypothesis that endogenous retroviruses have shaped and are essential for mammalian biology.

Further Reading

- Armezzani, A., Varela, M., Spencer, T.E., Palmarini, M., Arnaud, F., 2014. Menage a Trois: The evolutionary interplay between JSRV, enJSRVs and domestic sheep. *Viruses* 6, 4926–4945.
- Borobia, M., De las Heras, M., Ramos, J.J., *et al.*, 2016. Jaagsiekte retrovirus can reach Peyer's patches and mesenteric lymph nodes of lambs nursed by infected mothers. *Veterinary Pathology* 53, 1172–1179.
- Fan, H. (Ed.), 2003. *Jaagsiekte Sheep Retrovirus and Lung Cancer*. Berlin: Springer.
- Karagianni, A.E., Vasoya, D., Finlayson, J., *et al.*, 2019. Transcriptional response of ovine lung to infection with Jaagsiekte sheep retrovirus. *Journal of Virology* 93 (21).
- Maeda, N., Palmarini, M., Murgia, C., Fan, H., 2001. Direct transformation of rodent fibroblasts by jaagsiekte sheep retrovirus DNA. *Proceedings of the National Academy of Sciences of the United States of America* 98, 4449–4454.
- Martineau, H.M., Cousens, C., Ilmach, S., Dagleish, M., Griffiths, D., 2011. Jaagsiekte sheep retrovirus infects multiple cell types in the ovine lung. *Journal of Virology* 85, 3341–3355.
- Palmarini, M., Gray, C.A., Carpenter, K., *et al.*, 2001. Expression of endogenous betaretroviruses in the ovine uterus: effects of neonatal age, estrous cycle, pregnancy, and progesterone. *Journal of Virology* 75 (23), 11319–11327.
- Palmarini, M., Sharp, J.M., De las Heras, M., Fan, H., 1999. Jaagsiekte sheep retrovirus is necessary and sufficient to induce a contagious lung cancer in sheep. *Journal of Virology* 73, 6964–6972.
- Wootton, S.K., Halbert, C.L., Miller, A.D., 2005. Sheep retrovirus structural protein induces lung tumours. *Nature* 434, 904–907.

Japanese Encephalitis Virus (*Flaviviridae*)

Sang-Im Yun and Young-Min Lee, Utah State University, Logan, UT, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A.D.T. Barrett, Japanese Encephalitis Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00434-9](https://doi.org/10.1016/B978-012374410-4.00434-9).

Classification

Japanese encephalitis virus (JEV) is a species belonging to the largest genus *Flavivirus* (53 species) of the family *Flaviviridae*, which includes three other genera, *Pestivirus* (11 species), *Hepacivirus* (14 species), and *Pegivirus* (11 species), according to a report published by the International Committee on Taxonomy of Viruses in 2018. Like many other species within the *Flavivirus* genus, JEV can infect both arthropods (mosquitoes), which serve as vectors, and vertebrates (including humans), which act as animal hosts. JEV is a clinically important mosquito-borne zoonotic flavivirus representing the JEV antigenic complex, a group of virus species that show a high degree of interspecies cross-neutralization and phylogenetic relatedness. This group also includes West Nile (WNV), Murray Valley encephalitis, Saint Louis encephalitis, Usutu, Koutango, Yaounde, and Cacipacore viruses. Moreover, JEV is genetically related to other medically significant but antigenically distinct mosquito-borne flaviviruses, such as dengue (DENV), yellow fever (YFV), and Zika (ZIKV) viruses, as well as tick-borne flaviviruses, including tick-borne encephalitis (TBEV), louping ill, Langat, Powassan, Omsk hemorrhagic fever, and Kyasanur forest disease viruses.

Virion Structure

JEV is an enveloped virus, 50–60 nm in diameter. It comprises three virus-encoded structural proteins: capsid (C), membrane (M) or its precursor prM, and envelope (E). Conventional thin-section electron microscopy (EM) has revealed that the intracellular virions in JEV-infected cells have an inner electron-dense core, ~30 nm in diameter, surrounded by an electron-lucent narrow interspace and an outer electron-dense thin layer (**Fig. 1(A)**). The inner core of the virus is the nucleocapsid, a poorly-defined structure consisting of the viral genome and multiple C proteins. This nucleocapsid is surrounded by a host cell-derived lipid bilayer membrane that is coated with an outer protein shell containing two transmembrane proteins, prM/M and E. Based on the protein composition and architecture of this shell, two different forms of the virion are recognized: (a) extracellular mature virions (fully infectious), with a smooth protein shell containing M and E proteins; and (b) intracellular immature virions (non-infectious), with a spiky protein shell containing prM and E proteins. Occasionally, partially mature virions are released from JEV-infected cells; these forms are still infectious but presumably less so than completely mature virions; their protein shell appears partly smooth and partly spiky, depending on the local protein composition of either M + E or prM + E. Also, subviral particles are secreted from JEV-infected cells; these are rather heterogenous in size but generally smaller than the viral particles and contain only the prM/M and E proteins on the viral membrane, without the inner nucleocapsid.

The Extracellular Mature Virion

The structure of the ~50 nm-sized mature JEV has recently been solved at 4.3 Å resolution by cryo-EM and three-dimensional image reconstruction (PDB ID 5WSN), showing a stoichiometry and arrangement of the M and E proteins on the viral membrane that agrees well with all the previously reconstructed cryo-EM structures of mature DENV (PDB IDs 1K4R and 3J27), WNV (PDB ID 3J0B), ZIKV (PDB IDs 5IRE, 6CO8, and 5IZ7), and TBEV (PDB ID 5O6A). The outer protein shell of the mature JEV is made up of 60 icosahedral asymmetric units (**Fig. 1(B)**), each containing three antiparallel M:E heterodimers, arranged into 30 diamond-shaped “tiles” by positioning two asymmetric units in opposite orientations. Each tile therefore contains three E:M:M:E heterotetramers, with the two adjacent M proteins being hidden under the antiparallel E:E homodimer. Each E monomer consists of an N-terminal banana-shaped ectodomain, a membrane-proximal three helix-containing stem, and a C-terminal double-spanning membrane anchor. The E ectodomain is divided into three β -barrel domains (DI–DIII), with a hydrophobic fusion loop located at the distal end of DII and fitted into a pocket at the DI–DIII interface of the neighboring E ectodomain. The M protein consists of an N-terminal flexible loop, an amphipathic α -helix, and a C-terminal double-spanning membrane anchor. Overall, the outer protein shell of the mature virion is stabilized by multiple molecular interactions: (a) electrostatic and hydrophobic interactions between the M-loop and E-DI and DII; (b) hydrogen-bond interactions between the M-helix, E-DII, and the loop of a neighboring M; and (c) charge interactions between E-DI, DIII, and the stem region.

The Intracellular Immature Virion

The structure of the ~60 nm-sized immature JEV has not yet been determined at the atomic level, but those of immature DENV (PDB IDs 3C6D and 4B03), WNV (PDB ID 2OF6), YFV (PDB ID 1NA4), and ZIKV (PDB ID 5U4W) have already been reported at

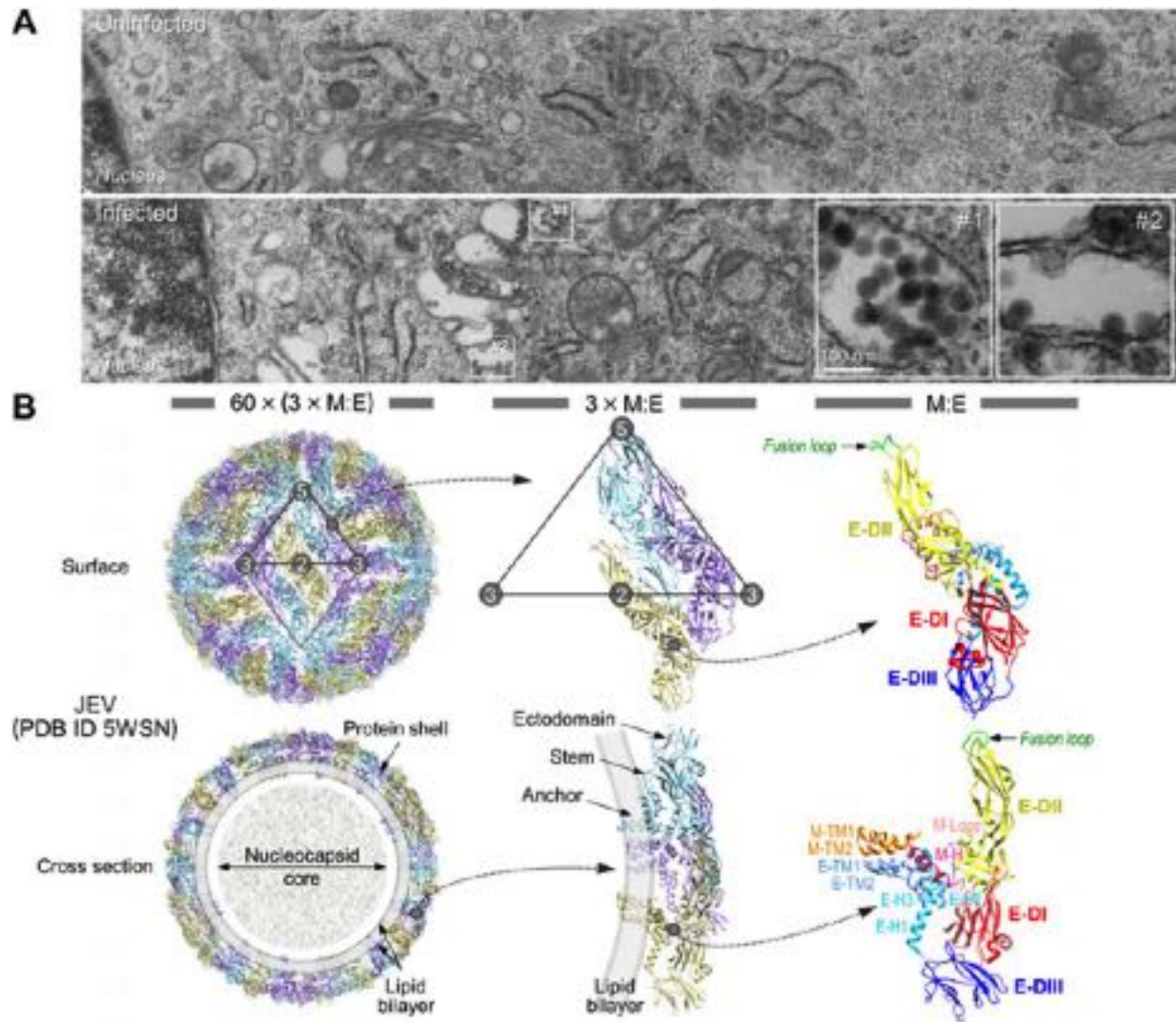


Fig. 1 Virion structure of JEV. (A) Thin-section EM images of BHK-21 cells uninfected or infected with JEV strain CNU/LP2 for 18 h. Insets #1 and #2 are the enlarged views of virus-induced vesicles containing the intracellular immature JEV particles. (B) Three-dimensional cryo-EM structure of the extracellular mature JEV particles (strain P3). The inner nucleocapsid core, a complex of the viral genomic RNA and multiple copies of the C protein, is enveloped by a lipid bilayer that is coated with an outer protein shell composed of 180 copies each of two membrane-anchored proteins, M and E. On the viral membrane, the 180 M:E heterodimers are organized into 60 icosahedral asymmetric units (open black triangle), each containing three antiparallel M:E heterodimers (medium purple, sky blue, and dark khaki), which then form 30 diamond-shaped tiles (open gray diamond) as two asymmetric units oriented in opposite directions. The 2-, 3-, and 5-fold axes of symmetry are marked with the corresponding numbers (2, 3, and 5). The E monomer comprises a three-domain-containing ectodomain (E-DI, red; E-DII, yellow; and E-DIII, blue), followed by a three-helix-containing stem (E-H1, E-H2, and E-H3, deep sky blue) and a two-transmembrane helix-containing anchor (E-TM1 and E-TM2, cornflower blue). The fusion loop (green) is located at the tip of E-DII. The M monomer comprises a flexible loop (M-Loop, pink), followed by an α -helix (M-H, hot pink) and a two-transmembrane helix-containing anchor (M-TM1 and M-TM2, orange). The pdb data file for the virion structure was retrieved from the Protein Data Bank (PDB ID 5WSN), and its virion structure was visualized using Chimera software.

high resolution. Given their close relatedness, immature JEV is likely to have an overall virus structure similar to that of these other immature flaviviruses. In all of the currently available cryo-EM structures of immature flaviviruses, the inner nucleocapsid is poorly defined, as is true of their respective mature virions; however, the outer protein shell is well defined, composed of 60 trimers of prM:E heterodimers, with each trimer appearing as a protruding spike on the surface of the viral membrane. Each spike has a tripod-like structure holding the three E ectodomains together, in each of which E-DII forms a tripod leg, and E-DI and DIII shape a tripod foot. In this structure, three hydrophobic fusion loops come together to form the tripod head, with each loop exposed at the apical tip of E-DII but capped by the N-terminal pr portion of prM, thereby preventing the virions from prematurely fusing with intracellular membranes during viral morphogenesis. The C-terminus of the pr fragment is connected to the N-terminal loop segment of M that extends downward toward the viral membrane along the lateral surface of E-DII.

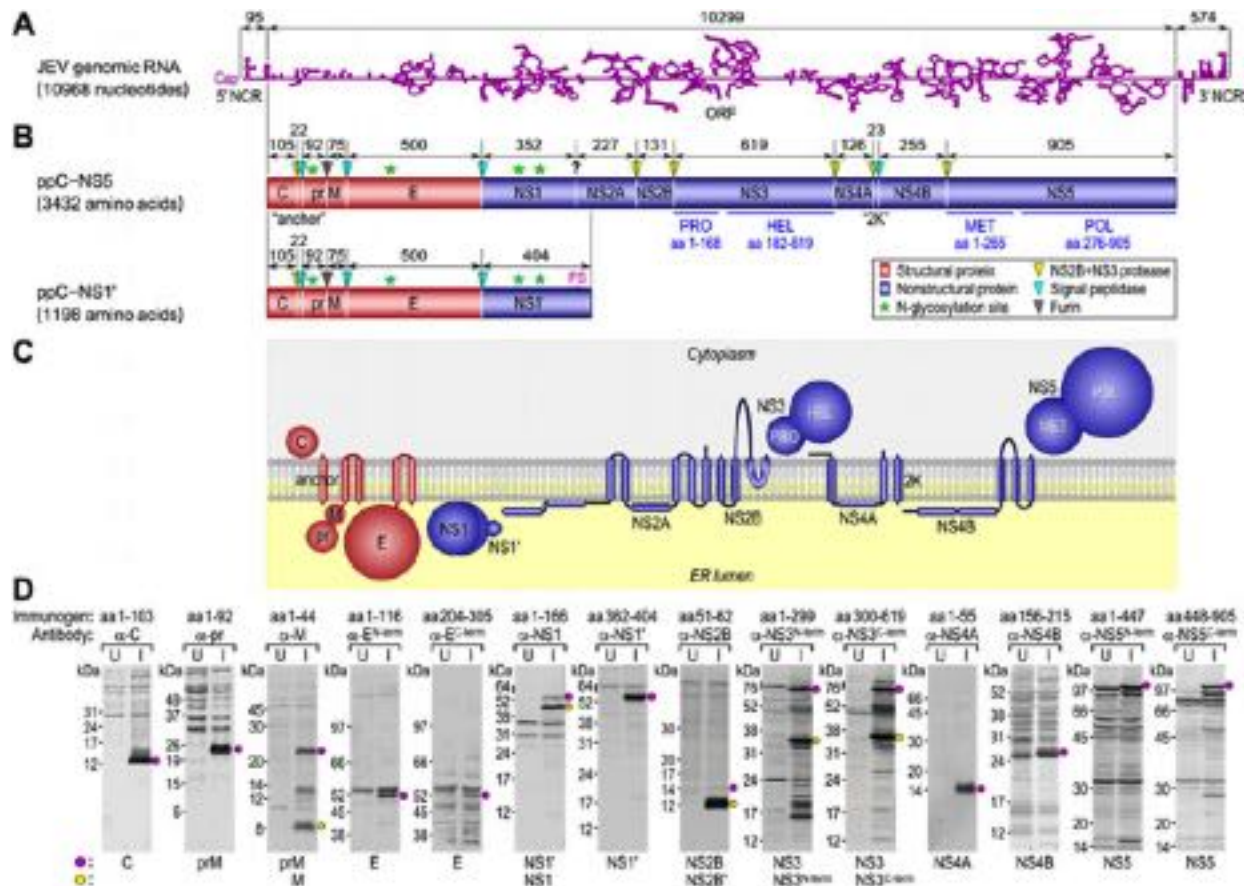


Fig. 2 Genome structure and gene expression of JEV. (A) Schematic representation of the single-stranded positive-sense RNA genome of JEV strain CNU/LP2. The genomic RNA has a cap structure at the 5' end, followed by a 5'NCR, an ORF, and a 3'NCR with no poly(A) tail at the 3' end. (B) Schematic representation of the two polyprotein precursors (ppC-NS5 and ppC-NS1') translated from the JEV single ORF. While the ppC-NS5 precursor is the full-length polyprotein, the ppC-NS1' precursor is a C-terminally truncated polyprotein produced by a -1 ribosomal frameshift (FS) occurring between codons 8 and 9 of NS2A. The two polyproteins are processed at the specific cleavage sites by cellular and viral proteases (yellow, cyan, and gray arrowheads) into three structural proteins (red) and at least seven nonstructural proteins (blue). Of these 10 mature proteins, two possess a bipartite structure: NS3, which has an N-terminal protease (PRO) domain and a C-terminal helicase (HEL) domain; and NS5, which has an N-terminal methyltransferase (MET) domain and a C-terminal RNA-dependent RNA polymerase (POL) domain. Green asterisks indicate four N-linked glycosylation sites located in the pr portion of prM (Asn¹⁵), E (Asn¹⁵⁴), and NS1/1' (Asn¹³⁰ and Asn²⁰⁷). (C) Predicted membrane-spanning regions and their orientation in all 10 major JEV proteins associated with the ER membrane. (D) Detection of viral proteins in BHK-21 cells infected with JEV strain CNU/LP2. At 18 h after infection, total lysates of uninfected (U) and JEV-infected (I) cells were examined by immunoblotting with each of 14 JEV region-specific rabbit polyclonal antibodies, as indicated. In each blot, purple and yellow dots indicate the major viral proteins reacting with the given antibody. Panels A–C are modified from a review article (Yun, S.I., Lee, Y.M., 2018. Early events in Japanese encephalitis virus infection: Viral entry. *Pathogens* 7, 68), and panel D is reproduced from a research article (Kim, J.K., Kim, J.M., Song, B.H., *et al.*, 2015. Profiling of viral proteins expressed from the genomic RNA of Japanese encephalitis virus using a panel of 15 region-specific polyclonal rabbit antisera: Implications for viral gene expression. *PLoS One* 10, e0124318), with permission.

Genome

Genome Structure

JEV contains a linear single-stranded positive-sense RNA genome, ~11,000 nucleotides in length (Fig. 2(A)). The 5' end of the genomic RNA has a type 1 cap and begins with the conserved dinucleotide AG (abbreviated m⁷GpppAmG, where the "m" indicates a methyl group added on the terminal nucleobase guanine at the N-7 position and the first nucleotide adenosine at the ribose 2'-O position). In contrast, the 3' end of the genomic RNA lacks a poly(A) tail but terminates instead with the conserved dinucleotide CU. In flaviviruses, a methylated 5'-cap on the genomic RNA plays multiple roles in preventing RNA degradation, promoting translation initiation through a canonical cap-dependent mechanism, and protecting viral RNAs from recognition by a cellular factor(s) that senses the presence of non-self RNAs; however, these viruses can also utilize a non-canonical translation initiation mechanism that depends minimally on the presence of a methylated 5'-cap. The viral genome consists of an ~95-nucleotide 5' noncoding region (NCR), an ~10,299-nucleotide open reading frame (ORF), and an ~574-nucleotide 3'NCR.

The 5'NCR and the following ~100-nucleotide 5'-terminal region of the ORF contain a cluster of *cis*-acting primary sequences and secondary/tertiary structures that are involved in regulating viral translation and RNA replication through a long-range intramolecular RNA–RNA interaction with the 3'NCR, which also comprises an array of *cis*-acting linear sequences and higher-order structures.

Gene Expression

The genomic RNA of JEV is infectious on its own, since it can be directly translated by the protein synthesis machinery of its host cells. Upon JEV infection, the single long ORF encoded in the viral genomic RNA is translated into a full-length polyprotein precursor (designated ppC-NS5) that is then cleaved *co*- and *post*-translationally to generate at least 10 functional proteins: C-“anchor”-prM-E-NS1-NS2A-NS2B-NS3-NS4A-“2K”-NS4B-NS5, where NS denotes “nonstructural” (Fig. 2(B) and (C)). Mainly based on previous research with YFV, DENV, and WNV, the flavivirus full-length ppC-NS5 is shown to be processed in a spatially and temporally controlled manner by four different proteases: (a) the host signal peptidase, which mediates cleavage at the junctions of anchor-prM, prM-E, E-NS1, and 2K-NS4B on the lumen side of the endoplasmic reticulum (ER) membrane; (b) the viral two-component serine protease NS2B + NS3, which recognizes the cleavage site at the junctions of C-anchor, NS2A-NS2B, NS2B-NS3, NS3-NS4A, NS4A-2K, and NS4B-NS5 on the cytoplasmic side of the ER membrane; (c) the host furin, which cleaves prM into the pr fragment and M protein in the *trans*-Golgi network (TGN); and (d) an as-yet unidentified host protease, which mediates the cleavage at the NS1-NS2A junction.

In the case of JEV, WNV, and presumably other members of the JEV antigenic complex, the viral single ORF is translated into not only the full-length ppC-NS5 but also a C-terminally truncated polyprotein precursor (designated ppC-NS1') by a programmed –1 ribosomal frameshift event occurring at codons 8–9 of NS2A (Fig. 2(B) and (C)). As with the full-length ppC-NS5, the truncated ppC-NS1' is also processed to yield C, prM/M, E, and a unique NS1', an extended form of NS1 that has 52 additional amino acids added to the C-terminus of the protein. This frameshifting takes place at a slippery heptanucleotide sequence from ³⁵⁵⁴Y_CCU_UUU (0 frame) to YCC_UUU_U (–1 frame), which is stimulated by a pseudoknot structure located six nucleotides downstream of the slippery sequence. The pseudoknot consists of two immediately adjacent, possibly coaxially stacked stem loops (SL1 and SL2), with a five-nucleotide loop region (³⁵⁸⁵YUGGC) of SL1 base-paired with its complementary sequence (³⁶²¹GCCAG) six nucleotides downstream of SL2.

Using a collection of 14 JEV region-specific rabbit polyclonal antibodies, we conducted a comprehensive immunoblot analysis that presented a genome-wide expression profile of the viral gene products accumulated in JEV-infected cells, enabling us to label all the mature viral proteins except NS2A, together with a significant number of related species that are presumably produced as a result of further viral protease-mediated processing, post-translational modification, and/or intracellular protein degradation (Fig. 2(D)). Described below are the 12 biologically active JEV proteins, with a short account of their structure and function in the viral life cycle.

Structural proteins

The structural proteins required for infectious virus production are C, prM/M, and E:

- (1) C, a highly basic ~12-kDa cytosolic protein, forms a compact homodimer, with each monomer folded into four α -helices (Fig. 3(A)). It binds to the viral genomic RNA for nucleocapsid formation through charge interactions, probably with both the unstructured N-terminal region (residues 1–25) and the last C-terminal α -helix that are rich in basic residues. For viral assembly, C associates with the cytoplasmic face of the ER membrane through hydrophobic interactions, possibly with the first and second α -helices. Interestingly, C is also found in the nucleolus of JEV-infected cells, and Gly⁴² and Pro⁴³ at the beginning of its second α -helix are critical for the nuclear and nucleolar localization that plays a role in promoting viral replication in a cell type-dependent manner (i.e., Vero vs. C6/36). Also, full-length C is further cleaved internally between Lys¹⁸ and Arg¹⁹ in the unstructured N-terminal region by the host cysteine protease cathepsin L, which is important for cell type-specific viral replication (e.g., N18 vs. Vero). Although the nature of their cell type dependency is poorly understood, both the nuclear/nucleolar localization and internal cleavage of the C protein affects JEV's virulence in mice.
- (2) prM, a glycosylated ~24-kDa membrane protein, is the precursor molecule of M; it has an N-glycan covalently linked to Asn¹⁵ in its pr portion, which has a β -barrel fold, as inferred from the crystal structure of the DENV pr fragment in complex with its E ectodomain (PDB ID 3C6E). In the intracellular compartments along the secretory pathway, prM acts as a chaperone to assist in the proper folding and oligomerization of the main surface glycoprotein E and the efficient secretion of viral particles. In the immature virion, the pr portion of prM shields the hydrophobic fusion loop of E to prevent premature fusion during transit through the secretory pathway as part of viral morphogenesis.
- (3) M, a nonglycosylated ~8-kDa membrane protein, is the proteolytic product of prM; it consists of an N-terminal 19-amino-acid flexible loop, a perimembrane α -helix, and a C-terminal double-spanning membrane anchor (Fig. 3(B)). In the mature virion, M is the minor component of a viral envelope complex E:M:M:E heterotetramer, in which the two adjacent M proteins are buried underneath the antiparallel E:E homodimer. A recent study has shown that a single Ile³⁶→Phe mutation introduced at the end of the M-helix has little effect on viral entry but fails to produce infectious particles in mammalian cells, although not in mosquito cells, and it results in full attenuation in mice. Despite this knowledge of its structural details, the functional role of M in viral replication, particularly at the early and late stages of JEV infection, remains unclear.

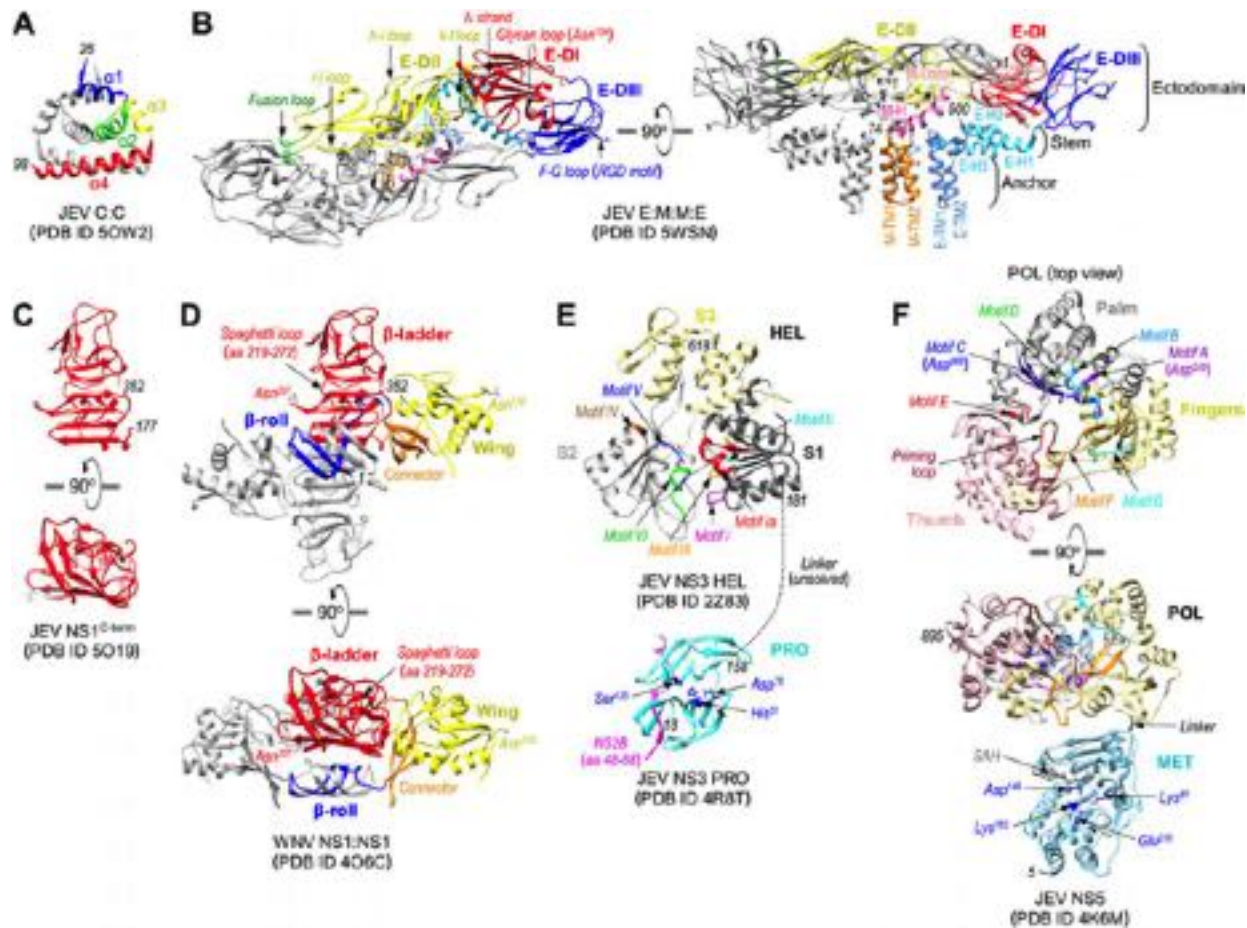


Fig. 3 Ribbon representation of JEV C, M, E, NS1, NS3, and NS5 proteins. Displayed are the atomic structures of a JEV C:C homodimer (A), a JEV E:M:E heterotetramer (B), a C-terminal fragment of JEV NS1 (C), a WNV NS1:NS1 homodimer as a reference for the as-yet experimentally unsolved JEV NS1:NS1 homodimer (D), the PRO and HEL domains of JEV NS3 (E), and a JEV NS5 monomer (F). The pdb data files for all the protein structures were retrieved from the Protein Data Bank (PDB) database, as indicated by their four-character identifiers, and their protein structures were visualized using Chimera software. If a dimer or tetramer is shown, only one monomer is used to color-code the structurally or functionally important domains/regions/motifs/residues. Black italic numbers indicate the amino acid positions of the N- and C-termini of each protein. A detailed description of the protein structures is provided in the main text.

- (4) E, a glycosylated ~53-kDa membrane protein, forms an antiparallel E:E homodimer lying horizontally on the surface of the mature virion, with each monomer composed of an N-terminal ectodomain, a membrane-proximal stem, and a C-terminal membrane anchor (**Fig. 3(B)**). The E ectodomain folds into three β -barrel domains (PDB IDs 3P54, 5MV1, and 5MV2): (a) DI, a centrally located globular domain, with an N-glycan covalently attached to Asn¹⁵⁴ in the E₀-F₀ or glycan loop and a stretch of closely distributed basic residues (Lys²⁷⁹–Lys²⁹⁷) on the I₀ strand and its flanking region; (b) DII, an elongated dimerization domain, with a glycine-rich hydrophobic peptide (Gly¹⁰⁰–Gly¹⁰⁹) in the c-d or fusion loop and several functionally important loops, such as h-i, i-j, and k-l loops; and (c) DIII, an immunoglobulin-like domain, with an RGD motif (Arg³⁸⁷–Gly³⁸⁸–Asp³⁸⁹) in the F-G loop. DIII is then connected to the stem region, which includes three α -helices lying nearly flat on the viral membrane, followed by the membrane anchor region, which includes two antiparallel α -helices traversing the viral membrane. E is the major component of the viral envelope complex E:M:M:E heterotetramer, which directs both receptor-mediated endocytosis and pH-dependent membrane fusion during virus entry into host cells. Therefore, E is the primary target for neutralizing antibodies during infection.

Nonstructural proteins

The nonstructural proteins involved in viral polyprotein processing, RNA replication, and assembly/release, as well as evasion of the host's innate immunity, are NS1/1', NS2A, NS2B, NS3, NS4A, NS4B, and NS5:

- (1) NS1, a glycosylated ~45-kDa multi-domain protein with two N-glycans each conjugated at Asn¹³⁰ and Asn²⁰⁷, exists not only as a membrane-associated cruciform homodimer in the secretory compartments and on the surface of JEV-infected cells but also as a secreted barrel-shaped homohexamer composed of three dimers with lipid molecules loaded inside. The

crystallized C-terminal half of JEV NS1 forms a central “ β -ladder” domain, with an N-glycan at Asn²⁰⁷ and a long “spaghetti loop” on one side (Fig. 3(C)), which is preceded by its as-yet uncrystallized N-terminal half that is likely folded into a small “ β -roll” dimerization domain on the other side of the central β -ladder and a bulging “wing” domain with an N-glycan at Asn¹³⁰ projecting from it, as inferred from the reported crystal structures of full-length NS1 of WNV (Fig. 3(D)), DENV (PDB ID 4O6B), and ZIKV (PDB IDs 5K6K and 5GS6). In these flaviviruses, the β -roll domain, together with a “connector” subdomain of the wing domain, creates a hydrophobic surface on one side of the dimer for membrane association and faces towards the interior of the hexamer for lipid association. Although the molecular mechanisms are not well understood, the cell-associated dimeric form of NS1 is known to be involved in viral RNA replication and particle production, whereas the secreted hexamer is associated with multiple components of the innate and adaptive immune systems, thereby modulating host immune responses. The secreted hexamer was originally described as a soluble complement-fixing agent present in the blood and tissues of infected hosts. Thus, NS1 is a major immunogen, capable of inducing protective immunity during flavivirus infection.

- (2) NS1', a glycosylated ~58-kDa multi-domain protein, is the frameshift product of NS1 that acquires a 52-amino-acid C-terminal extension corresponding to the N-terminal nine amino acids of NS2A before the frameshift site and the unique 43 amino acids in the shifted -1 reading frame afterwards. Although the structure of its C-terminal extension is undefined, NS1' is dispensable for viral replication *in vitro* but plays a role in enhancing viral virulence *in vivo*. Further investigation is required to elucidate the differential roles of NS1' (apart from the multiple known functions of NS1) in JEV biology, particularly related to viral pathogenesis and host antiviral responses.
- (3) NS2A, a hydrophobic ~25-kDa transmembrane protein, is predicted to have eight α -helices; five span the ER membrane, and the other three are exposed to the ER lumen. NS2A is a component of the viral replication complex that can inhibit the host's interferon (IFN)-mediated antiviral response in cells infected with JEV, WNV, or DENV. In the case of DENV, NS2A inhibits the RIG-I–MAVS signaling pathway by preventing TBK1/IRF3 phosphorylation. In JEV, NS2A is degraded by TRIM52, a tripartite motif-containing cellular protein with antiviral activity. In YFV, full-length NS2A (224 amino acids) is cleaved internally by the viral two-component serine protease NS2B + NS3 to generate its C-terminally truncated variant (190 amino acids). A subsequent mutagenetic study has shown that loss of this internal cleavage site in YFV blocks the production of infectious particles without altering the level of viral RNA replication; also, this defect can be restored by introducing a suppressor mutation in NS3, suggesting a role for NS2A and NS3 in viral morphogenesis. Similarly, mutations introduced into WNV NS2A block viral particle production. Further studies are warranted to elucidate the precise membrane topology and function of NS2A.
- (4) NS2B, a hydrophobic ~14-kDa transmembrane protein, has three potential membrane-spanning α -helices, with a central ~45-amino-acid hydrophilic loop (residues 51–95) located on the cytoplasmic side of the ER membrane that forms a stable complex with NS3. This hydrophilic loop serves to anchor the NS2B + NS3 complex to intracellular membranes and acts as a cofactor for the protease activity of NS3. Also, recent genetic and biochemical studies have shown that the transmembrane domains of JEV NS2B play a role in both viral RNA replication and particle formation, likely through interactions with the viral protein NS2A and the host factor SPCS1, a component of the signal peptidase complex. Moreover, it is intriguing to note that the ~14-kDa NS2B protein is barely detectable in JEV-infected cells; instead, its smaller variant, ~12-kDa NS2B', is predominantly found. The nature and functional significance of NS2B' in JEV biology are unknown.
- (5) NS3, a cytoplasmic ~69-kDa bi-domain protein, contains an N-terminal serine protease (PRO) domain, a flexible interdomain linker, and a C-terminal RNA helicase (HEL) domain (Fig. 3(E)). The PRO domain has a catalytic triad of His⁵¹-Asp⁷⁵-Ser¹³⁵ positioned along a groove on its surface and requires a hydrophilic cofactor region of NS2B for its protease activity, with a specific cleavage recognition sequence of two basic residues at positions P2 and P1 and a small unbranched amino acid at position P1'. As a member of the DEAH/D-box helicase family within superfamily 2, the HEL domain folds into three subdomains (S1–S3), with the S1-S2 interface containing seven sequence motifs: I (also known as Walker A; ¹⁹⁷G–²⁰¹T), Ia (²²²L–²³²M), II (also known as Walker B; ²⁸⁵D–²⁸⁸H), III (³¹⁶T–³²¹G), IV (³⁶²F–³⁶⁷K), V (⁴⁰⁹T–⁴¹⁷N), and VI (⁴⁵⁷Q–⁴⁶⁴R). These motifs are associated with three catalytically related activities: (a) RNA helicase, which unwinds RNA duplexes and remodels RNA-protein complexes in an ATP-dependent manner; (b) nucleoside triphosphatase (NTPase), which provides energy from NTP hydrolysis; and (c) RNA 5'-triphosphatase (RTPase), which removes the γ -phosphate of a nascent 5'-triphosphorylated RNA to generate its 5'-diphosphorylated form as an initial step for RNA capping. Interestingly, the full-length NS3 of JEV is further cleaved into smaller products, including an ~34-kDa NS3^{N-term} and an ~35-kDa NS3^{C-term}, but the functional significance of these cleavage products is unknown. In addition to its roles in viral RNA synthesis, the HEL domain is also involved in viral assembly/packaging of YFV and WNV. In JEV, NS3 is associated with cellular microtubules and the TSG101 protein, which are potentially involved in viral assembly. Moreover, NS3 can induce cell death via caspase-dependent and/or -independent pathways in the case of JEV, WNV, DENV, and Langkat virus. In DENV, the NS2B + NS3 protease inhibits the activation of type I IFN signaling by cleaving the human adapter protein MPYS (also known as STING and MITA), but not its murine version.
- (6) NS4A, a hydrophobic ~14-kDa transmembrane protein, consists of three potential α -helices, with the first and third spanning the ER membrane and the second facing the ER lumen. NS4A, together with NS4B, is believed to serve as a major scaffold capable of inducing membrane rearrangements during the formation of viral replication complexes. Recent studies have shown that a single Lys⁷⁹→Arg substitution in NS4A causes a defect in JEV RNA replication that can be rescued by introducing a compensatory Tyr³→Asn mutation in NS4B, suggesting that a genetic interaction between NS4A and NS4B is

crucial for JEV RNA replication. In YFV, NS4A has been found to interact with NS1, which is essential for viral RNA replication. JEV NS4A is also reported to interact with cyclophilin B during RNA replication, which can therefore be suppressed *in vitro* by cyclosporin A treatment. JEV NS4A is further thought to participate in the inhibition of IFN signaling, as seen in WNV and DENV, and its interaction with the host DEAD-box RNA helicase DDX42 has been implicated in viral pathogenesis.

- (7) NS4B, a hydrophobic ~27/25-kDa transmembrane protein, is predicted to have five α -helices; the first two are located in the ER lumen, and the other three traverse the ER membrane. NS4B appears as a doublet on gel electrophoresis under denaturing and reducing conditions, indicating that a post-translational modification occurs in the protein, although its nature and significance for JEV biology are unknown. In DENV, NS4B co-localizes with NS3 and double-stranded RNA in the perinuclear region, where viral RNA replication presumably takes place, and it specifically binds to NS3, promoting NS3 helicase activity. In other mosquito-borne flaviviruses (WNV, DENV, and YFV), NS4B is known to play an important role in innate immune evasion, inhibiting the JAK–STAT signaling pathway by reducing STAT1 phosphorylation and/or inhibiting the RIG-I–MAVS signaling pathway by blocking TBK1/IRF3 phosphorylation, thereby suppressing IFN gene expression and the activation of IFN-stimulated response element promoters. Further research is needed to clarify the structure and function of NS4B in JEV biology.
- (8) NS5, a cytoplasmic ~103-kDa bi-domain protein, contains an N-terminal S-adenosyl-L-methionine (SAM)-dependent methyltransferase (MET) domain, a flexible interdomain linker, and a C-terminal RNA-dependent RNA polymerase (POL) domain (Fig. 3(F)). The MET domain, crystallized with S-adenosyl-L-homocysteine (SAH, the demethylated form of SAM), has a catalytic tetrad of Lys⁶¹-Asp¹⁴⁶-Lys¹⁸²-Glu²¹⁸, with their side chains pointing toward the center of the domain. This domain is associated with two enzymatic activities: (a) guanylyltransferase (GTase), which transfers a guanosine monophosphate (GMP) moiety from GTP in a backward orientation to the 5' end of the 5'-diphosphorylated RNA generated by the NS3 RTPase, thereby creating a unique 5'–5' triphosphaste linkage (GpppA-RNA); and (b) methyltransferase (MTase), which transfers a methyl group from SAM sequentially to the terminal nucleobase guanine at the N-7 position and to the first nucleotide adenosine at the ribose 2'-O position, thereby creating a type 1 cap (m⁷GpppAm-RNA). Mutagenetic studies have shown that N-7 methylation is essential for viral translation and RNA replication, whereas 2'-O methylation is involved in evasion of the host's innate immunity. The POL domain adopts a classical "right-hand" fold comprising fingers, palm, and thumb subdomains with seven sequence motifs, i.e., A (⁵³³Y–⁵⁴⁵T), B (⁶⁰¹Q–⁶¹⁵F), C (⁶⁶²R–⁶⁷³K), D (⁶⁹⁰S–⁶⁹⁴K), E (⁷¹²P–⁷¹⁷H), F (⁴⁵³I–⁴⁷⁹M), and G (⁴⁰⁴K–⁴¹³A), as well as the "priming loop" (⁷⁹⁰V–⁸¹²E) that initiates viral RNA synthesis *de novo* with two catalytic Asp residues (Asp⁵³⁶ in motif A and Asp⁶⁶⁸ in motif C) located on the palm subdomain. As noted in JEV and other flaviviruses, NS5 can shuttle between the nucleus and the cytoplasm, suggesting that it has additional functions other than viral RNA synthesis. In DENV, this nucleocytoplasmic shuttling is mediated by the importin and exportin transport system through a nuclear localization signal (residues 322–370) and a nuclear export signal (residues 329–345), respectively; however, these two signals overlap each other at the interface between the MET and POL domains, where a potential binding site for NS3 is also located. Thus, the role of the nuclear localization of NS5 remains unclear. NS5 is also a potent antagonist of IFN-stimulated JAK–STAT signaling, capable of (a) suppressing STAT1/TYK2 phosphorylation (in JEV, WNV, and TBEV), (b) binding to STAT2 and promoting its degradation (in DENV, YFV, and ZIKV), or (c) interacting with type I and II IFN receptor complexes and blocking JAK1/TYK2 phosphorylation (in Langat virus). JEV NS5 can also suppress the induction of type I IFN by inhibiting the nuclear translocation of IRF3 and NF- κ B by interrupting their interaction with two nuclear transport proteins, importins α 3 and α 4.

Life Cycle

The life cycle of JEV and other flaviviruses can be divided into three major phases (Fig. 4): entry (pre-RNA replication), translation and RNA replication, and morphogenesis (post-RNA replication).

Viral Entry

JEV infects a wide range of cell types in mosquitoes, humans, and various animal species, demonstrating its broad host cell tropism. The first step in the infection process is the attachment of a virion to the surface of a susceptible cell (Fig. 4(A)). This step involves multiple molecular interactions, each characterized by the binding of a part of the viral surface glycoprotein E to a certain host factor. Thus far, three such interactions have been documented: (a) A string of closely spaced positively charged residues in E-DI binds to one of the negatively charged glycosaminoglycans, such as a heparan sulfate proteoglycan. (b) An N-linked glycan in E-DI binds to one of the membrane-bound mammalian C-type lectins (e.g., DC-SIGN/L-SIGN, MR, CLEC5A, or LSECtin) or one of the secreted mosquito C-type lectins (e.g., mosGCIL-7) that is subsequently captured by its cell-surface receptor. (c) A tripeptide RGD motif in E-DIII binds to one of the RGD-binding integrins, such as $\alpha_v\beta_3$. Since these interactions are rather nonspecific, they are likely to play a role in recruiting the virions on the cell surface to promote a specific interaction of the viral E glycoprotein with its as-yet unidentified bona fide receptor(s) that directs the virions to a receptor-mediated clathrin-dependent/independent endocytic pathway for internalization. The cellular proteins implicated in JEV entry to date include HSP70, HSP90, the 37/67-kDa laminin receptor, CD4, CD14, vimentin, and the low-density lipoprotein receptor; however, their precise roles remain to be determined.

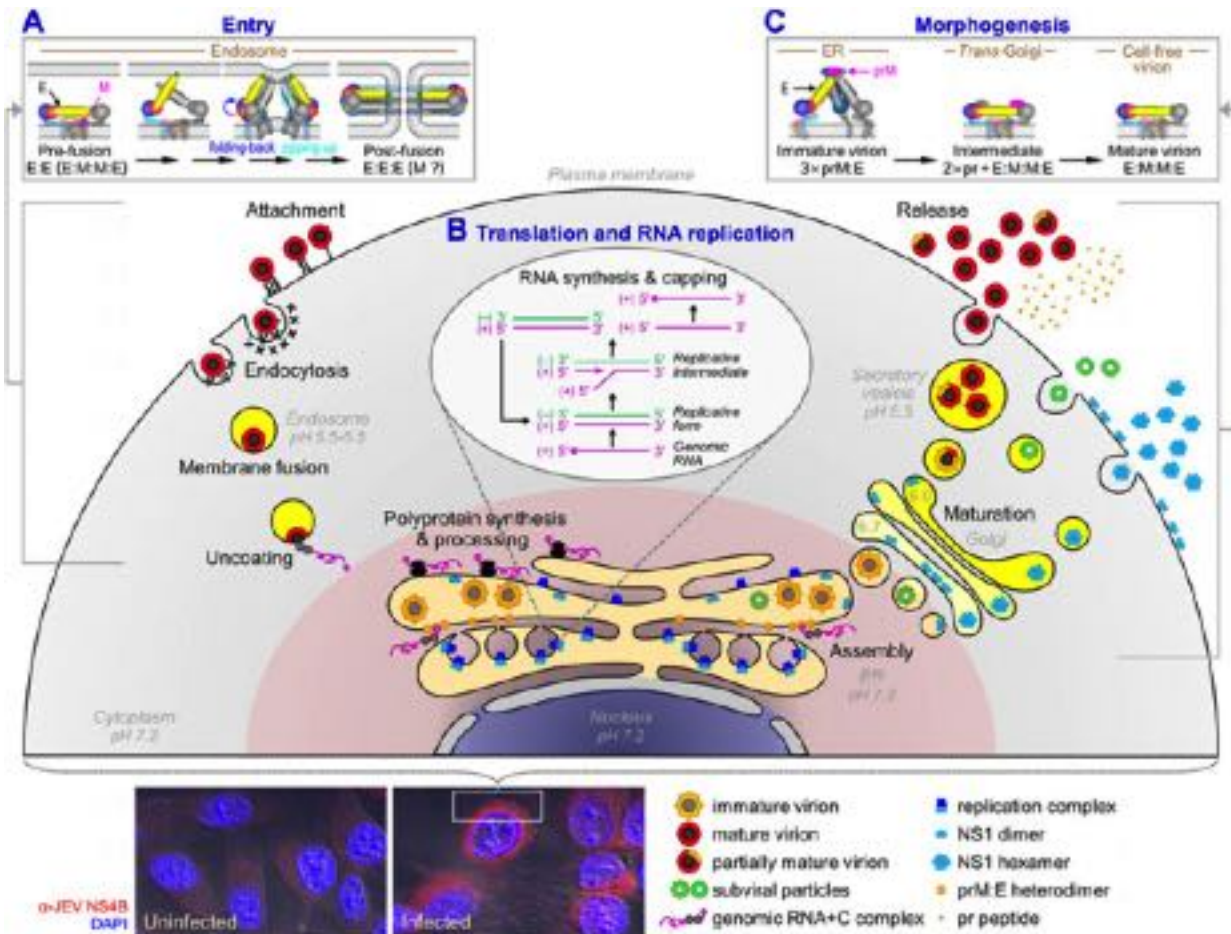


Fig. 4 Life cycle of JEV. Shown at the bottom are confocal microscopic images of BHK-21 cells uninfected or infected with JEV strain CNU/LP2 for 20 h, then immunostained with a primary rabbit anti-JEV NS4B antibody and a secondary Cy5-conjugated goat anti-rabbit IgG. Nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI). Illustrated above the immunofluorescent images are three phases of the JEV life cycle: (A) Entry phase, which includes the steps of viral attachment, endocytosis, membrane fusion, and uncoating at the early stage of viral replication. A rectangle-shaped inset depicts the low pH-induced structural reorganization of the viral envelope complex E:M:M:E heterotetramers occurring in an endosomal compartment. (B) Translation and RNA replication phase, which includes viral polyprotein synthesis and processing, as well as RNA synthesis and capping occurring in virus-induced specialized membranous compartments around the nucleus. An oval-shaped inset outlines the replication of the viral genomic RNA occurring in a semiconservative and asymmetric manner. (C) Morphogenesis phase, which includes the steps of viral assembly, maturation, and release at the late stage of viral replication. A rectangle-shaped inset depicts the low pH-triggered structural rearrangement of the icosahedrally arranged spikes ($3 \times \text{prM:E}$) on the surface of immature virions occurring during transit through the secretory pathway from the ER to the plasma membrane. The schematic diagram is modified from a review article (Yun, S.I., Lee, Y.M., 2018. Early events in Japanese encephalitis virus infection: Viral entry. *Pathogens* 7, 68).

Extensive structural and biochemical studies of other flaviviruses (DENV, WNV, and TBEV) have demonstrated that once the internalized virions reach an endosomal compartment, an acidic pH induces irreversible structural reorganization of the viral envelope complex E:M:M:E heterotetramers (Fig. 4(A), rectangle-shaped inset). This reorganization initially involves dissociating the metastable antiparallel E:E homodimers (the effect on the two adjacent M proteins is unknown), exposing their fusion loops to insert them into the inner leaflet of the endosomal membrane, and then reassociating the E monomers into stable parallel E:E:E homotrimers. This E trimerization requires the “folding back” of DIII against DII and the “zipping up” of the stem extended from DIII along the side of DII and toward the fusion loop embedded halfway into the endosomal membrane, thus driving the membrane fusion. This fusion process precedes a poorly understood uncoating step that may be mediated by one or more cytoskeletal structures (e.g., microtubules), resulting in the release of the viral nucleocapsid/genome into the cytoplasm.

Viral Translation and RNA Replication

After being released into the cytoplasm, the viral genomic RNA encoding a single long ORF is translated on ER-bound ribosomes into two polyprotein precursors, ppC-NS5 and ppC-NS1' (Fig. 4(B)). These precursors correspond to the full-length product and its C-terminally truncated form resulting from a -1 frameshift event occurring during the translational elongation step at codons 8

and 9 of NS2A, respectively (Fig. 2(B)). During and after translation, the two precursors are proteolytically processed by a combination of the viral serine protease (NS2B + NS3) and three host cell proteases (signal peptidase, furin, and an unknown protease) to produce three structural (C, prM/M, and E) and at least seven nonstructural (NS1/1', NS2A, NS2B, NS3, NS4A, NS4B, and NS5) proteins.

Based on recent electron tomography studies of several flaviviruses (DENV, WNV, ZIKV, and TBEV), a subset of the viral nonstructural proteins, particularly NS4A and NS4B, have been shown to be able to induce the proliferation and invagination of the ER to create clusters of ~90-nm-wide pouch-like vesicles (known as “vesicle packets”). Each vesicle often appears to have a pore connecting its interior with the cytosol (Fig. 4(B)). These vesicles can be seen in close proximity to highly folded membranes known as “convoluted membranes” or “paracrystalline arrays.” Inside these vesicles, most, if not all, of the viral nonstructural proteins, together with as-yet poorly defined host factors, are assembled on the membrane into a replication complex.

Within the complex, viral genomic RNA is replicated in a semiconservative and asymmetric manner by dynamic coordination of the two most conserved multienzymatic nonstructural proteins, NS3 and NS5 (Fig. 4(B), oval-shaped inset). Initially, the linear positive-sense genomic RNA is converted to a circular form through an intramolecular long-distance RNA–RNA interaction between the 5'- and 3'-terminal regions. The cyclized RNA serves as a template for the synthesis of its complementary negative-sense RNA, with the NS5 polymerase recognizing its promoter element (known as the stem-loop A) at the 5' end, and in turn, initiating negative-sense RNA synthesis at the 3' end. The nascent negative-sense RNA remains base-paired with the positive-sense RNA template, forming a double-stranded RNA molecule (called a “replicative form”). Subsequently, the negative-sense RNA serves as a template for the synthesis of multiple copies of new positive-sense RNAs (~10-fold more than negative-sense RNAs), each continuously displacing a preceding positive-sense RNA and thereby forming a partially single-stranded and partially double-stranded RNA molecule (called a “replicative intermediate”). The synthesis of both negative- and positive-sense RNAs is mediated by NS5 polymerase activity, coupled with the NS3 helicase and NTPase activities that are apparently required during the synthesis of positive-sense RNA.

During or shortly thereafter, the nascent positive-sense RNA undergoes RNA capping, which involves removal of the γ -phosphate from the 5' end to generate its 5'-diphosphorylated form through NS3 RTPase activity and subsequent transfer of a GMP group from GTP to the 5' end of the 5'-diphosphorylated RNA in a 5' – 5' triphosphate linkage via NS5 GTase activity. The 5'-capped RNA then goes through cap methylation, involving the initial transfer of a methyl group from SAM to the terminal nucleobase guanine at the N-7 position and then to the first nucleotide adenosine at the ribose 2'-O position by the NS5 MTase, thereby creating a type 1 cap at the 5'-end of the viral genomic RNA.

In addition to the three major RNA species (the single-stranded positive-sense genome, the double-stranded replicative form, and the partially single-stranded and partially double-stranded replicative intermediate), a set of subgenomic flavivirus RNAs (sRNAs) of ~0.2–0.5 kb is produced in cells infected with JEV or another flavivirus. These sRNAs correspond to the 3'-terminal region of the viral genomic RNA, produced by the cellular 5' → 3' exoribonuclease XRN1, which stops at a higher-order RNA structure within the 3'NCR. Primarily on the basis of previous work with other flaviviruses (e.g., WNV), it has been suggested that the accumulation of sRNAs in infected cells alters the homeostasis of cellular mRNA concentrations to promote viral replication and pathogenesis.

Viral Morphogenesis

Following viral translation and RNA replication, the assembly process is believed to begin in the vicinity of vesicle packets through the binding of viral C proteins to the newly made genomic RNA. This binding is followed by budding of the C protein-genomic RNA complex into the lumen of the ER, where the viral surface glycoproteins prM and E are incorporated (Fig. 4(C)). The budding process appears to be driven by oligomerization of the prM and E glycoproteins, which can lead to budding of subviral particles in the absence of the C protein-genomic RNA complex (nucleocapsid). Each of the nascent immature virions accumulated at the neutral pH of the ER lumen has an outer protein shell containing 60 icosahedrally arranged spikes; each spike comprises three prM:E heterodimers, with the three E proteins protruding from the viral membrane (Fig. 4(C), rectangle-shaped inset). As a notable feature of immature particles, the fusion loop at the apical tip of E-DII is protected from prematurely fusing with the ER membrane because it is covered with the globular pr portion of prM, rendering the particle noninfectious. These immature virions then undergo the maturation process during their transit through the secretory pathway to the plasma membrane, from which the mature virions are released.

Comprehensive structural and biochemical studies of other flaviviruses (DENV, WNV, and TBEV) have established that in the TGN, an acidic pH triggers an extensive structural rearrangement in the outer protein shell (Fig. 4(C), rectangle-shaped inset), converting the 60 protruding spikes into 60 asymmetric units that are organized into 30 diamond-shaped tiles each composed of three E:prM:prM:E heterotetramers (forming a total of 90 heterotetramers per virion). This structural change exposes the hidden cleavage site on prM for furin, a TGN-resident protease, and the furin-mediated cleavage of prM yields a soluble pr fragment and a membrane-anchored M protein. The resulting two pr fragments remain associated with the E:M:M:E heterotetramer at the acidic pH of the TGN, but they become dissociated from the mature virions when released into the extracellular space, which has a neutral pH. Both the fully and partially mature virions, together with the subviral particles and the NS1 hexamers, are exocytosed from virus-infected cells. A significant number of host factors are likely to be involved in viral morphogenesis, but at present they remain largely elusive.

Epidemiology

Geographic Distribution and Epidemiologic Patterns

JEV is the causative agent of Japanese encephalitis (JE), an acute encephalitis syndrome that is prevalent in the Asia-Pacific region, including South, East, and Southeast Asia, as well as the northern part of Oceania, affecting more than half of the world's population (Fig. 5). Suspected JE outbreaks were recognized in Japan as far back as the early 1870s. In 1924, the disease was initially named Japanese B encephalitis to distinguish it from encephalitis lethargica, known as Japanese A encephalitis. In 1934, a filterable agent from the brain of a deceased patient was used to demonstrate experimentally that it causes the encephalitis in non-human primates, and the virus (strain Nakayama) was then isolated in 1935.

For over half a century, JEV was mainly confined within originally defined boundaries in Asia. In the early 1990s, however, its geographic range expanded southward into the Torres Strait islands and the Cape York Peninsula in the northern Australia, westward into Pakistan, and eastward into Saipan in the western Pacific Ocean. Surprisingly, JEV RNA/proteins were detected in Italy from tissue samples of both healthy and dead birds in 1997–2000, as well as from a pool of *Culex pipiens* mosquitoes in 2010. More startlingly, a resident in Angola was found to be locally coinfecting with YFV and JEV in 2016. These pieces of recent evidence therefore indicate that JEV is circulating not only widely in the Asia-Pacific region but also possibly locally in Europe and Africa, with the potential for further spread.

JEV shows two epidemiological patterns, largely depending on regional climate (Fig. 5): (a) an endemic pattern, observed in the tropics, where viral transmission takes place throughout the year but increases during the rainy season; and (b) an epidemic pattern in the subtropical and temperate zones, in which viral transmission tends to occur seasonally during the summer and early fall. Although only 0.1%–4.0% of people infected with JEV are symptomatic, 50,000–175,000 clinical JE cases are estimated to

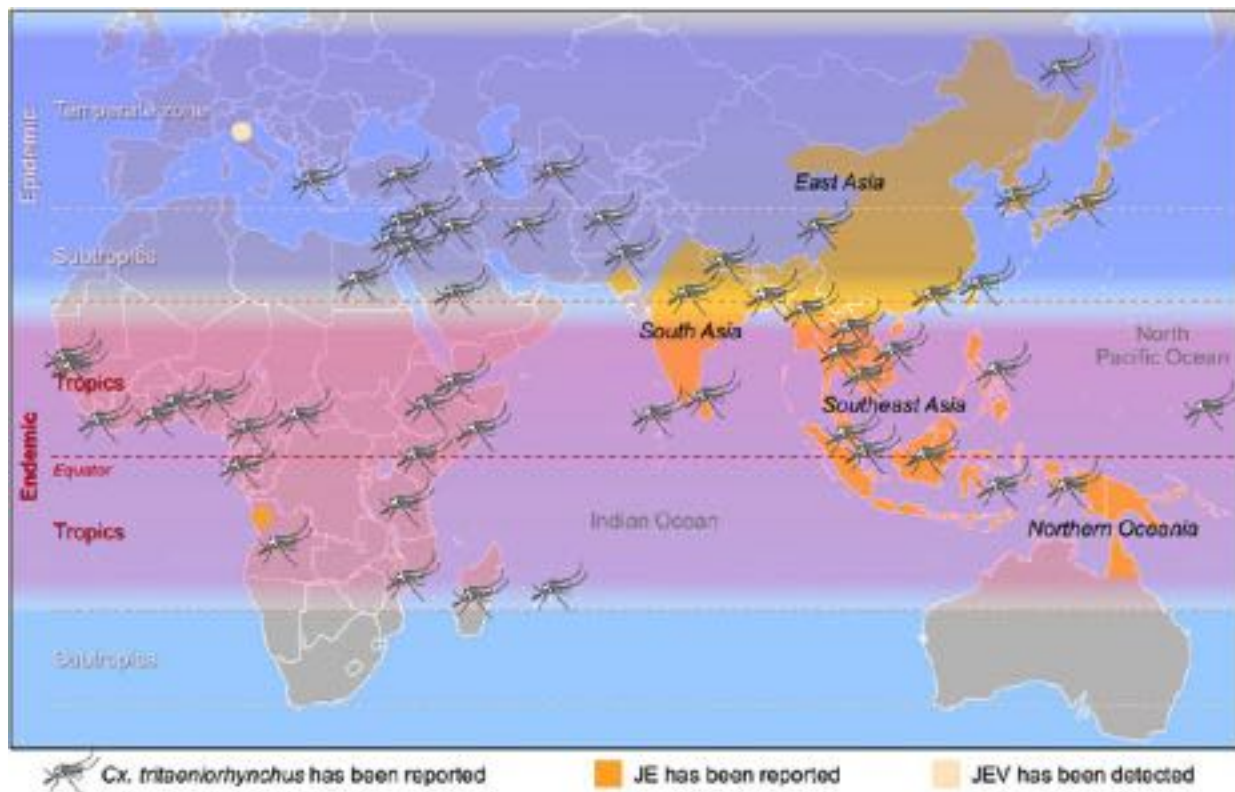


Fig. 5 Geographic distribution and epidemiologic patterns of JEV. Marked on the map are the countries or areas (mosquito) where *Cx. tritaeniorhynchus* has been reported, the countries or areas (orange) where JE has been reported in humans, and the areas (khaki) where JEV has been detected in birds and mosquitoes with no clinical JE case reported. Also indicated across the map are two major epidemiological patterns of JE: endemic (light red belt) and epidemic (light blue belt). The map is created based on the data from: (1) Jeffries, C.L., Walker, T., 2015. The potential use of *Wolbachia*-based mosquito biocontrol strategies for Japanese encephalitis. *PLoS Neglected Tropical Diseases* 9, e0003576. (2) Platonov, A., Rossi, G., Karan, L., *et al.*, 2012. Does the Japanese encephalitis virus (JEV) represent a threat for human health in Europe? Detection of JEV RNA sequences in birds collected in Italy. *Eurosurveillance* 17, 20241. (3) Preziuso, S., Mari, S., Mariotti, F., Rossi, G., 2018. Detection of Japanese encephalitis virus in bone marrow of healthy young wild birds collected in 1997–2000 in Central Italy. *Zoonoses and Public Health* 65, 798–804. (4) Ravanini, P., Huhtamo, E., Ilaria, V., *et al.*, 2012. Japanese encephalitis virus RNA detected in *Culex pipiens* mosquitoes in Italy. *Eurosurveillance* 17, 20221. (5) Simon-Loriere, E., Faye, O., Prot, M., *et al.*, 2017. Autochthonous Japanese encephalitis with yellow fever coinfection in Africa. *New England Journal of Medicine* 376, 1483–1485. (6) World Health Organization, Japanese encephalitis. Available from: <https://www.who.int/news-room/fact-sheets/detail/japanese-encephalitis>.

occur annually worldwide. The fatality rate can reach up to ~30% of JE cases, with a wide range of long-term neuropsychiatric sequelae developing in up to ~50% of JE survivors.

Genetic Diversity

All JEV isolates can be classified into one of the five genotypes (GI–GV), based on the complete nucleotide sequence of the viral genome; the sequence divergence is estimated to be as high as ~20% at the nucleotide level and ~10% at the amino acid level when their ORFs are compared. Of the five genotypes, GIII had been the major genotype detected from its first isolation in 1935 until the 1990s. In 2003, the first comprehensive phylogenetic study reported that the prevalence of the five genotypes varied by geographic locations in the Asia–Pacific area: (a) all five genotypes were found in the Indonesia–Malaysia region, (b) GI, GII, and GIII in the Thailand–Cambodia–Vietnam region, (c) GI and GIII in the Japan–Korea–China region, (d) GII and GIII in the Taiwan–Philippines region, (e) GIII in the India–Sri Lanka–Nepal region, and (f) GI and GII in the Australia–New Guinea region. It should be noted that since the early 2000s (or presumably even earlier), GIII has been gradually replaced by GI, which is now the dominant genotype, occasionally co-circulating with GIII in many Asian countries, including Thailand, Cambodia, Vietnam, Japan, Korea, China, and India. In the late 2000s and early 2010s, GV also emerged for the first time outside of the Indonesia–Malaysia region, in Korea and China.

With respect to the origin of JEV, it has been proposed that the virus initially evolved from its ancestral virus into the five genotypes in the Indonesia–Malaysia region, where they either circulate in that region or spread across the Asia–Pacific area. A closer look at the successful emergence and establishment of the currently dominant GI has suggested that it originated from the Indonesia–Malaysia region, diverged in the Thailand–Cambodia–Vietnam region, and then dispersed northward to southwestern China, from which it subsequently spread to other JE-affected countries (e.g., Japan, Korea, and Taiwan).

Although JEV presents only a single serotype, a considerable degree of antigenic variation has been noted not only between two different genotypes but also among strains of the same genotype. In line with this situation, relatively low levels of cross-neutralizing antibody response against heterologous genotypes (e.g., GI and GV) have been observed among people who previously received a GIII-based killed or live JE vaccine. Also, GI strains have been isolated from JE patients who had already been immunized with a GIII-based live JE vaccine. Thus, the genetic and antigenic variation of JEV may have a negative impact on the development of broadly effective JE vaccines.

Transmission

Vector-Borne Transmission

JEV is an arthropod-borne virus (arbovirus) transmitted by a hematophagous mosquito vector between animal hosts (Fig. 6). The most competent mosquito vectors are *Culex* species, of which *Cx. tritaeniorhynchus* is primarily responsible for the recurrence and expansion of JEV in Asia, and *Cx. annulirostris* mainly accounts for the introduction and spread of JEV in northern Australia. Notably, the primary vector *Cx. tritaeniorhynchus* is widely distributed not only in Asia, where JEV has been circulating regularly, but also in Europe and Africa, where JEV may have been introduced recently (Fig. 5). Other *Culex* species implicated in local JEV transmission include *Cx. annulus*, *Cx. bitaeniorhynchus*, *Cx. fuscocephala*, *Cx. gelidus*, *Cx. orientalis*, *Cx. pipiens*, *Cx. pseudovishnui*, *Cx. quinquefasciatus*, *Cx. sitiens*, *Cx. tarsalis*, and *Cx. vishnui*, from each of which the virus has been detected. JEV infection has also been documented in a small number of field-caught or laboratory-inoculated non-*Culex* species, such as four *Aedes* species (*Ae. albopictus*, *Ae. detritus*, *Ae. japonicus*, and *Ae. vexans*), three *Anopheles* species (*An. minimus*, *An. sinensis*, and *An. tessellatus*), one *Armigeres* species (*Ar. subalbatus*), and one *Mansonia* species (*Ma. uniformis*). However, these JEV-susceptible *Culex* and non-*Culex* mosquitoes are not all equally competent for JEV transmission, and thus their competency needs to be examined.

JEV is maintained in an enzootic cycle involving several vertebrate hosts (Fig. 6), of which suids (e.g., domestic pigs) and ardeid birds (e.g., herons and egrets) are known as the main virus-amplifying hosts, since they generally develop asymptomatic infection but produce a relatively prolonged high-titer viremia sufficient to infect engorging mosquitoes. In particular, domestic pigs play a major role in transmitting the virus to humans in many Asian countries, where pig farms are usually located near human residences; thus they can serve as a sentinel animal. In the mosquito-vertebrate-mosquito cycle of JEV, not only migratory birds but also hibernating/migrating bats are considered as reservoir hosts that can contribute to the overwintering and long-range dispersal of the virus. Another possible overwintering mechanism for JEV is transovarial transmission, passing the virus directly from a female mosquito to its eggs. On the other hand, humans and equids (e.g., horses and donkeys) are recognized as incidental dead-end hosts, since they typically produce a level of viremia insufficient for further transmission but develop clinical illness. Limited data from serosurveys, along with experimental infection studies in some cases, have suggested that JEV can cause subclinical infection in many other vertebrates (e.g., cows, goats, sheep, chickens, ducks, dogs, cats, rabbits, buffaloes, raccoons, and raccoon dogs), as well as in some reptiles and amphibians; however, their roles in JEV transmission remain undefined.

Non-Vector-Borne Transmission

Although JEV is a mosquito-borne pathogen, it can be transmitted in the absence of a mosquito vector via several non-vector-borne routes: (a) direct contact transmission, which has been demonstrated experimentally in pigs and mice; (b) aerosol transmission,

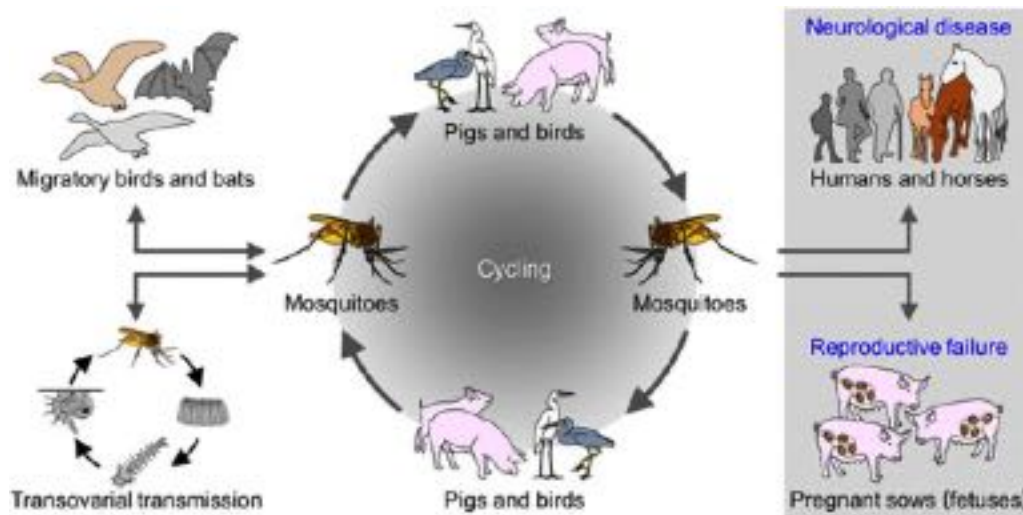


Fig. 6 Vector-borne transmission cycle of JEV. In nature, the virus is horizontally maintained primarily by culicine mosquito vectors between vertebrate animal hosts, with domestic pigs and ardeid birds serving as the most important virus-amplifying hosts/reservoirs. Typically, symptomatic infections are observed in humans, horses, and pregnant sows (fetuses). Also, an infected female mosquito is able to vertically transmit the virus to her eggs. The schematic drawing is modified from a review article (Yun, S.I., Lee, Y.M., 2014. Japanese encephalitis: The virus and vaccines. *Human Vaccines & Immunotherapeutics* 10, 263–279).

which has been shown in several laboratory animals, including rodents and squirrel monkeys; (c) transplacental transmission, which is known to occur commonly in pregnant sows but rarely in pregnant women and is reproduced experimentally in pregnant mice; (d) blood-borne transmission, which has been reported to occur in humans during blood transfusion and organ transplantation; and (e) artificial insemination, which is thought to be a potential mode of transmission from infected boars to breeding sows. Further studies are needed to address the impact of these instances of vector-free transmission on JEV ecology and epidemiology.

Clinical Features

JEV is a zoonotic pathogen that can cause a spectrum of neurological diseases in both people and livestock, such as horses, and reproductive problems in pigs (Fig. 6). In humans, the clinical signs after JEV infection typically appear within a week, but the incubation period may be as long as 2 weeks. While most infections are asymptomatic, some present a range of clinical manifestations, from mild febrile illnesses that may remain undiagnosed to severe neuroinvasive diseases that are often fatal, depending on the composite interaction between viral virulence factors and host risk factors (e.g., age and genetic makeup).

In particular, the potentially life-threatening neurological disease caused by JEV infection is JE, acute encephalitis that is characterized by extensive inflammation in the central nervous system (CNS). The encephalitis begins with prodromal flu-like symptoms, including fever, anorexia, nausea, vomiting, malaise, myalgia, headache, and back pain. During this period, liver dysfunction, gastric hemorrhage, and thrombocytopenia have also been observed in some patients. The prodromal phase may progress to an acute encephalitic phase, marked by decreased alertness and various neurological signs such as a mask-like expression, muscle rigidity, tremulous eye movements, coarse extremity tremors, abnormal involuntary movements, and pathological reflexes. Also, a polio-like acute flaccid paralysis has been described in JE patients. Patients with more severe and profound neurological signs during the encephalitic phase are more likely to experience long-term sequelae, such as persistent seizures, motor neuron weakness, cerebellar signs, extrapyramidal disorders, arm flexion deformities, leg hyperextensions, cognitive deficits, language impairment, learning disabilities, and behavioral problems.

Many animals are believed to be susceptible to JEV infection, but horses and pigs are known to show overt clinical signs of JEV infection. Like humans, horses may develop a variety of clinical conditions ranging from self-limiting illnesses with nonspecific signs (e.g., fever, lethargy, anorexia, and congested mucous membranes) to mild or severe forms of encephalitis with neurological signs (e.g., dysphagia, ataxia, wandering, neck stiffness, radial paralysis, vision impairment, aggressive behavior, and tremors). Occasionally, neurological deficits can persist for a long period of time after recovery. Although uncommon, encephalitis is also reported to occur in cows and can eventually lead to death.

In pigs, JEV infection during pregnancy commonly leads to reproductive losses, such as abortion, stillbirth, and mummification. Even though infected piglets are born alive, they often display neurological signs (e.g., tremor and convulsion), die soon after birth, and present with neuropathological lesions (e.g., hydrocephalus, cerebellar hypoplasia, and spinal hypomyelination). However, nonpregnant pigs generally show no symptoms or only a low-grade fever but occasionally develop neurological conditions, especially at a young age. In boars, JEV infection can disturb spermatogenesis, leading to a decrease in sperm count and motility.

Pathogenesis

JEV is a neurotropic virus that affects the CNS, highlighted by neuronal cell death. For all neurotropic viruses, virulence is determined by a set of viral factors that interact with their host targets during infection and define two pathogenic properties: (a) neuroinvasiveness, the ability of the virus to penetrate the CNS from peripheral sites where the virus is inoculated by a mosquito bite; and (b) neurovirulence, the ability of the virus to establish a lethal infection within the CNS. Over the years, significant progress has been achieved in determining the structure and function of the viral proteins that are critical for the various stages of JEV replication, but limited progress has been made in identifying the viral factors required for JEV virulence and the sequence of events involved in JEV pathogenesis.

Based on our current understanding of viral pathogenesis, JEV may replicate initially in leukocytes (e.g., lymphocytes, monocytes, and/or macrophages/dendritic cells) in peripheral lymphoid organs (e.g., lymph nodes and spleens) prior to penetration of the CNS. JEV can invade the CNS by crossing the blood-brain barrier (BBB) as cell-free or cell-associated virions or by directly infecting vascular endothelial cells. The replication of JEV in endothelial cells can compromise the integrity of the BBB. Within the CNS, JEV infects not only neurons but also non-neuronal cells, such as microglia and astrocytes. The viral replication in neurons directly causes cell death, which then activates the adjacent non-neuronal cells to produce a variety of inflammatory mediators (e.g., IL-1, IL-6, IL-8, TNF- α , MCP-1, IP-10, and RANTES) capable of promoting leukocyte migration and infiltration into the brain. A massive inflammatory response causes an overproduction of reactive oxygen/nitrogen species and cytokines (e.g., TNF- α), which then triggers bystander cell death in uninfected neurons.

Thus far, a multitude of viral genetic factors, including proteins and RNA elements, are shown to contribute to JEV virulence, but the viral surface glycoprotein E is believed to be the key factor that determines viral virulence, including both neuroinvasiveness and neurovirulence. Further research is required to define the viral factors required for JEV virulence, their functions, and the necessary virus-host interactions.

Diagnosis

Clinical Diagnosis

JE patients may have some abnormalities in (a) complete blood count, showing an elevated leukocyte count ($10 - 35 \times 10^9$ cells L⁻¹) with a high percentage of neutrophils; (b) cerebrospinal fluid (CSF) analysis, revealing an elevated opening pressure (> 250 mm), moderate pleocytosis ($10 - 100$ cells mm⁻³), and slightly increased protein level ($50 - 200$ mg dL⁻¹); and (c) electroencephalogram, displaying diffuse theta and delta waves, a burst suppression pattern, epileptiform activity, and alpha coma. The early clinical signs of CNS involvement include abnormal oculocephalic reflex, acute onset hemiparesis with hypertonia, and decorticate and decerebrate posturing.

Molecular imaging techniques such as X-ray computed tomography (CT), magnetic resonance imaging (MRI), and single-photon emission computed tomography (SPECT) are applicable to the diagnosis of JE. In JE patients, CT and MRI scans show brain lesions most often in the thalami, and less often in the basal ganglia, hippocampus, substantia nigra, cerebral cortex, cerebellum, brain stem, and white matter. MRI is generally more sensitive than CT at detecting brain lesions, particularly in the early phase of JE. SPECT scans can detect a significant level of hyperperfusion in the thalamus and putamen in the acute encephalitic stage, which can be extended to the frontal lobes at an advanced stage. Overall, these clinical diagnostics and neuroimaging are informative but nonspecific, so the etiological agent needs to be identified by laboratory diagnosis.

Laboratory Diagnosis

JEV infection is diagnosed by the detection of infectious virions, viral antigens/RNAs, and viral antibodies: (a) Infectious JEV can be isolated from clinical specimens (e.g., blood, CSF, and tissues) by inoculating them in a culture of susceptible primary cells or continuous cell lines, or more efficiently by injecting them into the brains of suckling mice. However, JEV isolation from clinical samples is generally challenging because of the low level of circulating virions, the rapid development of neutralizing antibodies, and the rapid clearance of transient viremia. (b) JEV-specific antigens can be identified using JEV-specific antibodies and a variety of immunological methods based on hemagglutination, coagglutination, immunofluorescence, or antigen-capture immunoassay. JEV RNAs can be amplified using JEV-specific primers and a range of molecular techniques, including conventional reverse transcription-polymerase chain reaction (RT-PCR), real-time quantitative RT-PCR, and reverse transcription loop-mediated isothermal amplification (RT-LAMP). (c) JEV-specific IgMs can be detected in the blood and CSF of JE patients by an IgM antibody-capture enzyme-linked immunosorbent assay (MAC-ELISA) or its derivatives, including a biotin-labeled antigen sandwich ELISA, a nitrocellulose membrane-based IgM-capture dot enzyme immunoassay, and an antibody-capture solid-phase radioimmunoassay. Other traditional serological tests such as hemagglutination inhibition, complement fixation, and neutralization can also be applied to the detection of JEV-specific IgM. It should be noted, however, that all the serology-based data for the detection of both JEV antigens and antibodies need to be interpreted with caution because there is some cross-reactivity with other flaviviruses.

Treatment

There is no clinically approved drug or therapy available for the treatment of JEV infection. Supportive care is the only current option to relieve the symptoms caused by JEV infection.

Prevention

JEV can be prevented by reducing the population of mosquito vectors (e.g., using insecticides), protecting people from mosquito bites (e.g., applying mosquito nets and repellents), and immunizing susceptible individuals with a vaccine. Of these measures, vaccination is the only one available for sustainable long-term protection against JEV. Thus far, four types of JE vaccines have been commercialized in various countries for humans: (a) The mouse brain-derived killed-inactivated vaccine JE-VAX was developed using the Japanese Nakayama strain, isolated from the brain of a patient in 1935. First licensed in Japan in 1954, JE-VAX was widely used internationally until the late 2000s, when the new cell culture-derived vaccines became available (see below). In Japan, a similar mouse brain-derived killed vaccine was also produced using the Chinese Beijing-1 (P1) strain, recovered from a patient's brain tissue in 1949 and distributed domestically from 1989 to the mid-2000s. (b) The cell culture-derived live-attenuated vaccine SA14-14-2 was produced in primary hamster kidney (PHK) cells using the SA₁₄-14-2 strain, an attenuated derivative of the virulent parental SA₁₄ strain isolated from a collection of *Cx. pipiens* larvae in China in 1954. Since its initial licensure in China in 1988, SA14-14-2 has become the most widely used JE vaccine in many Asian countries, including Cambodia, India, Laos, Myanmar, Nepal, South Korea, Sri Lanka, and Thailand. (c) The first cell culture-derived killed-inactivated vaccine was generated in PHK cells using the Chinese Beijing-3 (P3) strain, isolated during a JE outbreak in 1949. In China, this vaccine was licensed in 1968; it was then adapted by production in African green monkey kidney (Vero) cells, a method that was licensed in 1998 but has more recently been replaced by the PHK cell-produced live SA14-14-2 vaccine. Also, two other Vero cell-produced killed vaccines were commercialized: one produced with the Beijing-1 strain and marketed in Japan as JEBIK V or ENCEVAC in 2009–2011; and the other produced with the SA₁₄-14-2 strain and marketed since 2009 in many countries, including Australia, Canada, Hong Kong, India, Switzerland, the US, and Europe, under one of four trade names (IC51, IXIARO, JESPECT, and JEEV). (d) The cell culture-derived live chimeric vaccine ChimeriVax-JE (also known as IMOJEV, JE-CV, and THAIJEV) was created by genetically engineering the live-attenuated YFV 17D vaccine as a vector, replacing its prM-E coding region with the corresponding region of JEV SA₁₄-14-2. The Vero cell-produced ChimeriVax-JE vaccine is currently available in Australia and Thailand. In addition to humans, pigs and horses are also vaccinated in some countries to prevent JEV-related disease.

Given the fact that all of the discontinued and currently available JE vaccines are based on only one of four GIII strains (Nakayama, Beijing-1, Beijing-3, and SA₁₄-14-2), the genetic and antigenic variations among all five genotypes should be taken into consideration when developing next-generation JE vaccines, to provide broad protection against all the JEV strains belonging to different genotypes.

Further Reading

- Brinton, M.A., Basu, M., 2015. Functions of the 3' and 5' genome RNA regions of members of the genus *Flavivirus*. *Virus Research* 206, 108–119.
- Chao, L.H., Klein, D.E., Schmidt, A.G., Pena, J.M., Harrison, S.C., 2014. Sequential conformational rearrangements in flavivirus membrane fusion. *eLife* 3, e04389.
- Chen, S., Wu, Z., Wang, M., Cheng, A., 2017. Innate immune evasion mediated by *Flaviviridae* non-structural proteins. *Viruses* 9, e291.
- Dong, H., Fink, K., Zust, R., *et al.*, 2014. Flavivirus RNA methylation. *Journal of General Virology* 95, 763–778.
- Hasan, S.S., Sevvana, M., Kuhn, R.J., Rossman, M.G., 2018. Structural biology of Zika virus and other flaviviruses. *Nature Structural & Molecular Biology* 25, 13–20.
- Hegde, N.R., Gore, M.M., 2017. Japanese encephalitis vaccines: Immunogenicity, protective efficacy, effectiveness, and impact on the burden of disease. *Human Vaccines & Immunotherapeutics* 13, 1–18.
- Klema, V.J., Padmanabhan, R., Choi, K.H., 2015. Flaviviral replication complex: Coordination between RNA synthesis and 5'-RNA capping. *Viruses* 7, 4640–4656.
- Lindenbach, B.D., Murray, C.L., Thiel, H.-J., Rice, C.M., 2013. *Flaviviridae*. In: Knipe, D.M., Howley, P.M., Cohen, J.I., *et al.* (Eds.), *Fields Virology*. Philadelphia, PA: Wolters Kluwer Health, pp. 712–746.
- Luo, D., Vasudevan, S.G., Lescar, J., 2015. The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development. *Antiviral Research* 118, 148–158.
- Mackenzie, J.S., Barrett, A.D., Deubel, V., 2002. The Japanese encephalitis serological group of flaviviruses: A brief introduction to the group. *Current Topics in Microbiology and Immunology* 267, 1–10.
- Neufeldt, C.J., Cortese, M., Acosta, E.G., Bartenschlager, R., 2018. Rewiring cellular networks by members of the *Flaviviridae* family. *Nature Reviews Microbiology* 16, 125–142.
- Pijlman, G.P., Funk, A., Kondratieva, N., *et al.*, 2008. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host & Microbe* 4, 579–591.
- Turtle, L., Solomon, T., 2018. Japanese encephalitis – The prospects for new treatments. *Nature Reviews Neurology* 14, 298–313.
- Villordo, S.M., Carballeda, J.M., Filomatori, C.V., Gamarnik, A.V., 2016. RNA structure duplications and flavivirus host adaptation. *Trends in Microbiology* 24, 270–283.
- Wang, X., Li, S.H., Zhu, L., *et al.*, 2017. Near-atomic structure of Japanese encephalitis virus reveals critical determinants of virulence and stability. *Nature Communications* 8, 14.

Relevant Websites

- <https://www.cdc.gov/japaneseencephalitis/>
Centers for Disease Control and Prevention.
- https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/257/flaviviridae
International Committee on Taxonomy of Viruses.

https://www.aphis.usda.gov/animal_health/emergency_management/downloads/disease_strategy_jev.pdf

United States Department of Agriculture.

<https://www.who.int/news-room/fact-sheets/detail/japanese-encephalitis>

World Health Organization.

http://www.oie.int/fileadmin/Home/eng/Animal_Health_in_the_World/docs/pdf/Disease_cards/JAPANESE_ENCEPHALITIS.pdf

World Organization for Animal Health.

Kaposi's Sarcoma-Associated Herpesvirus (*Herpesviridae*)

Anne K Cordes, Hannover Medical School, Institute of Virology, Hannover, Germany

Thomas F Schulz, Hannover Medical School, Institute of Virology, Hannover, Germany and German Center for Infection Research, Hannover-Braunschweig Site, Braunschweig, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of T. Toptan, Y. Chang, P.S. Moore, Kaposi's Sarcoma-Associated Herpesvirus: General Features, Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:[10.1016/B978-0-12-801238-3.02612-X](https://doi.org/10.1016/B978-0-12-801238-3.02612-X) and E. Gellermann, T.F. Schulz, Kaposi's Sarcoma-Associated Herpesvirus: Molecular Biology, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00686-5](https://doi.org/10.1016/B978-012374410-4.00686-5).

Nomenclature

ALR	AIM2-like receptor	MSM	Men who have sex with men
ATP	Adenosine triphosphate	NEMO	NF- κ B essential modifier
AZT	Azidothymidine	NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B-cells
Bub1	Budding uninhibited by benzimidazoles 1	NLR	Nucleotide-binding oligomerization domain-like receptors
CDK-6	Cyclin-dependent kinase 6	NLS	Nuclear localization signal
CENP-F	Centromeric protein F	NPM	Nucleophosmin
cGAS	Cyclic GMP-AMP Synthase	NuMA	Nuclear mitotic apparatus protein
CIB1	Calcium and integrin-binding protein-1	Ori	Origin of replication
cKS	Classical Kaposi's Sarcoma	PAMP	Pathogen-associated molecular patterns
DBD	DNA-binding domain	PCNA	Proliferating cell nuclear antigen
DC-SIGN	Dendritic cell-specific ICAM-grabbing non-integrin	PCR	Polymerase chain reaction
DNA	deoxyribonucleic acid	PEL	Primary effusion lymphoma
ELISA	Enzyme linked immunosorbent assay	PI3-K	Phosphoinositide 3-kinase
EphA2R	Ephrin receptor A2R	PML	Promyelocytic leukemia
ESCRT	Endosomal sorting complexes required for transport	POD	Promyelocytic leukemia (PML) oncogenic domains
FAK	Focal adhesion kinase	PPR	Pathogen recognition receptors
HAART	Highly-active antiretroviral therapy	pRb	Retinoblastoma protein
HHV-8	Human herpesvirus-8	Pre-RC	Pre-replication complex
HIV	Human immunodeficiency virus	RFC	Replication factor C
IFIT	Interferon induced proteins with tetratricopeptide repeats	Rho	Ras homolog
IFN	Interferon	RLR	retinoic acid-inducible gene-I-like receptors
IKK	I κ B kinase	RNA	Ribonucleic acid
IL	Interleukin	RTA	Replication and transcription activator
IRF3	Interferon regulatory factor 3	RV	Rhadinovirus
K-bZIP	Basic domain-leucine zipper	SDS PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
KCP	Complement control protein	STING	Stimulator of interferon genes
kDa	Kilodalton	SUMO2	Small ubiquitin-related modifier 2
KICS	KSHV inflammatory cytokine syndrome	TBK1	TANK-binding kinase 1
KS	Kaposi's sarcoma	TLR	Toll-like-receptor
KSHV	Kaposi's sarcoma herpesvirus	TR	Terminal repeat
LANA	Latency-associated nuclear antigen	vbcl2	viral B cell lymphoma homolog
LBS	LANA binding site	vCyc	Viral cyclin
LUR	Long unique coding region	VEGFA	Vascular endothelial growth factor A
MCD	Multicentric Castleman's Disease	vFLIP	Viral Fas-associated protein with death domain-like interleukin-1 β -converting enzyme-inhibitory protein
MCM	Minichromosome maintenance	vGCR	Viral G protein-coupled receptor
MeCP2	Methyl CpG binding protein 2	vIAP	Viral inhibitor of apoptosis protein
MHC	Major histocompatibility complex	vIRF	Viral interferon regulatory factor
MIR	Modulator of immune recognition	vORF	Viral open reading frame
miRNA	Micro RNA	VR	Variable region
mRNA	Messenger RNA		

Glossary

Concatamer A long DNA molecule consisting of multiple identical subunits arranged in a head-to-tail linear conformation.

Episome A circular viral DNA.

Kaposi's Sarcoma An endothelial cell-derived tumor caused by KSHV.

Kaposi's Sarcoma-associated herpesvirus (KSHV/HHV-8) An oncogenic herpesvirus and the causative agent of Kaposi's Sarcoma, primary effusion lymphoma and multicentric Castleman's disease.

Latency The phase of the herpesvirus life cycle in which the virus does not replicate in a productive manner but persists in the host cell.

Multicentric Castleman's Disease (MCD) A group of rare lymphoproliferative disorders (Non-Hodgkin's lymphoma). The plasma cell variant of MCD is associated with KSHV.

Primary effusion lymphoma A B cell lymphoma (Non-Hodgkin's lymphoma), caused by KSHV that is characterized by a plasma cell differentiation profile and immunoglobulin mutations, indicating its origin in a post-germinal center B cell.

Tegument The space between virus capsid and envelope containing viral proteins.

Classification

Kaposi's sarcoma herpesvirus (KSHV) or human herpesvirus-8 (HHV-8) is a *Rhadinovirus* and the only member of this genus that can infect humans. Other members of the two rhadinovirus lineages (RV1 and 2) infect New and Old World primates. Rhadinoviruses (γ2-herpesviruses) belong to the subfamily *Gammaherpesviridae* in the family *Herpesviridae*. KSHV DNA was first discovered in Kaposi's sarcoma (KS) biopsies from AIDS patients suffering from KS in 1994. It is an oncogenic virus and Kaposi's sarcoma (KS) is one of the three known malignant diseases that can be caused by this virus. In addition to KS, these include two lymphoproliferative diseases, primary effusion lymphoma (PEL) and multicentric Castleman's Disease (MCD). In 2016, KSHV infection was also linked to the clinical syndrome with sepsis-like features and involving the overexpression of cytokines, which was named KSHV inflammatory cytokine syndrome (KICS).

Based on the sequence of the hypervariable regions (VR1 and VR2) of the K1 gene, seven different KSHV subtypes (A – F and Z) are known today. In addition, three variants (P, M, N) of the K15 gene at the other end of the viral genome occur in several KSHV subtypes and variants M and N are thought to be the result of recombination events with related rhadinoviruses. Some KSHV subtypes are typically found in certain geographic regions (e.g., subtype B in Africa, subtype D in Asia, subtype E in indigenous American populations).

Virion Structure

KSHV has an enveloped icosahedral capsid with 162 capsomers and a total diameter of about 150–200 nm. The capsid is surrounded by the tegument and the virus envelope which is studded with multiple glycoproteins (gB, gH, gL, gM, gN, ORF4 and gpK8.1A) that play a crucial role in virus binding and entry into the host cell.

Genome

The KSHV genome is made up of a double stranded DNA of 165–170 kb that consists of a long unique coding region (LUR) of 140 kb, which is flanked by multiple GC-rich terminal repeat subunits of 801 bp, resulting in a total length of the terminal repeat region of approx. 20–35 kb (Fig. 1). The viral LUR contains at least 86 viral open reading frames (vORFs), of which more than 20 are genes that were unique for KSHV when it was first sequenced and are therefore termed K genes (K1 – K15); since then similar genes have been found in other New World primate rhadinoviruses that were discovered subsequently. Many KSHV genes are conserved among some or all members of the herpesvirus family (ORF4 –ORF75). Furthermore, the virus genome encodes for 12 microRNAs (miRNAs) as well as for multiple non-coding RNAs and antisense RNAs (Fig. 1). Most of the viral genes are expressed during the lytic replication phase. Four KSHV genes and the miRNAs form the so called KSHV latency locus (ORF71, 72, 73, K12 and viral miRNAs) (Fig. 2), as they are expressed in latently infected cells. There are also viral genes that are expressed only in latently infected B cells (ORF K10.5), or are frequently expressed in tumor biopsies as part of a "relaxed latency" gene expression program (e.g., vIL6, K15).

In the virion, the KSHV genome is found in a linear form. After infection of the host cell and once the viral DNA has reached the nucleus, the viral DNA is circularized (episome) by host enzymes and associates with host cell histones. These chromatinized episomes are attached to the host chromosomes by the latency associated protein (LANA) and can thereby be transferred to latently infected dividing daughter cells.

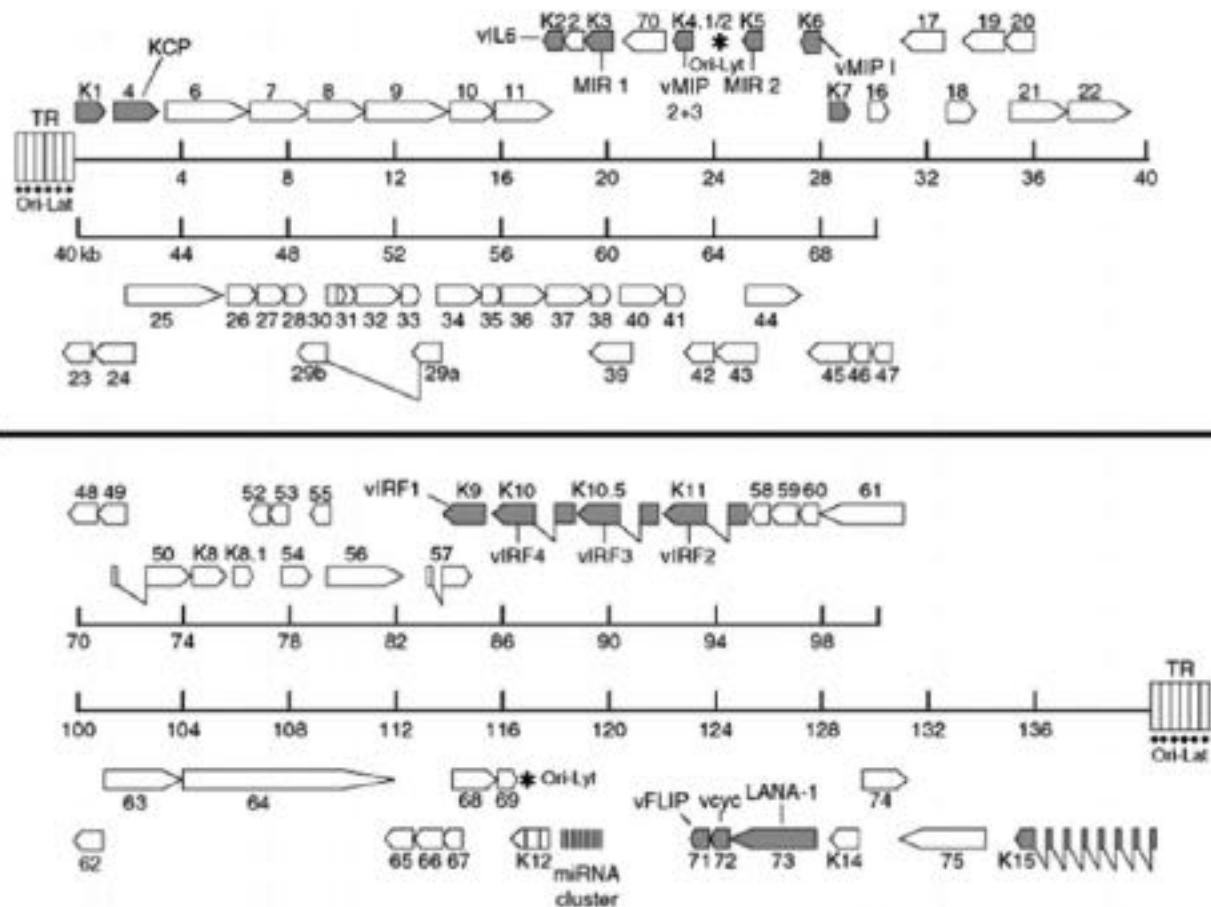


Fig. 1 Schematic map of the KSHV genome. Protein-coding ORFs are denoted ORF4–ORF75 and K1–K15, and the names of encoded proteins are given for some (see RefSeq NC_003409 and NC_009333). ORFs typical of Old World or New World Primate rhadinoviruses are shaded gray and other marked features are described in the text. The scale is in kbp. Reproduced from Gellermann, E., Schulz, T.F., 2008. Kaposi's sarcoma-associated herpesvirus: Molecular biology. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Elsevier. pp. 195–201.

Life Cycle

KSHV shows a broad host cell tropism. In vivo, the virus has been found in B cells, endothelial and epithelial cells, monocytes and also in keratinocytes. Two of the cell lineages play a major role in the pathogenesis of the three KSHV-associated tumors: infected endothelial cells are the origin of Kaposi's sarcoma, while infected B cells can progress to primary effusion lymphoma and multicentric Castlemann's disease.

Binding and Entry of the Virus to the Host Cell

The first interaction of the virus with its potential host takes place between glycoproteins that are embedded in the KSHV envelope and heparan sulfate proteoglycan on the cell surface. In addition, specific host cell glycoproteins on the host cell membrane serve as ligands for several viral glycoproteins, such as gH/gL, gpK8.1A, gM and gB. Host cell entry receptors include integrins $\alpha 3\beta 1$, $\alpha V\beta 3$ and $\alpha V\beta 5$, ephrin receptor A2R (EphA2R), DC-SIGN and xCT/CD98. In addition to attaching the virus to the cell surface, engagement of these cell surface receptors can induce an intracellular signal cascade that is important to facilitate virus entry and that may vary between different cell types. In HFF cells, the interaction of integrins with EphA2R induces the phosphorylation and activation of FAK, Src and PI3-K signal molecules. pPI3-K induces the activation of c-Cbl which in turn leads to polyubiquitination of EphA2R and the recruitment of Eps15 and AP2 resulting in the formation of clathrin-coated pits and uptake of the virus. Also in BJAB and HEK293 cells the uptake of the virus follows clathrin-mediated endocytosis. In HMVEC-d and HUVEC cells, interaction between integrins and xCT leads to activation of FAK, Src, PI3-K signal molecules, RhoA-GTPase and ROS signal molecules and to the recruitment of c-Cbl, CIB1, Crk and p130C as proteins. C-Cbl in turn induces monoubiquitination of integrins $\alpha 3\beta 1$ and $\alpha V\beta 3$ and leads to the translocation of these receptors and

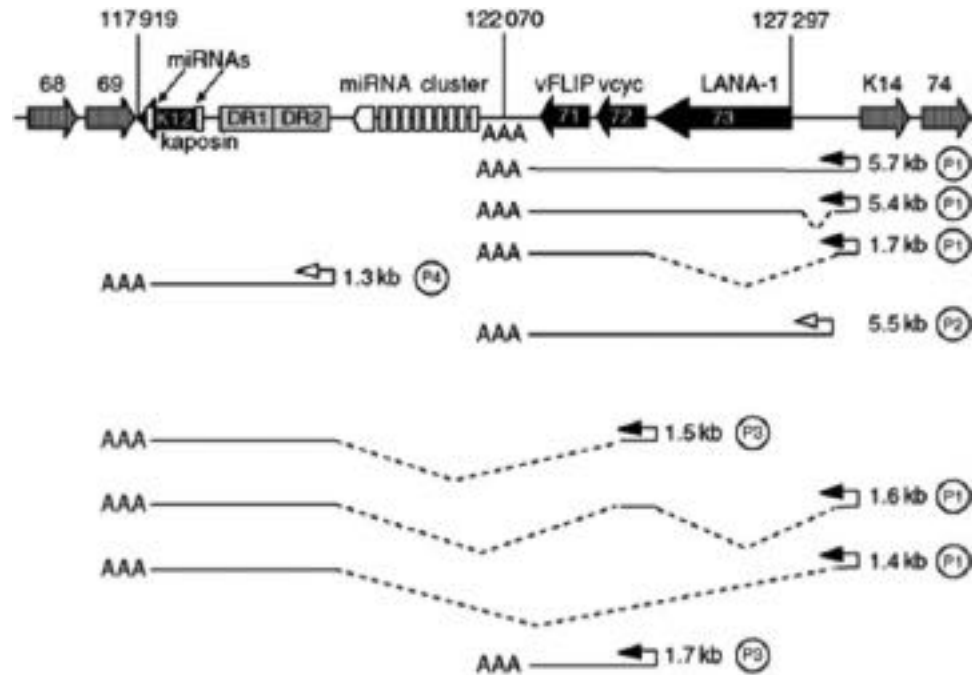


Fig. 2 The latency gene cluster in the KSHV genome. The KSHV latency locus encodes four proteins indicated by black arrows (K12/kaposin, ORF71/vFLIP, ORF72/vcyc, and ORF73/LANA-1), as well as 12 miRNAs (white boxes), and is flanked by lytic genes (striped arrows). Lytic promoters are indicated by white arrows and latent promoters by black arrows. Readthrough of a viral polyadenylation signal located at nucleotide 122,070 leads to miRNA expression. P1 (127880/86) gives rise to a precursor RNA that is spliced to produce 5.7 and 5.4 kb tricistronic (ORF71, ORF72, and ORF73) and 1.7 kb bicistronic (ORF71 and ORF72) mRNAs. Additionally, P1 drives two transcripts that could function as precursors for the miRNAs. As soon as RTA is expressed, P2 (127610) is induced and gives rise to a 5.5 kb mRNA encompassing all three ORFs. Lytic transcription from P4 (118758) leads to a 1.3 kb transcript coding for K12/kaposin. P3 (123751/60) controls a 1.7 kb transcript encoding vFLIP and vcyc and a 1.5 kb transcript that has the potential to encode the miRNAs. Sequence coordinates are derived from GenBank NC_003409. Reproduced from Gellermann, E., Schulz, T.F., 2008. Kaposi's sarcoma-associated herpesvirus: Molecular biology. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Elsevier. pp. 195–201.

xCT together with KSHV into the lipid raft region where the endosomal sorting complexes required for transport (ESCRT) is recruited; this in turn induces a signal cascade resulting in macropinocytosis of the virus. In THP-1 cells (monocytes), endocytosis was found to be mediated by clathrin as well as by caveolin.

Transport of the Virus in the Host Cytoplasm and Entry of Viral DNA Into the Nucleus

Following endocytosis/pinocytosis the virus lipid envelope fuses with the endosome membrane which releases the viral capsid into the cytoplasm. The viral capsid then takes advantage of the cellular microtubule system and host cell dynein motor proteins in an ATP dependent retrograde manner. Rho-GTPases that were activated by FAK and integrins acetylate host microtubules which in turn enable the transport of the viral capsid towards the nucleus. Once the virus has reached the nucleus, the double stranded linear viral DNA enters into the nucleus through a nuclear pore.

After entry of the viral DNA into the nucleus, there is a short phase in which a small number of lytic genes are expressed, among them ORF50 that encodes the Replication and Transcription Activator (RTA), the key regulator of the productive (lytic) replication cycle. The ensuing short burst of lytic replication is quickly switched off and it is thought that epigenetic mechanisms, in particular the gradually increasing deposition of histones with repressive chromatin marks are responsible for this transition to a latent stage of infection.

Latency

During latency most viral genes are silenced. Only a small number of genes are expressed from the latency locus which consists of ORFs 71, 72, 73, K12 and viral miRNAs (Fig. 2). These latent genes are under the control of a latent promoter located upstream of ORF73 as well as promoters located downstream of ORF73; several splicing events help to generate mRNAs for ORF73, ORF72/71 and K12, as well as for the precursor RNA that gives rise to the viral miRNAs (Fig. 2). An internal ribosomal entry site between ORF72 and ORF71 allows translation of ORF71 from the bi-cistronic ORF71/ORF72 mRNA (Fig. 2). In addition, a lytic promoter located within an intron immediately upstream of ORF73 is active and can direct LANA expression during the lytic replication cycle. In latently infected B cells ORF K10.5 is also expressed. The proteins encoded by these viral genes are crucial for the

maintenance of latency, since their silencing by siRNA inhibits growth of KSHV-infected primary effusion cell lines. These gene products and their specific functions will be discussed in the following.

ORF71 (K13) encodes for the protein vFLIP (viral Fas-associated protein with death domain-like interleukin-1 β -converting enzyme-inhibitory protein). Similar to cellular FLIPs, vFLIP inhibits death receptor-mediated apoptosis by inhibiting caspase 8 and 10. However, it is also a strong activator of the NF- κ B pathway as a result of its ability to bind to IKK γ /NEMO, an important NF- κ B pathway regulator. Activation of the NF- κ B pathway by vFLIP is important for cell proliferation and cell survival.

vCyclin is the protein encoded by ORF72. The protein is known to associate with the host cyclin-dependent kinase 6 (CDK-6). Together, these proteins phosphorylate a variety of substrates (pRb protein, histone H1, CDK inhibitor, nucleophosmin (NPM), p21 and p27 (Kip1)) resulting in the regulation of cell cycle and cell proliferation. Expressed on its own, vCyclin causes DNA damage. It is thought that the antiapoptotic property of vFLIP, which is translated simultaneously with vCyclin from a bi-cistronic mRNA (see above), may mitigate these properties of vCyclin.

The latency associated nuclear protein (LANA) is encoded by ORF73 and is a constitutively expressed protein that plays a major role in maintaining and replicating the latent viral episome in latently infected cells. Although ORF73 is part of the major latency locus that during latency is under the control of a latent promotor, in the lytic phase the transcription of LANA is inducible by RTA via a second, bidirectional promotor that lies downstream of the constitutively active promotor. This bidirectional promotor regulates the transcription of LANA and a bi-cistronic mRNA for ORF14 and the viral chemokine receptor homolog, vGCR, which is translated in the opposite direction from LANA.

The LANA protein contains an acidic, highly repetitive internal repeat domain that is flanked by an amino terminal domain (amino acid 1–312) and a basic carboxy terminal domain (amino acid 931–1162 in the prototypic KSHV strain) and has a deduced molecular mass of about 130 kDa, but an apparent molecular weight on SDS PAGE gels of approximately 220–250 kDa. Due to the varying length of the repetitive internal repeat domain in different KSHV strains, the length of the protein differs in different KSHV isolates.

During latency, LANA is mainly found in the nucleus, where the full-length protein is found in characteristic nuclear microdomains referred to as “LANA nuclear speckles”. The nuclear localization signal (NLS) is located in the N-terminal domain of the protein (amino acid 24–30). LANA is known to form dimers via its C-terminal DNA-binding domain (DBD). These LANA dimers attach the viral episome to host cell chromosomes, with the N-terminal domain binding to histones H2A and H2B on host chromosomes and the C-terminal domain binding to the LANA binding site within the terminal repeat sequence of the KSHV genome. This LANA-mediated linkage of the episome to the host chromosome is vital for segregation and the persistence of the virus in the host cell during latency. LANA is also required for the replication of the latent episome during the S-phase of the cell cycle by recruiting cellular factors such as minichromosome maintenance (MCM) proteins. In addition to these basic roles as origin-binding protein, replication of the latent viral episome and its partitioning to daughter cells during mitosis, LANA has also been shown to act as a transcriptional regulator and is involved in other cellular nuclear processes, such as the modulation of DNA repair mechanisms and of innate immune responses.

The ORF K12 encodes three different proteins: Kaposins A, B and C. Kaposin A and B have been shown to effect endothelial cells. Kaposin A induces cell proliferation and Kaposin B has been shown to induce RhoA dependent angiogenesis. Moreover, Kaposin B stabilizes cytokine mRNAs resulting in an increase in cytokine expression. The role of Kaposin C is yet unknown.

The major latency locus further comprises viral microRNAs (miRNAs). These miRNAs are short single stranded RNA fragments of 19–23 nucleotides in length. The miRNA cluster of the major latency locus comprises 12 pre-miRNAs that are processed into 25 mature miRNAs that are expressed during latency. Even though they are non-coding, these viral miRNAs are important for establishing latency in several aspects. Some interfere with cellular signal cascades such as IFN signaling (miR-K12-11) or the NF κ B pathway (miR-K1 targets I κ B α , an NF κ B regulator) and this may result in an inhibition of the lytic phase. Other miRNAs (miR-K2 and miR-K5) have been shown to be involved in KSHV induced angiogenesis through an induction of the expression of VEGFA and may thus contribute to KS oncogenesis. Furthermore, miR-K9-5p, miR-K12-7 and miR-K12-9 have been found to directly inhibit the expression of ORF 50 (RTA) and thereby prevent the switch from latency into the lytic phase.

ORF K10.5 encodes vIRF3, one of four viral interferon regulatory factor homologs, and is expressed constitutively in latently infected B cells. Also termed LANA-2, this protein has been shown to inhibit the NF κ B pathway by interacting with the β subunit of the IKK complex and was found to disrupt the promyelocytic leukemia (PML) oncogenic domains (PODs) by increasing the level of SUMO2-ubiquitin modification of the PML subunit of PODs. Furthermore, vIRF3 is known to inhibit p53 induced apoptosis and to interfere with the cellular interferon type 1 response.

Replication of the Latent KSHV Episome

In PEL cell lines, 30–80 copies of viral DNA persist in each latently infected cell; however, the number of latent viral genomes in an infected primary endothelial or B cell in vivo is not known. To avoid a loss of viral genomes during cell division each viral genome is replicated once per cell cycle and the viral genomes are then distributed equally onto the two daughter cells. This replication and distribution requires an accurate multi-step process.

DNA replication is initiated in the G1 phase of the cell cycle by the binding of LANA to the origin of replication (ori-P) via its LANA binding sites (LBS1–3) which in turn triggers the formation of a pre-replication complex (pre-RC) that associates at ori-P. The interaction of LANA with multiple other proteins is required to promote the viral DNA replication at ori-P and the

maintenance of the viral genome. The interaction of LANA with replication factor C (RFC) ATPase enables the loading of the sliding clamp protein PCNA onto the DNA which in turn promotes the DNA polymerase in its functions. Furthermore, LANA and PCNA form a complex together with the cellular mitotic kinase Bub1 that enhances the monoubiquitination of PCNA that also results in an increase in episome replication and maintenance. Moreover, the interaction of LANA with cellular Topoisomerase-2 β was described. The recruitment of these proteins by LANA to the TR region was reported to be crucial for viral DNA replication.

The replication of the viral DNA not only starts at ori-P, but several other origins of replication are expected to exist. At least one additional ori was identified and is thought to be LANA independent.

Genome Segregation

During mitosis, the latent virus genomes are tethered to host chromosomes to ensure a controlled segregation. LANA is a protein that plays a central role in this process. LANA provides the direct link between the histones (H2A and H2B) on mitotic chromosomes and the viral TR region. In addition, its interaction with other host chromatin proteins such as MeCP2 (Methyl CpG binding protein 2) has been reported to contribute to this tethering of viral episomes to mitotic chromosomes. Furthermore, LANA has been shown to recruit several other cellular proteins that are known to be involved in the control of mitosis: these include nuclear mitotic apparatus protein (NuMA), centromeric protein F (CENP-F), kinetochore protein, Bub1, and these may also play a role in promoting the segregation of viral episomal DNA.

Lytic Activation

The persistence of latent episomes in dividing cells with the help of the mechanisms described in the preceding paragraph is essential for the persistence of the virus in its host. However, the distribution of latent episomes dividing daughter cells is not 100% effective. It is therefore thought that reactivation of the virus from latency and the occasional production of viral progeny contributes to the long-term persistence of KSHV in infected individuals by periodically reinfecting new cells. In addition, infectious viral progeny is needed for the virus to spread to new hosts. Many more viral genes are expressed during the lytic than during the latent phase. The lytic gene expression pattern follows the classical pattern of immediate-early, early and late viral genes. The transcription of late genes occurs after the amplification the viral genome by the herpesviral DNA polymerase and is therefore susceptible to inhibitors of this viral enzyme, such as foscarnet. In contrast, early gene expression occurs in cells treated with DNA polymerase inhibitors. The early proteins are synthesized between 10 to 24 h, and the late lytic genes transcribed about 48 h, after the induction of the lytic phase. Several stimuli (e.g., chemical compounds like histone deacetylase inhibitors, hypoxia, oxidative stress) can induce this switch from latency to the lytic phase.

Crucial for the lytic phase is the immediate-early lytic protein RTA (regulator of transcriptional activation), a nuclear protein encoded by ORF 50. Expression of RTA is sufficient to induce the lytic cycle and also required in many stimulation conditions. RTA-independent modes of reactivating KSHV have however been described. In addition, a limited lytic gene expression program involving e.g., the vIL6 protein does not depend on RTA. RTA can activate both its own promotor and can also stimulate the transcription of multiple other lytic proteins that are necessary for lytic DNA replication and the assembly of new virions. The amplification of the viral DNA starts at two origins of lytic replication *ori-Lyt-L* (located between K4.2 and K5) and *ori-Lyt-R* (located between K12 and ORF71) and is initiated by the formation of the pre-replication complex that is induced by binding of RTA and K8 (K-bZIP) that again recruit further proteins. Binding of this pre-replication complex to *ori-Lyt* leads to a change in conformation of the viral DNA and to the formation of the replication initiation complex (complex of further viral and cellular proteins). The DNA amplification itself takes place via a rolling circle mechanism that leads to the synthesis of KSHV genome concatemers. Cleavage at the TR region leads to linear DNA genomes of defined length which are then packaged into new KSHV capsids that are assembled in the nucleus. The capsids most likely leave the nucleus via an envelopment (gaining an envelope by passing through the inner nuclear membrane)-de-envelopment (losing the envelope while passing through the outer nuclear membrane)-mechanism. Re-envelopment of the mature virus particle is poorly understood but is thought to take place after assembly of the virus particles in the Golgi apparatus by budding.

Interaction of KSHV With the Host Immune System

Infection of a host cell with KSHV leads to activation of the innate immune response via an interaction between pathogen-associated molecular patterns (PAMPs) and pathogen recognition receptors (PRRs) that results in an increased expression of interferons and pro-inflammatory cytokines. Several PRRs (e.g., TLRs and RLRs that activate RNA and/or DNA sensor proteins and NLRs and ALRs that can form an inflammasome complex) are known and PRRs can vary between different cell types. Together the combination of innate responses results in the expression of a range of different effector molecules.

During latency, KSHV evades the host immune system by expressing only a limited number of virus proteins. Some of these are able to modulate the immune system: thus LANA has been shown to interfere with the processing and MHC-mediated presentation of peptides to T cells. To evade the host immune system during viral lytic replication, KSHV possesses multiple mechanisms. Two viral proteins, modulator of immune recognition 1 and 2 (MIR1; MIR2), which are expressed in the early phase of lytic replication, interfere with MHC molecules and thereby abrogate the presentation of viral antigens on the cell surface.

Furthermore, several KSHV proteins, including LANA and ORF52, antagonize cGAS and thereby abrogate the activation of STING, TBK1 and IRF3 that results in a reduced expression of interferons. In addition, vIRF1, vIRF3, RTA and ORF45 also inhibit the interferon- and/or NF κ B-induced cytokine expression. Moreover, ORF63 inhibits NLRP1 and NLRP3 which results in a reduced production two pro-inflammatory cytokines IL-1 β and IL-18. KSHV ORF4 encodes a complement regulatory protein (KCP) that antagonizes the deposition of activated complement factors C3b and C4b on the surface of the infected cell.

There are also examples of KSHV taking advantage of the innate immune response to promote latent persistence. Thus vIRF2 suppresses lytic KSHV replication by activating certain cellular interferon-induced genes such as IFIT 1–3, which in turn restrict lytic replication.

Epidemiology

The prevalence of KSHV varies in different geographic regions of the world. Serological studies employing either immunofluorescence assays on latently infected or lytically induced PEL cell lines or ELISAs based on recombinant proteins or synthetic peptides have shown that the seroprevalence of KSHV in HIV-negative blood-donors in Western and Northern Europe as well as North America is generally less than 3%, although different results have been obtained with different serological assays. However, in Sub-Saharan Africa the sero-prevalence in HIV-negative adults is much higher and in the range of 35%–50%. Seroprevalence rates may be even higher in some East African countries, where even young children (0–4 years) have been reported to show seroprevalence rates of up to 25%. In countries bordering on the Mediterranean Sea, such as parts of Italy, Turkey and Israel seroprevalence rates in the range of 10%–20% have been reported, with considerable regional variation.

Sexual orientation and lifestyle can affect KSHV seroprevalence, which is higher among men who have sex with men (MSM) and individuals with a high number of sexual partners. Other infections may increase the risk of infections, in particular HIV, but also malaria and helminths, which are considered risk factors for KSHV infection in East Africa and may explain the higher seroprevalence rates in rural populations in this geographic region.

These epidemiological findings suggest that sexual contact provides an important route of infection among adults. KSHV can be detected in saliva by PCR and transmission by saliva is considered an important route of transmission, including during sexual contact. Prior to puberty, having a KSHV-seropositive sibling is a very strong risk factor for a child to be infected with KSHV. Transmission via saliva represents a likely route of the high prevalence in children living in KSHV-endemic areas. In addition, transmission via breast milk and under birth have been postulated. Studies in populations in endemic countries also suggest that there may be a genetic predisposition to become infected with KSHV before puberty. KSHV is thought to also be transmissible via blood transfusions and there is some evidence that intravenous drug use may represent a risk factor.

Clinical Features

Kaposi's Sarcoma

Kaposi's sarcoma (KS) lesions often appear as dark red spots or nodules on the skin, but also the mucosa in the oral cavity as well as internal organs may be affected. In the early stages, the skin alterations are referred to as patch lesions. These progress to raised plaques and then firm nodules, which can grow, disturb the lymph drainage and result in edema as well as ulcerate. Kaposi's sarcoma metastasizes rarely, and individual distinct nodules are thought to have resulted from the independent infection of multiple precursor cells. KS lesions can infiltrate and destroy the local tissue.

Epidemiologically, Kaposi's sarcoma can be subdivided into four forms that carry a different clinical prognosis. Classical Kaposi's sarcoma mainly occurs in elderly men. It is often found on the lower limbs and shows only a slow progression. Due to the patients' higher ages and slow tumor progression, classic KS is not normally the cause of death for these patients.

Post-transplantation or iatrogenic Kaposi's sarcoma usually occurs more than a year after an organ transplant and is associated with a suppressed immune system.

The African endemic form of KS occurs in East and Central Africa, is not associated with HIV, and can run a clinically more aggressive course than classic KS, with the involvement of internal organs such as the lung and intestine.

The HIV-associated form, defined as KS in an HIV-positive patient, is today the most common form. Its severity can vary, but rapid progression with the extensive involvement of multiple visceral organs is often seen and can rapidly lead to the death of affected patients. Thanks to effective antiretroviral combination therapy, HIV-associated KS can be controlled in most patients; however, KS may emerge or reemerge and progress dramatically in rare patients whose HIV viral load is well controlled.

Primary Effusion Lymphoma

The KSHV associated non-Hodgkin's Lymphoma PEL occurs mainly in HIV infected men. Clinically, it manifests itself as a pleural or abdominal effusion containing high numbers of malignant cells, often in the absence of an obvious solid tumor mass, which may later be found to line the pleural or abdominal cavity. Progression is often rapid and the disease is often fatal.

Multicentric Castleman's Disease

MCD also belongs to the group of non-Hodgkin's Lymphomas. KSHV is mainly associated with the plasma cell variant of MCD. As in the case for PEL, the plasma cell variant of MCD is a rare disease that often occurs in combination with HIV infection. It is associated with splenomegaly, lymphadenopathy and polyclonal hyperimmunoglobulinemia.

Pathogenesis

KSHV is an oncogenic virus that causes Kaposi's sarcoma and the two HIV associated non-Hodgkin's Lymphomas PEL and MCD. Several viral proteins, as well as the KSHV miRNAs, are thought to be involved in its oncogenic properties, but their precise contribution to the development of the three KSHV-associated tumors is still under investigation and likely varies in different tumor entities. Based on siRNA silencing experiments and experiments in transgenic mice it is thought that the three latent viral proteins LANA, vCyc and vFLIP play an essential role in the development of KSHV-associated lymphoid neoplasia. Some of their properties discussed above are likely to play a role in this context, in particular the ability of LANA to antagonize cellular DNA repair pathways including p53-dependent responses, the vCyc-mediated dysregulation of the cell cycle and DNA damage and the anti-apoptotic properties of vFLIP. In addition, viral proteins that are expressed during the early stages of the lytic replication cycle or as part of a "relaxed latency" transcriptional program are thought to be important for the development of these three tumors. Transcriptome analyses of KS tumors have shown that, in addition to the viral genes located in the latency locus (LANA, vCyc, vFLIP, K12, viral miRNAs; see above), others, in particular the viral protein encoded by ORF K15, vIL6, vGCR are frequently expressed in KS tumors. Experimental evidence obtained for these three proteins in tissue culture experiments or transgenic mice suggests that they may contribute to the atypical angiogenesis seen in KSHV-infected lesions, B-cell proliferation and/or the growth of vascular tumors. In addition, several other viral proteins, including the K1 protein, several of the viral interferon regulatory factor homologs (vIRFs) and the anti-apoptotic viral proteins vBcl2 and vIAP have cell biological properties that are compatible with a role in oncogenesis.

Kaposi's Sarcoma

Kaposi's sarcoma is the most common tumor caused by KSHV. It has its origin in endothelial cells and, in advanced tumors, is characterized by an abundance of KSHV-infected atypical endothelial spindle cells. Infiltration of inflammatory cells is another typical histological feature. Transcriptome and immunohistochemistry studies using antibodies to viral proteins indicate that viral gene and protein expression in Kaposi's sarcoma lesions can vary. While viral genes encoded in the latency locus (see above) are most frequently expressed, other commonly expressed genes include for example the viral K15 gene and the number of endothelial spindle cells staining for the K15 protein varies in different tumors. Expression of other viral genes or proteins linked to oncogenesis (see above), including for example K1, vGCR and vIL6 has been documented in fewer cells in some tumors.

Primary Effusion Lymphoma

This rare lymphoproliferative disorder is characterized by immunoglobulin mutations in B cells, indicative of a post-germinal center differentiation stage. These cells accumulate mainly in the fluid of the pleural, pericardial or abdominal cavity. In some cases, a solid tumor mass can be found lining the pleural or peritoneal membranes. The expression of CD138/syndecan-1 is typical for PEL cells and indicates their plasma cell – like differentiation. LANA, vCyc, vFLIP, vIRF3 and viral miRNAs were found to play a major role in the proliferation of PEL cells and by implication in PEL pathogenesis. PEL cells carry multiple episomal (and in some cases integrated) KSHV genomes (about 50–100 KSHV genome copies per cell in established PEL cell lines). The majority of PEL cells are latently infected. The proportion of cells that have spontaneously switched on the lytic replication cycle varies in different PEL tumors and cell lines and ranges from less than 1% to 10%–15% of these KSHV infected cells.

Multicentric Castleman's Disease

The overproduction of cytokines, especially interleukin-6 and -10, as well as the increase in vascular endothelial growth factor seems to be responsible for the progression of the disease. MCD is often accompanied by productive viral replication with high viral loads detectable in the peripheral blood of affected patients. Lytic KSHV replication is therefore thought to play an important role in the pathogenesis of this disease. Attempts to treat MCD with a combination of high-dose ganciclovir and AZT (azidothymidine) have been successful in some cases, underlining the importance of viral lytic replication at least in some cases (see below).

KSHV Inflammatory Cytokine Syndrome

The KSHV Inflammatory syndrome (KICS) was first described in 2016 by Polizzotto *et al.* as a KSHV related clinical appearance that seems to be associated to Kaposi's Sarcoma and HIV co-infection and that goes along with a systemic inflammation and a

wide variety of symptoms (e.g., gastrointestinal disturbance, edema, respiratory symptoms, anemia and others) and that leads to an increase in mortality. KSHV lytic replication is thought to play an important role in the pathogenesis of this disease.

Diagnosis

Kaposi's Sarcoma

The diagnosis is currently based on the clinical appearance of KS lesions on the skin, the standard histopathology of KS biopsies, combined with the immuno-histochemical detection of KSHV LANA in histological sections of KS tumors.

Primary Effusion Lymphoma

The diagnosis is mainly based on cytological or histopathological examinations, immunohistochemical staining for LANA and the detection of KSHV DNA in the tumor cells.

Multicentric Castleman's Disease

The diagnosis is mainly based on histological examinations of lymph node biopsies, including the immunohistochemical staining for KSHV LANA.

Treatment

In tissue culture, KSHV replication can be inhibited by several competitive inhibitors of the herpesviral DNA polymerase, such as foscarnet, ganciclovir, brivudine and cidofovir. However, in infected patients, these drugs are not effective against established KSHV-associated tumors, supporting the notion that the latent viral gene expression program and latent viral replication, which does not rely on the viral DNA polymerase, plays an essential role in the pathogenesis of KS, PEL and MCD. Interestingly, pre-emptive treatment of HIV-infected MSM with ganciclovir to prevent disease caused by human cytomegalovirus reduces the number of KS in these patients, indicating that productive replication of KSHV contributes to the development of this tumor, possibly by increasing the number of KSHV-infected cells.

However, in the case of MCD, which is frequently associated with KSHV lytic replication and an increased KSHV load in peripheral blood (see above), a combination treatment of high dose ganciclovir and AZT has been shown to be effective in some patients. The rationale behind this regimen is that ganciclovir and AZT, both prodrugs that need to be activated by phosphorylation, are phosphorylated, respectively, by the viral ORF36 kinase and the viral thymidine kinase to generate cytotoxic drugs that are thus preferentially active in KSHV-infected cells.

Kaposi's Sarcoma

Classical Kaposi's sarcoma

The classical Kaposi's sarcoma (cKS) is responsive to radiation and local or systemic chemotherapies. No antiviral therapies with activity against cKS are in use.

Post-transplantation or iatrogenic Kaposi's sarcoma

The tumor can completely regress if the immunosuppressive medication is stopped. If it is not possible to discontinue the patients' immunosuppressive therapy, the tumors can be treated with radiation therapy.

HIV associated Kaposi's sarcoma

When treating the HIV associated form, it is crucial to reduce the HIV load and to increase the CD4 + -cell count of the patients by HAART (highly-active antiretroviral therapy). This therapy can lead to a regression of the tumor. If the HAART therapy is not sufficient though, an additional chemotherapy based on liposomal daunorubicin is the appropriate choice. A few studies have reported that the tyrosine kinase inhibitor imatinib (Gleevec) may be partially effective and can cause disease regression in a few cases.

Primary Effusion Lymphoma

The therapy includes a combined chemotherapy along with the treatment of the HIV infection. However, the prognosis of patients suffering from PEL is poor, with an average survival time of about six months after diagnosis.

Multicentric Castleman's Disease

MCD can be treated with chemotherapy. Experimental approaches to therapy include a combination therapy with high dose ganciclovir and AZT (see above) as well as with antibodies directed against IL-6 and its receptor.

Prevention

No Vaccination Against KSHV is Available Today

Since the development of KSHV-associated tumors is often associated with uncontrolled HIV infection, an effective HAART therapy should be pursued. However, HIV-associated KS, MCD and PEL can occur in patients with a well-controlled HIV viral load.

Further Reading

- Aneja, K.K., Yuan, Y., 2017. Reactivation and lytic replication of kaposi's sarcoma-associated herpesvirus: An update. *Frontiers in Microbiology* 8, 613.
- Dittmer, D.P., Damania, B., 2016. Kaposi sarcoma-associated herpesvirus: immunobiology, oncogenesis, and therapy. *Journal of Clinical Investigation* 126 (9), 3165–3175.
- Kumar, B., Chandran, B., 2016. KSHV entry and trafficking in target cells-hijacking of cell signal pathways, actin and membrane dynamics. *Viruses* 8 (11).
- Kumar, B., Roy, A., Veettil, M.V., Chandran, B., 2018. Insight into the roles of E3 ubiquitin c-Cbl ligase, ESCRT machinery, and host cell signaling in Kaposi's sarcoma-associated herpesvirus entry and trafficking. *Journal of Virology* 92 (4).
- Mariggio, G., Koch, S., Schulz, T.F., 2017. Kaposi sarcoma herpesvirus pathogenesis. *Philosophical Transactions of the Royal Society London B: Biological Sciences* 372 (1732).
- Purushothaman, P., Dabral, P., Gupta, N., Sarkar, R., Verma, S.C., 2016. KSHV genome replication and maintenance. *Frontiers in Microbiology* 7, 54.
- Schulz, T.F., Cesarman, E., 2015. Kaposi sarcoma-associated herpesvirus: Mechanisms of oncogenesis. *Current Opinion in Virology* 14, 116–128.
- Uppal, T., Jha, H.C., Verma, S.C., Robertson, E.S., 2015. Chromatinization of the KSHV genome during the KSHV life cycle. *Cancers* 7 (1), 112–142.
- Veettil, M.V., Bandyopadhyay, C., Dutta, D., Chandran, B., 2014. Interaction of KSHV with host cell surface receptors and cell entry. *Viruses* 6 (10), 4024–4046.
- Yan, L., Majerciak, V., Zheng, Z.M., Lan, K., 2019. Towards better understanding of KSHV life cycle: From transcription and posttranscriptional regulations to pathogenesis. *Virologica Sinica* 34 (2), 135–161.

Marburg and Ravn Viruses (*Filoviridae*)

Courtney Woolsey, Thomas W Geisbert, and Robert W Cross, The University of Texas Medical Branch, Galveston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

Epidemiology

Marburg virus (MARV) was first identified in 1967 from a contemporaneous cluster of cases of severe hemorrhagic fever in Germany and the former Yugoslavia (now Serbia). Laboratory workers became infected after processing African Green Monkey (*Chlorocebus aethiops*) tissues for the production of poliovirus vaccines. The monkeys were freshly imported from Uganda, suggesting the virus was of African origin. Detailed virological analysis including electron microscopy demonstrated a unique filamentous morphology and a lack of serological cross-reactivity to known viruses at the time, indicating this was a member of a new family of virus, a filovirus. In subsequent years, a relatively small number of cases and sporadic outbreaks of Marburg virus disease (MVD) surfaced primarily in the vicinity of the greater Lake Victoria area in Central Africa, including the eastern Democratic Republic of Congo, Uganda, and Kenya. Case fatality rates (CFR) ranged from 27% to 100% for these episodes (Fig. 1 and Table 1). In 2004–2005, the largest and most lethal outbreak of MVD occurred outside of this region in northwestern Angola, wherein over 250 cases were confirmed with a ~90% case fatality rate (CFR). The high mortality rate suggested something unique in this outbreak or enhanced transmissibility or virulence of the virus variant responsible. In 2008, exported cases in the Netherlands and United States were reported from travelers that visited a bat cave in the Lake Victoria region of Africa and later returned to their home countries prior to experiencing symptoms of MVD (Table 1). Subsequent cases in 2012, 2014, and 2017 have been limited to Uganda.

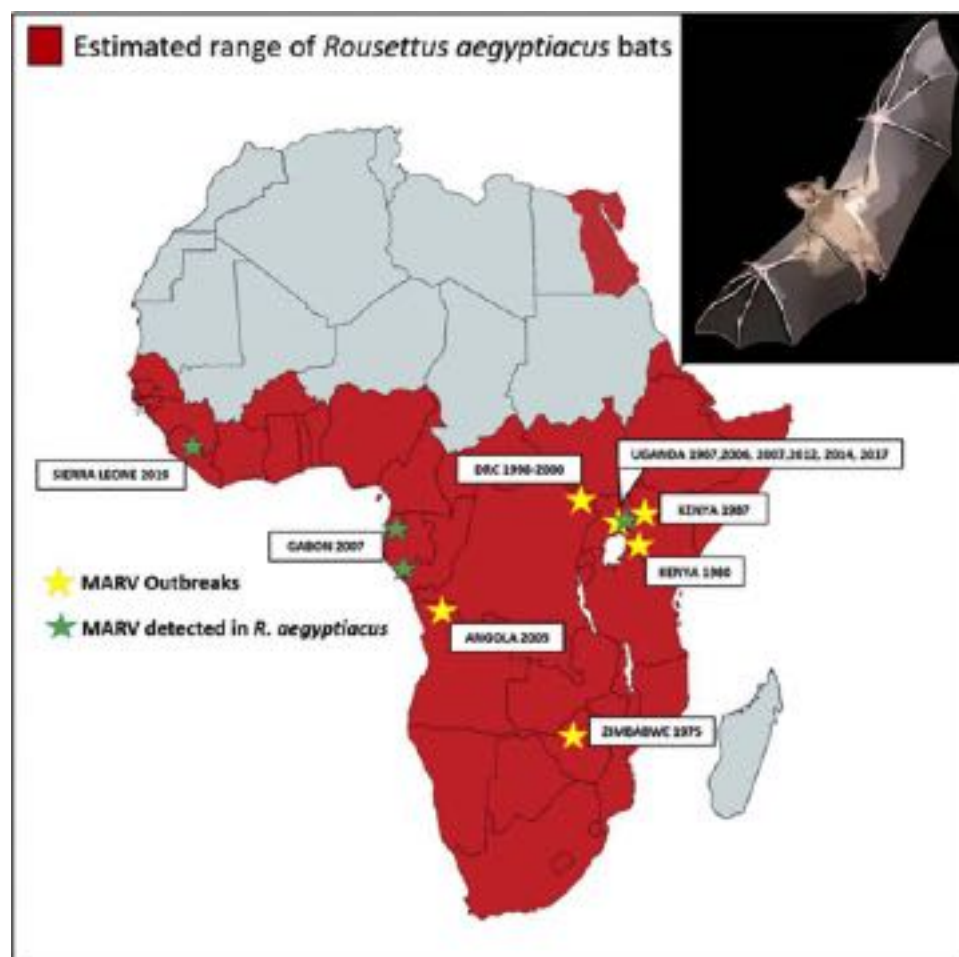


Fig. 1 Distribution of *Rousettus aegyptiacus* bats in Africa. Red areas depict estimated range of *Rousettus aegyptiacus* bats. Yellow stars indicate confirmed MARV outbreaks. Green stars indicate sites where MARV has been detected in wild *Rousettus aegyptiacus* populations. Inset: *Rousettus aegyptiacus* in flight. Photo courtesy of Jonathan Towner (CDC).

Table 1 Outbreaks of Marburg virus disease

Year	Location	Cases/Fatalities (CFR)	MARV/RAVV
1967	Germany (Marburg, Frankfurt), Yugoslavia (now Serbia-Belgrade) ex Uganda (importation of African green monkeys)	31/7 (23%)	MARV
1975	South Africa (Johannesburg) ex Zimbabwe (visited Chinhoyi caves)	3/1 (33.3%)	MARV
1980	Kenya (visited Kitum Cave in Mt. Elgon National Park/Nzoia; nosocomial transmission)	2/1 (50%)	MARV
1987	Kenya (visited Kitum Cave in Mt. Elgon National Park/Mombassa)	1/1 (100%)	RAVV
1988	U.S.S.R. (Koltsovo)	1/1 (100%)	MARV
1990	U.S.S.R. (Koltsovo)	1/0 (0%)	MARV
1998–2000	Democratic Republic of Congo (a continuum of overlapping outbreaks in workers at a gold mine and their families in the Durba-Watsa region)	154/128 (83.1%)	MARV
2004–2005	Angola (Uige province)	252/227 (90.1%)	MARV
2007	Uganda (Kakasi Forest Reserve, Kitaka gold mine in Kamwenge District)	4/1 (25%)	MARV
2008	Netherlands, United States ex Uganda (visited Python Cave in Queen Elizabeth National Park)	2/1 (50%)	MARV
2012	Uganda (Kabale)	15/4 (27%)	MARV
2014	Uganda (Kampala)	1/1 (100%)	MARV
2017	Uganda (Kween District)	3/3 (100%)	MARV

Most outbreaks of filovirus disease are associated with at least some type of exposure to infected wildlife, through the capture, preparation, and/or consumption of bushmeat, or visiting or working in areas inhabited by bats. The strongest evidence for the role of bats in the maintenance of filoviruses in nature is serological and molecular evidence of MARV in sub-Saharan African bats indicating previous or active infection. In 2009, MARV was first isolated from a colony of Egyptian rousette bats (ERBs), *Rousettus aegyptiacus*, cave-dwelling fruitbats ubiquitous to sub-Saharan Africa (Fig. 1). Genetic sequence comparisons of viruses isolated from ERBs in the Kitaka Cave of Uganda closely resembled the viruses responsible for an outbreak among mine workers that had been working in the same cave. Later reports indicating active infection of ERBs in Sierra Leone and high seropositivity among ERBs in South Africa should serve as a warning that MARV is not solely endemic to the Lake Victoria region of Africa. Studies addressing the mechanisms of maintenance of the virus in colonies of *R. aegyptiacus* support a seasonal cycle of enzootic pulses that align with breeding and birthing cycles that create periods of increased risk of human spillover.

Human-to-human transmission has been documented in nearly every outbreak of MARV and is a result, in part, of a failure to abide by universal barrier precautions. The virus is detectable in all bodily fluids and excretions. Virus titers tend to increase as the disease progresses. These features put health care workers at an elevated risk for nosocomial transmission. Animal studies have suggested that as little as 1 virion of MARV is sufficient to start the infection and induce a lethal disease. Further, the virus can persist in immune-privileged sites including the testes, a troubling issue when one considers persons who have survived the infection but continue to shed the virus. Indeed, sexual transmission has been documented in outbreaks of MVD.

Classification

MARV is taxonomically classified within the order *Mononegavirales*, a large group of enveloped, nonsegmented, negative-sense, single-stranded RNA viruses known to cause significant disease in humans. Within this order, MARV belongs to the family of *Filoviridae*, which comprises five genera according to the latest International Committee for Taxonomy of Viruses (ICTV) report (See Relevant Websites Section): *Ebolavirus*, *Cuevavirus*, *Striavirus*, *Thamnovirus*, and *Marburgvirus*.

While there is a large degree of speciation for *Ebolavirus*, only a single species of *Marburgvirus* exists: *Marburg marburgvirus*. Bayesian phylogenetic analyses of full-length genome sequences indicate two distinct evolutionary lineages within this sole species corresponding to MARV and Ravn virus (RAVV). These viruses vary by ~22% at the nucleotide level. MARV is further subdivided into two major clades. One clade consists of Ugandan isolates, including the Poppinga isolate from the 1967 Germany outbreak and bat isolates from Python Cave in Queen Elizabeth National Park (the location in which the Dutch and American tourists likely acquired MVD). Other isolates within this clade include a single specimen from Durba, DRC (1999), the Kenyan Musoke isolate (1980), and all of the available specimens from the 2004–2005 Angola epidemic. The other clade consists of a single 1975 isolate from Zimbabwe (Ozolin isolate), most of the isolates acquired during the 1998–2000 DRC outbreak, and several 2007–2009 Ugandan specimens (a Kitaka miner and Kitaka mine and Python cave bats). Recent isolates from ERBs in South Africa and Zambia, as well as the 2014 and 2017 Ugandan human cases, are related to the same clade. Bayesian coalescent phylogenetic estimates suggest the most recent common ancestor of MARV originated ca. 240 years ago, whereas RAVV emerged 58 years ago. The molecular evolutionary rate for MARV is estimated to be 5.67×10^{-4} nucleotide substitutions/site/year, which is slower than in Ebola and Reston viruses, but more rapid than in the Sudan virus.

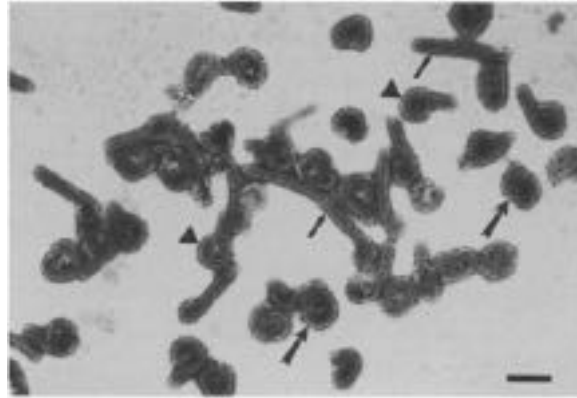


Fig. 2 Negatively contrasted MARV virions recovered from the fluid of infected MA-104 rhesus macaque cells (scale bar = 360 nm). Courtesy of Thomas W. Geisbert.

Virion Structure

The nomenclature for filoviruses was inspired by their distinctive thread-like appearance (latin *filum*, thread). In tissue culture, MARV is highly pleomorphic with filamentous, shepherd's crook, U-shaped, 6-shaped, circular (tori), or branched forms (Fig. 2). Each virion is composed of a tubular host-derived envelope with ~10 nm peplomer protuberances studding its surface and an interior ribonucleoprotein complex, or nucleocapsid (NC), surrounded by a matrix layer. For filamentous and six-shaped particles, a striated density exists between the electron-dense envelope and the NC central core, which is partially visible with other particle forms. The NC has an outer diameter of 45–50 nm that runs the length of the particle and a central axis of 19–25 nm. CryoEM studies demonstrate MARV filamentous particles have a mean diameter of 91 ± 6 nm and a mean length of 892 ± 63 nm, which is shorter than that of Ebola virus (EBOV) particles. These dimensions are slightly larger than prior estimates presumably due to shrinkage from conventional EM thin-section processing, i.e., dehydration and plastic embedding, as well as truncation during sectioning. Particles as long as 14 μ m have been observed; however, the majority of virions measure ~1 μ m and this size corresponds with optimal infectivity in human and monkey tissues. Length discrepancy is attributed to the unique ability of filoviruses to express multiple copies of their genome in a single particle. Empty particles devoid of NC are frequently evident, and are smaller in diameter with varying lengths. The buoyant density of MARV in a potassium tartrate gradient is ~1.14 g/ml.

Morphological heterogeneity has previously hampered efforts to resolve the ultrastructure of MARV, particularly the ribonucleoprotein complex. Cryo-electron tomography combined with sub-tomogram averaging has since provided a well-defined structural arrangement of the NC. The NC is composed of five proteins that collectively package the viral RNA genome: the nucleoprotein (NP), viral protein 35 (VP35), viral protein 30 (VP30), viral protein 24 (VP24), and the polymerase catalytic domain (L). MARV possesses a highly flexible, left-handed, helical NC with a pitch of 7.52 ± 0.19 nm and boomerang-shaped protrusions of variable symmetry with 13.8–15.8 (average 14.96) protrusions per turn. Two lobes of density exist for each NC: an innermost layer corresponding to NP subunits that enwrap the viral RNA and an outer layer composed of a ring of VP35-VP24 heterodimer bridges that independently interact with the C-terminus of NP monomers to form the protrusions. Each turn contains 29.92 copies of NP and each NP monomer packages 6.0 ± 0.2 RNA bases. The NC interacts weakly with the matrix/envelope layers in a pliable, “Velcro-like” fashion. The NC has a pointed and barbed end. Virions bud from filopodia-like cell extensions with the pointed end facing outward.

Genome

Filoviruses have the largest genomes in the *Mononegavirales* order. MARV has a marginally larger genome (~19.1 kb) than EBOV (~18.9 kb). The negative-sense, single-stranded RNA genome consists of seven monocistronic genes that encode seven respective proteins: NP-VP35-VP40-GP-VP30-VP24-L (Fig. 2). Each gene contains a single open reading frame (ORF), 3' and 5' untranslated regions (UTRs) flanking the ORFs, and a highly conserved UAAUU transcription start/stop signal. In contrast to EBOV, MARV expresses a single mRNA molecule per gene. Cotranscriptional editing of the EBOV GP gene results in the synthesis of several GP mRNA species. On the contrary, deep sequencing of *in vitro* and *in vivo* virus populations suggests an increased coding capacity for the Angola variant of MARV. Quasispecies analysis, or the examination of mutant population dynamics that are generated upon replication of RNA viruses, identified hot spots in NP and L mRNAs, as well as U-to-C and A-to-G substitutions in the 3'-untranslated region (UTR) of the NP mRNA. Clustered U-to-C substitutions indicate potential cellular adenosine deaminase (ADAR) activity within the viral genomic RNA. The functions of these editing sites have yet to be determined (Fig. 3).

To enable transcription and replication, the viral RNA-dependent RNA polymerase (L and VP35 proteins form the RdRP) recognizes *cis*-acting regulatory elements located at the 3' leader and 5' trailer ends of the genome. The RdRP uses a start-stop mechanism to transcribe mRNAs sequentially; consequently, it produces more positive-sense mRNA transcripts at the 3' end of the

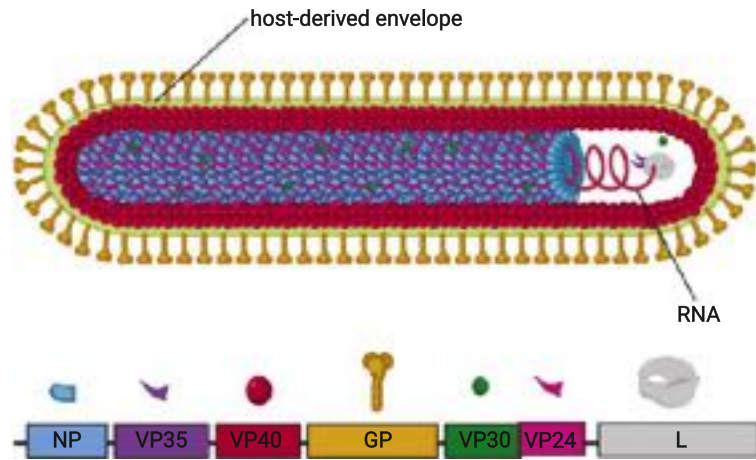


Fig. 3 MARV genome and structural organization. Each gene color corresponds to the encoded viral protein. The gene order is arranged 3' to 5'. Intergenic regions are denoted by a straight line. Note the VP30-VP24 gene junction contains an overlapping segment.

Table 2 Summary of physical and functional attributes of MARV proteins

Protein	Gene (nt)	Protein (aa)	Molecular weight (kDa)	Localization	Function
Nucleoprotein (NP)	2796	695	96	Inner nucleocapsid	Core component of the nucleocapsid, RNA encapsidation, budding, vital for transcription and replication
Viral protein 35 (VP35)	1557	329	32	Outer nucleocapsid	Polymerase cofactor, outer layer nucleocapsid component, antagonism of innate immunity
Viral protein 40 (VP40)	1405	303	38	Envelope matrix	Matrix protein, assembly, budding, host adaptation, IFN antagonism
Glycoprotein (GP)	2846	681	170–200	Envelope spike	Entry, fusion, immune evasion, cytotoxicity
Viral protein 30 (VP30)	1249	281	28	Nucleocapsid	Nucleocapsid formation
Viral protein 24 (VP24)	1287	253	24	Outer nucleocapsid	Outer layer nucleocapsid component, budding, cytoprotective responses, host adaptation, modulation of transcription and replication
Polymerase (L)	7745	2331	~ 200	Nucleocapsid	Catalytic component of the RNA-dependent RNA polymerase

Abbreviations: nt, nucleotides; aa, amino acids; kDa, kilodaltons; IFN, interferon. Gene and ORF lengths were derived from the GenBank Accession Number NC_001608 MARV-Musoke reference sequence.

genome than at the 5' end. Short intergenic regions segregate genes, except for an overlapping region at the VP30-VP24 gene junction. The structure of the overlapping segments are exclusive to filoviruses as they do not resemble other members of the *Mononegavirales* order. The filovirus replication promoter has a bipartite structure similar to paramyxoviruses. While MARV and EBOV do not have genome lengths that are a multiple of six, each NP protomer is in contact with six nucleotides in the NP-RNA complex implying they abide by the "rule of six" similar to paramyxoviruses. The first element of the MARV replication promoter is at the 3' end of the genome and is predicted to adopt a secondary stem-loop structure. The second promoter element consists of a (UN₅) x 3 hexamer motif located within the 3' UTR of the NP gene. The 5' non-coding region of the genome is thought to contain the antigenomic replication promoter.

MARV mRNA transcripts are translated into individual proteins that contribute to the structural integrity of the virion or serve a role in transcription, replication, or immune evasion (Table 2).

NP

The primary role of NP is to encapsidate the viral genomes and antigenomes. This function protects the viral RNA from degradation and detection by host pattern recognition receptors (PRRs). Purified NP self-assembles into a loosely coiled helix that resembles the NC structure, confirming its importance in NC formation. A coiled-coil motif in a central region of NP is thought to facilitate this helical arrangement. The abundance of NP determines the RdRP switch from transcription to replication and is governed by phosphorylation levels. NP also participates in budding by recruiting endosomal sorting complexes required for

transport (ESCRT) protein, Tsg101, via a C-terminal late PSAP domain motif. Recruitment of Tsg101 is required for actin-driven transportation of NC to viral inclusion sites in filopodia.

VP35

VP35 along with the large protein (L) catalytic domain forms the RdRP complex. Oligomerization of VP35 is required for the polymerase cofactor activity. VP35 assists in transcription and replication by inhibiting NP oligomerization and RNA binding, allowing the polymerase to access the RNA template. VP35 participates in immune evasion by blocking IFN gene expression and dendritic cell maturation. MARV VP35 is a comparatively less-efficient antagonist of IFN signaling than EBOV VP35. However, RAVV VP35 IFN-suppressing activity is as potent as that of EBOV VP35.

VP40

The matrix protein, VP40, is the most abundant protein in the virion and coats the inner viral membrane. VP40 is responsible for the filamentous morphology of MARV. Transfection of VP40 into cells results in the release of virus-like particles (VLPs) that bud from the cell membrane, emphasizing its vital role in assembly and budding. VP40 recruits the glycoprotein (GP) to budding sites at the plasma membrane and forms a flexible, weak interaction with the nucleocapsid. A PPPY late domain motif in the VP40 N-terminus interacts with proteins of the host ESCRT machinery, namely tumor susceptibility gene 101 (Tsg101) and neural precursor cell expressed developmentally down-regulated 4 E3 ubiquitin ligase (NEDD4), to facilitate particle release. VP40 can also hijack the cellular coat protein complex II (COPII) vesicular transport system to enhance cell egress. BCL2-associated athanogene 3 (BAG3), a host chaperone, was found to inhibit EBOV and MARV VP40-associated budding. VP40 also regulates MARV replication and transcription.

VP40 is a species-specific virulence factor that plays an integral part in pathogenesis. Serial adaptation of MARV and RAVV is required to produce lethal disease in immunocompetent rodents. However, STAT2 knockout Syrian hamsters are susceptible to infection with various wild-type variants, and humanized or IFN- α /beta receptor knockout mice are susceptible to disease with non-adapted filovirus. Mutations in VP40 are required for adaptation to mice and guinea pigs. Many of these amino acid changes correspond with the ability of the virus in these hosts to antagonize the type I IFN response. MARV VP40 interferes with the IFN response by blocking the phosphorylation of Janus kinases to inhibit IFN α - or beta-induced tyrosine phosphorylation of STAT1 and STAT2.

GP

MARV GP is the sole surface protein and is expressed on the virion envelope as homotrimeric spikes. GP performs the following functions: attachment, entry, fusion, and immune evasion. These vital roles make GP an attractive target for vaccines and therapeutics. The precursor GP undergoes various posttranslational modifications in the endoplasmic reticulum (ER) including glycosylation, phosphorylation, and acylation. N-linked and mucin-type O-linked glycans that form the glycan cap and mucin domain can contribute up to ~50% of the apparent molecular mass. Following modification, GP is cleaved by furin or furin-like proteases in the trans-Golgi network into two disulfide-linked subunits, GP1 (160 kDa) and GP2 (38 kDa). The GP1 ectodomain facilitates receptor binding, whereas the transmembrane subunit, GP2, mediates fusion.

The GP of MARV can subvert the immune system via a myriad of mechanisms. GP can counteract its sensitivity to antiviral tetherin, which normally restricts the release of virions. In addition, putative immunosuppressive motifs in GP were shown to induce lymphocyte death and cytokine suppression. Finally, MARV GP can sterically shield cell surface proteins, e.g., major histocompatibility complex class I, Fas, and integrin β 1, to enhance viral replication and spread.

VP30

VP30 associates with the nucleocapsid via NP binding following phosphorylation and is essential for replication. EBOV and MARV VP30 share high structural similarity. While EBOV VP30 is important for viral transcription and replication, the role of MARV VP30 is not fully understood as it is not required for these functions in a minigenome system. Interestingly, VP30 is required for successful recovery of infectious MARV from a full-length cDNA clone. The mechanism is possibly attributed to nucleocapsid maturation, yet VP30 is not essential for NC-like particle formation. *In vitro* RNA interference (RNAi) silencing of VP30 reduces synthesis of viral proteins and particle release.

VP24

VP24 helps form the outer NC layer and is involved in the maturation of nucleocapsids. This protein functions as an interface between nucleocapsids and budding sites. VP24 also regulates transcription and replication. Unlike EBOV VP24, MARV VP24 is

unable to block type I IFN signaling. However, an extended β -sheet of MARV VP24 recruits Kelch-like ECH-associated protein 1 (Keap1), a suppressor of the antioxidant response and regulator of nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) signaling. Presumably, this leads to transcription of cytoprotective genes in cells to prolong infection.

L

The major and catalytic component of the MARV RdRP, L protein, is essential for both transcription and replication. MARV L is hypothesized to carry out RNA synthesis, capping, and polyadenylation of viral mRNAs.

Life Cycle

The MARV life cycle begins when GP binds to lectins (e.g., DC-SIGN, L-SIGN, hMGL ASGP-R, LSECtin) and glycosaminoglycans on the surface of host cells. Other attachment factors include the receptor tyrosine kinase Axl, and T cell Ig mucin (TIM) proteins TIM-1 and TIM-3. The latter group of proteins is thought to interact with phosphatidylserine molecules on the viral envelope. Next, MARV virions enter cells by a macropinocytosis-like mechanism and are compartmentalized in an endosome. The GP1 subunit is cleaved by an endosomal cysteine protease to remove heavily-glycosylated domains, allowing GP1 to bind the entry receptor, Niemann-Pick C1 (NPC1). Other surface proteins may also serve as receptors to enhance the infection. Endosomal acidification causes the MARV GP2 subunit to undergo a pH-dependent conformational change, releasing the fusogenic loop and permitting fusion of the viral and late endosomal membranes. Following nucleocapsid release into the cytosol, the genome is uncoated and mRNAs are sequentially transcribed, co-transcriptionally capped, and polyadenylated by the viral RdRP. Host cell machinery then translates these positive-sense mRNAs into viral proteins. The genome serves as a template to generate positive-sense antigenomes. Antigenomes then serve as templates to produce progeny genomes that are packaged into virions or transcribed to produce more virus-specific mRNAs. The abundance of nucleoprotein determines when the polymerase switches from gene transcription to genome replication. For assembly, nucleocapsids are transported along actin filaments from viral inclusion bodies to budding sites with the aid of viral and cellular proteins, such as VP40 and ESCRT machinery. GP is recruited via a tubulin-dependent process. The lipid envelope is host-derived and acquired from the virion budding off the host cell membrane.

Clinical Features

MVD symptoms in humans are primarily based upon clinical data acquired during outbreaks in Germany, Yugoslavia, DRC, and Angola. There are three stages of the disease: a generalization phase, an early organ phase, and late organ phase or convalescent period. The incubation period ranges from 3 to 21 days with an average of 5–10 days, depending on the route of transmission and infectious dose. The generalization phase (day(s) 1–4) is characterized by fever, severe headache, malaise, myalgia, pharyngitis, and gastrointestinal symptoms (nausea, abdominal pain, diarrhea, anorexia). These symptoms may persist for the entire course of the infection. By day 4–5, patients develop a maculopapular rash, the earliest distinguishing feature indicating infection by a filovirus. Other common symptoms include thrombocytopenia, leukopenia, and lymphadenopathy. At days 5–13 (early organ phase), multiple organs are affected including the liver, kidneys, and the pancreas. Most patients show hemorrhagic manifestations at this point, such as petechiae, ecchymoses, bloody diarrhea, and mucosal bleeding. Infected individuals may appear irritable, aggressive, and confused. Patients may also experience dyspnea, increased vascular permeability, edema, or conjunctival infection. From day 13 on, patients enter the late organ phase characterized by preagonal symptoms such as severe disseminated intravascular coagulation (DIC), multiorgan failure, coma, shock, and eventually death. Alternatively, survivors enter an extensive convalescent period characterized by partial amnesia, sweating, myalgia, exhaustion, sweating, and peeling of skin at rash sites.

Pathogenesis

Animal models have advanced the understanding of virus infection and dissemination. Filovirus pathogenesis studies have largely been conducted in rodents and NHPs. The latter animal model most accurately recapitulates human infection, as rodents do not typically exhibit the immunological or hemorrhagic manifestations of MVD. MARV particles enter the body through compromised skin or mucosal membranes and infect monocytes, macrophages and dendritic cells (DCs). These early target cells then migrate to regional lymph nodes and spread through the lymphatic system to major organs. The liver and spleen are preferred sites of replication and contain high numbers of monocytes and macrophages. In the late stages of the disease, nearly every organ is affected. Monocytes and macrophages appear highly activated and secrete reactive oxygen species and proinflammatory cytokines/chemokines, such as IL-1- β , IL-6, IL-8, TNF- α , and MCP-1. These soluble factors recruit other inflammatory cells and increase vascular permeability. The secretion of tissue factor by macrophages contributes to coagulopathy and disseminated

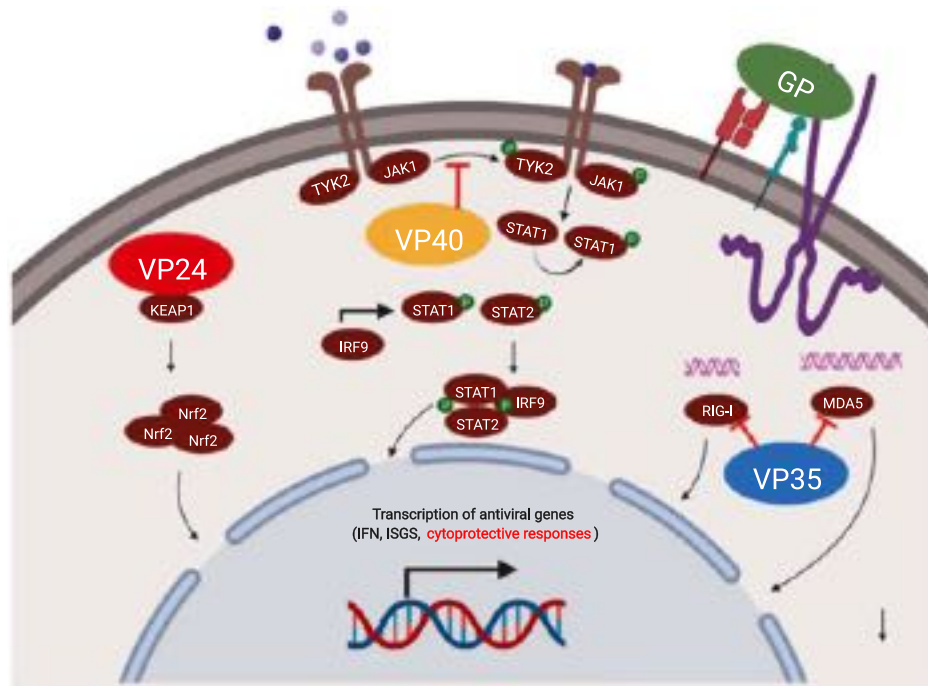


Fig. 4 Virulence factors for MARV. Binding of cytokines to the cell surface receptor activates the JAK/STAT pathway. This interaction increases the kinase activity of JAKs, promoting tyrosine phosphorylation of the receptor complex and mobilization of STAT proteins and association with IRF9. STAT proteins then form hetero- or homodimers and translocate into the nucleus to induce transcription of antiviral genes. MARV VP40 blocks STAT signaling by preventing the phosphorylation activity of JAKs. In the cytoplasm, RIG-I and MDA-5 recognize short and long viral dsRNA, respectively. A complex of proteins then signals recruitment of interferon regulatory factor (IRF) proteins that migrate to the nucleus to stimulate the production of ISGs and IFN-beta. MARV VP35 prevents immune signaling by interfering with these pathways. VP24 binds cellular Keap1, leading to Nrf2 accumulation and persistent upregulation of cytoprotective responses. GP sterically shields epitopes and host surface proteins of cells, e.g., MHC-I and integrins, to induce immunopathology. MHC-I is essential for presenting epitopes to killer T cells and inducing cellular immunity. GP can also antagonize the IFN-inducible protein tetherin. Tetherin traps virus particles at the cell surface to prevent them from being released and infecting other cells. Abbreviations: JAK, Janus kinase; STAT, signal transducer and activator of transcription; RIG-I, retinoic acid-inducible gene I protein; MDA-5, melanoma differentiation-associated protein 5; IFNs, interferons; ISGs, interferon-stimulated genes.

intravascular coagulation (DIC). Consequences of DIC include widespread deposition of fibrin resulting in ischemia, hemolytic anemia, as well as hemorrhagic diathesis due to consumption of platelets and clotting factors. In culture, infected human primary macrophages weakly upregulate T cell costimulatory molecules CD40 and CD80, fail to upregulate CD83, and only express low levels of CD86 and HLA class I and II molecules. In contrast to monocytes and macrophages, MARV-infected dendritic cells (DCs) upregulate co-inhibitory molecules and fail to undergo maturation or elicit cytokine production. Costimulatory molecules of DCs are downregulated, and infected cells fail to induce proliferation of allogenic T cells. The lack of support from DCs is thought to contribute to lymphocyte apoptosis and an impaired adaptive response. Direct interaction with viral proteins or induction of Fas death receptor pathways may additionally advance filovirus-induced lymphocyte apoptosis. Due to extensive necrosis and apoptosis of lymphocytes in secondary lymphatic tissues, an appropriate adaptive immune response is delayed or not induced. Other cell types permissive to MARV infection include endothelial cells, fibroblasts, hepatocytes, and adrenal cortical and medullary cells. In contrast, neutrophils and lymphocytes are spared. Infected endothelial cells are observed in low numbers in the NHP model suggesting that vascular changes are caused by paracrine signaling. Coagulopathies are likely exacerbated by an impaired synthesis of clotting factors in the liver, given the prominent pathology seen in this organ. Additionally, MVD causes necrosis of the adrenal glands leading to decreased steroid synthesis. As steroids help to control blood pressure, MARV infection may eventually lead to hypovolemia. The systemic virus spread and replication, dysregulation of the immune response, coagulopathies, and hypotension ultimately result in shock and multiorgan failure.

MARV proteins broadly dysregulate host antiviral responses to cause pathogenesis (Fig. 4). An IFN-induced cellular protein, tetherin, blocks the release of several viruses including filoviruses from the cell surface at lipid-raft sites. To counter this response, MARV GP antagonizes tetherin restriction to allow egress of virus particles. Furthermore, GP sterically shields epitopes and host surface proteins of cells, e.g., MHC-I and integrins, which contributes to the lack of an adaptive response. Steric shielding of host proteins is more pronounced for MARV Angola GP than for MARV Musoke GP, indicating its potential role in variant-specific pathogenicity. MARV VP24 has an unusual role in that it interacts with host Keap1 to modulate cytoprotective antioxidant pathways and cell survival pathways, presumably to prolong infection. Two viral proteins are considered as the main players in

MARV IFN antagonism: VP40 and VP35. MARV VP40 blocks tyrosine phosphorylation of JAK proteins to disrupt JAK/STAT signaling, whereas MARV VP35 prevents the activation of RIG-I and MDA-5 signaling. In turn, MARV viral proteins downregulate several ISGs. MARV VP40 strongly suppresses ISG54, ISRE, IRF1, and IFN-beta induction in cells stimulated with type I IFNs or by a potent and broad spectrum ISG-inducer, Sendai virus. MARV VP35 also attenuates the expression of these genes, but to a much lesser extent compared to EBOV VP35. This difference is likely attributed to a differential dsRNA binding ability of each virus protein. MARV VP35 coats the backbone of dsRNA to inhibit RLR activation, whereas EBOV VP35 can also bind the blunt ends of dsRNAs (end capping). At a higher multiplicity of infection (MOI) of Sendai virus, MARV VP35 and VP40 were less effective IFN antagonists.

As mentioned, most studies show that MARV infection suppresses innate immunity. MARV-infected immortalized human liver cells (Huh7) secrete low levels of IFN, and treatment of infected cells with polyinosinic-polycytidylic acid (poly I:C) or exogenous IFN does not lead to an expected transcription of classical ISGs. If type I or II IFN is added to cultured cells prior to filovirus infection, viral replication is substantially reduced. Accordingly, supplemental IFN-alpha or beta ameliorates disease in mice and prolongs survival, but does not prevent death in NHPs. These results suggest that at least some ISGs restrict filovirus replication and that MARV has a decreased ability to suppress antiviral gene expression once a host initiates an immune response. Transcriptomic analyses demonstrate that virus infection leads to substantial immune activation including ISGs, upregulation of complement system genes, and genes involved in neutrophil and monocyte recruitment. Thus, activation of innate immunity is dysregulated or only temporarily delayed. Alternatively, only uninfected bystander cells may initiate an antiviral response. Temporal characterization of gene expression in MARV-Angola-infected cynomolgus macaques further supports this concept. Upregulation of common innate response and pro-inflammatory genes in peripheral blood are observed early after infection (days 1–3) and sustained throughout the disease course. Many of these genes function in RLR and IFN signaling, for instance, MX1, RIG-I (DDX58), PARP14, STAT1, IRF3, IRF7, IRF9, IFITs, RNA helicases, and ISG15. However, increases in IFN-alpha and IFN-beta gene expression are not detected, nor are genes involved in other innate defense pathways, such as complement activation or antigen presentation. ISG upregulation does not appear to reduce viral load, or postpone death or clinical signs of the disease in these animals, proving MARV-Angola tolerates a high level of IFN-mediated antiviral activity. It is possible that MARV may weaken or overcome the effects of ISG activation by interfering with alternative signaling pathways or modulating key proteins in these antiviral pathways. For example, it has also been shown that infection of macaques leads to early signaling of T helper 2 (Th2; humoral response)-associated genes, IL-4 and IL-5, and serum levels of immunosuppressive IL-10. No appreciable changes in T helper 1 (Th1; cell-mediated response) cytokines, such as IFN-gamma and TNF-alpha, are noted until late in the disease. Th1 cytokines may contribute to a favorable prognosis since the survivors of a 2012 MARV outbreak in Uganda exhibited a Th1 response. IL-4 and IL-10 suppress Th1 type cytokine, IFN-gamma and IL-2 production, so it is possible that these cytokines exacerbate pathogenesis by skewing immunity towards a detrimental Th2 response. Thus, MARV-mediated immunopathology is multifaceted.

Diagnosis

Early in the course of the disease, MVD cases present as a non-specific, febrile illness which can easily be confused with the symptoms of other infections endemic to central Africa including influenza, malaria, and Dengue fever. This is a major confounder especially in the early phases of an outbreak as there would otherwise be limited suspicion for MARV infection unless certain risk factors (e.g., exposure to bats) were addressed upon presentation to the health center. If MARV infection is suspected, isolation of the patient is paramount to allow for expedient containment and safe testing under proper containment. Antigen-capture enzyme-linked immunosorbent assay (ELISA), polymerase chain reaction (PCR), and IgM-capture ELISA testing have been routinely utilized to confirm MVD. Virus isolation is often considered as the “gold standard” of any diagnosis; however, the biosafety level 4 (BSL4) facilities needed to safely cultivate filoviruses are rarely available. In cases of an ongoing outbreak investigation, the use of IgG-capture ELISA can be useful for testing recovered individuals as well as their contacts for estimation of asymptomatic cases within an outbreak. The rapid progression of disease in MVD coupled to the resource-limited environment of most filovirus outbreaks often complicates and delays the diagnosis. Unfortunately, some patients may succumb prior to seeking health care or before receiving a confirmed diagnosis. In the case of deceased patients, immunohistochemistry on autopsy tissues, virus isolation, or PCR from cadaver samples may be performed.

Treatment

Currently there are no regulatory-agency approved therapeutics or treatments for MVD. Palliative support should be offered under strict barrier conditions to reduce the risk of nosocomial infections. Based on the resources available and abilities of the treatment facility, patients should be provided with fluids and electrolyte replacement to prevent dehydration and maintain proper oxygenation and blood pressure. The risk for secondary infections is high due to an impaired immune system; attempts to treat any complicating infections should be made. Several experimental treatments have been in development over the last few decades and have shown promise in animal models. Commercial interest has impeded further development, although nucleoside analog

phosphorodiamidate morpholino oligomers (PMOs) have entered phase I human clinical trials for safety testing (see Relevant Websites section).

Although rodent studies may yield proof-of-concept efficacy and inform dosing regimens, protective immunity in these models may not translate to the more robust NHP models. As the latter models most reliably recapitulate the pathophysiology observed in humans, only NHP studies are discussed here. Some of the first interventions against MVD involved immunomodulatory treatment with IFN or an anticoagulant (recombinant nematode anticoagulant protein c2 (rNAPc2). These approaches proved unsuccessful in the NHP model. Subsequent medical countermeasure evaluations, such as twice-daily delivery of a nucleoside analog (BCX4430), or daily administration of phosphorodiamidate morpholino oligomers (PMOs), protected 100% of NHPs up to 2 or 4 days, respectively. However, the majority of animals exhibited clinical signs and had detectable viremia with these antivirals. There is also evidence for the usefulness of antibody therapy against MVD. Three doses of convalescent purified IgG from the serum of immunized NHP survivors fully protected rhesus macaques from a lethal MARV-Ci67 challenge and prevented 67% from disease when given up to 2 days post-infection.

Presently, the most encouraging therapeutic candidates against the most pathogenic Angola variant of MARV (associated with 90% case fatality rate) are stable nucleic acid-lipid particles (SNALPs) and MR191-N monoclonal antibodies. These treatments can rescue NHPs from MVD and lethality after viremia and clinical signs have already developed. SNALPs employ short-interfering RNA (siRNA) technology, which relies on host RNA interference (RNAi) machinery to target and degrade viral transcripts. Multiple doses of SNALPs directed against MARV NP RNA protected rhesus monkeys from a highly lethal 1000 PFU Angola challenge whenever treatment was delayed up to 5 days. Remarkably, treatment with MR191-N antibodies has a similar therapeutic window, but requires only two doses. Intravenous (i.v.) administration of MR191-N protects 100% of NHPs from advanced stages of disease when treated on days 4 and 7, and 80% of macaques when treated on days 5 and 8. MR191-N was isolated from a human survivor and is a neutralizing antibody that blocks the receptor binding site.

The use of live attenuated vaccines as post-exposure prophylaxis has also shown limited therapeutic potential in NHP models, although the treatment window is shorter. This possibility might be of interest as some MVD vaccine candidates utilize a similar vaccine platform as the recombinant Vesicular stomatitis virus (rVSV)-based Ebola vaccine, which was deployed and highly effective at preventing disease during the 2013–2016 West African and 2018–2020 EBOV outbreaks. An rVSV construct expressing a Musoke variant GP was 100% effective when administered at 20–30 min after a 1000 PFU challenge of MARV-Musoke, ~83% effective at 24 h, and 33% effective at 48 h. For high and low-dose challenges with the most pathogenic Angola variant, a total of 25% and 80–89% of treated macaques survived, respectively (Table 3).

Table 3 Summary of post-exposure treatment studies for Marburg virus disease in the non-human primate model

Treatment	Time post-exposure	Treatment doses	MARV challenge	Survival	Illness	Viremia
IFN-beta	1 h	15	~1000 PFU Musoke i.m.	33%	100%	100%
rNAPc2	10 min	15	~1000 PFU Angola i.m.	17%	100%	100%
BCX4430	1–48 h	26–30	1275 PFU Musoke s.c.	83%–100%	0%–33%?	100%
PMOplus (pool)	30–60 min	14	~1000 PFU Musoke s.c.	100%	100%	100%
PMOplus (NP)	1 h – 4 d		i.m.	83%–100%	100%	83%–100%
Convalescent IgG (purified)	15–30 min 48 h	3	~1000 PFU Ci67 i.m.	100%	0%–33%	0%–33%
siRNA SNALPs	30–45 min	7	~1000 PFU Angola i.m.	100%	50%	33%
	24 h			100%	25%	0%
	48 h			100%	50%	0%
	3 d			100%	50%	17%
	4 d			100%	100%	75%
	5 d			50%	100%	100%
	3 d	7	~1000 PFU RAVV i.m.	100%	25%	25%
	6 d			100%	100%	100%
MR191-N mAbs	4–5 d	2	~1000 PFU Angola i.m.	80%–100%	80%–100%	100%
	5 d	2	~1000 PFU RAVV i.m.	100%	60%	100%
rVSV	20–30 min	1	~1000 PFU Musoke i.m.	100%	60%	0%
	24 h			80%	17%	0%
	48 h			33%	67%	80%
	20–30 min	1	~1000 PFU Angola i.m.	25%	100%	100%
	20–30 min	1	~50 PFU Angola i.m.	80%–89%	22%–40%	33%–60%

Abbreviations: IFN, interferon; rNAPc2, recombinant nematode anticoagulant protein c2; PMO, phosphorodiamidate morpholino oligomers; NP, Marburg virus nucleoprotein; IgG, immunoglobulin G; siRNA, short interfering RNAs; SNALPs, stable nucleic acid lipid particles; mAbs, monoclonal antibodies; rVSV, recombinant Vesicular stomatitis virus; RM, rhesus macaques; CM, cynomolgus macaques; PFU, plaque-forming units; MARV, Marburg virus; RAVV, Ravn virus; i.m., intramuscular; s.c., subcutaneous. Illness is defined as an animal having fever and/or showing significant clinical signs of a disease.

Prevention

Similar to other viral hemorrhagic fever infections, secondary, or human-to-human, transmission is a common feature of MVD in humans. Therefore, suspected and confirmed patients should only be addressed under strict barrier techniques utilizing, at minimum, a universal precautions approach. Furthermore, efforts to ensure targeted public health outreach and education of affected and surrounding communities should be undertaken to ensure proper awareness of immediate risks associated with MVD. Cases of laboratory-acquired infections were noted in the former Soviet Union (Table 1) underscoring the need for stringent high containment practices to be in place when manipulating the virus in the laboratory setting. Accordingly, only basic diagnostic efforts limited to molecular or antigen based detection of inactivated samples should be attempted at containment levels below BSL-4.

While several promising vaccine candidates are currently in development, there are no vaccines or prophylactic measures approved by regulatory agencies. An overview of prospective vaccines for MVD is shown in Table 4.

Originally, scientists explored formalin-inactivated whole MARV particles as a vaccine. This approach failed to induce a protective immune response in rhesus macaques. Thereafter, Venezuelan equine encephalitis virus replicons expressing the Musoke variant GP, NP, or a combination of both antigens, were evaluated. Replicon-based vaccines protected 67–100% of cynomolgus monkeys against a high dose (8000 PFU) subcutaneous (s.c.) Musoke challenge, but did not protect NHPs against RAVV. Another vaccine expressing MARV VLPs fully protected cynomolgus macaques and was cross-protective against Musoke and Ci67 variants, as well as against RAVV. It would be of interest to see whether VLPs are also effective against the most pathogenic variant, the Angola strain, as this was not reported.

The safety profile of DNA subunit vaccines makes them an attractive option. However, DNA vaccines afforded only partial protection of cynomolgus macaques against MVD using MARV GP as a primary immunogen. Specifically, three doses of vaccine resulted in 67% and 100% survival of monkeys against a 1000 PFU i.m. Musoke and Angola challenge, respectively. All immunized animals became sick when challenged with the Musoke variant and approximately half of them when challenged with the Angola variant, indicating lack of sterile immunity for this vaccine platform.

The most promising vaccines against MARV-Angola use a live attenuated vaccine platform. These include recombinant Adenovirus 5 (rAd5) and VSV (rVSV) vectors, which both employ MARV GP as an immunogen. Both vaccines are safe, immunogenic, and require a single injection to elicit complete protection. Moreover, these vectors provide cross-protection against multiple MARV variants. Added benefits of the rVSV-based vaccine are its proven defense against aerosol infection and post-exposure treatment potential. Vaccination with rAd5 or rVSV results in robust antibody production and cellular responses that are thought to contribute to protection. Safety is a significant issue for any replication-competent vaccine, particularly in the immunocompromised, a major concern in an HIV-endemic region. Nevertheless, immunization with a rVSV expressing EBOV GP resulted in a merely transient viremia in simian/human immunodeficiency virus-infected macaques, and NHPs intrathalamically inoculated with MARV GP- or EBOV GP-expressing rVSVs lacked neurovirulence. Another concern for live attenuated vaccines was that pre-existing immunity against the vector would influence protective efficacy. However, this concern is not

Table 4 Summary of vaccine studies for Marburg virus disease in the non-human primate model

<i>Vaccine</i>	<i>Immunogen</i>	<i>Vaccine doses</i>	<i>MARV challenge</i>	<i>Survival</i>	<i>Illness</i>
Inactivated MARV	Irradiated whole virus	1?	200 LD ₅₀ Popp parenteral	50%	50%
VEEV replicon	GP (Musoke)	3	8000 PFU Musoke/RAVV s.c.	100%	0%?
	NP (Musoke)	3	8000 PFU Musoke s.c.	67%	100%
	GP + NP (Musoke)	3	8000 PFU Musoke s.c.	100%	0%
	GP + NP (Musoke)	3	RAVV	0%	?
	GP + NP + VP40 (Musoke) QS-21 Adjuvant	3	1000 PFU Musoke/Ci67/RAVV s.c.	100%	0–33%
VLPs	GP (Musoke)	3	1000 PFU Musoke s.c.	67%	100%
	GP (Angola)	4	1000 PFU Angola i.m.	100%	50%
	GP (Angola)	3 + 1	1000 PFU Angola i.m.	100%	25%
DNA plasmid	GP (Angola)	1	1000 PFU Angola i.m.	100%	0%
	Blend GP (Z + S + Ci67 + RAVV)	2	1000 PFU Musoke/Angola s.c.	100%	0%
	+ Z NP				
DNA plasmid prime + rAd5 boost	GP (Musoke)	1	1000 PFU Musoke/RAVV/Angola i.m.	100%	0%
	GP (Musoke)	1	1000 PFU Musoke aerosol	100%	0%
rAd5					
rVSV	GP (Musoke)	1	1000 PFU Musoke/RAVV/Angola i.m.	100%	0%
	GP (Musoke)	1	1000 PFU Musoke aerosol	100%	0%

Abbreviations: MARV, Marburg virus; VEEV, Venezuelan equine encephalitis virus; VLPs, virus-like particles; rAd5, recombinant Adenovirus 5; rVSV, recombinant Vesicular stomatitis virus; RM, rhesus macaques; CM, cynomolgus macaques; GP, glycoprotein; NP, nucleoprotein; VP40, viral protein 40; Z, Zaire ebolavirus; S, Sudan ebolavirus; RAVV, Ravn virus; PFU, particle-forming units; s.c., subcutaneous; i.m., intramuscular. Illness is defined as an animal having fever, viremia, and/or exhibiting significant clinical signs of disease. A question mark (?) indicates the data was unclear or not provided in the literature.

relevant since previous vaccination with a Lassa virus GP-expressing vaccine did not abrogate immunity when NHPs were sequentially immunized with an EBOV GP-expressing rVSV and subsequently challenged with EBOV. The rVSV vaccine is also durable, with protection lasting as long as 14 months after immunization. While the scarcity of MVD outbreaks and lack of commercial interest have hindered further vaccine development, vaccines in phase I human clinical trials include a DNA plasmid vaccine (VRC-MARDNA025–00-VP) and the multivalent Modified Vaccinia Ankara - Bavarian Nordic MVA-BN(R)-Filo vaccine (see Relevant Websites section).

Although MARV outbreaks are generally rare, those that do surface are generally associated with high morbidity and mortality rates (**Table 1**). Preventative measures to limit exposure should be taken with extreme stringency. Given the clear association of MARV with fruit bats, avoidance of known bat roosts or wildlife that commonly comes in contact with these species or their excreta (i.e., non-human primates) is advised unless proper personal protective procedures are carried out.

Further Reading

- Albariño, C.G., Wiggleton Guerrero, L., Spengler, J.R., *et al.*, 2015. Recombinant Marburg viruses containing mutations in the IID region of VP35 prevent inhibition of Host immune responses. *Virology* 476, 85–91.
- Bharat, T.A., Riches, J.D., Kolesnikova, L., *et al.*, 2011. Cryo-electron tomography of Marburg virus particles and their morphogenesis within infected cells. *PLoS Biology* 9, e1001196.
- Cross, R.W., Mire, C.E., Feldmann, H., Geisbert, T.W., 2018. Post-exposure treatments for Ebola and Marburg virus infections. *Nature Reviews Drug Discovery* 17, 413–434.
- Feldmann, H., Sanchez, A., Geisbert, T.W., 2013. *Filoviridae: Marburg and Ebola Viruses*. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA, USA: Lippincott Williams & Wilkins.
- Kortopeter, M.G., Dierberg, K., Shenoy, E.S., Cieslak, T.J., Medical Countermeasures Working Group of the National Ebola Training and Education Center's (NETEC) Special Pathogens Research Network (SPRN), 2020. Marburg virus disease: A summary for clinicians. *International Journal of Infectious Diseases* 99, 233–242.
- Mehedi, M., Groseth, A., Feldmann, H., Ebihara, H., 2011. Clinical aspects of Marburg hemorrhagic fever. *Future Virology* 6, 1091–1106.
- Messaoudi, I., Amarasinghe, G.K., Basler, C.F., 2015. Filovirus pathogenesis and immune evasion: Insights from Ebola virus and Marburg virus. *Nature Reviews Microbiology* 13, 663–676.
- Olejnik, J., Mühlberger, E., Hume, A.J., 2019. Recent advances in marburgvirus research. *F1000Research* 8.
- Pattanaik, A.K., Whitt, M.A., 2015. *Biology and pathogenesis of rhabdo- and filoviruses*. New Jersey: World Scientific.
- Towner, J.S., Khristova, M.L., Sealy, T.K., *et al.*, 2006b. Marburgvirus genomics and association with a large hemorrhagic fever outbreak in Angola. *Journal of Virology* 80, 6497–6516.

Relevant Websites

- [https://clinicaltrials.gov/ct2/show/NCT02891980?cond=Marburg + Virus + Disease&draw=2&rank=6](https://clinicaltrials.gov/ct2/show/NCT02891980?cond=Marburg+Virus+Disease&draw=2&rank=6)
A Safety Trial to Test MVA-BN(R)-Filo and Ad26.ZEBOV Vaccines in Healthy Volunteers.
- www.ictv.global/report/filoviridae
Filoviridae.
International Committee on Taxonomy of Viruses (ICTV).
- <https://vhfimmunotherapy.org/>
Viral Hemorrhagic Fever Immunotherapeutic Consortium.

Measles Virus (*Paramyxoviridae*)

Roberto Cattaneo, Mayo Clinic, Rochester, MN, United States

Michael McChesney, University of California, Davis, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of R. Cattaneo, M. McChesney, Measles Virus, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:[10.1016/B978-0-12-801238-3.02619-2](https://doi.org/10.1016/B978-0-12-801238-3.02619-2).

Glossary

Oncolytic virotherapy The treatment of cancer patients based on the administration of replication-competent viruses that selectively destroy tumors but leave healthy tissue unaffected.

Subacute sclerosing panencephalitis (SSPE) A rare but always lethal brain disease that can occur several years after acute measles.

Syncytia Fused cells with multiple nuclei characteristics of measles virus infection.

Classification

Measles virus (MeV) is an enveloped nonsegmented negative-strand RNA virus of the order *Mononegavirales*, family *Paramyxoviridae*, genus *Morbillivirus*. Measles is the only medically relevant *Morbillivirus*; *Paramyxoviridae* of other genera include prevalent mumps and parainfluenza viruses, and emerging bio-threats to humans and economically important livestock like Nipah and Hendra viruses. Among negative-strand RNA viruses, the *Paramyxoviridae* are defined by a membrane fusion apparatus that can trigger fusion of viral and cell membranes at neutral pH.

The genus *Morbillivirus* consists of closely related, but host-specific viruses that cause age-old diseases of humans and animals like measles, rinderpest, and canine distemper. In addition, newly recognized morbilliviruses of aquatic mammals cause emerging diseases. The defining characteristics of the *Morbilliviruses* is that the attachment protein hemagglutinin (H) does not have neuraminidase activity, but interacts with two different receptors: first, the primary receptor signaling lymphocyte activation molecule (SLAM, CD150), and then the adherens junction protein, nectin-4. Morbillivirus infection of SLAM-expressing immune cells explains immunosuppression, and infection of upper airways epithelial cells expressing nectin-4 accounts for extremely efficient transmission of the virus.

Virion Structure

MeV particles are enveloped by a lipid bilayer derived from the plasma membrane of the cell in which the virus was grown. **Fig. 1** illustrates a MeV particle (top), which components are color-coded as and mapped on the viral genome (bottom). The particles of MeV and the other members of the *Paramyxoviridae* have been visualized as pleomorphic or spherical, depending on the methods used for their purification. Their diameter ranges from 120 to 300–1000 nm, implying that their cargo volume may differ by a factor 30. Indeed, large particles can contain multiple encapsidated genomes, as deduced initially from sedimentation and ultraviolet inactivation studies. Three RNA genomes are drawn schematically in the MeV particle of **Fig. 1**.

In cultivated cells, MeV particles accumulate below the plasma membrane, and the ratio of intracellular to secreted infectious virus is approximately 10:1. Therefore ultrastructural studies have been performed mainly in virus-infected cells. **Fig. 2** visualizes by cryo-electron tomography a MeV particle being assembled in infected HeLa cells. Glycoprotein spikes (H and F proteins) that extend approximately 8–12 nm from the surface of the membrane are inserted into the viral envelope. The two viral glycoproteins form the membrane fusion apparatus: the H protein contacts the cellular receptors, whereas F fuses the viral and cellular membranes. The F protein spike is trimeric, whereas the H protein forms covalently linked dimers that form noncovalently linked tetramers. In the assembly process, the viral glycoproteins are preferentially incorporated into nascent viral particles, whereas the majority of host proteins are excluded.

The matrix (M) protein is observed as an organized two-dimensional para-crystalline array associated with the inner leaflet of the plasma membrane. It is the assembly organizer, and bridges the envelope with the ribonucleocapsids (RNP). The helical RNP is the replicative complex: the RNA genome is tightly encapsidated by a helically arranged nucleocapsid (N) protein. Two other proteins, a polymerase (L for large) and a polymerase cofactor (P for phosphoprotein), are associated with the RNP.

In monolayers of most immortalized cell lines, MeV infections spread mainly through fusion of an infected cell with recipient cells expressing an appropriate receptor, rather than through particle budding. Receptors on recipient cells trigger the viral membrane fusion apparatus expressed on the surface of infected cells to form fusion pores; pore expansion then results in coalescence of plasma membranes, and formation of large multinucleated syncytia. This mechanism promotes the spread of progressively larger genome populations, without selection for individual genomes. From this vantage point, the MeV infectious “unit” appears to be highly complex and diverse. Indeed, genetic analyses of MeV infections have provided independent evidence of multi-genome MeV transmission.

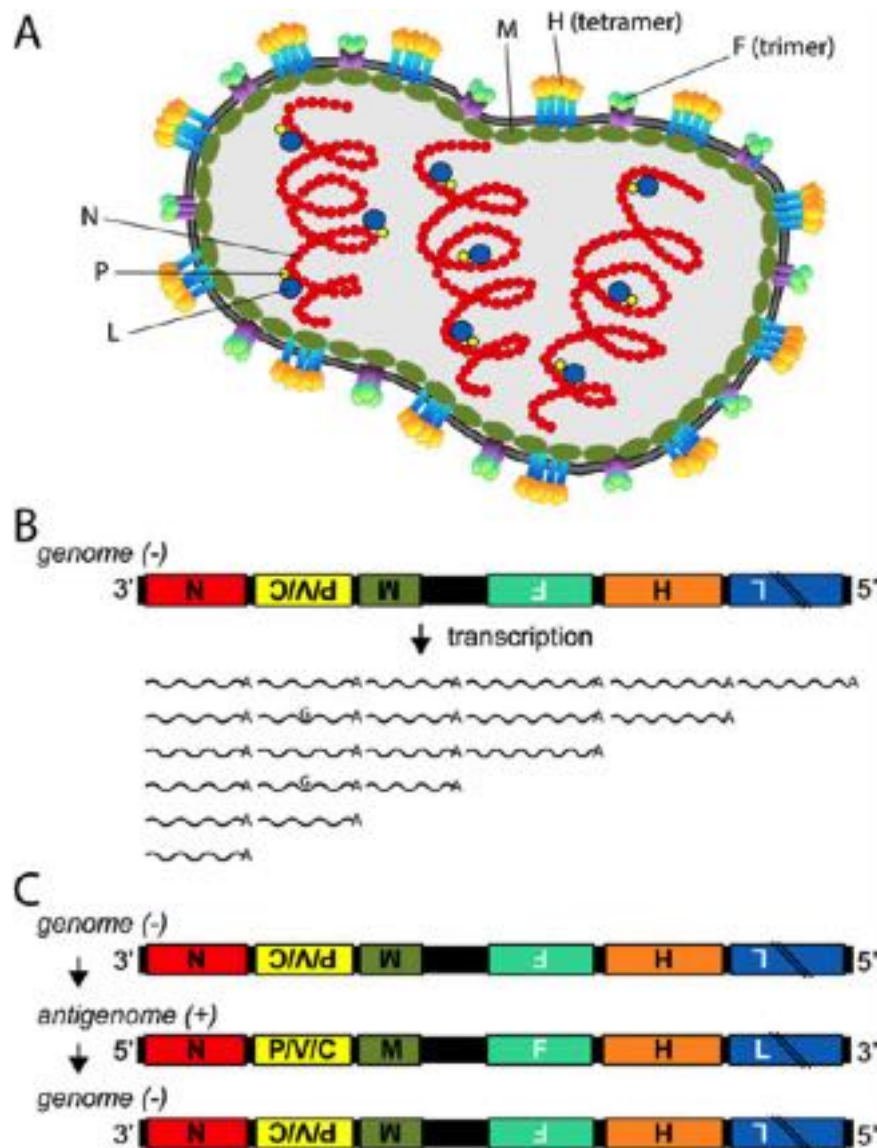


Fig. 1 Diagram of a MeV particle, and of MeV transcription and replication. (A) The particle is drawn with its six main components: the nucleocapsid (N) that covers the genomic RNA and interacts with the phosphoprotein (P), and polymerase (large, L), forming the ribonucleocapsid complex; the fusion (F) and hemagglutinin (H) proteins forming the membrane fusion apparatus; and the matrix (M) protein controlling particle assembly as well as transcription and membrane fusion. Particles can contain multiple encapsidated genomes, of which three are drawn schematically in the particle. (B) The encapsidated negative strand RNA genome, which serves as template for transcription. The coding regions of the proteins are color-coded, non-coding regions are black. Transcription starts with a short, uncapped leader RNA (not shown) from the 3' end of the genomic RNA; this is followed by the transcription of 5' capped and polyadenylated mRNAs (wavy lines), which encode the viral proteins. The polymerase complex stops at the end of each transcript, polyadenylates (A), ignores the 3-nucleotide intergenic region, and restarts transcription. However, re-initiation of transcription is not always successful, resulting a gradient in the direction of 3' to 5'. In the middle of the P mRNA, the polymerase complex adds a single G residue (G) to about half of the transcripts, allowing access to the V reading frame. The C reading frame is accessed by an alternative ribosome initiation codon. (C) During replication, the polymerase complex ignores the transcription stop/start signals, rendering a full-length antigenomic RNA, which is also encapsidated. The synthesized antigenome serves then as template for the synthesis of additional copies of genomic RNA. Reproduced from Cattaneo, R., Donohue, R.C., Generous, A.R., Navaratnarajah, C.K., Pfaller, C.K., 2019. Stronger together: Multi-genome transmission of measles virus. *Virus Research* 265, 74–79.

Genome

The 15,894 nucleotides (nt) MeV negative-strand RNA genome begins with a 56 nt 3' region, called leader, and ends with a 40 nt region, called trailer. These control regions are essential for transcription and replication and flank the six genes. The term gene refers here to contiguous transcription units separated by three untranscribed nucleotides. There are six genes coding for eight

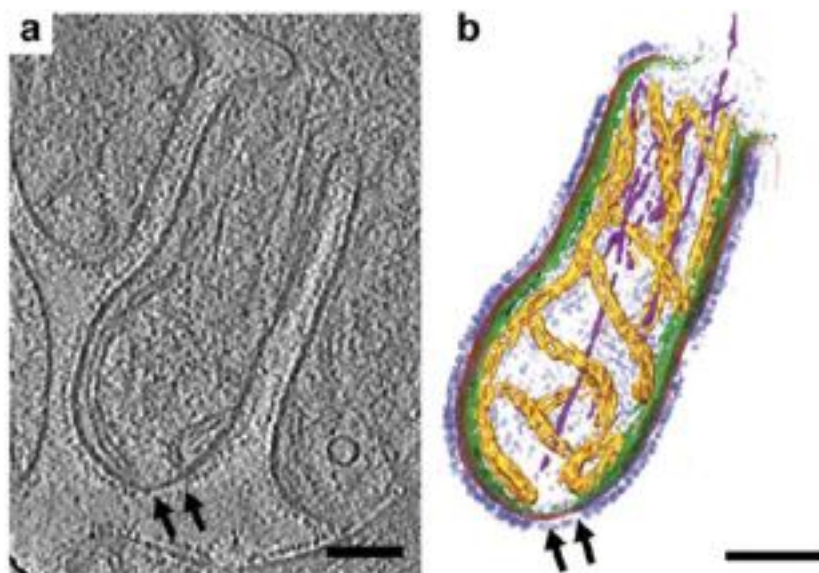


Fig. 2 Cryo-electron tomography of a MeV assembly site in HeLa cells. (a) Tomographic slice (1.18 nm thick) of the assembly site. (b) Three-dimensional reconstruction of the data. Color coding: glycoproteins (purple), viral membrane (red), M (green), RNP (gold), and actin filaments (magenta). Black arrows indicate areas where M is absent on the viral membrane. Scale bars are 100 nm. Reproduced from Ke, Z., Strauss, J.D., Hampton, C.M., *et al.*, 2018. Promotion of virus assembly and organization by the measles virus matrix protein. *Nature Communications* 9, 1736.

proteins, in the order (positive strand): 5'-N-P/V/C-M-F-H-L-3' (Fig. 1, bottom). The P gene uses overlapping reading frames to code for three proteins, P, V, and C. Complete sequences of several MeV wild type and vaccine strains have been published in GenBank.

Replicative Complex

The first gene codes for the N protein. Each N protein interacts with and covers 6 nt, and about 13 N proteins constitute a turn in the RNP helix (Fig. 3). RNPs are formed when N is expressed in the absence of other viral components, suggesting that N–N interactions drive RNP assembly. Two domains have been identified in the N protein: a conserved amino-terminal N_{CORE} (about 400 amino acids) and a variable carboxyl-terminal N_{TAIL} (about 100 amino acids). N core is essential for self-assembly, RNA binding, and replication activity, whereas N tail interacts with a carboxyl-terminal domain of the P protein (Fig. 3, upper right). N protein exists in at least two forms in infected cells: one associated with RNA in an RNP structure, and a second unassembled soluble form called N0 that may encapsidate the nascent RNA strand during genome and antigenome replication.

The second gene codes for three proteins implicated in transcription or innate immunity control: P, V, and C proteins. Two of these proteins have a modular structure: the 231 amino-terminal residues of V are identical with those of P, but its 68 carboxyl-terminal residues are translated from a reading frame accessed by insertion of a pseudo-templated G residue through polymerase stuttering. This V domain is highly conserved in *Paramyxoviridae*, with cysteine and histidine residues binding two zinc atoms per protein. The main function of the V protein is to counteract the innate immune response: V interacts with both STAT2 and MDA5, thereby antagonizing both type I interferon signaling and interferon production.

The main function of the P protein (Fig. 3, red) is to support viral replication and transcription; it is an essential component of the polymerase, and of the protein complex mediating RNA encapsidation by N0 protein. The P protein amino-terminal segment, identical with the V protein amino-terminus, is phosphorylated on serine and threonine residues and it contains regions of high intrinsic disorder, possibly facilitating interactions with multiple proteins. P self-assembles as a tetramer through a central region in its unique domain, and interacts with N_{TAIL} through its carboxyl-terminus.

The third protein expressed from the P gene is called C. Its reading frame is accessed by ribosomes initiating translation 22 bases downstream of the P AUG initiation codon. The MeV C protein not only inhibits the interferon response but also enhances infectivity.

The third, fourth, and fifth genes code for envelope associated proteins that are discussed below in the context of receptor recognition and membrane fusion. The sixth and last gene codes for the RNA-dependent RNA polymerase (large protein, L), that possesses all enzymatic activities necessary to synthesize mRNA: nucleotide polymerization, capping and methylation, and polyadenylation. L adds poly-A tails to nascent viral mRNAs co-transcriptionally, by stuttering on a stretch of U residues occurring at the end of each viral gene. Sequence comparison identified six highly conserved domains in the L protein that were tentatively

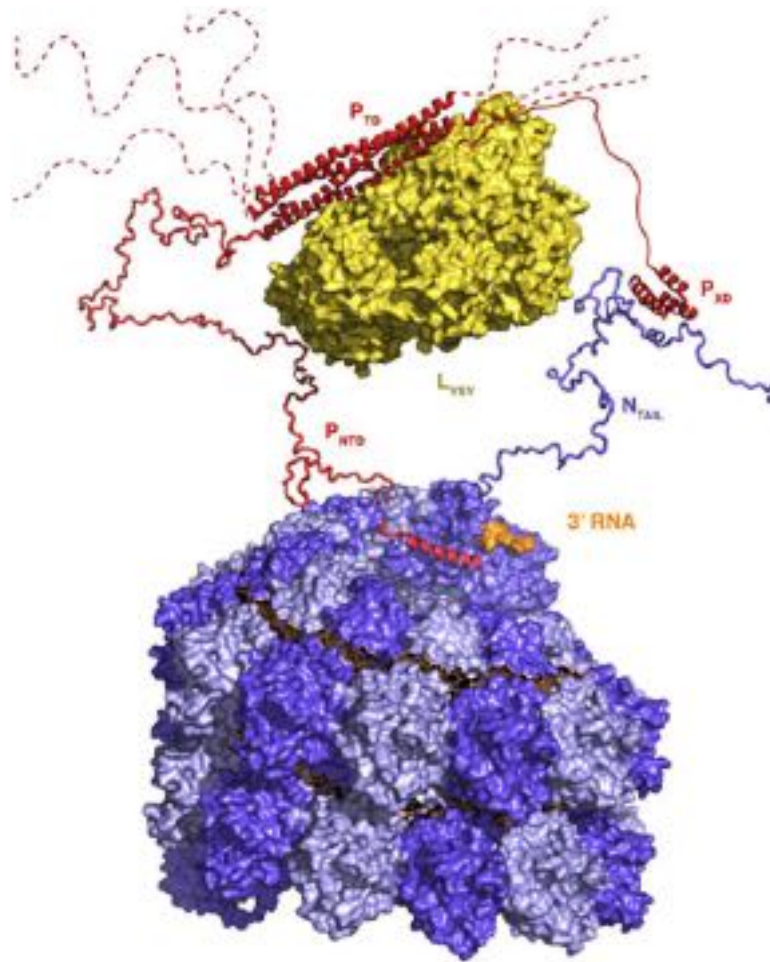


Fig. 3 Proteins involved in MeV replication, and possible positioning of the N, P and L proteins. The solvent-accessible 3' end of the RNA genome (orange surface) is positioned near the amino-terminus of the first N protomer (adjacent N protomers are shown in different shades of blue). Only N_{TAIL} of the terminal N protomer is shown for simplicity. The protomer can bind P via two possible mechanisms, via an interaction between the N_{TAIL} molecular recognition element and the XD domain of P, and via the N-terminal peptide of P (P_{NTD}). Reproduced with permission from Desfosses, A., Milles, S., Jensen, M.R., *et al.*, 2019. Assembly and cryo-EM structures of RNA-specific measles virus nucleocapsids provide mechanistic insight into paramyxoviral replication. *Proceedings of the National Academy of Sciences of the United States of America* 116, 4256–4264.

assigned different catalytic functions. **Fig. 3** uses the polymerase of vesicular stomatitis virus (L_{VSV}, yellow) to represent the MeV L protein, for which no atomic resolution structure is available.

Transcription and Replication

After cell entry, the polymerase transcribes the viral genome with a sequential “stop–start” mechanism. The polymerase accesses the genome through an entry site located near its 3' end, transcribes the first gene (N) with high processivity, polyadenylates the N mRNA, and reinitiates P mRNA synthesis. The frequency of re-initiation is less than 100%, resulting in a gradient of transcript levels; N is transcribed at the highest levels, the most promoter-distal L gene at the lowest.

The gene order and transcription strategy are fundamental characteristics that MeV shares with all other *Paramyxoviridae*. The negative-strand genome is then used to synthesize positive-strand “antigenomes” that produce more genomes, completing one amplification cycle. Amplification produces the genomic templates for secondary transcription.

Envelope Complex

The third gene codes for the M protein, a moderately hydrophobic protein that can associate with membranes. M also binds the RNP, and interacts with the F and H protein cytoplasmic tails, modulating cell fusion. M is the assembly factor and may also drive viral particle release.

The fourth gene codes for the F protein. An interesting peculiarity of the F mRNA is its long (almost 500 nt) 5' untranslated region (Fig. 1). The function of this region has not been characterized, but it may regulate protein expression. The F protein is synthesized as a precursor F₀, which folds into a trimer that is cleaved into two subunits, F₁ and F₂ by the ubiquitous intracellular protease, furin. Only activated trimers can unfold and cause membrane fusion, a process which occurs after the attachment protein H contacts a receptor and transmits the fusion triggering signal to the F trimer.

The fifth gene codes for the attachment protein. Since this protein hemagglutinates certain nonhuman primate red blood cells, but lacks neuraminidase activity, it is called H instead of HN. Other Paramyxoviridae use sialic acid as their receptors, and they need neuraminidase to destroy receptor activity while budding from a host cell. Morbilliviruses do not use sialic acid as a receptor and do not need a neuraminidase. The nature and cell-specific expression of the two cellular MeV receptors, SLAM and nectin-4, explains the tissue specific tropism on MeV infections.

Life Cycle

MeV begins its circuit through selected organs of the human body within alveolar macrophages and dendritic cells (DC), which express the primary receptor SLAM. These cells first transfer the infection through the epithelial barrier, and then spread it into lymphoid tissues (Fig. 4, left panel). SLAM was originally identified on activated B and T lymphocytes, but it is also expressed constitutively on immature thymocytes, memory T cells, and certain B cells. Subsets of other cell types, including monocytes and DC, also express SLAM. This cellular distribution overlaps with the susceptibility of different cell types to wild-type MeV infection. Another strong argument for the central role of SLAM in MeV tropism is the fact that three morbilliviruses (MeV, canine distemper virus, and rinderpest virus) enter cells via SLAM (human, canine, or bovine, respectively). Experiments in macaques revealed that the earliest target cells after intratracheal MeV inoculation are DCs and alveolar macrophages.

More recently nectin-4, also called poliovirus receptor-like-4 (PVRL4), was shown to serve as the receptor for MeV spread in the upper airway epithelium (Fig. 4, right panel). This adherens junction protein interacts with H with five times higher affinity than SLAM. Nectin-4 sustains basolateral entry of MeV, and of all animal morbilliviruses examined, into upper airway epithelial cells, including those of the trachea.

Contrary to the infections with other respiratory viruses, MeV enters the airway epithelium "en masse": several days after host-to-host transmission, highly infected immune cells synchronously deliver large amounts of virus to the upper airways, for secondary amplification. Since the upper airways are the anatomical location most useful to support particle aerosolization, this two-phase mechanism of host invasion may account for the extremely contagious nature of MeV infection.

The live attenuated MeV vaccine strain, Edmonston, can also use the regulator of complement activation membrane cofactor protein (MCP; CD46) as a receptor. The primary function of CD46 is to bind and promote inactivation of the C3b and C4b complement products, a process protecting human cells from lysis by autologous complement, a function that requires ubiquitous expression. Only tissue-culture adapted MeV interact with CD46, and indiscriminate cell entry through this protein correlates with MeV attenuation. A chimeric MeV expressing a vaccine strain H protein, which binds to CD46, is attenuated in the nonhuman primate, and productive infection is still confined to SLAM-expressing cells in vivo.

Epidemiology

Humans in all age groups are susceptible to measles, but because the immunity of natural infection is protective life-long, measles occurs only once in a lifetime, as documented by Panum in the 19th century. This clearly differs from infections with other respiratory tract viruses (e.g., respiratory syncytial virus or influenza). Morbidity from measles complications is highest in infants and immunocompromised hosts (non-immune, pregnant women or patients with cancer or AIDS).

Measles remains one of the most contagious of infectious diseases, with attack rates approaching 99% in susceptible populations. Using several epidemiological modeling studies in the recent past, a consensus on the level of herd immunity required to prevent a measles epidemic is up to 94%. One explanation for the high level of contagion is that, contrary to the infections with other respiratory viruses, MeV enters selectively the upper airway epithelium after strong amplification of the infection in immune cells. MeV is then transmitted by respiratory aerosol from infected hosts by coughing and sneezing in the prodromal phase through several days after onset of the measles rash.

In the past, measles epidemics targeted young children. Since the widespread vaccination campaigns began in 1963, measles outbreaks have occurred in older children and young adults. The shifting pattern of measles outbreaks may be due to failure to vaccinate, and to variable rates of primary or secondary vaccine failures.

Clinical Features

Measles, one of the most contagious infectious diseases of man, was recognized clinically by the rash and other signs from early historical times. It was differentiated from smallpox by Rhazes in the tenth century, who noted that it could be more fatal than smallpox. Transmitted by aerosol droplets, the infection has an incubation period of 7–10 days, with onset of fever, cough, and

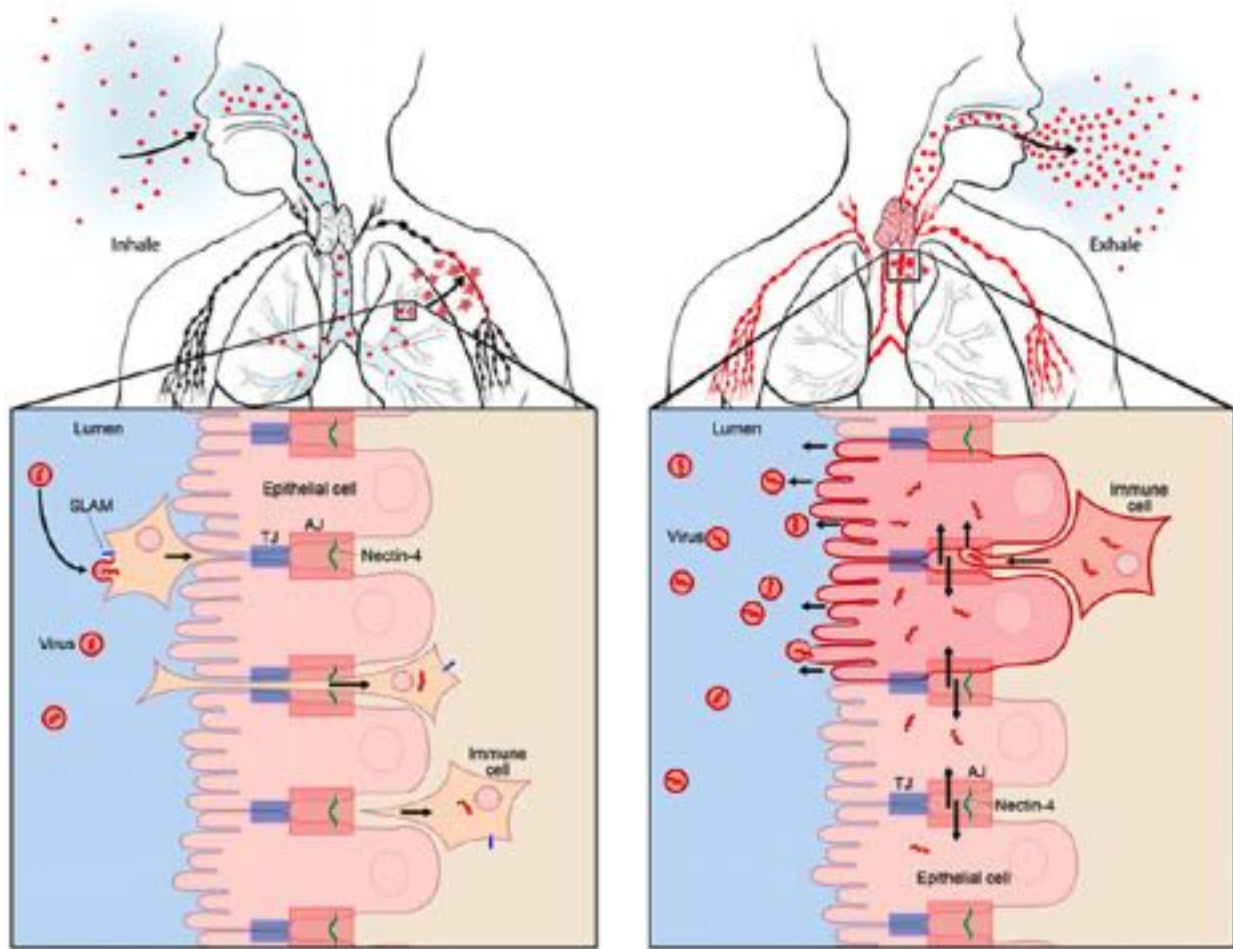


Fig. 4 *How MeV spreads in its host.* Measles and the other morbilliviruses have developed an immunosuppressive host invasion strategy: the airways epithelium is initially bypassed within carrier immune cells (left panel and inset). After replication in cells of lymphoid organs, epithelial invasion occurs from the basolateral side (right panel and inset). *Left half: bypass of the airway epithelium.* MeV (red dots), once inhaled, infects alveolar macrophages or dendritic cells that express SLAM (blue protein on the surface of an immune cell, top left). Virus-carrying immune cells traverse the airways epithelium (inset, top to bottom) and deliver the infection to the local lymph nodes (red to indicate infection). The infection spreads rapidly to the primary lymphoid organs, thymus and spleen, where the virus efficiently replicates. *Right half: basolateral entry and spread through the airway epithelium.* Inset, right side: after viral amplification in lymphoid organs, circulating immune cell delivers the infection to epithelial cells. The plasma membrane of the infected cells is drawn as thick red line to indicate the presence of viral glycoproteins. The membranes of the immune and epithelial cell fuse when the viral attachment protein binds to nectin-4 in the adherens junction (AJ, green, shown only in the lower epithelial cells). Then the infection spreads laterally in the epithelium via adherens junction (black arrows). Infected epithelial cells (top two cells with red plasma membranes) release progeny viral particles in the airway lumen. Preferential nectin-4 expression in the tracheal epithelium (shown in red) facilitates host exit at a strategic location for efficient aerosol particle formation, and explains extremely efficient transmission of MeV. Apical secretion of MeV particles is inefficient, but infectious centers detaching from the epithelium may also be expelled by coughing and sneezing. TJ: tight junction. Reproduced with permission from Mateo, M., Generous, A., Sinn, P.L., Cattaneo, R., 2015. Connections matter — How viruses use cell–cell adhesion components. *Journal of Cell Science* 128, 431–439.

coryza, followed in about 4 days by the skin rash which begins on the face and spreads to the whole body. The skin rash fades in about 5 days and clinical resolution is usually uneventful.

Measles is typically an infection of childhood and protective immunity is lifelong, such that a second case of measles in a child or adult is highly unusual. Prior to widespread vaccination against measles in the 1960s, the infection had a case–fatality rate of less than 5% in children, higher in infants and in children in low income countries, where fatality rates of up to 20% can still occur.

Complications

Complications include pneumonia, encephalitis, otitis media, blindness, and secondary infections by common bacteria and viruses. In low income countries, common complications are diarrhea and wasting. “Atypical measles”, a severe MeV infection that occurred following the use of whole, inactivated virus vaccine from 1963 to 1967, was characterized by high and persistent fever, a

different body rash that resembles Rocky Mountain spotted fever, and lobar pneumonia with effusions. This unusual clinical manifestation is discussed in the section on Pathogenesis.

Compared to infections by other respiratory viruses, measles has more serious clinical potential, possibly because the other infections are usually confined to the respiratory tract and not systemic; but measles in man causes less disease and death than infections due to the most closely related morbilliviruses, rinderpest in cattle, or canine distemper in dogs.

Immunosuppression

Latent tuberculosis may be reactivated following MeV infection. This predictable complication of MeV-induced immunosuppression was initially studied by von Pirquet, who observed that the tuberculin skin test response is suppressed during MeV infection. MeV-induced immunosuppression has an onset near to the peak of viremia, such that the primary immune responses to MeV are not impaired, but the immunosuppression is severe and lasts for several months to years.

The severe and sustained nature of immunosuppression may be mainly the result of strong replication of the virus in SLAM-expressing cells, which include T and B memory cells. Infections by other morbilliviruses are also associated with immunosuppression and this complication was a hallmark of epidemic disease in harbor seals that resulted in the discovery of a new morbillivirus, the phocine distemper virus. In the extreme case, immunosuppression with massive lymphocyte depletion is fatal in canine distemper virus-infected ferrets and rinderpest-infected cattle.

Immune Responses

The primary immune responses to MeV, initially IgM antibody response, type 1 CD4 and CD8 + T-cell responses, followed by neutralizing IgG antibodies, are completely effective in controlling viral replication and resolution of the infectious process. As originally observed by Macfarlane Burnet, both primary and secondary immunodeficiencies that impair T cell responses, for example, the DiGeorge syndrome or advanced HIV infection, are a significant risk for failure of the host to control MeV infection, resulting in persistent infection and death or serious disease in the lower respiratory tract and the central nervous system. In contrast, deficiency of the antibody response does not impair the immune control of MeV replication.

These fundamental observations in antiviral immunity were confirmed by experimental infections of rhesus monkeys with MeV. CD8 + T cell depletion of monkeys at the time of viral inoculation resulted in prolonged viremia until the T cells repopulated, but depletion of B cells had no significant effect on MeV infection. A persistent and/or a fatal MeV infection has occurred in HIV-infected children, with CD4 + T cell counts being below 200/ μ l, who contracted measles. This happened to a young, HIV-infected man who was vaccinated with live, attenuated MeV vaccine as required for college admission. However, the risk of a serious disease form of measles in immunocompromised hosts (50% or greater case fatality) outweighs the risk of vaccination, and measles vaccination is recommended for HIV-infected infants and children unless they have severe immunodeficiency with low, age-adjusted CD4 + T cell counts in blood.

Persistent Infections

Persistent MeV infection is rare, but it can result in giant cell pneumonia or two different neurological diseases: measles inclusion body encephalitis and subacute sclerosing panencephalitis (SSPE). In SSPE, MeV infection spreads slowly by cell contact within the central nervous system; however, the virus is typically defective in its expression of M and/or F proteins and the infection does not lead to the formation of progeny virus particles. Another rare neurologic complication, acute demyelinating encephalomyelitis, not due to continuing viral replication in the brain, is associated with autoimmunity to myelin. Several other diseases, including multiple sclerosis, inflammatory bowel disease and certain behavioral disturbances, have been linked anecdotally to MeV infection, but causal connections have not been established.

Pathogenesis

The histopathological hallmark of MeV infection is the formation of syncytia, or multinucleated giant cells. Syncytial cells are not unique to MeV infection but they are characteristic. MeV infects cells of ectodermal, endodermal, and mesenchymal origin and syncytial cells have been observed in all of these cell types. Multinucleated giant cells were observed in lymphoid organs and in mononuclear cell aggregates in many tissues, including the inflamed lung, where they are referred to as Warthin-Finkeldey giant cells. These syncytia are of lymphocyte, macrophage, dendritic, or reticular cell origin. Syncytial cells are also readily observed in epithelia, including the columnar epithelium of the trachea and bronchi, the stratified squamous epithelium of the skin and buccal mucosa, and the transitional epithelium of the urinary bladder and urethra. Endothelial syncytial cells were observed in small pulmonary arteries of monkeys infected with MeV. Eosinophilic cytoplasmic and nuclear inclusion bodies can be seen in measles giant cells. The syncytial cells of measles are not long lived and they disappear with resolution of the infection.

The major pathologic changes in measles infection are due to inflammation and necrosis followed by tissue repair without fibrosis. Secondary infections by bacterial or other viral pathogens are common and they alter the pathologic process accordingly, especially in the respiratory and gastrointestinal tracts. Pathology in the lower respiratory tract is mainly peribronchiolar inflammation and necrosis with a mild exudate, but interstitial pneumonitis with mononuclear cell infiltrates may occur. In the brain and central nervous system, perivascular mononuclear infiltrates can occur with a few necrotic endothelial cells, microglia, and neurons. In lymph nodes, spleen, and thymus, the major changes are mild to moderate lymphocyte depletion and multinucleated giant cells. Lymph node and splenic follicular hyperplasia are not seen in primary measles infection but are present if the host was previously infected or vaccinated against measles.

The unexpected occurrence of “atypical measles” following vaccination with whole-inactivated virus in the early 1960s and subsequent infection with wild-type virus is thought to be due to an aberrant, anamnestic host immune response resulting in an Arthus reaction or delayed hypersensitivity. Similar immunopathology was seen in children exposed to respiratory syncytial virus (RSV) following vaccination with whole-inactivated RSV. Atypical measles was experimentally induced in monkeys vaccinated with a whole-inactivated MeV vaccine followed by challenge with a wild type virus. Compared to monkeys vaccinated with a live, attenuated MeV vaccine, the aberrant immune responses associated with atypical measles resulted in immune complex deposition and eosinophilia. A marked skewing of T cell cytokine responses toward type 2 response, with abnormally high interleukin 4 production was also observed.

Diagnosis

As measles became a rare illness in regions with high vaccine coverage where the virus has been eliminated or controlled, the diagnosis of MeV infection became a challenge. The typical punctate, maculopapular skin rash, and the prodromal signs and symptoms are not specific for this viral infection and measles diagnosis requires differentiation from other pathogens causing exanthems in children, including Epstein-Barr Virus and cytomegalovirus infections, rubella, dengue, Zika virus infection, scarlet fever, typhus, toxoplasmosis, meningococcus, and staphylococcus.

An enanthem with small white lesions on the buccal, labial, or gingival mucosal (Koplik's spots) is considered specific for MeV infection and it typically occurs a few days before the skin rash. A buccal swab smeared on a glass slide, fixed and stained (e.g., Wright-Giemsa), may show multinucleated syncytia of epithelial cells. Secondary infections due to MeV-induced immunosuppression are caused by the common pathogens in a geographic region. Serological diagnosis of MeV infection is based on early IgM response in blood by antibody-capture ELISA or by other methods detecting increase in serum IgG antibodies between paired serum specimens.

MeV can be cultured from peripheral blood mononuclear cells and from oral or nasal aspirates. Rescue of wild-type MeV is readily done using B95a cells, or Vero cells expressing human SLAM. The virus can be detected by reverse transcription polymerase chain reaction (RT-PCR) from these samples and from urine. Genetic sequencing of MeV isolates worldwide has yielded clustering of viral genomes into ca. 24 distinct genotypes, based on a variable carboxyl-terminal region of the N protein, which have been used to track imported cases of MeV infection and to distinguish wild type virus from vaccine strains. As of 2018, only 4 MeV genotypes have been circulating globally (Fig. 5). The World Health Organization (WHO) hosts a laboratory manual and website with descriptions of measles diagnostic samples and assays.

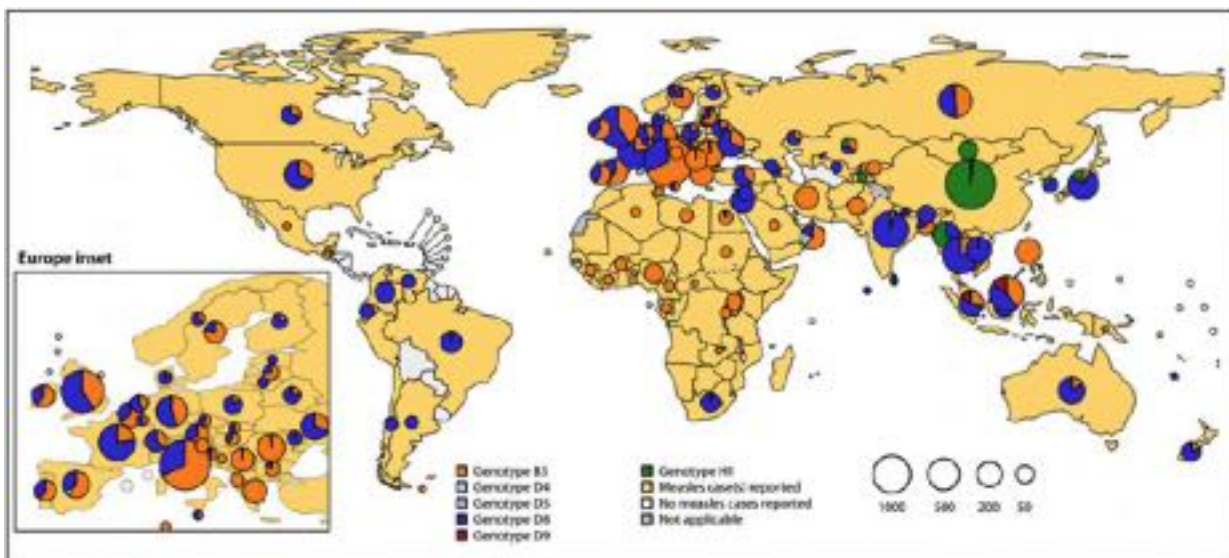


Fig. 5 Global distribution of MeV genotypes, 2016–2018. The size of the circles reflects the numbers of identifications reported for each genotype. Source: World Health Organization. Reproduced from Brown, K.E., Rota, P.A., Goodson, J.L., *et al.*, 2019. Genetic characterization of measles and Rubella viruses detected through global measles and Rubella elimination surveillance, 2016–2018. *Morbidity and Mortality Weekly Report* 68, 587–591.

Treatment

No specific antiviral treatment is recommended for the typical case of measles in a child or adult. In addition to supportive care, vitamin A can prevent disease complications and WHO recommends its preventive use. In a measles outbreak, passive antibody, in the form of human immune globulin given by the intramuscular or intravenous route, can protect non-immune, immunocompromised hosts from serious complications of measles. The antiviral, ribavirin, may attenuate severe measles pneumonia or disease complications in an immunocompromised patient. Novel viral polymerase inhibitors may become available to treat severe measles in the near future.

Prevention

A live attenuated MeV strain, called Edmonston, developed by John Enders and further attenuated by Maurice Hilleman in the 1950s, was licensed in the US in 1963. In 1971, the US government licensed Merck to produce MMR (for Measles, Mumps and Rubella) vaccine by combining MeV with mumps and rubella live, attenuated vaccine viruses. MMR remains one of the safest and most effective human vaccines. Some other attenuated measles vaccine strains are used in other countries. It is remarkable that for this RNA virus, vaccine failure due antibody escape has never occurred.

Measles remains one of the leading causes of vaccine-preventable diseases and death in the world. The WHO and other international bodies have proposed the global eradication of MeV in the twenty-first century, following the current campaign to eradicate poliovirus. This is an achievable goal as humans are the only known host species for this virus, although nonhuman primates are susceptible hosts. Regional elimination of MeV has been achieved: MeV circulation has been interrupted for decades in the Americas, and in many European countries, by the vaccination of children with the live, attenuated vaccines in current usage. However, decline of vaccination coverage has led to recent measles re-emergence in Europe and North America.

Many countries are planning to increase vaccination coverage as they progress toward targeted reduction of measles mortality, or elimination of transmission. Recent experience in high and low income countries has shown that the maintenance of effective herd immunity against MV requires a two-dose vaccination strategy. Because of the highly contagious nature of this infection, about 95% of children require vaccination to prevent the circulation of MeV in a population of more than several hundred thousand. Greater than 90% measles vaccine coverage has been achieved in many countries but more than 25% of countries have not reached this level of vaccination yet. The HIV pandemic presents a potential obstacle to MeV eradication globally, especially in Africa and Asia, but HIV-infected children that are not immunocompromised can be safely vaccinated with live, attenuated MeV (see above). The importation of MeV into countries where elimination has been achieved is a very real challenge for diagnosis and control which has been successfully met by molecular genetic methods of rapid viral genomic sequencing and taxonomy.

MeV-Based Vaccines and Cancer Therapeutics

The live, attenuated measles vaccine, which has an outstanding efficacy and safety record, is being developed as a pediatric vaccine eliciting immunity against additional microbial infections. Vectored MeV-expressing the hepatitis B surface antigen (HBsAg) has been generated by reverse genetics. One of these vectored MeV vaccines induced protective levels of HBsAg antibodies while protecting rhesus macaques against measles challenge. Another vectored MeV expressing the capsid and envelope protein from chikungunya virus is currently in a phase III clinical trial. Another advantage of immunization with a di- or multivalent MV-based vaccine is the delivery of an additional immunization safely without additional cost.

The knowledge gained from basic research has also been applied to the development of vectors for targeting and eliminating cancer cells. Oncolytic virotherapy is the experimental treatment of cancer patients based on the administration of replication-competent viruses that selectively destroy tumor cells but leave healthy tissue unaffected. MeV is one of several human vaccine strains, or apathogenic animal viruses, currently being genetically modified to improve oncolytic specificity and efficacy. These modifications include targeting cell entry through designated receptors expressed on cancer cells. Moreover, cancer cell specificity can be enhanced by silencing the expression of proteins that counteract the interferon system, which is often non-functional in cancer cells. Clinical trials of MV-based oncolysis for ovarian cancer, myeloma, and glioma are ongoing.

Further Reading

- Bankamp, B., Hickman, C., Icenogle, J.P., *et al.*, 2019. Successes and challenges for preventing measles, mumps and rubella by vaccination. *Current Opinion in Virology* 34, 110–116.
- Griffin, D.E., 2013. Measles virus. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 1042–1069.
- Katz, S.L., 2004. Measles (rubeola). In: Gershon, A.A., Hotez, P.J., Katz, S.L. (Eds.), *Krugman's Infectious Diseases of Children*, eleventh ed. Philadelphia, PA: Mosby, pp. 353–372.
- Lamb, R.A., Parks, G.D., 2013. *Paramyxoviridae*. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 957–995.
- McChesney, M.B., Miller, C.J., Rota, P.A., *et al.*, 1997. Experimental measles. I. Pathogenesis in the normal and the immunized host. *Virology* 233, 74–84.

- Miest, T.S., Cattaneo, R., 2014. New viruses for cancer therapy: Meeting clinical needs. *Nature Reviews Microbiology* 12, 23–34.
- Mühlebach, M., Mateo, M., Sinn, P.L., *et al.*, 2011. Adherens junction protein nectin-4 (PVRL4) is the epithelial receptor for measles virus. *Nature* 480, 530–533.
- Navaratnarajah, C.K., Generous, A.R., Yousaf, I., Cattaneo, R., 2020. Receptor-mediated cell entry of paramyxoviruses: Mechanisms, and consequences for tropism and pathogenesis. *Journal of Biological Chemistry* 295, 2771–2786.
- Panum, P., 1939. Observations made during the epidemic of measles on the Faroe Islands in the year 1846. *Medical Classics* 3, 829–886.
- Pfaller, C.K., Bloyet, L.-M., Donohue, R.C., *et al.*, 2020. The C protein is recruited to the measles virus ribonucleocapsid by the phosphoprotein. *Journal of Virology* 94, e01733.
- Rota, P.A., Moss, W.J., Takeda, M., *et al.*, 2016. Measles. *Nature Reviews Disease Primers* 2, 16049.
- von Pirquet, C., 1908. Das Verhalten der kutanen Tuberkulinreaktion während der Masern. *Deutsche Medizinische Wochenschrift* 34, 1297–1300.

Relevant Websites

- [https://www.who.int/immunization/monitoring_surveillance/burden/laboratory/manual/en/Immunization, Vaccines and Biologicals.](https://www.who.int/immunization/monitoring_surveillance/burden/laboratory/manual/en/Immunization,Vaccines%20and%20Biologicals)
- <https://www.mayoclinic.org/diseases-conditions/measles/symptoms-causes/syc-20374857>
Measles.
- <https://www.cdc.gov/measles/index.html>
Measles (Rubeola).

Molluscum Contagiosum Virus (*Poxviridae*)

Joachim J Bugert and Rosina Ehmann, Bundeswehr Institute of Microbiology, Munich, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J.J. Bugert, Molluscum Contagiosum Virus, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:[10.1016/B978-0-12-801238-3.02621-0](https://doi.org/10.1016/B978-0-12-801238-3.02621-0).

Nomenclature

H&E stain Histological staining method using the two colorants hematoxylin and eosin

Glossary

Acanthoma Benign tumor of the stratum spinosum of the dermal epithelium.

Curettage Scoop-like surgical instrument for the removal of tissue by scraping or scooping.

Detritus Accumulation of dead biological material, e.g., dead cells from tissue.

Florid Disease manifestation in its fully developed form.

Histology The study of microscopic structures of tissues often with the help of fixed and stained sections.

Lobule Anatomical term for a clear division or extension of an organ.

Stratum spinosum Prickle cell layer in the dermal epithelium.

Virion Structure

MCV particles have a brick-shaped ovoid morphology similar to other vertebrate poxviruses with an average size of 360×210 nm and the tubular core protein structures typical for MCV ([Fig. 1](#)).

History

"Molluscum contagiosum" was first described as a separate disease entity and transmissible infection of the human skin by T. Bateman (1778–1821) in 1814. W. Henderson subsequently observed viral inclusion bodies in microscopical sections of MCV infected skin, which he named "molluscum bodies", in 1841, also known as "Henderson–Paterson bodies". Molluscum bodies were later compared to "Borrel bodies", inclusion bodies found in avipoxvirus associated infections, and a viral etiology was therefore proposed by Goodpasture, Woodruff and King around 1930.

MCV was first transmitted between human volunteers using sterile filtered (cell free) extracts of human MCV lesion material in the laboratories of Juliusberg, Wile and Kingery in the early 20th century.

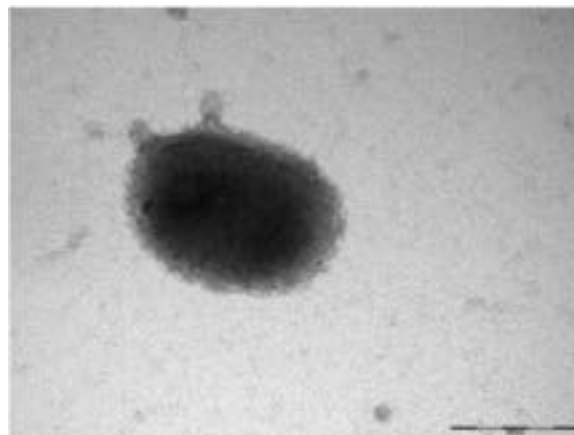


Fig. 1 MCV electron microscopy. Ammonium molybdate negative stain of an MCV particle demonstrating the typical core protein pattern and envelope structure (scale bar 200 nm). Electron microscopy: Bugert and Hobot, Cardiff University School of Medicine, 2005.

Genome

The genomes of sequenced MCV isolates vary slightly in size around 190,000 bases and show the classical structure of poxviruses with a “conserved core” harboring the genes that code for the proteins of the replication and transcription machinery. The conserved core region is flanked by genes that target interactions with the host immune system. The left and right genome termini contain “inverted terminal repeats”, important structures needed for the replication process of poxviruses. To date there are 15 full length genomes available for MCV subtypes 1 and 2. Two other subtypes named 3 and 4 are rarely described and so far have not been characterized on the genome level. MCV subtypes 1 and 2 are genetically very similar and all subtypes 1–4 of MCV cause the same clinical picture. MCV is predicted to encode a total of 170–181 proteins with many of them not yet characterized.

Life Cycle

MCV is an extremely specialized virus that targets only human keratinocytes. After attachment of MCV particles to target cells and subsequent entry, genomic DNA is released into the cytosol by a process called uncoating. mRNA synthesis of viral genes is regulated into three temporal phases: early, intermediate and late. This allows the virus to first produce viral proteins that interact with the host immune system to create a favorable environment for virus replication and transcription in the cell while proteins needed for the assembly of complete virus particles may be produced at a later time point of infection. After formation of progeny virus the particles have to undergo complex maturation steps until they are released from the cell ready to initiate the next replication cycle (Fig. 2).

Interestingly, MCV does not replicate in standard cell culture systems. It causes cytopathogenic effects in human fibroblasts and expresses early genes in a large range of human and animal cells. However, the infection does not yield virus in vitro, and seems to be abortive due to what appears to be a second stage uncoating defect (Fig. 2).

Epidemiology

MC is a frequent disease with worldwide distribution. Reported incidence rates vary between 0.14%–7% and seem to be highest in children aged 2–5 years. However, infection can occur at any age and all ethnicities are affected. Hot and humid climates might favor infection with MCV. Serological studies in several countries suggest that MCV has a surprisingly high prevalence in many populations. Non-neutralizing antibodies are found in 5%–30% of populations studied so far. The discrepancy to the considerably lower incidence rates could be explained by a high number of subclinical or very mild infections. Risk groups for MC

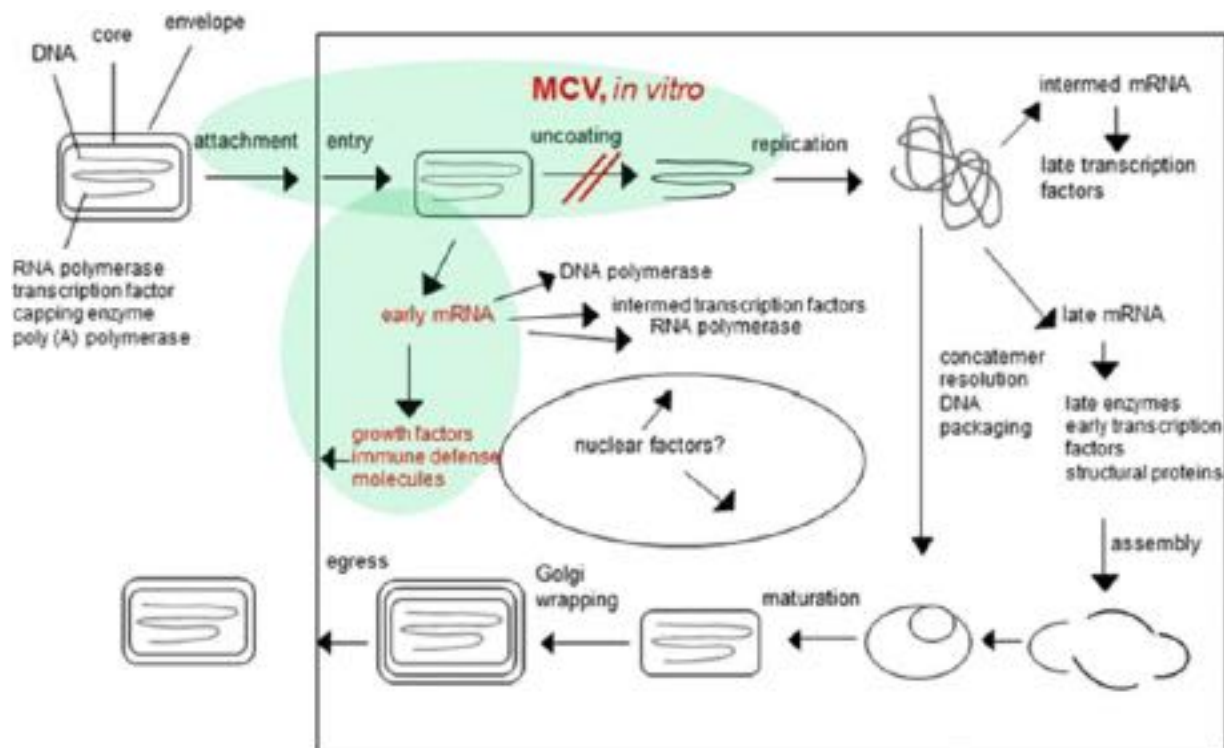


Fig. 2 MCV life cycle in standard single cell culture (green) on the background of the full poxviral lifecycle.

include groups of people living under crowded and poor hygienic conditions or closed communities like nurseries or military barracks, as well as immunocompromised individuals.

The disease is transmitted via smear infection with the virus-containing chalky substance released from MC lesions either directly between humans (self-inoculation of different skin areas of the same patient is also commonly seen) or indirectly via fomites. MC is also a sexually transmissible disease.

Productive MCV infection is presumably specific for humans. Attempts to transmit MCV to laboratory animals were unsuccessful, unless MCV infected human skin was transplanted into SCID mice. MC-like infections have been reported in various animal species, however, characterization on the molecular level was so far achieved only for a molluscum contagiosum-like virus which causes lesions in horses, donkeys and other equids. Though clearly related to MCV, this virus was found to form a separate species.

Clinical Features

After a variable incubation time of 2 weeks to 6 months MCV causes benign tumors of the human skin. Because of their limitation to the epidermal layers these tumors are classified as acanthomas. The typical epidermal tumors are described as pearly, pinkish protrusions with an average diameter of 1–5 mm (**Fig. 3**). The lesions have a solid consistency and a central indentation or umbilication. The indentation is formed by regions of cellular decay releasing a white chalky substance composed of large amounts of MC virions and cellular debris.

MC lesions are mainly located on the face, neck, arms and in the genital areas. Lips, tongue, and mucous membranes are rarely affected. MC is not known to infect the skin of the palms or the soles of feet. If MC affects the eyelid it often leads to subsequent conjunctivitis and keratitis by mechanical irritation.

Patients can have several hundred lesions which develop by self-inoculation over time. These lesions persist for several weeks to months and can last up to years in some cases. MC lesions are usually painless if not manipulated by scratching. Signs of inflammation often occur only shortly before regression of the skin proliferations. If fever occurs it is usually linked to bacterial superinfection of the lesions.

Immunosuppressed patients such as HIV-infected individuals or patients after splenectomy are frequently reported to develop higher numbers of MC lesions and often with a larger size (10 mm and more in diameter, also called molluscum giganteum). In such patients, MC lesions can be florid and often superinfected. This is known as “eczema molluscatum” in analogy to similar conditions observed in herpes simplex virus and vaccinia virus infections. While self-limiting in otherwise healthy individuals, MC may be persistent and hard to manage in the immunocompromised.

Pathogenesis

After transmission by smear infection, MCV enters basal keratinocytes and induces a marked hyperproliferation of the stratum spinosum. Infected cells form large eosinophilic inclusion bodies containing large numbers of virions that compress the nuclei of



Fig. 3 MCV clinical picture. Forearm of a patient with approximately 50 MC lesions, some of which are bacterially superinfected. The lesions are in different stages of development suggesting autoinoculation and local dissemination.

the keratinocytes, often dislocating them to the cellular periphery. Due to the hyperproliferation the epithelium starts to wrinkle and form lobules. After initial proliferation more and more cells in the upper layers of the epithelium die and form the chalky mass that contains large numbers of virions and cellular detritus. This zone of cellular decay leads to the central indentation from which the chalky mass can be readily expressed in mature lesions.

Healing of these lesions is often associated with signs of inflammation that include swelling and reddening.

A unique trait of MCV is its outstanding ability to evade and subvert the host immune system enabling the virus to persist for many weeks to months in the infected tissues. Several viral proteins play a major role in the molecular pathogenesis of MC. These include:

- (1) MC005L – interferes with the modulatory protein NEMO and its interaction with ubiquitin thereby inhibiting NF- κ B activation.
- (2) MC007L – targets retinoblastoma protein (pRb). Dysregulation of this protein is associated with tumor pathogenesis and enables the virus to perturb the cell cycle.
- (3) MC54L – exhibits sequence similarity to human interleukin 18 binding protein (IL-18BP) and acts as a decoy receptor for IL-18. This effectively decreases the IL-18 mediated synthesis of interferon gamma, activation of NK cells and Th1-mediated immune response.
- (4) MC66L – encodes a homolog of selenocysteine-containing glutathione peroxidase. This protects infected cells from peroxide radicals and ultraviolet-mediated apoptosis. It may also explain how MCV replicates exclusively in the epidermis. MC66L mediated anti-apoptotic activity may also explain the increased survival time of MCV-infected cells in the UV exposed epidermis, allowing MCV a longer productive cycle.
- (5) MC80R – this protein shows moderate homology to human major histocompatibility complex (MHC) class I and mediates ER-retention of host MHC I by a so far unique mechanism. ER-retention of MHC I leads to a reduced cell surface presentation of viral antigens to cells of the immune system.
- (6) MC132 – also inhibits NF- κ B activation, however, the target protein for interaction is the cellular transcription factor p65.
- (7) MC148R – encodes a secreted protein which has a homology with CC family chemokines. A candidate gene for a MC148R human homolog was discovered mapping to chromosome 9p13. Subsequent studies showed that MC148R protein is a highly selective antagonist to CCR8 in human (but not mouse) monocytes and lymphoid tissues, and can conceivably interfere with monocyte invasion and dendritic cell function. However, the observation that MC148R can inhibit allograft rejection in transgenic mice supported the theory that it may possess anti-inflammatory properties. MC148 is expressed early in the MCV lifecycle, and may provide an anti-inflammatory activity gradient in the transition zone between epidermal and dermal tissue as a form of stand-off defensive device for the virus before other mechanisms engage.
- (8) MC159L and MC160L – known as vFLIPs (virus Fas-linked ICE-like protease), are inhibitors of death receptor-induced apoptosis and programmed necrosis. They share homology with caspase-8 and caspase-10 in the tandem death effector domains (DEDs) at the amino termini. MC159L blocks apoptosis induced by FAS and TNF. A recent study on vaccinia virus infected MC159 transgenic mice revealed enhanced innate inflammatory reaction characterized by increased expression of the chemokine monocyte chemoattractant protein-1 (MCP-1) and infiltration of $\gamma\delta$ T cells into peripheral tissues. Thus, MC159L is considered a dual-function immune modulator. MC160 protein inhibits NF- κ B activation either by its DED-containing N-terminus binding to procaspase-8 or its C-terminal region interacting with Hsp90, leading to a degradation of IKK (I κ B kinase)1 and inhibition of phosphorylation of the p65 subunit.
- (9) MC163L – was found to be a regulator of apoptosis. Expression of this protein dampened cellular apoptotic responses by interaction with several pathways.

Diagnosis

The pathognomonic appearance of the umbilicated lesions with a central discharge of the chalky cellular detritus is often sufficient for clinical diagnosis. In doubt, histological sections stained with H&E can assist diagnosis. Histologic presentation of MC is characterized by hyperproliferation of the stratum spinosum with a typical formation of lobules. Large numbers of eosinophilic inclusion bodies can be found in the hyperproliferating keratinocytes and the cellular debris which is discharged from the central indentation (**Fig. 4**).

Other diagnostic methods include electron microscopy to demonstrate MCV particles, PCR and real-time PCR for MCV nucleic acid detection and ELISA for detection of antibodies against MCV.

Differential diagnoses include infectious (cutaneous cryptococcosis, furuncles, herpesvirus infections, histoplasmosis and papillomavirus infections) and non-infectious diseases (adnexal tumors, basal cell carcinoma, fibrous papules, naevi).

Treatment

In immunocompetent patients – especially in children – it is advisable to wait for spontaneous disease regression. Although MC lesions may persist for several months the disease is usually self-limiting and heals without scar formation. If treatment is needed because of cosmetic reasons or complications like bacterial superinfection, MC lesions in mechanically stressed locations (eyelid, inguinal region) or persistent infection in immunocompromised, there is a broad choice of surgical and medical treatment options available. Surgical intervention includes curettage, cryotherapy, laser treatment, electrosurgery or occlusive duct tape. Medical options can be divided into topical and systemic approaches. Topically applied drugs include acantholytic and keratolytic substances like

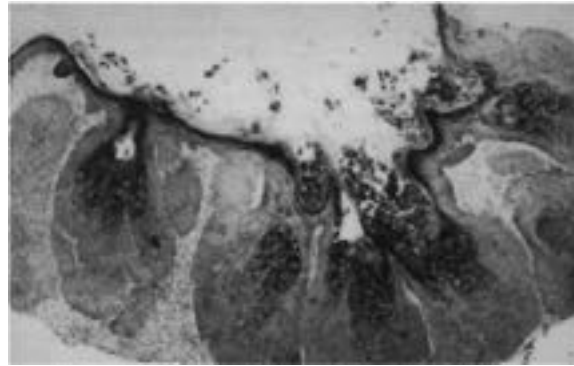


Fig. 4 Histologic presentation of a typical MC lesion. Histological section (HE stain) of an MC lesion demonstrates the typical lobular organization and hyperproliferative character of the skin affections in MC. Virus particles and lipid debris are discharged from the central indentation.

cantharidin, potassium hydroxide (KOH) and salicylic acid. Other topical solutions reported to alleviate MC lesions include drugs and plant extracts like podophyllotoxin, Australian lemon myrtle (*Backhousia citriodora*), diphencyprone, 5-aminolaevulinic acid and nitric oxide (NO). Systemic treatments include antiviral agents like Cidofovir, treatment of underlying HIV infection with antiretroviral drugs, as well as interferon- α . The immunomodulatory drug imiquimod received a lot of attention during the last few years, however, the results of clinical studies led to controversial conclusions and there is so far no reliable evidence for its effectiveness.

In general, it has to be stated that although MC is a relatively frequent disease, high-quality clinical studies to compare different therapeutic options for MC are scarce leaving no consensus on best practice. The 2017 Cochrane review on intervention for cutaneous molluscum contagiosum comes to the conclusion that there is no reliable evidence-based recommendation for the treatment of MC so far and that awaiting spontaneous resolution is favored if possible.

Prevention

There is no vaccine available. The best prevention is avoidance of smear infection by following good hygiene practices: covering of MC lesions with clothing or bandages, avoidance of scratching and manipulation of the papules.

Further Reading

- Bugert, J., Alikhan, A., Shwayder, T., 2019. Molluscum contagiosum. In: Hoeger, P., Kinsler, V., Yan, A. (Eds.), *Harper's Textbook of Pediatric Dermatology*, fourth ed. © 2019 John Wiley & Sons Ltd.
- Chen, X., Anstey, A.V., Bugert, J.J., 2013. Molluscum contagiosum virus infection. *The Lancet Infectious Diseases* 13, 877–888. doi:10.1016/S1473-3099(13)70109-9.
- Forbat, E., Al-Niaimi, F., Ali, F.R., 2017. Molluscum contagiosum: Review and update on management. *Pediatric Dermatology* 34, 504–515. doi:10.1111/pde.13228.
- Harvey, I.B., Wang, X., Fremont, D.H., 2019. Molluscum contagiosum virus MC80 sabotages MHC-I antigen presentation by targeting tapasin for ER-associated degradation. *PLoS Pathogens* 15 (4), e1007711. doi:10.1371/journal.ppat.1007711.
- Katz, K.A., Swetman, G.L., 2013. Imiquimod, molluscum, and the need for a better “best pharmaceuticals for children” act. *Pediatrics* 132, 1–3. doi:10.1542/peds.2013-0116.
- Katz, K.A., 2014. Imiquimod is not an effective drug for molluscum contagiosum. *The Lancet Infectious Diseases* 14, 372–373. doi:10.1016/S1473-3099(14)70728-5.
- Nichols, D.B., De Martini, W., Cottrell, J., 2017. Poxviruses utilize multiple strategies to inhibit apoptosis. *Viruses* 9, 215–250. doi:10.3390/v9080215.
- Senkevich, T.G., Koonin, E.V., Bugert, J.J., Darai, G., Moss, B., 1997. The genome of molluscum contagiosum virus: Analysis and comparison with other poxviruses. *Virology* 233, 19–42. doi:10.1006/viro.1997.8607.
- Sherwani, S., Farleigh, L., Agarwal, N., et al., 2014. Seroprevalence of molluscum contagiosum virus in German and UK populations. *PLoS One* 9, e88734. doi:10.1371/journal.pone.0088734.
- Shisler, J.L., 2015. Immune evasion strategies of molluscum contagiosum virus. *Advances in Virus Research* 92, 201–252. doi:10.1016/bs.aivir.2014.11.004.
- Trčko, K., Hošnjak, L., Kušar, B., et al., 2018. Clinical, histopathological, and virological evaluation of 203 patients with a clinical diagnosis of molluscum contagiosum. *Open Forum Infectious Diseases* 5. doi:10.1093/ofid/ofy298.
- van der Wouden, J.C., van der Sande, R., Kruithof, E.J., et al., 2017. Interventions for cutaneous molluscum contagiosum. *Cochrane Database of Systematic Reviews* 5, CD004767. doi:10.1002/14651858.CD004767.pub4.
- Zorec, T.M., Kutnjak, D., Hošnjak, L., et al., 2018. New insights into the evolutionary and genomic landscape of molluscum contagiosum virus (MCV) based on nine MCV1 and six MCV2 complete genome sequences. *Viruses* 10. doi:10.3390/v10110586.

Relevant Websites

- <https://www.cdc.gov/poxvirus/molluscum-contagiosum/index.html>
Molluscum Contagiosum.
- <https://4virology.net>
Viral Bioinformatics Research Centre.

Mumps Virus (*Paramyxoviridae*)

Stephen A Winchester and Kevin E Brown, Frimley Park Hospital, Frimley, United Kingdom and Immunisation and Counter-measures Division, Public Health England, London, United Kingdom

Crown Copyright © 2021 Published by Elsevier Ltd. All rights reserved

This is an update of B.K. Rima, W.P. Duprex, Mumps Virus, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:[10.1016/B978-0-12-801238-3.02623-4](https://doi.org/10.1016/B978-0-12-801238-3.02623-4).

History

The earliest description of mumps dates back to the fifth century BC by Hippocrates of a disease capable of spreading in humans with characteristic enlargement close to the ears and painful swelling of the testes common in young men. Infection in the central nervous system (CNS) and meninges in some cases of mumps was first noted by Hamilton in 1790. In 1934, Johnson and Goodpasture demonstrated the filterable nature of the causative agent and Koch's postulates were fulfilled by infection of human volunteers with virus propagated in the salivary glands of monkeys. The first vaccine against mumps was developed in 1946, and it protected macaques against infection.

Taxonomy and Classification

Mumps virion is a spherical or pleomorphic particle with glycoprotein projections 150-nm or more in diameter (**Fig. 1**). It is stable at 4°C for several days but its lipid membrane is sensitive to ether and other membrane-destroying reagents, and has haemagglutination and neuraminidase (HN) activities. It contains a nonsegmented negative-strand RNA genome and is classified in the family *Paramyxoviridae*, in the genus *Orthorubulavirus* within the *Rubulavirinae*. MuV is the prototype species, and two other human *Rubulavirinae*, parainfluenza virus 2 (hPIV2) and hPIV4 have been identified. Other rubulaviruses infect a range of vertebrates – for example, Mapuera virus infects bats and porcine rubulavirus neonatal pigs. Simian virus 5 (SV5) was originally isolated from rhesus and cynomolgus monkey kidney-cell cultures and thus designated as a primate virus. SV5 was sometimes referred to as canine parainfluenza virus although it has a more extended host range and has now been isolated from dogs, pigs and human bone marrow. Consequently, the virus has now been renamed PIV5.

MuV has only been isolated from humans, although other species including macaques, dogs, mice and ferrets can be experimentally infected. Recently, sequences of a closely related virus have been identified in samples from fruit bats.

Properties of the Virion

Paramyxoviruses consist of an inner ribonucleoprotein (RNP) core surrounded by a lipid bilayer membrane from which the glycoproteins protrude (**Fig. 1**). Electron microscopy (**Fig. 1(a)**) shows that MuV has a typical paramyxovirus structure with a lipid

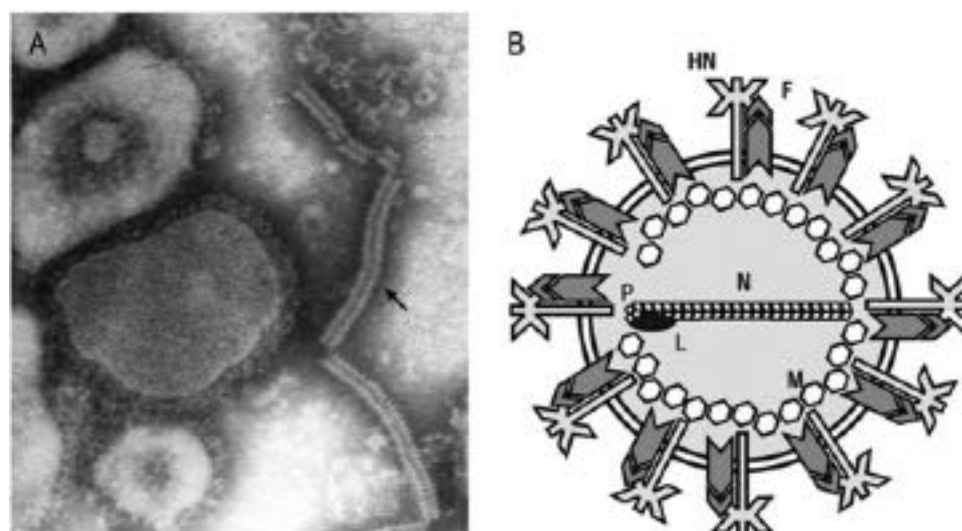


Fig. 1 The structure of the mumps virion. In electron micrograph (a) the arrow points to a nucleocapsid released from a virus particle showing the characteristic herringbone structure. Schematic representation (b) indicates the location of the major viral structural proteins (see **Fig. 2** for an explanation of the shapes and abbreviations).

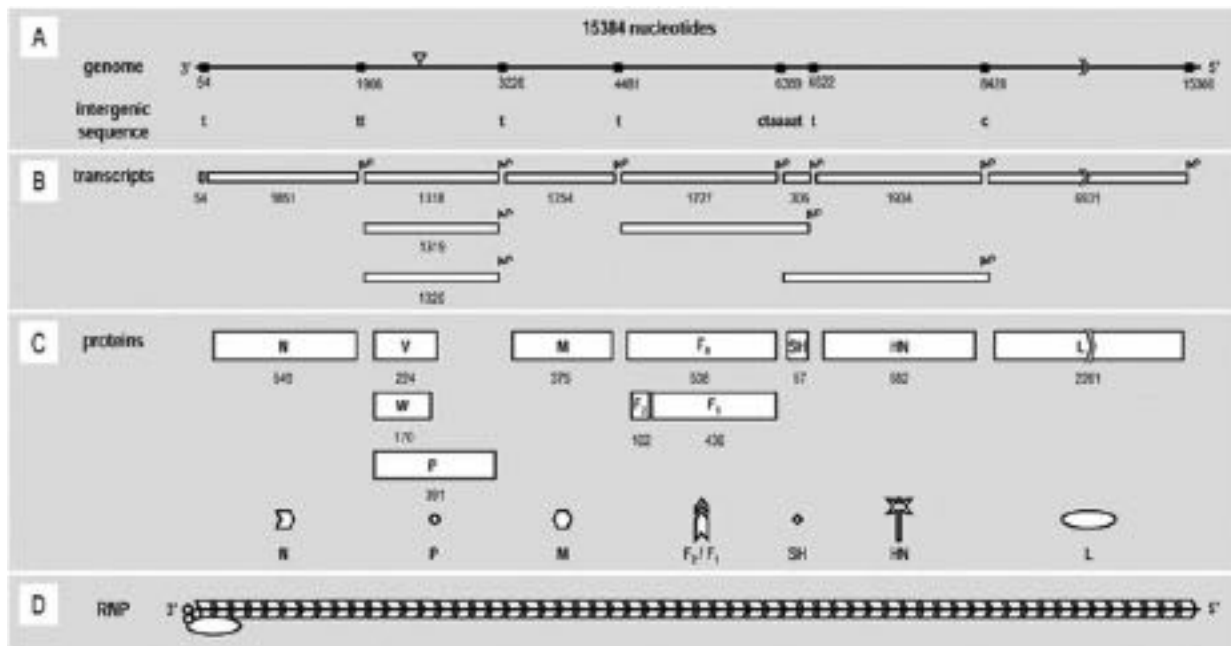


Fig. 2 Transcription, replication, and translation of the MuV genome. The mumps gene order, the size and the position of encoded gene, sequence (positive strand) of the Ig sequences (a); the major transcripts derived from the MuV genome and their sizes (b); the proteins and their sizes in numbers of amino acid residues and their schematic representations (c), and the replicative intermediates (d).

bilayer membrane surrounding an internal RNP complex, termed the nucleocapsid. The nucleocapsid displays the herringbone structure characteristic of *Paramyxoviridae* (arrow, Fig. 1(a)) and is approximately 1 μm in length with a diameter of 17 nm and an internal central core of 5 nm. Some pleomorphic particles have been reported to contain more than one RNP structure. The biological significance of such polyploid particles has not been investigated.

The RNP contains a negative (–) sense RNA genome covered with nucleocapsid (N) protein as well as a phospho- (P) protein and the large (L) protein. The RNP is surrounded by a lipid bilayer derived from the host cell in which the matrix (M) protein of the virus is embedded (Fig. 1(b)). This protein interacts with the internal core N protein and the cytoplasmic tails of the viral glycoproteins. Spikes (10–15 nm in length) protrude from the membrane and these contain the haemagglutinin–neuraminidase (HN) and the fusion (F) glycoproteins. The HN glycoprotein is probably a homo-tetramer and the F protein a homo-trimer. The ratio of HN and F glycoproteins in the spike complexes has not been elucidated.

The spikes have haemagglutination and neuraminidase activity and will haemolyse red cells from a range of different species. Hemolysis reflects the ability of the virus to fuse with infected cells.

Properties of the Genome

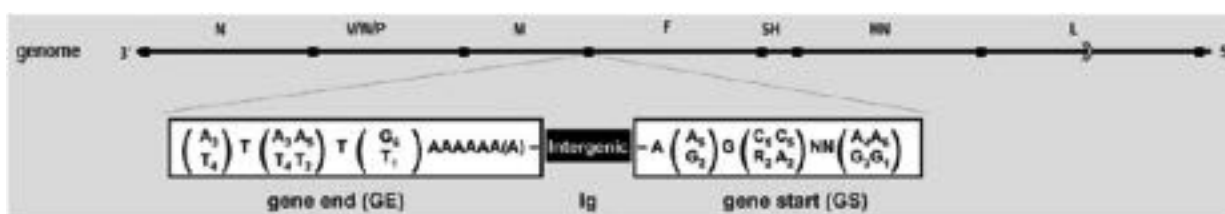
MuV has a single nonsegmented negative-strand RNA genome (Fig. 2(a)). The nucleotide sequences of the complete genome is 15,384 nucleotides in length. The MuV gene order and transcription of the genome is similar to other paramyxoviruses, with seven transcription units separated by intergenic (Ig) sequences. The basic unit of infectivity is the negative-stranded, encapsidated RNP (Fig. 2(d)). The virion is probably polyploid containing several RNPs.

Properties of the MuV Proteins

The properties of the MuV proteins are summarized in Table 1 and the number of amino acid residues in each viral protein is given in Fig. 2(c). Six structural proteins, namely, the N, P, M, and L proteins, as well as two glycosylated membrane-spanning proteins, the HN and F proteins, have been identified in MuV virions. At least two nonstructural proteins (V and W) are synthesized from transcripts of the V/W/P gene. The presence of a small hydrophobic (SH) protein has been demonstrated in MuV-infected cells using antisera to peptides derived from the deduced amino acid sequence. It is not clear whether the protein is incorporated into virions. At least one strain (Enders) expresses the SH gene as a tandem read-through transcript with the F gene in tissue culture. It is unlikely that the SH protein is translated from an F–SH bicistronic mRNA. The growth of Enders strain in tissue culture may not require the expression of SH. A recombinant virus lacking the SH protein has been shown to be capable of growing in several cell types, formally proving a nonessential nature of the protein for growth of the virus in cell culture.

Table 1 Properties of the proteins of MuV

Protein		Size (kDa)	Function
Nucleocapsid protein	N	68–73	Phosphorylated structural protein of RNP; possible role in regulation of transcription and regulation
Phosphoprotein	P	45–57	Associated with RNP; possible role in RNA synthesis
Large	L	>200	Polymerase activity; role in capping, methylation and polyadenylation
Matrix	M	39–42	Associated with inner side of membrane; interacts with N protein and HN and F glycoprotein
Fusion	F	65–74	F ₂ -F ₁ heterodimer activated by a proteolytic cleavage of F ₀ ; fusion of virion membrane with plasma membrane (with HN)
Haemagglutination-neuraminidase	HN	74–80	Glycoprotein with haemagglutination and neuraminidase activities; accessory role in fusion with plasma membrane
Small hydrophobic	SH	6	Membrane protein of unknown function. Sequence used to determine genotype
Non-structural V	V or NS1	23–28	Role in STAT1 and STAT3 degradation
Non-structural W	W or NS2	17–19	Unknown role

**Fig. 3** Consensus sequences of the seven gene start (GS) and gene end (GE) sequences. The various consensus sequences are shown as positive strand sequences. The specific intergenic sequences and their positions in the genome are shown in [Fig. 2\(a\)](#).

Physical Properties

The MuV virion is sensitive to heat and treatment with ultraviolet (UV) light. The virus is inactivated by 0.2% formalin and the presence of the lipid bilayer confers sensitivity to both ether and chloroform. Treatment with 1.5 M guanidine hydrochloride leads to a selective inactivation of neuraminidase but not haemagglutination activity of the virus, indicating that separate domains of the HN molecule are responsible for these two functions.

Replication

MuV infects a variety of cells *in vitro* but whether this is the case *in vivo* remains to be determined. The attachment of MuV is mediated primarily through the interaction of the HN glycoprotein with an α 2,3-linked sialic acid. After introduction of the RNP into the cell, primary transcription of the negative-stranded genome occurs within the cellular cytoplasm. This is mediated by the viral L and P proteins ([Fig. 2\(b\)](#)). The transcription complex recognizes the 3' end of the genome and transcribes the genes sequentially ([Fig. 2\(c\)](#)), stopping at each Ig sequence to synthesize polyadenylate (A_n) tails of the various mRNAs by repeated transcription on a poly(U) stretch in the genome. The Ig sequences are between 1 and 7 nt-in length ([Fig. 2\(a\)](#)) and they are flanked by highly conserved gene end (GE) and gene start (GS) signal sequences ([Fig. 3](#)). Co-transcriptional editing is responsible for the generation of mRNAs encoding the P and W proteins of MuV. Some of the viral proteins are modified post-translationally ([Table 1](#)). The HN glycoprotein forms oligomers during transport through the Golgi complex whereas the F glycoprotein rapidly matures simultaneously with its glycosylation. During replication of the genome the RdRp ignores the GE, Ig, and GS signals to generate one single positive-strand RNA molecule. The role(s) of V and W nonstructural proteins in the processes of replication, transcription, translation, and assembly has yet to be elucidated. Neither is it clear what is the role of SH protein in the virus life cycle.

Serologic Relationships and Variability

MuV is a monotypic virus, and tests with human sera indicate the existence of only a single serotype. Polyclonal sera from infected individuals show a low level of cross-reactivity with hPIV2 and PIV5. Variability between strains has been demonstrated using monoclonal antibodies against the HN and N glycoproteins showing the greatest diversity when single epitopes are examined.

WHO currently recognizes 12 genotypes of MuV, assigned with letters from A to N (excluding E and M) which are based on the nucleotide sequence of the highly variable SH gene. These are routinely used in molecular epidemiological studies (see below).

Evolution and Genetics

Human populations only became dense enough to sustain MuV from about 4000 years ago and, therefore, it has been suggested that the virus must have evolved from an animal reservoir. Recently closely related Rubulaviruses have been identified in a number of different bat species.

Neither temperature-sensitive nor any other conditional lethal mutants of MuV have been reported nor has recombination been described in any nonsegmented negative-strand RNA viruses. Although host range mutants have not been isolated, adaptation of the virus to grow in embryonated eggs or in chicken embryo fibroblasts requires a number of blind passages. Strains adapted in this way do not readily grow and fail to generate syncytia in mammalian cells in culture. As MuV is a neuropathogenic virus some clinical isolates have been adapted to grow in the CNS of experimental animals to study their neuropathogenicity. These viruses have been used in reverse genetics approaches to identify molecular determinants of neuropathogenesis. These main determinants appear to reside in the viral envelope proteins. At present neutralizing monoclonal antibody escape mutants of the HN glycoprotein are the only type of MuV mutants which have been described.

Epidemiology

Geographic and Seasonal Distribution

Prior to vaccination the peak incidence of mumps was in children of 5–7 years of age. The overall risk of infection in unvaccinated populations has been observed to be equal between genders but males are at increased risk of complications. Epidemics of mumps were seen to occur at intervals of 2–5 years. Seasonal variation occurs in temperate climates with highest incidence in February and March. The overall annual incidence in Europe in the pre-vaccine era (1977–1985) was 290 per 100,000 of the population. The seroprevalence of mumps IgG in a nonimmunized population of adults can be greater than 90%. Clinically apparent mumps infection following a previous infection may occur though an accurate history of previous infection can be difficult to establish. It is generally considered that wild-type infection gives rise to a lifelong immunity against the disease.

Transmission and Tissue Tropism

The virus replicates in the epithelium of the upper respiratory tract and the salivary glands and is transmitted in salivary droplets. Patients are infectious from 3 days before until approximately 4 days after the onset of clinical symptoms. Mumps can also cause viruria but this is not considered important in transmission. Transmission has been demonstrated by a direct contact with upper respiratory tract secretions and droplets and only occurs in the acute phase. It has been proposed that primary viraemia occurs following the replication of the virus in local lymphoid tissues. Viraemia may lead to subsequent replication in other target organs including the kidneys, thyroid glands and sexual organs. Salivary glands are commonly involved but not essential to pathogenesis.

Clinical Features of Primary Infection

A prodromal phase is usually followed by clinical mumps symptoms within 24–48 h which include low-grade fever, headache, malaise, myalgia and anorexia. The median incubation period is 19 days and can vary from 14 to 25 days. Infectivity is usually present from 48 h prior to symptoms until several days after onset. Asymptomatic infection occurs more frequently in adults with an estimated subclinical infection rate of 15%–20%.

Systemic Manifestations

Parotid Glands

The predominance of parotitis ([Fig. 4](#)) has ensured that other clinical features are commonly described as complications rather than systemic manifestations. Recovery from parotitis is usually swift and complete. Parotid swelling can obscure the angle of the mandible and the orifice of Stensen's duct can become erythematous and edematous. Lymphatic obstruction may rarely lead to presternal and supraglottic edema.

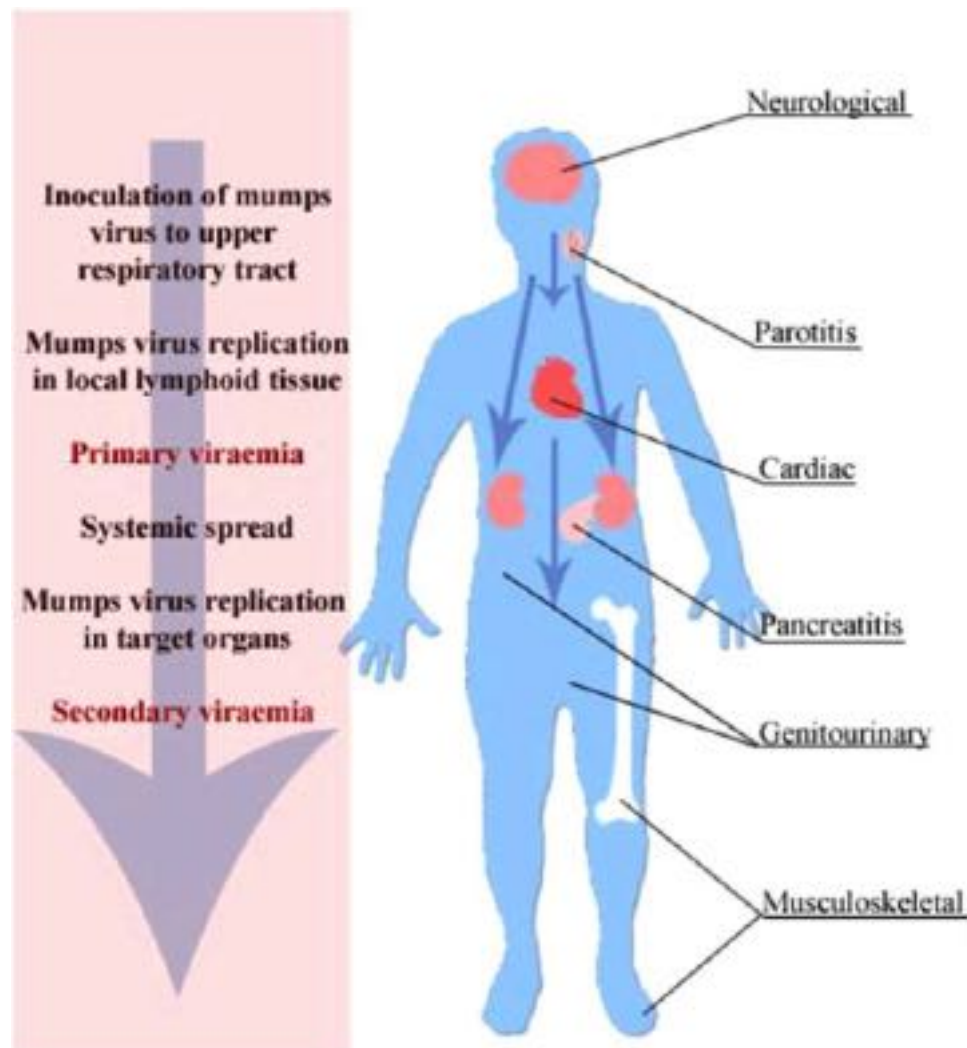


Fig. 4 Pathogenesis and systemic manifestations of MuV infection. *Parotitis*- Parotid swelling occurs in as many as 95% of individuals with symptoms; a further 75% of these subsequently develop swelling in the contra-lateral gland. *Neurological*- Aseptic meningitis is the most frequent of these manifestations occurring in as many as 1%–10% of infections. Encephalitis occurred in approximately 1 in 6000 patients that presented in the pre-vaccine era. Transient sensorineural deafness has been reported in adult military personnel whilst permanent deafness may occur in up to 1 in 15,000 cases of mumps. *Pancreatitis*- Pancreatitis has been reported in up to 4% of cases presenting to hospital. *Genitourinary*- Orchitis can occur in as many as 30% of post-pubertal males presenting with symptomatic mumps and close to 40% of these can develop bilateral swelling. Epididymitis is associated with 85% of cases of orchitis. Oophoritis is present in comparatively fewer post-pubertal females at a rate up to 7% with no subsequent infertility reported. Transient nephritis is relatively frequent though few cases go on to develop severe nephritis and thrombocytopenia. *Cardiac*- Myocarditis has been reported in 4%–15% of patients presenting to hospital with mumps though ECG findings may reflect symptoms of an acute infection. *Musculoskeletal*- Young male adults make up the largest proportion to suffer from monoarticular or polyarticular mumps arthropathy.

Neurological Disease

Aseptic meningitis has been reported as up to 3 times more frequently in men than in women and occurs in the presence or absence of parotitis. It is usually benign with no permanent neurological damage. Before the development of the currently used live-attenuated vaccine, MuV was the most common cause of viral meningitis and encephalitis in the USA. Encephalitis usually has a favorable prognosis, although minor neurological changes, such as learning and concentration impairments and sudden deafness are well-documented sequelae. Other neurological syndromes proposed to be associated with mumps included cerebellar ataxia, facial palsy, Guillain-Barré syndrome, hydrocephalus, poliomyelitis-like syndrome and transverse myelitis.

Pancreas

The clinical course of pancreatitis due to mumps virus is typically benign though cases of operable pseudocysts have been noted.

Genitourinary Tract

Orchitis occurs more frequently in post-pubertal males than oophoritis in post-pubertal females (Fig. 4). Late atrophy of testicular tissue may lead to infertility in approximately 17% cases of epididymitis and rarely sterility ensues. The risk of miscarriage is considered to increase in pregnant mothers with mumps infection but there is no evidence of the disease causing congenital malformations.

Cardiac Symptoms

The causative role of mumps virus in myocarditis is uncertain (Fig. 4). Fulminant fatal myocarditis with dilated cardiomyopathy has been described in rare instances.

Musculoskeletal System

Monoarticular and polyarticular mumps arthropathies have been described (Fig. 4).

Other Symptoms

Other organs or glands that may be affected include the eyes, gall bladder, lacrimal glands, liver and thyroid gland. Mumps virus has also been proposed to be associated with diabetes mellitus, haemophagocytic syndrome and thrombocytopenia.

Pathogenesis

The virus replicates primarily in the nasal mucosa and epithelium of the upper respiratory tract. After penetrating the draining lymph nodes, a transient viremia occurs and thereafter various target organs such as the salivary glands, the kidneys, pancreas, and the CNS are infected. Salivary glands are commonly involved but not essential to pathogenesis. Infection of the salivary glands produces parotitis, the most predominant clinical feature of the disease. Viral replication leads to tissue damage and the subsequent immune response causes inflammation and swelling of the gland. Dissemination into the kidneys can lead to a prolonged infection of this organ and viruria. Virus can be isolated from throat swabs, blood, saliva and urine. Involvement of the CNS may be as high as 50% of cases and parotitis is not required for this to occur. MuV can be readily detected in the cerebrospinal fluid (CSF) in cases of meningoencephalitis.

Pathological changes initiated by mumps virus are nonspecific. In the parotids, diffuse interstitial edema and serofibrinous exudate consisting predominately of polymorphonuclear leukocytes (PML) have been observed in the connective tissue and ducts. Degeneration of ductal epithelium can be seen as a result of accumulation of neutrophils and necrotic debris in the lumen. Involvement of the pancreas or testis may lead to a more marked interstitial hemorrhage and PML accumulation. This can lead to vascular compromise and eventual epithelial atrophy, hyalinization and fibrosis.

Mumps virus infection can induce the production of interferons and anti-mumps IgA, IgG and IgM antibodies as well as the activation of cell-mediated immunity. A viruria can frequently be detected though it is unclear whether this is due to a local replication of the virus in the kidneys or whether it is of haematogenous origin. Central nervous system (CNS) involvement occurs more frequently in the meninges with electroencephalograms revealing little deviation from normal. Although rare, mumps encephalitis with perivenous demyelination and microglial cell increase has been observed. Conversely, the death of neurons without demyelination has been demonstrated in an apparent case of primary mumps encephalitis. The contribution or extent of immune response-related damage to CNS in mumps encephalitis has been difficult to estimate.

Immune Response

It is not known whether the humoral or cell-mediated immune (CMI) response is the most important factor in the clearance of MuV infection. The correlates of protection in terms of antibody titer or CMI responses have not been determined. Both play a role although neither seems to be required exclusively for a successful control of the infection. Eleven days after infection, the humoral immune response is well established (Fig. 5) and the presence of neutralizing anti-MuV antibodies probably terminates viremia. Similarly, the appearance of IgA in saliva stops excretion of infectious MuV. The precise time at which MuV reactive cytotoxic T lymphocytes appear is unclear, although these have been demonstrated in both the blood from patients with the natural infection and in vaccinated individuals.

The magnitude and effectiveness of cell-mediated immune response may be related to the genetic human leukocyte antigen (HLA) background of the host. A complication in the development of the CMI response is the tropism of the virus for T and B cells. A reduction in CMI responses to antigens previously recognized has been observed, although the mechanism is unclear. This response is less severe and of shorter duration than the immunosuppression associated with a measles virus infection. The virus grows well in activated but not in resting T lymphocytes and infection of the CNS during MuV infection

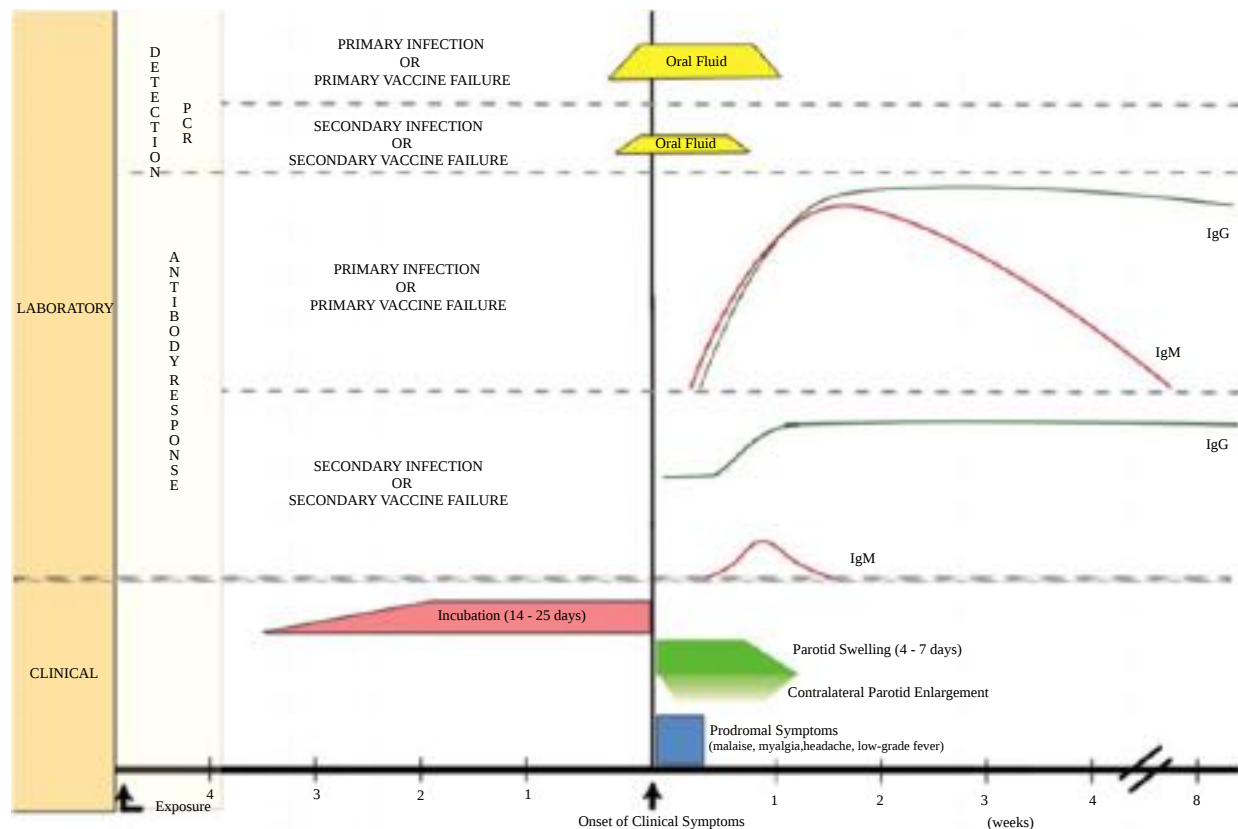


Fig. 5 The laboratory diagnosis of mumps virus infection. The clinical course of mumps virus infection and laboratory evidence to support a diagnosis of mumps through detection of viral RNA by PCR (oral fluid) or antibody response in a single (positive anti-mumps IgM antibodies) serum specimen or an increase in anti-mumps IgG antibodies between paired serum specimens.

is thought to occur via transfer of infected monocytes into the choroid plexus and the meninges. Perinatal exposure of the fetus to MuV via the placenta does not appear to lead to infection of the fetus. However, congenital infection can give rise to anomalous immune responses to MuV in the newborn child. In these children, CMI but no humoral immune responses are detectable. Neutralizing B cell epitopes have been defined in the HN glycoprotein of MuV but no other B or T cell epitopes have been detected.

Differential Diagnosis

The diagnosis of mumps is more readily suspected clinically with classical swelling of the parotid(s). Other infectious causes of parotitis comprise viral and bacterial infections including *Staphylococcus aureus* and atypical mycobacteria, though rarely described. A viral cause was identified by serologic testing in 14% (84/601) cases of mumps-like illness in children who were tested mumps-negative. The main viral aetiologies were Epstein-Barr virus (7%), parainfluenza virus types 1, 2 or 3 (4%) and adenovirus (3%). Other viruses that have implicated in parotitis include coxsackieviruses, HHV6B and influenza A viruses. Non-infectious causes of enlarged parotids include trauma, calculi, tumors, Sjögren syndrome and sarcoidosis. Symptomatic treatment of virus-related parotitis include thiazide diuretics, phenylbutazone and phenothiazines. In the absence of parotitis, symptoms or signs of mumps systemic manifestations may predominate and specific diagnosis can be obtained with laboratory analyzes.

Laboratory Diagnosis

The laboratory evidence to support a diagnosis of mumps commonly comprises detection of viral nucleic acids (RNA) and/or anti-MuV antibody response (IgM or significant rise in IgG titers) (**Fig. 5**). Serum and oral fluid (OF) specimens are routinely used for the determination of mumps virus antibody response. The advantages of OF rather than blood collection include accessibility, reduced hazard, cost and convenience. Mumps virus can usually be detected in OF 2 days prior to 5 days after the onset of clinical

symptoms by culture, immunofluorescence (IF) or polymerase chain reaction (PCR). Mumps virus may also be detected in urine during the first 2 weeks of mumps infection.

Mumps virus is present in CSF in the first 3 days and up to the 6th day of clinically apparent mumps meningitis/encephalitis. A lumbar puncture taken in patients with mumps meningo-encephalitis commonly reveals an elevated white blood cell count and normal or weakly elevated protein. A low glucose concentration in the CSF has been measured in up to 30% of patients which is more frequent than other viral meningitides. PML predominance in CSF has also been seen in approximately 20%–25% of patients. Increased CSF: serum anti-MuV IgG antibody ratios (e.g., > 1:50) may be indicative of an active CNS infection.

Detection of Mumps Virus

Virus culture has been considered the gold standard for mumps confirmation though it is time and resource intensive and presently it is not widely used any more. Mumps virus RNA can be detected by various nested reverse transcriptase PCR (RT-n-PCR) assays. In addition, real time ((RT)-PCR) has been used to detect MuV RNA in OF, CSF, respiratory and urine samples. The advantages of (RT)-PCR include greater sensitivity and specificity as well a significant reduction in the likelihood of contamination. Furthermore, quantitative real-time (RT)-PCR is increasingly being used in investigations of mumps infection to quantify viral load.

MuV RNA has been detected in patients with mumps infection and healthy children following immunization live mumps vaccines. A greater sensitivity of (RT)-PCR versus cell culture has been demonstrated in OF for mumps infection and in CSF for clinically apparent mumps CNS disease. However, a lower sensitivity of (RT)-PCR relative to cell culture has been observed in urine specimens and may be attributable to inhibition. SH gene sequencing has shown geographically distinct strains, and can distinguish between wild virus and vaccine strains.

Antibody Response

Standard enzyme-labeled immunoassays (ELISAs) are the preferred method for the detection of mumps virus IgM and IgG antibodies. The efficiency and lower complexity of the enzyme immunoassay (EIA) method to measure and quantify mumps virus IgM and IgG has ensured its predominant use over of a variety of other techniques.

There are capture or indirect assays to detect mumps virus-specific antibodies that use recombinant mumps nucleoprotein (rMuVN) as an antigen. Mumps virus IgM antibodies has been detected in CSF from patients with clinically apparent mumps meningitis but the sensitivity of this detection method has remained at a level of ca. 50%.

A significant rise in IgG antibody levels between an acute and convalescent sample can diagnose primary infection and the level of significant and diagnostic rise depends on the assay. Mumps virus IgG detection by EIA is not an international standard for protective immunity following infection or vaccination. A residual low-level serum IgG antibody level in adulthood following immunization with MMR vaccine as a child does not correlate with protection against mumps infection.

The measurement of mumps virus IgG avidity has been proposed to assist in the differentiation of primary and secondary mumps virus infection. The strength of binding of mumps virus IgG to antigen has been shown to increase with time following primary mumps infection. Commercially available assays have been used to evaluate primary and secondary vaccine failure during a mumps outbreak. Mumps virus IgG may be of low or borderline avidity in vaccinated individuals with mumps virus infection. There are currently no commercial mumps virus IgG avidity assays.

Neutralisation Tests

The plaque reduction neutralisation test (PRNT) and focus reduction neutralisation test (FRNT) are the most specific assays for measuring neutralizing antibodies against mumps virus. PRNT has been shown to detect antibodies below the level of detection of IgG binding antibody assays. FRNT has been validated against PRNT and the method is more accessible, faster, utilizes lower serum volumes and consumes less reagents than PRNT. A protective antibody level with PRNT or FRNT has not been determined. A recombinant mumps virus GFP expression inhibition assay has been described with ca. 95% correlation with a conventional CPE-based neutralisation assay. Ultimately, it is not clear whether a detectable level of mumps neutralizing antibody confers protection from natural infection.

Utilization of Assays

Primary Infection and Primary Vaccine Failure

The presence of mumps virus IgM provides laboratory evidence to support a diagnosis of primary mumps infection or primary vaccine failure. A documented history of mumps virus vaccination is necessary to distinguish a primary vaccine failure from a primary infection. Mumps virus IgM is most reliably detected 1 week after symptom onset, declines within 1 month and persists at

low levels for a minimum of 4 months. However, mumps virus IgM can also be detected with a secondary infection or a secondary vaccine failure. Moreover, the absence of mumps virus IgM in acute illness does not exclude an infection.

Mumps virus IgG seroconversion is consistent with primary infection although dependent upon collection of acute and convalescent samples with no established standard rise in IgG titer. The detection of mumps virus RNA in high levels is consistent with primary infection or primary vaccine failure although the levels are dependent on sample timing.

Secondary Infection and Secondary Vaccine Failure

Mumps virus IgM is usually present in low levels or absent in secondary infection or secondary vaccine failure. Mumps virus RNA detection is predominately detected with lower viral RNA amounts than in primary infection at an equivalent clinical time point. A documented history of mumps virus vaccination is necessary to distinguish secondary vaccine failure from secondary infection.

Prevention and Control

Adaptation of MuV to growth in embryonated eggs and chicken embryo fibroblasts allowed the development of live-attenuated vaccines. In 1946, Enders observed that adaptation of MuV to chicken cells was associated with the loss of virulence in monkeys. In the past, killed virus preparations were used as vaccines in humans, although these vaccines did not lead to a life-long protective immunity and their use has been discontinued.

Mumps virus vaccines are available as a monovalent vaccine or in combination with other vaccines such as the measles, mumps and rubella (MMR) with or without varicella (MMRV) vaccines. MMR vaccine is a live attenuated vaccine used commonly worldwide. The Jeryl-Lynne (JL) strain of MuV is most frequently used in the MMR vaccine and interestingly it has been shown to be comprised of two strains, the major (JL5) and the minor (JL2) components. JL vaccine has been in global use since the 1970s and an initial field-randomized controlled clinical trial reported an efficacy exceeding 95% with the monovalent vaccine. A low complication rate and subsequent eradication of mumps infection has been reported following its introduction in some countries. However, there is now evidence for the efficacy of JL mumps vaccine component to be as low as 62% during some epidemics.

There are 4 other major vaccine strains (Urabe, Leningrad-3, Leningrad-Zagreb and Rubini). There have been demonstrable increases in central nervous system (CNS) complications that have limited the use of the Urabe strain in the vaccine. However, the Urabe strain has, however, shown greater immunogenicity in inducing antibody responses (97%) compared to Jeryl Lynn strain (90%). The Leningrad-Zagreb strain is an attenuated form of the Leningrad-3 strain. The use of this strain as a vaccine has been limited due to complications including aseptic meningitis although this complication has not been found in other studies. The Rubini strain, when introduced, was shown to not achieve sufficient levels of protection from wild type virus and it is not recommended by the World Health Organization.

Mumps vaccination has substantially reduced the incidence of mumps infection worldwide. After successful licensing of the vaccine in 1967 in the USA, the incidence of mumps dropped from 76 to 2 cases per 100,000 population. A similar decrease in mumps incidence was found in many countries after introduction of the MMR vaccine. Clinically apparent mumps infection can occur following a previous immunization, often 10 or more years later, and outbreaks of mumps in teenagers and young adults are common in many countries with a childhood vaccination program. Mumps symptoms in these cases are milder, short-lived and rarely are complications other than transient parotitis reported.

Future Perspectives

Over the last 10 years, the development of reverse genetics systems for all members of the *Mononegavirales* has had a significant impact on our understanding of the pathogenesis of these viruses. Reverse genetic systems have been generated for PIV5, hPIV2, and MuV, and to date they have been used to begin to define the structural and functional relationships and the roles that various proteins play in attenuating these viruses. The opportunity to examine the contribution of individual proteins from neurovirulent strains, such as Urabe, to neuropathogenesis in animal models should help to identify virulence determinants, something which is important if it becomes necessary to develop new and safer MuV vaccines.

One of the key challenges for the future is to develop safe, non-neurovirulent vaccines that induce a long-lasting protective immune response. Correlates of protection are still not well understood, but the increasing number of mumps outbreaks in vaccinated populations show the need for such vaccines.

Further Reading

- Rubin, S.A., Sauder, C., Carbone, K.M., 2013. Mumps virus. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Wolters Kluwer/ Lippincott Williams and Wilkins Oxford University Press, pp. 1024–1041.
- Rima, B.K., 2010. Mumps: Epidemic parotitis. In: Warrall, D., *et al.* (Eds.), *Oxford Textbook of Medicine*, fifth ed. Oxford University Press, pp. 513–515.

Murine Leukemia and Sarcoma Viruses (*Retroviridae*)

George R Young and Kate N Bishop, Francis Crick Institute, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Endogenous virus A retroviral genome integrated within that of its host and inherited in a Mendelian fashion.

Exogenous virus Infectious viral particles transmitted between individuals or causing spreading infection within an individual.

Inbred mice Genetically invariant strains of mice developed through inbreeding programs.

Integration The process or action of a retrovirus inserting its reverse transcribed genome into the DNA of an infected cell.

Provirus A viral genome integrated within the DNA of an infected cell (not synonymous with “endogenous virus” as infections of somatic cells will not be vertically inherited).

Viral core The assembled structure in the center of the virion encompassing the capsid protein shell, viral nucleic acids, and enzymes.

Classification

Murine Leukemia Viruses (MuLVs or MLVs) are members of genus *Gammaretrovirus* of the family *Retroviridae*, subfamily *Orthoretrovirinae*. As with other retroviruses, MLV virions contain two copies of a non-segmented positive-sense single-stranded RNA genome and fall into the larger *Ortervirales* order. As single-stranded RNA viruses replicating through a DNA intermediate, alternative classification places them within Baltimore Group VI. MLVs exist both as exogenous viruses and as endogenous retroviruses (ERVs) within the genome of their host.

History

As judged by the fossil record of endogenous examples within species of *Mus*, MLVs have been infecting mice for at least 1.5 million years (my), although other gammaretroviruses are likely to have been circulating within their ancestors for considerably longer. The first breeding of fancy mice and of early inbred lines, primarily from stocks of *Mus musculus domesticus* wild-caught in China, revealed a variety of neoplastic and leukemic pathologies that, thanks to detailed record keeping, were observed to have familial associations. The search for the etiology of such heritable cancers led to the creation of numerous modern laboratory mouse lines, many bred for their specific resistance or susceptibility to disease. Using these inbred lines, a viral cause of mammary carcinoma, Mouse Mammary Tumor Virus (MMTV), was identified by John Bittner in 1936. Although MLVs were also quickly identified by the breeding of strains with high incidences of leukemia in the 1920s and 30s, it was not until 1957 that the first MLV was isolated by Ludwik Gross. MLVs with different properties – generally either T lymphoma-inducing (e.g., Gross- and Moloney-MLV) or erythroleukemia-inducing (e.g., Friend- and Rauscher-MLV) – were subsequently isolated, several of which remain actively used in research.

The deliberate selection of inbred lines created great diversity in the complement of endogenous MLVs between strains and, depending on relatedness, only around 30% of proviral loci are shared between common inbred mouse lines. Selection for inbreds with altered characteristics or phenotypic differences in turn elucidated the various impacts of MLVs and other mobile elements in genomes; indeed, many of the first characteristics mapped to genomic loci were determined to result from disruption of gene function due to MLV insertions, including *dilute* (*d*) in *Myo5a*, *hairless* (*hr*) in *Hr*, which remains named for the phenotype, and *rodless retina* (*rd1*) in the *Pde6b* gene. Work with MLVs revealed the means of retroviral replication by reverse transcription and formed the basis for much of the field of Retrovirology, for example early serological classification of MLVs defined various “group specific antigens”, for which the *gag* gene is named.

Genome and Structure

Mature MLV virions are spherical enveloped particles typically ~100–120 nm in diameter with a classical C-type morphology defined by viral assembly at the plasma membrane and an electron-dense polyhedral core in electron micrographs (Fig. 1, top), although overall size and core structures are pleomorphic. Virions budding from the cell membrane are initially immature, formed from assembled Gag polyproteins, with an electron-lucent core (Fig. 1, bottom). Gag is then cleaved by the virally encoded protease to allow the multimerization of CA hexamers and pentamers into the mature CA lattice that surrounds the viral core.

The MLV genome defines the prototypic genetic structure of all simple retroviruses with three genes (5′-*gag-pol-env*-3′), along with the order of the proteins within the polyproteins encoded by these genes: *gag* encodes the Gag polyprotein, cleaved to release matrix, p12, capsid, and nucleocapsid (MA-p12-CA-NC), *pol* encodes the Pol polyprotein, cleaved to release protease, reverse transcriptase, and integrase (PR-RT-IN), and *env* encodes the Env polyprotein, cleaved into surface and transmembrane (SU-TM)

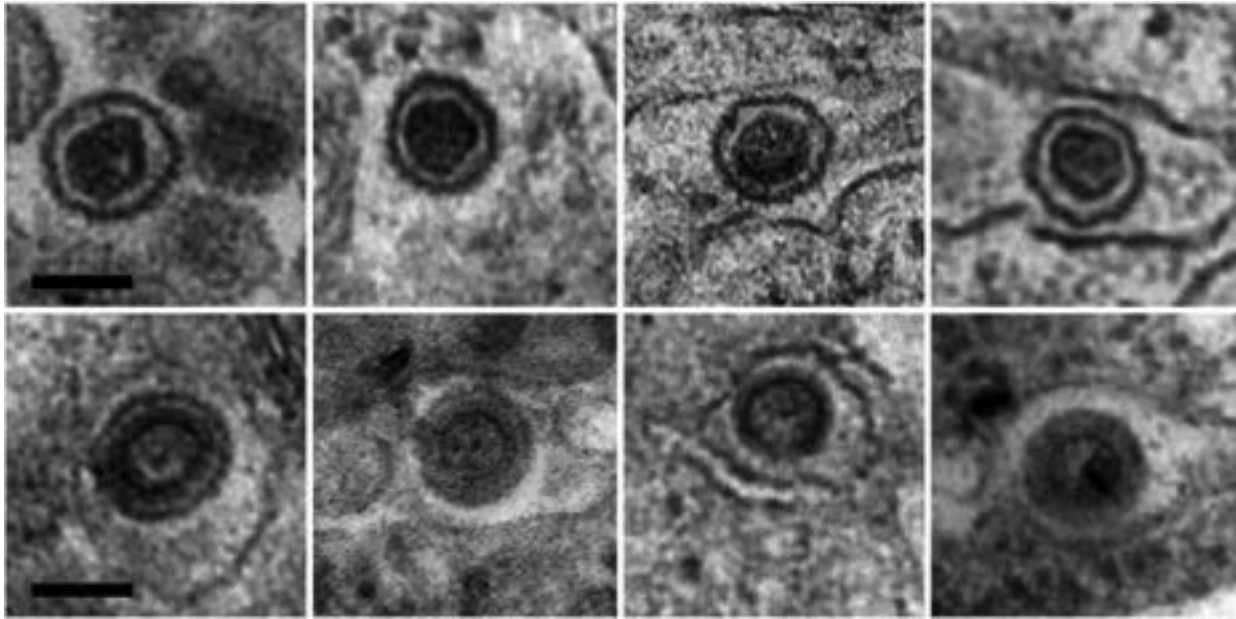


Fig. 1 Transmission electron micrographs of MLV particles within spleen and lymph node tumor sections following spontaneous disease development in immunodeficient C57BL/6 inbred mice. Top row, mature virions showing the electron-dense polyhedral core. Bottom row, immature virions showing the electron-lucent circular core. Images were acquired at 5000–12000 $\times$ magnification using a JEOL 100E $\times$ microscope and ORIUS 1000 CCD camera. Presented fields are standardized to 300 nm and scale bars show 100 nm.

(**Fig. 2**). Thus, products of *gag* contain all elements necessary for assembly and release of virions, *pol* the enzymes required for the process of retroviral replication, and *env* the structural elements required for target cell entry. The specific roles of these proteins are discussed in greater detail elsewhere.

Whilst *env* is expressed from a spliced RNA, the products of *gag* and *pol* are translated from an unspliced genomic transcript (**Fig. 2**). Initiation at the *Gag* N-terminus either gives rise to the *Gag* polyprotein or, due to read-through of the in-frame UAG (amber) stop codon, a *Gag-Pol* polyprotein. This differentiates MLVs from other retroviral genera, which typically produce the *Pol* polyprotein through ribosomal frameshifting. The efficiency of the read-through thus also dictates the ratio of *Gag*: *Pol* gene products, at around 20:1.

Certain MLVs also produce a longer, glycosylated, form of the *Gag* polyprotein (glyco-*Gag* or *gGag*), initiated upstream of the canonical start codon, which is not cleaved to yield the structural *Gag* proteins (*gPr80*, **Fig. 2**). *gGag* is transported to the cell membrane and shed from infected cells and, whilst *gGag* deficient viruses replicate well *in vitro*, such mutants exhibit reduced *in vivo* fitness. The role of *gGag* has yet to be fully elucidated but it has been proposed to overcome restriction factors including murine *Apobec3* and *Serinc5* (see below).

Although all retroviruses express *MA*, *CA* and *NC*, different viruses release different additional *Gag*-cleavage products. The functions of these proteins are less well characterized and they are usually named for their apparent molecular weight. MLV, like many other genera, releases a protein from between *MA* and *CA*, called *p12* (**Fig. 2**). This protein contains a four amino-acid motif referred to as the Late (L)-domain which recruits components of the cellular endosomal sorting complexes required for transport (ESCRT) machinery required for immature particles to bud from the plasma membrane. It also contains two functional domains that are important for viral infectivity. The N-terminal functional domain binds to the *CA* protein and stabilizes the *CA* lattice whilst the C-terminal domain interacts directly with host chromatin facilitating subsequent integration.

In the integrated, proviral form, the long terminal repeats (LTRs) flank the viral genome and contain the regulatory elements necessary for expression within the U3 region (**Fig. 2**). The LTR's complement of transcription factor binding sites dictate spatial and temporal expression patterns and, hence, too the mode of disease in MLV infection. Exchanging the LTR of Moloney-MLV for that of Friend-MLV, for example, switches its mode from T lymphomagenic to erythroleukemogenic, and *vice versa*.

Replication

The process of retroviral replication is discussed in detail elsewhere, yet MLVs differ from other retroviruses in certain key aspects. Significantly, where many retroviruses, such as HIV-1, possess mechanisms to gain access to the cell nucleus, allowing cell-cycle-independent infection, MLVs are incapable of infecting non-dividing cells. If target cells are arrested in S phase or at the G2/M boundary, although viral entry, reverse transcription, and formation of the pre-integration complex (PIC) can complete at normal rates, integration cannot occur, even after extended periods. Thus, MLV infection is believed to require dissolution of the nuclear envelope in order for the PIC to gain access to the host chromosomes and, accordingly, successful infection is highly dependent on the cycling state of target cells.

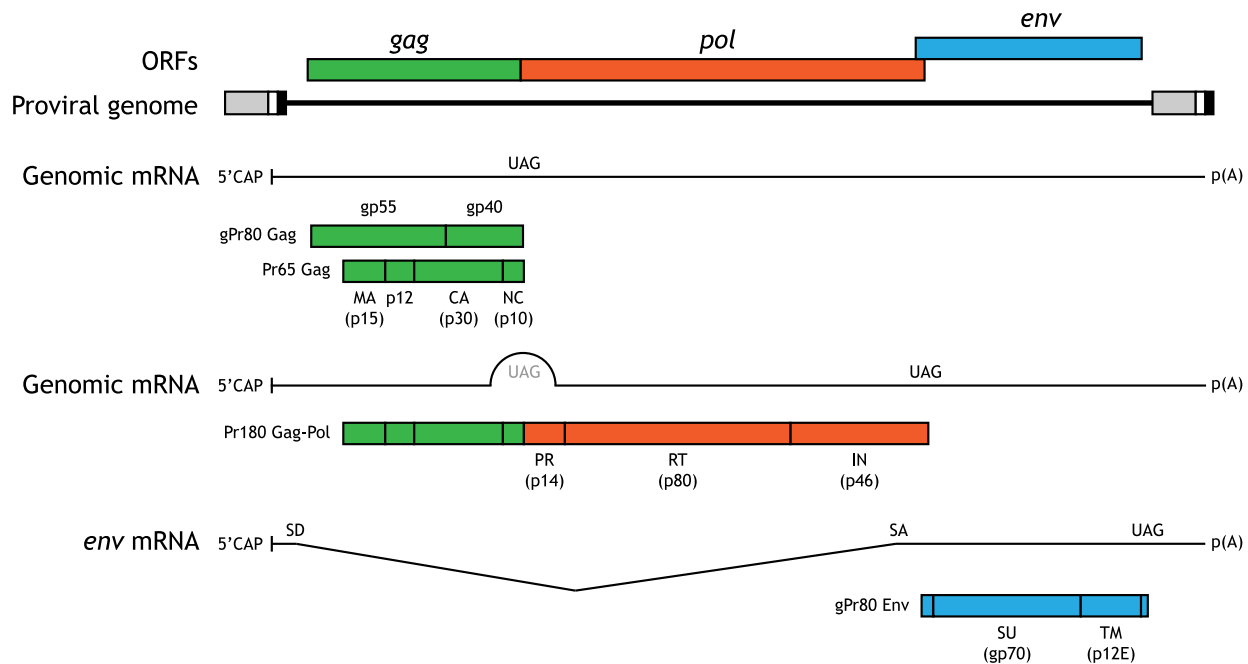


Fig. 2 To-scale schematic representation of the MLV genome detailing the protein products of the *gag*, *pol*, and *env* genes, their apparent molecular weights, and points of enzymatic cleavage. The proviral DNA genome includes long terminal repeats at either end (composed of U3 – gray, R – white, and U5 – black) that both result from and allow for reverse transcription. Genomic mRNAs may be translated to yield only the *gag* gene products (green) through termination at the first UAG stop codon or, alternatively, read-through of this codon allows the translation of *pol* gene products (orange) in the form of a Gag-Pol precursor. The *env* gene products (blue) are produced from a spliced mRNA with the indicated splice donor (SD) and splice acceptor (SA) points. All mRNAs have a 5' 7mG cap (5' CAP) and are polyadenylated (p(A)) at the 3' end. See main text for protein name two-letter abbreviations.

Once inside the cell nucleus, mechanisms of integration differ between retroviruses, resulting in varying target site selection preferences clearly visible with high-throughput sequencing. MLVs display a strong preference for integration within regions of active transcription and, specifically, for areas surrounding promoter-proximal CpG islands and transcription start sites. This differs significantly from HIV-1, where integration within transcriptional units (regions spanning from points of transcription initiation to the corresponding points of termination) is favored, but no specific enrichment is seen near transcription start sites and small reverse-associations are visible for CpG islands. Integration site selection in MLVs is mediated through IN binding host bromo-domain and extraterminal (BET) proteins.

Host-Virus Interactions

Host-virus interactions can be divided into the innate and the adaptive arms. Methods of innate control or resistance to MLV infection were some of the first subjects to be studied with inbred mice and the identification of *Friend Virus Susceptibility 1* (*Fv1*) in 1970 and its positional cloning in 1996 initiated the study of host “restriction factors”. Common laboratory inbred mice express either *Fv1^b* or *Fv1ⁿ*, each allele conferring resistance to virus of the opposing N- or B-tropism, determined by differences at residue 110 of the MLV CA. The range of *Fv1*-based restriction is now known to extend beyond the gammaretroviruses and its evolution and continued positive selection has occurred in tandem with the speciation of the Myodonta, over at least 45 my. *Fv1* derives from the *gag* gene of an ERV and likely blocks infection through a process involving its scaffolding over incoming retroviral cores. A more common co-option of ERVs is seen in the *Fv4*, *Rmcf*, and *Rmcf2* genes, each representing an MLV envelope glycoprotein (of ecotropic, polytropic, and xenotropic origins, respectively, see below) whose expression acts to prevent superinfection through interaction with the cognate cellular receptors.

As well as co-opting retroviral proteins in host defense, a number of other host genes play significant roles as restriction factors. The murine *Apobec3* cytidine deaminase inhibits MLV infection through both G-to-A hypermutation of the viral genome during reverse transcription and deaminase-independent mechanisms. Similarly to *Fv1*, *Apobec3* exhibits significant positive selection and different mouse lines possess different alleles. The allele of C57BL/6 has particularly strong antiviral activity due to the recent capture of an MLV LTR, acting as an enhancer element for increased expression, as well as loss of exon 5, which increases protein synthesis. Studies of the mutational profile of integrated proviruses reveals that hypermutation is a substantial burden to infecting viruses, both *in vitro* and *in vivo*. Another host factor, Tetherin (*Bst2*), acts to prevent virion release from the cell surface and can have significant impacts in *in vivo* infection settings.

Adult mice typically mount robust humoral responses to immunogenic epitopes within *env*, which are capable of inhibiting cell-to-cell spread of virus by neutralization and elimination. Both CD4⁺ and CD8⁺ T cells additionally respond to both Env and

Gag determinants and mediate direct killing, which has been shown to be relevant in limiting viral spread before the humoral response is mounted. After an initial period of viremia, infections within adults are typically controlled within 14–21 days, as exemplified by the Friend-MLV model system. Conversely, infection of neonates results in immunotolerance to viral antigens, persistent infection, and high levels of viremia. This is mediated through thymic deletion of Env-reactive T cells, rather than through B cell tolerization, and experimental introduction of T cell help allows infected mice to mount effective humoral responses. Although not causing complete tolerance, there is also evidence that background expression of MLV ERVs comparably shapes the repertoire of T cells able to respond to exogenous virus.

Pathology: Murine Leukemia Viruses

MLVs can be subdivided into different groups according to the host range of their envelopes. Ecotropic MLVs (eMLVs) exhibit tropism for murine cells but not for cells of other species and utilize the cationic amino-acid transporter mCat1 (*Slc7a1*) as a receptor. eMLVs split into three groupings phylogenetically: AKV-type, CasBrE-type, and HoMuLV-type, the latter two having only been isolated from wild mice. Xenotropic MLVs (xMLVs) exhibit tropism for non-murine cells, using the inorganic phosphate exporter Xpr1 (*Xpr1*) as a receptor. Polytopic and modified polytopic MLVs (p/mpMLVs), differentiated by variation within their LTRs, also utilize Xpr1, but can infect both murine and non-murine cells. Amphotropic MLV (aMLVs) exhibit the same wide host range as p/mpMLVs but utilize one or both of the phosphate transporter Pit1 (*Slc20a1*) and Pit2 (*Slc20a2*) paralogues as receptors. These, although having only been identified as pathogens of wild mice, bear strong phylogenetic similarities to several model MLVs, including both Friend- and Moloney-MLV.

The apparent historic escape of mXpr1 from xMLV envelope underlines the likely extended periods of receptor-envelope coevolution and suggest a certain plasticity in receptor usage over time. Indeed, endogenous MLVs with ecotropic host ranges have been identified within wild mice that use Smit1 (*Slc5a3*) in place of mCat1, as well as polytopic viruses using Pllp in place of Xpr1. Nevertheless, the global prevalence of retroviral pathogens of wild mice, including of MLVs, is largely unstudied. Indeed, whilst MLVs may be considered the primary burden, the evolution of *Fv1* would suggest that both MMTVs and as yet unidentified lentiviruses circulate within certain species of *Mus*.

In wild mice, exogenous MLVs are of low pathogenicity and disease occurs after long latency periods (these viruses are frequently described as “slowly transforming”, therefore), allowing for efficient horizontal transmission. Lesser routes of infection may be venereal, perinatal, transplacental, or through saliva but, primarily, transmission occurs through milk, from infected dams to nursed pups and the fostering of pups born of infected dams eliminates transmission. Whereas infection of adults rarely results in leukemogenesis for most viruses, neonatal infection frequently results in disease after 2–18 months, commonly in B/pre-B cells and visible within the spleen. Subsequent to initial viremia, a general requirement for MLV leukemogenesis is the spontaneous generation of recombinants, historically named mink cell focus-inducing (MCF) viruses due to their acquisition of a polytopic host range. MCF viruses form through individual or sequential recombination events with endogenous MLVs and are typified by wholly replaced or partially-recombinant *env* genes. Replicating at high titers, MCF viruses cause oncogenesis through random insertional mutagenesis – insertion near to, and the resulting dysregulation of, a cellular proto-oncogene (*c-onc*) – and it is common that resulting tumors are infected solely by MCF viruses, rather than with the initially-transmitted MLV. In the absence of MCF recombinants, oncogenesis may still occur, albeit at a reduced rate and with extended latency periods.

Pathology: Murine Sarcoma Viruses

Early passage experiments with MLVs gave rise to occasional instances of increased pathogenicity and of altered disease characteristics – neoplastic rather than leukemogenic modes of disease. Named Murine Sarcoma Viruses (MuSVs or MSVs), these viruses cause disease after far shorter latency periods than for MLVs and are hence described as “strongly/acutely transforming”. MSVs, while derived from MLVs, acquire cellular oncogenes (termed v-onc genes when carried by the virus) at the expense of viral sequence and are, hence, replication incompetent, requiring complementation from a helper virus for their spread. Exapted v-onc genes are capable of inducing rapid transformation. Examples include the HRas GTPase (Harvey-MSV *v-ras*, controlling cell proliferation, differentiation, adhesion, and migration), Abl1 tyrosine kinase (Abelson-MSV *v-abl*, controlling cell division, differentiation, and stress responses), and the transcription factor c-Fos (Finkel-Biskis-Jenkins-MSV *v-fos*, pairing with c-Jun to form the AP-1 complex, controlling genes influencing cell proliferation, survival, and polarity).

All known MSVs have been derived from experimental laboratory processes including the repeat passage of viral stocks or tumor extracts through susceptible mice, repeated neonatal inoculation, passage through alternate species, and host irradiation. As such, MSVs are experimental phenomena and are not believed to be a characteristic of disease within wild mice.

MLV Vectors

The study of MSVs showed that it was possible to complement defective particles *in trans*. This allowed for designed, defective, genomes to be packaged in the same way and for stable integration of vectored sequences. MLV-based retroviral vector systems are

now routinely used in research laboratories as a way of expressing introduced genes. These systems are single cycle, meaning that no new viral particles are produced following reverse transcription and integration of the encapsidated RNA. Such retroviral vectors have also been used in the clinic as gene therapy vectors for several immunodeficiency disorders. However, insertional mutagenesis resulted in malignancy in some cases and other systems may provide improved safety features.

Endogenous MLVs

Electron micrographs of both healthy and neoplastic tissues frequently reveal C-type retroviral particles as incidental observations and, indeed, common inbred mouse lines harbor between 50 and 80 complete MLV proviruses within their genomes, with a further number of intact *env*-deficient integrations, fragmented proviruses, and solo LTRs. Xenotropic, polytropic, and modified polytropic MLV, termed *Xmv*, *Pmv*, and *Mpmv* as endogenous loci, constitute the majority of endogenous MLV, with ecotropic MLV (*Emv* loci) being found in low numbers and in comparatively few inbred mice by comparison, for example the single *Emv2* locus of C57BL/6 mice.

The majority of endogenous MLVs are replication incompetent through point mutation, insertion, or deletion, although certain high-leukemia-incidence inbred mice (AKR and C58) notably carry infectious *Emv* loci. MLV loci are specifically targeted for epigenetic silencing through Trim28/Setdb1 by KRAB zinc-finger proteins (for example by Zfp809, which targets MLVs utilizing a reverse transcriptase proline-tRNA primer binding site (PBS^{Pro})) and, hence, suppression can be reversed by treatment with demethylating agents, including BrdU and 5-AzaC. Potentially as a result of gradual changes in the epigenetic landscape accumulated with time, expression of endogenous MLVs increases in aged mice, as well as varying between tissues as a result of differing milieus of transcription factors. Accumulation of MLV products can induce progressive autoimmune conditions, including spontaneous vasculitis and glomerulonephritis. Expression of select proviruses can also be induced by stimulation of cells with TLR agonists, including LPS, Poly(I:C), and Pam₃CSK₄. In conditions of immunodeficiency, activation of endogenous MLVs in this manner by products of commensal flora can result in their expression and recombinational repair, formation of MCF recombinants, and subsequent oncogenesis.

In 2006, a new gammaretrovirus was isolated from prostate cancer patient tissue. Named xenotropic murine leukemia virus-related virus (XMRV), this was potentially the third class of retrovirus to be pathogenic in humans. In 2009 XMRV was also linked to chronic fatigue syndrome. However, subsequent studies failed to find an association of XMRV with disease and, in most cases, failed to find the virus in human samples. In 2011, work from multiple laboratories revealed that XMRV was derived from recombination between two murine ERVs during serial passage of a human prostate tumor in nude mice to make the 22Rv1 and CWR-R1 cell lines in the mid-1990s. Indeed, it is likely that MLVs contaminate other cell lines historically passaged through mice. These studies separately highlighted that certain MLV RT-utilizing commercial reverse transcription kits harbor low levels of MLV DNA contamination.

See also: Avian Leukosis and Sarcoma Viruses (*Retroviridae*). Fish Retroviruses (*Retroviridae*). Jaagsiekte Sheep Retrovirus (*Retroviridae*). Structure of Retrovirus Particles (*Retroviridae*). The Role of Retroviruses in Cellular Evolution

Further Reading

- Coffin, J.M., Hughes, S.H., Varmus, H.E. (Eds.), 1997. *Retroviruses*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Groom, H.C., Bishop, K.N., 2012. The tale of xenotropic murine leukemia virus-related virus. *Journal of General Virology* 93, 915–924.
- Kozak, C.A., 2015. Origins of the endogenous and infectious laboratory mouse gammaretroviruses. *Viruses* 7, 1–26.
- Mukerjee, S., Thrasher, A.J., 2013. Gene therapy for PIDs: Progress, pitfalls and prospects. *Gene* 525, 174–181.
- Rein, A., 2011. Murine leukemia viruses: Objects and organisms. *Advances in Virology*. 403419.
- Skorski, M., Bamunusinghe, D., Liu, Q., Shaffer, E., Kozak, C.A., 2019. Distribution of endogenous gammaretroviruses and variants of the Fv1 restriction gene in individual mouse strains and strain subgroups. *PLoS One* 14, e0219576.
- Young, G.R., Yap, M.W., Michaux, J.R., Stepan, S.J., Stoye, J.P., 2018. Evolutionary journey of the retroviral restriction gene Fv1. *Proceedings of the National Academy of Sciences of the United States of America* 115, 10130–10135.

Newcastle Disease Virus (*Paramyxoviridae*)

Ben Peeters and Guus Koch, Wageningen Bioveterinary Research, Lelystad, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AAvV-1 The suggested nomenclature for AAvV-1 isolates is derived from that used for avian influenza virus isolates. At first mention, the isolate should be fully listed using the format: AAvV-1/Species/Country (state, city or more specific location)/isolate number, name, or other unique

identifier/year of isolation (e.g.: AAvV-1/Chicken/USA/LaSota/1946). Subsequently, this may be appropriately abbreviated (LaSota/1946). Certain isolates, in particular those used for vaccine production, are so well known that they often are listed by one name (LaSota, B1, Ulster, Herts33, etc.).

Glossary

ICPI Intracerebral pathogenicity index

Lentogenic Avirulent

Mesogenic Moderately virulent

OIE World Organisation for Animal Health (Office International des Epizooties)

Velogenic Very virulent

Classification

The taxonomic name for Newcastle disease virus (NDV) has recently changed from Avian paramyxovirus 1 (APMV-1) to Avian Avulavirus 1 (AAvV-1). Together with 18 other avian avulaviruses (AAvV-2 to -19) AAvV-1 forms the genus Avulavirus within the family Paramyxoviridae in the order Mononegavirales.

Phylogenetic analysis of genome sequences is the standard procedure to characterize AAvV-1 strains. The current generally accepted phylogenetic classification is based on the complete nucleotide sequence of the fusion protein (F) gene. Over time, AAvV-1 has undergone considerable genetic and antigenic drift. This genetic diversity is represented by groups of genetically related viruses denoted as genotypes and sub-genotypes. A system to define genotypes based on a mean evolutionary distance of 10% between groups is used to define genotypes. An evolutionary distance of 3%–10% is used to define sub-genotypes. Furthermore, a (sub-) genotype must consist of at least 4 epidemiologically-independent isolates, and the bootstrap value at branch points must be at least 60 (using the Neighbor Joining and the Maximum likelihood methods, with standard errors being calculated based on 500 bootstrap replicates). Strains of AAvV-1 can be grouped into 2 main classes, class I and class II. Class I viruses are mostly low pathogenic (lentogenic) viruses found in wild birds and belong to a single genotype. Class II viruses are much more diverse and currently 18 genotypes (with some containing sub-genotypes) can be discriminated.

Virion Structure

AAvV-1 virions consist of enveloped pleomorphic particles ranging in size from approximately 100 nm to more than 500 nm in diameter. The majority of virus particles contains one functional genome although particles containing two or three genomes have also been observed. The F protein and hemagglutinin-neuraminidase protein (HN) are anchored in the virion envelope which is derived from the membrane of the infected cell. The F and HN proteins participate in attachment and entry, while HN is also involved in the release of progeny virus from infected cells. The matrix protein (M) is located immediately beneath the viral envelope. It maintains the shape of the virus and assists in the packaging of newly assembled viruses. Three proteins are associated with the single-stranded viral RNA genome. The nucleoprotein (NP) covers the entire viral RNA to form the ribonucleoprotein (RNP), which is the template for transcription and replication by the viral RNA-dependent RNA polymerase (large protein; L) in conjunction with the phosphoprotein (P) which acts as cofactor of the polymerase ([Fig. 1](#)).

Genome

The genome of AAvV-1 consists of a non-segmented single-stranded RNA of negative polarity with a size of 15,186, 15,192 or 15,198 nucleotides (nt). The variation in genome size is due to a 6 nt deletion/insertion in the intergenic region between the NP and P genes, or a 12 nt deletion/insertion in the P gene. The AAvV-1 genome contains six open reading frames (ORF) which encode for NP, P, M, F, HN and L. Two additional, non-structural proteins, V and W, are generated by RNA editing during P gene transcription. The 3' and 5' ends of the genome comprise the non-transcribed leader (55 nt) and

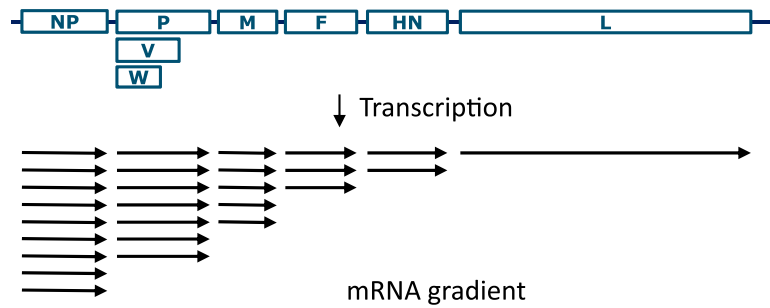


Fig. 1 AAVV-1 genome.

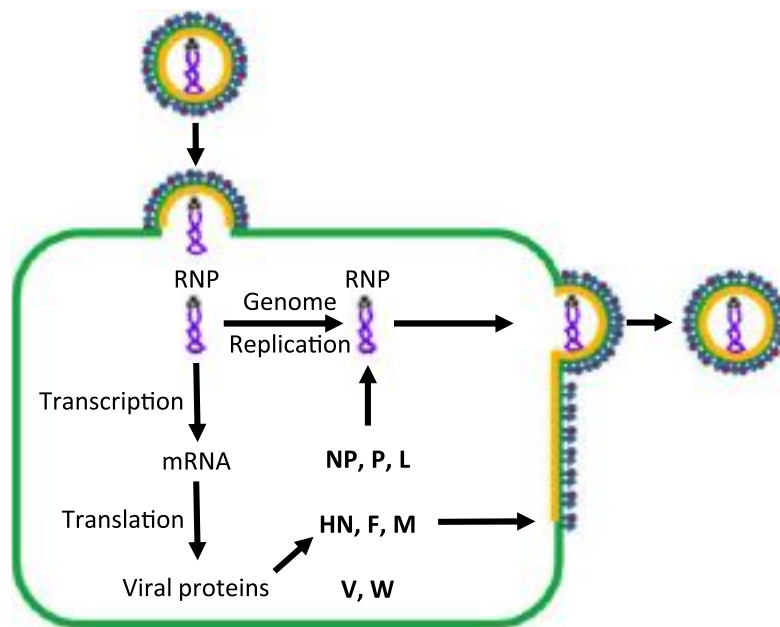


Fig. 2 AAVV-1 life cycle.

trailer (114 nt) regions. These regions contain regulatory sequences involved in transcription and replication of the genome. The terminal 10–20 nucleotides at the 3′ and 5′ ends are highly conserved and show complementarity to each other, suggesting an important role in genome replication. Transcription is initiated at the leader region and takes place by a sequential ‘stop-start’ mechanism in which the polymerase terminates transcription of the proximal gene and starts transcription of the distal gene. To this end, each gene is flanked by conserved transcription-initiation and transcription-termination sequences, the so called ‘gene-start’ and ‘gene-end’ boxes, which are separated by intergenic regions that vary in length from 1 to 47 nucleotides. Since not every stop results in a start event, the sequential transcription process gives rise to an mRNA gradient from the 3′ to the 5′ end of the genome, resulting in higher amounts of the structural proteins compared to the polymerase protein ([Fig. 2](#)).

Life Cycle

AAVV-1 infection is initiated by HN-mediated attachment of the virus particle to sialic acid-containing glycoproteins on the cell surface. Binding of HN to sialic acid triggers a conformational change in the F protein that leads to fusion of the viral envelope with the plasma membrane. It has been reported that NDV may also enter the host cell through receptor-mediated endocytosis. After entry, the RNP is released into the cytoplasm where the genomic RNA is transcribed by the L protein to produce capped and polyadenylated mRNAs for synthesis of the viral proteins. When sufficient amounts of the viral proteins have accumulated, the polymerase complex switches from transcription to replication of the genome. Newly formed genomic RNA is then covered by NP and associates with a few copies of the polymerase complex to form RNPs. The other components of the virus particle are transported to the cell membrane where they are assembled under the direction of the M protein, followed by encapsidation of the

RNP and subsequent budding of mature virions. Finally, the neuraminidase activity of the HN protein removes sialic acid from progeny virus particles, which facilitates detachment from the cell and prevents virus self-aggregation. The V protein is rich in cysteine and binds zinc while its carboxy-terminal portion has an anti-interferon activity, allowing the virus to counteract the innate immune response of the host.

Epidemiology

At least 241 species from 27 of the 50 orders of birds have been shown to be susceptible to AAvV-1 infection. However, clinical disease signs, if any, may vary considerably between different species. Virulent AAvV-1 strains have been isolated from all types of poultry, ranging from pigeons to ostriches. Chickens infected with virulent AAvV-1 show overt disease signs and often succumb to the infection, whereas ducks may not show clinical signs. In some poultry species the disease may be milder than in chickens and cause problems in diagnosis, e.g. in pheasants and ostriches. Furthermore, there is often a difference in the severity of disease between young and adult birds. AAvV-1 isolates have been frequently obtained from wild birds, especially migratory aquatic birds. Most of these isolates were of low virulence for chickens. Occasionally, virulent viruses have been detected in wild birds found dead near infected poultry, suggesting that spill-over from infected poultry to wild birds can occur. For example, in North America AAvV-1 causes periodic die-offs in cormorants, which may also represent a reservoir species for this virus.

Wild birds, captive birds and racing pigeons have been implicated in the introduction of virulent NDV into poultry in a number of outbreaks over the last two or three decades. Psittacine birds such as parrots have been shown to excrete virulent AAvV-1 intermittently for extremely long periods of time without showing clinical signs, emphasising the potential role of these birds in dissemination of AAvV-1. In the late 1970s a genotype VI AAvV-1 strain that had adapted to pigeons (known as PPMV-1) arose in the Middle East. This virus spread in racing and show pigeons throughout the world, resulting in a true panzootic (see below). The virus still remains enzootic in racing pigeons in many countries, posing a continuing threat to poultry.

AAvV-1 infected birds shed virus in oropharyngeal secretions and faeces. Susceptible birds may become infected by ingesting or inhaling contaminated material. Vaccinated chickens may shed velogenic AAvV-1 for 6–9 days after infection, while unvaccinated wild or captive exotic birds like parrots, cormorants, and pigeons may have prolonged shedding of virus without clinical signs.

The emergence of virulent AAvV-1 is a serious concern to the global poultry industry. AAvV-1 strains circulating in wild bird reservoirs are probably the major risk for the introduction of virulent AAvV-1 in poultry. Through ecological contacts and trade, these viruses can be transmitted from wild birds to domesticated poultry, where they cause clinical disease. Also, continuous circulation of lentogenic AAvV-1 among poultry forms a potential risk factor for the emergence of virulent AAvV-1. Furthermore, vaccine-derived AAvV-1 strains used in domesticated poultry have been reported to occur in wild birds, which further provides evidence of virus exchange between wild and domesticated birds. Thus, the continuous circulation and maintenance of AAvV-1 between wild and domesticated birds constitute a considerable threat for the emergence of virulent AAvV-1 strains. Introduction of virulent AAvV-1 into poultry has huge economic consequences in particular for exporting countries.

Since the emergence of velogenic AAvV-1, four major AAvV-1 panzootics have occurred. The first started around the mid-1920s in Asia and Europe and was initially caused by genotype II viruses and later also by genotype III and IV viruses. In this case, spreading to the rest of the world was relatively slow and took about twenty years. The second panzootic started in the late 1960s and was caused by genotype V viruses. This time the panzootic developed within four years presumably due to the global increase in large-scale commercial poultry production and the international trade of poultry and captive wild birds. The third panzootic was caused by genotype VI viruses and started in the early-1980s among racing pigeons but eventually affected several other bird species. The fourth panzootic was caused by genotype VII viruses and started in the mid-1980s, resulting in severe economic losses in a large number of countries in South East Asia, the Middle East, Europe, Africa, and America. Currently, genotype VII viruses still constitute the most rapidly evolving strains of the virus. The recent expansion in the geographic distribution and host range of sub-genotype VIIi strains may signify the start of the fifth panzootic.

Pathogenesis

Velogenic viruses can be further divided into viscerotropic velogenic (with a tropism for intestinal organs and tissues) and neurotropic velogenic (causing predominantly neurological signs). All mesogenic and velogenic AAvV-1 isolates require immediate notification to the World Organisation for Animal Health (OIE). Hence, pathotype identification is necessary for diagnosis of Newcastle disease (ND) in poultry. A common method to discriminate between different pathotypes is to establish the mean death time (MDT) in experimentally infected 9–10-day-old embryonated chicken eggs. Viruses are diagnosed as velogenic when the MDT is 40–60 hours, mesogenic when the MDT is 60–90 hours, and lentogenic when the MDT is > 90 hours. In case of a suspected outbreak, the pathogenicity test prescribed by the OIE is the intracerebral pathogenicity index (ICPI) in ten 1-day old SPF chickens. Velogenic strains have ICPI values from 1.4 to 2.0, while the values for mesogenic strains range from 0.7 to 1.4. Lentogenic isolates generally have ICPI values of 0.0–0.7.

The major determinant of AAvV-1 pathogenicity is the F protein. The capacity of the F protein to act as a fusion protein must be activated by cleavage of the precursor protein F₀ into F₁ and F₂. This cleavage is required for entry of AAvV-1 into host cells. Cleavage of

F₀ is mediated by cellular proteases and the type of protease that is capable of cleavage is dependent on the amino acid sequence motif around the cleavage site. The cleavage site of F₀ of lentogenic AAvV-1 is characterised by a monobasic amino acid sequence motif at the C-terminus of F₂ and a leucine at the N-terminus of F₁, ¹¹²G-R/K-Q-G-R↓L¹¹⁷. F₀ of lentogenic strains are cleaved mainly only extracellularly by trypsin-like proteases present in the respiratory and intestinal tract. Virulent strains have a multibasic amino acid sequence motif at the C-terminus of F₂ and a phenylalanine at the N-terminus of F₁, ¹¹²R/G/K-R-Q/K-K/R-R↓F¹¹⁷, and are cleaved intracellularly by ubiquitous furin-like proteases found in most host tissues. This difference in protease activation is the major determinant of systemic replication associated with severe disease of velogenic AAvV-1. Apart from the F protein, also other viral proteins may have a role in determining the pathogenicity of AAvV-1. The most notable example is the V protein which is involved in counteracting the host immune response. In general, the contribution of other viral proteins to pathogenicity is only minor when compared to the F₀ cleavage site.

The OIE has defined a Newcastle disease outbreak as an infection with any AAvV-1 that has an ICPI equal to or greater than 0.7. Detection of multiple basic amino acids near the cleavage site of the F protein is also acceptable to diagnose Newcastle disease. Viruses with a polybasic F cleavage site and phenyl alanine at position 117 are considered velogenic or mesogenic, while those with a monobasic F cleavage site and leucine at position 117 are considered lentogenic. However, if the analysis of the cleavage site of a suspected case reveals a lentogenic sequence, the ICPI test is still required by the OIE for the *in vivo* determination of pathogenicity.

Clinical Features

The clinical symptoms after an AAvV-1 infection can differ greatly, depending on the pathotype of the virus and/or vaccination or infection status of the birds. In fully susceptible birds, the most severe symptoms are observed with velogenic viscerotropic strains which can cause mortality rates up to 100%. Symptoms include conjunctivitis, nasal discharges, dyspnoea, diarrhoea, ruffled feathers, prostration, tremors, and paralysis. Throughout the digestive tract haemorrhages may be observed, especially in the proventriculus and the caecal tonsils. Necrosis may be observed in internal organs such as the spleen, liver, and gut associated lymphoid tissue.

Infections with velogenic neurotropic viruses are characterised by a mortality rate between 50% in older and 90% in younger susceptible birds. Mainly neurological and some respiratory clinical signs with almost no gastrointestinal involvement are observed. Typically, the affected birds show tremors, head twisting, opisthotonus and paralysis. Gross lesions are often absent but histologically, necrosis of Purkinje fibres as well as perivascular cuffing are often observed.

Infection with mesogenic strains is also associated with neurological and respiratory symptoms but with a low mortality rate. Clinical signs are associated with a drop in egg production and mild to moderate respiratory illness. Infection with lentogenic viruses are generally associated with mild respiratory signs or no evidence of clinical disease, in particular in young fully susceptible birds.

Diagnosis

Virus isolation is still regarded as the gold standard method for the diagnosis of an AAvV-1 infection. The choice of samples in live birds include cloacal and oropharyngeal swabs. Samples from dead birds should include tissue samples in addition to cloacal and oronasal swabs. The samples are inoculated into the allantoic cavity of 9- to 11-day-old specific pathogen and antibody free embryonated chicken eggs, and after 4–7 days of incubation a hemagglutination test (HA) is used to detect the presence of the virus in the allantoic fluid. Since other viruses also possess HA activity, it is necessary to further confirm the identity of the virus by either a hemagglutination inhibition test (HI) or molecular tests such as the polymerase chain reaction (PCR) or nucleotide sequencing.

Virus identification by quantitative polymerase chain reaction (qPCR) is faster and less cumbersome than conventional diagnostic techniques. In addition, it also provides equal or even greater sensitivity of virus detection than virus isolation. M-, L- and F-gene based qPCR assays have been described as standard methods for AAvV-1 screening. However only F-gene based PCRs can be used for pathotyping directly from clinical samples. AAvV-1 class I isolates may escape detection using F-gene based PCRs assays. Therefore, a multiplex M- and L-gene based PCR was proposed. For pathotyping, it is important to continuously monitor the genetic diversity of the evolving AAvV-1 isolates, so that the primers and probes can periodically be updated to identify all variants. In addition to disease identification and pathotyping, the qPCR assay can also be used for the quantification of viral load in different organs or swabs from vaccinated animals that have been challenged with a virulent AAvV-1 strain.

Serology is not very useful for the diagnosis of ND because serologic methods cannot differentiate antibodies induced after infection with AAvV-1 from those induced by vaccination. Nevertheless, serological tests are valuable tools to assess the humoral immune responses following vaccination. In countries that do not vaccinate, serology can confirm exposure to AAvV-1. In case of a suspicion based on clinical signs, a rise in HI titre may be an indication that exposure has occurred. The widest used serological test for AAvV-1 is the HI test which measures the ability of AAvV-1 specific antibodies to inhibit the agglutination of erythrocytes by the AAvV-1 particles. Virus neutralisation test (VNT) is yet another serological test that is useful for measuring AAvV-1 specific neutralising antibodies. Cross-reactivity of antibodies directed against other AAvV

serotypes in particular against AAvV-3 may complicate interpretation of HI results. Enzyme-linked immunosorbent assays (ELISA) are also used in AAvV-1 serology, and several commercially available ELISA kits based on viral proteins or whole virus antigen have been developed.

A tool that has recently revolutionised the diagnosis of infectious diseases is next-generation sequencing (NGS). NGS is not only important in tracking disease epidemics, but it also facilitates the rapid, sensitive, and specific detection and differentiation of mixed infections within a single host. Furthermore, it can also be used to detect low frequency variants within a mixed population which would escape detection using classical diagnostic tools. With sequencing services becoming cheaper and cheaper, in the near future, NGS is likely to become a routine tool for viral diagnosis.

Prevention

Tight biosecurity at the farm level and of the entire poultry production chain is critical to prevent the introduction of diseases to domestic poultry. Water and feed quality and pest and litter management are areas that need to be tightly controlled. All-in-all-out practices and routine disinfection and cleaning procedures between production rounds should be applied to reduce the risk of virus introduction. Biosecurity measures also should be employed in backyard flocks. Even with a good biosecurity practice, vaccination is still required to optimally protect the birds against economically devastating diseases such as ND.

Because vaccines in the field are not able to completely prevent vaccinated birds from being infected with velogenic AAvV-1, the role of vaccination in the control of ND is to prevent losses from morbidity and mortality. Vaccination increases the resistance to infection and decrease the number of birds that become infected and the quantity of virus being shed, thus reducing within- and between-flock transmission. The use of vaccines derived from naturally occurring highly immunogenic AAvV-1 strains of low or intermediate virulence with ICPI <0.4 has been quite successful and has limited the economic losses due to ND in many countries. To date, a number of lentogenic AAvV-1 strains such as LaSota, Ulster, B1, F, V4, and I2 are extensively used as live vaccines. Live vaccines derived from mesogenic viruses such as the Komarov and Mukteswar strains, have been used in the past to booster the immune response following priming with lentogenic viruses. However, for safety reasons and because of the new definition of ND adopted by the OIE, use of these vaccines is no longer allowed.

Live attenuated AAvV-1 vaccines applied by aerosol stimulate both mucosal and systemic immune responses. When applied correctly, a single dose of live AAvV-1 vaccine should be sufficient to induce an immune response capable of protecting 100% of the animals against clinical disease after challenge. However, more vaccinations are required to attain such levels in practice. Thus, shedding of virulent challenge virus via the cloacal and oropharyngeal routes may still occur. Replication and shedding of virulent virus can be substantially reduced when much higher doses of live vaccine are administered. However, because of vaccination reactions that can cause growth retardation and high cost of vaccination per bird, this is not an economically viable option. Thus, new approaches are required to tackle the problem of virus shedding associated with the use of conventional ND vaccines.

Live AAvV-1 vaccines are suitable for mass application via drinking water or spray. In addition, the vaccine virus may spread from vaccinated birds to suboptimally vaccinated ones, thereby contributing to overall herd immunity. In spite of all their benefits, live AAvV-1 vaccines have also shortcomings. Maternal antibodies in chickens up to the age of 3 weeks may interfere with the infection and replication of vaccine virus. After weaning of the maternal antibodies, vaccination may cause respiratory disease or cause growth retardation. These post vaccination reactions can predispose the birds to secondary infections by other pathogens. Furthermore, vaccines are produced using genotypes I or II seed strains which are phylogenetically divergent from currently prevalent genotypes. Hence, although the vaccines are still protective against clinical signs and mortality caused by any AAvV-1 isolate, their reduced ability to block shedding of contemporary viruses post challenge increases the risk of the continuous presence of the virulent virus in the environment. This is particularly more dangerous with genotype VII isolates, whose shedding from vaccinated birds is significantly higher than those of other genotypes. Therefore, considering the above limitations, live attenuated vaccines must be used with care and the state-of-the-art vaccines are urgently needed to address these weaknesses. Rationally designed vaccines targeting the prevailing genotypes, the so-called genotype-matched vaccines, are highly needed to overcome the shortcomings of classical vaccines. Reverse genetics-based live attenuated vaccines are probably the most promising candidates. Vector vaccines (see below) form another alternative approach having the advantage of in ovo application in the face of maternal immunity and claims of live-long protection.

Inactivated vaccines are not suitable for mass application because they have to be applied via the parental route, and thus individually. Furthermore, they do not replicate, and thus cannot spread horizontally among the birds. Individual application is labour intensive and thus expensive. The same nonreplicating feature, however, makes them safe with no risk of reversion to virulence. Inactivated AAvV-1 vaccines are generally poor inducers of mucosal or cell-mediated immune responses but induce high HI titres. Therefore, for best results, these vaccines are administered after initial priming and/or boosting with live vaccines in layer flocks of 15–17 weeks or older. Inactivated vaccines require the use of adjuvants which may cause undesirable reactions in vaccinated birds such as a drop in egg production. Furthermore, a withdrawal period is required before birds immunised with inactivated adjuvanted vaccines can be processed for human consumption. Therefore inactivated vaccines cannot be applied in broilers.

A promising strategy that can be used to combat pathogens is the use of recombinant viral vector vaccines. The most common vectors used in poultry are the vaccinia virus, fowl pox virus (FPV), herpes virus of turkeys (HVT), and Marek's disease virus (MDV). Because of their large double-stranded DNA genome, these viruses have a very high capacity for the insertion of foreign

genes. They are highly immunogenic, capable of inducing strong inflammatory innate immune responses, and they can be easily propagated on a large scale. Recombinant vaccinia virus expressing the F gene from AAvV-1 was shown to induce strong immunity that protected birds against virulent AAvV-1 challenge. Recombinant FPV-vectored ND vaccines generated by replacing the thymidine kinase gene with AAvV-1 F, HN or F and HN were shown to induce protective immunity against virulent AAvV-1 challenge in chickens. One of the advantages of an FPV-vectored vaccine is that it does not cause post vaccination respiratory reactions. However, the presence of anti-FPV antibodies in vaccinated chickens may dramatically interfere with the efficiency of this vector. Furthermore, this vaccine may not be suitable for young birds. This particular shortcoming can be overcome by the use of recombinant herpes virus vectored vaccines which can be used at the hatchery both in ovo in 18-day-old embryo's or in 1-day-old chickens. It elicits a strong and long lasting cell mediated and humoral immunity due to its ability to persist as a latent infection in the vaccinated chicken.

Further Reading

- Alexander, D.J., 2009. Ecology and epidemiology of newcastle disease. In: Capua, I., Alexander, D.J. (Eds.), *Avian Influenza and Newcastle Disease: A Field and Laboratory Manual*. New York: Springer, pp. 19–26.
- Ayala, A.J., Dimitrov, K.M., Becker, C.R., *et al.*, 2016. Presence of vaccine-derived newcastle disease viruses in wild birds. *PLOS ONE* 11 (9), e0162484.
- Cardenas-Garcia, S., Diel, D.G., Susta, L., *et al.*, 2015. Development of an improved vaccine evaluation protocol to compare the efficacy of Newcastle disease vaccines. *Biologicals* 43 (2), 136–145.
- Cattoli, G., Susta, L., Terregino, C., Brown, C., 2011. Newcastle disease: A review of field recognition and current methods of laboratory detection. *Journal of Veterinary Diagnostic Investigation* 23, 637–656.
- Czegledi, A., Ujvari, D., Somogyi, E., *et al.*, 2006. Third genome size category of avian paramyxovirus serotype 1 (Newcastle disease virus) and evolutionary implications. *Virus Research* 120, 36–48.
- Diel, D.G., da Silva, L.H., Liu, H., *et al.*, 2012. Genetic diversity of avian paramyxovirus type 1: Proposal for a unified nomenclature and classification system of Newcastle disease virus genotypes. *Infection, Genetics and Evolution* 12, 1770–1779.
- Dimitrov, K.M., Afonso, C.L., Yu, Q., Miller, P.J., 2017. Newcastle disease vaccines – A solved problem or a continuous challenge? *Veterinary Microbiology* 206, 126–136.
- Dimitrov, K.M., Ramey, A.M., Qiu, X., Bahl, J., Afonso, C.L., 2016. Temporal, geographic, and host distribution of avian paramyxovirus 1 (Newcastle disease virus). *Infection, Genetics and Evolution* 39, 22–23.
- Dortmans, J.C., Koch, G., Rottier, P.J., Peeters, B.P., 2011. Virulence of newcastle disease virus: What is known so far? *Veterinary Research* 42, 122.
- Miller, P.J., King, D.J., Afonso, C.L., Suarez, D.L., 2007. Antigenic differences among Newcastle disease virus strains of different genotypes used in vaccine formulation affect viral shedding after a virulent challenge. *Vaccine* 25, 7238–7246.
- Miller, P.J., Haddas, R., Simanov, L., *et al.*, 2015. Identification of new sub-genotypes of virulent Newcastle disease virus with potential panzootic features. *Infection, Genetics and Evolution* 29, 216–229.
- Miller, P.J., Koch, G., Newcastle Disease, 2013. *Diseases of Poultry*. Swayne, D.E. (Ed.), thirteenth ed. John Wiley & Sons, Inc.
- Palya, V., Kiss, I., Tatar-Kis, T., *et al.*, 2012. Advancement in vaccination against Newcastle disease: Recombinant HVT NDV provides high clinical protection and reduction challenge virus shedding, with the absence of vaccine reaction. *Avian Diseases* 56, 282–287.
- Peeters, B.P.H., De Leeuw, O.S., Koch, G., Gielkens, A.L.J., 1999. Rescue of Newcastle disease virus from cloned cDNA: Evidence that cleavability of the fusion protein is a major determinant for virulence. *Journal of Virology* 73, 5001–5009.
- Rauw, F., Gardin, Y., Palya, V., *et al.*, 2010. Improved vaccination against Newcastle disease by an in ovo recombinant HVT-ND combined with an adjuvanted live vaccine at day-old. *Vaccine* 28, 823–833.

Relevant Websites

- <http://agriculture.vic.gov.au/agriculture/pests-diseases-and-weeds/animal-diseases/poultry/newcastle-disease>
Agriculture Victoria – Newcastle disease.
- https://ec.europa.eu/food/animals/animal-diseases/control-measures/newcastle-disease_en
European Commission. Animal Diseases and control measures – Newcastle disease.
- <http://www.efsa.europa.eu/en/efsajournal/pub/477>
European Food Safety Authority – Opinion of the Scientific Panel on Animal Health and Welfare (AHAW) to review Newcastle disease focussing on vaccination worldwide in order to determine its optimal use for disease control purposes.
- <https://www.agric.wa.gov.au/livestock-biosecurity/newcastle-disease>
Government of Western Australia. Department of Primary Industries and Regional Development – Newcastle disease.
- <https://www.gov.uk/guidance/newcastle-disease>
Gov.uk – Newcastle disease: How to spot and report it.
- <http://www.cfsph.iastate.edu/DiseasesInfo/disease.php?name=newcastle-disease&lang=en>
The Center for Food Security and Public Health – Newcastle Disease.
- <https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/animal-disease-information/avian-influenza-disease/vnd>
United States Department of Agriculture – Virulent Newcastle Disease.
- <https://www.wur.nl/en/Research-Results/Research-Institutes/BioVeterinary-Research/Animal-diseases/Virology/Newcastle-disease-3.htm>
Wageningen University and Research – Newcastle Disease.
- <http://www.oie.int/en/animal-health-in-the-world/animal-diseases/Newcastle-disease/>
World Organisation for Animal Health – Newcastle Disease.
- http://www.oie.int/index.php?id=169&L=0&htmfile=chapitre_nd.htm
World Organisation for Animal Health – Terrestrial Animal Health Code. Infection with Newcastle Disease.

Orthobunyaviruses (*Peribunyaviridae*)

Alyssa B Evans, Clayton W Winkler, and Karin E Peterson, National Institutes of Health, Hamilton, MT, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of C.H. Calisher, Orthobunyaviruses, In Encyclopedia of Virology (Third Edition), Edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00623-3](https://doi.org/10.1016/B978-012374410-4.00623-3).

Glossary

Arbovirus A virus that is transmitted to vertebrates through the bite of a hematophagous (blood-feeding) arthropod.

Endemicity When a virus is found consistently in a region, generally determined by measuring neutralizing antibody responses in the primary animal reservoir for that particular virus.

Orthobunyavirus Genus of viruses in the family *Peribunyaviridae*, order *Bunyvirales*. They are enveloped viruses with tri-segmented single stranded negative-sense

RNA genomes. These segments are L (large), M (medium) and S (small).

Reassortment When one or more segments of one orthobunyavirus is combined with segments from another virus to form a new reassortant virus (e.g., the M segment from Batai virus reassorted with the L and S segments from Bunyamwera virus to generate Ngari virus).

Transovarial The transmission of a virus or other pathogen from an organism, such as a mosquito, to its offspring by infection of the eggs in the ovary.

Classification

Serological Classification

The orthobunyaviruses are enveloped, negative-sense, single-stranded RNA viruses containing three genomic segments, the Large (L), Medium (M), and Small (S) segments (**Fig. 1(A)**). Traditionally, the orthobunyaviruses have been classified into serogroups via serological assays such as complement fixation, hemagglutination (HI), and neutralizing antibody assays. However, classification into serogroups is complicated by the tri-segmented nature of the genome of these viruses. Different serological assays target different genome segments, which can lead to contradicting results. Complement fixation assays interact with the Nucleocapsid protein encoded by the S segment, while HI and neutralization assays interact with the Gn and Gc envelope glycoproteins encoded by the M segment (**Table 1**). Although none of the serological tests interact with proteins generated by the L segment, differences between M and S segments can affect serological testing. Related segmented viruses can exchange genome segments in a process called reassortment (**Fig. 1(B)**). Reassortant viruses may be determined to fall into two different serogroups depending on the serological test used in classification. Even without the process of reassortment, viruses within a serogroup can group differently depending upon the serological test used. Despite these caveats, serogroups are a useful way of determining antigenic relatedness and are an important consideration in diagnostic testing. At least 20 orthobunyavirus serogroups have been identified or proposed (**Table 2**). The Bunyamwera, California, and Simbu serogroups contain many of the most significant human and livestock pathogens of the *Orthobunyavirus* genus. The orthobunyaviruses do not have any cross-reactivity with other genera of bunyaviruses.

Phylogenetic Analysis by Sequencing

With the increased capacity and decreased cost of sequencing technology, classification of viruses by phylogenetic relationships has become more common, and the orthobunyaviruses are no exception. Based on genetics, distinguishing characteristics of the *Orthobunyavirus* genus include specific consensus terminal nucleotide sequences and the M segment which encodes the envelope proteins and a nonstructural protein as a precursor polyprotein in the order of Gn-NSm-Gc. As with serology, different results can be obtained depending on the genome segment(s) used for the analysis. Additionally, the viruses/strains used in studies and methods of phylogenetic analysis can yield differing results between studies. However, in the phylogenetic analyses that have been performed, genetic relationships tend to group with serological classifications. The advantage of sequencing over serology is that all three segments can be compared. Despite advances in the technology, sequencing is still more difficult, time-consuming, and expensive than serology. However, sequencing information is an important tool for evaluating virus genetic relatedness and assessing reassortment between orthobunyaviruses, particularly in cases where serology analysis results in conflicting or inconclusive classification. According to the International Committee for the Taxonomy of Viruses 9th Report, when sequencing data is available, orthobunyaviruses are categorized as different species when there is more than a 10% difference in amino acid sequences of the nucleocapsid protein encoded by the S segment. However, this classification is limited because it does not account for the genetic diversity in the M and L segments.

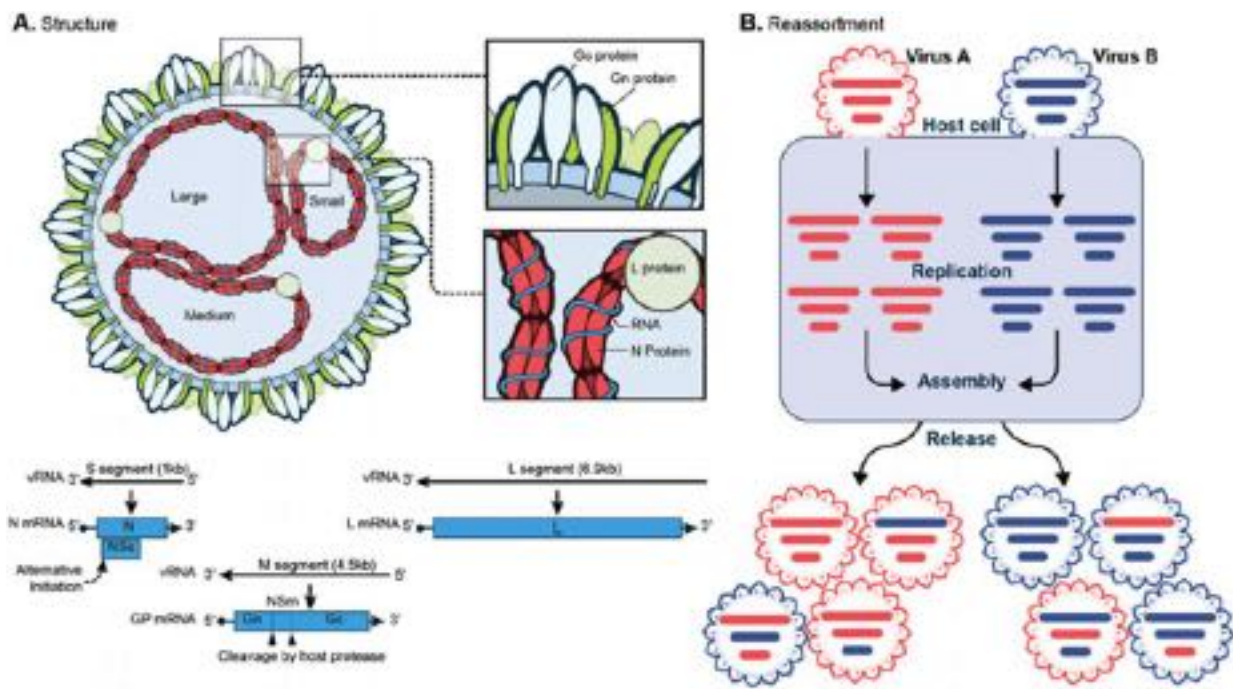


Fig. 1 Structure, genomic makeup and reassortment potential of orthobunyaviruses. (A) Orthobunyaviruses are enveloped spherical viruses with two envelope proteins (Gn and Gc) in the membrane, which surrounds three negative strand RNA segments of genome (L, M, S), associated with the N and L proteins. The genome and encoded proteins are shown below. (B) Due to the tri-segmented nature of orthobunyaviruses, when two viruses infect the same cell the individual segments of each viral genome can be reassorted and produce eight possible different viruses. These include the original two parental strains as well as a multiple combinations of L, M, and S segments.

Table 1 Orthobunyavirus proteins of Bunyamwera virus prototype

Genome segment	Protein	Amino acids	Function
Large	L protein	2238	RNA-dependent RNA-polymerase: Mediates transcription and replication Endonuclease domain: involved in cap-snatching from host mRNAs
Medium	Glycoprotein(n) (Gn)	296	Envelope glycoprotein; chaperone for Gc trafficking to the Golgi
	Glycoprotein(c) (Gc)	146	Envelope glycoprotein: Class II fusion protein
	Nonstructural protein M (NSm)	956	May play a role in virion assembly and budding; not strictly required for replication in vitro
Small	Nucleocapsid (N)	233	Encapsidates genomic and anti-genomic RNAs to form ribonucleoproteins
	Nonstructural protein S (NSs) ^a	101	Inhibits host RNA Polymerase II; Type I IFN-antagonist; not required for replication Inhibits apoptosis in BUNV, but increases apoptosis in LACV

^aNot all orthobunyaviruses encode NSs, including members of the Anopheles A, Anopheles B, and Tete serogroups.

Genome

Genome and Encoded Proteins

Orthobunyaviruses contain tri-segmented, single-stranded, negative-sense RNA genomes (Fig. 1(A)). The Large (L) segment is the largest at ~6.9 kb and encodes the L protein, which functions as the viral RNA-dependent RNA-polymerase (RdRp, L protein) and contains an endonuclease domain (Table 1). The Medium (M) segment is ~4.5 kb and encodes a polyprotein in the order Gn-NSm-Gc. This protein is proteolytically processed into envelope glycoproteins (Gn and Gc) and a nonstructural protein (NSm) by internal signal peptides and host signal peptide peptidase. The Small (S) segment is ~1 kb and encodes the nucleocapsid (N) protein. Most orthobunyaviruses also encode a second nonstructural protein on the S segment (NSs) in an overlapping reading frame with N. However, some orthobunyaviruses, including members of the Anopheles A, Anopheles B, and Tete serogroups, do

Table 2 Orthobunyavirus family members ICTV recognized species; group virus not listed as ICTV species

<i>Virus</i>	<i>Serogroup</i>	<i>Disease</i>	<i>Vector</i>	<i>Host</i>	<i>References</i>
<i>Anopheles A virus</i>	Anopheles A				Hughes <i>et al.</i> (2017)
Lukuni virus	Anopheles A				Hughes <i>et al.</i> (2017)
<i>Tacaiuma virus</i>	Anopheles A				Hughes <i>et al.</i> (2017)
<i>Anopheles B virus</i>	Anopheles B				Mohamed <i>et al.</i> (2009)
Boraceia virus	Anopheles B				Mohamed <i>et al.</i> (2009)
<i>Bakau virus</i>	Bakau		Ticks		Zeller <i>et al.</i> (1989)
Ketapang virus	Bakau				Zeller <i>et al.</i> (1989)
Nola virus	Bakau				Zeller <i>et al.</i> (1989)
Tanjong Rabok virus	Bakau				Zeller <i>et al.</i> (1989)
Telok Forest virus	Bakau				Zeller <i>et al.</i> (1989)
<i>Anadyr virus</i>	Bunyamwera		mosquitoes		Shchetinin <i>et al.</i> (2015) (in Russian)
<i>Anhembi virus</i>	Bunyamwera		mosquitoes		de Souza Lopez <i>et al.</i> (1975)
<i>Batai virus</i>	Bunyamwera		mosquitoes		Dutuze <i>et al.</i> (2018)
<i>Birao virus</i>	Bunyamwera		mosquitoes		Arbocat
<i>Bozo virus</i>	Bunyamwera		mosquitoes		Arbocat
<i>Bunyamwera virus</i>	Bunyamwera	livestock, avian, humans: mild; fever, joint pain, rash	mosquitoes		Dutuze <i>et al.</i> (2018)
<i>Cache Valley virus</i>	Bunyamwera	sheep: embryonic and fetal death, stillbirths, congenital defects; human disease (rare)	mosquitoes		
Calovo virus	Bunyamwera		mosquitoes		Zeller <i>et al.</i> (1989)
<i>Fort Sherman virus</i>	Bunyamwera		mosquitoes		Arbocat
Germiston virus	Bunyamwera		mosquitoes		Calisher (1983)
<i>Ilesha virus</i>	Bunyamwera	humans: mostly febrile, occasional association with meningoencephalitis and hemorrhagic fever	mosquitoes		Pachler <i>et al.</i> (2013)
<i>Kairi virus</i>	Bunyamwera	not reported	mosquitoes		Soto <i>et al.</i> (2009)
Lokern virus	Bunyamwera		mosquitoes		Meyers <i>et al.</i> (2015)
<i>Maguari virus</i>	Bunyamwera		mosquitoes		Pollitt <i>et al.</i> (2006)
<i>Main Drain virus</i>	Bunyamwera	experimental infection in sheep: severe musculoskeletal and nervous system malformations, death in fetuses	mosquitoes		Tangudu <i>et al.</i> (2018); Arbocat
Ngari virus*	Bunyamwera	livestock, avian, humans: fatal hemorrhagic fevers in humans and ruminants	mosquitoes		Briese <i>et al.</i> (2006)
Northway virus	Bunyamwera		mosquitoes		Meyers <i>et al.</i> (2015)
Playas virus	Bunyamwera		mosquitoes		Calisher <i>et al.</i> (1983)
<i>Potosi virus</i>	Bunyamwera		mosquitoes		Harrison <i>et al.</i> (1995)
<i>Tensaw virus</i>	Bunyamwera	human encephalitis	mosquitoes		Tangudu <i>et al.</i> (2018)
<i>Bwamba virus</i>	Bwamba	humans: febrile illness, hemorrhagic complications			Groseth <i>et al.</i> (2014)
Pongola virus	Bwamba				Groseth <i>et al.</i> (2014)
<i>California encephalitis virus</i>	California	neuroinvasive in humans	mosquito		Evans and Peterson (2019)
<i>Jamestown Canyon virus</i>	California	neuroinvasive in humans	mosquito		Evans and Peterson (2019)
<i>Keystone virus</i>	California	Rash in 1 human	mosquito		Evans and Peterson (2019)
<i>La Crosse virus</i>	California	neuroinvasive in humans	mosquito		Evans and Peterson (2019)
<i>Lumbo virus</i>	California		mosquito		Evans and Peterson (2019)
<i>Melao virus</i>	California		mosquito		Evans and Peterson (2019)
<i>San Angelo virus</i>	California		mosquito		Evans and Peterson (2019)
<i>Serra do Navio virus</i>	California		mosquito		Evans and Peterson (2019)
<i>Snowshoe hare virus</i>	California	neuroinvasive in humans	mosquito		Evans and Peterson (2019)
<i>Tahyna virus</i>	California	mostly flu-like, neuroinvasive	mosquito		Evans and Peterson (2019)
<i>Trivittatus virus</i>	California		mosquito		Evans and Peterson (2019)
Inkoo virus	California	neuroinvasive in humans	mosquito		Evans and Peterson (2019)
South River virus	California		mosquito		Evans and Peterson (2019)
Chatanga virus	California	neuroinvasive in humans	mosquito		Evans and Peterson (2019)
Morro Bay virus	California		mosquito		Evans and Peterson (2019)
Infirmatus virus	California		mosquito		Evans and Peterson (2019)
Achiote virus	California		mosquito		Evans and Peterson (2019)
Jerry Slough virus	California		mosquito		Evans and Peterson (2019)

Table 2 Continued

<i>Virus</i>	<i>Serogroup</i>	<i>Disease</i>	<i>Vector</i>	<i>Host</i>	<i>References</i>
<i>Acara virus</i>	Capim				(Arbocat)
<i>Benevides virus</i>	Capim				Shchetinin <i>et al.</i> (2015)
<i>Benfica virus</i>	Capim				Shchetinin <i>et al.</i> (2015)
<i>Bushbush virus</i>	Capim				Shchetinin <i>et al.</i> (2015)
<i>Capim virus</i>	Capim				Shchetinin <i>et al.</i> (2015)
<i>Guajara virus</i>	Capim				Shchetinin <i>et al.</i> (2015)
<i>Juan Diaz virus</i>	Capim				Shchetinin <i>et al.</i> (2015)
<i>Moriche virus</i>	Capim				Shchetinin <i>et al.</i> (2015)
<i>Enseada virus</i>	Enseada (proposed)				de Souza <i>et al.</i> (2016)
<i>Alajuela virus</i>	Gamboa	Not reported	mosquitoes	birds	Chiang <i>et al.</i> (2018)
<i>Calchaqui virus</i>	Gamboa	Not reported	mosquitoes	birds	Chiang <i>et al.</i> (2018)
<i>Gamboa virus</i>	Gamboa	Not reported	mosquitoes	birds	Chiang <i>et al.</i> (2018)
<i>Pueblo Viejo</i>	Gamboa	Not reported	mosquitoes	birds	Chiang <i>et al.</i> (2018)
<i>San Juan virus</i>	Gamboa		mosquitoes		Chiang <i>et al.</i> (2018)
<i>Soberania virus</i>	Gamboa	Not reported	mosquitoes	birds	Chiang <i>et al.</i> (2018)
<i>Apeu virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Bruconha virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Caraparu virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Gumbo limbo virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Itaqui virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Madrid virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Marituba virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Murutuca virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Nepuyo virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Oriboca virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Ossa virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Restan virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Vinces virus</i>	Group C				Nunes <i>et al.</i> (2005)
<i>Ananindeua virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Bertioga virus</i>	Guama				de Souza Lopez <i>et al.</i> (1975)
<i>Bimiti virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Cananeia virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Catu virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Guama virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Guaratuba virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Itimirim virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Mahogany Hammock virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Mirim virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Moju virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Timboteua virus</i>	Guama				Shchetinin <i>et al.</i> (2015)
<i>Guaroa virus</i>	Guaroa (proposed)	humans: febrile illness			Aguilar <i>et al.</i> (2010)
<i>Koongol virus</i>	Koongol				Shchetinin <i>et al.</i> (2015)
<i>Wongal virus</i>	Koongol				Shchetinin <i>et al.</i> (2015)
<i>Leanyer virus</i>	Leanyer (proposed)				Savji <i>et al.</i> (2011)
<i>Buffalo Creek virus</i>	Mapputta				Shchetinin <i>et al.</i> (2015)
<i>Gan Gan virus</i>	Mapputta				Shchetinin <i>et al.</i> (2015)
<i>Mapputta virus</i>	Mapputta				Shchetinin <i>et al.</i> (2015)
<i>Maprik virus</i>	Mapputta				Shchetinin <i>et al.</i> (2015)
<i>Murrumbidgee virus</i>	Mapputta				Shchetinin <i>et al.</i> (2015)
<i>Salt Ash virus</i>	Mapputta				Shchetinin <i>et al.</i> (2015)
<i>Trubanaman virus</i>	Mapputta				Shchetinin <i>et al.</i> (2015)
<i>Minatitlan virus</i>	Minatitlan				Calisher <i>et al.</i> (1983)
<i>Palestina virus</i>	Minatitlan				Calisher <i>et al.</i> (1983)
<i>Nyando virus</i>	Nyando	humans: febrile illness			Groseth <i>et al.</i> (2014)
<i>Botambi virus</i>	Olifantsvlei				Arbocat
<i>Olifantsvlei virus</i>	Olifantsvlei				Taxonomy of the order Bunyavirales
<i>Abras virus</i>	Patois				Shchetinin <i>et al.</i> (2015)
<i>Babahoyo virus</i>	Patois				Shchetinin <i>et al.</i> (2015)
<i>Pahayokee virus</i>	Patois				Shchetinin <i>et al.</i> (2015)

(Continued)

Table 2 Continued

<i>Virus</i>	<i>Serogroup</i>	<i>Disease</i>	<i>Vector</i>	<i>Host</i>	<i>References</i>
<i>Patois virus</i>	Patois				Aguilar <i>et al.</i> (2018)
Shark River virus	Patois				Shchetinina <i>et al.</i> (2015)
<i>Zegla virus</i>	Patois				Zeller <i>et al.</i> (1989); Aguilar <i>et al.</i> (2018)
<i>Aino virus</i>	Simbu	ruminants: hydranencephaly/arthrogryposis, abortion, stillbirths, congenital abnormalities of CNS	biting midges		Agerholm <i>et al.</i> (2015); Saeed <i>et al.</i> (2001)
<i>Akabane virus</i>	Simbu	ruminants: hydranencephaly/arthrogryposis, abortion, stillbirths, congenital abnormalities of CNS	biting midges		Agerholm <i>et al.</i> (2015); Saeed <i>et al.</i> (2001)
<i>Buttonwillow virus</i>	Simbu				Saeed <i>et al.</i> (2001)
<i>Cat Que virus</i>	Simbu				Zhang <i>et al.</i> (2015)
Douglas virus	Simbu				Saeed <i>et al.</i> (2001)
<i>Faceys paddock virus</i>	Simbu				Saeed <i>et al.</i> (2001)
<i>Ingwavuma virus</i>	Simbu				Saeed <i>et al.</i> (2001)
Inini virus	Simbu				Saeed <i>et al.</i> (2001)
<i>Jatobal virus</i>	Simbu				Aguilar <i>et al.</i> (2011)
Kaikalur virus	Simbu				Saeed <i>et al.</i> (2001)
<i>Manzanilla virus</i>	Simbu				Saeed <i>et al.</i> (2001)
<i>Mermet virus</i>	Simbu				Saeed <i>et al.</i> (2001)
<i>Oropouche virus</i>	Simbu	humans: febrile illness			Saeed <i>et al.</i> (2001)
Para virus	Simbu				Saeed <i>et al.</i> (2001)
<i>Peaton virus</i>	Simbu				Saeed <i>et al.</i> (2001)
<i>Sabo virus</i>	Simbu				Saeed <i>et al.</i> (2001)
<i>Sango virus</i>	Simbu				Saeed <i>et al.</i> (2001)
Sathuperi virus	Simbu				Saeed <i>et al.</i> (2001)
Shamonda virus	Simbu				Saeed <i>et al.</i> (2001)
<i>Schmallenberg virus</i>	Simbu	ruminants: teratogenic, congenital malformations, CNS disease	biting midges		Agerholm <i>et al.</i> (2015)
<i>Shuni virus</i>	Simbu				Saeed <i>et al.</i> (2001)
<i>Simbu virus</i>	Simbu				Saeed <i>et al.</i> (2001)
<i>Thimiri virus</i>	Simbu				Saeed <i>et al.</i> (2001)
Tinaroo virus	Simbu				Saeed <i>et al.</i> (2001)
<i>Utinga virus</i>	Simbu				Saeed <i>et al.</i> (2001)
Yaba – 7	Simbu				Saeed <i>et al.</i> (2001)
Bahig virus	Tete				Shchetinina <i>et al.</i> (2015)
<i>Batama virus</i>	Tete		Ticks		Mohamed <i>et al.</i> (2009)
Matruh virus	Tete				Shchetinina <i>et al.</i> (2015)
<i>Tete virus</i>	Tete				Mohamed <i>et al.</i> (2009)
Tsuruse virus	Tete				Shchetinina <i>et al.</i> (2015)
Weldona virus	Tete				Shchetinina <i>et al.</i> (2015)
Lednice virus	Turlock				Shchetinina <i>et al.</i> (2015)
<i>MPoko virus</i>	Turlock				Shchetinina <i>et al.</i> (2015)
<i>Turlock virus</i>	Turlock				Shchetinina <i>et al.</i> (2015)
Umbre virus	Turlock				Shchetinina <i>et al.</i> (2015)
Yaba – 1 virus	Turlock				Shchetinina <i>et al.</i> (2015)
Anhembi virus	Wyeomyia				Chowdhary <i>et al.</i> (2012)
<i>Cachoeira Porteira virus</i>	Wyeomyia				Chowdhary <i>et al.</i> (2012)
<i>Iaco virus</i>	Wyeomyia				Chowdhary <i>et al.</i> (2012)
<i>Macaua virus</i>	Wyeomyia				Chowdhary <i>et al.</i> (2012)
<i>Sororoca virus</i>	Wyeomyia				Chowdhary <i>et al.</i> (2012)
Taissui virus	Wyeomyia				Chowdhary <i>et al.</i> (2012)
<i>Wyeomyia virus</i>	Wyeomyia				Chowdhary <i>et al.</i> (2012)
<i>Bellavista virus</i>	Unclassified				Hang <i>et al.</i> (2016)
<i>Kaeng Khoi virus</i>	Unclassified				Groseth <i>et al.</i> (2014)
<i>Tataguine virus</i>	Unclassified	Humans: fever, headache, rash, joint pain			Shchetinina <i>et al.</i> (2015)
<i>Witwatersrand virus</i>	Unclassified	not reported			Shchetinina <i>et al.</i> (2015)
<i>Wolkberg virus</i>	Unclassified				Jansen van Vuren <i>et al.</i> (2017)

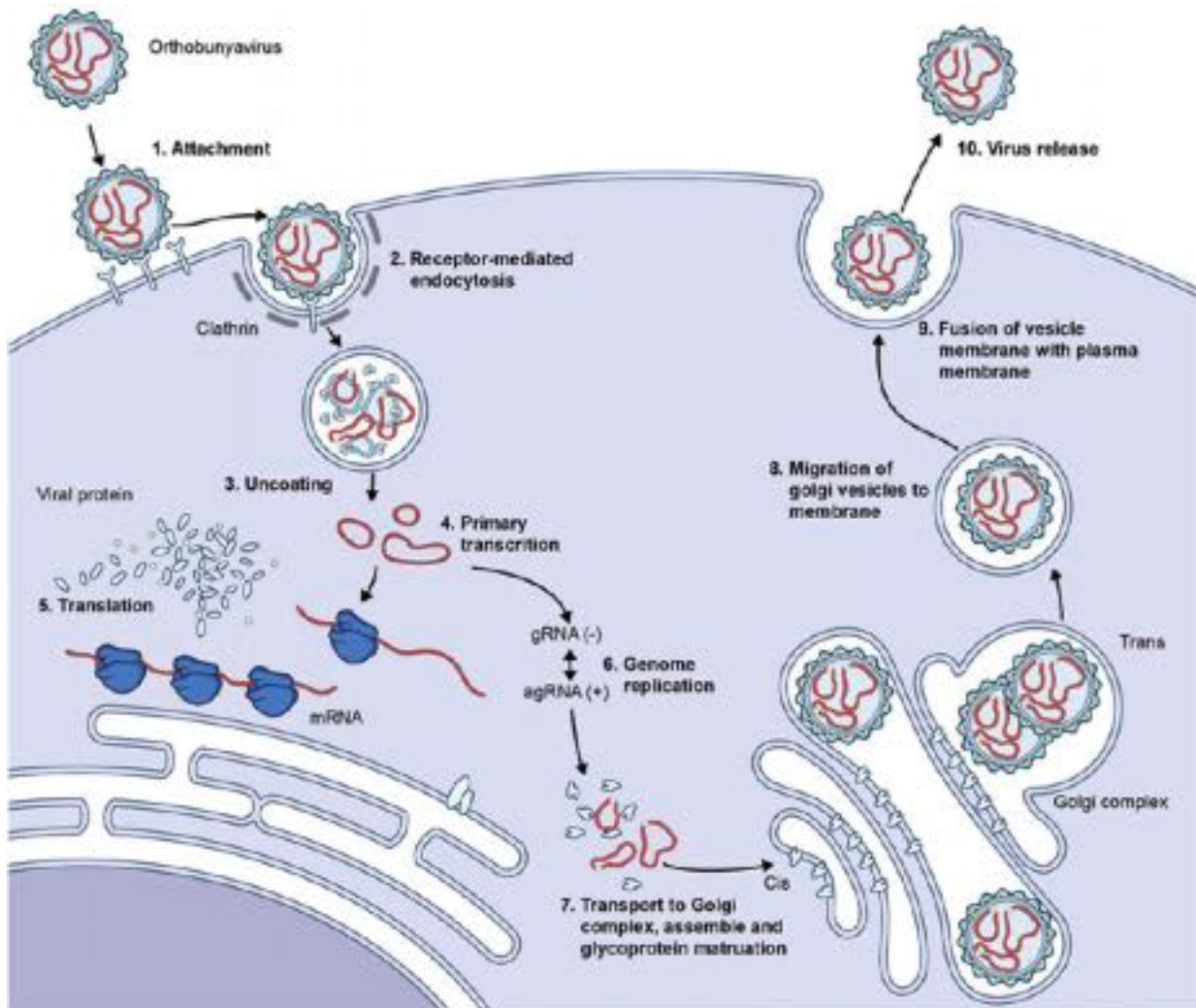


Fig. 2 Orthobunyaviruses enter host cells by receptor-mediated attachment of virus envelope (1) leading to clathrin-mediated endocytosis (2). The virus then uncoats (3) in endosomal vesicles and the RNA segments are released in the cytoplasm where they undergo primary transcription (4), translation of viral proteins (5) and genome replication (6). The components are transported to the Golgi complex (7) where virus assembly takes place. Virus particles then migrate to the membrane (8) where they fuse with the plasma membrane and are released via exocytosis.

not have NSs. In Bunyamwera virus (BUNV) and La Crosse virus (LACV), the NSs has been shown to antagonize the host antiviral response by targeting host Polymerase II, leading to impaired host transcription and interferon production.

Life Cycle

Most studies of orthobunyavirus replication have focused on just a select few viruses, primarily LACV and BUNV, the latter of which is the orthobunyavirus prototype. There is likely a great deal of diversity in the mechanisms used by different orthobunyaviruses for replication. From studies of LACV and BUNV, it has been shown that orthobunyaviruses utilize receptor-mediated endocytosis to enter cells (Fig. 2). The entry receptor(s) are currently not known for any of the orthobunyaviruses. Upon entry, orthobunyaviruses uncoat at the endosomal membrane, likely through conformational changes in Gc, which is the predicted orthobunyavirus fusion protein. The negative-sense genome segments are used as the template for the L protein RdRp to generate viral mRNA. The L protein contains an endonuclease domain that steals the 5' 7-methylguanylate capped ends of host mRNAs in order to prime transcription in a process called cap-snatching. Orthobunyaviruses require active host translation for viral transcription to occur. Genome replication uses full-length anti-sense ribonucleoproteins (RNPs) as templates for the RdRp in a primer-independent process to generate the full length negative-sense genome segments. The 5' and 3' ends of each genome segment are complementary and form a panhandle that circularizes each segment. The genome segments form RNP complexes through interactions with the N protein, and the panhandles are thought to associate with the L protein in the virion.

Several orthobunyaviruses have been shown to create Golgi-derived viral factories to facilitate virus replication through endosomal complex required for transport (ESCRT) machinery. Restructuring of the Golgi complex for viral replication likely helps facilitate efficient virus assembly, maturation, and budding, processes that occur at the membrane of the Golgi. Assembly and budding may be facilitated by NSm, however, in studies utilizing recombinant viruses, the importance of NSm for the production of infectious virus varies between orthobunyaviruses. The Golgi membranes express the viral glycoproteins. The Gn and Gc glycoproteins dimerize and form a tripod-like trimer of dimers on the envelope surface. Genomic RNPs are trafficked to these membranes where the UTRs mediate packaging. These interactions facilitate budding, then exocytosis releases virions from the cell. Once released from cells, the orthobunyavirus glycoproteins experience a final enzymatic cleavage maturation step that results in fully infectious virions. Virion size varies between orthobunyaviruses but is approximately 108 nm in diameter.

Transmission

Vector

Most isolations of orthobunyaviruses have been made from arthropods, and all orthobunyaviruses are thought to be carried and transmitted by hematophagous insects and/or arachnids. Orthobunyaviruses are found throughout the world, but the range of individual viruses is restricted by their arthropod vector and vertebrate host ranges. The majority of orthobunyaviruses utilize mosquito vectors. However, viruses in the Simbu serogroup are transmitted by biting midges, viruses in the Tete serogroup are transmitted by ticks, and one virus, Kaeng Khoi virus, has been found in bed bugs. Viruses from other serogroups have been isolated from non-mosquito insects, but their roles as vectors is unclear. For many orthobunyaviruses in the California serogroup, transovarial transmission in mosquito vectors has been well established as an over-wintering mechanism. However, evidence of transovarial transmission of other orthobunyaviruses is limited, so it is unclear if transovarial transmission is widely utilized as an overwintering mechanism for orthobunyaviruses.

Vertebrate Hosts

Vertebrate hosts have not been identified for many of the orthobunyaviruses. Some vertebrate hosts have been characterized for orthobunyaviruses, and range in diversity from birds to mammals. Most viruses in the California serogroup are transmitted by small mammals like hares, chipmunks, and squirrels. Bunyamwera serogroup viruses utilize a wide variety of vertebrate hosts including rodents, humans, and ruminants. Gamboa serogroup members appear to be maintained in a mosquito to bird transmission cycle, and Simbu viruses use a variety of mammals and birds. Due to the lack of evidence of transovarial transmission for many orthobunyaviruses, and the wide range of vertebrate hosts that the viruses have been associated with, most orthobunyaviruses are likely maintained in a vector to vertebrate host to vector transmission cycle.

Reassortment

Evidence of Reassortment

Reassortment is the process where related segmented viruses exchange genome segments creating novel reassortant viruses (**Fig. 1 (B)**). Reassortment among the orthobunyaviruses has been well documented between viruses as well as viruses within a serogroup. Based on phylogenetic analyzes, there is evidence of reassortment within several orthobunyavirus serogroups including the California, Guamá, Bunyamwera, and Simbu serogroups. Interestingly, the M segment is the most commonly reassorted segment between orthobunyaviruses, possibly because it encodes the envelope proteins that control virus entry.

Effects of Reassortment

Of particular interest and concern are reassortant viruses with increased pathogenicity compared to their parental viruses. In addition to the naturally occurring reassortant orthobunyaviruses, laboratory derived reassortant viruses have been generated between viruses in the California serogroup. Because many orthobunyaviruses share geographical ranges, vector species, and vertebrate hosts, the occurrence of novel reassortant orthobunyaviruses arising with enhanced pathogenicity is possible.

Epidemiology

Viral Discovery

The prototypical and namesake orthobunyavirus, BUNV, was first discovered in Uganda in 1943 as a result of *Aedes* mosquito surveillance studies looking for Yellow fever virus. This route of discovery is fairly typical as novel orthobunyaviruses have been discovered and isolated in surveys of known vector species, presumed vector species, host species and host parasite

species. Alternatively, novel viruses have also been identified as a result of human or animal disease. Because of the potential for reassortment of orthobunyaviruses, the expanding range of vector species, the ability of some orthobunyaviruses to colonize new vectors and the increasing numbers of novel human cases associated with orthobunyavirus infection, surveillance, identification, characterization and early detection of novel pathogenic viruses that could lead to an outbreak virus is paramount.

Surveying Viral Endemic Ranges and Tracking Viral Spread

Areas of endemicity for most orthobunyaviruses have been determined by a combination of vector and host molecular identification surveys, and seroprevalence studies. However, most molecular identification studies are limited by being short in duration and only determining the presence of a virus rather than establishing if that virus has a replicative life cycle in a putative endemic area. To supplement these survey studies, seroprevalence studies in host species, including wildlife, domesticated livestock and humans have also been completed. While these types of studies can be complicated by viral misidentification due to antibody cross-reactivity within and between orthobunyavirus serogroups, they do utilize easily accessible samples that can be monitored longitudinally to determine endemicity. Seroprevalence sampling is also a convenient and effective tool to monitor viral spread particularly for human disease-causing orthobunyaviruses. This technique is commonly employed by government agencies to monitor health risks to their population and provide public awareness. Such information is becoming increasingly important in the modern era of human global travel and expanding vector ranges.

Orthobunyavirus Outbreaks and Tracking

Due to the inherently high mutation rates of RNA viruses and the ability of orthobunyaviruses to reassort and acquire pathogenic features, periodic outbreaks and emergence of novel orthobunyaviruses occur. A recent example involves an outbreak of Schmallenberg virus, which is transmitted by biting midges, in sheep and cattle during 2011–2012 in Europe. The first cases reported were in December of 2011 in the Netherlands and Belgium, but shortly thereafter, cases were reported in France, Luxemburg, and Spain corresponding to the initial outbreak. In late May of 2012, new cases were reported in France, Germany and the UK indicating that the virus had reemerged following the winter. Early in 2012, a serological test was developed which allowed for the wide-spread tracking of the virus, which revealed the virus had spread to the entirety of northern Europe by early 2013. The speed with which this virus spread was surprising and was likely influenced by propagation by a broad-ranged, mobile vector, the shipping of livestock, and the availability of naïve hosts. Recent serological surveys have indicated that Schmallenberg virus continues to circulate at low levels in northern European livestock herds, but no further outbreaks have been reported.

Outbreaks of reassortant orthobunyaviruses have also occurred in humans. One such virus, Ngari virus, emerged at the end of the last century in East Africa. In 1988 and again in 1997 and 1998, outbreaks of severe hemorrhagic fever occurred in eastern Africa. While at the time all febrile illness cases were diagnosed as malaria, random serological surveys of patient samples suggest that more than 5000 cases could be attributed to a Bunyamwera serogroup orthobunyavirus infection. Initial molecular analysis of patient samples determined the causative virus was a reassortant of BUNV, as the S and L segments, but not the M segment, matched this virus. Further studies showed Ngari virus to be a novel reassortant with the L and S segments from BUNV and the M segment from Batai virus that had in fact been first discovered in 1979. Both BUNV and Batai viruses cause mild febrile disease in humans, though Batai virus can cause severe congenital malformations in ruminants as well. However, neither parental virus is known to cause severe hemorrhagic fever demonstrating how reassortment can lead to a more pathogenic virus and more severe human infection.

Another factor that influences outbreaks and the detection of disease is alterations in climate. For example, Ngari virus may have gone unrecognized as a human pathogen until much later had the outbreaks in the 1980s and 90s not occurred. Those years were associated with unusually high rainfall and increased mosquito breeding in the outbreak area which likely led to a large increase in reported human febrile cases. This increase in cases led researchers to perform retrospective molecular studies that identified Ngari virus and drew attention to its pathogenic potential in humans. It is possible that other orthobunyaviruses which possess pathogenic potential have yet to be discovered due to a low incidence of human cases.

The life cycle of orthobunyaviruses within a vector may also affect outbreaks. For some orthobunyaviruses, such as those in the California and possibly the Simbu serogroups, the viruses are transovarially transmitted, allowing them to over-winter in vector species and reemerge each year in endemic regions. This process is also likely aided by viral genetic factors that sustain infection throughout the life of the vector which allows for more host encounters. For other serogroup viruses, it is unclear how they reemerge in subsequent years, which suggests an unrecognized means of overwintering or a persistent infection of host species. One California serogroup virus, Snowshoe hare virus, has been isolated from snowshoe hares during the winter months when mosquitos are absent, suggesting long-term infection of the host may contribute to overwintering of this virus. However, for most orthobunyaviruses, persistent infection seems unlikely as immunity and seroconversion within host populations is typically widespread and quite effective at clearing infection.

Host and Viral Determinants of Disease-Species Barriers

Outbreaks can occur when a virus crosses a species barrier and infects a new, immunologically naïve species. Factors that control the crossing of this barrier include species-specific factors, such as cell receptors, elements of the immune response, and host factors required for viral replication. The orthobunyavirus NSs protein can specifically shut down mammalian, but not mosquito, host cell transcription by inhibition of the RNA polymerase II which allows the virus to suppress the innate immune response. This demonstrates species-specific antagonism that correlates with host disease as these viruses typically cause disease in mammalian hosts. Furthermore, studies of select ruminant-infecting versus human-infecting orthobunyaviruses have shown species-specific restriction by BST2/tetherin expression in the host, but it is unknown whether this is a major species determinant across all orthobunyaviruses. Laboratory evidence suggests that orthobunyaviruses that cause disease in humans (Oropouche) can reassort with those that cause disease in animals (Schmallenberg). Thus, depending upon the genetic factors of novel reassortant viruses, orthobunyaviruses could in principle jump species to cause an outbreak in a naïve host.

Clinical Features and Pathogenesis

Orthobunyaviruses can cause disease in animals and in humans (Table 2). Disease in animals is most commonly reported in livestock, although zoonotic transmission is known to be important for some viruses suggesting wild populations may suffer from disease as well. Human disease ranges from mild febrile illness to hemorrhagic fever to acute encephalitis. In some cases, infection can be fatal. Here we will discuss by serogroup several of the more recent and consistently relevant orthobunyaviruses that cause disease. Not all serogroups will be represented as some do not contain known disease-causing viruses.

Anopheles A

In humans, *Tacaiuma virus* is known to cause myalgia, arthralgia, headaches, chills and weakness that lasts from days to a week with a sudden onset. At least a transient viremia is associated with infection as virus has been isolated from the blood of an acute patient. Although the number of infections per year is unknown, seroprevalence studies show specific antibodies in 0.5%–1% of residents in endemic areas.

Bunyamwera

Bunyamwera virus (BUNV) is present on both the African and South American continents and known to infect both humans and animals. High viremia is known to occur in some mammals. Infection in ruminants can include spontaneous abortion and teratogenic effects. In Brazil in 2013, two horses were found to be infected with BUNV and both developed neurological disease and died. Symptoms in humans are typically mild including fever, arthralgia and rash, but young or immunocompromised individuals can develop severe encephalitis. In endemic areas such as sub-Saharan Africa, seroprevalence has been recorded as high as 82% in some areas.

Ngari virus, as discussed above, has caused several large hemorrhagic fever outbreaks in humans in East Africa. It has also been associated with hemorrhagic disease in small ruminants such as sheep and goats although to a lesser extent. Disease in humans is characterized by sudden onset of fever and headache followed by gastrointestinal and mucosal hemorrhage. Up to 0.3% of cases have been fatal.

Bwamba

Bwamba virus causes clinically significant, although rarely fatal disease in humans. Symptoms are typically acute (4–5 days) and include general malaise associated with fever, headache, arthralgia and widespread or focal myalgia. Skin rash is also common. Intestinal tract involvement also commonly occurs, typically as diarrhea and can progress to intestinal and mucosal hemorrhagic complications in rare cases. Viremia is short-lived (1–2 days) which may be in part why this virus has not been associated with large outbreaks. However, seroprevalence in central and southern Africa can be quite high suggesting the virus is widespread. Neutralizing antibodies have also been found in birds, nonhuman primates and livestock although no associated disease has been described.

California

Jamestown Canyon virus is endemic to North America and has been associated with human disease in both Canada and the USA. Since 2013, the number of reported human disease cases has been increasing dramatically, leading many to consider it an emerging pathogen. However, these numbers may appear inflated because in the same year, the Centers for Disease Control instituted testing for Jamestown Canyon virus as a standard practice for all suspected arbovirus cases. Symptoms of infection are commonly flu-like, involving fever and myalgia. Severe cases may involve meningitis or meningoencephalitis that can result in hospitalization and in rare instances death. In reported cases, neurological complications are more common than febrile illness

and typical occur in older adults (40–59 years). Dependent on region, seroprevalence ranges from ~4%–40% in the human population and as high as ~90% in deer in Nova Scotia. No disease has been described in animals.

La Crosse virus (LACV) is also endemic to North America and is currently the leading cause of pediatric arboviral encephalitis on that continent. It is estimated that 300,000 infections occur annually with approximately 70 of those resulting in encephalitis. Encephalitic disease is far more common in younger individuals (16 years or younger), although immunocompromised individuals are also at risk. In less severe cases, symptoms include fever lasting 2–3 days, nausea, vomiting, headache, lethargy and fatigue. In severe disease cases, hospitalization is typically required, and patients can experience paralysis, seizures and coma. In these cases, CT scan and MRI can reveal areas of hyperintensity associated with diffuse edema and inflammation in the brain and spinal cord. Abnormal electroencephalographic readings are also common. LACV disease is lethal in up to 1.9% of cases, but most patients appear to recover completely. However, long-lasting neurological sequelae has been reported in as many as 15%–30% of cases upon post-discharge checkup more than a year later.

While LACV viremia in humans has not been well studied, seroprevalence can range as high as ~23% in some endemic areas. Viremia is known to occur in small mammals such as squirrels and chipmunks which are thought to be important zoonotic factors. Laboratory infections of young mice have demonstrated they are susceptible to viral-induced neurological disease, however, disease in wild-animals or livestock populations has not been reported.

Simbu

Akabane virus is broadly distributed and has been shown to cause disease in ruminants in Australia, the Middle East, Africa, southeast Asia and Japan. Clinical disease involves congenital defects in the offspring of infected host species. Disease incidence is greatly influenced by the timing during pregnancy when the maternal host becomes infected. Infection early during pregnancy (80–150 days for cattle, 28–56 days for goats and sheep) can result in up to 80% of offspring having abnormalities. Infection later in pregnancy carries a much lower risk. Offspring which are infected late in pregnancy and do develop disease are typically born alive but are often uncoordinated, unable to stand, or have flaccid paralysis and are blind. These animals can also exhibit cavitation of the cerebrum that causes hydranencephaly. Offspring infected early in pregnancy commonly exhibit arthrogryposis and sometimes torticollis, kyphosis and scoliosis as a result of muscle atrophy resulting from motor neuron loss. These animals often are severely hydraencephalic which can cause midterm abortion. If the offspring is carried to term, they typically do not survive birth and may cause obstetric complications for the mother that results in infertility and potentially death. In adult animals, the virus typically does not cause disease except for a very small number of cases in Japan. Asymptomatic infection typically results in long-lasting immunity that can be verified serologically. No human or wild ruminant disease cases have been reported, however, mice and hamsters can be infected in the laboratory.

Diagnosis

Serological Assays

Most orthobunyavirus infections remain undetected as they are asymptomatic or do not induce sufficient disease to warrant diagnostic testing. Other than seroprevalence surveys, the presence of virus infection is only analyzed once clinical signs of disease are observed in livestock or in human cases. Clinical signs are generally the first preliminary diagnostic and are based on the relevant disease signs correlating with those known for that specific virus as well as the endemicity of the virus for that region. Once an orthobunyavirus is suspected as a causative agent, typically a virus-specific Enzyme-linked Immunoabsorbant Assay (ELISA) will be performed, followed by plaque reduction neutralization tests (PRNTs) on serum from individuals or animals to test for virus-specific antibodies against the suspected orthobunyavirus. Specific diagnosis of an individual virus can be complicated by the cross-reactivity of antibodies to orthobunyaviruses within a serogroup. Additionally, these assays usually require two tests; one taken when a patient or animal first presents with clinical signs (acute phase) to determine if an antibody response is present, which indicates an infection but not necessarily an active infection. A subsequent analysis is performed a few weeks later (convalescent phase) to determine if antibody titers have risen, which would be indicative of a recent or an ongoing infection. A confirmed case is determined based on the four-fold rule: if the suspected orthobunyavirus has a four-fold higher antibody titer than the other serogroup viruses tested, and there is at least a four-fold rise in antibody titer from the acute serum to the convalescent serum, then the case is confirmed. Therefore, this type of test often does not result in an immediate diagnosis, which may limit its usefulness for an acute disease case.

Polymerase Chain Reaction (PCR) Detection of Viral RNA

Recent strides have been made in developing PCR assays that detect individual strains of viruses based on clear genetic differences. Although this assay is far more specific than sero-conversion assays, it also has limitations, as detection requires the virus to be present in the tested sample. For several viruses, including those that cause encephalitis, the presence of clinical signs occurs once the virus has been cleared from the blood and is present primarily in the central nervous system. A negative test in serum or in CSF

for virus cannot be taken as an indication that the patient does not have that infection. Thus, PCR analysis may work for positive confirmation for a virus, but not to rule out a specific orthobunyavirus.

Treatment and Prevention

There is no known specific treatment for orthobunyavirus infections. In the cases of human encephalitis infection, palliative care is given to patients to aid in recovery. Instead, the primary focus is generally towards prevention. Vaccine studies, including the identification of T cell and B cell epitopes that are protective against different orthobunyaviruses, are being investigated. However, vaccine development can be lengthy and costly. Additionally, the low incidence of disease, anti-vaccine public sentiment and potential for a higher mutation rate for RNA viruses may be barriers for establishing effective vaccination against different orthobunyaviruses. Another avenue for prevention is limiting arthropod transmission of virus to the vertebrate host. This can be done by using aerial spray pesticides to control mosquitoes and ticks in outdoor environments, removal of stagnant water where mosquitoes like to lay eggs and using insect repellent when in areas that are frequented by arthropods.

Acknowledgments

The authors would like to thank Ryan Kissinger and Durbadal Ojha for the design of the figures. This work was supported by the National Institute of Allergy and Infectious Diseases, Division of Intramural Research.

References

- Agerholm, J.S., *et al.*, 2015. Virus-induced congenital malformations in cattle. *Acta Veterinaria Scandinavica* 57, 54.
- Aguilar, P.V., *et al.*, 2011. Iquitos virus: A novel reassortant Orthobunyavirus associated with human illness in Peru. *PLOS Neglected Tropical Diseases* 5 (9), e1315.
- Aguilar, P.V., *et al.*, 2018. Genetic characterization of the patois serogroup (Genus Orthobunyavirus; Family Peribunyaviridae) and evidence that estero real virus is a member of the genus orthonairovirus. *The American Journal of Tropical Medicine and Hygiene* 99 (2), 451–457.
- Aguilar, P.V., *et al.*, 2010. Guaroa virus infection among humans in Bolivia and Peru. *The American Journal of Tropical Medicine and Hygiene* 83 (3), 714–721.
- Briese, T., *et al.*, 2006. Batai and Ngari viruses: M segment reassortment and association with severe febrile disease outbreaks in East Africa. *Journal of Virology* 80 (11), 5627–5630.
- Calisher, C.H., *et al.*, 1983. Identification of hitherto unrecognized arboviruses from Ecuador: Members of serogroups B, C, Bunyamwera, Patois, and Minatitlan. *The American Journal of Tropical Medicine and Hygiene* 32 (4), 877–885.
- Chiang, J.O., *et al.*, 2018. Characterization of the gamboa virus serogroup (Orthobunyavirus Genus, Peribunyaviridae Family). *The American Journal of Tropical Medicine and Hygiene* 98 (5), 1502–1511.
- Chowdhary, R., *et al.*, 2012. Genetic characterization of the Wyeomyia group of orthobunyaviruses and their phylogenetic relationships. *Journal of General Virology* 93 (Pt 5), 1023–1034.
- de Souza Lopez, O., *et al.*, 1975. Bertioga (Guama group) and Anhembi (Bunyamwera group), two new arboviruses isolated in Sao Paulo, Brazil. *The American Journal of Tropical Medicine and Hygiene* 24 (1), 131–134.
- de Souza, W.M., *et al.*, 2016. Molecular characterization of Capim and Enseada orthobunyaviruses. *Infection, Genetics and Evolution* 40, 47–53.
- Dutuze, M.F., *et al.*, 2018. A review of Bunyamwera, Batai, and Ngari Viruses: Understudied orthobunyaviruses with potential one health implications. *Frontiers in Veterinary Science* 5, 69.
- Evans, A.B., Peterson, K.E., 2019. Throw out the map: Neuropathogenesis of the globally expanding California serogroup of orthobunyaviruses. *Viruses* 11 (9).
- Groseth, A., *et al.*, 2014. Molecular characterization of human pathogenic bunyaviruses of the Nyando and Bwamba/Pongola virus groups leads to the genetic identification of Mojuí dos Campos and Kaeng Khoi virus. *PLOS Neglected Tropical Diseases* 8 (9), e3147.
- Hang, J., *et al.*, 2016. Genome sequence of bellavista virus, a novel orthobunyavirus isolated from a pool of mosquitoes captured near iquitos, Peru. *Genome Announcement* 4 (6).
- Harrison, B.A., *et al.*, 1995. Isolation of potosi virus from *Aedes albopictus* in North Carolina. *Journal of the American Mosquito Control Association* 11 (2 Pt 1), 225–229.
- Hughes, H.R., *et al.*, 2017. Full genomic characterization of California serogroup viruses, genus Orthobunyavirus, family Peribunyaviridae including phylogenetic relationships. *Virology* 512, 201–210.
- Jansen van Vuren, P., *et al.*, 2017. Isolation of a novel orthobunyavirus from bat flies (*Eucampsipoda africana*). *Journal of General Virology* 98 (5), 935–945.
- Meyers, M.T., *et al.*, 2015. Management factors associated with operation-level prevalence of antibodies to cache valley virus and other bunyamwera serogroup viruses in sheep in the United States. *Vector-Borne and Zoonotic Diseases* 15 (11), 683–693.
- Mohamed, M., *et al.*, 2009. Viruses in the Anopheles A, Anopheles B, and Tete serogroups in the Orthobunyavirus genus (family Bunyaviridae) do not encode an NSs protein. *Journal of Virology* 83 (15), 7612–7618.
- Nunes, M.R., *et al.*, 2005. Molecular epidemiology of group C viruses (Bunyaviridae, Orthobunyavirus) isolated in the Americas. *Journal of Virology* 79 (16), 10561–10570.
- Pachler, K., *et al.*, 2013. Molecular characterization of the African orthobunyavirus Ilesha virus. *Infection, Genetics and Evolution* 20, 124–130.
- Pollett, E., *et al.*, 2006. Characterization of Maguari orthobunyavirus mutants suggests the nonstructural protein NSm is not essential for growth in tissue culture. *Virology* 348 (1), 224–232.
- Saeed, M.F., *et al.*, 2001. Phylogeny of the Simbu serogroup of the genus Bunyavirus. *Journal of General Virology* 82 (Pt 9), 2173–2181.
- Savji, N., *et al.*, 2011. Genomic and phylogenetic characterization of Leanyer virus, a novel orthobunyavirus isolated in northern Australia. *Journal of General Virology* 92 (Pt 7), 1676–1687.
- Shchetinin, A.M., *et al.*, 2015. Genetic and phylogenetic characterization of tataguine and witwatersrand viruses and other orthobunyaviruses of the anopheles A, Capim, Guama, Koongol, Mapputta, Tete, and Turlock serogroups. *Viruses* 7 (11), 5987–6008.
- Soto, V., *et al.*, 2009. Complete nucleotide sequences of the small and medium RNA genome segments of Kairi virus (family Bunyaviridae). *Archives of Virology* 154 (9), 1555–1558.

- Tangudu, C.S., *et al.*, 2018. Evidence that Lokern virus (family Peribunyaviridae) is a reassortant that acquired its small and large genome segments from Main Drain virus and its medium genome segment from an undiscovered virus. *Virology Journal* 15 (1), 122.
- Zeller, H.G., *et al.*, 1989. Electron microscopic and antigenic studies of uncharacterized viruses. II. Evidence suggesting the placement of viruses in the family Bunyaviridae. *Archives of Virology* 108 (3-4), 211–227.
- Zhang, J., *et al.*, 2015. Molecular Characterization and Seroprevalence in Pigs of SC0806, a Cat Que Virus Isolated from Mosquitoes in Sichuan Province, China. *Vector-Borne and Zoonotic Diseases* 15 (7), 423–431.

Further Reading

- Collins, A.B., Doherty, M.L., Barrett, D.J., Mee, J.F., 2019. Schmallenberg virus: A systematic international literature review (2011–2019) from an Irish perspective. *Iranian Journal of Veterinary Research* 72, 9.
- Dutuze, M.F., Nzayirambaho, M., Mores, C.N., Christofferson, R.C., 2018. A review of Bunyamwera, Batai, and Ngari viruses: Understudied orthobunyaviruses with potential one health implications. *Frontiers in Veterinary Science* 12, 69.
- Elliott, R.M., 2014. Orthobunyaviruses: Recent genetic and structural insights. *Nature Reviews Microbiology* 12, 673–685.
- Romero-Alvarez, D., Escobar, L.E., 2018. Oropouche fever, an emergent disease from the Americas. *Microbes and Infection* 20, 135–146.

Relevant Websites

- <https://www.cdc.gov/jamestown-canyon/index.html>
Jamestown Canyon virus.
CDC.
- <https://www.cdc.gov/lac/index.html>
La Crosse encephalitis.
CDC.
- https://www.aphis.usda.gov/animal_health/animal_diseases/schmallenberg/downloads/SBV_EE_February%202014.pdf
Schmallenberg.
USDA APHIS.

Parapoxviruses (*Poxviridae*)

Hanns-Joachim Rziha, Eberhard Karls University of Tübingen, Tübingen, Germany

Mathias Büttner, Leipzig University, Leipzig, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of D. Haig, A.A. Mercer, Parapoxviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00461-1](https://doi.org/10.1016/B978-012374410-4.00461-1).

Introduction

Parapoxviruses (PPVs) are epitheliotropic viruses found worldwide. The individual viruses generally exhibit a narrow host range and infect *via* scarified or damaged skin and give rise to pustular lesions of the skin and occasionally the buccal mucosa. These lesions are associated with low mortality and high morbidity. In addition to a narrow host range, most of the PPVs can also infect humans. However, the zoonotic potency of PPV is mainly restricted to an occupational infection of localized self-limited manifestation without generalization and life threatening pathology. Therefore, PPV never attained the attention of some Orthopoxviruses (OPV) especially Vaccinia virus (VACV) and Variola virus.

There are numerous historical references to diseases of domesticated animals such as sheep and cattle that we would now suspect to be the result of infection by PPVs. These references include Jenner's 'spurious' cowpox which is likely to have been caused by the PPV, pseudocowpox virus (PCPV). In the latter part of last century, reports appeared in the scientific literature which recognized the distinct identities of the diseases caused by members of this genus. Following an extensive study of contagious pustular dermatitis of sheep, Aynaud produced a report in 1923, which included the observation that the disease could be transmitted by a 'filterable' agent. The isolation of PPV in cell culture was reported from 1957 to 1963. Detailed reports of the transmission of each disease to humans appeared in 1933 (ORFV), 1963 (PCPV), and 1967 (bovine papular stomatitis virus (BPSV)). The first molecular analyses of PPV genomes appeared in 1979 followed in 1989 by the first description of PPV DNA sequence and in 2004 the reporting of the full genome sequence for two strains of ORFV and one of BPSV.

Taxonomy and Classification

Parapoxvirus (PPV) represents a genus of the subfamily *Chordopoxvirinae* of the family *Poxviridae*. Originally, classification of different PPV species was based on the host and/or the pathology of the disease and genomic restriction profiling, but now sequence homologies are particularly taken into account. Currently PPV encompasses four species including the type-species *Orf virus* (ORFV) with the synonyms contagious pustular dermatitis virus or contagious ecthyma virus, the *Bovine papular stomatitis virus* (BPSV), the *Pseudocowpox virus* (PCPV) synonymous with milker's nodule virus or paravaccinia virus, and the *Parapoxvirus of red deer* (PVNZ) first described in New Zealand but later also found in Europe. Tentative species of *Parapoxvirus* are Auzduk disease or camel contagious ecthyma virus, chamois contagious ecthyma virus, and Sealpox virus. Morphological virus particle similarity and limited partial DNA sequence homologies led to speculations about the presence of novel PPV, for instance the *Squirrel poxvirus* (SQPV) that, however, turned out to represent a new genus of *Chordopoxvirinae*. Whole viral genome sequencing will clarify such questions.

Host Range, Epidemiology, and Virus Propagation

Natural infection by ORFV has been reported in many wildlife and domestic small and larger ruminants such as bighorn, and thornhorn sheep, domestic and Rocky Mountain goat, Dall's sheep, chamois, ibex, Himalayan thar, musk-ox, reindeer, caribou, steenbok, camelids from which all are transferable to humans. Experimental inoculations have shown that monkeys are susceptible to ORFV but a wide range of other animals including mouse, rabbit, dog, cat, and domesticated chicken are resistant. BPSV and PCPV both establish infection in cattle and humans but all other species tested, including sheep, are resistant. Both, ORFV and PCPV can affect reindeer where ORFV infection caused the more severe lesions. The PPV of red deer induces only very mild lesions in sheep and so far was not found in other species. The PPV of seals has been reported in a range of seals and sea lions. PPVs do not produce lesions on the chorioallantoic membrane of the developing chick embryo.

The PPVs of cattle and sheep are widespread throughout the world, essentially wherever their host animals are kept. The viruses are maintained in populations by a combination of persistent (chronic) infection, frequent reinfection, and the environmentally resistant nature of the viruses.

ORFV shed in scab material can remain infective under dry conditions for lengthy periods (at least 4 months and possibly years) and infection of naive animals by virus persisting in heavily contaminated areas such as barns, yards, and sheep camps is likely to play a major role in maintaining the disease. One study has shown that if the scab material is ground up so as to release the virus, then exposure to field conditions quickly results in inactivation of ORFV. Several studies have shown that a productive



Fig. 1 Orf (*Ecthyma contagiosum*) in *Ovis aries*. Note the pustular lesions around the mouth and nares aggravated by bacterial super-infection.

infection can be established in animals which have recovered from a previous infection. Such reinfections result in lesions that are smaller and resolve more quickly than primary infections. This short-lived immunity is likely to contribute to the persistence of the disease.

In the case of PCPV, it is apparent that infection can be spread within dairy herds by contamination of milking machinery and milkers' hands. The introduction of procedures, which reduce damage to teats and improve general hygiene at milking, can control the spread of the disease.

The most widely used permissive cell culture systems are primary ovine or bovine cells derived from sources such as testis, skin biopsy and embryonic kidney, lung, and muscle. Passaging PPV in cell cultures rapidly (after six and more passages) leads to genomic alterations as a result of adaptation to *in vitro* conditions. There have also been reports of ORFV isolates adapted to growth in few cell lines, e.g., Vero cells, accompanied by substantial genomic changes and attenuation. Yields of infectious ORFV from cell culture tend to be 10- to 100-fold lower than those achieved with VACV.

Serologic Relationships

PPVs show extensive antigenic cross-reactivity, although monoclonal antibodies can be used to distinguish each of the species. There are also antigens shared with other poxvirus genera but there is no cross-protection between PPVs and either orthopoxviruses or capripoxviruses.

Clinical Features

PPVs cause inflammatory and/or proliferative lesions that are confined to the skin and oral mucosa with no evidence of systemic spread. Infection is initiated in abrasions and generally proceeds through an afebrile, self-limiting lesion that resolves within 3–9 weeks without leaving a scar. ORFV lesions (**Fig. 1**) are most generally seen around the mouth and nares; hence, the infection is commonly referred to as scabby mouth or sore mouth. Lesions are also observed on other parts of the body, for example, the coronet, udder, or vulva. Following experimental inoculation of scarified skin, lesions progress through erythema, papule, vesicle, pustule, and scab before resolving. Large, proliferative, tumor-like lesions can be observed. It is likely that these are a result of an immune impairment of the host animal.

Orf, Scabby Mouth, Contagious Pustular Dermatitis (CPD), Ecthyma Contagiosum

Lesions around the mouth can interfere with feeding or suckling and especially in young animals result in failure to thrive. Teat lesions can have similar effects through the inhibition of suckling. Lesions in sheep and goats can develop into tumor-like cauliflower proliferative erosions usually accompanied by secondary bacterial infection. Edema of the head and swelling of the regional lymph nodes are common but rather non-specific signs of severe progression. In severe cases, glossitis phlegmonosa and ulcerosa with secondary bacterial infections can lead to starvation especially in young animals. Most common is the *labial form* that is the basis for the name of the disease. Blisters and yellowish pustules that may reach pea size are formed on the lips and at the corners of the mouth and can extend up to the nose, ears and eyelids. The mild labial form heals within 3–6 weeks. Pustules can also develop on the udders of ewes shortly before the lambing period. Secondary bacterial infections cause complications. The *podal form* (scabby foot) can occur simultaneously with the labial manifestation or independently. Lesions develop at the

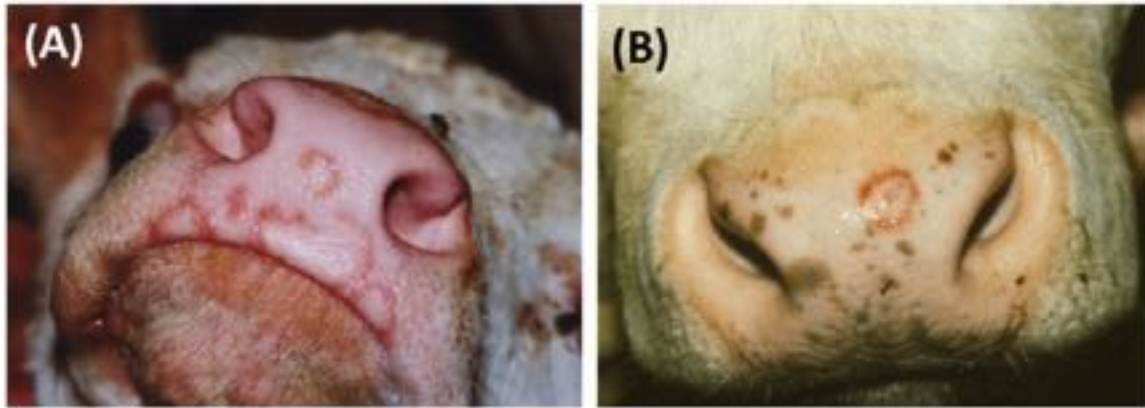


Fig. 2 Bovine Papular Stomatitis (BPS). (A) Cockade-like inflammatory zones at the muzzle. (B) Circle of inflammation around protected zone.

corner edges of the hooves, at the pasterns and in the hoof gaps. The ends of the digits are painful and lead to lameness and the refusal to stand. *The genital form* is less common. Typical pustule and crust formation occurs on the udder mostly developing into mastitis. Skin lesions also can occur on the inner leg, the labia or the prepuce.

Milker's Nodule, Paravaccinia, Pseudocowpox

It is probable that all PPVs are able to infect humans, although a human case of the PPV of red deer is not reported. Transmission to humans occurs readily although there is little evidence of human to human transmission. Nowadays human infections with ORFV are less frequent since milking of cows by hand became uncommon. Usually, Orf of man is an occupational disease of veterinarians, farmers, butchers and abattoir workers. Highest risk of infection is linked to handling of sheep fleece or wool and ritual slaughter of affected sheep, e.g. at religious feasts. Progression of the lesions is essentially as seen in sheep and cattle such that the infection is benign and confined to localized pustular lesions on the skin mainly at the hand and fingers at the points of virus entry. *Restitutio ad integrum* without leaving a scar usually occurs after a few weeks post infection. More severe progressive disease can occur in immune-compromised individuals. Severe reactions and proliferation, called giant Orf, have also been recorded in otherwise normal individuals in cases of burns and in cases of atopic dermatitis. *Erythema multififormae* reactions in the form of rashes on the backs of the hands and on the legs and ankles are common.

Bovine Papular Stomatitis (BPS)

Bovine papular stomatitis normally is a mild form of inflammation around the muzzle and mucous membrane of the mouth in large ruminants (Fig. 2). Horseshoe or coroner herds of inflammation are commonly seen on the muzzle often interrupted by zones of normal cells probably protected by interferon (ring zone phenomenon). However, more severe lesions can occur resulting in extension of confluent inflammation to the hard gum and far down to the esophagus. Signs of BPS may arise spontaneously in apparently healthy animals (especially young calves) without reports about skin damage or injury. Similar to infections with PPV of red deer subclinical persistence of the virus cannot be excluded. The trigger of clinical manifestation can be either unusual virulence of BPSV or immune suppression or both in synergy.

Parapox of Red Deer

Initially described in red deer of New Zealand it became evident that red deer in Europe and probably worldwide can become infected with this unique species of PPV. In Italy, clinical manifestation in red deer was seen as inflammation of the mouth whereas in Germany subclinical infection was diagnosed by viral DNA presence and virus isolation in cell culture. For red deer in New Zealand it is reported that PPV infection in the growing deer antler can affect antler growth and severely affect marketability of the product.

Physical Properties

Although exclusively tested with ORFV, PPVs seem to be resistant to desiccation and within scab material of infected animals; the dried viruses retain infectivity for at least 4 months and possibly years. Under laboratory conditions, infectivity can be maintained over many years. UV light, γ -irradiation, or heating at 56°C for more than 1 h can destroy infectivity of PPVs.



Fig. 3 Electron micrograph of a negatively stained characteristic ORFV particle showing the crisscross arrangement of the tubular core filaments ('ball-of-wool' appearance).

Virion Properties

The ovoid PPV virion has a length of 220–300 nm and a width of 147–170 nm and thereby are smaller than most of the brick-shaped virions of *Chordopoxvirinae*. PPV are surrounded by a tubule-like spiral structure resulting in the characteristic and distinct 'ball-of-wool' appearance of PPV by electron microscopy (Fig. 3).

Thin sections of virions reveal a lipoprotein bilayer surrounding a biconcave core and two associated lateral bodies. Further studies revealed ORFV to undergo morphogenesis similar to VACV. Thus, the intracellular mature virion (MV) is enveloped by a membrane derived from the endoplasmic reticulum. Subsequently two additional, probably Golgi-derived membranes form the wrapped virions (WV), which distribute by cell-to-cell infection. Eventually, the WV fuse with the cell membrane to release MVs as extracellular virions (EV). Under hostile conditions, e.g., non-permissive cells or adverse physical or chemical conditions, virus particles can lose their typical morphology showing unconventional shape such as disrupted arrangement of the spiral coil. Such particles may still be infectious.

PPV Genomes

The PPV core harbors a linear, double-stranded DNA genome with sizes of approximately 130 to 140 kbp encoding 124–134 predicted genes (Fig. 4), and an unusually high average G + C content of 63%–64%. The genome is terminally cross-linked by short, *ca.* 60 bases long single-stranded hairpin loops. Its genomic organization shows the typical poxviral structure: A central core region encoding essential genes conserved in position, spacing, and orientation, and variable genomic termini harboring genes dispensable for *in vitro* growth and involved in virus virulence and pathogenesis. Inverted terminal repeats (ITR) of varying sizes (2.6–3.5 kbp) embrace the genomic ends. During virus attenuation by cell culture passaging ITR have undergone genetic recombination between non-homologous segments, which resulted in transposition and/or duplication of DNA sequences. In addition, deletions of non-essential gene regions were found to create virus variants of reduced virulence. To date the complete genome sequences of 14 ORFV, 2 BPSV, and 2 PCPV strains have been published, and use of next generation genomic sequencing methods will increase their number in the near future (see also Diagnosis).

Viral Replication

As with all poxviruses, the PPVs replicate in the cytoplasm of the infected cell leading to a shut-down of the cellular protein synthesis. The viral cores not only harbor the DNA genome but also are able to produce early poxviral transcripts upon cell entry.

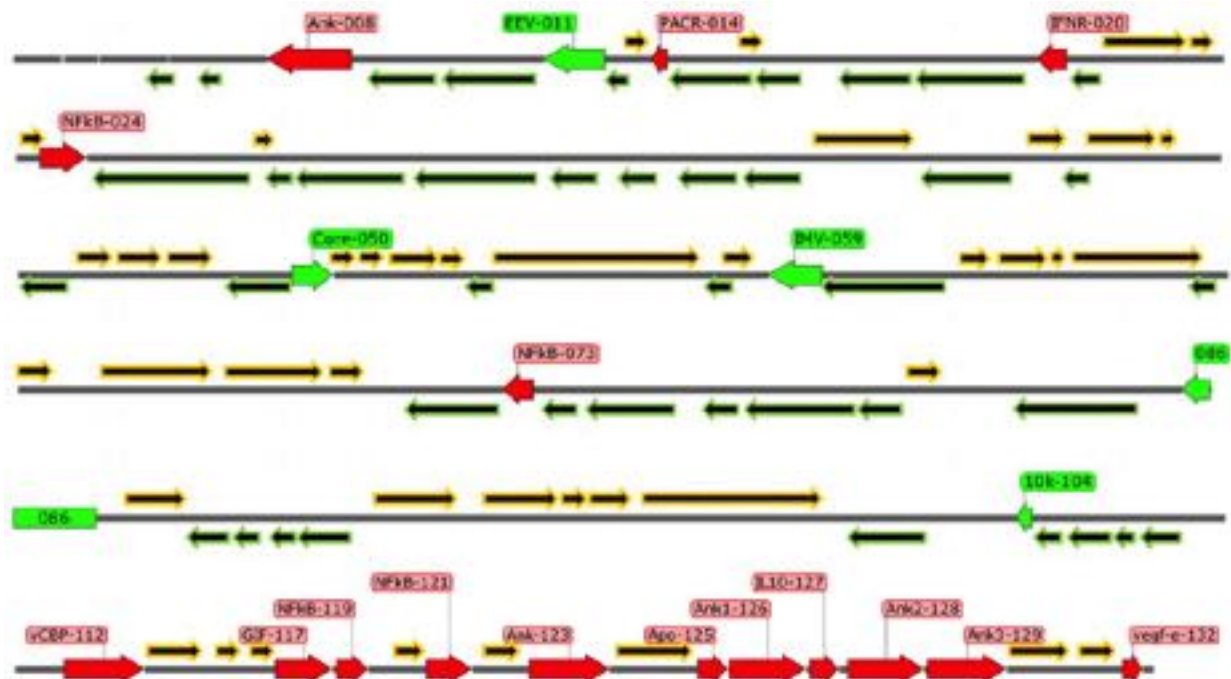


Fig. 4 Genomic map of ORFV strain B015. Arrows indicate gene transcription, rightward orientation above the line and leftward below the line. Colored in red are the virulence genes introduced in the text and in green those involved in virion maturation and mentioned in the text.

Early viral gene transcription is controlled by early promoters and according to their definition are also expressed in the presence of inhibitors of DNA synthesis. Sequences around the start and the end of PPV early genes contain A+T-rich control motifs very similar in structure and function to Orthopoxviruses. Analyses of vaccinia virus recombinants carrying large multigene DNA fragments of ORFV indicate the regulated cascade of early, intermediate, and late PPV gene expression is very comparable to Orthopoxviruses. PPV DNA replication begins approximately 4–8 h post infection (hpi) and viral particles are produced at 16–36 hpi continuing until at least 48 hpi.

Viral Proteins

SDS-polyacrylamide gel electrophoresis analyses of ORFV and PCPV virions resolve 30–40 polypeptides ranging in size from 10 to 220 kDa including 10–13 surface polypeptides. Significant antigenic overlap among PPV has been reported. However, pronounced cell association of the virus and the lack of additional specific antibodies limit the knowledge of electrophoretically demonstrable PPV proteins. For ORFV the most prominent polypeptides are found in the mol. wt. range of 30 and 56 kDa. The 39 kDa major envelope protein of ORFV represents an immuno-dominant antigen recognized by several independently derived monoclonal antibodies. This protein, formerly also named F1L, is encoded by the open reading frame ORF 059, and is a major heparin-binding protein potentially involved in binding to cells during virus entry. Another major structural, 37/42 kDa protein of ORFV exhibits strong amino acid similarity with the F13L gene product of the extracellular enveloped form of VACV. This B2L named ORFV protein is encoded by ORF 011 and found to induce antibody as well as T cell responses during ORFV infection. The product of ORF 104 is the so-called 10-kDa protein of ORFV, an orthologue of the VACV A27L gene product (14-kDa fusion protein). Removal of ORF 104 leads to disruption of the spiral tubular structure of ORFV. Recently, two other late core proteins were identified, which are encoded by ORF 050 and ORF 086, respectively.

Known or Putative Genes Involved in Pathogenesis and Virulence

Poxviruses, particularly Orthopoxviruses, contain extensive arrays of immunomodulatory genes to defend the virus against the host innate and acquired immune responses and to manipulate various intracellular signaling pathways, e.g., regulating apoptosis. Many of those immune evasion genes have been acquired during co-evolution of viruses with their hosts. During the last decade, those genes were increasingly identified also in PPV, especially in ORFV. Most of the factors influencing virulence, pathogenesis, or host range are dispensable for virus growth *in vitro* and are encoded in the genomic termini. In the following subsections, potential virulence factors of ORFV and their encoding open reading frame (ORF) numbers are briefly described.

VEGF-E – ORF 132

The first of these factors reported in ORFV was the functional homolog of the mammalian angiogenic vascular endothelial growth factor (VEGF). This early expressed, unspliced viral VEGF is unique for PPV. In contrast to the other members of the VEGF family it interacts only with VEGF receptor 2 (VEGFR-2), but not with either VEGFR-1 or VEGFR-3, and was classified as VEGF-E. Deletion of the ORFV VEGF-E gene proved its major role as a virulence determinant. VEGF-E can induce epidermal hyperplasia, which might contribute to scab formation and thereby continuously provide permissive cells for productive virus replication. Recent reports also indicate the role of VEGF-E in wound healing. Expression of ORFV VEGF-E can promote epidermal regeneration and reduce wound inflammation, and together with enhancing IL-10, skin repair in mice is positively influenced. In different PPV strains two VEGF-E variants, designated VEGF-NZ2 and VEGF-NZ7, can be distinguished. Despite these variants sharing only 41% amino acid identity, the structure of the functional protein domains is highly conserved. Since the VEGF-E gene displays a remarkably low G + C content as compared to the flanking gene regions, a more recent gene acquisition from the mammalian host is suspected.

Viral Interferon Resistance Protein – ORF 020

An interferon resistance protein (OVIFNR) is encoded by the early ORF 020 gene, a functional homolog of the VACV E3L gene. It inhibits the antiviral activity of interferon induced shutdown of cellular protein translation by binding to viral dsRNA and thus, preventing activation of the interferon-inducible protein kinase (PKR). In addition, OVIFNR acts as a PKR antagonist and prevents binding of PKR to the PKR activator. The gene product exists in 2 isoforms, the shorter, non-essential sh020, and the essential full-length 020 product, which is not dispensable for ORFV replication.

Viral dUTPase – ORF 007

At the left genomic side gene ORF 007 encodes a functional dUTPase. This virion-associated enzyme is suspected to be involved in ORFV virulence since it is missing in some attenuated ORFV strains.

Viral IL-10 – ORF127

This early gene of PPVs encodes a polypeptide with remarkable homology to mammalian, especially ovine IL-10. The mammalian IL-10 is a pleiotropic cytokine known to induce immune-stimulating or -suppressive effects in various cell types. IL-10 can prevent the synthesis of other cytokines, such as IFN- γ in blood monocytes or TNF α , IL-1 β and IL-8 in macrophages and keratinocytes. Additionally, T-helper 1 cell activation can be indirectly suppressed by decreased antigen processing and presentation or recruitment of dendritic cells (DCs). Several reports indicate that ORFV IL-10 displays a similar range of activities as its cellular counterpart. Immature DCs accumulate at the site of ORFV infection, which is possibly a result of ORFV IL-10-mediated blocking of DC maturation as well as decreasing interferon- γ production. This might be one factor contributing to the short-term ORFV-specific immune response. The secreted ORFV IL-10 represents a virulence determinant, a recombinant ORFV lacking the gene is attenuated compared to wild-type or to ORFV IL-10 gene-reconstituted virus in sheep. As mentioned above, together with the ORFV VEGF-E, the ORFV IL-10 can regulate skin repair, accelerate wound healing, and limit local skin inflammation.

Viral GM-CSF Inhibitory Factor (GIF) – ORF 117

The gene ORF 117 encodes a viral protein, which is a secreted dual inhibitor both of granulocyte macrophage colony-stimulating factor (GM-CSF) and of interleukin-2 (IL-2), and is therefore termed GM-CSF/IL-2 inhibition factor (GIF). Due to its dimeric structure, GIF can bind two molecules of the ovine but not of human target cytokines. Among other properties, GM-CSF regulates the recruitment, differentiation, and activation of macrophages, neutrophils, and DCs. Activated T cells are an important source of GM-CSF during immune responses to pathogens. IL-2 is produced predominantly by T cells and stimulates the expansion and activation of T cells and natural killer cells. Consequently, interaction of GIF with these cytokines can help ORFV to escape the host cellular immune attack. The GIF protein is conserved in ORFV strains and also expressed in other PPVs, though there is only approx. 40% amino acid identity.

Chemokine Binding Protein (vCBP) – ORF112

The early gene ORF 112 encodes a viral chemokine-binding protein, structurally and functionally related to the type II CC-chemokine binding proteins (CBP-II) of the Orthopoxvirus and Leporipoxvirus genera. CBP-II regulates recruitment of T cells, monocytes/macrophages, and DCs to sites of infection. In addition, vCBP binds to lymphotactin, a C-chemokine attracting T and B cells, and neutrophils. By disrupting chemokine gradients, the vCBP may inhibit immune cell trafficking to the site of infection and prevent key aspects of the anti-PPV immune response. After deletion of ORF 112 the knock-out mutants are attenuated in sheep showing its critical role in virulence and pathogenesis of ORFV. It seems reasonable to assume that the PPV encoded IL-10,

GIF, and CBP act in concert to counteract innate immune responses of the host. In addition, BPSV express a vCBP sharing 40% sequence identity with that of ORFV and which exhibits comparable inhibition of chemokines.

NF- κ B Inhibitors – ORF 002, 024, 073, 119, 121

Among others, the nuclear factor kappa B (NF- κ B) signaling pathway plays a vital role in regulating innate immunity, inflammation, programmed cell death (apoptosis), and keratinocyte proliferation. Therefore, not only poxviruses evolved strategies to counteract this essential regulatory pathway. To date five ORFV encoded polypeptides have been identified *in vitro* to target NF- κ B pathway signaling factors. The ORFV encoded inhibitors display no homology to those from other poxviruses. Gene ORF 002 encodes an early-late protein found in the nucleus of the infected cells and binds to NF- κ B-p65. Knock-out of this non-essential gene has no effect on ORFV virulence. ORF 002 is present in different ORFV and PCPV strains, but seems to be lacking in BPSV. The non-essential, early gene ORF 024 has no influence on virulence or disease severity. This gene product interferes with the core element of NF- κ B pathway, the IKK kinase complex. The product of the late gene ORF 073 was the first poxviral virion-associated NF- κ B inhibitor. The virion component ORF 073 acts transiently very early in ORFV infection without the need of *de novo* protein synthesis. Infection of sheep with an ORFV 073 knock-out mutant demonstrated the role of this non-essential gene in pathogenesis. The ORF 119 late gene product represents another virion protein unique for PPV. Similar to the ORF 073 protein, it also acts very early in ORFV infection, but unlike 073 it is also found late in infection in the cell nucleus. Elimination of ORF 119 gene resulted in ORFV attenuation in sheep. Interestingly, the loss of ORF 119 is found in attenuated ORFV strains after cell culture passaging. Finally, ORF121 encodes an early, cytoplasmically expressed inhibitor of NF- κ B signaling by counteracting the translocation of NF- κ B-p65 into the nucleus. This dispensable ORFV gene contributes to reduction of epidermal pathology and decreased cellular inflammatory response in the sheep skin.

Poxvirus APC/C Regulator (PACR) – ORF 014

The multicomponent anaphase promoting complex or cyclosome (APC/C) is a dominant regulator of the cell cycle. The poxvirus APC/cyclosome regulator (PACR) protein is expressed from the ORFV gene ORF 014. PACR seems to be incorporated into APC/C leading to its inhibition. It is suggested that PACR promotes an S-phase-like state of the infected cells favoring ORFV DNA replication. Although not essential for *in vitro* replication, ORF 014 deletion substantially reduces ORFV growth and plaque size. Amino acid comparison revealed a PACR homolog encoded by BPSV gene 013.

Ankyrin Repeat (ANK)/F-box Genes – ORFs 008, 123, 126, 128, 129

Ankyrin repeats (ANK) are protein consensus motifs crucial to facilitate protein-protein interactions, and are present in a plethora of cell proteins involved in cell–cell signaling, cytoskeleton integrity, regulation of transcription and cell cycle, inflammatory response, or protein transport. Those proteins are absent in most viruses except *Chordopoxviruses*. The poxviral ANK proteins share a common molecular structure including a protein motif at their C-termini, called F-box, which is also present in cellular proteins. This PRANC (Pox protein Repeats of Ankyrin C-terminal) domain is responsible for recruiting specific substrates to SCF1 ubiquitin ligase. For ORFV five ANK proteins are known and coded by ORF 008, 123, 126 (ANK-1), 128 (ANK-2) and 129 (ANK-3). The ORF 126 encoded ANK-1 protein co-localizes with mitochondria. A recent study demonstrated for the first time an influence of the five ORFV ANK proteins on the Hypoxia-inducible factor (HIF) pathway. The HIF pathway is not only crucial in the regulation of cellular responses to hypoxia, but also in angiogenesis- and anti-apoptotic programs. Upon ORFV infection, HIF gene expression is upregulated due to binding and sequestration of the cellular HIF inhibiting factor (FIH) by ORFV ANK proteins.

Inhibitor of Apoptosis – ORF 125

Programmed cell death or apoptosis is a highly effective innate response to eliminate infected cells. Thus, especially large DNA viruses evolved various immune evasion strategies limiting or preventing apoptosis. In addition to the anti-apoptotic effect of ORF 020 (described above), an anti-apoptotic protein is coded by the ORFV early gene ORF 125, which is essential for ORFV replication. This Bcl2-like inhibitor of apoptosis is directed to mitochondria to inhibit cytochrome C release, caspase activation, and cellular DNA fragmentation. Unlike other described poxviral Bcl2-like factors, the ORF 125 product targets cellular Bak. Recently, additional host proteins were identified to interact with ORF 125 indicating further biological functions.

Pathogenesis and the Host Immune Response to Infection

The histological sequence of events in the skin of sheep after ORFV infection is similar in primary and reinfection lesions, in spite of differences in the magnitude of the lesions and the time taken to resolve. Antibodies to ORFV envelope proteins have been used to detect ORFV antigen in epidermal keratinocytes, particularly those regenerating the damaged skin. Basal keratinocytes at the root of hair follicles can also contain virus. *In vitro* ORFV can persist in an infective state in keratinocytes for up to ten days. Some

infected cells show evidence of a ballooning-like degeneration. There is no evidence that ORFV infects other, non-epithelial cell-types *in vivo*. ORFV lesions often exhibit epidermal down-growths (rete formation) into the dermis. This is particularly marked in primary lesions. Another characteristic feature is extensive capillary dilation and proliferation. A-type intra-plasmatic eosinophil inclusion bodies can be seen in infected cells. By using its unique gene coding for vascular epidermal growth factor (VEGF-E, see also above), ORFV is able to promote endothelial cell proliferation and such create permissive cells for virus spread.

ORFV lesions contain a dense accumulation of immune and inflammatory cells underneath and adjacent to virus-infected cells. These include neutrophils, lymphocytes (T and B cells), and dendritic cells that express major histocompatibility complex class II antigens. A dense network of probably immature dendritic cells is characteristic of Orf lesions in sheep. The accumulating cells increase and decrease in number in parallel with the presence of virus in epidermal cells. The histology of human ORFV lesions is generally similar to that described in sheep. A comparison of ORFV and PCPV lesions in humans has not revealed any histopathological differences.

Immune Response to Infection

Typical for poxviruses PPVs induce a vigorous immune and inflammatory response in their hosts and have evolved to replicate in the presence of this response. In sheep experimentally infected with ORFV, studies of the skin and lymph draining into (afferent lymph) and out of (efferent lymph) local lymph nodes have demonstrated that activated CD4 + (helper) and CD8 + (cytotoxic) T cells, B cells, and antibodies are generated as part of the sheep-acquired immune response to infection. The cytokines generated in lymph in response to virus reinfection are typical of type 1 antiviral cell-mediated immune responses and include interleukin (IL) – 1 β , IL-2, tumor necrosis factor (TNF)- α , granulocyte-macrophage colony-stimulating factor (GM-CSF), interferon (IFN)- α , and IFN- γ . Studies of ORFV reinfection in sheep depleted of specific lymphocyte subsets or treated with the immunosuppressant drug cyclosporine-A indicated that at least CD4 + T cells and interferons are important components of the host-protective-response against infection. These studies also indicated that the cutaneous damage sustained during ORFV infection is due in large part to the virus rather than host immune-mediated responses. Sheep infected with ORFV mount detectable antibody responses to a small number of viral antigens but there is considerable individual qualitative and quantitative variation in the response. The lack of ORFV neutralizing antibody is exceptional. No apparent correlation between antibody titers and severity of viral lesions exists and passive antibody transfer does not confer protection against virus challenge.

Diagnosis

After a long period of restricted diagnosis based on clinical observation and morphological rating in electron microscopy, molecular methods have opened the capability of rapid and exact diagnosis even for differentiation at the species level. Most rapid differentiation from Orthopoxviruses still is the negative staining method of virions for electron microscopy; however, its low sensitivity requires high numbers of intact virus particles in lesion samples (about 10⁵ particles). Polymerase chain reaction (PCR), especially real time PCR, became a method of choice for PPV diagnosis because of the high sensitivity and specificity when appropriate primers are chosen. For PPV species resolution and molecular epidemiology less conserved genes or sequences at the terminal regions of the genomes can be targeted with a set of primers. Thus, even differentiation of ORFV isolates from sheep or goats is possible. Whole genome (WGS) or next generation sequencing (NGS) achieves most precise characterization and differentiation of PPV genomes. To avoid genome alterations caused by *in vitro* isolation of PPV, WGS is attempted to be applied directly out of lesion material as demonstrated with a case of PPV infection in seals. In general, differential diagnosis in affected domestic animals is very important since some diseases caused by notifiable infections such as Foot-and-mouth disease virus (FMDV), Bluetongue virus (BTV), BVDV /mucosal disease virus, Border disease virus, Malignant catarrhal fever virus, Lumpy skin and Sheep or Goat pox virus can cause similar lesions and must be excluded.

Natural antibody formation after PPV infection is highly restricted to the major envelope 39 kDa protein (see above). These antibodies do not possess neutralizing potency and in general, antibody detection in animals for diagnostic purpose is not constructive since PPV infections are so widespread. In the case of isolated PPV-free flocks of sheep or goats, antibody detection for status surveillance and maintenance makes sense. After infection of humans or for retrospective resolution of suspected human infections antibody detection by ELISA can be applied but whenever possible molecular diagnosis with fresh biopsies or lesion material should be preferred.

ORFV Vector Vaccines

ORFV has proven useful as a novel poxvirus vector platform for delivering heterologous antigens without the risk of a vigorous immune response against the vector backbone. The restricted host range, lack of systemic spread even in immuno-compromised animals, and induction of strong humoral and cellular immune responses to expressed antigens are excellent viral vector properties. Furthermore, even inactivated ORFV particles exhibit nonspecific immuno-modulatory and immune enhancing effects successfully applied in different pre-clinical disease models especially with non-permissive animals. Live ORFV recombinant

vaccines are capable of mediating excellent and long-term protective immunity against diverse viral infections. Most of the recombinants were created with the highly attenuated ORFV strain D1701-V by replacing its VEGF-E virulence gene by foreign antigen insertion. Additional new insertion sites have been established in this ORFV strain, which enables expression of at least three more different transgenes. Because an ORFV early promoter regulates transgene expression in strain D1701-V, no ORFV replication or production is needed to elicit protective immunity, even under non permissive conditions. Moreover, transgene substitution of the virulence genes ORF 024 or ORF 121 (see above) of the non-attenuated ORFV strain IA82 also allowed generation of protective recombinant vaccines, and deletion of ORF 113 as well as 116 allowed marker gene expression.

Oncolytic Potency

Besides the investigation of various other promising viruses, e.g., measles virus, among the poxviruses especially VACV, Myxoma virus, and ORFV are candidates for virotherapy of tumor diseases. The combined advantages of ORFV as a potent vector with immune stimulating efficacy and the capability to lyse tumor cells *in vitro* are favorable characteristics for future prospects in oncotherapy.

Prevention and Control

Vaccines against ORFV have been available for many years and are widely used to protect lambs against the debilitating effects of natural infection. These vaccines consist of live and essentially non-attenuated virus that is applied to a scratch on the skin of a leg. The ensuing infection does not interfere with feeding and provides significant protection against infection for some months. However, the scab derived from vaccination lesions is likely to contaminate the environment and contribute to the perpetuation of the disease. A few attenuated vaccine virus strains for parenteral application have been developed but they also induce only short lasting protection (up to six months), however the severity of reinfections can be effectively reduced. By analogy with many Orthopoxviruses including vaccinia virus inactivated ORFV does not mediate protective immunity as learned from many vaccination attempts. New vaccines that induce protection against Orf but do not shed infectious virus are highly desirable. This might be achieved by deleting genes encoding viral virulence determinants or by delivering the protective antigens of the virus in an appropriate way. Acyclic nucleoside analogs, such as acyclovir ACV or cidofovir CDV selectively interacting with poxviral DNA polymerase are effective against human PPV infection. Topical treatment with CDV resulted in complete regression even of giant orf lesion in immunocompromised patients.

Further Reading

- Fleming, S.B., Wise, L.M., Mercer, A.A., 2015. Molecular genetic analysis of Orf virus: A poxvirus that has adapted to skin. *Viruses* 7, 1505–1539.
- Friederichs, S., Krebs, S., Blum, H., *et al.*, 2014. Comparative and retrospective molecular analysis of Parapoxvirus (PPV) isolates. *Virus Research* 181, 11–21.
- Rziha, H.-J., Büttner, M., Müller, M., *et al.*, 2019. Genomic characterization of Orf virus strain D1701-V (Parapoxvirus) and development of novel sites for multiple transgene expression. *Viruses* 11, 127.
- Weber, O., Mercer, A.A., Friebe, P., *et al.*, 2012. Therapeutic immunomodulation using a virus – The potential of inactivated orf virus. *European Journal of Clinical Microbiology and Infectious Diseases* 32, 451–460.

Parechoviruses (*Picornaviridae*)

Sisko Tauriainen, University of Turku, Turku, Finland

Glyn Stanway, University of Essex, Colchester, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

cre Short RNA stem loop needed to generate the primer for PeV RNA replication.

Encephalitis Inflammation of brain tissue.

IRES (internal ribosome entry site) Region of the PeV genome, which is needed for initiation of translation.

Meningitis Acute inflammation of the protective membranes covering the brain and spinal cord, the meninges.

Meningoencephalitis An acute inflammation of the brain and meninges.

Myalgia Nonspecific pain or tenderness in muscles.

Myositis Inflammation of muscle tissue.

NPGP Motif associated with translation termination/reinitiation in several PeV species.

Orchiodynia Pain in the testicles.

Sepsis A systemic multi-organ infection, usually caused by bacteria. When it is caused by viruses, the condition is often called Sepsis-like disease or viral sepsis.

The first two Parechoviruses (PeVs) were discovered in a diarrhea epidemic in the 1950s and were named Echovirus 22 and Echovirus 23. They were classified as enteroviruses (EVs), but were reclassified as Parechoviruses and first renamed Human Parechoviruses (HPeV1 and HPeV2), but recently they have been renamed again as Parechovirus-A1 and Parechovirus-A2 (PeV-A1 and PeV-A2), respectively. It took several decades before the third type, PeV-A3, was discovered in 1999. Thereafter a further 16 types have been identified, increasing the total identified PeV-A types up to 19.

In addition to the PeVs found in humans, several PeVs infecting rodents have been detected. The PeV-Bs were first discovered in bank voles (*Clethrionomys glareolus*) in Sweden and were named as Ljungan virus. Another PeV was already detected in 1972 in an African wood mouse (*Hylomyscus* sp.), but the virus was later reclassified as an additional PeV species, PeV-C, based on the genome organization and sequence. The fourth species, PeV-D, contains a virus found in a pet ferret (*Mustela putorius furo*).

Classification

Parechoviruses are classified within the Realm *Riboviria*, Order *Picornavirales*, Family *Picornaviridae*, Genus *Parechovirus*. Phylogenetically, *Parechovirus* clusters with several other *Picornaviridae* genera, together forming a group currently called supergroup 4 (SG4), which in the future may formally be designated as a *Picornaviridae* Subfamily. Currently, genus *Parechovirus* consists of four species, *Parechovirus A* (previously called *Human parechovirus*), *Parechovirus B* (previously called *Ljungan virus*), *Parechovirus C* and *Parechovirus D* (PeV-A to D). Originally on the basis of serology, but now defined by amino acid identities within the virus capsid proteins, parechoviruses are grouped into types (originally serotypes). 19 PeV-A types are currently recognized. Five PeV-B types are known, while only one type has been identified for each of PeV-C and PeV-D.

Virion Structure

As in other picornaviruses, the parechovirus virion is composed of an icosahedral protein capsid enclosing the genomic RNA. The capsid is composed of 60 copies of each of three proteins, VP0, VP3, and VP1 (Fig. 1). These three proteins have the same core structure, an 8-stranded anti-parallel β -barrel. Copies of VP1 are clustered around the capsid 5-fold axis, while VP0 and VP3 alternate around the 3-fold and 2-fold axes. The average diameter of the virion is somewhat smaller than that of many picornaviruses and the surface is relatively flat, with no sign of the deep “canyon” that is the site of receptor binding in enteroviruses. However, in PeV-B five copies of a very large protrusion, made up of the C-terminal region of VP1, circle the 5-fold axis.

Parechoviruses were the first picornaviruses found to lack the cleavage of VP0 to give VP4 and VP2, and are now known to share this property with several more recently described *Picornaviridae* genera. In the picornaviruses where VP0 cleavage occurs, this takes place after assembly, as the final step of virion maturation and is necessary for stability and infectivity, so its absence in parechoviruses is intriguing. Another interesting feature of PeV is that the VP3 protein has a longer N-terminal region than most other picornaviruses and it contains several basic amino acids. When the structures of parechoviruses were first analysed, contacts between the capsid proteins and the genomic RNA were more evident than in other picornaviruses and extensive amounts of the RNA could be visualized, indicating essentially identical contacts between the RNA and capsid proteins. Contact regions include the VP3 N-terminus, which was found to be on the inside of the virion, and this protein may also serve to help neutralize the charge of the RNA to facilitate packaging. A number of RNA stem-loops, arrayed along the genomic RNA, and with a conserved

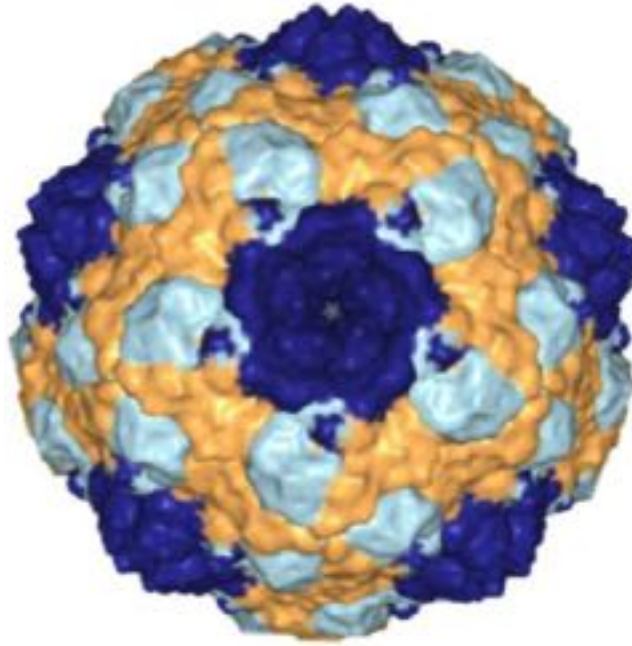


Fig. 1 Virion structure of PeV-A1. The icosahedral capsid is made up of 60 copies of 3 virus proteins VP0 (orange), VP1 (dark blue), and VP3 (pale blue). The image was generated from PDB ID: 4Z92 using NGL viewer and RCSB PDB. Reproduced from Kalynych, S., Pálková, L., Plevka, P., 2016. The structure of human parechovirus 1 reveals an association of the RNA genome with the capsid. *Journal of Virology* 90, 1377–1386, available from the RCSB PDB database, using NGL Viewer. Rose, A.S., *et al.*, 2018. NGL viewer: Web-based molecular graphics for large complexes. *Bioinformatics* 34 (21), 3755–3758, doi:10.1093/bioinformatics/bty419.

apical sequence motif, have been identified and may be the observed points of contact and possible packaging signals. The unusually ordered protein/RNA contacts are also seen in other picornaviruses where VP0 cleavage does not occur. This possibly indicates that these non-VP0 cleaving picornaviruses may follow a distinct assembly pathway where multiple capsid subunits align themselves along the RNA and then interact with one another to give an intact virus particle. Alternatively, a similar mechanism may occur in other picornaviruses, but some RNA/protein interactions are broken as a consequence of the structural change to the virus particle caused by VP0 cleavage, giving a less ordered RNA structure.

Parechoviruses are also unusual in lacking N-terminal myristoylation of VP0, thought to be important in interactions with membranes during early events in the infection of cells by many picornaviruses. This suggests that parechoviruses, and other picornaviruses lacking VP0 myristoylation, use fundamentally different pathways during entry and uncoating from picornaviruses with a myristoylated VP0/VP4.

Genome

Parechoviruses have a typical picornavirus genome organization (Fig. 2). A long (about 700 nucleotides in PeV-A) 5' untranslated region (5'UTR) is followed by a single open-reading frame (around 2200 codons) and a short 3'UTR (70–120 nucleotides). The genome is polyadenylated at the 3' end and a small protein, VPg, is covalently attached by an O-tyrosyl link to the 5' end. The genome is analogous to a cellular mRNA and it is translated to give a polypeptide, which is cleaved by a virus-encoded protease to give precursors and the final virus proteins. The picornavirus 5'UTR contains several RNA secondary structures, particularly within the extensive IRES domain, involved in translation of the RNA. Five structurally-distinct IRES types (I–V) have been identified among picornaviruses. PeV-A, -B, and -C have an IRES of type II, while PeV-D has a type IV IRES, suggesting that recombination has played a role in the evolution of the different genera. Clear evidence of recombination has been found within the large number of PeV-A isolates that have been sequenced and this may be of significance in the generation of epidemic or pathogenic strains.

The order of proteins encoded within the open reading frame is essentially conserved among picornaviruses. The parechovirus gene order is 1AB(VP0)-1C(VP3)-1D(VP1)-2A-2B-2C^{hel}-3A-3B(VPg)-3C^{pro}-3D^{pol} (Fig. 2). The more conserved proteins (2C^{hel}, 3C^{pro}-3D^{pol}) possess the typical motifs of a putative helicase, chymotrypsin-like protease and polymerase, respectively. 2B and 3A are more diverse among picornaviruses, but as in most picornaviruses both parechovirus proteins are highly hydrophobic. They are associated with intracellular membranes when over-expressed in mammalian cells. The PeV 2C is largely associated with lipid droplets. There are two variable loci within picornavirus genomes, where proteins of quite different types can occur, L, which is encoded before the capsid proteins in some viruses, and 2A. L does not occur in parechoviruses. The PeV-A 2A was found to be related to a group of cellular proteins and similar proteins are present at the 2A locus in several other picornaviruses. This type of

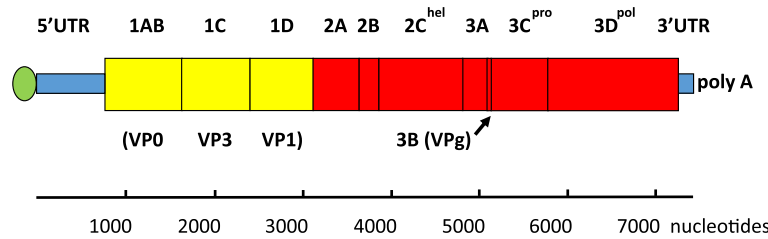


Fig. 2 The PeV genome as seen in the species PeV-A. The single stranded positive sense RNA genome is around 7300 nucleotides long in PeV-A and it encodes a single polyprotein. The open-reading frame is flanked by 5' and 3' untranslated regions (UTRs) and there is a genetically-encoded 3' poly A tail. A small protein, VPg, which is the product of the 3B region, is covalently attached to the 5' terminus (green circle). Following translation, the polyprotein is cleaved into several individual proteins. The structural proteins VP0, VP3, and VP1 are encoded by the 1AB, 1C, and 1D regions, respectively (yellow). The other non-structural proteins are encoded by the 2A-3D regions (red). The proteins encoded by the 2C, 3C, and 3D regions have helicase, protease and polymerase activities, respectively.

protein is termed Hbox/NC from two characteristic amino acid motifs, but its biochemical properties and function in virus replication remain largely obscure. The only property ascribed to the parechovirus 2A to date is binding to the 3'UTR of genomic and dsRNA, but its functional significance is not clear. An interesting difference of the other 3 *Parechovirus* species from PeV-A is that although these also encode an Hbox/NC type 2A protein, additionally they seem to possess a second type of 2A, a small protein characterized by an NPGP motif. Such a protein is associated with ribosome skipping, a translation termination/re-initiation event, leading to the polyprotein being produced in two pieces. These 3 species are said to encode 2A1 (NPGP type) and 2A2 (Hbox/NC type), although it is not clear whether 2A1 is a separate protein or if it remains part of the preceding protein, VP1, possibly accounting for part of the large 5-fold related protrusions seen in the PeV-B capsid structure.

Life Cycle

The cell and molecular biology of parechoviruses has not been studied extensively, but some specific details are known. Some of the PeV-A types have an RGD (arginine-glycine-aspartic acid) motif close to the VP1 C-terminus and this interacts with cell surface integrins during the entry process. Several RGD-dependent integrins have been found to be potentially utilized as the PeV receptor, including $\alpha v\beta 1$, $\alpha v\beta 3$, and $\alpha v\beta 6$, with most proof of $\alpha v\beta 1$ being the main receptor for PeV-A1. Entry of the virus into the cell seems to be via endocytosis into endosomes and acidification of the endosome seems to be vital for infection, presumably to trigger uncoating of the RNA. The receptors used by the members of PeV-A which lack an RGD motif (including PeV-A3) are not known, as is the case for viruses in PeV-B to -D, which also lack RGD.

Following the release of the genomic RNA, translation occurs and gives rise to the parechovirus polyprotein. Translation is initiated by a cap-independent, IRES-driven mechanism. This has been shown to depend on RNA structures within the type II IRES, particularly domains I and J/K. These probably serve as binding sites for proteins usually involved in cellular mRNA translation, as well as IRES-specific factors (ITAFS), but this has not been studied directly. It is assumed that all processing of the polyprotein in PeV-A is performed by 3C^{pro}, as 2A was shown not to have a protease activity, while PeV-B to -D have the additional "processing" activity of the NPGP-containing 2A1 protein.

RNA replication depends on stem-loops and pseudoknots close to the 5' terminus of the RNA, but little is known about the details of the process. As in other picornaviruses, the genome harbors a small RNA element, the cre, which is needed to template the 3D^{pol}-mediated addition of two uridine residues to VPg so it can be used as a primer in RNA replication. The picornavirus cre occurs in diverse locations in different genera or species and in PeV-A it is found in the VP0-encoding region, while in PeV-B it is in the 3B-encoding region. During infection, picornaviruses manipulate intracellular membranes and generate replication complexes formed from components of these membranes to produce nascent RNA. In parechovirus-infected cells, it has been reported that the Golgi disintegrates and the endoplasmic reticulum is modified. Replication complexes seem to contain Golgi markers, as well as the virus 2C protein, while some of the 2C locates to other membrane structures. In contrast to enteroviruses, PeV replication complex formation does not depend on COP-I, indeed COP-I is dispersed in PeV-infected cells. These results emphasize that there are significant differences between picornaviruses in intracellular replication events.

Later steps in infection, including assembly, whether there is some distinct pathway of maturation of the particle in the absence of VP0 cleavage and how the virus is released from the cell, have not been studied directly.

Epidemiology

PeV-As are very common and they infect mainly young children. Their seasonality depends much on the geographic location and may differ between the PeV types. However, seasonality is best known for PeV-A1 and PeV-A3, but a clear picture for other types is lacking. Most studies showed the peak of infections to be either in summer or in fall. Nordic countries have mostly a peak of

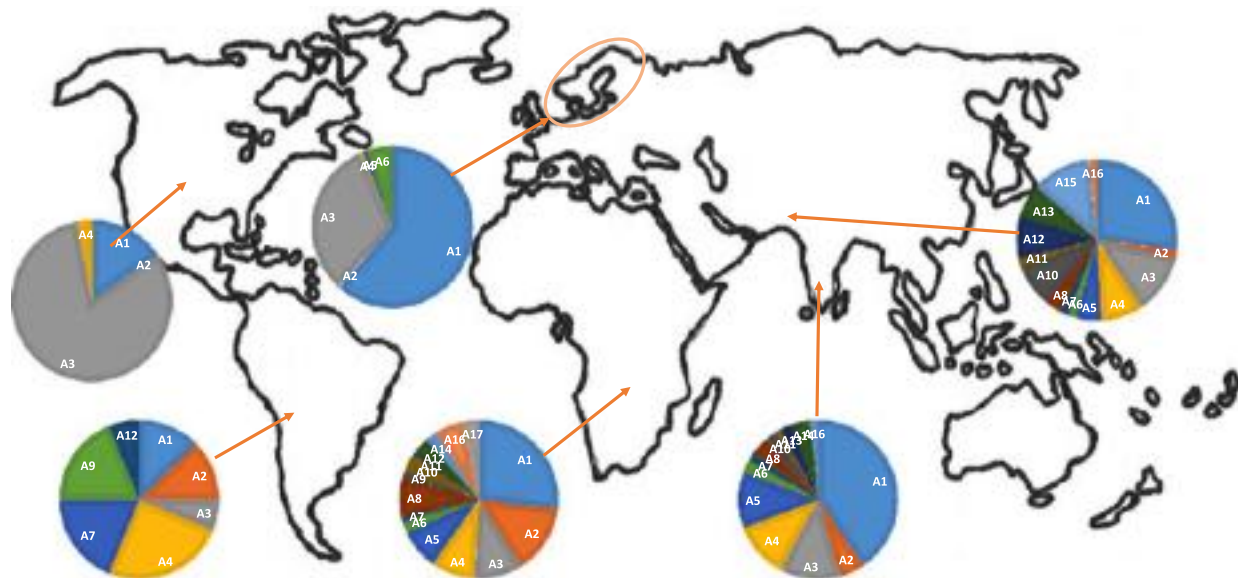


Fig. 3 PeV-A type distribution around the world. Pie charts show the type distribution of each PeV-A type, named in charts as PeV-A1 = A1, PeV-A2 = A2, etc. Studies from Malawi, Pakistan, and India are based on one or two large studies and represent background population. The Bolivian results are based on a small study population and thus does not likely represent real type distribution. The result from USA is based on clinical surveillance data and a small epidemiological series, thus PeV-A3 is overrepresented. In addition, the European data is partly based on a clinical data making PeV-A3 overrepresented. Distribution of viruses in charts: USA: 15% PeV-A1, 82% PeV-A3, 3% PeV-A4; Bolivia: 13% PeV-A1, 13% PeV-A2, 6% PeV-A3, 25% PeV-A4, 19% PeV-A7, 19% PeV-A9, 6% PeV-A12; Europe: 62% PeV-A1, 32% PeV-A3, 1% PeV-A4, 1% PeV-A5, 5% PeV-A6; Malawi: 27% PeV-A1, 14% PeV-A2, 10% PeV-A3, 9% PeV-A4, 8% PeV-A5, 2% PeV-A6, 1% PeV-A7, 8% PeV-A8, 2% PeV-A9, 3% PeV-A10, 1% PeV-A11, 3% PeV-A12, 2% PeV-A14, 6% PeV-A16, 3% PeV-A17; Pakistan: 27% PeV-A1, 2% PeV-A2, 12% PeV-A3, 8% PeV-A4, 5% PeV-A5, 2% PeV-A6, 2% PeV-A7, 3% PeV-A8, 8% PeV-A10, 2% PeV-A11, 8% PeV-A12, 7% PeV-A13, 12% PeV-A15, 2% PeV-A16; India: 40% PeV-A1, 5% PeV-A2, 12% PeV-A3, 11% PeV-A4, 11% PeV-A5, 3% PeV-A6, 2% PeV-A7, 4% PeV-A8, 3% PeV-A10, 1% PeV-A11, 3% PeV-A13, 4% PeV-A14, 2% PeV-A16. Not typed viruses were left out of the results.

infections in fall and Central European countries have a peak in summer. Furthermore, PeV-A3 showed a biannual cycle in several countries, such as in the Netherlands, Scotland, USA and Australia.

PeV-As have been detected in all continents. PeV-A1 is especially common and it mainly infects children under the age of 2 years. Most studies show PeV-A1 to be the dominating type, but there are big differences in its prevalence. European studies show up to 70%–90% of detected PeVs to belong to PeV-A1, in contrast to studies from Asia, South America, and Africa, which show PeV-A1 to represent 25%–60% of detected PeVs. This is the case in studies from either healthy population or samples from children with varying symptoms, such as diarrhea, respiratory symptoms or fever.

However, in studies done on samples from neonates with serious infections such as sepsis-like disease, meningitis or encephalitis, the most prevalent type is PeV-A3. There are several reports of PeV-A3 outbreaks around the world. In some regions, PeV-A3 seems to be very common and the virus circulates in the population with infections peaking every 2–3 years.

In Europe, USA, Australia, and Japan, the type distribution is limited to PeV-A1 to PeV-A6. Types PeV-A7 to PeV-A19 can be found in Asia, South America, and Africa (Fig. 3). However, there is no clear picture on the prevalence of these types.

Although PeV-A types are mostly detected in humans, several types have also been isolated from non-human primates. PeV-A types 1, 4, 5, 12, 14, and 15 have been detected in Rhesus macaques, which were in close contact with humans. PeV-A types are seen as human viruses, but additional studies are needed to evaluate the possibility of animal reservoirs of PeV-A viruses.

Seroprevalence

The seroprevalence of PeV-As differs very much between virus types and populations. However, serological studies have only been done in a few countries and they have mostly focused on PeV-A1. A study comparing the seroprevalence of PeV types 1–6 in Finnish and Dutch populations found PeV-A1 and PeV-A2 to be very common and the least common type was PeV-A3, with only minor differences between these populations for most types. Other studies do not support PeV-A2 to be common, and the virus is isolated only occasionally. In addition, PeV-A3 is found to be common in different populations, however, there might be a bias in the frequency of PeV-A3, since this type is the one seen in the clinics and therefore it is reported most often.

Despite differences seen in studies done in different regions, PeV-A1 is the most common type found around the globe, reaching seroprevalence up to about 82%–99% in adults. Types 4, 5, and 6 are also fairly common, with seroprevalences of 35%–75%. Studies from Japan show seroprevalences against PeV-A3 of 80%–85% in children of 5 years of age, but lower seroprevalence in adults. A recent study compared PeV-A3 seropositivity in three populations, the Netherlands, Australia and USA,

and the study showed prevalence in these populations ranging from 58% up to 83%, confirming the frequent findings of PeV-A3 in the clinics. Seroprevalences of types 7–19 are unknown.

Outbreaks

The first reported hospital outbreak caused by PeV-A1 was in 1964–1965 in a nursery of premature infants. The babies presented with upper and lower respiratory symptoms. After this, several outbreaks have been reported from different continents. After the discovery of PeV-A3, most reported outbreaks have been found to be caused by this type. This is quite understandable, since this is the most pathogenic type and may cause severe symptoms in neonates. So far the largest reported outbreaks have been in Japan starting from 2006 and in Australia from 2013, which are described below.

Japan

Japan has experienced several PeV-A3 outbreaks; these were detected in 2006, 2008, 2011, 2014, and 2016. The numbers of PeV-A3 positive cases have risen from the early epidemics (2006, 2008) with around 50 cases, to more than 300 in 2014 and 2016. This is either due to rising infection rates of PeV-A3 or just due to increased diagnostic activity in response to better awareness of the virus. Even in 2017, a year of lower PeV-A3 activity, there were about 100 cases detected. The outbreaks almost followed the biannual cycle seen elsewhere, except the outbreak in 2011, with a three-year gap on either side.

In addition to the severe central nervous system (CNS) infections and sepsis-like disease in neonates, these outbreaks manifested with myalgia and myositis both in children and adults. The association to myalgia and myositis has not been described elsewhere. It could be due better awareness among Japanese scientists and doctors of PeV-A3 infections, since this virus was first detected in Japan. However, there could also be some genetic or other factor in the Japanese population making them more susceptible to PeV-A3 caused myalgia.

Australia

So far, the biggest and best-documented epidemic is the one from Australia in 2013 to 2014. The succeeding epidemics occurred in 2015 to 2016 and 2017 to 2018. All epidemics had their peak in November or December, which are summer months in the southern hemisphere. This pattern fits the earlier finding from the Netherlands and UK, where PeV-A3 has a biannual seasonality and most cases were seen in summer months.

The 2013/4 epidemic started in October and lasted until February 2014. More than 198 cases were identified and all typed samples were of PeV-A3 type. The epidemic was detected in New South Wales and it spread nationwide. The epidemic confirmed many earlier findings, such as most affected children were under the age of 3 months (93%) and both boys and girls were infected.

The 2015/6 epidemic spread over a larger area from Western Australia to Sidney and the island of Tasmania. It had a slightly longer duration, from September 2015 to May 2016. The 2017/8 epidemic was first noted in Victoria, spreading thereafter to Queensland and subsequently nationwide. In both the 2015/6 and 2017/8 epidemics, more than 200 infected individuals were identified.

Parechovirus Species B-D

The first characterized non-human PeV was Ljungan virus (a member of PeV-B) found in bank voles (*Clethrionomys glareolus*) in Sweden near the Ljungan River. Now five PeV-B types have been identified in several rodents including montane vole (*Microtus montanus*), southern red-backed vole (*Myodes gapperi*), yellow-necked mice (*Apodemus flavicollis*) and Eurasian red squirrels (*Sciurus vulgaris*), as well as in bird feces in Japan. PeV-Bs have so far been detected in several European countries, USA and Japan. Information from other regions is lacking.

There has been speculation on whether PeV-B is able to infect humans and cause several different diseases, such as myocarditis, type 1 diabetes and intrauterine fetal death. However, no clear proof of human infections has been shown. PeV-B has never been detected from human stool samples, despite several large studies, with 2000–4000 screened samples. Seroprevalence studies have shown a significant proportion of positivity, but as there is no other evidence of PeV-B being able to infect humans, it is speculative whether this positivity is due to cross-reactive antibodies against antigenically related viruses. Thus, it still remains to be proven if PeV-Bs can infect humans and cause disease in humans. Other species (PeV-C, PeV-D) have so far only been detected in rodents.

Clinical Features

The clinical picture of PeV-A1 and PeV-A3 is fairly well described in the literature. For other types the information is limited or lacking. One reason for this may be that these viruses mostly cause asymptomatic or very mild infections and therefore they are not seen in the clinics, this is probably true for types PeV-A2 and PeV-A4 to PeV-A6. For types PeV-A7 to PeV-A19, the reason could be that these types are mostly found in Asia, South America and Africa and studies from these continents, apart from Japan, are scarce.

PeV-A1 was originally found to mostly cause fairly mild gastrointestinal and respiratory symptoms. However, in studies done after 1990 the proportion of PeV-As detected in these cohorts is quite low. Other symptoms that PeV-A1 has been connected to

include fever, cough, rash, and otitis media. PeV-A1 has also been detected in serious infection cases throughout the decades. Severe infections include sepsis-like disease, meningitis, encephalitis, Guillain-Barré syndrome, paralysis, myocarditis, and hepatitis. Most of these findings are individual detections or case reports. A recent study from Iran found that all viruses of 12 positive CSF samples were PeV-A1, which is a rare observation since after its discovery, PeV-A3 has been reported to be almost exclusively the CNS infection-causing type.

PeV-A3 is the type, which causes most severe infections and these are the best documented. Despite this, most PeV-A3 infections are mild or asymptomatic. Most common severe infections are sepsis-like disease, meningitis, encephalitis or meningoencephalitis. In Japan, PeV-A3 has caused myalgia in adults, but reports from other countries are lacking. In adults, clinical infection with any PeV-A is quite rare and most severe cases are seen in babies below the age of 3 months.

In the 2013/4 Australian epidemic the most common symptoms in severe PeV-A3 infections were: fever, tachycardia, tachypnea and irritability/pain, which were present in more than 90% of patients. The next most common were rash and poor feeding (over 70%), followed by poor perfusion, diarrhea and lethargy, with each being present in more than 40% of patients. Hepatitis was seen in 33% of studied patients. Additionally, some patients (less than 15%) presented with upper respiratory symptoms, abdominal distention, vomiting, edema, seizures, apnea/preapnea or myoclonic jerks. The patients were described to be “red, hot, and angry”. Other studies have quite frequently reported seizures in PeV-A3 infected children, but definite numbers are hard to give with small patient numbers.

In addition, to having quite severe symptoms at the acute phase, it has been shown that many children suffering from PeV-A3 caused encephalitis have long-term sequelae. After one year of follow-up 7 of 8 children showed at least some developmental concern. Earlier studies already suggested long-term sequelae, but with no follow-up and low patient numbers, this could not be confirmed. Now, after the Australian epidemic there are follow-up studies, which will provide more information on long-term consequences. More studies are needed to clarify the whole picture of PeV-A3 caused diseases.

PeV-A3 has caused myalgia and myositis in Japanese children and adults. Symptoms are mostly muscle weakness and pain in arms and legs, sometimes even preventing patients from walking, as well as fever, loss of appetite, sore throat and diarrhea. Most adult male patients presented also with orchiodynia. Most of these cases had elevated creatine phosphokinase (CPK) levels, but no elevation of C-reactive protein (CRP). Most of the myalgia patients have been between ages 20–60 and are predominantly male. One report suggests the cause of myalgia to be an elevation of IL-6 by the PeV-A3 infection, since it is known that IL-6 causes myalgia, but this theory needs further studies. So far, the patients have recovered in about a week and no long-term sequelae have been reported.

PeV-A types 2, 4–6 are rare in clinical cases. There was an outbreak of PeV-A4 in Helsinki in 2012 with a sepsis-like disease. This virus was discovered to be a recombinant, but it is not known if the recombination was the cause of more severe symptoms. Mostly PeV-A types 4–6 are seen in fever of unknown causes, upper respiratory infections and diarrhea. There are a few reported cases where PeV-A6 has been connected to severe CNS infections, but number of cases have been very low. In addition, the first detection of PeV-A6 was from a child with Reye's syndrome, but the connection to this disease has not been confirmed.

The clinical symptoms of PeV-A types 7–19 are mostly unknown and the types are rarely found. They have been detected in stool samples of children with or without diarrhea in Asia (except Japan), South America and Africa but the numbers are still low. It is possible that these types mostly cause asymptomatic infection and are thus seen rarely in the clinics. It is just as likely that these types are undiagnosed in developing countries.

Virus Genetics and Association to Disease

The Australian epidemic-causing PeV-A3 strain was shown to be a recombinant virus. This was also the case for the sepsis-like disease-causing PeV-A4 strain in Finland. Despite genetic analyses comparing PeV-As involved in CNS-infections, sepsis-like disease or myalgia to those causing minor symptoms, no clear evidence of genetic differences explaining differences in the pathogenicity have been identified.

Pathogenesis

Research on how PeV-As cause disease is scarce. It is not even known what cell types PeV-As infect in the CNS. White matter damage in the brain of encephalitis cases has been seen, but this does not necessarily mean that these cell types are directly infected. In cell culture settings it has been demonstrated that PeV-As can infect many different human cell lines, such as neuronal cells (SH-SY-5Y), astrocytes (T98G, U373-MG), lung (A549) and colon epithelial cells (HT-29).

Neutralizing antibodies are important in preventing secondary PeV-A infections and in protecting new-born babies in their first months of life by maternal antibodies. As PeV-A1 is very prevalent, and according to some studies almost all mothers are positive for PeV-A1 neutralizing antibodies, it is not surprising that severe PeV-A1 caused infections are not often seen in neonates. This is in contrast to what is found for PeV-A3, where a good proportion of mothers were negative for PeV-A3 antibodies (30%–90%). This difference in the presence of neutralizing antibodies could be one explanation why most severe infections seen in neonates are caused by PeV-A3. However, this does not explain why older children do not get severe infections. It is not known if the brains of an infant is more susceptible to PeV-A infections than the brains of an older child.

Diagnosis

Earlier, when PeVs were still typed as enteroviruses and the diagnosis was done in cell cultures, PeV-A1 and PeV-A2 were detected as part of enterovirus (EV) diagnostics. When molecular methods were introduced to enterovirus diagnosis, PeV diagnosis went from bad to worse, as EV RT-PCR does not detect PeVs.

PeV-A diagnosis these days is mostly done by RT-PCR followed by sequencing to establish the PeV-A type. The diagnostic RT-PCR is based on primers amplifying a conserved region in the 5' UTR and the sequencing is based on either the VP1 region or the VP3/VP1 junction. Mostly those laboratories that have introduced PeV in their routine screening use in-house RT-PCRs and there are numerous publications describing the methods. However, today there are also commercial PeV-A nucleic acid detection kits available, either as a PeV-A specific RT-PCR or PeV-A is included in multiplex PCR panels for the detection of meningitis or respiratory pathogens.

Although viral culture and isolation is an important tool for further studies, the diagnosis should not be based on cultures. PeV-A3, which is the most clinically important PeV-A, does not grow very well in the cells mostly used for diagnostic viral culture and even successful growth does not always produce a clear cytopathic effect. Furthermore, PeV-A types 7–19 might not be detected, since it is not known whether they grow in cell cultures, if they produce CPE or which cells should be used for virus culture.

Diagnosing PeV-As has proven beneficial in several studies and the outcome of knowing the infective agent has reduced the length of stay in hospitals and unnecessary use of antibiotics.

Treatment and Prevention

There are no specific antiviral medication or vaccines against PeV-As. There is an urgent need for medication since a great proportion of children suffering of severe PeV-A sepsis-like disease or CNS infection need intensive care. Recovered patients are often left with long-term problems and some infections even lead to death. Medication could help in reducing the number of children left with long-term sequelae.

The problem with a preventive vaccine would be to decide when to vaccinate, since most children get the severe infection in their first weeks of life. The best option may be to vaccinate mothers against PeV-A3, which in the best case would lead to protective maternal antibodies.

At the moment the only treatment option is intravenous administration of immunoglobulins (IVIG). This has been successfully used in clinical settings, but the success of treatment depends on the presence of specific antibodies, which could vary from batch to batch. Successful treatment was accomplished in a child suffering from PeV-A1 associated dilating cardiomyopathy. PeV-A1 is very common in the population and therefore IVIG is almost guaranteed to be positive for PeV-A1 antibodies. In contrast, according to several studies, seropositivity against PeV-A3 is not very high in European countries and therefore IVIG might lack PeV-A3 antibodies and thus not work.

There is a clear need for antiviral medication against PeV-A3, but research to develop medication is not very active. Only a few *in vitro*-studies have been published. Studies have revealed potential compounds such as: 2-C-methylcytidine (2'CMC), a nucleoside analog, as well as itraconazole (ITZ) and posaconazole (PSZ), which are FDA-approved antifungal compounds. Of these 2'CMC reduced PeV-A1 and PeV-A3 growth in cell culture, but ITZ and PSZ only reduced the growth of PeV-A3. However, these substances could be clinically useful as the need for medication is the greatest for PeV-A3.

Several compounds known to inhibit EV replication *in vitro* have been tested, but these compounds show no effect on PeVs. This leads to the conclusion that the PeV-As lifecycle differs greatly from the EV lifecycle and therefore more studies are needed for targeted drug design against PeV-As to be possible. Furthermore, the finding that ITZ and PSZ inhibited PeV-A3 growth but had no effect on PeV-A1 suggests that there may be significant differences in the lifecycle of different PeV-A types.

Further Reading

- Aizawa, Y., Izumita, R., Saitoh, A., 2017. Human parechovirus type 3 infection: An emerging infection in neonates and young infants. *Journal of Infection and Chemotherapy* 23, 419–426.
- Britton, P.N., Jones, C.A., Macartney, K., Cheng, A.C., 2018. Parechovirus: An important emerging infection in young infants. *The Medical Journal of Australia* 208, 365–369.
- Britton, P.N., Khandaker, G., Khatami, A., *et al.*, 2018. High prevalence of developmental concern amongst infants at 12 months following hospitalised parechovirus infection. *Journal of Paediatrics and Child Health* 54, 289–295.
- Brouwer, L., Karelehto, E., Han, A.X., *et al.*, 2019. High frequency and diversity of parechovirus A in a cohort of Malawian children. *Archives of Virology* 164 (3), 799–806.
- de Crom, S.C., Rossen, J.W., van Furth, A.M., Obihara, C.C., 2016. Enterovirus and parechovirus infection in children: A brief overview. *European Journal of Pediatrics* 175, 1023–1029.
- Kalynych, S., Pálková, L., Plevka, P., 2016. The structure of Human Parechovirus 1 reveals an association of the RNA genome with the capsid. *Journal of Virology* 90, 1377–1386.
- Kolehmainen, P., Siponen, A., Smura, T., *et al.*, 2017. Intertypic recombination of human parechovirus 4 isolated from infants with sepsis-like disease. *Journal of Clinical Virology* 88, 1–7.
- Mizuta, K., Aoki, Y., Komabayashi, K., *et al.*, 2018. Parechovirus A3 (PeV-A3)-associated myalgia/myositis occurs irrespective of its genetic cluster: A longitudinal molecular epidemiology of PeV-A3 in Yamagata, Japan between 2003 and 2016. *Journal of Medical Microbiology* 68 (3), 424–428.
- Olijve, L., Jennings, L., Walls, T., 2017. Human parechovirus: An increasingly recognized cause of sepsis-like illness in young infants. *Clinical Microbiology Reviews* 31, 1–17.

- Shakeel, S., Dykeman, E.C., White, S.J., *et al.*, 2017. Genomic RNA folding mediates assembly of human parechovirus. *Nature Communications* 8, 5.
- van der Linden, L., Wolthers, K.C., van Kuppeveld, F.J., 2015. Replication and inhibitors of enteroviruses and parechoviruses. *Viruses* 7, 4529–4562.
- Zheng, L., Wang, F., Huang, J., Xin, H., 2015. Evaluation of the association of zoonotic Ljungan virus with perinatal deaths and fetal malformation. *Birth Defects Research, Part C* 105, 81–85.

Relevant Websites

- <http://www.europic.org.uk/>
Europic '79.
- <http://www.picornastudygroup.com/>
Picornaviridae Study Group.
- <http://www.picornaviridae.com/>
Picornavirus.
- <https://talk.ictvonline.org/taxonomy/>
Taxonomy
International Committee on Taxonomy of Viruses.
- <https://viralzone.expasy.org/>
ViralZone root
ExPASy.

Parvoviruses of Carnivores, and the Emergence of Canine Parvovirus (*Parvoviridae*)

Colin R Parrish and Ian EH Voorhees, Cornell University, Ithaca, NY, United States

Susan L Hafenstein, Pennsylvania State University, Hershey, PA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Carnivora An order of eutherian mammals that share a common ancestor, which includes wolves, dogs, cats, raccoons, bears, weasels, hyaenas, seals, and walruses.

Translesional polymerases Specialized DNA polymerases that are expressed in cells that contain damaged DNA. This damage repair mechanism is error-prone.

Classification (Compact)

Family – *Parvoviridae*

Genus – *Protoparvovirus*

Species – *Carnivore protoparvovirus 1*

Despite infecting a broad range of host species within the Order Carnivora, all *Carnivore protoparvovirus 1* variants share over 97%–98% nucleotide sequence identity, forming a single distinct clade (Fig. 1). The most closely related viruses known outside this group are only about 60% or less identical (*Ungulate PPV 1* and the *Rodent PPV 1* and *Rodent PPV 2*). Within the *Carnivore protoparvovirus 1* clade, the two major subclades comprise the viruses related to FPV, and those within the lineage of viruses that emerged in dogs in the 1970s which are identified as CPV-variants no matter which hosts they are isolated from (Fig. 2).

Virion Structure and Genome

Like other members of the family *Parvoviridae*, the *Carnivore protoparvoviruses 1* virion consists of a single linear strand of DNA of about 5000 nucleotides in length packaged within a non-enveloped icosahedral protein capsid. The capsid is initially assembled from 60 copies of a mixture of two different forms of a capsid protein (virus protein 1 and 2 (VP1 and VP2)). The two forms are produced by alternative splicing of the same message, sharing a common C-terminal sequence of 568 residues, with the VP1 form containing an additional 145 residue N-terminal domain. VP2 makes up a majority (about 90%) of the 60 capsid proteins, and in a full (DNA containing) capsid around 24 residues may be cleaved from the N-terminus by host proteinases to form VP3. The robust capsid is a compact $T = 1$ icosahedron that is ~26 nm in diameter with a prominent raised region or spike surrounding each threefold symmetrical axis, an elevated cylindrical pore-like structure around each fivefold axis formed by 5 antiparallel beta strands, and depressions around the fivefold axis and also spanning the two-fold axis.

The genome of the virus is composed of two major coding regions, with the NS1 and NS2 proteins encoded within the left-hand end of the genome (the 3'-end of the minus sense DNA strand that is packaged into these viruses). The VP1 and VP2 proteins are encoded by right hand ORFs. The messenger RNAs that encode the NS and VP proteins are expressed from the P4 and P38 promoters, respectively. Like the VP forms, the NS1 and NS2 are produced by alternative splicing. Additionally, a small (58–68 residue) alternatively translated non-structural protein (SAT) is encoded within the VP1/2 mRNA, overlapping the sequence near the VP2 N-terminus.

Life Cycle

Infection of host cells requires that the virus capsid bind to the transferrin receptor type-1 (TfR) on the surface of the cell, which mediates rapid uptake into the cell through clathrin-mediated endocytosis. The TfR is a large dimeric receptor that is expressed on the surface of all cells, where its normal function is to bind iron-loaded (holo) transferrin, and to transport that into the endosomal system of the cells, where the iron is released at the low pH of the early or recycling endosome, after which the TfR and iron-free (apo) transferrin recycle back to the cell surface. The TfR is assembled as a dimer of three domains, the apical domain, the protease-like domain, and the helical domain that forms the dimer interface (Fig. 3(a)).

Once inside the cell the capsid enters the cytoplasm, and the ssDNA (likely without the capsid) enters the nucleus, where the 3'-end of genome in a palindromic hair-pinned form acts as a template for DNA synthesis to generate a dsDNA form. Genome replication occurs via a variation of the rolling-hairpin replication process, that involves nicking of the double stranded replicative form of the DNA by the viral NS1 protein. The 3'-end of the DNA generated is then extended by one or more of the host cell DNA polymerases to create a new double strand, displacing the second strand when replicating on double stranded

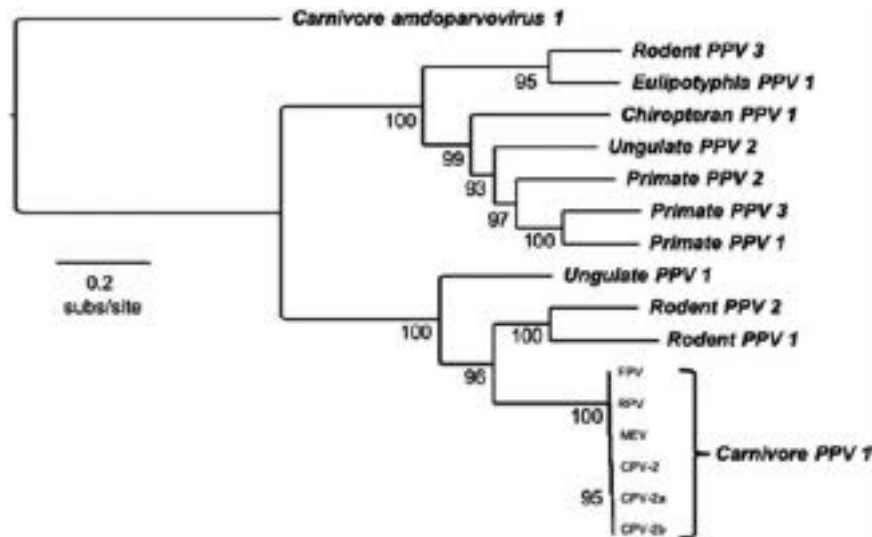


Fig. 1 Evolutionary relationships among species within the genus *Protoparvovirus*. A related parvovirus, *Carnivore amdoparvovirus 1*, within the genus *amdoparvovirus* is used as an evolutionary outgroup. For each species, the nearly full genome NCBI Ref sequences was used. In addition, multiple isolate sequences were included for *Carnivore PPV 1* to display the key variants within the species. Numbers represent node support (out of 100). GTR PhyML tree. 100xBS replicates. Abbreviations: PPV = protoparvovirus; FPV = feline panleukopenia virus; RPV = raccoon parvovirus; CPV = canine parvovirus.

DNA. While host DNA polymerase delta appears to be the primary polymerase involved in the viral genome replication, other canonical DNA polymerases or translesional polymerases may also play roles in viral genome replication when they are present. Replication produces both positive and negative strand viral genomes; however, negative strands are preferentially packaged by these viruses (other members of the Parvoviridae may package either strand with similar efficiency). Packaging is mediated by the helicase domain of the viral NS1 protein, and for these viruses the ssDNA DNA is fed in a 3'-to-5' orientation into the preassembled capsid through one of the fivefold capsid pores.

Epidemiology

Carnivore protoparvovirus 1 was first described as the etiological agent of severe enteric disease in cats as early as the 1920s. By the 1940s, the virus was recognized in number of wild and domestic non-canine carnivores around the world, and referred to as feline panleukopenia virus (FPV), mink enteritis virus (MEV), or raccoon parvovirus (RPV) depending on the host infected. In early 1978, a variant of those viruses (named canine parvovirus (CPV)) or canine parvovirus type-2 (CPV-2) emerged to cause disease in domestic and wild dogs and that reached pandemic proportions by the end of that year. Today, natural infections by various strains of these viruses are found in cats and dogs, as well as domestic or captive populations mink, foxes, and raccoon dogs, as well as in most major groups of wild terrestrial carnivores. It is unclear whether Pinnipeds (seals, sealions and relatives) or bears are susceptible to infection or to clinical disease by these viruses. Occasionally indications of infection of other hosts have been reported, including reports in captive rhesus macaques, and in insectivores such as hedgehogs.

Testing for viral DNA or antibodies indicates that *Carnivore protoparvovirus 1* is very widespread in nature among susceptible carnivore populations. Likewise, phylogenetic studies of the viruses and their transmission reveals widespread global dissemination of virus genotypes as there are close similarities between viruses that are derived from distant locations. Such rapid global dissemination was well documented during the original emergence and spread of CPV in dogs in 1978 and 1979, as dogs and related canids (coyotes and wolves) throughout the world became infected over a period of a few months.

There are also significant DNA sequence similarities between the viruses found infecting different host species, suggesting frequent cross-species transmission (Fig. 2). Domestic dogs and cats remain the largest host populations for these viruses around the world by a wide margin, and are assumed to maintain the most common virus strains in wide circulation. It is likely that the viruses in other hosts mostly originate from spill-over infections of viruses from cats or dogs, perhaps followed by local transmission within that new host - there is evidence of viral lineages in wild raccoons and in farmed mink and arctic foxes, likely because those populations are large enough to reveal outbreaks and to sustain transmission for weeks or months.

As mentioned above the viruses appear to be very widespread in nature due to the very high levels of shedding that occurs in the feces of infected animals, and the stable capsid which can persist in an infectious form in the environment for weeks or even months. Most puppies or kittens acquire maternal antibodies to the virus from their mother, mostly in the form of colostral IgG,

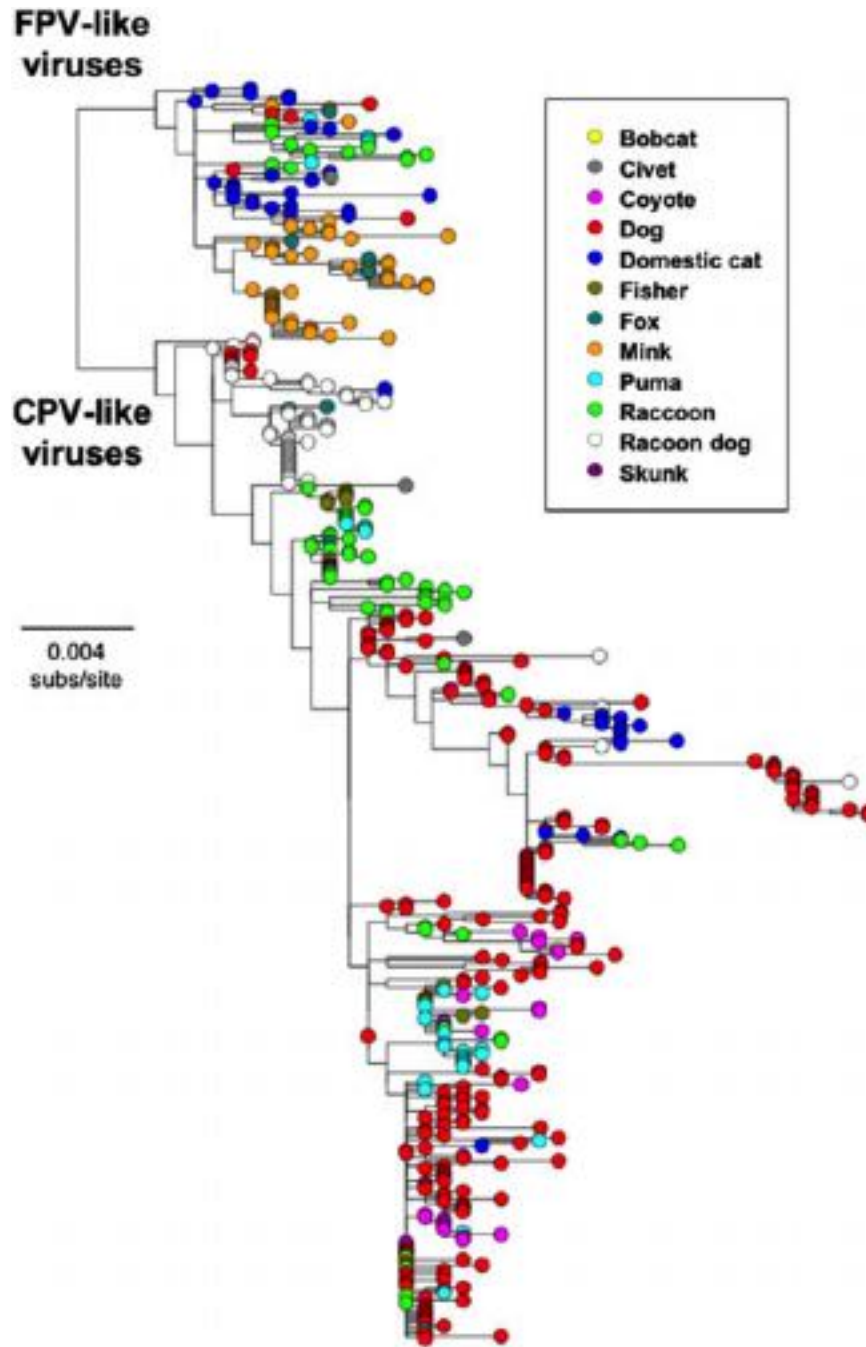


Fig. 2 Evolutionary relationships among viral variants within *Carnivore protoparvovirus 1* collected from various host species. Tree tips colored according to host species.

protecting them from infection by virus for the first 8–16 weeks of life (depending on the levels of maternal antibodies transferred). For domestic animals, those maternal antibodies also prevent infection by vaccines, such that the young become receptive to vaccine and to wild virus in the environment at about the same time. There has therefore continued to be relatively high levels of parvovirus infection in young animals even in the face of widespread application of vaccines, and the virus also infects most wild animals in the early months of life. Both FPV-like and CPV-like strains of virus continue to circulate in parallel, with most viruses in cats being FPV-like, and all strains in dogs being CPV-derived viruses.

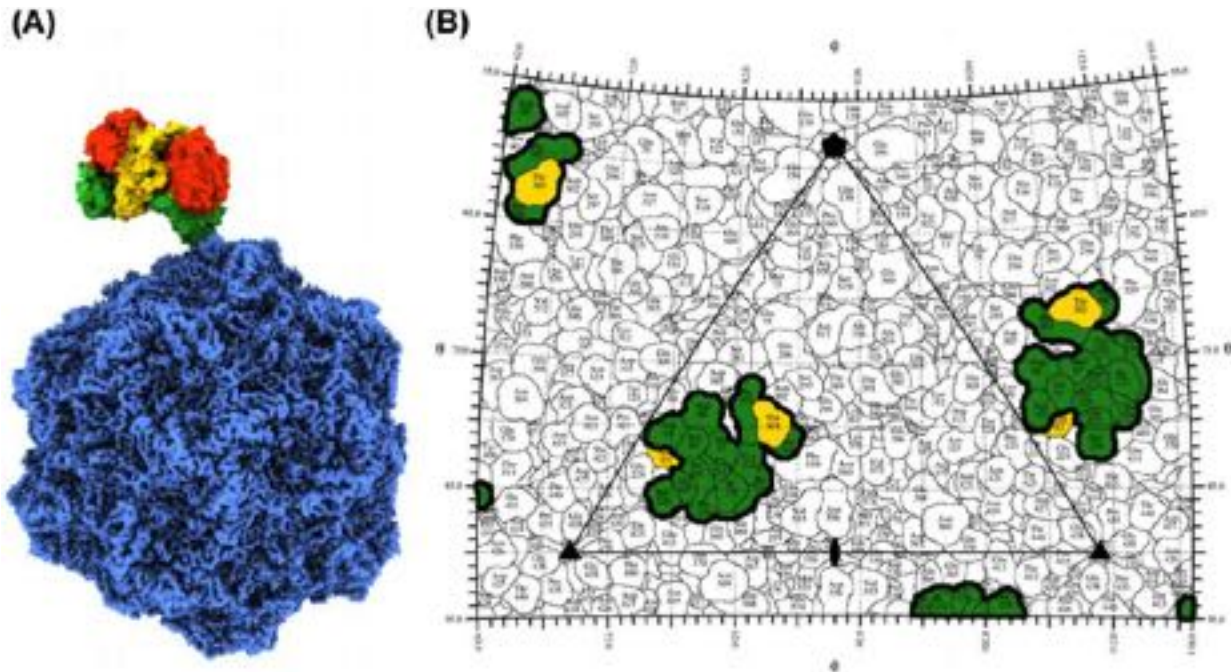


Fig. 3 (A) Structure of the CPV capsid (colored blue) in complex with host TfR (colored red = protease domain; yellow = helical domain; and green = apical domain) determined by cryo-electron microscopy. (B) Roadmap of CPV capsid with the TfR footprint indicated (bold outline). Changes that occurred at VP2 residues 301 and 93 (colored in yellow) expanded the host range of the virus, enabling binding and entry of host cells using the canine TfR.

Host Range Variation, Its Control, and the Emergence of CPV

Host range variations are often associated with small numbers of amino acid differences in the viral capsid, which affect the structure and interaction with the host TfR. When viruses or viral DNA are recovered from various hosts (e.g., raccoons, foxes, and mink), they often show host range variation that appears to be related to differences in binding to the host receptor. The best-documented example of host range variation was the acquisition of the canine host range by the ancestor of the CPV lineage. When FPV viruses were compared to the earliest CPV for which sequences are known (from 1978), at least 2 key mutations were identified that alter residues in the viral capsid and together control the ability of the virus to bind to the canine TfR.

Mutagenesis studies and the cryo-electron microscopy structures of the TfR from a host related to the dog – the black backed jackal – in complex with the CPV-2 capsid, have provided high resolution structural detail and context to this host-range acquisition. Those revealed that the TfR bound the capsid through a relatively small interface (Fig. 3), with parts of two loops of the TfR apical domain interacting with portions of three loops exposed on the surface of the CPV capsid. The canine TfR differs from the related forms found in other carnivore hosts in that it has an N-linked glycosylation in the apical domain, which blocks the binding of viruses that are similar to the FPV. The mutations in the virus that resulted in it gaining the canine host range occur within or adjacent to the TfR footprint (Fig. 3(b)), altering the flexibility of the capsid loops that interact with the TfR. These changes allowed the modified receptor to interact with the apical domain around the additional glycosylation site, mediating cell infection. The original strain of CPV, termed CPV type-2, spread world-wide during a period of around 1 year, when it was replaced by a variant form (identified as CPV- type-2a) which was essentially the CPV-2 genotype, but had acquired an additional group of 5 mutations which altered sites on the surface of the capsid, further expanding the virus's host range.

Pathogenesis and Clinical Features

The pathogenesis of the viruses are all similar, and they infect through the oral or oro-nasal routes, and likely establish their initial infections in the tonsils and retro-pharyngeal lymphoid tissues. Virus spreads free in the plasma, as well as likely by binding to or infecting lymphocytes, and therefore infects cells in many secondary tissues. A defining feature of the pathogenesis of these viruses is that they only replicate in cells that are undergoing mitosis, because the polymerases, nucleotides and other cellular components required for viral DNA replication are only expressed in cells that are dividing. This also results in a strong age-relationship to the disease due to the reduced proportions of dividing cells as animals age. Two diseases are observed only after the virus infects neonatal animals, due to the populations of cells that are replicating shortly after birth. Those diseases are cerebellar hypoplasia in cats due to virus replication in and killing of the germinal cells of the cerebellum, thereby resulting in permanent ataxia, and viral

myocarditis in dogs due to virus infection of myocardial cells, which results in a slowly developing myocarditis where host responses target the infected myocytes, resulting in cardiac failure a few weeks after infection. Both of those diseases are rare because most kittens or puppies are now protected during that age by maternal antibodies. In animals that are infected at two months of age or later (after maternal immunity has waned), many develop only mild or subclinical infections. However, a proportion of infected animals develop severe disease due to viral replication in the rapidly dividing cells that regenerate the epithelium of the small intestinal villi. The epithelial cells are continuously turned over, and loss of the regenerating cells results in the loss of epithelial integrity, so that infected animals develop severe diarrhea and vomiting, as well as loss of the intestinal barrier.

Diagnosis

There are many diagnostic tests that can be applied to detect these viruses, or to provide evidence of their infection. The virus may be detected in feces using point-of-care ELISA tests that involve antibody capture of the viral particles from feces (or sometimes from pharyngeal swabs), followed by a second reagent that detects the bound virus. That test depends on the presence of high levels of free virus in the feces, so sensitivity is lost after host antibodies are produced that bind to the virion, or once viral shedding falls to low levels. Other tests include virus isolation into tissue cultures of feline cells (which are susceptible to all strains of virus), or PCR detection of the viral DNA. The PCR test is very sensitive and can detect the DNA of actively replicating virus, as well as DNA in the tissues that represents previous viral infection but where infectious virus is no longer present. Quantitative PCR may be used to distinguish between those circumstances, as false-positive DNA tests are otherwise very common since most animals have some residual DNA.

Treatment

No specific treatment is known that effectively targets the virus or its replication. The cerebella hypoplasia and myocarditis that develop after neonatal infection are irreversible once infection and replication of the virus occurs. For animals with enteritis the treatment is mainly supportive, including replacing the fluid and electrolyte losses from the diarrhea and vomiting. Antibiotics may be given to forestall secondary infections, but their efficiency is marginal in most cases.

Prevention

The most effective prevention involves the use of live attenuated vaccines that are given by parenteral routes, and which replicate systemically and induce the production of highly protective immunity that appears to last for the life of the animal. However, like the natural infection, the vaccine is blocked by the presence of maternal antibodies, so that it is necessary to wait until animals are between 8 and 16 weeks of age when the maternal immunity has waned sufficiently to allow virus infection and replication. Since animals may be exposed to wild type viruses from the environment during the same period, continuing disease is seen even when vaccines are used diligently.

Further Reading

- Barrs, V.R., 2019. Feline Panleukopenia: A re-emergent disease. *Veterinary Clinics of North America: Small Animal Practice* 49, 651–670.
- Carmichael, L.E., 2005. An annotated historical account of canine parvovirus. *Journal of Veterinary Medicine. B, Infectious Diseases and Veterinary Public Health* 52, 303–311.
- Lee, H., Callaway, H.M., Cifuentes, J.O., *et al.*, 2019. Transferrin receptor binds virus capsid with dynamic motion. *Proceedings of the National Academy of Sciences of the United States of America* 116, 20462–20471.
- Parker, J.S., Parrish, C.R., 1997. Canine parvovirus host range is determined by the specific conformation of an additional region of the capsid. *Journal of Virology* 71 (12), 9214–9222.
- Parrish, C.R., 1995. Pathogenesis of feline panleukopenia virus and canine parvovirus. *Baillière's Clinical Haematology* 8, 57–71.
- Reed, A.P., Jones, E.V., Miller, T.J., 1988. Nucleotide sequence and genome organization of canine parvovirus. *Journal of Virology* 62 (1), 266–276.
- Voorhees, I.E.H., Lee, H., Allison, A.B., *et al.*, 2019. Limited intra-host diversity and background evolution accompany 40 years of canine parvovirus host adaptation and spread. *Journal of Virology* 94, e01162.

Polioviruses (*Picornaviridae*)

Philip D Minor, St. Albans, United Kingdom

© 2021 Published by Elsevier Ltd.

Introduction

A funerary stele from about 1300 BCE currently in the Carlsberg museum in Copenhagen shows the priest Rom with a withered single limb and down-flexed foot typical of motor neuron destruction caused by poliovirus infection. It is considered to be the first documentary evidence for an infectious disease of humans. Two highly effective vaccines were developed in the 1950s, and a major program of the World Health Organization (WHO) is underway which may make poliomyelitis the second human disease to be eradicated globally after smallpox.

The Virus Genome

The virus contains a single strand of messenger sense RNA of about 7500 nt. A long highly structured 5' noncoding region of about 740 nt which serves as an internal ribosomal entry site (IRES) is followed by a single open reading frame and terminates in a 3' noncoding region of *c.* 70 nt followed by a polyadenylate tract. The 5' end of the genome is covalently linked to a virus-encoded protein termed VPg. The single open reading frame encodes the structural proteins which make up the capsid (collectively termed the P1 region) followed by regions encoding the nonstructural proteins P2 and P3. P1 is divided into VP1, VP3, and VP0 (VP2 plus VP4). P2 is cleaved into 2A^{pro}, 2B, and 2C, and P3 into 3A, 3B (or VPg), 3C^{pro}, and 3D^{pol} by virus-encoded proteases. 3D^{pol} is the viral polymerase, but all proteins in the nonstructural part of the genome play a role in RNA replication, which also depends on RNA structural elements and host cell membranes which are extensively rearranged in the course of infection. A significant structural element involved in replication is the stem loop designated *cre* (cis acting replication element) which is essential to the initiation of genome replication but can be moved to any position by molecular manipulation to any site in the genome where it remains functional, as mentioned below.

The Virus Particle

The infectious virus particle consists of 60 copies each of VP1, VP2, VP3, and the smaller protein VP4 surrounding the viral genome; VP2 and VP4 are generated by autocatalytic cleavage of VP0 in the last stage of the maturation of the virus particle. The proteins are arranged with icosahedral symmetry such that VP1 is found at the pentameric apex of the icosahedron and VP2 and VP3 alternate around the pseudo-sixfold axes of symmetry of the triangular faces of the icosahedron. VP4 is located internally about the pentameric apex and is myristylated. The pentameric apex of the particle is surrounded by a dip or canyon into which the cellular receptor for poliovirus fits. The atomic structure of the virus was solved in 1985.

The Poliovirus Receptor

Although some strains of type 2 poliovirus are able to infect normal mice, only the higher primates and Old World monkeys are susceptible to infection by all serotypes as they possess a specific poliovirus receptor that is required for infection, and this species restriction is one of the factors that makes polio eradication theoretically possible. The receptor molecule was identified in 1986 and it is a three-domain membrane protein of the immunoglobulin superfamily termed CD155. It is necessary and sufficient for the infection of cells *in vitro*, and transgenic mice carrying the gene for the poliovirus receptor can be infected by all poliovirus types, developing paralysis when infected with wild-type virus. Such mice have been developed as alternatives to monkeys in the safety testing of live poliovirus vaccines.

The tissue distribution of the receptor does not explain the targets of infection of the virus and it is possible that innate immunity plays a significant role in its highly specific tropism in normal individuals.

Classification of *Picornaviridae*

The members of the family *Picornaviridae* to which poliovirus belongs are small nonenveloped viruses of about 30 nm in diameter when hydrated, essentially almost featureless when examined by electron microscopy and containing a single strand of an RNA molecule of messenger sense. The family is currently split into sixty three genera and one hundred and forty seven species on the basis of sequence similarities.

Table 1 Species within the genus *Enterovirus*

Species	Serotype
Human enterovirus A	Coxsackievirus A2, 3, 5, 7, 8, 10, 12, 14, 16 Enterovirus 71
Human enterovirus B	Coxsackievirus B 1–6, coxsackievirus A 9 Human echovirus 1–7, 9, 11–21, 24–27, 29–33 Human enterovirus 69
Human enterovirus C	Poliovirus 1, 2, 3 Coxsackie virus A 1, 11, 13, 15, 17–22, 24
Human enterovirus C	Human enterovirus 68, 70
Bovine enterovirus	Bovine enterovirus 1, 2
Porcine enterovirus A	Porcine enterovirus 8
Porcine enterovirus B	Porcine enterovirus 9, 10

The genus *Enterovirus*, of which poliovirus is the archetypal species, is currently further subdivided into twelve species of which A to D and the three rhinovirus species infect humans. Poliovirus is classified as a Human enterovirus type C. The classification of the enterovirus species and serotypes is shown in [Table 1](#).

The sequences of the capsid region of the viruses within a species define a virus type while sequences of other regions of the genome do not. This suggests that viruses within a particular species can exchange genome segments more or less freely.

The comparison of the sequence of field poliovirus isolates has proved a very valuable tool in the WHO program to eradicate polio. Viruses from the same geographical region tend to cluster in terms of their sequence largely independent of the year of isolation. Thus, it is possible to identify a virus from a case as indigenous to the region where it was isolated or whether the virus was an importation from another geographic region. Similarly, as virus circulation is restricted by effective vaccination programs, the variety of sequences found declines, indicating progress in the vaccination program long before the disease is eradicated. Finally, it is possible to identify strains as wild type or vaccine derived. The rate at which the sequence of a vaccine or wild-type virus lineage drifts in an epidemic or during chronic infection of immunodeficient individuals unable to clear the virus is amazingly constant at about 2%–3% silent substitutions per year, or 1% for all substitutions. Thus, the comparison of viral sequences can date the common ancestor of two polioviruses accurately and give an estimate how long they have been circulating.

Pathogenesis and Disease

Poliomyelitis gets its name from the specificity of the virus for the motor neurons that form the gray matter (*polios* and *myelos*, Greek for gray and matter, respectively) of the anterior horn. Despite the obvious physical signs of muscle atrophy and motor neuron degeneration, few cases of poliomyelitis are identifiable in the literature before the end of the nineteenth century, and it is believed that it was extremely rare. However, at this time, the disease began to occur in large epidemics, initially in Scandinavia (particularly Sweden), and then in the USA, most commonly affecting young children (hence the alternative name of infantile paralysis). The legs are more commonly affected than the arms, and paralysis tends to occur in one limb rather than symmetrically.

Poliovirus occurs in three antigenically distinct serotypes designated 1, 2, and 3, such that infection with one serotype confers solid protection only against other viruses of that serotype. Infections primarily occupy the gut, and most are entirely silent, but in a small number of cases they lead to a systemic infection (the minor disease) 3–7 days post exposure, characterized by fever, rash, or sore throat. Depending on the strain or type of virus, usually less than 1% of all infections lead to a more serious major disease or poliomyelitis which develops on average 7–30 days after infection. Spinal poliomyelitis resulting in lower limb paralysis or bulbar poliomyelitis in which the breathing centers are affected occurs when the lower or upper regions, respectively, of the spinal cord are affected. In the mid-twentieth century, about 10% of the polio cases lead to death, 10% of cases recovered without sequelae and 80% had permanent residual paralysis. Encephalitis may occur but it is a rather rare complication. Meningitis, also known as abortive or nonparalytic poliomyelitis, occurs at a rate of about 5% of poliomyelitis cases.

Infection is mainly fecal–oral, although the virus can also be transmitted by the respiratory route from the throats of infected individuals. The primary site of infection in the gut is not known but may be the lymphoid tissues, specifically the Peyer's patches and tonsils or the mucosal surfaces in the gut or throat; infectious virus can be found in the local lymph nodes and in more distal mesenteric lymph nodes but it is not clear whether the virus actually replicates there. The Peyer's patch-associated M-cells are believed to be the major infected cell types in the gut.

Viremia may occur about 7 days after infection, corresponding roughly to the appearance of the minor disease. It is believed to seed sites including the peripheral and central nervous systems resulting in the major disease. Thus, antibodies should prevent poliomyelitis by blocking viremia; the protective effect of passively administered antibodies was shown in the 1950s. This explains the change in the epidemiology of the disease in the beginning of the twentieth century as hygiene improved, and children were exposed to poliovirus infection later in life when maternal antibody levels had declined to levels no longer able to confine the infection to the gut.

Vaccines

Inactivated polio vaccine (IPV) was developed by Dr. Jonas Salk. It consists of wild-type virus that has been treated with low concentrations of formalin sufficient to inactivate the virus without affecting its antigenic properties to a great extent. The vaccine was first licensed in 1955 in the USA and its use reduced the number of cases by over 99%. It has been shown to be capable of eradicating the disease in certain countries, notably Scandinavia and the Netherlands. However, there remained controversy over whether a nonreplicating vaccine could eradicate the disease and the virus altogether in low income countries where exposure is more intense, and it was possible that live-attenuated vaccines imitating natural infection might give better protection against infection by wild-type viruses. The live-attenuated vaccines given orally (oral polio vaccine, OPV) that formed the basis of the eradication campaign were developed by Albert Sabin and introduced in the early 1960s. The Sabin vaccines have been shown to be able to interrupt epidemics and to break transmission chains if used correctly, resulting in the eradication of the virus in entire regions and possibly eventually the world.

It was also known that the Sabin vaccines altered in the vaccine recipients, becoming more neurovirulent particularly in the case of the type 2 and 3 components and that in rare instances, now estimated at about 1 case per 750,000 first-time vaccinees, the vaccine could cause poliomyelitis in recipients and their close contacts. Therefore, there is increasing usage of IPV in developed countries. WHO have recommended the inclusion of a dose of IPV in all immunization programs. Both IPV and OPV contained a single representative of each of the three serotypes.

Attenuation of the Sabin Vaccine Strains

The genome of poliovirus is of positive (messenger) RNA sense and therefore the viral RNA is infectious. Complete genomes of the virulent precursors of the Sabin vaccine strains or isolates from vaccine-associated cases can be cloned, sequenced, and mutations introduced or segments exchanged. Genomes can be synthesized *de novo*. Virus recovered from RNA transcribed from the modified plasmids was tested in monkeys or in transgenic mice carrying the poliovirus receptor. It was possible to identify two differences between Leon, the virulent precursor of the Sabin type 3 vaccine strain which would attenuate Leon or would de-attenuate the Sabin strain, one in the 5' noncoding region, the other in capsid protein VP3. Similarly, there were two major attenuating mutations in the type 2 strain, one in the 5' noncoding region and another one in capsid protein VP1. Other attenuating mutations with weak effects may also exist. The situation with type 1 poliovirus was more complicated. Again, there was one mutation in the 5' noncoding region but several others throughout the capsid region. All three of the 5' noncoding mutations affect the highly structured region shown in Fig. 1.

All 5' noncoding mutations have been shown to affect the initiation of protein synthesis, all involve allowed but altered base pairing, and all revert or are suppressed in vaccine recipients when sequential isolates are made. Excretion of virus following vaccination with OPV continues on an average for 4 or 5 weeks and the reversions observed occur early in the excretion period, usually within 1 or 2 days for the type 3 strain and within a week for the other two types. Other changes also take place, including the reversion or more usually direct or indirect suppression of the VP3 mutation in the type 3 strain.

It is possible to alter the base pairing in domain 5, exchanging GC or GU base pairs for AU to stabilize the structure genetically while altering its thermodynamic stability. Thus, where a GU base pair is replaced by an AU base pair two mutations will be required to generate the more thermodynamically stable GC. The thermodynamic stability of domain 5 has been shown to correlate with the virulence of the virus so that in principle a virus of any desired degree of attenuation can be generated by changing appropriate base pairs; this forms the basis of some novel vaccines which are in development as described later.

In addition, the viruses recombine with each other at high frequency. The type 3 strain in particular is usually excreted by vaccinees as a recombinant from about 11 days post immunization with the part of the genome encoding the nonstructural genes from the type 1 or type 2 strains. Complex recombinants with portions from different serotypes are also common, although their selective advantage is not known. The adaptation of the virus to the gut is therefore rapid and subtle and by a variety of mechanisms including reversion, second site suppression, and suppression of the phenotype of poor growth by enhancing fitness by an entirely unrelated route, and may involve mutation or recombination. Once adapted, virus excretion persists typically for several weeks. The drift in the viral sequence over prolonged periods of time referred to above occurs in addition to these selected changes, and appears to be a consequence of a purely stochastic process.

The Global Polio Eradication Program

For many years, it was thought that OPV could not be effective in tropical countries, and despite large-scale use its impact was in fact minimal. The reasons put forward included interference by other enteric infections, and many other factors, including breast-feeding. The most plausible explanations are that the vaccine used in routine vaccination programs was probably poorly looked after and therefore inactive, and the epidemiology of the disease was different in tropical and temperate climates. In countries of Northern Europe, poliomyelitis is highly seasonal, occurring essentially only in summer. Thus, a routine vaccination campaign in which children are immunized at a specific age is able to reduce the number of susceptible individuals during the winter when the virus is not freely circulating, so that circulation of the wild-type virus in the summer is impaired. Eventually, the wild-type virus

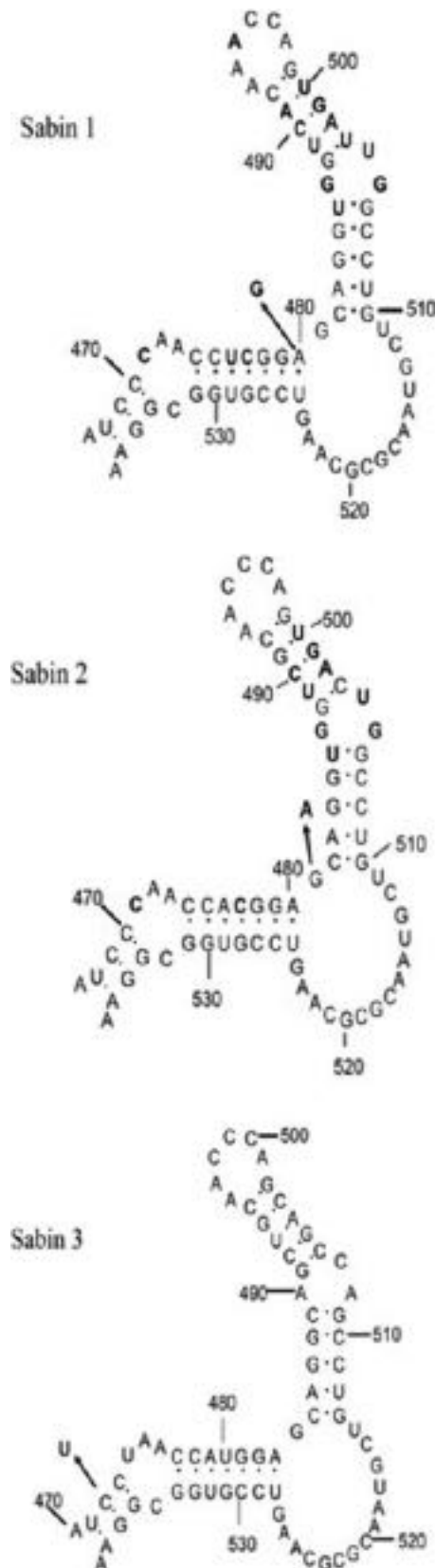


Fig. 1 Structure of RNA domain in the 5' noncoding region of the poliovirus genome involved in attenuating the neurovirulence of the Sabin live-attenuated vaccine strains. Top: type 1; center: type 2; bottom: type 3. Differences from the type 3 sequence are shown in bold, illustrating the general conservation of the base-paired structures. Base changes involved in attenuation are shown by arrows for the three types. Note that for types 1 and 3, the changes result in weaker interactions in the secondary structure but allowed base pairing compared to the wild-type sequence.

dies out. In tropical climates, however, exposure is less seasonal, so that there is no respite during which immune populations can be built up by immunization; it remains a matter of chance whether a child is exposed first to the vaccine or wild-type virus, and the impact on reducing the disease is correspondingly weaker. The strategy required is to immunize large proportions of the population at once. This approach was successfully followed in the 1960s in the USA in the southern states, where the climate is more tropical and routine immunization was less effective. However, it was not until the 1980s that the strategy was used in developing countries in South America when vaccine was given in mass campaigns termed National Immunization Days (NIDs). The success of the program led to a resolution in 1988 by WHO to eradicate polio from the world by the year 2000. While this goal has still not been achieved at the time of writing, Pakistan and Afghanistan are currently the only two countries from which the wild type viruses have never been eliminated. Northern India proved extremely difficult and some children contracted poliomyelitis after 10 or more immunizations with OPV. One approach adopted was to use monovalent vaccine instead of the trivalent form containing all three serotypes. The rationale is that in the trivalent vaccine the different serotypes compete with each other so that the required serotype may not infect and induce immunity, whereas a monovalent type 1 vaccine used in an area where this is the problem serotype will be effective. There is evidence that this gives faster immune responses, and it was first used in Egypt where it eradicated the remaining type 1 strains. India was declared free of polio in 2014 after three years without a case. The whole of Africa was declared free of wild type poliovirus of all three serotypes in 2020.

The type 2 component was removed from all OPVs in 2016 as no naturally occurring case of poliomyelitis caused by wild type 2 has been identified globally since 1999 and the type 2 component was the main source of circulating vaccine derived polioviruses (cVDPVs) which were a major cause of residual polio cases in the world.

While there are only two countries in which poliomyelitis has not been eradicated, many countries have suffered reintroductions. This is shown in Fig. 2, which shows cases of the regions where polio has not been completely eliminated.

The first examples of this phenomenon were in central Africa. A decision was taken because of shortage of funds to concentrate immunization efforts in Nigeria which had many cases (Fig. 2(a)). Thus, immunization in the surrounding countries suffered, and at the same time resistance to immunization grew in northern Nigeria because of local concerns about supposed contaminants in the vaccine. During the period when immunization ceased, there was a resurgence in polio in Nigeria, which spread to adjacent countries where the immunization activities had been reduced (Fig. 2(b)). As a result, polio spread from Nigeria across much of central Africa, being brought under control eventually by coordinated NIDs across most of the continent, an operation of unprecedented scale. Shortly after this, with polio still endemic in Nigeria, the annual pilgrimage to Mecca resulted in the introduction of polio from northern Nigeria into Yemen and Indonesia (Fig. 2(c)). In 2006, polio was introduced into several African countries from northern India (Fig. 2(d)).

The difficulties of eradication are hard to overestimate, and are associated with the fact that most infections are entirely silent. Despite this, as stated above, there has been no case of poliomyelitis caused by a naturally occurring type 2 virus since 1999 and there have been no cases caused by wild type 3 virus since 2012.

It is clear that as long as one country remains a source of the virus the world remains at risk. This complicates the strategies to be followed once there is some confidence that the virus has in fact been eradicated in the wild, which are further complicated by the fact that OPV is a live vaccine derived from wild-type virus that is able to change in infected vaccinees, from whom virus may be isolated for significant periods of time.

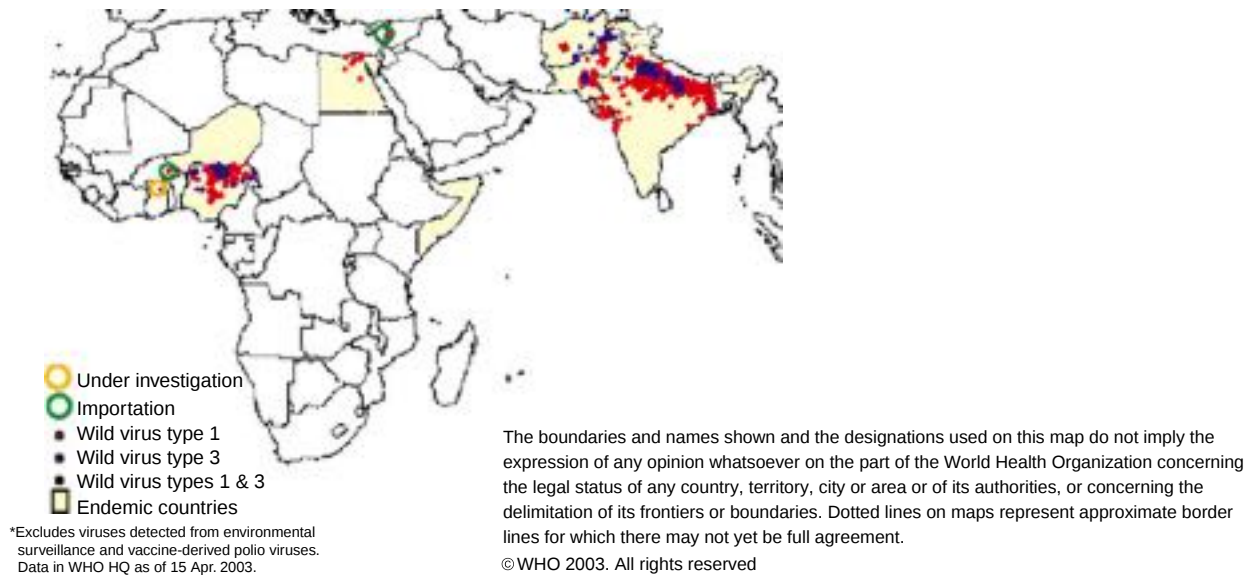
Vaccine-Derived Poliovirus

Vaccine-associated paralytic poliomyelitis (VAPP) cases were recognized from the first use of the Sabin vaccine strains, although their unambiguous identification required molecular methods applied in the 1980s. VAPPs occur in primary vaccinees at a rate of about 1 per 750,000 but can also occur in contacts and those previously vaccinated. The occurrence of such cases, while rare, indicates that the vaccine strains can revert to virulence in recipients and that it is possible for the vaccine strains to spread from person to person.

In 2001, it was recognized for the first time that the Sabin vaccine strains could be the cause of outbreaks of poliomyelitis when about 22 cases were identified in Hispaniola, comprising Haiti and the Dominican Republic. Sequencing of the strains showed that they were very closely related to the type 1 Sabin vaccine strains and not to the previous endemic strains of which the last had been isolated 20 years earlier. Moreover, the sequence diversity indicated that the outbreak strains had been circulating unnoticed for *ca.* 2 years, as they differed from the vaccine strain by about 2% overall. In addition, they were shown to be recombinant strains with a major portion of the nonstructural regions of the genome from viruses identified as species C enteroviruses unrelated to the vaccine strains. At least three different genomic structures were identified. Subsequently, it was reported that between 1988 and 1993 all supposed wild type 2 strains isolated from poliomyelitis cases in Egypt were in fact heavily drifted vaccine-related strains. Outbreaks in the Philippines in 2001, two in Madagascar in 2002 and 2005, one in China in 2004, and one in Indonesia in 2005 and many others have been recorded. The most likely explanation is that vaccination programs become less vigorous once polio is eliminated from a country and other health issues take priority, so that while administration of OPV continues, the coverage is less than 100% giving perfect conditions for the selection of transmissible strains. For reasons that are not understood, almost all such strains to date have been recombinants with unidentified species C enteroviruses and most are type 2, followed by type 1, with type 3 being rare. So far, the outbreaks have been limited and easy to control; this was particularly shown in Indonesia where a type 1 outbreak from a virus introduced from Nigeria occurred at the same time as an outbreak of circulating vaccine-derived

(a)

Wild poliovirus*, 15 Apr. 2002–14 Apr. 2003



(b)

Wild poliovirus*, 12 Apr. 2004–12 Apr. 2005

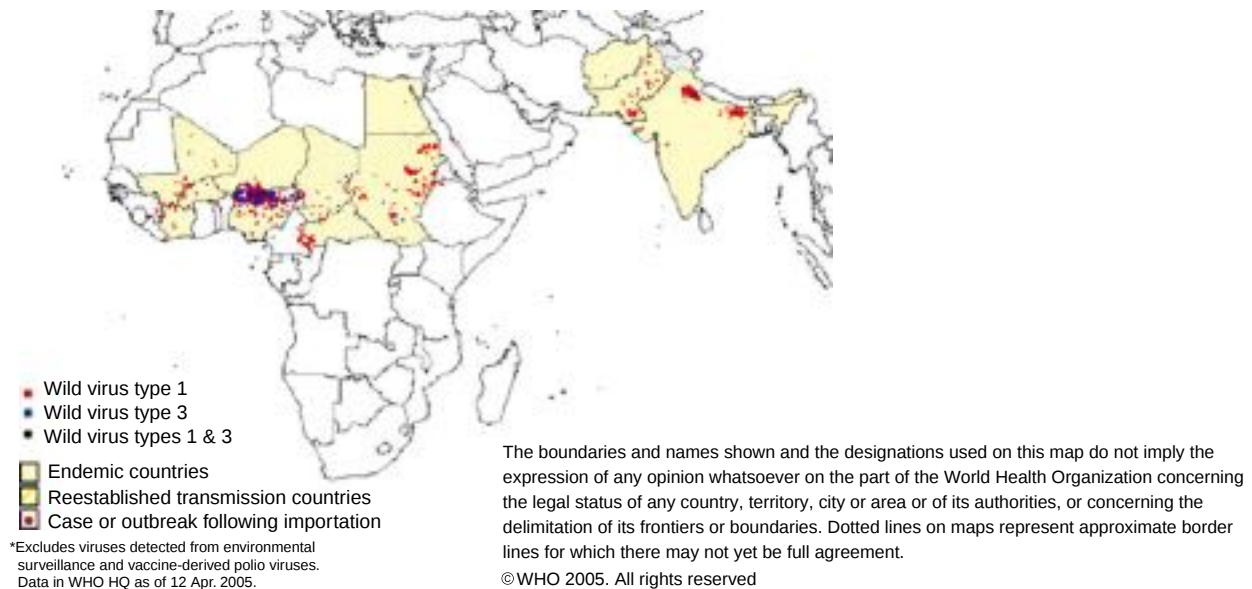
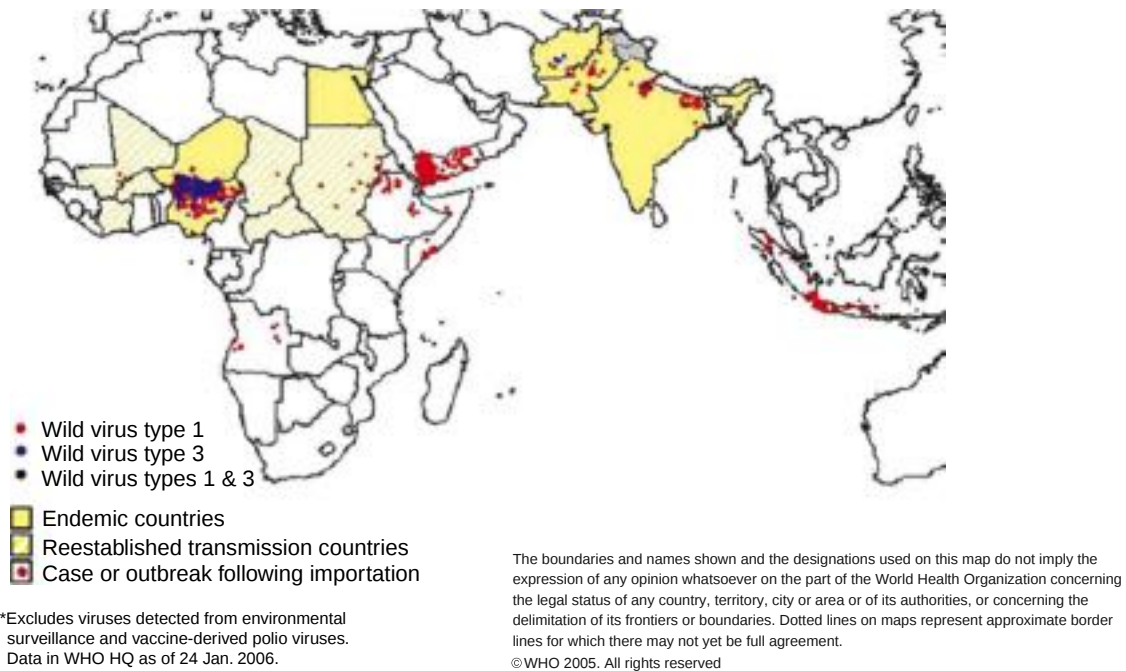


Fig. 2 Occurrence of cases of poliomyelitis in remaining areas of infection from 2002 to 2006. World Health Organization (WHO).

(c) Wild poliovirus*, 25 Jan. 2005–24 Jan. 2006



(d) Wild poliovirus*, 30 Aug. 2005–29 Aug. 2006

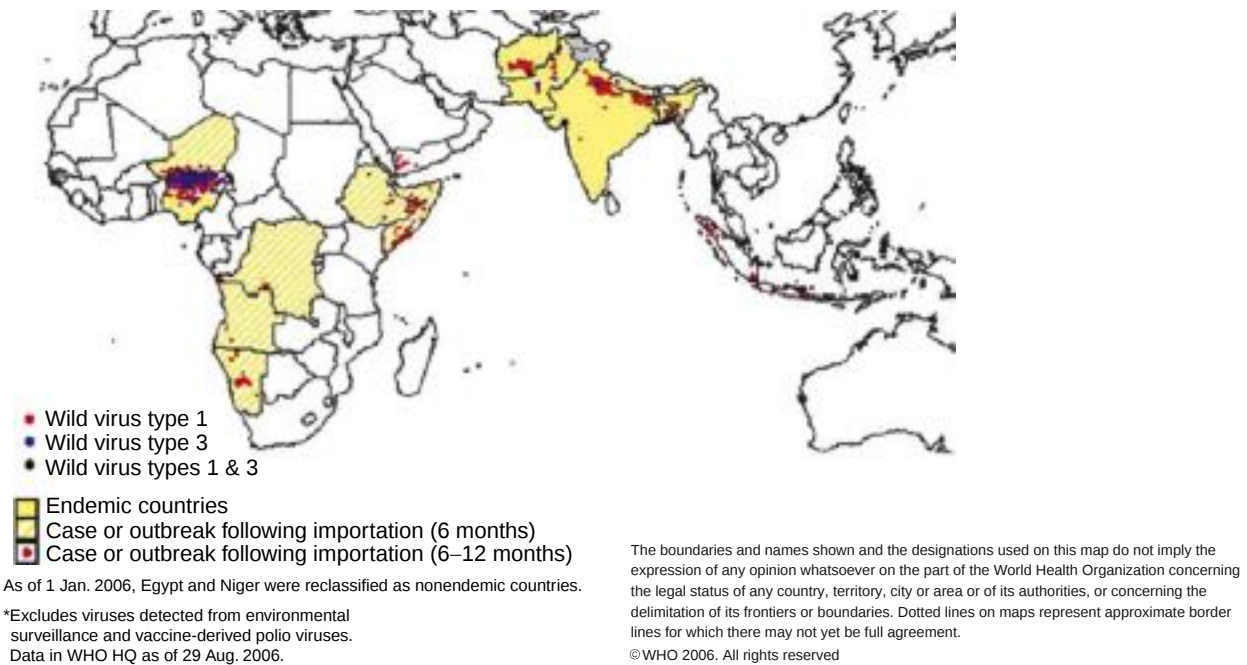


Fig. 2 continued.

poliovirus (CVDPV). The wild-type outbreak persisted longer. Virologically, there is no real reason why vaccine-derived viruses should not become as aggressive as the wild-type viruses from which they were derived, although this does not seem to have happened at the time of writing. The occurrence of such viruses is clearly a problem when cessation of vaccination is considered and was a major factor in the decision to remove the type 2 component of OPV.

A second issue is the occasional case of long-term excretion of polioviruses by individuals deficient in humoral immunity. The viruses are termed immunodeficient vaccine-derived polioviruses (IVDPVs). Probably only a few percent of these hypogammaglobulinemia individuals will become long-term excretors of virus even if given OPV, and most stop shedding virus spontaneously after a period, which may be 2 or 3 years. Occasionally virus excretion may continue for much longer; in one instance isolates were available for 19 years, and the sequence data and medical history of the patient suggest that virus has been shed for well over 30 years. The patient remains entirely healthy, although the virus is highly virulent in animal models. No treatment has been successfully applied to such individuals, although there are claims for a success with an antiviral compound in one case. It is of interest that the mechanism by which virus is cleared is still not known. An infected individual can be expected to excrete the virus for over 4 weeks and will continue to do so after an immune response is detectable. Immune-deficient patients can stop excreting virus spontaneously with no evidence of an immune response and while termination of excretion correlates in general with a rise in fecal anti-poliovirus IgA levels, this is neither necessary nor sufficient.

Cessation of Vaccination

Wild type 1 poliovirus currently circulates only in Pakistan and Afghanistan and wild type 2 and 3 have been eliminated from the entire world. Once the wild virus has been eradicated vaccination in its current form is hard to justify. WHO have produced very stringent guidelines (the Global Action Plan, third edition, or GAP3) on containment of poliovirus as this becomes a major concern as eradication approaches.

In 2016 the type 2 component of oral polio vaccine was withdrawn on the grounds that there had been no wild type 2 polio since 1999, the type 2 component interfered with immunization with the other two strains in OPV and the type 2 component was responsible for most cases of vaccine associated poliomyelitis and circulating vaccine derived strains. However, withdrawal was problematic in so far as there were still type 2 cVDPVs, so a fall back policy was needed which took the form of supplies of monovalent type 2 OPV intended to deal with future cVDPV outbreaks should they arise. This strategy failed in so far as outbreaks continued to occur and the programmatic response caused more type 2 cVDPV to be generated.

A more radical approach is therefore being adopted in which a new live attenuated type 2 vaccine has been developed that is designed to be more genetically stable than the Sabin strain, based on the understanding gained of the molecular basis of attenuation. Two candidates were developed. Both contain modified domain 5 structures in which base pairs are modified to fine tune the thermodynamic stability to produce the desired degree of attenuation by substituting weaker AU base pairs for GC pairs as mentioned earlier. One candidate also includes a high fidelity RNA polymerase to reduce the generation of mutations and in this candidate the *cre* was moved to a region to the 5' side of domain 5. This last manipulation means that the attenuating domain 5 structure cannot be replaced by a single recombination event without losing *cre* and thus the ability to replicate. Overall, the virus should therefore be genetically more stable than the Sabin strain. The other candidate possessed the attenuated domain 5 but was produced by synthesis of the genome and deoptimizing codon usage so that virus production and replication is far less efficient. Both candidates have been evaluated in clinical trials and shown to be immunogenic and genetically stable over the time covered. Vaccine usage based on at least one of the strains will begin in 2021.

In addition, OPV usage has stopped in many countries and been replaced by the use of Inactivated Polio Vaccine. This switch has taken place in Europe, the whole of North America and much of central and Southern America as well as in Australia and New Zealand. The production of IPV requires the growth and inactivation of live poliovirus which may be a problem if the virus escapes the manufacturing facility. There are stringent requirements for handling poliovirus but a safer strain would be advisable. A strain designated S19 in which domain 5 has been seriously weakened by base pair exchanges so that virus growth is seriously hampered has been developed. The strain is unable to infect animal models and is genetically very stable making it safe for use. Its use in production is being explored.

These developments give hope that cessation of vaccination in its current form can occur safely and that both wild type and vaccine derived polioviruses can be eradicated. The progress will need to be monitored closely.

Further Reading

- Bodian, D., 1955. Emerging concept of poliomyelitis infection. *Science* 122, 105–108.
- Fine, P.E., Carneiro, I.A., 1999. Transmissibility and persistence of oral polio vaccine viruses: Implications for the global poliomyelitis eradication initiative. *American Journal of Epidemiology* 150, 1001–1021.
- Hammon, W.D., Coriell, L.L., Wehrle, P.F., *et al.*, 1953. Evaluation of Red Cross gammaglobulin as a prophylactic agent for poliomyelitis. Part 4: Final report of results based on clinical diagnosis. *JAMA* 151, 1272–1285.
- Kew, O., Morris-Glasgow, V., Landarverde, M., *et al.*, 2002. Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. *Science* 296, 356–359.
- Minor, P.D., 1992. The molecular biology of poliovaccines. *Journal of General Virology* 73, 3065–3077.

- Minor, P.D., John, A., Ferguson, M., Icenogle, J.P., 1986. Antigenic and molecular evolution of the vaccine strain of type 3 poliovirus during the period of excretion by a primary vaccinee. *Journal of General Virology* 67, 693–706.
- Minor, P.D., Nathanson, N., Ahmed, R., Gonzalez-Scarano, F., *et al.*, 1997. *Viral Pathogenesis*. Philadelphia: Lippincott-Raven Publishers, pp. 555–574.
- Nkowane, B.U., Wassilak, S.G., Orenstein, W.A., 1987. Vaccine associated paralytic poliomyelitis in the United States: 1973 through 1984. *JAMA* 257, 1335–1340.
- Paul, J.R., 1971. *A History of Poliomyelitis*. New Haven, CT: Yale University Press.
- Sabin, A.B., 1956. Pathogenesis of poliomyelitis: Reappraisal in the light of new data. *Science* 123, 1151–1157.
- Sabin, A.B., Ramos-Alvarez, M., Alvarez-Amesquita, J., *et al.*, 1960. Live orally given poliovirus vaccine: Effects of rapid mass immunization on population under conditions of massive enteric infection with other viruses. *JAMA* 173, 1521–1526.

Relevant Website

<http://www.polioeradication.org>

Global Polio Eradication Initiative: GPEI.

Porcine Reproductive and Respiratory Syndrome Virus and Equine Arteritis Virus (*Arteriviridae*)

Jianqiang Zhang, Iowa State University, Ames, IA, United States

Alan T Loynachan, University of Kentucky, Lexington, KY, United States

Gregory W Stevenson and Jeffrey J Zimmerman, Iowa State University, Ames, IA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Chronic persistent infection An infection in which the virus continues to replicate in the host, despite an adaptive immune response. In contrast to “latent persistent infection”, chronic infection is eventually cleared by the immune system.

Clinical signs Objective evidence of disease perceptible to an observer. In contrast, “symptoms” are subjective sensations described by the affected person.

Half-life The time required to inactivate one-half of a viral population. Half-life is situation-specific, i.e., dependent on the environment (temperature, humidity, pH, light, etc.).

Pathogenicity The quality or state of being able to produce disease. Sometimes incorrectly used as a synonym for virulence.

Sensitivity, analytical For any assay, the relationship between the concentration of the target in the sample and the probability of detecting the target. It is sometimes (incorrectly) described as the limit of detection.

Sensitivity, diagnostic For any assay, the probability of a positive test result when testing a true-positive specimen.

Specificity, analytical The degree to which an assay is selective for the target of interest. Assays that lack analytical specificity will cross-react with other targets (false positive).

Specificity, diagnostic For any assay, the probability of a negative test result when testing a true-negative specimen.

Virulence A measure of pathogenicity. For example, if mortality is the criterion, virulence can be expressed as the “case-fatality rate.”

Classification

Porcine reproductive and respiratory syndrome (PRRS) was first reported in the late 1980s in the U.S. and early 1990s in Europe. The etiologic agent, PRRS virus (PRRSV), was identified in Europe (Type 1, PRRSV-1) and North America (Type 2, PRRSV-2) in 1991. Equine arteritis virus (EAV) was first isolated from the lung tissue of an aborted fetus during an epidemic of abortions and respiratory disease in pregnant mares in Bucyrus, Ohio, USA, in 1953.

The nomenclature is in transition, but the current taxonomic classifications of PRRSV and EAV are given in [Table 1](#). Both are members of the family *Arteriviridae* in the suborder *Arnidovirineae* within the order *Nidovirales*. However, PRRSV-1 (species name *Betaarterivirus suid 1*) and PRRSV-2 (species name *Betaarterivirus suid 2*) respectively belong to the subgenera *Eurpobartevirus* and *Ampobartevirus* in the genus *Betaarterivirus* and the subfamily *Variarterivirinae*. In contrast, EAV (species name *Alphaarterivirus equid*) belongs to the genus *Alphaarterivirus* in the subfamily *Equarterivirinae*.

PRRSV strains are highly genetically diverse, especially in open reading frame (ORF) 5 and non-structural protein (nsp) 2 regions. Based on ORF5, PRRSV-1 and PRRSV-2 vary by ~44% in nucleotide sequence; within PRRSV-1 and PRRSV-2, nucleotide sequence variations are approximately 30% and 21%, respectively. Phylogenetic analyses of ORF5 sequences classified PRRSV-1 viruses into 4 subtypes and PRRSV-2 viruses into 9 lineages.

EAV field strains are also diverse, although to a lesser extent than PRRSV. EAV field isolates from North America and Europe differ approximately 15% in nucleotide sequences, with the major differences located in the nsp2, ORF3, and ORF5 regions. Phylogenetic analyses of ORF5 sequences classified EAV isolates into North American and European clades, with the latter further divided into European subgroup-1 (EU-1) and European subgroup-2 (EU-2).

Table 1 Taxonomy of EAV and PRRSV

Order	Suborder	Family	Subfamily	Genus	Subgenus	Species	Virus name
Nidovirales	Arnidovirineae	Arteriviridae	Equarterivirinae	Alphaarterivirus		Alphaarterivirus equid	Equine arteritis virus (EAV)
			Variarterivirinae	Betaarterivirus	Eurpobartevirus	Betaarterivirus suid 1	Porcine reproductive and respiratory syndrome virus 1 (PRRSV-1)
					Ampobartevirus	Betaarterivirus suid 2	Porcine reproductive and respiratory syndrome virus 2 (PRRSV-2)

Note: Based on the International Committee on Taxonomy of Viruses report released in 2018.

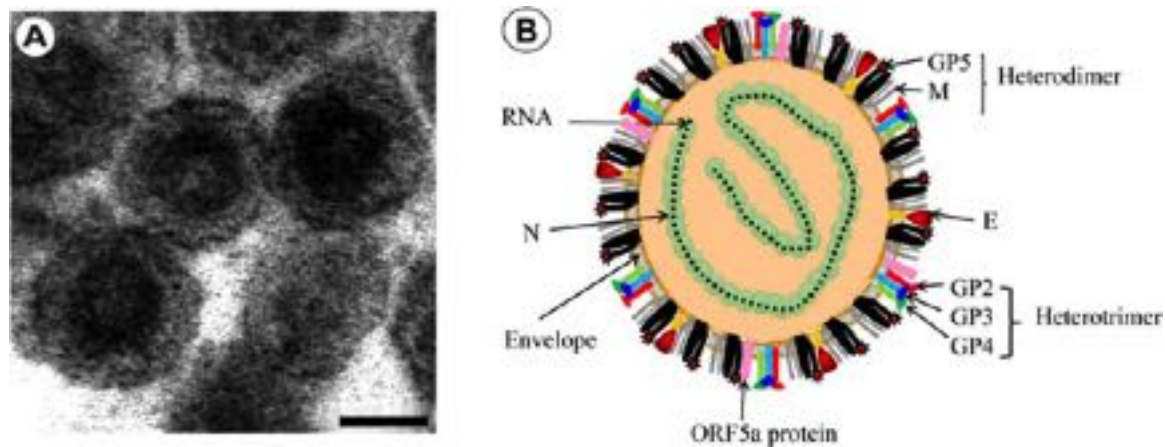


Fig. 1 Electron micrograph of arterivirus particles and schematic diagram of arterivirus virion. (A) Transmission electron micrograph of extracellular porcine reproductive and respiratory syndrome virus (PRRSV) particles. Bar = 25 nm. (B) Schematic diagram of equine arteritis virus (EAV) virion. Image adapted from (A) Snijder, E.J., Meulenberg, J.J.M., 1998. The molecular biology of arteriviruses. *Journal of General Virology* 79, 961–979 with permission. (B) Balasuriya, U.B.R., Go, Y.Y., MacLachlan, N.J., 2013. Equine arteritis virus. *Veterinary Microbiology* 167, 93–122 with permission.

Virion Structure

Both PRRSV and EAV virions are spherical, with a diameter of 50–60 nm and inapparent surface projections (**Fig. 1(A)**). The virion includes a nucleocapsid core composed of an infectious RNA genome and nucleocapsid (N) protein, surrounded by a lipid-containing envelope with 7 viral-encoded proteins (**Fig. 1(B)**). It has been assumed that the arteriviruses have an icosahedral nucleocapsid core, but recent cryo-electron tomography studies suggest that the PRRSV nucleocapsid core is pleomorphic. The N protein interacts with the viral RNA during virus assembly and is the most abundant structural protein in arterivirus virions. The non-glycosylated membrane (M) protein and the glycosylated GP5 protein are major envelope proteins that form disulfide-linked heterodimers. The small non-glycosylated envelope protein (E), glycosylated GP2, GP3, GP4 proteins, and the newly discovered ORF5a protein are minor envelope proteins. The GP2/GP3/GP4 glycoproteins are covalently associated and form a heterotrimeric complex on the virion surface.

The buoyant density of infectious arterivirus particles is ~ 1.18 – 1.22 g/cm³ in cesium chloride and ~ 1.13 – 1.17 g/cm³ in sucrose. Based on a study of 4 PRRSV-2 isolates, virus half-life was found to be highly temperature dependent: 155 h at 4°C (39°F), 84.5 h at 10°C (50°F), 27.4 h at 20°C (68°F), and 1.6 h at 30°C (86°F). Half-life estimates are not available for EAV, but infectious virus was detected up to 75 days at 4°C (39°F), 48–72 h at 37°C (99°F), and 20–30 min at 56°C (133°F). Both PRRSV and EAV are stable at -70°C (-94°F) for years without significant loss of infectivity. PRRSV is stable at neutral pH but is inactivated by low and high pH (<6 or >7.5); this may apply to EAV as well, although it has not been experimentally investigated.

Genome Organization

PRRSV and EAV have similar genomic organization, i.e., positive-sense, single-stranded, 3'-polyadenylated RNA (~ 15 kb for PRRSV and ~ 12.7 kb for EAV), presumably with a cap structure at the 5' end. The genome includes 5' untranslated region (UTR), 10 ORFs (ORFs1a, 1b, 2a, 2b, 3, 4, 5a, 5b, 6 and 7), and 3' UTR (**Table 2**; **Fig. 2**). Based on EAV studies, the 5' UTR is involved in translation, replication and transcription. An RNA hairpin in EAV 3' UTR is required for RNA synthesis and is essential for a pseudoknot interaction with a hairpin in the N protein gene.

ORF1a and ORF1b occupy approximately 5' three-quarters of the genome. ORF1a is translated into replicase polypeptide (pp) 1a, which is further processed by viral encoded proteinases (papain-like proteinases and serine proteinase) into nsp1–8 (**Table 2**; **Fig. 2**). ORF1b translation involves a -1 programmed ribosomal frameshifting (PRF) which extends pp1a into pp1ab. The ORF1b-encoded portion of pp1ab is further cleaved by viral-encoded serine proteinase into nsp9–12 (**Table 2** and **Fig. 2**). Some important functional domains conserved in arterivirus ORF1b-encoded protein include the RNA-dependent RNA polymerase in nsp9, putative multinuclear zinc-binding domain and RNA helicase in nsp10, and the NendoU endoribonuclease domain in nsp11 (**Fig. 2**). Recently, a conserved small ORF that overlaps the nsp2-encoding region of ORF1a in the +1 frame was identified in PRRSV, but not in EAV genome, and translation of this small ORF involves a -2 PRF mechanism.

The 3' one-quarter of the genome includes ORFs 2a, 2b, 3, 4, 5a, 5b, 6, and 7, which respectively encode structural proteins E (EAV) or GP2 (PRRSV), GP2 (EAV) or E (PRRSV), GP3, GP4, ORF5a, GP5, M, and N (**Table 2**; **Fig. 2**). These structural proteins are not directly translated from genomic RNA; rather, they are translated from subgenomic (sg) mRNAs (see Life Cycle).

Table 2 Nonstructural and structural proteins of EAV and PRRSV

ORF	Protein	Predicted nucleotide position			Predicted protein		
		EAV ^a	PRRSV-1 ^b	PRRSV-2 ^c	EAV ^a	PRRSV-1 ^b	PRRSV-2 ^c
5' UTR		1–224	1–221	1–189			
ORF1a	pp1a	225–5408	222–7412	190–7701	Met1-Asn1727 (1727 aa)	Met1-Cys2396 (2396 aa)	Met1-Cys2503 (2503 aa)
ORF1b	–	5405–9751	7406–11785	7695–12071			
ORF1a/b	pp1ab	225–5405, 5405–9751	222–7406, 7406–11785	190–7695, 7695–12071	Met1-Val3175 (3175 aa)	Met1-Pro3854 (3854 aa)	Met1-Asn3960 (3960 aa)
	nsp1	255–1004	222–1376	190–1338	Met1-Gly260 (260 aa)	Met1-Gly385 (385 aa)	Met1-Gly383 (383 aa)
	nsp2	1005–2717	1377–4610	1339–4926	Gly261-Gly831 (571 aa)	Ala386-Gly1463 (1078 aa)	Ala384-Gly1579 (1196 aa)
	nsp3	2718–3416	4611–5300	4927–5616	Gly832-Glu1064 (233 aa)	Ala1464-Glu1693 (230 aa)	Gly1580-Glu1809 (230 aa)
	nsp4	3417–4028	5301–5909	5617–6228	Gly1065-Glu1268 (204 aa)	Gly1694-Glu1896 (203 aa)	Gly1810-Glu2013 (204 aa)
	nsp5	4029–4514	5910–6419	6229–6738	Ser1269-Glu1430 (162 aa)	Gly1897-Glu2066 (170 aa)	Gly2014-Glu2183 (170 aa)
	nsp6	4515–4580	6420–6467	6739–6786	Gly1431-Glu1452 (22 aa)	Gly2067-Glu2082 (16 aa)	Gly2184-Glu2199 (16 aa)
	nsp7	4581–5255	6468–7274	6787–7563	Ser1453-Glu1677 (225 aa)	Ser2083-Glu2351 (269 aa)	Ser2200-Glu2458 (259 aa)
	nsp8	5256–5405	7275–7409	7564–7698	Gly1678-Asn1727 (50 aa)	Ala2352-Cys2396 (45 aa)	Ala2459-Cys2503 (45 aa)
	nsp9	5256–5405, 5405–7333	7275–7406, 7406–9328	7564–7695, 7695–9617	Gly1678-Glu2370 (693 aa)	Ala2352-Glu3036 (685 aa)	Ala2459-Glu3143 (685 aa)
	nsp10	7334–8734	9329–10654	9618–10940	Ser2371-Gln2837 (467 aa)	Gly3037-Glu3478 (442 aa)	Gly3144-Glu3584 (441 aa)
	nsp11	8735–9391	10,655–11,326	10,941–11,609	Ser2838-Glu3056 (219 aa)	Gly3479-Glu3702 (224 aa)	Gly3585-Glu3807 (223 aa)
	nsp12	9392–9748	11,327–11,782	11,610–12,068	Gly3057-Val3175 (119 aa)	Gly3703-Pro3854 (152 aa)	Gly3808-Asn3960 (153 aa)
ORF2a	E (EAV); GP2 (PRRSV)	9751–9954	11,796–12,545	12,073–12,843	E (67 aa)	GP2 (249 aa)	GP2 (256 aa)
ORF2b	GP2 (EAV); E (PRRSV)	9824–10507	11,801–12,013	12,078–12,299	GP2 (227 aa)	E (70 aa)	E (73 aa)
ORF3	GP3	10,306–10,797	12,404–13,201	12,696–13,460	163 aa	265 aa	254 aa
ORF4	GP4	10,700–11,158	12,946–13,497	13,241–13,777	152 aa	183 aa	178 aa
ORF5a	ORF5a	11,112–11,291	13,499–13,630	13,778–13,933	59 aa	43 aa	51 aa
ORF5b	GP5	11,146–11,913	13,494–14,099	13,788–14,390	255 aa	201 aa	200 aa
ORF6	M	11,901–12,389	14,087–14,608	14,375–14,899	162 aa	173 aa	174 aa
ORF7	N	12,313–12,645	14,598–14,984	14,889–15,260	110 aa	128 aa	123 aa
3' UTR		12,646–12,704	14,985–15,098	15,261–15,411			

^aBased on EAV VBS strain (ATCC VR-796) (GenBank accession number # DQ846750).

^bBased on PRRSV-1 strain Lelystad (GenBank accession number # M96262).

^cBased on PRRSV-2 strain VR2332 (GenBank accession number # PRU87392).

The PRRSV and EAV N proteins are phosphorylated and the N-terminal domain of the N protein is presumed to interact with the genomic RNA during nucleocapsid assembly. M and GP5 are two major envelope proteins. The nonglycosylated M protein lacks an N-terminal signal sequence and is presumed to span the membrane three times. M protein forms heterodimers with GP5 protein, probably by the formation of a disulfide bridge between a Cys residue in the M ectodomain and a Cys residue in the GP5 ectodomain. The glycosylated GP5 protein contains an N-terminal signal sequence, a hydrophobic region spanning the membrane three times, and a cytoplasmic domain. The PRRSV and EAV GP5 ectodomain is an important target for neutralizing antibodies. E, GP2, GP3, GP4, and ORF5a protein are minor envelope proteins. The disulfide-linked GP2-GP3-GP4 trimers have been identified in both PRRSV and EAV; in addition, the disulfide-linked GP2-GP4 dimers are found in EAV.

All major (N, GP5 and M) and minor (E, GP2, GP3, and GP4) structural proteins are required for the production of infectious progeny virus, as demonstrated by individually knocking out the expression of each structural protein. Knocking out ORF5a protein expression led to the production of infectious progeny virus with a small plaque phenotype and significantly reduced virus

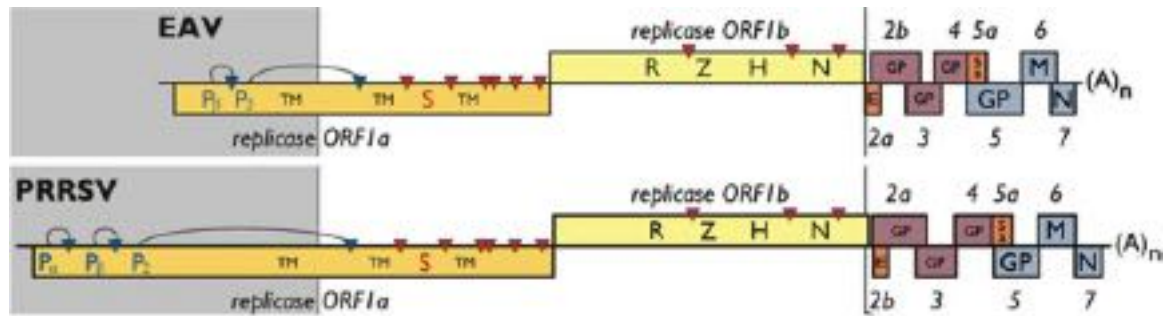


Fig. 2 Genome organization of equine arteritis virus (EAV) and porcine reproductive and respiratory syndrome virus (PRRSV). In open reading frame (ORF) 1a, the papain-like proteinase domains (P) in the nsp1-nsp2 region and their predicted cleavage sites are indicated with blue triangles. In nsp4, the main viral protease (serine proteinase, S) and its predicted cleavage sites are indicated with red triangles. "TM" indicates the putative transmembrane domains in nsp2, nsp3, and nsp5 of the ORF1a-encoded polyprotein. In ORF1b, the four most conserved replicase domains are indicated: the RNA-dependent RNA polymerase (R) in nsp9, putative multinuclear zinc-binding domain (Z) and RNA helicase (H) in nsp10, and the NendoU endoribonuclease domain (N) in nsp11. Minor structural proteins E (encoded by ORF2a in EAV and ORF2b in PRRSV), GP2 (ORF2b in EAV and ORF2a in PRRSV), GP3 (ORF3), GP4 (ORF4) and ORF5a (encoded by ORF5a) and major structural proteins GP5 (ORF5), M (ORF6) and N (ORF7) are indicated. Modified with permission from Snijder, E.J., Kikkert, M., 2013. Arteriviruses. In: Knipe, D.M., Howley, P.M., (Eds.), Fields Virology, sixth ed. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 859–879.

titer. Knocking out minor structural proteins (E, GP2, GP3, and GP4) produced non-infectious subviral particles consisting of N, GP5, M, and the genomic RNA, suggesting that major structural proteins N, GP5, and M are essential and sufficient for forming virion scaffold (virus-like particle), although such particles lack infectivity.

Life Cycle

The virus infection cycle starts with virus attachment and entry into host cells. This process involves the interaction between the viral proteins and receptor(s) on the host cells.

In pigs, PRRSV primarily infects monocytes, monocyte-derived macrophages, and monocyte-derived dendritic cells. Continuous cell lines MARC-145 and CL2621, both derived from African green monkey kidney MA-104 cells, are also permissive to PRRSV infection. The scavenger receptor CD163, expressed on both porcine macrophages and MARC-145 cells, can alone mediate virus attachment and entry (including internalization and uncoating). Sialoadhesin (Sn, or CD169), which is expressed on macrophages, but absent on MARC-145 cells, can enhance virus internalization, but cannot mediate uncoating. There are reports that glycosaminoglycans (heparan sulfate) on the cell surface may interact with sialic acids on the PRRSV virion surface and are implicated in the initial binding step, but heparan sulfate is not absolutely required for PRRSV entry.

In horses, EAV infects equine endothelial cells, monocytes, macrophages, and a small subpopulation of CD3⁺ T cells. *In vitro*, EAV can infect a broad range of continuous cell lines from various host species, including baby hamster kidney (BHK-21), African green monkey kidney (Vero, MARC-145), rabbit kidney (RK-13), and mouse C2C12 cells. Recent studies suggest that equine CXCL16 (EqCXCL16) protein, encoded by the CXCL16 gene located on equine chromosome 11, acts as a receptor for EAV entry into some cell populations in horses. However, EAV may also use other as-yet-unidentified molecule(s) for entry into equine cells. That is, EqCXCL16 protein is not expressed on non-equine host cell lines permissive to EAV infection, thus it remains to be determined if CXCL16 orthologue proteins are present on these cell lines and act as receptors for EAV or if EAV uses other as-yet-unidentified receptors to infect these cell lines.

The PRRSV and EAV GP5 and M proteins were initially hypothesized as viral ligands for interacting with cell receptors. However, studies with chimeric viruses indicated that GP5 and M ectodomains are not the determinants for PRRSV and EAV cell tropism. Rather, PRRSV/EAV chimeric virus studies suggested that the minor envelope proteins GP2, GP3, and GP4, in combination with the E protein, may be responsible for interacting with the cellular receptor(s). In fact, some studies suggested that PRRSV GP4 and GP2 proteins interact with the receptor CD163.

As shown in Fig. 3, after binding to receptor(s) on host cells, PRRSV and EAV are postulated to enter through standard clathrin-dependent receptor-mediated endocytosis, with the viral nucleocapsid released into the cytosol following endosome acidification and membrane fusion. Positive-sense genomic RNA is directly translated into two large replicase polyproteins, pp1a and pp1ab. Posttranslational processing of pp1a and pp1ab involves a complex proteolytic cleavage cascade mediated by three (EAV) or four (PRRSV) ORF1a-encoded proteinases (EAV: papain-like proteinases in nsp1 β , nsp2, and serine proteinase in nsp4; PRRSV: papain-like proteinases in nsp1 α , nsp1 β , nsp2, and serine proteinase in nsp4) (Fig. 2). Viral nonstructural proteins then assemble into a replication and transcription complex (RTC) associated with vesicular double-membrane (double membrane vesicles, DMV) derived from the endoplasmic reticulum. The RTC directs synthesis of both genome-length and subgenome

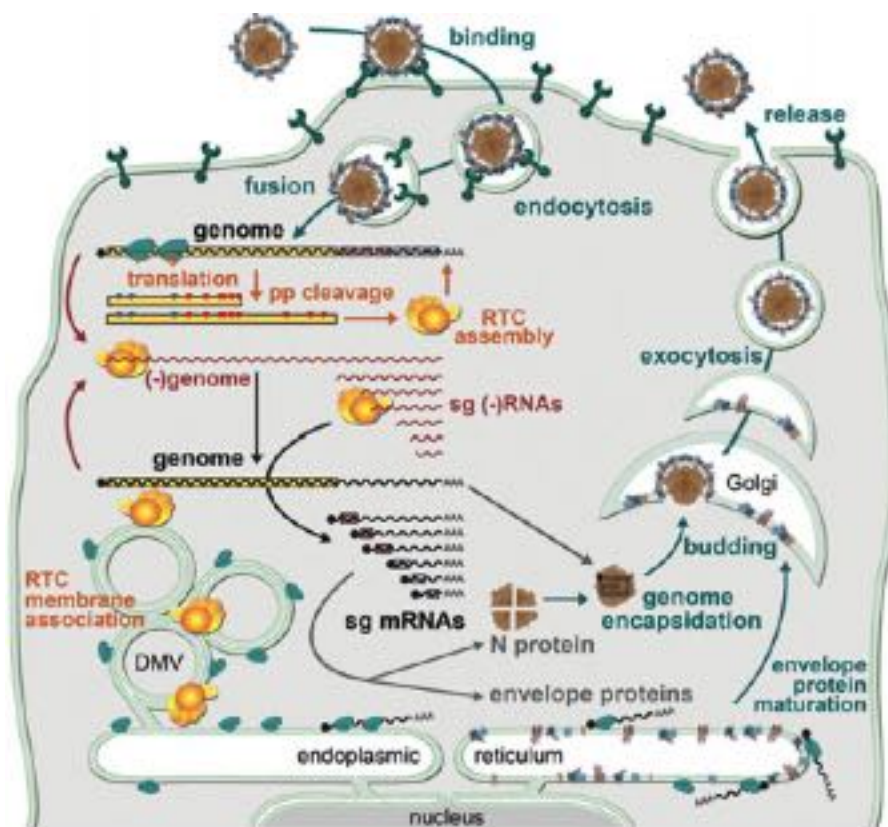


Fig. 3 Overview of arterivirus infection cycle. Following entry by receptor-mediated endocytosis and release of the genome into the cytosol, genome translation yields the replicase polyproteins pp1a and pp1ab which are cleaved by multiple internal proteases into nonstructural proteins. The viral nonstructural proteins assemble into a replication and transcription complex (RTC) on double membrane vesicles (DMV) to engage in minus-strand RNA synthesis. Both genome-length and subgenome (sg)-length minus strands are produced, with the latter serving as templates for the synthesis of sg mRNAs required to express the structural protein genes residing in the 3'-proximal quarter of the genome. Novel genomes are packaged into nucleocapsids that become enveloped by budding from smooth intracellular membranes, after which the new virions leave the cell using the exocytic pathway. Modified with permission from Snijder, E.J., Kikkert, M., 2013. Arteriviruses. In: Knipe, D.M., Howley, P.M., (Eds.), Fields Virology, sixth ed. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 859–879.

(sg)-length minus strands. The genome-length minus strand serves as a template for replication of genomic RNA. Nested sets of sg-length minus-strand RNAs serve as templates for generating positive-strand sg mRNAs from which structural proteins are translated. Novel genomes are encapsidated by N protein to form nucleocapsids that become enveloped when budding into the lumen of the smooth endoplasmic reticulum and/or Golgi complex. The formation of GP5/M heterodimer is a primary determinant of virus budding. After budding, new virions accumulate in intracellular vesicles and are then transported to the plasma membrane where the progeny virus is released.

Epidemiology

Pigs (*Sus scrofa*) and collared peccary (*Pecari tajacu*) are susceptible to PRRSV. The susceptibility of other species within the superfamily *Suoidae* (families *Suidae* and *Tayassuidae*) is likely, but it has not been determined. Members of family *Equidae*, i.e., horses, donkeys, zebras, as well as hybrid crosses (mules and hinnies), are susceptible to EAV. Neither virus is infectious for humans, i.e., zoonotic, and therefore do not affect public health.

Both viruses have a global distribution, with some exceptions. PRRSV is not present in Oceania, parts of South America (Argentina, Brazil), parts of Scandinavia (Finland, Norway, Sweden), and Switzerland. EAV has not been reported in Japan, Iceland, and New Zealand.

Both PRRSV- and EAV-infected animals shed virus in oronasal secretions, urine, semen, and perhaps briefly in feces. Shedding in boar (PRRSV) and stallion (EAV) semen presents the risk of long distance virus transmission to susceptible females through the use of artificial insemination, as has been documented in the literature.

PRRSV transmission can occur by intranasal, intramuscular, oral, intrauterine, and vaginal routes of exposure. Indirect transmission occurs by exposure to contaminated fomites (equipment, instruments, clothing, medications) or via aerosols. Parenteral

exposure (breaks in the skin) is a particularly efficient mode of transmission. Because PRRSV is present in oropharyngeal fluids for weeks, the dominance behaviors associated with mixing pigs can result in rapid transmission via bites, scrapes, and abrasions. Aerosol transmission is dependent on the presence of meteorological conditions able to support infectious virus and compatible with the movement of virus in space, e.g., low temperature, moderate levels of relative humidity, a rising barometric pressure, low sunlight levels, and slow, directional winds.

EAV transmission is primarily via respiratory and venereal routes. Aerosol transmission occurs via infectious droplet particles originating either from the respiratory tract or from the environment, e.g., urine splashes. Extensive EAV transmission can occur among groups of horses brought together for sales, racetracks, fairs, etc.

Vertical transmission, i.e., transplacental transmission to fetuses, is a feature of both viruses when females become infected in late gestation. PRRSV-naïve females infected in late gestation may farrow virus-infected piglets and will shed virus in mammary secretions throughout lactation. Likewise, EAV-susceptible mares infected in late gestation may produce congenitally-infected foals.

Both viruses can produce chronic persistent infections. Infectious PRRSV has been detected in individual animals at 175 days post exposure (DPE). The mechanism(s) by which the PRRSV persists in the face of an active immune response is unknown, but does not depend on the age of the pig at the time of exposure and does not involve evasion of immunity through *in vivo* viral mutation. Similarly, EAV carrier stallions may shed infectious levels of EAV in semen for weeks to years (see EAV Pathogenesis).

Clinical Features

PRRSV

Pigs of all ages are susceptible to PRRSV infection. Clinical signs are highly variable, but may include pyrexia (39–41°C), anorexia, lethargy, dyspnea, and variable hyperemia or cyanosis of skin on the extremities. Affected pigs may be leukopenic, anemic, and thrombocytopenic during the acute stage of the infection. The expression of clinical disease reflects the virulence of the PRRSV initiating the infection in combination with various co-factors, e.g., pig immune status, concurrent viral or bacterial co-infections, and herd management practices. Frequent recombination and mutation of PRRS viruses has led to the emergence of variants of atypically high virulence (ATV), including PRRSV-2 variants in North America, PRRSV-1 variants in Europe, and PRRSV-2 variants in Asia (“highly pathogenic” or HP-PRRSV). Some so-called “neurotropic” viruses may induce nervous signs, including head tilt, ataxia, lateral recumbency, and/or convulsions. Mortality ranges from 0% to 10% depending on variant virulence and the presence of co-factors, but may be higher with ATV viruses.

Infection in breeding-age boars may be inapparent or result in reduced libido and semen quality. Shedding of virus in semen can result in PRRSV transmission to sows through natural service or artificial insemination. Infection in pregnant females may result in transmission of PRRSV to the developing fetuses. Due to individual fetal placentation, some or all of the fetuses will be infected. Abortion rates may be 10%–50% (“abortion storms”). Aborted litters contain a mix of normal pigs, weak and small pigs, stillborn pigs, and partially or completely mummified pigs. If brought to term, litters may include viable, but PRRSV-infected, piglets. Transplacental transmission is most efficient in the third trimester, but some ATV PRRS viruses are able to cross the placenta with moderate efficiency during the first and second trimesters, resulting in aborted litters composed of smaller mummified fetuses, or irregular returns to heat without observed pregnancy loss.

In commercial herds, clinical signs range from subclinical to severe respiratory and reproductive diseases. In acute outbreaks in PRRSV-naïve herds (populations without prior immunity), clinical signs may be observed in all ages. More typically, swine herds are endemically infected with one or more PRRSV variants, in which case clinical signs are observed periodically in susceptible members of the population, i.e., weaned pigs (nursery-grower pigs) with declining levels of maternal immunity and replacement breeding stock (gilts or boars) that lack adequate immunity against the PRRS viruses circulating in the herd.

EAV

Clinical signs associated with EAV infection vary by strain virulence, animal age, route of infection, host susceptibility, and environmental conditions. Only one EAV serotype is recognized, but field isolates can be separated into those that produce subclinical infection (EAV KY63, EAV PA76, EAV KY77, and EAV CA95), mild disease (EAV SWZ64, EAV AUT68, EAV IL94, and EAV CA97), or moderate-to-severe disease (EAV KY84, EAV AZ87, EAV IL93, EAV PA96, and Bucyrus strain). The vast majority of adult horses are subclinically infected, but highly virulent strains can rarely result in death of adult animals. Pyrexia (up to 41°C) and leucopenia are the most common clinical signs. Less commonly, horses will develop depression, anorexia, nasal congestion and discharge, lymphadenopathy, mild-to-severe dyspnea, edema of the limbs, ventrum, scrotum, mammary gland, periorbital regions, and intermandibular space, conjunctivitis, corneal opacity, diarrhea, stiffness, icterus, and oral and cutaneous papules.

Infection in stallions can result in subfertility for 6–7 weeks post infection due to diminished libido and decreased semen quality. Genetically susceptible stallions can become persistently infected and continuously shed virus in semen. Acutely infected pregnant mares are at increased risk of abortion and 10%–71% will abort during the second to tenth month of gestation. Neonatal infection can present as severe dyspnea due to interstitial pneumonia and older foals are susceptible to progressive pneumoenteric syndrome. Most adult horses spontaneously recover, but young and debilitated animals may develop significant disease and die.

Pathogenesis and Immunity

PRRSV Pathogenesis

PRRSV infection consists of three stages: acute infection, persistence, and elimination. Acute infection is characterized by extensive viral replication in lung and lymphoid tissues and viremia (21–35 days). Persistence is marked by the resolution of viremia and progressively less PRRSV replication in lymph nodes and tonsils. Elimination begins when viral shedding ceases and is complete when virus is cleared from all tissues. The entire course of the infection can be ≥ 175 days and the detection of carrier animals is problematic in the latter stages of the infection.

PRRSV replicates in a subset of monocyte-derived cells that display CD163 and sialoadhesin (Sn) receptors, including pulmonary alveolar macrophages and pulmonary intravascular macrophages in the lung, monocyte-derived macrophages in lymphoid tissues, and dendritic cells and monocyte-derived macrophages that reside in perivascular locations in most organs. CD163 is required for PRRSV binding, internalization, and replication; but co-expression of Sn augments virus internalization. Most PRRS viruses replicate in CD163+Sn+ cells, but ATV variants may also replicate in CD163+Sn- cells, a factor that likely contributes to the greater breadth and severity of clinical disease, lesions, and mortality.

PRRSV-infected macrophages are killed by necrotic or apoptotic mechanisms, but the majority of cell death in PRRSV-infected tissues is due to indirect killing of uninfected cells (bystander apoptosis). In the lung, cells killed by bystander apoptosis are mostly macrophages, with fewer lymphocytes and alveolar pneumocytes. In lymph nodes and thymus, cells killed by bystander apoptosis are mostly lymphocytes, with fewer macrophages. ATV PRRSV variants induce higher levels of bystander apoptosis than less virulent variants. The reduction in the number, phagocytic capacity, and bactericidal capacity of pulmonary alveolar macrophages and pulmonary intravascular macrophages is thought to be responsible for an increased incidence of septicemia in PRRSV-infected pigs (*Streptococcus suis* and others) and to opportunistic bacterial bronchopneumonia. PRRSV infection alters innate immunity and inflammatory/immuno-regulatory cytokines. Secretion of pro-inflammatory cytokines from PRRSV-infected macrophages, including TNF- α , IL-1, and IL-6, promote the influx and activation of leukocytes, increased microvascular permeability, and induction of systemic effects, such as pyrexia, anorexia, and lethargy. Even higher levels of these and other pro-inflammatory cytokines in HP-PRRS viruses are suggested as a cause of death similar to “macrophage-activation-syndrome” or “cytokine-storms” in humans.

Reproductive failure usually occurs in females infected during the third trimester of gestation. During the first two trimesters, CD163+Sn- macrophages predominate in the placenta, but in the third trimester the number of highly permissive CD163+Sn+ macrophages continually increases. Thus, in the third trimester, the virus is able to efficiently transit the maternal-fetal-interface and encounter PRRSV-permissive macrophages in the fetal placenta. In acutely-infected gestating females, PRRSV is present in the maternal endometrium, primarily in CD163+Sn+ macrophages. These infected cells undergo direct apoptosis and also induce bystander apoptosis in both the maternal endometrium and interfacing fetal placenta, including endometrial epithelial cells and placental trophoblasts. This results in the direct apposition of PRRSV-infected, macrophage-rich, endometrial lamina propria and fetal placental stroma. At the maternal-fetal-interface for a particular fetus, the degree of direct and bystander apoptosis is directly proportional to the number of PRRSV-infected macrophages in the maternal endometrium, the concentration of PRRSV in the fetus' thymus, and the risk for fetal death. PRRSV replicates to its highest concentration in fetal thymus, tonsils, and lymph nodes, thereby contributing to fetal death, or if pregnancy continues to term, the birth of PRRSV-infected piglets. PRRSV ATV variants induce higher levels of apoptosis at the maternal-fetal-interface and in fetal tissues, and are able to replicate in CD163+Sn- macrophages; features that explain, at least in part, the higher rates of abortion associated with ATVs across a broader range of gestational ages.

PRRSV Immunity

PRRSV-induced alterations of immuno-regulatory cytokines and in the numbers of various lymphocyte subsets cause a prolonged delay in the implementation of an effective adaptive immune response. Regardless, PRRSV infection induces an immune response that slowly eliminates the virus and establishes memory that may (*or may not*) protect against other PRRS viruses. Non-neutralizing antibodies against PRRSV structural and nonstructural proteins appear 7–10 DPE. Neutralizing antibodies appear 21–28 DPE, i.e., concurrently with the resolution of viremia. Given their role in cell tropism and interactions with cell receptors, it has been hypothesized that neutralizing antibodies target glycoproteins GP2, GP3, and GP4, but data confirming this hypothesis have not yet been produced. Despite the fact that their role in controlling PRRSV is not clearly established, neutralizing antibodies are the best predictor of the level and duration of viremia. Interferon- γ ELISPOT has demonstrated a consistent T-cell response to PRRSV, but its significance is uncertain. Overall, the protracted development of neutralizing antibody is credited for the prolonged PRRS viremia and the delayed cell-mediated immune response for the persistence of PRRSV in lymphoid tissues.

EAV Pathogenesis

Virus transmission can occur by either venereal or respiratory routes of exposure. Within 24 h of respiratory exposure, EAV replicates in alveolar macrophages and bronchiolar epithelial cells. The virus localizes in tonsils and regional lymph nodes within 48 h and then rapidly disseminates throughout the host by cell-associated viremia. Viremia lasts 3–19 DPE and nasal shedding

occurs for 2–21 DPE. Once viremic, the virus replicates in macrophages, endothelial cells, and vascular smooth muscle myocytes. The combination of direct virus-mediated injury to endothelial cells and leukocyte infiltration results in vascular necrosis and panvasculitis. The resultant vascular injury and increased permeability contribute to perivascular hemorrhage and edema that grossly manifests as the characteristically observed clinical sign. Virus clearance typically occurs by 28 DPE in concurrence with a rising serum neutralizing antibody response. Strain virulence and host genetic factors greatly influence the host proinflammatory immune response and disease severity. EAV virulence determinants are complex and involve both nonstructural and envelope proteins. EAV exhibits tropism for CD14⁺ monocytes through interactions with GP2, GP3, GP4, GP5, and M envelope proteins and CD3⁺ lymphocytes through interactions with GP2, GP4, GP5, and M envelope proteins.

Host genetic factors can contribute to the clinical outcome. Horses can be separated into EAV-susceptible or -resistant phenotypes based on the *in vitro* susceptibility of their CD3⁺ lymphocytes. Based on genome-wide association studies, the susceptible phenotype was associated with a common, genetically dominant haplotype located on equine chromosome 11. Further investigation identified two allelic variants of chemokine ligand 16 (CXCL16) with strong associations to the susceptible and resistant phenotypes. The susceptible CXCL16 phenotype has EAV receptor activity and the resistant phenotype lacks receptor activity. Susceptible vs resistant phenotype horses differ in proinflammatory (tumor necrosis factor α , interleukin (IL)-1 β , IL-6, and IL-8) and immunomodulatory cytokine mRNA expression. In horses possessing the resistant phenotype, clinical signs are more severe, but clearance of the virus is rapid. Stallions exhibiting the susceptible phenotype are more prone to develop persistent infection. The CXCL16 phenotypes vary by breed and correlate to observed variations in breed EAV seroprevalence.

Persistently infected carrier stallions serve as the natural reservoir of EAV. Persistent infection is testosterone dependent and occurs in 10%–70% of infected stallions and sexually mature colts. The susceptible CXCL16 CD3⁺ T lymphocyte phenotype is at higher risk for developing the EAV carrier state. Carrier stallions may never eliminate infection or may eventually clear infection after weeks, months, or years. In persistently infected stallions, the virus localizes to the reproductive tract where it primarily resides in vimentin⁺ stromal cells and lymphocytes in the ampullae of the vas deferens and to a lesser degree in other accessory sex organs. Microscopic lesions in persistently infected stallions consist of mild-to-moderate lymphoplasmacytic inflammation in the accessory sex organs. Long-term carrier stallions show no clinical signs of infection and have no alterations in semen quality. Carrier animals retain high serum neutralizing antibody levels and have significant local mucosal antibodies, but EAV persists and is continually shed into semen. EAV immune evasion mechanisms are suspected to play a role.

Pregnant mares vertically transmit EAV to the fetoplacental unit. Within the fetus and placenta, the virus replicates in a wide array of cell lineages, but rarely results in fetoplacental pathology. Fetal pathology, when present, consists of necrotizing vasculitis, tissue hemorrhage and edema, and effusion. Abortion is hypothesized to occur as a result of myometrial vasculitis and subsequent placental hypoxia and detachment, or by direct viral replication within the fetus resulting in fetal stress and expulsion.

EAV Immunity

EAV infection induces protective immunity. Complement-fixing and neutralizing antibodies develop 1–2 weeks following infection. Complement-fixing antibodies remain for 8 months, but neutralizing antibodies remain for years. EAV-infected horses develop immune responses to GP5, N, and M structural proteins and non-structural proteins 2, 4, 5, and 12. Foals acquire passive immunity from immune mares, and colostrum-associated virus neutralizing antibodies remain for 2–7 months after birth.

Diagnosis

PRRSV

Acute PRRS is suggested by clinical signs and lesions, but a definitive diagnosis requires laboratory confirmation based on evaluation of specimens, e.g., lung, lymphoid tissues, bronchoalveolar lavage fluids, serum, semen, and/or oral fluids. Preferred specimens in cases of reproductive failure include fetal lung and thymus, serum collected from dams shortly after abortion, and “processing fluid” from live-born littermates or cohorts. Processing fluid is a composite sample consisting of the blood-serum exudate recovered from tissues collected at the time of castration and/or tail-docking of neonatal piglets.

Lesions in lungs and lymph nodes include lymphoplasmacytic bronchiointerstitial pneumonia with necrotic intra-alveolar macrophages and variable hyperplasia of type II pneumocytes. Lymph nodes are enlarged and edematous with variable necrosis of lymphoid follicles. In addition to lesions in lungs and lymph nodes, sows with reproductive signs often have lymphoplasmacytic perivascular endometritis. Lesions are inconsistent, mild, and observed in a minority of PRRSV-infected fetuses.

PRRSV antigens can be visualized in frozen tissue sections by fluorescent antibody staining or in paraffin-embedded tissue sections prepared from tissues fixed in 10% neutral-buffered formalin by immunohistochemical staining. Monoclonal antibodies specific for the highly conserved nucleocapsid protein are typically used. Both fluorescent antibody and immunohistochemical staining are less analytically sensitive than reverse-transcription polymerase chain reaction (RT-PCR).

Virus isolation in cell culture is not routinely used for diagnosis, but may be performed when viable PRRSV is needed for further analyses. PRRSV is heat-labile, so samples should be immediately chilled (4°C) and maintained on ice until tested. The preferred antemortem sample is serum, but lung and lymphoid tissue samples collected at necropsy may also be used. PRRSV is fastidious and inoculation of several cell types maximizes the chances of isolation, e.g., primary porcine alveolar macrophages,

African monkey kidney cells (MA-104, CL-2621, MARC-145), and an immortalized pulmonary alveolar macrophage cell line, ZMAC-1.

Most commonly, RT-PCR is used to confirm PRRSV infection. Reverse transcription loop-mediated isothermal amplification and *in situ* hybridization assays have been described for detection of PRRSV RNA, but are not widely used. Commercial PRRSV RT-PCRs are analytically sensitive, analytically specific, quantitative, and compatible with high-throughput testing. PRRSV is detected in serum 6–48 hours post exposure, with peak virus concentration 4–14 DPE, reaching 1×10^2 – 10^5 TCID₅₀ per ml of serum or gram of lung tissue, but as high as $\geq 1 \times 10^8$ TCID₅₀ in ATV variants. Notably, RT-PCR testing of serum will not detect carrier animals in the persistence or elimination phases of infection. Individual specimens may be “pooled” by combining roughly equal volumes into one sample for PRRSV rRT-PCR testing. This approach reduces testing costs, but the inclusion of virus-negative specimens will dilute the concentration of virus and increase the probability of false-negative RT-PCR results.

Interpretation of PRRSV RNA assays is complicated by the use of modified-live (MLV) PRRSV vaccines. Like field viruses, MLV vaccine viruses replicate in vaccinated animals, or in non-vaccinated animals from vaccinated populations, for a prolonged period. Sequencing of ORF5, the major envelope protein gene, or whole genome sequencing with phylogenetic analysis, are used to differentiate MLV viruses from wild-type viruses and, in epidemiological studies, to describe and compare viruses in farms, production systems, and geographic regions.

A variety of tests have been described for the detection of PRRSV antibodies, e.g., virus neutralization, fluorescent microsphere immunoassay, immunoperoxidase monolayer assay, and indirect fluorescent assay. Commercial enzyme-linked immunosorbent assays (ELISAs) for the detection of antibody in serum and/or oral fluid specimens are commonly used because they can provide high diagnostic sensitivity and specificity and are compatible with high throughput testing. Most commercial ELISAs detect IgG antibodies against both PRRSV-1 and PRRSV-2 nucleocapsid antigens, but some are species-specific.

A disadvantage of antibody testing is that current assays cannot differentiate between antibodies produced in response to wild-type infection vs MLV vaccination. An advantage of antibody testing is that antibody is detected for ≥ 6 months and positive results reveal prior PRRSV exposure, even in RT-PCR-negative animals. Periodic population sampling (serum or oral fluids) and antibody testing can provide a cost-effective approach for monitoring PRRSV circulation or vaccination.

EAV

EAV infection may be subclinical, variable in clinical presentation, or resemble other infectious and non-infectious equine diseases. Therefore, a definitive diagnosis of EAV requires confirmation based on laboratory testing of clinical specimens. Appropriate specimens from acutely infected horses include nasopharyngeal swabs, nasal washings, conjunctival swabs, and whole blood collected in anticoagulant, e.g., ethylenediaminetetraacetic acid (EDTA) or sodium citrate. Swabs should be placed in a suitable viral transport medium. Fresh and formalin-fixed lung, liver, spleen, thymus, and lymph nodes should be collected from animals that have died and, in addition, both allantochorion and amnion should be collected from aborted fetuses. Fresh samples should be held at 4°C and submitted to the laboratory by overnight delivery. Freezing of samples should generally be avoided.

Virus isolation and RT-PCR can be attempted on samples collected from acutely infected horses, aborted fetoplacental units, and semen from carrier stallions. Rabbit kidney (RK-13) cells are optimal for EAV isolation, but Vero cells and equine lung cells can also be used. Virus isolation on RK-13 cells is the World Organization for Animal Health (OIE) standard for the detection of EAV in semen and is the recommended test for international trade. Virus isolation has comparable sensitivity to RT-PCR, but in contrast to RT-PCR, cannot produce rapid results. Various RT-PCR assays (single step, nested, real-time, and insulated isothermal) have been developed to target specific genes (ORFs 1b, 3, 4, 5, 6, and 7), but have varying degrees of analytical sensitivity and specificity. Reagent kits can also differ in analytical sensitivity. A recently developed RT-insulated isothermal PCR assay was shown to have a diagnostic sensitivity of 100% and specificity of 98.33% and overall diagnostic accuracy similar to virus isolation.

Frozen or formalin-fixed tissues from deceased animals and fetoplacental units can be used to visualize virus by immunohistochemistry and *in situ* hybridization. Immunohistochemistry should be conducted using monospecific polyclonal serum against EAV or monoclonal antibody derived from the highly conserved EAV nucleocapsid protein. Conventional RNA *in situ* hybridization was less sensitive than immunohistochemistry or an *in situ* hybridization signal amplified system (RNAscope®) using oligonucleotide probes targeting ORFs 5, 6, 7, and the 3′ untranslated region.

Serum antibody assays have been developed to indirectly identify horses exposed to EAV and to detect acutely infected animals. The virus neutralization test is the World Organization for Animal Health (OIE) standard for the serologic diagnosis of EAV. Both diagnostically sensitive and highly specific, a diagnosis of acute infection requires the demonstration of a significant increase, e.g., ≥ 4 -fold increase, in neutralizing antibody titers in serum samples taken 3–4 weeks apart from the animal of interest. Complement fixation, indirect fluorescent assay, agar gel immunodiffusion, various ELISAs, and a microsphere immunoassay have also been described, but variation in viral antigen and serum can affect test performance.

In stallions, a diagnosis of persistent EAV infection is based on a combination of the virus neutralization test and the detection of EAV in semen. Stallions can be serologically screened for neutralizing antibody because all carrier stallions have virus neutralization titers $\geq 1:8$. Persistence is then confirmed by virus isolation or detection of EAV nucleic acid in the sperm-rich fraction of semen or by test breeding 2 seronegative mares and assessing them for seroconversion within 4 weeks post-breeding.

Prevention and Control

There are no specific anti-viral treatments for EAV or PRRSV. In clinically-affected animals, supportive veterinary care and the use of anti-inflammatory agents and/or antimicrobials may be warranted. Overall, the focus must be on prevention and control, i.e., stopping the transmission of virus to susceptible animals using appropriate biosecurity measures and reducing the clinical effect of infection through the use of vaccines.

Prior to introduction into herds, animals should be quarantined and tested to determine infection status. Thereafter, animals and/or population should be monitored routinely for changes in infection status. Knowledge of true infection status should be an important consideration in management, husbandry, and biosecurity decisions. Shedding of virus by infected animals in body secretions and excretions leads to contamination of the environment with infectious virus, i.e., housing, trucks, trailers, transport equipment, and other fomites. Both PRRSV and EAV can be inactivated by lipid solvents and by common disinfectants and detergents. Vaccines may be useful in managing EAV and PRRSV, but their limitations and recommendations for use should be fully appreciated.

Further Reading

- Balasuriya, U.B.R., Carossino, M., Timoney, P.J., 2018. Equine viral arteritis: A respiratory and reproductive disease of significant economic importance to the equine industry. *Equine Veterinary Education* 30, 497–512.
- Balasuriya, U.B.R., Go, Y.Y., MacLachlan, N.J., 2013. Equine arteritis virus. *Veterinary Microbiology* 167, 93–122.
- Sarkar, S., Chelvarajan, L., Go, Y.Y., *et al.*, 2016. Equine arteritis virus uses equine CXCL16 as an entry receptor. *Journal of Virology* 90, 3366–3384.
- Snijder, E.J., Kikkert, M., 2013. Arteriviruses. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 859–879.
- Snijder, E.J., Kikkert, M., Fang, Y., 2013. Arterivirus molecular biology and pathogenesis. *Journal of General Virology* 94, 2141–2163.
- Zimmerman, J.J., Dee, S.A., Holtkamp, D., *et al.*, 2019. Porcine reproductive and respiratory syndrome viruses (porcine arteriviruses). In: Zimmerman, J.J., Karriker, L.A., Ramirez, A., *et al.* (Eds.), *Diseases of Swine*, eleventh ed. John Wiley and Sons, Inc, pp. 685–708.

Relevant Websites

- <http://www.oie.int/en/standard-setting/terrestrial-manual/access-online/>
OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals.
- <http://www.oie.int/international-standard-setting/terrestrial-code/access-online/>
OIE Terrestrial Animal Health Code.

Prions of Vertebrates

Jonathan DF Wadsworth and John Collinge, UCL Institute of Prion Diseases, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J.D.F. Wadsworth, J. Collinge, Prions of Vertebrates, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:10.1016/B978-0-12-801238-3.02648-9.

Glossary

Codon 129 polymorphism There are two common forms of *PRNP* encoding either methionine or valine at codon 129; a major determinant of genetic susceptibility to and phenotypic expression of prion disease.

Conformational selection model A hypothetical model which explains transmission barriers on the basis of overlap of permissible conformations of PrP^{Sc} (prion strains) between mammalian species. Although a wide range of mammalian PrP^{Sc} conformations may be possible, only a subset will be compatible with each individual PrP primary structure.

Molecular strain typing A means of rapidly differentiating prion strains by biochemical differences in PrP^{Sc}.

Prion The infectious agent causing prion diseases.

Prion incubation period The interval between exposure to prions and the development of neurological signs of prion disease; typically months even in rodent models and years to decades in humans.

Prion protein (PrP) A glycoprotein encoded by the host genome and expressed in many tissues but especially on the surface of neurons.

Prion strain Distinct isolates of prions originally identified and defined by biological characteristics which breed true in inbred mouse lines.

PRNP The human prion protein gene; mouse gene is designated *Prnp*.

Protein-only hypothesis That prions lack a nucleic acid genome, are composed principally or solely of abnormal isomers of PrP (PrP^{Sc}), and replicate by recruitment of host PrP^C.

PrP Prion protein.

PrP^C The normal cellular isoform of PrP rich in α -helical structure.

PrP^{Sc} The “scrapie” or disease-associated isoform of PrP which differs from PrP^C in its conformation and is generally found as insoluble aggregated material rich in β -sheet structure.

Subclinical infection A state where host prion propagation is occurring but which does not produce clinical disease during normal lifespan; essentially a carrier state of prion infection.

Transmission barrier This describes the observation that transmission of prions from one species to another is generally inefficient when compared to subsequent passage in the same host species.

Introduction

The prion diseases are a closely related group of neurodegenerative conditions which affect both humans and animals. They have previously been described as the subacute spongiform encephalopathies, slow virus diseases, and transmissible dementias, and include scrapie in sheep, bovine spongiform encephalopathy (BSE) in cattle, chronic wasting disease in deer and elk, and the human prion diseases, Creutzfeldt–Jakob disease (CJD), variant CJD (vCJD), Gerstmann–Sträussler–Scheinker disease (GSS), fatal familial insomnia (FFI), and kuru. While rare in humans, prion diseases are an area of intense research interest. This is first because of their unique biology, in that the transmissible agent appears to be devoid of nucleic acid and to consist of a post-translationally modified host protein. Secondly, because of the ability of these and related animal diseases to cross from one species to another, sometimes by dietary exposure, there has been widespread concern that the exposure to the epidemic of BSE poses a distinct threat to public health in the United Kingdom and other countries. The extremely prolonged and variable incubation periods of these diseases, particularly when crossing a transmission barrier, means that it will be some years before the full consequences of human exposure to bovine prions can be predicted with confidence. In the meantime, we are faced with the possibility that significant numbers in the population may be incubating this disease and that they might pass it on to others via blood transfusion, blood products, tissue and organ transplantation, and other iatrogenic routes.

Aberant Prion Protein Metabolism is the Central Feature of Prion Disease

The nature of the transmissible agent in prion disease was a subject of heated debate for many years. The understandable initial assumption that the causative agent of “transmissible dementias” must be some form of virus was challenged by the failure to directly demonstrate such a virus (or indeed any immunological response to it) and by the remarkable resistance of the transmissible agent to treatments that inactivate nucleic acids. These findings led to suggestions that the transmissible agent may be devoid of nucleic acid and might be a protein. Subsequently in 1982, Prusiner and co-workers isolated a protease-resistant sialoglycoprotein, designated the prion

protein (PrP), that was the major constituent of infective fractions and was found to accumulate in affected brain. The term prion (from proteinaceous infectious particle) was proposed by Prusiner to distinguish the infectious pathogen from viruses or viroids and was defined as “small proteinaceous infectious particles that resist inactivation by procedures which modify nucleic acids.”

Initially, PrP was assumed to be encoded by a gene within the putative slow virus thought to be responsible for these diseases; however, amino acid sequencing of part of PrP and the subsequent recovery of cognate cDNA clones using an isocoding mixture of oligonucleotides led to the realization that PrP was encoded by a single-copy chromosomal gene rather than by a putative viral nucleic acid. Following these seminal discoveries, a wealth of data has now firmly established that the central and unifying hallmark of the prion diseases is the aberrant metabolism of PrP, which exists in distinct conformational states with different physicochemical properties. The normal cellular form of the protein, referred to as PrP^C, is a highly conserved cell surface glycosylphosphatidylinositol (GPI)-anchored sialoglycoprotein that is sensitive to protease treatment and soluble in detergents. In contrast, disease-associated isoforms of PrP are found only in prion-infected tissue and composed of detergent-insoluble infectious polymeric PrP assemblies some of which acquire protease-resistance and are classically designated as PrP^{Sc} (the scrapie isoform). Due to its physicochemical properties, the precise atomic structure of the infectious particle or prion is still undetermined. The essential role of host PrP for prion propagation and pathogenesis is demonstrated by the fact that knockout mice lacking the PrP gene (*Prnp*^{0/0} mice) are entirely resistant to prion infection, and that reintroduction of PrP transgenes restores susceptibility to prion infection in a species-specific manner.

Human Prion Diseases are Biologically Unique

Human prion diseases are biologically unique and can be divided etiologically into inherited, sporadic, and acquired forms. Approximately 85% of cases of human prion disease occur sporadically as Creutzfeldt–Jakob disease (sporadic CJD) at a rate of roughly 1 case per million population per year across the world, with an equal incidence in men and women. The etiology of sporadic CJD is unknown, although hypotheses include somatic *PRNP* mutation, or the spontaneous conversion of PrP^C into PrP^{Sc} as a rare stochastic event. Polymorphism at residue 129 of human PrP (encoding either methionine (M) or valine (V)) powerfully affects genetic susceptibility to human prion diseases. About 38% of Europeans are homozygous for the more frequent methionine allele, 51% are heterozygous, and 11% homozygous for valine. Homozygosity at *PRNP* codon 129 predisposes to the development of sporadic and acquired CJD and is most strikingly observed in vCJD. At the time of writing all but one pathologically proven cases of vCJD studied so far have been homozygous for codon 129 methionine of *PRNP* with one heterozygous patient confirmed in 2016.

Approximately 15% of human prion diseases are associated with autosomal dominant pathogenic mutations in *PRNP*. How pathogenic mutations in *PRNP* cause prion disease is yet to be resolved, however, in most cases the mutation is thought to lead to an increased tendency of PrP^C to form a pathogenic PrP isoform. However experimentally manipulated mutations of the prion gene can lead to spontaneous neurodegeneration without the formation of detectable protease-resistant PrP. These findings raise the question of whether all inherited forms of human prion disease invoke disease through the same mechanism, and in this regard it is currently unknown whether all are transmissible by inoculation. Because of the extensive phenotypic variability seen in inherited prion disease and its ability to mimic other neurodegenerative conditions, notably Alzheimer’s disease, *PRNP* analysis should be considered in all patients with undiagnosed dementing and ataxic disorders. Recently, a novel clinical and pathological phenotype associated with a Y163X mutation in *PRNP* was identified which generates a prion protein systemic amyloidosis, characterized by slow disease progression, diarrhea, autonomic failure and neuropathy. This newly recognized novel disease phenotype indicates that *PRNP* analysis should also be considered in the investigation of unexplained chronic diarrhea associated with a neuropathy or an unexplained syndrome similar to familial amyloid polyneuropathy.

Although the human prion diseases are transmissible diseases, acquired forms have, until recently, been confined to rare and unusual situations. The two most frequent causes of iatrogenic CJD occurring through medical procedures have arisen as a result of implantation of dura mater grafts and treatment with human growth hormone derived from the pituitary glands of human cadavers. Less frequent incidences of human prion disease have resulted from iatrogenic transmission of CJD during corneal transplantation, contaminated electroencephalographic (EEG) electrode implantation, and surgical operations using contaminated instruments or apparatus. The most well-known incidences of acquired prion disease in humans resulting from a dietary origin have been kuru that was caused by cannibalism among the Fore linguistic group of the Eastern Highlands in Papua New Guinea, and more recently the occurrence of vCJD in the United Kingdom and some other countries that is causally related to human exposure to BSE prions from cattle. Incubation periods of acquired prion diseases in humans can be extremely prolonged, and it remains unclear how many people within the UK and elsewhere may have become infected and what proportion of these individuals may go on to develop clinical disease rather than remaining as asymptomatic carriers.

The Protein-Only Hypothesis of Prion Propagation

Despite extensive investigation, no evidence for a specific prion-associated nucleic acid has been found. Instead, a wide body of data supports the idea that infectious prions consist principally or entirely of an abnormal isoform of PrP. PrP^{Sc} is derived from PrP^C by a post-translational mechanism and neither amino acid sequencing nor systematic study of known covalent post-translational modifications have shown any consistent differences between PrP^C and PrP^{Sc}. The protein-only hypothesis, in its current form, argues that prions propagate by means of seeded protein polymerization, a process that involves the recruitment of PrP monomers to fibrillar

assemblies of PrP followed by fission of the polymer to produce more “seeds” and that this conversion involves only conformational change. However, the underlying molecular events during infection that lead to the conversion of PrP^C to prototypical PrP^{Sc} and how prototypical PrP^{Sc} accumulation leads to neurodegeneration remain poorly defined. Notably, in multiple prion strain/host combinations it now appears that the majority of disease-related PrP and prion titer may be destroyed by proteases under conditions that are typically employed to detect prototypical PrP^{Sc}. These data suggest that the term PrP^{Sc}, often used synonymously with prion infectivity, should be restricted to material as classically biochemically defined (infectious detergent-insoluble, protease-resistant PrP assemblies). Such prototypical PrP^{Sc} constitutes a small (variable) fraction of total disease-related PrP isoforms and its proportional contribution to both infectivity and neurotoxicity remains unclear. Resolving the structural relationship between protease-sensitive and protease-resistant disease-related infectious PrP assemblies and whether these simply comprise different-sized aggregates of essentially the same PrP conformer has yet to be established. It is now clear that a full understanding of prion propagation will require knowledge not only of the structure of PrP^C and prototypical PrP^{Sc} assemblies but also the role of other protease-sensitive disease-related PrP isoforms that are generated during the course of prion disease pathogenesis.

Structure and Putative Function of PrP^C

PrP is highly conserved among mammals, has been identified in marsupials, birds, amphibians, and fish, and may be present in all vertebrates. It is expressed during early embryogenesis and is found in most tissues in the adult with the highest levels of expression in the central nervous system, in particular in association with synaptic membranes. PrP is also widely expressed in cells of the immune system. As a GPI-anchored cell surface glycoprotein, it has been speculated that PrP may have a role in cell adhesion or signaling processes, but its precise cellular function has remained obscure.

Mice lacking PrP as a result of gene knockout (*Prnp*^{0/0} mice) show no gross phenotype; however, these mice are completely resistant to prion disease following inoculation and do not replicate prions. *Prnp*^{0/0} mice do however show subtle abnormalities in synaptic physiology, peripheral myelination and in circadian rhythms and sleep. While the relative normality of *Prnp*^{0/0} mice was thought to result from effective adaptive changes during development, data from *Prnp* conditional knockout mice suggest this is not the case. These mice undergo ablation of neuronal PrP expression at 9 weeks of age and remain healthy without evidence of neurodegeneration or an overt clinical phenotype. Thus, acute loss of neuronal PrP in adulthood is tolerated and the pathophysiology of prion diseases appears to be unrelated to loss of normal PrP function in neurons.

Nuclear magnetic resonance (NMR) measurements and crystallographic determination of normal PrP structures from numerous mammalian species, including human PrP, show that they have essentially the same conformation. Following cleavage of an N-terminal signal peptide, and removal of a C-terminal peptide on addition of a GPI anchor, the mature PrP^C species consists of an N-terminal region of about 100 amino acids which is unstructured in the isolated molecule in solution and a C-terminal segment, also around 100 amino acids in length. The C-terminal domain is folded into a largely α -helical conformation (three α -helices and a short antiparallel β -sheet) and stabilized by a single disulfide bond linking helices 2 and 3. There are two asparagine-linked glycosylation sites. The N-terminal region contains a segment of five repeats of an eight-amino-acid sequence (the octapeptide-repeat region), expansion of which by insertional mutation leads to inherited prion disease. While unstructured in the isolated molecule, it seems likely that the N-terminal region of PrP may acquire coordinated structure *in vivo* through coordination of either Cu²⁺ or Zn²⁺ ions.

Structural Properties of PrP^{Sc}

PrP^{Sc} is extracted from affected brains as highly aggregated, detergent insoluble material that is not amenable to high-resolution structural techniques. However, Fourier transform infrared spectroscopic methods show that PrP^{Sc}, in sharp contrast to PrP^C, has a high β -sheet content. PrP^{Sc} is covalently indistinguishable from PrP^C but can be distinguished from PrP^C by its partial resistance to proteolysis and its marked insolubility in detergents. Under conditions in which PrP^C exists as a detergent-soluble monomer and is completely degraded by the nonspecific protease, proteinase K, PrP^{Sc} exists in aggregated assemblies with the C-terminal two-thirds of the protein showing marked resistance to proteolytic degradation leading to the generation of N-terminally truncated fragments of di-, mono-, and nonglycosylated PrP (Fig. 1).

Due to the difficulty in performing structural studies on native PrP^{Sc} assemblies considerable international effort has instead focused on attempting to produce infectious synthetic β -sheet-rich forms of recombinant PrP which may be amenable to NMR or crystallographic structure determination. Direct *in vitro* mixing experiments were first attempted to produce PrP^{Sc} *in vitro*. In such experiments PrP^{Sc} was used in excess as a seed to convert PrP^C to a protease-resistant form, designated PrP^{Res}. While there are now many historical examples in the literature of conditions that generated PrP^{Res}, such reactions were not able to demonstrate *de novo* production of prion infectivity. However, the discovery of a protein misfolding cyclic amplification (PMCA) system, established that substantial amplification of PrP^{Res} and prion infectivity can be achieved in the absence of living cells. Although the generation of infectious prions from recombinant PrP alone has been reported the low prion titer of such preparations has so far precluded meaningful structural analyses. Consequently, the requirement to isolate and study *ex-vivo* prions continues.

Progress in understanding prion structure has been severely hindered by the difficulty of isolating relatively homogeneous prion particles from infected tissue and unequivocally correlating infectivity with composition and structure. However new methods for isolating exceptionally pure, high-titer infectious prion preparations from mouse brain have recently been developed

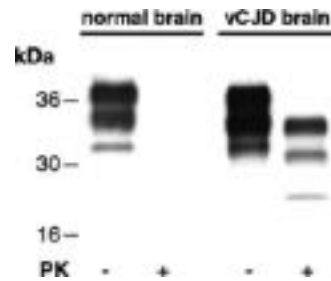


Fig. 1 PrP analysis by immunoblotting. Immunoblot analysis of normal human brain and vCJD brain homogenate before and after treatment with proteinase K (PK). PrP^C in both normal and vCJD brain is completely degraded by proteinase K, whereas PrP^{Sc} present in vCJD brain shows resistance to proteolytic degradation leading to the generation of N-terminally truncated fragments of di-, mono-, and non-glycosylated PrP.

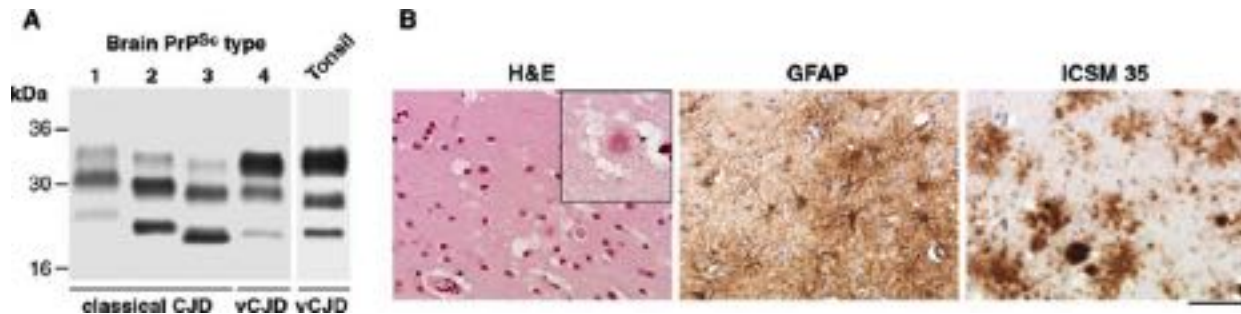


Fig. 2 Characterization of disease-related prion protein in human prion disease. (a) Immunoblots of proteinase K-digested tissue homogenate with monoclonal antibody 3F4 showing PrP^{Sc} types 1–4 in human brain and PrP^{Sc} type 4t in vCJD tonsil. Types 1–3 PrP^{Sc} are seen in the brain of classical forms of CJD (either sporadic or iatrogenic CJD), while type 4 PrP^{Sc} and type 4t PrP^{Sc} are uniquely seen vCJD brain or tonsil, respectively. (b) Brain from a patient with vCJD showing spongiform neurodegeneration following hematoxylin- and eosin staining (H&E), reactive proliferation of astroglial cells following staining with a monoclonal antibody recognizing glial-fibrillary acidic protein (GFAP), and abnormal PrP immunoreactivity following immunohistochemistry using anti-PrP monoclonal antibody ICSM 35 (ICSM 35). Scale (main panels) = 100 μ m. Inset, high-power magnification of a florid PrP plaque. Courtesy of Professor Sebastian Brandner.

which showed that PrP^{Sc} in these preparations is assembled into rod-like assemblies (PrP rods) that faithfully transmit prion strain-specific phenotypes when inoculated into mice. Sensitive cell culture infectivity assays demonstrated that the PrP rods are intrinsically infectious and that aggregates of the rods behave as discrete infectious particles in cell culture. Electron tomography of negatively-stained prion rods from multiple prion strains revealed a common three-dimensional architecture comprising a pair of short, intertwined fibers, each with a double helical repeating substructure, separated by a distinct gap of 8–10 nm in width. The architecture of the infectious PrP rods, which cause lethal neurodegeneration, readily differentiates them from all other protein assemblies so far characterized in other neurodegenerative diseases.

Prion Disease Pathogenesis

Microscopic examination of the central nervous system of humans or animals with prion disease reveals typical characteristic histopathologic changes, consisting of neuronal vacuolation and degeneration, which gives the cerebral gray matter a microvacuolated or “spongiform” appearance, and a reactive proliferation of astroglial cells (Fig. 2). Demonstration of abnormal PrP immunoreactivity, or more specifically biochemical detection of PrP^{Sc} in brain material by immunoblotting techniques, is diagnostic of prion disease (Figs. 1 and 2) and some forms of prion disease are characterized by deposition of amyloid plaques composed of insoluble aggregates of PrP. The histopathological features of vCJD are remarkably consistent and distinguish it from other human prion diseases with large numbers of PrP-positive amyloid plaques that differ in morphology from the plaques seen in kuru and GSS in that the surrounding tissue takes on a microvacuolated appearance, giving the plaques a florid appearance (Fig. 2).

Although the pathological consequences of prion infection occur in the central nervous system and experimental transmission of these diseases is most efficiently accomplished by intracerebral inoculation, most natural infections do not occur by these means. Indeed, administration to sites other than the central nervous system is known to be associated with much longer incubation periods, which in humans may extend to 50 years or more. Experimental evidence suggests that this latent period is associated with clinically silent prion replication in lymphoreticular tissue, whereas neuroinvasion takes place later. The M-cells in the intestinal epithelium appear to mediate prion entry from the gastrointestinal lumen into the body, and follicular dendritic cells

(FDCs) are thought to be essential for prion replication and for accumulation of disease-related PrP assemblies within secondary lymphoid organs. B-cell-deficient mice are resistant to intraperitoneal inoculation with prions probably because of their involvement with FDC maturation and maintenance. However, neuroinvasion is possible without FDCs, indicating that other peripheral cell types can replicate prions. The interface between FDCs and sympathetic nerves represents a critical site for the transfer of lymphoid prions into the nervous system; however, the mechanism by which this is achieved remains unknown. Distinct forms of prion disease show differences in lymphoreticular involvement that may be related to the etiology of the disease or to divergent properties of distinct prion strains. For example, the tissue distribution of PrP^{Sc} in vCJD differs strikingly from that in classical CJD with uniform involvement of lymphoreticular tissues. Depending upon the density of lymphoid follicles, PrP^{Sc} concentrations in vCJD peripheral tissues can vary enormously, with levels relative to brain as high as 10% in tonsil or as low as 0.002% in rectum. In contrast, in sporadic CJD, PrP^{Sc} has only been irregularly detected by immunoblotting in noncentral nervous system tissues at very low levels. Tonsil biopsy is used for diagnosis of vCJD, and to date it has shown 100% sensitivity and specificity for diagnosis of vCJD at an early clinical stage, although some patients show scanty deposition of abnormal PrP and a large number of follicles may have to be examined by immunohistochemistry. Anonymous studies of discarded surgical lymphoreticular tissues (tonsil and appendix) have estimated that the overall prevalence of BSE-related prion infection within the UK may be 1 in 2000 of the population. The demonstration of extensive lymphoreticular involvement in the peripheral pathogenesis of vCJD raises concerns that iatrogenic transmission of vCJD prions through medical procedures may be a major public health issue. Disturbingly, cases of transfusion-associated vCJD prion infection have occurred. Prions resist many conventional sterilization procedures and surgical stainless steel-bound prions transmit disease with remarkable efficiency when implanted into mice. While effective methods for prion decontamination of surgical instruments and medical equipment have been reported these have yet to be effectively implemented. Notably, there is now considerable evidence for human transmission of amyloid-beta protein pathology and cerebral amyloid angiopathy resulting from discontinued medical practices involving treatment with human cadaveric pituitary-derived growth hormone or cadaveric dura mater grafting. These findings suggest that it is now important to systematically re-evaluate the potential risks for iatrogenic transmission of proteopathic seeds from individuals with other neurodegenerative diseases.

Prion Strains

A major problem for the “protein-only” hypothesis of prion propagation has been to explain the existence of multiple isolates, or strains, of prions. Prion strains are distinguished by their biological properties: they produce distinct incubation periods and patterns of neuropathological targeting (so-called lesion profiles) in defined inbred mouse lines. As they can be serially propagated in inbred mice with the same *Prnp* genotype, they cannot be encoded by differences in PrP primary structure. Usually, distinct strains of conventional pathogen are explained by differences in their nucleic acid genome. However, in the absence of such a scrapie genome, alternative possibilities must be considered.

Support for the contention that prion strain specificity may be encoded by PrP itself was provided by study of two distinct strains of transmissible mink encephalopathy prions which can be serially propagated in hamsters, designated “hyper” and “drowsy”. These strains can be distinguished by differing physicochemical properties of the accumulated PrP^{Sc} in the brains of affected hamsters. Following limited proteolysis, strain-specific migration patterns of PrP^{Sc} fragments on western blots are seen which relate to different N-terminal ends of PrP^{Sc} following protease treatment implying differing conformations of PrP^{Sc}. Distinct PrP^{Sc} conformations are now recognized to be associated with numerous other prion strains and, similarly, different human PrP^{Sc} isoforms propagate in the brain of patients with phenotypically distinct forms of CJD.

The different fragment sizes seen on Western blots, following treatment with proteinase K, suggests that there are several different human PrP^{Sc} conformations, referred to as molecular strain types. These types can be further classified by the ratio of the three PrP bands seen after protease digestion, corresponding to N-terminally truncated cleavage products generated from di-, mono-, or nonglycosylated PrP^{Sc}. Four types of human PrP^{Sc} have now been commonly identified in sporadic and acquired CJD using molecular strain typing (Fig. 2), although much greater heterogeneity seems likely. Polymorphism at *PRNP* residue 129 dictates the propagation of distinct PrP^{Sc} types in humans. To date, the repertoire of PrP^{Sc} conformations that can be stably propagated by human PrP with 129 methionine or 129 valine remains unknown. In the inherited prion diseases specific pathogenic PrP point mutations dictate the propagation of prion strains that are distinct from those generated from wild-type human PrP.

The identification of strain-specific PrP^{Sc} structural properties facilitates an etiology-based classification of human prion disease by typing of the infectious agent itself. Stratification of human prion disease cases by PrP^{Sc} type enables rapid recognition of any change in relative frequencies of particular PrP^{Sc} subtypes (for example in relation to either BSE exposure patterns or iatrogenic sources of vCJD prions). However, efforts to produce a unified international classification and nomenclature of all human PrP^{Sc} types has been complicated by the fact that the N-terminal conformation of some PrP^{Sc} subtypes seen in sporadic CJD can be altered *in vitro* via changes in metal-ion occupancy or solvent pH. Although agreement is yet to be reached on methodological differences, nomenclature and the biological importance of relatively subtle biochemical differences in PrP^{Sc}, there is strong agreement between laboratories that a significant proportion of the phenotypic diversity in human prion disease relates to the propagation of distinct human prion strains defined by distinct PrP^{Sc} conformations and assembly states. However because host genetic makeup and numerous other factors may also significantly influence prion disease phenotype, it is expected that the actual

number of distinct human prion strains may be considerably less than the number of identified phenotypes. Detailed transgenic modeling therefore remains crucial to establishing how many human prion strains exist.

The hypothesis that alternative conformations or assembly states of PrP provide the molecular substrate for clinicopathological heterogeneity seen in human prion diseases (and that this relates to the existence of multiple human prion strains) has been strongly supported by transmission experiments to conventional and transgenic mice. Transgenic mice expressing only human PrP with either valine or methionine at residue 129 have shown that this polymorphism dictates both the propagation of distinct human PrP^{Sc} conformers and the occurrence of associated patterns of neuropathology and these data provide the molecular basis for *PRNP* codon 129 as a major locus influencing both prion disease susceptibility and phenotype in humans. Biophysical measurements suggest that this powerful effect of residue 129 on prion strain selection is likely to be mediated via its effect on the conformation of PrP^{Sc} or its precursors or on the kinetics of their formation, as it has no measurable effect on the folding, dynamics, or stability of PrP^C. These data are consistent with a conformational selection model of prion transmission barriers and strongly support the “protein only” hypothesis of infectivity by suggesting that prion strain variation is encoded by a combination of PrP conformation and assembly state. Heterozygosity at codon 129 is thought to confer resistance to prion disease by inhibiting homologous protein-protein interactions essential for efficient prion replication while the presence of methionine or valine at residue 129 controls the propagation of distinct human prion strains via conformational selection. Kuru imposed strong balancing selection on the Fore population essentially eliminating *PRNP* 129 homozygotes. Elderly survivors of the kuru epidemic, who had multiple exposures at mortuary feasts, are, in marked contrast to younger unexposed Fore, predominantly *PRNP* 129 heterozygotes. Recently a novel protective PrP variant with a coding change at residue 127 (glycine to valine) has been found to colocalize with kuru exposure. Variants at codons 127 and 129 of *PRNP* demonstrate the population genetic response to an epidemic of prion disease and represent a powerful episode of recent selection in humans. Remarkably, mice expressing only human PrP 127 valine appear to be as resistant to prion disease as PrP null mice and understanding the structural basis of this effect may provide critical insight into the molecular mechanism of mammalian prion propagation.

Neuronal Cell Death in Prion Disease

Although various mechanisms have been proposed to explain neuronal death in prion disease, the precise structure of the infectious agent and the cause of neuronal cell death in prion disease remains unclear. While PrP^C is absolutely required for prion propagation and neurotoxicity, knockout of PrP^C in adult brain and in embryonic models has no overt phenotypic effect, effectively excluding loss of PrP^C function in neurons as a significant mechanism in prion neurodegeneration. Notably, however, there is also considerable evidence that argues that prototypical PrP^{Sc} and indeed prions may not themselves be highly neurotoxic. Instead, it is now hypothesized that the neurotoxic entity may comprise a distinct toxic monomeric or oligomeric PrP species designated PrP^L (for lethal). The steady-state level of PrP^L could determine the rate of neurodegeneration with subclinical prion infection states generating PrP^L at levels below the threshold required for neurotoxicity. Direct support for this hypothesis has been demonstrated by depleting endogenous neuronal PrP^C in mice with established neuroinvasive prion infection. This depletion of PrP^C reverses early spongiform change and progression to clinical prion disease despite the slow accumulation of extraneuronal PrP^{Sc} to levels seen in terminally ill wild-type mice. These data establish that propagation of non-neuronal PrP^{Sc} is not pathogenic, but arresting the misfolding of PrP^C within neurons during scrapie infection prevents prion neurotoxicity. Importantly, this model also validates PrP^C as the key therapeutic target in prion disease. Anti-PrP monoclonal antibodies cure prion-infected cells readily and passive immunotherapy with monoclonal antibodies that bind PrP^C has shown powerful therapeutic effects in several animal models. Based upon this research, and proceeding with extreme caution under very tightly controlled conditions, administration of a humanized anti-PrP monoclonal antibody to a limited number of CJD patients has now been approved. Although PrP^{Sc} has long been considered as a target for chemotherapy, drugs interacting with PrP^{Sc} are likely to be prion strain-specific and may only target a specific subset of PrP^{Sc} conformers resulting in propagation of drug resistant prions.

Future Perspectives

Mammalian prions exist as multiple strains which produce characteristic phenotypes in defined hosts. How this strain diversity is encoded by an apparently protein-only agent remains one of the most interesting and challenging questions in biology, with significant evolutionary implications. The principles of protein conformation-based inheritance established from studying mammalian prions have now been formally demonstrated by elegant studies with analogous systems in yeast models indicating the very wide potential relevance of understanding prion biology. The novel pathogenic mechanisms involved in prion propagation appear highly relevant to other neurological and non-neurological illnesses. Indeed, advances in understanding prion neurodegeneration are already casting considerable light on related mechanisms in other, more common, neurodegenerative diseases such as Alzheimer's and Parkinson's disease. While the protein-only hypothesis of prion propagation is supported by compelling experimental data, and is able to encompass the phenomenon of prion strain diversity, detailed structural studies remain problematic. However, recent advances made with prion isolation together with improved methods in cryo-electron microscopy and high-speed vertically-oriented-probe force microscopy suggests that obtaining a high resolution prion structure will be achievable.

See also: Prions of Yeast and Fungi

Further Reading

- Asante, E.A., Smidak, M., Grimshaw, A., *et al.*, 2015. A naturally occurring variant of the human prion protein completely prevents prion disease. *Nature* 522, 478–481.
- Collinge, J., 2016. Mammalian prions and their wider relevance in neurodegenerative diseases. *Nature* 539, 217–226.
- Collinge, J., Clarke, A.R., 2007. A general model of prion strains and their pathogenicity. *Science* 318, 930–936.
- Gill, O.N., Spencer, Y., Richard-Loendt, A., *et al.*, 2013. Prevalent abnormal prion protein in human appendixes after bovine spongiform encephalopathy epizootic: Large scale survey. *BMJ* 347, f5675.
- Jaunmuktane, Z., Mead, S., Ellis, M., *et al.*, 2015. Evidence for human transmission of amyloid β pathology and cerebral amyloid angiopathy. *Nature* 525, 247–250.
- Mead, S., Gandhi, S., Ayling, H., *et al.*, 2013. A novel prion disease associated with diarrhoea and autonomic neuropathy. *New England Journal of Medicine* 369, 1904–1914.
- Mead, S., Stumpf, M.P., Whitfield, J., *et al.*, 2003. Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. *Science* 300, 640–643.
- Mead, S., Whitfield, J., Poulter, M., *et al.*, 2009. A novel protective prion protein variant that colocalizes with kuru exposure. *New England Journal of Medicine* 361, 2056–2065.
- Prusiner, S.B., 1998. Prions. *Proceedings of the National Academy of Sciences of the United States of America* 95, 13363–13383.
- Purro, S.A., Farrow, M.A., Linehan, J., *et al.*, 2018. Transmission of amyloid- β protein pathology from cadaveric pituitary growth hormone. *Nature* 564, 415–419.
- Sandberg, M.K., Al-Doujaily, H., Sharps, B., *et al.*, 2014. Prion neuropathology follows the accumulation of alternate prion protein isoforms after infective titre has peaked. *Nature Communications* 5, 4347. doi:10.1038/ncomms5347.
- Terry, C., Harniman, R.L., Sells, J., *et al.*, 2019. Structural features distinguishing infectious ex vivo mammalian prions from non-infectious fibrillar assemblies generated in vitro. *Scientific Reports* 9, 376.
- Terry, C., Wenborn, A., Gros, N., *et al.*, 2016. Ex vivo mammalian prions are formed of paired double helical prion protein fibrils. *Open Biology* 6, 160035. doi:10.1098/rsob.160035.
- Wadsworth, J.D., Asante, E.A., Collinge, J., 2010. Review: Contribution of transgenic models to understanding human prion disease. *Neuropathology and Applied Neurobiology* 36, 576–597.
- Wenborn, A., Terry, C., Gros, N., *et al.*, 2015. A novel and rapid method for obtaining high titre intact prion strains from mammalian brain. *Scientific Reports* 5, 10062. doi:10.1038/srep10062.

Pseudorabies Virus (*Herpesviridae*)

Thomas C Mettenleiter and Barbara G Klupp, Friedrich-Loeffler-Institute, Greifswald-Insel Riems, Germany

© 2021 Elsevier Ltd. All rights reserved.

History

The first description of symptoms due to pseudorabies virus (PrV) infection came from infected cattle ('mad itch'), although pigs have later been found to represent the natural hosts of this virus. Hence, Suid Alphaherpesvirus 1 (SuHV-1) is the taxonomic name of PrV. Throughout this article, we will use the term PrV as this is the name most used in the scientific community. Symptoms of PrV infection in swine range from respiratory signs to severe central nervous system disorders. In other susceptible species productive infection is nearly invariably fatal and characterized by severe central nervous symptoms which prompted its designation as pseudorabies due to the rabies-like clinical picture. A typical symptom in species other than pigs is extensive pruritus which resulted in the name 'mad itch' to describe the disease present in cattle in the United States in the first half of the 19th century. In 1902, the Hungarian veterinary pathologist Aladár Aujeszky reported successful isolation of the infectious agent from a diseased ox, a dog and a cat, and differentiated it from rabies. The agent could be passaged in rabbits reproducing the typical symptoms. Guinea pigs and mice were also found to be susceptible, whereas chicken and doves were resistant. Thus, the illness has become widely known as Aujeszky's disease (AD). However, it was not until 1931 that Richard Shope established the identity of the 'mad itch' agent with an infectious agent widely present in domestic pig holdings in the US. In Germany, Erich Traub was the first to cultivate PrV in vitro in porcine organ explants in 1933. One year later, Albert Sabin published his findings of a serological relationship between PrV and herpes simplex virus (HSV) resulting in the inclusion of PrV into the herpesvirus group.

Classification

PrV belongs to the *Alphaherpesvirinae* subfamily of the *Herpesviridae* within the order *Herpesvirales*. Originally based on serological studies and later confirmed by molecular biological analyses and comparison of deduced amino acid sequences of homologous proteins, it was shown to be most closely related to bovine herpesvirus 1 (BoHV-1) and equine herpesvirus 1 (EHV-1), and also to varicella-zoster virus (VZV). This prompted its assignment to the *Varicellovirus* genus within the *Alphaherpesvirinae*. Its current correct taxonomic designation is *Suid Alphaherpesvirus 1* following the binary nomenclature for viruses. Alphaherpesviruses are characterized by a rapid lytic replication, a pronounced neurotropism with establishment of latency in sensory ganglia, and a broad host range. All this applies especially to PrV.

Virion Structure

PrV particles exhibit typical herpesvirion morphology with a diameter of ca. 180 nm. At the centre is the genomic DNA which is enclosed in an icosahedral capsid. The capsid is surrounded by a structure designated as tegument which equals the matrix of RNA viruses but is significantly more complex. Finally, the virion envelope which contains virally encoded glycosylated and non-glycosylated proteins anchored in the lipid bilayer encloses the nucleocapsid and tegument ([Fig. 1](#)).

Genome

The PrV genome consists of a double-stranded linear DNA. Complete genomic sequences have been published which comprise between ca. 140 and 145 kbp encoding a minimum of 70 open reading frames which all exhibit homology to genes in related alphaherpesviruses ([Fig. 2](#)). The genome contains a long (U_L) and short (U_S) unique region with the latter being bracketed by inverted repeats resulting in two possible isomeric forms of the genome with inverted U_S regions relative to the U_L . This arrangement has been designated as a class D herpesvirus genome. So far, three functional origins of replication have been mapped, two in the inverted repeats and one in the middle of the unique long region. Compared to the genomes of other alphaherpesviruses, which generally are largely collinear in their gene arrangement, the PrV genome specifies an inversion of ca. 40 kbp which encompasses genes homologous to the UL27 to UL44 genes of herpes simplex virus ([Fig. 2](#)). Although a similar inversion is also present in the genome of infectious laryngotracheitis virus, a distantly related avian alphaherpesvirus, its biological significance is unclear. A list of the identified PrV genes and function of encoded proteins is shown in [Table 1](#). Wherever possible, PrV genes have been named after their homologs in herpes simplex viruses (HSV). However, the UL3.5 gene of PrV which has homologs in other alphaherpesviruses such as varicella-zoster virus or bovine and equine herpesvirus 1 is absent from the HSV genome. In contrast, PrV does not specify homologs to the UL45, US5, US10, US11 and US12 genes of HSV. Approximately half of the PrV genes are considered as 'nonessential', which indicates that they are dispensable for viral replication, at least in cultured

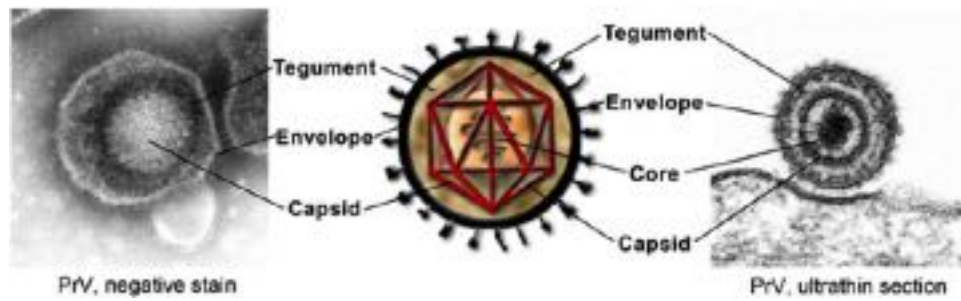


Fig. 1 The PrV virion. A schematic diagram of the PrV virion is presented between an electron micrograph of a negatively stained PrV virion (left) and a thin-sectioned virus particle (right). The location of the virion subcomponents core, capsid, tegument and envelope is indicated. Spikes at the envelope represent viral glycoproteins.

cells. Also, it is estimated that about half of the protein products are located in the virion and can be considered structural components of the virus particle.

Capsid

As a typical herpesviral capsid the icosahedral PrV capsid is composed of 161 capsomers consisting of 955 copies of the major capsid protein, the product of the UL19 gene (see below). Homologs of the HSV pUL18 and pUL38 which form triplexes connecting and stabilizing the capsomers as well as the pUL35 protein located at the tip of the hexons have been identified. Also present are homologs for the pUL6 portal protein which forms a dodecameric channel for package and release of the viral genome at one pentonal site. Intimately associated with the capsid is the capsid vertex specific complex (CVSC) consisting of the UL17 and UL25 gene products.

Tegument

The tegument of herpesviruses is a complex structure which, in the case of PrV, contains in excess of 15 virally encoded proteins. It has become clear that tegument formation is an important step in the morphogenesis of the PrV virion requiring a network of partially redundant protein-protein interactions. The tegument can be divided into capsid-proximal and envelope-proximal parts. The capsid-proximal tegument is comprised of the UL36 gene product, the largest protein in PrV consisting of 3108 amino acids, which physically interacts with the UL37 gene product. From cryoelectron image reconstructions of herpesvirus particles, the UL36 gene product is thought to contact the capsid. On the other hand, the UL46, UL47, UL48 and UL49 gene products are easily stripped from the capsid together with the envelope and, thus, are located in the envelope-proximal tegument. Correlating with these findings, the PrV UL49 gene product has been shown to interact with the intracytoplasmic carboxytermini of the gE and gM envelope proteins. Both parts of the tegument may be connected by the UL48 gene product. Tegument proteins enter the cell after fusion of the virion envelope and the cellular plasma membrane during entry, and prime the cell for virus production. The alphaherpesvirus UL48 gene products are strong transactivators of viral immediate-early gene expression, whereas the UL41 proteins possess endoribonucleolytic activity to degrade preexisting cellular mRNAs. Interestingly, cellular proteins have also been detected in the PrV tegument including cellular actin, annexins and heat shock proteins. Their biological role is unclear.

Envelope

Receptor binding proteins as well as major immunogens are located in the viral envelope. More than 10 envelope constituents have been identified in PrV. Most of them are modified by the addition of carbohydrates and, thus, represent glycoproteins. Several type I, type II and type III PrV membrane proteins have been described (Table 1). Since early nomenclature of PrV glycoproteins was somewhat confusing, it has been agreed to name them after their HSV counterparts. Several of these glycoproteins form complexes such as the homo-oligomeric gB, and the heterodimeric gE/gI, gH/gL, gK/pUL20 and gM/gN. Glycoprotein N and the gM/gN complex which is conserved throughout the *Herpesviridae* have for the first time been identified in PrV as has the gK/pUL20 interaction. Nonglycosylated membrane proteins include the pULs9, pUL43 and pUL56 gene products (see Table 1). The non-structural gG is proteolytically cleaved and released from infected cells.

Replication Cycle

PrV is amongst the most studied animal herpesviruses and has become a major model virus to understand fundamental herpesvirus biology. PrV infection of host cells starts with the interaction of the envelope glycoprotein C (gC) with cell-surface

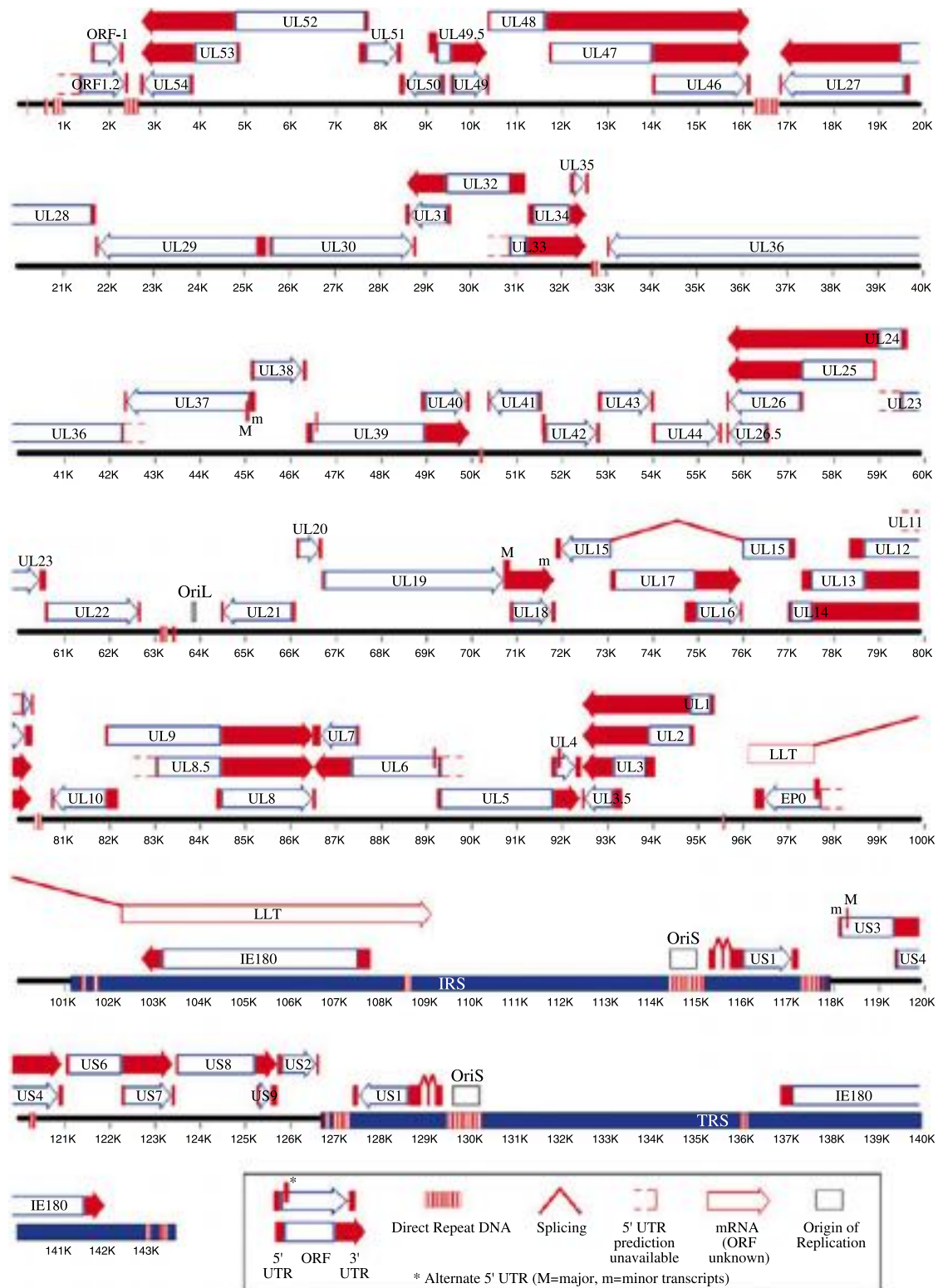


Fig. 2 The PrV genome. The transcript and gene organization as deduced from the complete genomic sequence is shown. The linear form of the PrV genome comprises the unique long sequence, the internal repeat region, the unique short sequence and the terminal repeat. The predicted locations of PrV open reading frames (see Table 1), 5'- and 3' nontranslated regions, DNA repeats, splice sites and origin of replication are shown. Reprinted with permission from the American Society for Microbiology from Klupp, B.G., Hengartner, C., Mettenleiter, T.C., Enquist, L.W., 2004. Journal of Virology 78, 424-440.

Table 1 PrV ORFs

<i>Protein</i>	<i>ORF location^a</i>	<i>Length (aa)</i>	<i>MW (kD)</i>	<i>Alias</i>	<i>Function/property^b</i>	<i>Virion subunit^c</i>
pORF1.2	1370-2254	294	31.0		Membrane protein, pUL56 family; vesicle trafficking	V (E)
pORF1	1631-2254	207	21.8		Membrane protein, pUL56 family; vesicle trafficking	V (E)
pUL54	3873-2788r	361	40.4	ICP27	Early protein; gene regulation	NS
pUL53	4890-3952r	312	33.8	gK	Type III membrane protein; glycoprotein K; complex with pUL20; viral egress	V (E)
pUL52	7733-4845r	962	103.3		Helicase-primase primase subunit; DNA replication	NS
pUL51	7720-8430	236	25.0		Tegument protein; complexed with pUL7; viral egress (secondary envelopment)	V (T)
pUL50	9390-8584r	268	28.6	dUTPase	Deoxyuridine triphosphatase; nucleotide metabolism	NS
pUL49.5	9314-9610	98	10.1	gN	Type I membrane protein; glycoprotein N; complexed with gM	V (E)
pUL49	9648-10,397	249	25.9	VP22	Tegument protein; interacts with C-terminal domains of gE & gM	V (T)
pUL48	10,461-11,702	413	45.1	VP16/αTIF	Tegument protein; gene regulation (trans activator); viral egress (secondary envelopment)	V (T)
pUL47	11,803-14,055	750	80.4	VP13/14	Tegument protein; viral egress (secondary envelopment)	V (T)
pUL46	14,074-16,155	693	75.5	VP11/12	Tegument protein; signaling	V (T)
pUL27	19,673-16,911r	920	100.2	gB	Type I membrane protein; glycoprotein B; viral entry (fusion); cell-cell spread	V (E)
pUL28	21,700-19,544r	718	78.5	ICP18.5	DNA packaging terminase subunit 1; associated with pUL15, pUL33 & pUL6	pC
pUL29	25,383-21,850r	1177	125.1	ICP8	Single-stranded DNA-binding protein; DNA replication	NS
pUL30	25,699-28,845	1048	115.3		DNA polymerase catalytic subunit; DNA replication; complexed with pUL42	NS
pUL31	29,581-28,766r	271	30.4		Tegument protein; primary virion tegument; nuclear egress; interacts with pUL34	pV (T)
pUL32	30,986-29,574r	470	51.6		DNA encapsidation; efficient localization of capsids to replication compartments	V (?)
pUL33	30,985-31,332	115	12.7		DNA Cleavage/encapsidation; associated with pUL28 & pUL15	NS
pUL34	31,491-32,279	262	28.1		Type II nuclear membrane protein; primary virion envelope protein; nuclear egress; interacts with pUL31	pV (E)
pUL35	32,334-32,645	103	11.5	VP26	Capsid protein	V (C)
pUL36	42,472-33,146r	3108	326.7	VP1/2	Tegument protein; interacts with pUL37 & pUL19; capsid transport	V (T)
pUL37	45,269-42,510r	919	98.2		Tegument protein; interacts with pUL36; capsid transport	V (T)
pUL38	45,326-46,432	368	40.0	VP19c	Capsid triplex subunit; complexed with pUL18	V (C)
pUL39	46,763-49,135	790	86.5	RR1	Large subunit of ribonucleotide reductase; nucleotide metabolism	V (?)
pUL40	491,45-50,056	303	34.4	RR2	Small subunit of ribonucleotide reductase; nucleotide metabolism	V (?)
pUL41	51,655-50,558r	365	40.1	VHS	Gene regulation (inhibitor of gene expression); virion host cell shut off factor	V (?)
pUL42	51,784-52,938	384	40.3		DNA replication; DNA polymerase processivity subunit; complexed with pUL30	?
pUL43	52,998-54,119	373	38.1		Type III membrane protein; unknown; inhibits membrane fusion in transient assays	V (E)
pUL44	54,186-55,628	480	51.3	gC	Type I membrane protein; glycoprotein C; viral entry (virion attachment); binds to heparan sulfate	V (E)
pUL26.5	56,695-55,859r	278	28.2	VP22a	Scaffold protein; substrate for pUL26; capsid morphogenesis	pC
pUL26	57,433-55,859r	524	54.6	VP24	Scaffold protein; proteinase; capsid morphogenesis	pC; V (?)
pUL25	59,071-57,467r	534	57.4		Capsid associated protein; complexed with pUL17 (CVSC); DNA packaging and nuclear egress	V (C)
pUL24	59,679-59,164r	171	19.1		Unknown	NS
pUL23	59,672-60,634	320	35.0		Thymidine kinase; nucleotide metabolism	NS
pUL22	60,770-62,830	686	71.9	gH	Type I membrane protein; glycoprotein H; viral entry (fusion); cell-cell spread; complexed with gL	V (E)
pUL21	66,226-64,649r	525	55.2		Capsid associated protein; complexed with pUL16	V (T)?
pUL20	66,333-66,818	161	16.7		Type III membrane protein; viral egress; complexed with gK	V (E)
pUL19	66,905-70,897	1330	146.0	VP5	Major capsid protein; forms hexons und pentons	V (C)
pUL18	71,063-71,953	296	31.7	VP23	Capsid triplex subunit 2; complexed with pUL38	V (C)
pUL15 (Ex2)	73,285-72,149r	735	79.1		DNA packaging terminase subunit; DNA cleavage/encapsidation; interacts with pUL33, pUL28 & pUL6	pC

(Continued)

Table 1 Continued

Protein	ORF location ^a	Length (aa)	MW (kD)	Alias	Function/property ^b	Virion subunit ^c
pUL15 (Ex1)	77,235-76,165r					
pUL17	73,336-75,129	597	64.2		DNA-packaging tegument protein; DNA cleavage/encapsidation; complexed with pUL25 (CVSC)	V (C)
pUL16	75,156-76,142	328	34.8		Tegument protein; complexed with pUL11 & pUL21	V (T)?
pUL14	77,234-77,713	159	17.9		Unknown	NS
pUL13	77,683-78,879	398	41.1	VP18.8	Serine/threonine protein kinase	?
pUL12	78,845-80,296	483	51.3		Deoxyribonuclease; DNA processing	NS
pUL11	80,254-80,445	63	7.0		Myristylated tegument protein; viral egress (secondary envelopment); complexed with pUL16&pUL21	V (T)
pUL10	82,105-80,924r	393	41.5	gM	Type III membrane protein; glycoprotein M; viral egress (secondary envelopment); C-terminus interacts with pUL49; complexed with gN; inhibits membrane fusion in transient assays	V (E)
pUL9	82,104-84,635	843	90.5	OBP	DNA replication origin-binding helicase; DNA replication	NS
pUL8	84,632-86,683	683	71.2		Helicase-primase subunit; DNA replication; part of pUL5/pUL8/pUL52 complex	?
pUL7	87,649-86,849r	266	29.0		Tegument protein; virion formation and egress; complexed with pUL51	V (T)
pUL6	89,471-87,540r	643	70.3		Capsid portal protein; DNA packaging; docking site for terminase	V (C)
pUL5	89,470-91,974	834	92.1		Helicase-primase helicase subunit; DNA replication; part of pUL5/pUL8/pUL52 complex	NS
pUL4	92,033-92,470	145	15.8		Nuclear protein; unknown	NS
pUL3.5	93,320-92,646r	224	24.0	V57	Tegument protein; viral egress (secondary envelopment)	V (T)
pUL3	94,030-93,317r	237	25.6		Nuclear protein; unknown	NS
pUL2	95,036-94,086r	316	33.0	UNG	Uracil-DNA glycosylase; DNA repair	NS
pUL1	95,484-95,014r	156	16.5	gL	Glycoprotein L; membrane anchored via complex with gH; viral entry (fusion); cell-cell spread	V (E)
EP0	97,884-96,652r	410	43.8	ICP0	Gene regulation (trans activator of viral and cellular genes); early protein	V (?)
IE180 (IRS)	107,695-103,343r	1450	148.9	ICP4	Gene regulation; immediate-early protein	V (?)
IE180 (TRS)	137,044-141,396					
US1 (IRS)	116,181-117,275	364	39.6	RSp40/ ICP22	Immediate-early protein; gene regulation	V (?)
US1 (TRS)	128,558-127,464r					
US3 (minor)	118,239-119,405	388	42.9	PK	Minor form of the serine/threonine protein kinase	NS
US3 (major)	118,401-119,405	334	36.9	PK	Major form of the serine/threonine protein kinase; viral egress (nuclear egress)	V (T)
US4	119,465-120,964	499	53.7	gG	Type I membrane protein; glycoprotein G	secreted
US6	121,145-122,347	400	44.3	gD	Type I membrane protein; glycoprotein D; viral entry (receptor binding protein)	V (E)
US7	122,368-123,465	365	38.7	gL	Type I membrane protein; glycoprotein L; cell-cell spread; complexed with gE	V (E)
US8	123,569-125,302	577	62.4	gE	Type I membrane protein; glycoprotein E; cell-cell spread; complexed with gL; C-terminus interacts with pUL49	V (E)
US9	125,360-125,656	98	10.6	11K	Type II membrane protein; axonal transport	V (E)
US2	125,879-126,649	256	27.7	28K	Prenylated tegument protein; unknown	V (T)

^aNumbering starts at +1 on the U_L end of the genome. r indicates ORF encoded on reverse strand; AccNo. JQ809328.1.

^bFunction/Property as demonstrated for the PrV and/or HSV-1 homolog.

^cV (C): virion capsid component; V (T): virion tegument component; V (E): virion envelope component; V (?): virion component of unknown subviral localization; pV: primary enveloped virion precursor component (not found in mature virion); NS: non structural protein; pC: present in intranuclear capsid precursor forms but not found in mature virion;?: unknown.

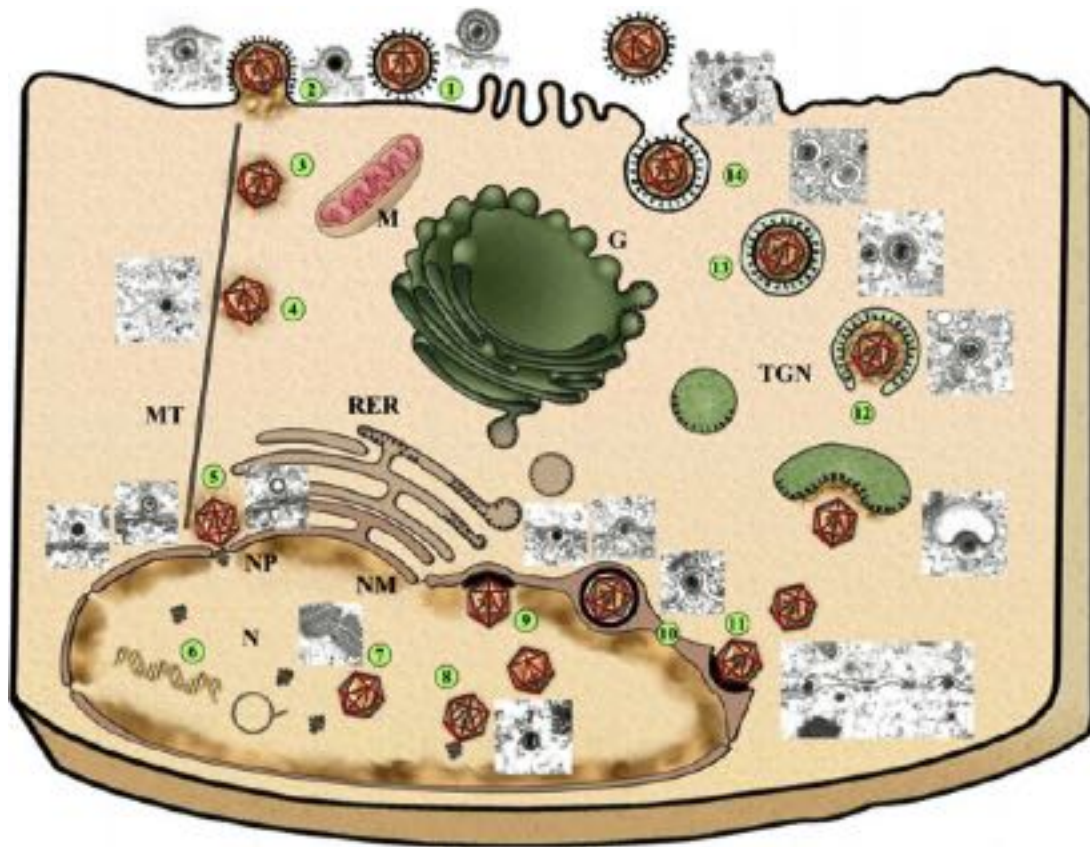


Fig. 3 The PrV replication cycle. A diagram of the replication cycle of PrV is shown together with electron micrographs showing the respective stages. After attachment (1) and penetration (2), capsids are transported to the nucleus (3) via interaction with microtubuli (4) docking at the nuclear pore (5) where the viral genome is released into the nucleus. Here, transcription of viral genes and viral genome replication occurs (6). Concatemeric replicated viral genomes are cleaved to unit-length molecules during encapsidation (7) into preformed capsids (8) which then leave the nucleus by budding at the inner nuclear membrane (9) followed by fusion of the envelope of these primary virions located in the perinuclear cleft (10) with the outer nuclear membrane (11). Final maturation then occurs in the cytoplasm by secondary envelopment of intracytosolic capsids via budding into trans-Golgi derived vesicles (12) containing viral glycoproteins (black spikes) resulting in an enveloped virion within a cellular vesicle. After transport to the cell surface (13), vesicle and plasma membranes fuse releasing a mature enveloped PrV particle from the cell (14).

heparansulfate-containing proteoglycans (Fig. 3). This interaction is beneficial but not essential for the second step which involves binding of the essential envelope gD to one of its cellular receptors, e.g., nectin. Interestingly, HSV and bovine herpesvirus 1 also use heparansulfate and nectin for attachment. However, gD-negative and even gC- and gD-negative infectious PrV mutants have been isolated which indicates that infection can also occur by other routes. Interestingly, these mutants harbour mutations in gB and gH. Penetration, i.e., fusion of viral envelope and cellular plasma membrane, requires the essential gB and gH/gL which represent glycoproteins conserved throughout the *Herpesviridae* indicating a common herpesvirus membrane fusion mechanism. The capsid with associated inner tegument is then transported to the nuclear pore via microtubules where it docks and releases the viral genome through one vertex via the nuclear pore into the nucleus of the cell. Empty capsids may remain bound to the nuclear pore for a considerable time. This cascade of events can be alleviated by transfection of naked genomic viral DNA. In the nucleus, transcription of viral genes is initiated by expression of the major immediate-early (IE) gene resulting in the translation of a 180 kDa immediate-early protein (IE180). Although IE180 has long been considered the only immediate-early protein of PrV, studies using specific inhibitors have identified that the US1 (RSp40) mRNA is also expressed with immediate-early kinetics. Like other herpesvirus IE proteins, PrV IE180 is a potent transcriptional activator that induces the expression of viral early (E) genes. These encompass genes encoding enzymes involved in nucleotide metabolism (e.g., UL23 = thymidine kinase; UL2 = Uracil-DNA-glycosylase; UL39/40 = Ribonucleotide-Reductase; UL50 = dUTPase), DNA replication (UL30/UL42 = DNA-polymerase and associated factor; UL5 = helicase; UL8/UL52 = primase; UL29 = ssDNA-bdg. protein; UL9 = origin-binding protein) as well as two protein kinases (pUS3, pUL13). DNA replication, whether exclusively via rolling circle or involving intra- and inter-molecular recombination and branching, results in the formation of head-to-tail fused concatemers of the genome. Finally, late genes encoding primarily virion structural proteins are expressed. After translation in the cytosol, capsid proteins are transported into the nucleus for capsid assembly and DNA packaging. Capsid assembly is morphologically similar in all herpesviruses: capsids containing the major capsid protein pUL19, triplex proteins pUL18 and pUL38, hexon-tip protein pUL35, and portal protein

pUL6 assemble autocatalytically with a protein scaffold consisting of the UL26.5 gene products. Packaging occurs via the unique portal at one vertex comprising 12 molecules of the pUL6 protein. Genome-length molecules are cleaved from concatemeric replication products during packaging which requires the terminase complex (pUL15, UL28, pUL33) as well as pUL32. The CVSC of pUL17 and pUL25 stabilizes the capsid and is essential for triggering primary envelopment. It is located at the outside of the capsid. Egress of herpesvirus capsids from the nucleus occurs by budding at the inner leaflet of the nuclear membrane followed by de-envelopment (fusion) at the outer leaflet. For primary envelopment, the conserved heterodimeric nuclear egress complex (NEC) consisting of pUL31 and pUL34 proteins has been shown to be crucial in all three subfamilies of herpesviruses. Crystal structures for the HSV-1, PrV and human cytomegalovirus (HCMV) NEC show a conserved arrangement with extensive interfaces between the complex partners. The NEC which is localized in the nuclear membrane recruits cellular protein kinase C which phosphorylates and, thereby, dissociates nuclear lamins allowing access of nascent capsids to the inner nuclear membrane. Primary enveloped virions in the perinuclear space also contain the NEC which constitutes part of the primary envelope (pUL34 is a type II transmembrane protein) and tegument (pUL31). The pUS3 protein kinase is also part of the primary PrV virion. pUS3 has been detected in primary and mature PrV virions, whereas pUL31 and pUL34 proteins are absent from mature virus particles. The mechanisms of de-envelopment are unclear. However, in the absence of the nonconserved and nonessential pUS3, primary enveloped virions accumulate in the perinuclear space demonstrating the participation of this protein in nuclear egress. Within the cytoplasm, virion morphogenesis is completed by tegumentation, final envelopment and transport of mature virus particles for release at the plasma membrane. Tegumentation apparently starts at two sites, the capsid and the future envelopment site. At the capsid, the conserved tegument protein pUL36, the largest gene product of PrV with 3108 amino acids, interacts with pUL37. At the future envelopment site, i.e., at vesicles derived from the Golgi apparatus, the carboxytermini of the type I membrane protein gE and the type III gM bind the pUL49 tegument protein. Presumably, pUL48 links the two parts of the tegument which drives budding of tegumented virions into these trans-Golgi vesicles containing viral glycoproteins to yield mature virions within a vesicle. PrV gM has been shown to relocate other viral and cellular proteins to the trans-Golgi and, thus, may be involved in assembling the envelope proteins. On the other hand, the conserved pUL11 is thought to be involved in directing tegument proteins to the envelopment site. Finally, virion-containing vesicles move to the cell surface, a process in which pUL20 is involved, where plasma and vesicle membranes fuse resulting in release of infectious particles. Apparently, the gK/pUL20 complex inhibits an immediate re-fusion of released virions with the cell they just left.

Besides infection by free virions, infectivity can also be transmitted via direct cell-cell transmission. Although several virion proteins required for penetration (gB, gH/gL) are also required for direct cell-cell spread, its mechanism remains enigmatic. Interestingly, direct cell-cell spread of PrV does not require the presence of the receptor-binding gD nor does syncytium formation after transient expression of gB and gH/L which contrasts the situation in other alphaherpesviruses such as HSV.

Clinical Features and Pathogenesis

PrV is able to productively infect most mammals with the exception of higher primates including humans. However, primate and human cells are infectable in cell culture and the reason for the natural resistance is not clear. Equids and goats are also rather resistant but may be infected experimentally. In addition, pseudorabies has been reported in many species of wild mammals, including wild boar, feral pigs, coyotes, raccoons, rats, mice, rabbits, deer, badgers, and coatimundi. It is so far unknown whether these animals play a role in farm-to-farm transmission of PrV. In susceptible species other than porcines, infection is invariably fatal and animals die from severe neuronal disorders.

After PrV infection of the natural host, the clinical outcome varies depending on the age of the animal, the virulence of the virus and the route of infection. In nature infection occurs predominantly oro-nasally, although genital transmission may also take place especially in feral pigs. After replication in epithelial cells the virus gains access to neurons innervating the facial and oropharyngeal area, in particular the olfactory, trigeminal and glossopharyngeal nerves. By fast axonal retrograde transport the virus reaches the cell bodies of infected neurons where either lytic or latent infection ensues (see below). By a short viremia PrV may be disseminated to other organs where the virus replicates in epithelia, vascular endothelium, lymphocytes, and macrophages.

Neonate piglets become prostrate and die quickly, often without nervous signs. In slightly older piglets, severe CNS disorders are characterized by incoordination, twitching, paddling, tremors, ataxia, convulsions and/or paralysis whereas itching is only rarely present (Fig. 4). Mortality in piglets until 2–3 weeks of age may be as high as 100% resulting in severe losses. Piglets between 3 and 6 weeks may still exhibit neurological signs and a high morbidity, but mortality is usually reduced. Infection in older pigs induces primarily respiratory symptoms like coughing, sneezing and heavy breathing resulting from viral replication in and destruction of pulmonary epithelium. Despite the absence of overt nervous signs, virus gains access to neurons and remains latently established in the olfactory bulb, trigeminal ganglia, and brain stem or, after venereal transmission, in the sacral ganglia. PrV infection of pregnant sows may result in abortion, or the delivery of stillborn or mummified fetuses due to endometritis and necrotizing placentitis with infection of trophoblasts.

In non-porcine species, PrV infection is strictly neurotropic and invariably fatal, sometimes after a rapid, peracute course without any preceding overt clinical signs. Pruritus is a lead symptom of PrV infection in these species which, in particular in rabbits and rodents, may result in violent itching and automutilation. Often the death of mice, rats, cats or dogs on farms is a warning sign for the presence of PrV prior to symptoms in pigs.



Fig. 4 Neurological symptoms of PrV infection in piglets. The animals show ataxia (A), convulsions and paralysis (B) which ultimately lead to death.

Transmission occurs via virus-containing body fluids like nasal and genital secretions which gain access to unprotected epithelial surfaces within the respiratory or genital tract. Airborne transmission is efficient only at short range. Carnivores become infected by ingesting contaminated meat. After primary replication in epithelial cells the virus enters the endings of sympathetic, parasympathetic or sensory and motor neurons innervating the area of primary replication. Infection occurs most likely by the same mechanism as outlined above for cultured cells. De-enveloped and partially de-tegmented virus particles are retrogradely transported to the neuronal cell body, where DNA replication and formation of progeny virions ensues. Primarily complete enveloped virions within transport vesicles are then translocated to the synapse for transsynaptic transfer. Depending on the virulence of the virus and the age and immune status of the host, infection may not proceed beyond the first neuronal level, i.e., ganglia directly innervating the affected peripheral site. However, virus may also spread to the brain resulting in ganglioneuritis and encephalitis. Lymphocytes can also become infected by PrV and infection of these cells may help viral spread within the body playing an important role in infection of the fetus. However, the percentage of infected cells in the blood is rather low, even during acute infection, and difficult to detect. A major target organ for latency in swine are tonsils, and tonsil biopsies allow reliable detection of virus either by molecular biological techniques or virus isolation.

There are no specific gross lesions of Aujeszky's disease. Only in piglets there may be necrotizing tonsillitis, rhinotracheitis or proximal esophagitis. Other lesions commonly seen include pulmonary edema, necrotizing enteritis, and multifocal necrosis of the spleen, lung, liver, lymph nodes, and adrenal glands. Histologically, PrV causes a nonsuppurative meningoencephalitis and paravertebral ganglioneuritis. The grey matter is especially affected, and infected neurons or astrocytes may present acidophilic intranuclear inclusions. The presence of viral antigen can be visualized by immunostaining and viral genomes can be detected by in situ hybridization. PrV infected cells usually show more or less extensive degeneration and necrosis due to the lytic viral replication. Whether apoptosis induced by PrV infection also plays a role in vivo is unclear. Mainly a T-cell reaction of the immune system induces ganglioneuritis, polio- or panencephalitis with foci of gliosis contributing to the loss of neuronal function. The described extraneural lesions in pigs or acute myocarditis in carnivores might provide additional explanations for the fatal outcome of infections in which virus can not be recovered from the brain.

Immunity

Live as well as inactivated vaccines induce efficient protective immunity against AD. Antibodies against a number of viral structural and nonstructural proteins have been detected in infected animals and virus-neutralizing murine monoclonal antibodies have been isolated. Antibody responses are primarily directed against the major surface glycoproteins including gB, gC, gD and gE as well as the secreted gG. The most potent complement-independent virus neutralizing antibodies are directed against gC, gD and, to a lesser extent, gB, and subunit vaccines consisting of gB, gC, gD, as well as anti-idiotypic anti-gD antibodies and heterologous vectors expressing gC or gD elicit protective immunity. In contrast, anti-gE antibodies require complement for neutralization, and anti-gG antibodies have no neutralizing ability at all. In diagnostic assays, antibodies against whole virus or specific for gB are being used to serologically detect PrV infection. Major targets for cell-mediated immunity in pigs are primarily gC and to a lesser extent gB.

Although the numerous elaborate immune evasion mechanisms of beta- and gammaherpesviruses including expression of virokines and viroceptors have not been found in alphaherpesviruses, these viruses still interact with the immune system to evade its activity. Like other alphaherpesvirus gC proteins, PrV gC binds species-specifically to porcine complement component C3, and

the gE/gI complex binds the Fc portion of porcine IgG. The secreted gG may bind chemokines, thereby impairing intercellular signaling. Moreover, infection of cells by PrV results in downregulation of MHC-I antigen presentation, and envelope glycoproteins present at the plasma membrane of infected cells are internalized by as yet unknown factors resulting in a paucity of antigens presented to the immune system at the cell surface. PrV gN inactivates the transporter which translocates processed peptides for loading onto MHC-I molecules into the endoplasmic reticulum. The combined action of these mechanisms may give the virus an edge over the immune system facilitating establishment of latency and further virus spread.

Latency

Like other herpesviruses, PrV has the capacity to become latent in neurons. During latency, the genome persists largely quiescent in a presumably circular form. Expression is restricted to one region, the LAT gene, which encompasses part of the inverted repeat and adjoining U_L region. It is located antiparallel to the gene for the major immediate-early protein IE 180 and the EP0 gene (Fig. 2). The LAT gene is transcribed into three different RNAs of 8.4, 8.0 and 2.0 kb. The 8.4 and 2.0 kb species are derived from a larger precursor by splicing. During latency, only the 8.4 kb RNA is produced from a separate promotor which, apparently, is active only under latent conditions, whereas the 8.0 and 2.0 species are transcribed during lytic infection. Detailed RNA analyses revealed the existence of eleven micro-RNAs derived from the LAT which play a role in latency and virulence of PrV. Like other herpesviruses, PrV encodes proteins which are able to suppress apoptotic cell death which is a prerequisite for the establishment of latency. The PrV Us3 protein kinase has been demonstrated to mediate this function in porcine fetal trigeminal neurons. Like other alphaherpesviruses, PrV establishes latent infections predominantly in neuronal tissues such as the trigeminal or sacral ganglia. However, tonsils have also been identified as sites of PrV latency.

Epidemiology and Prevention

In the 20th century PrV has become a pathogen distributed worldwide with the exception of Australia, Canada and the Scandinavian countries. In particular in major swine producing areas, PrV infection caused significant economic losses amounting to hundreds of million of dollars making it one of the most devastating pig diseases. Control and eradication of PrV infection in pigs relied on two strategies. In areas with a low prevalence of infection, serological identification and consequent elimination of seropositive animals resulted in the successful eradication of AD e.g., from the United Kingdom, Denmark and East Germany. PrV infection can be diagnosed by detecting either viral nucleic acid using polymerase chain reaction (PCR) or the infectious agent (antigen detection by immunofluorescence or virus isolation). PCR is also suitable for detecting latent viral genomes. A PrV specific immune response in live animals can be confirmed using various serological assays such as virus neutralisation test, latex agglutination test and ELISA systems based on either complete virus particles or distinct viral antigens, e.g., gB. To reduce disease in areas of high infection prevalence, vaccination using either live-attenuated or inactivated vaccines has also been used. However, vaccination does not result in sterile immunity, and vaccinated animals may still be infected with and carry the virus but these carriers were no longer identifiable by serological analysis. This problem has been solved by the advent of so-called 'marker' vaccines. This concept provided a break-through in animal disease control, and serves as a blueprint for control of other infectious diseases. It was based on the finding that several immunogenic envelope glycoproteins of PrV such as gC, gE and gG (see above) are not required for productive replication and, thus, can be deleted from the viral genome without abolishing virus replication. These gene-deleted strains can be produced easily in conventional cell systems and can be administered as inactivated as well as

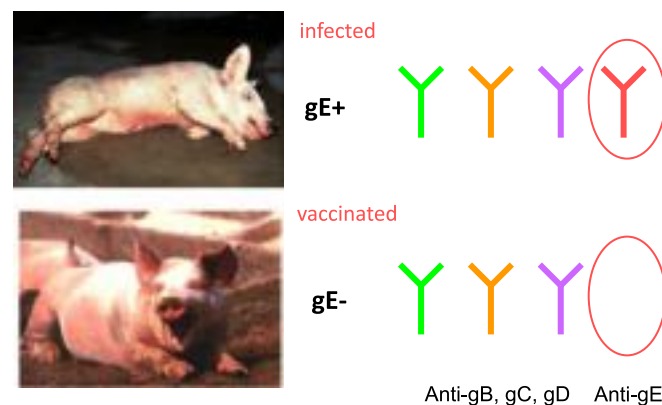


Fig. 5 The principle of DIVA or marker vaccination. Whereas after wild-type infection antibodies are produced against all immunogenic viral proteins, after vaccination with a gene-deleted virus antibodies against the missing gene product (circled) will not be formed. Thus, presence or absence of these antibodies is used to differentiate between infected and vaccinated animals.

modified-live vaccines. In fact, gene-deleted PrV strains were the first genetically engineered live-virus vaccines ever licensed. Thus, PrV has pioneered modern vaccinology. Whereas animals vaccinated with these vaccines do not mount an immune response to the missing gene product, wild-type virus infection invariably results in seroconversion for the differentiating antigen. Serological assays (ELISA) have subsequently been developed that allow easy and sensitive detection of antibodies against these marker proteins resulting in the unambiguous identification of wild-type virus infected animals despite vaccination (Fig. 5). Thus, virus circulation can be reduced by vaccination, and, subsequently, infected animals which still harbour field virus can be identified and eliminated resulting in a cost-efficient eradication. This 'differentiating infected from vaccinated animals' (DIVA) approach used first in the eradication of PrV, has been a breakthrough in animal disease control and is now widely accepted and practiced on other infectious diseases such as bovine herpesvirus 1 infection, classical swine fever and foot-and-mouth disease. Its application resulted in the eradication of PrV from heavily infected West Germany within ten years, and also recently succeeded in eliminating PrV infection from USA and New Zealand pig herds. Although European wild boar and American feral pigs also harbour PrV, there is no epidemiological link between PrV in wild boar and domestic pigs in Europe since clearly different viral strains have been isolated from either. However, infected feral pigs may represent a source of infection of domestic pig holdings in the Southern part of the USA. Recently, a surge in PrV infections has been noticed in China despite extensive vaccination. It has been attributed to the emergence of hypervirulent PrV strains, although the molecular basis for this increased virulence remains unclear.

Pseudorabies Virus as a Tool in Neurobiology

Like other alphaherpesviruses, PrV exhibits a distinct neurotropism invading the CNS via peripheral nerves. While wild-type strains of PrV may spread within the CNS both, laterally and transsynaptically, attenuated PrV mutants have been identified which, under proper assay conditions, travel more or less exclusively along nerves and are transported transsynaptically. This particular property has prompted an increasing use of PrV as a suitable transneuronal tracer to label neuronal connections in experimental animal models being useful in mice and rats for elucidating detailed neuroanatomical networks. The virus used most in these studies is the Bartha strain of PrV, a modified-live vaccine strain which had been attenuated by the Hungarian veterinarian Adorjan Bartha by multiple passages in embryonated chicken eggs and chicken embryo fibroblasts. Molecular biological analyses demonstrated that this strain carries several lesions compared to wild-type PrV: it lacks the gE, gI and Us9 genes, contains a mutation in the signal sequence for gC, specifies attenuating mutations in the UL21 gene, and expresses a UL10 gene product (gM) which is not glycosylated due to mutation of the N-glycosylation site. The glycoprotein deletions and the UL21 mutation have been shown to be most important for the observed attenuation. Recently, genetically engineered Bartha-derivatives expressing a range of chromogenic or fluorescent marker proteins have been used in tracing studies. Moreover, elegant mutants that express their markers only under specific conditions, e.g., in transgenic cells or animals expressing cre-lox recombinase under control of tissue-specific promoters have added another possibility of tissue-specific labelling.

Future Perspectives

PrV is a most fascinating virus with several interesting properties. The availability of conventional and genetically engineered marker vaccines allows effective and cost-efficient disease control campaigns which have been shown to result in the eradication of virus and disease from animal populations. Although PrV infection is still widespread, in particular in certain areas in Eastern Europe and Asia, concerted efforts could result in the elimination of the disease at the global level. Beyond its importance as the causative agent of a relevant animal disease, it is an ideal tool to study basic mechanisms of herpesvirus biology with the enormous advantage of an experimentally accessible natural virus-host system by infection of pigs. Moreover, its broad host range allows the use of other well-defined animal models for neuroanatomical, immunological and molecular biological studies. Since PrV grows exceedingly well in tissue culture, it is also well suited for detailed analysis of the requirements for replication of other (alpha) herpesviruses. Thus, PrV will remain under intensive scrutiny.

Acknowledgments

We thank Thomas Müller and Jens Teifke for help with the manuscript and Harald Granzow and Mandy Jörn for Figs. 1 and 3.

Further Reading

- Freuling, C.M., Müller, T.F., Mettenleiter, T.C., 2018. *Veterinary Microbiology* 206, 3–9.
 Klupp, B.G., Hengartner, C., Mettenleiter, T.C., Enquist, L.W., 2004. *Journal of Virology* 78, 424–440.
 Mettenleiter, T.C., 2016. *Viruses* 8, 266. doi:10.3390/v8100266.
 Mettenleiter, T.C., Ehlers, B., Müller, T., Yoon, K.-J., Teifke, P.T., 2012. In: Zimmerman, J.J., Karriker, L.A., Ramirez, A., Schwartz, K.M., Stevenson, G.W. (Eds.), *Diseases of Swine*, tenth ed. John Wiley & Sons, pp. 421–446.
 Pomeranz, L., Reynolds, A.E., Hengartner, C.J., 2006. *Microbiology and Molecular Biology Reviews* 69, 462–500.

Rabbit Hemorrhagic Disease Virus and European Brown Hare Syndrome Virus (*Caliciviridae*)

Lorenzo Capucci, Patrizia Cavadini, and Antonio Lavazza, The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

© 2021 Elsevier Ltd. All rights reserved.

Classification

Rabbit hemorrhagic disease virus (RHDV) and European brown hare syndrome virus (EBHSV) belong to the *Lagovirus* genus within the *Caliciviridae* family. RHDV and EBHSV are among the most pathogenic and contagious animal viruses, causing an acute and almost always lethal hepatitis in their main hosts, respectively, RHD in the European rabbit (*Oryctolagus cuniculus*) and EBHS in brown hare (*Lepus europaeus*). These viruses cause disease only in adults while animals under 7–8 weeks of age are resistant, as the infection is subclinical. In 2010 a third lagovirus emerged, which is related to RHDV but shows significantly different characteristics, hence the proposed name of RHDV type 2 (RHDV2). Namely, RHDV2 (1) infects and causes RHD in rabbits but also an EBHS-like disease in hares, with a degree of susceptibility that is species dependent; (2) causes RHD also in very young rabbits (2–3 weeks of age onwards); and (3) has a specific antigenic profile that allows to overcome RHDV-induced immunity. Also, during its spread in Europe, RHDV2 became more virulent, approaching that of RHDV.

The development and application of serological ELISA for RHD on farmed and wild rabbits led to suspect that under the tip of the viral iceberg (i.e., pathogenic lagoviruses) there were nonpathogenic RHDV-related viruses. Since mid-1990s the employment of genomic techniques allowed the discovery of two main lineages of nonpathogenic lagoviruses, the so-called rabbit caliciviruses (RCVs): RCV-E identified in Europe, and RCV A-1 first identified in Australia but present also in rabbit populations in Europe. RCVs cause subclinical infections, have a tissue tropism for intestinal epithelia cells and behave as enteric viruses. Similarly, a hare nonpathogenic lagovirus, called hare calicivirus (HaCV), was recently discovered in brown hares.

Virion Structure

RHDV and EBHSV are spherical particles with an outer diameter of 40 nm, whose structure is defined by characteristic cup-shaped depressions (Fig. 1). The 180 VP60 subunits that comprise the capsid of lagoviruses are organized as 90 dimers, each appearing as an arch-like capsomer. Particles with a different morphology are detectable in liver and spleen of rabbits that died as a result of chronic RHD (see “Clinical features”). This is a smooth particle of around 28–30 nm without the cup-shaped depression and relative protrusions, corresponding to the S domain only.

The RHDV capsid is resistant to low pH but it is disassembled to subunits at pH > 10. The assembled VP60 is also resistant to trypsin.

RHDV VP60 is assembled into three domains, the N-terminal arm, aa 1–65 (NTA), the shell, aa 66–229 (S), the protrusion, aa 238–579 (P), and a short hinge (aa 230–237) that connects S and P. The S domain shares high sequence and structural homology with those of other caliciviruses.

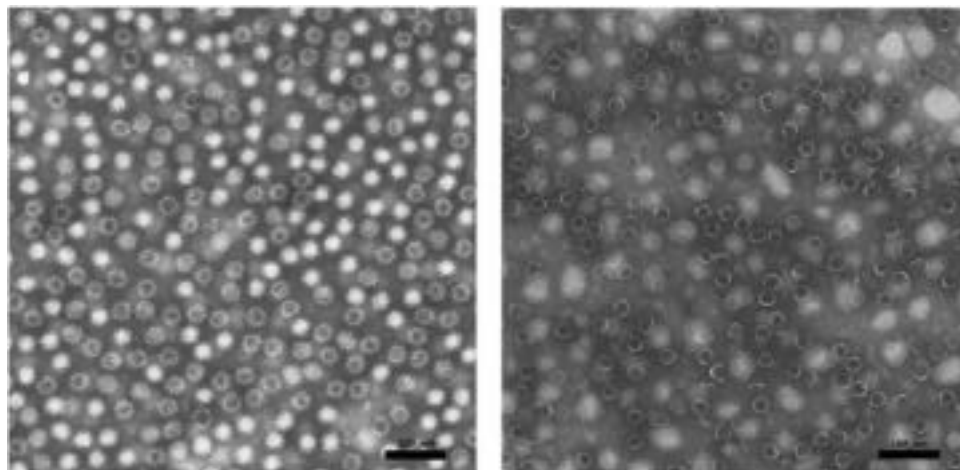


Fig. 1 Electron micrographs of purified RHDV particles. Micrographs show the typical calicivirus morphology (left) and smooth particles (right), constituted only by the S domain of VP60, purified from the liver of a rabbit died due to chronic RHD (right). Negative staining (NaPT 2%)
Bar = 100 nm.

The P domain consists of subdomains P1 surmounted by the P2. The external part of P2 is formed by seven loops of various lengths. Indeed, the most exposed surface loop L1 (aa 300–314), which exhibits high sequence variations among RHDV isolates, stimulates production of antibodies that can protect rabbits from RHD. X-ray crystallography studies showed that loop 7 contains the key residues of the binding pocket of the histo-blood group antigens, considered anchors/receptors on lagomorph epithelial cells for all lagoviruses (see below). Part of these residues on loop 7 interact also with the protein nuclein, a nuclear protein present also on the surface of cell membranes, indicated as cell receptor for RHDV2.

Genome

The genomes of Lagoviruses are composed of single positive stranded polyadenylated RNA ~7.4 kb long (Fig. 2). An additional RNA species of 2.2 kb, collinear with the 3' end of the gRNA and designated subgenomic RNA (sgRNA), is also produced during replication. sgRNA usually contributes to the production of high levels of structural proteins required during the intermediate and late stages of infection (e.g., VP60 and VP10).

The genomic structure is composed of two slightly overlapping open reading frames (ORFs): ORF1 encodes for a large polyprotein that is cleaved by a viral protease into seven nonstructural (NS) proteins and the major capsid protein VP60; ORF2 produces a minor structural protein, VP10, that might regulate virus replication and virion release from infected host cells. In addition, both genomic RNA and sgRNA are polyadenylated at the 3' end and covalently linked to a small virus-encoded protein (VPg) at the 5', which is essential for the translation of viral RNA.

The genetic similarity between RHDV and EBHSV is approximately 70% for the VP60 protein gene but it increases up to 80% at the amino acid level. Based on phylogenetic analysis performed with the VP60 coding sequences, the lagoviruses could be divided in two genogroups (GI and GII) that can be subdivided into six genotypes (GI.1, GI.2, GI.3, GI.4, GII.1, and GII.2), further subdivided into variants (four variants for GI.1 and three variants for GII.1) (Fig. 3).

Since the identification in 2010 of RHDV2 (GI.2), several complete genome sequences of the circulating strains were obtained and phylogenetic analysis showed frequent recombination events. In particular, two types of recombinant strains were detected in Europe and Australia, both with the structural proteins VP60 and VP10 originating from strains GI.2, and different NS proteins

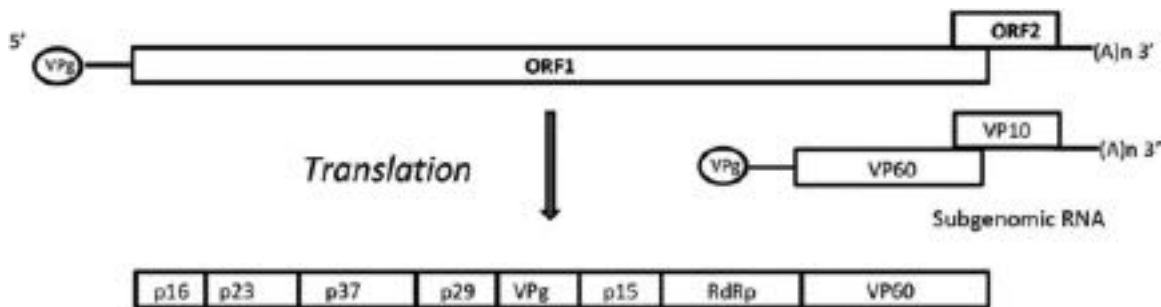


Fig. 2 Genome organization of lagoviruses. The genome has a VPg protein attached to the 5' end of the gRNA and sgRNA and a 3' poly A tail. The positive sense RNA genome is organized into two ORFs. The ORF1 polyprotein undergoes autocatalytic processing due to the activity of the 3C-like protease.

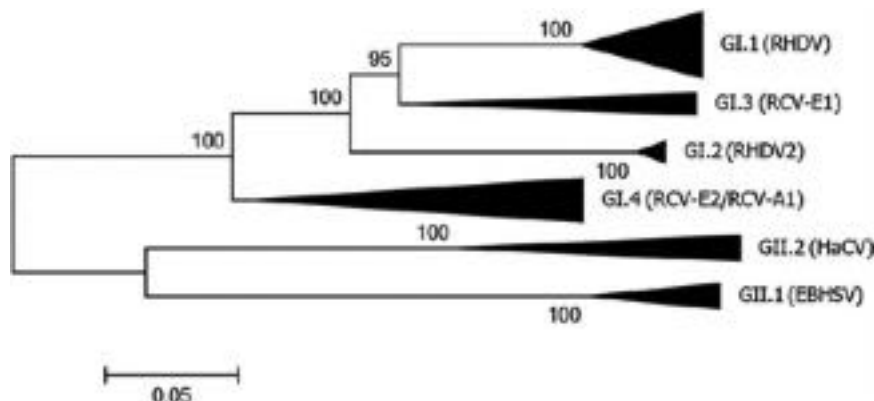


Fig. 3 Genetic relatedness of lagoviruses. Phylogenetic tree of lagoviruses including both pathogenic (RHDV, RHDV2, and EBHSV) and nonpathogenic viruses (RCV E1, RCV E2, RCV A1, and HaCV). The coding sequences of the genes encoding VP60 of representative lagoviruses were used to construct the tree. Bootstrap values greater than 60% are shown.

originated respectively from the virulent variant GI.1b (no more circulating in Europe) or the nonpathogenic strain GI.4 identified in Australia. The evolutionary rate of VP60 gene for RHDV has been estimated to be 2.77×10^{-3} substitutions per site per year and the estimation of the evolutionary time scale places the date of origin of RHDV between 1970 and 1981 and hence only shortly before its first description in 1984. Similar studies performed for RHDV2 circulating in Europe identified a mean substitution rate of 3.93×10^{-4} substitutions per site per year, and place the RHDV2 origin on July 2008, just two years before its first identification in France. The mean rate of nucleotide substitution for EBHSV ranges from 2.6 to 3.6×10^{-3} subs/site/year and the estimated time scale places its origin between 1969 and 1980, which fits with its circulation a few years before the original diagnosis in 1980.

Epidemiology

The pathogenic lagoviruses independently emerged on at least three distinct occasions. EBHSV and RHDV emerged in the early 1980s, but while EBHS in hares was first noticed in Scandinavian countries, RHD in rabbits was reported in China in 1984 and two years later in Europe. In a few years RHDV caused devastating epidemics worldwide where the European rabbit is present. Few RHD outbreaks occurred in the USA, South America, and Canada where rabbit populations are very scarce. In 1995 RHDV was released into Australia, as a new possible rabbit biological control agent, like the Myxoma virus in the 50s, starting an epizootic that continues to this day.

A first antigenic subtype of RHDV, named RHDVa (GI.1a), was identified in Europe in 1996. However, VP60 sequence of RHDVa of China isolates deposited in Genbank, dating back to 1985, confirms that the emergence of RHDV occurred in that country.

About 30 years later, the third emergency occurred in Europe, firstly noticed in France both in wild and farmed rabbits that succumbed to RHD although they were vaccinated against RHDV. The distinctive features of RHDV2 favored its rapid and wide diffusion in Europe and Australia, almost replacing RHDV. Particularly, RHDV2 diffused in an almost immunologically naïve rabbit population, with a high RHDV herd immunity largely ineffective against RHDV2. In addition, a broader usage of the HBAGs as anchors to the cells in comparison with RHDV enlarged the potential host susceptibility, further considering that RHDV2 kills also kittens. More recently, RHDV2 outbreaks were registered also in North America in colonies of European rabbits.

EBHS is a disease of the brown hare (*Lepus europaeus*) that occurred in Europe some years before the appearance of RHD. In fact, the earliest confirmed case was from 1981 in Sweden. Thereafter it has been reported in many European countries but never outside Europe. Cases of EBHS are observed also in mountain hare (*Lepus timidus*) and Italian hare (*Lepus corsicanus*); however, they were limited to the areas where these species live in sympathy with brown hare.

In relation to the presence of RCVs, several studies have been performed in Australia to understand the influence of RCV-A1 on the diffusion of the virulent lagoviruses. As RCV-A1 is antigenically distinct from RHDV and even more from RHDV2, rabbits infected by RCV-A1 are poorly protected against RHD.

RHDV and EBHSV spread very quickly and the infection can occur mainly through the oral routes. Transmission could be direct due to exposure to infected or dead animals or indirectly by means of fomites, contaminated vegetation, food, bedding, water, equipment, tools, vehicles, etc. Indeed, considering the stability of lagoviruses in the environment, they could be easily passively transmitted by other animals, like carnivores, scavenger birds, insects (flies), and even humans.

The virus is excreted in feces, urine, and respiratory secretions. The virus has been detected also in bone marrow, where it remains infectious for several months, as well as in rabbit meat, where the virus is able to survive freezing. Thus, importation of contaminated rabbit meat from subclinically infected, regularly slaughtered rabbits is considered a likely way of transmission of lagoviruses to new areas/continents.

Clinical Features

RHD and EBHS usually appear abruptly and the clinical evolution can be peracute, characterized by sudden death, acute, or subacute/chronic. Peracute and acute forms are prevalent where animal populations are fully susceptible to the diseases.

Animals with the acute form survive fairly longer (2–4 days), with signs of dullness/apathy, anorexia and prostration, neurological signs (e.g., incoordination, opisthotonos, convulsions, squeals, and paddling) and respiratory signs (e.g., dyspnea and a terminal, bloodstained, foamy nasal discharge). With respect to RHD, EBHS has a slightly longer course (death occurs at 3–5 days p.i.); the number of hares showing subclinical/chronic signs is higher (30%–50%) and mortality rate is lower (40%–60%) compared with RHD infected rabbits. In wild hares, changes in behavior could be observed (e.g., lack of fear, dullness, jumping into the air, circling, incoordination and convulsion excitement, respiratory distress) before death.

Rabbit and hares with the subacute/chronic form of the disease survive longer (over 6–8 days p.i.) and develop severe generalized jaundice, accompanied by weight loss and lethargy. Death usually occurs due to liver dysfunction, but some animals may survive, readily recovering in few days.

Pathogenesis and Host Immunity

RHD and EBHS can be experimentally reproduced by inoculation of seronegative adult animals with infected liver homogenates. The oro-nasal route permits to study the natural course of infection and its pathogenesis. In experimentally infected rabbits, RHDV

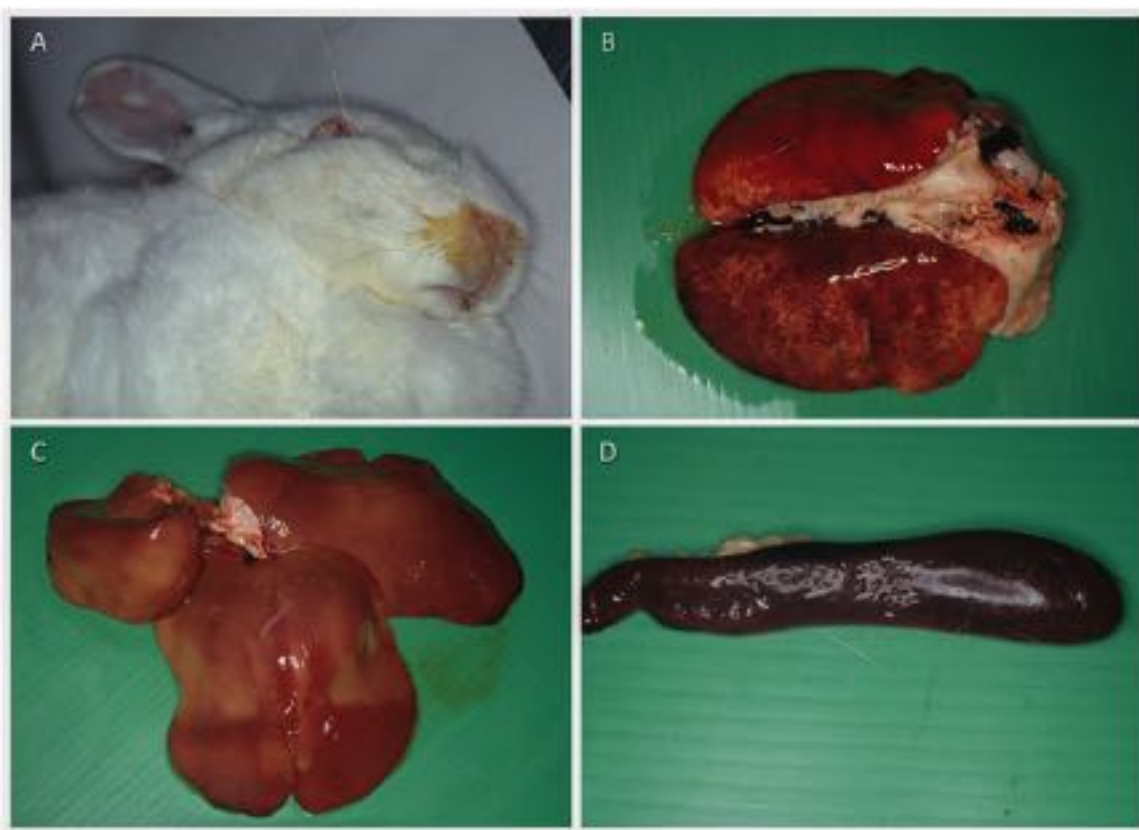


Fig. 4 Typical pathological lesions associated with RHD. (a) Rabbit died as a result of an acute form of RHD showing frothy, bloody pollution of the nostrils area; (b) lung congestion and edema with multifocal hemorrhages; (c) enlarged, degenerated and discolored liver; (d) spleen congestion and enlargement.

is detected in the epithelial cells of the nasal cavity, nasopharynx, and in the salivary glands at 12–24 h p.i. At 24–48 h p.i. the virus is detected in the hepatocytes, where it progressively replicates, inducing liver necrosis and inflammatory changes until the death of the animal. The virus induces high-level of viremia. In the terminal stages of the disease, disseminated intravascular coagulation and ischemic necrosis in various organs (including the central nervous system), shock, and hemorrhages are observed. In rabbits with subacute/chronic disease (up to 7–10 days p.i.), the viral load is in general lower, and the liver necrosis is focal. Mononuclear phagocytic cells in blood and organs (e.g., liver Kupffer cells) contain viral antigen, according to their role in the viral clearance.

At necropsy, both in case of RHD and EBHS, a frothy, bloody pollution of the nostrils area and, in case of subacute/chronic disease, an icteric discoloration in the pinnae of the ears and of external mucosae could be observed. Indeed, the most severe lesions are liver necrosis and degeneration (pale, swollen, friable, yellowish-brown in color, and marked lobular pattern) and spleen congestion and enlargement, accompanied by poor blood coagulation and petechial hemorrhages in almost all organs (Fig. 4).

The histopathological changes observed in rabbits infected with RHD and hares with EBHS are similar. These include necrotic hepatitis with multifocal necrosis, fibrosis and calcification (mainly in hares), follicular karyorrhexis and red pulp necrosis of spleen and lymphoid tissue, edema, microthrombi, hyperemia, hemorrhages, leukocyte infiltration and calcification in lungs, trachea, kidneys, and SNC.

Young animals infected with either RHDV or EBHSV show only a subclinical infection, with the innate immune system seemingly playing a key role for this distinct clinical outcome as opposed to the lethal hepatitis displayed in adult animals. In fact, rabbit kittens treated with a corticosteroid with pleiotropic actions and then challenged with RHDV developed RHD in 100% of cases.

In addition, a recent genome-wide study of liver transcripts from adult and young rabbits showed that kittens infected with RHDV displayed an increased level of expression of multiple genes encoding components of the innate immune response (particularly those associated with MHCII genes) compared with adult rabbits. In contrast, the same genes were downregulated in kittens during RHDV2 infection. The authors proposed a model where the innate immune system of young rabbits is able to respond rapidly to pathogens, including RHDV, whose replication is hindered in liver (very few hepatocytes become infected). Conversely, RHDV2 prevents the innate immune response, replicating at high level also in the liver of kittens, so causing RHD.

The final failure of the liver functions, resulting in animal death, occurs a few hours before the appearance in the blood of high levels of virus-specific IgM, which triggers the rapid removal of the virus from the blood stream. The huge quantity of IgM-virus

immunocomplexes is mainly conveyed to the liver and spleen to be cleared, and this could contribute to cause the chronic form of the disease. In addition to specific IgM, also virus-specific IgA is detectable in the serum of rabbits at 3–4 days p.i. Conversely, virus-specific IgG appears in serum 7–8 days p.i., becomes the Ig dominant class in few weeks, and remains at detectable levels for over one year. In case of lagovirus infection in disease resistant animal (i.e., RHDV and EBHSV in young, RCVs in rabbits, HaCV in hares) the kinetic of immunoglobulin classes is similar, but with serum titers 20–40 times lower.

Animals vaccinated with inactivated vaccine develop an IgM response in 6–8 days, followed by IgG at 10–14 days with titers similar to those found in rabbits infected by nonpathogenic virus. Specific IgG remain detectable for several months, while usually IgA are not found, presumably because parenteral vaccination of a killed virus does not involve the mucosal immune system, main producer of IgA. Conversely, seropositive rabbits, when infected with a wild type virus develop a transient peak of IgA, detectable in serum for 1–2 months, due to the replication of the virus at the mucosal level. Experimental infections demonstrated that even very low levels of specific IgG protect rabbits from RHD. Altogether, available data indicate that humoral immunity is the main defense mechanism in lagomorphs against lagoviruses-induced diseases.

Diagnosis

Both RHD and EBHS are characterized by typical clinical signs as well as by characteristic gross lesions but confirmation by laboratory diagnosis is strongly recommended.

Initially, the hemagglutination (HA) test using human red cell type “0” and immunoelectron microscopy were largely used for detection of RHDV but these tests were soon replaced by techniques such as reverse transcription polymerase chain reaction (RT-PCR) and enzyme-linked immunosorbent assay (ELISA).

Liver and spleen are the best organs to use for diagnosis as they contain very high levels of virus. Hence, ELISA methods using polyclonal or specific monoclonal antibodies are very reliable, being able to detect the agent at high dilutions of the tissue homogenate. For chronic cases of the disease, tests such as HA and ELISA are less reliable due to the presence of high levels of virus-IgM immunocomplexes.

Since RT-PCR tests can be positive also in convalescent animals for over two months, in recently vaccinated rabbits and in animals with ongoing not pathogenic RCV infections, false positive RHD or EBHS results are possible. Consequently, the use of ELISA or of quantitative real time RT-PCR is strongly recommended for a reliable diagnosis of the diseases.

ELISA-based methods are commonly used as serology tests for seroepidemiological surveys investigating the prevalence of infection or vaccine efficacy. Detection of lagovirus-specific antibodies is based on three different ELISA-based assays:

- (1) Classical ELISA (spELISA): the purified virus (or virus-like particles) adsorbed directly to the solid phase. Since adsorption to the plastic causes capsid denaturation with exposure of the internal structure (i.e., the S domain highly conserved among all lagoviruses), spELISA is the best method to detect the subset of cross-reactive antibodies. This ELISA is recommended to identify newly infected animals that have not been previously infected by any lagovirus (pathogenic or not).
- (2) Competition ELISA (cELISA): the virus–antibodies reaction occurs in the liquid phase where the virus is in a native conformation and exposes only external antigenic determinants (i.e., present on the VP60 P2 sub domain). Consequently, cELISA is optimal to detect the subset of specific and protective antibodies in the serum. Testing positive RHDV, RHDV2, and EBHSV sera with homologous and heterologous cELISAs provides a virus-specific antigenic profile.
- (3) Isotype ELISAs (isoELISA): distinct methods to detect virus-specific IgM, IgA, and IgG. These methods can be used in seroepidemiological studies on wild rabbit populations since they allow the distinction of seropositive rabbits in those who have been recently infected (high IgM titers), those recently reinfected (positive IgA titers) or those with an older infection (only IgG positive).

Prevention and Control

Eradication of RHD is unfeasible in regions where wild European rabbits are present and where there are high numbers of small backyard farms, usually lacking biosecurity approaches. Effective measures to prevent RHD could be adopted only in farmed, domestic/pet, or laboratory animals through the application of vaccination programs.

Several vaccines for RHDV2 as well as RHDV/RHDVa are commercially available, prepared from clarified inactivated liver suspension of experimentally infected rabbits (Lagovirus do not replicate in *in vitro* systems) used in conjunction with adjuvants (incomplete mineral oil or aluminum hydroxide).

Vaccines prepared with a specific lagovirus (RHDV/RHDVa or RHDV2) are poorly cross-protective against infections by other lagoviruses. Combined vaccination with both antigenic virus strains is highly advisable unless there is a clear understanding of the epidemiological situation of the region where the vaccine is sought.

There are no commercial vaccines for EBHS, but when EBHS outbreaks occurs, autogenous killed virus vaccines can be produced from livers of the hares infected at the beginning of the outbreak, to immunize the remaining apparently unaffected hares.

Since protection induced by vaccination usually lasts for over 1 year, most vaccine manufacturers recommend a single vaccination, from 25 to 30 days of age onwards, with annual boosts. However, a booster vaccination after 3–4 weeks and then a

further vaccination after 6–12 months depending on epidemiological situation is recommended to ensure a good level of protection.

RHD is an OIE-listed disease, hence reporting the disease to veterinary authorities is compulsory in most countries.

When an RHD outbreak in a farm is laboratory confirmed and the agent typed, all the surviving animals, including young rabbits, should be immediately vaccinated. In addition, direct prophylaxis actions should be adopted, including cleaning and disinfection of premises, equipment, plants, and instruments; safe removal of carcasses and fomites; and control of movement of people vehicles and animals. A quick response is vital for containing outbreaks in disease-free regions and preventive actions should be extended to rabbit farms surrounding the outbreak.

Treatment

Given that immunity starts very quickly, vaccination can also be considered an effective post-exposure emergency measure, especially when a homologous vaccine is used and depending from the farm organization.

Serotherapy through the parenteral administration of a homologous hyperimmune antiserum has been shown to be effective in limiting the spread of the disease and reducing economic losses. This short-lived, passively acquired immunity, can protect animals that have not yet developed clinical signs. Other treatments are currently limited to supportive care especially in pet rabbits.

Further Reading

- Abrantes, J., van der Loo, W., Le Pendu, J., Esteves, P.J., 2012. Rabbit haemorrhagic disease (RHD) and rabbit haemorrhagic disease virus (RHDV): A review. *Veterinary Research* 43, 12–31.
- OIE's Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, 2012. Chapter 2.6.2. Rabbit Haemorrhagic Disease, seventh ed., version adopted in May 2016.
- Bárcena, J., Verdagué, N., Roca, R., *et al.*, 2004. The coat protein of rabbit hemorrhagic disease virus contains a molecular switch at the N-terminal region facing the inner surface of the capsid. *Virology* 322, 118–134.
- Cooke, B.D., 2014. Australia's War Against Rabbits. The Story of Rabbit Hemorrhagic Disease. CSIRO Publishing.
- Eden, J.S., Read, A.J., Duckworth, J.A., Strive, T., Holmes, E.C., 2015. Resolving the origin of rabbit hemorrhagic disease virus: Insights from an investigation of the viral stocks released in Australia. *Journal of Virology* 89, 12217–12220.
- Frollich, K., Lavazza, A., 2007. European brown hare syndrome. In: Alves, P.C., Ferrand, N., Hacklaender, K. (Eds.), *Lagomorph Biology*. Berlin: Springer, pp. 253–262.
- Lavazza, A., Capucci, L., 2007. How many calicivirus are there in the rabbit? A review on RHDV and correlated viruses. In: Alves, P.C., Ferrand, N., Hacklaender, K. (Eds.), *Lagomorph Biology*. Berlin: Springer, pp. 263–278.
- Le Gall-Reculé, G., Lavazza, A., Marchandeau, S., *et al.*, 2013. Emergence of a new lagovirus related to rabbit haemorrhagic disease virus. *Veterinary Research* 44, 81–94.
- Le Pendu, J., Abrantes, J., Bertagnoli, S., *et al.*, 2017. Proposal for a unified classification system and nomenclature of lagoviruses. *Journal of General Virology* 98, 1658–1666.
- Lopes, A.M., Breiman, A., Lora, M., *et al.*, 2017. Host specific glycans are correlated with susceptibility to infection by lagoviruses, but not with their virulence. *Journal of Virology* 92, e017559-17.
- Lopes, A.M., Capucci, L., Gavier-Widén, D., *et al.*, 2014. Molecular evolution and antigenic variation of European brown hare syndrome virus (EBHSV). *Virology* 468–470, 104–112.
- Lopes, A.M., Dalton, K.P., Magalhães, M.J., *et al.*, 2015. Full genomic analysis of new variant rabbit hemorrhagic disease virus revealed multiple recombination events. *Journal of General Virology* 96, 1309–1319.
- Neave, M.J., Hall, R.N., Huang, N., *et al.*, 2018. Robust innate immunity of young rabbits mediates resistance to rabbit hemorrhagic disease caused by lagovirus europaeus GI.1 But Not GI.2. *Viruses* 10, 512–534.
- Neimanis, A., Larsson Pettersson, U., Huang, N., Gavier-Widén, D., Strive, T., 2018. Elucidation of the pathology and tissue distribution of lagovirus europaeus GI.2/RHDV2 (rabbit haemorrhagic disease virus 2) in young and adult rabbits (*Oryctolagus cuniculus*). *Veterinary Research* 49, 46–61.
- Wang, X., Xu, F., Liu, J., *et al.*, 2013. Atomic model of rabbit hemorrhagic disease virus by cryo-electron microscopy and crystallography. *PLoS Pathogen* 9, e1003132.
- Zhu, J., Miao, Q., Tang, J., *et al.*, 2018. Nucleolin mediates the internalization of rabbit hemorrhagic disease virus through clathrin-dependent endocytosis. *PLoS Pathogen* 14, e1007383.

Rabbit Myxoma Virus and the Fibroma Viruses (*Poxviridae*)

Peter J Kerr, University of Sydney, Sydney, NSW, Australia and CSIRO Health and Biosecurity, Black Mountain Laboratories, Canberra, ACT, Australia

© 2021 Elsevier Ltd. All rights reserved.

Glossary

European rabbits *Oryctolagus cuniculus*: this species evolved in the Iberian Peninsula and southern France. All the domestic and farmed breeds of rabbit including laboratory rabbits are derived from the original wild populations. European rabbits have been spread around the world and established in the wild in Australia and New Zealand as well as parts of the Americas and many islands.

Feral Feral populations are previously domestic animals that have established in the wild.

Host-pathogen coevolution The dynamically opposed processes of natural selection acting on a virus to maximize its transmission and the host species to control virus replication.

Species-jump The shift of a virus into a novel host-species in which it was not previously established.

Myxoma Virus and the Fibroma Viruses

Classification

Family: *Poxviridae*; subfamily: *Chordopoxvirinae*; Genus: *Leporipoxvirus*; Species: *Myxoma virus* (type species); *Rabbit fibroma virus* (Shope fibroma virus); *Squirrel fibroma virus*; *Hare fibroma virus*.

Virion Structure

The virions are indistinguishable from those of other poxviruses such as *Vaccinia virus*: brick shaped approximately $300 \times 250 \times 75$ nm with convex ends and a disordered surface tubular structure. As with other poxviruses, two infectious forms of virus are formed: intracellular mature virions, which in thin section electron micrographs have an external lipoprotein membrane surrounding a nucleoprotein core and two lateral bodies and are released on cell rupture and extracellular enveloped virions that are surrounded by an additional cell-derived lipid envelope containing virus-encoded proteins not present in the intracellular mature form.

Genome

Like all poxviruses, the leporipoxvirus genome consists of double stranded DNA with terminal inverted repeats (TIR) and closed single strand hairpin termini. The Lausanne strain is the type sequence of myxoma virus with a genome of 161,777 base pairs (bp) including TIRs of 11,577 bp. There are 158 unique open reading frames defined; 11 of which are duplicated in the TIRs. Full genome sequences have also been obtained for the Kazza strain of rabbit fibroma virus and the MSW strain of Californian myxoma virus.

Life Cycle

Myxoma virus replication is similar to that of other chordopoxviruses, the best characterized being the orthopoxvirus vaccinia virus. The key steps in replication are cell attachment and entry followed by early gene expression, uncoating of the viral core, DNA replication, intermediate and late gene expression and viral assembly all of which occur in the cytosol of the infected cell.

No specific cell surface receptor has been identified for myxoma virus. Following attachment of the mature virion, uptake is triggered via macropinocytosis, which requires activation of multiple cell-signalling pathways and dynamic actin rearrangements. This delivers the virion into the endocytic pathway. Subsequent fusion of the viral membrane and the endosomal membrane releases the viral nucleoprotein core into the cytoplasm where it is transported to the perinuclear region. Enveloped virus presumably requires different attachment factors and must disrupt its extra envelope membrane to expose the fusion entry complex, 11 highly conserved viral proteins assembled on the mature viral membrane, before entry can occur. This disruption may possibly occur at the cell surface or within the endosome.

Gene expression is temporally regulated by viral transcription factors and promoters. Early gene expression occurs within the viral core from packaged transcription machinery. This is followed by uncoating of the genome. Viral DNA replication, intermediate and late gene transcription and gene translation and viral assembly occur in membrane-bound, discrete viral factories that form around the unpackaged core. DNA replication involves formation of long concatemers that are resolved into unit length genomes by the viral Holliday junction resolvase. Intermediate and late gene transcription occurs from the newly synthesised DNA.

Viral assembly is a complex process involving membranes, probably derived from the endoplasmic reticulum, that complex with viral scaffold proteins and expand to form crescents that grow into spheroids containing the viral genome and core proteins and called immature virions. Subsequent proteolytic processing, disulphide bond formation and the assembly of surface membrane proteins forms the mature virion consisting of the nucleoprotein core and two lateral bodies enclosed within the viral surface lipoprotein membrane. Mature virions are infectious and released upon cell lysis. Some mature virion particles can acquire an extra double membrane, likely from Golgi membranes, together with additional viral envelope proteins, to form wrapped virions. These wrapped virions are transported to the cell plasma membrane where exocytosis occurs with the loss of one layer of the double membrane to form extracellular enveloped virions which are considered important in cell-cell spread and dissemination within the host.

Epidemiology

The natural history of the leporipoxviruses can be considered in two parts. Firstly, in their “natural hosts” and geographic range (Table 1) and secondly, the species jump of the South American type of myxoma virus into a novel host, the European rabbit (*Oryctolagus cuniculus*), and expansion of its geographic range into Australia, Europe and new areas of South America.

With the exception of hare fibroma virus, the known leporipoxviruses are native to the Americas. All appear to have a relatively narrow host range in which the viruses typically cause cutaneous fibromas. Transmission relies on biting arthropods, usually fleas or mosquitoes, which pick up virus on their mouthparts when they probe through the virus-rich epidermis overlying the fibroma. Virus is then passively transmitted from the mouthparts when the vector feeds on another animal. The virus does not replicate in the vector.

Myxoma Virus

Myxoma virus causes a lethal disease of European rabbits termed myxomatosis because of the mucinous character of the skin lesions on incision. This disease was first described following an outbreak in rabbits housed in a laboratory in Montevideo, Uruguay in 1896. European rabbits are not native to the Americas but have been widely introduced and are present as both domestic and feral populations with widespread feral populations in Argentina and Chile.

Subsequent studies in Brazil showed that myxoma virus in this region most likely circulates in the native forest rabbit or tapeti *Sylvilagus minensis* (also referred to as *S. brasiliensis minensis* and *S. brasiliensis*). In this species, the virus induces a localized cutaneous fibroma 1–2 cm in diameter at the inoculation site. Infection appears to have little impact on the host although more generalized disease was reported in juvenile tapetis. The fibroma is cleared by the host immune response within a few weeks.

Myxomatosis is predominantly a disease of European rabbits but it occasionally occurs in European brown hares (*Lepus europaeus*), Iberian hares (*L. granatensis*) and mountain hares (*L. timidus*). In addition, North American *Sylvilagus* species such as *S. nuttallii* (mountain cottontail) and *S. audubonni* (desert cottontail) are susceptible to experimental infection with South American myxoma virus and could potentially transmit virus. However, their natural geographic range does not overlap with the South American virus.

Myxoma virus has been isolated from European rabbits in Central America and Colombia with some isolates antigenically related to Brazilian and others to Californian myxoma virus. The virus has been deliberately released in Chile and has also spread into Argentina.

Californian Myxoma Virus

A virus related to South American myxoma virus circulates in the brush rabbit, *Sylvilagus bachmani* on the west coast of the United States and Mexico. Hairless, well delineated fibromas 1–2 cm in diameter occur at the base of the ears, on the legs or even the lip.

Table 1 The Leporipoxviruses and probable natural hosts

Virus	Natural host	Geographic range	Disease
Myxoma virus	Forest rabbit (tapeti) (<i>Sylvilagus menensis</i>) in South America Brush rabbit (<i>S. bachmani</i>) in North America	Brazil; Argentina West coast of USA and Mexico	Localized cutaneous fibroma
Rabbit fibroma virus	eastern cottontail (<i>S. floridanus</i>)	Eastern and Central North America	Localized cutaneous fibroma
Squirrel fibroma virus	Grey squirrel (<i>Sciurus carolinensis</i>) ^a	Eastern North America	Cutaneous fibroma; may be generalized in young squirrels and affect internal organs
Hare fibroma virus	European brown hare (<i>Lepus europaeus</i>) ^b	Europe	Localized cutaneous fibroma

^aA poxvirus that antigenically cross-reacted with Californian myxoma virus has been described from a fibroma on a Western grey squirrel (*Sciurus griseus griseus*).

^bAn uncharacterized poxvirus has been demonstrated by electron microscopy in fibromas on hares (*Lepus capensis*) from Kenya.

More than one fibroma may occur on an individual. Following experimental inoculation, fibromas take several weeks to develop and begin to scab 5 weeks or so after infection before being cleared by the host immune response. Occasionally, fibromas can persist for up to 3 months.

Outbreaks of myxomatosis occur in European rabbits on the west coast of the USA and Mexico. Some strains of Californian myxoma virus cause a fulminant form of myxomatosis with death occurring before the typical clinical signs of myxomatosis can develop.

The virus appears well adapted to the brush rabbit. Experimentally, other *Sylvilagus* species may be infected with Californian myxoma virus but titres of virus in the lesions are too low for vector transmission. European brown hares (*Lepus europaeus*) are susceptible to infection. Brush rabbits could be infected with South American myxoma virus but titres in the lesion were insufficient for mosquito transmission.

Myxomatosis in European Rabbits

The lethality of myxomatosis in European rabbits, together with the species-specificity of the virus, suggested that myxoma virus might be used as a biological control for this introduced pest species in Australia. Experimental field trials in 1950 led to the unexpected spread of myxoma virus across a large rabbit-infested area of the south-eastern part of the continent causing enormous numbers of deaths. The virus has circulated in the Australian rabbit population ever since. Subsequently, in 1952, a landowner in France inoculated two wild rabbits with a separate strain of myxoma virus and from this introduction the virus spread and established throughout the wild and farmed rabbit populations of Europe and Britain. As in Australia, massive population crashes occurred. Thus from a relatively obscure origin in the Americas, myxomatosis had emerged as a major disease of domestic and wild rabbits.

These separate introductions on two continents initiated one of the best documented examples of host-pathogen coevolution following a species jump by a pathogen. The initially released viruses had case fatality rates in excess of 99% with death typically occurring 8–14 days after infection. This meant that there was only a short time period between developing transmissible virus titres in the lesions and death of the rabbit. With no natural host reservoir, virus was rapidly selected for slightly lower virulence, which allowed the infected rabbit to survive for longer in an infectious state and so increased the likelihood of vector transmission. These moderately attenuated viruses had case fatality rates ranging from 50% to 99%. At the same time the very high population mortality rate meant that resistance to myxomatosis was strongly selected in the wild rabbit populations. Because of the short generation interval in rabbits this resistance could be measured in real-time. At one Australian study site, case fatality rates for a particular virus strain dropped from 90% to 26% over seven generations. So both the virus and the rabbit were undergoing rapid natural selection. Resistance was slower to emerge in Europe and Britain although strong resistance now exists in wild rabbits. The rabbits are not resistant to actual infection but control of virus replication by the innate immune system occurs distal to the inoculation site. Resistance can be overcome by viruses of higher virulence or manipulation of the innate immune response. It is likely that resistance in the wild rabbit population has selected for higher virulence (when measured in non-resistant rabbits) in the circulating virus strains to maintain transmissibility since, unlike in the natural host species, highly attenuated viruses are poorly transmitted from European rabbits.

Clinical Disease in European Rabbits

Two clinical forms of myxomatosis are described. A nodular or myxomatous form characterized by a large elevated primary cutaneous lesion at the inoculation site and multiple secondary cutaneous lesions – this was the typical disease seen following the introduction of myxoma virus in Australia and Europe. Subsequently, a second form of disease with minimal cutaneous lesions, sometimes termed amyxomatous, has emerged.

In the nodular form, the initial clinical sign is a slight redness and swelling at the inoculation site 2–3 days after inoculation. Elevated rectal temperature $> 40^{\circ}\text{C}$ can occur from day 4 onwards. Depending on the virus strain, this primary lesion gradually enlarges and may become as much as 2 cm in height and 4–5 cm in diameter by 10–12 days after infection. It sits loosely within the skin and late in infection may erode and ulcerate at the surface with haemorrhage and scabbing. From 4–5 days after infection, the ano-genital area becomes increasingly red and swollen, the eyelids redden at the edges and start to thicken and the ears swell especially at the base. By 8–12 days after infection the head can be extremely swollen with drooping ears and swollen, closed eyelids often accompanied by a mucoid or mucopurulent conjunctival discharge (Fig. 1). Ano-genital swelling can be quite extreme at this stage. Secondary elevated cutaneous lesions up to 2 cm in diameter are present on the body, legs, ears and head. Swollen upper respiratory tract mucosa together with mucoid or mucopurulent discharge in the nasal passages can cause respiratory tract obstruction and dyspnoea. Death typically occurs around this point although the proximate cause of death is unclear. Viruses of lower virulence cause a more protracted although essentially similar clinical course; secondary bacterial infection can be a major cause of mortality in these cases. Inappetence or anorexia and significant weight loss occurs but some rabbits can mount a successful immune response and survive infection. Recovered rabbits have long-term immunity to reinfection.

Vector transmission can occur from cutaneous lesions and other sites where high titres of virus occur in the epidermis such as eyelids and base of the ears. Virus is shed in most secretions and contact transmission can occur via conjunctival or respiratory tract inoculation but not orally.

In the amyxomatous form, no or minimal cutaneous lesions are seen but otherwise a similar clinical course to the nodular form occurs with swollen head, eyelids and ano-genital region and death occurring 2–3 weeks after infection (Fig. 2). Secondary



Fig. 1 Domestic rabbit with severe nodular myxomatosis. Figure is from: Kerr, P.J., Cattadori, I.M., Rogers, M.B., *et al.*, 2017. Genomic and phenotypic characterization of myxoma virus from Great Britain reveals multiple evolutionary pathways distinct from those in Australia. *PLOS Pathogens* 13, e1006252, under the terms of the Creative Commons Attribution License.



Fig. 2 Domestic rabbit with amyxomatous form of myxomatosis.

bacterial infection may be important in disease outcome. Transmission occurs between rabbits in direct contact but not via aerosol; the role of vectors is unclear. Viruses causing amyxomatous disease in laboratory rabbits have also been described from Australia and Britain. Some of these viruses cause a peracute syndrome manifesting as pulmonary oedema and haemorrhage in multiple tissues 9–16 days after infection. In these rabbits, rectal temperatures were elevated briefly early in infection before returning to the normal range and then often spiking to very high levels shortly before death. The Californian myxoma viruses can also cause a peracute form of myxomatosis with minimal cutaneous lesion development and death in 7–9 days after infection. In these forms of the disease, infected rabbits may simply be found dead and in those animals the early clinical signs may not be noticed.

Pathogenesis of Myxomatosis in European Rabbits

Following intradermal inoculation, myxoma virus initially replicates in MHC-II positive cells in the dermis and the dermal-epidermal interface. From here it spreads to infect the cells of the epidermis and to the draining lymph node where it can be detected within connective tissue of the subcapsular sinus and subsequently within paracortical lymphocytes. From the draining lymph node, virus spreads throughout the body, probably within infected lymphocytes as there is no free virus in the blood, to distal lymphoid tissues and cutaneous and mucocutaneous sites such as eyelids. In males, high titres may occur in testis and epididymis.

At the inoculation site, keratinocyte hypertrophy and hyperplasia is followed by ballooning degeneration of cells and vesicle formation within the thickened epidermis (**Fig. 3**). Oedema and the accumulation of large amounts of mucinous material in the dermis causes disruption of the collagen and hair follicles. Few inflammatory cells are seen in the epidermis where most of the virus is present. Deeper in the dermis there may be large numbers of neutrophils particularly around small blood vessels, which are frequently distorted by “myxoma cells,” large stellate cells, probably fibroblasts in origin, which appear to bud through and disrupt the endothelium allowing extravasation of red blood cells. Virus is present in the myxoma cells and within endothelial cells. Late in

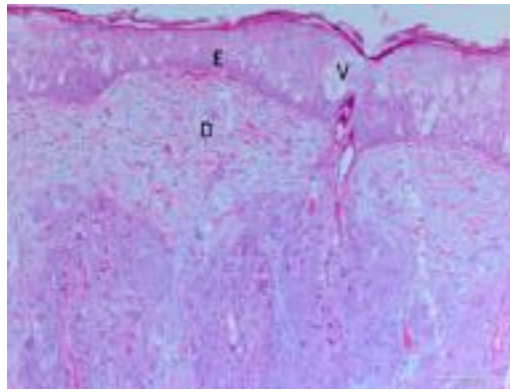


Fig. 3 Histopathology of primary lesion. Proliferation and hypertrophy of cells in the epidermis (E) with ballooning degeneration and beginnings of vesicle formation (V). The underlying dermis (D) is disrupted by pale staining mucinous material.

infection, there may be complete loss of the epidermis with formation of a thick scab. Intracytoplasmic eosinophilic inclusions may sometimes be prominent in degenerating epidermal cells but are not always seen in haematoxylin and eosin stained sections.

The histological appearance of secondary cutaneous lesions, eyelids and ears is similar to that at the primary lesion but usually less advanced at the time of death.

Disruption of lymphoid tissue architecture with depletion of lymphocytes, destruction of follicles and proliferation of reticuloendothelial cells is a major feature of the acute phase of the disease particularly in lymph nodes; the spleen is less consistently affected and even some lymph nodes may have normal areas. Large numbers of neutrophils may be present in affected lymph nodes.

In the amyxomatous form of myxomatosis, the lesion at the inoculation site can include minor hypertrophy and hyperplasia of epidermal cells with little deposition of mucin within the dermis and limited disruption of collagen and blood vessels. Despite the limited pathology the titres of virus at these sites can be very high. In the peracute form of the disease, neutrophils are rarely seen in any tissues. However, large numbers of bacteria may be present throughout the tissues, within blood vessels or macrophages but without an obvious inflammatory response. High titres of virus can occur in liver and lung.

Myxoma virus is profoundly immunosuppressive. Over 40 genes encode proteins with demonstrated or probable roles in manipulating or suppressing host anti-viral responses. These include proteins that suppress type I and type II interferon responses, inhibit multiple steps in inflammatory pathways, inhibit tumour necrosis factor, inhibit activation of macrophages and T cells, competitively bind pro-inflammatory chemokines, downregulate MHC-I molecules from the surface of infected cells and inhibit cell death pathways. All of these proteins must have evolved in the *Sylvilagus* host species where the local innate and adaptive immune responses are sufficiently suppressed to allow continued replication at the inoculation site. This is despite the development of an adaptive immune response demonstrated by resistance to superinfection within a few days and neutralizing antibody in the serum. In the European rabbit these proteins facilitate dissemination and generalization of the virus with replication in multiple tissues and ultimately death of the rabbit.

Diagnosis

Diagnosis of the acute or chronic nodular forms of myxomatosis in European rabbits is relatively straightforward since the clinical appearance is almost unique. More chronic forms of myxomatosis may be seen in wild rabbits where genetic resistance alters the outcome of infection. In these cases, large raised, grey, hairless swellings can be present over the head, ears and body and the affected rabbit often appears emaciated.

At necropsy, in the nodular form the main findings are on the external body and have been described above. Internally, there may be little to see beyond swollen lymph nodes and testes. The spleen may or may not be enlarged. There is usually little or no gross pathology in the lung, liver and other internal tissues. In chronic infections or recovering animals, lymph nodes and spleen may be massively enlarged and significant fat loss and emaciation will be apparent.

For the amyxomatous form, the external lesions may be relatively minor although head and ano-genital swelling and some swelling of the eyelids will usually be apparent. In the peracute form, there may be massive pulmonary oedema with swollen, wet lungs and fluid and froth filling the trachea and bronchi (Fig. 4). In some cases there will be visible haemorrhages over the lung surface and through the parenchyma. The liver may be enlarged and firm but this is less consistent. Haemorrhage may also be present in the wall of the appendix and caecum or in the hind legs. Subcutaneous oedema of hind legs and rump is common. Lymph nodes tend to be very small.

Differential diagnosis includes *Leporid herpesvirus 4*, which can also cause swelling of the head, blepharoconjunctivitis and death. The peracute forms of myxomatosis must be distinguished from other causes of acute death such as rabbit haemorrhagic disease, bacterial septicaemias, coccidiosis and poisonings for pest control. Low virulence forms of amyxomatous virus may circulate undiagnosed in rabbit farms and be confused with bacterial respiratory disease.

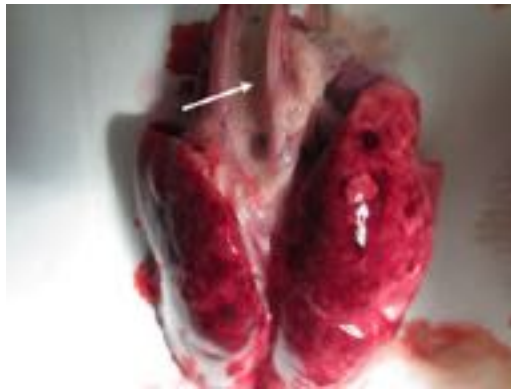


Fig. 4 Pulmonary oedema in peracute death. Arrow indicates froth filling the trachea which has been opened along its length.

In the acute stage of infection high titres of virus are present in lymph nodes, eyelids, testis and cutaneous lesions. At later stages, cutaneous lesions and eyelids are a more reliable source of virus. Conjunctival swabs are a readily collected source of virus for diagnosis from the live rabbit. Polymerase chain reaction (PCR) using DNA extracted from tissues or conjunctival swabs is the simplest and fastest means to confirm diagnosis. Electron microscopy can demonstrate typical poxvirus particles.

Myxoma virus can be readily cultured in cell lines such as RK-13. In these cells, virus replication causes raised proliferative foci of cells with formation of multinucleate syncytia followed by monolayer destruction. The virus remains predominantly cell associated. Myxoma virus can also be grown on the chorioallantoic membrane of embryonated chicken eggs. Virus identity can be confirmed by PCR or immunostaining.

The histology of skin lesions is reasonably characteristic. Immunostaining with specific antibody can be used on impression smears or frozen or fixed histological sections to detect myxoma virus.

Serum antibodies to myxoma virus can be detected as early as 8–10 days after infection by ELISA or virus neutralization assay.

Treatment

There is no specific treatment for myxomatosis. Many wild rabbits survive infection; some domestic rabbits infected with attenuated viruses will recover although illness may be prolonged for several months. General supportive care with attention to hydration and nutrition may assist survival. Antibiotics may be needed to control secondary bacterial infection.

Prevention of Myxomatosis

Vaccination and quarantine are the main means of preventing myxomatosis in domestic and farmed rabbits.

Three types of live vaccine are available. No inactivated vaccines have been successfully developed.

- (1) Heterologous live vaccines. Infection with rabbit fibroma virus provides cross-protection from myxoma virus. Following the spread of myxoma virus in Europe, vaccines based on rabbit fibroma virus were widely used in farmed and domestic rabbits and attempts also made to vaccinate wild rabbits.
- (2) Attenuated live myxoma virus vaccines. These have been developed by serial passage of virulent myxoma viruses in cell culture. Examples include the Borghi strain based on the MSD strain of Californian myxoma virus and the SG-33 strain based on an attenuated field strain from France.
- (3) Recombinant live myxoma virus. This vaccine was released commercially in 2012 and expresses the *Rabbit haemorrhagic disease virus type 1* (RHDV-1) capsid protein providing dual protection against both myxomatosis and rabbit haemorrhagic disease due to RHDV-1 (but not RHDV-2). Insertion of the RHDV capsid protein gene into the myxoma virus genome has been done such that critical virulence genes have been disrupted, producing an attenuated myxoma virus vaccine.

Vaccination is not permitted in Australia.

Introduction of myxoma virus into rabbit farms has occurred in new stock incubating the virus even though these were reportedly vaccinated. Vaccination and quarantine of any introductions should be routine. In countries where myxoma virus circulates in wild rabbits, protection from mosquitoes and fleas is critical to protect unvaccinated rabbits. Rabbits suspected of having myxomatosis should be culled or isolated until diagnosis is confirmed. Virus can be readily transmitted between in-contact rabbits and on contaminated water bottles or feeders. The virus is potentially transmitted in semen from infected males.

Rabbit Fibroma Virus

Rabbit fibroma virus circulates in eastern cottontails, *Sylvilagus floridanus*, which are widely distributed in the eastern and central United States, north into Canada and south into Central America. In the eastern cottontail, rabbit fibroma virus causes cutaneous fibromas, 1.5–2 cm in diameter, appearing as a raised hairless dome or flat plaque, most often on the feet or legs and not firmly attached to underlying tissues. More than one fibroma may be present. It takes 3–4 weeks after inoculation for the fibroma to mature into a readily transmissible form with high titres of virus in the surface epidermis. In immunologically mature cottontails the fibroma is cleared by the immune response after 6–8 weeks. In very young suckling cottontails fibromas can grow extremely large and cause death with virus also present in internal tissues. When slightly older kittens are infected, the fibroma can persist in an infectious state for as long as nine to ten months, despite the presence of neutralizing antibody in the serum and resistance to superinfection within a few days of the initial infection. This persistence would allow overwintering in the absence of mosquito vectors.

Histologically, in cottontails the fibroma consists of collagen, fibroblasts including stellate “fibroma cells” and fibrocytes. The overlying epidermis is thickened with hyperplasia, hypertrophy and degeneration of keratinocytes. Pegs of epidermis extend deep into the fibroma. Eosinophilic intracytoplasmic inclusions may be present in degenerating keratinocytes. Neutrophils and occasional lymphocytes are present within the fibroma while the underlying tissue can be massively infiltrated with lymphocytes.

In European rabbits, inoculation of rabbit fibroma virus induces a cutaneous fibroma at the inoculation site. This persists for 2–3 weeks before clearance by the immune response. In immunosuppressed or immunologically immature suckling rabbits more generalized fibromatosis and death can occur. Rabbit fibroma virus is rarely able to be experimentally transmitted by vectors from European rabbits, despite reaching very high titres in the fibroma. Localized outbreaks of rabbit fibroma virus occur in domestic European rabbits housed outdoors in the United States presumably spread by mosquitoes from local cottontail populations.

Rabbit fibroma virus can be cultured in cell-lines such as RK-13 where it causes similar cytopathic effect to myxoma virus.

Squirrel Fibroma Virus

Squirrel fibroma virus affects juvenile grey squirrels, *Sciurus carolinensis* in the eastern United States. Affected squirrels may have a few well-separated cutaneous fibromas or generalized fibromatosis with dozens of coalescing fibromas covering the body and nodules in lung, liver, kidney, spleen and lymph nodes. Fibromas range from a few mm to 3 cm in diameter. This highly variable host response was confirmed experimentally with some suckling squirrels developing generalized disease following inoculation while others in the same litter only had a single fibroma at the inoculation site. Adult squirrels appear refractory to infection: experimentally only 50% developed short-lived nodules at the inoculation site and seroconverted. Histologically, the fibromas resemble those caused by rabbit fibroma virus in cottontails.

Woodchucks (*Marmota monax*) can be infected experimentally although young animals are easier to infect than adults. Thickened nodules develop in the skin and subcutaneous tissues at the inoculation site. Squirrel fibroma virus replicates poorly in European rabbits and does not replicate in eastern cottontails.

Squirrel fibroma virus can be cultured in primary squirrel kidney cells and rabbit embryo fibroblasts. It replicates poorly in primary rabbit kidney cells and does not replicate in chick embryo fibroblasts.

Fibromas have also been described in fox squirrels (*Sciurus niger*) and porcupines (*Erethizon dorsatum*) from the eastern USA but these were not further characterized. A poxvirus, which cross-reacted antigenically with Californian myxoma virus, has been reported from a western grey squirrel (*Sciurus griseus griseus*). However, the virion morphology resembled a parapoxvirus. Squirrel fibroma virus is distinct from squirrel poxvirus, which occurs in grey and red squirrels (*Sciurus vulgaris*).

Hare Fibroma Virus

Hare fibroma virus causes cutaneous fibromas in European brown hares (*Lepus europaeus*). Little is known of the natural history of the virus, although occasional large outbreaks are recorded. One or more fibromas 1–3 cm in diameter raised above the skin are seen, usually on feet or ears. In most cases these scab and resolve over 1–3 months. Outbreaks have been reported in Germany, France and Italy.

Hare fibroma virus replicates poorly in European rabbits but will form subcutaneous nodules in new born animals. It forms pocks on the chorioallantoic membrane of chicken embryos and plaques on primary rabbit kidney cell monolayers but does not replicate in RK-13 cells.

Fibromas containing poxvirus particles have also been described in cape hares (*Lepus capensis*) from Kenya. Unfortunately, further characterization of this virus was not possible.

Further Reading

Cameron, C., Hota-Mitchell, S., Chen, L., *et al.*, 1999. The complete DNA sequence of myxoma virus. *Virology* 264, 298–318.

Fenner, F., Fantini, B., 1999. Biological control of vertebrate pests. The history of myxomatosis – An experiment in evolution. New York: CAB International.

- Fenner, F., Ratcliffe, F.N., 1965. Myxomatosis. Cambridge: Cambridge University Press.
- Grilli, G., Piccirillo, A., Pisoni, A.M., *et al.*, 2003. Re-emergence of fibromatosis in farmed game hares (*Lepus europaeus*) in Italy. *Veterinary Record* 153, 152–153.
- Kerr, P.J., 2012. Myxomatosis in Australia and Europe: A model for emerging infectious diseases. *Antiviral Research* 93, 387–415.
- Kerr, P.J., Cattadori, I.M., Liu, J., *et al.*, 2017. Next step in the ongoing arms race between myxoma virus and wild rabbits in Australia is a novel disease phenotype. *Proceedings of the National Academy of Sciences of the United States of America* 114, 9397–9402.
- Kilham, L., Herman, C.M., Fisher, E.R., 1953. Naturally occurring fibromas of grey squirrels related to Shope's rabbit fibroma. *Proceedings of the Society for Experimental Biology and Medicine* 82, 298–301.
- Shope, R.E., 1932. A transmissible tumour like condition in rabbits. *Journal of Experimental Medicine* 56, 783–802.
- Spiesschaert, B., McFadden, G., Hermans, K., Nauwynck, H., Van de Walle, G.R., 2011. The current status and future directions of myxoma virus, a master in immune evasion. *Veterinary Research* 42, 76.
- Waller, D.O., McFadden, G., Evans, D.H., 1999. The complete genome sequence of Shope (rabbit) fibroma virus. *Virology* 264, 319–343.

Relevant Website

<http://www.oie.int/en/animal-health-in-the-world/animal-diseases/Myxomatosis/>
Myxomatosis: OIE - World Organisation for Animal Health.

Rabies and Other Lyssaviruses (*Rhabdoviridae*)

Ashley C Banyard, Animal and Plant Health Agency, Addlestone, United Kingdom; University of West Sussex, Falmer, United Kingdom; and St. George's Medical School, University of London, London, United Kingdom

Anthony R Fooks, Animal and Plant Health Agency, Addlestone, United Kingdom; University of Liverpool, Liverpool, United Kingdom; and St. George's Medical School, University of London, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Anterograde Directed or moving forward.

Apoptosis The death of cells which occurs as a normal and controlled part of an organism's growth or development.

Autophagy It is a normal physiological process in the body that deals with destruction of cells in the body. It maintains homeostasis or normal functioning by protein degradation and turnover of the destroyed cell organelles for new cell formation. During cellular stress the process of autophagy is upregulated and increased.

Paraesthesia An abnormal sensation, typically tingling or pricking ('pins and needles'), caused chiefly by pressure on or damage to peripheral nerves.

Pleomorphic A variable appearance or morphology.

Pruritis An itching sensation.

Retrograde Directed or moving backwards.

Sylvatic rabies The sylvatic rabies disease is transmitted by wildlife, particularly foxes and wolves.

Classification (Compact)

Lyssaviruses are classified within the family Rhabdoviridae within the order Mononegavirales. Within the lyssavirus genus, there are 16 different viral species, classified as separate entities according to divergence across genomic sequence. The genus currently includes: Aravan lyssavirus (ARAV), Australian bat lyssavirus (ABLV), Bokeloh bat lyssavirus (BBLV), Duvenhage lyssavirus (DUVV), European bat –1 lyssavirus (EBLV-1), European bat –2 lyssavirus (EBLV-2), Gannoruwa bat lyssavirus (GBLV), Ikoma lyssavirus (IKOV), Irkut lyssavirus (IRKV), Khujand lyssavirus (KHUV), Lagos bat lyssavirus (LBV), Lleida bat lyssavirus (LLEBV), Mokola lyssavirus (MOKV), Rabies lyssavirus (RABV), Shimoni lyssavirus (SHIBV) and West Caucasian bat lyssavirus (WCBV). These 16 viruses are accepted by the International Committee on Taxonomy of Viruses ([Fig. 1](#)). Two further novel lyssaviruses, Kotalahti bat lyssavirus (KBLV) and Taiwanese Bat Lyssavirus (TBLV), have been reported and group genetically within the lyssavirus genus but remain as tentative species until fully characterised.

Virion Structure

All lyssaviruses exhibit a bullet-like virion morphology with an envelope derived from the plasma membrane of the infected host cell. One end of the bullet shape is conical, the other flat with approximate dimensions of ~75 nm and a length of ~180 nm, depending on the virus.

Genome

The lyssavirus genome consists of a single negative stranded (–) RNA molecule, consisting of approximately 12,000 nucleotides that encode 5 viral proteins with a conserved gene order; nucleocapsid (N), phosphoprotein (P), matrix (M), glycoprotein (G) and large (L) polymerase protein. The N, M and L proteins are conserved in structure and length across each of the lyssaviruses, whilst the P and G proteins vary in length. The conserved gene order has been proposed to have a role in the control of protein expression. The genomes of lyssaviruses are encapsidated in their entirety with the N protein and as such the viral RNA is protected from host cell based degradation. This encapsidated state also means that replicative mechanisms that might result in genetic recombination are considered to be unlikely although they have been reported in isolated studies.

Life Cycle

In the host, lyssaviruses infect peripheral nerves, often via muscle cells, at the motor endplate of neuromuscular junctions. The virus binds to a cellular receptor and enters endosomal transport vesicles via receptor-mediated endocytosis. Several cellular receptors have been proposed including the nicotinic acetylcholine receptor (nAChR), neuronal cell adhesion molecule (NCAM) and the p75 neurotrophin receptor (p75NTR). Once in neurons the virus travels in a retrograde fashion utilising microtubules

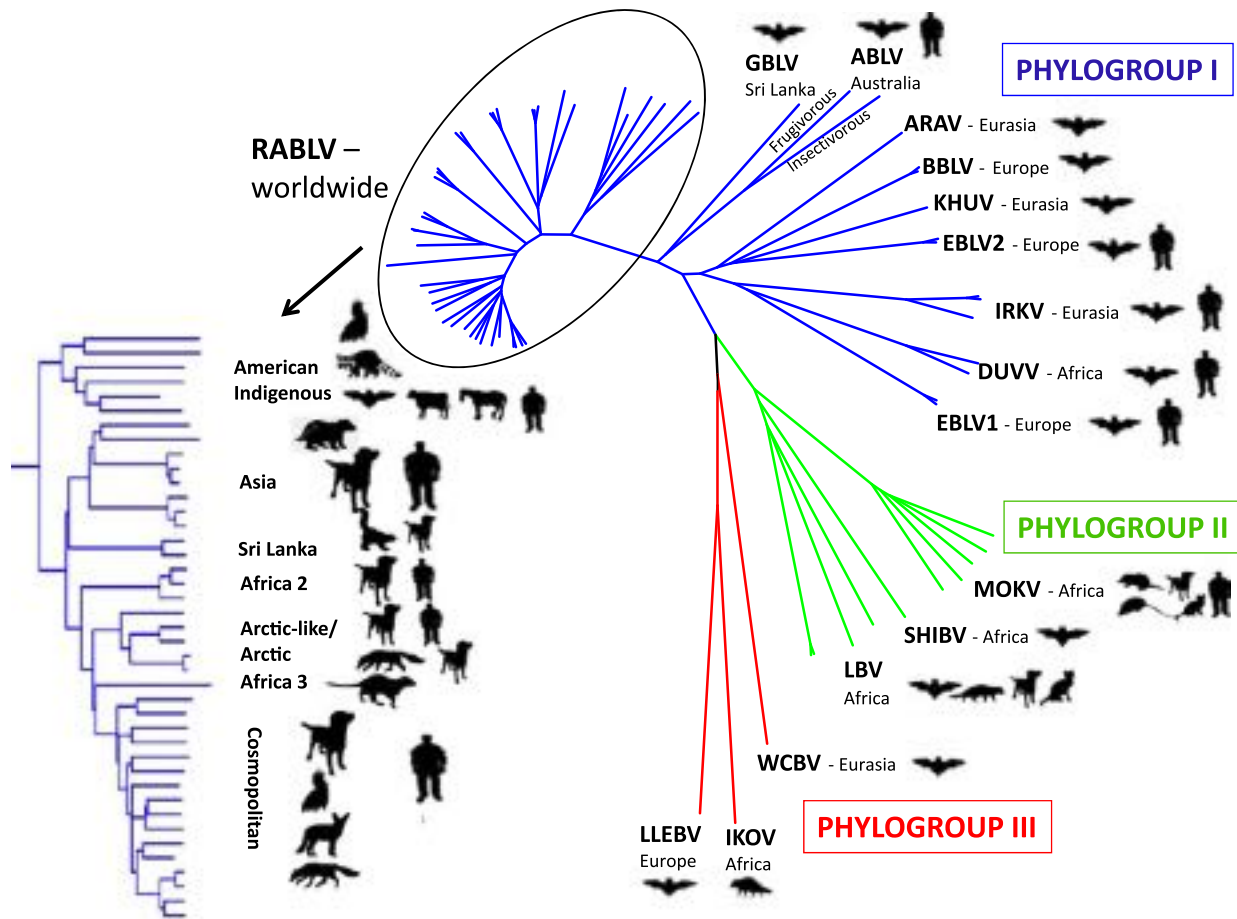


Fig. 1 Phylogenetic analysis of globally circulating lyssaviruses.

along axonal processes. Primary transcription occurs, leading to the production of viral proteins, after ribonucleoproteins (RNPs), a viral unit consisting of the viral RNA, N and P are released into the cytoplasm of the infected cell. Primary transcription initiates when the polymerase complex binds to the genome promoter at the 3' proximal end of the infecting genome. N, P and L interact in as yet undefined ratios to initiate primary transcription. As the transcriptase complex processes nascent mRNA it stutters at distinct intergenic boundaries, with reinitiation of transcription only occurring through binding at the 3' terminus of the negative sense genome. This results in a transcriptional gradient whereby the 3' proximal genes are produced to the highest levels with downstream genes being expressed at consecutively lower levels as the genome is traversed by the transcriptase complex. Following round of primary transcription, at some point during infection, the viral transcriptase complex switches to a replicative mode that enables the viral machinery to transcribe through the intergenic control regions present at each gene boundary to generate a full length antigenome positive sense (+) RNA. This positive sense RNA is a replicative intermediate that is fully encapsidated by N and serves as the template for nascent negative sense genome RNA synthesis. This secondary transcription occurs where viral RNA synthesis occurs in inclusion bodies (IBs), sites where viral proteins accumulate. As new RNPs are formed within IBs, they are transported to post-synaptic membranes where the nascent virions are assembled and transmitted in a process that depends on the presence of viral G on the budding membrane. Moving transynaptically through the peripheral nervous system the virus eventually enters the central nervous system (CNS) and once in the brain, massive viral replication occurs causing clinical disease. From the brain the virus then spreads centrifugally being excreted in saliva facilitating the onward transmission of the virus.

Epidemiology

Human rabies is a neglected zoonosis, especially in countries where the virus is endemic such as across Africa and Asia. In endemic areas, it is estimated that a human fatality from rabies occurs every 10–20 min, with 40%–50% of the victims being children less than 15 years of age. Children are thought to be disproportionately affected because of their small size and frequent interactions with domestic or free-roaming dogs. In endemic countries, it has been proposed that rabies be considered a neglected paediatric disease. The estimated annual 59,000 deaths worldwide do not capture the true burden of this disease. A profound lack of surveillance due to limited health care services, including an absence of laboratory capabilities and both medical and veterinary

infrastructures, results in under-reporting, which is further confounded by cultural, religious and social taboos. Furthermore, human rabies can present with symptoms that are common to other clinically indistinguishable conditions (for example, Guillain-Barré syndrome and cerebral malaria) and as such, in the absence of laboratory confirmation of disease, rabies is likely misdiagnosed. Consequent inaccuracies in reporting the burden of disease means that rabies is often not considered a high priority disease in endemic settings.

Lyssaviruses are capable of infecting all mammalian species. The development of serological techniques and later genetic methods of virus characterisation led to the understanding that multiple lyssaviruses exist that have been characterised as 16 distinct viral species. Whilst the archetypal lyssavirus, RABV, is most commonly associated with dogs, its evolution is thought to originate in bats and most lyssavirus species are now thought to be maintained by Chiropteran hosts with few exceptions. Where multiple isolates have been described, the chiropteran hosts and geographical range are relatively well documented. However a number of lyssaviruses have been characterised following only a single isolation and so their epidemiology and true association with different bat species remains speculative. It is likely that RABV evolved from a bat-associated progenitor and that numerous cross-species transmission events from bats to terrestrial carnivores resulted in the distribution of closely related but genetically distinct classical rabies viruses described. Across the globe, seven primary lineages of RABV have been described and within each lineage, numerous variants often described as causing clinical disease within a particular mammalian host across a certain geographical range, have emerged. The most widespread lineage has been described across several continents including Europe, the Americas, Africa and Asia, being maintained by both dog populations as well as by several terrestrial wildlife species including foxes, raccoon dogs, skunks and jackals. Classical rabies is found regularly in bat species across the Americas, with over 40 species of bat being reported as potential transmitters of disease to the human and terrestrial carnivore populations. From an economic perspective the vampire bat rabies lineage causes rabies within livestock, as well as the human population, and as such is considered of great economic significance in areas of vampire bat habitation.

Genetic analyses of RABV across the globe have demonstrated that natural landscape features such as mountain ranges and waterways are frequent barriers to epizootic spread. Man-made roads and human mediated movement of animals, often accompanying large-scale human migrations can, however, facilitate disease dispersion. Habitat fragmentation may also affect the epidemiology of the disease and spread of the virus. Genetic characterisation has also demonstrated where human translocation of animals has introduced rabies lineages in to new communities of both human and animals. The latter is particularly evident where the virus has then established a cycle of infection within a new geographical location. The use of vaccination campaigns, parenteral for domestic dog populations, and oral for wildlife populations has eliminated dog rabies across Europe, the United States and Canada and substantially reduced its prevalence in Latin America, with a resulting decrease in incidence of human rabies in these areas.

Within Europe there are six lyssaviruses, all are considered pathogenic and capable of causing the disease, rabies: the European Bat Lyssaviruses (EBLV-1 and EBLV-2) that circulate in Chiropteran hosts, the latter being endemic within UK bat populations; BBLV that has been detected in Natterer's bats in Germany and France; Kotohlahti Bat lyssavirus for which only genetic evidence has been reported in a Brandt's bat in Finland; WCBV isolated from a Bent Winged bat in the Caucasus mountains and LLEBV isolated in Bent winged bats in both Spain and France. WCBV and LLEBV are both genetically, highly divergent from all other European lyssaviruses. Alongside European viruses, three Eurasian viruses have been characterised, all isolated from bats; Aravan Virus (ARAV) in Kyrgyzstan, Khujand (KHUV) Virus isolated in Tajikistan and Irkut Virus (IRKV) was isolated in Eastern Siberia. Across the rest of the Old World, several other lyssaviruses have been described including within Africa: LBV isolated from a number of different bat species across sub-Saharan Africa; MOKV which is prevalent across Africa and has been associated with two recorded human cases; DUVV, that was initially isolated from a fatal human bat bite case in Kenya in 1971 and subsequently from a number of fruit bats in South Africa; and SHIBV that was first isolated in 2009 from a Commerson's leaf-nosed bat in Kenya. Most recently, IKOV was isolated from an African Civet in Tanzania in 2010 and although discovered in a terrestrial carnivore is believed to be of bat origin. Within Australasia, only Australian Bat Lyssavirus (ABLV) has been reported and has been isolated from five different bat species since its initial isolation in 1996. In Asia only GBLV has been reported in Sri Lanka in Fruit bats in 2014 and 2015.

Clinical Features

Human and animal rabies are often characterised into two classical forms: furious (often termed encephalitic) and paralytic. The factors determining the progression into either form are not understood, and whilst some specific symptoms are often reported, case definition can only generally be concluded during the neurological phase (Fig. 2). Incubation periods vary, although most patients develop symptoms between 20 and 90 days following exposure, however, there have been reports of extreme cases of incubation periods ranging between 14 and 19 years although the possibility of a secondary infection cannot be eliminated in such reports. Regardless, it is clear that the incubation period following infection can vary greatly depending on a range of factors. Due to the neurotropic nature of the virus, a bite/site of infection that is highly innervated (e.g., face and hands) can result in a reduced incubation period or faster progression of disease than following infection at poorly innervated sites. Alongside this, the presence/absence of permissive receptors in the tissue surrounding the infection site can affect viral entry and spread, hence increasing the incubation period. The dose of infective virus also plays a role in determining incubation period as does the immune status and age of an individual infected. Certainly, incubation periods in paediatric cases are shorter than those in adults, presumably due to their

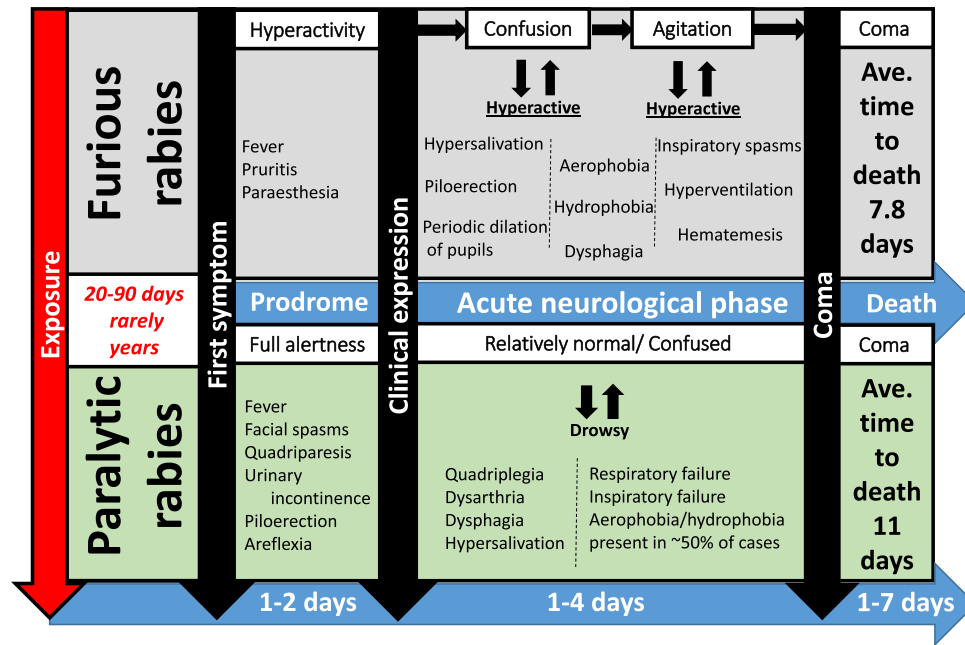


Fig. 2 Clinical disease progression in rabies.

smaller stature. Incubation periods following exposure to other lyssavirus are very poorly defined as exposure times are often difficult to determine. In particular it is recognised that often bat bites may go totally unnoticed leading to cryptic cases of rabies. A case of EBLV-1 infection in a human recorded an incubation period of 45 days. Further to this, a case of EBLV-2 infection in a bat handler reported a 19 week incubation period, however, the patient had a documented history of bat bites which may have confounded determination of the incubation period. DUVV infection has been reported to have an incubation period of around 4 weeks whereas ABLV may have a much longer incubation period as a case in 1998 reported onset of clinical disease 20 months after exposure to a bat. Within the chiropteran host, even more protracted prodromes have been described with a bat in a rehabilitation center developing clinical disease 9 months after being received and testing positive for EBLV-2.

From a human perspective, initial symptoms occur as the virus initiates replication in the dorsal root ganglia of neurons at the bite site. These symptoms can include pain, paraesthesia and pruritus which then develop to the neurological phase often characterised by more pronounced clinical signs. Where furious rabies develops clinical disease can include transient agitation, hypersalivation and hydrophobia whilst the paralytic form of disease can manifest with muscle weakness leading to paralysis. Regardless, both forms result in to coma and death with no treatments identified that can halt the progress of disease once clinical signs have developed.

Pathogenesis

The most common route of infection with a lyssavirus is via the mechanistic action of a bite from an infected animal which contains live virus in its saliva. Historically, this mechanism was first described 200 years ago by experimental work in 1804 by Zinke. Alongside this early observation, it was later determined that these viruses are unable to cross an intact dermal barrier. However, if the skin is broken, the risk of infection is greatly increased. Lyssaviruses can also infect hosts via mucous membranes including the nasal lining, oral cavity, conjunctivae, external genital organs and the anus. Rare reports of airborne transmission have not been scientifically corroborated. Human-to-human transmission has been reported very rarely with transmission of bodily fluids being implicated as the source of virus. Alongside this, there have been occurrences of infection through transplantation of infected tissue. A case in 2004 resulted in three rabies fatalities when an infected donor supplied a liver section, a kidney and a section of artery to three separate recipients. Within 30 days of transplantation all three recipients had returned to the hospital with rapidly progressing neurological disease, which was diagnosed *post mortem* as rabies infection. A further case in Germany resulted in the deaths of three individuals following organ donation and corneal transplantation has also facilitated infection in the recipient following surgery.

Whilst it is appreciated that lyssaviruses are highly neurotropic, the mechanism behind viral entry of neurons at peripheral inoculation sites remains largely undefined. The location of nicotinic acetylcholine receptors (nAChR) in the vicinity of post-synaptic membranes on muscle cells may support infection prior to neuronal invasion. It is then thought that virus replication in the musculature and consequent virus egress into synaptic clefts may facilitate neuroinvasion via receptors at pre-synaptic axonal membranes. Once inside a cell the virus encounters host immune effector mechanisms that attempt to

clear the infection. Critically, RABV virulence resides in the virus's ability to evade the localised immune response alongside retaining the viability of the neurons it has invaded following infection. The demonstration of immune effector cells within the CNS have suggested that RABV replication in the CNS occurs in the face of a fully immunocompetent cellular environment although it has long been stated that the immunocompromised nature of the CNS enables ease of virus replication. From a host response perspective, experimental models have demonstrated that wildtype and fixed rabies strains induce different inflammatory markers both in the periphery and within the CNS. Critically, virus clearance from the brain, that predominantly correlates with an increase in B cell infiltration and antibody production in the CNS, is reduced during infection with pathogenic RABV isolates. This supports the concept that adaptive immune responses are impaired in natural RABV infections. Further, blood–brain barrier permeability to the infiltration of immune cells is important for successful virus clearance, with enhanced permeability being associated with an increase in viral clearance of attenuated RABV in experimental models. Canine models of infection have suggested that in line with the pathogenicity observed, the immune response in the CNS is limited with wildtype RABV infection having a minimal effect on the permeability of the blood brain barrier. Certainly, permeability of the blood–brain barrier is of vital importance for viral clearance. Interestingly, virus-neutralizing antibodies (VNAs) have been reported in sera and cerebral spinal fluid during the clinical stages of human infection but appear unable to clear virus alone. Patients that develop neutralizing antibodies early during the disease course are considered potentially favorable candidates for aggressive treatment of the disease. Interestingly, almost all reported human rabies survivors were diagnosed on the basis of neutralizing antibodies in CSF or serum with no other viral products being detected in submitted samples. Immune effectors appear restricted by elements of the viral protein repertoire including poor induction of host type I IFN responses, apoptosis of induced antigen-activated T-cells and reduced T cell proliferation and apoptosis. A further mechanism by which RABV may replicate alongside minimal pathological changes has been proposed through M protein influenced autophagy. This may enable preservation of brain and spinal cord function despite abundant viral replication. Apoptotic responses appear to be common following experimental infection with attenuated viruses.

Diagnosis

Ante-mortem laboratory tests can only confirm infection: a negative *ante-mortem* test result cannot be used to rule out infection. Notably, no clinical feature is pathognomonic for rabies. Hydrophobia or aerophobia (fear of drafts or fresh air) provide a strong indication of RABV infection but laboratory diagnosis is always required to confirm rabies. A number of diagnostic tests are available to diagnose RABV infection but the use of any given test depends on the laboratory capability of the local infrastructure to support diagnostic testing (Fig. 3). Typically, rabies diagnosis is performed on *post mortem* material. However, where available, clinical rabies can be diagnosed *ante-mortem* with testing of either nuchal (nape of the neck) skin biopsies or saliva enabling detection of viral RNA.

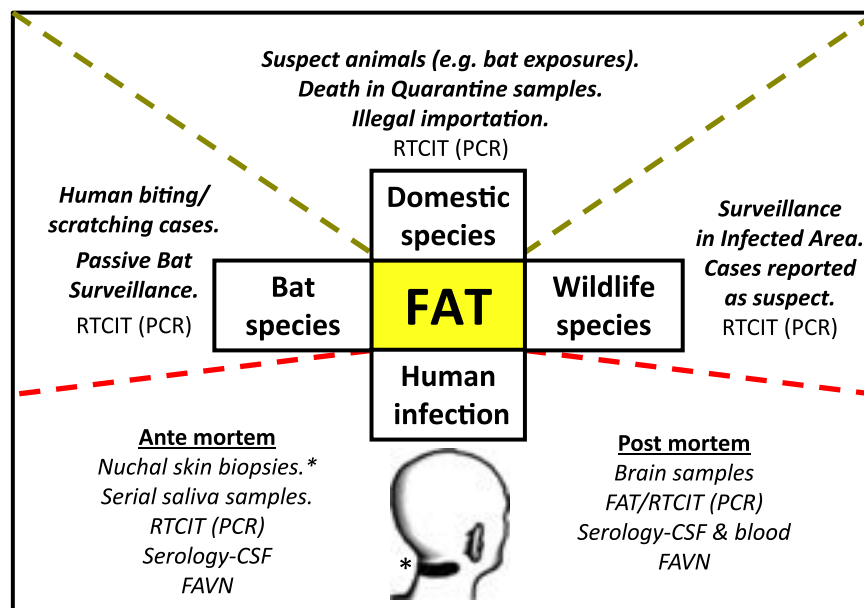


Fig. 3 Rabies diagnostic testing in laboratories with capacity to perform recommended tests. Samples are submitted from a variety of sources. The frontline test on post mortem material is the FAT. The RTCIT, and where unavailable the MIT, are used as secondary confirmatory tests. In many laboratories, molecular tools such as PCR are used as a more rapid secondary test platform.

Live virus has been occasionally isolated from human saliva and salivary excretion can be detected before the onset of clinical disease in experimental canine models. Molecular testing using RT-PCR has proven to be a rapid, highly sensitive and specific tool for ante-mortem diagnosis. However, testing saliva samples should involve testing multiple samples as intermittent secretion of the virus into this fluid means that negative results does not prove absence of infection.

The invariably fatal nature of rabies infection means that almost all laboratory confirmations of rabies are made *post mortem*. Samples of brain tissue are optimal for *post mortem* diagnosis with the fluorescent antibody test (FAT) detecting virus antigen in brain material as the diagnostic test recommended by both the WHO and the OIE. Efficient diagnosis by FAT depends on the quantity and quality of the sample in terms of autolysis and decomposition. In addition, the requirement of expensive equipment and secondary antibody conjugate to detect virus antigen does not allow to use FAT in resource limited settings where, in most cases, the virus is endemic. The development of the direct "rabies immunoperoxidase test" (dRIT) has overcome the problems associated with conducting FAT in resource limited settings. Other antigen detection tests, including lateral flow devices (LFDs) have been developed but require further development and validation before being useful as diagnostic for an invariably fatal disease.

Molecular tools, such as reverse transcription-polymerase chain reaction (RT-PCR) have been developed that have many advantages over traditional tests and, through the highly sensitive and specific detection of viral nucleic acid, are slowly replacing classical antigen based assays. One-step closed tube RT-PCR systems have been developed that dramatically reduce the risk of contamination. RT-PCR assays that can detect RNA of many lyssaviruses are available and are replacing the exiting confirmatory tests for rabies diagnosis in animals. In comparison to the FAT, RT-PCR methods are less prone to subjective interpretation and have the added benefit that they can identify the virus species, following sequencing of the derived amplicon, a useful tool in cases with undefined history of potential exposure events. Molecular assays are now recognised by the OIE; however, these tests need to be performed in well controlled environments to avoid false positive results and, as a consequence, the standardization of the assays and utilization of appropriate controls are required before these tools can be used for routine *post mortem* diagnosis of rabies in animals in resource-limited settings (WHO, 2013). Most *post mortem* diagnoses are currently based on the previously recommended techniques including FAT, tissue culture isolation and mouse inoculation tests: adoption of new reliable and validated tools (dRIT, RT-PCR) have improved diagnostic options for rabies diagnosis in endemic areas where empirical data are lacking and the burden of the disease is disproportionately high.

The use of serological assays for detection of a current rabies infection is generally limited to post-vaccinal assessment of serostatus. The OIE recommends the rapid fluorescent focus inhibition test (RFFIT) and the fluorescent antibody virus neutralisation (FAVN) test for monitoring of antibodies to allow both international travel in companion animals and evaluation of human vaccination. Both of these tests measure the ability of antibodies in suspect sera to neutralise live virus. Once the virus is applied to the sera, fluorescent antibody staining is used to detect any virus that has not been neutralised by the sera. In order to quantify the virus neutralizing antibody, standardised reference sera from the WHO and OIE are used as a comparison. These tests are widely considered as the gold standard in serological assays but they have some logistical restrictions as the use of live virus requires containment facilities which may not be available in resource limited settings. This has lead to efforts to develop serological assays that do not require the use of live virus.

Novel enzyme-linked immunosorbent assay (ELISA) techniques are available that are capable of the detection of rabies specific antibodies in serum from both humans and animals. However, the current ELISAs have a major drawback in that they are unable to conclusively show the presence of neutralizing antibodies within a sample.

Treatment

No tools exist to prevent the clinical phase of lyssavirus infection and so the development of antivirals for rabies remains an urgent priority. Prevention of rabies disease is best achieved through multiple activities including: preventing transmission from dogs to humans via dog vaccination campaigns; promoting awareness of rabies to avoid exposure; ensuring availability of post exposure rabies immunoglobulin and vaccine; and encouraging responsible dog ownership. It is anticipated that dog-mediated human rabies will be confined to history with the goal of eliminating transmission of the virus between dogs and humans targeted for 2030. For this target to be achieved improved diagnostic tools, improved veterinary and medical infrastructures, cheaper and efficacious biologicals (vaccines, new antivirals and HRIG), and increased education and awareness of rabies is required. Where disease has developed the outlook for treating clinical rabies remains poor. Despite reports of promising molecules from reports of *in vitro* studies, following attempts with the same molecules *in vivo*, no molecules have shown sufficient promise to warrant further investigation. However, developing novel antiviral treatments remains a major challenge as both viral factors, including the neurotropic nature of the virus, and host factors, such as variable stages of presentation, exposure history and immune response, preclude progress. Any candidate molecules need to be able to cross the blood-brain barrier, disrupt and potentially halt virus replication without negatively impacting on host neurons and also inhibit detrimental host responses to RABV infection. Strategies reported have included ultrasound and microbubbles, which have been used to try and increase permeability of the BBB to facilitate the entry of therapeutics to the optimal site for action. Other tools, such as small interfering RNA (siRNA) or microRNAs (miRNA) that bind to and target RNA species for degradation within the cells have demonstrated some suppressive activity on virus replication *in vitro*. Some broad spectrum antiviral molecules including Favipiravir (6-fluoro-3-hydroxy-2-pyrazinecarboxamide) have demonstrated an *in vitro* effect although again *in vivo* studies have shown limited use of these molecules. Other

delivery methods including stable nucleic acid lipid particles and nanoengineered particles with si-RNAs or miRNAs are being assessed for utility as rabies antiviral molecules. The urgent need for rabies antiviral treatments means that if any benefit can be demonstrated with drugs licensed for human use in experimental models, there is good rationale for assessing them in human rabies cases where consent is given.

Prevention

Despite the high number of annual human deaths from rabies, the disease can be readily prevented with the use of vaccines. Rabies vaccines have been available since 1885, when Louis Pasteur established that immunisation with desiccated nerve tissue from an infected rabbit prevented disease in a young boy who had been attacked by a rabid dog. Since the dawn of vaccination, safe and efficacious rabies vaccines have been developed. Human rabies vaccines are administered intramuscularly or intradermally with the primary goal of vaccination being the induction of virus neutralizing antibodies directed against the viral G protein. This response provides protection against RABV infection although the location and mechanisms by which neutralizing antibodies prevent viral replication remain unknown. Human rabies vaccines are based on inactivated purified viruses and as such are poorly immunogenic and often require three doses given within a 4-week period to reliably induce protective titers of neutralizing antibodies. The development of inactivated, lyophilized, cell culture-based vaccines with increased antigenicity has led to a reduction in the number of doses needed to elicit a sufficient neutralizing antibody response. This in turn has reduced the number of clinic visits required and has improved complete vaccine administration through increased patient uptake. All licensed rabies vaccines have superb safety records and adverse post-vaccination events are rarely reported.

As with most other vaccine formulations for infectious diseases, pre-exposure immunisation (PrEP) using standard vaccines is recommended for travellers to rabies endemic areas. The invariably fatal nature of the disease and the strong association in endemic areas with infection of children under 15 years of age has led some to suggest that PrEP should be considered as part of a recommended paediatric vaccination strategy. PrEP regimens consist of four vaccine doses being administered on days 0, 7, 14 and 21 or 28. In endemic areas, economic limitations restrict the availability of vaccines and as such extensive PrEP coverage is not generally considered possible. As such post exposure prophylaxis (PEP) is generally relied upon following a potential exposure event. Rabies is unusual in that, even following a potential exposure event, vaccination is effective and can prevent the development of disease as long as its administered in a timely manner. However, before vaccination is considered, primary wound care including the irrigation of wounds with soapy water; thorough scrubbing of affected areas to remove foreign bodies; and the use of iodine based antiseptics can significantly reduce the likelihood of establishing a productive infection. Suturing the wounds should be avoided. Then, where available, and individuals are aware of its utility, PEP can be sought. PEP administration is guided by WHO recommendations that define responsive activities depending on the degree of exposure with category I and II exposures warranting vaccination alone whilst severe exposures, termed category III, requiring rabies immunoglobulin (RIG) alongside vaccination. In exposed unvaccinated individuals, the severity of exposure dictates if the vaccine has to be combined with a rabies immunoglobulin (RIG). The application of RIG directly to the wound serves to neutralise any live virus present in the wound area and prevent any further live virus infection of exposed cells and prevents spread of the virus in the time before vaccination elicits sufficient immunity, generally considered to be up to 14 days. RIG is generated from both human vaccinees (HRIG) and vaccinated horses (Equine- ERIG). RIG administration is based on weight with 20 IU/kg for HRIG and 40 IU/kg for ERIG being recommended by the WHO. It is also suggested that all of the calculated dose of RIG should be infiltrated into wounds, with any residual RIG should be injected intramuscularly at a distant site to the wound. Certainly, whilst ERIG is safe, easier to produce than HRIG, effective and increasingly available in rabies endemic countries, the search for alternatives continues. To try and extend the use of available RIG in endemic regions studies have looked at their application to wounds alone, with any residual product being reserved for future patients rather than being injected at distant sites. Vaccination is performed after RIG administration at a site distant to the wound itself. Vaccination for PEP can be time-consuming and costly, requiring multiple clinic visits and studies have tried to reduce PEP vaccination regimens to one week, making PEP more affordable and accessible. In cases of licks or minor scratches, several vaccination regimens are approved for intramuscular PEP and a two site 4-dose intradermal regimen has been adapted in some regions. PEP should be administered rapidly following a potential exposure to increase the probability of preventing clinical disease. Although the clinical disease development is almost 100% preventable with timely PEP, failures most often due to alterations in PEP administration, have been reported. The majority of exposed individuals do not receive PEP when required or they are administered PEP without RIG. Both vaccines and RIG are costly and in some countries patients often are expected to pay for required biologicals, medical supplies and treatment alongside covering all costs associated with travel to health care facilities.

The role of vaccines in protection against RABV is well defined with correlates of protection against the virus, in the form of neutralizing antibody titres of or above 0.5 International Units (IU), being defined as protective. Serological responses to lyssaviruses are targeted towards the trimeric lyssavirus G protein, the primary antigen on the surface of the virion that is responsible for cell receptor binding and entry. However the divergence of lyssaviruses has led to the categorisation of viruses within the genus into three phylogroups based on their reactivity against sera from vaccinated individuals (see below). Differences within G give an indication of the antigenic diversity within the genus although the lack of an atomic structure for this protein does not allow to identify the key antigenic determinants at the amino acid residue level. Human rabies vaccines

are formulated from inactivated preparations of whole virus and it is widely accepted that a protective titer of 0.5 international units (IU)/ml is sufficient for protection against RABV. Serum from vaccinated individuals has been utilised to determine cross neutralisation of different lyssavirus species demonstrating that vaccine derived sera does not neutralise all of the lyssavirus species. Phylogroups broadly define the relationship between existing vaccines and neutralisation with phylogroup I viruses (RABV/ABLV/ARAV/BBLV/DUVV/EBLV-1/EBLV-2/GBLV/IRKV/KHUV) being neutralised whilst phylogroup II (LBV/MOK/SHIBV) and phylogroup III (LLEBV/IKOV/WBCV) viruses failing to be neutralised. Still, the epidemiology of these divergent lyssaviruses is ill defined and only 13 human fatalities have been directly attributed to infection with lyssaviruses other than classical rabies virus as such the burden is undefined. Attempts are being made to increase the valency of vaccine preparations to enable cross lyssavirus protective responses to be induced although exact requirements for any pan-lyssavirus vaccine remain to be defined.

Conclusions

Aims to reduce the human burden from rabies are ongoing with the OIE targeting the elimination of dog-mediated rabies by 2030. A first steps toward human rabies elimination is to ensure a globally available diagnostic capability to accurately confirm infection with rabies or a lyssavirus. This requires laboratory networks capable of performing diagnosis across resource limited regions where the virus is endemic. In the future, approaches to rabies diagnosis should combine laboratory confirmation with point-of-care testing in the field. Studies are ongoing to develop novel diagnostic tools. Alongside this, endemic regions need to ensure that rabies is a notifiable disease to ensure that laboratory networks and government bodies develop the infrastructure to accurately diagnose and report the disease. Several pathways to rabies prevention have been developed through the WHO, OIE and other international organisations.

Elimination of RABV from dogs populations requires extensive in country veterinary capabilities to enable a sustained approach of vaccination of companion animals, and control and vaccination of free-roaming and community dogs. Estimated costs required to eliminate dog-mediated rabies in all endemic countries have been estimated at US\$6.3 billion with most human rabies cases being prevented through control, licensing and vaccination of dog populations in endemic regions. Alongside parenteral vaccination of dogs, the oral vaccination of free-roaming dog populations has been proposed although assessment of vaccine coverage using oral formulations remains difficult. Lessons learned from extensive and highly successful oral vaccination campaigns to protect wildlife species has demonstrated difficulties in achieving adequate seroconversion even though significant successes have been made.

From a prevention perspective, clearly adoption of rabies vaccination as part of a paediatric regimen in endemic areas would reduce the human burden of disease significantly. The existing recommendation is that PEP, in the form of vaccination alone, is still given to exposed individuals who received PrEP. Hence, there are still costs associated with exposure events but the requirement for RIG would reduce dramatically. RIG constitutes the highest cost in PEP and therefore a vaccine only approach would make a significant difference in the costs associated with rabies. The number of animal bites considered to be potential exposure events is often high in regions where rabies is endemic. It is estimated that PEP is given to > 15 million individuals annually to prevent the establishment of disease. Optimization of awareness to ensure exposed individuals practice thorough wound cleaning and development of vaccines to reduce production costs and doses required could improve prospects for exposed individuals in endemic areas drastically. Despite all the challenges, the 2030 eradication goal will promote global initiatives to encourage international organisations, governments, charities, non-governmental bodies and potentially pharmaceutical companies to develop strategies to achieve global human rabies elimination.

Further Reading

- Abela-Ridder, B., Knopf, L., Martin, S., *et al.*, 2016. The beginning of the end of rabies? *The Lancet Global Health* 4, e780–e781.
- Banyard, A.C., Fooks, A.R., 2017. The impact of novel lyssavirus discovery. *Microbiology Australia* 38, 18–21.
- Bourhy, H., Reynes, J.-M., Dunham, E.J., *et al.*, 2008. The origin and phylogeography of dog rabies virus. *Journal of General Virology* 89, 2673–2681.
- Dodet, B., Korejwo, J., Briggs, D.J., 2013. Eliminating the scourge of dog-transmitted rabies. *Vaccine* 31, 1359.
- Fooks, A.R., Cliquet, F., Finke, S., *et al.*, 2017. Rabies. *Nature Reviews Disease Primers* 3.
- Hampson, K., Coudeville, L., Lembo, T., *et al.*, P., 2015. Estimating the global burden of endemic canine rabies. *PLoS Neglected Tropical Diseases* 9, e0003709.
- Hemachudha, T., Ugolini, G., Wacharapluesadee, S., *et al.*, 2013. Human rabies: Neuropathogenesis, diagnosis, and management. *The Lancet Neurology* 12, 498–513.
- Johnson, N., Cunningham, A.F., Fooks, A.R., 2010. The immune response to rabies virus infection and vaccination. *Vaccine* 28, 3896–3901.
- Lafon, M., 2005. Rabies virus receptors. *Journal of Neurovirology* 11, 82–87.
- Mittrabhakdi, E., Shuangshoti, S., Wannakrairot, P., *et al.*, 2005. Difference in neuropathogenetic mechanisms in human furious and paralytic rabies. *Journal of the Neurological Sciences* 238, 3–10.
- Rupprecht, C., Kuzmin, I., Meslin, F., 2017. Lyssaviruses and rabies: Current conundrums, concerns, contradictions and controversies. *F1000Research* 6.
- Rupprecht, C.E., Nagarajan, T., Ertl, H., 2016. Current status and development of vaccines and other biologics for human rabies prevention. *Expert Review of Vaccines* 15, 1–19.
- Schnell, M.J., Mcgettigan, J.P., Wirblich, C., Papaneri, A., 2010. The cell biology of rabies virus: Using stealth to reach the brain. *Nature Reviews Microbiology* 8, 51–61.
- Wallace, R.M., Undurraga, E.A., Blanton, J.D., Cleaton, J., Franka, R., 2017. Elimination of dog-mediated human rabies deaths by 2030: Needs assessment and alternatives for progress based on dog vaccination. *Frontiers in Veterinary Science* 4.
- WHO, 2013. WHO Expert Consultation on Rabies, Second Report. WHO Technical Report Series No. 982. Geneva: World Health Organization.

Relevant Websites

www.rabiesalliance.org

Global Alliance for Rabies Control: Home.

www.missionrabies.com

Mission Rabies – Taking up the fight against rabies in India.

http://web.oie.int/boutique/index.php?page=ficprod&id_produit=1641&lang=en

OIE standards and guidelines related to trade and poultry diseases.

www.oie.int/animal-health-in-the-world/rabies-portal

Rabies Portal: OIE – World Organisation for Animal Health.

<http://www.who.int/rabies/resources/who-cds-ntd-nzd-2018.03/en/>

WHO Laboratory Techniques in Rabies.

www.who.int/rabies/en

WHO – Rabies – World Health Organization.

Respiratory Syncytial Virus (*Pneumoviridae*)

Tiffany King, Nationwide Children's Hospital, Columbus, OH, United States and The Ohio State University College of Medicine, Columbus, OH, United States

Tiffany Jenkins, Nationwide Children's Hospital, Columbus, OH, United States and The Ohio State University, Columbus, OH, United States

Supranee Chaiwatpongsakorn, Nationwide Children's Hospital, Columbus, OH, United States

Mark E Peeples, Nationwide Children's Hospital, Columbus, OH, United States and The Ohio State University College of Medicine, Columbus, OH, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of Tracy J. Ruckwardt, Peter L. Collins, Barney S. Graham, Human Respiratory Syncytial Virus, In Reference Module in Biomedical Sciences, Elsevier Inc, 2018, doi:[10.1016/B978-0-12-801238-3.02599-X](https://doi.org/10.1016/B978-0-12-801238-3.02599-X).

Nomenclature

Metapneumovirus Proper name for the human and avian viruses most closely related to the orthopneumoviruses

Mononegavirales Order of negative strand virus

Orthopneumovirus Proper name of human, bovine and ovine RSV and pneumonia virus of mice

Paramyxoviridae Family that included the Pneumoviridae before 2016

Pneumoviridae Family that includes orthopneumoviruses and the metapneumoviruses

Glossary

Antigenome Positive sense single strand copy of the genome, used the template for producing genomes.

F protein Fusion glycoprotein.

G protein Attachment glycoprotein.

Genome Negative single strand RNA encodes all the viral mRNAs and the positive strand antigenome.

L protein Large RNA-dependent RNA polymerase protein.

M2-1 protein Small protein that enables the polymerase to transcribe mRNA without releasing from the nucleocapsid.

M2-2 protein Small protein that inhibits replication at the expense of replication.

Minigenome System launched from plasmids that replaces all the viral genes with a marker gene that is used to study viral replication and mRNA transcription.

N protein Nucleocapsid protein.

NS1 protein Inhibits the initial interferon response.

NS2 protein Inhibits the amplification of the interferon response.

Nucleocapsid Helical complex of the genome or antigenome covered by the nucleocapsid protein.

P protein phosphoprotein interacts with the nucleocapsid and the L protein.

Post-F Structure of the F protein after it is triggered to refold.

Pre-F Structure of the F protein before it is triggered to refold, causing membrane fusion.

Replicon Self-replicating RSV genome that lacks the RSV F and G genes and so cannot produce virions.

RSV Respiratory syncytial virus

SH protein Small hydrophobic glycoprotein.

History

Chimpanzee coryza agent was first recognized in 1955 in chimpanzees with respiratory symptoms housed in an animal facility and in one of the animal handlers. Nasal samples transmitted the disease to other chimpanzees and caused cytopathology in cell culture. Human adolescents and young adults had serum antibodies to the agent. The virus was soon isolated and characterized. Human respiratory syncytial virus (RSV) is now recognized as the most important cause of lower respiratory tract infection in infants and young children and among the most frequent causes for hospitalizations in childhood. With the advent of highly sensitive RT-PCR assays, RSV has also been recognized as a major cause of "influenza-like illness" in the elderly, responsible for approximately one-third of the excess deaths each winter that had been traditionally attributed to influenza (Falsey *et al.*, 1992; Walsh *et al.*, 2004).

RSV Disease

RSV spreads through aerosolized droplets and contaminated surfaces, and typically requiring close contact with an infected individual. On average, the incubation period lasts 3–5 days before clinical signs become apparent: upper respiratory tract infection include rhinorrhea, sneezing, coughing, and occasionally a low-grade fever. Some 25%–40% of RSV infections in young children progress to acute lower respiratory tract infections (ALRI), in most cases bronchiolitis and/or pneumonia. Children with

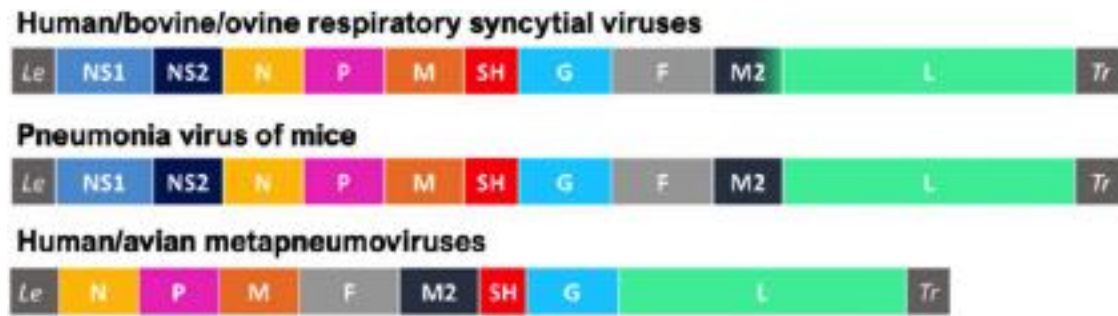


Fig. 1 *Pneumoviridae* genome organization.

ALRI may develop signs of tachypnea, inspiratory crackles, and expiratory wheezing. Recovery occurs 7–12 days after the onset of clinical symptoms except in severe cases, in which cough and wheeze may progress to dyspnea and hypoxia. The disease can be lethal when the infant experiences exhaustion from increased and persistent respiratory effort, resulting in respiratory failure. The decision to hospitalize a child is often based on the presence of a hypoxemic state. Treatments may involve oxygen supplementation and supportive care but severe cases may require mechanical ventilation.

Fatal RSV infections are rare in the developed world. They involve peribronchiolar infiltrates consisting of lymphocytes, plasma cells, and macrophages, with neutrophils comprising a larger portion of the inflammation early in the course of infection. There is often necrosis and sloughing of bronchiolar epithelial cells into the bronchiolar lumen. Obstruction can occur in the small bronchioles caused by the combination of enhanced mucus production, sloughed epithelial cells and inflammation. Despite the middle name of the virus, syncytia are rarely observed histologically in cases of human RSV but are a prominent feature of bovine RSV (bRSV) infections. T cell deficient individuals may develop a “giant cell pneumonia”, where syncytia formation is more readily apparent, but this is not a feature in immunocompetent individuals. Infection with RSV and RSV-induced ALRI early in life has been associated with the development of asthma in early adolescence (Schwarze *et al.*, 1997; Sigurs *et al.*, 1995; Welliver and Duffy, 1993).

Infections of infants and young children with RSV are the most frequent cause of hospitalization for children, responsible for over 570,000 annual hospitalizations and a major economic burden in the USA. Globally, fatal RSV infections are estimated to be 48,000–75,000 per year. However, the vast majority of these RSV-related deaths occur in developing countries, where ready access to medical care is not readily available. Worldwide, RSV is second only to malaria as a leading cause of death in the first year of life.

Throughout life, older children and adults are infected multiple times with RSV, developing symptoms of a common cold (Hall *et al.*, 1991). The inability to prevent subsequent infections is likely due to the inability of the virus to induce a strong memory immune response. In individuals who are immunosuppressed or 65 and older with weakening immune systems, RSV can be life threatening. RSV rivals influenza viruses each winter season as a cause of “excess deaths” with as many as 14,000 deaths in the USA, often in elderly individuals with underlying health problems.

Classification

Nonsegmented negative sense viruses like RSV, with an RNA genome and envelope are included in the order Mononegavirales. The RSV genome is approximately 15.2 kb in length with 10 gene units separated by intergenic regions (IR) of variable lengths. The genome is protected in a helical nucleocapsid structure formed by the nucleocapsid (N) protein. The *Pneumoviridae* family (Fig. 1) includes the orthopneumovirus genus, human, bovine and ovine RSV, as well as the related pneumonia virus of mice. The metapneumovirus genus includes the more distantly related human and avian metapneumoviruses. The avian metapneumoviruses are divided into three types, A, B and C. A and B are primarily European and C is primarily North American. Human metapneumovirus is most highly related to avian metapneumovirus C, the North American strain, suggesting that the transition from birds to humans took place in America.

All members of the *Pneumoviridae* infect the respiratory tract only, causing respiratory disease. They also share many of the same genes, including N, phosphoprotein (P), matrix (M), small hydrophobic (SH), attachment glycoprotein (G), fusion (F), the large (L) polymerase, and the M2 gene, encoding two proteins. The M2-1 and M2-2 proteins enable transcription and regulate replication, respectively. The nonstructural, NS1 and NS2, genes are found in the orthopneumoviruses, but not in the metapneumoviruses (Fig. 1). The source of these nonstructural genes is not clear as no other virus has similar genes and no obviously similar genes have been found in other organisms.

Interestingly, the F-M2 gene block follows the SH/G gene block in pneumoviruses but precedes the SH/G gene in metapneumoviruses. It apparently switched position in the divergence between the two groups, a change that would require recombination. Unlike positive strand viruses that replicate their RNA genomes outside the context of a nucleocapsid and have high rates of recombination due to polymerase copy choice directed by hybridization, recombination is exceedingly rare among the Mononegavirales, likely because the nucleocapsid covers both the genome and antigenome RNAs preventing hybridization from nascent RNA/polymerase complexes. But despite the rarity of such a recombination event, the appearance of the heavily

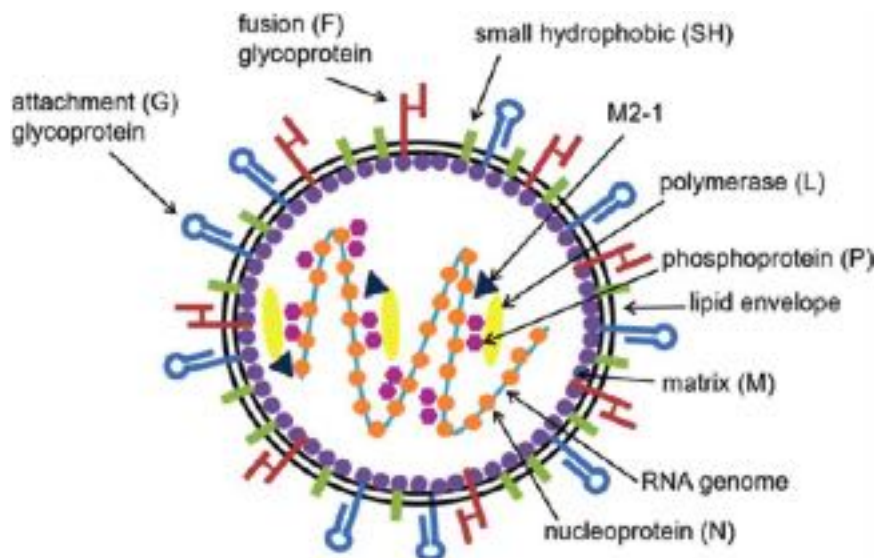


Fig. 2 RSV virion organization. Modified from the dissertation of Sara Johnson, The Ohio State University, 2015.

glycosylated G protein in the *Pneumovirinae* likely represents another such rare recombination event. The G protein is related to the highly glycosylated mucins that decorate the surface of the respiratory epithelium, suggesting that an ancestral *Pneumoviridae* virus acquired this gene from the cells in which RSV replicates.

The members of the *Pneumoviridae* have many similarities with the members of the *Paramyxoviridae*, with whom they were classified until recently. The two families share their N, P, M, F, and L proteins. A few paramyxoviruses have a small hydrophobic (SH) protein, but most do not, and none of them have homologs to the nonstructural proteins, NS1 and NS2, or M2 proteins. Furthermore, the *Pneumoviridae* attachment protein, G, is completely different from the hemagglutinin/neuraminidase (HN) attachment protein of most members of the *Paramyxoviridae*, which bind to sialic acid as its receptor and cleave sialic acid from its terminal position on glycans, releasing these viruses from their receptor.

Virus Entry

The Virion

The RSV virion incorporates 8 of the 11 viral proteins the virus produces. As shown in [Fig. 2](#), the nucleocapsid structure containing the RSV genome is wrapped by N protein and associated with the L/P polymerase complex and M2-1 protein. That ribonucleoprotein complex is surrounded by a lipid membrane. The M protein underlies the membrane, and the three viral glycoproteins are incorporated into the membrane. The shape of the RSV virion can be spherical, but is more often rod-shaped.

The RSV virion infects cells by binding to a receptor on the cell surface primarily with its G protein, and its F protein causes the virion membrane to fuse with the cell membrane. Once the membranes fuse, the nucleocapsid with its associated polymerase, is spilled into the cytoplasm where the L/P polymerase complex with help from the M2-1 protein produce 10 viral mRNAs that are translated into 11 viral proteins. The L/P polymerase complex replicates the RSV genome via an antigenome intermediate that is also encapsidated by the N protein in a helical nucleocapsid. The three viral glycoproteins are produced in the ER where the F and G are decorated with N-linked glycans that are processed as they move through the Golgi complex. In addition, the G protein is decorated with ~35 O-linked glycans in the trans-Golgi. Both the G and F proteins arrive on the cell surface where they are joined by the M protein underlying the membrane, the nucleocapsid with its L/P/M2-1 polymerase complex. New virions are formed by budding through the plasma membrane. Most virions produced in cell lines (~90%) remain associated with the cells, whereas in primary well differentiated human bronchial epithelial (HBE) cultures grow at the air-liquid interface and form tight junctions, they are released efficiently and apically.

Historically, most studies on RSV replication have been performed in HEP-2, HeLa or Vero cells, immortal cell lines that grow well in the laboratory as submerged cultures. However, RSV attachment to these cells, particularly the receptor recognized by the G protein, is different from its receptor in HBE cultures, an ideal model for the airway epithelium where RSV specifically infects the ciliated cells.

The F Glycoprotein

RSV enters a target cell by fusing its membrane with the membrane of the cell, a process mediated by the fusion (F) protein. The F protein is a highly conserved, type I transmembrane glycoprotein trimer. It is produced as a precursor, F0, in the ER where it is also

modified by the addition of 5 N-linked glycans. During passage through the Golgi these glycans are matured, and the F0 protein is cleaved, activating its fusion function. The transmembrane part of the protein, F1, is linked by two disulfide bonds to F2. Like many other class I viral fusion proteins, the *Pneumoviridae* F protein is a homotrimer and each monomer is activated by proteolytic cleavage, orthomyxoviruses by a furin-like protease. However, unlike the other *Pneumoviridae* F proteins, the RSV F protein is cleaved at two sites releasing a 27 amino acid peptide known as pep27. Pep27 does not appear to remain associated with the F protein. Following cleavage, the F monomers assemble into functional trimers. The pep27 of human RSV has no known function, but the pep27 of the bovine RSV functions as a tachykinin-like peptide, promoting smooth muscle contraction and perhaps mediating other proinflammatory responses.

The RSV F protein arrives at the cell surface as a fully function, metastable prefusion F (pre-F) trimer (Fig. 3, left) that undergoes major structural changes to accomplish membrane fusion. The process of refolding into the postfusion F (post-F) protein (Fig. 3, right) causes membrane fusion, thereby initiating virus infection. It is not clear what triggers refolding, but when pre-F is triggered, much of the top part of the molecule undergoes dramatic refolding, almost doubling the length of the F protein. The apical surface of the pre-F trimer contains 4 short α -helices and two β -sheets but during the refolding process the loops between these α -helices and β -sheets, as well as the β -sheets themselves, transform into one very long α -helix that is tipped with the highly hydrophobic fusion peptide at the N-terminus of F1 that had been revealed by cleavage in the trans-Golgi.

The three long α -helices of the trimer form a coiled coil and their three fusion peptides insert into the target cell membrane thereby linking the viral membrane with the target cell membrane. The F protein trimer then folds in half, bringing its N-terminal and C-terminal regions together and allowing the three C-terminal α -helices to lock in place in the grooves of the long N-terminal helix to form the very stable “6-helix bundle” (6HB) of the post-F structure. In so doing, the virion and plasma membranes are brought together and their lipids begin to mix, initiating the membrane fusion process. In traditional laboratory cultured cell lines, the F protein on the infected cell surface enables that cell to fuse with neighboring cells, resulting in the classical multinucleated giant cell or ‘syncytial’ pathology that gives RSV its middle name.

Six major antigenic sites have been identified on the surface of the pre-F protein by monoclonal antibody binding. Two of these sites at the top of pre-F are lost during the transition from pre-F to post-F (Fig. 3, red and orange) because these are the regions that refold dramatically following triggering. The remaining antigenic sites are present, to varying degrees on both pre-F and post-F. No antigenic site is uniquely post-F. All these observations are consistent with the rearrangement of the top part of the pre-F protein, while the lower portion of the pre-F protein remains largely intact. The potency of neutralizing monoclonal antibodies follows the same trend: antibodies to the epitopes unique to the pre-F form are, approximately 10 times more efficient at neutralizing the virus than the antibodies that bind both pre-F and post-F. This much higher potency likely reflects interference with triggering or refolding of this dynamic region of the pre-F protein, thereby blocking its ability to cause fusion and initiate infection.

It is not clear what triggers the RSV pre-F protein to refold, leading to membrane fusion. The RSV pre-F protein does not require its attachment protein to trigger and cause fusion. All paramyxovirus pre-F proteins do. While the G protein is not essential for infection in vivo, the F protein is capable of mediating at least some attachment and infection in immortalized laboratory cells, ~10-fold less efficiently than in the presence of G, likely because of the much more efficient cell binding ability of the G protein that would position the F protein close to the cell membrane. Several proteins, including nucleolin, EGFR, ICAM-1 and TLR4, interact with the F protein and could be receptors for pre-F. Binding to one of these receptors might be the trigger that initiates pre-F refolding. It is also possible that contact with a more common molecule or structure on the plasma membrane of a target cell such as a charged glycan or a lipid membrane triggers the F protein.

While most evidence points to the plasma membrane as the point of RSV entry, there are reports that RSV is taken into a cell by micropinocytosis and fusing with intracellular membranes. Low ionic conditions have been shown to trigger the structural conversion of pre-F to post-F and initiate cell to cell fusion. Low ionic strength conditions might be provided in intracellular vesicles, enabling entry at that site.

The G Glycoprotein

The G glycoprotein is the primary attachment protein of RSV. It is a type II transmembrane protein that is rich in Ser, Thr and Pro, and highly modified by the addition of glycans, similar to the mucins in the respiratory tract where RSV resides. The initial G protein molecular weight is 32 kDa but its final form in immortalized cells is 90–100 kDa. The first modifications, palmitoylation and N-linked glycosylation, occur during translation in the endoplasmic reticulum resulting in a 45 kDa species. Palmitoylation modifies cysteines within the cytoplasmic N-terminal domain while N-linked glycosylation occurs in the two hypervariable mucin-like regions of the ectodomain. Once in the Golgi, O-linked glycans are added to these mucin-like regions, resulting in a fully glycosylated 90–100 kDa glycoprotein. The range in its size is likely due to minor variations in the many attached glycan chains. Finally, the G protein reaches the plasma membrane where it is incorporated into budding virions.

The N and O-linked glycosylation sites reside in two large mucin-like domains that flank a conserved central region and cysteine noose. Three to 5 N-linked glycans are added, depending on the RSV strain and ~35 O-linked glycans modify the many Ser and Thr residues in the mucin-like domains. The G protein produced in HBE cultures is much larger, 180 kDa, glycoprotein, when evaluated under reducing conditions. It is not yet clear if this larger size is due to non-disulfide, covalent dimerization or to additional post-translational modifications.

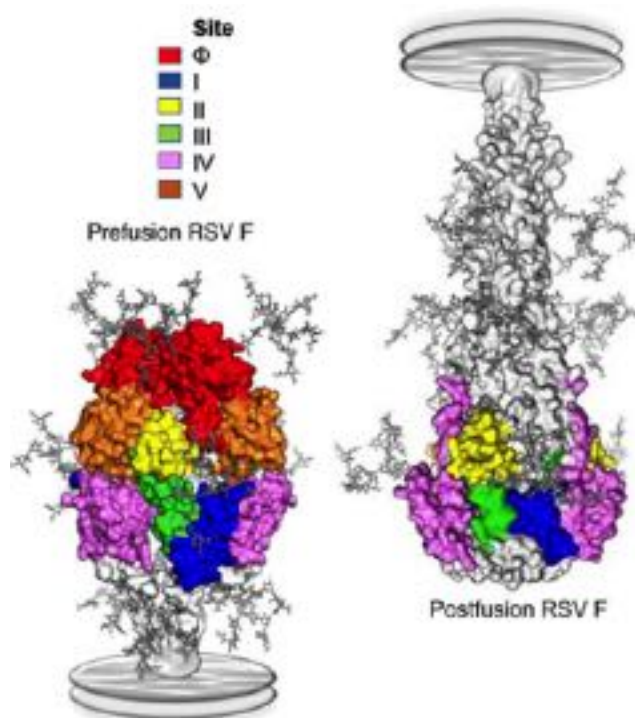


Fig. 3 RSV F protein trimer with its antigenic sites indicated on both its pre-F and post-F forms. Pre-F is the metastable, functional F protein which is anchored in the virion membrane. Post-F is the post-triggered, refolded F protein that brought the virion and cell membranes together and caused them to fuse. Note that the blue, purple, yellow and green antigenic sites remain relatively intact following refolding while the orange and red sites disappear, refolding into long helices that form the core of the long stalk of post-F that trimerize and insert into the target cell membrane via their hydrophobic N terminus. The stalk of the pre-F protein then swings upwards to contribute its three helices to the post-F stalk, forming the extremely stable 6-helix bundle. Modified from Graham, B.S., 2017. Vaccine development for respiratory syncytial virus. *Current Opinion in Virology* 23, 107–112.

Unlike the F glycoprotein whose sequence is highly conserved (88%), the G glycoprotein is highly variable, with as little as 48% conservation between RSV subgroups. A portion of the C-terminal hypervariable region of the G glycoprotein has been used to define two antigenic subgroups, A and B, and further divisions into genotypes that are useful for epidemiology. The A and B subgroups have not diverged enough to be recognized as serotypes because antibodies raised against one subgroup are able to neutralize viruses from the others subgroup. However, cross-subgroup neutralization is not as potent as intra-subgroup neutralization, indicating that these differences could be important in protection.

The only region that is completely conserved in the G protein is a 13 amino acid central conserved domain which is followed by a mostly conserved sequence containing 4 completely conserved cysteines that form a “cysteine noose” (Fig. 4). The C-terminal side of this noose includes a CXXXC (CX3C) motif. Only one other CX3C motif is found in Genbank and that belongs fractalkine, a chemokine which interacts with its receptor CX3CR1 on immune cells to induce chemotaxis. G protein has, likewise, been shown to induce chemotaxis in immune cells through its interaction with CX3CR1. CX3CR1 has also been detected on the ciliated cells in primary HBE cultures and may be the receptor for RSV on ciliated cells, its target cell. It is not clear how this interaction would occur because the disulfide bonding pattern of the cysteine noose of G differs from that of fractalkine resulting in a different structure for the CX3C in the G protein.

A generally conserved region of the G glycoprotein, C-terminal to the cysteine noose, is a stretch of positively charged amino acids. In immortalized cell lines, these residues interact with the heparan sulfate attached to transmembrane proteoglycans (HSPG), which serve as the receptor for RSV infection in these cells. Heparan sulfate is an unbranched polysaccharide chain of a repeating disaccharide, hexuronic acid and hexosamine. The polysaccharide chain is modified by the addition of sulfate groups, giving the heparan sulfate a negative charge. It is the charge–charge interaction between the cluster of positively charged amino acids in the G protein with the negatively charged sulfate groups on the HSPG that enables RSV to attach to these cells. Many other viruses that infect immortalized cells also use this “non-specific” mechanism to attach and initiate infection.

Some of the G glycoprotein produced during infection is released from infected cells in a soluble form. This “soluble G” is produced by initiating translation at the second AUG of the G protein, reaching the cell surface. The result is a G protein whose N-terminus includes part of the hydrophobic transmembrane domain and contains the full ectodomain of G. It appears to be released from the cell by cell surface proteases. Approximated 20% of the G glycoprotein expressed in cultured cells is in a soluble

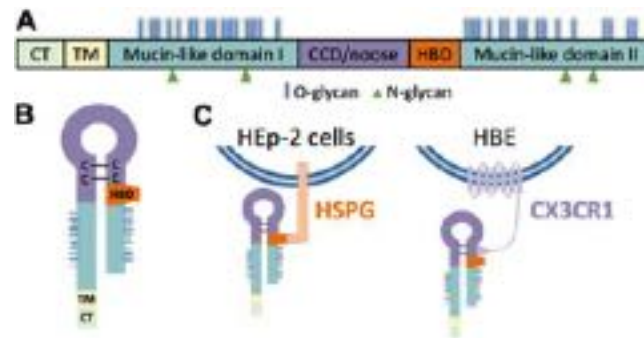


Fig. 4 The RSV G protein and its roles in attachment. (A) G protein domains and predicted glycosylation sites (4 N-linked, green triangles; ~35 O-linked, blue horizontal lines). (B) General 2D structure with two disulfide bonds creating a “cysteine noose”. (C) Attachment sites for heparan sulfate proteoglycans on immortalized cell lines and for CX3CR1 on primary well differentiated human bronchial epithelial cultures.

form. Soluble G may play an immunomodulatory role via its interaction with the CX3CR1 receptor on immune cells. Soluble G could also be a decoy for neutralizing antibodies, although most neutralizing antibodies (90%–99%) are against the F protein, not the G protein. However, it remains possible that non-neutralizing antibodies to the G protein may suppress RSV infection in another manner such as antibody dependent cellular cytotoxicity or complement mediated lysis.

As described above, in immortalized cells G attaches to the negatively charged heparan sulfate proteoglycans (HSPG) to initiate infection via a stretch of positively charged residues C-terminal to the cysteine noose. Soluble HS interferes with this interaction, inhibiting RSV infection of immortalized cells but not infection of HBE cultures, suggesting that RSV uses a different receptor *in vivo*. HSGP is not detectable on the apical surface of the ciliated cells that RSV infects. CX3CR1, a chemokine receptor which interacts with the CX3C motif is found on the apical surface of ciliated cells in HBE cultures. A G-specific monoclonal antibody, 131-2G, which binds to G in the region of the CX3C motif does not prevent RSV from infecting immortalized cells, but it does prevent RSV from infecting HBE cultures. These findings confirm that RSV uses different receptors in standard laboratory cells compared to primary HBE cultures and are consistent with CX3CR1 being a receptor for RSV *in vivo*.

Genome Transcription and Replication

The result of fusion between the virion membrane and the cell membrane is that the RSV nucleocapsid enters the cytoplasm of a target cell, its associated polymerase complex produces viral mRNAs, antigenomic and genomic RNAs in localized regions that grow into viral inclusion bodies in the cytoplasm. Viral genes are transcribed into mRNAs for each gene in order, from the 3' to 5' end of the genome. As the polymerase reaches a gene end (GE) sequence, it repeatedly copies the poly U tract in the GE, generating a poly A tail before releasing the mRNA. The polymerase then scans through the intergenic region (IR), 1–52 nt long. Some of the polymerase molecules release while crossing the IR, creating a gradient of reduced gene expression for each subsequent gene.

Eventually a sufficient amount of N protein is produced to encapsidate antigenome transcripts, producing full-length antigenome nucleocapsids. Finally, encapsidated antigenomes are copied many times, beginning at the ‘trailer complement’ sequence at the extreme 3' end of the antigenome to produce full-length genomes that are encapsidated. These new encapsidated genomes either reenter the replication cycle or are incorporated into new virions during the budding process at the plasma membrane. Deletion of the M2-2 gene from the RSV genome results in enhanced mRNA transcription and reduced genome replication indicating that the M2-2 protein encourages the opposite, genome replication, but the mechanism is not known.

The template for mRNA transcription and replication is the genome-nucleocapsid complex, not free genomic RNA. The nucleocapsid is composed of either the genomic or the antigenomic RNA and over 2000 N proteins in a long helical structure that looks like a string of vertebrae. One “vertebra” represents one revolution of the helix of 10 N proteins, each holding 7 nucleotides of genomic RNA in the groove between its N- and C-terminal domains.

The viral L (large) protein is an RNA-dependent RNA polymerase that is complexed with a tetramer of the P (phosphoprotein). This L/P complex transcribes the 10 mRNAs and the antigenome from the genome template, and the genome from the antigenome template. The determination of whether the L/P complex will produce a full-length viral antigenome or the series of 10 viral mRNAs takes place in the leader sequence, the promoter at the extreme 3' end of the genome. The leader sequence contains two overlapping viral promoters, one for transcription and one for replication. To produce mRNA, the L protein, initiates transcription at the third and fourth nucleotides of the leader sequence. Starting at that position rather than the first and second nucleotides of the leader results in the transcription of a 25 nt “leader transcript” that terminates and is released, allowing the L/P complex to recognize the GS sequence and begin the process of transcribing the mRNAs.

The L polymerase protein is a doughnut-shaped structure that complexes with a tetrameric P protein. To transcribe the genomic RNA into complementary mRNA, the C-terminal domains of the P proteins each pull an N protein off the RNA template, allowing the RNA to pass through the polymerase, followed by the N proteins being replaced. The L/P protein complex remains associated

with the nucleocapsid, scanning until it finds the first GS signal where it reinitiates transcription, this time producing the NS1 mRNA. The M2-1 protein associates with the L/P complex involved in mRNA transcription, functioning as an “anti-termination” factor that prevents premature termination of viral mRNA transcripts. The lengthening transcript passes through the polyribonucleotidyltransferase (capping enzyme) domain on the opposite side of the L protein, where a cap is added to its 5′ end.

The next step in mRNA processing is the addition of two methyl groups, one to the cap and one to the first nt of the mRNA by the methyltransferase domain of the L protein. This portion of the L protein has not yet been visualized. The addition of these methyl groups is important to make the viral mRNAs appear to be cellular mRNAs and avoid the induction of interferon. Finally, the L/P/M2-1 polymerase slips on the poly-U tract of the GE signal, repeatedly copying it to produce the poly A tail and eventually releases the NS1 mRNA.

The polymerase continues along the nucleocapsid template by sliding through the IR and re-engaging at the next, NS2, GS sequence where it transcribes, caps, methylates, polyadenylates, and releases NS2 and each of the downstream viral mRNAs in sequence. During transit through each IR, some of the polymerase molecules fall off the template such that the level of transcription of each downstream gene is less than the one before it. Another less frequent occurrence is that the polymerase misses the GE sequence, continuing transcription through to the following GE. The result is a bicistronic “read-through” transcript. These read-through events may help more polymerase molecules reach the downstream genes.

If instead of initiating transcription at the third nucleotide, the L/P complex initiates transcription at the first two nucleotides, the resulting full-length antigenome is produced. If the concentration of N protein in the local area is high enough, as the complementary “antigenome” RNA strand exits the polymerase, it is encapsidated by newly synthesized N proteins, guided into position by the P protein. Once antigenome nucleocapsids are produced, they are used as the template to produce copies of the genomic RNA that are also encapsidated. The trailer complement promoter on the antigenome appears to be stronger than the leader promoter because the number of genomes in an infected cell greatly outnumber the number of antigenomes.

All but one of the RSV genes encode a single protein. The M2 gene encodes two proteins, M2-1 and M2-2, which are translated from the same mRNA by initiating translation at two different sites and in different, but partially overlapping reading frames. The M2-L gene junction is unique among the junctions in RSV. The GS for the L gene precedes the GE of the M2 gene. To transcribe the L gene, the polymerase must terminate the M2 mRNA at its GE and backtrack by retrograde scanning to find the GS site for the L gene to initiate its transcription. This type of GE and GS placement is used by other negative strand viruses, such as Ebola, at many of its gene junctions.

Non-Structural Viral Proteins

The NS1 and NS2 proteins are not present in virions but are produced in the infected cell. The NS1 protein plays an important, but as yet undefined role in RSV replication. Both NS1 and NS2 proteins have roles in efficiently evading innate immune responses. NS1 interferes with the induction of interferon by preventing the interaction between RIG-I and MAVS. NS2 interferes with the amplification phase of interferon signaling by cleaving the STAT-2 protein that is involved in transmitting the signal from the interferon receptor to the nucleus to activate antiviral genes and enhance interferon production. The M2-2 protein is also non-structural. As discussed above, its expression enhances viral genome replication at the expense of mRNA production.

The Third Viral Glycoprotein

The SH protein is the third viral glycoprotein, along with F and G, is expressed on the virus surface. Although SH is not required for RSV replication *in vitro* or *in vivo*, two distinct functions have been identified for SH. First, the SH protein appears to be a “viroporin”, a virus-produced membrane channel, that forms pentameric structures within membranes increasing permeability to ions. Second, SH has an immune modulatory role that protects infected cells from TNF- α mediated apoptosis. SH inhibits TNF- α by: (1) inhibiting of TNF- α production by infected cells; and (2) inhibiting of downstream TNF- α signaling that would have led to phosphorylation of p65 by NF- κ B. It remains unclear if these functions are related and their significance to RSV pathogenesis.

Animal Models of RSV

Human RSV is host-specific and replicates poorly in most other species, which is thought to be due to the inability of the virus to impede the interferon response in species other than humans. The paucity of animal models that accurately recapitulate human RSV infection and disease has been a hindrance in the understanding of RSV pathogenesis and immunity.

Chimpanzees are the only non-human species in which RSV causes respiratory symptoms and have been used in the past to test vaccine candidates. However, chimpanzees are no longer available for such studies in the United States and many other parts of the world. Other non-human primates such as African green monkeys and Rhesus macaques can be infected with RSV but demonstrate few if any symptoms or pathology.

The most common small animal model used to study RSV is the mouse, particularly BALB/c. A murine model lacking signal transducer and activator transcription 1 (STAT1), and thus an interferon-response, is more susceptible to RSV infection. Attempts to develop a “mouse-adapted” strain of RSV through serial passage have been unsuccessful, with the virus failing to gain the ability

to replicate to high titers even after numerous passages in vivo. Hamsters, guinea pigs, ferrets, and cotton rats have all been used as alternatives to mice for RSV studies because they are more susceptible to RSV.

The cotton rat (*Sigmodon hispidus*) has become one of the most widely-used and accepted rodent models for RSV investigation, and is approximately 100-fold more permissive for RSV infection than the BALB/c mouse. The cotton rat was used as a model to study formalin-inactivated RSV (FI-RSV) vaccine-enhanced disease because it develops a pulmonary disease that more-closely reflects of the disease in humans. The predictive quality of the cotton rat is so high that the FDA-approved prophylaxis for RSV, palivizumab (*Synagis*), proceeded to clinical trials based on the safety and efficacy studies in the cotton rat model alone, bypassing non-human primate studies.

Alternative animal models for RSV research include the highly related bovine RSV (bRSV) that naturally causes respiratory infection in calves, resulting in tachypnea, oculonasal discharge, hypoxemia, and fever. The bovine respiratory tract offers more parallels to the human respiratory system than the murine model, as both human and bovine respiratory tracts are lined predominately by ciliated epithelium, contain abundant submucosal glands, and have both palatine and nasopharyngeal tonsils, all of which are features not present in the mouse respiratory tract. Alternatively, pneumonia virus of mice (PVM) is a orthopneumovirus that naturally infects and induces severe respiratory disease in mice. However, PVM is not as closely-related to human RSV as is bRSV.

Protection From RSV Disease

Vaccination – The First Attempt

In the mid-1960s the first attempt to develop an RSV vaccine was launched. Using the “Salk” approach that had been very successful for poliovirus vaccination, RSV was grown in cell culture and inactivated with formalin. Children were vaccinated before RSV season with no ill effects. However, after natural RSV infection, 80% of the vaccinated children were hospitalized with enhanced RSV disease (ERD) and two children died. Only 5% of the children in the control group who had received parainfluenza virus vaccine were hospitalized and none experienced ERD (Kim *et al.*, 1969). The cause has not been clearly identified but evidence for an enhanced Th2 response, the response to cell culture contaminants, poorly neutralizing antibodies that recognized the post-F rather than pre-F protein, and complement deposition in the airways have all been presented as potential explanations.

Passive Immunity and Prophylactic Therapy

Maternal antibodies (Abs) are transmitted to the fetus in utero from approximately 26 weeks of gestation up until birth, and levels of these Abs in infants directly correlate to Ab titers in the mother. However, the half-life of antibodies is 21–26 days and maternal Abs usually become undetectable by 4 months of life. Infants born prematurely receive fewer transplacental Abs and are therefore at a higher risk for severe hRSV infection. Injection with a humanized monoclonal Ab, palivizumab (*Synagis*), has been used to supplement maternally derived Abs in infants born prematurely or with other conditions that make them more susceptible to RSV infection. However, palivizumab levels also wane with time and must be replenished once a month for 5 months to maintain an effective neutralizing titer. The cost of this treatment is high and so is not used for all at risk infants.

Palivizumab binds to site 2 on both the pre-F and post-F protein (Fig. 3) and neutralizes RSV. Currently, a monoclonal Ab is being tested in infants that uniquely recognizes a pre-F site (Fig. 3), neutralizing RSV approximately 10 times more efficiently than palivizumab. In addition, this pre-F Ab contains several mutations in its heavy chain that triples its half-life. It may, therefore, be possible to protect infants during their most vulnerable first winter season with a single injection before the RSV season (November–March).

Maternal Vaccination

Approximately 60% of severe cases of RSV occur in infants, less than 6 months old. Active vaccination during this time, especially during the first 4 months of life is not likely to be effective because the immune response of infants that young is poor. An alternative, immunization of pregnant women during their third semester, is currently being tested. Such a vaccine would likely protect the pregnant woman from RSV disease and allow RSV neutralizing antibodies to cross the placenta, supplement the mother’s own antibodies in protecting the infant after birth.

Because antibodies that uniquely recognize pre-F are the most effective for neutralizing RSV, pre-F would seem to be an ideal RSV vaccine, but pre-F is metastable, and readily converts to post-F. However, solving the structure of the pre-F protein enabled the prediction of mutations to stabilize it. Testing, selection, combining and retesting has resulted in several stabilized versions of the RSV F protein that are currently in clinical trials as a vaccine for pregnant women.

Live Attenuated Vaccines

RSV live attenuated vaccines (LAV) are replicating RSV viruses with mutations that weaken their ability to replicate and cause disease. This approach was very successful in developing the current vaccines against measles, mumps, rubella, chickenpox, and polio (Sabin) which are all given by intramuscular injection. Similarly, the live attenuated influenza virus vaccine that is given intranasally has been effective. Live attenuated RSV vaccines would also be given intranasally.

It has been difficult to find live RSV vaccine candidates that do not cause symptoms but are still protective. The Laboratory of Infectious Diseases at the National Institute of Allergy and Infectious Diseases (NIAID) in the USA has been leading the development of RSV LAVs. They have generated many vaccine candidates and tested 14 of them in sero-positive adults, sero-positive children and sero-negative children. Two of the promising candidate vaccines in phase 1 trials have now proceeded to phase 2 trials.

Live attenuated vaccines must be produced in one of three cell lines approved by the World Health Organization for that purpose. Of these, Vero, a cell line derived from African green monkey kidneys, produces the most RSV and is therefore the best choice for vaccine production. However, when RSV is produced in Vero cells, much of its ~95 kDa G protein is cleaved by cathepsin L resulting in a truncated, 55 kDa protein. This Vero-grown virus is 5-fold less infectious in HBE which would reduce the efficiency of vaccine production by 5-fold. However, the location of this cleavage site has been identified and mutated to prevent cleavage, allowing enhanced vaccine virus yields which should enhance manufacturing efficiency.

Vectored RSV Vaccines

Several RSV vaccine candidates are based on virus vectors that are non-replicating (adenovirus) or poorly replicating (Modified Vaccinia Ankara (MVA) and Sendai virus, a paramyxovirus of mice). All these vectors express the RSV F protein and in some cases a stabilized pre-F protein, and particularly to the pre-F protein, are much more neutralizing than antibodies to the G protein, at least by the standard neutralization assays. Some of these vectors also express the G protein or internal viral proteins such as N and M2 in order to induce T cell immunity capable of killing RSV-infected cells during an RSV infection.

Treatment of RSV: Antiviral Agents

Once an individual is infected with RSV, a vaccine is not useful. RSV infections might be halted by small molecule antiviral agents capable of inhibiting RSV replication. Attempts to develop such drugs have been mounted over the past 40 years but have not yet resulted in an approved product. Blind screening of chemical libraries for small molecules that inhibit RSV infection or cell-cell fusion resulted in compounds that bind to the F protein and prevent it from refolding. Although these fusion inhibitors showed promising early results, most have not reached clinical trials due to safety concerns and poor pharmacokinetic properties. However, one of them, lumicitabine is currently being evaluated in hospitalized adults in clinical trials. Other efforts have targeted the RSV polymerase, or L protein, including nucleoside analogs and non-nucleoside inhibitors. Multiple companies now have effective candidate drugs that inhibit RSV replication and are bioavailable. As with all acute infections, the difficulty is in diagnosing an RSV infection early enough for an inhibitor to have a significant effect on symptoms.

References

- Falsey, A.R., Treanor, J.J., Betts, R.F., Walsh, E.E., 1992. Viral respiratory infections in the institutionalized elderly: clinical and epidemiologic findings. *Journal of the American Geriatrics Society* 40, 115–119.
- Hall, C.B., Walsh, E.E., Long, C.E., Schnabel, K.C., 1991. Immunity to and frequency of reinfection with respiratory syncytial virus. *Journal of Infectious Diseases* 163, 693–698.
- Kim, H.W., Canchola, J.G., Brandt, C.D., *et al.*, 1969. Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. *American Journal of Epidemiology* 89, 422–434.
- Schwarze, J., Hamelmann, E., Bradley, K.L., Takeda, K., Gelfand, E.W., 1997. Respiratory syncytial virus infection results in airway hyperresponsiveness and enhanced airway sensitization to allergen. *Journal of Clinical Investigation* 100, 226–233.
- Sigurs, N., Bjarnason, R., Sigurbergsson, F., Kjellman, B., Bjorksten, B., 1995. Asthma and immunoglobulin E antibodies after respiratory syncytial virus bronchiolitis: A prospective cohort study with matched controls. *Pediatrics* 95, 500–505.
- Walsh, E.E., Peterson, D.R., Falsey, A.R., 2004. Risk factors for severe respiratory syncytial virus infection in elderly persons. *The Journal of infectious diseases* 189, 233–238.
- Welliver, R.C., Duffy, L., 1993. The relationship of RSV-specific immunoglobulin E antibody responses in infancy, recurrent wheezing, and pulmonary function at age 7–8 years. *Pediatric Pulmonology* 15, 19–27.

Further Reading

- Collins, P.L., Karron, R.A., 2013. Respiratory syncytial virus and metapneumovirus. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. vol. 1. Lippincott Williams & Wilkins: Philadelphia, pp. 1086–1123.

- Collins, P.L., Hill, M.G., Camargo, E., *et al.*, 1995. Production of infectious human respiratory syncytial virus from cloned cDNA confirms an essential role for the transcription elongation factor from the 5' proximal open reading frame of the M2 mRNA in gene expression and provides a capability for vaccine development. *Proceedings of the National Academy of Sciences of the United States of America* 92, 11563–11567.
- Johnson, S.M., McNally, B., Willette, M., *et al.*, 2015. Respiratory syncytial virus uses CX3CR1 as a receptor on primary human airway epithelial cultures. *PLOS Pathogens* 11, e1005318.
- McLellan, J.S., Chen, M., Joyce, M.G., *et al.*, 2013. Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science* 342, 592–598.
- Shi, T., McAllister, D.A., O'Brien, K.L., *et al.*, 2017. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: A systematic review and modelling study. *Lancet* 390 (10098), 946–958.

Relevant Websites

https://path.azureedge.net/media/documents/RSV-snapshot-2020_03_26_High_Resolution_PDF_1.pdf
RSV Vaccine and mAb Snapshot.

<https://www.cdc.gov/rsv/index.html>
RSV.
Home.
Respiratory Syncytial Virus.
CDC.

Rhinoviruses (*Picornaviridae*)

Matti Waris, University of Turku, Turku, Finland

Olli Ruuskanen, Turku University Hospital, Turku, Finland

© 2021 Elsevier Ltd. All rights reserved.

This is an update of N.W. Bartlett, S.L. Johnston, Rhinoviruses, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:10.1016/B978-0-12-801238-3.02658-1.

Nomenclature

AOM Acute otitis media

CDHR3 Cadherin related family member 3

CPE Cytopathic effect

cre Cis-acting replication element

C_t Cycle threshold

EV Enterovirus

ICAM-1 Intracellular adhesion molecule 1

IFN Interferon

IL Interleukin

IRES Internal ribosomal entry sites

LDLR Low-density lipoprotein receptor

RV Rhinovirus

T_m Melting temperature

UTR Untranslated region

VP Viral protein

Glossary

AOM Acute otitis media, infection of the middle ear.

CPE Cytopathic effect, microscopically detectable effect of viral infection in infected cell.

Cre Cis-acting replication element; the action on a genetic element on the activity elsewhere in the same nucleic acid molecule.

IRES Internal ribosomal entry sites, secondary structure on the 5' UTR of positive-sense RNA viruses for ribosome contact.

Myristoylation Covalent attachment of a derivative of the fatty acid myristic acid.

ORF Open reading frame, gene region that encodes a protein.

SYBR Green Cyanine dye forming a highly fluorescent complex with dsDNA.

Uridylation Post-translational addition of one or more uridines.

UTR Untranslated region, gene region not encoding a protein.

Classification

Rhinoviruses (RVs) form three species A, B, and C within the *Enterovirus* genus of *Picornaviridae* family. The current total number of types is 169 (Table 1). Initially, 100 RV serotypes were numbered as new serotypes were identified. Phylogenetic analysis divided them into two groups, RV-A and RV-B, and the types were assigned with their original serotype number, e.g., A1, A2, B3 ... B6, A7, etc. New genetically defined RV-A and RV-B types from number 100 on are assigned the next available number for the species in question, i.e. A100–A109 and B100–106. There are currently 80 types in RV-A and 32 in RV-B. Type A87 is missing since it was reclassified as enterovirus D68. Types A44, A95, A98 have also been deleted, because they were reassigned as variants of A29, A8, and A54, respectively. Analysis of clinical specimens positive for RV by RT-PCR but unculturable by traditional methods revealed a third rhinovirus species, type C. RV-C types are numbered starting from C1, and their current total is presently 57. As agreed by the International Committee on Taxonomy of Viruses in 2012, the sequence identity is > 70% within a species, and > 87% in VP1 gene or > 90% in VP2/VP4 gene region within a type. Current status of the picornavirus taxonomy including approved RV types can be found at “See Relevant Websites section”.

Virion Structure

Picornaviruses are small, non-enveloped viruses with an icosahedral protein capsid shielding an RNA genome. The diameter of the virus particle is about 30 nm. The outer part of the capsid is formed from 60 copies each of three proteins VP1, VP2, and VP3 in pseudo T = 3 symmetry (Fig. 1(A)). A further 60 copies of a fourth, internal capsid protein VP4 is connected with the externum. A mature virion is formed from 12 pentamers, each pentamer containing five protomers with one copy of each of the structural proteins. VP4 protein undergoes post-translational modification with a hydrophobic amino-terminal myristic acid. A micelle of 5 VP4 units, in contact with the internal N-terminus of VP1, reside under the vertex of the core pentamer. Between species, there are differences in the capsid surface topographies due to differences in VP tertiary structures (Fig. 1(B–C)).

Genome

Picornaviruses are single-stranded RNA viruses with positive polarity and 7.1–7.2 kb long genome. A single open reading frame (ORF) is flanked by untranslated regions (UTRs) with a polyA tail at the 3' end and a small viral protein of the genome (VPg) at the

Table 1 Rhinovirus species and types. Group refers to receptor usage (see life cycle)

Species	Group	Types
RV-A	Major	A7–13, A15–22, A24, A28, A32–34, A36, A38–41, A43, A45, A46, A50, A51, A53–61, A63–68, A71, A73–82, A85, A88–90, A94, A96, A100–109
RV-A	Minor	A1, A2, A23, A25, A29–31, A47, A49, A62
RV-B	–	B3–6, B14, B17, B26, B27, B35, B37, B48, B52, B69, B70, B72, B79, B83, B84, B86, B91–93, B7, B99, B100–106
RV-C	–	C1–57

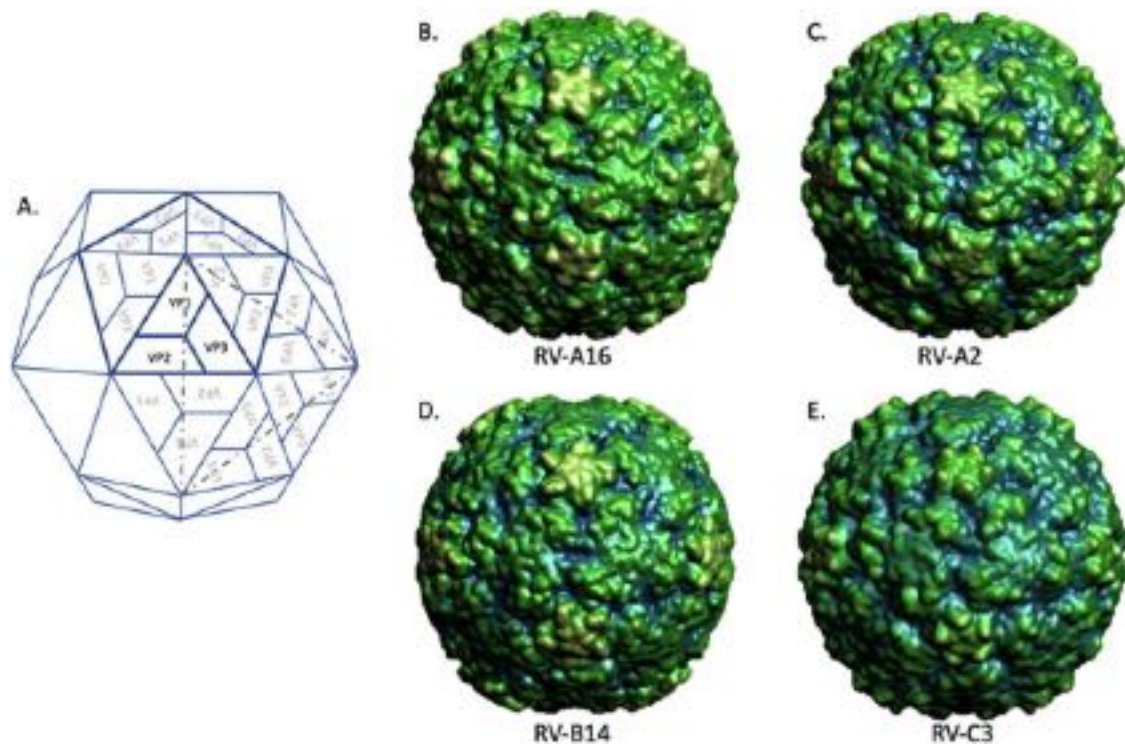


Fig. 1 Rhinovirus structure. (A) Schematic presentation of the RV capsid structure showing a protomer with three viral proteins VP1, VP2, and VP3 as a triangle and a pentamer of five protomers (thick blue lines). The capsid is formed from 12 pentamers. The triangular face of an icosahedron is indicated with dashed gray lines. (B–E) Capsid structure images of different RV species and major (RV-A16) and minor (RV-A2) receptor group RV-As. VP1 proteins around the pentamer vertex are discernible by the yellow coloring of heights. Blue color indicates depths. Predicted RV capsid structures were obtained under CC by 4.0 from McErlean *et al.* (2015): PLoS One, doi:10.1371/journal.pone.0001847.g005. Type names were updated according to the current classification.

5'-end (Fig. 2). The about 600 bases long 5' UTR contains secondary and tertiary structures acting as internal ribosomal entry sites (IRES). PolyA tail is required for infectivity and the 40–50 bases long 3' UTR for efficient replication, while VPg is dispensable in *in vitro* and *in vivo* experimental settings.

Life Cycle

Receptor

All RV-B types and the major group of RV-As use intracellular adhesion molecule 1 (ICAM-1) as their receptor, while the minor group of 10 RV-As use low-density lipoprotein receptor (LDLR) family proteins (Table 1). RV infection up-regulates the synthesis of membrane-bound ICAM-1 and down-regulates that of the soluble ICAM-1, thereby facilitating the epithelial cell infectivity. Cadherin-related family member 3 (CDHR3), mainly expressed in airway ciliated epithelial cells, has been identified as a putative receptor for RV-Cs, based on CDHR3 expression vector-transfected HeLa cell line, HeLa-E8. Experimental data on virus-receptor interactions involving CDHR3 has been described for types C2, C15, C40, C41, and C45. In primary human airway epithelial cells,

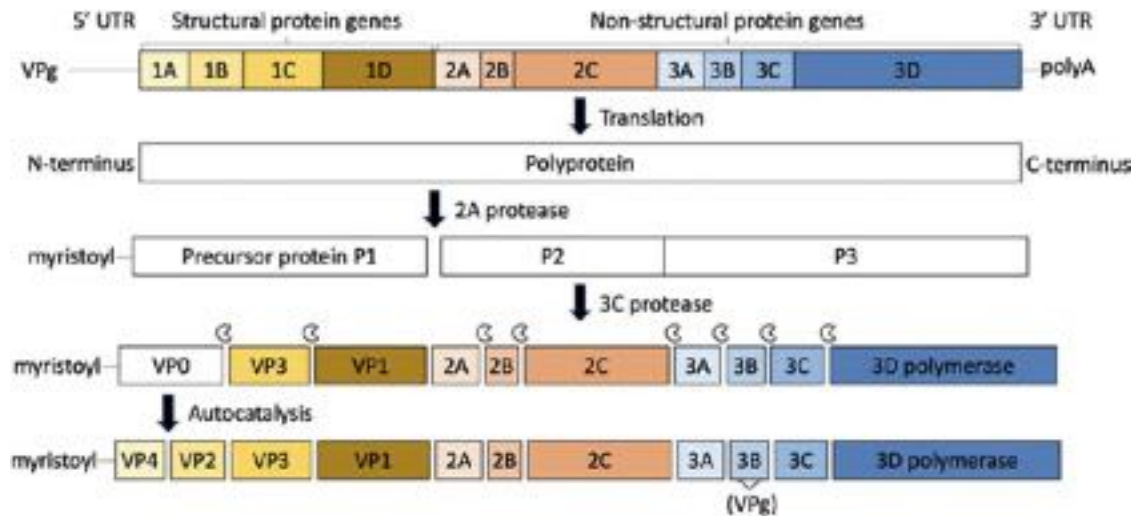


Fig. 2 Schematic illustration of rhinovirus genomic organization and protein formation. The genome is translated into a large polyprotein that is proteolytically processed through several precursor states to 11 proteins. First, protease activity of 2A co-translationally cleaves off the structural proteins' precursor P1. Then, subsequent cleavages take place through the action of 3C protease, except the final autocatalytic cleavage of VP0 to VP4 and VP2, turning the provirion into a mature virion.

Table 2 Receptor usage of rhinoviruses

Species	Prototype	Receptor ^a	Binding site on the virus capsid
RV-A (major)	A16	ICAM-1	Canyon at the 5-fold axis, VP1 and VP3
RV-A (minor)	A2	LDLR	Dome at the 5-fold axis, VP1
RV-B	B3	ICAM-1	Canyon at the 5-fold axis, VP1 and VP3
RV-C	C15	CDHR3	Surface groove at 3-fold axis, mainly VP2 and VP3

^aICAM-1, intracellular adhesion molecule 1; LDLR, low-density lipoprotein receptor; CDHR3, cadherin-related family member 3.

knock-out of CDHR3 reduced RV-C15 replication by 80%, while binding of the virus on the cells was not affected. Most of the RV-C/CDHR3 receptor-related work have been done with a growth adapted variant of the type C15 virus. Cryo-electron microscopy studies indicate different binding sites on the virion surface for each of the receptor molecules (Table 2).

Entry

Receptor-mediated endocytosis of both RV-A receptor groups and RV-Bs occurs predominantly via clathrin-coated pits in HeLa cells, but for ICAM-1-mediated endocytosis in other cells alternative routes may be in use. Since, in contrast to LDLR, ICAM-1 has no known internalization signals, co-receptor(s) are likely to be involved with ICAM-1.

Uncoating

All events of RV replication take place in the cytoplasm, where the viral RNA must be released from the endosome. The release is triggered by low endosomal pH inducing conformational changes in the capsid proteins. For minor receptor group viruses, the acidic trigger is direct and independent of dissociated LDLR. For major receptor group viruses, it is indirect and in synergy with the action of ICAM-1. VP4 and the N-terminus of VP1 induce membrane permeability through pore formation, resulting in RNA release. Empty virus capsids are transferred to lysosomes and degraded.

Translation

Guided by IRES at the 5' UTR, the viral RNA is translated as mRNA by host cell ribosomes and other cellular translation factors into a single polyprotein of 250 kD. The polyprotein is proteolytically processed via precursors (Fig. 2). First, the protease activity of 2A cleaves the polyprotein P1 to release it from the entire polyprotein. Then, protease 3C sequentially cleaves all the remaining several intermediate precursors, except VP4 and VP2, that remain as the precursor protein V0 until autocatalytic cleavage in the newly formed capsid packaged with viral RNA.

Replication

The RNA-dependent RNA polymerase protein 3D is mainly responsible for the replication of RV genome, facilitated by other non-structural proteins. To initiate the replication, 3D polymerase uridylylates the VPg protein using cis-acting replication element (cre) as the template. Cre has a hairpin RNA structure, and its location in the genome is species-specific: within 2A in RV-A, 1D (VP1) in RV-B, and 1B (VP2) in RV-C. Two adenosines in the loop of cre serve the uridylation. Uridylated VPg acts as the primer for RNA replication at the 3' polyA tail. The negative-sense RNA copy is then used as the template for positive-sense copies, resulting in an average ratio of 50 copies of (+)RNA to one copy of (−)RNA.

Assembly

Structural proteins VP1, VP3 and VP0 form protomers, which organize to pentamers and subsequently form the capsid. When the newly synthesized genomic RNA of the virus enters, the final maturation of the capsid takes place by autocatalytic cleavage of VP0 into VP2 and VP4, rendering the particle infectious. VP4 is myristoylated, which, in analogy with other enteroviruses, likely has a role in the final proteolytic step of maturation.

Release. In susceptible, cultured cell-lines, RV infection causes a strong cytopathic effect (CPE), indicating lytic release of nascent viruses. Ciliated epithelial cells are found in excretions of RV infected individuals regardless of symptom severity. On the other hand, studies of differentiated airway epithelial cells show only little CPE, also suggesting non-lytic RV release with an unknown mechanism.

Epidemiology

RV is circulating in the populations all year round, with typical peaks of high prevalence in early autumn and late spring. In year-round surveys with multiplex PCR detection methods, RV is the most common virus in all age groups, constituting 30%–50% of the findings. The epidemic seasons mirror each other in temperate zones of northern and southern hemispheres but are less clear near the equator. The autumn increase of RV prevalence coincides with the start of a new school year. When other respiratory viruses become more common during winter, the number of RV infections decrease, or their share of all infections is reduced and *vice versa* in the spring (Fig. 3). Thus, RV seems to be highly transmissible but sensitive to interference by other virus infections. Compared to influenza, respiratory syncytial virus (RSV), parainfluenza viruses and metapneumovirus, RV is more often detected as a co-infection with other viruses. These epidemiological observations could be explained by RV being sensitive to but a relatively poor inducer of host interferon (IFN) system.

RV-A is usually the most prevalent species, sometimes overtaken by RV-C, but RV-B is less common, representing rarely more than 10% of the findings. Some longitudinal studies have implicated RV-A12 and RV-A78 as more prevalent types. In addition to them, RV-C15 and RV-C2 are detected more frequently than others in general. Still, further studies over more extended periods of time must be performed to pinpoint these or any other types as more transmissible or more pathogenic.

Pathogenesis

Symptomatic RV infection begins with coryza, rhinorrhea or nasal obstruction. The neutrophilic inflammatory response to the infection leads to increased vascular permeability and mucin secretion. An uncomplicated RV infection is limited to the upper

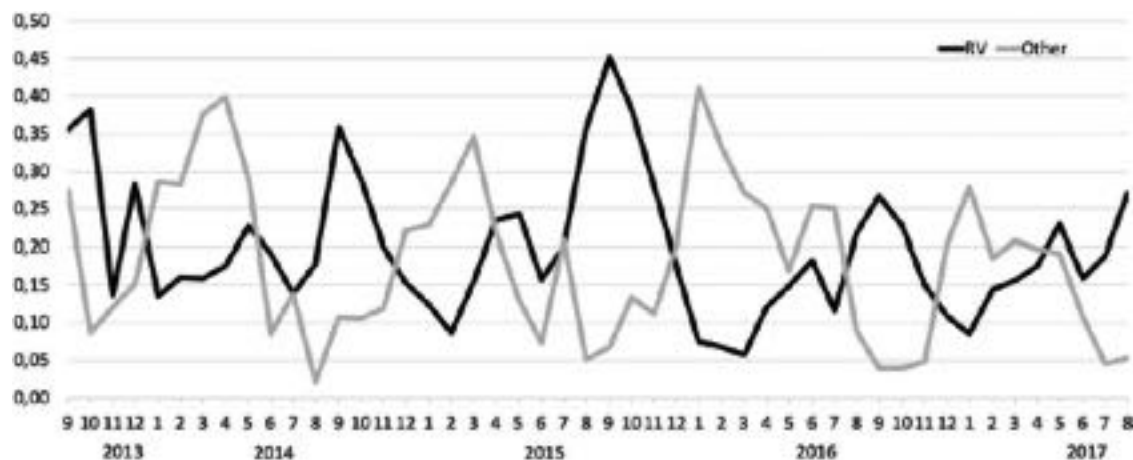


Fig. 3 Seasonality of rhinovirus infections. The graph shows the monthly incidence of RV as compared to selected other respiratory viruses, including influenza A and B virus, respiratory syncytial virus, parainfluenza virus types 1–4, and metapneumovirus. The incidence was calculated as findings per total number of monthly specimens in multiplex RT-PCR. Overall, 7436 specimens were tested; 1385 were positive for RV and 1440 for the other viruses. Data from the Clinical Microbiology laboratory of Turku University Hospital.

respiratory epithelium, but it is not effusively cytotoxic even in lower airways. However, there is a disruption in the epithelial barrier function, which may open the window for other opportunistic pathogens.

Neutrophils are attracted to the infection site mainly by chemokines interleukin (IL)-8 and IP-10, which are hallmarks of a symptomatic RV infection. Other proinflammatory cytokines induced by the host innate immune response include IL-6 and RANTES. Induction of type I and III IFNs is detectable through an expression of IFN stimulated genes and proteins, such as myxovirus resistance protein A and viperin, both locally and systemically. However, RV does not induce particularly strong IFN response that is often associated with fever. IFNs, in turn, are effective against RV infection. Poor IFN response is associated with prolonged RV infections in patients with a common variable immunodeficiency. Reduced *in vitro* IFN production has been found in studies on RV infection in airway epithelial cells from patients with asthma, cystic fibrosis, and chronic obstructive pulmonary disease (COPD).

Asymptomatic RV infection differs in the host transcriptome analysis from the symptomatic infections, which are characterized by overexpression of IFN-, inflammation-, neutrophil- and monocyte-related genes. Asymptomatic infections generally have lower viral loads.

Clinical Features

RV is the most commonly detected respiratory virus in all age groups, which makes it also the most common cause of any acute infection in humans. Children acquire their first RV infection early, and by two years of age, practically all have been infected at least once, with a mean annual rate of 3.5 infections. In adults, RV infections account for approximately 20%–50% of respiratory tract infections annually. In practice, over 90% of the subjects suffer at least one RV infection in a year. Species A and C are more common and may cause more severe illnesses early in life than RV-B.

Asymptomatic Infections

Many studies have shown that RV may be more common in asymptomatic subjects than in cases with respiratory symptoms. In analogy with COVID-19, this may have significance for transmission of the infection, but this phenomenon has remained poorly studied. RV RNA can be detected before, during, and after symptomatic infection, in subclinical infection, or may rarely be innocent contamination. Different studies have identified RV RNA in 12%–50% of asymptomatic children. In adults, asymptomatic respiratory infections, mostly caused by RV, are detected in 4%–6% of the subjects. It is essential to recognize that in otherwise healthy subjects, RV is not known to induce chronic illness. For these reasons, it has been concluded that RV RNA finding signifies most probably actual infection with or without symptoms.

Common Cold

Compared with other respiratory viruses, RVs have a dominant role as causative agents of upper respiratory tract infections in young children, adults and older people. The incubation period of RV common cold is two days. The most common symptoms of RV-induced common cold include rhinorrhea, nasal congestion, cough and sore throat. The symptoms are most severe during the first four days of illness and last, on an average, 8–10 days. In the older people, the median duration of the illness is 16 days. RV RNA positivity lasts up to four weeks after an acute RV infection, but the infectious period is likely to be closer to that of the most substantial symptoms.

Acute Otitis Media (AOM)

RSV has earlier been considered as a major predisposing factor for the development of AOM. Studies with RT-PCR suggest a dominant role of also RVs. In large Finnish Otitis Media Cohort Studies, 32%–42% of AOM episodes were associated with RV infections. RV RNA has been detected in 17%–41% of the middle ear fluid specimens from children with AOM.

Bronchiolitis and Recurrent Wheezing

RV is associated in up to one-quarter of bronchiolitis cases in infants, second only to RSV found in about two-thirds of them. RSV dominates in infants and RV in children older than 12 months. Interestingly, RV-infected children had a significantly shorter length of hospital stay as compared with children with RSV bronchiolitis. Innate immune responses inversely correlate with the RV disease severity. Several studies have shown that infants with a history of bronchiolitis will have significantly more often recurrent wheezing than controls without bronchiolitis. RV detection in the first wheezing episode is the most critical risk factor, followed by positive family history for asthma or atopy. Severe episodes of bronchiolitis increase the odds of early asthma. Asthma at seven years has been diagnosed in 52% or 15% of children, who have suffered RV or RSV bronchiolitis, respectively. There is evidence that early prednisolone treatment of RV-induced wheezing may reduce recurrent wheezing. This requires further confirmation, as prednisolone treatment does not achieve a similar benefit in bronchiolitis caused by RSV.

Asthma

RV infections are responsible for the majority of asthma exacerbations in children and adults. Up to 90% of asthma exacerbations in children are associated with virologically defined natural respiratory virus infections. RV is the most common finding and is detected in approximately two-thirds of the cases. Children with asthma and IgE-mediated allergic sensitization are more prone to RV infections. Polymorphisms in CDHR3 and chromosome locus 17q12–21 are associated with increased risk of RV wheezing. There is evidence that asthmatic patients have more frequent RV, more severe, and longer-lasting lower respiratory tract infections than non-asthmatics. Many studies have unveiled the mechanisms of increased susceptibility to RV infections in asthmatic patients. They include changes in airway epithelium, defective RV-induced secretion of IFN- β , IFN- λ , IL-10 and probably IFN- α , and impaired alveolar macrophage function. RV infection during the first weeks of life is a significant risk factor because it may permanently mold the response to subsequent infections and increase the risk of asthma. RV-A and RV-C are more likely to induce wheezing than RV-B.

Pneumonia

Many different studies suggest that RVs are important causative agents of pneumonia both in children and adults, and in immunosuppressed subjects. RV has been identified from induced sputum, tracheal brushing, bronchoalveolar lavage and lung tissue samples, strongly suggesting that it can cause lower respiratory tract infections. In studies on childhood community-acquired pneumonia RV is detected in one-fifth of the cases. The clinical profile of RV infections has included pneumonia in 11%–53% children admitted to hospital. RVs are commonly detected in children with severe pneumonia also in low-income countries. In a cornerstone study among 2320 adults with radiographic evidence of pneumonia, viral pathogens were detected in 589 of the cases and out of those RV was detected in 194 cases. RV may also be the causative agent of severe pneumonia in adults.

Chronic Obstructive Pulmonary Disease (COPD)

Respiratory viruses are involved in 40%–50% of acute exacerbations of COPD, which is the fourth leading cause of death worldwide. RV is the most common virus causing COPD exacerbations. Viral exacerbation is associated with a more severe illness than exacerbation with no virus. Experimental RV infections in patients with mild COPD have confirmed the causative role of RV in exacerbations. There is evidence that patients with COPD may have impairment in the immune response to RV and COPD patients may have chronic RV infections as also seen in patients with immunodeficiency diseases and lung and hematopoietic stem cell transplant recipients.

Diagnosis

As with other respiratory viruses, there are no clinical features of RV infection, allowing a specific diagnosis without laboratory tests. Nucleic acid detection is the only method for universal RV detection. Virus culture is less sensitive in general and not available for wild type RV-C strains. The significant number of different strains has thwarted the development of antibody reagents for diagnostic antigen detection. For the same reason, antibody detection for serological RV diagnostics in clinical use is a futile task.

Sampling

For the nucleic acid detection or culture, a nasopharyngeal swab is the specimen of choice, but any sample containing respiratory excretions, including nasal swab, nasopharyngeal or tracheal aspirate, (induced) sputum and bronchoalveolar lavage may be used. For culture, the specimen should be immediately placed into virus transport medium, for PCR, a swab can be transported even in a dry test tube for more flexible downstream processing in the laboratory. RV can be detected in respiratory tract tissue biopsies. Although RVs do not survive acidic pH of a normal stomach, they can be often detected in the feces of children.

Nucleic Acid Detection

All members of *Enterovirus* genus, including human enterovirus (EV) and RV species, have highly conserved motifs at the 5' UTR. These enable universal amplification of EVs and RVs with a single primer pair in RT-PCR. The optimal primer site is at the distal part of the 5' UTR (Fig. 4). Using advanced probe design, it is also possible to simultaneously differentiate RVs from EVs in a real-time RT-PCR assay. Using only dsDNA dye, like SYBR Green and melting curve analysis, it is possible to make a differential diagnosis with 80% accuracy, due to lower average melting temperatures of RVs. Additional conserved sites closer to the proximal part of the 5' UTR are applicable but result in longer amplicons with less efficient amplifications or less specific primers and probes. Many commercial assays do not differentiate between RVs and EVs, which is suboptimal, since it may delay early recognition of EVs in respiratory samples or lead to overestimation of EVs in stool samples. The copy number or threshold cycle (C_t) determination may elucidate clinical relevance of RV detection in co-infections.

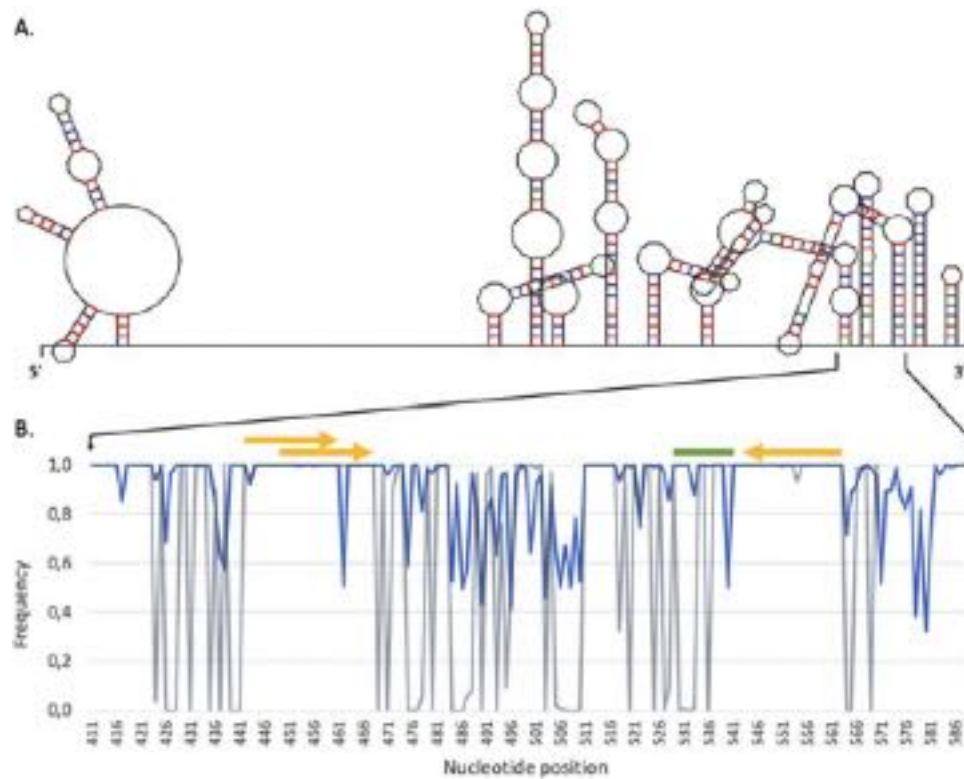


Fig. 4 Rhinovirus 5' UTR target site for RT-PCR. (A). 5' UTR with predicted secondary structure. The prediction was performed for RV-A1 genome (GenBank D00239.1) using mfold (unafold.rna.albany.edu). Base pairs: red line, C-G; blue line, A-U; green line, G-U. The coding sequence for 1A (VP4) starts at nucleotide 623. (B). Nucleotide homology at the optimal 5' UTR target site for RT-PCR. Frequency of the identical nucleotides of RVs (blue line) and EVs (gray line) as compared to the consensus RV sequence. The positions refer to RV-A1 genomic sequence (GenBank D00239.1). Sites for universal RV and EV primers are indicated by orange arrows and a target site allowing design of probes differentiating between RVs and EVs by a green line.

Culture

Routine culture of RV-A and RV-B strains can be carried out in heteroploid HeLa cells and primary diploid fibroblast cell lines MRC-5 and WI-38. RV-C strains can only be cultured in respiratory organ culture. HeLa-E8 cells over-expressing CDHR3 seem not to be applicable for universal RV isolation in cell culture. Since 30%–50% of circulating strains are representatives of RV-C species, culture is not recommended as the primary diagnostic method for RV infections. RV isolation is usually performed in roller tube cultures at 33–35°C for the best yield, but the multiplication of wild type viruses is not restricted at 37°C. A blind passage of week-old cultures increases the yield. A specific diagnosis of a positive culture requires confirmation by RT-PCR. A high C_t value in RT-PCR, corresponding to low amounts of virus in a cell culture supernatant, may indicate traces of non-replicating seed virus, abortive infection, or slow adaptation of the virus. Former methods for differentiation between cultivable RV and EV, by acid lability test (RVs are sensitive to $\text{pH} < 5$) or serological typing, are nowadays mostly not in use.

Typing

Identification of the strain in individual infections has no clinical value. Still, it is of great interest from an epidemiological point of view, which may also reflect on drug development. While the whole genome sequencing is developing, single amplicon sequencing, by Sanger or NGS methodology, is used for RV typing. To cover the high number of different RV types, highly degenerate primers targeting gene regions for VP1 or VP4/2 are used. Typing directly from RNA of clinical specimens usually has up to 75% success rate depending on the amount of virus and primability of the strain in question. The resulting sequence is compared with those in the GenBank, and the type is revealed among the nearest matches, following the rules explained above in the Classification section. At the species level, the more sensitive 5' UTR sequencing can be used, but there is some ambiguity in 5' UTR between RV-A and C species. On the other hand, 5' UTR sequence may strengthen the VP gene sequencing results. High resolution melting temperature (T_m) analysis of 5' UTR amplicons can be used to discern whether RV in sequential samples from the same individual represents the same RV type (same T_m) or subsequent infections with multiple types (different T_m s).

Treatment

No antiviral drugs have been approved for clinical use in HRV infections. Interestingly, the monoclonal anti-IgE antibody, omalizumab, reduces RV infections and shedding of RV in asthmatic children. Azithromycin, a macrolide antibiotic, reduces *in vitro* replication of RV. Furthermore, azithromycin increases RV-induced IFN and IFN-stimulated gene expression. In a controlled clinical study, azithromycin reduced the duration of RV-induced episodes of asthma-like symptoms in young children.

Prevention

There is no vaccine against RV infections. Avoidance of exposure to RV is the best way to prevent infection. A major route of transmission is probably from hands of infected subjects to an intermediary surface or directly to the fingers of the susceptible recipient, who is then infected by self-inoculation of the nose or eye. Handwashing with soap is considered better than an alcohol sanitizer because the effectiveness of the latter is questionable against the non-enveloped RV. Airborne RV has also been identified in exhaled breath and coughs of children and adults with an acute respiratory illness. Thus, aerosol transmission is a potential mode of transmission. Physical distancing, covering coughs and sneezes or using face mask may be recommended, as with any other respiratory infection, to prevent the spread of the infection.

Further Reading

- Altman, M.C., Beigelman, A., Ciaccio, C., *et al.*, 2020. Evolving concepts in how viruses impact asthma: A work group report of the microbes in Allergy Committee of the American Academy of Allergy, Asthma & Immunology. *Journal of Allergy and Clinical Immunology* 145, 1332–1344.
- Bartlett, N., Wark, P., Knight, D. (Eds.), 2019. *Rhinovirus infections: Rethinking the Impact on Human Health and Disease*. Academic Press: Elsevier.
- Jacobs, S.A., Lamson, D.M., George, K.S., Wash, T.J., 2013. Human rhinoviruses. *Clinical Microbiology Reviews* 26, 135–162.
- Jain, S., Self, R.G., Wunderink, R.G., *et al.*, 2015. Community-acquired pneumonia requiring hospitalization among U.S. adults. *New England Journal of Medicine* 373, 415–427.
- Leung, N.H.L., Chu, D.K.W., Shiu, E.Y.C., *et al.*, 2020. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nature Medicine* 26, 676–680.
- McIntyre, C.L., Knowles, N.J., Simmonds, P., 2013. Proposals for the classification of human rhinovirus species A, B and C into genotypically assigned types. *Journal of General Virology* 94, 1791–1806.
- Miller, K.E., Linder, J., Kraft, D., *et al.*, 2016. Hospitalization and outpatient visits for rhinovirus-associated acute respiratory illness in adults. *Journal of Allergy and Clinical Immunology* 137, 734–743.
- Oliver, M.E., Hinks, T.S.C., 2020. Azithromycin in viral infections. *Reviews in Medical Virology*. e2163. doi:10.1002/rvm.2163.
- Osterback, R., Tevaluoto, T., Ylinen, T., *et al.*, 2013. Simultaneous detection and differentiation of human rhino- and enteroviruses in clinical specimens by real-time PCR with locked nucleic acid probes. *Journal of Clinical Microbiology* 51, 3960–3967.
- Paul, A.V., Wimmer, E., 2015. Initiation of protein-primed picornavirus RNA synthesis. *Virus Research* 206, 12–26.
- Peltola, V., Waris, M., Kainulainen, L., Kero, J., Ruuskanen, O., 2013. Virus shedding after human rhinovirus infection in children, adults and patients with hypogammaglobulinaemia. *Clinical Microbiology and Infection* 19, E322–E327.
- Ritchie, A.I., Wedzicha, J.A., 2020. Definition, causes, pathogenesis, and consequences of chronic obstructive pulmonary disease exacerbations. *Clinics in Chest Medicine* 41, 421–438.
- Ruuskanen, O., Waris, M., Ramilo, O., 2013. New aspects of human rhinovirus infections. *The Pediatric Infectious Disease Journal* 32, 553–555.
- Zlateva, K., van Rijn, A.L., Simmonds, P., *et al.*, 2020. Molecular epidemiology and clinical impact of rhinovirus infections in adults during three epidemic seasons in 11 European countries (2007–2010). *Thorax* 75, 882–890.

Relevant Websites

- www.picornaviridae.com
Picornavirus Home.
- <https://virologydownunder.blogspot.com/2015/04/rhinoviruses-rvsa-primer.html>
Rhinoviruses (RVs)...a primer.

Rift Valley Fever Virus and Other Phleboviruses (*Phenuiviridae*)

Tetsuro Ikegami, The University of Texas Medical Branch at Galveston, Galveston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ABIN2	A20-binding inhibitor of NF- κ B 2	NSm	Nonstructural gene encoded by the bunyavirus M-segment
ALT	Alanine aminotransferase	NSs	Nonstructural gene encoded by the bunyavirus S-segment
AMTV	Arumowot virus	ORF	Open reading frame
AST	Aspartate aminotransferase	PKR	Double-stranded RNA-dependent protein kinase
BCX4430	Galidesivir	PTV	Punta Toro virus
CPK	Creatinine phosphatase	RIG-I	Retinoic acid-inducible gene I
DC	Dendritic cells	RNP	Ribonucleocapsid
DC-SIGN	Human dendritic cell-specific intercellular adhesion molecule 3 grabbing non-integrin	RVF	Rift Valley fever
DC-SIGNR	Liver/lymph node-specific intercellular adhesion molecule 3 grabbing nonintegrin (L-SIGN)	RVFV	Rift Valley fever virus
eIF2 α	Eukaryotic initiation factor 2 α	SAP30	Sin3A associated protein 30
ELISA	Enzyme-linked immunosorbent assay	SFNV	Sandfly fever Naples virus
ENSO	El Niño Southern Oscillation	SFSV	Sandfly fever Sicilian virus
GTV	Guertu virus	SFTSV	Severe Fever with Thrombocytopenia Syndrome virus
HRTV	Heartland virus	S-segment	Small segment
IFN	Interferon	STAT	Signal transducer and activator of transcription
IFN-AR	Interferon- α/β receptor	T-705	Favipiravir
IRF-3	Interferon regulatory factor 3	TBK1	TANK-binding protein 1
LAMP	Loop-mediated isothermal amplification	TFIIH	Transcription factor II H
LDH	Lactate dehydrogenase	TOSV	Toscana virus
L-segment	Large segment	TPL2	Tumor progression locus 2
M-segment	Medium segment	TSI-GSD-200	A formalin-inactivated RVF vaccine under clinical trials for human use (U.S.)
NIH	National Institutes of Health	UUKV	Uukuniemi virus
		VLP	Virus-like particle

Glossary

78-kD protein A protein encoded by the RVFV M-segment. It is expressed via cotranslational cleavage of a glycoprotein precursor from the first initiation codon.

Ambi-sense A viral coding strategy encoding two different genes in negative- and positive-sense RNA, respectively.

Antigenic complex A group of phleboviruses, which show high cross-reactivity via complement fixation test and hemagglutination inhibition test.

Cap-snatching Use of 5' short fragments of capped cellular mRNAs to prime viral mRNA synthesis. Cleavage of cellular mRNA is mediated via an endonuclease activity of viral L proteins.

Category A-C Priority pathogens The National Institutes of Health in the United States of America list and classify pathogens based on emerging threat to the national security and public health.

DC-SIGN A C-type lectin protein expressed on the surface of dendritic cells or macrophages. DC-SIGN can bind to specific types of carbohydrates and the interaction can lead to phagocytosis.

Floodwater *Aedes* spp. Floodwater *Aedes* spp. mosquitoes in floodplain habitat lay eggs in moist or waterlogged soil. Eggs survive drought and hatch upon rainfall or flooding.

Select agent Biological agents and toxins regulated under the Federal Select Agent Program in the United States of America, which potentially pose a severe threat to public health and safety.

Transcription factor II H TFIIH play important roles in transcriptional initiation via RNA polymerase II and DNA nucleotide excision repair. TFIIH is comprised by ten subunits, XPB, XPD, p52, p44, p62, p34, p8, cdk7, cyclin H, and MAT1.

Introduction

The family *Phenuiviridae*, within the order *Bunyavirales*, consists of 15 genera: *Beidivirus*, *Goukovirus*, *Horwuvirus*, *Hudivirus*, *Hudovirus*, *Mobuvirus*, *Phasivirus*, *Pidchovirus*, *Tenuivirus*, *Wenrivirus*, *Wubeivirus*, *Laulavirus*, *Kabutovirus*, *Banyangvirus*, and *Phlebovirus*.

Viruses belonging to the former 12 genera are not known to cause any diseases in vertebrates. Many viruses belonging to the genus *Phlebovirus* are transmitted via arthropod vectors, such as sandflies, mosquitoes, or ticks, and they cause mild to severe diseases in humans and/or in animals. Some tick-borne phleboviruses were reclassified into the *Banyangvirus* or *Kabutovirus* genera, which were newly made in 2018–2019, are also described in this article. At least 100 distinct phleboviruses have been characterized and named, although only 10 species have been officially approved via the International Committee for Taxonomy of Viruses: *Bujaru*, *Candiru*, *Chilibre*, *Frijoles*, *Mukawa*, *Punta Toro*, *Rift Valley fever*, *Salehabad*, *Sandfly fever Naples*, and *Uukuniemi phlebovirus*. These 10 phleboviral species can be distinguished via four-fold differences in the two-way cross-neutralization test. The remaining phleboviruses have not been officially assigned to species. Such phleboviruses can be further grouped into antigenic complexes via serological cross-reactivity using a complement fixation test and a hemagglutination inhibition test. In addition to serological tests, phylogenetic analyses can be additionally applied to the grouping of phleboviruses.

Rift Valley fever (RVF) is a mosquito-borne zoonotic disease caused by Rift Valley fever virus (RVFV), which is endemic to sub-Saharan Africa, but has spread to Egypt, Madagascar, the Comoros, Yemen, and Saudi Arabia. RVF is one of the most pathogenic phleboviral diseases affecting ruminants and humans, and it has been classified as a Category A Priority Pathogen by the United States (U.S.) National Institutes of Health (NIH), and as a Blueprint priority disease by the World Health Organization. In response to an increasing fear of bioterrorism, RVFV is also designated as an overlap select agent by the U.S. Department of Health and Human Services and the U.S. Department of Agriculture, meaning that its possession and handling are strictly monitored in the U.S.

Phlebotomus fever (three-day fever) is a disease caused by Sandfly fever Naples virus (SFNV) and Sandfly fever Sicilian virus (SFSV), which are transmitted via sandflies in Eurasia and Africa. Although the febrile disease is self-limiting, outbreaks may affect the health status of soldiers or travelers for up to two weeks before a complete recovery from the illness occurs. Moreover, Toscana virus (TOSV), which was first isolated from sandflies in Italy in 1971, has spread into Europe and North Africa. Aseptic meningitis or meningoencephalitis caused by TOSV are more severe than the typical phlebotomus fever. In the New World, phlebotomus fever can be caused by several different phleboviruses, such as the Punta Toro (PTV), Chagres, and Candiru viruses. Furthermore, febrile illness associated with a viral meningitis can be caused by a tick-borne Bhanja virus, which is widely distributed throughout Eurasia and Africa. Uukuniemi virus (UUKV) is a tick-borne phlebovirus that is not pathogenic to humans. This virus has been used as a model for phleboviral virology studies.

In the past decade, the *Severe Fever with Thrombocytopenia Syndrome phlebovirus* (SFTSV), a highly pathogenic tick-borne phlebovirus, was identified in East Asia (China, Korea, Japan, Vietnam, Taiwan). A genetically similar pathogenic tick-borne phlebovirus was also identified in the U.S. Those pathogenic tick-borne phleboviruses were reclassified from the genus *Phlebovirus* into the genus *Banyangvirus*, which now includes *Huaishangyang banyangvirus* (formerly SFTSV), *Heartland banyangvirus* (HRTV), and *Guertu banyangvirus* (GTV). Due to public health concerns, both SFTSV and HRTV are classified as Category C Priority Pathogens in the U.S.

Virion Structure, Genome, and Strategy of Replication

The phleboviral genome consists of three single-stranded RNA segments: negative-sense Large (L)- and Medium (M)-segments, and an ambi-sense Small (S)-segment (Fig. 1). The L-segment encodes a single open reading frame (ORF) for the RNA-dependent RNA polymerase and the M-segment encodes a single ORF for a precursor protein, which can be co-translationally cleaved into Gn and Gc envelope glycoproteins. Some phleboviruses also encode non-structural M (NSm) proteins upstream of the Gn proteins. The S-segment encodes an ORF for the nucleoprotein (N) in the negative-sense RNA, and an ORF for the non-structural S (NSs) protein in the positive-sense RNA. Phleboviral virions are spherical or pleomorphic. In virions, ribonucleocapsids (RNP), which consist of viral RNA encapsidated with N proteins, are surrounded by a viral envelope, which comprises a host-derived lipid bilayer and Gn/Gc proteins. RVFV virions are 90–110 nm in diameter, with surface spikes 10–18 nm in length. Cryoelectron tomography has further characterized the surface arrangement of hollow-like capsomers composed of Gn/Gc heterodimers and/or homodimers arranged in $T = 12$ icosahedral symmetry. The $T = 12$ icosahedral symmetry of capsomers was also found in virions of a tick-borne UUKV, which is 125 nm in its mean diameter, with surface spikes 12–15 nm in length.

Virions attached to cell surface receptors are internalized via endocytic pathways. One receptor molecule for phleboviruses is the human dendritic cell-specific intercellular adhesion molecule 3 grabbing non-integrin (DC-SIGN), which is a C-type lectin expressed on the surface of a specific subset of dendritic cells (DCs) and macrophages. Binding to specific types of mannose-rich or fucosylated glycans leads to an uptake of ligands via endocytosis. Other molecules, such as DC-SIGNR (L-SIGN), heparan sulfate proteoglycans, and the non-muscle myosin heavy chain IIA (NMMHC-IIA), are also important for the attachment of some phleboviruses or banyangviruses. Endocytosis involved in phleboviral entry is mediated via caveolae or clathrin, depending on the virus. In late endosomes, low pH leads to an exposure of hydrophobic fusion loops as a result of structural changes in the Gc protein (class II fusion protein). Fusion loops can then be inserted into cellular endosomal membranes, triggering the release of RNP into the cytoplasm.

Once RNP is released into the cytoplasm, L proteins associated with incoming virions start the transcription of viral mRNAs (primary transcription). L proteins cleave host cytoplasmic mRNA, typically 10–20 nt from the 5' termini, and use those oligonucleotides as primers to synthesize their own transcripts (cap-snatching). Viral mRNA is not polyadenylated at the 3' terminus, and the lengths of transcripts are shorter than those of viral genomic RNA templates because transcriptional termination occurs

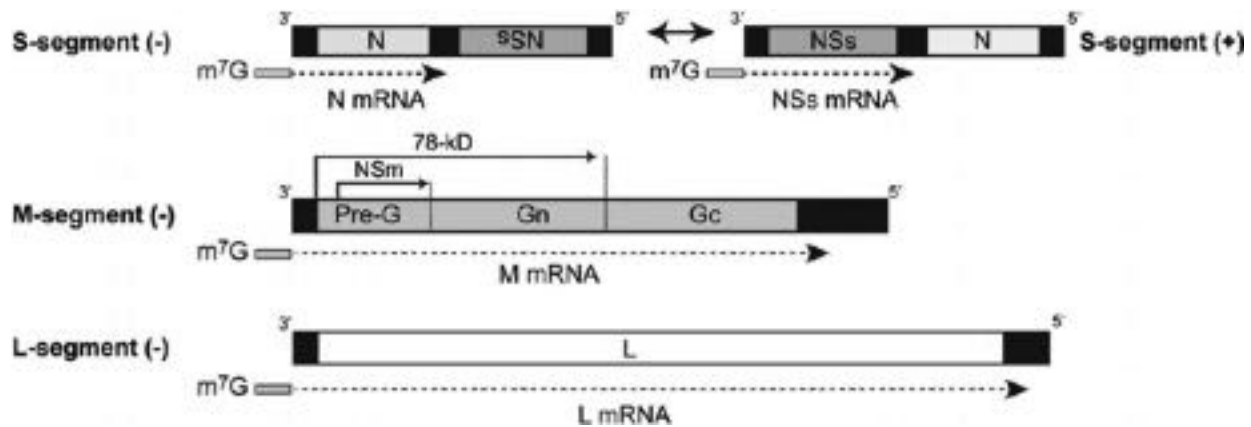


Fig. 1 Genome structure of phleboviruses. The schematics represent the genome structure of Rift Valley fever virus (RVFV). Viral mRNAs are transcribed via cap-snatching of host mRNA. N and NSs mRNA are transcribed from negative- and positive-sense S-segment, respectively. The RVFV M mRNA can be translated into two precursors, which generate (i) 78-kD + Gc, and (ii) NSm + Gn + Gc. Tick-borne phleboviruses do not encode NSm gene, whereas glycoprotein precursors have not been characterized for most phleboviruses. Newly identified tick-borne phleboviruses (e.g., Blacklegged tick phleboviruses) encode a large N gene and a small NSs-like open reading frame in S-segment.

within untranslated regions in L-, M-, and S-segments. The positive-sense S-segment of RVFV or UUKV can also be packaged into virions, which allows the accumulation of primary transcripts of NSs mRNA from virion-associated positive-sense S-segments, together with N mRNA, during primary transcription. With an increase in cellular levels of N and L proteins, the replication of viral genomic RNA occurs. The sequences of the 3' and 5' termini of L-, M-, and S-segments are complementary, and thus the RNPs of L-, M-, and S-segments form panhandle structures. The panhandle structures are considered to support encapsidation of RNA with N proteins and promote the recruitment of L proteins for viral genomic RNA replication.

RVFV N proteins can be assembled into a hexameric ring, structurally displaying both a RNA binding cleft and a N oligomerization groove, which allows the multimerization of N proteins along with genomic RNA. Viral RNA synthesis occurs via genomic RNA encapsidated with N proteins. Accumulation of genomic RNA during viral RNA replication can also amplify viral mRNA synthesis (secondary transcription), followed by further accumulation of viral proteins. The assembly of viral components occurs in the Golgi complex or in the ER-Golgi intermediate compartment. Although phleboviruses do not encode any matrix proteins, the cytoplasmic domain of Gn play a role in interactions with RNP in the Golgi. The Golgi retention signal is also encoded surrounding the Gn cytoplasmic tail. The Gc protein cytoplasmic tail at the C-terminus is very short, which encodes a presumable ER retrieval signal. Co-expression of Gn and Gc proteins leads to the co-localization of both Gn and Gc proteins in the Golgi. Mutagenesis of the Gc cytoplasmic tail could lead to mislocalization of Gc into the ER, even in the presence of Gn proteins, indicating that the Gc cytoplasmic tail plays a role in the dimerization of Gn and Gn proteins, or in the trafficking of Gc proteins to the Golgi. The formation of virus-like particles (VLPs) occurs with the expression of Gn and Gc proteins. In the case of RVFV, however, the presence of RNP facilitates the production of VLPs. In an artificial over-expression system of RVFV Gn and Gc proteins using insect Sf9 cells, VLPs consisting of either Gn, Gc, and N proteins, or Gc and N proteins could be produced without RNP.

Functions of Nonstructural Proteins

NSs proteins, which are expressed from the S-segment, are the major virulence factors for pathogenic phleboviruses and banyangviruses, and they share low similarity in amino acid level. RVFV NSs proteins form filamentous structures in the nuclei of infected cells, sequestering p44 protein and promoting post-translational degradation of p62 protein, both of which are among ten subunit proteins of transcription factor (TF) IIH, which is essential for cellular transcription via RNA polymerase II. The expression of RVFV NSs proteins can thus induce general transcriptional suppression, including that of the IFN- β gene. RVFV NSs filaments also bind to Sin3A Associated Protein 30 (SAP30) and form a repressor complex on the IFN- β gene promoter, which can be another mechanism inhibiting IFN- β gene expression in RVFV-infected cells. RVFV NSs proteins also localize to the cytoplasm and promote post-translational degradation of dsRNA-dependent protein kinase (PKR), which is a cellular sensor of dsRNA and triphosphated ssRNA. PKR bound to RNA substrates exposes kinase-active domains, which can phosphorylate the eukaryotic initiation factor (eIF) 2 α subunit and lead to the shutoff of host and viral protein syntheses. Because RVFV NSs proteins can promote the degradation of PKR, infected cells can maintain active viral protein synthesis due to a lack of PKR-mediated eIF2 α phosphorylation.

TOSV NSs proteins also inhibit the expression of IFN- β gene, while promoting the posttranslational degradation of PKR and the retinoic acid-inducible gene I (RIG-I), but not that of TFIIF p62. SFTSV NSs proteins form inclusion bodies in the cytoplasm, which sequester interferon regulatory factor 3 (IRF3) via binding to TANK-binding protein 1 (TBK1). These inclusion bodies of SFTSV NSs proteins also sequester cellular signal transducer and activator of transcription (STAT) 1 and STAT2. Accordingly, both

the induction and signaling pathways for the IFN- β gene can be inhibited by SFTSV NSs proteins. Moreover, the SFTSV NSs proteins bind to the A20-binding inhibitor of NF- κ B 2 (ABIN2), which in turn inhibits the suppression of tumor progression locus 2 (TPL2) and p105 via ABIN2, and leads to the upregulation of anti-inflammatory genes, such as IL-10. HRTV NSs protein does not form inclusion bodies in the cytoplasm, but it binds to TBK-1, which hinders the association of TBK1 with IRF3. Other pathogenic phleboviruses, such as PTV and SFSV, also inhibit the expression of IFN- β gene, but these inhibitor mechanisms have so far remained unknown. The NSs protein of Arumowot virus (AMTV) inhibits the upregulation of IFN- β gene in murine cells, whereas the AMTV NSs proteins can be posttranslationally degraded via proteasomal pathway in human cells, indicating varied stability of NSs proteins in different host species.

The NSm protein is encoded upstream of the GnGc coding region in the M-segment, although tick-borne phleboviruses or banyangviruses do not encode the NSm gene. The role of NSm proteins is largely unknown in most phleboviruses. The RVFV M-segment encodes five in-frame AUGs upstream of the GnGc coding region, and cellular ribosomes can initiate translation from the 1st and the 2nd AUGs via leaky scanning. The first AUG encodes a precursor protein of 78 kD, whereas the second AUG encodes a precursor that can be co-translationally cleaved into the NSm, Gn, and Gc proteins. RVFV lacking four of the five AUGs can still replicate in cells, indicating that 78 kD or NSm proteins are dispensable in the viral life cycle. In contrast, specific knockout of the 78 kD protein may result in poor viral transmission in mosquitoes. In one study, RVFV 78 kD proteins were incorporated into virions in mosquito C6/36 cells, but not in Vero cells, indicating an interesting biological difference between mosquito and mammalian cells. Furthermore, RVFV NSm proteins localize to the mitochondrial outer membrane via basic amino acids upstream of a putative transmembrane domain at the C-terminus and suppress apoptosis. Abolishing of both the 78kD and NSm proteins can partially attenuate RVFV.

The Viruses

Rift Valley Fever Virus

RVF is now endemic in sub-Saharan Africa, Madagascar, the Comoros, Egypt, Saudi Arabia, and Yemen, and it is primarily maintained via floodwater *Aedes* spp. mosquitoes (e.g., subgenera *Neomelanimon*, *Aedimorphus*) (Table 1). RVFV can be vertically transmitted into offspring via transovarial transmission, and viruses are considered to maintain their infectivity in drought-resistant eggs for several years. Other mosquito species also play a role in the amplification of RVFV via susceptible animals in nature (e.g., mosquitoes of the genera *Culex*, *Coquillettidia*, *Anopheles*, *Eretmapodites*, *Mansonia*). In endemic areas, animals such as African buffalos, domestic ruminants, bats, rodents, and other wild mammals are likely involved in the maintenance of RVFV. It has been suggested that increased biting of susceptible animals by RVFV-infected mosquitoes can trigger RVFV amplification and epidemics. In terms of increasing mosquito populations, the establishment of habitats favorable for the hatching of infected mosquitoes (e.g., heavy rain, irrigation) is important. The prediction of RVF outbreaks linking to heavy rainfall in climates during the El Niño Southern Oscillation (ENSO) condition is one of the important efforts for allowing early preparation for RVF outbreaks in eastern Africa.

RVFV isolates can be grouped into lineages A – G (alternatively, lineages A – N). The maximum genetic diversity of L-, M-, and S-segments among RVFV strains is 4%, 5%, and 4%, respectively, at the nucleotide level; and 1%, 2%, and 1%, respectively, at the deduced amino acid level. High similarity among RVFV isolates allows cross-neutralization using antisera raised against one RVFV strain.

An RVF outbreak is a major public health threat, and can also cause devastating economic damage by affecting animal industry. For example, a single RVF outbreak in South Africa in 1951 resulted in an estimated 500,000 abortions and 100,000 deaths among sheep. During RVF outbreaks, high rates of abortions occur in sheep (up to 100%), goats (up to 100%), and cattle (up to 40%). Newborn lambs and goat kids less than one week old are highly susceptible to RVFV and, due to an acute or peracute liver necrosis, case fatality rates are typically higher than 90%, whereas those of newborn calves range from 20% to 70%. Adult ruminants are more resistant to RVFV infection than are newborn animals, and the disease is associated with hepatic necrosis, resulting in case fatality rates of 5%–30% for sheep and goats, and 10% for cattle. Camels are also susceptible to RVFV, although the clinical signs vary a lot according to different reports (e.g., asymptomatic, icterus, ocular discharge, blindness, hemorrhage, nervous symptoms, abortion). RVFV does not usually cause diseases in pigs, horses, dogs, cats, or birds.

Humans can be infected with RVFV via the bite of an infected mosquito, or through an exposure to the bodily fluids of infected animals. Infected humans exhibit a biphasic febrile illnesses. An estimated 18,000–200,000 people were infected during the first RVF outbreak in Egypt in 1977–1978, and 598 patients died. Although the overall case fatality rate of humans is 1%–2%, the case fatality rate among hospitalized patients was estimated to be 34% during the RVF outbreak in Saudi Arabia in 2000–2001. The incubation period is typically 4–6 days, and clinical signs start abruptly. These include malaise, severe chills, severe headache, nausea, weakness, and fever (38.8–39.5°C). Clinical signs can be reduced in 3–4 days, although severe clinical signs can recur thereafter. In the convalescence phase, most patients recover without sequela in two weeks or more, although some patients experience neurological disorders, thrombosis, or vision loss weeks or months later. Some patients suffer from an acute hemorrhagic syndrome, including macular rash, ecchymosis, bleeding, or disseminated intravascular coagulation.

RVF can be diagnosed based on the detection of elevated IgM and IgG specific to RVFV via an enzyme-linked immunosorbent assay (ELISA), using paired serological samples. The detection of viral neutralizing antibody can improve the specificity of

Table 1 Phleboviruses in the old world (mosquito-, sandfly-borne, and other non tick-borne viruses)

Antigenic complex	Virus	Isolations	Geographic distribution	Diseases
Sandfly Fever Naples	Sandfly Fever Naples	Sandflies (<i>Phlebotomus papatasi</i> , <i>P. perniciosus</i> , <i>P. perfiliewi</i> and others)	South Europe, North Africa, Middle East, South and Central Asia	(Human) Phlebotomus fever
	Toscana	Sandflies (<i>P. perfiliewi</i> and <i>P. perniciosus</i> , and others)	Italy, Spain, Portugal, France, Malta, Croatia, Bosnia-Herzegovina, Greece, Turkey, Cyprus, Morocco, Algeria, Tunisia	(Human) Phlebotomus fever, aseptic meningitis, meningoencephalitis
	Massilia	Sandflies (<i>P. perniciosus</i>)	Portugal, France	Unknown
	Granada	Sandflies (<i>P. perniciosus</i>)	Spain	(Human) Seroprevalence in Spain
	Arrabida	Sandflies (<i>Phlebotomus</i> spp.)	Portugal	Unknown
	Fermo	Sandflies (<i>P. perfiliewi</i>)	Italy	Unknown
	Tehran	Sandflies (<i>Phlebotomus</i> spp.)	Iran	Unknown
	Zerdali	Sandflies (<i>Phlebotomus</i> spp.)	Turkey	Unknown
	Punique	Sandflies (<i>P. perniciosus</i> , <i>P. longicuspis</i>)	Tunisia	Unknown
	Gordil	Striped grass mice (<i>Lemniscomys striatus</i>) Gerbil (<i>Tatera kempi</i>)	Central African Republic	Unknown
	Saint Floris	Gerbil (<i>Tatera kempi</i>)	Central African Republic	Unknown
Sandfly Fever Sicilian	Sandfly Fever Sicilian	Sandflies (<i>P. papatasi</i> and others)	South Europe, North Africa, Middle East, South and Central Asia	(Human) Phlebotomus fever
	Sandfly Fever Turkey	Sandflies (<i>Phlebotomus</i> spp.)	Turkey	(Human) Phlebotomus fever, aseptic meningitis, meningoencephalitis
	Sandfly Fever Cyprus	Human	Cyprus	(Human) Phlebotomus fever
	Corfou	Sandflies (<i>P. neglectus</i>)	Greece	Similar viral RNA was detected from a patient with encephalitis
	Toros	Sandflies (<i>Phlebotomus</i> spp.)	Turkey	Unknown
	Dashli	Sandflies (<i>Sergentomyia</i> spp.)	Iran	Unknown
Salehabad	Salehabad	Sandflies (<i>Phlebotomus</i> spp.)	Iran	Unknown
	Arbia	Sandflies (<i>P. perfiliewi</i> , <i>P. perniciosus</i>)	Italy	Unknown
	Medjerda	Sandflies (<i>Phlebotomus</i> spp.)	Tunisia	Unknown
	Adana	Sandflies (<i>Phlebotomus</i> spp.)	Turkey	Unknown
	Alcube	Sandflies (<i>Phlebotomus</i> spp.)	Portugal	Unknown
	Ponticelli	Sandflies (<i>Phlebotomus</i> spp.)	Italy	Unknown
	Arumowot	Mosquitoes (<i>Culex</i> , <i>Mansonia</i> spp.) Gerbils (<i>Tatera kempi</i>) Wild rats (<i>Arvicanthis niloticus</i> , <i>Thomomys macmillan</i>) Shrew (<i>Crocidura</i> sp.) Striped grass mice (<i>Lemniscomys striatus</i>) Birds (<i>Turdus libyanus</i>)	Kenya, Sudan, Uganda, Ethiopia, Nigeria, Senegal, Central African Republic, Rhodesia, Zimbabwe, South Africa	(Human) Seroprevalence in Somalia, Sudan, and Egypt
	Odrenisrou	Mosquitoes (<i>C. albiventr</i>)	Ivory Coast	Unknown
Rift Valley fever	Rift Valley fever	Mosquitoes (<i>Aedes</i> , <i>Culex</i> , <i>Coquillettidia</i> , <i>Anopheles</i> , <i>Eretmapodites</i> , <i>Mansonia</i> spp.)	Sub-Saharan Africa, Egypt, Madagascar, Saudi Arabia, Yemen	(Human) Febrile illness, hemorrhagic fever, encephalitis, blindness (Ruminant) Febrile illness, abortion (Newborn lambs, goat kids) Acute hepatitis
Karimabad	Karimabad	Sandflies (<i>Phlebotomus</i> spp.)	Iran	(Human) Seroprevalence in Iran, Azerbaijan, Uzbekistan, Tajikistan, Pakistan

(Continued)

Table 1 Continued

Antigenic complex	Virus	Isolations	Geographic distribution	Diseases
	Gabek Forest	Spiny mouse (<i>Acomys cahirinus</i>) Hedgehog (<i>Atelerix albiventrix</i>) Bushbaby (<i>Galago senegalensis</i>) Gerbil (<i>Tatera kempi</i>) Wild rat (<i>Arvicanthis niloticus</i>) Striped grass mice (<i>Lemniscomys barbarus</i>)	Sudan, Nigeria, Benin, Senegal, Central African Republic	Unknown
Unclassified virus	Salanga virus	Wild rodent (<i>Aethomys medicatus</i>)	Central African Republic	Unknown

serological diagnosis. To confirm the presence of pathogenic RVFV strains, it is important to isolate RVFV in cultured cells (e.g., Vero or BHK cells), or detect viral RNA via quantitative reverse transcriptase (RT)-PCR, RT-Loop-Mediated Isothermal Amplification (LAMP), Recombinase Polymerase Amplification, or next-generation sequencing, although the period of viremia is short in ruminants and humans. Post-mortem diagnosis can be made via the detection of RVFV-specific antigens or viral RNA, associated with histopathological lesions.

Vaccination is likely the most effective way of preventing the outbreak of RVF. There are several licensed RVF vaccines for veterinary use in endemic countries, but no licensed RVF vaccines for humans are currently available. The live-attenuated Smithburn vaccine was developed via serial intracerebral passages of the pathogenic Entebbe strain in the 1950s. This vaccine is manufactured in South Africa and Kenya, and more than 46 million doses have been sold. Although long-term protective immunity can be induced via a single dose of the Smithburn vaccine, drawbacks include potential abortion or fetal malformation in pregnant animals. The live-attenuated Clone 13 vaccine originated from a plaque variant of the 74HB59 strain in the Central African Republic and the vaccine virus has an in-frame 69% truncation of the NSs gene, which is a major virulence factor for RVF. Accordingly, this strain induces host type I IFN responses and it is avirulent in most immunocompetent animals. The Clone 13 vaccine has been licensed in several African countries, and more than 28 million doses have been given in South Africa. Three different inactivated RVF vaccines are also available in Africa: the formalin-inactivated RVFV strain (South Africa), the formalin-inactivated RVFV Menya/Sheep/258 strain (Egypt), and the Binary Ethylenimine-inactivated RVFV ZH501 strain (Egypt). In the U.S., the live-attenuated MP-12 vaccine was generated via serial passaging of the pathogenic ZH548 strain in human lung diploid MRC-5 cells in the presence of a chemical mutagen, 5-fluorouracil. The MP-12 vaccine shows attenuation via point mutations in the L-, M-, and S-segments, and it exhibits protective efficacy in livestock ruminants. Vaccination of ewes in the second or third trimester can confer protective immunity not only for vaccinated ewes, but also for newborn lambs via colostrum containing the neutralizing antibody. The MP-12 vaccine has been conditionally licensed for veterinary use upon RVFV introduction in the U.S.

As mentioned above, there are currently no licensed human RVF vaccines. In the U.S., the formalin-inactivated Entebbe strain was developed as a human RVF vaccine candidate (TSI-GSD-200), and its safety and immunogenicity has been tested in clinical trials. This vaccine elicits protective immunity via three primary doses, and the maintenance of immunity requires one or more booster doses. A live-attenuated MP-12 vaccine has also been studied in clinical trials for potential human use. In comparison to the formalin-inactivated vaccine, this vaccine can induce long-term protective immunity via a single dose, with no major adverse effects reported in clinical trials. Several other vaccine candidates are being characterized for safety, immunogenicity, and efficacy at the research level.

There are no effective treatment regimens for RVF patients. Ribavirin, a guanosine analog, has been shown to effectively limit RVFV replication in animal models. In mice, ribavirin can limit acute liver disease, but not viral encephalitis, in a postexposure treatment regimen, whereas favipiravir (T-705), another guanidine nucleoside analog, showed protective efficacy against late-onset RVFV encephalitis in mice. The combined use of ribavirin and favipiravir could thus improve the protective efficacy of RVF in the mouse model. Galidesivir (BCX4430), an adenosine nucleoside analog, also showed protective efficacy in mice. In Japan, favipiravir has been approved for the treatment of influenza, whereas it is under a phase 3 clinical trials for influenza in the U.S. BCX4430 has also been studied for safe dosage in phase 1 clinical trials. A number of antiviral candidates have been studied in animal models, but most hits have not been further tested in clinical trials.

Sandfly Fever Naples and Sandfly Fever Sicilian viruses

During the World War II, phlebotomus fever (three-day fever, Pappataci fever) was caused by either SFNV or SFSV among soldiers who were deployed and stayed in the Mediterranean region. SFNV and SFSV are widely distributed in southern Europe, North Africa, the Middle East, and southern and central Asia, and are transmitted via sandflies (*Phlebotomus papatasi*, *P. perniciosus*, *P. perfiliewi*, and others). Phlebotomus fever typically occurs between May and October, when female sandflies feed on human blood. The incubation period for this disease is typically 3–5 days. Phlebotomus fever is characterized by an abrupt onset of fever, headache, myalgia, nausea, retro-orbital pain, conjunctival injection, rash, or leukopenia, which typically lasts for three days. The

febrile illness is self-limiting and patients can fully recover without sequelae. Since the 1980s, few reports of SFNV detection or isolation have been made from human or sandfly samples, with no known reason. The diagnosis of phlebotomus fever can be made via serological tests for IgM, IgG, and neutralizing antibodies, using sera from the acute and convalescent phases. Although the detection of viral RNA or viral isolation can better specify the genotype of the phlebovirus, the short viremic phase may hamper sensitive detection of viruses. Most cases of phlebotomus fever do not require specific treatment.

Toscana Virus

TOSV was first isolated in Italy in 1971, and its presence has been further confirmed in many European and African countries (Table 1). Phylogenetically, TOSV genome sequences can be clustered into three lineages: Lineage A (Italy, France, Algeria, Tunisia, Turkey), lineage B (France, Spain, Portugal, Morocco, Turkey), and lineage C (Croatia, Greece). While the vectors of sandfly viruses are *P. perfliewi* and *P. perniciosus*, TOSV has also been isolated from other species, including *Sergentomyia minuta*, *P. longicuspis*, *P. sergenti*, and *P. neglectus*. The incubation period of TOSV infection is 3–7 days, and the disease is characterized by an abrupt onset of fever, headache, myalgia, nausea, and due to viral aseptic meningitis, neurological signs such as neck rigidity, Kernig sign, reduced Glasgow Coma Scale score, tremors, paresis, and nystagmus. Most cases are self-limiting, and the patients usually fully recover within 7 days without sequelae. In rare cases, TOSV infection leads to a severe meningoencephalitis, with clinical signs such as seizure, paresis, aphasia, hearing loss, deep coma, hydrocephalus, orchitis, or intravascular coagulation. Severely affected patients may exhibit long-term sequelae, such as aphasia, blurred vision, deafness, severe headache, or depression. In Tuscany, the seroprevalence of TOSV increases in an age-dependent manner, and 20–30 cases of TOSV meningitis are reported annually, particularly during the summer season. The diagnosis of TOSV infection can be made serologically, as well as via detection of TOSV RNA in blood or cerebrospinal fluid samples. Patients with meningitis or meningoencephalitis may require supportive care in a hospital.

Punta Toro and Cocle Viruses

PTV is transmitted by sandflies (*Lutzomyia* spp.) in Panama (Table 2). The Adames strain of PTV was isolated from a febrile entomologist in Darien Province in 1972, and the Balliet strain was isolated from a febrile soldier undergoing a jungle warfare training in Colon Province in 1966. The genetically-related Cocle virus can also cause febrile illness in humans. In an investigation of acute dengue in Panama in 2009, 13.4% of suspected acute dengue patients, who were negative for dengue virus, were positive for PTV RNA, indicating that PTV infection is an important cause of non-dengue febrile illnesses in Panama. Clinical signs of PTV and Cocle virus infection have been characterized as fever, chills, myalgia, severe headache, retro-orbital pain, and arthralgia, with full recovery without sequelae. Syrian hamsters infected with the Adames strain of PTV show anorexia, ruffled fur, and lethargy, with pathological findings including hepatic and splenic necrosis, and hemorrhage in the duodenum. In contrast, inbred mice exhibit age-dependent and strain-dependent susceptibility to the Adames strain. In rodent models, the Balliet strain did not show notable pathogenicity. The PTV infection model has been used as a surrogate experimental model for the pathogenesis of RVF, as well as for antiviral screening.

Phleboviruses in the Candiru Antigenic Complex

The Candiru antigenic complex consists of at least 13 viruses (Candiru, Morumbi, Serra Norte, Alenquer, Itaituba, Jacunba, Ariqueles, Oriximina, Turuna, Mucura, Maldonado, Echarte, and Nique viruses) found in Brazil, Peru, or Panama (Table 2). Six viruses (Candiru, Morumbi, Serra Norte, Alenquer, Maldonado, and Echarte) have been isolated from febrile patients. Clinical signs of illnesses are considered similar to those of phlebotomus fever.

Chagres Virus

The Chagres virus has been isolated from febrile humans and sandflies in Panama (Table 2). Clinical signs of Chagres virus infection include fever, severe headache, myalgia, arthralgia, and retro-orbital pain lasting for a couple of days. Seroprevalence studies were exhibited in populations in Panama in 1965, with especially high seroprevalence among residents of San Miguel Island.

Huaiyangshan Banyangvirus (Formerly Severe Fever With Thrombocytopenia Syndrome Virus: SFTSV)

SFTS is endemic in northeast to southeast Asia, including China, the Korean Peninsula, and Japan, whereas it was also found in Taiwan, and Vietnam (Table 3). SFTSV was first isolated in 2009, during an investigation of an etiological agent causing unusually high fatality rates in humans via thrombocytopenia, fever, gastrointestinal symptoms, or leukocytopenia in China's Hubei and Henan provinces. Transmission of SFTSV occurs via a bite of an infected tick (e.g., *Haemaphysalis longicornis*, *Amblyomma testudinarium*, *Rhipicephalus microplus*, *Ixodes nipponensis*). In China, SFTS cases occur from March to November, with a peak between May and July. After hatching, larvae of *H. longicornis* attach to hosts and feed on blood for several days (between June and October).

Table 2 Phleboviruses in the new world (mosquito-, sandfly-borne, and other non tick-borne viruses)

Antigenic complex	Virus	Isolations	Geographic distribution	Human diseases
Candiru	Candiru	Human	Brazil	Phlebotomus fever
	Morumbi	Human	Brazil	Phlebotomus fever
	Serra Norte	Human	Brazil	Phlebotomus fever
	Alenquer	Human	Brazil	Phlebotomus fever
	Maldonado	Human	Peru	Phlebotomus fever
	Echarte	Human	Peru	Phlebotomus fever
	Itaituba	Opossum (<i>Didelphis marsupialis</i>)	Brazil	Unknown
	Jacunba	Rodent (<i>Myoprocta acouchy</i>)	Brazil	Unknown
	Ariquemes	Sandflies (<i>Lutzomyia</i> spp.)	Brazil	Unknown
	Oriximina	Sandflies (<i>Lutzomyia</i> spp.)	Brazil	Unknown
	Turuna	Sandflies (<i>Lutzomyia</i> spp.)	Brazil	Unknown
	Mucura	Mosquito (<i>Anopheles triannulatus</i>)	Brazil	Unknown
	Nique	Sandflies (<i>Lutzomyia</i> spp.)	Panama	Unknown
Punta Toro	Punta Toro (Adames strain)	Human	Panama	Phlebotomus fever
	Punta Toro (Balliet strain)	Human	Panama	Phlebotomus fever
	Cocle	Human	Panama	Phlebotomus fever
	Capira	Sandflies (<i>Lutzomyia</i> spp.)	Panama	Unknown
	Campana	Sandflies (<i>Lutzomyia</i> spp.)	Panama	Unknown
	Letica	Sandflies (<i>Lutzomyia</i> spp.)	Columbia	Unknown
	Buenaventura	Sandflies (<i>Lutzomyia</i> spp.)	Columbia	Unknown
Bujaru	Bujaru	Rodent (<i>Proechimys guyannensis oris</i>)	Brazil	Unknown
	Munguba	Sandflies (<i>Lutzomyia umbratilis</i>)	Brazil	Unknown
Tapara	Tapara	Sandflies	Brazil	Unknown
	Urinara	Sandflies	Brazil	Unknown
Frijoles	Frijoles	Sandflies (<i>Lutzomyia</i> spp.)	Panama	Unknown
	Joa	Sandflies (<i>Lutzomyia</i> spp.)	Brazil	Unknown
Aguacate	Aguacate	Sandflies (<i>Lutzomyia</i> spp.)	Panama	Unknown
	Armero	Sandflies (<i>Lutzomyia</i> spp.)	Columbia	Unknown
	Durania	Sandflies (<i>Lutzomyia</i> spp.)	Columbia	Unknown
	Ixcanal	Sandflies (<i>Lutzomyia</i> spp.)	Guatemala	Unknown
Chilibre	Chilibre	Sandflies (<i>Lutzomyia</i> spp.)	Panama	Unknown
	Cacao	Sandflies (<i>Lutzomyia</i> spp.)	Panama	Unknown
Rift Valley fever	Belterra	Sandflies (<i>Lutzomyia</i> spp.), Mosquitoes (<i>Sabethes chloropterus</i>)	Brazil	Unknown
	Salobo		Brazil	Unknown
	Icoaraci	Mosquitoes (<i>Aedes</i> , <i>Anopheles</i> , <i>Culex</i> , <i>Sabethini</i> spp.), Sandflies (<i>L. flaviscutellata</i>)	Brazil	Unknown
Others	Chagres	Sandflies (<i>Lutzomyia</i> spp.), Mosquitoes (<i>Sabethes chloropterus</i>)	Panama	Phlebotomus fever
	Mariquita	Sandflies (<i>Lutzomyia</i> spp.)	Columbia	Unknown
	Itaporanga	Mosquitoes (<i>Culex</i> spp.), Opossum (<i>Caluromys</i> spp.), Birds (<i>Thamnophilus aethiops</i>)	French Guiana, Trinidad, Brazil	Unknown
	Rio Grande	Southern Plains Woodrat (<i>Neotoma micropus</i>)	United States (Texas)	Unknown
	Ambe	Drain fly (<i>Psychodidae</i> sp.)	Brazil	Unknown
	Anhangá	Sloth (<i>Choloepus brasiliensis</i>)	Brazil	Unknown
	Urucuri	Guyenne spiny rat (<i>Proechimys guyannensis</i>) Long-tailed spiny rat (<i>Proechimys longicaudatus</i>)	Brazil	Unknown

Larvae then molt and become nymphs, which then feed on the blood of hosts for another several days (between March and September) after surviving the winter without feeding. Nymphs become adults after dropping off of the host and molting. Adults feed on host blood for 7–14 days (between April and September) and then lay up to 2500 eggs. One study showed that microinjection of SFTSV into *H. longicornis* resulted in transovarial transmission of SFTSV into larvae, nymphs, and adults; ticks in

Table 3 Tick-borne phleboviruses/Banyangviruses

Genetic group	Virus	Isolations	Geographic distribution	Diseases
SFTS/HRTV (Banyangvirus)	SFTS	Ticks <i>Haemaphysalis longicornis</i> <i>Amblyomma testudinarium</i> <i>Rhipicephalus microplus</i> <i>Ixodes nipponensis</i>	China, Korean Peninsula, Japan, Vietnam, Taiwan	(Humans) Febrile illness with thrombocytopenia, leukopenia, multiorgan dysfunctions (Cheetahs, cats) Hemorrhage in digestive tract, thrombocytopenia, decreased leukocyte counts
	Guertu Heartland	Ticks (<i>Dermacentor nuttalli</i>) Ticks (<i>Amblyomma americanum</i>)	China (Xinjiang) United States	Unknown (Humans) Febrile illness with thrombocytopenia, leukopenia
SFTS/HRTV (unclassified)	Hunter Island group	Ticks (<i>Ixodes eudyptidis</i>)	Australia (Albatross Island)	Unknown
	Malsoor	Bats (<i>Rousettus leschenaultii</i>)	India	Unknown
Bhanja (Phlebovirus)	Bhanja	Ticks (<i>Haemaphysalis</i> , <i>Amblyomma</i> , <i>Rhipicephalus</i> , <i>Hyalomma</i> , <i>Dermacentor</i> spp.) Hedgehog (<i>Atelerix albiventris</i>), Squirrel (<i>Xerus erythropus</i>)	Central and southern Europe, southern and central Asia, Africa	(Humans) Febrile illness, meningoencephalitis (Young ruminants): CNS diseases
	Palma	Ticks (<i>Haemaphysalis punctata</i>)	Portugal	Unknown
	Forecariah	Ticks (<i>Rhipicephalus geigy</i>)	Guinea	Unknown
	Kismayo	Ticks (<i>Rhipicephalus pulchellus</i>)	Somalia	Unknown
	Razdan	Ticks (<i>Dermacentor marginatus</i>)	Armenia	Unknown
	Lone Star	Ticks (<i>Amblyomma americanum</i>)	United States (Kentucky)	Unknown
Kaisodi (Phlebovirus)	Kaisodi	Ticks (<i>Haemaphysalis spinigera</i>) Bird (<i>Zoothera citrina</i>)	India	Unknown
	Lanjan	Ticks (<i>Dermacentor auratus</i> , <i>Haemaphysalis</i> spp., <i>Ixodes granulatus</i>)	Malaysia	Unknown
	Silverwater	Ticks (<i>Haemaphysalis</i> <i>Leporispalustris</i>), Snowshoe hare (<i>Lepus americanus</i>)	Canada (Ontario), United States (Wisconsin, Alaska)	Unknown
	Khasan	Ticks (<i>Haemaphysalis longicornis</i>)	Russia	Unknown
Uukuniemi (Phlebovirus)	Uukuniemi (S23 strain)	Ticks (<i>Ixodes ricinus</i>), Rodent (<i>Apodemus flavicollis</i>), Birds (<i>Turdus merula</i> and others)	Northern and eastern Europe (Finland, Norway, Poland, Czechoslovak, Hungary), Russia	Unknown
	EgAN1825–61	Bird (<i>Phylloscopus trochilus</i>)	Egypt	Unknown
	Chizé	Ticks (<i>Ixodes frontalis</i>)	France	Unknown
	Zaliv Terpeniya	Ticks (<i>Ixodes putus</i>)	Russia	Unknown
	Fin V707	Ticks (<i>Ixodes uriae</i>)	Norway	Unknown
	Murre	Bird (<i>Uria aalge</i>)	United States (Alaska)	Unknown
	RML-105355	Ticks (<i>Ixodes uriae</i>)	United States (Alaska)	Unknown
	Sunday Canyon	Ticks (<i>Argas cooley</i>)	United States (Texas)	Unknown
	Manawa	Ticks (<i>Argas abdussalami</i>)	Pakistan	Unknown
	Grand Arbaud	Ticks (<i>Argas reflexus</i>)	France	Unknown
	Gissar	Ticks (<i>Argas reflexus</i>)	Tajikistan	Unknown
	Rukutama	Ticks (<i>Ixodes signatus</i>)	Russia (Far East)	Unknown
	Komandory	Ticks (<i>Ixodes uriae</i>)	Russia (Far East)	Unknown
	Precarious Point	Ticks (<i>Ixoides uriae</i>)	Australia (Macquarie Island)	Unknown
	Catch-me-cave	Ticks (<i>Ixoides uriae</i>)	Australia (Macquarie Island)	Unknown
	Oceanside	Ticks (<i>Ixoides uriae</i>)	United States (Oregon)	Unknown
	St. Abb's Head	Ticks (<i>Ixoides uriae</i>)	Scotland, UK	Unknown
	Ponteves	Ticks (<i>Argas reflexus</i>)	France	Unknown
	Tunis	Ticks (<i>Argas reflexus</i>)	Tunisia	Unknown

(Continued)

Table 3 Continued

Genetic group	Virus	Isolations	Geographic distribution	Diseases
Mukawavirus (Phlebovirus)	Mukawa	Ticks (<i>Ixodes persulcatus</i>)	Japan	Unknown
Kabutovirus (Kabutovirus)	Kabuto Mountain	Ticks (<i>Haemaphysalis flava</i>)	Japan	Unknown

all three developmental stages can transmit SFTSV into mice via saliva. Moreover, tick colonies derived from SFTSV-infected *H. longicornis* collected in China also showed evidence of transovarial and transstadial transmission.

In addition to the reservoir role of ticks, other animals may serve as amplifying hosts for SFTSV. High seroprevalence of SFTSV antibodies has been shown in sheep, goats, cattle, dogs, and chickens in China. Moreover, viral RNA has been detected in cattle, deer, wild boars, hedgehogs, common shrews, fulvous harvest mice, house mice, and cats. It is therefore likely that those animals may be involved in viral amplification or in the life cycle of *H. longicornis*. It has also been suggested that the geographical distribution of *H. longicornis* is similar to the flyway route of migratory birds, which can harbor this tick, across China, Korea, and Japan. In Japan, a 50-year-old woman who was bitten by a stray cat died from SFTS without any evidence of a tick bite, which suggested cat-to-human transmission of SFTSV. Human-to-human transmission of SFTSV can also occur via exposure to the bodily fluids of an infected patient. Typically, people who have had direct contact with the blood of an index case (e.g., cleaning blood from the face or body, removing tubes or needles), including that of a cadaver, without using personal protective equipment (e.g., gloves, eye protection, mask, gown) could contract SFTS.

SFTSV strains can be phylogenetically clustered into multiple clades, with the L-, M-, and S-segments grouped into genotypes A – F. The majority of SFTSV strains belong to genotypes A, D, and F, which are commonly found in the mainland China. Genotype B is found in Japan, South Korea, and the Zhoushan islands in China, indicating transmission of this genotype among those three countries. Due to the co-circulation of several genotypes, reassortant strains among the A, C, D, and F genotypes have also been isolated. The amino acid similarity among the GnGc proteins of those genotypes is higher than 97%, and SFTS patient sera can cross-neutralize different SFTSV genotypes.

The case fatality rate of SFTS patients is typically 6–33%, with 343 deaths among 5360 laboratory confirmed cases in China from 2011 to 2016; 56 deaths among 172 cases in South Korea from 2013 to 2015; and 394 SFTS cases with estimated 27% case fatality rate in Japan from 2013 to 2019. SFTSV was also found in blood samples from two patients (27- and 28-year-old) in Vietnam and a patient (70-year-old) in Taiwan. Elderly individuals are more susceptible to SFTS, and deaths occurs more frequently in individuals older than 50 years of age. In China, nearly 97% of the abovementioned SFTS patients were farmers, and up to 21% of them could recall receiving a tick bite within the previous two weeks. The incubation period of SFTSV is typically 6–14 days, followed by an initial fever stage, which lasts for 5–11 days and involves high fever (38°C – 41°C), myalgia, headache, fatigue, anorexia, nausea, vomiting, and/or diarrhea. During this period, thrombocytopenia, leukocytopenia, and lymphadenopathy can accompany the clinical signs. At 7–13 days after the disease onset, surviving patients start showing signs of recovery based on increasing platelet counts or a reduction in viral loads. Conversely, disseminated intravascular coagulation and multiple organ dysfunction/failure may occur in severely ill SFTS patients in this stage of the disease. Moreover, hemorrhagic manifestations (e.g., ecchymosis, pulmonary or gastrointestinal bleeding, infarctions), necrotizing lymphadenitis, elevation of liver enzymes (associated with secondary hepatocyte injury other than viral infection), elevation of muscle enzymes, or neurological manifestations (e.g., apathy, lethargy, muscular tremor, convulsions, coma) may also occur in severely ill patients. Bone marrow hematopoietic cells in SFTS patients do not show apparent cytological changes, indicating that thrombocytopenia and leukocytopenia may not occur via direct damage to those cells. The median time from the onset to death is typically 8–9 days. The convalescence phase typically starts 11–19 days after the onset of the disease, and recovery occurs within 3–4 weeks after the onset. Viremia in SFTS patient blood can be detected for 2–3 weeks after the onset of the disease, and it peaks around 6 days after onset.

SFTS can be diagnosed based on the detection of viral L-, M-, or S-segment RNA via quantitative RT-PCR in serum samples. Alternatively, a RT-LAMP assay can be used to detect SFTSV RNA. Serum IgM or IgG antibodies specific to SFTSV proteins can be used to evaluate seroconversion during an SFTSV infection. The presence of neutralizing antibodies against SFTSV also strengthens the serological diagnosis of SFTSV infection.

Little is known about animal diseases caused by natural SFTSV infections. In a zoo in Japan, two cheetahs exhibited anorexia, vomiting, hemorrhage, thrombocytopenia, decreased leukocyte counts, and elevations of alanine aminotransferase (ALT), aspartate aminotransferase (AST), lactate dehydrogenase (LDH), creatinine phosphatase (CPK), and total bilirubin. One of these animals died on day 4 and the other one died on day 7. Post-mortem examination revealed bleeding in the esophagus and ulcers in the gastrointestinal tract. SFTSV was isolated, and viral genome sequences were clustered into a clade with those of SFTSV previously isolated from humans in the same region.

Animal models of SFTS are critical for understanding pathogenesis and allowing the development of vaccines and antivirals. Rhesus macaques experimentally infected with SFTSV developed mild fever, decreased counts of platelets and white blood cells, and elevations of ALT, AST, or LDH, mimicking clinical signs exhibited by patients with nonlethal, mild SFTS. Viral RNA in blood peaked at days 3–5, and it was undetectable by 7 days post infection (dpi). In contrast, cynomolgus macaques inoculated with SFTSV did not show notable clinical signs or viremia, except for a temporary decrease in platelets. Immunocompetent ferrets are uniquely susceptible to

SFTSV infection in an age-dependent manner. Upon SFTSV infection, older ferrets (≥ 4 years old) showed an elevated body temperature, thrombocytopenia, reduced white blood cell counts, and elevated AST and ALT, finally succumbing to infection by 8 dpi. Young ferrets (≤ 2 years old), however, survived the infection without notable clinical signs. Experimental infection of 0.5–2 years old cats with SFTSV via intravenous route resulted in a lethal SFTS-like disease with marked leukopenia and thrombocytopenia in four out of six cats, in which viral antigens were detected in immunoblast-like cells and macrophages in lymphoid tissues.

Immunocompetent adult Balb/c mice, Syrian hamsters, and goats have not shown notable clinical signs in response to experimental SFTSV infection. In contrast, adult C57BL/6 mice inoculated with SFTSV exhibit transiently decreased platelet and white blood cell counts, and elevations of AST, ALT, and blood urea nitrogen, although all infected mice survived the infection without changes in body weight. Interferon α/β receptor (IFN-AR) knockout C57BL/6 or 129/Sv mice are highly susceptible to SFTSV, and typically succumb to infection within 7 dpi. The disease can vary to some extent depending on the virus strains. Pathological findings include gastric and intestinal distensions, lymphoid depletion, and necrosis of lymph nodes or spleen. Viral antigens or RNA are found in reticular cells in the spleen, epithelial cells in the stomach or intestines, or mononuclear cells in various organs. Similarly, STAT2 knockout Syrian hamster strain is highly susceptible to SFTSV infection; death often occurring within 6 dpi when using 10 PFU of SFTSV. The disease in hamsters was characterized by thrombocytopenia, elevations of serum AST and ALT levels associated with multifocal neutrophilic hepatitis, and lymphoid necrosis in the spleen, but leukopenia was less evident in this hamster model.

No licensed SFTS vaccines are currently available for humans. SFTSV GnGc proteins are the targets of neutralizing antibodies, and passive transfer of neutralizing antibodies may protect IFN-AR knockout mice from lethal SFTSV challenge. Further research is required to identify safe and immunogenic vaccine candidates that can induce protective immunity in target populations. Presently, there are no approved treatment regimens for SFTS patients. IFN-AR knockout mice have been used to evaluate potentially effective antivirals. Ribavirin has been shown to partially protect mice from lethal SFTSV infection. In contrast, favipiravir (T-705, per os or intraperitoneal inoculation) could fully protect IFN-AR knockout mice or STAT2 knockout hamsters from lethal SFTSV infection. Further preclinical and clinical trials will potentially lead to the approval of antivirals against SFTS infections.

Heartland Banyangvirus (HRTV)

Another tick-borne pathogenic banyangvirus, HRTV, was first identified in 2009 in two patients in Missouri in the U.S. As of 2018, over 40 cases have been identified in the midwestern and southern U.S., in the states of Missouri, Kansas, Oklahoma, Arkansas, Illinois, Indiana, Kentucky, Tennessee, Georgia, and North Carolina (Table 3). In northern Missouri, serological prevalence of HRTV infection was 0.9%, indicating that some infected patients do not seek for medical care. Among the ten documented clinical cases (patients were 50–80 years of age), eight patients were hospitalized, three of whom died. Nine patients had recently been exposed to tick bites. HRTV has been isolated from the lone star tick (*Amblyomma americanum*), which is thus considered to be the vector for HRTV. Lone star ticks prefer their second developmental phase in the forest, and altogether they require three hosts in their life cycle. In Missouri, the seasonal peak of each stage is as follows: larvae from July to September, nymphs from May to August, and adults from May to July. HRTV RNA has been detected in different pools of depleted, host-seeking nymphs, but not in adults or larvae, indicating that transstadial transmission may occur from infected larvae to nymphs, but it is likely less frequent from nymphs to adults, or from adults to eggs. A serological survey of wild and domestic animals in Missouri in 2012–2013 revealed that neutralizing antibodies specific to HRTV were detected in northern raccoon, horse, white-tailed deer, dogs, and Virginia opossums, indicating that larval or nymphal lone star ticks may be newly infected via viremic blood from those animals. In a large serological survey across the U.S., samples seropositive for HRTV were identified in raccoons (Missouri, Tennessee, Kentucky, Indiana, Texas), white-tailed deer (Florida, Georgia, North Carolina, Vermont, New Hampshire, Maine), coyotes (Illinois, Kansas), and moose (New Hampshire).

The incubation period of HRTV infection is not precisely known, but clinical signs of illness typically start within two weeks after a tick bite. Clinical signs of non-lethal HRTV infections are characterized by fever, fatigue, headache, anorexia, nausea, myalgia, non-bloody diarrhea, leukopenia, thrombocytopenia, and mild to moderate elevations of serum AST and ALT levels. Patients exhibited persistent fatigue or impaired short-term memory, yet they tended to be fully recovered 4–6 weeks after the onset of symptoms. In two lethal cases, erythrophagocytosis with histiocytic hyperplasia was noted in the bone marrow, spleen, and lymph nodes, indicating the occurrence of hemophagocytic lymphohistiocytosis. HRTV antigens were detected in mononuclear phagocytic cells in those patients, showing erythrophagocytosis in the bone marrow. Moreover, one of the two patients also exhibited viral antigens in multiple tissues, including thalamus, liver, gallbladder, pancreas, heart, lung, intestine, kidney, and testes. Acute gastritis, gastric and duodenal ulcers, hemorrhage in the colon, acute pyelonephritis, atherosclerosis in cardiac arteries and aorta, and systemic amyloidosis were also observed in a different HRTV-infected patient. HRTV infection can be diagnosed by detection of viral RNA in blood or serological tests, including a plaque reduction neutralizing test and an IgM or IgG detection assay (e.g., microsphere immunoassay). There are no licensed vaccines or antivirals against HRTV infection.

To model HRTV infection, several animal species have been tested for susceptibility to viral infection. No detectable viremia occurs in wild raccoons, goats, New Zealand white rabbits, Syrian hamsters, C57BL/6 mice, cynomolgus macaques, or chickens inoculated with HRTV. In contrast, IFN- $\alpha/\beta/\gamma$ receptor knockout Ag129 mice succumbed to infection in a dose-dependent manner. Clinical signs of infected Ag129 mice include ruffled fur, hunched back, hematochezia, and ocular discharge, with viremia progressively increasing until the death. Post-mortem examination showed enlarged pale spleen, hepatic hemorrhage, enlarged

gall bladder, and/or excess peritoneal fluid, with HRTV antigens detected in mononuclear cells in the spleen, hepatic sinusoid, or renal interstitium. STAT2 knockout Syrian hamsters infected with HRTV exhibited viremia and a transient reduction of body weight, and mild elevations of alkaline phosphatase, but most animals survived the challenge. Leukocytopenia and thrombocytopenia are not evident in this model. Histopathological examination of one moribund hamster showed neutrophilic inflammation in the spleen, liver, lung interstitium, heart, and lymph nodes.

Bhanja Virus

Bhanja virus was first isolated from *Haemaphysalis intermedia* ticks feeding on a paralyzed goat in India in 1954. Later, this virus was also isolated from African hedgehogs and African ground squirrels. Bhanja virus is widely distributed throughout central and southern Europe, southern and central Asia, and Africa, with samples isolated in India, Italy, Slovakia, Armenia, the Balkan region including Bulgaria, Romania and Croatia, and Africa including the Central African Republic, Senegal, and Nigeria (Table 3). Accidental laboratory infections with Bhanja virus have been characterized by a sudden onset of myalgia, arthralgia, frontal headache, and photophobia. Clinical signs are then cleared in 2–10 days. Clinical signs exhibited by a patient in Zagreb, who was naturally infected with Bhanja virus, were characterized by an initial phase of fever, back pain, photophobia, and vomiting, followed by an unconscious phase with fever, meningitis, and muscle hypertonia, with gradual recovery over 45 days. Seropositivity has also been observed in man, sheep, goat, and cattle in Europe.

Other Phleboviruses

There are at least 100 distinct phleboviruses, and evidence of related human or animal disease has not been identified for many of them. Most phleboviruses have been isolated from sandflies, mosquitoes, ticks, or mammals. Transmission patterns and associations with human or animal illnesses are largely unknown for these phleboviruses. The current classification, determined via antigenic cross-reactivity using the complement fixation test and plaque reduction neutralization test, involves the following antigenic complexes: Sandfly Fever Naples, Sandfly Fever Sicilian, Salehabad, Rift Valley fever, Karimabad, Candiru, Punta Toro, Bujaru, Tapara, Frijoles, Aguacate, Chilibre, SFTS/Heartland, Bhanja, Kaisodi, and Uukuniemi. Genetic grouping had been alternatively applied to those that could not be readily analyzed by antigenic characterization.

The Sandfly Fever Naples antigenic complex consists of SFNV and TOSV, as well as the Massilia, Granada, Arrabida, Fermo, Tehran, Zerdali, and Punique viruses, among others. The Massilia and Granada viruses share highly similar L- and S-segments, but the M-segment sequences of these two viruses are distinct, indicating that genetic reassortment may have created these two phleboviruses. Seropositivity for the Granada virus was found in healthy humans in the Granada area.

The Sandfly Fever Sicilian antigenic complex consists of SFSV, as well as the Sandfly Fever Turkey (SFTV), Sandfly Fever Cyprus (SFCV), Corfou, Toros, and Dashli viruses. SFCV and SFTV were isolated during outbreaks of phlebotomus fever in Cyprus in 2002 and in Turkey in 2008, respectively. The SFTV infection also caused meningitis or encephalitis in some patients. In Greece, viral RNA similar to that of Corfou virus was detected in cerebrospinal fluid of a patient with an acute encephalitis, and this unisolated viral strain was provisionally named Chios virus. Toros and Dashli viruses were isolated from sandflies, whereas human illnesses associated with these viruses have not been reported.

The Salehabad phlebovirus antigenic complex consists of at least Salehabad phlebovirus, Arbia virus, Adria virus, Medjerda virus, Adana virus, Alcube virus, Ponticelli virus, Arumowot virus, and Odrenisrou virus. There is no clear evidence of a human disease associated with these phleboviruses. A partial L-segment RNA of Adria virus was, however, detected in Greece in the blood of a hospitalized febrile 2.5-year-old child who exhibited fever, vomiting, and tonic seizures. Adria virus has never been isolated, yet its RNA has been detected in two pools of sandflies in Albania. Arumowot virus (AMTV), which is proposed to be a member of the Salehabad antigenic complex, has shown seropositivity in humans in Somalia, Sudan, and Egypt, as well as in sheep in Burkina Faso. AMTV is transmitted by mosquitoes and distributed throughout several African countries that overlap with the RVF endemic area. In contrast to RVFV, AMTV has not been experimentally found to cause any lethal diseases in adult mice, hamsters, or sheep. The functional incompatibility of respective N, L, and GnGc proteins also hampers the formation of reassortant strains between RVFV and AMTV.

Guertu banyangvirus (GTV), in the SFTS/Heartland antigenic group, was isolated from *Dermacentor nuttalli* ticks in Guertu county in Xinjiang, China in 2014, and it shows a distinct but close phylogenetic relationship with other SFTSV genotypes. GTV and SFTSV can be cross-neutralized by each other's antisera due to conserved antigenic epitopes. Seropositivity has been shown in populations in Xinjiang, although associated clinical diseases have not been characterized in humans. In contrast, Hunter Island Group virus, which was isolated from *Ixodes eudyptidis* ticks parasiting in shy albatrosses on Albatross Island, Australia, and Malsoor virus, which was isolated from wild bats in India, are phylogenetically close to SFTSV/HRTV, yet they have not been shown to cause human or animal diseases.

The Bhanja antigenic group consists of at least Bhanja virus, Forecariah virus, Kismayo virus, Razdan virus, Palma virus, and Lone Star virus. Little is known about the pathogenicity of Palma, Forecariah, Kismayo, and Razdan viruses in humans and animals. Lone Star virus was isolated in 1967 from lone star ticks (*Amblyomma americanum*) collected from woodchucks in Kentucky, but no association with a human disease has been identified.

The Kaisodi and Uukuniemi antigenic complexes consist of tick-borne phleboviruses distributed worldwide, with no clear association with human or animal diseases. Antigenically, the Uukuniemi group has at least two complexes – the Uukuniemi

antigenic complex and the Murre antigenic complex. The former includes at least Uukuniemi virus (S23 strain), EgAN1825–61 virus, Chizé virus, Zaliv Terpeniya virus, and Fin V707 virus, and the latter includes at least Murre virus, RML-105355 virus, and Sunday Canyon virus. The Manawa, Grand Arbaud, and Precarious Point viruses are antigenically distinct from the others. Other Uukuniemi-like phleboviruses, such as the Catch-me-cave, Oceanside, Ponteves, and St. Abbd Head viruses, have not been sufficiently characterized to allow their antigenic grouping. The Kabuto Mountain virus and Mukawa virus, which were isolated from ticks in Japan, are classified as independent species within the *Kabutovirus* and *Phlebovirus* genera, respectively.

A recent study showed further genetic diversity of tick-borne phleboviruses. Many novel phleboviruses (e.g., Blacklegged tick phleboviruses) have not been successfully grown in culture cells, although viral RNA (L- and/or S-segments) has been amplified from total RNA collected from ticks. Interestingly, M-segment RNA could not be detected in those studies. Moreover, N genes appear to be longer, and NSs genes much shorter, than those of other phleboviruses. There are a number of phleboviruses for which there is only limited information regarding their pathogenicity or transmission to humans or animals. Further studies are needed to better understand the life cycle of each phlebovirus and their potential impact on public health.

Further Reading

- Abudurexiti, A., Adkins, S., Alioto, D., *et al.*, 2019. Taxonomy of the order *Bunyavirales*: Update 2019. *Archives of Virology* 164, 1949–1965.
- Alkan, C., Bichaud, L., de Lamballerie, X., *et al.*, 2013. Sandfly-borne phleboviruses of Eurasia and Africa: Epidemiology, genetic diversity, geographic range, control measures. *Antiviral Research* 100, 54–74.
- Brault, A.C., Savage, H.M., Duggal, N.K., *et al.*, 2018. Heartland virus epidemiology, vector association, and disease potential. *Viruses* 10, 498.
- Gai, Z., Liang, M., Zhang, Y., *et al.*, 2012. Person-to-person transmission of severe fever with thrombocytopenia syndrome bunyavirus through blood contact. *Clinical Infectious Diseases* 54, 249–252.
- Matsuno, K., Weisend, C., Travassos da Rosa, A.P., *et al.*, 2013. Characterization of the *Bhanja* serogroup viruses (*Bunyaviridae*): A novel species of the genus *Phlebovirus* and its relationship with other emerging tick-borne phleboviruses. *Journal of Virology* 87, 3719–3728.
- Mendoza, C.A., Ebihara, H., Yamaoka, S., 2019. Immune modulation and immune-mediated pathogenesis of emerging tickborne banyangviruses. *Vaccines* 7, 125.
- Nunes-Neto, J.P., Souza, W.M., Acrani, G.O., *et al.*, 2017. Characterization of the *Bujaru*, *frijoles* and *Tapara* antigenic complexes into the sandfly fever group and two unclassified phleboviruses from Brazil. *Journal of General Virology* 98, 585–594.
- Palacios, G., da Rosa, A.T., Savji, N., *et al.*, 2011a. Aguacate virus, a new antigenic complex of the genus *Phlebovirus* (family *Bunyaviridae*). *Journal of General Virology* 92, 1445–1453.
- Palacios, G., Tesh, R., Travassos da Rosa, A., *et al.*, 2011b. Characterization of the *Candiru* antigenic complex (*Bunyaviridae*: *Phlebovirus*), a highly diverse and reassorting group of viruses affecting humans in tropical America. *Journal of Virology* 85, 3811–3820.
- Palacios, G., Savji, N., Travassos da Rosa, A., *et al.*, 2013a. Characterization of the *Salehabad* virus species complex of the genus *Phlebovirus* (*Bunyaviridae*). *Journal of General Virology* 94, 837–842.
- Palacios, G., Savji, N., Travassos da Rosa, A., *et al.*, 2013b. Characterization of the *Uukuniemi* virus group (*Phlebovirus*: *Bunyaviridae*): Evidence for seven distinct species. *Journal of Virology* 87, 3187–3195.
- Palacios, G., Tesh, R.B., Savji, N., *et al.*, 2014. Characterization of the *Sandfly fever Naples* species complex and description of a new Karimabad species complex (genus *Phlebovirus*, family *Bunyaviridae*). *Journal of General Virology* 95, 292–300.
- Palacios, G., Wiley, M.R., Travassos da Rosa, A.P., *et al.*, 2015. Characterization of the *Punta Toro* species complex (genus *Phlebovirus*, family *Bunyaviridae*). *Journal of General Virology* 96, 2079–2085.
- Swanepoel, R., Coetzer, J.A.W., 2004. Rift valley fever. In: Coetzer, J.A.W., Tustin, R.C. (Eds.), *Infectious Diseases of Livestock with Special Reference to Southern Africa*, second ed. Cape Town: Oxford University Press, pp. 1037–1070.
- Tokarz, R., Williams, S.H., Sameroff, S., *et al.*, 2014. Virome analysis of *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* ticks reveals novel highly divergent vertebrate and invertebrate viruses. *Journal of Virology* 88, 11480–11492.

Roseoloviruses: Human Herpesviruses 6A, 6B and 7 (*Herpesviridae*)

Katherine N Ward, University College London, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Allogeneic Refers to the genetic differences between individuals of the same species.

Antibody avidity The binding strength between antibody and antigen. In primary infection the first IgG antibodies produced are of low avidity but with time the immune response matures, specific B cell clones undergo somatic mutation and IgG antibody of higher avidity is produced.

Astrocytes Brain cells that form a supporting framework for neurons. Other properties include proliferation after damage.

Atherosclerosis A disease characterized by thickening and loss of elasticity of arterial walls.

CD4, CD134, CD46 and CD8 molecules These are some of the cell surface antigens of leukocytes also known as CD (cluster of differentiation) antigens. As lymphocytes differentiate, they express different patterns of CD molecules on their cell surface, which can aid in determining the type and maturation stage of the cells being examined.

Chromosomal integration of HHV-6 (CIHHV-6) A condition in which the complete HHV-6 genome is integrated into the host cell chromosome in every nucleated cell in the body and transmitted from mother or father to child in the germ line.

Encephalitis Inflammation of the brain causing a reduced level of consciousness together with other symptoms such as fever and fits. Usually caused by an infection.

Episome An episome is an extrachromosomal replicating genetic element that is not necessary for cell survival. Viral episomes can persist latently in the host cell but have the potential for virus replication.

Febrile fits or convulsions An epileptic seizure occurring in childhood associated with a febrile illness. The peak age of occurrence is in the second year of life.

Febrile status epilepticus A serious complication of febrile fits, convulsions lasting more than 30 min – a medical emergency.

Founder effect The reduced genetic diversity, which results when a population is descended from a small number of colonizing ancestors.

Graft-versus-host disease Reaction occurring when T-lymphocytes present in a graft recognize and attack host cells.

Hematopoietic stem cell transplant (HSCT) Colloquially referred to as bone marrow transplant. A transplant of self-renewing stem cells that are capable of giving rise to all of the formed elements of the blood (i.e., leukocytes, erythrocytes and platelets).

Horizontal transmission The spread of an infectious agent from one individual to another.

Limbic encephalitis Encephalitis with characteristic signs and symptoms originating in the hippocampus in the temporal lobe of the brain.

MHC (major histocompatibility complex) A genetic region encoding molecules involved in antigen presentation to T lymphocytes.

Monoclonal antibodies Homogeneous antibody derived from a single B lymphocyte and therefore each antibody molecule is identical to all others.

Peripheral blood mononuclear cells (PBMC) Any peripheral blood cell having a round nucleus, i.e., lymphocytes and monocytes.

Peripheral Occurring outside of the primary lymphoid organs.

Telomeres Caps at the ends of human chromosomes, which are composed of repeats of a DNA sequence of 6 base pairs. The telomere is maintained (copied) by a specialized enzyme that protects against loss of genetic information.

Temporal lobe epilepsy A chronic disorder of the nervous system characterized by seizures that originate in the temporal lobe of the brain.

Temporal lobe The lower lateral lobe of the cerebral hemisphere; contains the hippocampus.

Vertical transmission Genetic transmission from one generation to another. Also extended by others to include transmission of infection from one generation to the next i.e., through the placenta.

Virus reactivation A term usually applied to herpesviruses which describes a switch from latent infection to productive also known as active infection, i.e., virus replication.

Introduction

In 1986 human herpesvirus 6 (HHV-6) was discovered by accident in lymphocyte cultures designed for the isolation of HIV. The virus was initially named human B lymphotropic virus but was later found to infect and replicate in T lymphocytes (also known as T cells), at which point the name was changed to its present form. Identification of another closely related T lymphotropic virus, HHV-7, followed in 1990. Soon after HHV-6 was discovered, it was realised that viral isolates fell into two readily distinguishable groups, variants A and B, which differed in their molecular, biological and epidemiological properties. The two groups showed different abilities to grow in tissue culture of different T cell lines, specific immunological reactivity with monoclonal antibodies, distinct patterns of restriction endonuclease sites, and specific and conserved interstrain variations in their DNA sequences. In 2011, in the face of accumulating evidence, the International Committee on Taxonomy of Viruses declared that the two variants met the formal definition of separate herpesvirus species, namely HHV-6A and HHV-6B. In this article, wherever possible, the two

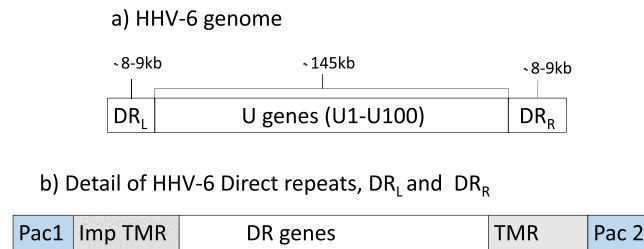


Fig. 1 Overview of HHV-6 genome. (a) Schematic representation of the HHV-6 genome (not to scale). U unique region, DR_L direct repeat left, DR_R direct repeat right. (b) Schematic representation of the DR (not to scale). TMR telomeric repeat, imp TMR imperfect telomeric repeat, pac1 and pac2 packaging sequences.

species of HHV-6 (A and B) are distinguished and the more general term HHV-6 is reserved for studies where either the distinction was not made or there is no clear point to be elucidated.

No disease is causally linked to HHV-6A, and its natural history is unknown. In contrast, primary or first infections with HHV-6B and HHV-7 are ubiquitous in early childhood sometimes causing *exanthem subitum* (roseola infantum); subsequent viral latency gives the potential for reactivation, transmission and disease. HHV-6B reactivation is well established as a cause of encephalitis principally after hematopoietic stem cell transplant (HSCT).

A unique feature of HHV-6 amongst the human herpesviruses is the persistence of the viral genome in human chromosomal telomeres in a minority of individuals with the capacity for transmission from one generation to the next. Telomeres are the essential capping structure of chromosomes that play roles in cancer and ageing raising the possibility that disruption of telomeres by the acquisition of HHV-6 nucleic acid may cause consequential disease. Importantly, the phenomenon of viral integration with its accompanying very high levels of HHV-6 DNA throughout the body has resulted in inappropriate association with disease. It also confounds laboratory diagnosis of active HHV-6 infection potentially resulting in the unnecessary use of antiviral therapy. It was originally considered that chromosomally integrated virus remained latent in the host without reactivation but there is now increasing evidence for viral reactivation and consequent disease. Much remains to be discovered about this fascinating phenomenon.

Biology of the Viruses

Viral Genomics

The virus species, HHV-6A, HHV-6B and HHV-7, belong to the roseolovirus genus of the betaherpesvirus subfamily, so named because of the propensity of HHV-6B and HHV-7 to cause the childhood illness, roseola infantum (*exanthem subitum*). The only other human betaherpesvirus is the more distantly related human cytomegalovirus (HCMV).

Many roseolovirus genes are homologs of core genes shared by all herpesviruses, most of which encode structural proteins or proteins involved in virus replication. The viruses also contain genes unique to betaherpesviruses and those specific to roseoloviruses; encoded proteins include transcription factors controlling the gene expression cascade (see below "Virus Structure and Replication").

Open reading frames (ORF) are named with a prefix of U or DR (Fig. 1 and see below "HHV-6A and HHV-6B"). Table 1 gives a selected list of HHV-6 and HHV-7 genes and gene products. Notably major functions of many roseolovirus genes are known only by inference from that of their homologs in HCMV or other herpesviruses. Only a few genes unique to roseoloviruses have been studied functionally including an immunoevasin encoded by U21, the gQ1 and gQ2 glycoproteins encoded by U100 and lastly the U94 parvovirus rep gene homolog, which is specific to HHV-6A and HHV-6B.

HHV-6A and HHV-6B

HHV-6A (strains U1102 and GS) and HHV-6B (strains Z29 and HST) are 160–162 kilobases (kb) in size. Their complete nucleotide sequences have been published; both genomes are co-linear consisting of a long unique (U) region, containing the majority of genes, bracketed by terminal direct repeats on the left and right (DR_L and DR_R) yielding the arrangement DR_L-U-DR_R (Fig. 1(a)). The few open reading frames (ORFs) in the DR are prefixed 'DR' and those in the unique region U1-U100 from the left to the right hand of the sequence.

The DRs are bounded by multiple copies of short sequences that are either identical to the 6 base pair repeat sequences found in human telomeric DNA or imperfect copies of the telomere repeat sequence (TMR or impTMR). At the extreme ends of each DR are "pac" sequences for cleavage and packaging of unit length genomes during virus replication (Fig. 1(b)).

As to the molecular differences between HHV-6A and HHV-6B underlying their recent classification as separate viral species, HHV-6A is predicted to have 102 genes and HHV-6B 97. Not all HHV-6A genes have counterparts in HHV-6B and vice versa. The two genomes have an overall nucleotide sequence identity of 90%. Divergence of specific sequences (e.g., the immediate early, IE, region) is higher than 30% and there are clear functional differences between the IE1 (U90) genes of HHV-6A and HHV-6B. Moreover, even highly conserved sequences with homology higher than 95%, e.g., U94, as well as divergent genes such as U83, are

Table 1 Selected HHV-6 and HHV-7 genes and gene products

<i>Gene/Gene product^a</i>	<i>HHV-6A</i>	<i>HHV-6B</i>	<i>HHV-7</i>	<i>HCMV^b</i>	<i>Core herpesvirus</i>
<i>Nucleic acid metabolism</i>					
U69 ^c /protein kinase confers ganciclovir sensitivity	+	+	+	+	+
U86/ immediate early protein – IE-A (IE2)	+	+	+	+	–
U90/IE-A (IE1)	+	+	+	–	–
U94 ^d /rep protein	+	+	–	–	–
<i>Structural proteins</i>					
U11/tegument phosphoprotein	+ (p100) ^{e,f}	+ (101K) ^{e,f}	+ (pp89) ^f	+	–
U14/tegument phosphoprotein	+	+	+ (pp85) ^f	+	–
U57/major capsid protein	+	+	+	+	+
<i>Virus attachment/membrane fusion proteins</i>					
U39/glycoprotein B (gB)	+	+	+	+	+
U48/glycoprotein H (gH)	+	+	+	+	+
U82/glycoprotein L (gL)	+	+	+	+	+
U100 ^c /glycoproteins Q ^g	+	+	+	–	–
<i>Immunomodulatory molecules</i>					
U12/chemokine receptor	+	+	+	+	–
U21/down regulator of MHC class 1	+	+	+	–	–
U51/chemokine receptor	+	+	+	+	–
U83/potential viral chemokine	+	+	–	–	–

^aKnown function by experimental validation or by analogy with homologous gene.^bHuman cytomegalovirus.^cHomolog of CMV U97.^dHomolog of adeno-associated virus type 2 (AAV-2) *rep* gene.^eOnly 80% amino acid identity between HHV-6A and HHV-6B.^fMajor antigenic target for human antibody response.^gSpliced envelope glycoproteins: gQ1 and gQ2 for HHV-6A and HHV-6B and gQ for HHV-7. Likely target for neutralizing antibodies.

characterized by specific amino acid signatures, which permit distinction between HHV-6A and HHV-6B. At present, the full implications of genomic and proteomic differences between the two species are unknown although it is clear that such differences may have profound biological effects.

In contrast, there is very little variability within strains of HHV-6A or HHV-6B. For instance, the IE1 gene is highly conserved (> 95%) within clinical isolates of each species.

HHV-7

The full sequences of three isolates of HHV-7 (strains JI, RK and UCL-1) have been published. The genome is arranged in the same way as HHV-6 with a U region flanked at either end by DR regions bounded by telomeric repeats but is smaller than that of HHV-6A/B, namely 144–153 kb with fewer genes, namely 86. Over 90% of the ORFs of HHV-7 are similar in amino acid sequence to those of HHV-6A and HHV-6B. There are only a small number of HHV-6 genes without a counterpart in HHV-7 and vice versa.

Virus Structure and Replication

In common with all herpesviruses, HHV-6 and HHV-7 have a linear double-stranded DNA genome surrounded by an icosahedral capsid separated by amorphous material, the tegument, from an outer envelope containing viral glycoproteins. HHV-6 and HHV-7 are characterized by growth in T lymphocytes, although as described below (Roseolovirus tropism) they can infect other cell types. Replication is slow in vitro with the production of swollen (cytomegalic) cells and results in cell death.

The entry of roseoloviruses into the cell requires attachment of viral glycoproteins to the target cell receptor, which is CD46 for HHV-6A, CD134 for HHV-6B and CD4 for HHV-7. In the case of HHV-6A and HHV-6B, virus attachment involves a complex of glycoproteins gH/gL/gQ1/gQ2 whereas the virus attachment protein or complex for HHV-7 likely involves gH, gL and gQ. Following attachment, fusion between the viral envelope and the cell membrane occurs by a poorly understood mechanism, which may involve gB and gH. The viral nucleocapsid is then transported through the cytoplasm to the nucleus, likely using the microtubule network.

As for other herpesviruses, once viral DNA is released in the nucleoplasm, the gene expression cascade commences. Roseolovirus genes are expressed in a temporally ordered manner starting with immediate early (IE) genes from the IE-A locus, which

comprises two genetic units, IE1 and IE2, with transcription occurring in the absence of de novo protein synthesis. Dependent on prior synthesis of IE proteins, regulation of gene expression follows starting with early (E) genes (encoding enzymes dedicated to nucleotide metabolism and DNA synthesis) and followed by late (L) genes (encoding virus proteins for structure, e.g., viral capsid proteins, and virus attachment).

During active infection, HHV-6 and HHV-7 genomes are predicted to change from linear to a circular, episomal form with replication of the genome through a rolling circle mechanism. A long continuous DNA molecule known as a concatemer is generated, which contains multiple copies of the viral genome joined end to end. This molecule is cleaved into single viral genomes, which is packaged into nucleocapsids thanks to specific cleavage packaging signals present in the DR regions of the genome. The capsids exit the nucleus, acquiring an intermediate envelope by budding through the inner part of the nuclear membrane, are de-enveloped by fusion with the external part of this membrane and appear coated with tegument in the cytoplasm. This is followed by acquisition of the final envelope carrying viral glycoproteins and transport to the cell membrane in a Golgi-derived vesicle. Mature viruses are released from the cell by exocytosis.

Roseolovirus Cell Tropism

Tropism is the predilection of a virus to invade and reproduce in a particular cell type and is a determinant of its propensity to cause disease.

HHV-6A and HHV-6B

The major IE genes of HHV-6A and HHV-6B have promoters that are highly active in T cells reflecting their principal tropism for activated CD4⁺ T lymphocytes. Primary isolation of either species from a human sample requires co-cultivation with T lymphocytes derived from umbilical cord blood mononuclear cells or peripheral blood mononuclear cells (PBMC). Subsequently some isolates have been adapted to replicate in CD4⁺ T cell lines; HHV-6A strains replicate in HSB-2 and JHhan cells whereas HHV-6B grows in the less differentiated MOLT-3 T cell line. Both HHV-6A and HHV-6B can also infect cultured glial cells of central nervous system origin such as astrocytes, although HHV-6A has the greater capacity being capable of productive replication whereas HHV-6B merely persists at low level. These differences in cell tropism are at least partially explicable because the cellular receptors for the two virus species are different.

The HHV-6A cell receptor, CD46, is a ubiquitous molecule (also known as membrane cofactor protein), which is a regulator of the complement cascade. This molecule is expressed on all nucleated cells but not all cell types are permissive for HHV-6A replication indicating a lack of specific cellular factors important for the viral life cycle.

The HHV-6B cell receptor, CD134, is expressed on activated but not resting naïve T lymphocytes and a secondary co-stimulatory molecule for T cell differentiation.

HHV-7

Tropism is more restricted for this virus, which infects T cells expressing its receptor CD4 and replicates efficiently in the T cell line, Sup-T1.

Chromosomal Integration of HHV-6A and HHV-6B

HHV-6 chromosomal integration was first described in 1993 in 3 individuals with unusual HHV-6 DNA levels so high in PBMC that they could be detected by Southern blotting rather than the very much more sensitive technique of PCR. Fluorescent in situ hybridization (FISH) using HHV-6 specific probes identified integrated viral sequences close to, or in, the telomeres of the short arm of chromosome 17 in each case. Subsequently, the telomeric location was confirmed by direct sequencing of junctions between viral and host cell DNA recovered by PCR from patient's cells. Study of a number of Japanese and other families throughout the world indicated vertical transmission from one generation to the next. Such integration is unique amongst the human herpesviruses in that it occurs naturally in vivo.

It is now clear that chromosomally integrated HHV-6 (CIHHV-6), formerly regarded as an oddity, is in fact relatively common, being found inherited in about 1% of humans. CIHHV-6 is alternatively termed inherited CIHHV-6 (iCIHHV-6). In each case of CIHHV-6, one copy of HHV-6A or HHV-6B is integrated into a chromosomal telomere in every nucleated cell in the body because of Mendelian inheritance. Viral integration is not limited to chromosome 17 as telomeric integration sites have been identified on several different chromosomes by FISH (see Fig. 2 for an example). Integration is normally restricted to a particular chromosome per individual but very rarely two sites, if inherited from both parents.

Studies of a few dozen CIHHV-6 genomes have identified a founder effect suggesting that chromosomal integration into the germline is an extraordinarily rare event i.e., most CIHHV-6 individuals acquired their virus in the remote past thousands of years ago. This hypothesis needs further confirmation by nucleic acid sequencing of many more HHV-6 genomes.

Mechanism of chromosomal integration

Exactly how HHV-6 integrates into the ends of chromosomes is unknown but this region is notoriously unstable. Proteins associated with the telomeric DNA sequence at human chromosome ends carry out and regulate the maintenance and protection

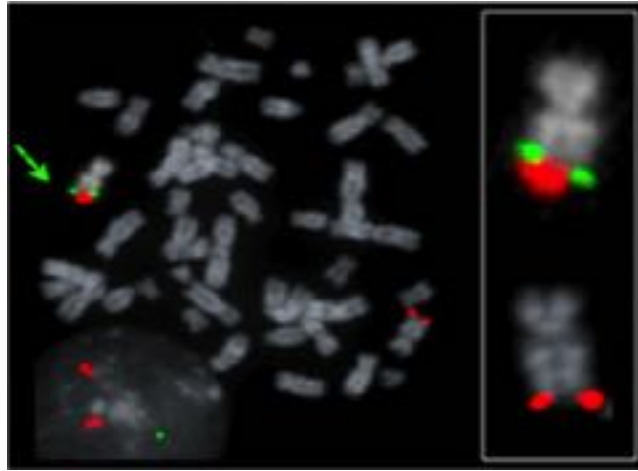


Fig. 2 Fluorescent in situ hybridization (FISH) mapping of HHV-6 chromosomal integration site 9q. Example of dividing cells showing the green signal from the HHV-6 specific probe on one homolog of chromosome 9 (arrow) and the red signal from the 9q telomeric probe on both homologs. Inset right: two enlarged images of the homologs of chromosome 9; the upper image shows superimposed signals from the HHV-6 and 9q telomeric probes and the lower the other homolog with signals from the 9q telomeric probe. Inset lower left: Probes as above; interphase nucleus with uncondensed nucleic acid. CIHHV-6 chromosomally integrated HHV-6.

of telomeres by means of DNA repair and homologous recombination. It is therefore presumed that homologous recombination between the telomere repeat sequences in the viral genome and those present at the ends of human chromosomes is responsible for HHV-6 integration.

Another herpesvirus, Marek's disease virus of chickens, has telomeric repeat sequences similar to HHV-6. Marek's virus is found integrated into the telomeres of avian chromosomes in newly harvested lymphoma cells. Being closely related to HHV-6, HHV-7 also has telomeric repeats at the extremities of the genome but there is no evidence for vertical transmission of this virus suggesting that although such repeats might be necessary they may not be sufficient for integration. It is therefore hypothesized that the HHV-6 U94 rep gene, which is unique amongst the herpesviruses, might be the additional missing factor. This gene codes for a protein, which shares homology with the REP68/78 protein of the adeno-associated parvovirus type 2 (AAV-2) and which is known to be essential for AAV-2 chromosomal integration. Despite this promising hypothesis, U94 is not necessary for the chromosomal integration of HHV-6A into several different cell lines, suggesting that this gene is dispensable for HHV-6A/B integration at least under these conditions.

Viral Latency

HHV-6 and HHV-7 persist in the host throughout life following primary infection with the possibility of subsequent reactivation. Like all herpesviruses, their life cycle has two phases. One is the lytic phase, in which infection of a permissive cell results in coordinated expression of most genes, DNA replication, viral assembly, and release of progeny viruses. Alternatively, the virus infects a cell and becomes latent. Latency is a reversible state in which viral DNA (without viral replication) remains in the cell's nucleus, with limited gene expression.

HHV-6 and HHV-7 can be detected in a small proportion of the PBMC of healthy individuals by PCR when sufficient quantities of DNA are tested. The site of latency for HHV-6A is unknown. Candidates for HHV-6B include monocytes and early bone marrow progenitor cells. For HHV-7, CD4⁺ T helper cells are the proposed site of latency.

HHV-6 and HHV-7 persistence probably includes not only a latent state in cells, such as PBMC, with virions only produced during episodes of reactivation, and chronic replication with continuous or frequent but intermittent production of infectious virus. Salivary glands are a candidate site for chronic replication of HHV-6B and HHV-7 (see "Transmission" below).

State of viral genome during latency

Roseolovirus latency is poorly defined in molecular terms. The assumption was that similar to other human herpesviruses, such as herpes simplex virus, the roseolovirus genomes would be in the form of covalently closed circular episomes. However, the presence of roseolovirus episomes has never been documented. Our understanding of HHV-6A and HHV-6B latency and persistence in the host has undergone significant revision in the light of studies characterising CIHHV-6. It is not formally proven, but entirely possible such integration is an important property of HHV-6 that is the basis of latency after post-natal infection.

Reactivation from latency

In the case of HHV-6B, differentiation of bone marrow progenitor cells might lead to viral reactivation as occurs with the related betaherpesvirus, HCMV. Virus reactivation is common after immunosuppression accompanying organ transplant indicating the

importance of T lymphocytes in controlling HHV-6B virus infection. HHV-6B reactivation occurs after allogeneic HSCT and the disease syndrome, DRESS (see below "HHV-6 and DRESS"), both conditions in which CD134, the receptor for HHV-6B, is upregulated facilitating virus replication.

CIHHV-6 and the possibility of reactivation

The majority of CIHHV-6 genomes contain a full set of intact viral genes and therefore might have the capacity for viral gene expression and full reactivation. There is limited evidence for this possibility in tissue culture after stimulation with drugs such as histone deacetylase inhibitors. Importantly, symptomatic reactivation has been reported in a few cases of immunocompromised patients (see below "CIHHV-6 and Potential for Disease in the Immunocompromised").

It is presumed that for the virus to reactivate from its chromosomal latent state, the genome must revert from linear to circular form so that virus replication can take place. The molecular mechanism for this is currently unknown, although extrachromosomal circular HHV-6 molecules have been reported in a study of human telomeres that carry an integrated form of HHV-6.

Epidemiology

Occurrence

HHV-6B and HHV-7 primary infections

First infection with HHV-6B occurs in almost all children worldwide by the time they are two years old, the peak age of acquisition being between 9 and 21 months. Primary HHV-7 infection is similarly ubiquitous also occurring early in life but usually somewhat later than the latter over the first 5 or 6 years of life.

HHV-6A

In contrast to HHV-6B and HHV-7, natural primary infection with HHV-6A has never been identified and the virus itself is only rarely detected.

Chromosomal integration

In contrast to the almost universal postnatal acquisition of HHV-6B, both HHV-6A and HHV-6B are found chromosomally integrated in a small proportion of the human population (as described above). Although HHV-6A is rare in the general population, it is commonly found chromosomally integrated; one third of HHV-6A cases have CIHHV-6A.

Tissue Distribution

The distribution of HHV-6A in the body differs from that of HHV-6B. Thus, HHV-6B is commonly found in saliva and blood whereas HHV-6A has only rarely been found in these sites. HHV-6B is also the predominant species found in brain tissue.

In vivo HHV-7 is commonly found in blood and saliva but is rarely been found in brain tissue.

Transmission

Horizontal transmission of HHV-6 and HHV-7

It is presumed that HHV-6B and HHV-7 are transmitted between individuals via the saliva of infected individuals. As with the other human herpesviruses, close contact is required i.e., with older siblings and parents rather than casual contacts.

Vertical transmission of HHV-6A and HHV-6B

Two possibilities for the route of congenital infection with these viruses were investigated by Hall and colleagues in Rochester, USA, one being germline, vertical transmission, and the other a conventional infectious process whereby virus crosses the placenta. The results showed that the majority if not all congenital infections are due to Mendelian inheritance of CIHHV-6. One-third of the positives were HHV-6A as opposed to the control post-natal infections all of which were with HHV-6B. The suggested possibility that a small proportion of cases were due to reactivation of virus from its latent chromosomally integrated state and crossing the placenta to infect the fetus remains to be formally proven.

Transmission of HHV-6 by organ donation

Since HHV-6 replicates in T lymphocytes and is likely latent in monocytes and bone marrow progenitor cells, it is to be expected that the virus can be transmitted via blood transfusion and organ transplant.

HHV-6B transmission has been well documented in an infant who received an HSCT from his non-identical twin brother, who had *exanthem subitum* at the time of marrow donation. There are also reports of primary HHV-6 infection acquired by infants after liver transplant from their HHV-6 seropositive mothers.

If an organ donor has CIHHV-6, then integrated virus will be transmitted to the recipient. In the case of allogeneic HSCT from a donor with CIHHV-6A or CIHHV-6B the result is an extremely high HHV-6 DNA load in the blood of the engrafted recipient, which must always be distinguished from active infection (see below 'CIHHV-6 and misdiagnosis of active infection').

Immune Responses

For HHV-6, the tegument proteins, p100 for HHV-6A and 101K for HHV-6B, encoded by U11, are a major target of the antibody response. However, the HHV-6A and HHV-6B proteins have an 80% amino acid identity and substantial antibody cross-reactivity exists. Antibody responses to HHV-7 are directed predominantly against the tegument proteins pp85 and pp89, encoded by U14 and U11 respectively. There is a low level of cross-reactivity between the antibody response to HHV-6 and HHV-7 (see below "HHV-6 and HHV-7 Laboratory Tests and Diagnosis of Infection").

As with other viruses, roseoloviruses are susceptible to neutralisation by antibodies able to block binding to their receptors on host cells. Neutralizing antibodies have been described for a number of HHV-6 glycoproteins including gB, the gH/gL complex and gQ. For HHV-7, neutralizing antibodies to the HHV-7 gQ homolog have been reported.

Whilst inside the cell the virus is effectively hidden from the immune response especially whilst latent. However, viral proteins produced during active infection are at risk of being processed for presentation by major histocompatibility complex (MHC) class I molecules to CD8⁺ cytotoxic T cells with subsequent destruction of the infected cell.

HHV-6, HHV-7 and Immunomodulation

HCMV, the prototypical human betaherpesvirus, encodes many gene products that function to subvert host defense mechanisms. This virus is particularly adept at interfering with antigen processing and presentation by MHC class 1 to cytotoxic T cells. Both HHV-6 and HHV-7 infect CD4⁺ T helper lymphocytes, the cornerstone of the adaptive immune response, and it seems likely that in common with HCMV, HHV-6 and HHV-7 have genes that function to facilitate the avoidance of immune surveillance during primary infection, latency and reactivation. So far, this topic remains largely unexplored but a few genes have been identified that encode immunomodulatory molecules (see [Table 1](#) for some examples). In particular, HHV-6A and HHV-6B encode a down regulator of MHC Class 1.

Based on the above limited evidence, it has been suggested that infection with either HHV-6 or HHV-7 (but primarily the former) interferes with the functioning of the host immune response, and therefore contributes to overall mortality by enhancing the pathogenic effects of agents such as HIV and HCMV or exacerbates graft-versus-host-disease. Further studies are required to establish the clinical impact, if any, of immunomodulation due to HHV-6 and HHV-7.

HHV-6A, HHV-6B and HHV-7 Disease Associations

The association of HHV-6A, HHV-6B and HHV-7 with disease is fraught with difficulty since HHV-6B and HHV-7 are universal in the human population from an early age and both primary infection and subsequent reactivation/reinfection may be asymptomatic. The problem is to distinguish infection from infection that has induced disease. In the case of HHV-6 A and HHV-6B, matters are further complicated by the need to differentiate HHV-6A from HHV-6B and to determine whether CIHHV-6 is present.

HHV-6 in particular has been alleged to cause many different diseases in both immunocompetent persons and those with cellular immunodeficiency but without compelling evidence for an etiological role. No disease has been causally linked to HHV-6A. The only examples of proven disease are HHV-6B and HHV-7 as the cause of exanthem subitum, and HHV-6B as the cause of encephalitis after organ transplant, particularly after HSCT.

CIHHV-6 and Disease

HHV-6 integration can occur in several distinct chromosomes but is invariably telomeric. It is well established that the self-renewal potential of cells is directly proportional to telomere length and telomerase activity. The shortest chromosomal telomere, not average telomere length, is critical for cell viability and chromosome stability. Several diseases are linked with telomere dysfunction and/or telomerase mutations such as cancer, hematopoietic dysfunction, pulmonary fibrosis, liver and cardiovascular disease. Telomeric integration of HHV-6 therefore has the potential to disrupt normal telomere function leading to disease.

CIHHV-6 and potential for disease in the immunocompetent

Neurodevelopment is known to be adversely affected by some viral infections. HHV-6 is closely related to HCMV, which is a major cause of central nervous system disease in the newborn child. A prospective study of children born with CIHHV-6, and possible neurodevelopmental consequences suggested such infection may be associated with cognitive impairment at 12 months of age. Because CIHHV-6 is present in only 1% of births, large population studies are needed.

In a groundbreaking study of almost 20,000 persons in a large general population in Canada, CIHHV-6 was identified as a risk factor for angina pectoris. It was suggested that coronary artery disease might be the result of viral integration leading to telomeric instability, namely deletions and shortening. Alternatively, endothelial cells with integrated HHV-6 might express viral proteins and/or produce virions leading to chronic inflammation and atherosclerosis. Other much smaller studies have not identified a significantly increased frequency of CIHHV-6 in a range of hematological malignancies. Notably no study so far has taken account of the likelihood that integration into different chromosomal sites might have different pathological consequences.

CIHHV-6 and potential for disease in the immunocompromised

There is limited evidence of symptomatic reactivation of CIHHV-6 in three different reports. One in a child with X-linked severe combined immunodeficiency after HSCT, a second in a patient treated with a histone deacetylase inhibitor and a third in a patient who received a liver transplant from a donor with CIHHV-6. Despite the above case of reactivation with accompanying morbidity post-HSCT, this has not been reported in the few other cases where CIHHV-6 was identified in the donor or recipient and the frequency and type of diseases caused by CIHHV-6 in HSCT recipients remain unknown.

HHV-6B and HHV-7 Primary Infection and Disease

Exanthem subitum

This usually benign rash illness of early childhood begins with a high fever followed by a rose pink, non-pruritic, macular rash; the rash is commonly mistaken for measles, rubella or an allergic reaction to antibiotics. The disease was given its name because of the sudden appearance of the rash (subitum means sudden in Latin). Primary infection does not cause exanthem subitum in all affected children but most have some symptoms ranging from fever, to febrile convulsions, to febrile status epilepticus (FSE). The severity of disease depends on the population studied. Notably, 20% of cases of fever in children aged 6–8 months that come to medical attention are due to primary HHV-6B infection.

Encephalitis

In the USA and the United Kingdom, primary HHV-6B and to a lesser extent primary HHV-7 infections are an important cause of complex febrile convulsions and FSE. In Japan HHV-6B is also reported as a cause of encephalitis with an incidence estimated as 5.5/100,000 infants. This difference may be due to a genetic susceptibility of the Japanese. One possibility is that the thermolabile phenotype of carnitine palmitoyltransferase 2, which occurs only in Japanese, is a predisposing factor for encephalitis, with it being inactivated during the high fever characteristic of primary HHV-6B infection; the enzyme is involved in the production of ATP during mitochondrial fatty acid oxidation and its dysfunction would produce an energy crisis.

FSE and temporal lobe epilepsy

Primary HHV-6B and HHV-7 infections are a common cause of FSE. Although the short-term outcome of FSE is good, it has long been debated whether this condition predisposes in the long-term to temporal lobe epilepsy (TLE), also known as mesial temporal lobe epilepsy. TLE is one of the most common and intractable form of seizure disorders. FSE may cause hippocampal injury and HHV-6B DNA has been identified in samples of resected temporal lobe in astrocytes from patients with TLE. Whilst the role, if any, of HHV-6B and HHV-7 in epilepsy remains unresolved, it is noteworthy that after HSCT HHV-6B limbic encephalitis (see below “HHV-6B Encephalitis After HSCT”) predisposes to TLE.

HHV-6 and HHV-7 Disease in Older Children and Adults

Delayed primary infection

Almost all children are infected with HHV-6B by their second birthday and primary infection delayed until after this point must be extremely rare. On the other hand, HHV-7 encephalitis due to delayed primary HHV-7 infection, defined as at or beyond 6 years of age, proven by antibody testing and exclusion of other causes has been reported in one adult and two adolescents. It is tempting to hypothesize that, as when primary Epstein-Barr virus (EBV) infection is delayed beyond early childhood, the consequences of infection are more severe, i.e., infectious mononucleosis in the case of EBV.

HHV-6 and HHV-7 and the possibility of encephalitis

After primary infection, HHV-6 and HHV-7 persist in the host and, when reactivated, there is the potential to cause disease. Encephalitis is a rare but serious disease in immunocompetent older children and adults. The most common cause is herpes simplex virus and the key diagnostic test is detection of specific viral DNA in cerebrospinal fluid (CSF). Similarly, both HHV-6A and HHV-6B DNA have been concluded to be a cause of encephalitis based on the finding of viral DNA in the CSF. However, the diagnosis is not straightforward due to the phenomenon of CIHHV-6; in this condition, inevitably HHV-6 DNA will be present in amounts equal to the number of leukocytes present in CSF but it is not necessarily an indicator of viral replication. Formal proof of suspected HHV-6 encephalitis requires evidence of active virus infection and cell destruction in the brain and this, to date, is lacking in immunocompetent older children and adults.

HHV-6B and DRESS (drug rash with eosinophilia and systemic symptoms)

This disorder is also known as drug-induced hypersensitivity syndrome (DHIS) or drug hypersensitivity syndrome (DHS). DRESS is characterized by a severe multiorgan hypersensitivity reaction that appears 3–6 weeks after exposure to certain therapeutic agents. Rising HHV-6B DNA levels in blood occur after rash onset together with increasing amounts of HHV-6-specific IgG antibody. However, it is unclear whether the virus plays a causal role in this syndrome rather than being a consequence of immune activation.

HHV-6 and multiple sclerosis

Multiple sclerosis is a progressive demyelinating disease characterized by an immune-mediated focal breakdown of myelin sheaths around neuronal axons, causing impaired conduction of nerve impulses and symptoms ranging from blurred vision to paralysis. Viral infections are likely candidates given their associations with encephalitis and demyelination. HHV-6B antigen, a marker of active viral protein expression was first localized to myelin-producing oligodendrocytes in multiple sclerosis lesions in 1995. Since then large numbers of conflicting reports have explored the role of HHV-6 in multiple sclerosis. These studies have utilized a myriad of diagnostic techniques in different sample populations rendering data interpretation challenging.

Alzheimer's disease

There is growing evidence that inflammation plays an important role in this disease, with chronic brain infection by neurotropic agents being one plausible cause of neuroinflammation. There is provocative new evidence that HHV-6A and HHV-7 RNA levels are high and correlate with amyloid plaque and neurofibrillary tangle densities as well as clinical dementia ratings in brain samples from patients with Alzheimer's disease. Although this suggests that a relationship might exist between these viruses in the brain and Alzheimer's disease, direct involvement in causing the disease is not supported and must remain speculative.

HHV-6B and Disease After Organ Transplant

Since HHV-6B infection is almost universal in older children and adults, it is not surprising that HHV-6B reactivation is common after solid organ and HSCT. Such reactivations occur in about half of all patients within 2–4 weeks of transplant. Hepatitis, pneumonitis, bone marrow suppression and encephalitis have been reported for liver, lung and heart transplant patients but a synergistic pathological role for HCMV could not be ruled out in all cases. As discussed below, a causative relationship for HHV-6B is most convincing for HSCT patients.

HHV-6B encephalitis after HSCT

Not surprisingly, HHV-6B reactivation is common early after allogeneic HSCT because of the accompanying immunosuppression. Causal association with disease is fraught with difficulty as the described cases are often complex involving several different possible pathological mechanisms or infections. Proof remains elusive for proposed links of HHV-6B to hepatitis and pneumonitis. There is some support for a connection to acute graft-versus-host disease and bone marrow suppression but only fever with rash and encephalitis are agreed to be caused by HHV-6B.

Encephalitis is rare after allogeneic HSCT but the most common cause is HHV-6B, which carries significant morbidity and mortality; individuals at highest risk are those recipients with poor T cell function, i.e., the donor was mismatched as regards histocompatibility, the presence of acute graft-versus-host disease and an unrelated cord blood transplant; the latter is a particularly high risk. In the first described case of HHV-6B encephalitis there was evidence of infection, namely viral protein in astrocytes and neurons, with necrosis and demyelination in the diseased area of the brain. The virus shows a predilection for the hippocampus and patients characteristically present with post-transplant limbic encephalitis (PALE) – the clinical picture includes depressed consciousness, convulsions, confusion and disorientation, often with short-term memory loss. In cerebrospinal fluid, as well as HHV-6 DNA, there may be evidence of inflammation – leukocytes and elevated protein levels. Mortality is high but the virus may also have more subtle effects; a recent prospective study has reported a correlation between HHV-6 reactivation and impaired cognitive function after allogeneic HSCT in the absence of overt encephalitis.

HHV-7 and Disease after Organ Transplant

Although HHV-7 is capable of reactivation in immunosuppressed patients there is little if any evidence of consequent disease.

HHV-6 and HIV

HHV-6 is also frequently detected in patients with HIV but unlike in transplant patients HHV-6 is not considered as a significant cause of morbidity and mortality in this patient population.

HHV-6 and HHV-7 Laboratory Tests and Diagnosis of Infection

Unfortunately, antibody tests cannot distinguish between HHV-6A and HHV-6B due to antigenic cross-reaction between these two closely related species. Virus detection is therefore the most useful test but distinction must be made between viral chromosomal integration versus the very much more usual acquisition of active infection by the classical process of primary infection.

Antibody Detection – HHV-6 and HHV-7

The first tests to be developed used indirect immunofluorescence and these remain the gold standard. Such tests allow evaluation of amount of HHV-6 or HHV-7 IgG antibody and in sequential serum samples, and when modified to detect IgG antibody avidity can distinguish primary (low avidity) from long-standing infection (high avidity). Although different less labor-intensive tests now exist, in particular enzyme immunoassay, systematic comparisons have rarely been carried out. Testing for roseolovirus antibodies is rarely available in diagnostic laboratories.

Primary HHV-6B

IgM antibodies can be detected 5 days after the onset of infection and IgG antibodies by day 7 the latter increasing for up to 3 weeks and persisting thereafter. HHV-6B DNA appears transiently in the blood but subsides with the appearance of antibodies.

Primary HHV-7

Although less well studied the immune response is similar to that of HHV-6; some cases are accompanied by a rise in HHV-6 antibodies due to limited antigen cross reaction between the two viruses.

Virus detection – HHV-6A, HHV-6B and HHV-7

Roseoloviruses can be grown by coculture of a sample of an individual's PBMC with phytohemagglutinin-stimulated cord blood lymphocytes. Infected cells can be individually identified by the use of specific monoclonal antibodies or nucleic acid tests. Nowadays virus isolation tests are not used diagnostically and real time quantitative PCR that distinguishes between HHV-6A and HHV-6B DNA is the preferred assay, although no quantitative threshold has been formally established to differentiate latent infection from active viral replication.

CIHHV-6 and Misdiagnosis of Active Infection

In chromosomal integration, HHV-6 DNA can be found not only in whole blood but also in "cell-free" samples such as serum or plasma and CSF, which contain human chromosomal DNA released from damaged cells during sample preparation. For most viruses detection of significant levels of viral DNA in the blood, especially in serum or plasma, can be taken as evidence of active infection but HHV-6 is an exception as such high amounts may instead be due to viral integration into host leukocyte chromosomes.

For example, after HSCT from a donor with CIHHV-6 there will be asymptomatic elevation of HHV-6 DNA load in the recipient rising to characteristically and strikingly high levels that correlate with leukocyte engraftment. In the past, misunderstanding of this has led to the inappropriate use of antiviral therapy but in CIHHV-6 even prolonged exposure will be ineffective in reducing viral DNA levels. On the other hand, if the stem cell recipient has viral chromosomal integration rather than the donor, the converse occurs as the recipient's leukocytes are replaced by donor leukocytes and the level of HHV-6 DNA in blood drops as donor leukocytes engraft.

Diagnosis of CIHHV-6

CIHHV-6 should always be suspected if there is an abnormally high HHV-6 DNA load in peripheral blood. CIHHV-6 can be confirmed by evidence of one copy of viral DNA/cellular genome. Comparison of two quantitative real-time PCR results (one for HHV-6 and one for a human gene present in all nucleated cells) is acceptable albeit with a significant margin of error due to inherent assay imprecision. Droplet digital PCR is the most accurate method as it gives an absolute number. Other possibilities for confirmation are HHV-6 DNA in either hair follicles or nails; sites where the virus is only found in CIHHV-6, or by FISH demonstrating HHV-6 integrated into a human chromosome.

HHV-6 and HHV-7 and Antiviral Therapy

HHV-6 and HHV-7 are not susceptible to acyclovir as they lack a gene for thymidine kinase but they are sensitive to ganciclovir as they encode a kinase (gene U69) able to phosphorylate this drug. In vitro studies show that all human betaherpesviruses are susceptible to ganciclovir, foscarnet and cidofovir but all have significant side effects. Despite this, these antiviral agents are

commonly and effectively used to treat HCMV infections in the immunocompromised, especially organ transplant patients. Of the three roseoloviruses, it is only HHV-6B that is proven to cause serious disease worthy of antiviral therapy. To date there have been no controlled trials of ganciclovir, foscarnet or cidofovir therapy for HHV-6B. However, there is enough evidence in the literature of reduction of amount of HHV-6B DNA in blood or simply a clinical response to treatment to support recommendation of ganciclovir or foscarnet for treatment of HHV-6B encephalitis; no recommendation can be made for cidofovir. It is also important to remember that antiviral therapy will not reduce the HHV-6 DNA integrated into host chromosomes.

Concluding Remarks

Whereas the natural history of HHV-6A remains unknown, the pathogenic potential of HHV-6B and HHV-7 is well understood for both viruses as a cause of febrile fits in young children and for HHV-6B as the major cause of encephalitis after HSCT. Issues still to be resolved are a proposed link of FSE to TLE, and effective prevention and treatment strategies, such as new antiviral drugs and immunotherapy, to improve the dismal outcome of HHV-6B encephalitis after HSCT.

HHV-6 reactivation has been claimed as a cause of many different diseases although for most there is no compelling evidence for an etiological role. In some cases, HHV-6 may play an immunomodulatory role but alternatively may merely be a bystander reactivated by inflammation or immune reactivation. In other cases, association with active infection and disease may have been wrongly suggested by the phenomenon of HHV-6 chromosomal integration. It is also important any study of disease associations distinguishes HHV-6A from HHV-6B by the use of appropriate tests so that the role of HHV-6A in particular can be more fully explored.

Studies to determine whether CIHHV-6 is significantly associated with disease are at an early stage. Although relatively rare in the human population, it is estimated that 40–70 million individuals worldwide carry a chromosomally integrated copy of HHV-6. Understanding the full pathogenic potential of CIHHV-6 at the level of different chromosomes and distinguishing HHV-6A from HHV-6B will necessitate very large prospective studies.

Further Reading

- Aimola, G., Beythien, G., Aswad, A., Kaufer, B.B., 2020. Current understanding of human herpesvirus 6 (HHV-6) chromosomal integration. *Antiviral Research* 176, 104720. doi:10.1016/j.antiviral.2020.104720.
- Clark, D.A., 2016. Clinical and laboratory features of human herpesvirus 6 chromosomal integration. *Clinical Microbiology and Infection* 22, 333–339.
- Collin, V., Flamand, L., 2017. HHV-6A/B integration and the pathogenesis associated with the reactivation of chromosomally integrated HHV-6A/B. *Viruses* 9, 160. doi:10.3390/v9070160.
- Gravel, A., Dubuc, I., Morissette, G., *et al.*, 2015. Inherited chromosomally integrated human herpesvirus 6 as a predisposing risk factor for the development of angina pectoris. *Proceedings of the National Academy of Sciences of the United States of America* 112, 8058–8063.
- Hall, C.B., Caserta, M.T., Schnabel, K., *et al.*, 2008. Chromosomal integration of human herpesvirus 6 is the major mode of congenital human herpesvirus 6 infection. *Pediatrics* 122, 513–520.
- Hall, C.B., Long, C.E., Schnabel, K.C., *et al.*, 1994. Human herpesvirus-6 infection in children. A prospective study of complications and reactivation. *New England Journal of Medicine* 331, 432–438.
- Huang, Y., Hidalgo-Bravo, A., Zhang, E., *et al.*, 2014. Human telomeres that carry an integrated copy of human herpesvirus 6 are often short and unstable, facilitating release of the viral genome from the chromosome. *Nucleic Acids Research* 42, 315–327.
- Krug, L.T., Pellett, P.E., 2014. Roseolovirus molecular biology: recent advances. *Current Opinion in Virology* 9, 170–177.
- Pellet, P.E., Ablashi, D.V., Ambros, P.F., *et al.*, 2012. Chromosomally integrated human herpesvirus 6: Questions and answers. *Reviews in Medical Virology* 22, 144–155.
- Schwartz, K.L., Richardson, S.E., Ward, K.N., *et al.*, 2014. Delayed primary HHV-7 infection and neurologic disease. *Pediatrics* 133, e1541–e1547.
- Ward, K.N., 2005. The natural history and laboratory diagnosis of human herpesviruses-6 and -7 infections in the immunocompetent. *Journal of Clinical Virology* 32, 183–193.
- Ward, K.N., 2014. Child and adult forms of human herpesvirus 6 encephalitis: Looking back, looking forward. *Current Opinion in Neurology* 27, 349–355.
- Ward, K.N., Hill, J.A., Hubacek, P., *et al.*, 2019. Guidelines from the 2017 European conference on infections in leukaemia for management of HHV-6 infection in patients with hematologic malignancies and after hematopoietic stem cell transplantation. *Haematologica* 104, 2155–2163.
- Ward, K.N., 2013. Human herpesviruses-6 and -7 (HHV-6A, HHV-6B and HHV-7). *Encyclopaedia of Life Sciences*. Chichester: John Wiley & Sons, Ltd, doi:10.1002/9780470015902.a0023616. <http://www.els.net>.
- Yamanishi, K., Mori, Y., Pellett, P.E., 2013. Human herpesviruses 6 and 7. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed., 2. Wolters Philadelphia PA: Lippincott Williams & Wilkins, pp. 2058–2079.

Rotaviruses (*Reoviridae*)

Juana Angel and Manuel A Franco, Pontifical Javeriana University, Bogota, Colombia

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J. Angel, M.A. Franco, H.B. Greenberg, Rotaviruses, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:10.1016/B978-0-12-801238-3.02662-3.

Glossary

Antigenemia Presence of viral antigen in blood.

CFTR Cystic fibrosis transmembrane conductance regulator, a chloride channel localized in the apical membrane of epithelial cells implicated in secretory diarrhea.

Genotype Specific genetic makeup of one or more viral genes determined by sequence comparison.

Intussusception Pathological event in which the intestine acutely invaginates upon itself and becomes obstructed, followed by local necrosis of gut tissue.

Serotype Significant differences in the antigenic composition of the neutralizing antigens, VP4 and VP7 in the case of rotavirus.

Transcytosis Active transport by which polymeric IgA and IgM antibodies are transported from the basolateral to the lumen of the intestine by the polymeric Ig receptor.

Classification

Using electron microscopy, rotaviruses were discovered as the etiological agents of epizootic diarrhea of infant mice (EDIM) in 1963 and of calf scours in 1969. Employing this same technique, Ruth Bishop identified a rotavirus (RV) in intestinal biopsies of children with diarrhea in 1973. Rotaviruses are classified in the genus *Rotavirus* of the family *Reoviridae* that comprises icosahedral, nonenveloped viruses with segmented, double-stranded RNA (dsRNA) genomes. Based primarily on sequence and antigenic differences of VP6, rotaviruses are classified into ten groups (A–J). Most human pathogens fall into groups A, B, C, and H. The information presented in this article is mostly limited to group A rotaviruses which are, by far, the most prevalent in humans. Serotypes within each group are defined by epitopes in VP7 (glycoprotein, G types) and VP4 (protease, P types) that induce neutralizing antibodies. As the genes encoding these two proteins segregate independently during genome segment reassortment, a binary serotyping system has been developed to identify isolates.

Genotypes, determined by nucleic acid sequence similarity of genes encoding VP7 and VP4, and serotypes, determined by antigenic similarity in VP4 or VP7, as tested by neutralization assays, are generally equivalent for VP7. Currently, 36 G genotypes have been described among group A rotaviruses. G1, G2, G3, G4, G9, and G12 constitute the most prevalent G serotypes of humans detected globally and appear to be equally virulent. For VP4, direct relationship between genotypes and serotypes is unclear and, therefore, a dual system for P classification is in use. The RV classification-working group endorsed the following nomenclature: first, the G serotype/genotype (“X” if it is unknown), second, the P serotype, and lastly, the P genotype denoted in brackets. At least 51 P genotypes, that have been further grouped into 5 serogroups (P[I]–P[V]), and 14 serotypes have been described. The predominant circulating human RV strains encode the P[4], P[6], and P[8] genotypes. P[4] and P[8] genotypes correspond to two subtypes of the P1 serotype (P1A and P1B, respectively) that share some cross-reactive epitopes.

Virion Structure

Rotaviruses were given their name because, when examined by classical electron microscopy, they appear as wheel (*rota*)-shaped, 70 nm particles. However, by cryoelectron microscopy (a method that permits visualization of the viral spikes), the viral diameter is 100 nm, with an icosahedral structure $T = 13I$ (Fig. 1), and formed by three concentric layers of proteins. The core comprises three viral structural proteins (VP) and the 11 segments of dsRNA in the viral genome: 120 copies of the scaffolding protein (VP2), and anchored to each segment of dsRNA are the RNA-dependent RNA polymerase (VP1) and the guanylyl transferase and methyltransferase (VP3) (Fig. 1). The intermediate layer is made up of 260 trimers of the structural protein VP6, the most abundant and most antigenic viral protein. Particles with these protein layers are non-infectious but transcriptionally active and called double layered particles (DLP). Their surface has 132 channels of three types that penetrate into the interior of the capsid. These channels are important during viral replication, allowing exchange of compounds in aqueous solution to the inside of the capsid and the export of nascent RNA transcripts. The external layer comprises 260 trimers of the glycoprotein VP7, which depend on bound calcium ions for stability, and 60 trimeric spikes formed by VP4.

Several RV structural proteins and nonstructural proteins (NSPs) have been crystallized, permitting detailed molecular studies of viral physiology. Using this method, the rearrangement of the viral spike protein (VP4) upon trypsin cleavage (a process that greatly enhances viral infection) has been characterized: The cleavage of VP4 by trypsin into VP8* and VP5* (the proximal portion), which remain non-covalently associated with each other on the virion surface, generates a “rigidification” of the spike.

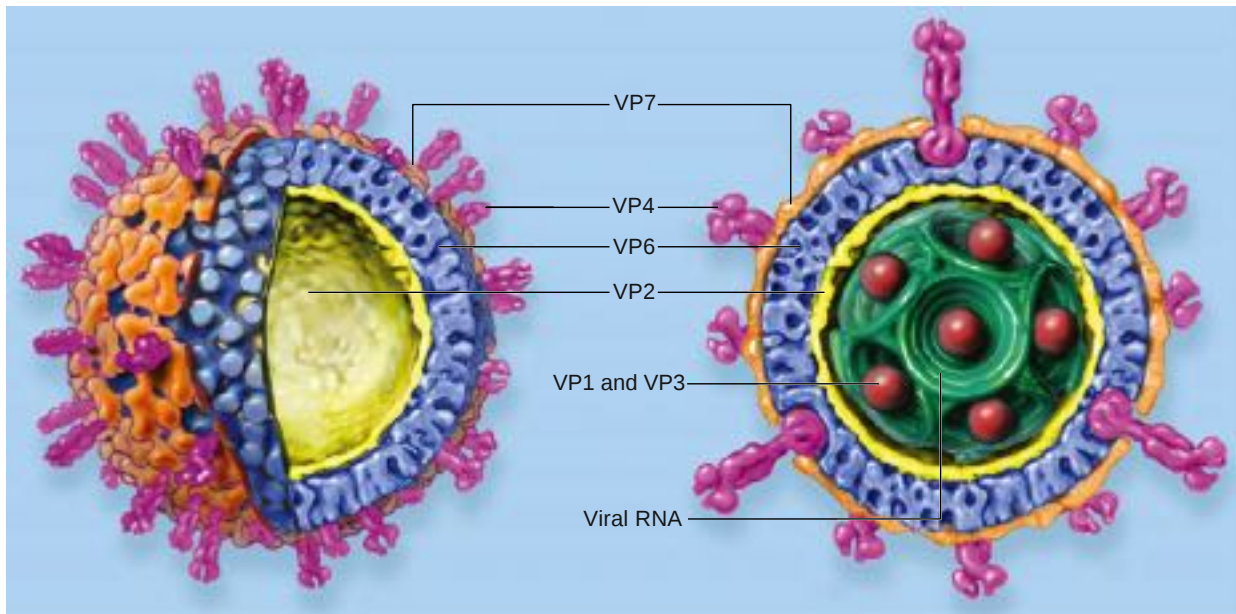


Fig. 1 An artist's reconstruction based on cryoelectron microscopy studies of an RV particle. Shown are the seven structural proteins, and the viral RNA. Reproduced with permission from Andrew Swift, Swift Illustration.

These changes resemble the conformational transitions of membrane fusion proteins of enveloped viruses. Structure-based epitopes for neutralizing antibodies have been identified on the two RV outer layer proteins: for VP7 two unique areas are targeted by neutralizing antibodies, some of which stabilize the capsid and prevent viral uncoating. The viral hemagglutinin VP8* contains at least four virus-neutralizing epitopes and antibodies directed against these epitopes probably block virus attachment. Antibodies directed at VP5* appear to neutralize the virus by inhibiting cell entry during capsid uncoating.

Genome

The RV genome comprises 11 dsRNA segments that code for six structural proteins and six nonstructural proteins. The size of segments varies from 0.6 to 3.3 kbp. Each segment contains a single long open reading frame (ORF), with the exception of segments 9 and 11, each of which may contain two ORFs. The 5' and 3' ends of the RNA segments have noncoding regions that differ between rotaviruses from groups A, B, and C. These sequences are important in transcription, replication, and reassortment of the virus genome.

The viral RNA itself is not infectious. For this reason, the engineering of recombinant rotaviruses has been very difficult, and this has impeded functional analysis of the viral RNA noncoding sequences and the viral proteins. This problem was partially solved by the development of a cell-free system that supports the synthesis of dsRNA from exogenous mRNA, and by gene silencing using small interfering RNA (siRNA) to specifically inhibit translation of viral proteins. In addition, an entirely plasmid-based reverse genetics system for introduction of site-specific mutations into the dsRNA genome of infectious RV has been very recently developed.

Rotaviruses are highly variable. Three mechanisms for generating genetic diversity have been identified: point mutation, genome segment reassortment, and recombination. Rotaviruses have a high rate of mutation and it has been estimated that, on average, at least one mutation occurs during each genome replication. Animal and human group A rotaviruses can undergo genome segment reassortment *in vitro* and *in vivo*. Reassortment originates in cells simultaneously infected with two different rotaviruses strains and results in progeny viruses that have a combination of genes from each parental strain. Natural reassortment *in vivo* appears to influence the serotypic diversity in humans, especially in less developed countries (see below). Recombination and related rearrangements of viral RNA segments (e.g., partial gene duplication or deletions) probably play a minor role in generating viral diversity but may be important in longer-term viral evolution.

Life Cycle

RV entry into host cells is a multistep process and several molecules have been identified as RV receptors or co-receptors. However, the process has been shown to vary in different viral strains. For example, for several animal RV strains, the first binding step involves the interaction of VP8* with sialic acid, while some human RV strains appear to bind initially to GM1 ganglioside. As a second step, RV binds to the integrin $\alpha 2 \beta 1$ in an interaction mediated by the integrin-binding motif DGE in VP5*. In addition to these two interactions, integrins $\alpha v \beta 3$ and $\alpha x \beta 2$, and the heat shock protein hsp70 have also been shown to be involved at a later step of RV cell entry. These interactions probably induce conformational changes in the two viral surface proteins, allowing viral

entry. It seems that the association of some of these molecules with rafts is important for viral entry. A specific interaction between A-type histo-blood group antigen (HBGA) and human RV VP8* to promote RV cell attachment was observed. This interaction seems important since rotaviruses of the five described P genogroups (P[I]–P[V]), differentially infect humans or animals: P[II] rotaviruses infect humans, while P[III] rotaviruses infect both humans and animals, and viruses of both genogroups have different binding patterns to the type A antigen. Although the clinical significance of this discovery is incompletely understood, persons with Lewis-negative secretor-positive HBGA phenotypes have an augmented susceptibility to P[8] RV strains, compared with non-secretor individuals. Also, genetic differences in HBGA expression may explain the susceptibility of neonates to some RV genotypes and the high prevalence of P[6] RV infections in Africa. Furthermore, research on the importance of HBGA in RV vaccine take and efficacy is ongoing.

Although it was initially proposed that rotaviruses enter by direct penetration, current models favor the hypothesis that virus entry is by endocytosis. Viral particles then traffic to the early endosome, where disassembly of VP4 and VP7 – or viral decapsidation – is postulated to occur, due to low Ca^{2+} concentrations. This event functions as a signal for VP5* rearrangement that allows permeabilization of the membrane of the endosome. Then, the viral transcriptase (DLPs) gains access to the cytoplasm and begins to synthesize viral mRNAs. The mRNAs produced by the transcriptase are exact copies of each genome segment, with a 5'-terminal type 1 methylated cap structure and without a 3'-terminal poly(A) sequences.

The viral mRNAs are translated, giving rise to the structural and nonstructural proteins necessary to complete the viral replication cycle. NSP3 is reported to shut off the synthesis of cellular proteins and induces the preferential translation of viral proteins. These proteins accumulate in the cytoplasm in an electron dense region called the viroplasm (constructed mainly by NSP2 and NSP5), where the viral genome is replicated and the assembly of progeny double-layered particles takes place. The mechanism by which one viral particle assembles with each and only one of the 11 RNA segments is unknown. Synthesis of the dsRNAs occurs following the packaging of viral mRNAs into intermediate precursors of double-layered particles.

Assembled double-layered particles and probably VP4 interact with NSP4, which is synthesized by ribosomes associated with the endoplasmic reticulum (ER), and bud into the ER lumen. In this organelle, the double-layered particles acquire a transient lipid membrane. Then, in a very poorly understood process (probably related to the high Ca^{2+} levels of the ER), the viral particles acquire VP7, locking VP4 into place, and lose the transient enveloping membrane, giving rise to the triple-layered particle.

The physiological mechanism of exit of the mature triple layer viral particle from the cell is unknown. In polarized Caco-2 cells (intestinal epithelial cells derived from a human colon adenocarcinoma), RV is released, at least in part, before cell death using a vesicle-associated vectorial transport system, which bypasses the Golgi apparatus and lysosomes, to the apical pole. However, rotaviruses are lytic viruses, and also exit the cell after cell lysis. The mechanism of cell death induced by RV is incompletely understood. Results in polarized Caco-2 cells and in vivo studies in mice suggest that it is by viral induced cell apoptosis.

Up until recently, most in vitro cellular models of RV infection comprised transformed cell lines and animal RV. Novel insights in human virus-cell interactions are being achieved using human intestinal enteroids produced from self-organizing and self-renewing stem cells, isolated from crypts of intestinal samples, that differentiate to enterocytes, goblet cells, enteroendocrine cells, and Paneth cells, modeling human gastrointestinal epithelium.

Epidemiology

Before the introduction of RV vaccines, a meta-analysis presented in 2008 estimated that RV-induced GE caused around 453,000 deaths of children under 5 years old, accounting for 5% of children's deaths of all causes and 37% of those attributable to GE worldwide. More than 90% of these deaths occur in developing countries of sub-Saharan Africa, and South Asia. Although deaths due to RV are rare in developed countries, the incidence of viral infection is the same as in developing countries and health costs associated with RV disease are considerable. In the United States, for example, prior to the introduction of vaccine, an estimated of 58,000–70,000 rotavirus-associated hospitalizations had occurred each year and cost-effectiveness studies clearly justified the use of RV vaccines. After introduction of RV vaccines, norovirus has become the principal cause of medically consulted visits due to GE in the United States.

The multisite MAL-ED cohort study of children aged 0–2 years, performed at eight low-resource settings in different continents, assessed the etiology of diarrhea of any severity using quantitative PCR for an extended range of enteropathogens. The results published in 2017 showed that in the first and second year of life the incidence of diarrhea was 304.9 and 242.7 episodes per 100 child-years, respectively. RV was the principal pathogen identified in the first year of age and the fourth in the second year of age (23.9 and 17.7 episodes per 100 child-years, respectively). In two of eight settings RV vaccine had been introduced to the national vaccine programs, and the total incidence of RV disease was less than half, compared with that of settings where vaccination was absent (4.2% versus 8.9%).

Overall, strain variability is less than would be expected from a random assortment of G and P genotypes since human RV strains belonging to G1, G3, and G4 serotypes are preferentially associated with P[8] while G2 serotype strains are most frequently associated with P[4] genotypes. Besides, six strains of RV represent more than 90% of worldwide circulating species A RV in children necessitating medical attention: G1P[8], G2P[4], G3P[8], G4P[8], G9P[8], and G12P[8].

In addition to these more prevalent strains, rotaviruses with unusual serotypes circulate and can be identified sporadically in developed and particularly in developing countries. Several serotypes may coexist within a community, but, especially in temperate climates, each season is usually dominated by a single serotype that may change from season to season. The global distribution of

rotaviruses varies over continents. Prior to the introduction of RV vaccines G1P[8] represented over 70% of RV infections in North America, Europe, and Australia, but only about 30% of the infections in South America and Asia, and 23% in Africa. These differences in geographic distribution can probably be explained by variations in sanitary and climatic conditions and/or closer contact of individuals with animal rotaviruses in areas with more RV diversity. From 2000–2007 the most prevalent RV strain in India was G12, and during this period other strains such as G2P[11] and G3P[11] were infrequently identified. G and P associations of a probable zoonotic origin have also been detected in Africa. From 2000–2010 G12 strains emerged and G2P[4] reemerged in Latin America and the Caribbean countries.

In several cases, human rotaviruses have been found that have genes of animal RV origin (natural reassortment), adding a further dimension to strain diversity. Even though rotaviruses have an important host range restriction, there is rising evidence of zoonotic transmission to humans, although animal origin strains rarely, if ever, persist in a human population. Notably, host range restriction has been exploited for the development of several viral vaccines.

In the temperate zones of the world (mostly developed countries), prior to vaccine introduction, RV infection occurred primarily during epidemic peaks in the cooler months of the year. A yearly wave of rotaviral illness spread over the United States; beginning in the southwest in November and terminating in the northeast in March. A similar phenomenon has been reported to occur in Europe. This pattern has been markedly attenuated in the US since the introduction of vaccine and is unseen in countries within 10° of latitude from the Equator (mostly developing countries), where epidemics occur year-round. Use of mathematical models of RV transmission dynamics suggest that in the United States spatiotemporal variations in birth rate may have explained RV seasonality, and that for countries localized in the tropics, high birth rates and RV transmission, instead of environmental factors, may explain this lack of seasonality. There has been substantial disruption of the RV circulation patterns after the introduction of RV vaccines.

Rotaviruses are highly contagious and spread easily but, unlike certain human bacterial infections that occur disproportionately in developing countries, the prevalence of RV infection is the same in developed and developing countries, implying that sanitation and hygiene are ineffective measures for disease control.

Clinical Features

Rotaviruses are highly contagious since approximately 10^{11} particles per gram of feces are excreted, only 10 or less particles are necessary to infect a new host, and they are very resistant to ambient conditions. In addition, viral excretion in most children lasts for up to 10 days and, in some children, may extend for 2 months after onset of infection. Rotaviruses are mainly transmitted by an oral–fecal route although, in some cases, a respiratory route has been suggested.

In developing countries, the peak incidence of RV disease occurs in children between 6 and 11 months of age. In contrast, in developed countries, the highest incidence is observed in older children (2 years old). This difference is probably related to differences in sanitation in the different settings, but also influence by the patterns of annual viral circulation. Notwithstanding this variation, RV incidence is similar in both developing and developed countries but mortality is mainly observed in developing countries, presumably due to limited access to appropriate health care. The relative protection of infants younger than 2 months of age, which occurs worldwide, could be related to the presence of protective maternal antibodies. Up to 50% of adults caring for children with RV diarrhea can become infected and, of these, 50% develop disease symptoms, which are generally mild.

The primary clinical syndrome caused by RV infection is acute gastroenteritis. After a short incubation period of 48 h or less, children frequently present with vomiting that lasts for 1–2 days. The vomiting is often accompanied with fever (37.9°C or greater). Subsequently, or at the same time, a watery diarrhea appears, (commonly more than six loose stools in 24 h) and high severity, lasting 5–6 days and, if left untreated, frequently induces dehydration. It is estimated that up to 50% of RV infections in children are asymptomatic but some of these may represent second or third exposures.

Children attending day care institutions are at high risk of developing RV-induced diarrhea. Although RV infections in neonatal care units have been classically described as asymptomatic, probably due to the presence of maternal antibodies, severe symptomatic outbreaks have been described. RV strains that induce nosocomial infections in neonatal nurseries are generally different from those circulating in the community. RV infection can cause severe and prolonged disease in children with primary immunodeficiencies, some of whom develop systemically disseminated infections. Acquired immunodeficiency also predisposes to severe RV disease in bone marrow- and liver-transplanted children. The role of RV-induced disease in immunosuppressed adults with HIV seems, at present, less important.

Pathogenesis

Rotaviruses are very well adapted to their host: they replicate very efficiently, sterilizing immunity is not developed and, despite an important host range restriction, many animal hosts exist (Table 1). These characteristics help to explain the high viral prevalence, and suggest that the prevention of severe disease is an appropriate goal for vaccination.

Important viral antigenemia and some level of viremia are observed in the initial phase of RV-induced diarrhea in children and animals. In mice, extraintestinal viral replication occurs commonly during homologous and some heterologous RV infections. Also in this model, the level and location of extra intestinal replication varies between RV strains and low-level replication can occur in several leukocytes subsets. Elevation of hepatic transaminases has been detected in 20% of children with RV gastroenteritis. However, the clinical relevance of these findings is unclear and, in children and in animals, the bulk of RV replication occurs in the mature villus tip cells of the small bowel.

Table 1 Virus-associated factors that contribute to high viral prevalence and reinfections

Characteristic	Comments
Natural infection does not generate sterilizing immunity	Goals of vaccination are to decrease severe disease but not to prevent infection
Multiple animal hosts exist	Eradication does not seem feasible
Short incubation period (1–2 days)	Does not allow time for the recall of high levels of immune effector mechanisms
The entry cell is the same as the cell used for viral replication	Does not allow time for the recall of high levels of immune effector mechanisms
Virus is excreted in high quantities (up to 10^{11} pfu g ⁻¹ of feces)	High levels of viral dissemination in the environment
Up to 30% of children excrete antigen up to 57 days after onset of diarrhea	High level of viral dissemination
High rate of viral mutation and gene reassortment	May permit the virus to evade the immune system. Currently unproven
Over 50% of infections are asymptomatic	RV well adapted to the human host

Note: Reprinted from Franco, M.A., Angel, J., Greenberg, H.B., 2006. Immunity and correlates of protection for rotavirus vaccines. *Vaccine* 24 (15), 2718–2731, with permission from Elsevier.

RV-induced diarrhea probably occurs by multiple mechanisms that vary, depending on the animal species analyzed. Pathological findings in the small intestine of children with RV diarrhea include: shortening and atrophy of the intestinal villi, enterocyte vacuolization with distended cisternae of the endoplasmic reticulum, and mononuclear infiltration in the intestinal lamina propria. However, in children a direct relationship between the extent of histopathology and disease has not been demonstrated. This finding suggests that RV diarrhea can occur without important enterocyte death.

In pigs, RV infection induces an intestinal lactase deficiency and increased lactose in feces induces an osmotic diarrhea. Elevated levels of lactose are also observed in the feces of some RV infected children. Lactase deficiency could be due to RV-induced enterocyte destruction or to alteration in the synthesis or metabolism of disaccharidases. In mice, two mechanisms of diarrhea, independently of enterocyte destruction, have been identified. NSP4, and a derived peptide of NSP4, induce diarrhea in mouse pups but not adult animals, making it the first viral enterotoxin described. However, its role in RV-induced diarrhea in other species remains uncharacterized. Also in mice, RV activates the intestinal autonomous neural system and increases the secretion of water and electrolytes, as well as the intestinal motility. In fact, RV infection and NSP4 can induce secretion of 5-hydroxytryptamine (serotonin) from enteroendocrine cells in humans and mice, which can activate enteric nerves of the small intestine leading to increased intestinal motility. Racecadotril, an inhibitor of the enteric nervous system, is somewhat efficacious in treating RV-induced diarrhea in children, suggesting that this mechanism plays at least some role in human diarrhea.

Diagnosis

Initial efforts to isolate wild-type rotaviruses in tissue culture were unsuccessful and even today it is a complex procedure. For this reason, diagnosis is more commonly made using commercial ELISAs that detect viral antigen (mostly VP6) in the feces. Rotavirus is shed in very large amounts making the ELISA a highly effective, inexpensive, simple and accurate diagnostic assay. For epidemiological studies, human RV strains present in feces can be grown in cell culture in MA104 (African green monkey kidney) and Caco-2 cells using trypsin for enhancement of viral growth during the culture. Trypsin cleaves VP4 and increases infectivity in culture. Genotype characterization of RV strains is mainly by reverse transcription-polymerase chain reaction (RT-PCR). RV detection by electron microscopy is only conducted in research laboratories, especially when other viral pathogens are also being investigated.

Treatment

RV diarrhea is self-limited and treatment is aimed at reducing symptoms until the immune response resolves the infection. Children with mild diarrhea are treated by oral rehydration. Those presenting moderate to severe dehydration may require intravenous rehydration. In cases of severe disease, treatment with probiotics, preparations that contain antibodies against RV, Racecadotril (an inhibitor of the enteric nervous system), and Nitasoxanide (a drug of unknown mechanism of action) have been shown to accelerate resolution of the disease. However, it is unclear whether these interventions truly provide significant advantages to ill children.

Immune Response and Prevention

Immunity to RV is incompletely understood and animal models have been important in the acquisition of our current knowledge. Pepsin and gastric acid seem to be important host defense factors against RV infection, since these factors inactivate rotaviruses in adult mice but not in suckling mice. In addition, the innate immune response and interferons (IFN), in particular, seem to mediate an antiviral effect, and viral mechanisms for evading this response involving NSP1 have been described. However, mice lacking T and B cells become chronically infected with RV, suggesting that the adaptive immune system is essential for viral elimination. Also, children lacking T and B cells, or only B cells have been shown to shed virus chronically.

Several recent studies in mice have addressed the role of IFNs in the immune response against rotaviruses: RV infection of intestinal cells of mouse pups induces expression of mRNA of type I and III IFNs, treatment of mice with these IFNs reduced RV

replication, and in parallel, enhanced RV replication was detected in mice lacking receptors for these IFNs. In agreement with these findings, it has been shown that intracellular RV replication is recognized by RIG-I and/or MDA5, which activate the transcription factors IRF3 and NF- κ B, that in turn induce expression of type I IFN and interferon-secreted genes (ISGs). After RV induced production of IFN, it signals, in an autocrine fashion, through STAT1/STAT2/IRF9 complexes to further induce type I IFNR expression, more IFN and ISG expression, and an antiviral state. Moreover, RV has developed numerous mechanisms to evade the type I IFN response: RV NSP1 triggers degradation of IRF3 and inhibition of NF- κ B activity; RV inhibits the function of STAT1 activation in vitro and in vivo in mice, by an unknown mechanism, possibly inhibiting all types of IFN responses; it significantly depletes IFN-type I, II, and III receptors in vitro and in vivo in mice. When human intestinal enteroids were used, type III IFN-regulated response was the predominant transcriptional pathway stimulated by human RV infection, which is, however, unable to restrict viral replication.

A recent study showed that RV activates another component of the innate immune response, the inflammasome, and Nlrp9b, specifically expressed in murine intestinal epithelial cells, was recognized as the NOD-like receptor (NLR) that restricts murine RV infection.

Intestinal antibodies seem to be the principal mechanism that mediates protection against viral reinfection. In agreement with the fact that most viral replication occurs in the intestine, the localization of virus specific B cells in the intestine seems important for their capacity to mediate protection. It is postulated that local IgA antibodies can mediate expulsion of rotaviruses inside the enterocytes and exclusion (to avoid de novo infection of enterocytes) of rotaviruses in the gut lumen. Neutralizing antibodies to VP4 and/or VP7 can block enterocyte infection directly when present in the gut lumen (exclusion). In mice, anti-VP6 non-neutralizing polymeric IgA may bind virus VP6 during transcytosis from the basolateral membrane of enterocytes to the gut lumen and "expulse it". These antibodies may also inhibit RV transcription intracellularly. In addition, antibodies to NSP4 may block diarrhea, but not infection, by blocking the enterotoxic property of the molecule. However, the antiviral effects of antibodies against VP6 and NSP4 have only been shown to be protective in mice. In piglets, a model that is probably closer to humans than mice, the presence of antibodies against VP4 and/or VP7 seems necessary for protection.

Although local antibodies appear to be the principal mechanism that protects against viral infection, T cells also directly mediate antiviral immunity, at least in mice. CD4 + T cells are also essential for the development of more than 90% of the RV-specific intestinal IgA, and thus their presence seems critical for the establishment of protective long-term memory responses. Moreover, murine RV-specific CD8 + T cells are involved in the timely resolution of primary RV infection and can mediate short-term partial protection against reinfection.

In children from middle- to high-income countries, primary RV infections are generally the most severe, with severity decreasing as the number of reinfections increases. Complete protection against moderate-to-severe illness is achieved after two natural RV infections, whether symptomatic or asymptomatic. However, in a cohort study in India, where close to 22% of deaths due to RV worldwide occur and half of the children were infected by 6 month of age, the severity of diseases decreased between the second and third infection, and 79% of protection against moderate-to-severe illness was observed only after the third infection.

In agreement with the results of animal models mentioned previously, total serum RV IgA, induced by a primary infection, seems to be the best but imperfect correlate of protection against subsequent reinfections. Primary RV infection induces homotypic-neutralizing antibodies to VP7 and VP4 and some level of heterotypic immunity.

The lack of understanding of RV immunity limits our capacity to develop new RV vaccines. Due to the relative inaccessibility for study of the human intestinal immune tissue, cells involved in the RV immune response have mainly been studied in peripheral blood. It is hypothesized that at least a fraction of these lymphocytes are recirculating to and from the intestine, and thus may reflect intestinal immunity. In blood of children with acute RV infection, plasmablasts and plasma cells (secreting mostly RV specific IgM) are found, and these cells express intestinal homing receptors. In the convalescence phase, RV-specific B cells are mainly memory cells, some of which express intestinal homing receptors. In children and adults with acute infection, RV-specific CD4 and CD8 T that secrete gamma interferon circulate, but in unexpectedly low frequencies. It is speculated that the limited induction of RV-specific T cells can be related to the occurrence of delayed viral clearance and symptomatic reinfections in a subset of individuals. Class II T cell tetramers with RV specific epitopes have been described and used to show that a subset of circulating RV-specific CD4 are probably originated in the intestine.

The first commercially available RV vaccine was a simian-human reassortant tetravalent vaccine (RotaShield). Although this vaccine was highly effective, it was withdrawn from the market 1 year after its introduction due to its temporal association with low levels of intussusception (1 in 10,000 children). Two new RV vaccines received worldwide recommendation by the World Health Organization and have now been licensed for use in more than 100 countries. The three dose vaccine produced by Merck (RotaTeg™, RV5) contains mono-reassortants of a bovine virus with G1, G2, G3, G4, and P1A[8] human RV genes. The two dose vaccine produced by GlaxoSmithKline (GSK, Rotarix™, RV1) is a human-attenuated G1P1A[8] virus. In trials that involved over 60,000 infants from middle- to high-income countries, both of these vaccines showed to be safe and to provide over 70% protection against any RV diarrhea and over 98% protection against severe RV infection. As observed for other oral vaccines, both RV vaccines are less efficacious, with 44% protection, against severe RV gastroenteritis in settings with high under-5 mortality rates. Importantly, protection against severe disease generally decreased between the first and second year after vaccination.

A systematic review of the effectiveness of RV1 and RV5 in the first decade of global postlicensure confirmed that vaccine effectiveness in routine programmatic use depends on under-5 mortality rates. The median vaccine effectiveness against RV disease

were: RV1 84% (range, 19%–97%) and RV5 90% (range, 63%–100%) in low-mortality countries; RV1 75% (range, –2% to 94%) in medium-mortality countries; and RV1 57% (range, 18%–69%) and RV5 45% (range, 43%–92%) in high-mortality countries. The median RV1 and RV5 effectiveness against RV infections requiring hospitalizations were higher: 88% (range, 70%–95%) and 94% (range, 83%–100%), respectively. A decline of 35% in RV induced mortality has been observed from 2008 to 2011 in Mexico, and other countries, after introduction of RV vaccination nationwide.

The lower efficacy of oral vaccines in low-income settings remains incompletely understood. To increase RV vaccine effectiveness in these settings, investigators have performed clinical trials with variations in the age of administration, timing of breastfeeding, improvement of nutrition status, and supplementation with zinc, which have, unfortunately, shown none or modest improvement. Even though concomitant administration of live oral polio and RV vaccination is recommended, it interferes with RV vaccines immunogenicity, and consequently, some improvement may arise in settings transitioning from oral polio formulations to parental vaccines.

Rotavirus vaccines postlicensure studies revealed: (a) herd immunity appears to be generated following RV vaccination in high- and middle-income countries; (b) changes in the age of individuals infected with RV (increase in unvaccinated children 6–16 years of age or individuals > 70 years of age) (c) changes in the seasonal pattern of RV disease. Also, in some settings, there is evidence that RV vaccines deploy different immunological pressures that impact, to some extent, the diversity of circulating RV strains. However, the fact that RV1 and RV5 induce wide protection against homotypic (strains with genotypes similar to those included in the vaccine formulation) and heterotypic (strains with genotypes different to those included in the vaccine formulation) strains has been corroborated in a great number of different settings. More studies are necessary to establish if waning of vaccine effectiveness after the first year of life is occurring in the different settings, which may have important consequences for vaccination programs, since a booster dose of rotavirus vaccine may become necessary.

In post-vaccination surveillance programs, a very low (1–6 per 100,000) increased risk of intussusception has been observed in RV1 or RV5 vaccinated children, primarily in the first week after the first dose of vaccines in several high- and middle-income countries. However, health authorities have generally concluded that the benefits of RV vaccination outweigh these small potential risks. Some evidence indicates that the increased risk of intussusception after RV vaccination may be lower in low-income countries.

In China, a lamb RV strain based vaccine (LLR, Lanzhou Institute of Biological Produces) was licensed in 2000, and in Vietnam a human G1P[8] strain based vaccine (Rotavin™, PolyVac) was licensed in 2012. Rotavac™ (Bharat Biotech), a three-dose vaccine containing a natural human-bovine reassortant 116E neonatal strain (G9P[11]) was prequalified by WHO in 2018, allowing its procurement by UNICEF for global use. Rotasiil™ (Serum Institute of India), a three-dose pentavalent vaccine composed of bovine-human VP7 reassortants of the UK strain developed at The NIH, was also recently prequalified by WHO. A neonatal RV3-BB strain based vaccine, generated in Australia, is being evaluated for neonatal vaccination. In addition, current studies in animal models and people are also focused on the development of third generation nonreplicating RV vaccines, such as recombinant RV proteins or virus-like particles.

Further Reading

- Angel, J., Steele, A.D., Franco, M.A., 2014. Correlates of protection for rotavirus vaccines: Possible alternative trial endpoints, opportunities, and challenges. *Human Vaccines & Immunotherapeutics* 10 (12), 3659–3671.
- Bines, J.E., At Thobari, J., Satria, C.D., *et al.*, 2018. Human neonatal rotavirus vaccine (RV3-BB) to target rotavirus from birth. *The New England Journal of Medicine* 378 (8), 719–730.
- Burnett, E., Jonesteller, C.L., Tate, J.E., Yen, C., Parashar, U.D., 2017. Global impact of rotavirus vaccination on childhood hospitalizations and mortality from diarrhea. *The Journal of Infectious Diseases* 215 (11), 1666–1672.
- Crawford, S.E., Ramani, S., Tate, J.E., *et al.*, 2017. Rotavirus infection. *Nature Reviews Disease Primers* 3, 17083.
- Kanai, Y., Komoto, S., Kawagishi, T., *et al.*, 2017. Entirely plasmid-based reverse genetics system for rotaviruses. *Proceedings of the National Academy of Sciences of the United States of America* 114 (9), 2349–2354.
- Kotloff, K.L., Platts-Mills, J.A., Nasrin, D., *et al.*, 2017. Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques. *Vaccine* 35 (49 Pt A), 6783–6789.
- Lee, B., Dickson, D.M., deCamp, A.C., *et al.*, 2018. Histo-blood group antigen phenotype determines susceptibility to genotype-specific rotavirus infections and impacts measures of rotavirus vaccine efficacy. *The Journal of Infectious Diseases* 217 (9), 1399–1407.
- Lopez, S., Sanchez-Tacuba, L., Moreno, J., Arias, C.F., 2016. Rotavirus strategies against the innate antiviral system. *Annual Review of Virology* 3 (1), 591–609.
- Nair, N., Feng, N., Blum, L.K., *et al.*, 2017. VP4- and VP7-specific antibodies mediate heterotypic immunity to rotavirus in humans. *Science Translational Medicine* 9 (395).
- Platts-Mills, J.A., Liu, J., Rogawski, E.T., *et al.*, 2018. Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: A reanalysis of the MAL-ED cohort study. *The Lancet Global Health* 6 (12), e1309–e1318.
- Saxena, K., Simon, L.M., Zeng, X.L., *et al.*, 2017. A paradox of transcriptional and functional innate interferon responses of human intestinal enteroids to enteric virus infection. *Proceedings of the National Academy of Sciences of the United States of America* 114 (4), E570–E579.
- Shah, M.P., Dahl, R.M., Parashar, U.D., Lopman, B.A., 2018. Annual changes in rotavirus hospitalization rates before and after rotavirus vaccine implementation in the United States. *PLoS One* 13 (2), e0191429.
- Tate, J.E., Burton, A.H., Boschi-Pinto, C., Parashar, U.D., N. World Health Organization–Coordinated Global Rotavirus Surveillance, 2016. Global, regional, and national estimates of rotavirus mortality in children <5 Years of Age, 2000–2013. *Clinical Infectious Diseases* 62 (Suppl 2), S96–S105.
- Tate, J.E., Yen, C., Steiner, C.A., Cortese, M.M., Parashar, U.D., 2016. Intussusception rates before and after the introduction of rotavirus vaccine. *Pediatrics* 138 (3).
- Zhu, S., Ding, S., Wang, P., *et al.*, 2017. Nlrp9b inflammasome restricts rotavirus infection in intestinal epithelial cells. *Nature* 546 (7660), 667–670.

Relevant Websites

<https://www.cdc.gov/rotavirus/about/symptoms.html>

CDC.

<https://www.path.org/diarrheal-disease/>

Diarrheal Disease

PATH.

<http://rotacouncil.org/>

Rotacouncil.

<https://rega.kuleuven.be/cev/viralmetagénomics/virus-classification/rcwg>

Rotavirus classification working group: RCWG.

<https://www.who.int/immunization/diseases/rotavirus/en/>

WHO

Rotavirus.

Rubella Virus (*Picornaviridae*)

Annette Mankertz, Robert Koch-Institute, Berlin, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of T.K. Frey, Rubella Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00499-4](https://doi.org/10.1016/B978-012374410-4.00499-4).

Glossary

Birth defects Malformations or abnormalities developed during gestation that are apparent at, or soon after, delivery.

Congenital infection Infection of the fetus during gestation as a result of virus passage through the placenta.

Elimination program Vaccination program designed to decrease the number of susceptible individuals in order to eliminate transmission of a pathogen in a geographic region and limit circulation after imports from other regions.

Enveloped virion Virus particle with a lipid bilayer membrane, or envelope, surrounding the capsid or core. The envelope is usually derived from host cell membranes and contains virus-specific glycoproteins decorating its outer surface.

Icosahedral capsid In the virus particle, the proteinaceous shell surrounding the virus genome – in this case, with symmetry of an icosahedron (appears quasi-spherical or round in electron micrographs).

Live, attenuated vaccine A vaccine formulation using an infectious variant of the virus that has been attenuated or

weakened by repeated passage in cell culture or in alternate hosts. Such vaccines recapitulate the infection, including induction of a complete adaptive immune response, but do not cause disease due to a loss of pathogenic features.

Persistent infection Infection characterized by continued presence of the virus, often despite the induction of an adaptive immune response.

Plus-strand RNA virus Virus with a single-stranded RNA genome that is of messenger RNA sense and can be directly translated to produce virus-specified proteins.

Serodiagnosis Detection of virus-specific antibodies for diagnosis of infection or verification of the immune status. Briefly the presence of IgM antibodies can indicate an acute infection, while presence IgG can be taken as a token of previous infection or immunity.

Systemic infection Infection that spreads from local site entry via the lymphatic system and blood stream to one or more target internal organs or tissues.

Introduction

Rubella virus (RV) infection leads to a benign disease with rash and fever in children and adults. In contrast, infection during the first trimester of pregnancy can result in a miscarriage, fetal death, stillbirth, or infants with malformations, known as a congenital rubella syndrome (CRS). Norman Gregg, an Australian ophthalmologist, first reported the association of congenital cataracts as a consequence of gestational rubella in 1941, establishing RV as a major teratogen. Vaccination programs administering live-attenuated vaccines have been successful in greatly reducing the incidence of both rubella and CRS, particularly in the Americas, where rubella elimination was verified in 2015. In the end of 2014, 140 countries had introduced national rubella vaccination programs (54 had not) and a global coverage with rubella vaccine had increased to 47%. This does not prevent the birth of approximately 110,000 children with CRS annually, as estimated by WHO.

Classification

The classification of RV is currently under reassessment by the ICTV: Until 2019 RV was the only member of the genus Rubivirus (family *Togaviridae*) which comprises mainly arthropod-borne alphaviruses. This classification was based on similarities in virion structure, gene expression strategy, and some sequence similarity between conserved regions of the alphavirus genome and their counterparts in the rubivirus genome. However, RV differs from alphaviruses in its mode of transmission, genome organization, virion structure and also in sequence relationship. Thus, RV has been moved to the family of *Matonaviridae*.

The rubella virion consists of a single-stranded, plus-sense genomic RNA with a variable length of approximately 9760 to 9778 nucleotides. Phylogenetic analysis of RV is based on a reference sequence within the E1 gene (nt 8731–9469). The World Health Organization (WHO) has proposed a standardized taxonomy for RVs which consists of two clades 1 and 2, and 13 genotypes within these two clades. The sequences in these two clades differ by approximately 8%. Clade 1 includes ten genotypes and is distributed worldwide. Clade 2 contains three genotypes and appears to be restricted mostly to Asia. Genotypes 1E, 1G, 1J and 2B currently dominate global circulation.

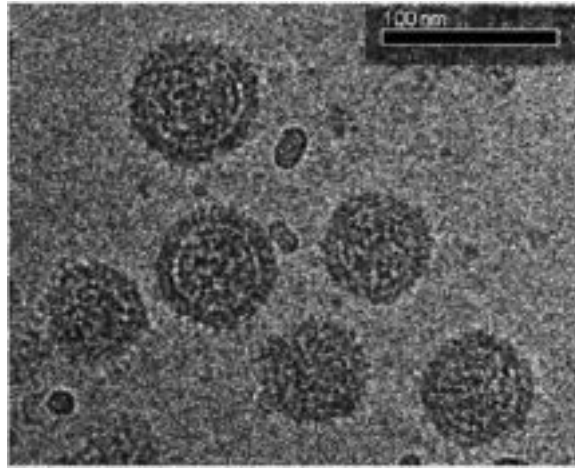


Fig. 1 Cryoelectron micrograph of rubella virions. Unlike conventional negative staining, the virions have a uniform structure when visualized by this technique; however, pleomorphism of particle diameter and the gap between the core and envelope remains apparent. Courtesy of Teryl Frey (unpublished data).

Host Range and Virus Propagation

RV has no known natural host other than humans. No reliable animal model exists for the study of RV pathogenesis. RV replicates in a number of laboratory cell culture lines; however, in most of these no cytopathic effects (CPEs) are usually observed. Continuous cell lines commonly used to propagate RV include Vero (African green monkey kidney), RK-13 (rabbit kidney), BHK-21 (baby hamster kidney), HaCaT cells (human keratinocytes). Also HUVEC (primary fetal endothelial cells derived from human umbilical vein) can be used. In all cell lines in which RV replicates, persistent infections are readily established and maintained, whether or not the cell line exhibits a functional interferon system. Infectious virus is released into the media for several subcultures of infected cells. Persistent RV infection in cell culture cannot be cured by the inclusion of neutralizing antibodies in the culture medium because virus budding occurs at intracytoplasmic locations. Thus, RV is highly adapted for persisting infection.

Properties of the Virion

Rubella virions are 60–70 nm spherical particles composed of an electron-dense core separated from the lipid envelope by an electron-lucent zone. Rubella virions exhibit a marked degree of pleomorphism (**Fig. 1**). The virion has a density of 1.18–1.19 g/ml, whereas isolated capsids have a density of 1.44 g/ml. The genome is contained within a quasi-spherical core or capsid, composed of a single virus-encoded structural protein, C. The C protein (~34 kDa) is present as disulfide-linked homodimers. Although presumed to be icosahedral with a $T = 4$ symmetry, the precise structure of the RV capsid has not been reported.

The capsid is surrounded by a lipid envelope in which two virus-specified glycoproteins, E1 and E2, are embedded, which form the virion spikes. E1 has a molecular weight of 59 kDa whereas E2 is a heterogeneous species ranging from 44 to 50 kDa due to differential glycosylation. Both E1 and E2 appear to be primarily in the form of heterodimers which are easily disrupted by routine preparation techniques. The higher-order architecture of the virion spikes is unclear. E1 is more exposed on the virion surface than is E2, contains both the viral hemagglutinin and receptor site, and is also immunodominant in terms of humoral immunity.

Rubella virions are stable at physiological pH values and can be frozen at temperatures below -20°C for years, without a significant loss of infectivity. Live, attenuated vaccine virus is stored in lyophilized form and must be reconstituted before use. Rubella virions are susceptible to most commonly used inactivating agents, such as formaldehyde, UV light, and lipid solvents.

Genomic Organization

The RV genomic RNA is approximately 9762 nucleotides in length. It contains a 5' terminal cap structure and a 3' terminal poly(A) tract. A distinctive feature of the genomic RNA is that it contains 30% guanine residues and 39% cytosine residues, the highest G + C content of all RNA viruses. The genome encodes two long, nonoverlapping open-reading frames (ORFs) plus untranslated regions (UTRs) at the 5' and 3' ends and between the ORFs (**Fig. 2**). The 5' proximal ORF (NS) encodes nonstructural proteins which are translated into a 2116-amino-acid precursor that is proteolytically cleaved into two proteins, 150 kDa (P150) and 90 kDa (P90), which are at the N- and C-termini of the ORF, respectively. The cleavage is mediated by a papain-like cysteine protease located at the C-terminus of P150. P150 and P90 are responsible for virus RNA replication: P150 contains a domain predicted to have methyl/guanylyltransferase activity (responsible for forming the cap structure at the 5' end of the genomic and

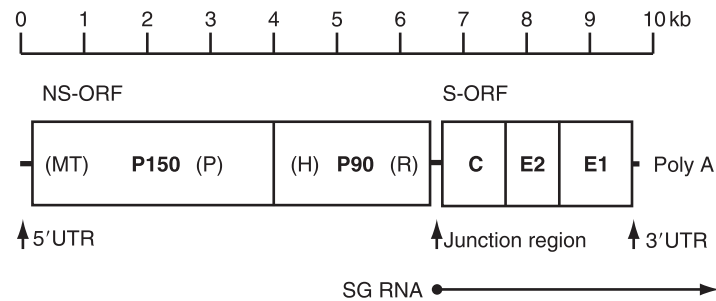


Fig. 2 Coding strategy of the RV genome. Shown is a schematic representation of the RV genomic RNA with untranslated regions (UTRs) drawn as solid black lines and coding regions (ORFs) as open boxes (NS-ORF, nonstructural protein ORF; S-ORF, structural protein ORF). Within each ORF, the coding sequences for the proteins processed from the translation product of the ORF are delineated and, in addition, within the NS-ORF, the locations of motifs associated with the following activities are indicated: methyl/guanylyltransferase (MT), protease (P), helicase (H), and RNA-dependent RNA polymerase (R). The sequences encompassed by the subgenomic RNA (SG RNA) are also shown. The scale at the top of the diagram is in kilobases.

subgenomic RNAs) in addition to the protease, whereas P90 contains both a helicase domain and an RNA-dependent RNA polymerase (or replicase) domain. The 3' located ORF (S) encodes the structural proteins. The 1063-amino-acid product is cleaved by a cellular protease into the virion proteins C, E1 and E2. The order of the virion protein genes within the ORF is 5' C-E2-E1 3'. The S-ORF is translated from a subgenomic RNA synthesized in infected cells and containing the sequences from the start site through the 3' end of the genome. An infectious clone for RV has been developed, which is beneficial for studies based on site-directed mutagenesis. It is a cDNA copy of the viral RNA under the control of promoter contained in a plasmid. Since RV has a plus-sense genome, in vitro transcripts from the plasmid will initiate virus replication following transfection into susceptible cells.

Intracellular Replication Cycle

Myelin oligodendrocyte glycoprotein (MOG) has been identified as a cellular receptor for RV. Since this molecule is expressed mainly on nervous tissue and MOG-independent infection has also been demonstrated, other not yet identified receptors are thought to be responsible for virus entry. Following attachment to the receptor, the virus is taken into the cell by receptor-mediated endocytosis. In the reduced pH environment of the endocytic vesicle, fusion between the viral envelope and the vesicular membrane occurs, releasing the capsid and genomic RNA into the cytoplasm of the cell.

The genomic RNA is translated to produce the NS protein precursor which is cleaved into P150 and P90 (Fig. 3). These proteins then use the genomic RNA as a template for the synthesis of a genome-length, minus-sense RNA, which is subsequently used as the template for the synthesis of both the genomic and the subgenomic RNAs. Synthesis of the subgenomic RNA is initiated by internal recognition of sequences on the genome-length, minus-sense RNA template. Host-cell proteins are likely involved in the replication process and, interestingly C protein seems to be as well. RNA synthesis is asymmetric in infected cells in that more of the plus-sense species than the minus-sense RNA is produced. The uncleaved NS protein precursor is active in minus-strand RNA synthesis, while the cleaved P150/P90 complex appears to be active in only plus-strand RNA synthesis. Thus the activity of the NS protease through mediating this cleavage is important in regulating plus- and minus-sense RNA synthesis. RV RNA synthesis occurs in cytopathic vacuoles of lysosomal origin, in infected cells.

In the structural protein ORF S, E2 and E1 are immediately preceded by hydrophobic signal sequences that direct translation of secreted and membrane-associated proteins into the lumen of the endoplasmic reticulum (ER). Therefore, following translation of the C sequences within the ORF, the E2 signal sequence mediates association of the translation complex with the ER. C-E2 cleavage is mediated by signal endopeptidase, which functions in the lumen of the ER to remove signal sequences from secreted and membrane-associated proteins. Unlike the proteins of the alphaviruses, RV C protein does not have autocatalytic protease activity. Following cleavage, the E2 signal sequence remains associated with C. Similarly, the E1 signal sequence maintains the association of the translational complex with the ER, signalase mediates the E2-E1 cleavage, and the E1 signal sequence remains attached to E2. Soon after synthesis, heterodimerization of E2 and E1 occurs in the lumen of the ER. The three-dimensional folding of E1 appears to be a complicated process requiring intramolecular disulfide-bond formation by all 20-Cys residues in the ectodomain of the protein. Both E1 and E2 acquire high-mannose glycans in the ER; E1 contains three potential glycosylation sites and E2 contains four and all appear to be utilized. The sites of O-glycosylation of the E2, which accounts for the size heterogeneity of E2, are not known. E1 contains an ER retention signal that is only overridden once conformational folding is complete, after which the E1-E2 heterodimer migrates to the Golgi. In the Golgi the N-glycans of both E2 and E1 are modified to complex form, although modification is not complete and the extent of modification on both proteins is heterogeneous. Modifications of the O-glycans on E2 also occur and E2 contains a Golgi retention signal, indicating that the Golgi is the preferred site of viral budding in infected cells. However, late in infection E1 and E2 migrate to the cell surface and budding can also occur on the plasma membrane in some cell lines.

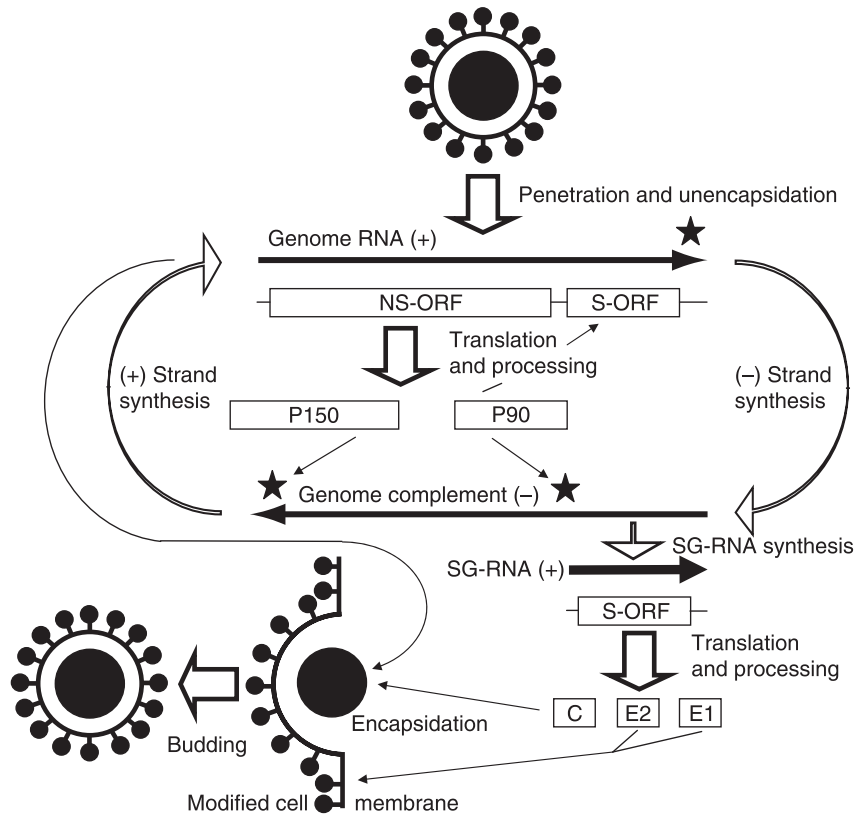


Fig. 3 Replication strategy of RV. The plus-sense genome and subgenomic RNA are represented by solid black arrows indicating plus polarity; beneath each of them, the ORFs that they contain are shown as open boxes. The minus-sense genome RNA complement, represented as a dotted arrow, is used solely as a template for the two plus-sense RNA species. Putative cis-acting sequences on each RNA, which are recognized by the virus RNA-dependent RNA polymerase to initiate synthesis of complementary RNAs, are marked with stars. The general functions of the virus proteins are indicated by arrows (e.g., P150 and P90 functioning as the RNA-dependent RNA polymerase by interacting with cis-acting sequences on the viral RNA species and synthesizing complementary strands).

RV capsid morphogenesis occurs in association with cell membranes. The association of RV capsid protein with membranes is probably mediated by the E2 signal sequence, which is retained at the COOH terminus of C. In fact, C may associate with the E2–E1 heterodimer and migrate as a passenger on the cytoplasmic side of vesicles transporting E2–E1 from the ER to the Golgi and among the Golgi stacks. Unlike the alphaviruses, whose capsids accumulate in infected cells, RV capsids only become visible in association with deformed, thickened membranes that appear to be in the early process of budding. A putative encapsidation signal has been localized near the 5' terminus of the genomic RNA. The C protein is phosphorylated and phosphorylation/dephosphorylation by cell enzymes is proposed as the regulator of the process by which the genome is unpacked following entry (phosphorylation) and encapsidated later in infection following replication (dephosphorylation). Interestingly, in cells in which the complete S-ORF is expressed in the absence of genomic RNA, virus-like particles form and are secreted and these have the same morphology and isopycnic density as the virions do.

RV replicates in the cytoplasm of the infected cell. None of the virus proteins exhibit any involvement with the nucleus during infection and RV infection does not appear to inhibit cell macromolecular synthesis in any grossly detectable manner; however, perturbations of specific macromolecular products and induction of specific genes occurs. Microscopically, RV-infected cells appear similar to uninfected cells; however, rearrangements of cellular cytoskeletal elements and organelles such as mitochondria have been reported. Moreover, RV infection induces a shift of the bioenergetic state of epithelial cells (Vero and A549) and human umbilical vein endothelial cells to a higher oxidative and glycolytic level. RV reportedly inhibits growth in primary human cell cultures in part due to an inhibition of mitosis; however, the virus has no reproducible effect on the growth of stable cell lines. In those cell lines which exhibit CPE (Vero and RK-13 cells), cell death is due to apoptosis, although the extent of apoptosis seems to vary according to virus strain and cell line infected.

Genetics and Evolution

The RV genome is extremely stable. Currently, more than 80 genomes of independent strains of RV have been fully sequenced and all are nearly identical in terms of the size of the genome as well as the coding and noncoding regions within the genome. The only

exceptions are short deletions occasionally encountered in the junction region. A unique feature of RV evolution is that changes to G and C are selected for, indicating an adaptive advantage of the high G + C content of the genome.

There are no known close relatives of RV, and the origin of the virus prior to its introduction into the human population is unknown. Except for short stretches at the 5' end of the genome and at the subgenomic promoter site, RV and the alphaviruses share no nucleotide homology. Accordingly, these two genera are only distantly related and RV has been recently regrouped to the *Matonaviridae* family. Phylogenetic analysis of the NS proteins indicates that RV is more closely related to hepatitis E virus than to alphaviruses and this dissimilarity is borne out by differences in the order of motifs in the NS-ORF. Thus, it is hypothesized that the evolution of the togaviruses may have been more complicated than simple divergence from a common ancestor and probably involved recombination events between progenitors of the current alphaviruses, RV, hepatitis E virus, and, possibly, certain plant viruses.

Serologic Relationships and Variability

RV is monotypic and immunological characterization of diverse strains, including both clade 1 and 2 viruses, has only revealed subtle antigenic differences which map to C or E2. As might be anticipated from the lack of serologic cross-reaction with other viruses like e.g., alphaviruses, there is no homology at the amino acid level between RV and alphaviruses within the virion proteins.

Epidemiology

Historically, RV was endemic worldwide. Over the past 50 years, vaccination programs have curtailed this distribution. In 2016, 22,361 rubella cases were reported to WHO, a 97% decrease from 670,894 cases reported in 2000. Before the advent of vaccination programs, rubella was considered a childhood disease in temperate zones, with seasonal peak occurrence in the spring and epidemics at 5–9-year intervals. However, in tropical zones the highest infection rates were in children under 5 years of age. RV is not as transmissible as measles virus and even during the epidemics, many susceptible individuals are spared. Thus, infection of adolescents and especially young females in childbearing age, a group of the population that should be protected against rubella infection, is not uncommon.

RV elimination has been verified in the Americas in 2015 but since the vaccination coverage decreases, it remains doubtful, whether this great success can be sustained. Other parts of the world experience large rubella outbreaks due to inadequate vaccination programs, groups that do not take vaccination e.g., for religious communities or ethnic minorities which are underserved by national health programs.

Transmission and Tissue Tropism

RV is transmitted between individuals by aerosol. The epithelium of the buccal mucosa provides the initial site for virus replication and the mucosa of the upper respiratory tract and nasopharyngeal lymphoid tissue serve as portals of virus entry. The virus is then spread locally by the lymphatic system, seeding regional lymph nodes where further virus replication occurs. After an incubation period of 7–9 days, virus appears in the blood. Viremia ceases with the onset of detectable rubella-specific antibody, shortly after the rash appears 2–3 weeks post infection. Patients are most infectious immediately before and during the rash phase. RV generally disappears from nasopharyngeal secretions within 4 days of the appearance of rash. Congenitally infected infants shed high loads of virus for 3–6 months following birth and are a source of transmission during that period. Reinfection with RV occurs, usually without clinical illness or virus shedding. There are very few described cases in which RV reinfection of pregnant women with well-documented immunity has resulted in CRS as the result of a two rather rare events: the mother's lack of immunity after vaccination and her exposure to rubella in a highly immune population during early pregnancy.

During pregnancy, placental tissues are very susceptible to infection. Placental infection results in scattered foci of necrotic syncytiotrophoblast and cytotrophoblast cells and evidence of damage to vascular endothelium. Following placental infection, virus can spread to the fetus but this does not always occur and RV is more often recovered from placental tissue than from fetuses. Once fetal infection occurs, virus spreads throughout the fetus and almost any organ may be infected. Severe fetal damage is only associated with infection during the first trimester of pregnancy; the rate of CRS is >85%, 25%, and 10% when infection occurs during the first, second, and third months, respectively. This is due to a combination of an apparent decline in the efficiency of placental transfer after the first trimester and a reduction in the ability of the virus to inflict damage after this time of gestational development.

Clinical Features of Infection

Rubella acquired in childhood or early adulthood is usually mild and it is estimated that up to 50% of rubella infections are clinically inapparent, making rubella surveillance for elimination programs very difficult. Symptomatic rubella encompasses

maculopapular rash, lymphadenopathy, low-grade fever, conjunctivitis, sore throat, and arthralgia. The rash is the most prominent feature and appears following an incubation period of 16–20 days. The rash begins as distinct pink maculopapules on the face that then spread over the trunk and distally onto the extremities. The maculopapules coalesce and the rash fades over several days. An associated posterior cervical and suboccipital lymphadenopathy is also characteristic. Infrequent complications include thrombocytopenia and post-infectious encephalitis. Acute polyarthralgia and arthritis following natural RV infections of adults are common and occur more frequently and with a greater severity in women than in men. Joint involvement is usually transient; however, chronic arthritis, persisting or recurring over several years, has been reported.

The clinical manifestations of CRS apparent at birth vary widely, most frequently including thrombocytopenic purpura (“blueberry muffin syndrome”), intrauterine growth retardation, congenital heart disease (patent ductus arteriosus or pulmonary artery or valvular stenosis), psychomotoric retardation, birth defects in the eye (cataract, glaucoma, retinopathy), hearing loss, and hepatomegaly and/or splenomegaly. Nearly 80% of CRS children show some type of neural involvement, particularly neurosensory hearing loss.

Most clinical manifestations of congenital rubella syndrome are evident at or shortly following birth. Some of these defects are transient. However, recognition of retinopathy, hearing loss, and mental retardation may be delayed in some cases. Progressive consequences of congenital rubella have become increasingly visible as CRS children from the 1964 epidemic have been followed longitudinally. These predominantly involve endocrine dysfunction (diabetes mellitus, which ultimately affects 40% of CRS patients, and thyroid dysfunction). A rare, fatal neurodegenerative disease, progressive rubella panencephalitis (PRP), described in CRS patients, bears superficial resemblance to subacute sclerosing panencephalitis associated with measles virus. Subsequently, PRP cases have been reported in individuals infected postnatally.

Pathogenesis, Pathology, and Histopathology

There is limited information on the pathogenesis of uncomplicated rubella because of the benign nature of the illness. With respect to the complications that can accompany acute rubella, the postinfectious encephalitis is thought to be autoimmune in nature since RV cannot be isolated from the cerebrospinal fluid or brain at autopsy. Interestingly, however, extensive inflammation and demyelination are not observed. In a few cases of rubella arthritis, the presence of RV in synovial fluid and/or cells has been demonstrated and therefore it is assumed that virus persistence is involved. However, considering the age and sex factors in the incidence of arthritis, it seems likely that immunopathological mechanisms also play a role. Human leukocyte antigen (HLA) class II haplotypes were associated with differences in predisposition of adult women to arthritis and arthralgia following rubella vaccination to experience arthritis and arthralgia and generally with responses to vaccination and viral infection.

Following fetal infection, virus can be isolated from practically every organ of aborted fetuses or infants who die soon after birth. However, only 1 in 10^3 to 1 in 10^5 cells are infected and it is not known how such a low infection rate leads to the profound birth defects exhibited in CRS. RV antigen was found in interstitial fibroblasts in the heart, adventitial fibroblasts of large blood vessels, alveolar macrophages, progenitor cells of the outer granular layers of the brain, and in capillary endothelium and basal plate in the placenta. Affected organs are usually small for gestational age and contain reduced numbers of cells. Considering the inhibitory effect of RV on primary cells, it is thought that virus infection early in organogenesis inhibits cell division, leading to both retardation and alteration in organ development. Moreover, RV infection has shown virus persistence continuing after birth, as evidenced by virus shedding, which generally ceases within 6 months of age. Whether virus persistence continues beyond cessation of shedding and plays a role in the delayed and progressive manifestations of CRS is not known. Transcriptome analysis have shown genes involved in cytokine production and cytokine regulation were differentially regulated in wild type RV and vaccine virus-infected cells. Increased synthesis of inflammatory cytokines following RV infection was verified as well as down-regulation of chemokine CCL14, which is implicated in supporting embryo implantation at the fetal-maternal interface. GO term-based cluster analysis of the down-regulated genes of HUVEC revealed an enrichment of the GO terms “sensory organ development”, “ear development” and “eye development”. When human induced pluripotent stem cells (iPSCs) were infected with RV followed by differentiation into cells of the three embryonic lineages (ecto-, meso-, and endoderm) as a cell culture model for blastocyst- and gastrulation-like stages, the RV-infected endodermal cells showed profound alterations of the epigenetic landscape including the expression level of components of the chromatin remodeling complexes and an induction of type III interferons.

Histologically, affected organs of CRS patients show a limited number of well-recognized malformations, with noninflammatory histopathology predominating. Particularly apparent are vascular lesions and focal destruction in tissues bordering these lesions. These lesions are likely to be due to virus replication in the vascular endothelium and damage to neighboring tissue may play a role in the pathogenesis of CRS. The neuropathology of CRS is of interest not only because of the defects manifest shortly after birth, but also because some CRS patients develop schizophrenia-like symptoms later in life. CRS brains are generally free of gross morphological malformations, with a common tendency toward microcephaly. Vascular damage, leptomenigitis, decreased number of oligodendroglial cells, and alteration of white matter are observed. Magnetic resonance imaging of a group of CRS adults with schizophrenia-like symptoms revealed specifically reduced cortical gray matter and enlargement of the ventricles, which were not previously observed characteristics of CRS-induced neuropathology. Interestingly, the comparative finding that non-CRS schizophrenia patients exhibit a pattern of brain dysmorphism similar to that found in CRS patients with schizophrenia-like symptoms supports the hypothesis that schizophrenia is developmental in nature (there is some evidence for a viral trigger to schizophrenia).

Immune Response and Serodiagnosis

Following an acute infection, anti-RV IgM antibodies are detectable a few days after rash onset and for four to six weeks afterwards, so that detection of IgM using ELISA is important for the diagnosis of an acute RV infection; in the succeeding weeks anti-RV antibodies in all immunoglobulin classes appear. The dominant early and persistent IgG response is in the IgG1 subclass and antibodies of this class persist indefinitely after natural infection of otherwise healthy individuals. The majority of the antibody response is directed to the E1 glycoprotein, with proportionally lesser amounts of the response directed against E2 or C; neutralizing and complement-fixing antibodies are induced as well. When investigating relationships between rubella vaccine-specific humoral and cellular immunity, a significant correlation between neutralizing antibodies and IFN- γ secretion was observed. Because of the importance of serology testing for rubella, many rubella tests are on the market but standardization between the tests is missing. This leads often to an apparent change in titer values (and the respective concern) but the phenomenon may be based solely on the use of two discordant ELISA tests.

RV-specific cellular immune responses are measurable within 1–2 weeks of onset of illness. Major histocompatibility complex (MHC)-restricted CD4 + epitopes have been mapped to all three of the virus structural proteins; however, CD8 + epitopes have thus far only been mapped to the C protein. HLA class I and II alleles both positively and negatively associated with antibody and lymphoproliferative responses following rubella vaccination have been determined.

Following fetal infection, the fetus produces IgM antibodies, detectable at 18–20 weeks of gestation, and maternal IgG antibodies cross the placenta. Both types of antibodies exhibit virus-neutralizing activity *in vitro*; however, these antibodies are not sufficient to resolve virus infection during gestation. As discussed above, the intracellular maturation of virus probably shields it from antibody recognition. After birth, the presence of IgM or a lack of decline of IgG antibodies are both considered indicative of a fetal infection as well as the shedding of virus that can be detected by RT-PCR. CRS infants exhibit various degrees of impairment in the cellular immune response to RV and it is thus a deficiency in this arm of the immune response that allows the virus to persist. Considering the cellular immune deficiency in CRS infants, it is curious that detectable virus persistence ends a few months after birth. The way by which virus persistence is cleared under these conditions is not understood.

Diagnosis

The common symptoms of acute rubella, maculopapular rash, low-grade fever, and lymphadenopathy are nonspecific and easily confused with other illnesses like e.g., measles, scarlatina, or infection with parvovirus B19. In compliance to WHO elimination programs every suspected rubella case should be confirmed by a laboratory diagnosis. Laboratory confirmation of a suspected case of rubella requires PCR detection of the genome in a throat swab taken within 5 days after onset of symptoms and the detection of IgM antibodies in an acute-phase serum specimen. The diagnosis of acute rubella is of utmost importance when an infection during early pregnancy is suspected. In these cases, the detection of IgM alone is not a reliable basis to decide whether an RV infection has occurred in early pregnancy; a specialized laboratory should be consulted for verification. Laboratory diagnosis will usually employ IgG avidity, and detection of the E2-reactive IgG antibodies in a Western blot to exclude an earlier infection. The risk of congenital infection is related to the gestational age at the time of maternal infection. When pregnant women are infected with rubella during the first 11 weeks of pregnancy, up to 90% of liveborn infants may suffer from CRS. There is no other intervention for congenital rubella than abortion. Diagnosis of CRS in a newborn is initially made on the basis of symptoms and confirmed in the laboratory by serological testing for the presence of IgM antibodies, by lack of decrease of IgG antibodies in serum specimens, and by detection of virus RNA. CRS infants shed high amounts of virus and therefore are a source of contagion. Comparison of serologic markers revealed that compared with healthy children, children with CRS have significantly higher RV-specific IgG and stronger C and E2 protein-specific antibody responses.

ELISA-based detection of IgG as a surrogate marker for immunity by occupational physicians is routine practice as well as for pre-pregnancy planning. IgG detection can also be used in lieu of proof of vaccination, which is required in many countries for the enrollment in school or university. Individuals found to be seronegative should subsequently be vaccinated. In case of women who are pregnant at the time of testing, vaccination should be offered directly post-partum.

Prevention and Control of Rubella

There is no specific treatment of RV infection, but since the infection is vaccine-preventable and humans are the only host, it is possible to eliminate rubella. Several live attenuated vaccines were developed and placed in use by 1970; the vaccine RA 27/3 is used in most countries. Most national immunization programs recommend two doses of the MMR or MMRV vaccine during childhood. If not already covered by a routine vaccination in childhood, women in reproductive age and health care workers should be protected against rubella.

Attenuated RV vaccines cause subclinical infection with transient viremia in susceptible patients. Natural transmission of vaccine virus has not been reported. The RA27/3 vaccine strain produces seroconversion in greater than 95% of recipients. Vaccine-induced titers are lower than those induced by a natural infection but a life-long immunity is mounted in most cases. Reinfections can occur but in contrast to primary infections, they do normally not endanger the unborn child. The RV vaccine is generally

administered to children in a trivalent form with the measles- and mumps-attenuated vaccines (MMR) or as a tetravalent vaccine with varicella virus vaccine (MMRV).

The RV vaccines have been most successful in terms of induction of immunity with the absence of side effects. However, two issues have arisen concerning rubella vaccination. The first is that the vaccine virus can cross the placenta and infect the fetus. However, registries of deliveries to women inadvertently vaccinated during early pregnancy revealed no reported congenital abnormalities. Similar findings were reported in subsequent studies conducted in other countries in conjunction with mass vaccination campaigns. Nevertheless, due to a theoretical risk vaccination during pregnancy is contraindicated and deferred until postpartum. Second, is the occurrence of arthralgia and arthritis following vaccination. Joint complications are nonexistent in children with the currently used rubella vaccines; however, transient arthralgia and arthritis is reasonably common among adult female vaccinees (or vaccine recipients). There have also been reports of chronic arthritis and related neurological involvement following vaccination of adult women. Although these complications are consistent with complications that can accompany natural rubella infection in adult females, recent studies have shown that the incidence of such vaccine-related complications is low and cannot be statistically differentiated from the incidence of similar symptoms in control, unvaccinated populations.

Since the inception of rubella vaccination in 1969 in the US, many countries have implemented rubella vaccination for children, health care workers and individuals working in community facilities as well as women in childbearing age. The cost of the vaccine, the general mildness of the disease compared to other infections like tuberculosis, malaria and HIV, and the nature of national public health infrastructures are factors impeding vaccination programs in developing countries. However, efforts among developing countries have accelerated, particularly in conjunction with ongoing measles elimination programs, and as of December 2016, 152 (78%) of 194 countries had introduced RV vaccine into the national immunization schedule. In addition to its efforts to initiate rubella vaccination programs in developing regions, the WHO established a global Measles and Rubella Laboratory Network that is indispensable in surveying the way to elimination. The WHO Regions of the Americas have eliminated rubella and CRS in 2010 and verified this success in 2015, while other regions like e.g., Europe and the Western Pacific have set goals for elimination of rubella and CRS by 2020.

Future

RV is a viral pathogen, that infects only humans. Due to the presence of a safe and efficient vaccine, global implementation of rubella vaccination in accordance with the WHO elimination goal is of vital importance to eliminate rubella and prevent CRS in the future. Moreover, RV and the attenuated vaccine are a fascinating model to unravel teratogenic effects and compare it with other agents causing congenital diseases like e.g., Zika virus, cytomegalovirus or toxoplasmosis. The follow-up of CRS patients can further teach us of a very special condition, e.g., the high incidence of diabetes in CRS patients which is statistically the strongest direct association between a specific human virus and an autoimmune disease. The mechanism of viral involvement, if any, in each of these diseases is not fully understood. Unfortunately, our present understanding of the disease mechanisms in the RV-related syndromes is hindered by the lack of a suitable animal model system that fully mimics the infection seen in humans, making the development of an animal model a research priority. RV is taxonomically unique and appears to have evolved as a recombinational hybrid of distantly related ancestor viruses. Thus, investigation of the molecular biology of RV likely will reveal novel replication strategies and yield insight into virus evolution.

Further Reading

- Best, J.M., Castillo-Solorzano, C., Spika, J.S., *et al.*, 2005. Reducing the global burden of congenital rubella syndrome: Report of the World Health Organization steering committee on research related to measles and rubella vaccines and vaccination, June 2004. *International Journal of Infectious Diseases* 192, 1890–1897.
- Bilz, N.C., Willscher, E., Binder, H., *et al.*, 2019. Teratogenic rubella virus alters the endodermal differentiation capacity of human induced pluripotent stem cells. *Cells* 8 (8), 870.
- Dimech, W., Grangeot-Keros, L., Vauloup-Fellous, C., 2016. Standardization of assays that detect anti-rubella virus IgG antibodies. *Clinical Microbiology Reviews* 29 (1), 163–174.
- Frey, T.K., 1994. Molecular biology of RV. *Advances in Virus Research* 44, 69–160.
- Geyer, H., Bauer, M., Neumann, J., *et al.*, 2016. Gene expression profiling of rubella virus infected primary endothelial cells of fetal and adult origin. *Virology Journal* 13, 21.
- Grant, G.B., Desai, S., Dumolard, L., Kretsinger, K., Reef, S.E., 2019. Progress toward rubella and congenital rubella syndrome control and elimination – Worldwide, 2000–2018. *Morbidity and Mortality Weekly Report* 68 (39), 855–859.
- Hobman, T.C., 2013. Rubella virus. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia: Lippincott Williams and Wilkins, pp. 1069–1100.
- Hyde, T.B., Sato, H.K., Hao, L.J., *et al.*, 2015. Identification of serologic markers for school-aged children with congenital rubella syndrome. *The Journal of Infectious Diseases* 212 (1), 57–66.
- Lambert, N., Strebel, P., Orenstein, W., Icenogle, J., Poland, G.A., 2015. 6. Rubella. *Lancet* 385 (9984), 2297–2307.
- Lazar, M., Perelygina, L., Martinez, R., *et al.*, 2016. Immunolocalization and distribution of rubella antigen in fatal congenital rubella syndrome. *EBioMedicine*. 3, 86–92.
- Ovsyannikova, I.G., Salk, H.M., Larrabee, B.R., Pankratz, V.S., Poland, G.A., 2015. Single nucleotide polymorphisms/haplotypes associated with multiple rubella-specific immune response outcomes post-MMR immunization in healthy children. *Immunogenetics* 67 (10), 547–561.
- Pitts, S.I., Wallace, G.S., Montana, B., *et al.*, 2014. Congenital rubella syndrome in child of woman without known risk factors, New Jersey, USA. *Emerging Infectious Diseases* 20 (2), 307–309.
- Tzeng, W.P., Matthews, J.D., Frey, T.K., 2006. Analysis of RV capsid protein-mediated enhancement of replicon replication and mutant rescue. *Journal of Virology* 80, 3966–3974.
- Vynnycky, E., Adams, E.J., Cutts, F.T., *et al.*, 2016. Using seroprevalence and immunisation coverage data to estimate the global burden of congenital rubella syndrome, 1996–2010: A systematic review. *PLoS One* 11 (3).

Saint Louis Encephalitis Virus (*Flaviviridae*)

William K Reisen, Lark L Coffey, and Daniele M Swetnam, University of California, Davis, CA, United States
Aaron C Brault, Centers for Disease Control and Prevention, Fort Collins, CO, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Culex Mosquito genus

Flaviviridae Viral family

Flavivirus Viral genus

Glossary

Bridge vector Arthropod capable of moving an infectious agent from an enzootic cycle to hosts typically not involved in the enzootic cycle.

Diapause Period of hormonally induced reproductive inactivity for insects typically triggered by shortening day length and cool temperatures (akin to hibernation of mammals).

Extrinsic incubation period Time period within the arthropod vector between the ingestion of an infectious agent and when the vector is capable of transmitting following biological amplification.

Gonotrophic cycle The period between blood feeding and the oviposition of eggs by an insect.

Vector competence The capacity of an arthropod to become infected with an infectious agent (typically through oral ingestion), disseminate the agent to other tissues and subsequently transmit through blood feeding, typically through salivary secretions.

Viremia Presence of virus in the peripheral circulation of a vertebrate host.

History

Saint Louis encephalitis virus (SLEV) probably has been present in the New World within enzootic cycles for thousands of years, although recent molecular genetic studies indicate the current lineages may have diverged as recently as 200 years ago from a common flaviviral ancestor within present day Mexico. The arrival of European settlers in the 1600s and the extensive agricultural development that followed greatly altered the landscape by clearing and irrigating vast areas of North America and establishing extensive urban centers. These changes likely increased the abundance of *Culex* mosquito species and avian hosts such as house finches (*Haemorrhous mexicanus*) and mourning doves (*Zenaida macroura*). Furthermore, the introduction of new avian hosts such as house sparrows (*Passer domesticus*) and vectors such as *Culex pipiens* complex mosquitoes, intensified human-vector mosquito contact, and likely increased the incidence of human infection; moreover, clinical diagnosis of disease caused by SLEV assuredly had been confused with other infections causing fever and/or central nervous system (CNS) disease during summer months.

During the summer of 1933, a major encephalitis epidemic with more than 1,000 clinical cases occurred in St. Louis, Missouri. These cases occurred during the middle of an exceptionally hot, dry summer and were concentrated within areas of the city adjacent to open storm water and sewage channels that produced a large number of *Culex* mosquitoes. A virus, later named “Saint Louis encephalitis virus” (SLEV), was isolated from a human brain specimen during an autopsy during this outbreak. The epidemiological features of this epidemic included the late summer occurrence of cases (especially in persons > 50 years of age), exceptionally warm temperatures during the summer months preceded by a wet spring, and elevated *Culex* mosquito abundance associated with poorly draining waste water systems. These features have remained the hallmarks of SLEV epidemics through the 21st century.

A multidisciplinary team of entomologists, vertebrate ecologists, epidemiologists and microbiologists from the University of California at Berkeley subsequently investigated an SLEV epidemic in the Yakima Valley of Washington State 1941–1942 and established the components of the summer transmission cycle, including the identification of wild birds as the primary vertebrate hosts and *Culex* mosquitoes as principal vectors. The isolations of SLEV from *Culex tarsalis* and *Cx. pipiens* mosquitoes were among the first of any virus from mosquitoes and stimulated the redirection of mosquito control in North America from *Anopheles* malaria vectors and pestiferous *Aedes* to *Culex* vectors.

A more thorough understanding the basic transmission cycle, an appreciation of the wide range of clinical symptoms, and the development of laboratory diagnostic procedures subsequently has provided an expanded view of the public health significance of SLEV, with epidemics or clusters of cases recognized annually throughout the United States. Wide geographic distribution and transmission since 1933 has resulted in > 1000 deaths, > 10,000 cases of severe illness and an estimated > 1,000,000 mild or subclinical infections. The largest documented SLEV epidemic occurred during 1975 in the Ohio River drainage, with > 2000 human cases documented. Other substantial human epidemics, involving up to hundreds of cases, have occurred in Missouri (1933, 1937), Texas (1954, 1956, 1964, 1966), Mississippi (1975), Florida (1977, 1990) and Louisiana (2001). Smaller outbreaks have been recognized in California (1952, 1989), New Jersey (1962), Arizona (2015) and several other states plus Ontario (1975), Canada, and Cordoba, Argentina (2005) (**Fig. 1**). Cases reported annually to the Centers for Disease Control and Prevention

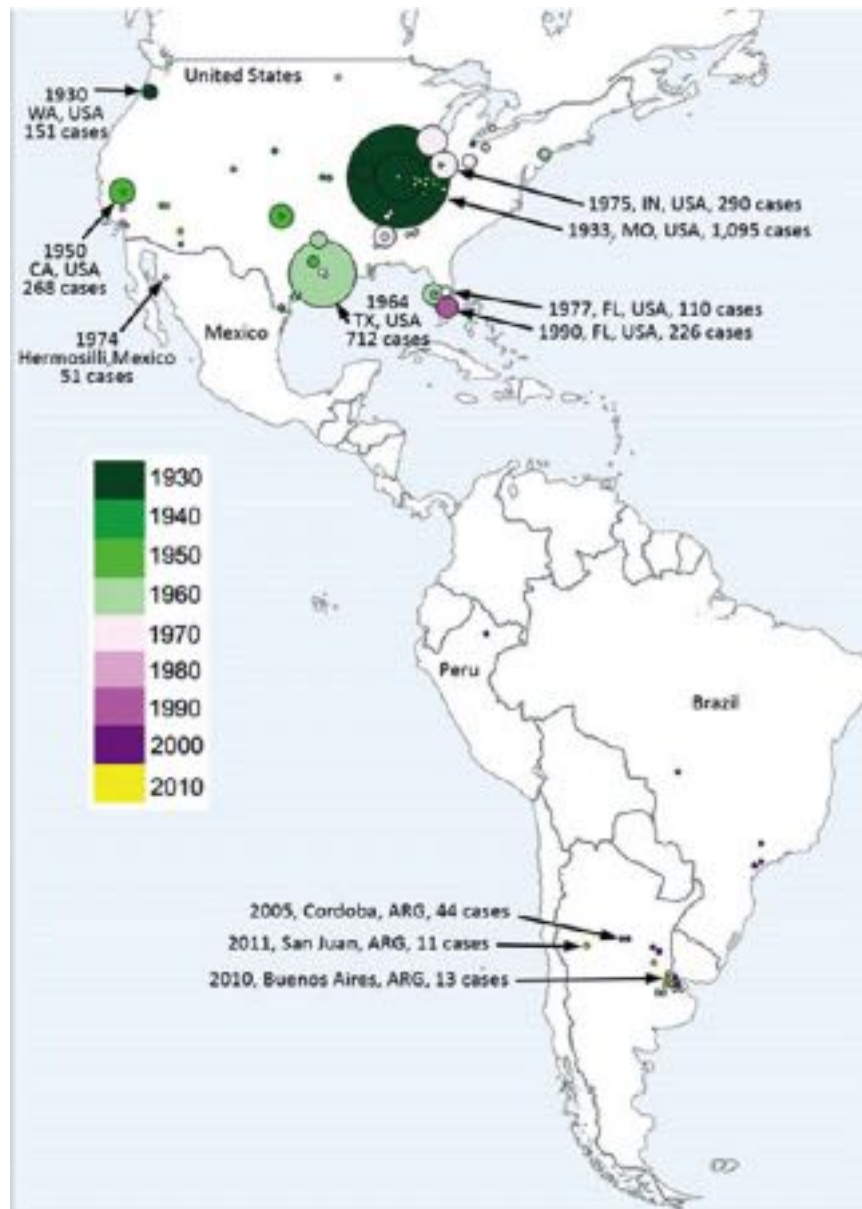


Fig. 1 Geographic distribution of historical Saint Louis encephalitis human cases reported in the Americas through November 2017. Dot size represents the number of human cases reported in each episode. Colors represent year of detection. Reproduced with permission from Diaz, A., Coffey, L.L., Burkett-Cadena, N., Day, J.F., 2018. Reemergence of St. Louis encephalitis virus in the Americas. *Emerging Infectious Diseases* 24 (12).

(CDC) from 1964 to 2018 are shown in [Fig. 2](#). Geographical distribution of cases can be seen throughout the continental United States with recurrent disease transmission centers existing along the Ohio River basin, Florida, Texas and California ([Fig. 3](#)).

Virion Structure/Genome

SLEV virions are approximately 50 nanometers in diameter and are comprised of a lipid envelope in which viral structural proteins, (pre)membrane (prM/M) and envelope (E) glycoproteins, are embedded. The structural proteins associate as heterodimers within the lipid envelope with $T = 3$ symmetry. The lipid envelope scaffolds an internal icosahedral capsid (C) protein constellation encompassing a single-stranded (monopartite), positive-sense viral genome of approximately 11 kb in length. The RNA genome is flanked by 5' and 3' untranslated regions ([Fig. 4](#)). A 5' m⁷GpppAmpN2 cap is present that protects the virus from endoribonuclease degradation. No polyadenylation is present at the 3' end of the genome; however, a non-coding sequence of variable length (~300 nucleotides) is present and single-stranded RNA structures are formed that block 5' – 3' endoribonucleases degradation forming subgenomic RNAs that are believed to serve as decoys for siRNA antiviral pathways in both vertebrate and

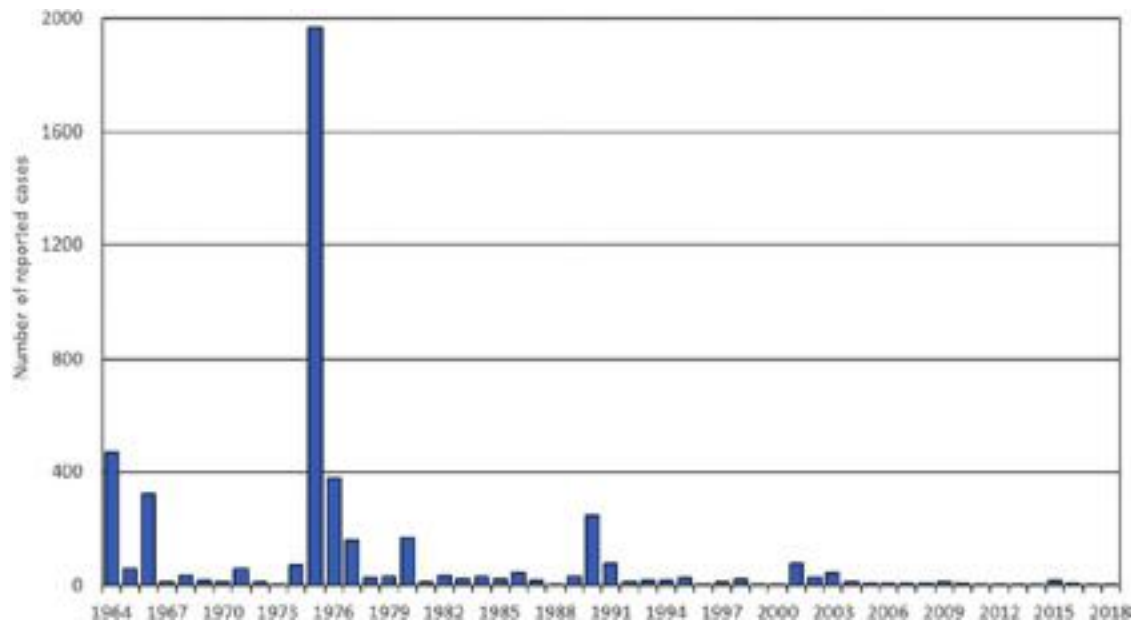


Fig. 2 Number of human Saint Louis encephalitis virus cases reported to the Centers for Disease Control and Prevention ArboNet system from 1964 to 2018 by year.



Fig. 3 Map of United States indicating the cumulative number of Saint Louis encephalitis virus cases per state between 1964 and 2018 (data from ArboNet).

invertebrate hosts. The 5' and 3' ends of the genome have conserved cyclization sequences that are involved in transcriptional and translational initiation. The coding region of the virus is expressed as one long translated polyprotein that is cleaved by host and viral proteases. The three structural proteins (C, prM and E) are coded at the 5' end of the genomes and the seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) are found in the 3' portion of the genome ([Fig. 4](#)). The nonstructural proteins principally serve to replicate viral RNA through RNA-dependent RNA polymerase (NS5) and helicase (NS3), process

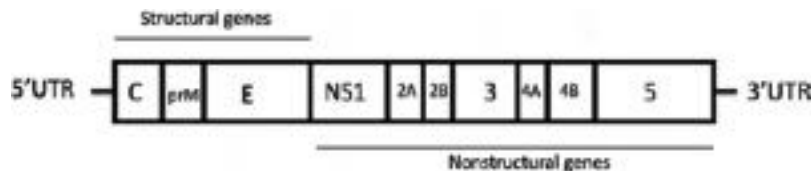


Fig. 4 Depiction of the genomic structure of the Saint Louis encephalitis virus. The 11 kb RNA genome is flanked by 5' and 3' untranslated sequences. The 5' structural proteins are delineated as well as the 3' non-structural protein genes.

translated viral proteins with a protease (NS2B/NS3) and cap viral RNA with methyltransferase (NS5). Nonstructural proteins are also involved in the antagonization of host-induced antiviral responses.

Classification/Genetic Characterization

Saint Louis encephalitis virus (SLEV) is a mosquito-borne flavivirus within the family *Flaviviridae*. Serologically, SLEV has been classified as a member of the Japanese encephalitis serocomplex with the most closely related viruses (West Nile, Japanese encephalitis, Usutu and Murray Valley encephalitis viruses). Subsequent to its initial isolation from a brain sample during the human encephalitis outbreak in St. Louis, Missouri in 1933, serological assays, which included mouse protection assessments with viral preparations incubated with convalescent serum dilutions, were performed that demonstrated that the isolated virus was antigenically distinctive from other viruses typically associated with human encephalitis cases during summer months, such as the equine encephalitis viruses, poliomyelitis and vesicular stomatitis viruses. Marked differences in the severity of SLEV epidemics stimulated interest in possible differences among isolates made over time and space. Detailed studies during the 1980s clearly demonstrated geographic variation among 43 different SLEV isolates using oligonucleotide fingerprinting and virulence in model vertebrate hosts. These strains were grouped into 6 clusters: (1) east central and Atlantic USA, (2) Florida epidemic, (3) Florida enzootic, (4) eastern USA, (5) Central and South America with mixed virulence, and (6) South America with low virulence. Changes in virulence were attributed, in part, to differences in mosquito vector competence and were supported by the historical presence or absence of human cases. Subsequent sequencing studies extended the understanding of SLEV genetics and provided further insight into patterns of geographical variation. Sequences of the envelope gene from SLEV strains isolated in California from 1952 through 1995 varied temporally and spatially, but indicated regional persistence in the Central Valley for at least 25 years as well as sporadic introduction and extinction. Studies in Texas using a single-strand conformation polymorphism technique showed that multiple SLEV strains circulate concurrently and remain highly focal, whereas other strains amplify and disseminate aggressively during some summers, but then disappear. Further analysis of sequences from 62 isolates made throughout the known geographical range of SLEV indicated that there have been 7 lineages that overlapped somewhat with the 6 groups previously defined using oligonucleotide fingerprinting: (I) western USA, (II) central and eastern USA and 3 isolates from Mexico and Central America, (III) one mosquito isolate from Argentina, (IV) 5 isolates from Panama mosquitoes, (V) South American strains plus an isolate from Trinidad, (VI) one Panama isolate from a chicken, and (VII) two isolates from Argentinian rodents. Subsequent re-analysis of 73 envelope gene sequences from GenBank basically supported these groupings, but had low support for separation of the Panama isolates. This analysis also indicated limited exchange between North and South American clades, especially California, but indicated repeated circulation among the Caribbean Islands, Mexico and the southern USA. Collectively, these data indicated that SLEV strains vary markedly in virulence and that the frequency and intensity of epidemics in the USA may be related to genetic selection by different host systems. Interestingly, transmission within the Neotropics appears to have given rise to and/or allowed the persistence of less virulent strains that rarely amplify to produce epidemic level transmission, a scenario duplicated by West Nile virus in the Americas.

Several changes have been proposed since the initial designation, including joining lineages IV and VI, and the addition of lineage VIII and the Palenque lineage (Fig. 5). Lineage I is geographically restricted to the western US, comprises genomes from 1950 to 1998 and can be divided into Clades IA and IB. Clade IA contains only genomes collected from Kern County, California between 1950 and 1970. Sequences belonging to Clade IB have been detected in the greater western US, including California from 1978 to 98, as well as New Mexico and Colorado in 1972, and Texas from 1966 to 1989 (Hale County, El Paso, and Dallas). Whereas lineage I genomes have only been detected in a small geographic area, lineage II is more widely distributed throughout North, South and Central America. Kramer and Chandler (2001) initially divided lineage II into six clades (IIA–IIF). Lineage I clade IIA consists of genomes from Jamaica in 1962, Brazil in 1968, and the central and eastern US from 1933 to 1996, including Florida, Kentucky, Mississippi, Missouri and Texas. Lineage II clade IIB contains sequences from Central America, specifically Guatemala in 1969 and 1978, and several genomes from the southeast portion of the US, including Florida in 1962, Tennessee in 1974 and the Gulf Coast region of Texas (Harris, Jefferson and Nueces Counties) between 1983 and 2003. The sequences of lineage I clade IIC, although restricted to the US, are split between the eastern (Tennessee, Florida and Maryland) and western (California) states and were from genomes detected between 1975–1977 and 1988–1992, respectively. Genomes belonging to lineage I clade IID were from 1962 to 1973 from Florida, Mexico and Panama. In its initial description by Kramer and Chandler (2001), lineage I clade IIE contained the two earliest SLEV sequences, Parton and Hubbard, from Missouri in 1933 and 1937, respectively. Later studies indicated that a third genome from the US (state unknown) during 1966, Laderle, also belonged to this

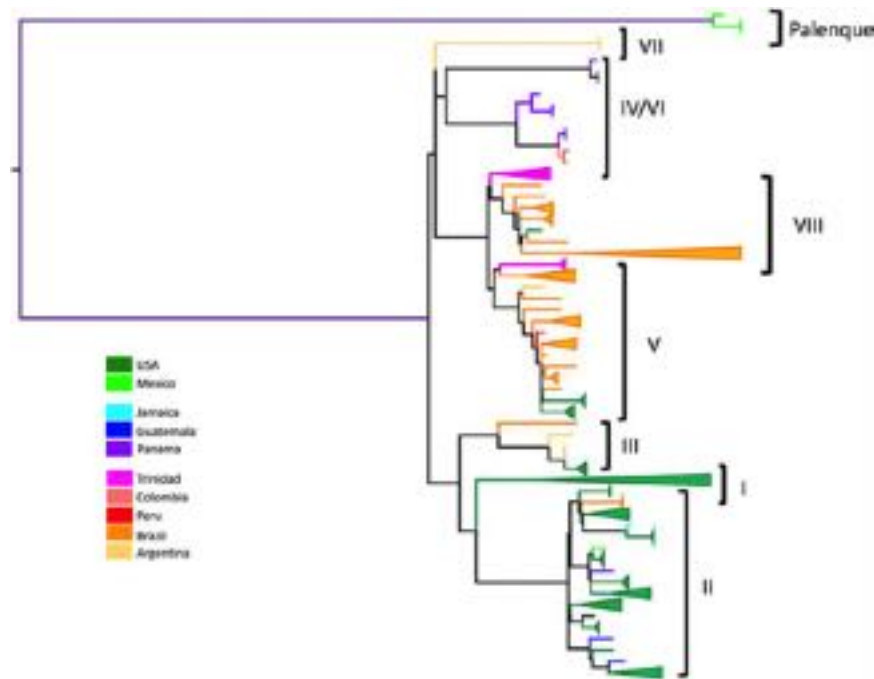


Fig. 5 Phylogenetic tree based on the envelope coding sequences of Saint Louis encephalitis viruses indicating the relationship of different geographic lineages (designated by roman numerals). Isolate locations are color-coded according to the legend.

clade. Finally, clade IIF was defined using one genome from Guatemala collected in 1969. However, several studies found evidence of sequencing inconsistencies in the Parton genome that altered the phylogenetic clustering pattern of clade IIE so that the three genomes clustered within clade IIA. Similarly, a revised Guatemala 1969 genome clustered within clade IIB. Finally, one additional clade, clade IIG, was identified in 2008 and contains only a single genome collected in Florida during 1962. Both lineages I and II have been associated with epidemics in North America. Lineage III was responsible for epidemic disease in Argentina and more recently, in the western US, and also was detected in Brazil during 2004.

Like lineages II and III, lineage V also has been detected throughout the Americas. In a pattern very similar to lineage III, lineage V historically was restricted to South America (Argentina in 1978, Brazil from 1960 to 1973, Peru in 1975, Trinidad in 1955) and subsequently expanded into the western US, including western Texas (El Paso) in 2002 and California between 1998 and 2003. The remaining lineages: IV, VI, VII, VIII and Palenque, are all geographically restricted. Lineage IV originally was defined by a single sequence collected in Panama during 1983; however, support for the separation of lineages IV and VI was very weak leading investigators to suggest combining the two lineages. Lineage VII contains two genomes from Argentina during 1966 and 1967 and the distantly related Palenque lineage that has only been detected in Palenque National Park in southern Mexico. Finally, lineage VIII appears to be restricted to the Para state in the northern Amazon region in Brazil and was detected between 1964 and 1984.

Taken together, several trends have been reported for SLEV across multiple lineages. Overall, there appears to be a consistent pattern of northward SLEV dispersal. In particular, the Gulf of Mexico region (bounded by latitudes 15°N and 30°N) appears to act as a major source for SLEV expanding into North America. South America, specifically Brazil and Argentina, also appears to be a source of SLEV moving into California (lineages III and V). A degree of isolation separating the eastern and western US has been reported in several lineages. For instance, lineage IIB in the Gulf Coast region of Texas clusters more closely with SLEV collected in the Eastern or Central US, whereas lineages IB, III and V from west and north Texas are related more closely to SLEV in California.

It is important to note that our current understanding of SLEV lineages is restricted by limited and inconsistent virus surveillance and sequencing. Additional SLEV sequences, especially from new locations, will likely refine our understanding of endemic SLEV circulation and dispersal events. Furthermore, while multiple studies have described the geographic distribution of each SLEV lineage, only one has expanded on their specific nucleotide or amino acid differences. That study defined 17 amino acid changes within the E protein associated with different lineages or clades. However, several of the amino acid changes were shared across multiple lineages and no defining amino acid changes were described for clade IID. Additional studies are needed to identify the genotypic differences among the SLEV lineages, especially outside of the E protein.

Life Cycle

St. Louis encephalitis virus is transmitted in enzootic transmission cycles between viremic avian hosts and *Culex* spp. mosquito vectors. Birds develop viremias of sufficient magnitude to infect mosquitoes with virus following bloodfeeding. Experimental data also suggests that nestlings develop elevated viremias and could be instrumental for seasonal enzootic amplification of the virus.

Although humans can show symptoms of viral infection, they serve as tangential or “dead end” hosts, not manifesting viremias of sufficient magnitude to infect additional mosquito vectors. Once exposed to virus through the ingestion of a viremic blood meal, *Culex* vectors’ midguts are infected and the virus disseminates through the open circulatory system (hemocoel) of the mosquito, infecting the salivary glands. Subsequent feeding by the mosquitoes and its associated salivation inject viral particles into another avian host. The time period in which the virus traverses from the midgut to the salivary glands is defined as the extrinsic incubation period, the duration of which is directly related to temperature.

Host Range

Although a wide variety of mosquitoes occasionally have been found infected in nature, three avian-feeding species within the genus *Culex* appear to be the most frequently infected and important arthropod hosts: *Culex pipiens* (including the subspecies *Cx. p. quinquefasciatus* at southern latitudes, *Cx. p. pipiens* at northern latitudes, and intergrades) in urban and periurban environments throughout North and South America, *Culex tarsalis* in irrigated agricultural settings in western North America including northern Mexico, and *Culex nigripalpus* in the southeastern U.S., the Caribbean and parts of the Neotropics. Although these species feed predominantly on birds, they also feed on mammals including humans, and therefore function as both maintenance and bridge vectors. Other *Culex* species such as *Culex stigmatosoma* in the west, *Culex restuans* and *Culex salinarius* in the east, and perhaps species in the subgenus *Melanoconion* in the Neotropics also may be important in local transmission. Ticks have been found naturally infected, but their role in virus epidemiology most likely is minimal.

The importance of avian host species appears to be related to vector *Culex* host selection patterns as well as to avian susceptibility to the virus. Species can be separated into those frequently, sporadically and never found infected in nature, and these groupings are related directly to their nocturnal roosting/nesting behavior and the questing behavior of *Culex* vectors. Wild birds do not develop apparent illness following experimental infection, but their viremia response varies markedly, depending upon virus strain, bird species and bird age. Viremia titers sufficient to infect mosquitoes typically are limited to 1–5 days post-infection. Based on serological surveys during or after epidemics, peridomestic passeriforms, including house finches, house sparrows, cardinals (*Cardinalis cardinalis*) and blue jays (*Cyanocitta cristata*), columbiforms, including mourning doves and rock doves (*Columba livia*) or domestic pigeons (*Columba livia domestica*), and galliforms, especially domestic chickens (*Gallus gallus*), seem to be infected most frequently. SLEV strains isolated from *Cx. pipiens* complex mosquitoes from the central and eastern U.S. produced elevated viremias in house sparrows, whereas strains isolated from *Cx. tarsalis* from the western U.S. were weakly viremogenic. Although host competence studies have been limited, the adults of few bird species seem to develop elevated viremias. However, nestling house finches, house sparrows and mourning doves produce high viremias that readily infect mosquitoes. Therefore, the nesting periods of multibrooded species may be critical for virus amplification. Regardless of their viremia response, most experimentally infected birds produce antibody and, although titers typically decay rapidly, these birds remain protected from reinfection for life. Although frequently antibody positive during serosurveys, SLEV infection does not produce elevated viremias or cause clinical illness in domestic animals, including equines, porcines, bovines or felines. Similar to wild birds, immature fowl <1 month old (including chickens and ducks) consistently develop sufficient viremia to infect mosquitoes, but do not develop clinical illness. Adult chickens (> 22 wks old) usually fail to develop a detectable viremia, and along with immature birds, develop long-lasting antibodies.

The response of wild mammals to natural or experimental infection varies. Serosurveys occasionally have shown higher SLEV prevalence in mammals than in birds, but these data could be confounded because mammalian hosts typically live longer than avian hosts and therefore have a longer history of exposure. Overall, the role of mammalian infection in SLEV epidemiology is complex and difficult to interpret. All reputed *Culex* vectors feed most frequently on avian hosts, occasionally on large mammals and lagomorphs, rarely on rodents, and almost never on bats.

Epidemiology

Human SLEV disease cases typically peak in late summer/early fall and have historically occurred during periods of warm and dry conditions that serve for the optimal development of *Culex* vectors. Furthermore, hot conditions shorten the extrinsic incubation period due to the faster rate of viral replication and reduce the duration of the gonotrophic cycle, increasing the rate of blood feeding and host contact. Transmission of SLEV is complex and requires that the virus replicate in and avoid the immune responses of alternating insect and vertebrate hosts under temperatures ranging from those found in diapausing mosquitoes to more than 43°C in febrile avian hosts. Annual transmission activity may be divided into overwintering, vernal and/or summer amplification, and autumnal subsidence periods.

Three possible mechanisms may explain the persistence of SLEV at temperate latitudes; however, few supportive field data are available. First, low level vertical passage of SLEV from infected females to their progeny has been demonstrated repeatedly in laboratory experiments. Although not detected for SLEV in nature, vertical transmission has been documented for other viruses in the Japanese encephalitis (JEV) complex, including JEV and West Nile virus (WNV). Second, *Cx. p. pipiens* females destined for diapause have been shown to take small blood meals during late summer and early fall without ovarian development. Two isolations of SLEV made from diapausing *Cx. p. pipiens* females collected resting during winter in Maryland were considered to have been infected by

this mechanism, although infection by vertical transmission also was possible. Third, *Culex p. quinquefasciatus* and *Cx. nigripalpus* do not enter reproductive diapause, remain reproductively active or quiescent during winter at southern latitudes and, depending upon ambient temperature, could maintain SLEV by continued, infrequent transmission among resident birds. Experimentally infected, reproductively active *Cx. p. quinquefasciatus* females have been shown to survive winter as gravid females and to then transmit SLEV to recipient birds throughout the following spring. SLEV also may persist over-winter within vertebrate host populations. Passeriform birds infrequently develop chronic infections that persist as long as a year following experimental infection. However, attempts to demonstrate natural relapse or to trigger recrudescence experimentally have not been successful. Flaviviruses, including SLEV, also have been isolated repeatedly from bats, and experimental infections in bats destined for hibernation have been maintained for 20 days at 10°C. When returned to room temperature, SLEV was detected in the brown fat and at low levels in the blood. These data indicated that bats could function as an over-wintering host. However, studies of mosquito host selection patterns indicated that bats rarely, if ever, are fed upon by host-seeking mosquitoes, raising the unanswered question of why so many field-collected bats have neutralizing antibodies. An alternative hypothesis to local persistence involves annual or periodic re-introduction of virus into northern latitudes from southern refugia. Long-distance movement of SLEV has been indicated indirectly from genetic evidence as well as by the re-appearance of SLEV after years of absence. For instance, genotype III SLEV that was identified in California and Arizona in 2015 is most closely related genetically to genotype III SLEV associated with the 2005 Cordoba, Argentina outbreak, indicating long-distance introduction of this virus to North America. Two possible hypotheses address re-introduction, but neither are well supported by field evidence. Many species of birds and some bats have long distance annual migrations that could allow the transport of virus from foci active during winter in southern latitudes or south of the equator to receptive areas north of the equator during spring. These vertebrate migrations typically are very consistent in their summer and winter destinations, and this would allow the same or similar genetic strains to re-appear each summer at the same locality. However, molecular genetic studies of North, Central and South American isolates indicate that they are relatively distinct, thereby implying infrequent genetic exchange. In addition, north-bound migratory birds do not seem to be frequently involved in transmission because they infrequently are found positive for virus or antibody.

Regardless of the persistence mechanism, summer enzootic amplification transmission in North America involves *Culex* mosquitoes and primarily birds in the orders Passeriformes and Columbiformes. Humans become infected tangentially to the primary cycle, do not develop viremias sufficient to infect mosquitoes, and are considered to be “dead end” hosts. Transmission appears to be initiated after *Culex* spp. vectors resume blood-feeding and reproductive activity, and ambient temperatures warm sufficiently to allow the replication of virus in the mosquito vector. Infection is acquired when a female *Culex* mosquito blood feeds on a viremic avian host. Virus imbibed within infectious blood meals taken in early in spring when ambient temperatures average <15°C may lay dormant until warm conditions or changes in mosquito physiology stimulate replication. Under warm temperatures, virus replicates rapidly, disseminates within the mosquito during the ensuing extrinsic incubation period, and then may be transmitted by bite after the female oviposits and attempts to imbibe a subsequent blood meal. The duration of the extrinsic incubation period is temperature-dependent and requires >10 days and perhaps 2 mosquito gonotrophic cycles when temperatures average 22°C. In contrast, the viremia response in susceptible avian hosts typically is of short duration, lasting 2–4 days. Four distinct transmission cycles of SLEV are defined by differences in the biology of the primary vector mosquito species and their distribution, and include: (1) rural North America, west of the Mississippi River transmitted by *Cx. tarsalis*, *Cx. quinquefasciatus* in southern California and *Cx. pipiens* in Washington; (2) rural and urban central and eastern North America transmitted by members of the *Cx. pipiens* complex; (3) Florida, Caribbean and parts of Central America transmitted by *Cx. nigripalpus*; and (4) urban and rural South America transmitted by *Cx. pipiens* complex and mosquitoes of other taxa.

Intensity of enzootic transmission and occurrence of new human cases always subsides rapidly during autumn. Cool evening temperatures slow the replication of SLEV within infected mosquito hosts, decreasing the efficiency of transmission and, concurrently, the combination of cool water temperature and shortening days during larval development initiates reproductive diapause (*Cx. tarsalis*, *Cx. p. pipiens*) or quiescence (*Cx. p. quinquefasciatus*, *Cx. nigripalpus*) in vector females emerging during fall. The fall mosquito population declines in abundance and divides into newly emerged females that do not routinely blood feed and survive the winter, and remnants of the summer population that continue reproductive activity, but usually fail to survive winter. The critical day length that triggers the onset of diapause in *Cx. p. pipiens* may occur in late summer at northern latitudes, markedly shortening the SLEV transmission season. During warm days, however, females may become infected when taking partial blood meals from viremic birds, survive winter and then transmit the virus after diapause is terminated by warm spring temperatures. *Cx. p. quinquefasciatus* does not undergo diapause, so that reproductive activity may continue through winter, albeit at a rate slowed by winter temperatures. Populations exploiting underground storm water systems for resting or for larval development may be exposed to relatively warm temperatures throughout winter.

Five factors have been associated with human risk of SLEV infection: (1) place of residence, (2) age, (3) occupation, (4) socioeconomic status and (5) climate. Clearly, the place of residence markedly affects the risk of infection, with geographic regions in the southern U.S. historically having the greatest numbers of human cases and greatest incidence of disease (Fig. 3). Based on experimental infection patterns in laboratory mice, virus strains from this geographical area also exhibit greater neurovirulence than strains from the western U.S. or South America. Because of greater mosquito abundance relative to humans and host selection patterns, urban residents seem to be at greater risk for SLEV infection than are rural residents; however, these conclusions may be confounded by protective immunity acquired early in life that may be greater among rural residents and by low apparent: inapparent case ratios that require a substantially large population to produce recognizable clusters of human cases.

In the absence of acquired immunity, clinical illness and fatality rates, but not necessarily infection rates, increase dramatically with age. Infection seems to occur equally among different age classes as indicated by the increase in antibody as a function of age in endemic areas and by cohort seroconversion rates determined after epidemics in previously unexposed populations. For example, using data following the 1964 Houston, Texas epidemic, seroprevalence rates remained similar among age cohorts, whereas the case-incidence rates increased from 8.2 per 100,000 for the 0–9 year old cohort to 13.5–27.6 for the 10–59 year old cohort and to 78.0 for the >60 year old group; apparent to inapparent ratios decreased concomitantly from 1:806 to 1:490–1:239 and to 1:85, respectively. Case-fatality rates among 2288 cases reported to the CDC from 1971 to 1983 increased from <6.7% for 0–64 year old age groups to 9.5% for the 65–74 year class to 18% for the >75 year groups.

In the western United States, where SLE historically has been a rural disease, infection risk was found to be greatest among male agricultural workers who frequently lived in suboptimal housing and worked at night. However, infection patterns during urban outbreaks indicated that attack rates were highest among elderly women. These data indicated that there may be differences in risk related to vector species, with elderly women infected most readily during urban outbreaks associated with the *Cx. pipiens* complex, and men working outdoors at greatest risk during rural outbreaks associated with *Cx. tarsalis*. Historically, socioeconomic status has been related closely to the distribution of cases during urban epidemics. Homes and municipal drainage systems frequently were not well maintained in low income neighborhoods, and this was related to the distribution of human cases, but not necessarily the occurrence of virus within the enzootic transmission cycle. Television and air conditioning ownership that brought people indoors during the evening *Culex* host-seeking period was found to reduce risk. Climate variability affects temperature and precipitation patterns, mosquito abundance and survival, virus replication in the vector, and therefore SLEV transmission. Annual changes based on the El Niño/southern oscillation in the Pacific sea surface temperature markedly alter precipitation and temperature patterns over the Americas and cycle with varying intensity at 3–5 year intervals. These cycles alter storm tracks that affect mosquito and avian abundance, the intensity and frequency of rainfall events, and ground water depth; all related to SLEV risk. Above normal temperatures have been especially necessary for northern latitude SLEV epidemics, because elevated temperatures are required for effective SLEV replication within the mosquito vector.

Pathogenesis/Clinical Features

Like many other flaviviruses within the Japanese encephalitis virus serocomplex that can elicit encephalitic syndromes such as WNV and Murray Valley encephalitis virus, human disease is typically recognized in more serious cases that manifest as aseptic meningitis, encephalitis and/or meningoencephalitis. However, the majority of human exposures result in asymptomatic infection. Clinical severity is more notable in individuals >65 years of age. Humans serve as incidental hosts and do not produce viremias sufficient to infect mosquitoes. Similar to most arboviruses that cause CNS disease, infection with SLEV does not result in a clear clinical picture in humans and most infections remain unrecognized, unless associated with an epidemic. When presented with such diverse symptoms, few physicians initially suspect SLEV, even in endemic areas. Most SLEV infections, especially in young or middle age groups, fail to produce clinical disease, and infected individuals rarely experience more than a mild malaise of short duration with spontaneous recovery.

In humans, clinical disease due to SLEV infection may be divided into 3 syndromes in increasing order of severity: (1) *febrile headache* with fever, headache possibly associated with nausea or vomiting, but no CNS illness; (2) *aseptic meningitis* with high fever and stiff neck; and (3) *encephalitis* (including meningoencephalitis and encephalomyelitis) with high fever, altered consciousness and/or neurological dysfunction. The onset of illness may be sudden (<4 days after infection) and acute, leading rapidly to encephalitis, or progressing gradually through all 3 syndromes. Symptoms may resolve spontaneously during any stage of the illness, with full recovery. Acute illness may be followed by “convalescent fatigue syndrome” in <50% of patients, with complaints of general weakness, depression, and the inability to concentrate that generally resolve within 3 years. Other sequelae include headache, disturbances in gait, and memory loss.

Transmission of SLEV from an infected mosquito into the epidermis results in initial infection of either Langerhans cells, epidermal cells, melanocytes or tissue macrophages as initial sites of viral replication based on studies with SLEV as well as other flaviviruses. Viral infected myeloid cells are trafficked to the draining lymph nodes where viral replication is associated with subsequent serum viremia and viral dissemination to alternative tissues including the peripheral and central nervous system. The mechanism(s) of neuroinvasion have not been conclusively demonstrated; however, axonal retrograde transport, introduction into the CNS through the olfactory bulb, a Trojan horse movement of SLEV through infected circulating monocytes and direct hematogenous passage through compromised point of the blood brain barrier have all be suspected as potential routes.

CNS pathology consists of necrosis of neurons and glia cells and inflammatory changes. Inflammatory changes typically are most important in slowly progressing or sublethal CNS disease and sequelae. Viral clearance is dependent upon a functional immune system and the rapid production of neutralizing antibody, which typically appears within 7 days of infection. Viral pathogenesis associated with SLEV infection is principally the result of a combination of viral replication within neural tissue and the subsequent necrosis associated directly with viral infection or the subsequent inflammatory process that neuronal infection elicits.

Diagnosis

The identification of human SLEV cases is primarily based on serological laboratory confirmation of patients exhibiting aseptic encephalitis. Detection of SLEV-specific reactive immunoglobulin is typically made by ELISA with subsequent confirmation by

plaque reduction neutralization testing. Testing of patient sera and/or cerebral spinal fluid for viral RNA by reverse transcription polymerase chain reaction (RT-PCR) within the first five days post-symptom onset also serves as a direct confirmation of SLEV infection.

Prevention

There are no approved antiviral or therapeutic treatment regimens for SLEV human infection and all medical care is supportive. As such, the principal preventive action is to reduce human exposure to mosquitoes. Reducing human/mosquito contact is principally managed by mosquito control activities focused on larviciding, with the implementation of adult mosquito control in situations in which active transmission necessitates targeting infected adult mosquitoes. Best results have been achieved using an integrated management approach that focuses on mosquito vector population suppression through habitat inspection and larviciding. Failure of larval management can be followed by emergency adult control focusing on eliminating infected adult mosquitoes and thereby reducing the force of transmission and preventing human infection. Even if a vaccine was available, protection of the human population by vaccination does not seem cost effective or prudent, because there is no human-to-human transmission, few human infections produce disease, and infection rates remain relatively low, even during epidemics. However, if regional infection rates were to become high, thereby placing selected cohorts at high risk for disease, then selective vaccination may be warranted. There currently is no approved commercial vaccine for SLEV, although vaccination against other flaviviruses such as JEV may impart some protection. Control of avian hosts such as house sparrows and pigeons in urban situations could be done, but this approach is not generally acceptable to the public. Notification of the public of infection risk through the media and the wide-scale use of personal protection through changes in behavior (staying indoors after sunset) and/or repellent application have been credited with reducing the number of infections during the 1990 epidemic in Florida.

Further Reading

- Auguste, A.J., Pybus, O.G., Carrington, C.V.F., 2009. Evolution and dispersal of St. Louis encephalitis virus in the Americas. *Infection, Genetics and Evolution* 9, 709–715.
- Curren, E.J., Hills, S.L., Fischer, M., Lindsey, N.P., 2018. St. Louis encephalitis virus disease in the United States, 2003–2017. *The American Journal of Tropical Medicine and Hygiene* 99 (4), 1074–1079.
- Day, J.F., 2001. Predicting St. Louis encephalitis virus epidemics: Lessons from recent, and not so recent, outbreaks. *Annual Review of Entomology* 46, 111–138.
- Diaz, A., Coffey, L.L., Burkett-Cadena, N., Day, J.F., 2018. Reemergence of St. Louis encephalitis virus in the Americas. *Emerging Infectious Diseases* 24 (12).
- Kopp, A., Gillespie, T.R., Hobelsberger, D., *et al.*, 2013. Provenance and geographic spread of St. Louis encephalitis virus. *mBio* 4 (3).
- Kramer, L.D., Chandler, L.J., 2001. Phylogenetic analysis of the envelope gene of St. Louis encephalitis virus. *Archives of Virology* 146, 2341–2355.
- Monath, T.P., 1980. *St. Louis Encephalitis*. Washington, D.C: American Public Health Association.
- Monath, T.P., Tsai, T.F., 1987. St. Louis encephalitis: Lessons from the last decade. *The American Journal of Tropical Medicine and Hygiene* 37, 40s–59s.
- Ortiz-Martinez, Y., Vega-Useche, L., Villamil-Gomez, W.E., Rodriguez-Morales, A.J., 2017. Saint Louis encephalitis virus, another re-emerging arbovirus: A literature review of worldwide research. *Le Infezioni in Medicina* 25 (1), 77–79.

Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (*Coronaviridae*)

Malik Peiris and Leo LM Poon, The University of Hong Kong, Pok Fu Lam, Hong Kong

© 2021 Published by Elsevier Ltd.

Glossary

Lymphopenia Reduction in the lymphocytes in the circulating blood below the normal range for age.

Pathognomonic Characteristic and diagnostic of a particular disease.

PEGylated A molecule conjugated with polyethylene glycol (PEG).

Radiological abnormalities Pathological findings observed by medical imaging procedures (e.g., chest X-ray).

Rhinorrhea Runny nose.

History

Coronaviruses OC43 and 229E were first discovered in the 1960 as causes of mild upper respiratory disease. The discovery that a novel coronavirus was the cause of severe acute respiratory syndrome (SARS) in 2003 led to increased activity of research on human coronaviruses, resulting in the discovery of NL63, HKU1 and, in 2012 the identification of Middle East respiratory syndrome (MERS)-coronavirus as a novel zoonotic disease with epidemic potential. Both SARS and MERS coronaviruses are identified by WHO as pathogens of greatest concern for global public health for which countermeasures are urgently needed.

Emergence of SARS

From November 2002 to January 2003, clusters of cases of an unusually severe atypical pneumonia were observed in Guangdong Province, China. The disease was characterized by the lack of response to conventional antibiotic therapy and the occurrence of clusters of cases within a family or health care setting. In retrospect, these were the first known cases of the disease that was later called severe acute respiratory syndrome (SARS). In January, the numbers of cases “atypical pneumonia” continued to increase with examples of “super-spreading incidents” that were to punctuate the course of the subsequent SARS epidemic.

On February 21, 2003, a health care worker in a hospital in the city of Guangzhou, the provincial capital of Guangdong, arrived in Hong Kong and checked in at Hotel M. He had treated patients with “atypical pneumonia” in Guangzhou and had been ill himself since 15 February. His one-day stay on the ninth floor at this hotel led to the infection of at least 17 other guests or visitors, some of whom traveled later to Hanoi, Toronto, Vancouver, Singapore, USA, Philippines, Guangzhou, and Australia. Five of these secondary cases initiated clusters of infection in Hanoi, Singapore, Toronto and two clusters of infection in Hong Kong. This was the most significant single event in the global spread of SARS, and arguably the most dramatic single event in the global spread of any infectious disease.

In response to the outbreaks in Hanoi and Hong Kong, on 12 March WHO issued a Global Health Alert warning of a pneumonia that was a particular risk to health care workers. Subsequently, Singapore and Toronto also reported clusters of cases. On 15 March, WHO issued a Travel Advisory. The new disease was named Severe Acute Respiratory Syndrome (SARS) and a preliminary case definition was provided. Within weeks, SARS had spread to affect 8096 patients in 29 countries across five continents with 744 fatalities, an overall case-fatality rate of 9.6%. Health care facilities served as a major amplifier of the infection, constituting 21% of all reported cases. By 21–24 March, the etiological agent of SARS was identified to be a novel coronavirus, subsequently named as SARS coronavirus (SARS-CoV). Early case detection and isolation of infected individuals reduced and interrupted SARS-CoV transmission throughout the world. By July 5, 2003, WHO announced that all chains of human transmission of SARS were broken and the outbreak was at an end. Although SARS was subsequently to re-emerge to cause limited human disease (and in one instance, limited human-to-human transmission) as a result of laboratory escapes and zoonotic transmission from live game animal markets of Guangdong in December 2003–January 2004, the human outbreak of SARS had been efficiently controlled.

Emergence of MERS: Almost a decade later, a virus was identified from a patient with a severe pneumonia in Saudi Arabia and later identified to be a novel betacoronavirus, initially named EMC, but later renamed as Middle East Respiratory Syndrome coronavirus (MERS-CoV). The report of this discovery via Promed on September 20, 2012 led to recognition of other cases, and to a retrospective recognition that a previously undiagnosed cluster of patients with pneumonia in Jordan in 2012 was linked to the same virus. As of October 2020, ca. 2600 confirmed cases of MERS with 880 deaths (34%) have been reported to WHO from 27 countries.

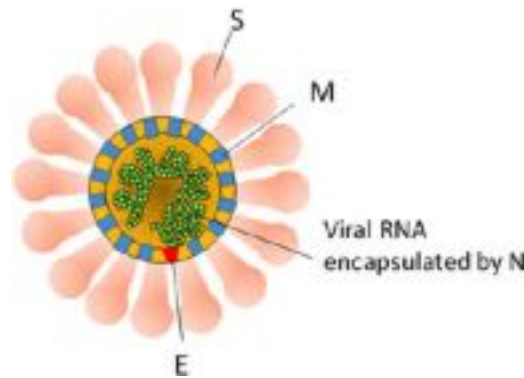


Fig. 1 Schematic drawing of a SARS or MERS viral particle.

Virology

SARS- and MERS-CoVs

Coronaviruses are currently genetically classified into 4 genera (α , β , γ , and δ) and a putative one (ϵ). SARS- and MERS-CoVs are now classified within the sub-genus Sarbecovirus and Merbecovirus, respectively, in the genus *Betacoronavirus* within the family *Coronaviridae*, under the order *Nidovirales*. Coronaviruses are enveloped positive-sense, single-stranded RNA viruses with a genome size of approximately 30 kb. Their virus particles are about 100–140 nm in diameter with a distinctive corona of petal-shaped spikes on the surface which is comprised of the spike glycoprotein (S) (Fig. 1). The S protein is in a trimeric form on the viral surface and it has an N-terminal subdomain (S1) which contains the motifs responsible for receptor binding. A more conserved C-terminal subdomain (S2), which contains heptad repeats and a coil-coil structure, is important in the membrane fusion process. The S1–S2 subdomains remain in a non-cleaved form in newly synthesized progeny viral particles and cleavage of S into S1 and S2 subunits is essential for the viral entry process. The envelope also contains a transmembrane glycoprotein M and in much smaller amounts, an envelope (E) protein. The M protein is a triple-spanning membrane protein and it has a key role in coronavirus assembly. The hemagglutinin-esterase (HE) glycoprotein, found in some betacoronaviruses, are absent in both SARS- and MERS-CoVs. The nucleocapsid protein (N) interacts with the viral genomic RNA to form the viral nucleocapsid. Within the cell, viral replication complexes are located at double-membraned vesicles or autophagosomes within the cell for viral RNA synthesis.

SARS- and MERS-CoV Genomes

The genome organization of SARS- or MARS-CoV is that of typical coronaviruses. The viral genomes of SARS- and MERS-CoVs have, respectively, at least 14 and 11 open reading frames (ORFs) (Fig. 2(A)). Both genomes code for 16 nonstructural proteins (nsp1–16) in their replicase genes (Fig. 2(B)). The capped genomic RNA encoding the replicase gene functions as mRNA to generate polyproteins 1a and 1ab. The translation of ORF1b is directed by a –1 ribosomal frameshift (RFS) signal that contains a nucleotide slippery sequence and an RNA pseudoknot. By contrast, the structural and accessory proteins are products derived from capped subgenomic RNA (sgRNA) which are synthesized by discontinuous RNA transcription.

The polyproteins 1a and 1ab generated from the replicase gene of SARS- or MERS-CoV are cleaved by a papain-like proteinase (part of nsp3) and a 3C-like proteinase (nsp5 or 3CLpro) to generate 16 nsps (Fig. 1(B)). As SARS-CoV is relatively well characterized, the biological functions of these nsps described below are primarily based on the findings of studies on SARS-CoV. Nsp12 is a primer-dependent RNA-dependent RNA polymerase (RdRp), whereas nsp8 is a noncanonical RdRp (nsp8) synthesizing primers utilized by nsp12. The N-terminus of Nsp12 has a domain (nidovirus RdRp-associated nucleotidyltransferase, niRNA) which is unique to nidoviruses. In addition, eight nsp7 and eight nsp8 subunits are able to form a hexadecamer with a hollow, cylinder-like structure. RNA-binding studies and the overall architecture of this nsp7–nsp8 complex suggest that it may encircle RNA and confer processivity of nsp12. The nsp9 is a single-stranded RNA-binding protein and is able to interact with nsp8. The dimerization of nsp9 is essential for virus replication. The nsp10 contains two zinc finger motifs and is suggested to be a regulator of vRNA synthesis. The nsp13 is a helicase and unwinds duplex RNA (and DNA) in a 5′-to-3′ direction. The nsp14, nsp15, and nsp16 have been shown to have 5′-to-3′ exonuclease, endoribonuclease, and 2′-O-ribose methyltransferase activities, respectively. These three proteins, together with nsp3 (see below) are distantly related to cellular enzymes involved in RNA metabolism. These observations may be relevant to viral RNA processing. The exonuclease activity of nsp14 is known to be essential for replication fidelity. The nsp10, 13, 14 and 16 are the subunits essential for forming the cap structure of viral RNA. The nsp3 contains multiple domains and it has 8 domains which are conserved in all coronaviruses (ubiquitin-like domain 1, hypervariable region, X domain, ubiquitin-like domain 2, papain-like protease 2 domain, zinc-finger domain and Y1 and CoV-Y domain). Apart from facilitating viral RNA synthesis and releasing nsp1–nsp3 from the polyprotein 1a and 1ab, nsp3 is also involved in various post-translation modifications (de-ADP-ribosylation, deubiquitination, and de-ISCylation) and evasion of host innate immunity. The biological functions of nsp1, nsp2, nsp4, nsp6, and nsp11 are not entirely clear. The nsp1 is reported to induce chemokine dysregulation and

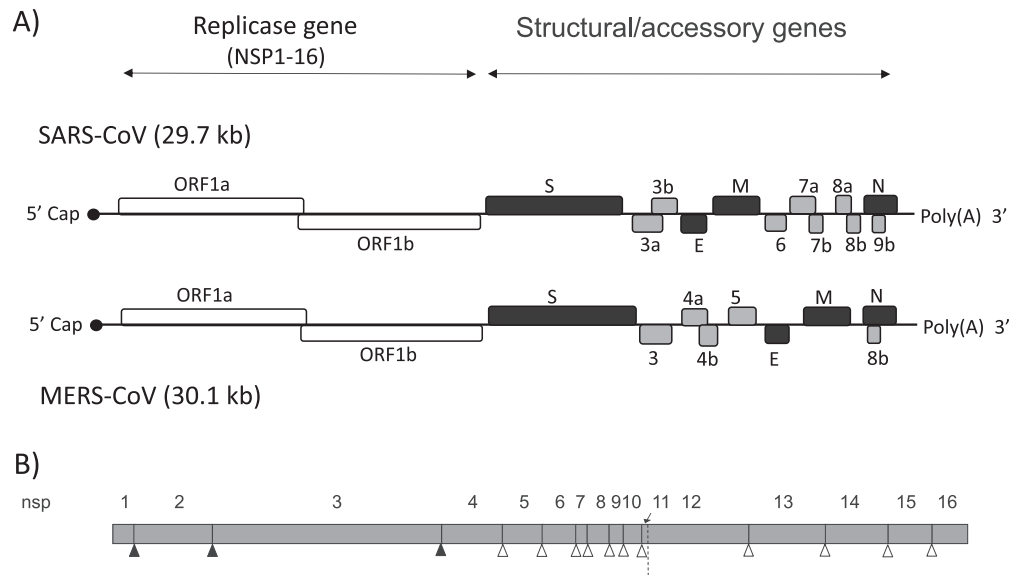


Fig. 2 SARS and MERS CoV genomes. (A) Genomic organization of SARS and MERS CoVs. The ORFs are expressed from the genomic RNA and a nested set of subgenomic mRNA (not shown) that all have a common leader sequence derived from the 5' end of the genome will be expressed. The genomic RNA and all sgRNA contain a 5' cap and a polyadenylated tail at the 3' end. (B) Domain organization of the proteins for ORF1ab. Black and white arrow heads represent the sites cleaved by papain-like and 3C-like proteinases, respectively. The ribosomal frameshift (RFS) site is highlighted by a broken line.

host mRNA degradation, and regulate host gene expression. The nsp2 is reported to be involved in mitochondrial biogenesis and intracellular signaling, but it is dispensable for virus replication. Both nsp4 and nsp6 contain a putative transmembrane domain and they are suggested to be essential for membrane rearrangement. The majority of nsps of MERS-CoV (nsp3, 5, 7–10, 12–16) are assumed to have similar biological functions, whereas the functions of other MERS-CoV nsps are not clear.

Both SARS- and MERS-CoVs have 5 structural proteins (S, E, M, ORF3a and N) encoded by the corresponding ORFs as indicated (Fig. 2(A)). The proteins are the basic protein components of the viral particles. Some studies suggest that ORF3a of SARS-CoV is also a structural protein. This protein can induce membrane rearrangement and activate NF- κ B and the NLRP3 inflammasome in infected cells. However, the precise function of this protein in these virions is not entirely clear.

Apart from the ORFs encoding the replicase and structural proteins, SARS-CoV and MERS-CoV genomes contain additional ORFs that code for accessory proteins (SARS-CoV: 3b, 6, 7a, 7b, 8a, 8b, and 9b and MERS-CoV: 3, 4a, 4b, 5 and 8b). The accessory proteins of SARS- and MERS-CoV have very different sequences, suggesting that the proteins may have very different biological functions. Genetically modified recombinant SARS- and MERS-CoVs without these accessory ORFs have been shown to be replication competent in cell cultures, indicating that the accessory ORFs may not be essential for virus replication *in vitro*. However, recombinant viruses with deletions in these regions are attenuated *in vivo*, suggesting that these proteins may have functions that are important for viral replication, immune evasion and pathogenesis *in vivo*.

The accessory proteins of ORF3b and ORF7a of SARS-CoV induce apoptosis in transfected cells. ORF3b is reported to act as an interferon antagonist during the infection. There is also evidence suggesting that the 7a protein is incorporated into virions. The protein encoded by ORF6 has been shown to inhibit the nuclear import of STAT1 and function as an interferon antagonist in SARS-CoV-infected cells. In addition, this protein is also reported to interact with the viral RNA transcription complex. These properties may be related to virus virulence. Interestingly, comparative sequence analysis of SARS-CoV isolated from palm civets (see below) and humans showed that all animal strains contained a 29-nucleotide (nt) sequence which is absent from most human strains obtained in the later phase of the SARS outbreak. The ORF8 in human SARS-CoVs encodes 8a and 8b proteins, whereas the corresponding ORF in the animal isolates encodes a single protein, known as the 8ab protein. These proteins from the animal and human ORF8 have differential binding affinities to various SARS-CoV structural proteins. Furthermore, the expression of E can be downregulated by 8b but not 8a or 8ab in infected cells. These observations may suggest that the 29-nt deletion modulates the replication or pathogenesis of the human SARS-CoV. The crystal structure of the 9b protein suggests that it may be a lipid binding protein and that it may be a virion-associated accessory protein. But its function is yet to be identified. Overall, these accessory proteins may have a role in viral replication and pathogenesis.

The biological functions of MERS-CoV accessory proteins, except ORF4a and ORF4b, are not clear. All the accessory proteins are dispensable for *in vitro* virus cultures. ORF4a and ORF4b of MERS-CoV are known to suppress innate sensing triggered by dsRNA. Thus, these proteins may have essential roles in virus pathogenesis. Interestingly, deletions in ORF3 and ORF4b sequence can be found in MERS-CoV circulating in West and North African dromedary camels, but not in those animals found in East Africa and the Arabian Peninsula (see below).

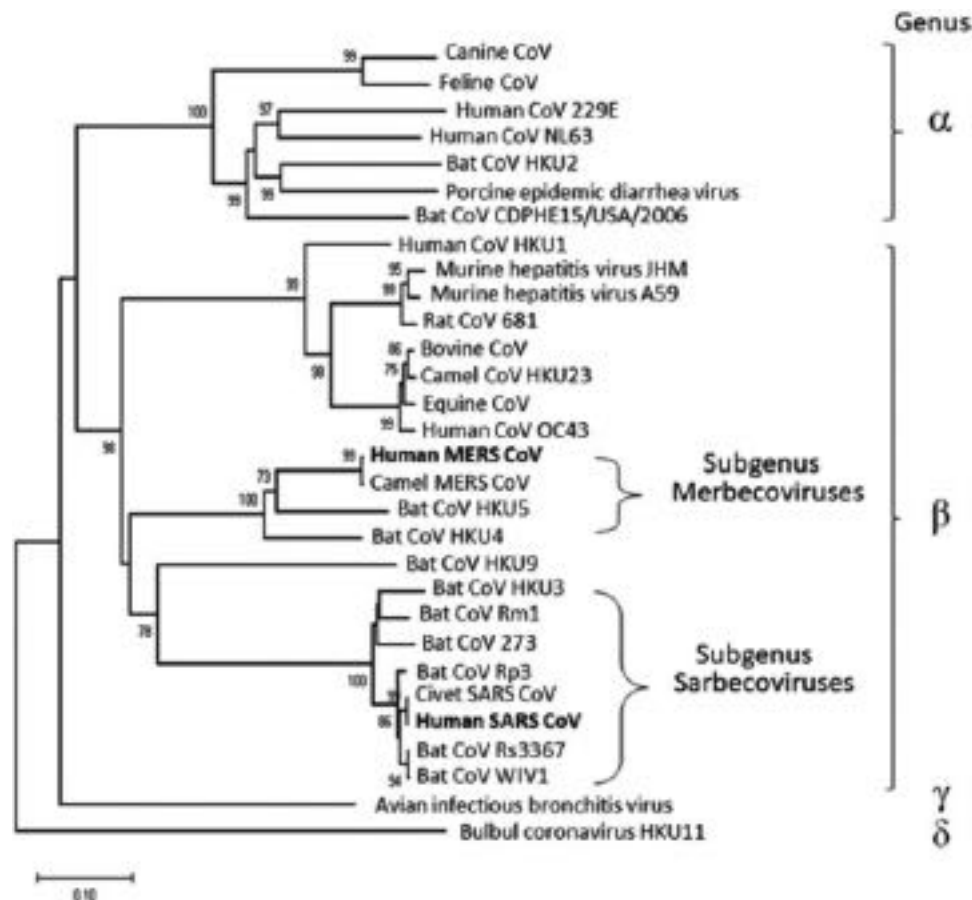


Fig. 3 Phylogenetic analysis of representative RNA-dependent RNA polymerase sequences from α , β , γ and δ coronaviruses. The phylogenetic tree was constructed by the neighbor-joining method and bootstrap values were determined with 500 replicates. Bootstrap values are shown as indicated. Scale bar indicates an estimated genetic distance.

Phylogeny

The phylogeny of nidoviruses is based on the concatenated 5 highly conserved domains in 3 nsps (nsp5: 3CLpro; nsp12: NiRAN and RdRp; nsp13: ZBD and HEL1). Currently, viruses under the same subgenus have patristic pairwise distances less than 0.22. As formal virus taxonomy analyses primarily focuses on (almost) full genome sequences, partial virus sequence fragments derived from potentially novel virus species are not considered in these analyses. Thus, virus diversity within each subgenus is expected to be underestimated. SARS-CoV, MERS-CoV and their related animal viruses are grouped into subgenera *Sarbecovirus* and *Merbecovirus*, respectively (Fig. 3).

The search for the precursor of SARS-CoV led to the discovery of a number of novel coronaviruses in both wild and domesticated animals. In particular, a vast number of alphacoronaviruses and betacoronaviruses were identified in bats. Interestingly, these recently discovered bat coronaviruses appear to be in evolutionary stable while many other mammalian coronaviruses still appear to be under evolutionary selection pressure, raising an intriguing possibility that bats may in fact be the precursors, not only of SARS-CoV, but also for many other mammalian coronaviruses. It is suggested that some other human coronaviruses, such as NL63 and 229E primarily originated from bat coronaviruses.

Genetic and phylogenetic analysis indicates that the viruses associated with the early phase of the human SARS outbreak are more closely related to the viruses found in palm civets and other small mammals in the live game animal markets in Guangdong. The genomes of viruses in the early phase of the human outbreak in 2003 were observed to be under strong positive selective pressure, suggesting that the virus was rapidly adapting to the new host, humans. Furthermore, SARS-CoV in civets was also found to be under strong positive selection pressure, supporting the view that civets were not the natural host of the precursor SARS-like coronavirus.

MERS-CoV found in humans and camels can be divided into 3 clades (A–C). Both clades A and B viruses were found in humans and camels from the Arabian Peninsula, whereas clade C viruses found in African camels are very diverse. The genomes of these viruses are genetically relatively stable, but some studies suggest that the S genes of these camel and human viruses are under positive selection pressure. This may be due to adaptive evolution of these viruses to achieve better attachment to the cells of their new hosts. Deletions in accessory genes of MERS-CoV were found in camel and human viruses. In addition, recombination

between MERS-CoV in camels or humans has been reported, for example, giving rise to the lineage 5 clade B viruses currently causing outbreaks of human disease in the Arabian Peninsula.

Virus Receptors

The S1 subunit of coronavirus spike proteins contains two distinctive structural domains, the N-terminal domain (NTD) and the receptor binding domain (RBD). The RBD binds to the major functional receptor to achieve viral attachment and entry. However, work on other coronaviruses indicate that the NTD may also have strain-specific binding affinity to other host proteins or sugar residues. Both SARS- and MERS-CoVs can bind to multiple non-functional receptors and it is generally assumed that the NTDs may also facilitate virus-host interactions.

The functional receptor for SARS-CoV in human cells is the angiotensin-converting enzyme 2 (ACE-2) which binds the RBD (amino acid residues 424–494) of the SARS-CoV S protein. While the human SARS-CoV S protein binds efficiently to both human and civet ACE-2, the civet-like SARS-CoV S protein binds efficiently to ACE-2 from civets but poorly to human ACE-2. These findings explain the increased human transmissibility of SARS-CoV in the later stages of the SARS outbreak. An observation that human SARS-CoV efficiently infects civets under experimental conditions, the poor virulence and transmissibility of re-emergent SARS in December 2003–January 2004 when humans are believed to have been infected with a civet-like SARS-CoV and the observation that there was approximately 20% sero-prevalence to SARS-CoV in those working in Guangdong live game animal markets although none of them had SARS-like disease. The spike protein of some bat SARS-like coronavirus binds to human ACE-2. Other cell-surface molecules such as L-SIGN, DC-SIGNR, DC-SIGN (CD209), and LSECtin may serve as binding receptors but they do not appear to be functional viral receptors in the absence of ACE-2. They may, however, promote cell-mediated transfer of the virus to other susceptible target cells.

The RBD of MERS-CoV S protein (AA residues 484–567) can bind to its human functional receptor, dipeptidyl peptidase-4 (DPP4). The RBD of human MERS-CoV is identical to that of dromedary camel virus. MERS-CoV can experimentally infect animals having DPP4 with conserved critical binding residues to those of human DPP4 (e.g., camelids, primates, and rabbits). By contrast, animals with critical DPP4 residues that are different to those of human DPP4 are not susceptible to MERS-CoV infection (e.g., mice, rats and ferrets). MERS-CoV is also known to bind to sialic acid, CEACAM5 and GRP78.

Epidemiology

Ecology, Animal Reservoir and Zoonotic Transmission

Both SARS- and MERS-CoVs are of animal origin, but their host origins and modes of zoonotic transmission differ (see below).

SARS

In the early phase of SARS epidemic (before the end of January 2003), 39% of patients with SARS in Guangdong had handled, killed, or sold wild animals or prepared and served them as food. However, such risk factors were found in only 2%–10% of cases from February to April 2003 when the virus had adapted to efficient human-to-human transmission. Thus, the early epidemiological evidence pointed to the live game animal trade as an interface for the emergence of the SARS-CoV. SARS-like coronaviruses were identified in a number of small mammalian species sold in the live game animal markets in Guangdong, including the palm civet (*Paguma larvata*), raccoon dog (*Nyctereutes procyonides*), and the Chinese ferret badger (*Melogale moschata*). A high proportion of individuals working in these markets had developed antibodies to SARS-CoV, although none of them had a history of the disease. Viruses isolated from the re-emergent SARS cases in Guangdong in December 2003–January 2004 were more similar to those found in civets in these markets, rather than to viruses causing the global outbreak in early 2003. These observations strongly implicated that the live game animal trade is the interface for interspecies transmission of a precursor animal SARS-like coronavirus to humans.

SARS-CoV is shed for weeks in experimentally infected palm civets but many of the other species appear to clear the virus more rapidly. While civets in live-animal markets tested positive for SARS-like coronavirus RNA and antibody, civets tested in the farms that supply these markets and those caught in the wild rarely had evidence of infection. Thus, although palm civets and other small mammals sold in these markets were likely the intermediate host that amplified and maintained the virus in these markets and provided the source of virus for repeated human exposure to zoonotic infection, they were not the natural reservoir of the precursor SARS-CoV. Several sarbecoviruses related to SARS-CoV have now been identified in *Rhinolophus* bats in China. Such bats were also sold live in these game animal markets. So far, none of the reported bat CoV is genetically identical to SARS-CoV. It is now believed that SARS-CoV is a recombinant virus of bat sarbecoviruses. Studies in Yunnan, China, has revealed several bat coronaviruses containing all genetic elements needed for generating SARS-CoV in a single geographical location.

MERS-CoV

Is enzootic in dromedary camels which is the main source of zoonotic infection. Bactrian camels have no evidence of infection. The majority of dromedary camels in the Middle East, Africa and Central Asia have serological evidence of MERS-CoV infection. Infection in dromedary camels is characterized as a mild upper respiratory infection. Virus shedding in camel herds is highly seasonal (winter months and the calving season) and is most commonly seen in calves although adults can get re-infected. Settings

where camels from different sources intermingle (camel markets, abattoirs, racing meets) are associated with higher rates of virus detection and may show less pronounced seasonality. Zoonotic transmission is stochastic and uncommon and the routes by which such transmission occurs are unclear. Direct contact with infected camel upper respiratory secretions is one presumed route of infection to humans. A gastro-intestinal route of infection has also been proposed, based on studies of virus tropism in the human gastrointestinal tract and experimental animal models. While virus RNA has been detected in camel milk, presumably derived by the suckling calf contaminating the udder and thence the milk. Viable virus has not been isolated from milk and it is likely that passive maternal antibody found in camel milk would neutralize infectious viruses.

Most zoonotic transmission appears to be asymptomatic or mild, while those individuals with underlying co-morbidities are more likely to manifest severe disease. Over 50% of patients with laboratory confirmed "primary" MERS infection (not explainable by potential contact with other MERS patients) had recent direct or indirect contact with camels. Case-control studies have confirmed direct exposure to dromedaries within two weeks of disease onset as a significant risk factor.

MERS-CoV transmission between humans can lead to outbreaks, especially within health care facilities. Transmission may also occur within households or other community settings but this is less frequent. MERS patients have been diagnosed in 27 countries to date, some associated with index cases acquiring infection in the Arabian Peninsula. Zoonotic MERS has only been reported in the Arabian Peninsula, not in Africa where the majority (>70%) of infected camels are found. The reason for this is at present unclear. MERS-CoV found in African camels is genetically highly similar but phylogenetically distinct from those found in camels in the Arabian Peninsula.

MERS-CoV neutralizing antibody was detected in archived serum samples collected in 1983, indicating that the virus has been circulating in camels for over 30 years. It is unclear whether dromedary camels are the primary reservoir of MERS-CoV. There are several merbecoviruses found in *Tylonycteris* and *Pipistrellus* bats, suggesting that the precursor(s) of MERS-CoV derives from bats but a closely related ancestor is yet to be found.

Human-to-Human Transmission

Both SARS and MERS-CoV appear to be more stable and remain viable for longer on surfaces and in aerosols than other enveloped respiratory viruses. Both viruses are more stable at low humidity and low temperature environment such as those found in air-conditioned spaces. This may explain why the major outbreaks of both have occurred in air conditioned environments such as modern hospitals. The findings are also compatible with epidemiological observations that transmission may occur via fomites, large droplet and aerosol routes. Aerosol generating procedures such as tracheal intubation, manual ventilation before intubation, non-invasive ventilation and tracheostomy have contributed to the transmission in health care settings. SARS-CoV is present in high titers in feces and this may well be a source of infection. The largest single outbreak of SARS in an apartment block in Hong Kong where over 300 individuals were infected from a single index case, is believed to have been caused by aerosols generated from infected feces via a faulty sewage system.

The incubation period for both SARS and MERS ranges from 2 to 14 days. While a asymptomatic infection is rare in SARS it appears to be more common in MERS. The majority of cases did not transmit disease at all and only a few patients accounted for a disproportionately large number of secondary cases, referred as "super-spreading events". While some host factors may play a role in these super-spreading events, in many cases there was a unique combination of host factors and environmental circumstances that facilitated transmission. In contrast to the high transmission rates in these super-spreading events and within hospitals, there was less evidence of secondary transmission within the family or households.

SARS-CoV viral load in the upper respiratory tract was lower in first few days of illness and it peaked in the beginning of the second week of the disease. This correlated with minimal transmission occurring in the first five days of illness and explains why health care-associated outbreaks was more common because hospitalization and use of aerosol generating procedures are more common towards the end of the first week of illness. Taken together, with the lack of asymptomatic infections the public health measures of aggressive case detection and isolation lead to efficient interruption of transmission of the SARS-CoV outbreak in 2003. The characteristics of the epidemiology of MERS is generally comparable to that of SARS, except that asymptomatic infections are more common in MERS.

The basic reproduction number (R_0) of SARS is estimated to range from 2 to 4 while that for MERS over the course of an outbreak is <1. However, there is considerable heterogeneity of effective reproduction number (R_t) during the course of an outbreak, being 1.3–5.4 in the early stages of a hospital outbreak followed by a decrease to <1 within 2–6 weeks.

Clinical Features

The clinical features of SARS and MERS are not pathognomonic. Fever, cough, headache, myalgia, nausea, diarrhea and malaise are commonly reported at early stages of the disease followed by progression to shortness of breath, acute respiratory distress syndrome (ARDS) and multiple organ failure in more severe cases. Rhinorrhea was uncommon in both diseases (Table 1). Lymphopenia was common in both diseases. Differences seen in MERS compared to SARS include a higher proportion of patients who are older and/or with co-morbidities and the presence of hypotension that lead to the need for vasopressors. In both diseases, symptoms may be atypical in immune-compromised patients. A contact history with SARS-CoV positive individuals was an important epidemiological risk factor for SARS. A travel history to the Arabian Peninsula, direct contact with dromedary camels or visiting health care facilities in these affected areas are important factors relevant to initiating specific diagnostic tests for MERS.

Table 1 Comparison of virological, clinical and epidemiological features of SARS and MERS

	SARS	MERS
Virus taxonomy	Genus betacoronavirus; subgenus Sarbecovirus	Genus betacoronavirus; subgenus Merbecovirus
Receptor	ACE2	DPP4
Animal reservoir and zoonotic source	Natural reservoir Rhinolophus bats, Zoonotic interface – live game animal markets	Natural reservoir: unknown. Natural host and source of zoonotic infection: Dromedary camels
<i>Clinical and epidemiological aspects (see notes below)</i>		
Health care workers	21%	Approx. 17%
Males	47%	Approx 63%
Fever	++++	++++
Headache	++	+
Myalgia	+++	++
Cough	+++	+++
Shortness of breath	++	++
Sore throat	+	+
Nausea/vomiting	+	+
Diarrhoea	+	+
Rhinorrhoea	(+)	(+)
<i>Co-morbidities</i>		
Diabetes	+	++
Hypertension	Not noted	++
Chronic lung disease	+	+
Chronic renal failure	<1%	+
Chronic heart disease	+	+
Malignancy	+	+

Approximate frequency of occurrence: + 1%–25%; ++ 26%–50%; +++ 51%–75%; ++++ 76%–100%.

Radiological abnormalities were observed in >60% of cases at initial stages of the disease and preceded lower respiratory tract symptoms in approximately 41% of patients with SARS. In MERS, chest radiographs or computed tomography scans showed multilobar airspace disease, ground glass opacities, and sometimes, pleural effusions, but some patients with mild or asymptomatic infection may show minimal changes in lungs. Liver dysfunction and lymphopenia are seen in both infections. High serum levels of chemokines (IL-8, CCL2, and CCL10) and pro-inflammatory cytokines have been seen in patients with SARS. In MERS, higher serum levels of IL-6 was associated with a severe disease.

The overall case fatality ratio for confirmed SARS and MERS cases has been 9.6% and 35%, respectively. Since MERS-CoV infections, unlike SARS, can often be mild and unrecognized by the health care system, this apparently higher fatality rate may be misleading. In both diseases, infections of children and young adults have been mild. Case-fatality rate increased progressively with increasing age in SARS and with the presence of co-morbidities such as diabetes, cardiac disease, obesity, chronic respiratory disease, end stage renal disease in the case of MERS.

Autopsy findings of those who died in the first 10 days of illness with SARS were diffuse alveolar damage, desquamation of pneumocytes, and hyaline membrane formation. Viral RNA was detected by quantitative polymerase chain reaction (PCR) at high copy numbers in the lungs, intestine and lymph nodes, and at lower levels in the spleen, liver and the kidneys. In the lungs of the patients who died within the first 10 days after disease onset, viral antigen and viral nucleic acid were detectable by immunohistochemistry and *in situ* hybridization in alveolar epithelial cells and to lesser extent in macrophages. Some studies also reported virus particles or viral RNA in other organs but these findings require independent confirmation.

There are limited autopsy studies on MERS. A diffuse alveolar damage was seen in a man who died approximately 12 days after onset of illness. MERS-CoV antigen was detected in pneumocytes and syncytial epithelial cells. No evidence of infection in the kidneys was observed. In an immunocompromised patient with a T cell lymphoma who died of MERS around 3 weeks after onset, histopathological examination showed necrotizing pneumonia, pulmonary diffuse alveolar damage, acute kidney injury, portal and lobular hepatitis and myositis with atrophic changes in the muscles. The brains and the heart were histologically normal. Virus infection was demonstrated in pneumocytes, pulmonary macrophages and renal proximal tubular epithelial cells.

Laboratory Diagnosis

Specimens from the lower respiratory tract such as sputum or endotracheal aspirates have higher viral load and give a higher diagnostic sensitivity than those from the upper respiratory tract for both SARS or MERS. Highly sensitive and specific real-time PCR assays for the detection of viral RNA remain the best choice for early diagnosis. For diagnosis of MERS, initial screening by

RT-PCR targeting the E gene followed by confirmatory tests with RT-PCR for the ORF1 gene region is recommended and only patients positive in both assays are diagnosed as confirmed cases of MERS.

While SARS-CoV RNA remains detectable in the respiratory secretions and feces for many weeks after the onset of illness, specimens rarely yield a virus isolate after the third week of illness, likely because of the emerging antibody responses. Viral RNA can be detected in feces, serum, and urine in patients with SARS, later in the illness. MERS-CoV RNA may also be detected in serum approximately in one-third of patients, especially those with a more severe disease; and in the stool of around 15% of patients, less often than that was seen in SARS. In both SARS and MERS, a negative PCR result in an individual specimen does not exclude the disease. Testing multiple specimens improves the diagnostic accuracy. Both viruses can be cultured from clinical specimens in Vero or Vero-E6 cells but the sensitivity of virus culture is much lower as compared to RT-PCR.

Immunofluorescence or neutralization tests have been used for serological diagnosis or sero-epidemiology of SARS-CoV. A MERS-CoV S1 ELISA is available for the detection of MERS-CoV antibody. Positive results are best confirmed by neutralization tests with live MERS-CoV or spike-pseudotyped virus. Seroconversion occurs during the second week of illness in both diseases and can provide reliable retrospective diagnosis. While a seroconversion is invariable in patients with a severe MERS disease, antibody responses may be marginal or absent in those with a mild MERS infection. Waning antibody responses need to be considered in sero-epidemiological studies. MERS-CoV-specific T cell response has also been used as an indicator of past infection that may complement conventional serology.

Animal Infection Models

Experimental animal models are important in the studies related to pathogenesis, transmission and the effectiveness of therapeutics and vaccines. Experimental animal models for both SARS and MERS are sub-optimal, especially in relation to pathogenesis and there are only few that serve to investigate transmission.

SARS

Experimental infection with SARS-CoV leads to virus replication in Cynomolgous and Rhesus macaques, African green monkeys and marmoset monkeys, with variable manifestations of the disease. Part of the variation in the development of the disease may relate to age (older animals develop more severe disease) or virus strains used.

Mice (BALB/c, C57/BL6, 129SvEv-lineage), Syrian golden and Chinese hamsters, ferrets, and cats are also susceptible to SARS-CoV. Only some of these animals develop pathological lesions in the lungs (ferrets, hamsters, marmosets, aged BALB/C mice). Interestingly, while SARS-CoV replicates in the lung of both young and aged (12–14 months) BALB/c mice, only aged mice manifest clinical symptoms and histological evidence of lung pathology. This is reminiscent of the disease in humans where children suffer only a mild form of the disease. Serial passages in mice has led to the development of virus strains (MA15) with increased pathogenicity in mice. Ferrets are susceptible to SARS-CoV infection and can transmit the virus to other ferrets. They develop fever and show variable histological changes in the lungs but there is no mortality. In most experimental models, viral titers peak early (within first 4 days) and the virus is cleared by the end of the first week, which is not comparable to the human disease. Furthermore, few animal models reproduce the gastrointestinal manifestations of the illness. While an ideal animal model for SARS pathogenesis or transmission is lacking, those that support viral replication (with or without clinical disease) are adequate ones for evaluating the efficacy of vaccines.

MERS

Mice, ferrets, guinea pigs, ferrets and hamsters are not suitable experimental models to study MERS-CoV because the DPP4 receptors of these species do not support virus attachment. Human DPP4 has been transduced into mouse lungs by adenovirus vectors and these mice are susceptible to infection and they develop lung pathology. Human DPP4 (hDPP4) transgenic mice are susceptible to MERS-CoV infection but because the expression of hDPP4 was not physiological in its tissue distribution, infection lead to unusual manifestations such as central nervous system disease, which is not comparable to the one seen in humans. On the other hand, hDPP4 knock-in mice have key regions of the mouse DPP4 replaced with amino acid residues of the hDPP4 and these mice are susceptible to infection. Adapting virus strains by passage in such mice led to virus strains that caused a disease.

Rabbits have DPP4 similar to that of humans and they are, therefore, susceptible to MERS-CoV infection. In contrast to humans, the distribution of DPP4 is predominantly in the upper respiratory tract and consequently, infected rabbits do not develop a lung disease. Marmosets are also susceptible to an infection and a disease although these findings have not been reproduced by different researchers. Rhesus macaques are susceptible to an infection and they manifest a mild form of disease. Dromedary camels, alpacas and llamas are all susceptible to experimental MERS-CoV infection and they can be used as experimental models for infection and transmission. Camels have been used to study the effectiveness of vaccines.

Pathogenesis

SARS

The primary mechanism of lung damage appears to be due to infection of type 1 and type 2 pneumocytes which are the key target cells of the virus. Type 2 pneumocytes are important in the repair of lung injury and infection of these cells can potentially impair the regenerative processes of the lung and aggravate the respiratory impairment. While mice deficient in NK, T or B lymphocytes display similar kinetics of viral replication to that seen in normal mice, infection of mice with defects in the STAT1 signaling pathway results in a more prolonged viral replication and a more severe disease. These findings indicate the importance of innate immune responses in the control of the infection, at least in the mouse. Infection of epithelial cells, macrophages, and myeloid dendritic cells fails to induce a type 1 interferon response although IFN-inducible genes are activated. Multiple accessory proteins mediate innate immune evasion leading to innate immune dysregulation. Patients with a severe SARS had higher plasma levels of IFN- γ , IL-1, IL-6, IL2, TGF β , CCL2, CXCL10, CXCL9, and IL8 compared to those with an uncomplicated disease. There is evidence of viral replication within intestinal epithelial cells but there is minimal cellular infiltrate or disruption of intestinal architecture and the pathogenesis of diarrhea in SARS remains unclear.

MERS-CoV

DPP4, the receptor for MERS-CoV is less abundant in upper respiratory tract epithelium but it is expressed in the epithelial cells of the distal airways and alveoli as well as in the kidneys, intestine, liver, thymus and the bone marrow. Patients with a chronic lung disease have an increased DPP4 expression. Limited autopsy studies and in ex vivo cultures of human lung specimens infected with MERS-CoV have identified evidence of virus infection in ciliated epithelial cells of the bronchioles, type 1 and 2 alveolar epithelial cells in the lung, alveolar macrophages, lung microvascular endothelial cells, renal proximal tubular epithelial cells and human primary intestinal epithelial cells or small intestine explants. The DPP4 distribution and tropism of the virus explains why MERS-CoV causes severe lower respiratory tract infections with minimal upper respiratory symptoms and also explains the mild renal dysfunction observed in patients with MERS and virus detection in the stool and urine in some patients. Unlike SARS-CoV, MERS-CoV can infect human dendritic cells and can infect activated T cells leading to apoptosis. Thus, MERS-CoV may impair adaptive immune responses. As with SARS, MERS-CoV has potent mechanisms for innate immune evasion which contribute to innate immune dysregulation and pathogenesis. Patients with a severe MERS had higher plasma levels of IL-6, IL8, CXCL10 and CCL5 compared those with a mild disease.

Treatment

In the absence of specific antiviral therapies of proven efficacy, good supportive care remains the major form of therapy for both SARS and MERS.

SARS

SARS emerged as a disease of unknown etiology and empirical therapeutic options were initially tested including broad spectrum antivirals and immunomodulators such as ribavirin, intravenous immune globulin, type 1 interferon, SARS convalescent plasma and corticosteroids. However, in the absence of controlled clinical trials, no conclusions can be drawn on the efficacy or adverse effects of these interventions.

A clinical trial of 400 mg lopinavir with 100 mg ritonavir orally every 12 h (added to an existing regimen of ribavirin and corticosteroid therapy) appeared to provide clinical benefit compared to historical controls. However, the lack of concurrent controls makes it difficult to draw conclusions. Similarly, a limited clinical trial of 13 patients using interferon alfacon-1 treatment showed a trend toward improved radiological and clinical outcomes, without, however, reaching a statistical significance. A meta-analysis of observational studies of passive immunotherapy for SARS suggested a decrease in mortality with therapy with convalescent plasma from patients with SARS, particularly if there was evidence of the presence of neutralizing antibodies to SARS-CoV.

Studies in primate models demonstrated prophylactic and therapeutic benefit for PEGylated recombinant interferon α -2b and from small interfering RNA therapies.

Anti-SARS-CoV activity has been demonstrated in vitro, and sometimes in experimental animal models, for several therapeutic substances already in clinical use for other diseases, including lopinavir–nelfinavir, glycyrrhizin, baicalin, reserpine and niclosamide. Screening of combinatorial chemical libraries *in vitro* has led to identification of potential inhibitors of the viral protease, helicase and spike protein-mediated entry.

MERS

Given the promising observational data with the use of convalescent antibody therapy with SARS, a similar approach was attempted for MERS. However, the poor convalescent antibody levels and waning immunity in patients with MERS and because

many patients are elderly or have co-morbid conditions, it turned out very difficult to collect adequate amounts of convalescent plasma for good clinical trials.

A human polyclonal antibody to MERS-CoV produced in transchromosomal cattle has been shown to be safe in phase 2 clinical trials (AIMS). Similarly, several anti-S monoclonal antibodies with neutralizing activity against MERS-CoV are under development and they have shown *in vivo* efficacy in transgenic mice and *Rhesus macaque* models. Some of these antibodies are proceeding to phase 1 clinical trials.

Corticosteroid treatment of MERS-CoV infected patients was not significantly associated with a difference in mortality but it was associated with a delay in MERS-CoV RNA clearance suggesting a lack of efficacy of corticosteroids against SARS-CoV infection.

Based on the data from *in-vitro* and *Rhesus macaque* challenge studies where high doses of ribavirin and IFN- α 2b administered within 8 h post challenge showed partial efficacy in reducing clinical symptoms and viral loads, varying combinations of ribavirin with several interferons (IFN- α 2b, IFN- α 2a, IFN- β 1a) have been used to treat patients with MERS. A retrospective cohort study showed no effect on mortality or viral load, after adjusting for confounding factors. It is now recognized that ribavirin concentrations required to inhibit MERS-CoV need to be far higher than what is clinically acceptable for humans.

A double-blind randomized trial comparing a combination of lopinavir/ritonavir and recombinant IFN- β 1b to placebo is presently ongoing in Saudi Arabia.

Vaccines and Immunity

SARS

A wide range of strategies have been explored for the development of SARS vaccines, including inactivated whole virus vaccines, subunit vaccines (baculovirus expressed S1 subdomain or the complete trimeric spike protein of the virus expressed in mammalian cells), DNA vaccines expressing S (full-length and fragments), N, M, or E proteins; and vectored vaccines based on modified vaccinia Ankara (MVA) virus, vesicular stomatitis virus, adenoviral vectors carrying S, M, or N proteins and attenuated parainfluenza virus type 3 vectored vaccines carrying S, E, M, and N proteins. Neutralizing antibody responses and where appropriate, cell-mediated immune responses have been measured as correlates of immunity. Some of these vaccines have been evaluated in experimental models by challenging with infectious SARS-CoV.

It has been shown that the S protein is the principal immunogen that induces antibody response with neutralization-mediated protection. Passive immunization with human monoclonal antibodies against the S protein has been successful at protecting mice and ferrets from an experimental challenge by reducing viral load in the lungs but not in the nasopharynx. However, S protein-based vaccines have also been shown to elicit Th2 immunopathology following experimental live virus challenge.

A newly emerged SARS outbreak will probably arise from an animal reservoir and it is, therefore, important to investigate cross-protection against animal SARS-like CoV. While human SARS-CoV S protein was neutralized by antibodies to the civet SARS-like virus, civet-like S protein was not neutralized by antibodies against the human SARS-CoV. On the other hand, antibodies against human SARS-CoV appeared to enhance the infectivity of the GD03 and SZ3 pseudo-typed viruses. The development of vaccines that can prevent re-emergence of SARS-CoV from its zoonotic reservoir remains a challenge.

MERS

There are no licensed vaccines against MERS so far, although there are some in clinical trials. Pre-clinical studies have shown that both antibody responses to the receptor binding domain of the spike protein and cellular immune responses are required for full protection against MERS-CoV infection. In spite of genetic heterogeneity of MERS-CoV in Africa, virus neutralization epitopes remain conserved and a protective neutralizing antibody-based vaccine to any MERS-CoV strain is likely to cross-protect against most MERS-CoV strains currently circulating in camels. MERS vaccines based on the virus spike protein, live-attenuated, viral vectored vaccines, DNA vaccines and virus-like particle-based vaccines are in the development and they have undergone initial dose-response tolerability and immunogenicity studies. T cell responses have been consistently observed in MERS survivors and they are more long-lasting than antibody responses. In pre-clinical models, intranasal vaccination with the N protein was shown to generate airway memory CD4 T cells responses that protected against a challenge against multiple CoV.

As with SARS, the possibility of disease enhancement by vaccination needs to be considered. Pulmonary Th2-immunopathology, associated with eosinophilic infiltration and increased pro-inflammatory cytokines in the lungs has been observed after immunization with an inactivated MERS vaccine followed by a wild-type MERS-CoV challenge in a transgenic mice model.

Vaccines have also been developed for vaccinating dromedary camels with an idea of reducing zoonotic infections. A vaccinia (MVA) vectored vaccine administered by an intranasal route was shown to protect camels from an experimental challenge. However, an experimental challenge was carried out within a few weeks of vaccination and it remains unclear whether a vaccine will provide protection for several years in field-settings. It is known that a natural infection does not protect seropositive camels against mucosal re-infection of the nasopharynx. It may be that mucosal immunity is more important than systemic serum antibody-mediated. On the other hand, it is possible that even if the vaccine is not providing sterilizing immunity, vaccination may lead to lower peak viral titers, shorter duration of virus shedding, reduced transmission within camel herds and reduce zoonotic transmission. An important aspect of acceptance of such a vaccine by camel owners is the fact that it also protects against camel pox, which is considered as a more important disease of camels than MERS-CoV infection which is a very mild illness in camels.

Conclusion

Given that the precursor viruses of SARS remain in bats, it is conceivable that SARS may return as a public health threat. However, the wild game animal markets which allowed the virus to transmit to humans are now prohibited in mainland China, thus reducing the possibility of potentially new SARS-like viruses to infect humans. However, the situation is very complex, which is evidenced by the emergence of COVID-19 pandemic caused by a SARS-CoV-2 which likely originated from wild game.

MERS-CoV continues to cause zoonotic infections, followed by outbreaks in health care facilities associated with transmission between humans. Such repeated outbreaks may provide the virus an opportunity to further adapt to humans in order to become a virus with a higher potential to infect humans. In addition, we do not know how extensive MERS-CoV infection is in camel-exposed humans in Africa where >70% of infected dromedary camels are found. It is important to understand the extent of zoonotic transmission and disease, if any, in Africa.

Further Reading

- Arabi, Y.M., Balkhy, H.H., Hayden, F.G., *et al.*, 2019. Middle East respiratory syndrome. *New England Journal of Medicine* 376, 584–594.
- Channappanavar, R., Perlman, S., 2017. Pathogenic human coronavirus infections: Causes and consequences of cytokine storm and immunopathology. *Seminars in Immunopathology* 39, 529–539.
- Fehr, A.R., Channappanavar, R., Perlman, S., 2017. Middle East respiratory syndrome: Emergence of a pathogenic human coronavirus. *Annual Review of Medicine* 68, 387–399.
- Peiris, J.S.M., Lai, S.T., Poon, L.L., *et al.*, 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361, 1319–1325.
- Peiris, J.S.M., Guan, Y., Yuen, K.Y., 2004. Severe acute respiratory syndrome. *Nature Medicine* 10 (12 Suppl), S88–S97.
- Song, Z., Xu, Y., Bao, L., *et al.*, 2019. From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses* 11, 59.
- Sikkema, R.S., Farag, E.A.B.A., Islam, M., *et al.*, 2019. Global status of Middle East respiratory syndrome coronavirus in dromedary camels: A systematic review. *Epidemiology and Infection* 147 (e84), 1–13. [2019].
- Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D.M.E., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine* 367, 1814–1820.

Relevant Website

<https://www.who.int/emergencies/mers-cov/en/>
WHO.
Middle East respiratory syndrome coronavirus (MERS-Cov).

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (*Coronaviridae*)

Malik Peiris, The University of Hong Kong, Pok Fu Lam, Hong Kong

© 2021 Published by Elsevier Ltd.

Glossary

Lymphopenia Reduction in the lymphocytes in the circulating blood below the normal range for age.

Reproduction number The number of secondary infections, on average, generated by one infected person.

RNA Ribonucleic acid.

Non-pharmaceutical interventions Public health measures other than drugs or vaccines that are used to contain an infectious disease outbreak.

Acute Respiratory Distress Syndrome (ARDS) Occurs when fluid builds up in the tiny, elastic air sacs (alveoli) in the lungs leading to reduced oxygenation of the blood.

CD4 T cells T cells bearing the cluster differentiation marker 4. These T cells have a helper function in both cell-mediated and antibody-mediated immune responses.

CD8 T cells T cells bearing the culture differentiation marker 8. These T cells have a cytotoxic function and recognize and kill other cells that express “foreign” antigens which may be of viral or malignant origin.

JAK inhibitors inhibitors of the Janus kinase (JAK)/signal transducers which are activators of the transcription (STAT) pathway are linked to various cytokines and are involved in a variety of immune-mediated and inflammatory diseases.

An unusual cluster of severe pneumonia was noted in Wuhan, China, in December 2019. The etiological agent was identified to be a novel coronavirus closely related, but not identical to, the virus that caused SARS in 2003. The virus was named SARS coronavirus 2 (SARS-CoV-2) and the disease coronavirus virus disease 2019 (COVID-19). Within months, the virus spread to cause a pandemic which has had catastrophic impacts on global health, economy and society. SARS-CoV-2, SARS-CoV-1 (causing SARS epidemic in 2003) and related viruses found in Rhinolophid bats are classified within the sub-genus Sarbecovirus, genus Betacoronavirus, Family Coronaviridae. Closely related viruses have been found in Rhinolophus genus bats, notably RaTG13 and RMYN02, which respectively, share 96.3% and 93% nucleotide identity with SARS-CoV-2. RMYN02 and SARS-CoV-2 both have furin cleavage sites within the spike protein while RaTG13 does not. The natural reservoir from which SARS-CoV-2 emerged is likely to be bats of the Rhinolophus genus. What remains unclear is whether there were intermediate hosts that facilitated transfer and adaptation of the precursor virus to humans. The receptor used by SARS-CoV-2 to gain entry to cells is angiotensin-converting enzyme 2 (ACE-2).

The median incubation period of SARS-CoV-2 infection is around 5 days (range 2–14 days) and the reproduction number (R_0) is estimated to be 2.5. Infected persons may transmit infection from 1 to 2 days prior to onset of symptoms to around 7–10 days after symptom onset. However, severely ill patients and immunocompromised individuals may be infectious for longer periods of time. Pre-symptomatic as well as asymptomatic infections may lead to transmission. The virus is transmitted via large respiratory droplets or respiratory aerosols, predominantly over close range (a few meters) although there are occasional instances of transmission over greater distances. Crowded indoor environments are more conducive to transmission and singing or loud speaking by infected individuals increases the risk of transmission. The virus remains viable for many hours on smooth surfaces (stainless steel, glass, plastic) but survival is much shorter on non-porous surfaces such as cloth or paper. Therefore, indirect transmission from contaminated surfaces via hands to eyes, nose or mouth may potentially contribute to transmission. While the virus RNA can be detected in feces for prolonged periods, infectious virus has infrequently been detected and the degree of infectiousness of feces remains unclear. Super-spreading events are prominent drivers of transmission. In the early stages of the pandemic, non-pharmaceutical interventions including case detection, isolation, contact tracing, quarantine, physical distancing, reduction of mobility and travel related measures were successfully used to reduce transmission.

Symptoms include fever or chills, cough, shortness of breath or difficulty in breathing, fatigue, muscle or body aches, headache, sore throat, congestion or runny nose, nausea, vomiting or diarrhea. Loss of smell or changed sense of taste is frequently reported and is associated with infection and damage of olfactory neurones in the nasopharynx. Progression of clinical disease may lead to hypoxia and acute respiratory distress syndrome (ARDS). Radiological changes include bilateral ground glass opacities and alveolar exudation. Lymphopenia with increased serum transaminase, C-reactive protein and d-dimer levels are commonly seen in severe cases. Progression of disease is associated with difficulty in breathing, leading to ARDS, sometimes leading to a fatal outcome.

The overall infection fatality risk increases progressively with age; those aged 15–44, 65–74 and >75 years having infection fatality risks of 0.03%, 3.1% and 11.6%, respectively; males roughly having twice the risk of females across age spectrums. COVID-19 infection in children and young adults is often mild or asymptomatic. The presence of co-morbidities including heart, respiratory, renal and liver diseases, cancer, diabetes and obesity increase the risk of severe infection and fatal outcome.

Molecular detection of SARS-CoV-2 RNA is the mainstay of diagnosis. Detection of viral protein (usually nucleoprotein) by rapid antigen detection tests give more rapid results and are sensitive in detecting specimens with high viral load who have highest transmissibility of infection. Antibody responses to multiple viral proteins (spike, nucleoprotein, ORF8) and virus neutralizing antibodies are progressively detectable towards the end of the first week after onset of symptoms, and are detectable in most

patients by the end of the third week of infection. Neutralizing antibodies target the spike protein and is protective. CD4 and CD8 T cell responses are also elicited following infection but their role in protection remains to be elucidated.

Direct viral damage as well as immunopathology contribute to pathogenesis, a hyper-inflammatory state being observed in severely ill patients. An intravascular coagulopathy also contributes to pathogenesis, often involving the microvasculature but sometimes leading to thrombosis of large blood vessels with poor prognosis.

Supportive care in the management of patients include provision of supplemental oxygen or mechanical ventilation as and when required. Randomised clinical trials are beginning to identify specific therapies with proven clinical efficacy and this is a fast-moving area of knowledge. There is emerging consensus for the beneficial use of corticosteroids in those patients who require supplemental oxygen or mechanical ventilation. The antiviral drug remdesivir improves time to recovery but does not appear to provide survival benefit when used by itself. However, a combination of remdesivir with immunomodulators (e.g., JAK inhibitors such as baricitinib) may provide improved benefit.

There has been a rapid progress in developing and evaluating COVID-19 vaccines. These have included protein subunit, viral vectored (e.g., adenoviral vectors) and RNA vaccines targeting the viral spike protein; and inactivated whole virus vaccines which elicit immune responses against the structural proteins of the virus. By the end of the year 2020 phase 3 trial data show acceptable levels of efficacy and safety with RNA and adenoviral vectored vaccines targeting the virus spike protein providing evidence that the viral spike is a protective antigen. In December 2020 some countries have started to vaccinate their populations. The duration of vaccine induced protection remains unknown. Most clinical trials evaluate protection from virologically confirmed symptomatic clinical disease and it is unclear whether there will be comparable impact in reducing transmission, a question of key relevance in disease control and population immunity.

Further Reading

- Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* 5, 536–544.
- Chin, A.W.H., Chu, J.T.S., Perera, M.R.A., *et al.*, 2020. Stability of SARS-CoV-2 in different environmental conditions. *Lancet Microbe* 1 (1), e10. doi:10.1016/S2666-5247(20)30003-3.
- Hu, B., Guo, H., Zhou, P., Shi, Z.L., 2020. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* 6, 1–14.
- Krammer, F., 2020. SARS-CoV-2 vaccines in development. *Nature* 586, 516–527.
- Leung, G.M., Cowling, B.J., Wu, J.T., 2020. From a sprint to a marathon in Hong Kong. *The New England Journal of Medicine* 382 (18), e45.
- Mann, R., Perisetti, A., Gajendran, M., *et al.*, 2020. Clinical characteristics, diagnosis, and treatment of major coronavirus outbreaks. *Frontiers of Medicine* 7, 581521. doi:10.3389/fmed.2020.581521.
- Vabret, N., Britton, G.J., Gruber, C., *et al.*, 2020. Immunology of COVID-19: current State of the Science. *Immunity* 52, 910–941.

Simian Immunodeficiency Virus (SIV) and HIV-2 (*Retroviridae*)

Phyllis J Kanki, Harvard T.H. Chan School of Public Health, Boston, MA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Beta chemokines A family of cellular proteins produced in response to acute or chronic inflammation to attract a variety of white blood cells. They bind to cellular targets by specific beta chemokine receptors, that are one of the members of the 7-transmembrane chemokine receptors. Beta chemokines include: CCR5, MIP1 α and MIP 1 β .

CCR5 A cell membrane protein expressed on several cell types including peripheral blood-derived dendritic cells, CD34 + hematopoietic progenitor cells and certain activated/memory Th1 lymphocytes. This receptor is well defined as a major co-receptor for primate lentiviruses, in conjunction with CD4 +, the major cellular receptor. CCR5 is implicated in susceptibility to HIV-1, HIV-2 and some SIV infections.

CD4 + T cell An immune cell, lymphocyte (white blood cell) characterized by the CD4 + antigen (protein) on its surface. This is a T lymphocyte considered to have a “helper” function to enhance the cellular immune response. The CD4 + is the primary receptor for SIV and HIV, and upon infection the virus can destroy the CD4 + cell. The drop in CD4 + T lymphocyte cells is a major determinant of the progression of SIV/HIV infection to AIDS.

CD4 + A large glycoprotein that is found on the surface of helper T lymphocyte cells, regulatory T cells, monocytes, and dendritic cells. Its natural function is as a co-receptor that assists the T cell receptor (TCR) to activate its T cell following an interaction with an antigen-presenting cell. CD4 + is the primary receptor used by SIV and HIV viruses to enter and infect host T cells.

Chemokine receptor A member of the family of seven-transmembrane G-protein-coupled protein molecules on the surface of certain cells that interact with specific chemokines to activate various cellular functions. Chemokine receptors on lymphocytes or monocytes bind to the envelope protein of SIV or HIV and serve as a coreceptor with CD4 (the primary cellular receptor for SIV and HIVs), allowing fusion and entry of the virus into the susceptible cell.

HIV-2 The second HIV virus discovered in registered sex workers in Dakar, Senegal in 1984, the virus is more closely related to the simian immunodeficiency virus (SIV). There are 9 lineages of HIV-2 described, although only A and B are

thought to transmit in human populations. Although HIV-2 can cause AIDS, it has a distinct epidemiology, lower rates of transmission and slower progression to disease.

Long-term non progressors HIV-1 infection and rate of disease progression is variable but the majority of infected patients in the absence of therapy develop AIDS. HIV-1 infected patients with greater than 8–10 years of infection with high CD4 + cell counts have been considered long-term non progressors. Long-term non progressors can be further characterized based on the stability and level of their plasma viral load.

Old World primates The first monkey species of the infraorder Anthroidea are thought to have arisen ~34 million years ago (Oligocene), New World monkeys appeared 30 million years ago most probably originating in Africa but moving to South and North America. The technical distinction between New World platyrrhines (flat sideways facing nostrils) and Old World catarrhines (downward facing nostrils) is the shape of their noses. The first apes lived in African forest ~20 million years ago evolving as a division within the Catarrhine suborder. Two ape genera are endemic in Africa: *Pan* (chimpanzee and bonobo) and *Gorilla* (gorilla). The monkey superfamily of ancestral catarrhine primates or Old World monkeys (family *Cercopithecidae*) in current day are represented by 71 species. Most Old World monkeys are limited to sub-Saharan Africa, although macaques (*Macaca* sp.) are found in Asia.

SHIV Recombinant SIV/HIV chimeric viruses generated in the laboratory, in which subsets of viral genes from one parent virus (SIV) were replaced with the homologous regions of the other (HIV). This has allowed researchers to study particular aspects of virus infection or immunity. SHIVs have been useful in HIV-1 vaccine development research.

SIV A primate lentivirus, first discovered in captive macaques with a simian AIDS syndrome. SIV is known to naturally infect ~40 African Old World primate species without apparent disease. Cross-species transmission events of ancestral SIV gave rise to the origin of HIV-1 and HIV-2.

TRIM5 α TRIM5 α belongs to the large family of tripartite motif (TRIM) proteins with many cellular functions including antiviral defense.

Introduction

The virus responsible for human acquired immunodeficiency syndrome (AIDS) was discovered in the early 1980s. This new virus was a member of the *Lentivirus* genus of the *Retroviridae* family of RNA viruses and later designated as Human immunodeficiency virus type 1 (HIV-1). The first simian lentivirus, Simian Immunodeficiency Virus (SIV) was identified in 1984 from immunodeficient captive rhesus macaques (*Macaca mulatta*), its relatedness to HIV was demonstrated by serologic cross-reactivity to HIV-1 viral antigens. In 1985, serum samples from registered female sex workers in Dakar, Senegal, demonstrated unusual immunoblot

reactivity to HIV-1 with strong reactivity to gag and pol antigens but minimal reactivity to the envelope proteins, suggesting infection with a related virus. The samples were subsequently tested for antibody reactivity to the recently described SIV and they reacted strongly with the envelope proteins, indicating that this new human virus was more closely related to SIV than to HIV-1. The close antigenic relatedness of both SIV and HIV-2 to the prototype HIV-1 virus prompted both the discovery and further classification of these related viruses.

SIV and HIV-2-Taxonomy and Classification

SIV and HIV are viruses belonging to the Lentivirus genus in the family of Retroviridae. The term lentivirus refers to the long incubation period that has been a hallmark of this group of viruses. There are 45 known primate lentiviruses which naturally infect various African Old World monkeys and apes. Phylogenetic analyses have revealed that HIV-1 and HIV-2 are human viruses that resulted from multiple cross species transmissions from monkey and ape SIVs. HIV-1 resulted from four independent cross-species transmissions from chimpanzees (SIVcpz) and gorillas (SIVgor) as well as recombination events. SIV cpz in humans gave rise to the HIV-1 groups M and N. The HIV-1 group O viruses are thought to have arisen from SIVgor but may also have been independently transmitted from a related SIVcpz. Within the SIVs, there is phylogenetic evidence that at least five of them arose from recombination events between two or more parental SIV strains.

HIV-2 in humans resulted from multiple transmissions of SIVsmm (simian immunodeficiency virus, sooty mangabey), this SIV grouping also includes the SIV viruses originally found in captive macaques with an immunodeficiency syndrome similar to human AIDS. Molecular clock studies have estimated the most recent common ancestor of HIV-2 was introduced into humans between 1920 and 1940. Since HIV-2's initial discovery, nine distinct lineages have been identified, only groups A and B are endemic, while the other lineages are represented by 1–2 reports. Individuals harboring non-A and non-B HIV-2 were asymptomatic at the time of the virus description, while the lack of secondary transmission suggests that not all cross-species transmission from SIVsmm monkeys have been successful from the virus' standpoint. Phylodynamic analysis of HIV-2 sequences indicate Guinea Bissau as the geographic origin of Group A and Ghana as the geographic origin for Group B. The global epidemiology of HIV-2 is supported by social-political events such as wars, colonialism, immigration patterns and other population movements.

SIV and HIV-2-Virion Structure

Primate lentiviruses have a common morphology with viral particles 80–110 nm long and 25–50 nm wide, consisting of an electron dense conical or wedge-shaped nucleoid composed of the viral capsid that houses a pair of single-stranded RNA genomes and viral enzymes. The virion is covered by a host cell derived membrane with viral envelope glycoprotein spikes. The envelope spike proteins are more prominent on SIV and HIV-2 compared with HIV-1.

SIV and HIV-2-Genome

The primate lentiviruses have a similar organization of open reading frames within the ~9 kb genome. The genome is flanked by palindromic 5' and 3' long terminal repeat sequences important to viral replication. There are three structural genes, gag (group specific antigen), pol gene encoding the polymerase including the protease, reverse transcriptase and integrase enzymes and env (envelope). There are five more accessory genes, vif (viral infectivity factor), tat (trans activator), and rev that are involved with viral transcription, while vpr, nef, vpx and vpu are variability represented depending on the SIV/HIV grouping ([Fig. 1](#)). The gag and pol messages are unspliced, the env singly spliced, and the major regulatory genes doubly spliced and derived from multiple messages.

SIV and HIV-2 target the CD4 bearing cells including helper lymphocytes, monocytes, macrophages and microglia cells in the brain. Several members of the seven-transmembrane G-protein-coupled receptor family function as coreceptors in association with CD4 to permit viral entry and infection. Different from HIV-1, both HIV-2 and SIVs demonstrate more promiscuous utilization of coreceptors including CCR1–5, CCR8–10, CXCR4–6, GPR1 and GPR15 and many others, some of them associated with specific SIV species. In the case of HIV-2, use of coreceptors other than CCR5 would impact the efficacy of antiretroviral therapy (ART) drugs whose mechanism of action is to block viral infection via CCR5 coreceptor usage.

SIV and HIV-2 Lifecycle

The replication cycle of SIV and HIV-2 include the following steps: (1) adsorption to the major receptor, CD4 and relevant coreceptor with subsequent fusion to the host cellular membrane, (2) penetration and uncoating; (3) reverse transcription is carried out by the viral reverse transcriptase, resulting in a DNA copy of the viral RNA genome; (4) transport of viral DNA to the nucleus; (5) integration of the viral DNA into the host cell DNA which is facilitated by the viral integrase; (6) transcription of the

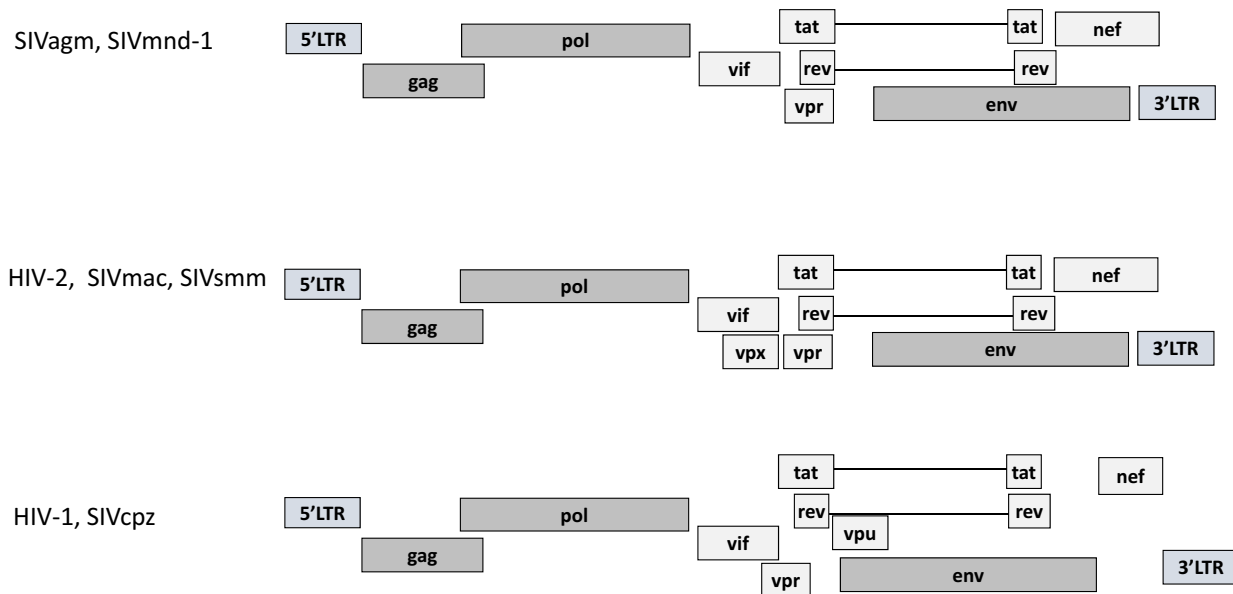


Fig. 1 Genomic structure of representative primate lentiviruses.

viral mRNA and splicing; (7) translation of viral structural and accessory proteins; (8) transport of the viral structural proteins to the cytoplasmic membrane and virus assembly; and (9) viral budding, virion maturation and release.

SIV Epidemiology

SIV was discovered in captive Asian rhesus macaques dying with a syndrome similar to human AIDS. Subsequent studies could not identify SIV in Asian macaques in the wild, whereas numerous African monkeys were found to have high prevalence rates of SIV. In 1986, studies to transmit leprosy from a captive sooty mangabey to a rhesus macaque resulted in simian AIDS, this solidified the evidence that SIV was not a natural infection in the Asian macaque and the artificial infection with the African SIVsmm was pathogenic.

Natural SIV has now been described in 40 different African Old World monkeys and anthropoid apes but not Asian Old World monkeys, such as macaques, in the wild. This suggests that the SIV emerged after the Old World primates began speciation within Africa. Although not extensively studied in the wild, it is considered that SIV in the African monkey and ape is typically non-pathogenic. This is supported by the high prevalence rates found in various serologic surveys, and high levels of viremia despite robust cellular and humoral immune responses to the virus. It has been postulated that a number of host immune mechanisms may distinguish natural SIV infection from the pathogenic HIV/SIV. In at least some African Old World monkeys like the African Green monkey and sooty mangabey, there appears to be host cell restriction to SIV virus entry, mediated by both CD4 and CCR5. Natural SIV infections also appear to resolve immune activation by the virus which thereby preserves central memory CD4 + T cells, the major target in development of immunodeficiency. The nef gene of certain natural SIVs have the ability to downmodulate CD3-TCR from the surface of CD4 + T cells, rendering them resistant to activation. This function has been lost in the SIVs from chimpanzees or HIV-1, apparently linked to the high levels of chronic immune activation and pathogenic outcomes.

SIV Clinical Features

SIV in the natural African primate host is generally non-pathogenic. However, SIV in the Asian macaque host causes an immunodeficiency syndrome marked by profound destruction of CD4 + T cells in the periphery and lymphoid organs throughout the body. The end-stage disease can include opportunistic infections, SIV mediated inflammatory disease, and cancer. The clinical course is dependent on the SIV strain and host macaque species with development of AIDS-like disease between 6 and 24 months. Similar to HIV-1 infection in humans, SIV replicates to high levels in the macaque host with viremia as high as 10^9 viral RNA copies per ml of plasma. The level of SIV RNA in the plasma is an indicator of clinical prognosis in the Asian macaque.

The gastrointestinal tract and other mucosal sites are a major site for SIV induced CD4 + T cell destruction with accompanying immune activation and lymphocyte proliferation. Polyclonal B-cell activation results in hypergammaglobulinemia similar to the human condition. Anemia, weight loss and diarrhea are common clinical features of the disease. Lymphomas are the most common neoplasia seen in SIV infected macaques. Unlike the pathogenesis of infection in the natural African primate, chronic immune activation and inflammation appear to contribute to disease pathogenesis. SIV crosses the blood-brain barrier and infects

cells of the macrophage lineage resulting in viral encephalitis. In addition, SIV antigen and antibody complexes in the kidney can result in immune complex glomerulonephritis.

SIV in the Asian macaque has been studied intensively as a model of HIV-1 infection and pathogenesis. Notably, SIV disease pathogenesis in the Asian macaque demonstrated the importance of lymphoid organ and gut-associated lymphocyte infection and destruction in disease pathogenesis. This model system has also been used to better understand viral and host parameters involved in HIV sexual transmission, including oral, anal and vaginal routes of infection. Anatomic and physiologic similarities of the macaque with humans has increased the relevance of the system to understand the genotypic and phenotypic viral characteristics involved in sexual transmission. Transmission of genetically diverse SIV populations has been shown in the macaque model similar to HIV infection in humans and recombination of viral variants has also been demonstrated after mucosal SIV exposure. SIV has also been useful to better understand HIV perinatal transmission particularly breast milk transmission. Female macaques can be SIV infected after delivery and the infant allowed to breast feed, this approach reduces the variables involved in fetal exposure and transplacental maternal antibody transfer. This has allowed researchers to better understand the dynamics of SIV and HIV breastmilk transmission.

There is significant heterogeneity in the disease progression dependent on the specific SIV strain, molecular clone or Asian macaque species host. In some instances, this diversity has been useful, such as the discovery of SIV variants with specific tropism for monocyte/macrophages that result in SIV meningoencephalitis. However, the differences between HIV-1 and SIV cellular tropism mediated by the CD4 + and coreceptors resulted in an imperfect model to study potential HIV vaccine candidates at the interface of the HIV-1 envelope and vaccine induced immune response. As a result, researchers developed the chimeric SHIV virus consisting of the SIV backbone with an HIV envelope in order to test for the generation of HIV neutralizing antibodies and its ability to protect from a SHIV challenge virus in the macaque host.

Immunity to SIV

SIV in the Asian macaque has also been extensively studied to better understand HIV immunity. HIV-1 could not infect Old World primate cells which led to the discovery of intrinsic host factors such as TRIM5 α with retrovirus-specific antiviral activity. During viral entry into a susceptible host, TRIM5 α binds the viral capsid blocking its function prior to viral integration into the host DNA. Other intrinsic antiviral factors including APOBEC3G, BST-2/tetherin and SAMHD1 were discovered through SIV studies and likely play a role in the viral host range and transmission of these viruses in nature.

In the Asian macaque SIV model there is little evidence for innate immune responses with only slight increases in type 1 interferon and interferon-stimulated gene levels at the site of infection. At the peak of viremia however there is a dramatic increase in innate antiviral immune response in all lymphoid tissues and SIV-specific CD8 + T cells are seen shortly after the peak viremia. The control of SIV replication by CD8 + lymphocytes has been demonstrated in depletion studies using antibodies to the α -chain of the CD8 complex, which is accompanied by corresponding increases in SIV viremia. Similar to HIV infection in humans, detailed studies of well characterized MHC-I-restricted T-cell epitopes in SIV infected macaques has shown targeting of specific SIV peptide sequences with the evolution of viral variants, some of which are capable of escaping the response and increasing the resultant viral population.

The B cell or humoral immune response also plays a role in controlling SIV replication as demonstrated by the transfer of high titer SIV specific-gamma globulin to inhibit SIV replication and slow the progression to disease. B-cell depletion studies have resulted in rapid uncontrolled SIV replication and a shortened course to SIV disease.

Epidemiology of HIV-2

In 1985 when HIV-2 was discovered, the number of AIDS cases reported in West Africa was quite low. The original Senegalese sex workers that had evidence of HIV-2 were healthy, and testing of hospitalized patients that might have AIDS or an AIDS-like disease revealed HIV-1 infection and often a connection to Central Africa but not HIV-2. In 1986, the isolation of HIV-2 (LAV-2) from an AIDS patient originating from Cape Verde, and subsequent AIDS case reports prompted the research community to fear a second global AIDS epidemic. Large scale serologic surveys conducted in healthy and hospitalized patients in West Africa showed varying rates of HIV-2 but the highest prevalence was in healthy adults rather than hospitalized patients. At that time, AIDS cases were reported at higher levels in Central Africa, yet there was no evidence of HIV-2 in AIDS or hospitalized patients from Zaire (DRC), Burundi, Tanzania, Kenya, Zambia and Cameroon.

Although HIV-2 infection in sub-Saharan Africa was concentrated in West Africa, Mozambique and Angola were notable exceptions. These countries although distant from West Africa, were often on the same Portuguese trade routes as Guinea Bissau and Cape Verde, both West African countries with some of the highest rates of infection. In the early 1990s, there were reports of HIV-2 in southwestern India and Goa, a former Portuguese colony. These disparate geographic sites shared economic and political ties with Portugal and some of the first HIV-2 cases in Europe were Portuguese military veterans who had served in the war of independence in Guinea Bissau (1963–1974). In Europe, other former West African colonial powers such as France and England began to detect cases of HIV-2

infection, often with a connection to West Africa. In the late 1980s, France was one of the first countries to screen for HIV-1 and HIV-2 in their blood banks; in current times HIV-2 represents 1%–2% of new HIV diagnoses in France. Portugal is another country with significant HIV-2, likely due to their strong political and economic ties to West African countries such as Guinea-Bissau and Cape Verde. In 2009, over 1800 HIV-2 infections had been identified, one third of which were native borne Portuguese without a known West African connection; suggesting local transmission consistent with an endemic infection.

Currently, longitudinal studies of HIV-2 infection in Guinea-Bissau, Senegal and the Gambia document the decline in HIV-2. In Guinea Bissau, the HIV-2 prevalence rate in the adult populations was 8.9% in 1987, but this was halved by 2007, with the highest rates of infection in the elderly. Modeling data from rural Caió in Guinea Bissau (1990–2007) predicts continued decline in HIV-2 prevalence and incidence with eventual extinction of the virus in this locale by the second half of the century.

Risk and Transmission of HIV-2

Similar to HIV-1, many of the risk factors for HIV-2 heterosexual transmission would include: unprotected sex with an HIV-2 infected partner, genital ulcer disease and lack of male circumcision. The prospective study of registered sex workers in Dakar began in 1985 and analysis of risk determinants of HIV-2 versus HIV-1 infection demonstrated virus-specific differences. The approximate log-linear relation of HIV-2 infection with increasing years of sexual activity was consistent with the hypothesis that the virus had been in the population for at least several decades.

The most common modes of transmission in HIV-2-endemic areas are heterosexual and perinatal transmission; since Senegal like most West African countries had evidence of both HIV-1 and HIV-2 infections, measurement and comparison of incidence rates for both viruses was possible. In this high-risk group, the heterosexual transmission of HIV-2 was significantly slower than that of HIV-1, with HIV-1 increasing 12-fold over an 8-year period, while HIV-2 remained stable. This strongly suggested differences in the heterosexual transmission of these two related immunodeficiency viruses. In different mathematical modeling studies the male-to-female transmission rates of HIV-2 and HIV-1 were estimated with a 3.5–8.9 difference between HIV-1 and HIV-2 in infectivity per sexual act with an infected partner. The mathematical model of the concomitant transmission of the two viruses within the same sexually active population suggested a positive association between pathogenicity and reproductive success, indicating that HIV-1 would competitively displace HIV-2 in the long term. In the study of both viruses in Dakar, Senegal, over more than 25 years, the decrease of HIV-2 prevalence and incidence was accompanied by the increase in HIV-1 prevalence and incidence.

Perinatal transmission of HIV-2 and HIV-1 has been studied in Guinea Bissau, Ivory Coast, France, and Senegal, with all demonstrating extremely low rates of HIV-2 perinatal transmission (0%–4% transmission) in contrast to that of HIV-1 (15%–45% transmission) in the absence of ART prophylaxis. In studies that measured perinatal transmission of both viruses, the rate of HIV-1 transmission was 10- to 20-fold higher than that of HIV-2 (Table 1).

Diagnosis of HIV-2

The close relatedness of HIV-2 to HIV-1 also meant that methods to diagnose the virus, through antibody, antigen, and even genetic material required specific assays. Blood bank screening for HIV-2 was instituted in the United States in 1990, following the lead of many European nations. As a result, the vast majority of commercial HIV ELISA assays utilize a combined antigen source that incorporates HIV-2 specific antigens along with HIV-1 antigens. However, confirmation of HIV-2 serostatus requires HIV-2 specific assays such as peptide specific assays or HIV-2 immunoblots. Immunoblots demonstrating a profile of antibody recognition of HIV-2 envelope proteins are typically used to confirm HIV-2 diagnosis.

Early on it was recognized that HIV-2 was less prevalent globally and therefore commercial companies were slow to develop HIV-2 specific diagnostics. This became even more pertinent, when the description of HIV-2 and HIV-1 dually infected individuals was first noted in West Africa; relying heavily on unequivocal type-specific diagnosis of both viruses with type-specific peptide assays or nucleic acid diagnosis. In recent times, most national serologic surveillance surveys have not distinguished these virus types, and it is likely that in some regions of the world such as West Africa, the serologic surveys are detecting both HIV-2 and HIV-1 without distinguishing them. As a result, much of what is known about the current status of HIV-2 infection even in West Africa is based on research reports.

Table 1 Key Characteristics of SIV, HIV-2 and HIV-1

	<i>Natural SIV</i>	<i>HIV-2</i>	<i>HIV-1</i>
Host	African Old World monkeys and apes	Humans	Humans
Geographic distribution	Sub-Saharan Africa species specific	West Africa-historical ties to West Africa	Worldwide
Plasma viral load	High	Low	High
Perinatal transmission	Low	0%–4% without ART	~25% without ART
Progression to AIDS	Absent	Rare – more slowly	High morbidity

Although numerous genetic tests for HIV-1 are approved and widely used in diagnosis and monitoring of therapy, this is lacking for HIV-2. Babies exposed to HIV-2 from their mothers would require an HIV-2 specific DNA test, which are not available except through a few research laboratories in the US and France. This has also compromised treatment efforts for HIV-2, where quantitation of HIV-2 viral load is required to monitor the success of ART therapy.

HIV-2 Clinical Features

Early case reports described HIV-2-infected people with disease consistent with an AIDS diagnosis. The disease characteristics, including tuberculosis, chronic diarrhea, and *Candida* infections, which were similar to diseases seen in HIV-1-associated AIDS in the same settings. Central nervous system involvement has also been described in HIV-2 AIDS cases. However, classical African AIDS comorbidities, such as tuberculosis, often have had only a weak epidemiological association with HIV-2, even in HIV-2-endemic areas. In addition, large serological surveys of healthy and AIDS patients failed to demonstrate a strong association of HIV-2 with AIDS and a prospective study was warranted.

The prospective cohort study of Dakar sex workers initiated in 1985 compared disease progression in women with known time of infection with HIV-2 or HIV-1, with follow-up exceeding 22 years, representing one of the longest HIV natural-history studies in the literature to date. The majority of HIV-2-infected women (85%) were AIDS-free after 8 years of HIV-2 infection. Using a definition of long-term non-progression employed for HIV-1 infection this would've classified 95% of HIV-2 infected women as long-term non-progressors. In the absence of ART therapy, HIV-2 infected individuals studied in France showed a slower rate of CD4 + cell decline (9 cells/ml/year) compared to HIV-1 infected individuals (49 cells/ml/year). In Guinea Bissau, the policeman cohort study has recently described a median time to AIDS for HIV-2 as 14 years as compared to a median time to AIDS of 6.2 years in HIV-1. Thus, it may be that HIV-2 infected individuals may follow a similar progression to AIDS as HIV-1 although much slower. Since this cohort is dominated by men in distinction from other HIV-2 natural history studies, the role of gender in disease progression warrants further study.

Pathogenesis of HIV-2

As studies continued to document HIV-2's distinct differences in epidemiology, transmission and pathogenicity, other pertinent research questions emerged to identify and characterize the viral and host immune mechanisms responsible. Evidence for a lower viral burden in HIV-2-infected individuals was first reported from the decreased ability to isolate virus and later on quantitative PCR based DNA and RNA studies. While levels of HIV-2 proviral DNA were similar to those of HIV-1 proviral levels, they failed to correlate with levels of viral RNA. Thus, it appears that significant differences occur upon expression, release, and/or maintenance of HIV-2 virions in the bloodstream. Plasma viremia comparisons revealed a 30-fold lower median HIV-2 viral load compared to HIV-1 infected women, irrespective of the length of time infected. This reinforced the concept that plasma viremia was linked to the differences in the pathogenicity of the two related HIVs.

HIV-1 viral integration mapping studies had demonstrated a major effect on viral transcription, however, mapping of HIV-2 integration sites during in vitro infection failed to identify a major difference between HIV-2 and HIV-1. However, accumulation of viral mRNA was attenuated in HIV-2 infection, relative to HIV-1. The differences in viral mRNA were consistent with the differences in plasma viral loads between HIV-1 and HIV-2, and suggested that lower plasma viral loads, and possibly the attenuated pathogenesis of HIV-2, could be explained by lower rates of viral replication in vivo.

HIV diversity is considered a major determinant in the virus' pathobiology. In studies of sequence diversity in the central HIV-2 envelope the rates of viral divergence and diversification were slower in individuals infected with HIV-2 compared to individuals with HIV-1. Viral evolution occurs slowly in HIV-2 infection, which is consistent with the slow disease progression, and supports the notion that viral evolution may be a relevant correlate for disease progression. In detailed analysis of the HIV-2 envelope, while the C2 and C3 regions appear to be well-exposed, similar to HIV-1, they appear to be structurally constrained and consequently mutations are thought to have a negative effect on virus fitness, this could represent a fundamental difference between the biology of the two HIVs and their interaction with the host immune response.

Since slower disease course appeared to be common in HIV-2 infection, it was reasonable to consider that certain subsets of the population would possess host characteristics that might predispose them to a more rapid disease course. Some human leukocyte antigen (HLA) class I alleles have been associated with slower or faster rates of HIV-1 disease progression, suggesting that the presented epitopes or the restricting HLA class I molecule may play a central role in determining the ability to control viral replication. A case-control study investigated possible associations between HLA and the risk of disease progression in HIV-2. The HLA class I status was molecularly typed in female sex workers from the Dakar cohort; HLA B35 was associated with lack of p26 antibodies and higher risk of disease progression. The same association was found for the class I haplotypes B35-Cw4 and A23-Cw7, similar to the association with HIV-1. The data showed that certain HLA molecules are associated with risk of disease progression in HIV-2; some of the alleles and haplotypes involved in susceptibility to disease are similar for both HIV-1 and HIV-2. Therefore, certain genetic factors may be shared by HIV-1 and HIV-2 with respect to susceptibility to enhanced disease progression.

Immunity to HIV-2

HIV-2 Humoral Immunity

The attenuated phenotype of HIV-2 infection *in vivo* has sparked considerable interest in understanding the immunopathogenesis of this particular HIV infection. Although certain viral determinants appear to be central to the lower replicative capacity *in vivo*, the virus appears uniquely immunogenic, which may be central to the explanation of the weaker, more attenuated phenotype of HIV-2 compared to HIV-1. The antibody response to most HIV-2 structural proteins occurs shortly after infection and is long lived. Antibodies to gag, pol, and env encoded antigens are routinely demonstrated with cross-reactivity in the gag and pol to HIV-1 analogous related proteins. IgG1 and IgG2 are the predominant antibody subclasses. Similar to HIV-1, the lack of antibodies to the p26 gag antigen was correlated with disease progression, but unlike the situation with HIV-1, this does not appear to be related to circulating p26 antigen. Also similar to the observations in HIV-1, antibody response to the tat viral protein appeared shortly after seroconversion and were stably maintained, while the majority of HIV-2 infected individuals have antibodies to HIV-2 tat, those that lack the antibodies are more likely to progress to clinical disease.

The low viremia, reduced transmission and slow disease progression of HIV-2, led to the interest in HIV-2 as a model of lentivirus control. The identification of potent and broadly neutralizing antibodies has been considered critical components for an effective HIV-1 vaccine. Autologous and heterologous neutralizing antibodies (NAb) responses in HIV-2 infection have been studied, where heterologous refers to an HIV-2 individual's sera to neutralize HIV-2 viruses other than their own. NAb to autologous HIV-2 was considered as a potential mechanism for HIV-2 viral control, however, studies failed to demonstrate a significant inverse relationship between NAb titers and plasma viremia. In multiple studies of the heterologous HIV-2 NAb responses, potent NAb responses were found with a significant positive association between these titers and HIV-2 plasma viral load. In individuals with higher viral loads suggestive of disease progression, the higher heterologous NAb production may reflect abnormal polyclonal B- cell activation, similar to what is seen in HIV-1. The development of NAb responses was analyzed in HIV-1 infected Dakar sex workers assessed longitudinally for up to 14 years with heterologous responses developing fully after 2.5 years of infection, this is similar to what is seen in the development of HIV-1 NAb responses. There was no correlation between titers of either heterologous or autologous NAb and clinical progression in the cohort and HIV-2 envelopes failed to demonstrate any evidence of escape from NAb selection pressure. Thus, HIV-2 appears to induce superior NAb titers which may result from preserved CD4 + T cell help that is maintained in HIV-2 infection. This is coupled with the lack of HIV-2 variability which allows for persistence of NAb epitopes that would promote higher affinity maturation by B cells.

Antibody-dependent complement-mediated inactivation by HIV-2 has not been well studied. One study compared the effect of complement on HIV-1 and HIV-2 and found a higher magnitude of anti-viral activity in HIV-2 compared to HIV-1 via activation of the classical complement pathway via IgG. Complement increased the HIV-2 antiviral activity 32-fold whereas the increase in HIV-1 antiviral effect was modest. This may result from intrinsic differences in the envelope spike proteins of HIV-2 versus HIV-1, where the HIV-2 envelope is more accessible with functional and stable envelope spike proteins.

HIV-2 Cellular Immunity

T cell activation in HIV-2 infection has been studied where CD4 + and CD8 + T cell populations were independently assessed with the hypothesis that distinct factors mediate immune activation in specific T cell subsets. CD8 + T cell immune activation has previously been strongly associated with HIV-1 disease progression. CD8 + T cell activation was significantly different between HIV-2 and HIV-1 infected individuals although both groups expressed significantly higher activation level compared to seronegative individuals. Unlike HIV-1, there was no association of CD8 + or CD4 + activation with HIV-2 viral load or CD4 + cell count.

Numerous studies have described robust gag specific CTL responses in HIV-2 infection, in one study, 88% of patients demonstrated CD8 + mediated responses. These were notable for their presence even in patients with undetectable HIV-2 viremia. The responses were found to be more polyfunctional in HIV-2 compared to HIV-1 as measured by secretion of IFN- γ , IL-2, TNF- α , MIP-1 β and CD107 α mobilization. HIV-2 CD8 + responses to HIV-1 was also demonstrated, HLA-B5801 positive subjects showed broad cross-recognition of various HIV-1 subtypes suggesting a response that tolerated extensive amino acid substitutions within the HLA-B5801-restricted HIV-1 and HIV-2 epitopes.

Treatment of HIV-2

There has been tremendous progress and success in the science and clinical testing of antiretroviral therapies (ART) for HIV-1 infection. Previous and even current guidelines for ART treatment of HIV-2 have often been based on HIV-1 without significant scientific evidence or clinical testing. This is at least partially due to the overall lower global prevalence of HIV-2, its African geographic localization and the paucity of AIDS-related HIV-2 cases in resource rich settings. There are also diagnostic challenges to the clinical monitoring of HIV-2 therapy. The lack of approved nucleic acid tests to confirm HIV-2 infection, monitor viral loads and characterize host-cell tropism are lacking. There has been no formal regulatory approval of ART drugs for HIV-2 treatment and

only in 2018 have results from small clinical trials been reported. Several other trials are ongoing which will help to determine the optimal first-line therapy regimen for HIV-2.

HIV-2 is considered susceptible to the class of nucleoside reverse transcriptase inhibitors (NRTIs) which are typically combined with non-nucleoside reverse transcriptase inhibitors (NNRTI) in most first-line ART regimens in Africa. Early in vitro studies indicated high level resistance of HIV-2 to NNRTI class of drugs. This is due to the known mechanism of action of this class of ART drugs where the drug binds to a specific subunit of the HIV-1 reverse transcriptase enzyme, that is not present in HIV-2. There is active research on protease inhibitors that may be effective in HIV-2 ART. There are currently 10 approved protease inhibitors for HIV-1 treatment, only four of these (ritonavir boosted -lopinavir, -darunavir and -saquinavir) are considered active for HIV-2 treatment due to naturally occurring secondary mutations or polymorphisms in the HIV-2 protease. Upon selection of primary resistance mutations to protease inhibitors, pre-existing secondary changes may explain the acquisition of multi-PI resistance phenotype. HIV-2 is also intrinsically resistant to the fusion inhibitor enfuvirtide. Maraviroc is an entry inhibitor that blocks HIV entry via the chemokine receptor, CCR5. While there is limited in vitro data suggesting that HIV-2 may be susceptible to this drug, it is anticipated that the virus' ability to utilize other non-CCR5 coreceptors may compromise in vivo efficacy.

HIV-2 treatment in West Africa has relied on the availability of drugs available for standard HIV-1 treatment. Historically, the typical first line ART regimens in Africa were NRTI and NNRTI based regimens. Since HIV-2 patients would not be susceptible to NNRTI drugs, they were initiated on therapy with NRTIs and available protease inhibitors. Reports from those HIV-2 infected West African patients indicated high clinical and virologic failure, mortality and loss to follow-up. Upon genotypic analysis, a significant proportion (30%–65%) of failing patients had developed multiclass NRTI and protease inhibitor resistance.

There are no current guidelines for second-line ART for HIV-2 infected patients. There are ongoing clinical trials to determine if drug resistance testing can provide clinical benefit in HIV treatment. In recent years the WHO guidelines for first line ART have included integrase inhibitors. Clinical studies support their use as first-line ART in HIV-2 and this has been bolstered by in vitro selection studies, biochemical analyzes and studies with site-directed HIV-2 mutants.

Prevention of HIV-2

Although HIV-2 transmission by sexual and perinatal transmission appears to be reduced compared to HIV-1, risk factor analysis confirms that methods of prevention for both viruses should be effective. Prevention of HIV-2 by sexual transmission would include protected sexual intercourse, reducing concomitant sexual transmitted diseases and male circumcision. In addition to sexual and perinatal transmission, blood borne transmission should also be considered. Screening of blood banks for HIV-2 has been instituted since the 1990s in most countries. While HIV-2 transmission through intravenous drug use is biologically possible, current epidemiology has not documented significant HIV-2 infection in intravenous drug users. Although these interventions have been extensively studied with HIV-1, they have not been studied with HIV-2 because the transmission rate has been relatively low and geographically isolated to West Africa.

Antiretroviral prophylaxis has been an effective means of preventing perinatal transmission of HIV-1 and would be considered a means of preventing HIV-2 perinatal transmission. Since the majority of HIV-2 infected pregnant women would have low HIV-2 plasma viral loads and reduced cervical shedding, their risk of perinatal transmission is considered to be significantly lower than HIV-1 infected woman, 0%–4% with or without prophylaxis. As discussed in the HIV-2 therapy section, there are certain classes of ART that are ineffective in blocking HIV-2 infection, so drugs for HIV-2 prophylaxis must take this into consideration. In the US, a regimen with two NRTIs and a ritonavir-boosted protease inhibitor or integrase strand inhibitor are recommended; NNRTIs should not be used since they are not active against HIV-2. It is recommended that all infants borne to an HIV-2 infected woman receive 4 weeks of zidovudine prophylaxis. Since breastmilk transmission is also potential, breast-feeding is not recommended in HIV-2 infection. Again, there are no clinical trials that have addressed the efficacy of these interventions to prevent HIV-2 perinatal transmission.

Interactions of HIV-2 and HIV-1 in Vivo

Early in the description of HIV-2 in West Africa, serologic surveys for both HIV-2 and HIV-1, found samples that reacted with both viruses. This created a diagnostic challenge and required the development of type specific antibody assays or nucleic acid tests to confirm the HIV-dual infection status of the individual. Isolation of both HIV-1 and HIV-2 viruses has been reported from selected HIV-dual cases and nucleic acid evidence has been demonstrated in most of these with the improvement of these assays. HIV dual in regions of the world that lacked evidence of HIV-2 infection alone, were considered to result from misinterpretation of laboratory data. However, in West Africa, where HIV-2 and HIV-1 infection co-exist, HIV-dual infection has been well documented.

HIV-2 Protection From HIV-1

Demonstrated differences in the infectivity and disease potential of HIV-2 compared to HIV-1 supported the notion that upon interaction, HIV-2 might protect from HIV-1 infection or disease, analogous to an attenuated virus vaccine model. In 1995, an

8-year study of the Dakar female sex worker cohort, the hypothesis that the attenuated phenotype of HIV-2 infection might protect from subsequent HIV-1 infection was addressed. HIV-1 infection in previous HIV negatives along with super-infection of HIV-2 infected was documented over the study period with both serology and nucleic acid assays. A Poisson regression model was used to estimate the independent effect of demographic, behavioral, and biologic variables on the risk of HIV-1 infection. Despite higher incidence of other sexually transmitted diseases, HIV-2 infected women had lower incidence of HIV-1 than seronegatives, with a statistically significant incidence rate ratio (IRR) of 0.32. When immunosuppression by reduced CD4+ cell count was accounted for, the IRR associated with HIV-2 was reduced further to 0.23. This analysis led to the conclusion that HIV-2 infection conferred a significant reduction (32%) in the subsequent risk of HIV-1 infection.

The generalizability of these findings was questioned by studies from other West African sites. In Ivory Coast, Guinea Bissau and the Gambia, studies originally designed as cross-sectional surveys were analyzed for short periods of longitudinal observation, as a result of their design they did not possess sufficient statistical power, capable only of detecting an extremely high protected fraction (>99%) of HIV-1 infection due to HIV-2 infection. Continued analysis of the Dakar cohort extended the observation period from the first published report to over 13 years, HIV-2 protection ranged from 52% to 74% depending on the method of analysis.

In 2012, results from a cohort of 233 HIV-dual and HIV-1 infected in Guinea Bissau policemen were reported, the cohort had been followed for ~20 years. They reported 53% slower progression to AIDS in individuals with dual infection compared to those with HIV-1 infection, as reflected by higher CD4+ cell counts. Dual infected individuals where HIV-2 infection had preceded HIV-1 had the longest progression time to AIDS, median time to AIDS estimated at 129 months (93–165) compared to singly HIV-1 infected individuals with a median time to AIDS of 68 month (60–76). This was accompanied by significantly higher CD4+ cell counts. This study was extended with further analysis to show that HIV dual infected individuals, where the HIV-2 infection had preceded HIV-1, had longer survival times compared to individuals with HIV-1 alone, measuring all-cause mortality as the outcome measure. This suggested that HIV-1 disease progression was inhibited by prior HIV-2 infection and again demonstrated the importance of in vivo studies to understand the pathogenesis of these related viruses. These results also seemed to confirm the 1995 findings from the Dakar cohort, and further strengthen the view that prior HIV-2 infection could modulate the pathogenicity of HIV-1 in vivo.

In vitro evidence for HIV-2 protection from HIV-1

Studies have described in vitro interactions of HIV-1 and HIV-2 that support the in vivo observations, these range from virus-virus interactions to potential immune mediated mechanisms for HIV-2 protection. Prior studies had reported that HIV-2 inhibits the replication of HIV-1 at the molecular level. This inhibition was selective, dose-dependent, and nonreciprocal. Though the exact mechanism remains to be defined, the inhibition appeared to be mainly due to an intracellular molecular event because it could not be explained solely on the basis of cell surface receptor mediated interference. The results supported the notion that the inhibition likely occurred at the level of viral RNA, possibly involving competition between viral RNAs for some transcriptional factor essential for virus replication.

Using an in vitro HIV-1 challenge system, PBMCs from HIV-2 infected women could not support replication of a CCR5-dependent HIV-1 virus compared with CXCR4-dependent virus. Resistance was transferable, CD8 dependent and strongly correlated with beta chemokine production in the media. All resistant cultures were rendered susceptible by addition of a cocktail of antibodies to beta chemokines. HIV-2 infection might dramatically influence beta chemokine production by enhancing it in magnitude and/or duration, thus enabling HIV-2 infected individuals to cope favorably with subsequent exposure to HIV-1. This is supported by studies demonstrating that binding of the HIV-2 envelope to the alpha chain of CD8 stimulates dramatic levels of beta-chemokine production in comparison to HIV-1 envelope stimulation activity. Further, unstimulated HIV-2 PBMCs had diminished surface CCR5 receptor expression on CD4+ T cells. In vitro up-regulation of the CCR5 receptor was readily demonstrated indicating that the receptor was functional. The in vivo down-regulation of the receptor was not correlated with activation markers (HLA-DR), beta chemokine levels or plasma viral load. It was postulated that the down regulation of the CCR5 receptor in HIV-2 infection contributed to slower disease course, by decreasing susceptibility of target cells for infection in conjunction with HIV-2 specific cellular immune responses. It is recognized that CTLs are able to lyse infected cells before progeny virions are produced and the production of granzyme particles will elicit beta chemokines therefore providing two mechanisms by which viral replication could be limited or prevented.

In the Guinea Bissau policeman cohort, HIV-1 disease progression was most attenuated in HIV-2-infected subjects who had stable asymptomatic disease. This suggested that HIV-1 was inhibited by immune responses mounted to HIV-2, where CD8+ gag-specific T cells had already been documented to be present in the majority of individuals with asymptomatic HIV-2 infection. Extensive cross-reactivity has been demonstrated between HIV-1 and HIV-2 epitopes, particular for CD4+ and CD8+ T-cells directed against the most conserved regions of the gag proteins between HIV-2 and HIV-1. It is therefore plausible that the potent and unusually high avidity T-cell responses elicited by HIV-2 contribute to the control of HIV-1 replication in vivo. Although HIV-2 infection is also characterized by very potent neutralizing antibody response that sometimes cross-neutralize HIV-1, the neutralization of HIV-1 is much less potent. HIV-2 infection results in a low level persistent viral infection, which elicits robust humoral, innate, and adaptive cellular immunity; cross-reactivity likely inhibits the replication of HIV-1 and ensuing disease progression. The mounting of the immune response to HIV-2 would be critical to the protective effect for both HIV-1 immunopathology and the development of disease.

Since its initial discovery, studies of HIV-2 in infected individuals in West Africa have provided compelling evidence for potential mechanisms that may be critical to the pathogenicity of HIV viruses in general. Observational and in vitro studies suggest that the attenuated HIV-2 virus infection may provide cross-protective immunity for HIV-1. Suggesting that one plausible explanation for the relatively low rates of HIV-1 infection may be due to the co-existence of HIV-2, a unique feature of West Africa. Unbiased, powerful studies, using sensitive and specific classification methods, will effectively address the generalizability of the observation of HIV-2's protective efficacy against subsequent HIV-1 infection described in West African studies over multiple decades. The investigation of virologic and immunologic mechanisms responsible for HIV-2 attenuated phenotype and protection from HIV-1 disease has significant implications for HIV pathogenesis studies and vaccine development.

Further Reading

- Bell, S.M., Bedford, T., 2017. Modern-day SIV viral diversity generated by extensive recombination and cross-species transmission. *PLoS Pathogens* 13 (7), e100646.
- Chahroudi, A., Bosinger, S.E., Vanderford, T.H., Pairedini, M., Silvestri, G., 2012. Natural SIV hosts: Showing AIDS the door. *Science* 335 (6073), 1188–1193.
- Esbjornsson, J., Mansson, F., Anders, K., *et al.*, 2012. Inhibition of HIV-1 disease progression by contemporaneous HIV-2 infection. *New England Journal of Medicine* 367, 224–232.
- Gottlieb, G.S., Raugi, D.N., Smith, R.A., 2017. 90–90–90 for HIV-2? Ending the HIV-2 epidemic by enhancing care and clinical management of patients infected with HIV-2. *Lancet HIV* 5, e390–e399.
- Kanki, P.J., Mboup, S., 2015. HIV-2: Lessons from the Dakar cohort. In: Hope, T., Richman, D.D., Stevenson, M. (Eds.), *Encyclopedia of AIDS*. New York, NY: Springer, pp. 1–17.
- Kirmaier, A., Krupp, A., Johnson, W.E., 2014. Understanding restriction factors and intrinsic immunity: Insights and lessons from the primate lentiviruses. *Future Virology* 9 (5), 483–497.
- Marlink, R., Kanki, P., Thior, I., *et al.*, 1994. Reduced rate of disease development with HIV-2 compared to HIV-1. *Science* 265, 1587–1590.
- Nyamweya, S., Hegedus, A., Jaye, A., *et al.*, 2013. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Reviews in Medical Virology* 23, 221–240.
- Travers, K., Mboup, S., Marlink, R., *et al.*, 1995. Natural protection against HIV-1 infection provided by HIV-2. *Science* 268, 1612–1615.
- VandeWoude, S., Apetri, C., 2006. Going wild: Lessons from naturally occurring T-lymphotropic lentiviruses. *Clinical Microbiological Reviews* 19, 728–762.
- Visseaux, B., Damond, F., Matheron, S., Descamps, D., Charpentier, C., 2016. HIV-2 molecular epidemiology. *Infection, Genetics and Evolution* 46, 233–240.
- Zhou, Y., Bao, R., Haigwood, N.L., Persidsky, Y., Ho, W.Z., 2013. SIV infection of rhesus macaques of Chinese origin: A suitable model for HIV infection in humans. *Retrovirology* 10, 89.

Relevant Websites

- <https://www.health.ny.gov/diseases/aids/providers/testing/docs/guidelines.pdf>
2018 Guidelines for use of the HIV Diagnostic Testing Algorithm for Laboratories.
- <https://www.uptodate.com/contents/clinical-manifestations-and-diagnosis-of-hiv-2-infection>
Infection.
- <https://www.sciencedirect.com/topics/medicine-and-dentistry/simian-immunodeficiency-virus>
Simian Immunodeficiency Virus.

Sindbis Virus (*Togaviridae*)

Satu Kurkela, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

arbovirus Arthropod-borne virus

BFV Barmah Forest virus

CHIKV Chikungunya virus

E1 Envelope glycoprotein 1

E2 Envelope glycoprotein 2

EEEV Eastern equine encephalitis virus

MAYV Mayaro virus

ONNV O'nyong-nyong virus

ORF Open reading frame

RRV Ross River virus

SINV Sindbis virus

VEEV Venezuelan equine encephalitis virus

WEEV Western equine encephalitis virus

Classification

Sindbis virus (SINV) is an arbovirus, which belongs to Western equine encephalitis (WEE) complex of the genus *Alphavirus* in the family *Togaviridae*. The phylogenetic relationships of alphaviruses resemble those of the suggested serocomplex division (SFV, WEE, EEE, VEEE). A phylogenetic division of SINV into six genotypes has been suggested, with African and European virus strains forming genotype 1, Australian, East Asian and Central Asian strains genotypes 2–4, the closely related Whataroa virus genotype 5, and the South-West Australian strains genotype 6 (Fig. 1). Alphaviruses that are pathogenic to humans can also be divided into arthropathic and neuropathogenic viruses. SINV is the only arthropathic alphavirus belonging to the WEE complex, while the others (Chikungunya, Ross River, Barmah forest, O'nyong-nyong, and Mayaro) belong to the SFV complex. SINV, alongside with e.g., SFV, is widely used as a model for viral encephalitis in mice, as an expression vector in cell biology, as well as in versatile gene therapy applications.

Virion Structure

SINV is a spherical virus of 70 nm in diameter with $T = 4$ icosahedral symmetry. SINV particle consists of a nucleocapsid core, a host cell derived lipid membrane, and a glycoprotein shell. The nucleocapsid core includes an RNA genome and 240 copies of the

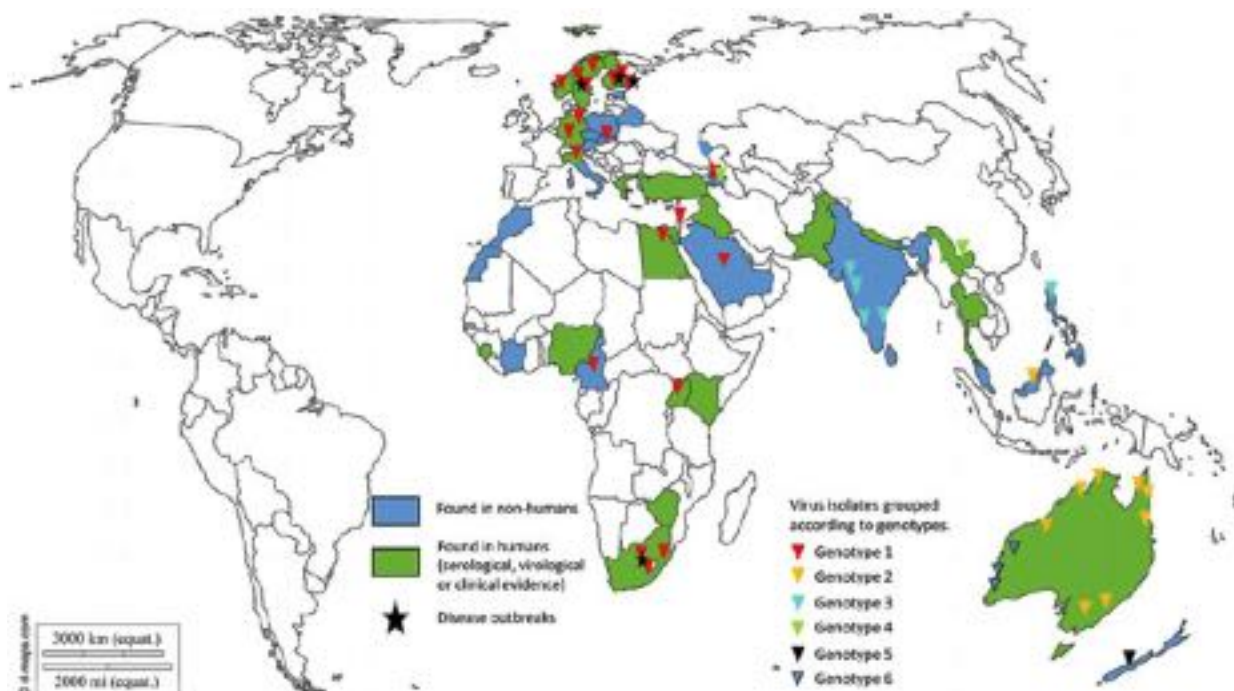


Fig. 1 Geographical distribution of SINV. Regions with serological, virological or clinical evidence of human infections are marked in green. Regions with signs of SINV in some specific animal species such as mosquitoes, birds, etc., are marked in blue. Black stars indicate regions with known SINV infection outbreaks. Colored triangles indicate virus isolates belonging to genotypes 1–6 shown in red, orange, cyan, green, black and blue, respectively. Base map from: <http://d-maps.com>. From Adouchief, S., Smura, T., Sane, J., Vapalahti, O., Kurkela, S., 2016. Sindbis virus as a human pathogen-epidemiology, clinical picture and pathogenesis. Reviews in Medical Virology 26, 221–241. The image is reproduced with the permission of the rights holder, John, Wiley & Sons.

Table 1 SINV proteins and their key functions

<i>Protein</i>	<i>Length (aa), size (kDa)</i>	<i>Functions</i>
nsP1	540 aa, ≈60 kDa	Methylation and capping of newly synthesized viral genomic and subgenomic RNA
nsP2	807 aa, ≈90 kDa	Protease, RNA helicase and RNA triphosphatase activity required for viral RNA replication and transcription Host translation shut-off
nsP3	556 aa, ≈60 kDa	Macrodomain with phosphatase activity and nucleic acid binding ability Interference of the innate antiviral mechanisms
nsP4	610 aa, ≈70 kDa	RNA-dependent RNA Polymerase
C	264 aa, ≈30 kDa	Formation of nucleocapsid with viral RNA
E3	64 aa, ≈7 kDa	Proper folding of pE2 pE2/E1 complex formation Transport of structural components to the site of budding Spike folding and spike activation for viral entry
E2	423 aa, ≈46 kDa	Receptor binding Receptor-mediated endocytosis Antigenic and neutralization epitopes
6K	55 aa, ≈6 kDa	Glycoprotein trafficking Virion assembly Virus particle formation Membrane permeability
E1	439 aa, ≈47 kDa	Fusion of the viral envelope with the host endosomal membrane during virus entry Antigenic and neutralization epitopes

capsid protein. The glycoprotein shell features 80 trimeric spikes consisting of three heterodimers of E1 and E2 transmembrane glycoproteins, projecting from the membrane. The ectodomain of E2 forms the outer portions of the spike, while the cytoplasmic tail interacts with the nucleocapsid. The ectodomain of E1 has a more tangential orientation within the viral membrane. E1 is responsible for membrane fusion, while E2 is responsible for the host cell receptor recognition ([Table 1](#)).

Genome

SINV has a linear single-stranded 11.7 kilobase RNA genome of positive polarity. The coding region consists of two open reading frames (ORF). The N-terminal ORF is translated directly from the 5' two-thirds of the genomic RNA, and encodes the nonstructural polyproteins P123 and P1234, from which the nonstructural proteins nsP1–4 are cleaved. The C-terminal ORF is expressed through a subgenomic mRNA, and encodes a structural polyprotein p130, from which the structural proteins C (capsid), E3, E2, 6K, and E1 are cleaved by cellular and virus-encoded proteases ([Fig. 2](#)).

Life Cycle

The host cell entry of SINV occurs by receptor-binding, followed by clathrin-mediated endocytosis. The low pH of the endosome triggers viral fusion, and nucleocapsid is released into the cytoplasm. A conformational rearrangement of E1 and E2 occurs, during which E2 dissociates from E1. Subsequently, E1 forms homotrimers and expose their fusion peptides, which drives the fusion of the viral membrane with the endosomal membrane. This results in the release of viral genome into the cytoplasm.

The nonstructural proteins are translated from the full-length genomic viral RNA to polyprotein P123, and specifically in SINV, by translational readthrough of an opal termination codon to polyprotein P1234 ([Fig. 2](#)). The polyproteins are processed by the protease located within nsP2 protein. nsP1 and nsP2 are involved in several enzymatic activities essential for RNA replication ([Table 1](#)). nsP4 functions as RNA-dependent RNA polymerase. The structural proteins are translated from a subgenomic RNA to a C-pE2-6K-E1 polyprotein, from which C protein is released by autoproteolysis. pE2 (precursor of E3 and E2) and E1 are translocated across the ER membrane and undergo post-translational modifications ([Fig. 2](#)). pE2 and E1 form heterodimers, which are processed and transported to plasma membrane. pE2 is cleaved by a furin-like protease to yield E3 and E2 proteins.

Viral RNA replication takes place in cytoplasmic vacuoles. The minus-strand synthesis predominates early in infection, and it is dependent on a replication complex composed of P123 and nsP4. The plus-strand synthesis requires a replication complex composed of individual nonstructural proteins, and it takes place later in infection. In the virus particle assembly, the nucleocapsid core is formed by the autoproteolytic process of the capsid protein. The trimers consisting of three heterodimers of E1 and E2 are transported to the plasma membrane, and interact with the nucleocapsid cores in the cytoplasm, forming progeny viruses. During the release from cells, virions acquire a membrane bilayer derived from the host cell plasma membrane.

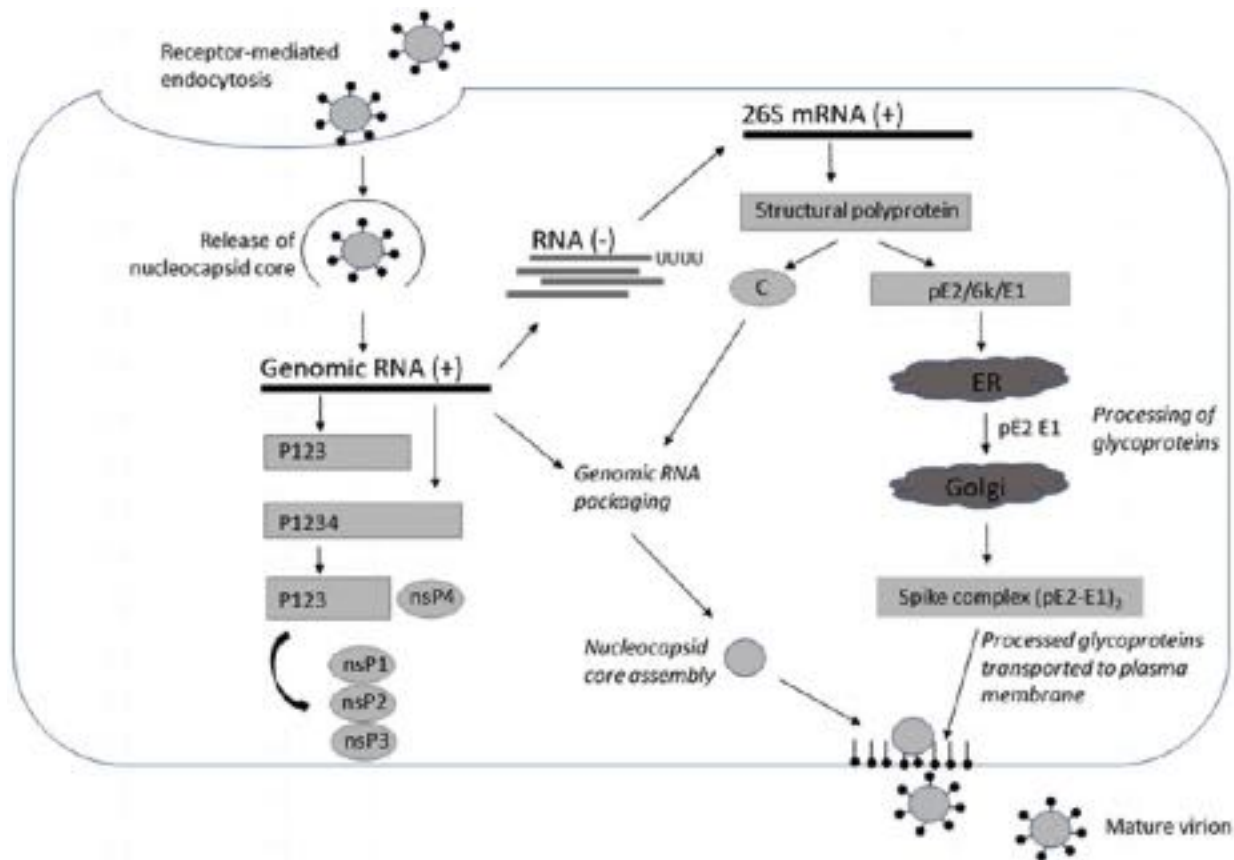


Fig. 2 Schematic representation of SINV life cycle, genome and proteins.

Epidemiology

SINV was first isolated in 1952 from a pool of *Culex pipiens* and *Cx. univittatus* mosquitoes in the village of Sindbis located in the Nile river delta near Cairo in Egypt. SINV was later isolated from humans in Uganda in 1961, however, without a reported link to a clinical illness. The virus was recovered for the first time from a clinically ill patient in South Africa in 1963 from skin lesions, and subsequently associated with a rash-arthritis syndrome. In 1960s, SINV disease outbreaks were reported in South Africa. At the same time, antibodies to SINV were first detected in Northern Europe, several Mediterranean countries, and the Volga delta region in humans and/or birds. In 1967, first clinical patients were noted in Sweden, while in 1974, a disease outbreak took place both in South Africa and in Northern Europe (Finland and Sweden). **Fig. 3** illustrates some of the key events in the history of SINV emergence.

Clinical SINV disease only occurs in geographically restricted areas with an unpredictable temporal appearance. Between 1974–2002, disease outbreaks in Finland took place in a seven year interval, during which the largest outbreak took place in 1995 with 1310 laboratory confirmed cases. Finland has the highest incidence of SINV disease (in epidemic years > 25/100,000/year), and approximately 5% of the Finnish population has been estimated to be seropositive for SINV. Data from Northern Europe indicates a higher SINV antibody prevalence in men than in women, while the incidence rates are higher in women than in men. Incidence rates in Northern Europe are highest in middle-aged adults.

Fig. 1 shows the geographical regions in which SINV has been detected in humans and some animal species, as well as areas of SINV outbreaks. SINV or antibodies to SINV have been detected in several parts of Europe, Africa, and Oceania, as well as sporadically in different parts of Asia. SINV is not present in the Americas. Constant reports of clinical disease appear primarily from Finland and Sweden, as well as from South Africa. In recent years, the presence of SINV has been reported in new geographical regions, such as Germany, Czech Republic, the United Kingdom, and Israel, where, however, clinical reports are rare or nonexistent. The restricted geographical occurrence of clinical SINV disease may be due to differences in clinical alert, virus ecology, as well as genetic factors involving both human populations and virus strains. In endemic areas, human infections coincide with activity of enzootic vector mosquitoes. In Northern Europe, enzootic vector mosquito species include at least *Cx. torrentium*, *Cx. morsitans* and *Cx. pipiens*. In South Africa, *Cx. univittatus* appears to play an important role as an enzootic vector for SINV. Bridge vectors to man have been suggested to include e.g., *Aedes cinereus* and *Ochlerotatus annupiles* species.

Birds have been identified as the natural hosts of SINV. Experimental SINV infection of different species in the Passeriformes, Galliformes and Anseriformes orders have shown that viremia reaches a level and duration which is sufficient to infect both the

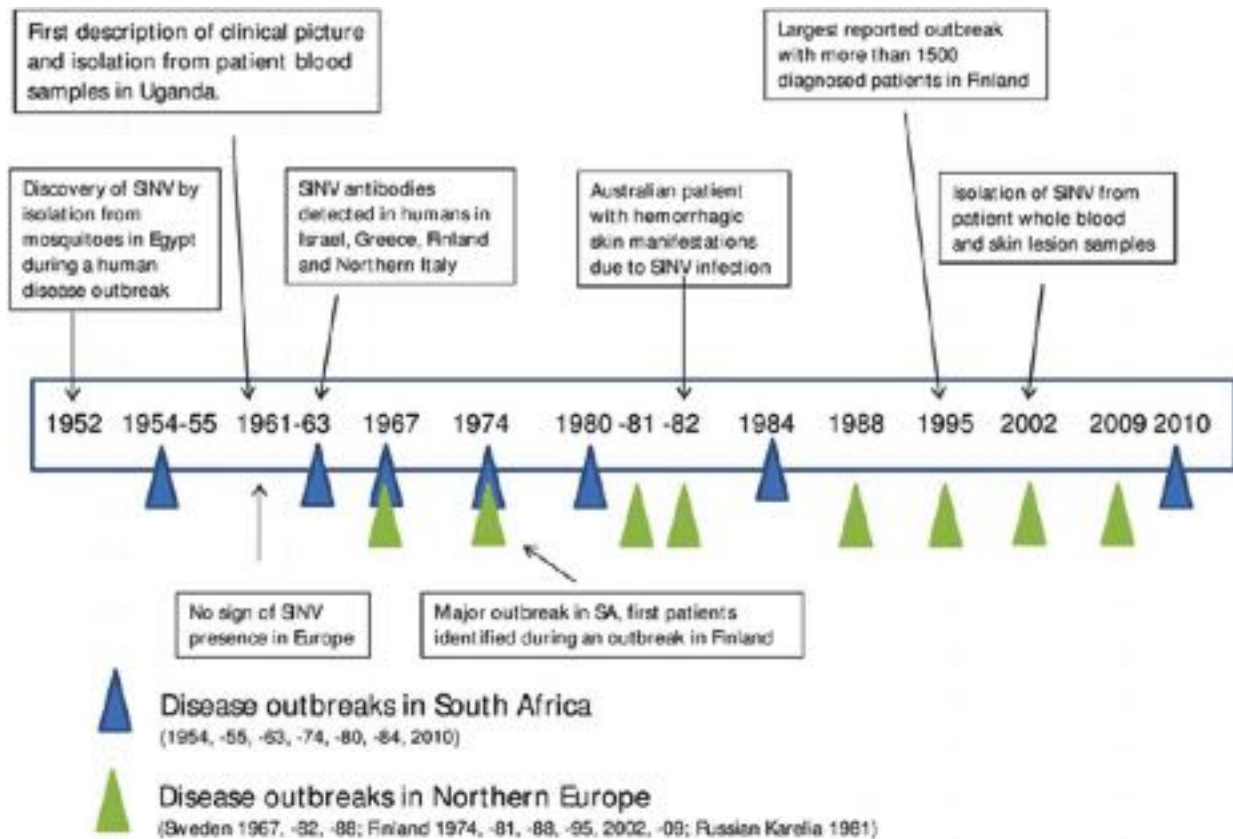


Fig. 3 Timeline of selected events in the history of Sindbis virus as a human pathogen. Arrows indicate disease outbreaks in South Africa (blue) and in Northern Europe (green). From Adouchief, S., Smura, T., Sane, J., Vapalahti, O., Kurkela, S., 2016. Sindbis virus as a human pathogen-epidemiology, clinical picture and pathogenesis. Reviews in Medical Virology 26, 221–241. The image is reproduced with the permission of the rights holder, John, Wiley & Sons.

enzootic vectors and the bridge vector mosquitoes. There is also a wide variety of bird species within these three orders in which SINV antibodies have been found in nature. In Sweden, the highest SINV seroprevalence has been detected in Fieldfare, Redwing, and Song thrush of the order Passeriformes, which have been suggested as the main amplifying hosts in that region. In Finland, also the role of Galliformes in SINV ecology has been proposed.

Migratory birds have been suggested to play a role in the spread of SINV to non-endemic areas. Field investigations have provided only limited information, but phylogeographic studies have provided further evidence on their role. SINV genotype 1 is the only genotype that has been associated with human outbreaks. Recent evidence suggests that genotype 1 has originated from central Africa. Genotype 1 was introduced to Sweden in 1920s as a single introduction, probably via migratory birds, and it subsequently dispersed to other parts of northern, eastern and central Europe in 1960s and 1970s. This data fits well with the epidemiological information available as to emergence of SINV disease and outbreaks.

Clinical Features

Human SINV disease is also known as Pogosta disease (Finland), Ockelbo disease (Sweden), and Karelian fever (Russia). In many respects, clinical SINV infection resembles that of CHIKV infection, but in a much milder form. SINV infection is frequently subclinical or even asymptomatic. Particularly in children, the disease is usually mild, and often presents without joint symptoms. There are no reports of fatal infections. In Northern Europe, 3%–6% of the clinically ill patients become hospitalized over the course of illness. Human SINV reinfections have not been reported, and it appears likely that a lifelong immunity follows the infection.

The incubation period of SINV infection is typically less than seven days. The hallmarks of an acute SINV infection are itchy, maculopapular exanthema (papules approximately 3 mm in diameter) over the trunk and limbs and mild fever, followed by muscle pain and joint symptoms, particularly in several large joints. Inflamed joints typically include ankles, fingers, wrists and knees, but often also hips, shoulders, and elbows. The joint pain, which often manifests as tenderness in movement and aching in rest can be fluctuating, but resides in the same articulations. Other manifestations include general malaise, headache, dizziness, nausea, and lymphadenopathy. Basic blood chemistry tests are typically within normal range. Extra-articular symptoms alleviate within a few weeks.

Most patients with a symptomatic SINV infection have a relatively mild and self-limiting course of illness. Yet, a considerable proportion will experience long term sequelae. Over that past few decades, several clinical investigations have been conducted in Northern Europe to better understand the frequency and severity of persistent joint manifestations after SINV infection. Prospective studies from Finland show that approximately 50% of patients have a varying degree of joint manifestations attributable to SINV infection 12 months post infection, and 25% for three years post infection. In a small proportion (<5%) of patients, the long term joint manifestation can be defined as arthritis in an examination by a professional rheumatologist. Female gender and higher age are risk factors for prolonged symptoms. Of note, pre-existing rheumatic and autoimmune diseases are more frequent among those patients who experience persistent manifestations.

The estimated public health burden from persistent joint manifestations after SINV infection is considerable in areas with high SINV seroprevalence. In Finland, for instance, there are specific geographical areas in which the seroprevalence of SINV can reach > 10%. In such areas, differential diagnosis of patients manifesting with unspecified joint symptoms should include consideration of long term sequelae of SINV infection.

Although SINV antibodies are found even in high frequency in various animals in many regions in the Old World, no clear evidence exists as to SINV pathogenicity in animals.

Pathogenesis

SINV has been shown to infect bone-associated connective tissue and to replicate in the skin in mouse models. SINV has also been isolated from human skin lesion biopsies. Pathological examination of patients' skin lesions reveals edema in the upper dermis with inactive T-lymphocytes and lymphoblasts, as well as macrophages present in the area. Primary human myoblasts and myotubes are susceptible to SINV infection in vitro, and signs of tissue regeneration following necrosis has been observed in a patient with chronic myalgia after SINV infection.

Major human antibody responses are targeted to the SINV structural proteins E1 and E2. Patients show a strong antibody response with particularly IgG3 subclass antibodies predominating. In the related CHIKV infection, IgG3 response protects patients from a chronic disease and mouse models suggest therapeutic effect from a passive transfer of patient anti-CHIKV antibodies. Elevated levels of type 1 interferons are seen in SINV mouse models, and also in patients infected with a related virus, namely CHIKV. Macrophages and macrophage induced cytokine response have been shown to play an important role in the immune response of SINV and other alphaviruses. Replication of SINV in human macrophages has been demonstrated, and SINV infection promotes macrophage activation, subsequently inducing the secretion of TNF-alpha, IL-1beta and IL6, as well as increased expression of matrix metalloproteinases 1 and 3, which may be associated with joint tissue damage.

Evidence suggests an association between alphavirus arthralgias and autoimmune diseases or their markers. Elevated antinuclear and anti-mitochondrial autoantibodies and rheumatoid factor have been observed in patients examined three years after SINV infection. Symptomatic SINV infection is also associated with HLA DRB1*01 allele, which is linked to e.g., rheumatoid arthritis.

Diagnosis

Even in endemic countries, the clinical diagnosis of SINV infection is not straightforward, and therefore a clinically suspected case should be confirmed in a laboratory. This is of importance not only to establish the correct diagnosis, but also to avoid unnecessary and ineffective therapies, and to provide information of the prognosis. Laboratory diagnostics are based on the detection of SINV IgM antibodies or seroconversion of IgG antibodies. IgM antibodies can be detected within 8 days and IgG antibodies within 11 days after the onset of disease. IgM antibodies commonly persist for several months post infection (> 30% after six months). IgG antibodies remain positive, likely for whole life as a sign of immunity. Immunoassays for SINV are only available in specialized laboratories, and they are based on in-house enzyme immunoassays and immunofluorescence techniques. Molecular diagnostic techniques are not readily applicable for laboratory diagnostics of human SINV disease due to the typically low level of viremia and a short viremic period in the acute phase. Virus isolation from blood or skin lesions has been successful only in rare occasions.

Treatment

Specific antiviral treatment is not available for SINV infection. Symptomatic treatment can be considered, including antihistamines for itching rash, and non-steroidal inflammatory drugs for joint pain. Intra-articular corticosteroids are sometimes used for persisting joint symptoms.

Prevention

Persons without immunity to SINV and who are exposed to bridge vector mosquitoes may become infected. Prevention is based on protective clothing and mosquito repellents in endemic areas with a known SINV circulation. Vaccines or prophylactic medication are not available.

Further Reading

- Adouchief, S., Smura, T., Sane, J., Vapalahti, O., Kurkela, S., 2016. Sindbis virus as a human pathogen-epidemiology, clinical picture and pathogenesis. *Reviews in Medical Virology* 26, 221–241.
- Hesson, J.C., Verner-Carlsson, J., Larsson, A., *et al.*, 2015. *Culex torrentium* mosquito role as major enzootic vector defined by rate of Sindbis virus infection, Sweden, 2009. *Emerging Infectious Diseases* 21, 875–878.
- Jalava, K., Sane, J., Ollgren, J., *et al.*, 2012. Climatic, ecological and socioeconomic factors as predictors of Sindbis virus infections in Finland. *Epidemiology and Infection* 141, 1857–1866.
- Kurkela, S., Manni, T., Myllynen, J., Vaheri, A., Vapalahti, O., 2005. Clinical and laboratory manifestations of Sindbis virus infection: prospective study, Finland, 2002–2003. *The Journal of Infectious Diseases* 191, 1820–1829.
- Ling, J., Smura, T., Lundström, J.O., *et al.*, 2019. Introduction and dispersal of Sindbis virus from Central Africa to Europe. *Journal of Virology* 93, e00620.
- Pietilä, M.K., Hellström, K., Ahola, T., 2017. Alphavirus polymerase and RNA replication. *Virus Research* 234, 44–57.
- Rupp, J.C., Sokoloski, K.J., Gebhart, N.N., Hardy, R.W., 2015. Alphavirus RNA synthesis and non-structural protein functions. *Journal of General Virology* 96, 2483–2500.
- Suhrbier, A., Jaffar-Bandjee, M.C., Gasque, P., 2012. Arthritogenic alphaviruses – An overview. *Nature Reviews Rheumatology* 8, 420–429.

Relevant Website

<https://www.ecdc.europa.eu/en/sindbis-fever/facts>
Facts about Sindbis fever.

Tick-Borne Encephalitis Virus (*Flaviviridae*)

Teemu Smura, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

Suvi Kuivanen, University of Helsinki, Helsinki, Finland

Olli Vapalahti, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

This is an update of T.S. Gritsun, E.A. Gould, Tick-Borne Encephalitis Viruses, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:10.1016/B978-0-12-801238-3.02668-4.

Classification

Tick-borne encephalitis virus (TBEV) is a flavivirus (genus *Flavivirus*, family *Flaviviridae*). Within the genus *Flavivirus*, TBEV is a member of a monophyletic tick-borne flavivirus group. The genus also contains three other ecological groups: the insect-specific flaviviruses (ISFV), the mosquito-borne flaviviruses (MBFV) and the flaviviruses with no known arthropod vectors (NKV). There are three well-established subtypes of TBEV: European (Eur), Siberian (Sib) and Far-Eastern (FE) subtypes that reflect phylogenetic and geographic relationships. TBEV-Eur is carried mainly by *Ixodes ricinus* ticks in central and north-eastern Europe, whereas TBEV-Sib and -FE are found mainly in *Ixodes persulcatus* ticks in an area reaching from north-eastern Europe to the Russian Far East, China and Japan. Louping ill virus and Spanish sheep encephalitis viruses form a sister clade to the European TBEV subtype. These are further related to Greek goat encephalitis virus and Turkish sheep encephalitis virus. Recently, divergent lineages of TBEV have been characterized from Eastern Siberia (Irkutsk and Transbaikali), Mongolia, Western Siberia (river Ob region) and Tibetan highlands. These have been proposed to form new subtypes of TBEV (Baikalian, Obskaya and Himalayan subtypes) (Fig. 1). Other tick-borne flaviviruses outside the TBEV clade described above include Omsk hemorrhagic fever, Langat virus, Alkhurma hemorrhagic fever virus and Powassan virus, the latter representing ecologically and clinically parallel pathogen to TBEV in North America.

Virion Structure

TBEV particles are approximately 50 nm in diameter. The virion consists of three structural proteins: capsid (C), precursor of membrane (prM) and envelope (E), as well as a positive-sense RNA genome and a lipid membrane derived from the host cell. (Fig. 2) Seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) are expressed in infected cells. The envelope is formed by E and M heterodimers that form tetramers in a head-to-tail manner giving the virion an icosahedral shape. (Fig. 2(D)) The capsid protein C encloses the genome that is approximately 11 kb long single-stranded positive-sense RNA. The genome has a single open reading frame of approx. 11,000 nucleotides flanked by 5' and 3' untranslated regions (UTRs).

TBEV Life Cycle

TBEV virions are internalized mainly via receptor-mediated endocytosis. In mammalian cells, laminin-binding protein (LBP) and the $\alpha V\beta 3$ integrin are major candidates for TBEV receptors, but less is known of the receptors in tick cells. The virions are stable in slightly basic conditions ($\sim$ pH 8), and the gradual acidification of the endosome causes the E homodimers to rearrange into homotrimers. Viral and endosomal membranes fuse and release the nucleocapsid into the cytoplasm. After nucleocapsid uncoating, the genome is translated into one long polyprotein of approximately 3400 amino acids at the ER membrane. The viral genome is replicated at membrane-associated replication complexes (RC) in the perinuclear region of the cell. The amino-terminal part of the polyprotein consists of three structural proteins followed by seven nonstructural proteins that are co- and post-translationally cleaved by several proteases, most importantly by the host signal peptidase and viral serine protease (NS2B/NS3). The individual proteins are released either on the luminal (prM, E, NS1 and NS4B) or the cytoplasmic side (NS2A, NS2B, NS3, NS4A and NS5) of the ER. (Fig. 3).

Viral Proteins

The capsid protein C (11 kDa) forms homodimers that enclose the viral genome into an electron dense core of ca. 30 nm. The precursor of membrane protein (prM) is cleaved by furin into 'pr' and membrane M (26 kDa) proteins prior to virion budding. The prM protein enables E protein to fold correctly into its native conformation. The envelope protein E (53 kDa) is a membrane protein that typically has 1 or 2 glycosylation sites, and functions in receptor binding and membrane fusion. E monomer consists of four domains (dI, dII, dIII and dIV). The putative receptor-binding sites reside in domain III, and the fusion peptide is located at the tip of domain II.

NS1 (46 kDa) is a multifunctional glycoprotein that exists as a dimer and a hexamer. It translocates into the ER by a signal sequence at the carboxy-terminus of E, and is cleaved by a host signal peptidase. Shortly after synthesis, the downstream cleavage

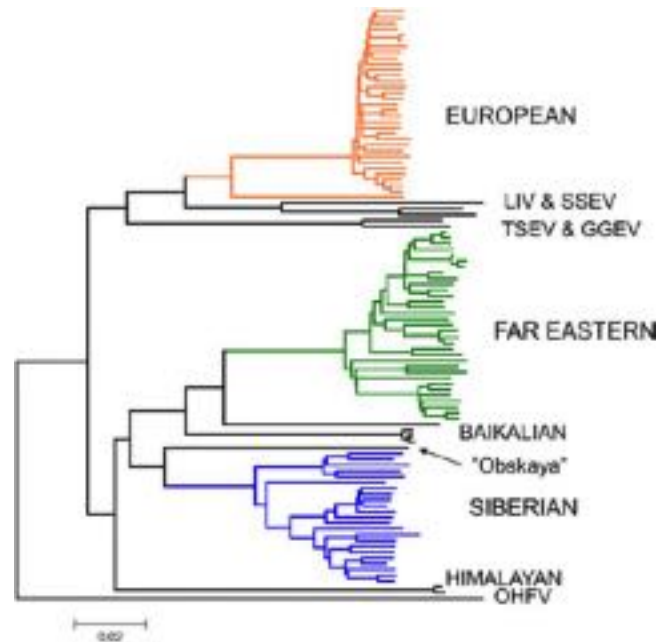


Fig. 1 A phylogenetic tree based on the complete coding regions of TBEV. Omsk hemorrhagic fever virus (OHFV) that is the closest relative to TBEV was included as an outgroup. The three well-established subtypes; European, Far Eastern and Siberian are shown in orange, green and blue colors, respectively. Of the proposed new subtypes, the two Baikalian subtypes form sister clades to the Far Eastern subtype while the proposed “Obskaya subtype” forms a sister clade to the Siberian subtype. Louping ill virus (LIV), Spanish sheep encephalitis virus (SSEV), Greek goat encephalitis virus (GGEV) and Turkish sheep encephalitis virus (TSEV) form sister clades to the European subtype of TBEV.

of NS1 from NS2A is catalyzed by an unknown ER-resident host protease. NS1 functions in replication, forming the replication complex (RC) together with NS4A and NS4B. Dimeric NS1 forms soluble hexamers in the Golgi network that are released outside of the infected cell, and possess immunomodulatory activities via interactions with the host complement system. NS1 shares 90% amino acid homology with different TBEV strains and as high as 44% homology with other flaviviruses. The extracellular form of NS1 elicits a strong immune response in the host in natural flavivirus infections.

NS2A is a small hydrophobic protein (22 kDa) that functions in virus replication and assembly. The downstream cleavage from NS2B occurs via the viral serine protease. NS2B is a membrane-associated protein (14 kDa) that functions as a key cofactor of the viral serine protease complex with NS3. NS2A and NS2B function in immunomodulation and together with NS3, NS2B has been shown to play a role in TBEV neuroinvasiveness. NS3 is a multifunctional cytoplasmic protein associated with membranes via its interaction with NS2B. NS3/NS2B catalyzes the autocleavage of NS2A/NS2B and NS2B/NS3 and the trans cleavage of NS3/NS4A and NS4B/NS5. NS3 possesses RNA helicase, nucleotide 5'-triphosphatase (NTPase) and RNA 5'-triphosphatase (RTPase) activities, and it associates with viral protein NS5 in replication. NS4A and NS4B are small integral membrane proteins (16 kDa and 27 kDa, respectively) connected via a conserved signal peptide 2k that serves as a signal sequence for NS4B. NS4A and NS4B function in RC formation and immunomodulation. NS5 (104 kDa) contains two distinct domains: the N-terminal S-adenosyl-methionine (SAM) transferase domain associated with the 5'-cap formation of newly synthesized viral RNA, and the C-terminal RNA-dependent RNA polymerase domain (RdRp). The interdomain region of NS5 is involved in RC formation. The genome and viral proteins are schematically described in Fig. 3. Immature virus particles are assembled in the ER from the structural proteins and the genome. The lipid envelope is gained by budding into the ER. pH and protease-dependent maturation of the virion occurs in the trans-Golgi network (TGN), where prM and E undergo major structural reorganization. Mature particles are formed once pr is cleaved from M. Three types of virions are released from infected cells by exocytosis: mature, partly mature and immature (Fig. 2(A)). Flavivirus-infected cells also release particles containing only the envelope proteins E and M and the lipid envelope. These virus-like particles (VLPs) are noninfectious and approximately 30 nm in diameter. The maturation and assembly of VLPs seem very similar to those of the virus: VLPs undergo low-pH induced rearrangement of envelope proteins and are able to undergo membrane fusion.

Epidemiology

TBEV is maintained in nature by ticks and their rodent hosts, typically in geographically restricted “TBEV foci” (Fig. 4). Although the virus has been detected from at least 22 tick species, *Ixodes ricinus* and *I. persulcatus* are considered to play the main role in virus transmission. Humans do not contribute to the circulation of TBEV, but can be infected as accidental (dead end) hosts. Ticks serve both as a reservoir and a vector for the virus. Ticks can obtain TBEV either by feeding on a viremic animal or by non-viremic

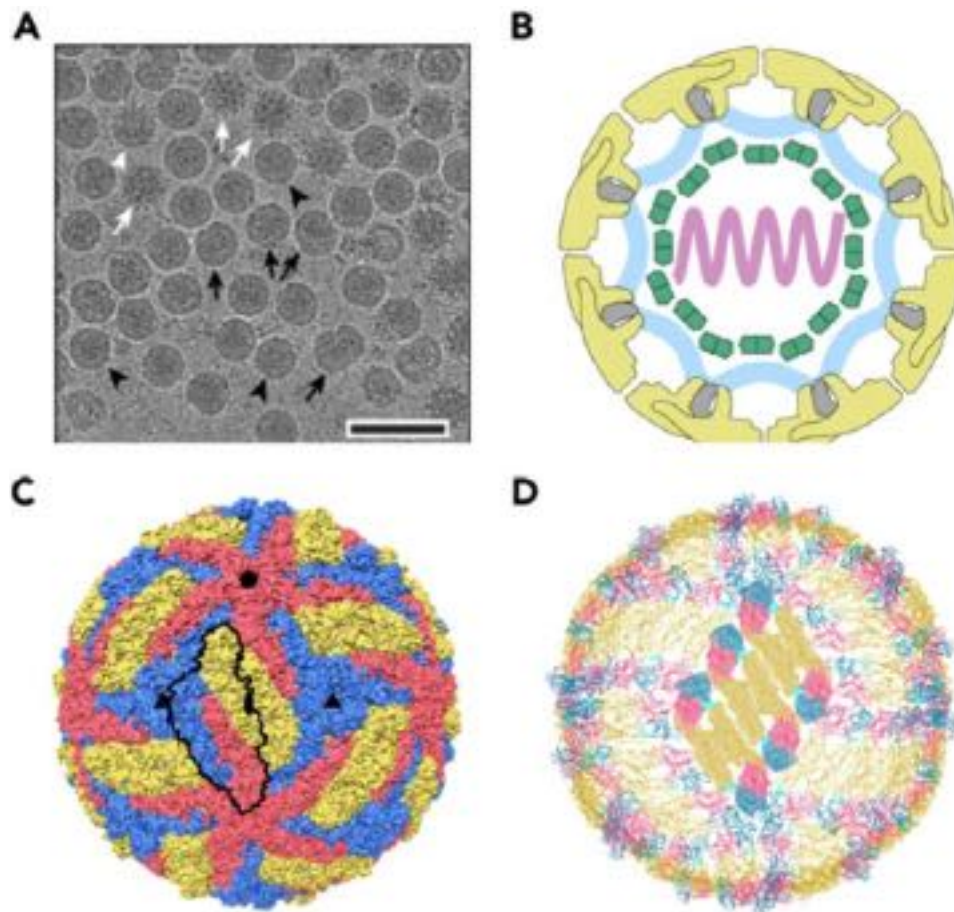


Fig. 2 Structure of the TBEV virion. (A) Electron cryo-micrograph of smooth mature particles (black arrowheads), immature (white arrows), partially mature (white arrowhead), and damaged (black arrows) TBEV particles. The scale bar is 100 nm. The image is courtesy of Dr. T. Füzik *et al.* Nat Commun. 2018 Jan 30;9(1):436. Structure of tick-borne encephalitis virus and its neutralization by a monoclonal antibody and is reproduced under a Creative Commons Attribution 4.0 International License. (B) Schematic representation of the TBEV virion. Multiple copies of the C protein (green) encapsulate the viral genome (lilac). The nucleocapsid is surrounded by a lipid membrane (light blue), in which E and M proteins (yellow and gray, respectively) are embedded; (C) Surface representation of the TBEV virion (wwPDB: 506A). An icosahedral asymmetric unit is outlined in black. The three E proteins within each asymmetric unit are shown in blue, red, and yellow. Symmetry axes are indicated by the black pentagon (five-fold), the triangles (three-fold), and the ellipse (two-fold); (D) Three E-M-M-E heterotetramers on the TBEV surface. Three domains of E are highlighted in red (I), yellow (II), and blue (III), and the fusion loop is highlighted in turquoise. E protein domain IV and M protein are not visible on the virion surface. Figure reproduced from Pulkkinen, L., Butcher, S.J., Anastasina, M., 2018. Tick-borne encephalitis virus: a structural view. *Viruses* 10 (7), 350. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

transmission between co-feeding ticks. During host infestation, Ixodes ticks tend to aggregate to certain preferred feeding sites. During such co-feeding events, the virus can transmit from one tick to another without a need for host viremia. Thereby, also animals with antibodies against TBEV can support TBEV circulation. Co-feeding transmission, especially between tick larvae and nymphs, is considered to amplify the spread of TBEV in the tick population. Among ticks, TBEV may also be transmitted transovarially from infected adult females to their offspring and transstadially, i.e., the tick stays infected during molting and can therefore carry the virus through all of its life span. These two mechanisms are likely to support the persistence of TBEV in the tick population. The prevalence of TBEV among the ticks of a given TBEV focus is usually low; typically only a few percent or less of ticks in a given TBEV focus are TBEV RNA positive (Fig. 4).

Since the efficient co-feeding transmission requires overlapping activity periods for tick larvae and nymphs, microclimatic conditions favoring such activity patterns are considered to play a major role in the susceptibility of a tick population for TBEV circulation. The immature life stages (larvae and nymphs) of Ixodes ticks feed mainly on small mammals, whereas the adult ticks mainly feed on larger mammals such as cervids. Considering the TBEV circulation, the most important mammalian species are considered to be the abundant rodent species of *Apodemus* and *Myodes* genera. Due to its dependence on the population dynamics of ticks and various mammalian species, as well as climatic factors suitable for co-feeding transmission, TBEV is typically found focally within the distribution area of the host ticks.

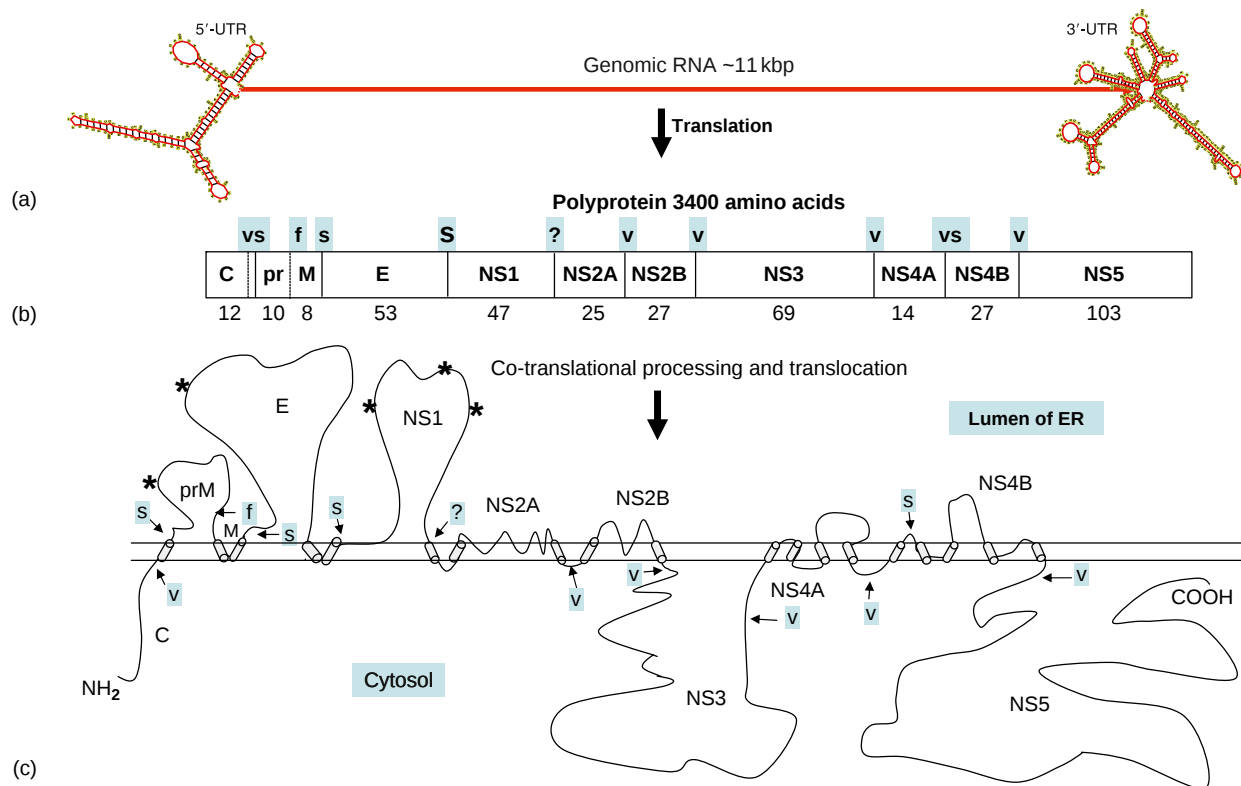


Fig. 3 Genome strategy of TBEV. (a) Genomic RNA is presented as a solid line on the top; the 5'- and 3'-UTR are depicted in the predicted conformations according to Gritsun, T.S., Venugopal, K., Zanotto, P.M., *et al.*, 1997. Complete sequence of two TBEV isolated from Siberia and the UK: Analysis and significance of the 5'- and 3'-UTRs. *Virus Research* 49, 27–39. (b) Translation and co-translational processing of flavivirus polyprotein. The flavivirus polyprotein is depicted as a bar, with specified individual proteins and their molecular masses (numbers below the bar). (c) Membrane topology of viral proteins in relation to the lumen of the ER and cytoplasm after the completion of co-translational processing and translocation. Adapted from Westaway, E.G., Mackenzie, J.M., Khromykh, A.A., 2003. Kunjin RNA replication and applications of Kunjin replicons. *Advances in Virus Research* 59, 99–140. Transmembrane domains are shown as cylinders. Glycosylation is indicated as (*). The polyprotein is processed by ER signalases (s) and viral-specific protease NS2B-NS3 (v). The cleavage of M from prM is carried out by the Golgi protease, furin (f) and cleavage between NS1 and NS2A is by an unknown (labeled as ?) ER protease.

It has been estimated that over 10,000 TBEV cases occur every year in northern Eurasia. Ixodes ticks are active in temperatures above 5 degrees, and in Central Europe, there is typically a two-peak distribution of cases in early and late summer following the activity pattern of ticks. Human behavior (e.g., outdoor recreational activities), socio-economic factors, awareness of the disease and vaccine coverage are considered to affect the incidence of the disease significantly. In Europe, the disease incidence is highest in Central and Eastern Europe. Recently, previously unknown TBEV foci have been detected in Western Europe (e.g., The Netherlands and England) and Northern Europe (e.g., Finland, Sweden, Norway and Denmark), suggesting that TBEV endemic areas may be spreading toward these directions. During the years 2000–2016, the number of TBE cases reported in Europe (excluding Russia) fluctuated between 2000 and 3500 cases per year. Over one third of these cases were reported in Czech Republic and Lithuania. The incidence of the disease is highest in Baltic countries; Estonia, Latvia and Lithuania. The majority of the notified cases occur among individuals of age 40–69, with increasing incidence towards the older (60–69) age group. The disease is more prevalent in males than females in all age groups. In Russia, more than 27,000 TBE cases were registered between the years 2007–2016, with highest incidences in the Western Siberian part of the country: Altai, Krasnoyarsk, Tomsk, Khakassia and Tuva. Compared to the 1990s, the incidence of TBE has decreased in Russia during the last two decades. The disease occurs also in China, mainly in the north-eastern regions including Inner Mongolia Autonomous Region, Heilongjiang province and Jilin province. While the virus and antibodies against it have been detected in ticks and reservoir animals in Korean peninsula and Japan, only a few cases have been reported from Japan and none from the Republic of Korea.

Clinical Features

TBEV is transmitted from the saliva of an infected tick within minutes after the tick bite. In addition, TBEV infection may occur via the gastrointestinal tract by unpasteurized milk from an infected animal. The route of infection as well as the infecting TBEV subtype may affect the clinical picture of the disease. European subtype TBEV infection with involvement of CNS is typically a biphasic illness. After an incubation period of on average 8 days (range 4–28 days), the first phase may present with general malaise, fatigue, headache, myalgia,

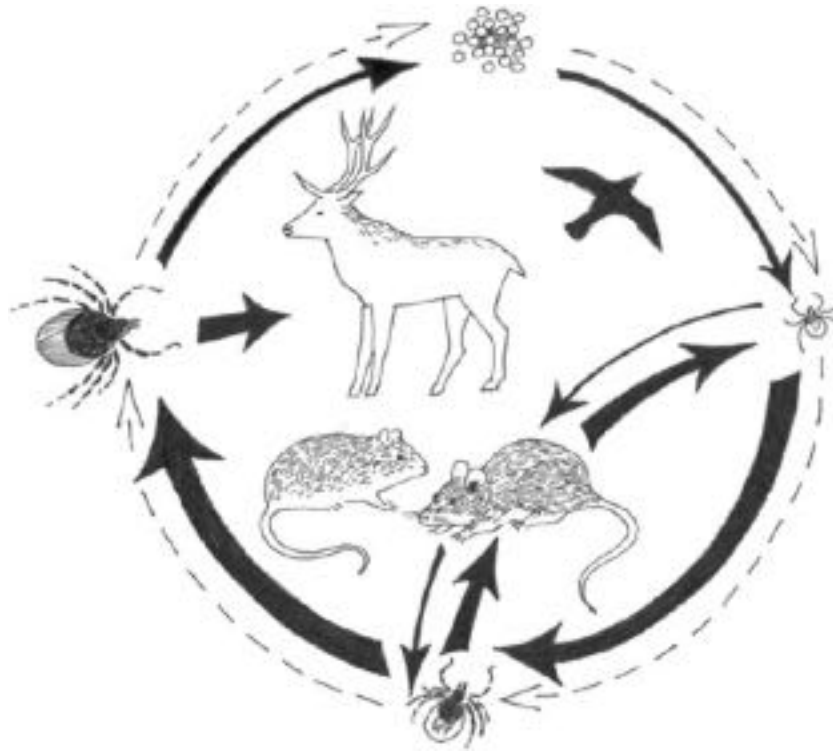


Fig. 4 TBEV ecological cycle. Dashed arrows depict the different stages of the *Ixodes ricinus* or *I. persulcatus* tick (top: eggs, right: larva, below: nymph, left: adult). Black arrows depict the presence and transtadial transmission of TBEV. The co-feeding transmission from infected nymphs to uninfected larvae, both of which typically feed on rodents, is supported by suitable environmental conditions, and enhances the circulation of TBEV. The virus may also transmit transovarially. Larger mammals such as cervids support larger tick populations and birds may transmit TBEV to establish geographically remote foci. Vertical transmission may also occur among rodents. (Picture by Marjukka Törmi).

anorexia, nausea, vomiting and low grade fever. This phase usually lasts for 2–4 days. For the majority of the patients, the disease is resolved after the first phase. However, after a lag period of approximately a week (1–20 days), about one third of the patients enter a second phase of the disease that manifests with high fever and headache. The severity and occurrence of neurological symptoms and manifestations vary and may include nuchal rigidity, sensitivity to light, diplopia, dizziness, cerebellar ataxia, disorientation, hallucinations, epileptic seizures, dysphasia and pareses, and in extreme cases, tetraparesis. The severity of the disease increases with age; the CNS manifestations are meningeal in the majority of those under 40, and above this age meningoencephalitis and myelitis myeloradiculitis occur more often.

Recovery from more severe forms of the disease may take months. 13%–36% of patients can have neuropsychiatric sequelae such as pareses, hearing and balance disturbances, headache, difficulties to concentrate, and depression. The severity and case fatality rate of the disease is considered to vary between the different TBEV subtypes. The case fatality rate in TBEV-Eur infections is <2%, whereas for TBEV-Sib infections the case fatality rate is 2%–8%. For the Far Eastern TBEV subtype, the case fatality rate has been reported to be up to 40% in some years, but it is unclear how many milder TBEV-FE cases may go unnoticed. The disease caused by the Siberian subtype of TBEV is often monophasic, may proceed more rapidly and may present with a chronic form. In fatal Siberian subtype infections, death usually occurs 3–7 days after onset of the disease. For the Far Eastern subtype, the typical first-phase symptoms are followed by a stiff neck, sensorial changes, visual disturbances and variable neurological dysfunctions. In fatal cases, death occurs within the first week after onset of illness.

Pathogenesis

Langerhans dendritic cells are the primary targets of vector-mediated flavivirus infections. Tick saliva facilitates the infection by modulating the innate immune responses of the first infected cells. Infected dendritic cells then migrate to draining lymph nodes, from which the infection spreads to other organs such as the spleen, liver and the kidneys. The hallmark of TBEV pathogenesis is the entry of the virus into the CNS and brain, characterized by breakdown of the blood-brain-barrier (BBB) and neuroinflammation. The BBB is a semi-permeable physical barrier between the blood circulation and the brain, consisting of epithelial cells, glial cells, neurons and pericytes. The inflammatory responses to TBEV may lead to increased permeability of the BBB, and the virus has been suggested to cross the BBB via either direct infection of the BBB cells, trans- and paracellular transport, or within migrating leukocytes. In addition, retrograde axonal transport may serve as a potential route for TBEV to enter the CNS.

Diagnosis

As the clinical picture of TBE is often similar to other viral CNS infections (caused by e.g., enteroviruses or herpes simplex virus 1) a laboratory confirmation is required to diagnose the disease. Anamnestic information of tick bites or potential exposure to ticks or unpasteurized milk during the tick season in TBEV endemic areas is the basis for suspicion of TBE. Notably, tick bites may often be unnoticed by the patient. It is also essential to know the TBEV vaccination history and possible exposure to other flaviviruses or flavivirus vaccines, particularly as a TBEV vaccinee may nevertheless have an acute TBE and the vaccination or other flavivirus infections affect the serological results and their interpretation.

The diagnostics is primarily based on detection of IgM and IgG antibodies from serum during the 2nd phase of TBE. Additional confirmation of CNS infection can be acquired through detection of these antibodies in cerebrospinal fluid. Typically, the antibodies are measured by EIA or IFA and the strongest reactivity can be detected against the viral E protein. Also reliable TBEV IgM immunochromatographic tests are available. Usually, viral RNA can no longer be detected during the 2nd phase of the disease, although it has been occasionally detected in early blood samples. On the other hand, during the first phase the patient rarely seeks medical care, and therefore, serum samples from the early stage of the infection are rarely available. However, from such samples, TBEV can be readily isolated, or detected by RT-PCR while antibodies are not yet present. In fatal cases, viral RNA is found and the virus can often be isolated from autopsy brain samples.

Although the diagnostics is in principle straightforward, since a single serum sample IgM and IgG test can usually provide the diagnosis, there are some caveats: (1) in TBEV Siberian subtype infections, a first sample from a patient with CNS symptoms may be antibody negative. This is possibly due to the faster invasion of the CNS by the Siberian TBEV subtype, without the first, systemic phase. In addition, the antibody positivity in CSF has preceded that of the serum in some cases. Therefore, in the TBEV endemic areas, where Siberian and/or Far Eastern subtypes are circulating, it is particularly important to have a second sample 1–2 weeks after the onset of illness. (2) The positive TBEV antibody reactivity can, in principle, be due to cross-reactive antibodies raised by infection with other flaviviruses. This is especially relevant in the areas where e.g., both West Nile virus (WNV) and TBEV circulate. In such circumstances a comparison of serum IgM levels or the neutralizing antibody titers for both viruses may be needed (3) In case of vaccine failure infections, IgM is seldom found in the first serum sample, while IgG levels may already be very high. However, in a second sample anti-TBEV IgM antibodies are usually positive. In all of these cases, reviewing anamnestic and clinical information, studying the CSF, neutralization tests, or testing for TBEV NS1 antibodies, as well as having consecutive samples are helpful.

Treatment

The treatment is supportive as currently there is no specific antiviral therapy for TBE. In Russia and Kazakhstan, post-exposure treatment with hyperimmune serum is provided, but this is no longer used elsewhere. Long term rehabilitation may be needed.

Prevention

TBE can be prevented by vaccination. Commercially available vaccines contain formalin-inactivated TBEV and are widely used across the TBE-endemic areas of Eurasia. Three doses are typically required for effective protection. For the vaccines used in Europe, the second dose is given after an interval of 1–3 months, and the third dose after an interval of 5–12 months, although accelerated protocols are also in use. The first booster vaccination is recommended 3 years later, and further boosters are recommended with e.g., 3–10 year intervals, depending on the age of the vaccinee. At population level, the vaccinations have shown to be effective: The highest vaccine coverage is in Austria, where nearly 90% of the population is vaccinated, and the case numbers have diminished despite increasing in neighboring countries and 98% of the remaining cases occur in the unvaccinated population. Despite full vaccination series, individual vaccine failure infections may occur. WHO recommends population-wide vaccination, if the pre-vaccination incidence is more than 5/100,000. It should be noted, however, that the risk to acquire TBEV infection may vary considerably even in a small geographical scale, due to the focal occurrence of the virus. In endemic areas, measures to minimize tick bites such as protective clothing and repellents as well as avoiding the consumption of unpasteurized milk products are important. Tick removal is an uncertain preventive measure, since the virus is transmitted soon after a tick bite. Updated surveillance of disease occurrence through diagnosed cases and One Health surveys, and increasing people's knowledge of TBE and vaccination are of key importance.

Further Reading

- Füzik, T., Formanová, P., Růžek, D., *et al.*, 2018. Structure of tick-borne encephalitis virus and its neutralization by a monoclonal antibody. *Nature Communications* 9 (1), 436. doi:10.1038/s41467-018-02882-0.
- Lindquist, L., Vapalahti, O., 2008. Tick-borne encephalitis. *Lancet* 371 (9627), 1861–1871.
- Pulkkinen, L., Butcher, S.J., Anastasina, M., 2018. Tick-borne encephalitis virus: A structural view. *Viruses* 10 (7), 350.

Ruzek, D., Avšič Županc, T., Borde, J., *et al.*, 2019. Tick-borne encephalitis in Europe and Russia: Review of pathogenesis, clinical features, therapy, and vaccines. *Antiviral Research* 164, 23–51. doi:10.1016/j.antiviral.2019.01.014.
 World health Organization, 2011. Vaccines Against Tick-Borne Encephalitis: WHO Position Paper. vol. 86. World health Organization. pp. 241–256.

Relevant Website

<https://www.ecdc.europa.eu/en/tick-borne-encephalitis>
 Tick-borne encephalitis.
 ECDC - Europa EU.

Transmissible Gastroenteritis Virus of Pigs and Porcine Epidemic Diarrhea Virus (*Coronaviridae*)

Qiang Liu and Volker Gerdts, University of Saskatchewan, Saskatoon, SK, Canada

© 2021 Elsevier Ltd. All rights reserved.

Classification

Transmissible gastroenteritis virus (TGEV) of pigs and porcine epidemic diarrhea virus (PEDV) are two porcine coronaviruses in the genus *Alphacoronavirus* of the *Coronavirinae* subfamily in the *Coronaviridae* family within the order *Nidovirales*. Based on sequence heterogeneity, PED viruses are tentatively categorized as “classical” and “emerging” strains. The classical strains include PED viruses identified between 1970s to 2010, whereas PED viruses isolated after 2010 are referred to as emerging strains. The emerging strains are further divided into “non-S INDEL (insertions and deletions)” and “S INDEL” strains on the basis of the spike (S) protein sequences and virulence in piglets. Another proposal classifies PED viruses into up to five genotypes.

Virion Structure

The enveloped virions of coronaviruses are spherical and/or pleomorphic with diameters of 144.8 ± 7.2 nm for TGEV (PUR46-MAD strain) and 95–190 nm for PEDV (CV777 strain). Viral envelope contains the spike (S), membrane (M), and envelope (E) proteins. Homotrimeric S protein complexes form the distinctive “corona-like” structure on the surface of the virions. Within the envelope, there exists a nucleocapsid consisting of the nucleocapsid (N) protein and viral genomic RNA.

Genome

Both TGEV and PEDV have a typical genomic organization of coronaviruses. The positive-sense, single-stranded RNA genome is approximately 28 kb in length with a 5′ cap structure and a 3′ polyadenylated tail. The coding sequence is flanked by untranslated regions (UTRs) at 5′ and 3′ ends. The N-terminal two-thirds of the genome contain one major open reading frame ORF1a encoding replicase polyprotein pp1a. A -1 frameshift just 5′ to the stop codon of ORF1a gives rise to ORF1b encoding a much longer replicase polyprotein pp1ab. These polyproteins are cleaved into 16 nonstructural proteins (nsps) mostly involved in viral RNA replication. The rest 10-kb genome codes for structural proteins S, M, E, and N, as well as accessory proteins with various functions (Fig. 1). TGEV encodes three accessory proteins, whereas PEDV encodes one.

Life Cycle

The life cycles of TGEV and PEDV consist of virion attachment and entry, viral RNA translation, viral RNA replication and transcription, virion assembly and release (Fig. 2). Virion attachment to a host cell requires the interaction between viral S protein with cellular receptors. Aminopeptidase N (APN) is a major receptor for TGEV. However, whether APN is an essential receptor for PEDV is debatable. PEDV S protein can bind to sialic acid on the cell surface that may mediate virion attachment. After receptor binding, the S protein undergoes proteolytic cleavage that in turn induces membrane fusion and virion entry. Viral RNA with a 5′ cap structure and a 3′ polyadenylated tail serves as an mRNA that is translated by cellular translation machinery to generate viral replicase proteins pp1a and pp1ab. Replication and transcription complexes formed by these proteins produce both genomic and subgenomic progeny RNA species. Structural and accessory proteins are translated from subgenomic RNAs. Nucleocapsids formed by the N protein and progeny viral genomic RNA are enveloped in the endoplasmic reticulum – Golgi intermediate compartment with the involvement of S, E, and M proteins. The assembled virions are transported to the cell surface and released.

Epidemiology

TGEV, first described in 1946 in USA, has been detected all over the world. PEDV was first isolated in 1978 in Belgium and is widespread in Europe and Asia ever since. The appearance of PEDV in North America was reported in 2013 where the virus continues to circulate in swine herds and there is a potential for new PEDV strains to emerge.

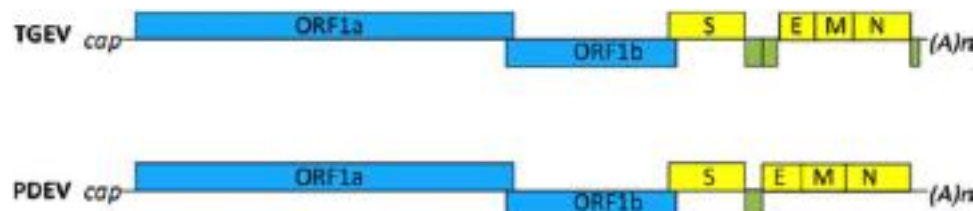


Fig. 1 Genome organization of TGEV and PEDV. Genes encoding structural proteins are presented in yellow. Putative accessory genes are shown in green. Nonstructural proteins encoded by ORF1a/b are presented in blue. Abbreviations: TGEV, transmissible gastroenteritis coronavirus; PEDV, porcine epidemic diarrhea virus; S, spike; E, envelope; M, membrane; N, nucleocapsid. Genomes have 5' cap and 3' poly A tail. Reproduced from Gerdts, V., Zakhartchouk, A., 2017. Vaccines for porcine epidemic diarrhea virus and other swine coronaviruses. *Veterinary Microbiology* 206, 45–51, with permission.

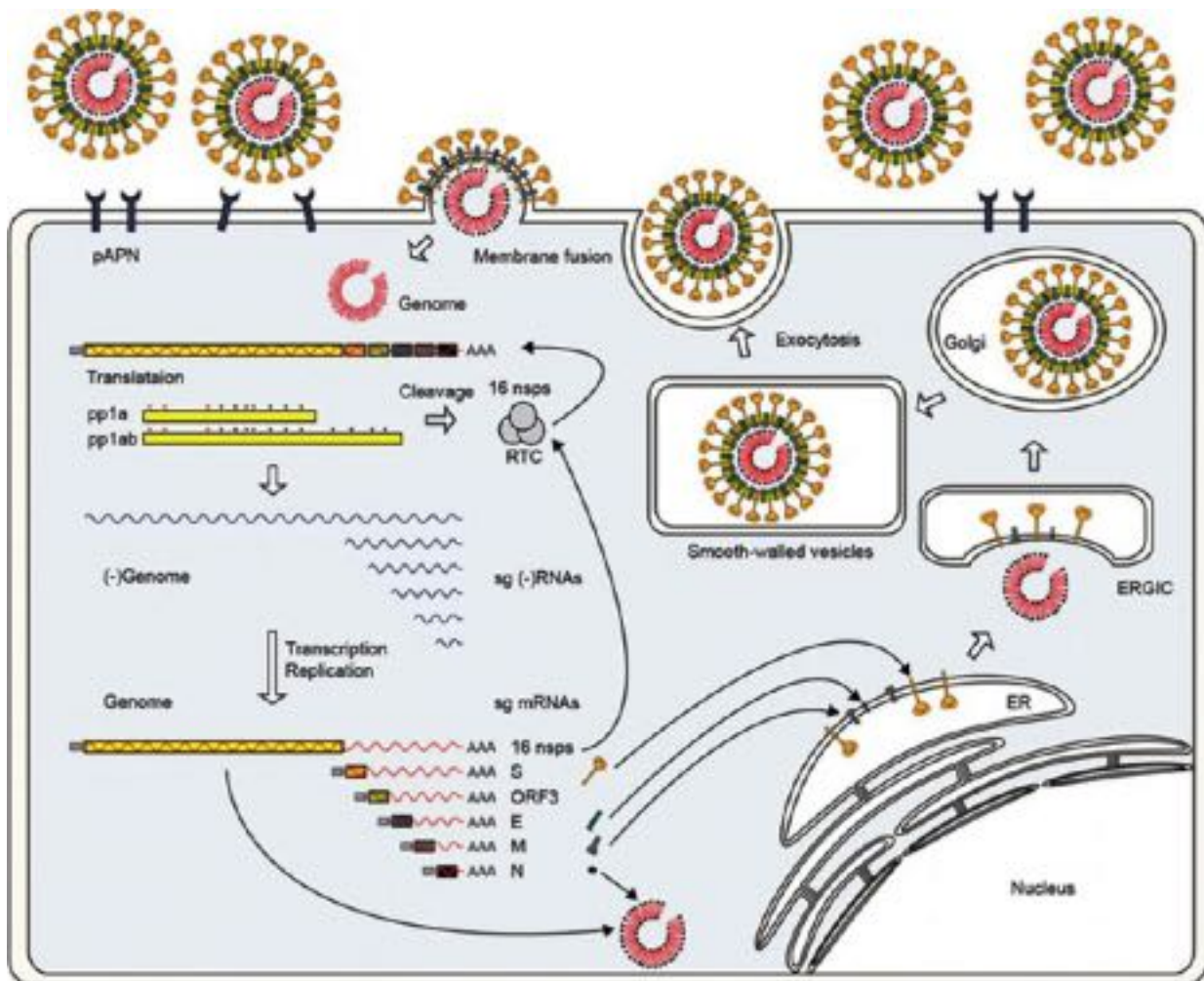


Fig. 2 PEDV replication cycle. PEDV binds a cellular receptor such as pAPN via the spike (S) protein. Penetration and uncoating occur after the S protein-mediated fusion of the viral envelope with the plasma membrane. Following disassembly, the viral genome is released into the cytoplasm and immediately translated to yield replicases pp1a and pp1ab. These polyproteins are proteolytically cleaved into 16 nsps comprising the replication and transcription complex (RTC) that first engages in the minus-strand RNA synthesis using genomic RNA. Both full- and sub genomic (sg)-length minus strands are produced and used to synthesize full-length genomic RNA and sg mRNAs. Each sg mRNA is translated to yield only the protein encoded by the 5'-most ORF of the sg mRNA. The envelope S, E, and M proteins are inserted in the ER and anchored in the Golgi apparatus. The N protein interacts with newly synthesized genomic RNA to form helical RNP complexes. The progeny virus is assembled by budding of the preformed RNP at the ER-Golgi intermediate compartment (ERGIC) and then released by the exocytosis-like fusion of smooth-walled, virion-containing vesicles with the plasma membrane. Reproduced from Lee, C., 2015. Porcine epidemic diarrhea virus: An emerging and re-emerging epizootic swine virus. *Virology Journal*, 12, 193, under BioMed Central license agreement.

Table 1 Vaccines for TGEV and PEDV

<i>Virus</i>	<i>Region/country</i>	<i>Vaccines in development</i>	<i>Commercial vaccines</i>
TGEV	North America	Recombinant proteins expressed in baculovirus, yeast, and plants; live attenuated vaccine; DNA vaccine	Live attenuated vaccines (mono, bi-, and trivalent for TGEV, rotavirus, and <i>E. coli</i>)
	Europe	Recombinant proteins expressed in baculovirus, yeast, and plants; live attenuated vaccine; DNA vaccine	Live attenuated vaccines (mono, bi-, and trivalent for TGEV, rotavirus, and <i>E. coli</i>)
	Asia	Recombinant proteins expressed in baculovirus, yeast, and plants; live attenuated vaccine	Inactivated vaccines (mono, bi-, and trivalent for TGEV, rotavirus, PEDV and/or <i>E. coli</i>); live attenuated trivalent for TGEV, PEDV, and porcine (China)
PEDV	North America	Recombinant proteins expressed in yeast and baculovirus; DNA vaccine; infectious clone for live attenuated vaccine;	Inactivated vaccine; recombinant alphavirus-based vaccine
	Europe	DNA vaccine	Inactivated vaccine
	Asia	Recombinant vaccines expressed in baculovirus, yeast, plants, <i>Lactobacillus casei</i> , <i>Salmonella typhimurium</i> and others	Inactivated bivalent TGEV and PEDV vaccine (China, PEDV strain CV777); live attenuated trivalent TGEV, PEDV, and porcine rotavirus (China, PEDV strain CV777); live attenuated vaccines (Japan, PEDV strain 83P-5; South Korea, PEDV strains SM98-1 and DR-13; Philippines, PEDV strain DR-13); inactivated vaccine (South Korea, PEDV strain SM98-1)

Note: Gerdts, V., Zakhartchouk, A., 2017. Vaccines for porcine epidemic diarrhea virus and other swine coronaviruses. *Veterinary Microbiology* 206, 45–51, with permission.

Clinical Features

Both TGEV and PEDV cause enteritis in pigs with very similar clinical symptoms. Major clinical signs include vomiting, watery diarrhea, dehydration, and weight loss. The mortality rate can reach greater than 90% and is inversely related to the age of pigs.

Pathogenesis

Villous enterocytes in small and large intestine are the major target cells of TGEV and PEDV infections. Viral infection causes cell death that results in villous atrophy followed by malabsorption, diarrhea, dehydration, anorexia, and eventually animal death. The molecular mechanisms for pathogenesis are not very well understood. There is evidence to suggest that viral proteins, such as spike and ORF3 proteins may affect viral virulence. At the molecular level, TGEV and PEDV have been shown to modulate multiple cellular processes/pathways including endoplasmic reticulum stress, cell cycle, and mitogen-activated protein kinase signaling.

Diagnosis

Because clinical signs cannot distinguish between TGEV and PEDV infections, additional assays are required for diagnosis. Common diagnostic assays include viral antigen detection by histoimmunocytochemistry and enzyme-linked immunosorbent assay (ELISA), viral RNA detection and genotyping by PCR and sequencing, virus isolation by cell culture, as well as antibody detection by serology.

Treatment

There is no specific treatment for TGEV and PEDV infections.

Prevention

Enhanced biosecurity procedures are a general means to contain the spread and prevent the entrance of viral infection in pig farms, but vaccination represents the most effective way in preventing TGEV and PEDV outbreaks. Because of high mortality in piglets, it has become a common practice to vaccinate sows in order to transfer lactogenic immunity to protect piglets from TGEV and PEDV infections. Live attenuated and inactivated virus vaccines have been developed for both TGEV and PEDV. Since the spike protein is the major immunogen, numerous technologies have been employed to express the spike protein. These include DNA vaccine, vectored vaccine, subunit vaccine, and dendritic cell-based vaccine. Experimental and commercial TGEV and PEDV vaccines are listed in [Table 1](#).

Further Reading

- Carman, S., Josephson, G., McEwen, B., *et al.*, 2002. Field validation of a commercial blocking ELISA to differentiate antibody to transmissible gastroenteritis virus (TGEV) and porcine respiratory coronavirus and to identify TGEV-infected swine herds. *Journal of Veterinary Diagnostic Investigation* 14, 97–105.
- Choudhury, B., Dastjerdi, A., Doyle, N., Frossard, J.P., Steinbach, F., 2016. From the field to the lab – An European view on the global spread of PEDV. *Virus Research* 226, 40–49.
- Crawford, K., Lager, K.M., Kulshreshtha, V., Miller, L.C., Faaborg, K.S., 2016. Status of vaccines for porcine epidemic diarrhea virus in the United States and Canada. *Virus Research* 226, 108–116.
- Cubero, M.J., Leon, L., Contreras, A., *et al.*, 1993. Transmissible gastroenteritis in pigs in south east Spain: Prevalence and factors associated with infection. *Veterinary Record* 132, 238–241.
- Delmas, B., Gelfi, J., L'haridon, R., *et al.*, 1992. Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. *Nature* 357, 417–420.
- Diel, D.G., Lawson, S., Okda, F., *et al.*, 2016. Porcine epidemic diarrhea virus: An overview of current virological and serological diagnostic methods. *Virus Research* 226, 60–70.
- Eleouet, J.F., Rasschaert, D., Lambert, P., *et al.*, 1995. Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. *Virology* 206, 817–822.
- Fehr, A.R., Perlman, S., 2015. Coronaviruses: An overview of their replication and pathogenesis. *Methods in Molecular Biology* 1282, 1–23.
- Gerdtts, V., Zakharthouk, A., 2017. Vaccines for porcine epidemic diarrhea virus and other swine coronaviruses. *Veterinary Microbiology* 206, 45–51.
- Hooper, B.E., Haelterman, E.O., 1969. Lesions of the gastrointestinal tract of pigs infected with transmissible gastroenteritis. *Canadian Journal of Comparative Medicine* 33, 29–36.
- Hsu, T.H., Liu, H.P., Chin, C.Y., *et al.*, 2018. Detection, sequence analysis, and antibody prevalence of porcine deltacoronavirus in Taiwan. *Archives of Virology* 163, 3113–3117.
- Hu Jr., X., Li Jr., N., Tian Jr., Z., *et al.*, 2015. Molecular characterization and phylogenetic analysis of transmissible gastroenteritis virus HX strain isolated from China. *BMC Veterinary Research* 11, 72.
- Ji, C.M., Wang, B., Zhou, J., Huang, Y.W., 2018. Aminopeptidase-N-independent entry of porcine epidemic diarrhea virus into Vero or porcine small intestine epithelial cells. *Virology* 517, 16–23.
- Jung, K., Saif, L.J., 2015. Porcine epidemic diarrhea virus infection: Etiology, epidemiology, pathogenesis and immunoprophylaxis. *The Veterinary Journal* 204, 134–143.
- Kocherhans, R., Bridgen, A., Ackermann, M., Tobler, K., 2001. Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* 23, 137–144.
- Langel, S.N., Paim, F.C., Lager, K.M., Vlasova, A.N., Saif, L.J., 2016. Lactogenic immunity and vaccines for porcine epidemic diarrhea virus (PEDV): Historical and current concepts. *Virus Research* 226, 93–107.
- Lee, C., 2015. Porcine epidemic diarrhea virus: An emerging and re-emerging epizootic swine virus. *Virology Journal* 12, 193.
- Li, B.X., Ge, J.W., Li, Y.J., 2007. Porcine aminopeptidase N is a functional receptor for the PEDV coronavirus. *Virology* 365, 166–172.
- Li, W., Luo, R., He, Q., *et al.*, 2017. Aminopeptidase N is not required for porcine epidemic diarrhea virus cell entry. *Virus Research* 235, 6–13.
- Li, W., Van Kuppeveld, F.J.M., He, Q., Rottier, P.J.M., Bosch, B.J., 2016. Cellular entry of the porcine epidemic diarrhea virus. *Virus Research* 226, 117–127.
- Lin, C.M., Saif, L.J., Marthaler, D., Wang, Q., 2016. Evolution, antigenicity and pathogenicity of global porcine epidemic diarrhea virus strains. *Virus Research* 226, 20–39.
- Masters, P.S.P., Stanley, 2013. *Coronaviridae*, *Fields Virology*, sixth ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
- Miyazaki, A., Fukuda, M., Kuga, K., Takagi, M., Tsunemitsu, H., 2010. Prevalence of antibodies against transmissible gastroenteritis virus and porcine respiratory coronavirus among pigs in six regions in Japan. *The Journal of Veterinary Medical Science* 72, 943–946.
- Nakagawa, K., Lokugamage, K.G., Makino, S., 2016. Viral and cellular mRNA translation in coronavirus-infected cells. *Advances in Virus Research* 96, 165–192.
- Niederwerder, M.C., Hesse, R.A., 2018. Swine enteric coronavirus disease: A review of 4 years with porcine epidemic diarrhoea virus and porcine deltacoronavirus in the United States and Canada. *Transboundary and Emerging Diseases* 65, 660–675.
- Pensaert, M.B., Bouck, P.D.E., 1978. A new coronavirus-like particle associated with diarrhea in swine. *Archives of Virology* 58, 243–247.
- Pineyro, P.E., Lozada, M.I., Alarcon, L.V., *et al.*, 2018. First retrospective studies with etiological confirmation of porcine transmissible gastroenteritis virus infection in Argentina. *BMC Veterinary Research* 14, 292.
- Rasmussen, T.B., Boniotti, M.B., Papetti, A., *et al.*, 2018. Full-length genome sequences of porcine epidemic diarrhoea virus strain CV777; Use of NGS to analyse genomic and sub-genomic RNAs. *PLoS One* 13, e0193682.
- Risco, C., Anton, I.M., Enjuanes, L., Carrascosa, J.L., 1996. The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. *Journal of Virology* 70, 4773–4777.
- Saif, L.J., Van Cott, J.L., Brim, T.A., 1994. Immunity to transmissible gastroenteritis virus and porcine respiratory coronavirus infections in swine. *Veterinary Immunology and Immunopathology* 43, 89–97.
- Shirato, K., Maejima, M., Islam, M.T., *et al.*, 2016. Porcine aminopeptidase N is not a cellular receptor of porcine epidemic diarrhea virus, but promotes its infectivity via aminopeptidase activity. *Journal of General Virology* 97, 2528–2539.
- Song, D., Moon, H., Kang, B., 2015. Porcine epidemic diarrhea: A review of current epidemiology and available vaccines. *Clinical and Experimental Vaccine Research* 4, 166–176.
- Stevenson, G.W., Hoang, H., Schwartz, K.J., *et al.*, 2013. Emergence of Porcine epidemic diarrhea virus in the United States: Clinical signs, lesions, and viral genomic sequences. *Journal of Veterinary Diagnostic Investigation* 25, 649–654.
- Subramaniam, S., Yugo, D.M., Heffron, C.L., *et al.*, 2018. Vaccination of sows with a dendritic cell-targeted porcine epidemic diarrhea virus S1 protein-based candidate vaccine reduced viral shedding but exacerbated gross pathological lesions in suckling neonatal piglets. *Journal of General Virology* 99, 230–239.
- Wesley, R.D., Woods, R.D., McKean, J.D., Senn, M.K., Elazhary, Y., 1997. Prevalence of coronavirus antibodies in Iowa swine. *Canadian Journal of Veterinary Research* 61, 305–308.
- Zuniga, S., Pascual-Iglesias, A., Sanchez, C.M., Sola, I., Enjuanes, L., 2016. Virulence factors in porcine coronaviruses and vaccine design. *Virus Research* 226, 142–151.

Vaccinia Virus (*Poxviridae*)

Yan Xiang and Rebecca K Lane, University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of G.L. Smith, Vaccinia Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd, 2008, doi:10.1016/B978-0-12374410-4.00525-2.

The Origin of Vaccinia Virus

Vaccinia is the name given to the smallpox vaccine that has been in use since the 19th century, long before the nature of the vaccine was defined and the virus bearing its name was isolated from it. The name (*vacca*, Latin for cow) reflects an age-old misconception that the vaccine was derived from cowpox. While the original smallpox vaccine used by Jenner in the 1790s is assumed to be cowpox virus (CPXV), the identity and source of the vaccines utilized during much of the history of smallpox vaccination were not well-characterized. Jenner and other early European vaccinators derived vaccination material from both cowpox and horsepox, and the vaccine stocks were propagated by serial passage in humans and animals. In 1939, it was discovered that the vaccine strains in use at that time were serologically different from CPXV, and subsequent studies confirmed that vaccinia virus (VACV) and CPXV are distinct virus species. The natural host and origin of VACV remain unknown, but the most probable ancestor of VACV is the horsepox virus (HSPV), considered to be a rodent virus that sporadically caused the now-extinct disease in horses. VACV is phylogenetically closer to HSPV than any other virus species. A commercial smallpox vaccine produced in the USA in 1902 was recently found to be 99.7% identical to HSPV with respect to the genome sequence.

Classification

VACV belongs to the genus Orthopoxvirus (OPXV) of the family *Poxviridae*. Orthopoxvirus consists of more than ten genetically and antigenically closely related species, including CPXV, HSPV, monkeypox virus and variola virus, the causative agent for smallpox. Infection by one orthopoxvirus species induces an immune response that can cross-protect against all orthopoxvirus species, making VACV an effective vaccine against both smallpox and monkeypox. As the prototype poxvirus, VACV has been studied most intensively. The genome of more than 100 different VACVs has been completely sequenced. The Copenhagen vaccine strain was the first one sequenced and hence its genome has been used as the reference. The genes in Copenhagen were named with the convention that uses a letter (from A to O) to denote the genome fragment that is produced from digestion with HindIII restriction endonuclease, a number (from 1 to 57) for the open reading frame (ORF) number within the fragment, and L or R to represent the direction of transcription. The Copenhagen gene names are also used for the corresponding genes in other VACV strains. The letter L or R of the gene name is dropped when referring to the encoded protein. The VACV strain that has been the standard for laboratory research is the Western Reserve (WR), which originated from passaging the New York City Board of Health strain (NYCBH) in mice and is neurovirulent in that host.

Virion Structure

Two distinct forms of infectious virus are produced from infected cells. The more abundant and stable form is the intracellular mature virion (IMV), now simply referred to as the mature virion (MV). The less numerous and more fragile form is the cell-associated and extracellular enveloped virion (CEV and EEV), now collectively referred to as the extracellular virion (EV). The virions are complex and barrel shaped, lacking icosahedral or helical symmetry. The virion dimensions are approximately $360 \times 270 \times 250$ nm, amongst the largest of animal viruses. The ultrastructure as well as molecular architecture of the virions have been studied extensively with electron microscopy and mass spectrometry.

The MV form has an outer membrane layer embedded with ~30 different viral envelope proteins, many of which are involved in virus attachment or entry. Some of the MV envelope proteins are known targets of neutralizing antibodies. Unlike envelope proteins from many other viruses, none of the MV membrane proteins is modified by glycosylation. This may be due to an unusual process of MV envelope biogenesis that is unlike the budding process used by other viruses (see the section titled "Assembly"). Lying within the envelope is a dumbbell-shaped virus core, flanked by two amorphous "lateral bodies" in the concavities of the core. The surface of the core contains a palisade of protein spikes as well as pores for transport of molecules. The core encapsulates the virus genome along with all the enzymes and factors required for transcribing the early subset of viral genes (see the section titled "Transcription and translation"). Additional core proteins play roles in virus structure or morphogenesis. In total there are approximately 50 core proteins.

The EV form is essentially an MV wrapped with an extra envelope embedded with at least six EV-specific proteins. MV and EV are subjected to neutralization by different sets of antibodies, due to their difference in antigen composition in the outer envelope.

The EV envelope proteins are mostly glycosylated, and two of those are known targets of neutralizing antibodies. Although produced in lower amount than MV in infected cells, EVs are critical for viral dissemination in infected hosts, as they are specialized in long-range spread.

Genome

The VACV genome consists of a linear, double-stranded DNA (dsDNA) of 177–200 kilobase pairs (kbp), tightly packed with approximately 200 genes. The two DNA strands are covalently linked together at the ends by hairpin loops into one polynucleotide chain. The regions at the two termini are inverted terminal repeats (ITR), whose sequences are reverse complement of each other. ITRs contain a few genes and a variable number of tandem repeats. A core set of ~90 VACV genes are conserved among nearly all vertebrate poxviruses, encoding essential proteins that function at various stages of virus replication, including entry, gene transcription, genome replication and virus assembly. Most of these genes are located within the central ~100 kbp of the genome. Outside this conserved central region, the rest of the genome is more variable between different VACV strains, encoding proteins that are usually nonessential for virus replication but often participate in some aspect of host interactions, such as inhibition of cell death or evasion of host immunity. The naked VACV genome by itself is not infectious, as a virus-encoded and virion-packaged transcription system is required for initiating transcription from the genome. However, the VACV genome, which has been cloned in a bacterial artificial chromosome, can be initiated to produce infectious virus by using a related, replication-defective poxvirus as the helper. A chemically synthesized genome has been used to resurrect the horsepox virus, which appears to have become extinct in nature.

Life Cycle

Entry

VACV is capable of entering all mammalian and even insect cells that have been tested, with the exception of perhaps some primary human lymphocytes. It is thus presumed that either the entry requires no specific cellular receptors or that the receptors are ubiquitous molecules present on nearly all animal cells. The entry of MVs has been most extensively studied and will be described first. Four VACV proteins on the MV envelope are known to mediate MV attachment to cells. Three bind glycosaminoglycans (chondroitin or heparin) on the cell surface, while one binds laminins of the extracellular matrix. However, none of these viral proteins is individually required for entry, as loss of any one of the proteins has no deleterious effect on entry. VACV can enter cells either by direct fusion at the plasma membrane or by endocytosis followed by fusion with the endosomal membrane (**Fig. 1**), preferring one over the other depending on the virus strain or cell type. For entry through the plasma membrane, MV uses a virus-encoded entry fusion complex (EFC) to fuse directly with the plasma membrane in a pH-independent manner. The EFC consists of eleven proteins, including nine that form the core complex and two that peripherally associate with the complex. All EFC members are essential for entry, as loss of any one of the EFC proteins causes an entry defect. For entry through the endosomal path, MV particles are first internalized by macropinocytosis or fluid phase endocytosis. The interaction of MV with the cell surface triggers a signaling cascade that induces membrane rearrangements and the formation of actin- and ezrin-enriched membrane protrusions that engulf the virus particles. The internalized virus particles escape the endosomes by fusing the viral envelope with endosomal membranes. This process is activated by endosomal acidification and also requires the EFC. A viral protein, A26, regulates this process by interacting with the EFC and suppressing fusion at neutral pH. The low pH environment of the endosome results in a conformational change of A26 that de-represses the fusion machinery. Strain-specific differences in A26 expression may influence the preferred entry path used by different VACV strains.

Similar to MV, EV can enter through both the plasma membrane and the endocytic pathway in an EFC-dependent manner. The binding of EV to cells triggers the disruption of the EV outer envelope via a ligand-induced, non-fusogenic mechanism. This reaction occurs at neutral pH and involves cellular polyanionic molecules and two EV glycoproteins. Once devoid of the outer wrapper, the underlying MV particles can enter cells essentially as free MV particles as described above.

Transcription and Translation

The transcription of VACV genes occurs in the cytoplasm and depends on a completely virus-encoded transcription system, including transcription factors, a multi-subunit DNA-dependent RNA polymerase, capping and methylating enzymes, and a poly (A) polymerase. The transcriptome and translome of VACV have been extensively studied with RNA-seq and Ribo-seq. The ~200 genes of VACV are classified into three temporal classes: early, intermediate, and late, with some genes belonging to more than one temporal class. Each gene class is defined by class-specific transcription factors recognizing promoters with characteristics of that class. The sequential transcription of different classes of genes in a “cascade” fashion is enforced by the staged expression of these different class-specific transcription factors. The transcription factors for early genes are products of late genes and packaged in the virion. In turn, early and intermediate genes code for the necessary transcription factors for intermediate and late gene class, respectively (**Fig. 1**). The switch of the transcription program from cellular to viral and from one temporal class to another is also

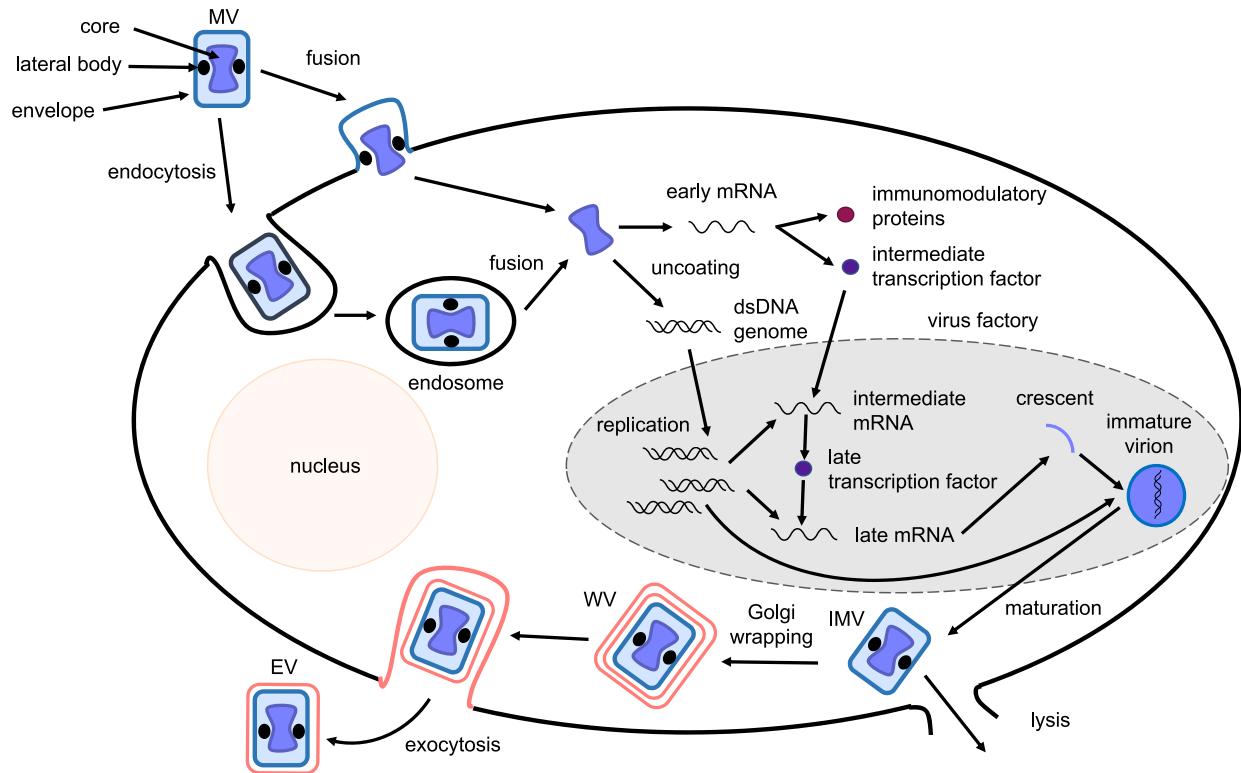


Fig. 1 Vaccinia virus life cycle. Mature virus (MV) virions enter cells by either fusion of the viral envelope with the plasma membrane or by endocytosis. Endocytosed virions escape to the cytoplasm by low pH-dependent fusion with the endosomal membrane. Once inside the cell, transcription of early mRNAs begins within the viral core and transcripts are exported into the cytoplasm for translation. The uncoating process then disassembles the core to release the viral genome, allowing for replication to proceed. Genome replication and transcription of intermediate and late genes take place in cytoplasmic virus factories. Transcription occurs in a temporally-regulated manner; early genes produce transcription factors for intermediate genes, and likewise, intermediate genes produce transcription factors for late genes. New virions begin to form as membrane crescents and assembling cores grow into a spherical immature virion (IV). Newly-replicated viral DNA is packaged into the IV. Condensation of the viral core, enzymatic cleavage of several proteins, and structural changes result in an intracellular mature virion (IMV). The IMV either remains in the cell until lysis or is wrapped with a double membrane from the trans-Golgi or endosomal vesicles to form the wrapped virion (WV). WV can exit cells by fusion of its outer envelope with the cell's plasma membrane, releasing double-enveloped extracellular virus (EV) by exocytosis.

facilitated by an increase in mRNA turnover mediated by two virus-encoded decapping enzymes, which remove the 5'-cap of mRNAs that protects them from a cellular 5-3' mRNA exonuclease.

As described above, the virion core is packaged with the complete transcription apparatus for early genes, so synthesis of early mRNAs does not require *de novo* transcription or translation and starts inside the virion cores immediately after cell entry. Nearly half of the VACV genes are transcribed early, and the mRNAs are extruded through pores in the virion core into the cytoplasm for translation. Early proteins include modulators for various cellular antiviral responses, enzymes and factors for genome replication, and transcription factors for intermediate genes. Some early viral proteins and the cellular ubiquitin proteasome system are required for a deliberate process of core uncoating, which simultaneously disrupts the transcription apparatus for early genes and releases the genome for replication. The replicated genome serves as the template for the transcription of intermediate and late genes (collectively referred to as post-replicative genes), which appears to be coupled with the translation at sites of genome replication. Post-replicative gene expression requires genome replication, probably because the genome becomes accessible to the newly synthesized transcription factors only after core uncoating and genome replication. The products of post-replicative genes include all virion components as well as regulatory proteins for virion morphogenesis.

The viral mRNAs do not undergo splicing but possess all the characteristic features of eukaryotic mRNA with a methylated cap at the 5' end and a poly(A) tail at the 3' end. Early VACV mRNAs have defined 3'-ends owing to the recognition of a specific transcription termination signal by the vaccinia termination factor. In contrast, no specific termination signal exists for post-replicative genes. Consequently, the post-replicative mRNAs are long with heterogeneous 3'-ends, leading to the formation of dsRNA by overlapping transcripts. The dsRNA represents a pathogen-associated molecular pattern that activates several innate antiviral pathways. As will be described later, VACV encodes proteins to specifically counteract the responses to dsRNA. Another unusual feature of post-replicative mRNAs is the presence of a poly(A) leader of heterogeneous lengths at the 5'-untranslated region. These poly(A) leaders are not derived from the DNA template, but are formed instead by RNA polymerase slippage. It has been shown that the 5' poly(A) leader confers translational advantage for the mRNAs.

Genome Replication

VACV early gene products generate a complete DNA replication apparatus in the cytoplasm, including a DNA polymerase, a helicase-primase, a single-stranded DNA-binding protein, a processivity factor, and a DNA ligase. All except for the DNA ligase are required for genome replication, as a cellular DNA ligase can also fulfill the apparently essential function of DNA ligation. In addition, a viral serine/threonine protein kinase and a viral uracil DNA glycosylase are required. The former phosphorylates a cellular protein that, in its unphosphorylated state, binds DNA and inhibits viral DNA replication. The latter is required for its interaction with the processivity factor but not for its enzymatic activity. Genome replication starts within 2 h after infection and takes place in discrete juxtannuclear areas of cytoplasm called virus “factories”, which are thought to be derived from individual infecting virions and devoid of most cellular organelles. Several models of DNA replication have been advanced, with a self-priming model being the more favored by the existing data. In this model, a nick is first introduced into a DNA strand near the terminal hairpin loop, creating an open 3' end which is extended by the polymerase to the end of the hairpin. The newly synthesized DNA segment then folds back to form a hairpin, essentially creating a primer for continuous DNA replication by strand displacement, until both strands of the DNA genome are traversed and the template is exhausted. An alternative model dictates that the genome replicates in a semi-conservative manner with leading and lagging strand synthesis, like eukaryotic DNA replication. This model is supported by reports of VACV DNA covalently linked to RNA, putative Okazaki fragments, and the essential role of the DNA ligase and primase. Regardless of the exact mechanism, the product of replication is a concatemer of multiple unit-length genomes connected by junctions that resemble Holliday junctions (HJ). The resolution of the concatemers into unit-length genomes is carried out by a viral HJ resolvase with similarity to the bacterial RuvC HJ resolvase.

Assembly

Virus assembly commences in the cytoplasmic factories following genome replication and post-replicative gene expression. The first viral structures discernible by electron microscopy are crescent-shaped membranes, which are highly unusual, open-ended membranes. The origin and biogenesis of the crescents are still debated, but the available evidence favors a model that involves the modification and scission of the ER membranes by viral proteins. Two MV envelope proteins and six regulatory viral membrane assembly proteins (VMAPs) are known to be required. One envelope protein, A17, is inserted into the ER membranes and functions similarly to cellular reticulons, which promote membrane curvature. VMAPs are thought to participate in the rupture of the ER and the stabilization of the resulting open membranes. A17 also interacts with a capsid-like protein, D13, which assembles a honeycomb-like scaffold on top of the membranes to give them a spherical shape. The crescents grow in length while engulfing the core proteins, eventually closing off to form the spherical immature virions (IVs). A complex of seven core proteins is required for the association of the core with the crescents. Before the membrane is completely sealed, the newly synthesized virus genome is also packaged into the IV, facilitated by a viral ATPase and a viral telomere-binding protein. Proteolytic cleavage of several virion proteins accompanies the maturation of the spherical IVs into the brick-shaped MVs. Particularly, the cleavage of A17 by a viral protease removes the capsid-like scaffold from the surface of MV. MV envelope proteins are nonglycosylated, probably because they are synthesized on viral membranes that are already well separated from the ER. A unique cytoplasmic redox system formed by three viral proteins catalyzes the unusual cytoplasmic disulfide bond formation in some MV envelope proteins, including the EFC proteins.

The majority of MV particles remain within the cell until it lyses, but a small fraction of them are transported on microtubules to near the microtubule organizing center where they are wrapped by trans-Golgi or endosomal vesicles that have been modified by the inclusion of several viral proteins. The wrapping of MV with two additional membranes produces the wrapped virion (WV) with a total of three membranes. WVs are transported to the cell surface on microtubules, facilitated by the interaction between microtubule motor proteins and two viral proteins on the surface of WV. The outer membrane of the WV fuses with the plasma membrane to externalize the virion via exocytosis, producing extracellular virions (EV) with two membranes ([Fig. 1](#)). Most of the EVs remain attached to the cell (CEV), while some are released (EEV), with the relative ratio influenced by both the cell type and some EV proteins. The interaction of EV with the surface of a VACV-infected cell triggers a signaling cascade that induces actin polymerization and the formation of finger-like actin tails right beneath the plasma membrane. These motile actin tails propel EVs away from the cell and into surrounding cells, preventing superinfection and facilitating efficient cell-to-cell spread of EVs.

Host Range

Although its natural hosts are unknown, VACV is capable of infecting a wide range of wild and experimental animals, including primates, ungulates and rodents. Cultured cells from an even broader range of animal species are permissive for VACV replication, with the exception of a few cell types, such as Chinese hamster ovary (CHO) cells. In rare cases when VACV does fail to infect, the block in replication occurs after cell entry at an intracellular step, an example being the block of intermediate protein synthesis in CHO cells. A variety of viral “host-range” genes that are essential for VACV replication only in some cell types have been identified. Of them, two pairs of functionally redundant genes, E3L/K3L and K1L/C7L, affect VACV replication in the widest range of mammalian cells. Many host range genes function by antagonizing host antiviral factors that would otherwise block VACV replication. The E3L/K3L pair mainly inhibits dsRNA-activated protein kinase R (PKR), while the K1L/C7L pair inhibits two interferon-inducible antiviral factors SAMD9 and SAMD9L. Host species-specific differences in PKR and SAMD9L are known to

affect VACV replication in cells of different species. For example, a species-specific difference in Chinese hamster SAMD9L results in resistance to inhibition by both K1L and C7L, preventing VACV from replicating in CHO cells.

Virulence

Naturally circulating VACV or its close relatives have been reported to cause infections in bovines and humans. In the Indian subcontinent, buffalopox virus (BPXV), a VACV sub-lineage, has sporadically caused pox lesions in Asian buffalos and their direct human contacts since the 1930s. Similarly, in south America, VACV-like viruses have repeatedly caused disease outbreaks in cattle and their human handlers since the late 1990s, having a significant impact on the dairy industry and public health in Brazil. VACV circulation has been observed in several species of marsupials and wild rodents in Brazil, suggesting that wild rodents are reservoirs or intermediate hosts for VACV transmission. Phylogenetic analyses suggest that the feral BPXV and Brazilian VACV-like viruses may derive from escaped smallpox vaccine viruses.

VACV pathogenesis has been studied extensively with mouse models. The intranasal or intradermal routes of infection are commonly used, as respiratory tract and direct skin contact represent the main transmission routes for natural OPXV infection. Intranasal infection usually causes a systemic illness in mice, manifesting in body weight loss and the spread of virus to many tissues. Intradermal infection causes a localized disease in mice, characterized by relatively healthy animals with skin lesions. The viruses do not spread from the skin and the lesions heal in about 2 weeks. Many VACV proteins have been found to contribute to virulence in the mouse models, including enzymes that are involved in nucleotide metabolism, proteins that affect EV production, and proteins that modulate the host antiviral response. Among them, the loss of proteins that are involved in EV production results in the greatest virus attenuation.

Immunomodulators

A growing list of VACV proteins have been found to modulate various aspects of the host immune system, particularly the innate immune system. These so-called immunomodulators are divided into the secreted proteins that act extracellularly and the intracellular factors that act within infected cells. The extracellular proteins include binding proteins for complement factors and various cytokines such as type I and type II interferons, IL-1 β , IL-18, TNF- α , and CC chemokines. The interaction of the specific viral binding protein prevents the corresponding immune molecule from binding to its cellular receptor and from triggering the downstream signaling event. The intracellular immunomodulators include inhibitors of apoptosis and the signal transduction pathways initiated by cell pattern recognition receptors and interferons. They also include inhibitors of some interferon-induced antiviral factors, such as PKR and SAMD9 (see the section titled "Host range"). The signaling pathways and antiviral factors are often modulated by multiple VACV proteins that target different points of the pathway or through different mechanisms. For example, PKR is inhibited by two VACV proteins, with one sequestering dsRNA that is needed for PKR activation and the other acting as a pseudo-substrate inhibitor of PKR.

VACV: The Smallpox Vaccine

Jenner's inquiry in the 1790s heralded a new era of infectious disease prevention. In honor of Jenner's discovery, "vaccination" is now used generally as the term for the preventive immunization procedure against any infectious disease. As the vaccine for smallpox, VACV is one of the most successful vaccines in human history and remains the only vaccine that successfully eradicated a human disease. During the global smallpox eradication campaign, several replication-competent VACV strains were used as the vaccine by different countries. They include the NYCBH strain (Americas), the Copenhagen strain (Denmark), the Lister strain (UK), the Ankara strain (Turkey), and the Tian Tan strain (China). However, vaccination with these strains was associated with a relatively high rate of adverse reactions, including generalized vaccinia, progressive vaccinia, eczema vaccinatum, and postvaccinial encephalitis. Some of these adverse reactions could be fatal and were mainly associated with individuals having an underlying immune deficiency. The vaccination of the general public ceased globally after the eradication of smallpox in the late 1970s. For many years afterwards, the only smallpox vaccine available for limited use in the US was the Dryvax vaccine, manufactured from the lymph fluid of calves' skin infected with the NYCBH strain. The heightened concern over smallpox bioterrorism after 2001 led to a concerted effort by the public and private sectors in developing vaccines and antivirals against smallpox. The first new smallpox vaccine approved in the US is ACAM2000, which was derived from a clone of the Dryvax vaccine and manufactured on the Vero monkey cell line. ACAM2000 has a similar safety profile as the Dryvax vaccine and is contraindicated in individuals with atopic dermatitis or immune deficiencies. The vaccine is administered by scarification of the skin using a bifurcated needle containing the vaccine, and a successful vaccination results in VACV replication at the inoculation site producing a pustule or "take". A safer vaccine, derived from the highly attenuated Modified vaccinia Ankara strain (MVA), was recently approved in the US for use against both smallpox and monkeypox. MVA was developed in the 1970s through passaging the Ankara strain in chick embryo fibroblast cells for more than 500 generations, resulting in the loss of nearly 30 kb of genomic information. MVA is replication-defective in most mammalian cells, and the MVA-based smallpox vaccine is approved for use even in people with atopic dermatitis or immune deficiencies. The vaccine is administered by two subcutaneous injections 4 weeks apart.

Smallpox was eradicated before the correlates of protective immunity to smallpox could be studied with modern immunology, but recent studies with animal models and proteome-wide immune profiling have provided a clear understanding of the immune responses to VACV immunization. A single immunization with live VACV elicits robust antibody and cytotoxic T cell responses to a wide array of viral proteins, which could last for decades in humans. Neutralizing antibodies play an essential role in protection against OPXV infection, while the T-cell response is essential for recovery from infection in the absence of antibodies. The antibody responses predominantly recognize structural, late VACV proteins, while the cytotoxic T cell responses predominantly recognize epitopes present in nonstructural, early VACV proteins. For optimal protection against OPXV infection, antibodies against both MV and EV are required. One protein (B5) is the major target of neutralizing antibodies against EV, while multiple proteins are the targets of MV neutralizing antibodies. None of the MV neutralizing antibodies are individually required for MV neutralization or dominant in all vaccinated individuals. Instead, the highly redundant neutralizing antibody responses may be a feature of the smallpox vaccine that ensures protection in very different human populations.

VACV: The Vaccine Vectors

VACV has been used as a vector for developing vaccine candidates for infectious diseases and cancer, and as an oncolytic virus to target human cancers. As a vector, VACV has the advantage of the relative ease of generating recombinant virus, a large capacity for accommodating exogenous DNA, a high level of expression of biologically active protein and a wide host range. For human vaccine development, VACV strains that are replication defective in mammalian cells, such as MVA and NYVAC, are often used. NYVAC derived from VACV Copenhagen by the specific deletion of 18 genes including the host range genes K1L and C7L (see the section titled “Host range”), which are essential for replication in mammalian cells. Consequently, NYVAC is highly attenuated, but remains capable of inducing strong immune responses. These vectors have been used for candidate vaccines against AIDS and malaria with some success in human clinical trials. A replication-competent VACV that expresses the rabies glycoprotein has been used as an oral based rabies vaccine for wildlife, which is responsible for significantly reducing the incidence of rabies in the US and Europe.

Antiviral Drugs

ST-246 (Tecovirimat) is the first antiviral drug that has been approved in the US for treating smallpox. It is a potent inhibitor of a highly conserved OPXV envelope protein essential for wrapping of MVs (see the section titled “Assembly”). ST-246 is effective against VACV in animal models and presumably also in humans. Another class of antiviral that has been in clinical trials for treating OPXV infection is Cidofovir and its derivatives. Cidofovir is an acyclic nucleoside analog licensed for treating human cytomegalovirus infection that also inhibits the poxvirus DNA polymerase.

Further Reading

- Condit, R.C., Moussatche, N., Traktman, P., 2006. In a nutshell: structure and assembly of the vaccinia virion. *Advances in Virus Research* 66, 31–124.
- Esparza, J., Schrick, L., Damaso, C.R., Nitsche, A., 2017. Equination (inoculation of horsepox): An early alternative to vaccination (inoculation of cowpox) and the potential role of horsepox virus in the origin of the smallpox vaccine. *Vaccine* 35, 7222–7230.
- Lane, J.M., Ruben, F.L., Neff, J.M., Millar, J.D., 1969. Complications of smallpox vaccination, 1968: National surveillance in the United States. *New England Journal of Medicine* 281, 1201–1208.
- McFadden, G., 2005. Poxvirus tropism. *Nature Review Microbiology* 3, 201–213.
- Moss Poxvirus, B., 2013. DNA Replication. *Cold Spring Harbor Perspectives in Biology* 5, a010199.
- Moss, B., 2013. Poxviridae. In: Knipe, D.M. (Ed.), *Fields Virology*, sixth ed., 2. Philadelphia, PA: Lippincott Williams and Wilkins, pp. 2129–2159.
- Moss, B., 2015. Poxvirus membrane biogenesis. *Virology*. 479–480, 619–626.
- Moss, B., 2016. Membrane fusion during poxvirus entry. *Seminars in Cell & Developmental Biology* 60, 89–96.
- Sánchez-Sampedro, L., Perdiguero, B., Mejías-Pérez, E., *et al.*, 2015. The evolution of poxvirus vaccines. *Viruses* 7, 1726–1803.
- Schmidt, F.L., Bleck, C.K., Mercer, J., 2012. Poxvirus host cell entry. *Current Opinion in Virology* 2, 20–27.
- Smith, G.L., Vanderplasschen, A., Law, M., 2002. The formation and function of extracellular enveloped vaccinia virus. *Journal of General Virology* 83, 2915–2931.
- Smith, G.L., Benfield, C.T., Maluquer de Motes, C., Mazzon, M., 2013. Vaccinia virus immune evasion: Mechanisms, virulence and immunogenicity. *Journal of General Virology* 94, 2367–2392.
- Veyer, D.L., Carrara, G., Maluquer de Motes, C., Smith, G.L., 2017. Vaccinia virus evasion of regulated cell death. *Immunology Letters* 186, 68–80.

Relevant Website

<https://www.viprbrc.org/brc/home.spg?decorator=pox>
(ViPR).
Poxviridae.
Virus Pathogen Database.

Varicella-Zoster Virus (*Herpesviridae*)

Jeffrey I Cohen, Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Published by Elsevier Ltd.

Glossary

Dermatome The area on the skin the is innervated by a single cranial or spinal nerve.

Varicella A disease also known as chickenpox that presents with a disseminated vesicular rash.

VariZIG Varicella immune globulin obtained from plasma of persons with antibodies to varicella-zoster virus.

Vesicle A small fluid filled blister on the skin.

Zoster A disease also known as shingles that usually presents with a dermatomal, vesicular rash.

Classification

Varicella-zoster virus (VZV) is a member of the Alphaherpesvirus subfamily of the family Herpesviridae. Other alpha-herpesviruses that infect humans include herpes simplex viruses (HSV) 1 and 2, and very rarely macacine herpesvirus 1 (B virus). All of these agents exhibit relatively short replicative cycles, destroy the infected cell, and establish latent infection in sensory ganglia.

Simian varicella virus (SVV) is the most closely related, well characterized virus to VZV. Natural infection with SVV can cause a varicella-like illness in Old World monkeys; the virus establishes latency in trigeminal ganglia and can spontaneously reactivate to cause a rash which can transmit the virus to naïve animals. SVV shares a nearly identical set of genes with VZV; the virus is not known to infect humans.

Virion Structure

It is not possible to obtain high titers of cell free VZV, so high resolution structural analysis of the virus has not been performed. Nonetheless, using electron microscopy the virus has been shown like other herpesviruses to contain DNA within a nucleocapsid that is surrounded by a tegument and viral envelope (**Fig. 1**). The nucleocapsid is derived from the major nucleocapsid protein (encoded by ORF 40) and, largely by analogy with other herpesviruses, proteins encoded by VZV ORFs 20, 21, 23, 33, 33.5, and 41. Like other alphaherpesviruses the nucleocapsid consists of 162 capsomeres arranged in an icosahedron. Several viral proteins have demonstrated to be localized in the tegument including proteins encoded by ORFs 4, 9, 10, 11, 47, 62, 63, and 66. Viral glycoproteins stud the membrane and include gB, gC, gE, gH, gI, gK, gL, gM, and gN.

Genome

The VZV genome consists of double-stranded DNA and like other herpesviruses is usually linear in the virion and circularizes in latently infected cells. It contains about 125,000 base pairs with a single unpaired nucleotide at both ends of the genome. The VZV genome consists of a unique long region (UL, which includes ORFs 1–61 and a portion of ORF0 [sometimes referred to as ORF S/L]), an internal repeat short region (IRS, which includes ORFs 62–64), a unique short region (US, which includes ORFs 65–68), a terminal repeat short region (TRS, which includes ORFs 69–71, and a portion of ORF 0), as well as very short inverted repeat regions (terminal and internal repeat long; TRL and IRL) (**Fig. 2**). The viral genome encodes at least 20 small noncoding RNAs as well as the viral latency associated transcript (VLT). The genome exists predominantly as two different isomers depending on the orientation of the unique short region. It has five repeat (R) regions. R1 is located in ORF11, R2 in ORF 14, R3 in ORF 22, R4 between ORF62 and ORF63, and R5 between ORF60 and ORF61. R3 is the largest of the repeats, approximately 1000 base pairs in length.

VZV encodes at least 70 genes, 5 of which are not conserved with HSV. These latter genes are ORF 0, 1, 2, 13, 32, and 57; ORF13 which encodes the thymidylate synthetase has a homolog in human herpesvirus 8, while the other four genes have homologs in equine alpha-herpesviruses.

VZV has the most stable genome of all the human herpesviruses. Comparison of over 1300 whole VZV genomes from throughout the world has resulted in classification of viruses into 7 clades (1 – 6, 9) and 2 provisional clades (VII, VIII). At least one whole genome sequence is required for classification into a clade, while a partial genomic sequence defines a provisional clade. Clade 1 contains European isolates including VZV Dumas (the first completed VZV genome), clade 2 contains Asian isolates including the parental VZV Oka strain and the Oka vaccine derived from the parental strain, clade 3 contains European isolates

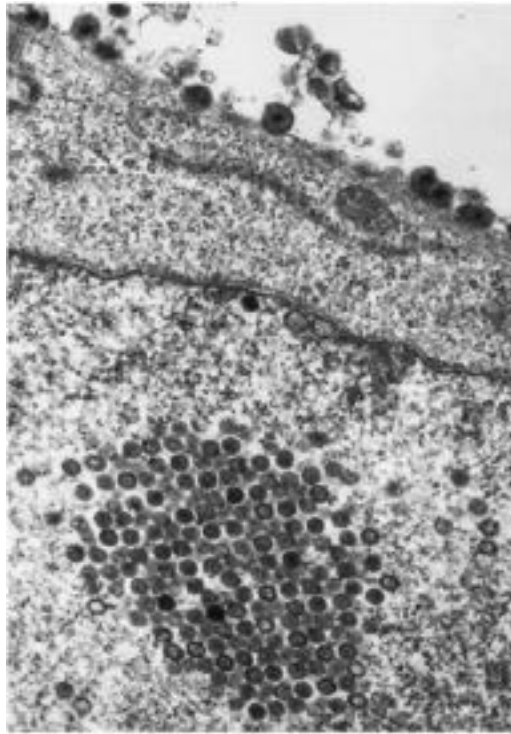


Fig. 1 Electron micrograph of varicella-zoster infected cell. Nucleocapsids in the nucleus (bottom) and virions budding from the cell membrane (top). Reproduced with permission from Cohen, J. I., Straus, S. E., 2001. Varicella-zoster virus and its replication. In: Knipe, D. M., Howley, P. M., Cohen, J. I., *et al.* (Eds.), *Fields Virology*. Philadelphia, Lippincott-Williams & Wilkins, pp. 2707–2730.

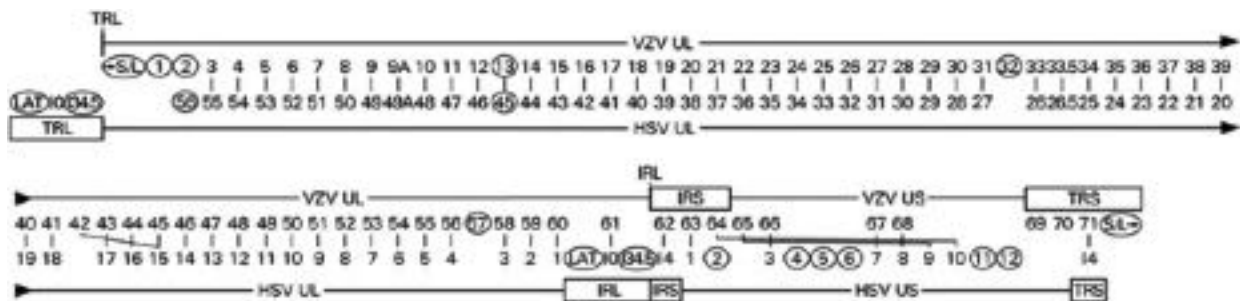


Fig. 2 Map of the varicella-zoster virus (VZV) genome (top) and the herpes simplex virus (HSV) genomes. Unique short (US), unique long (UL), terminal repeat (TR), internal repeat (IR), long (L), and short (S) regions are shown. Genes unique to varicella-zoster or herpes simplex are circled. Reproduced with permission from Cohen, J.I., Straus, S.E., Arvin, A.M., 2007. Varicella-zoster virus: Replication, pathogenesis, and management. In: Knipe, D.M., Howley, P.M., Cohen, J.I., *et al.* (Eds.), *Fields Virology*. Philadelphia, Lippincott-Williams & Wilkins, pp. 2773–2818.

including VZV HJO, clade 4 contains African isolates including VZV DR, clade 5 contains European isolates including VZV CA123, and clade 9 contain a United States isolate VZV KY037798.

Life Cycle

VZV infects epithelial cells in the oropharynx and infects T cells that circulate throughout the body. Virus spreads to the skin, resulting in the rash of varicella, but also to other tissues including sensory, autonomic, and cranial nerve ganglia where the virus establishes latency. In the skin, virus is present in vesicles which become pustular and break down. VZV is released from the skin to infect other persons. In vitro, the virus has been shown to infect CD4 and CD8 T cells, NK and NKT cells, B cells, and monocytes.

During initial infection, the virus enters host cells. The inability to obtain high titers of cell-free virus, has prevented researchers from identifying important steps in VZV entry and many steps are inferred from studies of the well characterized other human

alpha-herpesvirus HSV. Initially, the virus attaches to host cells by electrostatic interactions between viral glycoproteins and host cell glycosaminoglycans. Heparan sulfate on the surface of cells binds to VZV gB. The virus is thought to undergo endocytosis followed by fusion of the virus membrane to the host cell membrane. Three entry mediators, mannose-6 phosphate (cation independent), myelin-associated glycoprotein, and insulin degrading enzyme have been proposed as entry mediators, but these have not been confirmed in other laboratories. Glycoproteins gB, gE, gH, gL are all essential and as such are likely to have key roles in VZV infection. VZV gE is the most abundant glycoprotein on the virion and by analogy to the most prevalent glycoproteins in other herpesviruses is likely important for attachment to cells and binding to a receptor. By analogy with other herpesviruses, gH, and its chaperone gL, likely trigger activation of gB and all three glycoproteins are required for fusion of the virus membrane to the cell. The viral nucleocapsid is then uncoated and transported to the nucleus, where tegument proteins are released in the nucleus and the linear viral DNA circularizes and DNA is likely replicated through a rolling circle model.

Initially the immediate-early (IE) genes are transcribed and translated which include IE4, IE61, IE62, and IE63. Since it is difficult to obtain high titer cell-free VZV, the kinetics (IE, early, or late) of most viral genes have not been formally determined, but have been putatively assigned based on homology with HSV. VZV IE62 is the major viral transactivator and upregulates expression of IE4 and IE61 as well as putative early and late VZV genes. IE62 binds to cellular transcription factors USF, TFIIB, and SP1 as well as to TATA binding protein. VZV IE4 transactivates some viral IE, early, and late genes and binds to TFIIB and to the NF- κ B subunits p50 and p65. IE61 can upregulate expression of putative IE, early, and late VZV genes and can modulate expression of viral genes that are transactivated by IE4 or IE62. IE63 may down-regulate expression of some viral genes and binds to anti-silencing function-1 protein, RNA polymerase 2, and IE62.

VZV putative early genes encode proteins important for viral DNA replication. These include the large (ORF28) and putative small (ORF16) subunits of the viral DNA polymerase, the single-stranded DNA binding protein (ORF29), the origin binding protein (ORF51), the viral dUTPase (ORF8), the viral uracil DNA glycosylase (ORF 59), the large (ORF19) and small (ORF18) subunits of the viral ribonucleotide reductase, and the viral thymidine kinase (ORF36). The latter protein phosphorylates acyclovir resulting in inhibition of virus replication.

VZV putative late proteins include protein kinases, a protease (ORF33), nucleocapsid and tegument proteins, and glycoproteins. The viral protein kinases (ORF47 and ORF66) phosphorylate both cellular and viral proteins. Nucleocapsid proteins include the major nucleocapsid protein (ORF40), a 115 kDa protein (ORF21), a 30 kDa protein (ORF23), and the viral assembly protein (ORF33.5) which is thought to form a scaffold for assembly of the nucleocapsid. Tegument proteins include ORF10 which transactivates IE62 and the ORF11 RNA binding protein.

VZV encodes multiple glycoproteins important for virus entry and egress. VZV gE (ORF68) forms a heterodimer with gI (ORF67); the latter is important for trafficking of gE to the cell surface. gE is essential for virus infection and binds the Fc receptor of immunoglobulin. gH (ORF37) forms a heterodimer with gL (ORF60); the latter is important for maturation of gH and trafficking of gH to the plasma membrane. gH/gL are thought to activate gB (ORF 31) to allow the virus to fuse to the cell membrane. Other glycoproteins include gC (ORF14), gK (ORF5), and gM (ORF50) which forms a heterodimer with gN (ORF9A); of these three glycoproteins, only gK is essential for virus growth in vitro.

Virus egress from cells is not well understood. In cell culture no cell-free infectious virus is released, instead the virus spreads by cell-cell fusion. In vitro, virus is directed to pre-lysosomes where it is degraded. In contrast, in vivo virus is released through an alternative pathway, likely in cells not expressing the cation-independent mannose-6 phosphate receptor which otherwise binds to mannose-6 phosphate on the viral envelope and directs the virus for degradation in pre-lysosomes.

VZV establishes latency in neurons of the dorsal root ganglia (alongside the spine), cranial nerve ganglia, enteric ganglia (in the intestine), and autonomic (e.g., celiac and vagus nerve) ganglia. The virus enters dorsal root and cranial nerve ganglia either by the bloodstream or by retrograde transport from cutaneous nerves. Analysis of transcripts expressed in latently infected human ganglia obtained less than 9 h after death indicates that two transcripts are usually expressed in both the nucleus and cytoplasm of neurons- VZV ORF63 and VLT (VZV latency associated transcript). VLT RNA is expressed more abundantly than ORF63 RNA in latently infected human ganglia. ORF63 encodes IE63, an immediate-early protein in lytically infected cells, which is the homolog of HSV US1.5. In lytically infected cells VLT is spliced in different forms than in latently infected cells, and in these lytic cells VLT encodes a late protein; VLT protein was not detected in latently infected cells. In latently infected cells VLT is a multi-spliced unique transcript that is expressed anti-sense to ORF61. Somewhat surprisingly, not all neurons latently infected with VZV DNA express ORF63 or VLT transcripts. Expression of VLT inhibits ORF61 (a potent viral transactivator) expression in vitro which suggests that VLT could help the virus to maintain latency. Additional transcripts have been identified in latently infected ganglia including ORF4, ORF21, ORF29, ORF62, and ORF66. VZV microRNAs have not been detected in human ganglia. While numerous prior studies reported VZV proteins in latently infected human ganglia, more recent reports indicate that detection of some of these proteins may have been due to cross-reactivity of the antibodies with human proteins. Thus, at present VZV proteins are generally not thought to be expressed in latent infection. The viral genome is thought to be a circular episome exclusively in neurons.

Epidemiology

Infection with VZV occurs worldwide and prior to the development of the varicella vaccine over 95% of children in temperate countries were infected with the virus before age 15. Varicella occurs more commonly in the winter and spring, often occurring at a later age in tropical climates. Varicella may occur after exposure of susceptible persons to chickenpox or to herpes zoster. Over 95% of primary infections result in symptomatic chickenpox. VZV is transmitted by the respiratory route and virus has been detected by

PCR in room air from patients with varicella or zoster. Intimate, rather than casual contact is important for transmission. Chickenpox is highly contagious; about 60%–90% of susceptible household contacts become infected. In contrast, only 20%–30% of susceptible contacts of persons with zoster develop varicella. Patients with varicella are infectious from two days before the onset of the rash until all the lesions have crusted.

Primary infection with VZV results in viral replication in epithelial cells of the upper respiratory tract and oropharynx with lesions present on the respiratory mucosa. T cells subsequently become infected and transmit virus throughout the body including to the skin and to the dorsal root and cranial nerve ganglia where the virus establishes latency. Virus has been detected by PCR in the oropharynx and virus has been cultured from the blood early during varicella. Virus infection of epidermal cells in the skin is thought to be initially limited by interferon; later T cells trafficking in the skin may become infected and disseminate the virus to various organs throughout the body including the nervous system.

Zoster is due to reactivation of VZV in patients who have had prior chickenpox; some of these patients may not recall the primary infection. The virus usually reactivates from dorsal root or cranial nerve ganglia, travels down the axon, and replicates in the skin resulting in a rash in the dermatome innervated by the nerve. Zoster does not result from exposure to chickenpox or to other cases of zoster. In contrast to varicella, zoster occurs throughout the year. The risk of reactivating VZV due to varicella vaccine virus is lower than due to wild-type virus.

The risk of zoster rises steadily with age, and persons who live to age 85 have a 50% risk of reactivating the virus and developing zoster, if they do not receive the zoster vaccine. Immunocompromised patients, such as those with the acquired immunodeficiency syndrome (AIDS), hematologic malignancies, as well as persons who have received hematopoietic stem cell or organ transplants, or those on immunosuppressive medications that reduce cellular immunity have a high incidence of zoster. Viremia and subsequent cutaneous or visceral dissemination of lesions may occur in zoster, especially in immunocompromised patients. Injury to a nerve is also a risk factor for reactivation of the virus from the ganglia innervated by the nerve. The rate of zoster has been increasing in the last 60 years perhaps because of longer lifespans or better recognition of the disease. Recurrent zoster is uncommon; less than 4% of patients experience a second episode. Asymptomatic viremia has been detected in bone marrow transplant recipients and has been followed by recovery of cell mediated immunity.

Severe VZV infections have been reported in children and adults with heterozygous mutations in subunits of RNA polymerase III (POLR3A, POLR3C, POLR3E, or POLR3F). These patients have reduced induction of interferon after stimulation of cells with poly dA: dT or after VZV infection. Unlike HSV, the VZV genome is AT-rich, which may explain the severe CNS or pulmonary disease seen with VZV, but not HSV in persons with these mutations.

Clinical Features

The incubation period for chickenpox is 2 weeks, with a range of 10–21 days. The disease begins with fever and malaise, followed by a generalized vesicular rash (**Fig. 3**) which evolves into pustules before crusting. Lesions tend to appear first on the head and trunk and then spread to the extremities. New lesions usually follow viremic waves for three to five days and, in the normal host, most lesions are crusted and healed by two weeks. Lesions in different stages coexist in an individual. The disease is usually self-limited in the normal host.

Complications of varicella are more common in neonates whose mothers were first infected 5 days before to 2 days after the infant was born, in children with malnutrition, in immunocompromised patients (e.g., malignancy or immunosuppressive therapy), in pregnant women, and in older adults. These complications include bacterial superinfection of the skin, cerebellar ataxia, pneumonia, hepatitis, thrombocytopenia, and less commonly encephalitis, meningitis, purpura fulminans, and central nervous system vasculopathy (vasculitis of the cerebral arteries) which often presents as a stroke. Group A streptococcus or staphylococcus is the most common serious superinfection and can result in cellulitis, fasciitis, or bacteremia. Reye syndrome occurs in rare children who take aspirin to treat varicella fever. Severe complications occur more often in children with impaired T cell immunity such as those with leukemia, HIV, or congenital immune deficiencies or those receiving corticosteroids. Prior to the widespread use of the varicella vaccine in the United States, there were about 3–4 million cases of varicella each year with about 100 deaths; however, the number of cases has dropped sharply and death from varicella in the United States is very rare. Pregnant women with primary infection during the first trimester can give birth to infants with the congenital varicella syndrome, manifested by limb hypoplasia, scarring of the skin, central nervous system abnormalities, and cataracts. Pregnant women with primary infection during the third trimester are more likely to have complications of varicella, especially pneumonia.

Zoster usually presents with pain and dysesthesias 1–4 days before the onset of the vesicular rash. The rash is usually painful and confined to a single dermatome (**Fig. 4**) but may involve several adjacent dermatomes. The trigeminal or thoracic dermatomes are most often affected. Fever and malaise often accompany the rash. Vesicles often are pustular by day 4 and become crusted by day 10 in the normal host. Some patients have pain, but do not develop a rash which is referred to as zoster sine herpete.

Postherpetic neuralgia, manifested by pain that persists at least one month after the initial rash resolves can last for months to several years, and is the most common and disconcerting complication of zoster in the normal host. Postherpetic neuralgia is more common in the elderly. Ocular disease manifested by keratitis, episcleritis, iritis or glaucoma is not uncommon. Other less common ocular complications include acute retinal necrosis which can occur in apparently healthy persons, or progressive outer retinal necrosis which occurs in very immunocompromised persons. Less common complications include encephalitis, myelitis, the Ramsay-Hunt syndrome (lesions in the ear canal, with auditory and facial nerve involvement), facial weakness, bacterial superinfection, and pneumonitis. Patients can present with strokes concurrent with or within a few months after presenting with



Fig. 3 Child with disseminated rash of varicella. Courtesy of the Centers for Disease Control and Prevention.



Fig. 4 Adult with zoster in the ophthalmic branch of the trigeminal nerve (A) and vesicles and pustules with the rash of zoster (B). Reproduced with permission from Cohen, J. I., 2013. Herpes zoster. *The New England Journal of Medicine* 369, 255–263.

zoster due to central nervous system vasculopathy. Patients with impaired T cell immunity are more likely to develop disseminated disease with neurologic, ocular, or visceral involvement. Patients with AIDS have a high frequency of zoster and may develop recurrent or chronic disease with verrucous, hyperkeratotic skin lesions. Immunocompromised transplant recipients can present with viral hepatitis, pancreatitis, or disease involving other organs, sometimes in the absence of, or before development of a rash.

Pathogenesis

Varicella lesions are readily recognized in the skin and mucous membranes. However, similar lesions also occur in the mucosa of the respiratory and gastrointestinal tracts, liver, spleen, and any tissue, and remain unrecognized except in severe cases. With severe disease there is inflammatory infiltration of the small vessels of most organs. Zoster causes inflammation and necrosis of the sensory ganglia and its nerves, and skin lesions which are histopathologically identical to those seen with varicella.

Cutaneous lesions due to VZV begin with infection of capillary endothelial cells followed by direct spread to epidermal epithelial cells. The epidermis becomes edematous with acantholysis and vesicle formation. Mononuclear cells infiltrate the small

vessels of the dermis. Initially, vesicles contain clear fluid with cell-free virus, but later the vesicles become cloudy and contain neutrophils, macrophages, interferon, and other cellular and humoral components of the inflammatory response pathways. Subsequently, the vesicles dry leaving a crust that heals, usually without scarring.

Cells infected with VZV show eosinophilic intranuclear inclusions with multinucleate giant cell formation. These changes are not specific for VZV, as they are seen with herpes simplex and cytomegalovirus infections.

Infection with VZV elicits both a humoral and cellular immune response. The ability of VZV immune globulin (VariZIG) to attenuate or prevent infection in exposed children (see Prevention and Control, below) indicates that virus-specific antibody is important for protection from primary infection. The presence of VZV-specific IgG does not correlate, however, with protection from zoster. Antibody to VZV is often present by the time the rash of varicella first appears. Virus-specific IgM, IgG, and IgA are present within five days of symptomatic disease; however, only IgG persists for life. Antibodies to viral glycoproteins gE, gB, gH, and the immediate-early 62 protein have been detected during acute infection, and the titers of antibodies to these proteins are boosted during recurrent infection. The mere presence of antibody to VZV glycoproteins in children with leukemia who had received live varicella vaccine is not adequate to prevent breakthrough varicella or zoster.

Cellular immune responses to VZV are more important in recovery from acute varicella infection and for prevention of, and recovery from, zoster. The level of cellular immunity correlates with disease severity during acute varicella. Cytotoxic T cells that lyse virus-infected cells are present by two to three days after the onset of the rash of varicella. Cell-mediated immunity, as measured by lymphocyte proliferative response, is directed against cells expressing glycoproteins gE, gB, gH, gI, and gC, IE62, IE63, and other viral proteins. Interferon is present in VZV vesicles.

Most varicella infections result in lifelong immunity to reinfection. Second episodes of varicella are rare; these individuals tend to have reduced humoral and cellular immunity to VZV at the time of the second infection. Zoster is associated with a reduction in cellular immunity to VZV that, in the normal host, is partially restored in response to this recurrent infection. Recurrent zoster is uncommon, except in severely immune deficient patients, such as those with AIDS.

Diagnosis

Varicella and zoster are usually diagnosed clinically. Varicella typically presents with a disseminated rash with lesions present in various stages including vesicles, pustules, and crusts. Zoster usually presents with a similar rash, but with pain and skin lesions that are localized to one or two dermatomes and does not cross the midline. In cases where the diagnosis is less certain, scraping of skin lesions to unroof the vesicles or pustules followed by PCR for VZV DNA is diagnostic. PCR for VZV DNA is much more sensitive than staining the fluid for viral antigen or trying to culture the virus from the fluid. Detection of multinucleated giant cells in the specimen is consistent with a herpesvirus infection; however, it is not specific for VZV. In rare cases of zoster which present with visceral disease before, or in the absence of a rash, PCR of blood for VZV is used to make a diagnosis. Patients who present with stroke and are suspected of having VZV vasculopathy are usually diagnosed by PCR for VZV DNA in the cerebrospinal fluid or by demonstrating a significantly higher level of VZV antibody in the cerebrospinal fluid relative to the level of antibody in the blood.

Testing for VZV antibody in the blood is indicative of prior infection but is not useful for diagnosis of acute disease. Testing for antibody can be helpful in immunocompromised persons exposed to varicella or zoster, to determine if they have never been infected in the past, and thereby determining whether post-exposure prophylaxis is needed (see below).

Treatment

Treatment of varicella includes acetaminophen for fever and anti-pruritics when itching is problematic. While acyclovir is licensed for treatment of varicella, it is not recommended by the American Academy of Pediatrics for otherwise healthy children under age 13 since it shortens the duration of disease by only 1 day. Nonetheless, some physicians recommend acyclovir since it may reduce rare complications in children. Acyclovir is recommended for persons at higher risk of complications of varicella including persons 13 years or older, infants born to mothers with varicella at the time of delivery, immunocompromised children (including those on steroid therapy) regardless of their age, children with chronic pulmonary or skin disease, those taking aspirin (at higher risk of Reye's syndrome), and persons already with complications associated with varicella. It is often given to pregnant women with varicella in their third trimester who have an increased risk of severe disease. Acyclovir decreases visceral dissemination of varicella in the immunocompromised host. Oral valacyclovir has better bioavailability than oral acyclovir and is approved for treatment of varicella in persons aged 2 years or older; it is not recommended for pregnant women.

Acyclovir also reduces the risk of VZV spread of trigeminal zoster to the eye and modestly shortens the duration of zoster symptoms in the normal host. Analogs of acyclovir, such as famciclovir and valacyclovir, result in higher levels of antiviral activity than oral acyclovir and have been licensed for oral therapy of zoster in the United States. Treatment is given within 3 days of the onset of the rash, as well as if new lesions are continuing to occur. Antiviral therapy is recommended for persons with zoster who are age 50 or above, who are immune compromised, who have disease involving the eye or face, or have other complications associated with zoster, and those with moderate to severe pain. Acyclovir shortens the duration of lesions and reduces dissemination of the virus in immunocompromised patients; acyclovir reduces zoster-associated pain but does not reduce the rate of post herpetic neuralgia. Corticosteroids, when used early during zoster, reduce acute pain. Herpes zoster, particularly in elderly patients may lead to prolonged and severe postherpetic neuralgia.

Treatment of postherpetic neuralgia is difficult and often unsatisfactory, but many patients experience improvement with gabapentin, pregabalin, tricyclic antidepressant drugs like amitriptyline, or lidocaine patches. Acyclovir-resistant strains of VZV have been reported in patients with AIDS and transplant recipients; these infections are best treated with foscarnet.

Prevention

Passage of wild-type VZV in cell culture by Takahashi in 1974 led to attenuation of the virus and changes in its temperature sensitivity and infectivity for certain cell lines. Comparison of the complete nucleotide sequence of the Oka vaccine strain with its parental virus shows 42 nucleotide and 20 amino acid differences. Multiple nucleotide changes in several genes are responsible for attenuation of the vaccine for growth in skin. The Oka vaccine strain can be distinguished from wild-type strains by differences in restriction endonuclease patterns.

The live, attenuated varicella vaccine (Oka strain) was licensed in the United States in 1995 and is recommended for vaccination of healthy children age 1–12 and for older persons who have not been infected with the virus. Two doses of vaccine are recommended, the first at one year of age and the second at 4–6 years of age. The vaccine has also been combined as a single injection with the measles, mumps, and rubella vaccines. Varicella vaccine is about 90% effective in preventing varicella and more than 95% effective in preventing moderate or severe varicella. A mild papular rash, usually near the injection site, may occur during the first two weeks after vaccination. It is usually mild but can be severe if the vaccine is given to patients experiencing periods of profound cellular immune impairment. The live vaccine virus establishes neural latency and can reactivate. Zoster has been reported in vaccine recipients, especially those who are immunocompromised, but the rate appears to be lower than that following natural infection. The varicella vaccine virus has very rarely been transmitted to other persons, and only when the vaccine recipient has developed a rash. The vaccine is not given to persons on high dose immunosuppression, persons with hematologic malignancies, HIV patients with CD4 counts below 200 cell/ul, or pregnant women.

Two vaccines are available to prevent zoster. A high titer formulation of the live attenuated Oka vaccine virus is given as a single dose and is approved for persons age 50 or older. The vaccine reduces the frequency of zoster and postherpetic neuralgia by 51% and 67%, respectively, in healthy persons ≥ 60 years old. The vaccine retains its efficacy to prevent postherpetic neuralgia in persons older than age 70 but is less effective to prevent zoster in these older persons. The most common side effects are mild injection site reactions; no significant severe adverse reactions have been attributed to the vaccine. The vaccine is not given to persons on high dose immunosuppression, persons with hematologic malignancies, or HIV patients with CD4 counts below 200 cell/ul.

In 2017 a two-dose adjuvanted subunit vaccine was approved for prevention of zoster. The vaccine consists of VZV glycoprotein E and AS01_B adjuvant (saponin and monophosphoryl lipid A) in a liposomal formation. The subunit vaccine is 97% effective to prevent zoster in persons age 50–69, and 91% effective for preventing zoster and 89% effective for preventing postherpetic neuralgia in persons age 70 and older. The adjuvanted subunit has more side effects than the live attenuated zoster vaccine; 17% of persons receiving the subunit vaccine had systemic or injection site reactions sufficiently severe to temporarily prevent normal activities. These vaccine reactions included pain, redness, or swelling at the vaccination site, or generalized muscle pains, fever, headache, fatigue, or gastrointestinal symptoms. In 2018 the Advisory Committee for Immunization Practices at the Centers for Disease Control and Prevention gave a preferential recommendation for the zoster subunit vaccine over the live attenuated zoster vaccine for prevention of zoster and its complications.

Prevention of varicella can be achieved by restricting exposure to persons with varicella or zoster until all lesions have crusted. This is particularly important for hospital workers and immune deficient patients.

For persons exposed to varicella, three options are available- vaccination with the live, attenuated vaccine, giving immunoglobulin prophylaxis, or providing acyclovir to persons soon after exposure. Vaccination is preferred for postexposure prophylaxis of non-immunocompromised persons who were exposed to varicella within the prior three days. Vaccination is 70%–90% effective in this situation. Varicella-zoster immune globulin (VariZIG), prepared from human plasma, prevents or attenuates varicella in 90% of seronegative persons if given within four days of exposure to the virus. Recent recommendations indicate that it can be used up to 10 days after exposure to varicella or zoster. VariZIG has no effect in modifying zoster. VariZIG is recommended for individuals at high risk for developing severe varicella and is indicated in those (a) with recent, close contact to patients with varicella or zoster, (b) who are susceptible to varicella, and (c) who fall in a high-risk category. The latter include premature or certain newborn infants, pregnant women, and patients with congenital or acquired cellular immunodeficiencies. Oral acyclovir is an alternative to VariZIG in high risk VZV seronegative persons exposed to varicella or zoster. Acyclovir is estimated to be about 85% effective in preventing varicella when given for one week beginning 7–9 days after exposure.

Further Reading

- Abendroth, A., Arvin, A.M., Moflat, J.F., 2000. *Varicella-Zoster Virus*. Current Topics in Microbiology and Immunology. 342. New York: Springer.
- Arvin, A.M., Gilden, D., 2013. Varicella-zoster virus. In: Knipe, D.M., Howley, P.M., Cohen, J.I. *et al.* (Eds.), *Fields Virology*, sixth ed. Philadelphia: Lippincott Williams & Wilkins, pp. 2015–2057.
- Cohen, J.I., 2013. Clinical practice: Herpes zoster. *The New England Journal of Medicine* 369, 255–263.
- Cunningham, A.L., Lal, H., Kovac, M., *et al.*, 2016. Efficacy of the herpes zoster subunit vaccine in adults 70 years of age or older. *The New England Journal of Medicine* 375, 1019–1032.

- Depledge, D.P., Ouwendijk, W.J.D., Sadaoka, T., *et al.*, 2018. A spliced latency-associated VZV transcript maps antisense to the viral transactivator gene 61. *Nature Communications* 9, 1167.
- Dooling, K.L., Guo, A., Patel, M., *et al.*, 2018. Recommendations of the advisory committee on immunization practices for use of herpes zoster vaccines. *Morbidity and Mortality Weekly Report* 67, 103–108.
- Gershon, A.A., Breuer, J., Cohen, J.I., *et al.*, 2015. Varicella zoster virus infection. *Nature Reviews Disease Primers* 1, 15016.
- Jensen, N.J., Rivaller, P., Tseng, H.F., *et al.*, 2017. Revisiting the genotyping scheme for varicella-zoster viruses based on whole-genome comparisons. *Journal of General Virology* 98, 1434–1438.
- Leung, J., Harpaz, R., 2016. Impact of the maturing varicella vaccination program on varicella and related outcomes in the United States: 1994–2012. *Journal of the Pediatric Infectious Diseases Society* 5, 395–402.
- Markus, A., Golani, L., Ojha, N.K., *et al.*, 2017. Varicella-Zoster virus expresses multiple small noncoding RNAs. *Journal of Virology* 91, e01710–e01717. (9).
- Ogunjimi, B., Zhang, S.Y., Sørensen, K.B., *et al.*, 2017. Inborn errors in RNA polymerase III underlie severe varicella zoster virus infections. *Journal of Clinical Investigation* 127, 3543–3556.
- Oliver, S.L., Yang, E., Arvin, A.M., 2016. Varicella-Zoster virus glycoproteins: Entry, replication, and pathogenesis. *Current Clinical Microbiology Reports* 3, 204–215.
- Sadaoka, T., Depledge, D.P., Rajbhandari, L., *et al.*, 2016. In vitro system using human neurons demonstrates that varicella-zoster vaccine virus is impaired for reactivation, but not latency. *Proceedings of the National Academy of Sciences of the United States of America* 113, E2403–E2412.
- Warren-Gash, C., Forbes, H., Breuer, J., 2017. Varicella and herpes zoster vaccine development: Lessons learned. *Expert Review of Vaccines* 16, 1191–1201.
- Xing, Y., Oliver, S.L., Nguyen, T., *et al.*, 2015. A site of varicella-zoster virus vulnerability identified by structural studies of neutralizing antibodies bound to the glycoprotein complex gHgL. *Proceedings of the National Academy of Sciences of the United States of America* 112, 6056–6061.
- Zerboni, L., Sen, N., Oliver, S.L., Arvin, A.M., 2014. Molecular mechanisms of varicella zoster virus pathogenesis. *Nature Reviews Microbiology* 12, 197–210.

Variola and Monkeypox Viruses (*Poxviridae*)

Lalita Priyamvada and Panayampalli S Satheshkumar, Centers for Disease Control and Prevention, Atlanta, GA, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S. Parker, D.A. Schultz, H. Meyer, R.M. Buller, Smallpox and Monkeypox Viruses, In Reference Module in Biomedical Sciences, Elsevier Inc., 2014, doi:10.1016/B978-0-12-801238-3.02665-9.

Glossary

Exanthem Eruptive skin rash.

Papule Small, solid, elevated lesion of the skin that is often inflammatory.

Pock Pitted scar or mark caused by a pustule or pimple.

Prodrome Early sign or symptom of illness.

Pustule Small elevated pus-containing lesion of skin.

Sequelae Pathological condition occurring as a result of disease.

Vesicle Small cavity within the epidermis containing serous fluid.

History

Variola virus (VARV), the causative agent of smallpox, is perhaps the singular most devastating human pathogen in history. Since the earliest evidence of smallpox infections preserved in Egyptian mummies dating back to 1200 BCE, through thousands of years later, human beings have battled this highly virulent and contagious virus in numerous outbreaks worldwide. Entire communities have perished during smallpox epidemics, including an estimated 300–500 million individuals in the 20th century alone. Those who survived infection suffered disfiguring residual scarring or blindness. Between the 10th and 18th centuries, efforts to control the spread of smallpox mainly involved quarantine, and intranasal and intradermal variolation using powdered or pulverized smallpox lesions. Thereafter, following Jenner's inoculation studies using cowpox virus, vaccination emerged as a safer, more effective alternative in the fight against smallpox. In the 20th century, the generation of freeze-dried formulations extended the shelf life of vaccines, which allowed for greater distribution and reach worldwide. Widespread public health campaigns that involved ring vaccination and mass vaccination strategies were deployed around the globe. As a result of these extensive efforts, the number of new smallpox cases worldwide dwindled. The last naturally-occurring cases of smallpox were recorded in the Indian subcontinent in 1975 and in the Horn of Africa in 1977. In 1980, the WHO declared that smallpox had been eradicated worldwide.

In contrast to the ancient descriptions of VARV, monkeypox (MPXV) was discovered relatively recently in 1958. MPXV was first discovered when monkeys that were part of a polio research study developed pox-like vesiculopapular skin lesions. Although the clinical symptoms observed were virtually identical to smallpox, serial passage of the isolated virus in mice and rabbit skin distinguished the virus from VARV. The first human case of monkeypox was recorded a decade later in 1970 when a 9-year old boy from Zaire (now Democratic Republic of Congo, DRC) presented with pox lesions in a region that had previously been cleared of smallpox. It is highly probable that human MPXV cases occurred prior to 1970, but were previously misidentified as VARV infections due to the clinical similarities and geographical overlap between the two viruses. Although smallpox was eradicated in 1980, isolated MPXV cases and larger outbreaks have been reported over the past few decades. These have occurred mainly in Africa, but exported cases have also reached North America, Europe and Asia. A zoonotic disease, a majority of monkeypox cases to date have been caused by direct or indirect contact with infected rodents and primates. Today, MPXV remains endemic in the Congo Basin, and has taken the place of VARV as the most critical orthopoxvirus for human health.

Classification and Structural Morphology

VARV and MPXV are both members of the *Poxviridae* family of large double-stranded DNA viruses. Within the poxvirus family, they belong to the sub-family *Chordopoxvirinae*, which constitutes poxviruses that infect vertebrates, and are further classified into the Orthopoxvirus (OPXV) genus based on serological cross-reactivity with other species within the genus. Like other OPXVs, VARV and MPXV produce large brick-shaped particles that are approximately 140–260 nm in diameter and 220–450 nm long, containing a linear DNA genome approximately 200 kb in length. The genome, enclosed within an enveloped capsid, is monopartite and encodes around 200 genes. As is a classic feature of OPXVs, both VARV and MPXV exist in two distinct forms during viral replication; these forms are named intracellular mature virus (MV) and enveloped virus (EV). Although the same viral genome is shared between the two forms, the particles differ in the number of membranes surrounding the core. While MVs possess a singular viral membrane, EVs undergo an additional double-membrane wrapping step during the viral life cycle resulting in the presence of two viral membranes encapsulating the genome. As a consequence, MVs and EVs express different viral surface proteins, which further results in unique antigenic signatures differentiating the two forms of virions.

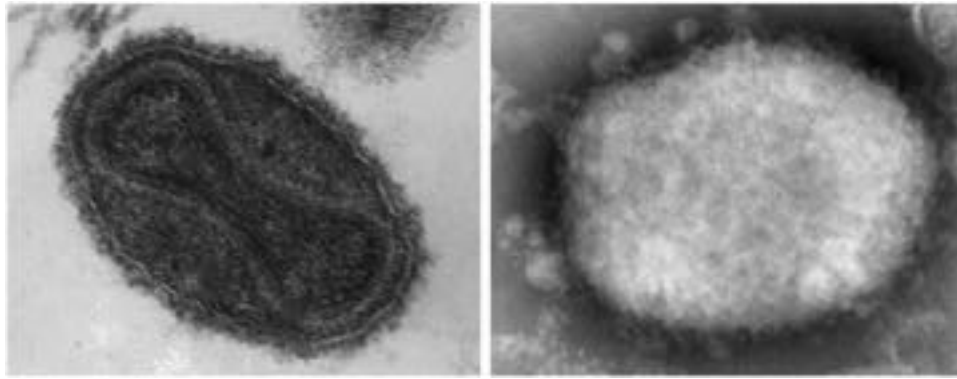


Fig. 1 Highly magnified images of variola and monkeypox viruses obtained using transmission electron microscopy. Image on left shows a variola virion with an internal dumbbell-shaped core. Image on right shows the “M” (mulberry type) monkeypox virion in human vesicular fluid. Photographs obtained from the CDC Public Health Image library, courtesy of Dr. Fred Murphy and Sylvia Whitfield (VARV), and Charles D. Humphrey, Tiara Morehead, and Russell Regnery (MPXV).

Electron microscopy (EM) images have revealed morphological details of the exterior and interior surfaces of the VARV and MPXV virions (**Fig. 1**). The outer surface of VARV and MPXV MV particles consists of a lipoprotein membrane with a knobby, mulberry-like appearance due to the presence of irregularly-shaped surface tubules. Enclosed within is a core wall surrounding the dumbbell shaped viral core, as well as two amorphous structures named lateral bodies. The viral core contains tightly compressed nucleoprotein and a DNA genome. The characteristic morphology of VARV and MPXV particles has made EM-based visualization a valuable confirmatory technique for VARV and MPXV infection diagnosis, both historically during the smallpox eradication campaign, as well as more contemporarily in recent monkeypox outbreaks.

Viral Life Cycle

Much of what is understood about the viral life cycle of OPXVs like VARV and MPXV stems from studies of the prototypic OPXV, vaccinia virus (VACV). The steps involved in viral entry and replication of OPXVs have been described in extensive detail elsewhere. Briefly, particles can enter cells either by fusion directly at the plasma membrane or through macropinocytosis. Proteins on the surface of EVs bind to host attachment receptors resulting in the rupture of the EV membrane. This exposes the MV membrane which expresses the entry-fusion complex required for virus-host cell fusion. Viral fusion releases the internal core into the cytoplasm where viral enzymes packaged within the core initiate early transcription and translation. Early proteins include immunomodulatory factors, as well as viral enzymes that facilitate DNA replication. As DNA replication occurs, intermediate and subsequently late genes are transcribed and expressed under regulation by stage-specific viral transcription factors. Late translation products include structural proteins required for virion assembly.

Following infection, all the steps involved in viral replication cycles of VARV and MPXV occur entirely within the cytoplasm. Particle assembly occurs within electron dense regions termed “virus factories” and begins with the formation of immature virions consisting of a newly generated DNA core and a singular viral membrane. The immature virions undergo maturation through proteolytic processing of core proteins mediated by viral encoded protease to generate MVs. Although entirely infectious, most of the MVs remain in the cytoplasm, and only a proportion undergo an additional wrapping event resulting in a double-membrane enveloped wrapped virus (WV). The WV is an intermediate between the MV and the cell-associated EV (CEV) or extracellular EV (EEV) forms, both of which are responsible for a majority of viral dissemination within a host during infection. Egress of the enveloped viral forms occurs through membrane fusion. Viral egress triggers the formation of actin tails which can propel viral particles from an infected cell towards neighboring cells. CEVs and EEVs are launched from the tips of microvilli protruding out of cell surfaces toward other cells within the microenvironment, at which point another round of infection can occur. MVs largely remain trapped within infected cells until cell lysis, following which they are released and are able to infect other cells. A simplified illustration of the OPXV life cycle is provided in **Fig. 2**.

Pathogenesis

The pathogenesis of VARV and MPXV are virtually identical with one important exception at the point of entry. While VARV was a solely human pathogen, monkeypox is a zoonotic disease. Therefore, the point of entry for most VARV infections during historic smallpox outbreaks occurred within the respiratory mucosa during person-to-person viral transmission. In contrast, a majority of human MPXV cases are caused by animal exposure and likely require contact with lesions on the skin or oral mucous membranes. After infection has occurred, both viruses travel to the closest lymph nodes through the lymphatic system and replicate in lymphoid tissue.

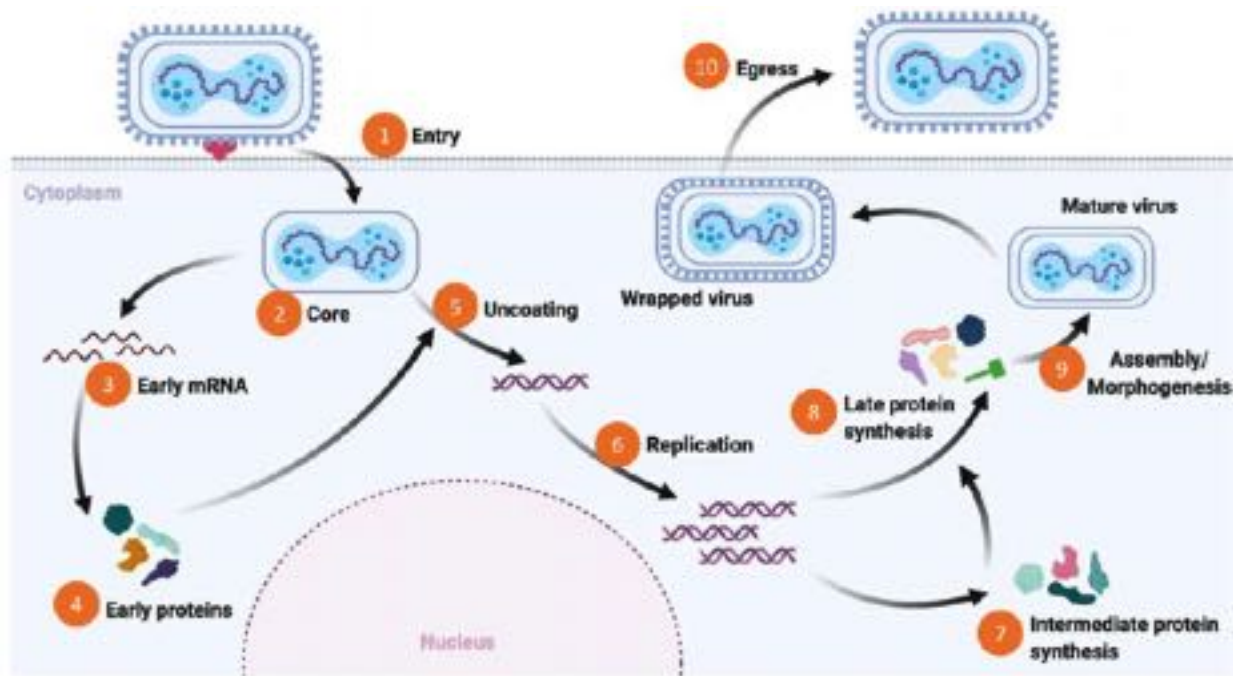


Fig. 2 Simplified illustration of the OPXV life cycle.

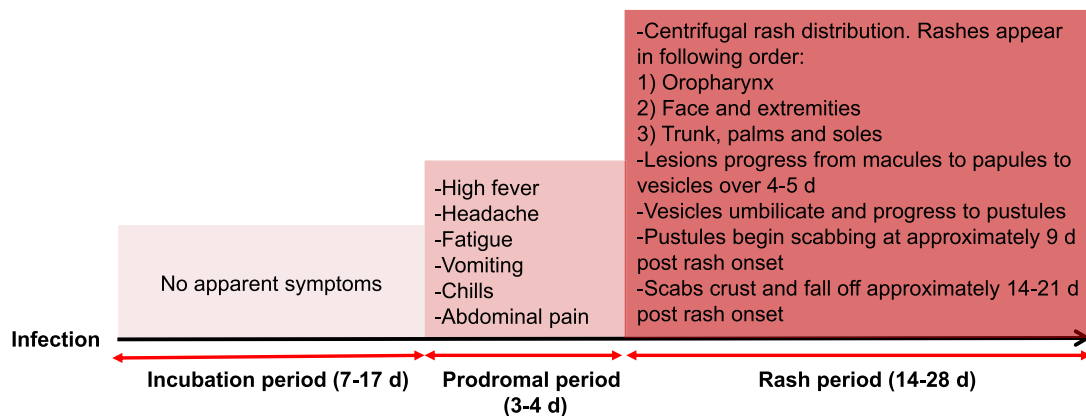


Fig. 3 Illustration showing the timeline of clinical stages and symptoms in smallpox and monkeypox disease. Days abbreviated as “d”.

Primary viremia is initiated as the virus enters the blood stream and migrates to other sites within the body. Between days 3–14 after infection, the virus spreads to other lymph nodes, the bone marrow, spleen, kidneys and liver. Secondary viremia occurs as the virus disseminates to the oral mucosa and skin. Viral replication in the blood vessels of the dermis causes the formation of lesions that eventually spread to the epidermis. This results in the appearance of pus-filled pocks that are characteristic of smallpox and monkeypox disease. With virus rapidly spreading throughout the body, a systemic inflammatory response is generated causing toxemia, which can lead to death.

Clinical Features

As with viral pathogenesis, the clinical features of the two orthopoxviral diseases are nearly indistinguishable but for a few differences. As shown in **Fig. 3**, human MPXV and VARV infection begin with an incubation period lasting approximately 7–17 days during which no clinical symptoms are observed. This is followed by a prodromal period characterized by high fever, malaise, fatigue and headaches, among other symptoms. In case of smallpox, individuals often suffer very severe fevers with body temperatures exceeding 40°C. During MPXV infection, fevers typically range between 38.5–40.5°C. Another notable difference during this period is the lack of severe lymphadenopathy in smallpox cases, which is often observed in human MPXV infections.

The prodromal period is followed by a decline in fever and the onset of rash lesions. Rashes usually appear centrifugally, starting on the throat and oral cavity which can cause difficulties in eating and drinking. Rashes then appear on the face, arms and legs, and finally on the trunk, palms and soles. Over the first week of the rash period, lesions progress from macules, to papules, to vesicles to pustules. The lesions in MPXV and VARV infections are hard and deep, progress slowly and can cause significant pain to infected individuals until they begin to crust. The rash period ends as pustules form scabs, and the scabs crust and fall off. Depending upon the severity of symptoms, an infected individual may succumb to disease during or after the rash period, or survive with sequelae including but not limited to scarring, vision loss or blindness and respiratory complications.

The clinical course and severity of disease in MPXV and VARV infections can depend on several factors such as age, immune status, infecting dose, route of infection and the infecting virus strain. The differences in disease severity after VARV infection are illustrated in Fig. 4.

Epidemiology

Smallpox

For thousands of years until its eradication in 1980, smallpox outbreaks caused widespread mortality and devastation. After the earliest descriptions of pox-like rashes on Egyptian mummies, independent records of human disease resembling smallpox appeared in China, India and Western Asia between the 4th and 10th centuries AD. Through colonization, the virus was brought to Europe and northern Africa, and later to the Americas and southern Africa in the 16th and 17th centuries. With the exception of Australasia, smallpox had become endemic globally by the turn of the 18th century. Through the next two centuries, vaccination efforts strengthened and brought about a large decline in the number of outbreaks worldwide. Nonetheless, smallpox remained a persistent problem for densely populated regions in Africa and Asia through the second half of the 20th century. Through the introduction of the Intensified Smallpox Eradication Program in 1967, and active engagement by global, national and local leaders in widespread immunization campaigns, additional countries eliminated smallpox over the next decade. In 1977, the last known natural case of smallpox occurred in Somalia. In 1980, the WHO declared that smallpox had been eradicated worldwide, and no new VARV infections have been recorded since.

Monkeypox

The first known case of human monkeypox was detected in Zaire (now DRC) in 1970. Over the next decade, an additional 54 cases were recorded in Western and Central Africa, 44 of which occurred in DRC. Interestingly, while no deaths were reported in West



Fig. 4 Clinical and epidemiological features of disease following VARV infection. Images showing Malignant, Ordinary and Modified smallpox, and Alastrim obtained from the CDC Public Health Image Library. Photographs by Dr. Dean, Dr. Robinson and Dr. Noble. Source of Hemorrhagic smallpox image: Herrlich, A., Mayr, A., Munz, E., Rodenwaldt, E., 1967. Die Pocken; Erreger, Epidemiologie und klinisches Bild, second ed., Stuttgart: Thieme.

Africa due to monkeypox during this time, the case mortality rate in DRC was 21%. Later studies demonstrated two distinct clades of MPXV based on genomic alignments and orthologous gene clusters, which were subsequently named for the regions in which they were isolated: Congo Basin and West Africa.

Despite the eradication of smallpox in 1980, monkeypox persisted in the Congo Basin. Over three hundred cases were reported between 1981–1986, partly due to increased interest in monkeypox surveillance in the region. Sero-surveillance data indicated that prior smallpox vaccination conferred >85% protection from MPXV during this period, and case-control studies revealed a large majority of infections occurred due to contact with wildlife. As smallpox vaccine-mediated immunity waned, several human monkeypox outbreaks occurred in DRC over the subsequent decades. The years 1996–1997 witnessed a prolonged outbreak of 88 possible cases, following which >2500 cases of suspected human MPXV infection were reported during 1998–2004.

The first human monkeypox outbreak outside Africa occurred in 2003 in the United States, as a result of the importation of infected animals from Ghana. It is hypothesized that West African rodents carrying MPXV transmitted the virus to indigenous rodents including prairie dogs, which eventually spread monkeypox to humans. In all, 37 confirmed human infections were identified, and no deaths were reported. In contrast, an MPXV outbreak occurring the same year in the Republic of Congo (ROC) involved one death and one case of severe sequelae out of a case total of 11 infected persons. Other significant outbreaks have occurred since, the first in DRC involving over 750 cases in 2006–2007, and more recently in Nigeria in 2017.

The 2017 Nigerian monkeypox outbreak has been the largest to date in the West African nation, occurring nearly 40 years after its last confirmed human case. To date, over 200 confirmed and suspected monkeypox cases have been identified, although the total disease burden remains unknown. In 2018 and 2019, a total of six MPXV cases were reported in the United Kingdom (4), Israel (1) and Singapore (1), all involving travelers from Nigeria who fell ill with monkeypox upon arrival. While five of these cases originated in Nigeria, the sixth was a nosocomial infection of a healthcare worker in the United Kingdom. In conjunction, ROC reported over 40 suspected human monkeypox infections in 2017–2018. The recent re-emergence of monkeypox in Western and Central Africa and increase in travel-related cases have elevated the importance of monkeypox as a growing public health threat to humans.

Diagnosis

The accurate diagnosis of OPXV infections, including VARV and MPXV, is complicated by a few factors. First, poxviral and non-poxviral exanthema can appear very similar, making it exceptionally challenging to identify the causative pathogen purely by clinical evaluation. This is illustrated in [Fig. 5](#), which shows a side-by-side comparison of pustular and maculopapular lesions caused by VARV and varicella zoster virus (causative agent of chickenpox and shingles). Second, prior vaccination with the smallpox vaccine may confound the serology-based diagnosis of an ongoing OPXV infection. Third, antigenic and structural similarities between VARV, MPXV and other OPXVs requires the use of advanced laboratory techniques and exclusive resources to differentiate between these viruses. As such, a combination of detailed case history including prior vaccination records, clinical examination and laboratory assays is critical for the accurate and efficient diagnosis of VARV and MPXV.

Due to the eradication of smallpox, clinical and laboratory testing to confirm VARV infection do not routinely occur in the present day. In case of MPXV, the symptoms associated with the prodromal and rash phases listed in [Fig. 3](#), as well as the presence of swollen lymph nodes are typical clinical features of monkeypox disease. Importantly, the physical characteristics of the lesions (monomorphy versus pleomorphy) and their distribution on the body (centrifugal versus centripetal) can help to distinguish between OPXVs and other non-poxviral exanthematous infections. Laboratory-based tests to distinguish between OPXV and non-OPXV infections include electron microscopy negative staining, immunohistochemistry, serological assays, and viral isolation, culture and amplification. Of the aforementioned techniques, only virus isolation and DNA-based detection can parse MPXV apart from other OPXVs at this time. As a result, the current laboratory criteria for confirmation of monkeypox require the PCR-based detection of MPXV DNA in clinical samples or a positive MPXV culture from virus isolated from a clinical specimen.



Fig. 5 Image showing lesions caused by VARV (left panel) and varicella zoster virus (right panel). The lesions have similar morphologies that appear undistinguishable to an untrained eye. Images obtained from the CDC Public Health Image Library. Photographs by Dr. Noble and Dr. K.L. Herrmann.

Prevention

The earliest efforts in smallpox prevention can be dated back to 10th century China, where dried pustular lesions from infected persons were applied to the nostrils of uninfected persons. This practice was observed to sometimes protect the recipients from smallpox. In the 17th and 18th centuries, “pox-laden” blankets were wrapped around uninfected children in India, with the intent of transmitting mild disease and conferring protective immunity. The instillation of smallpox pustular matter into the skin of uninfected persons, also known as variolation, was also practiced during this time in China, India and Turkey. Variolation was introduced to England in the 18th century and subsequently to multiple English colonies. With the introduction of variolation in new areas, many communities experienced a decline in smallpox incidence. However, safer alternatives were imperative, as variolation itself caused a case mortality rate of 0.5%–2%.

In 1796, Dr. Edward Jenner injected material from the vesicle of a milkmaid with cowpox into the arm of a smallpox non-immune boy. He then infected the boy with VARV and observed that the boy did not develop smallpox disease. This initial challenge experiment, an egregious violation of ethics in human research by today's standard, was the foundation of modern-day vaccination. Far safer than variolation, vaccination quickly became widespread in the 19th century. By the onset of the 20th century, a number of Western nations had entirely eliminated smallpox from their population through mass and ring vaccinations. Interestingly, the smallpox vaccines used during the 20th century WHO smallpox programs, also known as first generation smallpox vaccines, were formulated using VACV, not cowpox virus. The origin of VACV is unclear, and the cause for a switch from cowpox virus to VACV, if one even occurred is not well understood. The mostly widely used and studied first generation smallpox vaccine was Dryvax, based on the NYCBH (New York City Board of Health) strain of VACV. Others include Lister strain-based vaccines Pourquier and Lancy-Vaxina Berna. First generation vaccines were largely propagated in animal skin and lymph and although efficacious in controlling smallpox incidence, were associated with a wide range of serious post-vaccinial complications.

Second generation smallpox vaccines, including ACAM2000, Elstree-BN and VV Lister/CE, were manufactured using modern cell culture in accordance with Good Manufacturing Practices (GMP). These vaccines were formulated with either the same vaccine strains as first generation vaccines or their purified clonal variants. ACAM2000 was licensed in the United States in 2007, available for use under limited circumstances and part of the Strategic National Stockpile (SNS). Given that smallpox is no longer an imminent threat, the vaccine is only administered to individuals with a high risk of exposure, such as certain military, healthcare and laboratory personnel. ACAM2000 is a live replication-competent vaccine derived from a less virulent subpopulation of Dryvax vaccine NYCBH strain. The vaccination schedule requires a singular dose of vaccine administered percutaneously using a bifurcated needle as pictured in Fig. 6. Individuals with certain health conditions, including immune deficiency disorders, cardiac disease, chronic skin conditions such as eczema, as well as infants and pregnant women cannot receive ACAM2000 as a pre-exposure vaccine as its effectiveness relies on the local replication of VACV at the site of vaccination.

Third generation vaccines were developed as a response to the need for safer vaccines with fewer side effects. Towards the end of the smallpox eradication era, the Lister strain based LC16m8 and Modified Vaccinia Ankara (MVA) were evaluated for their safety and immunogenicity in several clinical trials. Both vaccines showed significantly safer post-vaccination health outcomes in their respective studies compared to first and second generation vaccines. The most recent smallpox vaccine to have received licensure and have been added to the SNS is the MVA-based third generation vaccine named Jynneos (also known as Imvanex, Imvamune or MVA-BN). Jynneos is a replication-deficient (in mammalian cells) vaccinia virus-based vaccine that was approved by the FDA for smallpox and monkeypox prevention in 2019. The approval of Jynneos is noteworthy for the following two reasons: first, Jynneos was shown to cause no adverse events even in recipients with immune deficiencies or exfoliative skin conditions. Since ACAM2000 is contraindicated for these groups of individuals, Jynneos would be a safe way to provide these susceptible groups with protective immunity in the event of a future smallpox outbreak. Second, Jynneos is the first and only FDA-approved monkeypox vaccine. Although the cross-protection and safety testing of Jynneos is currently ongoing in MPXV endemic regions, Jynneos is the only vaccine that is presently approved for prevention of monkeypox disease in the United States.



Fig. 6 Image showing the correct technique for smallpox vaccine administration using a bifurcated needle. Image obtained from the CDC Public Health Image Library. Photograph by James Gathany, 2002.

Treatment

At present, the only FDA-approved treatment for smallpox is the small molecule inhibitor tecovirimat (also known as ST-246, T-POXX). Approved in 2018 by the FDA, tecovirimat targets a conserved VARV gene involved in EV formation. A decrease in the production of EV particles, as caused by tecovirimat, reduces overall viral spread and associated pathology. The anti-VARV potency of tecovirimat has been demonstrated in several in vitro studies. Since VARV is a human pathogen that has been eradicated, other OPXV animal models involving smallpox-like illness have been used to evaluate the antiviral efficacy of tecovirimat, such as ectromelia virus in mice, rabbitpox virus in rabbits and MPXV in primates in accordance with FDA recommendations. Tecovirimat is also well tolerated with few side effects in healthy individuals. As it has not been evaluated in humans experiencing VARV infection, it is unknown whether tecovirimat will be effective in the event of a smallpox outbreak.

There are no approved treatments for monkeypox at this time. If an exposure occurs in the absence of prior smallpox or monkeypox vaccination, a clinical regimen involving vaccine immune globulin (VIG), tecovirimat and other investigational drugs currently in development, including Cidofovir and Brincidofovir (CMX001) may be considered.

Disclaimer

The conclusions, findings, and opinions expressed by authors contributing to this article do not necessarily reflect the official position of the U.S. Department of Health and Human Services or the Centers for Disease Control and Prevention.

Further Reading

- Beer, E.M., Rao, V.B., 2019. A systematic review of the epidemiology of human monkeypox outbreaks and implications for outbreak strategy. *PLOS Neglected Tropical Diseases* 13, e0007791.
- Chan-Tack, K.M., Harrington, P.R., Choi, S.Y., *et al.*, 2019. Assessing a drug for an eradicated human disease: US Food and Drug Administration review of tecovirimat for the treatment of smallpox. *The Lancet Infectious Diseases* 19, e221–e224.
- Damon, I.K., 2013. Poxviruses. In: Knipe, D.M., Howley, P. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Lippincott Williams & Wilkins.
- Durski, K.N., Mccollum, A.M., Nakazawa, Y., *et al.*, 2018. Emergence of monkeypox – West and Central Africa, 1970–2017. *Morbidity and Mortality Weekly Report* 67, 306–310.
- Fenner, F., Henderson, D.A., Arita, I., Jezek, Z., Ladnyi, I.D., 1988. *Smallpox and its Eradication*. Geneva: World Health Organization.
- Jezek, Z., Fenner, F., 1988. Human monkeypox. In: Melnick, J.L. (Ed.), *Monographs in Virology*, vol. 17. Basel: Karger.
- Mccollum, A.M., Damon, I.K., 2014. Human monkeypox. *Clinical Infectious Diseases* 58, 260–267.
- Merchlinsky, M., Albright, A., Olson, V., *et al.*, 2019. The development and approval of tecovirimat (TPOXX(R)), the first antiviral against smallpox. *Antiviral Research* 168, 168–174.
- Moss, B., 2013. Poxviridae: The viruses and their replication. In: Knipe, D.M., Howley, P. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Lippincott Williams & Wilkins.
- Pittman, P.R., Hahn, M., Lee, H.S., *et al.*, 2019. Phase 3 efficacy trial of modified vaccinia ankara as a vaccine against smallpox. *The New England Journal of Medicine* 381, 1897–1908.

Relevant Websites

- <https://www.cdc.gov/smallpox/clinicians/algorithm-protocol.html>
Evaluating Patients for Smallpox: Acute, Generalized Vesicular or Pustular Rash Illness Protocol.
- <https://www.who.int/csr/disease/smallpox/en/-WHO>
I Smallpox.
World Health Organization.
- [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/smallpox-variola-virus-infection-developing-drugs-treatment-or-prevention-guidance-industry-Smallpox-\(Variola-Virus\)-Infection-Developing-Drugs-for-Treatment-or-Prevention-Guidance-for-Industry](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/smallpox-variola-virus-infection-developing-drugs-treatment-or-prevention-guidance-industry-Smallpox-(Variola-Virus)-Infection-Developing-Drugs-for-Treatment-or-Prevention-Guidance-for-Industry)

Vesicular Stomatitis Virus and Bovine Ephemeral Fever Virus (*Rhabdoviridae*)

Peter J Walker, The University of Queensland, St. Lucia, QLD, Australia

Robert B Tesh, The University of Texas Medical Branch, Galveston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of H. Bourhy, A.J. Gubala, R.P. Weir, D.B. Boyle, Animal Rhabdoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00783-4

Glossary

Akt A serine/threonine-specific protein kinase that plays important roles in cellular signaling pathways (also known as protein kinase B).

AP1 A transcription factor (activation protein 1) that regulates cellular gene expression.

Epicardium The outer layer of heart tissue.

Fas A member of the tumor necrosis factor (TNF) receptor superfamily which has a central role in the regulation of apoptosis (also known as Apo1 or CD95).

Hypocalcemia A low level of calcium in the circulating blood.

JNK c-Jun N-terminal kinases are serine/threonine kinases and members of the mitogen-activated protein (MAP) kinase family involved in signal transduction.

Leucopenia A decreased total number of white blood cells in the circulating blood.

NF-κB Transcription factor that plays a key role in regulation of cellular gene expression during viral infection.

Pericardial fluid Fluid within a double-walled sac that contains the heart and the roots of the great blood vessels.

PI3K Phosphoinositide 3-kinases involved in signal transduction by phosphorylating the inositol ring of phosphatidylinositol.

Src Non-receptor tyrosine kinases (src is short for sarcoma) that play key roles in signal transduction from transmembrane receptor tyrosine kinases.

Sternal recumbency Reclined in a position of comfort on the chest bone.

Synovial membranes Connective tissue membranes lining the cavities of the freely movable joints.

Thoracic fluid Fluid in the chest cavity.

TURBS A 'termination upstream ribosome-binding site' that allows stop-start expression from consecutive or partly-overlapping open reading frames.

Classification

Apart from rabies, the animal rhabdoviruses that can cause disease in livestock fall into two genera of the family *Rhabdoviridae*, order *Mononegavirales*. Four viruses assigned to the genus *Vesiculovirus* cause vesicular stomatitis in cattle, horses and swine, and occasionally spill over to cause disease in humans: vesicular stomatitis Indiana virus (VSIV; species *Indiana vesiculovirus*), vesicular stomatitis New Jersey virus (VSNJV; species *New Jersey vesiculovirus*), Cocal virus (COCV; species *Cocal vesiculovirus*) and vesicular stomatitis Alagoas virus (VSAV; species *Alagoas vesiculovirus*). Antigenically, these represent two major serotypes, Indiana and New Jersey. The Indiana serotype is subdivided into four subtypes represented by VSIV (Indiana 1), COCV (Indiana 2), VSAV (Indiana 3) and Morreton virus (MORV; Indiana 4; species *Morreton vesiculovirus*). Antibodies to MORV has been detected in animals and humans but to date the virus has not been associated with disease. The term vesicular stomatitis virus (VSV), as commonly used in the literature, usually refers to vesicular stomatitis Indiana virus and/or vesicular stomatitis New Jersey virus.

Other vesiculoviruses that infect mammals, including humans, are Piry virus (PIRV; species *Piry vesiculovirus*) and Isfahan virus (ISFV; species *Isfahan vesiculovirus*), as well as Chandipura virus (CHPV; species *Chandipura vesiculovirus*) which has been associated with major outbreaks of an influenza-like illness and encephalitis in India. Eight other vesiculoviruses have been isolated from mammals or hematophagous insects but have not been associated with disease. These include American bat vesiculovirus (ABVV; species *American bat vesiculovirus*), Carajas virus (CARV; species *Carajas vesiculovirus*), Jurona virus (JURV; species *Jurona vesiculovirus*), Malpais Spring virus (MSPV; species *Malpais Spring vesiculovirus*), Maraba virus (MARV; species *Maraba vesiculovirus*), Perinet virus (PERV; species *Perinet vesiculovirus*), Radi virus (RADV; species *Radi vesiculovirus*) and Yug Bogdanovac virus (YBV; species *Yug Bogdanovac vesiculovirus*).

In the genus *Ephemerovirus*, bovine ephemeral fever virus (BEFV; species *Bovine fever ephemerovirus*) causes ephemeral fever in cattle and water buffalo. Kotonkan virus (KOTV; species *Kotonkan ephemerovirus*) is the only other member of the genus *Ephemerovirus* to have been associated with clinical disease, both naturally and in experimentally infected cattle. There is no evidence of human infection with either BEFV or KOTV. Ephemeroviruses are distinguishable antigenically in virus neutralization tests using polyclonal hyperimmune sera or ascites fluids. Six other ephemeroviruses have been isolated from healthy cattle and/or hematophagous insects but have not been associated with disease in animals or humans. These include Berrimah virus (BRMV; species *Berrimah ephemerovirus*), Kimberley virus (KIMV; species *Kimberley ephemerovirus*), Adelaide River virus (ARV; species *Adelaide River ephemerovirus*), Obodhiang virus (OBOV; species *Obodhiang ephemerovirus*), Koolpinyah virus (KOOLV; species *Koolpinyah ephemerovirus*) and Yata virus (YATV; species *Yata ephemerovirus*).

Virus Structure

VSV virions are enveloped, bullet-shaped particles, approximately 190 nm in length and 80 nm in diameter (**Fig. 1**). One end is conical and comprises approximately 25% of the length of virions; the cylindrical trunk comprises 75% of the length. Projections, 5–10 nm long and about 3 nm in diameter, form a honeycomb pattern on the surface. Internally, the nucleocapsid is 45 nm in diameter with cross-striations spaced at 4.5–5 nm. Defective-interfering particles with incomplete genomes are formed under some conditions in cell culture and appear as proportionally truncated virions. VSV particles are composed of approximately 74% protein, 20% lipid, 3% RNA and 3% carbohydrate. Cryo-electron microscopy and crystallographic studies have revealed the fine structure of purified VSV virions. The viral RNA is precisely complexed with the nucleoprotein (N) to form an inner helical nucleocapsid with the 3' end located in the conical tip of the bullet and the 5' end is at the base of the trunk. The inner nucleocapsid is encased in a second helical layer formed from the matrix protein (M) which then interacts with the glycoprotein (G) which projects through the lipid envelope to form trimeric spikes on the virion surface. Two other VSV proteins (P and L) are also components of the nucleocapsid but their precise structural locations have not yet been determined.

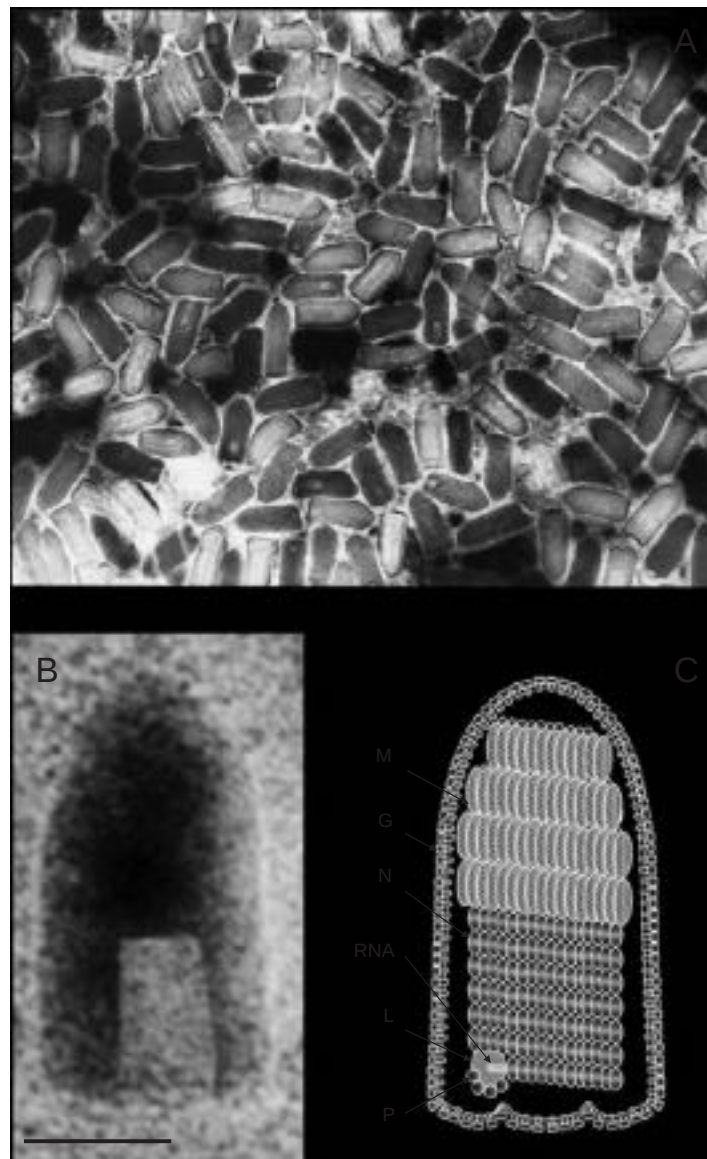


Fig. 1 (A) Negative stain, transmission electron micrograph of VSV virions. (B) Negative stain, transmission electron micrograph of a BEFV virion, illustrating the prominent axial channel. Bar = 50 nm. (C) Schematic illustration of a rhabdovirus virion. Structural proteins N, P, M, G and L, and the (–) ssRNA genome are indicated. The size, relative quantities and precise location of the proteins do not accurately reflect the content in virions. (A) Courtesy F.A. Murphy, CDC Public Health Image Library

BEFV virions are also bullet-shaped (Fig. 1). The virions (140–200 nm × 60–80 nm) display a prominent axial channel intruding from the base, a precisely coiled, helical nucleocapsid with 35 cross-striations at intervals of 4.8 nm, and an envelope with projections extending 8 nm from the surface. Truncated defective-interfering particles are frequently observed following passage of the BEFV in cell culture.

Genomes and Proteins

Like most other rhabdoviruses, vesiculoviruses and ephemeroviruses have non-segmented, negative sense (–), single-stranded (ss) RNA genomes (Fig. 2). Vesiculovirus genomes range in size from approximately 10.7 to 11.3 kb. The genomes feature 3' leader and 5' trailer regions in which the three terminal nucleotides (nt) are complementary (i.e., 3'-UGC...ACG-5', in negative sense). In the VSIV genome, the 3' leader is 47 nt in length and the 5' trailer is 57–59 nt in length. The 3' leader region is followed consecutively by five genes (*N*, *P*, *M*, *G* and *L*), each bounded by conserved transcription initiation (3'-UUGUC...) and transcription termination/polyadenylation (...ACUUUUUUUU-5') sequences. Each gene contains a long open reading frame (ORF) encoding a protein that forms a structural component of virions. In VSIV, the *P* gene also contains an alternative ORF that encodes two small proteins (*C'* and *C''*) from different initiation codons. Other vesiculoviruses have very similar genome organizations to VSIV.

Ephemerovirus genomes are significantly larger than those of vesiculoviruses, ranging in size from approximately 14.5 to 16.1 kb. They are similar in organization to the genomes of vesiculoviruses, including the same conserved complementary terminal trinucleotides, the same conserved transcription initiation and transcription termination sequences, and the five structural protein genes (*N*, *P*, *M*, *G* and *L*). However, ephemeroviruses have 3–5 additional genes located between the *G* gene and *L* gene, including the *G_{NS}* gene and the α gene. In the BEFV genome (14,900 nt), the α gene is followed by the β and γ genes. The α gene contains two consecutive long ORFs, $\alpha 1$ and $\alpha 2$ separated by a single nucleotide. Although not yet demonstrated experimentally, expression of the $\alpha 2$ ORF appears to be possible by a stop-start mechanism involving a conserved 'termination upstream ribosome-binding site' (TURBS) in the mRNA (i.e., 5'-UGGGA-3'). In KOTV, the α gene also contains two consecutive long ORFs ($\alpha 1$ and $\alpha 2$) separated by a single nucleotide and a TURBS sequence upstream of the ORF junction. The KOTV α gene is followed by β , γ and δ genes.

The five vesiculovirus structural proteins (*N*, *P*, *M*, *G* and *L*) are encoded in long ORFs in the corresponding genes. The *N* protein (~47 kDa) is a major component of virions. It associates with genomic RNA to form a ribonucleoprotein (RNP) which serves as the template for replication and transcription. The *N* protein also associates with anti-genomic RNA during replication but does not associate with the viral mRNAs. The *P* protein (~29 kDa) is a phosphoprotein that complexes with *N* to prevent both self-aggregation and binding of *N* to cellular RNA, and then acts as a chaperone to deliver *N* to complex with genomic or anti-genomic RNA. *P* is also a cofactor of the viral polymerase, interacting with the *L* protein to position it on the *N*-RNA template. The *L* protein (~241 kDa) is an enzyme that is responsible for RNA transcription and replication. It performs the functions of the RNA-dependent RNA polymerase (RdRp) as well as 3'-polyadenylation and 5'-capping of mRNA, and protein kinase activities. The *N*, *P* and *L* proteins, together with the viral RNA, are components of the inner helical nucleocapsid in virions. The *M* protein (~26 kDa) is also a major component of virions. It binds to *N* to form an outer helix layer, providing rigidity by clamping adjacent turns of the RNP. In infected cells, *M* is also

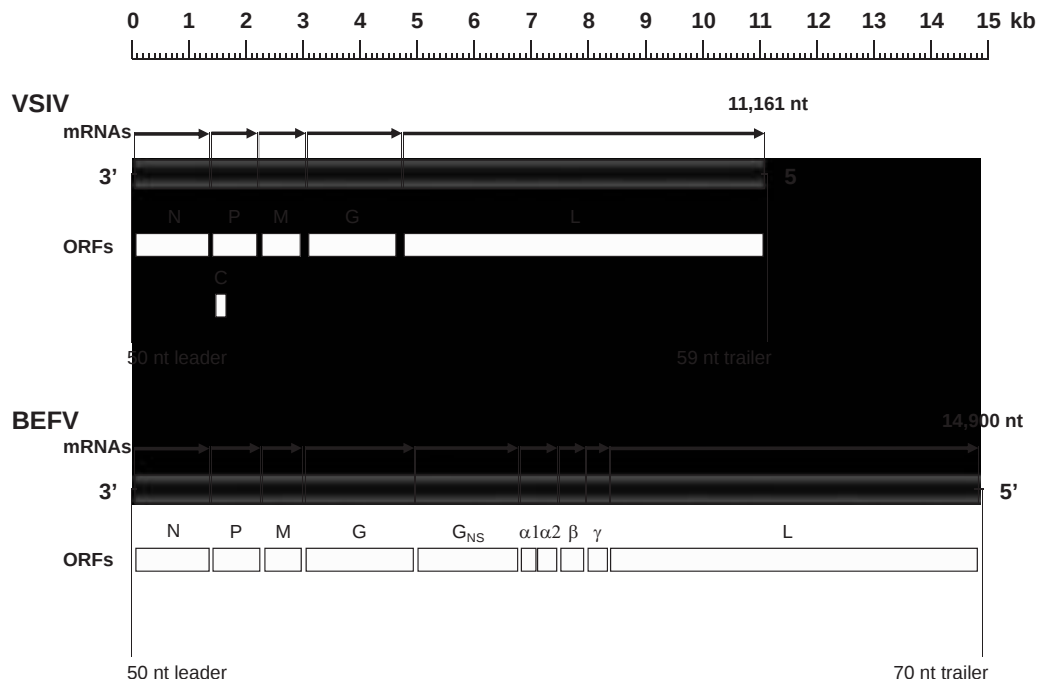


Fig. 2 Schematic illustration of the genome organizations and transcription strategies of VSIV and BEFV. Solid arrows indicate the major transcriptional products. Open boxes indicate long open reading frames (ORFs).

responsible for several cytotoxic effects associated with VSIV infection including cell rounding, regulation of apoptosis, inhibition of host cell transcription and inhibition of nuclear export of host RNAs. The G protein (~67 kDa) is a class I transmembrane glycoprotein which forms trimers that project through the lipid envelope. The C-terminal cytoplasmic domain of G interacts with the nucleocapsid-bound M protein layer to promote budding of virions at cellular membranes. The N-terminal ectodomain of G forms the outer layer of the virion where it mediates binding to cellular receptors, membrane fusion and entry into cells. The G protein also induces virus-neutralizing antibodies. Two small basic protein (C' and C'') encoded in vesiculovirus *P* genes are expressed in infected cells (~8 kDa and ~6.5 kDa, respectively) but they are not required for normal replication in mammalian cell cultures. Their roles are currently not known but they may be involved in pathogenesis of transmission by insect vectors. Evidence of their expression indicates that alternative ORFs as small as 165 nt (encoding proteins of 55 amino acids) are potentially functional in rhabdoviruses.

Ephemeroviruses also contains five structural proteins (N, P, M, G and L) corresponding approximately in size and basic functions to those of vesiculoviruses and most other rhabdoviruses. The ephemerovirus G_{NS} gene, which is located immediately following G gene, also encodes a class I transmembrane glycoprotein (G_{NS}) that shares moderate levels of sequence identity with the BEFV G protein, including a sub-set of the 12 cysteine residues that are variously conserved in animal rhabdovirus G proteins and form disulfide bridges that stabilize their folded structure. BEFV G_{NS} (~90 kDa) is expressed in infected cells but is not present in virions produced either in mammalian cells or insect cells and does not induce virus-neutralizing antibodies or pH-activated cell fusion. The function of G_{NS} is currently unknown. Ephemerovirus $\alpha 1$ proteins have the structural characteristics of class IA viroporins, featuring an N-terminal ectodomain containing multiple large aromatic residues (tyrosine, tryptophan and phenylalanine), a central transmembrane domain, and a C-terminal cytoplasmic domain that is rich in basic residues (lysine, arginine and histidine). BEFV $\alpha 1$ (~10.5 kDa) has been shown to localize in the Golgi complex and increase cellular permeability. The C-terminal cytoplasmic domain of BEFV $\alpha 1$ also interacts specifically with importin $\beta 1$ and importin 7, suggesting that it may also modulate components of nuclear trafficking pathways. Other potential accessory proteins ($\alpha 2$, β , γ and δ) encoded in the ephemerovirus G–L intergenic region have not yet been identified or characterized.

Life Cycle

Attachment, Entry and Uncoating

Most studies of rhabdovirus replication have been conducted by using VSV as a model (Fig. 3). Virions enter cells by endocytosis via clathrin-coated pits in a dynamin-2 dependent manner. Adsorption to cells is mediated by the G protein which binds to its

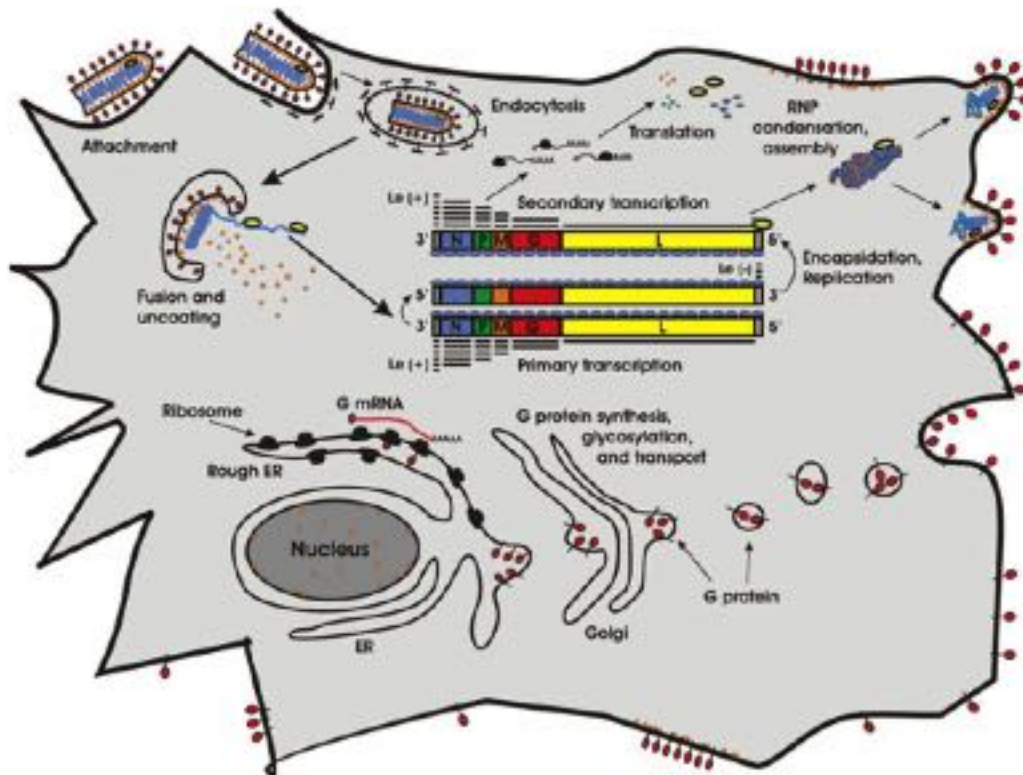


Fig. 3 A schematic illustration of the replication cycle of VSV. The replication cycle (described in the text) is depicted showing attachment of virus to the cell, internalization, release of the viral core into the cytoplasm, primary viral mRNA synthesis, mRNA translation, genomic replication, secondary viral mRNA synthesis, assembly and budding of infectious particles. Illustration courtesy of S.P. Whelan.

receptor on the cell surface. The low-density lipoprotein receptor (LDLR) serves as the primary receptor for VSV but other LDLR family members containing structurally homologous class A cysteine-rich repeats can serve as alternative receptors. Low pH within endocytic vesicles induces a conformational rearrangement of the G protein that exposes the fusion domain and triggers fusion between the endosomal and viral membranes, releasing viral nucleocapsids into the cytosol, either directly or via a secondary fusion involving an intraluminal vesicle. During this process, the M protein is shed from the surface of nucleocapsids and migrates to the nucleus to inhibit host cell transcription. Primary viral transcription occurs directly from the RNP complex comprising the L protein, P protein and the N-RNA template.

BEFV also enters cells via the clathrin-mediated, dynamin-2-dependent endocytic pathway and has been shown to require microtubules as well as Rab 5 and Rab 7, regulatory guanosine triphosphatases that associate with the sorting endosome and are involved in endosomal membrane fusion. To enhance virus entry, BEFV activates the Src-JNK-AP1 and PI3K-Akt-NF- κ B signaling pathways to upregulate clathrin and dynamin 2 expression. Although mediated through the viral G protein, the cellular receptor for BEFV entry is not yet known.

Gene Expression and Replication

VSV transcription occurs exclusively in the cytosol from the RNP complex, generating 5'-capped and 3'-polyadenylated mRNAs. Transcription from the (–) RNA genome is initiated by a sequence-specific promoter at the 3' end, first generating a short (47 nt) leader RNA (Le+) and then commencing mRNA synthesis at a conserved initiation sequence (3'-UUGUC-5') at the start of the N gene. N gene transcription terminates with polyadenylation of the mRNA at another conserved sequence (3'-AUACUUUUUUU-5') at the end of the N gene. The polymerase then recommences transcription at the next conserved initiation sequence and terminates at the next conserved termination/polyadenylation sequence, moving progressively in this fashion to the 5' end of the genome to produce 5 viral mRNAs. There is a decrease in mRNA copy numbers as the polymerase moves along the genome (N > P > M > G > L), apparently due to decreasing efficiency of re-initiation of transcription at each gene junction. The mechanisms of mRNA capping and polyadenylation differ from those employed for host mRNAs. Methylated caps are formed on the nascent 5'-triphosphorylated mRNAs (5'-pppApApCpApG-) in a multi-step process coordinated and catalysed by several active sites in the L protein: (1) although not formally demonstrated, a GTPase appears to produce GDP from GTP; (2) a GDP polyribonucleotidyl transferase links the L protein covalently via a histidine residue (H¹²²⁷) to the RNA to form an intermediate structure (5'-L-pApApCpApG-) and then transfers the 5'-monophosphorylated RNA to GDP (5'-GpppApApCpApG-); and (3) a methyl transferase then sequentially methylates the ribose 2'-O and guanosine N-7 positions in the cap structure (7^mGpppA^mpApCpApG-). Similarly, polyadenylation of VSV mRNAs is catalysed by the L protein in an unconventional pseudo-templated reaction in which the polymerase reiterates transcription in the conserved tract of U residues at the end of each gene. The capped and polyadenylated mRNAs are translated by cytoplasmic polysomes, except for the G mRNA which is translated by membrane-bound polysomes.

Following translation, RNA replication occurs in the cytoplasm, generating full-length (+) RNA and then full-length (–) RNA. Like transcription, replication is conducted by RNP complexes using the N-RNA template. However, during replication the RdRp initiates at the 3' end of the template, reads through transcriptional control signals and continues to terminate at the 5' end. Newly synthesized N, P and L proteins are required to simultaneously encapsidate the nascent antigenomic RNA and then genomic RNAs, ensuring that they are never naked in infected cells. Encapsidation is thought to protect the genome and antigenome from degradation and detection defensive responses by host cells. Nascent RNP complexes containing (–) sense genomic RNA undergo rounds of secondary transcription. Switching between transcription and replication is not well understood but involves the increasing accumulation of N protein to encapsulate nascent genomic and antigenomic RNA. According to a proposed 'single initiation stop-start model', interaction between N-P complexes and the 47 nt Le+ RNA modifies of the transcriptase, promoting read-through of the initiation and termination signals. However, several observations have raised issues with this model, leading to a proposed 'two polymerase entry model' in which the polymerase can initiate either at the genome 3' terminus for replication or internally at the start of the N gene for transcription. To enter full replication mode, this model requires suppression of Le (+) synthesis. In support of this model, it has been reported that there are indeed two distinct forms of RNP complex in infected cells, with the replication complex comprising only the P, L and the N-RNA template and the transcription complex sequestering several additional host proteins including elongation factor 1 α , heat shock protein 60 and RNA guanidinyl transferase.

Assembly and Budding

Assembly and budding of VSV from infected cells occur at the plasma membrane and involve interaction of the viral RNP complex with the M protein. M is synthesized as a soluble cytoplasmic protein. Three forms of M are generated from alternative in-frame initiation codons but only the full-length protein is associated with budding and assembly. A proportion of full-length M is transported to the inner leaflet of plasma membrane where it forms microdomains. The G protein is synthesized at membranes of the rough endoplasmic reticulum where the signal peptide is removed and the step-wise process of N-linked glycosylation commences. G is then trafficked via the Golgi complex to the plasma membrane where it forms separate microdomains from the M protein. The M and G microdomains appear subsequently to coalesce to establish the sites of virion assembly and budding from infected cells. By interacting with M, the RNP complex is condensed into a tightly packed helix. Late domain motifs (²⁴PPPY²⁷ and ³⁷PSAP⁴⁰) in the flexible N-terminal region of M then mediate the recruitment of host cell factors that are required for the late step

of virus egress. The ²⁴PPPY²⁷ domain interacts with the WW domains of E3 ubiquitin ligase Nedd4 and ³⁷PSAP⁴⁰ recruits Tsg101/vps23, a component of the ESCRT (endosomal sorting complex required for transport) complex. Recruitment of ESCRT to the site of viral budding facilitates multivesicular body (MVB) biogenesis as well as the constriction and severing the neck of the bud, allowing the release of mature virions.

Inhibition and Modification of Host Cell Functions

VSV deploy a range of strategies to enhance viral replication processes and inhibit host defensive responses, many of which are mediated through the three forms of the M protein (M1, M2 and M3). The M protein targets several cellular proteins to inhibit host gene expression. At one level, M shuts off transcription by host RNA polymerases I and II. In the case of RNA polymerase II, inhibition has been shown to be due to inactivation of transcription factor TFIID. The M protein also blocks nuclear-cytoplasmic export by interacting with the nucleoporin Nup98 and the mRNA export factor Rae1. Although not attributed directly to the M protein, host protein synthesis is also blocked directly during VSIV infection by inhibition of host translation initiation factors eIF2 and eIF4E. In addition to favouring viral replication, inhibition of host gene expression results in cell rounding due to disruption to the cytoskeleton, apoptosis and inhibition of the interferon response. The M protein also appears to directly target the host defensive response by interaction with LMP2, a chymotrypsin-like catalytic subunit of the immunoproteasome which is involved in the generation of MHC class I-compatible peptides.

BEFV activates the Src-JNK-AP1 and PI3K-Akt-NF- κ B signaling pathways to enhance virus replication (see above). BEFV has also been shown to induce apoptosis in infected cells through the activation of the Fas-mediated and mitochondria-mediated caspase-dependent pathways. Specific viral proteins involved in activation of these signaling pathways have not yet been identified. However, the C-terminal domain of the BEFV α 1 has been shown to bind importin β 1 and importin 7, suggesting that it may modulate components of nuclear trafficking pathways.

Epidemiology

Based on antigenic and genetic differences, vesicular stomatitis viruses are divided into two distinct serotypes, Indiana and New Jersey. The Indiana serotype is further divided into four subtypes: classical Indiana or Indiana 1 virus (VSIV), Cocal or Indiana 2 virus (COCV), Alagoas or Indiana 3 virus (VSAV) and Morreton or Indiana 4 virus (MORV). At present, all of these viruses are restricted to the Americas, although the known geographic distribution of each within the continents varies. Vesicular stomatitis New Jersey virus (VSNJV) and VSIV occur throughout both continents from Canada to Argentina. In temperate regions of the Americas, their occurrence is sporadic with seasonal activity during warmer months of the year. In tropical and subtropical regions of Mexico, Central and South America, VSNJV and VSIV are endemic. In contrast, COCV has only occasionally been reported from Trinidad and Brazil, VSAV from north-eastern Brazil, and MORV from eastern Colombia. Each of these viruses, except MORV, have been associated with vesicular disease in livestock.

Despite many years of research, the epidemiology of vesicular stomatitis is not well understood. Serologic studies indicate that in regions where these viruses occur, a wide variety of mammals (wild and domestic) as well as humans are naturally infected, although the precise mode of their infection is uncertain. Vesicular stomatitis viruses have been isolated from a variety of field-collected biting insects (sandflies, mosquitoes, black flies, midges and eye gnats) and they are presumed to be arthropod-borne. Experimental studies have demonstrated infection, replication, bite transmission and even vertical (transovarial) transmission of VSV in some of these insects, although how the insects acquire the virus in nature is puzzling. Some insects, such as black flies (*Simulium*), biting midges (*Culicoides*) and eye gnats (*Hippelates*) are probably infected during outbreaks by feeding on vesicular lesions on livestock; fluid and exudate from such lesions contain high titers of infectious virus. However, results of field and experimental studies indicate that equines, bovines, swine and a variety of wild mammals have low levels of viremia during acute infection with VSV. In experimental studies, the duration of VSV viremia in livestock and wild animals is quite transient and often undetectable, with the development of high titers of specific neutralizing antibodies in the serum within 5–7 days. This suggests that a hematophagous insect, such as a sandfly or a mosquito, that fed directly on the blood of an infected mammal would not ingest sufficient virus to become infected. Birds are relatively refractory to VSV infection and there is no good evidence for persistent or congenital VSV infection in mammals.

VSNJV, VSIV, COCV and MORV have all been isolated from biting arthropods collected at localities where no livestock cases of vesicular disease were occurring. This observation raises the question: how is VSV maintained in endemic regions, during inter-epidemic periods or at times when the presumed insect vectors are absent or in low numbers? The most likely explanation is vertical transmission (VT) in phlebotomine sandflies. Experimental VT has been demonstrated in sandflies with both VSNJV and VSIV. VT occurs for a number of arboviruses in their arthropod vectors. In the case of VSV, VT would explain how the virus survives in nature between epizootics or when most of the local animal population is immune.

During vesicular stomatitis epizootics in livestock, the virus can also be transmitted by direct contact (i.e., among animals feeding from the same container, by infected calves feeding on mother, virus-contaminated milking machines, or to animal handlers and veterinarians working with infected animals). This probably occurs by introduction of virus through cuts or scratches in the skin or buccal mucosa, since there is not good evidence of aerosol transmission in nature or excretion of virus in milk or excrement of infected animals.

In contrast to VSV, BEFV occurs across a very large geographic range encompassing tropical and sub-tropical regions from Africa, the Middle-East, Asia and Australia. The virus does not occur in the Americas, the Pacific Islands or Antarctica, or in Europe excluding eastern Thrace. The vertebrate host range includes cattle, water buffalo and yak, as well as a wide range of wild African ungulates. The disease in livestock occurs seasonally, principally in the summer and early autumn, and with the onset of the monsoon season in Asia. Outbreaks are usually associated with periods of high rainfall which precipitate the emergence of insect vectors in large numbers. BEFV can also spread in epizootics that follow the pattern of prevailing winds with a general southward movement in the southern hemisphere and northward movement above the Equator. Wind-borne movement of insect vectors is the likely mechanism of spread. Vectors include biting midges (*Culicoides* spp.) and mosquitoes. BEFV has been isolated from *Anopheles bancroftii*, *Culicoides brevitarsis*, *Culicoides coarctatus*, a pool of mosquitos of mixed species in Australia, and from a mixed pool of biting midges in Africa. The virus has also been recovered from biting midges and mosquitoes of several other species following experimental infection. The abundance and distribution of insects from which BEFV has been isolated suggests that several major vectors may be involved in transmission. However, several lines of evidence indicate that, in Australia, mosquitoes rather than biting midges serve as the principal vectors. There is no evidence of direct transmission of BEFV between cattle, even when encouraged by smearing nasal or ocular discharges on mucosal surfaces.

Molecular epidemiological studies, primarily using the deduced amino acid sequences of the BEFV G protein ectodomain, have identified three largely independent epidemiological systems (episystems) in East Asia, Australia and the Middle-East. However, the absence of BEFV sequence data Africa and most of southern and western Asia represents a large gap in our understanding of the phylogeography. In each of the major episystems, there is evidence of relatively rapid spread across vast distances and the continual evolution of new lineages. Recent evidence of the presence of viruses of the East Asian lineage in the Middle East suggests that trade in livestock can contribute to translocation of the virus globally.

Pathogenesis and Clinical Features

The disease vesicular stomatitis primarily affects livestock (horses, cattle and swine), although humans can also develop the illness when handling affected animals or working with the virus in the laboratory. Vesicular stomatitis is of economic importance to farmers because it causes weight loss, reduced milk production and mastitis in cows, and lameness in affected animals. In addition, there are national and international restrictions on shipment of animals and semen from affected regions to VSV-free areas. Another major reason for its importance is that the clinical signs of vesicular stomatitis are similar to those of foot-and-mouth disease. In bovines and swine, the two diseases can only be differentiated by laboratory tests.

The incubation period for vesicular stomatitis ranges from 2 to 8 days. In affected livestock, excessive salivation (drooling) may be the first sign of illness. Subsequently, vesicular lesions (blisters) develop on the lips, around nostrils, tongue, corners of the mouth and buccal mucosa. The vesicles rupture with grazing or eating, leaving painful ulcers, so the affected animals often stop eating, resulting in weight loss. Affected cows may also develop lesions on their teats; mastitis and reduced milk production often result. In addition, livestock also can develop vesicular lesions in the coronary band of their feet, causing swelling and sometimes loss of the hoof. Lameness may result, which in some animals such as thoroughbred horses can be a major loss for owners. Histopathologic examination of skin lesions shows intercellular edema in the malpighian layer and epithelial cell necrosis extending down to the basal cells with neutrophilic and monocytic inflammation. The virus destroys cells above the basal layer forming large vesicles which rupture leaving cutaneous ulcers and necrotic tissue (Fig. 4). Without complications, the ulcers heal and most livestock recover in about two weeks.

Serological surveys indicate that many livestock as well as wild mammals and humans residing in rural areas where VSV is endemic have neutralizing antibodies to one or more of the five VSV serotypes or subtypes. This suggests that many VSV infections are probably subclinical or unrecognized. The explanation for this apparent paradox may be the route of infection. Intracerebral or intranasal inoculation of VSNJV or VSIV into adult mammals usually results in encephalitis and death. In contrast, when livestock and most other adult animals are inoculated intramuscularly or intravenously, they show little sign of illness and rapidly develop



Fig. 4 Tongue of a horse infected with VSNJV, showing extensive ulceration and loss of epithelium. Photo courtesy of R.A. Bowen.

neutralizing antibodies to the virus. However, in bovines, equine or swine, if these viruses are inoculated intradermally or intralingually or are rubbed onto abraded skin or buccal mucosa, the animal usually develops vesicular lesions on and around the site of introduction. Experimental transmission studies of VSNJV by infected black flies to cattle have given similar results. When the infected insects fed around the mouth, nostrils or coronary band of cattle, vesicular lesions developed at the bite sites; however, when the infected flies were allowed to feed on the flank or neck of the animals, infection occurred but vesicles were not observed.

Most VSV human infections are also probably asymptomatic or unrecognized. Serologic surveys among people living in rural communities in Central and South America in regions where VSNJV, VSIV, and MORV are endemic have detected specific neutralizing antibody prevalence up to 90%. Furthermore, the prevalence of VSV antibody increased with age, suggesting continued exposure and infection. In contrast, reported VSV infections in persons who handle sick livestock or in laboratory personnel working with the virus have developed an acute febrile illness with headache, myalgia, chills, anorexia and weakness lasting 3–6 days. In these individuals, the time and method of exposure were known (a hand laceration contaminated with a sick animal's saliva; a needle stick; or possibly an aerosol with a high dose of virus). It is uncertain whether humans naturally infected in endemic areas have a similar but unrecognized febrile illness or have an asymptomatic infection. The route of infection and the virus dose could also influence the outcome of infection in humans.

Bovine ephemeral fever is a disease affecting cattle and water buffalo. There is also a high prevalence of BEFV neutralizing antibodies in yaks in Tibet. Clinical signs of BEFV infection can include a bi-phasic fever, salivation, ocular and nasal discharge, recumbency, muscle stiffness, lameness and anorexia. As suggested by the name, ephemeral fever usually manifests with rapid onset and rapid recovery, lasting only 1–3 days, but prolonged paralysis and ataxia can sometimes occur. During epizootics in Australia, infection and morbidity rates are commonly very high (sometimes approaching 100% of susceptible livestock) but mortality rates are low, rarely exceeding 1%. However, reports from East Asia suggest somewhat higher mortality rates (5%–10%) have occurred historically and, in recent years, there have been reports from Taiwan (2002), mainland China (2011) and Turkey (2012) of case-fatality rates ranging from 20% to 50%, suggesting there may have been a shift in virulence of some BEFV strains of Asian origin. The economic impacts of BEFV can be high due to cessation of lactation in dairy cattle, loss of condition in beef cattle and immobilization of water buffalo used for draft power.

Bovine ephemeral fever is principally an inflammatory disease. The incubation period is normally 2–4 days. Viremia usually persists for 1–3 days and peaks approximately 24 h before the onset of fever. The initial sites of infection are not known but the virus has been isolated from neutrophils and reticulo-endothelial cells of the lungs, spleen and lymph nodes. There is also evidence of infection in synovial membranes, epicardium and aorta, and in cells derived from synovial, pericardial, thoracic and abdominal fluids. There is not widespread tissue damage. The primary lesion is a vasculitis affecting the endothelium of small vessels of synovial membranes, tendon sheaths, muscles, fascia and skin. The onset of fever and other clinical signs is accompanied by marked leucopenia, relative neutrophilia, elevated plasma fibrinogen and elevated levels of cytokines including interferon α , interleukin 1 and tissue necrosis factor. There is also a significant hypocalcemia that is thought to be responsible for sternal recumbency.

Control and Treatment

Control of vesicular stomatitis is difficult, since knowledge of the ecology, mode of transmission and maintenance mechanisms of the viruses causing the disease is limited. If sandflies or some other arthropod are the natural reservoir of the virus, then VSV probably cannot be eradicated. However, control measures can be taken during outbreaks of vesicular stomatitis to stop or to reduce spread of the virus. While VSV does not appear in urine or feces of infected animals, it is present in high titers in their saliva and in exudate from the vesicles and ulcerative lesions. Thus, cleaning water pails and feeding troughs with a viricidal disinfectant (i.e., 10% solution of household bleach) will decrease virus transmission. Many outbreaks in dairy cows have been traced to contaminated milking equipment, so cows with vesicular lesions should be milked last, and the equipment disinfected between uses. Infected livestock also should be quarantined and separated from other animals in the herd. Likewise, animals from an infected farm should not be transported to other premises. Generous application of insecticides in barns and corrals with infected livestock will reduce flies and other insect pests that might feed on lesions of affected animals and mechanically transmit the virus to others. Persons handling sick animals should also wear disposable gloves to avoid transmitting the virus to other animals or infecting themselves. Dairy farmers sometimes don't report vesicular stomatitis in their herds; if quarantined, the milk can't be sold until the farm is declared free of the disease. However, even if not reported, owners can initiate the above control measures.

At present, there are no effective VSV vaccines available for livestock. Various inactivated VSV vaccines have been tested but immunity generally wanes after a few years and reinfection with the same serotype can occur.

Natural BEFV infection induces a strong neutralizing antibody response and apparently durable immunity. Following experimental infection, neutralizing IgG antibody appears 4–5 days after the onset of clinical signs and peaks within 1–4 weeks. Although there are some reports that cattle with high levels of neutralizing antibody can be susceptible to experimental challenge, other evidence suggests a good correlation between protection and neutralizing antibody. Colostral antibody has also been shown to protect calves against experimental challenge. High levels of cytokines circulate during the acute phase of infection but little is known of the role of innate or adaptive cell mediated immunity in recovery from infection or protection against natural or experimental challenge.

Several forms of live-attenuated, inactivated, subunit and recombinant BEFV vaccines have been reported and vaccines of varying format are produced for commercial use. Live-attenuated vaccines have been produced in mice and in cell cultures. In

general, live vaccines are relatively effective in inducing protection but require at least two doses in adjuvant to generate durable immunity. Inactivated vaccines have been produced by treatment of BEFV with formalin or β -propiolactone but they have generally poor efficacy. Consecutive vaccinations with live-attenuated and killed preparations have also been used with some success. A purified G protein subunit vaccine delivered in Quil A adjuvant has been shown experimentally to provide reliable protection following a two-dose treatment at an interval for 21 days. Recombinant BEFV vaccines employing the BEFV G protein delivered in vaccinia and capripox viral vectors have also been trialed.

The major clinical signs of bovine ephemeral fever can be treated very effectively with non-steroidal anti-inflammatory drugs, such as phenylbutazone, which can reduce the duration of the disease. Calcium injections given early in the course of disease may also be used to treat animals that are recumbent. Where possible, affected cattle should be provided with shelter, water and food, particularly in hot weather.

Further Reading

- Albertini, A.A.V., Baquero, E., Ferlin, A., Gaudin, Y., 2012. Molecular and cellular aspects of rhabdovirus entry. *Viruses* 4, 117–139.
- Gaudier, M., Gaudin, Y., Knossow, M., 2002. Crystal structure of the vesicular stomatitis virus matrix protein. *Proceedings of the National Academy of Science of the United States of America* 21, 2886–2892.
- Ge, P., Tsao, J., Schein, S., *et al.*, 2010. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* 327, 689–693.
- Joubert, D.A., Blasdel, K.R., Audsley, M.D., *et al.*, 2014. Bovine ephemeral fever rhabdovirus α 1 protein has viroporin-like properties and binds importin β 1 and importin γ . *Journal of Virology* 88, 1591–1603.
- Letchworth, G.J., Rodriguez, L.L., Del Cbarrera, J., 1999. Vesicular stomatitis. *The Veterinary Journal* 157, 239–260.
- Liang, B., Li, Z., Jenni, S., *et al.*, 2015. Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. *Cell* 162, 314–327.
- Lyles, D.S., 2013. Assembly and budding of negative-strand RNA viruses. *Advances in Virus Research* 85, 57–90.
- Lyles, D.S., Kuzmin, I.V., Rupprecht, C.E., 2013. *Rhabdoviridae*. In: Knipe, D.M., Howley, P.M., Cohen, J.L., *et al.* (Eds.), *Fields Virology*, sixth ed. Philadelphia: Lippincott Williams and Wilkins, pp. 885–924.
- Roche, S., Rey, F.A., Gaudin, Y., Bressanelli, S., 2007. Structure of the prefusion form of the vesicular stomatitis virus G protein. *Science* 315, 843–848.
- Rodriguez, L.L., Pauszek, S.J., 2012. Genus *vesiculovirus*. In: Dietzgen, R.F., Kuzmin, I.V. (Eds.), *Rhabdoviruses: Molecular Taxonomy, Evolution, Genomics, Ecology, Host-Vector Interactions, Cytopathology and Control*. Norfolk: Caister Academic Press, pp. 23–35.
- Tesh, R.B., Bolling, B.G., Beaty, B.J., 2016. Role of vertical transmission in mosquito-borne arbovirus maintenance and evolution. In: Vasilakis, N., Gubler, D.J. (Eds.), *Arboviruses: Molecular biology, Evolution and Control*. Norfolk: Caister Academic Press, pp. 191–217.
- Trinidad, L., Blasdel, K.R., Joubert, D.A., *et al.*, 2014. Evolution of bovine ephemeral fever virus in the Australian epizootic. *Journal of Virology* 88, 1525–1535.
- Walker, P.J., 2005. Bovine ephemeral fever in Australia and the world. *Current Topics in Microbiology and Immunology* 292, 57–80.
- Walker, P.J., Klement, E., 2015. Epidemiology and control of bovine ephemeral fever. *Veterinary Research* 46, e124.
- Whelan, S.P.J., Barr, J.N., Wertz, G.W., 2004. Transcription and replication of non-segmented negative-strand RNA viruses. *Current Topics in Microbiology and Immunology* 283, 61–119.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/rhabdoviridae
Rhabdoviridae.

West Nile Virus (*Flaviviridae*)

Fengwei Bai and Elizabeth Ashley Thompson, The University of Southern Mississippi, Hattiesburg, MS, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of Laura D. Kramer, Elizabeth Kauffman, West Nile Virus (*Flaviviridae*), In Reference Module in Biomedical Sciences, Elsevier Inc., 2017, doi:10.1016/B978-0-12-801238-3.02696-9.

Glossary

Autoimmune disease A disease caused by attacks from the own immune system.

Blood-Brain Barrier A highly selective semipermeable border in the capillaries that blocks the passage of certain substances in the blood into the brain and spinal cord tissue.

Encephalitis Inflammation of the brain.

Epidemiology The study deals with the incidence, distribution, and possible control of diseases and other factors relating to health.

Meningitis Inflammation of the protective membranes covering the brain and spinal cord.

Neuroinvasive disease Disease caused by a pathogen that infects nerve cells.

Pathogenesis The mechanism of how a disease develops.

Viremia The presence of viruses in the blood.

Virion The complete, infectious viral particle.

Classification

West Nile virus (WNV) is a member of the genus *Flavivirus* within the family of *Flaviviridae*, which also includes three other genera, i.e., *Pestivirus* (including bovine viral diarrhea virus), *Hepacivirus* (including hepatitis C virus, HCV), and a recently established genus *Pegivirus* (including human pegivirus-1). The family *Flaviviridae* is named after the Latin word *flavus* (yellow) referring to jaundice caused by yellow fever virus (YFV). *Flavivirus* genus is composed of more than 90 viruses, many of which are important mosquito-transmitted human pathogens, including WNV, dengue virus (DENV), Japanese encephalitis virus (JEV), St. Louis encephalitis virus (SLEV), Zika virus (ZIKV) and YFV. While most of these viruses are transmitted by mosquitoes and ticks, some transmit without known transmission vectors, such as HCV. Despite many flaviviruses can cause human diseases varying from fever to life-threatening hemorrhagic fever and encephalitis, some do not infect humans and other vertebrate animals, such as mosquito-specific flaviviruses.

Virion Structure

WNV has a relatively smooth protein surface (no projections or spikes), enveloped, spherical structure with icosahedral symmetry approximately 50 nm in diameter. Cryoelectron microscopy (cryo-EM) shows a concentric multilayer organization of the virus. The outermost layer contains Envelope (E) and Membrane (M) transmembrane proteins. A host-derived lipid bilayer appears non-spherical with the transmembrane components of the M and E proteins. WNV and other flaviviruses are icosahedral enveloped viruses composed of a lipid bilayer surrounding a nucleocapsid containing its genomic RNA complexed with multiple copies of C proteins (Fig. 1).

Genome and Viral Proteins

WNV has a positive-sense, linear, single-stranded (ss) RNA genome, which is approximately 11 kb long and is flanked by 5' and 3' non-coding stem-loop structures with a 5' cap (m7GpppAmp) and no poly-A tail, respectively. The coding region of the genome has a single open reading frame (ORF), which is translated into a polyprotein that is cleaved by viral and host proteases into three structural proteins (capsid [C], pre-membrane [PrM/M], and envelope [E]), and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (Fig. 2). However, certain studies suggest that additional proteins may also be produced through ribosomal frameshifting. While non-structural proteins play essential roles in the replication of viral genomic RNA, the structural proteins assemble the virion and mediate host receptor binding and viral entry during the virus life cycle (Table 1).

C Protein

The capsid (C) protein is composed of a large number of positively charged amino acid residues and multiple copies dimerize and tetramerize into a nucleocapsid that is encapsulated by a lipid bilayer derived from the host cell membranes. The N- and

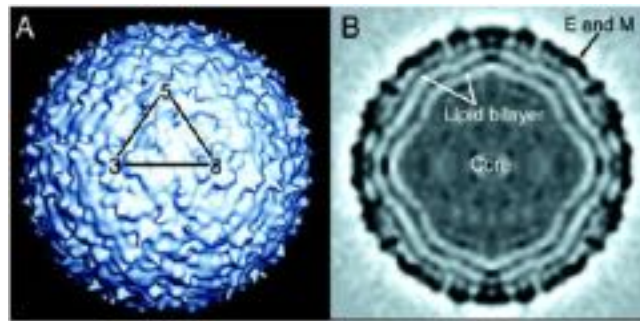


Fig. 1 WNV virion structure. A 17 Å structure of WNV virion determined by cryo-EM. (A) surface shaded view of the virus with one asymmetric unit of the icosahedron is shown as a triangle. The 5-fold and 3-fold icosahedral symmetry axes are shown. (B) A central cross section showing the concentric layers of density. Reproduced with permission from Mukhopadhyay, S., Kim, B.S., Chipman, P.R., Rossmann, M.G., Kuhn, R.J., 2003. Structure of West Nile virus. *Science* 302, 248.

C-terminal parts of C protein have been shown to facilitate RNA folding during viral replication through RNA chaperon activity to the capsid, and the central part of the C protein structure shows the presence of four α -helices likely involved in direct interaction with the viral RNA. In addition, it has also been associated with claudin protein degradation, which leads to the epithelial barrier damage and virus dissemination in the human host. While the C protein has a role in the induction of apoptosis, it is also suggested to block apoptosis through a phosphatidylinositol (PI) 3-kinase-dependent pathway.

prM/M

The prM protein of WNV is a short, glycosylated protein with a molecular weight of 20–26 kDa, which plays an essential role in viral particle formation and secretion. PrM chaperones E protein in the formation of immature virions and prevents a premature fusion during virus egress. The viral particles mature during virus egress through the secretory pathway, where cleavage of prM by a furin-like protease generates infectious particles, albeit the efficiency of the cleavage may be fairly low at approximately 30%. The immature prM-containing flavivirus particles are non-infectious, which may due to the requirement of the cleavage of prM to M to activate the membrane-fusion machinery of the virus.

E Protein

The envelope (E) protein is a transmembrane glycoprotein embedded into the lipid bilayer by a C-terminal α -helix. One hundred and eighty copies of E proteins, together with the same numbers of M proteins span through the lipid bilayer and cover the outer surface of the virion. Thus, E and M are the main antigens triggering neutralizing antibody responses in WNV infection and mediating the host receptor binding. Flavivirus E ectodomain is composed of three structural domains. i.e., DI, DII, and DIII that share a variable degree of homology among different flaviviruses. DI is a β -barrel structure containing a single glycosylation motif, while DII forms an elongated finger-like structure that contains an internal fusion loop, which is conserved among flaviviruses. DIII is an immunoglobulin (Ig)-like domain that has a major role in cellular receptor binding and viral entry. Although neutralizing epitopes have been described in all three domains, antibodies against DIII are generally considered to show higher neutralizing capacity compared to DI and DII-specific antibodies.

NS1

NS1 is a highly conserved protein sharing ca. 20%–40% identity and 60%–80% similarity among flaviviruses. The flavivirus NS1 protein is a multifunctional glycoprotein that can be found intracellularly, on the cell surface of infected cells, or secreted by infected cells as a hexamer and it circulates in the bloodstream. NS1 is a diagnostic target as it is secreted and circulates in patients' blood at an early stage of viral infection. The intracellular dimer form of NS1 facilitates viral replication and assembly, while the secreted hexameric NS1 downregulates viral immune responses. WNV NS1 has been shown to inhibit complement activation by binding the regulatory protein factor H, attenuate classical and lectin-dependent complement activation by directly interacting with C4, and antagonization of production of type I interferon. WNV NS1 has also been shown to induce hyperpermeability of brain endothelial cells and vascular leakage in the mouse brain, and induce hyperpermeability of human brain endothelial cells resulting in neurological complications and encephalitis. Besides NS1, WNV also produces an additional non-structural protein, NS1', a C-terminally extended product of NS1 generated as the result of a -1 programmed ribosomal frameshift, which leads to an addition of 52 amino acid (aa) residues at the C terminal of NS1 protein. The full-length NS1' may not be essential for virus replication in vitro or virulence in mice, but the sequence of the last 20 aa of NS1' appears to be responsible for its cellular retention.

NS2A and NS2B

NS2A and NS2B are small hydrophobic proteins. NS2A is a transmembrane protein involved in the production of virus-induced membrane structures and virion assembly, while NS2B is a cofactor of NS3 protease. WNV NS2A plays a role in the inhibition of type I interferon (IFN- α/β) production and it has been suggested to play a role in virus-induced cell apoptosis. NS2B binds to NS3 to function as a membrane anchor for the viral protease and is essential to NS3 proteolytic activity.

NS3

NS3 is a highly conserved protein across all flaviviruses and its N-terminal end contains a trypsin-like serine protease. When activated by binding to its cofactor NS2B, NS3 forms an NS2B-NS3 protease that cleaves the viral polyprotein into structural and nonstructural proteins. The NS2B-NS3 complex localizes to convoluted membranes, paracrystalline structures, and in the proliferating endoplasmic reticulum (ER), suggesting that these membranes are also involved in the proteolytic process. Besides its N-terminal protease functions, the C-terminus of NS3 possesses other important functions such as helicase, nucleoside triphosphatase, and RNA triphosphatase activities that are essential for viral replication. During viral replication, NS3 helicase separates the double-stranded (ds) RNA intermediates or disrupt secondary structures formed by the single-stranded (ss) RNA template. NS3 helicase has been shown to promote WNV resistance to antiviral action of 2',5'-oligoadenylate synthetase 1b (OAS1b). Recently, it was shown that the WNV NS3 helicase domain inhibits type I IFN signaling. The NS2B-NS3 protease of WNV, DENV, ZIKV, and JEV has been shown to interrupt human but not mouse, cGAS-STING pathway thus inhibiting IFN-mediated antiviral responses in humans.

NS4A and NS4B

NS4A and NS4B are small hydrophobic proteins that are linked by a 2K peptide composed of 23 amino acid residues. The complete cleavage of the polyprotein generates the 2K fragment that directs NS4B into the ER. Mutations in the 2K fragment resulted in resistance against the antiviral action of the IFN-inducible OAS1b during WNV infection. The NS4A and NS4B proteins have been related to membrane rearrangement in infected cells and to inhibit IFN signaling during WNV infections. NS4A and NS4B seems to be inducers of an unfolded protein response, which displays a high correlation in inhibiting IFN- α stimulated JAK-STAT signaling during WNV infection. Similar to HCV, WNV NS4A functions as a cofactor of NS3 to regulate the ATPase activity of the S3 helicase. The flaviviral NS4B appears to play an important role in viral replication by facilitating the formation of the viral replication complexes and in counteracting innate immune responses.

NS5

Flavivirus NS5 is the largest non-structural protein encoded by the viral genome. It contains an N-terminal methyltransferase (MTase) activity and a C-terminal RNA-dependent RNA-polymerase activity for viral genome replication. The NS5 MTase activity is responsible for the viral genome 5' capping, which mimics eukaryotic mRNA and allows to use of host protein synthesis machinery for efficient viral gene translation. Specifically, NS5 has been shown to sequentially catalyze methylation of a guanine N-7 residue followed by a ribose 2'O site on the 5' end of viral RNA. Like other RNA viruses, WNV NS5 RNA-dependent RNA-polymerase lacks a proof-reading capability; thus, the viral populations display a variable level of sequence diversity which may favor the selection of variants in response to selective immunological pressures. Besides the functions in virus replication, NS5 protein has been shown to function as a potent antagonist of type I IFN-mediated JAK-STAT signaling. Like in most flaviviruses, WNV NS5 is located in the cell cytoplasm at sites of virus replication. However, several studies have also shown that a significant proportion of NS5 of WNV (Kunjin strain), DENV, JEV, YFV, and ZIKV is found in the nucleus suggesting its important role during flavivirus infection. In DENV infection nuclear NS5 seems to downregulate interleukin-8 (IL-8) production and interfere with the host cellular mRNA splicing.

Life Cycle

WNV enters a host cell through receptor-mediated endocytosis after the E protein attachment to cell surface receptor(s). It appears that receptor recognition and attachment is mediated by multiple molecules in a combinatorial or consecutive manner. WNV and other flaviviruses make initial contact with the host cell by binding to glycosaminoglycans (GAGs), such as heparan-sulfate proteoglycans or syndecans. GAGs are long, anionic, unbranched polysaccharides located on the surface of eukaryotic cells and in the extracellular matrix. GAGs are prominently exposed on the cell surfaces of all tissues, providing an easily accessible primary receptor for viral adhesion by electrostatic interactions. Several molecules have been implicated as cellular receptors that mediate WNV entry and infection, including mammalian $\alpha v\beta 3$ integrin, DC-SIGN/DC-SIGNR, mannose receptor, and mosquito mosGCTL-1. However, like other flaviviruses, WNV entry process involves multi-protein interactions with a high degree of redundancy and a single, specific entry receptor might not exist. After attachment to the host cell surface receptors, WNV will be endocytosed into early endosomes with the pH dropping from neutral to slightly acidic and becoming more acidic during maturation to the late endosomes. The acidic pH in the late endosome triggers the conformation change of E protein leading to the

fusion of the viral lipid membrane with the endocytic membrane and the release of the viral RNA genome into the cell cytoplasm. Using the cellular translational machinery, the viral RNA is translated into a single polyprotein which is cleaved into ten viral proteins. Genomic positive-sense RNA is transcribed into complementary negative-sense RNA on the ER membranes, which serves as templates for the synthesis of 10- to 100-fold positive-sense RNA molecules. The newly synthesized positive-sense RNA is packaged in viral C protein to form the nucleocapsid, and bud into the cytoplasm via the Golgi complex. Immature virus particles are transported to the cell surface in exocytic vesicles acquiring an envelope consisting of a host-derived lipid bilayer and viral prM/ M and E proteins. The viral particles mature as cellular enzymes cleave the prM, resulting in the release of mature virus from the cell surface. Mature virions are released from the infected cell by exocytosis beginning at 10–12 h after the infection. (Fig. 2).

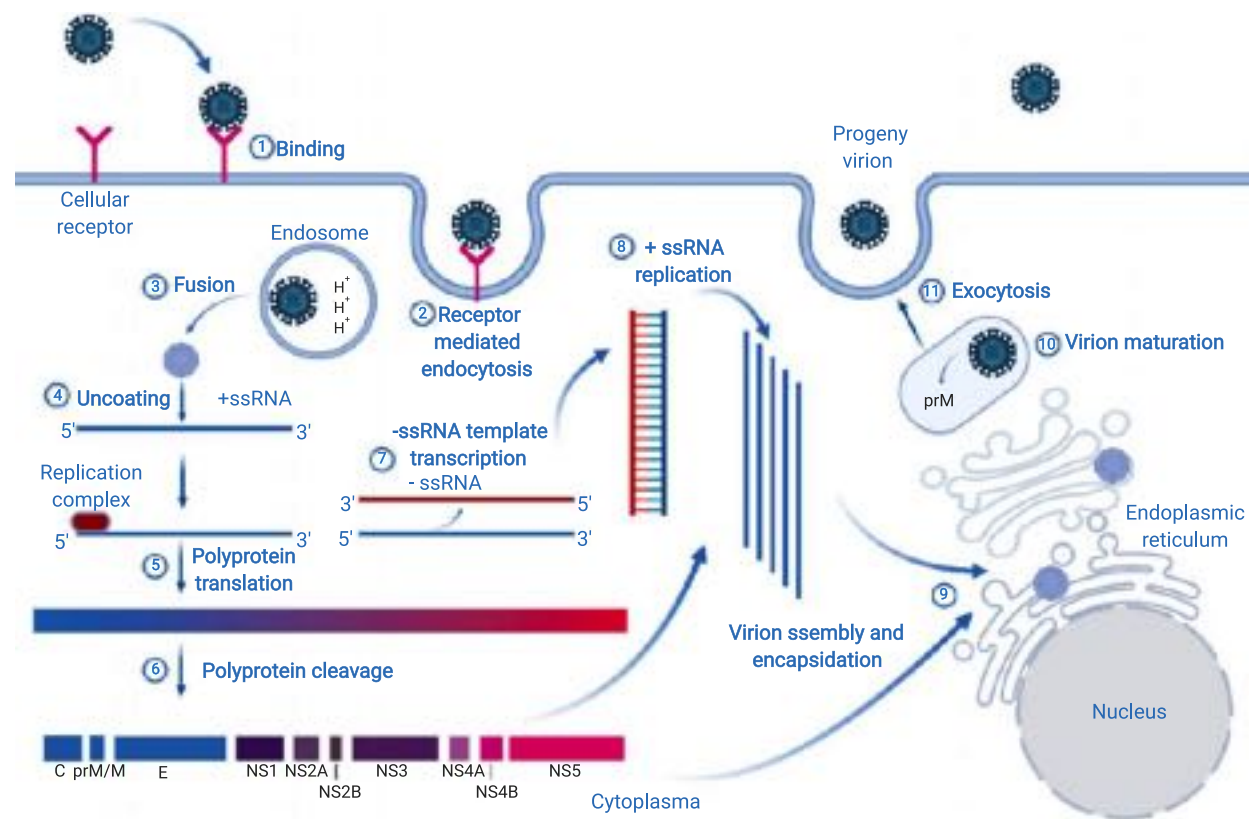


Fig. 2 WNV life cycle. 1. WNV particle binds to a cellular receptor on the cell surface; 2. The viral particle is enclosed in an endosome via a receptor-mediated endocytosis; 3. The acidic pH in the late endosome triggers the membrane fusion between the viral lipid membrane with the endocytic membrane and release nucleocapsid into the cell cytoplasm; 4. Uncoating of the nucleocapsid releases positive sense, single-stranded viral RNA genome (+ ssRNA); 5. Using the cellular translational machinery, the viral RNA is translated into a single polyprotein; 6. The polyprotein is cleaved into ten viral proteins; 7. + ssRNA is transcribed into complementary negative-sense RNA (-ssRNA); 8. Synthesis of + ssRNA using -ssRNA as a template; 9. The newly synthesized + ssRNA is packaged in viral C protein to form the nucleocapsid, and bud into the cytoplasm via the Golgi complex; 10. The immature virus is transported in an exocytic vesicle and acquires an envelope consisting of a host-derived lipid bilayer and viral prM/M and E proteins; 11. The viral particle matures as cellular enzymes cleave the prM, resulting in the release of mature virus from the cell surface.

Table 1 WNV proteins and their functions

Proteins	Functions
C	Nucleocapsid formation, viral RNA folding, regulation of cell apoptosis
prM/M	Viral particle formation and maturation
E	Viral surface protein, receptor attachment, entry
NS1	Viral replication and assembly, regulation of immune response
NS2A	Virion assembly, inhibition of IFN production, regulation of cell apoptosis
NS2B	Cofactor of NS2B-NS3 protease, polyprotein cleavage
NS3	Polyprotein cleavage, helicase, inhibition of IFN production, inhibition of antiviral response
NS4A	Membrane rearrangement, inhibition of IFN production, regulation of immune response, cofactor of NS3 helicase
NS4B	Membrane rearrangement, inhibition of IFN production, regulation of immune response, formation of the viral replication complexes
NS5	Methyltransferase, RNA polymerase, inhibition of IFN production, regulation of immune response

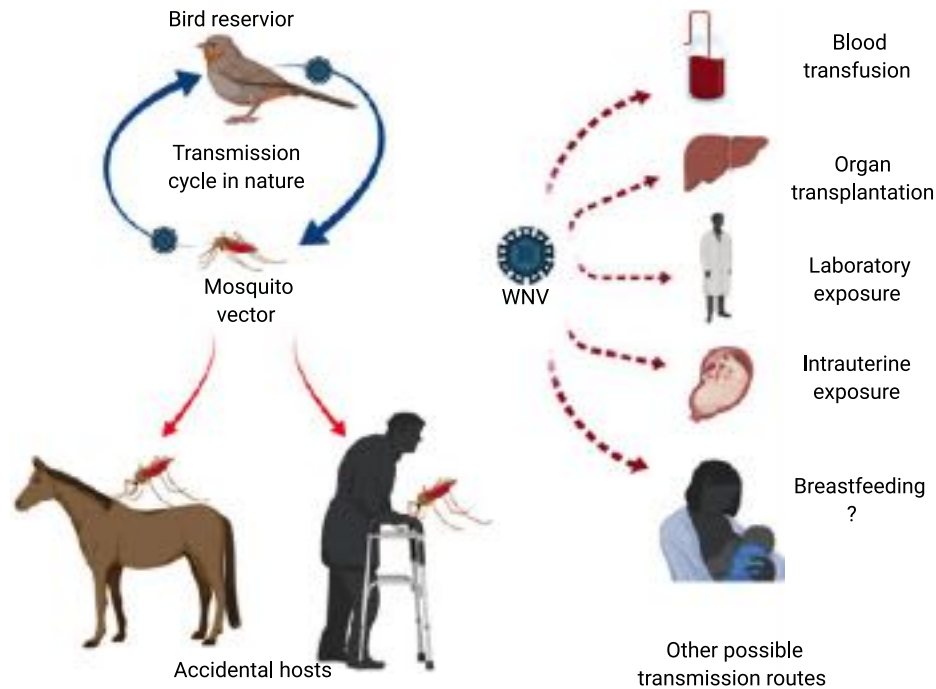


Fig. 3 WNV transmission cycle. WNV maintains its transmission cycle between *Culex* mosquitos and birds in nature. Infected mosquito bites can also transmit WNV to accidental hosts, such as humans and horses. Other than mosquito bites, WNV can also be transmitted to humans through other routes, such as blood transfusion, organ transplantation, laboratory or intrauterine exposure.

Epidemiology

WNV was first discovered in 1937 in West Nile Province Uganda Africa. In 1999, WNV gained entry into North America in New York City and within 3 years, the virus had spread to most of the continental states of the U.S. as well as the neighboring countries of Canada, Mexico and the Caribbean. WNV has now spread throughout the rest of the world except Antarctica. WNV is now considered as the most important causative agent of human viral encephalitis worldwide. In the last 20 years, WNV has been estimated to cause more than 6 million human infections, with over 24,000 neurological disease cases and 2300 deaths in the U.S. alone. In addition, a large outbreak occurred in Europe in 2018 involving over 2000 human cases in 15 countries. WNV transmission in nature is maintained in a cycle between mosquitos and a variety of bird species. The ability of different mosquito and bird species to acquire and transmit WNV is highly variable. In the U.S., the American robins (*Turdus migratorius*) and a few *Culex* mosquito species are thought to be the main host responsible for the maintenance and transmission of WNV. Humans, horses and other vertebrate animals can be infected with WNV through the bite of an infected mosquito but mammals are considered as “dead-end hosts” because the infection does not produce a viremia in a magnitude that can infect subsequently biting mosquitos. Besides mosquito bites, WNV transmission may also occur through blood transfusion, organ transplantation, breast-feeding, or intrauterine exposure, or accidental laboratory exposure (Fig. 3).

Clinical Manifestations

WNV infection in humans is predominantly asymptomatic (80%), but it may result in a spectrum of diseases in about 20% infected people, ranging from a febrile illness known as WNV fever to severe neuroinvasive diseases, including acute flaccid paralysis, meningitis, and encephalitis. In those individuals who develop symptoms, most have WNV fever, chills, malaise, headache, backache, myalgias, arthralgias, gastrointestinal symptoms (nausea, vomiting, or diarrhea) and maculopapular rash. Most WNV fever patients recover completely, but fatigue and weakness can persist for weeks or months. About 1 in 150 people who are infected with WNV develop a severe neuroinvasive disease affecting the central nervous system (CNS), such as flaccid paralysis, encephalitis, and meningitis. Symptoms of neuroinvasive disease include high fever, worsening headache, neck stiffness (nuchal rigidity), confusion, stupor, tremors, seizures, muscle weakness or paralysis, and focal neurological deficits. About 10% of people who develop a severe neuroinvasive disease die. Risk factors for severe encephalitis and death include advanced age, history of cardiovascular diseases, hypertension or a chronic renal disease, HCV infection, and immunosuppression. Elderly individuals are more susceptible to neurological diseases that may result in death. Those elderly individuals who survive, as many as 50% may develop post-illness morbidity. Although WNV can be cleared from the body within a few weeks, many WNV neuroinvasive disease survivors suffer a long-term post-infectious CNS symptom. Post-infectious CNS symptoms include post-exertion fatigue,

dizziness, altered sensation, arthralgias, impaired memory, difficulties in concentrating, depression, anxiety, sleep disruption, and recurrent headaches. Moreover, many WNV patients have developed autoinflammatory or autoimmune-related diseases, including classic autoimmune diseases, such as myasthenia gravis (MG). In recent years, the reported frequency of severe neurological illness associated with neuromuscular manifestations, post-infectious CNS symptoms, and autoimmune diseases has been increasing.

Pathogenesis

Mosquito transmitted WNV infection begins in the skin where the virus replicates initially in keratinocytes and Langerhans cells (LCs), the skin resident dendritic cells (DCs). After being activated by the WNV antigens, LCs can migrate to the draining lymph node, leading to T cell priming and activation. WNV-infected DCs produce large amounts of type I IFNs, however, IFN signaling is decreased in DCs of elderly individuals infected with WNV, which could partially explain why the elderly people are more susceptible to a clinical WNV infection. WNV enters the blood stream through the lymphatic drainage or directly by mosquito probing into blood vessels, resulting in a primary viremia that carries the virus to the peripheral organs, such as the spleen, the liver, and the kidneys, where the virus further replicates and re-enters into bloodstream creating a secondary viremia. WNV has also been shown to replicate efficiently in blood leukocytes, such as in neutrophils that may serve as a reservoir for WNV replication and dissemination. In mouse models, a sufficient magnitude of viremia is required for WNV to cross the blood-brain barrier (BBB) to infect the CNS. Although the detailed mechanisms are not completely understood, WNV has been suggested to enter the CNS through multiple pathways, including flow through the BBB tight junctions, direct infection of endothelial cells in the cerebral microvasculature, infection of olfactory neurons, infected leukocytes that “Trojan horse” WNV to the CNS, and/or direct axonal retrograde transport from infected peripheral neurons. Once in the brain, WNV can infect and replicate in various types of CNS residential cells, including neurons, astrocytes, and microglial cells. WNV infection results in necrosis or apoptosis of neuronal cells, which may be the consequence of WNV replication or activation of immune cells such as cytotoxic T cells, leading to neuroinvasive diseases.

Diagnosis

Currently, the diagnosis of WNV infection is mainly based on typical clinical symptoms and WNV-specific immunoglobulin M (IgM) response. The incubation period for WNV infection ranges from 2 to 14 days and the presence of specific anti-WNV IgM in serum or cerebrospinal fluid (CSF) using an IgM-capture enzyme-linked immunosorbent assay (MAC-ELISA) is typically used for the diagnosis. If necessary, plaque reduction neutralization test (PRNT) can be used to exclude the possibility of cross-reactivity with related flaviviruses. WNV-specific IgG ELISA is sensitive and can determine past exposure; however, the results need to be interpreted cautiously since false positives are possible due to cross-reaction with other closely related flaviviruses. Other laboratory techniques such as reverse transcriptase-polymerase chain reaction (RT-PCR) or RT-quantitative PCR (RT-qPCR) can be used to detect viral nucleic acid in serum, cerebrospinal fluid (CSF) and tissue specimens that are collected early in the course of illness; however, they are not commonly used for the diagnosis of WNV infection since the likelihood of detection of WNV RNA is low. Since WNV can be transmitted through blood transfusion, highly sensitive nucleic acid-amplification testing (NAT) assays have been developed for rapid screening of donated blood, which may also be used to complement IgM testing in the diagnosis of acute WNV infection within the first few days of clinical illness.

Treatment and Prevention

Although the licensing of several WNV vaccines for veterinary use have succeeded, there are still no approved human vaccines or specific antiviral treatment available against WNV. Despite of several clinical trials with multiple vaccine candidates, none have proceeded to the final phase of clinical testing yet. The main reasons for the lack of a human vaccine include lack of evidence of sufficient protective immunity; safety concerns in causing potential antibody-dependent enhancement responses to other flavivirus infections; difficulties in recruiting enough WNV cases for a phase III vaccine efficacy study, and economic considerations. Thus, the most effective way to prevent WNV infection is to avoid exposure to mosquitos by applying insect repellents, wear long-sleeved shirts and pants, and taking measures to control mosquitoes in indoors and outdoors. While the specific treatments are not currently available, patients with severe illness often need to be hospitalized to receive supportive care, such as intravenous fluids and pain medication. In addition, clinical researches have documented that administration of intravenous high dose steroids may result in significant clinical improvement in WNV patients who suffer from a severe neuroinflammatory disease. However, the therapeutic benefit of the immunosuppressive drugs in the treatment of WNV disease is still a question of debate.

Acknowledgment

This project was supported in part by the American Heart Association (AHA) under Award Number 17AIREA33671273 and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R15AI135893 to F.B.

Further Reading

- Bai, F., Thompson, E.A., Vig, P.J.S., Leis, A.A., 2019. Current understanding of West Nile virus clinical manifestations, immune responses, neuroinvasion, and immunotherapeutic implications. *Pathogens* 8.
- Campbell, G.L., Marfin, A.A., Lanciotti, R.S., Gubler, D.J., 2002. West Nile virus. *The Lancet Infectious Diseases* 2, 519–529.
- Colpitts, T.M., Conway, M.J., Montgomery, R.R., Fikrig, E., 2012. West Nile virus: Biology, transmission, and human infection. *Clinical Microbiology Reviews* 25, 635–648.
- Diamond, M.S., Pierson, T.C., Fremont, D.H., 2008. The structural immunology of antibody protection against West Nile virus. *Immunological Reviews* 225, 212–225.
- Hawkes, M.A., Hocker, S.E., Leis, A.A., 2018. West Nile virus induces a post-infectious pro-inflammatory state that explains transformation of stable ocular myasthenia gravis to myasthenic crises. *Journal of the Neurological Sciences* 395, 1–3.
- Kaiser, J.A., Barrett, A.D.T., 2019. Twenty years of progress toward West Nile virus vaccine development. *Viruses* 11.
- Kaiser, J.A., Wang, T., Barrett, A.D., 2017. Virulence determinants of West Nile virus: How can these be used for vaccine design? *Future Virology* 12, 283–295.
- Londono-Renteria, B., Colpitts, T.M., 2016. A brief review of West Nile virus biology. *Methods in Molecular Biology* 1435, 1–13.
- Mackenzie, J.S., Gubler, D.J., Petersen, L.R., 2004. Emerging flaviviruses: The spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nature Medicine* 10, S98–S109.
- Martin-Acebes, M.A., Saiz, J.C., 2012. West Nile virus: A re-emerging pathogen revisited. *World Journal of Virology* 1, 51–70.
- Mazeaud, C., Freppel, W., Chatel-Chaix, L., 2018. The multiples fates of the flavivirus RNA genome during pathogenesis. *Frontiers in Genetics* 9, 595.
- Murray, C.L., Jones, C.T., Rice, C.M., 2008. Architects of assembly: Roles of flaviviridae non-structural proteins in virion morphogenesis. *Nature Reviews Microbiology* 6, 699–708.
- Simmonds, P., Becher, P., Bukh, J., *et al.*, 2017. ICTV virus taxonomy profile: Flaviviridae. *Journal of General Virology* 98, 2–3.
- Ulbert, S., 2019. West Nile virus vaccines – Current situation and future directions. *Human Vaccines & Immunotherapeutics*. 1–6. doi:10.1080/21645515.2019.1621149.

Yellow Fever Virus (*Flaviviridae*)

Ashley E Strother and Alan DT Barrett, The University of Texas Medical Branch, Galveston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A.A. Marfin, T.P. Monath, Yellow Fever Virus, in Encyclopedia of Virology (Third Edition), Edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00535-5.

Nomenclature

3'UTR 3' untranslated region

ALT Alanine aminotransferase

AST Aspartate aminotransferase

C Capsid

E Envelope

ELISA Enzyme linked immunosorbent assay

FNV French neurotropic vaccine

FVV French viscerotropic virus

M Membrane

NS Nonstructural

prM pre Membrane

YEL-AND Yellow fever vaccine associated neurotropic disease

YF Yellow fever

YFV Yellow fever virus

Glossary

Biosafety Level The level of safety measures implemented when working with a pathogen. Levels range from 1 to 4 with Level 1 requiring the least precautions and Level 4 requiring the most.

Neurotropic disease Disease affecting the brain.

New World Refers to the Americas.

Old World Refers to Africa.

Viscerotropic disease Disease affecting kidneys, liver and heart.

Classification (Compact)

Yellow fever virus (YFV) is the prototype virus of the family *Flaviviridae* that takes its name from the Latin for yellow (*flavus*). The virus is a member of the genus *Flavivirus*, which presently contains 67 human and animal viruses. In 1901 the virus was the first to be shown to be transmitted by a mosquito. In 1927 YFV was isolated by monkey to monkey passage of blood from Mr. Asibi, a Ghanaian man who had a mild case of YF. Wild-type strain Asibi is the prototype strain of YFV.

On the basis of their ecology, flaviviruses have been termed arthropod-borne viruses, or “arboviruses”, because many are transmitted between vertebrate hosts by either mosquitoes or ticks. YFV was the first virus to be shown to be transmitted by mosquitoes. Carlos Finlay, a Cuban physician and scientist, proposed in 1881 that YF may be transmitted by mosquitoes rather than by a direct human contact. Subsequently, in 1900–1901 Major Walter Reed, a United States physician, led a team in Cuba composed of James Carroll, Aristides Agramonte and Jesse William Lazear that postulated and confirmed the theory that the virus was transmitted by a mosquito. Thus, YFV was the first virus to be shown transmitted by mosquitoes.

Neutralization tests based on epitopes on the envelope protein have been used to show that flaviviruses consist of eight serological subgroups together with 17 viruses that were not assigned to any serogroup due to weak or no serologic cross-reactivity with other members of the *Flavivirus* genus. Interestingly, although YFV is considered the prototype flavivirus, it is not assigned to a serogroup and is serologically distinct from all other flaviviruses identified to date.

YFV 17D vaccine strain was the first flavivirus genome to be sequenced. Genomic sequencing of flaviviruses has resulted in identification of genetic relationships that closely resemble those of serological relationships. On the basis of genetic relationships, YFV is closely related to nine other flaviviruses (Banzi, Bouboui, Edge Hill, Jugra, Saboya, Potiskum, Sepik, Uganda S and Wesselsbron viruses).

YF virus is classified as biosafety/containment level/risk group- 3 (i.e., indigenous or exotic agents with a potential to cause serious and potentially lethal infection).

Virion Structure

Virus particles are spherical with a lipid envelope and approximately 50 nm in diameter. The lipid envelope is derived from host cell membranes and constitutes approximately 20% of the total weight of the virus particle. Carbohydrates represent approximately 10% of the weight of virus particles and are found as glycolipids and glycoproteins; the envelope (E) protein is glycosylated at none, one or two sites, depending on the strain. The small 12 kd capsid (C) protein surrounds the genome of the virus and the envelope contains two proteins known as E and membrane (M). The E glycoprotein has a molecular weight of 53–59 kDa (depending on strain) and the M protein is 8 kDa. The M protein is thought to reside just below the E protein in virions. Two types

of virions are recognized; mature extracellular virions contain M protein while immature intracellular virions contain precursor M (prM) (27 kDa), which is glycosylated at one site in the “pr” region, that is proteolytically cleaved during maturation to yield M protein. The flavivirus virion is known to “breathe” between these two types of virions while mixtures of the two types are possible too. Cryo-electron microscopy has been used to determine the structure of a number of flavivirus virions but none have been solved for YFV. Recently, in 2019, the first structure of YFV E protein (vaccine strain 17D-204) has been determined and has been found to be similar to that of other flaviviruses.

Genome

The genome consists of a single-stranded, positive-sense (i.e., infectious) RNA of approximately 10,500–11,000 nucleotides with a single open reading frame of 10,233 nucleotides. The genome of the 17D-204 vaccine virus is 10,862 nucleotides in length but some strains have shorter or longer sequences due to variability in the length of the 3′ untranslated region (3′UTR). The 3′UTR is 511 nucleotides in length for 17D-204 vaccine virus and contains approximately 40 nucleotide repeats (RYF) and strains have either none, 1, 2 or 3 copies. In addition, some strains have a duplication of 126 nucleotides in the 3′UTR; thus, the variability in length of the genome. Although the genome is positive-sense RNA and functions as an mRNA there is no terminal poly (A) tract at the 3′ terminus. The 5′ terminus of the genome possesses a type I cap (m^7 GpppAmp) followed by the conserved dinucleotide AG. The 5′UTR is 118 nucleotides in length and there is very little sequence variation between any YFV strain sequenced to date. The genome encodes one single open reading frame that is co- and post-translationally processed by viral and cellular proteases to generate three structural proteins (capsid [C], membrane [M] and envelope [E]) and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) in the gene order 5′NTR-C-M-E-NS1-NS2A-NS3-NS4A-NS4B-NS5–3′NTR (Fig. 1). The

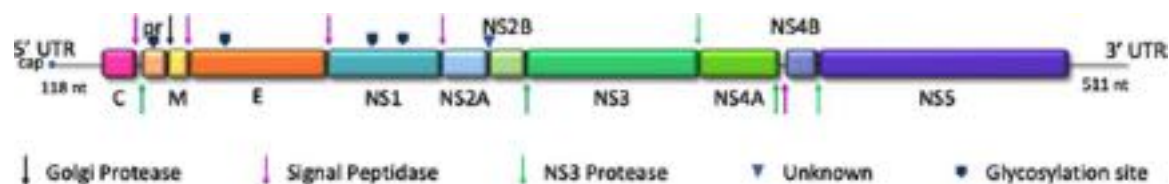


Fig. 1 Genome structure of yellow fever virus. Yellow fever has a genome length of 10,862 nucleotides with some variability between strains due to differences in length of the 3′ untranslated region (3′UTR). The 5′UTR has a length of 118 nucleotides and the 3′UTR has a length of 511 nucleotides for 17D vaccine. Cleavage sites are denoted by arrows corresponding to the enzyme responsible. Golgi protease cleavage is denoted by a black arrow, signal peptidase by a pink arrow, NS3 protease by a green arrow, and unknown enzyme cleavage is denoted by blue triangles. Glycosylation sites are denoted by blue tabs.

Table 1 Structural and Non-structural proteins. Columns show protein length, function, and immune response involvement of the proteins

Protein	Length (AA)	Function	Immune response
Capsid	121	Viral budding by gathering RNA into nucleocapsid	Inhibits Dicer to disrupt RNA silencing
prM	89	Chaperone for E protein to prevent premature maturation	Uncleaved prM-E may contribute to immune evasion; weak neutralizing epitopes
Membrane Envelope	75 493	Part of viral envelope along with E Part of viral envelope along with M; receptor binding; virion assembly	Induces mitochondrial- mediated apoptosis Determines neurovirulent phenotype of virus; cytotoxic T-cell epitope; neutralizing epitopes
NS1	352	Virus particle morphogenesis	Attenuation of TLR3 signaling; suppression of antibody dependent cellular toxicity and complement lysis of infected cells; cytotoxic T-cell epitope
NS2A	224	Virus particle morphogenesis	Reduction of IFN- β gene expression; inhibition STAT1 and STAT2 phosphorylation; cytotoxic T-cell epitopes; viroporin
NS2B	130	Co-factor for NS3 protease activity; induces membrane rearrangement	Inhibition STAT2 phosphorylation; cytotoxic T-cell epitopes; viroporin
NS3	623	Serine protease and helicase	Inhibition STAT2 phosphorylation; cytotoxic T-cell epitopes
NS4A	126	Co-factor for NS3 helicase activity; induces membrane rearrangement	Inhibition STAT1 and STAT2 phosphorylation; viroporin
NS4B	250	Scaffold for replication complex	Inhibition STAT1 and STAT2 phosphorylation; viroporin
NS5	905	RNA-dependent RNA polymerase; methyltransferase; RNA capping	Inhibition of Tyk2 and Jak1 phosphorylation; reduction in STAT2 gene transcription; cytotoxic T-cell epitopes

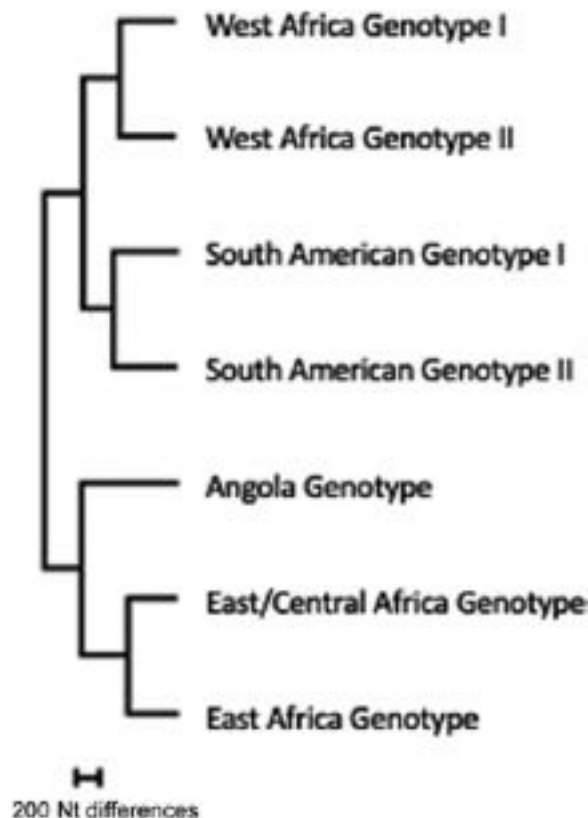


Fig. 2 Wild-Type Yellow Fever Virus Phylogeny. The yellow fever virus genomes are shown in their relation to one another in a Neighbor Joining tree. Nt = nucleotides.

seven NS proteins form the replication complex and have functions related to replication of the virus and interactions with the host immune response. The functions of the NS proteins have been determined and are shown in [Table 1](#).

Phylogenetic studies have identified seven genotypes of YFV ([Fig. 2](#)). There are five YFV genotypes in circulation within Africa: central/east Africa, east Africa, Angola, west Africa I, and west Africa II. Within South America, there are two genotypes: South America I and South America II. While genotypes differ up to 30% in nucleotide sequence, there is very little information on phenotypic differences, including virulence, between genotypes other than South America II genotype is associated with an enzootic transmission cycle whereas the other genotypes are associated with an epizootic transmission cycle.

Life Cycle

YFV has a life cycle that involves both vertebrate and arthropod hosts, with the mosquito acting as the invertebrate host and primates as the vertebrate hosts. Both vertebrate and arthropod hosts are necessary for maintaining the virus in nature.

Mosquitoes are infected when they bite and feed on the blood of a viremic primate. The blood contains virus such that when the mosquito feeds, the virus is taken up into the mosquito midgut. The virus multiplies in the mosquito midgut epithelium and crosses the midgut into the mosquito body into the haemolymph and subsequently disseminate other tissues, including the salivary glands and reproductive tract. When the virus-infected mosquito takes its next blood meal, virus present in the salivary glands is injected into a non-infected primate. The mosquito uses saliva to stop the blood clotting and so virus-infected saliva enters the primate and infection is initiated. It takes between 7 and 14 days for virus to be taken up during one feeding and spread to the salivary glands for transmission to a new primate (a process termed the extrinsic incubation period). If a non-virus-infected mosquito bites the primate during the viremic phase it becomes infected and the transmission cycle continues. The virus-infected primate has a short viremic phase, which is usually during the incubation phase in the primate where there are no clinical signs of infection.

In addition to the mosquito-primate transmission cycle, due to virus-infection of the mosquito reproductive tract, vertical (transovarial) transmission enables the virus to survive in mosquito eggs during the dry season. This is termed overwintering.

The initial events of the virus replication cycle involve the binding of virus to cell receptor(s) that is mediated by the E protein. The cellular receptor molecules have so far not been identified for YFV. Uptake of the virus particle into cells occurs via receptor-mediated endocytosis followed by pH-dependent membrane fusion activity to release the virus nucleocapsid into the cytoplasm.

Table 2 Mosquito and non-human primate species infected with yellow fever virus in Africa and South America

Country	Mosquitoes	Non-human Primates
South America	<i>Haemagogus</i> sp. <i>Sabethes chloropterus</i> <i>Aedes aegypti</i>	<i>Alouatta</i> sp. <i>Ateles</i> sp. <i>Cebus</i> sp. <i>Saimiri</i> sp. <i>Logothrix</i> sp. <i>Saguinus</i> sp. <i>Callithrix</i> sp. <i>Cercocebus</i> sp. <i>Erythrocebus</i> sp. <i>Cercopithecus</i> sp.
Africa	<i>Aedes africanus</i> <i>Ae. fuscifer</i> <i>Ae. vittatus</i> <i>Ae. luteocephalus</i> <i>Ae. simpsoni</i>	Family <i>Cercopithecidae</i> <i>Pan troglodytes</i> <i>Colobus</i> sp. All other African monkey species

The virus replicates in a variety of vertebrate and mosquito cell cultures. Host-cell macromolecular synthesis is not shut-off during YFV replication and is not decreased until cytopathic effect is evident late in the infection process.

The major animal hosts are many species of both New World and Old World primates (see [Table 2](#)). Note that although YFV is described as a mosquito-borne virus, there have been reports of isolation of the virus from *Amblyomma variegatum* ticks in central Africa indicating that the virus replicates in both mosquito and tick cells; however, there is no evidence of human disease caused by tick-borne infection.¹

Epidemiology

YFV is found in 46 countries in tropical South America and sub-Saharan Africa (see [Fig. 3](#)); the virus is not found in Asia. Using phylogenetic analyses, it has been inferred that YFV originated in central or east Africa and then moved westwards across Africa. It was introduced from Africa into the Americas between 400 and 500 years ago, consistent with the theory that the virus was introduced into the New World with the slave trade. Genetic studies have shown that South American viruses evolved from those originated from west Africa.

YFV exists in nature in tropical regions of Africa and South America and is transmitted between non-human primates by diurnally active tree hole-breeding mosquitoes in what is termed the sylvatic cycle. Humans infected as part of this sylvatic cycle have what is termed “jungle yellow fever” ([Fig. 4](#)). The monkey and mosquito species carrying YFV differ by geographic location ([Table 2](#)). The main vector in Africa is *Aedes africanus* while in South America it is *Haemagogus* species. Other mosquito species involved in transmission include *Ae. fuscifer*, *Ae. vittatus*, *Ae. luteocephalus* and *Ae. simpsoni* in Africa, and *Sabethes chloropterus* in South America. Primate species involved as vertebrate hosts in Africa do not tend to get fatal disease while in South America human outbreaks are preceded by epizootics where monkeys succumb to fatal infections. However, not all South America non-human primate species succumb to YFV infection. Urban yellow fever is the term given to a transmission cycle involving *Ae. aegypti* and humans, and takes place in cities as *Ae. aegypti* is a peri-domestic mosquito species ([Fig. 3](#)). Although urban YF is periodically reported in Africa, there have been very small numbers of urban YF cases in the Americas since 1942, even though *Ae. aegypti* has reinfested the majority of tropical South America during the last 60 years.

Countries in South America that have experienced outbreaks of YFV include Brazil, Paraguay, Argentina, Colombia, Bolivia, and Peru with annual activity reported in Brazil, Peru and Bolivia; YF activity is rare in Argentina and Paraguay. The majority of YF cases in the Americas are reported from Peru where the virus is endemic, although outbreaks in Brazil tend to involve higher numbers of cases with large outbreaks identified every 8–10 years. Since the year 2000, in Africa, YF outbreaks have occurred in Guinea, Liberia, Mali, Senegal, southern Sudan, Togo, Uganda, Burkina Faso, Côte d'Ivoire, Ethiopia, the Democratic Republic of the Congo, Angola, and Ghana. While YF activity is most often seen in west Africa, there are increasing numbers of outbreaks in east Africa. For example, the first outbreak in Ethiopia was in 1959, with another epidemic following in 1960 to 1962 followed by a gap of 50 years before an outbreak in 2013.

Not only does the monkey and mosquito species vary by geographic location but the genotype of YFV also varies by geographic location. West African genotype viruses are genetically distinct from those found in east and central Africa; the Angola genotype is the oldest genotype but has only been found in Angola with spillover into the Central African Republic in 2016. The two South American genotypes are geographically separated with South American genotype II found only in Bolivia and Peru as well as a small region of Brazil while South American genotype I found in the remainder of tropical South America. YFV is currently not found in central America and the last cases were in Panama in 1974. Nonetheless, the virus is still found in Trinidad.



Fig. 3 Yellow fever virus distribution in South America and Africa. Countries in yellow denote where yellow fever virus has been previously isolated.

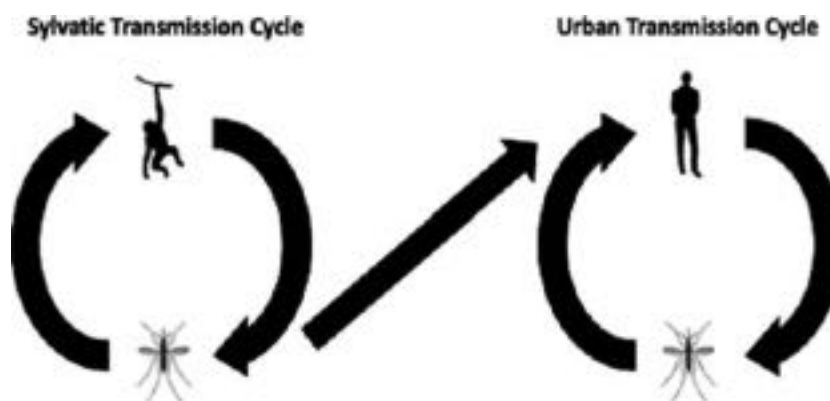


Fig. 4 Transmission cycles. The sylvatic cycle for yellow fever virus is shown on the left where virus circulates between non-human primates and mosquitoes. The middle arrow denotes a spill-over event where YFV can enter a human population and establish an urban transmission cycle between humans and mosquitoes, shown on the right.

The reason for the lack of YFV in Asia is not understood. The mosquito vector, *Aedes aegypti*, is found in Asia and laboratory studies indicate that Asian monkey species succumb to YFV infection and Asian strains of *Ae. aegypti* can transmit YFV. A number of hypotheses have been proposed to explain this phenomenon but none have been proven. In 2016 the first cases ($n = 11$) on YF were reported in China following infection of Chinese workers in Angola and returning home with clinical YF disease. No secondary cases were reported.

YF is considered to be a re-emerging disease due to the increase in the incidence of the disease since a large outbreak in Nigeria in the 1980s. It is estimated that there are approximately 150,000 cases with 80,000 deaths each year of which the vast majority ($> 90\%$) of cases occur in Africa. The case-fatality rate varies between outbreaks from 5% to 20%. There are frequent reports of YF in West Africa each year while there are occasional reports from East Africa; although East Africa has seen an increase in reports in recent years. The majority of YF activity in South America is found in the Amazon Basin with occasional reports from surrounding areas, including Trinidad. Approximately, 200 cases are reported from South America each year with a case-fatality rate of 65%. This high figure is in part due to the majority of cases to be in forest workers and to identification of the disease in some areas by histopathological examination of livers from fatal cases. This has led to a speculation that the true incidence of YF may be 10-fold greater than reported numbers of cases.

YF is considered to be a major public health problem and as such current YF activity is reported in the World Health Organization (WHO) *Weekly Epidemiological Record*, WHO webpages (see "Relevant Website section") and Pan American Health Organization (See "Relevant Website section") webpages are devoted to the current status of YF activity every few weeks. The most recent version of the International Health Regulations (IHR) was published in 2005 and are an international legal document for all 196 member states of the WHO. Because of the fear of international spread of the disease, YF is the only (obligatory) vaccine included in the International Health Regulations.

Clinical Features

As stated above, virus enters the primate via mosquito saliva. Virus multiplication takes place in cells near the point of infection and spreads to lymph nodes. Viremia allows the spread of the virus throughout the host, including spleen, bone marrow, heart, kidneys and liver, but not brain (i.e., wild-type YFV is viscerotropic). The liver is the target organ of the virus where virus initially infects Kupffer cells followed by hepatocytes. Injury to hepatocytes is due to direct virus-induced cell death rather than immunopathology and the disease is characterized (particularly for diagnostics) by eosinophilic degeneration (also known as “Councilman bodies”). Liver necrosis results in a loss of liver functions and as a consequence in severe YF infections is the skin the patients turn yellow; hence the name given to the virus. It should be noted that this is a minority of clinical cases. Clinical disease varies from a mild febrile influenza-like infection to a haemorrhagic fever and fatality in 5%–20% of cases. Fatal cases often show coagulative necrosis, which is characteristic of apoptosis (virus-infected cell cultures show apoptosis). A typical disease involves infection with the virus via a mosquito bite followed by an incubation period of 3–6 days during which viremia takes place such that an infected individual may be incubating the virus and not realizing he/she is infected. As stated above, if a non-infected mosquito bites the viremic human and becomes infected, there is an opportunity for the virus to spread to non-infected humans and continue the transmission cycle, and consequently disease; hence the concerns of YF outbreaks becoming major public health problems.

Clinical symptoms of the disease occur abruptly and include fever (39–40°C), headache, muscular pains, nausea and vomiting that last for the next 3–5 days. Following a brief period of “remission” for up to two days, the patient proceeds to the “intoxication” phase where fever, nausea, vomiting and jaundice take place for another 3–5 days. The latter phase includes an increase in serum transaminase levels (aspartate aminotransferase (AST) and alanine aminotransferase (ALT)) that are associated with liver damage. YF disease is characterized as a higher increase in AST compared to ALT levels. In severe clinical disease there is haemorrhagic fever with bleeding from different orifices, kidney disfunction (detected as increase in albuminuria), and dehydration. Death takes place 7–10 days after the first clinical symptoms. Survivors take 1–2 weeks to convalesce. It is important to note that a minority of cases (<25%) develop a severe disease form.

Virus infection results in a rapid appearance of an immune response. Neutralizing antibodies are detected from around day 7 post-infection and have been shown to correlate with elimination of the virus and improved clinical outcome. The role of the cellular immune response is incompletely understood at the present time. Wild-type infection induces protective immunity, shown by neutralizing antibodies, and there is no evidence of a second infection with demonstrable clinical infection, i.e., a individual can only have one episode of clinical YF disease.

Although primates are the vertebrate host of YF virus, the virus infects a number of animal species. The virus causes viscerotropic disease in the European hedgehog (*Erinaceus europaeus*), which was shown experimentally in the 1930s. Not surprisingly, limited experimental infection experiments have been done with this animal model. The virus causes neurotropic disease in immunocompetent mice, hamsters and guinea pigs usually only following intracerebral inoculation; however, wild-type YFV strains show viscerotropic disease in immunocompromised mice, especially interferon (IFN)- α/β or IFN- $\alpha/\beta/\gamma$ receptor-deficient mice (A129 and AG129, respectively). In addition, wild-type strains that are adapted to golden hamsters cause viscerotropic disease. A number of vertebrates (e.g., sloths) have been found to have antibodies against YF virus, but no viruses have been isolated and their role in the ecology of YF virus is unclear.

Diagnosis

The diagnosis of YF infection is usually made clinically based on typical symptoms such as fever, nausea and vomiting, and in yellow skin discoloration (jaundice) based on liver dysfunction. Skin and eyes appear to have a yellowish tint and there is associated abdominal pain and swelling. In the latter case, liver dysfunction due to the inability of the liver to break down bilirubin is usually an indicator of a severe YF case.

The diagnosis has to be confirmed using laboratory techniques due to the public health importance of a YF outbreak. This is only possible in a few laboratories that have appropriate capabilities (normally national reference laboratories).

Severe cases can be diagnosed by liver biopsy to identify necrosis or immunohistochemistry to determine if the infection is due to YFV. However, liver biopsy is normally undertaken on fatal cases due to the risk of bleeding in severe YF cases.

Viremia takes place only during the incubation period and, although the diagnosis can be made by either virus isolation or RT-PCR of serum samples, a negative result does not necessarily indicate that the disease is not a YFV infection. Serology by ELISA and neutralization test can be used for the diagnosis. Serological diagnosis is usually based on the detection of anti-YF virus IgM by enzyme linked immunosorbent assay (ELISA) or an antigen-capture-ELISA as detection of IgM specific to YFV is considered a good indicator of an acute YFV infection. IgG assays can be used but paired acute and convalescent serum samples are needed to show a four-fold or greater increase in IgG titers to suggest YFV infection. Neutralization tests are usually considered the “gold standard” for serologic assays but require suitable laboratories because of the need of live virus. However, serologic assays have problems due to the extensive cross-reaction between flaviviruses and as such serologic diagnosis has to be done with care.

Virus isolation can be undertaken from liver or serum specimens using mice, mosquitoes or cell culture, although this is only possible during the first 3–4 days of clinical symptoms of the disease when viremia is present. Intrathoracic inoculation of mosquitoes is the most sensitive technique but detection/isolation of infectious virus is rarely used these days as RT-PCR is

available. Because of RT-PCR there are only few virus isolates since 2000. Liver histopathology demonstrating necrosis of hepatocytes and formation of Councilman bodies can be used for the diagnosis, but this is usually done only for fatal cases.

It is important to note that in areas of Africa and South America where the disease is endemic, the initial clinical symptoms are similar to many other tropical diseases, particularly in mild cases. Also, progression to haemorrhagic fever does not always indicate YF as other viral haemorrhagic fevers are found in Africa (e.g., Lassa fever and ebola) and South America (e.g., Argentinian haemorrhagic fever) have similar symptoms and all are considered to be major public health problems. As such, accurate diagnosis of these cases is critical and must be able to identify the specific virus responsible to inform for public health interventions.

Treatment (When Applicable)

There is no specific antiviral treatment to control YF infections, rather supportive therapy is the common practice. Interferons have shown to have activity at controlling YFV infection in primates. The antiviral agent ribavirin has shown antiviral activity in cell culture but has had mixed results in animal studies. The major method of control is to use insecticides to control the mosquito vector. It should be pointed out that most YF cases take place in forest workers who have difficulties seeking treatment that is available in population centers.

Prevention

There is only one strategy to effectively control YF outbreaks and that is vaccination.

Two live attenuated vaccines were developed in 1930s. French workers took the wild-type strain French viscerotropic virus (FVV; isolated from a Lebanese man called Francoise Mayali) and passaged it 128 times in mouse brain to generate the French neurotropic vaccine (FNV). The commercial vaccine was used at passage 260. FNV was characterized by loss of viscerotropic disease, loss of vector competence in *Ae. aegypti*, but had enhanced neurotropism due to mouse brain passage. The vaccine was very immunogenic and was used in Franco-Africa until the early 1980s when it was discontinued due to the incidence of post-vaccinal encephalitis in children under the age of 14 (1 in 333).

The 17D vaccine was developed in 1930s by Max Theiler and colleagues and was responsible, in part, for him being awarded the 1951 Nobel Prize in Physiology or Medicine; indeed this is the only Nobel prize awarded for a vaccine. Theiler *et al.* took the wild-type strain Asibi and passaged it 18 times in embryonic mouse tissue, followed by 50 passages in minced chick embryo tissue, and finally 108 passages in minced chick embryo without nervous tissue (i.e., 176 passages) to develop the 17D attenuated variant. 17D was characterized by loss of viscerotropic disease, loss of vector competence in *Ae. aegypti*, and loss of neurotropism. Theiler tried to repeat the attenuation process but was never successful. The vaccine was tested in the late 1930s and was rapidly licensed.

17D YFV vaccine is produced in embryonated chicken eggs using technology that has changed little from the early 1940s when the seed-lot system was introduced. This was necessary as many versions of 17D were produced and with passage levels beyond 300 it was found that the vaccine became over attenuated and it was no longer immunogenic. In addition, human serum was initially used as a stabilizer and this resulted in some vaccinees contracting hepatitis B from contaminated serum.

Today, the strain 17D is used as three substrains known as 17D-204, 17D-213 and 17DD, and all are used as vaccines. 17D-204 was developed from the 204th passage of 17D while 17D-213 was developed from a the 213–77 seed virus of 17D-204 in 1977 and has an E protein glycosylation site that 17D-204 substrain does not have. 17D-204 and 17D-213 vaccines are used at passage 235–239. 17DD substrain was developed from the 195th passage of 17D and was passaged up to the 286th passage to produce the vaccine. Today the vaccine is only manufactured by six manufacturers. Institut Pasteur in Senegal, Sanofi Pasteur in France (Stamaril™), Wuhan Institute of Biological Products in China, Sanofi Pasteur in USA (YFVax™) use the 17D-204 substrain, Federal State Unitary Enterprise of Chumakov Institute in Russia uses 17D-213 substrain, and Bio-Manguinhos in Brazil uses 17DD substrain. The Chinese and Sanofi Pasteur (USA) vaccines are only used in domestic markets whereas the other four vaccines are prequalified by the WHO. Prequalification is a procedure by the WHO that ensures manufacturing sites, quality control laboratories and contract research organizations meet international standards for safety, quality and performance. Thus, the four prequalified vaccines are used internationally for vaccination programs in the 46 YF endemic countries. It should be pointed out that the two domestic vaccines are not-inferior, rather they have not been subjected to the international prequalification process. Recent years (2016–2018) has seen exhaustion of vaccine supplies due to increased outbreaks and fractional dosing has been used in emergency scenarios in Angola, Democratic Republic of Congo and Brazil. In situations of fractional dosing, one-fifth of vaccine dose is used for vaccination. Practically, this has involved giving 0.1 ml from a full 0.5 ml dose, i.e., one dose is used to immunize five individuals. All data indicates that this modified vaccination regimen was successful.

The 17D vaccine is one of the finest vaccines developed to date. Over 900 million doses have been distributed in the last 80 years. One dose of vaccine containing approximately 10,000 international units of vaccine induces protective immunity by 10 days post vaccination. The vaccine is given in a volume of 0.5 ml by a subcutaneous route. Until 2016, the WHO required booster vaccination every ten years but it was decided that protective immunity is probably life-long as demonstrated by studies in volunteers and that immunocompetent individuals probably do not need booster doses; however, booster doses are recommended for certain populations. There are occasional serious adverse events associated with the vaccine. There are allergic reactions

to egg antigens present in the vaccine and the vaccine is contraindicated in pregnant women, immunosuppressed individuals and children under the age of 6–9 months, depending on the country, due to the risk of post-vaccinal encephalitis. Two types of disease have been associated with the vaccine: yellow fever vaccine associated neurotropic disease (YEL-AND) and yellow fever vaccine-associated viscerotropic disease (YEL-AVD). Both occur at frequencies around 1 in 200,000–500,000, depending on the country studied. YEL-AND is associated with a reversion of the vaccine virus while YEL-AVD clinically resembles wild-type YF disease and appears to be due to unknown host factors. YEL-AND is rarely associated with a fatal outcome while YEL-AVD has a fatal outcome in approximately 60% of the cases. It should be noted that YEL-AND and YEL-AVD are only seen in primary vaccinees.

The molecular basis of attenuation and virulence of YFV has been investigated. The genome of wild-type strain Asibi has been compared to the 17D-204, 17D-213 and 17DD vaccine substrains, and there are 20 common amino acid substitutions between the 17D vaccine substrains and wild-type strain Asibi plus four nucleotides in the 3' non-coding region. Eight of the 20 amino acid differences are found in the E protein suggesting that this protein is involved in the attenuation of wild-type strain Asibi. Comparison of the genome of wild-type FVV with vaccine strain FNV also reveals 20 amino acid substitutions. Neither attenuation of Asibi nor FVV vaccine viruses involved nucleotide changes in the 5' non-coding region of the genome suggesting that this region is not involved in the attenuation of virulence of wild-type YF virus. Although Asibi and FVV were attenuated by very different routes (chicken and mouse tissue, respectively), both attenuation processes were found to share two common amino acid substitutions at M protein amino acid 36 and NS4B protein amino acid 95. It is unknown whether these common substitutions were fortuitous or are involved in the attenuation of viscerotropism of wild-type YF virus and/or loss of mosquito competence of the vaccine viruses.

The mechanism of protective immunity is poorly understood, although the production of neutralizing antibodies has been shown to be the correlate of protective immunity of 17D vaccine. Most neutralizing antibodies recognize epitopes of the E protein with some weak neutralizing antibodies recognizing the prM protein. Cell-mediated immunity correlates with nonstructural proteins, in particular NS3. Experimental studies have demonstrated that NS1 protein can induce protective immunity. The mechanism does not involve neutralization and is thought to be due to antibody-dependent cell-mediated cytotoxicity or complement-fixing activity of anti-NS1 antibodies.

Further Reading

- Barrett, A.D., Higgs, S., 2007. Yellow fever: A disease that has yet to be conquered. *Annual Review of Entomology* 52, 209–229.
- Monath, T.P., Vasconcelos, P.F.C., 2015. Yellow fever. *Journal of Clinical Virology* 64, 160–173.
- Philip, C.B., 1930. The experimental transmission of yellow fever by mosquitoes. *Science* 71 (1850), 614–615.
- Rice, C.M., Lenches, E.M., Eddy, S.R., *et al.*, 1985. Nucleotide sequence of yellow fever virus: Implications for flavivirus gene expression and evolution. *Science* 229 (4715), 726–733.
- Theiler, M., Smith, H.H., 1937. The use of yellow fever virus modified by in vitro cultivation for human immunization. *Journal of Experimental Medicine* 65 (6), 787–800.

Relevant Websites

- <https://www.paho.org>
Pan American Health Organization: PAHO/WHO.
- <https://www.who.int>
World Health Organization: WHO.
- <https://www.cdc.gov/yellowfever/index.html>
Yellow Fever
CDC.
- <https://medlineplus.gov/ency/article/001365.htm>
Yellow fever: MedlinePlus Medical Encyclopedia.
- <https://www.who.int/news-room/fact-sheets/detail/yellow-fever>
Yellow fever
World Health Organization.

Zika Virus (*Flaviviridae*)

Nikos Vasilakis, Shannan L Rossi, Sasha R Azar, and Irma E Cisneros, The University of Texas Medical Branch, Galveston, TX, United States

Cassia F Estofolete and Mauricio L Nogueira, São José do Rio Preto School of Medicine, São José do Rio Preto, Brazil

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Antibody-dependent enhancement Occurs when non-neutralizing antiviral proteins facilitate virus entry into host cells, leading to increased infectivity in the cells.

Arboviruses Arthropod-borne viruses. In the case of arboviruses that infect humans, mosquitoes are the most common vectors.

Congenital Zika syndrome A distinct pattern of birth defects and disabilities observed with ZIKV infection during pregnancy and may include microcephaly, ocular damage, congenital contractures (e.g., arthrogryposis or clubfoot) and hypertonía (restriction of body movement after birth).

Enzootic Refers to the maintenance of a transmission cycle among non-human vertebrate hosts in a specific geographical region. Sometimes it may also refer to the disease caused by pathogens that circulate enzootically.

Guillain-Barré syndrome A rare, autoimmune disorder in which a person's own immune system damages the nerves, causing muscle weakness and sometimes paralysis.

Hemogram Also known as a complete blood count is a broad screening panel that checks for the presence of any diseases and infections in the body. Hemogram tests mainly

the three components of the blood namely red blood cells, white blood cells and platelets.

Hyperendemic A geographic region where circulation of several pathogens (e.g., arboviruses) is observed throughout the years with a high incidence and/or prevalence affecting most or all age groups of the population.

Microcephaly A condition where a baby's head circumference is much smaller than the normal. It occurs due to genetic abnormalities or by exposure of the developing fetus to drugs, alcohol, certain viruses, and toxins that damage the developing brain tissue.

Sylvatic Refers to a form of a disease or virus that occurs in wild animals (in the case of Zika virus, in non-human primates) in the forest.

Vertical transmission Refers to transmission from parent to offspring via the ovaries.

Zone of emergence An ecological transition zone that is adjacent to sylvatic environments. Herein, prevailing ecological conditions, such as vector abundance and closeness for human contact, enable the cross-species transmission of viruses from animals to humans.

Introduction

Zika virus (ZIKV), a flavivirus closely related to West Nile, dengue and yellow fever viruses, remained in relative obscurity since its discovery in 1947 in Ziika Forest preserve, Uganda, during an investigation on the ecology of yellow fever and to identify additional arboviruses. In April of that year, one of the sentinel *Rhesus* monkeys placed into the canopy of the forest developed fever and viremia, leading to ZIKV isolation following intracranial inoculation of its blood into infant mice. Several months later, in January 1948, the virus was also isolated from the canopy-dwelling mosquito, *Aedes africanus*, collected at the same site. Characterization of the human Zika disease and isolation of the virus from the serum of a febrile patient occurred in Nigeria in 1954. The disease manifestations were characterized by fever, headache, diffuse joint pain, and in one case, mild jaundice. The observed mild febrile illness associated with natural infection was confirmed two years later during an experimental infection of a yellow fever-vaccinated human volunteer, who received the mouse-passaged virus isolated from one of the Nigerian patients during the 1954 outbreak. The first direct detection of ZIKV in Asia coincided the isolation of the virus from *Ae. aegypti* mosquitoes collected in Malaysia in 1966, suggesting active transmission by an urban vector. It took another 11 years for the identification and diagnosis of the first human infections in Asia, when seven patients in the island of Java, Indonesia, presented with fever, malaise, stomach ache, anorexia, and dizziness. Twenty years later, indirect evidence of ZIKV circulation among wild orangutans in Borneo was suggested.

Moreover several human serosurveys had suggested that ZIKV may have a more expansive presence in both Africa and Asia. However, one limitation of the serology-based assays used is their inability to demonstrate specificity for a particular virus exposure, as extensive crossreactivity exists among flaviviruses (e.g., dengue virus, Japanese encephalitis virus and others) known to circulate in these regions. However, despite both the direct and indirect lines of evidence described above, up to 2007 there were only 14 documented human cases of ZIKV infections. In 2007, ZIKV re-emerged to cause a series of epidemics in Africa (Gabon), Southeast Asia (Yap island), followed by French Polynesia and its introduction in the Americas as early as 2013 and quickly reaching pandemic levels throughout the Western Hemisphere (Fig. 1). It was declared by WHO a Public Health Emergency of International Concern (PHEIC) in February of 2016. The unprecedented numbers of people infected during the recent ZIKV outbreaks in the South Pacific and the Americas have been accompanied by the emergence of otherwise rare disease manifestations, notably congenital Zika syndrome (CZS) and Guillain-Barré syndrome (GBS).

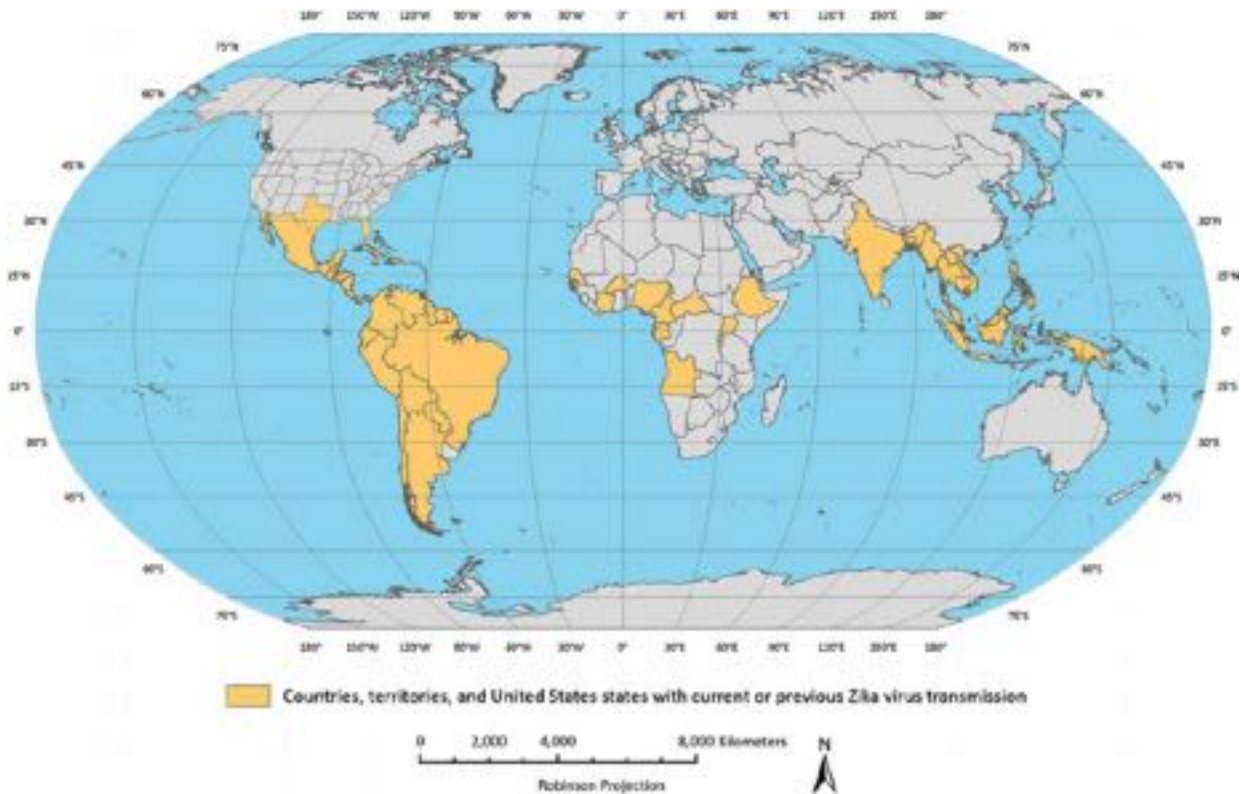


Fig. 1 Current map of global distribution of Zika virus. Courtesy of Dr. Michaela Bueneman, New Mexico State University.

Classification

Viruses belonging to the family *Flaviviridae* comprise a single stranded RNA genome of positive polarity and have a wide host range that includes both vertebrates and invertebrates, as well as a broad geographic distribution throughout the world, except Antarctica. They are recognized as pathogens with significant pathogenic manifestations in human and animal populations. Within this family there are five genera, *flavivirus*, *hepacivirus*, *jingmenvirus*, *pestivirus*, and *pegivirus*, in which the only recognized invertebrate hosts are mosquitoes and ticks. Arthropod-borne (arboviruses) members are exclusively found within the genus *Flavivirus*. Yet despite the extensive divergence among these viruses, they share important similarities in genome structure, virion morphology, and life cycle. There are more than 80 recognized species within the genus *Flavivirus*, which cluster within one of four major clades based upon the taxonomy of their hosts and their mode of transmission: (1) transmission among vertebrate hosts by mosquitoes (e.g., ZIKV), (2) transmission among vertebrate hosts by ticks (e.g., tick-borne encephalitis virus), (3) transmission among vertebrates without any known vector (no known vector) (e.g., Rio Bravo virus), and (4) restricted host range transmission among arthropods (insect-specific) without the involvement of vertebrates (e.g., cell-fusing agent virus). Viruses of public health importance (e.g., ZIKV, yellow fever and dengue viruses) cluster within the mosquito-borne group and belong to a subgroup primarily transmitted by *Aedes spp.* mosquitoes. As discussed below in Section “Ecology and Transmission Cycles” (*Ecology and transmission*), all of these viruses originated in sylvatic, enzootic cycles, in Asia and Africa respectively, maintained in non-human primates and forest-dwelling *Aedes spp.* mosquitoes, and have a history of successful emergence into sustained transmission among humans by the anthropophilic *Ae. aegypti* mosquitoes.

Virion Structure and Genome Organization

Mature ZIKV virions are spherical particles with a diameter of approximately 50 nm (500 Å). The particles are comprised of a host-derived lipid bilayer in which 180 copies of the envelope (E) protein are inserted and folded into a head-to-tail dimer confirmation. The viral prM serves as a scaffold during virion production and becomes cleaved to M protein during the maturation process. Within the particle, the RNA genome is encapsulated within the nucleocapsid comprised of capsid (C) protein (**Fig. 2(A)–(C)**).

As with other members of the genus *Flavivirus*, ZIKV has a linear, positive-sense, single-stranded RNA genome of about 11 kilobases in length, allowing for the genome to act directly as a messenger RNA within infected cells. Organizationally, the genome consists of a single open reading frame (ORF), coding for a polyprotein precursor that is cleaved by both viral and host

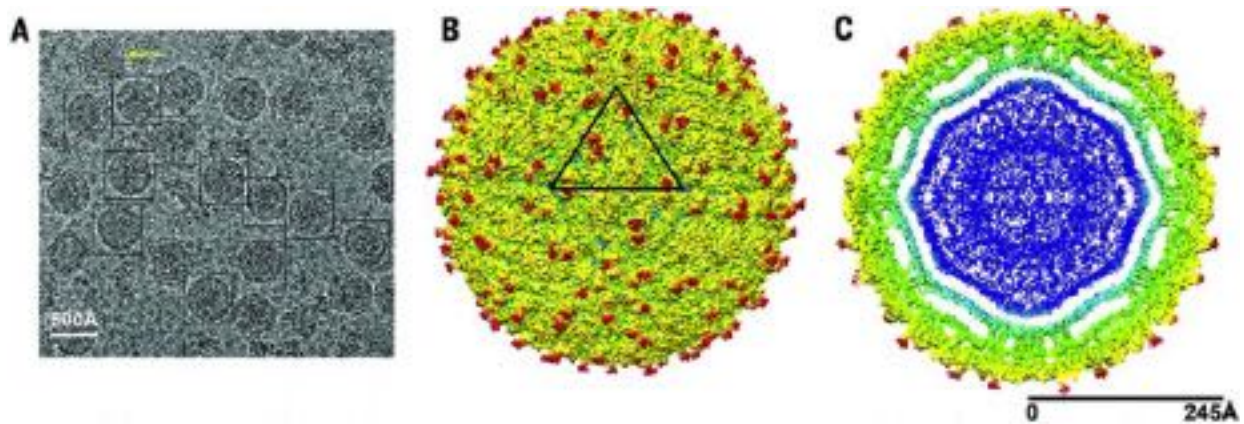


Fig. 2 The cryo-EM structure of Zika virus at 3.8 Å resolution. (A) Cryo-EM image of ZIKV showing the distribution of virion phenotypes, including mooth, mature virus particles (black box), and partially mature virus particles (yellow arrow). (B) Surface-shaded depth representation of ZIKV viewed down the icosahedral two-fold axis. The asymmetric unit is identified by the black triangle. (C) Cross-section of ZIKV showing the radial density distribution. From Sirohi, D., Chen, Z., Sun, L., *et al.*, 2016. The 3.8 Å resolution cryo-EM structure of Zika virus, *Science* 352 (6284), 467–470. Reprinted with permission from AAAS.

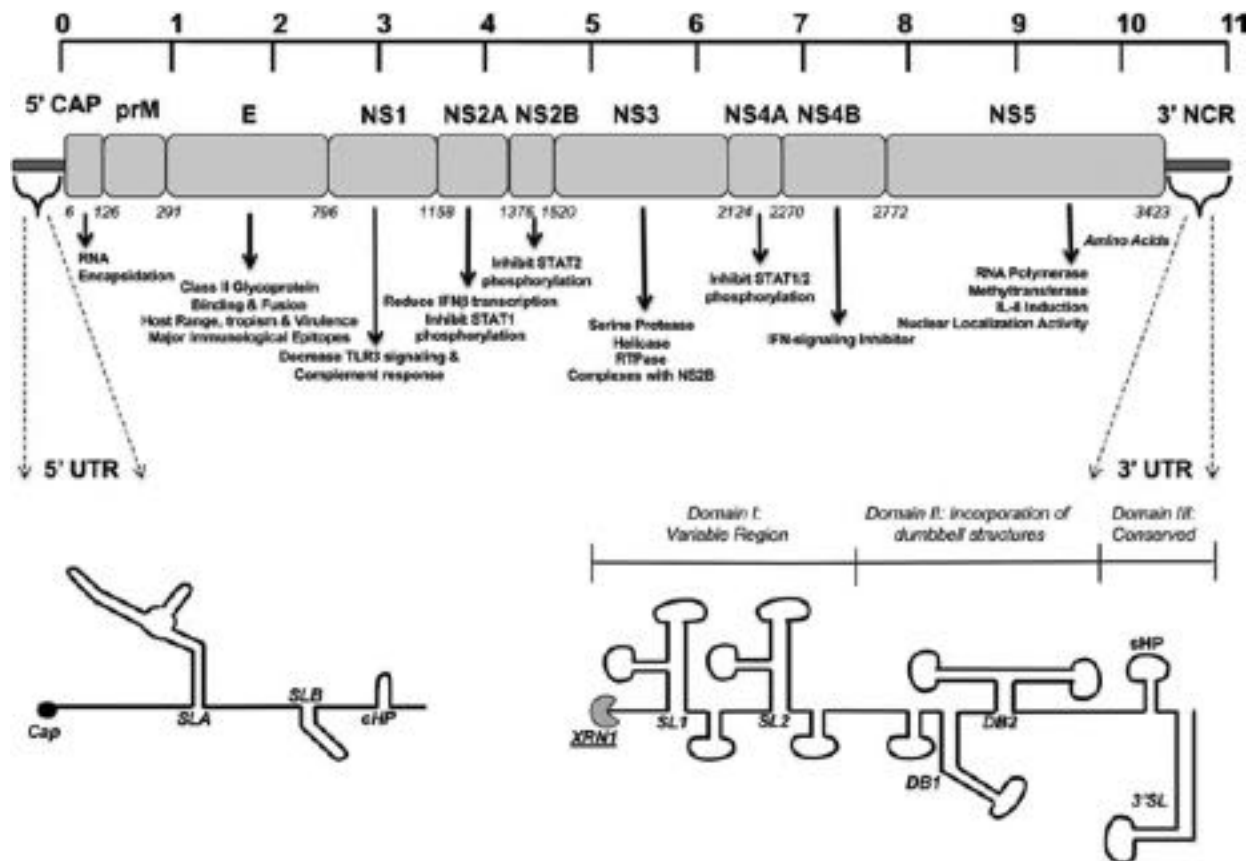


Fig. 3 The ZIKV genome and the major functions of the gene products. SLA, stem loop A; SLB, stem loop B; UAR, upstream AUG element; cHP, capsid coding region hairpin element; xRNA, exoribonuclease XRN1 resistant RNA; DB, dumbbell structures; RCS, repeated conserved sequence; SL, stem loop; SHP, small hairpin element.

proteases co- and post-translationally to produce three structural [capsid (C), pre-membrane/membrane (prM/M) and envelope (E), and seven nonstructural proteins (NS1, NS2A, NS2B, NSP3, NS4A, NS4B, and NS5)] (Fig. 3). The coding region of the genome is flanked on both 5' and 3' ends by non-coding or untranslated regions (NCRs/UTRs). The 5' region contains both a cis-acting element promoting vRNA replication as well as a type I methylated cap (m-7GpppAmp) responsible for the initiation of

translation while 3' UTR has a terminal 5'-CUOH-3' rather than a polyadenylated tail. Additionally, both 5' and 3' UTRs possess secondary structural elements that are critical to various viral processes such as translation, replication, packaging, and immune evasion (Fig. 3).

Flaviviral non-structural proteins perform a variety of functions covering genome assembly, genome transcription and innate immune evasion, many of which are discussed below. NS1 has multiple functions and multiple forms; the protein can be found as a membrane-associated monomer, plasma membrane-associated dimer and a secreted hexamer. NS1 is required for RNA replication, contributes to immune evasion and also is an immunogenic target for the adaptive immune response. As such, NS1 has been utilized for diagnostic serology tests. NS2B, NS4A, and NS4B are required for viral genome replication. NS3, along with the NS2B cofactor, form the RNA helicase and protease. NS5 functions as a methyltransferase and an RNA-dependent RNA-polymerase (RdRp). Many ZIKV non-structural proteins have secondary functions in directly antagonizing the innate immune response at the type I interferon (IFN) signaling level (Fig. 3).

Life Cycle

Ecology and Transmission Cycles

Like other arboviruses (e.g., dengue or yellow fever viruses), ZIKV exists in two ecologically distinct transmission cycles. The first, enzootic or sylvatic cycle, the virus circulates between canopy dwelling arboreal *Aedes* species (*spp*) mosquitoes and non-human primates, such as patas (*Erythrocebus patas*) or African green monkeys (*Chlorocebus aethiops*). Until 2007, our knowledge of ZIKV was derived almost exclusively from the sylvatic transmission cycle. As mentioned in the introduction, the first documented ZIKV isolation was from *Ae. (Stegomyia) africanus* mosquitoes in Uganda and subsequently in the Central African Republic (CAR) and Senegal, implying a central role of this species as the principal enzootic vector (the other being *Ae. (Stegomyia) luteocephalus*) of transmission (Fig. 4). ZIKV has been also isolated from several other arboreal *Aedes* spp., including *Ae. (Stegomyia) opok* in the CAR, *Ae. (Stegomyia) apicoargenteus* in Uganda, *Ae. (Stegomyia) luteocephalus* in Nigeria and Senegal, and *Ae. (Fredwardsius) vittatus*, *Ae. (Diceromyia) furcifer*, *Ae. (Stegomyia) aegypti formosus* in Côte d'Ivoire and Senegal, as well as *Ae. (Aedimorphus) dalzieli*, *Ae. (Aedimorphus) hirsutus*, *Ae. (Stegomyia) unilineatus*, *Ae. (Stegomyia) metallicus* (Fig. 4). A single report has suggested that non-Aedene species mosquitoes, such as *Anopheles* (*Meigen*) *coustani* and *Mansonia* (*Mansonioides*) *uniformis*, that inhabit various rural ecotypes may be vectors of ZIKV. Lastly, there is a report of ZIKV isolation from a single pool of *Culex* (*Culex*) *perfuscus* mosquitoes in Senegal, an observation that requires further verification (the controversy of *Culex* spp.

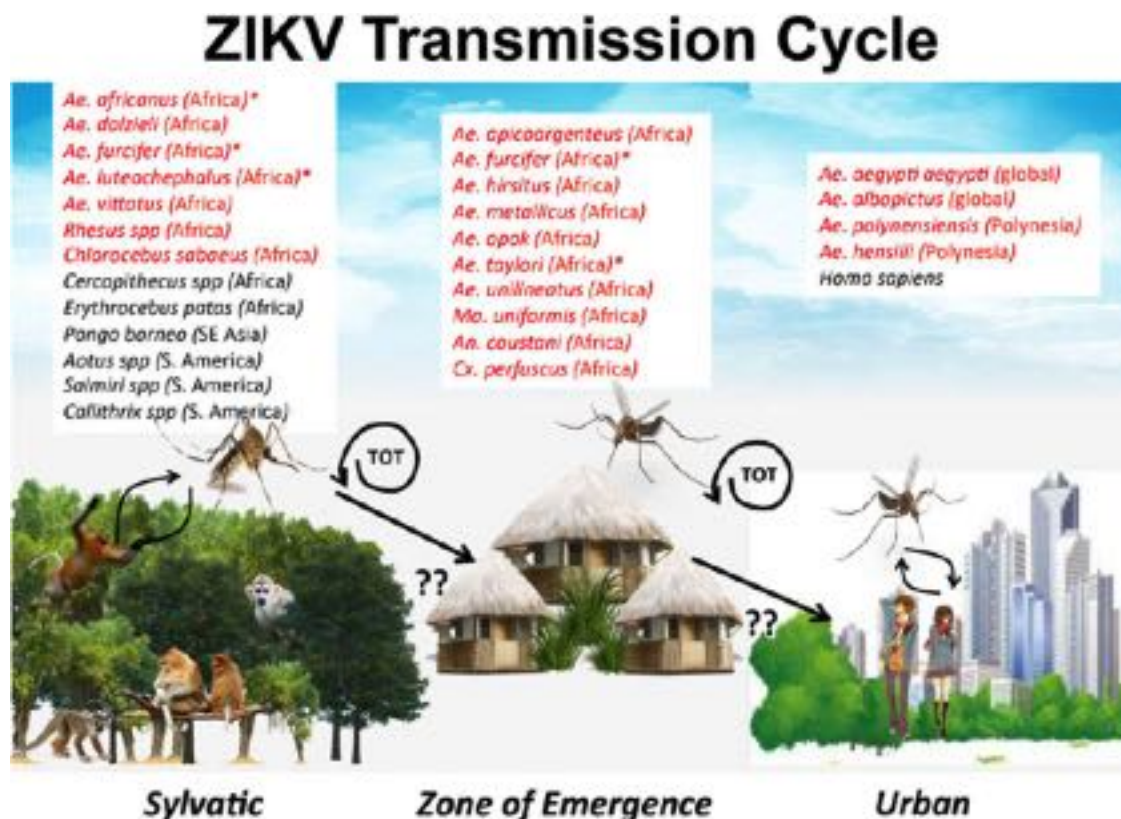


Fig. 4 Transmission cycles of Zika virus.

mosquitoes as vectors of ZIKV transmission will be examined in more detail in the urban transmission cycle below). The isolation of ZIKV from *Ae. (Fredwardsius) vittatus* mosquitoes sampled in a village within the 'zone of emergence' supports its putative role as a bridge vector of enzootic ZIKV strains into the human transmission cycle (Fig. 4).

The zone of emergence, is defined as an ecological transition zone that is adjacent to sylvatic environments, where humans habitate small agricultural settlements and frequently venture into the forest for hunting or logging and thus the potential for infection is high. The infection of humans by enzootic virus is referred to as a spillover event (Fig. 4). The risk, therefore, exists for sylvatic viruses to break away from the sylvatic cycle and perpetuate themselves indefinitely in a mosquito-human-mosquito transmission cycle in urban settings. Conversely, humans infected with ZIKV from an urban cycle have the potential to infect susceptible sylvatic mosquitoes and allow the virus to enter a previously "naïve" sylvatic environment, a phenomenon known as spillback (e.g., the spillback of YFV into the forests of South America after its importation into the New World during the slave trade or as recently demonstrated for ZIKV in various New World monkey species in Brazil).

The second, human or urban cycle, the virus circulates between humans and the anthropophilic *Ae. (Stegomyia) aegypti* mosquito, as well as the peridomestic *Aedes (Stegomyia) albopictus* mosquito, as documented by field surveillance and experimental studies. The latter was implicated as the main vector of peridomestic transmission in Gabon in 2007, and subsequent experimental studies supported the role of Asian *Ae. albopictus* populations as competent ZIKV vectors. It is a highly invasive species, fueled by global trade, and over the past 40 years has expanded significantly its global geographic distribution in tropical as well as temperate settings, thus becoming a significant ZIKV vector. Other minor vectors of transmission, such as *Ae. (Stegomyia) hensilli* and *Ae. (Stegomyia) polynesiensis*, may serve as secondary vectors in niche ecotypes (Fig. 4).

To explain the spectacular global spread of ZIKV, it was suggested that ZIKV may be utilizing other than *Aedes* spp. mosquitoes as vectors of transmission. A number of unsubstantiated reports around the world suggested that *Culex quinquefasciatus* mosquitoes may be competent vectors of ZIKV transmission. However, several experimental studies utilizing *Culex* mosquito populations from Brazil, Italy, Tunisia, and the U.S. demonstrated that these mosquitoes, as well as *Culex pipiens* are refractory to ZIKV infection and incapable of transmission. While factors such as the mosquito microbiome or genetic differences in various *Culex* spp. mosquito populations may contribute to the conflicting observations, several experimental studies have demonstrated that *Culex* spp. mosquitoes are incompetent vectors of ZIKV transmission, and no evidence of natural transmission during ZIKV outbreaks in the Americas, Southeast Asia, and Africa has been reported.

Other modes of transmission than mosquito-vertebrate-mosquito transmission have been documented to facilitate ZIKV transmission, although their contribution to transmission remains low to insignificant. Vertical (transovarial or transovum) transmission (VT) of mosquitoes had been demonstrated in both transmission cycles and may provide a mechanism for the maintenance of the transmission cycles in interepidemic periods. There is both field and laboratory evidence that many arboviruses of public health significance are vertically transmitted in their natural arthropod vectors, including ZIKV, dengue, Japanese encephalitis, yellow fever, and West Nile viruses.

Sexual transmission has also been demonstrated in ZIKV infections and can occur in both asymptomatic and symptomatic infections. It has been primarily observed in male-to-female and rather infrequently in female-to-male or male-to-male transmission. Current estimates suggest that 1% of reported ZIKV infections were acquired through sexual transmission, although assessing the true burden of this mode of transmission in endemic settings is quite challenging. However, ZIKV RNA has been detected in semen by molecular methods up to 370 days after onset of illness. To date there is no evidence to suggest that ZIKV presence in the semen leads to sterility, albeit alterations in semen and sperm quality have been observed.

Lastly, ZIKV blood-borne transmission has been documented through blood transfusion, and while ZIKV RNA has also been detected in breast milk, milkborne transmission has not been confirmed as a mode of transmission.

Virus Replication Cycle

The ZIKV replication cycle begins with the binding of the E protein to a putative receptor on host cells. The exact cellular receptor responsible for cognate binding differs based on the cell type in question, although AXL, Tyro3, and TIM1 appear to play a role in a large population of cells. Receptor binding results in internalization of the virion via receptor-mediated endocytosis. Following the acidification of the endocytic vesicle, the E glycoprotein undergoes a conformational change such that the viral envelope fuses with the endocytic membrane, allowing for the release of the + ssRNA genome into the host cytosol. Translation begins immediately in the cytosol upon release with a N-terminal ER localization signal on the forming polyprotein driving association of the ribosomes and polyprotein with the ER membrane (Fig. 5). This association with the ER membrane allows for insertion of the polyprotein into the ER membrane concurrently with translation, distributing prM, E, NS1, and parts of NS2A, NS4A, NS4B within the ER lumen, while C, NS3, and NS5 protrude into the cytoplasmic side of the ER. Stretches of NS2A, NS2B, and NS4B are localized within the ER bilayer (Fig. 5, inset). The cytosolic facing elements of the polyprotein are cleaved into functional proteins by the viral NS2B-NS3 protease complex, while host signal peptidase/signalase and other cellular proteases within the ER lumen cleave the luminal facing polyprotein elements. Cleavage of the polyprotein into functional effectors provides for the assembly of the NSPs into the replication complex and allows for subsequent propagation of the viral genome. NS5, the RNA-dependent RNA polymerase (RdRp) transcribes the positive sense genome to produce a dsRNA intermediate comprised of both the positive and negative sense genomes. The NS3 helicase unwinds the dsRNA intermediate producing both the positive and negative strand

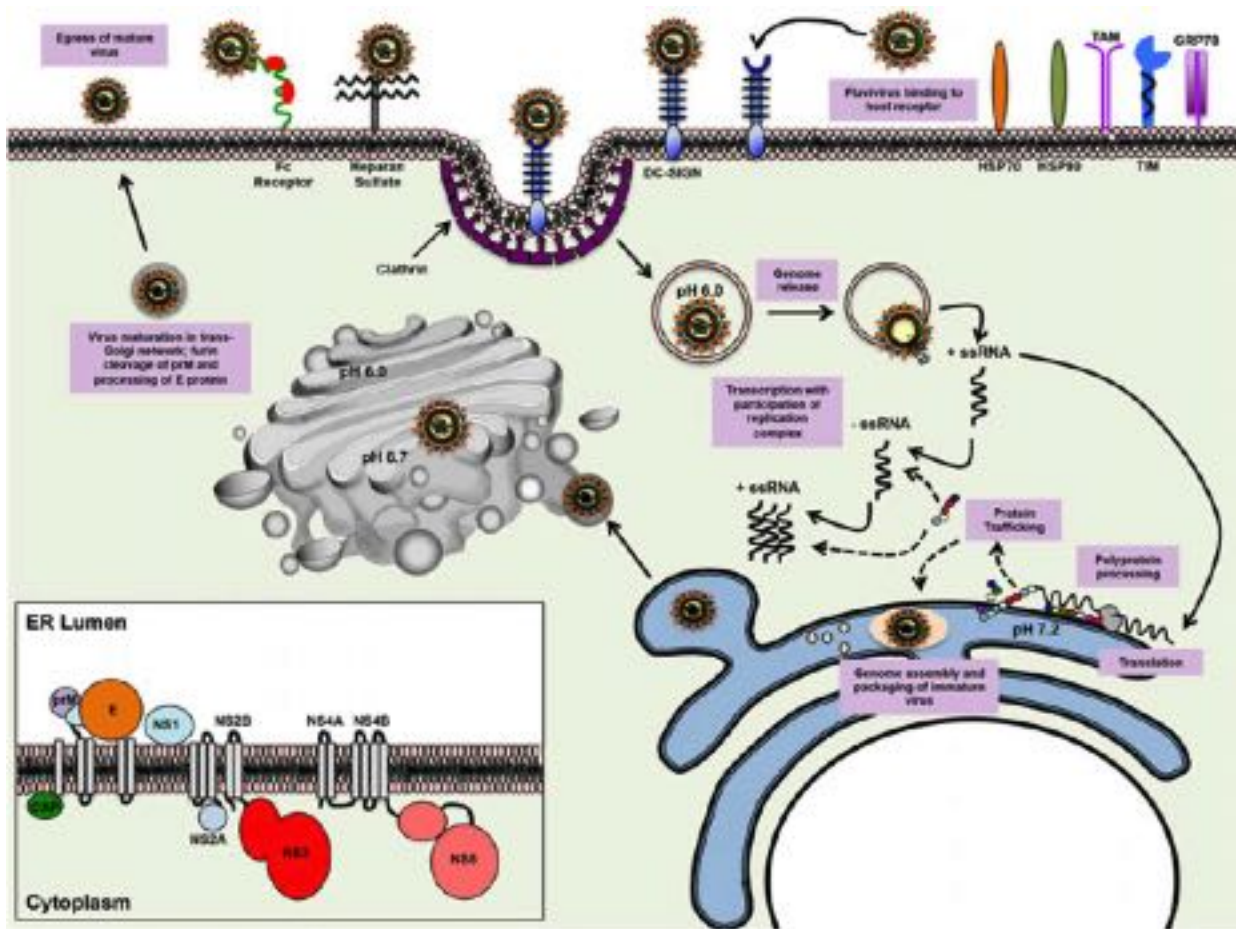


Fig. 5 Zika virus replication cycle.

genomes as ssRNAs, with the negative sense genome serving as the template for NS5 to produce subsequent dsRNA intermediates which are then unwound by NS3. Multiple copies of the positive sense genome are produced and receive their 5' m-7GpppAmp cap structures via the methyltransferase activity of NS5. The capped positive sense RNA genomes and structural proteins prM, C, and E assemble near the ER membrane and assemble into immature particles that bud into the ER lumen where the particles undergo glycosylation. Immature particles then bud out of the ER into the golgi complex and the Trans Golgi Network (TGN), where the acidic environment exposes the prM cleavage site and allows for furin protease to cleave the M protein. The functionally mature particles then exit the TGN and are released from the infected cell into the circulation via exocytosis (Fig. 5).

Epidemiology

Brief History

As indicated in the introduction, ZIKV was first identified and isolated in 1947 in the Ziika forest of Uganda from a sentinel rhesus macaque during a campaign by the Rockefeller Foundation to identify sylvatic yellow fever virus in the area. During the decades that followed, no major outbreaks associated with Zika disease were described, but seroprevalence studies showed evidence of unreported infections. Until the first major outbreak was described in 2007 in both Gabon and Yap Island, the virus remained in relative obscurity as up to that time there were only 14 documented cases of Zika virus infections were reported in the literature. Nonetheless, small focal outbreaks were noted throughout the 2000s into the mid 2010s in southeast Asia. In 2013, an outbreak affecting approximately 28,000 people was noted in French Polynesia, prior to its introduction in the New World shortly thereafter.

While the exact date of ZIKV introduction in the Americas is not known, coalescent phylogenetic analyses suggest that the introduction took place as early as in 2013. The hardest hit area was the State of Bahia, Brazil, where early prevalence estimates suggested 4.4 cases for every 1000 individuals, and served as the epicenter for the world's attention to congenital Zika syndrome (CZS), prompting retrospective epidemiological studies to conclude that CZS was also present in the 2007 epidemic of Yap Island, but went unnoticed. Once ZIKV reached a foothold in Brazil, it quickly spread throughout South America, Central America,

Mexico, and the Caribbean, estimated to have infected approximately 1 million people. At the time the causal link between ZIKV infection and birth defects among newborns prompted the WHO to declare in February of 2016 a PHEIC, an emergency that lasted till November 2016. As of January 2018, more than 3700 cases of congenital birth defects had been reported in the Americas. However, it is worth noting that accurate estimates of the burden of birth defects and GBS attributable to ZIKV infection have been hindered by the lack of systematic surveillance figures for these syndromes before and during the pandemic. As of the end of February 2020, autochthonous ZIKV transmission has been reported in 87 countries and territories in tropical and subtropical areas, including the states of Texas and Florida in the continental United States of America.

ZIKV Lineages and Disease Association

Phylogenetic analyses based on complete open reading frames reinforce the African origin of ZIKV which diverged into three major lineages: African, Asian and American lineages. The African lineage falls into two distinct groups: (1) the Uganda cluster, which is anchored by the prototype strain MR766 and (2) the Nigeria cluster, which includes strains isolated in Nigeria and Senegal. The Asian lineage is anchored by the prototype strain P6-740 isolated in Malaysia in 1966 and includes strains isolated throughout southeast Asia and Oceania, supporting the presence of the Asian lineage throughout Southeast Asia. Within this cluster, a new American lineage has emerged and includes strains sampled throughout the Western Hemisphere, as well as Italy and China, reinforcing the role of traveling and transportation in the rapid spread of these viruses in a global scale. To date, congenital birth defects have been only reported by ZIKV strains belonging to the Asian and the newly emerged American lineage. This may, however, be due to demographic factors i.e., that when the virus was newly introduced to Asia and the Americas many pregnant women were infected.

Potential Factors Contributing to ZIKV Emergence

The recent ZIKV pandemic has generated a number of baffling knowledge gaps centered around the paucity of recognized outbreaks since its discovery in 1947, as well as factors contributing to the emergence of the congenital and neurologic complications associated with infection (see Section “Clinical Features” below). Have these complications arisen due to new emerging phenomena or had they occurred previously but went undetected due to limited surveillance common in resource poor settings or infrequent transmission. Two non-mutually exclusive hypotheses sought to explain these knowledge gaps: (1) ZIKV underwent adaptive evolution to enhance infectivity of urban *Aedes (Stegomyia)* spp. vectors and enhance human viremia, which would not only enhance transmission efficiency but could increase the risk of transplacental fetal transmission. Various experimental studies demonstrated that *Aedes (Stegomyia)* spp. are capable of transmitting ZIKV, and their competence is dependent on the geographic origin of both mosquito populations and virus lineage, with African ZIKV strains typically being more infectious for both vectors as well as animal models than Asian or American strains. Furthermore, a number of mutations including A188V and T233A have been identified in the NS1 of American lineage strains. A188V has been associated with increased secretion of NS1 and an increased infectivity in mosquitoes. Furthermore, ZIKV strains with the A188V mutation are able to antagonize IFN- β , a key antiviral cytokine produced during viral infection, which is critical to mounting an effective innate immune response. Sequence analysis from a ZIKV-infected fetal brain identified an NS1 T233A mutation, which has been shown to destabilize the NS1 dimer; and (2) the stochastic introduction of ZIKV into large immunologically naïve populations allowing for uncontrolled spread. In order to address the immunity of the human populations in the Americas, it is important to realize that dengue virus (DENV), a closely related flavivirus shares antigenic similarities and has overlapping geographic distributions with ZIKV. The effect of pre-existing flavivirus immunity elicited by natural infection or vaccination on ZIKV infection outcomes is a highly debated subject. Several investigations have produced contradictory data with respect to whether DENV infection elicits a protective immune response against ZIKV infection or leads to exacerbation of the disease via antibody-dependent enhancement (ADE) mechanism. Most importantly, prospective studies in humans in Nicaragua and Brazil demonstrated that prior exposure to DENV reduced rather than enhanced the risk of ZIKV infection and disease.

Clinical Features

As in most flavivirus infections, ZIKV infection may present with a broad spectrum of manifestations ranging from asymptomatic, subclinical, mild febrile illness to severe zika disease and death. A significant proportion (up to 80%) of Zika virus infections is estimated to be asymptomatic, which may contribute to the underestimation of the public health impact as well as transmission burden, since they are not usually included in prospective surveillance studies and thus adverse outcomes associated with ZIKV infection may be underestimated.

Clinical manifestations of ZIKV infections usually present as not specific and can be easily misidentified with other arboviral infections prevalent in the area, most commonly dengue, chikungunya, and zika. The most commonly reported symptoms of infection include fever, conjunctivitis, headache, myalgia, pruritus, skin rash, retro-orbital pain, anorexia, vomiting, diarrhea, and abdominal pain. Symptoms last for several days up to a week. Although most ZIKV infections have been associated with a relatively mild illness, the neurovirulence and tissue tropism of ZIKV have been well documented, and several syndromes have



Fig. 6 Newborn exposed to ZIKV during pregnancy. Microcephaly in a newborn exposed to ZIKV during pregnancy with the skull partly collapsed (black arrows). Photograph Courtesy Dr. Drauzio Verella

been described, such as meningoencephalitis, GBS, uveitis, and recently, cerebrovascular events, including stroke and brain bleeding.

One of the major hallmarks of the ZIKV introduction in the Americas was the effects of ZIKV infection during pregnancy, such as microcephaly and other central nervous system (CNS) abnormalities, including brain volume decrease, ventriculomegaly, calcifications, craniofacial disproportion, occipital skin fold, hypertonia, seizures, dysphagia, and arthrogryposis, now described as Congenital Zika Syndrome (CZS) (Fig. 6), intrauterine growth restriction, ocular abnormalities, placental damage, and fetal death. Although several of the CNS implications are often observed in newborns with microcephaly in birth, several newborns of mothers exposed to ZIKV infection during pregnancy and born with normal head size, they may present later in life with several manifestations observed with congenital ZIKV infection, such as postnatal microcephaly, dysphagia, epilepsy, irritability, hypertonia, pyramidal and extrapyramidal signs with dystonic movement, grasp reflex, ataxia, and dyskinesia.

Given the global spread of ZIKV and the concomitant increase of the number of solid organ transplantations in developing countries, a concern among physicians arose how ZIKV infection may present in immunocompromised individuals (including HIV positive individuals). The spectrum of clinical manifestations in such groups and the exact evaluation of infection still remain unknown. Non-specific symptoms have been documented in solid organ transplant recipients exposed to ZIKV infections, including fever, myalgia, adynamia and hemogram abnormalities. Additionally, worsening liver or renal functions are common, requiring hospitalization. Reports of ZIKV infection in HIV patients have been shown mild symptoms, such as pruritic rash, myalgia, non-purulent conjunctivitis, arthralgia, prostration, and headache, followed by a complete recovery. Interestingly, clinical outcomes in ZIKV infected HIV positive individuals during pregnancy, may be similar to non-HIV positive individuals, characterized by placental alterations, such as chorioamnionitis, focal calcifications in the basement membrane of villous endothelium and focal villitis. In patients undergoing cancer treatment, the clinical manifestations of ZIKV infection are similar to those observed in other immunocompromised groups of patients, ranging from non-specific to death.

Pathogenesis

ZIKV-infected mosquitoes transmit the virus into susceptible humans during bloodmeal feeding. Virus is deposited into the skin and directly infects epidermal keratinocytes, dermal fibroblasts, dendritic cells and monocytes. Phosphatidylserine receptors protein families TIM and TAM (including Axl and Tyro3), as well as DC-SIGN, act as putative ZIKV cellular receptors to mediate virion entry. A transient viremia lasting approximately 3–4 days occurs shortly after infection but the virus or viral RNA can be isolated from a variety of bodily fluids, including, semen, urine, saliva and possibly tears, leading to human-to-human transmission in the absence of a mosquito vector. In the developing fetus ZIKV preferentially targets the brain, eyes, and gonadal tissues during all trimesters of pregnancy.

Infection in healthy adults is usually self-limiting but GBS, a rare neurological disease characterized by an acute inflammatory demyelinating polyneuropathy and acute motor axonal neuropathy, have been described. Based on evidence from murine models and human brain organoids, ZIKV shows a preference for infecting the neuronal progenitor cells (NPCs) over mature neurons, leading to apoptosis and potentially explaining how ZIKV causes devastating disease in the developing fetus. Murine studies also show a depletion of proliferating neuronal cells, a destruction of corticospinal pyramidal neurons, and altered neuronal migration and differentiation. In vitro studies further show infection of astrocytes and glial cells.

Other immune-privileged sites are favored by ZIKV including the eyes, testes, and placenta. Although ZIKV can infect the eye at any age, the developing fetal eye is susceptible to damage. Infection of the retinal pigment epithelial and Muller glial cells, as well as the optic nerve can be seen in ZIKV-infected mice. Uveitis and intraocular pressure can also result in vision problems.

Unlike many other arboviruses, ZIKV can cause a long-term infection in testicular tissue. ZIKV-specific RNA has been detected in human semen up to 180 days after infection. Animal models have been critical in understanding the pathogenesis in male gonads. Murine studies show high viral loads in testes, ZIKV-infected spermatozoa and the capacity to infect female mice via intercourse. Sertoli cells, necessary cell type for spermatogenesis, can become persistently infected with ZIKV. Sexual transmission, rather than arthropod transmission, may also contribute to enhanced infection *in utero*. Hofbauer macrophages, trophoblasts, and endothelial cells in the placenta support ZIKV infection and exposure of the fetus to infection. The mechanism through which ZIKV is able to invade immune-privileged sites remains unclear.

Diagnosis

As indicated above (clinical features section) clinical manifestations of acute ZIKV infections are often nonspecific. Therefore, definitive diagnosis relies on molecular or serologic testing. Molecular testing or nucleic acid amplification tests (NAAT) [e.g., reverse transcription polymerase chain reaction (RT-PCR)] requires the detection and amplification of virus-specific nucleic acids and is usually performed on whole blood or serum samples obtained during the first days following the onset of symptoms. Since ZIKV remains in urine for several days past the acute phase of the infection, it is recommended that paired blood and urine samples obtained within 2 weeks after the onset of illness are tested. A positive CRP test will confirm the presence of ZIKV RNA, suggesting an active infection, however, it does not indicate the presence of infectious virus, as the detection of ZIKV RNA in urine is common up to one year post-infection in the absence of an infectious virus.

The diagnosis through various serologic tests [immunoglobulin M or G enzyme-linked immunosorbent assay (IgM ELISA or IgG ELISA), plaque reduction neutralization test (PRNT), hemagglutination inhibition (HI), or complement fixation (CF)] is often complicated by the cross-reactivity observed by previous exposure to other flaviviruses, resulting in false-positive results. The screening test for an acute ZIKV infection for at least up to 4 weeks following the infection, is serum IgM ELISA. An increase in serum ZIKV IgG antibodies in paired serum specimens confirms the diagnosis. Confirmation of inconclusive IgM ELISA or IgG ELISA tests requires testing with the plaque reduction neutralization test (PRNT), a laborious and time-consuming test that is performed in highly specialized reference laboratories. The PRNT test is highly specific for up to 2 flavivirus exposures and becomes non-specific for multiple prior exposures or vaccination (greater than 2) to flaviviruses, which is often the case in hyperendemic areas.

The Centers for Disease Control and Prevention (CDC, USA) established guidelines to recommend a complex algorithm of laboratory and clinical testing during pregnancy. Diagnosis of ZIKV infection in pregnant women and fetus poses several challenges, as ZIKV RNA is transiently present in an infected mother and fetus despite the observation of prolonged viremia during pregnancy. As the laboratory diagnosis may be challenging, other methods such as serial ultrasound monitoring or magnetic resonance imaging (MRI) of the fetus during pregnancy is advised, however their effectiveness may be reduced by limitations to access and the level of the expertise of the radiologist.

Similarly, confirmation of congenital ZIKV infection in newborns poses challenges, as other causes of fetal anomalies include other infectious agents, such as toxoplasmosis, rubella, cytomegalovirus, herpes simplex, and syphilis (TORCH), genetic, toxic and metabolic disorders, which often render laboratory confirmation inconclusive.

Treatment

No specific antiviral therapy is available, however, supportive treatment can reduce morbidity and mortality. Furthermore, several compounds (bortezomib, ivermectin, daptomycin, and mycophenolic acid) have shown the ability to inhibit ZIKV replication *in vitro*, but none of them have yet been evaluated in clinical trials. For ZIKV induced Guillain-Barré syndrome, the therapeutic intervention includes the use of therapeutic plasma exchange or intravenous immune globulin (IVIG) therapy, which are equally effective in GBS. For infants affected with CZS, treatment recommendations are stratified according to the imaging and clinical observations and require care from a multidisciplinary team that monitors complications, developmental delays as well as supportive services to their families.

Prevention and Control

To date the most effective protection for the general population is avoiding mosquito bites, by the use of insect repellents (such as DEET, picaridin) and wearing long sleeved shirts and long pants while spending time outdoors, and taking steps to control mosquitoes both indoors and outdoors (use of screens in doors and windows, air-conditioning use and elimination of breeding sites). The documented prolonged presence of the virus in semen would require abstinence or protected sexual practices after suspected infection as a effective method of reducing the risk of sexual transmission. Current guidelines suggest for a 2 month abstinence or protected sexual intercourse for female partners with suspected infection and for 3 months if the partner with suspected infection is male. Furthermore, prevention of CZS would require avoidance of infection during pregnancy, delaying of pregnancy or avoiding travel to areas of active ZIKV transmission during pregnancy.

Moreover, controlling ZIKV outbreaks is a major challenge and the cornerstone of any successful strategy is based solely on interrupting its transmission cycle, which will require a multi-prong approach including vector control, antiviral therapy and vaccines. While vaccines and therapeutics will likely remain unavailable for a few years, the best prospects for controlling ZIKV will rely on reducing contacts between the vectors and susceptible humans. It is worth noting that recent attempts to control dengue virus, which shares a similar or identical urban transmission cycle, by relying on controlling its arthropod vectors has largely failed. Nonetheless, controlling mosquito populations could possibly be achieved using one or more of the following approaches: (1) protection of open water supply systems in settings where a reliable municipal supply is not available, and elimination of residential breeding sites, including tires and other residential refuse that fill with rainwater and serve as larval habitats. This approach generally relies on community engagement and personal responsibility, sometimes supplemented by penalties for allowing larval development to occur; (2) application of larvicides and adulticide fumigation inside households to eliminate larval and adult mosquitoes, respectively; (3) release of genetically modified male mosquitoes that express a dominant lethal gene, resulting in the death of all offspring from mating with wild females. This approach eliminates any risk for persistence of the transgene in nature and has demonstrated great success on a limited scale; (4) release of *Ae. aegypti* mosquitoes containing the endosymbiont bacterium *Wolbachia*, consequently allowing their spread through natural populations of mosquitoes and suppressing viral transmission by interfering with virus replication in the mosquito. This approach has resulted in significant reductions of disease incidence in all geographic regions applied to date. However, several concerns remain for the long term effectiveness of the latter 2 approaches, including the requirement of widespread release over vast geographic areas in order to overcome the limited flight range of the mosquitoes, the potential for the evolution of resistance mechanisms by either arboviruses and/or mosquitoes, and the enormous logistical, technical and financial challenges of scaling up; and (5) use of lethal traps designed to be inexpensive, relatively maintenance-free and highly effective in reducing *Ae. aegypti* populations. For any of the above methods to have a long-lasting effect will require the sustainable financial support of governments as well as buy-in and enforcement at the community level.

Presently, no clinically approved antiviral drug therapy is available for the treatment of ZIKV infection. However, recent studies have taken advantage of the wealth of knowledge accumulated over the past decade towards dengue drug discovery, to evaluate a number of repurposed approved drugs previously shown to have anti-flaviviral activity (e.g., bortezomib, ivermectin and mycophenolic acid), which also showed activity against ZIKV. Given that the prevention of CZS will be the major goal of anti-ZIKV drug therapy, an ideal drug candidate has to be able to inhibit ZIKV in both fetal brain and systemic sites. The development of a clinically effective antiviral drug may take a long time.

While there are no ZIKV licensed vaccines available, several approaches are currently pursued, including but not limited to live attenuated, inactivated and chimeric virus vaccines, as well as subunit vaccines representing ZIKV proteins, DNA vaccines expressing viral proteins, and other viral vectors expressing viral antigens. It should be noted that vaccine candidates based on DNA vectors, inactivated viruses, vectored viruses and live attenuated viruses have entered clinical trials. However, there are several major barriers moving forward, including: (1) the return of investment, vaccine development costs hundreds of millions of dollars and ZIKV infections have been significantly reduced following its global spread; (2) evaluating vaccine effectiveness in settings of the waning global incidence of ZIKV after the global epidemic, has delayed the implementation of phases 2 and 3 of clinical trials; and (3) fears of immune enhancement leading to more severe disease manifestations, between the interactions with immunity generated by other natural flavivirus infections, as well as vaccines available for dengue, yellow fever, and Japanese encephalitis. Additionally, the well-established association between ZIKV infection and GBS raises the concern that, if this manifestation results from an autoimmune triggering, any ZIKV vaccine could pose a similar risk. These fears will likely persist until the immunological mechanisms between the ZIKV and the host are better understood. Collectively, while several approaches are simultaneously being developed, effective countermeasures (vaccines and antiviral drug therapy) may take years for their final approval.

Further Reading

- Barrows, N.J., Campos, R.K., Powell, S.T., *et al.*, 2016. A screen of FDA-approved drugs for inhibitors of Zika virus infection. *Cell Host & Microbe* 20, 259–270.
- Boldescu, V., Behnam, M.A.M., Vasilakis, N., Klein, C.D., 2017. Broad-spectrum agents for flaviviral infections: Dengue, Zika and beyond. *Nature Reviews Drug Discovery* 16, 565–586.
- Brasil, P., Pereira JR., J.P., Moreira, M.E., *et al.*, 2016. Zika virus infection in pregnant women in Rio de Janeiro. *The New England Journal of Medicine* 375, 2321–2334.
- Caine, E.A., Jagger, B.W., Diamond, M.S., 2018. Animal models of Zika virus infection during pregnancy. *Viruses* 10.
- Gutierrez-Bugallo, G., Piedra, L.A., Rodriguez, M., *et al.*, 2019. Vector-borne transmission and evolution of Zika virus. *Nature Ecology & Evolution* 3, 561–569.
- Mlakar, J., Korva, M., Tul, N., *et al.*, 2016. Zika virus associated with microcephaly. *The New England Journal of Medicine* 374, 951–958.
- Poland, G.A., Ovsyannikova, I.G., Kennedy, R.B., 2019. Zika vaccine development: Current status. *Mayo Clinic Proceedings* 94, 2572–2586.
- Rodriguez-Barraquer, I., Costa, F., Nascimento, E.J.M., *et al.*, 2019. Impact of preexisting dengue immunity on Zika virus emergence in a dengue endemic region. *Science* 363, 607–610.
- Sarno, M., Sacramento, G.A., Khouri, R., *et al.*, 2016. Zika virus infection and stillbirths: A case of hydrops fetalis, hydranencephaly and fetal demise. *PLoS Neglected Tropical Diseases* 10, e0004517.
- Shan, C., Muruato, A.E., Nunes, B.T.D., *et al.*, 2017. A live-attenuated Zika virus vaccine candidate induces sterilizing immunity in mouse models. *Nature Medicine* 23, 763–767.
- Terzian, A.C.B., Zini, N., Sacchetto, L., *et al.*, 2018. Evidence of natural Zika virus infection in neotropical non-human primates in Brazil. *Scientific Reports* 8, 16034.

Relevant Websites

<https://www.cdc.gov/zika/index.html>

Zika Virus.
CDC.

<https://www.who.int/news-room/fact-sheets/detail/zika-virus>

Zika virus.
World Health Organization.

<https://www.niaid.nih.gov/diseases-conditions/zika-virus>

Zika Virus.
NIH: National Institute of Allergy and Infectious Diseases.

https://www.paho.org/hq/index.php?option=com_topics&view=article%20&id=427&Itemid=41484&lang=en

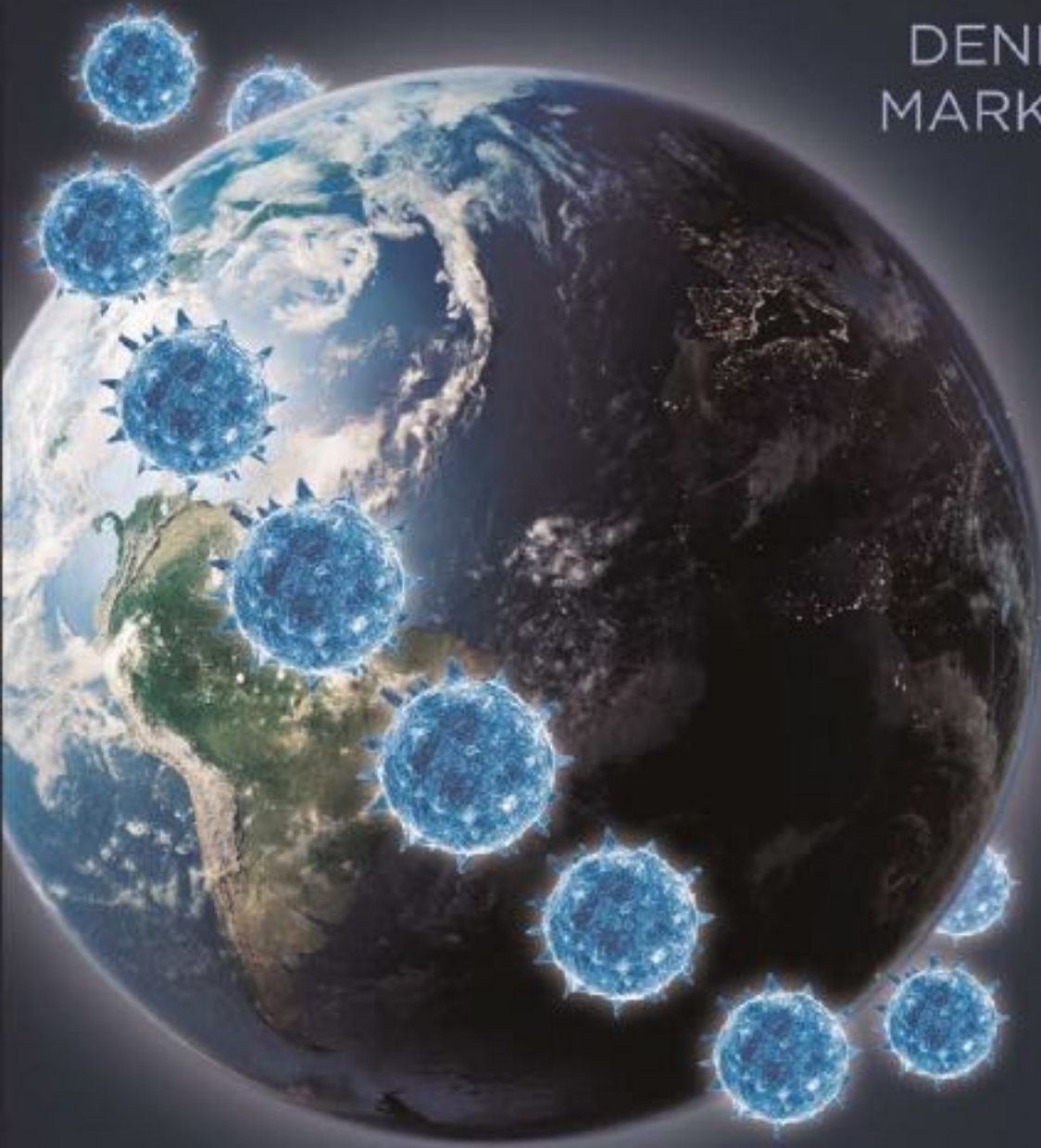
Zika.
PAHO/WHO.
Pan American Health Organization.

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Editors in Chief

DENNIS BAMFORD
MARK ZUCKERMAN



ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Volume 3

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

EDITORS IN CHIEF

Dennis H. Bamford

*Molecular and Integrative Biosciences Research Programme
Faculty of Biological and Environmental Sciences
University of Helsinki, Helsinki, Finland*

Mark Zuckerman

*South London Specialist Virology Centre
King's College Hospital NHS Foundation Trust
London, United Kingdom*

and

*Department of Infectious Diseases
School of Immunology and Microbial Sciences, King's College London Medical School
London, United Kingdom*

Volume 3



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2021 Elsevier Ltd. unless otherwise stated. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-12-814515-9

For information on all publications visit our
website at <http://store.elsevier.com>

		Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org		

Publisher: Oliver Walter

Acquisitions Editor: Priscilla Braglia

Content Project Manager: Katarzyna Miklaszewska

Associate Content Project Manager: Gayathri S.

Designer: Matthew Limbert

EDITORS IN CHIEF



Dennis H. Bamford, PhD, is Professor Emeritus of Virology at the Faculty of Biological and Environmental Sciences, University of Helsinki, Finland. He obtained his PhD in 1980 from the Department of Genetics, University of Helsinki. During 1981–1982 he was an EMBO postdoctoral fellow at the Public Health Research Institute of the City of New York, United States, and during 1983–1992 he worked as a Senior Scientist at the Academy of Finland. In 1993 he was appointed Professor of General Microbiology at the University of Helsinki. He was awarded the esteemed Academy Professorship twice, in 2002–2007 and 2012–2016, and he also served twice as the Director of the Finnish Center of Excellence (in Structural Virology, 2000–2005, and in Virus Research, 2006–2011). Prof. Bamford has had continuous external research funding (e.g., from several European Union, Academy of Finland, TEKES and Jusélius Foundation funds, as well as the Human Frontier Science Program). He is an EMBO member and has held several positions of trust in scientific and administrative organizations.

Prof. Bamford has published approx. 400 articles in international peer-reviewed journals in virology, microbiology, biochemistry, and molecular biology (36 of them in high impact journals). About half of the primary articles have been published with international collaborators showing high international integration. He has also been invited to give 56 keynote and plenary presentations in major international meetings. Prof. Bamford has supervised over 35 Master's and

over 40 PhD theses. Seven of his graduate students or post docs have obtained a professorship and a similar number have a principal investigator status. Prof. Bamford has studied virus evolution from a structure-centered perspective, showing that seemingly unrelated viruses, such as bacteriophage PRD1 and human adenovirus have similar virion architecture. When the corona virion architecture was gradually revealed, it was observed that its structural elements were close to those seen in RNA bacteriophage phi6 so that phi6 has been actively used as surrogate for pathogenic viruses - quite a surprise!



Dr. Mark Zuckerman is Head of Virology, Consultant Medical Virologist, and Honorary Senior Lecturer at South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust and King's College London Medical School, Department of Infectious Diseases, School of Immunology and Microbial Sciences in London, United Kingdom. His interests include the clinical interface between developing molecular diagnostic tests relevant to the local population of patients, respiratory virus infections, herpesvirus infections in immunocompromised patients and blood-borne virus transmission incidents in the healthcare setting. He has chaired the UK Clinical Virology Network, Royal College of Pathologists Virology Specialty Advisory Committee and Virology Examiners Panel and is a member of the Specialty Advisory Committee on Transfusion Transmitted Viruses. He is a co-author on four editions of the "Mims' Medical Microbiology" textbook, has written chapters in a number of other textbooks and has over 100 publications in international peer-reviewed journals and is an associate editor for two journals.

EDITORIAL BOARD

Editors in Chief

Dennis H. Bamford

Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Mark Zuckerman

South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust, London, United Kingdom and Department of Infectious Diseases, School of Immunology and Microbial Sciences, King's College London Medical School, London, United Kingdom

Section Editors

Claude M. Fauquet

St Louis, MO, United States

Michael Feiss

Department of Microbiology and Immunology, Carver College of Medicine, University of Iowa, Iowa City, IA, United States

Elizabeth E. Fry

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Said A. Ghabrial[†]

Department of Plant Pathology, University of Kentucky, Lexington, KY, United States

Eric Hunter

Department of Pathology and Laboratory Medicine, Emory University School of Medicine and Emory Vaccine Center, Emory University, Atlanta, GA, United States

Ilkka Julkunen

Institute of Biomedicine, University of Turku, Turku, Finland

Peter J. Krell

Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Mart Krupovic

Archaeal Virology Unit, Institut Pasteur, Paris, France

Maija Lappalainen

HUS Diagnostic Center, HUSLAB, Clinical Microbiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Hubert G.M. Niesters

Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Massimo Palmarini

MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

David Prangishvili

Institut Pasteur, Paris, France and Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

David I. Stuart

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom and Diamond Light Source, Didcot, United Kingdom

Nobuhiro Suzuki

Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

[†]Deceased.

SECTION EDITORS



Claude Fauquet received his PhD in biochemistry from University Louis Pasteur in Strasbourg, France in 1974. Dr. Fauquet joined the Institut de Recherche pour le Développement (IRD) and worked there as a plant virologist for 28 years, and served in Ivory Coast, West Africa for 14 years. In 1991, he founded the International Laboratory for Tropical Agricultural Biotechnology (ILTAB) at The Scripps Research Institute, CA, United States. ILTAB was then hosted by the Donald Danforth Plant Science Center, St. Louis, MO, from 1999 to 2012. In 2003, he co-founded the Global Cassava Partnership for the 21st Century (GCP21), which he directed until 2019 and which goal is to improve the cassava crop worldwide.

Dr. Fauquet is an international leader in plant virology including taxonomy, epidemiology, molecular virology, and in gene-silencing as an antiviral strategy. He was Secretary of the International Committee on Taxonomy of Viruses (ICTV) for 18 years and the editor of several ICTV Reports including the VIIIth ICTV Report in 2005.

He has published more than 300 research papers in reviewed journals and books. He is a fellow of the American Association for the Advancement of Science, of the American Phytopathological Society and a member of the St. Louis Academy of Sciences. In 2007, Dr. Fauquet was knighted “Chevalier de l’Ordre des Palmes Académiques” by the French Minister of High Education and Research.



Dr. Michael Feiss is Professor Emeritus in the Department of Microbiology and Immunology of the Carver College of Medicine at the University of Iowa, IA, United States. Dr. Feiss received his PhD in Genetics at the University of Washington followed by a postdoctoral traineeship in the laboratory of Dr. Allan Campbell at Stanford. Dr. Feiss is a microbial geneticist who studies virus assembly with an emphasis on how a DNA virus, bacteriophage lambda, packages viral DNA into the empty prohead shell. The lab investigates how sites in the viral DNA orchestrate the initiation and termination of the DNA packaging process. This work includes comprehensive examination of the DNA recognition sites. A related interest is study of terminase, the viral DNA packaging enzyme, including the functional domains for protein–DNA and protein–protein interactions. A second focus has been the roles of the bacterial host’s IHF and DnaJ proteins in the lytic life cycle of the virus. More recent work has involved a genetic dissection of the role of terminase’s ATPase center that powers translocation of viral DNA into the prohead. This interest in the ATP hydrolysis-driven packaging motor involves a multidisciplinary collaboration examining the kinetics of DNA packaging during individual packaging events. Finally, recent studies have also looked at how the packaging process has diverged among several lambda-like phages, including phages 21, N15, and Gifsy-1.



Elizabeth E. Fry is a senior postdoctoral scientist in structural biology at the University of Oxford, Oxford, United Kingdom, where she received her DPhil. for studies relating to the structure determination of Foot-and-Mouth Disease Virus. Specializing in structural virology, Dr. Fry has studied many virus/viral protein structures but her primary focus is on picornavirus structure and function, in particular receptor interactions and virus uncoating. She is particularly interested in rationally designing virus-like-particles as next generation vaccines to reduce the inherent risks in handling live viruses.



Said A. Ghabrial[†] received his BSc in 1959 from Cairo University, Cairo, Egypt, and his PhD from Louisiana State University, Baton Rouge, LA, United States, in 1965. Dr. Ghabrial did postdoctoral research at the University of California, Davis, CA, United States, before returning to Cairo, where he served as a plant virologist in the Ministry of Agriculture. He returned to the United States in 1970 to do postdoctoral research at Purdue University, West Lafayette, IN. In 1972, he joined the Plant Pathology Department at the University of Kentucky, Lexington, KY, United States, where he rose to the rank of professor in 1986 and worked until 2013.

Dr. Ghabrial has served as an associate and senior editor of *Phytopathology*. He served on the editorial boards of the *Encyclopedia of Virology*, 3rd edition and *Encyclopedia of Plant Pathology*, and edited a thematic issue of *Advances in Virus Research* on "Mycoviruses". He was a member of the American Phytopathological Society (APS) and the American Society for Virology (ASV); in July 2002 he was elected as a Fellow of the American Phytopathological Society. He also acted as Chair of the ICTV Subcommittee on Fungal Viruses in 1987–1993 and 2011–2014.

His long professional career allowed him to make many scientific achievements in phytopathology and virology. Among them are molecular dissection of a legume-infecting RNA virus, bean pod mottle virus (BPMV), development of BPMV-based vectors, discovery of a transmissible debilitation disease of the phytopathogenic ascomycete, *Helminthosporium victoriae* (*Cochliobolus victoriae*), establishment of a viral etiology of the *H. victoriae* disease, and advancement of structural biology of diverse fungal viruses.



Eric Hunter, PhD, is Professor of Pathology and Laboratory Medicine at Emory University, Atlanta, GA, United States. He serves as Co-Director of the Emory Center for AIDS Research and is a Georgia Research Alliance Eminent Scholar.

Dr. Hunter's research focus has been the molecular virology and pathogenesis of retroviruses, including human immunodeficiency virus. He has made significant contributions to the understanding of the role of retroviral glycoprotein structural features during viral entry and providing unique insights into the assembly and replication of this virus family. In recent years the emphasis of his research has been on HIV transmission and pathogenesis, defining the extreme genetic bottleneck and selection of viruses with unique traits during HIV heterosexual transmission. He has described the selection of fitter viruses at the target mucosa, a gender difference in the extent of selection bias, and a role for genital inflammation in reducing selection. His research has defined the impact of HIV adaptation to the cellular immune response on immune recognition and control of HIV after

transmission, as well as on virus replicative fitness in vitro and in vivo. Recent work highlights the roles that virus replicative fitness and sex of the host play in defining disease progression in a newly infected individual. His bibliography includes over 300 peer-reviewed articles, reviews, and book chapters. He has also been the recipient of four NIH merit awards for his work on retrovirus and HIV molecular biology.

Dr. Hunter served as the Editor in Chief of the journal *AIDS Research and Human Retroviruses* for 10 years. He was Chair of the AIDS Vaccine Research Subcommittee which is charged with providing advice and consultation on AIDS vaccine research to the National Institute of Allergy and Infectious Diseases and continues to serve on editorial boards for several academic journals and on external advisory committees for several government, academic, and commercial institutions.



Ilkka Julkunen graduated as an MD/PhD in 1984 from the Department of Virology, University of Helsinki, Helsinki, Finland. He worked as a postdoctoral research fellow at Memorial Sloan-Kettering Cancer Center in New York, United States, in 1986–1989, followed by positions as a senior scientist, group leader and research professor at Finnish Institute for Health and Welfare in 1989–2013. In 2013 he became a Professor of Virology at the University of Turku, Turku, Finland. The research interests of Dr. Julkunen have concentrated on innate and adaptive humoral immunity in viral and microbial infections. He has studied intracellular signaling and RIG-I and TLR-mediated activation of interferon system in human macrophages and dendritic cells and stable cell lines in response to human and avian influenza, Sendai, Zika and coronavirus infections. In addition, he has analyzed the downregulation of innate immunity by viral regulatory proteins from influenza, HCV, flavi-, filo- and coronaviruses. He has expertise in vaccinology, biotechnology and development of methods to analyze antiviral immunity, he has also been actively involved in research training and collaborations with biotechnological industry.

[†]Deceased.



Peter Krell started his career in virology early as a summer high school student working for the Canadian Forestry Service studying the resistance of nuclear polyhedrosis viruses (now called baculoviruses) to environmental exposure with Dr. Fred T. Bird at the Insect Pathology Research Institute in Sault Ste. Marie, ON, Canada. He received his BSc and MSc in biology from Carleton University studying the iridovirus *Tipula Iridescent Virus* with Dr Peter Lee, in Ottawa, the Canadian capital. For his PhD he headed east to Dalhousie University in Halifax, Nova Scotia on the Atlantic coast. In addition to enjoying the salt sea air, fresh cod, lobster and mussels, he studied the molecular biology of polydnaviruses under the guidance of Dr Don Stoltz. Heading south to Texas A&M University in College Station, TX, United States, as a Postdoctoral Fellow he worked with Dr. Max Summer, of baculovirus fame, and Dr. Brad Vinson continuing to study polydnaviruses, but also became steeped in the early days of molecular baculovirology. He then accepted a faculty position in the Department of Microbiology and Immunology at the University of Guelph in Guelph, ON, Canada. There he switched to baculovirus research, which was more tractable, due in part to available cell cultures and focused on viral DNA replication and functional genomics,

particularly on chitinase, cathepsin and ME53. In collaboration with Dr. Eva Nagy he studied molecular biology of different animal viruses, notably Fowl Avian adenoviruses and their development as vaccine vectors, but also on the birnavirus infectious pancreatic necrosis virus, the coronavirus porcine endemic diarrhea virus, fowlpox virus and the paramyxovirus Newcastle disease virus. He has been involved extensively with virus taxonomy, being active in the International Committee on Taxonomy of Viruses (ICTV) as member of the *Polydnaviridae* and *Baculoviridae* study groups, national representative of Canada on the ICTV, member of the Executive Committee for the ICTV and Chair of the ICTV Invertebrate Virus Subcommittee. In terms of governance, Peter Krell was President of the Canadian Society of Microbiology, Secretary and later President of the Society for Invertebrate Pathology, as well as being on the Editorial Boards of the *Canadian Journal of Microbiology* and the *ASM Journal of Virology*. While at the University of Guelph, he rose through the ranks to Professor and is currently University Professor Emeritus.



Mart Krupovic is the Head of the Archaeal Virology Unit in the Department of Microbiology at the Institut Pasteur of Paris, France. He received his MSc in Biochemistry in 2005 from the Vilnius University, Vilnius, Lithuania and PhD in 2010 in general microbiology from the University of Helsinki, Helsinki, Finland. His current research focuses on the diversity, origin, and evolution of viruses, as well as molecular mechanisms of virus–host interactions in archaea. He has published over 170 journal articles and serves as an editor or on the editorial boards of *Biology Direct*, *Research in Microbiology*, *Scientific Reports*, *Virology*, and *Virus Evolution*. He is also a member of the Executive Committee of the International Committee on Taxonomy of Viruses (ICTV) and chairs the Archaeal Viruses Subcommittee of the ICTV.



Maija Lappalainen, MD, PhD, Associate Professor of Clinical Microbiology, is the Head of Clinical Microbiology in the HUS Diagnostic Center, HUSLAB, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. In her thesis during the years 1987–1992 she studied the incidence and diagnostics of congenital toxoplasmosis. After PhD, her research interest has been in diagnostic clinical virology, viral hepatitis, respiratory infections, viral infections in the immunocompromised patients and viral infections during pregnancy.



Hubert G.M. Niesters (1958) studied biology and chemistry in Nijmegen, the Netherlands. After obtaining his PhD in Utrecht (Prof. dr. M. Horzinek and Prof. dr. B. van der Zeijst, 1987) on the molecular epidemiology of infectious bronchitis virus, he worked as a post-doctoral fellow with Prof. dr. Jim Strauss at the California Institute of Technology (Pasadena, United States) on the replication of Alphaviruses. He received a Niels Stensen fellowship (The Netherlands) and an E.S. Gosney fellowship (Caltech) during this period.

After returning to the Netherlands (1989), he became a research associate in medical microbiology at the Diagnostic Medical Center (Delft) but moved back to clinical virology as a senior research associate in 1991 at the Erasmus University Medical Center Rotterdam (Head Prof. dr. Ab Osterhaus). From 1993 to 2007, he was responsible for the molecular diagnostics unit. During this period, he was involved in the discovery and characterization of several new viruses and variants. In 2007, he became full professor and director of the Laboratory of Clinical Virology within the Department of Medical Microbiology at the University Medical Center Groningen and University of Groningen.

He has been actively involved in the implementation and development of new technologies like real-time amplification and automation within clinical virology. He has been focusing on molecular diagnostics and its use and the clinical value in a transplant setting, as well as in monitoring treatment of hepatitis viruses. Recently, his interest focuses on rapid regional epidemiology, automation including MiddleWare solutions for molecular diagnostics, as well as the cost-benefit of rapid point-of-impact molecular testing. Special interest is focused on raising awareness for the detection of enteroviruses (enterovirus D68) and its relationship with acute flaccid myelitis (AFM).

Since 2017, he is the Chair of the executive board of QCMD (Quality Control of Molecular Diagnostics, Glasgow). He is an auditor and team leader for the Dutch Council of Accreditation and Co-Editor in Chief of the *Journal of Clinical Virology*. He is an (co)-author of more than 250 peer-reviewed papers, chapters and reviews including emerging viruses, such as enterovirus D68 and hepatitis E virus (H-index 80).

For his entire work, he received in 2016 the “Ed Nowakowski Senior Memorial Clinical Virology Award” from the Pan American Society for Clinical Virology.



Massimo Palmarini is the Director of the MRC-University of Glasgow Centre for Virus Research and Chair of Virology at the University of Glasgow, Glasgow, United Kingdom. A veterinarian by training, his research programs focus on the biology, evolution and pathogenesis of arboviruses and the mechanisms of virus cross-species transmission. His work is funded by the MRC and the Wellcome Trust. Massimo Palmarini has been elected Fellow of the Academy of Medical Sciences, of the Royal Society of Edinburgh and of the Royal Society of Biology and he was a Wolfson-Royal Society Research Merit Awardee. He is a Wellcome Trust Investigator.



David Prangishvili, PhD, Honorary Professor at the Institut Pasteur, Paris, France, and Professor at Tbilisi State University, Tbilisi, Georgia, is one of the pioneers in studies on the biology of Archaea and their viruses. His scientific career spans ex-USSR (Institute of Molecular Biology, Moscow; 1970–1976), Georgia (Georgian National Academy of Sciences, Tbilisi; 1976–1991), Germany (Max-Planck Institute for Biochemistry, Munich; University of Regensburg; 1991–2004) and France (Institut Pasteur, Paris, 2004–2020). In the research groups headed by him, several dozens of new species and eight new families of archaeal viruses have been discovered and characterized, which display remarkable diversity of unique morphotypes and exceptional genome contents. The results of his research contribute to the knowledge on viral diversity on our planet and change the field of prokaryotic virology, leading to the notion that viruses of

hyperthermophilic Archaea form a particular group in the viral world, distinctive from viruses of Bacteria and Eukarya, and to the recognition of the virosphere of Archaea as one of the distinct features of this Domain of Life. David Prangishvili is a member of the Academia Europaea, the European Academy of Microbiology, and the Georgian National Academy of Sciences.



David I. Stuart is MRC Professor of Structural Biology in the Nuffield Department of Medicine, Oxford University, Oxford, United Kingdom, Life Science Director at Diamond Light Source and Director of Instruct-ERIC (pan-European organisation providing shared access to infrastructure and methods for structural biology). He has diverse interests in structural virology from picornaviruses, double-stranded RNA viruses and enveloped RNA viruses. His drive to develop structural techniques led to the determination of the structure of Bluetongue virus (1995) and then the first membrane containing virus, PRD1. More recently, he has been at the fore-front of bringing Cryo-EM technology to bear on virus structure determination and its future role in visualizing virus function in cellulose. In addition to basic science he has a strong commitment to structural vaccinology and the development of antiviral drugs.



Dr. Nobuhiro Suzuki, PhD, received his MSc (1985) in phytopathology and PhD (1989) in virology from Tohoku University in Sendai, Japan. Dr. Suzuki currently serves as a full Professor of the Institute of Plant Stress and Resources, formerly Institute of Plant Sciences and Bioresources at Okayama University and as an Editor of *Virus Research*, *Frontiers in Virology*, *Journal of General Plant Pathology*, *Virology Journal*, and *Biology*. He has also been Guest Editor to *PLoS Pathogens*, *PNAS*, and *mBio*, and an Editorial Board member of *Virology* and *Journal of Virology*.

Suzuki Laboratory focuses on characterization of diverse viruses infecting phytopathogenic fungi and exploration of their interplays. Recent achievements include the discovery of a neo-virus lifestyle exhibited by a (+)ssRNA virus and an unrelated dsRNA virus in a plant pathogenic fungus and of multilayer antiviral defense in fungi involving Dicer. Prior to coming to Kurashiki, Okayama Prefecture, he was a visiting fellow of the Center for Agricultural Biotechnology at the

University of Maryland Biotechnology Institute (UMBI), College Park, MA, United States, for 4 years (1997–2001) to study molecular biology of hypoviruses in the laboratory of Professor Donald L. Nuss. Before visiting UMBI, he served as an assistant professor and a lecturer of the Biotechnology Institute at the Akita Prefectural College of Agriculture, Japan, for 11 years (1988–1998) where he was engaged in a project on molecular characterization of rice dwarf phytoecovirus, a member of the family *Reoviridae*. He received awards from the Japanese Phytopathological Society of Japan and Japanese Society for Virology for his outstanding achievements in plant and fungal virology.

FOREWORD

I am delighted to write the foreword to this wonderful Fourth Edition of the *Encyclopedia of Virology*. The Third Edition was published in 2008, how the world has changed in the intervening years. The release of the updated fourth edition could not be more timely or more prescient. It is superb and a huge tribute to the authors, Elsevier the publisher, and to the brilliant editors, Dennis Bamford and Mark Zuckerman.

SARS-CoV-2 has dominated the world since it emerged in 2019 and affected every continent and every aspect of life. A reminder, if it were needed, of the impact of infectious diseases, the importance of virology and the vulnerability and interconnectivity of our world.

There is no doubt that with rapidly changing ecology, urbanization, climate change, increased travel, and fragile public health systems, epidemics and pandemics will become more frequent, more complex and harder to prevent and contain. Most of these epidemics will be caused by viruses, those we know about and maybe able to predict and some we do not know of that will emerge from animals, plants or the environment. Our changing climate will change the epidemiology of viruses, their vectors and the infections they cause, hence the critical importance of this totally revised Fourth Edition of the *Encyclopedia of Virology* which brings together research and an understanding of viruses in animals, plants, bacteria and fungi, the environment, and among humans. Never has a holistic, one-health understanding been more important.

That starts with an understanding of the fundamentals of virology, a field of science that has been transformed in the years since the Third Edition. An understanding transformed by embracing traditional fields of molecular and structural biology, genomics, and influenced by immunology, genetics, pharmacology and increasingly by epidemiology and mathematics. Events of 2020 and 2021 also show why it is so important to integrate within traditional virology an understanding of the animal and human health and behavior, of climate change and its impact on the ecology of viruses, plant sciences and vectors. And why we must understand the viruses we think we know well, and those viruses less extensively studied. Research is critical to this, research that pushes the boundaries of what we know, has the humility to seek answers to things we do not understand and shares that knowledge with the widest possible community. That research will be most exciting at the interface between disciplines, most impactful when dynamic, open, inclusive, global, and collaborative.

This is what the Fourth Edition of the *Encyclopedia of Virology*, the largest reference source of research in virology sets out to achieve. It is a wonderful contribution to a critical field of knowledge. It contains new chapters, every chapter revised and updated by a dedicated global community who have come together to provide what is a brilliant and inspiring reference. It is an honor to contribute in a very small way to the timely release of the Fourth Edition of the *Encyclopedia of Virology*.

Jeremy Farrar

PREFACE

The fourth edition of the Encyclopedia of Virology is encyclopedic, but we wanted to move away from an alphabetical list, apart from where it was more logical, to a vision that encompassed a different structure. Articles describing novel trends as well as original discoveries in specific subfields of virology have been distributed into a set of five volumes, namely Fundamentals of Virology, Human and Animal Viruses, Plant Viruses, Bacterial, Archaeal, Fungal, Algal and Invertebrate Viruses, and Diagnosis, Treatment and Prevention of Virus Infections.

We had hoped that the new edition would 'go viral' but it was ironic that the time to publication 12 years after the previous edition had been made a bit longer due to a virus infection. The world encountered a devastating global pandemic, COVID-19, caused by a new type of a coronavirus, SARS-CoV-2. Scientists in many disciplines all over the world started immediate efforts to discover solutions as to how to mitigate and stop the spread of the pandemic. Virology moved from being a highly specialized subject to one in which everyone became a virologist, proving just how significant the different aspects of virology are in terms of understanding the nature of viral infection. Since the previous edition, the growth in the field of general virology has been enormous, including huge advances in basic science, identification of novel viruses, diagnostic methods, treatment and prevention. Taking this into account, the introduction of the articles within the Encyclopedia are very timely and crucial for providing a wealth of knowledge of the latest findings in the field of virology to a vast range of people, whether school students, undergraduates, postgraduates, teachers, scientists, researchers, journalists and others interested in infections and the conflict between the host and the pathogen.

Pandemic viruses have become a serious public concern in the changing world. We can ask ourselves whether we have reached the point in which nature can no longer cope with the consequences of increased population density and human activities that are harmful to the environment. Although several pandemics have threatened mankind before, this COVID-19 pandemic has highlighted the massive adverse economic consequences towards the wellbeing of society and the importance of research in virology.

We aimed to produce a Major Reference Work that differs in approach to others and binds all the virology disciplines together. Chapters have been included on origin, evolution and emergence of viruses, environmental virology and ecology, epidemiology, techniques for studying viruses, viral life cycles, structure, entry, genome and replication, assembly and packaging and taxonomy and viral–host interactions. Information has been included on all known species of viruses infecting bacteria, fungi, plants, vertebrates and invertebrates. Additional topics include antiviral classification and examples of their use in management of infection, diagnostic assays and vaccines, as well as the economic importance of viral diseases of crops and their control.

This edition used viral classification according to the 9th Report of the International Committee on Taxonomy of Viruses published in 2012. Updating it to the 10th Report in 2020 was affected by the pandemic and can be found online at <http://ictv.global/report/>.

We wish to acknowledge the hard work, interest, flexibility and patience, during such difficult times both socially and professionally, of everybody involved in the process of writing this edition of the Encyclopedia of Virology, especially Katarzyna Miklaszewska, Priscilla Braglia, Sam Crowe and colleagues at Elsevier. We sincerely thank all the authors and section editors for their excellent contributions to this edition.

Book Cover Image: Viruses are obligate parasites and all cells have their own viruses increasing the total number of viruses to the estimated astronomical number of 10^{31} that extends the number of stars in the universe. The viral string illustrates how pandemic viruses surround the globe. The original picture was created by Dr. Nina Atanasova (Finnish Meteorological Institute and University of Helsinki) and amended by Matthew Limbert at Elsevier.

*Dennis H. Bamford
Mark Zuckerman*

HOW TO USE THE ENCYCLOPEDIA

Structure of the Encyclopedia

All articles in the encyclopedia are arranged thematically as a series of entries within subjects/sections, apart from volume 2 where there it was more logical to have articles arranged alphabetically.

There are three features to help you easily find the topic you are interested in: a thematic contents list, a full subject index, and contributors.

1. Thematic contents list: The alphabetical contents list, which appears at the front of each volume, lists the entries in the order that they appear in the encyclopedia.
2. Index: The index appears at the end of volume 5 and includes page numbers for quick reference to the information you are looking for. The index entries differentiate between references to a whole entry, a part of an entry, and a table or figure.
3. Contributors: At the start of each volume there is a list of the authors who contributed to all volumes.

LIST OF CONTRIBUTORS

Stephen T. Abedon

*The Ohio State University, Mansfield, OH,
United States*

Peter Abrahamian

*Agricultural Research Service, US Department of
Agriculture, Beltsville, MD, United States*

Jônatas S. Abrahão

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Florence Abravanel

*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Nicola G.A. Abrescia

*Center for Cooperative Research in Biosciences, Basque
Research and Technology Alliance, Derio, Spain;
Ikerbasque, Basque Foundation for Science, Bilbao,
Spain; and Center for Biomedical Research in the Liver
and Digestive Diseases Network, Carlos III Health
Institute, Madrid, Spain*

Gian Paolo Accotto

*Institute for Sustainable Plant Protection, National
Research Council of Italy, Torino, Italy*

Elisabeth Adderson

*St. Jude Children's Research Hospital, Memphis, TN,
United States and University of Tennessee Health
Sciences Center, Memphis, TN, United States*

Mustafa Adhab

University of Baghdad, Baghdad, Iraq

Alexey A. Agranovsky

Lomonosov Moscow State University, Moscow, Russia

Nasim Ahmed

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Maher Al Rwahnih

University of California, Davis, CA, United States

Olufemi J. Alabi

*Texas A&M AgriLife Research and Extension Center,
Weslaco, TX, United States*

Aurélien A. Albertini

*Institute for Integrative Biology of the Cell (I2BC),
French Alternative Energies and Atomic Energy
Commission, French National Center for Scientific
Research, Paris-Sud University, University of
Paris-Saclay, Gif-sur-Yvette, France*

Aleksandra Alimova

*The City University of New York (CUNY), School of
Medicine, The City College of New York, New York,
NY, United States*

Juan C. Alonso

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Imran Amin

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Stephanie E. Ander

*University of Colorado School of Medicine, Aurora, CO,
United States*

Danielle E. Anderson

Duke-NUS Medical School, Singapore, Singapore

Ida Bagus Andika

Qingdao Agricultural University, Qingdao, China

Ana C.d.S.P. Andrade

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Juana Angel

Pontifical Javeriana University, Bogota, Colombia

Vanesa Anton-Vazquez

King's College Hospital, London, United Kingdom

Guido Antonelli

Sapienza University of Rome, Rome, Italy

Josefa Antón

University of Alicante, Alicante, Spain

Nanako Aoki

*Tokyo University of Agriculture and Technology, Fuchu,
Japan*

Timothy D. Appleby

King's College Hospital, London, United Kingdom

Miguel Arenas

*Department of Biochemistry, Genetics and Immunology,
University of Vigo, Vigo, Spain and CINBIO
(Biomedical Research Center), University of Vigo, Vigo,
Spain*

Basil Arif

*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Vicente Arnau

*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain*

Gaurav Arya

Duke University, Durham, NC, United States

Leyla Asadi

University of Alberta, Edmonton, AB, Canada

Sassan Asgari

The University of Queensland, Brisbane, QLD, Australia

Nina S. Atanasova

*Finnish Meteorological Institute, Helsinki, Finland and
University of Helsinki, Helsinki, Finland*

Houssam Attoui

*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Silvia Ayora

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Walid Azab

Free University of Berlin, Berlin, Germany

Sasha R. Azar

*The University of Texas Medical Branch, Galveston, TX,
United States*

Fengwei Bai

*The University of Southern Mississippi, Hattiesburg, MS,
United States*

Dalan Bailey

The Pirbright Institute, Pirbright, United Kingdom

S.C. Baker

*Loyola University of Chicago, Maywood, IL,
United States*

Fausto Baldanti

*University of Pavia, Pavia, Italy and Scientific Institute
for Research, Hospitalization and Healthcare, San
Matteo Polyclinic Foundation, Pavia, Italy*

Logan Banadyga

*Public Health Agency of Canada, Winnipeg, MB,
Canada*

Ashley C. Banyard

*Animal and Plant Health Agency, Addlestone,
United Kingdom; University of West Sussex, Falmer,
United Kingdom; and St. George's Medical School,
University of London, London, United Kingdom*

Xiaoyong Bao

*The University of Texas Medical Branch, Galveston, TX,
United States*

Yiming Bao

*Beijing Institute of Genomics, Chinese Academy of
Sciences, University of Chinese Academy of Sciences,
Beijing, China*

Alan D.T. Barrett

*The University of Texas Medical Branch, Galveston, TX,
United States*

Diana P. Baquero

*Archaeal Virology Unit, Institut Pasteur, Paris, France
and Sorbonne University, Paris, France*

Moshe Bar-Joseph

*Agricultural Research Organization, Volcani Center, Bet
Dagan, Israel*

Rachael S. Barr

*Bristol Royal Hospital for Children, Bristol,
United Kingdom*

Ralf Bartenschlager

Heidelberg University, Heidelberg, Germany

David L.V. Bauer

Francis Crick Institute, London, United Kingdom

Oliver W. Bayfield

University of York, York, United Kingdom

Sally A. Baylis

Paul-Ehrlich-Institute, Langen, Germany

Philippa M. Beard

*The Pirbright Institute, Pirbright, United Kingdom and The
Roslin Institute, University of Edinburgh, United Kingdom*

Paul Becher

University of Veterinary Medicine, Hannover, Germany

Björn Becker

Saarland University, Saarbrücken, Germany

Karen L. Beemon

Johns Hopkins University, Baltimore, MD, United States

Martin Beer

Friedrich-Loeffler-Institute, Insel Riems, Germany

Jose Miguel Benito

*Health Research Institute of the Jiménez Díaz
Foundation, Autonomous University of Madrid and Rey
Juan Carlos University Hospital, Móstoles, Spain*

Mária Benkő

*Institute for Veterinary Medical Research, Center for
Agricultural Research, Budapest, Hungary*

Max Bergoin
National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada

Sabrina Bertin
Council for Agricultural Research and Economics,
Research Center for Plant Protection and Certification,
Rome, Italy

Shweta Bhatt
University of Copenhagen, Copenhagen, Denmark

Dennis K. Bideshi
California Baptist University, Riverside, CA,
United States and University of California, Riverside,
CA, United States

Yves Bigot
INRAE – French National Research Institute for
Agriculture, Food and Environment, Nouzilly, France

Richard J. Bingham
University of York, York, United Kingdom

Vera Bischoff
Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany

Kate N. Bishop
Francis Crick Institute, London, United Kingdom

Lindsay W. Black
The University of Maryland School of Medicine,
Baltimore, MD, United States

Romain Blanc-Mathieu
Institute for Chemical Research, Kyoto University, Kyoto,
Japan

Soile Blomqvist
National Institute for Health and Welfare, Helsinki,
Finland

Bryony C. Bonning
University of Florida, Gainesville, FL, United States

Lisa M. Bono
Rutgers, The State University of New Jersey,
New Brunswick, NJ, United States

Alexia Bordigoni
Aix-Marseille University, CNRS, IRD, Mediterranean
Institute of Oceanography, Marseille, France and
Aix-Marseille University, IRD257, Assistance-Publique
des Hôpitaux de Marseille, UMR Microbes, Evolution,
Phylogeny and Infections (MEPHI), IHU Méditerranée
Infection, Marseille, France

Mihnea Bostina
University of Otago, Dunedin, New Zealand

Maxime Boutier
University of Liège, Liège, Belgium

P.R. Bowser
Cornell University, Ithaca, NY, United States

Daniel Bradshaw
Public Health England, London, United Kingdom

Claude Bragard
University of Louvain, Louvain-la-Neuve, Belgium

Aaron C. Brault
Centers for Disease Control and Prevention, Fort Collins,
CO, United States

Nicolas Bravo-Vasquez
St. Jude Children's Research Hospital, Memphis, TN,
United States

Rob W. Briddon
University of Agriculture, Faisalabad, Pakistan

Thomas Briebe
Columbia University, New York, NY, United States

Paul Britton
The Pirbright Institute, Pirbright, United Kingdom

Thomas J. Brouwers
Athena Institute, VU Amsterdam, Amsterdam, The
Netherlands

Kevin E. Brown
Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom

Corina P.D. Brussaard
NIOZ Royal Netherlands Institute for Sea Research,
Den Burg, Texel, The Netherlands and Utrecht
University, Utrecht, The Netherlands

Harald Brüssow
Laboratory of Gene Technology, Department of
Biosystems, KU Leuven, Leuven, Belgium

Joachim J. Bugert
Bundeswehr Institute of Microbiology, Munich, Germany

Jozef J. Bujarski
Northern Illinois University, DeKalb, IL, United States
and Polish Academy of Sciences, Poznan, Poland

Laura Burga
University of Otago, Dunedin, New Zealand

Sara H. Burkhard
University Hospital of Zurich, Zurich, Switzerland

Cara C. Burns
Centers for Disease Control and Prevention, Atlanta,
GA, United States

Felicity Burt
University of the Free State, Bloemfontein, South Africa

Kerry S. Burton
Leamington Spa, United Kingdom

Sarah J. Butcher
University of Helsinki, Helsinki, Finland

Mathias Büttner
Leipzig University, Leipzig, Germany

Jesse Cahill
Sandia National Labs, Albuquerque, NM, United States

Marianna Calabretto
Sapienza University of Rome, Rome, Italy

Thierry Candresse
The National Research Institute for Agriculture, Food and the Environment, University of Bordeaux, Villenave d'Ornon, France

Alan J. Cann
University of Leicester, Leicester, United Kingdom

Lorenzo Capucci
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Irene Carlon-Andres
University of Oxford, Oxford, United Kingdom

José M. Casasnovas
National Center for Biotechnology, Spanish National Research Council (CSIC), Madrid, Spain

J.W. Casey
Cornell University, Ithaca, NY, United States

R.N. Casey
Cornell University, Ithaca, NY, United States

Sherwood R. Casjens
University of Utah, Salt Lake City, UT, United States

Antonella Casola
The University of Texas Medical Branch, Galveston, TX, United States

José R. Castón
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Carlos E. Catalano
University of Colorado Anschutz Medical Campus, Skaggs School of Pharmacy and Pharmaceutical Sciences, Aurora, CO, United States

Roberto Cattaneo
Mayo Clinic, Rochester, MN, United States

Patrizia Cavadini
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Supranee Chaiwatpongsakorn
Nationwide Children's Hospital, Columbus, OH, United States

Supriya Chakraborty
Jawaharlal Nehru University, New Delhi, India

Yu-Chan Chao
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Tyler P. Chavers
Centers for Disease Control and Prevention, Atlanta, GA, United States

Keping Chen
Jiangsu University, Zhenjiang, China

Xiaorui Chen
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Yanping Chen
Bee Research Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

Dayna Cheng
National Cheng Kung University, Tainan, Taiwan

Quentin Chesnais
University of Strasbourg, Colmar, France

Sotaro Chiba
Nagoya University, Nagoya, Japan

Wah Chiu
Stanford University, Stanford, CA, United States

David Chmielewski
Stanford University, Stanford, CA, United States

Irma E. Cisneros
The University of Texas Medical Branch, Galveston, TX, United States

Lark L. Coffey
University of California, Davis, CA, United States

Alanna B. Cohen
Rutgers University, New Brunswick, NJ, United States

Jeffrey I. Cohen
Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Seth Coleman
Rice University, Houston, TX, United States

Miquel Coll
*Institute for Research in Biomedicine, Barcelona, Spain
 and Institute for Molecular Biology of Barcelona,
 Barcelona, Spain*

John Collinge
*UCL Institute of Prion Diseases, London,
 United Kingdom*

Carina Conceicao
The Pirbright Institute, Pirbright, United Kingdom

Gabriela N. Condezo
*National Center for Biotechnology, Spanish National
 Research Council, Madrid, Spain*

Michaela J. Conley
*MRC-University of Glasgow Centre for Virus Research,
 Glasgow, United Kingdom*

Charles A Coomer
University of Oxford, Oxford, United Kingdom

Anne K. Cordes
*Hannover Medical School, Institute of Virology, Hannover,
 Germany*

Mauricio Cortes Jr.
*Department of Chemistry, College of Arts and Sciences,
 Fort Wayne, IN, United States*

Robert H.A. Coutts
University of Hertfordshire, Hatfield, United Kingdom

Jeff A. Cowley
CSIRO Livestock Industries, Brisbane, QLD, Australia

Robert W. Cross
*The University of Texas Medical Branch, Galveston, TX,
 United States*

Henryk Czosnek
The Hebrew University of Jerusalem, Rehovot, Israel

Håkon Dahle
*Department of Biological Sciences, University of Bergen,
 Bergen, Norway*

Janet M. Daly
*University of Nottingham, Sutton Bonington,
 United Kingdom*

Subha Das
Okayama University, Kurashiki, Japan

Indranil Dasgupta
University of Delhi, New Delhi, India

Sibnarayan Datta
*Defence Research Laboratory, Defence Research and
 Development Organisation (DRDO), Tezpur, Assam,
 India*

Amy Davis
*St Jude Children's Research Hospital, Memphis, TN,
 United States*

William O. Dawson
*Citrus Research and Education Center, Lake Alfred, FL,
 United States and University of Florida, Lake Alfred, FL,
 United States*

Erik De Clercq
*Rega Institute for Medical Research, KU Leuven,
 Leuven, Belgium*

Raoul J. de Groot
Utrecht University, Utrecht, The Netherlands

Juan C. de la Torre
*The Scripps Research Institute, La Jolla, CA,
 United States*

Marcelo De las Heras
University of Zaragoza, Zaragoza, Spain

Juliana Gabriela Silva de Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Athos S. de Oliveira
University of Brasília, Brasília, Brazil

Nicole T. de Stefano
University of Florida, Gainesville, FL, United States

Greg Deakin
NIAB-EMR, East Malling, United Kingdom

Philippe Delfosse
University of Luxembourg, Esch-sur-Alzette, Luxembourg

Natacha Delrez
University of Liège, Liège, Belgium

Tatiana A. Demina
*Molecular and Integrative Biosciences Research Program,
 Faculty of Biological and Environmental Sciences,
 University of Helsinki, Helsinki, Finland*

Ismail Demir
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

Zihni Demirbağ
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

X. Deng
*Loyola University of Chicago, Maywood, IL,
 United States*

Cécile Desbiez
*Plant Pathology Unit, INRAE – French National
 Research Institute for Agriculture, Food and
 Environment, Montfavet, France*

Christelle Desnues

Aix-Marseille University, CNRS, IRD, Mediterranean Institute of Oceanography, Marseille, France and Aix-Marseille University, IRD 257, Assistance-Publique des Hôpitaux de Marseille, UMR Microbes, Evolution, Phylogeny and Infections (MEPHI), IHU Méditerranée Infection, Marseille, France

Samantha J. DeWerff

University of Illinois at Urbana-Champaign, Urbana, IL, United States

Daniele Di Carlo

Sapienza University of Rome, Rome, Italy

Arturo Diaz

La Sierra University, Riverside, CA, United States

Alfredo Diaz-Lara

University of California, Davis, CA, United States

Ralf G. Dietzgen

The University of Queensland, St. Lucia, QLD, Australia

Michele Digiario

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Michael Dills

Montana State University, Bozeman, MT, United States

Wayne Dimech

National Serology Reference Laboratory, Fitzroy, VIC, Australia

Savithramma P. Dinesh-Kumar

University of California, Davis, CA, United States

Linda K. Dixon

The Pirbright Institute, Pirbright, United Kingdom

Valerian V. Dolja

Oregon State University, Corvallis, OR, United States

Aušra Domanska

University of Helsinki, Helsinki, Finland

Leslie L. Domier

Agricultural Research Service, US Department of Agriculture, Urbana, IL, United States

Pilar Domingo-Calap

Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, Valencia, Spain

Tatiana Domitrovic

Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Sarah M. Doore

Michigan State University, East Lansing, MI, United States

Lucy Dorrell

University of Oxford, Oxford, United Kingdom

Rosemary A. Dorrington

Rhodes University, Grahamstown, South Africa

Andor Doszpoly

Hungarian Academy of Sciences, Budapest, Hungary

Andreas Dotzauer

University of Bremen, Bremen, Germany

Simon B. Drysdale

St George's University Hospitals NHS Foundation Trust, London, United Kingdom and St George's, University of London, London, United Kingdom

Robert L. Duda

University of Pittsburgh, Pittsburgh, PA, United States

Carol Duffy

University of Alabama, Tuscaloosa, AL, United States

Siobain Duffy

Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

David D. Dunigan

University of Nebraska–Lincoln, Lincoln, NE, United States

Stéphane Duquerroy

University of Paris-Saclay, Orsay, France and Institut Pasteur, Paris, France

Bas E. Dutilh

Utrecht University, Utrecht, The Netherlands and Radboud University Medical Center, Nijmegen, The Netherlands

Michael Edelstein

Faculty of Medicine, Bar Ilan University, Ramat Gan, Israel

Herman K. Edskes

National Institutes of Health, Bethesda, MD, United States

Rosina Ehmann

Bundeswehr Institute of Microbiology, Munich, Germany

Toufic Elbeaino

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Joanne B. Emerson

University of California, Davis, CA, United States

Ann Emery

University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Christine E. Engeland
University Hospital Heidelberg and German Cancer Research Center, Heidelberg, Germany and Witten/Herdecke University, Witten, Germany

Luis Enjuanes
National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Katri Eskelin
University of Helsinki, Helsinki, Finland

Rosa Esteban
Institute of Biology and Functional Genomics, CSIC/University of Salamanca, Salamanca, Spain

Mary K. Estes
Baylor College of Medicine, Houston, TX, United States

Cassia F. Estofolete
São José do Rio Preto School of Medicine, São José do Rio Preto, Brazil

Alyssa B. Evans
National Institutes of Health, Hamilton, MT, United States

Øystein Evensen
Norwegian University of Life Sciences, Oslo, Norway

Alex Evilevitch
Department of Experimental Medical Science, Lund University, Lund, Sweden

Montserrat Fàbrega-Ferrer
Institute for Research in Biomedicine, Barcelona, Spain and Institute for Molecular Biology of Barcelona, Barcelona, Spain

Francesco Faggioli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Bentley A. Fane
University of Arizona, Tucson, AZ, United States

Brian A. Federici
University of California, Riverside, CA, United States

F. Fenner
Australian National University, Canberra, ACT, Australia

Isabel Fernández de Castro
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Giovanni Ferrara
University of Alberta, Edmonton, AB, Canada

Elvira Fiallo-Olivé
Institute for Mediterranean and Subtropical Horticulture “La Mayora”–Spanish National Research Council–University of Malaga, Algarrobo-Costa, Málaga, Spain

Andrew E. Firth
University of Cambridge, Cambridge, United Kingdom

Roland A. Fleck
King’s College London, London, United Kingdom

Ricardo Flores
Polytechnic University of Valencia, Higher Council of Scientific Research, Valencia, Spain

Ervin Fodor
University of Oxford, Oxford, United Kingdom

Anthony R. Fooks
Animal and Plant Health Agency, Addlestone, United Kingdom; University of Liverpool, Liverpool, United Kingdom; and St. George’s Medical School, University of London, London, United Kingdom

Patrick Forterre
Archeal Virology Unit, Institut Pasteur, Paris, France and French National Center for Scientific Research, Institute of Integrative Biology of the Cell, University of Paris-Saclay, Gif sur Yvette, France

Rennos Fragkoudis
University of Nottingham, Sutton Bonington, United Kingdom and University of Edinburgh, Edinburgh, United Kingdom

Manuel A. Franco
Pontifical Javeriana University, Bogota, Colombia

Giovanni Franzo
Department of Animal Medicine, Production and Health (MAPS), Padua University, Padua, Italy

Graham L. Freimanis
The Pirbright Institute, Pirbright, United Kingdom

Juliana Freitas-Astúa
Brazilian Agricultural Research Corporation (Embrapa) Cassava and Fruits, Cruz das Almas, Brazil

Elizabeth E. Fry
Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Marc Fuchs
Cornell University, Geneva, NY, United States

Tsutomu Fujimura
Institute of Biology and Functional Genomics, CSIC/University of Salamanca, Salamanca, Spain

- Kuko Fuke
Tokyo University of Agriculture and Technology, Fuchu, Japan
- Toshiyuki Fukuhara
Tokyo University of Agriculture and Technology, Fuchu, Japan
- To S. Fung
South China Agricultural University, Guangzhou, China
- Yahya Z.A. Gaafar
Julius Kuehn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany
- Toni Gabaldon
Barcelona Supercomputing Center-National Center for Supercomputing, Institute of Research in Biomedicine, and Catalan Institution for Research and Advanced Studies, Barcelona, Spain
- Morgan Gaïa
University of Paris-Saclay, Evry, France
- José Gallardo
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain
- Hernan Garcia-Ruiz
University of Nebraska–Lincoln, Lincoln, NE, United States
- Juan A. García
National Center for Biotechnology-Spanish National Research Council, Madrid, Spain
- Matteo P. Garofalo
The University of Texas Medical Branch, Galveston, TX, United States
- Yves Gaudin
Institute for Integrative Biology of the Cell (I2BC), French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, Paris-Sud University, University of Paris-Saclay, Gif-sur-Yvette, France
- Andrew D.W. Geering
The University of Queensland, St. Lucia, QLD, Australia
- Thomas W. Geisbert
The University of Texas Medical Branch, Galveston, TX, United States
- Andrea Gentili
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy
- Volker Gerdts
University of Saskatchewan, Saskatoon, SK, Canada
- Said A. Ghabrial[†]
Department of Plant Pathology, University of Kentucky, Lexington, KY, United States
- Clément Gilbert
Evolution, Genomes, Behavior and Ecology Laboratory, CNRS University of Paris-Sud UMR 9191, IRD UMR 247, Gif-sur-Yvette, France
- Robert L. Gilbertson
University of California, Davis, CA, United States
- Efstathios S. Giotis
Imperial College London, London, United Kingdom and University of Essex, Colchester, United Kingdom
- Laurent Glais
French Federation of Seed Potato Growers/Research, Development, Promotion of Seed Potato, Paris, France and Institute for Genetics, Environment and Plant Protection, Agrocampus West, French National Institute for Agriculture, Food and Environment, University of Rennes 1, Le Rheu, France
- Miroslav Glasa
Biomedical Research Center, Slovak Academy of Sciences, Bratislava, Slovakia
- Ido Golding
University of Illinois at Urbana-Champaign, Urbana, IL, United States
- Esperanza Gomez-Lucia
Complutense University of Madrid, Madrid, Spain
- Zheng Gong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
- Andrea González-González
University of Florida, Gainesville, FL, United States
- Michael M. Goodin
University of Kentucky, Lexington, KY, United States
- Alexander E. Gorbalenya
Leiden University Medical Center, Leiden, The Netherlands
- Paul Gottlieb
The City University of New York (CUNY), School of Medicine, The City College of New York, New York, NY, United States
- M.-A. Grandbastien
INRAE – French National Research Institute for Agriculture, Food and Environment, Versailles, France

[†]Deceased.

Meritxell Granell
*National Center for Biotechnology, Madrid, Spain and
 Institute of Chemical Research of Catalonia (ICIQ),
 Tarragona, Spain*

Patrick L. Green
*The Ohio State University, Columbus, OH,
 United States*

Sandra J. Greive
University of York, York, United Kingdom

Diane E. Griffin
*Johns Hopkins Bloomberg School of Public Health,
 Baltimore, MD, United States*

Jonathan M. Grimes
University of Oxford, Oxford, United Kingdom

Nigel Grimsley
*Integrative Biology of Marine Organisms Laboratory,
 Banuyls-sur-Mer, France and Sorbonne University,
 Banuyls-sur-Mer, France*

Bruno Gronenborn
*Institute for Integrative Biology of the Cell, CNRS,
 University of Paris-Sud, CEA, Gif sur Yvette, France*

Julianne H. Grose
Brigham Young University, Provo, UT, United States

Scott Grytdal
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

Duane J. Gubler
Duke-NUS Medical School, Singapore, Singapore

Peixuan Guo
*College of Pharmacy, The Ohio State University,
 Columbus, OH, United States*

Tongkun Guo
*Beijing Institute of Genomics, Chinese Academy of
 Sciences, Beijing, China*

Anne-Lise Haenni
*Institut Jacques Monod, French National Center for
 Scientific Research, Paris Diderot University, Paris,
 France*

Susan L. Hafenstein
*Pennsylvania State University, Hershey, PA,
 United States*

Ahmed Hafez
*Biotechvana, Valencia, Spain; Pompeu Fabra University,
 Barcelona, Spain; and Minia University, Minya, Egypt*

Marie Hagbom
Linköping University, Linköping, Sweden

Sébastien Halary
*National Museum of Natural History, UMR 7245
 CNRS/MNHN Molécule de Communication et
 Adaptation des Micro-organismes, Paris, France*

Aron J. Hall
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

John Hammond
*Floral and Nursery Plants Research, Agricultural
 Research Service, US Department of Agriculture,
 Beltsville, MD, United States*

Rosemarie W. Hammond
*Agricultural Research Service, US Department of
 Agriculture, Beltsville, MD, United States*

Virginia Hargest
*St Jude Children's Research Hospital, Memphis, TN,
 United States and University of Tennessee Health
 Science Center, Memphis, TN, United States*

Scott J. Harper
Washington State University, Prosser, WA, United States

Balázs Harrach
*Institute for Veterinary Medical Research, Center for
 Agricultural Research, Budapest, Hungary*

Masayoshi Hashimoto
The University of Tokyo, Tokyo, Japan

Muhammad Hassan
University of Agriculture, Faisalabad, Pakistan

Asma Hatoum-Aslan
University of Alabama, Tuscaloosa, AL, United States

Philippa C. Hawes
The Pirbright Institute, Pirbright, United Kingdom

Janelle A. Hayes
*University of Massachusetts Medical School, Worcester,
 MA, United States*

Guijuan He
Virginia Tech, Blacksburg, VA, United States

Klaus Hedman
*University of Helsinki, Helsinki, Finland and Helsinki
 University Hospital, Helsinki, Finland*

Albert Heim
Hannover Medical School, Hanover, Germany

Gary L. Hein
*University of Nebraska–Lincoln, Lincoln, NE,
 United States*

Manfred Heinlein
*IBMP-CNRS, University of Strasbourg, Strasbourg,
 France*

Mercedes Hernando-Pérez
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Carmen Hernández
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Etienne Herrbach
University of Strasbourg, Colmar, France

Stephen Higgs
*Biosecurity Research Institute, Kansas State University,
Manhattan, KS, United States*

Bradley I. Hillman
Rutgers University, New Brunswick, NJ, United States

Deborah M. Hinton
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Judith Hirsch
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Jody Hobson-Peters
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Natalie M. Holste
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Jin S. Hong
Kangwon National University, Chunchon, South Korea

Margaret J. Hosie
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Olivia G. Howell
University of Alabama, Tuscaloosa, AL, United States

Liya Hu
Baylor College of Medicine, Houston, TX, United States

Zhaoyang Hu
Jiangsu University, Zhenjiang, China

Kuan-Ying A. Huang
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Yu Huang
Peking University, Beijing, China

Natalia B. Hubbs
Hanover College, Hanover, IN, United States

Elisabeth Huguet
*Research Institute on Insect Biology, French National
Center for Scientific Research, University of Tours,
Tours, France*

Roger Hull
John Innes Centre, Norwich, United Kingdom

Kiwamu Hyodo
Okayama University, Kurashiki, Japan

Eugénie Hébrard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

Martin Hölzer
University of Jena, Jena, Germany

Tetsuro Ikegami
*The University of Texas Medical Branch at Galveston,
Galveston, TX, United States*

Niina Ikonen
*Finnish Institute for Health and Welfare, Helsinki,
Finland*

Cihan İnan
*Department of Molecular Biology and Genetics,
Karadeniz Technical University, Trabzon, Turkey*

İkbal Agah İnce
*Department of Medical Microbiology, Acibadem
University School of Medicine, Istanbul, Turkey*

Katsuaki Inoue
Diamond Light Source, Didcot, United Kingdom

Toru Iwanami
Tokyo University of Agriculture, Tokyo, Japan

Jacques Izopet
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Fauziah Mohd Jaafar
*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Andrew O. Jackson
China Agricultural University, Beijing, China

Daral J. Jackwood
*The Ohio State University/OARDC, Wooster, OH,
United States*

Jean-Rock Jacques
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Tiffany Jenkins
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University, Columbus,
OH, United States*

Jeffrey D. Jensen
Arizona State University, Tempe, AZ, United States

Daohong Jiang
Huazhong Agricultural University, Wuhan, China

Zhihao Jiang
China Agricultural University, Beijing, China

Allison R. Jilbert
The University of Adelaide, Adelaide, SA, Australia

Peng Jing
*Department of Chemistry, College of Arts and Sciences,
Fort Wayne, IN, United States*

Xixi Jing
Central South University, Changsha, China

Meesbah Jiwaji
Rhodes University, Grahamstown, South Africa

Kyle L. Johnson
*The University of Texas at El Paso, El Paso, TX,
United States and CQuentia, Memphis, TN, United
States*

Welkin E. Johnson
Boston College, Chestnut Hill, MA, United States

Ian M. Jones
*University of Reading, Reading, United Kingdom and
London School of Hygiene and Tropical Medicine,
London, United Kingdom*

Ramon Jordan
*Agricultural Research Service, US Department of
Agriculture, Beltsville, MD, United States*

Thomas Joris
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Ilkka Julkunen
*Institute of Biomedicine, University of Turku, Turku,
Finland*

Sandra Junglen
Charité - University Medicine Berlin, Berlin, Germany

Masanori Kaido
Kyoto University, Kyoto, Japan

Laura Kakkola
University of Turku, Turku, Finland

Hannimari Kallio-Kokko
*University of Helsinki and Helsinki University Hospital,
Helsinki, Finland*

Nassim Kamar
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Phyllis J. Kanki
*Harvard T.H. Chan School of Public Health, Boston,
MA, United States*

Peter Karayiannis
University of Nicosia, Nicosia, Cyprus

Henry M. Kariithi
*Kenya Agricultural and Livestock Research Organization,
Nairobi, Kenya*

Brian A. Kelch
*University of Massachusetts Medical School, Worcester,
MA, United States*

Karen E. Keller
*Horticultural Crops Research Unit, Agricultural Research
Service, US Department of Agriculture, Corvallis, OR,
United States*

Japhette E. Kembou-Ringert
University of Tel Aviv, Tel Aviv, Israel

Peter J. Kerr
*University of Sydney, Sydney, NSW, Australia and
CSIRO Health and Biosecurity, Black Mountain
Laboratories, Canberra, ACT, Australia*

Tiffany King
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University College
of Medicine, Columbus, OH, United States*

Andrea Kirmaier
Boston College, Chestnut Hill, MA, United States

Thomas Klose
Purdue University, West Lafayette, IN, United States

Barbara G. Klupp
*Friedrich-Loeffler-Institute, Greifswald-Insel Riems,
Germany*

David M. Knipe
Harvard Medical School, Boston, MA, United States

Nick J. Knowles
The Pirbright Institute, Pirbright, United Kingdom

Guus Koch
*Wageningen Bioveterinary Research, Lelystad,
The Netherlands*

Renate Koenig
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

Susanne E. Kohalmi
*The University of Western Ontario, London, ON,
Canada*

Hideki Kondo
Okayama University, Kurashiki, Japan

Jennifer L. Konopka-Anstadt
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Eugene V. Koonin
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States and
National Institutes of Health, Bethesda, MD, United States*

Marion P.G. Koopmans
Erasmus Medical Center, Rotterdam, The Netherlands

Richard Kormelink
*Wageningen University and Research, Wageningen,
The Netherlands*

Ioly Kotta-Loizou
Imperial College London, London, United Kingdom

Peter J. Krell
*Department of Molecular and Cellular Biology,
University of Guelph, Guelph, ON, Canada*

Mart Krupovic
Archaeal Virology Unit, Institut Pasteur, Paris, France

Andreas Kuhn
University of Hohenheim, Stuttgart, Germany

Jens H. Kuhn
*National Institutes of Health, Frederick, MD,
United States*

Richard J. Kuhn
Purdue University, West Lafayette, IN, United States

Suvi Kuivanen
University of Helsinki, Helsinki, Finland

Ranjababu Kulasegaram
*Guy's and St Thomas' NHS Foundation Trust, London,
United Kingdom*

Raghavendran Kulasegaran-Shylini
*Department of Pathogen Infection, Faculty of Infectious
and Tropical Diseases, London School of Hygiene and
Tropical Medicine, London, United Kingdom*

Gaurav Kumar
University of Delhi, New Delhi, India

Manish Kumar
Jawaharlal Nehru University, New Delhi, India

Gael Kurath
*US Geological Survey, Western Fisheries Research
Center, Seattle, WA, United States*

Satu Kurkela
*University of Helsinki and Helsinki University Hospital,
Helsinki, Finland*

Wan-Chun Lai
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Kevin Lamkiewicz
University of Jena, Jena, Germany

Rebecca K. Lane
*University of Texas Health Science Center at San
Antonio, San Antonio, TX, United States*

Andrew S. Lang
*Memorial University of Newfoundland, St. John's, NL,
Canada*

Daniel Carlos Ferreira Lanza
Federal University of Rio Grande do Norte, Natal, Brazil

Maija Lappalainen
*HUS Diagnostic Center, HUSLAB, Clinical
Microbiology, University of Helsinki and Helsinki
University Hospital, Helsinki, Finland*

Katherine LaTourrette
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Chris Lauber
*TWINCORE – Center for Experimental and Clinical
Infection Research, Hannover, Germany*

Antonio Lavazza
*The Lombardy and Emilia Romagna Experimental
Zootechnic Institute, Brescia, Italy*

C. Martin Lawrence
Montana State University, Bozeman, MT, United States

Hervé Lecoq
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Young-Min Lee
Utah State University, Logan, UT, United States

Kristen N. LeGault
University of California, Berkeley, CA, United States

James Legg
*International Institute of Tropical Agriculture, Dar es
Salaam, Tanzania*

Anne Legreve
University of Louvain, Louvain-la-Neuve, Belgium

Petr G. Leiman
The University of Texas Medical Branch, Galveston, TX, United States

Stanley M. Lemon
Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Sebastian Leptihn
Zhejiang University-Edinburgh University Institute, Zhejiang University, Haining, China

Dennis J. Lewandowski
University of Florida, Lake Alfred, FL, United States

Sébastien Lhomme
Toulouse University Hospital, Toulouse, France and Toulouse University Paul Sabatier, Toulouse, France

Dawei Li
China Agricultural University, Beijing, China

Guangdi Li
Central South University, Changsha, China

Guoqing Li
Huazhong Agricultural University, Wuhan, China

Yi Li
Peking University, Beijing, China

Zhefeng Li
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Zhenghe Li
Zhejiang University, Hangzhou, China

Jia Q. Liang
South China Agricultural University, Guangzhou, China

Sebastian Liebe
Institute of Sugar Beet Research, Göttingen, Germany

João Paulo Matos Santos Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Bruno Lina
HCL Department of Virology, National Reference Center for Respiratory Viruses, Institute of Infectious Agents, Croix-Rousse Hospital, Lyon, France and Virpath Laboratory, International Center of Research in Infectiology (CIRI), INSERM U1111, CNRS—UMR 5308, École Normale Supérieure de Lyon, University Claude Bernard Lyon, Lyon University, Lyon, France

Walter Ian Lipkin
Columbia University, New York, NY, United States

Jan G. Lisby
Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

Ding X. Liu
South China Agricultural University, Guangzhou, China

Qiang Liu
University of Saskatchewan, Saskatoon, SK, Canada

Sijun Liu
Iowa State University, Ames, IA, United States

Carlos Llorens
Biotechvana, Scientific Park University of Valencia, Valencia, Spain

L. Sue Loesch-Fries
Purdue University, West Lafayette, IN, United States

George P. Lomonosoff
John Innes Centre, Norwich, United Kingdom

L. Letti Lopez
The University of Texas at Austin, Austin, TX, United States

Alan T. Loynachan
University of Kentucky, Lexington, KY, United States

Garry A. Luke
University of St. Andrews, St. Andrews, United Kingdom

M. Luo
University of Alabama at Birmingham, Birmingham, AL, United States

Juan J. López-Moya
Center for Research in Agricultural Genomics and Spanish National Research Council, Barcelona, Spain

Che Ma
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Stuart A. MacFarlane
The James Hutton Institute, Invergowrie, United Kingdom

Saichetana Macherla
J. Craig Venter Institute, La Jolla, CA, United States

Kensaku Maejima
The University of Tokyo, Tokyo, Japan

Fabrizio Maggi
University of Pisa, Pisa, Italy and University of Insubria, Varese, Italy

Melissa S. Maginnis
The University of Maine, Orono, ME, United States

Edgar Maiss
Leibniz University Hannover, Hannover, Germany

Kira S. Makarova
National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, United States

Ariana Manglli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Annette Mankertz
Robert Koch-Institute, Berlin, Germany

Pilar Manrique
The Ohio State University, Wexner Medical Center, Columbus, OH, United States

Shahid Mansoor
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Marco Marklewitz
Institute of Virology, Charité – University Medicine Berlin, Berlin, Germany

Giovanni P. Martelli[†]
University of Bari Aldo Moro, Bari, Italy

Darren P. Martin
University of Cape Town, Cape Town, South Africa

Robert R. Martin
Horticultural Crops Research Unit, Agricultural Research Service, US Department of Agriculture, Corvallis, OR, United States

Manuel Martinez-Garcia
University of Alicante, Alicante, Spain

Francisco Martinez-Hernandez
University of Alicante, Alicante, Spain

Natalia Martín-González
Autonomous University of Madrid, Madrid, Spain

Joaquín Martínez Martínez
Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Manja Marz
University of Jena, Jena, Germany

Andrea Marzi
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, United States

Hema Masarapu
Sri Venkateswara University, Tirupati, India

Chikara Masuta
Hokkaido University, Sapporo, Japan

Carlos P. Mata
University of Leeds, Leeds, United Kingdom

Jelle Matthijnsens
Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

Claire P. Mattison
Centers for Disease Control and Prevention, Atlanta, GA, United States and Cherokee Nation Assurance, Arlington, VA, United States

William McAllister
Rowan University School of Osteopathic Medicine, Stratford, NJ, United States

Alison A. McBride
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Michael McChesney
University of California, Davis, CA, United States

Elaine McCulloch
Quality Control for Molecular Diagnostics (QCMD), Glasgow, United Kingdom

Andrew J. McMichael
University of Oxford, Oxford, United Kingdom

Alexander McPherson
University of California, Irvine, CA, United States

Irene K. Meki
French National Center for Scientific Research, Montpellier, France

Ulrich Melcher
Oklahoma State University, Stillwater, OK, United States

Tomas A. Melgarejo
University of California, Davis, CA, United States

Michael J. Melzer
Department of Plant and Environmental Protection Sciences, University of Hawaii, Honolulu, HI, United States

Luiza Mendonça
University of Oxford, Oxford, United Kingdom

Xiang-Jin Meng
Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Peter P.C. Mertens
University of Nottingham, Sutton Bonington, United Kingdom

[†]Deceased.

Thomas C. Mettenleiter
Friedrich-Loeffler-Institute, Greifswald-Insel Riems, Germany

Philip D. Minor
St Albans, United Kingdom

Ali Mirazimi
National Veterinary Institute, Uppsala, Sweden and Karolinska Hospital University, Huddinge, Sweden

Nischay Mishra
Columbia University, New York, NY, United States

Edward S. Mocarski
Emory University School of Medicine, Atlanta, GA, United States

Florian Mock
University of Jena, Jena, Germany

Volker Moennig
University of Veterinary Medicine, Hannover, Germany

Ian J. Molineux
The University of Texas at Austin, Austin, TX, United States

Aderito L. Monjane
Norwegian Veterinary Institute, Oslo, Finland

Jacen S. Moore
University of Tennessee Health Science Center, Memphis, TN, United States

Marc C. Morais
The University of Texas Medical Branch, Galveston, TX, United States

Cristina Moraru
Institute for Chemistry and Biology of the Marine Environment, Oldenburg, Germany

Hiromitsu Moriyama
Tokyo University of Agriculture and Technology, Tokyo, Japan

Sergey Y. Morozov
Lomonosov Moscow State University, Moscow, Russia

Thomas E. Morrison
University of Colorado School of Medicine, Aurora, CO, United States

Léa Morvan
University of Liège, Liège, Belgium

Benoît Moury
Plant Pathology Unit, INRAE – French National Institute for Agriculture, Food and Environment, Montfavet, France

Muhammad Mubin
University of Agriculture, Faisalabad, Pakistan

Nicolas J. Mueller
University Hospital of Zurich, Zurich, Switzerland

Emmanuelle Muller
The French Agricultural Research Center for International Development, Joint Research Units–Biology and Genetics of Plant-Pathogen Interactions, Montpellier, France and Biology and Genetics of Plant-Pathogen Interactions, University of Montpellier, The French Agricultural Research Center for International Development, French National Institute for Agricultural Research, Montpellier SupAgro, Montpellier, France

John S. Munday
Massey University, Palmerston North, New Zealand

Jacob H. Munson-McGee
Montana State University, Bozeman, MT, United States and Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Hacer Muratoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Kenan C. Murphy
University of Massachusetts Medical School, Worcester, MA, United States

Ugrappa Nagalakshmi
University of California, Davis, CA, United States

Keizo Nagasaki
Kochi University, Nankoku, Japan

Nazia Nahid
GC University, Faisalabad, Pakistan and University of Agriculture, Faisalabad, Pakistan

Venugopal Nair
The Pirbright Institute, Pirbright, United Kingdom

Remziye Nalçacıoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Shigetou Namba
The University of Tokyo, Tokyo, Japan

Rubab Z. Naqvi
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Rachel Nash
The Pirbright Institute, Surrey, United Kingdom

C.K. Navaratnarajah
Purdue University, West Lafayette, IN, United States

Maria A. Navarrete-Muñoz
Biotechvana, Madrid, Spain; Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid; and Rey Juan Carlos University Hospital, Móstoles, Spain

Jesús Navas-Castillo
Institute for Mediterranean and Subtropical Horticulture "La Mayora"—Spanish National Research Council—University of Malaga, Algarrobo-Costa, Málaga, Spain

Muhammad S. Nawaz-ul-Rehman
University of Agriculture, Faisalabad, Pakistan

Christopher L. Netherton
The Pirbright Institute, Pirbright, United Kingdom

Thu V.P. Nguyen
Baylor College of Medicine, Houston, TX, United States

Annette Niehl
Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

Hubert G.M. Niesters
Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Jozef I. Nissimov
University of Waterloo, Waterloo, ON, Canada

Norman Noah
London School of Hygiene and Tropical Medicine, London, United Kingdom

Mauricio L. Nogueira
São José do Rio Preto School of Medicine, São José do Rio Preto, São Paulo, Brazil

Johan Nordgren
Linköping University, Linköping, Sweden

C. Micha Nübling
Paul-Ehrlich-Institute, Langen, Germany

Visa Nurmi
University of Helsinki, Helsinki, Finland

Donald L. Nuss
University of Maryland, Rockville, MD, United States

M. Steven Oberste
Centers for Disease Control and Prevention, Atlanta, GA, United States

Hiroyuki Ogata
Institute for Chemical Research, Kyoto University, Kyoto, Japan

Ane Ogbe
University of Oxford, Oxford, United Kingdom

Hanna M. Oksanen
Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Graziele P. Oliveira
Federal University of Minas Gerais, Belo Horizonte, Brazil

Francesco Origi
University of Bern, Bern, Switzerland

Nikolaus Osterrieder
Free University of Berlin, Berlin, Germany

Robert A. Owens
Beltsville Agricultural Research Center, Beltsville, MD 20705, United States

Emine Özsahin
University of Guelph, Guelph, ON, Canada

Sergi Padilla-Parra
University of Oxford, Oxford, United Kingdom; Department of Infectious Diseases, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom; and Randall Division of Cell and Molecular Biophysics, King's College London, London, United Kingdom

Joshua Pajak
Duke University, Durham, NC, United States

Massimo Palmarini
MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Amanda R. Panfil
The Ohio State University, Columbus, OH, United States

Marcus Panning
Institute of Virology, Freiburg University Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Vitantonio Pantaleo
National Research Council, Research Unit of Bari, Bari, Italy

Anna Papa
Aristotle University of Thessaloniki, Thessaloniki, Greece

Nikolaos Pappas
Utrecht University, Utrecht, The Netherlands

Hanu R. Pappu
Washington State University, Pullman, WA, United States

Kristin N. Parent
*Michigan State University, East Lansing, MI,
United States*

Colin R. Parrish
Cornell University, Ithaca, NY, United States

A. Lorena Passarelli
Kansas State University, Manhattan, KS, United States

Basavaprabhu L. Patil
*ICAR–Indian Institute of Horticultural Research,
Bengaluru, India*

Jade Pattyn
University of Antwerp, Antwerp, Belgium

T.A. Paul
Cornell University, Ithaca, NY, United States

Lillian Pavlik
*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Susan L. Payne
Texas A&M University, College Station, TX, United States

Michael N. Pearson
The University of Auckland, Auckland, New Zealand

Mark E. Peeples
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University College of
Medicine, Columbus, OH, United States*

Ben Peeters
*Wageningen Bioveterinary Research, Lelystad,
The Netherlands*

Joseph S.M. Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Malik Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Judit J. Péntzes
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Miryam Pérez-Cañamás
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Marta Pérez-Illana
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Jaume Pérez-Sánchez
*Institute of Aquaculture Torre de la Sal, Spanish
National Research Council, Castellon, Spain*

Jean-Marie Peron
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Karin E. Peterson
National Institutes of Health, Hamilton, MT, United States

Karel Petrzik
*Biology Center CAS, Institute of Plant Molecular
Biology, České Budějovice, Czech Republic*

Mahtab Peyambari
*Pennsylvania State University, State College, PA,
United States*

Sujal Phadke
J. Craig Venter Institute, La Jolla, CA, United States

Hanh T. Pham
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Mauro Pistello
University of Pisa, Pisa, Italy

Daniel Ponndorf
John Innes Centre, Norwich, United Kingdom

Leo L.M. Poon
The University of Hong Kong, Pok Fu Lam, Hong Kong

Welkin H. Pope
University of Pittsburgh, Pittsburgh, PA, United States

Minna M. Poranen
University of Helsinki, Helsinki, Finland

Claudine Porta
*The Pirbright Institute, Pirbright, United Kingdom and
University of Oxford, Oxford, United Kingdom*

Samuel S. Porter
*National Institute of Allergy and Infectious Diseases,
National Institutes of Health, Bethesda, MD,
United States and University of Maryland, College Park,
MD, United States*

Frank A. Post
*King's College Hospital NHS Foundation Trust, London,
United Kingdom*

Nils Poulicard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

David Prangishvili
*Institut Pasteur, Paris, France and Ivane Javakhishvili
Tbilisi State University, Tbilisi, Georgia*

B. V. Venkataram Prasad
Baylor College of Medicine, Houston, TX, United States

Lalita Priyamvada
Centers for Disease Control and Prevention, Atlanta, GA, United States

Simone Prospero
Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

Elisabeth Puchhammer-Stöckl
Medical University of Vienna, Vienna, Austria

Jianming Qiu
University of Kansas Medical Center, Kansas City, KS, United States

S.L. Quackenbush
Colorado State University, Fort Collins, CO, United States

Killian J. Quinn
King's College Hospital, London, United Kingdom

Diego F. Quito-Avila
Department of Life Sciences, ESPOL Polytechnic University, Guayaquil, Ecuador

Frank Rabenstein
Julius Kühn Institute, Quedlinburg, Germany

Sheli R. Radoshitzky
United States Army Medical Research Institute of Infectious Diseases, Frederick, MD, United States

Saleem U. Rahman
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Mbolarinosy Rakotomalala
FOFIFA, Antananarivo, Madagascar

Norma Rallon
Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid and Rey Juan Carlos University Hospital, Móstoles, Spain

Robert P. Rambo
Diamond Light Source, Didcot, United Kingdom

Bertha Cecilia Ramirez
The Institute for Integrative Biology of the Cell, The French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

María D. Ramos-Barbero
University of Alicante, Alicante, Spain

Chris M. Rands
University of Geneva Medical School and Swiss Institute of Bioinformatics, Geneva, Switzerland

Venigalla B. Rao
The Catholic University of America, Washington, DC, United States

Janne J. Ravantti
University of Helsinki, Helsinki, Finland

Mandy Ravensbergen
Wageningen University and Research, Wageningen, The Netherlands

Georget Y. Reaiche-Miller
The University of Adelaide, Adelaide, SA, Australia

D.V.R. Reddy
International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

Vishwanatha R.A.P. Reddy
The Pirbright Institute, Pirbright, United Kingdom

Juan Reguera
Aix-Marseille University, French National Center for Scientific Research, Marseille, France and French National Institute of Health and Medical Research, Marseille, France

William K. Reisen
University of California, Davis, CA, United States

Jingshan Ren
University of Oxford, Oxford, United Kingdom

Renato O. Resende
University of Brasilia, Brasilia, Brazil

Peter A. Revill
The Peter Doherty Institute of Infection and Immunity, Royal Melbourne Hospital, Melbourne, VIC, Australia

Félix A. Rey
Institut Pasteur, Paris, France

Lara Rheinemann
University of Utah, Salt Lake City, UT, United States

Simone G. Ribeiro
Embrapa Genetic Resources and Biotechnology, Brasília, Brazil

Daniel Rigling
Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Cristina Risco
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Efraín E. Rivera-Serrano

Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Cécile Robin

INRAE – French National Research Institute for Agriculture, Food and Environment, UMR BIOGECO, Cestas, France

Rodrigo A.L. Rodrigues

Federal University of Minas Gerais, Belo Horizonte, Brazil

Elina Roine

University of Helsinki, Helsinki, Finland

Maria R. Rojas

University of California, Davis, CA, United States

Marilyn J. Roossinck

Pennsylvania State University, State College, PA, United States

Vera I.D. Ros

Wageningen University and Research, Wageningen, The Netherlands

Cristina Rosa

Pennsylvania State University, University Park, PA, United States

Hanna Rose

Leibniz University Hannover, Hannover, Germany

David A. Rosenbaum

University of Florida, Gainesville, FL, United States

Shannan L. Rossi

The University of Texas Medical Branch, Galveston, TX, United States

Michael G. Rossmann[†]

Purdue University, West Lafayette, IN, United States

L. Roux

University of Geneva Medical School, Geneva, Switzerland

Simon Roux

US Department of Energy Joint Genome Institute, Walnut Creek, CA, United States

J. Rovnak

Colorado State University, Fort Collins, CO, United States

David J. Rowlands

University of Leeds, Leeds, United Kingdom

Polly Roy

Department of Pathogen Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom and University of Reading, Reading, United Kingdom

Aaron P. Roznowski

The University of Texas at Austin, Austin, TX, United States and University of Arizona, Tucson, AZ, United States

Luisa Rubino

Institute for Sustainable Plant Protection, National Research Council, Bari, Italy

Olli Ruuskanen

Turku University Hospital, Turku, Finland

Eugene V. Ryabov

USDA, Agricultural Research Service, Beltsville, MD, United States

Martin D. Ryan

University of St. Andrews, St. Andrews, United Kingdom

Ki H. Ryu

Seoul Women's University, Seoul, South Korea

Hanns-Joachim Rziha

Eberhard Karls University of Tübingen, Tübingen, Germany

Sead Sabanadzovic

Mississippi State University, Starkville, MS, United States

Roghayeh Safari

Cellular and Molecular Epigenetics (GIGA), Liège, Belgium and Molecular Biology (TERRA), Gembloux, Belgium

Azeez Sait Sahul Hameed

C. Abdul Hakeem College, Melvisharam, India

Nicole Samies

University of Alabama at Birmingham, Birmingham, AL, United States

Carmen San Martín

National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Ruth-Anne Sandaa

Department of Biological Sciences, University of Bergen, Bergen, Norway

Hélène Sanfaçon

Agriculture and Agri-Food Canada, Summerland, BC, Canada

[†]Deceased.

Rafael Sanjuán
*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia-CSIC, Valencia, Spain*

Neeraja Sankaran
Utrecht University, Utrecht, The Netherlands

Fernando Santos
University of Alicante, Alicante, Spain

Cecilia Sarmiento
Tallinn University of Technology, Tallinn, Estonia

Takahide Sasaya
*National Agriculture and Food Research Organization,
Fukuyama, Japan*

Preethi Sathanantham
Virginia Tech, Blacksburg, VA, United States

Panayampalli S. Satheshkumar
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Yukiyo Sato
Okayama University, Kurashiki, Japan

Andreas Sauerbrei
Jena University Hospital, Jena, Germany

Eugene I. Savenkov
*Swedish University of Agricultural Sciences, Uppsala,
Sweden and Linnean Center for Plant Biology, Uppsala,
Sweden*

Carita Savolainen-Kopra
*National Institute for Health and Welfare, Helsinki,
Finland*

Kay Scheets
*Oklahoma State University, Stillwater, OK,
United States*

Uffe V. Schenider
*Copenhagen University Hospital Hvidovre, Hvidovre,
Denmark*

Richard H. Scheuermann
*J. Craig Venter Institute, La Jolla, CA, United States;
University of California, San Diego, CA, United States;
La Jolla Institute for Immunology, La Jolla, CA,
United States; and Global Virus Network, Baltimore,
MD, United States*

Manfred J. Schmitt
Saarland University, Saarbrücken, Germany

James E. Schoelz
University of Missouri, Columbia, MO, United States

Jason R. Schrad
*Michigan State University, East Lansing, MI,
United States*

Declan C. Schroeder
*University of Reading, Reading, United Kingdom and
University of Minnesota, St. Paul, MN, United States*

Stacey Schultz-Cherry
*St. Jude Children's Research Hospital, Memphis, TN,
United States*

Thomas F. Schulz
*Hannover Medical School, Institute of Virology,
Hannover, Germany and German Center for Infection
Research, Hannover-Braunschweig Site, Braunschweig,
Germany*

Catherine A. Scougall
The University of Adelaide, Adelaide, SA, Australia

Kimberley D. Seed
University of California, Berkeley, CA, United States

Joaquim Segalés
*Departament of Animal Health and Anatomy, Faculty of
Veterinary Medicine, Autonomous University of
Barcelona, Barcelona, Spain; Animal Health Research
Center (CReSA) – Institute of Agrifood Research and
Technology (IRTA), Campus UAB, Barcelona, Spain;
and OIE Collaborating Center for the Research and
Control of Emerging and Re-emerging Swine Diseases in
Europe (IRTA-CReSA), Barcelona, Spain*

Mateo Seoane-Blanco
National Center for Biotechnology, Madrid, Spain

Madhumati Sevvana
Purdue University, West Lafayette, IN, United States

Kazım Sezen
*Department of Biology, Karadeniz Technical University,
Trabzon, Turkey*

Arvind Sharma
Institut Pasteur, Paris, France

Sumit Sharma
Linköping University, Linköping, Sweden

James M. Sharp
*University of Zaragoza, Zaragoza, Spain and Edinburgh,
United Kingdom*

Qunxin She
Shandong University, Qingdao, China

Keith E. Shearwin
The University of Adelaide, Adelaide, SA, Australia

Hanako Shimura
Hokkaido University, Sapporo, Japan

Reina S. Sikkema
Erasmus Medical Center, Rotterdam, The Netherlands

Aaron Simkovich
Agriculture and Agri-Food Canada, London, ON,
Canada and The University of Western Ontario,
London, ON, Canada

Peter Simmonds
University of Oxford, Oxford, United Kingdom

Tarja Sironen
University of Helsinki, Helsinki, Finland

Susanna Sissonen
Finnish Institute for Health and Welfare, Helsinki,
Finland

Michael A. Skinner
Imperial College London, London, United Kingdom

Douglas E. Smith
University of California, San Diego, La Jolla, CA,
United States

Melvyn Smith
Viapath Analytics, Specialist Virology Centre, King's
College NHS Foundation Trust, London, United
Kingdom

Thomas J. Smith
The University of Texas Medical Branch, Galveston, TX,
United States

Teemu Smura
Helsinki University Hospital and University of Helsinki,
Helsinki, Finland

Eric J. Snijder
Leiden University Medical Center, Leiden,
The Netherlands

Gisela Soboll Hussey
Michigan State University, East Lansing, MI,
United States

Maria Söderlund-Venermo
University of Helsinki, Helsinki, Finland

Merike Sõmera
Tallinn University of Technology, Tallinn, Estonia

Eun G. Song
Seoul Women's University, Seoul, South Korea

Milan J. Sonneveld
Erasmus University Medical Center, Rotterdam,
The Netherlands

Beatriz Soriano
Biotechvana, Scientific Park University of Valencia and
Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain

Thomas E. Spencer
University of Missouri, Columbia, MO, United States

Pothur Sreenivasulu
Sri Venkateswara University, Tirupati, India

Ashley L. St. John
Duke-NUS Medical School, Singapore, Singapore

David K. Stammers
University of Oxford, Oxford, United Kingdom

John Stanley
John Innes Centre, Colney, United Kingdom

Glyn Stanway
University of Essex, Colchester, United Kingdom

Thilo Stehle
University of Tuebingen, Tuebingen, Germany and
Vanderbilt University School of Medicine, Nashville, TN,
United States

Gregory W. Stevenson
Iowa State University, Ames, IA, United States

Lucy Rae Stewart
Agricultural Research Service, US Department of
Agriculture, Wooster, OH, United States

C.C.M.M. Stijger
Wageningen University and Research Center, Bleiswijk,
The Netherlands

Peter G. Stockley
University of Leeds, Leeds, United Kingdom

David Stone
Weymouth Laboratory, Centre for Environment, Fisheries
and Aquaculture Science, Weymouth, United Kingdom

Ashley E Strother
The University of Texas Medical Branch, Galveston, TX,
United States

Sundharraman Subramanian
Michigan State University, East Lansing, MI,
United States

William C. Summers
Yale University, New Haven, CT, United States

Liyang Sun
Northwest A&F University, Yangling, China

Wesley I. Sundquist
University of Utah, Salt Lake City, UT, United States

Petri Susi
University of Turku, Turku, Finland

Curtis A. Suttle
University of British Columbia, Vancouver, BC, Canada

Nobuhiro Suzuki
Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

Lennart Svensson
Linköping University, Linköping, Sweden and Karolinska Institute, Stockholm, Sweden

Ronald Swanstrom
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Daniele M. Swetnam
University of California, Davis, CA, United states

Moriah L. Szpara
Pennsylvania State University, University Park, PA, United States

Keisuke Tabata
Heidelberg University, Heidelberg, Germany

Anna Taglienti
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Naoki Takeshita
Tokyo University of Agriculture and Technology, Fuchu, Japan

Kana Takeshita Urayama
Tokyo University of Agriculture and Technology, Fuchu, Japan

Michael E. Taliansky
The James Hutton Institute, Dundee, United Kingdom

Pan Tao
The Catholic University of America, Washington, DC, United States

Jacqueline E. Tate
Centers for Disease Control and Prevention, Atlanta, GA, United States

Satyanarayana Tatineni
Agricultural Research Service, US Department of Agriculture, Lincoln, NE, United States and University of Nebraska–Lincoln, Lincoln, NE, United States

Sisko Tauriainen
University of Turku, Turku, Finland

Norbert Tautz
University of Luebeck, Luebeck, Germany

Paulo Tavares
Institute for Integrative Biology of the Cell, CEA, CNRS, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

Nicholas M.I. Taylor
University of Copenhagen, Copenhagen, Denmark

Xu Tengzhi
University of California, Davis, CA, United States

Raquel Tenorio
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Robert B. Tesh
The University of Texas Medical Branch, Galveston, TX, United States

Vaskar Thapa
Pennsylvania State University, State College, PA, United States

John E. Thomas
The University of Queensland, Brisbane, QLD, Australia

Julie A. Thomas
Rochester Institute of Technology, Rochester, NY, United States

Lynn C. Thomason
Frederick National Laboratory for Cancer Research, Frederick, MD, United States

Elizabeth Ashley Thompson
The University of Southern Mississippi, Hattiesburg, MS, United States

Jeremy R. Thompson
Cornell University, Ithaca, NY, United States

Antonio Tiberini
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Peter Tijssen
National Institute of Scientific Research – Armand-Frappier Health Research Centre, Microbiology and Immunology, Laval, QC, Canada

Yuji Tomaru
Japan Fisheries Research and Education Agency, Kanagawa, Japan

Laura Tomassoli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Ruben Torres
National Biotechnology Center–Spanish National Research Council, Madrid, Spain

Jia Q. Truong
The University of Adelaide, Adelaide, SA, Australia

Erkki Truve[†]
Tallinn University of Technology, Tallinn, Estonia

Chih-Hsuan Tsai
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Roman Tuma
University of Leeds, Leeds, United Kingdom and
University of South Bohemia, České Budějovice, Czech Republic

Topi Turunen
Infectious Disease Unit, Espoo, Finland and Finnish
Institute for Health and Welfare, Helsinki, Finland

Reidun Twarock
University of York, York, United Kingdom

Ioannis E. Tzanetakis
University of Arkansas, Fayetteville, United States

Antti Vaheri
University of Helsinki, Helsinki, Finland

Eeva J. Vainio
Natural Resources Institute Finland (Luke), Helsinki, Finland

Anna M. Vaira
Institute for Sustainable Plant Protection, National
Research Council of Italy, Torino, Italy

Steven M. Valles
Center for Medical, Agricultural and Veterinary
Entomology, Agricultural Research Service, US
Department of Agriculture, Gainesville, FL,
United States

Adrián Valli
National Center for Biotechnology-Spanish National
Research Council, Madrid, Spain

Rodrigo A. Valverde
Louisiana State University Agricultural Center, Baton
Rouge, United States

Pierre Van Damme
University of Antwerp, Antwerp, Belgium

Rene A.A. van der Vlugt
Wageningen University and Research Center,
Wageningen, The Netherlands

Bernard A.M. Van der Zeijst
Leiden University Medical Center, Leiden,
The Netherlands

Koenraad Van Doorslaer
University of Arizona, Tucson, AZ, United States

James L. Van Etten
University of Nebraska–Lincoln, Lincoln, NE,
United States

Suzanne van Meer
University Medical Center Utrecht, Utrecht,
The Netherlands

Monique M. van Oers
Wageningen University and Research, Wageningen,
The Netherlands

Mark J. van Raaij
National Center for Biotechnology, Madrid, Spain

Marc H.V. Van Regenmortel
University of Strasbourg, Strasbourg, France

Piet A. van Rijn
Wageningen Bioveterinary Research, Lelystad, The
Netherlands and North-West University, Potchefstroom,
South Africa

Alain Vanderplasschen
University of Liège, Liège, Belgium

Dana L. Vanlandingham
College of Veterinary Medicine, Kansas State University,
Manhattan, KS, United States

Olli Vapalahti
Helsinki University Hospital and University of Helsinki,
Helsinki, Finland

Mark Varrelmann
Institute of Sugar Beet Research, Göttingen, Germany

Nikos Vasilakis
The University of Texas Medical Branch, Galveston, TX,
United States

Michael Veit
Free University of Berlin, Berlin, Germany

Česlovas Venclovas
Vilnius University, Vilnius, Lithuania

H. Josef Vetten
Julius Kühn Institute, Braunschweig, Germany

Marli Vlok
University of British Columbia, Vancouver, BC, Canada

Anne-Nathalie Volkoff
Diversity, Genomes and Insects-Microorganisms
Interactions, National Institute of Agricultural Research,
University of Montpellier, Montpellier, France

Ian E.H. Voorhees
Cornell University, Ithaca, NY, United States

Alex Vorsters
University of Antwerp, Antwerp, Belgium

[†]Deceased.

Jonathan D.F. Wadsworth
UCL Institute of Prion Diseases, London, United Kingdom

Peter J. Walker
The University of Queensland, St. Lucia, QLD, Australia

Paul Wallace
*Quality Control for Molecular Diagnostics (QCMD),
Glasgow, United Kingdom*

Aiming Wang
*Agriculture and Agri-Food Canada, London, ON,
Canada*

Jen-Ren Wang
National Cheng Kung University, Tainan, Taiwan

Lin-Fa Wang
Duke-NUS Medical School, Singapore, Singapore

Nan Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiangxi Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiaofeng Wang
Virginia Tech, Blacksburg, VA, United States

Katherine N. Ward
University College London, London, United Kingdom

Matti Waris
University of Turku, Turku, Finland

Ranjit Warrier
Purdue University, West Lafayette, IN, United States

Daniel Watterson
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Marta L. Wayne
University of Florida, Gainesville, FL, United States

Friedemann Weber
*FB 10 – Institute for Virology, Justus Liebig University
Giessen, Giessen, Germany*

Sung-Chan Wei
*Institute of Molecular Biology, Academia Sinica, Taipei,
Taiwan*

Robin A. Weiss
University College London, London, United Kingdom

Tao Weitao
*Southwest Baptist University, Bolivar, MO,
United States*

Kerstin Wernike
Friedrich-Loeffler-Institute, Insel Riems, Germany

Rachel J. Whitaker
*University of Illinois at Urbana-Champaign, Urbana, IL,
United States*

K. Andrew White
York University, Toronto, ON, Canada

Anna E. Whitfield
*North Carolina State University, Raleigh, NC,
United States*

Richard Whitley
*University of Alabama at Birmingham, Birmingham,
AL, United States*

Reed B. Wickner
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Luc Willems
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Brian J. Willett
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Alexis Williams
*The University of Texas Medical Branch, Galveston, TX,
United States*

Stephen A. Winchester
*Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom*

Clayton W. Winkler
*National Institutes of Health, Hamilton, MT,
United States*

Stephan Winter
*Leibniz Institute – DSMZ – German Collection of
Microorganisms and Cell Cultures GmbH, Braunschweig,
Germany*

William M. Wintermantel
*Agricultural Research Service, US Department of
Agriculture, Salinas, CA, United States*

Jennifer Wirth
Montana State University, Bozeman, MT, United States

Yuri I. Wolf
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States*

Thorsten Wolff
Robert Koch Institute, Berlin, Germany

Blaide Woodburn
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Michael E. Woodson
The University of Texas Medical Branch, Galveston, TX, United States

Courtney Woolsey
The University of Texas Medical Branch, Galveston, TX, United States

Chien-Fu Wu
Tokyo University of Agriculture and Technology, Fuchu, Japan

Mingde Wu
Huazhong Agricultural University, Wuhan, China

Songsong Wu
Huazhong Agricultural University, Wuhan, China

Yan Xiang
University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

Jiatao Xie
Huazhong Agricultural University, Wuhan, China

Zhuang Xiong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Hajime. Yaegashi
Institute of Fruit Tree and Tea Science, NARO, Morioka, Japan

Mehtap Yakupoğlu
Trabzon University, Trabzon, Turkey

Yasuyuki Yamaji
The University of Tokyo, Tokyo, Japan

Meng Yang
China Agricultural University, Beijing, China

Teng-Chieh Yang
Scarsdale, NY, United States

Qin Yao
Jiangsu University, Zhenjiang, China

Tianyou Yao
Baylor College of Medicine, Houston, TX, United States

Nobuyuki Yoshikawa
Iwate University, Morioka, Japan

George R. Young
Francis Crick Institute, London, United Kingdom

Lawrence S. Young
University of Warwick, Coventry, United Kingdom

Mark J. Young
Montana State University, Bozeman, MT, United States

Ry Young
Texas A&M University, College Station, TX, United States

Isaac T. Younker
University of Alabama, Tuscaloosa, AL, United States

Qian Yu
School of Life Sciences, Jiangsu University, Zhenjiang, China

Sang-Im Yun
Utah State University, Logan, UT, United States

Fauzia Zarreen
University of Delhi, New Delhi, India

Francisco M. Zerbini
Federal University of Viçosa, Viçosa, Brazil

Dong-Xiu Zhang
University of Maryland, Rockville, MD, United States

Jianqiang Zhang
Iowa State University, Ames, IA, United States

Junjie Zhang
Texas A&M University, College Station, TX, United States

Long Zhang
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Pan Zhang
Central South University, Changsha, China

Peijun Zhang
University of Oxford, Oxford, United Kingdom and Electron Bio-Imaging Centre, Diamond Light Source, Didcot, United Kingdom

Tao Zhang
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Yongliang Zhang
China Agricultural University, Beijing, China

Zhenlu Zhang
Shandong Agricultural University, Tai'an, China

Lixia Zhou
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Ling Zhu
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Heiko Ziebell
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

John Ziebuhr
*The Queen's University of Belfast, Belfast,
United Kingdom*

Jeffrey J. Zimmerman
Iowa State University, Ames, IA, United States

Falk Zucker
*Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany*

CONTENT OF ALL VOLUMES

Editors in Chief	v
Editorial Board	vii
Section Editors	ix
Foreword	xv
Preface	xvii
Guide to Use	xix
List of Contributors	xxi

VOLUME 1

The Virus as a Concept – Fundamentals of Virology

A Brief History of Virology <i>David J Rowlands</i>	3
The Origin of Viruses <i>Patrick Forterre and Morgan Gaïa</i>	14
The Virocell Concept <i>Patrick Forterre</i>	23
Virus Taxonomy <i>Jens H Kuhn</i>	28
The Greater Virus World and Its Evolution <i>Eugene V Koonin and Valerian V Dolja</i>	38
The Virus Species Concept <i>Peter Simmonds</i>	47
Genetic Diversity and Evolution of Viral Populations <i>Rafael Sanjuán and Pilar Domingo-Calap</i>	53
Mechanisms of RNA Virus Evolution <i>Lisa M Bono and Siobain Duffy</i>	62
Mechanisms of DNA Virus Evolution <i>Moriah L Szpara and Koenraad Van Doorslaer</i>	71
Paleovirology <i>Clément Gilbert</i>	79
Evolution Steered by Structure <i>Nicola GA Abrescia</i>	87

Pairwise Sequence Comparison in Virology <i>Tao Zhang, Zheng Gong, Tongkun Guo, Zhuang Xiong, and Yiming Bao</i>	100
Computational Analysis of Recombination in Viral Nucleotide Sequences <i>Miguel Arenas</i>	108
Phylogeny of Viruses <i>Alexander E Gorbalenya and Chris Lauber</i>	116
Virus Bioinformatics <i>Nikolaos Pappas, Simon Roux, Martin Hölzer, Kevin Lamkiewicz, Florian Mock, Manja Marz, and Bas E Dutilh</i>	124
Metagenomics in Virology <i>Simon Roux, Jelle Matthijnsens, and Bas E Dutilh</i>	133
Database and Analytical Resources for Viral Research Community <i>Sujal Phadke, Saichetana Macherla, and Richard H Scheuermann</i>	141
Classification of the Viral World Based on Atomic Level Structures <i>Janne J Ravantti and Nicola GA Abrescia</i>	153
Isolating, Culturing, and Purifying Viruses With a Focus on Bacterial and Archaeal Viruses <i>Katri Eskelin and Hanna M Oksanen</i>	162
High Throughput Sequencing and Virology <i>Graham L Freimanis and Nick J Knowles</i>	175
Single-Virus Genomics: Studying Uncultured Viruses, One at a Time <i>Manuel Martinez-Garcia, Francisco Martinez-Hernandez, and Joaquín Martínez Martínez</i>	184
Biophysical Characterizations in the Solution State <i>Robert P Rambo and Katsuaki Inoue</i>	191
Virus Crystallography <i>Jonathan M Grimes</i>	199
Advanced Light and Correlative Microscopy in Virology <i>Sergi Padilla-Parra, Charles A Coomer, and Irene Carlon-Andres</i>	208
Atomic Force Microscopy (AFM) Investigation of Viruses <i>Alexander McPherson</i>	218
Cryo-Electron Microscopy (CEM) Structures of Viruses <i>David Chmielewski and Wah Chiu</i>	233
Analysis of Viruses in the Cellular Context by Electron Tomography <i>Peijun Zhang and Luiza Mendonça</i>	242
Mathematical Modeling of Virus Architecture <i>Reidun Twarock</i>	248
Principles of Virus Structure <i>Madhumati Sevvana, Thomas Klose, and Michael G Rossmann[†]</i>	257
Structures of Small Icosahedral Viruses <i>Elizabeth E Fry, Jingshan Ren, and Claudine Porta</i>	278
Structural Principles of the Flavivirus Particle Organization and of Its Conformational Changes <i>Stéphane Duquerroy, Arvind Sharma, and Félix A Rey</i>	290
Reoviruses (Reoviridae) and Their Structural Relatives <i>Liya Hu, Mary K Estes, and B V Venkataram Prasad</i>	303

[†]Deceased.

Structures of Tailed Phages and Herpesviruses (<i>Herpesviridae</i>) <i>Montserrat Fàbrega-Ferrer and Miquel Coll</i>	318
Adenoviruses (<i>Adenoviridae</i>) and Their Structural Relatives <i>Gabriela N Condezo, Natalia Martín-González, Marta Pérez-Illana, Mercedes Hernando-Pérez, José Gallardo, and Carmen San Martín</i>	329
Negative Single-Stranded RNA Viruses (<i>Mononegavirales</i>): A Structural View <i>Juan Reguera</i>	345
Structure of Retrovirus Particles (<i>Retroviridae</i>) <i>David K Stammers and Jingshan Ren</i>	352
Structure of Helical Viruses <i>C Martin Lawrence</i>	362
Giant Viruses and Their Virophage Parasites <i>Rodrigo AL Rodrigues, Ana CdSP Andrade, Grazielle P Oliveira, and Jônatas S Abrahão</i>	372
Viral Replication Cycle <i>AJ Cann</i>	382
Viral Receptors <i>José M Casasnovas and Thilo Stehle</i>	388
Bacterial and Archeal Virus Entry <i>Minna M Poranen and Aušra Domanska</i>	402
Nonenveloped Eukaryotic Virus Entry <i>Ian M Jones and Polly Roy</i>	409
Enveloped Virus Membrane Fusion <i>Aurélié A Albertini and Yves Gaudin</i>	417
Genome Replication of Bacterial and Archaeal Viruses <i>Česlovas Venclovas</i>	429
Viral Transcription <i>David LV Bauer and Ervin Fodor</i>	439
Translation of Viral Proteins <i>Martin D Ryan and Garry A Luke</i>	444
Recombination <i>Jozef J Bujarski</i>	460
Assembly of Viruses: Enveloped Particles <i>CK Navaratnarajah, R Warriar, and RJ Kuhn</i>	468
Assembly of Viruses: Nonenveloped Particles <i>M Luo</i>	475
Virion Assembly: From Small Picornaviruses (<i>Picornaviridae</i>) to Large Herpesviruses (<i>Herpesviridae</i>) <i>Ling Zhu, Nan Wang, and Xiangxi Wang</i>	480
Genome Packaging <i>Richard J Bingham, Reidun Twarock, Carlos P Mata, and Peter G Stockley</i>	488
Virus Factories <i>Isabel Fernández de Castro, Raquel Tenorio, and Cristina Risco</i>	495
Release of Phages From Prokaryotic Cells <i>Jesse Cahill and Ry Young</i>	501
Virus Budding <i>Lara Rheinemann and Wesley I Sundquist</i>	519

Vesicle-Mediated Transcytosis and Export of Viruses <i>Efraín E Rivera-Serrano and Stanley M Lemon</i>	529
Vector Transmission of Animal Viruses <i>Houssam Attoui, Fauziah Mohd Jaafar, Rennos Fragkoudis, and Peter PC Mertens</i>	542
The Human Virome <i>Alexia Bordigoni, Sébastien Halary, and Christelle Desnues</i>	552
Epidemiology of Human and Animal Viral Diseases <i>Michael Edelstein</i>	559
Zoonosis, Emerging and Re-Emerging Viral Diseases <i>Janet M Daly</i>	569
Antiviral Innate Immunity: Introduction <i>Friedemann Weber</i>	577
Humoral and T Cell-Mediated Immunity to Viruses <i>Ane Ogbe and Lucy Dorrell</i>	584
Antigenicity and Antigenic Variation <i>Kuan-Ying A Huang, Xiaorui Chen, Che Ma, Dayna Cheng, Jen-Ren Wang, and Wan-Chun Lai</i>	597
Antigen Presentation <i>Andrew J McMichael</i>	601
Defense Against Viruses and Other Genetic Parasites in Prokaryotes <i>Kira S Makarova, Yuri I Wolf, and Eugene V Koonin</i>	606
Defective-Interfering Viruses <i>L Roux</i>	617
Ecology and Global Impacts of Viruses <i>Joanne B Emerson</i>	621
The Role of Retroviruses in Cellular Evolution <i>Andrea Kirmaier and Welkin E Johnson</i>	627
The Role of Bacteriophages in Bacterial Evolution <i>Chris M Rands and Harald Brüssow</i>	633
Viruses and Their Potential for Bioterrorism <i>Dana L Vanlandingham and Stephen Higgs</i>	644
The Use of Viral Promoters in Expression Vectors <i>Ian M Jones</i>	652
Oncolytic Viruses <i>Laura Burga and Mihnea Bostina</i>	658
Biotechnology Approaches to Modern Vaccine Design <i>George P Lomonosoff and Daniel Ponndorf</i>	662
Viruses: Impact on Science and Society <i>Neeraja Sankaran and Robin A Weiss</i>	671

VOLUME 2

Viruses as Infectious Agents: Human and Animal Viruses

Adenoviruses (<i>Adenoviridae</i>) <i>Balázs Harrach and Mária Benkő</i>	3
---	---

African Horse Sickness Virus (<i>Reoviridae</i>) <i>Piet A van Rijn</i>	17
African Swine Fever Virus (<i>Asfarviridae</i>) <i>Linda K Dixon, Rachel Nash, Philippa C Hawes, and Christopher L Netherton</i>	22
Akabane Virus and Schmallenberg Virus (<i>Peribunyaviridae</i>) <i>Martin Beer and Kerstin Wernike</i>	34
Alphaviruses Causing Encephalitis (<i>Togaviridae</i>) <i>Diane E Griffin</i>	40
Anelloviruses (<i>Anelloviridae</i>) <i>Fabrizio Maggi and Mauro Pistello</i>	48
Animal Lentiviruses (<i>Retroviridae</i>) <i>Esperanza Gomez-Lucia</i>	56
Animal Morbilliviruses (<i>Paramyxoviridae</i>) <i>Carina Conceicao and Dalan Bailey</i>	68
Animal Papillomaviruses (<i>Papillomaviridae</i>) <i>John S Munday</i>	79
Astroviruses (<i>Astroviridae</i>) <i>Virginia Hargest, Amy Davis, and Stacey Schultz-Cherry</i>	92
Avian Hepadnaviruses (<i>Hepadnaviridae</i>) <i>Allison R Jilbert, Georget Y Reaiche-Miller, and Catherine A Scougall</i>	100
Avian Herpesviruses (<i>Herpesviridae</i>) <i>Vishwanatha RAP Reddy and Venugopal Nair</i>	112
Avian Influenza Viruses (<i>Orthomyxoviridae</i>) <i>Nicolas Bravo-Vasquez and Stacey Schultz-Cherry</i>	117
Avian Leukosis and Sarcoma Viruses (<i>Retroviridae</i>) <i>Karen L Beemon</i>	122
Bluetongue Virus (<i>Reoviridae</i>) <i>Raghavendran Kulasegaran-Shylini and Polly Roy</i>	127
Borna Disease Virus and Related Bornaviruses (<i>Bornaviridae</i>) <i>Susan L Payne</i>	137
Bovine Leukemia Virus (<i>Retroviridae</i>) <i>Thomas Joris, Roghaiyeh Safari, Jean-Rock Jacques, and Luc Willems</i>	144
Bovine Viral Diarrhea, Border Disease, and Classical Swine Fever Viruses (<i>Flaviviridae</i>) <i>Paul Becher, Volker Moennig, and Norbert Tautz</i>	153
Capripoxviruses, Parapoxviruses, and Other Poxviruses of Ruminants (<i>Poxviridae</i>) <i>Philippa M Beard</i>	165
Chikungunya Virus (<i>Togaviridae</i>) <i>Thomas E Morrison and Stephanie E Ander</i>	173
Circoviruses (<i>Circoviridae</i>) <i>Giovanni Franzo and Joaquim Segalés</i>	182
Coronaviruses: General Features (<i>Coronaviridae</i>) <i>Paul Britton</i>	193
Coronaviruses: Molecular Biology (<i>Coronaviridae</i>) <i>X Deng and SC Baker</i>	198

Crimean-Congo Hemorrhagic Fever Virus and Nairoviruses of Medical Importance (<i>Nairoviridae</i>) <i>Ali Mirazimi, Felicity Burt, and Anna Papa</i>	208
Dengue Viruses (<i>Flaviviridae</i>) <i>Ashley L. St. John and Duane J. Gubler</i>	218
Ebola Virus (<i>Filoviridae</i>) <i>Andrea Marzi and Logan Banadyga</i>	232
Enteroviruses (<i>Picornaviridae</i>) <i>Carita Savolainen-Kopra, Soile Blomqvist, and Petri Susi</i>	245
Enveloped, Positive-Strand RNA Viruses (<i>Nidovirales</i>) <i>L. Enjuanes, AE Gorbalenya, RJ de Groot, JA Cowley, J Ziebuhr, and EJ Snijder</i>	256
Epstein-Barr Virus (<i>Herpesviridae</i>) <i>Lawrence S. Young</i>	267
Equine Herpesviruses (<i>Herpesviridae</i>) <i>Gisela Soboll-Hussey, Nikolaus Osterrieder, and Walid Azab</i>	278
Equine, Canine, and Swine Influenza (<i>Orthomyxoviridae</i>) <i>Janet M. Daly and Japhette E. Kembou-Ringert</i>	287
Feline Calicivirus (<i>Caliciviridae</i>) <i>Margaret J. Hosie and Michaela J. Conley</i>	294
Feline Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>Brian J. Willett and Margaret J. Hosie</i>	300
Fish and Amphibian Alloherpesviruses (<i>Herpesviridae</i>) <i>Maxime Boutier, Léa Morvan, Natacha Delrez, Francesco Origi, Andor Doszpoly, and Alain Vanderplasschen</i>	306
Fish Retroviruses (<i>Retroviridae</i>) <i>TA Paul, RN Casey, PR Bowser, JW Casey, J Rovnak, and SL Quackenbush</i>	316
Fish Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Gael Kurath and David Stone</i>	324
Foot-and-Mouth Disease Viruses (<i>Picornaviridae</i>) <i>David J. Rowlands</i>	332
Fowlpox Virus and Other Avipoxviruses (<i>Poxviridae</i>) <i>Efstathios S. Giotis and Michael A. Skinner</i>	343
Hantaviruses (<i>Hantaviridae</i>) <i>Tarja Sironen and Antti Vaheri</i>	349
Henipaviruses (<i>Paramyxoviridae</i>) <i>Lin-Fa Wang and Danielle E. Anderson</i>	355
Hepatitis A Virus (<i>Picornaviridae</i>) <i>Andreas Dotzauer</i>	362
Hepatitis B Virus (<i>Hepadnaviridae</i>) <i>Peter Karayiannis</i>	373
Hepatitis C Virus (<i>Flaviviridae</i>) <i>Ralf Bartenschlager and Keisuke Tabata</i>	386
Hepeviruses (<i>Hepeviridae</i>) <i>Xiang-Jin Meng</i>	397
Herpes Simplex Virus 1 and 2 (<i>Herpesviridae</i>) <i>David M. Knipe and Richard Whitley</i>	404

History of Virology: Vertebrate Viruses <i>F Fenner</i>	414
Human Boca- and Protoparvoviruses (<i>Parvoviridae</i>) <i>Maria Söderlund-Venermo and Jianming Qiu</i>	419
Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (<i>Coronaviridae</i>) <i>Ding X Liu, Jia Q Liang, and To S Fung</i>	428
Human Cytomegalovirus (<i>Herpesviridae</i>) <i>Edward S Mocarski</i>	441
Human Immunodeficiency Virus (<i>Retroviridae</i>) <i>Blaide Woodburn, Ann Emery, and Ronald Swanstrom</i>	460
Human Metapneumovirus (<i>Pneumoviridae</i>) <i>Antonella Casola, Matteo P Garofalo, and Xiaoyong Bao</i>	475
Human Norovirus and Sapovirus (<i>Caliciviridae</i>) <i>Sumit Sharma, Marie Hagbom, Lennart Svensson, and Johan Nordgren</i>	483
Human Papillomaviruses (<i>Papillomaviridae</i>) <i>Alison A McBride and Samuel S Porter</i>	493
Human Parainfluenza Viruses (<i>Paramyxoviridae</i>) <i>Elisabeth Adderson</i>	502
Human Pathogenic Arenaviruses (<i>Arenaviridae</i>) <i>Sheli R Radoshitzky and Juan C de la Torre</i>	507
Human Polyomaviruses (<i>Papillomaviridae</i>) <i>Melissa S Maginnis</i>	518
Human T-Cell Leukemia Virus-1 and -2 (<i>Retroviridae</i>) <i>Amanda R Panfil and Patrick L Green</i>	528
Infectious Bursal Disease Virus (<i>Birnaviridae</i>) <i>Daral J Jackwood</i>	540
Infectious Pancreatic Necrosis Virus (<i>Birnaviridae</i>) <i>Øystein Evensen</i>	544
Influenza A Viruses (<i>Orthomyxoviridae</i>) <i>Laura Kakkola, Niina Ikonen, and Ilkka Julkunen</i>	551
Influenza B, C and D Viruses (<i>Orthomyxoviridae</i>) <i>Thorsten Wolff and Michael Veit</i>	561
Jaagsiekte Sheep Retrovirus (<i>Retroviridae</i>) <i>James M Sharp, Marcelo De las Heras, Massimo Palmarini, and Thomas E Spencer</i>	575
Japanese Encephalitis Virus (<i>Flaviviridae</i>) <i>Sang-Im Yun and Young-Min Lee</i>	583
Kaposi's Sarcoma-Associated Herpesvirus (<i>Herpesviridae</i>) <i>Anne K Cordes and Thomas F Schulz</i>	598
Marburg and Ravn Viruses (<i>Filoviridae</i>) <i>Courtney Woolsey, Thomas W Geisbert, and Robert W Cross</i>	608
Measles Virus (<i>Paramyxoviridae</i>) <i>Roberto Cattaneo and Michael McChesney</i>	619
Molluscum Contagiosum Virus (<i>Poxviridae</i>) <i>Joachim J Bugert and Rosina Ehmann</i>	629

Mumps Virus (<i>Paramyxoviridae</i>) <i>Stephen A Winchester and Kevin E Brown</i>	634
Murine Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>George R Young and Kate N Bishop</i>	643
Newcastle Disease Virus (<i>Paramyxoviridae</i>) <i>Ben Peeters and Guus Koch</i>	648
Orthobunyaviruses (<i>Peribunyaviridae</i>) <i>Alyssa B Evans, Clayton W Winkler, and Karin E Peterson</i>	654
Parapoxviruses (<i>Poxviridae</i>) <i>Hanns-Joachim Rziha and Mathias Büttner</i>	666
Parechoviruses (<i>Picornaviridae</i>) <i>Sisko Tauriainen and Glyn Stanway</i>	675
Parvoviruses of Carnivores, and the Emergence of Canine Parvovirus (<i>Parvoviridae</i>) <i>Colin R Parrish, Ian EH Voorhees, and Susan L Hafenstein</i>	683
Polioviruses (<i>Picornaviridae</i>) <i>Philip D Minor</i>	688
Porcine Reproductive and Respiratory Syndrome Virus and Equine Arteritis Virus (<i>Arteriviridae</i>) <i>Jianqiang Zhang, Alan T Loynachan, Gregory W Stevenson, and Jeffrey J Zimmerman</i>	697
Prions of Vertebrates <i>Jonathan DF Wadsworth and John Collinge</i>	707
Pseudorabies Virus (<i>Herpesviridae</i>) <i>Thomas C Mettenleiter and Barbara G Klupp</i>	714
Rabbit Hemorrhagic Disease Virus and European Brown Hare Syndrome Virus (<i>Caliciviridae</i>) <i>Lorenzo Capucci, Patrizia Cavadini, and Antonio Lavazza</i>	724
Rabbit Myxoma Virus and the Fibroma Viruses (<i>Poxviridae</i>) <i>Peter J Kerr</i>	730
Rabies and Other Lyssaviruses (<i>Rhabdoviridae</i>) <i>Ashley C Banyard and Anthony R Fooks</i>	738
Respiratory Syncytial Virus (<i>Pneumoviridae</i>) <i>Tiffany King, Tiffany Jenkins, Supranee Chaiwatpongsakorn, and Mark E Peebles</i>	747
Rhinoviruses (<i>Picornaviridae</i>) <i>Matti Waris and Olli Ruuskanen</i>	757
Rift Valley Fever Virus and Other Phleboviruses (<i>Phenuiviridae</i>) <i>Tetsuro Ikegami</i>	765
Roseoloviruses: Human Herpesviruses 6A, 6B and 7 (<i>Herpesviridae</i>) <i>Katherine N Ward</i>	778
Rotaviruses (<i>Reoviridae</i>) <i>Juana Angel and Manuel A Franco</i>	789
Rubella Virus (<i>Picornaviridae</i>) <i>Annette Mankertz</i>	797
Saint Louis Encephalitis Virus (<i>Flaviviridae</i>) <i>William K Reisen, Lark L Coffey, Daniele M Swetnam, and Aaron C Brault</i>	805
Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (<i>Coronaviridae</i>) <i>Joseph SM Peiris and Leo LM Poon</i>	814

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (<i>Coronaviridae</i>) <i>Malik Peiris</i>	825
Simian Immunodeficiency Virus (SIV) and HIV-2 (<i>Retroviridae</i>) <i>Phyllis J Kanki</i>	827
Sindbis Virus (<i>Togaviridae</i>) <i>Satu Kurkela</i>	837
Tick-Borne Encephalitis Virus (<i>Flaviviridae</i>) <i>Teemu Smura, Suvi Kuivanen, and Olli Vapalahti</i>	843
Transmissible Gastroenteritis Virus of Pigs and Porcine Epidemic Diarrhea Virus (<i>Coronaviridae</i>) <i>Qiang Liu and Volker Gerds</i>	850
Vaccinia Virus (<i>Poxviridae</i>) <i>Yan Xiang and Rebecca K Lane</i>	854
Varicella-Zoster Virus (<i>Herpesviridae</i>) <i>Jeffrey I Cohen</i>	860
Variola and Monkeypox Viruses (<i>Poxviridae</i>) <i>Lalita Priyamvada and Panayampalli S Satheshkumar</i>	868
Vesicular Stomatitis Virus and Bovine Ephemeral Fever Virus (<i>Rhabdoviridae</i>) <i>Peter J Walker and Robert B Tesh</i>	875
West Nile Virus (<i>Flaviviridae</i>) <i>Fengwei Bai and Elizabeth Ashley Thompson</i>	884
Yellow Fever Virus (<i>Flaviviridae</i>) <i>Ashley E Strother and Alan DT Barrett</i>	891
Zika Virus (<i>Flaviviridae</i>) <i>Nikos Vasilakis, Shannan L Rossi, Sasha R Azar, Irma E Cisneros, Cassia F Estofolete, and Mauricio L Nogueira</i>	899

VOLUME 3

Viruses as Infectious Agents: Plant Viruses

An Introduction to Plant Viruses <i>Roger Hull</i>	3
Emerging and Re-Emerging Plant Viruses <i>Sabrina Bertin, Francesco Faggioli, Andrea Gentili, Ariana Manglli, Anna Taglienti, Antonio Tiberini, and Laura Tomassoli</i>	8
Emerging Geminiviruses (<i>Geminiviridae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, and Muhammad Mubin</i>	21
Movement of Viruses in Plants <i>Manfred Heinlein</i>	32
Plant Antiviral Defense: Gene-Silencing Pathways <i>Vitantonio Pantaleo, Chikara Masuta, and Hanako Shimura</i>	43
Plant Resistance to Viruses: Engineered Resistance <i>Marc Fuchs</i>	52
Plant Resistance to Viruses: Natural Resistance Associated With Dominant Genes <i>Mandy Ravensbergen and Richard Kormelink</i>	60

Plant Resistance to Viruses: Natural Resistance Associated With Recessive Genes <i>Masayoshi Hashimoto, Kensaku Maejima, Yasuyuki Yamaji, and Shigetou Namba</i>	69
Plant Viral Diseases: Economic Implications <i>Basavaprabhu L Patil</i>	81
Retrotransposons of Plants <i>M-A Grandbastien</i>	98
Vector Transmission of Plant Viruses <i>Etienne Herrbach and Quentin Chesnais</i>	106
Viral Suppressors of Gene Silencing <i>Hernan Garcia-Ruiz</i>	116
Virus-Induced Gene Silencing (VIGS) <i>Xu Tengzhi, Ugrappa Nagalakshmi, and Savithramma P Dinesh-Kumar</i>	123
Alfalfa Mosaic Virus (<i>Bromoviridae</i>) <i>L Sue Loesch-Fries</i>	132
Alphaflexiviruses (<i>Alphaflexiviridae</i>) <i>Sergey Y Morozov and Alexey A Agranovsky</i>	140
Alphasatellites (<i>Alphasatellitidae</i>) <i>Rob W Briddon and Muhammad S Nawaz-ul-Rehman</i>	149
Amalgaviruses (<i>Amalgaviridae</i>) <i>Ioannis E Tzanetakis, Sead Sabanadzovic, and Rodrigo A Valverde</i>	154
Badnaviruses (<i>Caulimoviridae</i>) <i>Andrew DW Geering</i>	158
Banana Bunchy Top Virus (<i>Nanoviridae</i>) <i>John E Thomas</i>	169
Barley Yellow Dwarf Viruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	176
Bean Common Mosaic Virus and Bean Common Mosaic Necrosis Virus (<i>Potyviridae</i>) <i>Ramon Jordan and John Hammond</i>	184
Bean Golden Mosaic Virus and Bean Golden Yellow Mosaic Virus (<i>Geminiviridae</i>) <i>Francisco M Zerbini and Simone G Ribeiro</i>	192
Beet Curly Top Virus (<i>Geminiviridae</i>) <i>Robert L Gilbertson, Tomas A Melgarejo, Maria R Rojas, William M Wintermantel, and John Stanley</i>	200
Beet Necrotic Yellow Vein Virus (<i>Benyviridae</i>) <i>Sebastian Liebe, Annette Niehl, Renate Koenig, and Mark Varrelmann</i>	213
Benyviruses (<i>Benyviridae</i>) <i>Annette Niehl, Sebastian Liebe, Mark Varrelmann, and Renate Koenig</i>	219
Betaflexiviruses (<i>Betaflexiviridae</i>) <i>Nobuyuki Yoshikawa and Hajime Yaegashi</i>	229
Betasatellites and Deltasatellites (<i>Toleusatellitidae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, Muhammad Hassan, and Muhammad Mubin</i>	239
Bluner-, Cile-, and Higreviruses (<i>Kitaviridae</i>) <i>Diego F Quito-Avila, Juliana Freitas-Astúa, and Michael J Melzer</i>	247
Brome Mosaic Virus (<i>Bromoviridae</i>) <i>Guijuan He, Zhenlu Zhang, Preethi Sathanantham, Arturo Diaz, and Xiaofeng Wang</i>	252

Bromoviruses (<i>Bromoviridae</i>) <i>Jozef J Bujarski</i>	260
Bymoviruses (<i>Potyviridae</i>) <i>Annette Niehl and Frank Rabenstein</i>	268
Cacao Swollen Shoot Virus (<i>Caulimoviridae</i>) <i>Emmanuelle Muller</i>	274
Carmo-Like Viruses (<i>Tombusviridae</i>) <i>Miryam Pérez-Cañamás and Carmen Hernández</i>	285
Cassava Brown Streak Viruses (<i>Potyviridae</i>) <i>Basavaprabhu L Patil</i>	293
Cassava Mosaic Viruses (<i>Geminiviridae</i>) <i>James Legg and Stephan Winter</i>	301
Caulimoviruses (<i>Caulimoviridae</i>) <i>James E Schoelz and Mustafa Adhab</i>	313
Cheraviruses, Sadwaviruses and Torradoviruses (<i>Secoviridae</i>) <i>Toru Iwanami and René AA van der Vlugt</i>	322
Citrus Tristeza Virus (<i>Closteroviridae</i>) <i>Moshe Bar-Joseph, Scott J Harper, and William O Dawson</i>	327
Closteroviruses (<i>Closteroviridae</i>) <i>Marc Fuchs</i>	336
Comoviruses and Fabaviruses (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	348
Cotton Leaf Curl Disease (<i>Geminiviridae</i>) <i>Nasim Ahmed, Imran Amin, and Shahid Mansoor</i>	355
Cowpea Mosaic Virus (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	364
Cucumber Mosaic Virus (<i>Bromoviridae</i>) <i>Judith Hirsch and Benoît Moury</i>	371
Dianthovirus (<i>Tombusviridae</i>) <i>Kiwamu Hyodo and Masanori Kaido</i>	383
Endornaviruses (<i>Endornaviridae</i>) <i>Toshiyuki Fukuhara</i>	388
Fimoviruses (<i>Fimoviridae</i>) <i>Toufic Elbeaino and Michele Digiaro</i>	396
Furoviruses (<i>Virgaviridae</i>) <i>Annette Niehl and Renate Koenig</i>	405
Geminiviruses (<i>Geminiviridae</i>) <i>Jesús Navas-Castillo and Elvira Fiallo-Olivé</i>	411
Hordeiviruses (<i>Virgaviridae</i>) <i>Zhihao Jiang, Meng Yang, Yongliang Zhang, Andrew O Jackson, and Dawei Li</i>	420
Idaeoviruses (<i>Mayoviridae</i>) <i>Robert R Martin and Karen E Keller</i>	430
Ilarviruses (<i>Bromoviridae</i>) <i>Aaron Simkovich, Susanne E Kohalmi, and Aiming Wang</i>	439

Luteoviruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	447
Machlomovirus and Panicoviruses (<i>Tombusviridae</i>) <i>Kay Scheets</i>	456
Maize Streak Virus (<i>Geminiviridae</i>) <i>Darren P Martin and Aderito L Monjane</i>	461
Nanoviruses (<i>Nanoviridae</i>) <i>Bruno Gronenborn and H Josef Vetten</i>	470
Necro-Like Viruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Giovanni P Martelli[†]</i>	481
Nepoviruses (<i>Secoviridae</i>) <i>Hélène Sanfaçon</i>	486
Ophioviruses (<i>Aspiviridae</i>) <i>Anna M Vaira and John Hammond</i>	495
Orthospoviruses (<i>Tospoviridae</i>) <i>Renato O Resende and Hanu R Pappu</i>	507
Ourmiaviruses (<i>Botourmiaviridae</i>) <i>Gian Paolo Accotto and Cristina Rosa</i>	516
Papaya Ringspot Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	520
Pecluviruses (<i>Virgaviridae</i>) <i>Hema Masarapu, Pothur Sreenivasulu, Philippe Delfosse, Claude Bragard, Anne Legreve, and DVR Reddy</i>	528
Pepino Mosaic Virus (<i>Alphaflexiviridae</i>) <i>Rene AA van der Vlugt and CCMM Stijger</i>	539
Plant Reoviruses (<i>Reoviridae</i>) <i>Yu Huang and Yi Li</i>	545
Plant Resistance to Geminiviruses <i>Basavaprabhu L Patil, Supriya Chakraborty, Henryk Czosnek, Elvira Fiallo-Olivé, Robert L Gilbertson, James Legg, Shahid Mansoor, Jesús Navas-Castillo, Rubab Z Naqvi, Saleem U Rahman, and Francisco M Zerbini</i>	554
Plant Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Ralf G Dietzgen, Michael M Goodin, and Zhenghe Li</i>	567
Plant Satellite Viruses (<i>Albetovirus, Aumaivirus, Papanivirus, Virtovirus</i>) <i>Mart Krupovic</i>	581
Plum Pox Virus (<i>Potyviridae</i>) <i>Miroslav Glasa and Thierry Candresse</i>	586
Poleroviruses (<i>Luteoviridae</i>) <i>Hernan Garcia-Ruiz, Natalie M Holste, and Katherine LaTourrette</i>	594
Pomoviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	603
Potato Virus Y (<i>Potyviridae</i>) <i>Laurent Glais and Benoît Moury</i>	612
Potexviruses (<i>Alphaflexiviridae</i>) <i>Ki H Ryu, Eun G Song, and Jin S Hong</i>	623

[†]Deceased.

Potyviruses (<i>Potyviridae</i>) <i>Adrián Valli, Juan A García, and Juan J López-Moya</i>	631
Quinviruses (<i>Betaflexiviridae</i>) <i>Ki H Ryu and Eun G Song</i>	642
Reverse-Transcribing Viruses (<i>Belpaoviridae, Metaviridae, and Pseudoviridae</i>) <i>Carlos Llorens, Beatriz Soriano, Maria A Navarrete-Muñoz, Ahmed Hafez, Vicente Arnau, Jose Miguel Benito, Toni Gabaldon, Norma Rallon, Jaume Pérez-Sánchez, and Mart Krupovic</i>	653
Rice Tungro Disease (<i>Secoviridae, Caulimoviridae</i>) <i>Gaurav Kumar, Fauzia Zarreen, and Indranil Dasgupta</i>	667
Rice Yellow Mottle Virus (<i>Solemoviridae</i>) <i>Eugénie Hébrard, Nils Poulicard, and Mbolarinosy Rakotomalala</i>	675
Satellite Nucleic Acids and Viruses <i>Olufemi J Alabi, Alfredo Diaz-Lara, and Maher Al Rwahnih</i>	681
Secoviruses (<i>Secoviridae</i>) <i>Jeremy R Thompson</i>	692
Sequiviruses and Waikaviruses (<i>Secoviridae</i>) <i>Lucy Rae Stewart</i>	703
Solemoviruses (<i>Solemoviridae</i>) <i>Cecilia Sarmiento, Merike Sõmera, and Erkki Truve[†]</i>	712
Tenuiviruses (<i>Phenuiviridae</i>) <i>Bertha Cecilia Ramirez and Anne-Lise Haenni</i>	719
Tobacco Mosaic Virus (<i>Virgaviridae</i>) <i>Marc HV Van Regenmortel</i>	727
Tobamoviruses (<i>Virgaviridae</i>) <i>Ulrich Melcher, Dennis J Lewandowski, and William O Dawson</i>	734
Tobraviruses (<i>Virgaviridae</i>) <i>Stuart A MacFarlane</i>	743
Tomato Leaf Curl New Delhi Virus (<i>Geminiviridae</i>) <i>Supriya Chakraborty and Manish Kumar</i>	749
Tomato Spotted Wilt Virus (<i>Tospoviridae</i>) <i>Hanu R Pappu, Anna E Whitfield, and Athos S de Oliveira</i>	761
Tomato Yellow Leaf Curl Viruses (<i>Geminiviridae</i>) <i>Henryk Czosnek</i>	768
Tombusvirus-Like Viruses (<i>Tombusviridae</i>) <i>K Andrew White</i>	778
Tombusviruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Kay Scheets</i>	788
Tritimoviruses and Rymoviruses (<i>Potyviridae</i>) <i>Satyanarayana Tatineni and Gary L Hein</i>	797
Triviruses (<i>Betaflexiviridae</i>) <i>Yahya ZA Gaafar and Heiko Ziebell</i>	805
Tymoviruses (<i>Tymoviridae</i>) <i>Rosemarie W Hammond and Peter Abrahamian</i>	818

[†]Deceased.

Umbraviruses (<i>Tombusviridae</i>) <i>Eugene V Ryabov and Michael E Taliansky</i>	827
Varicosaviruses (<i>Rhabdoviridae</i>) <i>Takahide Sasaya</i>	833
Virgaviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	839
Viroids (<i>Pospiviroidae</i> and <i>Avsunviroidae</i>) <i>Ricardo Flores and Robert A Owens</i>	852
Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	862

VOLUME 4

Viruses as Infectious Agents: Bacterial, Archaeal, Fungal, Algal, and Invertebrate Viruses

Bacterial Viruses

History of Virology: Bacteriophages <i>William C Summers</i>	3
Icosahedral Phages – Single-Stranded DNA (ϕ X174) <i>Bentley A Fane and Aaron P Roznowski</i>	10
Single-Stranded RNA Bacterial Viruses <i>Peter G Stockley and Junjie Zhang</i>	21
Enveloped Icosahedral Phages – Double-Stranded RNA (ϕ 6) <i>Paul Gottlieb and Aleksandra Alimova</i>	26
Membrane-Containing Icosahedral DNA Bacteriophages <i>Roman Tuma, Sarah J Butcher, and Hanna M Oksanen</i>	36
Tailed Double-Stranded DNA Phages <i>Robert L Duda</i>	45
Helical and Filamentous Phages <i>Andreas Kuhn and Sebastian Leptihn</i>	53
Replication of <i>Bacillus</i> Double-Stranded DNA Bacteriophages <i>Silvia Ayora, Paulo Tavares, Ruben Torres, and Juan C Alonso</i>	61
Lytic Transcription <i>William McAllister and Deborah M Hinton</i>	69
Lysogeny <i>Keith E Shearwin and Jia Q Truong</i>	77
Decision Making by Temperate Phages <i>Ido Golding, Seth Coleman, Thu VP Nguyen, and Tianyou Yao</i>	88
Mobilization of Phage Satellites <i>Kristen N LeGault and Kimberley D Seed</i>	98
Portal Vertex <i>Peng Jing and Mauricio Cortes Jr.</i>	105

Prohead, the Head Shell Pre-Cursor <i>Marc C Morais and Michael E Woodson</i>	115
Enzymology of Viral DNA Packaging Machines <i>Carlos E Catalano</i>	124
DNA Packaging: DNA Recognition <i>Sandra J Greive and Oliver W Bayfield</i>	136
DNA Packaging: The Translocation Motor <i>Janelle A Hayes and Brian A Kelch</i>	148
Biophysics of DNA Packaging <i>Joshua Pajak, Gaurav Arya, and Douglas E Smith</i>	160
Energetics of the DNA-Filled Head <i>Alex Evilevitch</i>	167
Bacteriophage Receptor Proteins of Gram-Negative Bacteria <i>Sarah M Doore, Kristin N Parent, Sundharraman Subramanian, Jason R Schrad, and Natalia B Hubbs</i>	175
Tail Structure and Dynamics <i>Shweta Bhatt, Petr G Leiman, and Nicholas MI Taylor</i>	186
Bacteriophage Tail Fibres, Tailspikes, and Bacterial Receptor Interaction <i>Mateo Seoane-Blanco, Mark J van Raaij, and Meritxell Granell</i>	194
Phage Genome and Protein Ejection In Vivo <i>Ian J Molineux, L Letti Lopez, and Aaron P Roznowski</i>	206
Dealing With the Whole Head: Diversity and Function of Capsid Ejection Proteins in Tailed Phages <i>Lindsay W Black and Julie A Thomas</i>	219
Jumbo Phages <i>Isaac T Younker and Carol Duffy</i>	229
CRISPR-Cas Systems and Anti-CRISPR Proteins: Adaptive Defense and Counter-Defense in Prokaryotes and Their Viruses <i>Asma Hatoum-Aslan and Olivia G Howell</i>	242
Bacteriophage: Therapeutics and Diagnostics Development <i>Teng-Chieh Yang</i>	252
Bacteriophage Vaccines <i>Pan Tao and Venigalla B Rao</i>	259
Bacteriophage Diversity <i>Julianne H Grose and Sherwood R Casjens</i>	265
Genetic Mosaicism in the Tailed Double-Stranded DNA Phages <i>Welkin H Pope</i>	276
Bacteriophages of the Human Microbiome <i>Pilar Manrique, Michael Dills, and Mark J Young</i>	283
Bacteriophage: Red Recombination System and the Development of Recombineering Technologies <i>Lynn C Thomason and Kenan C Murphy</i>	291
Nanotechnology Application of Bacteriophage DNA Packaging Nanomotors <i>Tao Weitao, Lixia Zhou, Zhefeng Li, Long Zhang, and Peixuan Guo</i>	302
General Ecology of Bacteriophages <i>Stephen T Abedon</i>	314
Marine Bacteriophages <i>Vera Bischoff, Falk Zucker, and Cristina Moraru</i>	322

Ecology of Phages in Extreme Environments <i>Tatiana A Demina and Nina S Atanasova</i>	342
---	-----

Archaeal Viruses

Diversity of Hyperthermophilic Archaeal Viruses <i>David Prangishvili, Mart Krupovic, and Diana P Baquero</i>	359
Euryarchaeal Viruses <i>Tatiana A Demina and Hanna M Oksanen</i>	368
Vesicle-Like Archaeal Viruses <i>Elina Roine and Nina S Atanasova</i>	380
Virus–Host Interactions in Archaea <i>Diana P Baquero, David Prangishvili, and Mart Krupovic</i>	387
Antiviral Defense Mechanisms in Archaea <i>Qunxin She</i>	400
Discovery of Archaeal Viruses in Hot Spring Environments Using Viral Metagenomics <i>Jennifer Wirth, Jacob H Munson-McGee, and Mark J Young</i>	407
Metagenomes of Archaeal Viruses in Hypersaline Environments <i>Fernando Santos, María D Ramos-Barbero, and Josefa Antón</i>	414
Extreme Environments as a Model System to Study How Virus–Host Interactions Evolve Along the Symbiosis Continuum <i>Samantha J DeWerff and Rachel J Whitaker</i>	419

Fungal Viruses

An Introduction to Fungal Viruses <i>Nobuhiro Suzuki</i>	431
Cross-Kingdom Virus Infection <i>Liyang Sun, Hideki Kondo, and Ida Bagus Andika</i>	443
Diversity of Mycoviruses in Aspergilli <i>Ioly Kotta-Loizou</i>	450
Evolution of Mycoviruses <i>Mahtab Peyambari, Vaskar Thapa, and Marilyn J Roossinck</i>	457
Mixed Infections of Mycoviruses in Phytopathogenic Fungus <i>Sclerotinia sclerotiorum</i> <i>Jiatao Xie and Daohong Jiang</i>	461
Mycovirus-Mediated Biological Control <i>Daniel Rigling, Cécile Robin, and Simone Prospero</i>	468
Mycoviruses With Filamentous Particles <i>Michael N Pearson</i>	478
Prions of Yeast and Fungi <i>Reed B Wickner and Herman K Edskes</i>	487
Single-Stranded DNA Mycoviruses <i>Daohong Jiang</i>	493
Structure of Double-Stranded RNA Mycoviruses <i>José R Castón, Nobuhiro Suzuki, and Said A Ghabrial[†]</i>	504

[†]Deceased.

<i>Ustilago maydis</i> Viruses and Their Killer Toxins <i>Alexis Williams and Thomas J Smith</i>	513
Vegetative Incompatibility in Filamentous Fungi <i>Songsong Wu, Daohong Jiang, and Jiatao Xie</i>	520
Viral Diseases of <i>Agaricus bisporus</i> , the Button Mushroom <i>Kerry S Burton and Greg Deakin</i>	528
Viral Killer Toxins <i>Manfred J Schmitt and Björn Becker</i>	534
Alternaviruses (Unassigned) <i>Hiromitsu Moriyama, Nanako Aoki, Kuko Fuke, Kana Takeshita Urayama, Naoki Takeshita, and Chien-Fu Wu</i>	544
Barnaviruses (Barnaviridae) <i>Peter A Revill</i>	549
Botybirnaviruses (Botybirnavirus) <i>Mingde Wu, Guoqing Li, Daohong Jiang, and Jiatao Xie</i>	552
Chrysoviruses (Chrysoviridae) - General Features and Chrysovirus-Related Viruses <i>Ioly Kotta-Loizou, Robert HA Coutts, José R Castón, Hiromitsu Moriyama, and Said A Ghabrial[†]</i>	557
Fungal Partitiviruses (Partitiviridae) <i>Eeva J Vainio</i>	568
Fusariviruses (Unassigned) <i>Sotaro Chiba</i>	577
Giardiavirus (Totiviridae) <i>Juliana Gabriela Silva de Lima, João Paulo Matos Santos Lima, and Daniel Carlos Ferreira Lanza</i>	582
Hypoviruses (Hypoviridae) <i>Dong-Xiu Zhang and Donald L Nuss</i>	589
Megabirnaviruses (Megabirnaviridae) <i>Yukiyo Sato and Nobuhiro Suzuki</i>	594
Mitoviruses (Mitoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	601
Mycoreoviruses (Reoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	607
Mymonaviruses (Mymonaviridae) <i>Daohong Jiang</i>	615
Narnaviruses (Narnaviridae) <i>Rosa Esteban and Tsutomu Fujimura</i>	621
Phlegiviruses (Unassigned) <i>Karel Petrzik</i>	627
Plant and Protozoal Partitiviruses (Partitiviridae) <i>Hanna Rose and Edgar Maiss</i>	632
Quadriviruses (Quadriviridae) <i>Hideki Kondo, José R Castón, and Nobuhiro Suzuki</i>	642
Totiviruses (Totiviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	648

[†]Deceased.

Yado-kari Virus 1 and Yado-nushi Virus 1 (Unassigned) <i>Subha Das and Nobuhiro Suzuki</i>	658
Yeast L-A Virus (<i>Totiviridae</i>) <i>Reed B Wickner, Tsutomu Fujimura, and Rosa Esteban</i>	664
Algal Viruses	
Algal Marnaviruses (<i>Marnaviridae</i>) <i>Marli Vlok, Curtis A Suttle, and Andrew S Lang</i>	671
Algal Mimiviruses (<i>Mimiviridae</i>) <i>Ruth-Anne Sandaa, Håkon Dahle, Corina PD Brussaard, Hiroyuki Ogata, and Romain Blanc-Mathieu</i>	677
Miscellaneous Algal Viruses (<i>Alvernaviridae, Bacilladnaviridae, Dinodnavirus, Reoviridae</i>) <i>Keizo Nagasaki, Yuji Tomaru, and Corina PD Brussaard</i>	684
Phycodnaviruses (<i>Phycodnaviridae</i>) <i>James L Van Etten, David D Dunigan, Keizo Nagasaki, Declan C Schroeder, Nigel Grimsley, Corina PD Brussaard, and Jozef I Nissimov</i>	687
Invertebrate Viruses	
An Introduction to Viruses of Invertebrates <i>Peter Krell</i>	699
Ascoviruses (<i>Ascoviridae</i>) <i>Sassan Asgari, Dennis K Bideshi, Yves Bigot, and Brian A Federici</i>	724
Baculovirus–Host Interactions: Repurposing Host-Acquired Genes (<i>Baculoviridae</i>) <i>A Lorena Passarelli</i>	732
Baculoviruses: General Features (<i>Baculoviridae</i>) <i>Vera ID Ros</i>	739
Baculoviruses: Molecular Biology and Replication (<i>Baculoviridae</i>) <i>Monique M van Oers</i>	747
Bidensovirus (<i>Bidnaviridae</i>) <i>Qin Yao, Zhaoyang Hu, and Keping Chen</i>	759
Bunyaviruses of Arthropods (<i>Myxoviridae, Nairoviridae, Peribunyaviridae, Phasmaviridae, Phuniviridae, Wupedeviridae</i>) <i>Sandra Junglen</i>	764
Dicistroviruses (<i>Dicistroviridae</i>) <i>Yanping Chen and Steven M Valles</i>	768
Entomobirnaviruses (<i>Birnaviridae</i>) <i>Marco Marklewitz</i>	776
Hytroviruses (<i>Hytroviridae</i>) <i>Henry M Kariithi and Irene K Meki</i>	780
Iflaviruses (<i>Iflaviridae</i>) <i>Bryony C Bonning and Sijun Liu</i>	792
Iridoviruses of Invertebrates (<i>Iridoviridae</i>) <i>İkbal Agah İnce</i>	797
Mesoniviruses (<i>Mesoniviridae</i>) <i>Jody Hobson-Peters and Daniel Watterson</i>	804

Nimaviruses (<i>Nimaviridae</i>) <i>Peter Krell and Emine Ozsahin</i>	808
Nodaviruses of Invertebrates and Fish (<i>Nodaviridae</i>) <i>Kyle L Johnson and Jacen S Moore</i>	819
Nudiviruses (<i>Nudiviridae</i>) <i>Yu-Chan Chao, Chih-Hsuan Tsai, and Sung-Chan Wei</i>	827
Parvoviruses of Invertebrates (<i>Parvoviridae</i>) <i>Judit J Péntzes, Hanh T Pham, Qian Yu, Max Bergoin, and Peter Tijssen</i>	835
Polydnaviruses (<i>Polydnaviridae</i>) <i>Anne-Nathalie Volkoff and Elisabeth Huguet</i>	849
Poxviruses of Insects (<i>Poxviridae</i>) <i>Basil Arif, Lillian Pavlik, Remziye Nalçacıoğlu, Hacer Muratoğlu, Cihan İnan, Mehtap Yakupoğlu, Emine Özşahin, İsmail Demir, Kazım Sezen, and Zihni Demirbağ</i>	858
Reoviruses of Invertebrates (<i>Reoviridae</i>) <i>Peter Krell</i>	867
Rhabdoviruses of Insects (<i>Rhabdoviridae</i>) <i>Andrea González-González, Nicole T de Stefano, David A Rosenbaum, and Marta L Wayne</i>	883
Sarothroviruses (<i>Sarothroviridae</i>) <i>Azeez Sait Sahul Hameed</i>	888
Solinviviruses (<i>Solinviviridae</i>) <i>Steven M Valles and Andrew E Firth</i>	892
Tetraviruses (<i>Alphatetraviridae, Carmotetraviridae, Permutotetraviridae</i>) <i>Rosemary A Dorrington, Tatiana Domitrovic, and Meesbah Jiwaji</i>	897

VOLUME 5

Diagnosis, Treatment and Prevention of Virus Infections

Diagnosis

Introduction to Virus Diagnosis and Treatment <i>Maija Lappalainen and Hubert GM Niesters</i>	3
Electron Microscopy for Viral Diagnosis <i>Roland A Fleck</i>	5
Serological Approaches for Viral Diagnosis <i>Klaus Hedman and Visa Nurmi</i>	15
A Brief History of the Development of Diagnostic Molecular-Based Assays <i>Hubert GM Niesters</i>	22
Sequencing Strategies <i>Sibnarayan Datta</i>	27
Validating Real-Time Polymerase Chain Reaction (PCR) Assays <i>Melvyn Smith</i>	35
Rapid Point-of-Care Assays <i>Jan G Lisby and Uffe V Schenider</i>	45

Standardization of Diagnostic Assays <i>Sally A Baylis, C Micha Nübling, and Wayne Dimech</i>	52
Quality Assurance in the Clinical Virology Laboratory <i>Paul Wallace and Elaine McCulloch</i>	64
Biosafety and Biosecurity in Diagnostic Laboratories <i>Hannimari Kallio-Kokko and Susanna Sissonen</i>	82
Screening for Viral Infections <i>Walter Ian Lipkin, Nischay Mishra, and Thomas Briebe</i>	91
Clinical Diagnostic Virology <i>Marcus Panning</i>	98
Virus Diagnosis in Immunosuppressed Individuals <i>Elisabeth Puchhammer-Stöckl and Fausto Baldanti</i>	105
Diagnosis; Future Prospects on Direct Diagnosis <i>Marianna Calabretto, Daniele Di Carlo, Fabrizio Maggi, and Guido Antonelli</i>	112

Treatment

Antiviral Classification <i>Guangdi Li, Xixi Jing, Pan Zhang, and Erik De Clercq</i>	121
Antiretroviral Therapy – Nucleoside/Nucleotide and Non-Nucleoside Reverse Transcriptase Inhibitors <i>Timothy D Appleby and Killian J Quinn</i>	131
Protease Inhibitors <i>Vanesa Anton-Vazquez and Frank A Post</i>	139
HIV Integrase Inhibitors and Entry Inhibitors <i>Daniel Bradshaw and Ranjababu Kulasegaram</i>	145
Management of Respiratory Syncytial Virus Infections (<i>Pneumoviridae</i>) <i>Rachael S Barr and Simon B Drysdale</i>	155
Management of Influenza Virus Infections (<i>Orthomyxoviridae</i>) <i>Bruno Lina</i>	160
Management of Herpes Simplex Virus Infections (<i>Herpesviridae</i>) <i>Nicole Samies and Richard Whitley</i>	175
Management of Varicella-Zoster Virus Infections (<i>Herpesviridae</i>) <i>Andreas Sauerbrei</i>	181
Treatment and Prevention of Herpesvirus Infections in the Immunocompromised Host <i>Sara H Burkhard and Nicolas J Mueller</i>	190
Management of Adenovirus Infections (<i>Adenoviridae</i>) <i>Albert Heim</i>	197
Management of Hepatitis A and E Virus Infection <i>Sébastien Lhomme, Florence Abravanel, Jean-Marie Peron, Nassim Kamar, and Jacques Izopet</i>	206
Management of Patients With Chronic Hepatitis B (<i>Hepadnaviridae</i>) and Chronic Hepatitis D Infection (<i>Deltavirus</i>) <i>Milan J Sonneveld and Suzanne van Meer</i>	217
Studying Population Genetic Processes in Viruses: From Drug-Resistance Evolution to Patient Infection Dynamics <i>Jeffrey D Jensen</i>	227

Virus-Based Cancer Therapeutics <i>Roberto Cattaneo and Christine E Engeland</i>	233
Prevention	
Surveillance of Infectious Diseases <i>Norman Noah</i>	247
Preparing for Emerging Zoonotic Viruses <i>Reina S Sikkema and Marion PG Koopmans</i>	256
Use of Immunoglobulins in the Prevention of Viral Infections <i>Leyla Asadi and Giovanni Ferrara</i>	267
Vaccine Production, Safety, and Efficacy <i>Thomas J Brouwers and Bernard AM Van der Zeijst</i>	281
Vaccines Against Viral Gastroenteritis <i>Scott Grytdal, Tyler P Chavers, Claire P Mattison, Jacqueline E Tate, and Aron J Hall</i>	289
Human Papillomavirus (HPV) Vaccines and Their Impact <i>Jade Pattyn, Pierre Van Damme, and Alex Vorsters</i>	295
Influenza Vaccination <i>Topi Turunen</i>	300
Polio Eradication <i>M Steven Oberste, Cara C Burns, and Jennifer L Konopka-Anstadt</i>	310
Subject Index	315

PLANT VIRUSES

An Introduction to Plant Viruses

Roger Hull, John Innes Centre, Norwich, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

This is a reproduction of R. Hull, History of Virology: Plant Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00589-6.

Introduction

History shows that the study of virus infections of plants has led the overall subject of virology in the development of several major concepts including that of the entity of viruses themselves. To obtain a historical perspective of plant virology, five major (overlapping) ages can be recognized.

Prehistory

The earliest known written record describing what was almost certainly a plant virus disease is a poem in Japanese written by the Empress Koken in AD 752 and translated by T. Inouye as follows:

In this village
It looks as if frosting continuously
For, the plant I saw
In the field of summer
The color of the leaves were yellowing

The plant, identified as *Eupatorium lindleyanum*, has been found to be susceptible to tomato yellow leaf curl virus, which causes a yellowing disease.

In Western Europe in the period from about 1600 to 1660, many paintings and drawings were made of tulips that demonstrate flower symptoms of virus disease. These are recorded in the Herbals of the time and in the still-life paintings of artists such as Johannes Bosschaert in 1610. During this period, blooms featuring such striped patterns were prized as special varieties leading to the phenomenon of 'tulipomania'. The trade in infected tulip bulbs resulted in hyperinflation with bulbs exchanging hands for large amounts of money or goods (Table 1).

In describing an experiment to demonstrate that sap flows in plants, Lawrence reported in 1714 the unintentional transmission of a virus disease of jasmine by grafting. The following quotation from Blair in 1719 describes the procedure and demonstrates that even in this protoscientific stage, experimenters were already indulging in arguments about priorities of discovery.

The inoculating of a strip'd Bud into a plain stock and the consequence that the Stripe or Variegation shall be seen in a few years after, all over the shrub above and below the graft, is a full demonstration of this Circulation of the Sap. This was first observed by Mr. Wats at Kensington, about 18 years ago: Mr. Fairchild performed it 9 years ago; Mr. Bradley says he observ'd it several years since; though Mr. Lawrence would insinuate as if he had first discovered it.

Recognition of Viral Entity

In the latter part of the nineteenth century, the idea that infectious disease was caused by microbes was well established, and filters were available that would not allow the passage of known bacterial pathogens. Mayer in 1886 showed that a disease of tobacco (*Mosaikkrankheit*; now known to be caused by tobacco mosaic virus; TMV) could be transmitted to healthy plants by inoculation with extracts from diseased plants; Iwanowski demonstrated in 1892 that sap from such tobacco plants was still infective after it had been passed through a bacteria-proof filter candle. This work did not attract much attention until it was repeated by Beijerinck who in 1898 described the infectious agent as *contagium vivum fluidum* (Latin for contagious living fluid) to distinguish it from contagious corpuscular agents. Beijerinck's discovery is considered to be the birth of virology. In 1904 Baur showed that an infectious variegation of *Abutilon* could be transmitted by grafting, but not by mechanical inoculation. Beijerinck and Baur used the term 'virus' in describing the causative agents of these diseases, to contrast them with bacteria; this term had been used as more or less synonymous with bacteria by earlier workers. As more diseases of this sort were discovered, the unknown causative agents came to be called 'filterable viruses'.

Between 1900 and 1935, many plant diseases thought to be caused by filterable viruses were described, but considerable confusion arose because adequate methods for distinguishing one virus from another had not yet been developed.

The original criterion of a virus was an infectious entity that could pass through a filter with a pore size small enough to hold back all known cellular agents of disease. However, diseases were soon found that had virus-like symptoms not associated with any pathogen visible in the light microscope, but that could not be transmitted by mechanical inoculation. With such diseases, the

Table 1 Tulipomania: the goods exchanged for one bulb of Viceroy tulip

4 t of wheat	4 barrels of beer
8 t of rye	2 barrels of butter
4 fat oxen	1000 lb cheese
8 fat pigs	1 bed with accessories
12 fat sheep	1 full dress suit
2 hogsheads of wine	1 silver goblet

criterion of filterability could not be applied. Their infectious nature was established by graft transmission and sometimes by insect vectors. Thus, certain diseases of the yellows and witches'-broom type, such as aster yellows, came to be attributed to viruses on quite inadequate grounds. Many such diseases are now known to be caused by phytoplasma or spiroplasma, and a few by bacteria or rickettsia.

The Biological Age

During most of the period between 1900 and 1935, attention was focused on the description of diseases using the macroscopic symptoms and cytological abnormalities as revealed by light microscopy, host ranges, and methods of transmission which were the only techniques available. The influence of various physical and chemical agents on virus infectivity was investigated, but methods for the assay of infective material were primitive. Holmes showed that the local lesions produced in some hosts following mechanical inoculation could be used for the rapid quantitative assay of infective virus. This technique enabled properties of viruses to be studied much more readily and paved the way for the isolation and purification of viruses a few years later. Until about 1930, there was serious confusion by most workers regarding the diseases produced by viruses and the viruses themselves. This was not surprising, since virtually nothing was known about the viruses except that they were very small. In 1931 Smith made an important contribution that helped to clarify this situation. Working with virus diseases in potato, he realized the necessity of using plant indicators, plant species other than potato, which would react differently to different viruses present in potatoes. Using several different and novel biological methods to separate the viruses, he was able to show that many potato virus diseases were caused by a combination of two viruses with different properties, which he named virus X (potato virus X, PVX) and virus Y (potato virus Y, PVY). As PVX was not transmitted by the aphid *Myzus persicae*, whereas PVY was, PVY could be separated from PVX. He obtained PVX free of PVY by needle inoculation of the mixture to *Datura stramonium* which does not support PVY. Furthermore, Smith observed that PVX from different sources fluctuated markedly in the severity of symptoms it produced in various hosts leading to the concept of strains.

An important practical step forward was the recognition that some viruses could be transmitted from plant to plant by insects. Fukushi recorded the fact that in 1883 a Japanese rice grower transmitted what is now known to be rice dwarf virus, RDV) by the leafhopper *Recelia dorsalis*. However, this work was not published in any available form and so had little influence. In 1922 Kunkel first reported the transmission of a virus by a planthopper; within a decade, many insects were reported to be virus vectors leading to the recognition of specific virus-vector interactions.

Since Fukushi first showed in 1940 that RDV could be passed through the egg of a leafhopper vector for many generations, there has been great interest in the possibility that some viruses may be able to replicate in both plants and insects. It is now well established that plant viruses in the families *Rhabdoviridae* and *Reoviridae* and the genera *Tenuivirus*, *Tospovirus*, and *Marafivirus* multiply in insects as well as in plants.

The Biochemical/Biophysical Age

Beale's recognition in 1928 that plants infected with TMV contained a specific antigen opened the age in which the biochemical nature of viruses was elucidated. In the 1930s Gratia showed that plants infected with different viruses contained different specific antigens and Chester demonstrated that different strains of TMV and PVX could be distinguished serologically.

The high concentration at which certain viruses occur in infected plants and their relative stability turned out to be of crucial importance in the first isolation and chemical characterization of viruses, because methods for extracting and purifying proteins were not highly developed. In the early 1930s, various attempts were made to isolate and purify plant viruses using methods similar to those that had just been developed for purifying enzymes. Following detailed chemical studies suggesting that the infectious agent of TMV might be a protein, Stanley announced in 1935 the isolation of this virus in an apparently crystalline state. At first Stanley considered that the virus was a globulin containing no phosphorus but in 1936 Bawden *et al.* described the isolation from TMV-infected plants of a liquid crystalline nucleoprotein containing nucleic acid of the pentose type. They showed that the particles were rod-shaped, thus confirming the earlier suggestion of Takahashi and Rawlins based on the observation that solutions containing TMV showed anisotropy of flow. Electron microscopy and X-ray crystallography were the major techniques used in early work to explore virus structure, and the importance of these methods has continued to the present day. Bernal and

Fankuchen applying X-ray analysis to purified preparations of TMV obtained accurate estimates of the width of the rods. The isolation of other rod-shaped viruses, and spherical viruses that formed crystals, soon followed. All were shown to consist of protein and pentose nucleic acid.

Early electron micrographs confirmed that TMV was rod-shaped and provided approximate dimensions, but they were not particularly revealing because of the lack of contrast between the virus particles and the supporting membrane. The application of shadow-casting with heavy metals greatly increased the usefulness of the method for determining the overall size and shape of virus particles but not structural detail. With the development of high-resolution microscopes and of negative staining in the 1950s electron microscopy became an important tool for studying virus substructure. From a comparative study of the physicochemical properties of the virus nucleoprotein and the empty viral protein shell found in TYMV preparations, Markham concluded in 1951 that the RNA of the virus must be held inside a shell of protein, a view that has since been amply confirmed for this and other viruses by X-ray crystallography. Crick and Watson suggested that the protein coats of small viruses are made up of numerous identical subunits arrayed either as helical rods or as a spherical shell with cubic symmetry. Subsequent X-ray crystallographic and chemical work has confirmed this view. Caspar and Klug formulated a general theory that delimited the possible numbers and arrangements of the protein subunits forming the shells of the smaller isodiametric viruses.

Until about 1948, most attention was focused on the protein part of the viruses. Quantitatively, the protein made up the larger part of virus preparations. Enzymes that carried out important functions in cells were known to be proteins, and knowledge of pentose nucleic acids was rudimentary. No function was known for them in cells, and they generally were thought to be small molecules primarily because it was not recognized that RNA is very susceptible to hydrolysis by acid, by alkali, and by enzymes that commonly contaminate virus preparations. In 1949 Markham and Smith isolated turnip yellow mosaic virus (TYMV) and showed that purified preparations contained two classes of particles, one an infectious nucleoprotein with about 35% of RNA, and the other an apparently identical protein particle that contained no RNA and that was not infectious. This result clearly indicated that the RNA of the virus was important for biological activity. Analytical studies showed that the RNAs of different viruses have characteristically different base compositions while those of related viruses are similar. About this time, it came to be realized that viral RNAs might be considerably larger than had been thought. A synthetic analog of the normal base guanine, 8-azaguanine, when supplied to infected plants was incorporated into the RNA of TMV and TYMV, replacing some of the guanine. The fact that virus preparations containing the analog were less infectious than normal virus gave further experimental support to the idea that viral RNAs were important for infectivity. However, it was the classic experiments in the mid-1950s of Gierer and Schramm, and Fraenkel-Conrat and Williams that demonstrated the infectivity of naked TMV RNA and the protective role of the protein coat. These discoveries ushered in the era of modern plant virology.

In the early 1950s Brakke developed density gradient centrifugation as a method for purifying viruses. Together with a better understanding of the chemical factors affecting the stability of viruses in extracts, this procedure has allowed the isolation and characterization of many viruses. The use of sucrose density gradient fractionation enabled Lister to discover the bipartite nature of the tobacco rattle virus genome. Since that time, density gradient and polyacrylamide gel fractionation techniques have allowed many viruses with multipartite genomes to be characterized. Their discovery, in turn, opened up the possibility of carrying out genetic reassortment experiments with plant viruses leading to the allocation of functions to many of the viral genes. Density gradient fractionation of purified preparations of some other viruses revealed noninfectious nucleoprotein particles containing subgenomic RNAs. Other viruses have been found to have associated with them satellite viruses or satellite RNAs that depend on the 'helper' virus for some function required during replication.

Further developments in the 1970s included improved techniques related to X-ray crystallographic analysis and a growing knowledge of the amino acid sequences of the coat proteins allowed the three-dimensional structure of the protein shells of several plant viruses to be determined in molecular detail.

For some decades, the study of plant virus replication had lagged far behind that of bacterial and vertebrate viruses mainly because there was no plant system in which all the cells could be infected simultaneously to provide the basis for synchronous 'one-step growth' experiments. However, following the initial experiments of Cocking in 1966, Takebe and colleagues developed protoplast systems for the study of plant virus replication. Although these systems had significant limitations, they greatly increased our understanding of the processes involved in plant virus replication.

Another important technical development has been the use of *in vitro* protein-synthesizing systems such as that from wheat germ, in which many plant viral RNAs act as efficient messengers. Their use allowed the mapping of plant viral genomes by biochemical means to begin.

The Molecular Biology Age

The molecular age opened in 1960 with the determination of the full sequence of 158 amino acids in the coat protein of TMV. The sequence of many naturally occurring strains and artificially induced mutants was also determined at about the same time. This work made an important contribution to establishing the universal nature of the genetic code and to our comprehension of the chemical basis of mutation.

Our understanding of the genome organization and functioning of viruses has come from the development of procedures whereby the complete nucleotide sequence of viruses with RNA genomes can be determined. In 1982 the genomes of both the first

plant RNA virus (TMV) and DNA virus cauliflower mosaic virus (CaMV) were sequenced. Since then, the genomes of representatives of all the plant virus genera have been sequenced and there are many sequences of virus species.

The late 1980s and 1990s was a period when molecular biological techniques were applied to a wide range of aspects of plant virology. These included the ability to prepare *in vitro* infectious transcripts of RNA viruses derived from cloned viral cDNA allowing techniques such as site-directed mutagenesis to be applied to the study of genome function and reverse genetics being used to elucidate the functions of viral genes and control sequences. These approaches, together with others such as yeast systems for identifying interacting molecules, the expression of viral genes in transgenic plants, and labeling viral genomes in such a manner that their sites of function within the cell are known, were revealing the complexities of the interactions between viruses and their hosts. Nucleotide sequence information has had, and continues to have, a profound effect on our understanding of many aspects of plant virology, including (1) the location, number, and size of the genes in a viral genome; (2) the amino acid sequence of the known or putative gene products; (3) the molecular mechanisms whereby the gene products are transcribed; (4) the putative functions of a gene product, which can frequently be inferred from amino-acid-sequence similarities to products of known function encoded by other viruses; (5) the control and recognition sequences in the genome that modulate expression of viral genes and genome replication; (6) the understanding of the structure and replication of viroids and of the satellite RNAs found associated with some viruses; (7) the molecular basis for variability and evolution in viruses, including the recognition that recombination is a widespread phenomenon among RNA viruses and that viruses can acquire host nucleotide sequences as well as genes from other viruses; and (8) the beginning of a taxonomy for viruses that is based on evolutionary relationships. On the host side, advances in plant genome sequencing are identifying plant genes that confer resistance to viruses.

During the 1980s, major advances were made on improved methods of diagnosis for virus diseases, centering on serological procedures and on methods based on nucleic acid hybridization. Since the work of Clark and Adams reported in 1977, the enzyme-linked immunosorbent assay (ELISA) technique has been developed with many variants for the sensitive assay and detection of plant viruses. Monoclonal antibodies against TMV lead to a very rapid growth in their use for many kinds of plant virus research and for diagnostic purposes. The late 1970s and the 1980s also saw the start of application of the powerful portfolio of molecular biological techniques to developing other approaches to virus diagnosis, to a great increase in our understanding of the organization and strategy of viral genomes, and to the development of techniques that promise novel methods for the control of some viral diseases. The use of nucleic acid hybridization procedures for sensitive assays of large numbers of samples and the polymerase chain reaction, also dependent on detailed knowledge of genome sequences, are being increasingly used in virus diagnosis. Most recently DNA chips are being developed for both virus diagnostics and studying virus infection.

In the early 1980s, it seemed possible that some plant viruses, when suitably modified by the techniques of gene manipulation, might make useful vectors for the introduction of foreign genes into plants. Some plant viruses have been found to contain regulatory sequences that can be very useful in other gene vector systems, notably the widely used CaMV 35S promoter. Another practical application of molecular techniques to plant viruses has been the modification of viral genomes so that products of interest to industry and medicine can be produced in plants. This is being done either by the introduction of genes into the viral genome or by modification of the coat protein sequence to enable epitopes to be presented on the virus.

Early attempts (early to mid-1900s) to control virus diseases in the field were often ineffective. They were mainly limited to attempts at general crop hygiene, roguing of obviously infected plants, and searching for genetically resistant lines. Developments since this period have improved the possibilities for control of some virus diseases. Heat treatments and meristem tip culture methods have been applied to an increasing range of vegetatively propagated plants to provide a nucleus of virus-free material that then can be multiplied under conditions that minimize reinfection. Such developments frequently have involved the introduction of certification schemes. Systemic insecticides, sometimes applied in pelleted form at the time of planting, provide significant protection against some viruses transmitted in a persistent manner by aphid vectors. It has become increasingly apparent that effective control of virus disease in a particular crop in a given area usually requires an integrated and continuing program involving more than one kind of control measure.

However, such integrated programs are not yet in widespread use. Cross-protection (or mild-strain protection) is a phenomenon in which infection of a plant with a mild strain of a virus prevents or delays infection with a severe strain. The phenomenon has been used with varying success for the control of certain virus diseases, but the method has various difficulties and dangers. In 1986 Powell-Abel and co-workers considered that some of these problems might be overcome by the application of the concept of pathogen-derived resistance of Sandford and Johnston. Using recombinant DNA technology, they showed that transgenic tobacco plants expressing the TMV coat-protein gene either escaped infection following inoculation or showed a substantial delay in the development of systemic disease. These transgenic plants expressed TMV coat-protein mRNA as a nuclear event. Seedlings from self-fertilized transformed plants that expressed the coat protein showed delayed symptom development when inoculated with TMV. Thus, a new approach to the control of virus diseases emerged. However, this approach revealed some unexpected results which led to the recognition that plants have a defense system against 'foreign' RNA. This defense system, initially termed post-translational gene silencing and now called RNA silencing or RNA interference (RNAi) was first recognized in plants and had, and is still having, a great impact on molecular approaches as diverse as disease control and understanding gene functions. Among the tools that have arisen from understanding this new phenomenon is virus-induced gene silencing which is being used to determine the functions of genes in plants and animals by turning them off.

However, the RNAi defense system in plants which targets double-stranded RNA, an intermediate RNA virus replication, raised the question of how RNA viruses replicated in plants. Studies on gene functions revealed that many plant viruses contain so-called 'virulence' genes. Many of these have been shown to suppress host RNA silencing, thus overcoming the defense system.

Suppression of gene silencing is widespread among plant viruses and examples are being found in animal viruses. Thus, as noted at the beginning of this article, plant virology is still providing insights onto phenomena applicable to virology in general.

See also: Plant Antiviral Defense: Gene-Silencing Pathways. Plant Viral Diseases: Economic Implications. Tobacco Mosaic Virus (*Virgaviridae*). Vector Transmission of Plant Viruses. Viral Suppressors of Gene Silencing. Virus-Induced Gene Silencing (VIGS)

Further Reading

- Baulcombe, D.C., 1999. Viruses and gene silencing in plants. *Archives of Virology* 15 (supplement), 189–201.
- Caspar, D.L.D., Klug, A., 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposium on Quantitative Biology* 27, 1–24.
- Clark, M.F., Adams, A.N., 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *Journal of General Virology* 34, 475–483.
- Fischer, R., Emans, N., 2000. Molecular farming of pharmaceutical proteins. *Transgenic Research* 9, 279–299.
- Hull, R., 2001. *Matthews' Plant Virology*. San Diego: Academic Press.
- Li, F., Ding, S.W., 2006. Virus counterdefense: Diverse strategies for evading the RNA silencing mechanism. *Annual Review of Microbiology* 60, 507–531.
- Lindbo, J.A., Dougherty, W.G., 2005. Plant pathology and RNAi: A brief history. *Annual Review of Phytopathology* 43, 191–204.
- Pavord, A., 1999. *The Tulip*. London: Bloomsbury.
- Van der Want, J.P.H., Dijkstra, J., 2006. A history of plant virology. *Archives of Virology* 151, 1467–1498.

Emerging and Re-Emerging Plant Viruses

Sabrina Bertin, Francesco Faggioli, Andrea Gentili, Ariana Manglli, Anna Taglienti, Antonio Tiberini, and Laura Tomassoli, Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

© 2021 Elsevier Ltd. All rights reserved.

This is an update of G.P. Martelli, D. Gallitelli, Emerging and Reemerging Virus Diseases of Plants, In Encyclopedia of Virology (Third Edition), Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00705-6.

Nomenclature

aa Amino acid(s)

asl Above sea level

BSC Bundle sheath cells

CP Coat protein or capsid protein

HC-Pro Helper component-protease

HR Hypersensitive response

HTS High-throughput sequencing

kb Kilobase

kbp Kilobase pair

kDa Kilo Dalton

MP Movement protein

NC Nucleocapsid

NSm Non-structural movement protein

NSS Non-structural RNA silencing suppressor

nt Nucleotide(s)

ORF Open reading frame

RB Resistance breaking

RdRp RNA-dependent RNA polymerase

sgRNA Sub-genomic RNA

ssRNA Single-stranded ribonucleic acid

Glossary

Emergence An emergent virus is a virus that has adapted and/or emerged as a new disease/pathogenic strain, with attributes facilitating pathogenicity in an environment normally not associated with that virus.

Isolate Unique sampling material out of which a given virus has been identified.

Pathosystem Subsystem of an ecosystem that is defined by the phenomenon of parasitism, in the present case, viruses.

Strain Virus strains are viruses that belong to the same species and differ in having stable and heritable biological, serological, and/or molecular characters.

Introduction

Plant viruses have been considered one of the most limiting biotic stresses, and their intrinsic and epidemiological features require specific containment measures aimed either to prevent (*i.e.*, use of virus-free germplasm), eradicate (*i.e.*, destruction of infected material) or control them (*i.e.*, control of vectors and use of resistance genes). Several outbreaks are recorded every year worldwide mainly due to: (1) new viruses (*i.e.*, first detected within the last five-ten years and persisting within restricted areas); (2) emerging viruses (*i.e.*, whose incidence has rapidly increased within the last decade); (3) re-emerging viruses (*i.e.*, viruses facing genetic evolution and host shift). The key factors driving the emergence of plant viruses include genetic variability of a viral population, changes in agricultural practices and/or distribution of insect vectors. Further, these factors are likely modified by the effects of climate change which, in recent years, is a major concern for plant pathologists proving to be a global problem of difficult resolution. According to the Food and Agriculture Organization (FAO), the climate change is among the driving factors of global hunger and malnutrition. It is reported that damage caused by plant pests is responsible for loss of 20%–40% of global food production. In view of this, it is easy to highlight how climate change, the emergence of new and more serious plant diseases and global malnutrition are closely related.

In addition, agricultural globalization, international trade and long-distance transport of plant materials, often associated with people migration flow, are relevant for virus spreading, leading to the introduction of viruses and their vectors to new geographical regions with unpredictable consequences at local level. In fact, epidemiological characteristics of viral pathosystems are variables depending on different agro-ecological contexts, hence the control strategy of a well-known disease in a new area can be quite challenging. For example, the introduction of a virus in a new region where highly susceptible hosts are usually cropped or, conversely, the import of such susceptible hosts in a region where the virus is already spread, can lead to severe viral outbreaks. Moreover, in such cases, early response from scientists, plant protection services and policy makers is needed, as delayed diagnosis, prevention and control measures can be inefficient.

For the above reasons, the list of new, emerging and re-emerging viruses (Table 1) needs to be often updated due to the quick environmental changes, global market and crop production. Some emergent viruses belong to important genera and families as well as the *Begomovirus* (*Geminiviridae*) within which some species suddenly increase in rate of spread, causing outbreaks and disease in new countries (*i.e.*, tomato leaf curl New Delhi virus, tomato yellow leaf curl viruses, cotton leaf curl viruses, East African

Table 1 Emerging and re-emerging viruses reviewed in the present Encyclopedia of Virology. Gray cells indicate other articles in the Encyclopedia

Order	Family	Genus	Species	Geographical spread	Status
Bunyavirales	Fimoviridae	Emaravirus	Rose rosette virus (RRV)	Canada, USA, India	Emerging
–	Tospoviridae	Orthotospovirus	Tomato spotted wilt virus (TSWV)	Worldwide	Re-emerging
Ortervirales	Caulimoviridae	Badnavirus	Grapevine Roditis leaf discoloration-associated virus (GRLDaV)	Europe: Greece, Italy, Croatia, Turkey	Emerging
–	–	–	Cacao swollen shoot disease (CSSD)	West Africa	Re-emerging
Tymovirales	Alphaflexiviridae	Potexvirus	Pepino mosaic virus (PepMV)	Worldwide	Emerging
–	Betaflexiviridae	Trichovirus	Grapevine Pinot gris virus (GPGV)	Europe, Americas, Asia, Australia	Emerging
–	Potyviridae	Ipomovirus	Cassava brown streak virus (CBSV)	Africa	Emerging
Unassigned	Virgaviridae	Tobamovirus	Cucumber green mottle mosaic virus (CGMMV)	Europe, USA, Canada, Australia	Emerging
–	–	–	Tomato brown rugose fruit virus (ToBRFV)	Middle East, Europe, Mexico, USA, China	Emerging
Unassigned	Nanoviridae	Babuvirus	Banana bunchy top virus (BBTV)	Africa, Asia	Emerging
–	Geminiviridae	Begomovirus	East African cassava mosaic virus (EACMV)	Africa, India, South East Asia	Emerging
–	–	–	Squash leaf curl virus (SLCV)	North and Central America, Saudi Arabia, Middle East, Egypt	Emerging
–	–	–	Tomato leaf curl New Delhi virus (ToLCNDV)	India, Pakistan, Bangladesh, Iran, Thailand, Indonesia, Spain, Italy, Morocco	Emerging
–	–	–	Tomato yellow leaf curl virus (TYLC)	Europe, Asia, Africa, North and Central America, Australia	Re-emerging

cassava mosaic viruses, etc.) or have the capability to produce variants by recombinant events (*i.e.*, tomato yellow leaf curl disease, East African cassava mosaic virus) causing enhancement of virulence and resistance breaking in plants. This article reports the main viral diseases, not treated in other articles, that are currently threatening farmers, scientific community, regulators, agricultural industry and even final consumers worldwide.

Vegetable Viruses

Tobamoviruses

Tobamovirus is a genus in the family *Virgaviridae*, reported to infect mainly tobacco, potato, tomato, and cucurbits. To date, thirty-seven virus species are included in this genus. According to molecular features and host range, four informal subgroups have been identified within this genus: brassicas, cucurbits, malvaceous and solanaceous group. *Tobamovirus* is considered one of the most important virus genera and the type species member, *Tobacco mosaic virus* has been the first virus ever discovered and characterized, leading to modern plant virology.

Cucumber Green Mottle Mosaic Virus

Geographical distribution

Cucumber green mottle mosaic virus (CGMMV) is an economically important pathogen infecting many cucurbit species. It was first described in cucumber plants in England in 1935. After a first period of slow diffusion, the virus has begun to cause disease more frequently after 1986. Since 2007, CGMMV represents a serious concern for the protected crops of cucumber and other cucurbits. In recent years, it rapidly has spread across European countries where it was already known and in new countries such as Bulgaria (2008) and Poland (2017). Outbreaks of CGMMV were also reported in Canada (2014), the USA (2014) and Australia (2015) with a rapid spread in different locations. To date, the virus occurs in almost all continents and over 40 countries, achieving a global distribution.

Causal agent and classification

CGMMV has a 6.4 kb single-stranded, positive sense RNA (+ ssRNA) genome encapsidated within 2000 molecules of the capsid protein (CP) resulting in a stiff, rod-shaped 300 × 18 nm particle. The genomic RNA contains four protein-coding open reading frames (ORFs). The first two putative ORFs encode a 129 kDa protein containing the methyl-transferase and helical domains, and a 186 kDa protein that contains the polymerase domain respectively. Both proteins are translated directly from the genomic RNA and are involved in RNA replication. ORF3 and ORF4 are translated from subgenomic (sg) RNAs and encode a 29 kDa protein involved in cell-to-cell movement (MP) and a 17.4 kDa CP respectively. Serological and molecular variability has been reported between European and Asian isolates.

Disease symptoms and yield losses

Symptoms caused by CGMMV vary depending upon cucurbit species and cultivars that are grown. In cucumber it causes leaf mottle mosaic, blistering, fruit discoloration, mottling and distortion. In watermelon, leaf mottling and mosaic are observed in the young leaves as well as brown necrotic lesion in the stems and peduncles (Fig. 1). Moreover, CGMMV causes a pulp deterioration called blood flesh disease and fruits lose their marketable value. In melon young leaves develop initial mottle and mosaic often disappearing in mature foliage. Fruits develop different degrees of malformation, mottling and surface netting. Pumpkin, squash and zucchini infected leaves can show leaf mottling, mosaic or may also be asymptomatic. Pumpkin fruits are always asymptomatic, while squash and zucchini fruits are externally symptomless but internally discolored and necrotic. In cucurbit seedlings symptoms are not visible but in severe infections from seed the cotyledons become yellow. Symptom development is influenced by environment, growing condition, plant growth stage at the time of infection or viral strain. Lower temperatures in early spring with low light intensity induce more severe symptoms than in summer. The damage caused by CGMMV consists in yield losses in the field and poor quality of the fruits. Serious damages are reported in cucumber and watermelon crops with yield losses up to 15% and 40%, respectively.

Epidemiology

Like other species belonging to the genus *Tobamovirus*, CGMMV is very stable and contagious. It can remain infectious for a long period in the soil contaminated with infected plant debris, in contaminated equipment surface, or on farming tools. It is easily transmitted through direct contact, seed, and other propagation material including grafting. Seed transmission is known to occur in eight different cucurbit species: watermelon, melon, cucumber, pumpkin, zucchini, bottle gourd and snake gourd. The virus contaminates the coat of the seed from infected plants. Studies reported different rates of seed transmission reaching up to 8% in bottle gourd, cucumber and watermelon. The transmission rate may decline rapidly after a few months during seed storage.

Contaminated seed stocks and infected seedlings can constitute the primary source of inoculum in a new area of production. Once introduced, the virus readily spreads mechanically during cultural operations, through contaminated instruments, hands and clothing, and through direct contact between the plants. CGMMV is transmitted also through irrigation. Insect transmission has been investigated but no insect vector is known to transmit the virus in a specific manner.

CGMMV host range is mainly restricted to the family Cucurbitaceae. The natural host range includes crops as *Benincasa hispida*, *Citrullus lanatus*, *Cucumis sativus*, *C. anguria*, *C. melo*, *Cucurbita maxima*, *C. moschata*, *C. pepo*, *Lagenaria siceraria*, *Luffa acutangula*, *L. cylindrica*, *Momordica charantia*, *Trichosanthes cucumerina*, and weeds as *Citrullus colocynthis*, *Ecballium elaterium* and *Mukia maderaspatana*. In the recent years, it has been naturally reported in non-cucurbit species in the families Amaranthaceae, Apiaceae, Boraginaceae, Portulacaceae, Euphorbiaceae and Solanaceae, which can act as reservoirs.

Control

Use of virus-free seeds and seedlings is essential to avoid virus introduction in new areas. As virus contaminates mainly the tegument, seed treatments based on dry heat (i.e., 70°C for 3 days) or disinfectants can be used without adversely affecting seed germination. In the seedling nurseries it is very important to practice stringent hygiene measures. Early detection and removal of infected plants is crucial to avoid the virus spread. Once the virus becomes established in a farm, all organic and contaminated disposable material should be destroyed by incineration or deep burial. All contaminated surfaces and irrigation systems should be cleaned with antiviral detergent or chemical disinfectant. Considering the high stability and infectivity of the virus, crop rotation avoiding any cucurbits for a minimum period of at least 24 months is recommended. Resistance to CGMMV has been reported in melon and cucumber, even if commercial varieties mainly display partial resistance (tolerance) resulting in a reduced symptom expression.



Fig. 1 Leaf mosaic and blistering of CGMMV on cucumber.

Tomato Brown Rugose Fruit Virus

Geographical distribution

Tomato brown rugose fruit virus (ToBRFV) was originally described in Jordan in 2015 and in Israel in 2014 (but reported only in 2017) on tomato plants harboring the tomato mosaic virus (ToMV) resistance *Tm-2²* gene. In 2018–19, ToBRFV outbreaks have been reported in countries on almost all the continents: Mexico, California (USA), China (Shandong district), Germany, Greece (Crete island), Italy, the Netherlands, Spain, Turkey and United Kingdom. In Mexico, Jordan and Italy ToBRFV was also reported on pepper plants not harboring *L* resistance to TMV. In view of its quick spreading and ability to overcome ToMV resistance, ToBRFV is considered an emerging pathogen, included in the EPPO Alert List in January 2019, potentially spreading worldwide, and it was suddenly regulated to contrast its introduction in free areas and spread where present. To date, ToBRFV was reported to be established in Mexico (53 provinces among 20 states), and in Israel and Jordan, where virus incidence can reach up to 100% in the greenhouse crop production.

Causal agent and classification

Tomato brown rugose fruit virus has structural and genomic features that are typical for the members of the genus *Tobamovirus*. The genome, encapsidated in a 300 nm long rod-shaped rigid particle, is constituted by a positive ssRNA of ~ 6400 nucleotides (nt) and consists of four ORFs. ORF1 and ORF2 encode two replication-related proteins of 126 and 183 kDa respectively, the latter being expressed by the partial suppression of the stop codon; ORF3 encodes the MP of 30 kDa, and ORF4 the CP of 17.5 kDa which is expressed via the 3' coterminal sgRNAs. Phylogenetic analysis showed that ToBRFV clustered distinctly from either ToMV or TMV clades. In addition, preliminary sequences analysis showed a low level of mutation rate among isolates of ToBRFV, and conserved nt sequences have been retrieved in isolates that have a different geographic origin.

Disease symptoms and yield losses

On tomato, the symptomatology includes chlorosis, mosaic and mottling on foliage, often leaf narrowing (needles) (Fig. 2) or more severe malformation; yellow and brown spots, uneven ripening and rarely rugose patches are observed on fruits which dramatically decrease their commercial value. Asymptomatic fruits were reported in several European countries and yield reduction is recorded due to a reduced number of branches and fruits. The occurrence and severity of symptoms vary by variety and environmental conditions (temperature and light intensity). The disease may result in yield losses of up to 70%. Plants for planting are generally asymptomatic. On pepper, foliar symptoms include deformation, yellowing and mosaic; fruits are deformed, with yellow or brown areas or green stripes.

Epidemiology

ToBRFV, as all tobamoviruses, is mechanically transmitted by contact. This transmission manner includes different patterns such as plant-to-plant contact, grafting, anthropic transmission by contaminated tools, hands and clothing during cultural practices (transplanting, pruning, staking, trellising, tying, spraying, and harvesting) as well as commercial packing and trade operations. ToBRFV was found in seed coats, and there is reason to suspect seed transmission early infecting emerging plantlets, but additional experimental trials should be carried out to confirm. Nevertheless, even if transmission rate from seed to seedling is low, further dissemination by contact (*e.g.*, during transplantation of seedlings or regular handling of the crop) allows a rapid spread within a glasshouse. In addition, ToBRFV could be carried by bumblebees (*Bombus terrestris*) and transmitted to healthy tomato plants during pollination (mechanically by visiting flowers of healthy tomato plants). As reported for all tobamoviruses, ToBRFV is characterized by very stable particles. In fact, preliminary studies showed that ToBRFV can survive in crop debris, in soil, and on implements, stakes, trellis wires, fruit containers, greenhouse benches and seedling trays for a long period (several weeks to months).

Tomato (*Solanum lycopersicum*) and pepper (*Capsicum* spp.) are the main hosts of ToBRFV. By experimental inoculations, symptoms were showed in *Nicotiana benthamiana*, *N. glutinosa*, *N. sylvestris*, *N. tabacum*, *Petunia hybrida*, *Datura stramonium*, *Chenopodium* spp. In addition, ToBRFV is reported to infect weeds such as *Chenopodium murale* and *Solanum nigrum*, that can



Fig. 2 Symptoms of mosaic and needles of ToBRFV on tomato.

therefore act as reservoirs. In experimental studies, aubergine (*Solanum melongena*) and potato (*S. tuberosum*) did not show symptoms after inoculation, and ToBRFV was not found when the plants were subsequently tested for latent infection detection.

Control

Regulations have been introduced to prevent entries of ToBRFV through seeds and/or plants for planting of tomato and pepper (i.e. EU Commission Decision 2019/1615; WTOs – Emergency phytosanitary measures). Certification of seeds by molecular tests and only seedling from certified seeds are required for import/export. Strict sanitation practices aiming to prevent ToBRFV establishment indoor (treatment of seeds, sanitation of clothing, tools and equipment, removal and destruction of infected debris) and its spread (packaging area out of production sites, sanitation of boxes for harvesting) are mandatory.

Orthospoviruses

Orthospovirus is a genus of the family *Tospoviridae* (order *Bunyavirales*). These thrips-transmitted viruses are amongst the most devastating plant viruses worldwide, causing severe economic damage to many vegetable crops, such as tomato and sweet pepper. The often wide and overlapping host ranges, emergence of resistance-breaking strains, circulative and propagative transmission by thrips, and difficulties in predicting their outbreaks cause high concern for their control.

Tomato Spotted Wilt Virus

The evolution of new tomato spotted wilt virus (TSWV) variants by mutation poses high concern for the capacity to adapt to tomato and pepper varieties carrying the resistance genes *Sw5* and *Tsw* respectively. For this reason, TSWV may be considered a re-emerging pathogen and included in the present article as a potential threat for the vegetable crops.

Geographical distribution

Tomato spotted wilt virus (TSWV) is present in all countries under temperate, tropical and subtropical climate conditions, being a serious and endemic pathogen since almost the last 20 years. Since the first description of “spotted wilt” disease in Australia in 1915, the spreading of TSWV became uncontrolled with outbreaks occurring in vegetable crops, such as tomato, lettuce and pepper, tobacco and in ornamentals especially in southern Europe. A threatening resurgence of TSWV occurred since 1980 in North America, 1987 in Western Europe and 1993 in Australia, leading to a worldwide distribution.

Causal agent and classification

TSWV, as other members of the *Orthospovirus* genus, is enveloped including additional protein package on its RNA genome, which is segmented in three portions named S, M and L, according to the size of the RNA molecules, 2.9 kb, 4.8 kb, and 8.9 kb, respectively. The L segment encodes putative RNA-dependent RNA polymerases (RdRp), such as replicase, transcriptase, nuclease, helicase, cap-binding and NTPase proteins responsible for several enzymatic functions of TSWV. The M and S RNAs have special structures where genomes are ambisense, both including parts of positive and parts of negative polarities. In particular, the M segment produces envelope glycoproteins (Gn-Gc) and in antisense orientation non-structural movement protein (NSm). The S segment has two ORFs encoding nucleocapsid (NC) protein and non-structural RNA silencing suppressor (NSS), with antisense and sense orientation, respectively.

Disease symptoms and yield losses

TSWV can induce a wide variety of symptoms on economically important ornamental plants, vegetables and industrial crops. The symptomatology expression is affected by plant species and cultivar, the developmental stage at the time of inoculation, the nutritional and environmental conditions (outdoor or greenhouse cultivations) as well as TSWV strain. Generally, symptoms on leaves include chlorotic or necrotic lesions, ring spots, concentric rings, line patterns together with crinkling, bronzing, distortion (curling) of leaves and a general plant disorder as wilting, stunting, mottling and necrosis. Symptoms on fruits usually could include irregular discoloration, yellow or brown flecks or rings, or necrotic lesions or rings on unripe fruits (Fig. 3). Affecting both yield and marketing quality, TSWV has been reported as the second most destructive viral pathogen in the list of economic damage caused by plant viruses. In fact, in tomato, it causes a yield and marketable value reduction up to 42% and about 95% respectively, with worldwide estimated losses exceeding one billion dollars annually. The economic impact of TSWV is great, due to its wide geographical distribution and to its broad host range. Crops in which important losses due to TSWV have been reported are tomato, pepper, lettuce, eggplant, papaya, French beans, celery and ornamental plants.

Epidemiology

TSWV is naturally transmitted by at least eight species of thrips, in a circulative and propagative manner. Analysis of factor determining vector competence showed *Frankliniella occidentalis* (Pergande) as the most efficient vector. In fact, the western flower thrips, *F. occidentalis*, contributed to the spreading and worldwide occurrence of TSWV starting from 80s. TSWV is characterized by a wide host range, and this relative lack of host specificity along with highly polyphagous nature of its vector are the two main factors that make TSWV one of the most widely distributed plant viruses. A recent report included more than 1000 plants species



Fig. 3 Fruit malformation and irregular discoloration in pepper infected by TSWV. Courtesy of M. Turina.

(exactly 1090) capable of being locally and/or systematically infected by TSWV. The species are from a single family of the class Pteropsida (Division X - Pteridophyta), 69 families of the class Dicotyledones (Division XII - Angiospermae) and 15 families of the class Monocotyledones (Division XII - Angiospermae). Most of the plants listed as hosts of TSWV belong to the family Asteraceae (247 species), Solanaceae (172 species) and Fabaceae (60 species).

Control

Prevention and control of the spread of TSWV is very difficult due to its vector which needs a complex integrated program itself for management. As for other plant viruses, virus-free planting material is a prerequisite to prevent TSWV in non-resistant cultivars. Cultural practices play an important role in controlling disease spreading as removing weeds and infected plants, destruction or removal of old crops. Additionally, density row planting, cultivar selection, date of planting, reflective mulches could have powerful effects on vectors, although optimum levels differ crop by crop, and by conditions of field and climate. Direct vector control by insecticides is needed but establishing appropriate rotations and employing products from different mode of action groups as essential to prevent the development of insecticide resistance on the target insect populations. An alternative indirect strategy to mitigate natural thrips density and damage is the use of predatory insects such as the minute pirate bugs (*Orius insidiosus*) and big eyed bugs (*Geocoris punctipes*), mainly successful in protected crops.

Host plant resistance is probably the most effective way to control this virus. Natural resistance to TSWV has been reported for different crops, including chrysanthemum, lettuce, pepper and tomato. Currently, two single dominant resistance genes against TSWV, *Sw-5* in tomato and *Tsw* in pepper, are known and both provide a hypersensitive response (HR) type of resistance. This means that resistant plants react with an HR against virus infection resulting in necrotic local lesions.

Resistance breaking strains

In recent years, increasing occurrence of TSWV strains characterized by a phenotype able to break resistance (RB strains) in resistant plants were recorded, even in natural conditions. Emergence of RB isolates of TSWV seems to be correlated with the extensive cultivation of resistant hybrids. From 2011 (Spain) resistant tomato plants seriously infected with TSWV were reported widely in most of the cultivation areas where TSWV is endemic and resistant varieties are largely used. In 2016, in California (USA) unusually early and severe symptoms of TSWV occurred in fields of TSWV-resistant (*Sw-5*) fresh market tomato cultivars. The occurrence of RB TSWV strains in California occurred 5–6 years after ever-increasing planting of *Sw-5* cultivars in hot spot areas, included in the local integrated pest management strategy. The extreme variability of TSWV isolates, joined with the possibility of the exchange of genetic information through reassortment of genome segments, is suggested as the main cause for the apparent readiness of the virus to adapt and overcome both natural and pathogen-mediated resistance.

Begomoviruses

The genus *Begomovirus* of the family *Geminiviridae* is the largest group of DNA plant-infecting viruses comprising 409 recognized species. Begomoviruses are transmitted by the whitefly *Bemisia tabaci* (Gennadius) and cause disease in dicotyledonous plants. In the warmer parts of the world, these viruses are responsible of significant yield losses for economically important vegetable and fiber crops. This genus comprises viruses of monopartite and bipartite single-stranded DNA genome encapsidated in geminate particles of $\sim 30 \times 18$ nm consisting of two incomplete icosahedra. Two groups of begomoviruses are reported: “New World” with bipartite genome and “Old World” comprising either bipartite or monopartite genomes.

Squash Leaf Curl Virus

Two strains named squash leaf curl virus-extended (SLCV-E) and squash leaf curl virus-restricted (SLCV-R) based on their host range and symptom severity have been identified. According to molecular analysis, the two strains were successively recognized as separate (but closely related) species: SLCV-E remained SLCV, while SLCV-R was named squash mild leaf curl virus (SMLCV).

Geographical distribution

Squash leaf curl disease (SCLD) was described for the first time in squash in California (1981) and in cultivated buffalo gourd in Arizona (1986). Since at least 1977, both SLCV and SMLCV were reported in the southwestern United States and northern Mexico and successively in Texas in watermelon plants. SLCV remains restricted in central and north America until 2000, when it was also found in Saudi Arabia and rapidly spread in the Middle East and other countries: Israel (2003), Egypt (2006), Jordan (2008), Palestine (2010), Lebanon (2012).

Causal agent and classification

SLCV is a typical “New World” begomovirus with bipartite genome DNA-A and DNA-B, each approximately 2.6 kb in size with different sequences except for an intergenic region (RI) also called as common region (CR) involved in replication initiation. DNA-A is essential for replication and encapsidation. DNA-A encodes five ORFs, one (AV1-coat protein) in the viral sense and the other four (AC1-replication initiator, AC2-transcriptional activator, AC3-replication enhancer protein and AC4-within the AC1) in the complementary sense. DNA-B encodes two ORFs, BV1- nuclear shuttle protein in viral sense and BC1- movement protein in complementary sense. DNA-B is essential in systemic movement of the virus and symptoms production.

Disease symptoms and yield losses

SLCV causes different symptoms according to the host, cultivar or growing stage. In the squash plants, it causes stunting, severe leaf curling, yellow mottling, blisters, and fruit deformation (Fig. 4). Similar symptoms can occur also in other cucurbits as melon and watermelon. In common beans it causes leaf curling, vein and stem necrosis. Leaf curling, yellowing and stunting of whole plants can occur also in *Malva parviflora*. In cucurbit crops, severe epidemics with almost 100% of virus incidence were reported in Israel. Higher incidence up to 95% was also reported from different growing areas in Jordan.

Epidemiology

The virus is transmitted in the field by *B. tabaci*, especially MEAM1 (formerly named biotype B) in a persistent circulative and non-propagative manner. Once acquired the virus is retained in the vector for a period of several weeks, often for the entire life during which the vector remains viruliferous to transmit the virus. Different weed species reported until now to be naturally infected by SLCV suggest that these weeds can serve as reservoir for the virus/vector during the year. In international trade the virus can be introduced through transplant movement and mainly by its vector *B. tabaci* on host plants, considering the wide host range of the vector and the virus persistence in the vector for a long time after acquisition (several weeks to whole life).

The virus infects plants among the Cucurbitaceae as squash (*Cucurbita pepo*), melon (*Cucumis melo*), watermelon (*Citrulus lanatus*), cucumber (*C. sativus*), giant pumpkin (*C. maxima*), pumpkin (*C. moschata*), buffalo gourd (*Cucurbita foetidissima*) and wild cucurbits (i.e., *Echallium elaterium*). In addition to cucurbits, the virus has been reported to infect common bean (*Phaseolus vulgaris*) and other weeds such as *Chenopodium murale*, *Datura stramonium*, *Prosopis farcta*, *Convolvulus* sp. and *Malva* sp.

Control

Outbreaks of the virus disease are associated with high population of *B. tabaci*. The virus control is closely related with the control of its vector *B. tabaci*, by chemical and physical methods in order to reduce whitefly density. Systemic insecticides can be applied to seedlings before transplanting and during the growing season. Row covers and reflective mulches can be used to protect plants from whiteflies early in the season. Other agricultural practices as a free cucurbit period of at least two-three months and weeds control can remove the reservoir hosts for virus and vector. To prevent the virus introduction and its spread in new areas the use of virus-free and whitefly-free transplants is very important. In cucumber, melon, squash and for some strains of the virus in watermelon, host resistance to SLCV has been identified but this resistance has not been widely introduced into commercial varieties. Commercial zucchini varieties with intermediate resistance (tolerance) to SLCV are available.



Fig. 4 Stunting and leaf malformation caused by SLCV in zucchini plant. Courtesy of G. Anfoka.

Grapevines Viruses

Grapevine Roditis Leaf Discoloration-Associated Virus

Geographical distribution

Grapevine Roditis leaf discoloration (GRLD) disease was first described in Greece in 1989 on 4-years old grapevines (*Vitis vinifera*) cv. Roditis, a red berry cultivar, grafted on 110R rootstock. The etiology of the disease remained unknown for more than 25 years until 2015, when a High-Throughput Sequencing (HTS) platform was used for sequencing siRNAs from samples of vine cv. Roditis showing typical GRLD symptoms. HTS analysis detected a DNA virus which had not been described yet; further studies showed that it was closely related to GRLD disease, hence it was given the name of Grapevine Roditis leaf discoloration-associated virus (GRLDaV). Since then, GRLDaV has been detected in a limited number of symptomatic and asymptomatic samples of different cultivars in Greece, Italy, Croatia and Turkey, mainly by HTS techniques; only in the latter country, positive samples were detected by Sanger sequencing of PCR amplicons. Since October 2018, GRLDaV has been added to the EPPO Alert List.

Disease symptoms and yield losses

Symptoms mostly appear on young leaves in late summer, consisting of yellow and/or reddish discolorations along the veins, on the interveinal areas or in wide or small surface area of the leaf blade (Fig. 5). Symptomatic leaves are also deformed, reduced in size and display abnormal venation and downward curling on the discolored sectors. Yield losses are associated with this disease, both in quantity and quality of grapes: in infected plants, bunches decrease in number and size and do not reach complete ripening and full color; a decrease in sugar content in yielded berries is also reported. In Italy (Apulia region) GRLDaV was detected in asymptomatic plants of 'Bombino nero'.

Causal agent and classification

GRLDaV is a member of the genus *Badnavirus*, family *Caulimoviridae* and displays the closest relationship with fig badnavirus 1. It is a circular dsDNA virus having non-enveloped bacilliform particles, $\sim 130 \times 30$ nm, and containing a genome of 6988 nt, which is organized in 4 ORFs. ORF3 encodes a polyprotein characterized by motifs involved in replication, encapsidation and movement function of the virus, while the other three ORFs encode proteins whose role is yet unknown.

Epidemiology

The only known natural host of GRLDaV is *V. vinifera* and graft transmission was confirmed; the virus is mechanically transmissible to grapevine and to some herbaceous test plants (*Chenopodium quinoa*, *Gomphrena globosa* and *Nicotiana benthamiana*). On



Fig. 5 Yellow discoloration of grapevine leaves by GRLDaV. Courtesy of N. Katis.

the basis of the transmission manner of other badnaviruses and the natural spreading of the disease observed in Greece in late 80s, transmission by mealybug vectors (family *Pseudococcidae*) is likely, but remains still unconfirmed. Such studies also assessed the disease incidence at very low level (less than 1%). To date, GRLDaV-infected grapevine propagative material remains the main way of virus spreading over long distances. The actual impact of GRLDaV on grapevine cropping is difficult to assess, due to the limited number of positive samples detected worldwide and insufficient information on its biology, epidemiology and distribution; nonetheless, due to the major cultural and economic importance of grapevine, surveys for the presence of GRLDaV in grapevine-growing countries are recommended.

Control

The good health status of propagative material is the main preventive measure to mitigate impact of virus diseases. In viticulture, the implementation of compulsory certification programs and the use of certified virus-free propagative material (scionwood and rootstocks) are playing a key role in a healthy grapevine production reducing the inoculum potential for further spread of the disease. To date, GRLDaV is not included in such programs but, in order to avoid its further diffusion, grapevine-growing countries where GRLDaV is present should consider extending regulation also for this virus.

Grapevine Pinot Gris Virus

Geographical distribution

Grapevine Pinot gris virus (GPGV) was first identified and characterized in northern Italy in 2012 where symptoms were observed since 2003, particularly in 'Pinot gris'. During the last years the disease was reported in many other European countries, in North and South America (Brazil, Canada, Chile, the United States), Asia (China, Georgia, Korea, Pakistan) and Oceania (Australia).

Disease symptoms and yield losses

The disease appears with chlorotic mottling, puckering and deformation of the leaves and causes a general growth disorder of plants, with delayed budburst and shortened shoot internodes (Fig. 6). Often, infected plants show most visible symptoms at the beginning of the vegetative revival (spring) while during summer they produce symptomless shoots and leaves. Reduced yield due to low grapes weight and number has been reported. The association between the symptomatology and the presence of GPGV is unclear, indeed the virus is often detected in symptomless plants. Moreover, GPGV causes a higher titratable berry acidity, so that wine produced from infected grapes presents altered flavor and aroma.

Causal agent and classification

GPGV is a positive-sense ssRNA virus belonging to the family *Betaflexiviridae*, genus *Trichovirus*. The viral genome is of 7259 nt and is organized in 3 ORFs: ORF1 encoding a polypeptide of 214 kDa including methyl-transferase, helicase and RdRp; ORF2 encoding the MP of 46 kDa and ORF3 the 22 kDa CP. Viral particles are filamentous and localized in the bundle sheath cells (BSCs) of the phloem parenchyma.

Epidemiology

The main way of GPGV spread seems to be related to the use of infected propagation material probably facilitated by using asymptomatic mother plants. GPGV is also transmitted by grafting and by the eriophyid mite *Colomerus vitis*, although this second pathway of diffusion seems to have a much lower incidence. *Silene latifolia* subsp. *alba* (Mill.) and *Chenopodium album* L. were found infected by the virus suggesting these weeds could act as reservoir plants.



Fig. 6 Growth reduction and yellowing of a GPGV infected grapevine plant.

Control

The unclear role of *C. vitis*, due to its low efficiency of transmission and limited mobility, does not suggest the control of the vector as a resolute measure to contrast the diffusion of the disease. The use and trade of infected propagation material are the main ascertained causes of rapid and global dispersal of GPGV worldwide. Therefore, a good control strategy should be focused on developing GPGV-free certified germplasm. To date, GPGV is not subjected to phytosanitary regulation in Europe, whereas it is considered a quarantine pathogen in Australia.

Staple Crops: Cassava Brown Streak Disease

Geographical distribution

Cassava (*Manihot esculenta* Crantz, family Euphorbiaceae) is vulnerable to at least 20 different viruses, of which those causing cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) are the most economically important. At the turn of the 21st century, CBSD re-emerged and rapidly spread at an alarming rate in both East and Central Africa, threatening the food security of millions of cassava farmers and attracting intense scientific interest. The first report of CBSD was from northern coastal areas of Tanzania in 1935. Next reports by 1950 noted that the affected areas were almost entirely restricted to coastal areas of East Africa from north eastern Kenya to northern Mozambique and it was believed that the disease did not spread at altitudes over 1000 m above sea level (asl). The first systematic countrywide assessment of CBSD was completed in 1994 and high CBSD incidences in areas of Tanzania, Mozambique and Malawi were reported. However, the view that CBSD is a lowland disease remained unchanged until 2004, when a significant spread of CBSD occurred in central and southern Uganda at altitudes above 1000 m asl. Significant increases in the incidence and distribution of the disease were recorded later from 2007 to 2011 in Uganda, western Kenya and the Lake Victoria zone of Tanzania. Additional reports have also been published in recent years from Rwanda (2011), Burundi (2011), eastern Democratic Republic of Congo (2012), Mayotte Island (2014) and southern Sudan (2017).

Causal agent and classification

Two species of viruses are known to cause CBSD: *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) which belong to the genus *Ipomovirus*, family *Potyviridae*. The viruses have characteristic pinwheel-like or cylindrical inclusions found in the phloem tissue and a positive-sense ssRNA genome of ~9 kb. The genome of CBSV is translated as a large polyprotein and autocatalytically cleaved by virus-encoded protease into 10 mature protein products: the serine protease/silencing suppressor (P1); the third protein (P3); two 6 kDa proteins (6K1 and 6K2); the cylindrical inclusion protein (CI); the viral genome-linked protein (VPg); the main viral protease (NIaPro); the viral RdRp (NIb); a putative pyrophosphatase (Ham1); the coat protein (CP). An additional P3N-PIPO (N-terminal of P3 fused with the Pretty Interesting Potyviridae ORF) protein is produced through ribosomal frameshifting. Genomic analysis has revealed that CBSVs share unusual features. First, CBSVs are the first members within the family *Potyviridae* that encode a single P1 proteinase and lack the multi-functional helper-component proteinase protein (HCPro). The HCPro activities (silencing suppressor, vector transmission and long-distance movement *in planta*) appear to have been replaced by silencing suppressor activity of P1. Moreover, CBSV and UCBSV encode novel Ham1 proteins with conserved motifs. The functions of Ham1 proteins are yet to be elucidated, but they are likely to provide essential functions in the lifecycle and pathogenicity of CBSVs and have a role in determining the rate of virus replication and movement as well as the severity of symptoms. Sequence analysis showed that CBSV and UCBSV share 76%–78% nt and 87%–90% aa identity respectively. These virus species are not limited to agro-ecological areas and frequently occur in mixed infections but the potential interaction between them is not currently understood.



Fig. 7 Spoilage of fresh roots in cassava by CBSD (left) and symptoms on old leaves of cassava (right). Courtesy of J. Legg.

Disease symptoms and yield losses

The aerial symptoms of CBSD in cassava include feathery chlorosis along the veins of the leaves or sometimes circular patches of chlorosis between the primary veins (Fig. 7, right), brown necrotic streaks on the stem and stem die-back in severe cases. Symptoms in the tuberous roots consist of a brown, corky necrosis of the starchy tissue, occasional radial constrictions (Fig. 7, left) and a reduction in the content of starch and cyanide. The two causal agents produce distinct symptoms on both cassava and indicator hosts: CBSV causes more severe root necrosis and feathery chlorosis along vein margins, which develops into chlorotic blotches, whereas UCBSV is responsible for circular chlorotic blotches between veins as well as leaf curling and blistering in indicator plants. The viral symptoms also depend on the virus isolate, variety of cassava, age of plant and environmental conditions.

Despite the distribution circumscribed to Africa, CBSD is reported by FAO as one of the most impacting diseases since cassava represents the main food intake in the whole continent. From some of the earliest studies of CBSD, it was noted that the disease can cause losses of crop production up to 70% due to reduced growth and spoilage of fresh roots in the most susceptible cassava cultivars. This lets to consider CBSD a re-emergent cause of serious cassava losses in East Africa. The further westward spread in Africa towards Nigeria, which is currently the world's largest producer of cassava, is a great concern.

Epidemiology

Cassava is the only known natural host of CBSVs, and no alternative crop or weed hosts have been reported. CBSV has been detected in the wild cassava relative *Manihot glaziovii* but the importance of this plant to CBSD epidemiology is not currently known. CBSVs can be mechanically transmitted to several herbaceous plant species that serve as indicator hosts, including *Petunia hybrida*, *Datura stramonium*, *Nicotiana benthamiana*, *N. tabacum*, *N. rustica* and *N. glutinosa*. CBSV and UCBSV are semi-persistently transmitted by the whiteflies *Bemisia tabaci*, *Trialeurodes vaporariorum* and *Aleurodicus dispersus*. Among these species, *B. tabaci* is considered the primary vector. Whiteflies acquire viruses in 5–10 min, retain them for up to 48 h and transmit them over relatively short distances. Even if the reasons of the sudden increase in CBSD incidence and geographical range in the last decade have not been clearly understood, it has been hypothesized that the disease is dispersed over long distances through the trade transportation of infected planting material or plants carrying viruliferous whiteflies. For instance, the primary causes of the increase of CBSD above 1000 m asl in the African Great Lakes region appears to be the concurrent dramatic increase of *B. tabaci* populations in that region.

Control

There is no cassava cultivar with high levels of CBSD resistance available to farmers. Anyway, the use of cultivars tolerant to CBSD is the more effective approach in reducing the economic impact. Conventional breeding programs produced the *M. esculenta* - *M. glaziovii* hybrid, known as 'Namikonga' in Tanzania or 'Kaleso' in Kenya, which reacts to CBSVs infection merely with mild disease symptoms and low incidence of root necrosis. However, the adoption of this hybrid that remains susceptible to CBSVs does not remove viral inoculum from the field. Genetic engineering approaches can offer promising potential for the control of CBSD, bypassing the appearance of undesirable traits in the hybrids and increasing the chances of CBSD and CMD combined resistances in cassava. The management of CBSD needs to pass through adequate phytosanitary measures. The production of "clean" stocks of planting material as well as its maintenance through rouging and selection of healthy stems for replanting are the major components of the CBSD control programs that are currently being implemented at a community level in the affected

regions of Africa. Moreover, it is very important to establish and enforce strict quarantine measures on the exchange of cassava germplasm between countries, regions and continents (*i.e.*, in Latin America) in order to prevent accidental introduction of CBSD to countries where it is currently absent. Finally, the control of whiteflies is critical to prevent the dissemination of cassava viruses over short distances.

Ornamental Viruses: Rose Rosette Virus

Geographical distribution

Rose rosette disease (RRD) was originally described in Canada in 1940 and in the United States in 1941, where it was first reported in California and Wyoming. During the last decade, a virus, therefore named rose rosette virus (RRV), was associated with this disease which rapidly spread to North-Central, South-Central and South-East USA with an exponential growth in incidence for the cultivated and garden roses (*e.g.*, in South-Central USA). In 2017, a first detection of RRV was reported in India.

Disease symptoms and yield losses

Several different symptoms have been observed in plants infected by RRV. The most recognized symptoms include fast elongation of new vertical and lateral shoots with intense red pigmentation followed by development of witches' brooms or clustering of small branches (Fig. 8). In the witches' broom, leaves are observed to be smaller, distorted, and may have a marked red pigmentation, although red pigmentation alone is not considered a consistent symptom. Other symptoms include flowering decrease, reduction and color alteration of petals, as well as petals and sepals switching to leaf-like tissue. Not all the rose varieties show all these symptoms and some of them (*e.g.*, witches' broom) look very similar to herbicide injury caused by glyphosate on roses. Infected rose plants often die within one to two years.

Causal agent and classification

The causal agent of the disease is Rose rosette virus (RRV), belonging to the genus *Emaravirus*, family *Fimoviridae*. Like other emaraviruses, RRV has double membrane-bound particles of 120–150 nm in diameter, and has multipartite, negative-sense ssRNA genome. The genome structure of emaraviruses displays significant variability. RRV was reported as a quadra-segmented virus and more recently three novel genomic RNA segments were discovered, one of which is predicted to be bicistronic.



Fig. 8 Rose rosette virus symptoms on a garden rose 'Firefighter'. Courtesy of F. Ochoa Corona.

Epidemiology

The disease is transmitted by the eriophyid mite *Phyllocoptes fructiphilus* and by grafting. RRV causes severe symptoms in *Rosa multiflora*, which can represent a common source of inoculum: rose plants cultivated downwind of infected multiflora are at risk for the vector dispersion by air currents from infected to healthy plants. The causal agent of rose rosette disease is not soil-borne; however, being a systemic pathogen, the virus may persist in RRV-infected root fragments that remain in the soil. The plants that regrow from these old root pieces, as in multiflora rose, can serve as an inoculum source for the vector and the virus can be transmitted to healthy plants.

Control

No effective control is available for rose rosette disease to recover infected plants, and there is no evidence of resistant cultivars so far. Early detection of the disease is a key measure for controlling the virus introduction in new areas. Any suspected rose should be removed and destroyed immediately as soon as presence of RRD is ascertained. In some areas, burning is permitted and can be used to destroy the diseased plants. If burning is not allowed in the area, plants should be bagged and removed. In the infected areas, chemical treatments with miticides can reduce the vector population and the risk of virus spread. For new planting beds, strict measures are suggested: use of healthy planting material, avoiding dense plantations, use of no-host barriers to limit wind dispersal of infectious mites, disinfection of pruning tools, thorough removal of symptomatic plants and ensure that infected plants do not have chance to regrow from old, infected root fragments.

Further Reading

- Aramburu, J., Marti, M., 2003. The occurrence in north-east Spain of a variant of tomato spotted wilt virus (TSWV) that breaks resistance in tomato (*Lycopersicon esculentum*) containing the *Sw-5* gene. *Plant Pathology* 52, 407.
- Babu, B., Washburn, B.K., Poduch, K., Knox, G.W., Paret, M.L., 2016. Identification and characterization of two novel genomic RNA segments RNA5 and RNA6 in rose rosette virus infecting roses. *Acta Virologica* 60, 156–165. doi:10.4149/av_2016_02_156.
- Bertazzon, N., Forte, V., Filippin, L., Borgo, M., Angelini, E., 2016. Evolution of the new grapevine disease of pinot gris and of grapevine pinot gris virus (GPGV). *BIO Web of Conferences* 7, 01042. doi:10.1051/bioconf/20160701042.
- Dombrovsky, A., Tran-Nguyen, L.T.T., Jones, R.A.C., 2017. Cucumber green mottle mosaic virus: Rapidly increasing global distribution, etiology, epidemiology, and management. *Annual Review of Phytopathology* 55 (1), 231–256. doi:10.1146/annurev-phyto-080516-035349.
- Laney, A.G., Keller, K.E., Martin, R.R., Tzanetakis, I.E., 2011. A discovery 70 years in the making: Characterization of the rose rosette virus. *Journal of General Virology* 92, 1727–1732. doi:10.1099/vir.0.031146-0.
- Luria, N., Smith, E., Reingold, V., et al., 2017. A new Israeli tobamovirus isolate infects tomato plants harboring *Tm-2²* resistance genes. *PLoS One* 12 (1), e0170429. doi:10.1371/journal.pone.0170429.
- Maliogka, V.I., Olmos, A., Pappi, P.G., et al., 2015. A novel grapevine badnavirus is associated with the Roditis leaf discoloration disease. *Virus Research* 203, 47–55.
- Martelli, G.P., Saldarelli, P., 2015. Phytosanitary challenges for the Mediterranean viticultural industry: Emerging grapevine viruses. *Watch Letter* 33, 4. https://www.ciheam.org/publications/174/016_-_Martelli.pdf.
- Patil, B.L., Legg, J.P., Kanju, E., Fauquet, C.M., 2015. Cassava brown streak disease: A threat to food security in Africa. *Journal of General Virology* 96, 956–968.
- Saldarelli, P., Gualandri, U., Malossini, U., Glasa, M., 2017. Chapter 17 – Grapevine pinot gris virus. In: Meng, B., Martelli, G.P., Golino, D.A., Fuchs, M. (Eds.), *Grapevine Viruses: Molecular Biology, Diagnostics and Management*. Springer International Publishing. doi:10.1007/978-3-319-57706-7_17.
- Saleem, A.M., Rana, J., Eman, H., Omar, M., Salam, A., 2013. Squash leaf curl virus (SLCV): A serious disease threatening cucurbits production in Palestine. *Virus genes* 48. doi:10.1007/s11262-013-1012-1.
- Tomlinson, K.R., Bailey, A.M., Alicai, T.T., Seal, S., Foster, G.D., 2018. Cassava brown streak disease: Historical timeline, current knowledge and future prospects. *Molecular Plant Pathology* 19 (5), 1282–1294.

Relevant Websites

- <https://www.cabi.org/isc/datasheet/16951>
Cucumber green mottle mosaic virus.
- <https://roserosette.org>
Rose Rosette.
- https://microbewiki.kenyon.edu/index.php/Rose_Rosette_Virus
Rose Rosette Virus.
- <https://www.cabi.org/isc/datasheet/15038>
Squash leaf curl virus (leaf curl of squash).
- <https://gd.eppo.int/taxon/tobr/v>
Tomato brown rugose fruit virus (TOBRFV).

Emerging Geminiviruses (*Geminiviridae*)

Muhammad S Nawaz-ul-Rehman, University of Agriculture, Faisalabad, Pakistan

Nazia Nahid, GC University, Faisalabad, Pakistan and University of Agriculture, Faisalabad, Pakistan

Muhammad Mubin, University of Agriculture, Faisalabad, Pakistan

© 2021 Elsevier Ltd. All rights reserved.

This is an update of C.M. Fauquet, M.S. Nawaz-ul-Rehman, Emerging Geminiviruses, In Encyclopedia of Virology (Third Edition), Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00705-6](https://doi.org/10.1016/B978-012374410-4.00705-6).

Nomenclature

CLCuD Cotton leaf curl disease

CMD Cassava mosaic disease

CMG Cassava mosaic geminivirus

CP Coat protein or capsid protein

nt Nucleotide(s)

SEA South East Asia

ssDNA single-stranded deoxyribonucleic acid

ToLCNDD Tomato leaf curl New Delhi disease

TYLCD Tomato yellow leaf curl disease

Glossary

Geminivirus A class of plant viruses having circular single stranded DNA genome encapsidated in icosahedral geminate particle.

Pandemic The outbreak of disease at large scale, across several countries.

Satellites A sub-viral agent composed of nucleic acids, which requires a helper virus for its key functions.

Introduction

The single stranded DNA (ssDNA) viruses of the family *Geminiviridae* are serious pathogens of agricultural crops in the tropical and sub-tropical regions of the world. The geminiviruses can be classified into 9 different genera and more than 440 species based on their genome organization, sequence similarity, insect vectors and specific plant hosts. These genera include, *Becurtovirus*, *Begomovirus*, *Capulavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Mastrevirus*, *Topocuvirus*, and *Turncurtovirus*. The genus *Eragrovirus* has only two member species, while the genus *Begomovirus* comprises hundreds of species. If we consider the number of species of each genus, the begomoviruses contribute the most to the economic impact of the family *Geminiviridae*. In the Old World, begomoviruses are generally associated with satellites, known as alphasatellites and betasatellites. The betasatellites encode for a pathogenicity determinant protein, while alphasatellites apparently do not play any significant role in disease development. In the past three decades, the number of geminivirus species described has tremendously increased. As begomoviruses have a more devastating impact on world's agriculture, they are therefore, the main focus of this article. The success of begomoviruses can be attributed to many factors such as climatic changes, introduction of susceptible host genotypes and the diversification of insect vector whitefly (*Bemisia tabaci*). Although, whitefly is the only vector for begomoviruses transmission, but their spread through stem cuttings and grafted material have also played an important role in their global spread for vegetatively propagated plant hosts. Begomoviruses can be monopartite or bipartite. The monopartite begomoviruses contain one circular molecule (DNA-A), coding for six different genes necessary for replication, transcription, and other important regulatory functions. While, bipartite viruses have two circular genomes (DNA-A and DNA-B). Where DNA-A of bipartite viruses is equivalent to genomic component of monopartite viruses, DNA-B encodes two necessary proteins for movement of the virus in the host. Because there is a huge diversity of begomoviruses infecting many crops, we have focused our interest only on four major diseases caused by begomoviruses: the Tomato yellow leaf curl disease, the Tomato leaf curl New Delhi disease, the Cassava mosaic disease, and the cotton leaf curl disease.

The Tomato Yellow Leaf Curl Disease

The tomato yellow leaf curl virus (TYLCV) was originally reported from the Jordan Valley in 1920. During the 1960s, the virus was reported from Israel infecting tomatoes. Since then, the tomato yellow leaf curl disease has been reported from all over the world throughout the tropical and sub-tropical regions. The tomato yellow leaf curl disease (TYLCD) is caused by several species of begomoviruses generally termed as Tomato yellow leaf curl virus-like (TYLCV-Like) viruses. There are at least thirteen different species of the TYLCV-like viruses around the globe. However, only the Tomato yellow leaf curl virus-Israel strain (TYLCV-IL) and tomato yellow leaf curl virus-Mild strain (TYLCV-Mld) have the broadest geographical distribution.

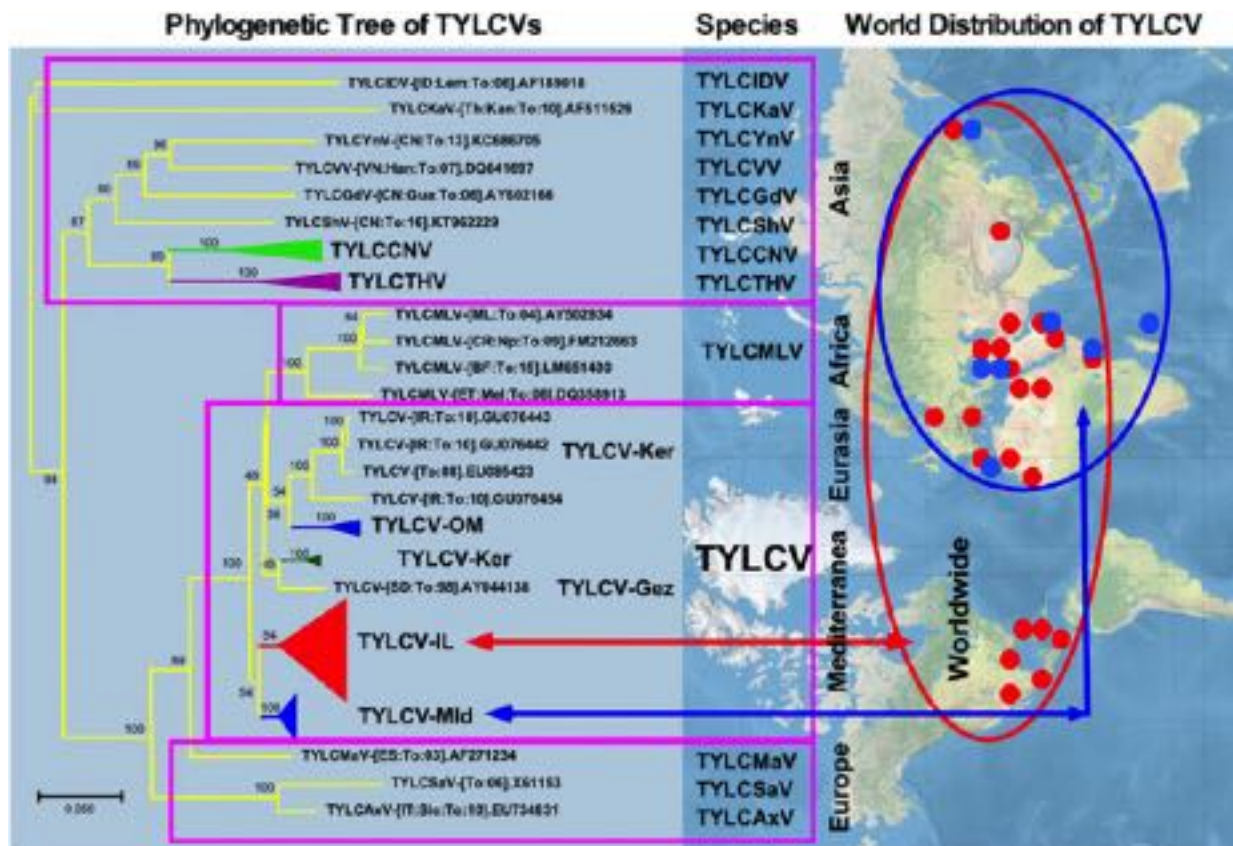


Fig. 1 The global spread of tomato yellow leaf curl virus (TYLCV): Tomato crop is infected by the at least 13 different begomoviruses, which collectively manifest yellow leaf curl disease. The phylogenetic tree represents biodiversity of all 13 begomoviruses. They include, tomato yellow leaf curl virus (TYLCV-Israel, mild, Gezira and Kerman strains), Tomato yellow leaf curl Mali virus (TYLCMLV), tomato yellow leaf curl Malaga virus (TYLCMaV), Tomato yellow leaf curl Sardinia virus (TYLCSaV), Tomato yellow leaf curl Axarquia virus (TYLAXV), Tomato yellow leaf curl Yunnan virus (TYLCYnV), Tomato yellow leaf curl Vietnam virus (TYLCVv), Tomato yellow leaf curl Guangdong virus (TYLCGdV), Tomato yellow leaf curl Shuangbai virus (TYLCShV), Tomato yellow leaf curl Thailand virus (TYLCTHV), Tomato yellow leaf curl China virus (TYLCCNV), Tomato yellow leaf curl Kanchanaburi virus (TYLCKaV) and Tomato yellow leaf curl Indonesia virus (TYLCIDV). The geographical map on left side indicates the spread of TYLCV-Israel and mild strain. Each strain is differentiated by different colors.

Geographical Distribution

TYLCD is reported from all the continents, in all tomato growing areas (Fig. 1). Initially, it was presumed that there were different species of begomoviruses infecting tomatoes in different parts of the world. But later, it was proved that TYLCV-IL strain has a worldwide distribution due to the modern trade between different countries in different continents. To date, the TYLCV-IL strain is found in Mediterranean Region, India, Japan, Australia, China, Cuba, Egypt, France, Greece, Korea, Iran, Italy, Mexico, Morocco, Puerto Rico, Sudan, Tunisia, and USA. The TYLCV-IL strain still continues to spread in the Old World, North and South America. Nearly 30% of the total world tomato production is produced in the region surrounding Israel, Jordan, and Iran therefore it is identified as the center for diversity of TYLCV-IL.

The Disease Symptoms, Host Range and Yield Losses

TYLCD can cause severe economic losses by reducing the tomato production. The disease symptoms are manifested by a stunted growth, upward curling of leaves and yellowing of plants (Fig. 2). During the early stages of infection, the leaf size is drastically reduced and plants remain stunted with no fruit development. In the epidemic conditions, the disease can cause 100% yield losses. Other than tomatoes, the TYLCV has been reported from several other hosts, such as *Solanum nigrum*, peppers, tobacco, common bean, soybean, and different weeds.

Causal Agent and Taxonomy

TYLCD is caused by mono- bipartite begomoviruses in the family *Geminiviridae*. There are several widely distributed tomato-infecting begomoviruses which are both monopartite or bipartite in nature, but TYLCV-IL is a true monopartite begomovirus. In



Fig. 2 Typical disease symptoms of Tomato yellow leaf curl virus on tomato in Jordan.

fact, all the TYLCV-Like viruses are monopartite begomoviruses, with the exception of tomato yellow leaf curl Thailand virus, which is a mono/bipartite virus. In some exceptional cases like the TYLCV-Oman strain and the Tomato yellow leaf curl China virus they are also associated with betasatellites. TYLCV has a typical genome of a monopartite DNA-A component with six different ORFs. TYLCV has evolved through inter and intra species recombination. There are six known strains of TYLCV in the Middle Eastern region, known as, TYLC-Israel, -Mild, -Oman, -Iran, -Kerman, and -Gezira. However, only the Israel strain spread all over the world and has the broadest geographical distribution and host range.

Epidemiology

All the begomoviruses including TYLCV are transmitted through whitefly (*Bemisia tabaci*). The literature suggests that TYLCV can be transmitted through B or Q whitefly biotypes. The whitefly ingests TYLCV when feeding on phloem tissues of infected tomatoes. The tomato leaves and fruit sepals can act as a reservoir of the virus for the plant. It is believed that TYLCV spread through infected fruit or nursery seedlings from the Middle East to the Americas. After the ingestion of TYLCV by whitefly, it can be transferred to the host plant after 8 h, whereas symptoms can appear after 2–4 weeks. TYLCV is an exception among all begomoviruses, because it can be transmitted through seeds in soybean. The viral replication was also observed in soybean seeds. The TYLCV virions can be observed in the eggs and nymphs of whiteflies. However, not all the whitefly biotypes have the equal capacity of *transovarial* transmission of TYLCV. The reasons for the fast spread of TYLCV can be attributed to trade related to agriculture among Middle Eastern countries and the rest of the world. Once, the TYLCV is introduced in an area it can spread very aggressively compared to other viruses in this particular region. The invasive spread of TYLCV can be due to high and diverse whitefly population, availability of susceptible plants and higher recombination rate among strains.

Control

TYLCD is very difficult to control. However, continuous breeding efforts have resulted in the introgression of resistant genes from wild tomato to cultivated tomatoes in the world. Until now, three different genes have been cloned from wild tomato sources. These include, *Ty-1*, *Ty-2*, and *Ty-3*. *Ty-1* and *Ty-3* are alleles of each other and encode an RNA dependent RNA polymerase, while *Ty-2* encodes an NB-LRR gene. Through breeding programs all three genes have been introgressed into cultivated tomato varieties. The *Ty-2* based resistance was overcome by TYLCV, while, *Ty-1* based resistance cannot sustain the high pressure of TYLCV infection. The resistance breakage ability of the virus poses a challenge to the plant breeders to continuously improve the germplasm to control TYLCV. Recently, transgenic plants employing the CRISPR/CAS technology have been developed, targeting Rep and coat protein (CP) genes of TYLCV. The results from transgenic plants are very promising. In the future, the natural and transgene based resistance may be combined to develop more durable resistant varieties.

The Tomato Leaf Curl New Delhi Disease

Tomato leaf curl New Delhi virus (ToLCNDV) is another tomato infecting bipartite begomovirus from the Old World. The disease symptoms include leaf curling, stunted growth, slight yellowing of leaves, and leaf puckering. ToLCNDV is known from in the Indian sub-continent from the last 3 decades. Initially, it was limited to the tomato crop in New Delhi, India, but later it spread to other crops, like potato and cucurbit species. Earlier, the ToLCNDV was reported as a typical bipartite virus, but now it is known to be able to capture several different DNA-B components and betasatellites. ToLCNDV infects more than 40 dicotyledonous species in the world. Currently, two different strains of ToLCNDV (namely, ToLCNDV-India and ToLCNDV-Spain) are dominant in the world.

Geographical Distribution

ToLCNDV infects different crops in Asia, Middle East, North Africa, and Europe. ToLCNDV was first identified from Northern India during 1990s. Later it was reported from Pakistan, Sri Lanka, Bangladesh, Malaysia, Iran, Taiwan, Thailand, and Indonesia. ToLCNDV is a fast spreading virus around the world. Recently, it was reported from Iran and after 2012 it spread westward to different European territories (Fig. 3). In 2012, it heavily infected cucurbits in the greenhouses and other field crops in Spain. Subsequently, it was reported from Tunisia and Italy. The isolates of Spain, Italy, and Tunisia are closely related to each other indicating their common origin. After 2017, the ToLCNDV has been reported from Algeria and Morocco. Right now (2018–2020) it continues to spread in Greece and Portugal.

The Disease Symptoms and Yield Losses

ToLCNDV is a major constraint in tomato production in the Indian subcontinent. Its host range is extended to 43 different hosts. It poses serious threats to vegetable and fibre crops. The tomato plants infected with ToLCNDV exhibit stunted growth, shortening and slight leaf yellowing, puckering and blistering of leaves. Subsequently, the internodes of plants remain shorter with impaired fruit setting, which leads to complete yield loss. In cucurbits, ToLCNDV induces blistering of leaves along with yellow mosaic pattern. The disease symptoms can be more severe with the adoption of betasatellites.

Epidemiology

Similarly to other begomoviruses, the ToLCNDV spread occurs through whitefly *B. tabaci*. Several whitefly species like, Asia-I, Asia-II and Middle East-Asia Minor-I spread the virus in South Asia. While in Spain, the Q1 biotype causes the spread of the virus in tomato, zucchini and melon. The virus-vector interaction studies demonstrated that ToLCNDV is transmitted in a circulative persistent transmission manner. Evidence also suggest that female whiteflies possess higher transmission efficiency. Interestingly, ToLCNDV is often found in mixed infection with other begomoviruses, such as Chili leaf curl virus or Bhindi yellow vein mosaic virus. During routine surveys, ToLCNDV is found with Chili leaf curl betasatellite, thus enhancing the pathogenicity and host range of the virus.

Component Capturing and Exchange of Helper Components

ToLCNDV is among the most unique bipartite viruses, which can co-exist in mixed infection and can exchange its DNA-B component with other begomoviruses of the region. For example, it is very common to find DNA-B of Tomato leaf curl Gujarat virus, Tomato leaf curl Palampur virus, Squash leaf curl China virus and Tomato leaf curl Ranchi virus in combination with the helper component of ToLCNDV DNA-A. It has also been shown that Croton yellow vein mosaic virus, Radish leaf curl virus, Tomato leaf curl Palampur virus, and Tomato leaf curl Gujarat virus can trans-replicate the DNA-B of ToLCNDV. These evidences suggest that ToLCNDV can result in serious agricultural losses due to its flexibility in component capturing.

Control

Fortunately, both breeding and genetic engineering approaches have led to the development of resistance against ToLCNDV in tomato and other important cucurbits. The breeders have developed the resistant lines through crossing of wild relatives of tomato with cultivated species. Just like TYLCV, the *Ty* loci have been exploited to develop resistance against ToLCNDV. The transgenic approaches included the use of AV1, AV2, and AC1 genes to target the ToLCNDV. However, recombination and component capturing can still result in serious outbreaks.

The African Cassava Mosaic Disease

The cassava mosaic geminiviruses (CMGs) in Africa are among the emerging geminiviruses and caused a pandemic in sub-Saharan African countries. The CMGs are known to infect the cassava crop in Africa for more than 100 years, but some of them emerged as major pathogens in the last 30 years. The cassava mosaic disease (CMD) spread in pandemic manner in the early 90 s in Uganda, East Africa, and then spread to other parts of Africa, mostly East and central Africa. More recently it was demonstrated that CMD was imported from the Indian subcontinent to South East Asia (SEA), and a rapid pandemic began about 3 years ago, invading 4 countries already and compromising the second most important production of cassava in the world. The spread of cassava pandemic throughout Africa is mainly due to the man-mediated transfer of infected material and the whitefly vector, while in SEA it is the international regional trade of ingected cuttings that is the driving force for the rapid spread of the disease. This African pandemic was responsible for very severe cassava crop losses and resulted in one of the worst modern famines on this continent. In the last twenty years, both the Cassava brown streak viruses (CBSV and UCBSV; RNA viruses of the genus *Ipomovirus*, family *Potyviridae*) and CMGs had resulted in very low cassava production in East and Southern Africa. The CMGs and CBSVs are spreading from East Africa to central Africa. Interestingly, both the viruses are transmitted through whiteflies that have been super abundant on cassava in the last 3 decades.

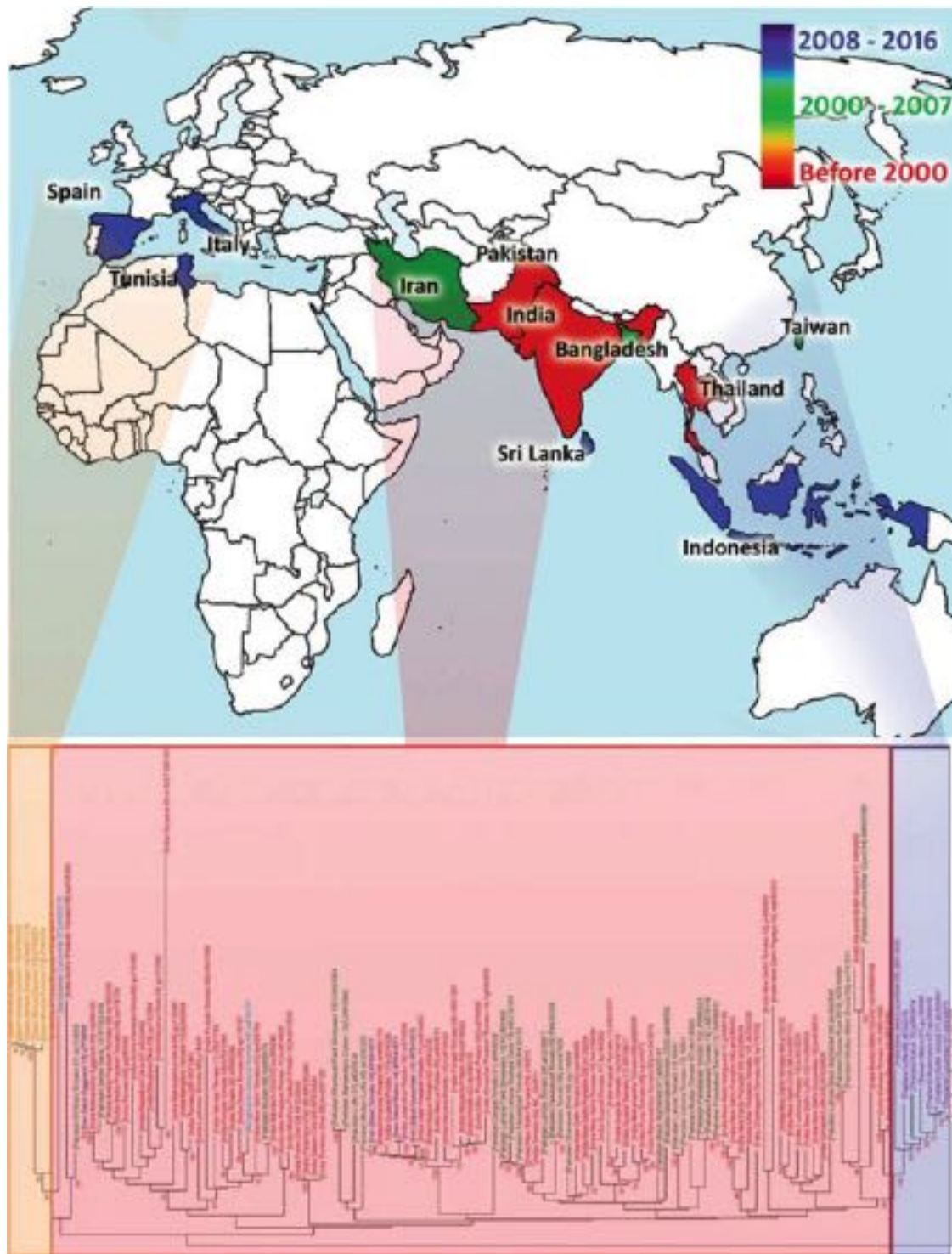


Fig. 3 Geographical distribution and phylogeny of ToLCNDV in the world. The distribution of Tomato leaf curl New Delhi virus (ToLCNDV) overtime. Countries are colored, according to the color index (top right), with respect to the first report of ToLCNDV in the countries. The phylogenetic tree at the bottom is based on ToLCNDV DNA-A sequences, and a spotlight on each cluster indicates the geographical region from which the isolates in that cluster originate. The photograph was previously published in Molecular plant pathology by Zaidi, S.S., Martin, D.P., Amin, I., Farooq, M., Mansoor, S., 2017. Tomato leaf curl New Delhi virus: A widespread bipartite begomovirus in the territory of monopartite begomoviruses. *Molecular Plant Pathology* 18 (7), 901–911. doi:10.1111/mpp.12481.

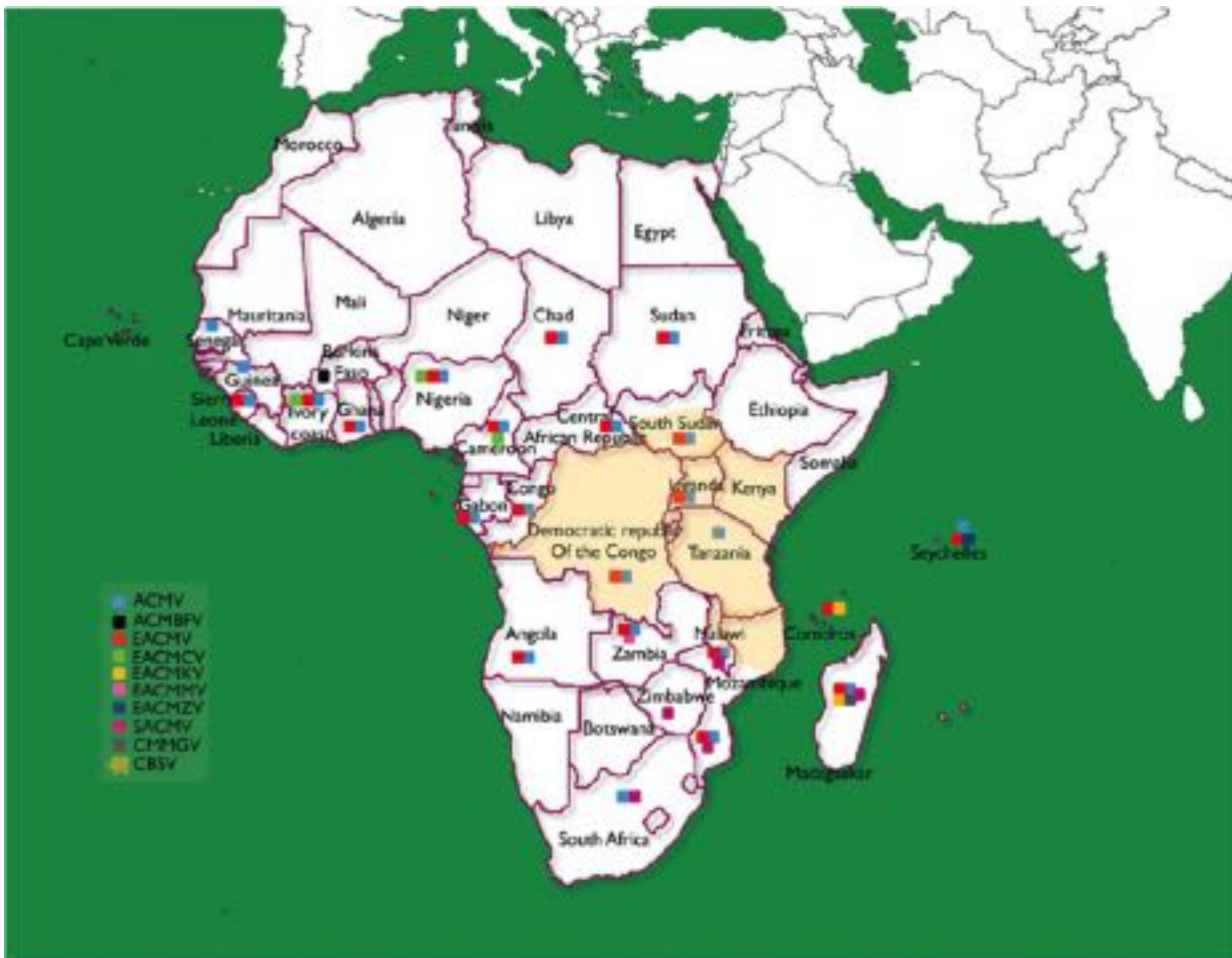


Fig. 4 Geographical distribution of cassava mosaic begomoviruses and cassava brown streak virus in Africa. The territory of each virus is marked with a unique colored box. The countries plagued by CBSV are shaded with yellow color. Cassava crop is infected by *African cassava mosaic virus* (ACMV), *African cassava mosaic Burkina Faso virus* (ACMBFV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Kenya virus* (EACMKV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV), *South African cassava mosaic virus* (SACMV), *Cassava mosaic Madagascar virus* (CMMGV), and *Cassava brown streak virus* (CBSV).

Geographical Distribution

CMD was observed in East Africa (Tanzania) since, 1894, several decades before the emergence of Cassava brown streak disease (CBSD- caused by CBSV and UCBSV). In 1994, the CMD spread from Uganda to neighboring countries surrounding the lake Victoria. The disease swept through Central Africa to West African countries (Congo, Gabon and south of Cameroon) (Fig. 4). The earlier sequencing data revealed the presence of the African cassava mosaic virus (ACMV) in the infected cassava plants. But the 1990s epidemic revealed the presence of a more severe strain of a different species, namely East African cassava mosaic virus Ugandan strain (EACMV-UG). The disease is currently present in more than 20 countries (Fig. 4). The East and West Africa are mainly dominated by ACMV and EACMV-UG strains respectively. While Southern and South East Africa are mainly dominated by several other CMGs and also by the two CBSD viruses. Although, the presence of ACMV and EACMV dominates in the whole African continent, the Island of Madagascar contains the largest number of CMGs (Fig. 4). In 2007, the non-cultivated cotton samples collected from Pakistan showed the presence of DNA-A component of ACMV. The DNA-B component of ACMV is still not discovered from Pakistan. The discovery of ACMV from non-cultivated cotton is surprising, not only because of the geographical distance, but also due to the host change.

The Disease Symptoms and Yield Losses

The CMGs can cause very severe damages to the cassava crop with annual losses of over US\$1 billion losses. During the epidemic period, the CMD resulted in a reduction of about 90% of the cassava production in Uganda which induced a severe famine and



Fig. 5 The disease symptoms of cassava plants infected with CMD. The health plant leaves do not show any mosaic pattern (a), The panel b represents the symptoms by EACMV and panel c represents the infection by ACMV. The synergism between both viruses results in candlestick syndrome in panel d.

thousands of people died due to hunger. The disease symptoms due to various strains can be easily identified. The symptoms due to EACMV-UG are more severe than ACMV (Fig. 5). The EACMV-UG infection results in reduction of leaf surface area, yellowing of lamina and downward curling of the leaves. The entire plant looks yellow and stunted in growth compared to healthy plants. The cassava plants infected with CBSD also show yellow symptoms on the old leaves but they never show the severity of EACMV-UG. Furthermore, the plants infected with CBSD have dry-brown necrosis on cassava tubers and are inedible. In the 90 s in Uganda, the cassava plants showed dual infection of ACMV and EACMV-UG with an extreme reduction of the growth of the plants and no root production. In the case of dual infected plants, the plant fails to produce any proper foliage; the stem turns into yellow-purple color and the infected plants showed a “candle stick syndrome”, and no root production whatsoever. In addition, cassava is propagated through stem cuttings and dually infected plants cannot be used for further propagation. The farmers in this part of the world, at that period in time, ran out of food and out of planting material!

Classification of Cassava Mosaic Begomoviruses

The CMD is caused by bipartite begomoviruses. Unlike TYLCV, the CMDs have two circular genomes (bipartite) i.e., DNA-A and DNA-B. As recombination is the driving force for geminivirus evolution, therefore CMDs are no exceptions. Indeed, the EACMV-UG strain is a recombinant for the CP region between EACMV and ACMV. The sequence analysis revealed that EACMV-UG strain gained ~400–550 nt from the CP of ACMV through recombination. Over the years, the molecular sequencing revealed nine different species of begomoviruses occurring in single or mixed infection. These species includes, *African cassava mosaic virus* (ACMV), *African cassava mosaic Burkina Faso virus* (ACMBFV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Kenya virus* (EACMKV), *South African cassava mosaic virus* (SACMV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV), and *cassava mosaic Madagascar virus* (CMMGV). The co-infection of different CMDs can lead to synergism and new recombinants are produced with variable virulence can arise.

Epidemiology

The literature suggests that CMGs mainly spread through infected cassava cuttings, however, whitefly (*Bemisia tabaci*) also played an important role in spreading the virus to other field crops. The transmission of CMGs occurs through whitefly in a circulative but non-propagative manner. It requires at least 8 h of circulation inside the whitefly before transmission into plants can occur. However, once the whitefly acquires the virus, it can remain viruliferous all its life, thus increasing the chance of viral spread for a long period of time and a longer distance. The East African cassava pandemic could be correlated to the synergism between ACMV and EACMV (Fig. 5) along with the explosion of “super abundant” whitefly populations. Recombination between different CMGs is extremely prevalent although it is not known how important this factor is in the disease spread. Recent research demonstrated that there are several whitefly biotypes infesting cassava in Africa and that these biotypes are geographically distributed.

Recent Outbreak of CMD in South East Asia

In 2016, the first report of the CMD, was published, describing the unequivocal identification of the Sri Lankan cassava mosaic virus (SLCMV) as responsible for the disease symptoms in the north east of Cambodia. Since that date, several surveys have been done in Cambodia, China, Vietnam, and Thailand, and as of January 2019, the disease was present in 10 provinces of Cambodia, 14 in Vietnam, and 7 provinces in Thailand (Fig. 6). There is no CMD case yet reported in Laos or Myanmar.

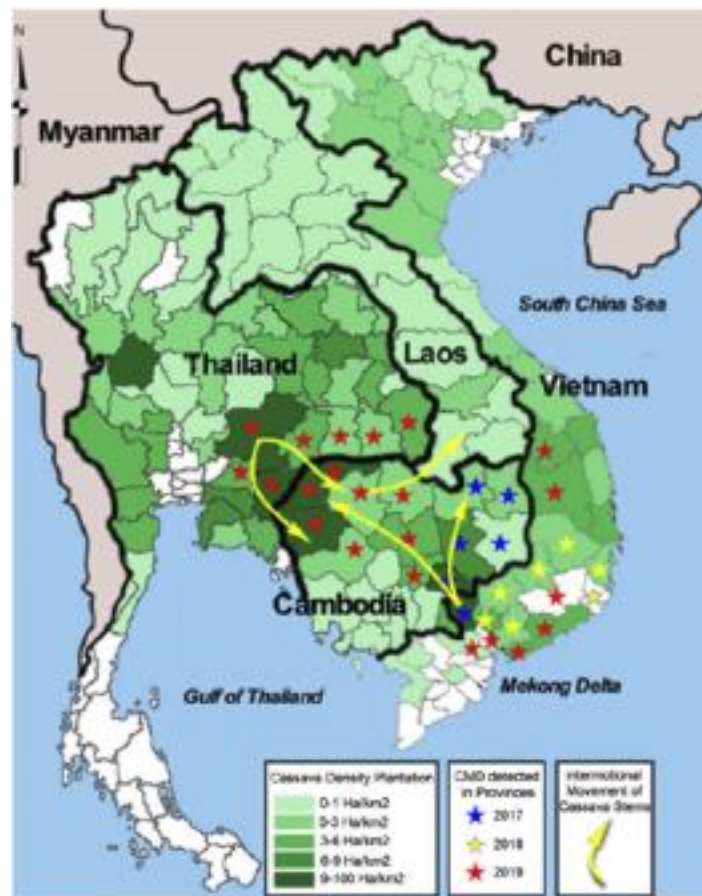


Fig. 6 Geographical distribution of cotton leaf curl geminiviruses. Cotton leaf curl disease is present in India, China, Pakistan and Philippines. The disease was first known from Pakistan and then spread to India, China and Philippines. The Chinese and Philippine's isolates are identical to each other. The dashed arrows represent direction of movement for cotton leaf curl Multan virus.

The presence of the same virus has also been reported from several locations of cassava collections in China. In 3 years, the disease reports extended from a single plantation in Cambodia to an estimated 44% of the surfaces cultivated in the whole region. The disease is transmitted in two ways: by the natural whitefly vector (*B. tabaci*, Aleyrodidae) and by the cassava cuttings, there is no transmission through the roots or the seeds. There is a very active cassava stake trade between several provinces of Cambodia and between Cambodia and Vietnam or Thailand. Although there is clearly insect transmission occurring everywhere, it is believed that currently the disease is mostly spread through cuttings. Considering the importance of cassava in the region (>55 million tons/year and >\$10 billion business) urgent action is needed to stop the spread and put CMD under control. To this effect the Global Cassava Partnership for the 21st Century - GCP21 and CIAT, organized a Regional Workshop in September 18–20, 2018, in Phnom Penh, Cambodia establishing a unique regional plan of control of CMD. This control plan lists 5 different topics; 1- Regional & National Coordination Mechanism, 2- CMD Awareness & Mitigation Extension, 3- Surveillance and Diagnostic, 4- Propagation and distribution of disease-free planting material, and 5- Varietal testing and breeding for CMD resistance. This recent emergence of CMD exemplifies the major role of international movement of infected material complemented by biological factors.

Control

In general terms, CMD can be controlled by three different ways. (1) Planting disease-free material to minimize the spread of infected material and crop losses, (2) Control of whitefly populations to delay the infection process till later growth stages, (3) Eradication of infected material to avoid the spread of infected plants through cuttings. Although CMD is known for more than a century, still the disease causes very significant losses in Africa and India. The earlier cassava breeding programs in East Africa led to the development of highly resistant germplasm against CMGs (multigenic type of resistance CMD1). Subsequently breeding programs in IITA in West Africa produced numerous CMD resistant cassava varieties between 1970 and 1990. In the 90 s another source of resistance (CMD2) of monogenic nature was identified in cassava landraces in West Africa, conferring extremely high levels of durable resistance to CMGs. SNP markers have been developed for CMD2 allowing its easy transfer to many more

genotypes of cassava. The introduction of resistant cassava germplasm in Uganda resulted in restoration of cassava yield in the late 90 s. However, due to their better processing and organoleptic properties, the farmers often prefer local susceptible landraces.

The Cotton Leaf Curl Disease in the Indian Sub-Continent

Geographical Distribution

The Cotton leaf curl disease (CLCuD) was first observed in the Multan region of Pakistan during late 1960s. The disease was limited to the surroundings of Multan until 1987. After 1987, the disease spread to the surrounding areas and appeared in epidemic form between 1992 and 1997. During early 2000, the disease spread to India, where it caused serious damage to cotton crops. Today, the disease is prevalent in Southeastern China, Philippines, Pakistan, and India. In Philippines, CLCuMuV infects hibiscus plants instead of cotton. Although there is a huge diversity among cotton leaf curl geminiviruses (CGs), the most devastating viruses are Cotton leaf curl Multan virus (CLCuMuV) and Cotton leaf curl Gezira virus (CLCuGeV-African origin). The CLCuMuV is prevalent in Pakistan, India, China, and Philippines, while CLCuGeV is present all over the Africa, wherever cotton and okra crops are grown. Indeed, CLCuGeV is now also present in Southern parts of Pakistan; however, its infection is limited on cotton crop which is dominated by Cotton leaf curl Multan virus.

The Disease Symptoms and Yield Losses

Like all other begomoviruses, the CLCuD is transmitted by whiteflies. So far there is no evidence for the seed transmission of the CLCuD. Upward or downward leaf curling, vein thickening, leaf enations and stunted growth are the main symptoms of infected plants (Fig. 7). Once the plant acquires CLCuD through whitefly inoculation, the infection is life long for the host. The infected plants remain stunted and dark green in color. Very few flowers are produced and they cannot produce fertile seeds. Over the past three decades, Pakistan had lost billions of dollars due to this disease. During the epidemic (1992–1997) Pakistan lost US\$5 billions. With the introduction of resistant varieties after the year 2000 the disease was fairly well controlled. However, a new out-break of the Burewala strain during 2004 resulted in the resistance breakdown. It was hypothesized that the Burewala strain is the most dominant virus on cotton field in Pakistan and India, however recent data showed that the Multan strain is re-emerging as a major pathogen on cotton.

Causal Agent and Disease Classification

The CLCuD is mainly caused by satellite associated monopartite begomoviruses. These satellites are known as betasatellites and alphasatellites. Both the satellites are approximately half the size of viral component and are encapsidated by the viral CP coded by the helper virus. The prevalent betasatellite is known as Cotton leaf curl Multan betasatellite (CLCuMuB). The DNA-A component of Cotton leaf curl geminiviruses alone cannot produce symptoms, until betasatellites are added. The CLCuMuB encodes a pathogenicity determinant protein. The second category of satellites associated with cotton geminiviruses are alphasatellites. They encode an alpha-Rep responsible for replication of alphasatellites and which is also a suppressor of gene silencing of the host, the natural virus defense mechanism of host. The origin of alphasatellites is considered to be from the family *Nanoviridae*. Apparently, the alphasatellites do not play an important role in the etiology of CLCuD. However, their relatively large diversity suggests that they have an important evolutionary role in disease epidemics. CLCuD is mainly caused by monopartite begomoviruses, but recently some bipartite viruses have also been reported from Pakistan. According to literature, there are at least 14 different begomovirus and one mastrevirus species, which infect the cotton crop in the Old World. CLCuMuV, CLCuGeV, *Cotton leaf curl Kokhran virus* (CLCuKoV), *Papaya leaf curl virus* (PaLCuV), and *Cotton leaf curl Alabad virus* (CLCuAlV) are frequently associated with CLCuD. However, *African cassava mosaic virus* (ACMV), *Okra leaf curl enation virus* (OLCEnV), *Tomato leaf curl New Delhi virus*



Fig. 7 Typical disease symptoms of cotton leaf curl disease in Pakistan. The healthy cotton plants have straight leaves with no curling (A). The infected plants have upward or downward leaf curling. Panel-B presents cotton infected by cotton leaf curl Multan virus, cotton leaf curl Multan betasatellite and *Gossypium darwinii* symptomless alphasatellite. The abaxial side of the infected leaf shows vein thickening and leaf like enations (panel-C).

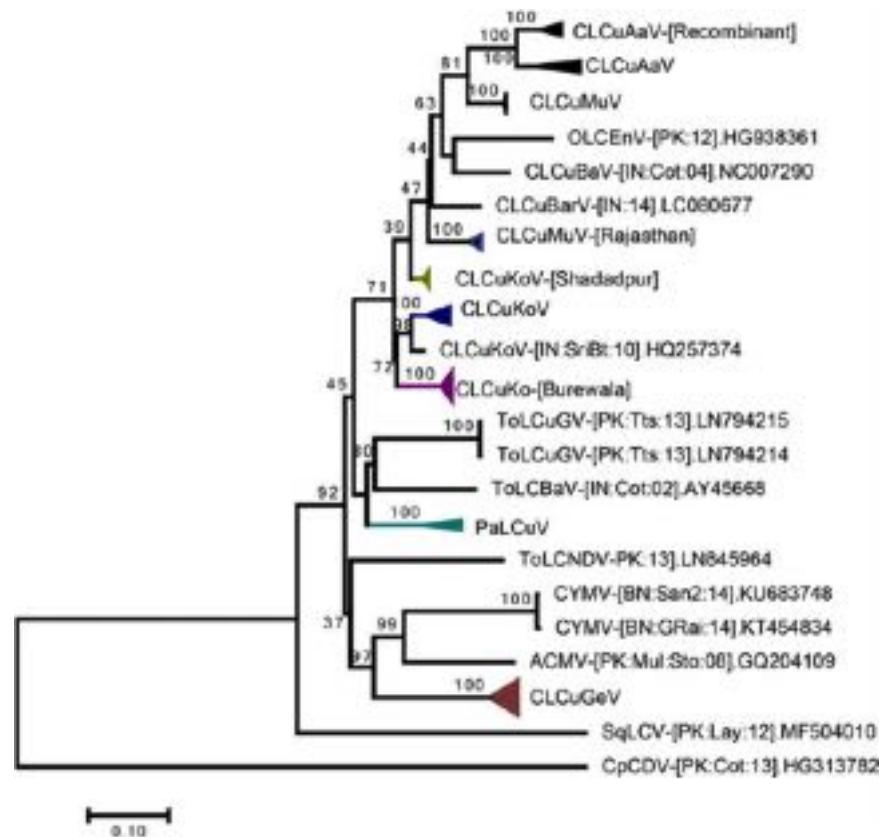


Fig. 8 Phylogenetic tree of cotton leaf curl geminiviruses. Cotton plant has been shown infected by 14 different begomoviruses species and one mastrevirus species. These species include, *Cotton leaf curl Gezira virus* (CLCuGeV), *African cassava mosaic virus* (ACMV), *Tomato leaf curl New Delhi virus* (ToLCNDV), *Tomato leaf curl Bangalore virus* (ToLCBaV), *Papaya leaf curl virus* (PaLCV), *Okra leaf curl enation virus* (OLCEnV), *Cotton leaf curl Bangalore virus* (CLCuBaV), *Cotton leaf curl Barasat virus* (CLCuBaV), *Cotton leaf curl Kokhran virus* (CLCuKoV), *Cotton leaf curl Multan virus* (CLCuMuV), *Cotton leaf curl Alabad virus* (CLCuAaV), *Tomato leaf curl Gujarat virus* (ToLCuGuV) *Cotton yellow mosaic virus*, *Squash leaf curl virus* (SqLCV) and a mastrevirus, *Chickpea chlorotic dwarf virus* (CpCDV). Shadadpur, Burewala and Rajasthan strains are recombinants of CLCuMuV and CLCuKoV.

(ToLCNDV), *Tomato leaf curl Gujarat virus* (ToLCuGuV), *Squash leaf curl virus* (SqLCV), *Cotton yellow mosaic virus* (CYMV), *Cotton leaf curl Bangalore virus* (CLCuBaV), *Cotton leaf curl Barasat virus* (CLCuBV), *Tomato leaf curl Bangalore virus* (ToLCBaV), and *Chickpea chlorotic dwarf virus* (CpCDV- a mastrevirus) are rarely found on cotton crop (Fig. 8).

Epidemiology

Upland cotton (*Gossypium hirsutum*) was introduced in Pakistan by the East India Company from Southern Mexico. The CLCuD is not prevalent in Southern Mexico. The disease is also not seed transmissible, therefore it can be presumed that viruses causing leaf curl disease in cotton in Pakistan shifted from a local plant host to cotton. The major spread of the disease occurred when the exotic germplasm was cultivated during the late 1980s in Pakistan. Indeed, the variety S12 (imported from Texas) proved to be hyper-susceptible to the CLCuD and resulted in the explosion of the disease in a very short period of time. Even after removing the variety S12 from the field, the CLCuD epidemic continued actively. The S12 variety acted as a trigger of the epidemic but it is whiteflies that spread the disease in the country. Due to year around presence of alternative hosts for cotton leaf curl geminiviruses; the threat of CLCuD remains permanent in Pakistan and India. As a matter of fact the begomoviruses from other crops like tomato, cassava, and chickpea are also invading cotton crop. After the year 2000 a recombinant Burewala strain dominated the cotton crop in Pakistan and India. But recently CLCuMuV has re-emerged on cotton in Pakistan.

Control

CLCuD can only be controlled through the use of resistant varieties and the control of whiteflies. Recently, virus resistant cotton varieties obtained through conventional breeding were introduced. It was an immediate success but more recently the resistance was broken down by CLCuMV strains resulting from recombination with other local geminiviruses. The recombinant strains, Burewala, Shadadpur and Rajasthan played an important role in resistance breakdown. Efforts have been made to control the

CLCuD through RNAi or CRISPR/CAS9 but to date there is still no data to support the use of transgene mediated resistance in cotton. The conventional breeding still remains the first and only option in controlling the CLCuD.

Conclusion

This article is documenting the emergence of 4 major geminivirus diseases in the world in the last 30 years, namely the TYLCD and ToLCNDV on tomato, the CMD on cassava and the CLCuD on Cotton. All three examples illustrate the importance of human intervention to trigger a new epidemic, but evidently are also showing the great importance of some biological factors; the greater effectiveness of TYLCV to infect tomatoes, the capacity of ToLCNDV to capture other components and satellites, the combination of two cassava geminiviruses to synergize, and the role of a hyper susceptible cotton variety. In addition, it is also showing the major role of whiteflies in each example to carry on the epidemic to vast surfaces and for many years.

The recent global spread of ToLCNDV on various hosts is also an alarming situation. The spread of ToLCNDV and TYLCV is a global concern and concrete efforts must be done to save crops from their worldwide epidemics.

Another important concern for the world's agriculture is the Squash leaf curl virus (SLCV) spread from California to Middle Eastern countries. If the other 4 examples in this article are all Old World geminiviruses that managed to invade the rest of the world, SLCV is a New World geminivirus that successfully invaded the Middle East region. Indeed, evidences of SLCV infection on cotton also exist in Pakistan. However, it still needs authenticated experiments to prove its level of infection on cotton. The emergence of such viruses is also a future concern.

Currently it is unclear why some strains spread more rapidly than others at the global level. For example there are several strains of TYLCV, but TYLCV-Mld is more wide spread than some others. Similarly, the Uganda strain of the East African cassava mosaic virus (EACMV-UG) in Africa and the cotton leaf curl Multan virus (CLCuMuV) and its betasatellite in Asia are more wide spread than other strains. The world spread of TYLCV can be attributed to trade of tomato seedlings from the Middle East to different parts of the world. Once, the virus is introduced in a new environment, the local whiteflies can quickly acquire and spread the virus to other hosts. The CMD spread in Uganda which was triggered by the synergism between two geminiviruses and the presence of a hyper abundant whitefly, is also attributed to human migration in Africa (Rwanda in the 90 s), carrying infected cassava cuttings for further propagation over a high mountain range. The recent CMD outbreak in SEA was initiated by the import of SLCMV from the Indian subcontinent, and hugely distributed by the clonal propagation in the region. If the initial pandemic of CLCuD in Pakistan was clearly triggered by the import of a super susceptible cotton variety from the US, local whiteflies then took the relay to spread the disease in the sun-continent and to many local hosts. However, the spread of CLCuD from Pakistan to China and Philippines is more intriguing. It is plausible that Pakistan and India sharing borders, therefore whiteflies could easily transfer the disease to the neighboring countries. However, there is a mountainous region between Pakistan and China and crossing of whiteflies is not possible. The transfer of infected ornamental plants could be a reason for the spread of CLCuMuV from Pakistan to China and Philippines. The Multan strain in Philippines is identical to the one isolated in China, therefore it can be hypothesized that CLCuMuV spread from China to Philippines.

In all these examples it is clear that humans, in combination with other biological factors, have played a vital role in the spread of highly pathogenic geminiviruses. In the future, such epidemics could be avoided by more strict quarantines measures and immediate proper agricultural policy control measures.

Further Reading

- Chen, W., Wosula, E.N., Hasegawa, D.K., *et al.*, 2019. Genome of the African cassava whitefly *Bemisia tabaci* and distribution and genetic diversity of cassava-colonizing whiteflies in Africa. *Insect Biochemistry and Molecular Biology* 110, 112–120.
- Jacobson, A.L., Duffy, S., Sseruwagi, P., 2018. Whitefly-transmitted viruses threatening cassava production in Africa. *Current Opinion in Virology* 33, 167–176.
- Lefevre, P., Martin, D.P., Harkins, G., *et al.*, 2010. The spread of tomato yellow leaf curl virus from the Middle East to the world. *PLoS Pathogens* 6, e1001164.
- Nawaz-ul-Rehman, M.S., Briddon, R.W., Fauquet, C.M., 2012. A melting pot of Old World begomoviruses and their satellites infecting a collection of *Gossypium* species in Pakistan. *PLoS One* 7, e40050.
- Saleem, H., Nahid, N., Shakir, S., *et al.*, 2016. Diversity, mutation and recombination analysis of cotton leaf curl geminiviruses. *PLoS One* 11, e0151161.
- Varsani, A., Roumagnac, P., Fuchs, M., *et al.*, 2017. Capulavirus and Grablovirus: Two new genera in the family Geminiviridae. *Archives of Virology* 162, 1819–1831.
- Zaidi, S.S., Martin, D.P., Amin, I., Farooq, M., Mansoor, S., 2017. Tomato leaf curl New Delhi virus: A widespread bipartite begomovirus in the territory of monopartite begomoviruses. *Molecular Plant Pathology* 18 (7), 901–911. doi:10.1111/mpp.12481.

Movement of Viruses in Plants

Manfred Heinlein, IBMP-CNRS, University of Strasbourg, Strasbourg, France

© 2021 Elsevier Ltd. All rights reserved.

This is an update of P.A. Harries, R.S. Nelson, Movement of Viruses in Plants, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00728-7.

Glossary

5'CAP The five-prime cap is a modified nucleotide at the 5' end of RNA molecules. It prevents RNA degradation by exonucleases and promotes RNA translation.

Actin filaments (microfilaments) A component of the cytoskeleton formed from polymerized G-actin monomers.

Chaperones Cellular proteins that assist in the correct folding or unfolding of proteins and in the assembly or disassembly of protein complexes.

COPI/COPII Cytoplasmic protein complexes (coatamers) that coat vesicles transporting proteins from ER to Golgi (COPII), and from Golgi to the ER (COPI).

Endosomes Membrane-bound compartments that function in the sorting and delivery of material internalized from the cell surface or destined for transport from the Golgi to the lysosome or vacuole.

ER exit sites Specialized zones of the ER engaged in COPII-mediated exit of newly synthesized proteins from the ER into vesicles for transport from ER to Golgi.

ESCRT The “endosomal sorting complexes required for transport (ESCRT)” machinery consists of cytoplasmic protein complexes involved in membrane remodeling.

GPI anchor Short glycolipid (glycosylphosphatidylinositol or glycosylphosphatidylinositol) fused to the C-terminus of a protein during posttranslational modification and tethering the protein to the outer leaflet of the PM.

Microtubules A component of the cytoskeleton composed of hollow tubes assembled from α -tubulin/ β -tubulin dimers.

Multivesicular bodies (MVB) Single membrane-bound organelles with intraluminal vesicles that mediate protein cargo transport from the trans-Golgi network (TGN) to vacuoles for degradation.

Pattern-triggered immunity (PTI) Innate immunity based on the recognition of conserved microbial signatures (patterns) by pathogen recognition receptors at the PM or in the cytoplasm.

Reticulons Proteins residing at the ER and playing a role in promoting membrane curvature.

Introduction

In order for a plant virus to infect its host systemically, it must be capable of hijacking the host's cellular machinery to replicate and move from the initially infected cell. Plant viruses require wounding, usually by insect or fungal vectors or mechanical abrasion, for an infection to begin. Once inside a cell, virus particles have to disassemble to release the viral genome which, then, can be transcribed (DNA viruses) and translated (DNA and RNA viruses) to initiate its replication and movement. Some of the viral products are required for virus movement and interact with host factors (proteins or membranes) to carry out this function. Virus movement in plants is tightly linked to virus replication and can be broken down into three distinct steps: (1) intracellular movement, (2) intercellular movement and (3) systemic movement. Intracellular movement refers to processes that allow the virus to reach plasmodesmata (PD), tiny (50 μ m in diameter) channels in the plant cell walls that connect neighboring cells, whereas intercellular movement refers to the process of virus movement through PD allowing the spread of infection between cells. By intercellular movement from-cell-to-cell and by reiteration of infection and viral genome replication and the targeting of PD in each cell, the virus finally reaches cells of the vascular system (Fig. 1). Once the virus has entered the sugar-transporting phloem sieve elements, the virus moves systemically with the flow of photoassimilates, which directs the infection efficiently to the young, sugar-importing sink tissues before also source tissues are slowly infected. Upon delivery by the phloem to a tissue distant from the original infection site, the virus exits the vasculature and resumes cell-to-cell movement via PD in the new tissue. Although virus movement generally occurs by viral exploitation of the symplasmic intercellular communication network established by PD and the connected phloem (Fig. 1), it is important to realize that some viruses, like Turnip mosaic virus (TuMV) and Rice yellow mottle virus (RYMV), also exploit the water-transporting xylem vessels, but this will not be further elaborated here.

The mechanism of virus movement in plants has been studied with a wide range of virus genera, including, but not limited to, tobamoviruses, potexviruses, hordeiviruses, comoviruses, nepoviruses, potyviruses, tombusviruses, tospoviruses, and geminiviruses. Although viruses generally move between cells with support of their Movement Proteins (MPs) (Fig. 2), the cellular pathways and host factors (proteins and membranes) used by viruses for movement show strong differences (Fig. 3). Several studies using GFP and other fluorescent proteins as visible markers have provided important insights into the dynamic changes in the subcellular distribution, accumulation, and movement patterns of viral and host proteins and helped to decipher the different processes occurring during virus infection in plant cells. The resulting observations are summarized in numerous review articles and highlight the roles of membranes and the associated cytoskeleton during virus replication and movement. This article will describe the three steps of virus movement in plants and will focus on model viruses within genera that provide the most

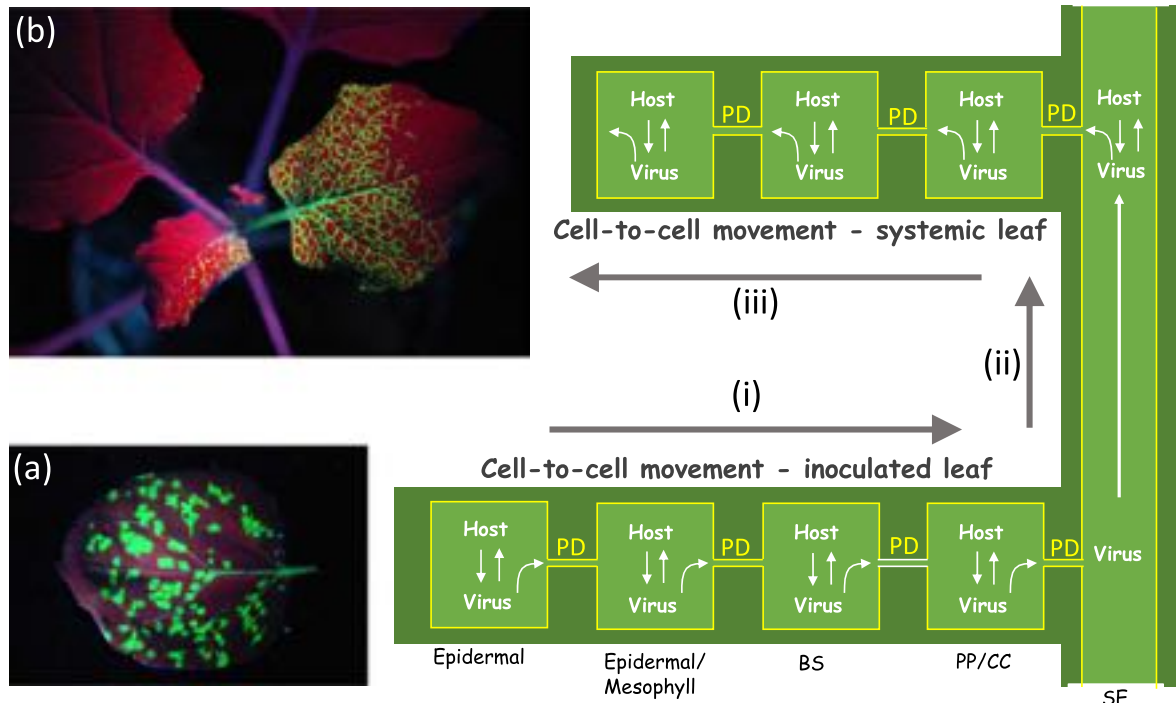


Fig. 1 Local and systemic virus movement in infected plants. In the inoculated leaf, viruses move cell-to cell (i) and expand the local sites of infection (a) until they reach the phloem sieve elements (SE) in the leaf vein vasculature by which they move with the flow of photoassimilates over long distance to sink tissues (ii). Upon unloading from the SE, viruses move again cell-to-cell (iii) to spread infection throughout the newly infected systemic leaves (b). Successful infection depends on compatible interactions between the virus and its host (arrows) to allow the virus to replicate its genome and to target PD (curved arrow, intracellular movement) in each cell it enters. (a) Local sites of infection by TMV tagged with green fluorescent protein (TMV-GFP). (b) a systemic leaf showing TMV-GFP as it spreads infection from the veins into the surrounding cells. BS, bundle sheath; PP/CC, phloem parenchyma/companion cell.

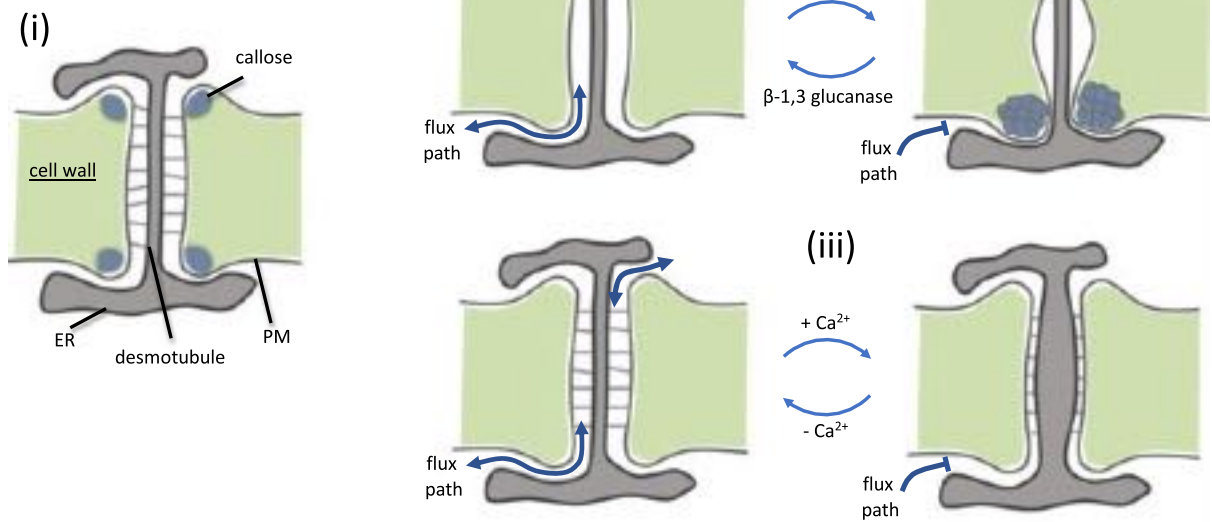
information on the subject. This article also considers the virus-triggered defense responses of the host that the virus must overcome to ensure its efficient cell-to-cell spread.

Intracellular Movement

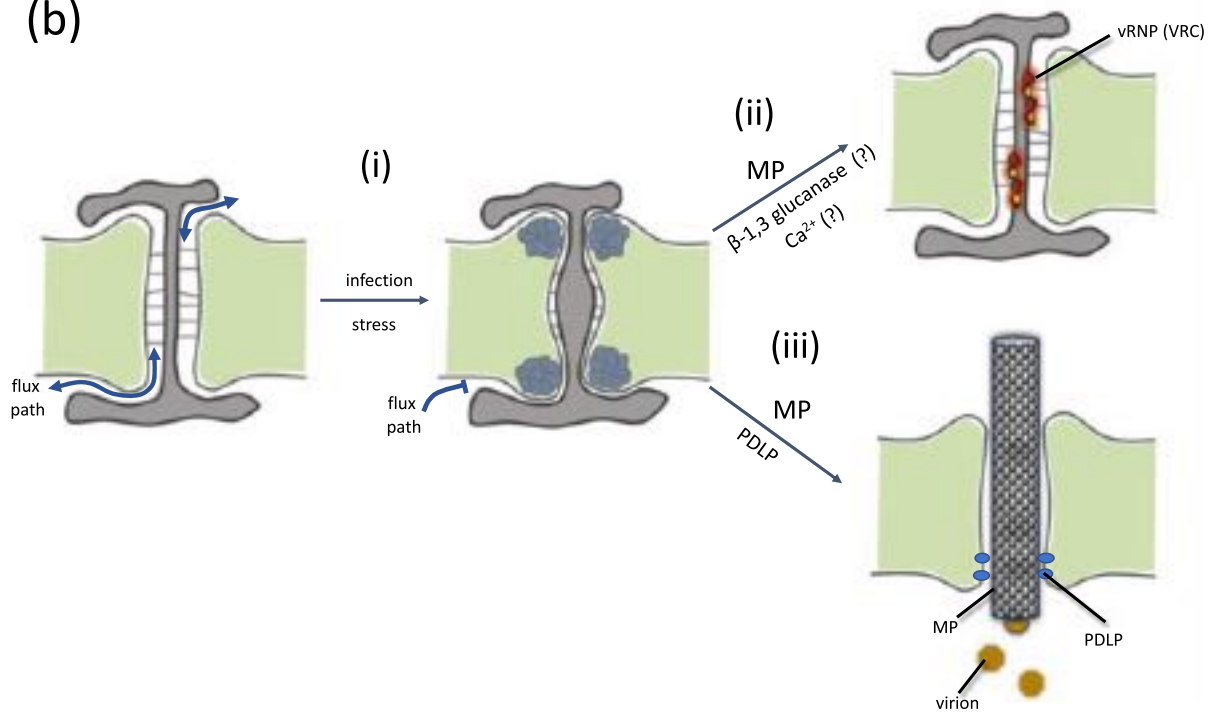
Although the process of virus movement can be complex and requires support by the coordinated activity of several virus- and host-encoded proteins, many viruses achieve their movement with the help of virus-encoded 'movement proteins (MP)'. Pioneering studies with the MP of TMV and later with other MPs demonstrated that these proteins bind single-stranded RNA and modify the size-exclusion limit (SEL) of PD (Fig. 2). Moreover, upon introduction into cells by microinjection, or by transient expression using microparticle bombardment or agroinfiltration, MPs interact with PD, move between cells and may co-transport co-injected nucleic acids, thus indicating that their function is independent of infection. The ability of MP to alter PD aperture is usually correlated with the accumulation of the protein at PD. Nevertheless, plants that express the MP and are able to complement the movement function of movement-deficient TMV mutants do not show obvious growth defects, suggesting the existence of mechanisms that tightly control the ability of MP to modify PD aperture and thus prevent the continuous trafficking of signaling macromolecules between cells. Microinjection studies indeed demonstrated that the MP of TMV modifies the SEL of PD only transiently in cells at the spreading front of infection, thus confirming that MP is tightly regulated. The MP of TMV has several amino acids that are phosphorylated *in vivo* and that may play a role in regulating MP functions. Consistent with this type of posttranslational control, PD are associated with several kinases and a specific kinase, plasmodesmal-associated kinase (PAPK), was shown to specifically phosphorylate the C-terminal phosphorylation sites of MP *in vitro* and also to be active on a subset of other MPs and non-cell-autonomous plant transcription factors.

The ability of MPs to bind nucleic acids is sequence-independent, which raises the important question of how these proteins find their viral RNA cargo in infected cells. Since the MPs are produced by translation of the viral genome, they may simply attach co-translationally to viral genomes in their vicinity. Consistently, the MPs of several RNA viruses, including Tobacco mosaic virus (TMV), Potato virus X (PVX), Brome mosaic virus (BMV) or Red clover necrotic mosaic virus (RCNMV) were shown to colocalize with membrane-associated inclusions that harbor their viral replication complexes (VRC). VRCs can be highly organized structures involving membranes and cytoskeletal elements, and in which the binding of MP to viral RNA involves specific mechanisms, for

(a)



(b)



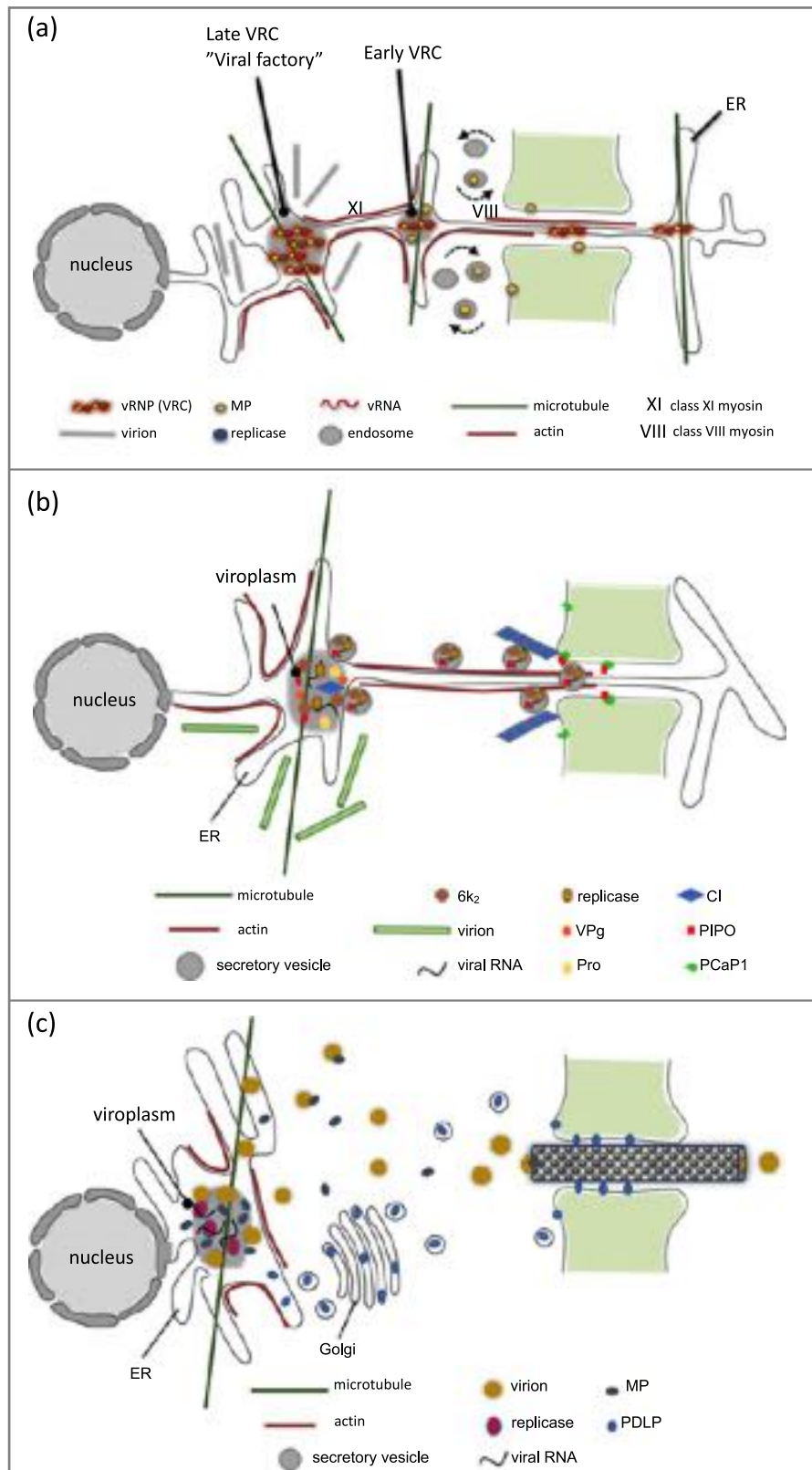
example interaction with other viral proteins, as exemplified by VRCs formed by RCNMV. Another important question is whether MPs transport viral genomes as part of subcomplexes that are derived from anchored VRCs or whether infection moves between cells in the form of mobile VRCs. Studies with TMV and other viruses have shown that VRCs develop at multiple sites in the infected cell. Moreover, VRCs can be stationary or mobile, and their subcellular formation and movement is associated with the spread of the virus. Studies with PVX have shown that VRCs can associate with PD and load replicated viral genome into PD for movement, while observations with TMV suggested that TMV spread infection by movement through PD as a wholly intact VRC. Thus, it appears that viruses develop VRCs for replication as well as for movement and that virus replication and virus movement are tightly linked processes. Movement to PD or even between cells as a VRC ensures the vicinity of MPs to the viral genome and may explain the apparent lack of evolution of any sequence specificity in the binding of MPs to nucleic acids.

Thus, how do viruses develop their VRCs and transport them to PD? TMV infection is initiated by attaching the viral RNA that entered the cell with its 5'CAP to the ER membrane. The RNA is then replicated before the virus can spread its RNA genome into adjacent cells. Time-resolved *in vivo* studies showed that TMV replication and, thus, the formation of VRCs occurs at distinct sites of the cortical ER. While some of the VRCs remain anchored at these sites and are destined for the development of ER-associated, virion-producing viral factories, other VRCs detach from their anchorage sites soon after formation and act as mobile RNA particles for intra- and intercellular transport (Fig. 3(a)). The formation of VRCs at distinct ER sites probably reflects the presence of localized mechanisms by which the virus recruits host factors and membranes for VRC formation. *In vivo* studies have shown that the membrane-associated mechanisms of VRC formation and movement are tightly associated with the cytoskeleton, thus reflecting the viral exploitation of the microtubule- and actin-dependent functions in plant endomembrane organization. The distinct sites of VRC formation by TMV overlap with sites of immobile ER at which the cortical ER-actin network intersects with microtubules. The literature suggests that such sites are attached to the plasma membrane (PM), thus potentially constituting ER-PM contact sites. In providing vertical vicinity between the PM and the ER and lateral vicinity between the ER and the associated cytoskeleton these sites may function as a cortical nexus at which actin- and microtubule-driven transport pathways for the trafficking of membranes, protein complexes and organelles converge. TMV, and potentially other viruses, may have evolved to target these junctions as sites at which membranes and host factors required for VRC formation can be recruited, concentrated, and organized. The MP of TMV has strong affinity for binding to the surfaces of both membranes and microtubules, and also interacts with microtubule-associated proteins such as tubulin dimers, γ -tubulin, and end-binding protein 1 (EB1). While microtubules may support VRC attachment and the structural stabilization of the growing VRC, the ER-associated actin network supports the recruitment of membranes and host factors to these sites with support of myosin motor proteins.

Many viruses associate with the ER for replication like TMV, but several other viruses use different locations to establish their VRCs. For example, Carnation Italian ringspot virus (CIRV) and Melon necrotic spot virus (MNSV) form factories from the outer membrane of mitochondria, whereas the tonoplast is implied to play a role in the replication of Alfalfa mosaic virus (AIMV) and Tomato mosaic virus (ToMV). Tomato bushy stunt virus (TBSV) induces the formation of multivesicular bodies (MVB) from peroxisomes and several other viruses show associations with chloroplasts. However, given that Golgi membranes, chloroplasts, peroxisomes, and also the tonoplast are tightly associated with the ER-actin network, ER membranes and associated vesicle trafficking as well as the above-described roles of the cytoskeleton in supporting membrane trafficking and anchorage may represent basic principles of membrane and host factor recruitment playing a role also in the association of viruses with these organelles.

Tight linkage between virus replication and movement is particularly important if more than a single viral RNA must be moved between cells to spread infection. Unlike TMV, RCNMV has a bipartite genome but also forms ER-associated VRCs and encodes two N-terminally overlapping replicase subunits similar to TMV. To ensure replication of RNA2 together with RNA1, the smaller replicase subunit (p27) of the replicase complex, which is associated with RNA 1, binds to RNA2 through

Fig. 2 Regulation of PD and modification by viruses. (a) PD structure and regulation. (i) PD are channels (ca. 50 μm in diameter) in the plant cell walls that provide cytoplasmic and membrane continuity between adjacent cells. Structurally they are concentric cylinders with the PM lining the cell wall and the ER forming a narrow rod called 'desmotubule' in the center. The desmotubule is linked to the PM by spoke-like extensions from all its sides. The cell wall regions surrounding the PD orifices at the PD neck regions contain deposits of callose (β -1,3 glucan). (ii) and (iii) Transport through PD occurs through the cytoplasmic space delimited by the PM and the desmotubule. This size of this space determines the size exclusion limit (SEL), thus the maximal size of the molecules that are allowed to pass from one cell to another. The SEL is regulated by two potential mechanisms. The first mechanism (ii) relies on the regulation of the callose deposits. Callose synthesis by callose synthases increases the size of the callose deposits in the cell wall, thereby narrowing the space between the PM and the desmotubule and, thus, reducing the SEL. Degradation of callose by β -1,3-glucanase has the opposite effect leading to an SEL increase. The second mechanism is based on the hypothesis that the spoke-like extensions between the desmotubule and the PM are actually formed by tether proteins that regulate their tether length in response to Ca^{2+} . PD are indeed regulated by Ca^{2+} . Elevated and reduced cytosolic Ca^{2+} levels were correlated with a decrease and increase in PD SEL, respectively. However, callose synthase holoenzyme complexes are also regulated by Ca^{2+} . Thus, further studies are needed to substantiate this model. (b) Modification of PD during viral infection. PD are associated with signaling complexes that sense the presence of pathogens and cause callose accumulation and PD closure (i). Virus-encoded MPs modify PD and allow virus movement irrespective of the physiological status of the cells by two different basic mechanisms. MPs that support virus movement in the form of a VRC/vRNP, as exemplified by the MP of TMV, modify the SEL by inducing the degradation of callose deposits, presumably by recruitment or activation of β -1,3, glucanases (ii). In contrast, tubule-forming viruses like CaMV, GFLV, or CPMV have MPs with the capacity to form a tubule-like structure within the PD channel allowing whole virion particles to move between cells. MP tubule formation depends on the interaction of MP with members of the Plasmodesmata Located Protein (PDL) family and leads to the displacement of the desmotubule and, thus, to the disruption of the ER-connectivity between cells (iii).



recognition of a Y-shaped RNA element (YRE) and also binds the complex to the membrane. Unlike the MP of TMV, the MP of RCNMV has no affinity for binding to the ER. Instead, the localization of this MP to the VRC rather depends on replicated RNA1. Interestingly, translation of MP from RNA2 depends on replicated RNA2. This way, virus movement can occur only after a functional VRC has been assembled and both viral RNAs have been replicated. The binding of the MP to the VRC is supported by the chloroplastic protein glyceraldehyde-3-phosphate dehydrogenase (GAPDH), which also binds to the smaller replicase subunit (p27). The role of chloroplastic proteins in virus replication and movement has also been noted for other viruses like TMV and Bamboo mosaic virus (BaMV). As chloroplastic proteins and salicylic acid (SA) produced in chloroplasts are implicated in the regulation of PD, it may not be surprising if chloroplasts play a yet uncharacterized role in coordinating virus replication and intercellular transport through PD. The interaction of the small subunit replicase of RCNMV with the ER and the resulting membrane perturbations involved in the formation of the VRCs are linked to interactions of p27 with components of the secretory pathway, such as ADP ribosylation factor 1 (ARF1), the target of the secretory pathway inhibitor Brefeldin A (BFA). Consistently, virus replication is inhibited by treatment of infected cells with BFA or by inhibition of the secretory pathway protein SAR1, indicating that RCNMV replication and, thus, also cell-to-cell movement depend on membrane trafficking within the COPI/COPII-dependent secretory pathway. These and many other studies indicate that the deformation and recruitment of membranes for virus replication is a rather complex process that generally involves viral interference with lipid metabolism and protein regulation, targeting and transport. Studies using heterologous expression in yeast have shown that the formation of replication vesicles by tombus- and bromovirus replicase proteins involves reticulons as well as ESCRT (Endosomal Sorting Complexes Required for Transport) and MVB (multivesicular endosome) proteins, thus implying a role of endosomal pathways and of proteins shaping the structure and curvature of membranes.

RCNMV exemplifies a relatively simple system in which only two parts of the viral genome must be replicated and moved between cells to spread infection. A more extreme case is represented by the multipartite nanoviruses, for example *Faba bean necrotic stunt virus* (FBNSV), which consists of eight genome segments. Recent studies to determine how the replication and cell-to-cell movement of eight genome segments can be coordinated led to the finding that the different segments of the virus do not occur in the same cells but in different cells. This exciting observation challenges the current view that virus replication must rely on single cells. Rather, nanoviruses, and potentially also other viruses with segmented genomes, may hijack groups of plant cells for replication and use intercellular trafficking of the virus-encoded products for functional complementation between segments and their coordinated replication and movement.

Given that virus replication and movement for many viruses are tightly connected, there must be a mechanism to determine which of the viral genomes stay in the cell to form virion progeny and which of the viral genomes are transported to PD to spread infection into adjacent cells. TMV does not require its coat protein (CP) for cell-to-cell movement and, during the time course of infection, the CP is expressed later than MP, thereby ensuring that virions are formed after the virus already spread infection between cells. The above-described studies with TMV but also studies involving other viruses like PVX and TuMV have shown that the VRC-containing membrane compartments occurring during infection are divided into small, mobile compartments associated with virus movement and larger, sessile compartments that occur later and function as viral factories that produce virion progeny. Although the two different compartments can often be observed in the same cells and may communicate with each other, the larger virion-producing compartments of TMV and also those of RCNMV and PVX rather mature after virus movement has occurred.

The mechanisms by which viruses transport their mobile VRCs to PD differ (Fig. 3) and many of the differences correlate with the mechanism by which the respective viruses modify PD to facilitate intercellular movement (see below). Because the ER membrane is continuous through PD between adjacent cells, many viruses target PD simply by transporting their VRCs along the ER membrane. However, several viruses, particularly the ‘tubule-forming’ viruses disrupt the ER continuity between cells (Fig. 2) and, therefore, depend on other, or additional, pathways to target the PD. Non-tubule-forming viruses can utilize the ER membrane because it is a fluid structure allowing embedded proteins and protein complexes to move by

Fig. 3 Examples of different PD targeting pathways used by viruses. (a) TMV forms VRCs at microtubule-associated sites of the cortical ER-actin network and targets the PD by diffusion along the ER. The targeting of the MP to PD and the spread of infection depends on class XI and class VIII myosin motors. While some early VRCs undergo movement and spread infection to new cells, other “early VRCs” stay behind and become virus factories that produce virion progeny. When class VIII myosins are inhibited, the MP accumulates in the PM and cannot enter PD. The concentration of MP within PM rafts at PD may involve endosomal recycling. (b) TuMV forms replication vesicles induced by the viral 6K2 protein that bud off from the ER at ER exit sites in a COPII-COPI-dependent manner and then move along actin filaments to reach PD. Cylindrical inclusions formed by the cylindrical inclusion protein (CI) acts as a docking point for 6K2 and for a viral protein involved in TuMV movement, called P3N-PIPO. The P3N-PIPO has the capacity to modify the SEL of PD. Thus, following docking, the cylindrical inclusion may guide P3N-PIPO and the 6K2-associated VRCs to the PD channel. (c) Tubule-forming viruses like GFLV, CaMV, and CPMV have MPs that assemble to hollow tubules that allow the cell-to-cell transport of whole virion particles. Tubule formation disrupts the ER connectivity between cells. PD targeting, tubule assembly, and virus movement depend on interaction of the MPs with members of the PDLF family at PD. The targeting of PDLF to PD depends on the ER-Golgi secretory pathway and, thus, virus movement also depends on this pathway. The cellular pathway by which GFLV targets its MP and virions to PD is not known. More is known for CaMV (but not shown here), which forms stationary and motile virion factories. The motile virion factories are transported to PD along actin filaments and deliver virion particles to the PD channel in a process involving viral proteins present in the motile complex and the MP at PD. Transport of virions through the tubules formed by the tubule-forming viruses depends on the interaction between the viral coat protein subunits of the virion and the MP within the tubule. The mechanism that drives virion transport through the tubule and whether this process may involve non-viral proteins remains to be elucidated.

diffusion. Moreover, similar to the ER-associated movements of organelles, such as Golgi complexes, mitochondria and peroxisomes, also viruses enhance the efficiency and directionality of their transport with the help of myosin motor proteins. By blocking the pathway of myosin, achieved by expression of actin-binding protein or by blocking specific myosin motors using myosin RNA silencing or expression of dominantly inhibiting myosin mutants, it was demonstrated that TMV depends on class XI and class VIII myosins for transport. Whereas class XI myosins support the transport of the TMV VRCs along the ER-actin network, class VIII myosins play a role at the PM and in the interaction of the MP with PD (Fig. 3(a)).

ER-mediated targeting to PD has also been demonstrated for many other viruses. However, while the mechanism of ER-mediated transport is rather simple for TMV, which encodes only one MP, the mechanism is more complex for viruses that encode more than one MP. For example, the rod-shaped hordei-like (hordei-, pomo-, peclu-, and beny-) viruses and potexviruses encode three MPs in overlapping open reading frames, the triple gene block (TGB). The general model for movement of these viruses involves the binding of viral RNA by TGB1 and the targeting of the TGB1:RNA complex to PD with the help of TGB2 and TGB3, transmembrane proteins that are localized to the ER. Studies using hordeiviruses, pomoviruses, and potexviruses as models demonstrated that TGB3 contains the signals for targeting specific sites on the ER, for the recruitment of TGB2, for transporting the TGB2/TGB3 complex to ER sites in proximity to the PD, and for recruitment of TGB1 to these sites. Several observations indicated that the TGB2/TGB3 complex is transported by ER-derived, motile complexes along actin filaments. Similar to the motile TMV VRCs, also these motile TGB complexes may represent VRCs. PVX, for example, replicates in association with the ER and the motile complexes contain the viral replicase as well as ribosomes and virions. At later stages of infection, the TGB proteins also colocalize with viral RNA to ER-associated replication factories that produce virions. Several lines of evidence suggest a role of microtubules in the localized formation and transport of the TGB VRCs. Thus, although many details remain to be explored, the mechanism of TGB virus movement using ER-guided movement of motile VRCs, potentially microtubules, and stationary VRCs for the formation of viral factories, appears to be similar to the ER-guided mechanism used by TMV. However, it is not always clear whether motile complexes described as “ER-derived” are still part of the ER or whether these rather represent ER vesicles that bud off from the ER. The TGB3 of *Poa semilatifolia* virus (PSV, an hordeivirus) forms motile complexes upon inhibition of the COPII-dependent ER-Golgi pathway thus indicating that PSV and potentially other TGB viruses target PD by an ER-mediated route that does not require exit from the ER. However, this is different, for example, for Turnip mosaic virus (TuMV, a potyvirus), which forms replication vesicles induced by the viral 6K2 protein that bud off from the ER at ER exit sites in a COPII-COPI-dependent manner and then move along actin filaments to reach PD (Fig. 3(b)). The requirement of exit from the ER for this virus may be linked to the specialized mechanism by which this virus targets and modifies PD for movement and which may disturb the normal ER continuity and PD function between cells. Indeed, this virus encodes a cylindrical inclusion protein (CI) that forms a highly elaborated virus-induced transport structure at PD. The CI protein acts as a docking point for 6K2 and for a viral protein involved in TuMV movement, called P3N-PIPO. P3N-PIPO has the capacity to modify the SEL of PD. Thus, following docking, the cylindrical inclusion my guide P3N-PIPO and the 6K2-associated VRCs to the PD channel.

Alternative routes for the targeting of the virus to PD are indeed required for the already mentioned “tubule-forming viruses”, which alter the inner structure of PD by forming a transport tubule assembled by MP and disrupt the ER continuity between cells (Fig. 3(c)). Viruses encoding tubule-forming MPs can be found in several virus families, such as the *Secoviridae*, *Caulimoviridae*, *Badnaviridae*, *Tospoviridae*, *Ilarviridae*, and *Bromoviridae*. One example is Cauliflower mosaic virus (CaMV), a pararetrovirus which replicates its dsDNA genome by reverse transcription of a terminally redundant RNA template. Infected cells form electron dense inclusion bodies in which the P6 protein of the virus facilitates the translation of the polycistronic viral RNA, and viral and host proteins are recruited for replication and virion assembly. In a manner similar to the virus factories formed by TMV, the inclusion bodies form in association with the ER and microtubules and can either be large, functioning as stationary “virion factories”, or small, functioning as motile complexes that target the PD. To reach PD, the motile complexes are transported along actin filaments facilitated by a protein called CHUP1 (Chloroplast Unusual Positioning Protein) that usually anchors chloroplasts to actin microfilaments and is required for the movements of chloroplasts in response to light. At the PD, the inclusion body delivers virions to the PD channel, in a process potentially involving interaction of P6 with the PD-localized MP (P1) and another P6-interacting protein called AtSRC2.2. Interestingly, the MP of the virus (P1) targets PD independently of the P6 inclusion bodies. As shown for CaMV and also for Grapevine fanleaf virus (GFLV) (Fig. 3(c)), the PD targeting by MP, efficient tubule assembly by MP and virus infection supported by MP depend on interaction of the tubule-forming MP with PD-localized proteins (PDL), a multigene protein family that localizes to PD via the ER-Golgi secretory pathway. The secretory pathway consists of interacting and highly dynamic organelles that move intracellularly with support of the actin cytoskeleton. Consistently, the PD targeting of PDL and, consequently, that of the tubule-forming MP is reduced by inhibition of the secretory pathway (e.g., with the chemical inhibitor Brefeldin A, BFA), of actin filaments (e.g., with the chemical inhibitor Latrunculin A, Lat A), or of myosins (e.g., with the chemical inhibitor 2,3 butane-dione-monoxime, BDM). Moreover, PDL and MP targeting to PD as well as GFLV movement and tubule formation by the GFLV MP depend on class XI myosins. Microtubules also play a role in this pathway, since GFLV MP tubules were shown to form at ectopic sites rather than at PD when microtubule inhibitors were present.

So far, the described mechanisms of intracellular transport involve membranes of the ER and the secretory system. However, several viruses also interact with endosomal pathways for movement. For example, TGB2 and TGB3 proteins do not move with viral RNA into adjacent cells but are recycled by endosomal membrane trafficking for further rounds of transport. Endosomal recycling may also play a role for the MP of TMV (Fig. 3(a)) and other MPs that interact with synaptotagmin A (SYTA, SYT1), one of five SYTs in *Arabidopsis*. SYTA was shown to localize to PM-derived endosomes at discrete areas at the PM, and to play an important role in endocytosis as well as in MP-mediated cell-to-cell transport. TMV movement may indeed involve a PM-associated function, since the MP showed strong accumulation at the PM when its targeting to PD was blocked by inhibition

of class VIII myosins. This observation may be related to the notion that PD contain PM rafts and that the MP of TMV and also the MPs of other viruses may be part of such rafts when present in PD. Thus, it appears conceivable that SYTA-associated endosomal recycling maintains the association of MP with membrane rafts and that myosin VIII is required to anchor the rafts at PD. Several MPs contain specific sorting motifs that are required for sorting into clathrin-coated endosomal vesicles. Mutation of such motif within the P1 MP of CaMV allowed the protein to accumulate at the PM but abolished its recruitment to PD thus again indicating that an endosomal recapture pathway plays a role in the recruitment of MPs from the PM to PD.

Intercellular Transport

PD are cytoplasmic communication channels in the cell walls between adjoining cells and play pivotal roles during plant development and in the interaction of plants with biotic and abiotic factors by allowing the cell-to-cell transport of water, metabolites, nutrients and hormones as well as of informational macromolecules such as RNAs, proteins and protein-RNA complexes. Structurally, they are concentric structures in which a cytoplasmic sleeve is delimited by the centrally located rod of the ER (the desmotubule) on one side and the PM that lines the outer diameter of the channel on the other. PD are associated with signaling components such as receptor proteins and receptor protein kinases as well as signal relaying components (*e.g.*, the already mentioned PDLs) by which the PD constantly adjust their permeability in response to developmental and environmental cues. The PM of PD contains rafts or other types of microdomains enriched in specific lipids and which bind and concentrate proteins carrying lipophilic modifications (*e.g.*, a glycosyl-phosphatidyl-inositol (GPI) “anchor”) that are enriched in PD. The SEL of PD, thus the upper limit of the size of molecules transported by PD, is determined by the degree of callose deposition in the cell wall surrounding the PD neck regions (**Fig. 2(a)**). Callose synthesis and accumulation by callose synthases in the cell wall leads to closure of the cytoplasmic compartment by forcing the PM against the desmotubule, whereas the degradation of callose by beta-1,3-glucanase opens the cytoplasmic space for intercellular transport. It has been proposed that the SEL of PD may also be regulated by proteins that tether the desmotubule to the PM and are able to regulate their tether length, and, thus, the size of the cytoplasmic space. The SEL of PD is regulated in response to Ca^{2+} waves or the sensing of pathogens. For example, one protein of the PDL family, PDL5, was shown to regulate the SEL of PD by stimulation of callose synthesis through callose synthase 1 (CalS1) and CalS8. Moreover, PDL5 is required for regulating PD permeability induced by bacterial pathogen or by accumulation of salicylic acid (SA), a hormone playing central roles in stimulating immunity against biotrophic pathogens, including viruses.

To overcome the gating of PD, the virus-encoded MPs modify PD in different ways (**Fig. 2(b)**). Whereas the MP of TMV and the MPs of probably most other viruses cause a transient modification of the PD SEL, the MPs of ‘tubule forming viruses’ assemble a hollow tubule within PD and thereby drastically alter the internal structure and function of the PD channel. The tubules formed within PD allow the cell-to-cell transport of whole virion particles, whereas viruses that modify the PD SEL move between cells rather in a non-virion form (*e.g.*, as VRCs) and may even be independent of CP, as in the case of TMV. Virion entry into tubules depends on direct or indirect interactions between the CP of the capsid and MP residues exposed in the inner tubule wall. Whereas the CPs and MPs of Cowpea mosaic virus (CPMV) and GFLV interact directly, the binding of the CP of CaMV to the MP (P1) is mediated through another viral protein (P3). The CaMV CP also binds to P6. Thus, both P6 and P3 may bind to the virion surface to assist in the transfer of virions from P6 inclusion bodies to MP (P1) within the tubule. Given that tubules are assembled by MP produced in the infected cells and have a given length, it seems likely that the tubules are dynamic structures in which MP units and associated virions that have docked to the tubule treadmill through the tubule length and are released at the distal end. However, whether tubules are indeed dynamic or rather static structures remains to be demonstrated.

The mechanism by which other MPs like that of TMV modify the PD SEL is not known. Although electron micrographs indicated that the TMV MP forms fibrillar substructures within PD cavities, it remains unknown whether these structures are functional. Other studies indicate that viruses modify the PD SEL by direct or indirect interaction with one or more beta-1,3-glucanases (GLU) and thereby direct the degradation of callose. The expression of specific GLU genes was shown to increase during infection and the efficiency of TMV movement was affected when the level of GLU expression was altered, thus confirming the hypothesis that a callose-degrading enzyme activity may regulate virus movement. Consistently, a specific glucanase, AtBG_{pap}, which localizes to PD, was identified. However, although an atbg_{pap} mutant exhibited reduced virus movement associated with increased PD-localized callose levels, it still remains to be clarified whether viruses like TMV indeed operate GLU activities for virus movement or whether they rather act by preventing infection-triggered callose synthesis. The MP of TMV and the TBG1 protein of PVX were shown to bind specific ankyrin repeat-containing proteins (TIP1-3, ANK). The tobacco homolog ANK was shown to facilitate virus movement and its co-expression with MP resulted in its recruitment from the cytoplasm to PD, in reduced callose levels at PD and increased intercellular diffusion of MP. Thus, ANK and related factors may be recruited by MP to facilitate virus movement. However, the exact mechanism through which these proteins act in virus movement remains to be further studied.

More recent studies indicated that the MP harbors a specific PD-localization signal (PLS) in its N-terminus. Interestingly, this domain is recognized by the MP-interacting protein SYTA/SYT1, which was already mentioned as a host protein required for the cell-to-cell movement of TMV and of several other viruses. Although the SYT proteins may play a role in Ca^{2+} -sensitive membrane trafficking, their similarity to mammalian synaptotagmin (E-Syt) and yeast tricalbin (Tcb) families suggests that they (also) act as membrane tethers. Thus, in addition to its function in endosomal trafficking, SYTA is now also recognized as an ER-anchored protein that tethers the ER to the PM at ER:PM contact sites and has critical functions in stabilizing the ER network. In binding MP, SYTA may play overlapping roles both in endosomal recycling of MP to maintain its enrichment at the cell periphery and also in

anchoring the MP-containing TMV VRCs to the microtubule-associated ER:PM contact sites mentioned above. Moreover, given the close proximity between the desmotubular ER and the PM within PD, it has been proposed that also the PD may represent ER:PM contact sites. In this case, SYTA may take part in the tethering of the desmotubule to the PM and thereby controlling the cytoplasmic space and, thus, the PD SEL by changing its tether length in response to Ca^{2+} (Fig. 2(a)). SYTA and other markers of ER:PM contact sites (e.g., VAP27) were found to localize to PD and to interact with PD-localized reticulons, which is consistent with this hypothesis. The finding that SYTA binds the PLS of MP and is required for the MP targeting to PD suggests that SYTA recruits or binds MP at PD. The association of MP with SYTA at PD may also imply the possibility that MP modifies PD SEL through regulating the length of SYTA tethers.

Several studies indicate that the PD SEL may also be directly or indirectly controlled by actin filaments. Consistent with a role of actin in controlling PD SEL, the MPs of TMV and of *Cucumber mosaic virus* (CMV) were shown to exhibit actin severing activity *in vitro*. Moreover, the stabilization of actin filaments by treatment with phalloidin prevented the ability of MP to increase the PD SEL *in vivo*. Although these observations suggest that MPs manipulate actin to control the PD permeability, it remains to be shown whether these MPs indeed interact with actin *in vivo* and, if so, whether the actin severing activity of these MPs occurs at PD or at other locations in the cell. Nevertheless, class VIII myosins were localized to PD, and such myosins were shown to be required for the interaction of the TMV MP with PD. Future studies may reveal whether or not PD indeed contain an actomyosin system and whether this system has a role in controlling the aperture of PD.

Virus movement through PD likely requires the activity of virus-encoded chaperones. This is illustrated by the virally encoded Hsp70 chaperone homolog (Hsp70h) of *Beet yellows virus* (BYV). This virus has a particularly large RNA genome and forms very long and flexible virions. Hsp70h is one of five virus-encoded proteins required for transport. This protein targets PD and is also found in actin-associated motile granules. Consistent with a role of the granules in intracellular transport, the targeting of Hsp70h to PD depends on intact actin filaments and class VIII myosins. The Hsp70h is a component of the filamentous capsid and its ATPase activity is required for virus cell-to-cell movement. These findings led to a model where Hsp70h mediates virion assembly and, once localized to the PD, actively translocates the virion from cell to cell via an ATP-dependent process. A role of viral proteins in plasmodesmal translocation of virus is also supported by the requirement of NTPase activities in other viral proteins for virus movement, such as in the TGBp1 helicase of PMTV and the CP of PVX. Moreover, also the helicase domains within the N-terminally overlapping replicase subunits of TMV are required for cell-to-cell movement. It seems plausible that the helicase activity found in several viruses is necessary for the remodeling of viral RNA during its passage through the PD.

As has been already mentioned, the process of virus movement through PD may also be regulated by posttranslational modifications of MP. In addition to its control by phosphorylation, the MP of TMV is also controlled by ubiquitinylation and degradation by the 26S proteasome. ER stress caused by viral replication and MP production leads to the activation of the AAA-ATPase CDC48, which binds and extracts the ubiquitinated MP from the ER and guides it into the cytoplasm for degradation. While this CDC48-dependent process of removing MP from the ER transport pathway may represent a plant defense response, this mechanism may also be part of a viral strategy to inhibit further virus movement and allow the virus to switch infection from an early phase focused on virus movement to a late phase focused on virus replication and virion production.

Systemic Transport

To achieve systemic movement, the infecting virus needs to pass through different cell types of the vascular veins until it reaches the phloem sieve elements for long-distance movement into other parts of the plant (Fig. 1). The phloem transports and distributes photosynthetic assimilates from source regions in mature leaves to sink regions that require carbon for growth and development. In addition to sugars the phloem allows the long-distance transport of many solutes, hormones, proteins and RNAs. Thus, once in the sieve element, the virus can exploit this mass flow from source to sink for its transport. The vascular veins are defined as major and minor, depending on their structure, location, branching patterns, and function. The vascular tissues of each vein, for example in tobacco, consist of cell types of the phloem, thus the sieve elements and adjacent companion cells together with phloem parenchyma cells, and cell types of the xylem, thus the xylem vessels and xylem parenchyma cells. All these cell-types are surrounded by cells of the bundle sheath which delimits the vascular tissue against the *mesophyll* tissue. Plant species differ in their strategy of loading sugars from the *mesophyll* into the phloem and this difference is reflected by the number of PD that connect the bundle sheath cells and the companion cells. In symplastic loaders, thus plant species that load sugars through PD, many PD are located at this interface, whereas in apoplastic loaders, thus plant species in which sugars are transported across the cell wall by specific transporter proteins located in the PM, only few PD are present at this interface. Since tobacco is an apoplastic loader, TMV is apparently able to find these few PD to enter the vascular tissue and cause systemic infection in this species. Moreover, whereas all species examined to date have PD between companion cells and sieve elements, PD are rather infrequent between the sieve element and either phloem parenchyma or bundle sheath cells. Thus, the sieve element is isolated from the rest of the leaf and the majority of molecules must enter the sieve element through companion cells, which therefore act as a major gatekeeper for phloem entry. Viruses use both minor or major veins to enter and exit the phloem in source and sink tissues, respectively. To reach the phloem, viruses recapitulate mechanisms of cell-to-cell movement. However, the ability to move long-distance depends on additional mechanisms. TMV, for example, moves cell-to-cell without CP but requires CP for systemic movement. Without the CP, the virus can reach bundle sheath and vascular parenchyma cells of the minor veins by cell-to-cell movement, but shows reduced accumulation in the companion cells compared to the wild type virus. This requirement of CP may

be consistent with several other findings in the literature indicating a requirement of specific interactions between virus and host factors at the boundary between different cell types. Moreover, because the phloem sieve elements are devoid of ribosomes and therefore lack the machinery to support virus replication, the viral genome may also require specific protection to reach the distant tissues. Virion particles of several viruses have been detected in the phloem, thus suggesting that virions are the most prevalent transport form of viruses in the phloem. For viruses that move cell-to-cell in a non-virion form like TMV, the exact site of virion formation is not known. While virion particles of several viruses have been detected in the PD that connect companion cells with sieve elements, the virions of CMV were only found in sieve elements thus suggesting that the virions are formed in the parietal, ER membrane-containing layer of the sieve element. Interestingly, MPs can enter the sieve element and may also be translocated to distant sites to exert an influence on PD in these regions. PD in sink tissues have larger SELs than PD in source tissues which facilitates phloem unloading. Thus, MPs that have entered the phloem stream could spread into and between cells of sink tissues and thereby prime these tissues for the subsequent export and transport of the virus.

Although long-distance movement appears to require encapsidation of the virus genome, umbraviruses like Groundnut rosette virus (GRV) do not encode a CP. Systemic movement of GRV depends on its ORF3 protein, which targets Cajal bodies in the nucleolus to recruit the nucleolar protein fibrillarin and to assemble complexes with viral RNA able to move systemically. Viroids are single-stranded, circular RNA molecules that even do not encode any protein but can nevertheless cause systemic infection. Thus, all proteins required for viroid replication and transport must be of plant host origin. The RNA is folded into a series of loops and bulges and studies on Potato spindle tuber viroid (PSTVd) and mutational analysis demonstrated that several of them are required for cell-to-cell and/or phloem entry and systemic transport. Certain phloem proteins that bind to viroids have been identified but their functional significance in viroid movement remains to be investigated.

Several viruses, for example, viruses belonging to the aphid-transmitted *Luteoviridae* family, are phloem-limited because they can enter the phloem but fail to be unloaded. The restriction of virus movement of these phloem-limited viruses usually occurs at some point outside the sieve element-companion cell (SE-CC) complex, suggesting that the limiting step in tissue invasion lies in the cell-to-cell post-phloem movement rather than in the ability to exit from the SE-CC.

Virus Movement and Plant Defense Responses

Successful systemic infection critically depends on the ability of the infecting virus to suppress the defense reactions of the plant. Replicating viruses trigger RNA silencing that targets viral RNA for cleavage and/or translational repression. RNA silencing is initiated by dsRNA produced during replication. This dsRNA is cleaved to small-interfering RNAs (siRNAs) that guide Argonaute protein-containing RNA-Induced Silencing Complexes (RISC) to further cleave viral RNA in an RNA sequence-dependent manner. Viruses encode a Viral Suppressor of RNA silencing (VSR) that binds small RNAs or interferes with another step of the RNA silencing pathway. Because plants use small RNAs and RNA silencing pathways to control gene expression, VSR activities can interfere with normal plant development and, thereby, play an important role virus-induced pathogenicity. Virus-derived and other small RNAs are mobile and may move ahead of the virus to immunize cells ahead of infection. The level and systemic spread of antiviral siRNAs is enhanced by an amplifying pathway involving RNA-dependent RNA polymerase 6 (RDR6). Enhanced production of antiviral siRNAs by RDR6 and their spread into sink tissues plays an important role in immune surveillance in meristems and derived young tissues. Viruses with weak or deleted VSR functions that are unable to suppress this amplified antiviral RNA silencing may fail to cause a sustained systemic infection. More recent studies have shown that in addition to triggering RNA silencing, dsRNAs also act as potent elicitors of Pattern-Triggered immunity (PTI), which inhibits virus propagation as well. Future studies may show how the two types of dsRNA-triggered plant host responses are coordinated during infection and whether they play synergistic roles in antiviral defense. A third layer of antiviral defense involves resistance (R) genes that cause Effector-triggered immunity (ETI) against specific viruses. ETI can lead to a hypersensitive response (HR) involving cell death, high levels of the immunity hormone SA, and the establishment of systemic acquired resistance (SAR). An important example is the tobacco N gene, which triggers an HR upon recognition of the VSR effector of TMV (the small subunits of the TMV replicase). An example for ETI without induction of an HR is provided by the RTM1 gene in Arabidopsis. The lectin-like RTM1 protein acts specifically in the phloem to restrict the long-distance movement of potyviruses, such as Tobacco etch virus (TEV), Lettuce mosaic virus (LMV), and Plum pox virus (PPV). The protein forms a complex with proteins encoded by other RTM resistance genes and likely targets the potyviral particle or a CP-containing ribonucleoprotein complex.

Further Reading

- Epel, B.L., 2009. Plant viruses spread by diffusion on ER-associated movement-protein-rafts through plasmodesmata gated by viral induced host beta-1,3-glucanases. *Seminars in Cell and Developmental Biology* 20, 1074–1081.
- Ghoshal, B., Sanfacon, H., 2015. Symptom recovery in virus-infected plants: Revisiting the role of RNA silencing mechanisms. *Virology* 497–480, 167–179.
- Gouveia, B.C., Calil, P., Machado, J.P., Santos, A.A., Fontes, E.P., 2017. Immune receptors and co-receptors in antiviral innate immunity in plants. *Frontiers in Microbiology* 7, 2139.
- Heinlein, M., 2015. Plasmodesmata: Channels for viruses on the move. *Methods in Molecular Biology* 1217, 25–52.
- Heinlein, M., 2015. Plant virus replication and movement. *Virology* 479–480, 657–671.
- Kleinow, T., 2016. Plant-virus interactions: Molecular biology, intra- and intercellular transport. Kleinow, T. (Ed.), Heidelberg: Springer.
- Niehl, A., Heinlein, M., 2011. Cellular pathways for viral transport through plasmodesmata. *Protoplasma* 248, 75–99.

- Niehl, A., Peña, E.J., Amari, K., Heinlein, M., 2013. Microtubules in viral replication and transport. *Plant Journal* 75, 290–308.
- Niehl, A., Wyrsh, I., Boller, T., Heinlein, M., 2016. Double-stranded RNAs induce a pattern-triggered immune signaling pathway in plants. *New Phytologist* 211, 1008–1019.
- Peña, E.J., Heinlein, M., 2013. Cortical microtubule-associated ER sites: Organization centers of cell polarity and communication. *Current Opinion in Plant Biology* 16, 764–773.
- Pitzalis, N., Heinlein, M., 2018. The roles of membranes and associated cytoskeleton in plant virus replication and cell-to-cell movement. *Journal of Experimental Botany* 69, 117–132.
- Reagan, B.C., Ganusova, E.E., Fernandez, J.C., McCray, T.N., Burch-Smith, T.M., 2018. RNA on the move: The plasmodesmata perspective. *Plant Science* 275, 1–10.
- Sambade, A., Brandner, K., Hofmann, C., *et al.*, 2008. Transport of TMV movement protein particles associated with the targeting of RNA to plasmodesmata. *Traffic* 9, 2073–2088.
- Schoelz, J.E., Angel, C.A., Nelson, R.S., Leisner, S.M., 2016. A model for intracellular movement of *Cauliflower mosaic virus*: The concept of the mobile virion factory. *Journal of Experimental Botany* 67, 2039–2048.
- Sevilem, I., Yadav, S.R., Helariutta, Y., 2017. Plasmodesmata: Channels for intercellular signaling during plant growth and development. *Methods in Molecular Biology* 1217, 3–27.
- Sicard, A., Pirolles, E., Gallet, R., *et al.*, 2019. A multicellular way of life for a multipartite virus. *Elife* 8, e43599.
- Tilsner, J., Linnik, O., Louveaux, M., *et al.*, 2013. Replication and trafficking of a plant virus are coupled at the entrances of plasmodesmata. *Journal of Cell Biology* 201, 981–995.
- Tilsner, J., Nicolas, W., Rosado, A., Bayer, E.M., 2016. Staying tight: Plasmodesmal membrane contact sites and the control of cell-to-cell connectivity in plants. *Annual Review in Plant Biology* 67, 337–364.
- Waigmann, E., Heinlein, M., 2007. *Viral Transport in Plants*. Plant Cell Monographs. vol. 7. Heidelberg: Springer.
- Wan, J., Cabanilla, D.G., Zheng, H., Laliberté, J.F., 2015. *Turnip mosaic virus* moves systemically through both phloem and xylem as membrane-associated complexes. *Plant Physiology* 167, 1374–1388.
- Wang, A., 2015. Dissecting the molecular network of virus-plant interactions: The complex roles of host factors. *Annual Review of Phytopathology* 53, 45–66.

Plant Antiviral Defense: Gene-Silencing Pathways

Vitantonio Pantaleo, National Research Council, Research Unit of Bari, Bari, Italy

Chikara Masuta and Hanako Shimura, Hokkaido University, Sapporo, Japan

© 2021 Elsevier Ltd. All rights reserved.

This is an update of G. Szitty, T. Dalmay, J. Burgyan, Plant Antiviral Defense: Gene Silencing Pathway, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00733-0](https://doi.org/10.1016/B978-012374410-4.00733-0).

Glossary

AGOs Argonaute (AGO) proteins bind to siRNAs and cleave target RNAs in plants as part of the RNA-induced silencing complex (RISC). *Arabidopsis thaliana* has ten AGOs with specific functions, which are summarized as follows: AGO1, antiviral silencing; AGO2, DNA methylation, antiviral silencing; AGO3, antiviral silencing; AGO4, DNA methylation; AGO5, antiviral silencing; AGO6, DNA methylation; AGO7, antiviral silencing; AGO8, induction of defenses against herbivores (insects); AGO9, DNA methylation; and AGO10, antiviral silencing, but also a singular pro-viral role.

DCLs Dicer-like proteins (DCLs, RNase III-like class of proteins) cleave double-stranded RNAs (dsRNA) to generate mostly 21- to 24-nt siRNAs or miRNAs. *A. thaliana* has four DCLs with specific functions that act synergistically: DCL1, biogenesis of 21-nt miRNAs, 21-nt nat-siRNAs; DCL2, biogenesis of 22-nt siRNAs for antiviral silencing, 24-nt nat-siRNAs; DCL3, biogenesis of 24-nt siRNAs (het-siRNAs) for RdDM; and DCL4, biogenesis of 21-nt siRNAs (24-nt in rice) for antiviral silencing.

RNA drug Viruses can be controlled through cellular antiviral RNA silencing (av-RNAi) against target viral genes by spraying double-stranded RNAs (dsRNAs), which are

processed into viral siRNAs (vsiRNAs) in the applied plant cells.

siRNAs Short double-stranded RNAs ranging in size from 21- to 24-nt long generated by Dicer-like proteins (DCLs). siRNAs include: vasiRNA, virus-activated siRNAs (21- and 22-nt) derived from endogenous RNAs; vsiRNA, viral small-interfering RNAs derived from viral RNAs; nat-siRNA, natural antisense transcript-derived siRNAs (21- or 24-nt); ta-siRNA, trans-acting siRNAs targeting mostly transcription factor genes; pha-siRNA, phased siRNAs (21- to 24-nt) from endogenous RNAs triggered by a particular miRNA; piRNA, piwi-interacting RNAs (21-nt) that suppress transposable elements (TEs) in the germline (to date not found in plants); het-siRNA, heterochromatic siRNAs (24-nt) generated by DCL3; easiRNA, epigenetically associated siRNA (21-nt), derived from pericentromeric TEs due to the transient loss of RNA-dependent DNA methylation (RdDM) in sperm.

Viral silencing suppressors (VSRs) Most plant viruses encode proteins that interfere with various steps in the host RNA silencing machinery. The main mechanism can be explained by their ability to bind to dsRNAs and siRNAs. Some VSRs can target the host DNA methylation machinery to inhibit host resistance modulated by epigenetics.

Introduction

RNA silencing or RNA interference (RNAi) refers to a conserved gene silencing process mediated by short RNAs, such as short-interfering RNAs (siRNAs) and micro RNAs (miRNAs). RNAi is one of the most important mechanisms in regulating gene expression and repressing transposable elements (TEs) at the transcriptional or posttranscriptional level. Across the eukaryotic kingdom, with the exception of some yeast species including *Saccharomyces cerevisiae*, the importance of RNAi-based gene regulation has been recognized as implicated in central developmental processes, maintaining genome integrity, and resilience to biotic and abiotic environmental inputs. Furthermore, plants synthesize viral siRNAs (vsiRNAs) and virus-activated siRNAs (vasiRNAs) as a means of defense in response to viral invasion.

RNAi was first described in the free-living nematode model *Caenorhabditis elegans*. It has been considered one of the most striking biological discoveries of the 1990s. Studies on RNAi have been readily and markedly supported by the genomic and Next Generation Sequencing (NGS) era in model plants; thus, studies in the field of small noncoding RNAs have advanced at a dramatic pace. It is now clear that RNAi is diverse, and in the last decade scientific efforts have been dedicated to extending the acquired knowledge from model plants to crops. Moreover, progress is being made to combine knowledge to the mechanisms and functions or to the applications of RNAi for sustainable plant protection.

Diverse Gene-Silencing Pathways in Plants

After 30 years since the discovery of RNAi in plants, most of the related knowledge is still rooted in analyses of the model plant *Arabidopsis thaliana*. RNAi pathways in *Arabidopsis* show considerable conservation of the molecular components: (1) long perfect or imperfect double-stranded RNAs (dsRNAs) are processed by DCLs into 21- to 24-nucleotide (nt) siRNAs; (2) siRNAs can be incorporated into members of

Argonaute (AGO), the core proteins of RNA-induced silencing complex (RISC); (3) sRNAs guide RISC to target nucleic acids in a sequence-specific manner and exert their function in transcriptional or post-transcriptional gene silencing (TGS and PTGS, respectively). Across eukaryotes, plants have the most complex RNA-based control of gene expression.

PTGS is usually directed against endogenous mRNAs and includes miRNA, trans-acting siRNA (ta-siRNA), natural antisense transcript-derived siRNA (nat-siRNA), and secondary phased siRNA (pha-siRNA) pathways. Instead, TGS is directed against components of the genome, including genes, repetitive elements and TEs.

The microRNA (miRNA) pathway. miRNAs derive from specific noncoding genes denoted as MIR genes, which are transcribed by RNA polymerase II (Pol II) into long single-stranded primary transcripts (pri-miRNA). pri-miRNAs exhibit typical Pol II cap structures at their 5' termini and poly(A) tails at their 3' termini and often contain introns. pri-miRNA transcripts adopt a fold-back stem-loop structure that is processed into a mature miRNA duplex by DCL1.

The processing of pri-miRNA into the miRNA duplex has been described in detail and requires:

- Cap-binding protein 20 (CBP20) and CBP80;
- zinc finger protein SERRATE;
- dsRNA-binding protein/HYPONASTIC LEAVES 1 (DRB1/HYL1);
- Forkhead-associated (FHA) domain-containing protein DAWDLE (DDL);
- TOUGH protein (TGH);
- Proline-rich protein SICKLE (SIC); and
- RNA-binding protein MODIFIER OF SNC1, 2 (MOS2).

miRNA duplexes in nuclei undergo a maturation process. Both strands of an miRNA duplex are methylated at their 3' terminal nucleotide by the RNA methyltransferase HEN1, an S-adenosyl methionine (SAME)-binding methyltransferase; the methyl group protects miRNAs from polyuridylation and degradation at the 3' termini.

Most methylated miRNA duplexes are exported to the cytoplasm by the exportin-5 homolog HASTY. One strand of the miRNA duplex acts as a guide strand and is selectively loaded onto an AGO-containing RISC, whereas the other strand, the passenger strand (miRNA*), is discarded from the complex and rapidly degraded. Most miRNAs associate with AGO1, but specific associations with AGO2, AGO7 and AGO10 have also been reported.

miRNAs are mainly 21-nt long, and mRNA cleavage is assumed as the common approach for miRNA-mediated gene regulation in plants. In addition to regulating RNA degradation, miRNAs occasionally direct DNA methylation or inhibit translation.

miRNAs regulate endogenous gene expression during development and environmental adaptation. Thousands of miRNAs have been identified in *Arabidopsis* and other plant species. Many of them are conserved and species-specific miRNAs. The list of currently known plant miRNAs can be found in miRbase (see "Relevant Websites section"), which has been recently updated.

The trans-acting siRNA pathway. ta-siRNAs derive from *TRANS ACTING siRNA (TAS)* genes, which are transcribed by Pol II into long non-coding, single-stranded (ss)RNAs that contain one or two specific miRNA target sites. TAS RNAs are targeted by miRNA-containing AGO1 (or AGO7)-RISC, and the cleaved TAS noncoding RNA is converted into dsRNA by the suppressor of gene silencing-3 (SGS3) and host RNA-directed RNA polymerase 6 (RDR6). Subsequently, DCL4 assisted by dsRNA-binding protein 4 (DRB4) processes the dsRNA to generate a population of 21-nt ta-siRNAs in phase with the miRNA-guided cleavage site. Similar to most miRNAs, ta-siRNA duplexes are methylated by HEN1. One strand of the ta-siRNA duplex associates with RISC to guide the cleavage of target mRNAs. Some ta-siRNAs regulate juvenile-to-adult transition or, more generally, transcriptional factors involved in plant response to environmental stimuli.

The natural cis-Antisense transcript-derived siRNA pathway. In plant genomes there are many genes that can be transcribed from complementary DNA strands at the same locus. If these genes are co-expressed, they produce overlapping sense/antisense transcripts. The overlapping region initiates the biogenesis of primary 24-nt nat-siRNAs through the action of a DCL (i.e., DCL2). Primary nat-siRNAs are loaded onto a yet unidentified AGO protein to direct the cleavage of the constitutively expressed complementary transcript. In a second step, the cleaved transcript is converted into dsRNA in a RDR6- and SGS3-dependent manner. This RNA amplification step may extend beyond the overlapping region to form outer siRNAs. Further DCL1-dependent processing of the newly synthesized dsRNA will generate 21-nt nat-siRNAs in phase with the primary cleavage. Similar to miRNAs and ta-siRNAs, nat-siRNA duplexes are methylated by HEN1.

nat-siRNAs were discovered in *A. thaliana* grown under salt stress conditions. However, recent genome-wide analyses have shown the widespread existence of overlapping sense/antisense transcripts, which raises the possibility that nat-siRNAs could be major effectors of gene regulation in the plant-adaptive protection mechanism in response to either abiotic or biotic stress.

Phased-siRNAs. The ability of miRNAs to indirectly regulate gene expression through the production of secondary siRNAs is an aspect that has become more evident only recently. Indeed, mRNAs cleaved by miRNA-loaded RISC can result in dsRNA synthesis, having the targeted transcript as template (like TAS ncRNAs). The newly synthesized dsRNA molecule is primarily the substrate for DCL4 to generate a new population of "secondary siRNAs"; these have a phased pattern similar to the ta-siRNA population, with siRNAs being produced at intervals of 21-nt (or 24-nt in rice) from the miRNA cleavage site. Secondary siRNAs can have a dramatic impact on gene regulation. They can act in *cis*, amplifying the silencing effect on their own target, or in *trans*, promoting a regulatory cascade if they derive from a member of large gene families. Nucleotide-binding site-leucine-rich repeat (NBS-LRR) clusters of pathogen resistance R-genes and pentatricopeptide repeats have been found subject to such regulatory cascades.

Heterochromatic and epigenetically activated siRNA Pathways. The majority of endogenous plant siRNAs consist in 24-nt siRNAs derived from heterochromatin, i.e. chromosome regions known to be extremely poor in genes and rich in repeated sequences, DNA transposons and silent retroviruses (retro-transposons), collectively named TEs. The biogenesis of heterochromatic (het)-siRNAs depends on plant DNA-dependent RNA polymerases Pol IV (p4) and Pol V (p5) and, therefore, het-siRNAs are also known as p4/p5-siRNAs. It has been proposed that p4/p5-dependent ssRNA transcripts from heterochromatic DNA loci generate dsRNAs through the action of RDR2, in partnership with other host factors. The long dsRNAs are processed by DCL3 into 24-nt-long het-siRNAs that are, in turn, methylated by HEN1. het-siRNAs associate with AGO4, AGO6, and AGO9, which then trigger TGS at RNA-dependent DNA methylation (RdDM) targets. This results in a silent heterochromatin, the prevention of TE evasion and preservation of genome stability.

In *Arabidopsis*, pericentromeric retrotransposons and thousands of DNA transposons lose RDR2-directed RdDM during pollen grain formation. Epigenetically reactivated TEs are preferentially targeted by more than 50 miRNAs bound to AGO1. TE transcripts engage RDR6 to form dsRNAs allowing DCL4 to produce 21-nt epigenetically activated (ea) siRNAs. easiRNAs trigger cascade PTGS of TE transcripts during the reprogramming of germlines, a critical window for the plant's life cycle. However, the transient loss of RdDM in sperm is restored after fertilization by maternal 24-nt het-siRNAs; therefore, easiRNAs are recognized as playing a role in fertility and male/female compatibility.

Another class of het-siRNAs is "Needed for RDR2-independent DNA methylation" (NERD) siRNAs. NERD is a class of proteins unique to plants that contain GW motifs known to bind AGO proteins. NERD has been shown to mediate TGS by recruiting AGO2-bound 21- to 22-nt siRNAs at specific genomic sites (i.e., those unmethylated at H3 lysine 9). This alternative TGS pathway is AGO4/RDR2 independent but RDR6/AGO2 dependent.

Antiviral RNA Silencing

Viral infection in plants requires replication in the infected cells, cell-to-cell movement, and long-distance spread of the virus. To defend themselves against viral infections plants use RNAi, a mechanism that fights parasitic nucleic acids. Invasive viruses and subviral infectious entities (satellites and viroids) are triggers of antiviral (av)-RNAi, which in turn targets viral genomes or transcripts for functional inactivation. av-RNAi results in a potent response of viral restriction using components of endogenous RNA silencing pathways. The av-RNAi mechanism in plants is best understood by examining the model plant *A. thaliana*, where the main players of RNA-silencing pathways have first been identified. The four DCLs (DCL1 to 4), ten AGO proteins (AGO1 to 10) and six RDR proteins (RDR1 to 6) are involved in av-RNAi depending on the type of plant virus. Different nucleic-acid content, genome organization and replication strategies influence the av-RNAi activated by plants.

From viral long dsRNAs to viral siRNAs. Viral dsRNA. The av-RNAi machinery is first triggered by viral dsRNAs at the level of infected plant cells. The origin of viral dsRNA is diverse. Av-RNAi is differentially activated depending on viral species (Fig. 1). Endornaviruses and reoviruses possess dsRNA genomes that trigger the RNAi response. In geminiviruses, dsRNA arises from the bidirectional transcription of overlapping genes. It is a common belief that most viruses with a positive (+) or negative (−) single-stranded ssRNA genome (approximately 75% of known plant viruses) also produce dsRNAs when replicating in infected cells. However, replication occurs in specific and protected membrane niches in the plant cell that releases solely viral ssRNA. This indicates that dsRNA replication intermediates (which have not been detected *in vivo* so far) are not essential to activate av-RNAi. Conversely, viral ssRNAs are potent triggers of av-RNAi. dsRNA signals of viral ssRNAs (viral genomes, sub-genomes, and transcripts) can anneal and form dsRNA motifs, also referred to as dsRNA-like structured ssRNAs.

DCLs. Primary viral dsRNA signals, either long perfect or imperfect, are recognized and processed by DCLs associated with DRBs into 21- to 24-nt RNA duplexes called primary viral small-interfering (vsi)RNAs. *A. thaliana* DCLs (DCL1 to 4) are all involved in primary vsiRNAs. Indeed, the dsRNAs derived from cytoplasmic RNA viruses are diced by DCLs 4 and 2, while those derived from viruses establishing minichromosomes, i.e., geminiviruses and caulimoviruses, are processed by DCLs 1 and 3.

RDRs. DCLs 2 and 4 are also deputed to process RDR-dependent viral dsRNAs into secondary vsiRNAs. Indeed, combinations of RDR activities contribute to the formation of secondary dsRNAs of viral origin. Viral aberrant ssRNAs (i.e., *bona fide* truncated viral RNAs lacking 3'-polyA or 5'-MG 7-CAP) are recruited by RDRs and converted into long perfect dsRNAs. RDR-dependent viral dsRNAs are processed by DCLs into 21- to 24-nt secondary vsiRNAs, which are structurally indistinguishable from primary vsiRNAs.

vsiRNAs (viral small-interfering RNAs). 21-nt vsiRNAs are predominantly associated with positive (+) strand RNA viruses and made by DCL4 (which also produces endogenous ta-siRNAs, see above). 22-nt vsiRNAs produced by DCL2 (also involved in nat-siRNAs, see above) in the presence of DCL4 often constitute less than 20% of the total vsiRNA population in infected tissues. Infection with DNA viruses induces a more abundant production of DCL3-derived, 24-nt vsiRNAs than 21- and 22-nt vsiRNAs in *A. thaliana*. DCL1 is active in viruses with DNA genomes and only in the absence of other DCLs. Primary and secondary vsiRNA duplexes are methylated by HEN1 at the 3'-termini of 2'-OH groups, analogously to endogenous functional siRNAs (see above).

RISCs in av-RNAi. AGOs are the core components of as yet incompletely characterized RISCs (see above). AGO/RISC-mediated av-RNAi is one of the most important components of the plant's immune response against viruses. In the last decade, major efforts were dedicated to accurately describe the antiviral RISC in plants. Besides *in vivo* studies, mainly based on reverse genetic approaches, *in vitro* systems have been recently launched and appeared to be valuable tools to define and characterize AGOs with

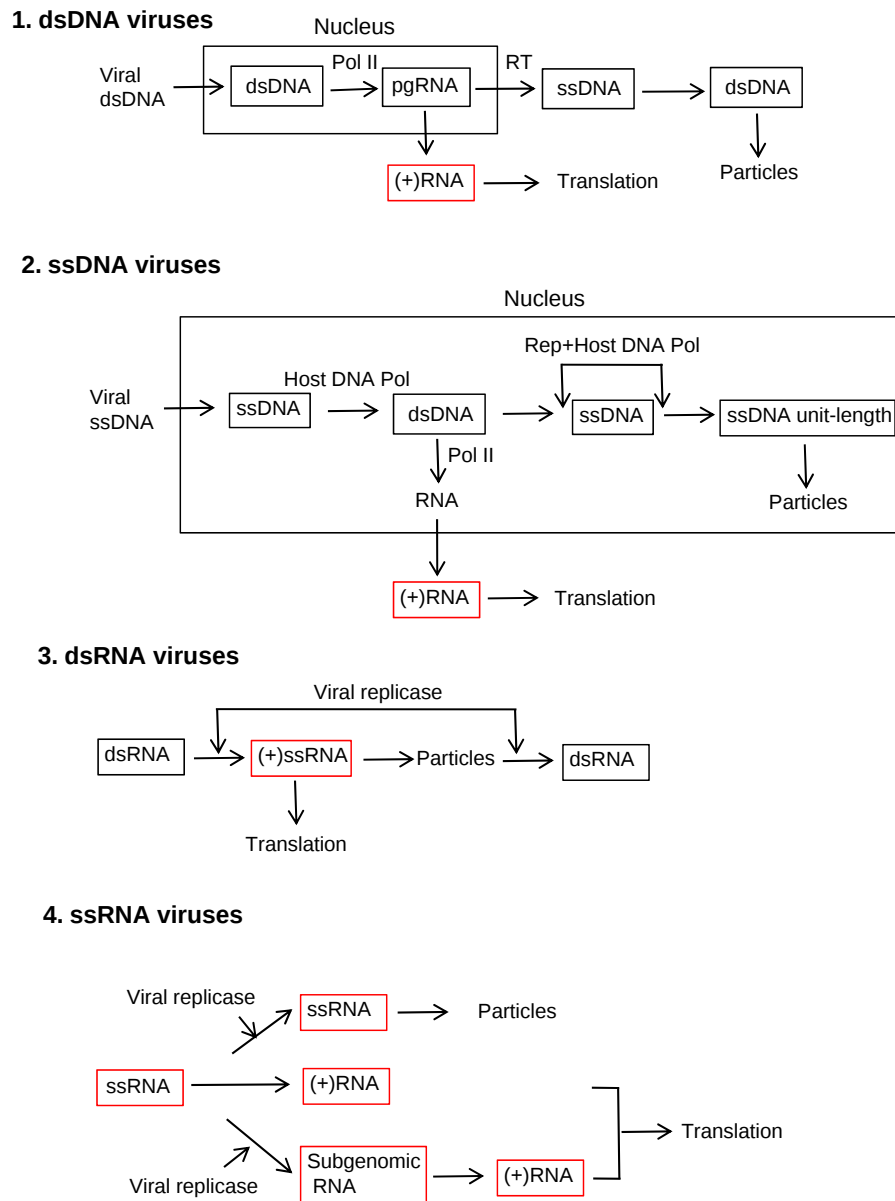


Fig. 1 Viral RNA species targeted by av-RNA silencing in host cells. Plant viruses, including DNA viruses, have RNA states in their transcription and replication cycles. Such RNAs are mainly cleaved by Argonaute (AGO) in RNA-induced silencing complex (RISC) or subjected to translational inhibition, though rarely. Double-stranded RNA (dsRNA) viruses may be subjected directly to Dicer cleavage, but their RNA transcripts (ssRNAs) seem to be a major target of AGO because dsRNAs are suspected to be formed in viral particles. Although dsDNA and ssDNA viruses use their RNA transcripts for replication in the nucleus, such RNAs become a target of RNA silencing when translated in cytoplasm.

antiviral functions. For instance, cytoplasmic extracts of evacuated *Nicotiana tabacum* BY2 protoplasts enabled the reconstitution of av-RNAi with a defined AGO protein and (an) exogenous siRNA(s) of choice.

Both *in vivo* and *in vitro* studies have clarified that of the ten AGO proteins that were first identified in *A. thaliana*, AGO1, –2, –3, –5, –7, and –10 have antiviral RNA-silencing activity. Upon incorporation of 21-nt or 22-nt-long vsiRNAs, these AGOs exhibited cleavage activity of targeted RNAs of viral origin and the capacity to inhibit viral replication. The most highly characterized antivirally acting AGO1 and AGO2 proteins contribute to the removal of viral or subviral RNA entities by endonucleolytic cleavage of the target RNA in a vsiRNA-directed, sequence-specific manner.

AGO4 has also been shown to have antiviral activity against ssDNA viruses, such as geminiviruses, through RdDM. Indeed, geminiviral mRNAs from bi-directional transcription are subject to av-RNAi similarly to RNA viruses. In addition, upon incorporation of 24-nt-long vsiRNAs, AGO4 promotes RdDM of viral chromatin leading to transcriptional gene silencing as a potent antiviral defense.

Either length or the 5'-terminal nucleotide has a strong impact on vsiRNA sorting into specific antiviral AGOs. For example, vsiRNAs of 21- and 22-nt have a 5'-terminal U co-immunoprecipitate with AGO1, whereas those that have a 5'-terminal A

co-immunoprecipitate with AGO2. Accordingly, vsiRNAs that are effective with AGO1 contain a U at the 5'-termini, whereas those effective with AGO2 contain an A at the 5'-termini.

As for endogenous regulatory sRNAs (see above), after the incorporation of a vsiRNA duplex into an AGO protein, the passenger strand is removed while the remaining guide(g) strand directs the RISC to the cognate viral RNA.

Function of vsiRNAs as a means of symptom development. In plants infected with viruses or viroids, vsiRNAs are abundantly produced. vsiRNAs induce av-RNAi and, in some instances, symptoms in infected plants. If vsiRNAs share high degrees of sequence similarity (or complementarity) with host genes, they can oftentimes induce specific symptoms through the host RNAi pathway.

One prime example of vsiRNA-mediated gene silencing is the bright yellowing symptom on tobacco infected with a satellite RNA (Y-sat) of cucumber mosaic virus (CMV). Y-sat-derived siRNAs can induce silencing of the chlorophyll biosynthetic gene (ChlI), leading to yellow chlorosis. Viroid vsiRNAs have also been implicated for a long time in specific symptom induction, yet their molecular mechanisms remain unknown. Viroid-induced symptoms seem to be a consequence of complex physiological disorder by viroid infection where viroids interact with many host factors. However, some target gene candidates of viroid vsiRNAs have been demonstrated to be associated with symptom expression; these include, for example, the chloroplast heat-shock protein 90 (cHSP90) gene by peach latent mosaic viroid (PLMVd), the chloride channel protein-*b*-like gene (CLC-*b*-like) and 40S ribosomal protein S3a-like gene (RPS3a-like) by potato spindle tuber viroid (PSTVd).

Recent advances in NGS and computational analyses have enabled the identification of vsiRNAs *in vivo* (denoted as degradome analysis) or *in silico*, which target host genes as well as the corresponding viral RNAs. By virtue of this technical development, vsiRNA-mediated host gene silencing leading to cleavage of the target transcripts has been proposed in many virus–host interactions, *e.g.*, tobacco mosaic virus (TMV)-*Arabidopsis*, cauliflower mosaic virus (CaMV)-*Arabidopsis*, and sugarcane mosaic virus (SCMV)-maize, as well as in grapevine with multiple viral infections.

Widespread silencing of host genes and other regulatory cascades induced by viral infections. In addition to vsiRNAs, plants produce a newly discovered class of endogenous siRNAs, the 21- and 22-nt virus-activated siRNAs (vasiRNAs). The production of vasiRNAs from thousands of *Arabidopsis*-encoding transcripts is induced by at least two distinct species of RNA viruses, *i.e.*, CMV and turnip mosaic virus (TuMV). The genetic requirements of vasiRNAs are distinct from those of miRNAs, vsiRNAs, ta-siRNAs and pha-siRNAs; although they are DCL4-dependent, they require RDR1 for biogenesis. Upon incorporation into AGO2, vasiRNAs direct the widespread silencing of host genes *in cis*. The observations performed to date suggest that the widespread silencing of host genes (that largely include factors responsive to biotic and abiotic stimuli) directed by vasiRNAs plays a role in the virus resistance activated during antiviral silencing in a manner that is independent of the antiviral activity of vsiRNAs.

Unlike the unique symptom induced by Y-sat, typical symptoms such as leaf mosaics, plant stunting and abnormal leaf development may not be simply explained by a specific vsiRNA, but rather by the result of all combined molecular effects mediated by miRNAs, vsiRNAs and vasiRNAs.

Viral and Plant RNA-Silencing Suppressors

Plant virus genomes embed genetic information that is surprisingly restricted but sufficient to coordinate the entire viral life cycle including cellular infection, replication processes, invasion of the host and spread into the environment. Additionally, viral structural or non-structural proteins are programmed to protect viral genomes against av-RNAi. Such proteins are widely known as viral silencing suppressors (VSRs). Initially, sRNA binding has been shown to be the prevalent and unifying mechanism of suppression activity at least in model systems such as p19 of tombusviruses, HC-Pro of potyviruses, and p21 of closteroviruses. In the last decade, the attention devoted by the scientific community to VSRs presented an extremely rich and diverse scenario: each step or player of the plant RNAi machinery emerged as a putative target of VSRs. It was also revealed that certain VSRs have multiple targets. The VSRs and silencing players in host immunity are summarized in [Table 1](#).

Many VSRs have been described as pathogenicity determinants and, indeed, the transient or constitutive overexpression of VSRs resulted in developmental abnormalities in plants. These capacities were attributed to the binding affinities of VSRs to the respective miRNAs and, accordingly, to the alteration of miRNA accumulation and functions. However, because VSRs differentially bind to miRNAs and the important silencing players such as DCL1, AGO1 and AGO2 are regulated by miRNAs, viral symptom development is not simply explained through miRNA sequestration by VSRs. In this perspective, the influence of VSRs may have a significant impact on endogenous RNAi during the early stages of infection.

Many endogenous RNA silencing suppressors (endoRSSs) have been identified in plants. It has been well documented that stress response, differentiation and development in plants are controlled by tuning the gene expression at the right time using miRNAs. In contrast to VSRs, which are produced to destroy host RNA silencing, endoRSSs probably play a roll in breaking miRNA-mediated gene regulation to remove excessively generated miRNAs. For example, it has been demonstrated that RNA silencing suppressed by FRY1 and XRN4 would be important for shaping root architecture.

Antiviral Silencing in Crops

In some virus-host couples (woody plants, agro-ecosystems, forests) that have cohabited for long periods, av-RNAi should be more properly considered as a mechanism of surveillance ensuring the coexistence of both partners and the absence of symptoms, with

Table 1 Target sites of VSRs to promote viral infection

Target sites	VSRs examples	Remarks
<i>Direct PTGS suppression</i>		
1. Generation of siRNA/miRNA		
(1) dsRNA binding		
Long dsRNA	NSs (TSWV), 2b (CMV), P38 (TCV-CP)	
ds-siRNA	HC-Pro (poty), 2b, P19 (TBSV), 126K (TMV)	
	V2 (begomo)	
ds-miRNA	2b, 126K	
(2) ssRNA binding		
ss-siRNA/miRNA binding	AC4 (begomo), NS3 (RSV), P10 (GVA), V2	V2 binds only 24-nt ss-siRNA
(3) DCL downregulation		
DCL3, DCL4	P38	Antagonistic effect of DCL1 upregulation
DCL1	2b	
(4) DCL binding		
DRB4 (DCL4 cofactor)	P6 (CaMV)	
2. siRNA/miRNA methylation		
(1) HEN1 binding	HC-Pro	
(2) No SAM supply to HEN1	HC-Pro, bC1	bC1 is produced by bsatellite of TYLCV
3. Target RNA cleavage		
(1) AGO binding		
AGO1	HC-Pro, 2b, p38, AC2 (begomo)	
AGO2	p38	
AGO4	2b	
(2) AGO downregulation		
AGO1	HC-Pro, 2b, 126K, P38	Downregulated by miR168 to control AGO1
(3) AGO degradation		
AGO1, AGO2, AGO4	P25 (potex), P0 (polero)	
4. mRNA Translation		
(1) Ribosome binding	HC-Pro, P6	
5. Secondary siRNA amplification		
(1) RDR6 binding		
SGS3 (RDR6 cofactor)	P25, V2, P2 (RSV), CMV-CP	
(2) RDR6 downregulation	HC-Pro	
<i>Indirect PTGS suppression mediated by host factors</i>		
XRN4 (exoribonuclease 4) binding	VPg (poty), HC-Pro	Reduce antiviral function of XRN4
WEL1 activation	AC2/AL2	Enhance WEL1 exonuclease activity
rgsCaM binding	HC-Pro	Stimulate endogenous RSS
DCP2 (decapping protein 2) binding	VPg, HC-Pro	Reduce antiviral function of DCP2
vasiRNAs suppression	2b	Reduce RDR1-dependent production of vasiRNAs
<i>TGS suppression</i>		
1. DNA methylation		
(1) ADK binding	AL2/L2 (begomo)	ADK, adenosine kinase
(2) SAHH binding	HC-Pro, β C1	SAHH, S-adenocyl-L-homocystein hydrolase
(3) SAMS binding	HC-Pro	SAMS, S-adenocyl-L-methionine syntase
(4) SAM decarboxylase I binding	C2 (begomo)	SAM, S-adenocyl-L-methionine

periodical outbreaks. The study of RNA silencing and av-RNAi mechanisms in the *Arabidopsis* model system, in which genetic dissection is possible, has helped the scientific community to accelerate the identification of biological gene functions. Upfront basic research has demonstrated that the av-RNAi process holds the key to future technological applications for the sustainable production of food (see also below, in this section). Indeed, av-RNAi is also effective in crop plants, which are important from an economic standpoint. Genome sequencing projects involve all crop types, including cereals and legumes, and are being extended to minor crops, ancient varieties or to wild species as a source of resistance to plant pathogens. In addition, image-based phenotyping is extremely effective to quantify the progression of biotic impacts and to associate it to genetic traits. The time is ripe to focus on av-RNAi in crop systems allowing us to contrast viruses (as well as other pathogens and pests) with the situation in crop plants.

In *Arabidopsis*, the existence of multiple paralogs of RNA-silencing-associated proteins (*e.g.*, ten AGOs, six RDRs and four DCLs) may explain the diversification of siRNAs and gene-silencing pathways; in crops the layer of complexity is more pronounced. The majority of crop plants possesses an allopolyploid genome, as they often undergo whole-genome duplication through the recombination of two or more genomes from progenitors. Paralogs of DCL and AGO genes are often present in crops,

along with the recruitment of different accessory proteins and co-factors. This has led to functional diversification or specialization in different crops, adding layers of complexity.

In rice (*Oryza sativa* L.), only AGO1 and AGO18 expressed activity against viruses. Rice AGO1 binds vsRNA to form a RISC for resistance against rice stripe virus (RSV) and rice dwarf virus. AGO18 expression is almost undetectable, but is highly induced upon RSV infection. Overexpression of AGO18 leads to increased expression of AGO1, which together confer broad-spectrum resistance against viruses in rice.

Nicotiana benthamiana (Domin) is considered either a model plant, since it is a permissive host for many plant viruses, or a typical case of allopolyploidization which is common in the Solanaceae family (e.g., tomato, potato, eggplant, pepper). The NbAGO1 gene is duplicated, and both homeologs (generically called “duplicates” or “isoforms”) retain the capacity to be transcribed into mRNAs and mainly differ in one 18-nucleotide insertion/deletion (indel). The indel is located eight nucleotides upstream of the target site of miR168, which is an important modulator of AGO1 expression. During infection with cymbidium ringspot virus, miR168 is up-regulated; however, the indel affects miR168-guided regulation and AGO1 isoforms, both of which appear to have specific involvement in symptom development. Similarly to *N. benthamiana*, tomato (*Solanum lycopersicum* L.) has AGO1 isoforms that are differently regulated in the case of tomato yellow leaf curl Sardinia virus (TYLCSV) infection. Moreover, functional indels affect the miRNA regulation of several transcriptional factors and NBS-LRR genes during viral infection.

In tomato a large family of disease-resistance proteins with NBS-LRR has been detected, and siRNA datasets revealed the presence of a regulatory cascade affecting NBS-LRRs. The initiator of the cascade is miR482, which targets mRNAs for NBS-LRR and triggers mRNA decay as well as the production of secondary siRNAs in an RDR6-dependent manner. These secondary siRNAs target other NBS-LRR mRNAs and continue the regulatory cascade. The miR482-mediated silencing cascade is suppressed by VSRs in tomato infected with viruses, increasing expression of the mRNAs targeted by miR482 or secondary siRNAs. This process allows pathogen-inducible expression of NBS-LRR proteins and contributes to a novel layer of defense against pathogen attack and a non-race-specific plant immunity induced by viral infection. Certain secondary siRNAs from NBS-LRR upon TYLCSV infection have been recognized to target and silence tomato genes involved in the photosynthetic pathway, partially explaining the phenotype plasticity in plants as well as the appearance of yellowing symptoms in the viral pathosystem.

Viral siRNA-Mediated Antiviral RNAi in Mammalian Cells

Paralogs of Dicer and AGO genes have emerged via duplication during the evolution of all eukaryotes, including mammals. Along with the recruitment of different accessory proteins and co-factors, this process has led to functional diversification or specialization in different organisms. For example, insects encode two DCL genes – one mediating miRNA biogenesis and the other being responsible for siRNA (and vsRNA) biogenesis. Conversely, mammals only encode a single DCL gene that generates both miRNAs and siRNAs. AGO2 is responsible for siRNA-mediated target RNA cleavage in insects, whereas AGO1 mediates miRNA-dependent gene silencing. Mammals, on the other hand, encode four AGO genes, all of which are involved in miRNA-guided gene silencing; only AGO2 is capable of cleaving target RNAs to mediate siRNA-dependent RNAi.

In plants, vsRNAs mainly induce the destruction of invading RNAs, leading to antiviral immunity. This is also the case in fungi and invertebrate cells. On the other hand, because mammalian antiviral defense heavily relies on interferon-mediated production of antiviral proteins, whether RNAi can indeed induce antiviral immunity in mammalian cells (especially somatic cells) has long been a matter of debate. Although numerous studies have reported the existence of siRNA-mediated av-RNAi in mammalian cells (not necessarily somatic cells), some questions still remain unanswered. It is clear that RNAi is involved in regulating mammalian viruses by host cellular miRNAs and that TEs have been found to be regulated by TE-derived siRNAs as well as by piwi-interacting RNAs (piRNAs) in both undifferentiated and somatic cells. However, to highlight the importance of av-RNAi in mammalian cells several factors must be carefully considered; for example: (1) whether VSRs is strong enough to reduce viral siRNA levels; (2) whether VSR can suppress other types of antiviral immune response; (3) whether the components in the RNAi pathway are well expressed in somatic cells; and (4) how cellular miRNAs contribute to the observed antiviral response.

View of Applications of av-RNAi

RNA drugs: novel non-transgenic and pesticide-free approaches for plant protection. Recovery phenotype in plants was described almost 80 years ago. In brief, primary virus-infected leaves manifested severe viral symptoms, whereas the upper leaves showed attenuated symptoms and became immune to secondary infection of the same or similar viruses in terms of viral sequence. av-RNAi in plants explained the recovery phenotype and supported the notion of searching for applicative solutions to boost natural av-RNAi. Thus, plant viruses became tools to examine RNA vaccination. RNA vaccination is attractive for several reasons: (a) RNA molecules can be applied transiently, i.e., not in the context of transgenic plants; (b) sRNAs are naturally produced in plants and other eukaryotes to regulate endogenous gene expression and mediate defense responses and crosstalk among organisms, with the potential of being extended to other pathogens including fungi, pests and insects; and (c) RNA vaccination enables rapid and flexible solutions against emerging pathogen or pest resistance by simple alterations of a target sequence.

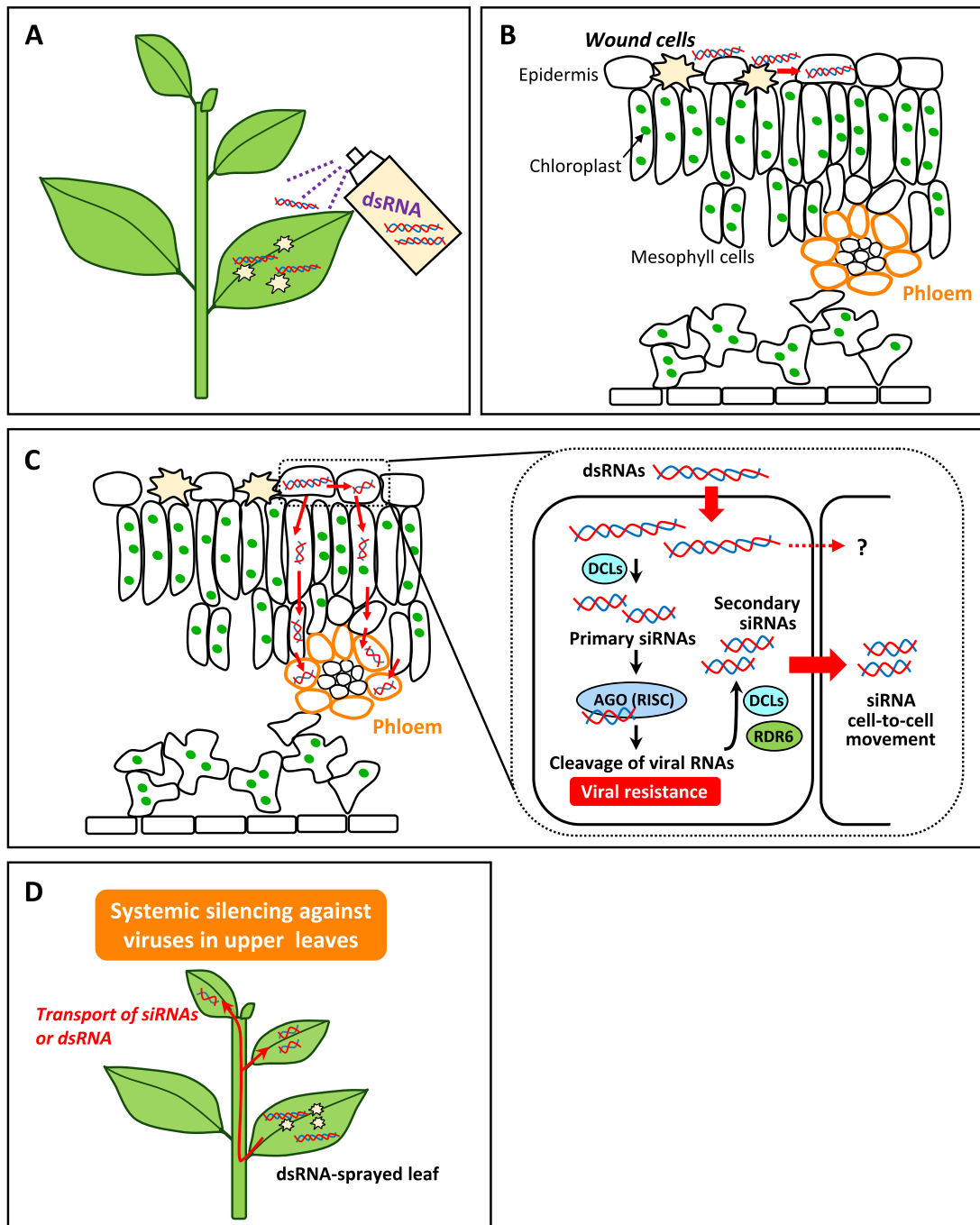


Fig. 2 Mechanism through which exogenous viral double-stranded RNAs (dsRNAs) sprayed onto leaves induce antiviral RNA silencing (from A to D). (A) dsRNA is sprayed onto leaves. (B) The applied dsRNA penetrates into cells through wound cells. (C) The dsRNA is then cleaved into short-interfering RNAs (siRNAs) by cellular Dicer-like proteins (DCLs). The siRNAs generated bind to Argonaute (AGO), which is a component of RNA-induced silencing complex (RISC), to guide AGO in targeting viral RNAs. Secondary siRNAs amplified by RNA-directed RNA polymerase (RDR) 6/DCL4 spread to neighboring cells. The transport of original dsRNA is not clear. (D) siRNAs (or dsRNAs) can be transported through phloem to induce systemic silencing in upper non-sprayed leaves.

dsRNA-based vaccination has been successfully applied against RNA viruses such as pepper mild mottle virus, tobacco etch virus, alfalfa mosaic virus, CMV, TMV, and zucchini yellow mosaic virus. dsRNAs can be easily applied to plants by spraying (Fig. 2).

Recognition of viral dsRNA by DCLs in av-RNAi produces vsRNAs to direct virus-specific resistance. In addition, viral dsRNA is being considered as a broadly conserved pathogen-associated molecular pattern (PAMP) by DCLs. PAMP involves the activation of

transcriptional signaling cascades (see vasiRNAs above and other endogenous secondary siRNAs) and could be used to confer broad-spectrum, non-specific pathogen resistance.

However, central to this application is the assumption that exogenous dsRNA does not show any non-specific side effects relating to mismatched off-target hybridization, protein binding to nucleic acids or activation of non-controlled cascades. A direct application of siRNAs ensures more specificity; the experience to date is positive and promising. Research, however, is currently focusing on aspects of stability, safety, efficiency and costs, which taken together will influence the diffusion of RNA drugs in agriculture.

Virus-induced transgenerational memory. Plant viruses can induce transient, stable, and heritable changes in gene expression without changes in genome sequences. For instance, plant-virus systems have led to the discovery and investigation of many aspects of plant epigenetics.

av-RNAi can, indeed, involve a transgenerationally inherited epigenetic modification. Thus, av-RNAi infection may induce heritable phenotypic variation that influences plant fitness in response to recurring pathogen invasions. For application purposes, its epigenetic mechanisms are far from being completely understood. However, science is not far from developing specific viral vectors that can be used to infect progenitors, conferring an inherited phenotype to the next generation. This would open the window to non-transgenic, virus-free engineered seeds.

vsiRNAs for viral metagenomics: from diagnostics to environmental studies. Given the RNAi-based plant antiviral and viral counter-defense strategies, av-RNAi often appears as a mechanism of surveillance ensuring the co-existence of both partners and the absence of symptoms. Indeed, thanks to the support of far-reaching NGS technology, we can detect new vsiRNAs as a result of av-RNAi even in non-symptomatic plants.

Viruses are incisively abundant on Earth and recognized as drivers of global processes. They also represent vast genetic and biological diversity on the planet. Viruses are obligate entities that infect the cells of living organisms, co-evolve with them and migrate from one host species to another. In plants, viruses or subviral entities can cause diseases that have serious bio-economical impacts, leading to the adoption of measures of prevention or resistance. However, in most cases, plant viruses are silent or asymptomatic in their hosts and continue to be regulators of plant gene expression, often through RNAi. Recent findings, for example, reinforce the notion that viruses are beneficial in complex ecosystems when used for specific purposes and when they do not cause diseases in agro-ecosystems. Indeed, plant viruses have been used to enhance the beauty of ornamental plants. They confer resilience under conditions of drought or cold temperatures. Genome-integrated viral forms can confer cross-protection against severe strains, support evolution by intra- or inter-specific horizontal gene transfer (plant-to-plant through virus), and can raise the fitness level of a crop in terms of seed production by attracting pollinators.

The great potential that plant viruses have in the agriculture and forestry sectors requires: (1) a greater focus on our biased notions of viruses as pathogens; (2) a deeper understanding of the mechanisms of how plant viruses regulate gene expression; and (3) closing the gap and providing insights into the composition and structure of environmental viral communities. av-RNAi includes steps of host-dependent amplification vsiRNAs, which in turn are molecular markers of viral infections. This unique feature allowed the identification of otherwise undetectable novel infectious entities in plants, present at a very low level of replication, using homology-independent approaches.

Further Reading

- Baulcombe, D.C., Dean, C., 2014. Epigenetic regulation in plant responses to the environment. *Cold Spring Harbor Perspectives in Biology* 6, a019471.
- Guo, Z., Li, Y., Ding, S.W., 2019. Small RNA-based antimicrobial immunity. *Nature Reviews Immunology* 19 (1), 31–44.
- Maillard, P.V., Ciaudo, C., Marchais, A., *et al.*, 2013. Antiviral RNA interference in mammalian cells. *Science* 342, 235–238.
- Martinez de Alba, A.E., Elvira-Matlot, E., Vaucheret, H., 2013. Gene silencing in plants: A diversity of pathways. *Biochimica et Biophysica Acta* 1829, 1300–1308.
- Mitter, N., Worrall, E.A., Robinson, K.E., Xu, Z.P., Carroll, B.J., 2017. Induction of virus resistance by exogenous application of double-stranded RNA. *Current Opinion in Virology*. 49–55. doi:10.1016/j.coviro.2017.07.009.
- Roossinck, M.J., 2015. A new look at plant viruses and their potential beneficial roles in crops. *Molecular Plant Pathology* 16, 331–333.
- Rosa, C., Kuo, Y.W., Wuriyanghan, H., Falk, B.W., 2018. RNA interference mechanisms and applications in plant pathology. *Annual Review of Phytopathology* 56, 581–610.
- Schuck, J., Gursinsky, T., Pantaleo, V., Burgyan, J., Behrens, S.E., 2013. AGO/RISC-mediated antiviral RNA silencing in a plant in vitro system. *Nucleic Acids Research* 41, 5090–5103.

Relevant Websites

<http://www.mirbase.org>
miRBase.

Plant Resistance to Viruses: Engineered Resistance

Marc Fuchs, Cornell University, Geneva, NY, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AGO Argonaute
amiRNA Artificial microRNA
CP Coat protein or capsid protein
CRISPR Clustered regularly interspaced short palindromic repeats
CAS CRISPR-associated proteins
DCL Dicer-like proteins with a dsRNA-specific endoribonuclease activity
DNA Deoxyribonucleic acid
dsRNA Double-stranded RNA
miRNA MicroRNAs

mRNA Messenger RNA
nt Nucleotide
PTGS Post-transcriptional gene silencing
RdRp RNA-dependent RNA polymerase
RISC RNA-induced silencing complex
ssRNA Single-stranded RNA
RNA Ribonucleic acid
RNAi RNA interference
siRNA Small interfering RNA
ta-siRNA Trans-acting siRNA
VSR Viral suppressor of RNA silencing

Glossary

Argonaute is a family of proteins that bind small noncoding RNAs, including miRNAs and siRNAs, and are guided to RNA targets through sequence complementarity, leading to the cleavage or translation inhibition of mRNAs, including viral RNAs.

CRISPR-Cas Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) is a prokaryotic adaptive immune system that can be leveraged for genetic material to be added, removed, or altered at particular locations in the plant genome, or to directly target viral DNA or RNA molecules.

MicroRNAs are a class of small non-coding ~21–24 nt RNAs that are involved in the regulation of gene expression at the post-transcriptional level by degrading target

mRNAs, including viral RNAs, and/or inhibiting their translation.

RNA interference is a biological process in which RNA molecules inhibit gene expression or translation by neutralizing target mRNA molecules, including viral RNAs.
Small interfering RNA is a class of 20–25 nt non-coding dsRNA molecules.

The RNA-induced silencing complex is a ribonucleoprotein complex that incorporates one strand of a miRNA or a siRNA duplex for RNAi.

Trans-acting siRNA are siRNA that repress plant gene expression through post-transcriptional gene silencing.

Transgenic plant is a plant with a genome modified by genetic engineering techniques to insert, remove or knockout genetic elements.

Introduction

Plant viruses can cause severe damage to crops by substantially reducing vigor, yield and product quality. A number of approaches are strategically implemented to mitigate the impacts of plant viruses. For example, quarantine measures and certification of seeds and propagative planting units are used to limit the introduction of virus diseases in production fields. Cultural practices, rogueing, control of vectors and cross-protection based on mild virus strains or benign satellite RNA are selected to reduce secondary spread of viruses in areas where epidemics can be problematic. However, the ideal and most effective approach to control viruses relies on the use of resistant crop cultivars. Host resistance genes have been extensively exploited by traditional breeding techniques although only a limited number of commercial crop cultivars exhibit practical resistance to viruses.

The advent of biotechnology through plant transformation, the application of the concept of pathogen-derived resistance, and the exploitation of the antiviral pathways of RNAi have opened new avenues for the development of virus-resistant crop cultivars, in particular when resistant material with desired horticultural characteristics has not been developed through conventional breeding or when no host resistance sources are known. RNAi is a conserved and RNA-dependent gene silencing process that is induced by dsRNA expression to target the cleavage of viral RNA molecules and confer resistance to virus infection via PTGS. In addition to the conventional expression of RNAi in transgenic plants to achieve virus resistance, RNAi can be triggered by the exogenous application of *in vitro*-produced dsRNA derived from viral sequences. More recently, genome editing via CRISPR/Cas has proven valuable to disrupt host susceptibility genes for the improvement of crop cultivars or directly target viral DNA or RNA for cleavage.

The past three decades have witnessed the development of numerous virus-resistant transgenic plants. Some crop cultivars engineered for virus resistance have been successfully tested in the field but only a few have been commercialized, including summer

squash and papaya which have been adopted for more than 20 years in the United States. Crop plants engineered for virus resistance offer numerous benefits to agriculture and cause extraordinarily limited, if any, hazard to the environment and human health.

A Historical Perspective on Engineered Resistance Against Plant Viruses

Genetic engineering expanded the scope of innovative approaches for virus control in crops by providing new tools to achieve resistance in otherwise susceptible plants. The first approach to confer resistance to viruses in plants resulted in the mid-1980s from the development of efficient protocols for plant transformation and the application of the concept of pathogen-derived resistance. This concept was initially conceived to achieve resistance by transferring and expressing a dysfunctional pathogen-specific molecule into the host genome to inhibit the pathogen. As a potential application of the concept, a segment of the virus' own genetic material was hypothesized to somehow protect a plant against virus infection by acting against the virus itself. The first demonstration of pathogen-derived resistance against a plant virus was with Tobacco mosaic virus (TMV) in 1986. This seminal work revealed that transgenic tobacco plants expressing the TMV CP gene showed a substantial delay in symptom development or did not display symptoms following inoculation with TMV particles. This type of resistance was known as CP-mediated resistance. It was successfully applied against numerous plant viruses.

In the early 1990s, RdRp, and then movement protein, protease, satellite RNA, defective interfering RNAs, or non-coding viral sequences were described to confer virus resistance in plants. Virus-derived transgene constructs included full-length, untranslated, and truncated coding and non-coding fragments, in sense or anti-sense orientation. It became soon apparent that more or less any viral sequence could provide some level of resistance to virus infection in transgenic plants. Remarkably, transcription of the virus-derived transgene rather than expression of a protein product was shown to be required for resistance. This pioneering work progressively led to the discovery of RNAi as a key antiviral defense of plants against viruses. Summer squash was the first virus resistant RNAi crop to be commercially released in 1996. Resistance to multiple viruses was achieved in summer squash cultivars by pyramiding several RNAi constructs derived from individual viruses.

Antiviral factors other than sequences derived from viral genomes have also been used to achieve resistance to viruses in plants. Production of antibodies, 2'-5' oligoadenylate synthase, ribosome inactivating proteins, dsRNA-specific RNases, dsRNA-dependent protein kinases, cysteine protease inhibitors, nucleoprotein-binding interfering aptamer peptides, and pathogenesis-related proteins has been reported to protect plants against virus infection.

Over the past three decades, resistance was engineered against numerous viruses with diverse taxonomic affiliations and various modes of transmission by insect vectors in many plant species and crop plants, essentially via RNAi. Further refinements of RNAi for virus resistance consisted of engineering amiRNAs to target a virus sequence and protect plants from virus infection. More recently, genome editing via CRISPR/Cas has proven valuable for engineered virus resistance by modifying host recessive resistance genes for crop improvement or directly degrading viral DNA or RNA.

Molecular Underpinnings of Engineered Resistant to Plant Viruses

Several mechanisms underlying engineered virus resistance have been described. Early studies suggested that expression of a transgene protein product seemed to be required for resistance. For example, resistance to TMV in tobacco plants was initially related to the expression level of the viral CP. Interference with an early step in the virus infection cycle, *i.e.*, disassembly of TMV particles, was hypothesized to explain CP-mediated resistance, maybe by inhibition of viral uncoating or reduction of protein-protein interactions between transgene-expressed CP and challenge virus. Resistance could be overcome by high doses of inoculum and was not very effective against virus RNA inoculation.

Next, resistance was shown to be RNA- rather than protein-mediated because untranslatable virus CP transgenes protected plants against virus infection. In addition, resistance was correlated to actively transcribed transgenes but low steady-state levels of transgene mRNA, indicating a sequence-specific post-transcriptional RNA-degradation system. RNA degradation was shown to target transgene transcripts and viral RNA with high nucleotide sequence identity. This was the first indication of the occurrence of PTGS as a manifestation of RNAi, a mechanism that inhibits gene expression in a sequence specific manner.

RNAi is a conserved eukaryotic mechanism that is critical for host defense against viruses and has become an approach of choice to engineer resistance to viruses in transgenic crops. A conserved feature of RNAi in plants is the processing of dsRNA into siRNA by DCL enzymes. Briefly, RNAi is initiated by dsRNA molecules and controlled by RISC. Long dsRNA molecules, for instance dsRNA corresponding to viral sequences, are cleaved into short dsRNA fragments of 20–25 nt siRNAs duplexes. The passenger strand of the siRNA duplexes is degraded and the guide strand is incorporated into RISC by DCL endoribonuclease proteins. Hybridization of the guide siRNA strand with a complementary RNA sequence such as a viral RNA molecule induces cleavage by AGO2, the catalytic component of RISC. This nucleotide sequence-based cleavage results in the degradation of the viral RNA target, hence resistance. Duplexes of siRNAs result from dsRNA derived from RNA virus replication, synthesis by plant RdRp, transcripts and hairpin-loop structures of ssRNAs.

Plants appear to have a set of four DCL proteins to process and generate siRNAs. DCL1 mainly processes imperfectly base-paired fold-back precursors to produce miRNAs. DCL2 generates 22 nt siRNAs and can produce secondary viral siRNAs. DCL3 generates 24 nt siRNAs that are associated with silencing of transposons and repetitive elements, and are involved in defense against DNA viruses. DCL4 generates 21 nt siRNAs that have major roles in antiviral defense against RNA viruses, and is involved in the generation of ta-siRNAs that regulate gene expression.

As an alternative to RNAi transgenic plants, exogenous application of *in vitro*-produced dsRNA to initiate RNAi via sprays or use of carrier compounds, including layered double hydroxide clay nanosheets, among other means of delivery, confers resistance to virus infection. This strategy was even shown to inhibit insect-mediated virus transmission.

To counter the RNAi plant defense, viruses have evolved RNAi-suppressor proteins called VSR. VSRs interfere at different steps of the antiviral RNAi pathways by sequestering siRNA, protecting dsRNA from degradation or interacting directly with host enzymes involved in RNAi. As a result of VSR expression, RNAi-mediated resistance is impaired.

RNAi can sometimes be unpredictable with regard to the degree of virus resistance. Different factors may be involved in deceiving levels of resistance or even loss of resistance. Nucleotide sequence divergence between a dsRNA transgene construct and cognate sequence in a virus population, the emergence of new virus strains in terms of the evolutionary dynamic of virus populations, VSRs of cognate or heterologous viruses, environmental conditions with fluctuating growing temperature during a cropping season can all affect the efficacy of RNAi at protecting plants from virus infection.

Another class of non-coding regulatory RNAs, miRNAs, has been exploited to engineer virus resistance in plants. Expression of modified miRNA precursors for which intrinsic plant specific nucleotides are replaced with virus-specific nucleotides results in the production of amiRNAs that target viral RNA sequences, resulting in resistance via RNAi. Although amiRNAs have been used for engineering resistance to a wide range of plant viruses, this strategy awaits field evaluation.

Engineered plant virus resistance can also be achieved by using CRISPR-Cas, a prokaryotic adaptive immune system which can be reprogrammed into a gene targeting technology. An RNA-guided Cas nuclease protein cleaves the substrate viral DNA or RNA at specific target sites, leading to their degradation. The specificity of the cleavage is governed by base complementarity between the CRISPR RNA and the target DNA or RNA molecule. A number of Cas proteins with sequence-specific nuclease activity have been identified to target DNA, RNA or both. Different CRISPR-Cas platforms have been successfully leveraged to engineer resistance to DNA viruses or RNA viruses *in planta*. CRISPR-Cas has also been used to knockout the host recessive eIF4E gene in cucumber and confer broad virus resistance. It will be interesting to see how this promising technology performs under field conditions.

Field Resistance and Commercial Adoption

Summer squash line ZW-20 expressing the CP gene of Zucchini yellow mosaic virus (ZYMV) and Watermelon mosaic virus (WMV), and summer squash line CZW-3 expressing the CP gene of ZYMV, WMV and Cucumber mosaic virus (CMV) were the first disease resistant transgenic plant events commercially released. Summer squash CZW-3 plants are highly resistant to mechanical inoculation of viruses in the greenhouse (Fig. 1(A)) or to aphid-mediated virus transmission in the field (Fig. 1(D)). Summer squash ZW-20 plants are highly resistant to mechanical inoculation in the field (Fig. 1(C)) or aphid-mediated virus transmission under field conditions in which no insecticides were applied to control aphid vector populations (Fig. 1(F)). Transgenic summer squash plants produce fruits of marketable quality (Fig. 1(E) and (H)) whereas most fruits of control plants are malformed and discolored (Fig. 1(E) and (G)).

Numerous transgenic crop plants such as cereal, vegetable, fruit, legume, flower, and forage crops, among others, that express RNAi constructs, virus-derived gene constructs or other antiviral factors have been developed and tested under field conditions (Table 1). A few transgenic crops engineered for virus resistance have been commercially released in the United States and the People's Republic of China (Table 1). In the United States, summer squash lines ZW-20 and CZW-3 have been commercialized in 1996. Several squash types and cultivars have been developed by crosses and back crosses with the two initially deregulated lines. Papaya expressing the CP gene of Papaya ringspot virus (PRSV) has been deregulated and commercialized in Hawaii in 1998. PRSV-resistant papaya was the first transgenic fruit crop to be commercially released in the United States. PRSV-resistant transgenic papaya cultivars are also commercialized in the People's Republic of China. Several potato expressing the RdRp gene of Potato leafroll virus or the CP gene of Potato virus Y were deregulated in 1998 and 2000. However, soon after their release, these potato lines were withdrawn from the market.

Benefits of Virus-Resistant Transgenic Plants to Agriculture

Virus-resistant transgenic plants offer numerous benefits to agriculture, particularly in cases where no genetic source of resistance has been identified or host resistance is challenging to transfer into elite cultivars by traditional breeding approaches due to genetic incompatibility or linkage to undesired traits. In such cases, engineered resistance may be the only viable option to develop virus-resistant cultivars. Engineered resistance may also be the only approach to develop cultivars with multiple sources of resistance by pyramiding several RNAi constructs.

Benefits of virus-resistant transgenic plants are of economic importance because yields are increased and quality of crop products improved. For example, the adoption of PRSV-resistant transgenic papaya cultivars resulted in a similar production level than before PRSV became epidemic in Hawaii, providing a steady source of income to growers. Benefits are also of epidemiological importance because virus-resistant transgenic plants limit virus infection rates by restricting challenge viruses, reducing their titers, or inhibiting their replication and/or cell-to-cell or systemic movement. Therefore, lower virus levels reduce the frequency of acquisition by vectors and subsequent transmission within and between field sites. Consequently, virus epidemics are substantially limited. In addition, benefits of virus-resistant transgenic plants are of environmental importance because chemicals directed to

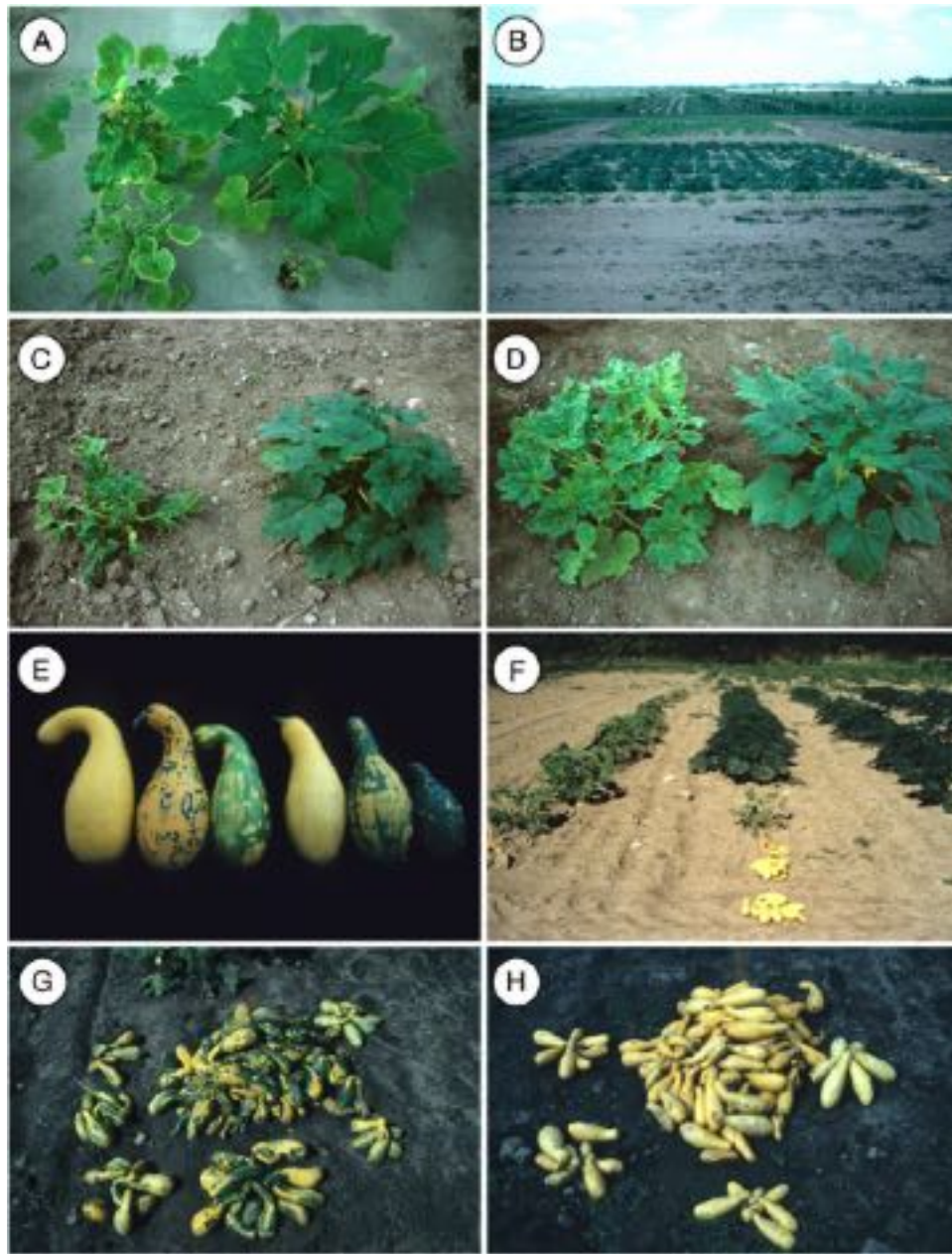


Fig. 1 Reaction of transgenic summer squash ZW-20 expressing the CP gene of Zucchini yellow mosaic virus (ZYMV) and Watermelon mosaic virus (WMV), and of transgenic summer squash CZW-3 expressing the CP gene of ZYMV, WMV and Cucumber mosaic virus (CMV) to virus infection. (A) Resistance of transgenic CZW-3 to infection by CMV, ZYMV and WMV (upper right), and susceptibility of conventional plants to single infection by CMV (lower right), ZYMV (upper left) and WMV (lower left); (B) Fields of transgenic ZW-20 (foreground) and conventional summer squash plants (background) surrounded by a border row of mechanically-inoculated non-transgenic plants; (C) Transgenic ZW-20 (right) and conventional (left) plants subjected to mechanical inoculation of ZYMV and WMV; (D) Transgenic CZW-3 (right) and conventional (left) plants mechanically inoculated with ZYMV, WMV and CMV; (E) Close-up of a fruit of transgenic ZW-20 (extreme left) and fruits of conventional summer squash plants; (F) Transgenic CZW-3 (middle and right rows) and conventional (left row) summer squash subjected to aphid-mediated inoculation by ZYMV, WMV and CMV; (G) Fruit production of conventional summer squash under conditions of infection by ZYMV and WMV; and, (H) Fruit production of transgenic summer squash ZW-20 under conditions of infection by ZYMV and WMV.

control virus vector populations are reduced or not necessary. Restricting the reliance on chemicals directed to arthropod, fungal and nematode vectors of plant viruses is important for sustainable agriculture. Finally, benefits of virus-resistant transgenic plants are of social importance. In Hawaii, growing papaya was not viable anymore prior to the adoption of PRSV-resistant transgenic papaya despite huge efforts to eradicate infected trees in order to limit the propagation of the virus. The impact of PRSV was so severe that some growers abandoned their farms and had to find new jobs to secure an income. PRSV-resistant transgenic papaya saved the papaya industry and strengthened the social welfare of local communities in Hawaii.

Table 1 Examples of virus-resistant transgenic crops that have been tested in the field or commercially released. This list is not exhaustive

<i>Crop category</i>				
<i>Crop common name</i>	<i>Crop scientific name</i>	<i>Resistance to virus</i>	<i>Genus</i>	<i>Family</i>
<i>Cereals</i>				
Barley	<i>Hordeum vulgare</i>	Barley yellow dwarf virus	<i>Luteovirus</i>	<i>Luteoviridae</i>
		Wheat dwarf virus	<i>Mastrevirus</i>	<i>Geminiviridae</i>
		Barley stripe mosaic virus	<i>Hordeivirus</i>	<i>Virgaviridae</i>
Canola	<i>Brassica napus</i>	Turnip mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
Maize	<i>Zea mays</i>	Maize dwarf mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Maize chlorotic dwarf virus	Waikavirus	Secoviridae
		Maize chlorotic mottle virus	Sobemovirus	Sobemoviridae
		Sugarcane mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
Oat	<i>Avena sativa</i>	Barley yellow dwarf virus	<i>Luteovirus</i>	<i>Luteoviridae</i>
Rice	<i>Oryza sativa</i>	Rice stripe virus	<i>Tenuivirus</i>	<i>Phenuiviridae</i>
		Rice yellow mottle virus	<i>Sobemovirus</i>	<i>Sobemoviridae</i>
		Rice hoja blanca virus	<i>Tenuivirus</i>	<i>Phenuiviridae</i>
		Rice black-streaked dwarf virus	<i>Fijivirus</i>	<i>Reoviridae</i>
		Rice dwarf virus	<i>Phytoreovirus</i>	<i>Reoviridae</i>
		Rice gall dwarf virus	<i>Phytoreovirus</i>	<i>Reoviridae</i>
		Rice black streaked dwarf virus	<i>Fijivirus</i>	<i>Reoviridae</i>
		Rice grassy stunt virus	<i>Tenuivirus</i>	<i>Phenuiviridae</i>
		Rice ragged stunt virus	<i>Oryzavirus</i>	<i>Reoviridae</i>
		Rice tungro bacilliform virus	<i>Tungrovirus</i>	<i>Caulimoviridae</i>
Wheat	<i>Triticum aestivum</i>	Barley yellow dwarf virus	<i>Luteovirus</i>	<i>Luteoviridae</i>
		Wheat streak mosaic virus	<i>Tritimovirus</i>	<i>Potyviridae</i>
		Wheat yellow mosaic virus	<i>Bymovirus</i>	<i>Potyviridae</i>
		Triticum mosaic virus	<i>Poacevirus</i>	<i>Potyviridae</i>
		Barley stripe mosaic virus	<i>Hordeivirus</i>	<i>Virgaviridae</i>
<i>Vegetables</i>				
Cucumber	<i>Cucumis sativus</i>	Cucumber mosaic virus	<i>Cucumovirus</i>	<i>Bromoviridae</i>
		Papaya ringspot virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Squash mosaic virus	Comovirus	Secoviridae
		Watermelon mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Zucchini yellow mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Cucumber vein yellowing virus	<i>Ipomovirus</i>	<i>Potyviridae</i>
Lettuce	<i>Lactuca sativa</i>	Lettuce mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Lettuce necrotic yellows virus	<i>Tenuivirus</i>	<i>Phenuiviridae</i>
		Mirafiori lettuce big-vein virus	<i>Ophiovirus</i>	<i>Aspiviridae</i>
Pepper	<i>Capsicum</i>	Cucumber mosaic virus	<i>Cucumovirus</i>	<i>Bromoviridae</i>
		Tobacco etch virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Tomato mosaic virus	<i>Tobamovirus</i>	<i>Virgaviridae</i>
		Potato virus Y	<i>Potyvirus</i>	<i>Potyviridae</i>
		Pepper mild yellow mottle virus	<i>Tobamovirus</i>	<i>Virgaviridae</i>
Squash ^a	<i>Cucurbita pepo</i>	Cucumber mosaic virus ^a	<i>Cucumovirus</i>	<i>Bromoviridae</i>
		Papaya ringspot virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Squash mosaic virus	Comovirus	Secoviridae
		Watermelon mosaic virus ^a	<i>Potyvirus</i>	<i>Potyviridae</i>
		Zucchini yellow mosaic virus ^a	<i>Potyvirus</i>	<i>Potyviridae</i>
Sugar beet	<i>Beta vulgaris</i>	Beet necrotic yellow vein virus	<i>Benyvirus</i>	<i>Benyviridae</i>
		Beet western yellows virus	<i>Polerovirus</i>	<i>Luteoviridae</i>
Tomato	<i>Solanum lycopersicum</i>	Beet curly top virus	<i>Curtovirus</i>	<i>Geminiviridae</i>
		Cucumber mosaic virus	<i>Cucumovirus</i>	<i>Bromoviridae</i>
		Tobacco mosaic virus	<i>Tobamovirus</i>	<i>Virgaviridae</i>
		Tomato mosaic virus	<i>Tobamovirus</i>	<i>Virgaviridae</i>
		Impatiens necrotic spot virus	Orthotospovirus	Tospoviridae
		Tomato spotted wilt virus	Orthotospovirus	Tospoviridae
		Ageratum yellow vein Malaysia virus	<i>Begomovirus</i>	<i>Geminiviridae</i>
		Tomato yellow leaf curl virus	<i>Begomovirus</i>	<i>Geminiviridae</i>
<i>Fruits</i>				
Alemow	<i>Citrus macrophylla</i>	Citrus tristeza virus	<i>Closterovirus</i>	<i>Closteroviridae</i>
Banana	<i>Musa</i> sp.	Banana bunchy top virus	<i>Babuvirus</i>	<i>Nanoviridae</i>
Grapefruit	<i>Citrus paradisi</i>	Citrus tristeza virus	<i>Closterovirus</i>	<i>Closteroviridae</i>

Table 1 Continued

Crop category				
Crop common name	Crop scientific name	Resistance to virus	Genus	Family
Grapevine	<i>Vitis</i> sp.	Grapevine fanleaf virus	<i>Nepovirus</i>	<i>Secoviridae</i>
Melon	<i>Cucumis melo</i>	Cucumber mosaic virus	<i>Cucumovirus</i>	<i>Bromoviridae</i>
		Papaya ringspot virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Squash mosaic virus	<i>Comovirus</i>	<i>Secoviridae</i>
		Watermelon mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Zucchini yellow mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Cucurbit yellow stunt disorder virus	<i>Crinivirus</i>	<i>Closteroviridae</i>
		Watermelon silver mottle virus	<i>Orthotospovirus</i>	<i>Tospoviridae</i>
Mexican lime	<i>Citrus aurantifolia</i>	Citrus tristeza virus	<i>Closterovirus</i>	<i>Closteroviridae</i>
		Citrus psorosis virus	<i>Ophiovirus</i>	<i>Aspiviridae</i>
Passion fruit	<i>Passiflora edulis</i>	Cowpea aphid borne mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Passion fruit woodiness virus	<i>Potyvirus</i>	<i>Potyviridae</i>
Papaya ^{a,b}	<i>Carica papaya</i>	Papaya ringspot virus ^{a,b}	<i>Potyvirus</i>	<i>Potyviridae</i>
Pineapple	<i>Ananas comosus</i>	Pineapple wilt-associated virus	<i>Ampelovirus</i>	<i>Closteroviridae</i>
Plum	<i>Prunus domestica</i>	Plum pox virus	<i>Potyvirus</i>	<i>Potyviridae</i>
Raspberry	<i>Rubus idaeus</i>	Raspberry bushy dwarf	<i>Idaeovirus</i>	Unassigned
		Tomato ringspot virus	<i>Nepovirus</i>	<i>Secoviridae</i>
Strawberry	<i>Fragaria</i> sp.	Strawberry mild yellow edge virus	<i>Potexvirus</i>	<i>Alphaflexiviridae</i>
Sweet orange	<i>Citrus sinensis</i>	Citrus tristeza virus	<i>Closterovirus</i>	<i>Closteroviridae</i>
		Citrus psorosis virus	<i>Ophiovirus</i>	<i>Aspiviridae</i>
Tamarillo	<i>Cyphomandra betacea</i>	Tamarillo mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
Walnut	<i>Juglans regia</i>	Cherry leafroll virus	<i>Nepovirus</i>	<i>Secoviridae</i>
Watermelon	<i>Citrullus lanatus</i>	Cucumber mosaic virus	<i>Cucumovirus</i>	<i>Bromoviridae</i>
		Cucumber green mottle mosaic virus	<i>Tobamovirus</i>	<i>Virgaviridae</i>
		Watermelon mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Zucchini yellow mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Papaya ringspot virus	<i>Potyvirus</i>	<i>Potyviridae</i>
<i>Legumes</i>				
Bean	<i>Phaseolus vulgaris</i>	Bean golden mosaic virus	<i>Begomovirus</i>	<i>Geminiviridae</i>
Clover	<i>Trifolium repens</i>	Alfalfa mosaic virus	<i>Alfamovirus</i>	<i>Bromoviridae</i>
		White clover mosaic virus	<i>Potexvirus</i>	<i>Alphaflexiviridae</i>
Cowpea	<i>Vigna unguiculata</i>	Mungbean yellow mosaic India virus	<i>Begomovirus</i>	<i>Geminiviridae</i>
		Cowpea severe mosaic virus	<i>Comovirus</i>	<i>Secoviridae</i>
		Cowpea aphid-borne mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
Groundnut	<i>Arachis hypogaea</i>	Peanut clump virus	<i>Pecluvirus</i>	<i>Virgaviridae</i>
		Groundnut rosette virus	<i>Umbravirus</i>	<i>Tombusviridae</i>
Pea	<i>Pisum sativum</i>	Alfalfa mosaic virus	<i>Alfamovirus</i>	<i>Bromoviridae</i>
		Bean leafroll virus	<i>Luteovirus</i>	<i>Luteoviridae</i>
		Bean yellow mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Pea enation mosaic virus	<i>Umbravirus</i>	<i>Tombusviridae</i>
		Pea seed-borne mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Pea streak virus	<i>Carlavirus</i>	<i>Betaflexiviridae</i>
Peanut	<i>Arachis hypogaea</i>	Tomato spotted wilt virus	<i>Orthotospovirus</i>	<i>Tospoviridae</i>
		Groundnut rosette assistor virus	unassigned	<i>Luteoviridae</i>
		Peanut stripe virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Tobacco streak virus	<i>Ilarvirus</i>	<i>Bromoviridae</i>
Soybean	<i>Glycine max</i>	Alfalfa mosaic virus	<i>Alfamovirus</i>	<i>Bromoviridae</i>
		Soybean mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
		Bean pod mottle virus	<i>Comovirus</i>	<i>Secoviridae</i>
		Southern bean mosaic virus	<i>Sobemovirus</i>	<i>Solemoviridae</i>
		Soybean dwarf virus	<i>Luteovirus</i>	<i>Luteoviridae</i>
		Bean yellow mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
<i>Flowers</i>				
Chrysanthemum	<i>Chrysanthemum indicum</i>	Tomato spotted wilt virus	<i>Orthotospovirus</i>	<i>Tospoviridae</i>
		Chrysanthemum virus B	<i>Carlavirus</i>	<i>Betaflexiviridae</i>
Dendrobium	<i>Encyclia cochleata</i>	Cymbidium mosaic virus	<i>Potexvirus</i>	<i>Alphaflexiviridae</i>
Gladiolus	<i>Gladiolus</i> sp.	Bean yellow mosaic virus	<i>Potyvirus</i>	<i>Potyviridae</i>
Impatiens	<i>Impatiens walleriana</i>	Impatiens necrotic spot virus	<i>Orthotospovirus</i>	<i>Tospoviridae</i>
Poinsettia	<i>Euphorbia pulcherrima</i>	Poinsettia mosaic virus	<i>Tymovirus</i>	<i>Tymoviridae</i>

(Continued)

Table 1 Continued

Crop category				
Crop common name	Crop scientific name	Resistance to virus	Genus	Family
Forage				
Alfalfa	Medicago sativa	Alfalfa mosaic virus	Alfamovirus	Bromoviridae
Grass				
Sugarcane	Saccharum sp.	Sugarcane mosaic virus	Potyvirus	Potyviridae
		Sugarcane yellow leaf virus	Polerovirus	Luteoviridae
		Sorghum mosaic virus	Potyvirus	Potyviridae
Tuberous root				
Cassava	Manihot esculenta	African cassava mosaic virus	Begomovirus	Geminiviridae
		East African cassava mosaic virus	Begomovirus	Geminiviridae
		Cassava brown streak virus	Ipomovirus	Potyviridae
		Sri Lankan cassava mosaic virus	Begomovirus	Geminiviridae
		Uganda cassava brown streak virus	Ipomovirus	Potyviridae
Potato	Solanum tuberosum	Potato virus A	Potyvirus	Potyviridae
		Potato virus X	Potexvirus	Alphaflexiviridae
		Potato virus Y	Potyvirus	Potyviridae
		Potato virus S	Carlavirus	Betaflexiviridae
		Potato leafroll virus	Polerovirus	Luteoviridae
		Tobacco rattle virus	Tobravirus	Virgaviridae
		Tobacco vein mottling virus	Potyvirus	Potyviridae
		Sweet potato	Ipomea batatas	Sweet potato feathery mottle virus
Sweet potato	Ipomea batatas	Sweet potato chlorotic stunt virus	Crinivirus	Closteroviridae
		Sweet potato virus G	Potyvirus	Potyviridae
		Sweet potato mild mottle virus	Potyvirus	Potyviridae
Fiber				
Cotton	Gossypium hirsutum	Cotton leaf curl Kokhran virus	Begomovirus	Geminiviridae
		Cotton leaf curl Multan virus	Begomovirus	Geminiviridae
		Cotton leaf curl Multan betasatellite	Betasatellite	Telocusatellidae

^aCommercially released in the USA.^bCommercially released in the People's Republic of China.

Environmental and Human Health Safety Issues

The insertion and expression of virus-derived genes or gene fragments in plants has raised concerns on their environmental and human health safety. Potential risks relate primarily to the occurrence and outcomes of heterologous encapsidation, recombination, gene flow, toxicity and allergenicity. It is important to keep in perspective that these three phenomena did not arise with transgenic plants expressing viral genes or gene fragments; they occur in conventional plants too. Therefore, it is critical to determine if they occur in transgenic plants beyond baseline events in conventional plants.

Heterologous encapsidation refers to the encapsidation of the genome of a challenge virus by CP subunits expressed in a transgenic plant containing a viral CP gene. Since the CP carries determinants for vector specificity, among other features, the properties of field viruses may change. Therefore, it is conceivable that an otherwise vector non-transmissible virus could become transmissible through heterologous encapsidation and infect otherwise non-host plants. Extensive work has shown that heterologous encapsidation in transgenic plants expressing virus CP genes should be considered negligible in regard to adverse environmental effects, given this phenomenon is known to occur in conventional plants subjected to mixed virus infection. It should also be kept in perspective that any change in vector specific is a single generation rather than a permanent event because they are not perpetuated in the virus genome progeny.

Recombination refers to template switching between viral transgene transcripts and the genome of a challenge virus. Resulting recombinant viruses may have a chimeric genome consisting of a segment from the challenge viral genome and another segment from viral transgene transcripts. It is argued that recombinant viruses may have identical biological properties as their parental lineages or new biological properties such as changes in vector specificity, expanded host range or increased pathogenicity. Since recombination alters the genome of challenge viruses, new properties of chimeric viruses will be stably transmitted to and perpetuated within the virus progeny. Comprehensive laboratory studies documented the occurrence of recombination in transgenic plants expressing viral genes. However, extensive field studies revealed a limited significance of recombination in transgenic plants expressing viral genes in regard to adverse environmental effects. This is because the stringency of selective pressure applied to the challenge virus is a critical factor in the recovery of recombinant viruses. Conditions of high selective pressure enhance the creation of recombinant viruses. In contrast, limited, if any, recombinant viruses are found to detectable level

under conditions of low or no selective pressure. The latter conditions prevail under field conditions in which plants are infected by functional, not defective, viruses.

Gene flow refers to the pollen-driven movement of transgenes from a virus-resistant transgenic plant into a non-transgenic compatible recipient plant, for example, a free-living relative. Hybrids resulting from gene flow can acquire and express virus-derived transgenes, and become resistant to the corresponding viruses. Subsequently, plants acquiring viral resistance traits can have a competitive advantage, exhibit increased fitness and eventually become more invasive. Movement of viral transgene constructs through pollen flow has been documented from a few virus-resistant transgenic crops, including summer squash, into free-living relatives under experimental field conditions. Hybrids between transgenic and free-living plants exhibited increased fitness under conditions of intense disease pressure. In contrast, under conditions of low disease pressure, no difference was observed between hybrids and free-living plant in terms of growth and reproductive potential. It is important to keep in perspective that gene flow with virus-resistant transgenic plants should not be perceived riskier than the equivalent situation with virus-resistant conventional plants.

Safety assessment studies on human health have documented no evidence of toxicity and allergenicity in crop cultivars engineered for virus resistance or food derived thereof. To fully grasp the significance of these findings, it is critical to recognize that plant virus proteins or sequences are not known to code for toxins, create new allergens or change the expression of endogenous proteins to produce allergens. In addition, foods derived from virus-resistant transgenic summer squash and papaya have been consumed by millions of people for more than 20 years, with no reported ill effects or legal cases related to human health.

So far, there is no compelling evidence to indicate that transgenic plants expressing virus genes or gene fragments increase the frequency of heterologous encapsidation, recombination, or food toxicity and allergenicity beyond background events. Similarly, there is little evidence, if any, to infer that transgenic plants expressing virus genes or gene fragments alter the properties of existing virus populations or create new viruses that could not arise naturally in conventional plants subjected to multiple virus infection. Finally, consequences of gene flow should be assessed on a case-by-case approach to make sound decisions on the safe release of crop cultivars engineered for virus resistance.

Conclusions

Virus diseases remain a huge problem for a sustainable and efficient agriculture. Engineered resistance via RNAi and genome editing has expanded the scope of innovative approaches for the management of virus diseases by providing new tools to develop resistant crop cultivars and increasing opportunities to implement effective and sustainable management strategies. The past three decades have witnessed an explosion in the development of virus-resistant transgenic plants. Lessons from field experiments and commercial release of virus-resistant summer squash, papaya, and potato have demonstrated that benefits outweigh by far risks to the environment and human health. Nonetheless, despite remarkable progress, less than a handful of virus-resistant transgenic crops have been commercially adopted. A timely release of new virus-resistant transgenic crop cultivars would be desirable because viruses continue to be devastating.

Further Reading

- Baulcombe, D., 2019. How virus resistance provided a mechanistic foundation for RNA silencing. *The Plant Cell* 31, 1395–1396.
- Dalakouras, A., Wassenegger, M., Dadami, E., *et al.*, 2019. GMO-free RNAi: Exogenous application of RNA molecules in plants. *Plant Physiology*. doi:10.1104/pp.19.00570.
- Dong, O.X., Ronald, P.C., 2019. Genetic engineering for disease resistance in plants: Recent progress and future perspectives. *Plant Physiology* 180, 26–38.
- Fuchs, M., Gonsalves, D., 2007. Safety of virus-resistant transgenic plants two decades after their introduction: Lessons from realistic field risk assessment studies. *Annual Review of Phytopathology* 45, 173–202.
- Gaffar, F.Y., Koch, A., 2019. Catch me if you can! RNA silencing-based improvement of antiviral plant immunity. *Viruses* 11, 673. doi:10.3390/v11070673.
- Gottula, J., Fuchs, M., 2009. Toward a quarter century of pathogen-derived resistance and practical approaches to engineered virus resistance in crops. In: Loebeinstein, G., Carr, J.P. (Eds.), *Natural and Engineered Resistance to Plant Viruses*. *Advances in Virus Research* 75. Elsevier, pp. 161–183.
- Kalinina, N.O., Khromov, A., Love, A.J., Taliansky, M.E., 2019. CRISPR applications in plant virology: Virus resistance and beyond. *Phytopathology*. doi:10.1094/PHYTO-07-19-0267-IA.
- Kalid, A., Zang, A., Yasir, M., Li, F., 2017. Small RNA-based genetic engineering for plant viral resistance: Application in crop protection. *Frontiers in Microbiology* 8, 43.
- Lindbo, J.A., Falk, B.W., 2017. The impact of “coat protein-mediated virus resistance” in applied plant pathology and basic research. *Phytopathology* 107, 624–634.
- National Academies of Sciences, Engineering, and Medicine, Engineering, and Medicine, 2016. *Genetically engineered crops: Experiences and prospects*. Washington, DC: The National Academies Press, doi:10.17226/23395.
- Powell Abel, P., Nelson, R.S., De, B., *et al.*, 1986. Delay of disease development in transgenic plants that express the tobacco mosaic virus coat protein gene. *Science* 232, 738–743.
- Pooggin, M.M., 2017. RNAi-mediated resistance to viruses: A critical assessment of methodologies. *Current Opinion in Virology* 26, 28–35.
- Rosa, C., Kuo, Y.-W., Wuriyanghan, H., Falk, B.W., 2018. RNA interference mechanisms and applications in plant pathology. *Annual Review of Phytopathology* 56, 581–610.
- Schmitt-Keichinger, C., 2019. Manipulating cellular factors to combat viruses: A case study from the plant eukaryotic translation initiation factors eIF4. *Frontiers in Microbiology* 10. doi:10.3389/fmicb.2019.00017.
- Tricoli, D.M., Carney, K.J., Russell, P.F., *et al.*, 1995. Field evaluation of transgenic squash containing single or multiple virus coat protein gene constructs for resistance to cucumber mosaic virus, watermelon mosaic virus 2, and zucchini yellow mosaic virus. *Nature Biotechnology* 13, 1458–1465.

Relevant Websites

<http://www.isaaa.org/>
ISAAA.

Plant Resistance to Viruses: Natural Resistance Associated With Dominant Genes

Mandy Ravensbergen and Richard Kormelink, Wageningen University and Research, Wageningen, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

This is an update of P. Moffett, D.F. Klessig, Plant Resistance to Viruses: Natural Resistance Associated with Dominant Genes, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00469-6.

Nomenclature

AGO Argonaute protein
Avr Avirulence
CC Coiled-coil structure
DCL Dicer-like
dsRNA Double-stranded RNA
EDS Enhanced disease susceptibility
HR Hypersensitive response
LRR Leucine-rich repeat
MP Movement protein
mRNA Messenger RNA
NB-ARC Nucleotide binding adaptor

NLRC Nod-like receptor
nt Nucleotide(s)
PAD Phytoalexin deficient
R Dominant resistance
RDR RNA-dependent RNA polymerase
RISC RNA-induced silencing complex
RITS RNA-induced transcriptional silencing
RNAi RNA interference
TIR Toll and interleukin receptor
VIGS Virus-induced gene silencing
VSR Viral suppressor of RNA silencing
ZAR Hopz-activated resistance

Glossary

Apoptosome An oligomerized form of animal R proteins, for example of APAF1 and CED4, leading to activation of apoptotic initiator caspases.

Avirulence gene Any pathogen gene whose protein product induces a resistance response in a plant possessing a matching resistance gene, thereby rendering the pathogen unable to establish an infection.

Gene-for-gene resistance A genetic paradigm whereby the outcome of a plant–pathogen interaction is determined by the genotype of both organisms. Plant disease resistance proteins confer resistance to pathogens possessing a corresponding avirulence gene.

Hypersensitive response A type of programmed cell death induced by a resistant plant in response to pathogen infection. Cell death is normally contained to infected cells.

Inflammasome An oligomerized form of animal R proteins, for example of NLRC4, leading to activation of inflammatory caspases.

NB-LRR protein Plant disease resistance proteins containing a central nucleotide-binding domain (NB) and C-terminal leucine-rich repeat (LRR) domain. Also known as NBS-LRR, NB-ARC-LRR or NLR proteins.

Resistance gene Plant genes which confer resistance to specific pathogens. Resistance genes confer resistance to pathogens possessing a corresponding avirulence gene.

Resistosome An oligomerized form of activated plant NB-LRRs, recently identified for ZAR1, bearing structural resemblance to animal apoptosomes and inflammasomes.

RNA interference A broad antiviral defense system used by plants, animals, and fungi. It recognizes double-stranded RNA as a cue of viral presence and induces degradation or silencing of viral transcripts or genomes.

Virulence gene Any pathogen gene whose function contributes to increased pathogen proliferation or disease symptoms on an infected plant.

Introduction

Host plants possess polymorphic resistance to different strains or races of pathogens, which is controlled by a single gene. Disease resistance controlled by recessive genes is often a passive form of resistance wherein the pathogen is unable to utilize the host-cell machinery. Dominant resistance (*R*) genes in plants appear to actively recognize the presence of specific pathogens and initiate responses to counteract infection. Usually, this involves a type of programmed cell death termed hypersensitive response (HR). However, this resistance is dependent on the pathogen possessing a corresponding gene that matches the *R* gene. These pathogen avirulence (*Avr*) genes are also dominant over their cognate virulent (*avr*) alleles. Since the outcome of an attempted infection is dependent on the genotype of both host and pathogen, this phenomenon is known as gene-for-gene resistance. Gene-for-gene resistance was first described over 60 years ago, with the discovery of various *R* genes against specific races of flax rust, caused by the fungus *Melampsora lini*, among different flax genotypes. Ever since, gene-for-gene resistance has been observed for many plant pathogens including fungi, oomycetes, bacteria, insects, nematodes, and viruses.

R Gene Products

R genes have been genetically characterized for many years and have been incorporated into breeding programs of multiple crops. To study the structure and function of the *R* gene encoded proteins, numerous *R* genes have been cloned. These proteins belong to a small number of protein classes including transmembrane proteins possessing extracellular leucine-rich repeat (LRR) domains as well as a small number of intracellular serine/threonine kinases. The most abundant class of *R* genes encodes intracellular proteins containing a central NB-ARC (nucleotide binding adaptor shared by APAF-1, certain *R* gene products, and CED-4) domain as well as a C-terminal LRR domain. Additionally, these proteins contain an N-terminal signaling domain, either displaying homology to the Toll and interleukin receptors (TIR domain) of animals or containing a coiled-coil structure (CC). These proteins are often denoted as CC-NB-LRR or CNL, and TIR-NB-LRR or TNL, collectively known as NB-LRR or NLR. Here, CC-NB-LRR and TIR-NB-LRR, collectively NB-LRR, will be used.

Sequencing projects have revealed that plant genomes contain large numbers of genes encoding NB-LRR proteins. For example, the *Arabidopsis* genome contains 149 such genes, whereas the cassava genome encodes 327, 99 of which are lacking an LRR domain. The polymorphic nature of host plant resistance is attributed to the polymorphic NB-LRR gene loci, with different alleles recognizing different pathogens or pathogen races. Furthermore, the proteins encoded are highly divergent both within and between species, presenting a large potential for recognition of pathogen diversity. *R* genes recognizing a specific pathogen could be identified by testing many genotypes of a certain species. For crop species, resistance sources are often identified in, and introgressed from, wild relatives.

Molecular studies have shown that recognition specificity is in most cases determined by the LRR domain of NB-LRR proteins. In certain cases, this could be demonstrated through swapping of LRR domains between two similar NB-LRR proteins. Correspondingly, the LRR domain is the most polymorphic region of NB-LRR proteins. Bioinformatics studies have suggested that the LRR domains are under diversifying selection, suggesting that this class of *R* genes is under selective pressure. There do not appear to be any obvious features that distinguish between those *R* genes that recognize, for example, bacteria versus those that recognize viruses. Within *Arabidopsis*, for example, the *HRT*, *RCY1* (*Resistance to CMV* (Y)), and *RPP8* genes are alleles of the same locus, but confer resistance to Turnip crinkle virus (TCV), Cucumber mosaic virus (CMV), and the oomycete *Hyaloperonospora parasitica*, respectively. Similarly, in potato, gene-duplication events have led to highly similar gene paralogues, *Rx1* and *Gpa2*, which confer resistance to Potato virus X (PVX) and to the nematode *Globodera pallida*, respectively. Additionally, PVX can be engineered to express a bacterial *Avr* gene and a plant expressing the corresponding NB-LRR gene will be resistant to the recombinant virus. Thus, most NB-LRR genes probably have the potential to evolve specificity to a variety of different pathogen types, allowing for maximum adaptability. As such, the mechanisms of recognition and initiation of disease-resistance responses are likely to be similar between *R* proteins that recognize viruses and *R* proteins that recognize other types of pathogens.

R Gene Responses to Viruses

Many different types of pathogen proteins are either synthesized in the host cell by intracellular pathogens (viruses) or delivered into the cytoplasm by the various secretion systems of bacteria, fungi, or oomycetes. Here, they can be recognized by NB-LRR proteins. In absence of a corresponding *R* gene, *Avr* proteins generally act as virulence factors, protect the pathogen from host defense responses, actively inhibit host defenses, or are involved in replication of the pathogen. Pathogen *avr* alleles can overcome *R*-gene-mediated resistance either by mutation or by elimination of the appropriate *Avr* gene. However, this often leads to a loss of fitness. Given their small genomes, this is particularly acute for viruses as most viral genes are crucial for viral fitness and often have multiple functions. An overview of cloned NB-LRR genes of which the *Avr* determinant has been identified is presented in Table 1. Anti-viral NB-LRR genes that have not been cloned yet or have an unknown elicitor can be found in recent extensive reviews, for example by de Ronde *et al.*

One of the most studied antiviral *R* genes is the *N* gene. It was introduced into commercial tobacco cultivars from *Nicotiana glutinosa* via interspecific crosses as a means of conferring resistance to tobamoviruses, primarily Tobacco mosaic virus (TMV). The *N* gene was among the first *R* genes to be cloned and encodes a typical TIR-NB-LRR protein. Like several TIR-NB-LRR-encoding genes, the *N* gene produces two different transcripts via alternative splicing, one encoding the full protein and a shorter version encoding a protein lacking the LRR domain. The role of these truncated proteins is unclear; plants transgenic for versions of the *N* gene that cannot produce both splice forms are compromised in their resistance to TMV. The *N* gene confers resistance against a wide range of tobamoviruses, with the exception of a virus originally designated TMV-ob and renamed Obuda pepper virus (ObPV). The *Avr* effector of the *N* gene resides in the TMV viral replicase protein, as demonstrated through gene swaps between TMV and ObPV. Expression of a helicase domain (p50) of the replicase is sufficient to induce HR in *N* plants. The HR induced by *N* in response to TMV is visible as small necrotic lesions at the site of infection. In this case, the virus is able to replicate to a limited degree, spreading to cells surrounding the initially infected cell before the onset of cell death. Initiation of HR requires only the presence of a matching pair of *R* and *Avr* proteins. In the past decades it has become evident that cell death caused by HR does not equate the resistance response. Follow-up studies indicated that the *N* gene based antiviral response leads to a translational arrest of viral transcripts by a process involving Argonaute 4 (AGO4), ultimately preventing virus accumulation and spread. Additionally, in cases of HR-induction with *Rx1*, this response can be knocked out without affecting *Rx1*-mediated resistance against PVX. Similar observations have been made by others and suggest that the actual resistance response is different from an HR, although often both are triggered. It follows that HR continues to serve as a valuable indication of NB-LRR mediated resistance.

Table 1 Overview of cloned and characterized antiviral NB-LRR genes. Columns list the plant species in which the *R* gene has been introgressed (plant), the domain facilitating oligomerization, the viral Avr determinant and the resistance response

<i>R</i> gene	Encoded protein	Plant	Homo-oligomerization domain	Virus	Avr determinant	Resistance response ^a
N	TIR-NB-LRR	Tobacco	NB-ARC	Tobamoviruses	Helicase/p50 subunit	HR
Rx1	CC-NB-LRR	Potato	— ^b	PVX	CP	ER
Rx2	CC-NB-LRR	Potato	ND	PVX	CP	ER
Sw-5b	CC-NB-LRR	Tomato	NB-ARC	Tospoviruses (TSWV, GRSV, TCSV)	NSm	ER, HR
HRT	CC-NB-LRR	Arabidopsis	ND	TCV	CP	HR
RCY-1	CC-NB-LRR	Arabidopsis	ND	CMV	CP	HR
Tm2	CC-NB-LRR	Tomato	ND	Tobamoviruses	30 kDa MP	ER
Tm22	CC-NB-LRR	Tomato	ND	Tobamoviruses	30 kDa MP	ER

^aCertain resistance genes condition more than one type of response depending on viral isolate, genetic background, or environmental conditions.

^bRx1 was found to heterodimerize with RanGAP2, there is no evidence for homodimerization.

Note: ND, not determined.

For some plant-virus interactions, HR is observed although a strict gene-for-gene relationship has not been defined due to a lack of variability in either the host or the pathogen. For example, eggplant undergoes a typical HR response when inoculated with tobamoviruses. A compatible interaction was not found either because the corresponding *R* gene is present in all cultivars of eggplant, or because not enough genetic material has been tested. This raises the possibility that some cases of non-host resistance, where all accessions are resistant to all isolates of a pathogen, may in fact be controlled by the same mechanisms as gene-for-gene resistance.

Aside from the typical HR, an interaction between an Avr determinant and an *R* gene encoded protein can result in two other comparable responses: extreme resistance (ER) and systemic necrosis (SN). ER is broadly defined as a lack of visible response to virus infection. In an ER response there is no necrosis and little or no virus accumulation can be detected. SN is caused by the capability of the virus to evade the initial resistance response and spread to systemic tissues. However, the accumulation of virus here eventually leads to the initiation of cell death. This response is not sufficient to contain the virus, causing a trailing HR that eventually leads to collapse of the plant. The differences between these three responses, HR, ER, and SN, are suggested to be quantitative rather than qualitative, as illustrated by the following examples.

The *Rx1* and *Rx2* CC-NB-LRR genes in potato confer an ER response to PVX, upon recognition of the same domain of the coat protein (CP). However, co-expression with the PVX CP results in an HR. The ER responses induced by *Rx1* and *Rx2* are thus not qualitatively different from HR, but rather eliminate the source of the Avr protein before it accumulates to HR inducing levels. Varying recognition efficiency of an NB-LRR protein between Avr effectors of different viral strains can result in a differential response, as in the case of Sw-5b. This tomato CC-NB-LRR recognizes the NSm protein of Tomato spotted wilt virus (TSWV), Tomato chlorotic spot virus (TCSV), and Groundnut ringspot virus (GRSV). Depending on the strain used, either HR or ER responses can be seen. The *Arabidopsis* CC-NB-LRR gene *HRT* (HR to TCV) recognizes the TCV CP, but in roughly 10% of cases the virus escapes the initial HR and replicates in systemic tissue, leading to SN. In absence of a second recessive gene, *mt* (regulates resistance to TCV), this effect is much more pronounced. Although as of yet unidentified, this gene is expected to elevate *HRT* transcript levels in the highly resistant line Di-17. Additionally, the *Rsv1* gene of soybean recognizes the viral Avr determinant of different Soybean mosaic virus (SMV) strains with varying efficiency, leading to ER, systemic infection, or SN, depending on the viral isolate. In the case of the *I* TIR-NB-LRR gene in bean, defense responses range from ER to HR to SN depending on the potyvirus isolate. The defense response is both dependent on temperature and has a gene dosage effect. At 23°C plants homozygous for the *I* gene (*I/I*) confer ER to Bean common mosaic virus (BCMV), whereas heterozygous plants (*I/i*) confer an HR-type of resistance. At 34°C, SN is seen on *I/i* plants, but to a lesser degree on *I/I* plants.

Thus, the different responses to viruses by resistant plants likely represent a continuum of quantitatively different responses. The degree of response can be modulated by the relative efficiency of recognition of the Avr determinant, *R* gene dosage, temperature, and genetic background.

Mechanisms of Recognition

Given the relatively straightforward genetic relationship between *R* and *Avr* genes, it was originally proposed that this might represent a receptor–ligand interaction. Since LRR domains often act as protein–protein interaction domains, it was proposed that the variable LRR domains of NB-LRR proteins would bind to different Avr proteins and this binding would activate defense responses. However, initial difficulties in demonstrating such interactions led to the formulation of indirect models of recognition. The guard hypothesis states that, since many *Avr* genes are also virulence genes, *R* proteins would ‘guard’ their cellular targets. The pathogen effector would modify cellular proteins, which is in turn detected by *R* proteins. Several examples of indirect recognition have been studied. For example, when expressed in *Pseudomonas syringae*, the Avr proteins AvrB and AvrRpm1 are both recognized

by the *Arabidopsis* CC-NB-LRR protein resistance to *P. syringae* pv *maculicola* 1 (RPM1). RPM1 interacts with rpm1 interacting protein 4 (RIN4), a cellular protein which in turn can interact with both AvrB and AvrRpm1. This latter interaction is thought to activate the RPM1 protein. The Avr protein AvrRpt2 from *P. syringae* is a cysteine protease and is recognized by the *Arabidopsis* CC-NB-LRR protein resistant to *P. syringae* 2 (RPS2). RPS2 also binds RIN4 and becomes activated when RIN4 is cleaved by AvrRpt2. In the absence of RIN4, for example through genetic ablation, RPS2 is constitutively activated. In this case, RIN4 thus appears to function as a negative regulatory protein.

At the same time, however, there is some indication that certain Avr/R protein pairs may interact directly. Yeast two-hybrid studies have shown Avr proteins interacting with the rice Pi-Ta, *Arabidopsis* RRS1 (resistant to *Ralstonia solanacearum* 1), and the flax L NB-LRR proteins which confer resistance to rice blast fungus, *R. solanacearum*, and flax rust, respectively. Direct interactions have also been reported between the viral R protein N and its cognate elicitor p50, surprisingly mediated by the TIR domain of N. However, a great deal about these interactions remains unclear, such as whether the other Avr proteins interact with the LRR domain, another part of the protein, or a combination thereof.

Two traits desired by plant breeders related to recognition are broad-spectrum resistance and durability. Often, *R* genes are deployed that provide good resistance but are quickly overcome by new strains of a pathogen. The durability of an *R* gene may be affected by different aspects of the pathogen such as dispersal rates and mechanisms, occurrence of sexual reproduction, and variable repertoires of virulence factors. However, since viruses cannot dispense with any of their limited number of genes, overcoming resistance must occur through sequence changes. Several cases have been studied where the virus *Avr* gene has mutated to a point it is no longer recognized by the corresponding *R* gene. However, this often results in a loss of viral fitness due to the detrimental effect of these mutations, as illustrated by the examples below.

The CC-NB-LRR encoding genes *Tm-2* and *Tm-22* have been introgressed from different *Lycopersicon peruvianum* accessions into tomato to control TMV and Tomato mosaic virus (ToMV). These genes confer resistance to nearly all strains of tobamoviruses through recognition of the viral movement protein (MP). Molecular cloning has revealed that *Tm-2* and *Tm-22* are different alleles of the same gene. They differ by only four aa, two of which are in the LRR domain. Despite this similarity, *Tm-22* has proven to be more durable than *Tm-2*. Strains of ToMV that overcome *Tm-2* or *Tm-22* have mutations in different regions of the MP. The *Tm-22* gene seems more durable because the resistance-breaking mutations have a very negative impact on virus fitness, resulting in reduced viral load in plants infected with these strains. The *Tm-22*-breaking strains cause very severe symptoms on *Tm-22* plants so that infected plants can be easily recognized and eliminated from the field or greenhouse. Other *R* genes that are very durable are *N*, *Rx1*, and *Ry*. There are no strains of Potato virus Y (PVY) that overcome *Ry* and although ObPV can overcome *N*, it is not widespread. Likewise, *Rx1*-breaking PVX strains have reduced virulence and are found only in South America. Less durable genes include *Nx* and *Nb*, where PVX strains that overcome this resistance are more widespread. Additionally, TSWV strains breaking Sw-5b resistance have been reported in several tomato-growing areas.

Not all mutations lead to a loss of fitness for the virus, for example in the case of differential recognition by the *N'* gene of tobacco. The *N'* gene confers resistance to several tobamoviruses including ToMV, but not TMV. Although *N'* has not been cloned, the Avr determinant has been shown to be the viral CP. Curiously, mutations can be introduced into the TMV CP that result in a gain of recognition by *N'*, but which appear to compromise CP function. These mutations are hypothesized to interfere with the tertiary or quaternary structure of the CP, exposing an *N'* recognition site previously masked. Differential recognition of viral Avr proteins by *R* proteins can also be seen for the products encoded by the allelic pepper genes *L1-L4*, also recognizing the CP of tobamoviruses. Certain virus strains are able to evade recognition of *L1*, and thus *L1*-mediated resistance, some overcome *L1* and *L2*, and some overcome *L1*, *L2*, and *L3*. However, the *L4* gene is very broad spectrum, controlling all known tobamoviruses.

Signaling Mechanisms

Studies with *Rx1*, *HRT*, *Sw-5b* and *N* have shown that the different domains of NB-LRR proteins undergo physical intramolecular interactions with the LRR and N-terminal domains interacting with the NB domain. These interactions appear to condition an auto-repressed state and alterations thereof are associated with activation of the protein.

In the presence of PVX CP, the interdomain interactions of *Rx1* are disrupted and the *N* protein has been shown to undergo oligomerization in the presence of p50. Both *N* and *Rx1* shuttle between the nucleus and the cytoplasm, and both depend on their nuclear localization in order to induce a defense response. Interestingly, p50 is not required in the nucleus to elicit the *N* mediated defense response, but does so even when restricted to the cytoplasm, and *Rx1* can only be activated in the cytoplasm. Similarly, the barley NB-LRR *MLA10*, and the *Arabidopsis* NB-LRRs *RRS1*, *RPS4*, and *SNC1* (suppressor of *npr1-1*), are all at least to some extent localized to the nucleus. Of these, *MLA10*, *RPS4*, and *SNC1* show a reduced immune response upon induced relocation from the nucleus to the cytoplasm. Taken together, these NB-LRRs provide evidence towards a nuclear signaling component, concurrent with one of the first steps of the immune response pathway being modification of transcription.

In vitro studies have shown that *Rx1* binds unselectively to DNA after activation through association with the PVX CP. It induces melting and bending of DNA, leading to structures resembling transcription initiation complexes. *Rx1* most likely interacts with a Golden2-like (Glk) transcription factor, as demonstrated by its binding to *Nicotiana benthamiana* Glk1 in a transient transformation assay. In *Arabidopsis*, Glk1 and Glk2 are involved in a defense response against CMV, possibly in a redundant manner. In a double mutant, induction of defense related genes is hampered. In the current *Rx1* resistance model, *Rx1* and Glk1 associate in a

complex and bind to DNA through Rx1 mediated DNA binding activity. Upon infection, the CP induces conformational changes in the Rx1-Glk1 complex, exposing the Glk1 DNA binding domain, previously blocked by the Rx1 CC domain. Glk1 binds stronger to its consensus DNA binding motifs, providing specificity to the induced transcriptional changes in the defense response. Concomitant with the role of Glk1 as an immune activating transcription factor, overexpression of Glk1 without Rx1 is sufficient to induce ER to PVX.

It is presumed that other R proteins induce a signal transduction cascade leading to disease resistance by interacting with signal adaptor molecules as well. Although these molecules have yet to be identified, a number of proteins required for NB-LRR function are known. TIR-NB-LRR proteins, including N, have a general requirement for the EDS1 (enhanced disease susceptibility 1) protein and often also require the EDS1-interacting protein PAD4 (phytoalexin deficient 4). Both of these proteins have lipase signature domains, of which no exact function is known. Given that the *N* gene is not effective above 32°C, it is interesting to note that *EDS1* expression levels are greatly reduced at higher temperatures in *Arabidopsis*. Likewise, the barley MLA NB-LRR proteins accumulate to greatly reduced levels at high temperatures. These phenomena may at least partially explain the temperature sensitivity of some *R* genes.

A number of studies have shown that during the resistance response, kinase cascades are activated, and transcription is reprogrammed substantially. Studies using virus-induced gene silencing (VIGS) have implicated several protein kinases, including MEK1, MEK2, NTF6, SIPK, WIPK, and NPK1, in *N* gene-mediated resistance. Additionally, successful *N* mediated resistance to TMV requires Beclin 1, Atg3 (autophagy related 3), Atg7, and Pi3k/Vps34 (class III phosphoinositide 3-kinase). These proteins are involved in the autophagy pathway, one of the major cellular degradation pathways in eukaryotes. Silencing these genes leads to programmed cell death beyond the infected tissue, thus for *N* mediated resistance the autophagy pathway appears involved in limiting the HR to the site of infection. Autophagy has been implicated in many plant processes, including innate immunity. Furthermore, the mammalian nod-like receptor 2 (NLRC2) activates antiviral innate immune signaling, using autophagy to regulate inflammatory responses. A requirement was also shown for the TGA and WRKY families of transcription factors as well as the individual transcription factors MYB1 (myb domain protein 1) and the TGA-interacting protein NPR1 (nonexpresser of *pr* genes 1). These proteins and protein families have also been implicated in gene-for-gene resistance to other types of pathogens. Gene-for-gene resistance may also involve targeted proteolysis since silencing the expression of certain proteins involved in regulated protein degradation, including a subunit of the COP9 (constitutive photomorphogenesis 9) signalosome, SKP1 (S phase kinase-associated protein 1 phase kinase-associated protein 1) and COI1 (coronatine insensitive 1), also compromises *N*-mediated resistance. To date, the only genes whose silencing has been shown to break Rx1-mediated resistance are *Hsp90* (*heat shock protein 90*) and *Sgt1* (*suppressor of the G2 allele of skp1*). Along with the *Sgt1* binding protein Rar1 (required for MLA12 resistance 1), these proteins are required for the function of a number of NB-LRR proteins. The *Hsp90*, *Sgt1*, and *Rar1* proteins are likely chaperoning proteins in order to allow proper protein folding.

The plant hormone ethylene plays a role in *R*-gene-mediated antiviral responses. RCY1 functions at reduced efficiency in *Arabidopsis* containing the ethylene-desensitizing mutations *ein3* and *etr1*. Salicylic acid (SA) also plays a role in *R*-gene-mediated resistance. The involvement of SA is often studied using SA biosynthetic mutants in *Arabidopsis* (such as *eds5*) or by using plants transgenically expressing the bacterial *nahG* gene, which encodes a salicylate hydroxylase that degrades SA. The *eds5* mutation and *nahG* transgene both decrease the efficiency of RCY1-mediated resistance and completely abrogate HRT-mediated resistance. When tobacco plants expressing *N* and *nahG* are infected with TMV, the resulting HR lesions are larger, and necrosis can spread beyond the infected leaf. Similarly, in *nahG* transgenic tomato the normal ER response of Tm-22 becomes a spreading HR phenotype in the inoculated leaf, although the virus does not spread systemically. The Rx1-mediated response, however, is not affected by the *nahG* transgene. SA may enhance *R* gene responses rather than directly targeting the virus, and its requirement for resistance may depend on the strength of the *R*-gene-mediated response.

In agreement with this, SA application results in reduced lesion size in the *N*-mediated response to TMV and an enhancement of HRT-mediated resistance. Cell death during the HR induces increased levels of SA. Thus, cell death may indirectly prime and/or augment *R*-gene-mediated antiviral responses. It is clear that cell death itself is not the only mechanism by which virus is eliminated as ER does not involve cell death, and in the *N*-TMV interaction, TMV can be detected outside the area of necrosis. Furthermore, in the interaction between Cauliflower mosaic virus (CaMV) and *Nicotiana spp.*, resistance is controlled by a single dominant gene. However, whether or not this resistance is accompanied by an HR is controlled by a separate gene. A clear understanding of how viruses are cleared from resistant plant cells requires further study.

Resistosome

Although multiple individual NB-LRR genes and their corresponding Avr effectors have been both identified and studied, the pathway of defense response induction in plants remains elusive. In animal systems, these mechanisms are often elucidated to a higher extent. Translating this information to plant systems could provide a valuable resource. In animal systems, one of the key factors required for resistance mediated by several *R* genes was found to be oligomerization. The highly structured complexes are referred to as either an inflammasome or apoptosome. Three well-studied examples include the human R protein apoptotic protease-activating factor 1 (APAF1), the nematode R protein cell death protein 4 (CED-4), and the human protein NLRC4. Within plants, although a similar resistance mechanism was hypothesized in 2010, only one R complex has been identified, as recent as 2019. This work involved the CC-NB-LRR protein hopz-activated resistance 1 (ZAR1), conferring resistance within *Arabidopsis* and

N. benthamiana to disease caused by the bacteria *Xanthomonas campestris*. The formation of the ZAR1 complex, termed resistosome, is essential for resistance induced by the *X. campestris* effector AvrAC.

The resistosome structurally resembles the inflammasome and the apoptosome. All three consist of a wheel-like basis, and the NB-ARC domain is similarly positioned between ZAR1, APAF1, and NLRC4 to facilitate oligomerization (Fig. 1). Furthermore, although ZAR1 oligomerization requires certain co-factors, these are not part of the interacting domains between the sub-units, similar to the co-factor cytochrome C in the Apaf1 apoptosome. Important differences include the organization of the N-terminal domains. The Apaf1 apoptosome and the NLRC4 inflammasome contain disorganized caspase recruitment domains (CARDs), in the absence of certain caspases. These caspases are interacting partners required for downstream processing of the defense signal. However, the α helices in the CC domain of the ZAR1 resistosome are highly structured. Furthermore, the number of sub-units contributing to the complex differs from five in the ZAR1 resistosome, to seven for the Apaf1 apoptosome, nine for the CED4 apoptosome, and eleven for the NLRC4 inflammasome.

Assembly of the ZAR1 resistosome requires the presence of two other proteins: resistance-related kinase 1 (RKS1) and PBS1 like protein 2 (PBL2), as well as the exchange of an ADP for ATP. Current evidence suggests ZAR1 and RKS1 form a dimer within the cytoplasm of the plant cell. Upon infection by the *X. campestris* bacteria, the AvrAc protein modifies the host protein

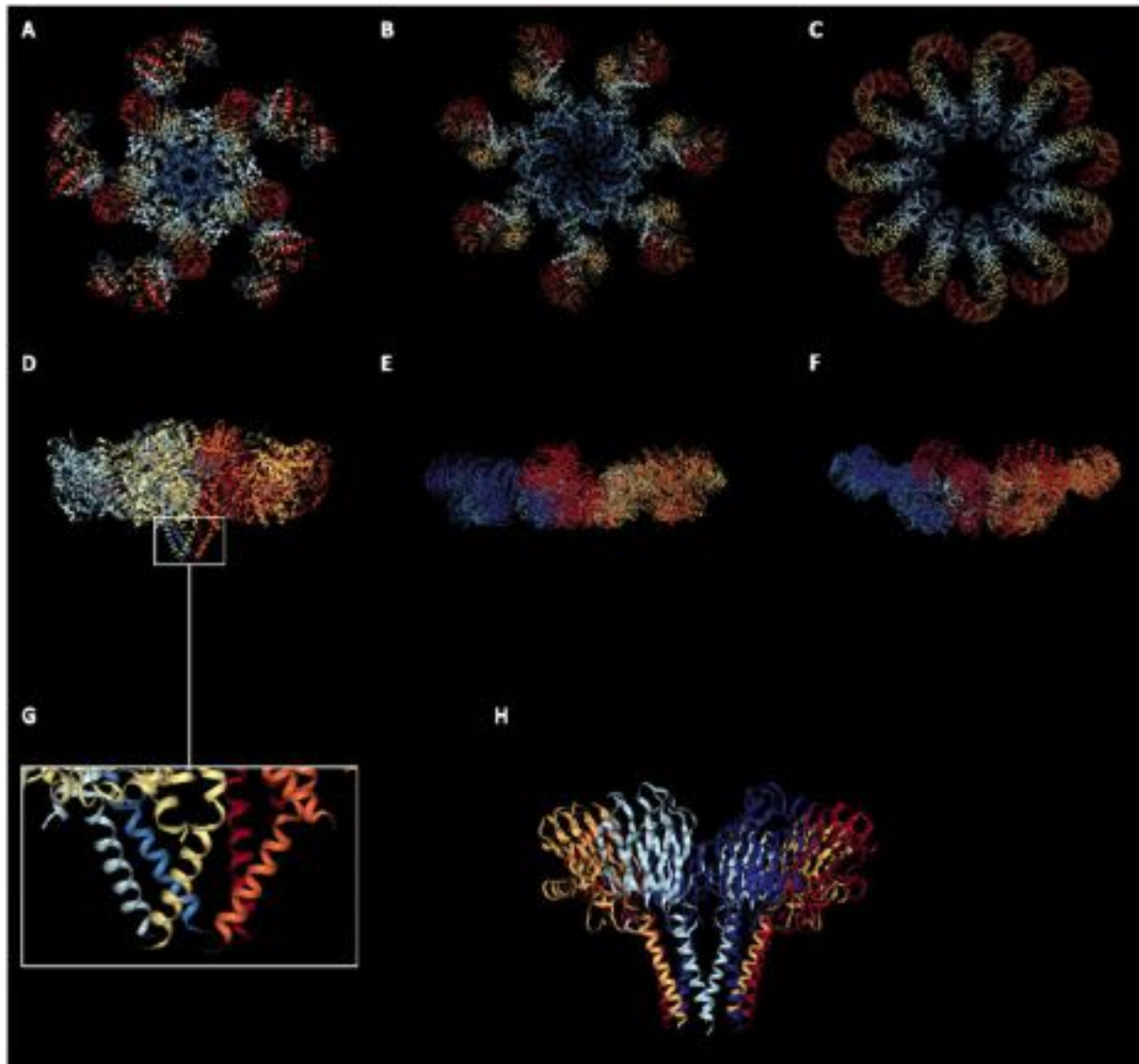


Fig. 1 Protein structure of the ZAR1 resistosome (A, D, G), the APAF1 apoptosome (B, E), the NLRC4 inflammasome (C, F), and the actinoporin FraC (H). Colors indicate the transition from N (blue) to C terminal (red) (A–C) or the different subunits in the complex (D–H). A–C show top views of the complexes, and D–H show side views. G shows a detail of the α helical bundle of the ZAR1 complex in D, for comparison to the α helical bundle of FraC in H.

PBL2 by the addition of a uridine monophosphate. This modified PBL2 binds to the ZAR1-RKS1 dimer, promoting the release of ADP from this complex. The subsequent binding of ATP leads to the oligomerization of this complex in a resistosome consisting of five ZAR1-RKS1-PBL2 subunits. Interestingly, the oligomerization towards the resistosome seems to differ from the activation and assembly of the human NLRP3 inflammasome, where a potassium efflux seems required in almost all cases. Although the ZAR1-RKS1 dimer is only to a very limited degree localized in the plasma membrane, resistance mediated by the resistosome complex requires its presence in the plasma membrane. To understand the translocation of the resistosome, a closer look must be taken to the rearrangements of the different ZAR1 domains during activation.

The NB-ARC domain of ZAR1 can be further divided in a NB domain, HD1 (helicase domain 1), and WHD (winged helix domain). Rearrangements within the CC domain lead to the exposure of the α 1 helix, from a position buried and contacting the LRR and WHD domain, to a position protruding from the rest of the protein. Upon oligomerization, the α 1 helices form a highly structured funnel shape. This most likely facilitates the association to the plasma membrane. Mutation of several aa in this α 1 helix, as well as adding a FLAG peptide, leads to a loss of plasma membrane association. Although these perturbations of the α 1 helix lead to loss of ZAR1 resistance, they do not affect the formation of the ZAR1 resistosome. Interestingly, this funnel shaped structure in the ZAR1 resistosome resembles that of the actinoporin fragaceatoxin C (FraC), although the proteins bear little homology. Actinoporins like FraC are pore-forming complexes found in sea anemone species, used both in defense and as a hunting strategy. FraC attaches to the plasma membrane, and oligomerizes into a complex consisting of nine subunits. This probably changes the protein conformation leading to the detachment of the N terminal α helix, and subsequent insertion of the α helices into the membrane, forming a highly structured α helical bundle. After the transition from a prepore to a pore, the membrane is no longer capable of maintaining ion balance. The presence of the nonselective hole leads to an efflux of ions and small molecules, depending on the pore size, and ultimately leads to cell injury or death. Aside from the similarities in structure between the α helical funnels of the ZAR1 resistosome and the FraC actinoporin complex, their size is also comparable, at 33 and 35 ångström, respectively. This could indicate that the ZAR1 funnel is also membrane spanning, rather than membrane inserted. Following this hypothesis, the ZAR1 resistosome could function through disturbance of the membrane integrity in a similar manner as the FraC complex, leading to localized cell death and limitation of pathogen spread. This is supported by the observation that mutating the inner surface of the ZAR1 resistosome funnel abolishes disease resistance. In animal cells, a similar route of using pore complexes to mediate a cell death response exists. For example, in humans and mice, inflammatory caspases oligomerize upon activation into an inflammasome. One such example is caspase 11, which cleaves gasdermin D (GSDMD), in a C-terminal and N-terminal part. After cleavage, GSDMD-NT oligomerizes within the cell membrane and forms pores, leading to cell death. For now, the exact resistance mechanism of ZAR1 remains unidentified, and future research will have to show whether ZAR1 functions as a pore complex as hypothesized, or through a different mechanism.

Non-NB-IRR Resistance

Although by far most identified *R* genes belong to the NB-LRR class, other classes of *R* genes should not be ignored or overlooked. In fact, breeding for resistance mediated by NB-LRR genes may prove less durable than breeding for resistance conferred by non-NB-LRR genes. This inherent disadvantage attributed to NB-LRRs is caused by the gene-for-gene relationship NB-LRRs have with pathogen Avr determinants. This puts a relatively large selection pressure on the pathogen. Any mutation in the pathogen effector leading to less recognition by the NB-LRR increases the pathogen's fitness considerably, as it is now able to infect a new host genotype. The likelihood of the frequency of this mutation rising within the population is therefore considerable, and the breeder would have to start over again by re-introducing resistance within the crop plant. Non-NB-LRR *R* genes could help alleviate this disadvantage, as often the principle of gene-for-gene resistance does not apply here. Instead, the underlying mechanism of resistance involves a more complex interplay of the pathogen and host, or the perception of more generic signals of the pathogen. Although less emphasis is placed on non-NB-LRRs *R* genes, they represent interesting sources to confer more durable (and sometimes broader) crop resistance.

One such example can be found within the major plant anti-viral RNA interference (RNAi) defense mechanism. RNAi is triggered by a highly conserved cue of viral presence in the plant cell: double stranded RNA (dsRNA). These dsRNAs are formed during viral replication, either through overlapping open reading frames (ORFs), from highly structured viral mRNAs, or as a replication intermediate of RNA viruses. During the life cycle of a healthy plant cell, dsRNA is rarely formed, rendering this form of RNA a distinctive signal to the plant cell of pathogen presence. Dicer-like (DCL) enzymes cleave these RNA duplex molecules in small fragments, typically 20–25 nt in size. One strand of these small interfering RNAs (siRNAs) is then loaded into an AGO containing RNA-induced silencing complex (RISC). Any RNA complementary to this strand is targeted for degradation, giving rise to aberrant RNA molecules, or translational arrest. During an amplification cycle, α type RNA dependent RNA polymerases (RDRs) synthesize a full-length dsRNA molecule using the single-stranded aberrant RNA molecule as a template. This full-length dsRNA is cleaved by one of the DCLs, generating a population of secondary siRNAs. These can either be used again to target complementary RNAs for silencing or spread through the plasmodesmata between plant cells, leading to immunization beyond the infected cell. Upon infection by a DNA virus, siRNAs are alternatively loaded on an RNA-induced transcriptional silencing (RITS) complex, containing a distinct AGO protein. This complex targets complementary DNA for methylation, leading to a block of transcription of viral DNA.

Table 2 Overview of cloned antiviral non-NB-LRR genes. The resistance mechanism is indicated, as well as the species in which the *R* gene is used (plant), and the viral elicitor

Resistance mechanism	<i>R</i> gene	Encoded protein	Plant	Virus	Elicitor
RNAi	Ty-1	RDR γ	Tomato	TYLCV	dsRNA
Restricted systemic movement	RTM1–5 ^a	Jacalin, HSP20, MATH, ND, ND	Arabidopsis	Potyviruses (TEV, PPV, LMV)	ND
Lectin-mediated resistance	JAX1	Jacalin	Arabidopsis	Potexviruses (PIAMV, PVX, Asparagus virus 3, White clover mosaic virus)	Replicase
Inhibition RNA replication	Tm-1	TIM-barrel-like domain	Tomato	ToMV	130K and 180K replication proteins

^aRTM mediated resistance requires the presence of all 5 genes.

Recently, *Ty-1*, one of the major dominant resistance genes against the geminivirus TYLCV (Tomato yellow leaf-curl virus) in tomato, has been cloned, and was shown to encode a member of the γ subclass of RDRs (Table 2). Previously, members of this class had not been assigned any function. The main difference between the well characterized α type and the γ type lies within the catalytic core domain, the α type containing a DLDGD motif and the γ type a DFDGD motif. *Ty-1* has been shown to confer resistance, still allowing very low levels of viral DNA replication, by enhancing the RNAi response. Specifically, it strengthens transcriptional gene silencing of viral DNA. As opposed to an NB-LRR recognizing one pathogen Avr effector, *Ty-1* potentially entails a more durable *R* gene and confers resistance to a wider range of mono-/bipartite geminiviruses, besides TYLCV. The achilles heel of *Ty-1*-mediated resistance lies in co-infecting plant viruses with the ability to interfere with antiviral RNAi using proteins called viral suppressor of RNAi (VSR).

Another interesting example can be found in the *Arabidopsis* *R* genes *Restricted tobacco etch potyvirus* (RTM) and *Sieve element lining chaperone 1* (SLI1). RTM mediated resistance requires the presence of five genes, termed RTM1–5. Together, these genes limit the spread of TEV, Plum pox potyvirus (PPV), and Lettuce mosaic virus (LMV) through the plant to systemic tissues. RTM1 encodes the lectin protein jacalin. Several members of this gene family have been shown to provide resistance to insects, fungi, and viruses. RTM2 is an unusual small heat shock protein, containing a C-terminal transmembrane domain. It is neither inducible by heat treatment, nor providing tolerance towards high temperatures. RTM3 belongs to a gene family of unknown function, whose members contain a meprin and TRAF homology (MATH) domain and a C-terminal CC domain. RTM4 and RTM5 remain uncharacterized. Although the exact resistance mechanism is unknown, resistance breaking strains rely on changes in the N-terminal region of the CP. No interaction has been shown between the CP and RTM1–3. Both RTM1 and RTM2 are shown to localize to sieve tubes, rendering it likely that the resistance acts in phloem tissues. Oligomerization of several different factors seems required, as demonstrated by both the direct interaction of RTM1 and RTM3, and the presence of several domains involved in protein-protein interactions in RTM1–3.

An alternative method of introducing resistance against viruses is by combatting vectors transmitting these viruses between plants. *SLI1* is an example of an *R* gene that could be used in such a manner. It has been recently identified and confers resistance in *Arabidopsis* against the generalist aphid *Myzus persicae* through restriction of both feeding time and phloem ingestion rate. *SLI1* limits the growth of the aphid population on the host plant, and in this manner *SLI1* could aid in limiting virus spread throughout the plant population. As opposed to RTM2, *SLI1* confers tolerance to higher temperatures. *SLI1* is lined around the entire sieve elements, including the sieve plate, and specifically around mitochondria in a peripheral strip of cytoplasm. Interestingly, RTM2 and *SLI1* are homologs belonging to the same superfamily of heat shock protein 20 (HSP20)-like chaperones.

Outlook

With the recent discovery and elucidation of the ZAR1 resistosome a new chapter has opened on the functioning of NB-LRR resistance proteins. Whereas earlier research has indicated that the actual resistance response is different from HR, findings on the ZAR1 resistosome, most likely functioning as a membrane pore, implies that cell death is an essential and final stage of the resistance response/mechanism. This (again) forces us to re-visit our views on the resistance mechanism. If the hypothesized function of ZAR1 holds true, how can HR and the so-called actual resistance response be uncoupled? Maybe on this point the resistance mechanism consists of sequential layers, with the Avr being the initial trigger to all. The first layer(s) could be pathogen-specific and the formation of a resistosome represents the final stage that is generic and always leads to cell death, irrespective of the type of pathogen. This also would explain why in most cases HR occurs concomitant with the induction of NB-LRR-gene mediated resistance. The lack of HR in several cases could then be a quantitative rather than qualitative matter, similar to the differences observed between HR, ER, and SN responses. Alternatively, this could result from mutations that prevent the *R* protein from oligomerization but not compromise the earlier, upstream steps of the resistance response.

Although speculative, the formation of different oligomers as observed with animal *R* proteins, like inflammasomes and apoptosomes, also raises the question on the existence of different resistosomes in plants. In light of this it is interesting to remark

on differences between plant NB-LRRs. For ZAR1 the CC domain is not only essential for oligomerization, but its overexpression also leads to HR. This is compromised by N-terminal fusions or -deletions at the CC domain. In contrast, for several other NB-LRR genes, not CC but overexpression of NB-ARC (e.g., Sw-5b) or NB (e.g., Rx1) triggers HR. In addition, the folding structure of the ZAR1 CC domain is different from those of stem rust resistance 33 (Sr33), Rx1 and MLA10. This indicates a possible classification of NB-LRRs into subgroups. On this point, a recent study has shown that several NB-LRRs, termed sensors, require the presence of a second NB-LRR, termed helper. These helpers can be divided in three classes: *NRC* (*NB-LRR required for cell-death*), *ADR1* (*Activated disease resistance 1*), and *NRG1* (*N required gene 1*). Interestingly, these helpers can be (partially) redundant, as in the case of NRC2, 3, and 4, while several sensors can share the same helper. Silencing all three NRCs affects, among others, resistance mediated by Sw-5b and Rx1. Subsequent challenge of NRC silenced plants carrying *Rx1* with PVX results in SN, instead of the usual ER response. This suggests NB-LRRs can be divided into at least two sub-groups, those that act without helper and those that require a helper. This could also explain why domains from different NB-LRR proteins are not all functionally equivalent in signal transduction. Building on this, a recent study revealed an N-terminal 21 aa motif, termed MADA, that is conserved among the NRC helper proteins and folds into an α helix similar to the $\alpha 1$ helix of ZAR1. Additionally, this motif is found in approximately 20% of the 988 CC-NB-LRRs tested. The MADA motif is essential for the HR response, and upon switching with the corresponding region in ZAR1 the helper NB-LRR NRC4 is still capable of conferring cell-death. Whether NB-LRRs that need a helper differ from ZAR1 in the formation of a distinct resistosome remains an intriguing question to be solved.

Despite a strong interest for dominant *R* genes, interest in non-NB-LRR genes lags behind. Their identification being more difficult and elaborate than that of NB-LRR genes has so far limited the number of antiviral non-NB-LRR genes characterized. Nevertheless, non-NB-LRR *R* genes hold a strong promise towards the future, due to their more durable nature and broader resistance spectrum.

Further Reading

- Adachi, H., Contreras, M., Harant, A., *et al.*, 2019. An N-terminal motif in NLR immune receptors is functionally conserved across distantly related plant species. *eLife* 8, e49956.
- Bentham, A., Burdett, H., Anderson, P.A., Williams, S.J., Kobe, B., 2016. Animal NLRs provide structural insights into plant NLR function. *Annals of Botany* 119 (5), 698–702.
- Carr, J.P., Murphy, A.M., Tungadi, T., Yoon, J.Y., 2019. Plant defense signals: Players and pawns in plant-virus-vector interactions. *Plant Science* 279, 87–95.
- de Ronde, D., Butterbach, P., Kormelink, R., 2014. Dominant resistance against plant viruses. *Frontiers in Plant Science* 5, 307.
- Jubic, L.M., Saile, S., Furzer, O.J., El Kasmi, F., Dangl, J.L., 2019. Help wanted: Helper NLRs and plant immune responses. *Current Opinion in Plant Biology* 50, 82–94.
- Qu, F., 2010. Plant viruses versus RNAi: Simple pathogens reveal complex insights on plant antimicrobial defense. *Wiley Interdisciplinary Reviews: RNA* 1 (1), 22–33.
- Wang, J., Hu, M., Wang, J., *et al.*, 2019. Reconstitution and structure of a plant NLR resistosome conferring immunity. *Science* 364 (6435), eaav5870.
- Wu, C.H., Abd-El-Halim, A., Bozkurt, T.O., *et al.*, 2017. NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences of the United States of America* 114 (30), 8113–8118.
- Zhang, L., Chen, S., Ruan, J., *et al.*, 2015. Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization. *Science* 350 (6259), 404–409.
- Zhou, M., Li, Y., Hu, Q., *et al.*, 2015. Atomic structure of the apoptosome: Mechanism of cytochrome c-and dATP-mediated activation of Apaf-1. *Genes & Development* 29 (22), 2349–2361.

Plant Resistance to Viruses: Natural Resistance Associated With Recessive Genes

Masayoshi Hashimoto, Kensaku Maejima, Yasuyuki Yamaji, and Shigetou Namba, The University of Tokyo, Tokyo, Japan

© 2021 Elsevier Ltd. All rights reserved.

This is an update of C. Caranta, C. Dogimont, Plant Resistance to Viruses: Natural Resistance Associated with Recessive Genes, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00735-4](https://doi.org/10.1016/B978-012374410-4.00735-4).

Nomenclature

aa	amino acid(s)	NGS	next generation sequencing
Avr	avirulence	nt	nucleotide(s)
Cas	CRISPR-associated proteins	ORF	open reading frame
CI	cylindrical inclusion	PABP	poly(A)-binding protein
CITE	cap-independent translation enhancer	PDIL5-1	protein disulfide isomerase like 5-1
cPGK2	chloroplast phosphoglycerate kinase 2	Poly(A)	polyadenylated
CRISPR	Clustered regularly interspaced short palindromic repeats	QTL	quantitative trait locus
dsRNA	double-stranded RNA	RdRp	RNA-dependent RNA polymerase
EMS	ethyl methane sulfonate	RIL	recombinant inbred line
ER	endoplasmic reticulum	SNP	single-nucleotide polymorphism
GM	genetically modified	ssRNA	single-stranded RNA
gRNA	genomic RNA	TALEN	transcription activator-like effector nuclease
HR	hypersensitive response	TILLING	targeting induced local lesions in genomes
IRES	internal ribosomal entry site	tRNA	transfer RNA
KO	knockout	UPR	unfolded protein response
mRNA	messenger RNA	Vpg	viral protein genome-linked
nCBP	novel cap-binding protein	VPS41	vacuolar protein sorting 41
		ZFN	zinc-finger nuclease

Glossary

Allele Allele is one of the distinct forms of a gene at a single locus. Show different effect on the phenotype from the other forms.

Avirulence Genetic traits for a pathogen to be impaired by genetically determined host resistance and to cause an incompatible (no disease) interaction.

Dominant allele An allele expressing its phenotypic effect even when heterozygous with a recessive allele.

Genome editing Genome editing is a technology for the modification of genomic DNA sequence (deletion, substitution, and insertion) using sequence-specific nucleases.

Paralog Paralogs are a group of genes generated by duplication events within a genome. Orthologs, which are evolved from a common ancestral gene in different species, retain evolutionally conserved functions, whereas paralogs generally evolve new functions.

Pleiotropic mutation A mutation showing multiple distinct effects.

Positional cloning Determining the location of a gene-of-interest on the chromosome without any idea of the biological function of that gene.

Quantitative trait loci Correspond to genomic regions associated with the phenotypic variation of a quantitative trait.

Recessive allele An allele whose phenotypic effect is not expressed in a heterozygote.

TILLING A reverse genetics approach that relies on the ability of a special enzyme to detect mismatches in normal and mutant DNA strands when they are annealed. It can therefore detect single point mutations of the type induced by chemicals conventionally used in mutation breeding programs.

Introduction

To defend against plant pathogenic microbes, plants deploy resistance machinery, which can be inherited genetically. Genetic resistance, also referred to as host or cultivar resistance, is determined by a specific genotype in the resistant host variety, relative to other varieties possessing distinct genotypes that confer susceptibility. Employing host resistance machinery through the conventional breeding has long been one of the most effective and sustainable ways to control diseases caused by plant pathogenic microbes including viruses.

In terms of genetics, host resistance can be classified as monogenic, digenic, or polygenic ones, governed by one, two, or several genes, respectively. Most monogenic and digenic types of resistance are characterized by qualitative resistance phenotypes, such as the complete absence of pathogen proliferation. In contrast, most polygenic resistance is phenotypically quantitative, underlined by several resistance genes generally called quantitative trait loci (QTLs). QTL mapping is the main technique used to identify genetic markers statistically linked to the resistance phenotype, in which the recombinant inbred lines (RILs) derived from the crosses between parents with different resistance traits are used to collect the phenotypic data. In practice, the development of genetic markers tightly linked to resistance traits has been quite valuable for the conventional breeding of resistant plant varieties.

Host resistance can be inherited in a dominant or recessive manner. Almost half of the host resistance against plant viruses that have been characterized in natural plant ecotypes are inherited recessively. This high ratio of recessive inheritance is specific to plant viruses and has not been observed for other plant pathogens, such as oomycetes, fungi, and bacteria: this feature is closely related to the distinct infection cycles of plant viruses, as will be discussed later.

The degree of recessively inherited resistance ranges from partial to complete inhibition of pathogen invasion. Some plants with partial resistance show “tolerance,” in which symptom attenuation and/or no crop yield loss is observed, even under complete systemic infection by the pathogens. More phenotypic variation is observed for recessive resistance, than for dominant resistance, which mainly involves pathogen confinement at the initial infection site and is frequently associated with localized cell death, also known as hypersensitive response (HR). Recessive resistance can interrupt viral infection at any stage of the infection cycle, including during viral protein translation, viral genome replication, and cell-to-cell and systemic movement.

Resistance breakdown by the emergence of a virus strain with genome mutations frequently accompanies recessive resistance in agricultural fields. Comparative genetic analysis using resistance-breaking viral strains can be used to identify the viral avirulence (Avr) factors that determine recessive resistance. Functional analysis of the interaction between the viral Avr factor and the plant resistance gene product promotes our understanding of the molecular dialog between viruses and their host plants, and contributes to the development of new breeding strategies based on natural recessive resistance to avoid resistance breaking.

In this article, we demonstrate the phenotypic and molecular characteristics of natural virus recessive resistance in plants, and describe recent advances in this field.

Overview of Naturally Occurring Recessive Resistance

Recessive resistance is a highly relevant trait for virus-resistant crop breeding. The ratio of recessive to dominant resistance to plant viruses is much higher than those of resistance to pathogens from other kingdoms, including oomycetes, fungi, and bacteria. Examples of recessive resistance found in nature, drawn from a survey of the literature, are shown in [Table 1](#). Recessive resistance to viruses has been observed in dicots and monocots. It has been studied most intensively in the dicots families Fabaceae, Cucurbitaceae, and Solanaceae, and in the monocots barley and rice. Recessive resistance is rarely found in polyploid plants, likely due to the presence of dominant gene copies with redundant function. However, one allopolyploid plant, *Brassica juncea* (mustard), encodes recessive resistance gene *retr03*, which has been found at a higher than expected ratio in 7 resistant landraces among 35 tested.

Inherently recessive resistance has been characterized for viruses of diverse taxa, with various types of RNA- or DNA-based genome structure. Many of these viruses belong to the genus *Potyvirus*. Potyviruses are single-stranded RNA viruses that account for about 70% of the recessive resistance in a wide range of plant species, together with other related viruses encoding the viral protein genome-linked (VPg) protein, which binds to the 5′ end of their genomic RNA as a cap analog. *Potyvirus* is the largest genus of phytopathogenic viruses; it contains many virus species that cause severe crop yield losses worldwide. The frequency of recessive resistance varies among virus families and genera: it is almost 50% in the genus *Potyvirus*, but rare in other virus genera, such as *Tobamovirus*. This inconsistency is likely to relate to virus infection strategies and selection pressure due to viral infection in natural plant habitats. Thus, the frequency of recessive resistance for a specific plant–virus interaction may be determined by a trade-off between the fitness costs of recessive mutations of resistance genes and the costs of viral infection to healthy plant growth.

Several resistance loci control distinct strains of single virus species, or two or more distinct viral species. In barley (*Hordeum vulgare*), two distinct alleles, *rym4*, and *rym5*, at a single recessive locus confer resistance to several bymoviruses, including Barley mild mosaic virus (BaMMV), Barley yellow mosaic virus (BaYMV), and Barley yellow mosaic virus-2 (BaYMV-2). Alleles at the *pvr2* locus in pepper (*Capsicum* spp.) are responsible for resistance to several potyviruses, including Potato virus Y (PVY), Tobacco etch virus (TEV), and Pepper mottle virus (PepMoV). Another example is pea (*Pisum sativum*) with two well-documented recessive resistance loci against several potyviruses. At the first locus are *bcm* against Bean common mosaic virus (BCMV), *cyr1* against Clover yellow vein virus (CIYVV), *mo* against Bean yellow mosaic virus (BYMV) and Watermelon mosaic virus (WMV), *sbm-2* against Pea seed-borne mosaic virus (PSbMV), and *pmv* against Pea mosaic virus (PMV). At the second locus are *sbm-1*, *sbm-3*, and *sbm-4*, conferring resistance to three PSbMV strains, *cyr2* against CIYVV, *wlv* against BYMV, and *mo1*¹ and *mo1*² against Lettuce mosaic virus (LMV). With recent advances in the molecular understanding of natural recessive resistance, discussed below, the broad resistance spectrum of these loci has been attributed to the action of a single gene with a pleiotropic effect.

Although recessive resistance is usually governed by a single gene, it is controlled by two distinct loci in several plant–virus pairs. In common bean (*Phaseolus vulgaris*), resistance against BCMV pathotypes requires two complementary genes at the *bc-u* locus and at one of the three strain-specific loci (*bc-1*, *bc-2*, and *bc-3*). Interestingly, in pepper, only the *pvr2* locus controls resistance against potyviruses including PVY, whereas the *pvr6* and *pvr2* loci are required for recessive resistance to another

Table 1 Virus recessive resistance found in hosts

Genus – Family/Virus (acronym)	Host plant	Resistance gene(s) or loci ^a	Resistance phenotype ^a
<i>RNA viruses with a 5' VPg and 3' poly(A) tail</i>			
<i>Bymovirus - Potyviridae</i>			
Barley mild mosaic virus (BaMMV)	<i>Hordeum vulgare</i> (barley)	<i>rym4</i> , <i>rym5</i> ± 15 <i>rym</i> on 5chr.	CR PR-CR
Barley yellow mosaic virus (BaYMV)	<i>Hordeum vulgare</i> (barley)	<i>rym4</i> , <i>rym5</i> ± 15 <i>rym</i> on 5chr.	CR PR-CR
Barley yellow mosaic virus-2 (BaYMV-2)	<i>Hordeum vulgare</i> (barley)	<i>rym5</i> <i>rym1</i> , <i>rym11</i>	CR
<i>Potyvirus - Potyviridae</i>			
Bean common mosaic virus (BCMV)	<i>Pisum sativum</i> (pea) <i>Phaseolus vulgaris</i> (common bean)	<i>bcm</i> <i>bc-1</i> , <i>bc-1</i> ² , <i>bc-2</i> , <i>bc-2</i> ² , <i>bc-3</i> , <i>bc-u</i> (≠ locus)	No SI
Bean yellow mosaic virus (BYMV)	<i>Phaseolus vulgaris</i> (bean) <i>Pisum sativum</i> (pea) <i>Vicia mungo</i> (mungbean)	<i>cyv</i> <i>mo</i> 2 genes	No SI SL
Bean yellow mosaic virus-W (BYMV)	<i>Pisum sativum</i> (pea)	<i>wlv</i>	No local or SI
Bidens mottle virus (BiMV)	<i>Lactuca sativa</i> (lettuce)	<i>bi</i>	
Blackeye cowpea mosaic virus (BICMV)	<i>Vigna unguiculata</i> (cowpea)	<i>bcm</i>	CR
Celery mosaic virus (CeMV)	<i>Apium graveolens</i> (celery)	<i>cmv</i>	
Clover yellow vein virus (CIYVV)	<i>Phaseolus vulgaris</i> (bean) <i>Pisum sativum</i> (pea)	<i>desc</i> <i>cyv1</i> <i>cyv2</i> (≠ locus)	Cell R Cell-to-cell Mvt
Lettuce mosaic virus (LMV)	<i>Glycine max</i> (soybean) <i>Lactuca sativa</i> (lettuce)	<i>d-cv</i> <i>mo1</i> ¹ , <i>mo1</i> ²	Red Acc Red Acc
Moroccan watermelon mosaic virus (MWMV)	<i>Cucumis sativus</i> (cucumber)	<i>mwm</i>	SL
Papaya ringspot virus-W (PRSV)	<i>Cucurbita ecuadorensis</i> <i>Cucumis sativus</i> (cucumber)	2 genes <i>prsv-1</i> <i>prsv</i> ²²⁴⁵	Red Acc Red Acc Red Acc
	<i>Cucurbita moschata</i> (squash)	<i>prv</i>	SL
	<i>Citrullus amarus</i> (watermelon)	<i>prv</i>	
Passionfruit woodiness virus (PWV)	<i>Pisum sativum</i> (pea)	<i>pwv</i>	CR
Pea seed-borne mosaic virus (PSbMV)	<i>Lens culinaris</i> (lentil bean) <i>Pisum sativum</i> (pea)	<i>sbv</i> <i>sbm1</i> , <i>sbm3</i> , <i>sbm4</i> <i>sbm2</i> (≠ locus)	Red Acc Cell R
Peru tomato virus (PTV)	<i>Solanum lycopersicum</i> (tomato)	Monogenic	
Pepper mottle virus (PepMoV)	<i>Capsicum annuum</i> (pepper) <i>Capsicum chinense</i> (pepper)	<i>pvr3</i> <i>pvr1</i>	Long-distance Mvt Cell R
Pepper veinal mottle virus (PVMV)	<i>Capsicum annuum</i> (pepper)	<i>pvr2</i> ² + <i>pvr6</i> (digenic)	CR
Plum pox virus (PPV)	<i>Arabidopsis thaliana</i>	<i>rpv1</i>	Long-distance Mvt
Potato virus A (PVA)	<i>Solanum tuberosum</i> subsp. <i>andigena</i>	<i>ra_{adg}</i>	Vascular Mvt
Potato virus Y (PVY)	<i>Capsicum annuum</i> (pepper)	<i>pvr2</i> ¹ , <i>pvr2</i> ² <i>pvr3</i> (≠ locus) Several QTLs	Cell R CR PR
	<i>Capsicum chinense</i> (pepper)	<i>pvr1</i> (allelic to <i>pvr2</i>)	Cell R
	<i>Solanum habrochaites</i> (tomato)	<i>pot-1</i>	CR
Tobacco etch virus (TEV)	<i>Nicotiana tabacum</i> (tobacco) <i>Capsicum annuum</i> (pepper) <i>Capsicum chinense</i> (pepper) <i>Solanum habrochaites</i> (tomato)	<i>va</i> <i>pvr2</i> ² <i>pvr1</i> <i>pot-1</i>	Cell-to-cell Mvt Cell R Cell R CR
Tobacco vein mottling virus (TVMV)	<i>Nicotiana tabacum</i> (tobacco)	2 genes linked	Long-distance Mvt
Turnip mosaic virus (TuMV)	<i>Nicotiana tabacum</i> (tobacco) <i>Brassica campestris</i> (cabbage) <i>Brassica rapa</i> (turnip)	<i>va</i> 2 genes 1 to 3 genes <i>trs</i> <i>retr01</i> <i>retr02</i> <i>retr03</i>	Cell-to-cell Mvt No SI No SI Red Acc

(Continued)

Table 1 Continued

Genus – Family/Virus (acronym)	Host plant	Resistance gene(s) or loci ^a	Resistance phenotype ^a
Watermelon mosaic virus (WMV)	<i>Cucumis melo</i> (melon)	monogenic <i>wmv</i> ¹⁵⁵¹	Red Acc; Long-distance Mvt To
	<i>Cucumis sativus</i> (cucumber)	<i>wmv</i> -2 <i>wmv</i> -3 + <i>wmv</i> -4 (digenic)	SL cotyledons SL
	<i>Pisum sativum</i> (pea)	<i>mo</i>	No SI
	<i>Arabidopsis thaliana</i>	<i>rmv1</i>	
	<i>Cucumis sativus</i> (cucumber)	<i>wmv</i> ⁰²²⁴⁵	
	<i>Citrullus lanatus</i> (watermelon)	2 genes	PR
	White lupin mosaic virus (WLMV)	<i>Pisum sativum</i> (pea)	<i>wlv</i>
Zucchini yellow mosaic virus (ZYMV)	<i>Cucumis sativus</i> (cucumber)	<i>zym</i>	Red Acc
	<i>Citrullus lanatus</i> (watermelon)	<i>zym</i> -CH	No SI
	<i>Citrullus mucospermus</i> (watermelon)	<i>zym</i> -FL	
Zucchini yellow fleck virus (ZYFV)	<i>Citrullus lanatus</i> (watermelon)	monogenic	No SI
	<i>Cucumis sativus</i> (cucumber)	<i>zyf</i>	No SI
<i>Comovirus</i> - <i>Secoviridae</i>			
Cowpea severe mosaic virus (CPSMV)	<i>Vigna unguiculata</i> (cowpea)	Oligogenic (3 genes)	CR
<i>Waikavirus</i> - <i>Secoviridae</i>			
Rice tungro spherical virus (RTSV)	<i>Oryza sativa</i> (rice)	<i>tsv</i> -1 <i>tsv</i> -2 (≠ loci)	No SI
<i>RNA viruses with a 5' VPg and non-polyadenylated 3' end</i>			
<i>Enamovirus</i> - <i>Luteoviridae</i>			
Bean leafroll virus (BLRV)	<i>Pisum sativum</i> (pea)	<i>lr</i>	SL
		<i>lrv</i>	SL
<i>Polerovirus</i> - <i>Luteoviridae</i>			
Beet western yellows virus (BWYV)	<i>Lactuca sativa</i> (lettuce)	<i>bwy</i>	Red Acc
Cucurbit aphid-borne yellows virus (CABYV)	<i>Cucumis melo</i> (melon)	<i>cab</i> -1 + <i>cab</i> -2 (digenic)	CR
<i>Sobemovirus</i> - <i>Solemoviridae</i>			
Rice yellow mottle virus (RYMV)	<i>Oryza sativa</i> (rice)	<i>rymv</i> 1	CR
		Several QTLs	PR
	<i>Oryza glaberrima</i> (rice)	<i>rymv</i> 1 <i>rymv</i> 2	CR
<i>RNA viruses with capped 5' end and non-polyadenylated 3' end</i>			
<i>Alfavirus</i> - <i>Bromoviridae</i>			
Alfalfa mosaic virus (AMV)	<i>Medicago sativa</i> (alfalfa)	<i>am</i> -1	
<i>Bromovirus</i> - <i>Bromoviridae</i>			
Cowpea chlorotic mottle virus (CCMV)	<i>Glycine max</i> (soybean)	2 genes	Long-distance Mvt
<i>Cucumovirus</i> - <i>Bromoviridae</i>			
Cucumber mosaic virus (CMV)	<i>Cucumis melo</i> (melon)	Polygenic with a major QTL (<i>cmv</i> 1)	SL
	<i>Cucurbita pepo</i> (pumpkin)	2 unlinked genes	SL
	<i>Capsicum frutescens</i> (pepper)	Oligogenic (2–3 genes)	No SI
	<i>Capsicum annuum</i> (pepper)	<i>cmr</i> 2	
<i>Tobamovirus</i> - <i>Tobamoviridae</i>			
Cucumber green mottle mosaic virus (CGMMV)	<i>Cucumis melo</i> (melon)	Polygenic	SL
		<i>cgmmv</i> -1 <i>cgmmv</i> -2	
<i>Umbravirus</i> - <i>Tombusviridae</i>			
Groundnut rosette virus (GRV)	<i>Arachis hypogaea</i> (peanut)	2 genes	SL
<i>Tospovirus</i> - <i>Bunyaviridae</i>			
Tomato spotted wilt virus (TSWV)	<i>Solanum lycopersicum</i> (tomato)	<i>sw</i> 2, <i>sw</i> 3, <i>sw</i> 4	PR
<i>RNA viruses with unmodified 5' and 3' ends</i>			
<i>Carmovirus</i> - <i>Tombusviridae</i>			
Turnip crinkle virus (TCV)	<i>Arabidopsis thaliana</i>	<i>rrt</i> + <i>HRT</i> (digenic)	HR
Melon necrotic spot virus (MNSV)	<i>Cucumis melo</i> (melon)	<i>nsv</i>	Cell R
	<i>Nicotiana benthamiana</i>	(no locus name)	Non-host R
<i>Luteovirus</i> - <i>Luteoviridae</i>			
Barley yellow dwarf virus (BYDV)	<i>Hordeum vulgare</i> (barley)	<i>ryd</i> 1	PR
<i>Idaeovirus</i>			
Raspberry bushy dwarf virus (RBDV)	<i>Rubus idaeus</i> (raspberry)	Monogenic	CR
<i>dsRNA virus</i>			

Table 1 Continued

Genus – Family/Virus (acronym)	Host plant	Resistance gene(s) or loci ^a	Resistance phenotype ^a
<i>Fijivirus</i> - <i>Reoviridae</i>			
Maize rough dwarf virus (MRDV)	<i>Zea mays</i> (maize)	Polygenic with a major QTL (<i>qMrdd1</i>)	
Mal de Rio Cuarto virus (MRCV)	<i>Zea mays</i> (maize)	Polygenic with a major QTL (<i>qMrdd1</i>)	
Rice black-streaked dwarf virus (RBSDV)	<i>Zea mays</i> (maize)	Polygenic with a major QTL (<i>qMrdd1</i>)	
<i>DNA viruses</i>			
<i>Begomovirus</i> - <i>Geminiviridae</i>			
Bean golden yellow mosaic virus (BGYMV)	<i>Phaseolus vulgaris</i> (common bean)	<i>bgm-1</i>	Red S
		<i>bgm-2</i> (≠ loci)	Red S
	<i>Phaseolus coccineus</i> (scarlet runner bean)	<i>bgm-3</i> (≠ loci)	Red S
Potato yellow mosaic virus (PYMV)	<i>Solanum pimpinellifolium</i>	Digenic	No SI
Tomato yellow leaf curl virus (TYLCV)	<i>Solanum habrochaites</i>	Oligogenic	CR
	<i>Solanum lycopersicum</i> (tomato)	A major QTL, <i>ty-5</i> , and four minor QTLs	Red S
Tomato leaf curl virus (ToLCV)	<i>Solanum chilense</i>	<i>tgr-1</i>	Cell-to-cell Mvt
Tomato chlorotic mottle virus (ToCMoV)	<i>Solanum lycopersicum</i> (tomato)	<i>tcm-1</i>	No SI
Cabbage leaf curl virus (CaLCuV)	<i>Arabidopsis thaliana</i>	<i>gip-1</i> (Pla-1 ecotype)	No Acc Inoc leaves; R to BCTV and TYLCV
<i>Curtovirus</i> - <i>Genimiviridae</i>			
Beet curly top virus (BCTV)	<i>Arabidopsis thaliana</i>	Monogenic R loci in three ecotypes, Ms-0, Pr-0, and Cen-O	To or No SI
<i>Tungrovirus</i> (<i>Caulimoviridae</i>)			
Rice tungro bacilliform virus (RTBV)	<i>Oryza sativa</i> (rice)	Polygenic	To, Red Acc

^aAbbreviations: CR, complete resistance; chr, chromosomes; PR, partial resistance; ≠, independent; SL, symptomless; Cell R, cellular resistance (resistance at a single cell level); To, tolerance; SI, systemic infection; Mvt, movement; Red, reduced; Acc, accumulation; Inoc, inoculated; HR, hypersensitive response; S, symptoms; R, resistance.

potyvirus, Pepper vein mottle virus (PVMV). These phenomena imply that the pyramiding of independent resistance loci in a single cultivar, if possible, would broaden the spectrum of recessive resistance.

Among a number of studies performing the genetic analysis on resistances governed by QTLs, some of them have been proved to be recessive. For example, quantitative recessive resistance in a melon cultivar to Cucumber mosaic virus (CMV; *Cucumovirus*, family *Bromoviridae*) is conferred by polygenes including the major QTL *cmv1*. In melon, the major QTL *wmv*¹⁵⁵¹, together with three minor QTLs on different chromosomes, regulates recessive resistance against WMV. The major QTL *qMrdd1*, which confers recessive resistance against maize rough dwarf disease caused by a group of viruses in the genus *Fijivirus*, family *Reoviridae*, has also been mapped in maize. The major QTL *rymv*-QTL1 in rice modulates quantitative recessive resistance to Rice yellow mottle virus (RYMV; *Sobemovirus*, family *Solemoviridae*). Interestingly, *rymv*-QTL1 was mapped to the same locus as the monogenic recessive resistance gene *rymv2*, indicating that the same gene can be involved in quantitative and qualitative resistance.

Cloned Recessive Resistance Genes

Plant viruses rely heavily on a number of host factors, also referred to as susceptibility factors, in host cells to replicate their genomes and spread into healthy tissues. In host cells, viruses must simultaneously deal with multiple defense responses by plant innate immune systems. Due to such inherent nature of virus infection strategies in host cells, recessive resistance is hypothesized to be genetically explained by loss-of-function (null) or point mutations of a gene encoding a host factor involved in viral infection, and of a negative regulator of defense responses. Both parts of this hypothesis appear to be supported by recent developments in the cloning and the characterization of recessive resistance genes found in natural plant germplasm (Table 2).

Eukaryotic translation initiation factor 4E (eIF4E)-mediated recessive resistance has been observed in a wide range of plant species, including dicots and monocots. The possible involvement of eIF4E in plant viral infection was first documented by the isolation of an isoform of eIF4E (eIFiso4E) as an interactor of the VPg protein of Turnip mosaic virus (TuMV) in a yeast two-hybrid assay. eIF4E-mediated resistance has been discovered in various crop species via the candidate gene approach. Additionally, there are also distinct classes of the recessive resistance genes in terms of their intrinsic functions in host plants, including both dicots and monocots, against a wide range of viruses. Elucidation of the characteristics and mechanisms of the natural recessive resistance would create a foundation for the development of the new plant breeding strategies for virus resistance. Recent findings imply the existence of an unexpected regulatory mechanism for natural recessive resistance, rather than a simple working hypothesis

Table 2 Cloned recessive resistance genes against plant viruses

Genus – Family/Virus (acronym)	Locus or allele	Host plant	Resistance Gene	Avr gene
<i>Potyvirus</i> - <i>Potyviridae</i>				
Bean yellow mosaic virus (BYMV)	<i>wlv</i>	<i>Pisum sativum</i>	<i>eIF4E</i>	VPg
Clover yellow vein virus (CIYVV)	<i>cyv2</i>	<i>Pisum sativum</i>	<i>eIF4E</i>	P1, P3N-PIPO
Clover yellow vein virus (CIYVV)	<i>desc</i> (= <i>bc-3</i> & <i>cyv</i>)	<i>Phaseolus vulgaris</i>	<i>eIF4E</i>	
Lettuce mosaic virus (LMV)	<i>mo1</i> ¹	<i>Lactuca sativa</i>	<i>eIF4E</i>	CI, VPg, NIa
Lettuce mosaic virus (LMV)	<i>mo1</i> ²	<i>Lactuca sativa</i>	<i>eIF4E</i>	CI, VPg, NIa
Pea seed-borne mosaic virus (PSbMV)	<i>sbm1</i> (= <i>wlv</i> & <i>cyv2</i>)	<i>Pisum sativum</i>	<i>eIF4E</i>	VPg
Pepper vein mottle virus (PVMV)	<i>pvr6</i> (+ <i>pvr2</i>)	<i>Capsicum annuum</i>	<i>eIFiso4E</i>	
Plum pox virus (PPV)	<i>rpv1</i>	<i>Arabidopsis thaliana</i>	<i>cPGK2</i>	
Potato virus Y (PVY)	<i>pvr2</i> ¹	<i>Capsicum annuum</i>	<i>eIF4E</i>	VPg
Potato virus Y (PVY); Tobacco etch virus (TEV)	<i>pvr2</i> ²	<i>Capsicum annuum</i>	<i>eIF4E</i>	VPg
Potato virus Y (PVY); Tobacco etch virus (TEV)	<i>pot-1</i>	<i>Solanum habrochaites</i>	<i>eIF4E</i>	VPg
Potato virus Y (PVY); Tobacco etch virus (TEV); Pepper mottle virus (PepMoV)	<i>pvr1</i>	<i>Capsicum chinense</i>	<i>eIF4E</i>	
Turnip mosaic virus (TuMV)	<i>retr01</i> , <i>retr02</i>	<i>Brassica rapa</i>	<i>eIFiso4E</i>	VPg
Turnip mosaic virus (TuMV)	<i>retr03</i>	<i>Brassica rapa</i>	<i>eIF2Bβ</i>	
Watermelon mosaic virus (WMV)	<i>rwm1</i>	<i>Arabidopsis thaliana</i>	<i>cPGK2</i>	
<i>Bymovirus</i> - <i>Potyviridae</i>				
Barley mild mosaic virus (BaMMV); Barley yellow mosaic virus (BaYMV)	<i>rym4</i>	<i>Hordeum vulgare</i>	<i>eIF4E</i>	VPg
BaMMV; BaYMV; Barley yellow mosaic virus – 2 (BaYMV – 2)	<i>rym5</i>	<i>Hordeum vulgare</i>	<i>eIF4E</i>	VPg
BaMMV; BaYMV; Barley yellow mosaic virus – 2 (BaYMV – 2)	<i>rym11</i>	<i>Hordeum vulgare</i>	PDIL5–1	
<i>Cucumovirus</i> - <i>Bromoviridae</i>				
Cucumber mosaic virus (CMV)	<i>cmv1</i>	<i>Cucumis melo</i>	VPS41	MP
<i>Sobemovirus</i> - <i>Solemoviridae</i>				
Rice yellow mottle virus (RYMV)	<i>rymv1</i>	<i>Oryza sativa</i>	<i>eIFiso4G</i>	VPg
Rice yellow mottle virus (RYMV)	<i>rymv2</i>	<i>Oryza glaberrima</i>	CPR5	
<i>Carmovirus</i> - <i>Tombusviridae</i>				
Melon necrotic spot virus (MNSV)	<i>nsv</i>	<i>Cucumis melo</i>	<i>eIF4E</i>	3' UTR
Melon necrotic spot virus (MNSV)	–	<i>Nicotiana benthamiana</i>	<i>eIF4E</i>	3' UTR
Melon necrotic spot virus (MNSV)	(EcoTILLING)	<i>Cucumis zeyheri</i>	<i>eIF4E</i>	
<i>Waikavirus</i> - <i>Secoviridae</i>				
Rice tungro spherical virus (RTSV)	<i>tsv1</i>	<i>Oryza sativa</i>	<i>eIF4G</i>	
<i>Begomovirus</i> <i>Geminiviridae</i>				
Tomato yellow leaf curl virus (TYLCV)	<i>ty-5</i>	<i>Solanum lycopersicum</i>	<i>Pelota</i>	

explained by null mutation of a single susceptibility factor. A deeper understanding of recessive resistance in natural germplasm will be valuable for durable resistance applications in crop species.

Characteristics of eIF4E-Mediated Recessive Resistance

The first molecular cloning of a naturally occurring recessive resistance gene revealed that *eIF4E* gene is responsible for *pvr2*-mediated resistance against PVY and TEV in pepper (*Capsicum annuum*). Although a number of susceptibility factors has been identified from naturally resistant accessions and artificially induced mutants, and as interactors of viral proteins in various plant species, the most-studied naturally recessive resistance genes remain *eIF4E*, its interacting partner *eIF4G*, and their isoforms (hereafter *eIF4Es*).

The protein eIF4E is commonly involved in canonical translation of mRNA in eukaryotic cells. eIF4E physically interacts with eIF4G scaffold proteins to form the eIF4F translation initiation binary complex. This complex initially binds to the m⁷G cap structure at the 5' end of mRNA through the eIF4E subunit. Subsequently, eIF4G recruits several other initiation factors, such as poly(A)-binding protein (PABP); eIF4A, a helicase for the unwinding of RNA secondary structures; and eIF3, a multi-subunit complex that binds to the 40S ribosomal subunit (Fig. 1). The eIF4F complex induces the close interaction of both termini within the mRNA molecules in a process termed circularization, which promotes efficient translation.

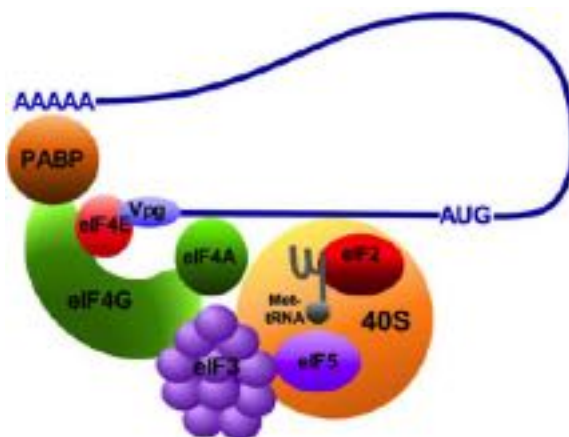


Fig. 1 The eukaryotic translation initiation complex. eIF, eukaryotic initiation factor; PABP, poly(A)-binding protein; 40S, 40S ribosomal subunit. Reproduced from Robaglia, C., Caranta, C., 2006. Translation initiation factors: A weak link in plant RNA virus infection. *Trends in Plant Science* 11(1), 40–45.

Plants possess two isoforms of eIF4E: eIFiso4E, and novel cap-binding protein (nCBP). eIF4E and eIFiso4E form the distinct protein complexes eIF4F and eIFiso4F with eIF4G and eIFiso4G, respectively. Although these complexes can have partially redundant functions in the protein translation machinery, each complex selectively recruits specific mRNA species for translation. Genes encoding the components of the eIF4F complexes constitute a small gene families in plants. For example, in the model plant *Arabidopsis*, three genes (*eIF4E1*, *eIF4E2*, and *eIF4E3*) belongs to the eIF4E subfamily, one gene belongs to eIFiso4E; one gene belongs to eIF4G; two genes (*eIFiso4G1* and *eIFiso4G2*) belongs to the eIFiso4G subfamily. Thus, the paralogous genes in each subfamily showing high degrees of sequence similarity are predicted to be functionally redundant in the translation machinery.

The protein eIF4Es may have biological functions other than translation initiation. eIF4E and/or eIFiso4E have been deduced to be involved in the nuclear export of a subset of mRNA molecules. The eIFiso4F complex may regulate microtubule dynamics through the physical interaction of eIFiso4G with microtubules. In addition, each component of the two eIF4F complexes had been demonstrated to be a susceptibility factor for some plant viruses, especially those encoding VPg. In contrast, although nCBP has a sequence motif related to eIF4E and less cap-binding activity than eIF4E and eIFiso4E, its intrinsic function in the plant translation machinery and interacting partners remain unknown. However, nCBP has been recently reported to be a susceptibility factor for several potexviruses (family *Alphaflexiviridae*) in *Arabidopsis*, and of Cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UCBSV; genus *Ipomovirus*, family *Potyviridae*) in cassava (*Manihot esculenta*).

As mentioned above, more than half of recessively inherited resistance alleles are associated with potyviruses. Similarly, the majority of viruses influenced by eIF4E-mediated resistance are potyviruses. This dominance of potyviruses may be consistent with the relevance of this large virus group in agricultural production. Given the essential role of eIF4Es in translation and the lack of their own translational machinery for plant viruses, all plant viruses very likely utilize the eIF4Es for the translation of viral factors. Nevertheless, a limited range of virus species is impaired by mutation of one of these translation initiation factors, suggesting that some potyviruses have a fitness advantage in their infection cycles characterized by specific interaction with one of the eIF4Es possessing redundant functions. In contrast, as mentioned above, resistance to PVMV in pepper requires both *pvr6* and *pvr2-eIF4E*. *pvr6* encodes a loss-of-function variant of the *eIFiso4E* gene, suggesting that PVMV can recruit both eIF4E and eIFiso4E to establish infection.

Except for the *pvr6* resistance gene in pepper, almost all natural recessive resistance alleles related to *eIF4Es* possess only several point mutations in their coding regions, which do not interfere with the intrinsic function of eIF4E variants in plant cells. To date, several surveys of natural variation in *eIF4E* alleles in various crop cultivars indicated that the nonsynonymous substitutions correlated with the resistance phenotype are predominantly found at the non-conserved aa residues near the cap-binding pocket; these substitutions occurs in two adjacent regions on the surface of the predicted three-dimensional eIF4E structure. Functional resistance alleles with subtle aa substitutions are also known in other viruses affected by eIF4E-mediated recessive resistance, such as bymoviruses, RYMV, and Melon necrotic spot virus (MNSV). In pepper, lettuce, and pea, several allelic series show distinct resistance specificity against different virus species and pathotypes. These series represent multiple combinations of point mutations at a small number of residues, enabling multiple distinct resistance specificities.

Another common feature of eIF4E-mediated resistance is that VPg, a viral protein covalently attached to the 5' end of viral RNA, has been annotated as an Avr factor of bymoviruses and sobemoviruses, and potyviruses. A few other viral factors, including the P1 cistron of CLYVV in pea and the cylindrical inclusion (CI) protein of LMV in lettuce, have been characterized as Avr proteins. Remarkably, the Avr determinant of MNSV for *nsv-eIF4E*-mediated recessive resistance in melon can be mapped to 3' cap-independent translation enhancer (3'-CITE) at 3' untranslated region of the viral genome, but not to viral proteins. Interestingly, the incompatibility of the MNSV 3'-CITE with eIF4E may also explain non-host resistance against MNSV in *Nicotiana benthamiana*.

The roles of eIF4Es in plant–virus interactions have long been studied using *Arabidopsis*. Similarly, *Arabidopsis* mutants of *eIF4Es* introduced by ethyl methane sulfonate (EMS) treatment or transposon knockout (KO) show phenotypes characterized by loss of

susceptibility to several plant RNA viruses, including potyviruses, CMV, and Turnip crinkle virus (TCV; *Carmovirus*, family *Tombusviridae*). Such phenotypes are genetically equivalent to natural recessive resistance found among crop cultivars. However, no natural *Arabidopsis* ecotype deploying eIF4Es-mediated recessive resistance has been identified to date, implying that *Arabidopsis* has a greater fitness advantage during probable interactions with viruses relying on eIF4Es than conferred by recessive resistance mutation in *eIF4Es*. Functional specificities to eIF4Es of particular viral strains appear to be distinct between *Arabidopsis* and crop species. For instance, whereas TEV and LMV rely on eIFiso4E to infect *Arabidopsis*, they rely on eIF4E to infect pepper, tomato, and lettuce. These results highlight the significance of the basic research to reveal specific interactions among eIF4Es and viruses in crop species and model plants for the development of recessive resistance cultivars.

Molecular Mechanisms of eIF4E-Mediated Recessive Resistance

Due to the early cloning of *eIF4Es* as recessive resistance genes, which have the most dominance among cloned resistance genes, the molecular mechanism of recessive resistance governed by *eIF4Es* has been the most intensively studied so far. The two main topics of this research area have been the molecular characterization of viral Avr factors, mainly the potyvirus VPg protein, and the functional analysis of natural allelic variation in *eIF4Es* in crops and wild relatives. These studies have provided profound insight into the principles and diversity of molecular mechanisms of eIF4E-mediated recessive resistance.

In most cases, it is well known that *eIF4E* resistance alleles fails to physically interact with the potyvirus VPg protein (Fig. 2). Given that point mutations in *eIF4E* resistance alleles are mainly found near the cap-binding pocket, eIF4E-mediated recessive resistance against potyviruses is plausibly due to the disruption of eIF4E–VPg interaction. Point mutations of eIF4Es may abolish the recruitment by VPg, instead of the 5' cap structure of intrinsic mRNA for translation of the potyvirus genome. This hypothesis is consistent with the finding that eIF4G, as well as eIF4E, is a susceptibility factor for CIYVV in *Arabidopsis*, suggesting that the eIF4F complex is involved in potyvirus infection. However, a few inconsistencies were found in the correlation of eIF4E–VPg interactions with resistance phenotypes, implying the existence of another distinct mechanism for resistance against at least a few potyviruses that is functionally independent of such interaction. Alternatively, it is possible the methods used to detect eIF4E–VPg interactions, such as the yeast two-hybrid assay, may not perfectly represent these biological interactions in host cells.

Intriguingly, each plant–virus pair shows diversity of infection steps and the extent to which viral infection is impaired in eIF4E-mediated resistance against potyviruses (Table 1). Different levels of recessive resistance against a virus are frequently observed on distinct alleles at a single locus in a host. This may also support the hypothesized diversity of potyvirus infection strategies, some of which are independent of eIF4E–VPg interaction. Alternatively, eIF4E may be involved in multiple steps of the potyvirus infection cycle. Thus, distinct *eIF4E* alleles may differentially change the eIF4E function at each step in a qualitative or quantitative manner, generating diversity within the resistance phenotype.

The translation of viral proteins from genomic RNAs of PVY and TuMV had been proposed to depend on the internal ribosomal entry site (IRES), suggesting an alternative hypothesis that the eIF4E–VPg interaction does not directly control genomic RNA translation. Rather, VPg may cause host translation shutoff via interaction with eIF4E. Such activity promotes the release of translation initiation factors and/or ribosomes from host mRNA translation, thereby directing eIF4E to translate viral proteins. Since replication is tightly linked to translation during the plant viral infection cycle, eIF4E–VPg interaction had also been proposed to control replication, rather than translation. Indeed, the VPg region of the 6K2-VPg-Pro polyprotein has been demonstrated to be recruited into the viral replication complex (VRC) in TuMV-infected cells.

A few exceptions to the role of VPg as an Avr determinant of recessive resistance have been found. For example, the breaking of *mol1*-mediated resistance is attributed to a single aa mutation of the LMV CI, which is involved in potyviral replication and systemic movement. Recessive resistance mediated by *cyl2* is also overcome by an aa substitution of the CIYVV P1 cistron. In the sobemovirus RYMV, virulent VPg protein has been shown to interact with eIFiso4G, but not with eIF4E. These results indicate

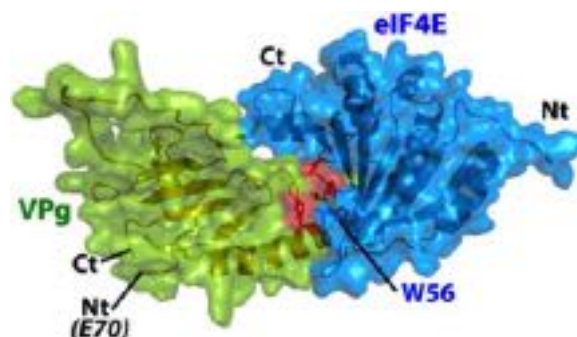


Fig. 2 Predicted three-dimensional structural model of the PVY VPg (green) and eIF4E (blue) complex. W56 in eIF4E protein indicates the tryptophan residue that interacts with the cap structure. Nt, N-terminus; Ct, C-terminus. Reproduced from de Oliveira, L.C., Volpon, L., Rahardjo, A.K., *et al.*, 2019. Structural studies of the eIF4E-VPg complex reveal a direct competition for capped RNA: Implications for translation. *Proceedings of National Academy of Science of the United States of America* 116, 24056–24065.

qualitative differentiation in the mechanism of resistance mediated by *eIF4E*, even between closely related viruses. VPg may not be the sole interactor with *eIF4Es*; other viral factors together with VPg, including the polyprotein intermediates containing the VPg domain, may also play essential roles in this interaction.

The *eIF4Es*-mediated recessive resistance has also been found to be functional against viruses with distinct genomic structures, such as the lack of VPg, clearly suggesting a resistance mechanism distinct from the one mediated by *eIF4Es* against VPg-containing viruses. As mentioned above, *nsu-eIF4E*-mediated resistance against MNSV is attributed to the 3'-CITE sequence, which is an Avr factor. A resistance-breaking strain of MNSV had been shown to gain the ability to interact with *eIF4E* from resistant melon. Although direct evidence for the molecular resistance mechanism remains lacking, *eIF4E* may facilitate circularization of MNSV genomic RNAs via interaction with 3'-CITE.

Large-scale surveys of natural polymorphism of *eIF4E* alleles in several crops and their wild relatives have been conducted to explore the genetic variation and evolutionary patterns of *eIF4E* recessive resistance loci. KO mutations of *eIF4E* has not been found among natural accessions in these surveys. Although artificial development of an *eIF4E* KO mutant in *Arabidopsis* is possible, this may be due to the moderate greenhouse conditions normally used in *Arabidopsis* research. In a survey using the wild pepper chiltepin (*Capsicum annuum* var. *glabriusculum*), which is undergoing early domestication, *eIF4E* resistance alleles were frequently found in cultivated chiltepin populations. The same trend has also been reported for bymovirus resistance in barley, where possible *eIF4E* resistance alleles are predominantly encoded in domesticated barley accessions. In contrast, a survey of melon wild relatives successfully found several natural alleles of *eIF4E* genes, but no correlation of these alleles with the resistance phenotype. These findings indicate that human management of plant populations exerts profoundly strong selection pressures on the diversity and evolution of recessive resistance genes.

Each potyvirus species has evolved to specifically recruit one *eIF4E* for viral infection, despite functional overlap among *eIF4Es* in the host translational machinery. Therefore, if possible, simultaneous KO of *eIF4Es* is desirable for the engineering of broad-spectrum resistance against viruses. However, several lines of evidence suggest that double mutation of *eIF4Es* is harmful to plant development. In *Arabidopsis*, the *eif4e/eifiso4e* double-KO mutant shows a lethal phenotype; in tomato, functional repression of two *eIF4E* paralogues, *eIF4E1* and *eIF4E2*, by RNA silencing induces severe dwarf phenotypes.

Given the functional relevance of *eIF4Es* in host translation, it is plausible that simultaneous mutation of two *eIF4Es* cause severe defects in plant growth. Despite these results, recent studies have suggested that natural resistance alleles would solve the problem attributed to the intrinsic function of *eIF4Es* to create a broad resistance spectrum. Tomato *pot-1-eIF4E1*-mediated resistance, which is attributed to a small number of *eIF4E1* aa mutations, impairs infection by PVY and TEV strains. In contrast, KO mutation of *eIF4E1* in tomato yields a much narrower spectrum of resistance. Furthermore, the broad spectrum of resistance found in natural alleles can be reconstituted by artificially combining KO mutations of *eIF4E1* and *eIF4E2* in tomato with severe growth defects, indicating that *eIF4E2* also has the capacity to participate in potyvirus infection, and that *pot-1-eIF4E* negatively regulates *eIF4E2* in the potyvirus infection cycle without no detriment to plant growth. Consistently, overexpression of the *eIF4E1* resistance allele in tomato provides dominant resistance to potyviruses, even in the presence of the intrinsic *eIF4E1* susceptible allele. These results suggest that *eIF4E1* resistance alleles negatively control *eIF4E2* with redundant function to obtain a broad spectrum of resistance. This may coincide with the fact almost all *eIF4E* resistance alleles identified to date sustain only several point mutations, but not null mutations.

New Recessive Resistance Genes Found in Natural Diversity

In the past decade, several distinct classes of recessive resistance genes other than *eIF4Es* have been discovered from a wide range of plant species. A few of these genes likely have functions similar to those of *eIF4Es* in recessive resistance, due to their known intrinsic functions in the translation machinery. Others appear to have a distinct recessive resistance mechanism, involving a negative defense regulator in recessive resistance. These newly identified recessive resistance genes show different specificities of resistance to plant viruses than those mediated by *eIF4Es*. Given candidate gene approaches have helped identify *eIF4E* resistance alleles in various crops, recent progress on the new recessive resistance genes will further promote to identify the resistance genes responsible for other plant-virus interactions, and will increase the potential of recessive resistance (in terms of the resistance spectrum and the gene repertoire) in the development of resistant cultivars in a wide range of crops.

In barley, although the recessive resistance genes *rym4/rym5*, which are alleles of *eIF4E*, are frequently overcome by BaMMV and BaYMV resistance-breaking isolates, cultivars with the *rym11* locus are highly resistant against both virus isolates. The *rym11* locus has been demonstrated to encode an allele of *protein disulfide isomerase like 5-1* (*PDIL5-1*), a distinct type of recessive resistance gene. A natural polymorphism survey of *PDIL5-1* genes suggested that most *rym11* resistance alleles are likely the results of frequent interactions with highly divergent forms of BaMMV and BaYMV in eastern Asia. *PDIL5-1* is an endoplasmic reticulum (ER)-localized chaperone in plants and animals; it plays an essential role in the unfolded protein response (UPR), a cellular machinery that alleviates unfolded protein overload in ER. Although the mechanism of *rym11-PDIL5-1*-mediated recessive resistance remains unknown, downregulation of other UPR components commonly impairs efficient viral infection in several plant-virus pathosystems, implying that they are promising targets for recessive resistance to a wide range of viruses.

Distinct classes of recessive resistance genes against potyviruses have also been cloned. Single-point mutation of eukaryotic translation initiation factor 2B β (*eIF2B β*) is responsible for the *retr03* resistance locus against TuMV in allopolyploid mustard. *eIF2B β* is a subunit of the *eIF2B* complex, which mediates guanine nucleotide exchange in *eIF2* in the later step of *eIF4E* in translation initiation. The role of *eIF2B β* in TuMV infection, although still unknown, may be as a host factor because of the

intrinsic function related to eIF4E in translation initiation. Similarly, a susceptibility factor EXA1 for potexviruses, identified from loss-of-susceptibility mutant of *Arabidopsis*, contains a putative eIF4E-binding motif. These findings imply that the factors functionally related to eIF4E are promising targets for recessive resistance. *eIF2B β* recessive resistance alleles are frequently found in several mustard cultivars, albeit those with allopolyploid genomes, suggesting the relevance of this gene in mustard–TuMV interaction. Another distinct type of recessive potyvirus resistance gene is nuclear-encoded *chloroplast phosphoglycerate kinase 2* (*cPGK2*), validated in *Arabidopsis* ecotypes, independently designated WMV resistance loci *rw1*, and *rpv1* against Plum pox virus (PPV). *cPGK2* gene homologs are found in a diverse range of plants, including dicots and monocots. Consistently, knockdown of *cPGK* genes in *N. benthamiana* compromises efficient accumulation of Bamboo mosaic virus (BaMV; *Potexvirus*, family *Alphaflexiviridae*). In addition, *cPGK2* is predicted to recruit BaMV genomic RNA to chloroplasts for BaMV replication. Although the exact molecular mechanism of *cPGK2*-mediated resistance in *Arabidopsis* ecotypes remains uncertain, *cPGK2* may be a general susceptibility factor for at least a subset of potyviruses and other distantly related viruses.

A mutation in *vacuolar protein sorting 41* (*VPS41*) gene has been demonstrated to be involved in resistance against CMV, mediated by the recessive allele *cmv1* in melon. *VPS41* plays a role in membranous vesicle trafficking from the late Golgi to the vacuole in eukaryotic cells. Intriguingly, in cucumber, the gene encoding another vesicle trafficking component, *VPS4*, has been proposed as a strong candidate of the recessive resistance locus *zym* against Zucchini yellow mosaic virus (ZYMV; *Potyvirus*, family *Potyviridae*). Although evidence for a specific molecular mechanism for *cmv1-VPS41*-mediated resistance is insufficient, several viruses appear to commonly recruit components of the vesicle trafficking machinery for efficient infection.

In an African rice (*Oryza glaberrima*) accession that is resistant against Rice yellow mottle virus (RYMV; *Sobemovirus*, family *Solemoviridae*), the *rymv2* resistance gene encodes an allele of the rice homolog of the *Arabidopsis* *CPR5* gene, whose dominant form plays a negative role in the plant defense response. Seven other African rice accessions also encode *rymv2* resistance alleles. Because of the *CPR5* function in *Arabidopsis* defense responses, *rymv2-CPR5*-mediated resistance is highly likely the first example of the activation of defense responses by recessive mutation of a negative defense regulator gene.

Remarkably, the recessive resistance locus *ty-5* in tomato has also been identified to confer resistance to a plant DNA virus, Tomato yellow leaf curl virus (TYLCV; *Begomovirus*, family *Geminiviridae*). The *ty-5* locus has been reported to encode an allele of *Pelota*, whose protein product is implicated in the recycling phase of translation, the latest step of protein biosynthesis. Although the resistance mechanism mediated by *Pelota* remains unclear, this result reinforces the relevance of the host translational machinery in recessive resistance despite significant differences in infection strategies among DNA and RNA viruses.

Collectively, the identification of these distinct classes of recessive resistance genes will broaden the potential for recessive resistance application to the plant–virus pathosystem, for which this type of resistance has never been described.

Toward the Application of Recessive Resistance in Agriculture

In the last ten years, due to the increase in the acquisition of genomic information and the development of techniques for the analysis of gene functions, greater numbers of studies have identified several classes of recessive resistance genes and provided mechanistic insight on recessive resistance. Loss of susceptibility to viral infection is genetically equivalent to natural recessive resistance. Therefore, random mutagenesis using crops and model plants would extend the potential to identify unknown susceptibility factors and find new ways to utilize host factors beyond the genetic variations found in naturally occurring cultivars. Model plants further facilitate the isolation of loss-of-susceptibility mutants and subsequent identification of responsible genes, due to the availability of high-quality whole-genome sequence, simple genetic approaches, and many other analytical tools. However, it should be noted that the generation of polyploid plant mutants with desired phenotypes by random mutagenesis is often difficult due to the presence of redundant genes.

Conventional breeding techniques based on crossbreeding and molecular marker selection has long been used to introduce the recessive resistance phenotypes from one cultivar to another elite cultivar. This method has contributed to the alleviation of crop yield losses caused by devastating viruses. However, the stable introduction of resistant traits to establish resistant cultivars through conventional techniques or by the targeting of a specific gene using random mutagenesis and screening are substantially technically challenging and time consuming.

Targeting induced local lesions in genomes (TILLING) is a technology applied to rapidly identify a mutant plant with genetic modification in a target gene by combining random mutagenesis by ethyl methanesulfonate (EMS) and single-nucleotide polymorphism (SNP) detection in a specific gene. TILLING has been successfully applied to establish a null allele of *eIF4E* gene in tomato showing loss of susceptibility to potyviruses. As discussed above, the range of resistance conferred by the null *eIF4E* allele is narrower than that of the natural functional allele *pot1*. Thus, at least when *eIF4E* is the gene of interest, functional resistance alleles are preferable to a null allele for obtaining a broad spectrum of resistance. EcoTILLING is a derivative of TILLING that is used to isolate the genetic variation of a target gene from that in natural cultivars. Using the EcoTILLING approach, several natural allelic variations of *eIF4E* gene have been found in wild relatives of melon; however, they do not correlate with potyvirus resistance phenotypes. Considering the results of other surveys of natural allelic polymorphism, EcoTILLING for natural resistance alleles may be performed more effectively with collections of crop cultivars, rather than their wild relatives.

Significant technological and conceptual progress is currently being made in further promoting the application of recessive resistance and loss of susceptibility for resistant crop breeding. Sometimes, the elimination of an undesired trait that is tightly linked to the resistance phenotype using conventional breeding techniques is difficult. However, the recent emergence of

next-generation sequencing (NGS) technology and genome editing could pave the way for the rapid introduction of recessive resistance phenotypes to crop cultivars as intended.

NGS technology promotes access to the whole-genome sequences of all organisms. Genomics-based crop breeding combined with NGS is expected to enable the rapid establishment of new crop cultivars with desirable traits. NGS technology, which have rarely been applied to antiviral crop breeding using natural variants, would be quite useful for the identification of loci including major resistant QTLs (e.g., QTL-seq) in naturally resistant variants, as well as for crop breeding, to introduce resistant loci into elite cultivars.

Genome editing technology is based on sequence-specific nucleases, including zinc-finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), and clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 9 (Cas9) in the system referred to as CRISPR/Cas. These technologies, which have emerged during the last two decades to enable mutagenesis and insertion at specific genomic sites in eukaryotic genomes, are now widely used in diverse crop species, as well as model plants. Targeted mutagenesis, made possible by genome editing, can overcome potential problems caused by existing technologies, such as the time-consuming aspect of random mutagenesis and potential tight linkage of undesired traits in crossbreeding.

One advantage of genome editing for crop breeding is that the original transgenes can be removed after editing via segregation. This feature permits the removal of bottlenecks for social acceptance of traditional genetically modified (GM) crops. The CRISPR/Cas9 system was applied to modify *eIF4E* genes in *Arabidopsis*, cucumber, and rice. As expected, the resulting mutant plants, which were free of transgenes, exhibited recessive resistance to potyviruses and waikaviruses. Multiple gene editing has been demonstrated in model plants *Arabidopsis* and rice, and is thus possible in polyploid plants. Genome editing technologies can be also applied to develop loss-of-susceptibility mutants by modifying host factors identified through basic research on plant-virus interactions. In this sense, genes encoding host factors with loss-of-susceptibility mutations in *Arabidopsis* and other plant species can be regarded as potential recessive resistance genes. As a proof of concept, nCBP, known as an *eIF4E* isoform, was isolated in cassava as an interactor with VPg proteins of two ipomoviruses, CBSV and UCBSV. CRISPR-Cas9-mediated editing of two cassava *nCBP* genes successfully generated a resistant phenotype. Considering that nCBP is a susceptibility factor for potexviruses in *Arabidopsis*, *nCBP* gene is a promising target of genome editing to obtain recessive resistance with distinct spectrum from *eIF4E*-mediated resistance. Furthermore, genome editing technologies are evidently compatible with the application of current knowledge of recessive resistance in antiviral resistance breeding of various crop species, including polyploids. Derived from the original CRISPR/Cas9 system, precise base editing has been accomplished using a new technology termed Target-AID, in which a nuclease-deficient Cas9 mutant is fused with activation-induced deaminase (AID), which catalyzes C to T mutations. This technology has been utilized to confirm that base editing of the *eIF4E* gene mimicking the pea *sbm1* allele achieved recessive resistance against CIYVV in *Arabidopsis*, where no natural *eIF4E* resistance allele has been found.

Conventionally, recessive resistance traits must be introduced into new crop species from closely related species that exhibit recessive resistance and can be crossbred. Since some crop species lack recessive resistance traits, random mutagenesis has been the only method for establishment of a resistant mutant in which a susceptibility gene is knocked out. However, in tomato, the *eIF4E1* KO mutant shows a much narrower potyvirus resistance spectrum than does the natural *pot1-eIF4E1* resistance allele. Considering this finding together with other results, researchers have hypothesized that the *pot1* resistance allele, which maintains function in translation, negatively regulates *eIF4E2* with redundant function in viral infections. Based on this insight, the researchers tested whether introgression of synthetic *eIF4E* (*eIF4E^R*) alleles, rationally-designed from the pea *sbm1* gene in trans-species manner, produces a broad resistance spectrum in the background of double KO of *eIF4E/eIFiso4E* (*eif4e^{KO}/eifiso4e^{KO}*) in *Arabidopsis*. The *Arabidopsis* genotype *eIF4E^R/eif4e^{KO}/eifiso4e^{KO}* confers no growth defect and shows a broad spectrum of resistance to several potyviruses, and more significantly, to WMV and one strain of Beet western yellows virus (BWYV; genus *Polytivirus*, family *Luteoviridae*), to both of which no recessive resistance has been documented in either KO mutant, *eif4e^{KO}* and *eifiso4e^{KO}*, in *Arabidopsis*. These results suggest the new crop breeding concept of developing a synthetic *eIF4E* allele that mimics natural resistance alleles from different species will retain intrinsic function in translation and shows a broad resistance spectrum in crops in which natural recessive resistance cannot be obtained by conventional breeding. However, it should be noted that the *Arabidopsis* genotype *eIF4E^R/eif4e^{KO}/eifiso4e^{KO}* does not show recessive resistance against all viruses tested. Thus, to control the viruses for which recessive resistance alleles have not been found yet, the identification of natural resistance alleles for these viruses may be necessary.

Conclusions

In this article, we addressed the characteristics of natural recessive resistance found in diverse plant species, the isolation and functional features of recessive resistance genes, and recent advances toward the application of recessive resistance in a wide range of crop species. Indeed, some of the recessive resistance genes summarized in Table 1 are currently utilized to control plant virus diseases in agriculture. Table 1 shows a strong bias toward recessive resistance against RNA viruses, perhaps due partly to research focus. Recent progress in the isolation of recessive resistance genes suggests the continuing relevance of *eIF4Es* as highly dominant recessive resistance genes, as well as the functional diversity of recessive resistance. Further research efforts should explore more diverse classes of recessive resistance genes in the near future. An unexpected regulatory mechanism of natural *eIF4E*-mediated recessive resistance has been proposed, such that a functional *eIF4E1* resistance allele likely negatively regulates *eIF4E2* with redundant functions. Recently emerged genome editing technologies are anticipated to be applied to the molecular breeding of recessive resistance in crops. In particular, the target-AID method will facilitate precise base editing, and is highly suitable for the design of a synthetic resistance allele mimicking natural resistance alleles in a trans-species manner.

As also discussed in this article, genome editing technologies are clearly promising tools for the application of recessive resistance to various crop species. However, social acceptance of genome-edited crops is uncertain in many parts of the world, partly because evidence supporting the safety of genome-edited crops seems to be insufficient due to still developing technologies. Indeed, although some genome-edited crops have been approved without restriction in the USA, the Court of Justice of the European Union decided to make genome-edited crops subject to the same regulations as conventional GM crops. Further basic research is essential to enhance the utility of recessive resistance and support the safety of genome editing technologies for recessive resistance-based crop breeding.

Further Reading

- Bastet, A., Robaglia, C., Gallois, J.L., 2017. eIF4E resistance: Natural variation should guide gene editing. *Trends in Plant Science* 22, 411–419.
- Bastet, A., Lederer, B., Giovinazzo, N., *et al.*, 2018. Trans-species synthetic gene design allows resistance pyramiding and broad-spectrum engineering of virus resistance in plants. *Plant Biotechnology Journal* 16, 1569–1581.
- de Oliveira, L.C., Volpon, L., Rahardjo, A.K., *et al.*, 2019. Structural studies of the eIF4E-VPg complex reveal a direct competition for capped RNA: Implications for translation. *Proceedings of National Academy of Science of the United State of America* 116, 24056–24065.
- Fraser, R.S.S., 1990. The genetics of resistance to plant viruses. *Annual Review of Phytopathology* 28, 179–200.
- Hashimoto, M., Neriya, Y., Yamaji, Y., Namba, S., 2016. Recessive resistance to plant viruses: Potential resistance genes beyond translation initiation factors. *Frontiers in Microbiology* 7, 1695.
- Kang, B.C., Yeam, I., Jahn, M.M., 2005. Genetics of plant virus resistance. *Annual Review of Phytopathology* 43, 581–621.
- Robaglia, C., Caranta, C., 2006. Translation initiation factors: A weak link in plant RNA virus infection. *Trends in Plant Science* 11, 40–45.
- Sanfaçon, H., 2015. Plant translation factors and virus resistance. *Viruses* 7, 3392–3419.
- Truniger, V., Aranda, M.A., 2009. Recessive resistance to plant viruses. *Advances in Virus Research* 75, 119–159.
- Wang, A., Krishnaswamy, S., 2012. Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *Molecular Plant Pathology* 13, 795–803.

Plant Viral Diseases: Economic Implications

Basavaprabhu L Patil, ICAR–Indian Institute of Horticultural Research, Bengaluru, India

© 2021 Elsevier Ltd. All rights reserved.

This is an update of G. Loevenstein, Plant Virus Diseases: Economic Aspects, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, <https://doi.org/10.1016/B978-012374410-4.00470-2>.

Nomenclature

Ha Hectares
mHa Million hectares
mt Million tonnes
SP Stem pitting
SSA Sub-Saharan Africa
ssDNA Single strand DNA

ssRNA Single strand RNA
SY Seedling yellows
t Tonnes
t/Ha Tonnes per hectare
USA United States of America
WTGs Whitefly-transmitted geminiviruses

Glossary

Non-persistent manner A mode of virus transmission characterized by short acquisition and inoculation times (from seconds to minutes).

Non-propagative manner A mode of virus transmission in which the virus is not internalized by the vector: the virus is carried externally by attachment to the mouthparts or foregut of the vector.

Pseudo-recombination Component exchange between two geminiviruses. The usual mechanism of pseudo-recombination is by a process known as ‘regulon grafting’: the A component of a virus donates its common region by recombination to the B component of a second virus being captured. This results in a new dependent interaction between two components.

Introduction

Estimates of losses as a result of pest or pathogen attack in crop plants are mostly lacking. Whatever meager amount of data is available is not necessarily accurate nor up-to-date. Such data from different countries assessed by different methods should be taken as only a rough estimate. Nevertheless, the available data gives an approximate idea on the tremendous losses to the food and fiber production for the increasing global population.

During 1988–1990 a team of Scientists from Germany, supported by the European Crop Protection Association, published a comprehensive report on pest-induced crop losses on major food and cash crops. In the four principal cereal crops, rice, wheat, barley, and maize, the yield losses due to diseases were in the range of 10%–16% which translates to approx. \$64 billion. When four more crops (potatoes, soybeans, cotton, and coffee) were taken into account the monetary estimate of yield loss increased to \$84 billion, with an estimated harvests worth \$300 billion in 1988–1990. However, this analysis did not cover some important food crops, such as cassava, millet, sorghum, and the horticultural crops, cultivated in developing countries, which are also significantly affected by viral diseases. It was also revealed that the pests and diseases accounted for pre-harvest losses of 42% of the potential value of output during 1988–1990, with 15% due to insect damage and 13% each to weeds and pathogens. An additional 10% of the potential yield was lost post-harvest.

Among different plant pathogens, the fungal pathogens cause maximum damage with an estimated loss of 40%–60%, followed by the viral diseases causing a loss of 10%–15% and the rest of the losses are caused by bacterial pathogens, phytoplasma, and nematodes. The above data should be cautiously interpreted, as such survey data are mere rough estimates and probably the selection of crops was different and fewer viral diseases were considered. In several crops the yield losses caused by different viral diseases can range from 0% to 100% (Table 1). For example, in one of the surveys conducted in eight African countries, the yield losses because of cassava mosaic disease (CMD) were estimated to be in the range of 30%–40%. Barley yellow dwarf viruses (BYDVs) has a global distribution and infect over 150 species of the family Poaceae, including staple cereals such as wheat, barley, oats, rye, rice, and maize. A number of studies have demonstrated the devastating effect these viruses may have on cereals and the seed yield reductions are in the range of 15%–40% depending on the virus isolate and the cereal crop species.

The economic impact caused by viral diseases is not reduction in yield alone, but also deterioration in the quality of the plant produce, such as deformations in peaches affected by Plum pox virus (PPV) or squash with Zucchini yellow mosaic virus (ZYMV) infection. In addition, major economic aspects to be considered are the expenses incurred in production of virus-indexed propagation material, eradication programs, control of insect vectors, and breeding for resistance to viruses. The viral epidemics and the subsequent economic impact are more prominent in tropical and subtropical regions with continuous vegetation round the year, which enables persistence and proliferation of the viruses and their vectors. In contrast, continuous herbaceous vegetation is interrupted during winters

Table 1 Estimated yield or monetary losses due to viral diseases in different crop plants

<i>Crop</i>	<i>Country/Region</i>	<i>Viral name (Acronym)</i>	<i>Virus taxonomy (Genus/Family)</i>	<i>Monetary/Yield losses</i>
<i>Cereals</i>				
Barley, oat, and wheat	Worldwide	Barley yellow dwarf viruses (BYDVs)	<i>Luteovirus/Luteoviridae</i>	Up to 50% yield loss
Maize	Worldwide	Cereal yellow dwarf viruses (CYDVs)	<i>Polerovirus/Luteoviridae</i>	11%–12% yield loss
	Africa	Maize streak virus (MSV)	<i>Mastrevirus/Geminiviridae</i>	17%–71% yield loss
	USA	Maize dwarf mosaic virus (MDMV)	<i>Potyvirus/Potyviridae</i>	42%–75% yield loss
Rice	South and South East Asia	Rice tungro bacilliform virus (RTBV) and Rice tungro spherical virus (RTSV)	<i>Tungrovirus/Caulimoviridae</i>	US\$1.5 billion
	Africa	Rice yellow mottle virus (RYMV)	<i>Waikavirus/Secoviridae</i>	
	South America	Rice hoja blanca virus (RHBV)	<i>Sobemovirus/Solemoviridae</i>	84%–97% yield loss
Wheat	USA	Wheat streak mosaic virus (WSMV)	<i>Tenuivirus/Phenuiviridae</i>	25%–50% yield loss
			<i>Tritimovirus/Potyviridae</i>	27%–75% yield loss
<i>Root and Tuber Crops</i>				
Cassava	Africa & Indian Subcontinent	African cassava mosaic viruses (ACMVs), and East African cassava mosaic viruses (EACMVs)	<i>Begomovirus/Geminiviridae</i>	US\$1.2–2.3 billion per year
Potato	East Africa	Cassava brown streak viruses (CBSVs)	<i>Ipomovirus/Potyviridae</i>	US\$75 million
	Worldwide	Potato leafroll virus (PLRV)	<i>Polerovirus/Luteoviridae</i>	50%–60% yield loss
	UK	Potato virus X (PVX), Potato virus Y (PVY), and Potato leafroll virus (PLRV)	<i>Potexvirus/Alphaflexiviridae</i>	US\$5.5 million
			<i>Potyvirus/Potyviridae</i>	Up to 80% Yield loss
			<i>Polerovirus/Luteoviridae</i>	
	Worldwide	Potato virus X (PVX)	<i>Potexvirus/Alphaflexiviridae</i>	5%–15% yield loss
	Eastern Europe	Potato virus S (PVS)	<i>Carlavirus/Betaflexiviridae</i>	3%–20% yield loss
	Eastern Europe	Potato virus M (PVM)	<i>Carlavirus/Betaflexiviridae</i>	40%–75% yield loss
Sweet potato	Worldwide, East Africa in particular	Sweet potato feathery mottle virus (SPFMoV) and Sweet potato chlorotic stunt virus (SPChSV)	<i>Potyvirus/Potyviridae</i>	50%–90% in combined infections
			<i>Crinivirus/Closteroviridae</i>	
<i>Vegetables</i>				
Cucurbits	Worldwide	Cucumber mosaic virus (CMV)	<i>Cucumovirus/Bromoviridae</i>	10%–20% yield loss
	Worldwide	Zucchini yellow mosaic virus (ZYMV)	<i>Potyvirus/Potyviridae</i>	Up to 100% yield loss
Tomato	Worldwide	Tomato yellow leaf curl viruses (TYLCVs)	<i>Begomovirus/Geminiviridae</i>	93%–100% yield loss (in Jordan)
	Worldwide	Tomato leaf curl New Delhi virus (ToLCNDV)	<i>Begomovirus/Geminiviridae</i>	93%–100% yield loss (in Jordan)
	Worldwide	Tomato spotted wilt virus (TSWV)	<i>Orthotospovirus/Tospoviridae</i>	US\$1 billion in 1994
<i>Fruits</i>				
Banana	Worldwide	Banana streak viruses (BSVs)	<i>Badnavirus/Caulimoviridae</i>	Up to 82% yield loss
	Worldwide	Banana bunchy top virus (BBTV)	<i>Babuvirus/Nanoviridae</i>	US\$50 million in India during 2007–2010
	Asia	Banana bract mosaic virus (BBRMV)	<i>Potyvirus/Potyviridae</i>	30%–70% in India & Philippines
Cacao	Africa	Cacao swollen shoot viruses (CSSVs)	<i>Badnavirus/Caulimoviridae</i>	up to 100% yield loss
Citrus	Worldwide	Citrus tristeza virus (CTV)	<i>Closterovirus/Closteroviridae</i>	Since 1919, 80 million citrus trees eradicated
Grape	France	Grapevine fanleaf virus (GFLV)	<i>Nepovirus/Secoviridae</i>	US\$100 million
Papaya	Worldwide	Papaya ring spot virus (PRSV)	<i>Potyvirus/Potyviridae</i>	Up to 70% yield loss
Stone fruits	Worldwide	Plum pox virus (PPV)	<i>Potyvirus/Potyviridae</i>	US\$10 billion during 1976–2006
<i>Legumes</i>				
Groundnut	Africa	Groundnut rosette virus (GRV)	<i>Umbravirus/Tombusviridae</i>	US\$100 million
		Groundnut rosette assistor virus (GRAV)	<i>Unassigned/Luteoviridae</i>	
	India	Groundnut bud necrosis virus (GBNV)	<i>Orthotospovirus/Tospoviridae</i>	US\$89 million
	India/West Africa	Peanut clump viruses (PCVs)	<i>Pecluvirus/Virgaviridae</i>	US\$38 million
Legumes	Worldwide	Bean yellow mosaic virus (BYMV)	<i>Potyvirus/Potyviridae</i>	US\$300 million (in India)
Pigeonpea	India	Pigeonpea sterility mosaic viruses (PPSMVs)	<i>Emaravirus/Fimoviridae</i>	US\$300 million
Soybean	Worldwide	Soybean mosaic virus (SbMV)	<i>Potyvirus/Potyviridae</i>	8%–25% yield loss

Table 1 Continued

<i>Crop</i>	<i>Country/Region</i>	<i>Viral name (Acronym)</i>	<i>Virus taxonomy (Genus/Family)</i>	<i>Monetary/Yield losses</i>
<i>Cash Crops</i>				
Cardamom	India	Cardamom mosaic virus (CarMV)	<i>Macluravirus/Potyviridae</i>	70%–100% yield loss
Cotton	India/Pakistan	Cotton leaf curl viruses (CLCVs)	<i>Begomovirus/Geminiviridae</i>	68%–71% yield loss US\$5 billion in Pakistan, during 1992–1997
Sugarcane	Worldwide except Guyana and Mauritius	Sugarcane mosaic virus (SCMV)	<i>Potyvirus/Potyviridae</i>	30%–80% yield loss
Tobacco/ Tomato	Worldwide	Tobacco mosaic virus (TMV)	<i>Tobamovirus/Virgaviridae</i>	~1% yield losses in Tobacco and Up to 20% in Tomato

in the temperate regions and also in arid regions with dry hot summers, impacting negatively the vector population and subsequent virus transmission is significantly reduced. Furthermore the spread and prevalence of viral diseases is much higher in plants that are vegetatively propagated, such as cassava, potato, and citrus, than in those that are propagated through seeds. With a few exceptions, a majority of viruses are not seed transmitted, because of lack of vascular or the plasmodesmatal connection between the embryo and its parent plant. In vegetatively propagated plants, when the seed tubers, cuttings, bud wood, bulbs, runners, *etc.*, are derived from infected plants viruses are transmitted to the subsequent generations. This is in addition to spread of the virus by vectors, which may occur in both cases. In the last two decades the whitefly transmitted viruses have been on the raise in several crops, including vegetables, often becoming a major constraint of production. In this article, it is not possible to cover the economic impact of all plant viral diseases, however we have made an effort to highlight the impact of most important virus diseases on economically important crops (**Table 1**).

Cereal Viral Diseases

Globally wheat is cultivated in an area of 218 mHa with a total production of 772 mt. Diseases caused by viral pathogens are the major threat to early planted wheat in the southern Great Plains of United States (USA). The Wheat streak mosaic virus (WSMV, genus *Tritimovirus*, family *Potyviridae*) transmitted by wheat curl mite causes grain and forage yield losses that significantly affect profits from wheat production. The yield losses in wheat crop due to WSMV ranged from 27% to 75%. WSMV shows synergistic interactions with other viruses (specifically Triticum mosaic virus and High plains virus), further contributing to yield reductions. WSMV infects some of the most agriculturally important members of the family Poaceae, including wheat, corn, rye, oats, barley, sorghum, millets, and the grassy weeds that may act as alternate hosts. Barley yellow dwarf virus (BYDV; genus *Luteovirus*, family *Luteoviridae*), transmitted by aphids, is another important virus known to infect economically important crops such as barley, oats, wheat, maize, triticale, and rice. The yield losses caused by BYDV infection can vary widely depending on the viral strain, time of infection and rate of spread. Maximum yield losses up to 50% occur from early infections. Brome mosaic virus (BMV, genus *Bromovirus*, family *Bromoviridae*) is another economically important virus infecting barley and other cereal crops. BMV is also known to co-infect triticale along with WSMV. As per a study conducted in Ohio (USA), BMV could reduce the yield by as much as 61% in soft red winter wheat.

Corn cultivated in a global area of 197 mHa, is also affected by more than 25 viral diseases. The total estimated corn loss (bushels) by different viral diseases in the United States and Canada, during 2012–2015 was estimated to be 23.6 million bushels and of this ~46% was by Maize dwarf mosaic virus (MDMV, genus *Luteovirus*, family *Luteoviridae*) alone. Maize streak virus (MSV; genus *Mastrevirus*, family *Geminiviridae*) transmitted by six leafhopper species, is one of the most devastating viruses that occurs throughout Africa on corn. MSV epidemiology is controlled by environmental conditions on its vector, resulting in epidemics every 3–10 years. During the epidemic years, disease incidence can vary from a few infected corn plants per field, with little associated yield loss, to 100% infection rates and complete yield loss (**Fig. 1**).

Root and Tuber Crops Viral Diseases

Cassava Viral Diseases

Cassava is an important food security crop for small holder and marginal farmers of developing countries. Cassava is a climate resilient crop with exceptional drought tolerance and is a staple food for close to one billion people in the tropics. Globally, cassava is cultivated in an area of 26 mHa, with a total production of 292 mt and of this total area, 76% area is in African continent itself accounting for a total production of 178 mt. However, because of its vegetative propagation, cassava is affected by at least 20 different viruses worldwide. In sub-Saharan Africa, these viruses are persistently posing a big threat to food security, and resulting in a monetary loss of more than \$1 billion annually. In particular, the two groups of whitefly transmitting cassava infecting viruses, namely begomoviruses (genus *Begomovirus*, family *Geminiviridae*) with single strand DNA (ssDNA) as their genome and Cassava brown streak viruses (genus *Ipomovirus*, family *Potyviridae*) with single strand RNA (ssRNA) as their genome,

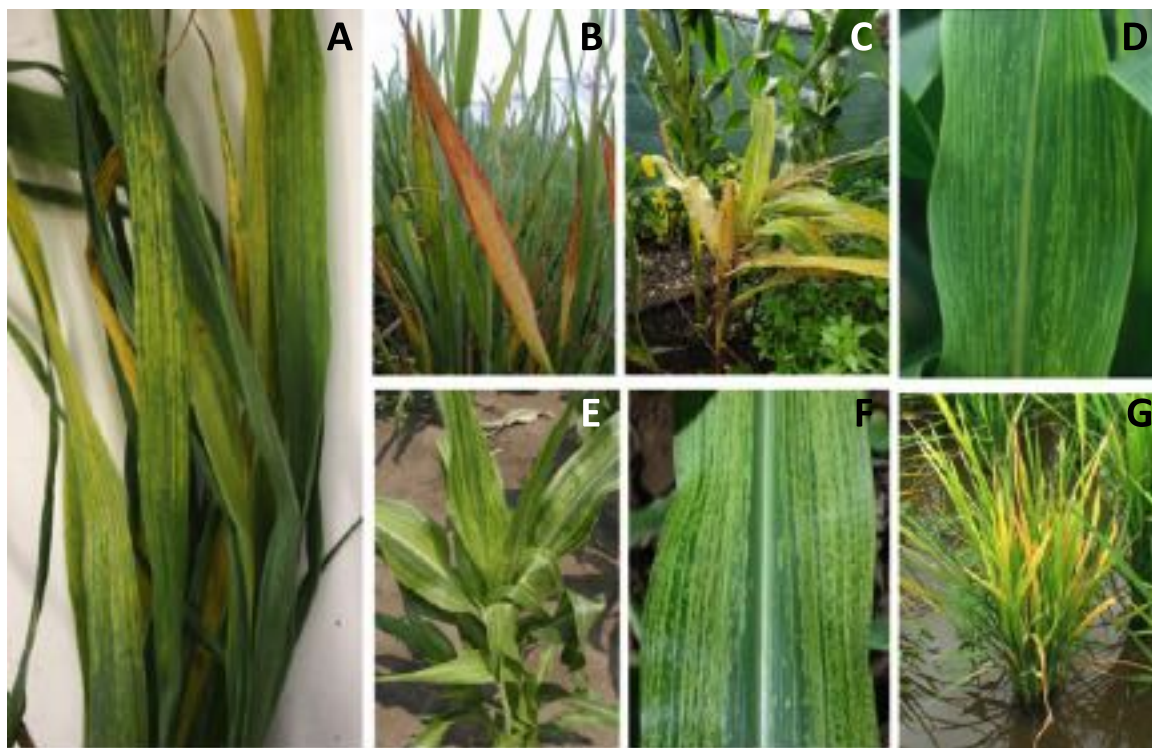


Fig. 1 Characteristic symptoms of important cereal viral diseases: (A) Wheat streak mosaic virus (WSMV) symptoms on wheat, (B) Barley yellow dwarf virus (BYDV) symptoms on barley, (C) Corn plant infected with a severe strain of Maize dwarf mosaic virus (MDMV), (D) Symptoms of Maize dwarf mosaic virus (MDMV) on corn leaf, (E) Corn infected with Maize streak virus (MSV), (F) Symptoms of Maize streak virus (MSV) on a leaf corn, (G) Rice tungro disease (RTD) on rice. (B) Source: Brian Olson, Oklahoma State University, Available at: Bugwood.org. (C) Source: <https://krex.k-state.edu/dspace/bitstream/handle/2097/20917/KSUL0006Claflini37262.jpg?sequence=1&isAllowed=y>. (E) Source: <https://alchetron.com/Maize-streak-virus>. (G) Source: IRRI.

have been responsible for the severe pandemics in Africa. The cassava infecting geminiviruses cause cassava mosaic disease (CMD) and the cassava infecting ipomoviruses cause cassava brown streak disease (CBSD). Currently, eleven distinct species of begomoviruses are involved in CMD, of these 9 are reported from Africa and 2 from the Indian subcontinent. In Africa, these 9 begomoviruses can be broadly categorized into African cassava mosaic viruses (ACMVs) and East African cassava mosaic viruses (EACMVs). ACMV and EACMV-like viruses can either occur singly or as mixed infections causing more severe symptoms. The yield losses in different cassava cultivars, in different African countries range from 20% to 95%. Estimates done in mid-1990s, revealed that the yield losses in the entire African continent was 12–23 mt of fresh edible roots per year, worth about \$1.2–2.3 billion. Starting from 1988 till date, an unusually severe form of CMD pandemic is spreading across East and Central Africa, affecting food supplies and creating a famine in Uganda in the late 90s. In 1999, this CMD pandemic had affected at least nine countries in East and Central Africa, encompassing an area of 2.6 million square kilometers and causing an estimated loss of \$1.9–2.7 billion per year. This severe form of CMD has been attributed at first to a recombinant strain of cassava mosaic virus. Later studies revealed that synergistic interactions between ACMV and a Ugandan strain of EACMV on mixed infection was mainly responsible for this severe form of CMD. The local cassava cultivars were highly susceptible to this recombinant strain and due to heavy losses many farmers abandoned cassava cultivation, resulting in destabilization of food security in Uganda. It was estimated that each year an area of c. 60,000 ha of cassava yielding 600,000 t of edible roots worth \$60 million was lost due to CMD. Thus CMD was considered as the most damaging plant viral disease in the world, causing famine and death of thousands of people in Uganda which lost 90% of its production by 1998.

In 2010, the Science journal published a report on seven most dangerous plant diseases in the world that can significantly affect the food security. Of all the plant diseases, CBSD was the only viral disease to be listed in this report. Although the first report of CBSD was made in 1930s from coastal region of Tanzania, in contrast to CMD, CBSD received much less attention, since it was restricted to the coastal low lying areas of East Africa, it appeared sporadically and did not cause any significant loss. However, post 2004, CBSD has been spreading across East and Central Africa at an alarming rate, with reports of CBSD incidence in highland areas too and hence risking the food security of millions of farmers in Africa.

One of the estimates on incidence of CBSD, by National Agricultural Research Organization (NARO, Uganda), put the overall incidence at 16% in 2008 and 29% in 2009, clearly indicating the rapid expansion of the new CBSD epidemic. Shortly after the reports of spread in Uganda, similar observations were made in western Kenya and the Lake Victoria zone of Tanzania. A survey in 19 districts of Tanzania revealed a steep increase in CBSD incidence from 6% in 2006 to 12% in 2007 and in 2008 it was 32%.

Most of the increase in disease incidence was attributed to greater levels of disease in districts in which CBSD was already present by 2006 (12 of 19 districts).

In addition to the reduced growth of the cassava plants as a result of CBSD infection, significant losses occur due to spoilage of harvested roots due to necrotic rots. There have been very few quantitative assessments of yield losses and one such assessment done in southern coastal Tanzania revealed that the losses could occur to an extent of 70% in susceptible cassava cultivars. As the CBSD infected cassava plants mature, the root symptoms can further increase and hence farmers opt for early harvest to prevent further damage of roots. The effects on roots could be variable, such as: reduction in the quality of roots caused by pitting, constrictions and root necrosis, as well as yield reduction through reduction in the number and weight of tuberous roots. An estimate done in 2001, indicated an overall loss of 20%–25%, equivalent to a monetary value of \$6–7 million. An effort was made to estimate the economic impact due to CBSD in Africa through development of a framework. This revealed that there are some of the undocumented elements of CBSD yield loss, such as additional labor cost incurred to separate necrotic from non-necrotic portions of affected roots and the deleterious effects on starch quality of non-necrotic portions of affected roots. A total loss of \$75 million was estimated for all the eight countries of East and Central Africa affected by CBSD.

For control of both CMD & CBSD, the main emphasis has been on the development of virus resistant cassava cultivars, phytosanitary practices such as using cuttings from healthy plants, elimination of infected plants, and production of virus-free tissue-cultured plant propagules. CMD can also be controlled by introgressing one of these three sources of resistance: polygenic recessive resistance, termed CMD1, the dominant monolocus type, named CMD2, and the recently identified CMD3 locus. Application of pesticides to stop the spread of CMD by controlling the whitefly vectors has not been successful.

Potato Viral Diseases

Globally, potato is one of the major crops cultivated in an area of 19 mHa with a total production of 388 mt per year. Since potato is mainly vegetatively propagated, viral diseases are major constraint for cultivation of potato and the yields are reduced beyond 50%. Since potato is an economically important crop maximum research was devoted to it in the first half of the 20th century. Several new tools and techniques such as, serology, meristem cultures, and advanced microscopy were developed for management of potato infecting viruses. More than 37 viruses are known to naturally infect potato, with Potato virus Y (PVY; genus *Potyvirus*, family *Potyviridae*), Potato virus X (PVX; genus *Potexvirus*, family *Alphaflexiviridae*), Potato leafroll virus (PLRV; genus *Polerovirus*, family *Luteoviridae*), Potato virus S (PVS; genus *Carlavirus*, family *Betaflexiviridae*), Potato virus M (PVM; genus *Carlavirus*, family *Betaflexiviridae*), Potato aucuba mosaic virus (PAMV; genus *Potexvirus*, family *Alphaflexiviridae*), and Potato mop-top virus (PMTV; genus *Pomovirus*, family *Virgaviridae*) being the most economically important and widely spread. Additionally, Potato spindle tuber viroid (PSTVd; genus *Pospiviroid*, family *Pospiviroidae*) also cause major damages.

PLRV, present across the globe, is reported to cause 50%–60% yield losses in potato and this virus is tuber borne and mainly vectored by aphids in a circulative and non-propagative manner. Whenever potato plants are co-infected with PLRV and PVX or PVY, yields are reduced to an extent of 70%.

PVY first described in 1930s, has a global distribution and has been known for several decades as a threat to seed, ware and processed potatoes. In addition to potato it is also known to infect several other host plants in Solanaceae family, such as tobacco, tomato, and pepper, causing yield losses up to 80%. PVY is transmitted by more than 40 aphid species in a non-persistent manner. The isolates of PVY are highly variable at the biological, serological, and molecular levels. Due to lack of source of resistance to PVY and its plant-to-plant transmission through daughter tubers, the main control strategy used is primarily based on certification of seed production. A severe group of PVY strains named as PVY_{NTN} are reported from Europe, Japan, and the US, and cause a severe form of disease in potato. The infected tubers develop superficial rings that later are sunken and necrotic. During storage these necrotic damages often become more conspicuous, and can affect 90% of tubers in susceptible cultivars. In spite of the latest improvements in diagnostic methods, routinely applied methods are unable to accurately characterize isolates responsible for tuber necrosis. Currently there is no serological assay, which can distinguish between PVY_{NTN} and other PVY strains. Hence, there is no efficient means to manage the risks of epidemics caused by the emerging necrotic variants of PVY. The global potato market is worth several billion dollars and highly competitive, with poor understanding of the PVY-host interactions involved in induction of necrotic symptoms and inefficient diagnostic tools have led to a situation in which necrotic strains of PVY can potentially cause huge agronomic and economic losses. Furthermore, when PVY occurs along with PVX the yield losses are much more drastic.

Crop yields can be further reduced by 5%–15% whenever plants are secondarily infected by PVX and PVS. PVM and PVS are common in countries of Eastern Europe. PVM can result in 40%–75% loss in tuber yield and it generally occurs in complex with other viruses (especially PVS). While with infection by PVS isolates tuber yield may be reduced by 3%–20%. Plants propagated from infected stocks, for several generations, may suffer more severe losses. Thus, for example, during the years 1993–1995, potato yields in Kazakhstan were extremely low, averaging 9 t/Ha. In a pilot project funded by USAID, micro-tubers obtained from meristems of potatoes screened for virus, were used, which resulted in a low infection rate in the elite seed tubers. These elite seeds enhance the yields by about 90% when compared with tuber yields from commercial fields in the neighborhood.

Potato yields in erstwhile Soviet Union countries were low, averaging 11 t/Ha during 1993–1995 as there was no reliable system for providing certified virus-free tuber seeds, and farmers used seeds from their own farms. Reduced yields are mostly due to virus diseases carried over in the planting material. Thus, Loshitsky and Belorusky 3N cultivars developed by the Belarusian Research Institute for Potato Growing (BRIP) yielded 30–50 t/Ha in the experimental fields, while Belarus' average potato yields

were in the range of 12.2 and 16.3 tonnes per hectare. Assuming that the viral diseases in potato cause 20% damage, that would be translated to a total loss of 18 mt of potato in former USSR with 6 mHa under potato cultivation, with a total production of 70 mt. For any estimate for monetary implication, it is also important to consider the expenses involved in seed potato certification schemes. The main objective of such schemes is to produce true-to-type disease-free (mainly viruses) potato seed tubers. Such schemes mostly include *in vitro* produced plantlets for the initial increase of pathogen-tested clonal selections, including production of mini-tubers. Price of certified tuber seeds are almost double the price for ware potatoes.

Sweet Potato Viral Diseases

Sweet potato (*Ipomoea batatas*) is cultivated in both tropical and warm temperate regions and is the seventh most important food crop in terms of production. In 2017, globally sweet potato was cultivated in an area of 9 mHa, with a total production of 113 mt, with China as the leading producer. Sweet potato is called 'poor man's crop', since most of the production is done on a small or subsistence level and can be cultivated without much inputs such as fertilizers and irrigation, and is one of the crops that can survive the famines. Sweet potato produces more biomass and nutrients per hectare than any other food crop in the world. Sweet potato is a major crop in East Africa and thousands of villages depend on sweet potato for food supply and the Japanese use it when typhoons demolish their paddy fields. Yields vary greatly in different areas or even fields in the same region. The average yields in African countries is about 4.7 t/Ha, while the yields in Asia are significantly higher, averaging to 18.5 t/Ha. Such large variation in yields is attributed to variation in quality of the propagation material. Sweet potatoes are vegetatively propagated from vines, root slips (sprouts) or tubers, and generally the farmers in African countries use vines from their own fields for propagation year after year. Thus, use of virus infected material for propagation results in reduced yield and often these fields are infected with several viruses, thereby compounding the effect on yields. Some of the viruses that are reported to infect sweet potato are Sweet potato feathery mottle virus (SPFMV; genus *Potyvirus*, family *Potyviridae*), Sweet potato mild mottle virus (SPMMV; genus *Ipomovirus*, family *Potyviridae*), Sweet potato latent virus (SPLV; genus *Potyvirus*, family *Potyviridae*), and Sweet potato chlorotic stunt virus (SPCSV; genus *Crinivirus*, family *Closteroviridae*). Infection by both SPFMV and SPCSV causes the very severe sweet potato virus disease (SPVD) and it is the most important disease of sweet potato in Africa. SPVD can cause 50%–80% yield losses in Uganda and Kenya. In a 3-years field study in Cameroon, SPVD incidence resulted in a yield loss of 56%–90%, especially in the susceptible cultivars. Similarly, 78% yield reductions have been reported from field trials in Nigeria due to SPVD. Yield reduction of 50% were reported from a 2-year field experiment in Israel, in plots planted with SPVD-infected cuttings, while infection by SPFMV or SPCSV alone did not have any significant effect on yield. In China, the average yield losses due to sweet potato viral diseases are of over 20% and the infection rate in the Shandong province can range from 5% to 41%. In countries where virus-free planting material is provided, for instance in the USA and Israel, markedly higher yields of 16.3 and 30 t/Ha respectively, are obtained. Cucumber mosaic virus (CMV) alone does not infect sweet potato plants that are healthy and virus free, however, superinfection by CMV can cause a complete crop failure when sweet potatoes are carrying SPCSV.

Most effective way to control viral diseases of sweet potato is through supply of virus-indexed healthy planting material, which is being practiced in Israel and in the Shandong province of China. Planting virus-tested material increased yields by at least by 100% in Israel, while in China the yield increase ranged from 22% to 92%. In Israel use of certified material costing 3 US cents per cutting, prepared by special nurseries, is a common practice. Initially, farmers buy about 30% of planting material required to plant their fields and later use the cuttings from this to fill in the rest. Yields are high and stable when virus-free material is used. In African countries virus-indexing programs operate only on a limited scale, since sweet potatoes are mainly grown as a food security crop, and not as a commercial one. Resistance breeding programs might be an answer for introgressing SPVD resistance and such programs are in operation in Uganda. One has to see if these improved sweet potato cultivars have stable, durable and broad spectrum resistance. An example for unstable resistance was reported when several clones resistant to SPFMV from the International Potato Center (CIP, Peru) were found to be susceptible when Israeli and Ugandan isolates of SPFMV were used for screening (Fig. 2).

Vegetable Viral Diseases

Chilli Viral Diseases

Chilli or hot pepper (*Capsicum annum* L., family Solanaceae) is economically most important and valuable spice and vegetable crop throughout the world. Chilli is native to New Mexico and the Portuguese introduced chillies in Indo-Pak sub-continent from Brazil in 16th century. India is the world's largest producer, consumer and exporter of chillies in the world, contributing 40% of the total world production. Other major chillies producing countries are China, Bangladesh, and Peru.

Chilli is vulnerable to several pathogens, including the viral diseases and till date 65 viruses are reported to infect chilli resulting in heavy yield losses. Mostly, the viruses are spread or transmitted by infected seeds, insects and or by the mechanical tools. Most of the viral diseases are difficult to diagnose due to significant overlap of symptomatology. Symptoms produced by different viruses are mosaic pattern on leaves, yellowing, ring spots, leaf deformation, curling of leaves, and stunting of plants. Common viral diseases are; Alfalfa mosaic virus, Beet curly top virus, Cucumber mosaic virus, Pepper mottle virus, Tobacco mosaic virus, Chilli leaf curl virus, and Pepper mosaic virus, which are transmitted by aphids and whitefly vectors. Chilli leaf curl viral disease

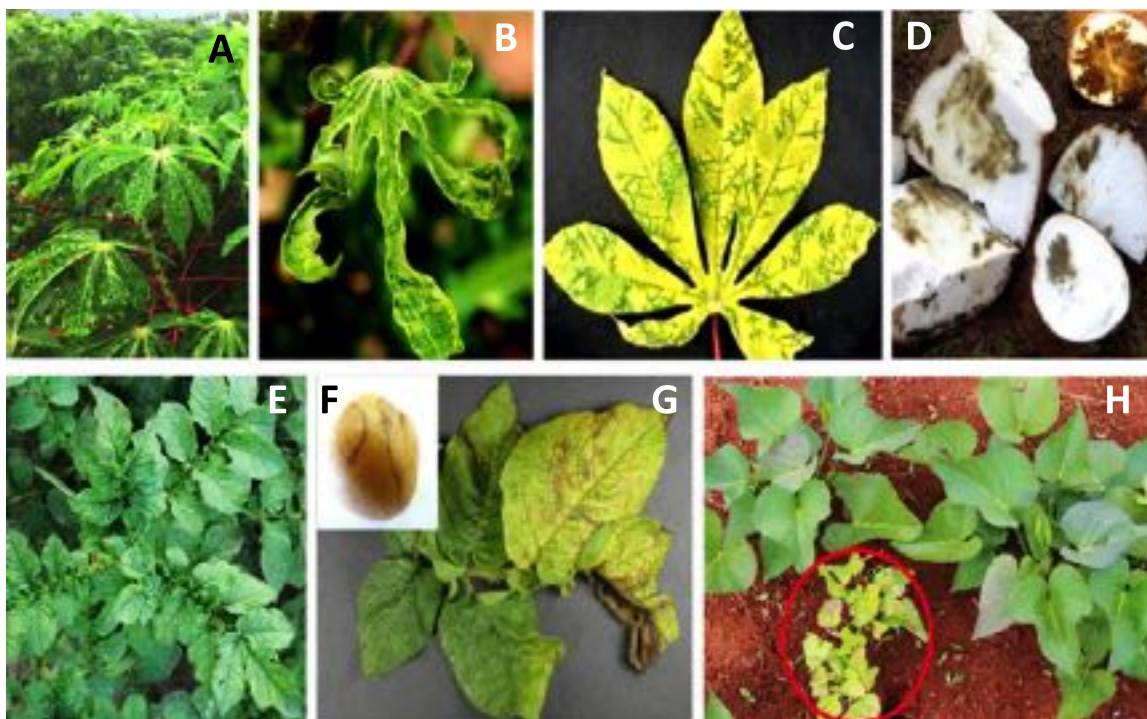


Fig. 2 Characteristic symptoms of important root and tuber crops viral diseases: (A) African cassava mosaic virus (ACMV), on cassava, (B) East African cassava mosaic virus (EACMV), on cassava leaf (Courtesy of James Legg), (C) Cassava brown streak virus (CBSV), on an old cassava leaf, (D) Cut roots of cassava infected with Cassava brown streak virus (CBSV), (E) Potato virus Y (PVY); symptoms on potato leaves, (F) Potato virus Y (PVY); symptoms on potato tuber, (G) Necrosis induced by Potato virus Y (PVY); on potato leaf, (H) Sweet potato virus disease (SPVD) on sweet potato (in the red circle) and healthy plants. (A) Source: <https://www.labmanager.com/news/2019/04/gene-editing-technology-may-produce-resistant-virus-in-cassava-plant#.XgOqPy2ZPyk>. (B) Courtesy of James Legg. (C) Courtesy of Claude Fauquet. (E) Source: <https://cpb-us-e1.wpmucdn.com/blogs.cornell.edu/dist/5/6962/files/2018/02/Severe-mosaic-symptoms-2g3v8aq.jpg>. (F) Source: Bruce Watt, University of Maine, Available at: [Bugwood.org](http://bugwood.org). (G) Source: Bruce Watt, University of Maine, Available at: [Bugwood.org](http://bugwood.org). (H) Source: <http://biosafetykenya.go.ke/images/download/conferencematerials/6thannualpresentations/Proposal-presentaion-GMO-NBA-presentation.pdf>.

(ChilCVD, genus *Begomovirus*, family *Geminiviridae*) persistently transmitted by whiteflies, is the most destructive virus disease in terms of incidence and yield loss, causing a yield loss of 53% or more. The disease can be identified by typical upward leaf curling, crinkling, puckering and reduction in leaf area along with stunting of whole plants. It is transmitted by the whitefly *Bemisia tabaci* in persistent manner. The first report of *Chilli leaf curl disease virus* in India was in 1954, which was later reported in frequent intervals. However, after 2005 the virus complex has been emerging rapidly across the Indian subcontinent.

Cucurbit Viral Diseases

Worldwide, cucurbits are known to be infected by large number of viruses causing significant economic losses. Cucurbit viruses represent a complex and changing pathosystem and cucumber alone is affected by at least 153 viruses. Virus infection in cucurbits generally result in three broad categories of symptoms: (1) Mosaic in leaves, often associated with deformations: leaf size reduction, rugosity, and enations; (2) Yellows, generally affecting the older leaves and later spreading to younger leaves, often accompanied by leaf thickening and reduction in fruit production, but not the fruit quality; (3) Necrosis as necrotic spots on leaves or as generalized necrosis resulting in death of plant, that occur in certain virus/strain and host/cultivar combinations, particularly with potyviruses. Fruits generally don't mature and may develop necrotic symptoms. Often in the field there are mixed infections of two or more viruses, that lead to combination of above three symptom types.

Cucumber mosaic virus (CMV, genus *Cucumovirus*, family *Bromoviridae*), transmitted by aphids, known since beginning of 20th century, is the first mosaic disease reported in cucurbits in the U.S. CMV causes most significant economic losses in melon, cucumber and squash, but is rarely found in watermelon. CMV has a broad host range, that infects more than 775 different plant species representing more than 86 different families. Two major CMV subgroups have been identified based on the serological and molecular properties. A diversity of CMV strains are also reported that differ in symptomatology, host range, virulence towards resistance genes, and aphid transmissibility. Epidemics of CMV were found to be very rapid in Southern Europe and generalized in melons, while in zucchini and squash were often slow to develop symptoms and limited to a few plants.

Epidemics are occasionally reported from temperate regions such as north-eastern U.S. or central Europe. Several distinct species of the *Potyvirus* genus (family *Potyviridae*) infect cultivated cucurbits. Three of them, Papaya ringspot virus (PRSV), Watermelon mosaic virus 2 (WMV2), and Zucchini yellow mosaic virus (ZYMV) have a global spread, and cause significant yield reductions in melon, cucumber, squash, and watermelon. Epidemics of PRSV are common in tropical and subtropical regions where the wild cucurbit species act as reservoirs for the virus. While, WMV2 has been reported more than 50 years ago from across the world.

Zucchini yellow mosaic virus (ZYMV) is one of the first cucurbit 'emerging' viruses that threatened the cucurbit industries across the globe in the early 1980s. In addition to mosaic and deformations on leaves and fruits of all cultivated cucurbits, various "bump" shapes are observed on mature fruits, making them unmarketable. First reported from Italy and France, within a decade's time ZYMV has spread to the major cucurbit producing regions of the world. ZYMV has also colonized new regions or islands. Economic effect of the virus can be dramatic, early infections can result in complete yield loss.

LIYV is another major virus prevalent in cucurbits, lettuce, and sugarbeets. Despite the atypical shift in vectors populations, LIYV caused major economic losses due to its prevalence in lettuce, sugarbeets, and melons. In 1981 alone, the estimated losses were US\$20 million. Squash leaf curl virus (SCLV) and Watermelon chlorotic stunt virus (WmCSV) are transmitted by *B. tabaci* in a persistent manner are two other important viruses that cause significant damage to cucurbits.

Okra Viral Diseases

Okra or bhendi or ladies finger (*Abelmoschus esculentus*) is an economically important vegetable crop grown in tropical and subtropical parts of the world and India is the leading producers of okra in the world. However, cultivation of okra is constrained by very high incidence of Bhendi yellow vein mosaic virus (BYVMV) disease, transmitted through whitefly. BYVMV infects the crop at all growth stages and causes 50%–100% yield loss, as well as reduction in quality, if the plants get infected within 20 days after germination. In India, Bhendi yellow vein mosaic disease (BYVMD) is one of the most economically important diseases of bhendi/okra and is caused by a monopartite begomovirus Bhendi yellow vein mosaic virus (BYVMV; genus *Begomovirus*, family *Geminiviridae*) associated with Bhendi yellow vein betasatellite (BYVβ). The first report of BYVMV was from Mumbai in 1924.

Tomato Viral Diseases

Globally, tomato (*Solanum lycopersicum*) is cultivated in an area of 4.8 mHa, with a total production of 182 mt and with an estimated value of over \$40 billion in 2017. In terms of economic value, tomato shares approx. 70% of the total value of fresh vegetables produced worldwide. At least 136 different viruses are known to infect tomato, whereas this number is significantly lower for other vegetable crops, for example, 49 viruses are reported to infect pepper (*Capsicum annuum*), 46 infect melon (*Cucumis melo*), 53 infect lettuce (*Lactuca sativa*), 54 infect potato (*S. tuberosum*), and 44 infect eggplant (*S. melongena*). Only cucumber (*C. sativus*) is infected by more viruses than tomato (153).

Years of intensive breeding for yield enhancement in tomato, ignoring the resistance traits has reduced the genetic base for virus resistance in cultivated tomato. Further intensified cultivation of tomato, with large areas under tomato mono-cropping might have favored proliferation of tomato viruses and their vectors. *Begomoviruses* and other whitefly-transmitted viruses are invading into new regions, and other emerging viruses such as new tospoviruses and Tomato torrado virus (ToTV; genus *Torradovirus*, family *Secoviridae*) are rapidly spreading over larger geographic areas. In this article, we will restrict to economic impact caused by tomato leaf curl begomoviruses and Tomato spotted wilt virus (TSWV).

The spotted wilt disease of tomatoes was first reported from Australia in 1915 and later demonstrated to be caused by a virus and transmitted by thrips. The causative agent, was named as Tomato spotted wilt virus (TSWV; genus *Tospovirus*, family *Bunyaviridae*). TSWV causes a range of symptoms, such as chlorotic/necrotic rings and flecking on leaves, stems and fruits. Early infection of TSWV can lead to one-sided growth, drooping of leaves reminiscent of vascular wilt, and stunting or death of the tomato plant. Subsequently, these infections produce unmarketable fruits with conspicuous chlorotic/necrotic ringspots. Control of the TSWV vector thrips is almost impossible due to their high fecundity, propensity to develop insecticide resistance and due to their vast host range. TSWV can replicate within the thrips vectors, which is a characteristic feature of many membrane-bound bunyaviruses. The worldwide dispersal of the TSWV vector Western flower thrips (*Frankliniella occidentalis*) resulted in re-emergence of TSWV as a major agricultural pathogen in 1980s. In 1994, the monetary losses by menace of TSWV, across the globe was estimated to be over \$1 billion annually. Till date TSWV is considered to be economically the most important virus due to its global presence and a broad host range with ability to infect more than 800 plant species, including tomato, pepper, lettuce, peanut, and chrysanthemum.

Tomato yellow leaf curl disease (TYLCD) is another major constraint for cultivation of tomato. TYLCD is widely distributed in Asia, sub-Saharan Africa, the Caribbean Islands, Australia, and the USA. More than 57 different geminivirus species generally called as Tomato yellow leaf curl virus (TYLCV, genus *Begomovirus*, family *Geminiviridae*) are reported to infect tomato. In the United States the whitefly-transmitted geminiviruses (WTGs) appeared first in early 1990s, accounting for 20% estimated yield loss. Whereas in the Latin American countries, namely, Brazil, Cuba, Guatemala, Mexico, and Costa Rica, yield losses were in the range

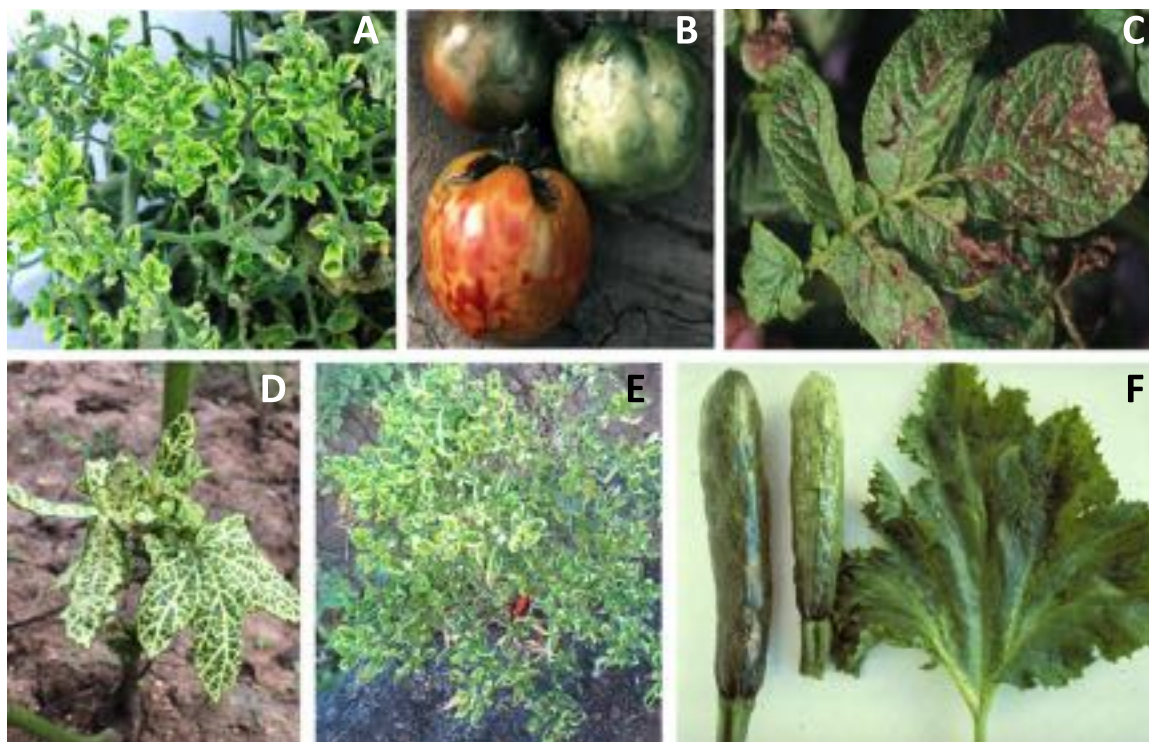


Fig. 3 Characteristic symptoms of important vegetable viral diseases: (A) Tomato yellow leaf curl virus (TYLCV) on tomato, (B) Tomato spotted wilt virus (TSWV) on tomato fruits, (C) Tomato spotted wilt virus (TSWV) on tomato leaf, (D) Okra yellow vein mosaic virus (OYVMV) on okra, (E) Chilli leaf curl disease (ChLCD) on chilli, (F) Zucchini yellow mosaic virus (ZYMV) on *Cucurbita pepo*, fruits (left) and leaf (right). (A) Source: <https://plantpath.ifas.ufl.edu/plant-virus-profiles/images/Picture1.jpg>. (B) Source: Whitney Cranshaw, Colorado State University, Available at: Bugwood.org. (C) Source: Whitney Cranshaw, Colorado State University, Available at: Bugwood.org. (F) Source: http://www.pestnet.org/fact_sheets/zucchini_yellow_mosaic_202.htm.

of 30%–100%. The estimated losses during 1989–1995 in Dominican Republic were to the level of \$50 million. Another study has revealed that the Tomato leaf curl viruses (ToLCVs) cause an annual yield loss of \$140 million in Florida, USA. Although it is possible to chemically control the whiteflies, but this may not be possible when there are large population sizes of whiteflies. However, under protected cultivation, physical barriers such as greenhouses enveloped by net screens are effective, while ‘floating barriers’ of perforated polyethylene sheets stretched over tomato fields prevent the landing of whiteflies. Such measures, increase the costs of cultivation and it may be difficult for small and marginal farmers to take up such protected cultivation. Hence, cultivation of virus resistant/tolerant tomato varieties is economically more feasible and breeding programs for introgression of virus resistant traits are in progress.

Tomato leaf curl New Delhi virus (ToLCNDV) (genus *Begomovirus*, family *Geminiviridae*) is a major constraint in tomato production and is economically most important disease affecting tomato in Indian subcontinent. The first report of association of ToLCNDV with tomato leaf curl disease on solanaceous crops was from India in 1995. However, in the last few years ToLCNDV has significantly widened its host range and has moved to new geographical areas, including the Middle East, the western Mediterranean Basin, and some of the European countries. Recombination and pseudo-recombination are the major factors for diversification and adaptive evolution of ToLCNDV. Emergence and epidemics of ToLCNDV has resulted in increased yield losses in several important crops, including tomato. Recent reports from Tunisia, Italy, and Spain indicate infection of ToLCNDV isolates in several cucurbit species. Sequence analysis revealed that a novel strain of ToLCNDV (ToLCNDV-ES) is spreading in the western Mediterranean region, which probably evolved through recombination from a highly diverse population of ToLCNDV. Several reports have demonstrated the involvement of monopartite begomoviruses in the major disease complexes affecting various crops in the “Old World”. However, ToLCNDV a bipartite begomovirus is the only exception to this, which causes diseases in a number of plant species of the families Solanaceae, Cucurbitaceae, Euphorbiaceae, Fabaceae, and Malvaceae. Furthermore in addition to vegetable crops, ToLCNDV has been reported to cause damage to fiber crop species and was also found to infect a number of weeds. Currently, ToLCNDV is known to infect 43 different plant species in Bangladesh, India, Indonesia, Iran, Italy, Malaysia, Pakistan, Sri Lanka, Spain, Taiwan, Thailand, and Tunisia. Co-existence of satellite DNAs (betasatellites) along with ToLCNDV is also reported from natural infections and are known to modulate its pathogenesis, including host range (Fig. 3).

Fruit Virus Diseases

Banana Viral Diseases

Banana is an important source of carbohydrate and is cultivated in more than 130 countries. Banana is also a major source of income for millions of farmers from tropical regions of the world. In 2017, total global area under banana cultivation was 5.6 mHa, with a total production of 114 mt. The Asian continent has the largest area under banana cultivation with more than 50% share in total world area. In 2011, the total export value of banana was estimated to be \$895 million. Banana is propagated vegetatively through the use of suckers and this practice has been one of the main cause for outbreaks of several diseases and pests around the world, especially the viruses. At least 20 different viral species belonging to five different families are known to infect banana globally. The economically most important viruses are: Banana streak viruses (BSVs; genus *Badnavirus*, family *Caulimoviridae*), Banana bunchy top virus (BBTV; genus *Babuvirus*, family *Nanoviridae*), Banana bract mosaic virus (BBRMV; genus *Potyvirus*, family *Potyviridae*), and Cucumber mosaic virus (CMV, genus *Cucumovirus*, family *Bromoviridae*).

Worldwide, Banana bunchy top disease (BBTD) is the most devastating disease caused by BBTV and transmitted by the aphid vector *Pentalonia nigronervosa*. This disease was first reported in 1889 from Fiji and an outbreak of BBTD occurred in 1953 resulting in loss of 641,000 bunches. Later, BBTD has spread to more than 36 countries in South Pacific, Asia, and Africa. However, there are no records of BBTD in the New World except in Hawaii (USA). In the 1920s, due to the bunchy top disease epidemics, the Australian banana industry in New South Wales had collapsed, and the total area under banana cultivation was reduced by 90%. Similarly, in Southern Queensland banana production was reduced by more than 95%. In India, BBTD is known to cause heavy losses. One of the most significant BBTD outbreaks was in highly prized hill banana cv. Virupakshi, in the state of Tamil Nadu (India), which reduced the cultivation area from 18,000 to 2,000 Ha. Several epidemics of BBTD during 2007–2010 in Koduru, Andhra Pradesh, and Jalgaon, Maharashtra, India, led to an annual loss of \$50 million. An outbreak of BBTD in Pakistan led to huge reduction in area under banana cultivation from 60,000 to 26,000 Ha. None of the farmer's cultivated banana varieties are resistant to BBTV and production of disease-resistant bananas by conventional breeding is very difficult. Hence, transgenic development of BBTV resistant bananas is the most viable option to achieve resistance to BBTD.

Banana streak disease caused by Banana streak virus (BSV) is also one of the important viral disease of banana which has a wider presence in banana plantations throughout the world. The streak disease was first observed in the Nieky Valley on the Ivory Coast in 1958 and later it was reported from Morocco where almost every established plantation was affected. Banana streak disease is now reported to occur in over 43 countries of Africa, Asia, Australia, Europe, Oceania, and tropical America.

Another viral disease called Banana bract mosaic disease, caused by the Banana bract mosaic virus (BBRMV), was first reported in most of the banana cultivars, from Philippines (island of Mindanao) in 1979 and thought to be different from all other recognized viruses of banana. Later, it was found widespread in entire Philippines. Subsequently the banana bract disease was discovered in other Asian countries including India, Sri Lanka, Samoa, Thailand, and Vietnam. In Latin America, BBRMV was first reported from Colombia. However, there are limited reports available on the economic impact of the disease. In 1997, up to 40% yield losses were reported from Mindanao island of the Philippines. Yield losses in the range of 30%–70% are also reported from India and the Philippines.

Cacao Swollen Shoot Disease

Cacao swollen shoot disease (CSSD) caused by Cacao swollen shoot virus (CSSV; genus *Badnavirus*, family *Caulimoviridae*), was first reported from Ghana in 1936 and is endemic to West Africa covering Togo, Ivory Coast and Nigeria. CSSD is semi-persistently transmitted by mealy bug (*Planococcoides njalensis*) vectors and this disease reduces cacao yield in first year of infection by almost 25%, in second year the yields are reduced by 50%, and in next 3–4 years the tree dies from die back. CSSD economically most important disease, since cacao has high industrial value and this disease has resulted in drastic reduction in cocoa production in Ghana. Since its first report in 1936, more than 200 million cacao trees have been eradicated from about 130,000 Ha. Yield of cacao bean were reduced by 86% during 1945–1950. In 1953, the 50 million diseased trees represented a capital depreciation of £25 million since the report of CSSD. During 2006–2010, over 28 million infected trees were removed and globally today CSSV is responsible for 15% of total cacao crop loss. This disease caused much rural disconnect and political uproar in Ghana. In response to the menace caused by CSSD, Ghana has launched one of the most ambitious eradication program with high investments. In the late 1940s, the British colonial authorities in Ghana enforced an eradication policy for removal of infected cacao trees. This met with a serious opposition, resulted in the rise of a political party led by Kwame Nkruma, who accused the authorities of attempting to destroy farmers' income.

Some of the strategies employed to control this disease are: use of insecticides, biological control of the mealybugs, breeding for resistant cocoa cultivars, and cross-protection with mild strains of CSSV, but the main practice is to rogue the infected and adjacent trees. CSSV has several alternate hosts, such as *Cola gigantean*, and mere removal of infected trees would not solve the CSSD menace and it is important to remove the alternate hosts too.

Citrus Tristeza Virus

Citrus tristeza virus (CTV) (genus *Closterovirus*, family *Closteroviridae*) is the causal agent of devastating epidemics that changed the course of the citrus industry. There are mainly three distinct syndromes caused by CTV, named as tristeza, stem pitting (SP), and

seedling yellows (SY), which is due to differential interaction of CTV with different scions and rootstocks from different regions. Tristeza disease is a decline syndrome caused by infection of CTV on different citrus species such as sweet oranges, mandarins, grapefruits and limes, which are propagated on rootstocks derived from sour orange or lemon.

During 1940–1960, about 30 million citrus trees grafted on sour orange rootstocks were lost in Argentina and Brazil. In 1980s, about 6.6 million trees were lost in Venezuela, and ~10 million trees were lost in Florida and other Caribbean countries. In these regions CTV is mainly transmitted by the brown citrus aphid, *Toxoptera citricida*. In Spain alone CTV took a toll of 10 million citrus trees and several million more were lost in California (USA), Israel, and other areas. In these areas CTV is being spread by the melon aphid, *Aphis gossypii*. In the presence of both the aphid vectors, the spread of CTV is greatly enhanced, as in the case of Florida (USA). Significant yield losses have been documented from several regions and the monetary losses have been estimated. However, these estimates do not include the expenses incurred on removal of dead trees and replanting with healthy plants. Despite the loss of millions of citrus trees in Brazil and Spain, in the long run the citrus production was not affected. In Brazil, in a span of about 40 years, the citrus production increased from 1.7 to 20 mt in 1990s. This would not have been possible without replanting new orchards on tolerant rootstocks, along with budwood control.

Some of the strategies employed for control of CTV damage are quarantine and budwood certification programs, rouging of infected trees, use of tristeza-tolerant rootstocks, or cross protection with mild isolates. The choice of the management practice mostly depends on CTV incidence, the virus strains present and host varieties cultivated in an area. Introgression of virus resistance genes into cultivated genes through conventional breeding is not feasible, while genetic engineering for virus resistance has given variable results. Long distance spread of CTV occurs through movement of infected planting material and then short distance local transport is facilitated by different aphid species and is not known to be seed transmitted. Infection of CTV is mostly restricted to species of two genera of Rutaceae family and the virus is then localized to the phloem cells. The most devastating epidemics of CTV occurred in Argentina (1930), Brazil (1937), California (1939), Florida (1951), Spain (1957), Israel (1970), and Venezuela (1980). Outbreaks of tristeza have also been reported from Cyprus (1989), Cuba (1992), Mexico (1995), the Dominican Republic (1996), and Italy (2002).

In addition to loss of trees, CTV is also indirectly associated with loss of sour orange root stocks with very good agronomic traits. Further, the use of tristeza-tolerant rootstocks has led to the emergence of new problems such as damage from soil salinity or alkalinity, waterlogging, or vulnerability to soil fungi or graft-transmissible pathogens such as citrus blight, Citrus sudden death-associated virus or Citrus tatter leaf virus. Regardless of the rootstock used, the affected trees showed stunting, SP, low yield and poor quality of fruits. Isolates of CTV causing SP, that were initially restricted to regions of Asia, Australia, South Africa, and South America, were subsequently reported from California, Florida, and the Mediterranean region, although at lower frequency.

Implementation of mandatory certification program for maintenance and distribution of virus-free material for commercial use is the best way to control CTV or any other graft transmissible disease. One such publicly operated self-supportive program called 'clean stock program' in Florida charge a fee of \$2 for registration of validated or parent trees. Cross-protection is widely practiced in Brazil, Australia, and South Africa. In particular, this has been highly successful in Brazil for CTV strains causing SP on sweet orange and grapefruit. By 1980, over 8 million trees of Pera sweet orange were cross-protected in Brazil and this is still practiced. In California, an eradication program for tristeza is in place and during 2000–2001 the Central California Tristeza Eradication Agency surveyed approximately 31,400 acres of citrus orchards and carried ELISA tests for about 375,000 individual plants. Undoubtedly this program deferred the spread of tristeza, but huge costs were incurred to remove the sick trees. In Israel, during 1970–1977, a tristeza suppression program was in place, wherein about 300,000 tests were made using indicator plants and electron microscopy and eventually 1700 trees diagnosed with CTV infection were removed. Later in 1979–1980, indexing of 1.25 million of citrus trees by ELISA revealed CTV infection in only 0.13% of the tested trees. Hence such eradication programs delayed the outbreak of a citrus tristeza pandemic in Israel by 15–20 years.

Eighty-five years of epidemics of CTV have made a devastating economic impact on the citrus farms and the industries dependent on it. Worldwide, CTV has taken a toll of more than 80 million citrus trees, mostly in South Africa since 1910, Argentina (10 million) and Brazil (6 million) since 1970, and the US. (3 million) since 1950. Dispersal of CTV is a continuous process and it is feared that the future losses from tristeza may be much more than the losses that have occurred till date. There is also a rapid dispersal of the CTV vector *T. citricida* in Mexico and the Mediterranean basin, with large area under citrus cultivation using the sour orange rootstock, could facilitate the spread of tristeza. The replacement of declining citrus trees with new trees on tristeza tolerant rootstocks can only restore the production of citrus. In the last 40 years, new isolates of SP are being periodically reported in new regions and is going to be a more enduring problem. Cross-border movement of new citrus varieties has resulted in invisible spread of severe SP isolates and is likely to continue despite the quarantine measures.

Papaya Ring Spot Virus

Worldwide papaya is cultivated in an area of 0.44 mHa with a total production of 13.2 mt and 45% of the total production is contributed by India alone. Several pests and pathogens are known to affect papaya (*Carica papaya*) plants, of which viral diseases are the most damaging ones. Of the several viral diseases Papaya ringspot virus (PRSV, genus *Potyvirus*, family *Potyviridae*) inflicts maximum damage to papaya and has a worldwide presence. Other viruses that are known to infect papaya are geminiviruses causing leaf curl symptoms, Papaya meleira virus (PMeV; Unassigned in the family *Totiviridae*), Papaya mosaic virus (PapMV;

genus *Potexvirus*, family *Alphaflexiviridae*), and Papaya lethal yellowing virus (PLYV; genus *Sobemovirus*, family *Solemoviridae*). PRSV is transmitted by several different aphid species in a non-persistent manner. There are no known resistance genes against PRSV in papaya, but tolerance to PRSV is conferred in a quantitative manner. In 1950s, Oahu island in Hawaii (USA) was the main papaya-growing area, however because of the heavy PRSV infection, papaya cultivation was shifted to fertile lands of Puna district (Hawaii, USA) where PRSV was absent. By 1970s, 95% of Hawaiian papaya were cultivated in Puna, but a potential threat of PRSV outbreak remained, since Hilo was just 30 km away.

In 1992, for the first time PRSV was reported from commercial plantings of papaya in Puna district on Hawaii island. Despite all the efforts to minimize the spread of PRSV, more than 50% of Puna was severely affected by PRSV by 1994. Subsequently the Hawaii Department of Agriculture (HDOA) abandoned all the eradication programs. By 1998, most of the papaya plants in Puna were PRSV infected and papaya production was reduced to about half of the 1992 level (12,000 vs 24,000 t of fruits). In 1992, Puna produced 95% (24,000 t) of Hawaii's papaya; but after the entry of PRSV in that region, the production decreased significantly and by 1998 Puna produced only 75% (12,000 t) of Hawaii's production. Papaya production in Puna was reduced to half in 1998, compared to 1992. Clearly, PRSV had a devastating effect on papaya production in 1998. To maintain its markets, especially in mainland of USA and Japan, Hawaii had to keep up production as much as possible. In addition to significant yield losses, PRSV infection severely affected the quality of the produce and the papaya fruits harvested were of inferior quality.

Transgenic papaya overexpressing the coat protein sequence of PRSV were developed by researchers from University of Hawaii (USA) and released as papaya varieties "SunUp" and "Rainbow". These varieties are now widely cultivated in the Hawaii islands of United States and is the most successful field application of transgenic technology. Since the release of PRSV resistant transgenic papaya, the papaya production in Puna showed an increasing trend, starting from 15,400 t in year 2000 to 18,300 t in 2001. Furthermore, papaya production in Puna accounted for 84% of the production in 2002, compared to the low of 65% in 1999. In 2004, Rainbow accounted for 88% of the total area of papaya in Puna (Hawaii, USA).

Plum Pox Virus (Sharka) Disease of Stone Fruits

Plum pox disease, also called as "sharka" is one of the most devastating viral diseases of stone fruits resulting in huge yield losses and a major economic impact. This disease is caused by Plum pox virus (PPV), a positive ssRNA virus and belonging to the genus *Potyvirus* and family *Potyviridae*. In addition to plums (*Prunus domestica*), PPV also infects other stone fruits such as, peaches (*Prunus persica*), apricots (*Prunus armeniaca*), nectarines (*Prunus persica* var. *nucipersica*), almonds (*Prunus dulcis*), sweet cherries (*Prunus avium*) and tart cherries (*Prunus cerasus*). PPV also infects important ornamental and wild *Prunus* species, including those used in traditional medicine: myrobalan, American plum, dwarf flowering almond, and blackthorn. Further, approx. 60 plant species representing eight families are recognized as experimental host plants of PPV. Nevertheless, PPV also infects plants that are not members of genus *Prunus*.

The first report of sharka disease was from plum orchards in Bulgaria during 1915–1918. During 1932–1960, the disease disseminated to north and east from Bulgaria into Yugoslavia, Hungary, Romania, Albania, Czechoslovakia, Germany, and Russia. Initially, sharka disease was mainly found in plums and apricots and since the 1980s has been reported in peaches. About a decade ago, about 100 million stone fruit trees in Europe were reported to be infected. The susceptible varieties could result in 80%–100% yield losses. In Eastern and Central Europe, susceptible plum cultivars exhibit premature fruit drop and bark splitting, and sweet cherry fruits develop chlorotic and necrotic rings and fruit drop occurs prematurely. Although the fruit trees don't die from PPV infection, but the fruits drop prematurely and are unfit for consumption and marketing due to their bitter taste. In 1968, the estimated yield losses due to sharka disease in Bulgaria was 30,000 t. In last several decades, the fruit drop production was to an extent of 80% in Czech Republic, and there was a drastic reduction in number of plum trees from 18 million to 4 million. During 1998–2002, due to PPV infection, about 69,000 stone fruit trees were removed in Emilia–Romagna region of Italy; and about \$450,000 is spent annually for the removal and replanting of trees. An expenditure of \$17 million was incurred for removal of ~1.5 million trees in Spain during 1988–2008. PPV is predominantly present in south east of France and all the symptomatic trees are eliminated and if infection levels are more than 10%, the entire orchard is destroyed. Such operations have resulted in removal of about 27,000 trees in 1992 and 100 ha of plantations (mainly peach) were eliminated in 1993. During 1973–1990, about 91,000 trees were estimated to be destroyed in France. PPV is transmitted by aphids in a non-persistent manner, however the spread in Europe is mainly due to movement of infected nursery material from the Balkans. Vegetative propagation of host plants is the main cause for efficient dissemination of PPV both locally and globally. General susceptibility of hosts, with scarce resistance sources identified, has largely frustrated international resistance breeding efforts.

For most of the 20th century PPV was limited to Europe, however in the past 20 years PPV infection has been reported from Africa, South & North America, and Asia. Currently almost all the major plum production areas are now affected by PPV at varying degrees. Huge efforts and costs are involved in the control and eradication of PPV in several countries. Though successful in a few cases, such efforts have generally slowed the progression of PPV onslaught, but certainly has not stopped it. In 2006, the combined costs for control of Sharka disease were estimated to be \$10 billion over the past 30 years worldwide. Large number of efforts are made for genetically engineering PPV resistance in plums, resulting in successful field trial of the PPV resistant HoneySweet transgenic plum. Dereglulation and cultivation of HoneySweet transgenic plum in USA, is one of the few example of virus-resistant transgenic crops developed to marketability (Fig. 4).

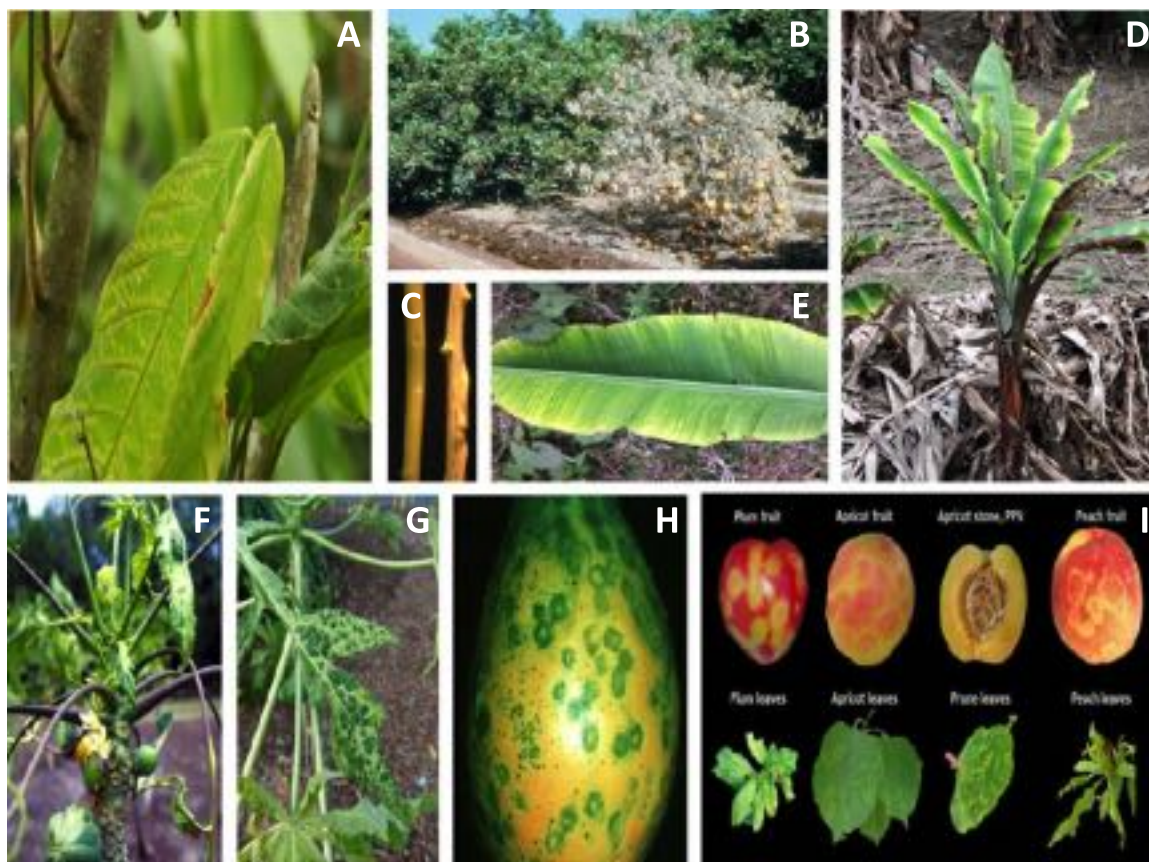


Fig. 4 Characteristic symptoms of important fruit viral diseases: (A) Cocoa swollen shoot virus (CSSV) on young cocoa tree branches and leaf, (B) Citrus tristeza virus (CTV) on citrus tree, (C) Stem pitting symptom induced by Citrus tristeza virus (CTV) on a citrus branch, (D) Banana bunchy top virus (BBTV) on banana, (E) Banana streak virus (BSV), on banana leaf, (F) Papaya ring spot virus (PRSV), symptoms on a tree, (G) Papaya ring spot virus (PRSV), symptoms on a fruit, (H) Papaya ring spot virus (PRSV), symptoms on a fruit, (I) Plum pox virus (PPV) on several fruits and leaves of plum, apricot and peach. (A) *Source*: wikimedia. (B) *Source*: L. Navarro, Instituto Valenciano de Investigaciones Agrarias, Available at: Bugwood.org. (D) *Source*: Rui map Zheng, Available at: Bugwood.org. (H) *Source*: <https://caes.ucdavis.edu/news/articles/2014/01/science-provides-facts-for-hawaii-gmo-debate>. (I) *Source*: European and Mediterranean Plant Protection Organization, Available at: Bugwood.org.

Legume Viral Diseases

A wide range of legumes (Fabaceae: Papilionoideae) are cultivated across the globe as an important source of proteins. These legume crops are severely affected by a range of viral diseases that result in reduced grain yield, poor quality seeds and increased cost of cultivation for phytosanitary activities and disease control, thus making significant impact on economic losses. The majority of the legume infecting viruses are transmitted by insect vectors, and many of them are also seed transmitted. Viral diseases are the major biotic constraints to legumes production, especially in the tropical and subtropical regions. Cultivated food legumes are vulnerable to infection by at least 150 viruses, belonging to different plant viral genera/families. Begomoviral diseases alone in legumes is known to cause an annual estimated yield losses of \$300 million in India. The economic impact of viral diseases in most important legumes/pulse crops, such as soybean, groundnut, cowpea, pigeonpea, and mungbean are listed in **Table 1**.

Worldwide, about 70 different viral species are reported to infect soybean, of which, Soybean mosaic virus (SbMV; genus *Potyvirus*, family *Potyviridae*), Tobacco ring spot virus (TRSV; genus *Nepovirus*, family *Secoviridae*), Peanut bud necrosis virus (PBNV; genus *Tospovirus*, family *Bunyaviridae*), Tobacco streak virus (TSV; genus *Ilarvirus*, family *Bromoviridae*), Soybean dwarf virus (SbDV; genus *Luteovirus*, family *Luteoviridae*), and some begomovirus species are considered to be economically important. SbMV is reported to cause a yield loss in the range of 8%–50%, and can reach up to 100% in severe outbreaks and when coinfecting with Bean pod mottle virus (BPMV; genus *Comovirus*, family *Secoviridae*), Cowpea mosaic virus (CPMV; genus *Comovirus*, family *Secoviridae*), Abutilon mosaic virus (AMV; *Begomovirus*), and TRSV. Several different begomovirus species are associated with the soybean yellow mosaic disease (SYMD). In India alone, the combined yield losses due to SYMD were estimated to exceed \$300 m in mungbean, black gram, and soybean.

At least 32 viruses are reported to naturally infect groundnut, of which TSWV, PBNV, TSV, Peanut clump virus (PCV; genus *Pecluvirus*, family *Virgaviridae*), Indian peanut clump virus (IPCV; genus *Pecluvirus*, family *Virgaviridae*), PeMoV, PSStV, CMV, and Groundnut rosette disease (GRD) caused by a complex of 3 viral agents, Groundnut rosette assistor virus (GRAV), Groundnut rosette virus (GRV, genus *Umbravirus*, family *Tombusviridae*), and a satellite RNA, are considered to be economically important. GRD is the most damaging disease in sub-Saharan Africa (SSA) and many GRD epidemics were reported from Africa, which has caused enormous crop losses.

Spotted wilt disease of groundnut caused by TSWV, can lead to 100% yield loss and this disease is reported from North and South Americas, several African countries, and Australia. In 1998, a similar disease has been reported from Asia, but was shown to be caused by PBNV, a distinct tospovirus. Groundnut bud necrosis disease (GBND) first reported from India, is also prevalent in Sri Lanka, Nepal, Myanmar, Thailand, and parts of China. GBND is caused by a distinct tospovirus, PBNV (synonym Groundnut bud necrosis virus, GBNV). PBNV also infects mungbean, urdbean, soybean, cowpea, pea, and lablab bean under field conditions.

Stem necrosis disease (SND), previously assumed to be caused by PBNV, was later found to be caused by TSV in India. TSV is known to occur on groundnut in India, Pakistan, South Africa, and Brazil and has a broad host range that includes sunflower (*Helianthus annuus*), soybean, mungbean, black gram, cowpea, and sunnhemp (*Crotalaria juncea*). TSV continues to expand its host range. In year 2000, SND was prevalent in nearly 225,000 Ha of erstwhile Andhra Pradesh state of India, which resulted in crop losses exceeding \$65 million. Peanut clump disease (PCD) is prevalent in India, Pakistan, and West Africa, and the causal virus is referred as IPCV in Indian subcontinent and PCV in Africa, both of which cause indistinguishable symptoms. In 2003, worldwide the economic losses due to PCD was estimated to exceed \$38 million. Both IPCV and PCV, infect pigeonpea and other economically important cereal crops. Since the casual viruses are seed transmissible in groundnut, cereals, and millets, PCD has quarantine importance.

Cowpea is most widely cultivated legume in SSA, and is the second most important food legume after groundnut and more than 80% of cowpea production is from West Africa. More than 140 viruses are reported to infect cowpea and of these 20 viruses are known to be economically important. Except CGMV (Cowpea golden mosaic virus) and CCMV (Cowpea chlorotic mottle virus), all the other cowpea viruses are known to be seed transmitted. Pigeonpea is another important grain legume crops, predominately cultivated in the Indian subcontinent, and also in Southern and Eastern Africa, the Caribbean, and China. About 15 viruses are reported to infect pigeonpea, of these, Pigeonpea sterility mosaic emaravirus (PPSMV, genus *Emaravirus*, family *Fimoviridae*) is most important, followed by the whitefly-transmitted begomoviruses. Sterility mosaic disease (SMD) caused by PPSMV is a serious constraint for pigeonpea cultivation in South Asia, with an estimated annual loss of over \$300 million in India alone.

Of several viruses infecting mungbean and urdbean, yellow mosaic caused by begomoviruses, leaf curl caused by PBNV, and leaf crinkle caused by Urdbean leaf crinkle virus (ULCV) are economically most important. In India, both Mungbean yellow mosaic Indian virus (MYMIV) and ULCV are prevalent in both mungbean and urdbean, and are often synergistic leading to yield losses over 90%.

Viruses that are transmitted through legume seeds serve as primary source of virus in virus ecology and disease epidemiology in cultivated legumes and have quarantine importance. Further the multiplication and spread of insect vectors is facilitated by intensive cropping and changes to cropping patterns as a result of increased irrigation and overuse of pesticides. Occurrence and the drastic expansion of the host range of tospoviruses and TSV in particular are major concerns. Some of the strategies employed for management of legume viruses are: selection and planting of virus-free seeds, modification of agronomic practices, chemical, physical, and biological control of viral vectors, and planting of virus-resistant crop varieties developed through conventional or non-conventional crop improvement strategies (Fig. 5).

Cash Crops Viral Diseases

Sugarcane Viral Diseases

Sugarcane is the largest crop production in the world with more than 1800 mt in 2017 on a surface of nearly 26 mt in the world. Worldwide, sugarcane is affected by diverse number of viruses, mainly because of its vegetative propagation. Sugarcane mosaic disease caused by Sugarcane mosaic virus (SCMV; genus *Potyvirus*, family *Potyviridae*) is economically the most important of all and is seen in almost all the major sugarcane-growing countries. The first report of sugarcane mosaic was in 1892, and was subsequently reported from other countries growing sugarcane, maize, and sorghum, namely, Bangladesh, Indonesia, Thailand, Malaysia, China, the United States, Pakistan, Argentina, Australia, Brazil, Cameroon, Colombia, Congo, Cuba, Denmark, Ethiopia, Egypt, France, Germany, Iran, Israel, Mexico, the Netherlands, and Vietnam. The first report from India was in 1921, since then, it invariably infected all the sugarcane cultivars. During 1914–1916, a severe epidemic of sugarcane mosaic occurred in Argentina, resulting in up to 80% losses in sugar production. Later, this crisis was overcome by replacing susceptible varieties of sugarcane with the resistant ones (POJ varieties) from Indonesia (Java). In mid-1920s, epidemic of sugarcane mosaic threatened the sugar industry in both Brazil and the United States, resulting in severe economic crisis in the sugar industry. Subsequently, the disease was controlled by replacing the susceptible varieties with mosaic resistant hybrids. However, later in 1971–1972, it was revealed in Brazil that the mosaic was found in tolerant hybrids with 100% infection, causing a loss of 18% in susceptible clones. Despite the report of sugarcane mosaic in India, it was ignored initially, however later, it was reported to have severe impact on popular sugarcane varieties in the field. Studies showed that the virus infection significantly reduced the number of millable canes, net CO₂

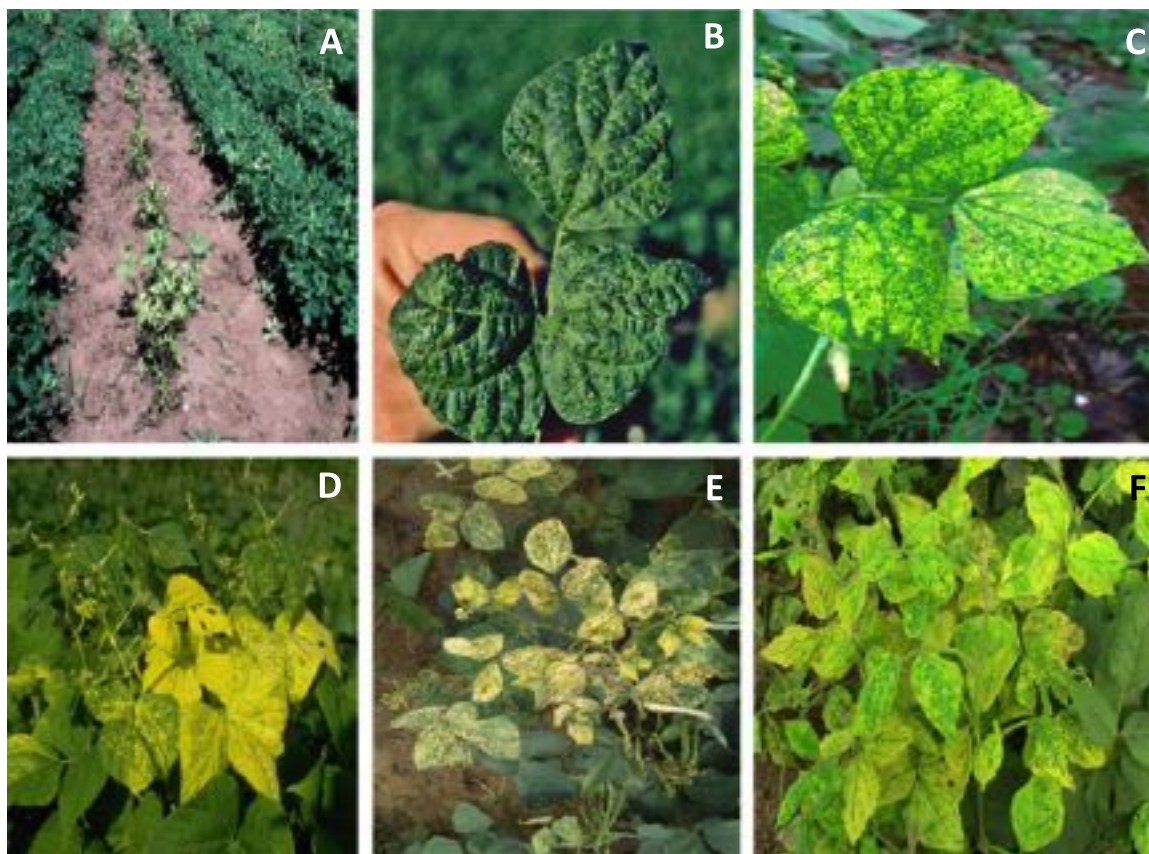


Fig. 5 Characteristic symptoms of important legume viral diseases: (A) Groundnut rosette disease (GRD) on groundnut (center) and healthy plants (left and right), (B) Bean common mosaic virus (BCMV) on bean leaf, (C) Bean yellow mosaic virus (BYMV) on bean leaf, (D) Bean golden mosaic virus (BGMV) on bean, (E) Mungbean yellow mosaic disease (MYMD) on mungbean, (F) Soybean mosaic virus (SbMV) on soybean. (A) Source: <http://www.dpvweb.net/dpv/showfig.php?dpvno=345&figno=03>. (B) Source: <https://www.invasive.org/browse/detail.cfm?imgnum=5364007>. (C) Source: Dr. Parthasarathy Seethapathy, Tamil Nadu Agricultural University, Available at: Bugwood.org.

assimilation rate, cane growth parameters and other post-harvest juice quality parameters. Extensive surveys conducted at different sugar factories in Indonesia, revealed that about 30% of sugarcane fields were affected by the mosaic disease.

Sugarcane bacilliform virus (SCBV; genus *Badnavirus*, family *Caulimoviridae*) causing leaf fleck disease in sugarcane was first reported from Morocco and now it shows its presence in all the major sugarcane-growing countries. Studies revealed that the *S. officinarum* clones recorded more than 90% SCBV incidence, and other *Saccharum* species also showed the presence of SCBV. SCBV infection is presumed to have originated from Papua New Guinea (Southeast Asia), the center of sugarcane origin, and from there, it might have spread to other areas through the germplasm movement. SCBV is distributed throughout the world in both tropical and subtropical regions and can result in economic losses in the range of 10%–90% in severely infected crops. *Sugarcane yellow leaf virus* (SCYLV; family *Luteoviridae*) causing yellow leaf disease (YLD) is also one of the major threat to sugarcane cultivation worldwide and the commercial sugarcane cultivars of major sugarcane-growing countries are reported with 100% yield incidence in Florida, Brazil, India, etc.

Ornamentals and Orchids Viral Diseases

In 2004, the global floricultural produce was estimated to be \$75 billion, for which Netherlands, USA and Japan were the major contributors. Virus infections in ornamental plants have enhanced the esthetic value in certain cases as well as have resulted in significant monetary losses in several other cases. One of the first uses of esthetic effects of plant virus infection was the production and marketing of highly valued variegated tulips, infected with Tulip breaking virus. Other popular examples are the attractive mosaic patterns on the leaves of Abutilon infected with Abutilon mosaic virus, the variegated flowers of camellia infected with Camellia yellow mottle virus, and the yellow-veined Japanese honeysuckle infected with Honeysuckle yellow vein mosaic virus. In contrast, there are also hundreds of viruses that are known to infect the ornamental plants and cause severe economic losses. Viral diseases reduce the esthetic value and marketability of ornamental plants. Large scale vegetative propagation through virus infected material such as cuttings, bulbs, and corms has significantly contributed for spread of viral diseases in ornamentals. Tospoviruses transmitted by the thrips are known to cause maximum damage to the ornamental plants and their incidence is increasing year by



Fig. 6 Characteristic symptoms of important cash crops viral diseases: (A) Cotton leaf curl disease (CLCuD) on cotton, (B) Sugarcane mosaic virus (SCMV) on sugarcane leaf, (C) Tobacco mosaic virus (TMV) on tobacco plant, (D) Cymbidium mosaic virus (CymMV) on orchid plant, (E) Dendrobium mosaic virus (DMV) on orchid leaf, (F) Odontoplossum ring spot virus (ORSV) on orchid leaf. (A) Courtesy of Suresh Kunkalikar. (C) Source: <https://burleytobaccoextension.ca.uky.edu-content-tobacco-mosaic-virus-tmv>. (D) Source: <http://www.orchidboard.com/community/pests-and-diseases/92602-cymbidium-mosaic-virus.html>. (E) Source: flicker. (F) Source: <https://www.flickr.com/photos/scotnelson/33983505945>.

year mainly because of international trade and movement of propagative material. TSWV and INSV (Impatiens necrotic spot virus) are the two most important tospoviruses infecting the ornamental plants. In India, Chrysanthemum diseases caused by Chrysanthemum virus B (CVB; genus *Carlavirus*, family *Betaflexiviridae*), Cucumber mosaic virus, Tomato aspermy virus and Gladiolus diseases caused by Cucumber mosaic virus (all three viruses; genus *Cucumovirus*, family *Bromoviridae*) are of economic importance.

More than 30 viruses are reported to infect different orchids in various geographical locations. Orchid viruses generally produce mosaics, ringspots, chlorosis, chlorotic and necrotic sunken patches on the leaves, deformation and a drastic reduction in the size of flowers and overall stunting of the plant. The orchid plants, *Phalaenopsis* and *Dendrobium* in particular, are commercially valuable ornamental plants with a global market. Unfortunately, orchid plants are highly susceptible to viruses such as Cymbidium mosaic virus (CymMV; genus *Potexvirus*, family *Alphaflexiviridae*) and Odontoglossum ringspot virus (ORSV; genus *Tobamovirus*, family *Virgaviridae*), posing a major threat and serious economic loss to the orchid industry worldwide. Among the orchid viruses, CymMV and ORSV are considered most important due to their worldwide occurrence and severity of the symptoms in several orchid genera. They reduce the general vigor of the plant and affect the flower quality, thereby reducing the marketability and incurring severe economic losses. These two viruses also occur naturally as mixed infections in several orchid species (Fig. 6).

Further Reading

- Gonsalves, D., 1998. Control of papaya ringspot virus in papaya: A case study. *Annual Review of Phytopathology* 36, 415–437.
- Legg, J.P., Kumar, P.L., Makesh Kumar, T., et al., 2015. Cassava virus diseases: Biology, epidemiology, and management. *Advances in Virus Research* 91, 85–142.
- Loebenstein, G., Thottappilly, G., 2003. *Virus and Virus-Like Diseases of Major Crops in Developing Countries*. Amsterdam: Elsevier/Kluwer Academic Publishers.
- Loebenstein, G., Katis, N., 2014. Control of plant virus diseases: Seed-propagated crops. *Advances in Virus Research* 90, 530.
- Loebenstein, G., Katis, N., 2015. Control of plant virus diseases: Vegetatively-propagated crops. In: *Advances in Virus Research*, first ed., 91. Elsevier, p. 332.
- Orke, E.C., Dehne, H.W., Schonbeck, F., Weber, A., 1994. *Crop Production and Crop Protection: Estimated Losses in Major Food and Cash Crops*. Elsevier.
- Patil, B.L., Legg, J.P., Kanju, E., Fauquet, C.M., 2015. Cassava brown streak disease: A threat to food security in Africa. *Journal of General Virology* 96, 956–968.

- Patil, B.L., 2018. Genes, Genetics and Transgenics for Virus Resistance in Plants. Norfolk: Caister Academic Press, doi:10.21775/9781910190814.
- Scholthof, K.B., Adkins, S., Czosnek, H., *et al.*, 2011. Top 10 plant viruses in molecular plant pathology. *Molecular Plant Pathology* 12 (9), 938–954.
- van der Zaag, D.E., 1987. Yield reduction in relation to virus infection. In: de Bolas, J.A., van der Want, J.P.H. (Eds.), *Viruses of Potatoes and Seed-Potato Production*. Wageningen: Pudoc, pp. 146–150.
- Waterworth, H.E., Hadidi, A., 1998. Economic losses due to plant diseases. In: Hadidi, A., Khetarpal, R.H., Koganezawa, H. (Eds.), *Plant Virus Disease Control*. St. Paul: American Phytopathological Society Press, pp. 1–13.

Relevant Websites

<http://faostat.fao.org>
FAOSTAT.

Retrotransposons of Plants

M-A Grandbastien, INRAE – French National Research Institute for Agriculture, Food and Environment, Versailles, France

© 2008 Elsevier Ltd. All rights reserved.

This is a reproduction of M.-A. Grandbastien, Retrotransposons of Plants, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00544-6.

Glossary

Env Envelope protein of retroviruses and errantiviruses, containing a transmembrane domain and a host receptor-binding domain, and mediating formation of infectious virions and targeting of host receptor cells.

Env-like Coding domains possessing some features similar to those of the env domains of retroviruses (such as a transmembrane domain), but for which no similar function has been shown as yet; env-like domains are found in addition to the gag-pol domain of many plant LTR retrotransposons.

Gag (group-specific antigen) Nucleic acid-binding protein forming the structural core of the retrovirus virion core and of the cytoplasmic VLP of retrotransposons.

Int Integrase function ensuring the insertion of the double-stranded DNA daughter copy in the genome; also sometimes referred to as endonuclease (endo).

LTR (long terminal repeat) Long sequence found repeated in the same orientation at both ends of retroviruses and LTR retrotransposons and containing promoter and regulatory sequences involved in transcription, that starts in the 5' LTR and terminates in the 3' LTR.

PBS (primer binding site) Short sequence found downstream to the 5' LTR, and involved in the priming of the (–)-DNA strand synthesis during reverse transcription.

PPT (polypurine tract) Short AG-rich sequence found upstream to the 3' LTR, and involved in the priming of the (+)-DNA strand synthesis during reverse transcription

Prot Protease protein involved in the cleavage of the functional proteins units from the (gag) pol polyprotein produced by the retrovirus or the retrotransposon.

RT Reverse transcriptase protein ensuring the synthesis of the double-stranded daughter DNA copy from the RNA template produced by the integrated provirus or retrotransposon; an associate ribonucleaseH (RNaseH) function ensures the concomitant degradation of the RNA template.

VLP (virus-like particle) Cytoplasmic intermediate analogous to the virion core of retroviruses; the VLP is constituted by structural gag proteins associated to two copies of the RNA template, together with enzymatic proteins (RT, int); the reverse transcription of the RNA template is thought to occur within the VLP.

Introduction and General Classification

Transposable elements are ubiquitous mobile DNA sequences found in both prokaryotic and eukaryotic genomes. For a long time, they were considered as 'parasite' DNA, but have now been shown to be major components of plant genomes, where they can represent up to 80% of the bulk of large cereal genomes. Transposable elements, with a few notable exceptions, encode the functions involved in their mobility and are classified into two main classes with radically different transposition mechanisms. Class I elements or retrotransposons transpose via an RNA intermediate that is reverse-transcribed into a daughter copy DNA and re-inserted into the genome, while class II elements or DNA transposons move directly from DNA to DNA. Retrotransposons thus belong to the large class of reverse-transcribing elements (or retroelements) that multiply by transferring their genetic information from RNA to DNA, and are close relatives to viral entities such as animal retroviruses or pararetroviruses infecting both animal and plant hosts. However, in contrast to true infectious viruses, which propagate only functional genomes, retrotransposons are vertically transmitted intrinsic components of host genomes. Their integrated genomes are thus submitted to evolutionary drift, with various mutations and restructurations destroying their functionality, and defective copies can still be amplified, sometimes at surprisingly high levels, via functional related copies. Plant retrotransposons are thus found in a tremendous variety of structural variants, including highly defective and deleted versions, making their classification a difficult task. In addition, retrotransposons are composed of modules that frequently show incongruent phylogenies, and this intrinsic chimerical nature often obscures classifications based on phylogenies.

There are two main categories of retrotransposons: LTR retrotransposons that contain two long terminal repeats (LTRs), and non-LTR retrotransposons often referred to as retroposons. All functional copies encode as basic modules the structural and functional proteins required for the retrotransposition cycle (summarized in Fig. 1). These include a structural RNA-binding protein (gag or gag-like), and a pol domain encoding the reverse transcriptase (RT) ensuring the synthesis of the double-stranded daughter DNA copy from the RNA template, as well as in most cases a protein ensuring insertion of the double-stranded DNA daughter copy in the genome (referred to as integrase for most elements). A number of plant LTR retrotransposons contain an additional coding domain currently termed env-like, although, as discussed below, functional and structural analogies to retroviral env proteins remain to be demonstrated.

LTR retrotransposons have been further classified in Ty1/*copia*-type elements and Ty3/*gypsy*-type depending on the order of the coding domains. Non-LTR retrotransposons include LINEs (long interspersed nuclear elements), which are elements carrying

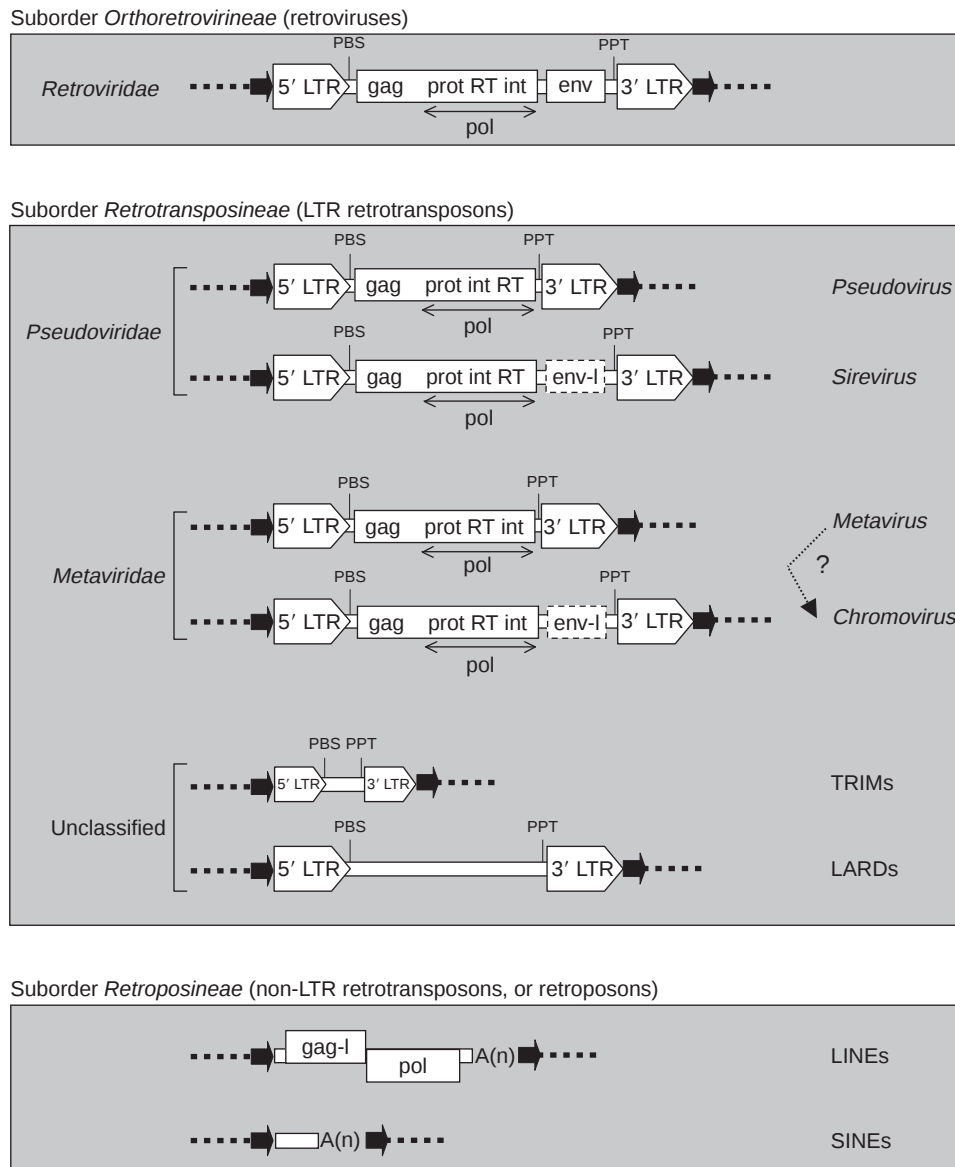


Fig. 1 Structural diversity of retrotransposons found in plant genomes. The structure of a retrovirus has been provided for reference. The gag and pol domains are encoded in one or two frames, depending on the element, and have been represented here as a single ORF for simplification. Family names are indicated on the left and genus names on the right. LTR, long terminal repeat; gag, nucleic acid-binding protein forming the structural core of the retrovirus virion core and of the cytoplasmic VLP of retrotransposons; prot, protease protein involved in the cleavage of the functional proteins units from the (gag) pol polyprotein; int, integrase function ensuring the insertion of the double-stranded DNA copy in the genome; RT, reverse transcriptase ensuring the synthesis of the double-stranded DNA copy from the RNA template; env, envelope protein ensuring infectiosity of the retrovirus; env-I, env-like coding domain possessing structural features similar to env, but for which no similar function has been shown yet (a dotted box framing the env-like domain indicates that elements within the genus may or may not contain the domain); PBS, primer binding site; PPT, polypurine tract. Arrows flanking elements represent the target site duplication. The question mark linking the genera *Metavirus* and *Chromovirus* indicates that the definition of *Chromovirus* as an additional genus within the family *Metaviridae* has not been included in the official classification, with the result that many chromoviruses are still considered to be metaviruses.

coding sequences, and SINEs (short interspersed nuclear elements), small noncoding elements of a few hundred base pairs that exploit the transposition machinery of LINEs to ensure their amplification. LTR retrotransposons are by far the most abundant transposable elements in plant genomes, while LINEs appears somewhat less represented, in contrast to mammalian genomes. SINEs are also abundant in plant genomes, although their small size prevents them from playing a large structural role.

The continuous discovery of new structural variants of these major categories has made it necessary to revise and refine previous classifications. Retrotransposons have been included in a recent classification proposed for all reverse-transcribing elements, and based on viral nomenclature with type species forms defined for some genera (Table 1). All retroelements have been defined as belonging to class *Retroelementopsida*, presently divided in two orders, order *Retrovirales* that encompasses retroviruses, pararetroviruses, and LTR

Table 1 Classification of retroelements

	Type	A few examples in Viridiplantae ^a
Order Retrovirales		
Suborder Orthoretrovirineae		
Family Retroviridae (retrovirus)		Not yet
Suborder Pararetrovirineae (pararetrovirus)		
Family Hepadnaviridae		Not yet
Family Caulimoviridae		CaMV, BSV, integrated EPRVs (banana BSV; petunia PVCV, tobacco TVCV, <i>Nicotiana</i> NtoEPRV, and Ns EPRV)
Suborder Retrotransposineae (LTR retrotransposons)		
Family Pseudoviridae (Ty1/copia retrotransposons)		
Genus Pseudovirus	Ty1	Ta1 (<i>Arabidopsis</i> X1329), Tnt1 (tobacco X13777), Tst1 (potato X52387), Tto1 (tobacco D83003), WIS2 (wheat X63184), BARE1 (barley Z17327), Hopscotch (maize U12626), Tto1 (tobacco D83003), Tos17 (rice D85876), RIRE1 (rice D85597), Art1 (<i>Arabidopsis</i> Y08010), Retrofit (rice U72726), Melmoth (kale Y12321), Tgmr (soybean U96748), Stonor (maize AF082134), Panzee (pigeon pea AJ000893), Tpv2 (bean AJ005762), LERE1 (tomato AF275345), Retrolyc1 (<i>S. peruvianum</i> AF228701), AtRE1 (<i>Arabidopsis</i> AB021265), TLC1 (<i>S. chilense</i> AF279585), Toto1 (tomato AF220602), Angela (wheat DQ666286)
Genus Hemivirus	Copia	Osser (<i>Volvox</i> X69552)
Genus Sirevirus	SIRE1	SIRE1 (soybean AY205608), ToRTL1 (tomato U68072), Endovir (<i>Arabidopsis</i> AY016208) Opie-2 (maize U68408), PREM2 (maize U41000), Osr7 (rice AP002538), Osr8 (rice AC021891)
Family Metaviridae (Ty3/gypsy retrotransposons)		
Genus Errantivirus	gypsy	Not yet
Genus Metavirus	Ty3	Athila (<i>Arabidopsis</i> X81801), Cyclops (pea AJ000640), RIRE8 (rice AB014740), RIRE2 (rice AB030283), BAGY2 (barley AJ279072), Cinfu1 (maize AF049110), Grande1 (maize X97604), Piggy1 (pea AY299398), Ogre (pea AY299398), Diaspora (soybean AF095730) Genus <i>Chromovirus</i> ? del1 (lilium X13886), IFG7 (pine AJ004945), Reina (maize U69258), RIRE3 (rice AB014738), RIRE7 (rice AB033235), Galadriel (tomato AF119040), tekay (maize AF448416), Beetle1 (beet AJ539424)
Genus Semotivirus	BEL	Not yet
Genus ?	DIRS	Not yet
Unclassified		LARDs: Sukkula (barley AY054378), Squiq (rice AY355293), Spip (rice AY355292), Dasheng (rice) TRIMs: (potato AJ276865), Katydid (<i>Arabidopsis</i>), mini-Toto1 (tomato X58273)
Order Retrales		
Suborder Retroposineae (non-LTR retrotransposons)		
	L1	Cin4 (maize Y00086), del2 (lilium Z17425), Isabelle (maize AF326781), Karma (rice AB081316), Bali1A (rapeseed AF525305)
	Alu	TS (tobacco D17453), S1 (rapeseed L76840), Au (<i>Aegilops</i> AB046134)
Suborder Retronineae (bacterial retrons)		

^aElements have been classified according to year of publication or sequence submission to Genbank databases. Note that in several cases, only partial sequences are available. Extensive data on various plant elements can be found at: Repbase, a database of repetitive DNA elements of all organisms, the TIGR Plant Repeat Database and the Triticeae-specific TREP database. Retroelements were classified according to Hull R (2001) Classifying reverse transcribing elements: A proposal and challenge to the ICTV. *Archives of Virology* 146: 2255–2261; and Boeke JD, Eickbush T, Sandmeyer SB, and Voytas DF (2004) *Pseudoviridae/Metaviridae* In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, and Ball LA (eds.) *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, pp. 397–420. San Diego, CA: Elsevier Academic Press.

retrotransposons, and order *Retrales* that contains non-LTR retrotransposons and bacterial retrons. While order *Retrales* has not been further organized at the present time, order *Retrovirales* has been divided in several suborders, families, and genera.

Suborder *Orthoretrovirinae* contains the family *Retroviridae*, or true retroviruses, of which no example is known in plants at the present time.

Suborder *Pararetroviridae* contains two families of pararetroviruses, one of which (*Caulimoviridae*) is represented in plants. Pararetroviruses represent sensibly different lineages of retroelements, and are thought to derive from preexisting viruses having acquired a reverse transcription function, presumably from retrotransposons of the family *Metaviridae*. Many plant genomes also contain vertically transmitted integrated endogenous pararetroviral sequences (EPRVs), which represent integrated derivatives of pararetroviruses, although the infectious counterpart has not always been characterized. Plant pararetroviruses are not discussed here.

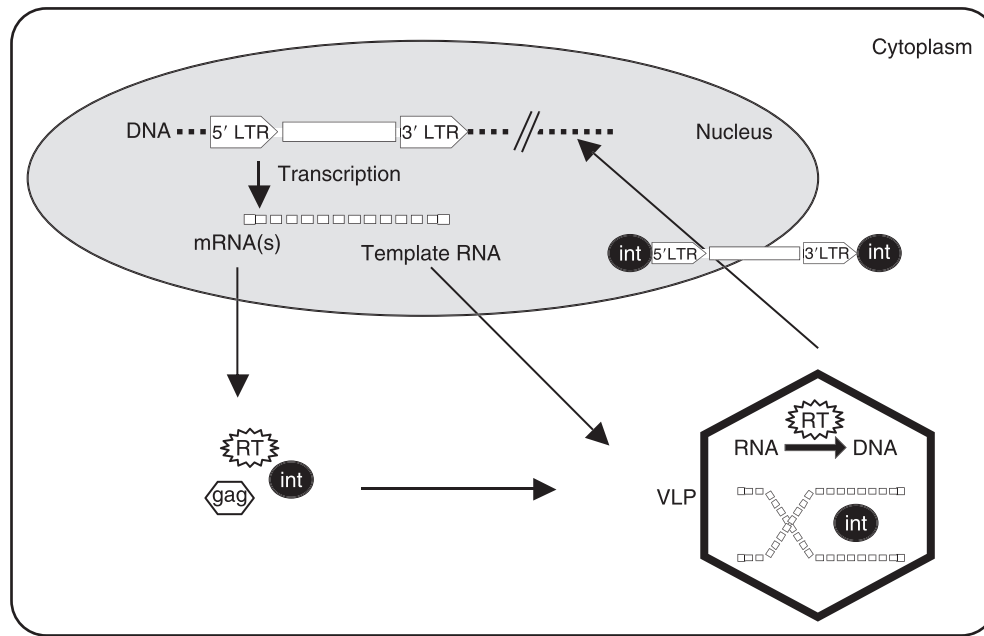


Fig. 2 The amplification cycle of LTR retrotransposons. Transcription begins in the 5' LTR and ends in the 3' LTR, and produces both the RNA template used for reverse transcription and mRNAs translated into proteins. The gag protein binds to the RNA template and forms a cytoplasmic virus-like particle (VLP). Together with the RNA template, VLPs encapsulate the RT that will produce the double-stranded DNA copy, and the integrase (int) involved in transfer of the linear DNA copy to the nucleus and its insertion in the genome.

Suborder *Retrotransposineae* contains LTR retrotransposons, and has been divided in family *Pseudoviridae* (Ty1/*copia* retrotransposons) and family *Metaviridae* (Ty3/*gypsy* retrotransposons), based on the organization of the pol domain.

The Life Cycle of LTR Retrotransposons

High Similarities with Intracellular Steps of the Retroviral Cycle

LTR retrotransposons represent by far the most abundant plant retrotransposons. They share striking similarities with animal retroviruses, both in their structure (Fig. 1) and in their replication cycle (Fig. 2). Like integrated retroviral forms (termed proviruses), LTR retrotransposons are terminated by two LTR sequences. As the reverse transcription process generates two identical LTRs, with mutations accumulating subsequently, the rate of divergences between LTRs of a given copy is often used as a molecular clock to date its insertion. The gag and pol coding domains are found between the LTRs. Other typical features include short sequences known as PBS (primer binding site) and PPT (polypurine tract). The PBS is found immediately downstream the 5' LTR, and is involved in priming of the (–)-DNA strand synthesis during reverse transcription. It is generally complementary to the 3' end of a tRNA recruited for the priming of DNA synthesis (tRNA^{met} for most *Pseudoviridae*, more variable for *Metaviridae*). The PPT is a short AG-rich sequence found immediately upstream to the 3' LTR, and involved in the priming of the (+)-DNA strand synthesis. Inserted copies are typically bounded by short direct repeat of the target host site, generated upon insertion, and usually 5 bp long for plant LTR retrotransposons.

Transcription begins in the 5' LTR, ends in the 3' LTR, and produces both the RNA template used for reverse transcription and the mRNA(s) that will be translated into proteins, after cleavage of the gag-pol polyprotein by the protease (prot) function. The gag protein binds to the RNA template and forms a cytoplasmic virus-like particle (VLP) analogous to the retroviral virion core. Together with the RNA template, VLPs encapsulate the RT that will produce the double-stranded DNA copy, and the integrase (int) that will be involved in transfer of the linear DNA copy to the nucleus and its insertion in the genome. This amplification cycle includes multiple steps of control, with transcription being absolutely crucial for retrotransposition. As a consequence, the mode of amplification of a particular element will be conditioned by its transcriptional regulation.

The amplification cycle of LTR retrotransposons is thus closely related to the intracellular steps of the retroviral cycle, with strong similarities in the reverse transcription and integration processes, as well as similar involvement of a cytoplasmic encapsulated intermediate. It has to be noticed that the amplification of non-LTR retrotransposons is sensibly different, with reverse transcription occurring at the integration site. The reality of the retroviral-type cycle of LTR retrotransposons has been demonstrated in yeast, but not yet in plants. However, VLPs have been detected for the barley BARE1 element. In addition, it has been demonstrated very recently that plant elements of the genus *Sirevirus* (see below) encode unusually large Gag domains extended in their C-terminus, and that this Gag extension binds to LC8/LC6 dynein proteins. Although the biological significance of this interaction remains to be determined, it is very similar to the binding of some human retroviruses' Gag protein to LC8 proteins to

ensure intracellular movement of virion cores along microtubules. Such observations reinforce the strong similarities between the life cycle of retroviruses and LTR retrotransposons.

The major difference between retroviruses and LTR retrotransposons resides in the infectious potentialities of the former, mediated by the *env* function that allows them to produce extracellular virions and infect host receptor cells. *Bona fide* retrotransposons, on the other hand, are not infectious and their amplification cycle is exclusively intracellular. However, we shall see later that this assumption should be treated with care and that the delimitation between retroviruses and retrotransposons is increasingly fuzzy, possibly even in plants.

It is now considered as certain that retroviruses are derivatives of some LTR retrotransposon *Metaviridae* lineages. LTR retrotransposons are themselves derived, with LINE-like ancestors, from the assembly of various modules together with an RT-like function derived from ancestral cellular functions used for the transition of genetic information from RNA to DNA at the dawn of life. Retroviruses have evolved from LTR retrotransposons through transduction of *env* domains ensuring their ability to exit the cell. Although the process can be reversed (e.g., with vertebrate endogenous retroviruses having lost their *env* functionality and reversed to a restricted intracellular life cycle), retroviruses can be seen as the ultimate evolutionary step in freeing of cellular genetic information outside of the cell, and they basically represent LTR retrotransposons that have succeeded in life. It is not clear whether this process has occurred only once during the evolution of *Retroviridae*; however, it is now increasingly clear that a number of their LTR retrotransposons' left-behind cousins have repeatedly attempted to join them in acquiring an extracellular life and that some of them have truly succeeded, albeit not yet in plants.

Evolution as Quasispecies-Like Populations

Other similarities of retrotransposons with retroviral characteristics include their high level of sequence variability. This has been best illustrated for elements related to the tobacco Tnt1 element, which are found in many species of the family Solanaceae, and evolve via a high variability of the LTR region carrying regulatory features (U3 region). Species of the genus *Nicotiana* contain Tnt1 elements that are classified into several groups (termed subfamilies) which are closely related, except in their U3 regions that differ strongly. Members of each subfamily are detected within each species, and their U3 divergence thus predates the *Nicotiana* radiation. However, the respective proportions of each subfamily differ in each species, indicating that each host species has preferentially amplified specific subfamilies. Similar evolutionary patterns have been detected in other Solanaceae species, such as tomato, pepper, or aubergine, which contain elements closely related to *Nicotiana* elements, except in the U3 region, and which can also be further subdivided in subfamilies based on the nature of the U3 region only. Interestingly, the different tobacco Tnt1 subfamilies are each expressed in response to different stimuli, within the same host genome. The very ancient Tnt1 element has thus evolved in Solanaceae a large range of highly related progeny populations that have gained new regulatory sequences and new expression patterns.

This strategy is highly similar to the evolutionary patterns of RNA viruses and retroviruses, generating populations of closely related but different genomes referred to as quasispecies. This variability has been attributed to the high-error-prone process of reverse transcription, due the lack of proofreading repair activity of RNA polymerases and reverse transcriptases. This allows viral populations to evolve very rapidly when environmental conditions change and endow retroviruses with high adaptative capacity. Thus, LTR retrotransposons similarly evolve continuums of closely related sequences, which can be defined as quasispecies-like. This evolutionary pattern allows them to vary their amplification conditions, as regulatory controls in the U3 will determine the conditions in which the element will be transcribed and amplified. The preferential amplification of specific subfamilies in each host species suggests that optimal patterns of amplification may have been selected and maintained in each host genome, possibly depending on its own evolutionary, reproductive, and environmental history.

The Different Genera of LTR Retrotransposons

The Family *Pseudoviridae* (Ty1/*copia* Retrotransposons)

Pseudoviridae are differentiated from all other LTR retrotransposons and from retroviruses by an unusual organization of the pol domain, with the *int* domain placed before the RT domain (Fig. 1). This confirms that the different LTR retrotransposon families have evolved by independent acquisition of modular functions. Historically, Ty1/*copia* retrotransposons have been the first LTR retrotransposons characterized in higher plants, and the best-known plant LTR retrotransposons belong to this family, including the tiny handful of elements for which mobility has been demonstrated. *Pseudoviridae* have been classified in the genera *Pseudovirus* (type species Ty1 of yeast), *Hemivirus* (type species *copia* of *Drosophila*), and the recent genus *Sirevirus* (type species, SIRE1 of soybean).

Pseudoviruses and hemiviruses differ by the fact that the former use a full tRNA^{met}i to prime synthesis of the (–)-DNA strand, while the latter use a half tRNA^{met}i. So far, no hemivirus has been referenced in plants, and all LTR retrotransposons usually – and incorrectly – referred to as *copia*-type retrotransposons in fact classify in the genus *Pseudovirus*. Sireviruses have so far been identified only in plants. They are phylogenetically well separated from the two other genera, and display some very interesting characteristics. In particular, several members of the genus have acquired additional ORFs carrying additional functions that present some

similarities to the env domain of retroviruses, such as a transmembrane domain (TMD). This is, for instance, the case with SIRE1, ToRTL1, and Endovir (Table 1). A number of other elements classified as sireviruses, such as Opie-2, PREM-2, Osr7, or Osr8 (Table 1), do not carry additional ORFs with typical env-like features. However, all of them carry a significant amount of noncoding DNA after the end of the pol domain, from a few hundred base pairs to over 1 kbp, while members of the genera *Pseudovirus* and *Hemivirus*, such as Tnt1, Ty1, or *copia*, separate the pol domain from the 3' LTR by a few dozen nucleotides or less. This suggests that the env-less sireviruses may also have contained similar additional ORFs that have decayed over evolutionary times.

The Family *Metaviridae* (Ty3/gypsy Retrotransposons)

The family *Metaviridae* is characterized by an organization of the pol domain identical to that of retroviruses, with the int protein encoded at the end of pol, after the RT domain (Fig. 1). Metaviruses appear to be basal to both retroviruses and pseudoviruses, and the family *Metaviridae* therefore encompasses a large variety of members, whose classification is probably not at its final stage yet. Although metaviruses have been characterized in plants more recently than pseudoviruses, they have been found to be more abundant in several plant genomes.

The family *Metaviridae* comprises the genera *Errantivirus* (type species *gypsy* of drosophila), *Metavirus* (type species Ty3 of yeast), and *Semotivirus* (type species BEL of drosophila) (Table 1). To these three genera are often added the DIRS retrotransposons, originally described from *Dictyostelium discoideum*, that show differences with other metavirus elements. DIRS do not carry typical LTRs, that is, direct repeats in the same orientation, and do not encode an int. Instead, they are bounded by inverted long repeats and integration in the host genome is ensured by a tyrosine-recombinase. In spite of these differences, they are more related to metaviruses, both in terms of RT sequence and organization of the pol domain, with the tyrosine-recombinase placed downstream to the RT domain. They could represent chimerical elements having acquired an RT domain from some metaviruses, and are placed by several authors as a putative additional genus (still unnamed) in this family. No DIRS, semotivirus, or errantivirus has yet been characterized in plants, with the totality of so-called *gypsy*-like retrotransposons of plants classified in the genus *Metavirus* (Table 1).

Errantiviruses are grouped together on the basis of the presence of an additional env-like domain, which ensures a real envelope function for the two infectious *gypsy* and ZAM elements of drosophila. *Gypsy* and ZAM are the first examples of true retroviruses outside of the vertebrate subphylum. However, their phylogenies place them quite apart from retroviruses, and they are more closely related to other LTR retrotransposons of the family *Metaviridae*. They do presumably represent recent progresses of LTR retrotransposons toward an extracellular life cycle.

The separation of errantiviruses and metaviruses has been interpreted by many authors as based on the lack of an env or env-like domain in metaviruses. However, the situation is more complex. Many plant elements presently classified in the genus *metavirus* indeed do not encode for additional ORFs after the pol domain (such as del1, IFG7, Reina, RIRE3, RIRE7, Galadriel, tekay, and Beetle-1; Table 1). In contrast to env-less sireviruses mentioned above, these elements do not show significant stretches of additional sequences between the pol domain and the 3' LTR (only a few dozens at the most, often very few), indicating that they may never have contained any additional ORF. However, a large number of plant elements presently classified in the genus *Metavirus* do contain an additional ORF with typical env-like features (such as Athila, Cyclops, RIRE8, RIRE2, BAGY2, Cinfu1, Grande1, Piggy1, or Ogre; Table 1). The two groups of elements cluster separately in most phylogenetic studies, indicating that they could be considered as two different genera of the family *Metaviridae*. An additional genus, *Chromovirus*, has recently been proposed within the family *Metaviridae*, based on the presence of a chromodomain in the integrase. However, it has not been included in the current official classification. Interestingly, all plant metaviruses devoid of additional coding domains seem to be chromoviruses. It is therefore likely that the classification of *Metaviridae* will be reconsidered in future, to separate the two groups of metaviruses and recognize the existence of the chromodomain. Interestingly, the soybean Diaspora element is devoid of any additional DNA between pol and the 3' LTR; however it clusters with env-containing elements, suggesting that its original env-like domain has been deleted. Therefore, the presence/absence of an env-like domain would not be a relevant criterion in the redefinition of genera within the family *Metaviridae*.

Unclassified LTR Retrotransposons

A number of LTR retrotransposons have recently been discovered that cannot be classified into any of the above genera due to their lack of coding sequences. The TRIMs (terminal repeat retrotransposons in miniature) consist of small LTRs framing a short central domain of noncoding sequences, while the LARDs (large retrotransposon derivatives) are larger defective derivatives that contain a large internal domain without significant similarities to retrotransposon coding domains. Both TRIMs and LARDs share many of the structural characteristics of LTR retrotransposons, such as RT priming sites PBS and PPT, and are highly amplified, suggesting efficient transactivation by functional retrotransposons. For several LARDs, sequence identities have been found with LTRs of canonical LTR retrotransposons (e.g., Dasheng with RIRE2, Squip with RIRE8, Spip with RIRE3), suggesting that they have been derived from these elements, and may eventually be classified together. Similarly, some TRIM-like small elements are derived from canonical LTR retrotransposons, such as the tomato mini-Toto1 element, derived from the Toto1 pseudovirus. In addition, a number of intermediate situations are frequently found, such as TRIM derivatives containing larger additional unknown DNA, a

possible intermediate situation between TRIMs and LARDs. The existence of TRIMs and LARDs and various intermediates indicates that plant genomes contain a large range of retrotransposons with structural features much more diverse and complex than previously thought. As long as the features required in *cis* for amplification are conserved (e.g., LTRs, PBS, and PPT, as well as a few internal sequences involved in packaging), and as long as the LTR promoter remains transcriptionally active to ensure the production of an RNA template, these defective or atypical elements can be amplified to high levels using the structural and enzymatic retroviral functions provided by their original functional parent, or possibly by other related LTR retrotransposons.

Env-Like Functions in Plants?

Although plant retrotransposons with additional ORFs of the *env* type have been detected, the potential role of such a function in infectivity remains very hypothetical in plants. Retroviral *env* glycoproteins carry a number of typical structures, such as a TMD mediating formation of infectious virions and a host receptor-binding domain, mediating targeting of host receptor cells. No significant homology to any retroviral *env* gene has been shown for any plant element; however, such identification is highly unlikely due to the intrinsic high variability shown by animal retroviral *env* proteins. As the definition of *env*-like functions in plant retrotransposons is mostly based on the presence of TMDs, the issue of the naming of these additional ORFs has been hotly debated. TMDs can also be found in various proteins, and the *env* denomination has a specific meaning related to anchoring of the protein in the membrane envelope of extracellular virion particles and to recognition of receptors on target cells to ensure viral infection after membrane fusion. Quite likely, such mechanisms are rarely used by plant viruses for host-to-host transmission, due to the cell wall barrier. Plant retrotransposons carrying additional ORFs related to functions used by most plant viruses to ensure host-to-host transmission, for example, via vector insects, have not yet been found. However, examples of enveloped viruses do exist in plants, and retrotransposon-encoded *env*-like proteins could be involved in anchoring of VLPs to cellular membranes, playing a role in cell-to-cell transmission via plasmodesmata and possibly in transmission by insects.

Furthermore, conservation of *env*-like domains has been observed between elements from different plant genera within both the *Metaviridae* and the *Pseudoviridae* (SIRE1 of soybean and Endovir of *Arabidopsis*). In addition, conserved splice acceptor sequences are detected 5' of *env*-like domains of several plant metaviruses, and production of an *env*-like spliced subgenomic RNA has been demonstrated for the barley BAGY-2 element, a mechanism similar to those used by retroviruses to ensure production of the *env* protein. In contrast, sireviruses such as SIRE1 use stop codon suppression to express *env*-like protein. The existence of such specific expression mechanisms strongly suggests that *env*-like domains are not mere incidental transduction of cellular functions, as shown for the maize Bs1 element, and that acquisition of additional coding domains with TMDs, whether in a quest or not for extracellular life, may play an important role in the plant LTR retrotransposon life cycle. In the search of a better definition, and in spite of lack of evidence for any retroviral *env*-type function, it has been proposed to maintain the *env* denomination due to its vernacular long-term use: for example, the nucleocapsid protein used to build up the VLP core particle is commonly termed *gag* in all retrotransposons, by analogy to the *gag* 'group-specific antigen' region of retroviruses, although such functional denomination does not make sense in plants.

A striking observation is that LTR retrotransposons generally have been very efficient in acquiring *env*-like domains, and have done so many times. For instance, additional ORFs with *env*-like features have been observed both for Ty1/*copia* and Ty3/*gypsy* retrotransposons, suggesting that they have been acquired independently during evolution. In addition, the *Metaviridae* contain clear examples of acquisition of *env* function from different sources, for example, from baculoviruses for *gypsy* of *Drosophila*, and from phleboviruses and herpesviruses for the semotiviruses Cer of *Caenorhabditis elegans* and Tas of *Ascaris lumbricoide*s, respectively. This indicates that LTR retrotransposons have repeatedly been able to hijack viral functions leading to extracellular autonomy, and there is no conceivable reason for plants to be an exception.

The Nomenclature Issue

Rules for naming individual elements also continue to be hotly debated. Earlier retrotransposons have been named according to authors' personal fancies (e.g., Athila, Cyclops, Opie, Ogre) or based on the plant species they were isolated from, but following no defined rule (e.g., Tnt1 for Transposon of Nicotiana Tabacum, BARE-1 for Barley Retrotransposon, WIS-2 for Wheat Insertion Sequence). An official nomenclature has recently been proposed for all retroelements, with three letters used for the species, up to three (exceptionally four) for the element and 'V' for virus. In this nomenclature, the yeast Ty1 and *Drosophila copia* elements have been renamed SceTy1V and DmeCopV, respectively, and Tnt1 and BARE1 have been renamed NtaTnt1V and HvuBV, respectively. However, this nomenclature has so far not been implemented by the plant scientific community, most likely because it does not solve the major problem arising from the presence of highly related elements in different plant species or genera, which is a general rule in plants.

In addition, most transposable element populations, and especially retrotransposons, are generally composed of populations of closely related sequences. Defining whether a particular copy is related to an already known element, or is different, is thus often a difficult task. A consensus position has been to consider that copies that show above 75% identity over most of their length (not taking into account deletions or insertions of unrelated material) are variants of the same element. However, many elements have been discovered and named independently. For instance, the wheat WIS2 and Angela elements are quite similar to the barley BARE1 element, with internal nucleotide sequences 75–80% identical between BARE1 and WIS2 and over

80% identical between BARE1 and Angela. Similarly, *Solanum* subsection *Lycopersicon* counterparts of the tobacco Tnt1 element have been named Retrolyc1 (first identified in *S. peruvianum*) and TLC1 (identified in *S. chilense*), although Retrolyc1 and Tnt1 nucleotide sequences are over 85% identical and TLC1 and Retrolyc1 are 93% identical. This highlights the difficulties inherent to a nomenclature based on host species names.

A proposition was made to replace species-based names by first names, a trend that seems to be developing at present, at least in the plant field. More specifically, it has been proposed that female first names should be used for retroelements and male first names for class II DNA transposons, although this particular proposal is not really implemented. Such type of nomenclature should alleviate the problem of the presence of related elements in different hosts, provided authors make sure their favorite first name has not been used yet. Attempts to unify the nomenclature (as well as the general classification) of transposable elements of Triticeae are currently coordinated by Dr. Thomas Wicker, who curates the TREP database for Triticeae repeats. This should lead to the redefinition of consensus guidelines that could be fruitfully followed by the entire scientific community for transposable elements of all plant genera.

See also: Caulimoviruses (*Caulimoviridae*). Movement of Viruses in Plants. Plant Reoviruses (*Reoviridae*). Vector Transmission of Plant Viruses

Further Reading

- Boeke, J.D., Eickbush, T., Sandmeyer, S.B., Voytas, D.F., 2004. Pseudoviridae/Metaviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press, pp. 397–420.
- Casacuberta, J.M., Vernhettes, S., Audéon, C., Grandbastien, M.-A., 1997. Quasispecies in retrotransposons: A role for sequence variability in Tnt1 evolution. *Genetica* 100, 109–117.
- Grandbastien, M.-A., Audéon, C., Bonnivard, E., *et al.*, 2005. Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Special Issue: Retrotransposable Elements and Genome Evolution. *Cytogenetic and Genome Research* 110, 229–241.
- Havecker, E.R., Gao, X., Voyas, D.F., 2004. The diversity of LTR retrotransposons. *Genome Biology* 5, 225.
- Havecker, E.R., Gao, X., Voytas, D.F., 2005. The sireviruses, a plant-specific lineage of the Ty1/copia retrotransposons, interact with a family of proteins related to dynein light chain 8. *Plant Physiology* 139, 857–868.
- Hull, R., 2001. Classifying reverse transcribing elements: A proposal and challenge to the ICTV. *Archives of Virology* 146, 2255–2261.
- Kalendar, R., Vicent, C.M., Peleg, O., Ananthawat-Jonsson, K., Bolshoy, A., Schulman, A.H., 2004. Large retrotransposon derivatives: Abundant, conserved but nonautonomous retroelements of barley and related genomes. *Genetics* 166, 1437–1450.
- Kordis, D., 2005. A genomic perspective on the chromodomain-containing retrotransposons: Chromoviruses. *Gene* 247, 161–173.
- Lucas, H., Yot, P., Lockhart, B.E.L., Capy, P., 2004. Taxa of viruses, virus-like and subviral agents infecting Poaceae, class '*Retroelementopsida*'. In: Lapierre, H., Signoret, P.A. (Eds.), *Virus and Virus Diseases of Poaceae (Graminae)*. Paris: INRA Editions, pp. 279–303.
- Malik, H.S., Henikoff, S., Eickbush, T.H., 2000. Poised for contagion: Evolutionary origins of the infectious ability of invertebrate retroviruses. *Genome Research* 10, 1307–1318.
- Peterson-Burch, B.D., Wright, D.A., Laten, H.L., Voytas, D.F., 2000. Retroviruses in plants? *Trends in Genetics* 16, 151–152.
- Vicent, C.M., Kalendar, R., Schulman, A.H., 2001. Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. *Genome Research* 11, 2041–2049.
- Wicker, T., Sabot, F., Hua-Van, A., *et al.*, 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8, 973–982.
- Witte, C.P., Le, Q.H., Bureau, T., Kumar, A., 2001. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proceedings of the National Academy of Sciences, USA* 98, 13778–13783.
- Wright, D.A., Daniel, F., Voytas, D.F., 2001. *Athila4* of *Arabidopsis* and *Calypso* of soybean define a lineage of endogenous plant retroviruses. *Genome Research* 12, 122–131.

Relevant Websites

- <http://www.girinst.org>
Genetic Information Research Institute (GIRI).
- <http://www.tigr.org>
The TIGR Plant Repeat Databases, J. Craig Venter Institute (JCVI).
- <http://wheat.pw.usda.gov>
The Triticeae Repeat Sequence Database (TREP), GrainGenes, Agricultural Research Service, USDA.

Vector Transmission of Plant Viruses

Etienne Herrbach and Quentin Chesnais, University of Strasbourg, Colmar, France

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S. Blanc, Vector Transmission of Plant Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00433-7](https://doi.org/10.1016/B978-012374410-4.00433-7).

Nomenclature

HC Helper component

HC-Pro Helper component-protease

kDa Kilo Dalton

ORF Open reading frame

Glossary

Helper component Virus-encoded nonstructural protein required for vector transmission of a virus.

Hemolymph Physiological fluid that circulates in the body cavity, or hemocoel, of arthropods in direct contact with their inner organs.

Horizontal transmission Transmission of a pathogen from one host to another within the same generation, as opposed to vertical transmission.

Phytoviruses Viruses infecting plants.

Piercing-sucking insects Insects adapted to sap or blood feeding, with the mouthparts in form of long chitin stylets

able to pierce and penetrate tissues and cells, and allow ingestion of their content.

Protist A unicellular eukaryotic organism.

Readthrough protein A protein expressed by two open reading frames separated by a stop codon that is bypassed during translation.

Vector Mobile organism acquiring a pathogen on an infected host and inoculating it in a healthy one.

Vertical transmission Transmission of a pathogen from the parent(s) to the offspring, usually through the germline.

Introduction

Plant-infecting viruses, here referred to as phytoviruses, are known since the last two decades of the 19th century. To date, over 900 species of phytoviruses have been described and their number is steadily increasing as a result of latest virological technologies. Unlike animal viruses, phytoviruses infect immobile hosts; moreover, they must cross the plant cuticula and the cell wall in order to infect living plant cells. The majority of phytoviruses have evolved a close relationship with mobile organisms, called 'vectors', which help them in their plant-to-plant propagation; this biological phenomenon is coined 'horizontal transmission'. However, some very stable viruses, such as tobamoviruses, are able to infect plants by leaf contact, or by wounds, and require no vector. Many viruses are transmissible through seed and/or pollen, and the majority can be transferred through vegetative multiplication; these routes are referred to as 'vertical transmission'. Many virus species can use both horizontal and vertical transmissions. Finally, for many viruses (e.g., in genera *Velarivirus*, *Foveavirus*, or *Hordeivirus*), the natural vector, if any, remains unknown. The present article aims at summarizing the current knowledge on the various strategies used for vector transmission of plant viruses.

Plant Virus Vectors

The vast majority of vector organisms belong to phytophagous insect taxa, mainly sap feeding members of the order Hemiptera (aphids, leaf- and planthoppers, whiteflies, coccoids). Several chewing insects in the order Coleoptera also are phytovirus vectors. Other important vectors belong to phytophagous Eriophyid and Tenuipalpid mites, to root-ectoparasitic Longidorid and Trichodorid nematodes, and to soil-dwelling plant-infecting Chytrid fungi and Plasmidio-phorid protists.

The transmission process includes several successive steps. 'Acquisition' refers to the uptake of virions while the vector feeds on an infected plant, whereas 'inoculation' occurs during a subsequent feeding on another plant. 'Latency' is the time lapse between virus acquisition and time when a viruliferous (i.e., virus bearing) vector becomes infective (i.e., able to inoculate the virus). 'Retention' time refers to the time period during which the vector remains infective. These criteria served to define the typology of the strategies involved, i.e., the transmission modes as described below.

Transmission of Plant Viruses by Insects

History of the Typology of the Transmission Modes

Vector transmission of a plant virus was first demonstrated in Japan in 1893 for the leafhopper-transmitted Rice dwarf virus (*Phytoreovirus*, *Reoviridae*). Interestingly, this discovery occurred shortly after that of insects as vectors of animal viruses. Pioneer researchers soon noticed that all virus–vector associations do not follow the same pattern, due to the specific biology of each of the three partners, i.e., plant, virus, and the vector. The typology of these ‘transmission modes’ was then defined for aphid vectors; however, it applies quite well to other vectors with a piercing-sucking apparatus, but not to beetles, equipped with biting-chewing mouthparts.

The first typology, proposed in 1939, was based on the optimal times for acquisition and inoculation and defined the ‘non-persistent’ (seconds to minutes) and the ‘persistent’ (hours to days) modes. This typology was completed in 1956 with the semi-persistent mode, with intermediary features (i.e., minutes to hours). In 1962, the second typology coined the terms circulative and stylet-borne viruses, according to whether the infectivity, and therefore the virus, is maintained or lost through molting (trans-stadial passage) and to whether the virus is detected or not in the insect hemolymph. In practice, circulative corresponds to persistent, and stylet-borne to non-persistent; however, the semi-persistent mode does not fit in this typology. In 1977, it was proposed to unify these classifications on the basis of the retention site of the virions. Thus, in the non-circulative (or externally-borne) mode, the virions are retained at the cuticular lining of stylets and/or foregut, without being transported across any epithelium, and therefore lost during molt; this includes the semipersistent mode, also named ‘foregut-borne’ according to the retention site. In the ‘circulative’ (or internally-borne) mode, the virions are internalized across the gut epithelium into the hemolymph, and from whence across the salivary gland epithelium, the circulation within the hemolymph occurring as virions. Circulative viruses are further classified as propagative or non-propagative, according to their ability to replicate or not in the vector’s organs. Furthermore, a few propagative circulative viruses are capable of a mother-to-progeny transovarial transmission.

The correspondence between feeding behavior of aphids and other hemipteran vectors, and transmission modes is well established. Aphid feeding behavior is characterized by two kinds of stylet insertion into plants. They first proceed to one or usually several brief probes (i.e., within seconds) in epidermal and mesophyll cells, during which cell content is sampled and tasted (host selection behavior) and saliva is expelled into these cells; after this phase the aphid may settle or leave the plant. If it settles, the stylets continue a deeper path into plant tissues, mainly intercellularly, to finally reach the phloem vessels, where the sap ingestion starts. Non-circulative viruses are acquired and inoculated during short probes in epidermal cells or phloem cells, in respectively non- and semi-persistent modes, whereas circulative viruses are generally acquired and inoculated during ingestion from, and salivation into phloem sieve elements. The feeding behavior of other hemipteran vectors differs somewhat from that of aphids. For example, coccoids (scale insects) make no epidermal probes, therefore transmit viruses in the semipersistent mode only. Moreover, the stylet pathway of whiteflies to the phloem is mainly intercellular, whereas that of leafhoppers, with less flexible stylets, pierce straight through mesophyll cells on hosts and non-hosts. For all hemipterans, patterns of feeding behavior vary according to host plant suitability. Many hemipterans can also ingest xylem sap, though this behavior is irrelevant to virus transmission, as xylem-located viruses are unknown.

Circulative Transmission

Circulative Non-Propagative Transmission

This strategy is very specific of phytoviruses, as it has ever been reported for the transmission of neither arthropod-transmitted animal virus (arboviruses), nor non-viral plant pathogens. The best studied pathosystems following this mode are aphids–luteovirids, leafhoppers–mastreviruses, and whiteflies–begomoviruses. Other pathosystems using this mode include aphids–nanovirids and –capulaviruses, and probably many of those involving fungus and protist vectors.

Case of Aphid-Transmitted Luteovirids

These viruses of the family *Luteoviridae*, that are causal agents of several economically important diseases in arable and vegetable crops worldwide, share the property of being phloem-restricted in their host-plants and transmitted by aphids (family *Aphididae*) during phloem feeding (Fig. 1(A)). The route of virions inside the vector has been extensively studied using cellular and molecular biology approaches. The widely accepted model is as follows. Virions are acquired into the vector along with ingested phloem sap. Then, in compatible aphid species–virus species combinations, the particles are retained at putative receptors borne by the luminal side of gut epithelium, then invaginated into gut cells and released at the basal side of these cells (‘transcytosis’) into the hemolymph. From there they diffuse, probably passively, until they reach accessory salivary glands, whose cells transport them by transcytosis from basal to apical side into the saliva (Fig. 2). They can then be inoculated during salivation into the phloem tissue of a new host. No virus particle has ever been observed in any other organ than the gut or the accessory salivary glands, nor has any viral replication ever been reported.

During transcytosis in epithelial cell of the gut and the salivary glands, virions observed using microscopy are always seen inside membranous vesicles, either alone, or grouped in clusters or tubular arrays, suggesting that the particles follow intracellular trafficking routes. In the gut, the transcytosis occurs at either the midgut, the hindgut, or both, depending on the aphid and virus species concerned. The gut barrier seems less specific than the salivary one, since viral particles are detected in the hemolymph of some non-vector species. Specificity is thought to rely on viral ligands that interact with receptors borne on the plasmalemma of

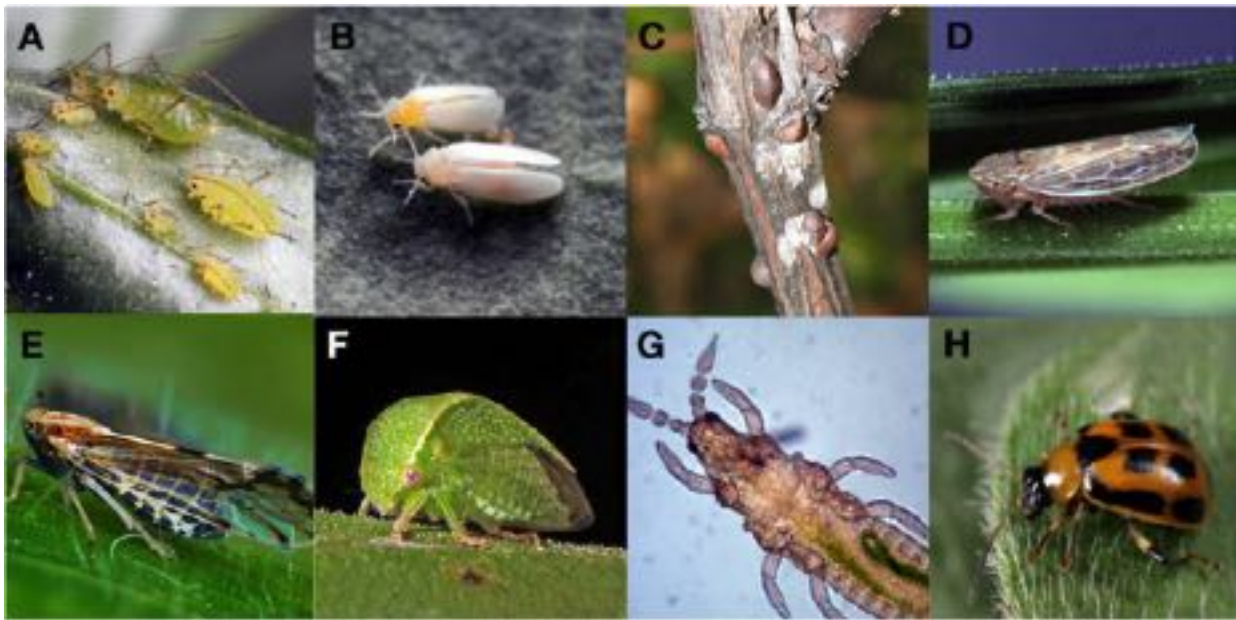


Fig. 1 Main insect vectors of plant viruses. (A) *Acyrtosiphon pisum* (Aphididae); (B) *Bemisia tabaci* (Aleyrodidae); (C) *Parthenolecanium persicae* (Coccidae); (D) *Psammotettix alienus* (Cicadellidae); (E) *Peregrinus maidis* (Delphacidae); (F) *Spissistilus festinus* (Membracidae); (G) *Frankliniella occidentalis* (Thripidae); (H) *Epilachna varivestis* (Coccinellidae). Photographs: (A)–(B) Courtesy of Ian Wright, UC Riverside USA; (C) Courtesy of Gérard Hommay, INRAE Colmar France; (D) Courtesy of Emmanuel Jacquot, INRAE Montpellier France; (E) Courtesy of Antoine Franck, CIRAD France; (F) Courtesy of Ken Childs, USA; (G) Courtesy of Dominique Blancard, INRAE Bordeaux France; (H) Courtesy of Bryan Cassone, Brandon University Canada.

traversed cells of the gut then the salivary gland. Luteovirids have two structural proteins: the major coat protein and a minor readthrough protein composed of the coat protein at the N-terminus followed by the readthrough domain which is extended outside the particle and indispensable for the interaction with putative receptors. Recently, alanyl aminopeptidase N and ephrin receptor proteins have been shown to act as a virion receptor at the gut level, whereas other proteins are suspected to do so. During virion diffusion in the hemolymph, the readthrough protein of at least some species interacts with a chaperone protein (homolog to *Escherichia coli* GroEL) secreted by endosymbiotic bacteria; this interaction is believed to protect the particle from degradation and/or assist its transfer to the salivary glands.

Case of Leafhopper-Transmitted Mastreviruses

Unlike luteovirids, mastreviruses (*Geminiviridae*) infect mesophyll cells and are transmitted by mesophyll- and phloem-feeding leafhoppers (*Cicadellidae*) (Fig. 1(D)). Acquired virions mainly accumulate in the epithelia of the filter chamber and of the midgut, though without replication, from whence they reach the salivary glands. Unlike luteovirids, the gut provides the main selective barrier, as a non-vector can become infective upon microinjection of virions into its hemolymph, as well as upon a simple puncture of the gut wall with a needle. A probing time of less than a minute may be enough for the vector to acquire the virus, indicating that acquisition occurs in the mesophyll while puncturing cells. Conversely, inoculation, requiring longer times, occurs in phloem. Virion entry into the gut cells, mainly at the filter chamber and the anterior part of the midgut, is thought to use lipid-raft-mediated endocytosis, and the viral capsid protein is thought to bear the transmission determinants.

In the same family, the genera *Curtovirus* and *Topocuvirus* are also transmitted according to the circulative non-propagative mode, by respectively leafhoppers and treehoppers (*Membracidae*) (Fig. 1(F)). A treehopper has been reported as a vector of a *Grabovirus* member, with a still undetermined transmission mode. Finally, the recently defined genus *Capulavirus* is transmitted by aphids.

Case of Whitefly-Transmitted Begomoviruses

The genus *Begomovirus* is the genus with the largest number of species in the family *Geminiviridae*, with over 400 species described, mainly in the tropics and subtropics where they cause severe diseases to crops. The vector of all these species is exclusively the whitefly *Bemisia tabaci* (*Aleyrodidae*) (Fig. 1(B)). However, *B. tabaci* now appears as a vast complex of biotypes and 'cryptic' species, with a great invasive potential worldwide; moreover, this complexity is associated to a wide viral genetic diversity and a strong proneness to viral recombination.

Like all geminivirids, begomoviruses are reported to follow the circulative non-propagative transmission mode. However, a propagative and a transovarial transmission of Tomato yellow leaf curl virus (TYLCV) DNA by *B. tabaci* have been reported, though this

Table 1 Families and genera of plant viruses and their vectors. Virus taxa with no known vector are not mentioned

Family	Genus	Vector	Transmission mode
<i>Alphaflexiviridae</i>	<i>Allexivirus</i>	Mite	
	<i>Potexvirus</i>	Aphid ^a , chytrid?, cycloraphan fly ^b	
<i>Aspiviridae</i>	<i>Ophiovirus</i>	Chytrid	
<i>Betaflexiviridae</i>	<i>Carlavirus</i>	Aphid, leafhopper, whitefly ^b	Non-circulative
	<i>Trichovirus</i>	Mite	
	<i>Vitivirus</i>	Coccoid, aphid ^c	Non-circulative
<i>Benyviridae</i>	<i>Benyvirus</i>	Plasmidiophorid	
<i>Bromoviridae</i>	<i>Alfavirus</i>	Aphid	Non-circulative capsid strategy
	<i>Anulavirus</i>	Thrips ^d	
	<i>Bromovirus</i>	Beetle, aphid ^a	
	<i>Cucumovirus</i>	Aphid	Non-circulative capsid strategy
	<i>Ilarvirus</i>	Thrips ^d	
<i>Caulimoviridae</i>	<i>Badnavirus</i>	Coccoid, aphid, lacebug ^b	Non-circulative
	<i>Caulimovirus</i>	Aphid	Non-circulative helper strategy
	<i>Tungrovirus</i>	Leafhopper	Circulative propagative
<i>Closteroviridae</i>	<i>Ampelovirus</i>	Coccoid	Non-circulative
	<i>Closterovirus</i>	Aphids	Non-circulative
	<i>Crinivirus</i>	Whitefly	Non-circulative capsid strategy
<i>Fimoviridae</i>	<i>Emaravirus</i>	Mite	
<i>Geminiviridae</i>	<i>Becurtovirus</i>	Leafhopper	Circulative non-propagative
	<i>Begomovirus</i>	Whitefly	Circulative non-propagative ^f
	<i>Capulavirus</i>	Aphid	Circulative non-propagative
	<i>Curtovirus</i>	Leafhopper	Circulative non-propagative
	<i>Grabovirus</i>	Treehopper	
	<i>Mastrevirus</i>	Leafhopper	Circulative non-propagative
	<i>Topocuvirus</i>	Treehopper	Circulative non-propagative
	<i>Turncurtovirus</i>	Leafhopper	Circulative non-propagative
<i>Kitaviridae</i>	<i>Cilevirus</i>	Mite	Circulative (propagative?)
<i>Luteoviridae</i>	<i>Enamovirus</i>	Aphid	Circulative non-propagative
	<i>Luteovirus</i>	Aphid	Circulative non-propagative
	<i>Polerovirus</i>	Aphid, whitefly ^b	Circulative non-propagative
<i>Nanoviridae</i>	<i>Babuvirus</i>	Aphid	Circulative non-propagative
	<i>Nanovirus</i>	Aphid	Circulative non-propagative (with helper factor)
<i>Peribunyaviridae</i>	<i>Tospovirus</i>	Thrips	Circulative propagative
<i>Phenuiviridae</i>	<i>Tenuivirus</i>	Planthopper	Circulative propagative
<i>Potyviridae</i>	<i>Bymovirus</i>	Plasmidiophorid	Circulative
	<i>Ipomovirus</i>	Whitefly	Non-circulative
	<i>Macluravirus</i>	Aphid	Non-circulative
	<i>Poacevirus</i>	Mite	
	<i>Potyvirus</i>	Aphid	Non-circulative helper strategy
	<i>Roymovirus</i>	Mite?	
	<i>Rymovirus</i>	Mite	
	<i>Tritimovirus</i>	Mite	
<i>Reoviridae</i>	<i>Fijivirus</i>	Planthopper	Circulative propagative
	<i>Oryzavirus</i>	Planthopper	Circulative propagative
	<i>Phytoreovirus</i>	Leafhopper	Circulative propagative
<i>Rhabdoviridae</i>	<i>Cytorhabdovirus</i>	Aphid, planthopper, leafhopper	Circulative propagative
	<i>Dichorhabdovirus</i>	Mite	Circulative (propagative?)
	<i>Nucleorhabdovirus</i>	Leafhopper, planthopper, aphid	Circulative propagative
	<i>Varicosavirus</i>	Chytrid	Circulative

(Continued)

Table 1 Continued

Family	Genus	Vector	Transmission mode
<i>Secoviridae</i>	<i>Cheravirus</i>	Nematode	
	<i>Comovirus</i>	Beetle	
	<i>Fabavirus</i>	Aphid	Non-circulative
	<i>Nepovirus</i>	Nematode, mite	Non-circulative capsid strategy
	<i>Sequivirus</i>	Aphid	Non-circulative
	<i>Torradovirus</i>	Whitefly	Non-circulative
	<i>Waikavirus</i>	Aphid, leafhopper	Non-circulative
<i>Solemoviridae</i>	<i>Sobemovirus</i>	Beetle, aphid, mirid, (thrips, diptera) ^g	Non-circulative?
<i>Tombusviridae</i>	<i>Alphanecrovirus</i>	Chytrid	
	<i>Aureusvirus</i>	Chytrid, plasmiophorid? ^b	
	<i>Avenavirus</i>	Chytrid	
	<i>Betacarmovirus</i>	Beetle	
	<i>Betanecrovirus</i>	Chytrid	
	<i>Dianthovirus</i>	Chytrid ^d	
	<i>Gammacarmovirus</i>	Chytrid	
	<i>Machlomovirus</i>	Beetle, thrips ^b	
	<i>Tombusvirus</i>	Chytrid	Non-circulative capsid strategy
	<i>Umbravirus</i>	Aphid	
	<i>Zeavirus</i>	Chytrid	
<i>Tymoviridae</i>	<i>Marafivirus</i>	Leafhopper	Circulative propagative
	<i>Tymovirus</i>	Beetle	
<i>Virgaviridae</i>	<i>Furovirus</i>	Plasmiophorid	
	<i>Pecluvirus</i>	Plasmiophorid	
	<i>Pomovirus</i>	Plasmiophorid	
	<i>Tobravirus</i>	Nematode	

^aOnly one viral species when assisted by the HC of a potyvirus.

^bOnly one viral species.

^cTwo viral species when assisted by a helper virus.

^dVirus transported along with pollen.

^eOnly one viral species, non-circulative mode, capsid strategy.

^fCirculative propagative transmission seems to occur in some specific virus-vector combinations.

^gRice yellow mottle virus (RYMV) can also be transmitted by orthopterans, hemi- and heteropterans, probably mechanically.

Sources: ICTV 9th and 10th Reports (<https://talk.ictvonline.org/>).

Note: Bragard, C., Caciagli, P., Lemaire, O., *et al.*, 2013. Status and prospects of plant virus control through interference with vector transmission. Annual Review of Phytopathology 51, 177–201. Brown, J.K., 2016. Vector-mediated Transmission of Plant Pathogens. St Paul, MN: American Phytopathological Society Press.

DNA is probably not infective. Moreover, the report of a venereal transmission of a specific TYLCV isolate in certain *B. tabaci* biotypes/species adds a further originality, unique among arthropod-transmitted phytoviruses. Globally, the begomoviruses–whiteflies interaction and its efficiency appear complex and diverse, possibly linked to a co-evolution between vector species or subspecies and a great diversity of viruses. Immuno-microscopic studies on several begomovirus species revealed para-crystalline structures, thought to be virions, in the filter chamber and salivary glands of the vector. Like for luteovirids, the participation of an endosymbiont-secreted chaperone protein during virion diffusion in the whitefly's hemolymph has been shown. The coat protein of begomoviruses, whose sequence is only slightly variable, determines the transmission specificity and its N-terminus is required for receptor-mediated endocytosis in the gut and salivary glands. Whitefly proteins, such as cyclophilin B, are known to be involved in TYLCV transmission.

Circulative Propagative Transmission

While circulative propagative mode is almost the general rule for arthropod-transmitted animal viruses (arboviruses), this strategy is also well known for many phytoviruses, such as the genera *Cytorhabdovirus* and *Nucleorhabdovirus* (*Rhabdoviridae*) (transmitted by leaf-, planthoppers, or aphids), *Fijivirus*, *Oryzavirus* (*Reoviridae*) and *Tenuivirus* (*Phenuiviridae*) (planthoppers), *Marafivirus* (*Tymoviridae*) and *Phytoreovirus* (*Reoviridae*) (leafhoppers), and *Tospovirus* (*Tospoviridae*) (thrips) (Fig. 1(G)). In these cases, virions infect gut cells, are then released into the hemolymph and infect various tissues and organs, including salivary glands. Rhabdoviruses infect the central nervous system, from whence they spread to other organs. The infection of the vector is very similar to the infection of an insect by an entomopathogenic virus, and the vector can therefore also be seen as a virus host, even though such vectors are usually much less impacted by the infection in terms of survival and fecundity. Remarkably, propagative phytoviruses belong (apart from marafiviruses) to families comprising also animal-infecting viruses, and the question of whether they may have evolved from insect viruses has been raised.

Non-Circulative Transmission

The non-circulative transmission is specific to phytoviruses, for which it is the most frequent case (> 50% of viruses), and unknown for vertebrate arboviruses. The virions are adsorbed at the luminal side of cuticula-lined tissues, either stylet food canal or/and foregut, without being internalized, and are therefore also called 'externally-borne'. This accounts for the loss of infectivity of the vector during molting and the shorter periods needed for efficient acquisition and inoculation (Table 2). As mentioned above, classical authors defined (1) the non-persistent mode, where the virions are acquired and inoculated from epidermal and mesophyll cells retained at the stylet level, and (2) the semi-persistent mode, where virions are acquired and inoculated from deeper cells, even in phloem, and/or retained at the foregut level (Fig. 2). However, it is sometimes uneasy to verify these features experimentally and this often vague distinction is less used today. Two hypotheses, though not mutually exclusive, have been put forward to explain the inoculation phase in this transmission mode. On one hand, the ingestion–egestion hypothesis states that the retained virions are regurgitated by the vector with the sampled plant sap during probing and feeding. On the other hand, the ingestion–salivation hypothesis, which is favored in many cases, postulates that the salivation into plant tissues is responsible to detach the retained virions and inoculate them into the host; this implies that, in the case of aphids at least, the retention site lies in the distal part of the stylets where the food and the salivary canals merge together.

Table 2 Characteristics of the different modes of vector transmission of plant viruses

Viruses	Non-circulative		Circulative	Circulative propagative
Mode	Non-persistent	Semi-persistent	Persistent non-propagative	Persistent propagative
Acquisition time	Seconds to minutes	Minutes to hours	Hours to days	Hours to days
Acquisition and inoculation sites	Epidermis	Phloem	Phloem	Phloem
Latency time	None	None	Hours to days	Days
Retention time	Minutes to hours	Hours to days	Several days	Life long
Inoculation time	Seconds to minutes	Minutes to hours	Hours to days	Hours to days
Passage through molt	No	No	Yes	Yes
Replication in vector	No	No	No	Yes
Transovarial passage	No	No	No	Possible for certain viruses
Vector specificity	Large	Medium	Narrow	Narrow

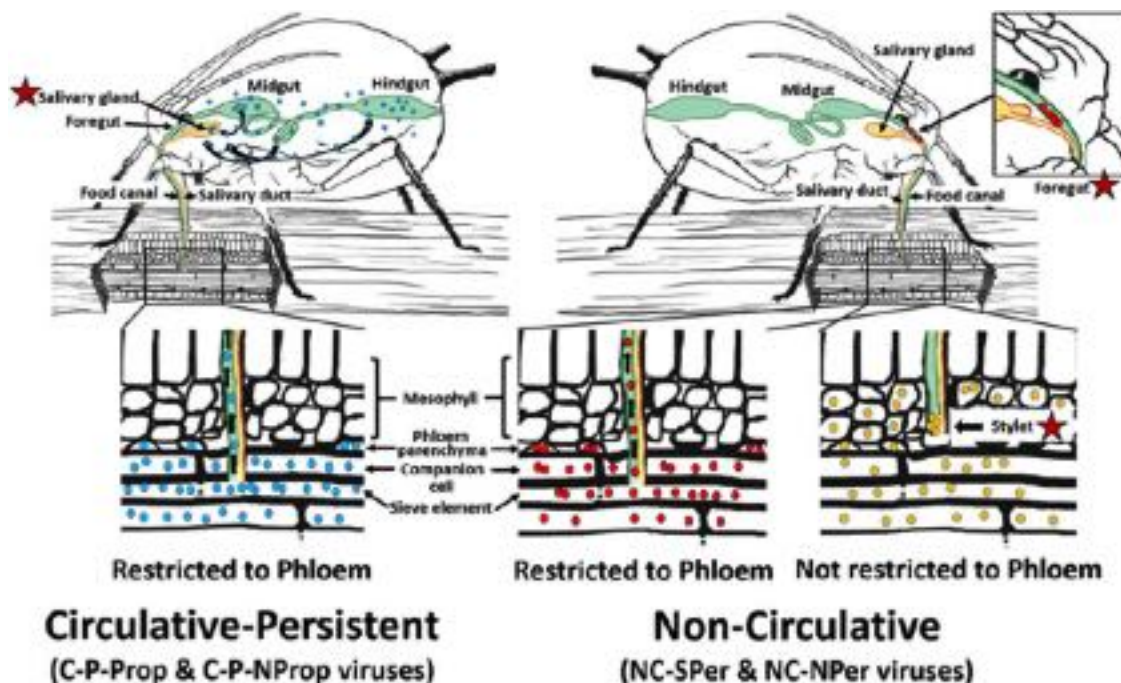


Fig. 2 Schematic representations of plant virus transmission mechanisms. Retention sites for each virus are indicated by a red star adjacent to the labeled area. C-P-Prop: circulative-persistent propagative viruses; C-P-NProp: circulative-persistent non-propagative viruses; NC-SPer: non-circulative semipersistent viruses; NC-NPer: non-circulative nonpersistent viruses. With the permission of Mauck et al. (2018). Evolutionary determinants of host and vector manipulation by plant viruses. *Advances in Virus Research* 101, 189–250.

Modern authors characterize the non-circulative mode according to the retention strategy of virus particles in the vector. In the 'capsid strategy' the virion is directly linked to cuticular sites in the vector, whereas the 'helper strategy' requires one or two extra-capsid peptides building a 'molecular bridge' between cuticular sites and virions. For both strategies, the existence of virus receptors in the vector, explaining the virus-vector specificity, has been postulated and in a few cases demonstrated. Remarkably, a helper factor has also been shown to be required in two circulative viruses, Faba bean necrotic stunt virus (FBNSV, *Nanovirus*, *Nanoviridae*) for its transfer within the aphid's hemocoel, and Rice stripe virus (RSV, *Tenuivirus*, *Phenuiviridae*) for its endocytosis into gut cells.

Capsid Strategy

Case of Aphid-Transmitted Cucumoviruses

This strategy is illustrated by the interaction between cucumoviruses and aphid vectors. The type-member Cucumber mosaic virus (CMV, *Cucumovirus*, *Bromoviridae*) is easily transmitted by aphids fed on purified virions through a stretched Parafilm™ membrane. This clearly demonstrates that virions alone are transmissible, without need for any other compound, indicating that the virion structure bears the determinant required for attachment to the cuticula. The viral retention site is likely in the stylet alimentary and/or common canals. Moreover, the three RNAs and the coat protein can be separated *in vitro*, then reassembled; thus, exchanging the components of transmissible and non-transmissible isolates confirmed the implication of the sole coat protein in transmission. Sequence modifications in this protein have been reported to modify vector specificity. Moreover, the vector-less Tobacco mosaic virus (*Tobamovirus*, *Virgaviridae*) can be aphid-transmitted when its RNA molecule is encapsidated within the capsid of a cucumovirus.

The same strategy operates in the aphid transmission of alfamoviruses and carlaviruses, as well as probably for whitefly-transmitted criniviruses; however, very few is known so far for other closterovirids.

Helper Strategy

Case of Aphid-Transmitted Potyviruses

This strategy was first described for members of the genus *Potyvirus* (*Potyviridae*). In this species-rich genus, most species are aphid-transmitted, many of them are also transmissible mechanically and some by seed. In terms of acquisition, retention, and inoculation times, potyviruses are transmitted similarly to cucumoviruses. However, purified potyvirus particles are not transmissible by aphids fed through a Parafilm™ membrane, albeit they are mechanically transmitted, showing that vector transmission requires an additional factor which is lost during purification process. Subsequent experiments proved that this factor is a virus-encoded nonstructural peptide, called helper factor or helper component (HC), released in infected cells and mediating the attachment of virions to cuticular sites during acquisition. For an efficient transmission, the HC must be taken up by the vector before or during virus acquisition, but not after. Comparison of transmissible and non-transmissible isolates or species, and mutagenesis and microscopy experiments revealed the role of the peptidic motifs, borne by the coat protein and the HC, in building a 'molecular bridge' between virion, HC and stylet cuticula. The HC is commonly designated as 'HCPro' in literature, since it also has a protease function. The precise retention site of the HC-virion complex in the vector's stylets likely comprises one or more cuticular protein(s).

Trans-complementation is another feature of the helper strategy of potyviruses. The HC of a potyvirus species can interact with the capsid of a related species acquired on the same plant or another one, due to sequence homologies, and thus assist its transmission; thus, this common phenomenon advantages the population genetics and evolution of potyvirus species.

Case of Aphid-Transmitted Caulimoviruses

Caulimoviruses (*Caulimoviridae*) are transmitted by aphids according to a particular helper strategy, that is best studied for Cauliflower mosaic virus (CaMV). Like for potyviruses, purified CaMV particles are not transmissible by artificially-fed aphids, indicating the requirement of an additional factor, later demonstrated to be the nonstructural transmission helper protein P2 expressed from the open reading frame (ORF) II of viral DNA. Moreover, a second helper peptide, the product P3 expressed from the ORF III and associated to the viral capsid, was proven indispensable in the transmission process. During cell sap ingestion, virus receptor(s) harbored on the stylets' inner cuticula retain either the complete P2-P3-virion complex simultaneously, or sequentially with P2 binding first to the stylets, followed by subsequent attachment of the P3-virion complex. In infected plant cells, CaMV virions accumulate and concentrate in specific inclusions called virus factories, whereas P2 and P3 accumulate in separate inclusions named transmission bodies. When the aphid stylets enter such cells, the transmission bodies are instantly (≤ 5 s) disrupted, and the freed P2 relocates to the cellular microtubules. Simultaneously, the virus factories release virions that join P2 on the microtubules and form transmissible virions that can be acquired efficiently by the vector. This remarkable phenomenon, coined 'transmission activation', is reversible after the stylet withdrawal. A similar reaction, i.e., vector-induced formation of transmissible virions, has recently been described for *Turnip mosaic virus*, a potyvirus, whose transmission also relies on a helper strategy.

Microscopy studies demonstrated that CaMV virions are retained at the tip of the aphid's stylet bundle, i.e., the common duct where the food and salivary canals converge, and more precisely at a particular part in the inner side of the maxillary stylets called acrostyle. Among cuticular proteins embedded in this chitinous organ, two proteins called stylin-01 and stylin-02, largely conserved among aphid species, have recently been shown to interact with either P2 or the transmissible complex P2-P3-virion, and are thus good candidate receptors for CaMV and possibly other aphid-transmitted non-circulative phytoviruses.

Transmission by Beetles

Vectors of plant viruses are known in phytophagous beetles within the Chrysomelid and Coccinellid families (Fig. 1(H)), and for members of the genera *Bromovirus* (Bromoviridae), *Comovirus* (Secoviridae), *Machlomovirus* (Tombusviridae), *Sobemovirus* (Solemoviridae), and *Tymovirus* (Tymoviridae). These viruses share isometric, stable and also mechanically transmissible particles. Virions are rapidly acquired when the vector's biting-chewing mouthparts damage plant tissues. The interaction, still poorly described, is likely ingestion-regurgitation and specific. Phytophagous beetles are devoid of salivary glands and coat their mandibles with an RNase-rich food regurgitate. The viruses transmitted display a certain resistance to RNases that may account for their vector specificity, and they are thought to translocate rapidly to unwounded cells and/or vascular tissues. In some cases, virions are detectable in the insect hemolymph within seconds, though whether this feature refers to a circulative mode or not is unknown. Some vectors can efficiently infect a plant after 'drinking' a purified virion suspension, indicating that the mere virion is required for transmission.

Transmission by Non-Insect Vectors

The transmission of several plant viruses relies on non-insect vectors (Table 1), such as mites, nematodes, and 'fungus-like' organisms (i.e., true fungi and protists).

Transmission by Mites

Two families of mites (*Acari*), *Eriophyidae* (gall mites), and *Tenuipalpidae* (false spider mites, flat mites), contain vectors of members of several virus genera (Table 1, Fig. 3(A)). However, these minute vectors have been less studied than insects, and the underlying mechanisms are poorly known, and only in some cases. For feeding, phytophagous mites empty plant epidermal cells, using a short stylet bundle. For *Brevipalpus* spp. (*Tenuipalpidae*) and *Citrus leprosis virus* (*Cilevirus*, *Kitaviridae*), virions were detected in midgut cells. Virions of the eriophyid (*Aceria tosichella*)-transmitted Wheat streak mosaic virus (WSMV, *Tritimovirus*, *Potyviridae*) were also detected in hemolymph and salivary glands. Whereas the transmission mode was long considered as non-circulative and semi-persistent, the observations mentioned above rather suggest a circulative (propagative?) interaction, but experimental proofs are lacking so far. Considering that tritimoviruses belong to the same family as aphid-transmitted potyviruses and possess an HCPro homolog ORF, experiments showed that HCPro of WSMV bears motifs required for transmissibility, even though the transmission modes (non-circulative for potyviruses) likely differ substantially between the two genera.

Transmission by Nematodes

Members of *Nepovirus* and *Cheravirus* in the family Secoviridae, and of *Tobravirus* (*Virgaviridae*) are transmitted by, respectively, longidorid and trichodorid nematodes. These nematodes feed on root cells using stylet-like organs called odontostyle and onchiostyle, respectively. The mode of transmission is described to be non-circulative, as virions are mainly adsorbed at the inner lining of the esophagus, at still unknown receptors. Despite the loss of virions during molting, *Xiphinema* spp. longidorids (Fig. 3(B)) harboring Grapevine fanleaf virus (GFLV, *Nepovirus*, Secoviridae) can remain viruliferous after at least 4 years rest in soil; thus, these slow-moving vectors provide the virus with a travel in time, rather than in space. The specific transmission of GFLV and probably other nepoviruses requires the sole capsid protein (capsid strategy), whereas that of tobraviruses needs one or two non-structural proteins (helper strategy). The viral determinants of GFLV are located around a canyon on the capsid surface and account for its vector specificity.

Transmission by Fungi and Protists

These soil-dwelling unicellular organisms, either chytrids (zoosporic Fungi or plasmidiophorid protists (*Cercozoa*) (Figs. 3(C) and (D)), have flagellate forms, called zoospores, that can enter the cytoplasm of root cells and from there colonize the plant. While doing so, these species may inoculate viruses of several genera (Table 1). Several mechanisms seemingly exist. The best documented associations are the chytrids *Olpidium* spp. transmitting tombusvirids and ophioviruses, and the plasmidiophorids *Polymyxa* spp. vectors of benyviruses and furoviruses. In the first association, during zoospore formation in infected host plants, virions bind to the zoospore external coat using fungal glycoproteins and specific regions in the capsid protein (homologous to non-circulative capsid strategy), leading to a swelling of virions. They are then released into the newly infested host cell upon cell wall digestion and fungal penetration. In the case of the protists *Polymyxa* spp. (Fig. 3(C)), the virions, or at least viral RNA-protein complexes, are endocytosed in the zoospore, presumably without replication (homologous to circulative non-propagative mode) and then inoculated into root cells with the zoospore cytoplasm. A 32 kDa viral protein was shown to be required for the transmission of a benyvirus by *Polymyxa* sp. Precise mechanisms of acquisition and inoculation are poorly known.

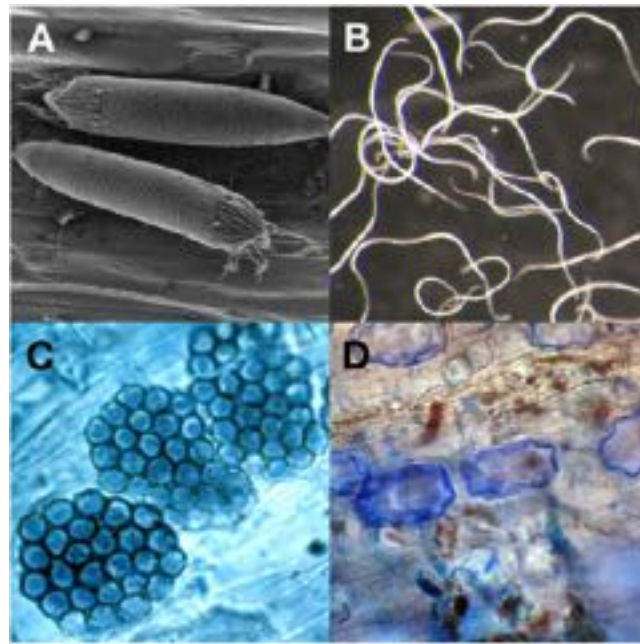


Fig. 3 Main non-insect vectors of plant viruses. (A) *Aceria tosichella* (Acari Eriophyidae); (B) *Xiphinema index* (Nematoda Longidoridae); (C) *Polymyxa betae* (Cercozoa Plasmidiophorida); (D) *Olpidium* sp. (Fungi Chytridiomycota). Photographs: (A) Courtesy of Anna Skoracka, Wielkopolska Centre for Advanced Technologies, Adam Mickiewicz University, Poznan, Poland; (B) Courtesy of Gérard Demangeat, INRAE Colmar France; (C) Courtesy of Olivier Lemaire, INRAE Colmar France; (D) Courtesy of Dominique Blancard, INRAE Bordeaux France.

Host and Vector ‘Manipulation’ by Plant Viruses

Plant virus infection can drastically alter host’s phenotype, including size, color, volatiles emission, primary and secondary metabolites composition. Historically, most studies have focused on exploring the effects of viruses on plants from an agronomic point of view, such as impacts on crop yield or quality. In recent years, it has been highlighted that virus-induced changes in plant physiology may also influence host–vector interactions and, thereby, the probability of virus transmission. Given that this crucial step of transmission obligately depends on vectors for many plant viruses, it has been proposed that vector-borne viruses evolved adaptive traits that induce host phenotypes, and effects on vectors that promote their own spread. Since then, numerous studies showed that virus infection in host plants modifies vector behavior, such as orientation and settling preferences, but also reproduction and survival, in ways that are generally conducive to transmission. These effects fall into the category of putative ‘indirect manipulations’, because, in these cases, the viruses modify the behavior and/or performance of their insect vector through phenotypic alterations of the shared host. Another category of effects, qualified as ‘direct’, corresponds in changes within the viruliferous vector’s orientation and feeding preferences.

In the early 1950s, it was reported for the first time a positive effect of plant infection on the fitness of an aphid vector. Thereafter, indirect effects of plant viruses on vectors, linked to induced volatiles and to changes in leaf attractiveness and suitability for the vector, have been reported with aphid vectors, and were shown to vary according to the transmission mode of the virus. Over the last two decades, these observations were substantially confirmed and documented by many published studies with aphids and other insect vectors (whiteflies, leafhoppers, thrips, etc.). Indeed, vector-borne viruses can be classified in different categories based on their associations with the host plant and the vector ([Table 1](#)). Some viruses are restricted to the phloem tissue of their host plant, virions are thus only acquired and inoculated by vectors achieving sustained phloem sap ingestion (i.e., during hours). In contrast, viruses that are not restricted to the phloem are acquired and inoculated most efficiently during brief epidermal or mesophyll intracellular probes and are generally retained within the vector only for seconds or minutes. In this case, if the vector initiates ingestion of phloem sap prior dispersal, virus particles might be lost, and viral transmission impeded. Consequently, short probing followed by rapid dispersal is the most-conducive vector behavior for non-phloem restricted virus transmission.

From this classification, it was predicted that viruses sharing the same transmission modes – and thus, benefiting from a similar sequence of attraction, vector probing, feeding, and dispersal behavior – should exhibit convergent indirect host-mediated effects on vectors. Indeed, most phloem-restricted viruses improve the nutritive quality of the host plant for vectors (i.e., improved survival and/or fecundity), which is expected in turn to lead to crowding and subsequent dispersal of viruliferous vectors. These viruses also increase the palatability of the infected host plant compared to healthy ones, which results in improved vector feeding behavior and retention, facilitating virion acquisition from the phloem tissue. For the non-phloem restricted viruses, infections often reduce plant quality for vectors, and tend to reduce the retention of vectors on infected plants, thus potentially promoting

the rapid dispersion of viruliferous vectors to healthy plants. Regardless of the type of virus, the vectors are preferentially attracted by the visual and volatile cues of infected plants over healthy plants, which is expected to promote virus spread.

Only a few studies have documented cases of direct effects of plant viruses on vector behavior and, for the moment, the mechanisms are poorly characterized. Phloem-restricted viruses with persistent-circulative transmission mode need to perform a route from the gut to the salivary glands of their vector before they can be inoculated into a new host. These intimate associations offer strong opportunities for direct interactions of viruses on the physiology of their vector. For this kind of virus, vector dispersal is favored only after sufficient virions have been acquired by the insect to become infective. Therefore, it is predicted that these viruses may have evolved mechanisms to encourage dispersal following sustained phloem sap ingestion on infected hosts, and to discourage viruliferous vectors from visiting infected hosts; these changes in vector preference are named 'conditional vector preferences'. This hypothesis was verified on aphids carrying two luteovirids, settling preferentially on uninfected plants, while non-viruliferous aphids preferred the infected plant. The same effect was observed with aphids having acquired virions from an artificial medium (thus overcoming any potential indirect plant-mediated effect). Some examples have also documented changes in viruliferous vectors feeding behavior, such as alteration of phases related to virus transmission (i.e., salivation, phloem sap intake, sustained ingestion of phloem sap), most of which are expected to enhance the probability of virus transmission.

Future Prospects

From a mechanistic point of view, basic information on the virus-vector interaction process is increasingly documented in several cases, but is lacking in many other cases. Further studies in various transmission modes will provide a better understanding of the complex virus-vector-plant interactions, including the viral 'manipulation' of the two other partners favoring disease spread, and the localization and identification of vector receptors used by the virus for its transmission. Novel knowledge is based on ongoing progress in high-throughput data analysis and microscopic technologies.

From an evolutionary point of view, the vector provides an obvious bottleneck for the virus, since only some genome variants encapsidated in transmissible virions will be transferred into a new host plant. This issue has started to be documented but requires more information to fully understand its implications.

From an agricultural point of view, vector-borne viruses represent a burden for most crops worldwide. Since virus-infected plants remain incurable, phyto-protection against viruses mainly relies on (1) use of resistant varieties, provided resistant genes are available, although resistance-breaking viral strains are susceptible to be selected, and (2) insecticide, or acaricide and nematicide, treatments directed against the vector, which have potential environmental drawbacks and can also lead to the selection of resistant vector populations. In case of perennial crops or seed-transmitted viruses, using virus-free material is also important, even though it does not prevent further contamination. Deeper knowledge of the transmission process and the viral epidemiology is expected to generate new strategies to protect plants from viruses using vector-targeted, and more environment-friendly and integrated practices. This includes the use of molecular and/or behavioral disruption of the virus-vector-plant interaction, and the adaptation of crop management to climate change.

See also: Betaflexiviruses (*Betaflexiviridae*). Bluner-, Cile-, and Higreviruses (*Kitaviridae*). Bromoviruses (*Bromoviridae*). Caulimoviruses (*Caulimoviridae*). Closteroviruses (*Closteroviridae*). Cucumber Mosaic Virus (*Bromoviridae*). Geminiviruses (*Geminiviridae*). Luteoviruses (*Luteoviridae*). Nanoviruses (*Nanoviridae*). Orthotospoviruses (*Tospoviridae*). Plant Reoviruses (*Reoviridae*). Plant Rhabdoviruses (*Rhabdoviridae*). Potyviruses (*Potyviridae*). Secoviruses (*Secoviridae*). Tenuiviruses (*Phenuiviridae*). Tobamoviruses (*Virgaviridae*). Tombusviruses (*Tombusviridae*). Tymoviruses (*Tymoviridae*). Virgaviruses (*Virgaviridae*).

Further Reading

- Blanc, S., Drucker, M., Uzzell, M., 2014. Localizing viruses in their insect vectors. *Annual Review of Phytopathology* 52, 403–425.
- Bragard, C., Caciagli, P., Lemaire, O., *et al.*, 2013. Status and prospects of plant virus control through interference with vector transmission. *Annual Review of Phytopathology* 51, 177–201.
- Brown, J.K., 2016. Vector-mediated Transmission of Plant Pathogens. St Paul, MN: American Phytopathological Society Press.
- Eigenbrode, S.D., Bosque-Pérez, N.A., Davis, T.S., 2018. Insect-borne plant pathogens and their vectors: Ecology, evolution, and complex interactions. *Annual Review of Entomology* 63, 169–191.
- Hogenhout, S.A., Ammar, E.D., Whitfield, A.E., Redinbaugh, M.G., 2008. Insect vector interactions with persistently transmitted viruses. *Annual Review of Phytopathology* 46, 327–359.
- Hull, R., 2014. Plant to plant movement. In: Hull, R. (Ed.), *Plant Virology*, fifth ed. Amsterdam: Academic Press, pp. 669–751.
- Mauck, K.E., Chesnais, Q., Shapiro, L.R., 2018. Evolutionary determinants of host and vector manipulation by plant viruses. *Advances in Virus Research* 101, 189–250.
- Wilson, J.R., DeBlasio, S.L., Alexander, M.M., Heck, M., 2019. Looking through the lens of 'Omics technologies: Insights into the transmission of insect vector-borne plant viruses. In: Bonning, B.C. (Ed.), *Insect Molecular Virology: Advances and Emerging Trends*. Poole: Caister Academic Press, pp. 113–144.
- Zhou, J.S., Drucker, M., Ng, J.C.K., 2018. Direct and indirect influences of virus-insect vector-plant interactions on non-circulative, semi-persistent virus transmission. *Current Opinion in Virology* 33, 129–136.

Viral Suppressors of Gene Silencing

Hernan Garcia-Ruiz, University of Nebraska–Lincoln, Lincoln, NE, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J. Verchot-Lubicz, J.P. Carr, Viral Suppressors of Gene Silencing, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00718-4](https://doi.org/10.1016/B978-012374410-4.00718-4).

Glossary

Antiviral immunity Is a group of mechanisms that protect cells from virus infection. They target viral proteins or nucleic acids.

Antiviral gene silencing Is the downregulation of viral RNA accumulation. It targets viral RNA for degradation or translational repression. Additionally, in DNA viruses, it prevents transcription.

Gene silencing Is a sequence-specific gene inactivation system mediated by small interfering RNAs (siRNA). It downregulates RNA accumulation without affecting DNA sequence. It is also known as RNA interference.

Pathogenicity determinants Are viral components, usually proteins, that promote virus infection in cells and cause developmental defects in infected organisms.

Post-transcriptional gene silencing Is the degradation or translational repression of mRNA. It targets both cellular and viral RNA.

Small interfering RNAs (siRNAs) Are double-strand RNAs that measure 21–24 nt in length. siRNAs are the specificity determinant of gene silencing and program argonaute proteins for specific degradation of RNA with sequence complementarity to the siRNA.

Suppressors of gene silencing Are proteins encoded by viruses that interfere with gene silencing. Their activity prevents degradation of viral RNA, enhance transcription of viral RNA, and alters silencing of cellular genes. These effects are in part responsible for developmental defects in virus infected organisms.

Transcriptional gene silencing Prevent inhibits transcription from chromosomes, plasmids or viral DNA mini chromosomes. It is directed by siRNAs guiding enzymatic DNA methylation.

Virus-induced gene silencing Virus infection induces silencing of cellular genes with sequence similarity to viral RNA.

Introduction

Gene silencing is a sequence-specific gene inactivation system mediated by small interfering RNAs (siRNA) that measure 21–24 nt. Gene silencing downregulates RNA accumulation without affecting DNA sequence and is also known as RNA interference. Gene silencing regulates gene expression at the transcriptional or post-transcriptional levels. Transcriptional gene silencing is directed by siRNAs guiding enzymatic DNA methylation and inhibits transcription from chromosomes, plasmids or viral DNA mini chromosomes. Post-transcriptional gene silencing is the nucleolytic cleavage and degradation or translational repression of mRNA (**Fig. 1**). The pathway and biological roles of gene silencing are conserved in eukaryotes.

Gene silencing regulates gene expression in a temporal and tissue-specific manner to control development, response to stress, and to preserve genome integrity. Additionally, in plants, nematodes, and insects, gene silencing is an essential component of antiviral immunity. In these organisms, plant viruses are inducers and suppressors of gene silencing. Our mechanistic understanding of gene silencing and virus-encoded gene silencing suppressors is derived mainly from the study of plant viruses. Accordingly, this article is focused on suppressors of gene silencing encoded by plant viruses.

Viruses use RNA to replicate, express their genes, or both. The RNA might be perceived by the cell and activates antiviral defense, including gene silencing and mRNA decay. DNA viruses are targeted by both transcriptional and post-transcriptional gene silencing. RNA viruses, viroids and satellite RNAs are targeted by post-transcriptional gene silencing. To overcome gene silencing and promote susceptibility, viruses encode suppressors of gene silencing that are essential for viruses to replicate and spread. Virus-encoded gene silencing suppressors interfere with antiviral defense, and have been identified in plant viruses, insect viruses, fungal viruses and in some viruses infecting humans. However, the antiviral role of gene silencing in mammals is only beginning to be elucidated.

In plant-virus interactions, the establishment of infection is determined by the balance between antiviral gene silencing versus virus evasion or suppression of antiviral defense. Endogenous and antiviral gene silencing have conserved and overlapping pathways. Accordingly, virus-encoded silencing suppressors affect both endogenous and antiviral gene silencing and their effects are in part responsible for symptom development in infected plants.

Antiviral Gene Silencing Pathway

Antiviral gene silencing is non-cell autonomous, initiates at the single cell level and spreads cell-to-cell and long distance through plasmodesmata and the plant vascular system. The pathway consists of four parts: Initiation, targeting, amplification, and systemic spread (**Fig. 2**). Upon initial detection, Dicer-like (DCL) proteins process viral dsRNA into virus-derived siRNA that associate with

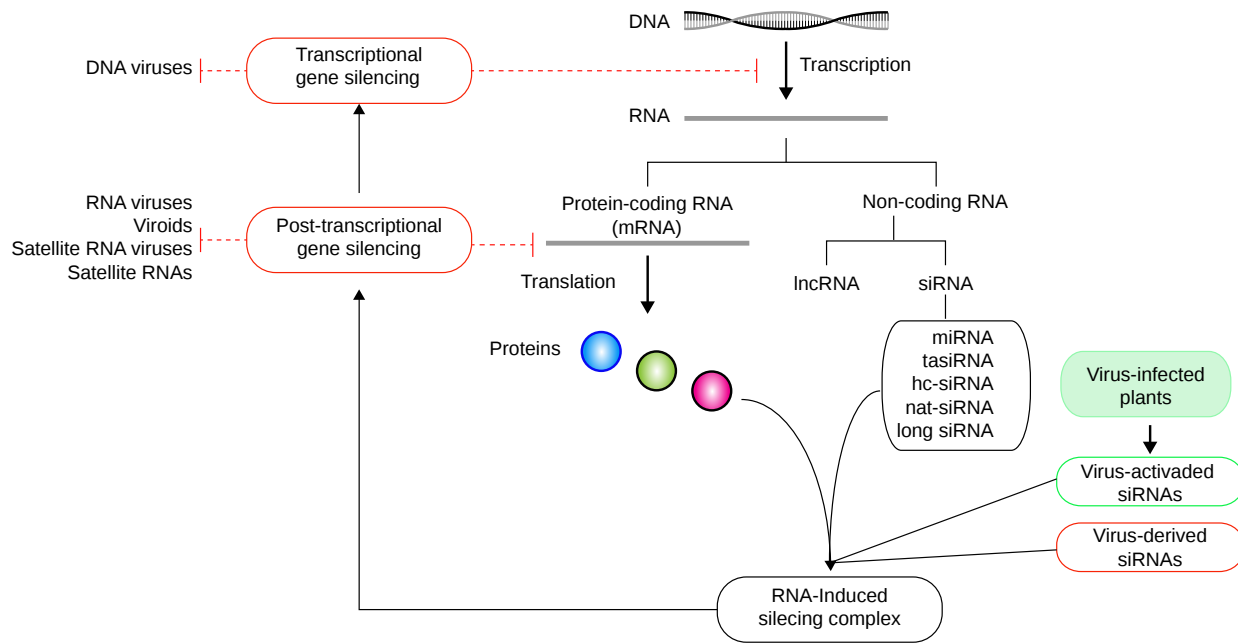


Fig. 1 Flow of genetic information and gene silencing in plants. RNA-induced silencing complexes are formed between effector proteins and small interfering RNAs (siRNAs) derived from non-coding transcripts. Gene silencing is a block in the flow of genetic information by preventing transcription (transcriptional gene silencing) or by inducing degradation or translational repression of protein-coding RNAs (post-transcriptional gene silencing). In addition to endogenous long non-coding RNAs and siRNAs, virus-infected plants accumulate virus-induced siRNAs and virus-derived siRNAs, which enter the RNA-induced silencing complex and program degradation of host or viral mRNA. Gene silencing is essential for normal development and antiviral immunity. It targets viruses at the transcriptional or post-transcriptional level. Adapted from Garcia-Ruiz, H., Ruiz, M.T.G., Peralta, S.M.G., Gabriel, C.B.M., El-Mounadi, K., 2016. Mechanisms, applications, and perspectives of antiviral RNA silencing in plants. *Mexican Journal of Phytopathology* 34, 286–307.

argonaute (AGO) proteins to form RNA-induced silencing complexes (RISC) that target viral mRNA, and potentially host mRNAs, for endonucleolytic cleavage or translational repression. Initial recognition of viral RNA is necessary but not sufficient to restrict plant virus infection. Restriction of infection requires silencing amplification by cellular RNA-dependent RNA polymerases (RDRs). After moving out of the infected cell, in recipient cells the silencing response is amplified. Cellular RDRs synthesize new viral dsRNA that is the substrate for the formation of secondary virus-derived siRNAs that move systemically ahead of the virus to establish an antiviral state.

DNA viruses express their genes through mRNA. Silencing is activated at the post-transcriptional level, resulting in the formation of virus-derived siRNAs that guide targeting of viral RNA. Furthermore, virus-derived siRNAs in association of AGO and RDR proteins guide methylation of viral DNA to prevent mRNA transcription (transcriptional gene silencing). Combined, the effects of both transcriptional and post-transcriptional gene silencing create an antiviral state that prevents virus replication and movement.

Small Interfering RNAs Direct Antiviral Immunity

Through their association with AGO proteins, virus-derived siRNAs are the specificity determinant guiding translational repression or cleavage of viral RNA with sequence complementarity. In a variation of the pathway targeting DNA viruses, silencing complexes formed by virus-derived siRNAs and AGO proteins direct methylation of viral DNA to prevent transcription. In both cases, the end result is the inhibition of virus replication and movement. In plant-virus interactions, several phenomena are mediated by gene silencing such as cross protection, symptom recovery, synergism, and non-host resistance.

Cross protection occurs when infection by a mild virus strain prevents subsequent infection by a second virus of the same or closely related species. siRNAs-derived from the mild virus strain direct targeting of viruses with sequence similarity. However, not all cross-protection cases are explained by gene silencing. In *Turnip crinkle virus*, and *Wheat streak mosaic virus*, suppressor infection exclusion is a form of cross protection mediated by viral proteins.

Symptom recovery has been described for several plant-virus combinations, such as *Arabidopsis thaliana* infected by *Oilseed rape mosaic virus* or *Tobacco rattle virus*. Infected plants accumulate high virus titers and develop visible symptoms. However, the upper leaves of those plants contain low virus titers and do not develop symptoms. Virus-derived siRNAs formed in the lower part of the plant move systemically to the meristem to direct viral RNA targeting and prevent virus invasion of the meristem. Leaves formed after that harbor virus-derived siRNAs that mediate virus downregulation. Interestingly, in recovered *A. thaliana* tissue, the suppressor from *Oilseed rape mosaic virus* lost its activity.

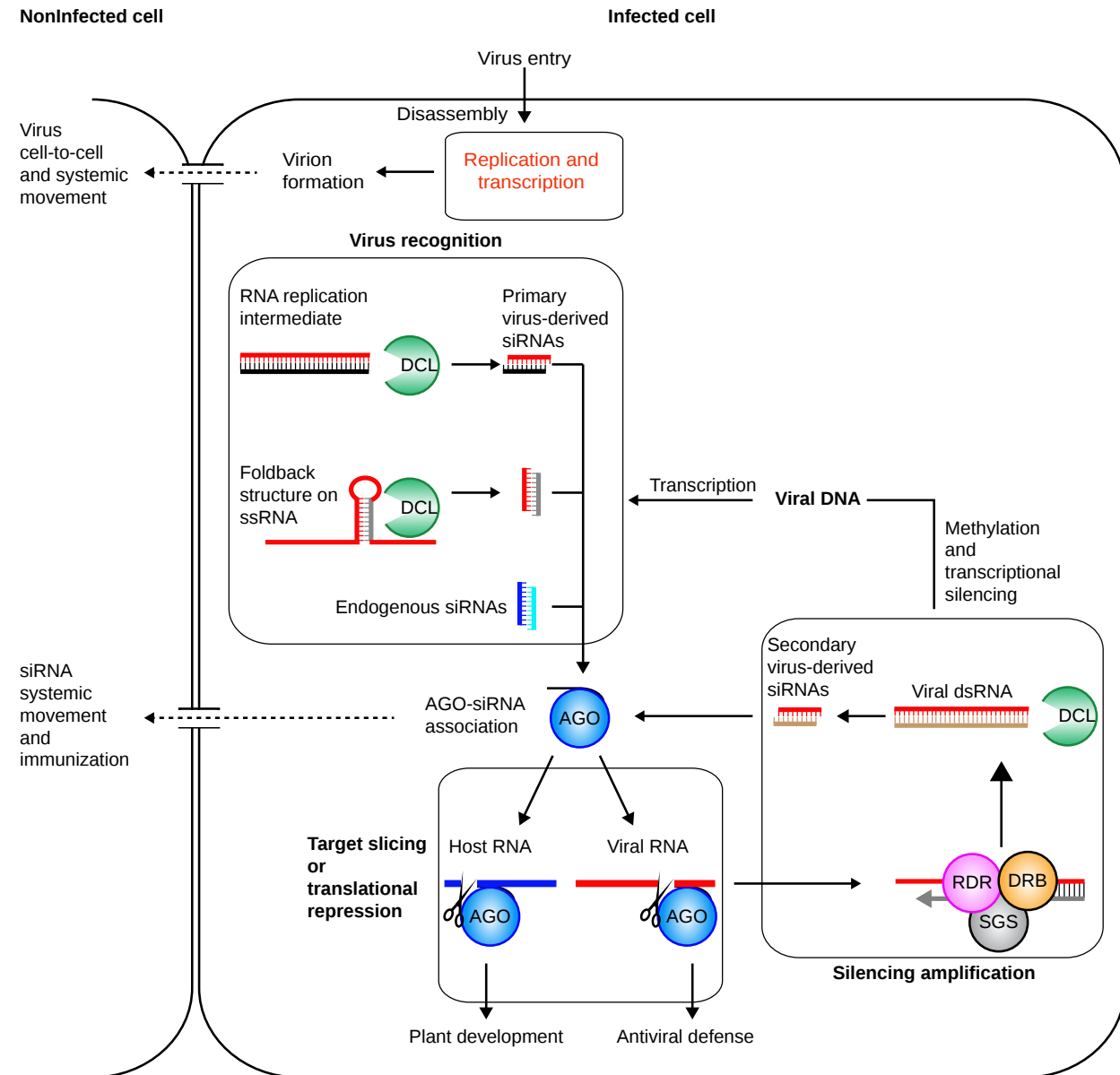


Fig. 2 Antiviral gene silencing and basic components in plants. In infected cells, viruses are recognized by dsRNA replication intermediates or self-complementary sequences in genomic RNA or mRNA. Dicer-like (DCL) proteins cut viral dsRNA to form primary virus-derived small interfering RNAs (siRNAs), that associate with argonaute (AGO) proteins and program endonucleolytic slicing and degradation or translational repression of viral RNA. Host transcripts with sequence complementarity to virus-derived siRNAs might be targeted. Slicing of viral RNA triggers amplification of antiviral RNA silencing that results in the formation of viral dsRNA by cellular RNA-dependent RNA polymerases (RDR) and processing by DCL to form secondary-virus-derived siRNAs. Silencing amplification triggers methylation of viral DNA which results in transcriptions silencing. Virus-derived siRNA move cell-to-cell and systemically to establish a state of antiviral immunity away from the initial infection site. Adapted from Garcia-Ruiz, H., Ruiz, M.T.G., Peralta, S.M.G., Gabriel, C.B.M., El-Mounadi, K., 2016. Mechanisms, applications, and perspectives of antiviral RNA silencing in plants. Mexican Journal of Phytopathology 34, 286–307.

Several mechanisms explain non-host resistance, such as the absence of pro-viral factors needed for replication or movement. However, antiviral gene silencing early in infection guided by endogenous siRNAs or virus activated siRNAs might target viral RNA to prevent the establishment of infection and plants show the phenotype of a non-host.

Viral Suppressors of Gene Silencing Promote Infection

Gene silencing is an essential component of antiviral immunity in plants, insects, and nematodes. To establish infection, viruses must protect themselves from antiviral defense responses such as RNA decay, gene silencing and protein-based

mechanism such as autophagy or ubiquitination. At least two mechanisms have been proposed to explain gene silencing: evasion and suppression. Positive-strand RNA viruses replicate in membrane bound compartments that sequester replication intermediates. Negative-strand RNA viruses and dsRNA viruses replicate in enveloped vesicles. These structures might prevent access to viral dsRNA by the gene silencing machinery. In an interesting variation of gene silencing, a non-coding region in *Cauliflower mosaic virus* produces large amounts of siRNAs that saturate the silencing machinery. This decoy role prevents targeting of viral RNA.

Mechanisms of Gene Silencing Suppression

To protect themselves from gene silencing and promote infection, most plant and some insect viruses encode silencing suppressors that inactivate gene silencing by multiple mechanisms (Table 1) discussed below.

Table 1 Representative virus-encoded gene silencing suppressors

<i>Virus</i>	<i>Suppressor</i>	<i>Silencing target</i>	<i>Other functions</i>
siRNA binding			
Potyruses	HC-Pro	siRNA monomers and dimers	Virus movement, symptom development, pathogenicity determinant
TBSV	P19	siRNA dimers	Virus movement, symptom development pathogenicity determinant
BYV	P21	siRNA dimers	Required for replication
TSWV	NSs	siRNAs	Pathogenicity determinant
WDV	Rep	siRNAs	Viral DNA replication
RGDV	PNS11	siRNAs	Pathogenicity determinant
ACMV	AC4	Single strand siRNAs	Viral synergism
Long dsRNA binding			
PoLV	P14	dsRNA and siRNAs	Symptom determinant
TCV	P38	dsRNA	Virion formation Virus movement
FHV	B2	dsRNA	NA
AGO degradation or inhibition			
CMV	2b	AGO1 AGO4	Virus movement
PVX	P25	AGO1	Virus movement
ToRSV	CP	AGO1	Virion formation
Polerovirus	P0	AGO1	Pathogenicity determinant
TCV	P38	AGO1 AGO2	Virion formation Virus movement
DCL inhibition			
RYMV	P1	DCL1 upregulation, DCL4	Pathogenesis, symptom development
TCV	P38	DCL1 upregulation, DCL4	Virion formation Virus movement
Silencing amplification inhibition			
RDV	PNS10	RDR6 downregulation	ND
TuMV PVA	VPg	RDR6 and SGS3 downregulation	RNA translation
CaMV	P6 (TAV)	DRB4 inactivation	Cell-to-cell movement, NPR1 downregulation, translational activator
RGDV	PNS12	NA	
Interference with siRNA methylation			
TMV	129K	HEN1	RNA replication
Tombusvirus	P19	siRNAs	Virus movement, symptom development
DNA methylation interference			
TYMV	P69	NA	Symptom determinant
TGMV BCTV	AL2 L2	Adenosine kinase inhibition	Pathogenesis
CaLCuV	AL2		
Activation of cellular transcription			
EACMV	AC2	Cellular transcription	Viral synergism
CaLCuV	AL2	Cellular transcription	Transcription activation
Decoy RNA			
CaMV	35S RNA translational leader	Overloading DCL	Translation, reverse transcription template

Viruses: African cassava mosaic virus (ACMV), East African cassava mosaic virus (EACMV), *Beet curly top virus* (BCTV), *Beet yellows virus* (BYV), *Cabbage leaf curl virus* (CaLCuV), *Cauliflower mosaic virus* (CaMV), *Cucumber mosaic virus* (CMV), *Flock house virus* (FHV), *Potato virus A* (PVA), *Potato virus X* (PVX), *Pothos latent virus* (PoLV), *Rice gall dwarf virus* (RGDV), *Rice dwarf phytoecovirus* (RDV), *Rice yellow mottle virus* (RYMV), *Tobacco mosaic virus* (TMV), *Tomato bushy stunt virus* (TBSV), *Tomato golden mosaic virus* (TGMV), *Tomato ringspot virus* (ToRSV), *Tomato spotted wilt virus* (TSWV), *Turnip crinkle virus* (TCV), *Turnip mosaic virus* (TuMV), *Turnip yellow mosaic virus* (TYMV), *Wheat dwarf virus* (WDV).

Small RNA Binding

Potyviral HC-Pro, tombusviral p19, and others, bind siRNAs of viral and cellular origin including microRNAs and secondary siRNAs. Binding by silencing suppressors prevents association of siRNAs with AGO proteins. Consequently, RISC complexes are not formed and viral RNA, and host mRNAs, are not targeted. As a consequence, host genes normally downregulated by siRNAs, accumulate to levels higher than normal and plants display malformation that resemble symptoms caused by virus infection.

Inhibition of siRNA Biogenesis

Several virus-encoded gene silencing suppressors interfere with the biogenesis of siRNA by at least two mechanisms: binding long dsRNA and altering homeostasis of core genetic components of the pathway (**Table 1**). Binding long dsRNA prevents access to DCL proteins and inhibits siRNA formation. Some suppressors induce degradation of cellular components needed for amplification of gene silencing, such as RDR6, SGS3 or DRB4. In both cases, the end result is a reduction in the biogenesis of virus-derived and endogenous siRNAs.

Homeostasis of Gene Silencing Components

Core components of gene silencing (DCL, AGO, RDR, and DRB proteins) are targeted for degradation by some virus-encoded gene silencing suppressors (**Table 1**). *Cucumber mosaic virus* 2b and polerovirus P0 induced degradation of AGO1. AGO1 mainly associate with microRNAs and is essential for normal plant development. AGO2 and AGO4, are also needed for plant development and are targeted for degradation by silencing suppressors.

Turnip crinkle virus P38 inhibits activity of DCL4, the first line of defense against viral RNA, and prevents biogenesis of DCL4-dependent siRNAs. Furthermore, DCL1 downregulates expression of DCL4 and DCL2, which have essential roles in siRNA biogenesis in both endogenous and antiviral gene silencing. Several viruses upregulate DCL1, consequently downregulating DCL4 and DCL2 to protect viral RNA from silencing.

Inhibition of Assembly or Activity of RNA-Induced Silencing Complexes

The catalytic components of RNA-induced silencing complexes (RISC) are binary complexes formed between AGO proteins and siRNAs. Several virus-encoded silencing suppressors inhibit assembly or activity of RISC complexes by binding siRNAs or targeting AGO proteins for degradation (**Table 1**). Interestingly, in a complementary mechanism 2b, the silencing suppressor encoded by *Cucumber mosaic virus* and related species, binds to AGO1 and prevents cleavage of target mRNA.

Inhibition of Silencing Amplification

A critical step in antiviral gene silencing is the systemic spread and amplification of silencing signals to establish a state of antiviral immunity away from the initially infected tissues (**Fig. 2**). The silencing signal consists of siRNAs, very likely in endosomes. After moving out of the initially infected cell, silencing is amplification by cellular RNA-dependent RNA polymerases (RDRs) and auxiliary proteins such as suppressor of gene silencing (SGS3) and double stranded RNA binding protein 4 (DRB4). Silencing amplification consists on the synthesis of new viral dsRNA that is processed by DCL proteins into secondary virus-derived siRNAs that move ahead of the virus to establish an antiviral state. Some virus-encoded gene silencing suppressors prevent spread or amplification of the gene silencing signal (**Table 1**). The mechanisms are dependent on siRNA sequestration, or induction of degradation of key components of secondary siRNA biogenesis, such as RDR6, SGS3 or DRB4.

Interference With DNA Methylation

Plants silence geminiviruses by methylating their mini chromosomes through transcriptional silencing. As counter-defense, some geminiviruses encode suppressors that interfere with DNA methylation (**Table 1**). AL2, encoded by *Cabbage leaf curl virus*, and L2 encoded by *Beet curly top virus*, interact with and inactivate adenosine kinase (ADK). ADK is a methyl cycle-associated enzyme required for the efficient production of S-adenosyl methionine (SAM), an essential methyltransferase co-factor. These viral suppressors interfere with methylation of both viral mini chromosomes and host chromosomes. The former promotes viral infection. The latter is associated to symptom development.

Activation of Cellular Transcription

AL2, encoded by *Cabbage leaf curl virus*, and AC2 encoded by *East African cassava mosaic virus* are a transcription factor that suppresses post-transcriptional gene silencing by upregulating plants genes that negatively regulate gene silencing: Werner exonuclease-like 1 (WEL1), and regulator of gene silencing-calmodulin-like protein (rgsCam). In the presence of WEL1 and rgsCam gene silencing is less effective, thus favoring virus infection.

Assays to Identify and Characterize Gene Silencing Suppressors

Viral suppressors of gene silencing have been identified and characterized using a standard transient assay in *Nicotiana benthamiana*. The assay consists in providing a transgene encoding green fluorescent protein (GFP) as a reporter gene. Accumulation of GFP is initially evaluated using a hand-held UV lamp. The GFP transgene is normally silenced by the plant in two days. However, when a silencing suppressor is provided, GFP expression remains active for 4–5 days. Both the GFP transgene and genes being tested are delivered using *Agrobacterium tumefaciens* carrying binary vectors. This assay measures suppression of transgene silencing in infiltrated leaves (local silencing), and *N. benthamiana* can be wild type or constitutively express a GFP transgene (16C). The 16C line was developed by Dr. David Baulcombe's group in 1998 and has been widely used to visualize local and systemic gene silencing. In 16C *N. benthamiana* plants, GFP is silenced in the infiltrated leaves and the silencing signal moves systemically. After 10–15 days of triggering local silencing, GFP silencing is evident in young leaves, as a result of systemic gene silencing. Assays using 16C *N. benthamiana* have been used to identify suppressors that interfere with movement or perception of the gene silencing signal.

In a variation of the assay, suppressor-deficient Turnip crinkle virus (TCV) tagged with GFP (TCV-GFP) is used instead of a single- or double-stranded GFP. TCV-GFP is unable to move cell-to-cell in *N. benthamiana* and is co-expressed with a gene of interest using *A. tumefaciens* carrying binary vectors. Silencing suppressors restore cell-to-cell movement to TCV-GFP and local infection foci visible to the naked eye are formed.

After initial identification in transient assays using single- or double-stranded GFP or TCV-GFP as reporter genes, silencing suppressors are further tested in a heterologous virus, most commonly *Potato virus X* (PVX). PVX expressing a heterologous silencing suppressor normally cause more severe symptoms than the wt PVX.

In a complementary approach, silencing suppressors have been expressed constitutively expressed in model plants such as *A. thaliana* or *N. benthamiana*. Examples are potyviral HC-Pro, CMV 2b, and polerovirus P0. Interestingly, plants expressing one of these silencing suppressors, display development defects that resemble AGO mutants or symptoms characteristic of virus infected plants. These observations support the model that silencing suppressors interfere with endogenous gene silencing and cause the developmental malformations perceived as symptoms in virus infected plants.

Using these assays, a large number of silencing suppressors have been identified in plant and insect viruses. Interestingly, proteins encoded by several human viruses have suppression activity in these assays. Vaccinia virus E3L and *Influenza virus* NS1 have gene silencing suppression activity. However, the biological role of this activity remains to be determined. Furthermore, *Dengue virus* proteins NS3 and NS4 are silencing suppressors. Since *Dengue virus* is transmitted by mosquitoes these proteins likely protect the virus from silencing in the mosquito.

Viral Suppressors of Gene Silencing Need Cellular Factors

Virus-encoded gene silencing suppressors are essential pathogenicity determinants necessary for viruses to interfere with gene silencing. Interestingly, some virus-encoded gene silencing suppressors interact with and need host factors to function. For example, RAV2, an ethylene-inducible transcription factor, is required for suppression of gene silencing by potyviral HC-Pro and TCV P38. In *N. benthamiana*, the calmodulin-like protein (Nbrgs-CaM) is an endogenous plant regulator of gene silencing that functions by repressing expression of RDR6. RDR6 and SGS3 participate in the biogenesis of secondary siRNAs necessary to amplify endogenous and antiviral gene-silencing signals against RNA viruses and geminiviruses. Infection of *N. benthamiana* by *Tomato yellow leaf curl China virus* and the associated β C1 DNA satellite induces expression of Nbrgs-CaM that in turn downregulates RDR6 expression, thus reducing antiviral defense mediated by gene silencing.

Evolution of Viral Suppressors of Gene Silencing

Most silencing suppressors are multifunctional and perform activities necessary for completion of the infection cycle, such as viral RNA replication, virus movement, virion formation or vector transmission (Table 1). Accordingly, virus-encoded silencing suppressors act through diverse mechanisms and are structurally diverse. Furthermore, even for suppressors with the same mechanism, their amino acid sequence is generally diverse. These features have limited the use of computational approaches to identify and characterize gene silencing suppressors and supports the model that silencing suppression has evolved multiple times, independently, through convergent evolution due to host pressure imposed through gene silencing as an antiviral mechanism.

Effect of Gene Silencing Suppressors on Plant Development

A wide range of biological processes are regulated through silencing mechanisms, including plant development, organ formation, and stress responses in eukaryotes. Thus, perturbation of silencing pathways leads to changes in host gene expression that alter normal development. This effect is, at least in part, responsible for symptom development in virus infected plants. As an example, the abaxial-adaxial leaf polarity is regulated by siRNA gradients moving between cells. AGO7 and AGO10 are expressed specifically

in the adaxial leaf primordial and in vascular tissues. AGO7 restricts activity of miR390 and directs a gradient of trans-acting small interfering RNAs from the adaxial to the abaxial side of developing leaves.

Virus-encoded silencing suppressors such as potyviral HC-Pro, orthotospoviral NSs and tombusviral P19 bind cellular miRNAs, and secondary siRNAs. Additionally, potyviral VPg induces degradation of SGS3 and RDR6. TCV P38 inactivates DCL4, while polerovirus P0 and CMV 2b induce AGO1 degradation (Table 1). Collectively, these effects prevent activity of cellular microRNAs and the biogenesis or activity of trans-acting secondary siRNAs. Normal activity of cellular miRNAs and trans-acting secondary siRNAs is essential for plant development. As a consequence of silencing suppression, host genes normally downregulated by siRNAs, accumulate to levels higher than normal and plants display malformation that resemble symptoms caused by virus infection.

Silencing Suppressors Mediate Viral Synergism

Viral synergism occurs when co-infection by two viruses results in disease with more severe symptoms than single virus infection. Most documented cases of viral synergism involve viruses in the family *Potyviridae* and depend on silencing suppressors HC-Pro or P1. Potyviral HC-Pro binds both endogenous and virus-derived siRNAs, their own and those derived from co-infecting viruses. siRNA binding by HC-Pro prevents targeting of viral RNA and results in higher accumulation of the non-potyvirus and symptoms are more severe than those of single infection. An example of economic importance is maize lethal necrosis, caused by the coinfection of Maize chlorotic mottle virus and Sugarcane mosaic virus. Maize lethal necrosis is a re-emerging disease that has reached epidemic proportion and is threatening food security in sub-Saharan Africa. In potato, co-infection by Potato virus X and Potato virus Y causes synergism. In pepper, co-infection by Cucumber mosaic virus and Pepper mottle virus causes synergism. Similarly, in papaya co-infection by Papaya mosaic and Papaya ringspot virus causes synergism. In all these cases, the non-potyvirus accumulates to higher levels in plants double infected, compared to plants single infected. Interestingly, accumulation of the potyvirus is similar in single or double-infected plants. The genetic determinant of synergism is potyviral silencing suppressor HC-Pro.

There is one documented case of synergism in DNA viruses. In cassava, a synergistic disease is caused by the co-infection of African cassava mosaic virus (ACMV) and East African cassava mosaic virus (EACMV). ACMV and EACMV encode AC4 and AC2, respectively, that suppress gene silencing by different mechanisms. Furthermore, it is the only documented case of synergism between viruses of the same genus.

Viral Suppressors in Support of Biotechnology

Several pharmaceutical products, including vaccines against human pathogens such as influenza viruses are produced in plants. The corresponding coding sequence is provided as a transgene, which often is turned off by gene silencing. In transient assays, this limitation has been resolved using virus-encoded gene silencing suppressors, which enhance gene expression and boost yield of the pharmaceutical product of interest.

Further Reading

- Burgyan, J., Havelda, Z., 2011. Viral suppressors of RNA silencing. *Trends in Plant Science* 16, 265–272.
- Chapman, E.J., Prokhnovsky, A.I., Gopinath, K., Dolja, V.V., Carrington, J.C., 2004. Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. *Genes & Development* 18, 1179–1186.
- Csorba, T., Kontra, L., Burgyan, J., 2015. Viral silencing suppressors: Tools forged to fine-tune host-pathogen coexistence. *Virology* 479–480, 85–103.
- Deleris, A., Gallego-Bartolome, J., Bao, J., *et al.*, 2006. Hierarchical action and inhibition of plant dicer-like proteins in antiviral defense. *Science* 313, 68–71.
- Diaz-Pendon, J.A., Ding, S.W., 2008. Direct and indirect roles of viral suppressors of RNA silencing in pathogenesis. *Annual Review of Phytopathology* 46, 303–326.
- Ding, S.W., 2010. RNA-based antiviral immunity. *Nature Reviews Immunology* 10, 632–644.
- Ding, S.W., Voinnet, O., 2007. Antiviral immunity directed by small RNAs. *Cell* 130, 413–426.
- Garcia-Ruiz, H., Carbonell, A., Hoyer, J.S., *et al.*, 2015. Roles and programming of Arabidopsis ARGONAUTE proteins during Turnip mosaic Virus Infection. *PLOS Pathogens* 11, e1004755.
- Incarbone, M., Dunoyer, P., 2013. RNA silencing and its suppression: Novel insights from in planta analyses. *Trends in Plant Science* 18, 382–392.
- Johansen, L.K., Carrington, J.C., 2001. Silencing on the spot: induction and suppression of RNA silencing in the Agrobacterium-mediated transient expression system. *Plant Physiology* 126, 930–938.
- Korner, C.J., Pitzalis, N., Pena, E.J., *et al.*, 2018. Crosstalk between PTGS and TGS pathways in natural antiviral immunity and disease recovery. *Nature Plants* 4, 157–164.
- Mlotshwa, S., Pruss, G.J., Vance, V., 2008. Small RNAs in viral infection and host defense. *Trends in Plant Science* 13, 375–382.
- Muhammad, T., Zhang, F., Zhang, Y., Liang, Y., 2019. RNA interference: A natural immune system of plants to counteract biotic stressors. *Cells* 8.
- Vanitharani, R., Chellappan, P., Pita, J.S., Fauquet, C.M., 2004. Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. *Journal of Virology* 78, 9487–9498.
- Zhang, X., Yuan, Y.R., Pei, Y., *et al.*, 2006. Cucumber mosaic virus-encoded 2b suppressor inhibits Arabidopsis Argonaute1 cleavage activity to counter plant defense. *Genes & Development* 20, 3255–3268.
- Zhao, J.H., Hua, C.L., Fang, Y.Y., Guo, H.S., 2016. The dual edge of RNA silencing suppressors in the virus-host interactions. *Current Opinion in Virology* 17, 39–44.

Virus-Induced Gene Silencing (VIGS)

Xu Tengzhi, Ugrappa Nagalakshmi, and Savithramma P Dinesh-Kumar, University of California, Davis, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of M.S. Padmanabhan, S.P. Dinesh-Kumar, Virus-Induced Gene Silencing (VIGS), In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00578-1](https://doi.org/10.1016/B978-012374410-4.00578-1).

Nomenclature

ABA	Absciscic acid	ORF	Open reading frame(s)
AGO	Argonaute	PDS	Phytoene desaturase
CHS	Chalcone synthase	PTGS	Post-transcriptional gene silencing
CP	coat protein or capsid protein	RdRp	RNA-dependent RNA polymerase
dsRNA	Double stranded RNA	RISC	RNA-induced silence complex
HSP90	Heat shock protein 90	RNAi	RNA interference
LIC	Ligation-independent cloning	SGN	Solanaceae genomics network
mRNA	Messenger RNA	UTR	Untranslated region
NLR	Nucleotide-binding leucine rich repeat	VIGS	Virus-induced gene silencing
NOS	Nopaline synthase	VPGD	VIGS phenomics and functional genomics database
		VPI	Vascular puncture inoculation

Glossary

Agrobacterium-mediated genetic transformation

Agrobacterium genetically transforms its host by transferring a well-defined DNA segment from its tumor-inducing (Ti) plasmid to the host-cell genome.

Agro-infiltration It is a method used in plant biology and especially lately in plant biotechnology to induce transient expression of genes in a plant, or isolated leaves from a plant, or even in cultures of plant cells, in order to produce a desired protein. In the method a *suspension* of *Agrobacterium tumefaciens* is introduced into a plant leaf by direct injection or by vacuum infiltration, where after the bacteria transfer the desired gene into the plant cells via transfer of T-DNA.

Argonaute It is a family of proteins that bind small noncoding RNAs, including miRNAs and siRNAs, and are guided to RNA targets through sequence complementarity, leading to the cleavage or translation inhibition of mRNAs, including viral RNAs.

Co-suppression phenomenon understood as post-transcriptional gene silencing, which can be triggered by introducing transgenes or a certain viruses expressing the target gene.

CRISPR-Cas Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated proteins (Cas) is a prokaryotic adaptive immune system that can be leveraged for genetic material to be added, removed, or altered at particular locations in the plant genome, or to directly target viral DNA or RNA molecules.

Dicer It is a multidomain ribonuclease that processes double-stranded RNAs to 21-nt small interfering RNAs during RNA interference and excises microRNAs from precursor hairpins.

Forward genetics Approaches used to identify directly genes responsible for a particular phenotype of an organism (DNA).

Functional genomics The study of genes with respect to the role they play within biological processes.

Knockdown Reduction in the expression of a gene.

Knockout Complete inhibition of gene expression.

MicroRNAs These are a class of small non-coding ~21–24 nt RNAs that are involved in the regulation of gene expression at the post-transcriptional level by degrading target mRNAs, including viral RNAs, and/or inhibiting their translation.

Post-transcriptional gene silencing Silencing of an endogenous gene caused by the introduction of a homologous dsRNA, transgene or virus. In PTGS, the transcript of the silenced gene is synthesized but does not accumulate because it is rapidly degraded.

Reverse genetics Approaches used to define the function of a gene or sequence of DNA within the context of the organism (RNA, proteins).

RNA interference It is a ribonucleoprotein complex that incorporates one strand of a miRNA or a siRNA duplex for RNAi.

RNA-induced silencing complex It is a ribonucleoprotein complex that incorporates one strand of a miRNA or a siRNA duplex for RNAi.

Small interfering RNA It is a class of 20–25 nt non-coding dsRNA molecules.

Trans-acting siRNA These are siRNA that repress plant gene expression through post-transcriptional gene silencing.

Introduction

Virus-induced gene silencing (VIGS) is an effective forward and reverse genetics approach in plants. VIGS uses the antiviral defense mechanism in plants to transiently downregulate specific transcripts through a recombinant vector. The VIGS vector, which carries a partial sequence from the gene to be silenced, is introduced into the plant cell via *Agrobacterium tumefaciens*-mediated transient transformation or using in vitro-generated transcripts. Double stranded RNA (dsRNA) molecules generated during viral infection trigger the RNA interference (RNAi) pathway; leading to the degradation of messenger RNAs (mRNAs) homologous to the dsRNA and subsequent downregulation of the corresponding gene. The first VIGS vector, based on Tobacco mosaic virus (TMV), was developed in 1995. The last two decades have seen a significant increase in the number of viruses that have been adapted for VIGS to study gene function in plants. VIGS has major advantages over the labor-intensive transgenic approaches including speed, efficacy and cost. These attributes have made VIGS a versatile functional genomic tool in a wide range of plant species to unravel gene functions involved in a variety of plant biological processes.

PTGS and VIGS

The term “cross-protection” was first described in the 1920s as a phenomenon of induced resistance in a host plant against one virus or viroid strain infection if the host plant had been previously infected with another milder strain of the same virus or viroid. However, the mechanistic basis of this phenomenon was not understood. In 1990, researchers attempted to develop an enhanced purple colored petunia flower by overexpressing a pigmentation gene *chalcone synthase* (CHS). Unexpectedly, the flowers of CHS overexpressing plants were either entirely white or a mottled purple/white. This phenomenon was named “co-suppression” and considered to be an oddity of petunia. In the following few years, the same phenomenon was reported in many other plant species. The “cross-protection” and “co-suppression” phenomenon were soon understood as post-transcriptional gene silencing (PTGS), which could be triggered by introducing transgenes or using viruses as gene carrier. Around the same time, a similar phenomenon was described in the fungus *Neurospora crassa* and named as “quelling”. This phenomenon has since been observed in many organisms, including nematodes, flies, and mammals; coming to be known more widely as RNA interference (RNAi). These cross-kingdom PTGS/quelling/RNAi are evolutionary conserved and induced by dsRNA.

The term VIGS was coined in 1997 and used to describe the symptoms recovery phenomenon of plants upon viral infection. As mentioned above, viral infection activates the plant RNAi-mediated antiviral defense mechanism, PTGS. The machinery involved in PTGS targets the viral genome as an antiviral defense. This was exploited as the basis of VIGS. Now, VIGS specifically refers to the reverse genetics technique for down-regulation of target gene expression in plants by recombinant viruses which carry a partial sequence of the target gene.

PTGS is activated upon virus infection in plants. During the viral replication process, dsRNA intermediates are generated. Synthesis of dsRNA is mediated by RNA-dependent RNA polymerase (RdRp). DNA viruses employ the host RdRp for the dsRNA synthesis, while RNA viruses encode their own RdRp for the dsRNA production. Viral dsRNAs are recognized as host aberrant RNAs and cleaved by the host plant's Dicer-like ribonuclease, producing short interfering RNA (siRNA) duplex molecules which are of 21–24 nt in length. The siRNA duplex can then be recognized and unpaired by Argonaute (AGO) protein into a guide strand and a passenger strand in an ATP-dependent manner. The Passenger strand is degraded while the Guide strand is loaded into a multi-protein complex known as the RNA-induced silencing complex (RISC). The RISC complex, harboring siRNA, scans the mRNA population inside the cell and specifically targets transcripts that are complementary to siRNA, resulting in cleavage of target transcript. As a result, the mRNA level of the target gene is reduced. The silencing signal can be further amplified by plant RNA-dependent RNA polymerase 6 (RDR6) and spread systemically throughout the plant, causing systemic gene silencing.

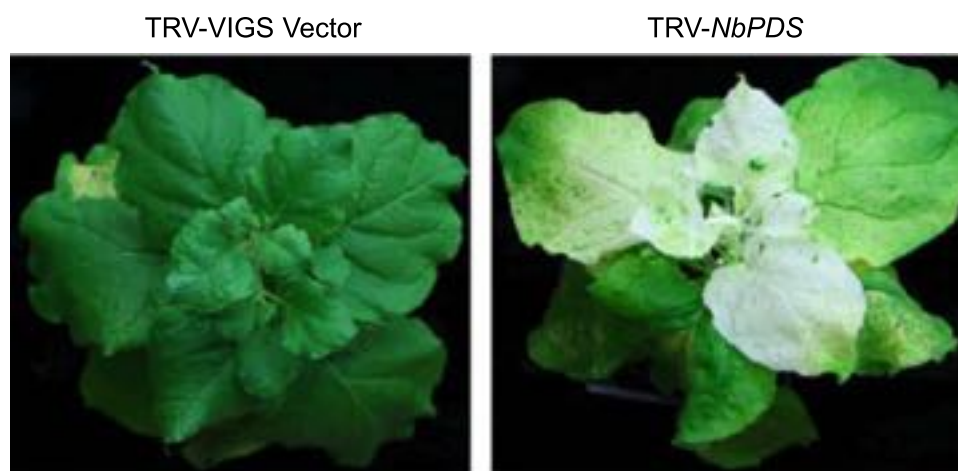
Establishment of a VIGS System

To develop a virus vector into a more efficient VIGS tool, the selected virus should have a broad host range and should cause only mild disease symptoms. In this way, the VIGS-vector-induced silencing phenotype will not be obscured by disease symptoms. The best VIGS vectors are characterized by high silencing efficiency as well as persistent silencing phenotype. Furthermore, uniform expression of a silencing phenotype, in a tissue-independent manner, is also an important factor that should be considered when choosing a virus for VIGS vector development.

In the last two decades, many plant viruses have been used to develop VIGS vectors. Most VIGS vectors are derived from RNA viruses, with a few DNA viruses and satellites having been used as well (Table 1). For development of a VIGS vector, the viral genome is cloned under the control of a constitutive promoter such as Cauliflower mosaic virus (CaMV) 35S promoter and a terminator such as nopaline synthase (NOS) terminator in a T-DNA binary vector. The VIGS T-DNA vector, containing the target gene insert, is transferred into plants via *Agrobacterium tumefaciens*-mediated transient plant transformation system or DNA of VIGS vector could be delivered by biolistic bombardment. In addition, transcripts of T7 promoter-based VIGS vector generated by in vitro transcription could be used for rub-inoculation onto plants. Agroinfiltration is the most widely used method for VIGS because it is inexpensive and highly reproducible. Syringe infiltration is commonly used to introduce the agrobacterium into the

Table 1 List of plant viruses used as VIGS vectors for gene silencing and their experimental plant hosts

<i>Virus</i>	<i>Genus/Family</i>	<i>Host species</i>
RNA virus		
Apple latent spherical virus (ALSv)	<i>Cheravirus Secoviridae</i>	<i>Chenopodium quinoa</i> , soybean, apple and <i>Nicotiana benthamiana</i>
Bamboo mosaic virus (BaMV)	<i>Potexvirus Alphaflexiviridae</i>	<i>Brachypodium distachyon</i>
Barley stripe mosaic virus (BSMV)	<i>Hordevirus Virgaviridae</i>	Wheat, barley
Bean pod mottle virus (BPMV)	<i>Comovirus Secoviridae</i>	Soybean
Brome mosaic virus (BMV)	<i>Bromovirus Bromoviridae</i>	Maize
Cucumber mosaic virus (CMV)	<i>Cucumovirus Bromoviridae</i>	<i>Nicotiana benthamiana</i> and maize
Cymbidium mosaic virus (CymMV)	<i>Potexvirus Alphaflexiviridae</i>	Orchid
Foxtail mosaic virus (FoMV)	<i>Potexvirus Alphaflexiviridae</i>	Barley, wheat, foxtail millet, and maize
Pea early browning virus (PEBV)	<i>Tobravirus Virgaviridae</i>	<i>Medicago truncatula</i> and pea
Potato virus X (PVX)	<i>Potexvirus Alphaflexiviridae</i>	<i>Nicotiana benthamiana</i>
Tobacco rattle virus (TRV)	<i>Tobravirus Virgaviridae</i>	<i>Nicotiana benthamiana</i> , tobacco, tomato, potato, pepper, <i>Arabidopsis</i> , cotton, poppy and others
Turnip yellow mosaic virus (TYMV)	<i>Tymovirus Tymoviridae</i>	Chinese cabbage and <i>Arabidopsis</i>
DNA virus		
African cassava mosaic virus (ACMV)	<i>Begomovirus Geminiviridae</i>	Cassava
Cabbage leaf curl virus (CaLCuV)	<i>Begomovirus Geminiviridae</i>	<i>Arabidopsis</i>
Cotton leaf crumple virus (CLCrV)	<i>Begomovirus Geminiviridae</i>	Cotton
Pepper huasteco yellow vein virus (PHYVV)	<i>Begomovirus Geminiviridae</i>	Pepper
Rice tungro bacilliform virus (RTBV)	<i>Tungrovirus Caulimoviridae</i>	Rice
Tomato golden mosaic virus (TGMV)	<i>Begomovirus Geminiviridae</i>	<i>Nicotiana benthamiana</i>
Satellite virus		
Satellite Bamboo mosaic virus (satBaMV)	<i>Potexvirus^a</i>	<i>Brachypodium distachyon</i>
Tomato yellow leaf curl China betasatellite (TYLCCNV)	<i>Begomovirus^a Geminiviridae</i>	Tomato

^aGenus of the helper virus.**Fig. 1** Silencing phenotype of the *phytoene desaturase* (*PDS*) gene in *Nicotiana benthamiana*. Four week old *N. benthamiana* plants infected with TRV-*NbPDS* show photobleaching phenotype (right panel) compared to the TRV-VIGS vector (left panel) infected plant. Pictures were taken two weeks post VIGS vector inoculation.

plants. However, other delivery routes such as spraying, vacuum infiltration and agro-drench have also been used to deliver VIGS vector containing T-DNA vectors in plants.

Virus-encoded silencing suppressor genes can reduce VIGS silencing efficiency and should be removed or modified in the VIGS vector. Some non-essential genes, for example, those that are not necessary for viral replication, cell-to-cell movement and systemic movement, could be removed from the viral genome while constructing a VIGS vector. To clone a target gene fragment into a VIGS vector, an appropriate position in the viral genome should be selected in such a way that it does not interfere with viral lifecycle. In some cases, a sub-genomic promoter from a related virus has been used in a VIGS vector to drive the expression of a target gene for silencing. To clone the target gene fragment into the VIGS vector, the restriction/ligation, gateway cloning, or ligation-independent cloning (LIC) strategies have been used. The optimal gene fragment length for efficient silencing of a target gene is between 200 and 500 bp. However, gene fragments as short as 23 nt are sufficient to knockdown target gene expression. The silencing

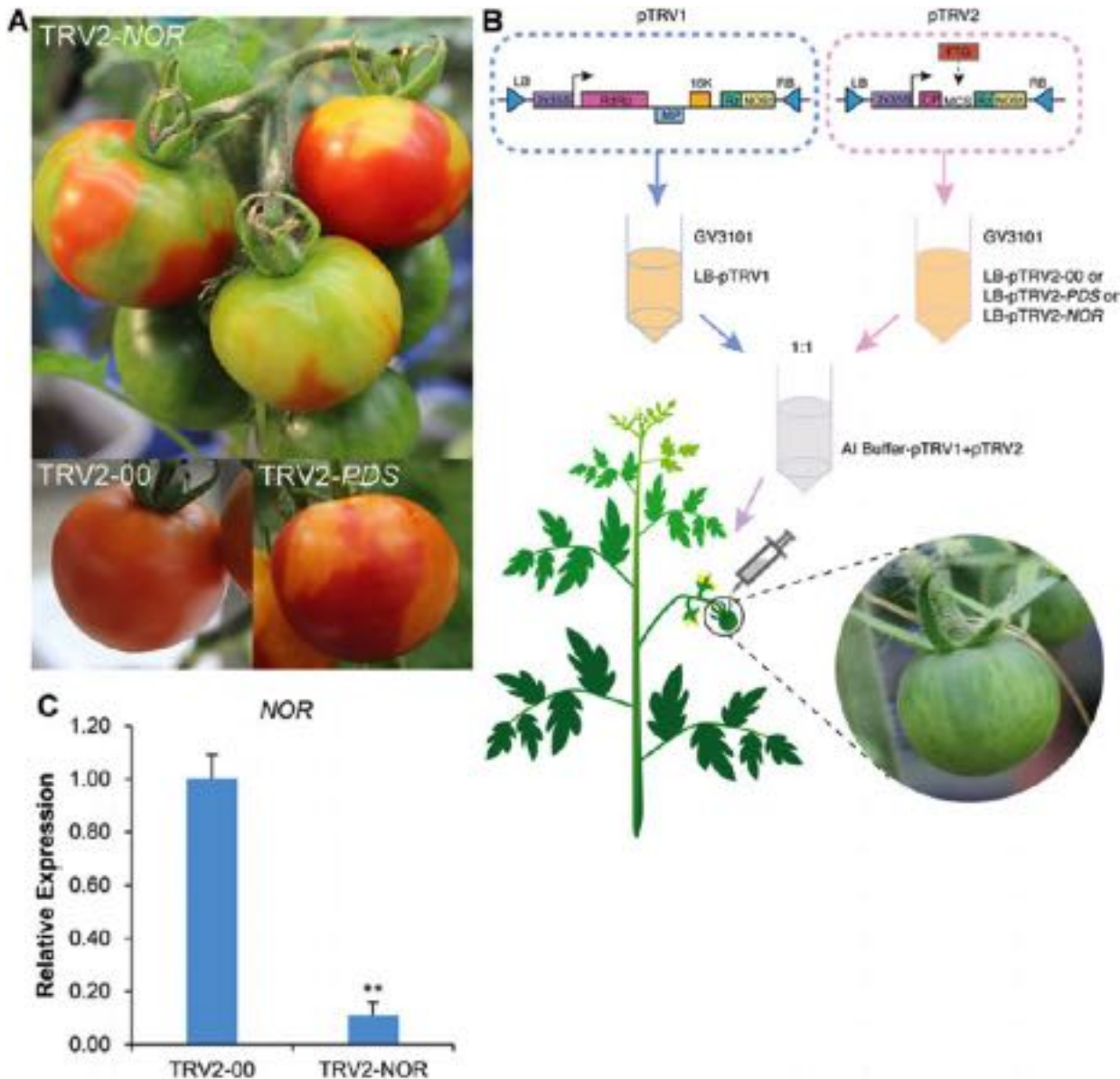


Fig. 2 VIGS assay applied to tomato fruit. (A) Phenotype of *Nor*-silenced fruit. Vector only (pTRV2-00) and *PDS*-silenced (pTRV2-*PDS*) fruits were used as the control. (B) Flow chart of the VIGS assay for tomato fruit. To construct the pTRV2 vector, a fragment of the target gene (FTG) was inserted into the multiple cloning site (MCS). pTRV1 and pTRV2 plasmids were independently transferred to *A. tumefaciens* (GV3101). The cells in the LB cultures containing pTRV1 and pTRV2 were harvested and mixed at a ratio of 1:1 after resuspension in *Agrobacterium* infiltration buffer (AI buffer); the mixture was used to infiltrate the carpel of the tomato fruit attached to the plant at 7–10 DPA. (C) The silencing efficiency of the *Nor* gene in AC tomato fruit at the red-ripe stage (RR). CaMV 35S promoter (35S), nopaline synthase terminator (NOST), coat protein (CP). Luria-Bertani medium (LB). Asterisks indicate a significant difference as determined using Student's t-test (**, $P < 0.01$). Reproduced Fig. 1 from Yuan, X.Y., Wang, R.H., Zhao, X.D., Luo, Y.B., Fu, D.Q., 2016. Role of the tomato non-ripening mutation in regulating fruit quality elucidated using iTRAQ protein profile analysis PLOS ONE 11, e0164335. Available at: <https://doi.org/10.1371/journal.pone.0164335>.

efficiency varies depending on the target region selected. Therefore, multiple fragments targeting different regions of the target gene of interest should be tested. In view of this, Solanaceae Genomics Network (SGN) VIGS Tool (See Relevant Websites Section) could be used to select an optimal target gene fragment for cloning into a VIGS vector.

VIGS Vectors

Many VIGS vectors developed to-date have been used for gene silencing in dicot plants. Since *Nicotiana benthamiana* is susceptible to a wide range of plant viruses and amenable to agro-infiltration it is widely used to test the efficacy of a newly developed VIGS vector. The

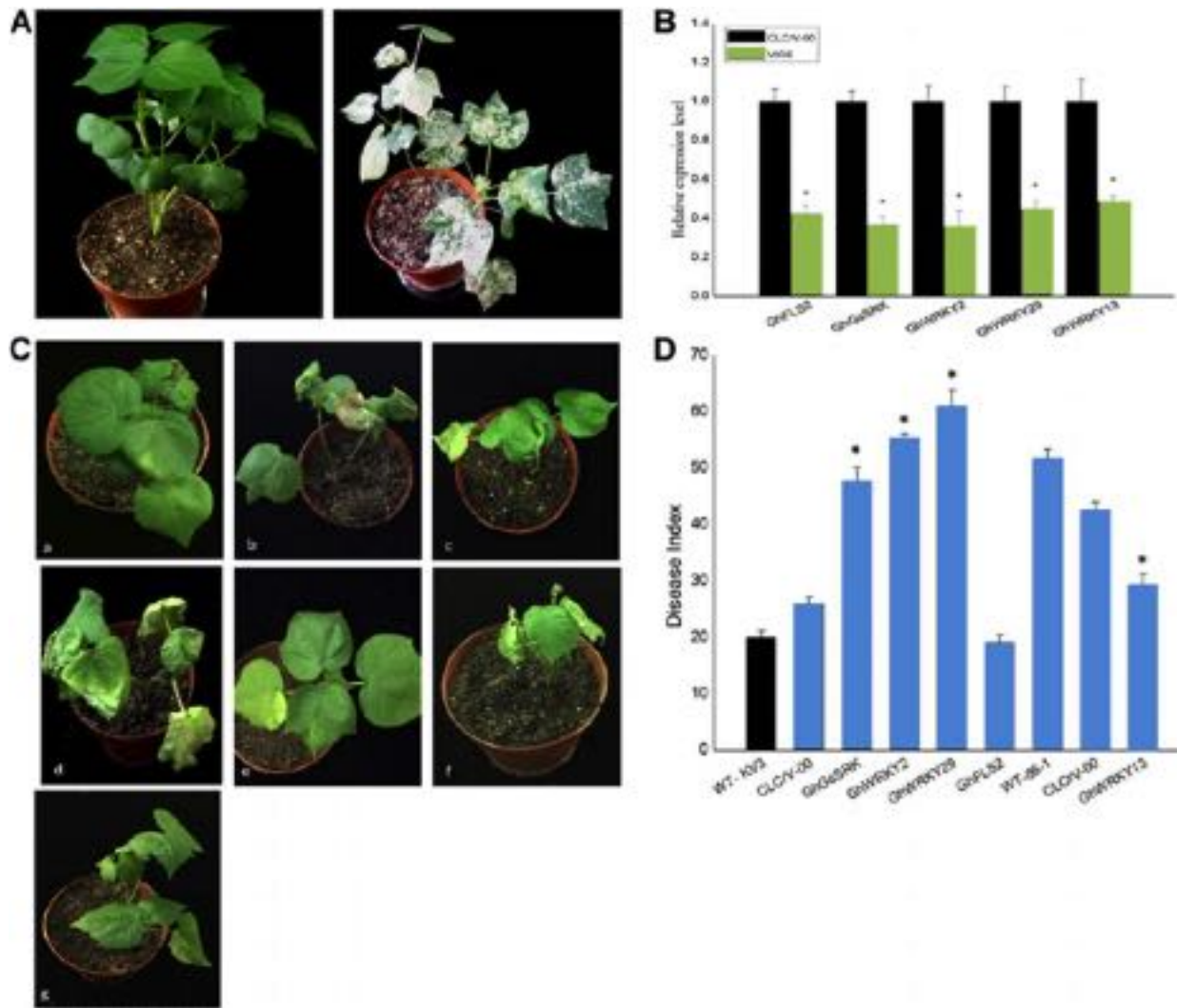


Fig. 3 The RLKs and WRKYs are required for resistance against *V. dahliae* in cotton via VIGS analysis. (A) Silencing of CLA1 gene in *G. hirsutum* variety Zhongzhimian KV3. Left, plant infiltrated with CLCrV-based empty vector (CLCrV-00); right, plant infiltrated with CLCrV-CLA1. (B) Relative expression levels of candidate gene in the silenced and non-silenced cotton plants at 20 dpi were determined through qRT-PCR. Asterisk represents the data point that was statistically different from the non-silenced ($p < 0.05$) analyzed by one-way ANOVA, using the SAS 8.1. The vertical bars indicate standard Deviation. (C) a, Negative control-silenced upland cotton (Zhongzhimian KV3, CLCrV-00); b, GhGsSRK-silenced Zhongzhimian KV3 plant; c, GhWRKY2-silenced Zhongzhimian KV3 plant; d, GhWRKY29-silenced Zhongzhimian KV3 plant; e, GhFLS2-silenced Zhongzhimian KV3 plant; f, Negative control-silenced upland cotton (86-1, CLCrV-00); g, GhWRKY13-silenced 86-1 plant. (D) The disease index (DI) after silencing different genes. The results are presented as mean $\pm$ standard deviation (SD) from three replicates with at least 20 plants per replicate. Black represents candidate gene silenced in Zhongzhimian KV3, and blue represents candidate gene silenced in 86-1. Asterisk represents the data point that was statistically different from wild-type and CLCrV-00 plants ($p < 0.05$) analyzed by one-way ANOVA, using the SAS 8.1. The vertical bars indicate standard Deviation. Reproduced Fig. 3 from Zhang, W., Zhang, H., Liu, K., *et al.*, 2017. Large-scale identification of *Gossypium hirsutum* genes associated with *Verticillium dahliae* by comparative transcriptomic and reverse genetics analysis. PLOS ONE 12 (8), e0181609. Available at: <https://doi.org/10.1371/journal.pone.0181609>.

Tobacco rattle virus (TRV)-VIGS vector is by far the most widely used VIGS vector due to its broad host range and high silencing efficiency in addition to its very mild disease symptoms. TRV is a bipartite, positive-strand RNA virus consisting of RNA1 and RNA2 genomes. Since the RdRp, movement protein, and 16K protein encoded by RNA1 are essential for TRV infection, the RNA1 is unmodified in the TRV-VIGS system. The cDNA corresponding to RNA1 is cloned under the control of a 35S promoter and a NOS terminator (TRV1). In RNA2, the coding region required for vector transmission has been removed and the cDNA corresponding to the remaining parts are cloned between a 35S promoter and a NOS terminator. In RNA2 vector (TRV2), a multiple cloning site, gateway cloning cassette, or ligation-independent cloning cassette has been introduced downstream of intergenic region after the coat protein (CP) open reading frame (ORF). TRV-VIGS has been used successfully for gene function studies in more than 25 plant species including a large number of plants in *Solanaceae* (Figs. 1 and 2), as well as in *Malvaceae* (Fig. 3), *Brassicaceae*, *Euphorbiaceae*, *Papaveraceae* and *Rosaceae*. In legumes, Bean pod mottle virus (BPMV) and Pea early browning virus (PEBV) have been used successfully as VIGS vectors.

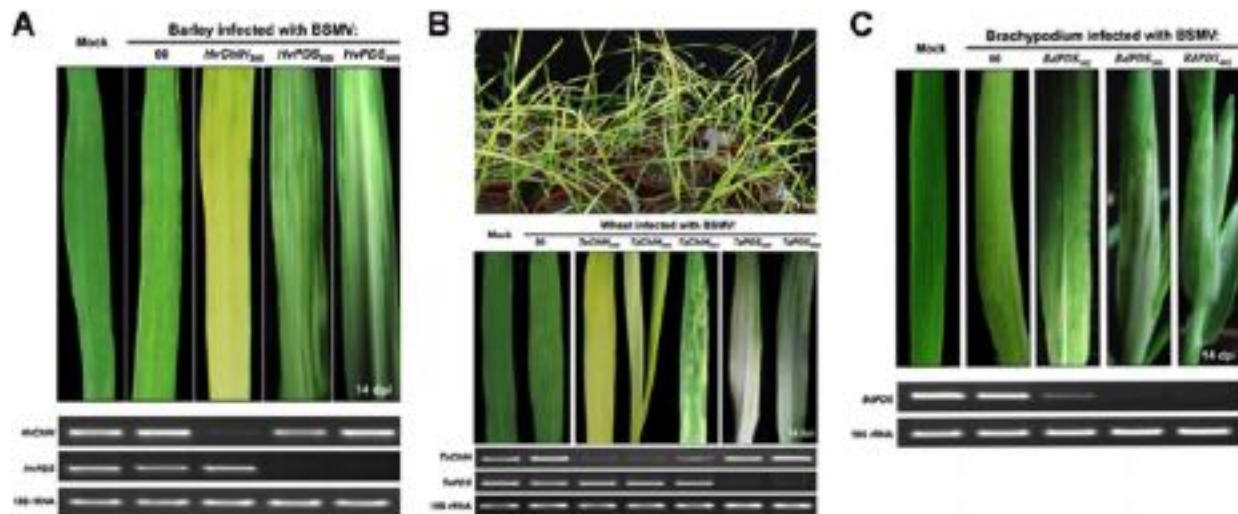


Fig. 4 *PDS* and magnesium chelatase subunit H (*ChlH*) gene silencing phenotypes in barley, wheat and *B. distachyon*. (A) Barley showing phenotypes typical of suppression of *ChlH* and *PDS* by BSMV:*HvChlH* or BSMV:*HvPDS* inserts of the indicated lengths. (B) Upper panel: Wheat inoculated with BSMV:*TaChlH*₂₅₀ showing more than 90% of inoculated plants developing the chlorotic phenotype associated with suppression of the *ChlH* gene. Middle panel: *PDS* and *ChlH* gene silencing phenotypes on wheat leaves infected with BSMV:*TaChlH* and BSMV:*PDS* derivatives with different length inserts. Bottom panel: PCR amplification of transcripts from leaves shown in the middle panel. (C) *B. distachyon* leaves showing effects of *PDS* suppression after infection with BSMV:*BdPDS* inserts. In these experiments, plants were inoculated at the two-leaf stage with infected *N. benthamiana* sap harboring BSMV derivatives targeting *PDS* and *ChlH* cognate genes from each species. Leaf photographs were taken at 14 dpi, and fragment lengths from each source are indicated as subscripts above each leaf. Relative transcript levels of *PDS* and *ChlH* genes in the leaves infected with the different BSMV derivatives are shown under the leaf photographs. RNA extracted from the leaves was subjected to semi-quantitative RT-PCR amplification (28 cycles for wheat and barley, 31 cycles for *B. distachyon*) with the gene-specific oligonucleotide primers. Amplified species-specific 18S rRNA served as internal controls for each species. Reproduced Fig. 6 from Yuan, C., Li, C., Yan, L., et al. 2011. A high throughput Barley stripe mosaic virus vector for virus induced gene silencing in monocots and dicots. PLOS ONE 6, e26468. Available at: <https://doi.org/10.1371/journal.pone.0026468>.

Very few vectors have been developed for gene silencing in monocot plants (Table 1). Monocots are generally recalcitrant to agro-infiltration. As such, mechanical inoculation methods such as rub-inoculation of crude sap from infected *N. benthamiana* plants, in vitro generated VIGS vector transcripts, or vascular puncture inoculation (VPI) of in vitro transcripts or virions has been used to introduce VIGS vectors into monocots. Brome mosaic virus (BMV) and Barley stripe mosaic virus (BSMV; Fig. 4) VIGS vectors have been used to silence genes in maize, wheat and barley. Recently, Foxtail mosaic virus (FoMV)-based VIGS vector has been used to silence genes in barley, wheat, foxtail millet, and maize and Cucumber mosaic virus (CMV)-based vector has been used to silence genes in maize.

Generally, *Phytoene desaturase* (*PDS*), a gene involved in carotenoid biosynthesis, has been used as a target to test the efficacy of newly developed VIGS vectors. Since carotenoid synthesis is important for photosynthesis and photoprotection, low levels of carotenoids result in a degradation of chlorophyll. This is manifested in a visible photobleaching phenotype in *PDS* silenced plants (Fig. 1). This visible feature makes *PDS* an excellent positive control for monitoring the silencing efficiency.

VIGS Application

VIGS vectors have been used widely as a reverse and forward genetic tool to unravel the functions of genes involved in various biological processes such as disease resistance, cell death, abiotic stress, cell signaling, flower development (Fig. 5), fruit development (Fig. 2), secondary metabolite biosynthesis, and evolutionary developmental biology.

The role of Heat Shock Protein 90 (HSP90) in Rx1 nucleotide-binding leucine rich repeat (NLR) class receptor-mediated immunity was uncovered in a forward genetics by screening nearly 5000 cDNAs in TRV-VIGS vector in *N. benthamiana* plants. The role of autophagy in immunity and cell death in plants was uncovered by screening TRV-VIGS library. Similar TRV-VIGS library was used to identify the role for ribosomal proteins, RPL12 and RPL19, in non-host disease resistance against bacterial pathogens in *N. benthamiana*.

Numerous studies have used VIGS as a reverse genetics tool to study gene function in the literature. NCBI PubMed database using the term "VIGS" lists 828 publications. Majority of these publications describe using VIGS to discover gene function in various plant biological processes. By far the most application of VIGS is in the disease resistance area (168 publications in NCBI). VIGS has been used to study function of genes involved in disease resistance in various plants including *N. benthamiana*, tomato, pepper, potato, soybean, wheat, barley, maize, cotton, cassava, and other plants.

In addition to biotic interaction, VIGS has been used to study abiotic stress responses. VIGS vectors based on TRV, BSMV, BPMV, PEBV, and Tomato yellow leaf curl China virus (TYLCCNV) have been used successfully to study genes function in abiotic

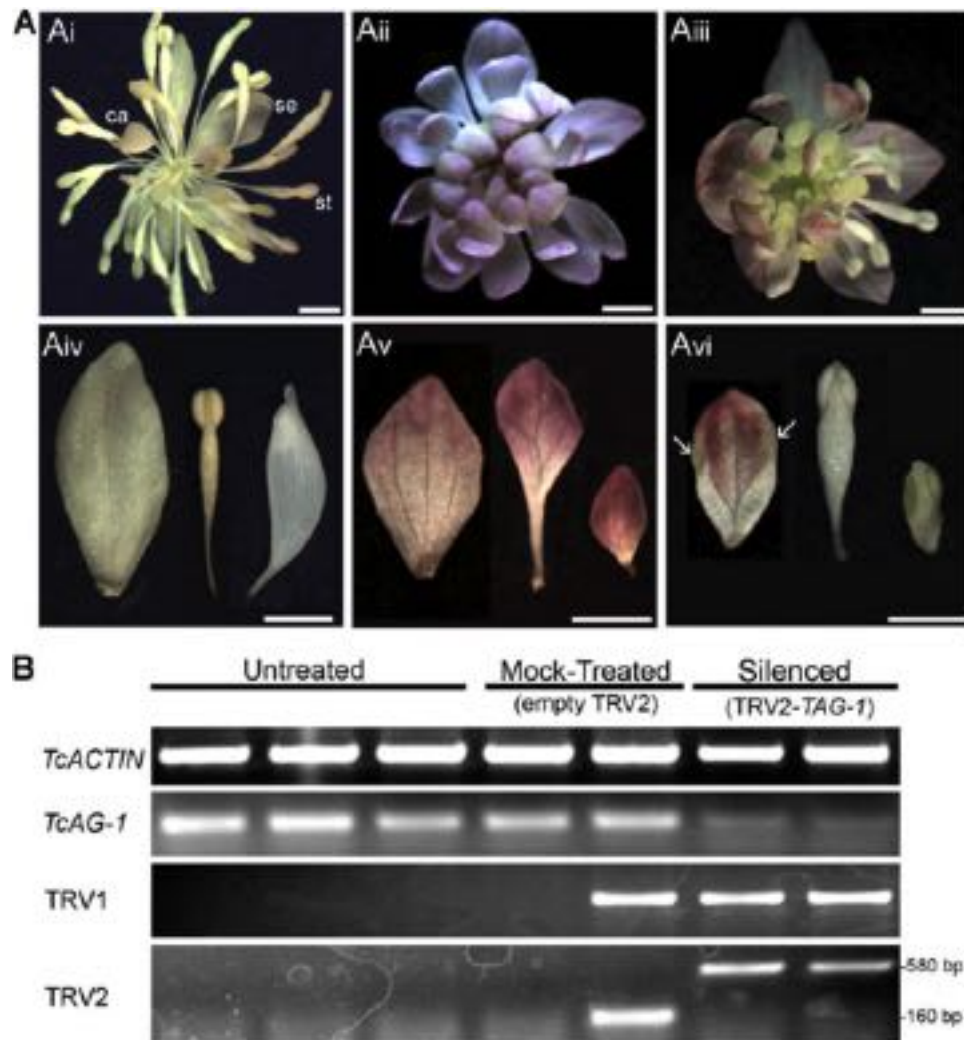


Fig. 5 Virus-induced gene silencing of *Thalictrum clavatum* AGAMOUS ortholog TcAG-1 results in homeotic floral phenotypes. (A) Flower silencing phenotypes of TcAG-1, relative to controls. Ai, Untreated flower of *T. clavatum* showing sepals (se), stamens (st) and carpels (ca); Aii: strongly silenced flower in TRV2-TAG-1 treated plant, showing an array of sepals and no stamens nor carpels, all reproductive organs have been homeotically converted to sepals; Aiii: intermediate phenotype with partial conversion of organs and some normal ones; Aiv: detail of dissected organs in an untreated flower (sepal, stamen, carpel, from left to right); Av: detail of all sepaloid dissected organs from a strong TcAG-1 silencing phenotype (from the outside to the inside of the flower, left to right); Avi: detail of sample chimeric organs, arrows point to anther tissue on the edges of an internal “sepal”. Scale bar = 1 mm. (B) Gene expression by Reverse Transcriptase (RT)-PCR in TcAG-1 silenced plants compared to controls. Untreated and mock-treated (empty TRV2) plants are compared to TRV2-TAG-1 treated plants showing strong homeotic conversions (Aii, Av). RT-PCR was performed with locus-specific primers to the housekeeping gene ACTIN (loading control); to the MADS box gene TcAG-1 and to the viral transcripts TRV1/TRV2. For TRV2: larger bands result from the presence of insert, the smaller band from an empty TRV2 (mock control). Reproduced Fig. 4 from Di Stilio, V.S., Kumar, R.A., Oddone, A.M., *et al.*, 2010. Virus-induced gene silencing as a tool for comparative functional studies in *Thalictrum*. PLOS ONE 5, e12064. Available at: <https://doi.org/10.1371/journal.pone.0012064>.

stress responses including drought, highlight, iron deficiency, oxidative stress, abscisic acid (ABA) response, salt stress, osmotic stress, dehydration and nutrient deficiency in crop plants such as tomato, pepper, pea, soybean, wheat, and barley.

Many groups have also used VIGS to study genes involved in various plant developmental pathways including flower and fruit development. Genes involved in flower development has not only been studied in model plants such as *N. benthamiana* and *Arabidopsis* but also in crop plants such as tomato, petunia, soybean, and cherry. In addition, TRV-based VIGS has been used in comparative studies involving flower development genes in basal eudicots such as *Aquilegia*, *Cysticapnos vesicaria*, *Papaver somniferum* (opium poppy) and *Eschscholzia californica* (California poppy); *Gerbera hybrida*, *Phalaenopsis* orchids, and rose. TRV-based VIGS has been used successfully to uncover the role of genes involved in fruit development in tomato, pepper, peach, and strawberry.

Recently, a database describing phenotypes resulting from silencing of approximately 1300 genes in *N. benthamiana* has been developed, VIGS phenomics and functional genomics database (VPGD; See Relevant Websites Section). Since genes in *N. benthamiana*

have homologs in other *Solanaceae* plants such as tomato, potato and pepper; one could infer the function of the corresponding genes in these crop plants.

Advantages and Limitations

VIGS is an efficient forward genetic approach to identify loss-of-function phenotypes of specific genes. TRV-VIGS requires less than 2 weeks to knockdown target gene expression in *N. benthamiana* plants. However, it usually takes 3–4 months to generate transgenic *N. benthamiana* with the traditional *Agrobacterium*-mediated genetic transformation; homozygous lines require additional time. Additionally, the transformation process is labor-intensive and expensive. The VIGS could be used to silence a single gene or even multiple genes. Since VIGS has been applied in a broad range of plant species, it is possible to conduct comparative studies using homologous genes in different plant species.

Many plants are highly recalcitrant to genetic transformation. For some genes in sexually propagated plants, the mutation or antisense-mediated knockdown is lethal. In these cases, VIGS can serve as a valuable alternative for gene function studies as it eliminates the need to establish a stable genetic transformation method. Silencing of a specific gene within a multi-gene family and silencing multi-copy genes in polyploid plants is also feasible by VIGS method by choosing the targeting sequence derived from the most conserved region among gene family members. Similarly, one could use 3' untranslated region (UTR) to specifically silence a single gene within a gene family.

VIGS, like any technique, has limitations. The silencing effect of VIGS is transient. However, silencing by TRV-VIGS has been shown to last for longer time. Another limitation is that VIGS fails to completely knockdown target gene expression. In some cases this imperfect knockdown may not manifest in a phenotypic change compared to the control plants. Since the VIGS vector is essentially an infectious plant pathogen, the silencing process along with virus infection may interfere with the plant metabolism which could affect the results of some studies. Therefore, it is always important to use negative control such as VIGS vector alone (Fig. 1) without any target insert in VIGS experiments to make sure that the observed phenotype is specifically due to the knockdown of a target gene. VIGS is an ideal method for preliminary identification of function of genes due to its quick and easy features. Following preliminary identification with VIGS, the most promising candidate genes could be further validated with other approaches.

Conclusion and Prospects

VIGS is an efficient forward and reverse genetic technique that exploits the plant's PTGS mechanism. It facilitates the down-regulation of single genes or gene families. A large number of plant viruses have been engineered as VIGS tools for gene function studies in a variety of plant species. In the last 20 years, VIGS has made a significant contribution to gene function studies in various plant processes. With the availability of more plant genome sequences in the post-genomic era, the improvement of existing VIGS vectors as well as the development of new VIGS vectors for additional plant species is required for VIGS-based functional genomic studies to achieve its full potential in plants.

Further Reading

- Bachan, S., Dinesh-Kumar, S.P., 2012. Tobacco rattle virus (TRV)-based virus-induced gene silencing. *Methods Molecular Biology* 894, 83–92.
- Burch-Smith, T.M., Anderson, J.C., Martin, G.B., Dinesh-Kumar, S.P., 2004. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *Plant Journal* 39, 734–746.
- Ding, X.S., Mannas, S.W., Bishop, B.A., *et al.*, 2018. An improved Brome mosaic virus silencing vector: Greater insert stability and more extensive VIGS. *Plant Physiology* 176, 496–510.
- Di Stilio, V.S., Kumar, R.A., Oddone, A.M., *et al.*, 2010. Virus-induced gene silencing as a tool for comparative functional studies in *Thalictrum*. *PLOS ONE* 5, e12064. Available at: <https://doi.org/10.1371/journal.pone.0012064>.
- Dommes, A., Gross, T., Herbert, D.B., Kivivirta, K.I., Becker, A., 2019. Virus-induced gene silencing: empowering genetics in non-model organisms. *Journal of Experimental Botany* 70, 757–770.
- Fu, D.Q., Zhu, B.Z., Zhu, H.L., Jiang, W.B., Luo, Y.B., 2005. Virus-induced gene silencing in tomato fruit. *Plant Journal* 43, 299–308.
- Jiang, C.Z., Chen, J.C., Reid, M., 2011. Virus-induced gene silencing in ornamental plants. *Methods in Molecular Biology* 744, 81–96.
- Kant, R., Dasgupta, I., 2019. Gene silencing approaches through virus-based vectors: speeding up functional genomics in monocots. *Plant Molecular Biology* 100, 3–18.
- Kumagai, M., Donson, J., Della-Cioppa, G., *et al.*, 1995. Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived RNA. *Proceedings of the National Academy of Sciences of the United States of America* 92, 1679–1683.
- Lacomme, C., 2015. Strategies for altering plant traits using virus-induced gene silencing technologies. *Methods in Molecular Biology* 1287, 25–41.
- Liu, Y., Schiff, M., Czymmek, K., *et al.*, 2005. Autophagy regulates programmed cell death during the plant innate immune response. *Cell* 121, 567–577.
- Liu, N., Xie, K., Jia, Q., *et al.*, 2016. Foxtail mosaic virus-induced gene silencing in monocot plants. *Plant Physiology* 171, 1801–1807.
- Lu, R., Malcuit, I., Moffett, P., *et al.*, 2003. High throughput virus-induced gene silencing implicates heat shock protein 90 in plant disease resistance. *EMBO Journal* 22, 5690–5699.
- Mei, Y., Zhang, C., Kernodle, B.M., Hill, J.H., Whitham, S.A., 2016. A Foxtail mosaic virus vector for virus-induced gene silencing in maize. *Plant Physiology* 171, 760–772.
- Nagaraj, S., Senthil-Kumar, M., Ramu, V.S., Wang, K., Mysore, K.S., 2016. Plant ribosomal proteins, RPL12 and RPL19, play a role in non-host disease resistance against bacterial pathogens. *Frontiers in Plant Science* 6, 1192.

- Ramegowda, V., Mysore, K.S., Senthil-Kumar, M., 2014. Virus-induced gene silencing is a versatile tool for unraveling the functional relevance of multiple abiotic-stress-responsive genes in crop plants. *Frontiers in Plant Science* 5, 323.
- Rosa, C., Kuo, Y.W., Wuriyanghan, H., Falk, B.W., 2018. RNA interference mechanisms and applications in plant pathology. *Annual Review Phytopathology* 56, 581–610.
- Senthil-Kumar, M., Mysore, K.S., 2011. New dimensions for VIGS in plant functional genomics. *Trends in Plant Science* 16, 656–665.
- Senthil-Kumar, M., Wang, M., Chang, J., *et al.*, 2018. Virus-induced gene silencing database for phenomics and functional genomics in *Nicotiana benthamiana*. *Plant Direct* 2, e00055.
- Wang, R., Yang, X., Wang, N., *et al.*, 2016. An efficient virus-induced gene silencing vector for maize functional genomics research. *Plant Journal* 86, 102–115.
- Yuan, C., Li, C., Yan, L., *et al.*, 2011. A high throughput Barley stripe mosaic virus vector for virus induced gene silencing in monocots and dicots. *PLOS ONE* 6, e26468. Available at: <https://doi.org/10.1371/journal.pone.0026468>.
- Yuan, X.Y., Wang, R.H., Zhao, X.D., Luo, Y.B., Fu, D.Q., 2016. Role of the tomato non-ripening mutation in regulating fruit quality elucidated using iTRAQ protein profile analysis. *PLOS ONE* 11, e0164335. Available at: <https://doi.org/10.1371/journal.pone.0164335>.
- Zhang, W., Zhang, H., Liu, K., *et al.*, 2017. Large-scale identification of *Gossypium hirsutum* genes associated with *Verticillium dahliae* by comparative transcriptomic and reverse genetics analysis. *PLOS ONE* 12 (8), e0181609. Available at: <https://doi.org/10.1371/journal.pone.0181609>.

Relevant Websites

<http://vigs.solgenomics.net/>
SGN-VIGS Sol Genomics Network.

<http://vigs.noble.org>
VIGS Database
Noble Research Institute.

Alfalfa Mosaic Virus (*Bromoviridae*)

L Sue Loesch-Fries, Purdue University, West Lafayette, IN, United States

© 2021 Published by Elsevier Ltd.

This is an update of J.F. Bol, Alfalfa Mosaic Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00635-X.

Nomenclature

aa Amino acid(s)

CP Capsid or coat protein

kb Kilobases; the size of a ssDNA or ssRNA molecule

kDa Kilodaltons; the size of a protein

MP Movement protein

MVBs Multi-vesicular bodies

nm Nanometer(s)

nt Nucleotide(s)

PABP Poly(A)-binding protein

Poly(A) Polyadenylated

TLS Transfer RNA-like structure

tRNA transfer RNA

UTR Untranslated region

Glossary

Multi-vesicular bodies Organelles bound by a membrane that contain intraluminal vesicles that mediate transport between the trans-Golgi network and vacuoles.

Poly(A)-binding protein Regulatory proteins that bind to the poly(A) tail of mRNAs.

T-DNA vector Plasmid with the T-DNA (transferred DNA) sequence of the Ti plasmid of *Agrobacterium tumefaciens*, which is transferred to plant cells where it can be transiently expressed or become integrated in the plant genome.

Transgenic P12 tobacco Tobacco plants that express P1 and P2 replicase proteins from nuclear genes inserted via a T-DNA vector.

History

Alfalfa mosaic virus (AMV) was identified in 1931 as the causal agent of an economically important disease in alfalfa. Purification of AMV around 1960 showed that virus preparations contained bacilliform particles of different length. Fractionation of these particles by sucrose-gradient centrifugation revealed that the four major components each contained a specific type of RNA, termed RNAs 1, 2, 3, and 4. Initially, an analysis of the biological activity of the viral nucleoproteins and RNAs resulted in a puzzle. A mixture of the three largest viral particles was fully infectious, but a mixture of RNAs 1, 2, and 3, purified from these particles, was not. At the RNA level, RNA4 was required in the inoculum to initiate infection. In 1971, it became clear that the AMV genome consisted of RNAs 1, 2, and 3, and that a mixture of these genomic RNAs became infectious only after addition of capsid or coat protein (CP) or its sub-genomic messenger, RNA4. High-affinity binding sites for CP were identified in the AMV RNAs in 1972 and could be localized to the 3' termini of the RNAs in 1978. From 1975 onwards it became clear that, similar to AMV, the ilarviruses required the CP to initiate infection. Moreover, the CPs of AMV and the ilarviruses could bind to the 3' termini of each other's RNAs, and could be freely exchanged in the initiation of infection. Currently, AMV is the type species of the genus *Alfavirus* and is classified together with the genus *Ilarvirus* in the family *Bromoviridae*.

Taxonomy and Classification

In addition to the genera *Alfavirus* and *Ilarvirus*, the family *Bromoviridae* contains the genera *Bromovirus*, *Cucumovirus*, *Anulavirus*, and *Oleavirus*. These viruses all have three positive-sense single-stranded genomic RNAs packaged into separate particles with sub-genomic, defective, or satellite RNAs. The particles are bacilliform or spherical. AMV has been placed into a separate genus from the other viruses because of its mode of transmission (aphids, pollen, and seed), its shape (bacilliform), and its very wide host range. In addition, AMV requires CP for the initiation of infection (termed 'genome activation'), whereas other members of the family do not, except for the ilarviruses. It has been suggested that the alfamoviruses and the ilarviruses be combined into a single genus, but this has yet to be resolved. The ilarviruses differ from AMV, however, by mode of transmission (thrips/pollen) and shape (irregular spheres). Many isolates of AMV are known which are closely related serologically and by nt sequence similarity. The complete nt sequences of AMV isolate 425 (Leiden [L] and Madison [M] isolates) were the first determined; GenBank now contains 11,602 full and partial sequences of many additional strains and isolates. The genetic diversity of the sequences is generally low with agro-ecological factors affecting their distribution.

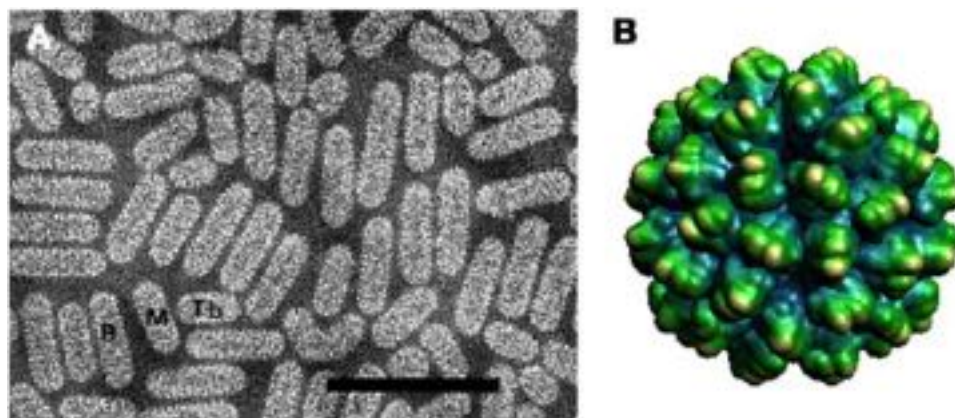


Fig. 1 Electron micrograph of Alfalfa mosaic virus (*Alfamovirus*, *Bromoviridae*). A. Bacilliform bottom particle, B, middle particle, M, and top particle Tb. Scale bar is 100 nm. (Courtesy of J. Bol). B. Cryo-electron micrograph of a $T=1$ AMV particle. Reproduced from <http://viperdb.scripps.edu>; Kumar, A., Reddy, V.S., Yusibov, V., *et al.*, 1997. The structure of Alfalfa mosaic virus capsid protein assembled as a $T=1$ icosahedral particle at 4.0-Å resolution. *Journal of Virology* 71, 7911–7916.

Particle Structure and Composition

Fig. 1 shows an electron micrograph of AMV. The four major classes of particles in AMV preparations are called bottom component (B), middle component (M), top component *b* (Tb), and top component *a* (Ta). B, M, and Tb are bacilliform and contain the genomic RNAs 1, 2, and 3, respectively. Ta contains two molecules of the sub-genomic RNA4 and can be subdivided into bacilliform Ta-b and spheroidal Ta-t particles. The bacilliform particles are all 19 nm wide and have lengths of 56 nm (B), 43 nm (M), 35 nm (Tb), and 30 nm (Ta-b) (**Fig. 1(A)**). The RNAs are encapsidated by a single type of CP, which in the case of strain AMV-L has a length of 220 aa (Mr 24.3 kDa). In solution, AMV CP occurs as dimers, which under appropriate conditions of pH and ionic strength form a $T=1$ icosahedral structures built from 30 dimers. This structure can be crystallized and has been studied by X-ray diffraction, cryoelectron microscopy, and image reconstruction methods (**Fig. 1(B)**). The CP was found to have the canonical eight-stranded β -barrel fold with the N- and C-terminal arms as extended chains. Dimer formation in the $T=1$ particle is based on the clamping of the C-terminal arms of the subunits.

From particle weight measurements and analysis of electron microscopic images, it was determined that the number of CP monomers in the major viral components is 60 p ($n \times 18$), n being 10 (B, 240 subunits), 7 (M, 186 subunits), 5 (Tb, 150 subunits), or 4 (Ta, 132 subunits). By gel electrophoresis, at least 13 minor components have been resolved which probably represent other n values and contain monomers of genomic RNAs, multimers of genomic RNAs or RNA4, or specific degradation products of RNA3. Although details of the arrangement of the protein monomers have not been established, electron microscopical studies indicate that the cylindrical parts of the bacilliform particles have a hexagonal surface lattice with dimers of CP associated with the twofold symmetry axes. The cylindrical portion of the particle is thought to be capped by two halves of an icosahedron by changing the axes from sixfold symmetry in the cylinder into axes of fivefold symmetry. Neutron scattering data suggested that the capsid structure of spheroidal Ta-t particles is represented by a deltahedron with 52-point group symmetry built from 120 subunits. The percentage of RNA in the particles decreases from 16.3 in B to 15.2 in Ta-b. The buoyant density in CsCl of the major components fixed with formaldehyde varies from 1.366 (Ta) to 1.372 (B) g cm^{-3} . The protein shell has an inner radius of 6.5 nm and an outside radius of 9.4 nm. The RNA is uniformly packed within the 6.5 nm radial limit, occupies about 20% of the interior volume available, and slightly penetrates the protein shell to stabilize the particles by protein–RNA interactions. The RNA is easily accessible to ribonucleases A and T1 through holes in the protein shell.

Genome Structure

The genome structure of AMV strain 425 is shown in **Fig. 2**. RNAs 1 and 2 encode the replicase proteins P1 and P2, respectively. P1 contains an N-terminal methyl-transferase-like domain and a C-terminal helicase-like domain, whereas P2 contains a polymerase-like domain. RNA3 is dicistronic and encodes the 5' proximal movement protein (MP) gene and the 3' proximal CP gene. MP and CP are both required for cell-to-cell transport of the virus. CP is translated from RNA4, which is identical in sequence to the 3' terminal 881 nt of RNA3. The length of the intercistronic region in RNA3 is 52 nt including the leader sequence of RNA4 of 36 nt. At the 5' end, all four AMV RNAs are capped. The organization of the leader sequence of RNA3 varies between strains. For strains M, S, L, and Y the length of this leader sequence is 240, 313, 345, and 391 nt, respectively. The increased length of the last three strains is due to the presence of direct repeats of 56 (S), 75 (L), or 149 (Y) nt. At their 3' termini, the viral RNAs contain a homologous sequence of 145 nt. The 3' 112 nt of this sequence may adopt two mutually exclusive conformations, one representing a strong CP binding site (CPB) and the other representing a transfer RNA-like structure (TLS) resembling the TLS of bromo- and cucumoviruses. The 112 nt CPB structure, consists of four hairpins

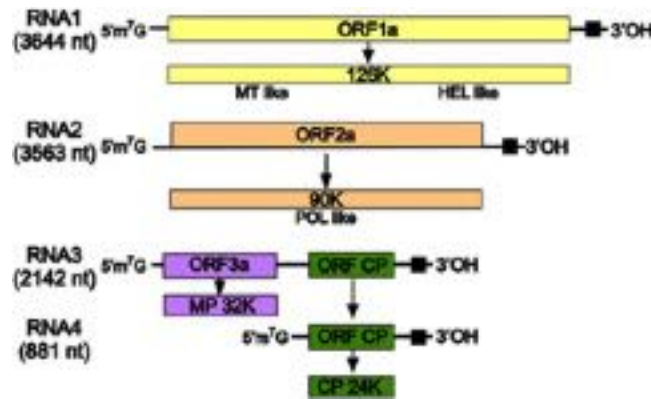


Fig. 2 Structure of the AMV genome. RNAs 1 and 2 are monocistronic and encode the replicase proteins P1 of ~110 kDa, and P2 of ~85 kDa. RNA3 is dicistronic and encodes a movement protein of ~34 kDa and coat protein of ~22 kDa. CP is translated from sub-genomic RNA4 (not shown). The gray squares represent the 3' terminal 112 nt of the RNAs, which can adopt either a tRNA-like structure or a structure with high affinity for CP. From: Bujarski *et al.*, 2019. Journal of General Virology 100, 1206–1207.

(designated A, B, C, and D from 3' to 5' end) flanked by tetranucleotide sequences AUGC (or a UUGC derivative). Base-pairing between a four nt sequence in the loop of the 5' proximal hairpin D and a four nt sequence in the stem of the 3' proximal hairpin A results in a pseudoknot interaction that generates the TLS structure. Structures similar to the AMV CPB and TLS conformations have been identified at the 3' termini of the RNAs of the ilarviruses. The TLS of AMV and ilarviruses cannot be charged with an aa as can the TLS of bromo- and cucumoviruses, but it can be recognized by the host enzyme that adds CCA to the 3' termini of cellular tRNAs (CTP/ATP: tRNA nucleotidyl transferase).

Interaction Between Viral RNA and Coat Protein

In vitro, the 3' untranslated region (UTR) of AMV RNAs can bind several dimers of CP. A minimal CP binding site consists of the 3'-terminal 39 nt of the RNAs with the structure 5'-AUGC-[hairpin B]-AUGC-[hairpin A]-AUGC-3'. This structure contains two overlapping binding sites for the N-terminus of the two subunits of a CP dimer. The consensus binding site in the RNA is UGC-[hairpin]-RAUGC (in which R is a purine). In addition to dimers of native CP, N-terminal peptides of CP can bind the 39 nt RNA fragment in a 2:1 ratio. The N-terminus of CP contains basic residues at positions 5, 6, 10, 13, 16, 17, 25, and 26, but only arginine-17 appears to be critical for binding of CP to RNA. This Arg residue is part of a Pro-Thr-x-Arg-Ser-x-x-Tyr (PTxRSxxY) RNA binding domain conserved among AMV and ilarvirus CPs.

A complex of the 3' terminal 39 nt RNA fragment and peptides corresponding to the N-terminal 26 aa of CP has been crystallized and its structure was solved to 3 Å resolution. Co-folding altered the structure of both peptide and RNA (**Fig. 3**). In the co-crystal, hairpins A and B are oriented at approximately right angles and each hairpin is extended by 2 bp formed between nt from adjacent AUGC sequences. If the AUGC motifs in the 39 nt fragment are numbered 1–3, starting from the 3' end of the RNA, hairpin B is extended by a duplex formed by base-pairing between the U and C residue of motif 3 and the A and G residue of motif 2. Similarly, hairpin A is extended by a duplex formed by base-pairing between the U and C residue of motif 2 and the A and G residue of motif 1. The two peptides in the co-crystal each form a helix with residues 12–26 ordered in peptide 1 and residues 9–26 ordered in peptide 2. The data has provided insight into the role of the PTxRSxxY-motif in RNA binding. *In vitro* selection of RNA fragments with high affinity for full-length AMV CP from a pool of randomized RNAs yielded fragments that maintained the unusual inter-AUGC base pairs observed in the crystal structure, but the primary sequences diverged from the wild-type RNA.

Translation of Viral RNA

Extension of the 3' end of AMV genomic RNAs with an artificial poly(A)-tail increased the basal level of infectivity of these RNAs 50-fold, compared to a 1000-fold increase caused by binding of CP to the RNAs. A role of CP in translation of viral RNAs was investigated by extension of the 3' end of a luciferase reporter RNA with the AMV 3' UTR [Luc-AMV]. CP stimulated translational efficiency of Luc-AMV 40-fold without affecting the half-life of Luc-AMV. GST-pull-down assays and Far Western blotting revealed that AMV CP specifically interacted with the eIF4G-subunit in the eIF4F initiation factor complex from wheat germ and with the eIFiso4G-subunit from the eIFiso4F complex in plants. This interaction allows the formation of a closed-loop structure for AMV RNAs similar to that formed by the interaction of poly(A) binding protein (PABP), the poly(A) tail of eukaryotic mRNAs, and the eIF4G-subunit during translation (**Fig. 4**). Therefore, AMV CP stimulated translation of viral RNAs by mimicking the PABP.

Protoplasts transfected with wild-type AMV RNA4 accumulated CP at a detectable level, although accumulation was 100-fold lower than in productively infected protoplasts. However, translation of an RNA4 mutant, R17A, which encodes a CP with

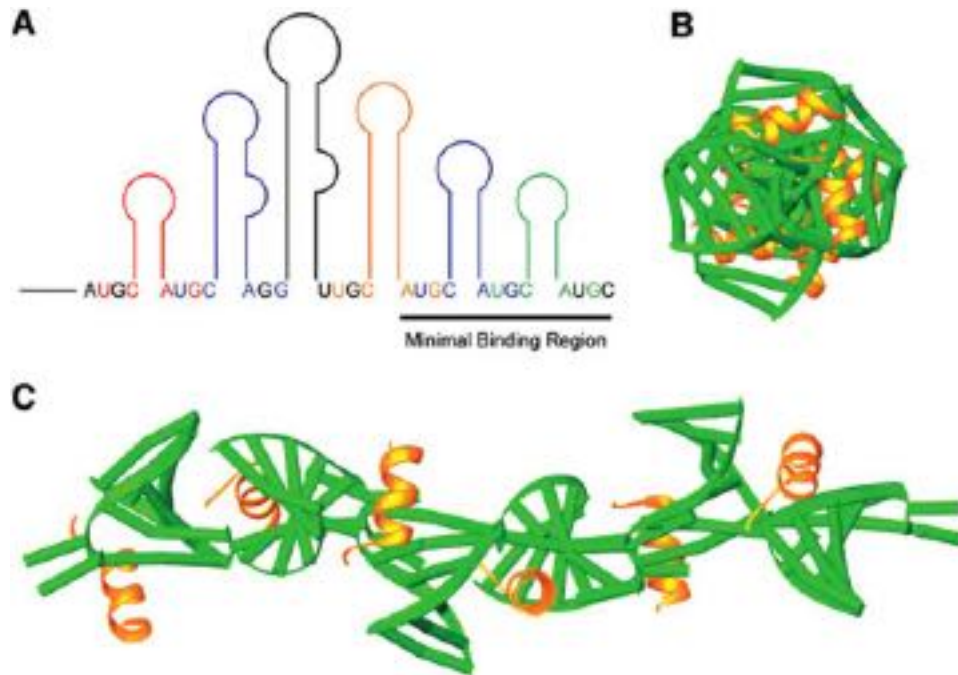


Fig. 3 Proposed structure of the entire 3' end of AMV RNA. A. Diagram of the 3' untranslated region (UTR) with color-coded U-C-hairpin-A-G CP-binding sites. B. End-on view of the simulated 3' UTR. C. View of the 3' UTR along its length. RNA is colored green; peptide is colored yellow. From Guogas, L.M., Filman, D.J., Hogle, J.M., Gehrke, L., 2004. Cofolding organizes alfalfa mosaic virus RNA and coat protein for replication. *Science* 306, 2108–2111.

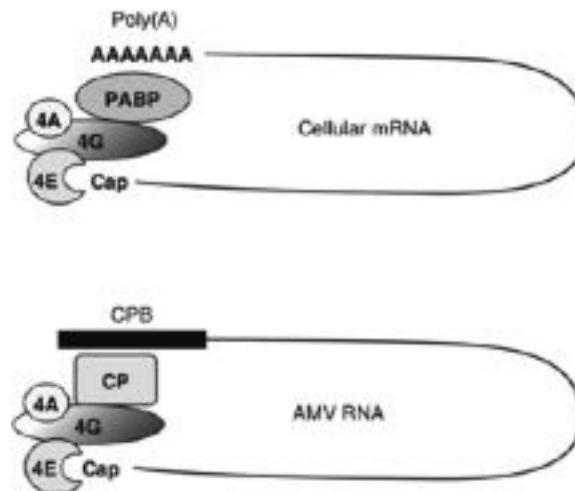


Fig. 4 Model for the role of coat protein in translation of AMV RNAs. Translation of cellular mRNAs is strongly enhanced by the formation of a closed-loop structure by interactions of the poly(A) binding protein (PABP) with the 3' poly(A) tail and with the eIF4G subunit of the eIF4F complex of initiation factors bound to the 5' cap structure (upper panel). AMV coat protein (CP) enhances translational efficiency of viral RNAs 40-fold by binding to the CP binding site (CPB) at the 3' end of the viral RNAs. The finding that CP also interacts with eIF4G indicates that CP mimics the function of PABP in formation of the closed-loop structure (lower panel).

arginine-17 replaced by alanine that is unable to bind to the 3' end of AMV RNAs, was below the detection level. Translation of mutant R17A could be rescued to wild-type levels by expression in *trans* of a CP that was functional in RNA binding. Also, translation of this mutant was rescued to wild-type levels by replacing its 3' UTR by the 3' UTR of brome mosaic virus (BMV) RNA4. BMV is the type species of the genus *Bromovirus* in the family *Bromoviridae* and its 3' UTR stimulates translation independently of CP, possibly by the binding of host factors.

Replication of Viral RNA

AMV replication complexes are associated with vacuolar membranes. Replicase protein P1 recruits P2 to multivesicular bodies, also called prevacuolar compartments. P1 and P2 localize to the tonoplast in the presence of full-length AMV RNA suggesting that functional replicase complexes assemble on multivesicular bodies (MVBs) then traffic to the tonoplast. Template-dependent replicase preparations were purified from infected normal and transgenic plants that constitutively express the replicase proteins. The purified replicase preparations, which specifically accept exogenous AMV RNA as template, were used to study viral minus-strand and plus-strand RNA synthesis *in vitro*. The identification in 1978 of high-affinity binding sites for CP at the 3' termini of AMV RNAs led to the hypothesis that binding of CP was required to permit initiation of minus-strand RNA synthesis. However, all experiments that were done with replicase preparations to test this hypothesis showed that CP was not required for the synthesis of AMV minus-strand RNA *in vivo* or *in vitro*. This led to the conformational switch model of AMV RNA replication.

An analysis of sequences of AMV RNAs, which direct minus-strand RNA synthesis by the purified viral replicase *in vitro*, showed that the entire 3' terminal homologous sequence of 145 nt is required for minus-strand promoter activity. This promoter consists of the TLS structure formed by the 3'-terminal 112 nt and a hairpin structure, termed hairpin E, between nt 112 and 145 from the 3' end of the RNAs. Hairpin E may be the primary element recognized by the viral replicase, and the TLS could serve to direct the replicase to the very 3' end of the template. If the TLS structure is disrupted by mutations affecting the pseudoknot or if the TLS is completely deleted, hairpin E directs initiation of minus-strand synthesis to a position located 5' from the hairpin. In the absence of the TLS, the mechanism of RNA synthesis directed by hairpin E is very similar to that directed by the sub-genomic promoter hairpin located in minus-strand RNA3, which directs plus-strand RNA4 synthesis. The finding that the sub-genomic promoter hairpin could be replaced by hairpin E without loss of infectivity illustrates the functional equivalence between the two hairpins. Although a knockout mutation of the CP gene does not affect AMV minus-strand RNA synthesis in infected protoplasts or in leaves expressing AMV replicase proteins, it results in a 100-fold drop in the accumulation of viral plus-strand RNAs. The expression of CP in leaves along with the replicase proteins not only enhanced the accumulation of replication-competent RNAs, but also the accumulation of replication-defective viral RNAs. Thus, it is possible that the stimulation of the accumulation of plus-strand RNAs by CP in infected cells reflects a role of CP in protection of the RNAs from degradation.

RNA3 can be replicated *in trans* by P1 and P2 proteins expressed from replication-defective cDNAs RNAs 1 and 2 in tobacco leaves or by P1 and P2 expressed in transgenic P12 tobacco. However, RNAs 1 and 2 are unable to make use of these expressed replicase proteins and are dependent for replication on their encoded proteins *in cis*. This requirement *in cis* explains why RNA3 can replicate in protoplasts of transgenic P12 plants without a requirement for CP, whereas replication of RNA1 or RNA2 in this system requires CP to enable translation of these RNAs. Moreover, replication of RNAs 1 and 2 is strictly coordinated. Mutations in either P1 or P2 ORF that is lethal to the function of the encoded proteins block the replication of both mutant and wild-type RNA1 or 2, but does not affect the replication of RNA3 by the expressed replicase. Apparently, replication of RNA1 controls replication of RNA2 and vice versa.

Virus Encapsidation and Movement

The high-affinity binding sites for CP at the end of RNAs 1 and 2 are dispensable for encapsidation of the RNAs. This was found by the expression of mutant RNAs 1 and 2 without their 3' UTRs along with wild-type RNA3 in leaves expressing replicase proteins. RNAs 1, 2, and 3 were all encapsidated into particles. Possibly, assembly of particles initiates on the identified internal CP binding sites in AMV RNAs.

The RNA3 encoded MP and CP are both required for cell-to-cell movement of AMV in infected plants. The MP is phosphorylated and forms tubular structures on the surface of infected protoplasts, suggesting that virus movement in plants involves transport of virus particles through tubules, which traverse the cell wall through modified plasmodesmata. Such a mechanism has been found for viruses of the families *Bromoviridae*, *Comoviridae*, *Caulimoviridae*, and *Bunyaviridae*. However, tubular structures filled with virus particles have not been observed in AMV infected leaf tissue. Moreover, a CP mutant defective in the formation of virions is able to move cell-to-cell at a reduced level. This suggests, along with information about other members of the *Bromoviridae*, that the transport of AMV moves cell-to-cell as viral ribonucleoprotein complexes that structurally differ from virions.

AMV MP is functional in the cell-to-cell transport of Tobacco mosaic virus, Brome mosaic virus, Prunus necrotic ringspot virus, Cucumber mosaic virus, and Cowpea mosaic virus. Conversely, the MPs of these viruses are competent in systemic transport of AMV as long as the MPs contain the last 44 CP-interacting aa of AMV MP at their C-termini. Analysis of AMV MP and those of other viruses in the family *Bromoviridae* suggests that the C-termini of these MPs enable specificity in virus transport by binding specific virus CPs and that the N-terminal portions of the MPs function in transport.

AMV MP interacts with *Arabidopsis* Patellin 3 (atPATL3) and Patellin 6 (atPATL6), which are implicated in membrane trafficking. The interaction interferes with the plasmodesmatal targeting of the MP and reduces AMV cell-to-cell movement. AMV replicated to higher levels in Patellin knockout mutants of *Arabidopsis* than in normal plants where replication is limited to fewer cells due to inhibition of cell-to-cell movement.

Role of Coat Protein in the AMV Replication Cycle

Natural infections by AMV are initiated by viral particles containing the genomic RNAs 1, 2, and 3, and sub-genomic RNA4. However, inoculation of plants in the laboratory with a mixture of the three genomic RNAs results in barely detectable levels of

AMV, but the infectivity of this mixture is increased 1000-fold by the addition of four to ten molecules of CP per RNA molecule. Alternatively, infectivity can be increased by including RNA4, the messenger for CP. It is assumed that AMV infection starts with the RNAs associated with one or more dimers of the viral CP. These RNA/CP complexes can be generated by partial disassembly of the particles in inoculated cells, by mixing purified viral RNAs and CP *in vitro*, or by translation of RNA4 in the inoculated cells and subsequent binding of *de novo* synthesized CP to the viral RNAs. The requirement of CP for translation of AMV RNAs was explained by the finding that AMV CP mimicks the function of the host cell poly(A) binding protein for efficient translation.

Two alternative models have been proposed for the role of CP in the replication of AMV RNA: a conformational switch model and a 3' organization model. The conformational switch model suggests that AMV RNA 3' ends form two mutually exclusive conformations, one that is extended and binds AMV CP and a second CP-free TLS pseudoknot conformation. This model proposes that CP enhances the translation of RNA 1 and 2 to make the replicase proteins and that the switch from translation to replication requires dissociation of CP from the 3' end of the RNAs to allow the formation of the TLS structure for minus-strand promoter activity. This model is based upon many experiments done *in vivo* and *in vitro* with mutant RNAs or replicase preparations, which found that CP was not needed for minus-strand synthesis. In contrast, the 3' organization model suggests that the CP organizes the 3' termini into compact structures for activation of early RNA replication. This was proposed based on the imperfect fit and uncertain functionality of the reported TLS at the 3' end of AMV RNAs and the crystal structure of the 3' terminal 39 nt of AMV RNA in complex with the RNA binding domain of the CP. This model proposes that a structurally organized RNA end is important for replication for all members of the family *Bromoviridae*. The compact structure is formed at the 3' end of some virus RNAs, such as those of BMV, by a TLS structure, while others, such as the 3' end of AMV and the ilarviruses, require CP binding to fold into the active structure.

Replicase proteins translated from inoculum RNAs 1 and 2 target the viral RNAs to vacuolar membranes where RNA synthesis takes place. The mechanism of the switch from minus-strand to plus-strand RNA synthesis is not yet clear. Although the promoter for plus-strand sub-genomic RNA4 synthesis is structurally similar to hairpin E in the minus-strand promoter, the promoters for plus-strand genomic RNA synthesis have not yet been characterized. The requirement of CP for efficient accumulation of plus-strand RNA in infected cells may reflect a role of CP in protection of the RNAs from degradation, but a role in plus-strand RNA synthesis has not yet been ruled out. There is growing evidence for AMV and other viruses in the family *Bromoviridae* that the many steps in the viral replication cycle are tightly linked.

Virus-Host Relationships

The AMV group is a large conglomerate of strains infecting a large number of susceptible hosts. This accounts for the tremendous range of symptoms displayed by AMV-infected plants. Mutations in the coat protein gene and 5' UTR of RNA3 have been shown to affect symptom formation in tobacco. Cytological modifications in AMV-infected plants occur only in cells of organs showing symptoms. In these cells, fragmentation of the ground cytoplasm and an increased accumulation of membrane-bound vesicles has been observed. Sometimes the lamellar system of chloroplasts is affected and invaginations of the nuclear membrane have been reported. The MP protein localizes to plasmodesmata. It interacts with atPATL3 and atPATL6, which are normally located at the cell periphery, to form complexes that were at the plasmodesmata and also at the cell periphery. Overexpression of atPATL3 and atPATL6 reduces the plasmodesmatal localization of AMV MP thereby reducing the spread of infection. When expressed alone in protoplasts, the P1 protein is found on MVBs, while the P2 protein is throughout the cytosol. When P1 and P2 are expressed together, both are found on MVBs; when both proteins are expressed with a functional AMV RNA, the proteins localize to the vacuolar membrane, the tonoplast, where replication takes place. Following encapsidation, virus particles are found throughout the cytoplasm with a few records of particles in the nucleus. Depending on the strain, the particles form different types of intracellular aggregates.

Recent studies have shown that the CP of AMV is multifunctional and interacts with a number of host proteins. Chloroplast proteins are increasingly found to be associated with virus proteins; *Arabidopsis* PsbP (atPsbP) protein is one example. This nuclear-encoded component of the oxygen-evolving complex of photosystem II was identified in a yeast two-hybrid screen to interact with AMV CP. Newly synthesized atPsbP was sequestered in the cytosol during virus infection, whereas it localizes to the chloroplasts in healthy cells. Overexpression of atPsbP inhibited virus replication suggesting that atPsbP has a role in plant defenses. AMV CP also interacts with the transcription factor, ILR3, in both tobacco and *Arabidopsis*. This interaction takes place in the nucleus where CP-interaction moves ILR3 to the nucleolus in which CP had previously been found. CP has a nucleolar localization signal in its N-terminus and interacts with importin- α to gain access to the nucleus. It is suggested that ILR3-CP interaction leads to a down-regulation of a host factor that is involved in host defense modulation. AMV CP also interacts with atALKBH9B, recently identified as the first demethylase described in plants. Such enzymes have broad regulatory roles in RNA biology and can remove m⁶A in virus RNAs to modulate infectivity of the virus. atALKBH9B also interacts with AMV RNA, whose sequence contains three of the common m⁶A consensus motifs. Down-regulation of atALKBH9B in infected plants increased the abundance of m⁶A in AMV RNA and decreased infection.

Host Range and Economic Significance

AMV occurs worldwide. Strains of this virus have been found in natural infections of about 305 plant species representing 22 families. The experimental and natural host ranges include over 600 species in 70 families. Recently, 68 *Arabidopsis thaliana* ecotypes were analyzed for their susceptibility to AMV infection. In total 39, ecotypes supported both local and systemic infection, 26 ecotypes supported only local infection, and three ecotypes could not be infected.

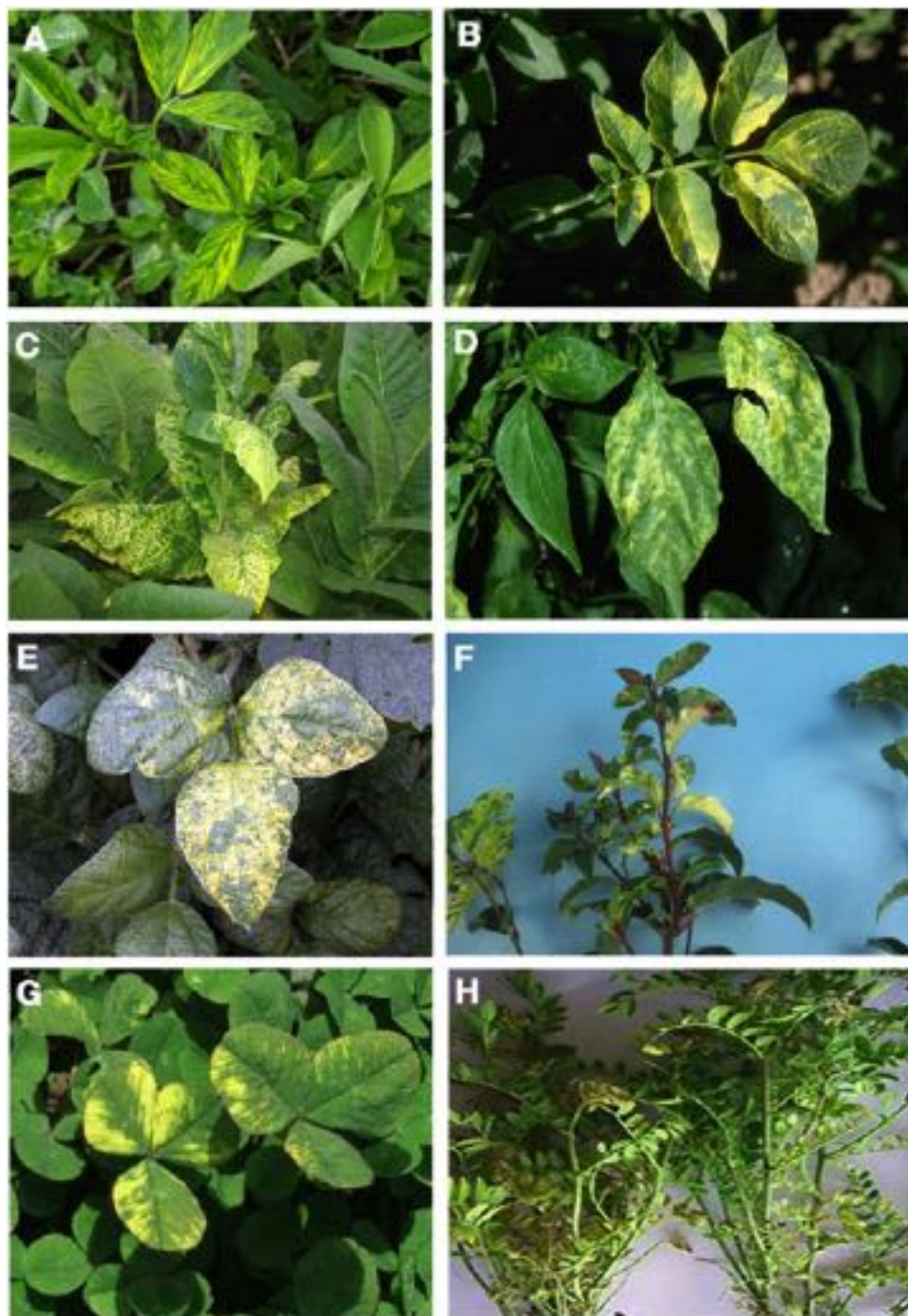


Fig. 5 Symptoms caused by AMV infection. (A) Interveinal chlorosis on alfalfa plants; (B) Mottle on potato; (C) Vein clearing on Burley tobacco; (D) Mosaic on pepper; (E) Mosaic on soybean; (F) Chlorosis on viburnum; (G) Interveinal chlorosis on white clover; (H) Tip necrosis on lentil. From: (A) South Dakota State University, extension.sdstate.edu; (B) Howard F. Schwartz, Colorado State University, Bugwood.org; (C) Kenneth Seebold, University of Kentucky, burleytobaccoextension.ca.uky.edu; (D) Whitney Cranshaw, Colorado State University, Bugwood.org; (E) Craig Grau, University of Wisconsin, cropprotectionnetwork.org; (F) Tom Creswell, Purdue University; (G) Barenbrug agriseeds, agriseeds.co.nz; (H) agriculture.vic.gov.au.

AMV causes significant losses in forage crops, can reduced winter survival, and increase the susceptibility to other pathogens. AMV is an economically important pathogen in temperate pulses (chickpeas, faba beans, field peas, and lentils), pasture legumes (alfalfa, and clovers), and may infect pepper, tobacco, tomato, soybean, potato, celery, weeds, and ornamental plants. In alfalfa, losses have been reported up to 30% and seed production up to 69%. In soybean, AMV infection along with soybean mosaic virus causes disease synergism with the development of severe symptoms. Such co-infection is of concern in areas where the soybean

aphid is present. Time of infection is important in yield loss caused by AMV. For example, late infection of faba bean can cause up to 45% of seed loss.

Symptoms caused by AMV vary depending upon the virus strain and host, but include: yellow mosaic and mottle, necrosis, stunting, leaf distortion, leaf rolling, vein-clearing, chlorotic spots, line patterns, and ringspots. The environment plays a role in symptom production. Infected field plants are more symptomatic in cool spring and autumn weather and may be symptomless in warm/hot summers. [Fig. 5](#) shows the symptoms for a variety of infected plants.

Transmission

AMV is easily mechanically transmissible. However, field spread occurs predominantly by aphid transmission. At least 15 aphid species are known to transmit the virus in a stylet-borne or non-persistent manner. Acquisition of the virus occurs within 10–30 s of feeding and is followed by immediate transmission without a latent period. The ability of the aphid to continue to transmit is lost within 1 h. The variability of individual aphid species in their capacity to transmit different AMV strains suggests a specific virus–vector relationship which is probably governed by the structural properties of the CP. Seed transmission of AMV has been reported for alfalfa and other plant species with rates of transmission varying from 0.1% to 50%. Transmission of the virus between plants by parasitic dodder has been observed with five *Cuscuta* species.

Epidemiology and Control

Genes conferring resistance to AMV have been reported in alfalfa (*amv-1* gene), soybean (*rav1* gene), and tomato (*am* gene). These, however, have not been deployed in commercial crops. Control of AMV in alfalfa is primarily achieved by using virus-free seed and avoiding reservoir hosts of the virus. Because AMV infection occurs naturally in many different plant species, control is difficult.

Transgenic tobacco, alfalfa, and white clover plants expressing the CP gene of AMV were found to be highly resistant to the virus when mechanically inoculated. The resistance was clearly protein-mediated because plants with the highest levels of CP accumulation were the most resistant. A mutation in the transgene that affected the N-terminal sequence of the encoded CP destroyed resistance to the wild-type virus but the mutant transgene conferred resistance to virus expressing the mutant CP. As yet, transgenic resistance to AMV has not been commercialized.

AMV is primarily spread by aphids. The incidence of AMV in alfalfa pastures can be as high as 50%–98%. Susceptible crops growing outdoors or in greenhouses close to alfalfa fields are at risk of infection. Because of AMV's wide host range, many weed species and ornamental plants are infected, which provides reservoirs for the aphids to acquire the virus. AMV is also seed-borne. It has been spread throughout the world by importation of infected alfalfa seed lots. AMV is also mechanically transmitted. It can be transmitted to healthy plants by handling infected plants or seedlings.

Control of AMV in alfalfa is primarily achieved by using virus-free seed and transplants, managing weeds, and avoiding reservoir hosts of the virus. Insecticides are largely ineffective in field management of AMV because aphids do not have to colonize a plant to acquire or transmit the virus; test feedings are sufficient. However, control of aphids early in the season may avoid buildup of populations. Various cultivars of crop plants may vary in susceptibility; however, there are no AMV-resistant commercial varieties.

See also: Brome Mosaic Virus (*Bromoviridae*). Bromoviruses (*Bromoviridae*). Cucumber Mosaic Virus (*Bromoviridae*)

Further Reading

- Balasubramaniam, M., Ibrahim, A., Kim, B.-S., Loesch-Fries, L.S., 2006. *Arabidopsis thaliana* is an asymptomatic host of Alfalfa mosaic virus. *Virus Research* 121, 215–219.
- Balasubramaniam, M., Kim, B.-S., Hutchens-Williams, H.M., Loesch-Fries, L.S., 2014. The photosystem II oxygen-evolving complex protein PsbP interacts with the coat protein of Alfalfa mosaic virus and inhibits infection. *Molecular Plant Microbe Interactions* 27, 1107–1118.
- Bol, J.F., 2005. Replication of alfamo- and ilarviruses: Role of the coat protein. *Annual Review of Phytopathology* 43, 39–62.
- Guogas, L.M., Filman, D.J., Hogle, J.M., Gehrke, L., 2004. Cofolding organizes Alfalfa mosaic virus RNA and coat protein for replication. *Science* 306, 2108–2111.
- Hull, R., 1969. Alfalfa mosaic virus. *Advances in Virus Research* 15, 365–433.
- Jones, R.A.C., 2012. Virus diseases of annual pasture legumes: Incidences, losses, epidemiology, and management. *Crop & Pasture Science* 63, 399–418.
- Krab, I.M., Caldwell, C., Gallie, D.R., Bol, J.F., 2005. Coat protein enhances translational efficiency of Alfalfa mosaic virus RNAs and interacts with the eIF4G component of initiation factor eIF4F. *Journal of General Virology* 86, 1841–1849.
- Kumar, A., Reddy, V.S., Yusibov, V., *et al.*, 1997. The structure of Alfalfa mosaic virus capsid protein assembled as a T=1 icosahedral particle at 4.0-Å resolution. *Journal of Virology* 71, 7911–7916.
- Martínez-Pérez, M., Aparicio, F., López-Gresa, M.P., *et al.*, 2017. *Arabidopsis* m⁶A demethylase activity modulates viral infection of a plant virus and the m⁶A abundance in its genomic RNAs. *Proceedings of the National Academy of Sciences of the United States of America* 114, 10755–10760.
- Navarro, J.A., Sanchez-Navarro, J.A., Pallas, V., 2019. Key checkpoints in the movement of plant viruses through the host. *Advances in Virus Research* 104, 2–63.
- Olsthoorn, R.C.L., Mertens, S., Brederode, F.T., Bol, J.F., 1999. A conformational switch at the 3' end of a plant virus RNA regulates viral replication. *EMBO Journal* 18, 4856–4864.
- Pallas, V., Aparicio, F., Herranz, M.C., Sanchez-Navarro, J.A., Scott, S.W., 2013. The molecular biology of ilarviruses. *Advances in Virus Research* 85, 140–181.
- Petrillo, J.E., Rocheleau, G., Kelley-Clarke, B., Gehrke, L., 2005. Evaluation of the conformational switch model for Alfalfa mosaic virus RNA replication. *Journal of Virology* 79, 5743–5751.

Alphaflexiviruses (*Alphaflexiviridae*)

Sergey Y Morozov and Alexey A Agranovsky, Lomonosov Moscow State University, Moscow, Russia

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
Avr Avirulence
CP Coat protein
DCL Dicer-like
ER Endoplasmic reticulum
HEL RNA helicase domain
kb Kilobases
kDa Kilo Daltons
MP Movement protein
MTR Methyl-transferase domain
nm Nanometer(s)
NMD Nonsense-mediated degradation pathways of mRNA decay

nt Nucleotide(s)
ORF Open reading frame
PAMP Pathogen-associated molecular pattern
PCNA Proliferating cell nuclear antigen
PD Plasmodesmata
POL RNA-dependent RNA polymerase domain
ROS Reactive oxygen species
satRNA Satellite RNA
SFI helicase Superfamily I helicase
sgRNA Sub-genomic RNA
TGB Triple gene block
VSR Viral suppressor of RNA silencing

Glossary

3'-co-terminal sub-genomic RNA Virus-specific RNA corresponding to the 3'-terminal portion of virus RNA genome.
Cap m⁷GTP moiety at the 5' end of the genomic virus RNA.
Cell-to-cell movement Substance movement between adjacent cells.
Coat protein Virion protein directly covering the virus genomic nucleic acid in the particle.
Endoplasmic reticulum Cell network of endomembrane tubules.

Long-distance movement Substance virus movement through the plant phloem system.
Movement protein Virus protein responsible for cell-to-cell movement.
Open reading frame Sequence of codons (usually not less than 15–20 codons) containing no termination triplets.
Triple gene block The module of three overlapping genes involved into cell-to-cell and long distance movement of the virus.

Introduction

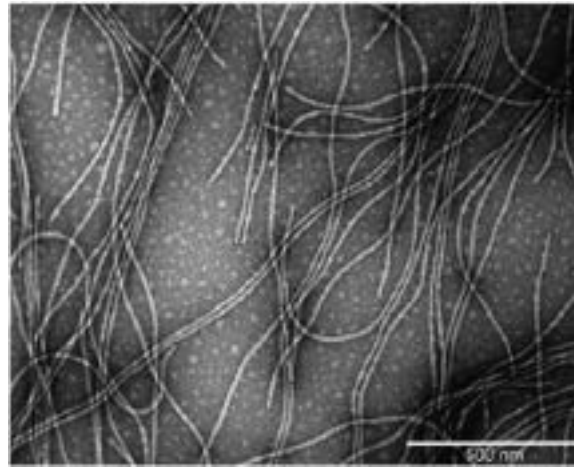
The family *Alphaflexiviridae* (order *Tymovirales*) contains seven genera (*Allexivirus*, *Botrexvirus*, *Lolavirus*, *Mandarivirus*, *Platypuvirus*, *Potexvirus*, and *Sclerodarnavirus*) and five species unassigned to a genus (Tables 1 and 2). These viruses have flexuous filamentous particles (Fig. 1) and infect higher plants or fungi. The 5'-capped and 3'- polyadenylated monopartite positive-stranded RNA genomes of the *Alphaflexiviridae* members are 5.9–9.0 kb in length. The 5'-terminal open reading frame (ORF) encodes the replicase with conserved methyltransferase (MTR), helicase (HEL), and RNA-dependent RNA polymerase (POL) domains, a configuration typical of the alphavirus-like superfamily of plant and animal viruses (Fig. 2). Most plant-infecting alphaflexivirus genomes encompass the module of three overlapping ORFs (Triple Gene Block, TGB) that are necessary for the cell-to-cell movement. Coat protein (CP) is encoded by the gene located at or close to the 3'-end of the genome (Fig. 2). Alphaflexiviruses produce 3'-coterminal sub-genomic RNAs that allow expression of the CP and other accessory proteins. Some alphaflexiviruses amplify to high yields in host plants, and are transmitted mechanically, whereas others have mites and insect vectors. Plant-infecting family members have been found in a wide range of mono- and dicotyledonous plants species, and the majority of them have only mild effects on their host.

Taxonomy, Phylogeny of Family Members, and Evolution

The genera of *Alphaflexiviridae* show close similarity of the encoded proteins (mostly the replicases) and share common features of the virion structure and genome organization. These properties are shown in Fig. 2 and Table 1. The affiliation with *Alphaflexiviridae* is based on the replicase amino acid sequence, and viruses from different genera usually have less than about 45% nt identity in the respective replicase genes.

Table 1 Some properties of the viruses in the genera of the family Alphaflexiviridae

Genus	Host	Virion length	Number of ORFs	Replicase size (kDa)	Coat protein size (kDa)
<i>Allexivirus</i>	Higher plants	800 nm	6	170–195	25–29
<i>Botrexvirus</i>	Fungi	720 nm	5	160	43
<i>Lolavirus</i>	Higher plants	640 nm	6	196	32
<i>Mandarivirus</i>	Higher plants	650 nm	6	187	34
<i>Platipuvirus</i>	Higher plants	800 nm	7	157	22
<i>Potexvirus</i>	Higher plants	470–580 nm	5	150–195	18–27
<i>Sclerodarnavirus</i>	Fungi	No virions	1	193	No CP


Fig. 1 Electron micrograph of negatively-stained virions of Alternanthera mosaic potexvirus. The bar = 500 nm. Courtesy of O.V. Karpova, Lomonosov Moscow State University. Department of Virology.

Among the plant-infecting Alphaflexiviridae five genera are currently recognized (*Potexvirus*, *Allexivirus*, *Lolavirus*, *Mandarivirus*, and *Platipuvirus*; Table 2). The most populated genus *Potexvirus* includes 38 members (Table 2). The next most populated genus *Allexivirus* contains 12 approved members. The genus *Lolavirus* contains two members, whereas each of the genera *Mandarivirus* and *Platipuvirus* includes only a single species. Two remaining fungus-infecting genera also include a single species (Table 2).

The single member of the genus *Botrexvirus* (*Botrytis virus X*) infects a filamentous fungus. The BVX genome lacks the movement protein genes, but encodes the CP and several unique proteins of unknown function. Another fungus-infecting genus *Sclerodarnavirus* is also represented by a single species whose genome contains only the replicase ORF (Fig. 2 and Table 2).

Phylogenetic trees based on the nucleotide and protein sequences depended on the genome region used for analysis. In a phylogenetic analysis of the whole replication protein, *Botrexvirus* and *Allexivirus* genera fall on a well-supported separated branch (Fig. 3). *Potexvirus*, *Mandarivirus* and *Lolavirus* genera form another branch in replicase phylogenetic tree, whereas *Platipuvirus* (DOSV) forms the most basal branch in this dendrogram (Fig. 3). If the genome nucleotide sequence is considered, the genomic RNAs of Alphaflexiviridae are most closely related to those of the other families in the order Tymovirales, namely Betaflexiviridae, Gammaflexiviridae and Tymoviridae. Strikingly, in this genome-based tree two genera of betaflexiviruses, namely *Vitivirus* and *Tepovirus*, show common root with the Alphaflexiviridae branch but not with the other members of Betaflexiviridae (Fig. 4). This is in line with replicase protein tree where *Vitivirus* and *Tepovirus* proteins form clear distinct branch separated from the rest of betaflexiviruses (Fig. 3).

The TGB proteins are related to the equivalent proteins encoded by some members of the family Betaflexiviridae and, more distantly, to those of the rod-shaped viruses in the family Virgaviridae (genera *Hordeivirus*, *Pecluvirus*, and *Pomovirus*).

Virion Structure

Typical alphaflexiviruses have flexuous filamentous particles (12–13 nm in diameter and up to 800 nm in length) with helical symmetry (pitch of ca. 3.4 nm) (Fig. 1). No lipids have been reported in the virions of alphaflexiviruses. Usually no carbohydrates are associated with the virions. The potexvirus and lolavirus CPs are reportedly glycosylated. The viral capsid of most members of the family (except lolaviruses) is composed of identical CP subunits ranging in size from 18 to 43 kDa. In allexiviruses, the 42 kDa protein (encoded in the third ORF from the 3' end; Fig. 2) is a minor virion component non-uniformly distributed along the particle. The particles of lolaviruses are composed of equimolar amounts the larger and smaller versions of the CP (32 kDa and

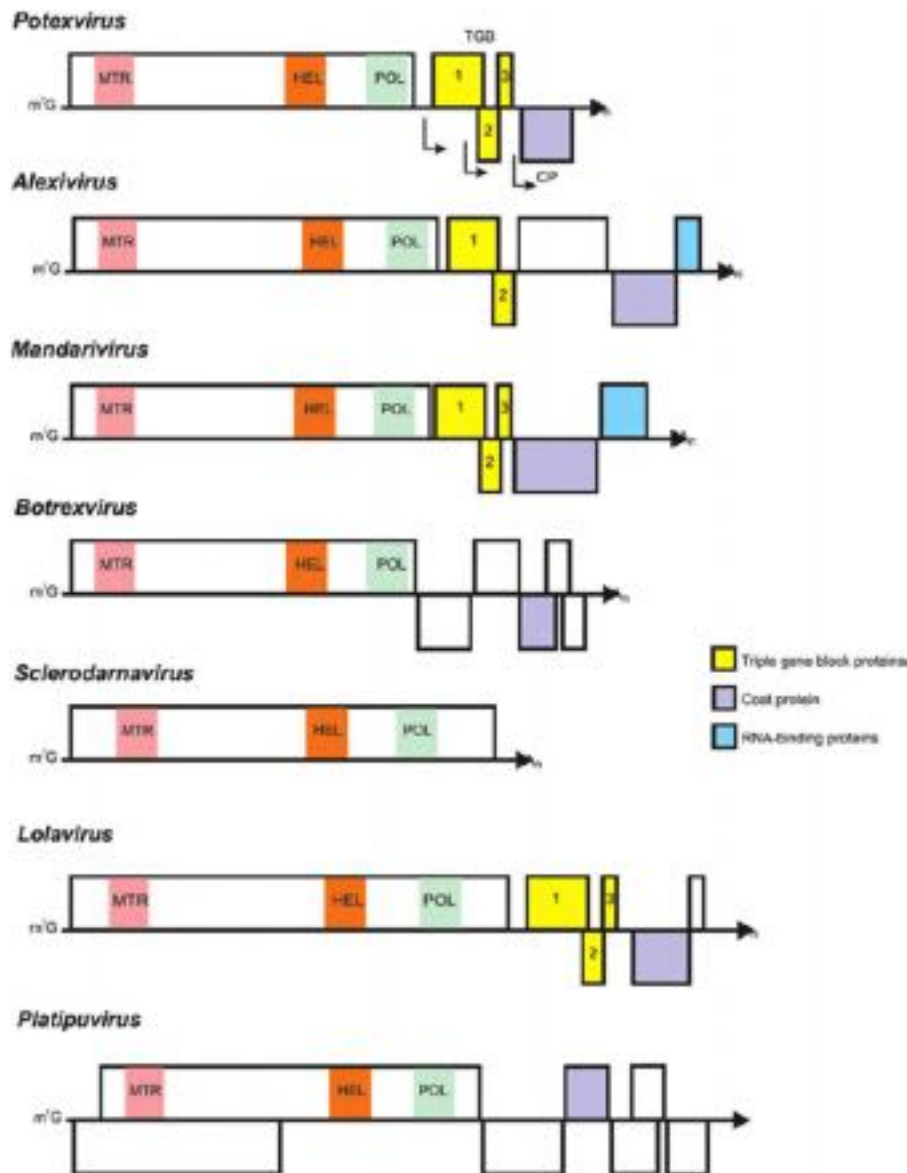


Fig. 2 Schematic representation of the genome structure of the *Alphaflexiviridae* (*Potexvirus*, Potato virus X; *Alexivirus*, Shallot virus X; *Mandarivirus*, Citrus yellow vein clearing virus; *Botrexvirus*, Botrytis virus X; *Sclerodarnavirus*, *Sclerotinia sclerotiorum* debilitation-associated virus; *Lolavirus*, Lolium latent virus; *Platipuvirus*, Donkey orchid symptomless virus). Arrowheads denote the 3'-termini of genomic RNAs; broken arrows – the start and direction of the potexvirus sub-genomic RNAs. Open reading frames (ORFs) are shown by boxes. MTR, HEL, and POL stand for the conserved methyl-transferase, NTPase/helicase, and RNA polymerase domains, respectively. TGB (triple gene block), CP (coat protein gene), and the genes for RNA-binding proteins are shown as boxes with different fills.

28 kDa, respectively). The 28 kDa CP seemingly originates from internal translation initiation of the coat protein gene. It should be noted that the sclerodarnavirus exists only as a capsid-less single-stranded and double-stranded RNA, making an intriguing parallel with capsid-less RNA agents associated with chestnut blight hypovirulence (family *Hypoviridae*) that code for the replicase orthologs of plant picorna-like viruses of the family *Potyviridae* (Fig. 2).

Genome Organization

Genomic RNAs of plant alphaflexiviruses typically have five or six ORFs. The 5' terminal ORF1 encodes the viral replicase (Fig. 2). The 3' proximal ORFs encode the movement proteins (TGB proteins), the CP, and an accessory zinc-binding and RNA-binding protein (Fig. 2). The last protein is encoded only in genera *Alexivirus* and *Mandarivirus*. ORF2 to ORF5 are expressed via the 3' co-terminal sgRNAs (Fig. 2).

Table 2 List of genera, virus species, acronyms, NCBI accession # and Genome length (nt) in the family *Alphaflexiviridae*. Type species are written in bold

Genus/Species name	Acronym	Accession #	Length (nt)
Potexvirus			
<i>Actinidia virus X</i>	AVX	NC028649	6888
<i>Allium virus X</i>	AIVX	NC012211	7118
<i>Alstroemeria virus X</i>	AlsVX	NC007408	7009
<i>Alternanthera mosaic virus</i>	AltMV	NC007731	6607
<i>Asparagus virus 3</i>	AV3	NC003400	6985
<i>Babaco mosaic virus</i>	BabMV	NC036587	6692
<i>Bamboo mosaic virus</i>	BaMV	NC001642	6366
<i>Cactus virus X</i>	CVX	NC002815	6614
<i>Cassava common mosaic virus</i>	CsCMV	NC001658	6376
<i>Cassava virus X</i>	CsVX	NC034375	5879
<i>Clover yellow mosaic virus</i>	CYMV	NC001753	7015
<i>Cymbidium mosaic virus</i>	CymMV	NC001812	6227
<i>Foxtail mosaic virus</i>	FoMV	NC001483	6151
<i>Hosta virus X</i>	HVX	NC011544	6528
<i>Hydrangea ringspot virus</i>	HdRSV	NC006943	6185
<i>Lettuce virus X</i>	LeVX	NC010832	7212
<i>Lily virus X</i>	LVX	NC007192	5823
<i>Malva mosaic virus</i>	MaMV	NC008251	6858
<i>Mint virus X</i>	MVX	NC006948	5914
<i>Narcissus mosaic virus</i>	NMV	NC001441	6955
<i>Nerine virus X</i>	NVX	NC007679	6582
<i>Opuntia virus X</i>	OpVX	NC006060	6653
<i>Papaya mosaic virus</i>	PapMV	NC001748	6656
<i>Pepino mosaic virus</i>	PepMV	NC004067	6450
<i>Phaius virus X</i>	PhVX	NC010295	5816
<i>Pitaya virus X</i>	PiVX	NC024458	6677
<i>Plantago asiatica mosaic virus</i>	PlaMV	NC003849	6128
<i>Potato aucuba mosaic virus</i>	PAMV	NC003632	7059
Potato virus X	PVX	NC011620	6435
<i>Schlumbergera virus X</i>	SchVX	NC011659	6633
<i>Senna mosaic virus</i>	SenMV	NC030746	6775
<i>Strawberry mild yellow edge virus</i>	SMYEV	NC003794	5966
<i>Tamus red mosaic virus</i>	TRMV	NC016003	6495
<i>Tulip virus X</i>	TVX	NC004322	6056
<i>Turtle grass virus X</i>	TGVX	NC040644	6278
<i>Vanilla virus X</i>	VVX	NC035205	6295
<i>White clover mosaic virus</i>	WCIMV	NC003820	5845
<i>Yam virus X</i>	YVX	NC025252	6158
<i>Zygocactus virus X</i>	ZyVX	NC006059	6624
Allexivirus			
<i>Alfalfa virus S</i>	AVS	NC034622	8349
<i>Arachis pintoi virus</i>	ApV	NC032104	7599
<i>Blackberry virus E</i>	BVE	NC015706	7718
<i>Garlic mite-borne filamentous virus</i>	GarMbFV	NC038864	762 ^a
<i>Garlic virus A</i>	GarV-A	NC003375	8660
<i>Garlic virus B</i>	GarV-B	NC025789	8336
<i>Garlic virus C</i>	GarV-C	NC003376	8405
<i>Garlic virus D</i>	GarV-D	NC022961	8424
<i>Garlic virus E</i>	GarV-E	NC004012	8451
<i>Garlic virus X</i>	GarV-X	NC001800	8106
Shallot virus X	ShVX	NC003795	8832
<i>Vanilla latent virus</i>	VLV	NC035204	7462
Lolavirus			
<i>Lolium latent virus</i>	LoLV	NC010434	7674
Mandarivirus			
<i>Citrus yellow vein clearing virus</i>	CYVCV	NC026592	7531
Indian citrus ringspot virus	ICRSV	NC003093	7560

(Continued)

Table 2 Continued

Genus/Species name	Acronym	Accession #	Length (nt)
Platypuvirus			
Donkey orchid symptomless virus	DOSV	NC022894	7838
Botrexvirus			
Botrytis virus X	BVX	NC005132	6966
Sclerodarnavirus			
Sclerotinia sclerotiorum debilitation-associated RNA virus	SSDaRV	NC007415	5470
Unassigned species			
<i>Insect-associated alphaflexivirus 3</i>	IaAIV-3	MN203145	796 ^a
<i>Euonymus yellow vein virus</i>	EuYVV	NC035190	7279
<i>Potexvirus</i> sp.	—	NC040842	5839
<i>Insect-associated alphaflexivirus 1</i>	IaAIV-1	MN203143	4165
<i>Insect-associated alphaflexivirus 2</i>	IaAIV-2	MN203144	868 ^a

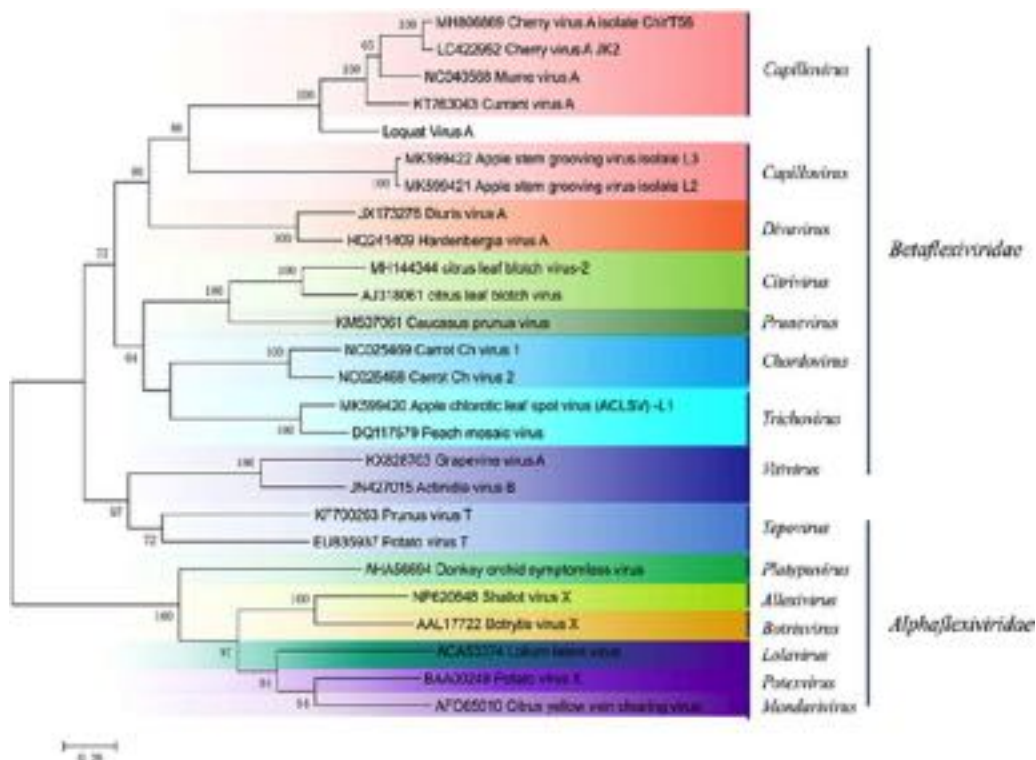
^apartial sequence.

Fig. 3 Maximum-likelihood tree derived from the complete amino acid sequences of the replicases of representative members of the families *Betaflexiviridae* and *Alphaflexiviridae*. Names of viruses and GenBank accession numbers are shown in the tree. Bootstrap analysis was done using 1000 replicates in MEGA7. The newly identified virus in this study is indicated by a red star. The scale bar represents the genetic distance. Accession number for Loquat virus A genome is MK936045. After Liu, Q., Yang, L., Xuan, Z., *et al.*, 2020. Complete nucleotide sequence of Loquat virus A, a member of the family *Betaflexiviridae* with a novel genome organization. *Archives of Virology* 165, 223–226.

Notable specific feature of *Allexivirus* members is the absence of TGB3 gene and the presence, in place of TGB3, of the gene for a large protein of unknown function (Fig. 2). Even more distinctive features in genome structure have been found in an unusual *Platipuvirus*, Donkey orchid symptomless virus (DOSV) (Fig. 2). The ORF1 that overlaps the replicase gene has marginal similarity with MPs of plant tymoviruses (family *Tymoviridae*). The replicase gene (ORF2) shares domain organization of the encoded protein with other members of the family *Alphaflexiviridae*. A 44 kDa protein encoded by ORF3 shows low level of identity with myosin. The putative 25 kDa movement protein (MP) (ORF7) shares low level of identity with the 3A-like MPs of some members of the family *Tombusviridae* (*Dianthovirus*).

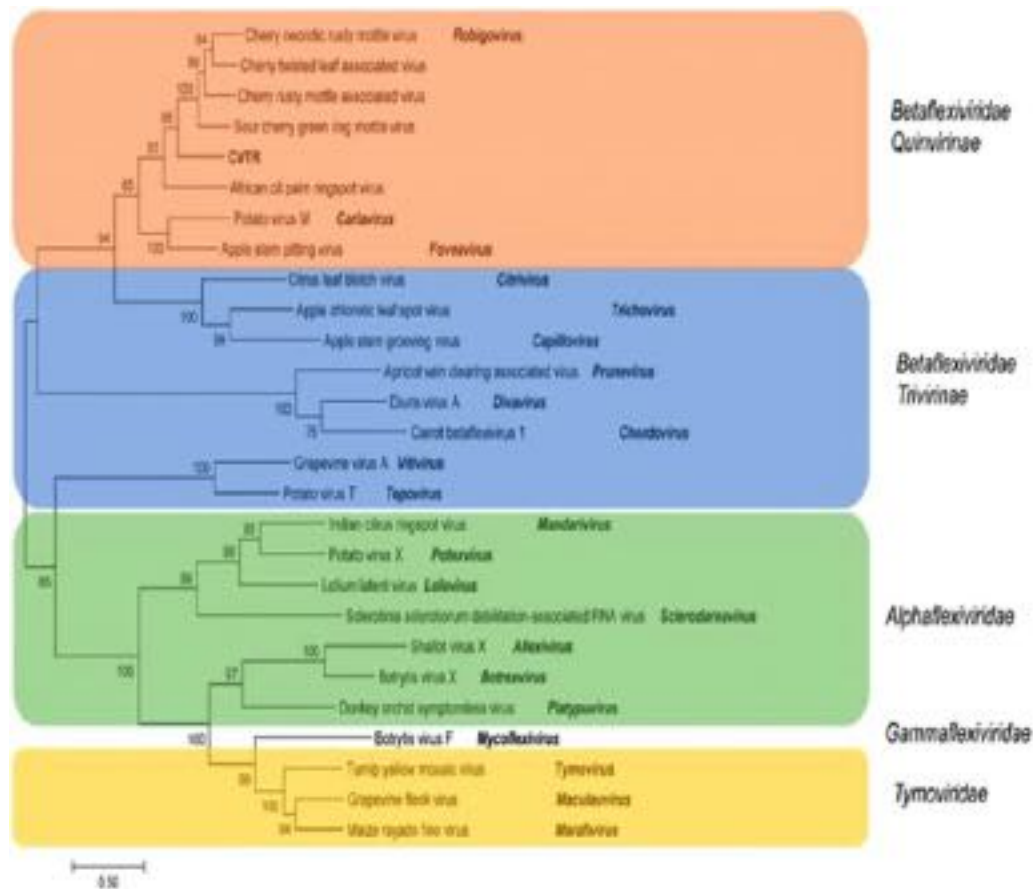


Fig. 4 Molecular Phylogenetic analysis of full-genome tymoviral reference sequences by Maximum Likelihood method based on the General Time Reversible model with a discrete Gamma distribution. The scale bar represents the genetic distance. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. After Çağlayan, K., Roumi, V., Gaze, L.M., *et al.*, 2019. Identification and characterization of a novel Robigovirus species from sweet cherry in Turkey. Pathogens 8, 57.

In genus *Potexvirus*, three 3' co-terminal sgRNAs are produced in infected plant cells (Fig. 2). The 2.1 kb species serves as mRNA for the TGB1 protein, whereas the 1.2-rb sgRNA directs expression of the TGB2 and TGB3 via leaky scanning. The 1.0-kb sgRNA encodes the coat protein. The genome organization of the TGB-containing alphaflexiviruses is quite similar to the TGB-containing betaflexiviruses belonging to the subfamily *Quinvirinae* (genera *Carlavirus*, *Robigovirus* and *Foveavirus*). Fungal alphaflexiviruses (genera *Botrexvirus* and *Sclerodarnavirus*) are drastically different with respect to their genome architecture. The capsid-less sclerodarnavirus encodes the replicase gene only, whereas *Botrytis virus X* contains six genes encoding replicase, CP and four proteins with unknown function (Fig. 2).

Properties and Functions of Gene Products

The ORF1 product contains the invariant array of protein domains involved in RNA capping (MTR), duplex unwinding (HEL), and RNA synthesis (POL) (Fig. 2). The POL domain of potexviruses is vital for the functioning of the viral replication complex. It contains the typical signature of plus-RNA viral RdRp – tripeptide GDD. Mutational analysis of potexvirus genomes unambiguously showed that during the catalysis of polymerization reaction this motif has the essential role. The POL domain of potexviruses also recognizes the cis-signals in the 3'-terminal region of both the positive and negative strands of RNA genome (see also below). These specific protein–RNA interactions allow correct initiation of the replication.

The replicative HEL domain of potexviruses locates in the middle of replicase protein and belongs to SFI helicase family. The purified HEL is able to unwind dsRNAs and to remove the γ -phosphate from nucleoside triphosphates. Both of these activities depend on Mg^{2+} or Mn^{2+} cations. Mutations of the SFI conserved motifs I, II, III, or VI resulted in inhibition of these activities. The RNA binding activity of HEL is cooperative and is based on conserved motifs I and II as well as on poorly conserved aa sequences.

The N-terminal viral replicase MTR domain was proposed to have an AdoMet-dependent guanylyl-transferase activity which transfers the methyl group of AdoMet to GTP, leading to m^7 GTP formation. The m^7 GMP moiety of m^7 GTP is then transferred to the 5' end of the genomic RNA via the covalent [m^7 GMP-enzyme] intermediate. Analogous cap formation mechanisms have been found also in other members of the alphavirus-like superfamily of positive strand RNA viruses. Importantly, the first two domains

of potexvirus replicase (MTR and HEL) work in a concert to form the 5' cap-structure on the nascent viral genomic and sub-genomic RNAs.

The potexvirus TGB1 product also contains an NTPase-helicase domain of SF1 with seven conserved motifs (I, Ia, II, III, IV, V, and VI). Motif I ("Walker A") includes typical GKS/T tripeptide, and motif II ("Walker B") is involved in binding of ATP and Mg^{2+} . The TGB1 protein is necessary for the cell-to-cell movement and also functions as a suppressor of RNA silencing. It seems that these activities of the TGB1 are uncoupled. In recent years it was documented that potexvirus TGB1 also participates in formation of virus replication factories in infected cells. Thus, the PVX TGB1 is involved in reorganizing actin cytoskeleton and endoplasmic reticulum (ER) into the membrane-bound virus replication complexes (VRC). Moreover, TGB1 participates in positioning the VRC close to plasmodesmata thus helping to establish connection between replication and movement.

The TGB2 and TGB3 proteins, containing blocks of hydrophobic residues, are associated with membranes derived from the ER. Potexvirus TGB2 protein has two predicted transmembrane domains that interact with ER membranes. It seems that TGB2 protein integrates into the ER membranes in a U-like structure, with the central part positioned in the ER lumen and the N- and C-termini located on the cytoplasmic side of the ER. Mutation analysis revealed that two conserved cysteine residues in the C-terminal region of potexvirus TGB2 protein are important for the cell-to-cell movement. Recently, it was found that PVX TGB2 may function as the molecular bridge allowing interaction between the VRC and TGB1, TGB3, and coat protein, thereby enhancing the efficiency of the virus spread in infected plant tissue.

Potexvirus TGB3 protein also has a conserved C(x₅)G(x₆₋₉)C sequence positioned at the C-terminus of a predicted transmembrane domain. This C-terminal sequence is exposed to the cytoplasmic side of the ER and contains a sorting signal that is necessary for TGB3 oligomerization and for the targeting of cortical ER.

Replication and Propagation

Cis-acting replication signals have been mostly studied using PVX and Bamboo mosaic virus (BaMV) as models. Replication of potexviruses is governed by the cis-signals at the 5' - and the 3' end of the genome (each representing a complex tertiary RNA structure). The synthesis of the negative strand and of the positive-strand genomic and sub-genomic RNAs requires long-distance interactions between the conserved sequences at the 3' ends of the positive and negative RNA strands, and the internal complementary octanucleotides. The consequence of the replication/transcription events is as follows: (1) using genomic RNA as template, the replicase recognizes the 3'-cis-signal base-paired with an internal octanucleotide and synthesizes the antigenomic (negative) RNA strand; (2) in the negative strand, the replicase interacts with the 3'-cis-signal base-paired with an internal octanucleotide and produces the progeny genomic RNA strands; (3) in the negative strand, contacts of the 3'-cis-signal with complementary octanucleotides localized, in 3' to 5' direction, prior to the genes for TGBp1, TGBp2, and CP, force the replicase to initiate internally and produce the respective sgRNAs.

Transmission, Host Range

Most members of the family infecting higher plants are transmitted by mechanical contacts, seeds, or grafting. Citrus yellow vein clearing mandarin virus (CYVCV) is spread from infected to healthy citrus trees by the whitefly *Dialeurodes citri* (Ashmead). Alexiviruses are transmitted by eriophyid mites. PVX causes mild mosaic or a symptomless infection in solanaceous hosts, depending on the virus strain. In *Gomphrena globosa* L., a diagnostic host, the virus induces formation of ring necrotic lesions. Another potexvirus, PepMV causes tomato fruit discoloration (marbling, blotchy ripening and flaming), leaf distortion, chlorosis, yellow spots, interveinal leaf yellowing (Fig. 5) and, occasionally, leaf and stem necrosis. In addition, CYVCV infection in lemon and sour orange is characterized by strong yellow vein clearing, leaf distortion, ringspot lesions, and veinal necrosis. The natural host range of the family members vary from narrow (lollaviruses, mandarinviruses, and the viruses infecting fungal hosts) to moderate (potexviruses).

Epidemiology and Control

The genus *Potexvirus* includes species that are responsible for serious crop damage (PVX, BaMV, Pepino mosaic virus, and Citrus yellow vein clearing virus), whereas the economical impact of other family members is limited or not important. Phytosanitary control measures are used to diminish the virus spread in vegetatively propagated crops. Application of cross-protection with mild viral strains, which proved effective against damaging crop pathogens like PepMV, has been proposed. Transgenic plants of tobacco, tomato, and orchid, resistant to some potexviruses have been produced; they may have a potential for virus control.

Virus-Host Relationships

Alphaflexiviridae members are replicated in the cytoplasm of infected host cells. Infection is accompanied by formation of ER-derived vesicles that serve as replication sites (virus fabrics). Other cytopathic effects are the fibrous, beaded, banded, or irregular aggregates in the cytoplasm and, occasionally, in the nucleus.



Fig. 5 Interveinal leaf yellowing symptoms after inoculation of RNA transcripts of PepMV infectious clone P-5-IYflc (variety Moneymaker). After Hasiów-Jaroszewska, B., Paeleman, A., Ortega-Parra, N., *et al.*, 2013. Ratio of mutated versus wild-type coat protein sequences in Pepino mosaic virus determines the nature and severity of yellowing symptoms on tomato plants, *Molecular Plant Pathology* 14, 923–933.

On the molecular level, virus-host interactions involve contacts of the replicase and the movement proteins of the plant-infecting alphaflexiviruses with a variety of host factors that are required for efficient infection. After entering the host cell, the virus RNA genome is used as template for translation and replication. Several host factors involved in the movement of the potexvirus were identified. First, potexvirus TGB3 exhibits the affinity to highly curved subdomains of cortical ER, targets membrane bodies at the cell periphery and directs the TGB2 protein to these structures. Second, a serine/threonine kinase-like protein from *N. benthamiana* (NbSTKL) was demonstrated to be upregulated in BaMV infection and to be critical for the cell-to-cell movement. It is suggested that NbSTKL modifies a specific protein factor that regulates plasmodesmata (PD) gating for virus trafficking. In addition, another kinase, casein kinase 2 α (CK2 α), which interacts with BaMV CP in PD, can be involved in releasing viral RNA from the movement RNP after the passage of the latter through PD. Third, RabGTPase-activating protein (GAP or NbRabGAP1), is involved in the intercellular and systemic movement of the BaMV potexvirus. Rabs, a family of small GTPases, is involved in all aspects of intracellular vesicle budding, targeting, docking, and fusion. It was shown that NbRabGAP1 activates one of the RabGTPases to support the activity of the TGB proteins in transferring the viral movement complex through plasmodesmata. Involvement of NbRabGAP1 in the BaMV movement suggests that this movement is associated with the vesicle turnover pathway.

Recently, several potential host factors affecting BaMV replication have also been identified. Among them, cytoplasmic 5' $\rightarrow$ 3' exoribonuclease NbXRN4 promotes potexvirus replication, most probably, by removing uncapped genomic and subgenomic RNAs that are produced during replication/transcription as aberrant transcripts. At the replication stage, the 3' untranslated region (UTR) of the potexvirus genomic RNA plays an important role in capturing several host factors assisting the RNA synthesis. Glutathione transferase U4 (NbGSTU4), chloroplastic PGK (chlPGK), and heat shock protein 90 (Hsp90), were found to interact with the 3' UTR of BaMV. The protein chlPGK interacts with 3' UTR and seems to be vital for transmitting viral RNA and the replicase, to chloroplasts. The advantage of BaMV RNA targeting to the chloroplast can be explained by isolation of the VRC from the host protection systems, i.e., RNA silencing. The large viral RNA-protein complex including viral RNA, replicase, and possible host factors enters into chloroplasts, assisted by the Hsp90 acting as chaperone in protein complex folding and protein translocation across membranes. NbGSTU4, a protein involved in antioxidant protection, also stimulates BaMV replication. As soon as BaMV RNA and VRC components enter chloroplasts, they face the danger of high levels of reactive oxygen stress (ROS), since the chloroplast is one of the main producers of ROS. Movement of NbGSTU4 into chloroplasts along with BaMV RNA and the bound glutathione may play a role in eliminating the effects of ROS. Importantly, NbGSTU4 can interact with 3' UTR at physiological glutathione concentrations. Additionally, the S-adenosyl methionine synthetase and the respiratory explosion oxidase homolog were found to stimulate the BaMV replication.

ER and myosins are employed in movement of potexviruses. Although the ER membrane allows the transport of membrane-associated molecules by diffusing within the lipid layer, the mobility of the ER membrane as well as the ER-directed transport of macromolecules is controlled by the ER-associated actin-myosin system. In accordance with this, the directed targeting of TGB proteins to PD and, consequently, the spread of infection also depends on actin and myosin activity.

Antiviral plant protection is also controlled by complex interactions of host and viral factors. For example, antiviral reactions may be associated with nonsense-mediated degradation pathways of mRNA decay (NMD), which is a conservative mRNA quality control mechanism and selectively eliminates aberrant transcripts containing unusually long non-coding 3' terminal sections. By recognizing such viral RNAs and directing them to degradation, NMD neutralizes viral infection and functions as a viral restriction mechanism. A genetic screen in *A. thaliana* found that inactivating mutations in the UPF1 gene, which is a key NMD factor, increase the susceptibility of plants to PVX infection.

Alphaflexiviruses produce dsRNA as an intermediate in replication, making them targets for RNA silencing mechanism initiated by Dicer-like ribonucleases (DCL). After formation, siRNAs bind to AGOs, which are part of RISCs, to form a molecular platform for RNA silencing mechanism. To counteract RNA silencing, alphaflexiviruses express RNA silencing suppressors (VSRs) that inhibit various steps of RNA silencing. Thus, the potexvirus TGBp1 and mandarivirus CP act as VSRs. The TGBp1 of several potexviruses has been shown to have VSR activity inducing destabilization of AGO proteins. On the other hand, the PLAMV TGB1 protein inhibits secondary siRNA production by targeting the RDR6 pathway and additionally affects the AGO4 transfer from nucleus to cytoplasm.

As alternative to RNA-targeted protective mechanisms, plants have developed a variety of resistance (R) proteins that can detect pathogen-associated molecular patterns (PAMPs) and induce protective reactions. Such detected effectors are called avirulence proteins (Avr). Most R-gene products have leucine-rich repeat domains (LRR) associated with the assumed nucleotide-binding site domain (NBS). Recent studies have identified molecular factors that affect the accumulation of NB-LRR and, consequently, the expression of resistance. Among them, the “susceptibility factor essential for potexvirus accumulation” (EXA1) protein containing the glycine-tyrosine-phenylalanine (GYF) domain was identified as a negative regulator of NB-LRR protein accumulation and, as a result, involved in plant immunity. It is interesting that mutations affecting GYF domain protein EXA1 are also involved in resistance to potexviruses. Plantago asiatica mosaic virus (PLAMV), Alternanthera mosaic virus (AltMV), and PVX show significantly suppressed ability to establish infection in the absence of EXA1. Some plants contain a single *EXA1* gene in their genomes. However, other plants, particularly *N. benthamiana*, soybean, and banana, encode multiple genes. It was found that EXA1 proteins are important factors for successful potexvirus infection in a variety of plant species.

Translation elongation factor eEF1a has been shown to regulate the RNA replication of potexviruses. As the eEF1a binding site in the 3′ UTR of BaMV genome overlaps with the RdRp recognition site, eEF1a can play a role in template switch that blocks viral replication during translation. NADP⁺-dependent isocitrate dehydrogenase, MAP kinase, phosphatase-like protein, and proliferating cell nuclear antigen (PCNA) also appear to suppress potexvirus replication.

Recently, it was shown that phosphorylation of REMORIN protein changes its membrane dynamics. PVX TGBp1 is able to activate protein kinase(s), which in turn phosphorylate REMORIN1.3. Since phosphorylated REMORIN is associated with PD closure by induction of callose deposition, this protein causes the reduction of PD permeability and thereby restriction of viral cell-to-cell movement.

Diagnosis

Diagnosis of viruses belonging to *Alphaflexiviridae* mostly relies on serological tests (enzyme-linked immunosorbent assay) and reverse transcriptase PCR methods, including the modifications of the latter like triplex PCR and droplet digital PCR.

Satellite RNAs Associated With Infections of *Alphaflexiviridae*

The only satellite RNAs (satRNAs) so far known to accompany the infections of *Alphaflexiviridae* are those associated with BaMV potexvirus. The satRNAs are variable in sequence, but have the 5′- and 3′ terminal hairpins resembling the cis-signals at the helper BaMV RNA termini. The vital functions provided by the helper BaMV to satRNAs are replication, encapsidation, and cell-to-cell movement. Co-replication of satRNAs has pronounced negative effect on the helper virus propagation. One of the BaMV satRNAs encodes a 20 kDa nonstructural protein which preferentially binds to the satRNA and promotes its long-distance transport in BaMV-infected *N. benthamiana*. The use of transgenic plants expressing satRNAs to control BaMV infection has been considered.

Further Reading

- Çağlayan, K., Roumi, V., Gaze, L.M., *et al.*, 2019. Identification and characterization of a novel Robigovirus species from sweet cherry in Turkey. *Pathogens* 8, 57.
- Hasiów-Jaroszewska, B., Paeleman, A., Ortega-Parra, N., *et al.*, 2013. Ratio of mutated versus wild-type coat protein sequences in Pepino mosaic virus determines the nature and severity of yellowing symptoms on tomato plants. *Molecular Plant Pathology* 14, 923–933.
- Lin, K.Y., Lin, N.S., 2017. Interfering satellite RNAs of bamboo mosaic virus. *Frontiers in Microbiology* 8, 787.
- Liu, Q., Yang, L., Xuan, Z., *et al.*, 2020. Complete nucleotide sequence of Loquat virus A, a member of the family *Betaflexiviridae* with a novel genome organization. *Archives of Virology* 165, 223–226.
- Martelli, G.P., Adams, M.J., Kreuze, J.F., Dolja, V.V., 2007. Family *Flexiviridae*: A case study in virion and genome plasticity. *Annual Review of Phytopathology* 45, 73–100.
- Park, M.R., Seo, J.K., Kim, K.H., 2013. Viral and non-viral elements in potexvirus replication and movement and in antiviral responses. *Advances in Virus Research* 87, 75–112.
- Ur Rehman, A., Li, Z., Yang, Z., *et al.*, 2019. The coat protein of Citrus yellow vein clearing virus interacts with viral movement proteins and serves as an RNA silencing suppressor. *Viruses* 11, 329.
- Verchot-Lubicz, J., Ye, C.M., Bamunusinghe, D., 2007. Molecular biology of potexviruses: Recent advances. *Journal of General Virology* 88, 1643–1655.

Alphasatellites (*Alphasatellitidae*)

Rob W Briddon and Muhammad S Nawaz-ul-Rehman, University of Agriculture, Faisalabad, Pakistan

© 2021 Published by Elsevier Ltd.

Nomenclature

CR-M Common region major
CR-SL Common region stem-loop
mRep Master replication-associated protein
ORF Open reading frame

PTGS Post-transcriptional gene silencing
Rep Replication initiator protein
ssDNA single-stranded deoxyribonucleic acid
TGS Transcriptional gene silencing

Glossary

Post-transcriptional gene silencing An epigenetic process for regulation of gene expression that results in the mRNA of a particular gene being destroyed. Post-transcriptional gene silencing is believed to protect the organism's genome from, among other things, transposons and viruses.

Rolling-circle replication A mechanism of DNA replication that synthesizes multiple copies of circular molecules of DNA. Rolling circle replication is initiated by an initiator protein that nicks one strand of a double-stranded, circular DNA molecule. The initiator protein remains bound to the 5' phosphate end of the nicked strand and the free 3' hydroxyl end is released to serve as a primer

for DNA synthesis using the unnicked strand as a template; replication proceeds around the circular DNA molecule, displacing the nicked strand as single-stranded DNA.

Satellite like molecule A self replicating molecule which depends on helper virus for movement, encapsidation and transmission.

Transcriptional gene silencing An epigenetic process for regulation of gene expression that results from DNA methylation and histone modifications. Transcriptional gene silencing is believed to protect the organism's genome from, among other things, transposons and integrated viruses.

Introduction

Viruses of the family *Nanoviridae*, some begomoviruses (genus *Begomovirus*, family *Geminiviridae*) and, in one case reported so far, a mastrevirus (genus *Mastrevirus*, family *Geminiviridae*) may be associated with additional circular ssDNA components that are known as alphasatellites. Recently coconut foliar decay, a severe disease of coconut palms in Vanuatu, has been shown to be associated with a multipartite ssDNA virus that is distinct from nanoviruses and geminiviruses and to be associated with multiple (possible) alphasatellites. The association with alphasatellites may thus extend to three families of plant-infecting ssDNA viruses.

Nanoviruses (family *Nanoviridae*) are multipartite ssDNA viruses that infect plants and are transmitted by aphids. The genomes of nanoviruses are composed of 6–8 circular components, each ~1100 nt in size, separately encapsidated in isometric particles. The component known as DNA-R encodes the master replication-associated protein (mRep) that is responsible for *trans*-replication of all *bona fide* virus components. mRep recognises the origin of replication (*ori*) which sits in the common region stem-loop (CR-SL), conserved between all *bona fide* virus components, and initiates rolling circle replication by nicking within the nonanucleotide sequence TAGTATTAC that forms part of the loop of a predicted stem-loop structure. The *bona fide* nanovirus components also contain a second conserved sequence known as the common region major (CR-M).

Geminiviruses (family *Geminiviridae*) are plant-infecting viruses with geminate (twinned quasi-icosahedral) particles that are transmitted by insect vectors. Begomoviruses (genus *Begomovirus*) have either monopartite or bipartite genomes and are transmitted by the whitefly *Bemisia tabaci* whereas mastreviruses (genus *Mastrevirus*) have monopartite genomes and are transmitted by specific leafhoppers.

The first alphasatellite was identified in association with the nanovirus *Subterranean clover stunt virus* in 1995. Sequencing of the genomic components of the virus highlighted the presence of more than one component encoding a replication-associated protein, a feature that was shown for all subsequently characterized nanoviruses. Realisation that the additional Rep-encoding components are satellites came with the demonstration that only the mRep has the capacity to direct replication of the other genome components using infectious clones. The first geminivirus-associated alphasatellite was identified in 1999 in cotton affected by Cotton leaf curl disease (CLCuD). *Cotton leaf curl Multan alphasatellite* was identified fortuitously, by shotgun-cloning, in the search for additional components responsible for the symptoms of CLCuD in cotton. The symptoms were subsequently shown to be encoded by a component we now know as a betasatellite (family *Tolycusatellitidae*; genus *Betasatellite*).



Fig. 1 Comparison of the structures nanovirus- and geminivirus-associated alphasatellites in comparison to the DNA-R genomic components of nanoviruses. The diagram shows the position of the replication-associated protein (Rep), the TATA box of the presumed Rep promoter, an adenine rich sequence (A-rich) in geminivirus-associated alphasatellites, as well as the common region major (CR-M) and common region stem-loop (CR-SL) of nanovirus DNA-R components that is conserved between all *bona fide* nanovirus genome components. All components have a predicted stem-loop structure (at position zero) with, in most cases, the nona-nucleotide sequence TAGTATTAC forming part of the loop. This likely forms the origin of virion-strand DNA replication that is nicked by Rep to initiate rolling-circle replication. Note that the DNA-R components of some nanoviruses contain an additional open reading frame (ORF) known as ORF 1.

Classification

The additional circular ssDNA components associated with nanoviruses and geminiviruses are classified in the recently established family *Alphasatellitidae*. The family encompasses two subfamilies, *Nanoalphasatellitinae* and *Geminialphasatellitinae*, to accommodate the nanovirus and geminivirus alphasatellites, respectively. At present the subfamily *Nanoalphasatellitinae* encompasses seven genera and 19 species, whereas the subfamily *Geminialphasatellitinae* encompasses 4 genera with 43 assigned species. Additionally a divergent possible alphasatellite, associated with coconut foliar decay disease, has been assigned to a species but not a subfamily. Within this classification structure alphasatellites are assigned based on sequence and phylogenetic relatedness.

Structure of Alphasatellites

In common with the DNA-R component of nanoviruses, the alphasatellites encode a Rep protein and have a predicted stem-loop structure with, in most cases, the nona-nucleotide sequence TAGTATTAC forming part of the loop (Fig. 1). Alphasatellites associated with nanoviruses have a size typical of nanovirus components (~1100 nt), but lack the conserved CR-SL and CR-M of their helper viruses and consequently are unable to *trans*-replicate the *bona fide* genome components of their helper viruses.

Typically the geminivirus-associated alphasatellites comprise ~1400 nt, being significantly larger than nanovirus components and nanovirus-associated alphasatellites. The size difference is due to the presence in geminivirus-associated alphasatellites of a sequence rich in adenine and a larger Rep-encoding gene (Fig. 1). The size difference is likely necessary for the movement and/or encapsidation of geminivirus alphasatellites by their helper viruses. Geminiviruses have a stringent size surveillance mechanism that operates during virus movement within the plant as well as size constraints for encapsidation. Being approximately half the geminivirus genome length (~1400 nt), geminalphasatellites may possibly be encapsidated in isometric (half geminate) virions. (Fig. 2).

Likely Origins of Alphasatellites

Based on sequence relatedness it is likely that nanoalphasatellites evolved from the DNA-R components of nanoviruses that were captured by a second nanovirus during a co-infection. A similar “capture following co-infection” mechanism may have resulted in the geminalphasatellites, although the progenitor DNA-R then needed to adapt, by size increase, for efficient maintenance by a geminiviruses. Consistent with this hypothesis it is interesting to note that, having been uncoupled from the mechanism which controls the relative titers of *bona fide* nanovirus components, alphasatellites of both nanoviruses and geminiviruses in many cases vastly exceed the titers of their helper virus component(s) in plants.

It is interesting to note that vast majority of geminivirus-associated alphasatellites from the Old World have been identified in association with monopartite begomoviruses infections that also include a betasatellite (family *Tolecasatellitidae*). In the New World, where only very few monopartite begomoviruses and no betasatellites have been identified, alphasatellites have only so far been reported to occur in association with bipartite begomoviruses. In contrast, nanoviruses have only been identified in the Old World, which may suggest that the association of alphasatellites with geminiviruses evolved in the Old World, although when this occurred is unclear.

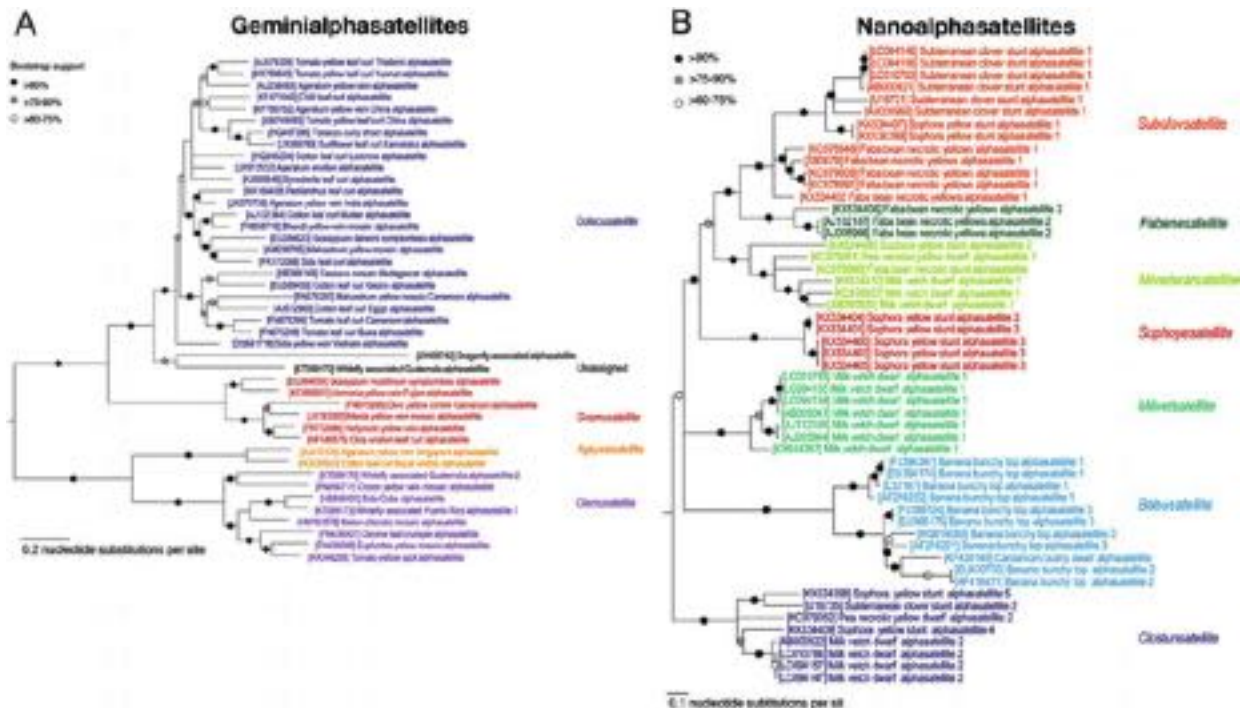


Fig. 2 (A) Maximum likelihood phylogenetic tree of representative geminivirus-associated alphasatellite sequences from each species. (B) Maximum likelihood phylogenetic trees of the nanovirus-associated alphasatellite sequences. Trees are inferred using IQ-TREE with TIM2 + F + I + G4 chosen as the best fit model. Branches with less than 60% bootstrap support have been collapsed. The Maximum likelihood phylogenetic tree on the left is a guide for genera demarcation with the cyan line showing the rough 67% genus demarcation threshold. Both trees reproduced from Briddon, R.W., Martin, D.P., Roumagnac, P., *et al.*, 2018. Alphasatellitidae: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. Archives of Virology 163, 2587–2600. doi:10.1007/s00705-018-3854-2.

Table 1 List of alphasatellites with subfamilies, genera, species, acronym and accession number. Type species in each genus is written in bold

Subfamily/Genus	Species	Acronym	Accession #
<i>Geminialphasatellitinae</i>			
<i>Ageyesisatellite</i>	Ageratum yellow vein Singapore alphasatellite	AYVSGA	NC_003414
	Cotton leaf curl Saudi Arabia alphasatellite	CLCuSAA	HG530543
<i>Clecrusatellite</i>	Cleome leaf crumple alphasatellite	CILCrA	NC_014646
	Croton yellow vein mosaic alphasatellite	CrYVMA	NC_013801
	Euphorbia yellow mosaic alphasatellite	EuYMA	NC_014630
	Melon chlorotic mosaic alphasatellite	MeCMA	NC_014379
	Sida Cuba alphasatellite	SiCUA	NC_021708
	Tomato yellow spot alphasatellite	ToYSA	NC_038993
	Whitefly associated Guatemala alphasatellite 2	WfaGUA2	NC_038973
	Whitefly associated Puerto Rico alphasatellite 1	WfaPRA1	NC_038975
<i>Colecusatellite</i>	<i>Ageratum enation alphasatellite</i>	AEA	NC_019547
	<i>Ageratum yellow vein alphasatellite</i>	AYVA	NC_033555
	<i>Ageratum yellow vein China alphasatellite</i>	AYVCNA	NC_023443
	<i>Ageratum yellow vein India alphasatellite</i>	AYVINA	NC_019546
	<i>Bhendi yellow vein alphasatellite</i>	BhYVA	NC_029313
	<i>Cassava mosaic Madagascar alphasatellite</i>	CMMGA	NC_018628
	<i>Chilli leaf curl alphasatellite</i>	ChLCuA	NC_039234
	<i>Cotton leaf curl Egypt alphasatellite</i>	CLCuEGA	AJ512960
	<i>Cotton leaf curl Gezira alphasatellite</i>	CLCuGeA	NC_013593
	<i>Cotton leaf curl Lucknow alphasatellite</i>	CLCuLuA	NC_015327
	Cotton leaf curl Multan alphasatellite	CLCuMuA	NC_018082
	<i>Gossypium darwinii symptomless alphasatellite</i>	GDarSLA	NC_039232
	<i>Malvastrum yellow mosaic alphasatellite</i>	MaYMA	NC_039239
	<i>Malvastrum yellow mosaic Cameroon alphasatellite</i>	MaYMCMA	NC_014907
	<i>Pedilanthus leaf curl alphasatellite</i>	PeLCuA	NC_033335
	<i>Sida leaf curl alphasatellite</i>	SiLCuA	NC_039240

(Continued)

Table 1 Continued

Subfamily/Genus	Species	Acronym	Accession #
	<i>Sida</i> yellow vein Vietnam alphasatellite	SiYVNA	NC_009563
	Sunflower leaf curl Karnataka alphasatellite	SLCuKaA	NC_019499
	<i>Synedrella</i> leaf curl alphasatellite	SyLCuA	NC_025244
	Tobacco curly shoot alphasatellite	TbCSA	NC_005057
	Tomato leaf curl Cameroon alphasatellite	ToLCuCMA	NC_014747
	Tomato yellow leaf curl China alphasatellite	TYLCuCNA	NC_043449
	Tomato yellow leaf curl Thailand alphasatellite	TYLCuTHA	AJ579359
	Tomato yellow leaf curl Yunnan alphasatellite	TYLCuYnA	NC_039229
<i>Gosmusatellite</i>	<i>Gossypium mustelinum</i> symptomless alphasatellite	GMusSLA	NC_038954
	Hollyhock yellow vein alphasatellite	HoYVA	NC_039084
	Mesta yellow vein mosaic alphasatellite	MeYVMA	NC_018573
	Okra enation leaf curl alphasatellite	OEnLCuA	NC_019944
	Okra yellow crinkle Cameroon alphasatellite	OKYCCMA	NC_014906
	Vernonia yellow vein Fujian alphasatellite	VeYVFJA	NC_039233
<i>Unassigned</i>	<i>Ageratum</i> leaf curl Cameroon alphasatellite	ALCuCMA	NC_014744
	Dragonfly associated alphasatellite	DfaA	NC_019498
	Parthenium leaf curl alphasatellite	PLCuA	NC_030226
	Tomato leaf curl Buea alphasatellite	ToLCuBuA	NC_038903
	Whitefly associated Guatemala alphasatellite 1	WfaGUA1	NC_038974
<i>Nanoalphasatellitinae</i>			
<i>Babusatellite</i>	Banana bunchy top alphasatellite 1	BBTA2	NC_038892
	Banana bunchy top alphasatellite 2	BBTA1	NC_038953
	Banana bunchy top alphasatellite 3	BBTA3	NC_038955
	Cardamom bushy dwarf alphasatellite	CaBuDA	NC_022919
<i>Clostunsatellite</i>	Milk vetch dwarf alphasatellite 2	MVDA2	NC_003639
	Pea necrotic yellow dwarf alphasatellite 2	PNYDA2	NC_038959
	Sophora yellow stunt alphasatellite 4	SYSA4	NC_039000
	Sophora yellow stunt alphasatellite 5	SYSA5	NC_038996
	Subterranean clover stunt alphasatellite 2	SCSA2	NC_003818
<i>Fabenesatellite</i>	Faba bean necrotic yellows alphasatellite 2	FBNYA2	NC_003567
<i>Milvetsatellite</i>	Milk vetch dwarf alphasatellite 3	MVDA3	NC_003640
<i>Mivedwarsatellite</i>	Faba bean necrotic stunt alphasatellite	FBNSA	NC_023881
	Milk vetch dwarf alphasatellite 1	MVDA1	NC_003638
	Pea necrotic yellow dwarf alphasatellite 1	PNYDA1	NC_038958
	Sophora yellow stunt alphasatellite 2	SYSA2	NC_038999
<i>Sophoyesatellite</i>	Sophora yellow stunt alphasatellite 3	SYSA3	NC_038998
<i>Subclovsatellite</i>	Faba bean necrotic yellows alphasatellite 1	FBNYA1	NC_024886
	Sophora yellow stunt alphasatellite 1	SYSA1	NC_038997
	Subterranean clover stunt alphasatellite 1	SCSCA1	NC_003814
<i>Unassigned</i>			
	Coconut foliar decay alphasatellite	CFDA	NC_001465

Life Cycle/Epidemiology

Alphasatellites replicate autonomously of their helper viruses in the nuclei of permissive host cells by rolling circle replication. For all other functions alphasatellites depend upon factors provided by their helper viruses including movement within/between cells, and encapsidation for transmission between plants. However, there appears to be little, if any, specificity for helper virus of a particular alphasatellite. For example, the geminalalphasatellite *Ageratum yellow vein Singapore alphasatellite* (AYVSGA; genus *Ageyesisatellite*), which was identified in association with the OW monopartite begomovirus *Ageratum yellow vein virus*, was shown to be maintained in plants by the NW bipartite begomovirus *Tomato golden mosaic virus* and the curtovirus (family *Geminiviridae*, genus *Curtovirus*) *Beet curly top virus* (BCTV). When maintained in plants by BCTV, AYVSGA co-insect transmitted by the leafhopper vector of BCTV, showing the vector of the alphasatellite to depend upon the vector specificity of the helper virus. The same work also showed AYVSGA not to be maintained in plants by the mastrevirus (family *Geminiviridae*, genus *Mastrevirus*) Chickpea chlorotic dwarf virus (CpCDV; earlier known as Bean yellow dwarf virus), possibly due to packaging constraints; the genome of CpCDV is significantly smaller than the genomes of the begomoviruses by which it is maintained.

Pathogenesis

The precise impact of alphasatellites on nanovirus and geminivirus infections remains unclear. For some nanovirus infections the presence of nanoalphasatellites has been shown to reduce nanovirus infectivity. This suggests that there is competition between the nanoalphasatellite and the *bona fide* nanovirus genomic DNAs for factors required for replication, systemic movement and/or encapsidation. For geminivirus infections the presence of geminialphasatellites has been shown to result in either attenuated or exacerbated symptoms. For infections with attenuated symptoms the DNA titer of either the virus genome or the associated betasatellite were found to be reduced in comparison to plants not inoculated with the alphasatellite.

The Rep proteins encoded by the geminialphasatellites *Gossypium darwinii* *symptomless alphasatellite* and *Gossypium mustelinum* *symptomless alphasatellite* suppress post-transcriptional gene silencing (PTGS), whereas those encoded by various other geminialphasatellites are suppressors of transcriptional gene silencing (TGS). This suggests that some, if not all, alphasatellites assist the helper virus in overcoming (suppressing) host resistance based on RNA interference ([Table 1](#)).

See also: Betasatellites and Deltasatellites (*Tolecusatellitidae*). Geminiviruses (*Geminiviridae*). Nanoviruses (*Nanoviridae*)

Further Reading

- Abbas, Q., Amin, I., Mansoor, S., *et al.*, 2019. The Rep proteins encoded by alphasatellites restore expression of a transcriptionally silenced green fluorescent protein transgene in *Nicotiana benthamiana*. *Virus Disease* 30 (1), 101–105.
- Boevink, P., Chu, P.W.G., Keese, P., 1995. Sequence of Subterranean clover stunt virus DNA: affinities with the geminiviruses. *Virology* 207, 354–361.
- Briddon, R.W., Martin, D.P., Roumagnac, P., *et al.*, 2018. *Alphasatellitidae*: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Archives of Virology* 163, 2587–2600. doi:10.1007/s00705-018-3854-2.
- Gronenborn, B., Randles, J.W., Knierim, D., *et al.*, 2018. Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded DNA virus representing a new taxon. *Scientific Reports* 8, 5698.
- Kumar, J., Kumar, J., Singh, S.P., Tuli, R., 2014. Association of satellites with a mastrevirus in natural infection: complexity of Wheat dwarf India virus disease. *Journal of Virology* 88, 7093–7104.
- Mansoor, S., Khan, S.H., Bashir, A., *et al.*, 1999. Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology* 259, 190–199.
- Mar, T.B., Mendes, I.R., Lau, D., *et al.*, 2017. Interaction between the new world begomovirus *Euphorbia yellow mosaic virus* and its associated alphasatellite: Effects on infection and transmission by the whitefly *Bemisia tabaci*. *Journal of General Virology* 98, 1552–1562.
- Nawaz-Ul-Rehman, M.S., Nahid, N., Mansoor, S., Briddon, R.W., Fauquet, C.M., 2010. Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. *Virology* 405, 300–308.
- Saunders, K., Bedford, I.D., Stanley, J., 2002. Adaptation from whitefly to leafhopper transmission of an autonomously-replicating nanovirus-like DNA component associated with ageratum yellow vein disease. *Journal of General Virology* 83, 909–915.
- Sicard, A., Yvon, M., Timchenko, T., *et al.*, 2013. Gene copy number is differentially regulated in a multipartite virus. *Nature Communications* 4.
- Timchenko, T., Katul, L., Aronson, M., *et al.*, 2006. Infectivity of nanovirus DNAs: Induction of disease by cloned genome components of *Faba bean necrotic yellows virus*. *Journal of General Virology* 87, 1735–1743.
- Zhou, X., 2013. Advances in understanding begomovirus satellites. *Annual Review of Phytopathology* 51, 357–381.

Amalgaviruses (*Amalgaviridae*)

Ioannis E Tzanetakis, University of Arkansas, Fayetteville, United States

Sead Sabanadzovic, Mississippi State University, Starkville, MS, United States

Rodrigo A Valverde, Louisiana State University Agricultural Center, Baton Rouge, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

CP Coat protein

dsRNA Double stranded RNA

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

RT-LAMP Reverse transcription loop-mediated isothermal amplification

RT-PCR Reverse transcription polymerase chain reaction

Glossary

Bicistronic Nucleic acid that codes for two proteins.

Nucleoprotein Protein protecting the genomic RNA of negative sense RNA viruses.

Ribosomal frameshift Translation strategy, quite widespread in viruses, that allows for the expression of two partially overlapping, non-in-frame ORFs in a single round of translation involving ribosome slipping by one base in either the 5' or 3' directions.

Classification

The family *Amalgaviridae*, part of the realm *Riboviria*, was recognized by the International Committee on Taxonomy of Viruses (ICTV) in 2014. The family includes viruses infecting plants and fungi (yeasts) with a non-segmented, bicistronic dsRNA genome of 3.1–3.5 kbp and is currently composed of two genera: *Amalgavirus* and *Zybavirus*. The genus *Amalgavirus* contains 9 officially recognized species while the genus *Zybavirus* has only one species. The type species for genus *Amalgavirus* is *Southern tomato virus*, and the type species for the genus *Zybavirus* is *Zygosaccharomyces bailii virus Z* (**Table 1**). There are also another 27 possible members of the family for which sequences are available and phylogenetic studies indicate that at least 12 of them could be classified in the genus *Amalgavirus* while one (*Antonospora locustae virus 1*, AnLoV-1) could be classified either in the genus *Zybavirus* or represent a completely new genus in the family (**Fig. 1**).

Virion Structure

Virus particles have never been observed for plant amalgaviruses under the electron microscope, but antibodies against the putative product of ORF1, with possible role of the viral CP, reacted with amorphous bodies in infected plants. Biochemical analysis points to protection of the genome by a protein. *In silico* analyses indicate that the putative CP of amalgaviruses is homologous to the nucleoprotein of animal and plant-infecting negative stranded RNA viruses suggesting the possibility of filamentous particles instead of the hypothesized spherical ones. The spherical particle hypothesis is based on *Zygosaccharomyces bailii virus Z* (ZbV-Z), where virus-like particles were purified from yeast. Notwithstanding, purification was performed using an isolate co-infected with a totivirus which is known to form spherical particles.

Genome

The genome of amalgaviruses is comprised of a single double stranded RNA (dsRNA) molecule of ~3.5 kbp (viruses in the genus *Amalgavirus*) or ~3.1 kbp in size (*Zygosaccharomyces bailii virus Z*, sole member of the genus *Zygovirus*). All viruses in the family code for two proteins (**Fig. 2**). The 5' proximal protein is presumed to be the CP of the virus. As noted above, the protein is homologous to the nucleoprotein of animal and plant-infecting negative strand viruses. The second protein, encoded by ORF2, is predicted to be expressed via a programmed +1 ribosomal frameshift and contains RdRp motifs most similar to those found in members of the family *Partitiviridae*. Putative +1 ribosomal frameshifting motif, UUU_CGN, present in all but two plant amalgaviruses was identified recently by extensive *in silico* analyses. Phylogenetic analyses have shown that RdRp of amalgaviruses is evolutionarily closer to that of partitiviruses than to counterparts expressed by members of the family *Totiviridae*. Overall, the genome appears to be an amalgam of the monopartite family *Totiviridae* and the multipartite family *Partitiviridae*.

Table 1 List of members in the family *Amalgaviridae*. Type species are written in bold. Accessions of complete genome sequences are provided

Genus	Species	Virus	Acronym	Accession #
<i>Amalgavirus</i>	<i>Allium cepa amalgavirus 1</i>	Allium cepa amalgavirus 1	AcAV1	NC_036580
	<i>Allium cepa amalgavirus 2</i>	Allium cepa amalgavirus 2	AcAV2	NC_036581
	<i>Blueberry latent virus</i>	Blueberry latent virus	BILV	NC_014593
	<i>Rhododendron virus A</i>	Rhododendron virus A	RhVA	NC_014481
	<i>Southern tomato virus</i>	Southern tomato virus	STV	NC_011591
	<i>Spinach amalgavirus 1</i>	Spinach amalgavirus 1	SpAV1	NC_035070
	<i>Vicia cryptic virus M</i>	Vicia cryptic virus M	VCV-M	NC_043095
	<i>Zoostera marina amalgavirus 1</i>	Zoostera marina virus 1	ZmAV1	NC_034614
	<i>Zoostera marina amalgavirus 2</i>	Zoostera marina virus 2	ZmAV2	NC_034615
<i>Zybvirus</i>	<i>Zygosaccharomyces bailii zybvirus</i>	Zygosaccharomyces bailii virus Z	ZbV-Z	NC_003874
Unassigned viruses		Anthoxanthum odoratum amalgavirus 1	AoAV1	NC_040432
		Anthoxanthum odoratum amalgavirus 2	AoAV2	BK010357
		Antonospora locustae virus 1	AIV1	NC_035189
		Camellia oleifera amalgavirus 1	CoAV1	NC_040433
		Capsicum annuum amalgavirus	CaAV-1	NC_040662
		Cistus incanus RNA virus 1	CiRV1	MG833407
		Cleome droserifolia amalgavirus	CdAV1	NC_040777
		Cucumis melo amalgavirus	CmAV	MH479774
		Erigeron breviscapus amalgavirus 1	EbAV1	NC_040592
		Erigeron breviscapus amalgavirus 2	EbAV2	NC_040593
		Festuca pratensis amalgavirus 1	FpAV1	NC_040663
		Festuca pratensis amalgavirus 2	FpAV2	NC_040778
		Festuca pratensis amalgavirus 3	FpAV3	BK010350
		Gevuina avellana amalgavirus 1	GaAV1	BK010351
		Lolium perenne amalgavirus 1	LpAV1	BK010352
		Medicago sativa amalgavirus 1	MsAV1	NC_040591
		Neurachne minor latent virus 1	NmLV1	MF325033
		Phalaenopsis equestris amalgavirus 1	PeAV1	NC_040590
		Pinus patula amalgavirus 1	PpAV1	BK010353
		Pterostylis amalgavirus	PdV	ND ^a
		Rubber dandelion latent virus 1	RdLV1	MF197380
		Rubber dandelion latent virus 2	RdLV2	MF197379
		Salicornia europaea amalgavirus	SeAV1	BK010404
		Salvia hispanica RNA virus 1	ShRV1	NC_040690
		Secale cereale virus	ScV	BK010349

^aND, no complete genome sequence available.

Life Cycle

Little is known about the life cycle of amalgaviruses. The plant-infecting members of the family are not mechanically, nor graft-transmissible. They are very efficiently transmitted through seed, reaching 100% transmission in the case of southern tomato virus (STV) and blueberry latent virus (BLV). ZbV-Z can be transmitted to other yeast species.

Epidemiology

Given the lack of a identifiable movement protein it is hypothesized that the plant-infecting members of the family move within their hosts only during cell division. No symptoms have been observed in hosts infected with any member of the genus *Amalgavirus* although there are reports of STV having a mutualistic interaction with tomato or affecting the expression of tomato miRNAs. The effect of ZbV-Z on its hosts remains unknown.

Pathogenesis

None known.

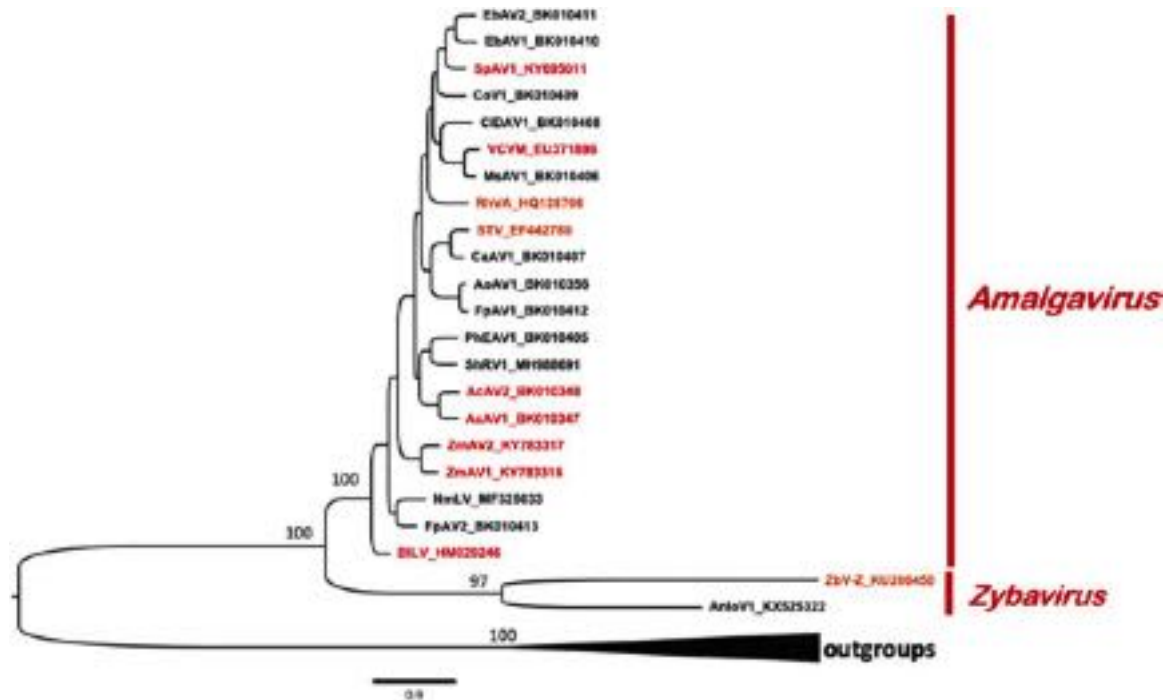


Fig. 1 Phylogenetic tree of recognized (red font) and putative members (black) of the family *Amalgaviridae*. The tree was reconstructed on amino acid sequences of viral RdRPs initially aligned with MUSCLE (v3.8.31) and curated by Gblocks (v0.91b) in order to remove ambiguous regions. The phylogenetic tree was reconstructed with the maximum likelihood method implemented in the PhyML program (v3.1/3.0 aLRT), applying LG + G substitution model, estimated to be best-fit for this dataset. Reliability for internal branch was assessed using the aLRT test (SH-Like). The tree was edited and visualized with FigTree (v1.44). Complete names of viruses used in the analyses are: *Allium cepa* amalgavirus 1 (AcAV1; BK010347), *Allium cepa* amalgavirus 2 (AcAV2; BK010348), *Anthoxanthum odoratum* amalgavirus 1 (AoAV1; BK010356), *Antoniospora locustae* virus 1 (AnloV1; KX525322), *blueberry latent virus* (BILV; HM029246), *Camellia oleifera* amalgavirus 1 (CoAV1; BK010409), *Capsicum annuum* amalgavirus 1 (CaAV1; BK010407), *Cleome droserifolia* amalgavirus 1 (CIDAV1; BK010408), *Erigeron breviscapus* amalgavirus 1 (EbAV1; BK010410), *Erigeron breviscapus* amalgavirus 2 (EbAV2; BK010411), *Festuca pratensis* amalgavirus 1 (FpAV1; BK010412), *Festuca pratensis* amalgavirus 2 (FpAV2; BK010413), *Medicago sativa* amalgavirus 1 (MsAV1; BK010406), *Neurachne minor* latent virus (NmLV1; MF325033), *Phalaenopsis equestris* amalgavirus 1 (PeAV1; BK010405), *rhododendron virus A* (RhVA; HQ128706), *Salvia hispanica* RNA virus 1 (ShRV1; MH988691), *southern tomato virus* (STV; EF442780), *spinach amalgavirus 1* (SpAV1; KY695011), *Vicia cryptic virus M* (VCVM; EU371896), *Zostera marina* amalgavirus 1 (ZmAV1; KY783316), *Zostera marina* amalgavirus 2 (ZmAV2; KY783317), *Zygosaccharomyces bailii* virus Z (ZbV-Z; KU200450). Amino acid sequences of RdRPs of *Nigrospora oryzae* unassigned RNA virus 1 (NoURV1; KT258976) and *Ustilagoidea vires* nonsegmented virus 1 (UvNV1; KJ605397) were used as outgroups. The two genera of the family, *Amalgavirus* and *Zybavirus*, indicated.

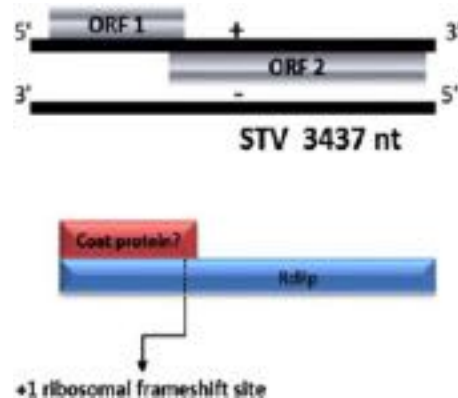


Fig. 2 Genome depiction and putative protein products of Southern tomato virus (STV), the type member of the genus *Amalgavirus*. The two open reading frames are coded in the same strand and translate to the putative coat protein (red block) and its fusion with the RNA dependent RNA polymerase (RdRp; blue).

Diagnosis

There are many protocols developed for the type species of the genus *Amalgavirus*, southern tomato virus, which include RT-PCR, quantitative RT-PCR, molecular hybridization and RT-LAMP. The other viruses in the family are primarily detected by virus-specific RT-PCR.

Treatment

Virus-elimination protocols such as thermo-/chemo-/cryotherapy have not been reported for any amalgavirus. Yeast can be cured from ZbV-Z using cycloheximide or elevated temperatures.

Prevention

The best prevention approach for all plant viruses including amalgaviruses is to start with clean propagative materials. Independent of true seed or clonal propagation, mother plants should be tested free of the respective amalgavirus.

See also: Fungal Partitiviruses (*Partitiviridae*). Totiviruses (*Totiviridae*)

Further Reading

- Depierreux, D., Vong, M., Nibert, M.L., 2016. Nucleotide sequence of *Zygosaccharomyces bailii* virus Z: Evidence for +1 programmed ribosomal frameshifting and for assignment to family *Amalgaviridae*. *Virus Research* 217, 115–124.
- Elvira-González, L., Medina, V., Rubio, L., Galipienso, L., 2020. The persistent Southern tomato virus modifies miRNA expression without inducing symptoms and cell ultra-structural changes. *European Journal of Plant Pathology* 156, 615–622.
- Elvira-Gonzalez, L., Puchades, A.V., Carpino, C., *et al.*, 2017. Fast detection of Southern tomato virus by one-step transcription loop-mediated isothermal amplification (RT-LAMP). *Journal of Virological Methods* 241, 11–14.
- Fukuhara, T., Tabara, M., Koiwa, H., Takahashi, H., 2020. Effect of asymptomatic infection with Southern tomato virus on tomato plants. *Archives of Virology* 165, 11–20.
- Isogai, M., Nakamura, T., Ishii, K., *et al.*, 2011. Histochemical detection of Blueberry latent virus in highbush blueberry plant. *Journal of General Plant Pathology* 77, 304.
- Martin, R.R., Zhou, J., Tzanetakis, I.E., 2011. Blueberry latent virus: An amalgam of the *Partitiviridae* and *Totiviridae*. *Virus Research* 155, 175–180.
- Nibert, M.L., Pyle, J.D., Firth, A.E., 2016. A +1 ribosomal frameshifting motif prevalent among plant amalgaviruses. *Virology* 498, 201–208.
- Puchades, A.V., Carpino, C., Alfaro-Fernandez, A., *et al.*, 2017. Detection of Southern tomato virus by molecular hybridisation. *Annals of Applied Biology* 171, 172–178.
- Sabanadzovic, S., Valverde, R.A., Brown, J.K., Martin, R.R., Tzanetakis, I.E., 2009. Southern tomato virus: The link between the families *Totiviridae* and *Partitiviridae*. *Virus Research* 140, 130–137.
- Schmitt, M.J., Neuhausen, F., 1994. Killer toxin-secreting double-stranded RNA mycoviruses in the yeasts *Hanseniaspora uvarum* and *Zygosaccharomyces bailii*. *Journal of Virology* 68, 1765–1772.

Relevant Websites

- https://viralzone.expasy.org/4956?outline=all_by_species
Amalgavirus
ViralZone.
- <https://talk.ictvonline.org/taxonomy/>
Taxonomy
ICTV.

Badnaviruses (*Caulimoviridae*)

Andrew DW Geering, The University of Queensland, St. Lucia, QLD, Australia

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

CP Capsid protein
dsDNA Double-stranded DNA
kbp Kilobase pairs
kDa Kilodaltons
MP Movement protein
NC Nucleocapsid
NID N-terminal intrinsically disordered

NLS Nuclear localization signal
nm Nanometre
ORF Open reading frame
pgRNA Pregenomic RNA
ssRNA Single-stranded RNA
tRNA Transfer RNA
uORF Upstream ORF
VAP Virion-associated protein

Glossary

Discontinuity Gap in the DNA strand.
Endogenous viral element Viral DNA integrated in the nuclear genome of the host organism and inherited from parents like any normal cellular gene.
Minichromosome Viral DNA wrapped around histone proteins in the nucleus.

Reverse transcription Use of an RNA template to direct synthesis of a complementary strand of DNA.
Virion The virus particle, consisting of a DNA genome surrounded by a protein shell and constituting the infective form of the virus.
Viroplasms Virus factories.

Introduction

The genus *Badnavirus* is a group of double-stranded (ds) DNA viruses with bacilliform particles in the family *Caulimoviridae*. They are referred to as 'reverse-transcribing viruses' as they produce a single-stranded RNA replicative intermediate, which acts as the template for reverse transcription back to dsDNA. Badnaviruses have been mainly identified in tropical regions and they have an extremely wide host range: a significant number of badnaviruses are pathogens of crop plants, such as banana, citrus, cacao, fig, grapevine, pepper, pineapple, raspberry, sugarcane, and sweet-potato, but many more are been found in ornamental plants such as bougainvillea, hibiscus, kalanchoe, lucky bamboo, pelargonium and wisteria. So far badnaviruses have been shown to be transmitted by two types of insect vector, aphids and mealybugs. A total of 68 species are currently recognized by the ICTV, and in addition there are six tentative species in the genus. There is also evidence in the form of endogenous viral elements (viral DNA that is integrated in the plant genome) that many more plant species have been exposed to badnaviruses throughout their evolution. The molecular species demarcation criteria of 80% identity of nucleotide sequences has been set 20 years ago, and now that so many more badnavirus complete genome sequences are available, this demarcation threshold should be revisited.

Classification

The genus *Badnavirus*, typified by *Commelina yellow mottle virus*, currently contains 68 officially recognized species and six tentative virus species (Table 1), making it the largest of all genera in the family *Caulimoviridae*. All members of the family *Caulimoviridae* have the conserved core of replication and structural proteins found within the order *Ortervirales*, called the gag-pol, which are names originating from the retrovirus literature and are abbreviations of 'group antigens' and 'polymerase', respectively. In badnaviruses, the gag gene encodes the major capsid protein (CP) and the pol gene, the replication-associated enzymes, namely the aspartic protease and reverse transcriptase/ribonuclease H1 proteins. The caulimoviruses are sometimes referred to as 'plant-infecting pararetroviruses' but the term 'pararetrovirus' has no taxonomic standing and is purely a descriptive term used to describe reverse-transcribing viruses that encapsidate a double-stranded DNA (dsDNA) form of the viral genome rather than single-stranded RNA (ssRNA) in the case of retroviruses. Caulimoviruses are united with all members of the *Ortervirales* by alternating between double-stranded DNA and single-stranded RNA during the replication cycle through cycles of transcription and reverse transcription. The RNA intermediate in the replication cycle is termed the pregenomic (pg) RNA. The key characteristic that differentiates the genus *Badnavirus* from all other members of the family *Caulimoviridae* is genome organization.

The criteria for demarcation of different badnavirus species is less than 80% nucleotide identity within the reverse transcriptase and ribonuclease H1 coding regions of ORF3, differences in host range and vector specificities. Evolutionary relationships within the genus are illustrated in Fig. 1.

Table 1 List of species, tentative species and unassigned viruses in the genus *Badnavirus*, family *Caulimoviridae*. Their acronym, accession # of an exemplar, natural hosts, geographical distribution, and their capacity to be endogenous are also indicated

Species	Acronym	Accession # of exemplar isolate	Natural hosts	Geographical distribution	Endogenous?
<i>Aglaonema bacilliform virus</i>	ABV	MH384837	<i>Aglaonema modestum</i>	USA	
<i>Banana streak GF virus</i>	BSGFV	AY493509	<i>Musa</i> spp.	All banana-growing countries	Yes
<i>Banana streak IM virus</i>	BSIMV	HQ593112	<i>Musa</i> spp.	Australia, China, Cuba, Guadeloupe, Indonesia, Kenya, Uganda	Yes
<i>Banana streak MY virus</i>	BSMYV	AY805074	<i>Musa</i> spp.	All banana-growing countries	Yes
<i>Banana streak OL virus</i>	BSOLV	AJ002234	<i>Musa</i> spp.	All banana-growing countries	Yes
<i>Banana streak UA virus</i>	BSUAV	HQ593107	<i>Musa</i> spp.	Kenya, Uganda	
<i>Banana streak UI virus</i>	BSUIV	HQ593108	<i>Musa</i> spp.	Uganda	
<i>Banana streak UL virus</i>	BSULV	HQ593109	<i>Musa</i> spp.	Uganda	
<i>Banana streak UM virus</i>	BSUMV	HQ593110	<i>Musa</i> spp.	Uganda	
<i>Banana streak VN virus</i>	BSVNV	AY750155	<i>Musa</i> spp.	China, Viet Nam	
<i>Birch leaf roll-associated virus</i>	BLRaV	MG686419	<i>Betula pubescens</i> , <i>Betula pendula</i>	Finland, Germany	
<i>Blackberry virus F</i>	BVF	KJ413252	<i>Rubus</i> hybrids	USA	Yes
<i>Bougainvillea chlorotic vein banding virus</i>	BCVBV	EU034539	<i>Bougainvillea spectabilis</i>	Brazil, China, Malaysia, Taiwan	
<i>Cacao bacilliform Sri Lanka virus</i>	CBSLV	MF642736	<i>Theobroma cacao</i>	Sri Lanka	
<i>Cacao mild mosaic virus</i>	CMMV	KX276640	<i>Theobroma cacao</i>	Puerto Rico, Trinidad and Tobago	
<i>Cacao swollen shoot CD virus</i>	CSSCDV	JN606110	<i>Theobroma cacao</i>	Cote d'Ivoire	
<i>Cacao swollen shoot CE virus</i>	CSSCEV	MF642720	<i>Theobroma cacao</i>	Cote d'Ivoire, Ghana	
<i>Cacao swollen shoot Ghana M virus</i>	CSSGMV	MF642724	<i>Theobroma cacao</i>	Ghana	
<i>Cacao swollen shoot Ghana N virus</i>	CSSGNV	MF642725	<i>Theobroma cacao</i>	Ghana	
<i>Cacao swollen shoot Ghana Q virus</i>	CSSGQV	MF642734	<i>Theobroma cacao</i>	Ghana	
<i>Cacao swollen shoot Togo A virus</i>	CSSTAV	AJ781003	<i>Theobroma cacao</i>	Indonesia (?), Ghana, Togo	
<i>Cacao swollen shoot Togo B virus</i>	CSSTBV	L14546	<i>Theobroma cacao</i>	Cote d'Ivoire, Togo	
<i>Cacao yellow vein banding virus</i>	CYVBV	KX276641	<i>Theobroma cacao</i>	Trinidad and Tobago	
<i>Canna yellow mottle associated virus</i>	CaYMaV	KX066020	<i>Canna indica</i>	USA	
<i>Canna yellow mottle virus</i>	CaYMV	KU168312	<i>Alpinia purpurata</i> , <i>Canna indica</i> , <i>Canna</i> hybrids, <i>Piper betle</i>	Austria, China, Italy, India, Kenya, Russia, The Netherlands, USA	
<i>Citrus yellow mosaic virus</i>	CIYMV	AF347695	<i>Citrus x aurantifolia</i> , <i>Citrus x jambhiri</i> , <i>Citrus x limonia</i> , <i>Citrus maxima</i> , <i>Citrus reticulata</i> , <i>Citrus x sinensis</i>	India	

(Continued)

Table 1 Continued

Species	Acronym	Accession # of exemplar isolate	Natural hosts	Geographical distribution	Endogenous?
<i>Codonopsis</i> vein clearing virus	CoVCV	MK048221	<i>Codonopsis lanceolata</i>	South Korea	
<i>Commelina</i> yellow mottle virus	CoYMV	X52938	<i>Commelina diffusa</i>	Guadeloupe	
<i>Dioscorea</i> bacilliform AL virus	DBALV	KX008573	<i>Dioscorea alata</i> , <i>Dioscorea bulbifera</i> , <i>Dioscorea cayennensis</i> , <i>Dioscorea esculenta</i> , <i>Dioscorea rotundata</i> , <i>Dioscorea transversa</i> , <i>Dioscorea trifida</i>	Brazil, Nigeria, Tonga, Vanuatu	
<i>Dioscorea</i> bacilliform AL virus 2	DBALV2	KY827395	<i>Dioscorea alata</i>	Papua New Guinea	
<i>Dioscorea</i> bacilliform ES virus	DBESV	KY827394	<i>Dioscorea dumetorum</i> , <i>Dioscorea esculenta</i>	Fiji, Nigeria	
<i>Dioscorea</i> bacilliform RT virus 1	DBRTV1	KX008574	<i>Dioscorea rotundata</i>	Nigeria	
<i>Dioscorea</i> bacilliform RT virus 2	DBRTV2	KX008577	<i>Dioscorea rotundata</i>	Nigeria, Samoa	
<i>Dioscorea</i> bacilliform SN virus	DBSNV	DQ822073	<i>Dioscorea sansibarensis</i>	Benin	
<i>Dioscorea</i> bacilliform TR virus	DBTRV	KX430257	<i>Dioscorea alata</i> , <i>Dioscorea trifida</i>	French Guyana, Guadeloupe	
Fig badnavirus 1	FBV1	JF411989	<i>Ficus carica</i>	USA	Yes
Gooseberry vein banding associated virus	GVBaV	JQ316114	<i>Ribes nigrum</i> , <i>Ribes rubrum</i> , <i>Ribes uva-crispa</i>	Bosnia and Herzegovina, Canada, Czech Republic, The Netherlands, United Kingdom, Croatia	
Grapevine badnavirus 1	GBV 1	MF781082	<i>Vitis vinifera</i>	Croatia	
Grapevine Roditis leaf discoloration-associated virus	GRLDaV	HG940503	<i>Ficus carica</i>	Croatia, Iran, Italy, Greece, Turkey	
Grapevine vein clearing virus	GVCV	JF301669	<i>Vitis vinifera</i>	USA	
Jujube mosaic-associated virus	JuMaV	KX852476	<i>Ziziphus jujube</i>	China	
<i>Kalanchoe</i> top-spotting virus	KTSV	AY180137	<i>Kalanchoë blossfeldiana</i>	USA	
Mulberry badnavirus 1	MBV1	LN651258	<i>Morus alba</i>	Lebanon, Iran	
Pineapple bacilliform CO virus	PBCOV	GU121676	<i>Ananas comosus</i> var. <i>comosus</i>	All pineapple-growing countries	
Pineapple bacilliform ER virus	PBERV	EU377672	<i>Ananas comosus</i> var. <i>erectifolius</i>	Australia	
Piper yellow mottle virus	PYMoV	KC808712	<i>Piper argyrophyllum</i> , <i>Piper attenuatum</i> , <i>Piper barberi</i> , <i>Piper betle</i> , <i>Piper colubrinum</i> , <i>Piper galeatum</i> , <i>Piper longum</i> , <i>Piper ornatum</i> , <i>Piper nigrum</i> , <i>Piper sarmentosum</i> , <i>Piper trichostachyon</i>	China, India, Indonesia, Sri Lanka, Thailand, The Philippines, Viet Nam	Yes
<i>Rubus</i> yellow net virus	RYNV	KM078034	<i>Rubus idaeus</i>	Canada, Bosnia and Herzegovina, Finland, United Kingdom, USA	Yes
<i>Schefflera</i> ringspot virus ^a	SRV	MN850490	<i>Schefflera</i> sp.	USA	
<i>Spirea</i> yellow leafspot virus	SYLSV	AF299074	<i>Spirea</i> spp.	USA	

<i>Sugarcane bacilliform Guadeloupe A virus</i>	SCBGAV	FJ824813	<i>Saccharum officinarum</i>	Guadeloupe, China	
<i>Sugarcane bacilliform Guadeloupe D virus</i>	SCBGDV	FJ439817	<i>Saccharum officinarum</i>	Guadeloupe, China	
<i>Sugarcane bacilliform IM virus</i>	SCBIMV	AJ277091	<i>Saccharum officinarum</i>	Australia, China, India	
<i>Sugarcane bacilliform MO virus</i>	SCBMOV	M89923	<i>Saccharum officinarum</i>	China, India, Morocco	
<i>Sweet potato pakakuy virus</i>	SPPV	FJ560943	<i>Ipomoea batatas</i>	Cameroon, China, Peru, South Korea, South Africa, Spain, Tanzania	
<i>Taro bacilliform CH virus</i>	TaBCHV	KP710178	<i>Colocasia esculenta</i> , <i>Xanthosoma</i> sp.	China, Ethiopia, Kenya, Tanzania, Uganda, USA	
<i>Taro bacilliform virus</i>	TaBV	AF357836	<i>Colocasia esculenta</i> , <i>Xanthosoma</i> sp.	Australia, Fiji, French Polynesia, Kenya, New Caledonia, Papua New Guinea, Tanzania, Uganda, Vanuatu	
<i>Wisteria badnavirus 1</i>	WBV1	KX168422	<i>Wisteria sinensis</i>	China	
<i>Yacon necrotic mottle virus</i>	YNMoV	KM229702	<i>Smallanthus sonchifolius</i>	South Korea	
<i>Camellia lemon glow virus</i>	CLGV	MN542417	<i>Camellia japonica</i>	USA	
<i>Cycad leaf necrosis virus</i>	CLNV	EU853709	<i>Zamia fischeri</i>	USA	
<i>Dioscorea bacilliform RT virus 3</i>	DBRTV3	MF476845	<i>Dioscorea rotundata</i>	Nigeria	
<i>Dracaena mottle virus</i>	DrMV	DQ473478	<i>Dracaena sanderiana</i>	China	Yes
<i>Epiphyllum mottle-associated virus</i> (syn. <i>Pitaya badnavirus 1</i>)	EpMoaV	MH396440	<i>Discocactus</i> spp., <i>Epiphyllum</i> spp. <i>Hylocereus polyrhizus</i> , <i>Opuntia</i> spp. <i>Schlumbergera</i> spp.	China, USA	
<i>Green Sichuan pepper vein clearing-associated virus</i>	GSPVCaV	MK371353	<i>Zanthoxylum schinifolium</i>	China	
<i>Ivy ringspot-associated virus</i> (syn. <i>Schefflera ringspot virus</i>) ^a	IRSaV	MN850490	<i>Hedera helix</i> , <i>Schefflera</i> sp.	South Africa, USA	
<i>Pagoda yellow mosaic associated virus</i>	PYMaV	KJ013302	<i>Styphnolobium japonicum</i>	China	
<i>Pelargonium vein banding virus</i>	PVBV	GQ428155	<i>Pelargonium x hortorum</i>	USA	
<i>Polyscias mosaic virus</i>	PoMV	MH475918	<i>Polyscias fruticosa</i>	USA, Nigeria	
<i>Tentative Species in the Genus</i>	Acronym	Accession # of exemplar isolate	Natural hosts	Geographical distribution	Endogenous?
<i>Agave badnavirus A</i>	ABVA	MH898467	<i>Agave tequilana</i>	USA	
<i>Aucuba ringspot virus</i>	AuRSV	LC487411	<i>Aucuba japonica</i>	Japan	
<i>Banana streak CA virus</i>	BSCAV	HQ593111	<i>Musa</i> spp., <i>Saccharum officinarum</i>	Australia, China, Guadeloupe, Kenya, Mauritius	
<i>Chestnut mosaic virus</i>	ChMV	MT269853	<i>Castanea crenata</i> x <i>C. sativa</i> , <i>Castanea dentata</i> , <i>Castanea mollissima</i> , <i>Castanea sativa</i>	China, France, Italy, USA	
<i>Dioscorea bacilliform virus A</i>	DBVA	MH898490	<i>Dioscorea composita</i>	China	
<i>Enset leaf streak virus</i>	ELSV	MF991909	<i>Ensete ventricosum</i>	Ethiopia	

^aOnly a partial genome sequence is available for *Schefflera ringspot virus* (MH475920) and this sequence is 99.1 identical to the genome sequence of *Ivy ringspot-associated virus* (MN850490), suggesting the two species are synonymous.

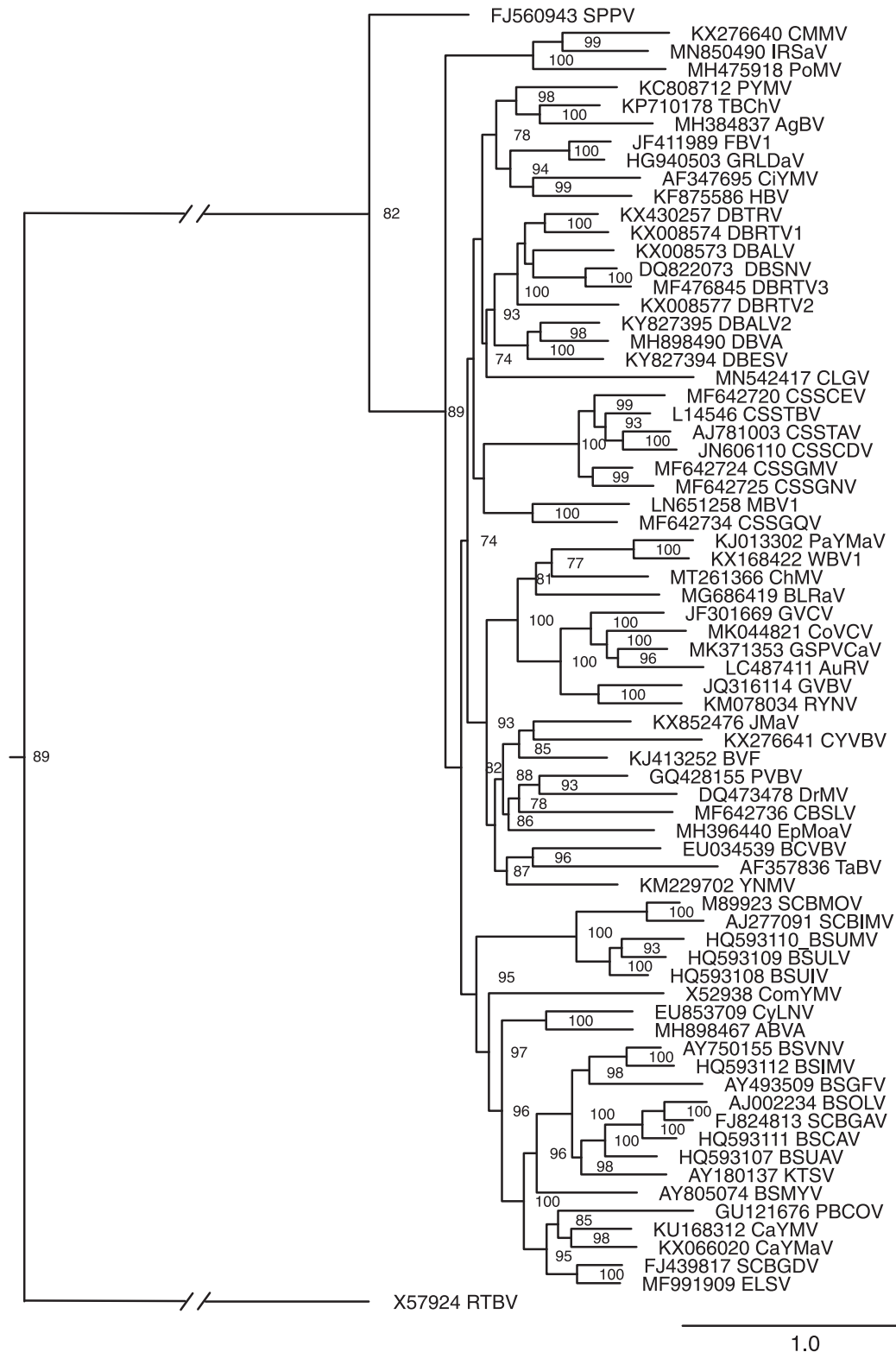


Fig. 1 Phylogram depicting relationships within the genus *Badnavirus*, based on a maximum likelihood analysis of a *pol* gene sequence alignment using the IQ-TREE web server. The *pol* gene sequence alignment of approved badnavirus species was downloaded from the ICTV *Caulimoviridae* Resources Page and amended with sequences of tentative badnavirus species that are available on GenBank. GenBank accession codes and virus acronyms (see [Table 1](#)) are provided at the tips of the branches; rice tungro bacilliform (RTBV) is the outgroup. Ultrafast bootstrap support values (%) from 1000 replicates are shown in the nodes of the branches (only values > 70% are shown).

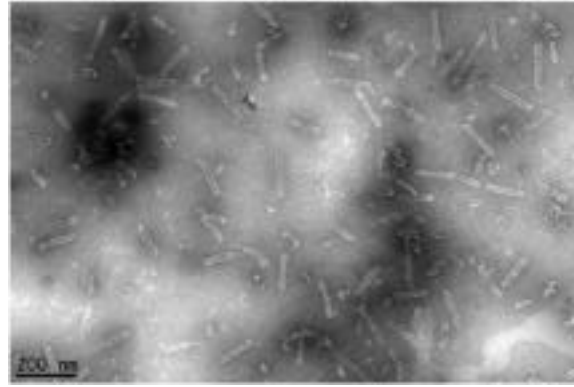


Fig. 2 Virions of Banana streak IM virus (BSIMV), negatively stained with 1% ammonium molybdate pH 7. Micrograph courtesy of K.S. Crew.

Virion Structure

Virions are bacilliform with parallel sides and rounded ends, and modal dimensions of 30 nm × 130 nm (Fig. 1). However, virions that are two to three times this modal length are sometimes observed, and these contain non-covalently bonded concatemers of viral genomic DNA that are up to four monomeric units in length. Two types of virion are observed in mulberry plants infected by Mulberry badnavirus 1 (MBV1), one that has normal dimensions and a second that is thinner (*c.* 20 nm diameter) and has a looser capsid structure. These aberrant virions contain a shorter form of the MBV1 genome.

The main structural protein of the virion is the capsid protein (CP), while a second protein, the virion-associated protein (VAP), is thought to decorate the surface of the virion by occupying the pores between the capsomers. At the C-terminus of the VAP, there is a proline-rich region that binds to double-stranded DNA in a sequence non-specific manner, and putatively interacts with the encapsidated DNA. Conversely, the N-terminus of the VAP is thought to be surface-exposed and contains a coiled-coil motif that could interact with other viral, host or insect proteins to mediate functions in the virus' life cycle such as cell-to-cell movement and vector transmission (Fig. 2).

Genome

The encapsidated form of the badnaviral genome is a non-covalently closed dsDNA molecule of 7.0–9.3 kbp. There are discontinuities in each strand of DNA, which arise as a consequence of synthesis of the DNA strands by reverse transcription. Synthesis of the negative strand of DNA is primed by a cytosolic initiator methionine tRNA ($tRNA^{met}$), and the discontinuity on this strand is adjacent to the 3' end of the homology to the $tRNA^{met}$. On the positive strand, the discontinuity is located within a polypurine-rich region within ORF3. By convention, position 1 of the genome is designated the first nucleotide of the $tRNA^{met}$ binding motif (5' T₁GGTATCAG.... 3') on the positive strand.

The majority of badnavirus genomes have three conserved open reading frames (ORFs), which are in different translational reading frames (Fig. 3). For the most part, the start codons for these ORFs are AUG but for some species, non-conventional start codons are utilized, particularly at the beginning of ORF1 (e.g., CUG for BSMYV ORF1). Other ORFs have been described but these are not found in all species and evidence that these ORFs are translated is lacking. It is typical for the stop codons of ORFs 1 and 2 to overlap the start codons of the succeeding ORFs. ORF1 encodes a small protein of *c.* 21–25 kDa that has no assigned function but is recognizable through the presence of a conserved domain, pfam07028. ORF2 also encodes a small protein of *c.* 15 kDa, the VAP. ORF3 encodes a large polyprotein of *c.* 197–232 kDa, with conserved movement protein (MP), capsid protein (CP) and nucleocapsid, aspartic protease, reverse transcriptase and RNase H domains in that order. All domains in the polyprotein bar the MP have strong sequence conservation at primary and secondary structural levels with paralogous proteins of retroviruses and long terminal repeat retrotransposons, and are classified in the same protein families.

Two variants of the typical badnavirus genome organization are known. MBV1 has a genome sequence with one long ORF, which appears to be the equivalent of a fusion of ORFs 1 and 3 in other badnaviruses based on the arrangement of conserved domains. The genome sequence of sweet potato pakakuy virus (SPPV), which contains four ORFs, represents yet another permutation of the typical badnavirus genome. ORF3 encodes the MP and CP, while ORF4 encodes the aspartic protease and reverse transcriptase/ribonuclease H1 proteins. Hence for SPPV, ORF3 in other badnaviruses is split in two.

The intergenic region in the genome between the end of ORF3 and the beginning of ORF1 in the typical badnavirus genome is long and contains transcription regulatory motifs. Between the end of ORF3 and the $tRNA^{met}$ motif, there is the equivalent of the cauliflower mosaic virus 35S promoter, identifiable by the presence of a TATA box. Due to prospects of using the badnavirus promoter for genetic engineering of plants, there has been considerable work done to identify the minimal promoter sequences and to identify cis-acting elements that control the specificity of tissue expression. Downstream of the promoter is a polyadenylation signal but only a small percentage of the RNA transcripts have a poly(A) tail.

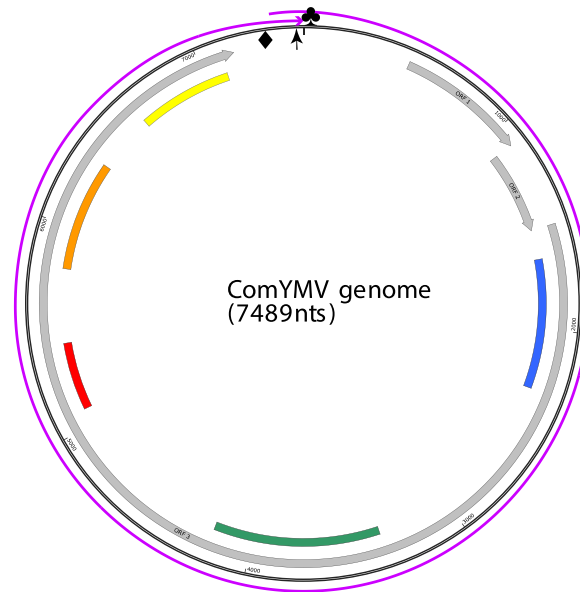


Fig. 3 Genome map for *Commelina* yellow mottle virus (ComYMV), type species of the genus *Badnavirus*. The black circle represents the genomic DNA, the purple circle is the pregenomic RNA and the gray curved boxes are the open reading frames. Marked on the genomic DNA are the positions of the TATA box (black diamond), polyadenylation signal (black arrow) and tRNA^{met}-binding site (black clover). The blue, green, red, orange and yellow curved boxes represent the movement protein, capsid protein, aspartic protease, reverse transcriptase and RNaseH domains on ORF 3, respectively.

Life Cycle

The replication cycle of badnaviruses has not been investigated to any great extent, and most knowledge is based on extrapolation from model plant viruses such as cauliflower mosaic virus (CaMV), and to a lesser extent, rice tungro bacilliform virus (RTBV).

Following entry of the badnavirus virion into the cell, the encapsidated viral DNA needs to be transported to the nucleus for replication to begin. A putative nuclear localization signal (NLS) is present in the basic, intrinsically disordered region of the NC domain of the CP. For RTBV, this NLS facilitates interaction between the CP and importin α but as the NLS is within the mature virion, is unlikely to play a role in docking of the virion to the nuclear pore but rather entry of the genomic DNA after either partial or complete virion decapsidation. There are two major CP isoforms that vary in length of the acidic N-terminus of the protein. It is not clear whether this variation in length is due to the presence of alternative aspartic protease cleavage sites or is a consequence of targeted degradation of the protein. With CaMV, the extended N-terminus inhibits nuclear targeting of the protein, ensuring that it is localized within the cytoplasm where virion assembly occurs.

Once the genomic DNA enters the nucleus, it is assumed that replication follows the pathway of CaMV, whereby the discontinuities in the genomic DNA are repaired to form a covalently closed circle, and the DNA then associates with histone proteins and RNA polymerase II to form transcriptionally active complexes called minichromosomes. A single, terminally redundant, slightly greater than genome-length RNA transcript is produced called the pregenomic RNA (pgRNA), and there is no evidence of subgenomic or spliced RNA transcripts. Following the replication model proposed for CaMV, it is thought that the RNA polymerase II enzyme begins transcription, ignores the polyadenylation signal on the first pass, continues transcribing around the circle and drops off the DNA template when it reaches the polyadenylation signal for the second time. Hence, the region between the transcription start site and the polyadenylation signal is duplicated at the 5' and 3' ends of the pgRNA.

It has been a mystery as to where in the cell, virion assembly occurs but recent studies suggest that badnaviruses produce viroplasm (virus factories) that are equivalent in function to those of CaMV but clearly different in composition because badnaviruses lack the major viroplasm matrix protein produced by caulimoviruses. These putative viroplasms have a characteristic structure comprising an undifferentiated core, an encircling lacuna where reverse transcription may occur, and an outside ring of differentiating virions. A gap is present in the outside ring, allowing the import of newly synthesized viral proteins from the surrounding ribosomes.

The pgRNA, which also acts as the messenger RNA, is polycistronic and the downstream ORFs are likely translated by a leaky ribosome scanning mechanism, whereby the start codons of ORFs 1 and 2 are in sub-optimal initiation contexts and are bypassed by the ribosome on a fraction of occasions in favor of the next downstream start codon. Consistent with this translation mechanism, internal AUG codons are rare to absent in ORFs 1 and 2. The 5' leader sequence of the pgRNA is long and folds into a stable hairpin structure, which is immediately preceded by a short upstream ORF (uORF). It is proposed that translation initiates at the start of the uORF and the ribosome then shunts across the hairpin structure to a landing site near the beginning of ORF1, where translation recommences. The uORF could regulate sorting of the ribosomes between the three ORFs to maintain balanced production of the different proteins.

The large polyprotein encoded by ORF3 must be post-translationally processed to release each individual component protein. Within the polyprotein is an aspartic protease, which belongs to the retrovirus aspartic protease family (pfam00077) and has a highly conserved active site (motif Asp-Ser/Thr-Gly). It is hypothesized that the aspartic protease autocleaves and then processes other proteins in the polyprotein. Based on studies with retroviruses, the active form of the protease is likely to be a β homodimer. Substrate sites for the enzyme are not strictly conserved at the amino acid sequence level and cleavage sites must be empirically determined. No badnaviral aspartic protease has been biochemically characterized and there has only been one attempt to map a cleavage site, at the N-terminus of the BSMYV capsid protein.

The final step in the badnavirus replication cycle, reverse transcription, is thought to follow the general model for CaMV. The (–) strand of DNA is first synthesized using the pgRNA as the template, the virus-encoded reverse transcriptase as catalyst and a tRNA^{met} molecule to prime the reaction. When the reverse transcriptase reaches the 5' end of the pgRNA, the nascent end of the DNA melts away from the RNA and anneals to the identical stretch of sequence at the 3' end of the pgRNA. Reverse transcription then continues around the circle until the original tRNA^{met}-binding site is encountered. At some point, the pgRNA and tRNA^{met} molecule must be removed, presumably through the action of the virus-encoded RNaseH enzyme. Synthesis of the (+) strand of DNA then begins, presumably using a host-encoded DNA polymerase and a primer that has yet to be characterized.

Transmission electron microscopy studies suggest that badnaviruses are trafficked between cells as whole virions and this is enabled by the formation of a tubule that displaces the contents of the plasmodesmum and increases its size exclusion limit. Badnaviruses produce an example of the 30K superfamily of viral MP, which is most closely related to those produced by nepo- and ideoviruses.

Epidemiology

The epidemiology of many of the badnaviruses is poorly understood, and most knowledge derives from the study of a small number of diseases with the largest economic impacts, such as cacao swollen shoot and banana streak disease.

The majority of badnavirus species are transmitted by mealybugs (family Pseudococcidae), which are most abundant in warm, moist climates or in greenhouses. Mealybugs inhabit protected sites on the plant such as the underside of plant leaves and stems, in leaf axils and the underneath leaf sheaths, and on the roots, and hence their presence may be overlooked. Where known, the virus-vector specificity of mealybug transmission is weak, although there is very little data on this topic due to the difficulties of rearing mealybugs and doing controlled transmission experiments. At least fourteen species of mealybug are capable of transmitting the cacao swollen shoot complex of viruses, although three species are considered the most important vectors, namely *Formicococcus njalensis*, *Planococcus citri* and *Ferrisia virgata*.

The most mobile growth stages of the mealybug are the first and second instars of nymph, and these are also known as 'crawlers'. Crawlers move over most parts of the plant, whereas older nymphs and adults will retreat to more protected positions on the plant. Transmission of badnaviruses can occur when either the crawlers actively cross leaf bridges between plants or are passively transported by the wind or carried by attendant ants. Ants have a strong mutualistic relationship with mealybugs, as they feed on their honeydew secretions and in turn, provide protection from parasites and predators and improve hygiene by cleaning up waste products. Mealybugs can be indirectly controlled by reducing ant populations using insecticides.

The second most important group of badnavirus vectors is aphids, and again these transmit the viruses in the semi-persistent manner. Only five species of badnavirus are known to be transmitted by aphids, namely chestnut mosaic virus (ChMV), gooseberry vein-banding associated virus (GVbAV), Rubus yellow net virus (RYNV), Spirea yellow leaf spot virus (SYLSV) and grapevine vein clearing virus (GVCV). All four viruses are closely related, and all originate from temperate regions of the Northern Hemisphere. Evolution of this novel virus-vector relationship may reflect the greater abundance of aphids in higher latitudes of Eurasia and North America relative to mealybugs.

The only other type of insect that can vector badnaviruses is the black pepper lace bug (*Diconocoris distanti*), which together with the citrus mealybug (*Planococcus citri*), is reported to transmit Piper yellow mottle virus (PYMoV); but this finding does need independent verification.

True seed transmission has been reported for several badnavirus species including PYMoV in black pepper, Kalanchoe top spotting virus (KTSV) in *Kalanchoë blossfeldiana*, SPPV in *Ipomoea batatas* and an unidentified species of taro bacilliform virus in *Colocasia esculenta*. A report that BSMYV is seed-transmitted in banana is ambiguous as there is the possibility that the infections in seedlings originated from activation of endogenous badnaviral elements. While seed transmission of many badnaviruses is theoretically possible, it is likely to be epidemiological insignificant as many of the host species are vegetatively propagated.

Where known, most badnaviruses have narrow host ranges in nature and alternative hosts are unlikely to play an important role in the virus epidemics but there are important exceptions. It is well documented that the indigenous rainforest trees in West Africa are the origins of the badnaviruses associated with cacao swollen shoot disease. When the rainforest was cleared for cropping, some trees were left to provide shade for the cacao plants, and these acted as primary sources of inoculum for the epidemics. The most important alternative host trees for cacao swollen shoot in Ghana are *Cola gigentia*, *Cola chlamydanthia* and *Ceiba pentandra*, all from the same Malvaceae family as *Theobroma cacao*. However, pockets of older, infected cacao trees in the same or neighboring plantations are now the most important sources of inoculum and disease control is dependent on operation of an effective roguing program.

Other examples of cross-species transmission of badnaviruses include banana streak CA virus, which as the name suggests, was first described from banana but appears to be more common in sugarcane. These two crops grow on the same types of soil, and often on the same or neighboring farms, and it is thought that infections in banana originate from nearby sugarcane plants. Canna yellow mottle virus (CaYMV), a virus that is ubiquitous in ornamental canna plants in the tropics, is responsible for diseases in betel nut (*Piper betel*) in Pakistan and ornamental ginger (*Alpinia purpurata*) in Hawaii. Finally, wild members of the Vitaceae family such as *Vitis rupestris* and *Ampelopsis cordata* are important alternative hosts of GVCV, which causes grapevine decline in the American Midwest. The frequency of cross-species transmission of these viruses is unknown but does not need to be common for the infections to proliferate as all crops are vegetatively propagated.

Clinical Features

Badnaviruses typically produce chlorotic streaks, mild mottles and sometimes ringspot patterns on the leaves of monocots, and mottle and mosaic patterns, vein clearing and ring spots on the leaves of dicots (Fig. 4). Plants affected by cacao swollen shoot disease have distinctive protuberances on the stems, caused by proliferation of the xylem, phloem and cortex cells.

Plants infected by badnaviruses may be asymptomatic or symptoms sporadic in occurrence, as best documented with the various species of the banana streak disease complex (BSD). The factors triggering symptom expression in banana are poorly understood but there is a clear trend of symptom severity increasing with maturity of the plant, reaching a maximum at about the time of bunch initiation. For BSD, virus titer is also positively correlated with symptom severity but infection often leads to cell death, causing the chlorotic streaks to turn necrotic, and virus titer to decline again.

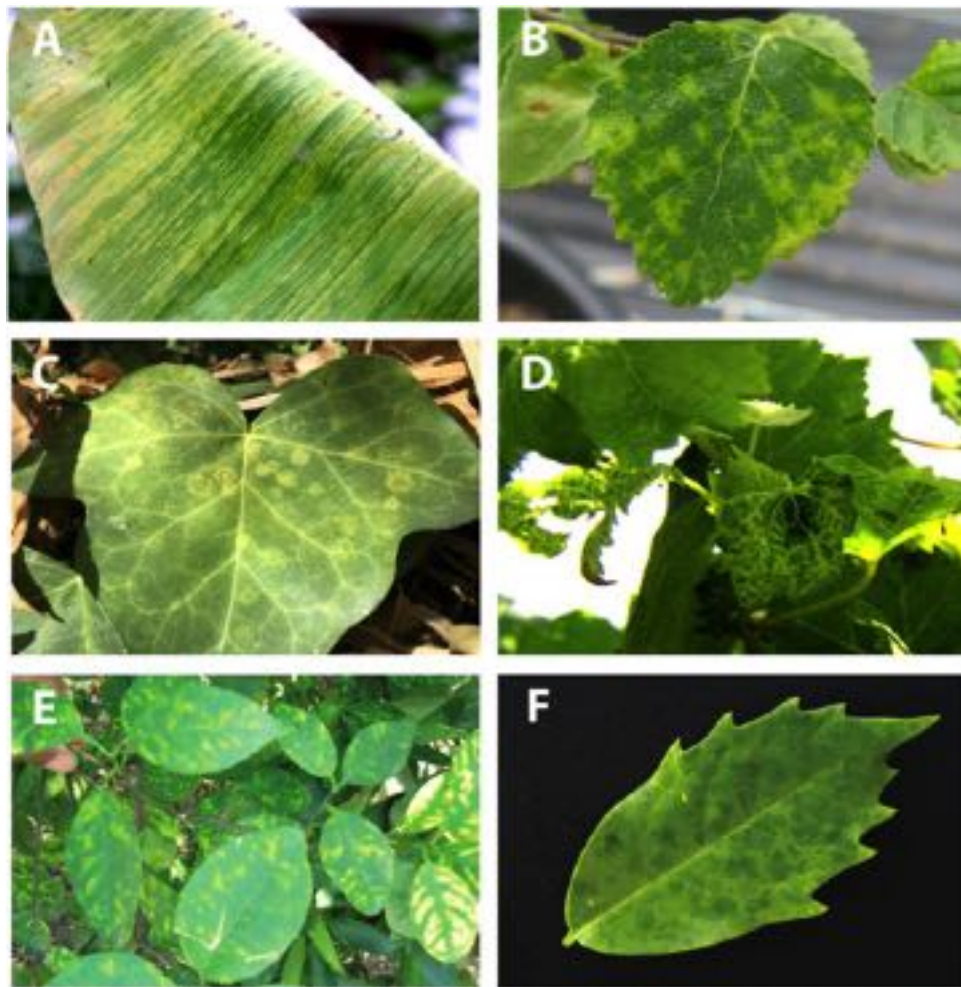


Fig. 4 Disease symptoms caused by badnaviruses: A, *Musa* sp. AAB group infected by banana streak MY virus (BSMYV) (photograph courtesy of A.D.W. Geering); B, *Betula pubescens* infected by birch leaf roll-associated virus (BLRaV) (photograph courtesy of A. Rumbou); C, *Hedera helix* infected by ivy ringspot-associated virus (IRSaV) (photograph courtesy of J.T. Burger); D, *Vitis vinifera* cv. Cabernet Franc infected by Grapevine vein clearing virus (GVCV) (photograph courtesy of W. Qiu); E, *Citrus x sinensis* infected by citrus yellow mosaic virus (CiYMV) (photograph courtesy of V.S. Baranwal); F, *Aucuba japonica* infected by Aucuba ringspot virus (AuRSV) (photograph courtesy of A. Uke).

Pathogenesis

There is scant information on how badnaviruses cause disease. The tissue tropism of badnaviruses is variable and dependent on the virus in question. Rubus yellow net virus infects all cell types of the raspberry leaf including the epidermis but is most common in the phloem companion cells. BSMYV most commonly occurs in the palisade and spongy mesophyll cells and rarely in the epidermal cells of the banana leaf. Cacao swollen shoot Togo B virus (CSSTBV) appears to be mainly vascular-limited, with most virions observed in the phloem companion cells and far less commonly in the xylem parenchyma cells.

Cytopathological studies suggest that badnaviruses trigger cell apoptosis although further studies are required to confirm these observations. Older infected cells exhibit severe plasmolysis, accumulate vacuoles, the chloroplasts become deformed, starch granules aggregate and the cells eventually lose all contents. The putative viroplasms can reach relatively large diameters and are easily visible by light microscopy using stains such as periodic acid-Schiff-toluidine blue.

One peculiar aspect of the biology of badnaviruses, particularly among plant-infecting viruses, is that new infections can arise following activation of viral sequences that are integrated in the nuclear genome of the host plant. These integrated sequences are commonly referred to as endogenous viral elements. Badnaviruses do not actively integrate into their host genome as do retroviruses but it is thought that the viral DNA is captured as a type of filler sequence during non-homologous or microhomology-mediated end-joining repair of double-stranded DNA breaks in the chromosomal DNA. Integration of viral DNA is part of a broader phenomenon of horizontal gene transfer involving all types of DNA, such as that from the plastid genomes.

The problem of endogenous badnaviral elements causing new infections was discovered in a quest to understand why new hybrid bananas generated by banana breeding programs around the world became infected with banana streak OL virus (BSOLV), despite the fact that the parent plants used to create the hybrids were healthy. A locus was discovered in one parent, cv. 'Obino l'Ewai' (*Musa* AAB group), which contains the complete genome sequence of BSOLV but divided in two fragments separated by a large scrambled region of fragments of non-contiguous and sometimes inverted viral sequence. An activation model was proposed involving two homologous recombinations, the first to excise the scrambled region and the second to join either end of the locus to form a circular, transcriptionally active form of the genome. This locus has now been better characterized and shown to be diallelic, with only one allele being infective. Activation is not spontaneous but triggered by external factors such as tissue culture propagation and interspecific hybridization. It is hypothesized that in a normal state, the endogenous badnaviral elements are transcriptionally silenced but certain stresses cause the transient release of this silencing.

The phenomenon of activation of endogenous badnaviral elements has now been extended to at least two more badnaviruses, banana streak GF virus (BSGFV) and banana streak IM virus (BSIMV) and as with BSOLV, activation involves recombinations to reconstitute complete viral genomes. All aforementioned replication-competent endogenous banana streak viruses are linked to the B genome of banana, which derives from *Musa balbisiana*. The A genome of banana (from *Musa acuminata*) also contains endogenous badnaviral elements but these are derived from different ancestral badnavirus species and all are regarded as being replication-defective due to the accumulation of mutations, rearrangements and deletions.

Endogenous badnaviral elements have now been detected in a broad range of plant species, both monocots and dicots. While it is not particularly surprising that badnaviral DNA is captured in the host genome, a more perplexing question is why the DNA has become fixed in the genome and has persisted for hundreds of thousands, even millions of years. It is generally assumed that the endogenous badnaviral elements confer a selective advantage to the plant, such as imparting resistance against the cognate infectious virus through activation of RNAi defense pathways. However, it is likely that the endogenous badnaviral elements could also help shape the transcriptome by, for example, contributing novel promoter elements or being a source of small non-coding RNAs. Furthermore, it is conceivable that badnaviral proteins are generated that may have been coopted for some other function in normal plant metabolism.

Diagnosis

Badnaviruses vary considerably in the difficulty of diagnosis, depending on the presence of endogenous badnaviral elements in the plant, as there is a need to distinguish these from encapsidated (exogenous) viral DNA. This problem is best exemplified by the various species of banana streak disease. When describing a new species of badnavirus, it is also essential to demonstrate that the DNA sequence derives from exogenous viral DNA.

The nuclear genome of cultivated banana contains multiple copies of endogenous badnaviral DNA, representing a large range of ancestral virus species. However, dessert banana such as the Cavendish subgroup of cultivars have an AAA genome, which is devoid of endogenous forms of Banana streak OL virus (BSOLV), Banana streak GF virus (BSGFV), Banana streak IM virus (BSIMV) and Banana streak MY virus (BSMYV). Hence, conventional PCR using species-specific PCR primers and total plant DNA extracts can be used for diagnosis. However, for hybrid plantain or dessert banana cultivars with AAB or ABB genotypes, conventional PCR cannot be used because of the problem of endogenous viral DNA giving false positive diagnostic results.

Two diagnostic technologies are routinely used to distinguish endogenous from exogenous viral DNA, these being immunocapture PCR and rolling circle amplification. The first step of immunocapture PCR is to trap virions on the sides of a PCR tube, and then to wash away other cellular components including plant nucleic acids. While this step does enrich for virions, plant DNA can bind directly to the tube wall and precautions need to be taken to prevent this from happening. Non-specific binding of DNA to the tube wall is a much slower process than the antibody-virion interaction, and limiting the immunocapture step to 3 h at room temperature helps prevent amplification of endogenous badnaviral elements. To alert the diagnostician of direct binding of

plant DNA, duplex PCRs can be done using primers against repetitive DNA sequences in the plant such as microsatellites. Another option is to do a DNase I digestion step after the immunocapture step, as the capsid shell protects the encapsidated DNA from nuclease degradation. One of the major constraints to the wider adoption of immunocapture PCR for diagnosis is the lack of availability of antibodies against many of the viruses but use of anti-peptide antibodies shows promise for rapidly resolving this problem. Some crop species such as banana, sugarcane and yam, are infected by a broad diversity of badnaviruses and universal badnavirus-group primers have been designed to cover most combinations of virus species.

Rolling circle amplification (RCA) using *Phi29* DNA polymerase is a very useful technique for characterizing badnaviral genomes and for the most part, the selected method has been to use a commercial RCA kit (TempliPhi™) that has been spiked with a mixture of degenerate and specific badnavirus primers to increase the sensitivity of detection. It is only the minichromosomal fraction of viral DNA that is amplified by RCA, as the discontinuities present in the encapsidated DNA cannot be bridged by the DNA polymerase. TempliPhi™ is a sequence independent amplification method, as DNA polymerization is primed with random hexamers, and this has sometimes created problems with interpreting results, as circular plant plastid DNA sequences may also be amplified. However, major improvements to the specificity of amplification by RCA have been made by formulating reaction mixtures from individual components, replacing the random hexamers with a panel of degenerate badnavirus primers, and optimizing amplification temperatures.

Prevention

The two foundations of badnavirus prevention are the use of virus-indexed planting material and the operation of disease roguing programs. Virus indexing is particularly important for crops such as banana, which are propagated by tissue culture and a single meristem can give rise to 500 progeny plants. For chronically-infected, vegetatively propagated plants, somatic embryogenesis and meristem-tip culture can be used to eliminate virus, and the success of the technique is enhanced by combining with chemotherapy. The use of insecticides to control mealybug vectors is generally considered to be ineffective due to the waxy, water-repellant coatings on the insect and the cryptic places where the insects feed but there may be a place for systemic insecticides in a management program.

A problem exists with hybrid (A x B genome) bananas and plantains as the practice of tissue culture in itself provides the trigger for activation of endogenous badnaviral elements and this is particularly a concern for the banana breeding programs as embryo rescue is integral to the breeding process. PCR diagnostic assays have now been developed to screen for replication-competent alleles of the endogenous banana streak viruses, and use of these diagnostic assays allows identification of cultivars that are particularly prone to becoming infected when propagated by tissue culture. The long term solution to the issue of endogenous badnaviral elements causing infection is to create plants that are devoid of replication-competent endogenous virus alleles by conventional breeding techniques or by inactivating these alleles using gene-editing techniques such as CRISPR/Cas9.

Cacao swollen shoot is undoubtedly the most economically important disease caused by badnaviruses, and no plant disease resistance is available and therefore the primary control strategy has been roguing, resulting in the destruction of over 200 million trees in Ghana since 1946. Mild strain cross protection does not protect plants from superinfection by virulent strains but does slow the rate of spread by nearly half. Oil palm and citrus barriers can also be used to slow the reintroduction of cacao swollen shoot into newly planted, healthy crops and at the same time provide an alternative source of income for the farmers.

Further Reading

- Bhat, A.I., Hohn, T., Selvarajan, R., 2016. Badnaviruses: The current global scenario. *Viruses* 8, 177.
- Chabannes, M., Baurens, F.-C., Duroy, P.-O., *et al.*, 2013. Three infectious viral species lying in wait in the banana genome. *Journal of Virology* 87, 8624–8637.
- Cheng, C.P., Tzafir, I., Lockhart, B.E., Olszewski, N.E., 1998. Tubules containing virions are present in plant tissues infected with Commelina yellow mottle virus. *Journal of General Virology* 79, 925–929.
- Diop, S.I., Geering, A.D.W., Alfama-Depauw, F., *et al.*, 2018. Tracheophyte genomes keep track of the deep evolution of the *Caulimoviridae*. *Scientific Reports* 8, 572.
- Krupovic, M., Blomberg, J., Coffin, J.M., *et al.*, 2018. *Ortervirales*: New virus order unifying five families of reverse-transcribing viruses. *Journal of Virology* 92, e00515-18.
- Medberry, S.L., Lockhart, B.E., Olszewski, N.E., 1990. Properties of Commelina yellow mottle virus's complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. *Nucleic Acids Research* 18, 5505–5513.
- Ngo, T.H., Webb, R., Crew, K.S., *et al.*, 2020. Identification of putative viroplasms within banana cells infected by banana streak MY virus. *Journal of General Virology*. doi:10.1099/jgv.0.001498.
- Pooggin, M.M., Ryabova, L.A., 2018. Ribosome shunting, polycistronic translation, and evasion of antiviral defenses in plant pararetroviruses and beyond. *Frontiers in Microbiology* 9, 644.
- Sukal, A.C., Kidanemariam, D.B., Dale, J.L., Harding, R.M., James, A.P., 2019. Assessment and optimization of rolling circle amplification protocols for the detection and characterization of badnaviruses. *Virology* 529, 73–80.
- Teycheney, P.-Y., Geering, A.D.W., Dasgupta, I., *et al.*, 2020. ICTV virus taxonomy profile: *Caulimoviridae*. *Journal of General Virology* 101, 1025–1026.
- Vo, J.N., Campbell, P.R., Mahfuz, N.N., *et al.*, 2016. Characterization of the banana streak virus capsid protein and mapping of the immune-dominant continuous B-cell epitopes to the surface-exposed N terminus. *Journal of General Virology* 97, 3446–3457.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/reverse-transcribing-dna-and-rna-viruses/w/caulimoviridae
Caulimoviridae.
Reverse Transcribing DNA and RNA Viruses.

Banana Bunchy Top Virus (*Nanoviridae*)

John E Thomas, The University of Queensland, Brisbane, QLD, Australia

© 2021 Elsevier Ltd. All rights reserved.

This is an update J.E. Thomas, Banana Bunchy Top Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00636-1](https://doi.org/10.1016/B978-012374410-4.00636-1).

Glossary

Cell-cycle link protein A plant virus protein which most probably subverts the cell cycle control of the host, forcing cells into DNA synthesis or S phase favorable to viral replication.

Circulative transmission Mode of transmission whereby the virus is internally borne in the insect vector but does not

replicate there. It is characterized by a latent period before which transmission cannot occur, and by retention of infectivity by the vector for a period of at least weeks.

Nuclear shuttle protein A virus-encoded protein that transports the viral single-stranded DNA, as a complex with the viral movement protein, to and from its site of replication in the nucleus and into adjacent cells.

Introduction

Banana bunchy top disease is the most economically important virus disease of banana and plantain (*Musa* spp.) worldwide, due to its devastating effect on crop yield, and the importance of banana and plantain as both a staple food and a major export commodity in much of the developing world. The causal agent is Banana bunchy top virus (BBTV), which has a multipartite circular ssDNA genome, encapsidated in small isometric virions, and is transmitted by the banana aphid, *Pentalonia nigronervosa*. Edible bananas are derived from wild seeded progenitors including predominantly subspecies of *M. acuminata* and *M. balbisiana*, which have a center of origin in South and South-East Asian-Australasian region. It is likely that BBTV and the aphid vector also originated within this area.

Taxonomy, Phylogeny, and Evolution

BBTV is a member of the genus *Babuvirus*, in the family *Nanoviridae*. Recently, two new virus species have been identified which are clearly distinct members of the genus *Babuvirus* (Fig. 1). *Abaca bunchy top virus*, which also infects *Musa* in South-East Asia, and *Cardamom bushy dwarf virus* causing Foorkey disease of cardamom in India, both have homologous components for all six genome components of BBTV. Additional species in this family are classified in the genus *Nanovirus*, and include *Faba bean necrotic yellows virus*, *Milk vetch dwarf virus* *Subterranean clover stunt virus*, *Black medic leaf roll virus*, *Faba bean necrotic stunt virus*, *Faba bean yellow leaf virus*, *Pea necrotic yellow dwarf virus* and *Pea yellow stunt virus*. Members of this family are phylogenetically related and share a similar particle morphology and multi-component circular, ssDNA genome. However, members of the two genera are not serologically related. Although both genera possess five components vital to replication and spread in common, some additional components present in one genus do not have a homologous component in the other genus. However, some of these components encode proteins on unknown function, and these unrelated proteins may share similar functions.

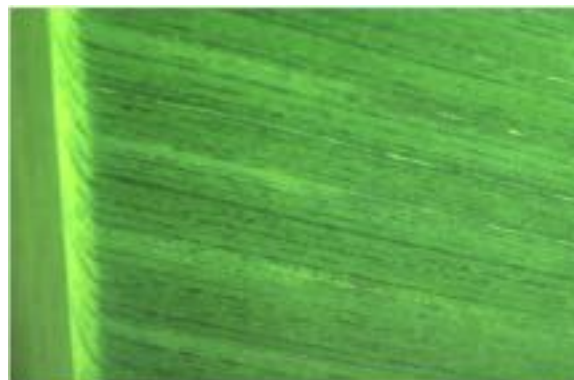


Fig. 1 Phylogenetic tree based on the amino acid sequences of babuvirus Rep proteins, using the nanovirus *Faba bean necrotic yellows virus* as an outgroup. The South-East Asian and Pacific/Indian Ocean subgroups of BBTV, with representative isolates, are circled in blue and red, respectively.



Fig. 2 Electron micrograph of BBTV virions, negatively contrasted with 1% ammonium molybdate.

BBTV isolates worldwide fall into two broad phylogenetic groups (Fig. 1), called the South Pacific (or Pacific/Indian Ocean) group (isolates from Australia, the Pacific islands the Indian sub-continent, Myanmar, Egypt and Sub-Saharan Africa) and the Asian (or South-East Asian) group (China, Philippines, Japan, Vietnam, Indonesia, Laos, Thailand and Taiwan). Sequence differences between the two groups are shared across all genome components. Within group sequence diversity is much greater for the Asian group suggesting a longer evolutionary period and possible evolutionary origin for the virus.

The coat protein gene of BBTV is highly conserved, with a maximum difference of 3% at the amino acid level between isolates. No serological differences have been detected between any isolates using polyclonal or monoclonal antibodies.

Recombination and reassortment has been detected between the two groups, supporting the temporal co-existence of the groups at one time. The most recent common ancestor of these BBTV sequences has been dated as occurring 1086 years ago (95% HPD 812–1399 years). The presence of the virus in many countries from the South Pacific group can be traced to introductions within the last century.

Virion Structure

The viral nature of bunchy top disease was established by CJP Magee in Australia in the 1920s. However, it was not until 1990 that the virus particles were first isolated, in part due to lack of a suitable herbaceous experimental host, its low titer in infected plants and its restriction to the phloem in the fibrous vascular tissue of *Musa*. Typical of the family *Nanoviridae*, BBTV has icosahedral particles, 18–20 nm in diameter with $T = 1$ symmetry (Fig. 2), a buoyant density of 1.29–1.30 g/cm³ in cesium sulfate, and a sedimentation coefficient of 46S.

Genome Organization and Functions of Gene Products

BBTV has a multi-component genome comprising six transcriptionally active components (Fig. 3, Table 1), each ca. 1 kb in size. DNA-R encodes two proteins, one the master replication initiation protein (Rep), which can initiate and terminate replication of DNA-R and the other genome components, and a second internal ORF of unknown function. All other components are monocistronic. DNA-S encodes the coat protein, demonstrated by N-terminal sequencing and the production of specific antibodies against the cloned and expressed gene. It also contributes to RNA silencing. The DNA-C gene product contains an LXCXE motif and binds to a retinoblastoma protein, indicating that it is a cell cycle link protein which stimulates viral DNA replication. It has also been linked to suppression of RNA silencing. DNA-M has been shown to be both a movement protein, and the major pathogenicity determinant and suppressor of RNA silencing. DNA-N encodes the nuclear shuttle protein, while the function of the protein encoded by DNA-U3 is unknown. The untranslated regions of all six components share two areas of extensive homology, both concerned with the rolling circle method by which BBTV replicates. A stem-loop common region (CR-SL) of 69 nt includes a nonanucleotide sequence (TATTATTAC) which is shared between plant-infecting circular ssDNA nanovirids and geminivirids, and which is the site of the origin of viral replication. It also contains three short repeated sequences (iterons F1, F2 and R1) which are the presumed binding site for the Rep. The major common region (CR-M) varies between 66 and 92 nt in length amongst the various components. Virions contain a heterogeneous population of DNA primers, around 80 nt in length, which bind to the CR-M and prime complementary strand synthesis. The relative concentrations of the genome components can vary enormously with DNA-S generally being the least abundant in plants, and DNA-N the highest, at 60 to >1660 times the copy number of DNA-S.

The intergenic regions of all six integral components of BBTV have been shown to have promoter activity. The highest activity in banana embryonic cells was shown by promoters from DNA-C and DNA-M, components thought to be intrinsic to the infection process. Studies on the DNA-N promoter demonstrated that expression in banana embryonic cells is limited to phloem tissue,

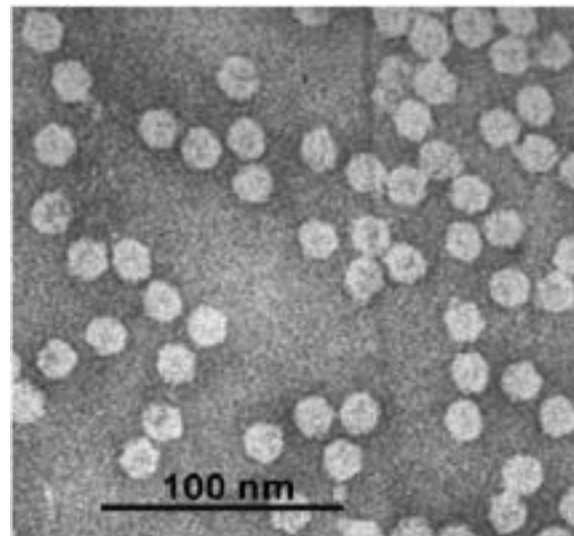


Fig. 3 Diagram of the integral genome components of BBTV, showing transcribed ORFs (black arrows) and other main genome features. From Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A., 2005. *Virus Taxonomy*. Eighth Report of the International Committee on Taxonomy of Viruses. San Diego, CA: Elsevier Academic Press.

Table 1 Size and function of the six integral components of BBTV

Genome component	Size (nt)	Size of encoded protein(s) (kDa)	Function
DNA-R	1111	(i) 33.6 (ii) 5.2	(i) Replication initiation protein (ii) Unknown
DNA-S	1075	20.1	Capsid protein
DNA-M	1043	13.7	Movement protein
DNA-C	1018	19.0	Cell-cycle link
DNA-N	1089	17.4	Nuclear shuttle protein
DNA-U3	1060	10.4	unknown

consistent with the circulative mode of transmission by the aphid vector and observations that histological effects were confined to the phloem and phloem parenchyma cells. Circulative transmission by aphid vectors usually involves specific feeding on the phloem tissue for virus acquisition. The CR-M and CR-SL are not essential for promoter activity. All essential elements of the promoter are located 3' of the stem-loop and within 239 bp of the translation start codon and within this region are an ASF-1 motif (TGACG), a hexamer motif (ACGTCA), *rbcS*-I-box (GATAAG), G-boxcore and the TATA box, all associated with promoter activity in other genomes.

Some isolates of BBTV, especially those from the South-East Asian group (see Taxonomy and Phylogenetic Relationships), contain additional Rep proteins called alphasatellites, similar in size to DNA-R, but capable of self-replication only. These molecules have a CR-SL, though the nonanucleotide sequence TAGTATTAC forming part of the loop is not conserved with the six integral DNA components. Unlike DNA-R, the alphasatellites lack the internal ORF, their TATA boxes are 5' of the stem loop and they generally lack the CR-M. Interestingly, the amino acid sequences of BBTV alphasatellites are more closely related phylogenetically to the alphasatellites associated with viruses in the related genus *Nanovirus* than they are to DNA-R of BBTV. The biological function of the alphasatellites is uncertain.

Host Range and Transmission

The symptoms of bunchy top disease in banana are characteristic, especially in the Cavendish subgroup of cultivars, and easily distinguished from all other virus diseases of banana. Plants can become infected at any stage of growth, and the initial appearance of symptoms can depend on the manner of infection. Plants that are aphid-inoculated usually develop first symptoms in the second or later leaves to emerge after inoculation, and there display a few dark green streaks and dots on the lower part of the lamina and on the petiole. These symptoms become more general on subsequent leaves. These streaks form hooks as they enter the midrib and are best viewed from the underside of the leaf, with transmitted light (Fig. 4). However, these dark streaks can be rare or absent in some cultivars. Successive leaves become shorter and narrower, and have a brittle lamina with upturned, chlorotic, ragged margins. Leaves fail to emerge fully, giving the plant a bunched appearance (Fig. 5). Plants derived from infected planting material (suckers, bits) develop severe symptoms from the first leaf to emerge.

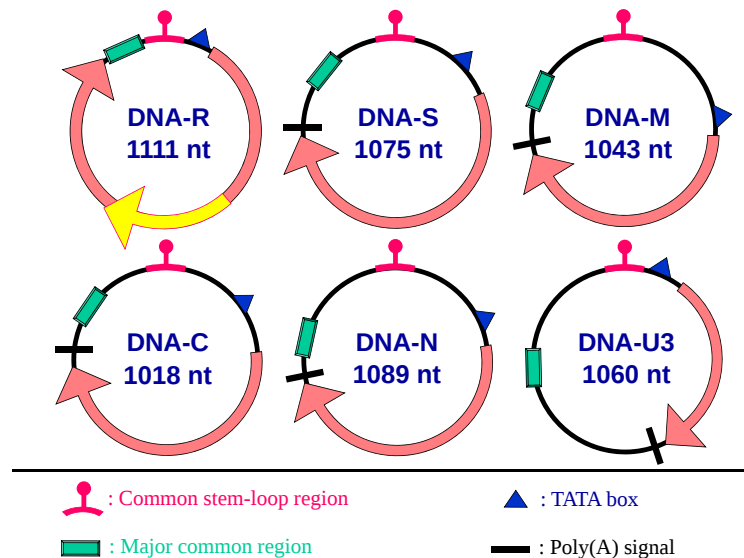


Fig. 4 Dark green dot-dash, hooking and vein clearing fleck symptoms on a Cavendish banana leaf.

Infected plants seldom produce a bunch, though if infected late in the current cropping cycle, a small, distorted bunch may result. With very late infections, the only symptoms to appear in the current season may be a few dark streaks on the tips of the flower bracts. No fruit is produced in subsequent years, and plants generally die within a couple of years.

From Taiwan, symptomless strains of BBTV, and mild strains that produce only limited vein clearing and dark green streaks have been reported. Interestingly, DNA-N could not be detected in the mild strain infection. Also, some field plants of the Cavendish subgroup cultivar Veimama from Fiji have been observed to initially show severe symptoms, then to recover and display few if any symptoms. This phenomenon has also been observed in the greenhouse in Cavendish cv Williams.

Confirmed hosts of BBTV are mostly confined to the family Musaceae. Known susceptible hosts include *Musa* species, cultivars in the Eumusa and Australimusa series of edible banana and *Ensete ventricosum*. Susceptible *Musa* species include *M. balbisiana*, *M. acuminata* ssp. *banksii*, *Musa acuminata* ssp. *halabanensis*, *M. acuminata* ssp. *longipetiolata*, *Musa acuminata* ssp. *malaccensis*, *M. acuminata* ssp. *zebrina*, *M. coccinea*, *M. jackeyi*, and *M. ornata*. *M. textilis*, *M. velutina*. There are some valid reports of hosts outside the Musaceae, in species in the Zingiberales, including *Alpinia purpurata*, *A. zerumbet*, *Canna indica*, *Colocasia esculenta* and *Hedychium coronarium* and from the Alismatales, *Colocasia esculenta*. However, using other virus isolates, not all tests to infect these hosts have been successful.

Although there are no confirmed reports of immunity to BBTV in *Musa* sp., there are many reports of differences in susceptibility and symptom expression between cultivars.

The banana aphid (*Pentalonia nigronervosa*) has a worldwide distribution in tropical and sub-tropical regions where banana is grown, and in 1925 was shown to be a vector of BBTV. Using morphological and molecular methods, *P. nigronervosa* f sp. *caladii* was resurrected as a distinct species, *P. caladii*. Both species are vectors of this virus. The host ranges of both aphid species show considerable overlap, though *P. nigronervosa* is found predominantly on *Musa* sp., while *P. caladii* occurs mostly on related hosts in the Zingiberales. On banana, they commonly colonize the base of the pseudostem at soil level and for several centimeters below the soil surface, beneath the outer leaf sheaths and newly emerging suckers.

Transmission by *P. nigronervosa* is of the persistent, circulative, non-propagative type, and the virus has been demonstrated to accumulate in the anterior midgut, hemolymph and principal salivary glands. Individuals from areas where BBTV is endemic and from where it is absent, both transmit the virus with equal efficiency. There is a minimum acquisition access period of 4 h, a minimum latent period of a few hours, and a minimum inoculation access period of 15 min. Aphids retain infectivity, after removal from a virus source, for at least 20 days and probably for life. Both nymphs and adults can acquire the virus, though the former a more efficient, and reported transmission rates for individual aphids are in the range 46%–67%. There is no evidence for transmission of BBTV to parthenogenetic offspring or for replication of the virus in the aphid.

BBTV is also efficiently transmitted in vegetative planting material, both conventional corms, corm pieces (bits) and suckers, and through micropropagation. All meristems from an infected corm will eventually become infected.

Virus–Host Relationships

BBTV is systemic within the banana plant, and following aphid inoculation, symptoms do not appear until at least two more new leaves have been produced (bananas produce single new leaves sequentially from a basal meristem). Histological examination suggests that BBTV is restricted to the phloem tissue, which shows hypertrophy and hyperplasia and a reduction in the

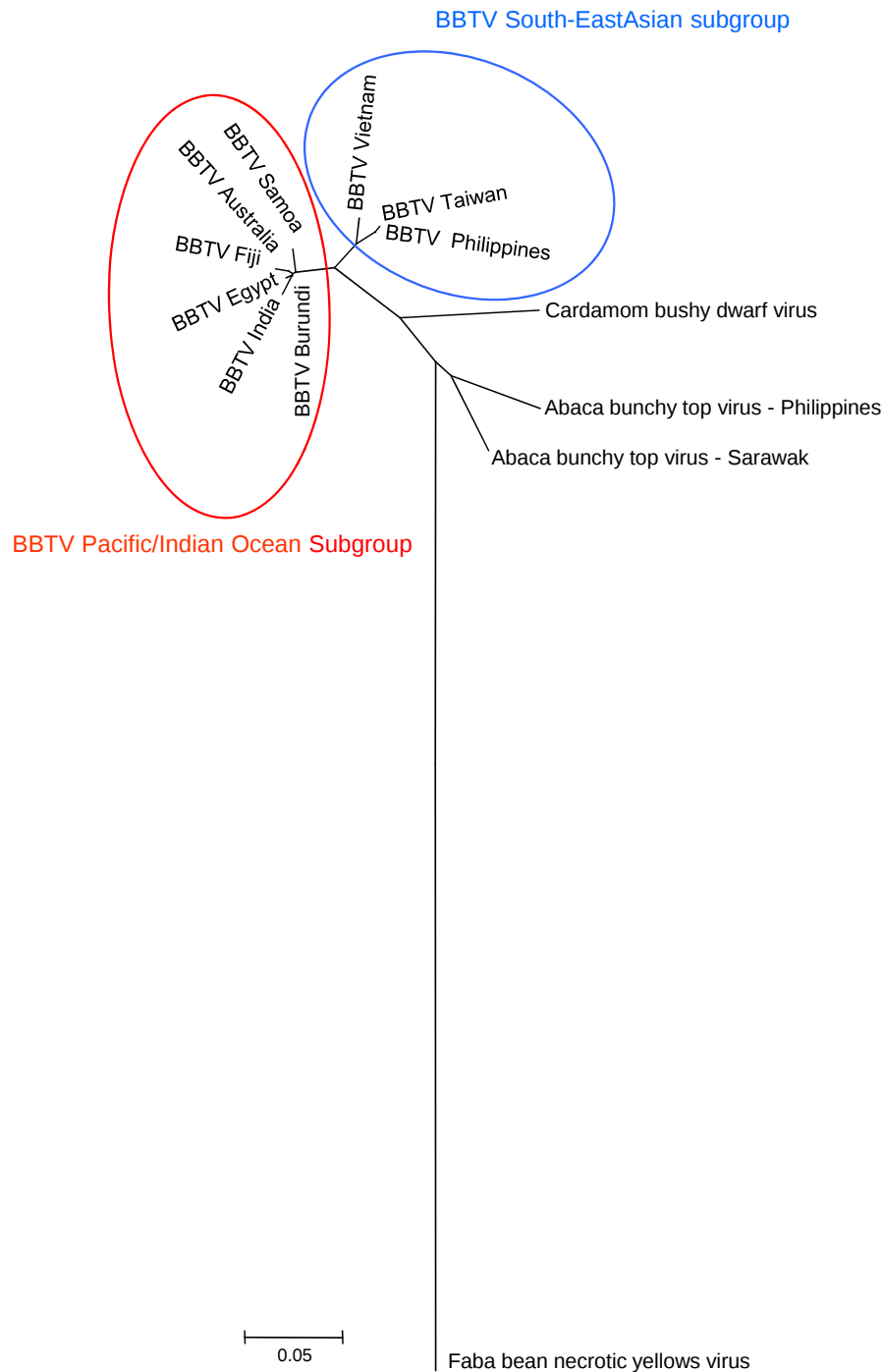


Fig. 5 Young BBTV-infected banana plants, showing stunting, and successively shorter narrow leaves with upturned, chlorotic margins. Leaves have failed to emerge fully, giving the plant a bunched appearance.

development of the fibrous schlerenchyma sheaths surrounding the vascular bundles. The cells surrounding the phloem contained abnormally large numbers of chloroplasts, resulting in the macroscopic dark green streak symptom.

It has been demonstrated through, the use of RNA probes and PCR, that BBTV replicates briefly at the site of aphid inoculation, then moves down the pseudostem to the basal meristem, subsequently infecting the newly formed leaves, the corm and the roots. The virus apparently does not replicate in leaves formed prior to infection, consistent with the lack of symptoms on these leaves and an inability to recover the virus from them via the aphid vector.

BBTV has been detected in most parts of the banana plant, including the leaf lamina and midrib, pseudostem, corm, meristems, roots, fruit stalk and fruit rind.

Diagnosis

The characteristic symptoms expressed by BBTV-infected plants allow symptomatology to be used to identify infected plants in field survey and control activities. However, BBTV is sometimes symptomless in tissue culture and the use of laboratory detection assays is desirable in the provision of clean planting material. Polyclonal and monoclonal antibodies to BBTV are used in ELISA to detect the virus in field and tissue culture plants and can detect the virus in single viruliferous aphids. Polymerase chain reaction (PCR) was shown to be about one thousand times more sensitive than ELISA or dot blots with DNA probes. Substances in banana sap inhibitory to PCR can be circumvented by simple extraction procedures or by immunocapture PCR. Nucleic acid-based loop mediated isothermal amplification (LAMP) and recombinase polymerase amplification (RPA) assays have also been described for the detection of BBTV.

Geographical Distribution

Historical Records and Possible Origins

The origin of BBTV is unclear. The first records were from Fiji in 1889, and though reports and photographs clearly indicate that it was present at least 10 years prior, soon after the establishment of an export industry based on Cavendish cultivars, evidence suggests that the disease did not originate there. Other early recorded are from Egypt (1901 – source unknown) and Australia and Sri Lanka both in 1913, and probably from planting material imported from Fiji.

Considerably more genetic diversity, and also a greater incidence of alphasatellites is found in BBTV isolates from South-East Asia. The aphid vector and the wild progenitors of modern edible bananas also originated in the South and South-East Asian-Australasian region, and the Cavendish cultivars of international trade and associated with early outbreaks of bunchy top disease, are thought to have originated in South-southern China/northern Vietnam. These factors lead to speculation that BBTV also originated and evolved in this region.

BBTV has a widespread, but scattered distribution in many of the banana-growing countries of the Asia-Pacific regions and Africa, but at present is not found in the Americas. In some countries the distribution is localized. For example, in Australia it is present only in southern Queensland and northern New South Wales, but not in the major production area of north Queensland. Banana bunchy top disease has been recorded in the following countries:

Africa: Angola, Benin, Burundi, Cameroon, Central African Republic, Congo Republic (Congo-Brazzaville), Democratic Republic of Congo (formerly Zaire), Egypt, Gabon, Malawi, Mozambique, Nigeria, Rwanda, South Africa, Zambia.

Asia: Bangladesh, Cambodia, China, Hong Kong, Indonesia, India, Iran, Japan (Ogasawara-gunto, formerly Bonin Island, Okinawa), Korea, Laos, Malaysia (Sarawak), Myanmar, Pakistan, Philippines, Sri Lanka, Taiwan, Thailand, Vietnam.

Oceania: Australia, Fiji, Guam, Kiribati (formerly Gilbert Islands), Marianas Islands (Guam, Saipan, Tinian and Rota), New Caledonia, Samoa, Tahiti, Tonga, Tuvalu (formerly Ellice Islands), USA (American Samoa, Hawaii), Wallis Island.

Epidemiology and Control of BBTV

Outbreaks of bunchy top disease can have a devastating effect on banana production, especially industries based on Cavendish cultivars. Production in Fiji fell by more than 80% from 1892 to 1895, primarily due to bunchy top disease. By 1925, around 10 years from the introduction of BBTV to Australia, the banana industry in northern New South Wales and southern Queensland had collapsed, with most plantations affected and production decreased by 90%–95%. Magee noted at the time “It would be difficult for anyone who has not visited these devastated areas to visualize the completeness of the destruction wrought in such a short time by a plant disease”.

More recently, a severe outbreak of banana bunchy top disease occurred in Pakistan. From 1991–1992, production area fell by 55% and total production by 90%, as a direct result of the disease. The disease has also recently appeared in Hawaii, New Caledonia, Benin, Nigeria, Mozambique and Malawi, all with severe impact in the affected areas.

The epidemiology of banana bunchy top disease is simplified by the occurrence of a single insect vector species and a limited host range for the virus, usually cultivated or feral edible bananas. Long distance spread is usually via infected planting material, and local spread via aphids and planting material.

Analysis of actual outbreaks of bunchy top disease in commercial banana plantations in Australia showed that the average distance of secondary spread by aphids was only 15.5–17.2 m, with nearly two-thirds of new infections less than 20 m from the nearest source of infection and 99% less than 86 m. Isolation of new plantations has a marked effect on reducing the risk of infection. New plantations situated adjacent to affected plantations had an 88% chance of recording infections in the first year. This was reduced to 27% if the plantations were separated by 50–1000 m, and to 5% if separated by more than 1000 m. The disease latent period (i.e., period from inoculation of a plant until an aphid can transmit the virus from this plant to another) is equivalent to the time taken for 3.7 new leaves to emerge from the plant. The actual time varies seasonally depending on the rate of leaf production.

Control strategies were devised by CJP Magee in the 1920s, and these measures still form the basis of the very successful control program in Australia today. The two major elements of the strategy are (1) exclusion of the disease from unaffected and lightly affected areas and (2) eradication of infected plants from both lightly and heavily-affected areas.

These measures require the participation of all growers and are unlikely to succeed if left to the goodwill of growers alone and are thus enforced by legislation in Australia. The measures include:

- (1) Registration of all banana plantations.
- (2) Establishment of quarantine zones.
- (3) Restrictions on the movement and use of planting material.
- (4) Regular inspections of all banana plantations for bunchy top.
- (5) Prompt destruction of all infected plants.
- (6) Ongoing education and extension programs for growers.

When adopted, these measures allowed the complete rehabilitation of the Australian banana industry. Sometimes total eradication of BBTV from a district has been achieved, but in most cases, incidence has been reduced to very low, manageable levels. Such successful control of bunchy top is rarely achieved in other countries, in most cases due to an inability to enforce an organized control program across whole districts.

There are no confirmed reports of immunity to BBTV in *Musa*. However, it has frequently been observed that there are differences in susceptibility between cultivars to both field and experimental infection. Edible bananas have diploid, triploid or tetraploid genomes containing, predominantly, elements of the *M. acuminata* (A), or *M. balbisiana* (B) genomes. Cultivars in the Cavendish subgroup (AAA genome), which dominates the international export trade, and many other A genome cultivars are highly susceptible and show severe disease symptoms. Interestingly, however, Gros Michel (AAA) displays resistance to the disease under both experimental inoculation and field conditions. Compared with highly sensitive cultivars such as Cavendish, the cultivar is less susceptible to aphid inoculation, contains a lower level of virions in infected plants, and symptoms are less severe and develop more slowly. These factors may contribute to a reduced rate of aphid transmission and field spread in plantations of Gros Michel, and introduction of this cultivar may explain the partial recovery of the Fijian banana industry after devastation of the Cavendish-based industry in the early 1900s. Field observations and glasshouse inoculations suggest that some B genome-containing cultivars are less susceptible to infection and/or display more limited symptoms, but this needs to be further investigated.

In glasshouse trials, the RNAi strategy for BBTV resistance based on DNA-R sequences has been shown to be effective in India and with DNA-M in Australia. In Malawi, initial field testing of transgenic bananas with RNAi-based BBTV resistance was less successful.

See also: Nanoviruses (*Nanoviridae*)

Further Reading

- Allen, R.N., 1987. Further studies on epidemiological factors influencing control of banana bunchy top disease, and evaluation of control measures by computer simulation. *Australian Journal of Agricultural Research* 38, 373–382.
- Amin, I., Ilyas, M., Qazi, J., *et al.*, 2011. Identification of a major pathogenicity determinant and suppressors of RNA silencing encoded by a South Pacific isolate of banana bunchy top virus originating from Pakistan. *Virus Genes* 42, 272–281.
- Burns, T.M., Harding, R.M., Dale, J.L., 1995. The genome organization of banana bunchy top virus: Analysis of six ssDNA components. *Journal of General Virology* 76, 1471–1482.
- Dugdale, B., Becker, D.K., Beetham, P.R., Harding, R.M., Dale, J.L., 2000. Promoters derived from banana bunchy top virus DNA-1 to –5 direct vascular-associated expression in transgenic banana (*Musa* spp.). *Plant Cell Reports* 19, 810–814.
- Kumar, P.L., Hanna, R., Alabi, O.J., *et al.*, 2011. Banana bunchy top virus in sub-Saharan Africa: Investigations on virus distribution and diversity. *Virus Research* 159, 171–182.
- Magee, C.J., 1953. Some aspects of the bunchy top disease of banana and other *Musa* spp. *Journal and Proceedings of the Royal Society of New South Wales* 87, 3–18.
- Magee, C.J.P., 1927. Investigation on the bunchy top disease of the banana. *Bulletin of the Council for Scientific and Industrial Research*. (30), 88.
- Niu, S., Wang, B., Guo, X., *et al.*, 2009. Identification of two RNA silencing suppressors from banana bunchy top virus. *Archives of Virology* 154, 1775–1783.
- Stainton, D., Martin, D., Muhire, B., *et al.*, 2015. The global distribution of banana bunchy top virus reveals little evidence for frequent recent, human-mediated long-distance dispersal events. *Virus Evolution* 1, vev009.
- Thomas, J.E., Dietzgen, R.G., 1991. Purification, characterization and serological detection of virus-like particles associated with banana bunchy top disease in Australia. *Journal of General Virology* 72, 217–224.
- Thomas, J., 2018. Diseases caused by viruses. In: Jones, D.R. (Ed.), *Handbook of Diseases of Banana, Abacá and Enset*. Wallingford: CABI, p. 632.
- Thomas, J.E., Smith, M.K., Kessling, A.F., Hamill, S.D., 1995. Inconsistent transmission of banana bunchy top virus in micropropagated bananas and its implication for germplasm screening. *Australian Journal of Agricultural Research* 46, 663–671.
- Vetten, H.J., Dale, J.L., Grigoras, I., *et al.*, 2012. *Nanoviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Amsterdam: Elsevier, Academic Press, pp. 395–404.
- Watanabe, S., Borthakur, D., Bressan, A., 2016. Localization of banana bunchy top virus and cellular compartments in gut and salivary gland tissues of the aphid vector *Pentalonia nigronervosa*. *Insect Science* 23, 591–602.
- Watanabe, S., Greenwell, A.M., Bressan, A., 2013. Localization, concentration, and transmission efficiency of Banana bunchy top virus in four asexual lineages of *Pentalonia* aphids. *Viruses* 5, 758–775.

Barley Yellow Dwarf Viruses (*Luteoviridae*)

Leslie L Domier, Agricultural Research Service, US Department of Agriculture, Urbana, IL, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update L.L. Domier, Barley Yellow Dwarf Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00637-3.

Glossary

Hemocoel The primary body cavity of most arthropods that contains most of the major organs and through which the hemolymph circulates.

Hemolymph A circulatory fluid in the body cavities (hemocoels) and tissues of arthropods that is analogous to blood and/or lymph of vertebrates.

Introduction

Barley yellow dwarf disease (BYD) is one of the most economically important virus diseases of cereals, and is found in almost every grain growing region in the world. Widespread BYD outbreaks in cereals were noted in the United States in 1907 and 1949. However, it was not until 1951 that a virus was proposed as the cause of the disease. The causal agents of BYD are obligately transmitted by aphids, which probably delayed the initial classification of BYD as a virus disease. Subsequently, BYD was shown to be caused by three species of Barley yellow dwarf virus (BYDV), a species of Cereal yellow dwarf virus (CYDV) and a species of Maize yellow dwarf virus (MYDV). Depending on the virulence of the virus strain, infection may contribute to winter kill in regions with harsh winters, induce plant stunting, inhibit root growth, reduce or prevent heading, or increase plant susceptibility to opportunistic pathogens and other stresses. Yield losses to wheat in the United States alone are estimated at 1%–3% annually, exceeding 30% in certain regions in epidemic years. The effects of BYD in barley and oats typically are more severe than in maize and wheat; sometimes resulting in complete crop losses. The existence of multiple strains and species of viruses that are transmitted in strain-specific manner made the viruses model systems to study interactions between viruses and aphid vectors in the circulative transmission of plant viruses. In addition, the compact genomes of the viruses have provided useful insights into the manipulation of host translation machinery by RNA viruses.

Taxonomy and Classification

The viruses that cause BYD are members of the family *Luteoviridae*, and were first grouped because of their common biological properties. These properties included persistent transmission by aphid vectors, the induction of yellowing symptoms in grasses, and serological relatedness. Different viruses are transmitted more efficiently by different species of aphids, a fact that was originally used to distinguish the viruses. Around 1960, the viruses were separated into five 'strains' (now recognized as distinct species) based on their primary aphid vectors. BYD-causing viruses transmitted most efficiently by *Sitobion* (formerly *Macrosiphum*) *avenae* were assigned the acronym MAV, for *Macrosiphum avenae* virus. Similarly, viruses transmitted most efficiently by *Rhopalosiphum maidis* and *Rhopalosiphum padi* were assigned the acronyms RMV and RPV, respectively. Viruses transmitted most efficiently by *Schizaphis graminum* were assigned the acronym SGV. Finally, vector-nonspecific viruses, that is, viruses transmitted efficiently by both *R. padi* and *S. avenae* were assigned the acronym PAV. The abbreviations now are assigned to BYDV-MAV, BYDV-PAV, BYDV-SGV, CYDV-RPV and MYDV-RMV. Subsequently, additional species were described, including a severe BYDV (BYDV-PAS) and a severe CYDV (CYDV-RPS). Prevalent BYD-causing viruses from China also have been named, including a *Luteovirus*, BYDV-GAV, which is a strain of BYDV-MAV transmitted by *S. graminum* and *S. avenae*, and a *Polerovirus* named wheat yellow dwarf virus-GPV (WYDV-GPV) transmitted by *R. padi* and *S. graminum*. Recently, two new BYDV species, BYDV-kerII and BYDV-kerIII, were identified in native grasses on the Keruelen islands in the Antarctic that are thought to have been introduced along with *R. padi* during the second half of the 20th century.

Based on genome organization and predicted amino acid sequence similarities, BYDV-MAV, BYDV-kerII, BYDV-kerIII, BYDV-PAS, and BYDV-PAV have been assigned to the genus *Luteovirus*, and CYDV-RPS, CYDV-RPV, MYDV-RMV and WYDV-GPV to the genus *Polerovirus*. The RNA-dependent RNA polymerases (RdRps) encoded by open reading frames (ORFs) 1 and 2 of luteoviruses are phylogenetically related to those of members of the family *Tombusviridae* (Fig. 1). In contrast, the predicted amino acid sequences of the RdRps encoded by ORFs 1 and 2 of poleroviruses are phylogenetically related to those of viruses in the family *Solemoviridae*. The two polymerase types are distantly related in evolutionary terms. For this reason, BYDV-SGV has not been definitively assigned to a genus because its genome has not been completely sequenced. The divergent origin of the RdRp sequences suggest that the genomic RNAs of luteoviruses and poleroviruses were formed by recombination between RNAs expressing a common set of structural and movement proteins and RNAs expressing two different sets of replication proteins. Because of these differences, it has been suggested that luteoviruses should be placed in the family *Tombusviridae* and poleroviruses in the family *Solemoviridae*.

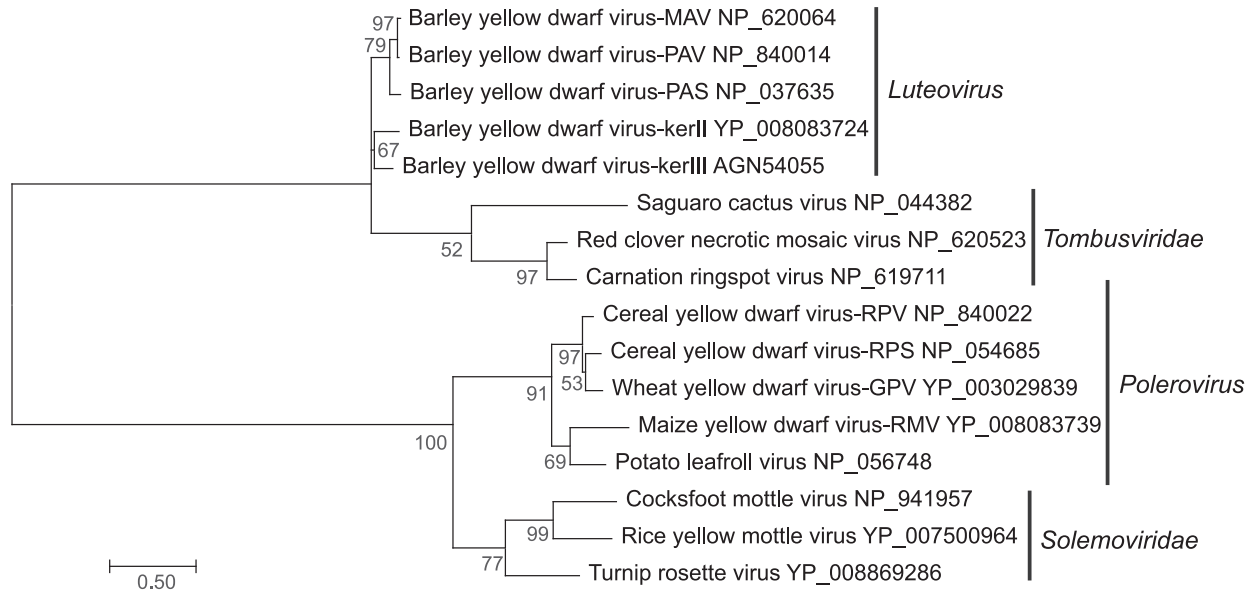


Fig. 1 Phylogenetic relationships of the predicted amino acid sequences of RNA-dependent RNA polymerases (RdRps) encoded by ORF2s of barley yellow dwarf (BYD)-causing viruses and selected members of the families *Solemoviridae* and *Tombusviridae*. The RdRps of BYD-causing luteoviruses are more similar to those of members of the family *Tombusviridae* than to those of BYD-causing poleroviruses. Similarly, the RdRp amino acid sequences of BYD-causing poleroviruses and Potato leaf roll virus (type member of the genus *Pterovirus*) are more similar to those of members of the family *Solemoviridae* than to those of BYD-causing luteoviruses. The resulting consensus tree from 500 bootstrap replications is shown. The numbers next to each node indicate the percentage of bootstrap replicates in which that node was recovered.

Virion Properties and Composition

All members of the family *Luteoviridae* have non-enveloped icosahedral particles with diameters of 25–28 nm (Fig. 2). Capsids are composed of major (22 kDa) and minor (65–72 kDa) coat proteins (CPs). The minor CP is formed by a carboxy-terminal extension to the major CP called the readthrough domain (RTD). According to X-ray diffraction and molecular mass analysis, virions consist of 180 protein subunits, arranged in $T = 3$ icosahedra. Virus particles do not contain lipids or carbohydrates, and have sedimentation coefficients $s_{20,w}$ (in Svedberg units) that range from 115 to 118 S. Buoyant densities in CsCl are approximately 1.4 g cm^{-3} . Virions are moderately stable, insensitive to freezing, and are insensitive to treatment with chloroform or nonionic detergents but are disrupted by prolonged treatment with high concentrations of salts. Table 1.

Genome Organization and Expression

Genomic RNAs of BYDV-causing viruses for which complete nucleotide sequences are available contain six to seven ORFs (Fig. 3). ORFs 1, 2, 3, 3a, 4 and 5 are shared among all the viruses. BYDVs lack ORF0. Genomic sequences of some BYDVs contain a small ORF, ORF6, downstream of ORF5. In poleroviruses, ORFs 0 and 1 and ORFs 1 and 2 overlap by more than 600 nt. In BYDVs, ORF1 overlaps ORF2 by less than 50 nt. In BYDV-causing virus genome sequences, ORF4 is contained within ORF3. An amber (UAG) termination codon separates ORFs 3 and 5.

The genomes of BYD-causing viruses have relatively short 5' ends and intergenic non-coding regions. ORFs 2 and 3 are separated by about 200 nt. The lengths of non-coding sequences downstream of ORF5 are very different between BYDVs and BYD-causing viruses. BYDV-PAV contains over 860 nt downstream of ORF5 compared to just 170 nt for CYDV-RPV.

The expression of BYDV-PAV RNA has been studied in detail and has revealed a complex set of RNA–RNA and RNA–protein interactions that are employed to express and replicate the virus genome. Less experimental data are available for BYDV-causing poleroviruses. However, expression and replication strategies and gene functions can be inferred from those of closely related poleroviruses, particularly Beet western yellows virus (BWYV) and Potato leaf roll virus (PLRV). ORFs 0, 1, and 2 are expressed directly from genomic RNAs. Downstream ORFs are expressed from subgenomic RNAs (sgRNAs) that are transcribed from internal initiation sites by virus-encoded RdRps from negative-strand RNAs and are 3'-coterminal with the genomic RNA. Since the initiation codon for ORF0 of poleroviruses is upstream of that of ORF1, translation of ORF1 is initiated by 'leaky scanning' in which ribosomes bypass the AUG initiation codon of ORF0 and continue to scan the genomic RNA until they reach the initiation codon of ORF1. The protein products of ORF2 are expressed only as a translational fusion with the product of ORF1. At a low frequency during the expression of ORF1, translation continues into ORF2 through a -1 frameshift that produces a large protein containing sequences encoded by both ORFs 1 and 2 in a single polypeptide. The frameshift is mediated by a

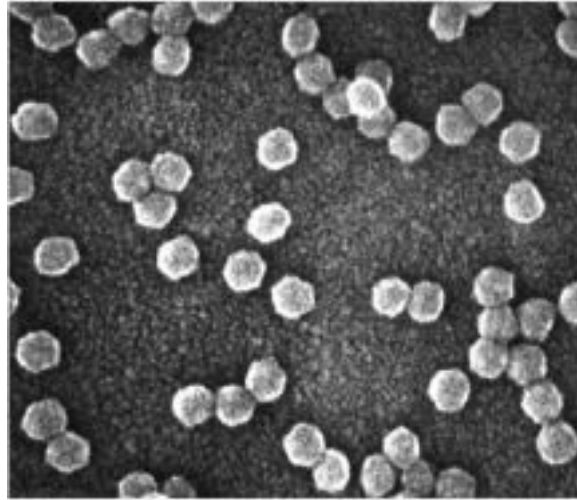


Fig. 2 Scanning electron micrograph of barley yellow dwarf virus-PAV particles, magnified $200,000\times$. Virions are c. 25 nm in diameter, hexagonal in appearance, and have no envelope and encapsidate a c. 5.6 kb molecule of single-stranded, positive-sense RNA. CYDV-RPV virions also encapsidate a 322-nucleotide satellite RNA that accumulates to high levels in the presence of the helper virus.

Table 1 Viruses causing barley yellow dwarf in cereals

Genus	Species (alternative name)	Abbreviation	Accession number ^a
<i>Luteovirus</i>	<i>Barley yellow dwarf virus KerII</i>	BYDV-KerII	NC_021481.1
	<i>Barley yellow dwarf virus KerIII</i>	BYDV-KerIII	KC559092.1
	<i>Barley yellow dwarf virus-MAV</i>	BYDV-MAV	NC_003680.1
	<i>Barley yellow dwarf virus-PAS</i>	BYDV-PAS	NC_002160.2
	<i>Barley yellow dwarf virus-PAV</i>	BYDV-PAV	NC_004750.1
	(<i>Barley yellow dwarf virus-RGV</i>) (rice giallume)		
<i>Polerovirus</i>	<i>Cereal yellow dwarf virus-RPS</i>	CYDV-RPS	NC_002198.2
	<i>Cereal yellow dwarf virus-RPV</i>	CYDV-RPV	NC_004751.1
	<i>Maize yellow dwarf virus-RMV</i>	MYDV-RMV	NC_021484.1
	<i>Wheat yellow dwarf virus-GPV</i>	WYDV-GPV	NC_012931.1
Unassigned	<i>Barley yellow dwarf virus-GAV</i>	BYDV-GAV	
	<i>Barley yellow dwarf virus-GPV</i>	BYDV-GPV	
	<i>Barley yellow dwarf virus-SGV</i>	BYDV-SGV	

^aAccession numbers beginning with NC_ represent reference genomic sequences.

“slippery heptanucleotide sequence” (in the form X XXY YYZ) and a downstream RNA secondary structure termed a pseudoknot that causes ribosomes to pause and then shift back one nucleotide before continuing translation in the new reading frame. In BYDV-PAV, frameshifting between ORFs 1 and 2 also is dependent upon the interaction of RNA sequences close to the site of frameshifting and a long-distance frameshift element (LDfE) located 4000 nt downstream in the 3′ non-coding region of genomic RNAs. Mutations that disrupt the interactions between these two distal regions suppress frameshifting and abolish RNA replication.

ORFs 3, 3a, and 4 are expressed from the 5′ terminus of sgRNA1 by a leaky scanning mechanism much like that used to express ORF1 of BYDV-causing poleroviruses. The 5′ terminus of sgRNA1 is located about 200 nt upstream of ORF3, and the 3′ terminus is co-terminal with the genome. ORF3a is initiated as a non-AUG codon. The product of ORF5 is expressed only as a translational fusion with the products of ORF3 by readthrough of the UAG termination codon at the end of ORF3. This produces a protein with the product of ORF3 at its amino terminus and the product of ORF5 at its carboxyl terminus. Readthrough is regulated by local and long-distance RNA-RNA interactions. BYDVs produce two additional sgRNAs. sgRNA2 contains ORF6, but sgRNA3 does not contain a predicted ORF.

While genomic RNAs of BYDV-causing poleroviruses contain 5′ VPgs that interact with translation initiation factors, BYDV-PAV RNA contains only a 5′ phosphate. Unmodified 5′ termini usually are recognized poorly for translation initiation. To circumvent this problem, a short sequence located in the noncoding region just downstream of ORF5 in the BYDV-PAV genome, called the BYDV translation enhancer (BTE), promotes efficient cap-independent translation initiation by directly interacting with eukaryotic translation initiation factor 4E and sequences near the 5′ termini of the genomic and sgRNA1.

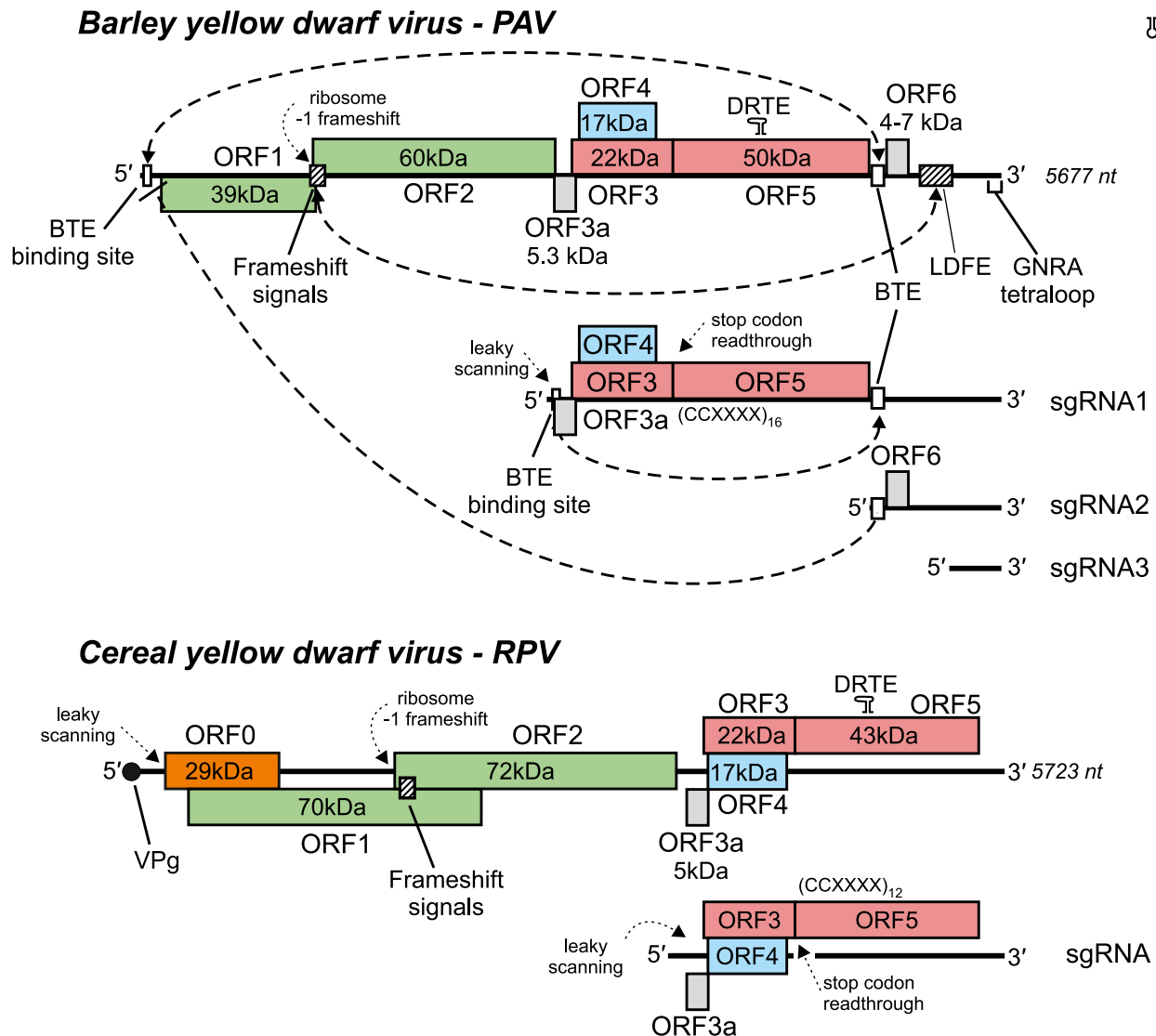


Fig. 3 Genome organizations of Barley yellow dwarf virus-PAV (BYDV-PAV) and Cereal yellow dwarf virus-RPV (CYDV-RPV). Individual open reading frames (ORFs) are shown as staggered open boxes. The predicted sizes of the protein products are indicated. The genome-linked protein (VPg) attached to the 5' terminus of CYDV RNA is indicated by a solid circle. Based on homology to other viruses ORF0 encodes a silencing suppressor and ORFs 1 and 2 encode replication-related proteins. ORFs 3 and 5 encode the major coat protein and readthrough domain, respectively. ORF3a, which is initiated at a non-AUG codon, encodes a protein required for long-distance movement. ORF4 encodes a protein required for virus cell-to-cell movement. The BYDV translation enhancer (BTE) facilitates translation initiation of BYDV-PAV genomic RNA and subgenomic RNA1 (sgRNA1). In both BYDV-PAV and CYDV-RPV, ORF2 is expressed as a translational fusion with the product of ORF1 via a -1 frameshift. In BYDV-PAV, frameshifting requires interaction between the 5' frameshift signals and the long-distance frameshift element (LDfE). Dashed lines indicate long-distance RNA–RNA interactions. GNRA tetraloops at the 3' terminus of BYDV-PAV genomic RNA are required for negative-strand RNA synthesis and virus replication.

Functions of BYDV and CYDV proteins have been ascribed based on homology to virus proteins with known functions and mutational characterization of protein coding regions. Similar to BWYV and PLRV, the proteins encoded by ORF0s of BYD-causing poleroviruses are inhibitors of local and systemic post-transcriptional gene silencing (PTGS). PTGS is an innate and highly adaptive antiviral defense found in all eukaryotes that is activated by double-stranded RNAs (dsRNAs), which are produced during virus replication. Furthermore, researchers noted a correlation between symptom severity and the strength of PTGS suppression of ORF0-encoded proteins of CYDV-RPV and CYDV-RPS. The genomes of BYDVs lack an ORF0, but the product of ORF4 functions as an inhibitor of systemic PTGS in those viruses. The ORF1-encoded proteins of BYD-causing poleroviruses contain the VPg and a chymotrypsin-like serine protease that is responsible for the proteolytic processing of ORF1-encoded polyproteins. The protease cleaves the ORF1 protein in trans to liberate the VPg, which is covalently attached to genomic RNAs. ORF2s, which are expressed as translational fusions with the product of ORF1, have coding capacities of 59–72 kDa and predicted amino acid sequences that are very similar to known RdRps and likely represent the catalytic portion of the viral replicase.

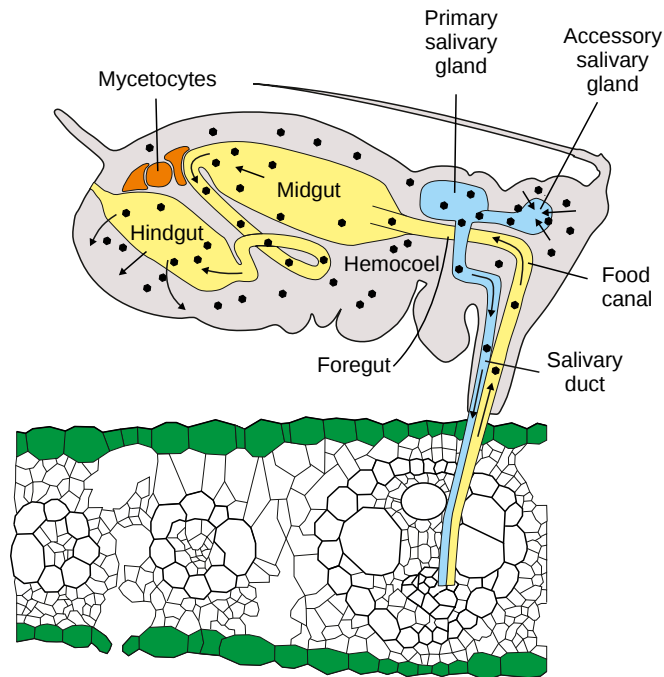


Fig. 4 Circulative transmission of BYD-causing viruses by vector aphids. While feeding from sieve tubes of an infected plant, an aphid (shown in cross section) acquires virions, which travel up the stylet, through the food canal, and into the foregut. Virions are actively transported across cells of the hindgut into the hemocoel and then passively diffuse through the hemolymph to the accessory salivary gland where they are again actively transported into the lumen of the gland. Once in the salivary gland lumen, the virions are expelled with the saliva into the vascular tissue of host plants. Viruses that are not transmitted by a particular species of aphid often accumulate in the hemocoel, but do not traverse the membranes of the accessory salivary gland.

ORF3a expresses highly conserved 4.8–5.3 kDa proteins that are essential for long-distance movement. ORF3 encodes the major 22 kDa CP. ORF5 has a coding capacity of 43–50 kDa, which is expressed only as a translational fusion with the product of ORF3 when translation reads through the termination codon at the end of ORF3 and continues through to the end of ORF5. The ORF5 portion of this readthrough protein has been implicated in aphid transmission and virus stability. Recombinant viruses that do not express ORF5 produce virions assembled from the major CP alone, which are not transmitted by aphid vectors and are less efficient in systemic infection of host plants than wild-type viruses. The amino-terminal portions of ORF5 proteins are highly conserved among BYD-causing viruses, while the carboxyl termini are much more variable. Proteins encoded by ORF4 are thought to facilitate intra- and intercellular virus movement because viruses with mutations in ORF4 are able to replicate in isolated plant protoplasts, but are deficient or delayed in systemic movement in whole plants.

Some BYDV genomic sequences contain an ORF6 downstream of ORF5. The predicted sizes of the proteins expressed by ORF6 range from 4 to 7 kDa. The predicted amino acid sequences of the proteins encoded by ORF6 are poorly conserved among BYDV-PAV isolates. Repeated attempts to detect protein products of ORF6 have been unsuccessful. In addition, BYDV-PAV genomes into which mutations have been introduced that disrupt ORF6 translation are still able to replicate in protoplasts. Based on these observations, it has been concluded that ORF6 is not translated *in vivo*. However, sgRNA2, which contains ORF6, has been shown to inhibit translation of capped and polyadenylated mRNAs, and hence, may function as a negative riboregulator of host translation.

Host Range and Transmission

BYD-causing viruses infect over 150 species of annual and perennial grasses in five of the six subfamilies of the Poaceae. The feeding habits of vector aphids have a major impact on the host ranges of virus species. Hence, the number of species naturally infected by the viruses is much lower than their experimental host ranges. As techniques for infecting plants with recombinant viruses have improved, the experimental host ranges of the viruses have been expanded to include plants on which aphid vectors would not normally feed. For example, BYDV-PAV and CYDV-RPV have been shown to infect *Nicotiana* species when inoculated using *Agrobacterium tumefaciens* harboring binary plasmids containing infectious copies of the viruses, which had not been described previously as experimental hosts for the viruses.

Viruses that cause BYD are transmitted in a circulative strain-specific manner by at least 25 aphid species. Circulative transmission of the viruses is initiated when the piercing-sucking mouthparts of aphids acquire viruses from sieve tubes of infected plants during feeding. Aphids that do not probe into and feed from the vascular tissues of infected plants do not transmit the

viruses. The virions travel up the stylet, through the food canal, and into the foregut (Fig. 4). After 12–16 h, virions then are actively transported across the cells of the hindgut into the hemocoel in a process that involves receptor-mediated endocytosis of the viruses and the formation of tubular vesicles that transport viruses through epithelial cells and into the hemocoel. Virions then passively diffuse through the hemolymph to the accessory salivary gland where virions must pass through the membranes of accessory salivary gland cells in a similar type of receptor-mediated transport process to reach the lumen of the gland. The accessory salivary gland produces a watery saliva, containing few or no enzymes, that is thought to prevent phloem proteins from clogging the food canal. Once in the salivary gland lumen, virions are expelled with the watery saliva into vascular tissues of host plants. Typically, hindgut membranes are much less selective than those of the accessory salivary glands. Consequently, viruses that are not transmitted by a particular species of aphid often are transported across gut membranes and accumulate in the hemocoel, but do not traverse the membranes of the accessory salivary gland. The specificity of aphid transmission and gut tropism has been linked to the RTD of the minor capsid protein. Even though large amounts of virions can accumulate in the hemocoel, there is no evidence for virus replication in their aphid vectors. Aphids may retain the ability to transmit virus for several weeks.

Genetic and biochemical studies have been conducted to identify aphid determinants of strain-specific transmission of BYDV-MAV and BYDV-PAV. Protein–protein and protein–virus interaction experiments were used to isolate two proteins from heads of vector aphids that bind BYDV-MAV that were not detected in non-vector aphids. These two proteins are good candidates for the cell-surface receptors that are thought to be involved in strain-specific transport of viruses into accessory salivary gland lumens. In addition, endosymbiotic bacteria that reproduce in specialized cells called mycetocytes in abdomens of aphids express chaperonin-like proteins that bind BYDV particles and the amino-terminal region of recombinant BYDV-PAV RTD proteins. However, the role of these proteins in aphid transmission is unclear since they are found in both vector and non-vector aphid species. Interactions of virus particles with these proteins seem to be essential for persistence of the viruses in aphids. The proteins may protect virus particles from degradation by aphid immune systems.

Replication

Like other viruses of the family *Luteoviridae*, BYD-causing viruses infect and replicate in sieve elements and companion cells of the phloem and occasionally are found in phloem parenchyma cells. The viruses induce characteristic ultrastructural changes in infected cells. BYDV-MAV, -PAV, and -SGV induce single-membrane-bound vesicles in the cytoplasm near plasmodesmata early in infection. Subsequently, filaments are observed in nuclei, and virus particles are first observed in the cytoplasm. In contrast, CYDV-RPV and MYDV-RMV induce double-membrane-bound vesicles in the cytoplasm that are continuous with the endoplasmic reticulum. Later, filaments and tubules form in the cytoplasm, and CYDV-RPV and MYDV-RMV particles are first observed in nuclei.

The subcellular location of viral RNA replication has not been determined unequivocally. However, early in infection, negative-strand RNAs of BYDV-PAV are first detected in nuclei and later in the cytoplasm, which suggests that at least a portion of the BYDV-PAV replication occurs in the nucleus. A nuclear location for replication is supported by the observation that the movement protein encoded by ORF4, which also binds single-stranded RNA, localizes to the nuclear envelope and is associated with virus RNA in nuclei of infected cells. Synthesis of negative-strand RNA, which requires tetraloop structures at the 3' end of BYDV-PAV genomic RNAs, is detected in infected cells before the formation of virus particles. Because tetraloops have been implicated in RNA–protein interactions, these structures could be binding and/or recognition sites for BYDV replication proteins. BYDV-PAV sgRNAs are synthesized by internal initiation of RNA synthesis on negative-strand RNAs from three dissimilar subgenomic promoters. Late in infection, the BTE near the 5' terminus of BYDV-PAV sgRNA2 inhibits translation from genomic RNA, which may promote a switch from translation to replication and packaging of genomic RNAs. In addition to genomic RNAs, CYDV-RPV replicates a satellite RNA by a rolling-circle mechanism that generates multimeric satellite RNAs that self-cleave to unit length.

Virus–Host Relationships

Visible symptoms induced by the viruses vary greatly depending on the genotype and developmental stage of the host and strain of the virus. The most common symptoms are stunting and chlorosis. While some infected plants display no obvious symptoms, most BYD-causing viruses induce characteristic symptoms that include stunting, leaves that become thickened, curled or serrated, and yellow, orange or red leaf discoloration, particularly of older leaves of infected plants. These symptoms result from phloem necrosis that spreads from inoculated sieve elements and causes symptoms by inhibiting translocation, slowing plant growth, and inducing the loss of chlorophyll. Symptoms may persist, may vary seasonally, or may disappear soon after infection. Temperature and light intensity often affect symptom severity and development. In addition, symptoms can vary greatly with different virus isolates or strains and with different host cultivars. Yield losses caused by BYD are difficult to estimate because the viruses are so pervasive, and symptoms often are overlooked or attributed to other agents. In Australia alone, losses in barley production have been valued at over 100 million US dollars annually. Plants infected with viruses causing BYD at early developmental stages suffer the most significant yield losses, which often are linearly correlated with the incidence of virus infection.

Epidemiology

BYD infections have been reported from temperate, subtropical, and tropical regions of the world. Even though the incidence of infections of individual viruses varies from year to year and can differ among annual and perennial hosts, BYDV-PAV usually is the most prevalent of the viruses causing BYD in small grains worldwide followed by CYDV-RPV or BYDV-MAV. The remaining BYD-causing viruses are typically much less prevalent. The viruses must be reintroduced into annual crops each year by their aphid vectors. Alate, that is, winged aphids may transmit viruses from local cultivated, volunteer, or weed hosts. Alternatively, alate aphids may be transported into crops from distant locations by wind currents. These vectors may bring the virus with them, or they may first have to acquire virus from locally infected hosts. In temperate regions of Europe and North America moderate and long-distance migration of viruliferous aphids is important to development of BYD epidemics. In Australasia, and other regions with Mediterranean climates, alate aphids usually transmit viruses from relatively close infected plants. Secondary spread of the viruses is often primarily by apterous, that is, wingless aphids. The relative importance of primary introduction of viruses by alate aphids and of secondary spread of viruses by apterous aphids in disease severity varies with the virus, aphid species, crop, and environmental conditions.

Studies have shown both vector and non-vector aphids species preferred to feed on noninfected plants after having acquired BYDV virus particles, while nonviruliferous aphids preferred to feed on BYDV-infected plants. This modification of aphid behavior associated with virus acquisition could enhance the spread of virus within plant populations.

In addition to crop plants, BYD-causing viruses native and invasive annual and perennial grasses. In some cases, prevalence of BYDV in native habitats was inversely correlated with distance from cultivated cereals. Grazing of herbivores on natural grass lands was also associated with increased incidence of BYDV infections. As observed on Keruelen islands in the Antarctic, the introduction of vector aphid species into new environments can increase incidence of virus infections that negatively impacts native plant communities.

Diagnosis

Accurate diagnosis of infections has been important in understanding the transmission and epidemiology of the viruses and developing control strategies for BYD. Because BYD symptoms resemble those caused by other biotic and abiotic factors, visual diagnosis is unreliable and other methods have been developed. Initially, infectivity, or biological, assays were used to diagnose infections. In bioassays, aphids are allowed to feed on infected plants and then are transferred to indicator plants. These techniques have also been used to determine vector specificities of viruses causing BYD and to identify viruliferous vector aphids in epidemiological studies. These techniques are very sensitive but can require several weeks for symptoms to develop on indicator plants. The viruses causing BYD are strongly immunogenic, which has facilitated development of genus- and even strain-specific antibodies that have been used extensively in BYD diagnosis. Because the viruses causing BYD are present in infected tissues at very low levels, mice have been used to produce monoclonal antibodies against the viruses. Mice typically require much less viral antigen per immunization than rabbits, and hybridoma cell lines that produce monoclonal antibodies can be stored for extended periods and used for many years, which further reduces the amount of antigen needed to produce diagnostic antibodies. Techniques have also been developed to detect viral RNAs from infected plant tissues by reverse transcription polymerase chain reaction, which can be more sensitive and discriminatory than serological diagnostic techniques. Even so, serological tests are the most commonly used techniques for the detection of infections because of their simplicity, speed, and relatively low cost.

Control

Planting of insecticide-treated seeds that protect emerging seedlings from aphid infestation has been shown to reduce losses caused by BYD in Africa, Australasia and North America. Foliar applications of insecticides on older plants typically have been less effective. Alternatively, planting of tolerant or resistant cereals has proved to be a much more cost-effective and sustainable management strategy for BYD. Breeding programs have successfully integrated genes conferring high levels of tolerance into barley and oat and to a lesser extent in wheat. Even though a limited number of single genes for BYD resistance/tolerance have been identified in cultivated barley and rice, in most instances, tolerance to BYD is conditioned by multiple genes in a quantitative fashion, which has made moving BYD tolerance into new plant lines challenging. Particularly in barley, molecular markers have begun to facilitate the process of breeding for BYD tolerance. Because of a lack of effective single-gene resistance in cultivated wheat, some researchers have moved BYD resistance genes from wheat grasses (*Thinopyrum intermedium* and *Thinopyrum ponticum*) into wheat, which have provided high levels of resistance. The lack of naturally occurring resistance in cereals to BYD has made transgene-mediated resistance very attractive. Transgenic barley and oat plants have been produced that express either intact or inverted-repeat regions of the BYDV-PAV genome that conferred high levels of resistance to BYDV-PAV and closely related viruses. However, the transgenic plant lines have not been released for use by growers possibly because of the expenses of obtaining regulatory approval and/or poor acceptance of genetically modified cereal crops by grower and consumer groups.

In many small grain growing regions, viruliferous aphids arrive at similar times each spring and fall even though sizes of the aphid populations can vary significantly from year to year. In these areas, it is sometimes possible to plant crops so that young,

highly susceptible plants are not in the field when the seasonal aphid migrations occur. However, crops planted later typically do not yield as well as those planted early in the growing season. Consequently, growers must weigh the probability of obtaining higher yields against possible yield losses caused by BYD. In some instances, biological control agents such as predatory insects and parasites have reduced aphid populations significantly.

See also: Luteoviruses (*Luteoviridae*). Solemoviruses (*Solemoviridae*). Tombusviruses (*Tombusviridae*)

Further Reading

- Crasta, O.R., Francki, M.G., Bucholtz, D.B., *et al.*, 2000. Identification and characterization of wheat-wheatgrass translocation lines and localization of Barley yellow dwarf virus resistance. *Genome* 43, 698–706.
- D'Arcy, C.J., Domier, L.L., 2000. Barley Yellow Dwarf. United States: American Phytopathological Society, Available at: <https://www.apsnet.org/edcenter/disandpath/viral/pdlessons/Pages/BarleyYelDwarf.aspx> (accessed 26.05.19).
- Falk, B.W., Tian, T., Yeh, H.H., 1999. Luteovirus-associated viruses and subviral RNAs. *Current Topics in Microbiology and Immunology* 239, 159–175.
- Fusaro, A.F., Barton, D.A., Nakasugi, K., *et al.*, 2017. The *Luteovirus* P4 movement protein is a suppressor of systemic RNA silencing. *Viruses* 9, 16.
- Gray, S., Gildow, F.E., 2003. Luteovirus-aphid interactions. *Annual Review of Phytopathology* 41, 539–566.
- Gray, S., Cilia, M., Ghanim, M., 2014. Circulative, "non-propagative" virus transmission: An orchestra of virus-, insect-, and plant-derived instruments. In: Maramorosch, K., Murphy, F.A. (Eds.), *Advances in Virus Research*, vol. 89. San Diego: Elsevier Academic Press Inc., pp. 141–199.
- Hershman, D.E., Johnson, D.W., 2011. Barley Yellow Dwarf – Plant Pathology Fact Sheet. PPFS-AG-SG-03 Cooperative Extension Service. United States: University of Kentucky, p. 5. Available at: <https://plantpathology.ca.uky.edu/files/ppfs-ag-sg-03.pdf> (accessed 26.05.19).
- Hulo, C., de Castro, E., Masson, P., *et al.* 2019. *Luteoviridae*. ViralZone. ExPASy bioinformatics resource portal. Switzerland. Available at: https://viralzone.expasy.org/609?outline=all_by_species (accessed 26.05.19).
- Kendig, A.E., Borer, E.T., Mitchell, C.E., Power, A.G., Seabloom, E.W., 2017. Characteristics and drivers of plant virus community spatial patterns in US west coast grasslands. *Oikos* 126, 1281–1290.
- Koev, G., Mohan, B.R., Dineshkumar, S.P., *et al.*, 1998. Extreme reduction of disease in oats transformed with the 5' half of the Barley yellow dwarf virus PAV genome. *Phytopathology* 88, 1013–1019.
- Miller, W.A., Jackson, J., Feng, Y., 2015. *Cis*- and *trans*-regulation of luteovirus gene expression by the 3' end of the viral genome. *Virus Research* 206, 37–45.
- Miller, W.A., Liu, S.J., Beckett, R., 2002. Barley yellow dwarf virus: *Luteoviridae* or *Tombusviridae*? *Molecular Plant Pathology* 3, 177–183.
- Miller, W.A., Rasochova, L., 1997. Barley yellow dwarf viruses. *Annual Review of Phytopathology* 35, 167–190.
- Miller, W.A., White, K.A., 2006. Long-distance RNA-RNA interactions in plant virus gene expression and replication. *Annual Review of Phytopathology* 44, 447–467.
- Ordon, F., Friedt, W., Scheurer, K., *et al.*, 2004. Molecular markers in breeding for virus resistance in barley. *Journal of Applied Genetics* 45, 145–159.
- Smirnova, E., Firth, A.E., Miller, W.A., *et al.*, 2015. Discovery of a small non-AUG-initiated ORF in Pteroviruses and Luteoviruses that is required for long-distance movement. *PLOS Pathogens* 11, e1004868.
- Xu, Y., Ju, H.J., DeBlasio, S., Carino, E.J., *et al.*, 2018. A stem-loop structure in potato leafroll virus open reading frame 5 (ORF5) is essential for readthrough translation of the coat protein ORF stop codon 700 bases upstream. *Journal of Virology* 92, e01544-17.
- Zhu, S., Kolb, F.L., Kaeppler, H.F., 2003. Molecular mapping of genomic regions underlying barley yellow dwarf tolerance in cultivated oat (*Avena sativa* L.). *Theoretical and Applied Genetics* 106, 1300–1306.

Bean Common Mosaic Virus and Bean Common Mosaic Necrosis Virus (*Potyviridae*)

Ramon Jordan, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

John Hammond, Floral and Nursery Plants Research, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

DAS-ELISA Double antibody sandwich ELISA

DIA Dot immunoassay

ELISA Enzyme-linked immunological assays

HC-Pro Helper component-proteinase

kb Kilobase

kDa Kilo dalton

McAbs Monoclonal antibodies

MP Movement protein

nm Nanometer(s)

nt Nucleotide(s)

ORF Open Reading Frame

RdRp RNA-dependent RNA polymerase

RT-PCR Reverse transcription-polymerase chain reaction

ssRNA Single-stranded RNA

UTR Un-translated region

Vpg Genome-linked protein

Glossary

Pathogroup A series of isolates of the same strain or pathotype (i.e., groups of isolates or strains that can be differentiated by their biological response on common bean cultivars carrying different combinations of resistance genes).

Pathotype An isolate or strain of a virus that is biologically distinct from other isolates of the same virus by virtue of differential host reactions.

Serogroup A series of isolates or strains of the same serotype.

Serotype An isolate or group of isolates that are distinguished from biologically related isolates by reaction (or lack of reaction) with key serological reagents such as defined polyclonal antisera or monoclonal antibodies.

Strain A genetic variant or subtype of a viral species.

History

Bean common mosaic virus and *Bean common mosaic necrosis virus* are species within the genus *Potyvirus*, family *Potyviridae*, and cause some of the most economically important diseases of legume crops worldwide. Yield losses due to Bean common mosaic virus (BCMV) and Bean common mosaic necrosis virus (BCMNV) may vary between 6 to 98% depending on the cultivar and time of infection. Both viruses occur essentially wherever bean and cowpea (including *Phaseolus*, *Vicia*, *Vigna*), lupin (*Lupinus*), pea (*Pisum*), peanut (*Arachis*), and soybean (*Glycine*) are grown. BCMV was first reported from the US in 1917, and the associated disease initially known as bean mosaic. It was renamed bean common mosaic in 1934 to differentiate it from bean yellow mosaic, caused by Bean yellow mosaic virus (BYMV), another potyvirus. Several pathotypes or strains of BCMV were distinguished in the 1970s by differential reactions of a number of bean cultivars, and in the early 1980s strains were further divided by serology into serotype A and serotype B. Serogroup A isolates were also biologically differentiated by temperature-insensitive induction of necrosis in bean cultivars carrying the dominant I gene. Peptide profiling of the coat protein (CP) and sequence analysis of the CP gene and 3'-untranslated region (UTR) demonstrated that the serotypes represented distinct viruses. Serotype A became known as bean common mosaic necrosis virus, while serotype B retains the name BCMV. 'Bean necrosis mosaic virus' was used for serogroup A isolates prior to acceptance of the species name bean common mosaic necrosis virus. BCMV and BCMNV also differ in particle length, CP size, and biological properties (see below). Complete genome sequences support the distinction between BCMV and BCMNV proposed using CP sequences and peptide profiles.

Taxonomy and Classification

The current criteria for differentiation of potyvirus species are <76% nt identity and <82% aa identity for the complete single ORF sequence and resulting polyprotein. The thresholds for species demarcation using nucleotide identity values for the individual coding regions range from 58% for the P1 coding region to 74%–78% for other regions. For the CP, the optimal species demarcation criterion is 76%–77% nt and 80% aa identity. Different species also often have distinct polyprotein cleavage sites. Using these criteria, *Bean common mosaic virus* and *Bean common mosaic necrosis virus* are distinct potyviral species; BCMV also now includes strains previously described by other names, including Azuki bean mosaic virus, Blackeye cowpea mosaic virus,

Dendrobium mosaic virus, Guar green sterile virus, Peanut chlorotic ring mottle virus, Peanut mild mottle virus, and Peanut stripe virus. At least 26 different strains of BCMV have been biologically authenticated as distinct on the basis of host range and host response, whereas only five biologically distinct strains of BCMNV have been recognized.

Potyviruses can be separated into at least 11 subgroups, including two 'supergroups', of which the *Bean common mosaic virus* supergroup is the largest, and the *Potato virus Y* (PVY) supergroup the second. Although 'supergroup' and 'subgroup' have no formal taxonomic meaning, it is a useful concept to describe a subset of virus species that cluster together in a phylogenetic tree (Fig. 1) and which may also have common serological properties.

BCMV and BCMNV, together with BSVA, CerMV, CABMV, CLLV, DsMV, EAPV, FVY, HarMV, IFBV, PMNV, PWV, SaLV, SaVY, SMV, TelMV, TrVY, WMV, WVMV, YBMV, ZaMMV, and ZYMV, form the BCMV supergroup (for full virus names, see Table 1). Some confusion exists in the literature over isolates described as Cowpea aphid-borne mosaic virus (CABMV); true CABMV isolates are distinct from BCMV, but some isolates initially described as CABMV were later shown to be synonymous with Blackeye cowpea mosaic virus (BlCMV), which is now recognized as a strain of BCMV.

Geographic Distribution

Many believe BCMV probably originated in Latin America, the center of diversity of *Phaseolus vulgaris*; others suggest South or East Asia. Regardless, BCMV has now spread worldwide and can be found wherever beans are grown, and in association with other bean types, as well as soybean, pea, and cowpea. Blackeye cowpea mosaic isolates are also found worldwide. Peanut stripe isolates are found in peanuts in Asia and the United States, and also infect lupin, soybean, and sesame. In many areas of the world BCMV is the most important potyvirus affecting beans, although in some areas BCMNV or BYMV may predominate. Before major efforts were made to eradicate BCMV from US germplasm collections, more than 60% of germplasm accessions were found to be infected.

BCMNV is thought to have evolved from BCMV in Central or Eastern Africa and is more important than BCMV in much of Central, Eastern, and Southern Africa, where BCMNV is common in a wide variety of wild and forage legumes as well as beans grown for human consumption. BCMNV has also been distributed throughout Africa, Europe, North and South America, largely through germplasm introductions, and has increased in importance in the United States partly as a consequence of deployment of resistance genes against BCMV (see below).

The increase in diversity and severity of bean-infecting geminiviruses in many regions has displaced BCMV and BCMNV as the most important viruses infecting beans in some areas.

Host Range and Transmission

BCMV naturally infects *P. vulgaris* (kidney bean), *P. acutifolius* (tepari bean), *P. atropurpureus*, *P. coccineus* (runner bean), *Glycine max* (soybean), *Macroptilium lathyroides* (horse gram), *Pisum sativum* (pea), *Rhynchosia minima*, *Vicia faba* (broad bean), *Vigna mungo* (black gram, urdbean), *V. radiata* (mung bean, green gram), *V. angularis* (azuki bean), and *V. unguiculata* (cowpea). Peanut stripe isolates naturally infect *Arachis hypogea* (peanut), *Lablab purpureus* (syn. *Dolichos lablab*; hyacinth bean), *Indigofera amoena*, *G. max*, *Lupinus albus* (lupin), *Pueraria phaseoloides*, *Sesamum* spp. (sesame), *Stylosanthes capitata*, and *S. craba*. Soybean isolates are adapted to soybean, and rarely infect peanuts. Dendrobium mosaic isolates infect *Dendrobium* orchids, while guar green sterile isolates infect guar (*Cyamopsis tetragonoloba*). In addition to its natural hosts, BCMV has a wide experimental host range, infecting about 100 species from 44 genera over nine families including *Amaranthaceae*, *Chenopodiaceae*, *Fabaceae*, *Solanaceae*, and *Tetragoniaceae*, although individual isolates may be much more restricted.

BCMNV naturally infects *P. vulgaris*, *P. lunatus*, *Centrosema pubescens*, *Crotalaria incana*, *Lablab purpureus*, *Senna bicapsularis*, *S. hirsuta*, *S. sophora*, and *V. vexillata*, and has been detected using a BCMNV-specific monoclonal antibody in naturally infected plants of *Albizia coriaria*, *Desmodium intortum*, *D. uncinatum*, *Rhynchosia resinosa*, *Tephrosia barbigera*, and *T. paniculata*. Experimental hosts include several other leguminous species, such as *Canavalia ensiformis*, *Crotalaria spinosa*, *Macroptilium lathyroides*, *Rhynchosia minima*, and *V. radiata*.

Both BCMV and BCMNV are transmitted by aphids in a non-persistent manner typical of viruses in the genus *Potyvirus*; the most important vector species in many countries are *Acyrtosiphon pisum*, *Aphis fabae*, and *Myzus persicae*. *Aphis craccivora* may be the most important vector in India. Minimum acquisition access and inoculation access periods are less than 1 min; no latent period is required between acquisition and inoculation. Aphid transmission is the cause of secondary transmission within the crop, but the primary means of introduction to the crop is through infected seed. Aphid transmission and infection rates tend to be higher when beans are grown under irrigation in dry regions.

Both BCMV and BCMNV are also seed-transmitted. Seed transmission varies with isolates and host species, as well as the timing of infection, but rates of up to 93% of infected seed may result from diseased plants, with erratic distribution of infected seed within individual pods. Plants infected after flowering typically do not yield infected seed, probably because the virus is not found in the embryo and cotyledons, and the virus may not be able to spread into these locations if flowering occurs prior to infection of the mother plant. Infection of the seed, and perhaps of the mother plant, can occur through infected pollen. Infectivity of the virus is retained during prolonged storage of seed. Dendrobium mosaic virus (a strain of BCMV) is transmitted by aphids and by vegetative propagation of *Dendrobium* orchids.

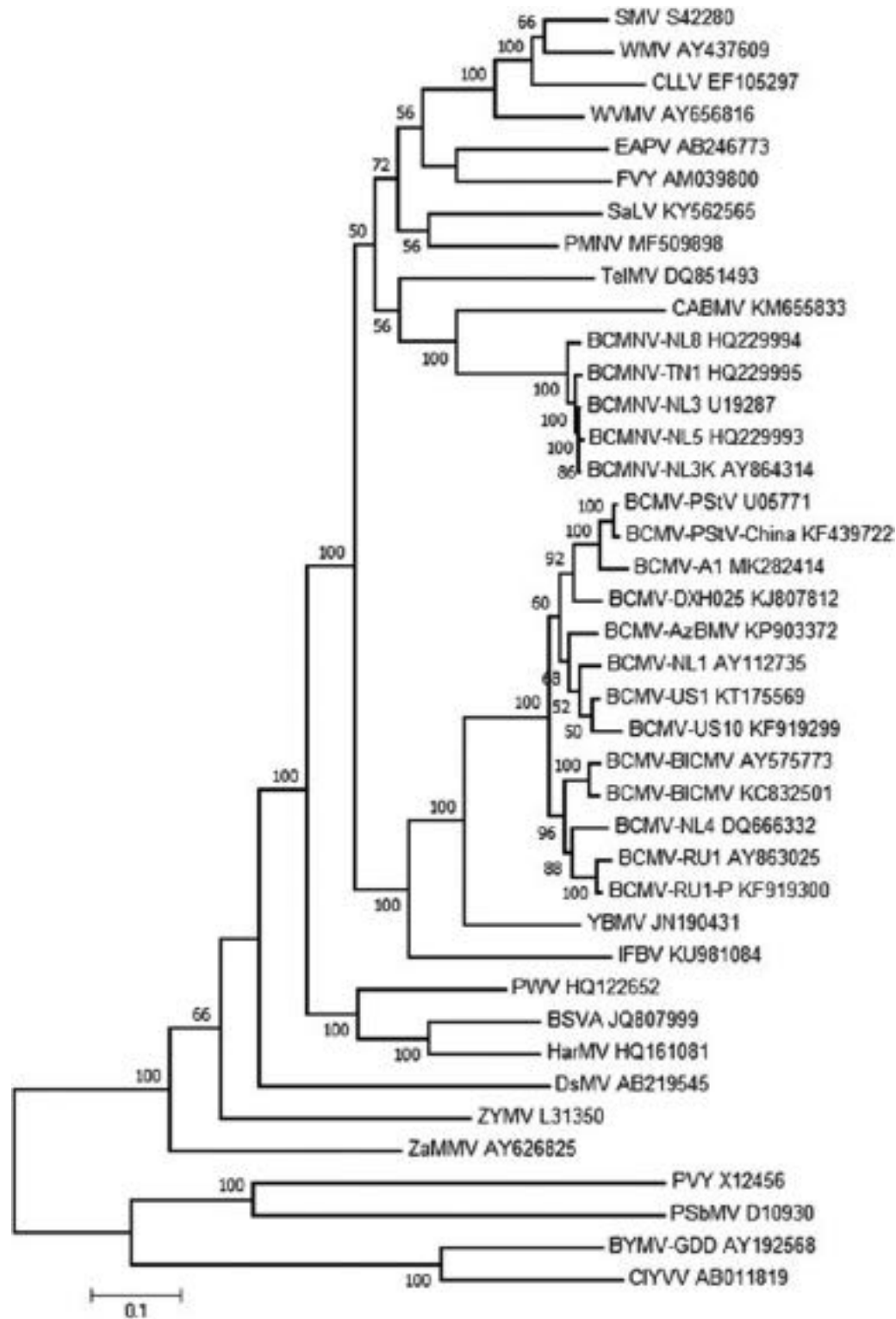


Fig. 1 Phylogenetic relationships of the Bean common mosaic virus subgroup. Phylogenetic relationship based on analysis of the polyprotein amino acid sequences of Bean common mosaic virus and Bean common mosaic necrosis virus with members of the BCMV subgroup, other legume-infecting potyviruses, and the genus *Potyvirus* type member, Potato virus Y (PVY). The tree was constructed by the neighbor-joining algorithm based on calculations from Clustal W pairwise amino acid sequence distances. The horizontal branch lengths are proportional to the genetic distance. The data set was subjected to 1000 bootstrap replicates. All nodes supported by >50% confidence values are shown. GenBank sequence sources are listed. Virus acronyms are as listed in Table 1; and, BYMV (Bean yellow mosaic virus), CIYVV (Clover yellow vein virus), and PSbMV (Pea seed-borne mosaic virus).

Table 1 The viruses included in the Bean common mosaic virus subgroup

<i>Virus species recognized as members of the bean common mosaic virus subgroup</i>	<i>Acronym</i>	<i>Representative GenBank Accession Number</i>
<i>Bean common mosaic virus</i>^a	<i>BCMV</i>	AY112735
Azuki bean mosaic virus ^b	(AzMV)	KP903372
Bean common mosaic virus serotype B	BCMV	KM023744
Blackeye cowpea mosaic virus	(BICMV)	KC832501
Dendrobium mosaic virus	(DeMV)	U23564
Guar green sterile virus	(GGSV)	AF045066
Peanut chlorotic ring mottle virus	(PCRMV)	Y11772
Peanut mild mottle virus	(PMMV)	Y11776
Peanut stripe virus	(PStV)	U05771
<i>Bean common mosaic necrosis virus</i>	<i>BCMNV</i>	HQ229994
Bean common mosaic virus serotype A	BCMNV	U19287
(Bean necrosis mosaic virus) ^c	(BNMV)	HQ229995
<i>Blue squill virus A</i>	<i>BSVA</i>	JQ807999
<i>Calla lily latent virus</i>	<i>CLLV</i>	EF105297
<i>Ceratobium mosaic virus</i>	<i>CerMV</i>	AF022442
<i>Cowpea aphid-borne mosaic virus</i>	<i>CABMV</i>	KM655833
Sesame mosaic virus	(SeMV)	U90326
South African passiflora virus	(SAPV)	S51666
<i>Dasheen mosaic virus</i>	<i>DsMV</i>	AB219545
(Vanilla mosaic virus) ^d	(VanMV)	KX505964
<i>East Asian passiflora virus</i>	<i>EAPV</i>	AB246773
<i>Fritillary virus Y</i>	<i>FVY</i>	AM039800
<i>Hardenbergia mosaic virus</i>	<i>HarMV</i>	HQ161081
<i>Impatiens flower break virus</i>	<i>IFBV</i>	KU981084
<i>Passionfruit woodiness virus</i>	<i>PWV</i>	HQ122652
<i>Paris mosaic necrosis virus</i>	<i>PMNV</i>	MF509898
<i>Saffron latent virus</i>	<i>SaLV</i>	KY562565
<i>Sarcochilus virus Y</i>	<i>SaVY</i>	AF185957
<i>Soybean mosaic virus</i>	<i>SMV</i>	S42280
<i>Telosma mosaic virus</i>	<i>TeIMV</i>	DQ851493
<i>Tricyrtis virus Y</i>^e	<i>TrVY</i>	AY864850
<i>Watermelon mosaic virus</i>	<i>WMV</i>	AY437609
Watermelon mosaic virus 2	(WMV-2)	D13913
Vanilla necrosis virus	(VNV)	L22907
<i>Wisteria vein mosaic virus</i>	<i>WVMV</i>	AY656816
<i>Yambean mosaic virus</i>	<i>YBMV</i>	JN190431
<i>Zantedeschia mild mosaic virus</i>	<i>ZaMMV</i>	AY626825
<i>Zucchini yellow mosaic virus</i>	<i>ZYMV</i>	L31350

^aNames in bold italic font are species currently recognized by the ICTV.

^bNames in regular font are no longer recognized as potyvirus species, but are indented to indicate their position as strains of the species listed above. Acronyms in parentheses are no longer valid as species, but refer to strains of the virus in bold font.

^c'Bean necrosis mosaic virus' was a name used for strains of the A serotype prior to ICTV designation of these isolates as strains of BCMNV.

^dVanilla mosaic virus was recognized in the ICTV Eighth Report as a tentative species in the genus Potyvirus; recent reports suggest that it is an isolate of DsMV.

^eTricyrtis virus Y appears to meet the criteria for distinct virus species, but has not yet been recognized as such by the ICTV; the acronym is indicated only in bold font.

Properties of the Virion and Genome

Properties of particles

BCMV and BCMNV are typical of viruses in the genus *Potyvirus* in that particle preparations contain a single sedimenting (154–158S) and buoyant density component (1.31–1.32 g cm⁻³ in CsCl). Virions are non-enveloped flexuous filaments, 12–15 nm wide, and 847–886 nm (BCMV) or 810–818 nm (BCMNV) long. Virions are composed of one single-stranded RNA (ssRNA) of c. 9600 nt for BCMNV and c. 10,000 nt for BCMV (accounting for ~5% of the particle weight) encapsulated in about 1700–2000 monomer units of one CP polypeptide species with a molecular mass of c. 30 kDa for BCMNV, and c. 33 kDa for BCMV (comprising ~95% of the particle weight); a molecule of the genome-linked protein NIa-VPg is bound the 5' end of the RNA. In gel electrophoresis the CP of BCMNV migrates with apparent Mr 33 kDa, and that of BCMV of Mr 34.5–35 kDa. Virus preparations that have undergone limited proteolysis also contain lower-molecular-weight peptides of apparent Mr 29–34 kDa.

Serological relationships

BCMV and BCMNV particles are strongly immunogenic and can be easily distinguished using polyclonal antisera and monoclonal antibodies (McAbs). Polyclonal antisera show only a very distant serological relationship between strains of these two viruses. BCMV- and

BCMNV-specific antisera are available commercially for use in standard serological tests such as enzyme-linked immunosorbent assay (ELISA). Early work with BCMV and homologous antisera showed that BCMV was closely related serologically to AzMV, BICMV, PStV, and several soybean infecting-isolates from Taiwan. As noted above, these viruses are now recognized as strains of BCMV. BCMV is distantly related to BYMV, Clover yellow vein virus (CIYVV), CABMV, PVY, Passion fruit woodiness virus (PWV), SMV, Tobacco etch virus (TEV), and WVMV.

Several broad-spectrum McAbs have been identified that detect all of the strains of BCMV, including AzMV, BICMV, DeMV, and PStV, as well as the BCMV subgroup member CABMV, but not with any strains of BCMNV. Other McAbs are reported which display various degrees of specificity for these viruses (e.g., can only detect CABMV or BCMV-PStV). While these McAbs can be used to discriminate among the BCMV isolates, biological diversity among these viruses may limit the use of many of these antibodies for diagnostic purposes. Most McAbs produced to BCMNV are specific only for strains of BCMNV (i.e., have no cross-reactivity with any strains of BCMV, nor other potyviruses).

Another series of McAbs were raised against three strains of BCMV (AzMV, BICMV, and PStV). Results of ELISA, immunosorbent electron microscopy, and western-blot analyzes indicated that these McAbs are specific for a series of epitopes located entirely on the virion surface. From western-blot analyzes of untreated and trypsin- as well as endopeptidase Lys-C-treated CPs, these virus-specific epitopes appear to be located exclusively on the amino-terminus of the CP. These McAbs were found to discriminate between strains of BCMV. These and other N-terminal targeted antibodies often enable clear distinction of strains in mixed infections, but a BCMV-PStV-specific McAb that reacts well with most BCMV-PStV blotch isolates failed to detect a necrotic isolate from Taiwan and several blotch isolates from Georgia, while a BCMNV-specific McAb misdiagnosed a naturally occurring genomic recombinant between BCMV and BCMNV.

Properties of the genome and replication

The ssRNA genome, which has a VPg protein covalently linked to its 5' end, is about 10,000 nt for BCMV and 9600 nt for BCMNV. The organization of the BCMV and BCMNV genomes is similar to other potyviruses, consisting of short untranslatable sequences at the 5' and 3' ends, and a single, long open reading frame (ORF). The ORF is translated into a single polypeptide: c. 3222 aa for BCMV and c. 3071 aa for BCMNV. The polypeptide undergoes co- and post-translational proteolytic processing by three viral-encoded proteinases to form ten individual gene products. Most of the polypeptide cleavage sites differ between BCMV and BCMNV, with the exception of the HC-Pro/P3 cleavage. The ten viral proteins (Fig. 2) include, in order from the 5' end of the genome, P1 proteinase; helper component-proteinase (HC-Pro); P3; a 6 kDa protein (6K1); cylindrical inclusion (CI); a second 6 kDa protein (6K2); the nuclear inclusion "a" (Nla)-VPg protein; Nla-proteinase; nuclear inclusion "b" (Nlb); and the CP. An additional protein is translated by a frameshift within the P3 gene, resulting in production of a fusion protein, P3N-PIPO. Genomic RNA replicates via production of a full-length negative-sense RNA.

Pathogenicity, Pathology, and Resistance Genes

Isolates of BCMV and BCMNV can be differentiated into at least nine pathotypes (0, and I through VIII) based on their reactions on a series of differential bean cultivars. In susceptible bean genotypes, BCMV and BCMNV induce highly similar symptoms including mosaic, dwarfing, chlorosis and leaf curling. The intensity and severity of the symptoms depend on various parameters including the virus strain, the bean cultivar, and the age and growing conditions (temperature) of the plant when infection occurs. The difference between the two viruses relies on the phenotypes generated on resistant cultivars. BCMV normally produces only mosaic symptoms in susceptible genotypes; BCMNV isolates can induce lethal systemic necrosis and plant death in bean cultivars carrying the incompletely dominant "I" gene. In most such cultivars the systemic necrosis develops at or above 20°C; in hypersensitive genotypes the response is restricted to necrotic local lesions at normal temperatures, and systemic or "black root" symptoms only at temperatures above 30°C. Hypersensitivity results in field resistance, and is genetically dominant; no seed transmission of BCMV occurs in genotypes with the dominant "I" gene. Other resistance genes (bc-1, bc-1², bc-2, bc-2², bc-3, and bc-u) condition resistance to particular pathotypes in various combinations; the combination of the dominant I gene with the strain-specific recessive genes, bc-1², bc-2², or bc-u protects against systemic infection and seed transmission of BCMNV. In the absence of the recessive strain-specific resistance genes effective against necrosis-inducing isolates of BCMNV, the I gene alone is not sufficient to protect against systemic infection by BCMNV in many regions of Africa. The viruses carry various combinations of six pathogenicity determinants (P0, P1, P1², P2, P2², and Px), which to date have not been correlated with particular viral genes. Pathotype is not correlated with either CP or 3' UTR phylogeny, nor is there evidence for determinants controlling systemic movement in the 3'-terminal region. Evidence from

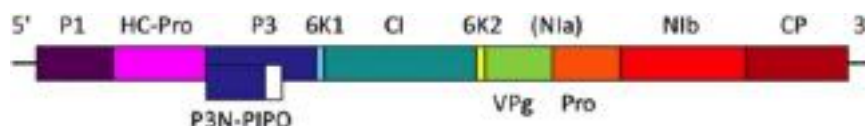


Fig. 2 Genome organization of BCMV and BCMNV. 5', 5' UTR; P1, P1 proteinase; HC-Pro, helper component-proteinase; P3, P3 protein; P3N-PIPO; 6K1, a 6 kDa protein; CI, cylindrical inclusion; 6K2, a second 6 kDa protein; Nla, nuclear inclusion "a", cleaved into VPg and Pro; VPg, genome-linked protein; Pro, proteinase; Nlb, nuclear inclusion "b"; CP, coat protein; 3', 3'-UTR. P3N-PIPO is expressed as a result of a frameshift within the P3 ORF.

natural recombinants highlights the importance of the potyviral and the P1-HC-Pro precursor protein(s) in pathogenesis and the breakage of genetic resistance (see below).

The incompletely dominant I gene was first identified in the cultivar “Corbett Refugee”, and the first recessive genes in “Robust” and “Great Northern 1” in the 1930s; strain-specificity of the recessive genes was identified in the 1950s, and the strain-unspecific gene bc-u shown in the 1970s to be necessary for expression of the strain-specific genes. Only one isolate has been identified so far that breaks resistance conferred by bc-3 (pathotype VIII); however, the gene originated in a single European breeding line, and it took many years to introgress bc-3 into genotypes adapted to the tropics. Cultivars carrying the dominant I gene may have increased susceptibility to Cowpea severe mosaic virus.

There is genetic linkage between certain seed colors and BCMV susceptibility, including red-mottled seed coat; the dominant I gene is closely linked to darker purple-mottled seed color. The lighter seed color associated with susceptibility is preferred in many cultures and commercial uses of the beans; this linkage can be partially avoided by selecting resistant types with commercially acceptable intermediate seed colors. Introgression of bc-3 can be problematic, as lines carrying the gene can be symptomless; serological or infectivity assays may be necessary to differentiate resistant and susceptible lines. Introgression of I and bc-3 has become simpler with development of molecular markers associated with each gene, allowing marker-assisted selection without virus challenge and phenotypic evaluation. The I gene is closely linked to a multigene family of TIR-NBS-LRR resistance genes conferring resistance to several related potyviruses and a comovirus.

In Latin America most black-seeded bean cultivars carry the I gene, and many additional bean lines bred at Centro Internacional de Agricultura Tropical and associated national programs also carry resistance. Many land races grown by traditional farmers, and particularly beans with light-colored seed coats, are susceptible; infection rates of up to 100%, and yield losses of 35%–98% are reported. In North America many bean varieties have effective resistance to BCMV, although many pinto, navy, and red-seeded types have only strain-specific resistance. However, recent breeding has yielded pinto and navy bean cultivars carrying both I and bc-3. BCMNV was previously rare in the US, but combined resistance to BCMV and BCMNV is now available in some cultivars of most bean types.

Many of the beans grown by small farmers in eastern and southern Africa are susceptible to both BCMV and BCMNV. The presence of necrosis-inducing strains of BCMNV, also found in wild legumes and species grown for fodder, means that the I gene cannot be deployed independent of recessive resistance genes, as infection by necrosis-inducing isolates could result in death of a significant proportion of plants in a field. In India infections of up to 100% are reported in bean, urdbean, mungbean, and cowpea, resulting in mottle, mosaic, blistering, stunting, and poor pod set. Many Asian bean types also lack resistance, due to linkage between susceptibility and desirable seed colors.

Peanut stripe strains of BCMV causing symptoms described as blotch, blotch-CP-N, blotch stripe, chlorotic line pattern, chlorotic ring mottle, mild mottle, necrosis, and stripe have also been differentiated on the basis of disease reactions on specific host genotypes. Losses of 23%–38% to peanut mild mottle have been documented in China. Peptide profiles of five symptom variants from Thailand that were not serologically distinct confirmed close relationships. No resistance was identified among over 8000 cultivated peanut genotypes in the germplasm collection at the International Crops Research Institute for the Semi-Arid Tropics. Some wild *Arachis* species have been identified as immune or highly resistant, including accession PI475998 of *A. cardenasii*, which could not be infected. Resistance genes have been identified and incorporated into resistant cultivars of *Pisum sativum*, *Lupinus angustifolius*, *Lablab purpureus*, and *Macroptilium lathyroides*.

Transgenic resistance has been demonstrated in peanut transformed with the CP gene of a peanut stripe isolate; a high level of resistance was observed in lines expressing either untranslatable or N-terminally truncated forms of the CP gene, suggesting that resistance resulted from RNA silencing (RNAi). While beans have been transformed for resistance to Bean golden mosaic virus (*Geminivirus*), and soybean for resistance to Soybean mosaic virus (*Potyvirus*), there are no reports of transgenic BCMV- or BCMNV-resistant bean lines.

Recombination and Variability

Although recombination between different isolates of BCMV had been reported for some years, no recombinants between BCMV and BCMNV were reported until recently. No recombinants were recovered from deliberate mixed infections of BCMV strain US-5 (pathogroup IV) and BCMNV strain NL-8 (pathogroup III) following many serial passages in either of two susceptible hosts. However, multiple recombinants were recovered within 28 days when the same strains were inoculated on opposite primary leaves of beans that were susceptible to one virus and resistant to the other; these recombinants fell into five different classes based on combinations of serotype (A, BCMNV; or B, BCMV) and pathogroup (IV, V, or VI). Subsequently, a naturally occurring strain of BCMNV (NL-3 K) with atypical responses on differential hosts was sequenced and shown to have a 5'-UTR and 5' region of the P1 gene almost identical to BCMV strain RU1, while the rest of the genome was almost identical to BCMNV strain NL-3 D. The P1 gene of the recombinant was larger than that of BCMNV, and similar to the BCMV P1. These results suggest that the P1 gene may play a significant role in pathogenicity and virulence, as previously suggested for BYMV.

The observed phenotypes of recombinants indicate that serotype, pathogenicity, and symptom expression are independent traits with separate determinants in the genome. The occurrence of recombination in bean genotypes with only recessive, strain-specific resistance genes suggests that this may be the primary cause of emergence of new resistance-breaking strains.

Much of the variability between isolates of either BCMV or BCMNV appears to be within the 5' portion of the genome, as has been shown both within and between other potyviruses. The P1 and P3 genes, and the N-terminal region of the CP gene are the most variable regions of the potyvirus genome, and may be involved in host–virus interactions. Differences in the 5' UTR and

N-terminus of P1 of both BCMV and BCMNV have been associated with symptom differences in common bean (*P. vulgaris*), and for BCMV in asparagus bean (*V. sesquipedalis*). Variability in the P1 gene is extensive, with as little as 61% identity at the aa level between BCMV strains, with most of the differences in the N-terminal domain. Differences in the CP gene result in clustering of BCMV isolates according to host species (bean, cowpea, peanut, asparagus bean, guar, *Dendrobium*, etc.) and to some extent by geographic origin within host clusters; peanut stripe isolates from China, Indonesia, and Thailand form separate clusters.

Diagnosis

Host range and symptomatology

Potyviruses can be detected and identified by a variety of techniques that are based on biological, cytological, biochemical, antigenic, and structural properties. With respect to BCMV and BCMNV, host range and symptomatology are not very useful because there are many viruses that infect bean, and significant differences in symptoms may be reproduced depending on the virus pathotype host genotype, and environmental conditions. Morphology of cytoplasmic inclusions has been useful in the identification of many potyviruses since inclusions can be virus-specific. However, inclusion morphology is not reliable for detection and differentiation of BCMV and BCMNV. BCMNV induces pinwheel and scroll types of cytoplasmic inclusions (subdivision I) in infected tissue. Most strains of BCMV, including BCMV-AzMV and BCMV-BICMV, also induce subdivision I type inclusions. However, PSTV and Taiwan soybean isolates induce subdivision IV cylindrical inclusions.

Serological techniques

BCMV- and BCMNV-specific polyclonal antisera and McAbs are useful for the detection and differentiation of these seed-borne viruses, and several methods have been used for detection in bean seed. When indirect ELISA and a dot immunoassay (DIA) with both McAbs and polyclonal antisera were evaluated, ELISA with McAbs proved the most sensitive method. However, BCMV antigen was found to be erratically distributed within the seed, with more than 50% of infected samples having detectable levels of BCMV antigen only in the seed coat. Tests on individual seeds proved unreliable and it is recommended that flour from a bulked sample of eight seeds be used when testing seed lots. Other assays include immunosorbent electron microscopy to detect BCMV in bean seed; or double antibody sandwich (DAS)-ELISA and bioassay on *Chenopodium quinoa* to test hydrated mature cowpea seed for presence of BCMV-BICMV. In these tests infectious virus was detected in cotyledons and sometimes in the embryonic axis of infected seed. Viral antigen was found in or on the testae, but very little infectious virus was present. Others have used direct antigen-coated-ELISA to detect BCMV-BICMV in cowpea seed.

Reverse transcription-polymerase chain reaction

The sensitivity, capacity, and potential of the PCR technique has been applied to BCMV and BCMNV. Several sets of BCMV- and BCMNV-specific RT-PCR primers have been designed and utilized to sensitively and accurately detect, differentiate, and characterize the two viral species. In one study using virus-specific CP primers, the specific geographical distribution of these viruses in Mexico was determined. In addition, the alignment of nine nucleotide sequences from cloned amplicons from the RT-PCR reactions for each viral species confirmed the identities of the viruses and was helpful in assigning them tentatively to pathogroups. Other primer sets have been developed and utilized, for example, in the successful detection and differentiation of the necrotic and blotch isolates of BCMV-PSTV, and in the differentiation of naturally occurring genomic BCMV/BCMNV recombinants. Considering the seed-borne nature of these viruses, RT-PCR has potential in quarantine programs for screening imported seed material and germplasm.

Prevention and Control

It is possible to eradicate BCMV and BCMNV from a particular production region, or at least to significantly reduce primary inoculum, by eliminating susceptible genotypes and planting only resistant varieties. There are few important non-crop reservoirs of inoculum, except in eastern and southern Africa, and infected seed are responsible for primary infections in the crop. Roguing of symptomatic plants is not recommended as a means of control, as it is highly likely that systemically infected plants without significant symptoms will remain to act as significant reservoirs for secondary infection by aphid transmission.

Exclusion of the virus is highly effective in areas where the virus is not present, but this may preclude introduction of new germplasm or cultivars without effective resistance. One or both viruses are almost universally present in seed lots of varieties that do not carry the I gene, and the majority of Latin American land race types do not possess any of the strain-specific recessive resistance genes. Most cultivars derived through extensive breeding programs carry the dominant I gene for resistance to BCMV and BCMNV. Increasingly bc-3 is being combined with I to yield cultivars with effective resistance to both BCMV and BCMNV.

Certified seed can be an effective means of control. A certified seed program in California with a limit of <0.5% BCMV has aided in the control of the disease, as resistant varieties are not yet available for all bean classes currently grown in California. Monitoring of BCMV for certification also aided in identification and containment of an outbreak of BCMNV introduced in seed of a navy bean cultivar.

Vector exclusion may aid in control, but as aphids rarely colonize common beans, migratory aphids are responsible for most transmission in many areas. As BCMV and BCMNV can be acquired and transmitted in less than 1 min access to plants, insecticide

treatment is unlikely to prevent, and may even encourage vector movement and transmission. Vector control through insecticide application may be effective in parts of Africa and Asia where aphids do colonize beans. Oil sprays are known to reduce transmission of stylet-borne viruses such as BCMV and BCMNV, but their lack of systemic action, and thus need for frequent reapplication to a growing crop, reduces the cost-effectiveness of their use.

See also: Recombination

Further Reading

- Adams, M.J., Antoniw, J.F., Fauquet, C.M., 2005. Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Archives of Virology* 150, 459–479.
- Berger, P.H., Wyatt, S.D., Shiel, P.J., *et al.*, 1997. Phylogenetic analysis of the *Potyviridae* with emphasis on legume-infecting potyviruses. *Archives of Virology* 142, 1979–1999.
- Chung, B.Y., Miller, W.A., Atkins, J.F., Firth, A.E., 2008. An overlapping essential gene in the *Potyviridae*. *Proceedings of the National Academy of Sciences of the United States of America* 105, 5897–5902.
- Dijkstra, J., Khan, J.A., 1992. A proposal for a Bean common mosaic subgroup of potyviruses. In: Barnett, O.W. (Ed.), *Potyvirus Taxonomy*. *Archives of Virology*, Suppl. 5 Springer, pp. 389–395.
- Drijfhout, E., 1978. Genetic interaction between *Phaseolus vulgaris* and Bean common mosaic virus with implications for strain identification and breeding for resistance. *Agricultural Research Reports No. 872*. Wageningen: Center for Agricultural Publishing and Documentation.
- Feng, X., Myers, J.R., Karasev, A.V., 2015. Bean common mosaic virus isolate exhibits a novel pathogenicity profile in common bean, overcoming the bc-3 resistance allele coding for the mutated eIF4E translation initiation factor. *Phytopathology* 105, 1487–1495.
- Gibbs, A.J., Ohshima, K., 2010. Potyviruses and the digital revolution. *Annual Review of Phytopathology* 48, 205–223.
- Higgins, C.M., Hall, R.M., Mitter, N., Cruickshank, A., Dietzgen, R.G., 2004. Peanut stripe potyvirus resistance in peanut (*Arachis hypogaea* L.) plants carrying viral coat protein gene sequences. *Transgenic Research* 13, 59–67.
- Larsen, R.C., Miklas, P.N., Druffel, K.L., Wyatt, S.D., 2005. NL-3 K strain is a stable and naturally occurring interspecific recombinant derived from *Bean common mosaic necrosis virus* and *Bean common mosaic virus*. *Phytopathology* 95, 1037–1042.
- McKern, N.M., Ward, C.W., Shukla, D.D., 1992. Strains of bean common mosaic virus consist of at least two distinct potyviruses. In: Barnett, O.W. (Ed.), *Potyvirus Taxonomy*. *Archives of Virology*. Springer, pp. 407–414. (Suppl. 5).
- Meziadia, C., Blanche, S., Geffroy, V., Pflieger, S., 2017. Genetic resistance against viruses in *Phaseolus vulgaris* L.: State of the art and future prospects. *Plant Science* 265, 39–50.
- Milne, R.G. (Ed.), 1988. *The Filamentous Plant Viruses*. *The Plant Viruses*, vol. 4. New York: Plenum.
- Mink, G.I., Silbernagel, M.J., 1992. Serological and biological relationships among viruses in the Bean common mosaic virus subgroup. In: Barnett, O.W. (Ed.), *Potyvirus Taxonomy*. *Archives of Virology*. Springer, pp. 397–406. (Suppl. 5).
- Morales, F.J., 1998. Present status of controlling bean common mosaic virus. In: Hadidi, A., Kheterpal, R.K., Koganezawa, H. (Eds.), *Plant Virus Disease Control*. St. Paul: APS Press, pp. 524–533.
- Providenti, R., Hampton, R.O., 1992. Sources of resistance to viruses in the *Potyviridae*. In: Barnett, O.W. (Ed.), *Potyvirus Taxonomy*. *Archives of Virology*. Springer, pp. 189–211. (Suppl. 5).
- Sengooba, T.N., Spence, N.J., Walkey, D.G.A., Allen, D.J., Femi Lana, A., 1997. The occurrence of bean common mosaic necrosis virus in wild and forage legumes in Uganda. *Plant Pathology* 46, 95–103.
- Shukla, D.D., Ward, C.W., Brunt, A.A., 1994. *The Potyviridae*. Wallingford: CAB International.
- Vallejos, C.E., Astua-Monge, G., Jones, V., *et al.*, 2006. Genetic and molecular characterization of the I locus of *Phaseolus vulgaris*. *Genetics* 172, 1229–1242.
- Vetten, H.J., Lesemann, D.-E., Maiss, E., 1992. Serotype A and B strains of bean common mosaic virus are two distinct potyviruses. In: Barnett, O.W. (Ed.), *Potyvirus Taxonomy*. *Archives of Virology*. Springer, pp. 415–431. (Suppl. 5).
- Worrall, E.A., Wamombe, F.O., Mukeshimana, G., *et al.*, 2015. Bean common mosaic virus and bean common mosaic necrosis virus: Relationships, biology, and prospects for control. *Advances in Virus Research* 93, 1–46.
- Wylie, S.J., Adams, M., Chalam, C., *et al.*, 2017. ICTV virus taxonomy profile: *Potyviridae*. *Journal of General Virology* 98 (3), 352–354. (accessed June 6, 2019). Available at: https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/potyviridae.

Bean Golden Mosaic Virus and Bean Golden Yellow Mosaic Virus (*Geminiviridae*)

Francisco M Zerbini, Federal University of Viçosa, Viçosa, Brazil

Simone G Ribeiro, Embrapa Genetic Resources and Biotechnology, Brasília, Brazil

© 2021 Elsevier Ltd. All rights reserved.

This is an update of F.J. Morales, Bean Golden Mosaic Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00695-6](https://doi.org/10.1016/B978-012374410-4.00695-6).

Glossary

HTS High throughput sequencing technologies, refer to methods that are able of sequencing multiple molecules in parallel, allowing hundreds of millions of molecules to be sequenced at a time.

Hybridization probe Labeled DNA or RNA molecule (radioactively or otherwise) that will hybridize with nucleic acids with complementary sequence and will detect their presence due to the label.

PCR Polymerase chain reaction is a technique used to make multiple (millions to billions) copies of a DNA segment *in vitro* using a thermostable DNA polymerase.

Resistance Relative ability of an organism to restrict the growth and attenuate the effects of a pathogen.

RFLP Restriction fragment length polymorphism is a method to examine the genetic variation of individuals by analyzing the pattern of the fragments generated by the digestion of a specific DNA segment with endonucleases.

RNA silencing A conserved evolutionary process active in a wide variety of eukaryotic organisms that can lead to inhibition of transcription or translation of a target gene in a sequence-specific way.

RNAi A process to control gene function by harnessing the expression of a dsRNA corresponding to the sequence of a target gene. The dsRNA triggers the RNA silencing mechanism resulting in the production of small RNAs that will induce the degradation of mRNA in a sequence-specific manner, hence blocking protein production.

Introduction

The common bean (*Phaseolus vulgaris* L.) is one of the most widely cultivated legume crops in the world. Its center of origin is located in the Americas, with two major domesticated gene pools known as Andean (from the highlands of Bolivia, Ecuador and Peru) and Mesoamerican (from Mexico and Central America). Common bean is a source of carbohydrates, protein and fiber and is a major component of the human diet in Latin America, Africa and Southeast Asia. As of 2016, the world's largest producers were Myanmar, India and Brazil. In the Americas the largest producers are Brazil, the United States and Mexico, but common bean is a major crop in every country in Central and South America.

Bean golden mosaic was first reported in the mid-1960s in Brazil by A.S. Costa as a whitefly (*Bemisia tabaci* Genn.)-transmitted disease. Similar diseases were described in several Central American and Caribbean countries during the late 1960s and early 1970s, and were named either "golden mosaic" or "golden yellow mosaic". In all countries from which it was reported, the disease was associated with an increase in the population of *B. tabaci*. The similar symptoms and association with the same vector led to the assumption that the same causal agent was responsible for the disease in all countries.

Plants affected by the disease usually display striking yellow mosaic symptoms (Fig. 1), named "golden" mosaic to differentiate from the less severe common mosaic (which is usually associated with mosaic and blistering symptoms without yellowing) and yellow mosaic diseases. In reality, it may be difficult to distinguish these three diseases based solely on symptoms in the field. Additional symptoms of bean golden mosaic may include downward leaf curling, leaf distortion and stunting. A significant level of flower abortion may occur in warmer climates, contributing to yield losses.

When first reported, bean golden mosaic was considered to be of "insignificant" economic importance. However, in less than ten years it became the main disease of common bean during the dry season (Jun–Sep) in south-central Brazil, mostly due to the exponential growth in the area planted with soybean, an excellent host of *B. tabaci*. Following the harvest of soybean (Feb–May), whitefly populations migrate massively to recently planted common bean fields, and the disease causes extensive losses. As soybean became a major crop in Brazil during the 1970s, bean golden mosaic reached alarming levels and the country's bean production was greatly affected. A similar scenario occurred in Argentina. Bean golden mosaic remains a major disease in these countries and the bean industry underwent several changes as a result, including relocation to regions which are less favorable to the disease and the large-scale adoption of central pivot irrigation (which eventually led to the emergence of soil-borne diseases such as white mold). The disease also continues to be of great relevance in other Latin American countries, where the increase in whitefly populations is associated with poor management practices leading to the emergence of insecticide-resistant populations and climate change.

Ultrathin sections of leaf tissue obtained from golden mosaic-affected bean plants in Brazil revealed the presence of isometric particles (20–25 nm) restricted to vascular tissues. Probably due to its phloem-restriction, the causal agent of bean golden mosaic



Fig. 1 Bright yellow mosaic (“golden” mosaic) symptoms in common bean (*Phaseolus vulgaris*) caused by Bean golden mosaic virus (BGMV). Photograph: R. Ramos-Sobrinho.

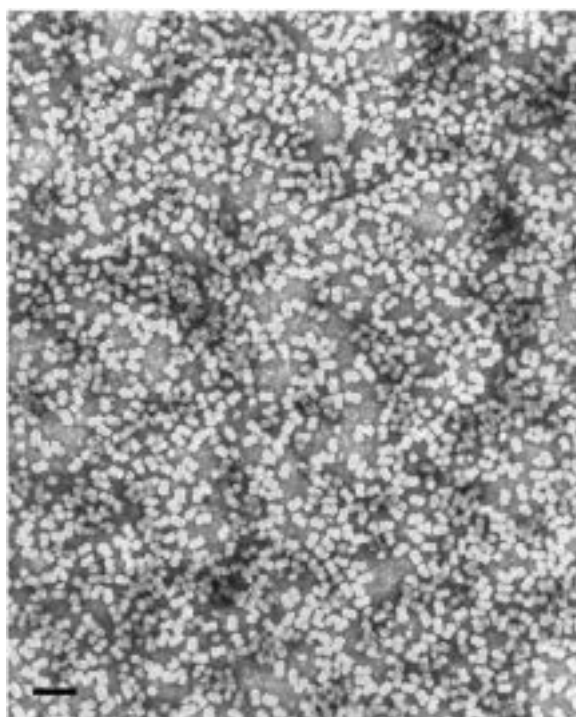


Fig. 2 Characteristic geminate particles of BGYMV. Bar = 20 nm.

in Brazil could not be transmitted by mechanical means or by seed. Conversely, the virus associated with the disease in Puerto Rico was readily sap-transmissible. This was an early indication that the similar diseases occurring in Brazil and in Central America/Caribbean were caused by distinct, albeit related, viruses. The isolation of whitefly-transmitted viruses that induced golden mosaic symptoms in common bean was achieved in 1976 by G. Galvez and M. Castaño at CIAT, Colombia, using a bean golden yellow mosaic virus (BGYMV) isolate from El Salvador, and in 1977 by R. Goodman at the University of Illinois, USA, using a BGYMV isolate from Puerto Rico. In both cases the isolated viruses had quasi-isometric particles with a unique geminate morphology (Fig. 2). The isolate from Puerto Rico was used to characterize the viral nucleic acid. Strikingly, the viral nucleic acid was identified as a single-stranded (ss) DNA molecule, the first reported ssDNA virus in plants. In 1981, the Puerto Rican BGYMV isolate was shown to possess a two-component ssDNA genome, with each molecule (ca. 2600 nt) separately encapsidated in a geminate particle.

In 1978, BGMV from Brazil and other BGMV/BGYMV-like viruses, as well as other similar ssDNA viruses transmitted by *B. tabaci* or leafhoppers (Homoptera: Cicadellidae) were included in the newly created Geminivirus group. BGMV was classified as a “subgroup II” geminivirus (whitefly-transmitted viruses with bipartite genomes), while the leafhopper-transmitted viruses with

monopartite genomes (such as Maize streak virus) were classified as “subgroup I” geminiviruses. In 1995 the leafhopper-transmitted geminiviruses were further divided into two subgroups, and BGMV became a member of “subgroup III” geminiviruses.

Studies performed by R. L. Gilbertson and D. Maxwell at the University of Wisconsin, USA in the early 1990s using BGMV isolates from Brazil and Central American/Caribbean countries provided definitive proof that two distinct geminiviruses were associated with bean golden mosaic disease in these regions. Broad-spectrum DNA probes were obtained which were capable of detecting all isolates tested (BGMV from Brazil, Guatemala and the Dominican Republic, plus an isolate of bean dwarf virus from Colombia) under low stringency hybridization conditions. A specific DNA probe was also prepared which detected only the isolate from Brazil, and another probe detected the isolates from Guatemala and the Dominican Republic (these two isolates could not be differentiated by nucleic acid hybridization even at high stringency hybridization conditions). These assays provided strong evidence that “BGMV” isolates from Brazil and Central America/Caribbean were actually distinct viruses. In the same year, a polymerase chain reaction (PCR)-based assay was used to obtain nucleotide sequences of a hypervariable region of the viral genome (the intergenic region between the MP gene and the common region of the DNA-B component). Nucleotide sequence identities between different BGMV isolates from the Dominican Republic (95% sequence identity amongst them), Guatemala (86%), BGYMV from Puerto Rico (75%) and BGMV from Brazil (46%) further suggested the existence of significant genetic variability, particularly between the Central American/Caribbean geminiviruses and the Brazilian BGMV isolate. In 1993, the complete nucleotide sequences of the Brazilian and Puerto Rican isolates were obtained. The percent nucleotide identities between these two viruses ranged between 71% and 82%, the latter value corresponding to the coat protein gene, which is the most conserved gene among whitefly-transmitted geminiviruses. These results clearly demonstrated that BGMV from Brazil was distinct from BGYMV from Puerto Rico. The former was classified as “type I” BGMV, and the latter as “type II” BGMV. In the same year, the isolates from Guatemala and the Dominican Republic were sequenced and shown to belong to type II BGMV, together with the isolate from Puerto Rico. In 2000 the Geminivirus group became the family *Geminiviridae*, including the genus *Begomovirus* (from bean golden mosaic). Type I BGMV isolates were classified as members of the species *Bean golden mosaic virus*, while type II BGMV isolates were classified as members of the species *Bean golden yellow mosaic virus*.

The bean golden mosaic disease observed in northwestern Mexico was shown in 1992 to be caused by another distinct begomovirus originally isolated in 1981 from squash in southwestern United States, named squash leaf curl virus (SLCV). Some SLCV isolates sampled from diseased common beans in Mexico had already changed significantly and were thus classified as members of a distinct species named *Bean calico mosaic virus*. However, SLCV can still be found in northwestern Mexico causing bean calico mosaic, a severe disease that often progresses from the characteristic golden mosaic into severe foliar bleaching and plant death. In Nicaragua, bean golden mosaic was shown to be caused predominantly by BGYMV, but also by squash yellow mild mottle virus and *Calopogonium* golden mosaic virus.

In 2010, a begomovirus detected in *Macroptilium lathyroides* and named *Macroptilium yellow spot virus* (MaYSV) was shown to infect common bean in the northeastern Brazilian state of Alagoas, inducing golden mosaic symptoms indistinguishable from those caused by BGMV and BGYMV. In a large-scale survey carried out in 2011–2012, MaYSV was the predominant begomovirus detected in common bean in that state, but was not detected in the midwestern states of Goiás and Minas Gerais. More recent surveys detected MaYSV in the states of Bahia and Pernambuco, also as the predominant begomovirus in common bean. The emergence of MaYSV adds an additional layer of complexity to the scenario of bean golden mosaic in Brazil.

Taxonomy, Phylogeny, and Evolution

Currently, viruses known to cause bean golden mosaic disease are classified in three species of the genus *Begomovirus* (family *Geminiviridae*): *Bean golden mosaic virus*, *Bean golden yellow mosaic virus*, and *Macroptilium yellow spot virus*. Isolates of *Bean golden mosaic virus* (BGMV) are found in Brazil and Argentina, while isolates of *Bean golden yellow mosaic virus* (BGYMV) are found throughout Central America and the Caribbean. It is remarkable that after more than 50 years since bean golden mosaic has been described, BGMV and BGYMV remain geographically isolated. This is probably due to the presence of large geographical barriers and the lack of seed transmissibility of both viruses. Isolates of *Macroptilium yellow spot virus* (MaYSV) have only been found in the northeastern Brazilian states of Alagoas, Bahia and Pernambuco.

Phylogenetically, begomoviruses are neatly divided into Old World and New World clades. Both BGMV and BGYMV are New World (NW) begomoviruses, but within the NW clade they are distantly related. Consistent with the geography-based (rather than host-based) grouping of begomoviruses, BGMV is phylogenetically close to other begomoviruses found in Brazil, such as tomato golden mosaic virus (TGMV) and *Sida micrantha* mosaic virus (SiMMV), while BGYMV is closer to other begomoviruses found in Central America and the Caribbean such as *Macroptilium yellow mosaic virus* (MaYMV).

In 1999, samples of common bean (*P. vulgaris*), lima bean (*P. lunatus*) and the weed *Leonurus sibiricus* were collected in bean-growing areas of Brazil, and viral fragments encompassing part of the Rep gene, the common region and part of the CP gene were amplified, cloned and sequenced. The *L. sibiricus* samples were infected by a distinct virus eventually named tomato yellow spot virus, which also infects tomato and common bean but is not associated with bean golden mosaic. All sequences obtained from common bean samples were closely related to the original BGMV isolate, with <7% nucleotide sequence divergence. This was low degree of genetic variability (unlike what is found for other begomoviruses) was attributed to the lack of BGMV-resistant bean cultivars that could place selection pressure upon the viral population. A second study in 2014 confirmed the low genetic variability of BGMV populations. This study compared BGMV and MaYSV populations obtained from three hosts (common bean,

lima bean and *M. lathyroides*), and reported a much higher degree of genetic variation in MaYSV isolates compared to BGMV. This was explained by numerous recombination events in MaYSV. Recombination was much less frequent in BGMV populations. Interestingly, the genetic variability of both BGMV and MaYSV was evenly distributed among isolates infecting the different hosts, indicating that genetic variability is an intrinsic viral property which is not affected by the host. Nevertheless, BGMV and MaYSV populations showed distinct distributions of genetic variation, with the BGMV population (but not the MaYSV population) being strongly structured by both host and geography, suggesting ongoing speciation. Purifying (negative) selection is the dominant force acting on BGMV populations. Positive selection is only rarely observed, and some positively selected sites in the coat protein gene could be associated with improving whitefly transmission.

No evolutionary studies have been carried out with BGYMV populations.

Virion Structure, Genome Organization, Properties and Functions of Gene Products, Replication and Propagation

Geminivirus particles are formed from two isometric, $T = 1$ capsids fused to form a geminate particle (18×30 nm), from which the family derives its name. The geminate capsid is formed by assembly of 110 subunits of a single CP, arranged as 22 pentamers (i.e., each $T = 1$ capsid is missing one pentamer). Virion structure has not been determined in detail for BGMV or BGYMV, but it is assumed to be similar to that determined for two other begomoviruses, African cassava mosaic virus (ACMV) and Ageratum yellow vein virus (AYVV). The CP would assume three distinct conformations, facilitating the formation of the interface between the two capsids. The solution structure of the AYVV capsid confirmed the long-standing assumption that each DNA component is encapsidated in a separate geminate particle.

Complete DNA sequences of several BGMV and BGYMV isolates were determined in the late 1980s/early 1990s, and genomic organizations are essentially identical for both viruses. Thus, only the one for the Brazilian isolate of BGMV will be described in detail. The DNA-A (RefSeq accession number NC_004042, derived from M88686) and DNA-B (NC_004043, derived from M88687) are 2617 and 2580 nt long, respectively. The two components share no significant sequence identity, except for an approx. 180-nt long region known as the common region (CR), which is highly conserved (>97% identity) between the two components. The CR sequence includes direct and inverted repeats that form a stem-loop structure, with the conserved 5'-TAATATTAC-3' nona-nucleotide in the loop. In addition, two direct repeats and one inverted repeat, known as iterons (high-affinity binding sites for the Rep protein), are present upstream from the stem-loop structure.

The DNA-A has five open reading frames (ORFs): one in the viral sense (CP; previously known as AV1 or AR1) and four in the complementary sense [Rep (AC1 or AL1), TrAP (AC2 or AL2), REn (AC3 or AL3) and AC4] (Fig. 3). The complementary sense ORFs are located in different reading frames, with the 3'-end of Rep overlapping the 5'-end of TrAP, and TrAP overlapping the 5' and central regions of REn. The Rep ORF overlaps the AC4 ORF in its entirety. CP and REn overlap in opposite directions in an AT-rich region and share four nucleotides at their C-termini, which contain their translational stop codons. The DNA-B contains two non-overlapping ORFs: one in the viral sense, the nuclear shuttle protein (NSP, also BV1 or BR1), and one in the complementary sense, the cell-to-cell movement protein (MP, also BC1 or BL1). All ORFs except REn and MP have TATA boxes within 100 nt of the start codon. Polyadenylation signals are located within the AT-rich regions at or near the 3'-ends of the CP, REn and NSP ORFs. The Rep, TrAP and REn ORFs are assumed to share a single polyadenylation signal located inside the 3'-end of REn, analogous to what was shown for TGMV.

The four complementary-sense genes in the DNA-A are involved in DNA replication (Rep and REn), suppression of host defenses (TrAP and AC4) and trans-activation of the viral-sense ORFs in both the DNA-A and DNA-B (TrAP). The CP, besides its structural role, is also necessary for whitefly transmission. Together, the two DNA-B-encoded proteins coordinate viral DNA movement from the nucleus to the cytoplasm (NSP) and cell-to-cell via plasmodesmata (MP).

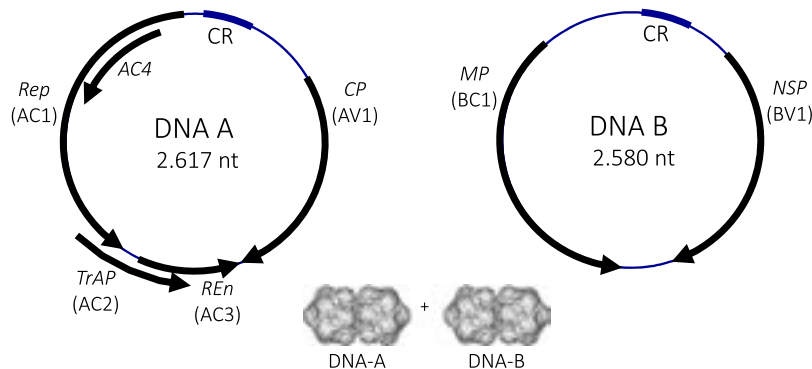


Fig. 3 Genome and virion structure of Bean golden mosaic virus (BGMV). The circles represent the two genomic components (DNA-A and DNA-B). The CP and NSP open reading frames are located in the viral-sense strand. Rep, TrAP, REn, AC4, and MP are located in the complementary-sense strand. CR, common region. Each genomic component is encapsidated in a separate geminate particle, as indicated below the genome structures. Thus, a virion is comprised of two geminate particles, one containing the DNA-A and the other containing the DNA-B.

Bipartite begomoviruses utilize both rolling-circle replication and recombination-dependent replication strategies. The common region contains the origin of rolling-circle replication, which is the conserved nona-nucleotide located in the stem-loop structure. The replication-associated protein (Rep) functions as a specific DNA binding protein that acts as both nuclease and ligase in the initiation and termination of the rolling-circle replication process.

In the infection process of begomoviruses, virus particles are inoculated into the plant by means of the insect vector, and the viral genome (ssDNA) disassociates spontaneously from the capsid. Viral DNA is then transported to the host cell nucleus, where the ssDNA is converted to a double-stranded dsDNA, referred to as the replicative form (RF). Indirect evidence, such as the need for local destabilization of dsDNA for initiation of rolling-circle replication in prokaryotes by strand-nicking enzymes, indicates that this is performed by host factors. The RF serves as a template for synthesis of new ssDNAs and viral mRNAs. The viral genome is replicated by the rolling-circle mechanism, similar to that used by bacteriophages phiX174 and M13, using the RF as a template. The origin of replication is the conserved nona-nucleotide (5'-TAATATTAC-3') located in the stem-loop structure in common region. Origin recognition is mediated by the high-affinity binding sites (iterons) located upstream of the stem-loop. After binding of Rep to viral DNA and stabilization of the complex formed by Rep, Ren and host factors, the Rep protein cleaves the nona-nucleotide (TAATATT//AC), initiating rolling-circle replication. Rep is the only viral protein which is essential for replication, with multiple roles. It functions as a DNA binding protein with sequence and structural requirements, and also as a nuclease and DNA ligase in the initiation and termination of the rolling-circle replication process (cleavage of the nona-nucleotide at the origin of replication; cleavage of the linear, multimeric DNA; ligation of the monomeric genome units).

As shown for other begomoviruses (Abutilon mosaic virus, African cassava mosaic virus, tomato golden mosaic virus and tomato yellow leaf curl virus), replication of BGMV and BGYMV may involve a recombination-dependent mechanism. In this mechanism, fragments of viral DNA are recovered from incomplete synthesis or from nucleolytic attack to create recombinant viruses.

Transmission, Host Range, and Epidemiology

BGMV and BGYMV have narrow host ranges. Host species are mostly restricted to the Fabaceae for both viruses. Among crop species, only *Phaseolus* spp. such as common bean (*P. vulgaris*) and lima bean (*P. lunatus*) are of significance. A.S. Costa and A.M. Almeida reported in 1979 that certain soybean (*Glycine max*) cultivars in Brazil are susceptible to BGMV. These susceptible cultivars are not considered important as virus sources for bean crops, but may probably serve as virus reservoirs between two consecutive bean plantings that coincide with the beginning and the end of the soybean growing season. The low degree of genetic variability of BGMV populations, which does not favor the emergence of new virus variants, may be one of the reasons why BGMV has never acquired economic relevance in soybean (in Brazil or in other countries).

Non-cultivated hosts of BGMV and BGYMV include *Macroptilium* spp., *Calopogonium* and wild *Phaseolus* spp. Early reports of BGMV infection in malvaceous and solanaceous hosts were probably due to mis-identification of the virus, and have not been confirmed since the use of molecular tools for virus identification became routine. It is noteworthy that fabaceous plant species used as cover crops, such as *Crotalaria* spp. and velvet bean (*Mucuna aterrina*), are non-hosts.

BGMV and BGYMV are not seedborne in any known host. BGMV from South America is not mechanically transmitted, while BGYMV from Central America and the Caribbean can be transmitted in this way with relative ease. Nevertheless, mechanical transmission is not relevant as far as virus spread in the field is concerned. The only significant form of natural transmission of both viruses in the insect vector, the whitefly *B. tabaci*.

Populations of this insect have always been recognized as genetically complex. In the past, *B. tabaci* was divided into biotypes. These are now recognized as distinct species, morphologically indistinguishable but with reproductive isolation. Up until the late-1980s, only indigenous species of *B. tabaci* such as New World 1 (NW1, previously known as biotype A) and New World 2, were present in the Americas. During the 1990s, multiple introductions of an aggressive species, *B. tabaci* Middle East-Asia Minor 1 (MEAM1, previously known as biotype B), occurred in the continent. The dissemination of *B. tabaci* MEAM1 brought dramatic changes in the incidence of begomoviruses in solanaceous crops, which are not normally colonized by *B. tabaci* NW1 but are excellent hosts for MEAM1. Although not to the same extent, MEAM1 also impacted bean golden mosaic disease. Both BGMV and BGYMV are transmitted with similar efficiencies by indigenous species and by MEAM1, but MEAM1 has a shorter life cycle and increased reproductive efficiency compared to NW1/2. Thus, MEAM1 populations reach much higher levels than NW1/2 populations, and virus transmission occurs at even higher levels. A second invasive species, *B. tabaci* Mediterranean (MED, previously known as biotype Q) has recently been introduced to South America, and is now present in southern and southeastern Brazil. The consequence of the presence of this species in the dissemination of bean golden mosaic disease is unknown, but a recent report indicated a higher efficiency of BGMV transmission by MED compared to MEAM1.

Bean golden mosaic, like other whitefly-transmitted diseases, is more severe under conditions of warm temperatures and low rainfall. In south-central Brazil, temperatures are warm during most of the year, and a pronounced dry season lasts from late June until late September. In addition, soybean crops are harvested from February until May, with massive migration of whitefly populations from soybean to other crops. As a result, June-Sep is the time of the year with the greater incidence and severity of bean golden mosaic in this region.

Virus-Host Relationships

BGMV was the object of pioneer studies on the cytopathology of geminiviruses. The 1978 work by K.S. Kim at the University of Arkansas and T. Shock and R. Goodman at the University of Illinois showed that these viruses induce major changes in the nuclear structure of infected cells, which resemble the cytopathology of animal cells infected with parvoviruses. The observed changes included the hypertrophy of the nucleolus, which eventually occupies most of the nuclear volume; the segregation of nucleolar components into granular and fibrillar regions composed of ribonucleoprotein; the presence of fibrillar rings in multiple numbers and sizes, which have been shown to contain DNA; and the appearance of viral particles either as loosely compacted aggregates or in hexagonally close-packed arrays. Virus particle aggregates were observed only in the nuclei of infected phloem cells.

Diagnosis

The correct identification of the virus infecting a plant is crucial for the selection of effective management and control procedures of the disease in the field. Bean golden mosaic disease, which is characterized by yellow and golden mosaic, leaf distortion, flower abortion, pod deformities, and stunting of plants is caused mainly by BGYMV in North and Central America and the Caribbean and by BGMV in South America. However, similar symptoms may also be induced in beans by about other 20 begomoviruses, for example, MaYSV in northeastern Brazil, tomato yellow vein streak virus in northwestern Argentina, Calopogonium golden mosaic virus in Costa Rica and Nicaragua, and also viruses from other genera and families such as alfalfa mosaic virus (fam. *Bromoviridae*, gen. *Alfamovirus*), bean yellow mosaic virus (*Potyviridae*, *Potyvirus*) and cowpea chlorotic mottle virus (*Betaflexiviridae*, *Carlavirus*) which may be present in different areas and countries. Therefore, it is not possible to identify the virus solely by relying on the type of symptoms displayed by the plant.

Different diagnostic methods may be used for begomovirus detection in bean plants. Even though antibody-based tests such as ELISA (enzyme-linked immunosorbent assay) and immunostrip assays are readily available for many plant viruses, specific antibodies and serological tests are lacking for bean-infecting begomovirus identification. Nucleic acid-based technologies are highly accurate and sensitive and are widely used for the detection of bean-infecting begomoviruses. BGYMV and BGMV were cloned and sequenced in the early 1990s, and the availability of genetic information of these and other begomoviruses allowed for the quick development of tests based on DNA molecular hybridization and polymerase chain reaction (PCR). Dot-blot and tissue blot are the most used assay formats in hybridization tests. Cloned viruses or virus genome fragments are used as probes in the hybridization tests, and the probe can be tailored to be a broad-spectrum probe and recognize several viruses or to be specific for a given virus, hybridizing only to that virus. Typically, a mixture of DNA-A components from different begomoviruses is used as a general probe under low stringency hybridization conditions, and the common region or a hypervariable fragment of the DNA-B of a particular bean-infecting begomovirus is used as a specific probe under high stringency hybridization conditions.

Currently, PCR is the most popular technique for the detection of bean-infecting begomoviruses due to its speed, sensitivity and accuracy. Different PCR set-ups, depending on the primers used, will provide general begomovirus detection or specific virus diagnosis. Usually, as a first step in identifying if a begomovirus is infecting a bean plant, very robust and effective universal degenerate primers designed in 1993 by M.R. Rojas and D. Maxwell at the University of Wisconsin are still used. Then, the amplicon is sequenced, permitting the initial identification of the begomovirus. Subsequently, back-to-back primers can be designed to recover the complete virus genome by inverse PCR for sequencing and definitive taxonomic classification. An additional set of specific primers can be designed to amplify part of the virus genome for specific diagnosis of that particular begomovirus in the samples.

More recently, alternative nucleic-acid based techniques that use isothermal amplification have been developed and are also employed for the identification of bean-infecting begomoviruses. RCA (rolling-circle amplification) uses random primers and a DNA polymerase with strand displacement properties, favoring ssDNA amplification into double-stranded products in a non-specific manner. RCA products can be cloned and sequenced allowing the identification and study of unknown viruses from symptomatic or asymptomatic samples. RCA is also valuable in discovering multiple virus infections in plants. Digestion of RCA products with restriction endonucleases and analysis of the restriction fragment length polymorphisms (RFLP) allows for the identification of mixed infections. Other diagnostic tools using isothermal amplification include loop-mediated isothermal amplification (LAMP) and recombinase polymerase amplification (RPL) assays. Protocols for LAMP and RPL are currently being developed for begomovirus diagnostics, and only RPL has been applied so far for BGYMV detection. However, these methods have great potential since they are sensitive, can use crude extracts and may be implemented to be used in the field.

High-throughput sequencing technologies (HTS), also known as next generation sequencing (NGS), have greatly advanced the detection and discovery of plant viruses. HTS coupled with appropriate bioinformatics is a very sensitive tool, having the potential to detect the whole virome of a sample, including known and unknown viruses. Hence, HTS has been considered an attractive alternative for diagnostic virology. For the identification of begomoviruses, diverse types of nucleic acid templates may be used as starting material in the sequencing process including RCA products, double-stranded RNA and small RNAs. The two latter ones enable the simultaneous detection of viruses with DNA and RNA genomes belonging to different genera and families, revealing complex multiple infections. Several HTS sequencing platforms are commercially available. However, the costs are still high for routine diagnosis.

Management

Bean golden mosaic disease is very challenging to control and manage. The disease occurs and is a serious problem in the tropical and subtropical regions of the Americas. These areas are also zones where the whitefly vector thrives, colonizing, besides common bean, other crops such as soybean, tomato, peppers, cucurbits, cotton, and numerous non-cultivated plants, which sometimes can act as virus reservoirs. Because these crops may be present in the field year-round, high populations of the whitefly are built and migrate to newly planted bean fields.

Chemical control using insecticides is the most used approach to control the whitefly vector and hence the infection caused by BGMV and BGYMV. However, the use of insecticides is not always efficient. Several types of molecules with different mechanism of action exist including nicotinoids and insect growth regulators. The over-employment of a determined insecticide chemistry over time may lead to the selection of resistant whitefly populations shortening the useful life of insecticides and, as a consequence, limiting the usefulness of chemical control. Thus, the alternation of products with distinct active ingredients should be implemented to preclude or delay resistance buildup. Severe epidemics and high yield losses are mainly recorded in the dry season when whitefly populations are high and may transmit begomoviruses to beans in the early phase of the cycle when the plants are young. To try to avoid serious yield reduction, a common practice is to treat the seeds and spray the fields with insecticide five to six times starting ten days after emergence, and then every five days until the plants are 40 days old. This type of schedule aims to control both adults and nymph stages, and as the whitefly population decreases, their migration to newly planted fields also tends to reduce. This scheme should take into consideration the insecticides' characteristics and are not appropriate for those that have longer residual activity.

Due to the limitations to the use of insecticides regarding the development of resistance in the whiteflies and environmental concerns, one of the best and sustainable options is the use of resistant cultivars. Many institutions throughout the Americas have dedicated great efforts in research programs focusing on breeding beans for resistance to golden mosaic disease. Screening attempts to identify common bean germplasm with disease immunity were unsuccessful. Thousands of bean accessions were evaluated under BGYMV and BGMV infection, and not a single genotype with full immunity was found. Some black-seeded genotypes displayed partial resistance or tolerance to BGYMV, and crosses between these genotypes rendered several lines denominated DOR that were released and cultivated, for example, in Guatemala. Further screening led to the selection of additional golden mosaic-resistant bean types with different seed colors. Genetic studies identified a few genes associated with the resistance in the different genotypes (*Bgp*, *bgm1* and *bgm2*). Pyramiding different genes into breeding lines allowed for the achievement of higher and more adequate levels of resistance and the introgression of these genes into resistant varieties. Cultivars with resistance to BGYMV were released in Honduras, Nicaragua, El Salvador, Costa Rica, Puerto Rico and Southern Florida with good acceptance by farmers.

Breeding for resistance to BGMV initiated in the 1970s in Brazil. Surveys for resistant genotypes and breeding efforts rendered bean lines with low levels of resistance that was controlled by polygenic and quantitative factors, and though a few cultivars with enhanced resistance were released, the resistance was not satisfactory under moderate to high BGMV incidence levels.

Because there were no good resistance sources to be exploited, bean improvement programs by crossing and selection of desired superior genotypes were unproductive in Brazil (the resistant cultivars planted in other Latin American countries are of different seed types which are poorly accepted by consumers in Brazil). Therefore, an alternative biotechnological approach was used to achieve BGMV resistance. The RNAi technology was explored to generate a transgenic bean line transformed with an intron-hairpin construction harboring a fragment of the BGMV *Rep* gene that expresses a *Rep*-derived double-stranded RNA and activates the RNA silencing mechanism in the plant. In this approach, the transgenic plant produces small interfering RNAs that induce the degradation of the BGMV-derived *Rep* mRNA, halting virus replication and disease progression. The transgenic line Embrapa 5.1 proved to be immune to BGMV, and after extensive risk assessment studies, this line was approved in 2011 by the Brazilian Biosafety Committee (CTNBio) for cultivation and human consumption in the country. The Embrapa 5.1 line was crossed with 'Carioca' type beans (the most widely consumed in Brazil), originating the cultivar BRS FC401 RMD which was tested in different environments, holding the resistance in the field since 2012. Even under high virus pressure, no symptoms or virus replication have been observed. BRS FC401 RMD is expected to be released for commercial production in 2020 and is recommended for cultivation in Central Brazil. It is important to point out that BRS FC401 RMD is most probably not resistant to other golden mosaic-inducing begomoviruses occurring in Brazil (MaYSV) or in other countries, and is not resistant to RNA viruses that infect beans such as bean rugose mosaic virus and cowpea chlorotic mottle virus. Hence, other measures still have to be employed to control those viruses.

Cultural practices may play a role in controlling bean golden mosaic disease. These include the sanitation of the field after harvest to eliminate any infected plants that could act as virus sources for the next crop; the avoidance of successive and continuous cultivation of the same crop in the same area; and prevention of virus transmission by sowing the new crop when the vector population and begomovirus availability are low. One control measure that is contributing to the reduction of begomovirus-associated disease in beans is the adoption of a host-free period. The host-free period usually spans from four to eight weeks after harvesting, when beans are not allowed to be cultivated. This type of measure is employed on a regional basis. Indeed, a bean-free period of four weeks in some areas in Brazil contributed to the reduction of the use of insecticides to control whiteflies and succeeded in reducing the losses due to BGMV infection. Similar success was achieved with an eight to twelve weeks of a bean-free period in the Dominican Republic in decreasing the incidence of BGYMV.

Further Reading

- Bird, J., Sanchez, J., Vakili, N.G., 1973. Golden yellow mosaic of beans (*Phaseolus vulgaris*) in Puerto Rico. *Phytopathology* 63, 1435.
- Costa, A.S., 1965. Three whitefly-transmitted diseases of beans in the state of São Paulo, Brazil. *FAO Plant Protection Bulletin*, 13, pp. 121–130.
- Faria, J.C., Aragão, F.J.L., Souza, T., *et al.*, 2016. Golden mosaic of common beans in Brazil: Management with a transgenic approach. *APS Features*. Available at: www.apsnet.org/edcenter/apsnetfeatures/Pages/GoldenMosaic.aspx.
- Gilbertson, R.L., Hidayat, S.H., Martinez, R.T., *et al.*, 1991. Differentiation of bean-infecting geminiviruses by nucleic acid hybridization probes and aspects of bean golden mosaic in Brazil. *Plant Disease* 75, 336–342.
- Kim, K.S., Shock, T.L., Goodman, R.M., 1978. Infection of *Phaseolus vulgaris* by bean golden mosaic virus: Ultrastructural aspects. *Virology* 89, 22–33.
- Morales, F.J., 2001. Conventional breeding for resistance to *Bemisia tabaci*-transmitted geminiviruses. *Crop Protection* 20, 825–834.
- Morales, F.J., Jones, P.G., 2004. The ecology and epidemiology of whitefly-transmitted viruses in Latin America. *Virus Research* 100, 57–65.
- Ramos-Sobrinho, R., Xavier, C.A.D., Pereira, H.M.B., *et al.*, 2014. Contrasting genetic structure between two begomoviruses infecting the same leguminous hosts. *Journal of General Virology* 95, 2540–2552.
- Rojas, M.R., Gilbertson, R.L., Russel, D.R., Maxwell, D.P., 1993. Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. *Plant Disease* 77, 340–347.
- Rojas, M.R., Hagen, C., Lucas, W.J., Gilbertson, R.L., 2005. Exploiting chinks in the plant's armor: Evolution and emergence of geminiviruses. *Annual Review of Phytopathology* 43, 361–394.
- Rojas, M.R., Macedo, M.A., Maliano, M.R., *et al.*, 2018. World management of geminiviruses. *Annual Review of Phytopathology* 56, 637–677.
- Silva, S.J.C., Castillo-Urquiza, G.P., Hora-Junior, B.T., *et al.*, 2012. Species diversity, phylogeny and genetic variability of begomovirus populations infecting leguminous weeds in northeastern Brazil. *Plant Pathology* 61, 457–467.

Beet Curly Top Virus (*Geminiviridae*)

Robert L Gilbertson, Tomas A Melgarejo, and Maria R Rojas, University of California, Davis, CA, United States
William M Wintermantel, Agricultural Research Service, US Department of Agriculture, Salinas, CA, United States
John Stanley, John Innes Centre, Colney, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J. Stanley, Beet Curly Top Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00696-8.

Glossary

Agroinoculation A method for inoculating plants with infectious clones of a virus inserted in between the T-DNA borders of a binary vector such that the T-DNA is inserted into the plant nucleus via a disarmed *Agrobacterium tumefaciens* strain.

Curly top disease A viral disease of plants characterized by stunting, up or downward curling and distortion of leaves, vein swelling, enations and purpling; the disease can be caused by genetically diverse geminiviruses, including *Beet curly top virus*.

Enation Virus-induced outgrowth from leaves (typically veins) or other plant tissues.

Endoreduplication DNA replication in the absence of cell division.

Etiology The cause of a disease.

Hemolymph The insect circulatory system.

Hyperplasia Unregulated cell division.

Hypertrophy Cell enlargement resulting from abnormal increase in cell size.

Non-propagative transmission Virus transmission mode in which the virus does not replicate in the vector.

Phloem-limited virus A virus that only infects cells and tissues of the phloem.

RFLP Restriction fragment length polymorphism, a molecular tool for fingerprinting and identifying DNA viruses.

Steckling A small plant composed mostly of the root of a biennial crop plant (e.g., carrot or sugar beet) that is usually exhumed, stored during the winter and replanted the next season for seed production.

Transovarial transmission Transmission of a virus vertically, through the insect egg.

Viruliferous The condition in which an insect vector is carrying a virus and can transmit the virus to an uninfected plant.

Introduction

Beet curly top virus (BCTV) is the type species of the genus *Curtovirus*, family *Geminiviridae*. Like all members of this family, BCTV has twinned quasi-icosahedral virions that measure $\sim 18 \times 30$ nm and a circular single-stranded DNA genome. BCTV is a phloem-limited virus that is transmitted by the beet leafhopper (BLH; *Circulifer tenellus* (Baker) (Homoptera: *Cicadellidae*) (Fig. 1(A)), and causes curly top disease of several economically important crops including common bean, pepper, sugar beet and tomato (e.g., Fig. 1(B–F)) in the western USA, north-central Mexico and countries of the Mediterranean Basin and the Middle East (e.g., Iran and Turkey), where it can cause substantial economic losses. Disease outbreaks are driven by populations of the beet leafhopper vector and favorable environmental conditions. In 2013, the disease caused an estimated US\$100 million loss to the California processing tomato industry. Research on various aspects of the virus has added important fundamental knowledge about geminiviruses and the complex epidemiology of plant viruses transmitted by insect vectors. BCTV and the beet leafhopper vector represent an early example of the introduction of an exotic virus-vector combination into a new geographic region, the New World (NW) through movement of plant materials (sugar beet roots and stecklings) by humans. The BCTV genome sequence provided early evidence of the important role of recombination in viral evolution. Studies of the etiology of curly top disease in different geographic regions revealed that this disease is caused by multiple genetically distinct geminiviruses. Curly top disease also serves as a success story for how the development and implementation of resistant varieties and other management strategies saved the sugar beet industry from economic collapse. Finally, the BCTV-BLH vector system may lead to the identification of the elusive receptor(s) involved in the persistent circulative transmission of plant viruses by their insect vectors.

History

The first reports of curly top disease were from sugar beet fields in the western USA in 1888. Along with the establishment and rapid growth of the sugar beet industry in this region (e.g., states of California, Idaho and Utah), came widespread and damaging outbreaks of curly top disease of sugar beet and other crops such as common bean, squash and tomato. Outbreaks of curly top disease caused substantial yield losses to sugar beet production in the western USA from ~ 1920 –1940 and the disease became a limiting factor in sugar beet production during this time, causing abandonment of major production areas and the closing of sugar

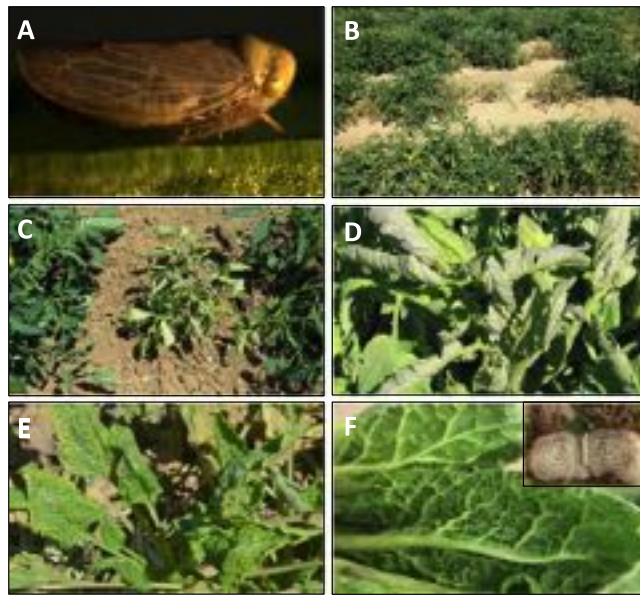


Fig. 1 (A) Beet leafhopper (*Circulifer tenellus*) feeding on a vein on the abaxial side of a sugar beet leaf infected with *Beet curly top virus* (BCTV); (B) Portion of a tomato field with plants infected with BCTV, showing severe stunting and yellowing and poor stand due to plants that have died from the disease; (C) BCTV-infected tomato plant showing typical symptoms of curly top including stunting, upward leaf curling and yellowing; (D) Close-up of foliar symptoms of curly top disease in tomato showing upward rolling, dull green to yellow coloration and swollen and purple veins; (E) BCTV-infected sugar beet plant showing foliar symptoms of curly top on the leaves, including yellowing, crinkling, and upward curling; (F) Abaxial side of leaves of a BCTV-infected sugar beet plant showing swollen veins and enations associated with curly top disease; and inset showing darkening of vascular tissue in the root of a sugar beet plant infected with BCTV and showing curly top symptoms.

refineries. Subsequently, curly top-like diseases were reported to occur in western Canada, north-central Mexico, Brazil, Argentina, the Caribbean Basin, Turkey, and Iran. Here, it is important to note that molecular characterization of the causal agent of curly top disease from different geographical regions has revealed that these disease symptoms can be caused by multiple genetically distinct geminiviruses, including some that are not in the genus *Curtovirus*. Thus, symptoms alone are insufficient for diagnosis, and molecular tests (PCR and sequencing) are needed to confirm the identity of the causal agent.

An association between curly top disease and the beet leafhopper (**Fig. 1(A)**) was noted, and evidence that this insect was a vector of curly top disease was first presented in 1909. These phloem feeding insects acquired and transmitted the causal agent of curly top rapidly (minutes), although longer periods of acquisition were needed for high rates of transmission. The curly top agent also persisted within the leafhopper vector, but it did not appear to replicate in the insect. Although the leafhopper transmission properties supported a viral etiology for curly top disease, the low level of accumulation (viral titer) of the causal agent in plants made purification difficult. This was mostly due to the virus exclusively infecting cells of the phloem. It was not until the 1970s that dimeric or twinned virus-like particles, measuring ~20–30 nm, were purified from curly top-infected plants. This suggested that the causal agent of curly top might be a geminivirus, and further evidence came from the development of curly top symptoms in plants exposed to beet leafhoppers that had fed on purified preparations of the twinned virus-like particles. These results indicated that curly top disease was caused by a geminivirus, which was named *Beet curly top virus*.

The viral genomic DNA was eventually cloned from DNA isolated from leaves of sugar beet plants with curly top disease symptoms, and the DNA sequence was determined in 1986. Analysis of the BCTV sequence confirmed that BCTV was a geminivirus, and also revealed that the virus has a hybrid genome composed of complementary-sense genes having greatest identities with those of whitefly-transmitted geminiviruses (genus *Begomovirus*) and virion-sense genes having highest identities with those of leafhopper-transmitted geminiviruses (genus *Mastrevirus*) (**Fig. 2**). This clearly indicated that recombination played a major role in the emergence of BCTV. The cloned genomic DNA of BCTV induced curly top symptoms in sugar beet plants, when delivered by agroinoculation, thereby fulfilling Koch's postulates for curly top disease. Progeny virions derived from the infectious BCTV clone were transmitted by beet leafhoppers and subsequent development of curly top symptoms in inoculated sugar beet plants confirmed leafhopper transmissibility of BCTV.

Classification and Nomenclature

Early investigations of host range, symptoms, and virulence suggested that curly top disease in the western USA was caused by a complex of viral strains/species. To some extent, this hypothesis has been supported by molecular (RFLP and sequencing) studies showing genetic diversity among the viruses causing curly top disease. In the western USA three strains were characterized based on

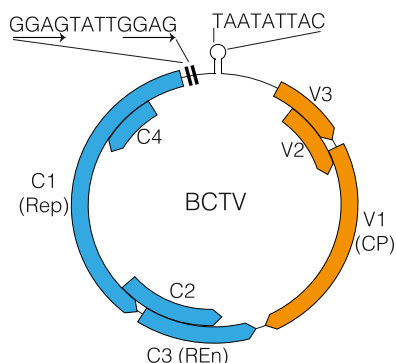


Fig. 2 The genome organization of *Beet curly top virus* (BCTV). The position and orientation of virion-sense (V) and complementary-sense (C) open reading frames (ORFs) are shown in relation to the intergenic region (IR) that contains the invariant nonanucleotide motif TAATATTAC and iterons (repeated sequences underlined with arrows) involved in replication-associated protein (Rep) recognition and the initiation of viral DNA replication. The recombinant nature of the BCTV genome is shown with the virion-sense genes having highest identities with those of leafhopper-transmitted geminiviruses (mastreviruses) shown in orange, and the complementary-sense genes having highest identities with those of whitefly-transmitted geminiviruses (begomoviruses) shown in blue. CP, capsid protein; and REEn, replication enhancer protein.

Table 1 History of the taxonomy and nomenclature for phenotypic and genotypic variants of *Beet curly top virus* (BCTV) and other curtoviruses before and after the advent of molecular tools

		<i>Molecular era</i> ^a					
<i>Before 1990</i>		<i>1990</i>		<i>1998</i>		<i>2014</i>	
<i>Species</i>	<i>Strains</i>	<i>Species</i>	<i>Strains</i> ^b	<i>Species</i> ^c	<i>Strains</i>	<i>Species</i> ^c	<i>Strains</i> ^b
BCTV	Various	BCTV	BCTV-Cal/Logan BCTV-Worland BCTV-CFH	BCTV BMCTV BSCTV SpCTV PeCTV PeYDV		BCTV	BCTV-Cal/Logan BCTV-Wor BCTV-Mld BCTV-Svr BCTV-SpCT BCTV-PeCT BCTV-PeYD BCTV-CO BCTV-LH71 BCTV-SvrPep BCTV-Kim1
		BCTV-HRCT		HrCTV		HrCTV SpSCTV	

^aMolecular tools employed: 1990, restriction fragment length polymorphism (RFLP); 1998, RFLP and DNA sequencing; and 2014, DNA sequencing and phylogenetic analysis.

^bStrain abbreviations: BCTV-Wor, BCTV-Worland; BCTV-Mld, BCTV-Mild; BCTV-Svr, BCTV-Severe; BCTV-SpCT, BCTV-spinach curly top; BCTV-PeCT, BCTV-pepper curly top; BCTV-PeYD, BCTV-pepper yellow dwarf; BCTV-CO, BCTV-Colorado; BCTV-LH71, BCTV-leafhopper71; BCTV-SvrPep, BCTV-severe pepper; BCTV-Kim1, BCTV-Kimberly1; BCTV-HRCT, BCTV-Horseradish curly top.

^cSpecies abbreviations: BMCTV, Beet mild curly top virus; BSCTV, Beet severe curly top virus; SpCTV, Spinach curly top virus; PeCTV, Pepper curly top virus; PeYDV, Pepper yellow dwarf virus; HrCTV, Horseradish curly top virus; and SpSCTV, Spinach severe curly top virus.

molecular and biological properties: BCTV-Cal/Logan, BCTV-Worland and BCTV-CFH, with the strain designation generally reflecting the geographic origin of these strains (Table 1). Another strain was recovered from a horseradish plant with brittle root disease in Illinois and was designated BCTV-HRCT. The first strain for which the complete sequence was determined was BCTV-Cal/Logan in 1986. As sequences of more BCTV strains and putative new curtoviruses became available, additional genetic diversity as well as differences in biological properties (e.g., host range, symptomatology in sugar beet plants, and transreplication) were revealed. The taxonomy of viruses causing curly top disease is now largely based on genome properties, e.g., genome-wide sequence identities, along with some consideration of biological properties of the viruses.

At the same time, the properties used to classify geminiviruses were also being defined and used to establish genera within the family *Geminiviridae*. The key properties that were used to classify geminiviruses were: genome organization, DNA sequence relationships, type of insect vector and host range of the virus. The viruses causing curly top disease in the western USA were placed into the genus *Curtovirus* (name derived from *curly top*) based upon (1) having a monopartite genome, (2) a high degree of sequence identity among members of the genus and divergence from sequences of species of other genera, (3) being transmitted by the beet leafhopper and (4) infecting dicotyledonous plants. There were originally four geminivirus genera: *Curtovirus*, *Begomovirus* (whitefly-transmitted and dicot-infecting),

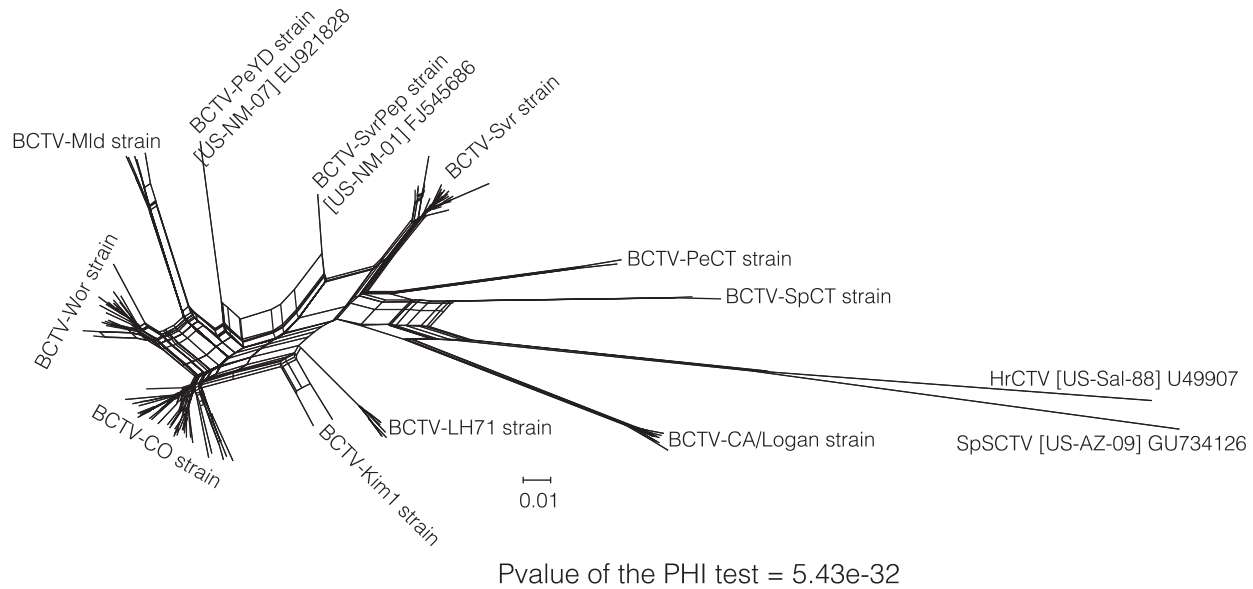


Fig. 3 Phylogenetic network generated with the complete genomes of 104 isolates of *Beet curly top virus* (BCTV) and illustrating the grouping of the isolates into the eleven recognized strains of the virus and their relationships with one another. The other two species in the genus *Curtovirus*, *Spinach severe curly top virus* (SpSCTV) and *Horseradish curly top virus* (HrCTV), were used as out-groups. The *P*-value of the pairwise homoplasy index (PHI) statistical test of recombination is shown on the bottom of the network.

Mastrevirus (leafhopper-transmitted and mostly monocot-infecting) and *Topocuvirus* (treehopper-transmitted and dicot-infecting). However, with the advent of inexpensive DNA sequencing and rapid methods for the amplification and cloning of geminivirus genomes, new types of geminiviruses have been identified and characterized. Thus, the family is now subdivided into nine genera with five new genera recently established: *Eragrovirus*, *Capulavirus*, *Grablovirus*, *Becurtovirus*, and *Turncurtovirus*. Here, it is important to note that the genera *Becurtovirus* and *Turncurtovirus* contain viruses, some of which are transmitted by the leafhopper *C. haematoceps*, that cause curly top-like disease symptoms in sugar beet and other crops. Thus, curly top disease symptoms can be caused by multiple genetically divergent geminiviruses from three genera. Therefore, molecular methods are needed to conclusively identify the virus(es) associated with the disease in a given geographical region.

The use of DNA sequences for classification and phylogenetic analysis of geminiviruses has provided important insight into viral evolution and the relationships among members of this large family of viruses. However, the relationships among some geminiviruses have been complicated by extensive recombination, especially those in the genera *Begomovirus* and *Curtovirus*. Thus, because recombination has played a major role in their evolution, it can be difficult to establish an exact lineage for some viruses, i.e., whether it represents the parent or progeny of a recombination event. Another important issue is where to establish the cut-off values for establishing species, strain and isolate relationships in different genera. With the large number of sequences now available, these values are typically established based upon genome-wide pairwise identities.

The original species demarcation value for curtoviruses was 89% identity, i.e., sequences having identities >89% to a known species represent isolates or strains of that species. This was the same species demarcation value used for begomoviruses. Other criteria that were considered in establishing curtovirus species were (1) lack of transreplication of known species mediated by the Replication-associated protein (Rep), (2) serological properties, and (3) biological properties (e.g., symptoms in sugar beet). Based upon these criteria, three BCTV strains were elevated to species in 1998: *Beet curly top virus* (type species, BCTV-Cal/Logan strain), *Beet mild curly top virus* (BCTV-Worland strain [BCTV-Wor]) and *Beet severe curly top virus* (BCTV-CFH strain) (Table 1). In addition, BCTV-HrCT was elevated to species (*Horseradish curly top virus*, HrCTV), as were at least five other curtoviruses associated with curly top disease in various crops (formally or informally). However, in 2014, the taxonomy of curtoviruses was revisited and, based upon genome-wide pairwise identities of all available curtovirus sequences, a new species demarcation value of 77% identity or greater was established, with little or no consideration of other properties. Based upon these new criteria, the genus *Curtovirus* currently includes only three species: the type species BCTV, HrCTV and the more recently described *Spinach severe curly top virus* (SpSCTV) (Fig. 3 and Table 1). The species BCTV is comprised of 11 recognized strains, which share <94% identity and, in some cases have different serological and biological properties (Fig. 3 and Table 1). Like BCTV-Cal/Logan, most of the recognized strains of BCTV have recombinant genomes (e.g., Fig. 4). HrCTV and SpSCTV also have recombinant genomes and lack a C3 ORF. The latter two species appear to have limited distribution and minimal economic impact, perhaps reflecting reduced fitness due to these genomic properties. For example, HrCTV was isolated in 1990 from a horseradish plant with brittle root disease in Illinois, USA, but Koch's postulates were not fulfilled for this disease nor has the virus emerged as a pathogen of economic importance. However, as the genomes of HrCTV and SpSCTV are composed mostly of curtovirus sequences and are well below the 77% species demarcation threshold, they clearly represent distinct curtovirus species.

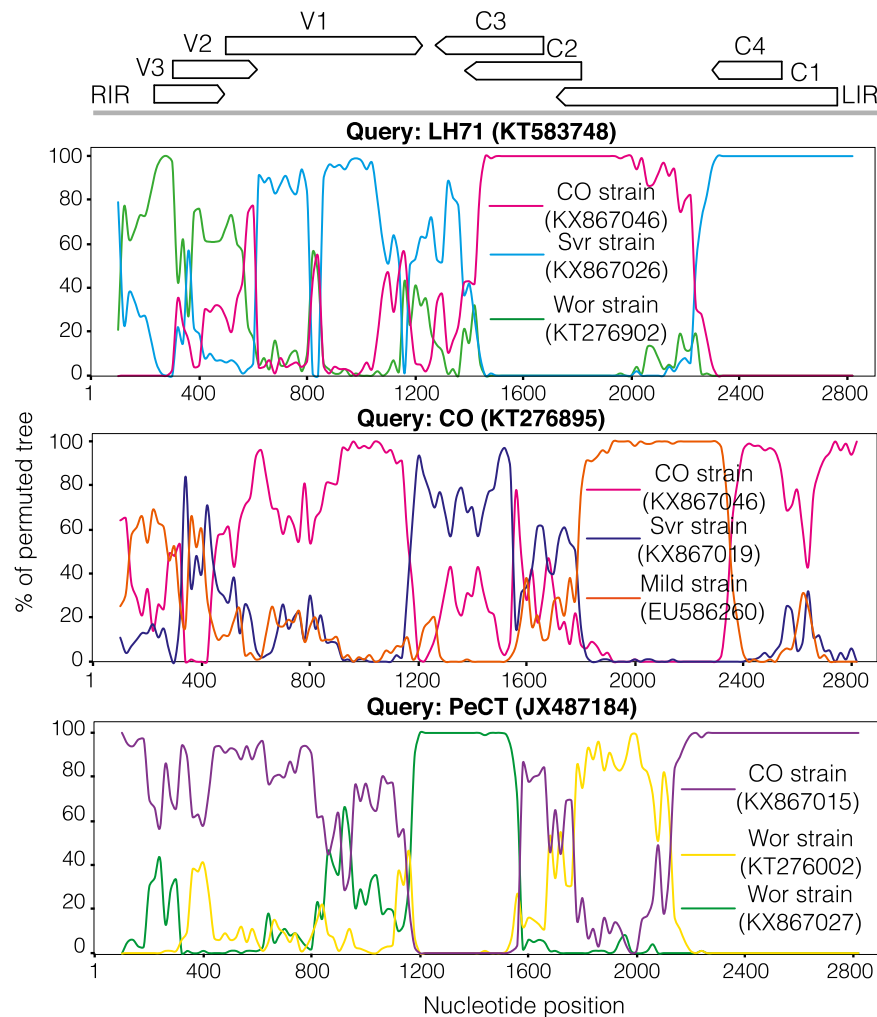


Fig. 4 Similarity plot analysis of complete genome sequences of three strains of *Beet curly top virus* (BCTV): BCTV-CO, BCTV-LH71 and BCTV-PeCT. Bootscan analysis was performed with a sliding window of 200 nucleotides and step size of 20 nucleotides with Neighbor-Joining phylogenetic trees with 100 bootstrap replicates inferred in each case. This analysis provides a schematic representation of the recombination profile of each of these strains (query) as revealed by comparisons with other BCTV sequences in the NCBI database. The horizontal arrows on the top indicate the position of the open reading frames (ORFs) encoded by the BCTV genome. LIR, left side of the intergenic region and RIR, right side of the intergenic region.

The Origin of BCTV and the Beet Leafhopper Vector: An Early Example of Global Transport of a Plant Pathogen and its Vector

Curly top disease continues to have the greatest economic impact on sugar beet and processing tomato production in the western USA. Soon after the disease became a major problem for sugar beet production, there were questions regarding the origin of BCTV and the beet leafhopper vector, i.e., were they indigenous to the NW or was one or both introduced from another region? Multiple lines of evidence suggest that BCTV originated in the OW and was introduced, along with the beet leafhopper, into the NW via movement of sugar beet propagative materials (e.g., roots and Stecklings) brought with settlers coming from the OW (e.g., Europe and the Mediterranean Basin). The evidence for the introduction of the beet leafhopper included (1) the presence of *C. tenellus* and its closest relative in the OW (Mediterranean Basin), (2) interbreeding between NW and OW beet leafhopper populations, and (3) transmission of the causal agent of curly top in the NW by beet leafhoppers from the OW. The origin of BCTV remained unclear as there were no reports of curly top disease in the OW. However, in 1955, curly top disease of sugar beet was reported from Turkey and, subsequently, from Iran. In 1998, an isolate of BCTV from sugar beet plants with curly top disease symptoms in Iran (BCTV-I) was reported to have a high level of nucleotide sequence identity with BCTV strains from the NW, including >98% identity with BCTV-CFH (now BCTV-Svr [US-SVR-Cfh]). Moreover, an infectious clone of BCTV-I induced symptoms in *N. benthamiana* and sugar beet that were indistinguishable from those induced by the infectious clone of BCTV-CFH. This suggested a common origin for BCTV in the NW and OW, though it did not resolve the origin of BCTV. Evidence for an OW origin of BCTV comes from (1) the high levels of resistance to curly top disease in wild beet species in the OW (where sugar beet originated), suggesting virus-host coevolution over a long period of time,

and (2) a portion of the recombinant genome of BCTV was derived from an OW leafhopper-vectored mastrevirus, consistent with emergence and evolution of BCTV in the OW. Although more BCTV genetic diversity has been documented in the NW, this could simply reflect the greater amount of research conducted on BCTV in the NW, or multiple introductions of BCTV strains from the OW and subsequent evolution mediated by mutation and recombination, which are facilitated during major curly top disease outbreaks. Taken together, these results support the hypotheses that both BCTV and the beet leafhopper vector originated in the OW and were introduced into the NW by the activities of humans. Unfortunately, the desert environment of the western USA was ideal for establishment of populations of beet leafhoppers and the associated BCTV.

Geographic and Seasonal Distribution

Curly top disease occurs widely throughout arid and semiarid regions of the western USA, where high populations of the beet leafhopper vector can occur. It is here that the epidemiology of the disease and the life cycle of the insect vector has been extensively investigated. At the end of the growing season (in the fall), adult female beet leafhoppers (some of which are viruliferous) migrate from agricultural fields to overwintering areas (e.g., foothills of mountain ranges located on the western side of the Central Valley of California), where they survive on perennial weeds. In late winter to early spring, these females lay eggs into stems and petioles of perennial and annual weeds. The resulting nymphs and adults of this new generation acquire BCTV as they feed in the phloem of infected weed reservoir hosts, most of which show few or no symptoms. Initially, the virus and vector survive on perennial weeds such as Russian thistle (*Salsola* sp.) and saltbush (*Atriplex* sp.), until winter rains lead to the emergence of annual weeds, such as filaree (*Erodium* sp.), peppergrass (*Lepidium* sp.) and plantain (*Plantago* sp.) on which the leafhoppers propagate. As these winter annuals become dry in mid- to late-spring, some combination of viruliferous and non-viruliferous adult beet leafhoppers migrate from the overwintering areas to agricultural areas (often valleys), where they transmit BCTV to crops including sugar beet (*Beta vulgaris*), tomato (*Solanum lycopersicum*), common bean (*Phaseolus vulgaris*), and pepper (*Capsicum annuum*), as well as to common agricultural weeds as they search for preferred hosts for reproduction. Depending on the geographic region, beet leafhoppers undergo 3–5 generations on weeds, such as wild mustards (*Brassica* sp.), London rocket (*Sisymbrium irio*), Shepherd's purse (*Capsella bursa-pastoris*) and goosefoot (*Chenopodium* sp.). As crops are harvested and summer annuals dry in the fall, some combination of viruliferous and non-viruliferous adult female leafhoppers migrate from agricultural areas back to their overwintering areas.

Changes in BCTV Strain Prevalence

BCTV strain diversity was investigated in samples of sugar beets with curly top disease collected from commercial fields in Texas in 1994 with restriction fragment length polymorphism (RFLP) analysis. The predominant strains were BCTV-Wor (previously BMCTV) and BCTV-Svr [US-SVR-Cfh] (previously BSCTV), with no detection of BCTV-Cal/Logan (type species). BCTV-Wor and BCTV-Svr were also the predominant strains detected in a more comprehensive survey of sugar beets with curly top symptoms collected from California, Colorado, Idaho, New Mexico, Oregon, Texas, Washington, and Wyoming in 1995. In a survey of tomato plants with curly top symptoms, beet leafhoppers and weed species in the Central Valley of California conducted between 2002 and 2008, BCTV-Wor and BCTV-Svr were prevalent, although BCTV-Cal/Logan was detected in a small number of leafhopper samples. Another survey of BCTV strains associated with curly top of sugar beet in the western USA was conducted from 2006–2007 and, again, BCTV-Wor and BCTV-Svr were prevalent, although a small number of plants from Idaho and Oregon were infected with BCTV-Cal/Logan. In these surveys, it was not uncommon to find plants infected with genotypic variants of these strains or co-infections of more than one strain.

Interestingly, a major shift in the BCTV population structure has been detected in recent years in sugar beet and processing tomato samples from the Pacific Northwest and the Central Valley of California, respectively. A survey of sugar beets with curly top symptoms from Idaho and Oregon was conducted between 2012 and 2015 in response to concerns about increased curly top disease severity. The results revealed that (1) BCTV-CO and BCTV-Wor had become the prevalent strains, (2) there was a re-emergence of BCTV-Cal/Logan (30% of isolates detected) and (3) there was a decline in the prevalence of BCTV-Svr. Although BCTV-CO and BCTV-Wor are closely related (Figs. 3 and 4), they were differentiated by whole genome sequencing. Following the 2013 curly top outbreak in processing tomatoes in the Central Valley of California (where there is currently little or no sugar beet production), surveys of tomato plants with curly top symptoms and symptomless weeds in the foothills conducted from 2014 to 2018 revealed that BCTV-CO and BCTV-LH71, a new recombinant strain initially identified from viruliferous leafhoppers (Fig. 4) were predominant, rather than BCTV-Wor and BCTV-Svr. This was particularly true with samples of tomatoes with curly top symptoms, whereas more strain diversity was detected in the weeds. Interestingly, in crop and weed samples from the Imperial Valley of California, where sugar beets are still grown, more BCTV-Svr was detected. In addition, two other BCTV strains were detected in pepper and tomato samples collected in California in 2014–2018, BCTV-spinach curly top (BCTV-SpCT) and BCTV-pepper curly top (BCTV-PeCT) (Table 1). Together, these results revealed a dramatic shift in strain prevalence in two different crops in two geographical regions, and the emergence of BCTV-CO as a major strain in both regions. Although the basis for this population shift in BCTV strains remains unknown, it may reflect changes in cropping patterns or in the predominant weed reservoirs surrounding agricultural fields and selection of more virulent or fit strains. For example, BCTV-CO and BCTV-LH71 appear to be better adapted to tomato.

Genome Organization and Gene Expression

Like all geminiviruses, BCTV has a genome of circular single-stranded DNA (ssDNA) that is encapsidated in twinned (geminate) quasi-icosahedral particles measuring approximately 18×30 nm. The BCTV genome is composed of a single genomic DNA of ~ 3000 nt that encodes seven genes distributed between the virion-sense (V) and complementary-sense (C) strands (also referred to as rightward [R] and leftward [L] strands in the literature), separated by an intergenic region (IR) of ~ 300 nt, which contains the origin of replication (Fig. 2). A nonanucleotide motif (TAATATTAC), which is highly conserved among all geminiviruses, is located within the IR and lies in between inverted repeat sequences with the potential to form a stem-loop structure. Short repeat sequences, termed iterons, are upstream of the stem-loop (Fig. 2). The complementary-sense genes encode the Rep protein (ORF C1) that is required for viral DNA replication, the replication-enhancer REn (ORF C3), and two proteins that contribute to viral pathogenicity (ORFs C2 and C4). The virion-sense gene products include the capsid protein (CP, ORF V1) and proteins involved in the regulation of the relative levels of viral ssDNA and double-stranded DNA (dsDNA) (ORF V2) and virus movement (ORF V3) (Fig. 2). The complementary-sense gene products are required during the early stages of infection, whereas the virion-sense ORFs encode proteins that are required later in infection, as the virus shifts to virion formation and movement within the plant.

The identification of BCTV virion- and complementary-sense transcripts is consistent with a bidirectional transcription strategy that, along with overlapping genes (in different frames), maximizes the number of genes encoded by a small genome (Fig. 2). The two most abundant complementary-sense transcripts are similar to their begomovirus counterparts, with the larger transcript mapping across all four ORFs and the smaller transcript across ORFs C2 and C3. Overall, the most abundant transcripts mapped across the virion-sense ORFs and downstream of two consensus eukaryotic promoter sequences. The most abundant transcript mapped to the ORFs V2 and V3, whereas a less abundant transcript mapped to ORF V1. Precisely how the viral proteins are translated from these transcripts is not yet understood, but the overlapping nature of both the virion-sense and complementary-sense ORFs may provide temporal control during the infection cycle.

Replication

The virion-sense ssDNA carried in the BCTV virion must gain access to the nucleus of phloem cells (e.g., companion and phloem parenchyma), where replication of viral DNA occurs. It is not known if virions or viral DNA released from disassembled virions enters the nucleus following introduction of virions by the beet leafhopper vector. Once in the nucleus, the virion-sense ssDNA is converted into dsDNA, mediated by host-derived short DNAs as primers and host DNA polymerases (note that geminiviruses do not encode a DNA polymerase). Fractionation of viral DNA forms by a combination of chromatography and two-dimensional gel electrophoresis has identified BCTV intermediates consistent with both rolling-circle and recombination-dependent replication strategies. During rolling-circle replication, Rep interacts with the iterons located in the IR and introduces a nick into the virion-sense strand of the dsDNA within the nonanucleotide motif (TAATATTAC). By analogy with the replication strategy of begomoviruses, Rep then forms a covalent bond with the 5' terminus of the nicked virion-sense strand, whereas the 3' terminus must be extended by a host DNA polymerase. The full-length nascent strand is nicked and ligated by Rep to produce a genomic circular ssDNA that either re-enters the replication cycle (i.e., be converted to dsDNA early in infection) or is encapsidated to form virions (later in infection).

The interaction between Rep and the origin of replication is defined by the amino acid sequence of the N-terminal region of Rep and the iteron sequences of the origin of replication. This interaction is highly specific and determines whether transreplication occurs between strains or species (i.e., the Rep of one strain/species mediates the replication of another strain/species). For example, the Rep of the BCTV-Cal/Logan strain functionally interacts with its own iteron, GGAGTATTGGAGT (iteron underlined) (Fig. 2), but not with those of BCTV-Wor [US-Mld-Wor], GGTGCTATCGGAG, and BCTV-Svr [US-SVR-Cfh], GGTGCTTTGGGTG. Conversely, the Rep proteins of BCTV-Wor and BCTV-Svr do not functionally interact with the iterons of BCTV-Cal/Logan, whereas the BCTV-Wor and BCTV-Svr iterons are sufficiently similar to allow functional interaction with the Rep proteins of these strains and, thus, transreplication. Comparison of Rep N-terminal amino acid sequences has revealed residues conserved between BCTV-Wor and BCTV-Svr and that differ from those of the BCTV-Cal/Logan Rep. These residues, referred to as the iteron recognition domain (IRD), are involved in iteron recognition. Whereas replication incompatibility between cis- and trans-acting elements may be considered to maintain the integrity of a particular species, this constraint can be overcome by a recombination event in which the functional module comprising the 5' terminus of the Rep ORF and the origin of replication is exchanged between incompatible viruses.

Indeed, in addition to mediating the emergence of the hybrid genome of the BCTV progenitor, recombination also has played a major role in the evolution and emergence of other BCTV strains, including BCTV-CO, BCTV-LH71, and BCTV-PeCT (Fig. 4). This is mediated by recombination-dependent replication, which typically occurs later in the infection cycle compared with rolling-circle replication, possibly due to higher DNA concentrations and the presence of ss- and dsDNA forms that mediate template switching and recombination. It remains to be established which, if any, viral proteins contribute to this process, although it is possible that Rep helicase activity could participate in both replication strategies. Additionally, the nick site for rolling-circle replication has been shown to be a recombinational hot spot. Thus, recombination-dependent replication likely explains the high frequency of recombinant geminiviruses.

Subgenomic-sized defective viral DNAs of diverse size (ranging from 800 to 1800 nt) and complexity are rapidly produced from the cloned BCTV genomic DNA in sugar beet, tomato, and *N. benthamiana*. Deletions occur within all ORFs and only the IR

sequence is conserved, consistent with its participation in viral DNA replication. BCTV-infected sugar beet plants grown in fields frequently contains such defective DNAs, indicating that they are also produced under natural conditions. They occur as both single- and double-stranded DNA forms but it is not known if they are encapsidated or have the ability to move systemically in the plant. However, in view of their rapid appearance, it is likely that they are produced *de novo* in every infected cell. Short repeat sequences of 2–6 nt at the deletion boundaries may specify the deletion endpoints and suggest that the defective DNAs are produced by recombination. These subgenomic-sized DNAs can interfere with infection, presumably by competing for a limited amount of Rep, and may have a biological role in modulating pathogenicity of the helper virus to encourage leafhopper feeding and virus transmission.

Virus Movement

The CP is a multifunctional protein that is targeted to the nucleus, probably via one or more nuclear localization signals (NLSs) in the N-terminus, where virion formation and accumulation occurs. In transient expression experiments in protoplasts and bombarded common bean hypocotyl tissues, a BCTV-Wor [US-Mld-Wor4] (previously BMCTV) CP-GFP fusion protein was targeted to the nucleus and formed ring-like structures, reminiscent of fibrillary rings observed in nuclei of BCTV-infected cells in ultra-structural studies. Late in infection, virions and viral nucleic acids form large aggregates or inclusion bodies in nuclei of BCTV-infected phloem cells.

The BCTV CP is essential for systemic infection and leafhopper transmission. BCTV-Wor [US-Mld-Wor4] CP alanine scanning mutants that retain the ability to form virions (mainly N-terminal mutations) were generally able to produce a systemic infection and were leafhopper-transmissible, whereas those unable to form virions (mainly C-terminal mutations) were not, despite being competent for replication. Thus, these results suggest that the C-terminal amino acids of the CP play an important role in virion assembly, and that long-distance movement of BCTV (in the plant) occurs in the form of virions, which is consistent with the observation of virus-like particles in phloem sap exudates of leaves of *Amsinckia douglasiana* plants with curly top symptoms.

ORF V2 mutants accumulate high levels of dsDNA and greatly reduced levels of ssDNA compared with the wild-type virus, implicating the V2 protein in the regulation of the levels of these viral DNA forms. However, the V2 protein does not have a typical NLS, and in transient expression experiments in protoplasts and bombarded hypocotyl tissues, a V2-GFP fusion protein was localized to the endoplasmic reticulum (ER), including having a perinuclear distribution. If the function of the V2 protein is to ensure that ssDNA is available in sufficient amounts for encapsidation and systemic movement late in infection, this may involve V2-mediated nucleocytoplasmic transport of an infectious form of the virus, e.g., virions or some other nucleoprotein complex. The possibility of an infectious non-virion form of BCTV was suggested based on results with alanine scanning mutant CP 49–51, which infected and caused curly top symptoms in *N. benthamiana* plants, but did not form virions and was not leafhopper transmissible.

ORF V3 mutants are competent for replication but produce only sporadic and symptomless infections with low levels of viral DNA in sugar beet and *N. benthamiana* plants, suggesting a role for the V3 protein in virus movement. In transient expression experiments in protoplasts and hypocotyl tissues, V3-GFP initially localized to the nuclear periphery and ER, similar to V2-GFP. However, V3-GFP subsequently was localized to motile ER-derived vesicle-like structures, which were observed around the nucleus, in the cytoplasm and at the cell periphery. It is tempting to speculate that this represents a transport mechanism, possibly delivering an infectious form of the virus to the cell periphery and plasmodesmata for limited cell-to-cell movement.

Finally, it is important to consider how the virions are acquired by leafhoppers during the feeding process. Although virions seem to be the main form in which BCTV moves in the phloem, it is not known if this occurs as individual virion particles or aggregates (e.g., intact or degraded parts of inclusion bodies). It is also not clear how virions enter the phloem. One hypothesis involves the collapse and death of BCTV-infected phloem cells and the release of cell contents, including inclusion bodies, into the developing phloem sieve elements, possibly in the shoot apex. These virions would then be available for acquisition by beet leafhoppers or long-distance movement in the phloem in a source-to-sink manner to infect new protophloem and other phloem cells in the newly developing root and shoot apices. In this ‘wave of infection’ model, the virus is continually infecting and colonizing the nucleate phloem progenitor cells in newly developing shoots.

Host Range and Pathogenesis

In contrast to most other geminiviruses that typically have a narrow host range, BCTV is reported to have an extremely wide host range, including over 300 species of dicotyledonous plants in 44 families, including members of the Chenopodiaceae, Compositae, Cruciferae, Leguminosae, and Solanaceae. Here, it is worth noting that the BCTV host range was established from early observations of diseased plants of undefined etiology, when it was difficult to precisely identify the causal virus(es) involved. Therefore, the reported host range of BCTV should be confirmed.

However, using agroinoculation and leafhopper transmission of infectious clones and progeny virions, respectively, relatively wide host ranges (at least among crops) were demonstrated for BCTV-Wor, BCTV-Svr, BCTV-SpCT, BCTV-Mld, BCTV-CO and other strains. Furthermore, the curly top symptoms that developed in these plants were similar to those described in these crops in the

field. BCTV-infected crop plants also had high concentrations (titers) of virus. In contrast, non-cultivated plants such as weeds typically show few or no symptoms when infected by BCTV, and such plants often have lower levels of virus. However, BCTV infects a wide variety of annual and perennial weed species and at rates as high as 20% (in the foothills of California), which serve as reservoirs of the virus, especially in leafhopper breeding areas. Moreover, even though the levels of virus are low in most infected weeds, the feeding characteristics of the beet leafhopper allows for concentration of the virus in viruliferous insects.

The products of BCTV complementary-sense ORFs C2 and C4 have been associated with pathogenesis. ORF C2 mutants induce severe symptoms and accumulate to wild-type virus levels in sugar beet and *N. benthamiana*. However, plants infected with these mutants have a greater propensity for recovery from severe symptoms compared with those infected with wild-type virus, suggesting that the C2 protein suppresses a plant defense response. Consistent with this hypothesis, transgenic *N. benthamiana* and *N. tabacum* plants constitutively expressing the C2 protein exhibit enhanced susceptibility to BCTV infection, as revealed by a shorter latent period before symptom development. However, when infected with BCTV, these transgenic plants did not develop more severe symptoms or accumulate higher levels of viral DNA compared with non-transformed control plants. The transgenic plants were also more susceptible to the begomovirus *Tomato golden mosaic virus* (TGMV) and the unrelated RNA virus *Tobacco mosaic virus*, indicating that the C2 protein affects a nonspecific host defense response. Consistent with these observations, the C2 protein has been shown to suppress post-transcriptional gene silencing (PTGS) and transcriptional gene silencing (TGS) in *N. benthamiana*. In this respect, the BCTV C2 resembles the AC2 homolog in the begomovirus TGMV. However, the two proteins share only limited homology, and the BCTV C2 protein does not appear to have transcriptional activator activity for virion-sense gene expression as has been shown for TGMV AC2. Despite this divergence, both BCTV C2 and TGMV AC2 proteins bind to, and inactivate, adenosine kinase (ADK), an enzyme required for 5'-adenosine monophosphate (AMP) synthesis. ADK activity is also reduced in BCTV-infected plants and transgenic plants expressing C2 protein. Regulation of adenosine levels by ADK plays a key role in the control of intermediates required for methylation, which is the underlying mechanism of TGS and provides a possible link between C2 protein activity and suppression of TGS targeting the viral genome. Both BCTV C2 and TGMV AC2 proteins also bind to, and inhibit, sucrose non-fermenting 1 (SNF1)-related nucleoside kinase. Reduction of SNF1 protein kinase expression in *N. benthamiana* results in an enhanced susceptibility phenotype resembling that associated with the expression of C2 and AC2 proteins, whereas SNF1 overexpression produces plants with reduced symptom severity. This activity is not directly linked to silencing suppression, suggesting that C2 protein inhibition of SNF1 activity affects a different plant defense pathway. BCTV ORF C3 mutants produce severe symptoms and accumulate to wild-type levels in *Nicotiana benthamiana* plants, but they induce only mild symptoms and accumulate low levels of viral DNA in sugar beet plants. This suggests that the BCTV RE_n may enhance viral DNA replication in a host-dependent manner.

BCTV typically induces upward curling of the leaves (Fig. 1(B-F)) and clearing, enations and swelling of veins on the abaxial side of leaves, particularly in leaves of sugar beet plants (Fig. 1(F)). Although the phloem-limited virus cannot access the apical meristem, it is possible that BCTV exploits undifferentiated cambium cells in the vascular bundles or nucleate protophloem cells in the root and shoot apices, as shown for the begomovirus *Bean dwarf mosaic virus*. BCTV infected tissues exhibit hyperplasia and hypertrophy within the phloem and adjacent parenchyma, and the abnormally dividing cells differentiate into sieve element-like cells. In older infected tissues, the affected phloem cells eventually become necrotic and collapse. Pathogenic effects occur in developing tissues only after maturation of sieve elements, consistent with the virus moving from source to sink cells with the flow of photoassimilates, either to newly developing shoots or roots, depending on where the leafhopper introduced the virus to the plant (there can be large amounts of BCTV in roots). Infected sugar beet leaves accumulate enhanced levels of sucrose, attributed to impaired transport resulting from disruption of the phloem. This is associated with a reduction in chlorophyll content, reduced activity of key photosynthetic enzymes and a concomitant reduction in the rate of photosynthesis and altered turgor pressure that causes an increase in mesophyll cell size.

Geminiviruses are masters at manipulating the plant cell cycle to produce an environment suitable for their replication and spread. This is achieved by the interaction of Rep with the plant homolog of retinoblastoma-related tumor-suppressor protein (pRBR), which relieves the constraint imposed by pRBR binding to the E2F transcription factor and allows the expression of E2F-responsive genes. BCTV infection induces cell division within vascular tissues, attributed to the action of the C4 protein. Sugar beet infected with ORF C4 mutants remain asymptomatic and *N. benthamiana* plants show an altered phenotype; neither host develops hyperplasia of the phloem or enations. Furthermore, transgenic *N. benthamiana* plants expressing C4 protein show abnormal plant development and tumorigenic growths. Consistent with this phenotype, C4 protein interacts with BIN2, a negative regulator of transcription factors in the brassinosteroid signaling pathway that controls cell division and tissue development. The reasons for the induction of cell division are not clear, particularly as the hyperplastic tissues often do not contain detectable levels of virus. However, it is likely that it does provide additional nucleate cells that can be infected, via cell division or a yet to be determined means of cell-to-cell movement, and produce virions that can be released into sieve elements for long-distance movement or acquisition by the beet leafhopper vector.

The host range of BCTV includes *Arabidopsis thaliana*, which represents an important resource for the study of virus-host interactions. Most susceptible ecotypes become stunted and develop enations on infected leaves. The Sei-O ecotype is hypersusceptible to BCTV-Svr and develops callus-like structures containing high levels of the virus, suggesting a virus-induced hormonal imbalance. Symptom development correlates with enhanced expression of the cell cycle gene *cdc2* and small auxin-upregulated RNA gene (*saur*). Hence, disruption of the phloem in infected tissues may affect auxin transport, causing localized increases that result in cell proliferation.

Transmission of BCTV by the Beet Leafhopper Vector

In nature, BCTV is transmitted, plant-to-plant, exclusively by the beet leafhopper (*C. tenellus*, Fig. 1(A)). BCTV is not transmitted by contact or mechanically (sap inoculation) and is not believed to be seed-transmitted, although seed transmission of BCTV in petunia has been reported. A search for additional insect vectors in the NW was unsuccessful, which indicated a high degree of specificity in the association of the curly top agent and the beet leafhopper. However, other species of *Circulifer* from the OW can vector BCTV, and *C. opacipennis* was reported to vector the curly top agent in the Mediterranean Basin. There is no evidence for strain specificity in transmission of BCTV by *C. tenellus*, and all strains tested to date have been leafhopper-transmissible. The mode of transmission is persistent circulative, and there is no evidence that BCTV replicates in the insect vector. The beet leafhopper is an efficient vector of BCTV and rapidly acquires the virus (in minutes) from the phloem of infected plants. Being a mobile, polyphagous phloem-feeding insect makes the beet leafhopper an ideal vector for BCTV. The leafhopper also prefers to feed from veins on the abaxial side of young- to medium-aged leaves, including the swollen veins of BCTV-infected plants (Fig. 1(F)), in which there are high concentrations of virions. It is not known if beet leafhoppers are preferentially attracted to plants with curly top symptoms, as has been reported for whiteflies (*Bemisia tabaci* species complex) and *Tomato yellow leaf curl virus* (TYLCV).

The pathway BCTV follows in the beet leafhopper was investigated with a PCR-based approach. BCTV-Wor [US-Mld-Wor4] was detected in the digestive tract of the insect after an acquisition access period (AAP) of 1 h, in the hemolymph after 3 h, and in the salivary glands after 4 h, from where it is reintroduced into plants with saliva during feeding. This is consistent with earlier reports estimating the minimum time between virus acquisition and the leafhopper becoming viruliferous (the latent period) to be 4 h. The feeding time necessary for transmission of the virus can be as short as 1 min, though longer feeding times (e.g., inoculation access periods of 24–48 h) provide higher rates of infection. Depending on the length of the AAP, leafhoppers can remain viruliferous for life, although the amount of virus retained and the ability of leafhoppers to transmit the virus declines with time when viruliferous leafhoppers are maintained on plants that are not hosts for the virus. This indicates that the virus is not replicating in the leafhopper. There is also no evidence for transovarial transmission of the virus, a common feature of viruses that replicate in their insect vector.

Several lines of evidence demonstrate that the CP is the determinant of leafhopper transmission specificity. First, BCTV-Wor [US-Mld-Wor4] CP mutants that failed to produce virions are not transmitted by *C. tenellus*. Second, a recombinant virus in which the CP coding sequence of the whitefly-transmitted begomovirus *African cassava mosaic virus* was replaced with that of BCTV produced virions that were transmissible by the beet leafhopper when injected into the hemocoel. Finally, the autonomously replicating nanovirus-like DNA-1 component, normally associated with the whitefly-transmitted begomovirus *Ageratum yellow vein virus*, can be maintained in sugar beet plants infected with BCTV, and is transmitted by the beet leafhopper. These results indicated that the ~1.4 kb DNA-1 molecule is encapsidated by the BCTV CP, although the mechanism is not known, i.e., are two ~1.4 kb ssDNAs or a single ~2.8 kb dimeric ssDNA form (a product of rolling circle replication) encapsidated.

The BCTV-beet leafhopper interaction was further investigated by immunofluorescence and immunolocalization, with antisera raised against *E. coli*-expressed BCTV-Wor [US-Mld-Wor4] CP. BCTV was detected in the mid- and hindgut of viruliferous beet leafhoppers, and distinct fluorescent punctate structures, far greater than the size of an individual virion, were observed on or in the gut walls. These may represent large aggregates of virions (possibly inclusion bodies) acquired from the phloem during feeding or some form of cooperative binding of virions during feeding to form aggregates following initial binding of a small number of virions to the gut wall. Immunolocalization studies revealed vesicle-like structures on the inside of the gut wall, possibly involved in endocytosis of BCTV virions. In the salivary glands, BCTV was mostly localized to the primary salivary gland. The BCTV-Wor alanine scanning mutant CP 25–28 formed virions, but was not leafhopper-transmitted. Dissection PCR and immunofluorescence and immunolocalization revealed that this mutant interacted with the gut in a similar fashion to wild-type virus and entered into the hemolymph, but was not detected in the salivary glands. Thus, this may reveal an amino acid motif of the CP that interacts with a receptor on the salivary gland surface. Together, these results suggest that surface features of the BCTV virion interact in a highly specific manner during circulative transmission through the beet leafhopper, possibly involving two receptor-mediated endocytosis events, i.e., during virus acquisition across the gut wall and movement into the salivary glands. An effort is ongoing to sequence the genome of the beet leafhopper with the goal of identifying the receptors involved in the transmission of BCTV and the microbes tomato big bud phytoplasma and citrus stubborn spiroplasma.

Experimental Systems for Infection of Plants by BCTV

BCTV is poorly (at best) transmitted by mechanical inoculation, which reflects the phloem-limited life style of the virus. Infection can be achieved using a fine-gauge needle (pin-pricking) to introduce virions or viral DNA inoculum directly into the phloem, but the efficiency is low. Particle bombardment is an efficient method for infection of the laboratory host *N. benthamiana* with BCTV, but very inefficient for infection of other hosts. The reason for the difference in host susceptibility to particle bombardment inoculation is not known, and BCTV is phloem-limited in *N. benthamiana*. It is possible that phloem cells susceptible to BCTV infection are more accessible in the leaves of the young (two to four leaf stage) *N. benthamiana* plants inoculated by particle bombardment.

In contrast, high rates of BCTV infection in numerous crops (common bean, cucurbits, pepper, sugar beet and tomato) can be achieved via agroinoculation, in which a multimeric infectious clone is inserted in between the T-DNA borders of a binary

plasmid. This plasmid is transformed into a disarmed strain of *Agrobacterium tumefaciens*, and bacterial cells carrying the binary plasmid are inoculated into stems of young plants just beneath the shoot apex by the needle-puncture method. The *A. tumefaciens* cells bind to wounded plant cells, presumably including nucleate progenitor cells of the phloem, and deliver the T-DNA containing the multimeric BCTV DNA into the nucleus. In the nucleus, the genomic BCTV ssDNA is released from the integrated T-DNA (in transformed cells) by the expressed Rep, which recognizes the IR sequences. The viral infection then proceeds as if the leafhopper had delivered virions into the phloem. Agroinoculation has proven to be a rapid and efficient method of inoculation for BCTV and other geminiviruses, and can be used for screening tomatoes and other crops for resistance to BCTV. This eliminates the need to rely on field plots and natural inoculum, the occurrence of which is highly unpredictable. Agroinoculation has limitations, including difficulties in generating agroinoculation systems for some strains (e.g., BCTV-CO), and the recalcitrance of some plants to *Agrobacterium*-mediated transformation, e.g., pepper.

Management of Curly Top Disease Caused by BCTV

Like many vector-borne diseases, management of curly top disease caused by BCTV has been challenging due to multiple factors. These include the sporadic nature of curly top outbreaks, difficulty in predicting 'bad' curly top years, the migratory nature of the vector and difficulty in managing leafhopper populations with insecticides. In years having severe outbreaks, plants are typically infected at an early stage of development, resulting in substantial yield losses. Resistance to curly top disease is not available in commercial varieties of a number of susceptible crops, including pepper and tomato. Therefore, differences in (1) crop production, (2) the biology of curly top disease in different crops, and (3) the availability of resistant varieties means that management approaches for different crops may vary. However, the most effective curly top disease management will be accomplished through an integrated pest management (IPM) approach. Below, some general strategies for curly top disease management and examples of IPM programs for sugar beet and tomato crops are presented.

Resistant Varieties

The planting of resistant varieties is one of the most effective and environmentally friendly management strategies for any disease. Sources of resistance to BCTV were identified in germplasm of common bean, flax, squash and sugar beet. In the case of sugar beet, mass selection of plants from fields with high incidences of curly top disease led to the identification and release of the first curly top resistant sugar beet variety, US1, in 1933. This variety became widely grown in the western USA, despite some horticultural shortcomings. The identification and selection of resistant sugar beet varieties has been facilitated by the use of breeding plots, known as curly top nurseries, in which large numbers of leafhoppers viruliferous for the most prevalent and virulent strains of BCTV are released. As a result of the high selection pressure in these nurseries, combined with the development of commercial hybrids and molecular breeding technologies, curly top resistance in sugar beet has advanced tremendously despite being a complex quantitatively inherited trait. Modern curly top resistant varieties have greater uniformity, higher yield and improved resistance. However, plants of these resistant sugar beet varieties will develop curly top symptoms and may suffer yield losses when infected at an early stage of development (two to four leaf stage). Therefore, these resistant varieties need to be part of an IPM program that addresses the susceptibility of young plants.

Curly top resistance was also found in a range of common bean varieties (e.g., Red Mexican type), and genetic studies identified a single dominant gene (*Bct-1*) associated with curly top resistance. Breeding efforts led to the release of the commercial curly top-resistant varieties Great Northern UI 15 and Red Mexican UI 34. Subsequently, snap bean varieties with curly top resistance were developed and released. These varieties have provided acceptable yields under high curly top pressure. There is a need to develop curly top-resistant varieties for other market classes of common bean. In terms of tomato and pepper, curly top resistance was not found in commercial cultivars. High levels of resistance were found in accessions of some wild species of tomato, including *Solanum chilense*, *S. habrochaites* and *S. peruvianum*. Unfortunately, introgression of this resistance into commercial tomato varieties has been difficult. Agroinoculation screening of tomato breeding lines with different combinations of *Ty* genes, which confer resistance to the begomovirus TYLCV, with BCTV-Svr and BCTV-LH71 revealed gene combinations, e.g., *Ty-1*, *Ty-2*, *Ty-3* and *Ty-5*, that provided high levels of curly top resistance. The tomato breeding Line 20 #12, carrying the *Ty-1*, *Ty-2* and *Ty-3* genes, has moderate resistance to curly top, which involves initial symptom development followed by recovery. This may indicate that TGS is targeting the viral genome for methylation.

Cultural Practices

Because young plants are more susceptible to infection and suffer the greatest yield losses, some degree of protection can be achieved by adjusting planting dates to avoid exposure of young plants to flights of viruliferous leafhoppers. Typically these flights occur in mid- to late-spring. Thus, early or late planting dates may allow for avoidance of peak leafhopper flights when plants are young, but the difficulty in predicting leafhopper flights from year-to-year is a challenge and populations vary considerably depending on weather conditions and the composition of host plants in overwinter areas. Avoiding the establishment of susceptible crops near leafhopper breeding areas is another strategy that can reduce curly top incidence, though this is limited by land

availability and options for planting crops that are not susceptible to curly top disease (e.g., garlic, melons, onions or small grains). Finally, over-seeding of tomatoes can reduce losses due to curly top disease by allowing for uninfected plants to compensate for yield losses due to infected plants, and to possibly make fields less attractive to migrating leafhoppers, which cue in on patches of brown and green.

Management of the Leafhopper Vector With Insecticides

The most common management strategy is to reduce populations of the beet leafhopper vector in order to reduce the incidence and spread of BCITV. This involves spraying contact insecticides such as malathion or dimethoate and delivering systemic insecticides such as neonicotinoids via sprays, soil drenches or drip irrigation. More recently, technologies have been developed for coating seeds with systemic insecticides. Spraying of contact insecticides can be carried out on weeds in leafhopper breeding grounds or directly on crops such as beans, sugar beets and tomatoes. Ideally, sprays are applied based on monitoring for leafhopper populations with sticky cards or sweep nets. In California, it has been shown that high incidences of curly top in tomato (>10%) require high populations of viruliferous leafhoppers in the spring. This information can inform growers if sprays are necessary. The California Department of Food and Agriculture runs the Curly Top Virus Control Program, the only remaining program that sprays breeding grounds of beet leafhoppers. In this grower-funded program, breeding areas in the rangelands of the foothills on the western edge of the Central Valley are surveyed with sweep nets for beet leafhoppers and malathion is applied when thresholds are reached. The efficacy of this program in reducing incidence of curly top disease in crops remains unclear. Parasites and predators of the beet leafhopper are not uncommon, but it has proven difficult to assess their sustainability or impact on leafhopper populations under natural conditions.

Sanitation

It is not possible to eradicate weed reservoirs in the vast areas that comprise the breeding areas for the beet leafhopper in the western USA. However, in the agricultural valleys where curly top disease is a problem, it is important to practice effective weed management in fields and in non-cultivated areas, as well as around canals and other bodies of water, to eliminate preferred hosts for reproduction of beet leafhoppers, such as Russian thistle (*Salsola sp.*). This can reduce beet leafhopper populations in areas near susceptible crops, as well as those that migrate to the foothills in the fall.

IPM for Curly Top Disease

IPM packages for management of curly top in sugar beet (a preferred host of the beet leafhopper on which it feeds and reproduces) and tomato (a non-preferred host that the leafhopper tastes and then moves on) are different. For sugar beet the IPM package frequently includes: (1) planting cultivars with curly top resistance; (2) early planting dates (to allow plants to reach six to eight true leaves before viruliferous leafhoppers arrive in the fields); (3) management of beet leafhopper populations in fields with insecticides based on an IPM approach, e.g., seed treatment with systemic insecticides followed by mid-season application of a contact insecticide; and (4) sanitation of harvested fields and of weeds on a regional basis. Tomato is not a preferred host of the beet leafhopper, and there are no commercially available resistant varieties. Thus, the IPM package may involve (1) modifying planting dates or the location of fields to avoid leafhopper flights and proximity to leafhopper breeding grounds or flight paths; (2) drenching tomato transplants in the greenhouse and at transplanting with a systemic insecticide; (3) management of beet leafhopper populations in fields with insecticides based on an IPM approach; and (4) sanitation of harvested fields and regional weed management.

Further Reading

- Bennett, C.W., 1971. American Phytopathology Society Monograph No. 7: The Curly Top Disease of Sugar Beet and Other Plants. St. Paul, MN: American Phytopathology Society.
- Chen, L.-F., Brannigan, K., Clark, R., Gilbertson, R.L., 2011. Characterization of curtoviruses associated with curly top disease of tomato in California and monitoring for these viruses in beet leafhoppers. *Plant Disease* 94, 99–108.
- Chen, L.-F., Gilbertson, R.L., 2016. Transmission of curtoviruses (Beet curly top virus) by the beet leafhopper (*Circulifer tenellus*). In: Brown, J.K. (Ed.), *Vector-Mediated Transmission of Plant Pathogens*. St. Paul, MN: American Phytopathological Society, pp. 243–262.
- Creamer, R., Luque-Williams, M., Howo, M., 1996. Epidemiology and incidence of beet curly top geminivirus in naturally infected weed hosts. *Plant Disease* 80, 533–535.
- Esau, K., Hoefert, L.L., 1978. Hyperplastic phloem in sugarbeet leaves infected with the beet curly top virus. *American Journal of Botany* 65, 772–783.
- Hanley-Bowdoin, L., Settledge, S.B., Orozco, B.M., Nagar, S., Robertson, D., 1999. Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Plant Sciences* 18, 71–106.
- Rojas, M.R., Macedo, M.A., Maliano, M.R., *et al.*, 2018. World management of geminiviruses. *Annual Review of Phytopathology* 56, 637–677.
- Soto, M.J., Chen, L.-F., Seo, Y.-S., Gilbertson, R.L., 2005. Identification of regions of the Beet mild curly top virus (family *Geminiviridae*) capsid protein involved in systemic infection, virion formation and leafhopper transmission. *Virology* 34, 257–270.
- Stanley, J., Markham, P.G., Callis, R.J., Pinner, M.S., 1986. The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO Journal* 5, 1761–1767.
- Stenger, D.C., 1998. Replication specificity elements of the Worland strain of beet curly top virus are compatible with those of the CFH strain but not those of the Cal/Logan strain. *Phytopathology* 88, 1174–1178.

- Strausbaugh, C.A., Eujayl, I.A., Wintermantel, W.M., 2017. Beet curly top virus strains associated with sugar beet in Idaho, Oregon, and a Western U.S. Collection. *Plant Disease* 101, 1373–1382.
- Varsani, A., Martin, D.P., Navas-Castillo, J., *et al.*, 2014. Revisiting the classification of curtoviruses based on genome-wide pairwise identity. *Archives of Virology* 159, 1873–1882.

Beet Necrotic Yellow Vein Virus (*Benyviridae*)

Sebastian Liebe, Institute of Sugar Beet Research, Göttingen, Germany

Annette Niehl and Renate Koenig, Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

Mark Varrelmann, Institute of Sugar Beet Research, Göttingen, Germany

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

+ ssRNA Positive-sense single-stranded RNA
ARF Auxin response factors
Aux/IAA Auxin/indole-3-acetic acid
BdMV Burdock mottle virus
BNYVV Beet necrotic yellow vein virus
BSBMV Beet soil-borne mosaic virus
CP Coat protein
CP-RT Read-through minor coat protein
ELISA Enzyme-linked immunosorbent assay
ER Endoplasmic reticulum
ncRNA Noncoding RNA

NES Nucleolar export signal
NLS Nucleolar localization signal
ORF Open reading frame
RdRp RNA-dependent RNA polymerase
RSNV Rice stripe necrosis virus
SCF Skp1-Cullin-F-box
siRNAs Small-interfering RNAs
TGB Triple gene block
TM Transmembrane domain
VSR Viral suppressor of gene silencing
Znf Zinc-finger domain

Glossary

Benyvirus Siglum derived from *Beet necrotic yellow vein virus*.

Rhizomania Root beardedness (root madness): Extensive proliferation of secondary rootlets at the expense of the main tap root.

Classification

Beet necrotic yellow vein virus (BNYVV) is the type species of the genus *Benyvirus*. This genus is the only one in the new family *Benyviridae* which comprises viruses with multipartite genomes encapsidated in rigid rod-shaped particles. The closest known relative is the *Beet soil-borne mosaic virus* (BSBMV). More distant relationships exist with *Burdock mottle virus* (BdMV) and *Rice stripe necrosis virus* (RSNV). Morphologically similar viruses are found in the family *Virgaviridae*. Similarities and dissimilarities between the viruses in the two families have been summarized in the 10th Report of the International Committee on Taxonomy of Viruses (ICTV).

Virion Structure

BNYVV virus particles are non-enveloped and rod shaped (Fig. 1). The particles have a diameter of about 20 nm and the length varies between 65 and 390 nm depending on the encapsidated RNA species. Deviations from the upper and the lower limits may be due to end-to-end aggregation or breakage of particles. The particles have a helical structure with a central canal. The pitch of the helix is estimated to be 2.6 nm. This helix has an axial repeat of four turns, involving 49 protein subunits of the 21 kDa major coat protein (CP). Each coat protein subunit occupies four nucleotides on the RNAs of BNYVV. Immunogold labeling revealed that the 75 kDa read-through minor coat protein (CP-RT) is predominantly localized close to one end of the virus particles. It had been believed that the CP-RT is necessary to initiate the encapsidation process, but recent investigations showed that it is dispensable for particle assembly and successful virus infection at least in the experimental host *Nicotiana benthamiana*.

Genome Organization and Functions of Gene Products

The BNYVV genome consists of four to five positive-sense single-stranded RNAs (+ ssRNAs) (Fig. 2). Each RNA is capped at the 5' end and possesses a 3' poly(A) tail. RNA1 (6.7 kb) contains one open reading frame (ORF) encoding a replicase protein of 237 kDa (P237), which is processed by a papain-like proteinase into one 150 kDa protein with methyltransferase, helicase and protease domains and one 66 kDa protein containing the RNA-dependent RNA polymerase. Localization studies revealed an association of the P237 protein as well as domains of the 150 kDa and 66 kDa replication-associated proteins with the endoplasmic reticulum (ER). The first ORF on RNA2 (4.6 kb) encodes the CP which is terminated by a leaky UAG stop codon allowing the translation of the readthrough protein CP-RT (P75). The C-terminal part of the CP-RT contains a KTER motif at position 553–556 which is essential for transmission by the vector *P. betae*. Several rounds of mechanical transmission can cause deletions within this region leading to a loss

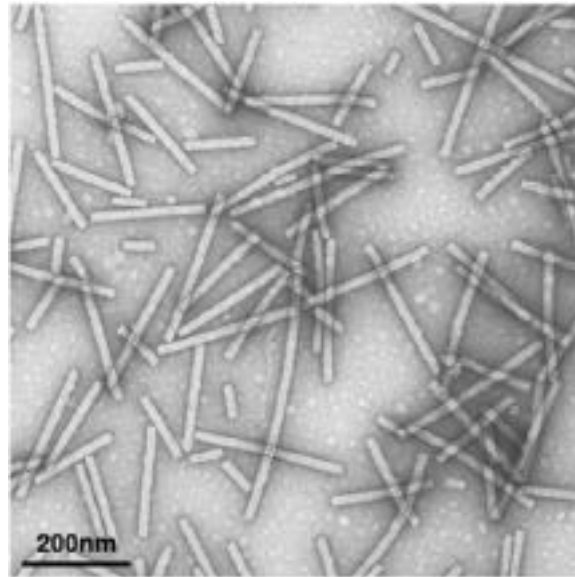


Fig. 1 Particles of Beet necrotic yellow vein virus in a purified preparation negatively stained with uranyl acetate. Courtesy D. E. Lesemann and J. Engelmann, BBA, Braunschweig, Germany.

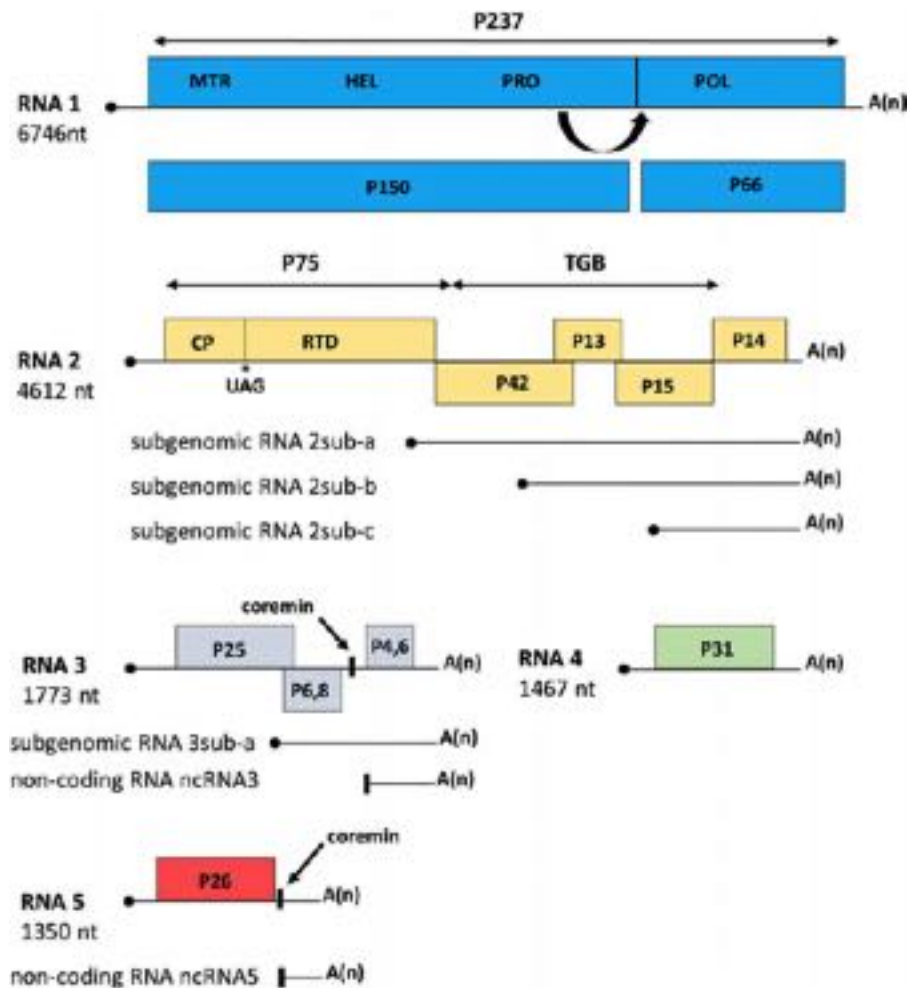


Fig. 2 Organization of the BNYVV genome and expression of BNYVV genes. Self-cleavage of the replicase protein (P237) is indicated by a black arrow. At 5' end, the m⁷Gppp is indicated by a black circle and the poly(A) tail at the 3' end by A(n). The suppressible leaky stop codon UAG is indicated by *. Abbreviations: methyltransferase (MTR); helicase (HEL); protease (PRO); RNA polymerase (POL); coat protein (CP); readthrough domain (RTD); triple gene block (TGB).

of vector transmissibility. It was shown that a mitochondrial targeting sequence within the CP-RT targets modified CP-RT expressing GFP and reporter proteins to the cytoplasmic face of mitochondria. Furthermore, four trans-membrane domains (TM1-TM4) have been identified by bioinformatic analysis, but their function is not clarified yet. The TM helices are apparently tightly packaged with ridge/groove arrangements between the helices and strong electrostatic association. It is supposed that these helices act as anchors to the mitochondrial membrane in plant cells, but might also facilitate the movement of virus particles across the fungal membrane separating the plasmodium from the cytoplasm. The next three ORFs on RNA2 encode the triple gene block (TGB) proteins 1–3 that are translated from subgenomic RNAs and responsible for cell-to-cell movement. TGB1 (P42) is derived from subgenomic RNA2a whereas TGB2 (P13) and TGB3 (P15) are translated from the bicistronic subgenomic RNA2b. P42 possesses an N-terminal nucleic acid-binding activity and helicase motifs at the C-terminus. Electron microscopy revealed that P13 and P15 localize to callose-containing cell wall thickenings which suggests plasmodesmal targeting by these two proteins. It is proposed that P13 and P15 provide a docking site for a P42-viral RNA complex which modifies plasmodesmal size exclusion limit to improve transit of the viral RNA. The last ORF on RNA2 is translated from the subgenomic RNA2c and encodes the viral suppressor of gene silencing (VSR). It is a 14 kDa cysteine-rich protein (P14) which inhibits the production of small-interfering RNAs (siRNAs), especially secondary siRNAs by interfering with RDR6-dependent siRNA generation. A zinc-finger domain (Znf) containing an embedded nucleolar localization sequence (NLS) has been shown to be crucial for silencing-suppression activity and protein stability. The protein accumulates in the nucleolus and in the cytoplasm, but its suppression activity is probably independent of the subcellular localization. Long-distance movement is abolished in viral mutants carrying a defective VSR.

The first two RNAs of BNYVV are sufficient for virus replication, systemic movement, and symptom development in the experimental host *N. benthamiana*, but characteristic disease symptoms in sugar beet are only induced when RNA3 (1.8 kb) is present. Thus, RNA3 carries the pathogenicity factor of BNYVV, encoded by the 25 kDa protein (P25). P25 possesses a nuclear localization signal (NLS) and nuclear export signal (NES) allowing the protein to shuttle between the nucleus and cytoplasm. The importance of this protein for viral pathogenicity is described in the section 'Virus-Host Relationships'. Downstream of P25 is another ORF located encoding for a 6.8 kDa protein called N gene, but its expression could not be detected under natural infection conditions. However, the N gene is translationally activated by the deletion of upstream sequences leading to a truncated RNA. Expression of this RNA induces necrotic lesions in the local lesion host *Tetragonia expansa*. The last ORF on RNA3 encodes a 4.6 kDa protein of unknown function that is presumably expressed from a subgenomic RNA (RNA-3 sub). Its expression has never been confirmed supporting the presumption that the ORF is cryptic. Besides encoding proteins, the RNA3 additionally has a non-coding function that is crucial for systemic movement. A 'core' nucleotide sequence (at nt position 1033–1257) was identified that mediates systemic movement independent of P25 expression in *Beta macrocarpa*. Within this region, a 20 nt long sequence stretch ('coremin') is crucial for systemic movement. In a study to decipher the function of the 'coremin' nucleotide sequence, a non-coding RNA3 (ncRNA3) was identified, which is a cleavage product that accumulates only when the coremin sequence is present. Production of the ncRNA3 appears to depend on XRN1-type, i.e., plant XRN4 exo-ribonucleases. It has been shown that the ncRNA3 acts synergistically to P14 as a weak second VSR and promotes efficient systemic host infection. The 1.4 kb RNA4 encodes a single 31 kDa protein (P31), which is essential for virus transmission by the vector *P. betae*. Some isolates of BNYVV (see section 'Epidemiology') possess an RNA5 (1.3 kb), containing a single ORF which codes for a 26 kDa protein (P26). The presence of this protein induces necrotic lesions in the host *Chenopodium quinoa* when it is expressed either from RNA5 or from an RNA3-derived replicon. The nucleo-cytoplasmic protein displays a strong transcriptional activity. Interestingly, the 20 nt coremin sequence is also present on RNA5 outside the ORF of P26, indicating a non-coding function of RNA5 during virus movement similar to RNA3.

Transmission

BNYVV is transmitted by the soil-borne ubiquitous plasmodiophorid *Polymyxa betae*, which infects sugar beet to complete its life cycle. *P. betae* can survive in the soil as sporosorus (cytosori), which are resting spores with a very thick and strong cell wall providing persistence in soil for many years under unfavorable conditions. When a host plant is present and sufficient soil moisture available, resting spores germinate and primary zoospores are released. Motile zoospores move through free water in the soil and start encysting on root hairs or epidermal cells of roots. Zoospores penetrate host cells and inject their cytoplasm. A multinucleate plasmodium is formed within the host cell and differentiates either into a zoosporangium releasing secondary zoospores (sporangial phase) or into a sporosorus forming resting spores (sporogonic phase). The secondary zoospores penetrate new root cells and induce plasmodium development. It is suggested that *P. betae* transmits BNYVV when zoospore cytoplasm is injected into the host cell. Likewise, the virus is acquired when a plasmodium develops within an infected host cell. Secondary zoospores released from this plasmodium will be viruliferous. Similarly, sporosorus developed within infected cells will likely release viruliferous primary zoospores. Due to the vector biology, BNYVV can be disseminated by soil adhering to agricultural machines, by irrigation or even by wind. All viral proteins of BNYVV were detected in resting spores and zoospores, however, whether BNYVV is able to replicate within its vector remains to be demonstrated.

Epidemiology

The host range of BNYVV is limited to certain members of the plant families *Amaranthaceae*, *Caryophyllaceae*, *Chenopodiaceae*, and *Tetragoniaceae*. The major natural hosts are *Beta vulgaris* (sugar beet) and *B. macrocarpa*, which are infected and systemically invaded

only when RNA1–4 are present. Spinach (*Spinacia oleracea*) was also identified as natural host, but is only of minor importance. The experimental host *N. benthamiana* is systemically infected in the absence of RNA3 and RNA4. Local lesion hosts for virus propagation are *C. quinoa* and *T. expansa*. Infection is restricted to small chlorotic or necrotic spots without systemic movement after mechanical inoculation. BNYVV was first described in 1950 in Italy and then spread rapidly into sugar beet growing areas worldwide. Restriction fragment length polymorphism, single-strand conformation polymorphism and phylogenetic analysis allowed the identification of four major virus types, namely the A-, B-, P-, and J-type. These different types are serologically indistinguishable, but their RNAs show sequence variability ranging between 1%–5%. The virus types are geographically differently distributed. The A-type is most prevalent with many reports in Europe, USA, and Asia. The B-type has been reported in Germany, France, and other central European countries. The P-type was first described in France in the area of Pithiviers (Center-Val de Loire) and was later discovered also in the UK, Kazakhstan, and a small area in Germany. A great diversity of BNYVV types can be found especially in the Pithiviers area. Within this region, the A-, B-, and P-type are present and occur frequently in mixed infections. Moreover, P-type RNA5 was found in plants infected only with A- or B-type RNAs 1–4 indicating the formation of reassortants. Although RNA5 increases viral aggressiveness, P-type isolates without RNA5 have been reported, indicating its dispensability for viral pathogenicity. The J-type was identified in Asia and has been distinguished from the P-type by sequence analysis. The evolutionary history of BNYVV has not yet been completely resolved, but several subpopulations are thought to have diverged from at least four original types of populations that developed in native hosts in East Asia long before sugar beet was cultivated. Afterwards, BNYVV populations from different origins were introduced into sugar beet cultivation areas and spread extensively. Sequence analysis indicates that the A- and P-type are more closely related than the A- and B-type. Therefore, it is suggested that the A- and B-type became separated before the P-type evolved from the A-type.

Disease Symptoms

BNYVV causes rhizomania. The term signifies root madness (Rhizo: root; Mania: madness). The disease is characterized by excessive lateral root formation ('root beardedness') (Fig. 3). Infected sugar beet plants can display also to some extent dwarfism with a taproot which is reduced in size, has a wine-glass like shape and low sugar content. A brownish discoloration of the vascular

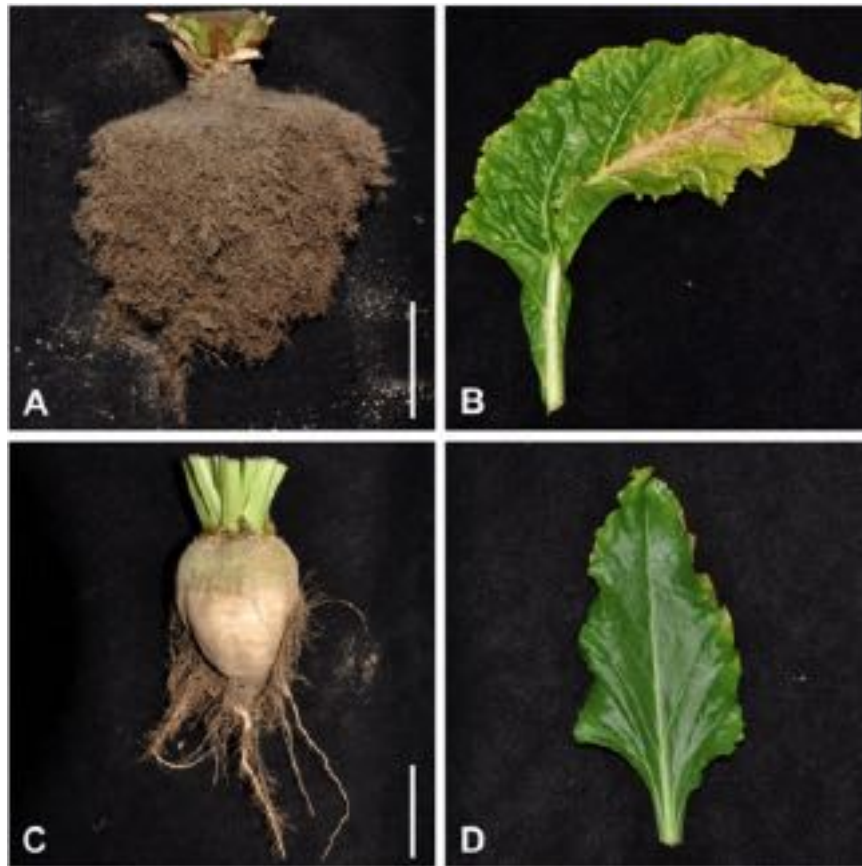


Fig. 3 Characteristic rhizomania disease symptoms on taproot (A) and leaf (B) from a BNYVV infected sugar beet plant compared to a healthy plant (C and D). The white bar indicates the scale (5 cm).

system is often seen. Leaves may become pale and have an upright position. Although symptoms are mostly confined to the root system, infected plants can display vein yellowing and necrosis on leaves after systemic movement of BNYVV to upper plant parts. The name BNYVV refers to these symptoms. Yield losses associated with disease occurrence can range between 50%–60% and are mainly due to reduced root weight and sugar content.

Virus–host Relationships

The pathogenicity of BNYVV is strongly linked to the activity of P25. The protein possesses a nucleo-cytoplasmic shuttling activity allowing the virus to manipulate many cellular processes. Upon BNYVV infection, plants display massive lateral root proliferation leading to the characteristic symptom of a “root beard” only in the presence of P25. Auxin as the major plant hormone controls an array of developmental processes including the development of lateral roots. Microscopic studies of sugar beet roots revealed that differentiated pericycle cells are converted into meristematic cells during BNYVV infection. Additionally, infected plants show elevated auxin concentration in the roots substantiating the hypothesis that BNYVV interacts with the auxin signaling pathway to induce lateral root proliferation. Transgenic *Arabidopsis thaliana* plants expressing P25 displayed also excessive lateral root formation and elevated auxin concentration, which strongly supports the role of P25 in viral pathogenicity. Recent investigations identified an auxin/indole-3-acetic acid (Aux/IAA) transcription factor in sugar beet that interacts with the pathogenicity factor P25 of BNYVV. Aux/IAA proteins are key regulators of the auxin signaling pathway in plants. They act as transcriptional repressors by suppressing the activity of auxin response factors (ARF). ARFs are transcription factors that regulate the expression of auxin-modulated genes responsible for lateral root formation. At low auxin concentration, the transcriptional activity of ARFs is suppressed by Aux/IAA proteins, whereas high auxin concentration promotes the degradation of Aux/IAA proteins, hence releasing ARFs. Aux/IAA proteins are nucleus-localized, whereas P25 shuttles between the nucleus and the cytoplasm. It has been shown that the nuclear localization of a specific Aux/IAA protein (AUX28) is disrupted by P25 leading to an accumulation of AUX28 within the cytoplasm. It is suggested that this interaction leads to the activation of auxin-modulated genes responsible for lateral root development. This hypothesis is further supported by the observation that the expression of genes downstream of Aux/IAA proteins involved in lateral root development are highly induced upon BNYVV infection. P25 is also responsible for the development of symptoms on leaves including vein yellowing and necrosis. Furthermore, P25 interacts with an F-box family protein that is part of the Skp1-Cullin-F-box (SCF) complex and mediates protein degradation. F-box proteins determine the selectivity of the SCF complex as they recruit target proteins. The exact function of the interaction of P25 with the F-box protein is yet unknown. Interestingly, a transcription activation domain could be mapped within the amino acid residues 103 and 160 of P25. This suggests that P25 can directly activate or suppress gene expression in sugar beet; however, target genes are still to be identified. Apart from its function in virus pathogenicity, P25 is also involved in the defense response of resistant cultivars (see section ‘Prevention’).

Diagnosis

Although BNYVV infection induces very characteristic symptoms, the occurrence of leaf symptoms is rare in the field making visual disease diagnosis very difficult. Therefore, serological and molecular detection tools are required for BNYVV diagnosis. Enzyme-linked immunosorbent assay (ELISA) is a common and reliable method to detect the virus within plant parts most likely to be infected (e.g., lateral roots), but BNYVV types are serologically undistinguishable. PCR techniques such reverse transcription-PCR and real-time PCR are also useful for detection using BNYVV specific primers. In addition, sequencing of (cloned) PCR products or next generation sequencing after RNA/dsRNA/virion extraction allows the differentiation of the different BNYVV types and identification of mutations associated with resistance breaking. Infestation of soil can be detected by means of bait plants which are tested for the presence of BNYVV with the aforementioned methods.

Prevention

Methods for controlling the vector *P. betae* by means of pesticides are currently not available. Although soil fumigation with pesticides can reduce the vector population in the soil, severe environmental damage and high economic costs prohibit an application. Resistance traits against the vector were found in *B. patellaris* and *B. procumbens*, but introduction of these traits into commercial sugar beet varieties failed. Therefore, growing of virus resistant cultivars is the only strategy to control rhizomania disease. In 1983, the Holly Sugar Company identified the first resistance gene effective against BNYVV (‘Holly’ resistance). Resistance is based on a single dominant gene (*Rz1*) located on chromosome III. Upon BNYVV infection, *Rz1* confers partial resistance by reducing virus replication. Infected plants do not show disease symptoms or yield loss but the virus can be still detected in lateral roots in low quantities. Most commercial varieties today contain the *Rz1* gene. Further analysis of wild-beet germplasm allowed the identification of a second resistance gene (*Rz2*) in *B. vulgaris* subsp. *maritima* WB42 originating from Denmark, which encodes a CC-NBS LRR (coiled-coil nucleotide binding site, leucine rich-repeat) resistance protein. The gene is also located on chromosome III, but *Rz2* appears to be based on a different mechanism than *Rz1*. The *Rz2* gene has also been

introduced into commercial varieties in combination with *Rz2*. Further resistance genes (*Rz3-Rz5*) located on chromosome III could be identified, but have not been introduced into commercial varieties yet.

With the introduction of the *Rz1* resistance gene into commercial varieties, the virus population was under a high selection pressure. Thus, reports about the emergence of resistance breaking strains in Europe and the USA appeared soon after the introduction of *Rz1*. Such strains can replicate to a higher level and induce disease symptoms in *Rz1* resistant plants, but the resistance breaking mutations are probably associated with a certain fitness loss in susceptible cultivars. Extensive sequence analysis of viral isolates revealed high sequence variability within a stretch of four amino acids (AS67–70) in the pathogenicity factor P25. Mutation analysis with infectious cDNA clones of RNA3 showed that a specific single amino acid substitution on position 67 already mediates *Rz1* resistance breaking. However, the resistance breaking ability for most of the reported sequence variants at AS67–70 is unknown and needs to be determined. Moreover, besides AS67–70, further mutations in P25 are presumably involved in resistance breaking. Interestingly, resistance breaking has so far been reported only for the BNYVV A- and P-type but not for the B-type. The P-type displays a low variability of AS67–70, but is still able to overcome the *Rz1* resistance. BNYVV P-type contains an additional RNA5 encoding the pathogenicity factor P26. It is believed that resistance breaking is mediated by RNA5 rather than by mutations in P25 at AS67–70. Hence, BNYVV has probably developed different strategies to overcome *Rz1* resistance.

See also: Benyviruses (*Benyviridae*)

Further Reading

- Bornemann, K., Varrelmann, M., 2013. Effect of sugar beet genotype on the Beet necrotic yellow vein virus P25 pathogenicity factor and evidence for a fitness penalty in resistance-breaking strains. *Molecular Plant Pathology* 14 (4), 356–364.
- Chiba, S., Kondo, H., Miyanishi, M., *et al.*, 2011. The evolutionary history of Beet necrotic yellow vein virus deduced from genetic variation, geographical origin and spread, and the breaking of host resistance. *Molecular Plant-Microbe Interactions* 24 (2), 207–218.
- Flobinus, A., Chevigny, N., Charley, P., *et al.*, 2018. Beet necrotic yellow vein virus non-coding RNA production depends on a 5' → 3' XRN exoribonuclease activity. *Viruses* 10 (3), 137.
- Flobinus, A., Hleibieh, K., Klein, E., *et al.*, 2016. A viral noncoding RNA complements a weakened viral RNA silencing suppressor and promotes efficient systemic host infection. *Viruses* 8 (10), 272.
- Galein, Y., Legrève, A., Bragard, C., 2018. Long term management of rhizomania disease – Insight into the changes of the Beet necrotic yellow vein virus RNA-3 observed under resistant and non-resistant sugar beet fields. *Frontiers in Plant Science* 9, 795.
- Gil, J.F., Liebe, S., Thiel, H., *et al.*, 2018. Massive up-regulation of LBD transcription factors and EXPANSINs highlights the regulatory programs of rhizomania disease. *Molecular Plant Pathology* 19 (10), 2333–2348.
- Koenig, R., Loss, S., Specht, J., *et al.*, 2009. A single U/C nucleotide substitution changing alanine to valine in the Beet necrotic yellow vein virus p25 protein promotes increased virus accumulation in roots of mechanically inoculated, partially resistant sugar beet seedlings. *Journal of General Virology* 90 (Pt 3), 759–763.
- Laufer, M., Mohammad, H., Christ, D.S., *et al.*, 2018. Fluorescent labelling of Beet necrotic yellow vein virus and Beet soil-borne mosaic virus for co- and superinfection experiments in *Nicotiana benthamiana*. *Journal of General Virology* 99 (9), 1321–1330.
- McGrann, G.R., Grimmer, M.K., Mutasa-Göttgens, E.S., Stevens, M., 2009. Progress towards the understanding and control of sugar beet rhizomania disease. *Molecular Plant Pathology* 10 (1), 129–141.
- Pakdel, A., Mounier, C., Klein, E., *et al.*, 2015. On the interaction and localization of the Beet necrotic yellow vein virus replicase. *Virus Research* 196, 94–104.
- Peltier, C., Hleibieh, K., Thiel, H., *et al.*, 2008. Molecular biology of the Beet necrotic yellow vein virus. *Plant Viruses* 2 (1), 14–24.
- Peltier, C., Klein, E., Hleibieh, K., *et al.*, 2012. Beet necrotic yellow vein virus subgenomic RNA3 is a cleavage product leading to stable non-coding RNA required for long-distance movement. *Journal of General Virology* 93 (5), 1093–1102.

Benyviruses (*Benyviridae*)

Annette Niehl, Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

Sebastian Liebe and Mark Varrelmann, Institute of Sugar Beet Research, Göttingen, Germany

Renate Koenig, Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

© 2021 Published by Elsevier Ltd.

This is an update of R. Koenig, Benyvirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00541-0.

Nomenclature

aa Amino acid(s)

CP Coat protein

CP-RT Coat protein-readthrough protein

ELISA Enzyme-linked immunosorbent assay

ER Endoplasmic reticulum

GFP Green fluorescent protein

kDa Kilodalton

ncRNA Non-coding RNA

NGS Next generation sequencing

NLS Nuclear localization signal

nm Nanometer(s)

nt Nucleotide(s)

ORF Open reading frame

PCR Polymerase chain reaction

PD Plasmodesmata

RdRp RNA-dependent RNA polymerase

RNA Ribonucleic acid

siRNAs Small-interfering RNAs

ssRNA Single-stranded ribonucleic acid

TGB Triple gene block

TM Trans-membrane domain

VSR Viral suppressor of gene silencing

Glossary

Benyvirus Siglum derived from beet necrotic yellow vein virus.

Rhizomania Root beardedness (root madness). Extensive proliferation of often necrotizing secondary rootlets at the expense of the main tap root.

History

Beet necrotic yellow vein virus (BNYVV) is the type species of the genus *Benyvirus*. Originally, it was classified as a possible member of the tobamovirus group, because its rod-shaped particles resemble those of tobamoviruses. In 1991, the rod-shaped viruses with multipartite genomes transmitted by species of the plasmodiophorid *Polymyxa* were separated from the monopartite tobamoviruses to form a new group, named furovirus (for 'fungus-transmitted rod shaped viruses'). *Soil-borne wheat mosaic virus* (SBWMV) became the type species of this new group; BNYVV and *Rice stripe necrosis virus* (RSNV) were listed as possible members. Molecular studies conducted in the following years revealed that the genome organization of the members belonging to this new virus group greatly differed from each other. Thus, eventually, four new genera were created, i.e., the genus *Furovirus* with SBWMV as type species, the genus *Benyvirus* with BNYVV as type species, the genus *Pomovirus* with *Potato mop top virus* as type species, and the genus *Pecluvirus* with *Peanut clump virus* as type species. The genus *Benyvirus* presently comprises four definitive species, i.e., BNYVV, *Beet soil-borne mosaic virus* (BSBMV), RSNV and *Burdock mottle virus* (BdMV), and two putative members: *Magnifera indica latent virus* (MILV) and *wheat stripe mosaic virus* (WhSMV). Moreover, a number of beny-like viruses were identified by high throughput sequencing in libraries prepared from several non-plant organisms: a study on invertebrate RNA virus diversity identified a benyvirus-like replication polyprotein gene, the source of this sequence was named *Hubei beny-like virus 1*. Similarly, sequences resembling those in benyviral replicase genes were found in *sclerotium rolfii* and *rhizoctonia solani* in studies aimed at identifying mycoviruses.

Virion Structure, Particle Properties and Associations of Viral Particles With Cellular Structures

Benyviruses are non-enveloped, rod-shaped viruses with multipartite (bi- to pentapartite) positive-single strand RNA (ssRNA) genomes. The RNAs are encapsidated with a coat protein (CP), and in the case of BNYVV also contain the CP-read-through (CP-RT) protein as minor CP in the capsid shell.

The diameter of benyvirus particles is approximately 20 nm. The rods have a central canal and usually show several length maxima ranging from approximately 80 to 400 nm depending on the RNA species encapsidated (Fig. 1). Additional length maxima may be due to end-to-end aggregation or breakage of particles. The right-handed helix of BNYVV particles has a 2.6 nm pitch with an axial repeat of four turns involving 49 subunits of the 21 kDa major CP, which consists of 188 aa. Each CP subunit

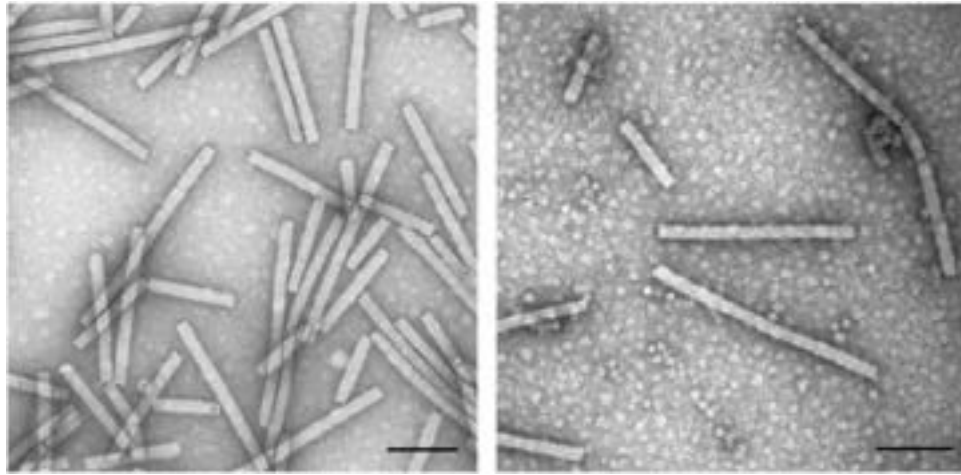


Fig. 1 Transmission electron micrographs of beet necrotic yellow vein virus (left) and beet soil-borne mosaic virus (right) particles negatively stained with uranyl acetate. Scale bars: 100 nm. BNYVV image (left): Courtesy of D.E. Lesemann and J. Engelmann, BBA, Braunschweig, Germany. BSBMV image: Modified from Kastirr, U., Richert-Pöggeler, K., 2018. Berichte aus dem Julius-Kühn Institut. Nr. 201, under the Creative Commons Attribution License (CC BY 4.0).

occupies four nucleotides on the BNYVV RNAs. The 75 kDa CP-RT protein has been detected by immunogold-labeling on one end of the particles in freshly extracted plant sap. Green fluorescent protein (GFP)-labeled particles of BNYVV localize to the cytoplasmic face of mitochondria early during infection, but later they relocate to semi-ordered clusters in the cytoplasm. In ultrathin sections, BNYVV particles are found scattered throughout the cytoplasm or occur in aggregates. Masses of particles with different densities arranged in parallel or angle-layer arrays may be formed.

Organization of the Genome and Properties of the Encoded Proteins

The functions of the BNYVV proteins and RNAs have been studied in detail. Hence, most knowledge on the genome organization of benyviruses is based on BNYVV; however, aa sequence homologies indicate functional similarities between benyvirus proteins.

The BNYVV genome consists of four to five positive-sense, ssRNAs, while the other benyviruses contain only two to four genomic RNAs (Fig. 2 and Table 1). Four genomic RNA segments have been identified for BSBMV, but only two for BdmV, RSNV, WhSMV and MILV. Each benyvirus RNA is capped at the 5' end and polyadenylated at the 3' end. On RNA1, all benyviruses possess one large ORF encoding the replicase. On RNA2, BNYVV, BSBMV, RSMV, BdmV, and WhSMV all contain six ORFs, of which the first two encode CP and CP-RT proteins, and the following three encode triple gene block (TGB)1, TGB2 and TGB3 movement proteins. The cysteine-rich proteins encoded by the 3' terminal ORF may encode the viral silencing suppressors (VSR), as a VSR function has been shown for BNYVV, BSBMV and BdmV cysteine-rich proteins. An RNA3 and RNA4 has only been identified for BNYVV and BSBMV, but some BNYVV isolates have an additional RNA5. RNA3 of BNYVV and BSBMV, RNA4 of BSBMV and RNA5 of BNYVV contain the 'coremin' sequence, which has been assigned RNA silencing suppressor activity and a function in virus systemic movement. RNA3-encoded BNYVV P25 and BSBMV P29 are the viral pathogenicity factors responsible for symptom development in their major host sugar beet; however, BSBMV RNA3-encoded P29 more resembles BNYVV RNA5-encoded P26 in sequence and function, as it resembles BNYVV P25. The 31 kDa and 32 kDa P31 and P32, encoded on RNA4 of BNYVV and BSBMV, respectively, are involved in vector transmission and symptom development. Moreover, BNYVV P31 has been shown to enhance silencing suppressor activity in roots of *Nicotiana benthamiana*.

The BNYVV large open reading frame (ORF) for the replicase protein encoded on RNA1 is cleaved autocatalytically by a papain-like proteinase. Virus replication may take place in association with the endoplasmic reticulum (ER), as the replication-associated proteins localize to this organelle. In *in vitro* systems, translation of the replication protein may start either at the first AUG at position 154 or at a downstream AUG at position 496–498. The resulting proteins of 237 and 220 kDa, respectively, contain in their N-terminal part methyl-transferase motifs, in their central part helicase and papain-like protease motifs, and in their C-terminal part RNA dependent RNA polymerase (RdRp) motifs.

BNYVV RNA2 contains six ORFs, *i.e.*, the CP gene, which is terminated by a suppressible UAG stop codon, the CP-RT protein gene, a TGB coding for proteins of 42, 13, and 15 kDa, and a gene coding for a 14 kDa cysteine-rich protein. The CP is the major capsid protein important for efficient encapsidation of the virus. Within the 75 kDa CP-RT protein, specific domains appear to be important for virion assembly and vector transmission. A KTER motif at aa position 553–556 of the 75 kDa CP-RT protein is essential for efficient transmission of the virus by *P. betae*. Upon repeated mechanical inoculation of the virus, deletions within the C-terminal part of CP-RT occur and the area containing this motif may be lost, which results in abolished vector transmissibility of the virus. Apart from the KTER motif, trans-membrane domains have been predicted in the CP-RT protein, which presumably aid

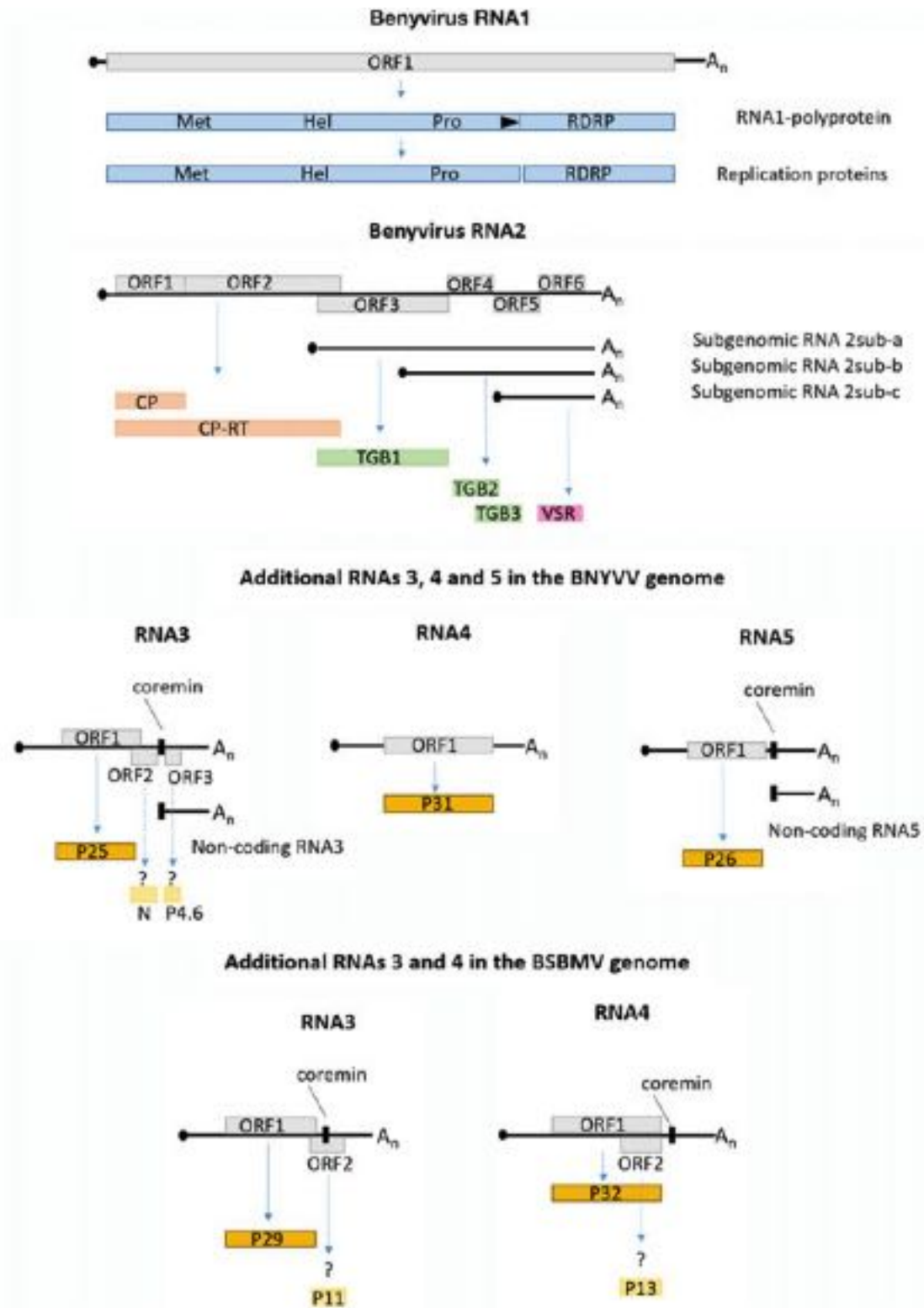


Fig. 2 Benyvirus genome structure. Benyviruses contain two to five genomic RNAs with a 5' guanosylmethionine cap (black circle) and a 3' poly A tail (A_n). Benyviruses share the genome structure of RNA1 and RNA2 and encode similar proteins on corresponding open reading frames (ORF). RNAs 3 and 4 encode proteins with partially overlapping functions. RNA5 is only present in BNYVV. RNA1: self-cleavage of the replicase polyprotein between protease (Pro) and RNA-dependent RNA polymerase (RDRP) domains is indicated by a black arrow. Met, methyltransferase domain; Hel, helicase domain. RNA2: Broken line between ORF1 and ORF2 indicates a read-through stop codon. CP, coat protein; CP-RT, CP-readthrough protein; TGB, triple gene block; VSR, viral silencing suppressor. Arrows indicate protein synthesis from coding sequences, broken arrows and question marks indicate that the proteins corresponding to the respective ORFs have not yet been detected.

Table 1 Benyvirus RNAs and encoded proteins

RNA	Features	BNYVV	BSBMV	RSNV	BdMV	WhSMV	MILV
RNA1	Accession number	D84410	JF513082	EU099844	AB818898	MH151795	KU140662
	Length	6746 bp	6679 bp	6634 bp	7038 bp	6583 bp	6725 bp
	ORFs	1	1	1	1	1	1
	Encoded proteins, function	237 kDa replicase ^a polyprotein 150 kDa replicase 66 kDa RDRP	238 kDa replicase ^a polyprotein	236 kDa replicase ^a poly-protein	249 kDa replicase ^a poly-protein	232 kDa replicase ^a poly-protein	234 kDa replicase ^a poly-protein
RNA2	Accession number	D84411	JF513083	EU099845	AB818899	MH151801	KU140663
	Length	4609 bp	4615 bp	4657 bp	4315 bp	4901 bp	2516 bp
	ORFs	6	6	6	6	6	3
	Encoded proteins, function ^b	75K CP-RT, virion assembly, vector transmission 54K CP, encapsidation 42K TGB1, movement 13K TGB2, movement 15K TGB3, movement 14K CRP, silencing suppression	75K CP-RT, virion assembly, vector transmission 21K CP, encapsidation 42K TGB1, movement 13K TGB2, movement 15K TGB3, movement 14K CRP, silencing suppression	74K CP-RT 20K CP 38K TGB1 12K TGB2 15K TGB3 17K CRP	66K CP-RT 20K CP 38K TGB1 12K TGB2 13K TGB3 13K CRP	81K CP-RT 19K CP 55K TGB1 14K TGB2 13K TGB3 3K	21K CP 39K helicase protein (TGB1?) 15K putative movement protein (TGB2?)
RNA3	Accession number	D84412	EU410955				
	Length	1774 bp	1720 bp				
	ORFs	3	2				
	Encoded proteins, Function	P25, pathogenicity factor N? ^c P4.6? ^c	P29, pathogenicity factor P11? ^c				
	Functional RNAs	Coremin, function in systemic movement and silencing suppression	Coremin, function in systemic movement and silencing suppression				

RNA4	Accession number	D84413	FJ424610
	Length	1465 bp	1730 bp
	ORFs	1	2
	Encoded proteins, Function	P31, transmission factor, symptom determinant, VSR?	P32, transmission factor
	Functional RNAs		P13? ^c Coremin, function in systemic movement and silencing suppression
RNA5	Accession number	D63936	
	Length	1320 bp	
	ORFs	1	
	Encoded proteins, Function	P26, pathogenicity factor	
	Functional RNAs	Coremin, function in systemic movement and silencing suppression	

^aExperimental evidence for autocatalytic cleavage of the replicase protein only exists for BNYVV.

^bProtein functions only experimentally verified for BNYVV and BSBMV.

^cProteins have never been detected in natural infection conditions.

association with distinct vector and host membranes during the viral life cycle. Three sub-genomic RNAs are derived from BNYVV RNA2. The first TGB protein (P42) is translated from sub-genomic RNA2a, the second and third TGB proteins (P13 and P15) from the bicistronic sub-genomic RNA2b, and the 14 kDa cysteine-rich protein from subgenomic RNA2c (Fig. 2). The three TGB proteins are necessary for cell-to-cell movement of the virus. The N-terminal part of the 42 kDa TGB1 protein has nucleic acid-binding activity; its C-terminal part contains helicase motifs. GFP-tagged P42 is targeted by P15 and P13 to punctuate structures associated with plasmodesmata (PD). It has been speculated that P15 and P13 provide docking sites for a P42–viral RNA complex at PD. P42 may alter the PD size exclusion limit to allow virus movement between cells. The 3′ proximal ORF encodes a cysteine-rich 14 kDa protein acting as an RNA silencing suppressor. P14 inhibits secondary siRNA production by interfering with RDR6-dependent siRNA synthesis.

BNYVV RNA3 encodes the 25 kDa P25 pathogenicity factor, which determines symptom severity, likely through reprogramming of auxin signaling within the host plant. In the major host sugar beet, P25 induces massive lateral root proliferation and vein yellowing in systemically infected leaves. The appearance of bright yellow local lesions can be observed on *Tetragonia expansa*. Symptoms seem to depend on the ability of P25 to nucleocytoplasmic shuttling, as mutants unable to shuttle between the nucleus and the cytoplasm fail to induce yellow lesions on *C. quinoa*. The following ORF on RNA 3 encodes a 6.8 kDa N gene, which is not expressed to detectable levels under natural infection conditions. However, expression of the protein is activated by the deletion of upstream sequences, which positions it closer to the 5′ end of RNA3. Expression of the N gene induces the formation of necrotic lesions in BNYVV infections and also when it is expressed from *Cauliflower mosaic virus*. The third ORF on BNYVV RNA3 encodes a protein of 4.6 kDa, which is presumably expressed from an abundantly present sub-genomic RNA; however, a function of this protein is not yet known. RNA3 also contains a functional non-coding RNA (ncRNA), which mediates systemic movement of the virus and functions complementary to P14 as VSR. The RNA3 20 nt ‘coremin’ sequence is responsible for accumulation of the non-coding RNA, as it stalls exoribonucleases of the yeast XRN1 or plant XRN4-type and inhibits RNA processing by the enzymes.

The BNYVV RNA4-encoded 31 kDa P31 protein increases the transmission rate of the virus by *P. betae* and increases symptom severity in *N. benthamiana* and in some Beta species. Moreover, experimental evidence suggests that P31 is able to suppress RNA silencing in roots.

BNYVV RNA5 only occurs in some BNYVV types (P-type and J-type, see section disease, host range and epidemiology below). It enhances symptom expression in sugar beet. Addition of P-type RNA5 to RNA1 and RNA2 results in the formation of necrotic local lesions in *C. quinoa*. P26, the only protein expressed by RNA5, exhibits a nucleo-cytoplasmic localization similar to RNA3-encoded P25 and has been proposed to act synergistically with P25 in symptom development. As RNA5 also contains the 20 nt ‘coremin’ motif it may, in analogy to RNA3, be involved in the systemic movement of the virus.

Blast searches of reference aa sequences of the different benyviruses revealed that the highest sequence similarity between the different members of the benyvirus family is found at the level of the replicase; BNYVV and BSBMV replicase proteins share 85% nt sequence identity. BdMV, RSNV, MILV, and WhSMV replicase proteins share 51%, 44%, 39%, and 31% aa sequence identity, respectively, with the BNYVV replicase protein. Also, the TGB proteins are relatively well conserved with 75%, 82%, and 66% nt identity between TGB1, TGB2 and TGB3 proteins, respectively, of BNYVV and BSBMV; 48%, 49%, and 38% nt identity between TGB1, TGB2, and TGB3 proteins, respectively, of BNYVV and BdMV, and 38%, 44%, and 30% nt sequence identity between TGB1, TGB2, and TGB3 proteins, respectively, of BNYVV and RSNV. The TGB1 and TGB2 proteins of BNYVV and WhSMV share 29% and 45% nt sequence identity, while no significant similarity is found between the TGB3 proteins of the two viruses. The TGB1 and TGB2 proteins of MILV show 27% and 39% nt sequence identity, respectively, with those of BNYVV. At the level of the CP, BNYVV and BSBMV share 59% nt sequence identity, while the CP of BdMV, RSNV, MILV, and WhSMV have 39%, 35%, 27%, and 26% nt sequence identity with BNYVV CP, respectively. The cysteine-rich proteins share only little sequence identity; here BNYVV and BSBMV have 43% aa sequence identity, BNYVV and RSMV 30% and BNYVV and BdMV 22% sequence identity.

Reassortment experiments showed that BNYVV and BSBMV RNA1 and RNA2 reassortants are viable and spread systemically in *N. benthamiana*. The small genomic RNAs 3 and 4 were also exchangeable and systemically infected *B. macrocarpa* and allowed vector transmission. Moreover, rhizomania-specific taproot symptoms in sugar beet could be induced by BSBMV in the presence of BNYVV RNA3 whereas BSBMV RNA3 in combination with BNYVV RNA1 and RNA2 failed to induce taproot symptoms. This underlines the close relationship between both viruses as well as the importance of RNA3 for viral pathogenicity.

Disease, Host Range, and Epidemiology

BNYVV is the causal agent of rhizomania, one of the most damaging diseases of sugar beet. Natural hosts of the virus are *B. vulgaris* ssp. and *B. macrocarpa*. Infections of sugar beet with BNYVV are mainly confined to the root system. Susceptible sugar beet varieties show an extensive proliferation of nonfunctional, necrotizing secondary rootlets, a condition described by the names ‘root beardedness’, ‘root madness’, or rhizomania (Fig. 3). The taproots are stunted, their shape tends to be constricted, and their sugar content is low. A brownish discoloration of the vascular system is often seen. Leaves may become pale and have an upright position. Under dry conditions, wilting is often observed due to insufficient water uptake by the roots. The upper parts of the plants are invaded only rarely by the virus. In infected leaves, the veins turn yellow, and occasionally become necrotic, a condition after which the virus was named (Fig. 3). In susceptible varieties, rhizomania may cause yield losses of 50% and more. Like other diseases caused by soil-borne organisms, rhizomania often occurs in patches within infected fields. BNYVV can be mechanically



Fig. 3 Benyvirus symptoms on sugar beet. Typical symptoms on roots and leaves of BNYVV (A) and BSBMV (B) infected sugar beet. Plants were mechanically inoculated with the virus (C) shows a root and a leaf of a healthy control plant.

transmitted to many species in the *Chenopodiaceae* and to some species in the *Amaranthaceae* and *Tetragoniaceae* as well as to *N. benthamiana*. *B. macrocarpa*, *S. oleracea* and *N. benthamiana* become infected systemically. BNYVV occurs in sugar beet-growing areas worldwide. Molecular analyses have revealed the existence of different BNYVV types (A type, B type, P type, and J-type) that cannot be distinguished serologically. The A type is most common and occurs in Europe, in the USA, and in East Asia. The B type is prevalent in Germany, France, and other central European countries. B-type-like BNYVV is also found in East Asian countries. The P type, which usually contains an RNA5 and causes especially severe symptoms, has only been found in restricted areas in central France (Pithiviers), the UK, Germany and in Kazakhstan. The J-type has been identified in Asia. The current BNYVV types seem to have evolved from at least four original populations that developed in native hosts in East Asia and spread from there. Naturally infected local hosts in different geographic regions may harbor various types of BNYVV. Once transmission from a native host to sugar beet has been successful, further spread in beet-growing areas by machinery, irrigation, or infested soil may be very rapid.

The symptoms caused by BSBMV on its natural host *B. vulgaris* are more variable than those produced by BNYVV. Infected roots may remain symptomless (Fig. 3) or occasionally show rhizomania-like symptoms. Infections of the upper parts of the plants occur more frequently than with BNYVV. Leaves may develop vein banding or faint mottling. When plants are dually infected with BSBMV and BNYVV, foliar symptoms appear more frequently. In general, BSBMV causes much less damage to sugar beet than BNYVV. It has been reported that, when present in mixed infections, BNYVV suppresses BSBMV. In addition, at the cellular level, BNYVV and BSBMV remain spatially separated. BSBMV can be transmitted mechanically to *C. quinoa*, *Chenopodium album*, and *T. tetragonioides*, all of which become infected locally, and to *Beta maritima*, which becomes infected systemically. BSBMV is widely distributed in the USA, but has so far not been found in other parts of the world. Analyses of polymerase chain reaction (PCR) products indicated that this virus is genetically much more variable than BNYVV. This may explain the variability in the symptoms produced by this virus.

RSNV was identified in 1983 in West Africa (Ivory Coast) as the causal agent of a long-known disease named 'rice crinkling disease'. The percentage of infection varies greatly between varieties and in different years. RSNV also occurs in additional African countries, so in Liberia, Nigeria, and Sierra Leone. Recently, RSNV was identified in Burkina Faso (2014), Benin (2015), and Mali (2017). In South America, the virus was first noticed in Columbia, where it causes a severe disease of rice named 'entorchamiento' that is characterized by seedling death, foliar striping, and severe plant malformation and is also present in Ecuador, Brazil and Panama. Recently, (2018) RSNV was also reported in Argentina. The virus can be transmitted mechanically to *C. quinoa*, *C. album*, and *Chenopodium amaranticolor* where it produces local lesions, but not to rice. *N. benthamiana* does not become infected.

BdMV has been isolated from leaves of naturally infected burdock (*Arctium lappa* L.) plants in Japan in 1970. Leaves as well as roots of this plant are common vegetables in Japan. BdMV causes mild chlorosis or mottling symptoms on leaves of burdock plants, the only known natural host of the virus. The symptoms can become masked in older plants. BdMV can be transmitted mechanically to *C. quinoa*, *C. murale*, *N. clevelandii*, and *N. rustica*, which are infected systemically, and to *B. vulgaris* var. *cicla* and var. *rapa*, *S. oleracea*, *Cucumis sativus*, and *T. expansa*, which are infected only locally, some of them only with difficulty. So far, BdMV has only been reported to occur in Japan.

WhSMV has only recently (2019) been isolated from wheat plants in Brazil showing chlorotic to necrotic stripes on the leaves, symptoms of soil-borne wheat mosaic disease typically associated with SBWMV and wheat spindle streak mosaic virus, and causes strong crop losses in Brazil.

Transmission

BNYVV, BSBMV and presumably also RSNV and WhSMV are transmitted by species of the obligate biotrophic root parasite *Polymyxa*. While direct evidence for the transmission of BNYVV and BSBMV by *Polymyxa betae* exists through transmission studies, the evidence, that RSNV and WhSMV are transmitted by *P. graminis* remains correlative, based on the consistent occurrence of the pathogen in virus infected roots. Whether the other members of the benyvirus family are also transmitted by plasmodiophorids is not yet known. RSNV-harboring cystosori may be carried in soil adhering to the surface of rice seeds, but true seed transmission has not been observed. Under experimental conditions, benyviruses are readily transmitted mechanically among test plants. However, attempts to transmit BNYVV mechanically from sugar beet rootlets to test plants may not always be successful.

The *Polymyxa* life cycle begins with the encystment of a zoospore on a host plant root epidermal or root hair cell, injection of the spore content into the host cytoplasm and the formation of a plasmodium inside the host cell. It is believed that virus transmission to the host and acquisition of virus from the host takes place during this plasmodial phase. The plasmodium then differentiates either into a zoosporangium or into resting spores; the mechanisms driving entry into a zoosporangial or sporogenic phase are not yet understood. BNYVV virus-like particles were found inside *P. betae* zoosporangia and proteins associated with viral replication and movement have been detected inside *P. betae* zoospores and resting spores, allowing the hypothesis that virus replication and movement may take place within *Polymyxa*. Whether *Polymyxa* may indeed be a host for the virus remains to be shown. Virus present in *Polymyxa* resting spores remains infective for many years. The spores may be distributed on agricultural equipment, by irrigation, or even by wind.

Diagnosis

The presence of symptoms typical for BNYVV is a good first indicator for benyvirus infection. Root symptoms, *i.e.*, rootlet proliferation and necrosis of vascular bundles are considered reliable indicators of infection. However, symptoms on the foliage of plants in the field can easily be confused with those resulting from other causes, *e.g.*, infection with other viruses, nematodes, fungi, or abiotic stress such as nitrogen deficiency. Thus, benyvirus infections are diagnosed most reliably by serological or molecular biological techniques, bioassays with indicator plants, electron microscopy or a combination of these techniques. Benyviruses are moderately to strongly immunogenic. Polyclonal antibodies have been obtained for BNYVV, BSBMV, and RSNV. BNYVV and BSBMV are only very distantly serologically related – some antisera may detect only one of the two viruses. Particles of RSNV and BdMV do not react with antisera raised against BNYVV. Monoclonal antibodies against BNYVV have been used for diagnostic purposes and for determining the accessibility of aa on the virus particles. Single-chain antibody fragments have also been produced and used to detect BNYVV. Enzyme-linked immunosorbent assay (ELISA), immunoelectron microscopy, tissue-print immunoblotting, western blot, and immunocapture PCR are serological methods commonly used for the detection of benyviruses. Molecular biology techniques based on RT-PCR or sequencing are also routinely applied for the diagnosis of benyviruses. Especially the identification of the putative new benyvirus member species has been achieved by next generation sequencing (NGS) technology. An advantage of PCR-based methods alone or in combination with sequencing of PCR products as well as of NGS is that they allow the discrimination of different virus types or strains, which are serologically indistinguishable and, in the case of NGS, allow the identification of viruses without the requirement of *a priori* sequence information. Electron microscopy techniques, *i.e.*, transmission electron microscopy, immunoelectron microscopy, or immunogold labeling are also possibilities to diagnose benyvirus infection. Bioassays relying on mechanical inoculation of indicator plants are also useful, though laborious. Infestation of soil may be detected by means of bait plants, which are planted into contaminated soil and tested for the presence of virus, usually by ELISA or with PCR-based methods.

Control

Polymyxa-transmitted plant viruses may survive in soil in the long-living resting spores of the vector for many years, probably even decades. The diseases caused by these viruses are therefore difficult to control. Chemical approaches, e.g., by soil treatment with methyl bromide, is neither efficient nor acceptable for economic and ecological reasons. Growing resistant or tolerant varieties currently represents the only practical and environmentally friendly means to lower the impact of these diseases on yield. BNYVV-tolerant or partially resistant sugar beet varieties are available which enable good yields also on infested fields. Their resistance is based mainly on the Rz1 gene from a sugar beet line ('Holly' resistance) and the Rz2 gene from the *B. maritima* line WB42. Rz1 resistance-breaking BNYVV strains have been reported, but presumably lose their advantage in susceptible varieties. The resistance breaking properties have been associated with aa in the P25 pathogenicity factor as well as with the presence of RNA5. Additional resistance genes (Rz3 to Rz5) have been identified, but have not yet been introduced into commercial varieties. Among the different known loci conferring resistance to BNYVV, all located on Chromosome III in *B. vulgaris*, only Rz2 has been identified to encode a CC-NBS-LRR (coiled-coil, nt binding site, leucine rich-repeat) resistance protein. These proteins typically act to restrict virus infection during effector-triggered immunity. Genetically modified sugar beet expressing various portions of the BNYVV genome have been shown to be highly resistant to BNYVV. Resistance against the vector has been detected in *B. patellaris* and *B. procumbens*, but attempts to develop agronomically acceptable sugar beet cultivars resistant to *P. betae* have so far failed.

The degree of resistance or tolerance to RSNV differs in different rice cultivars. A high degree of resistance is found in *Oryza glaberrima*. Studies with *O. glaberrima* x *O. sativa* introgression lines allowed the identification of an RSNV resistance locus on chromosome 11. This resistance locus has so far not been introduced into commercial varieties.

Apart from growing resistant plants, prevention methods consisting of stringent hygiene and decontamination procedures for contaminated material and instruments may help to limit further spread of the pathogen.

Similarities and Dissimilarities With Other Taxa

The morphology of benyviruses resembles that of other rod-shaped viruses, that is, of furoviruses, goraviruses, hordeiviruses, pecluviruses, pomoviruses, tobamoviruses and tobnaviruses. Phylogenetic relationships between benyviruses and viruses from other genera and similarities with other taxa have been extensively reviewed in the 10th report of the International Committee on Taxonomy of Viruses. The benyvirus genomes are multipartite with two to five RNAs. Tobamoviruses have monopartite, goraviruses, furoviruses, pecluviruses, and tobnaviruses bipartite, and hordeiviruses and pomoviruses tripartite genomes. The fact that the RNAs of benyviruses are polyadenylated differentiates them from the RNAs of the other viruses listed above. Opposed to the other rod-shaped viruses, which encode their replication-associated proteins in two ORFs, benyviruses have a single large ORF on their RNA1 coding for a polypeptide, which is cleaved post-translationally into two replication-associated proteins. The replication-associated ORFs of hordeiviruses are located on two different RNAs; in the case of furoviruses, gora-, peclu-tobamo-, and tobnaviruses, the ORFs coding for replication proteins are located on the same RNA, where ORF1 is terminated by a leaky stop codon and extends into an ORF2. The methyl-transferase, helicase, and RNA-dependent RNA polymerase motifs in the benyvirus replicase display higher similarity to e.g., *Hepatitis virus E* and *Rubella virus* than to other rod-shaped plant viruses, and database similarity searches of benyviral replicase sequences reveal similarities to sequences in plant-, fungal- and invertebrate-derived NGS sequencing libraries. Taken together, this supports a model in which benyviruses may have specialized from ancestors infecting a wide range of hosts.

Like hordei-, gora-, peclu-, and pomoviruses, benyviruses have their movement function encoded on a TGB. Sequence comparisons reveal similarities within the first and second TGB-encoded proteins of hordei- and pomoviruses as well as to carla- and potexviruses.

With respect to structural motifs in RNA, it is interesting to note that the 'coremin' motif present on BNYVV and BSBMV RNAs is also present on *Cucumber mosaic virus* RNAs. One of the functions assigned to this motif is stalling of host exonucleases, resulting in the production of ncRNAs with functions associated with virus movement and silencing suppression.

Further Reading

- Biancardi, E., Tamada, T. (Eds.), 2016. Rhizomania. Cham: Springer.
- Chiba, S., Hleibieh, K., Delbianco, A., et al., 2012. The benyvirus RNA silencing suppressor is essential for long-distance movement, requires both zinc-finger and NoLS basic residues but not a nucleolar localization for its silencing-suppression activity. *Molecular Plant-Microbe Interactions* 26, 168–181.
- Chiba, S., Kondo, H., Miyanishi, M., et al., 2011. The evolutionary history of Beet necrotic yellow vein virus deduced from genetic variation, geographical origin and spread, and the breaking of host resistance. *Molecular Plant-Microbe Interactions* 24, 207–218.
- Commandeur, U., Koenig, R., Manteuffel, R., et al., 1994. Location, size, and complexity of epitopes on the coat protein of Beet necrotic yellow vein virus studied by means of synthetic overlapping peptides. *Virology* 198, 282–287.
- D'Alonzo, M., Delbianco, A., Lanzoni, C., et al., 2012. Beet soil-borne mosaic virus RNA-4 encodes a 32 kDa protein involved in symptom expression and in virus transmission through *Polymyxa betae*. *Virology* 423, 187–194.
- Flobinus, A., Chevigny, N., Charley, P.A., et al., 2018. Beet necrotic yellow vein virus noncoding RNA production depends on a 5'→3' Xrn exonuclease activity. *Viruses* 10, 137.

- Flobinus, A., Hleibieh, K., Klein, E., *et al.*, 2016. A viral non-coding RNA complements a weakened viral RNA silencing suppressor and promotes efficient systemic host infection. *Viruses* 8, 272.
- Hehn, A., Fritsch, C., Richards, K.E., Guilley, H., Jonard, G., 1997. Evidence for *in vitro* and *in vivo* autocatalytic processing of the primary translation product of Beet necrotic yellow vein virus RNA 1 by a papain-like proteinase. *Archives of Virology* 142, 1051–1058.
- Jupin, I., Guilley, H., Richards, K.E., Jonard, G., 1992. Two proteins encoded by Beet necrotic yellow vein virus RNA3 influence symptom phenotype on leaves. *The EMBO Journal* 11, 479–488.
- Laufer, M., Mohammad, H., Christ, D.S., *et al.*, 2018. Fluorescent labelling of Beet necrotic yellow vein virus and Beet soil-borne mosaic virus for co- and superinfection experiments in *Nicotiana benthamiana*. *Journal of General Virology* 99, 1321–1330.
- Rahim, M.D., Andika, I.B., Han, C., Kondo, H., Tamada, T., 2007. RNA-4 encoded p31 of Beet necrotic yellow vein virus is involved in efficient vector transmission, symptom severity and silencing suppression in roots. *Journal of General Virology* 88, 1611–1619.
- Scholten, O.E., Jansen, R.C., Paul Keizer, L.C., De Bock, T.S.M., Lange, W., 1996. Major genes for resistance to Beet necrotic yellow vein virus (BNYVV) in *Beta vulgaris*. *Euphytica* 91, 331–339.
- Shi, M., Lin, X.-D., Tian, J.-H., *et al.*, 2016. Redefining the invertebrate RNA virosphere. *Nature* 540, 539.
- Tamada, T., Schmitt, C., Saito, M., *et al.*, 1996. High resolution analysis of the readthrough domain of Beet necrotic yellow vein virus readthrough protein: A KTER motif is important for efficient transmission of the virus by *Polymyxa betae*. *Journal of General Virology* 77, 1359–1367.
- Valente, J.B., Pereira, F.S., Stempkowski, L.A., *et al.*, 2019. A novel putative member of the family *Benyviridae* is associated with Soilborne wheat mosaic disease in Brazil. *Plant Pathology* 68, 588–600.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/benyviridae/648/genus-benyvirus

Genus: Benyvirus

Benyviridae

Positive-sense RNA Viruses.

Betaflexiviruses (*Betaflexiviridae*)

Nobuyuki Yoshikawa, Iwate University, Morioka, Japan

Hajime Yaegashi, Institute of Fruit Tree and Tea Science, NARO, Morioka, Japan

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)

AlkB Alpha-ketoglutarate-dependent dioxygenase

CP Coat protein or capsid protein

dsRNA double-stranded ribonucleic acid

Hel Helicase

kb Kilobases; the size of a ssDNA or ssRNA molecule

kbp Kilobase pairs; the size of a dsDNA or dsRNA molecule

kDa Kilo Daltons; the size of a protein

Met Methyl-transferase

MP Movement protein

NBP RNA-binding protein

nm Nanometer(s)

nt Nucleotide(s)

ORF Open reading frame

Pol RNA-dependent RNA polymerase

P-Pro Papain-like protease

Rep Replication initiator protein

S_{20,w} Corrected sedimentation coefficient; extrapolated in water at 20°C and infinitely diluted

sgRNA sub-genomic RNA

ssRNA single-stranded ribonucleic acid

TGB Triple gene block

Glossary

Alfavirus A genus in the family *Togaviridae*. The type species is *Sindbis virus*.

Helical symmetry A form of capsid structure found in many RNA viruses in which the protein subunits which interact with the nucleic acid form a helix.

Movement protein A virus-encoded protein which is essential for the cell-to-cell movement of the virus in plant tissues.

RNA silencing suppressor Viral protein that suppresses the antiviral RNA silencing response by the host.

Semi-persistent transmission The relationship between a plant virus and its arthropod vector which is intermediate between non-persistent transmission and persistent

transmission. It has the features of short acquisition feed and no latent period found in non-persistent manner, but the vector remains able to transmit the virus for periods of hours to days which is longer than the non-persistent manner.

Sub-genomic RNA A segment of RNA generated from a genomic RNA via an internal promoter that has the same 3' end as the genomic RNA, but has a deletion at the 5' end. The sub-genomic RNA makes it possible to efficiently translate the downstream open reading frame of the genomic RNA.

The 30K superfamily The group of movement proteins related to the 30 kDa ('30K') movement protein of Tobacco mosaic virus.

Introduction

Betaflexiviridae is a family of viruses in the order *Tymovirales* and contains viruses infecting plants and sharing a distinct lineage of alphavirus-like replication proteins. The viruses in this family have flexuous particles composed of a single stranded (plus-sense), polyadenylated RNA genome, and a single coat protein. In the ICTV 9th report (2009 release), six genera (*Capillovirus*, *Carlavirus*, *Citriivirus*, *Foveavirus*, *Trichovirus*, and *Vitivirus*) were established in this family. Now in 2019, two subfamilies and six new genera have been added, and the family *Betaflexiviridae* now consists of two subfamilies and 12 genera including 107 species, and 27 tentative species in the family.

Taxonomy and Phylogeny

The members of the family *Betaflexiviridae* are classified in two subfamilies *Quinvirinae* and *Trivirinae* and twelve genera: four genera; *Carlavirus*, *Foveavirus*, *Robigovirus* in the subfamily *Quinvirinae*, and eight genera; *Capillovirus*, *Chordovirus*, *Citriivirus*, *Divavirus*, *Prunovirus*, *Tepovirus*, *Trichovirus*, *Vitivirus*, and *Wamavirus* in *Trivirinae* in the subfamily *Trivirinae* (Table 1). Three species in the subfamily *Quinvirinae*; *Banana mild mosaic virus*, *Banana virus X*, and *Sugarcane striate mosaic-associated virus* are not assigned to any genera yet. The viruses in the subfamily *Quinvirinae* have the 'triple gene block' (TGB) as the proteins involved in cell-to-cell movement. On the other hand, the members in the subfamily *Trivirinae* have a single movement protein (MP) of the 30K superfamily (Table 2). The genus demarcation criteria in this family are less than 45% nt sequence identity and 40% aa identity in the replication-associated protein (Rep) and coat protein (CP), and the different species within the same genus should have less than 72% nt identity (or 80% aa identity) in the Rep and CP genes.

Table 1 Subfamilies, genera, and species in the family *Betaflexiviridae*. Type species are written in bold

Subfamilies	Genera	Species	Acronyms	Accession-#
Quinvirinae	Carlavirus	Carnation latent virus	CLV	NC038865
		<i>Aconitum latent virus</i>	AcLV	NC002795
		<i>American hop latent virus</i>	AHLV	NC017859
		<i>Atractylodes mottle virus</i>	AtMV	NC038966
		<i>Blueberry scorch virus</i>	BlScV	NC003499
		<i>Butterbur mosaic virus</i>	ButLV	NC013527
		<i>Cactus virus 2</i>	CV – 2	ND
		<i>Caper latent virus</i>	CapLV	NC043080
		<i>Chrysanthemum virus B</i>	CVB	NC009087
		<i>Cole latent virus</i>	CoLV	NC038322
		<i>Coleus vein necrosis virus</i>	CVNV	NC009764
		<i>Cowpea mild mottle virus V</i>	CPMMV	NC014730
		<i>Cucumber vein-clearing virus</i>	CuVCV	NC043081
		<i>Daphne virus S</i>	DVS	NC008020
		<i>Gaillardia latent virus</i>	GaILV	NC023892
		<i>Garlic common latent virus</i>	GarCLV	NC016440
		<i>Helenium virus S</i>	HVS	NC038323
		<i>Helleborus mosaic virus</i>	HeMV	NC043082
		<i>Helleborus net necrosis virus</i>	HNNV	NC012038
		<i>Hippeastrum latent virus</i>	HiLV	NC011540
		<i>Hop latent virus</i>	HpLV	NC00255
		<i>Hop mosaic virus</i>	HpMV	NC010538
		<i>Hydrangea chlorotic mottle virus</i>	HdLV	NC012869
		<i>Kalanchoe latent virus</i>	KLV	NC013006
		<i>Ligustrum necrotic ringspot virus</i>	LNRSV	NC010305
		<i>Ligustrum virus A</i>	LVA	NC031089
		<i>Lily symptomless virus</i>	LSV	NC005138
		<i>Melon yellowing-associated virus</i>	MYaV	NC038324
		<i>Mirabilis jalapa mottle virus</i>	MjMV	NC016080
		<i>Narcissus common latent virus</i>	NCLV	NC008266
		<i>Nerine latent virus</i>	NeLV	NC028111
		<i>Passiflora latent virus</i>	PLV	NC008292
		<i>Pea streak virus</i>	PeSV	NC027527
		<i>Phlox virus B</i>	PhIVB	NC009991
		<i>Phlox virus M</i>	PhIVM	NC04308
		<i>Phlox virus S</i>	PhIVS	NC009383
		<i>Poplar mosaic virus</i>	PopMV	NC005343
		<i>Potato latent virus</i>	PotLV	NC011525
		<i>Potato virus H</i>	PVH	NC018175
		<i>Potato virus M</i>	PVM	NC001361
		<i>Potato virus P</i>	PVP	NC009759
		<i>Potato virus S</i>	PVS	NC007289
		<i>Red clover vein mosaic virus</i>	RCVMV	NC012210
		<i>Sambucus virus C</i>	SVC	NC029087
		<i>Sambucus virus D</i>	SVD	NC029088
		<i>Sambucus virus E</i>	SVE	NC029089
		<i>Shallot latent virus</i>	SLV	NC003557
		<i>Sint-Jan onion latent virus</i>	SJOLV	NC043084
		<i>Strawberry pseudo mild yellow edge virus</i>	SPMYV	ND
		<i>Sweet potato C6 virus</i>	SPC6V	NC018448
		<i>Sweet potato chlorotic fleck virus</i>	SPCFV	NC006550
		<i>Verbena latent virus</i>	VeLV	NC043085
		<i>Yam latent virus</i>	YLV	NC026248
	Unassigned	<i>Alfalfa latent virus</i>	ALV	NC026616
		<i>Chrysanthemum virus R</i>	CVR	NC04070
		<i>Edelberry carlavirus</i>	EV	NC029086
		<i>Edelberry carlavirus A</i>	EVA	NC029085
		<i>Jasmine virus C</i>	JVC	NC030926
		<i>Narcissus latent virus</i>	NLV	NC008552
		<i>Opuntia virus H</i>	OV2	LS999823
		<i>Pepper virus A</i>	PVA	NC034376

Table 1 Continued

Subfamilies	Genera	Species	Acronyms	Accession-#	
	Foveavirus	Apple stem pitting virus	ASPV	NC003462	
		<i>Apricot latent virus</i>	ApLV	NC014821	
		<i>Asian prunus virus 1</i>	APV1	NC025388	
		<i>Asian prunus virus 2</i>	APV2	NC028868	
		<i>Grapevine rupestris stem pitting-associated virus</i>	GRSPaV	NC001948	
		<i>Grapevine virus T</i>	GVT	NC035203	
		<i>Peach chlorotic mottle virus</i>	PCMoV	NC009892	
		<i>Rubus canadensis virus 1</i>	RuCV — 1	NC019025	
		Unassigned	Apple green crinkle virus	AGCV	NC018714
	Asian prunus virus 3		APV3	NC028975	
	Robigovirus	Cherry necrotic rusty mottle virus	CNRMV	NC002468	
		<i>African oil palm ringspot virus</i>	AOPRV	NC012519	
		<i>Cherry green ring mottle virus</i>	CGRMV	NC001946	
		<i>Cherry rusty mottle associated virus</i>	CRMaV	NC020996	
		<i>Cherry twisted leaf associated virus</i>	CTLaV	NC024449	
		Unassigned	Banana mild mosaic virus	BanMMV	NC002729
			Banana virus X	BaVX	NC04308
			Sugarcane striate mosaic-associated virus	SCSMaV	ND
	Trivirinae	Capillovirus	Apple stem grooving virus	ASGV	NC001749
			<i>Cherry virus A</i>	CVA	NC003689
			<i>Currant virus A</i>	CuVA	NC029301
			<i>Mume virus A</i>	MuVA	NC040568
			Unassigned	Birch capillovirus	BCV
Yacon virus A		YVA		NC030657	
Chordovirus		Carrot Ch virus 1	CChV — 1	NC025469	
		<i>Carrot Ch virus 2</i>	CChV — 2	NC025468	
		Unassigned	Lettuce chordovirus 1	LeCV1	NC040627
Citrivirus		Citrus leaf blotch virus	CLBV	NC003877	
		Unassigned	Citrus leaf blotch virus 2	CLBV2	MH144344
Divavirus		Diuris virus A	DiVA	NC019029	
		<i>Diuris virus B</i>	DiVB	NC019030	
		<i>Hardenbergia virus A</i>	HarMV	NC015395	
		Unassigned	<i>Occimum basilicum RNA virus 1</i>	OBRV1	NC035462
Prunevirus		Apricot vein clearing associated virus	AVCaV	NC023295	
		<i>Actinidia seed borne latent virus</i>	ASbLV	NC040800	
		<i>Caucasus prunus virus</i>	CPV	NC038325	
Tepovirus		Potato virus T	PVT	NC011062	
		<i>Prunus virus T</i>	PrVT	NC024686	
		Unassigned	<i>Zostera virus T</i>	ZVT	MK514426
Trichovirus		Apple chlorotic leaf spot virus	ACLSV	NC001409	
		<i>Apricot pseudo-chlorotic leaf spot virus</i>	APCLSV	NC006946	
		<i>Cherry mottle leaf virus</i>	CMLV	NC002500	
		<i>Grapevine Pinot gris virus</i>	GPGV	NC015782	
		<i>Grapevine berry inner necrosis virus</i>	GINV	NC015220	
		<i>Peach mosaic virus</i>	PcMV	NC011552	
		<i>Phlomis mottle virus</i>	PhMV	NC043412	
		Unassigned	Fig latent virus 1	FLV — 1	FN377573
			Peach chlorotic leaf spot virus	PCLSV	MH084695
			Peach virus M	PeVM	MK012336
	Vitivirus	Grapevine virus A	GVA	NC003604	
		<i>Actinidia virus A</i>	AcVA	NC043087	
		<i>Actinidia virus B</i>	AcVB	NC016404	
<i>Arracacha virus V</i>		AVV	NC03426		
<i>Blackberry virus A</i>		BVA	NC040630		
<i>Grapevine virus B</i>		GVB	NC003602		

(Continued)

Table 1 Continued

Subfamilies	Genera	Species	Acronyms	Accession#
		<i>Grapevine virus D</i>	GVD	NC038326
		<i>Grapevine virus E</i>	GVE	NC011106
		<i>Grapevine virus F</i>	GVF	NC018458
		<i>Grapevine virus G</i>	GVG	NC040616
		<i>Grapevine virus H</i>	GVH	NC040545
		<i>Grapevine virus I</i>	GVI	NC037058
		<i>Grapevine virus J</i>	GVJ	NC040564
		<i>Heracleum latent virus</i>	HLV	NC039087
		<i>Mint virus 2</i>	MV – 2	NC043088
	Unassigned	<i>Agave tequilana leaf virus</i>	AtLV	NC034833
		<i>Arracacha latent virus V</i>	ALVV	KY451036
		<i>Grapevine virus K</i>	GVK	NC035202
		<i>Grapevine virus L</i>	GVL	MH248020
		<i>Grapevine virus M</i>	GVM	MK492703
	Wamavirus	Watermelon virus A	WVA	NC034377

Table 2 Distinguishing properties of sub-families and genera in the family *Betaflexiviridae*

Subfamily	Genus	Virion length (nm)	Genome size (kb)	No. of ORFs	Proteins				Unknown protein ^g
					Rep ^a	MP (s) ^b	CP ^c	RBP ^d	
<i>Quinvirinae</i>	<i>Carlavirus</i>	610–700	5.8–9	6	215–225	TGB	32–36	11–16	
	<i>Foveavirus</i>	800	8.4–9.3	5	230–250	TGB	28–44	–	
	<i>Robigovirus</i>	800	8–8.5	5	209–232	TGB	29–30	–	
<i>Trivirinae</i>	<i>Capillovirus</i>	640–700	6.5–7.5	2	210–245	30Ksf	25–27	–	
	<i>Chordovirus</i>	640–760	8	3	213	30Ksf	24	–	
	<i>Citriovirus</i>	960	8.7	3	227	30Ksf	41	–	
	<i>Dvavirus</i>	640	6.5–7.5	2	237	30Ksf	20	–	
	<i>Prunevirus</i>	960	8.7	4	193	30Ksf	25	16	
	<i>Trichovirus</i>	640–890	7.5–8	3 or 4	215–220	30Ksf	21–24	(15–16) ^f	
	<i>Tepovirus</i>	640	6.5	3	206	30Ksf	25	–	
	<i>Vitivirus</i>	725–785	7.6	5	190–200	30Ksf	18–22	11–14	19
	<i>Wamavirus</i>	–	8.4	4	212	30Ksf	27	–	25

^aReplication-associated protein (kDa).^bMovement protein either of the 30K superfamily (30Ksf) or a triple gene block (TGB).^cCoat protein (kDa).^dRNA-binding protein (kDa).^eThe hypothetical protein of unknown function (kDa).^fThe presence of RBP is depending on the virus species.

Phylogenetic analysis of the aa sequences of the Rep (POL) and CP of 106 members in the family *Betaflexiviridae* are shown in **Fig. 1(a)** and **(b)**. A phylogenetic tree constructed using POL sequences showed that the two subfamilies are separated and members in most genera are grouped into distinct clusters (**Fig. 1(a)**). As an exception, ASGV, a type species of the genus *Capillovirus*, and another three species (*Cherry virus A*, *Currant virus A*, and *Mume virus A*) are separated into different clusters. In a phylogenetic tree using the CP sequence (**Fig. 1(b)**), members of the genus *Foveavirus* are grouped into two different clusters which were divided into two subfamilies. Phylogenetic trees based on aa sequences of TGB1 proteins for the members of the subfamily *Quinvirinae* and MPs of the members of the subfamily *Trivirinae* are shown **Fig. 1(c)** and **(d)**, respectively. In the subfamily *Quinvirinae* -TGB1, members of the genus *Carlavirus* are divided into several clusters (**Fig. 1(c)**).

Members of the Family

The number of virus species within each genus varies greatly depending on the genus, the largest genus *Carlavirus* contains 53 species and the smallest ones *Citriovirus* and *Wamavirus* consist of only one virus species (**Table 1**). There is another 27 betaflexiviruses that are tentative members for some genera, and for which full sequences are already available (**Table 1**).

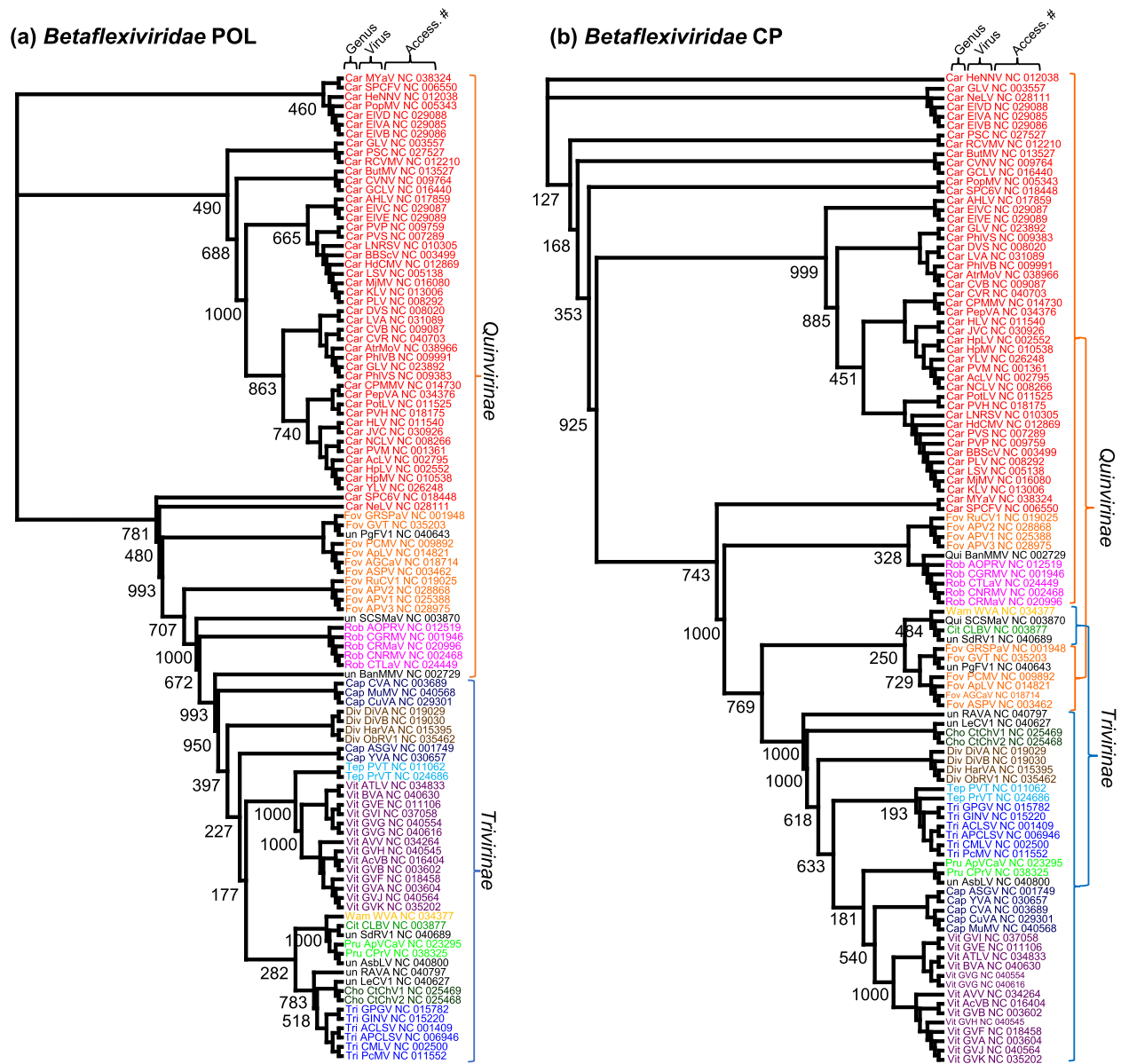


Fig. 1 Phylogenetic relationships among virus species in the family *Betaflexiviridae*. Phylogenetic tree was constructed by CLUSTAL W based on the amino acid sequences of the polymerase of viruses in the family *Betaflexiviridae* (1a), the coat protein of viruses in the family *Betaflexiviridae* (1b), the TGB1 protein of viruses in the subfamily *Quinvirinae*, and movement protein (MP) of viruses in the subfamily *Trinvirinae*. The POL amino acid sequences of 106 viruses classified in the family *Betaflexiviridae* (approved and tentative) were collected from Ref-seq database at July 2019. Virus species in each genus were selected from the virus reference sequence collection of the gene bank. The abbreviations of the viruses are listed in Table 1.

Virion Structure

Virions are non-enveloped, flexuous filamentous particles, 610–960 nm long and 10–13 nm in diameter, depending on the genus and species (Table 2). They have helical symmetry with a pitch of about 3.4 nm (range 3.3–3.7 nm). Electron micrographs of members of the genera *Potyvirus* (*Potyviridae*), *Potexvirus*, (*Alphaflexiviridae*), *Closterovirus* (*Closteroviridae*), *Carlavirus*, *Foveavirus*, *Capillovirus*, *Trichovirus*, and *Vitivirus* (the last five are in the family *Betaflexiviridae*) suggests that the flexibility and surface structure of the virus particles are not significantly different between Zucchini yellow mosaic virus (*Potyvirus*), Potato virus X (*Potexvirus*), and three carlaviruses (Potato virus S, Southern potato latent virus, and Strawberry pseudo mild yellow edge virus). On the other hand, the particle flexibility and cross-banding structure of Apple stem pitting virus (ASPV) (*Foveavirus*), Apple stem grooving virus

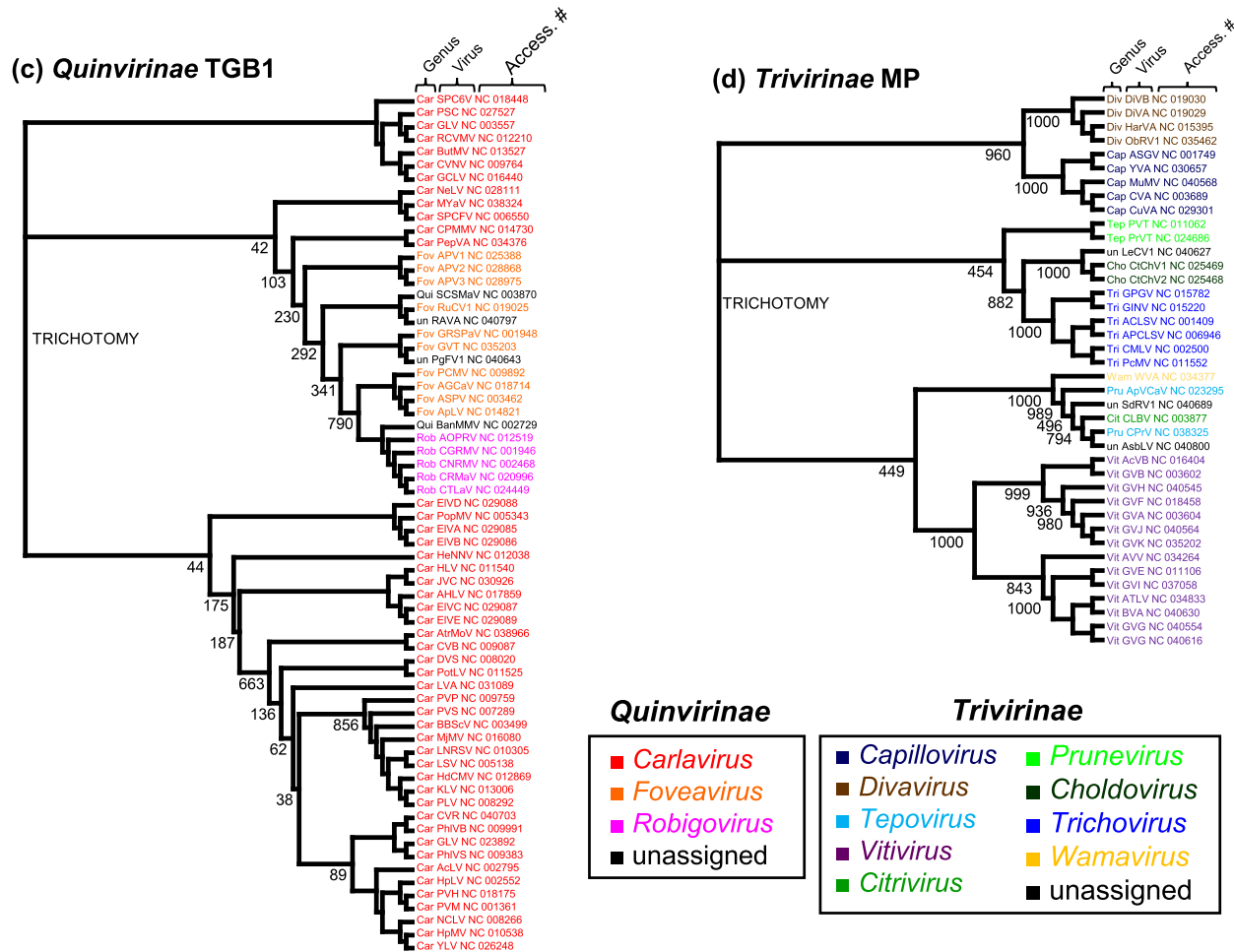


Fig. 1 Continue.

(ASGV) (*Capillovirus*), Apple chlorotic leaf spot virus (ACLSV) (*Trichovirus*), and Grapevine virus A (*Vitivirus*) resembles to those of Citrus tristeza virus (*Closterovirus*) (Fig. 2).

Virions of the members in the *Betaflexiviridae* sediment as a single species with a sedimentation coefficient ($S_{20,w}$) of 92–176S, depending on the genus and species. Virions are composed of a linear positive-sense, single-stranded RNA (ssRNA) of 5.8–9.3 kb which is 5%–6% by weight of the virion and a single coat protein of Mr of 18–44 kDa (Table 2). The 3' terminus of the genomic RNA has a poly(A)-tail and the 5' terminus has a cap structure.

The particles of ASPV (*Foveavirus*) readily form end-to-end aggregates with four prominent peaks appearing at 800 nm, 1600 nm, 2400 nm, and 3200 nm in length in purified preparation. ACLSV particles require the presence of divalent cations (for example, Mg^{2+}) to maintain the integrity of the quaternary structure. Potato virus T (*Tepovirus*) particles stained with uranyl formate showed cross-banding structure similar to ASGV (*Capillovirus*), whereas those stained with uranyl acetate showed 'criss-cross' and rope structure.

Genome Organization

The number of open reading frames (ORFs) encoded by viral genomes is between two and six depending on the genus (Table 2 and Fig. 3). In all species, the ORF1-encoded product has homologies with polymerase proteins of the 'alphavirus-like' supergroup of RNA viruses. This protein (Rep) (190–250 kDa) contains the conserved domains for methyl-transferase (Met), helicase (Hel), and RNA-dependent RNA polymerase (Pol) activity (Fig. 3). Most members also have a DNA alkylation damage repair protein (AlkB) and/or a papain-like protease (P-pro) domains between the Met and Hel (Fig. 3). Smaller ORFs encode the proteins involved in cell-to-cell movement, either a 'triple gene block' (TGB) (*Carlavirus*, *Foveavirus*, and *Robigovirus*) or a single MP of the '30K' superfamily (*Capillovirus*, *Chordovirus*, *Citrivirus*, *Divavirus*, *Prunevirus*, *Tepovirus*, *Trichovirus*, and *Vitivirus*) (Table 2 and Fig. 3). These are usually located following (3' proximal) the polymerase but in the genomes of capilloviruses and divaviruses, ORF2 (MP) is nested within the ORF1 (Fig. 3). For vitiviruses, an extra ORF with unknown function is present between the polymerase and MP genes. Watermelon virus A (*Wamavirus*) also have an additional protein of unknown function between MP and CP (Fig. 3). The CP

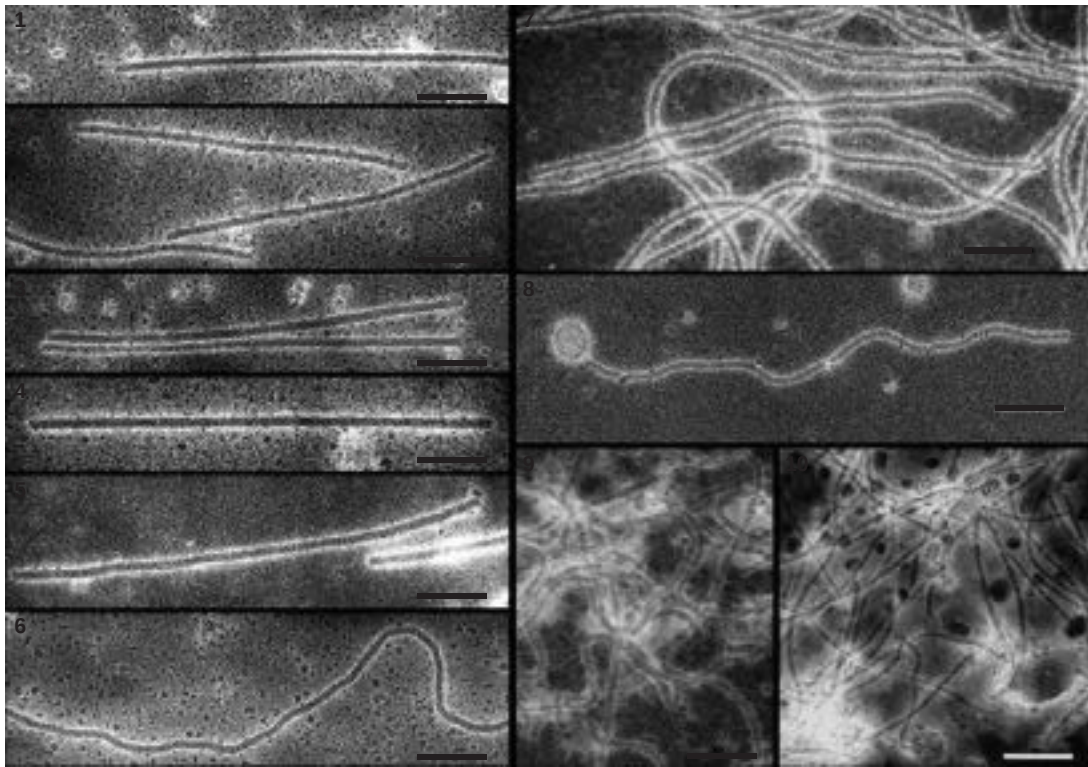


Fig. 2 Electron micrographs of (1) Zucchini yellow mosaic virus (*Potyvirus*), (2) Potato virus X (*Potexvirus*), (3) Potato virus S (*Carlavirus*), (4) Southern potato latent virus (*Carlavirus*), (5) Strawberry pseudo mild yellow edge virus (*Carlavirus*), (6) Citrus tristeza virus (*Closterovirus*), (7) Apple stem pitting virus (*Foveavirus*), (8) Apple chlorotic leaf spot virus (*Trichovirus*), (9) Grapevine virus A (*Vitivirus*), and (10) Apple stem grooving virus (*Capillovirus*). Bars represent 100 nm. (7) Courtesy of H. Koganezawa, (9) Courtesy of J. Imada.

gene always follows a single MP or TGB except a member of the genus *Wamavirus*. There is also another ORF at the 3' terminus in the genera *Carlavirus*, *Prunevirus*, and *Vitivirus* and in some species of *Trichovirus*. This latter ORF encodes a putative RNA-binding protein (NBP) with a zinc binding finger motif and the ability to bind nucleic acids (Fig. 3).

Properties and Functions of Gene Products

All members in the family *Betaflexiviridae* encode functional proteins for replication (Rep), cell-to-cell movement (a single MP or TGB), and protection of the genome RNA (CP). A small RNA binding protein is also encoded in all, or some members, in the genera *Carlavirus*, *Prunevirus*, *Vitivirus*, and *Trichovirus*. As an additional function, an RNA silencing suppressor activity is found in some proteins. For PVM in the genus *Carlavirus*, both a TGP protein 1 and a cysteine-rich 11K protein (ORF6) act as viral suppressor for local and systemic silencing, respectively. An ORF5 protein of Grapevine virus A (GVA) in the genus *Vitivirus* also suppresses local and systemic silencing. On the other hand, MPs of Citrus leaf blotch virus in the genus *Citrivivirus* and ACLSV in the genus *Trichovirus* have activity of silencing suppressors in which the former suppresses both local and systemic silencing, and the latter inhibits only systemic silencing.

Replication and Propagation

Members in the family *Betaflexiviridae* are alleged to replicate in the cytoplasm of infected cells. Virus particles are observed as aggregates in the cytoplasm but not frequently. Viral genomic RNA is infectious and serves as a viral messenger RNA. The ORF1 protein (Rep) is translated directly from the genomic RNA and probably the products are proteolytically processed by a P-Pro in Rep for maturation. The other ORFs located in the 3' terminus of the genome are expressed from sub-genomic mRNAs (sgRNAs). Northern blot hybridization analysis of virus-specific double-stranded (dsRNA) and ssRNAs in ASPV (*Foveavirus*)-infected tissues showed the presence of five distinct dsRNAs (ca. 9 kbp, 7.5 kbp, 6.5 kbp, 2.6 kbp, and 1.6 kbp) and three ssRNA (ca. 9 kb, 2.6 kb, and 1.6 kb) in infected tissues (Fig. 4). The slowest migrating ssRNA (9 kb) was equivalent to that of the ASPV genome,

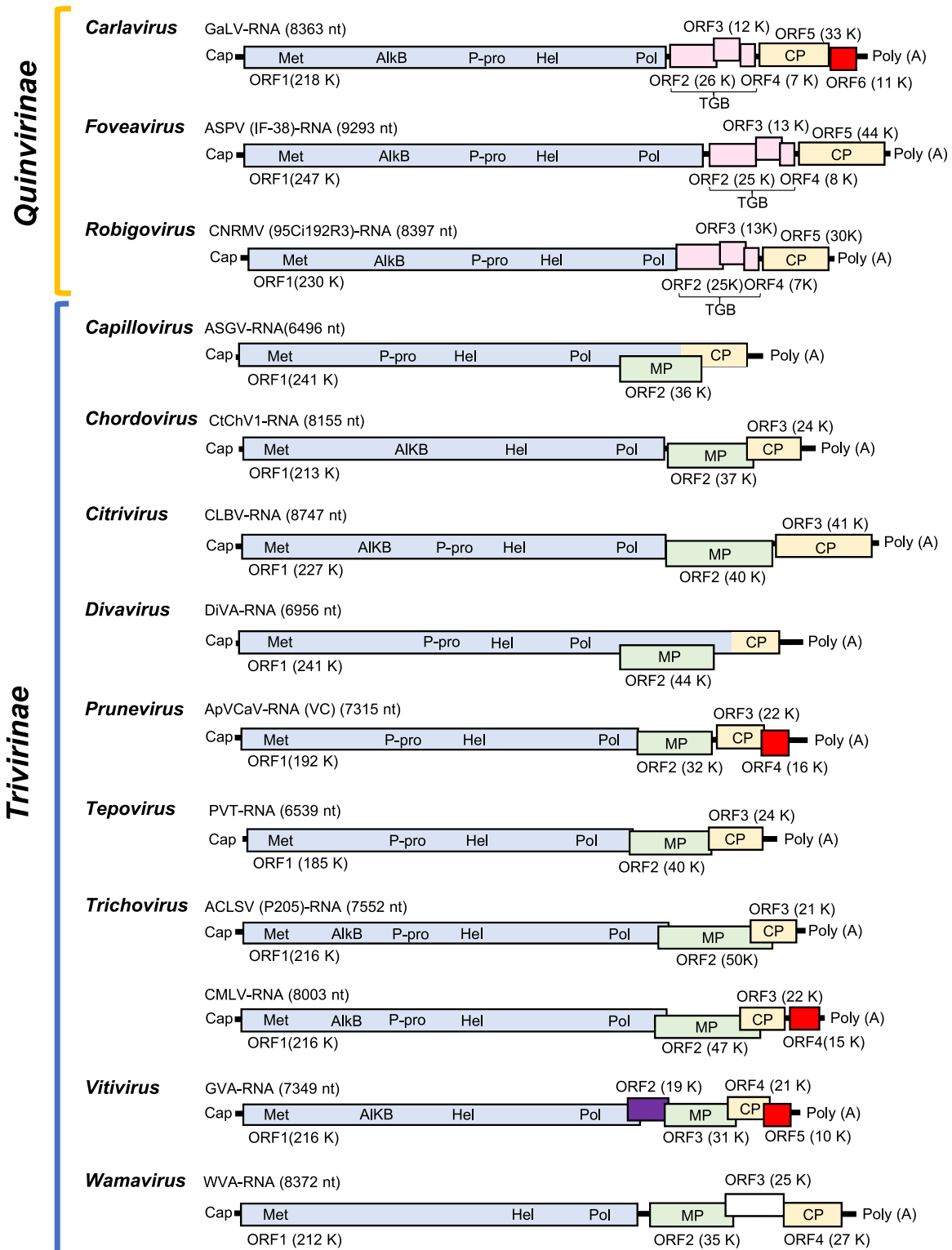


Fig. 3 Gemome organization of representatives of virus species from the twelve genera in the family *Betaflexiviridae*. GaLV, Garlic latent virus; ASPV, Apple stem pitting virus; CNRMV, Cherry necrotic rusty mottle virus; ASGV, Apple stem grooving virus; CtChV1, Carrot Ch virus 1; CLBV, Citrus leaf blotch virus; DiVA, Diuris virus A; ApVCaV, Apricot vein clearing associated virus; PVT, Potato virus T; ACLSV: Apple chlorotic leaf spot virus; CMLV, Cherry mottle leaf virus; GVA, Grapevine virus A; WVA, Watermelon virus A. Met, methyltransferase; AlkB, a DNA alkylation damage repair protein; P-pro, papain-like protease; Hel, nucleotide triphosphate-binding helicase; Pol, RNA dependent RNA polymerase; CP, coat protein; MP, movement protein; TGB, triple gene block.

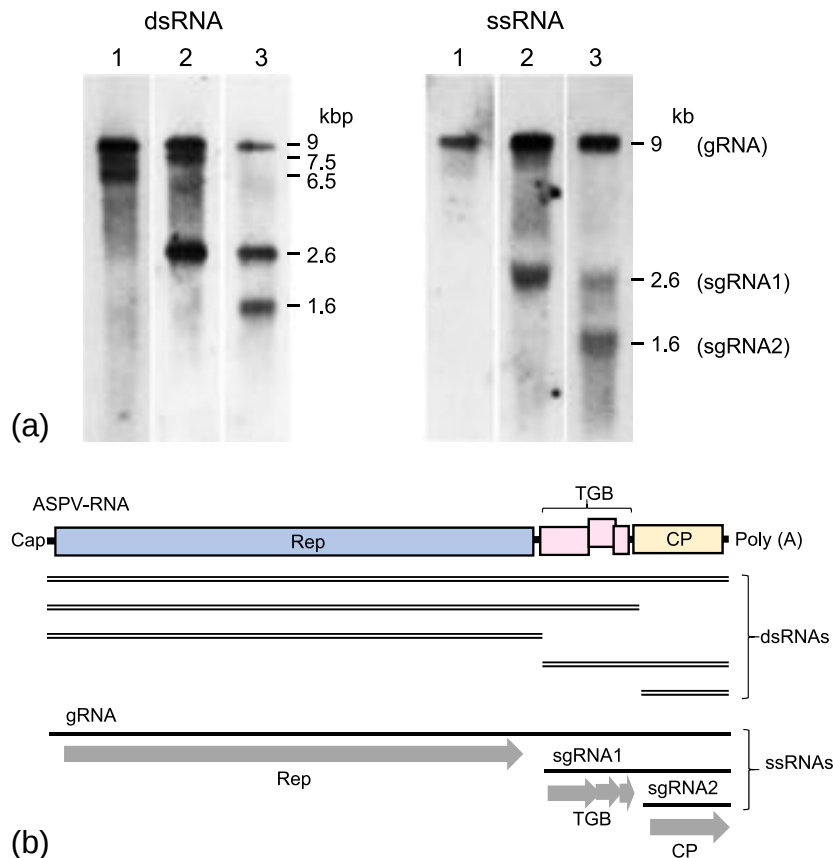


Fig. 4 (a) Northern blot analysis of double-stranded RNA (dsRNA) and single-stranded RNA (ssRNA) in leaves of *Nicotiana occidentalis* infected with Apple stem pitting virus (ASPV IF-38) using three different RNA probes specific to different sequences within the genomic RNA. Lane 1: a probe complementary to nt positions 6207 to 6683 (ORF1 region), lane 2; nt positions 6678 to 7447 (ORF2 region), and lane 3; nt positions 8717 to 9293 (CP and the 3' non-coding regions). gRNA, genomic RNA; sgRNA, sub-genomic RNA. (b) The predicted positions of ASPV-specific dsRNAs, genomic RNA (gRNA) for Rep, sub-genomic mRNA1 (sgRNA1) for TGB, and sgRNA2 for CP.

and other two ssRNAs (2.6 kb and 1.6 kb) are thought to be sgRNAs of TGB proteins (ORF2–4) and CP (ORF5), respectively. The dsRNA species with 7.5 kbp and 6.5 kbp, the function of which are unknown, are 5' co-terminal with genomic dsRNA. Similar patterns of dsRNAs and the 3' terminal sgRNAs are reported in infections with members of the genera *Capillovirus*, *Carlavirus*, and *Trichovirus*.

Transmission and Host Range

Most of the viruses in the family *Betaflexiviridae* are transmitted by mechanical inoculation, although the difficulty varies depending on the species. Transmission by grafting and dispersal through propagating materials is common in all species, especially in viruses that occur in woody fruit trees. Most members in the genus *Carlavirus* are transmitted by aphids in the non-persistent manner. Cucumber vein-clearing virus (*Carlavirus*) is transmitted by whiteflies. No vectors of any species have been reported for viruses in the genera *Foveavirus*, *Robigovirus*, *Capillovirus*, *Divavirus*, *Citriovirus*, *Chodovirus*, *Tepovirus*, and *Prunovirus*. However, some trichoviruses, i.e., Cherry mottle leaf virus and Grapevine berry inner necrosis virus are known to be transmitted by the peach bud mite *Eriophyes insidiosus* and the grape mite *Colomerus vitis*, respectively. For vitiviruses, both GVA and Grapevine virus B are transmitted in nature by several species of the pseudococcid mealybug genera *Planococcus* and *Pseudococcus* in a semi-persistent manner, whereas Heracleum latent virus is transmitted from naturally infected hogweed plants by aphids in a semi-persistent manner, which depends on a helper virus present in naturally infected plants. ASGV (*Capillovirus*) has been known to be transmitted through seeds to progeny seedlings of lily (1.8%) and *Chenopodium quinoa* (2.5%–60%). PVT (*Tepovirus*) is readily transmitted through true potato seed and pollen and spreads to tubers produced by infected plants. With the exception of most viruses in the genus *Carlavirus*, many viruses in this family use woody plants (fruit trees) as hosts, for example, rosaceae fruit trees for foveaviruses, robigoviruses, capilloviruses, prunoviruses, tepovirus, and trichoviruses, grapevines for foveaviruses, trichoviruses, and vitiviruses,

and citrus plants for capillovirus and citrivirus. Some viruses cause diseases in fruit trees, but many viruses are infecting fruit trees without showing obvious disease.

Geographical Distribution

Geographical distribution of viruses in the family *Betaflexiviridae* ranges from wide to restricted according the genus and the species. For example, ACLSV, ASGV, and APSV occurs in all countries where apples are grown, in contrast that GINV occurs only in Japan.

See also: Quinviruses (*Betaflexiviridae*). Triviruses (*Betaflexiviridae*)

Further Reading

- Adams, M.J., Antoniw, J.F., Bar-Joseph, M., *et al.*, 2016. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses. *Archives of Virology* 161, 2921–2949.
- Adams, M.J., Candresse, T., Hammond, J., *et al.*, 2012. Family *Betaflexiviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., *et al.* (Eds.), *Virus Taxonomy-Ninth Report on the International Committee on Taxonomy of Viruses*. Cambridge, MA: Elsevier Academic Press, pp. 920–941.
- Marais, A., Faure, C., Mustafayev, E., *et al.*, 2015. Characterization of new isolates of *Apricot vein clearing-associated virus* and of a new *Prunus*-infecting virus: Evidence for recombination as a driving force in *Betaflexiviridae* evolution. *Plos One*. doi:10.1371/journal.Pone.0129469.
- Rubino, L., Russon, M., De Stradis, A., *et al.*, 2012. *Tepovirus*, a novel genus in the family *Betaflexiviridae*. *Archives of Virology* 157, 1629–1633.
- Svanella-Dumas, L., Tsarmpopoulos, I., Marais, A., *et al.*, 2018. Complete genome sequence of Lettuce chardovirus 1 isolated from cultivated lettuce in France. *Archives of Virology* 163, 2543–2545.
- Villamor, D.V., Druffel, K.L., Eastwell, K.C., 2013. Complete nucleotide sequence of a virus associated with Rusty mottle disease of sweet cherry (*Prunus avium*). *Archives of Virology* 158, 1805–1810.
- Vives, M.C., Galipienso, L., Navarro, L., *et al.*, 2001. The nucleotide sequence and genome organization of Citrus leaf blotch virus: Candidate type species for a new virus genus. *Virology* 287, 225–233.
- Wylie, S.J., Li, H., Dixon, K.W., *et al.*, 2013. Exotic and indigenous viruses infect wild populations and captive collections of temperate terrestrial orchards (*Diuris* species) in Australia. *Virus Research* 171, 22–32.

Betasatellites and Deltasatellites (*Tolecusatellitidae*)

Muhammad S Nawaz-ul-Rehman, University of Agriculture, Faisalabad, Pakistan

Nazia Nahid, GC University, Faisalabad, Pakistan and University of Agriculture, Faisalabad, Pakistan

Muhammad Hassan and Muhammad Mubin, University of Agriculture, Faisalabad, Pakistan

© 2021 Published by Elsevier Ltd.

Nomenclature

A Adenosine

CP Coat protein or capsid protein

GFP Green fluorescent protein

JA Jasmonic acid

MAPK Mitogen activated protein kinase

NLS Nuclear localization signal

nt Nucleotide(s)

ORF Open reading frame

PASC Pairwise sequence comparison

RCR Rolling-circle replication

RDR6 RNA-dependent RNA polymerase 6

SCR Satellite conserved region

Glossary

Betasatellite ssDNA satellite molecules about 1300 nt long and coding for a pathogenicity determinant protein, which depends on a geminivirus helper for replication, movement, encapsidation and transmission.

Deltasatellite Small non-coding ssDNA satellites about 700 nt long, associated with helper begomoviruses.

Iterons High-affinity binding sites located upstream of the stem-loop of geminiviruses.

Pathogenicity determinant protein Any protein coded by virus or the satellite which determines the virulence of the pathogen.

Satellite conserved region A highly conserved region of ~ 80 nucleotides, which is integral part of all the betasatellites.

Satellite DNA Subviral agent consisting of ssDNA that becomes packaged in protein shells made from coat protein of the helper virus and whose replication, movement and transmission is dependent on that virus.

Introduction

Betasatellites are small circular sub-viral components associated with geminiviruses and are mainly confined to the Old World. The word “beta” was used to differentiate with the DNA-B component of begomoviruses and the word “satellite” was added because they are effectively satellites. Their discovery started from the identification of a satellite molecule associated with Tomato leaf curl virus in Australia (now classified under the genus *Deltasatellite* of the family *Tolecusatellitidae*). Betasatellites have very conserved sequence structure or genome organization which includes an adenosine rich (A-rich) region, a satellite conserved region (SCR) and a small open reading frame coding for a protein called β C1. Initially, they were associated with monopartite begomoviruses, but nowadays they are routinely identified in association with bipartite begomoviruses as well. So far, betasatellites have been limited to the Old World, however a recent discovery suggests that a defective form of an African-origin betasatellite (Cotton leaf curl Gezira betasatellite – lacking β C1) has been found in the United States.

The DNA-A component alone of begomoviruses, in many cases, is sufficient to induce the disease symptoms. Nevertheless, for an increasing number of begomoviruses, betasatellites are crucial to induce the typical disease symptoms. For example, Cotton leaf curl Gezira virus in Africa, Cotton leaf curl Multan virus in Pakistan and Tomato yellow leaf curl China virus in China require the presence of a betasatellite for disease development. Betasatellites are trans-replicated and encapsidated by their helper virus Rep protein and coat protein (CP) respectively. In the laboratory conditions Old World betasatellites can replace the functions of a DNA-B component. However, New World begomoviruses can also maintain the betasatellites without compromising the function of the DNA-B component.

The first betasatellite was characterized for the Ageratum yellow vein disease in 2000. The presence of the betasatellite increased viral symptoms in ageratum plants infected with the geminivirus. Subsequently several betasatellites were characterized from Asia and Africa. For several diseases in tomato, tobacco, cotton, chillies, and okra, the helper viruses do not require the absolute presence of a betasatellite for the disease development. To date, more than 1880 sequences of betasatellites are present in GenBank. The betasatellites not only interfere with the normal plant development, but they act as a suppressor of transcriptional and post-transcriptional gene silencing.

Classification and Nomenclature

Betasatellites and deltasatellites are satellite molecules, therefore dependent on a helper virus for their replication and/or other biological functions. Their classification is based on their genomic nucleic acid sequence. Since 2016, the species demarcation threshold for beta- and deltasatellites has been set to 91% nucleotide sequence identity. This species demarcation threshold will be soon revised as it seems

Table 1 List of species of the genera *Betasatellite* and *Deltasatellite* of the family *Tolecusatellitidae*. A list of betasatellite unassigned species is indicated. The acronym and isolate ID of the reference virus for each species, as well as the NCBI GeneBank accession number and the length of the satellite are also provided. Type species for each genus are written in bold

Genus/Species	Acronym-[isolate ID]	Accession #	Length (nt)
<i>Betasatellite</i>			
<i>Ageratum</i> leaf curl Buea betasatellite	ALCBueB-[CM: LI01: SatB33:09]	NC014746	1259 nt
<i>Ageratum</i> leaf curl Cameroon betasatellite	ALCCMB-[CM: Man:AMBF:06]	NC012557	1389 nt
<i>Ageratum</i> yellow leaf curl betasatellite	AYLCB-[PK: Fai4:00]	NC005046	1351 nt
<i>Ageratum</i> yellow vein betasatellite^a	AYVB-[SG:95]	NC003403	1347 nt
<i>Ageratum</i> yellow vein India betasatellite	AYVINB-[IN: Mad:03]	NC043428	1353 nt
<i>Ageratum</i> yellow vein Sri Lanka betasatellite	AYVLKB-[LK:Ag:03]	NC043429	1351 nt
<i>Alternanthera</i> yellow vein betasatellite	AIYVB-[VN: Hue:05]	NC009562	1344 nt
<i>Andrographis</i> yellow vein leaf curl betasatellite	AnYVLCB-[IN: Luc:10]	NC023876	1379 nt
<i>Bhendi</i> yellow vein mosaic betasatellite	BYVB-[IN: Mut:00]	NC014895	1373 nt
<i>Cardiospermum</i> yellow leaf curl betasatellite	CaYLCB-[SL:04]	NC010297	1338 nt
<i>Chili</i> leaf curl betasatellite	ChLCB-[PK: MC:97]	NC005048	1387 nt
<i>Chili</i> leaf curl Jaunpur betasatellite	ChLCJauB-[IN: Jau:07]	HM007103	1362 nt
<i>Chili</i> leaf curl Sri Lanka betasatellite	ChLCLKB-[LK: Mih:09]	NC038676	1371 nt
<i>Cotton</i> leaf curl Gezira betasatellite	CLCGezB-[SD: Dat:06]	NC006935	1348 nt
<i>Cotton</i> leaf curl Multan betasatellite	CLCMulB-[PK: Mul:U89:97]	NC009535	1346 nt
<i>Croton</i> yellow vein mosaic betasatellite	CroYVMB-[PK: Pun:06]	NC008579	1346 nt
<i>Eupatorium</i> yellow vein betasatellite	EpYVB-[JR: MNS2:00]	NC004515	1356 nt
<i>Eupatorium</i> yellow vein mosaic betasatellite	EpYVMB-[JP: Suya:03]	NC038677	1359 nt
<i>French bean</i> leaf curl betasatellite	FBLCB-[IN: Kan:11]	NC018091	1379 nt
<i>Hedyotis</i> yellow mosaic betasatellite	HeYMB-[VN: BinhDinh:13]	NC023015	1348 nt
<i>Honeysuckle</i> yellow vein betasatellite	HYVB-[UK: Nor1:99]	NC005052	1344 nt
<i>Honeysuckle</i> yellow vein mosaic betasatellite	HYVMB-[JP: Hy:04]	NC005953	1262 nt
<i>Malvastrum</i> leaf curl betasatellite	MaLCB-[CN: Gx87:04]	NC007711	1354 nt
<i>Malvastrum</i> leaf curl Guangdong betasatellite	MaLCGuB-[CN: Gu:11]	NC023896	1344 nt
<i>Mirabilis</i> leaf curl betasatellite	MiLCB-[IN: Him:13]	NC038924	1367 nt
<i>Momordica</i> yellow mosaic betasatellite	MamYMB-[BJ: 57:14:14]	KT454829	1367 nt
<i>Mungbean</i> yellow mosaic betasatellite	MYMB-[IN: Cowpea:12]	NC018869	1360 nt
<i>Okra</i> leaf curl Oman betasatellite	OLCOMB-[OM: Barka:12]	NC038678	1351 nt
<i>Papaya</i> leaf curl betasatellite	PaLCB-[IN: ND:03]	NC004706	1372 nt
<i>Papaya</i> leaf curl China betasatellite	PaLCCNB-[CN: Hn:14]	NC038679	1350 nt
<i>Papaya</i> leaf curl India betasatellite	PaLCINB-[IN: Pan:08]	HM143906	1333 nt
<i>Rhynchosia</i> yellow mosaic betasatellite	RhYMB-[IN: Pha:14]	NC038681	1359 nt
<i>Rose</i> leaf curl betasatellite	RoLCB-[PK: Fai:06]	NC024695	1349 nt
<i>Siegesbeckia</i> yellow vein betasatellite	SiYVB-[CN: FZ02:12]	NC038682	1359 nt
<i>Tobacco</i> curly shoot betasatellite	TbCSB-[CN: Yn35:01]	NC004546	1354 nt
<i>Tobacco</i> leaf curl betasatellite	TbLCB-[PK: Lah:04]	NC038891	1358 nt
<i>Tobacco</i> leaf curl Japan betasatellite	TbLCJPB-[JP: Miy:05]	NC009450	1350 nt
<i>Tobacco</i> leaf curl Patna betasatellite	TbLCPatB-[IN: Pusa:09]	NC038683	1338 nt
<i>Tomato</i> leaf curl Bangalore betasatellite	ToLCBanB-[IN: Ban:03]	NC038684	1377 nt
<i>Tomato</i> leaf curl Bangladesh betasatellite	ToLCBDB-[BD: Gaz:01]	NC014594	1371 nt
<i>Tomato</i> leaf curl betasatellite	ToLCB-[PK: RYK:97]	NC004715	1424 nt
<i>Tomato</i> leaf curl China betasatellite	ToLCCNB-[CN: Gx14:02]	NC004544	1339 nt
<i>Tomato</i> leaf curl Gandhinagar betasatellite	ToLCGanB-[IN: pToGNbH14:12]	NC023038	1365 nt
<i>Tomato</i> leaf curl Java virus betasatellite	ToLCJavB-[NP: R7: Pap:10]	NC005497	1360 nt
<i>Tomato</i> leaf curl Joydebpur betasatellite	ToLCJoyB-[BD: Gaz:05]	NC010236	1370 nt
<i>Tomato</i> leaf curl Laguna betasatellite	ToLCLagB-[PH: Lag2:06]	NC038685	1346 nt
<i>Tomato</i> leaf curl Laos betasatellite	ToLCLAB-[LA: Sav:01]	NC038925	1348 nt
<i>Tomato</i> leaf curl Malaysia betasatellite	ToLCMYB-[MY:13]	NC038686	1342 nt
<i>Tomato</i> leaf curl Nepal betasatellite	ToLCNPB-[NP: Jhapa]	NC038926	1350 nt
<i>Tomato</i> leaf curl Patna betasatellite	ToLCPatB-[IN: Pat:07]	NC012493	1349 nt
<i>Tomato</i> leaf curl Philippine betasatellite	ToLCPHB-[PH: Lag1:06]	NC009570	1349 nt
<i>Tomato</i> leaf curl Sri Lanka betasatellite	ToLCLKB-[LK]	AJ542493	1371 nt
<i>Tomato</i> leaf curl Yemen betasatellite	ToLCYEB-[YE: tob56:89]	NC018864	1352 nt
<i>Tomato</i> yellow leaf curl China betasatellite	TYLCCNB-[CN: Yn45:01]	NC019532	1335 nt
<i>Tomato</i> yellow leaf curl Rajasthan betasatellite	ToLCRajB-[IN: Raj:03]	NC038687	1371 nt
<i>Tomato</i> yellow leaf curl Shandong betasatellite	ToYLCShB-[CN: SDSG:14]	NC038688	1334 nt
<i>Tomato</i> yellow leaf curl Thailand betasatellite	TYLCTHB-[CN: Yn72:02]	NC004903	1337 nt
<i>Tomato</i> yellow leaf curl Vietnam betasatellite	TYLCVNB-[VN: Han:05]	NC009560	1356 nt

Table 1 Continued

Genus/Species	Acronym-[isolate ID]	Accession #	Length (nt)
<i>Tomato yellow leaf curl Yunnan betasatellite</i>	ToYLCYnB-[CN: tob:]	NC038689	1337 nt
<i>Vernonia yellow vein betasatellite</i>	VYVB-[IN: Mad:09]	NC013423	1364 nt
<i>Vernonia yellow vein Fujian betasatellite</i>	VYVFuB-[CN:09]	NC015928	1377 nt
Unassigned species			
<i>Datura leaf curl betasatellite</i>	DaLCB-[IN: Lak:Raj:11]	JQ693149	1042 nt
<i>Emilia yellow vein betasatellite</i>	EmYVB-[V: Hoa:06]	NC012666	1337 nt
<i>Leucas zeylanica yellow vein betasatellite</i>	LzYVB-[LK:06]	NC013424	1363 nt
<i>Lindernia anagallis yellow vein betasatellite</i>	LaYVB-[VN: Han:06]	NC009561	1346 nt
<i>Okra yellow leaf curl betasatellite</i>	OYLCB-[BF: Kaya:Sida30BA:14]	MK032309	1356 nt
<i>Pea leaf distortion betasatellite</i>	PeaLDB-[NP:10]	NC033618	1347 nt
<i>Sida leaf curl betasatellite</i>	SiLCB-[CN: Hai:Hn57:16]	NC007639	1365 nt
<i>Tobacco leaf chlorosis betasatellite</i>	TbLChB-[IN:10]	NC018935	1345 nt
<i>Tobacco leaf curl Yunnan betasatellite</i>	TbLCYnB-[CN: Yn:12]	NC005030	1349 nt
<i>Tomato leaf curl Yunnan betasatellite</i>	ToLCYnB-[CN: Yn:Yn4980:15]	NC040544	1361 nt
<i>Tomato yellow leaf curl Kanchanaburi betasatellite</i>	ToLCKanB-[ID: Kan:Egg: BL1:19]	MK936043	1323 nt
<i>Zinnia leaf curl betasatellite</i>	ZLCB-[TH: Pat:Tz:03]	NC005874	1354 nt
Deltasatellite			
<i>Croton yellow vein deltasatellite</i>	CrYVD-[IN: Mad:04]	AJ968684	739 nt
<i>Malvastrum leaf curl deltasatellite</i>	MalLCD-[PH: Mc1:12]	NC021929	673 nt
<i>Sida golden yellow vein deltasatellite 1</i>	SiGYVD:1-[CU:H1:09]	NC043200	687 nt
<i>Sida golden yellow vein deltasatellite 2</i>	SiGYVD:2-[CU:228H1:09]	NC043201	677 nt
<i>Sida golden yellow vein deltasatellite 3</i>	SiGYVD:3-[CU:412N1:10]	NC043202	686 nt
<i>Sweet potato leaf curl deltasatellite 1</i>	SPLCD:1-[ES: SBG51:09]	NC016536	662 nt
<i>Sweet potato leaf curl deltasatellite 2</i>	SPLCD:2-[VN: Mer:1764E13:09]	NC025220	733 nt
<i>Sweet potato leaf curl deltasatellite 3</i>	SPLCD:3-[PR: PR3:2:09]	NC043203	731 nt
Tomato leaf curl deltasatellite	ToLCD-[AU:96]	NC002743	682 nt
<i>Tomato yellow leaf distortion deltasatellite 1</i>	ToYLDD:1-[CU: Sida:404N1:10]	NC043204	685 nt
<i>Tomato yellow leaf distortion deltasatellite 2</i>	ToYLDD:2-[CU:603N1:11]	NC043205	694 nt

^aUnassigned betasatellite species.

that the distribution of pairwise comparison sequence (PASC) values indicate a clear difference between isolates, species and genera, pointing to a cut-off value closer to 73% for species demarcation. Currently the ICTV recognizes 61 species of betasatellites and 11 species of deltasatellites. In addition, 12 tentative species of betasatellites are considered for official recognition (Table 1, Fig. 1). Recently the family *Tolecusatellitidae* has been created to host betasatellites and deltasatellites. The word *Tolecusatellitidae* is coined from the first ever identified satellite in Australia (Tomato leaf curl virus-satellite). The nomenclature of betasatellites and deltasatellites species names is so far based on the name of the first helper geminivirus with which the satellite was identified, complemented, if necessary, with the geographical location of the first sample and a number if several species are identified on the same host with the same helper geminivirus. The family encompasses two genera, *Betasatellite* and *Deltasatellite*, and *Ageratum yellow vein betasatellite* and *Tomato leaf curl deltasatellite* are the type species for each genus respectively. Betasatellites and deltasatellites do not share any sequence similarity to each other nor with their helper viruses, except the nonanucleotide origin of replication (TAATATTAC).

Genome Organization

All the known betasatellites have conserved sequence features with an A-rich region (~160–280 nt), a satellite conserved region (SCR, ~80 nt) and a single open reading frame (ORF) coding for a protein called β C1 (~13.6 kDa) (Fig. 2). While deltasatellites lack the β C1 gene and are roughly half the size (~700 nt) of betasatellites (Fig. 2). The A-rich region is involved in the maintenance of betasatellite integrity, as its deletion can cause milder symptoms in the host plant. SCR is involved in interaction with the Rep protein of the helper virus, while β C1 is a multifunctional pathogenicity determinant protein, which interacts with several host factors to counter the host defense system by suppressing both transcriptional and post-transcriptional gene silencing. Deltasatellite do not code for any functional protein and are fully dependent on their helper virus.

Genetic Diversity and Center of Origin for Betasatellites

Betasatellites are widely distributed in the Old World mainly with monopartite begomoviruses. Due to the lack of sequence similarity with their helper viruses or other related viruses, the origin of betasatellites is unknown. Betasatellites are present in more than

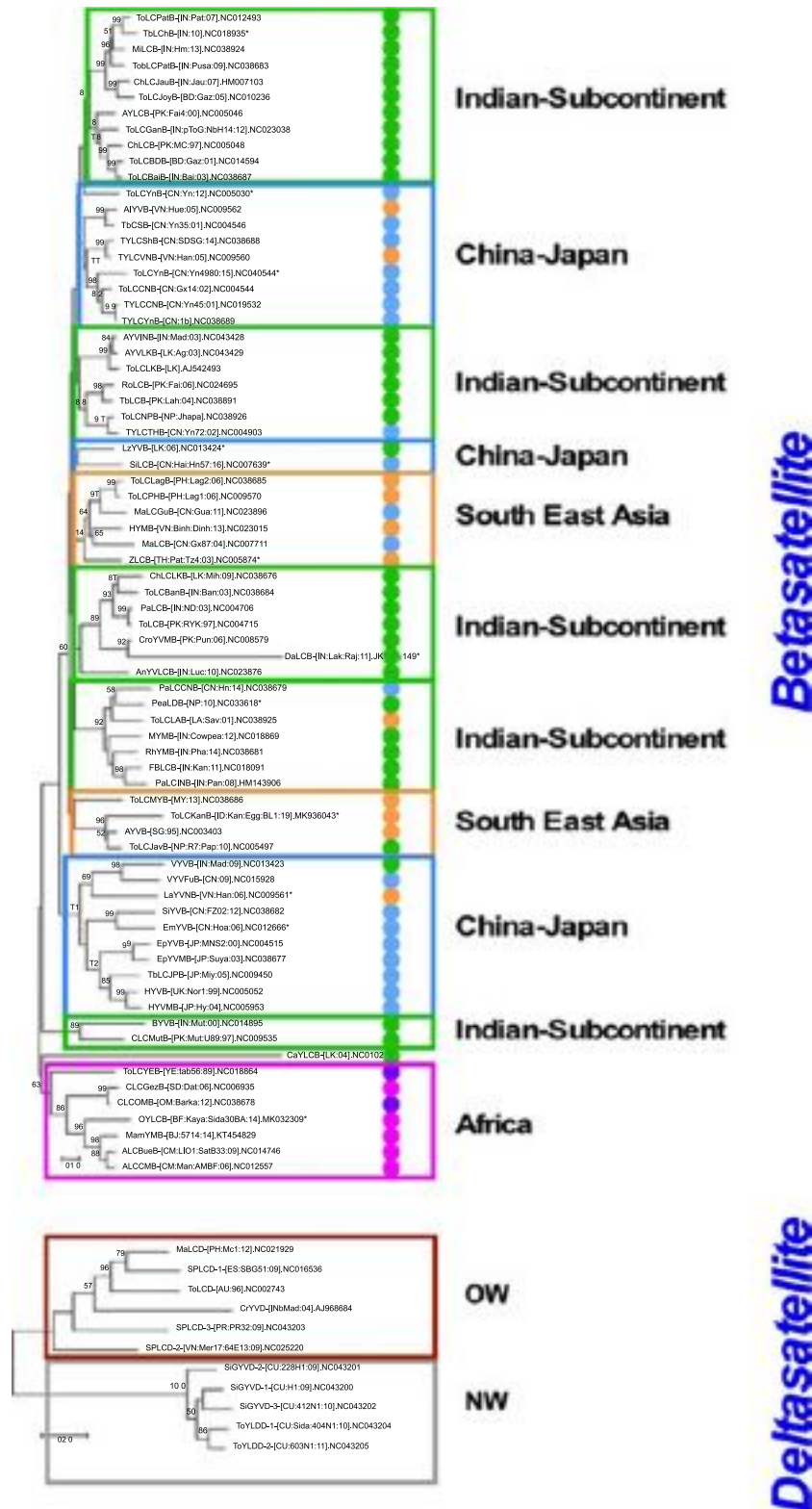


Fig. 1 Phylogenetic trees for betasatellites and deltasatellites in the family *Tolecosatellitidae*. The acronyms are listed in Table 1. Betasatellites have ~61 recognized species and 11 putative unassigned species (noted with *), while deltasatellites consist of 11 different recognized species. The reference sequences of each species were downloaded and aligned in MEGA software. The alignment was made through Clustal-W. Phylogenetic trees were constructed by neighbor joining method with 1000 bootstrap replications, and only scores above 50% were noted.

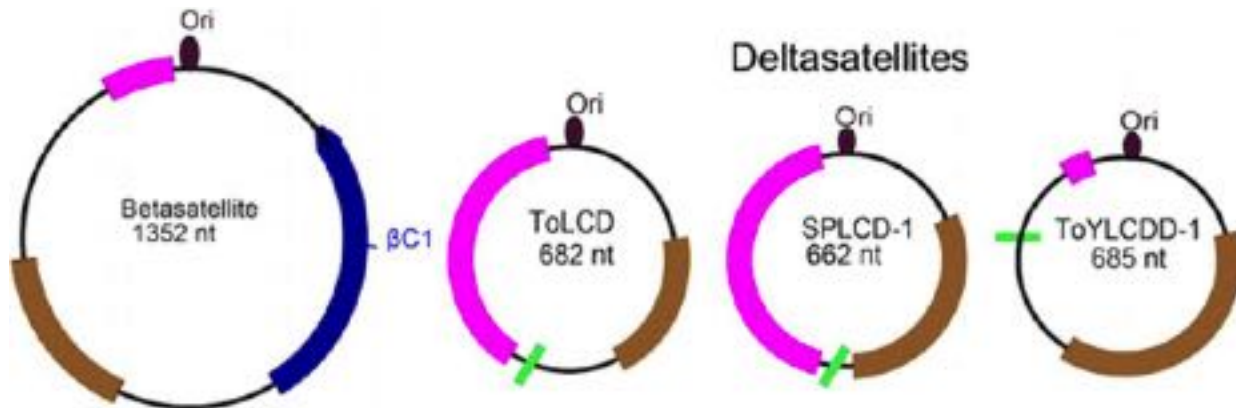


Fig. 2 Genome organization (left to right) of betasatellites and deltasatellites. Betasatellites are of ~1350 nt, with an A-rich region, a satellite conserved region and a small gene $\beta C1$. While, deltasatellites are of ~680 nt and do not encode any gene. The genome organizations of Tomato leaf curl deltasatellites (ToLCD) from Australia, Sweet potato leaf curl deltasatellite-1 (SPLCD-1) and Tomato yellow leaf curl distortion deltasatellite (ToYLCDD-1) from Cuba are presented. The nucleotides region mentioned with magenta colors represent the satellite-conserved region; brown color represents the A-rich region, while green parts represent the secondary stem loop in the sequence.

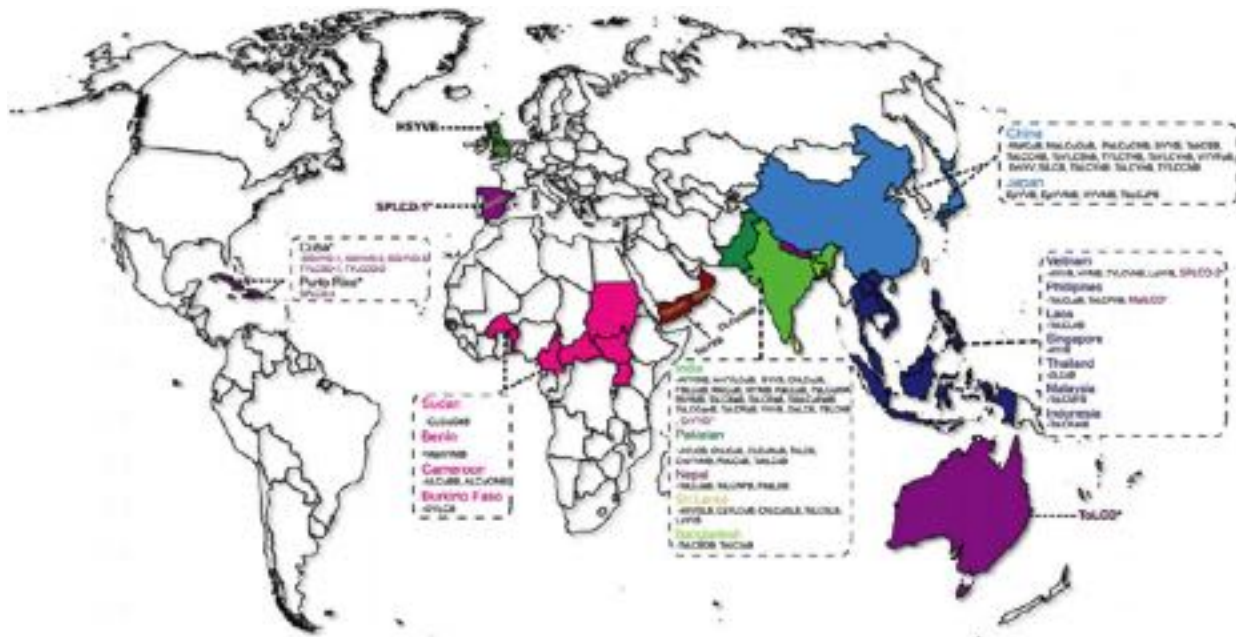


Fig. 3 Worldwide distribution for 61 species and 11 putative species of betasatellites and 11 species of deltasatellites. Most of the betasatellites species are present in the Indian sub-continent, while a second center of diversity appears in China and Japan. Betasatellites are limited to the Old World, while deltasatellites are present in both the Old World and the New World. (mentioned with *).

20 countries of Asia, Africa, and Europe and none have been identified in the New World (Fig. 3). Potentially, there are ~73 betasatellite species and they mostly belong to South, East, South East, and West Asia (67/73). Among these betasatellites, 35 species circulate in India, Pakistan, Bangladesh, Sri Lanka, and Nepal. While 19 betasatellite species are present in China and Japan. Another 11 distinct betasatellite species exist in Vietnam, Laos, Philippines, Singapore, Indonesia, and Malaysia. In addition, two different species infect okra and tomatoes in Oman and Yemen respectively. The whole African continent only hosts 5 different species of betasatellites that are phylogenetically close to the Oman and Yemen isolates (Fig. 1). While from Europe (United Kingdom) a single betasatellite species is known which was most probably imported from Japan (Fig. 1). A phylogenetic study carried with representatives of these 63 species reveals a very strong geographical distribution with 11 clusters almost 100% homogeneous for the regions mentioned above. Five clusters belong to the Indian subcontinent, 3 to the China-Japan region, 2 from south east Asia and 1 from Africa. Only 11 species are “misplaced” for having a perfect match. Therefore, considering the genetic diversity here described it can be concluded that the Indian sub-continent is possibly the center of origin of betasatellites (Figs. 1 and 3).



Fig. 4 Impact of betasatellite on symptom severity in *Nicotiana benthamiana* plants. The *Nicotiana benthamiana* plants inoculated with Cotton leaf curl Multan virus DNA-A alone show very mild symptoms (center), while the addition of the Cotton leaf curl Multan betasatellite induces very severe symptoms (right). Non-inoculated control plants are on the left.

Unlike betasatellites, deltasatellites have been also identified in the New World (Fig. 3), however it is believed that they were introduced through infected root crops such as sweet potato. A contrario to betasatellites, It is unclear what is the competitive advantage for the helper geminivirus to carry a deltasatellite along. Recent studies suggest that deltasatellites may reduce the geminivirus impact and thereby extend the life of the infected plant, thus allowing a greater period for the geminivirus transmission and survival of the host.

Life Cycle, Epidemiology

As mentioned earlier, the betasatellites are often found with monopartite begomoviruses. However, they are also found in combination with bipartite viruses such as Tomato leaf curl New Delhi virus, Tomato leaf curl Gujarat virus, and Sri Lankan cassava mosaic virus. In the laboratory conditions, betasatellites can also interact with the New World bipartite begomoviruses. Betasatellites are also known to interact with mastreviruses in wheat plants. These observations suggest that trans-replication of betasatellites is not very stringent in respect to the helper geminivirus. The experiments conducted on many economically important begomoviruses suggested that DNA-A alone could effectively produce the disease symptoms in *Nicotiana benthamiana* plants. However, in many other cases, like cotton leaf curl disease, Ageratum yellow vein virus and Eupatorium yellow vein virus, the betasatellites were necessary to induce the disease symptoms (Fig. 4). Further experiments suggested that the role of $\beta C1$ was important for disease induction. A mutated $\beta C1$ could not induce the proper disease symptoms in the inoculated plants, despite of its successful replication in the systemic leaves. The co-inoculation of betasatellite with bipartite begomoviruses like Tomato leaf curl New Delhi virus and Tomato leaf curl Gujarat virus resulted in increased viral titer and exacerbated symptoms. The presence of betasatellites in non-cultivated cotton species did not result in increased symptoms. Therefore it can be concluded that betasatellites result in severe disease symptoms in cultivated crops.

Normally bipartite begomoviruses are accompanied by a DNA-B component, which provides the cell-to-cell movement functions for the virus. The experiments conducted on Sri Lankan cassava mosaic virus and Tomato leaf curl New Delhi virus suggested that betasatellites could make DNA-B dispensable to the helper virus. The exact mechanism of virus movement in the absence of DNA-B is unknown. There are however few indirect evidences, which suggest the possible role of $\beta C1$ encoded by betasatellite in viral movement. The green florescent protein (GFP) tagged to $\beta C1$ of TYLCCNB and CLCMulB were transiently expressed in tobacco cells. The $\beta C1$ accumulated inside the nucleus and the cell periphery, suggesting its possible role in transport of DNA from nucleus to plasmodesmata. However so far, all the experiments for DNA-B replacement were conducted in laboratory conditions, and there is no such example in nature, where a bipartite virus was converted into monopartite virus by capturing a betasatellite, while loosing its DNA-B component. However, there are recent examples, where a cognate betasatellite was replaced by a non-cognate one. For examples, Tomato leaf curl Karnataka virus in India adopted Cotton leaf curl Multan betasatellite replacing Tomato leaf curl Karnataka betasatellite. TYLCV-IR is a monopartite begomovirus in the Middle East, but in Oman, it captured Tomato leaf curl betasatellite. A similar example of re-assortment also exists in Mali, where Cotton leaf curl Gezira betasatellite was found in combination with the monopartite Tomato yellow leaf curl Mali virus.

Replication of Betasatellites

The viruses belonging to family *Geminiviridae* multiply through rolling circle replication (RCR) mechanism. RCR starts by the binding of the Replication associated protein (Rep) at iteron sequences present at the upstream of the nonanucleotide origin of replication. Despite using the same Rep the trans-replication of betasatellites is very relaxed as compared to DNA-B component. A single betasatellite can be trans-replicated by several helper viruses. For example, Cotton leaf curl Multan betasatellite (CLCMulB) can be

maintained in *Nicotiana benthamiana* plants by Ageratum yellow vein virus and Eupatorium yellow vein virus, but Cotton leaf curl Multan virus shows specificity to its cognate betasatellite. The SCR of betasatellites present in the upstream of origin of replication appears to be the analogous of the common region of DNA-A components. As a matter of fact all the defective betasatellites or deltasatellites contain intact SCR indicating its vital role in their replication. The experiments conducted on Tomato leaf curl virus associated betasatellites suggested that the stem loop structure and the satellite conserved region are very crucial sequences for their replication. The betasatellites contain a primary stem loop structure (common to all geminiviruses) and a secondary stem loop structure ~380 nt upstream of origin of replication. The Rep binding affinity experiments suggested that the secondary stem loop structure is very crucial for betasatellites replication as its deletion can cause impaired replication. Despite of their relaxed trans-replication betasatellites have certain unknown specificities with their helper begomoviruses. The Rep binding affinity experiments suggested that helper components have better adaptation with their cognate betasatellites compared to non-cognate ones.

Functional Role of β C1 in Disease Development

The betasatellites strongly influence the pathogenicity of the helper begomovirus. The presence of betasatellites are resulting in very severe disease symptoms. The β C1 gene present in the complementary strand of betasatellite genomes encodes an important pathogenicity determinant protein for geminiviruses of the Old World. Most of the available information regarding β C1 is obtained from Tomato yellow leaf curl china betasatellite (TYLCCNB). The β C1 of TYLCCNB localizes in the nucleus and forms a multimeric complex due to two alpha helices present at the C-terminus of the protein. The self-interaction of β C1 is important to induce the disease symptoms. Indeed, a mutated version of β C1, that cannot make a multimeric complex, is unable to induce the typical disease symptoms. The nuclear localizing signal (NLS) of β C1 is also important in the disease induction. The deletion mutants for NLS in TYLCCNB resulted in its inability to induce typical disease symptoms.

The protein β C1 is a multitasking protein, which suppresses the transcriptional and post-transcriptional gene silencing, the whole hormone based defense system, and the ubiquitin proteome degradation system. It also inhibits the activity of kinases like mitogen activated protein kinases (MAPK). The pleiotropic interaction of β C1 with multiple pathways results in enhanced virulence and infection. The transgenic plants expressing the protein β C1 show severe disease-like symptoms. In the infected cells, β C1 interacts with the Asymmetric Leaves gene 1 (AS1), which can ultimately result in suppression of jasmonic acid (JA) production and its associated genes. The suppression of JA results in severe disease symptoms. The β C1 encoded by Radish leaf curl betasatellite (RaLCB) localizes in chloroplast and results in vein clearing symptoms. The symptoms were very severe in the presence of Tomato leaf curl New Delhi virus. The detailed biochemical analysis revealed that β C1 altered the starch accumulation, decreased the thylakoid contents and change the granum structure. β C1 counters the gene silencing initiated by plants as a defense against betasatellite and geminiviruses. The interaction of β C1 with Calmodulin-like (CaM-endogenous suppressors of gene silencing) proteins results in higher accumulation of the helper virus by compromising the RDR6 based gene silencing. β C1 is also a target of the ubiquitination based protein degradation pathway. It has evolved a system to hijack the ubiquitination machinery. The interaction of β C1 with host proteins resulting in reduction of polyubiquitinated proteins. Autophagy is another evolutionary conserved pathway, which is targeted by β C1. Autophagy is actively involved in controlling several viruses. Studies have shown that, autophagy pathways targeted and degraded the β C1. The silencing of autophagy related genes; *ATG5* and *ATG7* resulted in increased susceptibility to begomoviruses and betasatellites.

Resistance Against Betasatellites

So far, there is no example where geminiviruses/betasatellites have been controlled successfully at commercial level. However, limited success has been reported at the experimental levels. During the natural infection tobacco ring finger proteins (RFP- a functional E3 ubiquitin ligase) ubiquitinates and degrades the β C1 proteins. The impact of β C1 ubiquitination can be observed by attenuated symptoms. In the future, it is believed that exploiting such pathways to increase the plants immunity level against geminiviruses/betasatellites complexes, is possible. Recent studies conducted on Tomato yellow leaf curl virus demonstrated that *Ty-1* gene confers natural resistance against the virus. However, with the addition of a betasatellite the *Ty-1* based resistance was compromised and viral titer was increased. Due to the global spread of betasatellites the coordinated efforts between different countries are required.

Deltasatellites

Introduction

Deltasatellites were the first satellites identified with an example from Australia in association with Tomato leaf curl virus (ToLCV). Initially it was named ToLCV-Sat, but now it has been renamed as Tomato leaf curl deltasatellite (ToLCD). The ToLCD (682 nt long) did not share any significant sequence identity with the helper virus except for the origin of replication. The sequence variability of ToLCD is unknown due to the lack of sequences available in the GenBank. The ToLCD can be maintained by very distinct geminiviruses including Tomato yellow leaf curl Sardinia virus, African cassava mosaic virus and Beet curly top virus. From the New World, deltasatellites were first reported in 2012 from the Caribbean islands. To date, deltasatellites are known from India, Vietnam, Philippines, Spain, Australia, Cuba, and Puerto Rico. There are 11 distinct species of deltasatellites, which are phylogenetically not linked to each other (Table 1, Fig. 1).

Genome Organization of Deltasatellites

Deltasatellites do not encode any gene in their genomes. However, they share some common features. Most of the known deltasatellites contain a variable A-rich region, a nonanucleotide origin of replication (TAATATTAC) and a secondary stem loop structure. The deltasatellites from Australia and Spain, share common features of A-rich region and a satellite conserved region. However, the deltasatellites from the New World have very short satellite conserved region (**Fig. 2**).

Trans-Replication of Deltasatellites

Deltasatellites are not self-replicating molecules therefore they require helper viruses for their replication and systemic movement. Although, ToLCD can be maintained by several heterologous Old World geminiviruses, but it does not contribute to any symptom severity or increase of the helper virus titer. The experiments conducted on New World deltasatellites suggested that they cannot be trans-replicated by Old World begomoviruses or curtoviruses, however they can be efficiently maintained by New World bipartite and monopartite begomoviruses. Deltasatellites significantly reduce their helper component's accumulation. Deltasatellites do not encode any gene in their genome, therefore the biological significance and mode of action of deltasatellites is clearly not understood.

See also: Cotton Leaf Curl Disease (*Geminiviridae*). Emerging Geminiviruses (*Geminiviridae*). Geminiviruses (*Geminiviridae*). Nanoviruses (*Nanoviridae*)

Further Reading

- Amin, I., Hussain, K., Akbergenov, R., *et al.*, 2011. Suppressors of RNA silencing encoded by the components of the cotton leaf curl begomovirus-betasatellite complex. *Molecular Plant-Microbe Interactions* 24, 973–983.
- Briddon, R.W., Brown, J.K., Moriones, E., *et al.*, 2008. Recommendations for the classification and nomenclature of the DNA-beta satellites of begomoviruses. *Archives of Virology* 153, 763–781.
- Eini, O., 2017. A betasatellite-encoded protein regulates key components of gene silencing system in plants. *Molecular Biology* 51, 656–663.
- Fiallo-Olivé, E., Tovar, R., Navas-Castillo, J., 2016. Deciphering the biology of deltasatellites from the New World: Maintenance by New World begomoviruses and whitefly transmission. *New Phytologist* 212, 680–692. doi:10.1111/nph.14071.
- Gnanasekaran, P., Chakraborty, S., 2018. Biology of viral satellites and their role in pathogenesis. *Current Opinion in Virology* 33, 96–105.
- Gnanasekaran, P., Kishorekumar, R., Bhattacharyya, R.D., Kumar, V., Chakraborty, S., 2019. Multifaceted role of geminivirus associated betasatellite in pathogenesis. *Molecular Plant Pathology* 20 (7), 1019–1033. doi:10.1111/mpp.12800.
- Hassan, I., Orilio, A.F., Fiallo-Olivé, E., Briddon, R.W., Navas-Castillo, J., 2016. Infectivity, effects on helper viruses and whitefly transmission of the deltasatellites associated with sweepoviruses (genus *Begomovirus*, family *Geminiviridae*). *Scientific Reports* 6, 30204. doi:10.1038/srep30204.
- Hohn, T., Vazquez, F., 2011. RNA silencing pathways of plants: Silencing and its suppression by plant DNA viruses. *Biochimica et Biophysica Acta* 1809, 588–600.
- Khan, A.J., Mansoor, S., Briddon, R.W., 2014. Oman: A case for a sink of begomoviruses of geographically diverse origins. *Trends in Plant Science* 19, 67–70.
- Kumar, J., Kumar, J., Singh, S.P., Tuli, R., 2014. Association of satellites with a mastrevirus in natural infection: Complexity of Wheat dwarf India virus disease. *Journal of Virology* 88, 7093–7104.
- Kumar, R.V., Singh, A.K., Singh, A.K., *et al.*, 2015. Complexity of begomovirus and betasatellite populations associated with chilli leaf curl disease in India. *Journal of General Virology* 96, 3143–3158.
- Li, Y., Qin, L., Zhao, J., *et al.*, 2017b. SIMAPK3 enhances tolerance to Tomato yellow leaf curl virus (TYLCV) by regulating salicylic acid and jasmonic acid signaling in tomato (*Solanum lycopersicum*). *PLoS One* 12, e0172466.
- Li, F., Zhao, N., Li, Z., *et al.*, 2017a. A calmodulin-like protein suppresses RNA silencing and promotes geminivirus infection by degrading SGS3 via the autophagy pathway in *Nicotiana benthamiana*. *PLoS Pathogens* 13, e1006213.
- Lozano, G., Trenado, H.P., Fiallo-Olivé, E., *et al.*, 2016. Characterization of non-coding DNA satellites associated with Sweepoviruses (*Begomovirus*, *Geminiviridae*) – Definition of a distinct class of Begomovirus-associated satellites. *Frontiers in Microbiology* 7, 162.
- Mubin, M., Ijaz, S., Nahid, N., *et al.*, 2020. Journey of begomovirus betasatellite molecules: From satellites to indispensable partners. *Virus Genes* 56, 16.
- Nawaz-ul-Rehman, M.S., Fauquet, C.M., 2009. Evolution of geminiviruses and their satellites. *FEBS Letters* 583, 1825–1832.
- Nawaz-ul-Rehman, M.S., Nahid, N., Mansoor, S., Briddon, R.W., Fauquet, C.M., 2010. Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. *Virology* 405, 300–308.
- Pumplin, N., Voinnet, O., 2013. RNA silencing suppression by plant pathogens: Defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiology* 11, 745–760.
- Singh, A.K., Kushwaha, N., Chakraborty, S., 2016. Synergistic interaction among begomoviruses leads to the suppression of host defense-related gene expression and breakdown of resistance in chilli. *Applied Microbiology and Biotechnology* 100, 4035–4049.
- Verchot, J., 2016. Plant virus infection and the ubiquitin proteasome machinery: Arms race along the endoplasmic reticulum. *Viruses* 8, E314.
- Vinoth Kumar, R., Singh, D., Singh, A.K., Chakraborty, S., 2017. Molecular diversity, recombination and population structure of alphasatellites associated with begomovirus disease complexes. *Infection, Genetics and Evolution* 49, 39–47.
- Yang, X., Wang, Y., Guo, W., *et al.*, 2011a. Characterization of small interfering RNAs derived from the geminivirus/betasatellite complex using deep sequencing. *PLoS One* 6, e16928.
- Yang, X., Xie, Y., Raja, P., *et al.*, 2011b. Suppression of methylation-mediated transcriptional gene silencing by betaC1-SAHH protein interaction during geminivirus-betasatellite infection. *PLoS Pathogens* 7, e1002329.
- Zhang, T., Xu, X., Huang, C., *et al.*, 2015. A novel DNA motif contributes to selective replication of a geminivirus-associated betasatellite by a helper virus-encoded replication-related protein. *Journal of Virology* 90, 2077–2089.
- Zhou, X., 2013. Advances in understanding begomovirus satellites. *Annual Review of Phytopathology* 51, 357–381.
- Zhou, X., Xie, Y., Tao, X., *et al.*, 2003. Characterization of DNA beta associated with begomoviruses in China and evidence for co-evolution with their cognate viral DNA-A. *Journal of General Virology* 84, 237–247.
- Zubair, M., Zaidi, S.S., Shakir, S., Amin, I., Mansoor, S., 2017. An insight into Cotton leaf curl Multan betasatellite, the most important component of cotton leaf curl disease complex. *Viruses* 9, E280. doi:10.3390/v9100280.

Bluner-, Cile-, and Higreiviruses (*Kitaviridae*)

Diego F Quito-Avila, Department of Life Sciences, ESPOL Polytechnic University, Guayaquil, Ecuador

Juliana Freitas-Astúa, Brazilian Agricultural Research Corporation (Embrapa) Cassava and Fruits, Cruz das Almas, Brazil

Michael J Melzer, Department of Plant and Environmental Protection Sciences, University of Hawaii, Honolulu, HI, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Cytopathology The study of disease at the cellular level.

Symptomatology The symptom complex of a disease.

Introduction

Kitaviridae is a family of positive-sense RNA viruses with linear, segmented genomes and bacilliform to spherical virions. All current members of the family infect plants, often causing chlorotic and/or necrotic lesions in their natural hosts. These infections are typically non-systemic, making them highly unusual among plant viruses. Despite being non-systemic, diseases caused by these viruses can be highly destructive and of tremendous economic importance, with citrus leprosis being the most prominent. Citrus leprosis virus C, a causal agent of citrus leprosis, represents the most well-studied member and functional model for the family. The name *Kitaviridae* is derived from the renowned Brazilian virologist and electron microscopist Dr. Elliot W. Kitajima, in recognition of his extensive contributions to our knowledge on the fundamental aspects of kitavirus structure, cytopathology, and biology. The family was established by the International Committee on Taxonomy of Viruses (ICTV) in 2018.

Taxonomy, Phylogeny, and Evolution

The family *Kitaviridae* is currently composed of three genera, established by the ICTV, *Cilevirus*, *Higrevirus*, and *Blunervirus*. The genus *Cilevirus* derives its name from the type species *Citrus leprosis virus C*, the prevalent causal agent of citrus leprosis, one of the most aggressive and important citrus diseases in South and Central America. The genus *Higrevirus* is named after the type species *Hibiscus green spot virus 2* (HGSV 2), a virus associated with leprosis-like lesions in some citrus species and chlorotic lesions on the leaves of hibiscus species. To date, the virus and its associated diseases have only been reported in the Hawaiian Islands. The genus *Blunervirus* was named after its type species *Blueberry necrotic ring blotch virus*, which was discovered in 2010 from symptomatic highbush blueberry (*Vaccinium corymbosum*) plants in Georgia, United States. At the species level, demarcation criteria have only been established for the genus *Cilevirus* and include the extent of serological relationship; <85% amino acid identity for the proteome, natural and experimental host range reactions, and vector species and transmission. Recently, a threshold of 75% amino acid identity for the polyprotein encoded by RNA1 has been proposed for species demarcation in the genus *Blunervirus*.

Phylogenetic reconstructions using RNA-dependent RNA polymerase (RdRp) sequences, the standard for inferring relationships between RNA viruses, illustrate the family *Kitaviridae* as a monophyletic clade and sister taxon to the plant virus family *Virgaviridae*. The phylogenetic relationships that kitaviruses share with other viral taxa, however, are not this straightforward. For example, blueberry necrotic ring blotch virus (BNRBV) encodes two helicase (HEL) domains. Phylogenetic reconstructions using these two domains indicate that one (HEL-1) is most closely related to those of virgaviruses, and the other (HEL-2) most closely related to members of the plant virus family *Bromoviridae*. Furthermore, the methyltransferase (MTR) domain of BNRBV contains hallmark motifs that are found in bromoviruses. The predicted movement proteins (MP) of kitaviruses are also of different phylogenetic lineages, with blunerviruses and cileviruses possessing 3A MPs of the 30K superfamily, and the lone higrevirus possessing 'triple gene block' MPs. Taken together, the evolutionary history of kitaviruses is complex and highly indicative of recombination events between distinct viral taxa.

High-throughput metagenomic sequencing of arthropod and other invertebrate specimens has provided evidence of several unclassified viruses that appear to be closely related (and even interspersed) within the kitavirus clade when using the RdRp as the phylogenetic signal. Furthermore, a predicted structural protein designated p22 (*Blunervirus*), p23 (*Higrevirus*) or p24 (*Cilevirus*) is present in all kitaviruses (but no other plant virus taxa), and also in some of these unclassified invertebrate viruses (Fig. 1). This, combined with the inability of recognized kitaviruses to infect their natural plant host systemically, has led to speculation that the evolutionary progenitor(s) of the kitavirus clade might be an arthropod virus.

Members of the Family

The family *Kitaviridae* is sparsely populated, with only four recognized species among the three genera. Some tentative members have been partially described but are not currently recognized by the ICTV (Table 1).

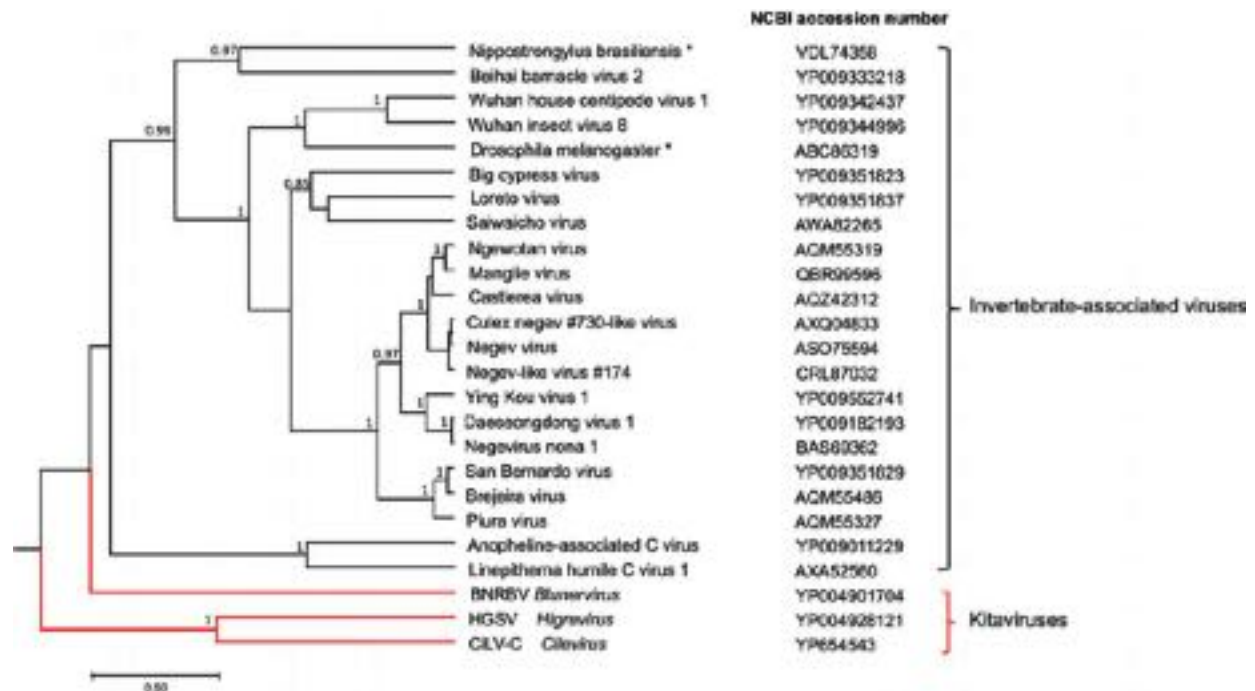


Fig. 1 Phylogenetic relationships across kitaviruses (BNRBV, HGSV-2 and CILV-C) and unclassified invertebrate viruses, based on hypothetical protein p22–24. A Bayesian tree is shown using LG + G + I protein evolution model. NCBI (National Center for Biotechnology Information) accession numbers are provided on the right. Taxa where p22–24 have been annotated as host proteins are indicated by the asterisk (*).

Table 1 Species and tentative species in the family *Kitaviridae*

Virus	Genomic segments	Genome length (kb)	Natural host
Genus <i>Blunervirus</i>			
Definitive species			
<i>Blueberry necrotic ring blotch virus</i> ^a	4	14.15	Blueberry
Tentative species			
Tea plant necrotic ring blotch virus	4	15.08	Tea
Genus <i>Cilevirus</i>			
Definitive species			
<i>Citrus leprosis virus C</i> ^a	2	13.73	Citrus, tropical spiderwort (<i>Commelina benghalensis</i>)
<i>Citrus leprosis virus C2</i>	2	13.71	Citrus, hibiscus
Tentative species			
Ligustrum ringspot virus	N/a	N/a	Privet
Passion fruit green spot virus	N/a	N/a	Passion fruit
Genus <i>Higraviruses</i>			
Definitive species			
<i>Hibiscus green spot virus 2</i> ^a	3	14.70	Citrus, hibiscus

^aType species.

Virion Structure

Virion morphology for members of the *Cilevirus* and *Higraviruses* genera has been determined by transmission electron microscopy (TEM). Infected cells contain congregations of bacilliform virions, which are typically detected in cytoplasmic vesicles associated with the endoplasmic reticulum. Virions range between 30 and 50 nm in width and 50–120 nm in length depending on the species (Fig. 2). Electron microscopy studies have not been done to investigate virion morphology of BNRBV (genus *Blunervirus*). However, TEM examination of ultrathin sections of symptomatic tissue infected with tea plant necrotic ring blotch virus, a proposed member of the genus, revealed cytoplasmic aggregates of spherical virus-like particles approximately 85 nm in diameter.

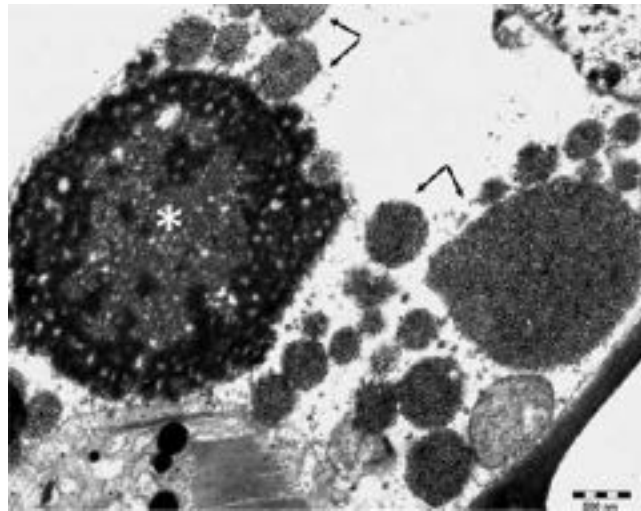


Fig. 2 Transmission electron micrographs of ultrathin sections of soybean (an experimental host) leaves infected with Citrus leprosis virus C (CiLV-C). Asterisk (*) indicates electron-dense viroplasm, and arrows show host-derived vesicle structures containing virus-like particles. Photo courtesy of Dr. E.W. Kitajima.

Genome Organization and Gene Product Function

Kitavirus genomes are multipartite and composed of linear, positive-sense RNA molecules (Fig. 3). The genomes of current cileviruses are bipartite and comprise a total length of 13.1 kb. RNA1 contains a large 5'-terminal open reading frame that encodes a replication-associated polyprotein containing MTR, HEL, and RdRp domains followed by an ORF encoding a 29 kDa protein predicted to be the capsid protein (CP). Homologs of this putative CP have not been found in other kitavirus genomes. RNA2 contains four ORFs, the first of which encodes p15, which appears to have a role in cellular remodeling for vesicle formation. The second ORF encodes p61, a predicted glycoprotein that also appears to remodel the host endoplasmic reticulum (ER). The third ORF encodes a 3A type MP, and the 3'-terminal ORF encodes p24, a predicted virion membrane protein. p24 is one of the few proteins encoded by all kitaviruses and is also present in some unclassified invertebrate viruses. The 5'-termini of the RNAs are capped and the 3'-termini are polyadenylated.

The genome of the current higrevirus species is tripartite and comprises a total length of 14.7 kb. RNA1 is similar to that of cileviruses but lacks the putative 29 kDa CP. RNA2 possesses a 50 kDa protein of unknown function followed by a triple gene block-like suite of proteins of 39, 9, and 6 kDa which putatively serves as the MPs of the virus. However, recent findings have suggested that proteins p39 and p9, acting in a 'binary movement block', are sufficient for virus intercellular movement. RNA3 possesses three ORFs of 33, 29, and 23 kDa. The first two are of unknown function, and p23 is a homolog of the cilevirus p24 and blunervirus p22. Higrevirus RNAs are polyadenylated on their 3'-termini.

Current blunervirus genomes are quadripartite and comprise a total length of 14.2 kb. RNA1 contains a single open reading frame (ORF), whose predicted protein contains MTR and HEL-1 domains. The single ORF predicted in RNA2 encodes the putative RdRp, with the HEL-2 domain at the N terminus. RNA3 comprises five short ORFs (ORF3a, 3b, 3c, 3d and 3e) putatively encoding proteins of unknown functions. One of these unknown proteins, p22, encoded by ORF3c, has homology to proteins of similar size in cileviruses (p24), higreviruses (p23) and unclassified invertebrate viruses. RNA4 has a single ORF coding for the putative MP, which possesses motifs conserved with the 3A family of the 30K superfamily of viral MPs (30Ksf). This protein is homologous to its counterpart in cileviruses. Non-coding regions (NCR) at 5'-ends vary in size, with an unusually long (> 350 nt) NCR in RNA4. NCRs at 3'-ends of BNRBV contain conserved nucleotide stretches. Although the presence of poly (A) tails were not documented for the original BNRBV genome, adenosine stretches of varying size were detected on 3'-termini of BNRBV-RL (a red-lesion strain discovered in Florida) and the putative new blunervirus Tea plant necrotic ring blotch virus (TPNRBV); suggesting that polyadenylation is a common feature in kitaviruses.

Replication and Propagation

The Cilevirus citrus leprosis virus C (CiLV-C) represents the most well-studied kitavirus and the main virus for which replication and propagation studies have been conducted. Some information on that regard has been obtained for Citrus leprosis virus C2 as well (CiLV-C2). CiLV-C RNA replication-associated proteins (MTR, HEL, and RdRp) are expressed from genomic RNA1 as a polyprotein. Other proteins involved with replication and propagation are expressed via 3'-terminal sub-genomic RNAs. Replication appears to occur in cytoplasmic vesicles formed through the modification of cortical endoplasmic reticulum (ER) by

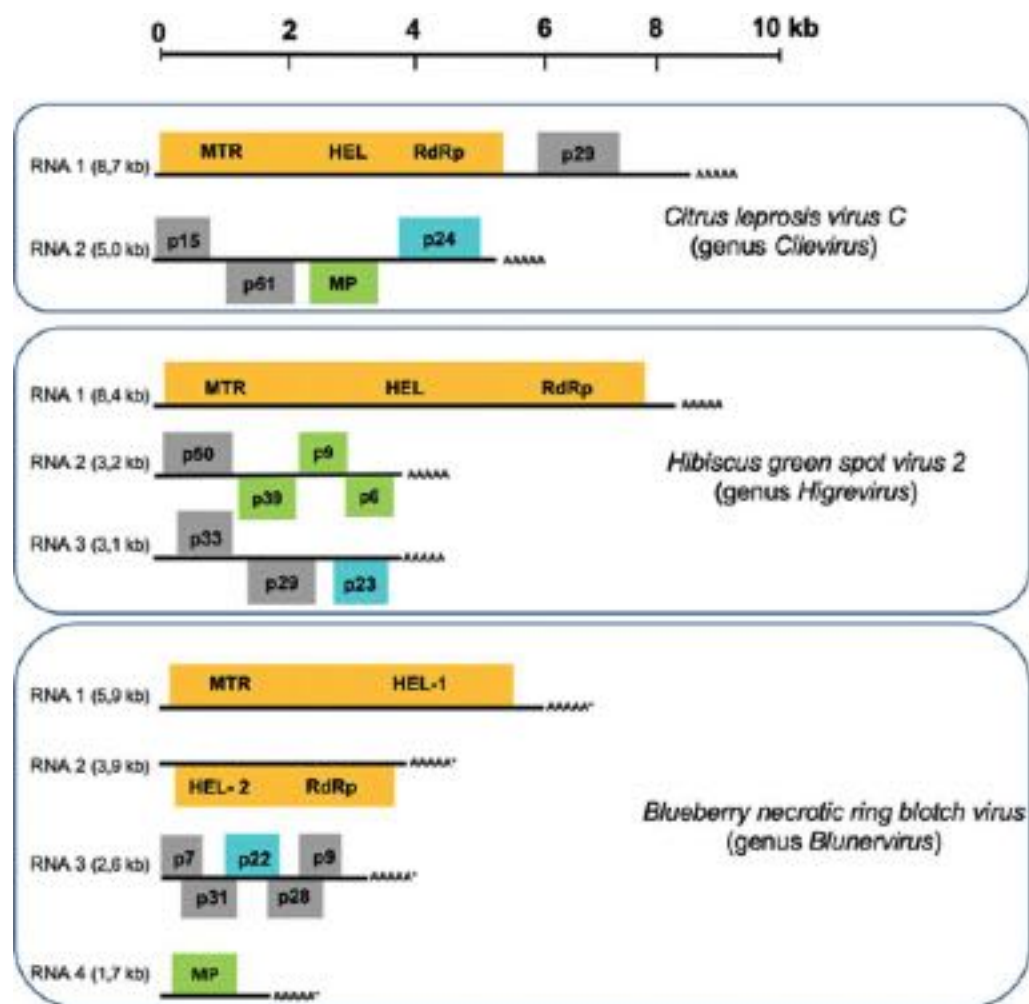


Fig. 3 Genomic organization of the type species for each genus in the family *Kitaviridae*. Boxes represent open reading frames (ORFs). ORFs are labeled with either the predicted molecular weight (kDa) of the protein product, or its function: MTR, methyltransferase; HEL, helicase; RdRp, RNA-dependent RNA polymerase; and MP, movement protein. Asterisk (*) indicates that the polyadenylated 3'-terminus has been reported for one isolate of the blueberry necrotic ring blotch virus, but not another. The scale at top represents approximate RNA length in kilobases (kb).

multiple virus-encoded proteins. Transfer of virus replication complexes and/or the viral ribonucleoproteins from the cytoplasmic vesicles to plasmodesmata may be facilitated by actin or ER trafficking. Viral MP then facilitate cell-to-cell movement through the plasmodesmata. Only non-systemic movement has been reported for any member of the family in plant hosts, although systemic movement has been shown for TPNRBV.

Transmission and Host Range

Cileviruses are vectored by polyphagous mites in the genus *Brevipalpus* in a persistent manner. There is molecular evidence for replication in these mites, but this has not been confirmed by electron microscopy, despite extensive efforts. Preliminary transmission studies indicate higreviruses are also transmitted by brevipalpus mites. Cileviruses and higreviruses have been found to naturally infect *Citrus* and *Hibiscus* spp., and experimental host range experiments using viruliferous mites have demonstrated that CiLV-C is able to infect dozens of plant species, including several of economic importance such as common bean and soybean.

Preliminary transmission experiments, both under field and greenhouse conditions, have demonstrated that the blunervirus BNRBV is transmitted by eriophyoid mites in the genus *Calacarus*. The specific mechanism by which the vector transmits the virus is not known. BNRBV has not been reported in any host other than highbush blueberry. TPNRBV, a putative blunervirus, was found to infect tea plants (*Camellia sinensis*), although no additional transmission or host range data is currently available.

Diseases, Epidemiology, and Control

Citrus leprosis disease has multiple causal agents, including CiLV-C and CiLV-C2, which are both members of the genus *Cilevirus*. Symptoms of infection include localized chlorotic and necrotic lesions on both leaves and fruit, which can lead to premature abscission. Citrus leprosis is extant only in South and Central American countries. Although historical reports of the disease in Florida, USA exist, it appears the pathogen at this location was an unrelated virus which causes nearly identical symptoms in citrus. In Brazil, citrus leprosis caused primarily by CiLV-C is often regarded as the most important virus disease of citrus. Management efforts to control the disease, which are primarily the application of acaricides to control vector populations, often reach \$40–80 M USD per year.

In Hawaii, the higrevirus Hibiscus green spot virus 2 (HGSV-2) is responsible for an isolated disease on some citrus varieties that resembles citrus leprosis. Unlike citrus leprosis, foliar lesions on citrus caused by HGSV-2 rarely become necrotic, and lesions on fruit are only visible on green, unripe fruit. HGSV-2 also causes lesions on hibiscus leaves which are most prominent during senescence. Severe infection may cause premature abscission of leaves which may be of concern to threatened and endangered hibiscus species native to Hawaii.

The blunervirus BNRBV was discovered as the causal agent of an unusual disease observed in southern highbush blueberry (*Vaccinium corymbosum* interspecific hybrids) plantings in Georgia, USA. Infected plants display irregularly shaped rings on the adaxial and abaxial leaf surfaces, which coalesce to form brown-to-black necrotic blotches, followed by premature defoliation. Since its first report in 2006, the disorder has become commonplace in other southern states including Florida, Mississippi, North Carolina and South Carolina. In Florida, a variant of the virus was characterized and found to induce red concentric rings, which develop across leaves without turning into necrotic spots or falling out prematurely. This strain is referred to as BNRBV-red lesion (RL) and shares an overall 90% nucleotide identity with BNRBV. The impact of the disease caused by BNRBV has not been reported.

None of the current kitaviruses can move systemically in plant hosts and are therefore not graft-transmissible. However, stem lesions or viruliferous mites present on propagative materials can serve as a reservoir for long-distant movement by humans. Unlike citrus and hibiscus, blueberry defoliates in winter, indicating a direct role of overwintering vectors and/or alternate (virus reservoir) hosts in the spread of BNRBV during subsequent seasons. This aspect of the virus life cycle has a direct implication in terms of disease management by controlling the vector.

Diagnosis

Determining the localization of virions and viroplasm by TEM was initially used to distinguish the cytoplasmic citrus leprosis viruses (CiLV-C and CiLV-C2) from the unrelated nuclear virus which caused similar symptoms. Today, PCR-based molecular approaches are used for kitavirus detection. Serological detection approaches are also available for CiLV-C and CiLV-C2, which include enzyme-linked immunosorbent assays (ELISA) and cellular analysis and notification of antigen risks and yields (CANARY) platforms.

Concluding Remarks

The kitaviruses are a small, but important group of emerging viruses that warrant study not only for control of the diseases they cause, but also to better understand the fundamentals of plant virology. The non-systemic nature of their infection opens intriguing opportunities for experimentation on virus movement and arthropod vector relationships. The putatively close evolutionary relationship with a growing number of unclassified invertebrate viruses suggests that in the future this taxon may not be restricted to plant viruses.

Further Reading

- Bastianel, M., Novelli, V.M., Kitajima, E.W., *et al.*, 2010. Citrus leprosis: Centennial of an unusual mite-virus pathosystem. *Plant Disease* 94, 284–292.
- Cantu-Iris, M., Harmon, P.F., Londono, A., Polston, J.E., 2013. A variant of Blueberry necrotic ring blotch virus associated with red lesions in blueberry. *Archives of Virology* 158, 2197–2200.
- Hao, X., Zhang, W., Zhao, F., *et al.*, 2018. Discovery of plant viruses from tea plant (*Camellia sinensis* (L.) O. Kuntze) by metagenomic sequencing. *Frontiers in Microbiology* 9, 2175.
- Lazareva, E.A., Lezhov, A.A., Komarova, T.V., *et al.*, 2017. A novel block of plant virus movement genes. *Molecular Plant Pathology* 18, 611–624. doi:10.1111/mpp.12418.
- Leastro, M.O., Kitajima, E.W., Silva, M.S., Resende, R.O., Freitas-Astúa, J., 2018. Dissecting the subcellular localization, intracellular trafficking, interactions, membrane association, and topology of Citrus leprosis virus C proteins. *Frontiers in Plant Science* 11 (9), 1299.
- Localí-Fabris, E.C., Freitas-Astúa, J., Souza, A.A., *et al.*, 2006. Complete nucleotide sequence, genomic organization and phylogenetic analysis of *Citrus leprosis virus* cytoplasmic type. *Journal of General Virology* 87, 2721–2729.
- Quito-Avila, D.F., Brannen, P.M., Cline, W.O., Harmon, P.F., Martin, R.R., 2013. Genetic characterization of Blueberry necrotic ring blotch virus, a novel RNA virus with unique genetic features. *Journal of General Virology* 94, 1426–1434.
- Robinson, T.S., Scherm, H., Brannen, P.M., Allen, R., Deom, C.M., 2016. Blueberry necrotic ring blotch virus in southern highbush blueberry: insights into in planta and in-field movement. *Plant Disease* 100, 1575–1579.
- Roy, A., Hartung, J.S., Schneider, W.L., *et al.*, 2015. Role bending: Complex relationships between viruses, hosts, and vectors related to citrus leprosis, and emerging disease. *Phytopathology* 105, 1013–1025.

Brome Mosaic Virus (*Bromoviridae*)

Guijuan He, Virginia Tech, Blacksburg, VA, United States

Zhenlu Zhang, Shandong Agricultural University, Tai'an, China

Preethi Sathanantham, Virginia Tech, Blacksburg, VA, United States

Arturo Diaz, La Sierra University, Riverside, CA, United States

Xiaofeng Wang, Virginia Tech, Blacksburg, VA, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of X. Wang, P. Ahlquist, Brome Mosaic Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00560-4](https://doi.org/10.1016/B978-012374410-4.00560-4).

Nomenclature

aa Amino acid(s)

CAT Chloramphenicol acetyltransferase

CP Coat protein or capsid protein

CPR5 Constitutive expresser of Pathogenesis-Related genes 5

ER Endoplasmic reticulum

ESCRT Components of the endosomal sorting complex required for transport

GFP Green fluorescent protein

kb Kilobases; the size of a ssDNA or ssRNA molecule

kDa Kilodaltons; the size of a protein

mRNA Messenger RNA

NCR Non-coding region

nm Nanometer(s)

nt Nucleotide(s)

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

RNAi RNA interference

tRNA Transfer RNA

UTR Untranslated region

VIGS Virus-induced gene silencing

Glossary

Host factor Host proteins, lipids, and metabolites subverted by the virus at multiple stages of the life cycle required for successful viral infection.

Spherule Single-membrane vesicular invaginations of host membranes induced by positive-strand RNA viruses in which viral replication proteins, RNA templates, and specific host factors are enriched and viral genome replication takes place. Other names used to refer to RNA replication competent spherules are viral replication compartments,

viral replication complexes, or viral replication mini-organelles.

Sub-genomic RNA A messenger RNA produced during viral replication to express downstream viral genes from a viral genomic RNA expressing multiple genes.

Tripartite RNA genome The viral genome is composed of three pieces of RNAs, which are all required for a successful systemic infection, may differ in length and in viral proteins encoded but may share similar sequences and/or structures of untranslated regions as well as 5' and/or 3' ends.

Introduction

Brome mosaic virus (BMV) is a positive-strand RNA virus that infects cereal plants, causing mosaic symptoms and stunting. BMV is the type member of the genus *Bromovirus* in the family *Bromoviridae*. BMV is also a representative member of the alphavirus-like super-family that include viruses infecting human, animal, and plants that share similar genome structures, gene expression strategies, and protein functions. BMV has been used as a model to study gene expression, RNA replication, virus-host interactions, recombination, and encapsidation by positive-strand RNA viruses. Major contributions have revealed insights and principles extending to many other viruses and to general cell biology. Among numerous advances, research in BMV defined the first ribosome binding sites in eukaryotic mRNAs, produced the first eukaryotic viral RNA-dependent RNA polymerase (RdRp) extract that was template specific, the first infectious transcripts from cloned cDNA, the first engineered RNA virus expression vector expressing foreign genes, the first definition of sub-genomic mRNA synthesis pathways and determinants, and the first demonstration of higher eukaryotic viral replication in yeast. BMV has contributed many insights into virion assembly, virus-host interactions, RNA recombination, and RNA replication, including parallels in the replication of positive-strand RNA, reverse-transcribing and dsRNA viruses.

Virion Structure and RNA Encapsidation

BMV forms non-enveloped virions ~28 nm in diameter. The outer capsid is composed of 180 copies of coat protein (CP) arranged with $T = 3$ quasi-icosahedral symmetry. Cryo-electron microscope reconstructions and subsequently x-ray crystallography showed that the BMV capsid structure is extremely similar to that of the related bromovirus Cowpea chlorotic mottle virus (CCMV). Intriguingly, some features of these capsids are dissimilar from other known isometric RNA virus capsids, but similar to capsids of the DNA-

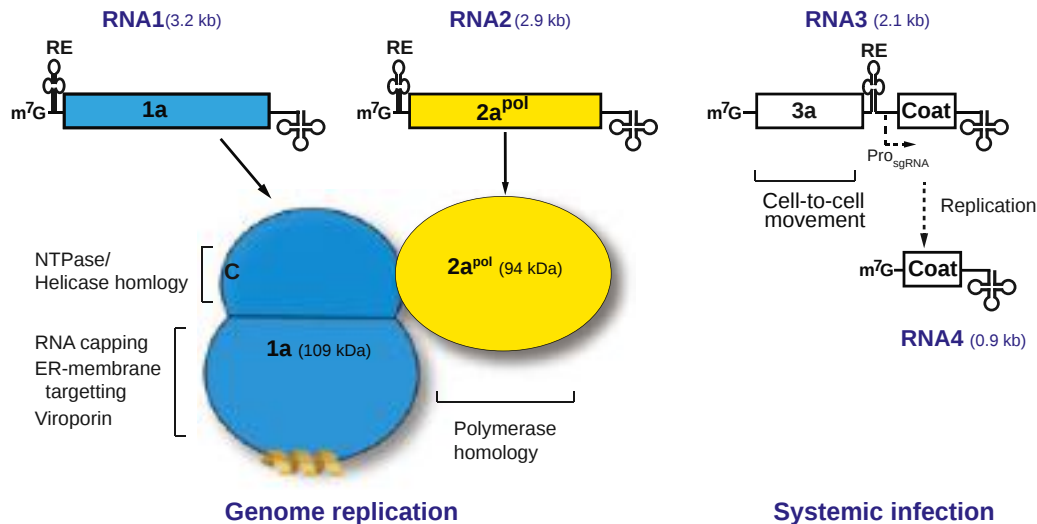


Fig. 1 Organization of the Brome mosaic virus (BMV) RNA genome. BMV has three genomic RNAs, RNA1, RNA2, and RNA3, and a sub-genomic mRNA, RNA4, encoded by the 3' portion of RNA3 as shown. Each of these four BMV positive-strand RNAs bears a 5' m⁷G cap and 3' tRNA-like structure (cloverleaf). The recruitment element (RE) in each viral genomic RNA and the promoter of sub-genomic RNA4 (Pro_{sgRNA}) in RNA3 are shown. Open boxes represent the genes for the viral proteins 1a and 2a^{pol}, which are required for RNA replication, and 3a and coat protein, which are required for systemic movement in host plants. Some specific functions and features of each viral protein are listed, including the amphipathic α -helices at the N-terminus of 1a. The N-terminus (bottom half) and C-terminus (top half) of 1a are labeled as N and C, respectively. BMV 1a interacts with 2a^{pol} via the C-terminus of 1a and the N-terminus of 2a^{pol}.

containing papovaviruses. These features include orientation of the core β -barrel nearly perpendicular to the capsid surface to form distinctively prominent hexameric and pentameric capsomeres, and linking of adjacent capsomere clusters by exchange of invading C-terminal arms. These shared features suggest that BMV CP and the CPs of polyoma- and papillomaviruses likely share a common ancestor. BMV virion RNA is arranged as an interior subshell inside the capsid, leaving a hollow virion center. The N-terminal 26 aa of the CP, which are highly basic and required for RNA packaging, interact with the RNA to neutralize its charge.

In addition to the predominant 180-subunit capsid, BMV CP can also assemble *in vivo* a 120-subunit capsid, composed of 60 CP dimers, first discovered in yeast and subsequently confirmed in infected plants. This assembly polymorphism is controlled *in vivo* by the RNA packaged, with BMV RNA2 packaged in 180-subunit capsids, while a small chimeric mRNA containing the CP open reading frame (ORF) as its only BMV sequence is packaged in 120-subunit capsids. Structural features shared by 120- and 180-subunit capsids imply that a common pentamer of CP dimers is an important intermediate in BMV virion assembly.

In vitro and *in vivo* encapsidation identified portions of the BMV 3a coding region whose deletion blocked RNA3 encapsidation and interfered with normal co-encapsidation of RNA3 and RNA4. Studies also implicated the 3' tRNA-like structure of BMV RNAs as a facilitator of encapsidation *in cis* or *trans*, possibly by nucleating CP interactions into productive assemblies such as pentamers of dimers, and showed that BMV RNA replication promotes selective encapsidation of viral RNAs, possibly by inducing coupled synthesis of viral RNA and CP in close proximity.

In 1971, buoyant density gradients were used to separate BMV virions into three classes, all three have identical capsids but package different RNAs (Fig. 1). Heavy virions contain a single copy of RNA1 (3.2 kb), medium density virions contain one copy each of RNA3 (2.1 kb) and RNA4 (0.9 kb), and light virions contain a single copy of RNA2 (2.9 kb). The three classes of viral particles are indistinguishable by electron microscopy, however, distinct capsid-RNA contacts influence the stability of the virions, viral RNA release, and BMV infection. Mapping of the contacts made between CP and the viral RNAs showed that RNA1 contacts CP differently than RNA2 or RNA3/4. Virions containing RNA1 were more sensitive to protease digestion and treatment of virions with RNase A revealed that RNA1 was preferentially degraded, while RNA2 or RNA3/4 were more resistant to RNase A digestion. The more facile interaction between RNA1 and the capsid proteins could ensure earlier and higher expression of BMV 1a, which is responsible for the formation of the viral replication compartments (see below in the Section of "Properties and Functions of Viral Gene Products"). The interaction of the viral RNAs with capsid is also influenced by the phosphorylation state of the N-terminal arm of the capsid protein. Interestingly, virions produced from different plant hosts had different post-translational modifications that impacted the thermostability of virions as well as the capsid-RNA interactions.

Genome Organization

Genome Structure, Gene Expression and Sequences

Interest in characterizing the protein(s) encoded by each of the four BMV virion RNAs motivated early *in vitro* translation studies with purified BMV virion RNAs. RNAs 1, 2, and 4 were characterized as monocistronic while RNA3 is dicistronic. Although early

infectivity studies showed that RNA3 encodes CP, tryptic analysis showed that the principle translation product of RNA3 was unrelated to CP, while RNA4 served as an excellent template for CP. Moreover, when mixtures of all BMV RNAs were added to wheat germ extracts, increasing amounts of RNA4 inhibited translation of genomic RNAs 1–3. These results showed that RNA4 is a sub-genomic CP mRNA derived from RNA3, and implied an elegant system of gene regulation by translational competition: early in infection when viral RNA concentrations are low, all viral proteins including nonstructural RNA replication factors are translated, while after the virion RNAs are sufficiently amplified, CP is preferentially translated to encapsidate these RNAs.

By the early 1970s, RNA bacteriophage studies had provided valuable information on prokaryotic translation initiation sites, including the finding that the first AUG initiation codon usually was 100 or more nucleotides from the RNA 5' end. In 1975, the first eukaryotic translation initiation site was characterized by isolating two fragments of BMV RNA4 that were efficiently bound by wheat germ ribosomes. RNA sequencing (a challenging, chromatography-based process at the time) revealed that the two fragments were 5' terminal, overlapping, and encoded the first 4 and 14 aa of CP, respectively. The most distinguishing feature was that the initiating AUG codon for CP began only 10 nucleotides from the m⁷G^{5'}ppp^{5'}Gp-capped RNA4 5' end, presaging the now well-known mechanistic linkage between most eukaryotic translation initiation and 5' mRNA caps.

The ~8.2 kb BMV genome sequence was completed in 1984. In good agreement with *in vitro* translation results, RNA1 and RNA2 encode single proteins 1a (109 kDa) and 2a (94 kDa), respectively, the 5' half of RNA3 encodes protein 3a (35 kDa), and the 3' half of RNA3 encodes for RNA4, the sub-genomic mRNA for CP (20 kDa) (Fig. 1). Later work discussed below showed that 3a is required for infection movement in plants. Comparisons with other emerging viral RNA and protein sequences quickly revealed that the BMV 1a and 2a proteins, already implicated in RNA replication by protoplast experiments, shared extensive amino acid sequence similarities with proteins encoded by a diverse set of plant and animal positive-strand RNA viruses. These similarities initially were recognized between BMV, Alfalfa mosaic virus, Tobacco mosaic virus and Sindbis virus, and subsequently were found to extend to many other viruses now grouped together as the alphavirus-like super-family. Similarities shared by these viruses include a polymerase domain in BMV 2a (hereafter 2a^{pol}) and RNA capping and RNA helicase-like domains in 1a (Fig. 1).

Noncoding Regions, *Cis* Signals, and Sub-Genomic Promoter

The first region of BMV RNAs to attract attention as a potential regulatory element was the 3' end. Synergistic work by multiple groups showed that the 3' non-coding regions (NCR) of BMV RNAs are highly conserved, contain multifunctional domains that direct negative-strand RNA synthesis, contribute to RNA encapsidation, translation and stability, and possess multiple tRNA-like features and functions.

Limited early sequence data showed that BMV RNAs 1–4 share a tRNA-like CCA_{OH} 3' end. All four BMV RNAs are selectively amino-acylated *in vitro* with tyrosine. Further studies showed that tyrosylated BMV RNAs interacted with translation elongation factor 1a and that BMV RNAs were tyrosylated *in vivo* during infection of barley protoplasts. BMV RNA 3' ends were also found to interact with (ATP, CTP):tRNA nucleotidyl transferase, which can add 3' CCA_{OH} ends to mature BMV RNAs or maintain incomplete BMV RNAs, thus acting as a primitive telomerase. Beginning in the 1970s, further sequencing, enzymatic structure probing, and three-dimensional computer modeling showed that the 3' ~200 nt of all BMV RNAs were strongly conserved and folded into an extended, tRNA-like structure with at least two alternate forms that differed in pairing nucleotides near the 3' with local or distal partners. Similarly conserved, highly structured 3' regions with related alternate forms were also found in other members of the family *Bromoviridae*.

In vitro and *in vivo* studies using RNA fragments and mutations showed that sequences within the 3'-terminal tRNA-like structure direct negative-strand RNA initiation for RNA replication. Recent results have also implicated the tRNA-like sequence in translation and encapsidation. Involvement in all of these functions led to early and continuing suggestions that the 3' region mediates co-regulation of the varied uses of BMV positive-strand RNAs to minimize conflicts between multiple essential processes.

A second class of elements combining tRNA-like features, replication signals and possible interaction with translation are the BMV template recruitment element (Figs. 1 and 2). Deletion analysis revealed that, in addition to 3' and 5' sequences required for negative- and positive-strand RNA initiation, BMV RNA3 replication *in vivo* requires a segment of ~150 nt in the 5' half of the intergenic non-coding region between the 3a and CP genes (Fig. 1). Subsequent studies showed that this region is required for a step prior to negative-strand RNA synthesis, and is necessary and sufficient for 1a to recognize and recruit an RNA to a membrane-associated, translationally inaccessible, nuclease-resistant state that appears to be the interior of the replication compartments, also known as spherules (see the Section "BMV 1a and Viral Replication Compartments"). Structure probing studies show that this intergenic RNA3 sequence folds into a long, bulged stem-loop, which presents at its apex the invariant sequence and structure of tRNA T Ψ C stem-loops. In plant and yeast cells, the appropriate BMV residues in this consensus are modified to T and Ψ , showing that, like the 3' end, this sequence interacts *in vivo* with tRNA-specific enzymes. Moreover, any mutations to this T Ψ C stem-loop mimicry abolish 1a-mediated template recruitment. Similar stem-loops with apical T Ψ C stem-loop regions are found at the extreme 5' ends of BMV RNA1 and RNA2, where they similarly direct 1a-mediated template recruitment (Fig. 1).

A common approach to express downstream gene(s) in a viral genomic RNA encoding more than one gene is to produce sgRNA(s) efficiently. *In vivo* and *in vitro* analyses of the BMV sub-genomic mRNA promoter have complemented well to reveal a core promoter within the 20 nt immediately upstream of the RNA4 start site, which directs low level but accurate initiation of sub-genomic mRNA. *In vivo*, the activity of this core promoter is greatly enhanced by upstream sequences that include an oligo(A) tract of variable, ~16–22 nt in length in the viral population as well as upstream, partially conserved repeats of core promoter

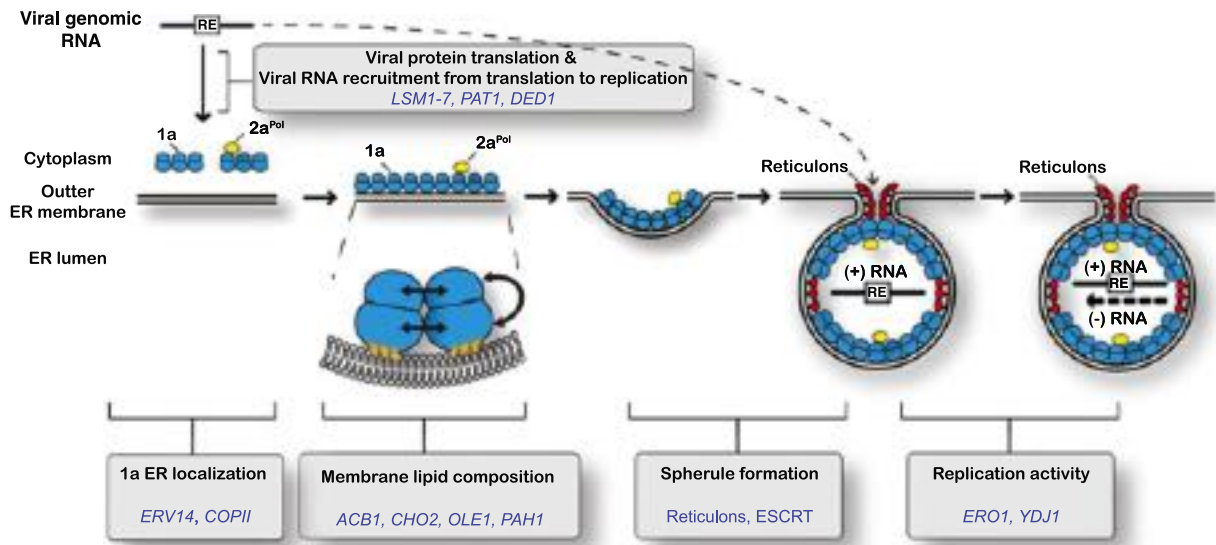


Fig. 2 Host factors required for the formation of functional BMV spherules. Host genes are involved in every step of BMV genome replication, including viral protein synthesis, recruitment of the viral RNA from ribosomes to the interior of spherules, targeting 1a to the nuclear ER membrane, membrane lipid composition and 1a-membrane interactions, spherule formation, and activation of replication activity. Host genes and pathways whose involvement in viral replication has been well characterized are listed.

sequences. The important role of the oligo(A) tract suggests that while negative-strand RNA can serve as a sub-genomic mRNA template *in vitro*, the natural *in vivo* template might be a dsRNA within which the oligo(A) provides a melting site to facilitate internal initiation.

As one important aspect of *cis* signals in BMV replication, minimal core promoters were defined and dissected for the synthesis of negative-strand, positive-strand, and sub-genomic RNA, using a variety of approaches. Among other results, mutational studies imply that the BMV RdRp-promoter interaction has the characteristics of an induced fit, wherein the RdRp has some tolerance to adjust its binding to a range of promoter variants as long as some key sequence features remain. This model potentially reconciles the specificity of BMV RNA synthesis with the ability of the RdRp to synthesize different forms of viral RNA from separate promoters with distinct primary sequences and secondary structures. DNA or DNA/RNA hybrid templates containing the key BMV promoter sequences can be recognized *in vitro* by BMV RdRp extracts and copied into RNA. Although the efficiency of copying DNA templates is ~15-fold lower than for BMV RNA templates, these results have significant potential implications for virus evolution.

Replication and Propagation

In Vitro and *In Vivo* Replication Studies

Positive-strand RNA viruses, like BMV, replicate their genomes in a completely RNA-dependent manner, producing a negative-strand RNA replication intermediate for each genomic RNA. Studies on BMV RNA replication were greatly advanced by the development and subsequent use of such tools as *in vitro* RdRp extracts and cultured plant protoplasts. In 1979, a virus-specific *in vitro* RdRp extract was developed from BMV-infected barley leaves that synthesized full-length negative-strand RNAs using added BMV virion RNAs as templates. This was the first eukaryotic *in vitro* RdRp preparation exhibiting a high level of template specificity, with other viral RNAs having less than 20% of the template activity of BMV RNAs. As noted in part below, this and similar BMV *in vitro* systems have been utilized to make many advances regarding promoters for positive- and negative-strand RNA synthesis, initiation mechanisms, and other issues.

In parallel to *in vitro* systems, cultured plant protoplasts provided a valuable substrate for *in vivo* replication studies due to their ability to be infected or transfected with nearly 100% efficiency with virions, virion RNAs, or *in vitro* transcripts from cloned viral cDNAs. For BMV, barley protoplast systems have allowed examining all aspects of BMV RNA replication, sub-genomic RNA synthesis, progeny RNA encapsidation, and the like. The highly synchronized infections obtained also allow detailed kinetic studies.

In the early 1990s, it was shown that BMV also would direct RNA replication, sub-genomic RNA synthesis, selective viral RNA encapsidation, RNA recombination and the like in the well-studied yeast, *Saccharomyces cerevisiae*. This host provides some of the same advantages as plant protoplasts together with rapid growth, a particularly small genome, well-characterized gene functions and cell biology, and powerful classical and molecular genetic tools including genome-wide arrays of isogenic yeast strains with each gene systematically modified by deletion, GFP-tagging, etc. Yeast cells expressing BMV RNA replication proteins 1a and 2a^{Pol} support the

replication of BMV genomic RNAs introduced by transfection or DNA-dependent transcription from plasmids or chromosomally integrated expression cassettes, duplicating all major features of replication in plant cells. To facilitate yeast genetic studies, BMV RNA replicons can express selectable or screenable markers and are transmitted to yeast daughter cells during cell division with 85%–90% efficiency, rivaling the transmission of yeast DNA plasmids. Some key findings identified and characterized in the BMV-yeast system have been confirmed in plants, including the enrichment of phosphatidylcholine at the viral replication sites in BMV-infected barley cells and an inhibition of BMV replication in BMV-infected *Nicotiana benthamiana* by overexpressing host gene *PAH1* (Phosphatidic Acid Hydrolase 1) or a dominant-negative mutant of host *SNF7* (Sucrose Non Fermenting 7) (see below in the Section “Host Factors in RNA Replication”). However, it should be noted that there are some differences in the replication features between plant and yeast cells as further discussed in the Section “BMV 1a and Viral Replication Compartments”.

Sub-Genomic mRNA Synthesis

Early observations that the nature of CP in BMV infections is dictated by RNA3 rather than RNA4, and that RNA4 was regenerated when omitted from BMV inoculum, were partially explained when sequencing revealed that RNA4 was encoded by the 3′ portion of RNA3. Nevertheless, whether RNA4 was produced from RNA3 by cleavage or any of several possible RNA synthesis pathways remained uncertain. In 1984, it was shown that a BMV RdRp extract produced sub-genomic RNA4 *in vitro* when supplied with negative-strand RNA3 templates, and that the product RNA4 could be labeled by γ -³²P-GTP incorporation, demonstrating *de novo* initiation. This first elucidation of a pathway for sub-genomic mRNA synthesis appears to provide a meaningful precedent for similar sub-genomic mRNA synthesis by many positive-strand RNA viruses, and an important foundation for understanding the diversity of alternate mechanisms that has begun to emerge with the demonstration of distinctly different sub-genomic mRNA synthesis pathways used by coronaviruses, nodaviruses and some other positive-strand RNA viruses.

RNA Recombination

In 1986, BMV was used to provide an early demonstration of RNA recombination in a plant virus. Subsequent work demonstrated many forms and features of inter- and intra-molecular, homologous and non-homologous RNA recombination in BMV. Mutations in BMV RNA replication proteins 1a and 2a^{pol} could alter the frequencies and distributions of crossover sites, implying that at least a significant portion of such recombination was a byproduct of RNA replication, as by template switching. These and other results show that RNA recombination is a major force for repairing BMV genomes damaged by the high mutation rates of viral RNA replication and other events, thereby contributing to BMV survival and adaptability. Subsequently, it was further demonstrated that the recruitment of parental RNAs to a shared replication compartment is a limiting step in intermolecular RNA recombination. These results support a model in which BMV genomic RNAs are individually recruited into replication compartments, and intermolecular recombinants arise only in replication compartments receiving multiple viral RNAs.

Properties and Functions of Viral Gene Products

BMV RNA3 encodes two proteins, 3a and CP, that are dispensable for RNA replication but are required for systemic spread (Fig. 1). Disruption of the 3a gene blocks cell-to-cell movement, limiting infection to individual, directly inoculated cells. The 3a protein shares multiple properties with cell-to-cell movement proteins of other viruses, including cooperative binding to single-stranded RNA, localization to the plasmodesmatal connections, and induction of virion-containing tubules from the surface of BMV-infected protoplasts. Disruption of the CP gene stops virus spread to uninoculated leaves. Whether local cell-to-cell spread occurs in the absence of CP depends on several factors including the 3a allele and the host plant.

Protoplast studies showed that BMV RNA replication and sub-genomic RNA4 synthesis require 1a and 2a^{pol} but not 3a or CP. The conserved central domain of 2a^{pol} shows similarity to RdRps encoded by picornaviruses and many other RNA viruses. The N-terminal 1a domain is related to alphavirus protein nsP1 and contains m⁷G methyltransferase and m⁷GMP covalent binding activities required for capping viral RNA *in vivo*. The C-terminal 1a domain has sequence similarity to super-family I NTPase/helicases and NTPase activity that is required for RNA template recruitment and RNA synthesis. The N- and C-proximal domains are separated by a short proline-rich sequence that may serve as a flexible spacer. As shown in several systems, the C-terminus of BMV 1a interacts with the N-terminus of 2a^{pol} and this interaction is required for the recruitment of 2a^{pol} into the viral replication compartments.

BMV 1a and Viral Replication Compartments

In yeast cells replicating BMV RNAs, the outer perinuclear endoplasmic reticulum (ER) membrane is invaginated towards the lumen to form numerous 50–80 nm spherular invaginations, which are referred to as spherules. Similar membrane invaginations are seen in plant cells infected by viruses from *Bromoviridae*, *Tombusviridae*, and *Tymoviridae* families, and in animal cells infected by viruses in and beyond the alphavirus-like super-family. However, in BMV-infected *N. benthamiana* cells, there are three types of polymorphic vesicles that are associated with cytoplasmic ER membranes but are absent from the perinuclear ER region. Type 1 vesicles are ~66 nm in diameter with some spherule-like structures that remain connected to the ER through neck-like structures and as such, type 1 vesicles resemble those formed in yeast cells. Type 2 vesicles are ~359 nm in diameter and likely are induced

by CP. Type 3 vesicles are most probably formed by fusing type 1 and type 2 vesicles. Some possible reasons for the differences in BMV-induced membrane rearrangements between yeast and *N. benthamiana* are likely due to the organization and lipid composition of the ER membranes in yeast and *N. benthamiana* or possible variation in relative protein expression levels and replication efficiency in each system.

Using immunogold-labeling and electron microscopy in yeast, replication proteins 1a, 2a^{pol}, and nascent BMV RNAs were found to be localized to the spherule interiors, which remain connected to the cytoplasm by a narrow neck that likely allows for the import of ribonucleotides required for replication and the export of offspring RNAs (Fig. 2). Immunogold and biochemical studies show that each spherule contains one or a few hundred 1a proteins and ~25-fold fewer 2a^{pol} proteins along with specific host proteins (Fig. 2 and see the Section on “Host Factors in RNA Replication” below). The structure, assembly, and operation of these spherular replication complexes have functional and perhaps evolutionary links to the replicative cores of retrovirus and dsRNA virus virions, which sequester viral RNA replication templates and their polymerases in a protein shell.

BMV 1a serves as the primary organizer to form active replication compartments: it invaginates the outer ER membranes into the ER lumen to form spherules, recruits 2a^{pol} via 1a-2a^{pol} interactions, recruits RNA templates into the interior of preformed spherules by recognizing the *cis*-element RE present only in viral genomic RNAs, and also interacts with and recruits specific host factors to the site of viral replication (Fig. 2, see below in the Section of “Host Factors in RNA Replication”). Spherular invaginations are formed when 1a is expressed in yeast cells in the absence of any other viral proteins or RNA templates. Localization of 1a to the perinuclear ER membrane requires an N-terminal amphipathic α -helix, termed Helix A (amino acids 392–407). NMR analysis shows that three leucine residues are deeply immersed into SDS micelles that mimicked a lipid bilayer. Substitutions of the three leucine residues with alanine prevent 1a from associating with the perinuclear ER membrane, and block spherule formation. In addition, Helix A plays critical roles in regulating the size and frequency of spherules as several individual substitutions in Helix A lead to similar viral replication defects including the formation of spherules that are more numerous in number but smaller in size.

A second amphipathic α -helix, Helix B (amino acids 415–433), plays a crucial role in 1a's function as a viroporin to permeabilize ER membranes to release oxidizing potential, which requires host thiol oxidase Ero1p (ER Oxidation 1), from the ER lumen to promote the formation of disulfide-bound 1a multimers, 1a's RNA capping activity, and active genome replication (See Ero1p in the Section “Host Genes Required for Viral Replication Activity”).

Host Factors in RNA Replication

As for many other viruses, the small size of BMV genome relative to the complexity of BMV replication suggests that many, if not most, steps in BMV RNA replication involve contributions from host factors. Indeed, genetic and systematic genome-wide screens have identified more than 100 genes that, when deleted, inhibited or enhanced BMV RNA replication by at least 3-fold. The mechanisms by which BMV usurps host factors to complete the various steps of its life cycle are illustrated in Fig. 2.

Host Genes That Regulate Lipid Metabolism and Membrane Composition

Since positive-strand RNA viruses replicate on intracellular membranes, cellular lipid synthesis and appropriate lipid composition are crucial for their replication. BMV replication increases total fatty acids per yeast cells by 33%. Mutation or deletion of the *OLE1* (OLEic acid requiring 1) gene, which encodes for the $\Delta 9$ fatty acid desaturase that converts saturated fatty acids to unsaturated fatty acids (UFAs), revealed that BMV RNA replication requires UFA levels five times higher than those required for normal yeast growth. Moreover, the gene products of *DOA4* (Degradation Of Alpha 4), *SPT23* (SuPpressor of Ty 23), and *MGA2* (Multicopy suppressor of GAM 2), which are involved in regulating the expression of lipid metabolism genes, including *OLE1*, are also critical for proper BMV RNA replication.

As major structural constituents of membranes, phospholipids are particularly important for the formation and activation of viral spherules. Phosphatidic acid is a precursor for both phospholipids and storage lipids. phosphatidic acid can be used to make phospholipids for membrane synthesis during active cell growth, or it can be converted to the storage lipid precursor diacylglycerol by Pah1p when cell growth reaches the stationary stage. Deleting *PAH1* led to a 2-fold increase in total phospholipids levels, resulting in twice as many spherules and a 3-fold increase in viral replication. Further studies showed that BMV replication leads to a 25% increase in the levels of phosphatidylcholine, which is predominantly enriched at the sites of BMV replication in both yeast and barley cells. At least in yeast, BMV 1a interacts with and recruits Cho2p (Choline requiring 2), an enzyme involved in the conversion of phosphatidylethanolamine into phosphatidylcholine, to the viral replication sites to promote phosphatidylcholine synthesis, and facilitates BMV replication. In addition, deleting *ACB1* (Acyl-CoA-Binding 1), another gene that regulates general lipid transport and synthesis, reduced BMV replication by up to 30-fold and it led to the formation of smaller but more abundant spherules compared to those in wild-type cells.

Host Genes That Induce/Stabilize Spherules

Host factors are also essential to form and/or maintain BMV-induced spherules. The reticulons are a family of proteins that stabilize highly curved tubules in the peripheral ER. Through an interaction with BMV 1a, the reticulons are relocated from the

peripheral ER into the interior of the spherules. The reticulons were shown to stabilize the membrane curvature in the neck of the spherules and to also regulate the size and number of spherules. Components of the endosomal sorting complex required for transport (ESCRT) are also essential for spherule formation. One group of ESCRT genes regulate BMV RNA replication and the frequency of spherule formation while a second group of genes regulate RNA replication independent of spherule formation. In particular, deleting *SNF7* significantly reduced viral replication and abolished detectable spherule formation. The current model suggests that BMV 1a initiates the invagination of the perinuclear ER, followed by recruitment of Snf7p and other ESCRT components to constrict the wide membrane rim, allowing the formation of the spherule body and neck. The reticulons in turn stabilize the curvature in the neck and regulate the size of the spherules.

Host Genes That Regulate Translation of Viral Proteins and 1a Targeting

Until recently the mechanism by which BMV 1a is targeted to the perinuclear ER was unclear. It was shown that Erv14p (ER-vesicle protein of 14 kDa), a cargo receptor in COPII (COat Protein complex II) coated vesicles, not only interacts with 1a but helps localize it to the perinuclear ER. Deletion of *ERV14* disrupts the proper distribution of 1a and significantly inhibits BMV RNA replication.

Host ribosomes recognize the genomes of positive-strand RNA viruses to initiate translation of the viral proteins. Once sufficient viral replication proteins accumulate, viral genomic RNAs are recruited out of the cellular translational machinery and into spherules to serve as templates for viral replication. For BMV, control of these two functions is tightly regulated by the host complex consisting of Pat1p (Protein associated with topoisomerase 1) and 7 members of the Lsm (Like SM) family: Lsm1p to Lsm7p. Lsm1–7-Pat1 complex binds to the 5' and 3' untranslated regions (UTR) of BMV RNAs to circularize the RNAs to facilitate translation. When BMV 1a is present, 1a binds to the viral RNA and interacts with the Lsm1–7-Pat1 complex, likely disrupting Lsm1–7-Pat1 mediated 3'-5' circularization, thus repressing translation to allow recruitment of the viral RNA to the spherules. Moreover, classical yeast genetic approaches identified *DED1* (Defines Essential Domain 1), an RNA helicase required for translation of all yeast mRNAs, to regulate translation of 2a^{pol} relative to other replication factors.

Host Genes Required for Viral Replication Activity

Host factors are also essential to activate the viral replication proteins after spherules have been assembled. Ydj1p (Yeast dnaJ 1), a J-protein co-chaperone of Hsp70 and Hsp90, is required to activate the BMV RNA replication complexes. In yeast with a mutation in the *YDJ1* gene, negative-strand RNA synthesis is inhibited even though 1a, 2a^{pol}, and RNA3 were recruited to ER membranes. The results suggest that the Ydj1p chaperone likely modulates 2a^{pol} folding or assembly into the complex. BMV RNA replication also depends on the ER luminal thiol oxidase Ero1p. 1a can act as a viroporin by permeabilizing ER membranes to release the Ero1p-generated oxidizing potential from the ER lumen to activate 1a's capping function via a new state of 1a multimerization involving oxidized linkages and potential conformational changes within the protein. 1a mutants lacking the viroporin function were able to recruit and stabilize positive-strand RNA3 templates but did not support significant RNA replication.

Transmission and Host Range

BMV transmission in plants is not well characterized. BMV can be transmitted in the field by human activities involved in agricultural production, such as machinery trampling. In laboratory settings, plants can be easily infected by mechanical inoculation using sap prepared from BMV-infected tissues, purified virions or virion RNAs, or *in vitro* transcripts from cloned viral cDNAs. Vectors transmitting BMV in the field are unknown. Spotted cucumber beetles and nematodes in the genus *Xiphinema* have been reported to be able to transmit BMV in the laboratory.

BMV replicates and encapsidates its RNAs in directly inoculated cells from a wide variety of plants, but has a fairly restricted host range for systemic infection of whole plants beyond monocotyledonous plants. The effective host range for BMV infection thus appears to be determined at the level of initiating or sustaining infection spread from the sites of primary infection.

Exchanging genomic RNAs, individual genes and gene segments among BMV strains and between BMV and other viruses shows that adaptation for infection spread in particular host plants depends not only on 3a and CP but also on features of RNA1 and RNA2. Host adaptation of 3a generally exerts the predominant effects on infection spread, and only a few amino acid changes in 3a are required to extend BMV host range from monocotyledonous to dicotyledonous plants. However, changes modulating the efficiency of systemic spread also map to 1a and 2a^{pol}. Such changes may alter systemic spread through host-specific effects on RNA replication, as by influencing the ability of the virus to replicate and spread faster than host defense responses. Alternatively, 1a and 2a^{pol} may possess additional functions, as for some C-terminal 2a^{pol} sequences that are dispensable for RNA replication but required for efficient systemic infection.

BMV primarily infects monocotyledonous cereal crops such as barley, maize, rice, wheat, and sorghum, as well as model plants *Brachypodium distachyon* and *Setaria viridis*. While *N. benthamiana* is a dicotyledonous host for BMV, *Arabidopsis thaliana* is generally considered as a non-host even though it allows BMV to infect in rare conditions. Screening numerous accessions and mutants of *Arabidopsis*, CPR5 (Constitutive expresser of Pathogenesis-Related genes 5) was identified as a key restriction factor for BMV systemic infection in *Arabidopsis*. Although *Arabidopsis cpr5* mutant plants show an enhanced resistance to multiple bacterial and oomycete

pathogens and several positive-strand RNA viruses compared to wt plants, they all support efficient systemic BMV infection. BMV RNA replication increased at the single-cell level by 6-fold in protoplasts prepared from *cpr5* mutant plants compared to wt. The allowance for BMV replication in various *cpr5* mutants is independent of host defense-related salicylic acid, jasmonic acid, and ethylene. *Arabidopsis* CPR5 is a component of nuclear pore complexes and serves as a selective barrier to restrict nuclear access of signaling cargos. However, the mechanisms whereby *Arabidopsis cpr5* mutants allow BMV to replicate efficiently in single cells and promote systematic infection are unclear. The role of CPR5 might be direct, or it could indirectly regulate specific downstream effectors that might be responsible for the non-host resistance in *Arabidopsis* and possibly other dicotyledonous plants.

BMV Application in Biotechnology

BMV was used to produce the first infectious transcripts from cloned RNA virus cDNA in 1984. Specially designed BMV cDNA clones were transcribed to produce capped *in vitro* transcripts of genomic RNAs 1–3, each with the natural viral RNA 5' end and only a few extra nucleotides at the 3' end. Mixtures of all three BMV RNA transcripts, but not their parent cDNA clones, were infectious to barley plants, a natural BMV host.

This ability to engineer the expression of infectious transcripts provided a means to manipulate the viral RNA genome at the cDNA level using recombinant DNA technology, which has subsequently proved applicable to many other RNA viruses. In one of the first applications of these new reverse genetics approaches, it was successfully demonstrated that foreign genes could be inserted into the viral genome while retaining the ability to replicate and express genes. Using a transcribable BMV RNA3 cDNA clone, the CP gene was replaced with the bacterial reporter gene chloramphenicol acetyltransferase (CAT). When *in vitro* transcribed and inoculated onto barley protoplasts with RNA1 and 2 transcripts, this RNA3 derivative was replicated and produced CAT activity at higher levels than previously achieved by DNA-based transformation. This first demonstration that RNA viruses can be engineered at the cDNA level showed that the viral RNA genome functions in a sufficiently flexible and modular fashion to tolerate even large changes such as whole gene replacements without substantial optimization, which has significant implications for virus evolution, biotechnology research and applications such as additional gene expression vectors.

RNA interference (RNAi) or gene silencing has been widely employed as a powerful tool to reveal gene functions in plants by degrading gene-specific transcripts. Virus-induced gene silencing (VIGS) knocks down gene expression of targeted cellular genes during virus infection, without a need to make transgenic plants. A BMV strain (fescue strain) isolated from tall fescue has been engineered as a VIGS vector by inserting a fragment of the target genes into the 3' UTR of RNA3. These BMV vectors have been successfully used in barley, maize, rice, and sorghum to understand gene function otherwise hard to achieve.

Future Perspectives

Through a variety of intrinsic features and the work of many investigators, research on BMV not only has advanced understanding of bromoviruses, but also has contributed significantly to general virology and molecular biology. Some of the challenges for the future include improved definition and analysis of distinct substeps in viral RNA synthesis including initiation, elongation, termination and capping; better characterization at molecular, cellular and tissue levels of the pathways and mechanisms involved in infection spread and the interplay of virus-directed processes and host defenses; and improved understanding of the linkages between different infection processes including regulated gene expression, RNA replication, encapsidation and spread. Since the mechanism of action has been characterized for several host genes involved in BMV replication, the manipulation of host genes via knockout, knockdown, targeted substitutions, or overexpression could potentially be used as a way to achieve viral control in crop plants.

See also: Alfalfa Mosaic Virus (*Bromoviridae*). Bromoviruses (*Bromoviridae*). Cucumber Mosaic Virus (*Bromoviridae*). Ilarviruses (*Bromoviridae*)

Further Reading

- Ahlquist, P., 2006. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nature Review Microbiology* 4, 371–382.
- Chaturvedi, S., Rao, A.L.N., 2018. Molecular and biological factors regulating the genome packaging in single-stranded positive-sense tripartite RNA plant viruses. *Current Opinion in Virology* 33, 113–119.
- Cheng, K.C., Sivakumaran, K., 2000. Brome mosaic virus, good for an RN virologist's basic needs. *Molecular Plant Pathology* 1, 91–97.
- Diaz, A., Wang, X., 2014. Bromovirus-induced remodeling of host membrane during viral RNA replication. *Current Opinion in Virology* 9, 104–110.
- Novelty, A.O., Ahlquist, P., 2003. Brome mosaic virus RNA replication: Revealing the role of the host in RNA virus replication. *Annual Review Phytopathology* 41, 77–98.
- Rao, A.L., Cheng, K.C., 2015. The brome mosaic virus 3' untranslated sequence regulates RNA replication, recombination, and viral assembly. *Virus Research* 206, 46–52.
- Szuba-Solinska, J., Bujarski, J.J., 2008. Insights into the single-cell reproduction cycle of members of the family *Bromoviridae*: Lessons from the use of protoplast systems. *Journal of Virology* 82, 10330–10340.

Bromoviruses (*Bromoviridae*)

Jozef J Bujarski, Northern Illinois University, DeKalb, IL, United States and Polish Academy of Sciences, Poznan, Poland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
CP Coat protein or capsid protein
ELISA Enzyme-linked immuno-sorbent assay
ER Endoplasmic reticulum
kb Kilobases; the size of a ssDNA or ssRNA molecule
kDa Kilodaltons; the size of a protein
MP Movement protein
Mr Relative molecular mass
NGS Next generation sequencing
nm Nanometer(s)
ORF Open reading frame(s)

PCR Polymerase chain reaction
RBD RNA binding domain
RdRp RNA-dependent RNA polymerase
RT-PCR Reverse transcription polymerase chain reaction
sgp Sub-genomic promoter
sgRNA Sub-genomic RNA
ssRNA Single-stranded ribonucleic acid
TLS Transfer RNA-like structures
UTR Untranslated region
VIGS Virus-induced gene silencing
VRC Virus replication complexes

Glossary

Cross Protection It describes a phenomenon in that an initial infection of a host plant with a mild strain of a virus induces resistance in that plant to the infection of another, closely related virus potentially protecting the plant from disease caused by a more virulent isolate.

Genome Activation Ilarviral genomic RNAs alone are unable to establish infection in plants, unless the coat protein is present. This function of the coat protein is termed genome activation and is specific for ilarviruses and closely related alfamoviruses. The event is triggered by binding of the coat protein RNA binding domain to the 3' terminus of genomic RNA.

RNA Silencing A fundamental, evolutionarily conserved and sequence-specific mechanism that is triggered by double-stranded RNA and regulates gene expression in

eukaryotes. It is a primary antiviral defense mechanism in plants and other living organisms.

RNA Silencing Suppressor A countermeasure to RNA silencing, often a protein encoded by a virus which interrupts a single, or multiple steps in the RNA silencing pathway such as binding to small interfering RNA and thereby preventing their incorporation into the RNA induced silencing complex.

Sub-genomic RNA A segment of RNA generated from a genomic RNA via an internal promoter that has the same 3' end as the genomic RNA, but has a deletion at the 5' end. The sub-genomic RNA makes it possible to efficiently translate the downstream open reading frame of the genomic RNA.

Tripartite Genome A viral genome consisting of three genomic fragments, which are encapsidated into three separate virions.

Introduction

The family *Bromoviridae* contains important genera of plant viruses, with host ranges varying from narrow to very wide, and infecting herbaceous plants, shrubs and even trees. Several of them are responsible for major epidemics in fodder crops such as tomato, cucurbits, bananas, or alfalfa. The members of the family *Bromoviridae* have spherical or bacilliform particles with a tri-segmented, positive-sense, single-stranded RNA (ssRNA) genome, packaged in separate virions. Bromovirids can be transmitted mechanically, via the pollen, seeds or by insect vectors.

As shown in [Tables 1](#) and [2](#), the *Bromoviridae* family includes six genera: *Alfavirus* (one member, type species: *Alfalfa mosaic virus*, AMV), *Anulavirus* (two members, type species: *Pelargonium zonate spot virus*, PZSV), *Bromovirus* (six members, type species: *Brome mosaic virus*, BMV), *Cucumovirus* (four members, type species: *Cucumber mosaic virus*, CMV), *Ilarvirus* (22 members, type species: *Tobacco streak virus*, TSV), and *Oleavirus* (one member, type species: *Olive latent virus 2*, OLV-2). Genus demarcation criteria are based on natural host range, method of transmission, detailed morphology and properties of particles, organization of RNA genome, replication schemes and producing defective RNAs and satellite RNAs.

The two prototype genera, *Bromovirus* and *Cucumovirus*, are the genera mostly related, with the latter being agriculturally important. Both bromo- and cucumoviruses share such properties like the molecular and genetic features of their tripartite RNA genome, the number of encoded proteins and similar virion structure. The computer-assisted comparisons of aa sequences reveal significant similarity among their RNA replication proteins, much beyond the presence of characteristic GDD motif for 2a or for helicase/transferase domains in 1a. More broadly, the replication proteins share aa sequence similarity within the alphavirus-like super-family of positive-strand RNA viruses, which includes numerous plant and important animal/human viruses. The type members of different genera, such as CMV, BMV and

Table 1 Main characteristics of the RNA genome in six genera of the family *Bromoviridae*^a

Genus	Acronym	RNA1	RNA2	RNA3	3' UTR	sgRNA/diRNA
Alfavirus						
<i>Alfalfa mosaic virus</i>	(AMV)	3644	2593	2037	Complex	–/–
Anulavirus						
<i>Pelargonium zonate spot virus</i>	(PZSV)	3383	2435	2639	Complex	–/–
Bromovirus						
<i>Brome mosaic virus</i>	(BMV)	3234	2865	2117	tRNA-like	+ / –
Cucumovirus						
<i>Cucumber mosaic virus</i>	(CMV)	3357	3050	2216	tRNA-like	+ / +
Ilarvirus						
<i>Tobacco streak virus</i>	(TSV)	3491	2926	2205	Complex	–/–
Oleavirus						
<i>Olive latent virus 2</i>	(OLV-2)	3126	2734	2438	Complex	+ / –

^aPartial sequence.

AMV, have and continue to constitute excellent models for molecular research on viral gene expression, RNA replication, virion assembly, RNA recombination, epidemiology or the role of cellular genes in basic virology.

Phylogeny and Biodiversity of the Family *Bromoviridae*

Although RNA1, 2 and 3 overall keep a great similitude in their sequences, clearly rearrangements of RNAs has been done for members of the Bromoviridae. Both RNA recombination but also segment reassortment played a major role as the sources of variation in shaping the bromovirids member groups, being important contributors to the evolutionary history of the family, especially for the genera *Bromovirus*, *Cucumovirus* and *Ilarvirus* (Figs. 1 and 2). However, doubts have been shed on the biological significance of the official taxonomy of the family *Bromoviridae*. To better understand the taxonomy, attempts have been made to reconcile the incongruences observed in the viruses' evolutionary radiation caused by recombination and reassortment. These two processes could create new genetic variability and then these primary variants would undergo further selection for functional genomes of individual viruses. Consequently, the variants generated by reassortment and recombination events must have been initially viable which represents the first selective filter while further directional selection fine tunes the newly created RNAs. RNA segment reassortment was probably common at the origin of the bromoviruses and cucumoviruses as well as at the origin of Alfalfa mosaic virus, American plum line pattern virus and Citrus leaf rugose virus. Furthermore, recombination analyzes done for each of the three genomic RNAs revealed very common crossovers within the members of the genera *Bromovirus*, *Cucumovirus* and *Ilarvirus*, but also mixed recombination involving species from different genera. It seems that bromoviruses and cucumoviruses did split from a common ancestor forming distinct clades due to crossover events in RNA3, whereas protein 2b promoted the selection of a CMV-TAV RNA1/2-RNA3 recombinant. In the 5' untranslated regions (UTR) of CMV RNA3 the sequence rearrangements have likely been the precursors of the radiation of three cucumovirus subgroups. The results illustrated in Fig. 2 confirm a clear separation between the genera *Bromovirus* and *Cucumovirus*, while the ilarviruses constitute their own cluster; two other genera (*Anulavirus* and *Oleavirus*) are more unique within the family. Although these results suggest that AMV should be included in the ilarviruses, its unequivocal assignment has yet to be resolved, especially because AMV differs from other ilarviruses with the mode of transmission, by aphids versus by pollen and thrips. Bromovirids are members of a larger alpha-like supergroup based upon both 1a and 2a proteins whereas 3a proteins cluster the bromovirids together with other viral groups into a separate pool of movement-associated proteins. In general, the constructed phylogenetic network not only reflects the initial genetic exchanges but also confirms the taxonomic status of the different genera within the family *Bromoviridae*, notwithstanding the phylogenetic disturbances caused by genetic exchange.

Virion Properties and Structure

Virions of bromovirids are non-enveloped, being either spherical or pseudo-spherical, with $T = 3$ icosahedral symmetries, and a diameter of 26–35 nm (genera *Anulavirus*, *Bromovirus*, *Cucumovirus* and *Ilarvirus*), whereas genera *Alfavirus* and some ilarviruses have bacilliform virions, of diameters 18–26 nm and lengths of 30–85 nm. In *Oleavirus* virions have different shape.

In genus *Bromovirus* three types of icosahedral particles are composed of 180 molecules of a 20 kDa CP, and encapsidate different RNA components: RNA1 (Mr c. 1.1×10^6), RNA2 (Mr c. 1.0×10^6) and RNA3 plus sgRNA4 (Mr c. 0.75×10^6 and 0.3×10^6). In addition to viral RNAs, BMV virions have been recently reported to package small amounts of host RNAs, with their

Table 2 List of genera and species in the family *Bromoviridae*. Type species are written in bold

Genus	Species	Acronym	GenBank accession no.		
			RNA 1 (P1)	RNA 2 (P2)	RNA 3 (MP and CP)
Alfavirus	Alfalfa mosaic virus	(AMV)	NC_001495	NC_002024	NC_002024
Anulavirus	<i>Amazon lily mild mottle virus</i>	(ALMMoV)	NC_018402	NC_018403	NC_018404
	Pelargonium zonate spot virus	(PZSV)	NC_003649	NC_003650	NC_003651
Tentative species	<i>Cassava Ivorian bacilliform virus</i>	(CsiBV)	NC_025482	NC_025483	NC_025484
Bromovirus	Broad bean mottle virus	(BBMV)	NC_004006	NC_004007	NC_004008
	Brome mosaic virus	(BMV)	NC_002026	NC_002027	NC_002028
	<i>Cassia yellow blotch virus</i>	(CsYBV)	NC_006999	NC_007000	NC_007001
	<i>Cowpea chlorotic mottle virus</i>	(CCMV)	NC_003543	NC_003541	NC_003542
	<i>Melandrium yellow fleck virus</i>	(MeYFV)	NC_013266	NC_013267	NC_013268
	<i>Spring beauty latent virus</i>	(SBLV)	NC_004120	NC_004121	NC_004122
Cucumovirus	Cucumber mosaic virus	(CMV)	NC_002034	NC_002035	NC_001440
	<i>Gayfeather mild mottle virus</i>	(GMMoV)	NC_012134	NC_012135	NC_012136
	<i>Peanut stunt virus</i>	(PSV)	NC_002038	NC_002039	NC_002040
	<i>Tomato aspermy virus</i>	(TAV)	NC_003837	NC_003838	NC_003836
Ilarvirus	Subgroup 1				
	<i>Ageratum latent virus</i>	(ALV)	NC_022127	NC_022128	NC_022129
	<i>Blackberry chlorotic ringspot virus</i>	(BChRSV)	NC_011553	NC_011554	NC_011555
	<i>Parietaria mottle virus</i>	(PaMoV)	NC_005848	NC_005849	NC_005854
	<i>Privet ringspot virus</i>	(PrRSV)	NC_027928	NC_027929	NC_027930
	<i>Strawberry necrotic shock virus</i>	(StNSV)	NC_008706	NC_008707	NC_008708
	Tobacco streak virus	(TSV)	NC_003844	NC_003842	NC_003845
	Subgroup 2				
	<i>Asparagus virus 2</i>	(AsV2)	NC_011808	NC_011809	NC_011807
	<i>Citrus leaf rugose virus</i>	(CiLRV)	NC_003548	NC_003547	NC_003546
	<i>Citrus variegation virus</i>	(CVV)	NC_009537	NC_009538	NC_009536
	<i>Elm mottle virus</i>	(EMoV)	NC_003569	NC_003568	NC_003570
	<i>Lilac ring mottle virus</i>	(LRMoV)	EU919668 ^a	NC_038777	NC_038776
	<i>Spinach latent virus</i>	(SpLV)	NC_003808	NC_003809	NC_003810
	<i>Tomato necrotic streak virus</i>	(ToNSV)	NC_039075	NC_039074	NC_039076
	<i>Tulare apple mosaic virus</i>	(TAMV)	NC_003833	NC_003834	NC_003835
	Subgroup 3				
	<i>Apple mosaic virus</i>	(ApMV)	NC_003464	NC_003465	NC_003480
	<i>Blueberry shock virus</i>	(BISV)	NC_022250	NC_022251	NC_022252
	<i>Lilac leaf chlorosis virus</i>	(LiLChV)	NC_025477	NC_025478	NC_025481
	<i>Prunus necrotic ringspot virus</i>	(PNRSV)	NC_004362	NC_004363	NC_004364
	Subgroup 4				
	<i>Fragaria chiloensis latent virus</i>	(FCLV)	NC_006566	NC_006567	NC_006568
	<i>Prune dwarf virus</i>	(PDV)	NC_008039	NC_008037	NC_008038
	Unassigned species				
	<i>American plum line pattern virus</i>	(APLPV)	NC_003451	NC_003452	NC_003453
	<i>Apple necrotic mosaic virus</i>	(ANMV)	NC_040469	NC_040471	NC_040470
	<i>Cape gooseberry virus 1</i>	(CGV1)	NC_040393	NC_040392	NC_040394
	<i>Humulus japonicus latent virus</i>	(HuJLV)	NC_006064	NC_006065	NC_006066
	<i>Tea plant line pattern virus</i>	(TPLPV)	NC_040435	NC_040436	NC_040437
Oleavirus	Olive latent virus 2	(OLV)	NC_003673	NC_003674	NC_003671

^aPartial sequence.

functions yet to be determined. The members of the genus *Cucumovirus*, in addition to three genomic RNAs, encapsidate two sub-genomic RNAs (sgRNAs) and a satellite RNA.

The crystal structures of both BMV and CCMV have been resolved showing very similar organization (Fig. 3), with the CP subunits folded into a beta-barrel core organized within the protruding both pentameric and hexameric capsomers. The interactions among hydrophobic aa residues stabilize the capsomers, and the hexameric subunits are further stabilized via interactions between N-terminal portions, where six short beta-strands form a tubule called beta-hexamer. Mutational analysis demonstrated that beta-hexamer was not required for virion formation but rather modulated virus spread *in planta*. In addition, the capsomers

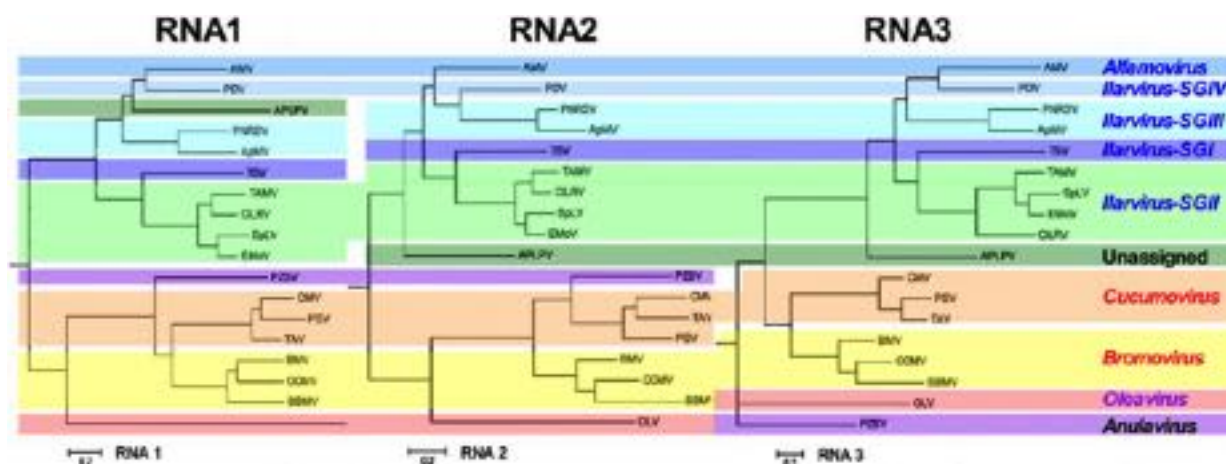


Fig. 1 Phylogenetic trees obtained for the three genomic segments RNA1, RNA2 and RNA3, by concatenating coding and non-coding regions and fitting a heterogeneous nucleotide-substitution model using CODEML. See Table 2 for virus names and abbreviations. The three trees have been aligned with a color-coded system for each genus to show similarities and differences. The trees are reproduced from Codoner, F.M., Elena, S.F., 2008. The promiscuous evolutionary history of the family *Bromoviridae*. *Journal of General Virology* 89, 1739–1747 with permission (modified by C. Fauquet).

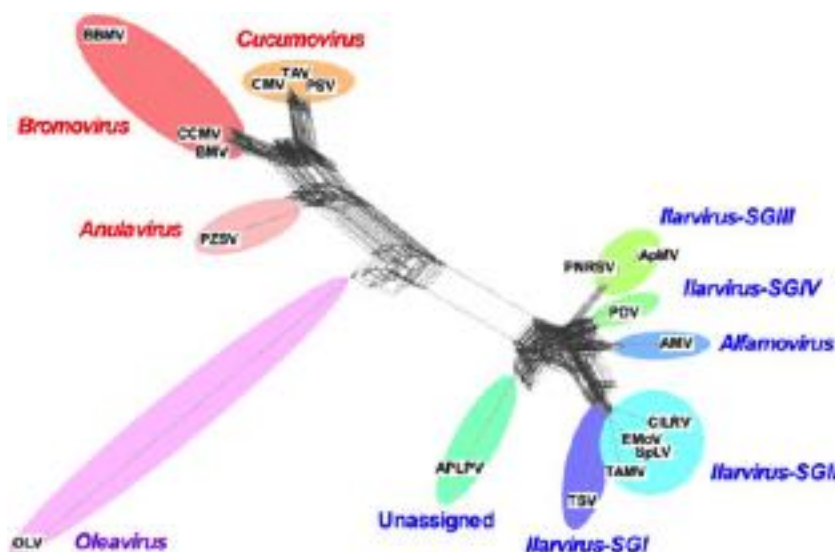


Fig. 2 Evolution of the family *Bromoviridae*. Unrooted phylogenetic network illustrating the evolutionary history of the family *Bromoviridae*. The different viral species in the family are linked to each other via multiple paths of interconnecting network rather than as single tree, which suggests the effects of recombination and reassortment events. Figure reproduced from Codoner, F.M., Elena, S.F., 2008. The promiscuous evolutionary history of the family *Bromoviridae*. *Journal of General Virology* 89, 1739–1747 with permission (modified by C. Fauquet).

are held together by interactions through C-terminal portions that extend radially from the capsid. The C-termini are anchored between the beta-barrel core and the N-proximal loop, and this interaction might be responsible for initiation of assembly of CCMV capsids. The structure of BMV is organized similarly to that of CCMV such that both capsids undergo well-studied reversible structural transitions where shifting pH from 5.0 to 7.0 causes capsid expansion. However, some CP mutations can further stabilize the capsids. Capsids are also stabilized by metals at multiple binding sites that coordinate the amino acids from adjacent CP subunits.

The packaged RNAs interact with the basic N-terminal aa of the CP in the torus-shaped sub-shell inside the BMV capsid so to neutralize the phosphate groups. Other sites of RNA interaction localize to the internally-proximal basic aa of the CP subunits. The RNA encapsidation signals have been mapped on BMV RNAs, especially to both the 3' UTR and a central sequence in RNA3. The co-packaging of sgRNA4 is contingent upon both RNA replication and translation of CP.

The detailed knowledge of CCMV capsids provided opportunities in nanotechnology, e.g., for the reversible pH-dependent gating, useful during the regulation of size-constrained biomimetic mineralization. The interior surface of CCMV capsids can be engineered of as differentially functionalized CP subunits, to act e.g., as a ferritin surrogate that spatially constrains the formation

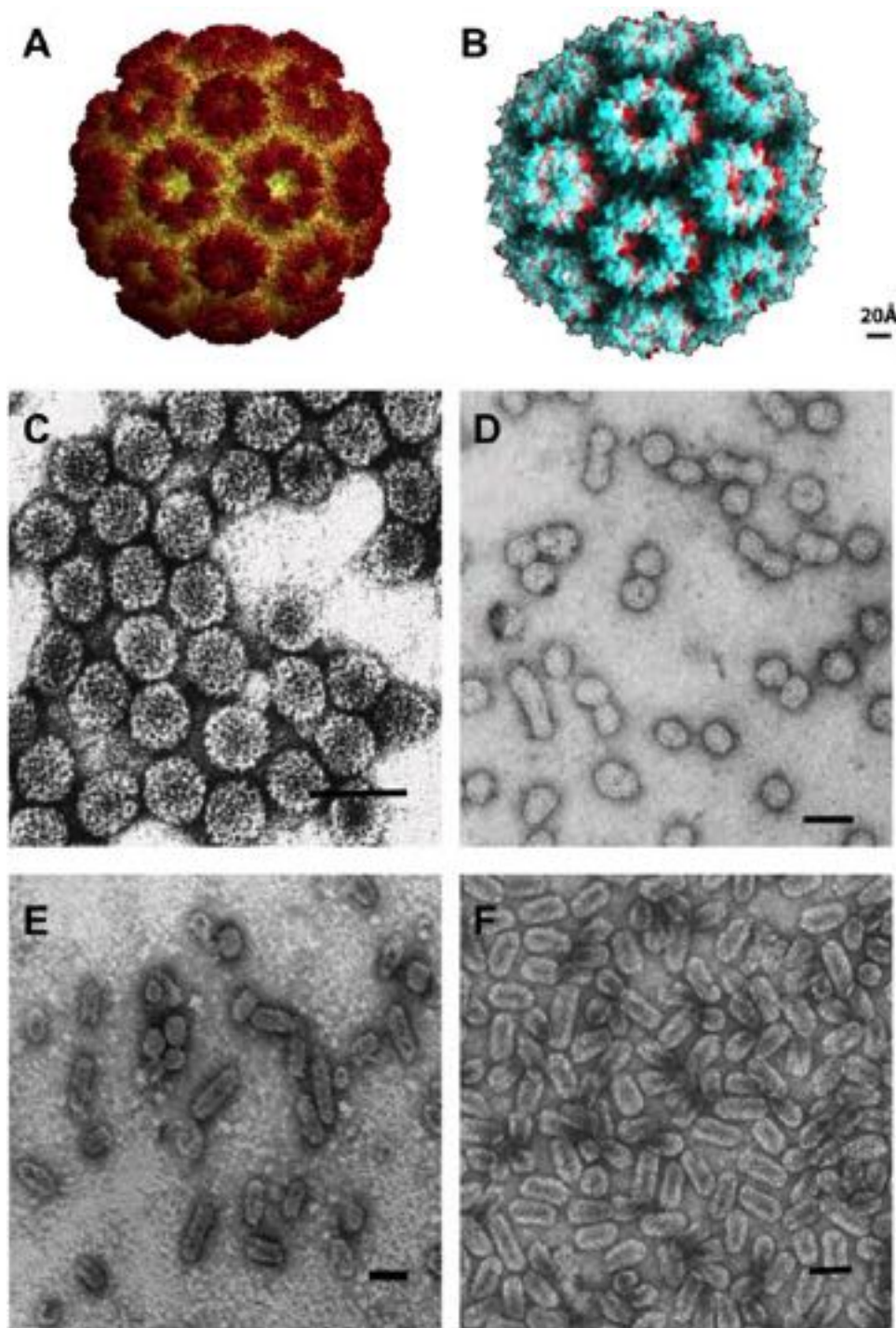


Fig. 3 Montage of transmission electron microscopic photographs and computer rendering of molecular particle structures of bromovirids. A & B. Surface capsid structure of the Brome mosaic virus (BMV) (A) and Cowpea chlorotic mottle virus (CCMV) (B). The hexameric and the pentameric structural elements are visible. Both pictures are from the virion picture collection at the web-site of the Institute for Molecular Virology at the University of Wisconsin-Madison. C–F. Negative-contrast electron micrograph of particles of (C) Brome mosaic virus (*Bromovirus*), (D) Prunus necrotic ringspot virus (*Ilarvirus*), (E) Alfalfa mosaic virus (*Alfavirus*), and (F) Olive latent virus 2 (*Oleavirus*). Photographs A–C IMV-Michigan State University, 1994 Thorben Lundsgaard, C.J. Woolston and Ed Rybicki, and Photograph D courtesy of A. Paredes, NCTR/ORR, Arkansas USA. Photographs E and F, courtesy of A. De Stradis, IPSP-CNR, Bari, Italy. Bars = 50 nm.

of iron oxide nanoparticles. Also, the electrostatically driven adsorption on Si and amine-functionalized Si as well as the fabrication of multilayer CCMV films have been reported. The electrostatically patchy protein cages of CCMV can be used to direct the assembly of super lattices for RNA encapsulation.

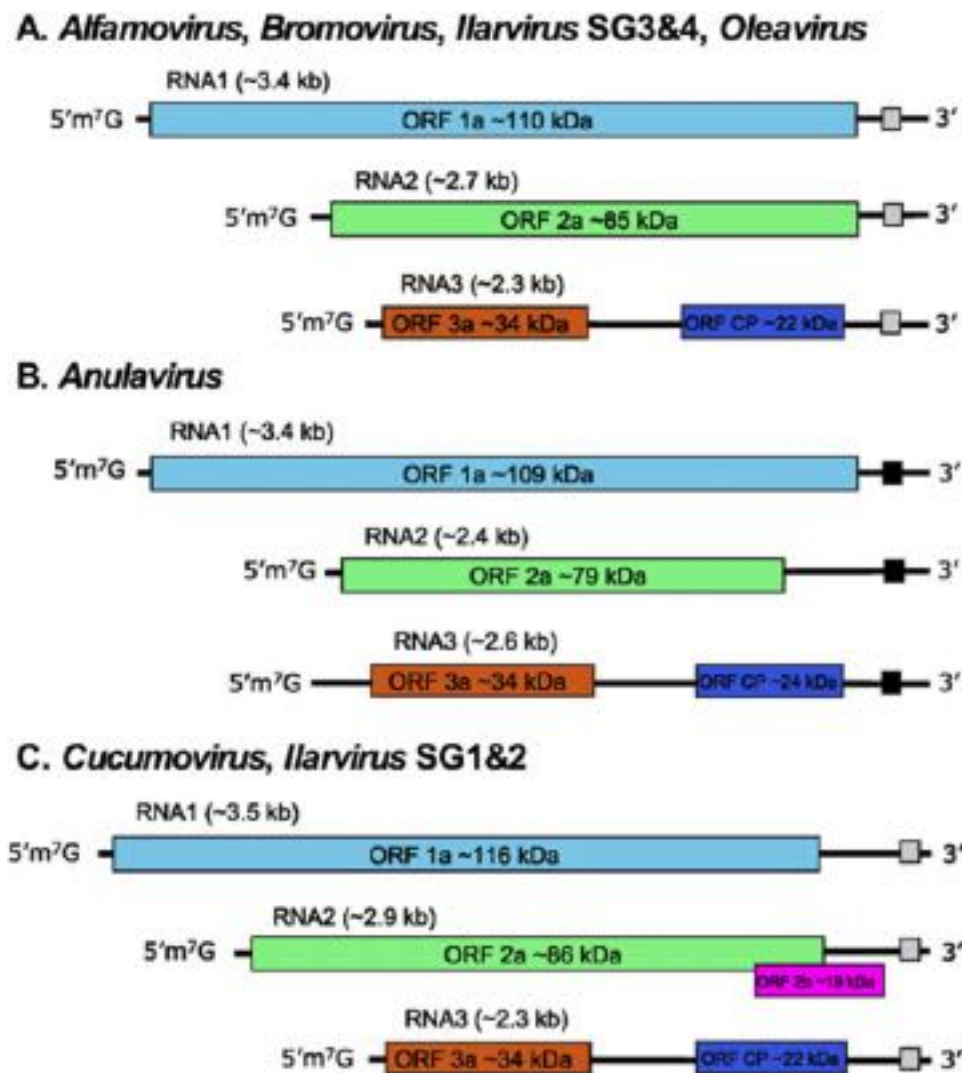


Fig. 4 Genome organizations representatives of the six genera of the family *Bromoviridae*: (A) genus *Alfamovirus*, *Bromovirus*, *Ilarvirus* subgroups 3 and 4 and *Oleavirus*. (B) genus *Anulavirus*. (C) genus *Cucumovirus* and *Ilarvirus* subgroups 1 and 2. The 3' termini form either tRNA-like (B) or complex structures (A, C) shown as black square boxes. Figure from Bujarski, J., Gallitelli, D., García-Arenal, F., *et al.*, 2019. ICTV Virus Taxonomy Profile: *Bromoviridae*. Journal of General Virology 100, 1206–1207.

Genome Organization and Expression

The total length of the *Bromoviridae* RNA genome is approximately 8 kb, with three RNA segments capped at the 5' terminus. Whereas the highly conserved within members, the not polyadenylated 3' termini fold into strong secondary structures. These structures are either aminoacylable tRNA-like (genera *Bromovirus* and *Cucumovirus*) or forming other not aminoacylated arrangements (genera *Alfamovirus*, *Anulavirus*, *Ilarvirus* and *Oleavirus*) (Table 1). RNAs 1 and 2 are monocistronic and they code, respectively, for viral replicase proteins 1a and 2a (Fig. 4). Protein 1a has two distinct domains of guanylyl transferase and helicase, and it is involved in anchoring the replicase complex to the endoplasmic reticulum membrane, and induces the formation of membranous vesicular mini-organelles called spherules where RNA replication occurs. In AMV the replicase proteins interact with the tonoplast. Protein 2a is the actual RNA-dependent RNA polymerase enzyme that interacts with protein 1a and synthesizes the vRNAs. Mutations and deletions in 1a/2a ORFs helped to identify the regions active in BMV RNA replication as well as regions responsible for interaction with the cellular membrane or for interactions between 1a/2a polypeptides. An active BMV replicase preparation has been extracted. The anulavirus encodes the smallest RdRp (2a) protein within the family. For cucumoviruses and in some ilarviruses RNA2 is dicistronic encoding a protein 2b, that is part of the C-terminal region of the protein 2a. Protein 2b was found to act as suppressor of RNA interference, being involved in the inhibition of viral gene silencing but also in systemic movement and affecting the symptoms.

RNA3 encodes the movement (MP) and coat (CP) proteins, the latter being translated from the subgenomic (sg)RNA4. BMV CP is a multi-functional protein. In addition to its structural/encapsidation role, it also coordinates the viral infection processes including (1) participation in the formation of replication factories, (2) repression of RNA replication but also translation, and (3) stimulation of BMV RNA accumulation at lower CP levels. Moreover, the BMV CP participates in RNA recombination events; an analogous function to that of BMV CP was assigned to nucleocapsid proteins in retroviruses and coronaviruses. The multiple functions of CP are exercised by effective binding to several distinct sites in RNA3, including the 3' non-coding hairpin, two central regions, and possibly at the 5' end. The contact BMV CP amino acids have been mapped at distinct parts of the CP monomers. Since BMV replicase complex also binds to most of these sites it has been postulated that the CP is involved in the regulation of BMV RNA replication such that the repression of RNA accumulation and translation occurs at higher levels of BMV CP, while stimulation of BMV RNA accumulation at the lower CP levels. While CP is usually translated from the encapsidated 3' sgRNA4, Olive latent virus 2 encapsidates a sgRNA with no apparent messenger activity whereas CP is translated from another non-encapsidated sgRNA4. The purified genomic RNAs are directly infectious but for some bromovirids the presence of CP is necessary (e.g., for AMV or ilarviruses).

The 32 kDa MP of bromovirids has a conserved RNA binding domain that binds to vRNA and assists with viral transport. There are two groups among the members of the family *Bromoviridae* regarding the cell-to-cell transport; those that do not require participation of virions and those that transport whole virions. In the first group only the MP-RNA entity is transported (non-virion transport) in a form of either the complex of MP and vRNA, or a triple complex of vRNA, MP and CP. For the first group, the prototype example is TMV whereas members of the genus *Bromovirus* have a similar mechanism, which has been demonstrated for CCMV. The MP of these viruses belongs to the 30K superfamily of MPs that appear to interact with cellular microtubules. In the second subtype, where the transported complex is vRNA-MP-CP, the typical example is CMV, a comovirus. In this case, the CMV MP exhibits the binding affinity to the actine microfilaments. The second group transports whole virions intercellularly through plasmodesmata inside the microtubules (virion transport). BMV and AMV are the members of two genera that are transported this way. Also for ilarviruses such as PDV or PNRSV, their MPs likely assist during translocation of the entire viral particles alongside the tubular structures. It appears that MPs of these viruses interact with virion CP subunits via a 44C-terminal key aa domain.

For long distance transport, most viruses require the CP which suggests that they are transported in the form of viral particles (virions). This has been demonstrated by showing that bromovirids require unmodified CP and the wild type C-terminus of MP for long distance spread, such as AMV, BMV and CMV. Very likely the CP-MP interactions enhance the systemic transport, independently of the mechanism of short distance (cell-to-cell) transport.

Genomic RNA Replication and Recombination

Replication of bromovirus RNAs are the best studied among the members of the family *Bromoviridae*. Only viral proteins 1a (helicase and methyl-transferase domains) and 2a (RdRp), but not the proteins 3a or CP, are required for RNA synthesis (Fig. 4), first demonstrated for BMV, both in plant and in yeast cells. The cytoplasmic RNA replicase complex localizes to the endoplasmic reticulum membranes called spherules. The extracted bromoviral RdRp preparations have allowed for mapping in vitro on three genomic BMV RNAs the promoter of (–) strand synthesis within the 3' UTR. The promoters of (+) strand synthesis have also been mapped to the 5' proximal non-coding region in BMV RNAs. Likewise, the sub-genomic promoter (sgp) has been localized as a 100 nt subset of the 250 nt intercistronic region in (–) strand in BMV RNA3, being responsible for synthesis (transcription) of sgRNA4. In addition, in the plus strand the 250 nt intercistronic sequence supports other functions including the efficient RNA3 replication, the maintenance of a proper ratio of (+) to (–) strands of RNA3, stabilization of RNA3 via interaction with protein 1a, synthesis (transcription) of sgRNA4, and the assembly of the active RNA replication complex. It also serves as an efficient RNA recombination hot spot. Some bromoviruses as well as togaviruses carry the internal poly(A) tract, as part of the intercistronic region of the RNA3 segment. Besides viral RNA sequences and viral proteins, a variety of essential host genes affecting BMV RNA replication have been identified, by using yeast knockout libraries, as apparently the yeast cells can support a complete replication cycle of this virus Fig. 5.

Both homologous and non-homologous RNA-RNA crossovers has been demonstrated between bromoviral RNAs during infection. Homologous recombination has also been shown during co-infection between two strains of BMV, with some distinct hot spots localized within both the coding and non-coding regions. The role in recombination of proteins 1a and 2a as well as of the CP have been demonstrated in BMV, suggesting the template switching as a likely mechanism for RNA recombination. For the cucumoviruses, the control of recombination frequency resides mainly in the 2a gene.

Members of the genera *Bromovirus* and *Cucumovirus* are capable of producing the defective (d)RNAs and defective-interfering (di)RNAs during infection (Table 1). In particular, strains of Broad bean mottle virus (BBMV) do accumulate RNA2-derived deletion variants after serial passages through broad bean. In BMV, both replicating and non-replicating truncated RNA2-derived artificial diRNAs have been shown to interfere with BMV RNAs in protoplasts. For CMV, several types of RNA3-derived diRNAs have been described.

Biology

The family *Bromoviridae* is one of the most important families of plant RNA viruses, with some members widely distributed in the world. In its entirety, the family has a wide host range (more than 10,000 species) and some members are causing agronomically



Fig. 5 Severe outbreaks of Cucumber mosaic virus (CMV; *Cucumovirus*) containing the necrogenic satRNA (A) and Pelargonium zonate spot virus (PZSV; *Anulavirus*) (B) in crops of canning tomato in southern Italy. Insets show disease symptoms on fruits. Photographs courtesy of ICTV.

important diseases. However, the host range of members of individual genera ranges from significantly narrow (genera *Bromovirus*, *Oleavirus*) to extremely broad (genus *Cucumovirus*). CMV can infect one of the largest number of plant species among plant viruses. Some of these viruses cause major disease epidemics in vegetables, fodder and fruit crops, e.g., in tomato, cucurbits, bananas, or alfalfa, and in fruit trees (ilarviruses). Different virus species are transmitted mechanically, via pollen/thrips, through seeds or by insect vectors like aphids or beetles. It has been speculated that lack of inefficient vectors evolved some bromovirids toward producing larger concentration of viral particles. For CMV, although the virus is prone to recombination, the recombinants are rare during infection, suggesting the presence of strong selection bottlenecks. No direct correlation between virus yield and symptom severity have been observed, but rather the symptoms seem to be associated with changes in specific regions in the RNA genome, as it has been mapped by using the natural strains of BMV, BBMV or CCMV.

Further Reading

- Bujarski, J., Gallitelli, D., García-Arenal, F., *et al.*, 2019. ICTV Virus Taxonomy Profile: *Bromoviridae*. *Journal of General Virology* 100, 1206–1207.
- Chaturvedi, S., Rao, A., 2018. Molecular and biological factors regulating the genome packaging in single-strand positive-sense tripartite RNA plant viruses. *Current Opinion in Virology* 33, 113–119.
- Codoner, F.M., Elena, S.F., 2008. The promiscuous evolutionary history of the family *Bromoviridae*. *Journal of General Virology* 89, 1739–1747.
- Diaz, A., Wang, X., 2014. Bromovirus-induced remodeling of host membranes during viral RNA replication. *Current Opinion in Virology* 9, 104–110.
- Gancarz, B.L., Hao, L., He, Q., Newton, M.A., Ahlquist, P., 2011. Systematic identification of novel, essential host genes affecting bromovirus RNA replication. *PLoS One* 6 (8), e23988.
- Kostiainen, M.A., Hiekkataipale, P., Laiho, A., *et al.*, 2013. Electrostatic assembly of binary nanoparticle superlattices using protein cages. *Nature Nanotechnology* 8, 52–56.
- Pallas, V., Aparicio, F., Herranz, M.C., Sanchez-Navarro, J.A., Scott, S.W., 2013. The molecular biology of ilarviruses. *Advances in Virus Research* 87, 139–181.
- Schoelz, J.E., Harries, P.A., Nelson, R.S., 2011. Intracellular transport of plant Viruses: Finding the door out of the cell. *Molecular Plant* 4 (5), 813–831.
- Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M., Bujarski, J.J., 2011. RNA-RNA recombination in plant virus replication and evolution. *The Annual Review of Phytopathology* 49, 415–443.
- Takeshita, M., Matsuo, Y., Suzuki, M., *et al.*, 2009. Impact of a defective RNA 3 from cucumber mosaic virus on helper virus infection dynamics. *Virology* 389, 59–65.
- Weber, P.H., Bujarski, J.J., 2015. Multiple functions of capsid proteins in (+) stranded RNA viruses during plant-virus interactions. *Virus Research* 196, 140–149.

Bymoviruses (*Potyviridae*)

Annette Niehl, Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

Frank Rabenstein, Julius Kühn Institute, Quedlinburg, Germany

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

bp Base pair

CI Cylindrical inclusion protein

CP Coat protein

dsRNA Double-stranded RNA

eIF4E Eucaryotic translation initiation factor 4E

ELISA Enzyme-linked immunosorbent assay

ER Endoplasmic reticulum

GFP Green fluorescent protein

kDa Kilodalton

MB Membranous body

Nla-Pro Nuclear inclusion protein a-proteinase

Nlb Nuclear inclusion protein b

nm Nanometer

nt Nucleotide(s)

ORF Open reading frame

PCR Polymerase chain reaction

RNA Ribonucleic acid

siRNAs Small-interfering RNAs

UTR Untranslated region

VPg Viral protein genome-linked

Glossary

Bymovirus From barley yellow mosaic virus.

Classification

The bymoviruses are single stranded positive-sense RNA viruses with a bipartite genome in the family *Potyviridae*.

Member Species

The *Bymovirus* genus currently comprises six member species, *Barley mild mosaic virus* (BaMMV), *Barley yellow mosaic virus* (BaYMV), *Oat mosaic virus* (OMV), *Rice necrosis mosaic virus* (RNMV), *Wheat spindle streak mosaic virus* (WSSMV), and *Wheat yellow mosaic virus* (WYMV). The type member of the genus is BaYMV. The currently unclassified *Soybean leaf rugose mosaic virus* seems to be a related virus species, since the C-terminal region of the Nlb gene exhibits about 60% amino acid sequence identity with barley mild mosaic virus and other bymoviruses.

Virion Structure, Serological- and Cytological Properties

Bymoviruses are flexuous filamentous particles of 250–300 nm and 500–600 nm in length and approximately 13 nm in width (Fig. 1(A)). The bipartite, positive-sense RNA genome is encapsidated by a single coat protein (CP). BaYMV is serologically related to WSSMV; WSSMV and WYMV appear to be serologically indistinguishable as neither polyclonal antisera or monoclonal antibodies are able to differentiate between them. OMV and BaMMV are serologically distinct from the other members of the genus. Bymoviruses induce pinwheel-like inclusions and endoplasmic reticulum (ER)-derived inclusions (membranous body, MB) with crystalline lattice-like or tubule dominated morphology (tube-MB) and lamellar structures (lamella-MB) in infected host cells (Fig. 1(B) and (C)). Electron tomography analysis and immuno-gold labeling revealed that both MB types contain viral proteins and components derived from ER membranes. The irregularly arranged lamellar-MB was composed of “rough” ER like structures, whereas regularly arranged tubule-MB was composed of “smooth” ER structures.

Nucleic Acid Properties and Differentiation of Bymoviruses

The bymovirus positive-sense genomic RNAs are approximately 7.5–8.0 kb (RNA1) and 3.5–4.0 kb (RNA2) in length. The RNAs contain the virus-encoded genome-linked viral protein VPg covalently bound to their 5' end and a poly-A tail. The WYMV poly-A tail in naturally infected wheat is variable in length and can even be absent. Recent experiments using *in-vitro* translation and replication showed that polyadenylation of the viral 3' ends impacted translation and negative strand RNA synthesis, indicating that polyadenylation may represent a means for the virus to fine tune translation or replication or may even represent a switch

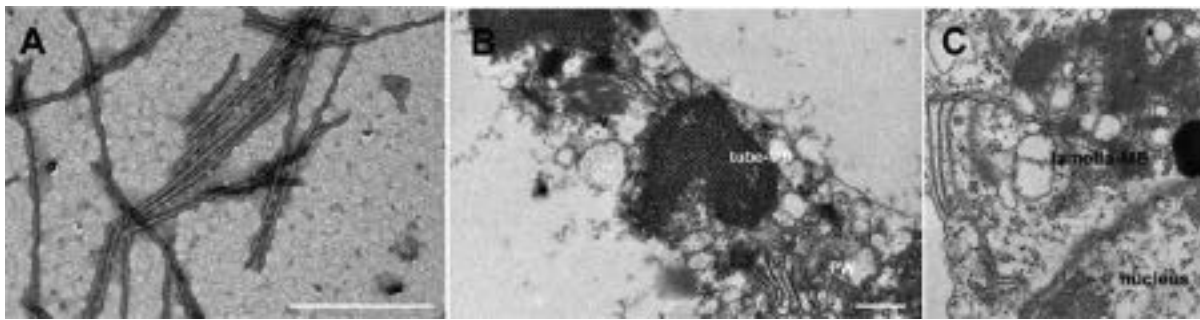


Fig. 1 Electron microscopy of bymoviruses and infection-derived cellular structures. (A) Transmission electron microscopy of a virus purification from BaMMV infected plants reveals flexuous particles of approximately 250–300 and 500–600 nm in length. (B) and (C) Ultra thin sections of BaYMV (A) and BaMMV (B) infected cells showing ER-derived membranous bodies (MB) of tubular (B) or lamellar morphology (C). V virus particles, PW pinwheel inclusion bodies, scale bars 500 nm.

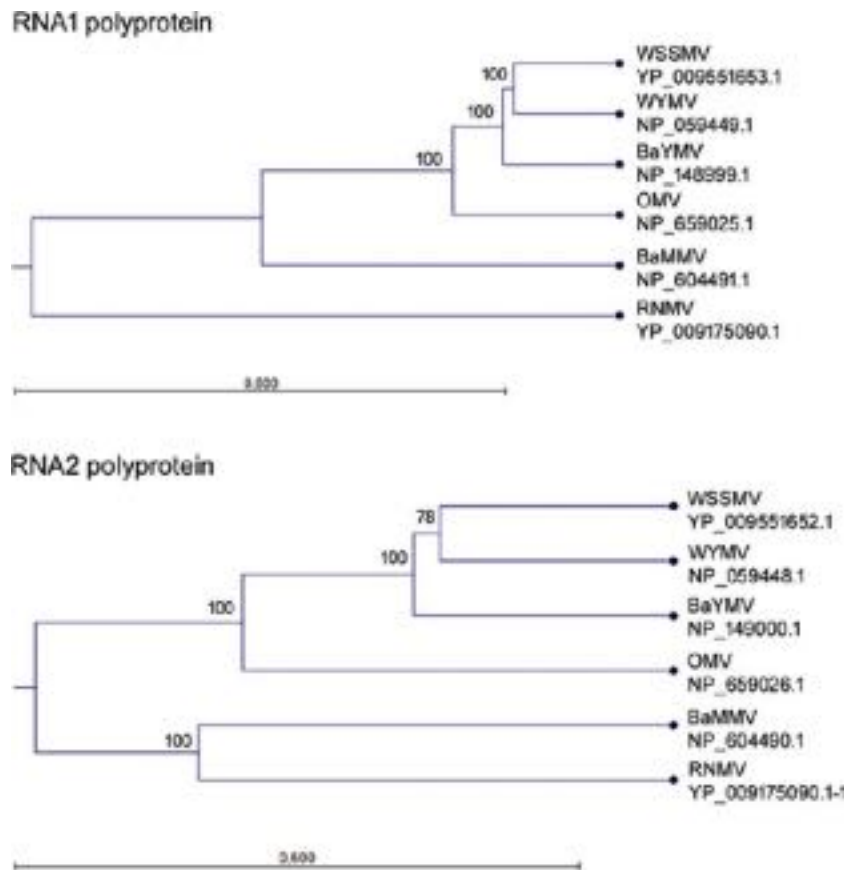


Fig. 2 Phylograms depicting relationships between the different bymovirus member species. Trees were made with alignments of polyprotein amino acid sequences in CLC main workbench 8.1.3 using the UPGMA algorithm, Jukes-Cantor distance measure and 100 bootstraps, indicated as numbers next to the branches.

between translation and replication of viral proteins. Interestingly, positive effects of cap-independent translation by the 5'UTRs (untranslated region) of WYMV RNA1 and RNA2 was counteracted by the presence of a poly-A tail at the 3'UTR of RNA1 and further increased by the presence of a poly-A tail at the 3'UTR for RNA2. Given that bymovirus resistance is at least in part based on VPg-eIF4 interaction (see below, paragraph on virus control), efficient cap-independent translation by bymoviruses due to a variable polyadenylation status may provide a possibility for the virus for host adaptation.

Sequence similarity comparisons of polyprotein sequences from bymovirus reference numbers published in the 10th ICTV report on the family *Potyviridae* (last revised in October 2018, see "Relevant Websites section") and the genebank reference sequence assembly data for WSSMV (RefSeq assembly: GCF_004117675.1) revealed relationships between polyproteins derived from RNA1 and RNA2 (Fig. 2). For RNA1 polyproteins, WYMV, WSSMV, BaYMV and OMV are more closely related to each other than to RNMV and BaMMV. For RNA2 polyproteins, RNMV and BaMMV form a subgroup, while WYMV, WSSMV, BaYMV and OMV form a different subgroup.

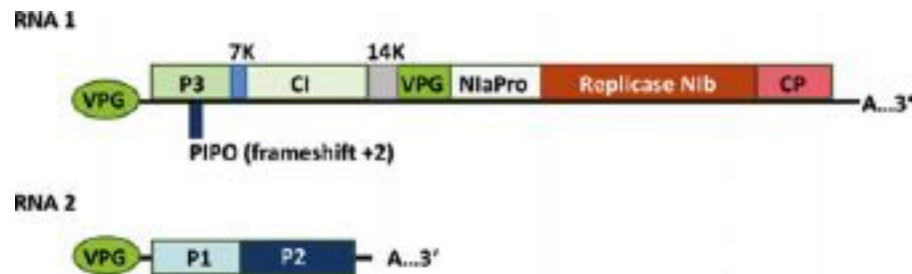


Fig. 3 Genome organization of bymoviruses. Black lines represent viral RNAs with VPg protein bound to the 5' end and variable poly A tail at the 3' end. Square boxes along the RNA represent encoded proteins and their relative localization.

Organization of the Genome and Properties of the Encoded Proteins

Bymovirus RNAs encode for polyproteins, which are proteolytically cleaved into functional proteins. RNA1 contains the typical potyvirus organization of genes except for P1 and HcPro proteins (Fig. 3). Bymovirus RNA2 encodes a P1 protein with amino acid sequence similarity to HcPro, which is restricted to the C-terminal cysteine protease domain of the protein, and a bymovirus-specific P2 protein. As the genome organization for RNA1 is shared among members of the *Potyviridae* family, it is likely that the proteins also have common functions. The large polyprotein derived from RNA1 is proteolytically cleaved into eight mature proteins, P3, 7K, CI (cylindrical inclusion) protein, 14K, VPg (viral protein genome-linked), nuclear inclusion protein a-proteinase (NIa-Pro), nuclear inclusion protein b (Nib), and CP (coat protein). The NIa-Pro is responsible for cleavage. Potyviral CI is thought to be involved in virus movement, VPg is a multifunctional protein involved in translation, replication and possibly movement, Nib is involved in replication, and CP is important for viral encapsidation. Apart from inferring viral protein functions from studies of other potyviruses, the functions of some of the proteins encoded by bymovirus RNAs have been studied in detail for WYMV. The small PIPO open reading frame (ORF) overlapping with the P3 ORF has been predicted in the WYMV genome. Bymovirus RNA2 encodes a polyprotein that is cleaved into the P1 and P2 proteins. The P1 protease cleaves at a site with the consensus sequence (Y,F,G)xG↓(A,N,S). P2 is believed to be important for virus transmission by its vector, *Polymyxa graminis*. VPg proteins are covalently attached to viral genomic RNA and involved in efficient translation of viral RNA. The WYMV VPg protein interacts with the other viral proteins P1 and CP, contains nuclear export and import signals and has been shown to localize to nucleus and cytoplasm during infection, indicating functions of the protein in both cellular compartments. Translocation to the cytosol depends on interaction of VPg with CP, however, the mechanism, by which CP may increase cytosolic accumulation of VPg remains unresolved. The viral VPg interacts with translation initiation factors of the eIF4E family to initiate translation and replication of viral proteins. Recently, WYMV CP has been shown to interact with TaHSP23.6, a wheat small heat shock protein and expression of a range of heat shock proteins was induced upon WYMV infection.

Bymovirus infection causes strong modifications of the ER membrane (membranous body (MB)). The inclusions derived from ER membranes produced during WYMV infection have been shown to contain multiple viral proteins. The integral membrane protein P2 interacts with itself, P1, P3, VPg, NIa-Pro, and CP, rearranges the ER and forms large aggregate structures, thus likely induces the formation of the inclusions by interacting with the ER and recruiting other viral proteins including P1, P3, NIa-Pro, and VPg, but not 7K, CI and CP into these inclusions. Moreover, P2 has been described to recruit host proteins into P2-derived inclusions. P2 interacts with the small GTPase Sar1 important for COPII-vesicle mediated ER-to-Golgi transport in P2-derived inclusions. The formation of the inclusions appears to depend on a functional secretory pathway. As the inclusions are associated with RNA synthesis, they likely represent sites of viral replication. Moreover, P2 appears to be important for transmission of the virus by its vector, *P. graminis*. Upon repeated mechanical inoculation, deletions within the 3' region of the P2 cistron occur for BaYMV, BaMMV and OMV. Two transmembrane domains have been predicted in bymovirus P2, as well as in *Furovirus*, *Pomovirus* and *Benyvirus* CP-Read-through protein (CP-RT). The second of these domains appears to be crucial for vector transmission and is lost in non-transmissible deletion mutants. For C-terminally truncated P2 forms replication still occurs in host cells, indicating that the C-terminal region of P2 is dispensable for virus replication. For BaYMV, the entire P2 protein appears to be dispensable for virus replication, but appears to be important for efficient systemic virus accumulation and symptom development. The peripherally ER membrane-associated WYMV P1 protein interacts with itself and with the other WYMV-encoded proteins, except 14K, and the WYMV integral membrane protein P3 localizes to the ER and Golgi and interacts with itself, P1, P2 and CP. P1 contains cysteine protease activity and believed to cleave the RNA2-derived polyprotein into the mature P1 and P2 proteins. In protoplasts, BaYMV P1 expression increases accumulation of CP, suggesting that P1 enhances RNA1 replication. BaYMV P1 also appears to be essential for systemic infection. Consistent with these functions and in analogy to potyviral P1/HcPro, bymoviral P1 may be involved in the suppression of RNA silencing; however, up to now, there is no experimental evidence for a role of bymoviral P1 as viral silencing suppressor. Nevertheless, that bymovirus double-stranded (ds)RNA replication intermediates are targeted by RNA silencing has been shown. RNA silencing differs between leaves and roots of WYMV infected plants in the way that more small interfering RNAs (siRNAs) – but also more virus- accumulate in roots versus leaves of the plants and that siRNAs in leaf but not root tissues demonstrate a bias towards 5' A or U, and may hence be incorporated into different Argonaute complexes. Moreover, Dicer-like gene expression seems to be differently regulated upon WYMV infection in leaves and in roots.

WYMV 7K was shown to localize to ER membranes when fused to GFP. The WYMV CI protein has been shown in reassortant experiments to determine the pathogenicity of virus infection. Three amino acids in the N-terminal region of CI involved in the change of pathogenicity were identified. As replication in protoplasts remained unchanged between different WYMV pathotypes, it was assumed that the reason for different pathogenicity between WYMV pathotypes is related to CI function in virus movement. Nucleotide sequences of several German WSSMV isolates used for virus resistance trials differ in their 14K protein region from the sequences of isolates found in France, Denmark, and the USA by a nona-nucleotide insertion. At present, it is not clear whether the amino acid changes in these virus isolates lead to biological effects and differences in resistance behavior.

Virus Transmission and Movement

The bymoviruses are transmitted by the soil-borne plasmodiophorid *P. graminis*. Despite challenging for some species of the genus, bymoviruses can also be transmitted mechanically. Transmission takes place after the virus-carrying zoospores of the vector penetrate host root epidermal or root hair cells; however, the detailed mechanism by which the virus crosses the zoosporangial plasmalemma to enter the host cell cytoplasm has not been illuminated. No evidence exists to demonstrate that the virus is able to replicate in its vector. WSSMV CP was detected inside *P. graminis* resting spores. Virus inside resting spores remains infectious over many years. Once transmitted to the host plant, the virus moves systemically. Whether the virus moves cell to cell and systemically as virion or as viral ribonucleoprotein complex in its host plants has not been determined.

Symptoms, Epidemiology and Host Range

Typical symptoms of bymovirus infection constitute of chlorotic mosaic of dark green areas on the younger leaves and chlorotic streaks on leaves (Fig. 4), growth depression, reduced vigor, fewer tillers and reduced grain yield. In the field, winter crops usually develop symptoms in late winter and early spring. Symptomatic plants are usually found in patches in the fields. Barley yellow mosaic disease caused by BaMMV and BaYMV was first reported in Japan in 1940. Since, it has also been detected in China, South Korea, and Europe including Belgium, Bulgaria, France, Germany, Greece, Hungary, Italy, the Netherlands, Poland, Russia, Spain, the UK, and Ukraine. In Japan, at least seven strains of BaYMV and two BaMMV strains are known, while in Europe currently two BaYMV and three BaMMV strains have been described. In China, at least six strains of BaYMV are probably present. Crop loss due to the barley-infecting bymoviruses can reach up to 50% in infected areas. Bymovirus-caused wheat yellow mosaic disease can be due to WSSMV and WYMV infection. WSSMV was first identified 1960 in Canada and ten years later in the USA, where it quickly spread throughout the country. In Europe, WSSMV is present in several countries, among them France, Italy, Germany, and Belgium. WSSMV has also been reported to occur in Africa. Wheat losses due to WSSMV have been reported to range between 20 and 25%. WYMV was first reported in Japan and also occurs in India and China. RNMV was first detected in Japan, and then in India in the late 1970th and causes yellow flecks and streaks on rice leaves. OMV was first reported in 1944 in the USA, where it is widely distributed. It is also present in Europe (England, France, Ireland, and Italy). Yield losses caused by oat mosaic disease of more than 50%, even to complete losses, have been reported. With respect to yield losses, it should be noted here that bymovirus infection often occurs as co-infection with furoviruses and the contributions of each virus species to the disease phenotype and associated crop losses have not been dissected. Thus, it remains unclear whether other virus factors contribute to the losses. Apart from other biotic parameters, also abiotic parameters such as temperature, humidity, soil conditions etc. contribute to the yield loss.

Under natural conditions, BaYMV and BaMMV infect only barley, WSSMV and WYMV wheat, OMV oat and RNMV rice. Interestingly, for RNMV mechanically inoculated onto Ludwigia, Corchorus, and Hibiscus species, growth promotion phenotypes have been described.



Fig. 4 Disease symptoms caused by bymoviruses. (A) Barley field showing patches of plants with barley yellow mosaic disease symptoms. BaYMV (B) and BaMMV (C) leaf mosaic symptoms on winter barley leaves.

Diagnosis

Depending on the cultivated variety and the environmental conditions, the severity of disease symptoms caused by bymovirus infection can vary greatly and normally, virus symptoms disappear with increasing temperatures. The frequent occurrence of mixed infections with other viruses makes a precise diagnosis based on the symptoms very difficult or impossible. Thus, immunoassays such as double-antibody sandwich (DAS)- and triple antibody sandwich (TAS)-ELISA are still most commonly used for practical diagnosis. Commercial test kits are available for all bymoviruses except for RNMV. The detection of virus by electron microscopy can be made specific through the use of immuno-gold labeled antibodies. For the detection and differentiation of the serologically closely related WSSMV and WYMV, a multiplex RT-PCR method was developed. Quantification and improvement of detection sensitivity for WYMV could be achieved by real-time quantitative RT-PCR. In addition, a sensitive reverse transcription loop-mediated isothermal amplification (RT-LAMP) method was developed for detection of WYMV and BaYMV. Reliable detection of BaYMV and BaMMV by real-time PCR has also been reported. Furthermore, a RT-PCR protocol to specifically detect OMV has been developed.

The application of High-Throughput Sequencing (HTS), also known as next-generation sequencing (NGS), has proven very successful also for virus detection of bymoviruses in field samples. In barley, the *rym4* gene confers resistance to BaMMV and the common BaYMV pathotype, BaYMV-1. However, a new virus pathotype designated BaYMV-2, which overcomes the *rym4*-controlled resistance, appeared in several European countries. In order to discriminate between BaYMV-1 and BaYMV-2 by molecular methods a dCAPS (derived Cleaved Amplified Polymorphic Sequences) tool was developed and successfully applied to track the spread of pathotype BaYM-2 in France.

It has been recently shown that deep small (s)RNA sequencing is a universal approach for virome reconstruction and RNAi characterization. This approach was used on dried barley leaves from field surveys. Illumina sequencing of sRNAs from plant samples identified besides BaYMV an endornavirus and, in one plant, a novel strain of Japanese soil-borne wheat mosaic virus (JSBWMV).

Control

Apart from preventive measures by decontamination of machinery and man, the planting of resistant crop varieties presently represents the only efficient means of control for bymovirus infection. Cultural practices remain ineffective due to the robustness of resting spores and soil treatments are ecologically unacceptable. While cereal varieties with resistance to *P. graminis* have not been identified, several crop varieties with resistance to one or more bymoviruses are available.

Against barley yellow mosaic disease caused by BaMMV and BaYMV, 19 *rym* resistance genes are known that are attributed to at least nine loci in the barley genome. 15 of these resistance genes are inherited in a recessive manner and are effective against one or more pathotypes or strains of the same virus or against more than one virus. Some of the resistance genes have been mapped and identified. One encodes for the eukaryotic translation initiation factor 4E (eIF4E). *rym4*, *rym5*, *rym6*, *rym_{Hor4224}*, and *rym_{Hor3298}* resistance based on eIF4E are allelic. Alleles confer resistance against different strains of BaYMV and/or BaMMV. The viral VPg interacts with eIF4 and is presumed to be important for virus translation, replication and movement. The second resistance gene encodes a protein disulfide isomerase like 5-1 (PDIL5-1). To our knowledge, nothing is known about the mechanisms underlying protein disulfide isomerase-based resistance to BaMMV and BaYMV. The monogenic resistances are relatively frequently broken and lead to the emergence of new virus pathotypes. Thus, apart from the identification of novel resistance genes and introduction of these genes into breeding programs, the currently most promising way to achieve durable resistance is by pyramiding resistance genes. In the case of WYMV, there is evidence that resistance is, similar to the barley infecting bymoviruses, eIF4E-mediated. Replication studies of WYMV in barley and wheat protoplasts indicated that expression of wheat eIF4E enabled the replication of WYMV in barley, which is otherwise restricted to wheat. Moreover, approaches in transgenic wheat overexpressing the viral replicase gene in antisense orientation resulted in durable resistance against WYMV. Against WSSMV, the resistant trait in commercial winter wheat varieties encodes a single dominant resistance gene on the long arm of chromosome 2D or is controlled in an additive dominant manner by two loci. WSSMV resistance probably limits virus spread from the roots to the shoots of the plants, as leaves of resistant varieties are readily mechanically infected and lead to systemic distribution of virus. OMV tolerance, in which virus accumulates to high titers but symptoms are mild or absent, has been identified in different *Avena* species and is inherited as a quantitative polygenic trait. Most of the Japanese paddy rice varieties can become infected when grown in soil infested with RNMV. To our knowledge, no resistance genes for RNMV have been identified so far. However, a study using transgenic plants overexpressing a HAP (heme activator protein) transcription factor revealed that the plants displayed a resistance phenotype towards RNMV infection by accumulating lower virus titer and the absence of visible disease symptoms. Moreover, infection induced expression of the transcription factor.

Further Reading

- Adams, M.J., 2002. Bymovirus. In: Tidona, C.A., Darai, G., Büchen-Osmond, C. (Eds.), *The Springer Index of Viruses*. Berlin, Heidelberg: Springer.
- Geng, G., Yu, C., Li, X., Yuan, X., 2019. Variable 3' polyadenylation of *Wheat yellow mosaic virus* and its novel effects on translation and replication. *Virology Journal* 16, 23.
- Ghosh, S.K., 1982. Growth promotion in plants by rice necrosis mosaic virus. *Planta* 155, 193–198.
- Hooper, G.R., Wiese, M.V., 1972. Cytoplasmic inclusions in wheat affected by wheat spindle streak mosaic. *Virology* 47, 664–672.

- Kühne, T., 2009. Soil-borne viruses affecting cereals – Known for long but still a threat. *Virus Research* 141, 174–183.
- Lapierre, H.D., Hariri, D., 2008. Cereal viruses: Wheat and barley. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Amsterdam: Elsevier, pp. 490–497.
- Li, L., Andika, I.B., Xu, Y., *et al.*, 2017. Differential characteristics of viral siRNAs between leaves and roots of wheat plants naturally infected with wheat yellow mosaic virus, a soil-borne virus. *Frontiers in Microbiology* 8, 1802. 1802.
- Li, H., Kondo, H., Kühne, T., Shirako, Y., 2016. Barley yellow mosaic virus VPg is the determinant protein for breaking eIF4E-mediated recessive resistance in barley plants. *Frontiers in Plant Science* 7, 1449. 1449.
- Ordon, F., Perovic, D., 2013. Virus resistance in barley. In: Varshney, R.K., Tuberosa, R. (Eds.), *Translational Genomics for Crop Breeding*. Wiley. doi:10.1002/9781118728475.ch5.
- Rolland, M., Villemot, J., Marais, A., *et al.*, 2017. Classical and next generation sequencing approaches unravel *Bymovirus* diversity in barley crops in France. *PLoS One* 12, e0188495.
- Stein, N., Perovic, D., Kumlehn, J., *et al.*, 2005. The eukaryotic translation initiation factor 4E confers multiallelic recessive Bymovirus resistance in *Hordeum vulgare* (L.). *The Plant Journal* 42, 912–922.
- Sun, L., Andika, I.B., Shen, J., Yang, D., Chen, J., 2014. The P2 of *Wheat yellow mosaic virus* rearranges the endoplasmic reticulum and recruits other viral proteins into replication-associated inclusion bodies. *Molecular Plant Pathology* 15, 466–478.
- Sun, L., Jing, B., Andika, I.B., *et al.*, 2013. Nucleo-cytoplasmic shuttling of VPg encoded by *Wheat yellow mosaic virus* requires association with the coat protein. *Journal of General Virology* 94, 2790–2802.
- Timpe, U., Kühne, T., 1995. In vitro transcripts of a full-length cDNA of a naturally deleted RNA2 of barley mild mosaic virus (BaMMV) replicate in BaMMV-infected plants. *Journal of General Virology* 76, 2619–2623.
- You, Y., Shirako, Y., 2010. Bymovirus reverse genetics: Requirements for RNA2-encoded proteins in systemic infection. *Molecular Plant Pathology* 11, 383–394.
- You, Y., Shirako, Y., 2013. Evaluation of host resistance to Barley yellow mosaic virus infection at the cellular and whole-plant levels. *Plant Pathology* 62, 226–232.

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/potyviridae/568/genus-bymovirus
 Bymovirus.
- <http://www.dpvweb.net/dpv/showdpv.php?dpvno=374>
 Barley yellow mosaic virus.

Cacao Swollen Shoot Virus (*Caulimoviridae*)

Emmanuelle Muller, The French Agricultural Research Center for International Development, Joint Research Units–Biology and Genetics of Plant-Pathogen Interactions, Montpellier, France and Biology and Genetics of Plant-Pathogen Interactions, University of Montpellier, The French Agricultural Research Center for International Development, French National Institute for Agricultural Research, Montpellier SupAgro, Montpellier, France

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Badnavirid Virus belonging to the genus *Badnavirus*, family *Caulimoviridae*.

Badnavirus Virus genus of plant double strand DNA virus encapsidated in bacilliform particles.

Pararetrovirus Reverse transcribing viruses that replicate through an RNA intermediate.

Retrotransposons Are genetic elements that can amplify themselves in a genome and are ubiquitous components of the DNA of many eukaryotic organisms.

RTase/RNaseH Enzymes responsible for the replication of the pararetrovirus group.

Introduction

Cacao swollen shoot disease (CSSD) is caused by Cacao swollen shoot virus (CSSV), member of the genus *Badnavirus*, family *Caulimoviridae*. CSSD can be regarded as the major viral disease on cacao and has been recognized as one of the most important diseases in West Africa limiting cacao production. The disease appeared only after the introduction of cacao in West Africa and is probably the result of a host jump from native reservoir plants. It was highlighted for the first time in Ghana in 1922 but was described and named in 1936, then in Nigeria in 1944, in Ivory Coast in 1946, in Togo in 1949, and in Sierra Leone in 1963. CSSV is naturally transmitted to cacao (*Theobroma cacao*) in a semi-persistent manner by several mealybug species. Extensive research on this emerging disease has been conducted since the 1940s in West African Cacao Research Institute (WACRI- now CRIG for Cacao Research Institute of Ghana) and then in other institutes, in virology, epidemiology, and genetic improvement of cacao tolerance to CSSD. Nevertheless yet, the disease has not been controlled and continues to spread into new areas.

Other viral diseases have been also described on cacao in Trinidad and Tobago and in Sri Lanka but associated with slightly different symptomatologies and have not caused epidemics similar to cacao swollen shoot disease. Complete viral genomes have been obtained recently from those diseases and correspond to clearly different badnavirid species: *Cacao yellow vein-banding virus* (CYVBV) and *Cacao mild mosaic virus* (CMMV) infecting cacao in Trinidad, *Cacao bacilliform Sri-Lanka virus* (CBSLV) infecting cacao in Sri Lanka.

Taxonomy

The genus *Badnavirus*, one of the eight genera of the family *Caulimoviridae*, are double stranded DNA (dsDNA) viruses with bacilliform particles that emerge mainly on tropical plants. The most commonly encountered on crops in addition to CSSV on cocoa are Banana streak viruses -BSV- on banana, Dioscorea bacilliform viruses -DBV- on yam and *Citrus yellow mosaic virus* -CIYMV- on citrus. The viruses member of this genus have an extremely wide host range: some viruses are found in many ornamental plants such as *Yucca*, *Dracaena*, *Kalanchoe*, *Hibiscus*, *Pelargonium*, *Bougainvillea*, *Cycas*, *Canna*, *Jasminum*, *Wisteria* and *Commelina* as well as in a significant number of cultivated plants, such as pineapple, fig, sweet potato, grapevine, black pepper, jujube, currant, raspberry, ginger, mulberry, and blackberry-ripe in addition to the major crops mentioned above (Banana, cocoa, and citrus). New badnavirids are regularly identified on new plant species such as *Sophora* from Japan, Yacon, birch, or Aralia Ming (*Polyscias fruticosa* L.) are found hosting badnaviruses. A total of 57 species are currently recognized by ICTV (Table 1) and many more are unassigned species in the genus (> 50). The description of some of these new viruses is certainly linked to the improvement of virus detection and testing techniques, rather than viruses a particular health problem for plant. It should be noted that the species demarcation of 80% identity of sequences has been set more than 15 years ago, and now that so many badnavirid sequences are available, this demarcation threshold should be revisited. It is remarkable that BSV, DBV, CSSV and even RTBV (*Tungrovirus*, *Caulimoviridae*) seem to have many species for a single disease in each case, or would it be possible that in the case of badnaviruses, or even caulimoviruses, that the species threshold might be much lower than 80%?

Host Range and Symptomatology

Experimental host range is limited to species of the large botanical family of Malvaceae but the principal host of the virus is *Theobroma cacao*. Symptoms on cacao trees are mostly seen in leaves, but stem and root swellings, as well as pod deformation also occur. In some

Table 1 List of members of the genus *Badnavirus* (*Caulimoviridae*). The list of assigned and unassigned species in the genus is indicated. The accession numbers for the reference complete genome sequences are indicated in the last column

Species in the genus <i>Badnavirus</i>	Abbreviation-Isolate	Accession #
<i>Aglaonema bacilliform virus</i>	ABV	MH384837
<i>Banana streak GF virus</i>	BSGFV	AY493509
<i>Banana streak IM virus</i>	BSImV	HQ659760
<i>Banana streak MY virus</i>	BSMyV	AY805074
<i>Banana streak OL virus</i>	BSOLV	AJ002234
<i>Banana streak UA virus</i>	BSUAV	HQ593107
<i>Banana streak UL virus</i>	BSULV	HQ593109
<i>Banana streak UI virus</i>	BSUIV	HQ593108
<i>Banana streak UM virus</i>	BSUMV	HQ593110
<i>Banana streak VN virus</i>	BSVNV	AY750155
<i>Birch leaf roll associated virus</i>	BLRaV	MG686419
<i>Blackberry virus F</i>	BVF	KJ413252
<i>Bougainvillea spectabilis chlorotic vein-banding virus</i>	BCVBV	EU034539
<i>Cacao bacilliform Sri Lanka virus</i>	CBSLV	MF642736
<i>Cacao mild mosaic virus</i>	CMMV	KX276641
<i>Cacao swollen shoot CD virus</i>	CSSCDV	JN606110
<i>Cacao swollen shoot CE virus</i>	CSSCEV	MF642719
<i>Cacao swollen shoot Ghana M virus</i>	CSSGMV	MF642724
<i>Cacao swollen shoot Ghana N virus</i>	CSSGNV	MF642725
<i>Cacao swollen shoot Ghana Q virus</i>	CSSGQV	MF642726
<i>Cacao swollen shoot Togo A virus</i>	CSSTAV	AJ781003
<i>Cacao swollen shoot Togo B virus</i>	CSSTBV	L14546
<i>Cacao yellow vein-banding virus</i>	CYVBV	KX276640
<i>Canna yellow mottle associated virus</i>	CaYMaV	KX0660020
<i>Canna yellow mottle virus-AP01</i>	CaYMV	MF074075
<i>Citrus yellow mosaic virus</i>	CiYMV	AF347695
<i>Commelina yellow mottle virus</i>	ComYMV	X52938
<i>Dioscorea bacilliform AL virus</i>	DBALV	X94576
<i>Dioscorea bacilliform AL virus 2</i>	DBALV2	KY827395
<i>Dioscorea bacilliform ES virus</i>	DBESV	KY827394
<i>Dioscorea bacilliform RT virus 1</i>	DBRTV1	KX008574
<i>Dioscorea bacilliform RT virus 2</i>	DBRTV2	KX008577
<i>Dioscorea bacilliform SN virus</i>	DBSNV	DQ822073
<i>Dioscorea bacilliform TR virus</i>	DBTRV	KX430257
<i>Fig badnavirus 1</i>	FBV 1	JF411989
<i>Gooseberry vein banding virus BC isolate</i>	GVBV-BC	HQ852250
<i>Grapevine roditis leaf discoloration associated virus</i>	GRLDaV	KT965859
<i>Grapevine vein clearing virus</i>	GVCV	JF301669
<i>Jujube Mosaic associated virus</i>	JuMaV	KX852476
<i>Kalanchoe top-spotting virus</i>	KTSV	AY180137
<i>Mulberry badnavirus 1</i>	MBV1	LN651258
<i>Pagoda yellow mosaic associated virus</i>	PYMaV	KJ013302
<i>Pineapple bacilliform CO virus</i>	PBCoV	GU121676
<i>Pineapple bacilliform ER virus</i>	PBErV	
<i>Piper yellow mottle virus</i>	PYMoV	KC808712
<i>Rubus yellow net virus</i>	RYNV	KF241951
<i>Schefflera ringspot virus</i>	SRV	
<i>Spiraea yellow leafspot virus</i>	SYLSV	
<i>Sugarcane bacilliform Guadeloupe A virus</i>	SCBGAV	FJ824813
<i>Sugarcane bacilliform Guadeloupe D virus</i>	SCBGDV	FJ439817
<i>Sugarcane bacilliform IM virus</i>	SCBIMV	AJ277091
<i>Sugarcane bacilliform MO virus</i>	SCBMOV	M89923
<i>Sweet potato pakakuy B virus</i>	SPPV	FJ560943
<i>Taro bacilliform CH virus</i>	TaBCHV1	KP710177
<i>Taro bacilliform virus</i>	TaBV	AF357836
<i>Wisteria badnavirus 1</i>	WBV1	KX168422
<i>Yacon necrotic mottle virus</i>	YaNMoV	KM229702

(Continued)

Table 1 Continued

Species in the genus <i>Badnavirus</i>	Abbreviation-Isolate	Accession #
<i>Unassigned species</i>		
Banana streak CA virus	BSCAV	HQ593111
Banana streak UJ virus	BSUJV	
Banana streak UK virus	BSUKV	
Banana streak PE virus	BSPEV	MN187554
Cacao swollen shoot Ghana T virus	CSSGTV	MN179342
Cycad leaf necrosis virus	CLNV	EU853709
Dracaena mottle virus 1	DrMV	DQ473478
Hibiscus bacilliform virus	HiBV	KF875586
Pelargonium vein banding virus	PVBV	GQ428155

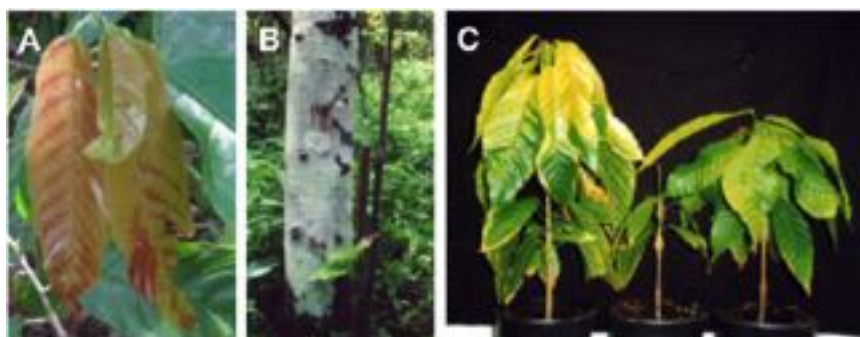


Fig. 1 (A) Symptoms of red-vein banding observed on young flush leaves of cocoa tree in Togo, Klotó area. (B) Symptoms of swellings on chupons of cocoa tree, in Togo, Litimé area. (C) Symptoms of swollen shoot observed on *Theobroma cacao* plantlets four months after agro-inoculation with the Togolese Agou 1 isolate. From left to right: a plantlet inoculated with the wild strain *Agrobacterium tumefaciens* LBA4404 and two plantlets inoculated with the recombinant *A. tumefaciens* bacteria LBA4404 (pAL4404, pBCPX2) containing CSSV insert.

varieties of cacao, particularly Amelonado cacao, reddening of primary veins and veinlets in flush leaves is characteristic (Fig. 1(A)). This red vein banding later disappears. There can be various symptoms on mature leaves, depending on the cacao variety and virus strain. These symptoms can include: yellow clearing along main veins; tiny pin-point flecks to larger spots; diffused flecking; blotches or streaks. Chlorotic vein flecking or banding is common and may extend along larger veins to give angular flecks.

Stem swellings may develop at the nodes, internodes, or shoot tips. These may be on the chupons, fans, or branches (Fig. 1(B)). Many strains of CSSV also induce root swellings. Stem swellings result from the abnormal proliferation of xylem, phloem, and cortical cells. Infected trees may suffer from partial defoliation initially due to the incompletely systemic nature of the infection. Ultimately, in highly susceptible varieties, severe defoliation and dieback occurs. Smaller, rounded to almost spherical pods may be found on trees infected with severe strains. Occasional green mottling of these pods is seen and their surface may be smoother than the surface of healthy pods.

Various isolates were described in Ghana as in Togo with a variability and a gradation in the type of symptoms observed. Severe strains of this virus can kill susceptible cacao trees within 2–3 years. They affect Amelonado cacao, widely considered to give the best-quality cacao, more seriously than Upper Amazon cacao and hybrids which have been selected for resistance to the virus. A few avirulent strains occur in limited, widely scattered outbreaks, usually inducing stem swellings only, and having little effect, if any, on growth or yield. As for other perennial crops, we observe a latency period between the time of infection and the time of symptom expression, which complicates control measures. Moreover, there are periods of remission during which symptoms are not visible.

The natural host range of the virus includes *Ceiba pentandra*, *Cola chlamydanthra*, *Cola gigantea* var. *glabrescens*, *Sterculia tragacantha*, and *Adansonia digitata* with associated symptoms of transient leaf chlorosis or conspicuous leaf chlorosis. Viral sequences belonging to different species have been detected in *C. pentandra* as well as in *Commelina* sp.

Transmission

Fourteen species of mealybugs (*Pseudococcidae* spp.), including *Planococcoides njalensis*, *Planococcus citri*, *Planococcus kenyae*, *Phenacoccus hargreavesi*, *Planococcus* sp. *Celtis*, *Pseudococcus concavocerrari*, *Ferrisia virgata*, *Pseudococcus longispinus*, *Delococcus tafoensis*, and *Paraputo anomalus*, have been reported to transmit CSSV. Only the nymphs of the first, second, and third larval stages and the adult females are able to transmit the virus. The virus does not multiply in the vector and is not transmitted to its progeny. CSSV is transmitted neither by seed nor by pollen. CSSV can infect cacao at any stage of plant growth. The virus is transmitted experimentally to susceptible species by grafting, particle bombardment, by agro-inoculation using transformed

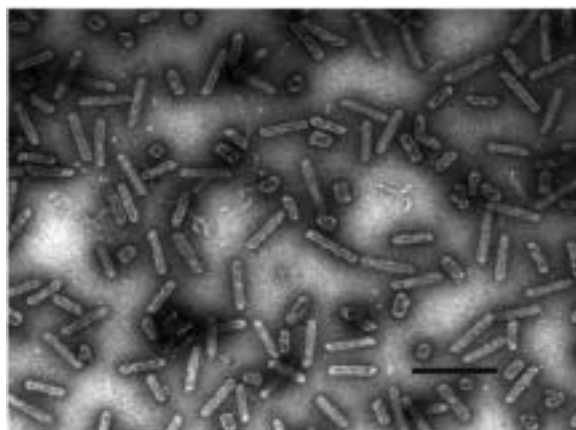


Fig. 2 Electron microscopic picture of virus particles of Cacao swollen shoot virus (CSSV), genus *Badnavirus*, family *Caulimoviridae*, 28×130 nm. The bar represents 200 nm.

Agrobacterium tumefaciens and with difficulty by mechanical inoculation. Seedlings usually produce acute red vein banding within 20–30 days and, 8–16 weeks later, swellings on shoots and tap roots (**Fig. 1(C)**).

The indigenous species are more difficult to infect than cacao, and mealybugs do not become infective as readily when feeding on them as when feeding on infected cacao.

Virion Structure

It was not until 1940 that through experimentation, the causative organism of the CSSD was identified to be a virus named CSSV. CSSV possesses small non-enveloped bacilliform particles and a double-stranded DNA genome (**Fig. 2**). The bacilliform particles measure $130 \text{ nm} \times 28 \text{ nm}$ in size and have been shown by dot-blot hybridization to occur in the cytoplasm of phloem companion and xylem parenchyma cells.

Genome Organization

Molecular characterization of CSSV had a boost in 1990, thanks to the improvement of the techniques of purification. The first complete genome of a CSSV isolate (Agou1 from Togo) was determined in 1993. Five putative open reading frames (ORFs) are located for this isolate on the plus strand of the 7.16 kbp CSSV genome (**Fig. 3**). ORF1 encodes a 16.7 kDa protein whose function is not yet determined. The ORF2 product is a 14.4 kDa nucleic acid-binding protein. ORF3 codes for a polyprotein of 211 kDa which contains, from its amino- to carboxyl-terminus, consensus sequences for a cell-to-cell movement protein, an RNA binding domain of the coat protein, an aspartyl proteinase, a reverse transcriptase (RTase), and an RNase H. The last two ORFs X (13 kDa) and Y (14 kDa) overlap ORF3 and encode proteins of unknown functions. More than 30 complete viral genomes have since been obtained by different technologies such as Sanger sequencing of PCR amplified full genome and Illumina sequencing. Their size varies from 6.8 kbp to 7.4 kbp. ORFY has been identified in all other complete genomes sequenced. ORFs 1, 2 and 3 are characteristic of all badnavirids unlike additional ORFs which are present only for some badnaviruses (CSSV, Citrus yellow mosaic virus –CiYMV-, Taro bacilliform virus –TaBV).

Phylogeny

Badnaviruses are highly variable at both the genomic and serological level, a feature which complicates the development of both molecular and antibody-based diagnostic tests. Moreover, CSSV isolates were for a long time classified according to the variability of the symptoms expressed on *T. cacao*, but it is not known if there is a correspondence between this variability of symptoms and the intrinsic molecular variability of the viruses.

A first study analyzed the molecular variability of the area of the ORF3 coding for N-terminal coat protein (CP) for 1A-like isolates from Ghana. Successive studies have analyzed molecular variability of CSSV isolates from Togo, Ghana and Côte d'Ivoire in the first part of ORF3 coding for the cell to cell movement protein. These analyses showed a high variability of CSSV isolates and permitted to define several groups (A to P) of sequences (**Fig. 4**). Groups Q, R and T were never detected with these primers. Recently the analysis of all complete viral genomes of isolates from those countries made it possible to describe more exhaustively the variability of this virus and to show that the disease is caused by a complex of many viral

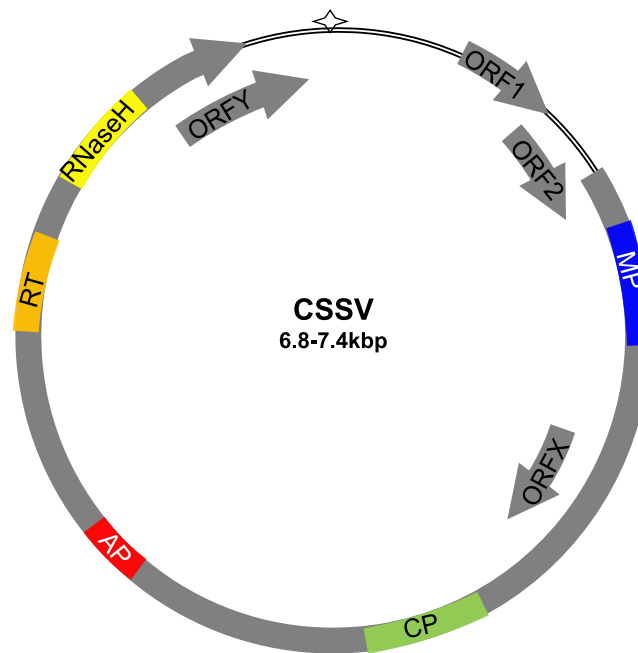


Fig. 3 Organization of the circular genome of the CSSV species. Gray arrows indicate the deduced ORFs 1–3, X and Y capable of encoding proteins larger than 9 kDa. In the ORF3, we indicate the position of the movement protein (MP); coat protein (CP); the RTase/RNaseH region, with the domains for Aspartyl Proteinase (AP); Reverse Transcriptase (RT); and Ribonuclease H (RNase H).

species. Results on phylogenetic relationships between the first Ghanaian and Togolese full sequences showed that these relationships were more influenced by their geographical origin than on whether the sequences originated from mild or severe isolates.

Table 2 shows the nt sequence identity of pairwise combinations of representative CSSV isolates of each molecular group in the RT/RNase H region of ORF3 (500 bp). Considering the 20% divergence threshold in the RTase/RNaseH region indicated by ICTV in a region of 1500 nt for the determination of new badnavirid species, we could define 8 species. Seven species are now recognized by ICTV: *Cacao swollen shoot Togo B virus* (CSSTBV, first described as CSSV, first isolate sequenced, Agou1), *Cacao swollen shoot Togo A virus* (CSSTAV), *Cacao swollen shoot CD virus* (CSSCDV), *Cacao swollen shoot CE virus* (CSSCEV), *Cacao swollen shoot Ghana M virus* (CSSGMV), *Cacao swollen shoot Ghana N virus* (CSSGNV), *Cacao swollen shoot Ghana Q virus* (CSSGQV). Another full genome has been obtained recently and potentially corresponding to the eighth species *Cacao swollen shoot Ghana T virus* (CSSGTV).

Recently a phylogenetic tree was built from all the available complete sequences of badnaviruses (**Fig. 5, Table 3**). It shows that the group of viral species associated with the cacao swollen shoot disease is polyphyletic since the MBV1, HiBV and CiYMV species are positioned at the junction between the two phyla of CSSV species TA, TB (groups B and C), CD, CE (groups E, F, G, H, J, K, L, P) GM and GN on the one hand and GQ (groups Q and groups Q and R) and GT on the other hand.

As has been observed for most viruses studied, that badnavirus populations contain recombinant viruses. This was illustrated for the first time for CSSV with the sequencing of isolate Wobe12 (type member of species CSSTAV). Wobe12 possesses an ORF1 close to ORF1 (96% nt identity) of isolate Agou1 (group C of species CSSTBV) but shares only 75% nt identity with the complete sequence of Agou1. Recombinations allow the emergence of new viral isolates and increases the variability of viral populations beyond what is observed by studying a single region of the genome. The sites of recombination in the genome and their frequency should be extensively investigated in the future to have a better understanding of virus evolution in epidemics.

Epidemiology: Geographical Distribution of Viral Molecular Diversity

Epidemiological characteristics of the disease derive directly from its natural transmission properties. The mealybug vectors (nymph of both sexes and adult females) spread the disease radially over short distances around the periphery of outbreaks by crawling through the canopy from infected trees to adjacent healthy trees. New outbreaks were shown to be occasioned by “jump spread” over greater distances by wind borne viruliferous mealybugs and mainly by the very active small first instar nymphs.

A systematic molecular characterization of viral isolates make it possible to study the molecular epidemiology of CSSV. Viral molecular diversity studied in the first region of the ORF3 for isolates from the three countries: Togo, Côte d’Ivoire and Ghana

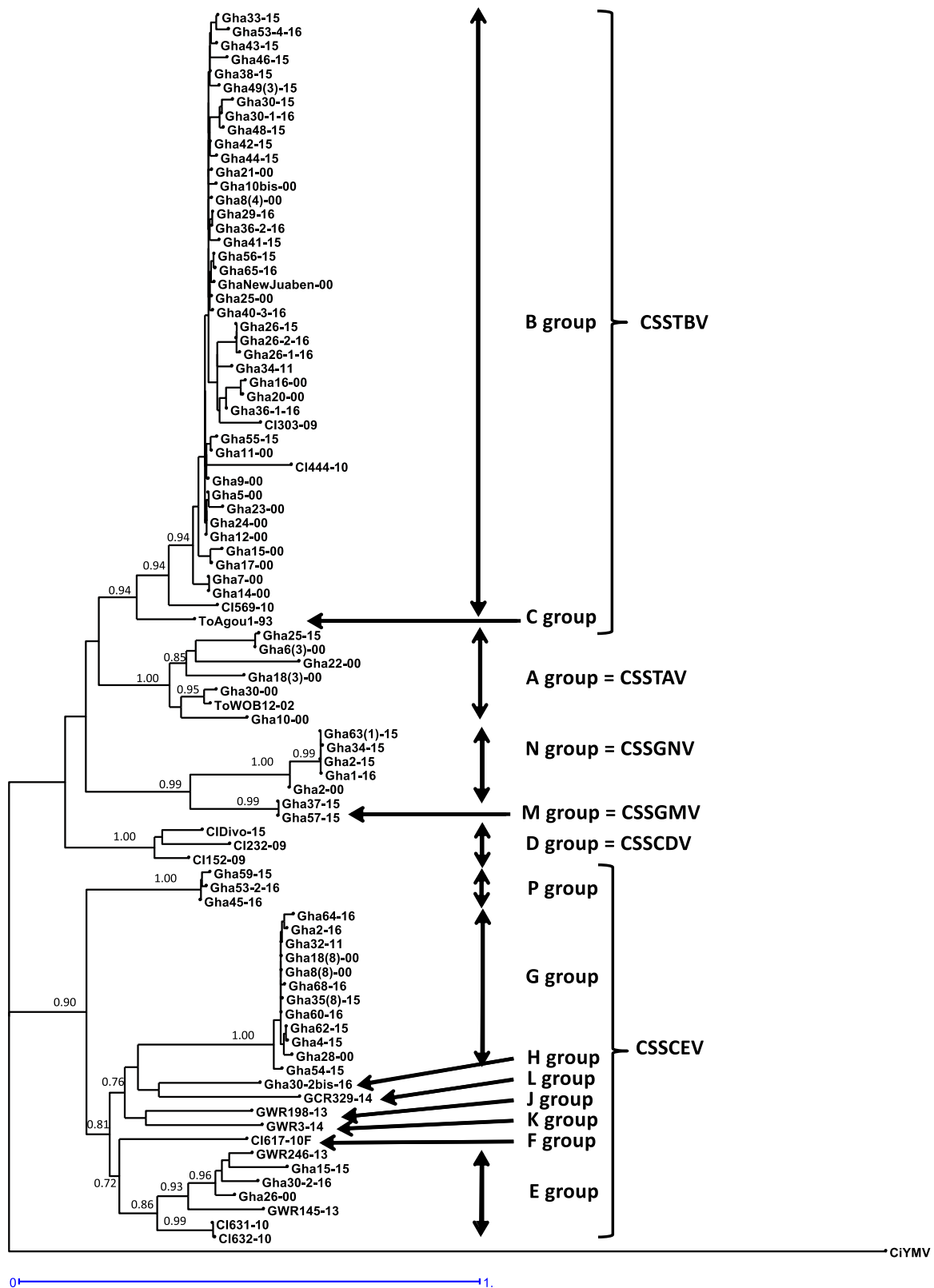


Fig. 4 Maximum likelihood phylogenetic tree of CSSV sequences based on alignment of the 5' end of open reading frame 3 (ORF3). Numbers on the branches represent the SH-aLRT (approximate likelihood ratio test) branch supports over 0.7. The names of the CSSV groups A, B, C, D, E, F, G, H, J, K, L, M, N and P are indicated, as well as the names of the corresponding species. The Citrus yellow mosaic virus sequence – CiYMV (AF347695) – is used as the outgroup. The names of sequences include the abbreviation of the country: CI for Côte d'Ivoire, G or Gha for Ghana, To for Togo, a sampling number along with a region code or a number corresponding to the clone number in brackets, and the year of sampling 1993–2016 coded as 93–16. When sequences belonging to different groups are obtained for the same isolate, the letter corresponding to the group is affixed to the name of the isolate.

Table 2 Nucleotide sequence identity of pairwise combinations of representative CSSV isolates of each molecular group in the RT/RNase H region of ORF3 (500 bp)

Species name	CSSV group	isolate name	A	B	C	D	E	F	G	J	K	L	P	M	N	Q	R	T
CSSTAV	A	ToWOB12-02	100	73	76	75	72	72	73	70	72	71	72	70	69	60	58	56
CSSTBV	B	GhaNew Juaben-00		100	89	72	73	75	73	70	76	74	72	74	70	59	57	59
	C	ToAgou1-93			100	71	73	74	75	70	76	74	73	74	71	60	59	61
CSSCDV	D	CI152-09				100	69	69	70	68	69	69	69	69	67	62	61	60
CSSCEV	E	CI632-10					100	83	83	82	86	82	85	72	71	59	56	60
	F	CI617-10						100	80	84	85	85	81	71	71	60	60	59
	G	Gha62-15							100	78	82	83	82	72	72	58	59	58
	J	GWR198-13								100	80	80	81	71	68	59	58	56
	K	GWR3-14									100	83	83	71	72	58	58	59
	L	GCR329-14										100	82	72	73	58	58	60
	P	Gha59-15											100	71	70	58	56	56
CSSGMV	M	Gha57-15												100	76	60	57	57
CSSGNV	N	Gha63-15													100	57	55	56
CSSGQV	Q	Gha1-00														100	76	60
	R	Gha60-16															100	62
CSSGTV	T	Gha2-00																100

has provided a picture of the geographical distribution of this diversity in these countries (Fig. 6). In most cases, sequences from the same farm/plot always belonged to the same group, but there were some instances where isolates belonging to two different groups were identified on the same plot/farm. The geographical distribution of the viral molecular diversity shows that some group (species TA or CD, or group C of species TB) have remained well localized in a single country or at the border between two countries, while other species (group B of species TB) have a more ubiquitous distribution in the different regions of the three countries. The ubiquitous nature of the B group may mean it is readily transmitted by mealybugs, probably because of higher concentrations in the cacao phloem than the other groups. This group from species TB tends to induce more easily the typical red vein banding symptom seen in young or flush leaves of infected cacao plants, since most of the samples collected in the Eastern Region containing only this group, whether young or mature, had the conspicuous red vein banding or vein clearing symptom.

When CSSV groups identified in Ghana are compared to those found in Côte d'Ivoire, only species TB and CE are common to both countries. In addition to these two species, Côte d'Ivoire has a species (CD) which have not been detected in Ghana up to now and Ghana also has species TA, GM, GN, GQ, group C (species TB) and groups G, J, K, L (species CE) which have not been identified or detected in Côte d'Ivoire yet. This distribution may reflect differences in cacao crop planting dates in the regions of these countries. These observations make it difficult to predict a single origin for CSSV in West Africa amongst the cacao growing neighboring countries and we can therefore make the assumption that the different putative CSSV species emerged from different sources at different times. In fact, the high variability observed within CSSV populations compared to its very short evolutionary history on cacao trees further suggests the existence of several parallel emergences, more likely by host shift from different native hosts to cacao trees in the various countries of West Africa. Studies on CSSV more exhaustively in adventitious plants or insect vectors would make it possible to better understand the geographical distribution and the epidemiologic factors that lead to the rapid expansion of the disease in some plots. However, the exact role played by these wild hosts in the epidemiology of CSSV is today difficult to investigate, because the prevalence of CSSD is greater in cacao trees than in these alternative host plants.

The diversity of CSSV observed in the CRIG Museum (a collection of viral isolates collected in Ghana since the 1940s) compared to cacao farms in Ghana is interesting from a historical point of view, because this range can be seen as a picture of the diversity of the CSSV population at the time the Museum was established. Indeed, it should be noted that this viral collection has not been regularly updated. The species CSSGQV, CSSGTV currently only detected in the CRIG Museum may have been present previously in cacao farms but were absent or only present at a frequency that avoided detection in the farms. Conversely, groups of the species CE (J, K and L) and group C only detected in cacao farms of the species TB appear to correspond to recently emerged groups in the cacao farms. Group C has been detected in Togo since 1993 but could have emerged either in Togo or in Ghana, since the establishment of the CRIG Museum. Species CE, is underrepresented in the CRIG Museum but present in a high proportion in field samples (29% of the total isolates characterized in Ghana) and appears to correspond to a species that has recently spread to cacao plots especially in the Western and Brong Ahafo regions of Ghana.

The complexity of the molecular diversity of viral species found in symptomatic cacao leaves means that there is a need to study the range of aggressiveness of the different species or groups. In future the symptomatology/aggressiveness associated with single strain infections arising from representative species should be explored as should the impact on hosts from mixed infections of these genomes.

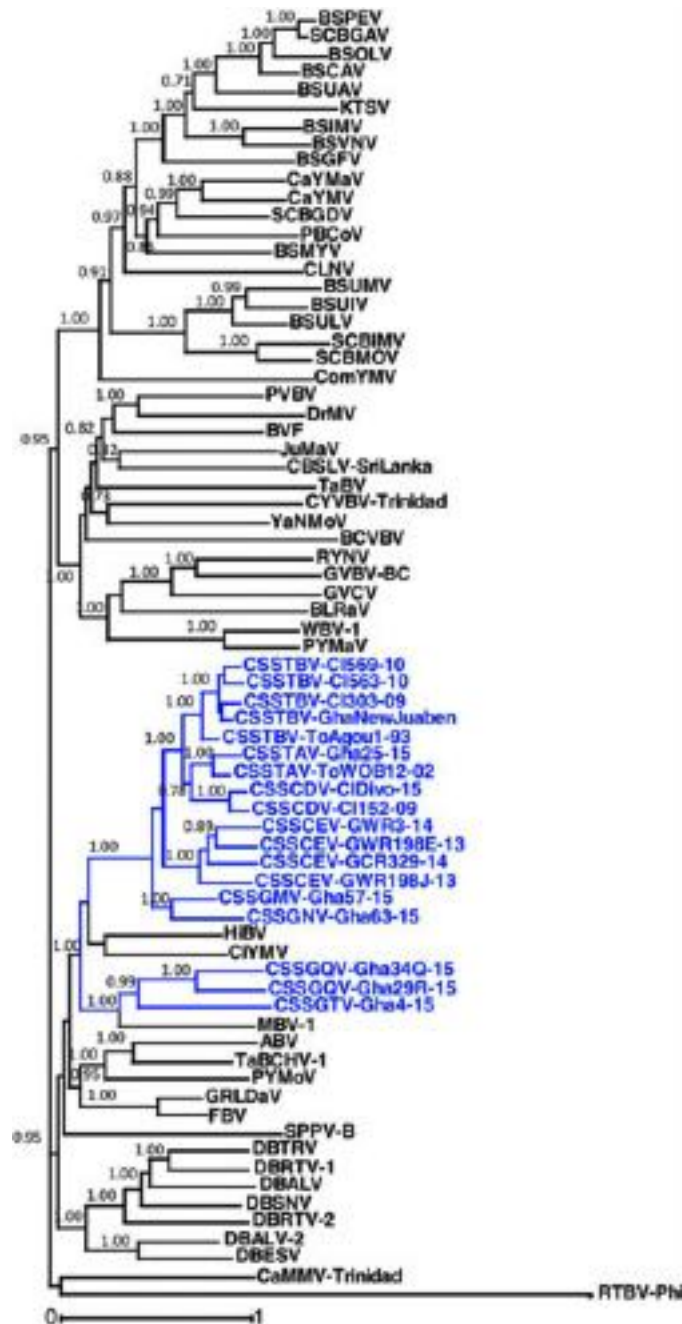


Fig. 5 Maximum likelihood phylogenetic tree generated by PhyML based on complete nucleotide sequences of badnaviruses. The blue sequences correspond to the viral sequences infecting the cocoa tree and associated with the cacao swollen shoot disease. Numbers on the branches represent the SH-aLRT (approximate likelihood ratio test) branch supports over 0.7. The scale bar represents the substitution number per base. The GenBank accession numbers of Badnavirus sequences used are listed in [Table 1](#) and the CSSV sequences are listed in [Table 3](#). Rice tungro bacilliform virus (RTBV-Phi – X57924) was used as outgroup.

Diagnosis

Serological Diagnostics

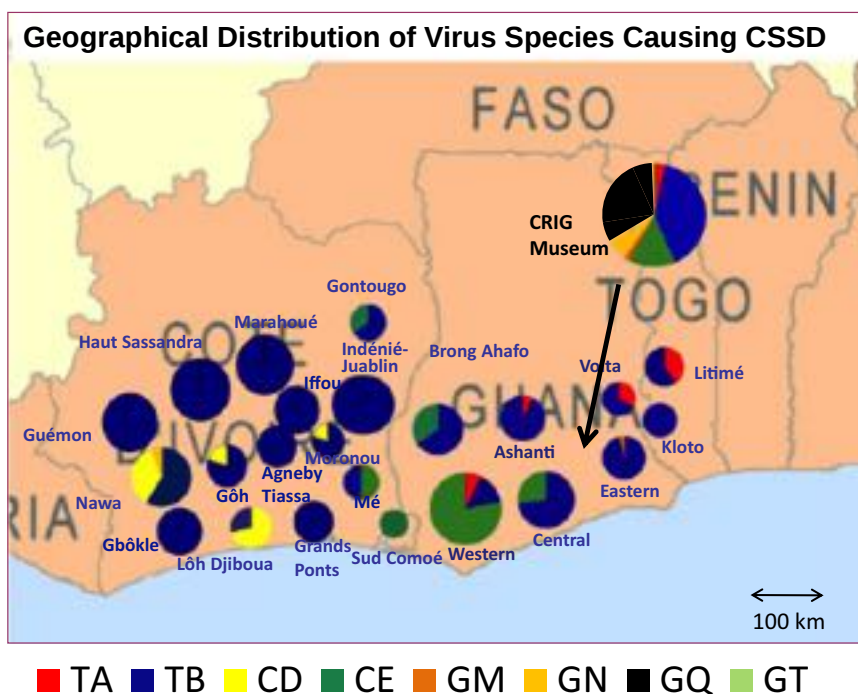
The virus is not strongly immunogenic; however, several antisera have been raised and shown to react with CSSV. Enzyme-linked immunosorbent assay (ELISA) has been used to detect CSSV but high background values were obtained and difficulties were found in detecting CSSV in plants suspected of having only a low virus titer. Immunosorbent electron microscopy has been used for detection and comparison of some isolates of CSSV in Ghana. Both of these techniques cannot detect latent infection.

Table 3 List of CSSV isolates with complete genome sequences and their NCBI accession number

<i>Cacao species in the genus Badnavirus</i>	<i>Abbreviation-Isolate</i>	<i>Accession #</i>
<i>Cacao swollen shoot CD virus</i>		
Cacao swollen shoot CD virus – CI152-09	CSSCDV-CI152-09	JN606110
Cacao swollen shoot CD virus – CIDivo-15	CSSCDV-CIDivo-15	MF642718
<i>Cacao swollen shoot CE virus</i>		
Cacao swollen shoot CE virus – GI632-10	CSSCEV-GI632-10	MF642719
Cacao swollen shoot CE virus – GWR198E-13	CSSCEV-GWR198E-13	MF642720,
Cacao swollen shoot CE virus – GWR198J-13	CSSCEV-GWR198J-13	MF642721
Cacao swollen shoot CE virus – GWR3-14	CSSCEV-GWR3-14	MF642722
Cacao swollen shoot CE virus – GCR329-14	CSSCEV-GCR329-14	MF642723
Cacao swollen shoot CE virus – GH64	CSSCEV-GH64	KX592572
Cacao swollen shoot CE virus – GH67	CSSCEV-GH67	KX592571
Cacao swollen shoot CE virus – CI275	CSSCEV-CI275	KX592573
Cacao swollen shoot CE virus – CI286	CSSCEV-CI286	KX592584
<i>Cacao swollen shoot Ghana M virus</i>		
Cacao swollen shoot Ghana M virus – Gha57-15	CSSGMV-Gha57-15	MF642724
Cacao swollen shoot Ghana M virus – NIG1	CSSGMV-NIG1	MH785297
Cacao swollen shoot Ghana M virus – NIG10	CSSGMV-NIG10	MH785300
Cacao swollen shoot Ghana M virus – NIG12	CSSGMV-NIG12	MH785301
Cacao swollen shoot Ghana M virus – NIG13	CSSGMV-NIG13	MH785302
Cacao swollen shoot Ghana M virus – NIG16	CSSGMV-NIG16	MH785303
Cacao swollen shoot Ghana M virus – NIG18	CSSGMV-NIG18	MH029282
Cacao swollen shoot Ghana M virus – NIG5	CSSGMV-NIG5	MH029281
Cacao swollen shoot Ghana M virus – NIG7	CSSGMV-NIG7	MH785298
Cacao swollen shoot Ghana M virus – NIG9	CSSGMV-NIG9	MH785299
<i>Cacao swollen shoot Ghana N virus</i>		
Cacao swollen shoot Ghana N virus - Gha63-15	CSSGNV-Gha63-15	MF642725
<i>Cacao swollen shoot Ghana Q virus</i>		
Cacao swollen shoot Ghana Q virus – Gha2-15	CSSGQV-Gha2-15	MF642726
Cacao swollen shoot Ghana Q virus – Gha34Q-15	CSSGQV-Gha34Q-15	MF642727
Cacao swollen shoot Ghana Q virus – Gha64-99	CSSGQV-Gha64-99	MF642729
Cacao swollen shoot Ghana Q virus – Gha40-15	CSSGQV-Gha40-15	MF642728
Cacao swollen shoot Ghana Q virus – Gha29-15	CSSGQV-Gha29-15	MF642730
Cacao swollen shoot Ghana Q virus – Gha34R-15	CSSGQV-Gha34R-15	MF642731
Cacao swollen shoot Ghana Q virus – Gha37-15	CSSGQV-Gha37-15	MF642732
Cacao swollen shoot Ghana Q virus – Gha3915	CSSGQV-Gha3915	MF642733
Cacao swollen shoot Ghana Q virus – Gha53-15	CSSGQV-Gha53-15	MF642734
Cacao swollen shoot Ghana Q virus – Gha54-15	CSSGQV-Gha54-15	MF642735
<i>Cacao swollen shoot Togo A virus</i>		
Cacao swollen shoot Togo A virus – Wobe12-02	CSSTAV-Wob12-02	AJ781003
Cacao swollen shoot Togo A virus – Gha25-15	CSSTAV-Gha25-15	MF642716
<i>Cacao swollen shoot Togo B virus</i>		
Cacao swollen shoot Togo B virus – Agou1-93	CSSTBV-Agou1-93	L14546
Cacao swollen shoot Togo B virus – Nyongbo2-00	CSSTBV-NB2-00	AJ534983
Cacao swollen shoot Togo B virus – N1A-00	CSSTBV-N1A-00	AJ609020
Cacao swollen shoot Togo B virus – Peki-00	CSSTBV-Peki-00	AJ609019
Cacao swollen shoot Togo B virus – New Juaben-00	CSSTBV-New Juaben-00	AJ608931
Cacao swollen shoot Togo B virus – CI303-09	CSSTBV-CI303-09	MN179343
Cacao swollen shoot Togo B virus – CI563-10	CSSTBV-CI563-10	MN179344
Cacao swollen shoot Togo B virus – CI569-10	CSSTBV-CI569-10	MF642717
Cacao swollen shoot Togo B virus – C1311	CSSTBV-CI311	KX592580
Cacao swollen shoot Togo B virus – CI134	CSSTBV-CI134	KX592578
Cacao swollen shoot Togo B virus – CI135	CSSTBV-CI135	KX592581
Cacao swollen shoot Togo B virus – CI215	CSSTBV-CI214	KX592577
Cacao swollen shoot Togo B virus – CI301	CSSTBV-CI301	KX592579
Cacao swollen shoot Togo B virus – CI44	CSSTBV-CI44	KX592574
Cacao swollen shoot Togo B virus – CIS2	CSSTBV-CIS2	KX592583
Cacao swollen shoot Togo B virus – CIS3	CSSTBV-CIS3	KX592576
Cacao swollen shoot Togo B virus – CIT5	CSSTBV-CIT5	KX592582

Table 3 Continued

<i>Cacao species in the genus Badnavirus</i>	<i>Abbreviation-Isolate</i>	<i>Accession #</i>
Cacao swollen shoot Togo B virus – GH75	CSSTBV-Gh75	KX592575
<i>Unassigned species</i>		
<i>Cacao swollen shoot Ghana T virus</i>		
Cacao swollen shoot Ghana T virus – Gha4-15	CSSGTV-Gha4-15	MN179342

**Fig. 6** Geographical distribution of the species complex associated with cacao swollen shoot disease in the different outbreaks in West Africa, (Côte d'Ivoire, Ghana and Togo).

The viro-bacterial agglutination (VBA) test has been found to be a useful test for detecting CSSV in leaf tissue from infected trees. Using this assay, CSSV can be detected in trees showing symptoms as well as in infected, but symptomless, trees. The immunocapture polymerase chain reaction (IC-PCR) technique was adapted to the only detection of CSSV-1A isolates from Ghana (CSSTBV species).

There is considerable strain variation among the many recognized CSSV isolates, some of which react only weakly with certain antisera. Using monoclonal antibodies, four serotypes of CSSV have been distinguished by ELISA analysis of 31 samples of the virus from different geographical locations in Ghana. The efficiency of serological diagnosis depends on the use of polyvalent antiserum able to detect all serotypes. Since the 2000s and the first results on viral molecular variability, no work has been published on the development of serological methods for CSSV but these methods should not be excluded for the implementation of a versatile and sensitive diagnosis.

PCR Diagnosis

For a reliable PCR diagnosis, it is necessary to design primers from conserved regions of the genome. Until 2000, only few badnaviruses were completely sequenced, and only the end of the ORF3, which contains the conserved motifs specific of RTase and RNaseH made it possible to obtain primers useful for diagnosis. After alignment of the first six full-length sequences of CSSV, polyvalent primers pairs were designed in two different parts of the genome, in the movement protein (MP, first part of ORF3) and in the RTase/RNaseH region (third part of ORF3) (Fig. 3) and were then improved to be able to detect new species that were subsequently identified. However, the presence in cacao genomes of retrotransposons containing also the motifs specific of RTase and RNaseH makes it difficult to develop specific CSSV primers in this area.

PCR-based diagnosis is able to detect CSSV not only in symptomless leaves of symptomatic plants, but also in symptomless plants as early as one week post-inoculation. As the sensitivity of diagnosis is better when young symptomatic host plant leaves are tested, it is recommended that validation of the diagnostic test is done using this type of material in the first instance, especially when testing isolates from new areas.

Given the heterogeneous distribution of the virus in the plant, especially in adult trees, this PCR diagnosis should be used primarily to check for infection of plants with suspicious symptoms, or of grafts or rootstocks.

Control of the Disease

Attempts at CSSV control in Ghana have required substantial financial and manpower inputs. The “cutting-out” policies in place in Ghana since the early 1940s which attempted to control cacao swollen shoot disease resulted in the removal of over 190 million trees up to 1988. Over ten million infected trees which still required ‘cutting out’ were identified in the field by 1990.

Control of the swollen shoot disease based only on ‘cutting-out’ campaigns has not been successful due to several factors including political and socioeconomic problems. The better strategy for dealing with the disease is to develop a combination of control measures in an integrated approach. Moreover, this approach should implicate cacao farmers as much as possible.

Mild strain protection is a possibility which is being investigated. Mild strains which appear to confer some protection against the severe strains from species CSSTBV are available and are being tested in the field in Ghana. However, only a closely related mild strain is able to protect against another strain, as has been demonstrated for the Citrus tristeza virus. Now that there are so many different CSSV species, it would seem that it would be impossible to protect the cacao trees against them all.

The control of insect vectors such as mealybugs is difficult to implement due to the excessive cost of the molecules to be used and their toxicity which results to residues in the cacao beans produced by treated trees.

The most effective replantation method is to clear whole areas and replant in large compact block because of the relatively small proportion of trees in the vulnerable peripheral areas. The possibility of isolating new cacao plantings from infected cacao by an isolation distance of 10 m or planting barriers with CSSV-immune crops should be considered. Barrier crops such as banana, plantain, oil palm, coffee, cola, and citrus can form a barrier to the movement of vectors. The ‘cutting out’ of the adventitious plants of the genus *Commelina* recognized as alternative hosts should also be done.

The use of resistant cacao is advocated for replantation as many of the new hybrids available in West Africa do have some partial resistance to CSSV. Replanting with resistant cacao trees, however, requires the installation of a protocol of effective screening for resistance. Severe isolates representative of the different molecular groups should be used as well as a suitable screening method. However, a standardized inoculation method is not yet available because particle bombardment is difficult to develop on a large scale and agro-inoculation needs biosafety confinement. Mealybug transmission tests can nevertheless be set up by the national research institutes of the different producing countries, taking all precautions in terms of confinement (insect proof greenhouses).

Intermediate quarantine facilities are at present hampered by the lack of suitable indexing methods for the virus particularly if the germplasm is to be moved as stem cuttings. New polyvalent PCR diagnosis could be easily tested in intermediate quarantine facilities and compared with the grafting procedure used for indexation on Amelonado cacao seedlings.

See also: Caulimoviruses (*Caulimoviridae*)

Further Reading

- Abrokwha, F., Dzahini-Obiatey, H., Galyuon, I., Osae-Awuku, F., Muller, E., 2016. Geographical distribution of Cacao swollen shoot virus molecular variability in Ghana. *Plant Disease* 100 (10), 2011–2017.
- Brunt, A.A., 1970. Cacao Swollen Shoot Virus. CMI/AAB Descriptions of Plant Viruses N°10. Wellesbourne: Association of Applied Biologists.
- CABI – Crop Protection Compendium, Crop protection Compendium, 2002. Cacao Swollen Shoot Virus. Wallingford: CABI Publishing.
- Castel, C., Amefia, Y.K., Djiekpor, E.K., Partiot, M., Segbor, A., 1980. Le swollen shoot du cacaoyer au Togo. Les différentes formes de viroses et leurs conséquences économiques. *Café, Cacao, Thé* 24 (2), 131–146.
- Fauquet, C., Mayo, M., Maniloff, J., Desselberger, U., Ball, L. (Eds.), 2005. *Virus Taxonomy: VIIIth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press.
- Hagen, L.S., Jacquemond, M., Lepingle, A., Lot, H., Tepfer, M., 1993. Nucleotide sequence and genomic organization of Cacao swollen shoot virus. *Virology* 196 (2), 619–628.
- Hagen, L.S., Lot, H., Godon, C., Tepfer, M., Jacquemond, M., 1994. Infection of *Theobroma cacao* using cloned DNA of Cacao swollen shoot virus and particle bombardment. *Molecular Plant Pathology* 4, 1239–1243.
- Jacquot, E., Hagen, L.S., Michler, P., *et al.*, 1999. In situ localization of Cacao swollen shoot virus in *Theobroma cacao*. *Archives of Virology* 144, 259–271.
- Lot, H., Djiekpor, E., Jacquemond, M., 1991. Characterization of the genome of Cacao swollen shoot virus. *Journal of General Virology* 72, 1735–1739.
- Muller, E., Sackey, S., 2005. Molecular variability analysis of five new complete Cacao swollen shoot virus genomic sequences. *Archives of Virology* 150, 53–66.
- Muller, E., 2016. Cacao swollen shoot virus (CSSV): History, biology, and genome. In: Bailey, Bryan A., Meinhardt, Lyndel W. (Eds.), *Cacao Diseases. A History of Old Enemies and New Encounters*. Cham: Springer International Publishing, pp. 337–358.
- Muller, E., Ravel, S., Agret, C., *et al.*, 2018. Next generation sequencing elucidates cacao badnavirus diversity and reveals the existence of more than ten viral species. *Virus Research* 244, 235–251.

Carmo-Like Viruses (*Tombusviridae*)

Miryam Pérez-Cañamás and Carmen Hernández, Institute for Plant Molecular and Cell Biology (Spanish National Research Council–Polytechnic University of Valencia), Valencia, Spain

© 2021 Elsevier Ltd. All rights reserved.

This is an update of F. Qu, T.J. Morris, *Carmovirus*. In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00365-4](https://doi.org/10.1016/B978-012374410-4.00365-4).

Nomenclature

aa Amino acid(s)

AGO Argonaute

CP Coat protein

diRNAs Defective interfering RNAs

dsRNA Double-stranded RNA

eIF Eukaryotic translation initiation factor

gRNA Genomic RNA

ICTV International Committee for Taxonomy of Viruses

kb Kilobase

kDa Kilodalton

MP Movement protein

mRNA Messenger RNA

NCR Non-coding region

nm Nanometer

nt Nucleotide(s)

ORF Open reading frame

Poly(A) Polyadenylated

RdRp RNA-dependent RNA polymerase

satRNA Satellite

sgRNA Sub-genomic RNA

spp Species (plural)

ssRNA Single-stranded RNA

tRNA Transfer RNA

VRC Viral replicase complex

Glossary

Defective interfering RNAs RNAs derived from the RNA genome of a virus in which a critical portion of such genome has been lost usually due to defective replication. These RNAs are associated to and depend on the parent virus for infection as it provides the lost factor.

Leaky-scanning Phenomenon in which a weak initiation codon triplet on mRNA is sometimes skipped by ribosome in translation initiation. The 40S ribosomal subunit continues scanning to further initiation codon. The weak initiation codon can be an AUG in a weak Kozak consensus context or a non-AUG codon.

Readthrough Consists of the “translation” of a stop codon, such that the synthesis of a larger proteins is allowed. It is a common phenomenon in viruses, and it has a regulatory function.

RNA silencing A mechanism of transcriptional and post-transcriptional gene regulation operating in many

eukaryotic organisms. It is mediated by 20–25 nt small RNAs produced from various RNA sources and functions in regulating RNA stability and translation as well as in chromatin modification underlying numerous developmental processes. It also plays a significant role in microbe–host interaction.

Satellite RNA A subviral RNA that depends on the co-infection of a host cell with a helper virus for its replication.

Silencing suppressor Virally encoded proteins from diverse families that interfere with host production or effectiveness of small RNAs involved in RNA silencing.

Sub-genomic RNA Viral RNA that derives from and is usually 3′ co-terminal to the viral genomic RNA and that is produced during infection to serve as mRNA for translation of internal and 3′-proximal genes.

Tombusvirid Member of family *Tombusviridae*.

Introduction

In 1988, a series of viruses forming polyhedral virions which enclosed monopartite single-stranded (ss) RNA genomes of ~4 kb and that were represented by Carnation mottle virus (CarMV) were described. Since then, more than twenty-five related viruses sharing significant sequence homology and overall gene arrangement have been identified and are currently known under the generic name of carmo-like viruses. Investigations in the last years have significantly improved our knowledge on the biological and molecular properties of carmo-like viruses which, moreover, has helped to refine their classification. In this article, summary and updated information on this group of infectious agents is provided.

Taxonomy, Classification, and Evolutionary Relationships

The International Committee for Taxonomy of Viruses (ICTV) currently recognizes sixteen genera within the family *Tombusviridae* which, in turn, are clustered into three subfamilies: *Calvusvirinae*, *Procedovirinae*, and *Regressovirinae*. Four out of the fourteen genera included in *Procedovirinae* comprise carmo-like viruses: *Alphacarmovirus*, *Betacarmovirus*, *Gammacarmovirus*, and *Pelarspovirus*. These genera show differential branching in phylogenetic trees based on their encoded RNA-dependent RNA polymerases (RdRps) (Fig. 1). The generic name “carmo” is derived from the first member of the group to be described, CarMV, currently being the type species of the genus *Alphacarmovirus*. Members of each of the four genera share recognizable, yet variable, sequence similarity with members of other genera of carmo-like viruses as well as with other genera of the family *Tombusviridae*. As other components of the

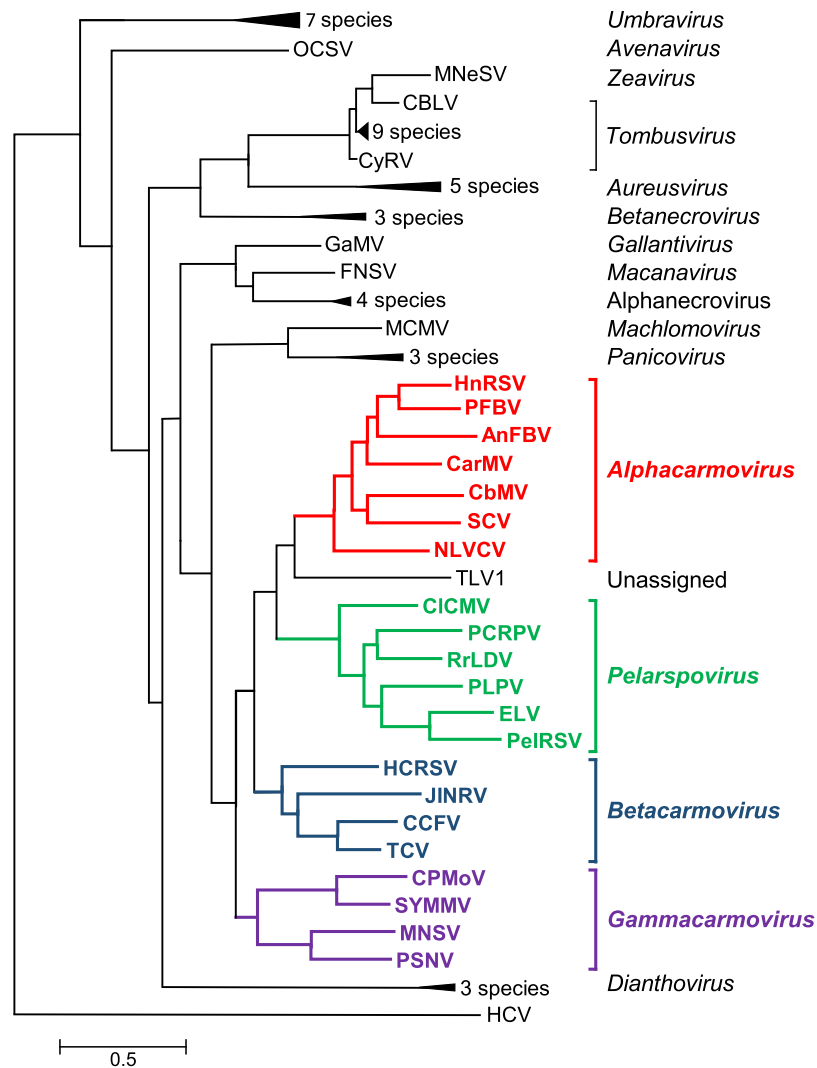


Fig. 1 Phylogenetic tree of the RdRps of tombusvirids. The alignment of was made using MUSCLE while tree was generated with the Maximum Likelihood (ML) algorithm in MEGA7 (<https://www.megasoftware.net/>). The analysis involved 64 amino acid sequences (those corresponding to the RdRps of all recognized tombusvirids). Condensed triangles mark monophyletic lineages. *Tombusviridae* subfamily names are on the right, and members of the four carmo-like virus genera are colored in red (*Alphacarmovirus*), green (*Pelarspovirus*), blue (*Betacarmovirus*), and purple (*Gammacarmovirus*). Virus names and acronyms are as follows: *Angelonia flower break virus* (AnFBV), *Calibrachoa mottle virus* (CbMV), *Cardamine chlorotic fleck virus* (CCFV), *Carnation mottle virus* (CarMV), *Clematis chlorotic mottle virus* (CICMV), *Cowpea mottle virus* (CPMoV), *Cucumber Bulgarian latent virus* (CBLV), *Cymbidium ringspot virus* (CyRSV), *Elderberry latent virus* (ELV), *Furcraea necrotic streak virus* (FNSV), *Galinsoga mosaic virus* (GaMV), *Hibiscus chlorotic ringspot virus* (HCRSV), *Honeysuckle ringspot virus* (HnRSV), *Japanese iris necrotic ring virus* (JINRV), *Maize chlorotic mottle virus* (MCMV), *Maize necrotic streak virus* (MNeSV), *Melon necrotic spot virus* (MNSV), *Nootka lupine vein clearing virus* (NLVCV), *Oat chlorotic stunt virus* (OCSV), *Pea stem necrosis virus* (PSNV), *Pelargonium chlorotic ring pattern virus* (PCRPV), *Pelargonium flower break virus* (PFBV), *Pelargonium line pattern virus* (PLPV), *Pelargonium ringspot virus* (PeIRSV), *Rosa rugosa leaf distortion virus* (RrLDV), *Saguaro cactus virus* (SCV), *Soybean yellow mottle mosaic virus* (SYMMV), *Trailing lespedeza virus 1* (TLV1), and *Turnip crinkle virus* (TCV). *Hepatitis C virus* (HCV) RdRp was used as outgroup.

family, carmo-like viruses have icosahedral virions of about 30 nm in diameter, with $T=3$ symmetry, that is composed of 180 coat protein (CP) subunits of about 37–42 kDa and a ssRNA genome ranging in size from 3.9 to 4.3 kb. In all cases, the genome consists of a unique positive (+) sense ssRNA molecule that contains, at least, five definitive open reading frames (ORFs). Genome organizations of *Alphacarmovirus*, *Betacarmovirus*, and *Gammacarmovirus* are very similar and can be exemplified by those of their type species, corresponding to CarMV, Turnip crinkle virus (TCV), and Melon necrotic spot virus (MNSV), respectively (Fig. 2). TCV is probably the best studied virus among carmo-like viruses. It was the first carmo-like virus for which infectious transcripts were produced from a cDNA clone of the genome and detailed information about genome functions and virion configuration has been obtained, in the latter case including crystal structure determination. TCV genome harbors five ORFs encoding proteins of about 28, 88, 8, 9, and 38 kDa from the 5'- to the 3' end, respectively (Fig. 2). Leaving aside some specific features that will be mentioned in the next sections, the genome organization of pelarspoviruses resemble those of other carmo-like viruses and can be illustrated by the one from its type species, *Pelargonium line pattern virus* (PLPV). PLPV genome contains five genes encoding proteins of about 27, 87, 7, 9.7, and 37 kDa from the 5'- to the 3' end, respectively (Fig. 2). Some carmo-like viruses, such as *Hibiscus chlorotic ringspot virus* (HCRSV, genus *Betacarmovirus*), have additional ORFs whose functions are not entirely clear.

To date, the nucleotide sequences of the 21 definitive carmo-like viruses have been determined (Table 1). These sequenced viruses share similar morphological and physicochemical properties with five additional viruses listed in Table 1, that are recognized as unassigned species of the family *Tombusviridae* by the ICTV. Moreover, genome sequences of four of these viruses have not yet been determined. Most carmo-like viruses are sufficiently distant from each other that they do not cross-react in standard RNA hybridization or serological tests.

Carmo-like viruses have properties in common with viruses belonging to other genera of the family *Tombusviridae*. Their particle structure and CP sequences are closely related to tombus-, aureus-, diantho-, macana-, gallanti-, and avenaviruses. Their RdRp genes share a remarkable level of homology with other tombusvirids and particularly with those belonging to subfamily *Procedovirinae*: *Machlomovirus*, *Panicovirus*, *Alphanecrovirus*, *Betanecrovirus*, *Aureusvirus*, *Gallantivirus*, *Macanavirus*, *Avenavirus*, *Zeavirus*, and *Tombusvirus*. RdRp genes of carmo-like viruses and of other tombusvirids also share similarity with more distantly related viruses, such as luteoviruses. These latter viruses currently belong to family *Luteoviridae* though their reassignment to the family *Tombusviridae* is expected in the near future. In a broader context, phylogenetic comparisons of viral RNA polymerase

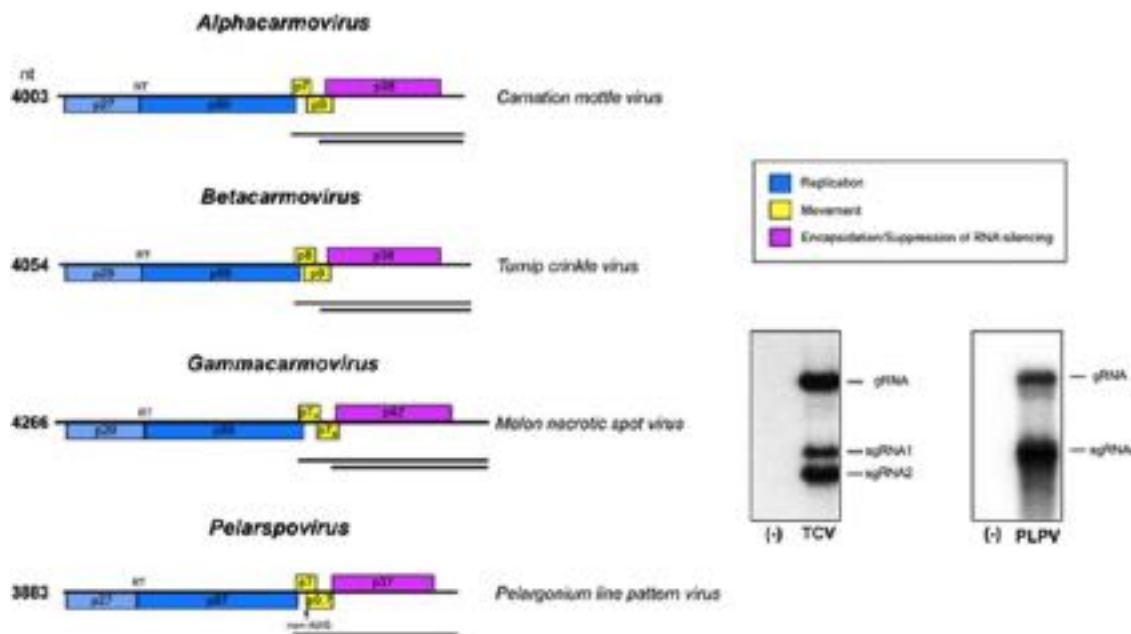


Fig. 2 Genome organizations of species of genera *Alphacarmovirus*, *Betacarmovirus*, *Gammacarmovirus*, and *Pelarspovirus* exemplified, respectively, by those of their type species, *Carnation mottle virus* (CarMV), *Turnip crinkle virus* (TCV) and *Melon necrotic spot virus* (MNSV), and *Pelargonium line pattern virus* (PLPV). The colored boxes represent open reading frames with the sizes of the encoded proteins, in kilodaltons, indicated within the boxes. Blue, yellow and pink boxes have been used, respectively, for genes encoding proteins involved in replication, movement and encapsidation/suppression of RNA silencing, as indicated in the inset at the right. The two proteins required for replication are translated from the genomic RNA (gRNA) with the viral RNA-dependent-RNA polymerase (86–89 kDa in the Figure) being translated by readthrough (RT) of an amber codon at the end of the 5'-proximal gene. Thin lines under genomes indicate sub-genomic RNAs (sgRNAs) produced by members of each genera to translate internal and 3' proximal genes. Dashed line of PLPV MP2 marks a noncanonical start codon. The bottom panel at the right corresponds to Northern blot results for detection of viral RNAs in TCV- (left) and PLPV- (right) infected plants to illustrate the typical pattern of accumulation of genomic and sub-genomic RNAs in genera *Alphacarmovirus*, *Betacarmovirus*, and *Gammacarmovirus*, in the first case, and genus *Pelarspovirus*, in the second.

Table 1 List of species in each genus composing the carmo-like viruses and unassigned species in the family *Tombusviridae*

Genus/Species	Virus acronym	Genbank accession
<i>Alphacarmovirus</i>		
<i>Angelonia flower break virus</i>	AnFBV	NC_007733.2
<i>Calibrachoa mottle virus</i>	CbMV	NC_021926.1
<i>Carnation mottle virus</i>	CarMV	NC_001265.2
<i>Honeysuckle ringspot virus</i>	HnRSV	NC_014967.1
<i>Nootka lupine vein clearing virus</i>	NLVCV	NC_009017.1
<i>Pelargonium flower break virus</i>	PFBV	NC_005286.1
<i>Saguaro cactus virus</i>	SCV	NC_001780.1
<i>Betacarmovirus</i>		
<i>Cardamine chlorotic fleck virus</i>	CCFV	NC_001600.1
<i>Hibiscus chlorotic ringspot virus</i>	HCRSV	NC_003608.1
<i>Japanese iris necrotic ring virus</i>	JINRV	NC_002187.1
<i>Turnip crinkle virus</i>	TCV	NC_003821.3
<i>Gammacarmovirus</i>		
<i>Cowpea mottle virus</i>	CPMoV	NC_003535.1
<i>Melon necrotic spot virus</i>	MNSV	NC_001504.1
<i>Pea stem necrosis virus</i>	PSNV	NC_004995.1
<i>Soybean yellow mottle mosaic virus</i>	SYMMV	NC_011643.1
<i>Pelarspovirus</i>		
<i>Clematis chlorotic mottle virus</i>	CICMV	NC_033777.1
<i>Elderberry latent virus</i>	ELV	NC_026239.1
<i>Pelargonium chlorotic ring pattern virus</i>	PCRPV	NC_005985.1
<i>Pelargonium line pattern virus</i>	PLPV	NC_007017.2
<i>Pelargonium ringspot virus</i>	PeIRSV	NC_026240.1
<i>Rosa rugosa leaf distortion virus</i>	RrLDV	NC_020415.1
<i>Species not assigned to a genus</i>		
<i>Ahlum waterborne virus</i>	AWBM	Unsequenced
<i>Bean mild mosaic virus</i>	BMMV	Unsequenced
<i>Cucumber soil-borne virus</i>	CSBV	Unsequenced
<i>Trailing lespedeza virus 1</i>	TLV1	NC_015227.2
<i>Weddel waterborne virus</i>	WWBV	Unsequenced

genes have identified the tombusvirids as a representative virus cluster for one of three RNA virus super-groups with relatedness to animal viruses of the family *Flaviviridae* and small RNA phages (*Leviviridae*).

Distribution, Host Range, Transmission, and Economic Significance

Carmo-like viruses are distributed worldwide and, with some exceptions, are reported to cause mild or asymptomatic infections in relatively restricted natural host ranges. Despite their usually moderate phenotypic effects, many of these viruses, although not all, accumulate to high concentrations in infected tissues.

Notwithstanding their narrow natural host ranges, carmo-like viruses have, in general, broad experimental host ranges that frequently include herbaceous plant species amenable to experimental manipulation. A number of carmo-like viruses have been identified in ornamental hosts. CarMV (genus *Alphacarmovirus*) is one of the most noteworthy, because its high incidence and wide geographical spread in cultivated carnations. PLPV (genus *Pelarspovirus*) and *Pelargonium flower break virus* (PFBV, genus *Alphacarmovirus*) are other examples of highly widespread viruses, in this latter case in *Pelargonium* spp., causing diseases either alone or in association with other viruses. Information on the prevalence of other carmo-like viruses affecting ornamentals, such as *Angelonia flower break virus* and *Calibrachoa mottle virus* (CMoV) (both in the genus *Alphacarmovirus*), HCRSV and *Japanese iris necrotic ring virus* (JINRV) (both in the genus *Betacarmovirus*), *Pelargonium chlorotic ring pattern virus*, *Pelargonium ringspot virus*, and *Rosa rugosa leaf distortion virus* (RrLDV) (the three of them in the genus *Pelarspovirus*), is more scarce. However, such prevalence is likely to be highly influenced by the distribution of the corresponding infected nursery stocks as most ornamental plants are vegetatively propagated. Illustrating the harmful effects that infection by these viruses may cause, the severe disease symptoms induced by HCRSV on ornamental hibiscus varieties has led to a massive removal of this type of plants from many recreational areas in Singapore. Interestingly, HCRSV also infects kenaf (*Hibiscus cannabinus* L.), an annual crop of major interest to the wood-pulp industry.

Legume species can also be naturally infected by several carmo-like viruses, such as *Nootka lupine vein clearing virus* (NLVCV, genus *Alphacarmovirus*), or *Cowpea mottle virus* (CPMoV), *Pea stem necrosis virus* (PSNV), and *Soybean yellow mottle mosaic virus* (SYMMV),

all three belonging to genus *Betacarmovirus*. The geographical distribution of each of these viruses does not seem very broad so far, with CPMoV being potentially the most important since it may seriously limit the cultivation of cowpeas that are important for food supply throughout Sub-Saharan Africa and particularly in Nigeria. On the opposite side, MNSV is considered to be an endemic virus in greenhouses and open-field systems for the production of cucurbits, significantly reducing yields and causing vast economical damage worldwide.

Two carmo-like viruses, TCV and *Cardamine chlorotic fleck virus* (CCFV) (genus *Betacarmovirus*), naturally infect *Brassicaceae* species. TCV is neither common, nor widespread in nature, whereas the known distribution of CCFV is limited to the Mount Kosiusko alpine region of Australia. Both are able to experimentally infect *Arabidopsis thaliana*, the most relevant model plant for basic plant research, in which they accumulate to extremely high concentrations, often approaching a level equivalent to 0.5% of the fresh weight of the plant tissue. *Honeysuckle ringspot virus* (HnRSV) and *Saguaro cactus virus* (SCV) (genus *Alphacarmovirus*), are presumably of little agricultural concern, and have been isolated from very particular hosts and areas, honeysuckle and San Luis Obispo (California) in the first case, and giant saguaro cactus from Sonoran Desert (Arizona), in the second. The infrequent isolation of these genetically similar viruses in remote locations around the world has prompted the speculation that ancestor carmo-like viruses may have been introduced into their natural hosts well before the last Ice Age and have since co-evolved in isolation in their diverse host plants.

Spread of carmo-like viruses may occur through infected soil, irrigation water, vegetative propagation, grafting and/or mechanical inoculation. In addition, transmission by biological vectors has been reported in some instances. Thus, TCV and CPMoV can be transmitted by beetles, PFBV by thrips and MNSV and PSNV by the root-inhabiting fungus *Olpidium bornovanus*. Seed transmission of carmo-like viruses is not well documented but it could also be an important factor for the distribution of, at least, MNSV and CPMoV.

Virion Structure and Assembly

The molecular structure of five carmo-like viruses (TCV, CarMV, CPMoV, HCRSV, and MNSV) has now been studied in detail, with TCV being the first studied by high-resolution X-ray crystallography. The detailed information about CP structure and inter-subunit interactions established that TCV and *Tomato bushy stunt virus* (TBSV, genus *Tombusvirus*) show remarkable structural conservation. In this regard, the common structural features shared by other members of the family *Tombusviridae* have been primarily deduced from alignment of the amino acid sequence of their CPs with those of TBSV and TCV. TCV consists of a $T=3$ icosahedral capsid of 180 subunits of the 38 kDa CP. The individual CP subunit folds into three distinct domains typical of CP subunit of tombusvirids. The relatively basic N-terminal R domain extends into the interior of the virus particle and presumably interacts with viral RNA. The R domain is connected by an arm to the S domain which constitutes the virion shell. The S domain is attached through a hinge to the P domain which projects outward from the virion surface. The resulting surface protrusions give the virions a bumpy appearance in electron micrographs.

TCV CP subunits are believed to form dimers in solution and during assembly. TCV is the only carmovirus for which detailed *in vitro* assembly studies have been performed. The virion has been shown to dissociate at elevated pH and ionic strength to produce a stable RNA–CP complex (rp-complex) and free CP subunits. Reassembly under physiological conditions in solution could be demonstrated using the isolated rp-complex and the soluble CP subunits. This rp-complex, consisting of six CP subunits tightly attached to viral RNA, could be generated *in vitro* and was shown to be important in selective assembly of TCV RNA. A model for assembly was proposed in which three sets of dimeric CP interact with a unique site on the viral RNA to form an initiation complex to which additional subunit dimers could rapidly bind. Preliminary characterization of the ‘origin of assembly’ for this virus identified two possible sites based on the identification of RNA fragments in the rp-complex protected from RNase digestion by CP. Further *in vivo* studies narrowed the assembly origin site to a bulged hairpin-loop of 28 nt within a 180 nt region at the 3′ end of the CP gene.

Genome Structure and Protein Functions

As mentioned above, complete nucleotide sequences of all virus species currently included in the genera *Alphacarmovirus* (7), *Betacarmovirus* (4), *Gammacarmovirus* (4), and *Pelarspovirus* (6) have been determined (Table 1). Comparative studies of the deduced ORFs revealed that all of these viruses encode a similar set of genes that are closely related and are present in the same gene order as illustrated in Fig. 2. Such set of genes is preceded by a 5′ non-coding region (NCR), whose length ranges between 5 nt (PLPV) to 134 nt (PSNV), and is followed by a 3′ NCR that varies from 224 nt (SCV) to 405 nt (NLVCV) in length. Distinct viral species do not share extensive sequence homology within these regions and none of them possesses neither a poly(A) tail nor a tRNA-like structure, in their 3′ end. In addition, the 5′ end of the genomes is not capped. The genome organization of the carmo-like viruses is quite compact and most of the identified ORFs are overlapping each other. As shown for one or more members of each genus, the product of the most 5′ proximal ORF (25–29 kDa) and its readthrough product (83–89 kDa, the RdRp) are essential for replication of the virus genome. In addition, all carmo-like viruses characteristically present two small ORFs in the middle of the genome that encode two proteins (6–13 kDa) involved in cell-to-cell movement (movement proteins or MPs). Both MPs have been proposed to function in a coordinate fashion, one recruiting the viral RNA and the other exploiting the plant

endomembrane system to promote intra- and inter-cellular movement of carmo-like viruses. The 3' proximal gene encodes the viral CP which varies from 37 to 42 kDa for the different viruses. Interestingly, as proven for several carmo-like viruses, including CarMV, MNSV, PFBV, PLPV, and TCV, this protein also acts as suppressor of RNA silencing, a potent antiviral mechanism in plants that is conserved in nearly all eukaryotic organisms. This function is performed by other viral products in other genera of the family *Tombusviridae*, underlining the uniqueness of the structural subunit of carmo-like viruses.

Although the genome organizations of all sequenced carmo-like viruses are very similar, those of pelarspoviruses are distinguished by some specific features such as the presence of a non-AUG start codon, instead of the canonical AUG, initiating the MP2 ORF and the lack of an AUG codon in any frame between the AUG initiation codon of MP1 and the CP ORFs. In addition, some individual carmo-like viruses display specific characteristics. For example, the two small central ORFs in MNSV (p7a and p7b) are connected by an in-frame amber codon that could result in the production of a 14 kDa fusion protein of the two ORFs by a readthrough mechanism. Similarly, a potential double readthrough could lead to translation of a C-terminal extended RdRp in the case of CarMV, HnRSV, PFBV, and JINRSV. On their side, genomes of CPMoV, HCRSV, SYMMV, and RrLDV contain a sixth ORF nested within the 3' proximal CP gene. Three-four additional small ORFs have also been predicted in RrLDV genome. Nevertheless, the biological functionality of most of these extra ORFs remains to be ascertained. Indeed, unequivocal proof of activity has only been provided for the novel ORFs that are nested within the RdRp and the CP gene, respectively, of HCRSV.

Replication and Gene Expression

Some carmo-like viruses, such as TCV, replicate to very high concentrations in protoplast infections, with the genomic (g) RNA accumulating to levels approaching that of the ribosomal RNAs. This facilitates identification and characterization of both *cis*-acting RNA elements and viral products involved in viral multiplication. For expression of genes located in internal and 3'-proximal positions of the gRNA, carmo-like viruses transcribe either one (genera *Alphacarmovirus*, *Betacarmovirus*, and *Gammacarmovirus*) or two (genus *Pelarspovirus*) 3'-coterminal sub-genomic RNAs (sgRNAs). In the first case, the larger sgRNA (~1.7 kb) functions as the messenger (m) RNA for the two MP genes utilizing a leaky-scanning mechanism whereas the smaller sgRNA (~1.5 kb) is the mRNA for the CP gene. When only one sgRNA (~1.7 kb) is produced, it serves as mRNA for translation of the two MP genes as well as the CP gene. Such translation is accomplished through leaky-scanning processes that are favored by the suboptimal translational context of the initiating AUG of the MP1 gene, the non-canonical start codon of MP2 gene (Fig. 2), and the lack of an AUG codon between the MP1 AUG and CP AUG. Both models for gene expression are supported by *in vitro* translation experiments performed with, at least, TCV, SCV, and PLPV. The number of sgRNAs (one or two) together with the monophyletic lineage of the encoded RdRps (Fig. 1) are the main criteria that define the distinct genera of carmo-like viruses.

As other viruses with RNA genome, carmo-like viruses are thought to replicate through a negative (–) strand intermediate because virus-specific double-stranded (ds) RNAs corresponding in size to the gRNA characteristically accumulate in infected plant tissue. As shown for TCV, the protein encoded by the 5' proximal ORF and its readthrough product are the only virus-encoded components of the polymerase complex. When expressed from two separate mRNAs, both viral products complemented in *trans* to enable the genome replication. Besides interactions between viral products, the assembly of viral replicase complexes (VRCs) involves intricate interactions between viral and host factors, including cellular membranes that serve as platforms for the formation of VRCs. Associations of the VCRs of PFBV and MNSV with mitochondrial membranes have been described. The VRCs also catalyze synthesis of sgRNAs most likely through a premature termination mechanism, as suggested from results with TCV and PLPV. This mechanism involves generation of a sg (–) strand on a genomic (+) template because of the presence of a structural motif in the latter that causes the RdRp to pause before it reaches the template 5' end. The sg (–) molecules serve as templates for generation of sg (+) RNAs. A membrane-containing extract prepared from evacuated protoplasts of uninfected *Arabidopsis* plants has been shown to faithfully produce both genomic and sub-genomic RNAs from a full-length TCV RNA template. Such an extract may be useful for future characterization of viral as well as host elements required for virus replication and transcription.

Numerous studies have demonstrated that, besides the virus-encoded proteins, a plethora of structural elements in the viral RNA play critical roles in the genome replication of carmo-like viruses. These elements are located throughout the entire viral genome and occur in both the (+) and (–) strands. Their roles range from promoters, enhancers or repressors for RNA replication and transcription, enhancers for translation, signals for readthrough translation and specificity determinants for virus assembly. The current knowledge of these RNA structures suggests that different secondary, or even tertiary, structural motifs, some of them mutually exclusive, are formed at the different stages of virus multiplication, and the highly coordinated nature of their formation ensures optimal utilization of the compact viral genome.

Satellites, Defective-Interfering RNAs

TCV is the only carmo-like virus in which replication of associated small subviral RNAs in infected plants has been characterized, and the situation for this virus is curiously complex. TCV infections are associated with defective interfering (DI) RNAs derived totally from the parent genome (e.g., RNA G of 342–346 nt), satellite (sat) RNAs of non-viral origin (e.g., RNAs D, 194 nt and F, 230 nt), and chimeric RNAs (e.g., RNA C of 356 nt) with a 5' region derived from sat RNA D and a 3' region derived from the 3' end of the TCV genome. All three types of small RNAs depend on the helper virus for their replication and encapsidation within

the infected plant. The different satellites and DI RNAs have been shown to affect viral infections in different ways. Both RNAs C and G intensify viral symptoms while interfering with the replication of the helper virus, while RNAs D and F seem to produce no detectable effects on either expression of symptoms or helper virus replication.

Virus–Host Interaction

Considerable progress has been made in the last two decades in our understanding of the molecular mechanisms of virus–plant interactions with significant contributions coming from studies utilizing TCV and its host plant *Arabidopsis*. Most notably, one ecotype of *Arabidopsis thaliana* has been determined to be resistant to TCV infection and the resistance gene (HRT) and its encoded protein have been characterized. Additional host factors that contribute to the resistance response have also been elucidated including the identification of a novel transcription factor (TIP) whose interaction with TCV CP is critical to prevent basal resistance. The determination that TCV CP is targeted by HRT resistance protein suggests that TCV CP plays a key role in combating antiviral defense mechanisms of the plant host, in addition to being required for virus assembly. Indeed, as mentioned above, distinct studies have established that TCV CP is a strong suppressor of RNA silencing, a host defense mechanism that targets invading RNA. TCV CP seems to develop this suppressor function through different mechanisms, one of which consists in interacting with Argonaute (AGO) proteins, the key effectors of RNA silencing. In addition, it has been reported that TCV CP requires RAV2, an ethylene inducible factor, for suppression of RNA silencing. The role of RNA silencing in the fight against TCV has been also evidenced by the high accumulation of suppressor-deficient TCV mutants in *A. thaliana* plants lacking either AGO1 or one or more of the four Dicer-like enzymes (dsRNA-specific endoribonucleases that play an essential role in triggering RNA silencing).

Besides host factors involved in the resistance response against TCV, a host factor that most likely assists the virus during its life cycle has been identified. This factor corresponds to eukaryotic translation initiation factor (eIF) 4G of *Arabidopsis* that has been found to augment TCV accumulation, presumably by promoting more efficient translation of viral genes. In case of MNSV, genetic and biochemical evidence showing that translation is eIF4E-dependent in melon has been obtained.

Similarly to that found for the CP of TCV, the CP of PLPV, which also acts as RNA silencing suppressor, interacts with AGO proteins. Interestingly, this PLPV viral product has been reported to interact as well with distinct importins alpha that are crucial nuclear-transport receptors. Such interaction seems to favor the accumulation of PLPV although the precise underlying molecular mechanism remains to be known.

Other relevant aspect of virus-host interactions refers to symptom elicitation. This question has been particularly tackled with TCV and its satellite RNA C. It has been shown that TCV CP is important in satellite RNA C interactions in the host plant. Normally, the presence of RNA C results in symptom intensification in TCV infections. However, when the TCV CP ORF is either deleted or replaced by the CCFV CP ORF, RNA C attenuates symptoms caused by the helper virus suggesting that CP either downregulates the replication of RNA C or enhances its own competitiveness.

Finally, the replicase gene of TCV has also been implicated in the symptom modification by satellite RNA C by two independent groups. The 3′ end of the TCV genome, a sequence common in TCV RNA, RNA C, and DI RNA G, was also suggested to be a symptom determinant.

Recent results also suggest that viral pathogenesis could be related with virus-induced alterations in the levels of long non-coding RNAs, which are involved in diverse and substantial biological processes, and/or of microRNAs, small RNAs that act as developmental regulators. In this context, floral structure disorders typically observed in TCV-infected *A. thaliana* plants have been correlated with upregulation of a long noncoding RNAs which, in turn, seems to lead to downregulation of a floral structure-related APETALA2 gene. However, changes in levels of regulatory RNAs induced by viral infections do not always correlate with phenotypic effects as shown with PLPV-infected *Nicotiana benthamiana* plants which, despite being asymptomatic, exhibited modifications in microRNA contents. These observations underline the complexity and the multiplicity of factors that may be involved in viral pathogenesis.

Further Reading

- Blanco-Pérez, M., Hernández, C., 2016. Evidence supporting a premature termination mechanism for sub-genomic RNA transcription in Pelargonium line pattern virus: Identification of a critical long-range RNA-RNA interaction and functional variants through mutagenesis. *Journal of General Virology* 97, 1469–1480.
- Castañó, A., Ruiz, L., Hernández, C., 2009. Insights into the translational regulation of biologically active open reading frames of Pelargonium line pattern virus. *Virology* 386, 417–426.
- Csorba, T., Kontra, L., Burgián, J., 2015. Viral silencing suppressors: Tools forged to fine-tune host-pathogen coexistence. *Virology* 479–480, 85–103.
- Hacker, D.L., Petty, I.T.D., Wei, N., Morris, T.J., 1992. Turnip crinkle virus genes required for RNA replication and virus movement. *Virology* 186, 1–8.
- Kachroo, P., Yoshioka, K., Shah, J., Dooner, H.K., Kleissig, D.F., 2000. Resistance to Turnip crinkle virus in *Arabidopsis* is regulated by two host genes and is salicylic acid dependent but NPR1, ethylene, and jasmonate independent. *Plant Cell* 12, 677–690.
- Komoda, K., Naito, S., Ishikawa, M., 2004. Replication of plant RNA virus genomes in a cell-free extract of evacuated plant protoplasts. *Proceedings of the National Academy of Sciences of the United States of America* 101, 1863–1867.
- Lommel, S.A., Martelli, G.P., Rubino, L., Russo, M., 2005. Tombusviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press, pp. 907–936.

- Koenig, R., 1988. Carnation mottle virus and viruses with similar properties. In: Morris, T.J., Carrington, J.C. (Eds.), *The Plant Viruses: Polyhedral Virions With Monopartite RNA Genomes* 3. New York: Plenum, pp. 73–112.
- Nagy, P.D., Pogany, J., Simon, A.E., 1999. RNA elements required for RNA recombination function as replication enhancers *in vitro* and *in vivo* in a plus-strand RNA virus. *The EMBO Journal* 18, 5653–5665.
- Navarro, J.A., Pallás, V., 2017. An update on the intracellular and intercellular trafficking of carmoviruses. *Frontiers in Plant Science* 8, 1801.
- Pérez-Cañamás, M., Blanco-Pérez, M., Forment, J., Hernández, C., 2017. *Nicotiana benthamiana* plants asymptotically infected by Pelargonium line pattern virus show unusually high accumulation of viral small RNAs that is neither associated with DCL induction nor RDR6 activity. *Virology* 501, 136–146.
- Pérez-Cañamás, M., Hernández, C., 2015. Key importance of small RNA binding for the activity of a glycine-tryptophan (GW) motif-containing viral suppressor of RNA silencing. *Journal of Biological Chemistry* 290, 3106–3120.
- Ren, T., Qu, F., Morris, T.J., 2000. HRT gene function requires interaction between a NAC protein and viral capsid protein to confer resistance to Turnip crinkle virus. *The Plant Cell* 12, 1917–1925.
- Russo, M., Burgyn, J., Martelli, G.P., 1994. Molecular biology of *Tombusviridae*. *Advances in Virus Research* 44, 381–428.
- Truniger, V., Miras, M., Aranda, M.A., 2017. Structural and functional diversity of plant virus 3'-cap-independent translation enhancers (3'-CITEs). *Frontiers in Plant Science* 8, 2047.

Relevant Websites

- <http://www.rcsb.org/structure/3ZX8>
3ZX8: Cryo-EM reconstruction of native and expanded Turnip Crinkle virus
RCSB PDB.
- <https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2560077>
Complete genomes: *Procedovirinae*
NCBI.
- <http://www.els.net/WileyCDA/ElsArticle/refId-a0022338.html>
eLS.
- <http://www.els.net/WileyCDA/ElsArticle/refId-a0000756.html>
eLS.
- https://viralzone.expasy.org/53?outline=all_by_species
Viralzone.

Cassava Brown Streak Viruses (*Potyviridae*)

Basavaprabhu L Patil, ICAR–Indian Institute of Horticultural Research, Bengaluru, India

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
amiRNAs Artificial microRNAs
Cas9 CRISPR-associated nuclease 9
CBSD Cassava brown streak disease
CMGs Cassava mosaic geminiviruses
CRISPR Clustered regularly interspaced short palindromic repeats
eIF4E Eukaryotic translation initiation factor 4E
HAM1h Homologue of HAM1

HCPPro Helper component proteinase
kDa Kilo dalton
NGS Next generation sequencing
Nia Nuclear inclusion protein a
Nib Nuclear inclusion protein b
nt Nucleotide(s)
NTP Nucleoside triphosphate
RNAi RNA-interference
siRNAs Small interfering RNAs
VPg Viral protein-genome linked

Glossary

Artificial microRNAs They are similar to micro RNAs, that are single-stranded, ± 21 nt long, and designed by replacing the mature miRNA sequences of duplex within precursor-miRNAs.
CRISPR-associated nuclease 9 A genome editing technology.
Eukaryotic translation initiation factor 4E An eukaryotic translation initiation factor involved in directing ribosomes to the cap structure of mRNAs.
Gene silencing A method to regulate gene expression to prevent the expression of a targeted gene.
Genome editing A method to edit the gene employing different editing technologies.
HAM1-like sequence A gene that protects against the mutagenic effects through detoxification of abnormal pyrimidine as well as purine nucleotides.
Helper component proteinase A multifunctional protein encoded by viruses of the genus *Potyvirus* and other viruses of the family *Potyviridae*, involved in aphid transmission, polyprotein processing, and suppression of host antiviral RNA silencing.
microRNA A type of small non-coding RNA of ± 21 nt in size involved in gene regulation and gene silencing.
Next generation sequencing A ultra-high throughput massively parallel sequencing technology.

Nuclear inclusion protein a It is one of the proteases encoded by potyviruses.
Nuclear inclusion protein b It is a RNA-dependent RNA polymerase encoded by potyviruses.
Phytosanitation It is one of the plant disease management strategy that involves removal and destruction of diseased plant parts to prevent further spread of the pathogen.
RNA silencing Sequence-specific regulation of gene expression triggered by double-stranded RNA, also termed RNA-interference (RNAi).
Silencing suppression Gene silencing suppression mechanism by certain specialized proteins encoded by plant viruses.
Small interfering RNAs A class of double-stranded non-coding RNA molecules, 20–25 base pairs in length, operating within the RNA interference pathway.
Transgenic plants Transgenic or genetically modified plants have genes that are derived from another species inserted into their genome through recombinant DNA technology.
Viral protein-genome linked A potyviral multifunction protein covalently linked to 5' end of viral RNA, which interacts with the eukaryotic translation initiation factor eIF4E, a cap-binding protein, which is imminent for virus infection.

Introduction

There are at least 20 different viruses that infect the tropical plant cassava (*Manihot esculenta* Crantz), of which the cassava infecting begomoviruses and the brown streak associated ipomoviruses are the economically most important viruses resulting in monetary loss exceeding US\$ 1 billion per year. Here in this article we intend to describe and discuss about the ipomoviruses associated with the cassava brown streak disease (CBSB) in Africa. Although the first report of CBSB was in 1930s from the coastal region of Tanzania, it did not receive much attention since it was restricted to the coastal lowlands of east Africa and appeared sporadically. However, post 2004, CBSB has been spreading to several parts of Africa putting the food security of millions of Africans at stake.

Aetiology, Host Range and Transmission

CBSB is known to affect cassava alone, while it has also shown its presence in the wild relative of cassava *Manihot glaziovii*. CBSB does not have any known alternate hosts and it is not known to infect weed species. However, under artificial inoculation

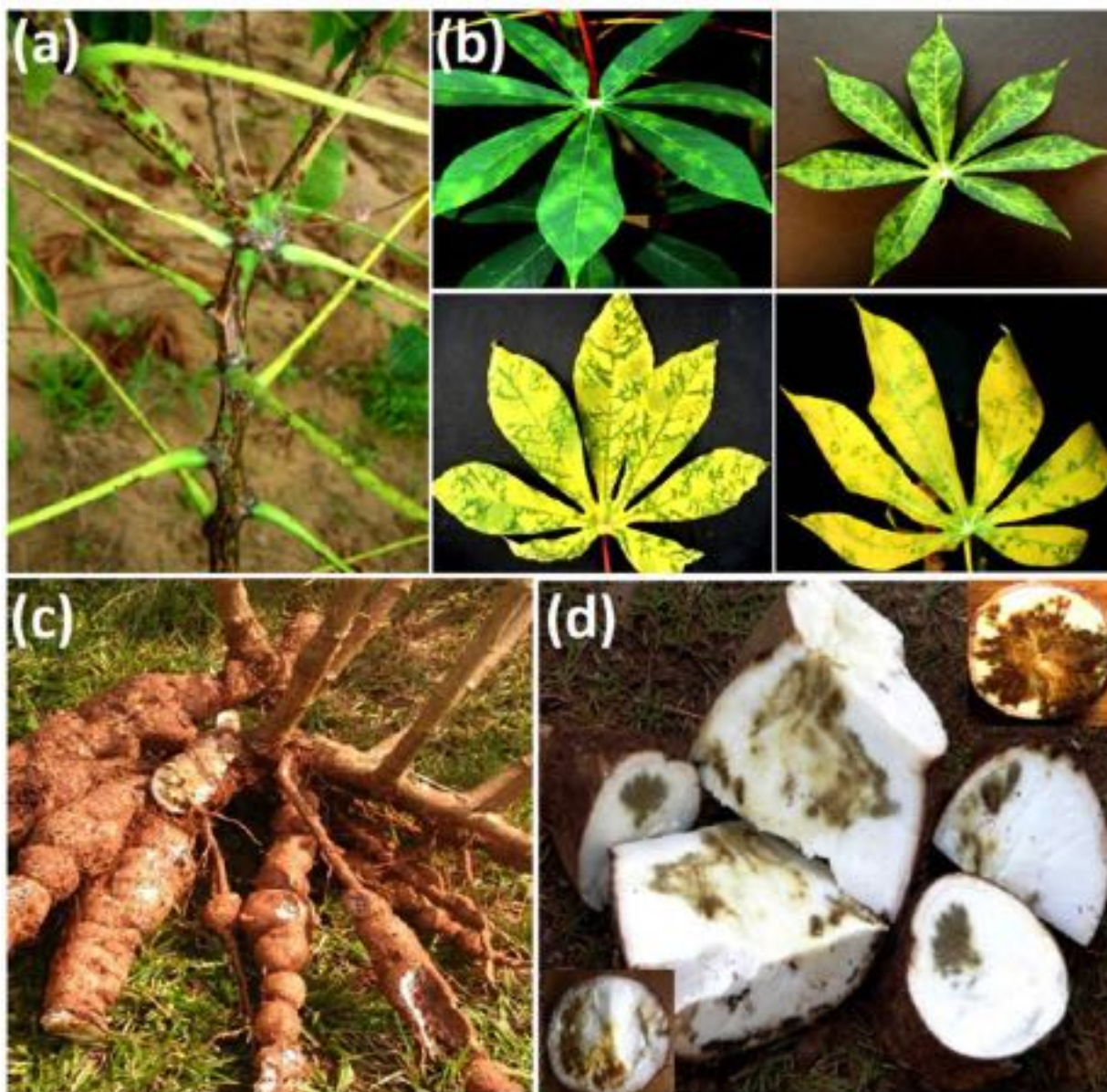


Fig. 1 Symptoms of CBSD on Root and Leaves of cassava. (a). Brown streaks on the young stem, (b) Different types of feathery chlorosis on leaves, (c) Constrictions on the root surface, (d) Necrosis in the roots. Adapted with minor modification from Patil, B.L., Legg, J.P., Kanju, E., Fauquet, C.M., 2015. Cassava brown streak disease: A threat to food security in Africa. *Journal of General Virology* 96, 956–968.

conditions, it can infect some of the herbaceous hosts and the model host *Nicotiana benthamiana*, widely used in virology studies. Two diverse and distinct ipomovirus species *Ugandan cassava brown streak virus* (UCBSV) and *Cassava brown streak virus* (CBSV) are known to cause CBSD, either as independent infection or sometimes as mixed infections. In this article both UCBSV and CBSV isolates are sometimes collectively referred as CBSVs. The most common symptoms on the foliage is feathery chlorosis along the veins of the oldest leaves, occasionally with circular patches of chlorosis in between the primary veins (Fig. 1(b)), while the stem shows characteristic brown necrotic streaks and stem die-back in severe cases (Fig. 1(a)). Young leaves do not show symptoms. However, CBSVs cause major damage in the tuberous roots, the edible part of cassava. CBSD affected tuberous roots show brown corky necrosis in the starchy tissue, resulting in reduction of starch and cyanide content, occasionally the tuberous roots develop radial constrictions (Fig. 1(c)). However, the symptoms can vary depending on the virus isolate involved, cassava genotype and its age, and the prevailing environmental conditions. In one of the studies involving mechanical inoculation of diverse CBSV/UCBSV isolates showed production of variable symptoms in both the natural host cassava and the model host *N. benthamiana*.

Comparison of symptom severity of CBSV and UCBSV indicated CBSV to be more virulent, causing necrosis in *N. benthamiana*, while sap inoculation of UCBSV induced mosaics and rugosity. Similar comparative studies in cassava by grafting of infected

scions showed CBSV to be more virulent than UCBSV, and the CBSV infected stem cuttings significantly reduced the cassava sprouting, as a result of higher virus accumulation in case of CBSV infection compared to infection by UCBSV. Agro-infiltration of infectious clones of cassava mosaic geminiviruses (CMGs) followed by sap inoculation of UCBSV indicated a synergistic interaction between the geminivirus and ipomoviruses in the case of *N. benthamiana*, however there are no reports of synergism in field situation.

Initially, *Bemisia afer* (Priesner-Hosny) was presumed to be the vector for CBSD transmission, since they were predominantly present on lower leaves of cassava, where the CBSD symptoms are prominent. Later, experiments showed that the whitefly (*Bemisia tabaci* (Gennadius)) was the vector transmitting CBSVs. It was capable of transmitting this ipomovirus in a semi-persistent manner unlike persistent transmission of the geminiviruses. The transmission studies done in the lab conditions showed that the CBSVs were acquired in 5–10 min and retained by the whiteflies for less than 48 h. A maximum transmission of 60% could be achieved by using 20–25 viruliferous whiteflies per plant and the virus spread in the field was maximum up to a distance of 17 meters per cropping season. This study clearly indicates that the whitefly vectors could transmit CBSVs semi-persistently and only for a short distance, while the long-distance transmission of CBSD was mostly through transportation and use of infected planting material.

Effective virus diagnostics is of utmost importance for accurate disease forecasting and management of any disease outbreaks. Early and accurate detection of CBSVs in infected cassava is a great challenge. There are several molecular methods for diagnostics of CBSD, of which ELISA and its modified versions are common. Using ELISA based techniques it is difficult to differentiate between CBSV and UCBSV. However, PCR based techniques such as RT-PCR, multiplex RT-PCR, Real Time PCR and loop-mediated isothermal amplification are developed to detect and differentiate CBSVs and UCBSVs. Sequences of several unknown viruses including CBSVs have been unraveled by next generation sequencing (NGS) strategies.

Current Status of CBSD in Africa

The first report of CBSD was from cassava cultivated in the Usambara mountains of Tanzania. Subsequent surveys showed that CBSD was prevalent in almost the entire coastal lowlands of East Africa, stretching from north-eastern border of Kenya, to the Tanzanian border with Mozambique, and lower altitudes of Malawi. CBSD incidence was not recorded in inlands above an altitude of 1000 m above sea level (a.s.l) and hence for a long time CBSD was considered as a low altitude disease. However, subsequent studies in 2004 have shown that CBSD was more widespread and also present at altitudes beyond 1000 m a.s.l., such as Uganda, western Kenya, north-western Tanzania, Burundi and the Democratic Republic of Congo.

A countrywide systematic survey for incidence of CBSD in Tanzania was done in 1994 and the maximum incidence was recorded from southern lowland coastal districts of Mtwara (36%) and Masasi (25%), while CBSD was completely absent in higher altitudes of north-western Tanzania. Contrastingly, at the same time CBSD symptoms were reported in cassava cultivated at altitudes 1200 m above sea level, in a location near Entebbe in central/southern Uganda and also from Tabora in north-western Tanzania. Except for these reports of occurrence of CBSD at higher altitudes the notion that CBSD is a low altitude disease remained unchanged until the time a major report on occurrence of CBSD was made from central/southern Uganda in 2004.

An unpublished estimate by NARO (Uganda) put the overall incidence of CBSD at 16% in 2008 and 29% in 2009, clearly depicting significant spread of CBSD in East Africa. Similar such reports on spread of CBSD were made in western Kenya and Lake Victoria zone of Tanzania. Surveys for CBSD were done in 19 districts of Tanzania and revealed a drastic increase in CBSD incidence in north-western regions of Tanzania.

In recent years, additional reports have been made from Rwanda, Burundi and the eastern Democratic Republic of Congo (DRC). In early 2000, symptoms similar to CBSD have also been reported from cassava cultivated in Bas Congo Province in western DRC, Mulanje Province in central Angola and parts of Madagascar. However, none of these reports has been confirmed for the presence of CBSV by using appropriate diagnostic tools. All of these surveys were primarily based on visual assessments of leaf symptoms, and they may be underestimates of the true level of CBSV infection, since under the unfavourable weather conditions the CBSD leaf symptoms may not be expressed.

Genome Organization and Gene Functions

The viruses involved in CBSD belong to the family *Potyviridae* and the genus *Ipomovirus*. The family *Potyviridae* is one of the largest families of plant viruses, consisting of six genera, classified based on their genome organization, sequence similarity and insect vectors involved in their transmission. Infected tissues with ipomoviruses show characteristic pinwheel-like structures or cylindrical inclusions within the phloem tissue, typical of potyviruses. Similar to other viruses of *Potyviridae* family, both the ipomoviruses CBSV and UCBSV have a positive-sense single stranded RNA genome with an average length of 9 Kb. All the viruses of the *Potyviridae* family except for the members of the genus *Bymovirus* have monopartite genome, which is translated into a single polyprotein which is later proteolytically autocleaved into ~10 mature proteins by three distinct proteases expressed by the virus itself.

Since the first report of CBSD in 1935, the first complete genome sequence of an ipomovirus involved in CBSD was published in 2009. This isolate from Tanzania sampled in 2007 was an Ugandan cassava brown streak virus isolate abbreviated as UCBSV-[TZ:MLB3:07] with a sequence length of 9069 nt, which was shorter than other ipomoviruses. This UCBSV genome has 134 nt and 226 nt untranslated regions (UTRs) on both of its ends referred as 5'UTR and 3'UTR respectively (Fig. 2(b)). These UTRs

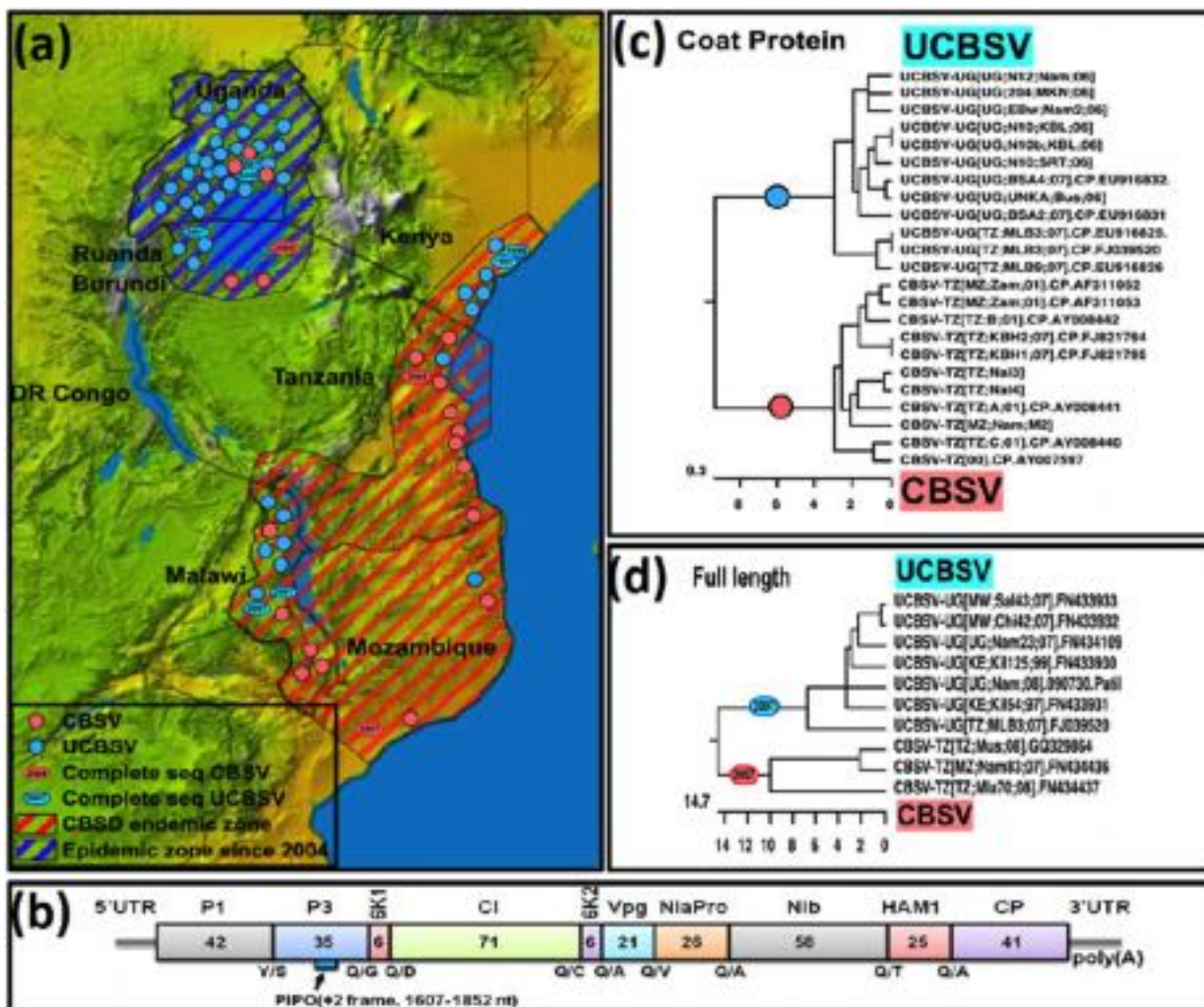


Fig. 2 (a) Depiction of endemic (red-striped) and epidemic zones (blue-striped) of CBSV and UCBSV in East Africa, (b) A representative picture showing the genome organization of CBSV/UCBSV encoding for 10 proteins (P1, first serine proteinase; P3, third protein; 6K1 and 6K2, two 6 kDa proteins; CI, cylindrical inclusion protein; VPg, viral protein genome-linked; NIa-Pro, main viral proteinase; NIb, replicase; HAM1, Maf/HAM1 pyrophosphatase homologue; CP, coat protein) and 3' and 5'UTRs, the poly(A) tail and PIPO translated by +2 frame shift within the P3. Two amino acids at the proteolytic cleavage sites of the polyprotein are shown below the polyprotein, and the estimated molecular mass (in kDa) for each protein is given in the box, (c, d) Phylogenetic analysis of selected CP sequences and the complete genome sequences of CBSV/UCBSV isolates. GenBank accession numbers are indicated and the scale bar represents the pairwise distance expressed as percentage dissimilarity. Adapted from Patil, B.L., Legg, J.P., Kanju, E., Fauquet, C.M., 2015. Cassava brown streak disease: A threat to food security in Africa. *Journal of General Virology* 96, 956–968.

encompass the regulatory sequences involved in viral gene/protein expression. Later a second distinct ipomovirus species named as Cassava brown streak virus (CBSV), with a sequence length of ~9008 nt, with similar genome organization as UCBSV, was also found to be involved in CBSV.

Unlike other members of the *Potyviridae* family, both the above ipomoviruses infecting cassava encode for a single P1 serine proteinase and are devoid of helper component cysteine proteinase (HCPro) (Fig. 2(b)). Such a high level of diversity in the P1 gene is characteristic of viruses of the *Potyviridae* family. The P1 of CBSVs are most similar those of another ipomovirus species called *Sweet potato mild mottle virus* (SPMMV). However, in contrast to CBSVs, the SPMMVs possess the HCPro, indicating it to be an evolutionary link between ipomoviruses and potyviruses. In contrast to both CBSVs and SPMMVs, another two ipomovirus species, namely *Cucumber vein yellowing virus* (CVYV) and *Squash vein yellowing virus* (SqVYV) possess two P1 serine proteinases (P1a and P1b). Silencing suppression studies done by individually expressing the CBSV genes has demonstrated P1 to be a silencing suppressor for CBSVs. The P1 of CBSVs encompass the basic LxKA and Zn-finger motifs, and the LxKA motif of P1 is diverse within different ipomoviruses, with exchanges of lysine (K) and arginine (R) at positions 2 and 3, and in the case of CBSVs, the P1 protein contains LRRs. Whereas in case of SPMMV although the P1 protein has the silencing suppression activity, the

HCPPro is also known to extend the durability of the silencing suppression. Whereas in the case of ipomoviruses lacking HCPPro, the P1b protein complements for the silencing suppression activity.

The other eight ORFs namely, P3 (the third protein), 6K1 and 6K2 which code for the two 6 kDa proteins, the cylindrical inclusion protein (CI), viral protein genome-linked (VPg), the main viral proteinase (NIa-Pro), the replicase (NIb) and the coat protein (CP) of CBSVs are more similar to the other ipomovirus sequences such as CVYV (*Cucumber vein yellowing virus*), SqVYV (*Squash vein yellowing virus*) and SPMMV. A recently reported potyviral protein P3N-PIPO expressing by 2 nucleotide frameshift within the P3 ORF is also found in CBSVs. Both CI and P3N-PIPO complex coordinate the intercellular movement of potyviruses by formation of plasmodesmata-associated structures.

In addition to the above described proteins, there are nine ORFs in CBSVs which are characteristic of the *Potyviridae* family. In addition, they contain an unusual ORF called as HAM1, encoding a polypeptide of 226 aa, flanked by proteolytic cleavage sites, and it is placed in between the replicase (NIb) and the CP. This additional ORF had homology with the Maf/HAM1 superfamily of proteins reported in several prokaryotes and eukaryotes, which are nucleoside triphosphate (NTP) pyrophosphatases. These proteins are known to reduce mutations by intercepting non-canonical forms of NTPs and thus prevent their incorporation into nucleic acids, which otherwise may result in unfavourable mutations. *Euphorbia ringspot virus* belonging to the *Potyvirus* genus is the only other viral species reported to contain an HAM1-like sequence. The presence of such an anti-mutator gene may have a beneficial role under oxidative stress conditions, during which mutations occur rampantly. Such a scenario may exist in the plants belonging to the Euphorbiaceae family, particularly the older leaves with higher levels of senescence where CBSVs are shown to accumulate. *Blackberry virus Y*, another species of the *Potyviridae* family is also known to contain such unusual heterologous sequences encompassing the AlkB domains within their P1 proteinase, which may also counter the occurrence of deleterious mutation. The presence of such AlkB domains are also known to be present in some of the viruses of the families *Flexiviridae* and *Closteroviridae*.

However, a more recent investigation about the properties of Ham1 protein of CBSVs by *in vitro* and *in vivo* overexpression has indicated Ham1 to be a necrosis determinant. *In vitro* enzyme assays showed that the Ham1 proteins of both CBSV and UCBSV had ITPase activities. However, deep sequencing experiments did not show any evidence of Ham1 mutagenic protection during infection in *N. benthamiana*. Deletion and mutation studies with an infectious clone for an isolate of CBSV from Tanzania showed that knock out of Ham1 in CBSV failed to produce necrotic symptoms in the model host *N. benthamiana*. Thus it was proposed that the Ham1 of CBSVs with highly conserved ITPase motifs may have a highly selectable function during infections in cassava and may represent a euphorbia host adaptation that could be targeted in antiviral strategies.

Evolution, Diversity and Distribution of Viruses Involved in CBSD

Phylogenetic analysis of the CBSV/UCBSV CP sequences from East Africa resulted in the formation of two distinct clusters, with all the CBSV CP sequences clustering together and the UCBSV CP sequences forming another distinct cluster (Fig. 2(c)). To date, 27 complete genome sequences of CBSVs are available in the NCBI database, and of these sequences, 11 are UCBSV isolates and 16 are isolates of CBSV. These sequences are isolated from 6 East African countries, 5 from Uganda, 14 from Tanzania, 3 from Mozambique, 3 from Comoros Islands, and one each from Kenya and Malawi (Fig. 2(a)). All the sequences are mostly of 9 Kb length, with minor variation in their sizes, the smallest one reported is 8748 nt long and the longest one is 9097 nt. Phylogenetic analysis of the full length sequences of CBSV/UCBSV also result in the formation of two main clusters corresponding to CBSV and UCBSV isolates and these major clades are further subdivided into minor clades. The cluster with UCBSV isolates has more sub-clusters than the CBSV species, though the genomes of CBSV isolates are genetically more diverse than those of UCBSV and the evolutionary rate of CBSV is considered to be much higher than UCBSV. Recent studies indicate that the accelerated rate of evolution is mainly contributed by *NIa*, *6K2*, *NIb* and *P1* sequences. *P1* was the highest contributor to diversity of CBSVs and phylogenetic analysis of N-terminal sequences of *P1* gene alone with advanced bioinformatics tools revealed the presence of a third distinct clade tentatively called as CBSV-Tanzania (CBSV-TZ).

Sequence comparison of the complete genomes of CBSV and UCBSV isolates revealed an identity range of 69.0%–70.3% at the nucleotide level and 73.6%–74.4% at the polyprotein amino acid sequence levels. Thus, in view of the sequence divergence between these two groups, two distinct virus species are recognized, CBSV and UCBSV (Fig. 2(d)). While the nucleotide sequence identities among the UCBSV isolates ranged from 87% to 99% and among the CBSV isolates it was 79%–95%. Recombination analysis using RDP4 for all the CBSV and UCBSV sequences indicated 11 recombination events within CBSV isolates and 3 in UCBSV. There was at least one recombination in each of the ten genes, with maximum recombinations (5) detected in CP followed by CI. However, till date no recombinations have been detected between the CBSV and UCBSV isolates, which is consistent with the previous reports on absence of recombination between diverse genomes of RNA viruses due to reduced fitness in the recombinants. The size of the gene sequences that encoded for CI, VPg and CP proteins was variable among isolates of CBSV and UCBSV.

The amino acid sequence identity for HAM1 sequences between CBSV and UCBSV was 55%, which was lowest among all gene sequences of the viruses. Such a low identity may indicate that both the Ham1 sequences may have been acquired from two different hosts at two different time points or they may have evolved rapidly than the other genes as evidenced by adaptive selection pressure on HAM1 of both CBSV and UCBSV. Sequence analysis revealed that there are 33 highly conserved amino acid

residues across HAM1 homologues from different organisms, indicating a strong negative selection on HAM1 and CP genes. The CP of CBSV was 9 aa longer than the CP of UCBSV and the size of P1 protein was also highly variable, with only ~60% aa identity between isolates of the two species. These size differences in the CP sequences may have an implication on their differential transmission by the whitefly vector. Although the transmission of CBSVs by whiteflies is well established under experimental conditions the differential transmission efficiency by whiteflies between the two viruses has not been significant. A recent report on detailed sequence analysis for several CBSV isolates revealed the presence of the DAG motif in their CP genes, however such a motif was not recorded in the UCBSV CP sequences. DAG sequences are present in the HCPro encoded by all the potyviruses which is an important determinant for aphid transmission. Hence it needs to be experimentally verified if aphids are capable of transmitting the CBSVs or if it is most probably an evolutionary mark in the CP of the ipomoviruses. The diversity in the P1 sequences may result in variation in their ability to suppress RNA silencing and which may have a correlation with the virulence of CBSV and UCBSV isolates.

Although in the fields there are no apparent symptom differences among the isolates of CBSV and UCBSV, under controlled conditions distinct phenotypes are produced by CBSV and UCBSV isolates. Mixed infections are found in regions where both CBSV and UCBSV are found. However, there is no evidence of synergistic interaction between the isolates of CBSV and UCBSV, unlike the interaction between different species of cassava infecting geminiviruses.

A RT-PCR based study in 2014, in eight major districts of Rwanda on distribution of CBSVs revealed that single infections of cassava by UCBSV were most common (74.2%) and were associated with wide range of symptom phenotypes. However, single infections by CBSV were predominant (15.3%) in plants showing CBSD symptoms on stem alone, while mixed infections of CBSV and UCBSV were in 10.5% of the total number of plants infected and were predominantly present in combinations of leaf + stem + root phenotypes. Further, phylogenetic analysis of a partial sequence of CP of all the isolates in this study indicated that in Rwanda there is only one type of CBSV while diverse types of UCBSVs are involved in CBSD.

Management and Control of CBSD

Phytosanitary practices are most important for virus free cultivation of cassava, because of its high vulnerability to virus transmission. In this regard the group led by Dr. James Legg based in the CGIAR institute IITA (Tanzania) is playing a major role in containing CBSD spread. CBSD is transmitted by the whitefly vectors for a shorter distance since the mode of transmission of these ipomoviruses is semi-persistent in nature. While long distance transmission of CBSD can only happen through long distance movement of vegetative propagules. Some of the important steps for control of CBSD are production of clean or disease free planting material, virus indexing of parent material and the pre-basic germplasm propagation, and regular roguing of diseased plants. Further, cooperation among group of farmers at a community level is important for a collective action and implementation of phytosanitary measures. Currently, such large scale initiatives are being implemented in Eastern and Southern Africa, and efforts are made to contain the regional spread of CBSD. Further, national and regional quarantine measures are enforced to prevent spread of CBSD during exchange of cassava germplasm.

Sources of CBSD Resistance and Their Introgression

Crop improvement through resistance breeding has been the most successful approach for control of crop diseases. Although, breeding for virus resistance in cassava was initiated in 1935 at Amani, Tanzania; it has been a major challenge since cassava is cross pollinated, with high level of heterozygosity. In this breeding programme, 46106/27 was the most CBSD resistant cassava variety developed by backcrossing *M. glaziovii* with *M. esculenta*. This is one of the most successful products of the research programme at Amani, presently cultivated by the farmers and this CBSD resistance still persists in the farmers' fields in Kenya, which locally is called as Kaleso.

Hundreds of SNP markers have been developed for CBSD tolerance and a linkage map has also been developed which is used for detection of CBSD tolerant quantitative trait loci (QTLs). Mapping populations developed by crossing CBSD-tolerant cultivar, Namikonga and a susceptible cultivar, Albert are being used for identification of QTLs associated with CBSD resistance. Now that cassava genome sequence is published, it could be used for identifying genes that control CBSD resistance and for development of new markers associated with CBSD resistance. Next generation sequencing technologies are also being employed for identification of genes associated with CBSD tolerance or the genes that respond to CBSD infection. Gene expression studies, using cassava varieties with differential resistance to CBSD revealed that at least 700 genes were overexpressing in the CBSD resistant variety Kaleso when compared to the susceptible variety Albert. Although none of these differentially expressed genes resembled a known disease resistance gene, many of them encoded enzymes or factors involved in hormone signalling or secondary metabolites pathways, that may be associated with disease resistance. However, none of the cassava genotypes/hybrids identified are immune to CBSD, but majority of them are tolerant since they show clear foliar symptoms, but necrosis in the roots is absent or delayed.

In one of the recent studies, 238 cassava germplasm accessions from South America were screened for CBSD resistance. There were 7 South American accessions that were completely free of CBSD symptoms and were negative for RT-PCR detection. The breeding lines DSC167 and DSC188 showed broad spectrum resistance to diverse isolates of CBSVs. This clearly reflects that

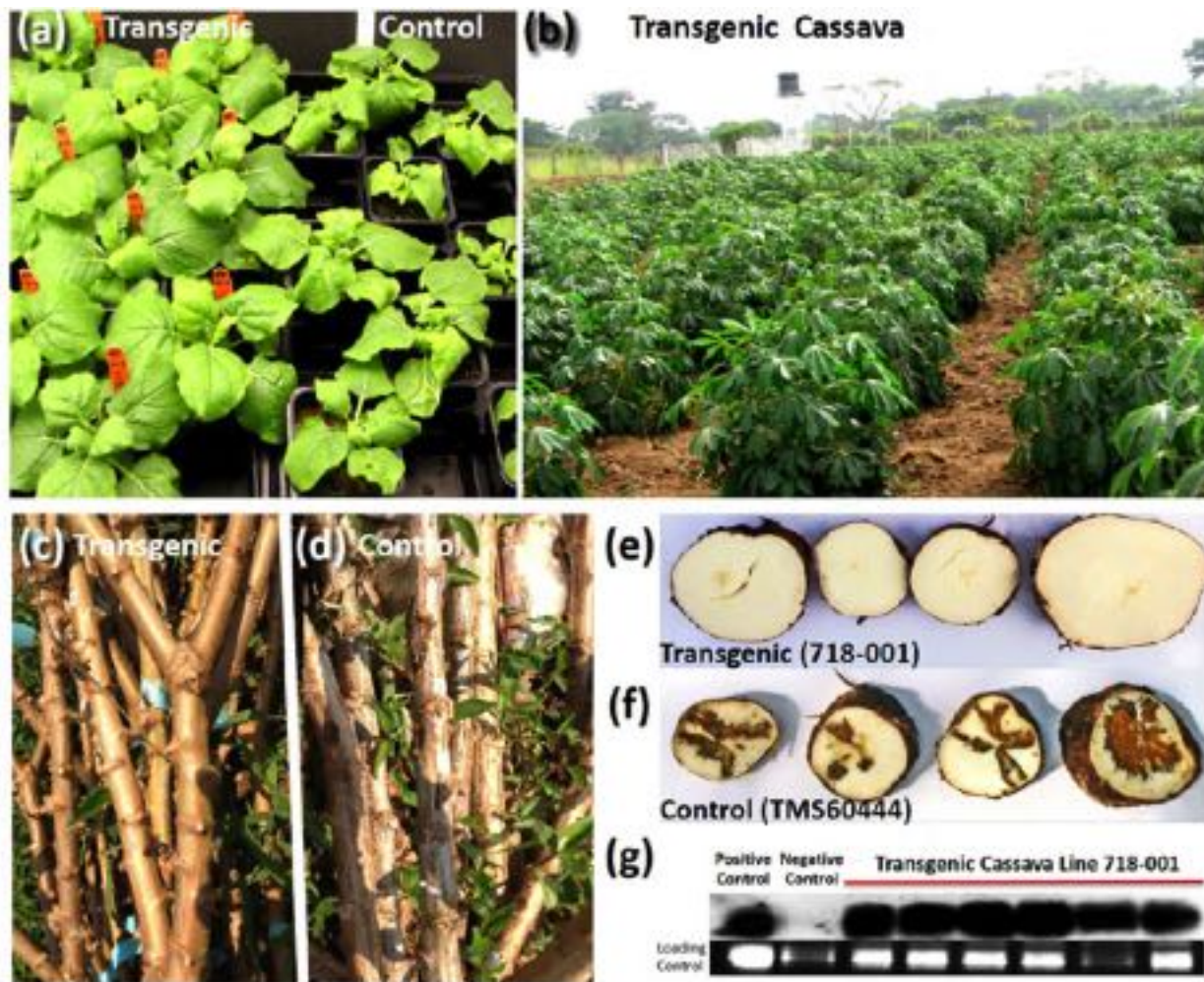


Fig. 3 (a) Screening of RNAi-based transgenic *N. benthamiana* for resistance against CBSVs (b) Confined field trials for RNAi-based transgenic cassava lines in Uganda, (c, d) Brown necrotic lesions seen on the stems of CBSV-infected control cassava plants, compared with non-infected transgenic cassava, (e, f) Cross section of cassava roots showing CBD symptom distribution in non-transgenic vs transgenic cassava plants, (g) accumulation of small interfering (si)-RNA in transgenic cassava as revealed by Northern hybridization technique. Adapted from Patil, B.L., Legg, J.P., Kanju, E., Fauquet, C.M., 2015. Cassava brown streak disease: A threat to food security in Africa. *Journal of General Virology* 96, 956–968.

the germplasm collections are potential reservoirs of several disease resistance genes and their conservation is critically important for resistance breeding programmes.

A project led by the International Institute of Tropical Agriculture (IITA) aims to ensure that the farmers in Africa have access to diverse disease-free improved cassava varieties with resistance to both CBD and CMD, and with consumer preferred traits. Studies have revealed that both additive and non-additive genetic effects are required for expression of CBD resistance trait, and of several crosses studied, the cassava variety Kaleso (Namikonga) had the best general combining ability for CBD resistance.

Engineering CBD Resistance

Although there are sources of resistance against CBD, it could be impossible to introgress CBD resistance while retaining the original agronomic/quality traits of cassava. Furthermore, the resistance reported against CBD is more of tolerance to virus infection, rather than immunity against the CBSVs. Hence, genetic engineering is an important option for controlling plant viruses and CBSVs in particular. There are several strategies used for control of viruses that infect plants, which will not be elaborated here, since it is an extensively reviewed topic. The first successful control of CBSV was demonstrated using RNA-interference (RNAi) strategy (Fig. 3), result confirmed later by other groups.

Similarly, CBD resistance has been accomplished by over-expression of artificial micro RNAs that targeted the CBSVs, however this work was restricted to the model host *N. benthamiana*. An interaction between the viral genome-linked protein (VPg) and host eukaryotic translation initiation factor 4E (eIF4E) isoforms is pre-requisite for a potyvirus to establish infection in a host cell.

Recently, the CRISPR/Cas9-mediated genome editing was employed to alter the sequence of cassava eukaryotic translation initiation factor 4E (eIF4E) isoforms, namely ncbp-1 and ncbp-2. When the ncbp-1/ncbp-2 cassava mutants were challenged with CBSV, there was a delay in symptom development, as well as reduced severity and incidence of storage root necrosis, which corresponded with reduced virus titer in storage roots. However, despite all these very positive results showing the possible control of CBSV through biotechnology, none of the putative commercial products have been distributed to farmers so far, due to the lack of a biotechnology official framework in East African countries to permit its use in safe conditions.

Cross Reading With Other Chapters

Cassava Mosaic Viruses (*Geminiviridae*); Potyviruses (*Potyviridae*).

See also: Cassava Mosaic Viruses (*Geminiviridae*). Potyviruses (*Potyviridae*)

Further Reading

- Anjanappa, R.B., Mehta, D., Okoniewski, M.J., *et al.*, 2018. Molecular insights into Cassava brown streak virus susceptibility and resistance by profiling of the early host response. *Molecular Plant Pathology* 19, 476–489.
- Beyene, G., Chauhan, R.D., Ilyas, M., *et al.*, 2017. A virus-derived stacked RNAi Construct confers robust resistance to Cassava brown streak disease. *Frontiers in Plant Science* 7, 2052.
- Gomez, M.A., Lin, Z.D., Moll, T., *et al.*, 2019. Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnology Journal* 17 (2), 421–434.
- Legg, J.P., Lava Kumar, P., Makesh Kumar, T., *et al.*, 2015. Cassava virus diseases: biology, epidemiology, and management. *Advances in Virus Research* 91, 85–142.
- Legg, J., Ndalahwa, M., Yabeja, J., *et al.*, 2017. Community phytosanitation to manage cassava brown streak disease. *Virus Research* 241, 236–253.
- Mbanzibwa, D.R., Tian, Y., Mukasa, S.B., Valkonen, J.P., 2009. Cassava brown streak virus (*Potyviridae*) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *Journal of Virology* 83, 6934–6940.
- Mbanzibwa, D.R., Tian, Y.P., Tugume, A.K., *et al.*, 2011. Evolution of cassava brown streak disease-associated viruses. *Journal of General Virology* 92, 974–987.
- Patil, B.L., Legg, J.P., Kanju, E., Fauquet, C.M., 2015. Cassava brown streak disease: A threat to food security in Africa. *Journal of General Virology* 96, 956–968.
- Patil, B.L., Ogwok, E., Wagaba, H., *et al.*, 2011. RNAi-mediated resistance to diverse isolates belonging to two virus species involved in Cassava brown streak disease. *Molecular Plant Pathology* 12, 31–41.
- Rwegasira, G.M., Momanyi, G., Rey, M.E., Kahwa, G., Legg, J.P., 2011. Widespread occurrence and diversity of Cassava brown streak virus (*Potyviridae: Ipomovirus*) in Tanzania. *Phytopathology* 101, 1159–1167.
- Taylor, N.J., Halsey, M., Gaitán-Solis, E., *et al.*, 2012. The VIRCA project: Virus resistant cassava for Africa. *GM Crops & Food: Biotechnology in Agriculture and the Food Chain* 3, 93–103.
- Tomlinson, K.R., Bailey, A.M., Alicai, T., Seal, S., Foster, G.D., 2018. Cassava brown streak disease: historical timeline, current knowledge and future prospects. *Molecular Plant Pathology* 19, 1282–1294.
- Winter, S., Koerbler, M., Stein, B., *et al.*, 2010. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *Journal of General Virology* 91, 1365–1372.
- Yadav, J.S., Ogwok, E., Wagaba, H., *et al.*, 2011. RNAi-mediated resistance to Cassava brown streak Uganda virus in transgenic cassava. *Molecular Plant Pathology* 12, 677–687.

Relevant Websites

- <https://www.cabi.org/isc/datasheet/17107>
Cassava brown streak viruses.
- <https://www.plantwise.org/knowledgebank/datasheet/17107>
Cassava Brown Streak Disease - Plantwise Knowledge Bank.
- <https://mel.cgiar.org/projects/424>
Consultative Group for International Agricultural Research.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/potyviridae/569/genus-ipomovirus
Genus: *Ipomovirus*.

Cassava Mosaic Viruses (*Geminiviridae*)

James Legg, International Institute of Tropical Agriculture, Dar es Salaam, Tanzania

Stephan Winter, Leibniz Institute – DSMZ – German Collection of Microorganisms and Cell Cultures GmbH, Braunschweig, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J.P. Legg, African Cassava Mosaic Disease, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00693-2](https://doi.org/10.1016/B978-012374410-4.00693-2).

Nomenclature

CMG Cassava mosaic geminivirus

CP Coat protein or capsid protein

CR Common region

dsDNA Double-stranded deoxyribonucleic acid

ELISA Enzyme-linked immuno-sorbent assay

IR Intergenic region

kb Kilobase

kbp Kilobase pair

kDa Kilo Dalton

MP Movement protein

NSP Nuclear shuttle protein

PCR Polymerase chain reaction

PTGS Post-transcriptional gene silencing

RCA Rolling-circle amplification

Ren Replication enhancer protein

siRNA Small interfering RNA

ssDNA Single-stranded deoxyribonucleic acid

TrAP Transcriptional activator protein

Glossary

Nonanucleotide A codon containing nine nucleotides.

Pandemic An outbreak of a disease over several countries or a large part of the world.

Pseudo-recombinant A viral infection involving complementary genome components from different virus species.

Taxonomy

Cassava mosaic viruses belong to the genus *Begomovirus* in the family *Geminiviridae* which comprises viruses with single-stranded circular DNA genomes that are transmitted by *Bemisia tabaci* (Gennadius) whiteflies. The cassava-infecting begomoviruses have two DNA components (DNA-A and DNA-B) of which the DNA-A genome sequences are used for taxonomic assignment of virus species. The species demarcation threshold is at 91% sequence identity, thus virus genomes having less than 91% of nucleotides on DNA-A that are identical to those of their closest neighbor are assigned to a unique species. Eleven cassava mosaic geminiviruses (CMGs) are currently recognized as distinct virus species by the ICTV (See Relevant Websites Section). These are: *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), *South African cassava mosaic virus* (SACMV), *Indian cassava mosaic virus* (ICMV), *Sri Lankan cassava mosaic virus* (SLCMV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Zanzibar virus* (EACMZV), *East African cassava mosaic Kenya virus* (EACMKV), *Cassava mosaic Madagascar virus* (CMMGV), and *African cassava mosaic Burkina Faso virus* (ACMBFV) (**Table 1**; **Fig. 1**). Several of these originated from recombination events that generated genomic diversities beyond species boundaries. The names given reflect the relatedness to their putative parental virus.

Intraspecific Diversity

New, recombinant cassava viruses described provide evidence for on-going diversification of CMG genomes. Interspecific recombination of DNA-A and DNA-B genome components of cassava viruses also include sequences from other begomoviruses. For example, genome portions of the monopartite Tomato leaf curl virus are present in ACMBFV. Similarly, the genome components of CMMGV include, besides elements from other CMGs, sequences from non-related monopartite begomoviruses infecting tomato in the Indian Ocean islands. Additionally, there is ample evidence for molecular diversity of CMG genomes from complete DNA-A and DNA-B genome sequences of virus isolates available in GenBank. Nevertheless, in spite of contrasting molecular characteristics, there is little evidence for unique biological features of each of the virus species in terms of their symptom phenotypes, natural host range, transmission parameters or the association of particular virus isolates with *B. tabaci* species that would provide additional information for the discrimination of virus strains or the identification of a particular virus species. The 11 acknowledged CMG species and the numerous isolates that are present in Africa and Asia are merely characterized by distinctive sequences of their DNA-A genome components.

Table 1 Cassava mosaic geminivirus species

<i>Virus</i>	<i>Reference</i>	<i>Genome</i>	<i>Occurrence</i>
African cassava mosaic virus	Bock and Woods (1983)	NC_001467 NC_001468	Africa
East African cassava mosaic virus	Swanson and Harrison (1994)	NC_004674 NC_004676	East Africa
Indian cassava mosaic virus	Hong <i>et al.</i> (1993)	NC_001932 NC_001933	India
Sri Lankan cassava mosaic virus	Saunders <i>et al.</i> (2002)	NC_003861 NC_003862	India, Sri Lanka, Vietnam, Cambodia, China
South African cassava mosaic virus	Berrie <i>et al.</i> (1998)	NC_003803 NC_003804	Southern Africa
East African cassava mosaic Cameroon virus	Fondong <i>et al.</i> (1998)	NC_004625 NC_004630	East and West Africa
East African cassava mosaic Malawi virus	Zhou <i>et al.</i> (1998)	NC_022645 NC_022644	East and southern Africa
East African cassava mosaic Zanzibar virus	Maruthi <i>et al.</i> (2004)	NC_004655 NC_004656	East Africa
East African cassava mosaic Kenya virus	Bull <i>et al.</i> (2006)	NC_011583 NC_001584	East Africa
Cassava mosaic Madagascar virus	Harimalala <i>et al.</i> (2012)	NC_017004 NC_017005	Madagascar
African cassava mosaic Burkina Faso virus	Tiendrebeogo <i>et al.</i> (2012)	HE616777 HE616778	West Africa

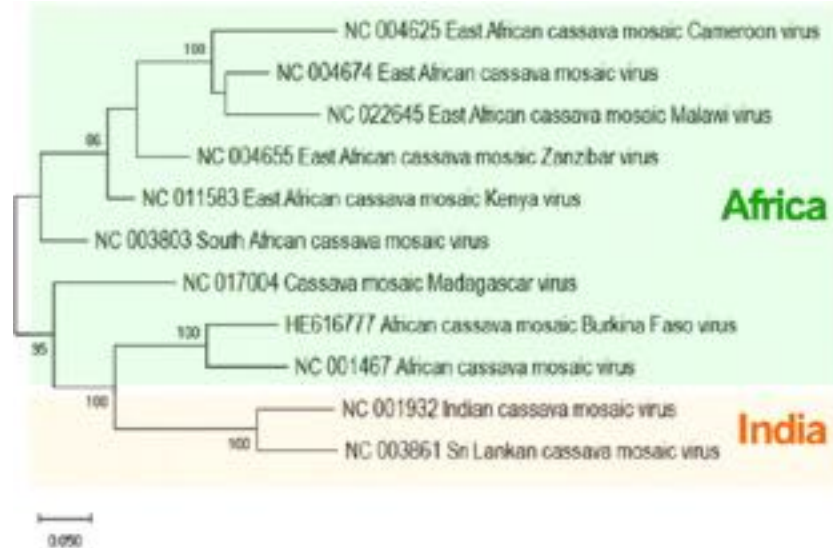


Fig. 1 Estimated phylogeny of DNA-A genome sequences of cassava mosaic geminivirus species within the genus *Begomovirus*, family *Geminiviridae*. The evolutionary history was inferred using the Maximum Likelihood method with the percentage of trees in which the associated taxa clustered together shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The evolutionary analyses were conducted in MEGA X.

History

Cassava (*Manihot esculenta* Crantz) is a root crop that is grown widely throughout the tropics, primarily for its value as a starchy staple food. From its origins in Latin America, cultivated cassava was introduced to Africa in the sixteenth century, and subsequently spread through much of Africa south of the Sahara. The first cultivation of cassava in Asia occurred in south India in the

19th century, before it spread more widely throughout South-East Asia and eventually as far as Southern China. Although initially used as a food crop, most cassava in Asia is now cultivated for industrial starch production.

None of the viruses infecting cassava in Latin America is known to have been carried to either Africa or Asia, but over time, the crop became infected by indigenous viruses. The first report of a virus-like disease in African cassava was made in 1894 from what is now northeastern Tanzania. The original German descriptor, 'Krausellkrankheit', made reference to the characteristic mosaic symptoms elicited in affected plants. It was not until just over a decade later, however, that the first firm indication was given that the disease had a viral etiology. In spite of these early reports, there seems to have been little concern about the impact of cassava mosaic disease (CMD) until the 1920s. Between 1929 and 1937, however, numerous reports were made of the spread and damaging effect on cassava crops of CMD from diverse locations throughout the continent, from the island of Madagascar off the southeastern shores of the African mainland, to Sierra Leone in West Africa. These developments provided the stimulus for the earliest concerted efforts to develop approaches to controlling the viruses that caused this damaging crop disease. Although substantial progress was made in the development of cassava varieties that were resistant to cassava mosaic-causing viruses in the 1930s and 1940s, the viruses themselves remained poorly understood, and it was not until 1983 that the first definitive study confirming the viral etiology of CMD was published. Geminiate virions were shown to encapsidate a bipartite genome of single-stranded circular DNA, leading to the designation of these viruses as geminiviruses in the genus *Begomovirus*. In Asia, symptoms of cassava mosaic disease were first observed from southern India in the 1950s, and subsequently from Sri Lanka in the 1980s. Since 2015, the disease has spread more widely, affecting Cambodia and Vietnam in South-East Asia.

Virion Structure

The two single-stranded circular DNA molecules (DNA-A and DNA-B) of CMGs are encapsidated within a virion having a unique architecture composed of two incomplete icosahedral particles (hemcapsids) fused to form a twinned, geminate particle from a single type of capsid protein. For each icosahedral half, 55 CP chains assemble into 11 capsomers. The near-atomic resolution of the virion structure of ACMV has been resolved by cryo-electron microscopy to 4.2 Å revealing that alternating conformations of the N-terminus of the CP confer stability of the twinned particle and most probably mediate the interaction with the encapsidated DNA (Fig. 2). This was further confirmed by similar studies with Ageratum yellow vein virus, and resolution at an even higher level

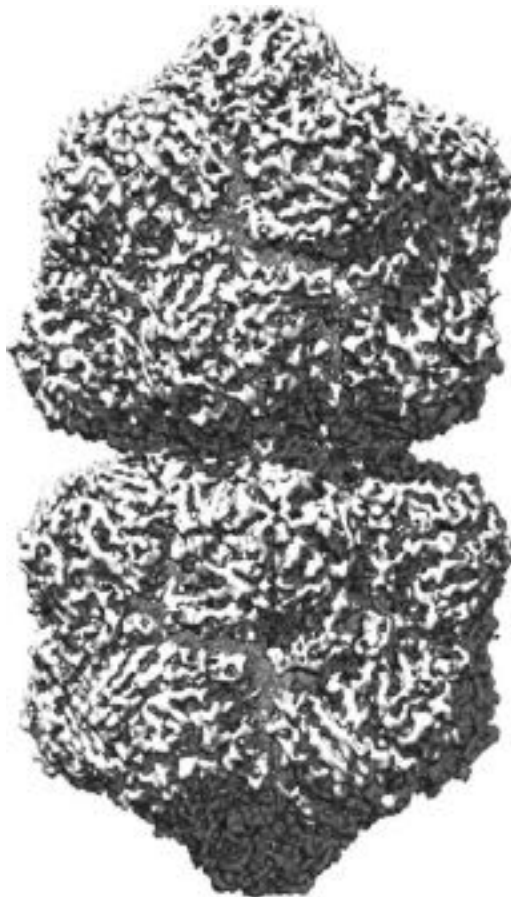


Fig. 2 Structure of African cassava mosaic virus determined by cryoelectron microscopy. Courtesy of Katharina Hipp.

(3.3 Å) showed that three distinct conformations of CP form the capsid and arrange the hemcapsid interface. The proposed model for geminivirus capsid assembly and DNA binding suggests that genomic DNA is crucial for the formation of the capsids and for directing the assembly of the capsomers and the virus particle. The high resolution model of the ACMV virion structure has also shown the position of amino acid side chains implicated in *B. tabaci* transmission of begomoviruses. The most critical, Q130 and N131, are located on the particle surface and are accessible to interact with vector insect receptors while other residues may influence particle stability and thus have indirect effects on virus transmission.

Genome

The eight genes on the DNA-A and DNA-B genome components are located in sense and complementary sense orientations on the circular covalently-closed single-stranded DNA molecules (Fig. 3). A conserved intergenic region (IR) of ca. 200 bp is shared between both genome components. This unique IR identifies DNA-B to be transreplicated by its cognate DNA-A only. Embedded in the IR is a stem loop comprising the TAATATTAC sequence that is highly conserved among almost all geminiviruses and which represents the origin of replication. While DNA-A can infect and replicate autonomously in a host plant it requires DNA-B for movement and systemic infection of the host. CMG genes are multifunctional and proteins are responsible for encapsidation and for directing the host machinery to provide cellular functions favorable for replication. In virion sense orientation on DNA-A, the structural protein gene CP (AV1) codes for the capsid protein, which is involved in whitefly transmission and transport, and which interacts with host proteins and binds to ss and double-stranded (ds)DNA. A functional CP, however, is not required for infectivity and movement. Thus, DNA-A with a defective or missing CP is still infectious and capable of systemic movement within its host, however its capacity for vector transmission is compromised. Upstream of CP, the AV2 gene also provides movement function. On the complementary-sense strand, the replication initiation protein Rep (AC1) initiates specific steps in DNA replication; TrAP (transcriptional activator protein)(AC2) and REn (replication enhancer protein)(AC3) further downstream of Rep, function as suppressors of post-transcriptional gene silencing (PTGS) and support efficient DNA replication respectively. AC4 is nested within Rep and involved in symptom expression and possibly in counter-defense. Genes on DNA-B enable virus movement within the cell and nucleus (nuclear shuttle protein, NSP)(BV1) as well as sustaining long distance spread and movement within the plant (movement protein, MP)(BC1).

Replication

Geminiviruses replicate in the nuclei of infected cells and depend on host DNA and RNA polymerases for their replication and transcription. Geminiviruses predominantly replicate in the phloem companion and parenchyma cells and some geminiviruses can also invade *mesophyll* cells but are never found in meristems. Thus, geminiviruses replicate in differentiated cells. Using host proteins the ssDNA genome is replicated by rolling circle- and recombination-mediated mechanisms in which double-stranded DNA intermediates are formed that are chromatinized with host histone proteins to form viral minichromosomes inside the nucleus. Rep recruits host factors, binds specifically to iteron sequences within IR and creates a nick in the conserved nonanucleotide TAATATTAC to initiate DNA replication. Rep also binds to a plant homolog of retinoblastoma protein to catalyze a series of intracellular reactions that move the cell from G to S phase to mobilize host factors essential for DNA replication. Double-stranded DNA (dsDNA) produced in the replication process then assembles under the control of TrAP and REn into

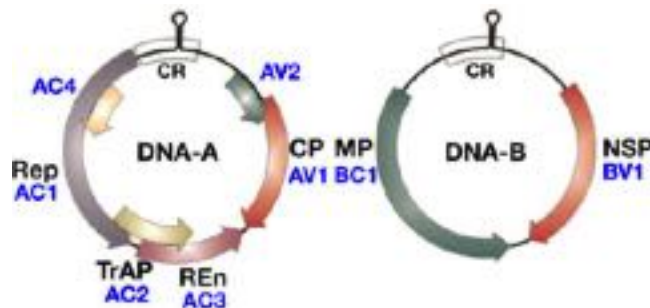


Fig. 3 Genome organization of bipartite cassava mosaic viruses. The proteins encoded by the components are indicated as follows: DNA-A component on the sense orientation encodes for the following proteins: CP or AV1; the coat protein, and the AV2 protein; and on the antisense orientation: Rep or AC1; replication associated protein, TrAP or AC2; the transcriptional activator protein, REn or AC3; the replication enhancer protein, and the AC4 protein. The DNA-B component encodes for the following proteins: NSP or BV1; the nuclear shuttle protein on the sense orientation and MP or BC1; the movement protein on the antisense orientation. The conserved region (CR) is the common region between the DNA-A and DNA-B components. For each component the conserved hairpin structure, containing the nonanucleotide sequence TAATATTAC within the loop structure, is shown at position zero.

transcriptionally active mini-chromosomes. The nuclear export is mediated by NSP, from nucleus to the cytoplasm, and movement out of the cell through plasmodesmata and long distance transport and systemic movement are facilitated by MP.

Transmission

CMGs, in common with all other members of the genus *Begomovirus*, are transmitted by the whitefly vector *B. tabaci*. Transmission can also be achieved through grafting, and relatively inefficiently through mechanical inoculation of indicator plants, but there is no seed-borne transmission. Cassava is normally propagated through the use of vegetative cuttings, and this is perhaps the most frequent source of infection in new crops under field conditions. *Bemisia tabaci* adults feed, mate, and reproduce preferentially on the upper newly-emerged leaves of cassava plants, and almost all transmission occurs here. Transmission is persistent, and ACMV has been shown to be retained by *B. tabaci* adults for up to 9 days. Transtadial, but not transovarial transmission, has been demonstrated, although the larval instars are unimportant in the epidemiology of CMGs in view of their sessile nature. Minimum periods for each of the stages of transmission for ACMV by *B. tabaci* adults are: acquisition (3 h), latent period (3 h), and inoculation (10 min). Inoculated plants begin to show symptoms of infection after 3–5 weeks. Varying levels of transmission efficiency have been reported, ranging from 0.3% to 10%, depending primarily on the nature of the CMG infection in the plants from which virus is acquired. There is evidence for a limited degree of co-evolutionary adaptation between the whitefly vector and CMGs, as Indian whiteflies are significantly better at vectoring Indian than African CMGs and vice versa. However, within Africa, there is currently no indication that whiteflies from a particular location are more efficient at vectoring locally-occurring CMGs than they are at vectoring CMGs from another part of the continent. There is, however, an important balance between the pattern of transmission and the nature of the virus infection. More severe infections, caused either by more virulent virus species/strains or by mixed virus infections, lead to a greater frequency of whitefly-borne infection but diminished propagation through cuttings by farmers. Conversely, more moderate or mild infections, caused by less virulent virus species/strains occurring in single infections, lead to less frequent whitefly-borne infection and increased propagation through cuttings by farmers. This balance dictates the epidemiological characteristics of CMG infection in any given location, area, or region.

Pathogenesis and Symptoms

Following initial infection of a previously uninfected cassava plant by a CMG, viral DNA moves through the phloem to the newly developing phloem and parenchyma cells immediately behind the meristem where rapid multiplication of virus particles takes place. Plants have developed defense responses to virus multiplication through the process of PTGS, but in response, CMG species/strains have developed various effective mechanisms to overcome this, and the degree of this effectiveness seems to be the main factor in determining the severity of disease resulting from infection. Both AC2 and AC4 have been shown to act as suppressors of PTGS. Significantly, in mixed ACMV + EACMV-like virus infections, the responses of the two viruses to host plant PTGS are complementary, leading to a synergistic interaction between the two viruses, greatly increased titers of both, and a concomitant increase in severity of the disease symptoms expressed in the plant. Molecular studies of these interactions have shown that the abundance of short interfering RNA (siRNA) molecules associated with the host plant PTGS response increases over time in pure ACMV infections, yet remains low over similar time periods for ACMV + EACMV-UG co-infections. Synergism has been described for ACMV + EACMCV mixtures in Cameroon, but its significance is greatest in the widely occurring ACMV + EACMV-UG mixed infections that cause severe CMD in many parts of East and Central Africa.

As CMG virions replicate in infected cells, chloroplasts become distorted and reduced in number leading to the appearance of a patchy mosaic of chlorotic leaf portions interspersed with normal green leaf tissue. The chlorotic mosaic patches have discrete, non-diffuse borders, but these may vary greatly in size, ranging from small portions of individual leaflets to the entire leaf (Fig. 4). In moderate to severe infections, leaves are distorted and reduced in size and the growth of the plant is stunted. The severity of symptoms is affected by several factors, including the virulence of the CMG strain, the occurrence of mixed virus infection, temperature, host susceptibility and the stage of growth of the plant. Varieties display a full range of responses to CMG infection, ranging from near immunity to extreme sensitivity resulting in plant death. The most frequent cause of severe symptoms is where ACMV and EACMV (or related viruses) occur in mixed infection. This is a consequence of the synergistic interaction between these viruses which leads to increased virus titer of both. There are currently no anecdotal or published records of synergistic interactions between ICMV and SLCMV in South Asia.

Plants infected through the cutting typically express symptoms in the first-formed leaves immediately on sprouting. Plants that sprouted without virus infection, but which were subsequently infected by the whitefly vector express symptoms only in leaves above the point of infection, although there is a latent period of 3–5 weeks between inoculation and the appearance of symptoms. These contrasting patterns of symptom expression make it possible to distinguish plants infected through the cutting from those infected during crop growth by the whitefly vector. Separating these two infection types has become an important facet of field research into the epidemiology of CMGs. Symptoms of CMD caused by ICMV and/or SLCMV in Asia vary in severity, depending on the factors described above, but the characteristics of the symptoms are the same as those of African CMD, as is the distinction between cutting-borne and whitefly-borne infection.

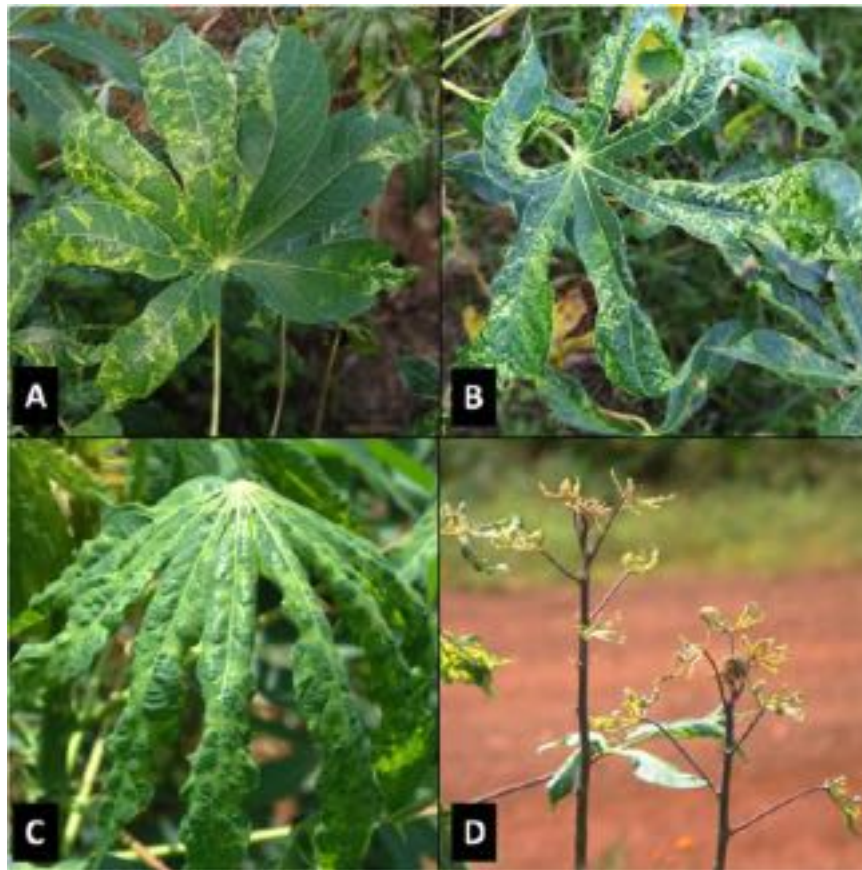


Fig. 4 Cassava mosaic disease (CMD) symptoms: (A) & (B) – Sri Lankan cassava mosaic virus in Vietnam; (C) – African cassava mosaic virus in Gabon; (D) – mixed infection of African cassava mosaic virus and East African cassava mosaic virus-Uganda in Burundi.

There are currently no reported cases of cassava varieties with immunity to infection by CMGs. However, there are many varieties that are highly resistant, and when these are infected, symptoms expressed are often confined to small portions of a few leaves. A common feature of resistant varieties is symptom recovery, as lower symptomatic leaves drop naturally and new foliage is asymptomatic. This phenomenon is linked to reversion, which is the absence of virus in a proportion of cuttings taken from an infected parent plant. This is thought to be a consequence of the incomplete systemicity of CMGs in resistant varieties.

Epidemiology

Diversity and Distribution

Distribution of the African CMGs

The earliest studies of CMG diversity and distribution used serological techniques to characterize variability, and then utilized that variability to develop diagnostic tests based on the use of monoclonal antibodies in the enzyme-linked immunosorbent assay (ELISA). Using these methods, two principal groups of CMGs were recognized that were subsequently confirmed as distinct species following the sequencing of DNA-A molecules. These were ACMV and EACMV. The earliest distribution maps of the African CMGs showed EACMV to occur in the coastal east African areas of Kenya, Tanzania, and Madagascar, as well as Malawi and Zimbabwe. ACMV, by contrast, occurred throughout the remainder of the cassava-growing areas of Africa, from South Africa and Mozambique in the southeast, to Senegal in the northwest. Significantly, at this time there was no reported zone of co-occurrence of the two virus species. With the increased use of polymerase chain reaction (PCR)-based diagnostics from 1990s onward, it became possible to identify differences not solely associated with the coat protein. This led to two important developments in the understanding of CMGs in Africa. First, several new species were identified, most of which were more closely related to EACMV than to ACMV. Second, it was shown that virus mixtures belonging to different species occurred frequently. A notable consequence of this finding was the concomitant evidence for the more widespread distribution of the EACMV-like viruses than had hitherto been recognized. New virus identifications included: SACMV (1998), EACMMV (1998), EACMCV (2000), EACMZV (2004), EACMKV (2006), ACMBFV (2012), and CMMGV (2012). Significantly, all but one (ACMBFV) of the CMGs occur in different parts of East and Southern Africa, while ACMV predominates in West Africa (**Fig. 5**). EACMCV, the only EACMV-like virus occurring in West Africa, is less frequent and often occurs in mixed infections together with ACMV. ACMV is absent from coastal areas of Kenya and

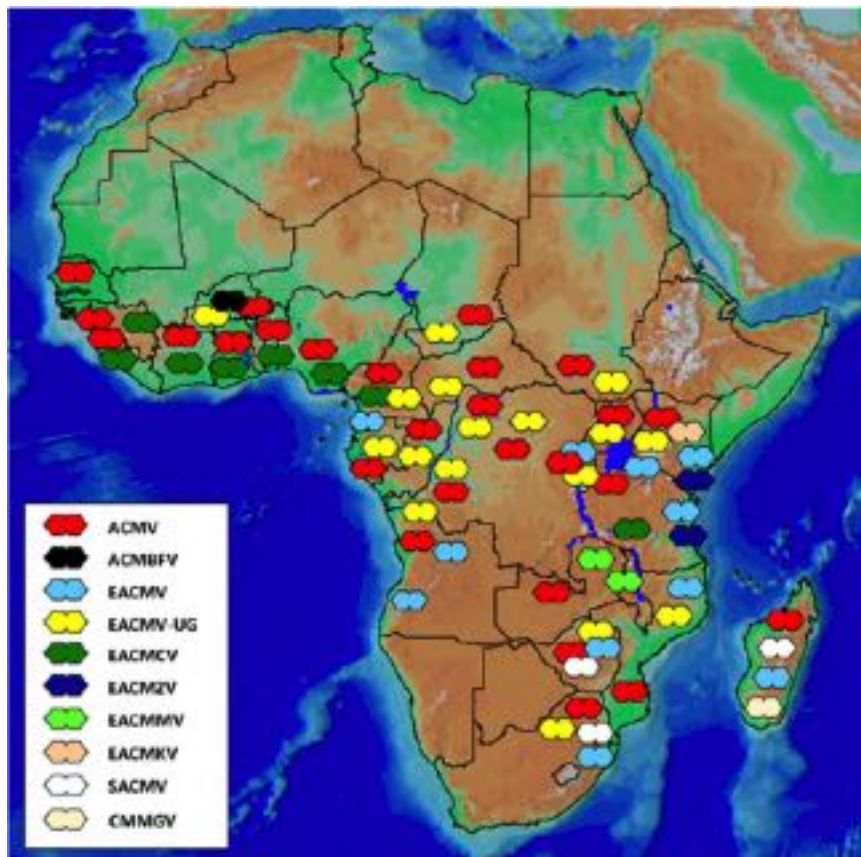


Fig. 5 Distribution of cassava mosaic geminiviruses in Africa: ACMV (African cassava mosaic virus); ACMBFV (African cassava mosaic Burkina Faso virus); EACMV (East African cassava mosaic virus); EACMV-UG (East African cassava mosaic virus-Uganda); EACMCV (East African cassava mosaic Cameroon virus); EACMZV (East African cassava mosaic Zanzibar virus); EACMMV (East African cassava mosaic Malawi virus); EACMKV (East African cassava mosaic Kenya virus); SACMV (South African cassava mosaic virus); CMMGV (Cassava mosaic Madagascar virus).

Tanzania. Since there has been very little CMG characterization in many of the cassava-growing countries of Africa, it seems likely that further variability within this group of viruses still remains to be revealed. An assessment of the data currently available has led to the conclusion that East Africa is the center of diversity for the EACMV-like CMGs and is probably the home for yet-to-be-identified wild hosts of the proto-CMGs that were first introduced to cassava by *B. tabaci* sometime between the earliest introductions of cassava to this part of Africa in the eighteenth century, and the first report of CMD in 1894.

Distribution of CMGs in Asia

ICMV and SLCMV are considered to be native to South Asia and have been recorded from cassava-producing areas of India and Sri Lanka for several decades. Early serology-based work identified ICMV from southern India. However, following the characterization of SLCMV as the causal agent of CMD in Sri Lanka in 2002, subsequent surveys in India showed that SLCMV was more widely distributed there than ICMV. The CMGs in India have been identified from Kerala, Karnataka, Tamil Nadu, Andhra Pradesh, Madhya Pradesh, and West Bengal states, a zone covering all of the major cassava-producing parts of the country. Although it has been suggested that SLCMV is a more aggressive virus than ICMV, and has been expanding its geographical coverage, the absence of systematic surveys conducted over several years means that the evidence for this remains circumstantial.

CMD was reported for the first time from Southeast Asia (SE Asia) from a single location in Ratanakiri Province in Northeastern Cambodia in 2015. The causal virus was SLCMV, and it was assumed that this new outbreak had arisen through the introduction of infected planting material from South Asia. Since that time there has been rapid spread of CMD and associated SLCMV through most of the major cassava-producing areas of Cambodia, as well as southern Vietnam. An isolated report of the disease from Ha Tinh Province in northern Vietnam presents a risk for spread to neighboring Laos, and there are also several reports of CMD in Thailand that have yet to be officially confirmed. The most recent confirmed reports of the occurrence of CMD in Asia come from the southern Chinese regions of Hainan and Fujian, in both cases coupled with positive diagnoses of SLCMV. Rapid spread of SLCMV through SE Asia is thought to be driven by extensive cross-border and intra-regional commercial trade in cassava stems, although there is also clear evidence of local spread through whitefly transmission. The progress of the SE Asian pandemic has also been exacerbated by the extreme susceptibility to CMG infection of virtually all of the cassava varieties currently being cultivated. Given the rapid spread of CMD through this region from 2015 to 2019, it seems likely that new occurrences will be reported from

Laos, Thailand and Myanmar in the near future, and there is also a greatly increased threat to other major-cassava producing countries in the region – Philippines and Indonesia.

Recombination and the CMD pandemic

The CMGs represent a very dynamic group of viruses, and evidence has been presented for the occurrence of both pseudo-recombinants, in which the DNA-A of one virus species co-replicates with the DNA-B of another species, as well as true recombinants, in which portions of the DNA-A or -B of one species have been spliced into the DNA-A or -B of another. In fact, all of the EACMV-like viruses other than EACMV show evidence for recombination events either with known or as yet unknown begomoviruses. One of the most important developments in the study of CMGs in Africa was the recognition that an unusually damaging strain, referred to as EACMV-UG, had arisen through a recombination event between EACMV and ACMV in which a 340 nt portion of ACMV AV1, had replaced the equivalent portion of the DNA-A of EACMV. The consequence of this was that ELISA-based diagnostic tests erroneously identified this strain as ACMV. This is one of the reasons why PCR-based diagnostic methods are now used almost exclusively for CMG monitoring work. EACMV-UG was associated with the rapid epidemic-like spread of CMD in East and Central Africa through the latter part of the twentieth and early 21st century. Countries affected by the severe CMD pandemic included: Angola, Burundi, Cameroon, Central African Republic, Democratic Republic of Congo (DRC), Gabon, Kenya, Republic of Congo, Rwanda, South Sudan, Tanzania, and Uganda. Localized identifications of EACMV-UG have also been made from Burkina Faso, Chad, Equatorial Guinea, South Africa, Swaziland, and Zimbabwe. The dynamic character of CMG diversity and distribution cannot be over-emphasized, and the propensity that this group of viruses shows to produce both naturally occurring pseudo-recombinants, as well as novel true recombinants, ensures that the patterns of diversity and distribution will continue to evolve. It is significant to note, however, that most of the CMG variability occurs amongst the EACMV-like viruses, which show a high propensity for recombination. This contrasts strongly with ACMV, for which all 311 publicly-available accessions have > 94% homology and can be considered to be a single strain of the same species.

Field-Level Epidemiology

Epidemiological studies of the CMGs can be broadly categorized into two groups: those that describe patterns of infection at the field level and those that relate to area or region-wide spread. In the case of the former, external sources of infection have been shown to be most important for determining the rate and quantity of infection in initially CMD-free plantings in the normal field environment. Gradients of infection occur in which new infections are most frequent on the windward sides of cassava crops and these gradients are matched by similar patterns of vector distribution. Multiple regression relationships have been used to describe the association between measures of inoculum pressure and final CMD incidence in test plots. Gompertz curves model patterns of infection increase in plots of CMG-susceptible cassava cultivars, and incidence increases rapidly to 100% over the first 3–6 months of growth. For resistant varieties, however, under similar conditions of inoculum pressure, rates of infection increase are much lower and final incidences typically range from 0% to 50%. Mathematical models have been generated that predict patterns of CMD spread under varied conditions of virus infection and vector abundance. The great majority of epidemiological research has focused on CMD in Africa, although several studies in India have confirmed that patterns of cutting-borne and vector-borne spread of CMGs are largely the same as those observed in Africa.

Regional Epidemiology

Following the earliest ‘first colonization’ descriptions of African CMD epidemics in the 1920s and 1930s, a few reports were made of rapid area-wide spread of severe CMD at other times during the twentieth century, some of the most notable of which were epidemics in Cape Verde and south-eastern Nigeria in the 1990s. Of much greater importance, however, has been the African CMD pandemic that was first reported from the northern-central part of Uganda in the late 1980s. CMD associated with the epidemic was unusually severe and was rapidly spread by super-abundant populations of *B. tabaci*. During the 1990s it became apparent that the zone affected by this severe CMD was expanding southwards at a rate of 20–30 km per year, and in 1997 molecular studies revealed that the severe disease phenotype was associated with the occurrence and spread of an ‘invasive’ recombinant CMG, EACMV-UG, commonly in mixed infection with the locally occurring, but now synergized ACMV. Regular monitoring surveys conducted throughout the East and Central African region through the 1990s and early part of the 21st century have used PCR-based diagnostics to map the spread of the EACMV-UG associated with this ‘pandemic’ of severe CMD. This work has given rise to first reports of EACMV-UG and resulting severe CMD in ten additional countries in East and Central Africa: Kenya (1996), South Sudan (1997), Tanzania (1999), DRC (1999), Republic of Congo (1999), Rwanda (2001), Burundi (2003), Gabon (2003), Angola (2008), and Cameroon (2010). Significantly, the pandemic ‘front’ advanced through north-western Tanzania from the Uganda border a distance of c. 400 km between 1999 and 2004. Although the severe CMD pandemic threatened to spread further through southern Africa and westwards to Nigeria, the period post-2010 saw a decline in new reports of spread. Evidence from Uganda in 2019 indicates that EACMV-UG has now largely been replaced by ACMV, although the reasons for this pattern of change are currently unclear. An important contributing factor to this amelioration in the CMD pandemic in Africa is likely to be the increasingly widespread adoption by farmers of CMD-resistant varieties.

Diagnosis

CMD symptoms are very characteristic, although they cannot be used to discriminate between the CMG virus species causing the infection. Additionally, symptoms are almost always expressed in infected plants at some stage during their period of growth, which means that symptoms are widely used as an indication of infection during surveillance activities and by regulatory authorities during certification inspections. The characteristic mosaic pattern of CMD symptoms has also made it possible to use machine learning and artificial intelligence to develop a smartphone app – Nuru AI – which allows untrained users to recognize CMD with a high degree of accuracy. Laboratory-based methods are required, however, to increase the accuracy of detection as well as to identify the various CMG species.

ELISA was the first of these approaches to be widely applied to CMG detection. Monoclonal antibodies allowed a distinction to be made between ACMV, EACMV, and ICMV, at a time when only these three species were recognized. The emergence of the severe CMD epidemic in Uganda, however, highlighted the weakness with this serology-based method, as EACMV-UG was misdiagnosed as ACMV by ELISA, since EACMV-UG was a recombinant virus with the coat protein of ACMV. From the late 1990s onwards PCR was used almost exclusively for the detection and identification of CMG species. PCR presents an easy and robust detection system for CMGs and specific primer sets and a diverse array of PCR protocols are available to achieve reliable detection of these viruses in cassava. To verify specificity, sequencing of PCR amplicons confirms virus identity. However, because of recombination at various positions across DNA-A, amplification of genome fragments may not fully resolve the diversity of DNA-A and the association of sequence to a particular virus species. Hence amplification followed by sequencing of the entire DNA-A genome component is required for reliable virus identification. This is achieved by using inverse, abutting PCR primers located in a conserved region of DNA-A (CP) to amplify the entire circular DNA-A molecule. A more general approach is rolling-circle amplification (RCA) using phi29 (Φ 29) DNA polymerase. Plant DNA including the circular virus DNA is amplified in an isothermal process using random oligonucleotide primers. Phi29 polymerase is highly processive and in this unique RCA process, small circular DNA molecules are replicated nearly infinitely. By enzymatic restriction of the RCA products using rarely-cutting restriction enzymes, CMG DNA fragments or complete virus genome components can be resolved and the entire begomovirus composition of a plant sample can be analyzed without prior knowledge of virus sequences. With this approach it is possible to detect unexpected and unknown begomoviruses, mixed infections and satellites. Sample preparation and DNA extraction are identical to other methods and commercial kits are available. RCA generates genome DNA sequences as concatemers. Virus identification can then be performed after restriction digestion, cloning and sequence analysis.

Economic Importance

CMD symptoms of leaf chlorosis, reduction of leaf size and plant stunting lead to a reduction in the quantity of photosynthetic assimilates channeled into the tuberous roots, and through this, a reduction in yield. The degree of yield loss varies greatly, depending primarily on the susceptibility of the cassava cultivar, the virulence of the CMG species/strain, and the stage of growth at which infection occurred (being most severe for cutting-infected plants). Typically, individual plants sustain yield losses ranging from 20% to 100%. The only study of the effects of specific viruses on yield showed average yield losses to be 42% for ACMV alone, 12% for EACMV-UG mild, 68% for EACMV-UG severe, and 82% for mixed ACMV + EACMV-UG. The relatively moderate losses attributable to ACMV infection, coupled with moderate to low incidences, have been the reasons for the apparent lack of concern through large parts of cassava-growing Africa about the impact of CMGs and CMD. This outlook changed markedly, however, following the emergence and regional spread of the CMD pandemic. Extensive surveys of CMD incidence, CMD severity, and the occurrence of the causal viruses have made it possible to make continent-wide yield loss estimates for CMD. The last of these (2006) provided estimates for losses in pandemic-recovered (16% loss), pandemic-affected (47%), and as yet unaffected (18%) countries in sub-Saharan Africa, leading to an overall loss figure of 34 million tons per year, which was equivalent to roughly a third of total African production of fresh cassava roots. Following the region-wide spread of the pandemic, major control programs were implemented, and anecdotal evidence suggests that this has resulted in reductions in CMD incidence and reduced losses. This declining impact of the disease in Africa is likely to be the reason why there have been no comprehensive assessments of losses since 2006.

For CMD in India, losses of up to 84% have been recorded at field level, although these are said to vary, as in Africa, depending on the varietal response, type of infection and virus species/species mixture causing the infection. There are currently no published records of overall losses due to CMD in South Asia or for the more recently affected countries of Southeast Asia.

Control

The most widely practiced approaches to controlling the CMGs that cause CMD include the deployment of host plant resistance and the use of cultural methods, particularly phytosanitation. More recently, considerable attention has been directed toward the use of genetic engineering techniques to produce transgenic virus-resistant cassava plants.

Host Plant Resistance

The potential value of introgressing virus resistance genes from wild relatives of cassava was recognized from the earliest days of CMD research in the 1920s/1930s, and inter-specific crosses combining cultivated cassava with Ceara rubber (*Manihot glaziovii*) were developed independently through breeding programs in modern-day Tanzania and Madagascar. F1 progeny were triple backcrossed with cultivated cassava to produce plants that combined acceptable food quality with significantly enhanced resistance to CMGs. Germplasm developed in this way formed the basis for the later continental breeding program run from the Nigeria-based International Institute of Tropical Agriculture (IITA), which from its establishment in 1967, to the present day, has developed thousands of CMD-resistant cassava clones. Many of these have been sent to cassava-producing countries in Africa for use either specifically for CMD management programs, or more generally for cassava development. Germplasm derived from the initial interspecific crosses uses the name prefix 'TMS' for 'Tropical Manihot species'. This resistance source is multigenic and has provided high levels of resistance which have been very durable when used against all CMGs and CMG combinations. Four distinct mechanisms of resistance are recognized: resistance to infection, resistance to virus multiplication, resistance to virus movement (leading to incomplete systemicity), and resistance of normal plant function to the effects of virus infection. During the 1990s, resistant landraces from West Africa, given the name prefix 'TME' for 'Tropical Manihot esculenta', were incorporated into the breeding program. These have been shown to possess alternative sources of resistance, one of the most important of which has been characterized through genetic analyses as a single dominant gene designated CMD2. Molecular marker approaches have been used subsequently to combine the multigenic *M. glaziovii*-derived resistance with CMD2. In recent years, improvements in sequencing technologies have enabled researchers to undertake genome-wide association studies. These have been used to identify major QTLs associated with CMD resistance and they have provided new tools to improve the efficiency of the breeding process.

Hundreds of CMG-resistant varieties have been released and disseminated throughout the cassava-growing areas of sub-Saharan Africa. However, their adoption by farming communities has been most widespread in CMD pandemic-affected countries, where they have provided the only effective means of restoring cassava production to pre-pandemic levels. In South Asia, sources of resistance have been drawn from the breeding work undertaken in Africa. Evaluations of seedling populations obtained from introduced germplasm have led to the development of varieties such as CMR1, CMR129, CR21-10, and CR43-11 which combine high starch content with strong resistance to ICMV and SLCMV. Some of this germplasm, together with other CMD-resistant varieties from IITA in Nigeria are currently being introduced to Southeast Asia as part of regional efforts to control the CMD outbreak there.

Cultural Methods

A range of cultural methods has been proposed for the control of CMGs. The methods most widely recommended have been the removal of infected plants (roguing) or the selection of disease-free planting material for the establishment of a new crop (selection). Crop isolation, adjusting crop disposition in relation to the prevailing wind, varying planting date, varietal mixtures, and intercropping cassava with other 'putative' protective crops have all been suggested at various times as potentially useful control options for CMGs. No convincing experimental evidence has yet been presented to confirm the value of any of these methods, however, and current field practice is restricted to selection and occasional roguing. Roguing is considered to be of value within the framework of institutional programs for the multiplication of CMD-resistant germplasm, in view of the requirement for the production of high-quality planting material. Experiments conducted in 'post-epidemic' areas of East Africa, first affected by the CMD pandemic five or more years previously, have provided clear evidence for the value of selection of CMD-free stems when choosing planting material. Local cultivars treated in this way provided equivalent yields to those of CMD-resistant varieties after two cropping cycles. A key drawback to the wider adoption of this approach, however, is the variability in effectiveness of the approach in relation both to the virus inoculum pressure of the location, as well as the relative susceptibility of the cultivar.

Systems for the phytosanitary control of CMGs have been greatly strengthened in several parts of Africa in recent years. Most notably, certification guidelines for cassava have been developed and appended to Seed Acts in countries such as Nigeria, Rwanda and Tanzania. These guidelines include maximum tolerance levels for CMD incidence. This has facilitated the certification of early generation cassava 'seed' (= planting material) and promoted the production of near virus-free planting material. The approach is helping to ensure that newly developed virus-resistant varieties are multiplied and disseminated in a healthy condition and it is reducing overall incidences of CMD.

Virus Resistance by Engineering

Engineering virus resistance by expression of virus sequences in transgenic plants is a strategy that is also used to achieve resistance against geminiviruses. The most common approach is to engineer RNAi (RNA interference) constructs as inverted-repeats of virus sequence which are spaced by an intron. Splicing of the intron upon transcription generates a "hairpin" dsRNA that is targeted by the plants RNA silencing machinery. RNAi triggered antiviral defense using constructs comprising Rep sequences is effective against Tomato yellow leaf curl virus (TYLCV) and against Bean golden mosaic virus (BGMV). The stability of this resistance was verified in field trials showing virus resistance under field conditions for several seasons and the feasibility of the

approach to provide protection in both crops. In Brazil, transgenic common beans resistant to BGMV were released in 2011 as a commercial variety. Thus RNAi strategies may protect plants against DNA viruses as shown also for other crops including cotton as well as cassava, by boosting the antiviral defense. The effectiveness of the RNAi protection depends on the choice of the viral sequence and the abundance of small interfering RNAs generated from the hairpin RNA transgene. In addition, temperature and the intrinsic interaction of virus and plant host influence the effectiveness of silencing and the level of resistance achieved. RNAi strategies have been used to induce gene silencing against cassava infecting geminiviruses. In all approaches, symptoms were attenuated and virus DNA accumulation was drastically reduced. Broad spectrum resistance against several CMGs e.g., ACMV, EACMCV, and SLCMV can be achieved when there is sufficient sequence homology between the siRNAs and transcripts of the viruses.

Several studies have shown the effectiveness of CRISPR-Cas9 to engineer resistance against the geminiviruses Tomato yellow leaf curl virus (TYLCV) and Beet curly top virus (BCTV). A single-guide RNA (sgRNA) carrying the TAATATTAC nonanucleotide origin of replication of geminiviruses was used to induce a broad spectrum virus resistance. In cassava, CRISPR-Cas9 was used to engineer resistance to ACMV targeting AC2 and AC3 with a single sgRNA. However, CRISPR-Cas9 did not show any difference to the wild-type susceptible plants because of the emergence of a cleavage mutant of ACMV in transgenic plants. The selection of cleavage resistant mutants, as shown for ACMV, or new virus mutants that evade the sequence-specific action of transgene derived RNAi in TYLCV transgenic plants or ACMV transgenic cassava demonstrate the challenges of any of the engineering strategies employing targeting sequences of a replicating virus. However, this can be addressed by using multiplexed CRISPR/Cas9 with multiple sgRNAs to target virus complexes, as shown for the control of Cotton leaf curl virus (CLCuV). With more information on functional networks and genes involved in vital virus functions, CRISPR-Cas9, RNAi or similar approaches will be increasingly used for the editing or interfering with expression of host genes to achieve virus resistance in cassava.

Biosafety considerations

Cassava plants transformed for resistance to CMGs have been tested under field conditions in several African countries (Kenya, Nigeria, Uganda) through confined field trial arrangements, with promising results confirming the effectiveness of the strategy. However, uncertainty in most African countries about the consequences of the widespread cultivation of GMOs has slowed the adoption of biosafety guidelines and meant that at the time of writing (2019) there are no transgenic cassava plants that have been evaluated under 'real-world' conditions in farmers' fields. In January 2019, however, the National Biosafety Management Agency of Nigeria made its first approval of a GM crop – podborer-resistant cowpea. Since Nigeria is the world's largest producer of cassava, and since its cassava crop is widely affected by CMD, this development offers significant promise for the future release of GM CMD-resistant cassava varieties. Furthermore, this is a move that could provide a signal for other countries similarly affected by cassava viruses to follow suit. Given the huge impact that CMD has on the production of one of the world's most important staple food crops, the widespread adoption of science-based strategies to control may be essential if the food security of hundreds of millions of cassava-dependent households in Africa and elsewhere is to be assured.

See also: Bean Golden Mosaic Virus and Bean Golden Yellow Mosaic Virus (*Geminiviridae*). Beet Curly Top Virus (*Geminiviridae*). Cassava Brown Streak Viruses (*Potyviridae*). Cotton Leaf Curl Disease (*Geminiviridae*). Emerging Geminiviruses (*Geminiviridae*). Geminiviruses (*Geminiviridae*). Maize Streak Virus (*Geminiviridae*). Plant Resistance to Geminiviruses. Tomato Leaf Curl New Delhi Virus (*Geminiviridae*). Tomato Yellow Leaf Curl Viruses (*Geminiviridae*)

Further Reading

- Bock, K.R., Woods, R.D., 1983. The etiology of African cassava mosaic disease. *Plant Disease* 67, 994–995.
- Dubern, J., 1994. Transmission of African cassava mosaic geminivirus by the whitefly (*Bemisia tabaci*). *Tropical Science* 34, 82–91.
- Dutt, N., Briddon, R.W., Dasgupta, I., 2005. Identification of a second begomovirus, Sri Lankan cassava mosaic virus, causing cassava mosaic disease in India. *Archives of Virology* 150 (10), 2101–2108.
- Fauquet, C., Fargette, D., 1990. African cassava mosaic virus: Etiology, epidemiology and control. *Plant Disease* 74, 404–411.
- Fondong, V.N., 2017. The search for resistance to cassava mosaic geminiviruses: How much we have accomplished, and what lies ahead. *Frontiers in Plant Science* 8, 408. doi:10.3389/fpls.2017.00408.
- Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D., Mansoor, S., 2013. Geminiviruses: Masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology* 11, 777–788.
- Harrison, B.D., Zhou, X., Otim-Nape, G.W., Liu, Y., Robinson, D.J., 1997. Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. *Annals of Applied Biology* 131, 437–448.
- Hipp, K., Grimm, C., Jeske, H., Bottcher, B., 2017. Near-atomic resolution structure of a plant geminivirus determined by electron cryomicroscopy. *Structure* 25, 1303–1309.
- Legg, J.P., Lava Kumar, P., Makesh Kumar, T., et al., 2015. Cassava virus diseases: Biology, epidemiology and management. *Advances in Virus Research* 91, 85–142.
- Legg, J.P., Owor, B., Sseruwagi, P., Ndunguru, J., 2006. Cassava mosaic virus disease in East and Central Africa: Epidemiology and management of a regional pandemic. *Advances in Virus Research* 67, 355–418.
- Mehta, D., Sturchler, A., Anjanappa, R.B., Zaidi, S.S., Hirsch-Hoffmann, M., Gruijssem, W., Vanderschuren, H., 2019. Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses. *Genome Biology* 20, 80.
- Otim-Nape, G.W., Bua, A., Thresh, J.M., et al., 1997. Cassava Mosaic Virus Disease in Uganda: The Current Pandemic and Approaches to Control. Chatham, UK: Natural Resources Institute, p. 65.
- Swanson, M.M., Harrison, B.D., 1994. Properties, relationships and distribution of cassava mosaic geminiviruses. *Tropical Science* 34, 15–25.

- Thresh, J.M., Otim-Nape, G.W., Thankappan, M., Muniyappa, V., 1998. The mosaic diseases of cassava in Africa and India caused by whitefly-borne geminiviruses. Review of Plant Pathology 77, 935–945.
- Wang, H.L., Cui, X.Y., Wang, X.W., *et al.*, 2016. First report of Sri Lankan cassava mosaic virus infecting cassava in Cambodia. Plant Disease 100, 1029.

Relevant Websites

- <https://gd.eppo.int/taxon/ACMV00>
African cassava mosaic virus (ACMV00)[Overview]
EPPO Global Database.
- <https://www.cabi.org/isc/datasheet/2747>
Cassava mosaic disease (African cassava mosaic)
CABI.
- <https://www.youtube.com/watch?v=rY8xyO44Z18>
Cassava Mosaic Disease (CMD) in Southeast Asia.
- <https://ciat.cgiar.org/event/cmd-sea/>
Cassava Mosaic Disease (CMD) in Southeast Asia
CIAT.
- <https://www.accessagriculture.org/cassava-mosaic-virus>
Cassava mosaic virus
Access Agriculture.
- <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/cassava-mosaic-virus>
Cassava Mosaic Virus
ScienceDirect.com.
- <https://www.dsmz.de>
DSMZ (homepage).
- <https://talk.ictvonline.org/files/master-species-lists/m/msl/8266>
ICTV Master Species List 2018b.v2
International Committee on Taxonomy of Viruses (ICTV).
- <https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses (ICTV).
- <https://www.plantvillage.psu.edu>
PlantVillage
Penn State.
- <https://www.iita.org>
The International Institute of Tropical Agriculture (IITA).

Caulimoviruses (*Caulimoviridae*)

James E Schoelz, University of Missouri, Columbia, MO, United States

Mustafa Adhab, University of Baghdad, Baghdad, Iraq

© 2021 Elsevier Ltd. All rights reserved.

This is an update of T. Hohn, Caulimoviruses: Molecular Biology, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00367-8](https://doi.org/10.1016/B978-012374410-4.00367-8).

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

dsDNA Double stranded DNA

IBs Inclusion bodies

kb Kilobase

kDa Kilo dalton

MP Movement protein

nt Nucleotide(s)

ORF Open Reading Frame

TAV Translational transactivator

Glossary

Agro-inoculation An inoculation technique in which *Agrobacterium tumefaciens* is used to deliver a full-length infectious clone of a virus into a plant cell.

Plasmodesmata A narrow channel of cytoplasm that functions as a bridge between two plant cells to facilitate movement of macromolecules.

Reverse transcriptase An enzyme that utilizes an RNA template for synthesis of DNA.

Ribosome shunt A translational mechanism in which ribosomes enter the RNA at the 5' end and scan for a short distance before being translocated to a downstream point.

Semi-persistent transmission Vector acquires the virus in minutes to hours, and can transmit to other plants for hours after the initial feeding.

Transgene A gene introduced into an organism through any one of a number of genetic engineering techniques.

Introduction

The members of the family *Caulimoviridae* are plant viruses that replicate by reverse transcription of an RNA intermediate and whose virions contain circular, double stranded DNA. They belong to the order *Ortervirales*, a virus order that encompasses five families of reverse transcribing viruses. Although caulimoviruses replicate by reverse transcription, integration into host chromosomes is not required for completion of their replication cycle. The circular DNA encapsidated in virions is not covalently closed, as it contains at least one discontinuity in each DNA strand, and these discontinuities occur as a consequence of the replication strategy of the virus. There are eight genera in this family and they can be divided into two groups based on virion morphology; the members of the genera *Caulimovirus*, *Petuvirus*, *Cavemovirus*, *Rosadnavirus*, *Solendovirus*, and *Soymovirus* contain viruses that form icosahedral particles that are largely found within amorphous inclusion bodies in the cell (**Fig. 1(A)** and **(B)**).

In contrast, the members of the genera *Badnavirus* and *Tungrovirus* form bacilliform particles that are not associated with inclusion bodies (**Fig. 1(C)**). Their virions are found in the cytoplasm either individually or clustered in palisade-like arrays.

Cauliflower mosaic virus (CaMV), the type species of the genus *Caulimovirus*, was the first of the plant viruses to be shown to contain double-stranded DNA (dsDNA) in its icosahedral virion in 1968. This discovery led to an extended investigation into its replication strategy throughout the 70's and early 80's, culminating in 1983 when it was shown that CaMV replicated through reverse transcription of an RNA intermediate. Perhaps because the genome of CaMV is composed of dsDNA, it was also the first of the plant viruses to be completely sequenced and cloned into bacterial plasmids in an infectious form.

In the early 1980s, CaMV was thought to have some promise as a vector for foreign genes in plants. However, the effort to convert CaMV into a vector was scaled back when it was shown that the virus genome could tolerate only small insertions of up to a few hundred base pairs of DNA. A few small genes, such as dihydrofolate reductase and interferon, were eventually expressed in plants via a CaMV vector, but other viruses have been shown to be much more versatile as vectors for foreign genes. Although the caulimoviruses have had only limited utility as plant virus vectors, they continue to have a great impact on plant biotechnology. The 35S promoter of CaMV is capable of directing a high level of transcription in most types of plant tissues. This promoter was used to drive expression of one of the first transgenes introduced into transgenic plants, and it is still widely used for expression of transgenes for both research and commercial applications. The promoter regions of several other caulimoviruses have also been evaluated for expression of transgenes in both monocots and dicots, and they can be good alternatives to the CaMV 35S promoter.

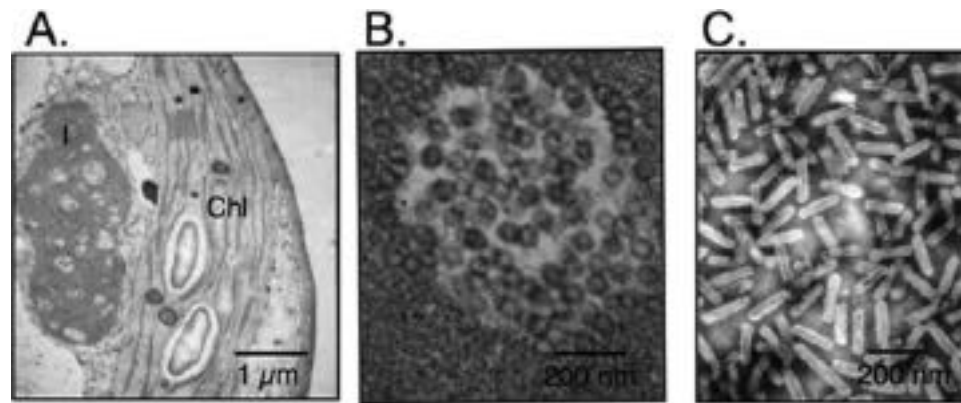


Fig. 1 Inclusion bodies and virions of members of the family *Caulimoviridae*. (A). Amorphous inclusion body of CaMV (I) adjacent to a chloroplast (Chl). Individual virions can be seen in the vacuolated regions of the inclusion body. (B). Icosahedral virions of CaMV visualized within an inclusion body. (C). Purified bacilliform virions of CoYMV. CoYMV photo courtesy of Ben Lockhart (University of Minnesota).

For many years, only plant viruses that had icosahedral virions of 50 nm in diameter were thought to have genomes composed of dsDNA. However in 1990, it was shown that the bacilliform virions of Commelina yellow mottle virus (CoYMV) contained circular, dsDNA, and *Commelina yellow mottle virus* became the type species for the genus *Badnavirus*. Nucleic acid isolated from the bacilliform virions was resistant to RNase and degraded by DNase. Furthermore, DNA treated with S1 nuclease revealed that the CoYMV genome contained at least two single-stranded discontinuities, a hallmark of the plant viruses that replicate by reverse transcription. In addition to CoYMV, DNA purified from Banana streak virus (BSV), Kalanchoe top-spotting virus (KTSV) and Canna yellow mottle virus, providing evidence that they should also be placed in the new genus *Badnavirus*. Currently, there are 57 species in the *Badnavirus* genus (Table 1), many of which cause economically important diseases in the tropics.

Soon after the discovery that CoYMV was a DNA virus, the bacilliform component of rice tungro disease was also shown to contain circular, dsDNA. Rice tungro is the most important virus disease of rice in South and Southeast Asia, with annual losses approaching \$1.5 billion dollars. The disease is caused by a complex of an RNA virus, called Rice tungro spherical virus (RTSV) coupled with the bacilliform dsDNA virus called Rice tungro bacilliform virus (RTBV). The genus *Tungrovirus* is distinguished from the genus *Badnavirus* because RTBV has one more open reading frame (ORF) than CoYMV and overall, the RTBV genome has only 20%–25% nt sequence identity with members of the genus *Badnavirus*. There is only one species in the genus *Tungrovirus*, although several isolates of the species *Rice tungro bacilliform virus* have been collected and sequenced.

Taxonomy, Phylogeny, and Evolution

The family *Caulimoviridae* consists of eight genera (Table 1), and they can be conveniently divided into two groups based on virion morphology. Viruses with icosahedral virions (Fig. 1(B)) include the genera *Caulimovirus*, *Petuvirus*, *Soymovirus*, *Rosadnavirus*, *Solendovirus*, and *Cavemovirus*. Viruses with bacilliform virions (Fig. 1(C)) include the genera *Badnavirus* and *Tungrovirus*. Genera can be further distinguished because of differences in genome organization (Fig. 2(A)) as well as phylogenetic analyses of proteins such as the reverse transcriptase, as illustrated in Fig. 2(B). All members of the family *Caulimoviridae* infect only plants. There are no animal or insect viruses in this family. Complete and partial nt sequences of members of the family *Caulimoviridae* are listed in Table 1. Only a single accession number is given for CaMV and RTBV, although multiple strains of each have been sequenced.

Virion Structure

The viruses in the genera *Caulimovirus*, *Cavemovirus*, *Soymovirus* and *Petuvirus* form non-enveloped, isometric particles that vary in size from 43 to 50 nm. The virion is composed of 420 subunits with a $T = 7$ structure. The viruses in the genera *Badnavirus* and *Tungrovirus* form non-enveloped, bacilliform particles that are 30 nm in width, but can vary in length from 60 to 900 nm. The length most commonly observed is 130 nm. Their structure is based on an icosahedron, in which the ends are formed from pentamers and the tubular section is made up of hexamers. The capsid also is loosely decorated with the 15 kDa virus-associated protein, which is encoded by ORF-III. The virions typically accumulate in amorphous inclusion bodies, which are composed primarily of a 62 kDa multifunctional protein encoded by ORF-VI. The virions of the family *Caulimoviridae* contain a single, circular dsDNA that is 7.2–9.3 kbp in length.

Table 1 Virus genera and virus species members in the family *Caulimoviridae*

Genus	Virus	Acronym	Host (Family)	Distribution	Accession #
<i>Caulimovirus</i>	Cauliflower mosaic virus ^a	CaMV	<i>Brassica</i> sp. (<i>Cruciferae</i>)	Worldwide	V00140
	Angelica bushy stunt virus	AnBSV	Angelica (<i>Apiaceae</i>)	South Korea	KU508800
	Atractylodes mild mottle virus	AMMV	Atractylodes (<i>Compositae</i>)	South Korea	KR080327
	Carnation etched ring virus	CERV	Carnation (<i>Caryophyllaceae</i>)	Worldwide	X04658
	Dahlia mosaic virus	DMV	Dahlia (<i>Compositae</i>)	Worldwide	KC967481
	Figwort mosaic virus	FMV	Figwort (<i>Scrophulariaceae</i>)	U.S.A.	X06166
	Horseradish latent virus	HRLV	Horseradish (<i>Cruciferae</i>)	Denmark	JX429923
	Lamium leaf distortion	LLDV	Lamium (<i>Lamiaceae</i>)	U.S.A.	EU554423
	Mirabilis mosaic virus	MiMV	Mirabilis (<i>Nyctaginaceae</i>)	U.S.A.	AF454635
	Rudbeckia flower distortion virus	RuFDV	Rudbeckia (<i>Asteraceae</i>)	U.S.A.	NC_011920
	Soybean Putnam virus	SPuV	Soybean (<i>Leguminosae</i>)	U.S.A.	JQ926983
	Strawberry vein banding virus	SVBV	Strawberry (<i>Rosaceae</i>)	Worldwide	X97304
	Thistle mottle virus	ThMoV	Thistle (<i>Compositae</i>)	Europe	
<i>Petuvirus</i>	Petunia vein clearing virus	PVCV	Petunia (<i>Solanaceae</i>)	Worldwide	U95208
<i>Soymovirus</i>	Soybean chlorotic mottle virus	SbCMV	Soybean (<i>Leguminosae</i>)	Japan	X15828
	Blueberry red ringspot virus	BRRSV	Blueberry (<i>Ericaceae</i>)	U.S.A.	AF404509
	Cestrum yellow leaf curling virus	CmYLCV	Cestrum sp. (<i>Solanaceae</i>)	Italy	AF364175
	Peanut chlorotic streak virus	PCSV	Peanut (<i>Leguminosae</i>)	India	U13988
<i>Cavemovirus</i>	Cassava vein mosaic virus	CsVMV	Cassava (<i>Euphorbiaceae</i>)	Brazil	U59751
	Sweet potato collusive virus	SPCV	Sweet potato (<i>Convolvulaceae</i>)	Worldwide	
<i>Rosadnavirus</i>	Rose yellow vein virus	RYV	Rose (<i>Rosaceae</i>)	U.S.A.	JX028536
<i>Solendovirus</i>	Tobacco vein clearing virus	TVCV	<i>Nicotiana</i> sp. (<i>Solanaceae</i>)	Worldwide	AF190123
	Sweet potato vein clearing virus	SPVCV	Sweet potato (<i>Convolvulaceae</i>)	Africa, Americas	NC_015228
<i>Badnavirus</i>	Commelina yellow mottle virus	ComYMV	Commelina (<i>Commelinaceae</i>)	Caribbean	X52938
	Aglaonema bacilliform virus	ABV	Aglaonema sp. (<i>Araceae</i>)	Southeast Asia	
	Banana streak GF virus	BSGFV	Banana (<i>Musaceae</i>)	Worldwide	AY493509
	Banana streak IM virus	BSIMV	Banana (<i>Musaceae</i>)	Worldwide	KJ013508
	Banana streak Mysore virus	BSMyV	Banana (<i>Musaceae</i>)	Worldwide	AJ805074
	Banana streak OL virus	BSOLV	Banana (<i>Musaceae</i>)	Worldwide	AJ002234
	Banana streak UA virus	BSUAV	Banana (<i>Musaceae</i>)	Worldwide	NC_015502
	Banana streak UI virus	BSUIV	Banana (<i>Musaceae</i>)	Worldwide	HQ593108
	Banana streak UL virus	BSULV	Banana (<i>Musaceae</i>)	Worldwide	HQ593109
	Banana streak UM virus	BSUMV	Banana (<i>Musaceae</i>)	Worldwide	HQ93110
	Banana streak VN virus	BSNV	Banana (<i>Musaceae</i>)	Worldwide	NC_007003
	Birch leafroll-associated virus	BLRaV	Birch (<i>Betulaceae</i>)	Europe	MG686420
	Blackberry virus F	BVF	Blackberry (<i>Rubus</i>)	U.S.A.	KJ413252
	Bougainvillea chlorotic vein banding virus	BsCVBV	Bougainvillea (<i>Nyctaginaceae</i>)	Worldwide	NC_011592
	Cacao bacilliform Sri Lanka virus	CBSLV	Cacao (<i>Malvaceae</i>)	Sri Lanka	MF642736
	Cacao mild mosaic virus	CaMMV	Cacao (<i>Malvaceae</i>)	Trinidad	KX276640
	Cacao swollen shoot CD virus	CSSCDV	Cacao (<i>Malvaceae</i>)	Africa	JN606110
	Cacao swollen shoot CE virus	CSSCEV	Cacao (<i>Malvaceae</i>)	Africa	MF642719
	Cacao swollen shoot Ghana M virus	CSSGMV	Cacao (<i>Malvaceae</i>)	Africa	MF642724
	Cacao swollen shoot Ghana N virus	CSSGNV	Cacao (<i>Malvaceae</i>)	Africa	MF642725
	Cacao swollen shoot Ghana Q virus	CSSGQV	Cacao (<i>Malvaceae</i>)	Africa	MF642728
	Cacao swollen shoot Togo A virus	CSSTAV	Cacao (<i>Malvaceae</i>)	Africa	AJ781003
	Cacao swollen shoot Togo B virus	CSSTBV	Cacao (<i>Malvaceae</i>)	Africa	L14546
	Cacao yellow vein banding virus	CYVBV	Cacao (<i>Malvaceae</i>)	Trinidad	KX276641
	Canna yellow mottle associated virus	CaYMAV-1	Canna (<i>Cannaceae</i>)	Worldwide	NC_030462
	Canna yellow mottle virus	CaYMV	Canna (<i>Cannaceae</i>)	Worldwide.	NC_038380
	Citrus yellow mosaic virus	CIYMV	Citrus sp. (<i>Rutaceae</i>)	India	NC_003382
	Dioscorea bacilliform AL virus	DBV	Yam (<i>Dioscoreaceae</i>)	Africa	X94576
	Dioscorea bacilliform AL virus 2	DBALV 2	Yam (<i>Dioscoreaceae</i>)	Africa	KY827395
	Dioscorea bacilliform ES virus	DBESV	Yam (<i>Dioscoreaceae</i>)	Africa	KY827394
	Dioscorea bacilliform RT virus 1	DBRTV1	Yam (<i>Dioscoreaceae</i>)	Africa	KX008574
	Dioscorea bacilliform RT virus 2	DBRTV2	Yam (<i>Dioscoreaceae</i>)	Africa	KX008577
	Dioscorea bacilliform SN virus	DBSNV	Yam (<i>Dioscoreaceae</i>)	Africa	NC_009010
	Dioscorea bacilliform TR virus	DBTRV	Yam (<i>Dioscoreaceae</i>)	Worldwide	KX430257

(Continued)

Table 1 Continued

Genus	Virus	Acronym	Host (Family)	Distribution	Accession #
	<i>Fig badnavirus 1</i>	FBV1	Fig (<i>Moraceae</i>)	Worldwide	JF411989
	<i>Gooseberry vein banding associated virus</i>	GVBAV	Ribes sp. (<i>Grossulariaceae</i>)	Worldwide	AF298883
	<i>Grapevine roditis leaf discoloration associated virus</i>	GRLDaV	Grapes (<i>Vitaceae</i>)	Greece	HG940503
	<i>Grapevine vein clearing virus</i>	GVCV	Grapes (<i>Vitaceae</i>)	U.S.A.	NC_015784
	<i>Jujube mosaic-associated virus</i>	JuMAV	Chinese Jujube (<i>Rhamnaceae</i>)	China	NC_035472
	<i>Kalanchoe top-spotting virus</i>	KTSV	<i>Kalanchoe</i> (<i>Crassulaceae</i>)	Worldwide	AY180137
	<i>Mulberry badnavirus 1</i>	MBV1	Mulberry (<i>Moraceae</i>)	Lebanon	NC_026020
	<i>Pagoda yellow mosaic associated virus</i>	PYMAV	Pagoda tree (<i>Fabaceae</i>)	China	NC_024301
	<i>Pineapple bacilliform CO virus</i>	PBCOV	Pineapple (<i>Bromeliaceae</i>)	Worldwide	GQ398110
	<i>Pineapple bacilliform ER virus</i>	PBERV	Pineapple (<i>Bromeliaceae</i>)	Worldwide	EU377672
	<i>Piper yellow mottle virus</i>	PYMoV	Pepper (<i>Piperaceae</i>)	Brazil, India, Asia	
	<i>Rubus yellow net virus</i>	RYNV	Raspberry (<i>Rosaceae</i>)	Eurasia, U.S.A.	AF468454
	<i>Schefflera ringspot virus</i>	SRV	<i>Schefflera</i> (<i>Araliaceae</i>)	Worldwide	
	<i>Spiraea yellow leaf spot virus</i>	SYLSV	<i>Spiraea</i> (<i>Rosaceae</i>)	Worldwide	AF299074
	<i>Sugarcane bacilliform Guadeloupe A virus</i>	SCBGAV	Sugarcane (<i>Poaceae</i>)	West Indies	FJ824814
	<i>Sugarcane bacilliform Guadeloupe D virus</i>	SCBGDV	Sugarcane (<i>Poaceae</i>)	West Indies	FJ439817
	<i>Sugarcane bacilliform IM virus</i>	SCBIMV	Sugarcane (<i>Poaceae</i>)	U.S.A., Morocco	AJ277091
	<i>Sugarcane bacilliform MO virus</i>	SCBMOV	Sugarcane (<i>Poaceae</i>)	U.S.A., Morocco	M89923
	<i>Sweet potato pakakuy virus</i>	SPPV	Sweet potato (<i>Convolvulaceae</i>)	Worldwide	NC_015655
	<i>Taro bacilliform CH virus</i>	TaBCHV-1	Taro (<i>Araceae</i>)	China	KP10178
	<i>Taro bacilliform virus</i>	TaBV	Taro (<i>Araceae</i>)	South Pacific	AF357836
	<i>Wisteria badnavirus 1</i>	WBV1	Wisteria (<i>Fabaceae</i>)	China	KX168422
	<i>Yacon necrotic mottle virus</i>	YV1	Yacon (<i>Asteraceae</i>)	Korea	KM229702
<i>Tungrovirus</i>	<i>Rice tungro bacilliform virus</i>	RTBV	Rice (<i>Poaceae</i>)	East Asia, China	X57924

^aThe type species of each genus is highlighted in bold font.

Genome Organization

Several genome features are common to all members of the family *Caulimoviridae*, in addition to their circular dsDNA genomes. For example, the dsDNA is not covalently closed; it has at least two discontinuities, but may have up to four. The most significant difference among the caulimoviruses concerns the arrangement and number of ORFs (Fig. 2). All of the ORFs are found on only one of the DNA strands and all of the viruses have at least one strong promoter that drives the expression of a terminally redundant mRNA. In addition, all of the viruses have a reverse transcriptase, and in most cases, it is located downstream from the coat protein (CP). Several active sites can be identified within the reverse transcriptase protein, including a protease, the core reverse transcriptase, and RNaseH activity. The reverse transcriptase of PVCV is distinguished from all other members of the family *Caulimoviridae* because it has the core features of an integrase function, in addition to other functions.

The genome organization of CaMV is the best characterized of the family and it is discussed in detail here. The CaMV genome is approximately 8000 bp in size and it contains three single-stranded discontinuities. One discontinuity occurs in the negative sense DNA strand and by convention, this is the origin of the DNA sequence. Two other discontinuities occur in the positive sense strand, one at nt position 1600 and a second at approximately nt position 4000. The virus genome consists of seven ORFs (Fig. 2(A)). ORF1 encodes a protein necessary for cell-to-cell movement (P1). ORFs 2 and 3 (proteins P2 and P3) are both required for aphid transmission. P3 also has an additional role in cell-to-cell movement. ORF4 encodes the CP, and ORF5 encodes a reverse transcriptase that also has protease and RNaseH domains. With the exception of *Petunia vein clearing virus* (PVCV), none of the reverse transcriptases of the caulimoviruses have evidence for an integrase function. The ORF6 product (P6) was originally described as the major inclusion body protein, but it also has a function as a translational transactivator (TAV). It physically interacts with host ribosomes to reprogram them for reinitiation of translation of the polycistronic 35S RNA. Furthermore, P6 is an important symptom and host range determinant. No protein product has been found for ORF7. Its nt sequence appears to play a regulatory role in aligning ribosomes for translation of other CaMV gene products.

Two transcripts are produced from the CaMV genomic DNA. The 19S RNA serves as the mRNA for P6, whereas the terminally redundant 35S RNA serves as a template for reverse transcription and as a polycistronic mRNA for ORFs 1–5. One feature common to many of the caulimoviruses is the ribosomal shunt mechanism of translation. The ribosomal shunt has undoubtedly evolved over time to compensate for the complexity and length of the leader sequence of the genomic RNA. In the case of CaMV, this leader sequence is approximately 600 bp in length and it contains up to nine short ORFs that vary in size from 9 to 102 nt. The complexity of the CaMV 35S RNA leader sequence is bypassed through the formation of a large stem-loop structure, which allows ribosomes to bypass most of the leader and to initiate translation at ORF7. CaMV utilizes several other strategies for expression of the 35S RNA, including splicing and reprogramming of host ribosomes by the TAV for re-initiation of translation.

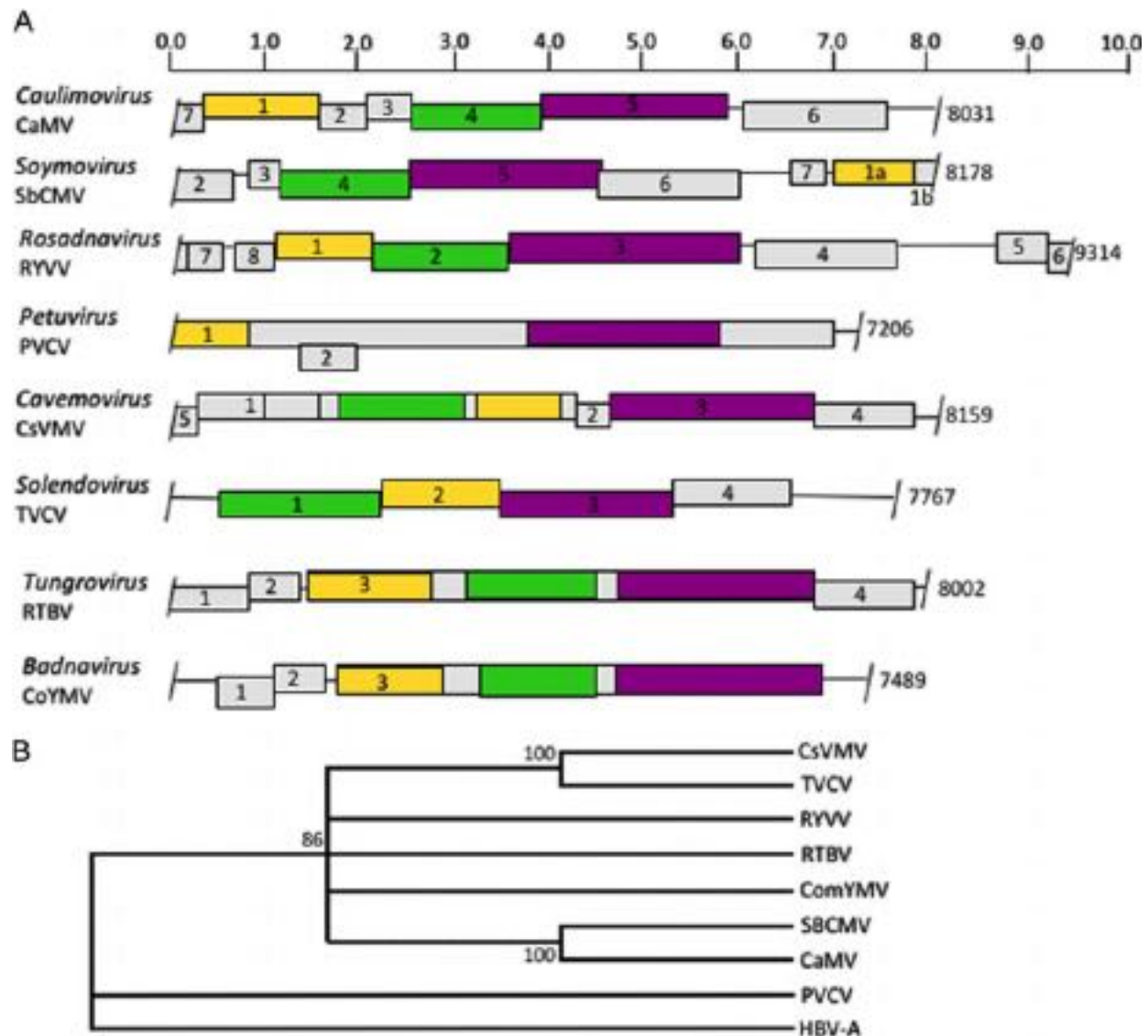


Fig. 2 Genome structure and phylogeny of members of the family *Caulimoviridae*. (A). Genome structure of representative viruses of the eight genera that comprise the family *Caulimoviridae*. Open reading frames (ORFs) are indicated by boxes, and are identified according to the information in the GeneBank nucleotide accession. The ORFs common to all genera include a movement protein (yellow), a capsid protein (green), and a protease/reverse transcriptase (purple). ORFs unique to a virus or whose function remains unidentified are in gray. The slashes on either end of the diagram (/) indicate where the genome is circular. The first base of the nucleotide sequence deposited in GenBank is located on the left, and the size of the genome is located on the right. Virus genera including *Petuvirus*, *Cavemovirus*, *Tungrovirus*, and *Badnavirus* produce polyproteins, which are presented as a single ORF. (B). Evolutionary relationships of the members of the family *Caulimoviridae*, based on the amino acid sequences of the reverse transcriptases. The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 3.32101427 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches. The evolutionary distances were computed using the p-distance method and are in the units of the number of amino acid differences per site. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 2020 positions in the final dataset. Evolutionary analyses were conducted in MEGA X. The outgroup for comparison is hepatitis B virus strain A (HBV-A).

Properties and Functions of CaMV Gene Products

Movement Protein (MP)

CaMV ORF-I encodes a 37 kDa protein dedicated to cell-to-cell movement. The CaMV movement protein (MP) dramatically reconfigures the host plasmodesma by removal of the desmotubule, leading to the formation of a tubule structure through which the virions move to adjacent cells. The MP is delivered to the plasma membrane through a vesicular transport pathway where it

makes its way to the plasmodesma. Excess MP is thought to be recycled back to the central vacuole. By contrast, the CaMV virion is thought to be delivered separately to the tubules by inclusion bodies formed from the P6 protein. Other plant viruses whose MPs are capable of forming tubules include the tospoviruses, comoviruses, nepoviruses and bromoviruses.

Aphid Transmission Factor (P2)

The product of ORF-II (or P2) is an 18 kDa protein that is required for transmission by aphids. P2 exists in cells in the form of inclusion bodies. When an aphid probes a leaf, this triggers the release of CaMV virions from the P6 inclusion bodies, the dissolution of the P2 inclusion bodies and the distribution of the P2 protein on host microtubules, where they come into contact with the virions. The N-terminus of the P2 protein is capable of binding to the aphid stylet, whereas the C-terminus of P2 binds to the CaMV P3 protein, which is located on the surface of virions. Consequently, CaMV virions are acquired by aphids through an association with the P2 and P3 proteins. Inactivation of the P2 protein abolishes aphid transmission but the mutant virus remains fully infectious through mechanical inoculation.

Virion-Associated Protein (P3)

The product of ORF-III, a 15 kDa virion-associated protein (P3) has key roles in both cell-to-cell movement and aphid transmission. P3 is found on the surface of virions and functions as a bridge between the virion and the MP and P2 proteins. The N-terminus of P3 binds to either to MP or to P2 to facilitate cell-to-cell movement or aphid transmission, respectively. Although P3 is located on the surface of virions, it is not uncommon for it to be released during purification of CaMV virions.

Capsid Protein (P4)

ORF-IV encodes the main capsid component (P4) has a molecular weight of 56 kDa and is processed into sizes of 37 and 44 kDa. The p37 protein is included within p44 and antibodies raised against virions react against both proteins. The p44 protein is phosphorylated by host casein kinase II, whereas p37 is not phosphorylated. Both proteins contain a Zn-finger motif, which is implicated in RNA binding. It is not clear which of p37 and p44 is the mature capsid protein and if one or both are required for full infection.

Protease/Reverse Transcriptase (P5)

The CaMV P5 protein is a polypeptide which is cleaved into a 15 kDa aspartic protease on the N-terminus and a 60 kDa reverse transcriptase/RNase H. With the exception of *Petunia vein clearing virus*, the reverse transcriptases of the caulimoviruses do not have an integrase function and do not integrate into host chromosomes as part of their replication cycle. Nonetheless, caulimovirus nt sequences are frequently found in the genomes of plants due to illegitimate recombination events.

Transactivator/Viroplasm (P6)

The P6 protein of CaMV has a size of 62 kDa and is a target for plant defenses and it plays a key role in multiple steps of the viral replication cycle. P6 is the matrix protein for the formation of the electron dense inclusion bodies (IBs) (**Fig. 1(A)**), which are the site for translation of the polycistronic 35S RNA, reverse transcription of the 35S RNA into DNA, and encapsidation of the viral DNA into virions. Nearly all of the CaMV virions accumulate and are retained within the P6 IBs, and this has led to its designation as a “virion factory”. The role of P6 in translation is especially noteworthy, because the P6 protein interacts with and modifies host ribosomes to facilitate translation of the 35S RNA, essentially reprogramming eukaryotic ribosomes to deal with a polycistronic message. This function, designated the translational transactivator (TAV), is unique in nature. New functions have recently been discovered for the P6 protein, including (1) movement of P6 inclusion bodies on microfilaments for delivery of virions to modified plasmodesmata for transit to adjacent cells, (2) modification of host defenses associated with SA and JA pathways, and (3) functioning as a gene silencing suppressor. Furthermore, P6 is an important pathogenic determinant. In susceptible hosts, P6 has been shown to be a prominent symptom determinant, responsible for the chlorosis symptom in turnips and a systemic necrosis symptom in *Nicotiana clevelandii*. P6 has also been shown to be a target for host resistance proteins present in *Nicotiana* species and *Arabidopsis*; recognition of the P6 protein by a host resistance protein triggers a cascade of plant defense responses that results in limitation of spread of CaMV. To date, P6 has been shown to physically interact with 14 host or viral proteins to carry out such diverse roles as translational transactivation, intracellular movement and silencing suppression.

Replication and Propagation

Caulimoviruses replicate by reverse transcription of an RNA intermediate, but integration of the viral DNA into host chromosomes is not required to complete the replication process. **Fig. 3** illustrates the replication of an icosahedral caulimovirus, but the same steps

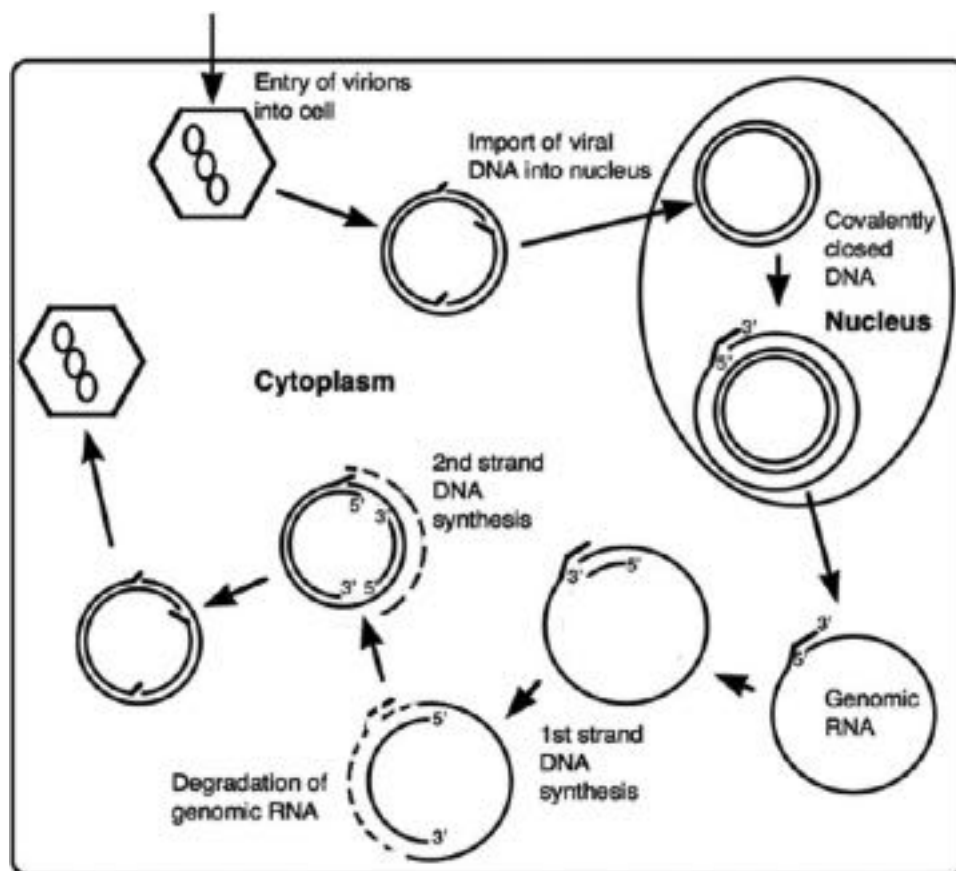


Fig. 3 Replication strategy of the members of the family *Caulimoviridae*.

are applicable for the bacilliform viruses. After virions enter a plant cell, the viral DNA becomes un-encapsidated and is transported into the nucleus. The viral DNA contains two to four single-stranded discontinuities, which form as a consequence of the reverse transcription process. Once in the nucleus, the single-stranded discontinuities are covalently closed and the DNA is associated with histones to form a minichromosome. The host RNA polymerase II is responsible for synthesizing an RNA that is terminally redundant; the sequence on the 5' end is reiterated on the 3' end. In the case of CaMV, the terminal redundancy is 180 nt in size, whereas the terminal redundancy in the CoYMV RNA is 120 nt. The terminally redundant RNA is transported out of the nucleus into the cytoplasm where it can either serve as a template for translation of viral proteins or as a template for reverse transcription. Reverse transcription is thought to occur in nucleocapsid-like particles.

First strand DNA synthesis is primed by a methionine (Met) tRNA that binds to a complementary sequence near the 5' end of the terminally redundant RNA. In the case of CaMV, the Met tRNA binds to a sequence approximately 600 nt from the 5' end of its 35S RNA. The virally encoded reverse transcriptase synthesizes DNA up to the 5' end of the terminally redundant RNA and the RNase H activity of the reverse transcriptase degrades the 5' end of the RNA. The terminal redundancies present in the genomic RNA provide a mechanism for a template switch, as the reverse transcriptase is able to switch from the 5' end of the genomic RNA to the same sequences on the 3' end of the genomic RNA and continue to synthesize the first strand of DNA.

The RNase H activity of the reverse transcriptase degrades the viral RNA template and small RNA fragments serve as the primers for second strand DNA synthesis. The RNA fragments bind to guanosine-rich tracts present in the first DNA strand, and these priming sites determine the positions of the discontinuities in the second DNA strand. A second template switch is required to bridge the gap from the 5' to the 3' end of the first DNA strand. The completion of the second DNA strand results in the formation of a dsDNA molecule that contains the characteristic single-stranded discontinuities.

Although caulimoviruses do not integrate into the host as part of their replication strategy, the sequences of several viruses have been detected in the genomes of their hosts. Furthermore, the integrated copies of several of these viruses, the Banana streak species, Tobacco vein clearing virus (TVCV) and PVCV, can be activated to yield episomal infections. All three viruses can form virions once episomal infections are initiated, but they differ in their capacity to be transmitted to other plants. Banana streak virus can be transmitted by mealybugs, whereas PVCV is transmitted only by grafting. TVCV and PVCV are transmitted vertically, through seed, but this probably involves only the integrated copies of the virus rather than the episomal forms.

The sites of integration are complex, as the viral sequences have undergone rearrangements, and multiple copies are arranged in tandem. Banana streak virus (BSV) is integrated at two loci in *Musa* chromosomes, TVCV is integrated at multiple loci in *Nicotiana edwardsonii* chromosomes, and PVCV is integrated into four loci in *Petunia hybrida* chromosomes. The mechanism that results in episomal infections remains to be elucidated, but there are some features common to all three viruses. First, the infections arise in interspecific hybrids. Consequently, each parent plant must contribute some factor to activate the integrated virus. Second, the infections arise after the hybrid has been exposed to some sort of stress. PVCV is activated when plants are exposed to water or nutrient stress, or the plants are wounded. TVCV infections arise in *N. edwardsonii* in the winter months in the greenhouse and are thought to be related to changes in light quality or duration. Integrated Banana streak virus sequences are activated in otherwise healthy *Musa* species when they are subjected to tissue culture. In each case, the virus is likely released through a series of recombination events or through reverse transcription of an RNA template.

Portions of other DNA viruses have been detected in plant chromosomes, but they have not been associated with episomal infections. This has led to speculation that plant DNA viruses might be capable of recombination in every infected plant, likely through a mechanism involving non-homologous recombination. However, since episomal forms of the caulimoviruses and badnaviruses have not been found in tissues fated for seed formation, the integrated forms would not appear in the next generation of plants. Consequently, the banana streak viruses, TVCV, and PVCV must have gained access to germline cells at some point such that their integrants would be passed through seed. This is likely to have been a relatively rare event.

Transmission, Host Range

Most members of the genus *Caulimovirus* are transmitted by as many as 27 species of aphids in a semi-persistent manner. The virions are acquired rapidly by the aphid and can be transmitted immediately upon acquisition. Virions can be maintained by the aphid for as little as 5 h up to three days, but are not retained after the aphid molts and are not passed on to aphid progeny. The protein products encoded by ORF-II (P2) and ORF-III (P3) are both required for aphid transmission. P3 forms a tetramer that binds to both the virion and to the C-terminus of P2. P2 is responsible for the binding of this complex to the aphid, as the N-terminus of P2 binds to a site in the aphid foregut. No insect vectors have been identified for species in the genera *Petuvirus*, *Rosadnavirus*, *Solendovirus*, *Soymovirus* or *Cavemovirus*.

None of the virions of the icosahedral viruses are transmitted through seed or pollen. Some viruses such as PVCV are transmitted from one generation to the next integrated into host chromosomes; at some point during the growth of the plant the integrated form of the virus is released from the chromosome and becomes capable of episomal replication. Most caulimoviruses are transmitted through mechanical inoculation, but there are a few exceptions, as PVCV, TVCV, Blueberry red ringspot virus, and Strawberry vein banding virus cannot be mechanically inoculated. In addition, most of these viruses are transmitted by grafting. In particular, Dahlia mosaic virus and Blueberry red ringspot virus infections in the field may be initiated through vegetative propagation or through grafting of infected plant material.

Badnaviruses are transmitted primarily by mealybugs, but a few are transmitted by aphids. Badnaviruses transmitted by mealybugs include the BSV species, the Sugarcane bacilliform viruses, Cacao swollen shoot virus (CSSV), Dioscorea bacilliform virus, Kalanchoe top spotting virus (KTSV), Piper yellow mottle virus (PYMoV), Taro bacilliform virus (TBV), and Schefflera ringspot virus. These viruses are transmitted in a semi-persistent manner and can be retained after molts, but do not multiply in the mealybug and are not transmitted to progeny. The badnaviruses shown to be transmitted by aphids include Gooseberry vein banding associated virus, Rubus yellow net virus (RYNV), Grapevine vein clearing virus, and Spiraea yellow leaf spot virus.

Other modes of transmission of the badnaviruses vary with the species. For example, CoYMV and KTSV can be mechanically inoculated, CSSV and PYMoV are mechanically inoculated with some difficulty, and RYNV and TBV have not been shown to be mechanically inoculated. The preferred method for inoculation of infectious badnavirus clones is agro-inoculation. Some of the badnaviruses are transmitted through seed, including the BSV, KTSV, and Mimosa bacilliform virus. KTSV is very efficiently transmitted by seed, with transmission rates from 60% to 90%, and is also transmitted in pollen.

RTBV is dependent on the RNA virus, RTSV, for its transmission. RTSV is vectored by a leafhopper, but causes only very mild symptoms. RTBV is only transmitted by the leafhopper in the presence of RTSV, but is responsible for the severe symptoms associated with the rice tungro disease. RTBV is not mechanically transmitted or carried in seed or pollen.

The host range of most caulimoviruses is fairly narrow, as in nature they generally infect plants within a single family (**Table 1**). Their experimental host range may extend to members of one or two other families, but in many instances, they may only infect a single genus of plants. There are a few exceptions. For example, the Sugarcane bacilliform viruses have a broader host range than most badnaviruses, as they can infect *Sorghum*, *Rottboellia*, *Panicum*, rice and banana. A limited host range can also be associated with similar limitations in the geographic distribution of the virus. For example, Soybean chlorotic mottle virus has only been recovered from a few samples in Japan. Perhaps the virus with the smallest geographic distribution is Stilbocarpa mosaic bacilliform virus, which has only been found on a single, small island in the subantarctic, midway between Tasmania and Antarctica. Other viruses, such as CaMV, Carnation etched ring virus and Dahlia mosaic virus are found worldwide, wherever their hosts are grown. Furthermore, the distribution of viruses that originate from integrated copies in their host's genomes, PVCV, TVCV, and the BSV species, are also closely aligned with the locations of their hosts. Interestingly, the icosahedral viruses of the caulimoviruses tend to infect hosts in temperate climates, whereas the bacilliform viruses of the badnavirus group are more likely to infect hosts in tropical or subtropical climates.

Virus-Host Relationships

Caulimoviruses induce a variety of systemic symptoms in their hosts, from chlorosis, streaking, and mosaics, to necrosis. The best-characterized pathogenicity determinant is the P6 protein of CaMV, as it has been shown to play a key role in the formation of chlorotic symptoms in turnips. This virulence function was first associated with P6 through gene-swapping experiments between CaMV isolates. It was confirmed when P6 was transformed into several species of plants, and in most cases, they exhibited virus-like symptoms. The P6 protein is also responsible for triggering systemic cell death in *Nicotiana clevelandii*, as well as a non-necrotic resistance response in *N. glutinosa*, and a hypersensitive resistance response in *N. edwardsonii*.

One feature that distinguishes the icosahedral viruses from the bacilliform viruses is that the former have the capacity to aggregate into amorphous inclusion bodies (Fig. 1(A)), whereas the latter do not form inclusions. The inclusion bodies formed by the icosahedral viruses are not bound by a membrane, can range in size from 5 to 20 μm , and occur in virtually all types of plant cells. The inclusions can be visualized with a light microscope in strips of epidermal tissue that has been stained with phloxine B. Close examination of CaMV inclusion bodies by electron microscopy reveals that there are actually two types. One type contains many vacuoles and consists of an electron-dense, granular matrix that is composed primarily of the P6 protein (Fig. 1(A)). A second, electron translucent type is made up of the P2 protein, a protein required for aphid transmission. Both types of inclusions are thought to have a role in the biology of the virus. The vacuolated inclusion bodies may be considered pathogen organelles, as they are thought to serve as the sites for replication of the viral nucleic acid, as well as translation of the 35S RNA and assembly of the virions. The electron translucent inclusions are considered to have a role in aphid transmission.

A second feature characteristic of the caulimoviruses is that the plasmodesmata of infected cells are enlarged enough to accommodate the virions, as electron micrographs have revealed the presence of CaMV and CoYMV virions in the enlarged plasmodesmata. For both CaMV and CoYMV, the alteration in size is mediated by their proteins required for cell-to-cell movement. In infected protoplasts, the CaMV P1 protein has been shown to induce the formation of tubular structures that extend away from the protoplast surface. It is hypothesized that virions are assembled in the cell and then are escorted to the enlarged plasmodesmata by the cell-to-cell movement proteins.

Further Reading

- Bak, A., Gargani, D., Macia, J.-L., *et al.*, 2013. Virus factories of Cauliflower mosaic virus are virion reservoirs that engage actively in vector transmission. *Journal of Virology* 87, 12207–12215.
- Bhat, A.I., Hohn, T., Selvarajan, R., 2016. Badnaviruses: The current global scenario. *Viruses* 8, 177. doi:10.3390/v8060177.
- Geering, A.D.W., Maumus, F., Copetti, D., *et al.*, 2014. Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. *Nature Communications* 5, 5269. doi:10.1038/ncomms6269.
- Hull, R., 1996. Molecular biology of rice tungro viruses. *Annual Review of Phytopathology* 34, 275–297.
- Krupovic, M., Blomberg, J., Coffin, J.M., *et al.*, 2018. *Ortervirales*: New virus order unifying five families of reverse-transcribing viruses. *Journal of Virology* 92, e00515–e00518.
- Leisner, S.M., Schoelz, J.E., 2018. Joining the crowd: Integrating plant virus proteins into the larger world of effectors. *Annual Review of Phytopathology* 56, 89–110.
- Lockhart, B.E.L., 1990. Evidence for a double-stranded circular DNA genome in a second group of plant viruses. *Phytopathology* 80, 127–131.
- Mollov, D., Lockhart, B., Ziesak, D.C., Olszewski, N., 2013. Complete nucleotide sequence of Rose yellow vein virus, a member of the family caulimoviridae having a novel genome organization. *Archives of Virology* 158, 877–880.
- Pooggin, M.M., Ryabova, L.A., 2018. Ribosome shunting, polycistronic translation and evasion of antiviral defenses in plant pararetroviruses and beyond. *Frontiers in Microbiology* 9, 459. doi:10.3389/fmicb.2018.00459.
- Richert-Pöggeler, K.R., Shepherd, R.J., 1997. Petunia vein-clearing virus: A plant pararetrovirus with the core sequences for an integrase function. *Virology* 236, 137–146.
- Schoelz, J.E., Leisner, S., 2017. Setting up shop: The formation and function of the viral factories of Cauliflower mosaic virus. *Frontiers in Plant Science* 8, 1832. doi:10.3389/fpls.2017.01832.
- Shepherd, R.J., Wakeman, R.J., Romanko, R.R., 1968. DNA in Cauliflower mosaic virus. *Virology* 36, 150–152.
- Staginnus, C., Richert-Pöggeler, K.R., 2006. Endogenous pararetroviruses: Two-faced travelers in the plant genome. *Trends in Plant Science* 11, 485–491.
- Uzest, M., Gargani, D., Drucker, M., *et al.*, 2007. A protein key to plant virus transmission at the tip of the insect vector stylet. *Proceedings of the National Academy of Sciences of the United States of America* 104, 17959–17964.

Cheraviruses, Sadwaviruses and Torradoviruses (*Secoviridae*)

Toru Iwanami, Tokyo University of Agriculture, Tokyo, Japan

Rene AA van der Vlugt, Wageningen University and Research Center, Wageningen, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

CP Coat protein or capsid protein

ELISA Enzyme-linked immunosorbent assay

MP Movement protein

nm Nanometer(s)

NTB Nucleotide-binding protein

PCR Polymerase chain reaction

Pol RNA-dependent RNA polymerase

PRO Protease

Pro-Co Protease cofactor;

Pro-glu Glutamic protease

RT-PCR Reverse transcription polymerase chain reaction

Vpg Viral protein genome-linked

Glossary

Enzyme-linked immunosorbent assay A commonly used detection method of any protein that uses specific antibodies against the protein. All viruses have proteins on their particle surface, and are applicable for ELISA, if specific antibodies are prepared and mixed with an enzyme that accelerate colorization in the presence of the viruses.

Reverse transcription polymerase chain reaction A laboratory technique that consists of reverse transcription of

an RNA molecule into a DNA molecule and subsequent amplification of the DNA molecule using polymerase chain reaction (PCR). Specific amplification of the DNA molecule is ensured by using a set of specific short DNA molecules (designated “primers”) that are complementary to the target DNA.

Classification

The genera *Cheravirus*, *Sadwavirus*, and *Torradovirus* are taxa in the subfamily *Comovirinae* within the family *Secoviridae*. Each genus is distinguished from the other genera by its specific genomic organization, biological properties and phylogenetic relations. The member species of the genus *Cheravirus* are *Apple latent spherical virus* (ASLV), *Arracacha virus B* (AVB), *Cherry rasp leaf virus* (CRLV), *Currant latent virus* (CLV), and *Stocky prune virus* (StPV). The member species of the genus *Sadwavirus* is *Satsuma dwarf virus* (SDV). The member species of the genus *Torradovirus* are *Carrot torradovirus 1* (CaTV1), *Lettuce necrotic leaf curl virus* (LNLCV), *Motherwort yellow mottle virus* (MYMoV), *Squash chlorotic leafspot virus* (SCLSV), *Tomato marchitez virus* (ToMarV) and *Tomato torrado virus* (ToTV) (Table 1 and Fig. 1).

Virion Structure

Cheraviruses, sadwaviruses, and torradoviruses have the same virion structure as other viruses in the family *Secoviridae*. Virions are non-enveloped, and 25–30 nm in diameter. Cheraviruses and torradoviruses have three capsid proteins. Sadwaviruses have two capsid proteins. All cheraviruses, sadwavirus, and torradoviruses have two genomic RNAs that are encapsidated in separate virions. Many virus preparations also contain empty particles.

Genome Organization

The genome organization of cheraviruses, sadwaviruses, and torradoviruses consists of two single-stranded (+)-sense RNA molecules and their genomic organization is similar to that of the other viruses in the subfamily *Comovirinae*. The distinguishing features in the genome of cheraviruses are three capsid protein genes, and a comparatively small RNA2 that apparently encodes a single polyprotein. The unique features in the genome of sadwaviruses are the extensive amino acid sequence similarity in the N-terminal regions of the polyproteins encoded by RNA1 and RNA2. A unique feature of the torradovirus genome is the presence of a small open reading frame (ORF1) that is located upstream of, and partially overlaps with the large ORF2 in RNA2. The protein potentially encoded by RNA2-ORF1 shows no apparent homologies with other known proteins (Fig. 2).

Table 1 List of all ICTV classified members of the genera *Cheravirus*, *Sadwavirus* and *Torradovirus* in the family *Secoviridae*, their taxonomic order, acronyms and GenBank sequence information. # Seq - number of sequences for each virus registered in NCBI Taxonomy Browser data from which a complete genome has been assembled to give a RefSeq assembly number (GCF_number). The last column to the right provides explanation if no RefSeq assembly number is available in the form of accession numbers for incomplete viral sequences. Type species are written in bold

Subfamily	Genus	Subgenus ^a	Species	Acronym	# Seq	RefSeq assembly	GenBank accession
Unassigned	<i>Cheravirus</i>	–	<i>Apple latent spherical virus</i>	ALSV	4	GCF_000861545	
			<i>Arracacha virus B</i>	AVB	14	GCF_000907115	
			<i>Cherry rasp leaf virus</i>	CRLV	10	GCF_000859565	
			<i>Currant latent virus</i>	CuLV	6	GCF_001503195	
			<i>Stocky prune virus</i>	StPV	4	–	DQ143874, DQ143875
	<i>Sadwavirus</i> ^c	<i>Satsumavirus</i>	<i>Satsuma dwarf virus</i>	SDV	33	GCF_000860985	
		<i>Stramovirus</i>	<i>Strawberry mottle virus</i>	SMoV	83	GCF_000850445	
			<i>Black raspberry necrosis virus</i>	BRNV	26	GCF_000867325	
		<i>Choliavirus</i>	<i>Chocolate lily virus A</i>	CLVA	4	GCF_000895075	
			<i>Dioscorea mosaic associated virus</i>	DMAV	4	GCF_001876835	
	<i>Torradovirus</i>	–	<i>Carrot torradovirus 1</i>	CaTV1	10	GCF_000927615	
			<i>Lettuce necrotic leaf curl virus</i>	LNLCV	4	GCF_002219685	
			<i>Motherwort yellow mottle virus</i>	MYMoV	4	GCF_002220005	
			<i>Squash chlorotic leaf spot virus</i>	SCLSV	35	GCF_002219665	
			<i>Tomato marchitez virus</i>	ToMarV	25	GCF_000879575	
			<i>Tomato torrado virus</i>	ToTV	148	GCF_000872965	

^asubgenera organization of the *Sadwavirus* genus has yet to be ratified by the ICTV.

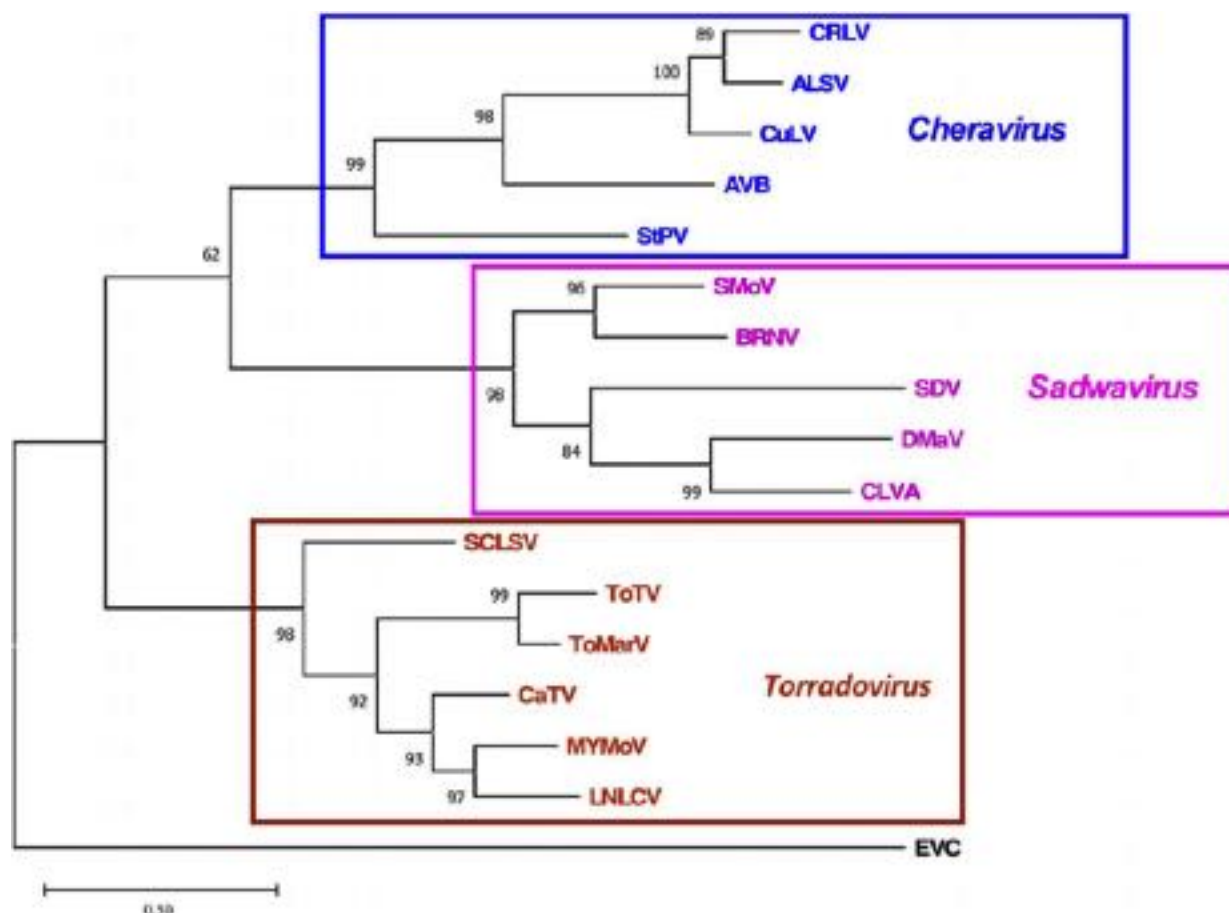


Fig. 1 Maximum-likelihood phylogenetic analysis of the cheraviruses, sadwaviruses, and torradoviruses using the amino acid sequence between protease domain and RNA polymerase domain in the RNA1 polyprotein. The percentage of trees in which the associated viruses clustered together in the bootstrap test (100 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The amino acid sequence between protease domain and RNA polymerase domain in the RNA1 polyprotein of Enterovirus C (EVC) was used as an outgroup. For acronyms of the other viruses, see Table 1.

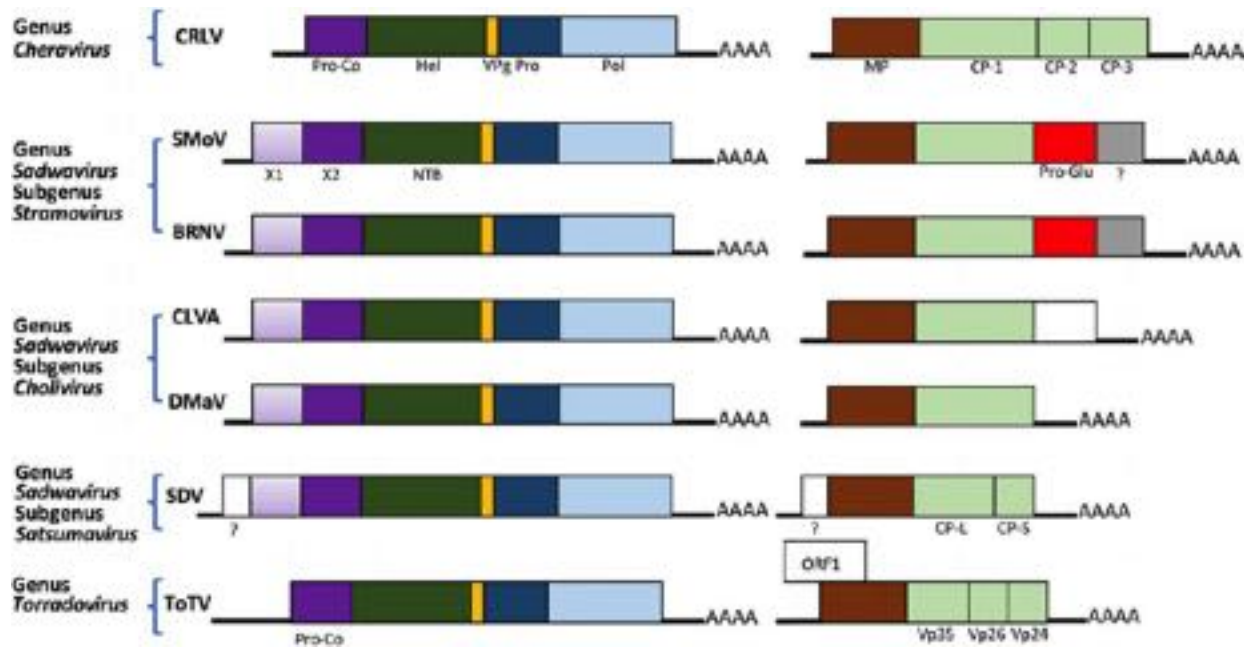


Fig. 2 Genomic organization of viruses in the genera *Cheravirus*, *Sadwavirus*, and *Torradovirus*. Schematic representation and the size of the genomes not reflecting their actual sizes. The figure is re-drawn from the figures published previously (Petrzik, *et al.*, 2016, Sanfacion *et al.*, 2009, van der Vlugt *et al.*, 2015) with modifications made by the authors. The same color box indicates the same domains in the polyprotein. For SMOV, the genomic organization of Canadian isolate SMOV-NsPer3 is shown here. The German isolate SMOV-1134 lacks the C-terminal protein domain of the RNA2 polyprotein. Abbreviations: CP, coat protein; CP-L and CP-S, large and small components of coat protein; MP, movement protein; NTB, nucleotide-binding protein (putative helicase); Pol, RNA dependent RNA polymerase; Pro, 3C-like protease; Pro-Co, protease cofactor; Pro-glu, glutamic protease; Vp-35, Vp-26 and Vp-24, viral protein (=coat protein) of about 35 kDa, 26 kDa, and 24 kDa; VPg, viral genome-linked protein. X1 domain was first reported in nepovirus and the function has not been characterized. X2 was also first reported for nepoviruses and it has sequence homologies with the Pro-Co of comoviruses. VPgs of viruses in the genera *Cheravirus*, *Sadwavirus*, and *Torradovirus* are putative.

Life Cycle and Epidemiology

Host Range

Natural hosts of cheraviruses include pome fruits, stone fruits, raspberry and currant as well as potato. The experimental host range of cheraviruses is broad or narrow, depending on the virus and strain. For example, CRLV, a cheravirus infects commercially cultivated cherry, apple, and potato, as well as wild herbaceous including dandelion, balsam root, and plantain. SDV, the type species of the genus *Sadwavirus* infects mainly citrus and occasionally *Daphniphyllum teijsmannii* and *Viburnum odoratissimum* var. *awabuki*. Torradoviruses were described to infect tomato or other crops like lettuce, carrot, squash and cassava, depending on virus species and strains. Little is known about other hosts but Tomato torrado virus (ToTV), the type species, infects over 20 weed species belonging to the family *Amaranthaceae*, *Chenopodiaceae*, *Solanaceae* etc., and the weeds apparently are ToTV sources for transmission to tomato (Figs. 3 and 4).

Geographic Distribution

Cheraviruses have been described from different countries worldwide; Apple latent spherical virus (Japan), Arracacha virus B (Peru), Cherry rasp leaf virus (the United States), Currant latent virus (the Czech Republic), and Stocky prune virus (France). Satsuma dwarf virus (*Sadwavirus*) occurs mostly in Japan but has been found also in China, Korea and Turkey. Tomato-infecting torradoviruses occur in some countries in Europa, in Australia, as well as some parts of North, Central and South America. LNLCV was found in the Netherlands and a CaTV1 torrado-like virus was found in the United Kingdom, France and Germany. MYMoV was detected only in Korea, while SLCV was detected only in Sudan.

Transmission

Among cheraviruses, CRLV is transmitted by a nematode (*Xiphinema americanum*) in the field and is seed-transmitted. Apple latent spherical virus, another cheravirus, is transmitted through seeds obtained from infected trees, and rarely through pollen. No vector



Fig. 3 Foliar symptoms of satsuma dwarf induced on satsuma mandarin. Satsuma mandarin showing small leaves with shortend internodes.



Fig. 4 Foliar symptoms induced on tomato with infection of ToTV. Necrotic spots surrounded by a light green or yellow area on young leaflets of ToTV-infected tomato. Photo courtesy by Martin Verbeek, Wageningen University and Research; copyright, Wageningen University and Research.

transmission was demonstrated for any sadwavirus, although field observation strongly indicated a soil-borne nature. Seed transmission of sadwavirus in citrus has not been tested, presumably because the main host of sadwavirus is a seedless satsuma mandarine, however it has been demonstrated in an experimental leguminous host. Tomato-infecting torradoviruses are the first spherical viruses reported to be transmitted by whiteflies (*Trialeurodes vaporariorum*, *Bemisia tabaci*, and *T. abutilonea*). No vector is known yet for carrot- and lettuce-infecting torradoviruses but these viruses are spreading in the open field, suggesting vector-involvement. Seed transmission of torradoviruses is very rare, if it occurs. Seed transmission of torradoviruses is still uncertain.

Clinical Features and Pathogenesis

Cheraviruses usually induce mild or no symptoms. Sadwavirus causes various foliar symptoms including smaller leaves, boat-shaped leaves with shortend internodes that lead to serious reduction in fruit quality and yield in citrus. Sadwavirus symptomlessly infects some wild trees including *D. teijsmannii* and *V. odoratissimum* var. *awabuki* near citrus fields. Some torradoviruses induce severe necrosis (in Spanish “torrado” meaning burned or roasted) on tomato leaves and fruits. The overall growth of the infected tomato plant is reduced and serious economic damage is inflicted. Torradoviruses infect other vegetables including carrot and lettuce, as well as herbs and wild plants.

Diagnosis

Cheraviruses, sadwaviruses, and torradoviruses are readily detected by RT-PCR. Generic RT-PCT that detect a wide range of viruses are available for cheravirus, sadwaviruses and tomato-infecting torradoviruses. Detection by ELISA using specific antibodies is possible for some cheraviruses, sadwaviruses and tomato-infecting torradovirus viruses.

Treatment

No treatment is available.

Prevention

Cheraviruses and sadwaviruses mainly infect fruit trees. In this case, movement of contaminated budwoods plays an important role in the long-distance dissemination. For example, occurrence of SDV in China, Korea and Turkey is obviously related to un-checked budwoods from Japan. Severe restriction on such practice is the first step of prevention. Some whitefly-transmitted torradoviruses affect tomato in greenhouses. In this case, damage would be alleviated by preventing whiteflies by suitable screens or sticky traps. Some level of genetic resistance appears to be available against ToTV. Possible sources of resistance for the other torradoviruses are not known.

References

- Petrzik, K., Koloniuk, I., Příbylova, J., Špak, J., 2016. Complete genome sequence of currant latent virus (genus *Cheravirus*, family *Secoviridae*). *Archives of Virology* 161, 491–493.
- Sanfaçon, H., Wellink, J., Le Gall, O., *et al.*, 2009. Secoviridae: A proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. *Archives of Virology* 154, 899–907.
- van der Vlugt, R.A.A., Verbeek, M., Dulleman, A.M., *et al.*, 2015. Torradoviruses. *Annual Review of Phytopathology* 53, 485–512.

Further Reading

- Adams, I.P., Glover, R., Souza-Richards, R., *et al.*, 2013. Complete genome sequence of Arracacha virus B: A novel cheravirus. *Archives of Virology* 158, 909–913.
- Candresse, T., Svanella-Dumas, L., Le Gall, O., 2006. Characterization and partial genome sequence of Stocky prune virus, a new member of the genus *Cheravirus*. *Archives of Virology* 151, 1179–1188.
- Iwanami, T., 2010. Properties and control of Satsuma dwarf virus. *Japan Agricultural Research Quarterly* 44, 1–6.
- James, D., Cieslinska, M., Pallás, V., *et al.*, 2017. Viruses, viroids, phytoplasmas and disorders of cherry. In: Quero-García, J., Iezzoni, A., Pulawska, J., Lang, G. (Eds.), *Cherries, Botany, Production and Uses*. Oxfordshire: CAB International, pp. 386–419.
- Kenten, R.H., Jones, R.A.C., 1979. Arracacha virus B, a second isometric virus infecting arracacha (*Arracacia xanthorrhiza*, Umbelliferae) in the Peruvian Andes. *Annals of Applied Biology* 93, 31–36.
- Koganezawa, H., Ito, T., 2011. Apple latent spherical virus. In: Hadidi, A., Barba, M., Candresse, T., Jelkmann, W. (Eds.), *Virus and Virus-Like Disease of Pome and Stone Fruits*. St. Paul, MN: American Phytopathological Society, pp. 23–24.
- Petrzik, K., Koloniuk, I., Příbylova, J., Špak, J., 2016. Complete genome sequence of currant latent virus (genus *Cheravirus*, family *Secoviridae*). *Archives of Virology* 161, 491–493.
- Rubio, M., Martínez-Gómez, P., Marais, A., *et al.*, 2017. Recent advances and prospects in Prunus virology. *Annals of Applied Biology* 171, 125–138.

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/secoviridae/583/genus-cheravirus
Cheravirus.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/secoviridae/584/genus-sadwavirus
Sadwavirus.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/secoviridae/585/genus-torradovirus
Torradovirus.

Citrus Tristeza Virus (*Closteroviridae*)

Moshe Bar-Joseph, Agricultural Research Organization, Volcani Center, Bet Dagan, Israel

Scott J Harper, Washington State University, Prosser, WA, United States

William O Dawson, Citrus Research and Education Center, Lake Alfred, FL, United States and University of Florida, Lake Alfred, FL, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of M. Bar-Joseph, W.O. Dawson, Citrus Tristeza Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00639-7](https://doi.org/10.1016/B978-012374410-4.00639-7).

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

dRNAs Defective RNAs

ELISA Enzyme-linked immunological assays

ER Endoplasmic reticulum

kb Kilobase

kDa Kilodalton

LAMP Loop mediated amplification

LMT Low molecular weight tristeza

MP Movement protein

nt Nucleotide(s)

ORF Open Reading Frame

PDR Pathogen derived resistance

RdRp RNA-dependent RNA polymerase

sgRNA Sub-genomic RNA

SL Stem-and-loop structures

SP Stem pitting

SY Seedling-yellow reaction

UTR Untranslated region

VLPs Virus-like particles

VPg Viral protein genome-linked

VRC Virus replication complexes

Glossary

Cross-protection Prevention of the symptomatic phase of a disease by prior inoculation with a mild or non-symptomatic isolate of the same virus.

Defective RNA Subviral RNA molecules lacking parts of the genome while maintaining the signals enabling their synthesis by the viral replication system.

Genome The complete genetic information encoded in the RNA of the virus, including both translated and non-translated sequences.

Sub-genomic RNA Shorter than full length genomic RNA molecules produced during the replication process to allow translation of open reading frames.

Transgenic pathogen-derived resistance Plants genetically transformed to harbor viral or other pathogen derived sequences expected to confer resistance against related virus isolates.

Introduction

The virions of Citrus tristeza virus (CTV) are long flexuous particles ($2,000 \times 12$ nm) with an unusual architecture that results from two different coat proteins, the major coat protein (CP) and minor coat protein (CPm) encapsidation approximately 97% and 3%, respectively, of the helical structures. Two additional viral proteins, p61 and p64 (HSP 70 homolog) are required for efficient assembly. The 19.3 kb single stranded RNA genome is the longest among the plant viruses. Genomic analysis showed a division into twelve open reading frames (ORFs), the organization of which is conserved among all isolates. The ORF1a encodes a 349 kDa polyprotein containing two papain-like protease domains plus methyltransferase-like and helicase-like domains. Translation of the polyprotein is thought occasionally to continue through the polymerase-like domain (ORF1b) by a +1 frameshift. Ten 3' ORFs are expressed by 3' co-terminal sub-genomic messenger RNAs. Infected cells contain a myriad of other sub-genomic RNAs. In addition, most CTV isolates contain one or more defective RNAs of different sizes and genomic rearrangements. Among the CTV translation products, at least 3 function as RNA silencing repressors, which are serving to protect the extremely large size CTV genome and correlated with unusual difficulty in conferring effective pathogen derived resistance to transgenic citrus plants.

This article is essentially reporting all the basic elements of the knowledge about CTV including CTV population genomic studies, as well as the use of bioinformatic tools, and the development of the CTV gene cloning vector, that allows elucidating the functional genomics of both CTV and its citrus hosts.

Although CTV is still considered to be among the most important plant viruses, the attention of the world's citrus growers to the significance of CTV research has apparently diminished. We may attribute this decline in interest, at least in part, to the efficacy of control measures introduced in modern citriculture which are based on (1) CTV-tolerant or resistant rootstocks; and (2) effective selection of mild CTV strains that are providing commercially acceptable cross-protection against severe stem-pitting CTV isolates. However, despite

the general decline in CTV research, the major reason that supports recent CTV research is related to the catastrophic damage experienced during the last 15 years by citrus growers in Brazil and Florida as a result of infection with “*Candidatus Liberibacter asiaticus*” bacteria. Finding effective means for controlling this bacterium, the causal agent of huanglongbing/greening (HLB) disease, is currently the focus of intense research efforts. Nevertheless, CTV-related problems have not been completely eliminated, and the danger of spread of severe CTV isolates, once the brown citrus aphid reaches the Mediterranean and Californian citrus, remains a continuous threat.

History

Over the last 90 years *Citrus tristeza virus* (CTV) has killed or rendered unproductive millions of trees throughout most of the world's citrus growing areas and hence it is rightfully considered as the most important virus of the world's largest fruit crop, citrus, hence the name “tristeza” which means “sadness” in Spanish and Portuguese. However, as with many other disease agents, the actual damage of CTV infections and the timings varied considerably at different periods and geographical regions. The origins of CTV infections remain unknown; the virus however existed for centuries in Asia as an unidentified disease agent, but growers in these areas adapted citrus varieties and rootstock combinations with resistance or tolerance to CTV infections. Out of Asia, citrus production moved to the Mediterranean region, primarily through the introduction of fruit and propagation of seed which does not transmit CTV to the resulting plantings; hence, the new cultivation areas started from seed sources remained free of the virus for centuries. However, the improvement in maritime transportation, allowing long distance transport of rooted citrus plants – led to the outbreak of the deadly citrus root rot disease caused by *Phytophthora sp.*, which was managed by adapting the more tolerant sour orange rootstock. The considerable horticultural advantages offered by this rootstock coincided with the large-scale expansion of plantings throughout the Mediterranean and American countries, and as a result citrus production in many areas was almost entirely based on that single rootstock. This decision had grave subsequent effects when CTV pandemics swept throughout the world, causing “quick decline” (death) of trees on this rootstock. Although the CTV problems were first noticed in South Africa and Australia where, by the end of the 19th century, the growers have found that sour orange was an unsuitable rootstock, it was only in the 1930s that the extent of this deadly disease problem manifested itself first in Argentina and shortly later in Brazil and California, with the death of millions of trees.

The virus-like nature of the tristeza-related diseases was demonstrated experimentally by graft and aphid transmission of the disease agent in 1946; however, description of CTV properties began with the seminal findings of thread-like particles associated with tristeza-infected trees. These unusually long and thin particles presented a challenging problem for isolation (purification), which was an essential step towards virus characterization. Development of effective purification methods and biophysical characterization of CTV and similar viruses led to their assignment to the *Closterovirus* group of elongated viruses. The association of infectivity with the thin-particle-enriched preparations was demonstrated early on; however, the unequivocal completion of Koch's postulates was only completed in 2001 with the mechanical infection of citrus plants with RNA transcripts of an infectious CTV cDNA clone amplified through serial passaging in *Nicotiana benthamiana* protoplasts.

Development of CTV purification methods paved the way to antibody preparations and to improved CTV diagnosis by ELISA which demonstrated considerable serological diversity among CTV isolates. Later, RNA extracts from CTV-enriched particle preparations were used for molecular cloning of cDNA molecules, which when used as probes confirmed considerable genomic variation among CTV isolates. Nucleic acid probes also demonstrated that plants infected with CTV and other closteroviruses contained many defective RNAs (dRNAs) in addition to the normal genomic and sub-genomic RNAs. Infected plants also contain unusually large amounts of dsRNAs corresponding to the genomic and sub-genomic RNAs. Because of the ease of purifying these abundant dsRNAs, they often have been the template of choice for producing cDNAs. The advent of cDNA cloning of CTV led to the sequencing of its genome and to description of the defected (d)RNAs.

Taxonomy and Classification

CTV belongs to genus *Closterovirus*, family *Closteroviridae*. The *Closteroviridae* family contains more than 30 plant viruses with flexuous, filamentous virions and with either mono- or bipartite single-stranded (one tripartite) positive sense RNA genomes. A recent revision of taxonomy of the *Closteroviridae* based on vector transmission and on phylogenetic relationships using three proteins highly conserved among members of this family (a helicase, an RNA dependent RNA polymerase, and a homolog of the HSP70 heat-shock proteins) led to defining of three genera: *Closterovirus* including CTV and other aphid-borne viruses with monopartite genomes, *Ampelovirus* comprising viruses with monopartite genomes transmitted by mealybugs, and *Crinivirus* that includes whitefly-borne viruses with bipartite or tripartite genomes. Closteroviruses are all phloem-limited viruses. Among the hallmarks of the group is the presence of a conserved five-gene-module that includes the four proteins involved in assembly of virions, the major (CP) and minor (CPm) coat proteins, p61, and p64, a homolog of the HSP70 chaperon. In addition, infected phloem-associated cells have clusters of vesicles considered specific to closterovirus cytopathology (Fig. 1).

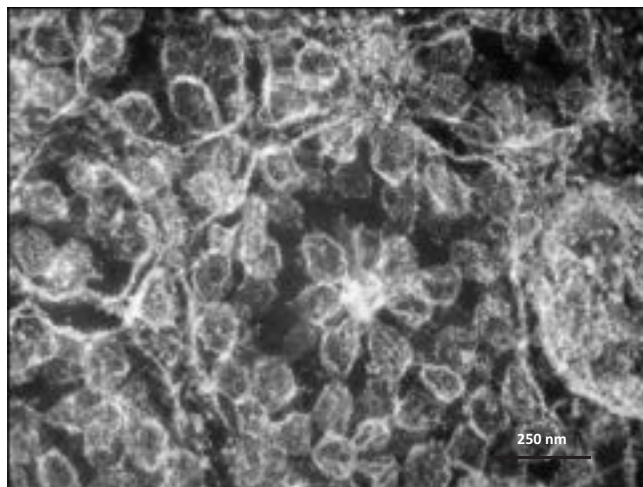


Fig. 1 An electron micrograph of a thin section from a Mexican lime (*Citrus aurantifolia*) young leaf vein showing typical symptoms of vein clearing following infection with the Floridian T36 isolate of Citrus tristeza virus. Note the clusters of vesicles typical of infections by closteroviruses. The bar represents 250 nm.

Geographic Distribution

Citrus is a common fruit crop in areas with sufficient rainfall or irrigation from the equator to about 41 degrees of latitude north and south. CTV is now endemic in most of the citrus growing areas, with only a few places in the Mediterranean basin and Western USA remaining free of CTV infections. The recent spread of the brown citrus aphid (*Toxoptera citricida*) to parts of Portugal and Spain are threatening most of the remaining CTV-free areas of the Mediterranean basin. One important aspect of CTV geographic distributions is that some area that have endemic isolates of CTV still do not have the most damaging stem-pitting isolates. With the spread of the vector comes the threat of extremely severe CTV isolates which so far have not spread to North America and Mediterranean countries.

Economic Costs of CTV

Control strategies varied at different geographic regions and periods, depending on the extent to which CTV was spreading within the newly infested areas, the sensitivity of the specific local varieties, and the economical availability of alternative crops to replace diseased groves. Most citrus growing countries that managed to remain free of CTV did so by enforcing sanitation practices to prevent virus spread. These included certification schemes aimed to ensure that citrus nurseries propagate and distribute to growers only CTV-free planting material. Other areas had far more costly and ambitious programs of eradication. These efforts were mostly only temporarily successful, mainly because of lack of long-term grower and governmental commitment to such costly operation. In areas where CTV was endemic, alternative strategies to enable continued commercial citrus production in the presence of CTV were developed. Indeed, long before the viral nature of the tristeza disease was realized, Japanese citrus growers were grafting CTV-tolerant Satsuma mandarins on the cold-tolerant and CTV-resistant trifoliate orange (*Poncirus trifoliata*) rootstock, thus enabling them to produce quality fruits despite endemic CTV infections. Similarly, change of rootstocks from the “quick declining” sour orange rootstocks to rough lemon rootstocks, and to less extent mandarin rootstocks, allowed the South African growers to produce good crops of oranges despite the presence of the most efficient aphid vector and infections with severe CTV strains. Similarly, half a century later the Brazilian citrus industry that was completely decimated during the 1940s, when all trees were grafted on sour orange rootstocks, was saved by adopting CTV-tolerant rootstocks and mild strain cross-protection.

Historically the costs of CTV epidemics were estimated to be of the order of tens and even hundreds of millions of dollars. These estimates, however, varied depending on actual market value of the lost production capacity, alternative uses of the land, and the time needed for the CTV-tolerant replants to enter production. There were also indirect costs resulting in poorer performance and/or sensitivity of some CTV-tolerant rootstocks used to replace the widely adapted sour orange rootstocks to other citrus diseases such as citrus blight and citrus sudden death.

Host Range and Cytopathology

CTV infects all species, cultivars and hybrids of *Citrus sp.*, some citrus relatives such as *Aeglopsis chevalieri*, *Afraegle paniculata*, *Fortunella sp.*, *Pamburus missionis* and some intergeneric hybrids. Species of *Passiflora* are the only non-Rutaceous hosts, infected naturally. The CTV decline isolates in trees on sour orange rootstocks (Fig. 2) are associated with the death of the phloem near the



Fig. 2 Symptoms of CTV on different citrus species: (A). decline of sweet orange on sour orange rootstock, (B & C) seedling yellows on Duncan grapefruit.

bud union, resulting in a girdling effect that may cause the overgrowth of the scion at the bud union, loss of feeder roots and thus drought sensitivity, stunting, yellowing of leaves, reduced fruit size, poor growth, dieback, wilting, and death. However, other virulent and damaging CTV strains cause stem pitting (SP), which results in pits in the wood under depressed areas of bark and are often associated with severe stunting and considerably reduced fruit production. The seedling-yellow reaction (SY) which includes severe stunting and yellowing on seedlings of sour orange, lemon and grapefruit is primarily a disease of experimentally inoculated plants but might also be encountered in the field.

CTV infection is closely restricted to the phloem tissues of infected plants that first show chromatic cells filled with either aggregates of virus particles and/or necrotized cytoplasm. Interestingly, the death of trees on the sensitive sour orange rootstock is not associated with increased amounts of chromatic cells, thus suggesting a different patho-physiological condition lead to the chromatic type of cell death, which could also be observed in tissues of sweet orange trees which are tolerant to CTV when growing on their own roots.

Nicotiana benthamiana has become an important experimental host for CTV since it was discovered that plants become systemically infected by *Agrobacterium tumefaciens* containing recombinant DNA plasmids of CTV (agro-inoculation).

The CTV Virions

The CTV virions are long flexuous particles, 2000 nm long $\times$ 10–12 nm in width (**Fig. 3**). The virions have helical symmetry with a pitch of 3–4 nm, about 8–9 capsids per helix turn, and a central hole of 3–4 nm. Unlike the virions of other elongated plant viruses that consist of cylindrical nucleocapsids made of a single CP, the CTV virions consist of bipolar helices with a long body and short tail. Immuno-electron microscopy showed antibodies to CPm attached to only one end of the virions, with the major part (>97%) of the virion is encapsidated by CP. Interestingly, the “tail” corresponds to the 5' end region of the viral genome and the tails of other closteroviruses have been associated with small amounts of p61 and HLP60h.

The CTV Genome

Fig. 4 shows a schematic presentation of the 19.3 kb CTV RNA. Generally, the CTV genome is divided almost equally into two parts, the 5' part consisting of ORF1a and 1b harboring the viral replication machinery and the 3' end half harboring 10 ORFs



Fig. 3 Particles of CTV, with a diameter of 10–12 nm, negatively stained with Uranyl acetate. Note their flexible thread like shape, considerably limiting the possibility of their isolation from infected citrus tissues. The bar represents 50 nm.

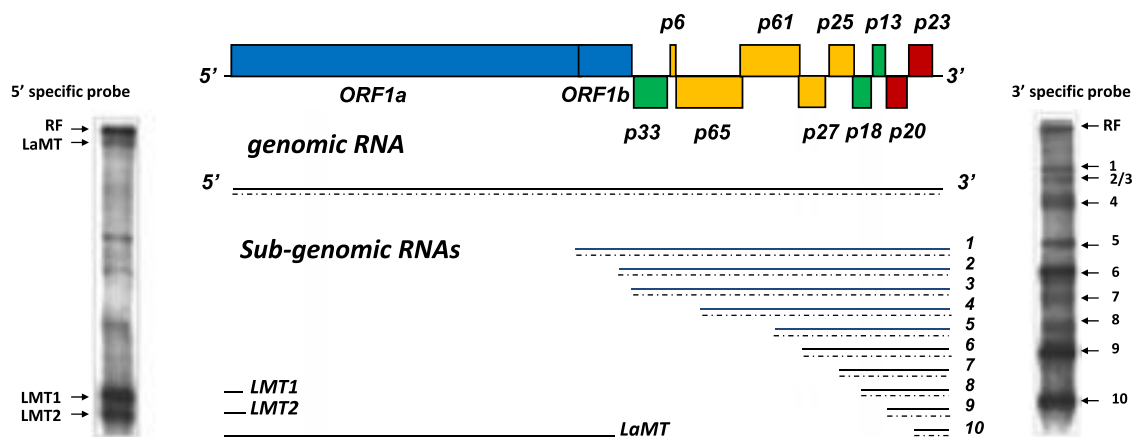


Fig. 4 A schematic presentation of the citrus tristeza virus (CTV) genomic (g) and subgenomic (sg) RNAs. The replication-associated ORFs are highlighted in blue, the five-gene structural and assembly block common to all closteroviruses in yellow, suppressors of silencing in red, and three genes on undetermined function in green. Lines shown on the left and right side below the genomic map indicate 5' large molecular single stranded (ss) 5' sgRNA (LaMT) and two low molecular ssRNAs (LMT1 and LMT2) and a nested set of 3' co-terminal sgRNAs, respectively. The left insert shows Northern blot hybridization of dsRNA enriched extracts from an Alemow plant infected with the VT strain of Citrus tristeza virus (CTV-VT) using Riboprobes specific to the 5' end of the viral RNA. Note the presence of the large replicative form (RF) molecules (upper band), LaMT and two abundant LMT molecules. The right insert shows hybridization of dsRNA with a 3' UTR probe. Note the hybridization bands of the different 3' sgRNAs indicated by arrows.

encoding a range of structural components and other gene products involved in virion assembly and host and vector interactions. CTV replicons containing only ORFs 1a and 1b plus the 5' and 3' non-translated ends (**Fig. 4**) fulfill all the requirements for efficient replication in protoplasts.

Interestingly, while the sequences of the 3' half of all CTV isolates that have been sequenced are 97% and 89% identical, the 5' half sequences often differ considerably more. For instance, the isolates T36 and VT show only 60%–70% identities for their 5' halves.

The Untranslated Regions (UTRs)

A remarkable feature of CTV isolates are the close identities (97%) of their 3' NTR primary sequences compared to the considerable divergence of their 5' UTR sequences. Computer assisted calculations, however, suggest that not only the 3' UTRs of different isolates fold into similar predicted structures, but surprisingly the dissimilar sequences of the 5' NTRs also are predicted to fold into similar secondary structures. The 3' replication signal of several CTV isolates was mapped to 230 nt within the UTR and predicted to fold into a secondary structure composed of 10 stem-and-loop (SL) structures. Three of these SL regions and a terminal 3' triplet, CCA, are essential for replication. Replacement of T36 3' UTR with 3' UTRs from other strains allowed

replication, albeit with slightly less efficiency, suggesting on the significance of the primary sequence of this part. The plus-strand sequence of the 5' UTRs from different strains are predicted to form similar secondary structures consisting of two stem-loops (SL1 and SL2) separated by a short spacer region, that were essential for replication and initiation of assembly by CPm.

Gene Functions

In infected cells, the 12 ORFs of CTV (**Fig. 4**) are expressed through a variety of mechanisms including: proteolytic processing of the polyprotein, translational frame shifting, and production of 10 3' co-terminal sub-genomic RNAs. The first two mechanisms are used to express proteins encoded by the 5' half of the genome while the third mechanism is used to express ORFs 2–11. The ORF1a encodes a 349 kDa polyprotein containing two papain-like protease domains plus methyltransferase-like and helicase-like domains. Translation of the polyprotein could also continue through the polymerase-like domain (ORF1b) by a +1 frameshift. These proteins along with the signals at the 5' and 3' ends of the genome are the minimal requirements for replication of the RNA. The function of p33 (ORF2) is required for superinfection exclusion and is required for infection of a subset of the viral host range. The next five gene products from the 3' genes include the unique signature block characteristic of closteroviruses, which consist of the small, 6 kDa hydrophobic protein (ORF3), 65 kDa cellular heat-shock protein homolog (HSP70h, ORF4), 61 kDa protein (ORF5), and the tandem pair of p27 (CPm, ORF6) followed by p25 (CP, ORF7). The four proteins with the exception of p6 are required for efficient virion assembly. The small hydrophobic p6 is a single-span trans-membrane protein not required for virus replication or assembly, but it is required for infection of all plants of the host range. CP is also a suppressor of RNA silencing. The function of p18 (ORF8) and p13 (ORF9) are required for infection of certain hosts. Protein p20 (ORF10) accumulates in amorphous inclusion bodies of CTV-infected cells and has been shown to be a suppressor of RNA silencing. p23 (ORF11) is a multifunctional protein with no homolog in other closteroviruses, that: (1) binds cooperatively both single-stranded and dsRNA molecules in a non-sequence-specific manner (2) contains a zinc finger domain that regulates the synthesis of the plus- and minus-strand molecules and controls the accumulation of plus-strand RNA during replication (3) is an inducer of CTV-like symptoms in transgenic *C. aurantifolia* plants and (4) is a potent suppressor of intracellular RNA silencing in *Nicotiana tabacum* and *N. benthamiana* and (5) controls the level of negative-stranded of genomic and sub-genomic RNAs. For additional p33 functions, see also Virus-Virus, interactions.

The CTV Sub-Genomic RNAs

The replication of CTV involves the production of large number of less than full-length RNAs. These include ten 3' co-terminal sg mRNAs, and ten negative-stranded sgRNAs corresponding to the ten 3' sg mRNAs, plus ten 5' co-terminal sgRNAs that apparently are produced by termination just 5' of each of the 10 ORFs (**Fig. 4**). The amounts of different sg mRNAs vary with the highest levels for sg mRNAs p23 and p20, located at the distal 3' end. Infected cells also contain abundant amounts of two other small 5' co-terminal positive-stranded sgRNAs of ~600 and 800 nt designated as low molecular weight tristeza (LMT).

Defective RNAs

Most CTV contain defective RNAs (dRNAs), which consist of the two genomic termini, with extensive internal deletions, resulting in different sizes of 5' and 3' sequences. CTV dRNAs accumulate abundantly in infections with wild type virus. The dRNAs contain different sizes of 5' and 3' sequences. The biological roles of CTV dRNAs remain obscure as no clear cut associations have been established with any of the dRNA.

Virus-Virus Interactions

The CTV isolates were categorized as strains based on biological properties. Major genetic differences between CTV isolates were reported using a CTV specific cDNA probe of the VT strain. Cross-protection is a practice in which a mild population of CTV is purposely pre-inoculated into trees to prevent infection and disease caused by endemic isolates. Cross-protection has been widely used to allow commercial citrus production in regions with severe endemic isolates of CTV. It works well when an effective protecting isolate is found, but this process is empirical and difficult. Most mild isolates fail to protect and finding the correct combination has consumed whole careers. The successful protecting isolate must affect a virus-virus interaction such that the protecting isolate interferes with establishment of infection and induction of disease symptoms. A component of this process is superinfection exclusion. It has been shown that previously established infections of CTV can completely exclude superinfection by another isolate. However, this is limited to isolates from the same strain of CTV. Challenges with isolates of different strains of CTV are not inhibited. This superinfection exclusion is dependent on the p33 gene and/or the 5' protease. For example, as the full-length virus can completely exclude an isolate from the same strain, CTV with the p33 gene deleted does not. Thus, one component of finding or creating a protecting virus against a pathogenic endemic isolate is to identify the sequence group (strain) of the cause of disease, challenging isolates and choose or create a mild isolate from the same strain.

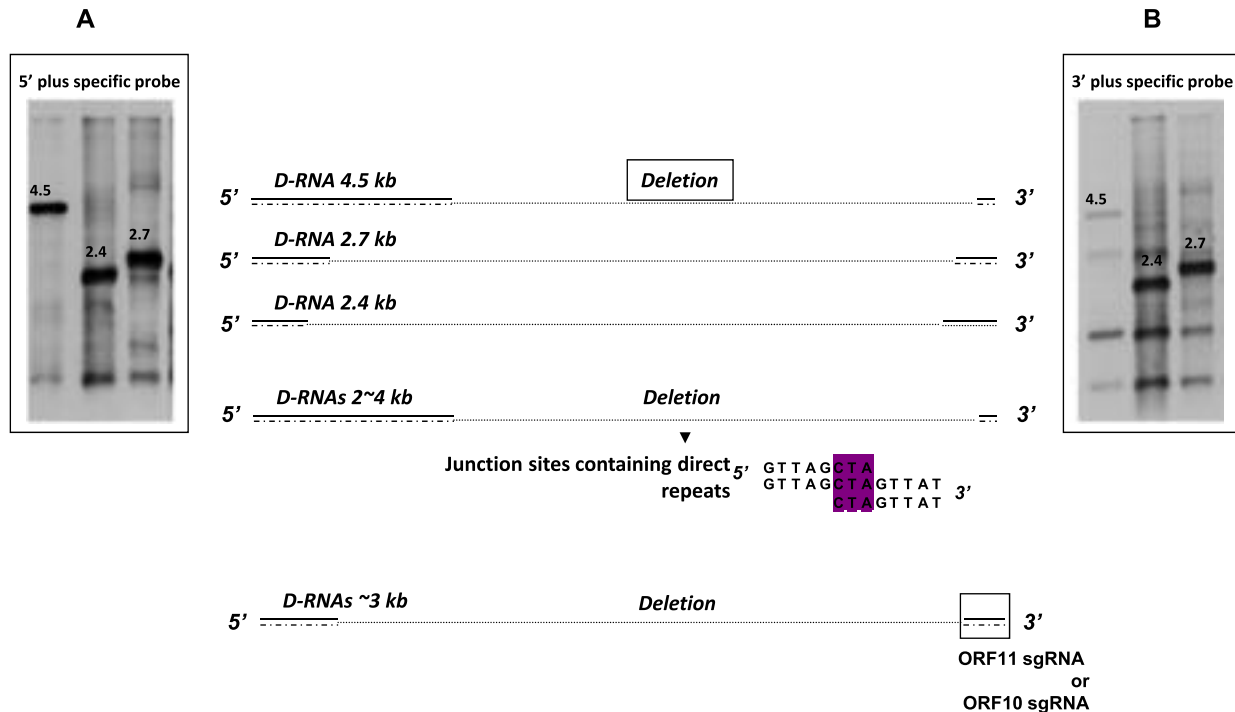


Fig. 5 A diagram of class 1 CTV defective (d) RNAs with three different sizes and class 2 CTV-dRNAs that show 3' moieties of the size and structure of the full-length sgRNA of ORF11. Inserts A and B show Northern hybridization of class 1-dsRNAs from three different CTV-VT sub isolates, hybridized with Riboprobes specific for the 5' and 3' ends of the CTV gRNA, respectively. Note the intense bands resulting of the hybridization with both probes of the abundantly present dRNAs. Lower part is a diagrammatic presentation of direct repeats at the junction sites of some class 1 dRNAs. The bottom part shows a schematic presentation of a class 2 dRNA, with a 3' terminus corresponding to the ORF11 sgRNA of CTV-VT.

However, since most endemic isolates are populations made up of components from several strains, in practice it is not that simple because so far effective protecting isolates also are population. Thus, this requires an understanding of the rules of populations. For example, a typical population might consist of 50% strain A, 20% strain B, 29% strain C, and 1% strain D.

CTV has at least seven characterized genotypes, defined by nucleotide sequence divergence of ~10%–20% from each other. **Fig. 5** shows a Phylogenetic tree of 60 CTV isolates belonging to different CTV strains (circles). These CTV genotypes differ in their infectivity in different citrus species, tropism and ability to move systemically, vector transmissibility and to some extent, pathogenicity. It should be noted, however, that there is variability in biological properties even between isolates or variants of the same genotype (**Fig. 6**).

As a virus with multiple, distinct genotypes, and a long-lived host system, it is unsurprising CTV isolates from field trees almost always consist of populations of multiple genotypes in mixture. The existence of multiple genotypes within an individual plant, and co-infection of the same cells also explains the frequency of recombination observed for CTV isolates. The CTV genotypes within individual plants both compete and interact with one another, reaching equilibrium based on their relative fitness vis-à-vis one another, and the host. A major rule of equilibrium of mixtures of genotypes in a population is determined by the host, which could explain why cross-protection often breaks down upon utilization of new citrus varieties. Also, individual CTV genotypes or variants are also capable of interacting via complementation to increase fitness and overcome tropism limitations as well as increasing the efficacy of aphid transmission.

Transmission

Although the virus is phloem-limited, mechanical inoculation can be done experimentally at relatively low efficiency by slashing the trunks of small citrus trees with blades containing sap extracts. In commercial groves, the virus is spread naturally by aphid vectors and by vegetative propagation in infected budwood. Long distance spread of the virus, particularly from country to country, had been through the long-distance transfer of plant propagation materials.

Several aphid species including *Aphis gossypii*, *A. spiraeicola*, *A. craccivora*, and *T. citricida* transmit CTV semi-persistently. The rate of transmissibility varies considerably between different virus isolates and different aphid species. Transmission efficacy of *T. citricida* is approximately 6 to 25 times more efficient than *A. gossypii*, and different CTV isolates or genotypes can differ dramatically, for example 40%–60% for the type isolate of strain T68 to less than 1% for type isolate or strain T36. However, it

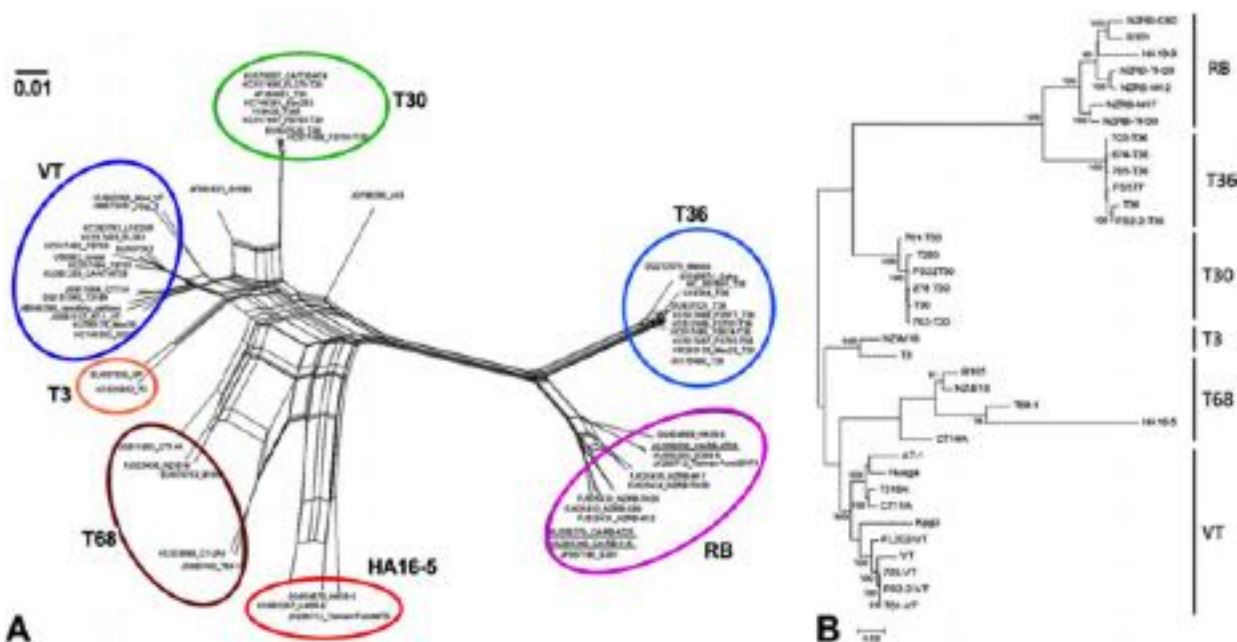


Fig. 6 (A) Phylogenetic tree and evolutionary analysis of 60 full-length citrus tristeza virus (CTV) genome sequences using the maximum-parsimony method. The CTV isolates belonging to different CTV strains are indicated by circles, with permission from the author Yokomi, R.K., Selvaraj, V., Maheshwari, Y., *et al.*, 2017. Identification and characterization of citrus tristeza virus isolates breaking resistance in trifoliate orange in California. *Phytopathology* 107, 901–908. (B) A phylogeny of displaying the major Citrus tristeza virus (CTV) genotypic groupings, generated using the maximum-parsimony method from complete genomic sequences.

should be noted that other isolates of the T36 strain are much more highly transmitted and the type isolate was originally isolated from a tree infected by aphid transmission. Apparently, more than 50 years of graft transmission of the type isolate of T36 resulted in loss of aphid transmission. Transmission efficacy can also be affected by the interaction of CTV genotypes in mixture, where the presence of highly transmissible variants can increase the transmission rate of poorly transmissible variants through complementation.

For the semi-persistent mode of transmission, the virion has been found to bind to the cibarium and foregut of the aphid vector via interaction of the p27, p61 and p65 proteins. Additionally, the p33 protein has also found to be involved in enhanced transmission.

CTV Control Measures

Strategies to control CTV have varied at different locations and periods and included (1) quarantine and budwood certification programs to prevent the introduction of CTV, (2) costly and ambitious eradication programs to contain situations of virus spread, (3) the use of CTV-tolerant rootstocks and mild (or protective) strain cross-protection, often also named preimmunization, (4) breeding for resistance, and (5) attempts to obtain resistance by genetic engineering. Mild strain cross-protection has been widely applied for millions of citrus trees in Australia, Brazil, and South Africa to protect against stem pitting of sweet orange and grapefruit trees. However, mild strain cross-protection has not provided field protection against CTV isolates causing quick decline of trees on the sensitive sour orange rootstock.

Useful resistance to CTV has been found in a citrus relative, *Poncirus trifoliata*. This resistance has been mapped and shown to be controlled primarily by a single genome region and sequenced. However, a specific gene has not been identified by transformation into susceptible citrus resulting in a resistance phenotype.

CTV Diagnosis

The outcome of CTV infections varies considerably depending the virulence of the prevailing virus isolates, and sensitivity of infected citrus varieties and rootstock combinations. Hence, the need for diagnosis is of utmost importance. Biological indexing of the disease by grafting sensitive citrus indicators has been the definitive assay, although it is costly and time consuming. It has largely been replaced with more rapid immunoassays (ELISA) that have been widely practiced for almost 30 years for a variety of CTV sanitation programs. Development of recombinant antigens considerably advanced diagnostic possibilities and production of monoclonal antibodies has allowed for more precise differentiation of isolates. PCR and combinations of immuno-capture PCR allow more sensitive CTV diagnosis. Recently, high-

throughput sequencing has proven an effective tool for the rapid identification of CTV isolates and populations. However, despite the development of better detection tools, none are effective for predicting the biological properties of new CTV isolates.

Attempts to Apply Transgenic Resistance to CTV

Conventional breeding for CTV resistance is a long, difficult, and inconvenient process mainly because most citrus varieties are hybrids of unknown parentage. Hence the considerable interest in applying pathogen derived resistance (PDR) to render both citrus rootstocks and varieties resistant to stem pitting isolates and decline causing CTV isolates. Yet, literally hundreds of independent transformations with a range of different configurations of CTV sequences has resulted in failure to obtain durable protection against CTV. These failures are especially frustrating since the control of severe stem pitting isolates is needed in several major citrus producing. Several reports on transgenic resistance to CTV infection including effective protection of plants expressing the three CTV silencing suppressor genes need still validation under field conditions. The lack of RNA silencing against CTV in citrus might be due to more effective suppression by the combination of the three suppressors of this virus.

Epilogue

As indicated in the preface, in much of the world's citrus-growing regions, CTV is now considered a manageable disease. Nevertheless, research is still needed to address several important questions such as the spread of severe CTV isolates and the brown citrus aphid into regions previously free of them, and a re-emergence of CTV as serious economic plant pathogen of citrus. We still do not know the origin of CTV; did it originate from one of the many wild relatives of domesticated citrus, or from another plant family and subsequently adapted to citrus?

CTV and its host interactions especially in Pummello, one of the three ancestors of citrus also remains to be elucidated, especially as both the highly CTV sensitive sour orange and the tolerant sweet oranges share both Pummello and mandarins as parents. Similarly questions of CTV genome variation are worth attention noting the considerable variation between the 5'-terminal regions of the genomes of T36 and VT strains raises the possibility of a long-time separation from the original source, does the parental strain still survive and, if so, could the related genome re-emerge as a new pathogen based on its genome and biology?

Finally, it is interesting to note that recent molecular studies of phloem-limited bacteria, such as the greening agent, indicated a drastic reduction in their genome size, compared to the majority of non-phloem-limited bacteria; whereas, CTV, with the largest genome of the family *Closteroviridae*, behaves just the opposite. The question thus remains why and how CTV adopted such a different survival strategy within the phloem vessels of its host plants.

See also: Closteroviruses (*Closteroviridae*). Plant Viral Diseases: Economic Implications

Further Reading

- Albiach-Martí, M.R., 2013. Chapter 1 – The complex genetics of citrus tristeza virus. In: Romanowski, V. (Ed.), *Current Issues in Molecular Virology – Viral Genetics and Biotechnological Applications*. InTech, pp. 1–25.
- Atta, S., Zhou, C., Zhou, Y., *et al.*, 2012. Distribution and research advances of citrus tristeza virus. *Journal of Integrative Agriculture* 11, 346–358.
- Bar-Joseph, M., Marcus, R., Lee, R.F., 1989. The continuous challenge of citrus tristeza virus control. *Annual Review of Phytopathology* 27, 291–316.
- Catara, A.F., Bar-Joseph, M., Licciardello, G., 2019. *Citrus Tristeza Virus: Methods and Protocols*. Springer. (ISBN 978-1-4939-9558-5).
- Chen, A.Y.S., Watanabe, S., Yokomi, R., Ng, J.C.K., 2018. Nucleotide heterogeneity at the terminal ends of the genomes of two California citrus tristeza virus strains and their complete genome sequence analysis. *Virology Journal* 15, 141.
- Dawson, W.O., Bar-Joseph, M., Garnsey, S.M., Moreno, P., 2015. Citrus tristeza virus: Making an ally from an enemy. *Annual Review of Phytopathology* 53, 137–155.
- Flores, R., Moreno, P., Falk, B., Martelli, G.P., Dawson, W.O., 2013. E-book on *Closteroviridae*. *Frontiers in Microbiology* 4, 411. doi:10.3389/fmicb.2013.00411. (PMID: 24409172).
- Folimonova, S.Y., 2013. Developing an understanding of cross-protection by citrus tristeza virus. *Frontiers in Microbiology* 4, 76.
- Harper, S.J., 2013. Citrus tristeza virus: Evolution of complex and varied genotypic groups. *Frontiers in Microbiology* 4, 93.
- Karasev, A.V., Hilf, M.E. (Eds.), 2010. *Citrus Tristeza Virus Complex and Tristeza Diseases*. APS Press. (SBN: 978-0-89054-378-8).
- Moreno, P., Ambrós, S., Albiach-Martí, M.R., Guerri, J., Peña, L., 2008. Citrus tristeza virus: A pathogen that changed the course of the citrus industry. *Molecular Plant Pathology* 9 (2), 251–268.
- Satyanarayana, T., Bar-Joseph, M., Mawassi, M., *et al.*, 2001. Amplification of Citrus tristeza virus from a cDNA clone and infection of Citrus trees. *Virology* 280, 87–96.
- Wallace, J.M., 1978. Virus and virus like diseases. In: Reuther, W., Calavan, E.C., Carman, G.E. (Eds.), *The Citrus Industry*, vol. 4. Berkeley, CA: Div. Agric. Sci. Univ. Calif., pp. 67–184.
- Yokomi, R.K., Selvaraj, V., Maheshwari, Y., *et al.*, 2018. Molecular and biological characterization of a novel mild strain of citrus tristeza virus in California. *Archives of Virology* 163, 1795–1804.

Closteroviruses (*Closteroviridae*)

Marc Fuchs, Cornell University, Geneva, NY, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa amino acid(s)
AlkB alpha-ketoglutarate-dependent hydroxylase domain
CP coat protein or capsid protein
CPm minor coat protein
ELISA enzyme-linked immunological assays
ER endoplasmic reticulum
HC-Pro helper component-protease
HEL helicase
HSP70h heat shock protein 70 homolog

kb kilobase
kDa kilo dalton
LAMP loop mediated amplification
L-Pro leader papain-like protease
MET methyltransferase
MP movement protein
nt nucleotide(s)
ORF open reading frame
RdRp RNA-dependent RNA polymerase
UTR un-translated region

Glossary

Ampelovirus One of the four genera in the family *Closteroviridae*. The name ampelo derives from the Greek *ampelos*, "grapevine", the natural host of grapevine leafroll-associated virus 3, the type species of the genus *Ampelovirus*.
Closterovirus Another genus in the family *Closteroviridae*. The name *clostero* derives from the Greek *kloster*, "spindle, thread" due to the size and appearance of very long, flexuous, and helically constructed filamentous virions.
Crinivirus It is another genus in the family *Closteroviridae*. The name *crini* derives from the Latin *crinis*, "hair", for the appearance of the very long thread-like particles.

RNA-dependent RNA polymerase An enzyme that catalyses the transcription of RNA from an RNA template.
tRNA-like structure An RNA sequence with a similar tertiary structure to transfer RNA.
Velarivirus It is the most recently identified genus in the family *Closteroviridae*. The name *velari* derives from the Latin *velari*, "cryptic or veiled", because *Grapevine leafroll-associated virus 7*, the type species of this genus, does not cause any apparent disease symptoms on its natural host.

Introduction

Plant viruses of the family *Closteroviridae* belong to the higher phylum *Kritrinoviricota* in the realm *Riboviria*. Their common features include the morphology and size of virions, the genome structure and expression, the presence of canonical conserved proteins, characteristic cytopathic structures, tissue tropism, transmission by specific hemipteran insects in a semi-persistent manner, and extreme difficulties at mechanical transmission to herbaceous hosts (Table 1).

Virions have a diameter of about 12 nm and lengths ranging from 650 nm for species with a fragmented genome to over 2000 nm for species with a monopartite genome (Table 2). Virions are helically constructed filaments with a pitch of the primary helix in the range of 3.4–3.8 nm, containing about 10 protein subunits per turn of the helix and showing a central hole of 3–4 nm (Fig. 1). Two types of coat protein (CP), the major CP and a CP analog or minor CP (CPm), are the most abundant protein components involved in the formation of most virions with CPm encapsidating the 600–700 nt 5'-terminal of the viral RNA and coating one extremity (75–100 nm) of the particles, thus forming a distinct structure referred to as rattlesnake or bipolar. The virus-encoded heat shock protein 70 homolog (HSP70h) of cell chaperones and the ~60 kDa proteins are also integral virion components. The ~60 kDa protein is required for incorporation of both HSP70h and CPm to virions, which also incorporate a 20 kDa protein that may form the tip segment of the virion head.

Species demarcation criteria used for all genera in the family *Closteroviridae* are particle size, size of the CP, genome structure and organization, vector species and specificity, cytopathological features, host range, as well as aa sequence of the RNA-dependent RNA polymerase (RdRp), CP, and HSP70h differing by more than 25%.

Classification

Plant viruses in the family *Closteroviridae* are divided into four genera: *Closterovirus* (monopartite genome), *Ampelovirus* (monopartite genome), *Velarivirus* (monopartite genome) and *Crinivirus* (bipartite or tripartite genome) (Table 2). Collectively these viruses are referred to as closteroviruses, the generic name for viruses in the family that was initially coined for this group of plant viruses by

Table 1 Common features among members of the family *Closteroviridae*

<i>Virions</i>	<i>Genome structure</i>	<i>Genome organization</i>	<i>Genome expression</i>	<i>Cytopathic structures</i>	<i>Tissue tropism</i>	<i>Transmission</i>	<i>Mechanical inoculation</i>
Long, helically constructed filamentous particles (650–2200 nm in length and 12 nm in diameter)	A very large positive sense, single-stranded RNA (12,000 to nearly 19,000 nt) with an organization distinct from those of other plant viruses	Two conserved gene modules encoding proteins associated with replication with a close phylogenetic relationship (ORF1a and 1b) and a quintuple gene module encoding proteins associated with replication, virion formation, cell-to-cell movement and suppression of RNA silencing; and a unique set of genes coding a heat shock protein 70 homolog, and for a duplicated, diverged copy of the coat protein (minor coat protein, CPm)	Proteolytic processing, +1 ribosomal frameshift, and the production of sub-genomic RNAs	Cytoplasmic aggregates of virions intermingled with single or clustered membranous vesicles	Phloem restriction with occasional location in the mesophyll and epidermis	Aphids, whiteflies, mealybugs and soft scale insects	Extremely difficult for a extremely limited number of virus species

Table 2 Distinguishing properties of the four genera in the family *Closteroviridae*

Genus	Virion length (nm)	RNA species (#)	Genome size (kb)	ORF (No.)	Replicase (kDa)	HSP70h (kDa)	CP (kDa)	CPm (kDa)	Vector
<i>Closterovirus</i>	1350–2000	1	14.5–19.3	8–12	349–367	65–67	22–25	24–27	Aphids
<i>Ampelovirus</i>	1400–2200	1	13.0–18.5	7–12	245–293	57–59	28–36	50–56	Mealybugs, soft scales
<i>Crinivirus</i>	650–850 and 700–900	2 or 3	7.8–9.1/7.9–8.5 8.0–5.3–3.9	9–13	267–280	62–65	28–29	53–80	Whiteflies
<i>Velarivirus</i>	1500–1700	1	16.2–16.9	9	260–270	62–69	34–46	69–76	None known

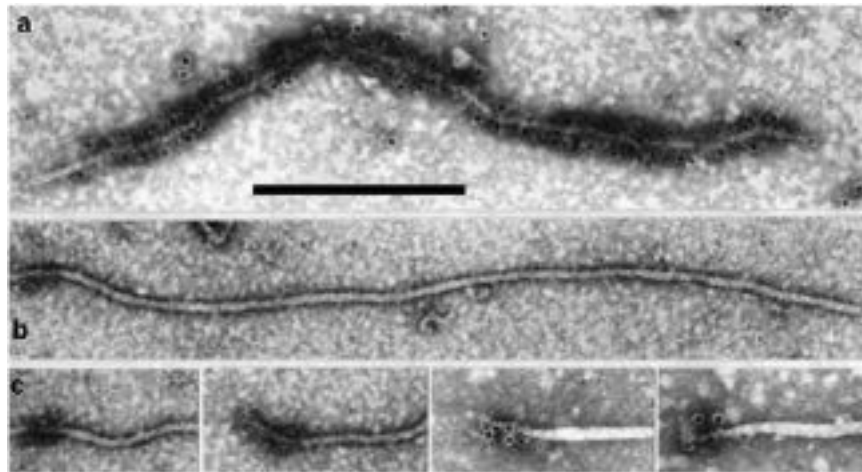


Fig. 1 Electron micrographs of virions of Beet yellows virus (BYV) from the genus *Closterovirus* that are negatively stained and decorated with an antiserum specific to the (a) coat protein (CP) -note that the CP minor (CPm) tail is not decorated-, and (b) CPm at the 75 nm tail of a virus particle. (c) Close-up of the 75 nm tail of four selected BYV particles decorated with an antiserum specific to CPm. Scale bars represent 300 nm. From Agranowsky, A.A., Lesemann, D.E., 2000. Beet yellows virus. Descriptions of Plant Viruses. Association of Applied Biologists. Available at: <http://www.dpvweb.net/dpv/showdpv.php?dpvno=377>, with permission.

the International Committee on Taxonomy of Viruses in 1976. In addition to classified viruses in the family, a few species are unassigned members in the family *Closteroviridae*: *Actinidia virus 1*, *Alligatorweed stunting virus*, *Blueberry virus A*, *Megakepasma mosaic virus*, *Mint vein banding-associated virus*, *Olive leaf yellowing-associated virus*, and *Persimmon virus B*.

The genus *Closterovirus* comprises 14 species with particles ranging from 1350 to 2000 nm in length, and a monopartite, positive-sense, single-stranded RNA genome, 14.5–19.3 kb in size, in which the CPm is uniquely located upstream of the CP. The natural transmission of closteroviruses is by aphids (**Table 2**). *Beet yellows virus* (BYV) is the type species of the genus *Closterovirus*.

The genus *Ampelovirus* comprises 13 species with particles ranging from 1400 to 2000 nm in length, and a monopartite positive-sense, single-stranded RNA genome, 13.0–18.5 kb in size. A wide variation in genome size and organization is typical for species in this genus. The natural transmission of ampeloviruses is by pseudococcid mealybugs and soft scale insects (**Table 2**). The type species of the genus *Ampelovirus* is *Grapevine leafroll-associated virus 3* (GLRaV-3).

The genus *Crinivirus* comprises 13 species transmitted by whiteflies. Virions usually have two modal lengths (650–850 and 700–900 nm) and a bipartite positive-sense, single stranded RNA genome with a size ranging from 7801 to 9127 nt (RNA-1) and from 7903 to 8530 nt (RNA-2) (**Table 2**). The type species of the genus *Crinivirus* is *Lettuce infectious yellows virus* (LIYV). Noteworthy, Potato yellow vein virus (PYVV) has a tripartite positive-sense, single-stranded RNA genome of 8035 nt (RNA-1), 5339 nt (RNA-2) and 3892 nt (RNA-3) in size (**Table 2**).

The genus *Velarivirus* comprises seven species with particles ranging from 1500 to 1700 nm in length, a monopartite, positive-sense, single-stranded RNA genome, 16–16.9 kb in size, and no known hemipteran vectors (**Table 2**). The type species of the genus *Velarivirus* is *Grapevine leafroll-associated virus 7* (GLRaV-7).

Genome Organization

The 5' end of the genome of viruses in the family *Closteroviridae* is likely capped, and the 3' end is not polyadenylated and does not possess a tRNA-like structure. The genome organization is conserved among species in the family although the number and relative

position of the different open reading frames (ORFs) differ between genera and/or individual viral species. However, the dual gene module ORF1a–ORF1b at the 5′ end of the genomic RNA invariably encodes the replication-associated proteins with conserved domains for a papain-like cysteine protease (L-Pro), a methyl-transferase (MET), a helicase (HEL), and an RNA-dependent RNA polymerase (RdRp). Downstream ORFs form a conserved five-gene module with few modifications among most members of the family. This quintuple gene module codes for, in the 5′→3′ direction, a 6K small hydrophobic protein, the HSP70h, a ~60 kDa protein, a CP, and a CPm.

The HSP70h and the ~60 kDa proteins are integral virion components present in all the sequenced members of the family. The functions postulated for HSP70h are mediation of cell-to-cell movement through plasmodesmata, involvement in the assembly of multi-subunit complexes for genome replication and/or sub-genomic RNAs synthesis, as well as assembly of virus particles. The ~60 kDa protein is required for incorporation of both HSP70h and CPm to virion heads. In general, CP and CPm show a significant degree of sequence conservation for each species in the family and the duplicate copies probably retain the general spatial folding and some crucial properties of the CPs. Noteworthy, a group of ampeloviruses [e.g., Grapevine leafroll-associated virus 4 (GLRaV-4) and related strains, Pineapple mealybug wilt-associated virus 1 (PMWaV-1) and Pineapple mealybug wilt-associated virus 3 (PMWaV-3), Air potato ampelovirus 1 (AiPoV-1), and Plum bark necrosis stem pitting-associated virus (PBNSPaV)] with the smallest genome in the family does not appear to possess a CPm; and grapevine leafroll-associated virus 1 (GLRaV-1) carries two copies of CPm.

The genome expression strategy of viruses in the family *Closteroviridae* is based on: (1) proteolytic processing of the polyprotein encoded by ORF1a; (2) +1 ribosomal frameshift for the expression of the RdRp domain encoded by ORF1b; and (3) expression of the downstream ORFs via the formation of a nested set of 3′ co-terminal sub-genomic RNAs (sgRNAs). Analysis of double stranded (ds) RNAs in infected plant tissue reveals very complex and variable patterns among species of the family, reflecting the different numbers and sizes of ORFs present in individual genomes and, in some cases, the existence of defective RNAs. The largest dsRNA is usually the replicative form of the entire genome and sgRNAs generate a range of smaller dsRNAs.

For species in the genus *Closterovirus* (Table 3), the order of the CP and CPm ORFs is reversed compared to that of species in the genera *Ampelovirus*, *Crinivirus*, and *Velarivirus*. BYV contains eight ORFs flanked by 5′ and 3′ untranslated (UTRs) of 107 and 181 nt, respectively, including an extra ORF (ORF2) encoding a 30 kDa polypeptide with no similarity to any other protein in databases (Fig. 2). Citrus tristeza virus (CTV) has 12 ORFs and UTRs of 107 nt at the 5′ end and 275 nt at the 3′ end. It differs from the BYV genome in having two L-Pro domains in ORF1a [so does Grapevine leafroll-associated virus 2 (GLRaV-2)], and two extra 3′ proximal ORFs (ORF9 and ORF11). The genome organization of Beet yellows stunt virus (BYSV) is intermediate between that of BYV and CTV with 10 ORFs and a 3′UTR 241 nt in size, suggesting that these three viruses might represent three distinct stages in the evolution of closteroviruses. Non-structural proteins common to all members of the genus are: (1) a large polypeptide (over 300 kDa) containing the conserved domains of L-Pro, MET, and HEL; (2) a ~50 kDa protein with all sequence motifs of viral RdRp; (3) a 6 kDa hydrophobic protein with membrane-binding properties; (4) the homolog of the cellular HSP70; and (5) a 55–64 kDa product, referred to as the ~60 kDa protein. Some of the structural and non-structural proteins function as suppressors of the RNA silencing plant defense mechanisms. For instance, CP, p20 and p23 proteins of CTV have suppressor activity, much the same as the homologs of p21 of BYSV, BYV, and GLRaV-2. Silencing suppressors contribute to the accumulation of virus particles and are important determinants of pathogenesis. The CTV p23 is a unique protein in the family with a nucleolar localization.

For species in the genus *Ampelovirus* (Table 3), Grapevine leafroll-associated virus 3 (GLRaV-3) has the largest genome of the family with 18,498 nt and 12 ORFs coding for the replication related proteins (ORFs 1a and 1b), two small hydrophobic proteins (6 kDa), the HSP70h, the ~60 kDa protein, CP, CPm, and five additional proteins of 21, 20, 20, 4, and 7 kDa in size, respectively (Fig. 3). The 5′ and 3′ UTRs are 737 and 277 nt in size, respectively. GLRaV-3 is a sub-group I ampelovirus. Subgroup II species in the genus have only seven ORFs and lack the CPm. For example, PMWaV-1 has a genome 13,071 nt in size, beginning with a 535 nt UTR at the 5′ end, followed by the ORFs expressing the replication related proteins, a 6 kDa hydrophobic protein, the HSP70h, the ~60 kDa protein, the CP, and a 24 kDa protein, respectively. A UTR 132 nt in size terminates the genome at the 3′ end. Structural and non-structural proteins are similar in type and function to those reported for the genus *Closterovirus* although ORF1a harbors, in addition to a L-Pro, an AlkB domain for RNA demethylation.

For species in the genus *Crinivirus* (Table 3), the genome is divided between two linear, positive sense, single-stranded RNAs (Fig. 4) although PYV is an exception with a tripartite genome. All RNA molecules are needed for infectivity and are separately encapsidated. RNA1 of LIYV contains three ORFs, i.e., the ORF1a–ORF1b complex coding for the replication-related proteins, plus a 3′-most ORF coding a 34 kDa protein with no similarity to any protein in databases. This ORF is similar in size and location to ORF2 of CTV and BYSV although the respective expression products are not related. RNA1 has 5′ and 3′UTRs of 97 and 219 nt, respectively. RNA2 has seven ORFs flanked by a 5′UTR of 326 nt and a 3′UTR of 187 nt. RNA2 contains the quintuple gene module, which, however, differs from that of members of the genera *Closterovirus*, *Ampelovirus*, and *Velarivirus* by the insertion of an extra gene (ORF4, p9-10) upstream of the CP gene. The ultimate ORF at the 3′ end of RNA2 of LIYV codes p26; this protein induces plasmalemma deposits, is localized in plasmodesmata, and is required for LIYV systemic plant infection. For PYV, RNA1 (8035 nt in size) is composed of three ORFs, i.e., the ORF1a–ORF1b complex and a 7 kDa hydrophobic protein containing a potential transmembrane helix; RNA2 (5339 nt in size) comprises five predicted ORFs that encode the HSP70h, a 7 kDa protein similar to a comparable protein of cucurbit yellow stunting disorder virus (CYSDV), the ~60 kDa protein, a 9.8 kDa product with no significant similarity to any other sequence in database, and the 28.2 kDa putative CP; and RNA-3 (3892 nt) has three potential ORFs coding for a protein 4 kDa in size with no counterpart with other proteins in the family and no significant sequence homology in databases, the 77.5 kDa CPm, and a 26.4 kDa protein present in other members of the genus. Sweet potato chlorotic stunt virus (SPCSV) and Tomato chlorosis virus (ToCV) have a particularly large CPm (75–80 kDa) compared to LIYV (53 kDa). Both genomic

Table 3 List of virus species in the four genera *Closterovirus*, *Ampelovirus*, *Crinivirus* and *Velarivirus*, in the family *Closteroviridae*. Their name, abbreviation and accession numbers of their genome sequence are indicated

Genus name	Species name	Acronym	NCBI accession #
<i>Closterovirus</i>	<i>Arracacha virus 1</i>	AV-1	MG919988
	<i>Beet yellow stunt virus</i>	BYSV	U51931
	<i>Beet yellows virus</i>	BYV	AF056575
	<i>Blackcurrant closterovirus 1</i>	BCCV1	MH267701
	<i>Carnation necrotic fleck virus</i>	CVFV	GU234166 ^a
	<i>Carrot yellow leaf virus</i>	CYLV	FJ869862
	<i>Citrus tristeza virus</i>	CTV	AF260651
	<i>Grapevine leafroll-associated virus 2</i>	GLRaV-2	AY881628
	<i>Mint virus 1</i>	MV-1	AY792620
	<i>Pistachio ampelovirus A</i>	PAVA	MF198462
	<i>Raspberry leaf mottle virus</i>	RLMoV	DQ357218
	<i>Rehmannia virus 1</i>	ReV-1	MH033657
	<i>Rose leaf rosette-associated virus</i>	RLRaV	KJ748003
	<i>Strawberry chlorotic fleck-associated virus</i>	SCFaV	DQ860839
	<i>Tobacco virus 1</i>	TV1	KT203917
<i>Ampelovirus</i>	<i>Subgroup I</i>		
	<i>Blackberry vein banding-associated virus</i>	BVBaV	KC904540
	<i>Fig leaf mottle-associated virus 2</i>	FLMaV-2	FJ473383 ^a
	<i>Grapevine leafroll-associated virus 1</i>	GLRaV-1	JQ023131
	<i>Grapevine leafroll-associated virus 3</i>	GLRaV-3	AF037268
	<i>Grapevine leafroll-associated virus 13</i>	GLRaV-13	LC052212
	<i>Little cherry virus 2</i>	LChV-2	AF531505
	<i>Pineapple mealybug wilt-associated virus 2</i>	PMWaV-2	AF283103
	<i>Pistachio ampelovirus A</i>	PAVA	MF198462
	<i>Subgroup II</i>		
	<i>Air potato ampelovirus 1</i>	AiPoV-1	MH206615
	<i>Grapevine leafroll-associated virus 4</i>	GLRaV-4	FJ467503
	<i>Pineapple mealybug wilt-associated virus 1</i>	PMWaV-1	AF414119
	<i>Pineapple mealybug wilt-associated virus 3</i>	PMWaV-3	DQ399259
	<i>Plum bark necrosis stem pitting-associated virus</i>	PBNsPaV	EF546442
<i>Crinivirus</i>	<i>Bean yellow disorder virus</i>	BYDV	EU191904, EU191905
	<i>Beet pseudoyellows virus</i>	BPYV	AY330918, AY330919
	<i>Cucurbit yellow stunting disorder virus</i>	CYSDV	AY242077, AY242078
	<i>Cucumber yellows virus</i>	CYV	AB085612, AB085613
	<i>Diodia vein chlorosis virus</i>	DVCV	GQ225585, GQ376201
	<i>Blackberry yellow vein-associated virus</i>	BYVaV	AY776334, AY776335
	<i>Lettuce chlorosis virus</i>	LCV	FJ380118, FJ380119
	<i>Lettuce infectious yellows virus</i>	LIYV	U15440, U15441
	<i>Potato yellow vein virus</i>	PYVV	AJ557128, AJ557129, AJ508757
	<i>Strawberry pallidosis-associated virus</i>	SPaV	AY488137, AY488138
	<i>Sweet potato chlorotic stunt virus</i>	SPCSV	AJ428554, AJ428555
	<i>Tetterwort vein chlorosis virus</i>	TwVSV	KR002686, KR002687
	<i>Tomato chlorosis virus</i>	ToCV	AY903447, AY903448
	<i>Tomato infectious chlorosis virus</i>	TICV	FJ815440, FJ814441
<i>Velarivirus</i>	<i>Areca palm velarivirus 1</i>	ArPV1	KR349464
	<i>Cordyline virus 1</i>	CoV-1	HM588723
	<i>Cordyline virus 2</i>	CoV-2	JQ599282 ^a
	<i>Cordyline virus 3</i>	CoV-3	JQ599283 ^a
	<i>Cordyline virus 4</i>	CoV-4	JQ599284 ^a
	<i>Grapevine leafroll-associated virus 7</i>	GLRaV-7	HE588185
	<i>Little cherry virus 1</i>	LChV-1	Y10237

^aPartial sequence.

RNAs of ToCV encode RNA silencing suppressors, e.g., the p22 protein in RNA1, and CP and CPM in RNA2. The p25 protein of CYSDV, a viral RNase III, and the p22 gene present in a few isolates of SPCSV also have RNA silencing suppressor activity.

For the virus species in the genus *Velarivirus* (Table 3), Little cherry virus 1 (LChV-1) has the largest genome of the genus with 16,934 nt and nine ORFs coding for the replication associated proteins (ORF1a with a L-Pro, a MET and a HEL; and ORF1b with a

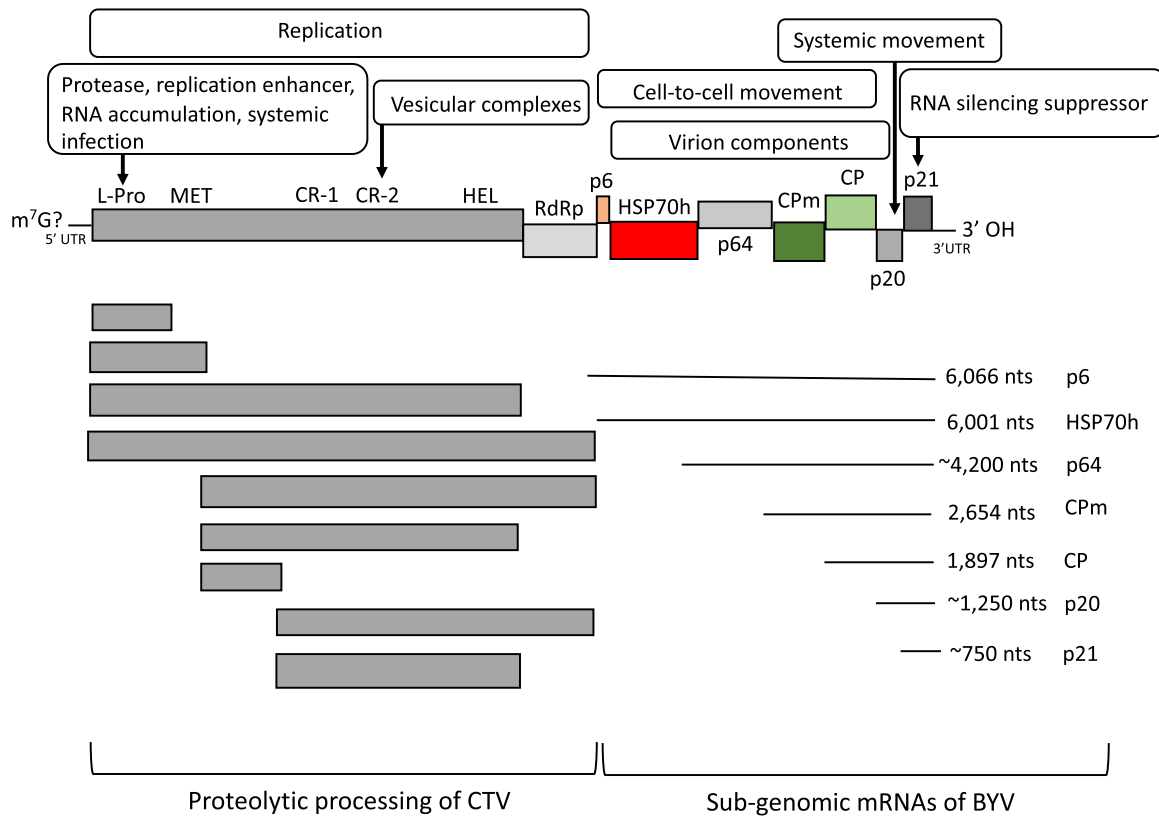


Fig. 2 Genome organization and strategy of replication of species in the genus *Closterovirus* with the relative position of the open reading frames, their expression products, and 3' nested sub-genomic RNAs (sgRNAs). Information on genomic organization and sgRNAs is provided for Beet yellows virus (BYV), the type species of the genus *Closterovirus*: UTR, untranslated region; L-Pro, leader papain-like cysteine proteinase; MET, methyltransferase; CR-1, central region 1; CR-2, central region 2; HEL, helicase; RdRp, RNA-dependent RNA polymerase; HSP70h, heat shock protein 70 homolog; ~60 kDa protein; CPm, minor coat protein; CP, coat protein. The five boxes under cell-to-cell movement represent the quintuple gene module conserved among most closteroviruses. Information on proteolytic processing of the polyprotein precursors is provided for Citrus tristeza virus (CTV).

RdRp), a small hydrophobic protein (4–8 kDa) with a transmembrane domain, the HSP70h, the ~60 kDa protein, CP, CPm, and two additional proteins 25 and 27 kDa in size (Fig. 5) (ICTV, see “Relevant Websites Section”). The genome of GLRaV-7 is 16,496 nt in size and has nine ORFs encoding structural and non-structural proteins similar in type and function to those of LChV-1.

Phylogenetic relationships based on HSP70h aa sequences reveal the grouping of species into four distinct genera with a few species unassigned to the family (Fig. 6). Similar trees are obtained when analyzing phylogenetic relationships based on CP or RdRp aa sequences. The genic diversity of species in the family *Closteroviridae* is primarily influenced by (1) strong negative selection for some viruses, (2) co-infection by genetically similar sequence variants in distant geographic regions, likely as a result of the extensive exchange of infected propagative planting material, (3) recombination between divergent sequence variants, (4) preferential accumulation of certain virus strains in relation to mixed infections, and (5) positive selection of sequence variants due to host change or transmission by insect vectors.

From an evolutionary point of view, viruses in the family *Closteroviridae* represent a monophyletic virus lineage that might have evolved from a smaller filamentous virus when higher plants have differentiated. This progenitor virus is hypothesized to be composed of three genes encoding replication-associated proteins, a protein (p6) with affinity for cell membranes, and a single CP, and then to have acquired the HSP70h and a ~60 kDa protein derived from a fusion of two domains, a N-terminal domain of unknown evolutionary provenance, and a duplicated CP domain. Under the pressure of additional modular evolutionary events, i.e., duplication of the CP gene, acquisition of diverse suppressors of RNA silencing and of additional genes acquired via horizontal gene transfer (e.g., papain-like cysteine proteinase and AlkB domains), this family ancestor gave rise to the progenitors of the four extant genera of the family. Species in the genus *Crinivirus* differentiated further by splitting their genome in two or three components.

Based on their extreme genomic stability, a few species in the family *Closteroviridae* have been engineered as gene expression and RNA interference vectors. For instance, BYV, LIYV, CTV, GLRaV-2, and GLRaV-3 have been engineered as vectors by elegantly inserting a heterologous sgRNA promoter into an optimal genomic location to drive the expression of the green fluorescence protein reporter, or silence host genes such as the phytoene desaturase via inverted-repeat sequences. The most notable success of a closterovirus-derived vector has been achieved with GLRaV-2.

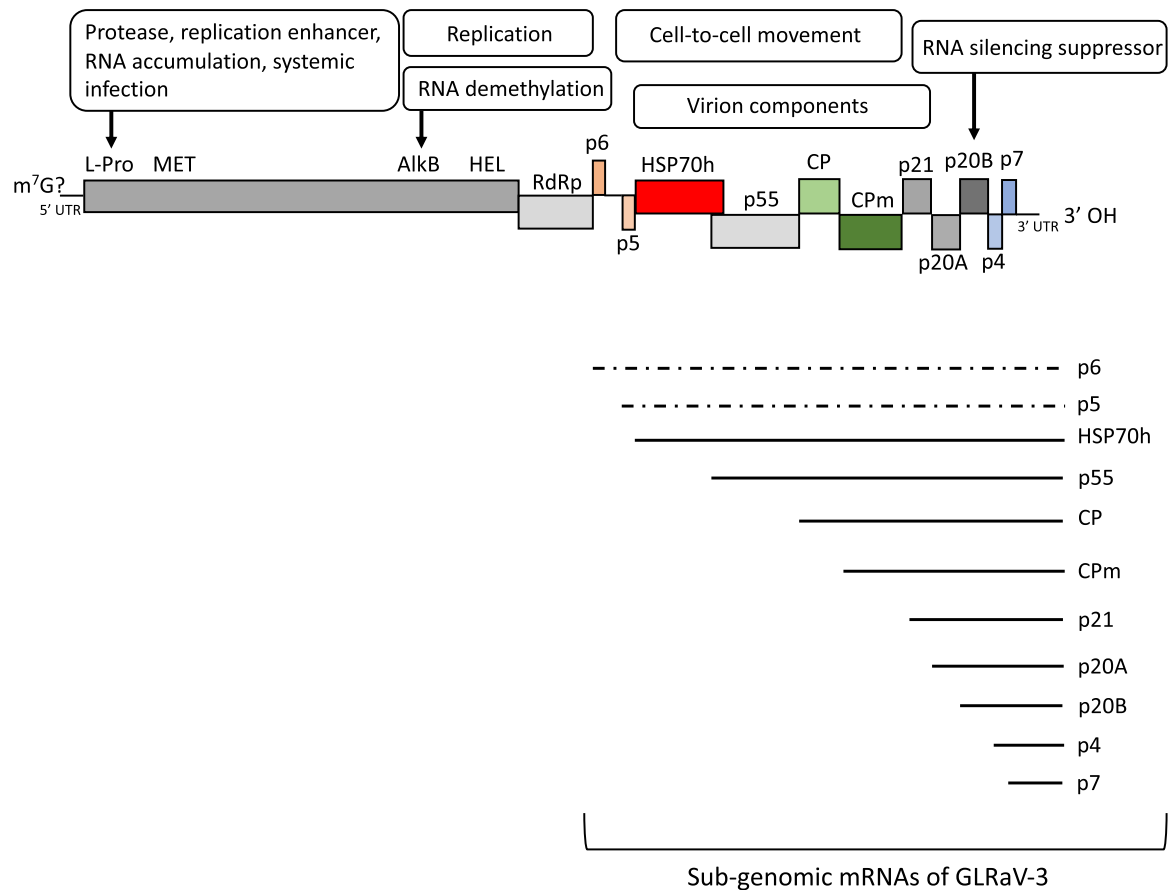


Fig. 3 Genome organization of Grapevine leafroll-associated virus 3 (GLRaV-3), the type species of the genus *Ampelovirus*, showing the relative position of the open reading frames and their expression products: UTR, untranslated region; L-Pro, leader papain-like protease; MET, methyltransferase; AlkB, alpha-ketoglutarate-dependent hydroxylase domain; HEL, helicase; RdRp, RNA-dependent RNA polymerase; HSP70h, heat shock protein 70 homolog; ~60 kDa protein; CP, coat protein; CPm, minor coat protein. The predicted sub-genomic RNAs (sgRNAs) are indicated below the genome map. Putative sgRNAs are indicated by dashed lines.

Infectious Cycle

Positive-sense RNA viruses, such as members of the family *Closteroviridae*, express their replication-associated proteins in infected plant cells following translation of the genomic RNAs. The replication complex then initiates the synthesis of complementary RNA strand while remodeling host membranes, thus creating cytoplasmic compartments that are referred to as virus factories. Such compartmentalization secures RNA replication from the action of nucleases and pattern recognition receptors that trigger host innate immunity responses while providing opportunities for exchange with the cytoplasm for the export of, for instance, newly synthesized genomic and sub-genomic RNAs.

The 3'UTR of the genomic RNA of species in the family *Closteroviridae*, particularly CTV, is highly conserved and contains several hairpin structures and a putative pseudoknot that is essential for replication, while the 5'UTR is variable but adopts a conserve secondary structure that is important for replication and proper encapsidation of the viral RNA. Replication occurs in the cytoplasm, possibly in association with endoplasmic reticulum-derived membranous vesicles and vesiculated mitochondria. Virus encoded polyproteins 1a and 1b direct membrane remodeling and formation of multi-vesicular replication factories. Polyprotein 1a of BYV contains a variable central region (CR) between the MTR and HEL domains of which a stretch (aa 1368–1432) induces the formation of 1-μm mobile globules originating from endoplasmic reticulum membranes. Part of the CR is conserved in all members of the genus *Closterovirus* and contains a predicted amphipathic helix (aa 1368–1385). This region may be involved in the biogenesis of closterovirus replication factories.

Ultrastructural modifications arise by membrane proliferation, degeneration and vesiculation of mitochondria, and formation of inclusion bodies. These are made up of aggregates of virions or membranous vesicles, or a combination of the two. Virions accumulate in conspicuous cross-banded fibrous masses or, more typically, in more or less loose bundles intermingled with single or clustered membranous vesicles. The vesicles contain a fibrillar network and derive either from the endoplasmic reticulum, or from peripheral vesiculation of mitochondria.

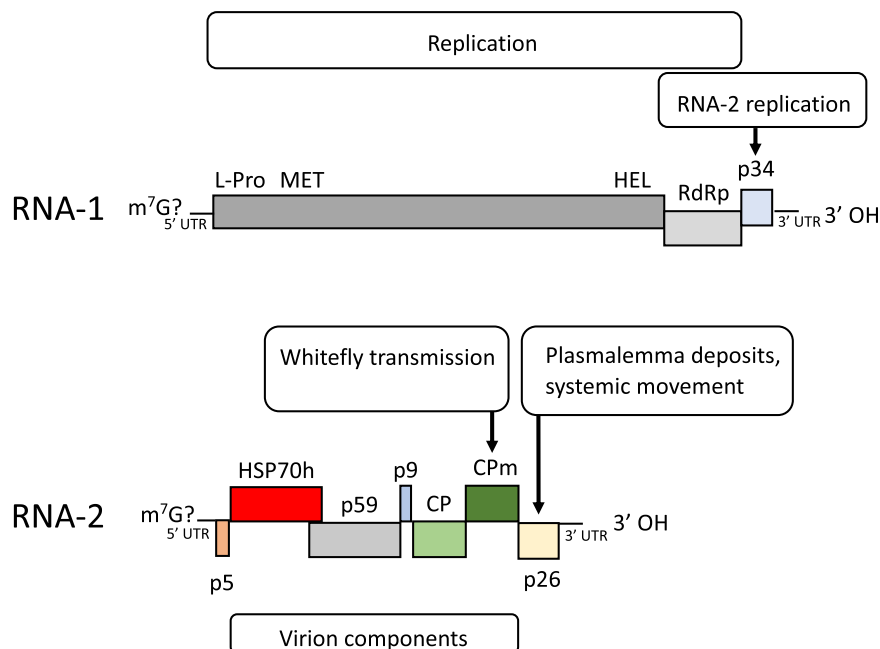


Fig. 4 Genome organization of Lettuce infectious yellows virus (LIYV), the type species of the genus *Crinivirus*, showing the relative position of the open reading frames and their expression products: UTR, untranslated region; L-Pro, papain-like protease; MET, methyltransferase; HEL, helicase; RdRp, RNA-dependent RNA polymerase; HSP70h, heat shock protein 70 homolog; ~60 kDa protein; CP, coat protein; CPm, minor coat protein.

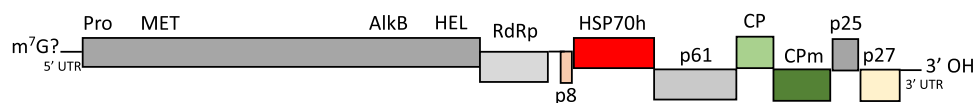


Fig. 5 Genome organization of Grapevine leafroll-associated virus 7 (GLRaV-7), the type species of the genus *Velarivirus*, showing the relative position of the open reading frames and their expression products: UTR, untranslated region; L-Pro, papain-like protease; MET, methyltransferase; HEL, helicase; RdRp, RNA-dependent RNA polymerase; HSP70h, heat shock protein 70 homolog; ~60 kDa protein; CP, coat protein; CPm, minor coat protein.

For GLRaV-3, putative 3' co-terminal sgRNAs are differentially expressed with four of them accumulating at high levels (CP, p20A, p20B, p21), two at intermediate levels (p4 and p7), and others at low levels (CPm, p55, HSP70h, and p5). This suggests temporal and quantitative regulation of sgRNAs transcription during plant infection. Each of these sgRNAs serves as a monocistronic messenger for translation of the corresponding 5'-proximal ORF. The 5' transcriptional start sites were determined for several GLRaV-3 sgRNAs. Additionally, the leader sequence of sgRNAs is variable in size and does not share conserved motifs.

For LIYV, the replication of both genomic RNAs is asynchronous as RNA1 genomic and sgRNAs accumulate before significant accumulation of RNA2 can be detected, suggesting that RNA1 likely replicates in *cis* while RNA2 replicates in *trans*. Knockout mutation studies showed that the single-stranded RNA-binding protein p34 that is encoded by RNA1 is a *trans* enhancer of RNA2 replication.

Epidemiology

Transmission of viruses in the family *Closteroviridae* by natural vectors is semi-persistent regardless of the type of vector, i.e., aphids (genus *Closterovirus*), whiteflies (genus *Crinivirus*), and pseudococcid mealybugs and soft scale insects (genus *Ampelovirus*). No insect vector is known for viruses in the genus *Velarivirus*. Additionally, some species in the family, particularly velariviruses, can be transmitted through dodder (*Cuscuta* spp.). Transmission through seeds has not been documented. Only a few species of the genus *Closterovirus* (CTV, GLRaV-2, BYV) are transmissible by mechanical inoculation to herbaceous hosts, though with difficulty, but none of the viruses in the genera *Ampelovirus*, *Crinivirus* and *Velarivirus* are mechanically transmissible.

Viruses in the family *Closteroviridae* that infect vegetatively propagated hosts are transmitted by grafting and, thus, disseminated over long distances with propagating material that is not adequately screened for their presence. The natural and experimental host ranges are usually restricted, except for a few members of the genus *Crinivirus*. Similarly, the geographical

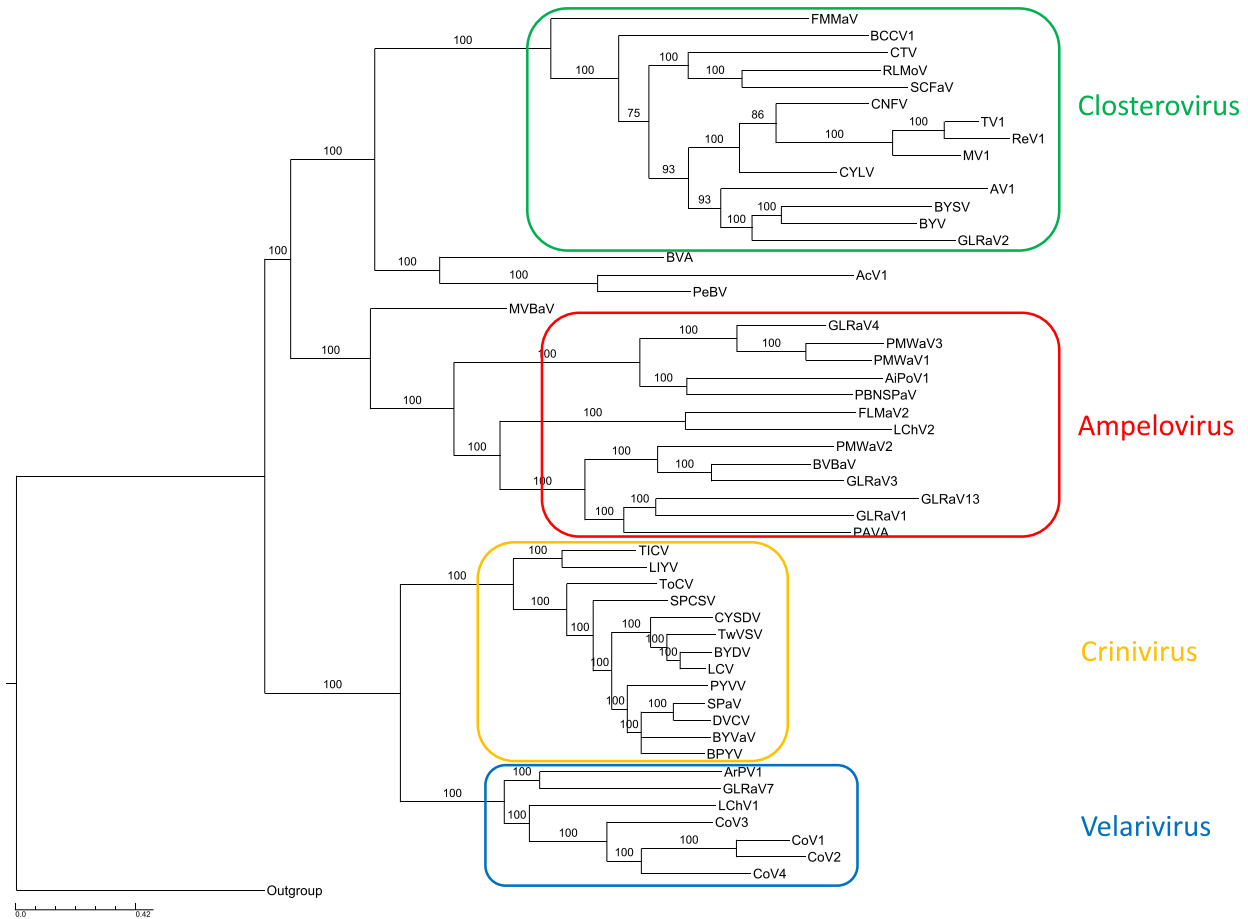


Fig. 6 Maximum likelihood phylogenetic tree showing the relationships among recognized members of the family *Closteroviridae* based on an alignment of the complete amino acid sequence of the heat shock protein 70 homolog using MUSCLE. The maximum likelihood tree was inferred using RAxML in the T-REX web server. Distances are proportional to branch lengths and the bar represents the genetic distance. The heat shock protein 70 from *Arabidopsis thaliana* (AEE75218) was used as outgroup. The GenBank accession number used for each virus is as follows: Actinidia virus 1 (AcV-1, KX857665), Air potato ampelovirus 1 (AiPV-1, MH206615), arracacha virus 1 (AV-1, MG919988), Blackcurrant closterovirus 1 (BCCV1, MH267701), Bean yellow disorder virus (BYDV, EU191904), Beet pseudoyellows virus (BPYV, AY330918), Beet yellow stunt virus (BSYV, U51931), Beet yellows virus (BYV, AF056575), Blackberry vein banding-associated virus (BVBaV, KC904540), Blueberry virus A (BVA, AB733585), Carnation necrotic fleck virus (CVFV, GU234166), Carrot yellow leaf virus (CYLV, FJ869862), Citrus tristeza virus (CTV, U16304), Cordyline virus 1 (CoV-1, HM588723), Cordyline virus 2 (CoV-2, JQ599282), Cordyline virus 3 (CoV-3, JQ599283), Cordyline virus 4 (CoV-4, JQ599284), Cucurbit yellow stunting disorder virus (CYSDV, AY242077), Diodia vein chlorosis virus (DVCV, CQ376201), Fig leaf mottle-associated virus 2 (FLMaV-2, FJ473383), Fig mild mottle-associated virus (FMMaV, FJ611959), Grapevine leafroll-associated virus 1 (GLRaV-1, JQ023131), Grapevine leafroll-associated virus 2 (GLRaV-2, JX513891), Grapevine leafroll-associated virus 3 (GLRaV-3, EU259806), Grapevine leafroll-associated virus 4 (GLRaV-4, FJ467503), Grapevine leafroll-associated virus 7 (GLRaV-7, HE588185), Grapevine leafroll-associated virus 13 (GLRaV-13, LC052212), Lettuce chlorosis virus (LCV, FJ380118), Lettuce infectious yellows virus (LIYV, U15440), Little cherry virus 1 (LChV-1, EU715989), Little cherry virus 2 (LChV-2, AF531505), Mint vein banding-associated virus (MVBaV, KJ572575), Mint virus 1 (MV-1, AY792620), Palm velarivirus 1 (ArPV1, KR349464), Persimmon virus B (PeBV, AB923924), Pineapple mealybug wilt-associated 1 (PMWaV-1, AF414119), Pineapple mealybug wilt-associated 2 (PMWaV-2, AF283103), Pineapple mealybug wilt-associated 3 (PMWaV-3, DQ399259), Pistachio ampelovirus A (PAVA, MF198462), Plum bark necrosis stem pitting-associated virus (PBNSPaV, EF546442), Raspberry leaf mottle virus (RLMoV, DQ357218), Rehmannia virus 1 (ReV-1, MH033657), Rose leaf rosette-associated virus (RLRaV, KJ7488003), Strawberry chlorotic fleck-associated virus (SCFaV, DQ860839), Potato yellow vein virus (PYVV, AJ557128), Strawberry pallidosis-associated virus (SPaV, AY488138), Sweet potato chlorotic stunt virus (SPCSV, AJ428554), Tetterwort vein chlorosis virus (TwVSV, KR002687), Tobacco virus 1 (TV1, KT203917), Tomato chlorosis virus (ToCV, AY903447), and Tomato infectious chlorosis virus (TICV, FJ815440).

distribution of viruses varies from restricted [Grapevine leafroll-associated virus 13 (GLRaV-13), Mint virus 1 (MV-1), Tobacco virus 1 (TV1)] to widespread (CTV, GLRaV-2, GLRaV-3, BYV).

The range of vectors for species in the genus *Closterovirus* varies from rather wide to restricted. For example, BYV is transmitted by at least 23 aphid species (*Myzus persicae* and *Aphis fabae* being the main natural vectors), CTV by seven species (*Toxoptera citricida* and *Aphis gossypii* being the most efficient vectors), MV-1 by a single aphid species while GLRaV-2 does not have any known vectors. Recent studies of specific interactions between CTV and its aphid vector, *Toxoptera citricida*, indicate the cibarium of the

aphid foregut as the virus retention site, and the critical role of a protein-carbohydrate complex for transmission with three viral proteins, i.e., the CPm, p61, and p65, binding to sugar moieties on the surface of the foregut.

For species in the genus *Ampelovirus*, the range of vectors varies from large to narrow. For instance, GLRaV-1 is transmitted by species of several genera of pseudococcid mealybugs (*Heliococcus*, *Phenacoccus*, *Pseudococcus*) and soft scale insects (*Pulvinaria*, *Neopulvinaria* and *Parthenolecanium*). Similarly, GLRaV-3 is transmitted by a wide range of pseudococcid mealybugs (*Planococcus*, *Pseudococcus*, *Heliococcus*, *Phenacoccus*) and soft scale insects (*Pulvinaria*, *Neopulvinaria*, *Parthenolecanium*, *Coccus*, *Saissetia*, *Parasaissetia*, *Ceroplastes*). To the contrary, vectors of PMWaV-1, PMWaV-2 and PMWaV-3 are two species of the genus *Dysmicoccus*, and LChV-2 is transmitted by *Phenacoccus aceris*.

Species in the genus *Crinivirus* are transmitted by whiteflies of the genera *Trialeurodes* and *Bemisia*. The specificity of transmission has been used for species differentiation. For example, viruses of group 1 [PYV, Blackberry yellow vein-associated virus (BYVaV), Beet pseudoyellows virus (BPYV) and Strawberry pallidosis-associated virus (SpaV)] are transmitted by *T. vaporariorum*, viruses of group 2 [ToCV, SPCSV, CYSDV, and Bean yellow disorder virus (BYDV)] by *B. tabaci*, whereas one of the viruses of group 3 is transmitted by *B. tabaci* (LIYV) and the other by *T. vaporariorum* (TICV). These three groups are sustained by comparative phylogenetic analyses of RdRp aa sequences.

Disease Symptoms

Viruses in the family *Closteroviridae* infect fruit, vegetable, tuberous, and berry crops, as well as herbs, flowering plants, shrubs, and weeds. They cause foliar discolorations (chlorosis, discoloration, or reddening) and rolling (Fig. 7), stem pitting and grooving on the woody cylinder of woody hosts, stunting, or symptomless infections. Disease symptoms also include small and late ripening fruits.

Pathogenesis

Plant infection by species in the family *Closteroviridae* is systemic but virions are usually found in the phloem (sieve tubes, companion cells and parenchyma), and only occasionally in the mesophyll and epidermis. Virus-host interactions have been

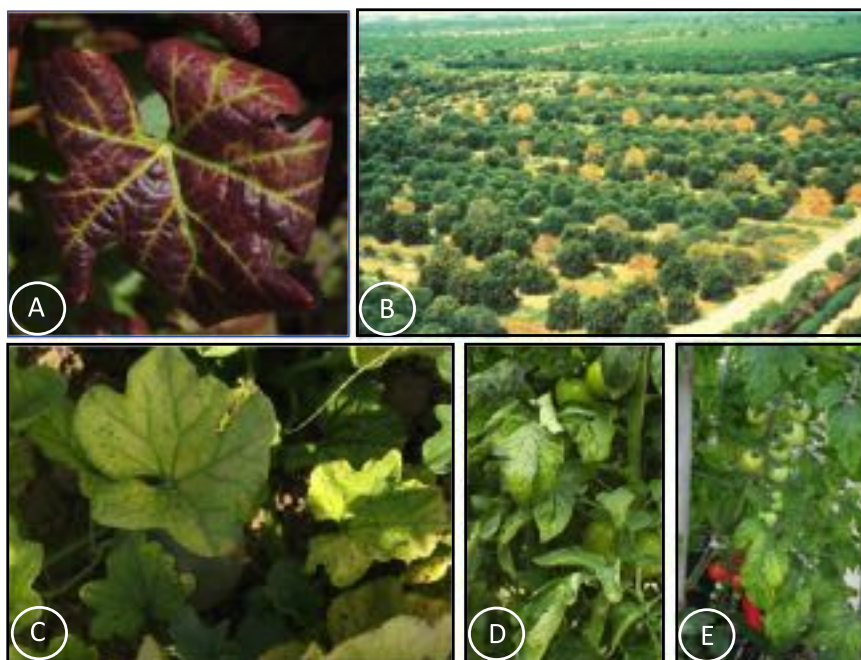


Fig. 7 Disease symptoms caused by selected viruses of the family *Closteroviridae* on their natural host. (A) Downward rolling and reddening of the leaf blade of *Vitis vinifera* cv. Pinot noir infected with Grapevine leafroll-associated virus 3 from the genus *Ampelovirus*, (B) Aerial view of declining and dead citrus trees grafted onto the sour orange rootstock in a grove infected with Citrus tristeza virus from the genus *Closterovirus*. Infected trees show necrosis at the graft union with reduced carbohydrate movement into the roots, and leaves fall off (courtesy of W.O. Dawson, University of Florida, USA), (C) Interveinal discoloration of melon leaves infected by Cucurbit yellow stunting disorder virus from the genus *Crinivirus* (courtesy of W. Wintermantel, USDA-ARS, Salinas, California, USA), (D) Discoloration of tomato leaves infected by Tomato chlorosis virus from the genus *Crinivirus* (courtesy of W. Wintermantel, USDA-ARS, Salinas, California, USA), and (E) Tomato infectious chlorosis virus from the genus *Crinivirus* (courtesy of W. Wintermantel, USDA-ARS, Salinas, California, USA). Note the similarities of symptomatology caused by ToCV and TICV, and by magnesium deficiency.

extensively studied for CTV. Protein p23 is a multifunctional RNA-binding protein with a putative zinc finger domain that accumulates in the nucleolus and plasmodesmata. It regulates the asymmetric balance of plus and minus RNA strands during replication, induces disease symptoms similar to those caused by CTV infection in certain hosts, and enhances systemic infection, and virus accumulation. Recently, it was shown that CTV uses the cytosolic glyceraldehyde 3-phosphate dehydrogenase via interaction with p23 to facilitate infection. In addition, p23 like p25 and p20 is a suppressor of RNA silencing. Proteins p33, p18, and p13 are dispensable for systemic infection of certain hosts, and involved in stem pitting disease symptoms.

Diagnosis

The visual identification of disease symptoms incited by viruses in the family *Closteroviridae* can be challenging due to confounding factors such as asymptomatic infections, virus co-infections, or similarities with nutritional deficiencies. Several techniques have been developed and applied to diagnose closteroviruses, including biological assays, serological assays, nucleic acid-based methods, and large-scale sequencing.

Biological indexing consists of chip budding a candidate plant to a disease susceptible, indicator plant. The indicator plant with the newly grafted material is then observed for the development of disease symptoms. This approach is labor intensive, time consuming, costly, and often not very robust. Biological assays are progressively replaced by more powerful diagnostic assay that identify viruses rather than diseases.

Serological diagnostic techniques such as enzyme-linked immunosorbent assays have been developed for simple, cost-effective and high-throughput testing. Although virion proteins are often moderately antigenic, serological reagents are available for many viruses in the family *Closteroviridae*, and most species within genera are serologically unrelated or distantly related to one another, for instance, GLRaV-1 and GLRaV-3, or GLRaV-1 and GLRaV-13, but no intergeneric serological relationship has been detected.

Nucleic acid-based diagnostic methods are increasingly used for the detection of closteroviruses. Such assays include reverse transcription (RT) polymerase chain reaction (PCR) and variations thereof (immunocapture-RT-PCR, spot RT-PCR, RT-qPCR), loop-mediated amplification of nucleic acid, and oligonucleotide microarrays. These assays are often more robust than serological assays. More recently, large scale sequencing is increasingly used for discovery of new viruses, including closteroviruses, and for an unbiased diagnosis of a virome in a plant sample without any prior knowledge of the pathogens present. This technology should facilitate the safe and rapid exchange of propagative plant material.

Treatment

Management of diseases caused by viruses in the family *Closteroviridae* is challenging, due mainly to the absence of curative practices. Strategies to mitigate the impact of viruses rely almost exclusively on preventive approaches.

Prevention

Screening nursery stocks for viruses is critical for the production of clean planting material and the establishment of clean production fields. Planting material derived from virus-tested nursery stocks minimizes virus incidence. Technological advances that enable the efficient, rapid identification and removal of virus-infected stocks facilitate disease management tactics. Additionally, reducing disease inoculum via, for instance, roguing, is critical to limit disease spread, especially in perennial crops.

Further Reading

- Al Rwahnih, M., Dolja, W.W., Daubert, S., Koonin, E.V., Rowhani, A., 2012. Genomic and biological analysis of Grapevine leafroll-associated virus 7 reveals a possible new genus within the family *Closteroviridae*. *Virus Research* 163, 302–309.
- Agranovsky, A.A., 2016. Closteroviruses: Molecular biology, evolution and interactions with cells. In: Gaur, R., Petrov, N., Patil, B., Stoyanova, M. (Eds.), *Plant Viruses: Evolution and Management*. Singapore: Springer.
- Dawson, W.O., Bar-Joseph, M., Garnsey, S.M., Moreno, P., 2015. Citrus tristeza virus: Making an ally from an enemy. *The Annual Review of Phytopathology* 53, 137–155.
- Dolja, V.V., Koonin, E.V., 2013. The closterovirus-derived gene expression and RNA interference vectors as tools for research and plant biotechnology. *Frontiers in Microbiology* 4. doi:10.3389/fmicb.2013.00083.
- Flores, R., Ruiz-Ruiz, S., Soler, N., *et al.*, 2013. Citrus tristeza virus p23: A unique protein mediating virus–host interactions. *Frontiers in Microbiology* 4. doi:10.3389/fmicb.2013.00098.
- Gushchin, V.A., Karlin, D.G., Makhonenko, A.V., *et al.*, 2017. A conserved region in the closterovirus 1a polypeptide derives extensive remodeling of endoplasmic reticulum membranes and induces motile globules in *Nicotiana benthamiana* cells. *Virology* 502, 106–113.
- Jarugula, S., Gowda, S., Dawson, W.O., Naidu, R.A., 2018. Development of infectious cDNA clones of Grapevine leafroll-associated virus 3 and analyses of the 5' non-translated region for replication and virion formation. *Virology* 523, 89–99.
- Killi, N., Harper, S.J., Alfaress, S., El Mohtar, C., Dawson, W.O., 2016. Minor coat and heat shock proteins are involved in the binding of Citrus tristeza virus to the foregut of its aphid vector, *Toxoptera citricida*. *Applied and Environmental Microbiology* 82, 6294–6302.
- Kiss, Z., Medina, V., Falk, B.W., 2013. Crinivirus replication and host interactions. *Frontiers in Microbiology* 4. doi:10.3389/fmicb.2013.00099.
- Kurth, E.G., Peremyslov, V.V., Prokhnovsky, A.I., *et al.*, 2012. Virus-derived gene expression and RNA interference vector for grapevine. *Journal of Virology* 86, 6002–6009.

- Maree, H.J., Almeida, R.P.P., Bester, R., *et al.*, 2013. Grapevine leafroll-associated virus 3. *Frontiers in Microbiology* 4. doi:10.3389/fmicb.2013.00082.
- Qiao, W., Medina, V., Kuo, Y.-W., Falk, B.W., 2018. A distinct, non-virion plant virus movement protein encoded by a crinivirus essential for systemic infection. *mBio* 9 (6), e02230.
- Rubio, L., Guerri, J., Moreno, P., 2013. Genetic variability and evolutionary dynamics of viruses of the family *Closteroviridae*. *Frontiers in Microbiology* 4. doi:10.3389/fmicb.2013.00151.
- Ruiz-Ruiz, S., Navarro, B., Peña, L., *et al.*, 2019. Citrus tristeza virus: Host RNA silencing and virus counteraction. In: Catara, A.F., Bar-Joseph, M., Licciardello, G. (Eds.), *Citrus Tristeza Virus: Methods and Protocols. Methods in Molecular Biology* 2015. New York: Humana, pp. 195–207.
- Tzanetakis, I.E., Martin, R.R., Wintermantel, W.M., 2013. Epidemiology of criniviruses: An emerging problem in world agriculture. *Frontiers in Microbiology* 4. doi:10.3389/fmicb.2013.00119.

Relevant Websites

<http://www.dpvweb.net/dpv/showdpv.php?dpvno=377>

Beet yellows virus

Show DPV and Refs in Frame.

https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/255/closteroviridae

Closteroviridae

Positive Sense RNA Viruses.

<https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=69973>

Complete genomes: Closteroviridae

NCBI.

Comoviruses and Fabaviruses (*Secoviridae*)

George P Lomonossoff, John Innes Centre, Norwich, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
Co-Pro Protease cofactor
CP Coat protein or capsid protein
kb Kilobase
kDa Kilo Dalton
MP Movement protein

nt Nucleotide(s)
ORF Open reading frame
RdRp RNA-dependent RNA polymerase
UTR Untranslated region
VLPs Virus-like particles
VPg Viral protein genome-linked

Glossary

Viral protein genome-linked A protein covalently attached to the 5' end of positive strand viral RNA and acting as a primer during RNA synthesis.

Virus re-assortment The mixing of the genome segments of a virus into new combinations.

Virus replication complexes Intracellular complexes formed by association with intracellular membranes that are sites for stages in the replication cycle of a virus.

β -barrel Beta-sheet composed of tandem repeats forming a closed structure.

Introduction

The genera *Comovirus* and *Fabavirus* are non-enveloped plant viruses that are classified within the family *Secoviridae*. They typically cause mottle or mosaic symptoms on susceptible hosts and are characterized by having a bipartite RNA genome with the individual components separately encapsidated in isometric particles (Fig. 1). Though having similar genetic and particle structures, the two genera are distinguished on the basis of their insect vectors.

Taxonomy and Classification

The genus *Comovirus* is named after the type species, *Cowpea mosaic virus* (CPMV). The genus *Fabavirus* derives its name from the host species (*Vicia faba*) from which the first member of the genus, *Broad bean wilt virus* (BBWV), was originally isolated as the causative agent of vascular wilt symptoms. At present, 15 and 7 distinct virus species are recognized by the International Committee for the Taxonomy of Viruses (ICTV) as being members of the genus *Comovirus* and *Fabavirus*, respectively (Table 1). Together with the genus *Nepovirus*, they are classified in the subfamily *Comovirinae* within the family *Secoviridae*. Comoviruses and fabaviruses are insect-transmitted by beetles and aphids, respectively, while nepoviruses are nematode-transmitted. The relationship between members of the two genera and their relationship to other members of the family *Secoviridae* is shown in (Fig. 2).

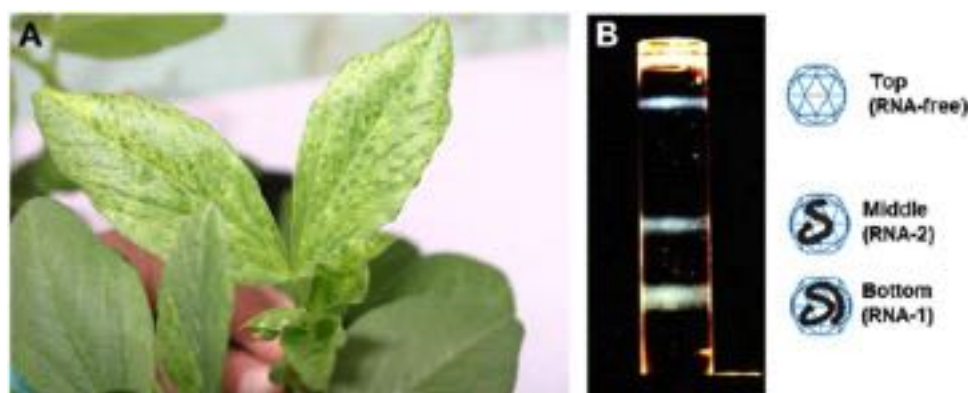


Fig. 1 Fundamental features of comoviruses and fabaviruses. (A) Symptoms of the fabavirus, BBWV-1, on *Vicia faba*. (B) The three-component nature of comoviruses and fabaviruses. The photograph shows fractionation of CPMV on a cesium chloride gradient. The positions of the three nucleoprotein components and their RNA contents are indicated. Photo in (A) courtesy of Dr. I. Ferriol, CRAG, Barcelona, Spain

Table 1 List of species members of the genera *Comovirus* and *Fabavirus* in the family *Secoviridae*. The genera type species are written in bold. The information regarding their molecular biology is indicated for each member

Genus/Species	Acronym	Genome sequence	Accession #	Infectious clones	Particle structure	Biotechnology use
Comovirus						
<i>Andean potato mottle virus</i>	APMoV	Complete	RNA 1: MN176101 RNA 2: MN176102			
<i>Bean pod mottle virus</i>	BPMV	Complete	RNA 1: NC003496/U70866 RNA 2: NC003495/M62738	YES	YES	YES
<i>Bean rugose mosaic virus</i>	BRMV	Complete	RNA 1: NC028139/KP404602 RNA 2: NC028146/KP404603	YES		
<i>Broad bean stain virus</i>	BBSV	Partial	RNA 1: NC043386/KJ746622 RNA 2: NC043385/FJ028650		YES	
<i>Broad bean true mosaic virus</i>	BBTMV	Complete	RNA 1: NC022004/GU810903 RNA 2: NC022006/GU810904			
<i>Cowpea mosaic virus</i>	CPMV	Complete	RNA 1: NC003549/X00206 RNA 2: NC003550/X00729	YES	YES	YES
<i>Cowpea severe mosaic virus</i>	CPSMV	Complete	RNA 1: NC003545/M83830 RNA 2: NC003544/M83309	YES		
<i>Glycine mosaic virus</i>	GMV					
<i>Pea green mottle virus</i>	PGMV					
<i>Pea mild mosaic virus</i>	PMiMV					
<i>Quail pea mosaic virus</i>	QPMV					
<i>Radish mosaic virus</i>	RaMV	Complete	RNA 1: NC010709/AB295643 RNA 2: NC010710/AB295644	YES		
<i>Red Clover mottle virus</i>	RCMV	Complete	RNA 1: NC003741/X64886 RNA 2: NC003738/M14913		YES	
<i>Squash mosaic virus</i>	SqMV	Complete	RNA 1: NC003799/AB054688 RNA 2: NC003800/AB054689			
<i>Ullucus virus C</i>	UVC					
Fabavirus						
<i>Broad bean wilt virus 1</i>	BBWV-1	Complete	RNA 1: NC005289/AB084450 RNA 2: NC005290/AB084451	YES		
<i>Broad bean wilt virus 2</i>	BBWV-2	Complete	RNA 1: NC003003/AF225953 RNA 2: NC003004/AF225954	YES		
<i>Cucurbit mild mosaic virus</i>	CuMMV	Complete	RNA 1: NC038760/EU881936 RNA 2: NC038759/EU881937			
<i>Gentian mosaic virus</i>	GeMV	Complete	RNA 1: AB084452 RNA 2: AB084453			
<i>Grapevine fabavirus</i>	GFabV	Nearly complete	RNA 1: NC039073/KX241482 RNA 2: NC039072/KX241485			
<i>Lamium mild mosaic virus</i>	LMMV	Complete	RNA 1: NC023016/KC595304 RNA 2: NC023017/KC595305			
<i>Prunus virus F</i>	PrVF	Complete	RNA 1: NC039077/KX269865 RNA 2: NC039078/KX269871			

In several instances, individual comoviruses and fabaviruses have been shown to exist in several different strains. For example, three distinct strains of *Red Clover mottle virus* (RCMV), termed -S, -N and -O, have been described. In the case of *Bean pod mottle virus* (BPMV) at least two distinct strains have been reported and these can reassort. Detailed molecular characterization of BPMV reassortants has revealed the presence of strains that are diploid for RNA1 and haploid for RNA2. The fabaviruses, BBWV-1 and BBWV-2 were originally classified as strains of the same species but are now considered to be distinct species.

Physical Properties of Viral Particles

The yields of comovirus particles from infected tissue are typically higher than those of fabaviruses, particles of which have a tendency to aggregate on purification. The thermal inactivation point of viral particles in plant sap is in the range 60–75°C and the particles have a longevity in sap of between 2 and 112 days at room temperature.

Comovirus and fabavirus preparations consist of non-enveloped isometric particles, approximately 30 nm in diameter, which can be separated into three components designated top (T), middle (M) and bottom (B) by centrifugation on sucrose density

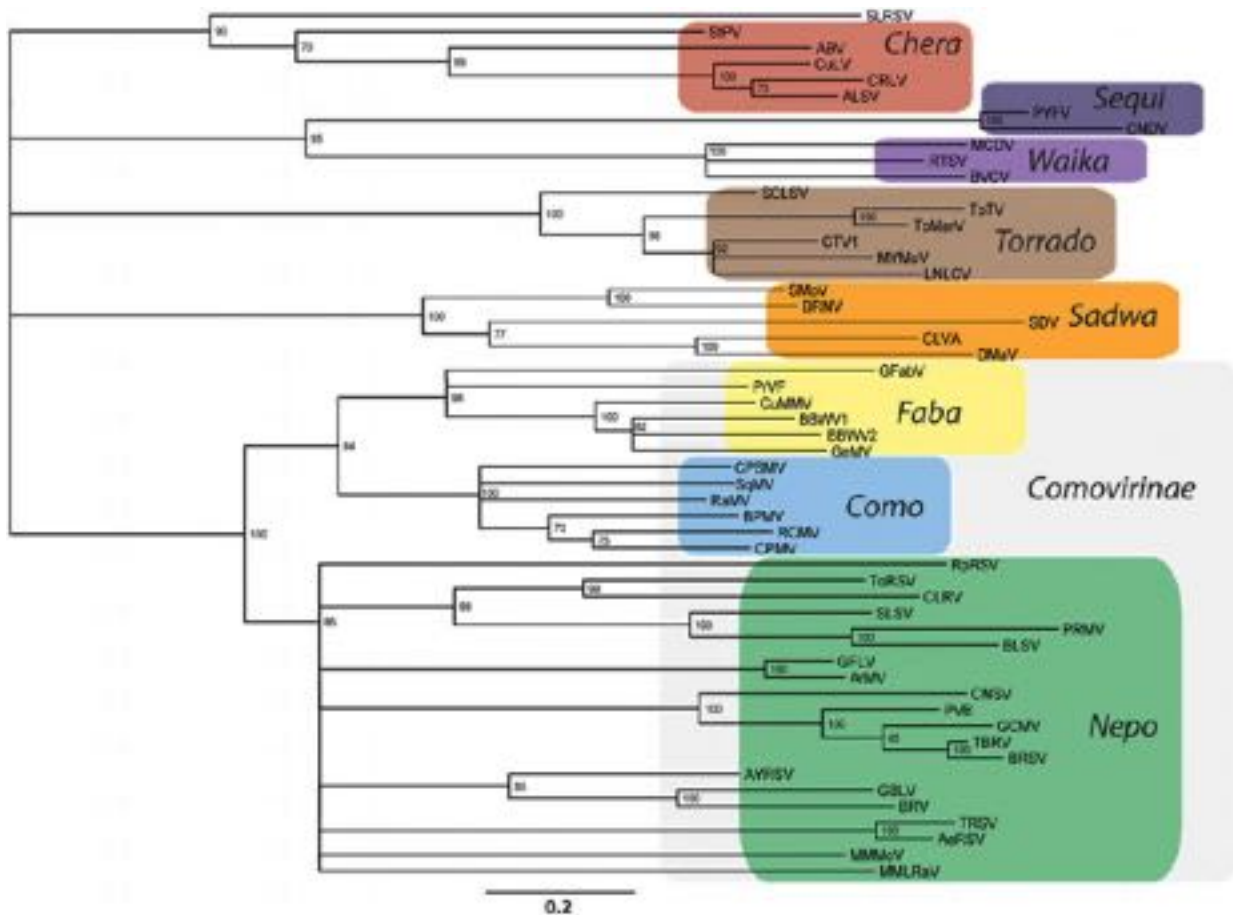


Fig. 2 Phylogenetic relationships within the family *Secoviridae*. The comoviruses and fabaviruses form two distinct branches within the subfamily *Comovirinae*, which includes the nepoviruses. Alignment figure courtesy of Prof. J. Thompson.

gradients (**Fig. 1(B)**). The exact sedimentation coefficients of the components vary slightly with the virus being analyzed but are typically in the range 50-60S for T, 90-100S for M and 110-120S for B components. The three components have identical protein compositions, containing 60 copies each of a large (L) and small (S) coat protein (CP). The sizes of the two proteins vary depending on the virus but lie in the ranges 37-49 kDa for the L protein and 18-26 kDa for the S protein. The difference between the three centrifugal components is in their RNA contents. Top components are devoid of RNA, while middle and bottom components each contain single molecules of RNA of approximately 3-4 kb and 6 kb, respectively. The two RNA molecules were originally termed middle (M) and bottom (B) component RNA after the nucleoprotein component from which they were isolated but are now termed RNA2 and RNA1, respectively (**Fig. 3(A)**).

Because of their differing RNA contents, the three components of comoviruses and fabaviruses also differ in density and can hence be separated by isopycnic centrifugation. Generally, T, M and B components have densities of approximately 1.29, 1.40 and 1.41 g ml⁻¹ respectively. However, the pattern obtained is often more complex than that seen with sucrose gradients and depends on the precise conditions used and virus being examined. For example, while BPMV gives the expected three components when centrifuged on cesium chloride gradients under a wide range of conditions, the bottom component of CPMV can be resolved into two forms of differing density under alkaline conditions.

Preparations of comoviruses are often not only centrifugally heterogeneous but can also be separated in two forms, fast and slow, electrophoretically. Both electrophoretic forms contain all three centrifugal components. The proportion of the two electrophoretic forms in a given viral preparation varies both with the time after infection at which the virus was isolated and the age of the preparation itself. Conversion of one form to the other is caused by loss of amino acids from the C terminus of the S protein. In the case of CPMV, this segment has been shown to be critical for virus assembly, and structural studies on Broad bean stain virus (BBSV) suggest that comoviruses share a common assembly mechanism. Though not studied extensively, it is likely that fabavirus particles also undergo a similar cleavage of the S protein.

In a number of instances, comovirus particles have been shown to contain polyamines, particularly spermidine. Such polyamines are believed to have a role in the neutralization of the negative charges of the RNA within the particles. It is the exchange of such polyamines for cesium ions that leads to the two forms of bottom component of CPMV seen under alkaline conditions.

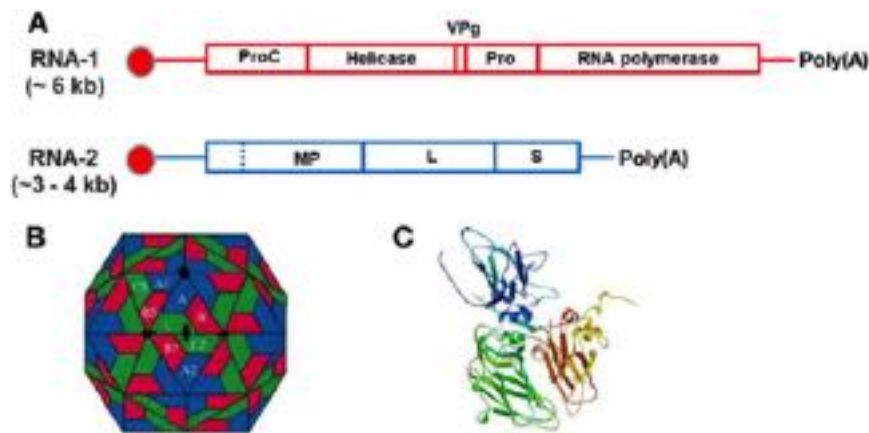


Fig. 3 Molecular details of comoviruses and fabaviruses. (a) Organization of the two single-stranded RNA molecules that comprise the genome. Both RNAs have a VPg (red circle) linked to their 5' ends and both are polyadenylated. RNA1 contains a single ORF that is processed by the 24K protease (Pro) to give the viral proteins required for RNA replication. The functions of the individual processed products are indicated. RNA2 encodes the structural CPs, Large (L) and Small (S) as well as the MP required for the movement of CPMV virions from cell-to-cell. The solid line indicates translation from the upstream initiation codon that results in an N-terminal extension to the MP, while translation from the downstream initiation codon indicated by the dotted line results in the MP. (b) Arrangement of the CP subunits of comoviruses and fabaviruses. The asymmetric unit consists of one single-domain S subunit (blue) and one two-domain L subunit (green and red), 60 copies of which are arranged with icosahedral symmetry to form a pseudo $T = 3$ ($P = 3$) capsid. (c) Ribbon diagram of the asymmetric unit, showing β -barrel structure of the L (green and red) and S (blue) subunits.

Viral Structure

X-ray crystallographic and cryo-EM studies on CPMV, BPMV, RCMV and BBSV have provided a detailed picture of the arrangement of the two viral CPs in the three-dimensional structure of comoviruses. Though, at the time of writing, there is no detailed structure of a fabavirus available, it is highly probable that their overall architecture is very similar. Comovirus virions are icosahedrally symmetrical, with 12 axes of fivefold and 20 axes of threefold symmetry and resemble a classic $T = 3$ particle (**Fig. 3(B)**). The two CPs taken together consist of three distinct β -barrel domains, two being derived from the L and one from the S protein. Thus, in common with the $T=3$ viruses, each CPMV particle is made up of 180 β -barrel structures. The S protein, with its single domain, is found at the fivefold symmetry axes and therefore occupies a position analogous to that of the A type subunits in $T = 3$ particles (**Fig. 3(c)**). The N- and C-terminal domains of the L protein occur at the threefold axes and occupy the positions equivalent to those of the C and B type subunits of a $T = 3$ particle respectively. Comovirus (and, by inference, fabavirus) capsids are also structurally homologous to those of the picornaviruses, with the N- and C-terminal domains of the L protein being equivalent to viral protein VP2 and viral protein VP3, respectively, and the S protein being equivalent to viral protein VP1. In the case of BPMV, ordered segments of RNA2 could be seen in the crystallographic structure of M components. RNA has also been detected in the cryo-EM structures of the M and B components of CPMV and in an unfractionated mixture of BBSV.

Genome Structure

The genomes of both comoviruses and fabaviruses consist of the two RNAs (RNA1 and RNA2) encapsidated in bottom and middle components. Both RNA molecules are positive-sense and both are required for an infection of whole plants. In the case of CPMV, RNA1 is capable of independent replication in individual plant cells but this results in the establishment of gene silencing in the absence of RNA2, which encodes a suppressor of gene silencing. Both genomic RNAs have a small basic protein (VPg) covalently linked to their 5' termini and are polyadenylated. The VPg is linked to the viral RNA via the β -hydroxyl group of its N-terminal serine residue. The VPg is not required for the viral RNAs to be infectious.

The complete nt sequences of both RNAs of most comoviruses and fabaviruses have been determined (**Table 1**); in addition, several partial sequences are also available. The two genomic RNAs of a given virus have no sequence homology apart from at the 5' and 3' termini. Full-length infectious cDNA clones of several viruses have been constructed (**Table 1**), allowing the genomes of these viruses to be manipulated.

Expression of the Viral Genome

Both genomic RNAs contain a single long open reading frame (ORF) which occupies over 80% of the length of the RNA. A combination of *in vitro* translation and protoplast studies has unraveled the basic mechanism of expression of the genomic RNAs

of CPMV, the type species of the genus *Comovirus*. More limited studies on other members of the group suggest that the mode of gene expression deduced for CPMV is applicable to all members of the genera.

Both RNAs of the comoviruses and fabaviruses are expressed through the synthesis and subsequent cleavage of large precursor polyproteins. On RNA1, initiation of translation occurs at the first AUG encountered on the sequence and results in the synthesis of a protein of approximately 200 kDa. This initial product undergoes rapid co-translational autoproteolysis to give proteins with apparent sizes of 32 and 170 kDa (the 32 and 170K proteins). The 170K protein undergoes further cleavages to give the range of virus-specific proteins, the functions of which are indicated in Fig. 2(A). The proteolysis of the polyprotein is catalysed by the region marked Pro in Fig. 2(A).

Initiation of translation of RNA2 occurs at two different positions on the RNA and results in the synthesis of two carboxy coterminal proteins. This double initiation phenomenon, which occurs as a result of “leaky scanning” of two in-frame AUG codons, is found with the RNA2 molecules of all comoviruses and fabaviruses. The distance between the two initiation codons varies significantly between different viruses, as does the sequence of the encoded aa. Both RNA2-encoded primary translation products are cleaved by the RNA1-encoded Pro to give the movement protein (MP) and a version with an N-terminal extension as well as the two viral CPs, L and S. Processing of the RNA2-encoded polyproteins at the site between MP and L CP also requires the presence of the protease cofactor (ProC) as well as Pro.

Functions of the Viral Proteins

Functions have been ascribed to most of the regions of the polyproteins encoded by both RNA1 and RNA2 based on extensive investigations with CPMV. Far less information is available for other comoviruses and fabaviruses, with what little that is available indicating that the analyses conducted with CPMV are generally applicable. Therefore, the reader is referred to the article on CPMV for a detailed description of the functions of the viral proteins.

Relationships With Other Viruses

Together with the genus *Nepovirus*, the genera *Comovirus* and *Fabavirus* are grouped within the subfamily *Comovirinae*, in the family *Secoviridae*. Consideration of genome structure and organization, translational strategy and aa homologies between the virus-encoded proteins has led to the family *Secoviridae* being grouped with the family *Picornaviridae*, within the order *Picornavirales*. Members of this order are all non-enveloped icosahedral positive-strand RNA viruses with 3'-polydenylated genomic RNAs which have the VPg covalently linked to their 5' ends. All members of the order have a similar mode of gene expression which involves the synthesis of large precursor polyproteins and their subsequent cleavage by a virus-encoded proteinase. The members of the order, all contain the similar gene order, membrane-bound protein-VPg-proteinase-polymerase, a feature shared with members of the filamentous family *Potyviridae*.

Use in Biotechnology

CPMV has been extensively used as a vector for the expression of foreign peptides and proteins in plants. The only other comovirus that has been deployed in biotechnology is BPMV. Several BPMV RNA2-based vectors have been developed which permit the insertion of foreign sequences for either protein expression or gene silencing in soybean. At present there are no reports of the use of a fabavirus in biotechnology.

Geographic Distribution

Individual comoviruses tend to have a somewhat restricted distribution. Clearly, the apparent distribution of a virus is governed, at least in part, by whether or not infection is diagnosed and reported. The following list, therefore, contains only those geographical areas where the presence of the virus has, to date, been confirmed. CPMV: Nigeria, Kenya, Tanzania, Japan, Surinam and Cuba; *Andean potato mottle virus* (APMV): Central and South America; BPMV: USA; *Bean rugose mosaic virus* (BRMV): Central America; BBSV: Europe and Africa; *Broad bean true mosaic virus* (BBTMV): Europe and North-west Africa; *Cowpea severe mosaic virus* (CPSMV): North, Central and South America; *Glycine mosaic virus*-GMV: Australia; *Pea mild mosaic virus* (PMiMV): New Zealand; *Quail pea mosaic virus* (QPMV): North and Central America; *Radish mosaic virus* (RaMV): USA, Japan and Europe; RCMV: Europe; *Squash mosaic virus* (SqMV): North Africa, Israel, China, Japan, Australia and North, Central and South America; *Ullucus virus C* (UCV): Peru and Bolivia.

Fabaviruses occur throughout the world and a wide variety of different isolates have been reported from different geographical regions have been reported.

Host Range

The host ranges of individual comoviruses tend to be rather narrow. Of the 15 members of the genus, ten (CPMV, BPMV, BRMV, BBSV, BBTMV, CPSMV, GMV, *Pea green mottle virus*-PGMV, PMiMV, QPMV and RCMV), in nature, infect predominantly, if not exclusively, legumes. The other four members of the group, APMV, RaMV, SqMV and UCV, infect predominantly members of the *Solanaceae*, *Cruciferae*, *Cucurbitaceae* and *Basellaceae* respectively.

By contrast, the experimental host range of Fabaviruses is wide: 328 species of 186 genera in 44 families have been listed as infectable by BBWV. *Lamium mild mosaic virus* (LMMV) has been less extensively tested but differs from BBWV in not infecting, for example, cowpea (*Vigna unguiculata*) or tomato (*Lycopersicon officinale* cv. Marmande).

Transmission

In nature, comoviruses are usually transmitted by leaf-feeding beetles, especially by members of the family Chrysomelidae. However, CPMV has been shown also to be transmitted by thrips and grasshoppers. The beetle vectors can acquire the virus by feeding for as little as one minute and can retain and transmit the virus for a period of days or weeks. Experimentally, all comoviruses are mechanically transmissible. In the case of BBSV, BBTV, CPMV, CPSMV and SqMV, transmission through seed has been reported.

Fabaviruses are transmitted by aphids in a non-persistent manner. The list of known vector aphids extends to some 20 species including *Acyrtosiphon pisum*, *Aphis craccivora*, *A. nasturtii* and *Macrosiphum euphorbiae*. However, there are unpredictable and unexplained differences between fabavirus isolates in the efficiencies of their transmission by different aphid species. There is only one report showing that BBWV is seed-transmitted. The transmission of fabaviruses by aphids is not known to require the presence in source plants of another 'helper' virus.

Epidemiology

The primary sources for infection of crop plants with comoviruses are transmission through seed and infection from wild hosts act as reservoirs of the viruses. It is also possible that virus which has overwintered in the beetle vector can act as a primary source of infection. Within-field spread is by the beetle vectors. Fabaviruses can be readily spread by their aphid vectors.

Symptomatology and Cytopathology

Generally, each virus has certain hosts in which the virus can spread systematically, causing mosaic or mottling symptoms, and other hosts in which the infection is confined to local lesions. Stunting of the host plant is sometimes observed, as is occasional systematic wilting and seed discoloration.

Infection of plant cells with comoviruses results in a number of characteristic cytopathological changes. These include the appearance of viral particles, either individually or as crystalline arrays, in the cytoplasm, a proliferation of cell membranes and vesicles in the cytoplasm, the appearance of amorphous inclusion bodies near or surrounding the nucleus and a variety of modifications to plasmodesmata.

Virions of all the fabaviruses occur in the cell cytoplasm of foliar mesophyll and epidermis. In some isolates, virions, virus-like shells and probably also capsid protein assemble into a variety of ordered arrays (tubules, rings, square patterns) that can be recognized by electron microscopy. Amorphous X bodies (spindle-shaped fibrous or crystalline inclusions visible using light microscopy) may have some diagnostic value.

Economic Importance

At present, BPMV, CPMV, CPSMV and RCMV are considered to be significant pathogens of legumes. Infection of soybeans with BPMV alone can cause yields to be reduced by 10%–17%. However, this figure can rise to 60% in plants doubly infected with BPMV and the potyvirus, Soybean mosaic virus. In Nigeria, infection of cowpeas with CPMV causes a great reduction in leaf area, flower production and yield. Infection of cowpeas with CPSMV has been shown to cause a 50% reduction in plant fresh weight and in the number and weight of pods. RCMV is of economic importance in forage production as it sometimes heavily infects clover. SqMV is a significant pathogen of cucurbits, problems associated with it being exacerbated by its high rate of seed transmission. Fabaviruses kill lettuce, bean and spinach plants and cause necrotic streaks in foliage of some pea cultivars. Symptoms in other crops are usually transient and have limited commercial significance – except in the People's Republic of China where complex intercropping patterns favor damaging epiphytotic also involving celery (*Apium graveolens*) and pakchoi (*Brassica chinensis*).

Further Reading

- Cooper, J.I., Lomonossoff, G.P., 2000. Fabaviruses. In: Maloy, O.C., Murray, T.D. (Eds.), *Encyclopedia of Plant Pathology*. New York, NY: John Wiley and Sons, Inc., pp. 445–446.
- Ferriol, I., Ferrer, R.M., Luis-Arteaga, M., *et al.*, 2014. Genetic variability and evolution of Broad bean wilt virus 1: Role of recombination, selection and gene flow. *Archives of Virology* 159, 779–784.
- Gergerich, R.C., Scott, H.A., 1996. Comoviruses: Transmission, epidemiology, and control. In: Harrison, B.D., Murrant, A.F. (Eds.), *The Plant Viruses 5: Polyhedral virions and bipartite RNA genomes*. New York: Plenum Press, p. 77.
- Ghabrial, S.A., 2018. My life and virus research journey. *Annual Review of Virology* 5, 1–32.
- Goldbach, R.W., Wellink, J., 1996. Comoviruses: Molecular biology and replication. In: Harrison, B.D., Murrant, A.F. (Eds.), *The Plant Viruses 5: Polyhedral virions and bipartite RNA genomes*. New York: Plenum Press, p. 35.
- Kobayashi, Y.O., Kobayashi, A., Nakano, M., *et al.*, 2003. Analysis of genetic relations between broad bean wilt virus 1 and broad bean wilt virus 2. *Journal of General Plant Pathology* 69, 320–326.
- Lecorre, F., Lai-Kee-Him, J., Blanc, S., *et al.*, 2019. The cryo-electron microscopy structure of Broad bean stain virus suggests a common capsid assembly mechanism among comoviruses. *Virology* 530, 75–84.
- Lin, T., Johnson, J.E., 2003. Structure of picorna-like plant viruses: Implications and applications. *Advances in Virus Research* 62, 167–239.
- Lisa, V., Boccardo, G., 1996. Fabaviruses: Broad bean wilt and allied viruses. In: Harrison, B.D., Murrant, A.F. (Eds.), *The Plant Viruses 5: Polyhedral virions and bipartite RNA genomes 1996*. New York: Plenum Press, pp. 229–250.
- Lomonossoff, G.P., Ghabrial, S.A., 2000. Comoviruses. In: Maloy, O.C., Murray, T.D. (Eds.), *Encyclopedia of Plant Pathology*. New York, NY: John Wiley and Sons, Inc., pp. 239–242.
- Pflieger, S., Blanchet, S., Meziadi, C., *et al.*, 2014. The "one-step" bean pod mottle virus (BPMV)-derived vector is a functional genomics tool for efficient overexpression of heterologous protein, virus-induced gene silencing and genetic mapping of BPMV R-gene in common bean (*Phaseolus vulgaris* L.). *BMC Plant Biology* 14, 232–248.
- Zhang, C., Ghabrial, S.A., 2006. Development of bean pod mottle virus-based vectors for stable protein expression and sequence-specific virus-induced gene silencing in soybean. *Virology* 344, 401–411.

Relevant Websites

- https://viralzone.expasy.org/728?outline=all_by_species
Comovirinae.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/secoviridae/589/genus
Comovirus.
- <http://viperd.b.scripps.edu/genuslist.php?genus=Comovirus>
Virus Genus Index
VIPERdb
Scripps Research.

Cotton Leaf Curl Disease (*Geminiviridae*)

Nasim Ahmed, Imran Amin, and Shahid Mansoor, National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S. Mansoor, I. Amin, R.W. Briddon, Cotton Leaf Curl Disease, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00703-2.

Nomenclature

aa Amino acid(s)
amiRNA Artificial microRNA
CP Coat protein or capsid protein
DAS Days after sowing
kb Kilobase
kDa Kilo Dalton
mRep Master replication-associated protein
nm Nanometer(s)
nt Nucleotide(s)

NW New world
OW Old world
PDR Pathogen derived resistance
PTGS Post-transcriptional gene silencing
Rep Replication initiator protein
SCR Satellite conserved region
ssDNA Single-stranded deoxyribonucleic acid
TGS Transcriptional gene silencing
TrAP Transcriptional activator protein

Glossary

Epidemic A wide spread of an infectious disease in the population at a specific time.

Host plant resistance The natural host plant species have the ability to control the parasite population. This is the most desired approach to control disease.

Koch's postulates The criteria first time developed in 1890 by Robert Koch that establishes the relationship between a disease and microbe. Following conditions must be fulfilled; the pathogen should be present in all symptomatic hosts, and after its isolation from the host it should be able to grow in pure culture, and when

reintroduced into a healthy host it must cause the disease. At the end, the pathogen must be isolated from this infected host.

Pathogen-derived resistance The type of resistance to specific pathogens that is created by engineering host to express genes or sequences derived from the pathogen, it interferes with the infection caused by this pathogen or its most related pathogen which has the same sequence.

Resistance breaking strain A strain of pathogen which breaks the resistance of host against a specific pathogen.

Introduction

Cotton leaf curl disease (CLCuD) is the most destructive disease of cotton (*Gossypium* spp. L.) and also for several other plants in the family *Malvaceae*. The disease is widespread in Africa (mainly Sudan, Egypt, Malawi, Nigeria and South Africa) and southern Asia where it is endemic across Pakistan and northwestern India. Recently, it is also been reported from southern China in non-cotton plants. Typical symptoms exhibited by CLCuD affected plants include vein thickening, upward or downward leaf curling, and on the main vein (underside the leaves) enations are formed that evolve into cup-shaped leaf-like structures (Fig. 1). CLCuD affected cotton plants are conspicuously greener than non-infected plants which is attributed to a higher density of chloroplasts in infected tissues. Symptoms of CLCuD vary between different varieties of cotton and also depend on the age of the infected plants. Plants infected at the early stage show severe symptoms like tightly rolled leaves, highly stunted growth and subsequently they produce no harvestable lint whereas plants which are infected at the late stage show just mild symptoms and their yield is not significantly reduced. CLCuD is caused by begomoviruses (genus *Begomovirus*, family *Geminiviridae*) in association with betasatellites (*Tolecusatellitidae*) and alphasatellites (*Alphasatellitidae*), all being transmitted by insect vector whiteflies (*Bemisia tabaci*). The disease complex is distinct between the two continents.

Since prehistoric times, cotton is being grown in almost all tropical and subtropical areas of the world and it is increasingly adapted to temperate climates. China, India, USA, Pakistan and Brazil are the major cotton producing countries in the world. The genus *Gossypium* consists of approximately 52 species, five of which are tetraploid, and all others are diploid. Only four species are cultivated which produce spinnable fiber for the textile industry. These include two allotetraploid species, *G. barbadense* (Pima/Egyptian cotton) and *G. hirsutum* and two diploid species, *G. arboreum* and *G. herbaceum*. *G. hirsutum* (also known as upland cotton) accounts for more than 90% of the cotton production in the world. Diploid cotton species are native to Asia and Africa and they were largely grown prior to *G. hirsutum* species that was introduced from the New world (NW). The Old world (OW) native species have natural resistance to CLCuD.



Fig. 1 Infected cotton (*Gossypium hirsutum*) plants showing typical symptoms of CLCuD. Symptoms include vein darkening, vein swelling, upward curling of the leaves and enations on the undersides of leaves (left). A severely infected plant showing a reduced leaf size, downward curling of leaf, leaf crumpling and severely stunted growth (right).

History of CLCuD

CLCuD in Africa

CLCuD was reported for the first time in 1912 from Nigeria, where it was sporadic and remained a minor problem. Its second outbreak occurred in the same country in 1924. Later in the year 1924, CLCuD was also reported from Sudan and in 1926 from Tanzania. In North-East Africa, (mostly Egypt and Sudan) species of cotton mostly cultivated is *G. barbadense*. CLCuD is endemic in these areas, although it is only sporadically a problem. An epidemic was reported in 1927–28 in Sudan which stimulated interest in this disease, and it was shown that whiteflies were responsible for its transmission. At that time extensive research established that the disease is also transmissible through graft and a virus-like agent was suspected as the causative agent although the actual causing agent was not identified. In those days the disease was controlled by imposing a period in which cotton was not grown and later they cultivated virus resistant *G. barbadense* varieties. CLCuD sporadically occurs in Africa but with no major economic losses.

CLCuD in Southern Asia

Prior to the 1980s, CLCuD was found sporadically in the Indian subcontinent. During the early-1990s, in Pakistan, the cotton production started to suffer heavy losses due to an epidemic of CLCuD. The disease was first reported near the city of Multan and soon it spread to almost all major cotton growing areas of Pakistan and also into the western part of India. Most probably this epidemic of CLCuD was attributed to the large-scale cultivation of two cotton varieties viz. CIM70 and S-12 which were high-yielding, but unfortunately highly susceptible to this local disease. *G. arboreum*, the native species of cotton is naturally immune to CLCuD, but this species does not produce cotton lint of high quality demanded by the processing industry. However, still farmers grow this species on a small scale. At the end of the 1990s, the cotton production reached again pre-epidemic levels, through the development and gradual adoption of tolerant and resistant cotton varieties. Unfortunately, this resistance did not last a long time, as it was broken in 2001, by the emergence of a resistance breaking combination of a virus strain of begomovirus and a betasatellite. This resulted in the 2nd epidemic of CLCuD in this region. Currently, there is no commercial cotton variety resistant to this disease complex. In 2015, begomoviruses associated with the first epidemic reappeared in cultivated cotton fields and the resistance breaking strain disappeared after 2015.

Etiology of CLCuD

In both CLCuD affected regions [Africa (Sudan and Egypt) and Indian subcontinent] the etiology of the disease has been determined. In both regions, it was found that the disease is caused by monopartite begomoviruses (genus *Begomovirus*, family; *Geminiviridae*) associated with single-stranded DNA satellite molecules viz., betasatellites (*Tolecusatellitidae*) and alphasatellites (*Alphasatellitidae*). Betasatellites encode a single protein β C1 which is a pathogenicity determinant, whereas alphasatellites play no apparent role in the etiology of CLCuD. However, it has been found in all cases of CLCuD in the field.

Elucidation of the Etiology of CLCuD

In Africa, since 2002 CLCuD is caused by a single species of begomovirus viz. *Cotton leaf curl Gezira virus* (CLCGeZV) in association with *Cotton leaf curl Gezira betasatellite* (CLCGeZB) and *Cotton leaf curl Gezira alphasatellite* (CLCGeZA). Although this disease complex remained permanently associated with CLCuD in Africa, recently and for the first time, in west and central Africa, a bipartite begomovirus, *Cotton yellow mosaic virus* (CYMV) was also reported in a non-cultivated cotton species i.e. *Gossypium raimondii*.

CLCuD caused huge losses to cotton industries in both Pakistan and India in the early 1990s. This initiated a deep investigation into the etiology of the CLCuD. Initially, several species of begomoviruses were isolated and sequenced from cotton in both countries but their infectivity was not tested in cotton. So, the Koch's postulate was not fulfilled at that time. Later, efforts were made to introduce one of these begomoviruses i.e., CLCMuIV into cotton which showed atypical and mild symptoms. These results indicated that either another component was required for the disease or CLCMuIV was not responsible for the CLCuD.

Begomoviruses are typically divided into two groups i.e., monopartite and bipartite based on their genomic components. The bipartite begomovirus has two component DNA-A and DNA-B approximately equal size (2.7 kb) whereas monopartite begomoviruses have single genomic components equivalent to DNA-A of bipartite begomoviruses. Old world monopartite begomoviruses are often also associated with DNA satellites referred to as betasatellites and alphasatellites.

During the first epidemic, it was troubling that CLCMuIV alone was unable to infect cotton to develop the typical symptoms of CLCuD and the same situation was noticed with several other begomoviruses. When CLCuD-affected plants were further analyzed then a novel additional, subviral component was identified. Similar components had been identified in association with several monopartite begomoviruses. These components had been named as DNA-1 (now known as alphasatellite). However, in the case of CLCuD, it was found that alphasatellites did not play an essential role in the disease. Nevertheless, this finding promoted the reinvestigation of subviral DNA molecules, which might be associated with monopartite begomoviruses. In the mean-time, with a different study done with the monopartite *Ageratum yellow vein virus* (AYVV), another sub-genomic DNA component, named DNA- β (betasatellite), was identified. Shortly after this discovery a similar component DNA- β was also identified associated with CLCMuIV. Inoculation of CLCMuIV with the associated betasatellite induced all typical symptoms of CLCuD in cotton, hereby establishing the etiology of the disease and Koch's postulate was fulfilled. Betasatellites encode a single protein referred to as β CI on its complementary strand which acts as a pathogenicity determinant.

Components of the CLCuD Complex

Begomoviruses Associated With CLCuD

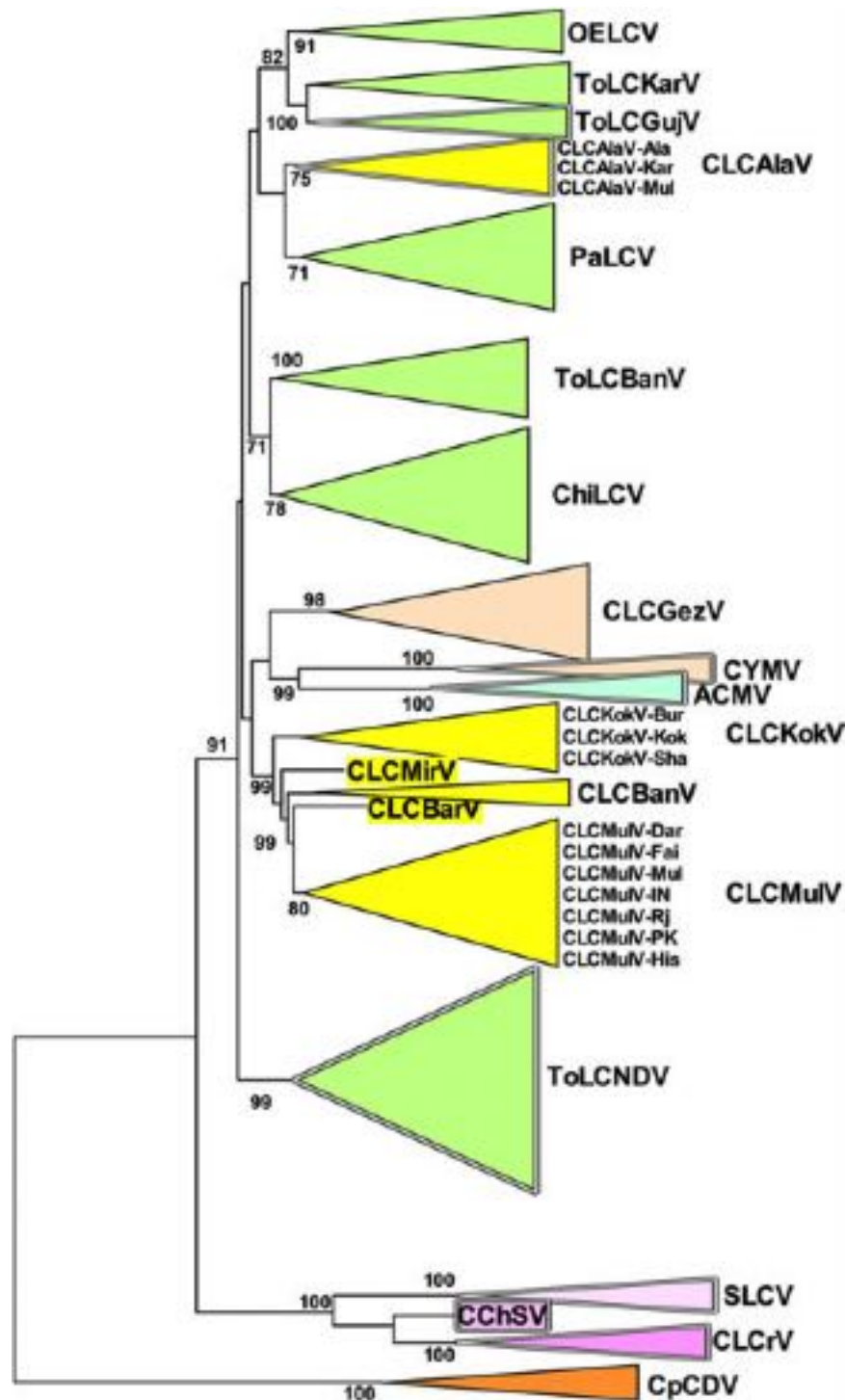
CLCuD in Africa

A single OW monopartite begomovirus viz. *Cotton leaf curl Gezira virus* (CLCGeZV), has been found in association with CLCuD in Africa. However, recently *Cotton yellow mosaic virus* (CYMV); a bipartite OW begomovirus has been isolated from non-cultivated cotton (*Gossypium raimondii*) for the first time in west and central Africa. CLCGeZV is more closely related to begomoviruses originating from the African Mediterranean region infecting malvaceous plants, and distantly related to begomoviruses associated with the Asian CLCuD. Begomoviruses are geographically related, supporting the distinct evolutionary origin of begomoviruses causing CLCuD in Asia and Africa. In 2005, CLCGeZV-EG (Egyptian strain) was isolated from cotton in Sindh, Pakistan, in a mixed infection, but its cognate betasatellite viz., CLCGeZB was not found, instead it was trans-replicating two asian betasatellites; CLCMuLB and Chili leaf curl betasatellite (ChLCB). Interestingly, until now, CLCGeZV has not been reported from Punjab in Pakistan and from India.

CLCuD in the Indian subcontinent

The first epidemic of CLCuD in the Indian subcontinent was associated with several distinct begomovirus species which are CLCMuIV, *Cotton leaf curl Alabad virus* (CLCAIaV), *Cotton leaf curl Kokhran virus* (CLCKokV), *Tomato leaf curl Bangalore virus* (ToLCBanV) and *Papaya leaf curl virus* (PaLCV). They were associated with the disease either in single or multiple co-infections (Fig. 2). Some of these viruses induced CLCuD in cotton plants when they were co-inoculated with CLCMuLB whereas, only ToLCBanV was able to cause tomato leaf curl disease in the absence of a betasatellite. However recently, it was shown to trans-replicate multiple betasatellites which resulted in the increased pathogenicity in tomato and a model plant. CLCBanV was reported from *G. barbadense* in southern India but it was not affecting the north of India where CLCuD was epidemic. In addition, experiments to demonstrate that it could infect cotton and cause CLCuD were not successful.

In the late 1990s scientists had developed some cotton resistant varieties which helped to restore the production to pre-epidemic levels. But, unfortunately in the year 2001, a resistance breaking begomovirus appeared. It was a recombinant of CLCMuIV and CLCKokV and referred to as CLCKokV-Burewala strain (CLCKokV-Bur). It was first identified from Punjab in Pakistan then also spread to north-western India and it became the second epidemic of CLCuD in the region. This strain had a truncated TrAP (AC2) protein which was speculated to cause the resistance breaking. However, since that time, the existence of CLCKokV-Bur with an intact C2 gene has been shown. Therefore, the reason for this strain to break the resistance and cause the second epidemic has not been elucidated.



In 2014, two bipartite begomoviruses have also been identified from cultivated cotton plants in Pakistan. The most important one is *Tomato leaf curl New Delhi virus* (ToLCNDV), which was frequently identified from cotton in both Sindh and Punjab provinces of Pakistan, whereas the other was *Tomato leaf curl virus* (ToLCV) reported from the same country. Moreover, the identification of *Chickpea chlorotic dwarf virus* (CpCDV), a *Mastrevirus* (*Geminiviridae*), and *Okra enation leaf curl virus* (OELCV), a typical begomovirus from infected cotton, has further complicated the situation. It has been shown experimentally that ToLCNDV enhances the accumulation of CLCMuB-Bur by synergistically interacting with CLCKokV-Bur in the infected plants. Moreover,

CLCMulB has been shown to be trans-replicated by the DNA-A component of ToLCNDV, in the absence of its DNA-B component, and causing typical symptoms of CLCuD in cotton. Interestingly, in 2015, multiple begomoviruses which were associated with the first epidemic of CLCuD in 1990s were found in cultivated cotton in Vehari, Punjab, in Pakistan and in the same year CLCMulV was exclusively identified from Punjab, in India. In both studies CLCKokV-Bur was not identified in cultivated cotton.

Identification of multiple begomoviruses with a single species of betasatellite in cotton while causing a single disease, suggests that these begomoviruses, also identified from many alternative hosts, most probably moved to cotton, where they could interact with CLCMulB to cause a unique cotton disease. Moreover, recently it was reported that non-cultivated cotton species from Multan, Punjab, in Pakistan (the region considered as hotspot for begomoviruses and satellites associated with CLCuD) were continuously infected by multiple begomoviruses, associated with multiple DNA satellites but with the unique identical betasatellite, CLCMulB. These plants are therefore reservoirs for a collection of begomoviruses and satellites. The presence of multiple begomoviruses makes it easy to overcome resistance against individual viruses and facilitates adaptation to other hosts. In addition, in the Indian subcontinent there are many other monopartite begomoviruses that are not found in cotton, possibly suggesting that these few species of begomoviruses recently adapted to cotton. Moreover, some begomoviruses that are not reported in the region, such as ToLCV, originating from Australia, and *Ageratum yellow vein virus* originating from southeast Asia, as well as Tomato yellow leaf curl virus (TYLCV), were shown capable experimentally to trans-replicate the CLCMulB (the unique betasatellite associated with CLCuD). The capability of CLCMulB to interact with a multitude of begomoviruses provides a huge flexibility for the cotton disease to occur and makes it a real threat to cotton production worldwide.

It has been shown experimentally that CLCMulV is efficiently transmitted by the whitefly species Asia II 1 as compared to Middle East Asia minor 1 (MEAM1), and Asia II 1 is the dominant species in major cotton growing areas of Pakistan and India. Moreover, the Asia II 1 whitefly genome has been sequenced, and compared to that of MEAM1, revealing variations in genes involved in insecticide resistance and virus transmission. These findings may explain the rapid spread of CLCuD but also its current limitation to the Indian subcontinent. If Asia II 1 was to become dominant in other parts of the world, most probably it may allow the rapid transfer of CLCuD associated begomoviruses and satellites to become a serious global threat. In addition, the Faisalabad strain of CLCMulV was recently isolated from *Hibiscus rosasinensis* and from okra in the southern China, but it did not become epidemic in China, most probably due to the local whitefly species.

With all this information, a third epidemic of CLCuD could be expected to occur soon, possibly with distinct begomoviruses, but with the same associated DNA satellite components (Fig. 3). A recent Pakistan report survey about whiteflies and begomoviruses from all major cotton growing areas (both Sindh and Punjab in Pakistan), indicates that the dominant whiteflies are of the species Asia II 1 and they all carry CLCMulV-Raj, a new strain of CLCMulV. The most recent isolates of CLCMulV from India and Pakistan indicate a novel clade in the phylogenetic tree of begomoviruses associated with CLCuD and this may be associated with a potential third epidemic of CLCuD. Similarly, the most recent strain of betasatellite viz., CLCMulB^{Veh} has been identified from whiteflies as well as cotton, so it may also be associated with a possible third epidemic of the disease.

CLCuD-Associated Alphasatellite Components

Alphasatellites (*Alphasatellitidae*) are the satellite components associated with disease complexes of monopartite begomovirus (Fig. 3). They are circular, single-stranded (ss)DNA molecules have about half the size (~1380 nt) of their helper viruses. With a single gene on virion-sense orientation which encodes a protein required for their replication (rolling-circle replication) initiation. The product of this gene is called replication associated protein (Rep) of the plant-infecting ssDNA viruses. In the cells of host plants, alphasatellites are capable of autonomous replication so for their replication they do not depend on their helper viruses. However, for movement within plants and transmission between plants, these molecules require the helper geminiviruses. Presumably, they are trans-encapsidated in the coat protein of the helper geminiviruses. Although alphasatellites are not apparently required to induce the disease symptoms in host plants it was however recently reported that Rep of alphasatellites may have a role in gene silencing mechanism in plants.

Fig. 2 Neighbor-joining phylogenetic tree of begomoviruses was reconstructed using bootstrap method with 1000 replicates (percentage bootstrap value is shown on the right corner of each branch when >50%). The size of the triangles is roughly proportionate to the number of isolates in the cluster. The color code is for identifying viruses originally described on cotton in the Indian subcontinent (yellow), or for newly identified viruses on cotton but from other origin in the Indian subcontinent (green), light brown is for African viruses, light blue is for ACMV identified in Pakistan but originally described in Africa, pink is from cotton viruses from America and orange is for a mastrevirus identified in cotton in Pakistan. The triangles with a double line are those for the bipartite viruses versus 1 line for the monopartite viruses. *Abbreviations used for different species and strains are as follows:* CChSV (Cotton chlorotic spot virus), ChiLCV (Chili leaf curl virus), CLCAIaV-Ala (Cotton leaf curl Alabad virus-Alabad), CLCAIaV-Har (Cotton leaf curl Alabad virus-Haryana), CLCAIaV-Kar (Cotton leaf curl Alabad virus-Karnal), CLCAIaV-Lob (Cotton leaf curl Alabad virus-Lobatum), CLCAIaV-Mul (Cotton leaf curl Alabad virus-Multan), CLCBaV (Cotton leaf curl Bangalore virus), CLCBarV (Cotton leaf curl Barasat virus), CLCGeV (Cotton leaf curl Gezira virus), CLCKokV-Bur (Cotton leaf curl Kokhran virus-Burewala), CLCKokV-Kok (Cotton leaf curl Kokhran virus-Kokhran), CLCKokV-Sha (Cotton leaf curl Kokhran virus-Shadadpur), CLCMirV (Cotton leaf curl Mirzapur virus), CLCMulV-Dar (Cotton leaf curl Multan virus-Darvinii), CLCMulV-Fai (Cotton leaf curl Multan virus-Faisalabad), CLCMulV-His (Cotton leaf curl Multan virus-Hisar), CLCMulV-PK (Cotton leaf curl Multan virus-Pakistan), CLCMulV-Raj (Cotton leaf curl Multan virus-Rajasthan), CLCrV (Cotton leaf crumple virus), CpCDV (Chickpea chlorotic dwarf virus), CYMV-DNA-A (Cotton yellow mosaic virus-DNA-A genome), OELCV (Okra enation leaf curl virus), PaLCV (Papaya leaf curl virus), SLCV (Squash leaf curl virus), ToLCBaV (Tomato leaf curl Bangalore virus), ToLCGujV-DNA-A (Tomato leaf curl Gujarat virus-DNA-A genome), ToLCNDV-DNA-A (Tomato leaf curl New Delhi virus-DNA-A genome).

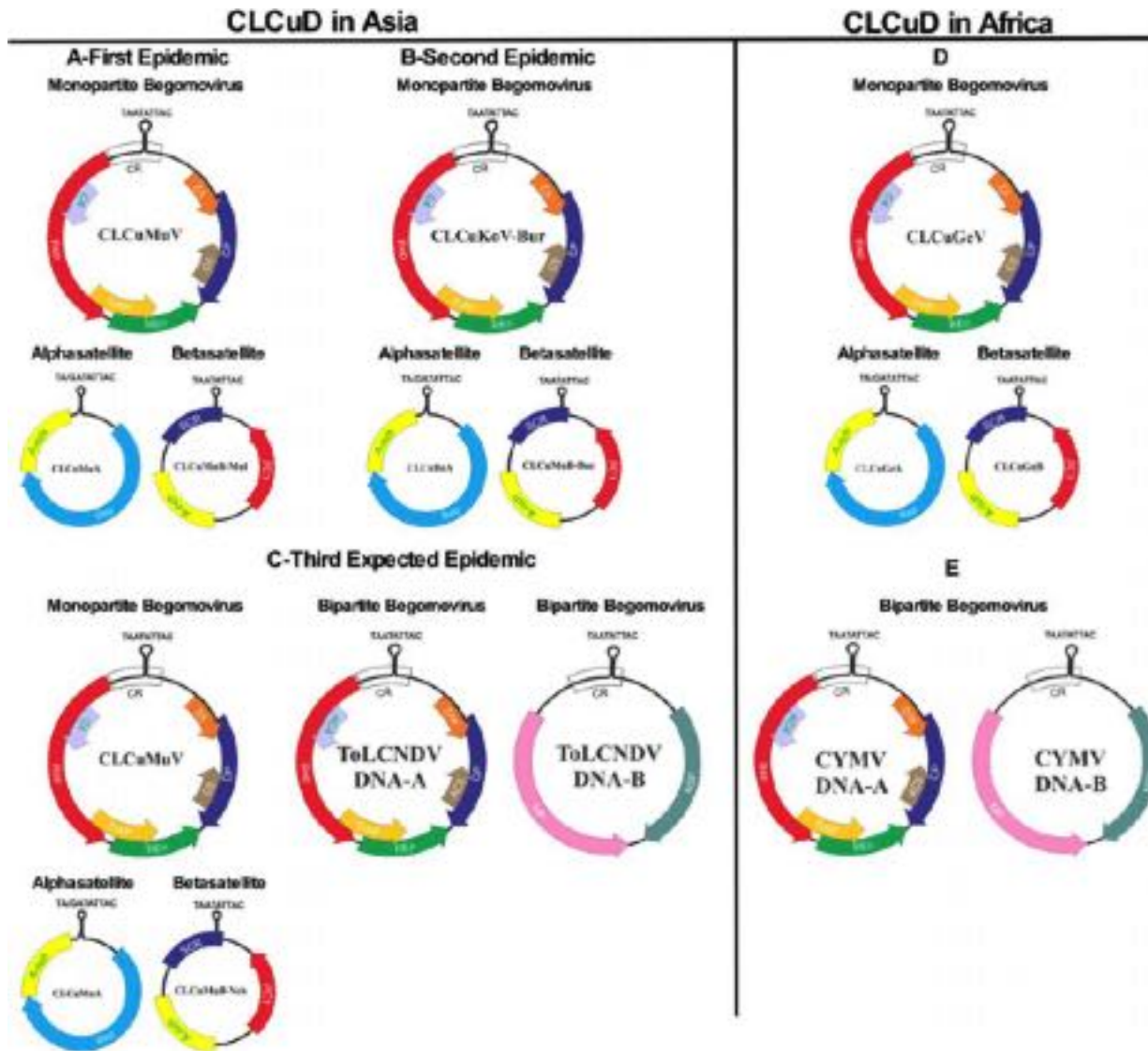


Fig. 3 Genomic organization of begomoviruses and associated alphasatellites and betasatellites. A; genome organization of a monopartite begomovirus and associated satellites in the first epidemic of CLCuD in Asia, B; genome organization of a monopartite begomovirus and associated satellites found in the second epidemic of CLCuD in Asia, C; genome organization of a monopartite begomovirus and associated satellites expected in a third epidemic (left), and genome organization of a bipartite begomovirus expected in a third epidemic (right), D; genome organization of a monopartite begomovirus and associated satellites found in Africa and E; genome organization of a bipartite begomovirus found in Africa. *Abbreviations*: CR; conserved region, Rep; Replication-associated protein, TrAP; Transcription associated protein, Ren; Replication Enhancer protein, CP; Coat Protein, MP; Movement protein, NSP; Nuclear Shuttle protein, β C1; Beta C1; SCR; Satellite Conserved Region. A-rich; Adenine rich region.

Surprisingly, the alphasatellite are more closely related to nanovirus genomic components that also encode a Rep protein. Nanoviruses (family *Nanoviridae*) are a group of plant-infecting, ssDNA viruses that are transmitted by aphids. Similarly, the master Rep (mRep) component which encodes the Rep required for replication of all nanovirid components. These satellite-like molecules and mRep components are however quite distinct from each other based on their genomic sequence and genome organization. The position of the promoter of these components is different from that of the mRep. It is believed that alphasatellites associated with begomoviruses have evolved directly from the satellite-like components of nanoviruses.

The major difference between alphasatellites and satellite-like components of nanovirus is their genomic size. The average size of alphasatellites is ~1380 bp, which is almost half the size of helper begomoviruses, whereas the genome of nanovirid components is typically of 1000–1100 bp. Most of this increased size is occupied by an adenine (A) rich region of about 200 bp. The function of this A-rich region it still not clear; it was observed that alphasatellites with deleted A-rich region were capable to replicate and move in plants in the presence of a helper begomovirus. There is an understated selection mechanism to maintain

the size of alphasatellites, most probably due to efficient encapsidation by the coat protein of their helper virus. Since the first identification of Cotton leaf curl Multan alphasatellite (CLCMulA) associated with the cotton begomovirus, multiple alphasatellites have been reported from CLCuD infected cotton in the Indian subcontinent. They include *Cotton leaf curl Multan alphasatellite* (CLCMulA), *Gossypium darwinii* symptomless alphasatellite (GDarSLA), *Ageratum conyzoides* symptomless alphasatellite (AConSLA), *Cotton leaf curl Burewala alphasatellite* (CLCBurA), *Okra enation leaf curl alphasatellite* (OELCA), *Ageratum yellow vein India alphasatellite* (AYVINA) and *Tomato leaf curl alphasatellite* (ToLCA). Interestingly, in North-East Africa, the only alphasatellite component reported with the CLCuD complex is *Cotton leaf curl Gezira alphasatellite* (CLCGezA).

CLCuD-Associated Betasatellite

Betasatellites are ssDNA satellites (genus *Betasatellite*, family *Tolecusatellitidae*) which are associated with many monopartite begomoviruses. Presumably, betasatellites are trans-encapsidated by the coat protein of helper viruses to allow movement in plants, as well as transmission by whiteflies. Betasatellites also depend on the helper geminivirus for their replication. In infectivity assays betasatellites were found responsible for the high titer of the helper virus, to levels found in their original host in the field. This showed that betasatellites are involved in either the replication of their helper virus, their systemic movement in the host, and/or suppressing the gene silencing host defense mechanism against viruses. The size of the betasatellites is about half (1350 bp) of the size of their helper begomovirus. Interestingly, they share no sequence homology with their helper virus except the stem-loop structure which contains the conserved nonanucleotide (TAATATTAC). This stem-loop structure is present in a satellite conserved region (SCR) which is about 100 nt and highly conserved among betasatellites. Betasatellites encode a single gene, the product of which is known as β C1 protein (~118 aa) which has a critical role in pathogenicity and symptoms determination as it alters several cellular processes of the host like ubiquitination, autophagy and suppression of gene silencing (post-transcriptional gene silencing (PTGS) or transcriptional gene silencing (TGS)). After the first discovery of the *Ageratum yellow vein betasatellite* (AYVB), they have been found from a diverse range of hosts. More than 1800 complete betasatellite genomes were sequenced and their comparison shows that they are highly diverse. Surprisingly, only a single species of betasatellite, *Cotton leaf curl Multan betasatellite* (CLCMulB) is associated with CLCuD. Both experimentally and in the field, this species was found to be trans-replicated by several distinct begomoviruses. When β C1 of CLCMulB is expressed in tobacco plants, either stably transformed or transient expression, it induces virus-like symptoms, which demonstrates that in the CLCuD complex, the β C1 is the most important pathogenicity and symptom determinant.

The second epidemic of CLCuD, which was caused by a recombinant begomovirus (from two different species), was also associated with a recombinant CLCMulB, in which CLCMulB^{Mul} (Multan strain) recombined with Tomato leaf curl betasatellite (ToLCB). In Sindh province of Pakistan, another novel strain of CLCMulB viz. CLCMulB^{Sha} (Shadadpur strain) was also found to be associated with CLCuD whereas in Punjab only CLCMulB^{Bur} was found. Most recently, a novel strain CLCMulB^{Veh} (Vehari strain) was identified from Vehari, Punjab, in Pakistan and from India. Identification of different strains of CLCMulB indicates the co-evolution of CLCMulB with helper begomoviruses. In Africa, Cotton leaf curl Gezira betasatellite (CLCGezB) is the only betasatellite found so far in association with CLCuD in Sudan and it is quite distinct from CLCMulB in the Indian subcontinent.

Control of CLCuD

Strategies to Control the CLCuD Complex

The CLCuD is a serious threat to cotton cultivation in the countries where whiteflies are a major pest. There are mainly two types of strategies i.e. a short term strategy, involving management practices and a long term strategy involving the development of resistant cotton varieties either by conventional or non-conventional methods.

Short Term Strategies

After the spread of CLCuD on a massive scale during the first epidemic, many short-term strategies were adopted to minimize the losses by reducing the whitefly population in the field. For instance, treatment of seeds to avoid early establishment of vector populations, spraying insecticide on cotton crop to control whitefly populations, eradication of weeds (alternative hosts of viruses), providing balanced dose of fertilizers and biological agents....

Application of selective insecticide is most important in controlling the whitefly population. Recently, the whole genome sequence of Asia II 1 showed genetic variance in insecticide resistance related genes as compared to MEAM1. Most probably, this may be the impact of misuse of insecticides to control whitefly population. Moreover, it is advisable to spray insecticide on cotton crop 70–90 DAS (days after sowing) to control the whitefly population as seed treatment protects from whitefly infestation up to 75 DAS in some reports. After this period, cotton plants are grown enough that they can escape the disease, and losses are reduced.

As CLCuD associated viruses are not seed-borne they depend upon the alternative hosts to survive the off season. Alternative hosts are many other plants like okra, tomato, tobacco, *ageratum*, *hibiscus* and many weeds. Avoiding the cultivation of alternative host in off season and proper eradications of weeds may reduce the disease incidence.

Further damage due to this disease can be reduced by applying a balanced dose of nutrients especially using potassium to boost the resistance against the disease whereas excessive use of nitrogen reduces the resistance. So, it is advised to use the balanced ratio of potassium and nitrogen which helps to reduce the disease severity.

Long Term Strategies

Genetics of host plant resistance to CLCuD

For durable resistance strategy to control the disease, it is most important to breed virus resistant varieties with enough genetic diversity. Before developing disease resistant varieties of cotton, the key components is designing a breeding plan for the genetic basis of disease resistance and its inheritance. Unfortunately, very little is known about the molecular basis of the disease resistance in cotton. *G. arboreum*, native species of Asia is naturally immune to the Asian CLCuD complex, but it is grown on a small scale because it produces a poor quality fiber. The transcriptomics study of *G. arboreum* under virus stress revealed multiple resistance mechanisms against CLCuD. The exotic species *G. hirsutum* produces good quality and quantity of fiber and is commercially cultivated on large scale. Unfortunately, this species showed limited resistance/tolerance to CLCuD and only two varieties, LRA5166 and CP15/2, have been used to develop resistance in the late 1990s. Losses due to CLCuD were successfully alleviated by cultivating these varieties and cotton production was regained to the level prior to the epidemic. Unfortunately, this resistance was broken by the evolution of novel strain of begomovirus viz., CLCKokV-Bur. Currently, all commercial varieties are highly susceptible to CLCuD. Recently, transcriptomic studies in *G. hirsutum* under whitefly transmitted virus stress provided some insights into disease susceptibility in cotton and predicted some important genes to be involved in the compatible interaction of CLCuD associated disease complex and cotton. These findings will have to be validated to pinpoint the genes that make *G. hirsutum* susceptible and *G. arboreum* resistant. Although some efforts were made in introgression virus resistance from *G. arboreum* into *G. hirsutum*, they were not successful in developing a resistant variety suitable for agriculture. It was also tried to transfer the desired traits from *G. hirsutum* to *G. arboreum* (after doubling its chromosomes) but unfortunately, it also remained unsuccessful in the development of commercial varieties. Finally, about 5000 accessions of *G. hirsutum* from USA and some introgression lines were screened in Pakistan against CLCuD. This work identified the “Mac-07” accession, and 95 other accessions, that remained asymptomatic under field conditions, hoping that a good source of resistance could be used in the near future.

Engineered Resistance to the CLCuD Complex

Because of the difficulty to identify and transfer useful genes of resistance to CLCuD from *G. hirsutum* germplasm, multiples strategies have been deployed using genetic engineering approach to develop cotton plants resistant to this viral disease and its insect vector whitefly. Generally, these strategies are of two types; firstly the “pathogen derived resistance” in which full length viral genes or different small conserved portion of the virus genome are used and secondly the ‘non-pathogen derived resistance’ in which genes from distantly related genetic source are used.

Pathogen Derived Resistance (PDR)

Introducing the genomic part (gene or its part) of virus that is conserved among different viruses into the host plant is the most desirable strategy for broad spectrum resistance to control disease. Following strategies have been adopted for the development of resistance in which genomic information of pathogen is used.

Antisense RNA technology

In this technology antisense RNA molecule suppresses the complementary target mRNA. When antisense RNA against Rep gene of the virus was introduced in cotton, it suppressed the replication of the virus, similarly in another study when the coat protein gene (AV1) was targeted by antisense RNA in transgenic cotton, it arrested the movement, replication and encapsidation of the target virus. Transgenic cotton was also developed using antisense AV2 (350vbp) gene and transgenic plants were resistant to CLCuD. Similarly, two truncated forms of the Rep gene were introduced in *G. hirsutum*, which inhibited the replication of both viral and satellite genomes. Recently antisense β C1 gene was also introduced in cotton and transgenic plants were symptomless under virus stress. However, none of these transgenic cotton lines were successful under field conditions.

RNAi technology

The principle of the RNAi technology utilizes the transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS) and has been deployed to develop resistance against different viruses, such as the African cassava mosaic virus (ACMV), the Mung bean yellow mosaic virus (MYMV) and many others. A 21 nt of V2 of CLCKokV-Bur strain was used to develop the artificial microRNA (amiRNA) and construct was transformed in *Nicotiana benthamiana* plants and transgenic plants showed resistance to CLCKokV-Bur. Recently, RNAi based construct against the V2 gene of CLCKokV-BuR was transformed in two cotton species viz. MNH-786 and VH-289 and transgenic cotton plants, when challenged to whitefly under control conditions, showed a low titer of

the virus for transgenic plants as compared to non-transgenic. It was concluded that these transgenic plants could be used in the fields to control the CLCuD but have so far not proven successful in the field.

Non-Pathogen Derived Resistance

In this strategy, genes from host or non-host are used to develop the resistance against the disease. For instance, genes which confer DNA binding proteins, antiviral antibodies, coat binding proteins etc. have been introduced in plants to develop resistance to CLCuD. Recently, Tma12 a protein identified from fern to confer resistance to whitefly, was transformed in cotton, and one of the transgenic lines showed increased resistance (> 99%) to whitefly. Moreover, this protein was non-toxic in rats which shows that this is safe to use in other crops.

Genome Editing Strategy: CRISPR/Cas9 Approach to Control CLCuD

CRISPR/Cas9 system has been exploited in many organisms for genome editing by transforming the Cas9 protein. One of the desirable properties of this system is that it can targets multiple sites, which can be utilized to control multiple viruses in plants. Resistance to different geminiviruses has already been achieved through this system. For example, genome of *Bean yellow dwarf virus* (BeYDV) was cut down when CRISPR–Cas9 system was applied in bean against this virus and resulted in reduced disease symptoms. Similarly, this system was used against TYLCV and *Beet curly top virus* (BCTV) and resulted in the reduced disease symptoms. In cotton this system is in progress to target CLCKokV-Bur and associated betasatellite genes and conserved regions. However, its success is yet to be demonstrated. Recently, it was reported that expressing Cas9 in cassava was unable to protect it from geminiviruses infections where viruses multiply faster than edited by Cas9 and this also helps in rapid evolution of viruses. This makes an alarming sign for future results of CRISPR/Cas9 use in cotton against CLCuD associated disease complex (as multiple sgRNAs are being used). In this scenario it is most important to explore the actual molecular mechanisms involved in compatible interactions of virus-host vector and identify some susceptible genes in cotton which can be knocked out by CRISPR/Cas9 system, or resistance genes for introgression in susceptible commercial varieties.

See also: Alphasatellites (*Alphasatellitidae*). Bean Golden Mosaic Virus and Bean Golden Yellow Mosaic Virus (*Geminiviridae*). Beet Curly Top Virus (*Geminiviridae*). Betasatellites and Deltasatellites (*Tolecusatellitidae*). Cassava Mosaic Viruses (*Geminiviridae*). Emerging Geminiviruses (*Geminiviridae*). Geminiviruses (*Geminiviridae*). Maize Streak Virus (*Geminiviridae*). Tomato Leaf Curl New Delhi Virus (*Geminiviridae*). Tomato Yellow Leaf Curl Viruses (*Geminiviridae*)

Further Reading

- Amin, I., Mansoor, S., Amrao, L., *et al.*, 2006. Mobilisation into cotton and spread of a recombinant cotton leaf curl disease satellite. *Archives of Virology* 151, 2055–2065.
- Asad, S., Haris, W.A.A., Bashir, A., *et al.*, 2003. Transgenic tobacco expressing geminiviral RNAs are resistant to the serious viral pathogen causing cotton leaf curl disease. *Archives of Virology* 148, 2341–2352.
- Briddon, R.W., 2003. Cotton leaf curl disease, a multicomponent begomovirus complex. *Molecular Plant Pathology* 4, 427–434.
- Briddon, R.W., Markham, P.G., 2000. Cotton leaf curl virus disease. *Virus Research* 71, 151–159.
- Briddon, R.W., Bull, S.E., Amin, I., *et al.*, 2004. Diversity of DNA-1: A satellite-like molecule associated with monopartite begomovirus-DNA β complexes. *Virology* 324, 462–474.
- Datta, S., Budhaliya, R., Das, B., *et al.*, 2017. Rebound of cotton leaf curl Multan virus and its exclusive detection in cotton leaf curl disease outbreak, Punjab (India), 2015. *Scientific Reports* 7 (1), 17361.
- Fauquet, C.M., Stanley, J., 2005. Revising the way we conceive and name viruses below the species level: A review of geminivirus taxonomy calls for new standardized isolate descriptors. *Archives of Virology* 150, 2151–2179.
- Leke, W.N., Khatabi, B., Mignouna, D.B., Brown, J.K., Fondong, V.N., 2016. Complete genome sequence of a new bipartite begomovirus infecting cotton in the Republic of Benin in West Africa. *Archives of virology* 161 (8), 2329–2333.
- Sattar, M.N., Iqbal, Z., Tahir, M.N., Ullah, S., 2017. The prediction of a new CLCuD epidemic in the old world. *Frontiers in microbiology* 8, 631.
- Sattar, M.N., Kvarnheden, A., Saeed, M., Briddon, R.W., 2013. Cotton leaf curl disease – An emerging threat to cotton production worldwide. *Journal of General Virology* 94 (4), 695–710.
- Sohrab, S.S., Kamal, M.A., Ilah, A., *et al.*, 2016. Development of cotton leaf curl virus resistant transgenic cotton using antisense BC1 gene. *Saudi Journal of Biological Sciences* 23 (3), 358–362.
- Zaidi, S.S.E.A., Shafiq, M., Amin, I., *et al.*, 2016. Frequent occurrence of tomato leaf curl New Delhi virus in cotton leaf curl disease affected cotton in Pakistan. *PLoS One* 11 (5), e0155520.
- Zubair, M., Zaidi, S.S.E.A., Shakir, S., *et al.*, 2017. Multiple begomoviruses found associated with cotton leaf curl disease in Pakistan in early 1990 are back in cultivated cotton. *Scientific Reports* 7 (1), 680.
- Zubair, M., Zaidi, S., Shakir, S., Amin, I., Mansoor, S., 2017. An insight into cotton leaf curl Multan betasatellite, the most important component of cotton leaf curl disease complex. *Viruses* 9 (10), 280.

Cowpea Mosaic Virus (*Secoviridae*)

George P Lomonosoff, John Innes Centre, Norwich, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)

Co-Pro Protease cofactor

CP Coat protein or capsid protein

ER Endoplasmic reticulum

kb Kilobase

kDa Kilo Dalton

MP Movement protein

nt Nucleotide(s)

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

UTR Untranslated region

VLPs Virus-like particles

VPg Viral protein genome-linked

Glossary

Affimers Small proteins that mimic antibodies.

Nanobiotechnology Refers to the intersection of nanotechnology and biology. Given that the subject is one that has only emerged very recently, bionanotechnology and nanobiotechnology serve as blanket terms for various related technologies.

Nanotechnology Also known as “nanotech” is manipulation of matter on an atomic, molecular, and supramolecular scale.

Viral protein genome-linked A protein covalently attached to the 5′ end of positive strand viral RNA and acting as a primer during RNA synthesis.

Virus replication complexes Intracellular complexes formed by association with intracellular membranes that are sites for stages in the replication cycle of a virus.

β-barrel Beta-sheet composed of tandem repeats forming a closed structure.

Introduction

Since its first isolation in 1959, Cowpea mosaic virus (CPMV) has served as a model icosahedral plant virus due to the high yield, ease of isolation and stability of its particles. As a consequence, it has been at the forefront of both genetic and structural studies of plant viruses and has been widely deployed in bionanotechnology. Thus, CPMV can be considered alongside viruses such as Tobacco mosaic virus (TMV) and Brome mosaic virus (BMV) as one of the “classic” plant viruses.

Taxonomy and Classification

CPMV is the type species of the genus *Comovirus* (which includes 14 additional members). The genus *Comovirus* is classified, together with the genera *Fabavirus* and *Nepovirus*, within the sub-family *Comovirinae* of the family *Secoviridae*. Within the subfamily the greatest affinity is between the genera *Comovirus* and *Fabavirus*, both of which are insect-, as opposed to nematode-, transmitted. On a wider scale, consideration of genome structure and organisation, translational strategy and aa homologies between the virus-encoded proteins has led to the family *Secoviridae* being one of the 6 families within the order *Picornavirales*. Historically, the family *Secoviridae* had been grouped within the same “superfamily” as the family *Potyviridae* but the two families are now separately classified.

Biology

CPMV was first isolated from infected cowpeas (*Vigna unguiculata*) in Nigeria in 1959. Subsequently it has been found to occur in Nigeria, Kenya, Tanzania, Japan, Surinam and Cuba. It causes typical mosaic symptoms on infected plants (Fig. 1). While its natural host is cowpea, it can infect other legumes and *Nicotiana benthamiana* has proven to be an extremely valuable experimental host. In nature, transmission is usually by leaf-feeding beetles, especially by members of the Chrysomelidae. However, CPMV has been shown also to be transmitted by thrips and grasshoppers. The beetle vectors can acquire the virus by feeding for as little as one minute and can retain and transmit the virus for a period of days or weeks. The virus does not, however, multiply in the insect vector. Experimentally, CPMV is mechanically transmissible.

In Nigeria, infection of cowpeas with CPMV causes a great reduction in leaf area, flower production and yield. Infected plant cells show a number of characteristic cytological changes. These include the appearance of viral particles, a proliferation of cell membranes and vesicles in the cytoplasm, and a variety of modifications to plasmodesmata.

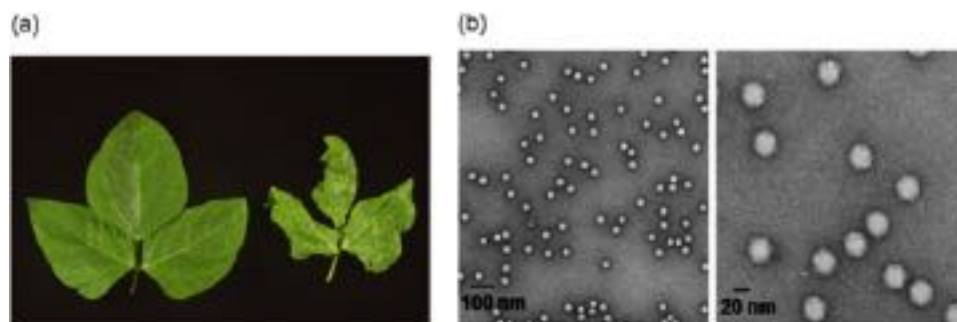


Fig. 1 CPMV symptoms and particles (A) Mosaic symptoms on the trifoliate leaves of a CPMV-infected *Vigna unguiculata* plant (right) compared to an uninfected plant (left). (B) Transmission electron micrographs of purified particles of CPMV at two different magnifications. The particles were negatively stained with 2% (w/v) uranyl acetate.

Physical Properties of Viral Particles

CPMV particles are icosahedral, non-enveloped with an average diameter of 28 nm (Fig. 1). Up to 2 g of particles can be obtained per kg of fresh infected cowpea tissue and they can be readily purified by polyethylene glycol precipitation and differential centrifugation. The particles are very stable with a thermal inactivation point in plant sap of 65–75°C and a longevity in sap of 3–5 days at room temperature. Once purified, the particles can be stored for prolonged periods at 4°C. The ease with which virus particles can be propagated, purified and stored has undoubtedly contributed to the early popularity of CPMV as an object of study.

Preparations of CPMV particles can be separated into three components on density gradients that are designated top (T), middle (M) and bottom (B), with sedimentation coefficients of 58S, 95S and 115S, respectively. The three components have identical protein compositions, containing 60 copies each of a large (L) and small (S) coat protein (CP), with sizes 42 kDa and 24 kDa, respectively, as calculated from the nucleotide sequence. The discovery, in 1971, that CPMV particles contained equimolar amounts of two different polypeptides suggested that the capsids had an architecture more similar to the animal picornaviruses than to other plant viruses which has been previously examined. This provided an early clue as to the common origins of plant and animal viruses.

The difference in the sedimentation behaviour of the three centrifugal components of CPMV lies in their RNA contents. Top components are devoid of RNA, while middle and bottom components each contain single molecules of positive-strand RNA of 3.5 and 6.0 kb respectively. The two RNA molecules were originally termed middle (M) and bottom (B) component RNA after the component from which they were isolated. However, more recently they have been referred to as RNA2 and RNA1, respectively. The three-component nature of CPMV preparations is summarized in Fig. 2. The determination of the component structure of the virus, and particularly the relationship between this and infectivity, was an important step in establishing the principle that plant viruses frequently have divided genomes, the individual components of which are separately encapsidated.

CPMV preparations are not only centrifugally heterogeneous but can also be separated in two forms, fast and slow, electrophoretically. Both electrophoretic forms contain all three centrifugal components. The proportion of the two electrophoretic forms in a given virus preparation varies both with the time after infection at which the virus was isolated and the age of the preparation itself. Conversion of one form to the other is caused by loss, through proteolysis, of 24 aa from the C terminus of the S protein.

Viral Structure

X-ray crystallographic and cryo-electron microscopy (cryo-EM) studies on CPMV, using either a natural mixture or individual components, have provided a detailed picture of the arrangement of the two viral CPs in the three-dimensional structure of the particle. Overall, the virions have icosahedral symmetry, with 12 axes of fivefold and 20 axes of threefold symmetry and resemble a classic $T=3$ particle. However, since there are two different types of CP subunit (L and S), their arrangement is usually referred to as pseudo $T=3$ or $P=3$. The two CPs taken together consist of three distinct β -barrel domains, two being derived from the L and one from the S protein. Thus, in common with the $T=3$ viruses, each CPMV particle is made up of 180 β -barrel structures. The S protein, with its single domain, is found at the fivefold symmetry axes and therefore occupies a position analogous to that of the A type subunits in $T=3$ particles. The N- and C-terminal domains of the L protein occur at the threefold axes and occupy the positions equivalent to those of the C and B type subunits of a $T=3$ particle respectively (Fig. 2). This detailed analysis confirmed the earlier suggestion that CPMV particles are structurally homologous to those of picornaviruses, with the N- and C-terminal domains of the L protein being equivalent to viral protein VP2 and viral protein VP3, respectively, and the S protein being equivalent to viral protein VP1; there is no equivalent of VP4. While no RNA could be seen in the crystallographic structures of either the M or B components of CPMV, cryo-EM analysis revealed that the encapsidated RNA forms concentric shells, with the strongest density being found directly beneath the twofold symmetry axes of the capsid. In common with other members of the order *Picornavirales*, the specific incorporation of the genomic RNAs into particles appears to be based on their ability to be replicated.

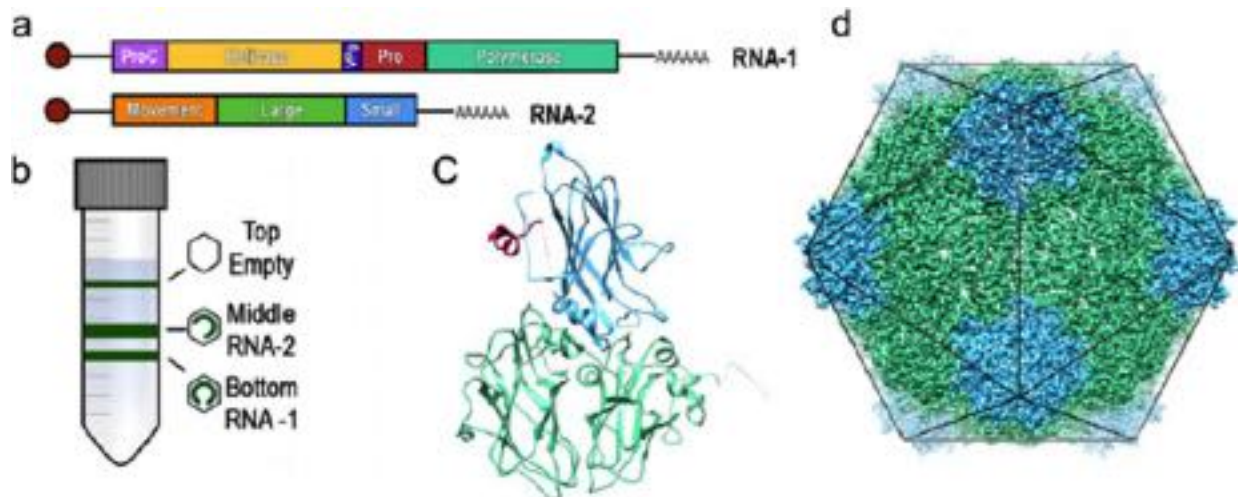


Fig. 2 Molecular structure of CPMV (a) Organisation of the two single-stranded RNA molecules that comprise the genome. Both RNAs have a VPg (red circle) linked to their 5' ends and both are polyadenylated. Excluding the poly(A) tail, RNA1 is 5889 nt in length and contains a single ORF that is processed by the 24K protease (Pro) to give the viral proteins required for RNA replication. RNA2 is 3481 nt in length and encodes the structural CPs, Large (L) and Small (S), as well as the protein required for the movement of CPMV virions from cell-to-cell. (b) Gradient separation of a natural mixture permits separation into three components. Empty CPMV particles sediment at the top (CPMV-T), RNA2-containing particles sediment in the middle (CPMV-M) and RNA1-containing particles sediment at the bottom of the gradient (CPMV-B). (c) An asymmetric unit of synthetic CPMV empty virus-like particles (eVLP) as revealed by cryo-electron microscopy. Both domains of the CP-L subunit are shown in green, while single domain of the CP-S is shown in blue. The C-terminal extension of the CP-S, visible in eVLPs, but not in particles from infectious virus particles, is coloured pink. (d) The icosahedral organisation of CPMV. Each of the 60 asymmetric units comprises one copy of the L subunit and the S subunit (coloured as in c). A view down the two-fold axis is shown. Reproduced from Hesketh, E.L., Meshcheriakova, Y., Dent, K.C., *et al.* 2015. Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. *Nature Communications* 6, 10113.

As well as the natural components, the structure of synthetic empty virus-like particles (eVLPs), that lack any detectable RNA, has been determined by both crystallography and cryo-EM. These particles were shown to be essentially identical to the natural components; however, it was possible to visualize some of the 24 aa that are normally cleaved from the C-terminus of the S protein. These structural studies, together with mutagenesis of the 24 aa segment, has enabled a model for the assembly of CPMV particles to be generated in which the 24 aa promote pentamer formation followed by the formation of RNA binding sites at the two-fold symmetry axes (Fig. 3). The presence of the C-terminal portion of the 24 aa segment on the particle surface is also important in permitting efficient systemic movement of the virus.

Genome Structure

Both M and B (but not T) components of a virus preparation are essential for infection of whole plants. As CPMV is a positive-strand RNA virus, a mixture of the genomic RNAs within the particles can also be used to initiate an infection. However, RNA1 is capable of independent replication in individual plant cells but this leads to the establishment of gene silencing, rather than a productive infection, in the absence of RNA2. Both genomic RNAs have a small basic protein (VPg) covalently linked to their 5' termini and both are polyadenylated at their 3' ends. The elucidation of the overall structure of the RNA segments once more underscored the similarity between CPMV and picornaviruses. However, unlike picornaviruses, the VPg is linked to the viral RNA via the β -hydroxyl group of its N-terminal serine residue rather than via a tyrosine. The VPg is not required for the viral RNAs to be infectious.

The complete nucleotide sequences of both genomic RNAs were reported in 1983, making CPMV one of the first RNA plant viruses to be completely sequenced. The length of the RNAs are 5889 and 3481 nt, for RNA1 and RNA2, respectively, excluding the poly(A) tails. The two genomic RNAs have no sequence homology apart from at the 5' and 3' termini. Full-length infectious cDNA clones of both RNAs of CPMV have been constructed, allowing the genome to be manipulated.

Expression of the Viral Genome

Both genomic RNAs contain a single long open reading frame (ORF) which occupies over 80% of the length of the RNA. A combination of *in vitro* translation and protoplast studies has unravelled the basic mechanism of gene expression of the virus. Both RNAs are expressed through the synthesis and subsequent cleavage of large precursor polypeptides (Fig. 2). On RNA1, initiation of

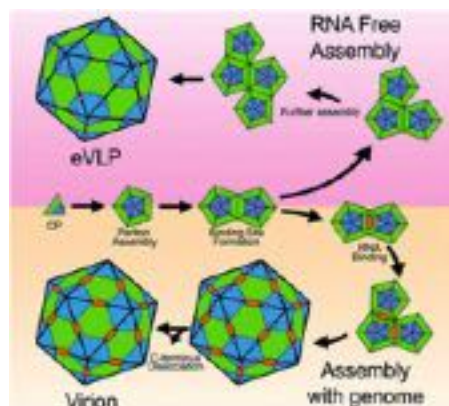


Fig. 3 A model for CPMV Assembly. The L (green) and S (blue) subunits associate to form a CP penton. The C-terminal extension of the S protein (pink circle) appears to act as 'molecular glue', stabilizing the formation of the penton due to extensive interactions with the neighbouring S subunit. In the absence of RNA (RNA-free assembly shown in the top half of the diagram), C-terminal cleavage is slow, and eVLPs are produced. Once an extensive protein: protein interaction network is formed, the eVLP structure is stable and no longer depends on the presence of the C-terminal extension. In the presence of genomic RNA, assembly follows a different path (bottom half of the diagram). Initially, pentons are formed in the same way as in the eVLP, using the C-terminal extension for stability. Two pentons interact to form a twofold axis, and thus the RNA-binding site appears to exist at that position. The model shows genomic RNA (represented as an orange rectangle) initially binding to the two pentons at this twofold axis, the first step in genome encapsidation, followed by the stepwise addition of pentons, to form RNA-containing capsids. Reproduced from Hesketh, E.L., Meshcheriakova, Y., Dent, K.C., *et al.* 2015. Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. *Nature Communications* 6, 10113.

translation occurs at the first AUG encountered on the sequence (at position 207) and results in the synthesis of a protein of approximately 200 kDa (the 200K protein). This initial product undergoes rapid co-translational autoproteolysis to give proteins with apparent sizes of 32 kDa (Protease co-factor; ProC) and 170 kDa (170K) proteins. The 170K protein undergoes further cleavages to give the range of virus-specific proteins shown in Fig. 2, specifically the 58K Helicase, VPg, the 24K protease and the 87K polymerase.

Initiation of translation of RNA2 occurs at two different positions on the RNA and results in the synthesis of two carboxy coterminal proteins, the 105K and 95K proteins. This double initiation phenomenon occurs through "leaky scanning". Synthesis of the 105K protein is initiated from an AUG at position 161, while initiation from an in-frame AUG at position 512 directs the synthesis of the 95K protein; RNA-2 also has an additional AUG (position 115) upstream of both these initiation sites but in a different reading frame. Both RNA2-encoded primary translation products are cleaved by the RNA1-encoded 24K protease to give either the 58K or the 48K protein (depending on whether it is the 105K or 95 K protein being processed) and the precursor two viral CPs, VP60, which is then subsequently processed to give the mature L and S CPs. Processing of the RNA2-encoded polypeptides, at least at the site between the 48K and L CP, has been shown to require the presence of the 32K ProC protein as well as the 24K protease.

Functions of the Viral Proteins

Functions have been ascribed to most of the regions of the polyproteins encoded by both RNA1 and RNA2 of CPMV. In most cases, however, it is not certain at what stage(s) in the cleavage pathway they manifest their activity.

In the case of RNA1, the 32K ProC protein, which is rapidly cleaved from the N-terminus of the 200K primary translation product, is a cofactor which modulates the activity of the virus-encoded 24K protease. As described above, the presence of ProC is required for the cleavage of the RNA2-encoded 105K and 95K proteins but is not essential for the cleavage of the RNA1-encoded 170K protein. It does, however, seem to play a role in determining the rate at which cleavage of the 170K protein occurs. When mutant RNA1 molecules carrying deletions in the region encoding ProC are translated *in vitro*, the rate of processing of the 170K protein is greatly increased, indicating that the 32K protein acts as an inhibitor of processing. This inhibition may be achieved through the interaction of the 32K with the 58K Helicase domain of the 170K protein. The mechanism by which the 32K protein enables the 24K proteinase to cleave *in trans* is unclear. It could be due to a direct association between the 32K protein and the 24K proteinase (or an intermediate containing the proteinase domain) or it could be an indirect result of the 32K protein modulating the cleavage of the 170K protein, thereby allowing a 24K-containing intermediate with *in trans* processing activity to accumulate.

The RNA1-encoded 58K Helicase protein is associated with cell membranes and contains a nucleotide-binding motif. A protein containing the aa sequence of the Helicase protein linked to VPg, is involved in rearrangements in the endoplasmic reticulum of CPMV-infected cells and acts in concert with ProC. The 24K protease carries out all the cleavages on both the RNA1- and RNA2-encoded polyproteins. Its proteolytic activity has been shown to be expressed in processing intermediates that contain its sequence. Indeed, it is not known whether the free form of the protein has any biological significance. Although the proteinase contains a

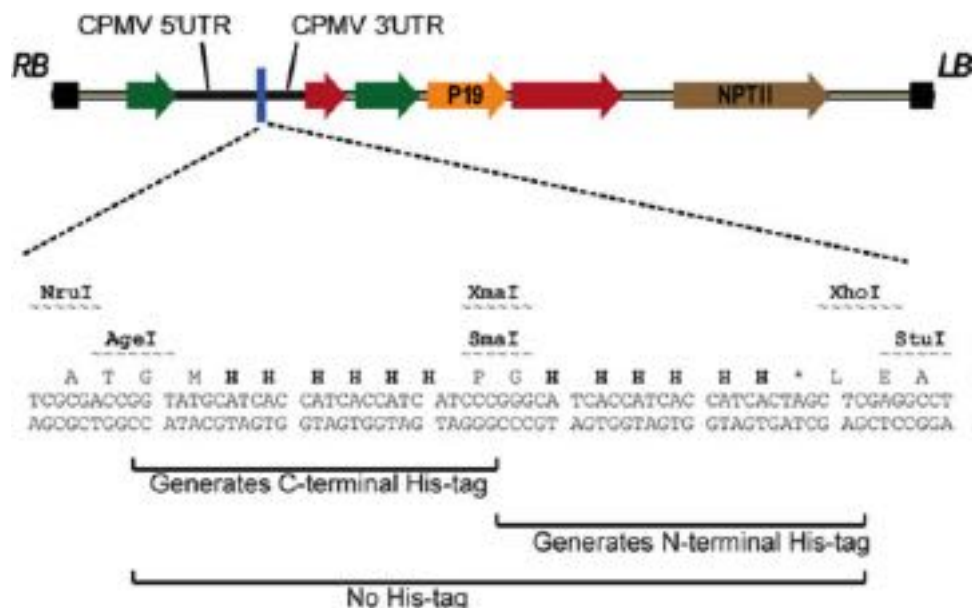


Fig. 4 Diagrammatic representation of the T-DNA region of the transient expression vector, pEAQ-HT showing the polylinker in detail. This polylinker allows the expression of the target protein with or without His-tags. The black boxes represent the T-DNA borders, the green arrows represent promoter sequences and the red arrows represent terminator sequences. The locations of the CPMV RNA2-derived 5' and 3' UTRs, between which the target sequence is inserted, is indicated. Reproduced from Sainsbury, F., Thuenemann, E.C., Lomonossoff, G.P., 2009. pEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal* 7, 682–693.

cysteine at its active site, it is structurally related to serine proteases, such as trypsin, rather than cellular thiol proteases, such as papain, a feature similar to the 3C proteinases of members of the family *Picornaviridae*. All CPMV cleavage sites have glutamine (Q) residue at the –1 position.

The 87K Polymerase contains all the sequences associated with it being the virus-encoded RNA-dependent RNA polymerase (RdRp), including the GDD sequence motif found in all such enzymes. It also has aa sequence homology to the 3D polymerases encoded by the picornaviruses. Consistent with this identification, mutation of the GDD motif to GAD or AAA reduces or abolishes replication, respectively. When replication complexes capable of elongating nascent RNA chains were isolated from CPMV-infected cowpea plants they were found to contain the 110K protein (Fig. 3), consisting of the sequence of 87K polymerase protein linked to the 24K proteinase, rather than the free 87K protein, and this 110K protein is often referred to as the viral “replicase”.

In the case of RNA2, the 48K protein, derived from processing of the 95K protein, is involved in potentiating the spread of the virus from cell to cell. This protein is found in tubular structures that are formed in the plasmodesmata of infected cells. Tubules extending into the culture medium can also be seen in protoplasts either infected with CPMV or transiently expressing the 48K protein. Virus particles can be seen within these tubules when protoplasts are infected with CPMV but not when only 48K protein is expressed. At present, no definite role has been assigned to the 58K protein which is produced by processing of the 105K protein. Elimination of AUG161, preventing production of the 105K protein, results in a substantial reduction of RNA2 replication. The mature viral CPs, L and S, are produced by cleavage of VP60 by the 24K protease. As well as protecting the genomic RNAs, capsid formation is essential for the virus to be able to spread from cell to cell through modified plasmodesmata and long-distance movement also requires capsid formation. An additional function in suppressing gene silencing is also provided by the C-terminal region of the S protein.

Replication

CPMV replicates to high level in infected cells. Replication is believed to involve the initial transcription of the incoming positive-sense RNA into minus-strands followed by initiation and synthesis of new plus-strands from the recently formed minus-strands. It has been shown that the 5' ends of both the plus- and minus-strands are covalently linked to the VPg, suggesting that this protein has an essential role in the initiation of RNA synthesis. There also appears to be a tight linkage between the translation of the viral RNAs and their replication and between replication and the encapsidation of the viral RNAs.

Replication of the viral RNAs has been shown to occur in the membranous cytopathological structures which are formed in the cytoplasm of cells during infection through the action of the RNA1-encoded 32K and 60K proteins. Both CPMV-specific double-stranded replicative form (RF) RNA and an enzyme activity capable of completing nascent RNA strands can be isolated

from such structures. Purified preparations of the enzyme activity contain the RNA1-encoded 110K replicase and two host-encoded proteins of 68 and 57 kDa. However, at present no enzymatic activity capable of initiating RNA synthesis *in vitro* has been described.

Use in Biotechnology

CPMV particles have been extensively used in bio and nanotechnology. In the first instance, antigenic peptides (epitopes) were genetically fused to exposed loops on the surface of the viral capsids using sites within both the L and S CPs. The resulting chimeric virus particles could be propagated in plants and the modified virions purified, provided the inserted peptide was not too long or excessively positively-charged. When injected into experimental animals, the modified particles could elicit the production of antibodies against the inserted epitope and in several instances were shown to confer protective immunity against the pathogen from which the epitope was derived. In addition to genetic modification, it is possible to chemically attach compounds to the surface of the particles via addressable aa such as lysine, glutamic and aspartic acid. Chemical attachment has the advantage that it allows moieties other than aa to be displayed on the surface of particles. It has also proved possible to combine both genetic and chemical methods to enable the mineralisation of virus particles. The development of a system for the production of eVLPs has also enabled CPMV particles to be loaded with foreign materials, such as drugs and heavy metals. The eVLPs have also proven useful for the selection of affimers which can subsequently be used to diagnose CPMV infections.

CPMV RNA2 has been extensively used as a gene vector for the expression of heterologous proteins in plants via transient expression. Initially, such applications relied on the ability of the modified RNA to be replicated by RNA1 to achieve high level expression. However, the deployment of suppressors of gene silencing and the removal of inhibitory AUG codons, coupled with efficient agro-infection technology, has led to the development of a series of RNA2-based expression vectors (the pEAQ series) that do not rely on replication. In this system, the gene to be expressed is positioned between a modified version of the RNA2 5'UTR, lacking the AUG codons upstream of AUG512, and the RNA2 3'UTR to create a synthetic mRNA cassette (Fig. 4). Transcription of the cassette is driven by the Cauliflower mosaic virus 35S promoter and terminated by the nos terminator. The whole transcription unit is placed between the left and right T-DNA borders within a binary plasmid together with a transcription unit for the P19 suppressor of gene silencing. Agro-infiltration is used to deliver the plasmids to plant tissue and expression is usually detected within one week.

The pEAQ expression system has been used for the expression of a wide variety of different proteins, including those capable of forming VLPs and enzymes for bio-transformations. The elimination of the need for replication means that it is possible to co-express multiple proteins within the same cell, thus allowing the manipulation of biosynthetic pathways.

See also: Cheraviruses, Sadwavirus and Torradoviruses (*Secoviridae*). Nepoviruses (*Secoviridae*). Secoviruses (*Secoviridae*). Sequiviruses and Waikaviruses (*Secoviridae*)

Further Reading

- Czapar, A.E., Steinmetz, N.F., 2017. Plant viruses and bacteriophages for drug delivery in medicine and biotechnology. *Current Opinion in Chemical Biology* 38, 108–116.
- Gergerich, R.C., Scott, H.A., 1996. Comoviruses: Transmission, epidemiology, and control. In: Harrison, B.D., Murrant, A.F. (Eds.), *The Plant Viruses 5: Polyhedral Virions and Bipartite RNA Genomes*. New York: Plenum Press, p. 77.
- Goldbach, R.W., Wellink, J., 1996. Comoviruses: Molecular biology and replication. In: Harrison, B.D., Murrant, A.F. (Eds.), *The Plant Viruses 5: Polyhedral Virions and Bipartite RNA Genomes*. New York: Plenum Press, p. 35.
- Hesketh, E.L., Meshcheriakova, Y., Dent, K.C., *et al.*, 2015. Mechanisms of assembly and genome packaging in an RNA virus revealed by high-resolution cryo-EM. *Nature Communications* 6, 10113.
- Hesketh, E.L., Meshcheriakova, Y., Thompson, R.F., Lomonosoff, G.P., Ranson, N.A., 2017. The structures of a naturally empty Cowpea mosaic virus particle and its genome-containing counterpart by cryo-electron microscopy. *Science Reports* 7, 539.
- Kruse, I., Peyret, H., Saxena, P., Lomonosoff, G.P., 2019. Encapsidation of Viral RNA in Picornavirales: Studies on Cowpea mosaic virus demonstrate dependence on viral replication. *Journal of Virology* 93.(pii: e01520-18).
- Lin, T., Johnson, J.E., 2003. Structure of picorna-like plant viruses: Implications and applications. *Advances in Virus Research* 62, 167–239.
- Lomonosoff, G.P., Evans, D.J., 2011. Applications of plant viruses in bio-nanotechnology. In: Palmer, K., Gleba, Y. (Eds.), *Current Topics of Microbiology and Immunology* vol. 375, pp. 61–87.
- Marsian, J., Lomonosoff, G.P., 2016. Molecular pharming – VLPs made in plants. *Current Opinion in Biotechnology* 37, 201–206.
- Peyret, H., Lomonosoff, G.P., 2015. When plant virology met *Agrobacterium*: The rise of the deconstructed clones. *Plant Biotechnology Journal* 13, 1121–1135.
- Sainsbury, F., Cañizares, M.C., Lomonosoff, G.P., 2010. Cowpea mosaic virus: The plant virus-based biotechnology workhorse. *Annual Review of Phytopathology* 48, 437–455.
- Sainsbury, F., Thuenemann, E.C., Lomonosoff, G.P., 2009. pEAQ: Versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal* 7, 682–693.
- Steinmetz, N.F., 2019. Biological and evolutionary concepts for nanoscale engineering: Viruses as natural nanoparticles have great potential for a wide range of nanoscale products. *EMBO Reports* 20 (8), (e48806).

Relevant Websites

<http://www.dpvweb.net/dpv/showdpv.php?dpvno=378>

Cowpea mosaic virus.

https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/secoviridae/589/genus-comovirus

Genus: Comovirus.

<http://viperdbscripps.edu/genuslist.php?genus=Comovirus>

Virus Particle ExploreR.

Cucumber Mosaic Virus (*Bromoviridae*)

Judith Hirsch and Benoît Moury, Plant Pathology Unit, INRAE – French National Research Institute for Agriculture, Food and Environment, Montfavet, France

© 2021 Elsevier Ltd. All rights reserved.

This is an update of F. García-Arenal, P. Palukaitis, Cucumber Mosaic Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00640-3](https://doi.org/10.1016/B978-012374410-4.00640-3).

Nomenclature

aa Amino acid(s)

AGO Argonaute proteins

CP Coat protein or capsid protein

CRISPR/Cas9 Clustered regularly interspaced short palindromic repeats/CRISPR associated protein 9

ELISA Enzyme-linked immunosorbent assays

ER Endoplasmic reticulum

kb Kilobase

kDa Kilo dalton

miRNA MicroRNA

MP Movement protein

nt Nucleotide(s)

ORF Open reading frame

RISC RNA-induced silencing complex

RdRp RNA-dependent RNA polymerase

RT-PCR Reverse transcription-polymerase chain reaction

satRNA Satellite RNA

sgRNA Sub-genomic RNA

siRNA Small interfering RNA

TILLING Targeting induced local lesions in genomes

tRNA transfer RNA

UTR Untranslated region

Glossary

Cross-protection Inhibition of systemic virus accumulation and disease by prior inoculation of plants with a mild or symptomless strain of the same virus.

Filiformism Narrowing of the leaf blade, often leading to symptoms referred to as shoestring.

Non-circulative A mode of virus transmission in which the virus does not need to circulate through the vector digestive system, hemocoel or salivary glands. The virus is carried by attachment to the mouthparts or foregut of the vector.

Non-persistent A mode of virus transmission characterized by short acquisition and inoculation times (from seconds to minutes).

Pseudo-recombination Reassortment of the genomic segments of two or more strains of a virus to generate novel combinations of the full complement of genomic segments.

Reassortant A virus resulting from reassortment (or pseudo-recombination), *i.e.*, exchange of entire genome segments between two or more parental strains.

Satellite RNA A subviral RNA dependent on a helper virus for both replication and encapsidation.

Tonoplast The membrane surrounding the central vacuole of plant cells.

Virulence The impact of a parasite on its host's health or fitness.

Yellowing Light green symptoms induced by virus infection affecting chlorophyll accumulation.

Introduction

Cucumber mosaic virus (CMV) is the type species of the genus *Cucumovirus* in the family *Bromoviridae*, a group of single-stranded positive-sense RNA viruses. It was first described in 1916 as a disease of cucurbits in the USA. CMV has a worldwide distribution and a host range of over 1000 plant species. CMV is one of the most important agronomic pathogens on vegetable and ornamental plants as well as some important fruit crops. The virus is transmitted horizontally by about 80 aphid species in a non-persistent and non-circulative manner. In some plants, CMV is also transmitted vertically through the seed with variable, but usually low, efficiencies. CMV has a segmented genome and is multipartite, meaning that the three main genomic RNAs are encapsidated in separate particles (virions). CMV virions are isometric particles of 29 nm in diameter ($T=3$ symmetry). The CMV genome encodes five proteins. The 1a and 2a proteins are involved primarily in virus replication occurring on tonoplast membranes, while the 2b, 3a, and 3b proteins are involved in virus cell-to-cell or systemic movement. The 2b protein is one of the best studied viral suppressors of plant defenses based on RNA silencing. Over 100 years of research on this agronomically important pathogen have yielded much knowledge on diverse aspects of virus biology, from epidemiology, ecology and evolution to the molecular mechanisms involved in virus-host interactions. Different isolates of CMV induce a variety of disease responses, which are attenuated by a satellite RNA (satRNA) associated with some isolates of CMV, although some satRNAs can enhance pathogenicity. Control of CMV is principally based on sanitary selection, prophylactic measures and use of resistant plant cultivars.

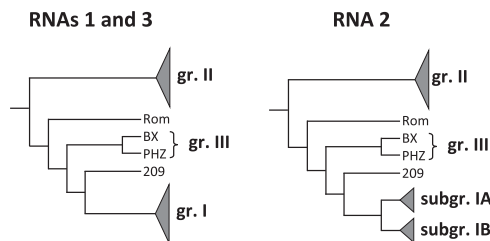


Fig. 1 Schematic topology of phylogenetic trees representing CMV genetic diversity from full-genome sequences. Whereas numerous isolates from subgroups IA and IB and group II have been entirely sequenced (indicated by gray triangles), only one or two isolates of group III or outliers (isolates Rom from rosemary, 209 from soybean and BX and PHZ from *Pinellia ternata*) have been sequenced. Only branches supported by bootstrap values >70% are depicted (branch lengths are not at scale). The tree was rooted with other members of the genus *Cucumovirus*.

Taxonomy and Phylogeny

CMV isolates differ from members of the other three species in the genus *Cucumovirus*, *Tomato aspermy virus* (TAV), *Peanut stunt virus* (PSV) and *Gay feather mild mottle virus* (GMMV), having only 50%–67% nt sequence identity, depending on the RNA and isolates being compared. Isolates of CMV are heterogeneous in symptoms, host range, transmission, serology, physicochemical properties, and nucleotide sequence of the genomic RNAs. CMV isolates can be classified into two major genetic groups, I and II. The nt identity between isolates belonging to these groups ranges from 69% to 77%, depending on the pair of isolates and the RNA segment compared, dissimilarity being highest for RNA2. Within a group, nt identity is above 88% for group I and above 96% for group II. Isolates of groups I and II can also be distinguished using monoclonal antibodies. RNA2 sequences of group I CMV isolates correspond to two monophyletic subgroups named IA and IB (Fig. 1). Subgroup IA, but not subgroup IB, is also monophyletic for RNA3, while RNA1 shows no clear division into subgroups IA and IB. This indicates that the different genomic segments have followed different evolutionary histories. Cross-protection occurs between isolates from all groups or subgroups. A few CMV isolates do not belong to groups I or II (Fig. 1).

Geographic Distribution

CMV isolates have a worldwide distribution, having been reported from both temperate and tropical regions. Most reported isolates belong to group I. Group II isolates are found more frequently in cooler areas or seasons of temperate regions. This has been associated with lower temperature optima for *in planta* virus accumulation shown for the few isolates characterized for this property. Most isolates in subgroup IB have been reported from East Asia, which is presumed to be the origin of this subgroup. Subgroup IB isolates have also been reported from other areas, for example, the Mediterranean region, California, Brazil, Africa and Australia. Those in the Mediterranean Basin could have been introduced recently from East Asia.

Host Range

Considering all isolates collectively, the host range of CMV includes 1071 plant species belonging to 100 different families, monocot or dicot. Of these plants, more than 500 species were found infected naturally, whereas the others were only found infected in laboratory conditions. Four families include more than 100 host species each: the *Solanaceae*, *Asteraceae*, *Fabaceae* and *Brassicaceae*. CMV infects most of the major horticultural crops as well as many weed species acting as reservoirs for the virus. However, considerable differences in host range breadth exist between individual CMV isolates. Most isolates have a broad host range while a few isolates have a very narrow host range, like certain isolates from soybean, lily or rosemary. These ‘soybean’ and ‘rosemary’ isolates, as well as isolates from *Pinellia ternata*, are phylogenetically distant from groups I and II (Fig. 1). Recent evidence has expanded the host range of CMV outside of the plant kingdom: a phytopathogenic fungus, *Rhizoctonia solani*, was described as a natural alternative host for CMV. This non-plant host may have an important impact on epidemiology of the disease, since the fungus could acquire CMV from an infected plant and transmit it to a healthy plant under laboratory conditions.

Symptoms and Diagnosis

Symptoms induced by CMV include leaf mosaic or mottling, vein-clearing, filiformism (the shoestring appearance of the leaf blade, a typical CMV symptom on tomato leaves), stunting, chlorosis and necrosis (including necrotic “oak-leaf” symptoms on older leaves) as well as decreased fruit or seed yield (Fig. 2). Pepper and cucurbit fruits, among others, can display symptoms such as mottling, mosaic, ring spotting or deformations. Symptom type and severity can vary greatly depending on the host species, the particular combination of plant and virus genotypes, the phenological stage at the time of infection, environmental conditions, as well as the presence of other

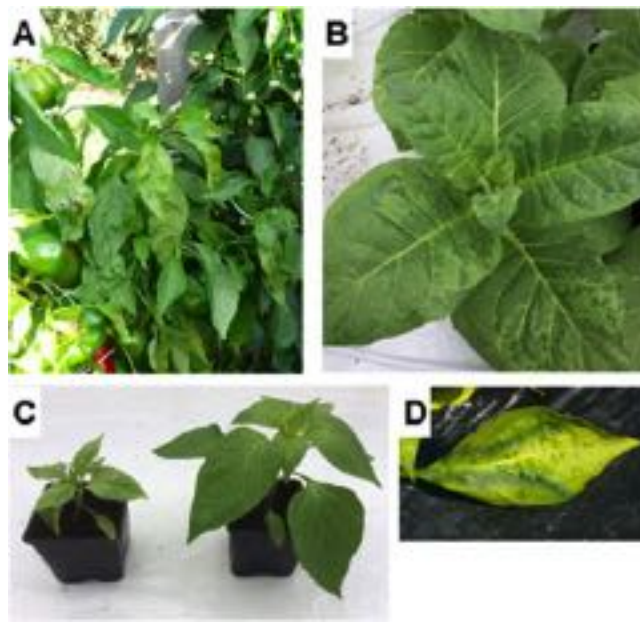


Fig. 2 Symptoms associated with CMV infection. (A) CMV-induced symptoms in a pepper field. (B) Mosaic symptoms on tobacco leaves infected with CMV I17F isolate (belonging to the IA subgroup). (C) Pepper plants mechanically inoculated with CMV I17F isolate displaying stunting and mosaic symptoms two weeks after inoculation (left infected plant, right mock-inoculated plant). (D) Necrotic oak-leaf pattern on naturally infected pepper leaf. Photographs Judith Hirsch, INRAE Avignon.

biotic factors such as satRNAs and other viruses, which can modulate the disease symptoms induced by CMV. CMV infection can also be symptomless. This is the case in many weed species which can serve as asymptomatic CMV reservoirs.

Host species which are systemically infected by most CMV isolates and commonly used as indicator plants include squash (*Cucurbita pepo*), cucumber (*Cucumis sativus*), tobacco (*Nicotiana tabacum*), as well as *N. benthamiana*, *N. clevelandii*, and *N. glutinosa*. *Chenopodium amaranticolor*, *C. quinoa* and cowpea (*Vigna unguiculata*) are local lesion hosts.

Serological methods like enzyme-linked immunosorbent assay (ELISA) have been successfully used for the large scale detection and diagnosis of CMV. Although molecular methods like reverse transcription-polymerase chain reaction (RT-PCR) are more sensitive, since CMV usually accumulates to high levels in infected plants, it is readily detected using ELISA, which makes this serological assay the method of choice for detecting CMV in plant samples. Another widely used serological diagnostic method is the lateral flow test (immune-strip tests based on this method are commercially available). Although lateral flow tests are relatively expensive, they constitute a rapid means of screening crops for the presence of CMV, particularly adapted for diagnostics in the field or greenhouse.

Cytopathology

Cytopathology associated with infection by CMV includes viral inclusions within the cytoplasm and vacuoles, or forming membrane-bound clusters in sieve elements. In some cases, angular inclusions corresponding to virus crystals can be seen in vacuoles by staining and light microscopy. CMV infection usually also leads to proliferation of cytoplasmic membranes, which originate from the plasma membrane, the endoplasmic reticulum (ER), or the tonoplast. Effects on the nucleus or nucleolus (e.g., vacuolation), usually due to virion accumulation, have been observed with some isolates of CMV. Similarly, some isolates have caused effects on mitochondria or chloroplasts. Yellowing isolates, in particular, lead to smaller and rounded chloroplasts that have fewer grana and starch granules. These various effects appear to be host specific. How virus-plant interactions lead to cytopathic effects remains unknown; however, as the virus expands from the initial site of infection, rings or zones of responses occur, in which the expression patterns of numerous plant genes are altered in spatial- and temporal-specific manners.

Virion Structure

Transmission electron microscopy observations allow visualization of the uniform isometric (roughly spherical) particles of CMV. The atomic structure of CMV has been determined to a 3.2 Å resolution using X-ray crystallography. CMV virions are icosahedral particles 29 nm in diameter. They are composed of 180 subunits of a single capsid protein (CP) of approximately 24 kDa, and contain approximately 18% RNA. RNA-protein interactions stabilize the virions and particle assembly requires a nucleic acid as a nucleating factor (no empty particles are formed). Virions are disrupted by sodium dodecyl sulfate and concentrated chloride salts, but can be

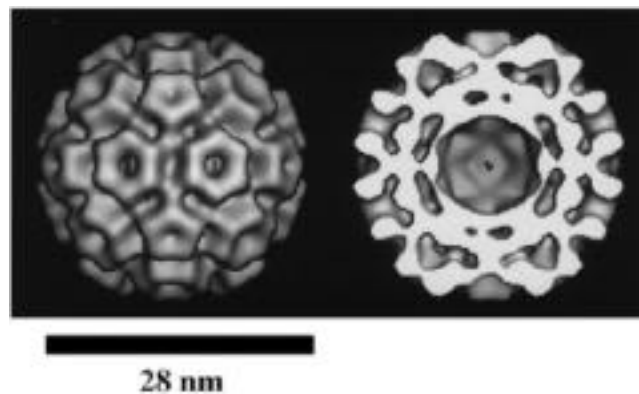


Fig. 3 Cryo-transmission electron microscopy reconstruction of CMV particles. Icosahedral CMV virions are composed of 12 pentamers and 20 hexamers. Reproduced from Wikoff, W.R., Tsai, C.J., Wang, G., Baker, T.S., Johnson, J.E., 1997. The structure of cucumber mosaic virus: Cryo-electron microscopy, X-ray crystallography, and sequence analysis. *Virology* 232 (1), 91–97.

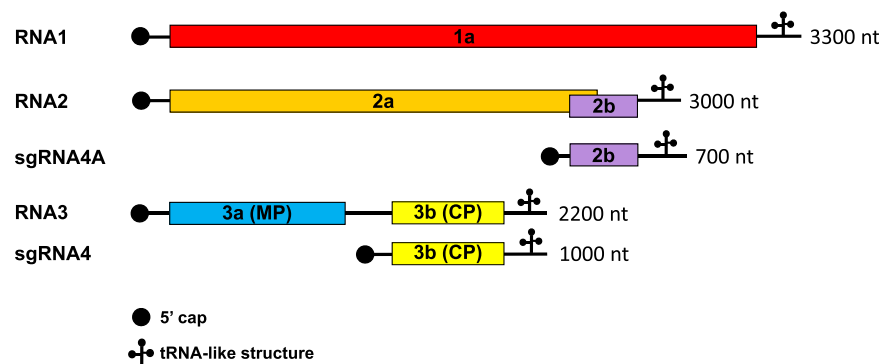


Fig. 4 Schematic representation of the genome organization of CMV. The CMV genome is composed of 3 genomic RNAs, RNAs 1, 2, and 3 encoding 5 proteins. Two sub-genomic RNAs, RNAs 4A and 4 are transcribed from the minus RNA2 and RNA3 strands, respectively. The open reading frames are depicted as colored bars. The approximate length of the RNAs are indicated. nt, nucleotides; sgRNA, sub-genomic RNA; MP, movement protein; CP, capsid protein.

reassembled *in vitro* by removing salts or sodium dodecyl sulfate. Each CP subunit adopts a canonical β -barrel structure, and all subunits assume one of three different configurations (A, B, or C). The A subunits cluster to form pentameric capsomers (or pentamers) and the B and C subunits form hexameric capsomers (or hexamers). The capsid is composed of 12 pentamers and 20 hexamers arranged with $T=3$ symmetry (Fig. 3). The N termini of the B and C subunits form a bundle of six α helices. This hexameric bundle constitutes a hydrophobic pore lined with leucine residues that extends down to the RNA-capsid interface. The first 30 residues of the CP (upstream of the α helix) are rich in positively charged aa, including 7 conserved arginine residues, that are proposed to form the interface with the RNA core. RNAs that can be encapsidated in CMV virions include genomic RNAs 1, 2, and 3, sub-genomic RNAs (sgRNAs) 4 and 4A, RNA5, transfer RNAs (tRNAs), as well as satellite and defective RNAs (see below).

Genome Organization

The genome of CMV consists of five genes distributed over three, single-stranded, positive-sense, capped genomic RNAs designated RNAs 1, 2, and 3 in descending order of size (Fig. 4). CMV is a multipartite virus, meaning that the three genomic RNAs are encapsidated in separate virus particles. RNA1 (~3.3 kb) encodes the 111 kDa 1a protein. RNA2 (~3.0 kb) encodes the 98 kDa 2a protein, as well as the 13–15 kDa 2b protein, which is translated from a ~0.7 kb sub-genomic RNA designated RNA 4A that is co-terminal with the 3' end of RNA2. The Open Reading Frame (ORF) expressing the 2b protein overlaps with the ORF encoding the 2a protein, but in a +1 reading frame. RNA3 (~2.2 kb) encodes the 30 kDa 3a protein, as well as the 25 kDa 3b protein, which is expressed from a ~1 kb sub-genomic RNA designated RNA4 that is co-terminal with the 3' end of RNA3. RNA4 is co-encapsidated with RNA3. The 3' untranslated regions (UTRs) of all three genomic and both sub-genomic RNAs are highly conserved, forming a tRNA-like structure as well as several pseudoknots. The 5' UTR regions of RNA1 and RNA2 are more conserved in sequence with each other than with those of RNA3. CMV also produces an RNA5 of unknown function, which is co-terminal with the 3' UTR regions of RNA1 and RNA2. RNA4A and RNA5 are only encapsidated by group II isolates. CMV

particles also encapsidate a low level of tRNAs, which have been reported in the literature as CMV RNA6. Although rarely reported, some isolates of CMV can also encapsidate defective RNAs derived from CMV RNA3.

CMV can also support satRNAs varying in size from 333 to 405 nt. These satRNAs do not encode any proteins. They are dependent upon CMV as the helper virus for both their replication and encapsidation, but have sequence similarity to the CMV RNAs limited to no more than 6–8 contiguous nts. The origin of satRNAs remains enigmatic. The *de novo* emergence of satRNAs after serial passages of CMV devoid of satRNA in tobacco plants suggests that they may have arisen from the host genome. Failure to detect long satRNA homologs in host plant genomes has led to an interesting hypothesis: that host genome-encoded small RNAs, possibly induced by viral infection, could be assembled by the viral replicase to form satRNAs. More than 180 satellite variants have been found associated with over 65 isolates of CMV from both of the CMV groups. These satRNAs usually reduce the accumulation of the helper viruses and on most hosts also reduce the virulence of CMV. However, certain satRNAs can enhance the disease induced by CMV in some hosts. In the case of tomato plants infected by CMV and certain satellites, this has led to systemic necrosis observed in the fields of several Mediterranean countries.

Properties and Functions of Gene Products

The CMV 1a and 2a proteins contribute to virus replication but also promote virus movement in several host species. The 2b protein is an RNA-silencing suppressor that interferes with the salicylic acid (SA) and jasmonic acid (JA) plant defense pathways and also influences virus movement in some hosts. The predominant mechanism by which the 2b protein inhibits antiviral silencing is by sequestering virus-derived small RNAs. Additionally, the 2b protein binds to Argonaute (AGO) proteins, the catalytic component of the RNA-induced silencing complex (RISC), and inhibits their RNA cleavage activity. The contribution of the 2b-AGO interaction to antiviral silencing is minor, likely through the inhibition of signal amplification mediated by host RNA-dependent RNA polymerases (RdRps). Conversely, the 2b-AGO interaction plays an important role in the induction of developmental symptoms, through the disruption of microRNA- (miRNA) mediated post-transcriptional gene silencing. The 3a protein is the major movement protein (MP) of the virus and is essential for both cell-to-cell and long-distance (systemic) movement. The 3b protein is the sole viral CP and is also required for cell-to-cell and systemic movement, although the ability to form virions is not a requirement for intercellular trafficking of the virus. All of the CMV-encoded proteins are RNA-binding proteins. Viral RNAs from different groups and subgroups can be exchanged to form novel viruses, allowing mapping of some phenotypes to specific RNAs. Higher-resolution mapping requires the use of biologically active cDNA clones for generating chimeras and site-directed mutants. By these approaches, a large number of genetic determinants of pathogenicity or transmission have been mapped ([Table 1](#)). All CMV RNAs and genes can be involved in infection and symptomatology, depending on the plant species or genotype.

The biochemical mechanisms responsible for the variety of symptoms induced by CMV have rarely been determined. In the case of an isolate of CMV that induces chlorotic symptoms in tobacco, interaction of the virus CP with the ferredoxin I protein precursor interferes with its import into the chloroplast and is therefore responsible for the symptoms.

Moreover, particular satRNAs can also contribute to CMV symptomatology, sometimes causing an attenuation of symptoms, a property that has been used commercially in tomato production. Attenuation of disease appears to result from the production of satellite-derived small interfering RNAs (siRNAs) that can either directly target the conserved 3'UTR region of the CMV genomic RNAs and trigger their degradation through the host RNA silencing machinery or interfere with the RNA silencing suppressor activity of the 2b protein by binding to it.

The CMV Infection Cycle

Replication

CMV replication takes place on the vacuole membrane (the tonoplast). Replication involves the 1a and 2a proteins of the virus and a number of host proteins. The purified CMV replicase contains the 1a and 2a proteins, as well as a 50 kDa host protein of unknown function. The 1a protein of CMV contains an N-terminal methyltransferase domain involved in capping of the RNAs, as well as a putative C-terminal helicase domain, presumed to be required for the unwinding of the viral RNAs during replication. The 2a protein possesses a GDD (Glycine-Aspartic acid-Aspartic acid) motif found in most RdRps. The N-terminal region of the 2a protein interacts with the C-terminal region of the 1a protein *in vivo* and *in vitro*. CMV replication is initiated by the binding of the tonoplast-associated replicase to the tRNA-like structure and various pseudoknots present in the 3'UTR region of the positive-sense CMV RNAs. Minus-sense RNAs are then synthesized from each of the genomic RNAs and act as templates for synthesis of new plus-stranded genomic RNAs. The minus-sense RNA2 and RNA3 also serve as templates for the synthesis of the two plus-sense sub-genomic RNAs (4 and 4A), through recognition of the sub-genomic promoters present on the minus-sense RNAs. The sub-genomic RNAs are not themselves replicated, but defective and satellite RNAs are replicated by the CMV replicase. Differences in the relative levels of accumulation of the various CMV genomic and satellite RNAs have been observed in different host species, which is probably due to host-specific differences in template copying.

Like most plant viruses, CMV encodes a very small number of proteins and relies on plant host factors to fulfill the different steps of its infection cycle. CMV replication is intricately dependent on various host proteins. Several host factors which support or

Table 1 Molecular determinants of pathogenicity and transmission identified in CMV

Phenotype	Plant	Genome region	Mutation (codon)
Symptoms			
Chlorosis, mosaic, or necrosis	<i>Nicotiana tabacum</i>	CP gene	129, 111, 124
Green mosaic, yellow chlorosis	<i>Nicotiana</i> spp.	CP gene	111, 124
Top necrosis	<i>Nicotiana glutinosa</i>	2b gene alone or 2b and 2a genes	
Necrotic etch	<i>N. tabacum</i>	RNA 3 (5' UTR and/or MP gene)	
Stunting	<i>N. glutinosa</i>	CP gene	193
Severe symptoms	<i>Cucurbita pepo</i>	RNA 1	
Stunting, filiformism	<i>N. tabacum</i>	RNAs 1 and/or 2	
Fast infection and severe symptoms	<i>C. pepo</i>	RNAs 1, 2, and 3	267 of 2a and 168 of 3a
Severe symptoms	<i>Datura stramonium</i> , <i>Gomphrena globosa</i> , <i>Nicotiana clevelandii</i> , <i>Nicotiana x edwardsonii</i> , <i>N. glutinosa</i> , <i>N. tabacum</i> , <i>Solanum melongena</i>	RNA 2 or RNAs 2 and 3	
White-yellow leaves, systemic necrosis, fruit necrosis	<i>N. tabacum</i> , <i>Solanum lycopersicum</i>	Satellite RNAs	nucleotides 148, 149, 153, 158, 170 (chlorosis)
Host range and interaction with plant resistance			
Infection	<i>Lilium</i> spp.	RNA 1 (5' UTR and/or 1a gene)	485–826 of 1a
Infection	<i>C. pepo</i> , <i>Phaseolus vulgaris</i> , <i>Pisum sativum</i>	RNA 2	
Infection at 27°C but not at 37°C	<i>Cucumis melo</i>	RNA 1	
Infection	<i>Vicia faba</i>	RNA 3	
Infection	<i>Zea mays</i>	CP gene	129, 162
Infection	<i>Tetragonia expansa</i> , <i>Momordica charantia</i> , <i>Physalis floridana</i>	CP gene	129
Infection	<i>Lactuca saligna</i> , <i>Raphanus sativus</i>	RNAs 2 and 3	
HR triggering	<i>Arabidopsis thaliana</i> (RCY1 gene)	CP gene	
HR triggering	<i>Vigna unguiculata</i> (Cry gene)	2a gene	631, 641
HR triggering	<i>Phaseolus vulgaris</i> (RT4–4 gene)	2a gene	
HR triggering	<i>N. tabacum</i>	1a gene	461
HR triggering	<i>Nicotiana</i> spp. (9 species)	CP gene	36
Resistance breakdown	<i>C. melo</i> (cmv1 gene)	MP gene	64–68, 81, 171, 195
Resistance breakdown	<i>Capsicum annuum</i> (Cmr1 gene)	1a gene	865, 896, 957, 980
Aphid transmission			
Transmission efficiency	<i>N. tabacum</i>	CP gene	25, 129, 162, 168, 214
Transmission efficiency	<i>N. tabacum</i>	2b gene	
Seed transmission			
Transmission efficiency	Legumes	RNA 1	
Others			
Satellite replication	<i>C. pepo</i>	RNA 1	

Abbreviations: CP, coat protein (*i.e.*, 3b protein); HR, hypersensitive reaction; MP, movement protein (*i.e.*, 3a protein); UTR, untranslated region.

restrict virus multiplication have been identified (Table 2). Although the exact roles of these proteins in virus replication are not always clear, they are proposed to act through regulation, assembly and/or anchoring of the CMV replication complex to the tonoplast. One such protein is glyceraldehyde 3-phosphate dehydrogenase (GAPDH) which is critical for CMV replication and appears to be involved in assembly of the viral replication complex by promoting the interaction between the 1a and 2a proteins.

Comparatively, very few plant factors involved in cell-to-cell or systemic propagation of the virus have been identified (see below and Table 2) although CMV presumably co-opts a number of host factors for these steps of its infection cycle.

Post-translational modifications of viral replication proteins play an important role in the regulation of CMV replication. For example, phosphorylation of the 2a protein prevents its association with the 1a protein and thus regulates assembly of the replicase complex. This mechanism probably plays an important role in the regulation of plus-strand and minus-strand synthesis and the asymmetry of CMV replication. Indeed, the kinetics of accumulation of positive and negative strands are distinct and accumulation of CMV positive strand RNAs is approximately 100-fold that of negative strand RNAs. Free 2a protein is present in the cytoplasm, while the 1a protein is strictly membrane-associated. The 2a protein alone is sufficient to synthesize positive strands from negative strand templates, while both the 1a and 2a proteins are required for synthesis of negative strands. Therefore the switch from negative to positive strand synthesis may be controlled by 2a phosphorylation, resulting in dissociation of 1a-2a complexes.

Positive-sense viral RNAs serve as templates for replication and translation of viral proteins and are encapsidated to form virions. The balance between replication, translation, encapsidation and probably also movement of these viral RNAs determines the efficiency of virus multiplication.

Cell-to-Cell and Long-Distance Movement

CMV moves cell to cell via plasmodesmata until it reaches the vasculature, when the virus moves systemically via the phloem. The viral RNAs move as a nucleoprotein complex between cells involving the 3a MP. Although the virus does not appear to migrate as virus particles, the CP is also necessary for cell-to-cell movement. This movement does not involve interactions with microtubules. Conversely, the actin cytoskeleton appears to be involved in CMV cell-to-cell movement (Table 2). The MP has been shown to interact with and sever actin filaments, an activity associated with increase of the plasmodesmal size exclusion limit. CMV appears to move between epidermal cells as well as from epidermal cells down to mesophyll cells toward vascular cells. The virus replicates in all of these cell types, but not in the sieve elements of the vasculature. Long-distance movement includes three main steps: viral loading into the phloem tissues, movement within sieve elements, and unloading from these elements into cells of the sink tissues. Virion assembly may take place inside sieve elements from RNAs and capsid protein moving from neighboring vascular cells. Virion assembly is necessary in some, but not all, species for systemic infection. A 48 kDa phloem protein (PP1) from cucumber phloem exudates interacts with CMV particles *in vitro* and increases virus particle stability. How the virus moves from the vasculature back to mesophyll and epidermal cells is unknown.

The 2b protein is also involved in local and systemic movement of the virus. The 2b protein has indirect effects on virus movement through its RNA silencing suppressor activity, as well as direct effects on cell-to-cell movement, independently of this suppressor activity, as has been revealed in an alanine scanning mutagenesis study that identified residues in the 2b protein that are required for local movement but not for silencing suppression.

The elucidation of resistance mechanisms targeted against systemic spread of CMV is revealing some of the probably manifold resistance barriers preventing long-distance movement of the virus. For example, a Piezo-like mechanosensitive cation channel which suppresses systemic spread of CMV has been identified in the model plant *Arabidopsis* (see also in the "Control" section the molecular mechanisms involved in resistance conferred by the *cmv1* resistance gene in melon). Conversely, other host proteins like Tco1 (tobacco CMV 1a-interacting protein 1) promote systemic spread of the virus (Table 2).

Plant proteins that are essential to the virus infection cycle can serve as targets to screen the plant natural genetic diversity in order to identify resistance alleles or to engineer new resistances, using technologies such as CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated protein 9) genome editing or TILLING (Targeting Induced Local Lesions in Genomes).

Transmission

Seed transmission of CMV has been reported in 41 plant species, mostly in the families Fabaceae (26 species) and Cucurbitaceae (seven species), with efficiencies varying from less than 1% up to almost 100%. The virus may be present in the embryo, endosperm, and seminal integuments, as well as in pollen. RNA1, possibly *via* the encoded protein 1a, affects the efficiency of seed transmission. Horizontal transmission of CMV is mediated by aphids (family *Aphididae*) in a non-persistent and non-circulative manner. This mode of virus transmission (also called stylet-borne) is characterized by short acquisition and inoculation times (from seconds to minutes). The stylets of the aphid become contaminated with the virus during probing of infected tissue, through a series of short intracellular stylet punctures of epidermal or neighboring cells. The virions are retained at sites at the tip of the aphid stylet, and released by salivation during probing of non-infected plants. Over 80 species of aphids have been reported to transmit CMV, *Aphis gossypii* and *Myzus persicae* being two efficient and most studied vectors. Transmission efficiency depends on several factors, particularly the specific combination of virus isolate and aphid species, and the

Table 2 Plant proteins having a positive or negative effect on one or more steps of the virus infection cycle or that interact with CMV proteins and play a putative role in assembly, anchoring or regulation of the CMV replicase complex

Plant protein	Plant species	CMV ligand(s)	Cellular function	Function in CMV infection cycle ^a	Method(s)
Virus multiplication					
APUM5 Pumilio protein	<i>Arabidopsis thaliana</i>	3'UTR of RNAs (group I only)	Post-transcriptional regulation through sequence-specific RNA binding	Translation repression, inhibition of replication?	T-DNA insertion mutant, 35S-APUM5 and APUM5-RNAi transgenic plants
BRP1	<i>Nicotiana benthamiana</i>	1a protein	Scaffolding functions in transcription complexes	Promotes replication	Co-IP followed by LC-MS/MS; BRP1-suppressed transgenic plants
CIPK12	<i>Nicotiana tabacum</i>	2a protein	Protein kinase	Phosphorylation of 2a protein; regulation of replication by preventing 2a interaction with the 1a protein?	Y2H screen using 2a N-terminal domain as bait; <i>in vitro</i> phosphorylation assay
CRT3	<i>Capsicum annuum</i> <i>N. benthamiana</i>	1a protein	Calcium-binding, endoplasmic reticulum chaperone	Required for replication	Y2H screen with the 1a helicase domain as bait; TRV-based VIGS
eIF4E and eIF4G	<i>A. thaliana</i>	Viral RNAs (5'cap?)	Initiation of translation	Required for virus multiplication	<i>cum1</i> and <i>cum2</i> EMS mutants
FDH	<i>C. annuum</i> <i>N. benthamiana</i>	1a protein	Metabolism (enzyme catalyzing the oxidation of formate to CO ₂)	Promotes replication	Y2H screen with 1a helicase domain as bait; TRV-based VIGS
GAPDH	<i>N. benthamiana</i> <i>A. thaliana</i>	Unknown	Glycolysis	Stabilization of 1a-2a protein interaction; required for replication	Co-IP with BRP1 protein followed by LC-MS/MS; T-DNA knock-out line; transient complementation by agroinfiltration
Tcoi2	<i>N. tabacum</i>	1a protein	Tyrosine protein kinase	Phosphorylation of 1a protein. Regulation of 1a function?	Y2H screen ; <i>in vitro</i> phosphorylation assay
TIP1 and TIP2	<i>A. thaliana</i>	1a protein	Tonoplast aquaporins (water channels)	Anchoring of 1a protein to the tonoplast?	Y2H targeted test
Tsip1	<i>N. tabacum</i>	1a and 2a proteins	Zinc finger protein with unknown function	Inhibition of replication	Y2H screen ; <i>Tsip1</i> overexpressor and RNAi transgenic plants
Virus cell-to-cell movement					
Actin	<i>N. tabacum</i>	MP	Cytoskeleton	Actin-severing by	<i>in vitro</i> binding

Table 2 Continued

			constituent	MP increases size exclusion limit of plasmodesmata	studies; microinjection assays
Virus systemic movement					
AtPiezo/ESC1	<i>A. thaliana</i>	Unknown	Mechanosensitive cation channel	Inhibition of viral translocation into and/or transport in the phloem	<i>enhanced susceptibility to CMV 1 (esc1)</i> EMS, T-DNA insertion and CRISPR/Cas9 mutants
CsAO4	<i>Cucumis sativus</i> <i>A. thaliana</i> <i>N. benthamiana</i>	MP	Apoplastic ascorbate oxidase	Facilitates viral spread at the early stages of infection	Y2H screen with MP protein as bait; <i>CsAO4</i> overexpressing transgenic lines; TRV-based VIGS
p48/PP1	<i>C. sativus</i>	virions	Constituent of P-protein filaments in phloem sieve elements	Stabilization of virions during translocation into and/or movement in phloem sieve elements	Increased resistance of virions to RNase A treatment
Tcoi1	<i>N. tabacum</i>	1a protein	Methyltransferase	1a protein methylation, promotion of systemic spread	Y2H screen with 1a protein as bait; Transgenic sense and antisense <i>Tcoi1</i> plants

^a Are not included in this table plant factors that are part of the antiviral silencing machinery or involved in its regulation, which are not specific to CMV.

Text in blue font: no functional data available concerning the role of the host protein in the CMV infection cycle.

Abbreviations: APUM5, arabidopsis pumilio 5; BRP1, bromodomain-containing protein 1; CIPK12, CBL-interacting protein kinase 12; Co-IP, co-immunoprecipitation; CRISPR/Cas9, clustered regularly interspaced short palindromic repeats/CRISPR associated protein 9; CRT3, calreticulin-3 precursor; CsAO4, *cucumis sativus* ascorbate oxidase 4; eIF, eukaryotic translation initiation factor; EMS, ethyl methane-sulfonate; FDH, formate dehydrogenase; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; LC-MS/MS, liquid chromatography-tandem mass spectrometry; MP, movement protein; PP1, phloem protein 1; RNAi, RNA interference; Tcoi, tobacco CMV 1a-interacting protein; TIP, tonoplast intrinsic protein; TRV, tobacco rattle virus; Tsi1, Tobacco stress induced 1(Tsi1) interacting protein; UTR, untranslated region; VIGS, virus-induced gene silencing; Y2H yeast two-hybrid.

accumulation of particles in the source leaf. Differences in transmissibility of various isolates by different aphid species are determined primarily by the virus coat protein and aa determinants for transmission have been mapped (**Table 1**). The aa positions that determine transmission are either exposed on the outer surface of capsids (e.g., aa position 129) or lay in the inner surface (e.g., position 162). Hence, the effect could be through direct interaction with receptors in the aphid mouth parts or by affecting particle stability. On the vector side, there is accumulating evidence for the role of cuticular proteins of the CPR (cuticular proteins with the Rebers and Riddiford Consensus) family in the transmission of plant viruses. Putative CMV receptors belonging to this family are present in the acrostyle, a small structure inside the fused food/salivary common canal at the tip of aphid stylets. One of these proteins has been shown to act as a probable caulimovirus receptor. However, there is no direct evidence of binding of CMV in the common canal, and validating the role of these cuticular proteins in CMV transmission will require further biochemical and functional studies.

The 2b protein of CMV also seems to have an indirect effect on virus transmission by manipulating the behavior of aphids. By altering the jasmonate signaling pathway and increasing the plant reactive oxygen species content, the 2b protein simultaneously enhances attractiveness of host plants for aphid vectors and increases probing and dispersal of the aphids, thus favoring spread of the virus.

Transmission efficiency also depends on the host plant, as shown by the resistance to transmission of melon genotypes carrying the *Vat* gene, which does not confer resistance to the virus directly, but impairs CMV transmission by *Aphis gossypii*.

Variation and Evolution

CMV populations share evolution patterns of many plant RNA viruses, including high mutation rates and high intensities of selection and genetic drift. Moreover, phylogenetic analysis showed that recombination and particularly reassortment have largely contributed to the diversification of CMV populations, as seen in other viruses with segmented genomes. Using infectious cDNA clones and experimental evolution, CMV mutation and recombination frequencies were estimated. Mutation frequency varied from 0.6×10^{-4} to 1.8×10^{-3} mutation per nt in a 12–15 days passage in plants. Significant differences in mutation frequency were observed between plant species and between CMV isolates. The recombination frequency was also highly variable between isolates. Indeed, depending on the parental isolates, from 1.8% to 28.2% of RNA3 recombinants were observed 15 days after inoculation with parental CMV clones. The 2a gene is a major determinant of recombination frequency differences. Recombination also seems to be facilitated by stem–loop structures in the RNA. The 3' UTR of CMV genomic RNAs is also a recombination hotspot. Recombinants in this region may have an increased fitness in some hosts, as shown for isolates infecting *Alstroemeria* spp. However, analyses of field populations of CMV show that recombinants are not frequent, suggesting that selection operates against most of them. A second mechanism of genetic exchange is reassortment of genomic segments also called pseudo-recombination. Reassortants exchanging any genomic segment have been obtained between different CMV isolates and multiply efficiently under experimental conditions. Natural CMV reassortants also have been described but are generally rare and seem to be selected against, as is the case for recombinants, which suggests co-adaptation of the different viral genes that, when disrupted, results in a decreased fitness. However, reassortants may reach higher frequencies as shown in pepper (*Capsicum* spp.) crops in Tunisia and South Korea, where they comprised 90% and 20%, respectively, of the CMV populations. Selection by the host plant and possibly by resistant cultivars may be responsible for this over-representation of reassortants.

Genetic diversity of CMV has been shown to be countered by genetic drift associated to population bottlenecks during plant colonization and horizontal transmission by aphids. For the latter, from one to two CMV infectious units are inoculated by a single aphid. Selection is also an evolutionary force that drastically restricts the diversity within virus populations. Sequence analysis have shown different evolutionary constraints for the different viral proteins. Protein 1a and the MP are the most constrained, whereas the 2b protein is the least constrained. The evolution constraint acting on the CP differs greatly between CMV groups and subgroups. Positively-selected aa positions have been detected in all CMV proteins and may be responsible for CMV adaptation to host plants or to the vector.

Analyses of the population structure of CMV in Spain and California indicate a metapopulation structure, with local extinctions and recolonizations, which suggests that population bottlenecks occur, probably associated with unfavorable seasons for the host plants and/or the aphid vectors. Interestingly, this is not the case for the population structure of the satRNA. Analyses in Italy and Spain during epidemics of CMV and the associated satRNAs have shown that the satRNAs have an undifferentiated population. The different population structures for CMV and its satRNAs indicate that the satRNAs expanded as a molecular parasite on the CMV population, rather than satellite expansion was linked to a particular CMV isolate.

Epidemiology and Control

CMV infects a wide range of food crops, ornamentals and wild plant species. Economic losses in crops are highest in field-grown vegetables and ornamentals, pasture legumes and banana. In recent times, CMV has caused severe epidemics in many crops, including necrosis of tomato in Italy, Spain, and Japan; mosaic and heart rot of banana worldwide; mosaic of melons in California and Spain; mosaic of pepper in Australia, California, Turkey and Tunisia; and mosaic of lupins and other legumes in Australia and the USA.

As for many plant viral diseases for which there are no curative methods, control of CMV begins with sanitary selection, planting or sowing virus-free material and prophylactic measures. In crops in which seed transmission is effective (e.g., pasture or fodder legumes), the primary inoculum for epidemics may be the seedlings from infected seeds and even a low frequency of seed infection may have dramatic consequences because of accelerated secondary spread of the virus. In most vegetable and ornamental crops, seed transmission does not occur or is negligible, and the primary inoculum must come from outer sources as other crops or weeds, which should be near the crop as aphid transmission is non-persistent. In the absence of crops during unfavorable seasons, infected perennial weeds or crops, and infected seeds from weeds act as inoculum reservoirs. Seed transmission has been shown to be important in several weed species from different regions. The dynamics of virus infection may differ largely in weeds and crops within a region, indicating that the relevant reservoirs and inoculum sources for crops need not be the most frequently infected weeds. In banana, secondary spread of infection within the crop is ineffective for most isolates, and alternative hosts are both primary and secondary inoculum sources for epidemics. Since many species of weeds act as reservoirs for the virus and many of these are asymptomatic hosts, it is important to remove them from the borders of fields. This also applies to removal of infected crop plants during the growing season.

A number of agricultural practices can be implemented to reduce CMV epidemics. For example, avoiding overlap of the growing cycles of CMV-susceptible crops, adapting sowing or planting dates in order to create a lag time between peaks of aphid flights and the susceptibility period of crop plants and using cultivar or crop mixing to benefit from inoculum dilution and barrier effects can slow down CMV epidemics. The use of mineral oils or reflective mulches can also affect aphid behavior and contribute to CMV control.

Table 3 Characteristics of CMV resistance genes or loci

Plant	Gene or allele name	Nature of encoded protein	Resistance mechanism
<i>Arabidopsis thaliana</i>	<i>RCY1</i>	NLR	HR
<i>Capsicum annuum</i>	<i>Cmr1</i>	–	Inhibition of systemic movement
<i>C. annuum</i>	<i>cmr2</i>	–	Inhibition of cell-to-cell movement
<i>C. annuum</i>	Several QTLs	–	Inhibition of infection, virus multiplication and systemic movement
<i>Capsicum frutescens</i>	At least 2 recessive genes	–	Inhibition of cell-to-cell and systemic movements
<i>Cucumis melo</i>	<i>cmv1</i>	VPS41	Inhibition of systemic movement of group II strains
<i>C. melo</i>	<i>Vat</i>	NLR	Resistance to <i>Aphis gossypii</i> -mediated inoculation
<i>Cucumis sativus</i>	Several QTLs or 1 recessive gene	–	Symptom attenuation
<i>Cucurbita moschata</i> , <i>Cucurbita pepo</i>	<i>Cmv-2</i>	–	High-level resistance
<i>Glycine max</i>	Three QTLs	–	Inhibition of systemic movement
<i>Lupinus luteus</i>	<i>Nam-1</i>	–	HR
<i>Nicotiana tabacum</i>	Several QTLs	–	Inhibition of infection, symptom attenuation
<i>Phaseolus vulgaris</i>	<i>RT4-4</i>	NLR	HR
<i>Solanum chilense</i>	<i>Cmr</i>	–	Symptom attenuation
<i>Solanum habrochaites</i>	polygenic	–	Tolerance to satRNA-induced lethal necrosis
<i>Solanum tuberosum</i>	Single major gene	–	Inhibition of systemic movement
<i>Vigna unguiculata</i>	<i>Cry</i>	–	HR
<i>Zea mays</i>	Single dominant gene	–	HR

Abbreviations: HR, hypersensitive reaction; NLR, nucleotide-binding and leucine-rich repeat-containing protein; QTL, quantitative trait locus; satRNA, satellite RNA; VPS, vacuolar protein sorting; –, unknown data.

Control of CMV can also be achieved by planting of resistant crop varieties. Genetic resources for CMV resistance are available in many crop species and their genetic determinism has been unraveled in a few of them (Table 3). However, there is generally a lack of high-level resistance and of major-effect resistance genes against CMV. A large number of resistance mechanisms are polygenic, which hampers their exploitation by breeders, and partial. They frequently slow down the systemic movement of the virus and reduce symptoms. Some of them are ontogenic, *i.e.*, they are only expressed when plants achieve a given developmental stage. Their spectrum of action is also frequently incomplete and covers a few CMV isolates or isolate groups.

A novel mechanism has been discovered for the resistance conferred by the *cmv1* gene in *Cucumis melo*. *cmv1* inhibits CMV systemic infection by restricting the virus to the bundle sheath cells and impeding its loading into the phloem. *cmv1* was cloned and shown to encode a novel class of resistance factors, the Vacuolar Protein Sorting 41 (VPS41) which is required for post-Golgi vesicle trafficking towards the vacuole. CMV might use VPS41 for its transport towards the phloem and a single aa mutation in the protein abolishes or reduces this transport and converts the VPS41 gene into a recessive resistance gene. However, the resistance is efficient only against group II CMV isolates, which reduces its agronomical interest.

In pepper (*Capsicum* spp.), several resistance sources and mechanisms have been described and exploited to breed resistant cultivars. In France, cultivars carrying a polygenic resistance were bred using a recurrent selection program involving several parental lines with different partial resistance mechanisms (inhibition of inoculation efficiency, decrease of virus multiplication and inhibition of systemic movement). Whereas all resistance sources were partial, the composite resistance resulting from the combination of these mechanisms is almost total. The resistance was efficient and durable in the field but is threatened by adapted CMV isolates, belonging to the IB subgroup, that have been introduced in the Mediterranean basin. In South Korea, the major-effect dominant resistance gene *Cmr1* was used in breeding programs and deployed in the field. It inhibits the systemic movement of CMV. However, as in the previous case, it is impaired by subgroup IB isolates. The *cmr2* recessive resistance gene seems to possess a broader spectrum of action and may be a valuable resource for resistance breeding in pepper.

In addition to resistance, mechanisms of plant tolerance against CMV have also been studied. Tolerance is defined as the ability of the host to limit the damage caused by a given virus load, as opposed to resistance, which is the capacity of the host to reduce virus load. Tolerance to CMV has been studied in detail in the model plant *Arabidopsis thaliana* but has also been observed in several agronomically important plants, including bean, cucumber, pepper and tomato. The effect of infection by different CMV isolates on biomass and viable seed production of different *Arabidopsis* accessions was evaluated. Tolerance to different CMV isolates was higher in accessions with longer life cycles and was associated with resource reallocation from growth to reproduction. This tolerance response is CMV-specific as it was not observed in response to infection by other more virulent viruses. Whether similar tolerance mechanisms exist in cultivated plants remains to be determined. As it exerts a weaker selection pressure on virus populations, tolerance is expected to be more durable than resistance. Thus, whatever the mechanisms involved, breeding CMV tolerant crops is a potential source of sustainable protection against CMV.

Other sources of resistance include the use of transgenic plants expressing either protein-mediated or RNA silencing-mediated resistance. Although most of these approaches have led to resistance to only members of one of the two major groups, the use of

pyramiding of viral segments from different groups offers the promise of obtaining a broad-spectrum resistance to CMV together with other viruses infecting the same crop species. Transgenic expression of satRNAs has also been used to confer resistance to CMV. This has met with success, but has raised concerns about favoring the emergence of a virulent pathogen, as did the use of mild isolates of CMV for cross-protection against severe isolates.

The use of insecticides to control the aphid vectors of the virus has met with only limited success, since the virus is transmitted in a non-persistent manner and thus the aphids can transmit the virus before being killed by the insecticide. Rather, insecticides are used to reduce aphid population size and thus reduce the spread of infections. Similar results were obtained through the biological control of aphid populations by the use of predators.

See also: Brome Mosaic Virus (*Bromoviridae*). Ilarviruses (*Bromoviridae*). Satellite Nucleic Acids and Viruses. Viral Suppressors of Gene Silencing

Further Reading

- Ben Tamarzizt, H., Montarry, J., Girardot, G., *et al.*, 2013. Cucumber mosaic virus populations in Tunisian pepper crops are mainly composed of virus reassortants with resistance-breaking properties. *Plant Pathology* 62, 1415–1428.
- Deshoux, M., Monsion, B., Uzest, M., 2018. Insect cuticular proteins and their role in transmission of phytoviruses. *Current Opinion in Virology* 33, 137–143.
- Giner, A., Pascual, L., Bourgeois, M., *et al.*, 2017. A mutation in the melon vacuolar protein sorting 41 prevents systemic infection of Cucumber mosaic virus. *Scientific Reports* 7, 10471.
- Jacquemond, M., 2012. Cucumber mosaic virus. *Advances in Virus Research* 84, 439–504.
- Lewsey, M.G., Murphy, A.M., Maclean, D., *et al.*, 2010. Disruption of two defensive signaling pathways by a viral RNA silencing suppressor. *Molecular Plant-Microbe Interactions* 23 (7), 835–845.
- Pagán, I., Alonso-Blanco, C., García-Arenal, F., 2007. The relationship of within-host multiplication and virulence in a plant-virus system. *PLoS One* 2 (8), e786.
- Palukaitis, P., García-Arenal, F., 2019. *Cucumber Mosaic Virus*. The American Phytopathological Society.
- Seo, J.-K., Kwon, S.-J., Choi, H.-S., Kim, K.-H., 2009. Evidence for alternate states of Cucumber mosaic virus replicase assembly in positive- and negative-strand RNA synthesis. *Virology* 383, 248–260.
- Wang, M., Smith, N.A., 2016. Satellite RNA pathogens of plants: impacts and origins – An RNA silencing perspective. *Wiley Interdiscip. Rev. RNA* 7, 5–16.
- Wu, D., Qi, T., Li, W., *et al.*, 2017. Viral effector protein manipulates host hormone signaling to attract insect vectors. *Cell Research* 27, 402–415.

Dianthovirus (*Tombusviridae*)

Kiwamu Hyodo, Okayama University, Kurashiki, Japan

Masanori Kaido, Kyoto University, Kyoto, Japan

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
CP Coat protein
gRNA Genomic RNA
kb Kilobase
kDa Kilo dalton
MP Movement protein

nt Nucleotide(s)
RdRp RNA-dependent RNA polymerase
sgRNA Sub-genomic RNA
SL Stem loop
ssRNA Single-stranded RNA
UTR Untranslated region
VRC Viral replicase complex

Glossary

Dicer Eukaryotic ribonuclease that cuts long dsRNAs and structured ssRNAs into ~20–25 bp fragments with 2-nt 5' extensions.

Frameshift A change of the reading frame, which results in a different translation and a different protein product.

Plasmodesmata (Singular: plasmodesma), channels traversing the plant cell walls, allowing for the cell-to-cell transport of substances. Plasmodesmata are internally coated by the endoplasmic reticulum.

Silencing suppressor (Suppressor of silencing): virally encoded proteins from diverse families that interfere with host production or effectiveness of small RNAs.

Introduction

The members of the family *Tombusviridae* have uncapped, single-stranded, positive-sense RNA [(+)ssRNA] genomes. The family *Tombusviridae* can be subdivided into three subfamilies: *Calvirusirinae*, *Procedovirinae*, and *Regressovirinae*. The subfamily *Regressovirinae* contains the genus *Dianthovirus*, the members of which have icosahedral virions and a genome consisted of two (+)ssRNA molecules. Although two genomic RNAs share viral and pro-viral host factors during infection, they exploit different strategies to regulate their translation, transcription, replication, packaging, and cell-to-cell movement.

Taxonomy and Classification

Dianthoviruses belong to the family *Tombusviridae* and the subfamily *Regressovirinae*. Of the 16 genera in the family, only the dianthoviruses have bipartite genomes. The genus *Dianthovirus* is composed of three viruses: *Carnation ringspot virus* (CRSV), *Red clover necrotic mosaic virus* (RCNMV), and *Sweet clover necrotic mosaic virus* (SCNMV) (Table 1). *Furcraea necrotic streak virus*, which was provisionally classified as a member of the genus *Dianthovirus*, is now classified as a member of the genus *Macanavirus* in the family *Tombusviridae*. The amino acid (aa) sequence of dianthovirus coat protein (CP) has high homology with those of the tombusviruses and aureusviruses that belong to the family *Tombusviridae*. Interestingly, the aa sequence of dianthovirus RNA-dependent RNA polymerase (RdRP) has higher homology with that of the luteoviruses that belong to the family *Luteoviridae*, than with those of the viruses in the family *Tombusviridae*.

Genome Organization

The genome of dianthoviruses consists of two single-stranded positive-sense RNA {(+)RNA} molecules, RNA1 (3.9 kb) and RNA2 (1.4 kb) (Fig. 1). RNA1 encodes two replication-associated proteins: an auxiliary 27 kDa protein (p27), an N-terminally overlapping 88 kDa RdRp (p88^{POL}); and the 37 kDa CP, which is the building block of the mature virion. A non-coding viral RNA, called SR1f, accumulates during dianthovirus infection. SR1f is generated from the 3'-untranslated region (UTR) of RNA1 through the action of a host 5' → 3' exoribonuclease. The 3'UTR of dianthovirus RNA1 contains a sequence that inhibits 5' → 3' exoribonucleolytic decay in *cis*. SR1f may suppress translation from RCNMV RNA1 and thus could play a regulatory role in infection by dianthoviruses. RNA2 encodes a 35 kDa movement protein (MP), which plays an essential role in viral cell-to-cell movement and systemic spread in plants.

Virions, Transmission, and Epidemiology

Similarly, to the viruses in other genera, members of the genus *Dianthovirus* have icosahedral virions of approximately 35 nm in diameter with *T* = 3 symmetry that consist of 180 CP subunits of approximately 36–37 kDa. RCNMV virions consist of two

Table 1 List of members of the genus *Dianthovirus*, the length of their genomes, and their accession numbers

Name	Abbreviation	Isolate	Genome segment length (nt)	Accession no.
Carnation ringspot virus	CRSV	^a	RNA1: 3840	NC_003530
			RNA2: 1403	NC_003531
		1.30	RNA1: 3843	LC461179
			RNA2: 1404	LC461180
		PV-21	RNA1: 3872	LC461178
			RNA2: 1405	LC461177
Red clover necrotic mosaic virus	RCNMV	Australia ^a	RNA1: 3851	LC460998
			RNA2: 1406	LC461176
		Canada	RNA1: 3890	NC_003756
			RNA2: 1448	NC_003775
Sweet clover necrotic mosaic virus	SCNMV	59 ^a	RNA1: 3890	AB034916
			RNA2: 1456	AB034917
		38	RNA1: 3876	NC_003806
			RNA2: 1449	NC_003807
			RNA2: 1446	S46027

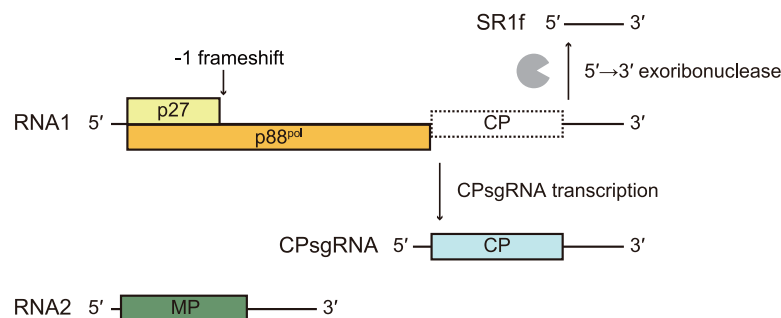
^aExample isolate.

Fig. 1 A schematic representation of the dianthovirus genome structure. The genome of dianthovirus consists of RNA1 and RNA2, which lack a 5'-cap structure and a 3'-poly(A) tail. RNA1 encodes an auxiliary replication protein p27 and an RNA-dependent RNA polymerase (RdRp) p88^{pol} which is translated via a -1 ribosomal frameshift mechanism. RNA1 also encodes a coat protein (CP), which is translated from sub-genomic RNA (CPsgRNA). SR1f, a small noncoding RNA, is generated from the 3'-untranslated region (UTR) of RNA1 by a host 5' → 3' exoribonuclease. RNA2 encodes a movement protein (MP) that is required for viral cell-to-cell movement.

distinct types: one contains one copy of each genomic RNA and the other contains four copies of RNA2. The origin of assembly sequence (OAS) is located in the MP coding region of RNA2. The stem-loop (SL) structure formed from the OAS is also known as the transactivator (TA), and is a multifunctional region that is required for the transcription of sub-genomic RNA and the replication of RNA2 (see the sections below that describe gene expression and RNA replication). RNA1 does not have an independent packaging signal and is co-packaged with RNA2 through an interaction with the SL structure in RNA2. The N-terminal 50 aa of RCNMV CP are rich in lysine residues and are essential for virion formation.

RCNMV virions are stable in conditions in which divalent cations such as calcium and magnesium ions are abundant. In contrast, the virions are unstable when divalent cations are depleted. Low-divalent-cation environments such as those found within the cytoplasm of a cell might be favorable for uncoating the virion and exposing the genomic RNAs for gene expression. The unique response of RCNMV virions to changes in divalent cation concentrations has been exploited for the development of plant viral nanoparticles for controlled delivery of the anticancer drug doxorubicin. In addition to medical applications, a study has shown that RCNMV nanoparticles improve the performance of abamectin, a potent nematocide that targets a broad spectrum of plant parasitic nematodes, by enhancing the poor mobility of abamectin in soil.

Dianthoviruses multiply to high concentrations in infected plants and the virions are stable *in vitro*. The release of dianthoviruses from infected roots into surrounding environments such as soil and drainage water has been reported. However, it has not been determined whether soil-inhabiting organisms transmit dianthoviruses in the field. *Olpidium* spp., nematodes, and bees have been proposed as vectors of RCNMV, CRSV, and SCNMV, respectively.

Gene Expression

After entry into host cells, the genomic RNA of (+)RNA viruses recruits host ribosomes to act as a template for translation. Like all viruses in the family *Tombusviridae*, the genomic RNAs of dianthoviruses lack both a 5'-cap structure and a 3'-poly(A) tail.

Thus, to produce their encoded proteins efficiently, dianthoviruses use cellular translational machinery in a non-canonical, cap-independent mechanism. RCNMV uses different modes for cap-independent translation of its bipartite genomic RNAs.

The protein p27 is directly translated from RNA1 in a cap-independent manner. Two *cis*-acting RNA elements play essential roles in the translation of RNA1: a 3'-cap-independent translation enhancer termed 3'TE-DR1 and an adenine-rich sequence (ARS) in the 3'UTR. 3'TE-DR1 is essential for recruiting the eukaryotic initiation factors eIF4F/eIFiso4F and ribosomes to RNA1. ARS, located 60 nucleotides upstream from 3'TE-DR1, recruits a host poly(A)-binding protein (PABP) to RNA1 through a direct RNA-protein interaction. The ARS-PABP interaction mediates the recruitment of eIF4F/eIFiso4F to 3'TE-DR1. Although 3'TE-DR1 can bind to both eIF4F and eIFiso4F, RNA1 preferentially utilizes eIF4F for its translation. Because the cap-independent translation of RNA1 is scanning dependent, ribosomes should be recruited to the 5-end of the viral RNA. However, the mechanism by which the 3'UTR of RNA1 coordinates the recruitment of ribosomes to the 5'end remains elusive.

The protein p88^{pol} is translated via -1 programmed ribosomal frameshifting (-1PRF) from RNA1. Three *cis*-acting RNA elements required for -1PRF have been identified: a shifty heptanucleotide sequence GGAUUUUU where the reading frame shifts, a highly structured bulged SL immediately downstream from the slippage site, and a small stable SL structure located between ARS and 3'TE-DR1. Intramolecular base-pairing between the bulge of the second element and the loop sequence of the third element mediates efficient -1PRF of RCNMV RNA1. This long-distance RNA-RNA interaction is likely to be conserved in all dianthoviruses.

CP is translated from sub-genomic RNA (CPsgRNA). CPsgRNA is likely to be generated via a premature termination mechanism, where the viral replicase terminates internally during negative-strand RNA {(−)RNA} synthesis. The 3'-truncated (−)RNA is then used as a template for CPsgRNA synthesis. CPsgRNA transcription requires intermolecular base-pairing between RNA2 (TA element) and RNA1 (TA-binding site located immediately upstream of the transcription start site), which is likely to be important for premature termination of the viral replicase complex (VRC) during (−)RNA synthesis. CPsgRNA is identical to the 3'-terminal 1.5 kb of RNA1 and therefore possesses the same 3'TE-DR1, which is also essential for translation from CPsgRNA.

MP is translated from RNA2. Unlike RNA1, RNA2 lacks *cis*-acting RNA elements such as 3'TE-DR1 and ARS. Instead, cap-independent translation of RNA2 seems to be coupled with RNA2 replication. That is, only the RNA2 that is generated *de novo* through replication seems to act as a template for translation. Interestingly, RNA2 preferentially utilizes eIFiso4F rather than eIF4F. The *cis*-acting RNA elements that recruit eIFiso4F and/or ribosomes to RNA2 have not yet been identified. A translation-silencer element has been identified in the basal stem structure of the 5'UTR (5'BS) of RNA2, and translation-enhancer activity is found in the 5'UTR other than 5'BS. Therefore, the 5'UTR may coordinate cap-independent and replication-coupled translation of RNA2. The molecular mechanisms underlying replication-coupled translation of RNA2 remain elusive.

RNA Replication

Replication of RCNMV takes place at the cytoplasmic surface of the host endoplasmic reticulum (ER) membrane. At the early phase of replication, RCNMV replicates at punctate structures on cortical ER. Later, such punctate structures disappear, and instead, a large aggregated ER structure emerges at the perinuclear region. p27 plays a key role in the biogenesis of the viral replication compartments. An amphipathic α -helix of p27 is essential for targeting p27 to ER membranes and the assembly of the VRC (termed the 480 kDa replicase complex). The p27-p27 and the p27-p88^{pol} interactions are prerequisite for the assembly of the 480 kDa replicase complex. The C-terminal half of p27 is responsible for the p27-p27 and the p27-p88^{pol} interactions, whereas the nonoverlapping region of p88^{pol} is responsible for the p27-p88^{pol} interaction.

After translation in the cytoplasm, the genomic RNA of (+)RNA viruses must be recruited to the viral replication compartment and used as a template for (−)RNA synthesis. To distinguish viral genomic RNAs from host-derived RNAs, viral replication proteins must have mechanisms to specifically recognize viral RNAs as replication templates. *Cis*-acting RNA elements that are important for replication have been identified in the 3'UTR of both genome segments. In RNA1, two SLs, termed SLDE and SLF, and the intervening sequence are likely to act as the core promoter for (−)RNA synthesis. The corresponding *cis*-acting RNA elements are also present in RNA2. Interestingly, RNA2 has additional RNA elements that are also crucial for (−)RNA synthesis: 5'BS in the 5'UTR, a TA element in the coding region of MP, and a Y-shaped RNA element (YRE) in the 3'UTR. Because RNA2 utilizes a replication protein supplied from RNA1, these additional *cis*-acting RNA elements probably contribute to recruiting viral replication proteins. Indeed, YRE is an essential *cis*-acting RNA element that is recognized by the p27 replication protein. The p27-YRE interaction is required for the recruitment of RNA2 to the sites of viral replication. In contrast, RNA1 has no RNA element that interacts with p27 and/or p88^{pol} supplied in *trans*, and p88^{pol} is only able to act *in cis* on RNA1 during replication. This *cis*-preferential requirement for p88^{pol} in the replication of RNA1 may be explained by the finding that p88^{pol} appears to bind to the 3'UTR of RNA1 in a translation-coupled manner. The translation-coupled binding of p88^{pol} may enable rigorous template selection of RNA1 without special *cis*-acting elements such as YRE.

Host Factors in Dianthovirus RNA Replication

Accumulating evidence has suggested that interactions between viral proteins and host factors are important determinants of successful viral infection. Recent proteomics approaches followed by functional studies have suggested that RCNMV hijacks a number of host proteins for its robust replication.

Heat shock protein 70 (Hsp70) and Hsp90 are abundant and highly conserved molecular chaperones that play central roles in protein homeostasis, including the folding of newly synthesized proteins and the assembly of macromolecular protein complexes. Like many other (+)RNA viruses, RCNMV requires these molecular chaperones for replication. p27 binds to Hsp70 and Hsp90 *in vitro* and *in vivo*. Hsp70 regulates the assembly of the 480 kDa viral replicase complex by preventing its aggregation, while Hsp90 promotes the p27–YRE interaction. Thus, RCNMV utilizes Hsp70 and Hsp90 to regulate different steps in RNA replication before starting (–)RNA synthesis.

One of the hallmark features of eukaryotic (+)RNA virus replication is the establishment of viral replication compartments, which are generated *de novo* during virus infection via massive rearrangement of host membranes and lipid metabolism. A host small GTPase ADP ribosylation factor 1 (Arf1) is a central regulator of the biogenesis of the COPI vesicle that mediates Golgi-to-ER retrograde vesicle trafficking. Arf1 is critical for RCNMV replication. p27 directly interacts with and recruits Arf1 from the Golgi to the viral replication compartment; in fact, RCNMV seems to hijack Arf1 function to establish the viral replication compartment. Another important host factor for RCNMV replication is phospholipase D (PLD). PLD mediates the biogenesis of phosphatidic acid (PA) through hydrolysis of phosphatidylethanolamine or phosphatidylcholine. The interaction of p27 with PA was demonstrated by a protein–lipid overlay assay, and RCNMV was shown to induce an approximately threefold accumulation of PA in infected plant leaves. p88^{pol} interacts with and recruits PLD α and PLD β of *Nicotiana benthamiana* to the viral replication compartment during RCNMV replication. PLD-derived PA is required for RCNMV replication, and exogenously added PA was shown to stimulate RNA replication in protoplasts and (–)RNA synthesis in plant-derived cell-free extracts. Therefore, RCNMV is likely to induce enrichment of PA at the viral replication compartment via PLD to establish a suitable environment for viral RNA replication.

It is known that many plant viruses induce production of reactive oxygen species (ROS) during infection. ROS are often associated with host defense against biotrophic pathogens such as fungi and bacteria. However, the role of ROS in plant–virus interactions has been obscure. RCNMV strongly induces intracellular ROS production in infected cells. Respiratory burst oxidase homolog (RBOH) is one of the major host-plant enzymes responsible for the production of ROS in response to various types of stresses. Interestingly, RCNMV replication protein p27 interacts with the RBOHB of *N. benthamiana* and recruits it to the viral replication compartment. In addition, p27 is sufficient for induction of ROS production in an RBOH-dependent manner and gene silencing or pharmacological inhibition of RBOH activity compromises viral replication. These findings suggest that RCNMV hijacks a host ROS-generating enzyme to boost viral replication. RCNMV appears to activate RBOH to mediate virus-induced intracellular ROS production. A subgroup II calcium-dependent protein kinase (CDPKiso2) of *N. benthamiana* phosphorylates the N-terminal region of RBOHB and activates its enzyme activity. CDPKiso2 is required for p27-triggered ROS production and viral RNA replication. Although p27 can form a complex with CDPKiso2, this interaction seems to be indirect. Instead, a host scaffold protein, receptor for activated C kinase 1 (RACK1), bridges the interaction of p27 with CDPKiso2. Indeed, RACK1 is required for the p27–CDPKiso2 interaction, p27-mediated ROS production, and viral RNA replication. Together, these findings suggest that RCNMV co-opts RACK1 to integrate CDPKiso2 into the viral replication compartment to promote RBOH activation and ROS production.

Cell-to-Cell and Systemic Movement

A 35 kDa MP encoded by RNA2 is essential for the cell-to-cell and systemic movement of dianthoviruses. The amino acid sequences of the MPs of dianthoviruses suggest that they belong to the 30K superfamily. RCNMV MP has the ability to bind single-stranded nucleic acids cooperatively. However, unlike the MP of tobacco mosaic virus (TMV), which also belongs to the 30K superfamily, RCNMV MP does not unfold single-stranded RNA molecules *in vitro*. Despite having such different characteristics, the MPs of TMV and RCNMV are functionally homologous because RCNMV MP can support the systemic infection of a movement-deficient mutant of TMV in *N. benthamiana*. Although it has not been determined whether dianthoviruses move cell-to-cell as virions or as RNA–MP complexes, the former is unlikely because the CP is dispensable for the cell-to-cell movement of dianthoviruses in several host plants.

RCNMV MP localizes at plasmodesmata (PD) and increases their size exclusion limit. In addition, MP also localizes to small punctate structures at the cortical ER that contain the VRCs during the early phase of viral infection. Colocalization of MP to the cortical VRCs is associated with the replication of RNA1, but not that of RNA2, and is required for the cell-to-cell movement of RCNMV. Glyceraldehyde 3-phosphate dehydrogenase A is involved in this colocalization event and is required for the efficient cell-to-cell movement of RCNMV in *N. benthamiana*.

Although a CP-deficient mutant of RCNMV is able to systemically infect *N. benthamiana* at lower temperatures such as 17°C, virion formation is required but not sufficient for the systemic movement of dianthoviruses.

RNA Silencing Suppression

The importance of RNA silencing mediated by small RNAs in the defense of various organisms, including plants, against viruses has been well established. To counteract antiviral RNA silencing, most plant viruses encode at least one viral silencing suppressor protein (VSR). RCNMV has at least two strategies to counteract host RNA silencing. Suppression of RNA silencing by RCNMV

closely correlates with viral (–)RNA synthesis and accumulation of the 480 kDa replicase complex, implying that RCNMV interferes with host RNA silencing in an RNA replication-associated manner. RCNMV replication also inhibits microRNA biogenesis that is catalyzed by the host RNaseIII-like enzyme dicer-like 1 (DCL1). Higher viral accumulation was observed during RCNMV infection in a *dcl1* mutant *Arabidopsis*. However, the role of DCL1 or its homologs in suppression of RNA silencing during RCNMV replication remains unknown. In addition to replication-associated RNA-silencing suppression, it has been reported that RCNMV MP has VSR activity. The functional domains required for VSR activity in MP differ from those required for cell-to-cell movement.

Further Reading

- Cao, J., Guenther, R.H., Sit, T.L., *et al.*, 2015. Development of abamectin loaded plant virus nanoparticles for efficacious plant parasitic nematode control. *ACS Applied Materials & Interfaces* 7, 9546–9553.
- Cao, J., Guenther, R.H., Sit, T.L., *et al.*, 2014. Loading and release mechanism of red clover necrotic mosaic virus derived plant viral nanoparticles for drug delivery of doxorubicin. *Small* 10, 5126–5136.
- Heinlein, M., 2015. Plant virus replication and movement. *Virology* 479–480, 657–671.
- Hiruki, C., 1987. The dianthoviruses: A distinct group of isometric plant viruses with bipartite genome. *Advances in Virus Research* 33, 257–300.
- Hyodo, K., Hashimoto, K., Kuchitsu, K., Suzuki, N., Okuno, T., 2017. Harnessing host ROS-generating machinery for the robust genome replication of a plant RNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 114, E1282–E1290.
- Hyodo, K., Nagai, H., Okuno, T., 2017. Dual function of a cis-acting RNA element that acts as a replication enhancer and a translation repressor in a plant positive-stranded RNA virus. *Virology* 512, 74–82.
- Hyodo, K., Suzuki, N., Okuno, T., 2019. Hijacking a host scaffold protein, RACK1, for replication of a plant RNA virus. *New Phytologist* 221, 935–945.
- Hyodo, K., Taniguchi, T., Manabe, Y., *et al.*, 2015. Phosphatidic acid produced by phospholipase D promotes RNA replication of a plant RNA virus. *PLoS Pathogens* 11, e1004909.
- Kaido, M., Abe, K., Mine, A., *et al.*, 2014. GAPDH – A recruits a plant virus movement protein to cortical virus replication complexes to facilitate viral cell-to-cell movement. *PLoS Pathogens* 10, e1004505.
- Kaido, M., Nagano, H., Omote, K., *et al.*, 2019. 5′ – Terminal stem-loop of carnation ringspot virus RNA1 is required for the efficient amplification of viral RNAs. *Virus Research* 265, 138–142.
- Okuno, T., Hiruki, C., 2013. Molecular biology and epidemiology of dianthoviruses. *Advances in Virus Research* 87, 37–74.
- Steckelberg, A.L., Akiyama, B.M., Costantino, D.A., *et al.*, 2018. A folded viral noncoding RNA blocks host cell exoribonucleases through a conformationally dynamic RNA structure. *Proceedings of the National Academy of Sciences of the United States of America* 115, 6404–6409.
- Tajima, Y., Iwakawa, H.O., Hyodo, K., *et al.*, 2017. Requirement for eukaryotic translation initiation factors in cap-independent translation differs between bipartite genomic RNAs of red clover necrotic mosaic virus. *Virology* 509, 152–158.
- Wang, A., 2015. Dissecting the molecular network of virus-plant interactions: The complex roles of host factors. *Annual Review of Phytopathology* 53, 45–66.

Endornaviruses (*Endornaviridae*)

Toshiyuki Fukuhara, Tokyo University of Agriculture and Technology, Fuchu, Japan

© 2021 Elsevier Ltd. All rights reserved.

This is an update of T. Fukuhara, H. Moriyama, Endornavirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00580-X](https://doi.org/10.1016/B978-012374410-4.00580-X).

Glossary

Cytoplasmic male sterility Male sterility caused by a cytoplasmic factor.

Endophyte A bacterium or fungus that lives within a plant for at least part of its life cycle without causing any apparent disease.

Horizontal transmission The transmission of infections (e.g., viruses) between members of the same species.

in silico An expression meaning “performed on computer” in reference to biological experiments.

Isogenic line Genetically identical pure-breeding group of individuals.

Mycorrhiza A symbiotic association between a green plant and a fungus.

Mycovirus A virus that infects fungal species.

Oomycete A fungus-like eukaryotic microorganism that is a filamentous, microscopic, and absorptive organism.

Open reading frame The part of a reading frame that has the ability to be translated into a protein.

Phytopathogen An organism that is pathogenic to plants.

Polyprotein A protein (produced by viruses) that is cleaved to produce a number of functional proteins.

Positive-sense, single-stranded RNA virus A virus that uses positive sense, single-stranded RNA as its genome.

RNA silencing The mechanism of transcriptional and post-transcriptional gene silencing mediated by 21–24 nt small RNAs.

RNA-dependent RNA polymerase An enzyme that synthesizes RNA from an RNA template.

Vertical transmission The transmission of parasites (e.g., virus) or symbionts from parents to children.

Introduction

Viruses are thought to affect the phenotype of their hosts, and plants infected with viruses usually exhibit visible disease symptoms. Thus, infection of crop plants and vegetables with viruses severely damages their yield and quality. Therefore, commercially available crops, vegetables, and fruits are believed to be virus-free, but they are often asymptotically infected with persistent viruses. As these plants are infected with persistent viruses and exhibit no symptoms, they are sold commercially.

Double-stranded RNAs (dsRNAs) have frequently been detected from various healthy plants. These dsRNAs are not transcribed from the host genome DNAs, and replicate (propagate) independent of their host genome DNAs as bacterial plasmids do. These dsRNA replicons have common properties that differ from those of conventional RNA viruses: (1) Most of these dsRNAs have no obvious effect on the phenotype of their host plants; (2) they are distributed throughout all tissues of the host plants at a stable low concentration; (3) they are transmitted to the next generation only through the gametes; (4) their horizontal transfer to other plants has never been proven. The size of these dsRNAs varies from 1.5 to 20 kbp.

Smaller dsRNAs (about 2.0 kbp) are often found with virus-like particles, and some of these dsRNAs have been classified as viruses in the family *Partitiviridae*. Partitiviruses have two unrelated linear dsRNA segments, each about 2.0 kbp in size, which encode a viral RNA-dependent RNA polymerase (replicase, RdRp) and a capsid (coat) protein (CP). These partitiviruses are typical persistent viruses that asymptotically infect crop plants and vegetables.

Larger dsRNAs (more than 10 kbp) are probably not associated with distinct virus-like particles, because no virus-like particles have been isolated in preparations obtained with various purification procedures. Thus, these large dsRNAs were previously referred to as RNA plasmids, enigmatic dsRNAs or endogenous dsRNAs. However, now these large dsRNAs are also recognized as persistent viruses, and named endornaviruses in the genus *Alphaendornavirus* of the family *Endornaviridae*.

Acute virus infection severely damages the yield and quality of crop plants and vegetables, so phytopathologic studies aimed at protecting crop plants and vegetables from virus infections have been conducted, and many acute plant viruses have been extensively studied. In contrast, because persistent viruses asymptotically infect host plants, studies on persistent plant viruses have been more limited than those on acute plant viruses. A typical persistent RNA viruses, endornaviruses, are the focus of this article.

Plant Endornaviruses

Linear dsRNAs of approximately 14 kbp in length were often found in rice (*Oryza sativa*), barley (*Hordeum vulgare*), melon (*Cucumis melo*), and bell pepper (*Capsicum annuum*). These dsRNAs were not transcribed from their host DNA genomes. Furthermore they propagate independent of their host genomes but have several common properties that are different from those of conventional viruses. Complete nucleotide sequences of these dsRNAs isolated from broad bean (*Vicia faba*), cultivated rice (*O. sativa*) and wild rice

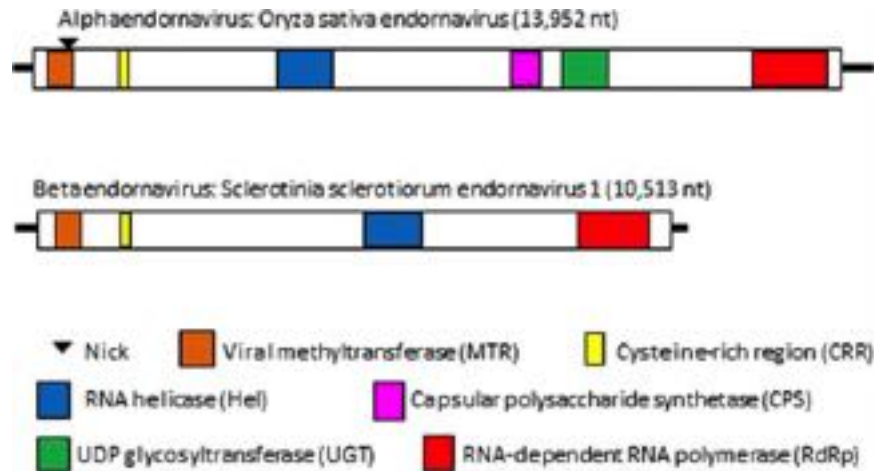


Fig. 1 Genome organizations of *Oryza sativa* endornavirus (OsEV), the type species of *Alphaendornavirus*, and *Sclerotinia sclerotiorum* endornavirus 1 (SsEV1), the type species of *Betaendornavirus*. The site-specific nick is only present in the coding strand of alphaendornaviruses.

(*O. rufipogon*) were reported about two decades ago. These dsRNAs encoded a single open reading frame (ORF) of approximately 4600 amino acid (aa) residues, in which the conserved viral RdRp and RNA helicase (Hel) domains were found (Fig. 1). Phylogenetic analyses of these two conserved domains indicated that these dsRNAs share a common ancestor with the alpha-like supergroup of positive-sense single-stranded RNA (ssRNA) viruses including many plant ssRNA viruses, such as Tobacco mosaic virus (TMV) and Cucumber mosaic virus (CMV). Consequently, the International Committee on Taxonomy of Viruses (ICTV) classified these dsRNAs as members of a new virus genus *Endornavirus* within a new virus family *Endornaviridae*, and named the three dsRNAs as *Vicia faba* endornavirus (VfEV), *Oryza sativa* endornavirus (OsEV) and *Oryza rufipogon* endornavirus (OrEV).

Deep-sequencing (RNA-seq) analyses have revealed asymptomatic infections of various viruses in crop plants and vegetables as well as wild plants. Some of these virus sequences have been recognized as members of the family *Endornaviridae*. To date, 14 endornaviruses have been found in plants, including barley, melon, bell and hot peppers (*C. annuum*), bottle gourd (*Lagenaria siceraria*), common bean (*Phaseolus vulgaris*), avocado (*Persea americana*) and others (Table 1).

The eighth and ninth reports of the ICTV classified the family *Endornaviridae* in the dsRNA virus group. However, an alternative opinion based on phylogenetic analyses claims that the family *Endornaviridae* should be classified in the positive-sense ssRNA virus group, and large dsRNAs of endornaviruses should be considered as their replication intermediates. Consequently, the tenth report of the ICTV, published recently, classifies the family *Endornaviridae* in the positive-sense ssRNA virus group.

Fungal Endornaviruses

A wide variety of viruses have been found in many fungi, including phytopathogenic and endophytic fungi and mushrooms. Fungal viruses are termed mycoviruses, and most mycoviruses contain dsRNAs as their genome. dsRNA mycoviruses are primarily classified into four families, *Partitiviridae*, *Totiviridae*, *Chrysoviridae*, and *Endornaviridae*, but novel mycoviruses recently discovered should be classified into a new genus and/or family. In general, mycoviruses persistently infect the host fungus and are transmitted vertically via host cell division, and infections are usually symptomless.

The violet root rot fungus, *Helicobasidium mompa*, occurs on various plants, and contains various sized dsRNAs. A large dsRNA (L1 dsRNA) in the V670 strain of *H. mompa* has been identified as a hypovirulence factor. The nucleotide sequence of this dsRNA encodes a single long ORF containing conserved RdRp and Hel domains exhibiting significant similarity to those encoded by plant endornaviruses described above, indicating that this dsRNA is a member of the family *Endornaviridae*. The virus was subsequently named *Helicobasidium mompa* endornavirus 1 (HmEV1). An increasing number of recent reports have described nucleotide sequences that likely belong to viruses of the family *Endornaviridae*, especially from plant pathogenic fungi.

Currently, the family *Endornaviridae* consists of two genera, *Alphaendornavirus* and *Betaendornavirus*, which are classified based on genome size and the presence of unique protein domains. Seven fungal endornaviruses, which have shorter genomes (< 12.6 kbp) than alphaendornaviruses and lack the UDP (uridine diphosphate) glucosyltransferase (UGT) domain present in the majority of alphaendornaviruses, have been classified as members of the genus *Betaendornavirus* (Fig. 1). To date, the genus *Alphaendornavirus* includes 20 viruses that infect plants, fungi, and oomycetes, whereas the genus *Betaendornavirus* includes seven viruses that infect ascomycete fungi (Table 1, Fig. 2).

Eight novel endornavirus-like sequences were recently identified from isolates of mycorrhizal (*Ceratobasidium*) fungi isolated from pelotons within root cortical cells of wild indigenous orchid species in Western Australia. The partial and complete sequences of viral RdRps shared low (9%–30%) identities with one another and with those of endornavirus RdRp described previously. Four putative viruses had longer genomes (> 20 kbp) than those of endornaviruses, and three had two ORFs within their genomes.

Table 1 List of species in the family Endornaviridae

Host type	Host species	Species name	Abbreviation
<i>Alphaendornavirus</i>			
Plant	Malabar spinach	<i>Basella alba endornavirus 1</i>	BaEV1
	Bell pepper	<i>Bell pepper endornavirus</i>	BPEV
	Melon	<i>Cucumis melo endornavirus</i>	CmEV
	Barley	<i>Hordeum vulgare endornavirus</i>	HvEV
	Hot pepper	<i>Hot pepper endornavirus</i>	HPEV
	Bottle gourd	<i>Lagenaria siceraria endornavirus</i>	LsEV
	Wild rice (W1714)	<i>Oryza rufipogon endornavirus</i>	OrEV
	Cultivated rice	<i>Oryza sativa endornavirus</i> ^a	OsEV
	Avocado	<i>Persea americana endornavirus 1</i>	PaEV1
	Common bean	<i>Phaseolus vulgaris endornavirus 1</i>	PvEV1
	Common bean	<i>Phaseolus vulgaris endornavirus 2</i>	PvEV2
	Common bean	<i>Phaseolus vulgaris endornavirus 3</i>	PvEV3
	Broad bean (447)	<i>Vicia faba endornavirus</i>	VfEV
	Winged bean	<i>Winged bean endornavirus</i>	WBEV1
	Mate	<i>Yerba mate endornavirus</i>	YmEV
Fungus	Powdery mildew fungus	<i>Erysiphe cichoracearum endornavirus</i>	EcEV
	Endophyte	<i>Grapevine endophyte endornavirus</i>	GEEV
	White root rot fungus	<i>Helicobasidium mompa endornavirus 1</i>	HmEV1
	Rhizoctonia	<i>Rhizoctonia cerealis endornavirus 1</i>	RcEV1
Protista	Phytophthora	<i>Phytophthora endornavirus 1</i>	PEV1
<i>Betaendornavirus</i>			
Fungus	Alternaria	<i>Alternaria brassicola endornavirus 1</i>	AbEV1
	Gray mold fungus	<i>Botrytis cinerea endornavirus 1</i>	BcEV1
	Gremmeniella	<i>Gremmeniella abietina endornavirus 1</i>	GaEV1
	Soil-borne fungus	<i>Rosellinia necatrix endornavirus 1</i>	RnEV1
	Sclerotinia blight fungus	<i>Sclerotinia minor endornavirus 1</i>	SmEV1
	White mold fungus	<i>Sclerotinia sclerotiorum endornavirus 1</i> ^a	SsEV1
	Truffle	<i>Tuber aestivum endornavirus</i>	TaEV

^aThe type species.

These two features of novel endorna-like viruses are unlike endornaviruses currently described, and they may challenge current taxonomic criteria for membership of the family Endornaviridae.

Endornaviruses From Other Eukaryotic Kingdoms

Plant pathogens of the genus *Phytophthora* have many of the same biological properties as fungi. However, on the basis of sequence similarities, plant pathogens of the genus *Phytophthora*, an oomycete, have been classified together with diatoms and brown algae into a protist group known as the Stramenopiles. Sequencing and phylogenetic analyses of a large dsRNA from a *Phytophthora* isolate from Douglas fir (*Pseudotsuga menziesii*) revealed that it encodes a single long ORF similar to those encoded by known plant and fungal endornaviruses. This was recognized as a member of the genus *Alphaendornavirus* and named *Phytophthora endornavirus 1* (PEV1).

Using an *in silico* bioinformatics approach to obtain full or partial cDNA sequences of genes and comparing them against known viral sequences in the NCBI (National Center for Biotechnology Information) Expressed Sequence Tag database, 119 novel virus-like sequences related to members of the families Partitiviridae, Totiviridae, Chrysoviridae, and Endornaviridae, were discovered. Among them one endornavirus-like sequence that exhibited significant similarity to the fungal endornavirus HmEV1 was identified from the sea louse (*Caligus rogercresseyi*). As the sea louse is a member of the phylum Arthropoda, this finding suggests that endornaviruses are also found in the animal kingdom.

Genome Organization

A recent ICTV report claimed, based on phylogenetic analyses, that endornaviruses are classified in the positive-sense ssRNA virus group, and that large dsRNAs should be recognized as their replication intermediates. However, ssRNA genomes of endornaviruses have never been isolated. Consequently, all cDNA cloning experiments were performed by using purified dsRNAs of

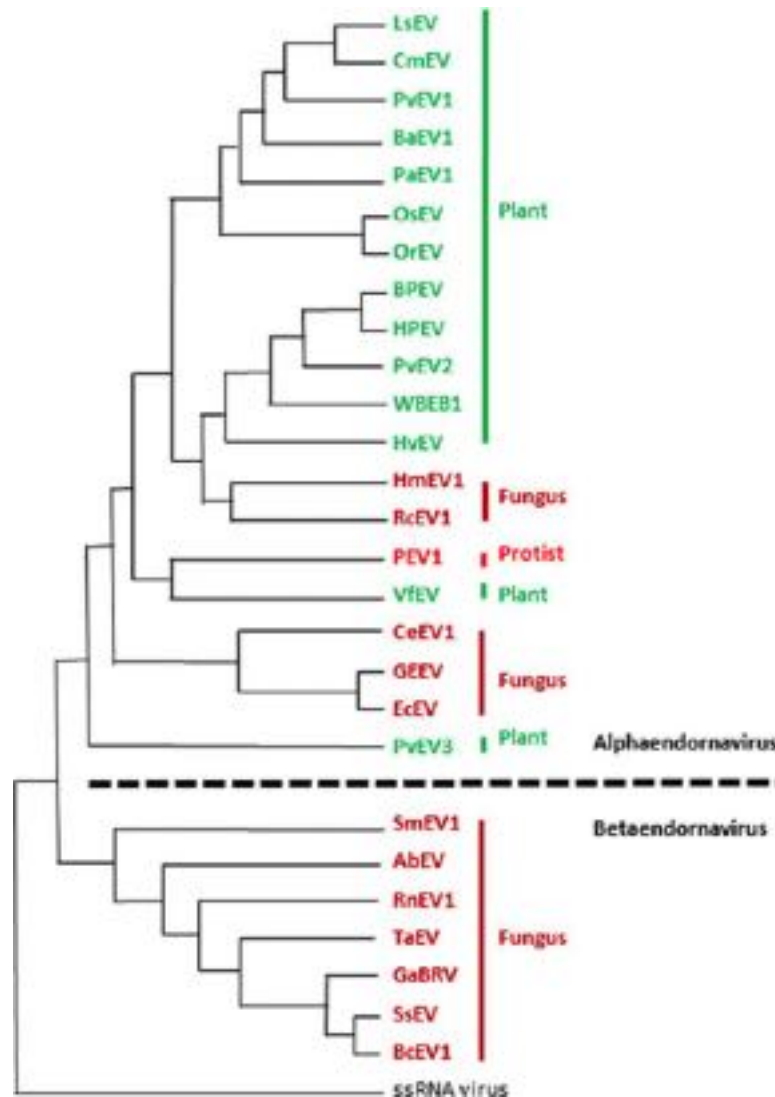


Fig. 2 Phylogenetic tree of selected endornaviruses, as modified from a tree constructed by Valverde *et al.* using the aa sequences of RdRp domains of endornaviruses. Abbreviations of virus names are indicated in Table 1. Reproduced from Valverde, R.A., Khalifa, M.E., Okada, R., *et al.*, 2019. ICTV virus taxonomy profile: Endornaviridae. Journal of General Virology 100 (8), 1204–1205.

endornaviruses as templates for cDNA synthesis. In this case, most endornavirus sequences were obtained from nucleotide sequences of cDNA clones. Recently, nucleotide sequences of endornaviruses have been obtained more easily from RNA-seq data of endornavirus-infected organisms.

The endornavirus genomes sequenced to date range from 9639 to 17,635 bp in length, and electron microscopic observations revealed that the dsRNAs of endornaviruses are linear (Fig. 3). They commonly encode a single ORF of 3148 to 5821 aa residues, and this long ORF is one of the unique molecular features of endornaviruses (Fig. 1). Conserved Hel and RdRp domains located in the central and C-terminal regions, respectively, of the long ORF are common features of all endornaviruses described to date (Fig. 1). Other conserved domains found in some, but not all, endornaviruses are viral methyltransferase (MTR), UGT, capsular polysaccharide synthetase (CPS), and cysteine-rich region (CRR) (Fig. 1).

Four plant endornaviruses [OsEV, OrEV, VfeEV and Bell pepper endornavirus (BPEV)], one fungal endornavirus (HmEV1), and one oomycete endornavirus (PEV1) belonging to the genus *Alphaendornavirus* were shown to contain a site-specific nick in the 5' region of the coding strand, which divides not only the coding strand but also the single long ORF (Fig. 1), but the nick has never been found in betaendornaviruses. The biological implications of this nick in the coding strand and the mechanism by which it is generated are unknown. However, because the divided coding strand can no longer be used as a template for non-coding strand synthesis (replication) or mRNA for translation of the protein, the nick must affect at least two important steps in the life cycle of endornaviruses. Therefore, endornavirus dsRNAs containing the nick may be remains (dead virus bodies) because they have a defect in translation and replication, even though these unique dsRNA molecules predominantly accumulate in the hosts infected by some alphaendornaviruses described above. The continuous ssRNA of the coding strand may play a vital role in the virus life cycle, but

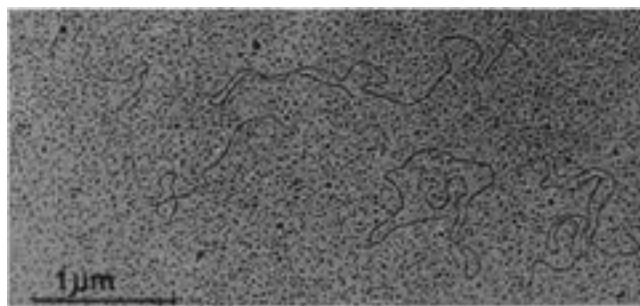


Fig. 3 Electron micrograph of dsRNA molecules of *Oryza sativa* endornavirus (OsEV). Bar represents 1 μm . Figure reproduced with permission from Fukuhara, T., Moriyama, H., Pak, J.K., Hyakutake, T., Nitta, T., 1993. Enigmatic double-stranded RNA in Japonica rice. *Plant Molecular Biology* 21, 1121–1130.

there are no reports describing the isolation of ssRNA molecules of endonaviruses. These two molecular features – the single long ORF and the site-specific nick – have not been reported in other known RNA viruses.

Gene Products

All endonaviruses reported to date encode a single long ORF of approximately 3100 to 5800 aa residues, in which several conserved protein domains are found. This suggests that this long ORF encodes a polyprotein that is cleaved to produce a number of functional proteins. Indeed, potyviruses, which have a single long ORF of approximately 3000 aa residues, encode a single polyprotein consisting of eight functional proteins (see Potato Virus Y, Potyvirus). Therefore, endonaviruses likely encode a polyprotein containing more than 10 functional proteins. Indeed, several conserved protein domains have been found in all endonaviruses (Fig. 1), as discussed below. However, these putative proteins have never been experimentally detected by a specific antibody using a technique such as Western blot.

Conserved motifs encoding RdRp and Hel are found in the same regions of long ORFs encoded by most endonaviruses (Fig. 1), and these two proteins are structurally similar to those encoded by ssRNA viruses such as CMV and TMV. Thus, these putative RdRp and Hel probably function in the replication of the genome RNA, as their homologues encoded by ssRNA viruses are essential for the replication of their genomes. MTR (a capping enzyme for viral genomic RNAs) also probably functions in the replication of endonaviruses, similar to their homologues encoded by ssRNA viruses, but the conserved MTR motif is not found in some endonaviruses.

The enzyme UGT is commonly found in eukaryotes. This enzyme catalyzes the addition of the glycosyl group from a UDP-sugar to a small hydrophobic molecule. UGTs have been identified in dsDNA viruses, including baculoviruses and nucleopolyhedroviruses from insects. In insect cells, the viral ecdysteroid UDP-glycosyltransferase inactivates the molting hormones (ecdysteroids) of the host insect by sugar conjugation. Although a UGT gene has never been found in the genomes of the alpha-like supergroup of ssRNA viruses, it was found in strains of two hypoviruses (fungal ssRNA viruses), *Cryphonectria hypovirus 3* and *4* (CHV3 and CHV4). A conserved motif significantly similar to that of UGT was found in some alphaendonaviruses (Fig. 1). A conserved motif similar to that of CPS is also found in some endonaviruses (Fig. 1). However, the functions of both putative UGT and CPS proteins in endonaviruses remain unknown.

The ORF of endonaviruses is very large (Fig. 1), so it likely encodes several other proteins in addition to those already discussed. The ORF must encode a polyprotein, so it must also encode one or more proteinases to cleave the putative polyprotein into functional units such as RdRp, Hel, MTR, UGT and CPS. A similar situation is well known in potyviruses, which encode a polyprotein of approximately 3000 aa residues containing two proteinases (see Potato Virus Y, Potyvirus). The conserved CRR located in the N-terminal half of endornavirus ORFs may function as a proteinase (Fig. 1), because many proteinases (also known as cysteine proteinases) containing a conserved cysteine residue in their catalytic domain have been reported so far.

CP and movement proteins (MP) are essential for conventional plant viruses such as TMV and CMV. CP is essential for viral horizontal transmission from one host plant to another, and MP is essential for the movement of viruses from one cell to surrounding (neighboring) cells and systemic infection in host plants. Systemic propagation (infection) within the host plant mediated by MP and horizontal transmission (infection) from one host to other hosts mediated by CP constitute a fundamental propagation strategy for acute plant viruses. In contrast, endonaviruses have a symbiotic relationship with the host plant. Therefore, both CP and MP are probably not necessary for persistent propagation of endonaviruses.

Replication (RdRp Activity)

A limited number of reports have demonstrated the functions (or enzymatic activities) of proteins encoded by endonaviruses. The RdRp activity of endonaviruses was initially reported from a study of VfEV in the “447” line of broad bean (*V. faba*).

Purified cytoplasmic membranous vesicles containing VfEV had RdRp activity, which required Mg^{2+} ions and the four nucleotide triphosphates, and was unaffected by inhibitors of cellular DNA-dependent RNA polymerases such as alpha-amanitin and actinomycin D. RdRp activity was also detected in the crude microsomal fraction of OsEV-harboring rice cells. Like the VfEV RdRp activity, this RdRp activity was also highest in the presence of all four nucleotide triphosphates and Mg^{2+} ions, and was resistant to inhibitors of DNA-dependent RNA polymerases. Treatment of the purified microsomal fraction with proteinase K plus deoxycholate suggested that the RdRp enzyme complex is located in vesicles. So far, RdRp activity has been found only in two endornavirus-harboring plants mentioned above. Therefore, the available data concerning endornavirus replication are limited. There is a need to identify and characterize the endornavirus- and/or host-encoded proteins that are involved in endornavirus replication.

Vertical Transmission

Although horizontal transmission of endornaviruses has never been reported in plants, these viruses are frequently found in a variety of plants. For example, OsEV and BPEV are commonly detected in many cultivars of *japonica* rice and bell pepper, respectively. To understand why OsEV is widely distributed in *japonica* rice cultivars, crossing experiments with OsEV-harboring and OsEV-free plants were carried out to examine the efficiency of vertical transmission of OsEV. OsEV-harboring and OsEV-free plants were found to coexist in the *japonica* rice cultivar Nipponbare, and these two isogenic lines could not be distinguished on the basis of their appearance. The results of crossing experiments indicated that the efficiency of OsEV transmission via pollen was more than 98%, and that of transmission via ovules was 100%. Consequently endornaviruses are efficiently transmitted to progeny plants via both pollen and ovules, though differential centrifugation and sucrose density-gradient centrifugation experiments revealed that OsEV localizes in the cytoplasm of host cells. The high efficiency of OsEV transmission via both pollen and ovules is likely responsible for the wide distribution of OsEV in many rice cultivars. Analyses of the F2 progeny plants of cv. Nipponbare indicated that the absence of OsEV was not associated with any particular gene(s) in the OsEV-free plants. The observed inheritance of OsEV was different from that of other cytoplasmic genetic elements (e.g., chloroplasts and mitochondria), which are usually inherited only via egg cells.

OsEV has been found in many cultivars of *japonica* rice but not in any cultivars of *indica* rice (a subspecies of *O. sativa*). *Japonica* and *indica* are distinguishable on the basis of their phenotype (e.g., grain shape) and genotype (e.g., the distribution of transposons in their genomes). To determine the reason why OsEV is not found in *indica* rice cultivars, reciprocal crosses between an OsEV-harboring *japonica* variety (cv. Nipponbare) and an OsEV-free *indica* variety (cv. IR26 or Kasalath) were performed. When cvs. IR26 and Nipponbare were used, efficient transmission of OsEV via ovule (93%) and pollen (89%) was observed. However, when cvs. Kasalath and Nipponbare were used, OsEV transmission efficiency to F1 progeny was 68% via ovule and 20% via pollen, and transmission to F2 progeny plants also followed a complicated, non-Mendelian inheritance pattern. These results suggest that OsEV is unstable in *indica* rice plants, which may lack one or more genes that are involved in the maintenance (replication) of OsEV.

Similar crossing experiments with endornavirus-harboring bell-pepper plants (cv. Yolo Wonder) and endornavirus-free plants (cv. Jalapeño M or Hungarian Wax) demonstrated that BPEV is also inherited maternally and paternally to the next generation. Thus, the propagation of OsEV and BPEV (and probably other plant endornaviruses as well) may depend on seed-mediated transmission, and endornaviruses may survive in cooperation with their host plants. No horizontal spread of plant endornaviruses has been observed in the field, and no potential vectors have been identified.

Horizontal Transmission

In general, fungal endornaviruses as well as other mycoviruses persistently infect the host fungus, and are transmitted vertically via host cell division. Recently, however, it has been reported that natural infections of a phytopathogenic soil-borne fungus, *Rosellinia necatrix*, with a novel endornavirus, *Rosellinia necatrix endornavirus 1* (RnEV1), can occur under experimental soil conditions. Furthermore, it has also been demonstrated that *Sclerotinia minor* endornavirus 1 (SmEV1), which is found in a phytopathogenic fungus *Sclerotinia minor*, was readily transmitted horizontally via hyphal contact to isolates of different vegetative compatibility groups of *S. minor*, and transmitted vertically as well. Therefore, some fungal endornaviruses are able to transmit horizontally as well as vertically.

Regulation of Copy Number

All plant endornaviruses reported so far are likely to be found in all tissues of host plants at a stable low concentration, suggesting that their propagation strategy is probably different from that of acute ssRNA viruses. However, studies on the copy number of endornaviruses are limited. OsEV in rice plants is the only report that studied the regulation of copy number of endornaviruses. A comparison of the relative amount of OsEV with genomic DNA in host rice plants of cv. Nipponbare showed that OsEV is maintained at approximately 100 copies per cell. No significant difference in OsEV concentration was found in seedlings, roots,

and mature leaves, indicating that a mechanism to regulate copy number may exist in host rice cells. In contrast, the copy number of OsEV in cultured rice cells was found to be about 10-fold higher than in the cells of seedlings. Although no changes in the copy number of OsEV was found in suspension cultures over the course of three years, copy number returned to the original low level in rice plants regenerated from cultured cells. Furthermore, the copy number of OsEV in pollen grains was found to be 50-fold higher than in seedlings.

The steady replication of plant endornaviruses before every host cell division (both mitosis and meiosis) and their efficient transmission to the next generation via ovules and pollen are essential for endornavirus propagation because horizontal transmission likely does not occur. The unregulated increase of endornavirus replication could cause disease or lead to the death of host plants, as occurs with acute virus infections. Conversely, a decrease in copy number could cause the virus to disappear from the host plant (germ cells). The mating of host plants provides an opportunity for endornavirus propagation, so an increase in copy number only in pollen grains is a reasonable strategy to ensure virus transmission, despite their cytoplasmic localization. Unlike almost all other viruses, endornaviruses are not able to transmit horizontally, but they are capable of vertical transmission.

Pathogenesis

Endornavirus-harboring crop plants and vegetables are commercially available, indicating that these plants are healthy and common cultivars. Current data also indicate that most endornaviruses found in plants, with the exception of VfEV, cause asymptomatic and persistent infections in the host plant. For instance, BPEV and OsEV have been frequently found in commercially available bell pepper fruits and rice grains, respectively. Consequently we usually consume these fruits and grains that contain endornaviruses.

In the model cultivar of cultivated rice, Nipponbare, molecular biological analyses identified OsEV-harboring and OsEV-free plants. However, these two isogenic lines did not exhibit distinguishable phenotypes, and even rice breeders and farmers were unable to distinguish them. If the OsEV-harboring line had a lower harvest yield than the OsEV-free line, breeders and/or farmers would recognize this phenotype and discard the lower-yielding line.

In commercial cultivars of bell pepper grown in the United States, an endornavirus was found at a prevalence rate of 100%. The overall appearance of the two bell pepper near-isogenic lines, BPEV-harboring and -free, was similar. However, the BPEV-free line had a significantly higher percentage of seed germination than the BPEV-harboring line. Plant height, number of fruits, and total fruit weight in plants of the BPEV-free line were higher than in plants of the BPEV-harboring line, but these differences were not statistically significant. Other characteristics between the two lines, such as, stem diameter, percentage of dry weight, fruit volume, chlorophyll, carotenoid, and anthocyanin content, were similar. These results suggest that BPEV appears to have a weak parasitic relationship with the host.

Recently, variations in physiological traits have been reported in eight lines of common bean (*P. vulgaris*) cv. Black Turtle Soup, four of which were double-infected with *Phaseolus vulgaris* endornavirus 1 (PvEV1) and 2 (PvEV2), and four of which were endornavirus-free. Plants from all eight lines were morphologically similar and did not show statistically significant differences in plant height, wet weight, number of days to flowering and pod formation, pods per plant, pod thickness, seed size, number of seeds per pod, and anthocyanin content. However, the endornavirus-infected lines had faster seed germination, longer radicles, lower chlorophyll content, higher carotene content, longer pods, and higher weight of 100 seeds, all of which were statistically significant.

Over four decades ago, a cytoplasmic male sterility (CMS) trait in the “447” line of broad bean (*V. faba*) and cytoplasmic spherical bodies (CSBs) found in the cytoplasm of 447 plants were reported. These CSBs contained a large dsRNA of approximately 17.6 kbp in length, and a correlation between the CMS trait and the presence of a CSB with a large dsRNA in the “447” strain was reported. This dsRNA was the genome of VfEV, which is the only plant endornavirus that affects the visible phenotype of the host.

In general, mycoviruses asymptotically infect host fungi. However, HmEV1, which was first reported as a fungal endornavirus, was identified as the hypovirulence factor from the strain V670 of the violet root rot fungus, *Helicobasidium mompa*. Recently, the endornavirus, SmEV1, was isolated from the hypovirulent strain LC22 of *S. minor*, and hypovirulence and associated traits of strain LC22 and SmEV1 were readily co-transmitted horizontally via hyphal contact to isolates of different vegetative compatibility groups of *S. minor*. Therefore, SmEV1 is also a hypovirulence-associated mycovirus with a wide spectrum of transmissibility and has the potential for biological control (virocontrol) of diseases caused by *S. minor*.

Therefore, most endornaviruses asymptotically infect their hosts. However, in combinations between specific endornaviruses and their hosts, endornaviruses are pathogenic to hosts and their infection affects the host phenotype.

Virus-Host Relationship (Host RNA Silencing Against Endornaviruses)

RNA silencing (RNA interference) is the process of sequence-specific post-transcriptional gene silencing triggered by dsRNAs. Most eukaryotes, such as fungi, insects, and plants, sense exogenous long dsRNAs, such as viruses, because replication intermediates of ssRNA viruses are long dsRNAs. They then activate the RNA silencing pathway for defense against virus infection. Extensive studies on the interactions between acute ssRNA viruses and host plants such as *Arabidopsis thaliana* revealed that RNA silencing functions as an innate defense system against virus infections in plants. In the RNA silencing pathway, a Dicer or Dicer-like (DCL) endoribonuclease cleaves long viral dsRNAs into viral small interfering RNAs (vsiRNAs), which then associate with an Argonaute

(AGO) protein. The vsiRNA-loaded AGO proteins then slice the viral genomic ssRNAs. However, many acute ssRNA viruses encode one or several viral suppressors of RNA silencing (VSR), which inhibits the host RNA silencing system, thus enabling the virus to overcome the host defense system and propagate systemically in the host plant.

Plant endornaviruses are likely maintained within the host plant for hundreds of generations, suggesting that the host plant regulates virus propagation and that some host factors control virus replication so that it is coordinated with host cell division. If host plants did not have such factors (proteins), those harboring endornaviruses would exhibit disease symptoms presumably due to unregulated virus propagation, or some of the plant's somatic or germ cells would lose endornaviruses. Therefore, host plants likely express factors that control endornaviruses as symbiont-like parasites. Candidate host factors for regulating endornaviruses could be proteins that constitute the RNA silencing machinery. Indeed endornavirus-derived vsiRNAs have been detected in host plants infected with OsEV, BPEV, PvEV1 and PvEV2, indicating that the host RNA silencing machinery recognizes endornaviral dsRNAs.

The copy number and inheritance of OsEV were examined in transgenic rice plants with a knock-down (KD) construct of genes for the RNA silencing machinery, *OsDCLs* or host RNA-dependent RNA polymerases. A low vertical transmission rate and the disappearance of OsEV during somatic cell divisions were observed in some *OsDCL2*-KD plants, suggesting that KD of a Dicer gene negatively affects the maintenance of OsEV in host plants. The host RNA silencing system may be necessary for persistent infection by endornaviruses because it provides stringent regulation of low copy number and efficient vertical transmission to the next generation. RNA silencing functions as a defense against acute virus infection, but it may also function as a host factor that maintains persistent infections of endornaviruses by keeping their copy number low.

Further Reading

- Boccardo, G., Lisa, V., Luisini, E., Milne, R.G., 1987. Cryptic plant viruses. *Advances in Virus Research* 32, 171–214.
- Brown, G.G., Finnegan, P.M., 1989. RNA plasmids. *International Review of Cytology* 117, 1–56.
- Dodds, J.A., Morris, T.J., Jordan, R.L., 1984. Plant viral double-stranded RNA. *Annual Review of Phytopathology* 22, 151–168.
- Fukuhara, T., 1999. Double-stranded RNA in rice. *Journal of Plant Research* 112, 131–138.
- Fukuhara, T., 2019. Endornaviruses: Persistent dsRNA viruses with symbiotic properties in diverse eukaryotes. *Virus Genes* 55, 165–173.
- Fukuhara, T., Gibbs, M.J., 2012. Family *Endornaviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. San Diego: Elsevier, pp. 519–521.
- Roossinck, M.J., 2018. Evolutionary and ecological links between plant and fungal viruses. *New Phytologist* 221, 86–92.
- Roossinck, M.J., Sabanadzovic, S., Okada, R., Valverde, R.A., 2011. The remarkable evolutionary history of endornaviruses. *Journal of General Virology* 92, 2674–2678.
- Tavantzis, S.M., 2001. *dsRNA Genetic Elements: Concepts and Applications in Agriculture, Forestry, and Medicine*. Boca Raton: CRC Press.
- Valverde, R.A., Khalifa, M.E., Okada, R., *et al.*, 2019. ICTV virus taxonomy profile: Endornaviridae. *Journal of General Virology* 100 (8), 1204–1205.

Fimoviruses (*Fimoviridae*)

Toufic Elbeaino and Michele Digiario, International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa	Amino acid(s)	MP	Movement protein
AAP	Acquisition access period	NGS	Next generation sequencing
CP	Coat protein	nt	Nucleotide(s)
DMB	Double-membraned body	ORF	Open reading frame
ELISA	Enzyme-linked immunosorbent assay	PCR	Polymerase chain reaction
ER	Endoplasmic reticulum	RdRP	RNA-dependent RNA polymerase
GP	Glycoprotein precursor	RNP	RNA ribonucleoprotein
HPD	High Plains disease	RRD	Rose rosette disease
kb	Kilobase	RT-PCR	Reverse transcription polymerase chain reaction
kDa	Kilo Dalton	ssRNA	Single-stranded RNA
		UTR	Untranslated terminal regions

Glossary

Double-membraned bodies Double-membraned proteinaceous structures enveloping the virus particles.

Introduction

Although they were associated with typical symptoms on their hosts, *i.e.*, chlorotic ringspot and mosaic (**Fig. 1**), the nature and identity of fimoviruses remained unknown for a long time. This was mainly due to the difficulty of isolating the fimoviruses by mechanical transmission onto herbaceous hosts and to the abnormal conformation of virus particles observed under the electron microscope, since their proteinaceous structures, known as double-membraned bodies (DMBs), was different from all other plant viruses known at the time.

Today, the family *Fimoviridae*, order *Bunyavirales*, includes only one genus (*Emaravirus*), which comprises nine species: *European mountain ash ringspot-associated emaravirus* (as a type species), *Actinidia chlorotic ringspot-associated emaravirus*, *Fig mosaic emaravirus*, *Pigeonpea sterility mosaic emaravirus 1*, *Pigeonpea sterility mosaic emaravirus 2*, *Redbud yellow ringspot-associated emaravirus*, *Raspberry leaf blotch emaravirus*, *Rose rosette emaravirus* and *High Plains wheat mosaic emaravirus*. It also includes several tentative emaravirus species: *Actinidia emaravirus 2* (AcEV-2), *Blackberry leaf mottle-associated virus* (BLMaV), *Camellia japonica associated emaravirus 1* (CjEV-1), *Camellia japonica associated emaravirus 2* (CjEV-2), *Jujube yellow mottle-associated virus* (JYMaV), *Lilac chlorotic ringspot-associated virus* (LiCRaV), *Palo verde broom virus* (PVBV), *Pistacia virus B* (PiVB), *Ti ringspot-associated virus* (TiRSaV), *Perilla mosaic virus* (PerMV) and *Woolly burdock yellow vein virus* (WBYVV).

All these viruses are characterized by a multipartite negative-sense ssRNA genome composed of four to eight segments (nine, in the case CjEV-1) contained within spherical virions. Most of these viruses are vectored in nature by eriophyid mites, which are known to have a highly specific relationship with their plant hosts, thus explaining the rather restricted host range that characterize them.

Classification

Fimoviruses are distantly related to orthospoviruses and orthobunyaviruses of the families *Tospoviridae* and *Peribunyaviridae*, respectively. They share with them the following common features: (1) a multipartite, negative-sense, ssRNA genome, which in the case of fimoviruses consists of four to eight segments; (2) a high sequence identity with orthologous proteins of members of the order *Bunyavirales* at equivalent genome positions in the first three RNAs (corresponding to L, M and S RNA segments); (3) the presence of five conserved motifs (A-E) in the amino acid sequence of their RNA-dependent RNA Polymerase (RdRP), similar to those in the L segment-encoded protein of members of the order *Bunyavirales*; (4) the virus particles are typically enveloped in membranes; (5) the presence of conserved nucleotide sequences at both the 5'- and 3'-termini of all genomic RNAs segments of fimoviruses, but not identical to those of other members of the order *Bunyavirales*.

Virion Structure

Fimoviruses are approximately spherical and enveloped virions with a diameter of 60–200 nm (**Fig. 2**). DMBs have an envelope consisting of a unit membrane approximately 12 nm thick, which is apparently acquired from the endoplasmic reticulum, containing proteinaceous

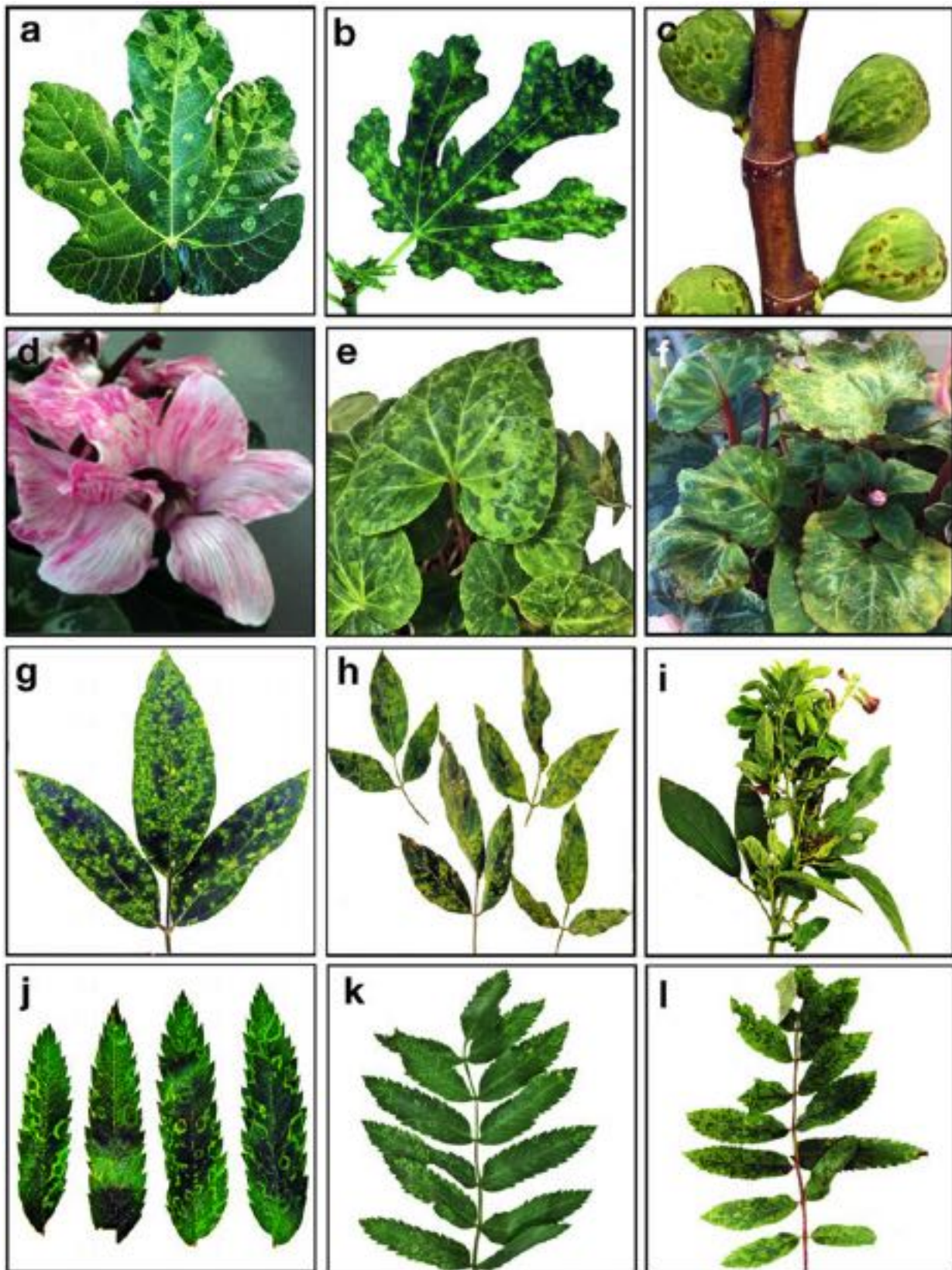


Fig. 1 Host symptoms caused by emaraviruses: FMV-infected plants showing chlorotic ringspots on leaves (a) and fruits (c), together with mosaic-like symptoms, *i.e.*, vein clearing and leaf mottling (b). FMV-infected cyclamen plants showing flower discoloration (d), mosaic (e) and leaf deformation (f). PPSMV-1-infected plants showing ringspot (g), chlorosis and mosaic (h), yellowing and sterility of flowers (i). EMARaV-infected plants showing ringspots (j), mottling (k) and mosaic (l) symptoms.

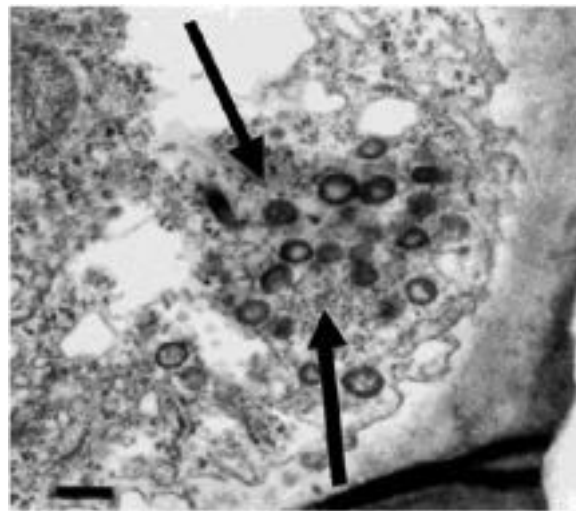


Fig. 2 Groups of double-membraned bodies (DMB) in the cytoplasm of cells from a fig seedling infected with FMV. Arrows point to the aggregates of convoluted electron-dense filamentous elements (DMBs). Bar = 200 nm.

material and fine fibrils. In ultrathin sections they are frequently found in association with amorphous electron-dense material. DMBs are present in all cell types except those of the vascular bundles and are rarely observed in bundle sheath parenchyma cells.

Genome

The virus genome comprises from four to eight segments of negative sense ssRNA ([Table 1](#)). Genomic RNAs are not capped or polyadenylated and all contain complementary sequences (13–20 nt, depending on the RNA segment) at their 5′- and 3′-termini conserved in all viral genomic RNA components. The lengths of 5′ and 3′ untranslated terminal regions (UTR) of each viral segment are very variable in the different species, ranging from 40 to 50 nt up to 600–700 nt in length. All emaraviruses have a similar genomic organization, and each RNA is composed of a single ORF in negative polarity ([Fig. 3](#)).

- (1) RNA-1 is 6981–7217 nt long and contains ORF1 that encodes a RdRP of a predicted molecular mass of 265–272 kDa (P1). In a delimited area of this ORF there are the five motifs also conserved in all bunyavirids, which correspond to the core polymerase module of the RdRP active site.
- (2) RNA-2 is 2054–2399 nt long and contains ORF2, encoding a glycoprotein precursor (GP) with a predicted molecular mass of 73–83 kDa (P2). P2 shares several conserved motifs, including one predicted cleavage site that may generate two larger glycoproteins, Gn (20–23 kDa) and Gc (52–53 kDa). A third smaller glycoprotein (Gs), consisting of 23 aa (2.6 kDa), has been predicted in the case of Actinidia chlorotic ringspot-associated virus (AcCRaV).
- (3) RNA-3 is 1220–1678 nt long and contains ORF3, encoding a nucleocapsid protein (NP) of 32–35 kDa (P3). In the case of High Plains wheat mosaic virus (HPWMoV), two RNA-3 consensus sequences encoding the NP were found, with 12.5% sequence divergence. This protein contains stretches of positively charged aa (BindN specificity, 80%) that are probably involved in RNA binding.
- (4) RNA-4 is 1342–1675 nt long (1154 nt in CjEV-2) and contains ORF4, encoding a putative movement protein (MP) of 38–43.6 kDa (P4), having structural elements resembling the consensus secondary structure of the 30K superfamily of plant virus MPs. In the case of European mountain ash ringspot-associated virus (EMARaV), the MP is encoded by the RNA-5 (# LR536382) and not by RNA-4 (# NC_013108).

More uncertainty surrounds the exact number and functionality of the other genomic RNAs that complete the emaravirus genome. Unlike the first four RNAs examined so far, the sequence homologies and the lengths of orthologs RNA-5 to RNA-8 in the different emaraviruses are much less significant. A ninth RNA (RNA-9) has been to date detected in the only CjEV-1. An exception is the high sequence identity (93%) shown in the RNA-6 of Pigeonpea sterility mosaic viruses 1 and 2 (PPSMV-1 and PPSMV-2), thus suggesting that the RNA-6 encoded protein is performing a conserved function in virus biology for both PPSMV-1 and 2. Instead of one ORF in all genomic RNAs of emaraviruses, unexpectedly, only RNA-6 of RRV has shown the presence of two different ORFs (7.4 kDa and 27.1 kDa).

Studies on the genetic variability within the population of different emaraviruses [Fig mosaic virus (FMV), Raspberry leaf blotch virus (RLBV), HPWMoV, Redbud yellow ringspot-associated virus (RYRSaV), PPSMV-1 and PPSMV-2] have shown significant differences in nucleotide sequences among isolates of different origins, which in the case of RNA-3 (NP) never exceeded 15%. This variability was generally lower than that found in RNA-1 and RNA-2 of FMV, whereas RNA-4 (MP) was more conserved than RNA-3 in the case of PPSMV-1 and PPSMV-2.

Table 1 List of members and tentative members of the Genus *Emaravirus* in the family *Fimoviridae*. Name, Acronym and Accession # of each genomic RNA is provided

Species names virus names	Acronym	Accession #							
		RNA-1	RNA-2	RNA-3	RNA-4	RNA-5	RNA-6	RNA-7	RNA-8
<i>Actinidia chlorotic ringspot-associated emaravirus</i>									
Actinidia chlorotic ringspot-associated virus	AcCRaV	NC038769	NC038770	NC038772	NC038771	NC038773			
<i>European mountain ash ringspot-associated emaravirus</i>									
European mountain ash ringspot-associated virus	EMARaV	NC013105	NC013106	NC013108	LR536382 ^a	NC013107 ^a	LR536383		
<i>Fig mosaic emaravirus</i>									
Fig mosaic virus	FMV	NC029562	NC029565	NC029563	NC029564	NC029566	NC029568		
<i>High Plains wheat mosaic emaravirus</i>									
High Plains wheat mosaic virus	HPWMoV	NC029570	NC029549	NC029550	NC029551	NC029552	NC029553	NC029554	NC029555
<i>Pigeonpea sterility mosaic emaravirus 1</i>									
Pigeonpea sterility mosaic virus 1	PPSMV-1	NC029575	NC029556	NC029557	NC029574	NC029569	KX363891		
<i>Pigeonpea sterility mosaic emaravirus 2</i>									
Pigeonpea sterility mosaic virus 2	PPSMV-2	NC030660	NC030662	NC030661	NC030663	NC030658	NC030659		
<i>Raspberry leaf blotch emaravirus</i>									
Raspberry leaf blotch virus	RLBV	NC029567	NC029558	NC029559	NC029560	NC029561	NC029571	NC029572	NC029573
<i>Redbud yellow ringspot-associated emaravirus</i>									
Redbud yellow ringspot-associated virus	RYRSaV	NC038852	NC038856	NC038854	NC038853	NC038855			
<i>Rose rosette emaravirus</i>									
Rose rosette virus	RRV	NC015298	NC015299	NC015300	NC015301	NC034979	NC034980	NC034981	
<i>Tentative Members</i>									
Actinidia emaravirus 2	AcEV-2	MK602171	MK602172	MK602173	MK602174	MK602175	MK602176		
Blackberry leaf mottle-associated virus	BLMaV	KY056657	KY056658	KY056659	KY056660	KY056661			
Camellia japonica associated emaravirus 1 ^b	CjEV-1	MN385573	MN385574	MN385575	MN385576	MN557024	MN557025	MN557026	MN557027
Camellia japonica associated emaravirus 2	CjEV-2	MN385577	MN385578	MN385579	MN385580				
Jujube yellow mottle-associated virus	JYMaV	MK305894	MK305895	MK305896	MK305897	MK305898	MK305899		
Lilac chlorotic ringspot-associated virus ^c	LiCRaV								
Palo verde broom virus	BPVBV	MF766024	MF766029	MF766034	MF766039				
Perilla mosaic virus	PerMV				LC430222 ^d				
Pistacia virus B	PiVB	MH727572	MH727573	MH727574	MH727575	MH727576	MH727578	MH727579	
Ti ringspot-associated virus	TiRSaV	MH223635	MH223636	MH223637	MH223638	MH223639			
Woolly burdock yellow vein virus (Arctium tomentosum virus)	WBYVV			JQ354894 ^d					

^aThe former RNA-4 of EMARaV reported in the GenBank is here reported as RNA-5 and *vice versa*.^bAn additional RNA segment (RNA 9, Accession #. MN557028) has been identified.^cNo sequences are available in GenBank.^dOnly the sequence of a partial RNA segment is available in GenBank.

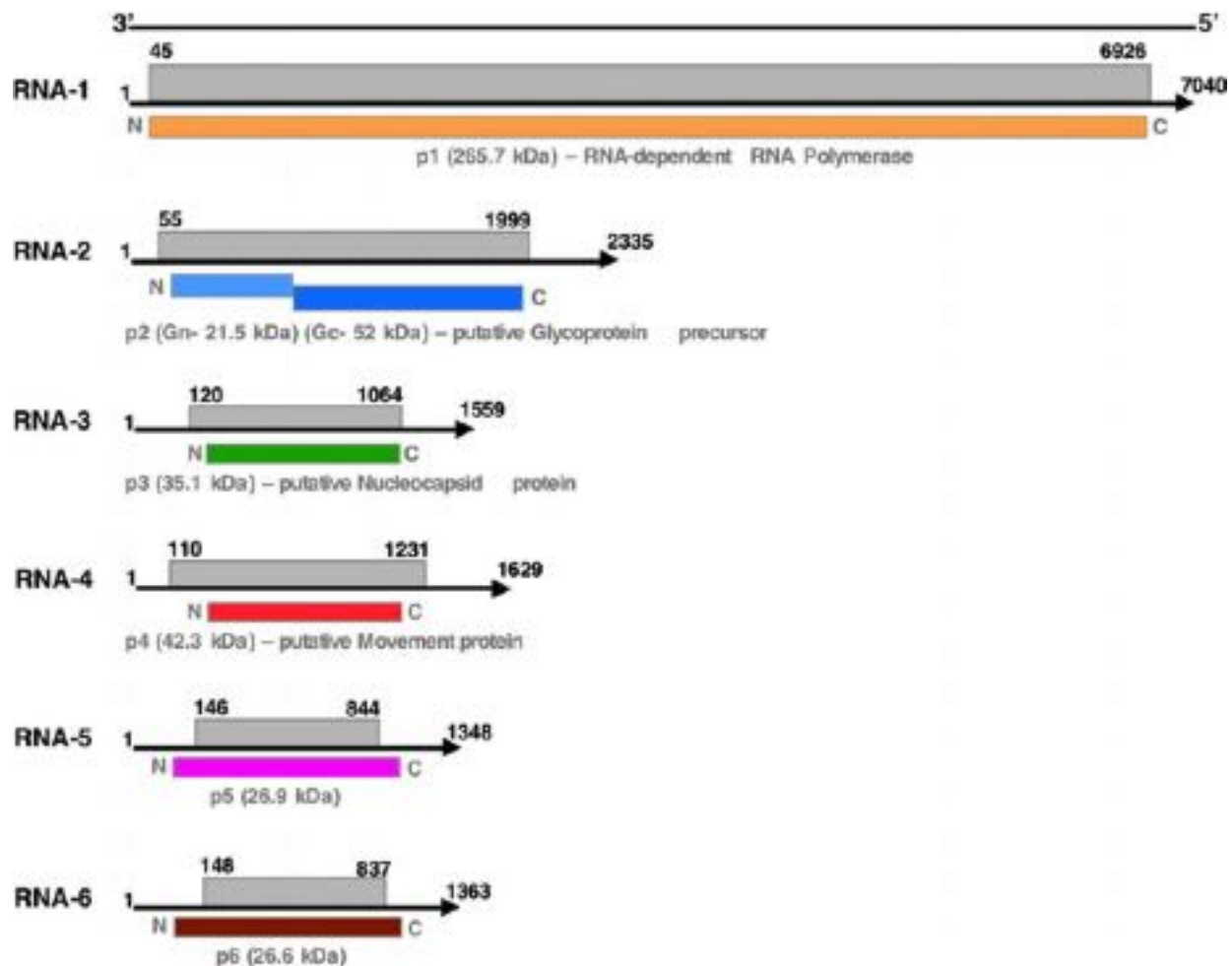


Fig. 3 Schematic representation of the organization of the six RNA segments constituting the genome of European mountain ash ringspot-associated virus (EMARaV); the type species of the genus *Emaravirus*. Expression products of each RNA (p1 to p6) are represented as dark gray boxes. The function and estimated molecular weight of each protein are reported below boxes. Arrows represent the virion-complementary sense RNA, from which the proteins are translated. Figure not drawn to scale.

In addition, interspecies recombination events were detected among PPSMV-1 and PPSMV-2 segments, two species commonly found in mixed infections in pigeonpea plants. Similar RNA segment reassortments were reported for FMV and BLMaV.

The phylogenetic trees inferred from the sequences of structural proteins (P1 to P4) of emaraviruses and putative emaraviruses in general displayed two clades (Fig. 4). The first group includes AcCRaV, AcEV-2, BLMaV, EMARaV, FMV, PiVB, PPSMV-1, PPSMV-2, RRV and RYRSaV, whereas the second group comprises HPWMoV, RLBV, PVBV, CjEV-1, CjEV-2, TiRSaV and JYMaV, among which the genomes of RLBV, HPWMoV and CjEV-1 have been reported to contain eight genomic RNA segments.

Life Cycle (Replication)

Like other negative sense ssRNA multipartite viruses, some fimoviruses are known to use “cap snatching” to initiate transcription and facilitate translation of their mRNAs, since, unlike the polymerases from non-segmented negative strand RNA viruses, they do not possess a capping activity. Cap-snatching involves binding of host capped mRNAs to the ribonucleoproteins (RNPs), cleavage of these RNAs close to the 5' cap by a viral endonuclease activity and use of the short-capped fragments as primers for viral mRNA transcription. As for other bunyavirids, the replication of fimoviruses probably occurs in the cytoplasm.

Epidemiology

Transmissibility by eriophyid mites is one of the most significant features shared by the different emaraviruses. The virus-eriophyid relationship is highly specific; it follows that only rarely the same viral species is transmitted by different eriophyid species and *vice versa*,

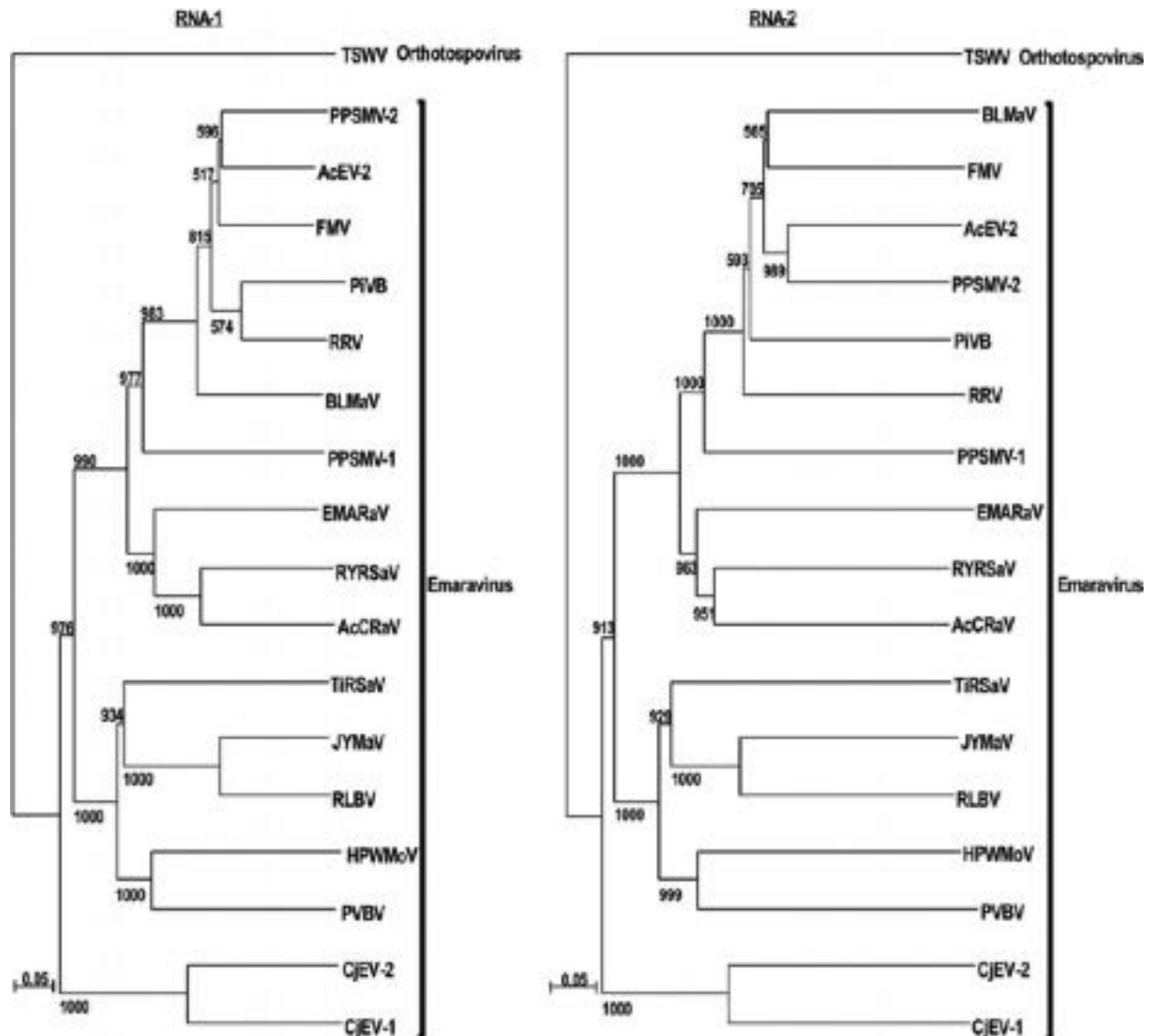


Fig. 4 Phylogenetic relationships among emaraviruses. Neighbor joining (NJ) phylogenetic trees generated on the deduced amino acid sequences of the putative RNA-dependent RNA polymerase (RdRP, RNA-1), glycoprotein precursor (GP, RNA-2), nucleocapsid protein (NP, RNA-3) and movement protein (MP, RNA-4). Orthologs of representative members in the genera *Emaravirus*, *Orthotospovirus* and *Orthobunyavirus* are included. Members and tentative members of the emaraviruses reported in the phylogenetic trees are the following: AcCRaV (Actinidia chlorotic ringspot-associated virus), AcEV-2 (Actinidia emaravirus 2), BLMaV (Blackberry leaf mottle-associated virus), CjEV-1 (Camellia japonica associated virus 1), CjEV-2 (Camellia japonica associated virus 2), EMARaV (European mountain ash ringspot-associated virus), FMV (Fig mosaic virus), HPWMoV (High Plains wheat mosaic virus), JYMaV (Jujube yellow mottle-associated virus), PIVB (Pistacia virus B), PPSMV-1 (Pigeonpea sterility mosaic virus 1), PPSMV-2 (Pigeonpea sterility mosaic virus 2), PVBV (Palo verde broom virus), RLBV (Raspberry leaf blotch virus), RRV (Rose rosette virus), RYRSaV (Redbud yellow ringspot-associated virus), TIRSaV (Ti ringspot-associated virus). RNAs accession numbers of emaraviruses are reported in [Table 1](#). TSWV (Tomato spotted wilt virus, AY070212; AY870390; KM365066) and BUNV (Bunyamwera orthobunyavirus, MH091919) were used to root the phylogenetic trees. The bar represents the number of amino acid replacements per site.

i.e., PPSMV-1 and PPSMV-2. This also explains the narrow range of host plants generally infected by each emaravirus. To date, the ascertained virus-vector transmission ratios are those between EMARaV with *Phytoptus pyri*, FMV with *Aceria ficus*, PPSMV-1 and PPSMV-2 with *Aceria cajani*, RLBV with *Phyllocoptes gracilis*, RRV with *Phyllocoptes fructiphilus*, HPWMoV with *Aceria tosichella*, and BPVBV with *Aculus cercidis*. In other cases (BLMaV and RYRSaV) transmission by eriophyid mites is highly suspected, but the exact species responsible has not yet been identified.

Spontaneous host plants in marginal and uncultivated areas contribute to preserving populations of eriophyids and viruses when the sensitive and elective host plant is temporarily unavailable. It has also been observed that eriophyid (*A. cajani*, *A. tosichella* and *P. fructiphilus*) populations are larger on infected plants than on healthy ones.

Eriophyid mites possess a short stylet, with which they can penetrate only epidermal and, at most, the underlying mesophyll cells. Therefore, mites can acquire the disease agent only if it is present in these plant cells. Transmission efficiency increases as the number of individual mites present on the host plant increases. In experimental trials, the efficiency of PPSMV-1 transmission by a single *A. cajani* mite was up to 53% but rose to 100% when more than 5 mites were used per plant. Moreover, *A. cajani* and *A. tosichella* acquired PPSMV-1 and HPWMOV, respectively, after a minimum acquisition access period (AAP) of 15 min (although 50% transmission of HPWMOV was achieved only with an AAP of 16 h) and inoculated virus after a minimum inoculation access period of 90 min. Starvation before or following virus acquisition can reduce these periods. Mites retain viruses for up to 13 h in *A. cajani*, 4 days in *A. tosichella*, and 10 days in *A. ficus*. No latent period was observed with PPSMV-1 in *A. cajani*, whereas a 6–7 h latent period was reported for FMV in *A. ficus*. These data suggest a semipersistent fimoviruses transmission mode. Trans-stadial transmission of the virus was ascertained for *A. ficus* and FMV, *A. cajani* and PPSMV, and *A. tosichella* and HPWMOV. The latter virus is transmitted by all stages of *A. tosichella* and is retained through the molt. However, adults cannot acquire the virus and therefore transmit only if they acquired the virus during their immature stages.

The mechanical transmission of fimoviruses onto herbaceous hosts is rare, and has been shown only for a few species, i.e., AccRaV onto *N. benthamiana*, PPSMV-1 and PPSMV-2 onto *N. benthamiana* and *N. clevelandii*, and TiRaV onto *N. benthamiana*, *N. tabacum* and some cucurbits. The leaf stapling technique was effectively used to transmit PPSMV-1 and PPSMV-2 from infected plants onto *Phaseolus vulgaris*, while mite transmission was used to transmit FMV to *Catharanthus roseus*.

Symptomatology and Economic Impact

Given the novel establishment of this virus family and the recent identification of many of its members, mainly thanks to the use of the innovative NGS technique, it is difficult to determine the actual economic impact of these viruses on crops. For many of them, their effective distribution, apart from the countries where they were first identified, remains unclear, except for FMV, which is ubiquitous. It seems certain, however, that the range of hosts naturally infected by each emaravirus species is rather restricted, due to the high specificity of the relationship between virus/vector/plant.

FMV is the main, if not the exclusive agent of Fig mosaic (FMD), a worldwide fig tree disease first described in 1933. FMV causes many disorders in fig trees: poor fruit quality, reduced productivity and stunted growth (Fig. 1). Given the widespread and great incidence of this disease, the economic impact of FMV should be considered as particularly relevant. Another natural host of FMV outside the genus *Ficus* is the species *Cyclamen persicum*, while members of family *Moraceae*, *Cudrania tricurpidata* and *Morus indica* can be artificially infected.

HPWMOV is the agent of High Plains disease (HPD), an economically important disease of wheat and maize, first found in the Great Plains region of the United States in 1993. Since then, HPD has been reported in different regions of the USA, Australia, and New Zealand. Symptoms on sweet corn consist of stunting, yellowing (chlorosis) and mosaic foliar patterns along veins, which progress into streaking patterns and necrosis. Infected wheat develops severe mosaic symptoms.

PPSMV-1 and PPSMV-2 are the agents of sterility mosaic (SMD), which is the most damaging disease of pigeonpea (*Cajanus cajan*) in the Indian subcontinent (India, Bangladesh, Nepal, Thailand, Myanmar, Sri Lanka and China). SMD is manifested by a wide array of symptoms on pigeonpea (Fig. 1). These include various mosaic patterns on leaves, small misshapen leaves, shoots proliferation, stunting and complete or partial cessation of flower production (sterility), but the disease is not lethal to the host. Plant infection by PPSMVs at an early stage can result in a 95%–100% loss of yield, while losses from late infection depend on the level of infection and range from 26% to 97%. In the 1970s it was estimated that this disease caused annual grain losses worth over \$300 million.

Rose rosette disease (RRD), induced by RRV, is one of the most destructive diseases of cultivated roses in North America, regardless of cultivar. Symptoms were described in the United States as early as 1941. Spread of the disease in the USA was linked to the introduction from Japan of the multiflora rose, an exotic plant used as a rootstock for ornamental roses. Symptoms include excessive lateral shoot growth, excessive thorniness, witches' broom, leaf proliferation and malformation, mosaic, red pigmentation, and eventually plant death.

EMARV mainly infects European mountain ash (*Sorbus aucuparia* L.), a widely distributed indigenous tree in many parts of Europe and the Middle East, widely present in forests, but also frequently cultivated in gardens, public green areas and along avenues. The virus mostly causes chlorotic ringspots and mottling on leaves (Fig. 1), leading to the slow dieback of affected trees within a few years. Its presence has been reported in Austria, the Czech Republic, Sweden, Finland, Germany, Russia, Norway, Poland and the United Kingdom. Other host plants of EMARV apart from *S. aucuparia* have been identified, such as various Rosaceae species and hybrids.

The impact of the other known emaravirus species appears less important; their diffusion seems to be relatively contained at present, their etiological meaning is still uncertain, and they attack ornamental or cultivated plants with less commercial value.

AccRaV infects kiwifruit trees (*Actinidia* spp.), causing leaf ringspots, vein yellowing and chlorotic spots.

BLMaV has been detected in many cultivated and wild blackberries affected by yellow vein, the major constraint for blackberry production in the southern United States.

RYRSaV has been detected in eastern redbud trees (*Cercis canadensis*) grown in Arkansas, which present yellow ringspot symptoms on the leaves. Grafting can artificially infect *Glicine max*, *Phaseolus vulgaris*, *Vigna unguiculata* and *Robinia pseudoacacia* plants.

RLBV is associated with severe raspberry leaf blotch disorder in wild and cultivated *Rubus* spp. plants in many European countries. It is a long-known disease characterized by large yellow blotches or rings on leaves, twisting of the leaves, distortion of leaf margins, leaf malformation, necrosis, and reduced vigor of the severely affected plants.

TiRSaV is a tentative emaravirus responsible for an emerging disease of *Cordyline fruticosa* in Hawaii and produces small chlorotic ringspots on leaves that often coalesce to form larger lesions.

PiVB is a tentative emaravirus infecting symptomless *Pistacia* plants in Turkey.

AcEV-2 is a tentative emaravirus affecting kiwifruit tree showing leaf mottle and chlorosis symptoms in China.

CjEV-1 and CjEV-2 are two tentative emaraviruses found in Italy in infected *Camellia* plants associated to symptoms of chlorotic and necrotic ringspots, deformations and yellowing on leaves, and deformations and color breaking of petals.

JYMaV is a tentative emaravirus found in China associated to Jujube yellow mosaic, a disease characterized by mottling, yellowing and distortion of leaves, malformation and discoloration of fruits, with necrotic areas mainly around the calyx.

LiCRaV is a tentative emaravirus found in China on ornamental shrub common lilac (*Syringa vulgaris* L.), associated to leaf chlorotic ringspots and mottling symptoms.

PVBV is a tentative emaravirus infecting the ornamental plant *Parkinsonia florida* in Arizona, causing witches' broom symptoms.

WBVV is a tentative emaravirus attacking *Arctium* spp. (*Asteraceae*) in Finland.

Diagnosis

In recent years there has been a sharp rise in the number of fimoviruses due to the impulse given using molecular techniques. Of the many such techniques available, Next-generation sequencing (NGS) technology is a powerful tool for the detection of unknown disease-associated viruses and has made a great contribution to the discovery of these viruses. Fimoviruses are easily discernable in the field because of the outstanding symptoms, *i.e.*, typical chlorotic ringspot and mosaic, that they induce in infected plants. However, when are present together with mixed viral infections, the specific recognition of their symptoms may be disorientating.

Their mechanical transmission to herbaceous hosts from infected sap, which is a simple way for recovering viruses, often is unsuccessful. This is due on the one hand to the non-transmissible nature of many of these viruses and on the other hand to the presence of plant inhibitors impeding the success of the inoculation.

In the absence of serological means for the detection of most fimoviruses, except for FMV and PPSMV-1, DNA-based molecular assays with all their various techniques remain the most suitable methods for detecting these viruses. The high number of RNA segments composing the genome and the presence of conserved motifs in many of their genes has led to the development of specific and degenerate primers, which have proved effective in RT-PCR to detect fimoviruses at the species and genus levels, respectively. Furthermore, TaqMan Realtime PCR and LAMP assays have also been developed for FMV and BLMaV, providing a greater sensitivity than conventional RT-PCR assay in the detection of both viruses. At the serological level, immunological detection is based on two different polyclonal antisera raised against the recombinant RNP of FMV and on purified virus particles of HPWMoV and PPSMV-1.

Treatment and Prevention

Given the impossibility of combating virus diseases in the field, control of fimoviruses is essentially preventive. It mostly consists of avoiding the introduction of infected plants that can act as a reservoir for the spread of viruses to healthy plants *via* vectors. Thus, rapid detection of infected plants and their immediate and complete removal, and the elimination of alternative host species, where these exist, (*e.g.*, multiflora roses for RRV, grass-type weeds and volunteer grain crops for HPWMoV) or of infected residues from previous crops (*e.g.*, for RRV and PPSMV) can be very useful in reducing the amount of inoculum.

Good spacing between plants so that leaves do not touch can prevent eriophyid mites from moving within a crop. Containment of the infection at very low levels in the field also serves to delay the onset of the disease and significantly attenuates the extent of symptoms, especially in herbaceous crops (*e.g.*, in the case of PPSMVs, RRV and HPWMoV).

Delayed planting of winter wheat reduces the chance for *Aceria tosichella* to transmit the virus and prevents an overwintering reservoir for HPWMoV. Early season planting of corn has been shown to reduce the effect of HPWMoV infections by allowing the plants to mature before the eriophyid vector becomes active.

The use of genetic resistance sources is one of the most promising approaches in preventing or lessening the impact of fimoviruses. Several varieties of HPWMoV-resistant maize have been identified, although there are no known resistant varieties of wheat, barley, oats or rye. Genetic resistance lines have also been identified regarding pigeonpea by evaluating the behavior of numerous varieties preserved in germplasm collections when infected with PPSMVs.

Control of the eriophyid vectors, mainly involving pesticides, is the other pillar in control of these viruses. Pesticide use alone entails a high level of risk because the eriophyid mite may develop resistance. Therefore, the most effective way to manage fimovirus diseases is a multi-tactic approach that includes both cultural and chemical practices.

Further Reading

- Buzkan, N., Chiumenti, M., Massart, S., *et al.*, 2019. A new emaravirus discovered in Pistacia from Turkey. *Virus Research* 263, 159–163.
- Di Bello, P.L., Laney, A.G., Druciarek, T., *et al.*, 2016. A novel emaravirus is associated with redbud yellow ringspot disease. *Virus Research* 222, 41–47.
- Elbeaino, T., Digiario, M., Alabdullah, A., *et al.*, 2009. A multipartite single-stranded negative-sense RNA virus is the putative agent of fig mosaic disease. *Journal of General Virology* 90, 1281–1288.
- Elbeaino, T., Digiario, M., Mielke-Ehret, N., Muehlbach, H.P., Martelli, G.P., 2018. ICTV Report Consortium. (2018) ICTV Virus Taxonomy Profile: *Fimoviridae*. *Journal of General Virology* 99 (11), 1478–1479.
- Elbeaino, T., Digiario, M., Uppala, M., Sudini, H., 2014. Deep sequencing of pigeonpea sterility mosaic virus discloses five RNA segments related to emaraviruses. *Virus Research* 188, 27–31.
- Elbeaino, T., Digiario, M., Uppala, M., Sudini, H., 2015. Deep sequencing of dsRNAs recovered from mosaic-diseased pigeonpea reveals the presence of a novel emaravirus: Pigeonpea sterility mosaic virus 2. *Archives of Virology* 160, 2019–2029.
- Laney, A.G., Keller, K.E., Martin, R.R., Tzanetakis, I.E., 2011. A discovery 70 years in the making: Characterization of the Rose rosette virus. *Journal of General Virology* 92, 1727–1732.
- McGavin, W.J., Mitchell, C., Cock, P.J.A., Kathryn, K.M., MacFarlane, S.A., 2012. Raspberry leaf blotch virus, a putative new member of the genus *Emaravirus*, encodes a novel genomic RNA. *Journal of General Virology* 93, 430–437.
- Mielke, N., Muehlbach, H.P., 2007. A novel multipartite negative-strand RNA virus is associated with the ringspot disease of European mountain ash (*Sorbus aucuparia* L.). *Journal of General Virology* 88, 1337–1346.
- Olmedo-Velarde, A., Park, A.C., Sugano, J., *et al.*, 2019. Characterization of Ti ringspot-associated virus, a novel emaravirus associated with an emerging ringspot disease of *Cordyline fruticosa* (L.). *Plant Disease* 103 (9), 2345–2352. doi:10.1094/PDIS-09-18-1513-RE.
- Peracchio, C., Forgia, M., Chiapello, M., *et al.*, 2019. A complex virome that includes two distinct emaraviruses is associated to virus-like symptoms in *Camellia japonica*. bioRxiv. doi:10.1101/822254.
- Tatineni, S., McMechan, A.J., Wosula, E.N., *et al.*, 2014. An eriophyid mite-transmitted plant virus contains eight genomic RNA segments with unusual heterogeneity in the nucleocapsid protein. *Journal of General Virology* 88, 11834–11845.
- Wang, Y., Zhai, L., Wen, S., *et al.*, 2020. Molecular characterization of a novel emaravirus infecting *Actinidia* spp. in China. *Virus Research* 275. doi:10.1016/j.virusres.2019.197736.
- Yang, C., Zhang, S., Han, T., *et al.*, 2019. Identification and characterization of a novel emaravirus associated with jujube (*Ziziphus jujuba* Mill.) yellow mottle disease. *Frontiers in Microbiology* 10 (1417), 1–12. doi:10.3389/fmicb.2019.01417.
- Zheng, Y., Navarro, B., Wang, G., *et al.*, 2017. Actinidia chlorotic ringspot-associated virus: A novel emaravirus infecting kiwifruit plants. *Molecular Plant Pathology* 18, 569–581.

Furoviruses (*Virgaviridae*)

Annette Niehl and Renate Koenig, Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of R. Koenig, Furovirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00408-8](https://doi.org/10.1016/B978-012374410-4.00408-8).

Nomenclature

bp	Base pair	nt	Nucleotide
CP	Coat protein	ORF	Open reading frame
CP-RT	Coat protein-readthrough protein	PCR	Polymerase chain reaction
ELISA	Enzyme-linked immunosorbent assay	PD	Plasmodesmata
ER	Endoplasmic reticulum	PME	Pectin methyl-esterase
GFP	Green fluorescent protein	RdRp	RNA-dependent RNA polymerase
HSP	Heat shock protein	RNA	Ribonucleic acid
kDa	Kilodalton	siRNAs	Small-interfering RNAs
MP	Movement protein	ssRNA	Single-stranded ribonucleic acid
NGS	Next generation sequencing	vRNP	Viral ribonucleoprotein
nm	Nanometer	VSR	Viral suppressor of gene silencing

Glossary

Furovirus Siglum from fungus-borne, rod-shaped virus.

Classification

Furoviruses are positive sense, single stranded RNA (ssRNA) viruses in the family *Virgaviridae*. Furoviruses have a bipartite genome and each RNA is encapsidated by a single SBWMV of coat protein (CP). *Soil-borne wheat mosaic virus* (SBWMV) is the SBWMV species of the genus *Furovirus*. The genus *Furovirus* presently comprises six species; *Soil-borne wheat mosaic virus*, *Japanese Soil-borne wheat mosaic virus* (JSBWMV), *Soil-borne cereal mosaic virus* (SBCMV), *Chinese wheat mosaic virus* (CWMV), *Oat golden stripe virus* (OGSV), and *Sorghum chlorotic spot virus* (SrCSV). In 2011, a barley-infecting furovirus named soil-borne barley mosaic virus (SBBMV) was identified. Because of high nt sequence identity to JSBWMV (approximately 94% on RNA2) and the fact that barley is infected as primary host species, we suggest that this virus is classified as strain of the JSBWMV. (Table 1).

Virion Structure

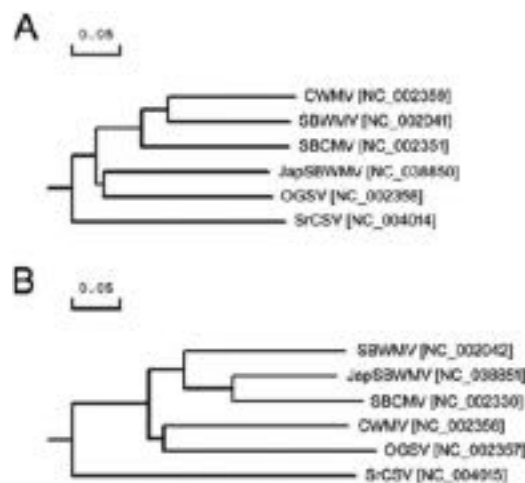
Furoviruses have bipartite genomes which are encapsidated in non-enveloped rod-shaped particles with a helical symmetry and a diameter of approximately 20 nm. The maximum length of the particles ranges between ca. 140–160 nm and 260–300 nm. Due to internal deletions in the CP-readthrough protein genes on RNA2, particles shorter than 140 nm may arise in naturally infected plants and in laboratory isolates at later infection stages or after passage of the virus from plant to plant, and may outcompete the 140–160 nm particles. The furovirus CP has a molecular mass of approximately 20 kDa. Virions occur scattered, or in aggregates or inclusion bodies in the cytoplasm and vacuole of infected cells. Inclusion bodies may be crystalline or consist of clusters of virus particles in association with microtubules. In plants doubly infected by *Sorghum chlorotic spot virus* (SrCSV) and potyviruses, SrCSV virions have been shown to bind to cylindrical inclusions induced by potyviruses present in the same cell.

Nucleic Acid Properties and Differentiation of Furoviruses

The nt sequences of the RNAs of all six furovirus member species have been determined. The RNAs are 5' capped and terminate in a tRNA-like structure. For SBWMV, it has been shown that the tRNA-like structure can be amino acetylated with valine in vitro. The furoviral tRNA-like structure is preceded in the 3' untranslated region by an upstream hairpin and an upstream pseudoknot domain with two to seven possible pseudoknots. The genome organization is rather similar for all furoviruses although there are considerable differences in the nt sequences (Fig. 1). The percentages of sequence identities between SBWMV and CWMV, SBCMV and JSBWMV,

Table 1 List of species in the genus *Furovirus*, family *Virgaviridae*. The SBWMV species is written in bold

Species name	Acronym	# segments	Length (nt)	Accession #
Chinese wheat mosaic virus	CWMV	2	(7147 nt) (3569 nt)	NC_002359 NC_002356
Japanese soil-borne wheat mosaic virus	JSBWMV	2	(3574 nt) (7226 nt)	NC_038851 NC_038850
Oat golden stripe virus	OGSV	2	(3232 nt) (7111 nt)	NC_002357 NC_002358
Soil-borne cereal mosaic virus	SBCMV	2	(7025 nt) (3683 nt)	NC_002351 NC_002330
Soil-borne wheat mosaic virus	SBWMV	2	(7099 nt) (3593 nt)	NC_002041 NC_002042
Sorghum chlorotic spot virus	SCSV	2	(3418 nt) (6878 nt)	NC_004015 NC_004014

**Fig. 1** Phylogenetic trees generated by the DNAMAN software for furoviral RNA1 (A) and RNA2 (B). The Genbank accession numbers are given in parentheses.

respectively, ranges from approximately 70%–77% for RNA1. OGSV and SCSV are more distantly related to SBWMV. For RNA2, SBWMV, JSBWMV, and SBCMV are most closely related. The percentage of sequence identities between the different members of the furoviruses for RNA2 ranges between 67% and 83%. Sequence dissimilarities justify that CWMV and SBCMV may be considered as separate species and that JSBWMV is considered a separate species rather than a strain of SBWMV (Fig. 1). However, the fact that reassortment experiments between RNA1 of JSBWMV and RNA2 of the SBWMV strain of SBWMV yielded an infectious progeny and the observation that the biological properties of the wheat-infecting furoviruses are very similar may be taken as evidence that SBWMV - and JSBWMV as well as CWMV and SBCMV may also all be regarded as distantly related strains of the same species. Moreover, differences at the nt level do not necessarily translate into differences at the aa level as exemplified with the CP sequences of the New York and Nebraska strains of SBWMV, which are 100% identical at the aa level, despite the fact that their coding regions differ in 68 nt positions.

Organization of the Genome and Properties of the Encoded Proteins

Furoviral RNA1 codes for two replication-associated proteins of approximately 150 kDa and 209 kDa and for a 37 kDa movement protein (MP) (Fig. 2). The shorter of the two replication-associated proteins contains methyl-transferase and helicase motifs. The longer one results from the read-through of a stop codon and contains the RNA-dependent RNA polymerase (RdRp) in addition to the methyl-transferase and helicase domains. Over-expression experiments using fluorescent protein tagged CWMV replicase showed that it localized to membrane-associated inclusions leading to the hypothesis that virus replication might occur on these host membranes. However, the identity of the host membranes was not further investigated. The CWMV replicase was also shown to interact with heat shock proteins (HSP) of the HSP70 family. The furovirus 37 kDa MP belongs to the '30K' superfamily of viral MPs. Experimental evidence for the function of the CWMV 37 kDa protein as MP came from studies showing that the green fluorescent

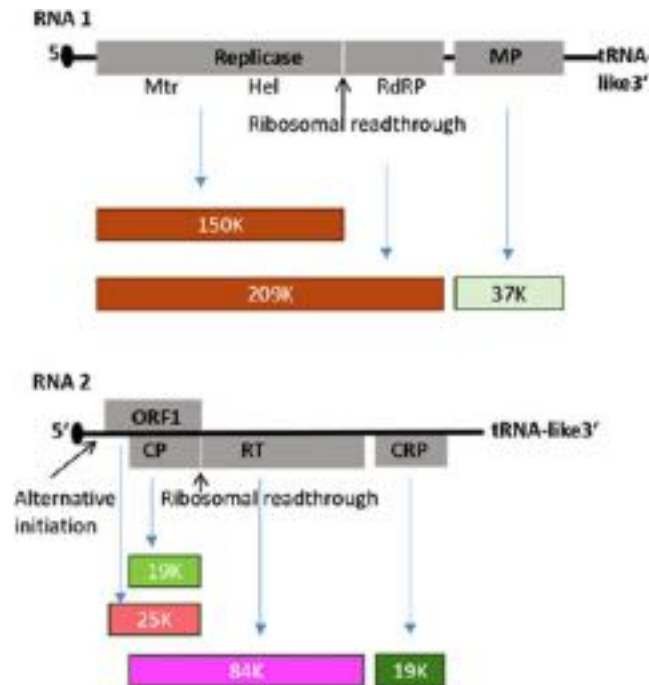


Fig. 2 Genome organization of SBWMV. The RNAs are capped and possess a t-RNA structure at the 3' end. RNA1 contains the genetic information for two replicase proteins, the 150K protein containing methyltransferase (Mtr) and helicase (Hel) domains and the 209K protein containing in addition the RNA-dependent RNA polymerase (RdRP) domain. A 37K MP is encoded by ORF2. RNA2 encodes for a 19K major CP, a 25K minor CP initiated at an upstream CUG start codon, a 84K CP-RT protein and a 19K cysteine-rich silencing suppressor.

protein (GFP)-labeled 37 kDa CWMV MP moves from cell to cell when bombarded into wheat leaf cells and can trans-complement the intercellular spread of potato virus X. Moreover, CWMV MP exhibits plasmodesmata (PD) labeling and forms endoplasmic reticulum (ER)-derived vesicular and aggregate structures. Export from the ER appears to be required for intracellular movement and depends on the secretory pathway. Moreover, the CWMV MP interacts with host pectin methyl-esterase (PME), a host cell-wall modifying enzyme, which has been shown to interact with a variety of MPs of the '30K' superfamily. In the case of tobacco mosaic virus, it has been shown that interaction between MP and PME is important for virus movement.

Furoviral RNA2 contains the genetic information for the approximately 20 kDa major coat protein (CP), an approximately 24 kDa minor CP initiated at an upstream CUG start codon, an approximately 84 kDa CP-readthrough (CP-RT) protein and an approximately 19 kDa cysteine-rich protein. Neither the CP with N-terminal extension nor the CP-RT are required for SBWMV and CWMV virion formation and systemic infection. The 24 kDa N-terminally extended CP may be incorporated into virus particles, as it has been detected in purified CWMV virions. For SBCMV and CWMV, the 24 kDa extended CP, but not CP itself, has been shown to interact with the CWMV 19 kDa RNA silencing suppressor. In the case of SBWMV, both CP forms were able to interact with CWMV P19. As silencing activity was not significantly affected by the interaction, a function of the interaction between forms of CP and P19 in silencing suppression cannot be deduced.

Despite dispensable for virion assembly and systemic infection, CP-RT can be associated with virions, as CWMV CP-RT was detected on the surface of virus particles by immunogold labeling. The CP-RT protein of furoviruses appears to be involved in vector transmission, similar to the CP-RT proteins of beny- and pomoviruses. Prolonged cultivation of field-infected plants, virus propagation by repeated mechanical inoculations or growth at elevated temperatures result in spontaneous deletions within the CP-RT domain. These mutants can no longer be transmitted by the vector. Two highly conserved trans-membrane domains within CP-RT proteins are thought to be involved in vector transmission by facilitating the transport of the virion or infectious viral ribonucleoprotein complex across the plasmodiophorid membrane into the host cell cytoplasm.

The CP-RT protein also appears to affect symptom development as spontaneous mutations occurring in SBWMV CP-RT or engineered mutations affecting expression of the CP-RT protein in infectious CWMV clones increased symptom severity. Interestingly, mutations affecting expression of the N-terminally extended minor CP slowed symptom development for CWMV and SBWMV, thus indicating that the different CP-variants and the ratio of the different CP-forms produced may be involved in fine-tuning the infection cycle.

The cysteine-rich 19 kDa protein acts as a suppressor of RNA silencing (VSR) and enhances symptom severity. For the CWMV protein it has been shown that it localizes to the ER and is capable to self-interact. Self-interaction, but not ER-localization correlates with VSR activity of the protein. Residues important for protein stability, silencing suppression activity, symptom development, as well as targeting to the ER have been identified. The exact mode of action of the furovirus VSR is as yet unknown; however, for the CWMV CRP, it appears not to be related with local siRNA production and accumulation. Instead, CWMV CRP protein appears to promote viral cell-to-cell spread, possibly by interfering with the spread of RNA silencing signals ahead of

infection, as movement of a heterologous virus, whose movement depends on the presence of a functional silencing suppressor to be provided in trans, is complemented in the presence of over-expressed CWMV CRP. The localization of the CWMV CRP protein to the ER may be important for symptom development.

Virus Transmission and Movement

Furoviruses are soil-borne. SBWMV has been demonstrated to be transmitted by the zoospores of *Polymyxa graminis*, a ubiquitous plasmodiophorid protozoan. *P. graminis* is probably also the vector for the other viruses in the genus. Polymyxa-transmitted viruses are believed to be taken up by the plasmodia of the vector in infected root cells. The multinucleate plasmodia, which are separated from the host cytoplasm by a distinct membrane, may either develop into zoosporangia from which secondary viruliferous zoospores are released. Alternatively, the plasmodia may develop into cytosori, in which resting spores mature. These resting spores are released into the soil from decaying plant material and may survive in the soil for many years even under extreme conditions. Spores are distributed by agricultural equipment, by irrigation or even by wind. Upon germination, they release primary zoospores transmitting the virus. Zoospores inject their contents into the cytoplasm of root cells and initiate infection by the formation of plasmodia. Zoospores treated with antisera to SBWMV or resting spores treated with 0.1 N NaOH or HCl retained their ability to transmit the virus, indicating that the infectious viral material is present inside the spores. Immunolabeling and in situ RNA hybridization studies have revealed the presence of SBWMV MP and RNA but not of SBWMV CP in the resting spores of *P. graminis*, indicating that SBWMV may be transmitted in form of viral ribonucleoprotein (vRNP) complexes. Moreover, the presence of viral RNAs and MP may indicate that the virus is able to replicate inside the vector. Transmission in form of a vRNP complex containing viral MP and RNA may have the advantage, that the RNA can be directly transported from the primarily infected cell into adjacent cells for the initiation of virus replication and thus, competition between virus and vector for reprogramming of host cell metabolism for replication may be avoided. The '30K' superfamily of MPs, to which the furovirus MPs belong, are known to mediate cell-to-cell and vascular transport of viruses by binding viral nucleic acids and carrying them through plasmodesmata and through the vasculature. However, with respect to systemic movement, SBWMV may use the xylem in order to move from infected roots to the leaves as suggested from light and electron microscopy experiments. The virus may enter primary xylem elements before cell death occurs. Moreover, the virus may move laterally between adjacent xylem vessels.

Symptoms, Epidemiology and Host Range

SBWMV is known since the early 1920s. It causes chlorotic leaf streaking, stunting, and severe losses of yield in winter wheat. SBWMV is widely distributed in the USA and Canada, mainly in the east and has been reported to occur in Europe in Great Britain, Germany, and Poland. It has also been described to occur in Brazil, Africa (Zambia), and Japan. Two types of SBWMV are described, the original "Nebraska-like" SBWMV and a "New York-like" SBWMV. Symptoms often occur in patches in the fields. SBWMV naturally infects wheat, barley, rye, and triticale. An additional strain of SBWMV naturally infecting barley and wheat, termed Japanese soil-borne wheat mosaic virus, has been identified in Japan (SBWMV-JP), France (SBWMV-FR), and Germany (SBBMV-DE). Diseases on wheat with symptoms similar to those described for SBWMV are caused by CWMV in China and SBCMV in Europe. Apart from wheat, SBCMV naturally also infects rye and triticale. SBCMV occurs in a number of European countries including Denmark, Great Britain, Belgium, Poland, Ukraine, France, Italy, Germany, and Turkey. SBCMV occurs in different SBWMV with different aggressivity on wheat. In Italy, France, England, Belgium, and Denmark, wheat becomes heavily infected, while SBCMV in Germany and Poland usually develops weaker symptoms in wheat.

OGSV causes yellow striping on leaves and has been detected in oats in Great Britain, France, and the USA. SBWMV, SBCMV, CWMV, and OGSV may be transmitted mechanically to some species of the *Chenopodium* and *Nicotiana* genus.

SrCSV was isolated in 1986 Kansas/USA. It is mechanically transmissible to maize where it produces a bright yellow mosaic and elongated ringspot symptoms several weeks after inoculation. Local infections are produced on mechanically inoculated *Chenopodium quinoa*, *C. amaranticolor*, and *Nicotiana clevelandii*. Attempts to transmit the virus back to *Sorghum bicolor* or to winter wheat either mechanically or by growing plants in soil, in which infected sorghum was growing, were unsuccessful.

Plants naturally infected by furoviruses are often also infected by bymoviruses, that is, Wheat spindle streak mosaic virus in Europe and North America, Wheat yellow mosaic virus in East Asia or Oat mosaic virus in Europe and the USA. The symptoms caused by these bymoviruses are very similar to those caused by the furoviruses, and both, the furoviruses and the bymoviruses are transmitted by *P. graminis*.

Diagnosis

Visible symptoms caused by furoviruses may vary depending on environmental conditions, such as moisture and temperature as well as dependent on the plant cultivar. Nutrient deficiencies, winter injury or other viruses (especially bymoviruses), may produce symptoms that are easily confused with those caused by furoviruses. Thus, serological or molecular biological techniques are the methods of choice for reliable virus detection. Serological tests such as the enzyme-linked immunosorbent assay (ELISA), tissue

print immunoassay, and immunoelectron microscopy as well as polymerase chain reaction (PCR) techniques are established for the specific detection and discrimination of the viruses. Real-time reverse transcriptase-PCR (RT-PCR) based on intercalation dyes and TaqMan probes have been developed for SBCMV and allow the quantification of virus titer within the tested plants. In the case of Next generation sequencing (NGS) it is possible to identify a virus in a sample without a priori sequence information.

Control

Furoviruses inside the long-living resting spores of *P. graminis* may survive for decades. Growing resistant or tolerant varieties currently represents the only practical and environmentally friendly means to lower the impact of these diseases on yield. Furovirus resistance appears to be quantitative and more than one gene appears to control the resistance phenotype. The only resistance genes known to date to be effective against furoviruses in wheat are the *Sbm1* gene located on chromosome 5DL and the *sbm2* gene located on chromosome 2BS. Both genes confer the so called “translocation resistance”, in which roots still become infected but the virus is inhibited to efficiently infect the shoots of the plants. Although disease symptoms are reduced in the resistant plants, virus ingress into the soil remains unaffected, thus necessitating the development of additional resistance strategies to control these viruses. Cereal genotypes with resistance to *P. graminis* have so far not been identified.

Similarities and Dissimilarities with Other Taxa

The morphology of furoviruses resembles that of other rod-shaped viruses, i.e., of benyviruses, pecluviruses, pomoviruses, hordeiviruses, tobnaviruses, and tobamoviruses. The CPs of all these viruses have a number of conserved residues. Furoviruses, pecluviruses and tobnaviruses have bipartite genomes. Furoviruses differ from pecluviruses by having their movement function encoded on a single ORF rather than a triple gene block. They also differ from pecluviruses as well as from tobnaviruses by having a CP-RT gene. The gene for their cysteine-rich protein is located on RNA2, whereas with pecluviruses and tobnaviruses it is located on RNA1. The cysteine-rich proteins of furoviruses, pecluviruses, and tobnaviruses act as suppressors of post-transcriptional gene silencing and are phylogenetically interrelated. Furoviruses encode two N-terminally overlapping replication-associated proteins (Fig. 2), which are related to those of pomoviruses, pecluviruses, tobnaviruses, and tobamoviruses. The 37 kDa MP of the furoviruses belongs to the ‘30K’ superfamily of MPs and relates the furoviruses to the dianthoviruses and the tobamoviruses, whereas the valine-accepting tRNA-like structure on the 3′ ends of furovirus RNAs relates them to the tymoviruses. Furoviruses like tobamoviruses have an upper pseudoknot domain in the 3′ untranslated regions of their genomic RNAs.

Further Reading

- Adams, M.J., Adkins, S., Bragard, C., *et al.*, 2017. ICTV virus taxonomy profile: *Virgaviridae*. Journal of General Virology. 1999–2000.
- An, H., Melcher, U., Doss, P., *et al.*, 2003. Evidence that the 37 kDa protein of *Soil-borne wheat mosaic virus* is a virus movement protein. Journal of General Virology 84, 3153–3163.
- Andika, I.B., Zheng, S., Tan, Z., *et al.*, 2013. Endoplasmic reticulum export and vesicle formation of the movement protein of *Chinese wheat mosaic virus* are regulated by two transmembrane domains and depend on the secretory pathway. Virology 435, 493–503.
- Bass, C., Hendley, R., Adams, M.J., Hammond-Kosack, K.E., Kanyuka, K., 2006. The *Sbm1* locus conferring resistance to *Soil-borne cereal mosaic virus* maps to a gene-rich region on 5DL in wheat. Genome 49, 1140–1148.
- Campbell, R.N., 1996. Fungal transmission of plant viruses. Annual Review of Phytopathology 34, 87–108.
- Dreher, T.W., 2009. Role of tRNA-like structures in controlling plant virus replication. Virus Research 139, 217–229.
- Driskel, B.A., Doss, P., Littlefield, L.J., Walker, N.R., Verchot-Lubicz, J., 2004. Soilborne wheat mosaic virus movement protein and RNA and wheat spindle streak mosaic virus coat protein accumulate inside resting spores of their vector, *Polymyxa graminis*. Molecular Plant-Microbe Interactions 17, 739–748.
- Goodwin, J.B., Dreher, T.W., 1998. Transfer RNA mimicry in a new group of positive-strand RNA plant viruses, the furoviruses: Differential aminoacylation between the RNA components of one genome. Virology 246, 170–178.
- Hao, Y., Wang, Y., Chen, Z., *et al.*, 2012. A conserved locus conditioning *Soil-borne wheat mosaic virus* resistance on the long arm of chromosome 5D in common wheat. Molecular Breeding 30, 1453–1464.
- Koenig, R., Bergstrom, G.C., Gray, S.M., Loss, S., 2002. A New York isolate of soil-borne wheat mosaic virus differs considerably from the Nebraska type strain in the nucleotide sequences of various coding regions but not in the deduced amino acid sequences. Archives of Virology 147, 617–625.
- Kühne, T., 2009. Soil-borne viruses affecting cereals – Known for long but still a threat. Virus Research 141, 174–183.
- Miyaniishi, M., Roh, S.H., Yamamiya, A., Ohsato, S., Shirako, Y., 2002. Reassortment between genetically distinct Japanese and US strains of *Soil-borne wheat mosaic virus*: RNA1 from a Japanese strain and RNA2 from a US strain make a pseudorecombinant virus. Archives of Virology 147, 1141–1153.
- Shirako, Y., Suzuki, N., French, R.C., 2000. Similarity and divergence among viruses in the genus *Furovirus*. Virology 270, 201–207.
- Sun, L., Andika, I.B., Kondo, H., Chen, J., 2013. Identification of the amino acid residues and domains in the cysteine-rich protein of *Chinese wheat mosaic virus* that are important for RNA silencing suppression and subcellular localization. Molecular Plant Pathology 14, 265–278.
- Te, J., Melcher, U., Howard, A., Verchot-Lubicz, J., 2005. Soilborne wheat mosaic virus (SBWMV) 19K protein belongs to a class of cysteine rich proteins that suppress RNA silencing. Virology Journal 2, 18. [18].
- Verchot, J., Driskel, B.A., Zhu, Y., Hunger, R.M., Littlefield, L.J., 2001. Evidence that soilborne wheat mosaic virus moves long distance through the xylem in wheat. Protoplasma 218, 57–66.
- Yamamiya, A., Shirako, Y., 2000. Construction of full-length cDNA clones to soil-borne wheat mosaic virus RNA1 and RNA2, from which infectious RNAs are transcribed in vitro: Virion formation and systemic infection without expression of the N-terminal and C-terminal extensions to the capsid protein. Virology 277, 66–75.

- Yang, J., Zhang, F., Xie, L., *et al.*, 2016. Functional identification of two minor capsid proteins from *Chinese wheat mosaic virus* using its infectious full-length cDNA clones. *Journal of General Virology* 97, 2441–2450.
- Yang, J., Zheng, S.-L., Zhang, H.-M., *et al.*, 2014. Analysis of small RNAs derived from *Chinese wheat mosaic virus*. *Archives of Virology* 159, 3077–3082.
- Ziegler, A., Klingebell, K., Papke, V., Kastirr, U., 2014. Quantification of *Wheat spindle streak mosaic virus* and *Soil borne cereal mosaic virus* in resistance testing of field samples of triticale using real-time RT-PCR. *Journal of Plant Diseases and Protection* 121, 149–155.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/virgaviridae/667/genus-furovirus

Genus: Furovirus

Virgaviridae

Positive-sense RNA Viruses.

Geminiviruses (*Geminiviridae*)

Jesús Navas-Castillo and Elvira Fiallo-Olivé, Institute for Mediterranean and Subtropical Horticulture “La Mayora”–Spanish National Research Council–University of Malaga, Algarrobo-Costa, Málaga, Spain

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
CP Coat protein or capsid protein
CR Common region
dsDNA Double-stranded deoxyribonucleic acid
EM Electron microscopy
HR Hypersensitive response
IPM Integrated pest management
IR Intergenic region
kb Kilobase
kDa Kilo dalton
MP Movement protein

NSP Nuclear shuttle protein
nt Nucleotide(s)
ORF Open reading frame
PCR Polymerase chain reaction
PTGS Post-transcriptional gene silencing
RCA Rolling circle amplification
RDR Recombination-dependent replication
RdRp RNA-dependent RNA polymerase
ssDNA Single-stranded DNA
TGS Transcriptional gene silencing
TrAP Transcriptional activator protein

Glossary

CRISPR-Cas The acronym for Clustered Regularly Interspaced Short Palindromic Repeat-CRISPR-associated protein, an adaptive immune system used by bacteria and archaea against viruses and mobile genetic elements, which has been adapted for genome editing in eukaryotes.

Emerging disease A disease of infectious origin that has appeared in a population for the first time, or that has rapidly increased in incidence or geographic range.

Geminate virion A type of twinned virions unique to geminiviruses consisting of two joined, incomplete T=1 icosahedra.

Nucleic acid satellites Subviral agents depending on helper viruses for their replication and/or transmission;

those associated to geminiviruses (alphasatellites, betasatellites and deltasatellites) are circular single stranded DNA molecules.

Pseudorecombination A particular type of genetic exchange which involves swapping of complete genome segments that make up the segmented viral genomes, as bipartite geminivirus genomes.

Recombination A process in which exchange of genome segments occurs between DNA (or RNA) molecules during replication.

Rolling circle amplification A technique that allows amplification of circular DNA molecules (as geminivirus genomes) with $\phi 29$ DNA polymerase and random primers under isothermal conditions.

Introduction

Geminiviruses (family *Geminiviridae*) are a large group of plant viruses that possess small circular single-stranded (ss) DNA genomes encapsidated in unique twinned (geminate) virions. Many geminiviruses cause economically important diseases to vegetable and fiber crops worldwide, although major losses are caused in tropical and subtropical regions. Many others are found infecting weeds and other wild plant species which, in turn, are potential reservoirs for geminivirus emergence. A geminivirus is thought to have been responsible for an unseasonal change in the appearance of eupatorium plants – autumnal yellowing in summer – described in a poem by the Empress Koken in Japan in 752 CE. Thus, geminiviruses are probably the cause of the first record of a viral disease in plants.

With nine recognized genera and a number of species approaching 500, the family *Geminiviridae* is revealed as the plant virus family with the greatest evolutionary success. Their members infect both dicot and monocot species and use as vectors a wide range of insects including whiteflies, aphids, leafhoppers, and treehoppers.

The present entry focuses on current knowledge about viruses in the family *Geminiviridae*, dealing with taxonomy, phylogeny, evolution, genome structure and expression, diagnosis, pathogenesis, epidemiology and control.

Taxonomy, Phylogeny, and Evolution

Genus demarcation criteria in the family *Geminiviridae* include host range (monocots or dicots), type of vector (aphids, leafhoppers, treehoppers, whiteflies), genome organization (monopartite or bipartite) and phylogenetic relationships. Currently, geminiviruses are classified in nine genera: *Becurtovirus*, *Begomovirus*, *Capulavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus*, briefly described below ([Table 1](#)).

Table 1 Classification of members of the family *Geminiviridae* into genera, with indication of type member, number of virus species, number of DNA genome components, transmission vector and host type

Genus	Type species	Acronym	Number ^a	Genome	Vector	Host type
<i>Becurtovirus</i>	<i>Beet curly top Iran virus</i>	BCTIV	3	Monopartite	Leafhoppers	Dicots
<i>Begomovirus</i>	<i>Bean golden yellow mosaic virus</i>	BGYMV	409	Monopartite/bipartite	Whiteflies	Dicots
<i>Capulavirus</i>	<i>Euphorbia caput-medusae latent virus</i>	EcmLV	4	Monopartite	Aphids	Dicots
<i>Curtovirus</i>	<i>Beet curly top virus</i>	BCTV	3	Monopartite	Leafhoppers	Dicots
<i>Eragrovirus</i>	<i>Eragrostis curvula streak virus</i>	ECSV	1	Monopartite	Unknown	Monocots
<i>Grablovirus</i>	<i>Grapevine red blotch virus</i>	GRBV	3	Monopartite	Treehoppers	Dicots
<i>Mastrevirus</i>	<i>Maize streak virus</i>	MSV	40	Monopartite	Leafhoppers	Monocots/Dicots
<i>Topocuvirus</i>	<i>Tomato pseudo-curly top virus</i>	TPCTV	1	Monopartite	Treehoppers	Dicots
<i>Turncurtovirus</i>	<i>Turnip curly top virus</i>	TCTV	2	Monopartite	Leafhoppers	Dicots

^aNumber of members according to the International Committee on Taxonomy of Viruses (ICTV) as per March 2019.

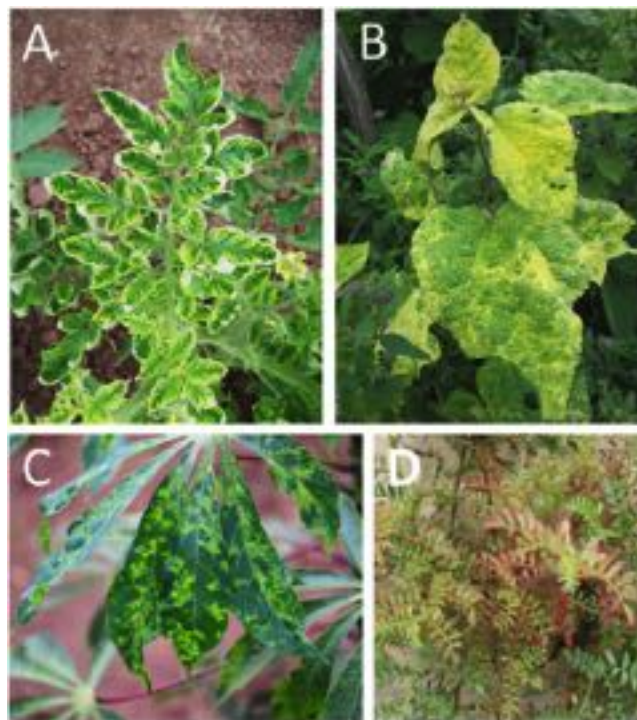


Fig. 1 Symptoms caused by the begomoviruses Tomato yellow leaf curl virus in tomato (A), the complex Tomato yellow leaf distortion virus and Tomato yellow leaf distortion deltasatellite 1 in the wild plant *Sidastrum micranthum* (B) and African cassava mosaic virus in cassava (C), and the mastrevirus Chickpea chlorotic dwarf virus in chickpea (D). (D) Reproduced from Kanakala, S., Kuria, P., 2019. Chickpea chlorotic dwarf virus: An emerging monopartite dicot infecting mastrevirus. Viruses 11, 5, CC BY 4.0 licence.

Becurtovirus

This genus includes three recognized species, *Beet curly top Iran virus*, *Exomis microphylla latent virus* and *Spinach curly top Arizona virus*. Members of all species are monopartite viruses transmitted by leafhoppers to dicot plants.

Begomovirus

This genus, with 409 accepted species, is the largest in the entire virosphere. Its members infect dicot plants and are transmitted by whiteflies of the *Bemisia tabaci* cryptic species complex. The genome can be either bipartite or monopartite. Begomoviruses usually induce severe symptoms in their hosts, including yellow/golden mosaics, leaf curl/deformation, yellow veins and stunting (**Fig. 1(A)–(C)**). Devastating viruses include members of the species *African cassava mosaic virus*, *Bean golden yellow mosaic virus*, *Cotton leaf curl Multan virus*, and *Tomato yellow leaf curl virus*. Three classes of circular ssDNA satellites have been described associated with begomoviruses: alphasatellites, betasatellites, and deltasatellites.

Capulavirus

This genus includes four species, members of which have monopartite genomes. Members of the species *Alfalfa leaf curl virus* are transmitted by the aphid *Aphis craccivora*.

Curtovirus

This genus includes three recognized species, including *Beet curly top virus*, whose members are economically important pathogens in North America and Iran and have the widest host range of any geminivirus (over 300 species reported in over 40 dicot plant families). Members are transmitted by leafhoppers.

Eragrovirus

This genus includes a single species, *Eragrostis curvula streak virus*, members of which have monopartite genomes. All known isolates have been found infecting the monocot plant weeping lovegrass (*Eragrostis curvula*) in South Africa. The vector of the only known eragrovirus remains unknown.

Grablovirus

This genus includes three species, members of which have monopartite genomes and infect *Vitis* and *Prunus* spp. Members of *Grapevine red blotch virus* are transmitted by the three-cornered alfalfa treehopper, *Spissistilus festinus*.

Mastrevirus

This genus include 40 species, members of which have monopartite genomes, infect mostly monocots and some dicots, and are transmitted by several species of leafhoppers. It is worth mention *Maize streak virus* and *Chickpea chlorotic dwarf virus*, members of which cause severe diseases of maize in Africa and chickpea (**Fig. 1(D)**) and other crops in Africa and Asia, respectively. Recently, members of *Wheat dwarf India virus* have been shown to have associated alphasatellites and betasatellites.

Topocuvirus

Members of the single species in this genus, *Tomato pseudo-curly top virus*, have monopartite genomes, infect dicots and are transmitted by the treehopper *Micrutalis malleifera*.

Turncurtovirus

This genus include two species, members of which have monopartite genomes and have been isolated from symptomatic turnip (*Brassica rapa*) or radish (*Raphanus sativus*) plants from Iran, being transmitted by the leafhopper *Circulifer haematocaps*.

Also, two unassigned species in the family *Geminiviridae* have been recognized, *Citrus chlorotic dwarf associated virus* and *Mulberry mosaic dwarf associated virus*.

Phylogenetic analysis of complete genome sequences (DNA-A sequences for bipartite genomes) from isolates of representative species shows that geminiviruses group in clusters corresponding to the nine existing genera (**Fig. 2**). In addition, they are grouped to some extent according to geographic distribution, at least within the begomoviruses, probably reflecting their evolutionary divergence as a consequence of isolation. Begomoviruses can be classified into four lineages: Old World (OW) begomoviruses, New World (NW) begomoviruses, sweepoviruses and legumoviruses. OW and NW begomoviruses are originated from the Old World (Africa, Asia, and Europe) and the New World (The Americas), respectively. OW begomoviruses can have monopartite or bipartite genomes whereas most of the NW begomoviruses have bipartite genomes. A few exceptions of monopartite begomoviruses natives to the NW exist, including tomato leaf deformation virus, reported from Peru and Ecuador. Legumoviruses are restricted to legumes from Asia and Africa and most of them have bipartite genomes. Sweepoviruses infect sweet potato (*Ipomoea batatas*) and other members of the family Convolvulaceae and have monopartite genomes. Sweepoviruses seem to have diverged prior to the separation between the Old World and the New World begomovirus branches, seeming to represent one of the earliest points of divergence among this genus.

Genetic diversity and evolution of geminiviruses are driven by mutation, recombination and pseudorecombination (reassortment of genome components) which contribute significantly to the appearance of new viral variants, increasing their potential of adaptation to different hosts and environmental conditions. Recombination, a process in which exchange of genome segments occurs between DNA strands during replication, is common among members of the family *Geminiviridae*, notably among members of the genus *Begomovirus*, contributing greatly to genetic diversification of viral populations and speciation. Replication of geminiviruses, in addition to a rolling-circle replication (RCR) mechanism, also involves a recombination-dependent replication (RDR) mechanism. RDR provides a tool by which damaged or incomplete geminivirus DNA could be recovered for productive infection by homologous recombination and converted into full-size genomic DNA. The existence of this mechanism might explain in part the extent at which recombination is shaping geminivirus populations. Recombination in geminiviruses has been shown to occur either at the strain, species, genus and family levels. Many studies have reported that an important number of begomovirus epidemics are directly linked to recombinant viruses, for example, cassava mosaic disease in Africa, tomato yellow leaf curl disease in the Mediterranean basin and cotton leaf curl disease in the Indo-Pakistan sub-continent.

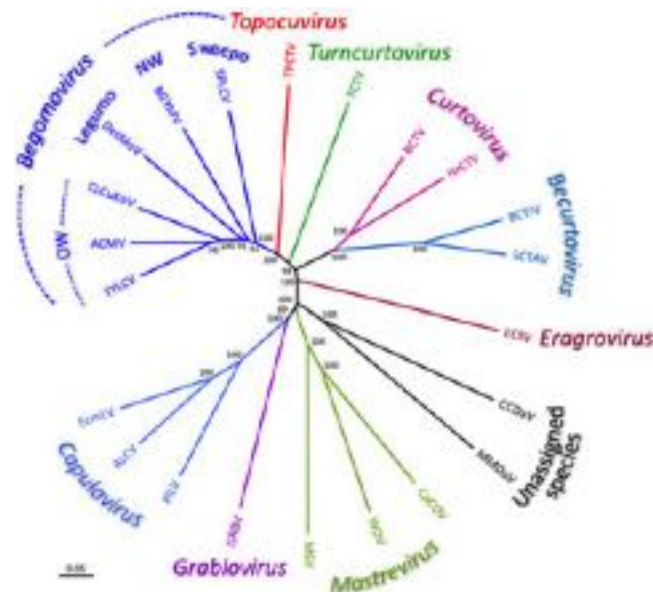


Fig. 2 Neighbor-joining phylogenetic tree of the full genome sequences (or DNA-A sequences for bipartite geminiviruses) of representative isolates of all genera (*Becurtovirus*, *Begomovirus*, *Capulavirus*, *Curtovirus*, *Eragrovirus*, *Grabovirus*, *Mastrevirus*, *Topocuvirus*, and *Turncurtovirus*) and unassigned species in the family *Geminiviridae*. OW (Old World), Legumo (legumoviruses), NW (New World) and Sweepo (sweepoviruses) are the four phylogenetic groups in the genus *Begomovirus*. Bootstrap values (1000 replicates) are shown. Viruses used to generate the phylogenetic tree are as follows: ACMV, African cassava mosaic virus; ALCV, Alfalfa leaf curl virus; BCTIV, Beet curly top Iran virus; BCTV, Beet curly top virus; BGYMV, Bean golden yellow mosaic virus; CCDaV, Citrus chlorotic dwarf associated virus; CLCuKov, Cotton leaf curl Kokhran virus; CpCDV, Chickpea chlorotic dwarf virus; DesMoV, Desmodium mottle virus; EcmLV, Euphorbia caput-medusae latent virus; ECSV, Eragrostis curvula streak virus; GRBV, Grapevine red blotch virus; HrCTV, Horseradish curly top virus; MMDaV, Mulberry mosaic dwarf associated virus; MSV, Maize streak virus; PILV, Plantago lanceolata latent virus; SCTAV, Spinach curly top Arizona virus; SPLCV, Sweet potato leaf curl virus; TCTV, Turnip curly top virus; TPCTV, Tomato pseudo-curly top virus; TYLCV, Tomato yellow leaf curl virus; WDV, Wheat dwarf virus. The bar below the tree indicates nucleotide substitutions per site.

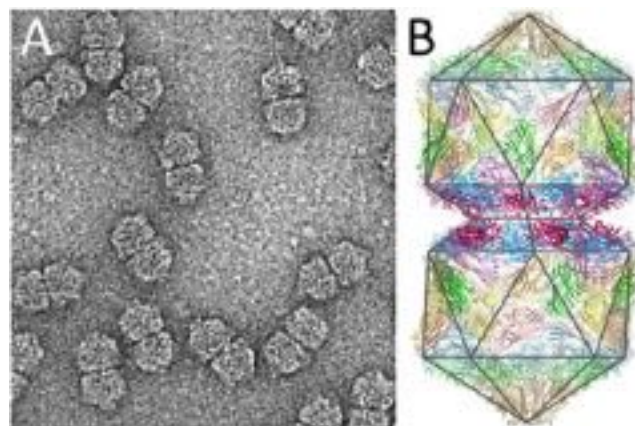


Fig. 3 (A) Electron micrograph of Maize streak virus virions negatively stained with uranyl acetate. (B) Complete atomic model for all 110 subunits in the capsid of Ageratum yellow vein virus obtained by cryo-electron microscopy at 3.3 Å resolution, with a polyhedral cage showing the symmetry of the particle. Reproduced with permission from (A) Zhang, W., Olson, N.H., Baker, T.S., *et al.*, 2001. Structure of the maize streak virus geminate particle. *Virology* 279, 471–477. (B) Hesketh, E.L., Saunders, K., Fisher, C., *et al.*, 2018. The 3.3 Å structure of a plant geminivirus using cryo-EM. *Nature Communications* 9, 2369, CC BY 4.0 licence.

Virion Structure

Geminiviruses have evolved a unique capsid structure. Virions are typically twinned or “geminate” particles of 22×30 nm which give the virus family its name. They consist of two joined, incomplete $T=1$ icosahedra (hemiscapsids) (Fig. 3). Cryo-electron microscopy (EM) has shown that virions of the begomovirus African cassava mosaic virus, and the mastrevirus maize streak virus containing 110 coat protein (CP) subunits organized as 22 pentameric capsomers. Cryo-EM structure at 3.3 Å resolution has further shown that the

capsid of the begomovirus *ageratum* yellow vein virus is built from three distinct conformations of a single CP, and that these conformational differences facilitate the formation of the interface between hemicapsids. Each twinned virion contains a single copy of circular ssDNA. Thus, for viruses with bipartite genomes, two virions containing different genomic components are required for infection. Defective genome components and ssDNA satellites associated with begomoviruses have also been found to be encapsidated.

Genome and Viral Proteins

Geminivirus genomes contain coding regions in both virion-sense and complementary-sense strands diverging from an intergenic region (IR) that includes the origin for RCR. Transcription is bi-directional, with independently controlled transcripts initiating within the IR. Multiple overlapping transcripts are used by geminiviruses for gene expression and transcript splicing is also used by members of the genus *Mastrevirus* and probably also by becurtoviruses, capulaviruses, and grabloviruses. The best-characterized genomes of geminiviruses are those of members of genera *Begomovirus*, *Curtovirus*, and *Mastrevirus*. Fig. 4 shows the genome organization of representative members of the nine genera in the family *Geminiviridae*. The genome of begomoviruses, the largest and more studied group in the family, is described more in detail below.

The genomes of bipartite begomoviruses consist of DNA-A and DNA-B components, each of 2.5–2.7 kb. Both genome components share an approximately 200 nt segment (common region, CR) within the IR that includes the origin of replication. The genomes of monopartite begomoviruses resemble the bipartite DNA-A component. DNA-A and monopartite genomes have six open reading frames (ORFs), two in the virion sense (AV1/V1 and AV2/V2) and four in the complementary sense (AC1/C1 to AC4/C4). DNA-B has two ORFs, the virion-sense BV1 and complementary-sense BC1. DNA-A and monopartite genomes encode a coat protein (CP, ORF AV1/V1), a protein putatively involved in viral movement (ORF AV2/V2, absent in bipartite New World begomoviruses), a replication-associated protein (Rep, ORF AC1/C1), a transcriptional activator (TrAP, ORF AC2/C2), a replication enhancer (REn, ORF AC3/C3), and C4 protein (ORF AC4/C4). DNA-B encodes a nuclear shuttling protein (NSP, ORF BV1) and a movement protein (MP, ORF BC1).

The CP is the only structural protein of geminivirus particles and, in addition to form the viral capsid, has been associated with other functions as insect transmission, shuttling of viral DNA into and out of the nucleus and cell-to-cell and systemic spread of the virus in for monopartite viruses. ORF AV2/V2 starts upstream ORF AV1/V1 and encodes for a protein whose exact function is uncertain, although in the case of at least some monopartite begomoviruses it is a movement protein. Rep is a multi-functional protein essential for viral DNA replication by initiating (endonuclease activity) and terminating (ligase activity) RCR. For that, Rep binds to dsDNA during origin recognition and introduces a nick in a highly conserved nonanucleotide (TAATATTAC) contained within a stem-loop structure that is part of the origin of replication. TrAP protein activates the transcription of the genes coding for CP and MP proteins. REn protein, although not essential for virus preplication, enhances viral DNA accumulation. ORF AC4/C4 is contained entirely within ORF AC1/C1, but in a different frame, and C4 proteins are the least conserved of all geminivirus proteins, having diverse functions including virus movement and are involved in symptom development. In addition to their specific functions, C4, TrAP and V2 proteins have been shown to suppress transcriptional (TGS) and post-transcriptional gene silencing (PTGS), whereas Rep protein suppresses TGS. In the DNA-B, BV1 encodes the NSP required for trafficking viral ssDNA between the nucleus and the cytoplasm in the form of a viral DNA–NSP complex. For cell-to-cell and long-distance movement, the NSP–viral DNA complex is trapped by the MP (encoded by BC1) in the cytoplasm and redirected to adjacent cells, where NSP directs the viral genome to the nucleus to initiate replication again.

DNA Satellites Associated With Geminiviruses

Three types of circular single-stranded DNA satellites are associated with begomoviruses, alphasatellites, betasatellites, and deltasatellites. Alphasatellites and betasatellites are about half size begomovirus genome components whereas deltasatellites are about half size alphasatellites and betasatellites. Alphasatellites are mainly associated with monopartite Old World begomoviruses but also with some bipartite New World begomoviruses. They encode a Rep protein needed for their replication with similarity with the Rep encoded by nanoviruses (family *Nanoviridae*). Betasatellites are associated with many monopartite Old World begomoviruses and are essential for induction of typical disease symptoms. The β C1 protein encoded by the betasatellites has important roles in symptom induction, suppression of transcriptional and post-transcriptional gene silencing, and can affect jasmonic acid responsive genes. Only in one case alphasatellites and betasatellites have been found associated with a *mastravirus*, Wheat dwarf India virus. Deltasatellites have been found associated with bipartite Old World begomoviruses, monopartite Old World begomoviruses and sweepoviruses. Deltasatellites do not encode for any protein and, although in some cases reduce the accumulation of the helper begomovirus, rarely modify the symptoms caused by the begomovirus. All these satellites depend on their helper begomoviruses for replication (except alphasatellites), cell-to-cell and systemic spread throughout the plant, encapsidation, and transmission to new host plants by insect vectors.

Diagnosis

There are a number of leaf symptoms typically associated with geminivirus – mainly begomovirus – infection, including yellow and golden mosaics, leaf curl and deformation and vein yellowing. Also, monocots infected with *mastraviruses* use to show typical



Fig. 4 Genome organization of representative members of the nine genera in the family *Geminiviridae*. ORFs are color-coded according to the function of their protein products (rep, replication-associated protein; ren, replication enhancer protein; trap, transcriptional activator protein; cp, capsid protein; mp, movement protein; nsp, nuclear shuttle protein). Old World bipartite begomoviruses contains ORF AV2, not shown in the figure. LIR, long intergenic region; SIR, short intergenic region; CR, common region. The hairpin which includes the origin of replication is indicated in the LIR. Reproduced with permission from Varsani, A., Roumagnac, P., Fuchs, M., *et al.*, 2017. Capulavirus and Grablovirus: Two new genera in the family Geminiviridae. Archives of Virology 162, 1819–1831.

leaf streaks and striates. However, these symptoms, although suggestive of geminivirus infection, may be caused by other viruses, and rarely they can be used to identify the specific virus involved. Thus, diagnosis based on symptoms alone is not sufficient and specific diagnostic tests must be used to confirm geminivirus infection. For some geminiviruses, such as the begomoviruses *African cassava mosaic virus* and *Tomato yellow leaf curl virus* or the curtovirus *Beet curly top virus*, antibodies against the coat protein have been generated to be used in serological tests. However, although in some cases antisera have been commercialized, they are not commonly used for diagnosis because the presence of common epitopes in the coat protein hinder them to be used for specific virus identification.

DNA-based methodologies have allowed the development of rapid and specific tests for detection and characterization of geminiviruses. From dot- and squash-blot hybridization using cloned geminivirus DNAs as probes to polymerase chain reaction (PCR) technologies, a wide range of methods have revolutionized the diagnosis of geminiviruses. PCR can be considered currently the method of choice for geminivirus detection due to sensitivity, specificity and rapidity. Also, PCR can be adapted to detection of a specific virus or a group of them. Sequencing of PCR products can be used later for precise identification of the virus. Loop-mediated isothermal amplification has been also used to develop tests for a number of geminiviruses.

Rolling circle amplification (RCA), which amplifies circular DNA molecules in a nonspecific manner using random primers and $\phi 29$ DNA polymerase, coupled with restriction enzyme digestion, has resulted recently in a valuable and powerful tool for cloning of uncharacterized geminiviruses. More recently, next-generation sequencing (Roche 454, Illumina, Ion Torrent, etc.) has become increasingly accessible, allowing characterization of mixed infections and readily discovery of novel geminiviruses. Nanopore-based platforms, including portable versions, are also being used in metagenomic studies applied to the discovery of geminivirus diversity.

Life Cycle

Geminiviruses, exemplified by begomoviruses, initiate infection when an insect vector (whitefly) carrying the virus feeds on the sap transported through the phloem of a healthy leaf and transmits virions to phloem-associated cells (Fig. 5). In the plant cell, viral ssDNA is released from virions and, since geminiviruses do not encode a DNA polymerase, it is copied by host DNA polymerases to generate double-stranded (ds) DNA. The dsDNA is transcribed by host RNA polymerase II, allowing the production of Rep. This protein initiates viral replication, which occurs by a combination of RCR and RDR mechanisms. Nascent circular ssDNA can be converted to dsDNA to re-enter the replication cycle. Later in infection, Rep represses its own transcription, leading to activation of transcriptional activator protein (TrAP) expression, which in turn activates coat protein (CP) and nuclear shuttle protein (NSP) expression. Circular ssDNA can then be encapsidated by CP into virions, which are available for insect vector acquisition. The infection is propagated inside the plant by the movement of viral DNA out of the nucleus into the next cell or the phloem through the action of two viral movement proteins, nuclear shuttle protein (NSP) and movement protein (MP).

An important phase of the begomovirus life cycle is the transmission between plants by the insect vector, the whiteflies of the *B. tabaci* complex. For that, begomoviruses need first to cross the midgut wall to reach the hemolymph and then accumulate in the primary salivary glands, from where they are secreted in the saliva to infect new plants. The transport across midgut wall and

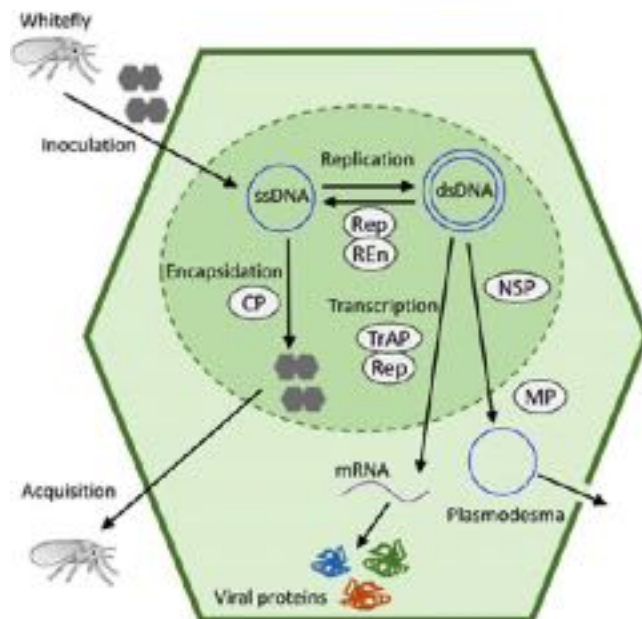


Fig. 5 The geminivirus life cycle, exemplified by begomoviruses. See main text for details.

through primary salivary glands are key steps determining differential transmission in different whitefly-begomovirus combinations. Although whether circulative begomovirus transmission is non-propagative (the virus does not replicate in the whitefly) or propagative (the virus replicates in the whitefly) has been a topic of debate, there is accumulating evidence that at least Tomato yellow leaf curl virus (TYLCV) is able to replicate within *B. tabaci*. Certain begomoviruses have been also shown to be transmitted transovarially to the descendance. From the virus side, the CP has been found to be the sole determinant of the persistent transmission of begomoviruses by *B. tabaci*. Inside the whiteflies, the transmission also involves certain endosymbiotic bacteria. For instance, it has been shown that transmission of TYLCV depends on chaperonin GroEL homologs produced by endosymbiotic bacteria, which have been proposed to protect the virus from degradation during passage through the hemolymph.

Pathogenicity

Although the information about how geminiviruses cause diseases in their host plants is very limited, there are a number of cases in which the mechanisms involved in the production of symptoms has been studied. Determinants of pathogenicity include C4, TrAP, V2, and NSP proteins encoded by different begomoviruses and curtoviruses. C4 encoded by the curtovirus Beet curly top virus, for example, is the major determinant of the characteristic vein swelling phenotype observed in infected plants caused by abnormal cell division. In this case, C4 induces the expression of RKP, a RING finger E3 ligase, which may be involved in cell cycle regulation. Another well studied case is that of NSP encoded by the begomovirus Tomato leaf curl New Delhi virus (ToLCNDV) which is an avirulence determinant inducing a hypersensitive response (HR) in tomato (*Solanum lycopersicum*) and tobacco (*Nicotiana tabacum*). ToLCNDV TrAP is also a pathogenicity factor which may counter NSP-induced HR in infected cells. The involvement of NSP in virus pathogenicity is supported by its interactions with membrane receptor kinases implicated in a wide range of signal transduction pathways. On the other hand, TrAP compromises the ability of COP9 signalosome (a protein complex that functions in the ubiquitin-proteasome pathway) to bind to Cullin-1 (CUL1), an essential component of the SCF (SKP1, CUL1/CDC53, F box proteins) ubiquitin E3 ligase complex; TrAP-COP9 interaction alters the cellular processes regulated by SCF complexes, including jasmonate signaling, thereby regulating host response to infection.

Control

Many factors affect the occurrence and prevalence of geminivirus diseases and the losses they cause to crops worldwide. As for plant viruses in general and particularly for those transmitted by an insect vector, control of diseases caused by geminiviruses would benefit from the application of integrated pest management (IPM) strategies, which take in account all those factors. Some of the general strategies that can be used with more or less success to manage geminivirus diseases and epidemics are considered below.

One of the most desirable situation is the availability of resistant or tolerant cultivars to a given geminivirus disease, as this can provide complete protection. These materials are obtained by conventional breeding methods by introgressing pertinent genes in commercial cultivars usually from wild relative species. Crops for which there are materials with effective resistance to specific begomovirus diseases include cassava (African cassava disease), cotton (cotton leaf curl disease) and tomato (tomato yellow leaf curl disease).

When resistant cultivars are not available, simple control measures can be taken in order to initiate crops with virus-free planting materials, thus reducing the amount of primary inoculum. In the case of vegetable propagated from seeds, whenever possible, transplants should be produced in greenhouses instead of in open fields to protect them from insect vectors. An extreme case of this approach is to protect the crop during the entire season, as it is done when high-value vegetables are grown in greenhouses. For crops propagated from cuttings, as cassava or sweet potato, mother plants should be selected based on absence of symptoms and molecular tests to discard infected plants. Another action to be taken in annual crops is to establish them at the moment when the source of inoculum and the populations of the insect vector are the lowest possible, thus delaying infection and providing protection to young, highly susceptible, plants. In some cases, mandatory host-free periods can be established in a given area. Such host-free periods have been implemented for example in Costa Rica and Brazil to protect tomato crops from begomovirus infections.

Just as vector insects play a key role in the life cycle of geminiviruses, their control is also essential when controlling the diseases that these viruses cause. A first measure, mostly used by farmers in many circumstances, is the use of insecticides to reduce vector populations. Options include systemic insecticides such as the neonicotinoids, contact insecticides, and insect growth regulators. An important limitation to the repeated application of the same active ingredient in a given area is that insect populations with high resistance to various insecticides can readily emerge, as it has been largely documented for certain species of the *B. tabaci* complex. Even though an insect population is susceptible to insecticide treatment, the effect may not be rapid enough to avoid virus transmission, thus limiting the practical effectiveness of this strategy for the control of geminivirus diseases.

An alternative to chemical insect vector control is biological control using natural enemies, mainly parasitoids and predators. For management of whiteflies and begomovirus diseases in protected crops, numerous commercial preparations are available including the parasitoids *Encarsia formosa* and *Eretmocerus eremicus* and the predators *Nesidiocoris tenuis*, *Macrolophus pygmaeus* and *Amblyseius swirskii*.

Two biotechnological approaches have also been evaluated to generate geminivirus-resistant crops, the production of transgenic plants and the most recent strategy of genome editing using the CRISPR-Cas system. Although many examples of transgenic

plants resistant to geminiviruses have been evaluated and proved to be effective, mostly using gene silencing by RNA interference, these plants have not been commercialized due to controversial public acceptance. Although not yet materialized, commercialization of transgenic bean plants resistant to the begomovirus Bean golden mosaic virus have been approved in Brazil. The CRISPR-Cas methodology has been evaluated against geminiviruses by editing the viral genome or host genes encoding factors required for the virus life cycle.

Further Reading

- Brown, J.K., Zerbini, F.M., Navas-Castillo, J., *et al.*, 2015. Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. *Archives of Virology* 160, 1593–1619.
- Fiallo-Olivé, E., Pan, L.L., Liu, S.S., Navas-Castillo, J., 2019. Transmission of begomoviruses and other whitefly-borne viruses: Dependence on the vector species. *Phytopathology* 110, 10–17.
- Fondong, V.N., 2013. Geminivirus protein structure and function. *Molecular Plant Pathology* 14, 635–649.
- García-Arenal, F., Zerbini, F.M., 2019. Life on the edge: Geminiviruses at the interface between crops and wild plant hosts. *Annual Review of Virology* 6, 411–433.
- Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D., Mansoor, S., 2013. Geminiviruses: Masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology* 11, 777–788.
- Hesketh, E.L., Saunders, K., Fisher, C., *et al.*, 2018. The 3.3 Å structure of a plant geminivirus using cryo-EM. *Nature Communications* 9, 2369.
- Lefevre, P., Moriones, E., 2015. Recombination as a motor of host switches and virus emergence: Geminiviruses as case studies. *Current Opinion in Virology* 10, 14–19.
- Lozano, G., Trenado, H.P., Fiallo-Olivé, E., *et al.*, 2016. Characterization of non-coding DNA satellites associated with sweepoviruses (genus *Begomovirus*, *Geminiviridae*) – Definition of a distinct class of begomovirus-associated satellites. *Frontiers in Microbiology* 7, 162.
- Navas-Castillo, J., Fiallo-Olivé, E., Sánchez-Campos, S., 2011. Emerging virus diseases transmitted by whiteflies. *Annual Review of Phytopathology* 49, 219–248.
- Rojas, M.R., Macedo, M.A., Maliano, M.R., *et al.*, 2018. World management of geminiviruses. *Annual Review of Phytopathology* 56, 637–677.
- Saunders, K., Bedford, I.D., Yahara, T., Stanley, J., 2003. The earliest recorded plant virus disease. *Nature* 422, 831.
- Varsani, A., Navas-Castillo, J., Moriones, E., *et al.*, 2014. Establishment of three new genera in the family *Geminiviridae*: *Becurtovirus*, *Eragrovirus* and *Turncurtovirus*. *Archives of Virology* 159, 2193–2203.
- Zerbini, F.M., Briddon, R.W., Idris, A., *et al.*, 2017. ICTV virus taxonomy profile: *Geminiviridae*. *Journal of General Virology* 98, 131–133.
- Zhou, X., 2013. Advances in understanding Begomovirus satellites. *Annual Review of Phytopathology* 51, 357–381.

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/geminiviridae
ICTV Report
Geminiviridae.
- <https://talk.ictvonline.org/taxonomy/>
International Committee on Taxonomy of Viruses
Taxonomy.

Hordeiviruses (*Virgaviridae*)

Zhihao Jiang, Meng Yang, Yongliang Zhang, Andrew O Jackson, and Dawei Li, China Agricultural University, Beijing, China

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J.N. Bragg, H.-S. Lim, A.O. Jackson, Hordeivirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00426-X](https://doi.org/10.1016/B978-012374410-4.00426-X).

Glossary

Autophagy A conserved eukaryotic mechanism that removes damaged or dysfunctional components.

CRISPR/Cas9 A genome editing tool that enables directly edit the genome by removing, adding or altering sections of the DNA sequence.

Cryo-EM (Cryo-electron microscopy) An electron microscopic technique that involves freezing and applies on protein structure analysis.

PTM (Posttranslational modification) A biochemical modification that occurs to amino acid(s) on a protein after the protein has been translated.

ROS (Reactive oxygen species) Chemically reactive oxygen radicals as well as non-radical derivatives of oxygen.

VIGS (Virus-induced gene silencing) A reverse genetics tools for gene function analysis that uses viral vectors carrying a target gene fragment to produce dsRNA which initiates the silencing of the target gene.

VSR (Viral suppressor of RNA silencing) Viral proteins which can counteract the host silencing-based antiviral process.

Introduction

The genus *Hordeivirus* has been assigned to the family *Virgaviridae*. Hordeiviruses represent a unique genus of serologically related viruses with rigid rods that are composed of 96% protein and 4% RNA. Although members of the genus *Hordeivirus* have similar particle structure, they collectively infect both monocots and dicots and exhibit considerable biological diversity. *Barley stripe mosaic virus* (BSMV), the type member of the genus, can be traced back to ~750 years ago at a site near the Nile River in modern-day Egypt, and the virus has subsequently spread around the world via global movement of seeds. BSMV has been known to cause serious worldwide disease problems in cultivated barley for more than 85 years, and during this time has also been isolated infrequently from wheat and wild oats. Because of the yield losses caused in barley, a number of strains of the virus have been isolated and their biological properties evaluated. The survival of BSMV in nature depends solely on direct plant-to-plant contact and seed transmission; consequently, virus spread can be prevented by planting virus-free seed. The advent of sensitive virus detection methods has permitted seed to be screened prior to planting, thus virus-free seed production has resulted in eradication of the virus in most areas of North America, Europe, and Asia.

Taxonomy and Characteristic

In addition to Barley stripe mosaic virus, three additional species, Poa semilatifolia virus (PSLV), *Lychnis* ringspot virus (LRSV), and Anthoxanthum latent blanching virus (ALBV) have been included in the genus *Hordeivirus* in the Eighth Report of the International Committee on Taxonomy of Viruses (ICTV). This assignment is based on serological relatedness of the coat proteins, genome organization, and sequence relatedness of BSMV, LRSV, and PSLV. The coat proteins of BSMV, LRSV, and PSLV are similar in size, but have different mobilities on polyacrylamide gels and are distantly related serologically. Serology and analyses of coat protein sequences indicate that BSMV and PSLV are more closely related to each other than to LRSV. Serological studies also show that ALBV is closely related to BSMV, and may be a strain of BSMV. However, ALBV has not been investigated by hybridization or sequence analysis to determine whether it is a strain of BSMV or is a distinct virus.

PSLV has been recovered on two occasions from native grasses in widely separated locations within Canada, and from one location in Hungary, throughout North America and Europe, so the virus may be relatively abundant amongst members of the *Poaceae* throughout North America and Europe. LRSV was first isolated from *Lychnis divaricata* seeds imported into the USA, and an additional mild strain of LRSV has been isolated in Hungary. Infectious cDNA clones of LRSV and PSLV have been constructed and used to compare some biological and biochemical properties with BSMV. ALBV has been described only in Great Britain and little is known about its properties. As is the case with BSMV, LRSV is highly seed transmitted, but seed transmission has not been reported for PSLV or ALBV. All of the viruses reach high concentrations in infected plants and are easily transmitted from plant to plant by mechanical contact.

In addition to their limited natural host ranges, numerous experimental hordeivirus hosts have been identified by mechanical inoculation and agro-infiltration of recombinant DNA clones. BSMV has been shown to infect several monocots and a few dicots, and recently, *Brachypodium distachyon*, wild emmer (*Triticum turgidum* var. *dicoccoides*), *Dasyphyrum villosum*, ginger (*Zingiber officinale* and *Z. zerumbet*) have been shown to support BSMV replication. Local lesions on *Chenopodium* species have been used for several

genetic studies of BSMV disease phenotype, and systemic *Nicotiana benthamiana* infections have provided a basis for analyses of virus life cycle and virus-host interactions in dicot hosts. PSLV infects several monocots while LRSV has been shown to infect several experimental dicot hosts, including several *Nicotiana* species and members of the family *Caryophyllaceae*. Reverse genetic systems for BSMV, PSLV, and LRSV have recently resulted in considerable information about biological properties of the viruses and the molecular biology of their infection processes.

Virus Particle Structure

The CP interacts with the genomic RNAs to form short rod-shaped particles and always undergo end-to-end aggregation during extraction *in vitro*. The length of BSMV virions range from about 110–150 nm, similar with LRSV but 20% shorter than PSLV. A cryo-EM structure of BSMV virion suggests that BSMV has an average of 23.2 subunits per helix turn and has two types of virions. A wide virion form has an additional CP per turn compared with the narrow form, and has specific contact differences between the CP subunits. The diameters of the wide and narrow virions are 22.4 and 21.6 nm, respectively, and the corresponding helical parameters are 106 and 111 subunits/period. In addition, the inter-subunit of the narrow virion has two more salt bridges than the wide virion. In agreement with their serological relatedness, the BSMV and PSLV coat protein sequences are more similar to each other than to the CP of LRSV. Although the CP shares several conserved motifs with other rod-shaped viruses, the hordeivirus CPs are closely related to those of members of the fungal-transmitted virus species *Peanut clump virus* (PCV) and *Indian peanut clump virus* (ICPV), but have limited similarity to *Soil-borne wheat mosaic virus* (SBWMV).

Genome Structure and Expression

Hordeiviruses are composed of three genomic (g) RNAs that have been designated α , β , and γ based on hybridization of the viral RNAs. The sizes of the BSMV α , β , and γ gRNAs vary between strains, with the ND18 strain sizes being 3.8, 3.2, and 2.8 kb, respectively. RNAs α and β are similar in size among strains of BSMV, but the sizes and complexity of the γ RNAs vary considerably. The PSLV and LRSV α , β , and γ gRNAs are similar in size (3.9, 3.6, and 3.2 kb, and 3.8, 3.0, and 2.7 kb, respectively) to those of BSMV.

Each of the hordeivirus RNAs has a 7-methylguanosine cap at the 5' terminus, and amongst the gRNAs within a species, the 5' untranslated regions (UTRs) that are necessary for replication have very little sequence similarity. Obvious sequence similarity is not evident among the 5' UTRs of the α and γ RNAs of BSMV, PSLV, and LRSV, but the RNA β components of the three viruses have considerable sequence similarity. The hordeiviruses gRNAs also contain a conserved 3' terminal non-translated sequence that is required for replication. The 3' termini are variable in length between the different viruses, ranging from 184 nt in LRSV to 290 ~ 330 nt in PSLV. The sequences form tRNA-like structures that in BSMV and PSLV are capable of binding tyrosine *in vitro*. Sequence comparisons indicate that the structures are very similar except that the 3'UTRs upstream of the conserved tRNA regions of the PSLV and BSMV RNAs have a large stem loop element and an array of possible pseudoknot stem loop regions that make the 3' UTR longer than that of LRSV. An internal poly(A) sequence of varying length is located directly after the stop codon of the 3' proximal gene of each RNA, and resides immediately upstream of the conserved 3' tRNA termini in the BSMV and LRSV genomes. Intriguingly, the poly(A) sequence is absent in the PSLV genome.

Hordeivirus gRNAs (Fig. 1) encode seven proteins. The replicase protein subunits are encoded by the α and γ gRNAs. RNA α encodes a single protein (α a) that contains methyltransferase and helicase domains that are highly conserved between Sindbis-like viruses. The bicistronic γ RNAs encodes the polymerase (γ a) subunit, which contains the conserved GDD motif found in RNA-dependent RNA polymerase (RdRp) proteins of RNA viruses, plus a 3' encoded cysteine-rich protein (γ b).

RNA β encodes the coat protein in the first open reading frame (ORF), and is followed by a series of overlapping ORFs termed the 'triple gene block' (TGB) that are also present in alexi-, beny-, carla-, fovea-, peclu-, pomo-, and potexviruses. The first ORF of the TGB encodes the TGB1 protein (formerly designated β b). The remaining two TGB ORFs encode the small hydrophobic proteins, TGB2 and TGB3 (formerly called β d and β c, respectively).

The α a, β a (coat protein), and γ a proteins encoded by gRNA α , gRNA β , and gRNA γ , respectively, are translated directly from the gRNAs (Fig. 1). Expression of the TGB proteins is mediated by transcription of two subgenomic (sg) RNAs, designated sgRNA β 1 and sgRNA β 2. TGB1 is translated from sgRNA β 1, and the other overlapping proteins, TGB2 and TGB3, are expressed from sgRNA β 2. Despite sequence predictions, *in vitro* translation assays, and barley protoplasts showing that the TGB2' protein is translated from sgRNA β 2, the TGB2' inactivated mutant has no detectable effects on BSMV infection in *N. benthamiana*. The γ b protein is translated from sgRNA γ . Each of the encoded proteins is expressed at different levels and times during the life cycle, and the three sgRNAs have similar variances in their abundance during replication. High-level constitutive expression of sgRNA γ occurs throughout the replication cycle, but expression of sgRNA β 1 and sgRNA β 2 is temporal. In addition, sgRNA β 1 is considerably more abundant than sgRNA β 2. The three sgRNA promoters have little obvious sequence relatedness; However, the individual BSMV, LRSV, and PSLV sgRNA promoters share a number of blocks of conserved sequence. Pairwise combinations indicate that the BSMV and PSLV promoters have the highest conservation, whereas the LRSV promoters have undergone more divergence. These results and the protein comparisons described below support existing evidence for a common origin of the hordeiviruses and buttress biological evidence suggesting that BSMV and PSLV are more closely related to each other than to LRSV. Additional comparisons of

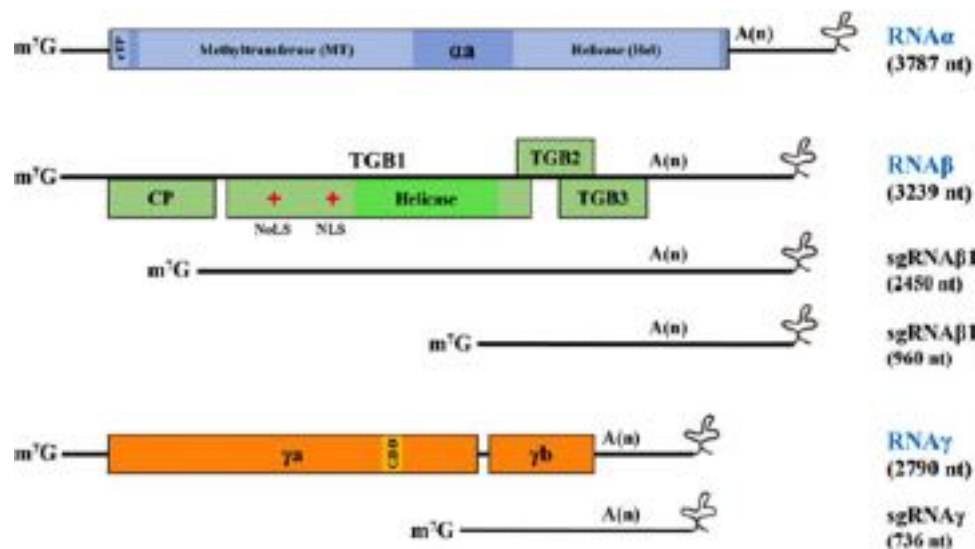


Fig. 1 The tripartite genome of BSMV and sgRNAs used for expression of the genes encoded by the β and γ RNAs. Each genomic (g) RNA and sub-genomic (sg) RNA contains a 5' cap structure (m^7G) and an internal polyadenylate sequence (A_n) preceding a 238 nt tRNA-like region that terminates with a UAA sequence. The three gRNAs are designated α , β , and γ and contain open reading frames (ORFs) represented by rectangular blocks. RNA α consists of a single ORF (α) that is translated directly from the gRNA. The α protein forms the 'helicase subunit' of the RNA-dependent RNA polymerase (RdRp) and contains characteristic chloroplast transit peptides (cTP), methyltransferase (MT) and helicase (Hel) motif signatures. RNA β encodes the 5' proximal coat protein (CP) and three overlapping triple gene block (TGB) proteins (TGB1, TGB2, and TGB3), that are each required for cell-to-cell movement. The CP ORF is translated directly from gRNA β and the TGB proteins are expressed from two subgenomic (sg) RNAs, sgRNA β 1 and sgRNA β 2, that are illustrated below gRNA β . TGB1 contains two positively charged regions, NoLS and NLS, toward the N-terminus of the protein and a Hel domain, whereas TGB2 and TGB3 are small hydrophobic transmembrane proteins. The bicistronic RNA γ encodes the 5' γ a GDD-containing polymerase protein subunit, and the 3' proximal γ b cysteine-rich pathogenesis protein. The γ a protein is translated directly from gRNA γ and the γ b protein is expressed from sgRNA γ b.

the TGB promoter sequences of several other viruses possessing TGB movement proteins have revealed no strong correlations in sequence or secondary structure between these putative promoter sequences and those of BSMV.

Replication of BSMV gRNAs requires the 5' and 3' terminal regions, but additional internal *cis*-elements are required for replication of each of the RNAs. RNA α replication is *cis*-preferential, with replication appearing to be coupled to translation of a functional α protein. The 117 nt intergenic region separating the β a (coat protein) ORF and the TGB1 ORF has a *cis*-acting function required for replication. In contrast, RNA γ replication is not dependent upon the 42 nt intergenic region separating γ a and γ b, but essential *cis*-acting elements are present in the first 500 nt of the γ a gene.

Function and Relatedness of the Hordeivirus Proteins

Replicase Proteins (α a and γ a)

BSMV gRNAs expressing α a and γ a are able to replicate in protoplasts. The α a and γ a proteins are essential subunits of the RdRp and are members of the tobamo-lineage of supergroup III RdRps. A histidine-tagged γ a protein recovered from infected barley fractionates with the α a protein and the recovered complex exhibits BSMV RNA-specific polymerase activity. Therefore, the α a and γ a proteins constitute the helicase and polymerase subunits of the RdRp complex.

The α a protein contains amino-terminal methyltransferase and carboxy-terminal NTPase/helicase domains. The methyltransferase domain likely functions in capping of the viral RNAs and the DEAD box helicase domain consists of at least six amino acid motifs that are conserved among the RdRps of different groups of RNA viruses. Software predictions show that α a has a chloroplast-targeted transit peptides (cTP) at the amino-terminus, and both subcellular localization assays and immuno-gold labeling experiments demonstrate that BSMV α a is localized on the chloroplast surface. A subcellular localization assay has also shown that LRSV α a has a chloroplast localization signal and targets to the chloroplasts when transiently expressed in *N. benthamiana*.

The γ a RdRp hordeivirus subunits contain a GDD polymerase motif toward their C-termini that is characteristic of the supergroup III polymerases, and this domain falls into the 'tobamo-lineage' among the plant viruses. Phylogenetic comparisons of hordeivirus polymerase subunits reveal more than 75% conservation within the genus and lower levels of conservation extend into members of the genera *Pecluvirus*, *Furovirus*, *Tobravirus* (> 65%), and *Tobamovirus* (> 30%). Sequence comparisons of the NTPase, helicase, and GDD motifs have revealed significant conservation among the polymerase subunits of Sindbis-like viruses (supergroup III). Among the plant

viruses within this group, the conserved domains of BSMV and PSLV are more closely related than those of LRSV, and these sequences also share close relatedness with viruses in the genera *Furovirus* and *Pomovirus*. Lower levels of sequence relatedness are observed when comparisons are extended to the genera *Tobravirus* and *Tobamovirus* and members of the family *Bromoviridae*.

Coat Protein

The coat protein (CP) is translated directly from gRNA β and is the most abundant of the viral proteins in infected plants. The CPs of all hordeiviruses are approximately 22 kDa in size, but exhibit different electrophoretic mobilities.

Studies of BSMV show that the CP is not essential for infectivity of two systemic hosts (barley and *N. benthamiana*) or *Chenopodium* species that form local lesions. An LRSV CP deletion mutant is also dispensable for systemic infection of *N. benthamiana*. Moreover, infections elicited by CP-deficient BSMV mutants appear to be more aggressive, and the phenotype is more severe and protracted in barley and *N. benthamiana* than derivatives expressing the CP. Plants infected with CP deletion mutants have serious mosaic and deformity symptoms in systemically infected leaves and appear to lack the recovery phase of infection that is usually observed in barley and *N. benthamiana* inoculated with BSMV. Molecular analysis also reveal that CP-deficient mutants have higher levels of viral RNA accumulation in barley and *N. benthamiana* than wild type BSMV, which suggests that the CP may affect replication through a feedback mechanism during the late stages of the virus life cycle when RNA replication switches to virion assembly.

TGB Movement Proteins

The hordeiviruses encode Class I TGB proteins that differ in several features from the Class II TGB proteins of *Potato virus X* (PVX) and other viruses of the family *Alphaflexiviridae*. Coordinated actions of each of the three TGB proteins are required for cell-to-cell and systemic movement. The proteins are expressed transiently and simultaneously during the early stages of infection and decline in abundance at later stages of infection.

TGB1

The TGB1 protein, encoded by the 5' proximal TGB gene, is expressed from sgRNA β 1 and accumulates to high levels early in infection. Among the hordeivirus species, the TGB1 proteins range from 50 to 63 kDa in size. These proteins can be distinguished from the ~25 kDa Class II TGB proteins of the *Alphaflexiviridae* by having substantially larger N-terminal domains preceding the helicase domain. The N-terminal half of the hordeivirus TGB1 proteins contains two positively charged regions rich in lysine and arginine residues. The N-terminal region also contains a nuclear localization signal (NLS) and a localization signal (NoLS), both of which are required for nuclear-cytoplasmic trafficking of TGB1 and viral cell-to-cell movement. The C-terminal half of the TGB1 protein contains an NTPase/helicase domain with seven conserved motifs (I, IA, II, III, IV, V, and VI) that are characteristic of superfamily I helicases of alpha-like viruses. The TGB1 NTPase/helicase domain is similar to the helicase domain present in the α a protein and is also essential for BSMV movement. The PSLV TGB1 protein has RNA helicase activity *in vitro* that is dependent on Mg²⁺ and ATP and can unwind RNA duplexes in both 5' - to - 3' and 3' - to - 5' directions. The BSMV and PSLV TGB1 proteins also bind ATP and dATP, and exhibit ATPase activity that maps to motifs I, IA, and II in the helicase domain.

The BSMV TGB1 protein has RNA-binding activity at high ionic strength and exhibits a high binding affinity for both double-stranded (ds) RNAs and single-stranded (ss) RNAs. RNA sequence specificity has not been detected in binding studies, and TGB1 has little detectable affinity for DNA. The TGB1 proteins of both BSMV and PSLV contain multiple RNA-binding regions, and in PSLV, the N-terminal half of TGB1 containing the positively charged regions is able to bind RNA under high-salt conditions, whereas the C-terminal half containing the helicase motif binds RNA only under low-salt conditions.

A BSMV TGB1/RNA ribonucleoprotein (RNP) complex has been purified from infected barley, demonstrating that TGB1 also binds to RNA *in vivo*. In addition, the TGB1 protein also participates in both homologous interactions and TGB3 binding. When GFP-TGB1 is transiently expressed in the absence of TGB2 or TGB3, the TGB1 protein is localized in cytoplasmic membranes. However, when GFP-TGB1 was expressed *in cis* during BSMV infection, fluorescence was evident on opposite sides of the cell wall, suggesting that TGB1 localizes to the plasmodesmata (PD). Due to its RNA-binding activities and presence in an RNP complex recovered from BSMV-infected plants, the TGB1 protein appears to function in formation of movement complexes that are transported to plasmodesmata and to adjacent cells during interactions with TGB2 and TGB3. Additionally, TGB1 interacts directly with a host factor, nucleolar protein fibrillarin (Fib2), and is a key component of a BSMV ribonucleoprotein (RNP) movement complex that functions in cell-to-cell movement.

TGB2 protein

Among the hordeivirus TGB proteins, the TGB2 protein contains the highest amount of sequence similarity within the hordeiviruses, and the degree of sequence similarity extends to the class I TGB2 proteins encoded by PVX and other members of the family *Alphaflexiviridae*. The TGB2 sequence contains two hydrophobic stretches and a conserved hydrophilic region that separates the hydrophobic residues. The hydrophobic regions are predicted to integrate into membranes with a U-shaped topology that directs the termini of the proteins to the cytoplasmic side of the membrane, and the strong conserved central portion of the protein is exposed to the ER lumen.

When expressed alone, the GFP-TGB2 protein is associated with the cortical ER where actin and TGB3 interactions result in dramatically increased filament thickening at the infection front of *N. benthamiana* leaves and motile vesicles resembling Golgi stacks that co-localize to plasmodesmata. However, in BSMV RNA α and RNA γ infections, ectopically expressed GFP-TGB2 proteins associated with chloroplasts, which suggests that TGB2 may have unknown roles in chloroplast infection processes.

TGB3 protein

The TGB3 protein is encoded by the 3' proximal ORF of sgRNA β 2 and is translated by leaky scanning of ribosomes past the TGB2 AUG, which is in a poor context for translation initiation. This mechanism normally produces about a 100:10:1 ratio of TGB1:TGB2:TGB3 during *in vitro* translation. The TGB3 proteins are the least conserved of the TGB proteins, and the class 1 and 2 TGB3 proteins are thought to have a polyphyletic origin, in contrast to the TGB2 proteins that are thought to have originated from a common ancestor.

The N-termini of the approximately 17 kDa TGB3 proteins from each of the hordeivirus species contains a stretch of conserved amino acids that includes one histidine and three cysteine residues. The TGB3 protein has two hydrophobic domains separated by a hydrophilic region that contains a conserved tetrapeptide (QDLN) sequence. The transmembrane regions are predicted to direct the TGB3 termini toward the ER lumen. This topology is quite distinct from that of the class II TGB3 proteins of the alphaflexiviruses, which have only a single membrane-spanning domain that results in localization of the C-terminus on the cytoplasmic side of the ER.

When expressed alone, TGB3 localizes to paired peripheral bodies at the cell wall (CW) that appear to lie on opposite sides of the PD channels connecting adjacent cells. The targeting of PSLV TGB3 to the CW requires the central hydrophilic and the C-terminal trans-membrane regions, whereas the BSMV TGB3 protein requires the five C-terminal residues for PD localization. Furthermore, protein interaction studies demonstrate that TGB3 binds both TGB1 and TGB2, and that TGB3 directs targeting of co-expressed TGB1 and TGB2 to paired peripheral bodies at the CW.

Pathogenesis Protein (γ b)

The 3' proximal ORFs of the hordeivirus γ RNAs encode cysteine-rich γ b proteins that range in size from 16 to 20 kDa. The hordeivirus γ b proteins share some structural similarities, but have little direct amino acid sequence similarity with the cysteine-rich proteins of tob- and carlaviruses. Studies of the BSMV ND18 strain demonstrate that γ b is not strictly required for infectivity in plants, but that it affects pathogenicity and virus accumulation. Extensive mutational analyses of γ b have revealed a number of distinct phenotypes that can ameliorate or exacerbate disease symptoms in barley, *Chenopodium* species, and *N. benthamiana*.

The majority of the cysteine residues in γ b are concentrated in the N-terminal half of the protein. These residues are clustered in two zinc-finger-like motifs, designated C1 and C2. A basic motif (BM) rich in lysine and arginine residues is located between the cysteine-rich regions. The BSMV γ b protein has been expressed and purified from *Escherichia coli* and shown to bind ssRNAs and zinc *in vitro*. The basic motif mediates RNA-binding activity and is involved in replication processes, and the C1, BM, and C2 regions each show independent zinc-binding activity. A coiled-coil domain near the C-terminus of the γ b protein mediates self-interactions that affect pathogenicity.

The hordeivirus γ b protein also has viral suppressor of RNA silencing (VSR) activities that interfere with host RNA silencing mechanisms, and recombinant BSMV mutants that lack γ b protein functions have reduced virulence. *Cis*- or *trans*- rescue with other VSR proteins, such as the *Tomato bushy stunt virus* (TBSV) P19 and Potyviral HC-Pro proteins, provide RNA silencing activities that partially restore BSMV virulence. The coiled-coil region of BSMV γ b is required for RNA silencing suppression, suggesting that homologous interactions of γ b are required for RNA silencing suppressor activity. In addition, individual site-specific mutations within the C1, BM, and C2 region of γ b also affect VSR activity of the protein, which suggests that RNA and metal binding are critical virulence components.

Fluorescence of the BSMV γ b-GFP fusion protein is dispersed throughout the cytoplasm, and subcellular fractionation of infected barley tissue and *N. benthamiana* leaf cells confirms that the protein is located in both the soluble and membrane fractions. In the presence of RNA α and RNA γ , a majority of the γ b proteins are recruited to the chloroplasts by interactions with the α a replication subunit protein. These interactions directly affect the BSMV replication cycle by binding progeny ssRNA, to facilitate enhanced dsRNA unwinding and BSMV replication. In addition, circumstantial evidence suggests that γ b may be involved in multiple steps of BSMV infection.

Cytopathology and Replication

Microscopy studies of BSMV-infected barley indicate that most of the visible changes in leaf appearance can be attributed to the disruption of organelles, most notably in chloroplasts of expanding barley and *N. benthamiana* leaves. Microscopy studies in *N. benthamiana* reveal cytoplasmic invaginations (CI) of the chloroplast outer membrane that contain ~50 nm diameter spherule structures within the invaginations that are characteristic of virus replication vesicles (VRCs) (Fig. 2). Several lines of evidence have confirmed that the BSMV α a replication protein has a major role in initiating chloroplast membrane remodeling to form the VCRs. At early stages of virion assembly, large numbers of virions accumulate in association with the CI's, suggesting that CI's may

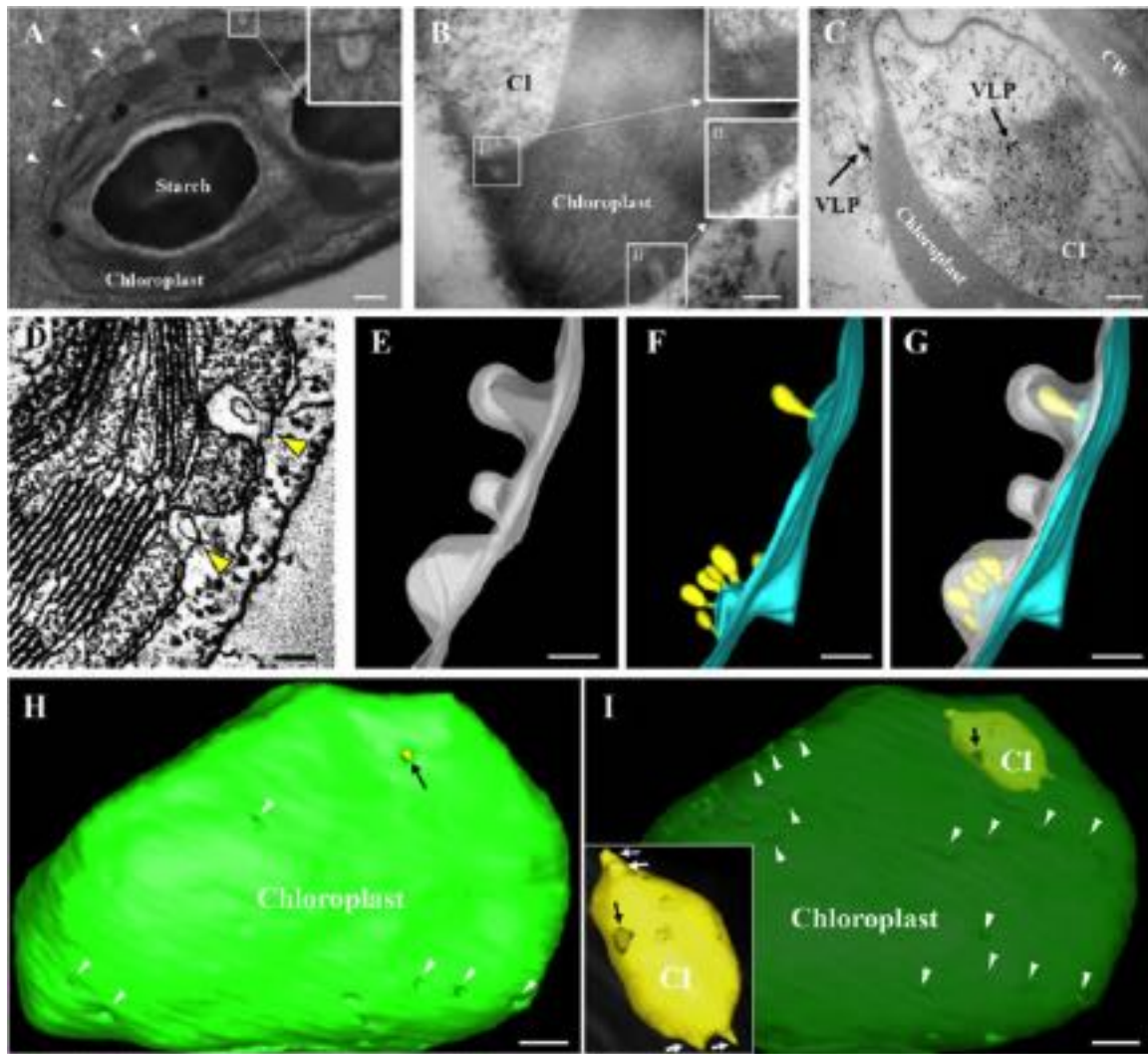


Fig. 2 Ultrastructural changes and 3D model of chloroplasts membrane rearrangement during BSMV infection of *N. benthamiana*. (A) Invaginations present around the chloroplast envelope (white arrowheads). The insert area illustrates a spherule magnification occasionally observed inside the invagination. Scale bar, 250 nm. (B) Immunogold labeling of dsRNA in BSMV-infected *N. benthamiana* cells. Insets I and II show magnifications of the corresponding boxed regions. Scale bar, 200 nm. (C) CP-specific antiserum showed specific labeling of virion-like particles (black arrows) in the cytoplasm and CIs. Scale bar, 200 nm. (D) Tomogram slices of altered chloroplast membranes from leaves of BSMV-infected *N. benthamiana*. The yellow arrowheads indicate the same spherules in different slices. Scale bar, 100 nm. (E–G) 3D model of remodeled chloroplast membranes induced by BSMV. Translucent white, inner chloroplast membrane; Light blue, outer chloroplast membrane; Yellow, spherules derived from the outer membrane. Scale bars, 100 nm. (H–I) A 3D surface rendering of the chloroplast tomogram. Green, chloroplast; Yellow, CI. To emphasize the CI (yellow) inside the chloroplast, the appearance of chloroplast shown in H was changed to translucent (I). White arrowheads indicate chloroplast envelope invaginations similar to those in Fig. 2G; black arrow point to the CI opening. The image inset in H is an enlarged view of the CI. Scale bars, 500 nm. This figure is adapted with permission from Jin, X., Jiang, Z., Zhang, K., *et al.*, 2018. Three-dimensional analysis of chloroplast structures associated with virus infection. *Plant Physiology* 176, 282–294. (© 2018 by the American Society of Plant Biologists).

provide sites for virion assembly, but at later stages of infection, so many virions are present in the cytoplasm and associated vacuoles that it is difficult to determine specific sites of accumulation.

A model for hordeivirus replication posits that virions are disassembled by host factors in the plant cell, followed by translation of RdRp subunits from RNA α and RNA γ (Fig. 3). The available evidence indicates that the α subunit replication protein localizes to chloroplasts via an N-terminal chloroplast localization signal and induces chloroplasts outer membrane invaginations. The α protein is then posited to recruit gRNAs, the γ subunit, γ b, and relevant host factors to the chloroplasts to facilitate replication. Current evidence suggests that the γ b protein is recruited to the VCRs by interactions with the methyltransferase (MT) hinge-like domain of the α protein. During replication, the γ b protein enhances the helicase activity of the α protein via its N-terminal ssRNA binding activity to facilitate efficient RNA unwinding and replication. Subsequently, progeny vRNAs in the VRCs interact with TGB1 proteins to assemble RNP movement complexes. These complexes participate in movement by interacting with TGB2

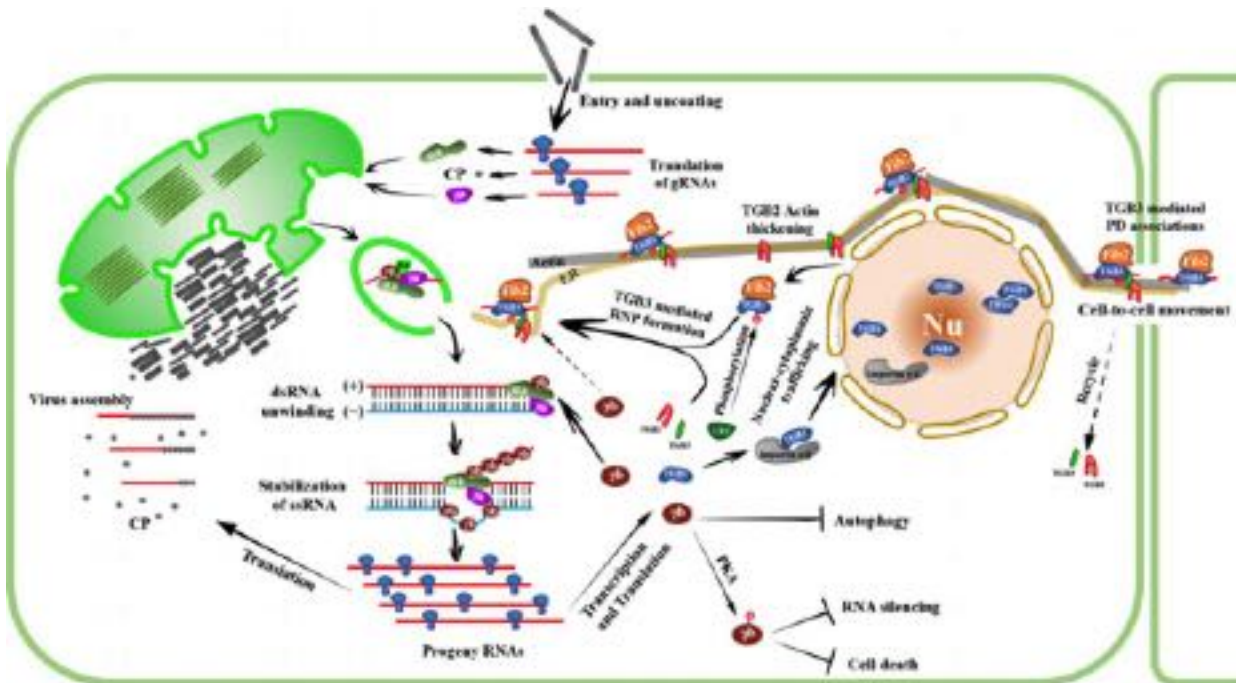


Fig. 3 Model of the hordeivirus infection cycle. Hordeivirus transmission from plant to plant occurs via mechanical or pollen transmission, and infections may be subsequently maintained by seed transmission through the embryo to germinating seedlings. Entry during mechanical abrasion is followed quickly by uncoating of gRNAs in the cytoplasm and translation of gRNAs α , β , and γ . Then, the translated α and γ proteins interact with gRNAs to form RNA-Dependent-RNA polymerase complexes that localize at chloroplast outer membranes where membrane invaginations are induced to form nascent VRCs. The RdRp then recruits relevant host factors (HF) to the VRCs to initiate robust gRNA replication. Production of negative- and positive-sense gRNAs results in formation of dsRNA replicative intermediate species. The RdRp also initiates transcription of sgRNAs from internal promoters on negative-strand RNA β and RNA γ to generate sgRNAs that function in translation of proteins encoded by the sgRNAs. The γ b protein, translated by sgRNA γ , functions as a helicase enhancer to facilitate RNA replication and subvert host autophagy processes. During the course of these events, phosphorylated γ b proteins suppress RNA silencing pathways and host cell death responses. The TGB1 protein has a nuclear phase during movement and relies on the importin α/β protein for some aspects of nuclear-cytoplasmic export and trafficking. The exported TGB1 protein interacts with the fibrillarin protein and binds to positive-sense gRNAs to create ribonucleoprotein (RNP) complexes. The TGB2 protein functions in ER/actin membrane thickening at the infection front, and binds to TGB3 proteins that interact with CK2 kinase phosphorylated TGB1-RNP complexes during transport to the plasmodesmata (PD). The TGB3 C-terminal amino acids have a central role in PD recognition and transit of RNP gRNAs complexes to adjacent cells, and the γ b protein may also be involved in some aspects of movement. The TGB2 and 3 proteins also may be recycled by the endocytic pathway to facilitate the movement of additional RNP complexes. As the later stages of infection progress, increasing amounts of CP are translated from gRNA β and encapsidate positive-sense gRNAs at chloroplast encapsidation sites to form progeny virions that accumulate in the cytoplasm.

and TGB3 to elicit actin remodeling that produces dramatically thickened filaments that associate with the endoplasmic reticulum. TGB2 and 3 actin interactions then facilitate transit of RNP complexes to peripheral vesicles and through the PD to adjacent cells. Molecular events involved in virion assembly are currently obscure, but it is likely that the CP has a major role in interacting with nascent gRNAs at CIs during the late stages of replication for virion assembly.

Pathogenesis

Natural mutants in both the non-coding and coding regions of BSMV RNAs have been shown to have strain-specific effects on pathogenesis. For example, the ND18 strain is unable to infect oats, whereas the CV42 strain is pathogenic in this host. ND18 virulence in oats can be engineered by altering a single amino acid in the α a protein, so this phenotype may result from specific associations of α a with a host protein. Some site-specific amino acids of BSMV are potential targets of host resistance genes. Most of the known resistance genes in barley and oats are recessive, but analysis of the inheritance of resistance in *B. distachyon* indicates that a single dominant gene, designated *Barley stripe mosaic virus resistance 1* (*Bsr1*), recognizes an arginine residue at position 390 (R390) and a threonine residue at position 392 (T392) of BSMV_{ND18} TGB1 and elicits the BD3–1 *Bsr1* hypersensitive response, whereas the BSMV_{NW} containing lysines at both positions does not.

Natural variation amongst BSMV strains has also facilitated identification of factors affecting seed transmission. Primary determinants influencing the efficiency of seed transmission reside in RNA γ sequences in the 5' UTR and in the γ b gene. In a fourth case of strain-specific virulence, ND18 is able to systemically infect *N. benthamiana*, whereas the Type strain is able to replicate and move in inoculated leaves, but is unable to invade systemically. Infections with the ND18 and the Type strains also result in

different local lesion phenotypes on *C. amaranticolor*. These aberrant phenotypes appear to be caused by the presence of a short ORF within an amino-terminal 372-nucleotide duplication in the 5' UTR of the Type strain that reduces γ a translation. However, the γ a replicase/host interactions contributing to these distinct movement properties have not been defined molecularly, and differences in the amino acid composition of the γ b proteins may also affect the disease phenotypes.

Advances in understanding pathogenesis have also been obtained by reverse genetic mutations. The results show that γ b is not strictly required for infectivity in plants, but that it has important pathogenicity effects. Individual site-specific mutations within the cysteine-rich region of γ b can also result in a variety of phenotypes ranging from striking white regions to necrotic streaks on systemically infected leaves of barley and *N. benthamiana*. Moreover, plants infected with site-specific γ b mutants have secondary effects that culminate in greatly reduced abundance of viral RNA accumulation. Although the CP is not required for infectivity in barley, *N. benthamiana*, or *Chenopodium* species, infections elicited by CP-deficient BSMV mutants appear to be more aggressive, because recovery is protracted and the symptoms are more severe than in the presence of the CP. These results suggest that in addition to being essential for formation of virions, the coat protein may have a critical role in the kinetics of replication late in infection. Overall, these analyses suggest that complex interactions may affect virulence and disease phenotype, and that the levels of replication may have important effects on movement, host range, and seed transmission.

Molecular genetic studies also suggest substantial specificity amongst genes encoded by BSMV, LRSV and PSLV RNAs, because heterologous strand reassortments of BSMV, LRSV, and PSLV gRNAs, fail to establish systemic infections in plants. However, substitution of the PSLV TGB for the BSMV TGB in RNA β results in systemic infections of *N. benthamiana*, *C. amaranticolor*, and wheat (*Triticum durum*), whereas similar LRSV TGB substitutions failed to establish systemic infections of any of these plants. Interestingly, BSMV infects all 4 hosts tested, whereas PSLV infects only wheat and LRSV infects *N. benthamiana* and *C. amaranticolor*, but not wheat. These results suggest that factors other than the TGB sequences *per se* contribute to host range, but this appears not to be a consequence of γ b because our recent studies show that heterogeneous substitutions of PSLV and LRSV γ b for BSMV γ b do not interfere with the ability of BSMV to infect *N. benthamiana*.

Hordeivirus-Host Interactions

γ b Interference With Host Defenses

RNA silencing pathways

Several lines of evidence indicate that a major function of γ b is to serve as an inhibitor of host defense mechanisms that target RNA replication. A VSR deficient γ b mutant compromises BSMV replication, but the γ b function can be partially complemented by expression of tombusvirus p19 and potyvirus HC-Pro VSRS *in cis* or *in trans*. Similarly, expression of the γ b protein is able to suppress silencing in heterologous virus-induced silencing systems. An agrobacterium-mediated transient silencing assay also provides strong support for a role of γ b in suppression of RNA silencing. In this assay, γ b interferes with silencing of GFP fluorescence induced by expression of dsRNAs. These results thus provide a persuasive argument that a major function of γ b is to counter host innate RNA interference defenses, as well as to assist in helicase unwinding during replication.

Autophagy responses

Autophagy is a conserved defense pathway against abiotic and biotic stresses including virus infection. However, plant virus counter-defense mechanisms that interfere with autophagy to promote viral infection have not been explored extensively. We recently found that BSMV infection suppresses autophagy and that expression of the γ b protein is sufficient to inhibit autophagy. Molecular analyses indicate that γ b interferes with interactions of the autophagy-related protein 7 (ATG7) with ATG8 in a competitive manner and the conversion of ATG8 to ATG8-PE. These findings reveal that the γ b protein subverts autophagy-mediated antiviral defenses by disrupting ATG7-ATG8 interactions to promote BSMV infection, and that ATG7 is a newly discovered target of pathogen effectors in the plant defense and virus counter-defense arms race.

ROS bursts

Reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2) or superoxide anions (O_2^-) are amongst the earliest cellular responses against pathogen infection. BSMV infection reduces both H_2O_2 and O_2^- elevation by ~ 10 fold in inoculated leaves compared with the empty vector. The γ b protein has a major role in reduction of ROS bursts by interacting directly with glycolate oxidase (GOX), which is localized in peroxisomes and is a key constitutive enzyme in the photorespiration pathway. These results demonstrate that γ b interacts with GOX to suppress peroxisomal ROS production and facilitate virus infection, and contributes to an understanding of roles played by ROS and plant ROS-producing enzymes during viral infection.

Post-Translational Modifications of BSMV Proteins

Post-translational modifications are ubiquitous in eukaryotic cells, and are defined by addition of a covalent group or an enzymatic modification of proteins after translation. Viruses have evolved many strategies to hijack host post-translational modification pathways to maximize their pathogenicity. During BSMV infection, protein kinase CK2 from barley and *N. benthamiana* phosphates the TGB1 protein mainly at Thr-401, and Thr-395 kinase docking sites. The TGB1_{T395A/T401A} mutation compromises BSMV cell-to-cell

movement, and the virus fails to develop systemic infections in barley and *N. benthamiana*. The phosphorylated TGB1 protein interacts more strongly than unphosphorylated TGB1 with the TGB3 protein, and this interaction promotes BSMV cell-to-cell movement. Post-translational modifications also affect γ b functions during BSMV infection. A recent study also indicates that γ b Ser-96 is phosphorylated *in vitro* and *in vivo* by a protein kinase A-like kinase. Non-phosphorylatable γ b substitution mutants (BSMV γ b_{S96A} and γ b_{S96R}) reduced BSMV accumulation and induced necrosis in barley, wheat, and *N. benthamiana*. Phosphorylation of the γ b protein at Ser-96 also enhances VSR activity by promoting 21-bp dsRNA binding activity. These results *in toto* suggest that phosphorylation may affect several aspects of pathogenesis and infection, including movement, cell death, and RNA silencing responses.

Hordeivirus Applications to Cereal Genomic Analyses

Virus-induced gene silencing (VIGS) has provided useful approaches to improve our understanding of the functional genomics of monocots and dicots, and several virus vectors have been developed for this purpose. BSMV has been developed into an effective VIGS vector by modifying sequences downstream of the γ b gene to enable expression of untranslatable foreign inserts. This strategy has been successfully applied to *N. benthamiana*, wheat, barley, and *B. distachyon* by targeting phytoene desaturase (*PDS*), magnesium chelatase subunit H (*ChlH*), and plastid transketolase (*TK*) genes for VIGS. Other BSMV VIGS applications have been developed to downregulate host genes known to be involved in disease resistance, and these have enabled identification of genes required for powdery mildew resistance in barley, and rust resistance in wheat. In addition, high levels of foreign proteins and chimeric proteins such as GFP proteins fused to the C-terminus of the γ b protein have been expressed from BSMV vectors. Heterologous γ b proteins less than 110 amino acids have been expressed during systemic infections, and the fusion proteins can be efficiently and constitutively expressed during infections of *N. benthamiana* and some monocots. These studies indicate that BSMV VIGS provides a powerful and robust system for gene silencing and gene overexpression from BSMV may have numerous applications for genetic analyses of a large number of cereal crops.

Future Perspectives

The broad host range of hordeiviruses and the information accumulated about virus genes involved in replication, movement, and responses to host resistance provides a rich resource for obtaining insight into the mechanics of infection. Considerable efforts have recently focused on BSMV replication and interactions between hordeivirus-encoded proteins and host factors. However, very little is known about the roles of host interactions that facilitate hordeiviruses replication cycles, switch from replication to RNP movement complexes that move across PD and into vascular tissues, or how host resistance pathways interfere with hordeivirus infection. An understanding of events leading to the establishment of replication factories and initiation and regulation of replication cycles requires additional information about the interplay between chloroplasts and cytoplasmic elements during the early stages of infection. An understanding of BSMV-chloroplast interactions may also reveal mechanisms whereby recessive resistance genes employed in agriculture affect BSMV replication. Additional biochemical and cell biological analyses of the TGB-encoded proteins and their interaction proteins could also hone our understanding of the cell-to-cell and vascular movement events that permit systemic hordeivirus infections, and how hordeiviruses invade vegetative tissue, seed, and pollen. Such investigations may also provide insights into global aspects of the subcellular transport network, and functions of organelles and PD functions in plant development.

New genome-editing techniques like CRISPR/Cas9 have developed rapidly over the past decade, and the use of hordeiviruses for novel applications of these techniques will accelerate future genomics methods. Hence, we envision that more sophisticated analyses of host genes, cellular interactions, and the ability to modulate their expression will accelerate the pace of future hordeivirus research findings.

Further Reading

- Adams, M.J., Adkins, S., Bragard, C., *et al.*, 2017. ICTV virus taxonomy profile: *Virgaviridae*. *Journal of General Virology* 98, 1999–2000.
- Bragg, J.N., Lim, H.S., Jackson, A.O., 2008. Hordeivirus. In: Mahy, B.W.J., Regenmortel, M.H.V.V. (Eds.), *Encyclopedia of Virology*. Oxford: Academic Press, pp. 459–467.
- Clare, D.K., Pechnikova, E.V., Skurat, E.V., *et al.*, 2015. Novel inter-subunit contacts in *Barley stripe mosaic virus* revealed by cryo-electron microscopy. *Structure* 23, 1815–1826.
- Cui, Y., Lee, M.Y., Huo, N., *et al.*, 2012. Fine mapping of the *Bsr1* *Barley stripe mosaic virus* resistance gene in the model grass *Brachypodium distachyon*. *PLoS One* 7, e38333.
- Holzberg, S., Brosio, P., Gross, C., Pogue, G.P., 2002. *Barley stripe mosaic virus*-induced gene silencing in a monocot plant. *The Plant Journal* 30, 315–327.
- Jackson, A.O., Lim, H.-S., Bragg, J., Ganesan, U., Lee, M.Y., 2009. Hordeivirus replication, movement, and pathogenesis. *Annual Review of Phytopathology* 47, 385–422.
- Jiang, Z., Li, Z., Yue, N., *et al.*, 2018. Construction of infectious clones of lychnis ringspot virus and evaluation of its relationship with barley stripe mosaic virus by reassortment of genomic RNA segments. *Virus Research* 243, 106–109.
- Jin, X., Jiang, Z., Zhang, K., *et al.*, 2018. Three-dimensional analysis of chloroplast structures associated with virus infection. *Plant Physiology* 176, 282–294.
- Lee, W.-S., Hammond-Kosack, K.E., Kanyuka, K., 2012b. *Barley stripe mosaic virus*-mediated tools for investigating gene function in cereal plants and their pathogens: Virus-induced gene silencing, host-mediated gene silencing, and virus-mediated overexpression of heterologous protein. *Plant Physiology* 160, 582–590.
- Lee, M.Y., Yan, L., Gorter, F.A., *et al.*, 2012a. *Brachypodium distachyon* line Bd3-1 resistance is elicited by the *Barley stripe mosaic virus* triple gene block 1 movement protein. *Journal of General Virology* 93, 2729–2739.
- Lim, H.-S., Lee, M.Y., Moon, J.S., *et al.*, 2013. Actin cytoskeleton and golgi involvement in *Barley stripe mosaic virus* movement and cell wall localization of triple gene block proteins. *The Plant Pathology Journal* 29, 17–30.

- Yang, M., Zhang, Y., Xie, X., *et al.*, 2018a. *Barley stripe mosaic virus* γ b protein subverts autophagy to promote viral infection by disrupting the ATG7-ATG8 interaction. *Plant Cell* 30, 1582–1595.
- Yuan, C., Li, C., Yan, L., *et al.*, 2011. A high throughput *Barley stripe mosaic virus* vector for virus induced gene silencing in monocots and dicots. *PLOS One* 6, e26468.
- Zhang, X., Dong, K., Xu, K., *et al.*, 2018. *Barley stripe mosaic virus* infection requires PKA-mediated phosphorylation of γ b for suppression of both RNA silencing and the host cell death response. *New Phytologist* 218, 1570–1585.
- Zhang, K., Zhang, Y., Yang, M., *et al.*, 2017. The *Barley stripe mosaic virus* γ b protein promotes chloroplast-targeted replication by enhancing unwinding of RNA duplexes. *PLOS Pathogens* 13, e1006319.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/virgaviridae/669/genus-hordeivirus
Genus: Hordeivirus.

Idaeoviruses (*Mayoviridae*)

Robert R Martin and Karen E Keller, Horticultural Crops Research Unit, Agricultural Research Service, US Department of Agriculture, Corvallis, OR, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A.T. Jones, H. Barker, Idaeovirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00539-2](https://doi.org/10.1016/B978-012374410-4.00539-2).

Nomenclature

aa Amino acid(s)

CP Coat protein

ELISA Enzyme-Linked Immuno Sorbent Assay

GFP Green fluorescent protein

HEL Helicase

kb Kilobase

kDa Kilo dalton

MP Movement protein

MTR Methyl-transferase

nt Nucleotide(s)

ORF Open reading frame

PCR Polymerase Chain Reaction

PVP Polyvinyl pyrrolidone

RdRP RNA-dependent RNA polymerase

RISC RNA-induced silencing complex

satRNA Satellite RNA

UTR Un-translated region

Glossary

Certification Program Is a comprehensive process established and authorized by a state or other governmental entity for the production of plants free of regulated pests and diseases. The regulations for each program define the program participation, plant production, plant identification and labeling, and quality assurance requirements.

ELISA An acronym for Enzyme-Linked Immuno Sorbent Assay, which is a serological test in which antibodies are used to detect plant viruses.

Foundation Block A foundation block is a group of plants that have been tested for viruses or other diseases and are maintained in isolation under conditions that prevent (re) infection. These blocks are considered Generation 1 (G1) stock within the meaning of the 'Generation level' concept recommended by the NAPPO Guidelines.

Heat therapy Is a method used to eliminate viruses in which plants are grown at 37°C–40°C for 4–6 weeks. This is used in combination with meristem tip culture for efficient elimination of most plant viruses.

Horizontal Transmission Transmission of a pathogen to individuals within the same generation. For pollen-borne plant viruses, this is virus movement from the pollen grain

to maternal tissue during the pollination process resulting in the infection of the pollinated plant.

Meristem tip-culture The process of removing pieces of plant 0.5 mm or less in length, consisting of a meristem and a few leaf primordia that is grown in vitro with the goal of generating a plant free of plant viruses.

Pollen-borne Carried on or in the pollen, plant viruses can be carried on the exterior of the pollen or internally in the pollen grain.

Polymerase Chain Reaction Is a laboratory detection technique for plant pathogens that amplifies a segment of DNA or RNA from the target organism (for example, a virus) many times by using short segments of DNA (primer) that are complimentary to the target nucleic acid. For RNA targets, reverse transcription is required to convert the RNA to DNA that is then used in PCR.

Tissue Culture The cultivation of plants (cells, tissues, or organs) under aseptic conditions in a synthetic medium in vitro.

Vertical Transmission Transmission of a pathogen from one generation to the next, for plant viruses this is the result of seed transmission.

Introduction

Until recently, Raspberry bushy dwarf virus (RBDV) was the sole recognized member of the genus *Idaeovirus*. More recently, Privet leaf blotch associated virus (PrLBaV) has been listed as a member of the genus. A novel virus from blackcurrant, Blackcurrant leaf chlorosis associated virus (BCLCaV), or Blackcurrant idaeovirus (BCIV) has been described by two groups (the acronym BCLCaV will be used here), which is phylogenetically most similar to PrLBaV and RBDV. A third virus, Japanese holly fern mottle virus has similarities in the RNA1 to that of RBDV, but its RNA2 is quite distinct from the other three viruses. A partially sequenced virus from citrus showed significant similarities to the coat protein of RBDV, but complete sequence information is lacking. The initial isolation of RBDV was from plants that exhibited stunting and proliferation, thus the name Raspberry bushy dwarf virus. RBDV was initially isolated from red raspberry by mechanical transmission from symptomatic plants to herbaceous hosts. It was shown later that these plants were coinfecting with Black raspberry necrosis virus and that RBDV in single infections did not cause the 'bushy dwarf' symptoms. However, in single infections, RBDV does induce

crumbly fruit in many cultivars of raspberry and in some cases leaf chlorosis. The virus genus name is from the species of the host (*Rubus idaeus*).

RBDV is serologically indistinguishable from Loganberry degeneration virus and synonymous with Raspberry yellows virus. It is also probably the same as Raspberry line pattern virus reported from Poland. Antisera has not been prepared to the other idaeoviruses. Studies on the molecular biology of RBDV have shown it to possess a novel combination of properties and was assigned as the sole known member in the genus *Idaeovirus*. Though RBDV and the other idaeoviruses have a bipartite genome, they are most similar to the ilarviruses that have tripartite genomes in the family *Bromoviridae*.

Geographical Distribution

RBDV probably occurs wherever red raspberry (*R. idaeus*, *R. strigosus*), black raspberry (*R. occidentalis*), or blackberry (*R. fruticosus*), are grown. The virus has been reported from Australasia, China, Eastern and Western Europe, North and South America, South Africa, and the former Soviet Union. Three strains of RBDV have been characterized based on serological or biological properties. The common or Scottish (S) isolates have been identified in most countries where raspberries are grown. The black raspberry (B) isolate from a North American black raspberry, *R. occidentalis* are serologically distinct from the S isolates. However, more data are needed to determine whether B isolates are restricted to *R. occidentalis*, or representative of isolates from North America, and whether S isolates can occur in *R. occidentalis* in nature. B isolates have only been found in North America in native *Rubus*. A third strain (RB) for resistance breaking, has limited distribution in Europe and was identified in North America for the first time in 2016. These isolates are indistinguishable from S isolates, except that they are able to infect cultivars that are resistant to the S isolates. A fourth strain has been partially characterized from *R. multibracteatus* collected in China, which has 23 aa substitutions in the coat protein (CP) when compared to the B isolate and is the most divergent isolate of RBDV reported to date. It is also distinguishable from D, S, and RB isolates using monoclonal antibodies. A fifth strain has been reported from grapevines in Slovenia and Hungary, which forms a separate cluster from the isolates from *Rubus* species, and can also be distinguished using monoclonal antibodies.

BCLCaV and BCIV were identified using high throughput sequencing from dsRNA purified from symptomatic plants. The distribution and population structure of the virus have not been studied. The virus isolates were from two different blackcurrant cultivars, 'Belaruskaya Slodkaya' held in the *Ribes* germplasm collection at the USDA-ARS National Clonal Germplasm Repository in Corvallis, Oregon, and 'Baldwin' from British Columbia, Canada. In both cases the plants exhibited a general chlorosis symptom. PrLBaV was identified using high throughput sequencing of small RNAs purified from privet (*Ligustrum* spp.) plants exhibiting yellow mosaic symptoms. Infectious chlorosis of privet, variegation of privet, mosaic of ligustrum are considered synonyms of yellow mosaic of privet. In parallel experiments with symptomless privet, the virus was not detected. Additionally, none of the other viruses reported to infect privet were detected in these experiments. There was no mention if there were any other viruses detected in the study that had not been previously reported in privet. The disease has been reported to be widespread in Europe based on symptoms, though it is not known if the multiple reports refer to the same disease or if there are multiple causal agents involved.

Natural and Experimental Transmission

In red raspberry and black raspberry, RBDV is readily transmitted to progeny seedlings via either the pollen or the ovule of infected plants and transmission is greatest when both parents are infected. If raspberry plants were deflowered they did not become infected when adjacent flowering plants of the same cultivar did become infected. Therefore, the flower is believed the route for natural transmission of the virus. Healthy plants can be infected after the first flowering season when planted close to virus sources. The rate of virus spread horizontally in the field varies greatly by cultivar. Horizontal transmission occurs as pollen tubes penetrate the stigma and virus is able to move into the stigmatic tissue, and from there move systemically to infect the plant. This transmission has been documented using pollen from RBDV infected *Rubus* to infect healthy *Rubus* plants. It was also shown that transmission to the maternal tissue requires pollen tube growth into the stigma, but does not require fertilization. Pollen from RBDV infected raspberry placed on stigmas of *Torenia fournieri* resulted in horizontal transmission of RBDV to the maternal tissue if the pollen tube penetrated the stigma. If the pollen was inactivated and not able to grow into the stigma transmission did not occur. Thus, horizontal transmission via infected pollen can occur between sexually incompatible species. RBDV is transmitted easily from *Rubus* by grafting to susceptible *Rubus* species and cultivars. However, graft-inoculated plants may need to go through a dormant season before the virus becomes fully systemic. RBDV is also transmissible from infected *Rubus* to herbaceous test plants by mechanical inoculation of sap extracts in aqueous solutions of either 2% nicotine or in 0.05 M phosphate buffer containing 1% polyvinyl pyrrolidone (PVP) (mol. wt. 6000–40,000 Da). Mechanical transmission from RBDV-infected herbaceous plants to susceptible red raspberry cultivars is possible, but with great difficulty. BCLCaV is graft transmissible to other blackcurrants and PrLBaV is graft transmissible to symptomless privet plants. In both cases, limited work on graft transmission has been reported, so symptoms in different cultivars or species is unknown. BCLCaV was transmitted mechanically to *Nicotiana benthamiana* and *N. occidentalis*. There are no reports of transmission of PrLBaV to herbaceous hosts.



Fig. 1 Top. Leaves of RBDV infected 'Caroline' red raspberry (A) and 'Marion' blackberry infected with RBDV, Blackberry calico virus, and Black raspberry necrosis virus (B). 'Marion' infected with the latter two viruses shows a fine line pattern and do not have fruit symptoms. Center: Symptoms of crumbly fruit of 'Meeker' red raspberry: (C). fruit from healthy plant (D). fruits from RBDV-infected plant. Bottom: Fruit of 'Marion' blackberry infected with Blackberry calico virus and Black raspberry necrosis virus; (E). plants without RBDV, (F). plants with RBDV infection. Photographs courtesy of R.R. Martin.

Disease Symptoms and Effects

Naturally Infected Plants

In nature, RBDV occurs in wild and cultivated *Rubus* species and cultivars. In single infections RBDV is often symptomless in raspberry and blackberry plants depending on cultivar. However, in sensitive raspberry cultivars and under certain environmental conditions, it can induce chlorosis. This disease is characterized by an initial yellowing of the main or minor veins of the lower leaves that can progress to affect the entire leaf (Figs. 1(A) and (B)). Some raspberry cultivars may also develop chlorotic/yellow rings or line patterns in leaves. RBDV infection also causes 'crumbly fruit' in some raspberry cultivars (Fig. 1). This condition arises from the abortion of some drupelets and the uneven development of others and frequently results in an abnormally shaped fruit, which, on picking, may disintegrate into individual drupelets or drupelet clusters. The same disease syndrome can be induced by other unrelated causes, such as poor pollination conditions, infection with other viruses, or genetic aberrations. Chlorosis and crumbly fruit is also observed in blackberry, but usually when RBDV is present in combination with other viruses (Fig. 1(C)–(F)). Additional symptoms have been attributed to RB isolates. In field studies an association was reported between infection with RBDV-RB and premature defoliation of fruiting canes, decreased vigor, leaf curling, necrosis, and death of laterals, and increased winter kill in some red raspberry cultivars that are immune to S isolates.

In commercial fields, RBDV rarely occurs in single infections. The virus name comes from symptoms in a plant that was co-infected with RBDV and Black raspberry necrosis virus. In single infections RBDV does not cause bushiness or dwarfing symptoms. In mixed infections with Raspberry leaf mottle virus the titer of RBDV in 'Meeker' red raspberry was increased approximately 400-fold over the titer of RBDV in single infections. In mixed infections of RBDV with Raspberry leaf mottle virus or Raspberry latent virus, raspberry plants growth was significantly reduced during the establishment year compared to plants infected singly with RBDV. The mixed infections also resulted in smaller fruit, more severe crumbliness of the fruit and lower yields, than in plants singly infected with RBDV in the first two fruiting years.

RBDV was detected in cultivars of white-fruited and red-fruited grapevines (*Vitis vinifera*) in Slovenia and Hungary. RBDV infected 'Laški Rizling' exhibited line pattern and mild leaf yellowing. In most cultivars RBDV infections were symptomless. RBDV is reported to be widespread in grapevines in Slovenia. Of 390 seedlings grown from seed collected from RBDV infected cv. Laška Rizling, all tested negative for RBDV, suggesting a very low levels or no seed transmission in grapevines. It would be curious to examine horizontal transmission in grapevines using pollen from infected plants. RBDV isolates from grapevine spread systemically in *Chenopodium murale*, which has not been observed with isolates from *Rubus* spp.

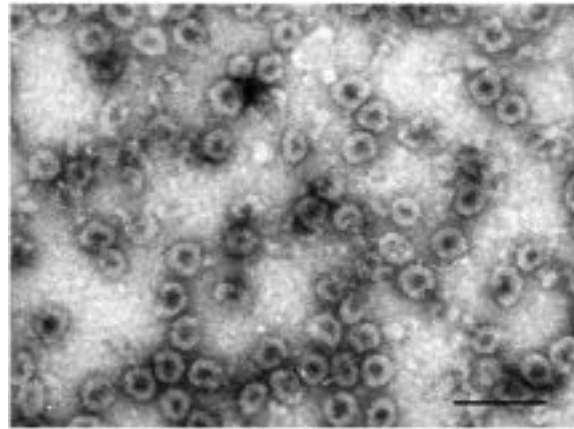


Fig. 2 Electron micrograph of a purified preparation of RBDV Particles stained with uranyl acetate. Scale=100 nm. Photographs courtesy of R.R. Martin.

PrLBaV was sequenced from plants exhibiting cream colored blotches, some ringspots, leaves with coalesced blotches leading to almost completely cream colored leaves and in some cases necrotic spots within the cream colored blotches. There was not any testing for the PrLBaV in a large number of symptomatic and asymptomatic plants to document an association between the symptoms and the presence of the virus. BCIV was sequenced from blackcurrant 'Belaruskaya Slodkaya' and BCLCaV was sequenced from blackcurrant 'Baldwin' both exhibiting a general chlorosis of the leaves.

Experimentally Infected Plants

RBDV has been transmitted by graft inoculation from *Rubus* to perennial plants in other genera. It infected *Fragaria vesca* and *Prunus mahaleb* seedlings without any symptoms, and induced pronounced chlorotic vein-banding and line patterns in *Cydonia oblonga* cv. C7-1. RBDV has a moderately wide host range in herbaceous test plants infecting over 50 plant species in 12 families of dicots. The idaeovirus from blackcurrants has been transmitted mechanically to *Nicotiana occidentalis* and *N. benthamiana*. There is no information on mechanical transmission of PrLBaV to herbaceous hosts.

Virion Structure

Chenopodium quinoa is the preferred propagation host for RBDV. Purified virus particles are about 33 nm in diameter, and appear quasi-isometric because they tend to collapse and deform on the electron microscope grid (Fig. 2). They appear more spherical in shape when fixed with formaldehyde. Particle preparations have A260/A280 of 1.62, suggesting that RBDV particles contain about 24% RNA. Purified preparations of RBDV particles typically yield a single major protein estimated by polyacrylamide gel electrophoresis to have a molecular weight of ca. 30 kDa. The particles contain three species of single-stranded RNA (ssRNA) of approximately 5.5 kb (RNA1), 2.2 kb (RNA2), and 1.0 kb (RNA3). RNA3 is a sub-genomic of RNA2 (Fig. 3). The particles are unstable and readily disrupt in the presence of 0.01% sodium dodecyl sulfate, indicating that they are stabilized by protein-RNA interactions.

Genome

The complete nucleotide sequences have been determined for the two genomic RNAs (RNA1 and RNA2) of RBDV isolates S and RB, BCLCaV, and PrLBaV. The genome organization of idaeoviruses is shown in Fig. 3, at this time the 12 kDa putative ORF has only been predicted in RBDV. RNA1 of RBDV (R15 strain), PrLBaV and BCLCaV are 5,449, 5377, and 5349 nt long, respectively. RNA1 codes for a multi-functional protein of ~19 kDa, which contains the methyl-transferase (MTR), helicase (HEL), and RNA-dependent RNA polymerase (RdRp) domains. RNA2 of RBDV (R15 strain), PrLBaV, and BCLCaV are 2,231, 2348, and 2280 nt long, respectively. RNA 2 codes for two proteins, a putative movement protein (MP) and a coat protein (CP) with a molecular masses of about ~39 kDa and ~30 kDa, respectively. The CP of RBDV is expressed via a sub-genomic RNA (RNA3, 946 nt), and this is likely the case for PrLBaV and BCLCaV. This predicted size corresponds to that of protein obtained from purified virus particles and that of the main polypeptide made by in vitro translation of RNA3. Fig. 4 shows phylogenetic relationship among the idaeoviruses and two closely related viruses.

RNA1 also contains a small ORF in a different translation frame that overlaps with the 190 kDa protein, which if expressed could code for a 12 kDa protein. The position of this ORF corresponds to that of the 2b gene in the tripartite genome of Cucumber

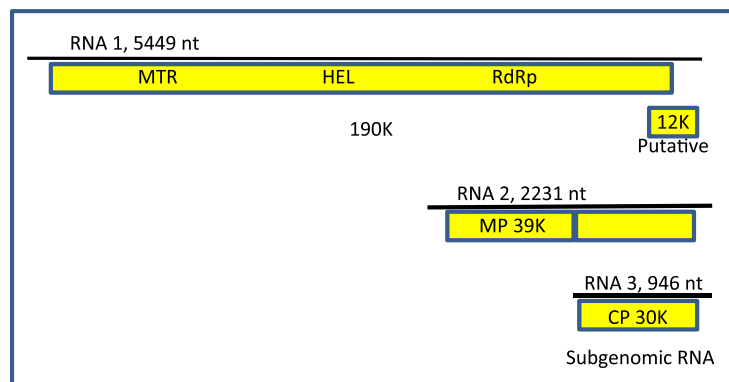


Fig. 3 Schematic showing the genome organization of RBDV. RNA molecules are represented by solid lines, the yellow blocks show the open reading frames (ORFs). The relative positions of the functional domains of the polyprotein encoded by RNA1 are MET (methyl-transferase), HEL (helicase) and RdRp (RNA-dependent RNA polymerase), RNA2 codes for the MP (movement protein), and RNA 3, a sub-genomic of RNA 2 codes for the CP (coat protein). PrLBaV and BCLCaV have similar organizations but lack the 12K putative open reading frame.

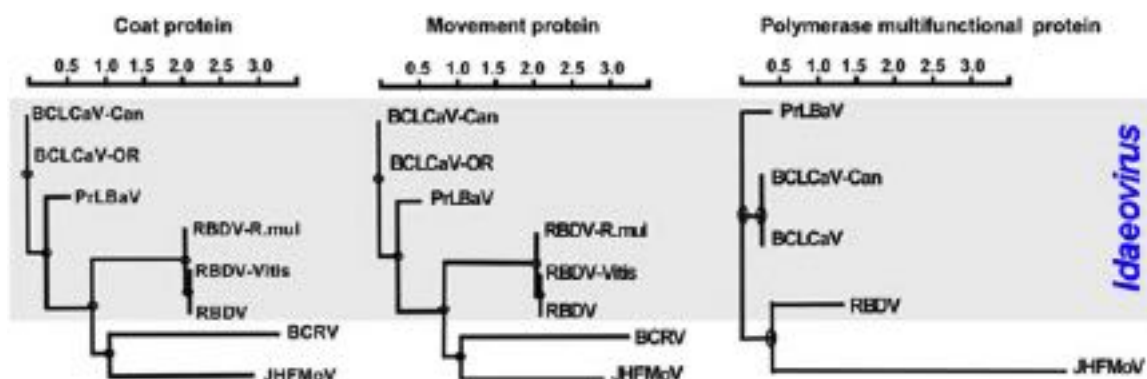


Fig. 4 Phylogenetic relationship of the idaeoviruses Raspberry bushy dwarf virus (RBDV), Privet leaf blotch associated virus (PrLBaV) and Blackcurrant leaf chlorosis associated virus (BCLCaV), the Japanese holly fern mottle virus (JHFMov) is also included in the polymerase protein since its RNA1 resembles those of idaeoviruses, and Blackberry chlorotic ringspot virus (BCRV), since idaeoviruses have nearest affinities to ilarviruses. Phylogenetic analysis was performed using the coat proteins (a), putative movement proteins (b) and polymerase proteins (c), the BCRV was not included in the polymerase phylogeny due to the multiple functional domains on two RNAs. The trees were generated by the maximum-likelihood PhyML 3.0, using the default parameters. Branch lengths are proportional to genetic distances between sequences and the scale represents substitutions per amino acid site. RBDV, red raspberry isolate (NC_003739, NC-003740); RBDV-Vitis, grapevine isolate (EU796086); RBDV-R.multi, *Rubus multibracteatus* isolate (DQ120126); BCLCaV-Can, Canadian isolate (KX838923, KX838924); BCLCaV-OR, NCGR isolate (KY399998, KY399999); PrLBaV (NC_031341, NC_031342); JHFMov, (FJ907327); and BCRV (NC_011554).

mosaic virus (CMV) that encodes a 10 kDa polypeptide and overlaps the C-terminal 69 codons of ORF 2a that encodes the RNA polymerase protein. The CMV 2b protein is important in long distance systemic movement and the suppression of post-transcriptional gene silencing, but it is not necessary for infectivity. However, experiments failed to demonstrate the activity of the 12 kDa RBDV protein in suppressing post-translational gene silencing. There is a report of RNA1 of RBDV replicating and moving systemic in the absence of RNA2, which suggests the expression of ORF 1a if it serves a long distance movement function similar to that of ORF 2a in CMV.

There are similarities among the 5'- or 3' terminal non-coding sequences of each of the genomic RNA species of RBDV, PrLBaV and BCLCaV. The 5' termini of both RNAs of the three viruses start with the same four nt 5'-AUAU-3'. The 3' terminal regions for RNA1 and RNA2 of each virus share 5-AACCCC-3', with BCLCaV having an extra C residue. Additionally, the 3' termini of RNA1 and RNA2 of each of the three viruses form four similar stem-loop structures. BCLCaV lacks the putative 12 kDa ORF on RNA1,

An isolate of RBDV from *Rubus glaucus* in Ecuador and the BCLCaV from Canada each had a concatenated form of RNA2 that contained an inverted CP gene. The RBDV isolate was cloned and sequenced from a dsRNA template using standard DOP-PCR and Sanger sequencing combined with RACE to obtain the terminal sequences. The dsRNA pattern on an agarose gel showed an unexpected band of about 3.2 kbp. The BCLCaV was sequenced using high throughput sequencing from a dsRNA template. The presence of this unusual RNA in two idaeoviruses suggest that there may be some secondary structure that leads to a rare formation of overhanging sequence that can be transcribed to yield this inverted repeat. In the case of RBDV it was found in a single plant,

Table 1 List of members of the genus *Idaeovirus* (top table). Percent amino acid identity between coat proteins and movement proteins sequences of Raspberry bushy dwarf virus (RBDV), Privet leaf blotch associated virus (PrLBaV), and Blackcurrant leaf chlorosis associated virus (BCLCaV) (bottom table)

Species names	Acronym	RNA1 accession #	RNA2 accession #
Raspberry bushy dwarf virus	RBDV	NC_003739	NC_003740
Privet leaf blotch associated virus	PrLBaV	NC_031341	NC_031342
Blackcurrant leaf chlorosis associated virus	BCLCaV	NC_034408	NC_034390
Tentative members			
Japanese holly fern mottle virus	JHFMoV	NC_013133	NC_013134
Virus	Coat Protein	Movement Protein	
	RBDV	PrLBaV	BCLCaV
RBDV	100	26	23
PrLBaV		100	61
BCLCaV			100

suggesting it may be maintained in vegetatively propagated material, but not encapsidated and therefore not pollen transmitted. It is not known if this is a common occurrence with BCLCaV since there are only two sources that have been sequenced.

Viral Proteins

The Coat Protein

The CPs of the idaeoviruses show no striking similarities with the CPs of other virus genera. An alternative to this sequence-based alignment program is to use secondary structure predictions to produce an alignment. This type of prediction with the RBDV CP predicted a long N-terminal sequence containing two α -helices and followed by a region containing eight β -sheets. If correct, this arrangement would resemble that of the CPs of many viruses that have isometric particles. The N-terminal parts of the CPs of RBDV and of viruses in the family *Bromoviridae* are relatively rich in basic aa, which are thought to be internal in virus particles and to be involved RNA encapsidation. The CPs of the ilarviruses and Alfalfa mosaic virus (AMV) are involved in 'genome activation' and required for virus replication. This phenomenon has also been reported for RBDV, where combining infectious clones of RNA1 and RNA2 gave very low level of infectivity but the addition of RNA3 greatly increased virus replication. With the ilarviruses and AMV addition of CP without adding the sub-genomic for the CP also was very effective at enhancing replication. RNA-binding domains have been shown to play a role in genome activation. In recent analyses using BindN several putative RNA-binding domains were identified in the CP of PrLBaV and similar domains were predicted in the CP of RBDV, though their involvement in genome activation has not been documented (Table 1).

The Movement Protein

RNA2 of RBDV corresponds in size and gene content to RNA3 of viruses in the family *Bromoviridae*. The ORF for the 39 kDa protein therefore corresponds in position to the ORF for the putative MP of members of the *Bromoviridae*. A limited sequence identity was detected in multiple alignment tests between the 39 kDa protein and the putative or actual MPs of several viruses. In further alignment tests of this sort, Mushegian and Koonin have detected a match between part of the 39 kDa protein sequence and a motif conserved in the MPs of some other viruses, such as Red clover necrotic mosaic virus (*Tombusviridae* family, *Dianthovirus* genus) and Soil-borne wheat mosaic virus (*Virgaviridae* family, *Furovirus* genus). These proteins were proposed to belong to a '30 kDa superfamily' of MPs. With PrLBaV the localization of the putative MP (38 kDa) in planta was studied by fusing green fluorescent protein (GFP) to its C-terminus. This fused protein was agro-infiltrated into leaves of *N. benthamiana*. When examined using confocal laser microscopy the GFP was accumulated at the plasmodesmata, which has been observed for other MPs. This provides strong evidence that the 38–39 kDa proteins coded by RNA2 are indeed MPs of the idaeoviruses and involved in virus trafficking.

The Replicase Protein

The sequence of the 190 kDa protein has similarities with replicase proteins of viruses in the Alphavirus superfamily. PFAM analysis identified three conserved domains in the predicted protein of the three idaeoviruses (RBDV, PrLBaV, and BCLCaV) coded by ORF1. The viral methyltransferases (MTR) belong to PF01660, which if found throughout the Alphavirus superfamily. The viral helicases (HEL) belong to PF01443 and the RdRp belong to PF00978. There was homology between parts of the idaeovirus replicase proteins and parts of the 183 kDa protein of Tobacco mosaic virus, but they showed greater similarity with the 90 kDa

protein (RdRp) of AMV (encoded by RNA2), and 125 kDa protein (HEL and MET), which are encoded by RNA1 of AMV. RBDV RNA1 therefore corresponds in size and gene functions to RNA1 plus RNA2 of the tripartite viruses in the family *Bromoviridae*.

Diagnosis and Detection

In *Rubus*, RBDV may induce yellowing, line patterns, or crumbly fruit, but such symptoms are unreliable for diagnosis because they may be due to infection with other disease agents, poor pollination, environmental conditions or poor nutrient management. Also, with many cultivars of raspberry (red and black) and blackberry RBDV is often asymptomatic. Detection and diagnosis therefore depend on bioassays, serological tests and/or reverse transcription PCR assays. RBDV can be detected reliably by ELISA. Many isolates of RBDV are readily detected by mechanical transmission to *C. quinoa* test plants, but this is most effective early in the season and is not as reliable as serological or RT-PCR assays. As the distribution of RBDV in individual *Rubus* plants can be erratic, leaves/leaflets for mechanical transmission or ELISA should be taken and pooled from at least three different nodes. For mechanical transmission newly emerging leaves should be used. For ELISA and RT-PCR, the youngest but fully expanded leaves are the optimal tissue for testing.

Serological Testing

Polyclonal antisera raised to S isolates have a moderate titer of ca. 1/512 in gel double diffusion tests. In agarose gel double-diffusion serological tests spur formation was used to document serological differences between S isolates from B isolates. Monoclonal antibodies to RBDV were able to distinguish three different epitopes of a Canadian isolate of RBDV. Each of the three monoclonal antibodies reacted with all isolates of RBDV from raspberry and blackberry. One of the monoclonal antibodies did not react with the isolates from grape or from *Rubus multibracteatus*. These monoclonal antibodies reacted in double antibody sandwich ELISA (DAS-ELISA) with all tested isolates from red raspberry, including B, S, and RB isolates. They are usually used in combination with a polyclonal antibody for trapping the virus onto the plate, the monoclonal antibodies as the second antibody after the virus has been trapped, and then followed with a goat anti-mouse conjugated antibody. This combination of antibodies are available commercially from several companies and can be used to detect RBDV with minimal equipment. There are not antibodies available for PrLBaV or BCLCaV, so testing for these two viruses is limited to biological assays or RT-PCR.

Reverse Transcriptase – PCR (RT-PCR)

There are more than 75 partial or full sequences of RBDV in GenBank, which is useful for designing detection primers that can capture the diversity of the virus. Detection primers have been designed for end-point RT-PCR and for quantitative RT-PCR. For most applications end-point PCR is used. Also, primers have been designed specific for RNA1 and for RNA2. There has been a report of RBDV RNA1 replicating and spreading systemic in raspberry plants. In this case, the ELISA tests and the end-point RT-PCR based on RNA2 sequences were negative, while the RT-PCR based on RNA1 was positive. It is not known if this is a common occurrence or was a rare event, but care should be taken to include some testing with primers specific for RNA1.

Therapy

Unlike many other viruses infecting *Rubus*, RBDV is not readily eradicated from *Rubus* plants by meristem tip-culture alone. However, heat treatment for several weeks at 36°C followed by meristem tip-culture can be used to eliminate RBDV from infected plants. Even with this combination, virus elimination of RBDV can be challenging. Cryotherapy has been used successfully to eliminate RBDV from raspberry, but again it was not very efficient. Chemotherapy combined with thermal therapy and meristem tip culture has shown some success in eliminating RBDV. RBDV appears to move very readily cell-to-cell and therefore can move into the meristematic dome more efficiently than most viruses.

Prevention and Control

The most important component of virus control in vegetatively propagated crops is to start with virus-tested planting stocks. This requires a clean plant production system. Such a system includes: (1) A facility to produce clean plants, hold the top-tier or G1 plants generally under protected culture (screenhouses to minimize vectors) or in long-term tissue culture at 4°C (**Fig. 5**), and that can retest G1 plants on a regular schedule to ensure they remain clean; (2) A certification scheme that performs audit testing and ensures best management practices are followed during the multiple stages of increase (G2 and G3 blocks) required between the G1 (Foundation Block) and the certified plants sold to growers, (3) Increase blocks that are monitored by the certifying agency and tested for viruses and other targeted pathogens as outlined in the certification standard; (4) A system to track the certified plants under the certification scheme so that disease issues can be traced-back or traced-forward as needed.



Fig. 5 Plants of 'Columbia Star' blackberry in gas permeable sealed bags. These plants can be held at 4°C for up to three years without any maintenance or risk of reinfection. Photographs courtesy of R.R. Martin.

Because RBDV is transmitted in association with infected pollen, the only methods of controlling it in crops are to grow resistant cultivars, or to plant healthy susceptible cultivars in isolation from possible sources of infection. If the virus is present in native *Rubus* spp., isolation from infected sources can be challenging in many growing regions. All black raspberry, blackberry, and blackberry/raspberry hybrids tested were susceptible to the S strains of RBDV. Many red raspberry cultivars are resistant, possibly immune, to S isolates (the most widespread isolates worldwide). This resistance is conferred by the presence of a single dominant gene, *Bu*. All *Bu*-containing cultivars are susceptible by grafting with RB isolates, though the cultivars, 'Heritage', 'Haida', 'Rannaya Sladkaya' were difficult to infect and did not become infected when exposed to the RB isolates under field conditions. There are cultivars that show 'field resistance' and are slower to become infected in the field to either the S or RB isolates. In some cases, cultivars that are susceptible to grafting infection with the S isolates remained RBDV-free in the field up to 20 years, even under conditions where adjacent plantings were nearly 100% infected after 5 years. The mechanism for this has not been studied, but if high fruit quality can be combined with slow to become infected, such cultivars could have a place in production systems. Breeding for field resistance to the RB strain may be a more realistic goal since resistance to this strain has not been identified. RBDV susceptibility has resulted in some cultivars with excellent fruit quality to be abandoned by growers.

Studies on this form of resistance indicated that its mode of inheritance was complex but seemed to depend on the presence of gene *Bu*, together with a second resistance component whose inheritance was probably multigenic. In field studies, some raspberry cultivars, including some that do not contain gene *Bu*, are more resistant than others to natural infection with RB isolates.

'Meeker' red raspberry has been developed with resistance to RBDV that makes use of RNAi. The plants were free of RBDV when challenged in the field for six years under extreme disease pressure when all 202 control 'Meeker' plants in the same test plots were infected after three years in the field. The resistance was stable and effective against graft inoculations that were repeated three times and the transgenic raspberry produced fruit of quality indistinguishable from wild-type 'Meeker'. Though the resistance is effective, the plants are on hold and applications for approval for human consumption to the Food and Drug Administration in the U.S. have not been made, pending a change in public perception on the use of this technology. Additionally, there has been 20 years of breeding since these transgenic lines were developed, and the standard for fruit quality and yield has surpassed that of 'Meeker'.

Relationships With Other Viruses

Idaeoviruses resemble viruses in the genus *Ilarvirus*, family *Bromoviridae* in having quasi-spherical particles that are easily distorted in preparations viewed in an electron microscope. Also, they are transmitted vertically and horizontally in association with pollen as are the ilarviruses. RNA1 of idaeoviruses contains the HEL, MET, and RdRp functions, while in the ilarviruses, RNA1 contains the HEL function and RNA2 contains the MET and RdRp functional domains. RNA2 of idaeoviruses resemble RNA3 of ilarviruses in the arrangement and sizes of its encoded gene products, and the generation of a 3' terminal sub-genomic RNA (sgRNA). RBDV is serologically unrelated to all recognized ilarviruses tested. Also, RBDV differs from ilarviruses in the sedimentation behavior of

its particles, in the number and sizes of its RNA molecules, and in having a bipartite genome. Taken together, these properties distinguish idaeoviruses from all other well-characterized viruses and support their status as a separate genus. The genus *Idaeovirus* has not been assigned to a family, but has many molecular properties in common with members in the family *Bromoviridae*, the major difference being that the multi-functional replicase protein of idaeoviruses is coded one RNA whereas in the family *Bromoviridae* these functional domains are coded by two RNAs. The idaeoviruses have bipartite genomes whereas the six genera in the *Bromoviridae* have tripartite genomes.

Further Reading

- Derrick, K.S., Beretta, M.J., Barthe, G.A., 2006. Detection of an idaeovirus in citrus with implication as to the cause of citrus blight. *Proceedings of the Florida State Horticultural Society* 119, 69–72.
- ICTV Virus Taxonomy, 2018. Release. Available at: <https://talk.ictvonline.org/taxonomy/> (accessed 08.07.2019).
- Isogai, M., Yoshida, T., Nakanowatari, C., Yoshikawa, N., 2014. Penetration of pollen tubes with accumulated Raspberry bushy dwarf virus into stigmas is involved in initial infection of maternal tissue and horizontal transmission. *Virology* 452–453, 247–253.
- James, D., Phelan, J., 2017. Complete genome sequence and analysis of Blackcurrant leaf chlorosis associated virus, a new member of the genus *Idaeovirus*. *Archives of Virology* 162, 1705–1709.
- Jones, A.T., Mayo, M.A., 1998. Raspberry bushy dwarf virus. *Descriptions of Plant Viruses* 360, 6. Available at: <https://www.dpvweb.net/dpv/showdpv.php?dpvno=360> (accessed July 2019).
- MacFarlane, S.A., McGavin, W.J., 2009. Genome activation by Raspberry bushy dwarf virus coat protein. *Journal of General Virology* 90, 747–753.
- MacFarlane, S.A., 2011. Genus *Idaeovirus*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier Academic Press, pp. 1173–1175.
- Martin, R.R., MacFarlane, S., Sabanadzovic, S., *et al.*, 2013. Viruses and virus diseases of *Rubus*. *Plant Disease* 97, 169–182.
- Navarro, B., Loconsole, G., Giampetruzzi, A., *et al.*, 2017. Identification and characterization of Privet leaf blotch-associated virus, a novel idaeovirus. *Molecular Plant Pathology* 18, 925–936.
- Pleško, I.M., Marn, M.V., Širca, S., Urek, G., 2009. Biological, serological and molecular characterization of Raspberry bushy dwarf virus from grapevine and its detection in the nematode *Longidorus juvenilis*. *European Journal of Plant Pathology* 123, 261–268.
- Quito-Avila, D.F., Ibarra, M.A., Alvarez, R., Peralta, E.L., Martin, R.R., 2014. A Raspberry bushy dwarf virus isolate from Ecuadorean *Rubus glaucus* contains an additional RNA that is a rearrangement for RNA2. *Archives of Virology* 159, 2519–2521.
- Thekke-Veetil, T., Ho, T., Postman, J.D., Tzanetakis, I.E., 2017. Characterization and detection of a novel idaeovirus infecting blackcurrant. *European Journal of Plant Pathology* 149, 751–757.
- Valverde, R.A., Sabanadzovic, S., 2009. A novel plant virus with unique properties infecting Japanese holly fern. *Journal of General Virology* 90, 2542–2549.

Iilarviruses (*Bromoviridae*)

Aaron Simkovich, Agriculture and Agri-Food Canada, London, ON, Canada and The University of Western Ontario, London, ON, Canada

Susanne E Kohalmi, The University of Western Ontario, London, ON, Canada

Aiming Wang, Agriculture and Agri-Food Canada, London, ON, Canada

© 2021 Elsevier Ltd. All rights reserved.

This is an update of K.C. Eastwell, Iilarvirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00642-7](https://doi.org/10.1016/B978-012374410-4.00642-7).

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

ELISA Enzyme-linked immuno-sorbent assay

ER Endoplasmic reticulum

kb Kilobases; the size of a ssDNA or ssRNA molecule

kDa Kilodaltons; the size of a protein

MP Movement protein

NGS Next generation sequencing

nm Nanometer(s)

ORF Open reading frame(s)

PCR Polymerase chain reaction

RBD RNA binding domain

RdRp RNA-dependent RNA polymerase

RT-PCR Reverse transcription polymerase chain reaction

sgRNA Sub-genomic RNA

TLS Transfer RNA-like structures

UTR Untranslated region

VIGS Virus-induced gene silencing

VRC Virus replication complexes

Glossary

Bicistronic A viral RNA that encodes two genes which are often separated by a non-coding intergenic region.

Cross protection It describes a phenomenon in that an initial infection of a host plant with a mild strain of a virus induces resistance in that plant to the infection of another, closely related virus potentially protecting the plant from disease caused by a more virulent isolate.

Genome activation Iilarviral genomic RNAs alone are unable to establish infection in plants, unless the coat protein is present. This function of the coat protein is termed genome activation and is specific for ilarviruses and closely related alfamoviruses. The event is triggered by binding of the coat protein RNA binding domain to the 3' terminus of genomic RNA.

RNA silencing A fundamental, evolutionarily conserved and sequence-specific mechanism that is triggered by double-stranded RNA and regulates gene expression in

eukaryotes. It is a primary antiviral defense mechanism in plants and other living organisms.

RNA silencing suppressor A countermeasure to RNA silencing, often a protein encoded by a virus which interrupts a single, or multiple steps in the RNA silencing pathway such as binding to small interfering RNA and thereby preventing their incorporation into the RNA induced silencing complex.

Sub-genomic RNA A segment of RNA generated from a genomic RNA via an international promoter that has the same 3' end as the genomic RNA, but has a deletion at the 5' end. The sub-genomic RNA makes it possible to efficiently translate the downstream open reading frame of the genomic RNA.

Tripartite genome A viral genome consisting of three genomic fragments, which are encapsidated into three separate virions.

Introduction

Iilarviruses are a group of isometric and labile viruses that are distributed worldwide and infect many agriculturally relevant herbaceous and woody hosts including fruit trees, vegetables, and ornamentals. For long time, the economic importance of ilarviruses has been underestimated as they are often considered as latent pathogens, particularly when they infect fruit trees. Since 1990s, a large number of ilarviruses have been molecularly identified, and this group of viruses has received more and more attention. In this article, we briefly discuss the classification, virion structure, genomic organization, and infection cycle of ilarviruses, summarize the current knowledge of their replication mechanisms and transmission modes, highlight the control strategies, viral symptoms and pathogenesis of the viruses, and finally introduce diagnosis of ilarviruses.

Table 1 A list of recognized species in the genus *Iilarvirus*, family *Bromoviridae*

Subgroup	Species	Acronym	GenBank accessions
Subgroup 1	<i>Ageratum latent virus</i>	ALV	JX463341
	<i>Blackberry chlorotic ringspot virus</i>	BCRSV	DQ091194
	<i>Parietaria mottle virus</i>	PMoV	AY496069
	<i>Privet ringspot virus</i>	PrRSV	KT290040
	<i>Strawberry necrotic shock virus</i>	SNSV	AY743591
	<i>Tobacco streak virus</i>	TSV	U75538
Subgroup 2	<i>Asparagus virus 2</i>	AV-2	EU919667
	<i>Citrus leaf rugose virus</i>	CiLRV	U17726
	<i>Citrus variegation virus</i>	CVV	EF584665
	<i>Elm mottle virus</i>	EMoV	U34050
	<i>Lilac ring mottle virus</i>	LiRMoV	EU919669
	<i>Spinach latent virus</i>	SpLV	U93193
	<i>Tomato necrotic streak virus</i>	TomNSV	KT779205
Subgroup 3	<i>Tulare apple mosaic virus</i>	TAMV	AF226161
	<i>Apple mosaic virus</i>	ApMV	AF174585
	<i>Blueberry shock virus</i>	BIShV	KF031038
	<i>Lilac leaf chlorosis virus</i>	LLCV	FN669168
Subgroup 4	<i>Prunus necrotic ringspot virus</i>	PNRSV	AF278535
	<i>Fragaria chiloensis latent virus</i>	FCiLV	AY707771
Unassigned	<i>Prune dwarf virus</i>	PDV	AF277662
	<i>American plum line pattern virus</i>	APLPV	AF235165
	<i>Humulus japonicus latent virus</i>	HJLV	AY500237

Classification

The *Iilarvirus* genus was first recognized during the third meeting of the International Committee on the Taxonomy of Viruses in 1975. The genus, along with five other genera including *Alfamovirus*, *Anulavirus*, *Bromovirus*, *Cucumovirus* and *Oleavirus*, belongs to the *Bromoviridae* family. Among the six genera, *Iilarvirus* is particularly closely related to *Alfamovirus*. The name “*Iilarvirus*” is in fact a siglum derived from isometric labile ringspot viruses, describing several characteristics of ilarviruses: the shape of virions, the fragility or labile nature of viral particles and lastly, the frequently observed ring spotting symptom on infected hosts. Currently, the genus consists of 22 recognized members (Table 1). Previously ilarviruses were sub-grouped mainly based on serological properties with antibodies against the coat protein (CP). Accumulated sequence data support the serological relationships for most ilarviruses. However, extensive phylogenetic analyses also identify some species that had been inappropriately classified possibly due to conserved amino acid (aa) sequences and secondary structures in their CPs. Currently, the 22 recognized ilarviruses have been phylogenetically classified into four sub-groups in addition two unassigned viruses (Table 1, Fig. 1).

Virion Structure

Virions of ilarviruses are most commonly found as quasi-isometric shapes (Fig. 2(B)). The capsid of the virus is comprised of 180 coat protein (CP) units arranged in a $T = 3$ symmetry, ranging in diameter from 26 to 35 nm. Additionally, some members, such as Prune dwarf virus (PDV) and Prunus necrotic ringspot virus (PNRSV) also produce bacilliform particles with a diameter of 18–26 nm and 30–85 nm in length. As the virion contains the encapsulated genomic RNA, the size of virions is often related to the length and amount of RNA contained. The regions of the CP in some members is implicated in the formation and degree of lability of the virion.

Genome

Members of the *Iilarvirus* genus have a tripartite genome consisting of three separate positive-sense single-stranded RNA molecules, i.e., RNA1, RNA2 and RNA3, all of which bear a 5′-7-methyl-G cap structure but lack a poly A tail. The 3′ ends of genomic RNAs code for hairpin loop structures potentially forming intricate pseudoknots. The genomic organization of Tobacco streak virus (TSV), the type member of the *Iilarvirus* genus is presented in Fig. 2(A). Among three ilarviral genomic RNAs, RNA1 is the largest

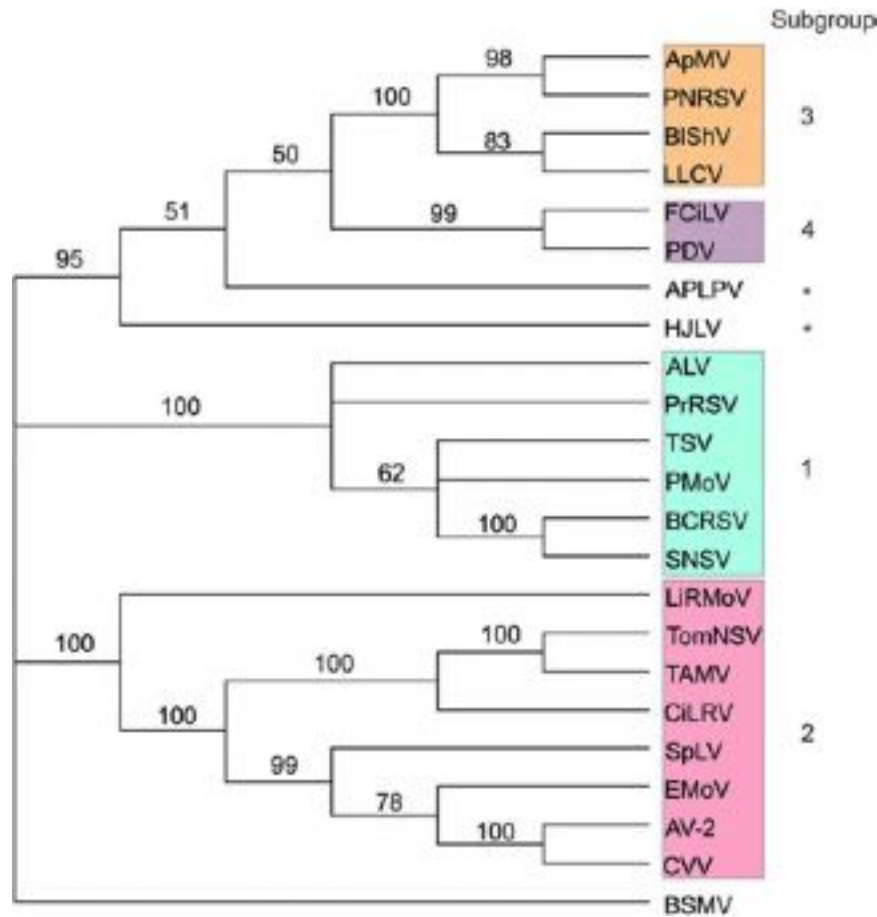


Fig. 1 Phylogram showing nucleotide sequence relationships of ilarvirus RNA dependent RNA polymerase (ORF2a) with Barley stripe mosaic virus (BSMV) included as an outgroup (Accession NC_003478). Subgrouping (shown at right) is based on phylogenetic data, host range and serology. *Species unassigned to subgroups. Numbers at branch nodes shows percentage of bootstrap replicates in which that branch occurs. Abbreviations and sequence accession numbers are listed in [Table 1](#).

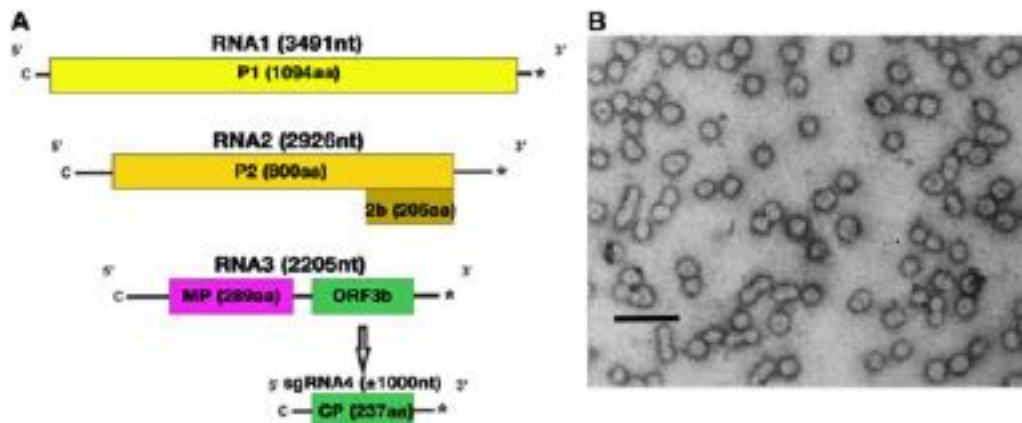


Fig. 2 (A) The genome structure of Tobacco streak virus, the type member of the *Iilarvirus* genus. Each single stranded positive sense RNA genomic fragment is shown and encoded proteins are shown as boxes with approximate nt and aa lengths, respectively. The complete genome of TSV isolate WC has a total length of 8622 nt. The P1 protein of TSV has 1094 aa with a molecular mass of 123.4 kDa. Encoded from ORF2a, the P2 protein has a sequence length of 800 aa and a mass of 91.5 kDa. The 2b protein has a sequence length of 205 aa, and a mass of 22.4 kDa. The movement protein (MP) of TSV has a length of 289 aa and a mass of 31.6 kDa. The coat protein (CP) encoded by ORF3b located in sgRNA4 is 237 aa in length and has a mass of 26.2 kDa. Each RNA fragment has a 5' cap shown as C and 3' tRNA-like structure shown as *. The large arrow denotes the transcription of sub-genomic RNA4. (B) A transmission electron micrograph showing virions of Prunus necrotic ringspot virus (PNRSV). Photograph courtesy of the Rothamsted Experimental Station, UK. The bar represents 100 nm.

genomic component, and contains a single open reading frame (ORF1), which encodes the P1 protein. This protein contains both methyltransferase (near the N-terminus) and helicase (near the C-terminus) domains. P1 may also contain a transmembrane domain at the C-terminus. It has been suggested that it is the P1 protein that recruits and anchors the viral RNA to the assembly site to support virus replication. Consistent with this assumption, subcellular localization studies suggest that P1 is associated with membranes of the vacuole, the endoplasmic reticulum (ER) and chloroplasts. Therefore, P1 is considered as an essential component of the viral replication complex (VRC). In subgroups 1 and 2, RNA2 is bicistronic with ORF2a and ORF2b. ORF2a encodes the P2 protein, which is the putative RNA dependent RNA polymerase (RdRp), whereas ORF2b codes for the 2b protein. P2 contains several conserved domains of RdRp such as the RNA binding domain and the GDD motif. Therefore, P2 is believed to be part of the VRC and directly responsible for biosynthesis of viral RNA. The 2b protein was originally proposed to be involved in systemic infection and was later proven to be a suppressor of post-transcriptional gene silencing to counteract against RNA silencing, a primary host antiviral defense mechanism. Members of subgroups 3 and 4 do not possess ORF2b. The third fragment, RNA3 is bicistronic and contains two ORFs, ORF3a and ORF3b. The former encodes a movement protein (MP), and the latter is expressed via a sub-genomic fourth RNA fragment (sgRNA4) for the translation of the CP. The ilarviral MP is a member of the 30K MP superfamily. Members of this superfamily are known to locate to the plasmodesmata (PD) and alter size exclusion limits of these cellular channels to assist virus cell-to-cell movement. The CP is a multifunctional protein and plays essential roles in different stages of the virus infection cycle such as virion formation and replication. A zinc finger motif has been identified in CPs of many plant viruses including some ilarviruses such as TSV and is presumed to be associated with nucleotide binding and virion formation. However, the fact that some ilarviruses such as PDV lack this motif supports the notion that this motif is not an absolute requirement for virion formation.

Infection Cycle

The infection cycle of ilarviruses may be artificially divided into several stages. As plant viruses usually have a small genome with limited coding capacity, they must hijack host pathways and factors to support their infection. Upon entry into host cells, ilarviruses undergo uncoating to remove the CP shell to expose the viral genome to the host translation machinery for translation on the rough ER. Subsequently, the P1 protein may target the membranes of preferred organelles such as the ER, vacuole and/or tonoplast and induces the formation of spherules or small vesicular structures where VRCs are assembled. The VRC catalyzes biosynthesis of the minus (−) strand RNA using the viral genomic RNA as template and further multiplies the (+) strand RNA using the newly synthesized (−) RNA as template. The (+) and (−) RNA strands are biosynthesized in an asymmetrical manner. During AMV replication, the ratio of (+) RNA versus (−) RNA is approximately 100:1. The progeny (+) RNAs are encapsidated by CPs to form virions. The newly synthesized RNA or virions may interact with MP, and the ribonucleoprotein complexes are transported intracellularly to the PD where they move from the primarily infected cells to neighboring cells. MP is definitely required for this process. For some ilarviruses, virion is not necessary for virus cell-to-cell spread, but the presence of CP is required for all ilarviruses to move intracellularly. For long-distance movement from the local infection to remote sites of the plant, the virus is transported from the mesophyll via bundle sheath cells, phloem parenchyma, and companion cells into phloem sieve elements and then following the source-to-sink flow of photo-assimilates and unloaded to sink tissues. Iilarviruses are likely transported for the long distance in the form of virions.

Replication

Most members in the *Bromoviridae* family possess transfer RNA-like structures (TLS) in the 3′ untranslated region (UTR) of their genomic RNAs, and these TLS regions interact with the cap at the 5′ end to circularize the viral RNA and promote the translation process. Different from these viruses, alfalfa mosaic virus (AMV) and ilarviruses lack TLS in their genomes. In this case, the CP is required to establish infection possibly by activation of viral protein translation and regulating transcription of new (+)/(−) viral RNA, a phenomenon termed genome activation. In lieu of TLS, this requirement of CP for virus replication is now a distinguishing feature of both *Alfavirus* and *Iilarvirus* genera. Viral replication mechanism has been studied more in depth for alfamoviruses than ilarviruses and due to the common feature of the CP requirement for genome activation has led to analogous conclusions for replication strategies. The process of genome activation is based on interactions between the viral CP and secondary structures found within the 3′ UTR of viral RNAs. A highly conserved RNA binding domain (RBD) containing crucial arginine (R) residue(s) is present in the N-terminal CP sequences of AMV and some ilarviruses, and the RNA-binding consensus motif in this RBD has the aa sequence Q/K/R-P/N-T-X-R-S-R/Q-Q/N/S-W/F/Y-A. Additional sequence analyses have identified a second putative RNA binding consensus sequence V(T/S)(R/N)RQ(S/R)RNA(A/R)RAAX(Y/F)R which is also conserved in at least six other ilarviruses. Some ilarviruses such as PNRSV do not have any of the two consensus RNA-binding sequences. However, the CP sequence in these viruses is R-rich. In the CP of PNRSV, there is a stretch of about 20 aa with five R residues that has the capacity to bind to the 3′ UTR of RNA3. In addition, the N-terminal region of the CPs of many ilarviruses including PNRSV possess a zinc finger motif that is believed to increase RNA binding affinity.

The process of genome activation is not species-specific, and some ilarviruses are able to use CPs from other ilarviruses to activate their genomes. This ability to reciprocally activate infectivity of the different viruses by their CPs can extend to the intergenus level. It

has also been shown that CP of AMV (an alfamovirus) and that of some ilarviruses can be interchanged to induce genome activation. In the absence of a translated CP, the presence of sgRNA4 is sufficient to result in an active genome as the CP is translated from this sub-genomic RNA. In the absence of sgRNA4 or CP, the transcription of viral RNA is preferred, however the addition of either one of these two elements leads to a shift towards translation. It is suspected that the binding of CP inhibits transcription, and that the RNA secondary structures enable binding by the CP. Complex secondary structures have been identified among sequenced ilarviruses, as the 3' UTRs of genomic RNAs possess conserved single stranded motifs of (A/U)(U/A/G)GC, which flank the regions potentially forming hairpin loop structures acting as binding sites for the CP. It has been suggested that the CP-RNA complex formed in ilarviruses are functionally analogous to the complex formed by poly A tract and the poly A binding protein in eukaryotes which is known to enhance translation of proteins. This has been further supported by the finding that the AMV CP does in fact bind to eukaryotic translation initiator complexes eIF4F and eIFiso4F.

Two models have been proposed to explain the role of the CP in the early stages of the ilarviral infection cycle. The conformational switching model proposes that the 3' UTR of AMV and ilarviruses can fold into two separate structures. The first is a pseudoknot which is functionally equivalent to the TLS and promotes the binding of viral RdRp and synthesis of negative strand RNA. The second structure is a linearized series of hairpin loops. This model proposes that early binding of CP causes a conformational change, or switch, unwinding the pseudoknot into a series of hairpin loops which cannot be bound by the RdRp. This model is supported by several lines of evidence from studies on AMV. The pseudoknot in the 3' UTR is essential for AMV replication and disruption of this structure inhibits viral replication. CP binding to the 3' UTR is inhibited when magnesium is added to stabilize the pseudoknot structures. The CP can bind to mutated 3' UTR sequences which are unable to form pseudoknots even after addition of magnesium. Consistent evidence has also been obtained from studies with ilarviruses. For example, the CP of PNRSV is unable to bind to the 3' UTR of genomic RNAs in the presence of magnesium. Overall, this model explains the role of CP in the regulation of RNA synthesis and protein translation.

The second model suggests that CP binding to the 3' UTR forms a complex critical for viral replication. This model contradicts the first model by predicting that the structure of the 3' UTR undergoes a different conformational change and CP binding to the 3' UTR creates a more compact structure. Structural analyses show that the unbound 3' UTR of AMV is flexible, supporting this model. When a short peptide identical in sequence to the AMV CP RNA binding domain is bound, the resulting CP-RNA complex is more rigid and compact due to base pairing between the AUGC repeats that flank hairpin loops. Modeling of the CP binding along the entire length of the AMV 3' UTR, shows that the predicted molecule has a compact, rod shaped structure with hairpins protruding from the center. In vitro studies have also shown that in absence of CP, the binding of the RdRp protein to labeled RNA is weak, and addition of CP significantly enhances the binding ability of RdRp to the labeled RNA. Therefore, CP binding to the 3' UTR promotes RNA transcription.

Transmission

Iilarviruses have several modes of transmission. The majority of ilarviruses have a relatively limited natural host range, with TSV as an exception which can infect a number of plant species naturally. All ilarviruses can infect a broad range of experimental herbaceous hosts via mechanical transmission. Iilarviruses naturally infecting herbaceous plants are easily transmitted mechanically via accidental damage during farming practices or by insects. Many ilarviruses naturally infect woody hosts such as stone fruit trees. However, the probability of mechanical transmission in the woody hosts is low. Perhaps the most common mode of ilarvirus transmission in these hosts is by vegetative propagation, which is a common horticultural practice. For example, grafting of scions to rootstocks is used for clonal propagation of woody fruit trees. In some cases, it is also used for herbaceous crop production (i.e., tomato, cucumber and watermelon) to take advantage of desirable traits of rootstocks such as disease resistance. In grafting, the use of infected material as either a root stock or scion facilitates the spread of viruses. Another method of vegetative propagation involves the direct planting of cuttings (i.e., alfalfa and hops), or division of crowns (i.e., rhubarb and asparagus). The propagation of new explants from an infected mother plant provides another simple yet common vector for virus spread. The finding that some ilarviruses are spread via seed poses a further challenge as rootstocks are commonly grown from seedlings. The spread of viruses by seed presents a challenge in the production of seed grown crops, and planting of uncertified seed stock has the potential to spread members of this genus at the beginning of a new growing season and even to new agricultural production regions.

Aside from transmission by vegetative propagation and seeds, ilarviruses may also transmit via plant-to-plant, i.e., hops, pollen grains and insects. For example, pollinating honeybees have been shown to transmit both PNRSV and PDV. The spread of TSV by thrips is well documented, and several species of thrips have been identified as vectors of ilarviruses. Feeding by thrips that carry virus-contaminated pollen, or feed on leaves covered by virus-contaminated pollen causes mechanical damage to the host plant, providing a point of entry for the virus.

Control

Treatment of any infections by plant viruses including ilarviruses is difficult. Therefore, methods to prevent viral infection should be implemented. Cross protection, also known as superinfection exclusion, describes the process in which the exposure to a mild

virus isolate induces the resistance against a more virulent, closely related isolate. The virus isolate selected for inoculation should be endemic to the growing region, ideally latent and showing minimal impact on the host. Cross protection does not offer broad resistance to unrelated viruses, and in fact, for some viruses, cross protection is only strain specific. A very successful example for the control of plant viruses using cross protection is to protect citrus from the devastating diseases caused by virulent isolate of *Citrus tristeza virus*, which belongs to the genus *Closterovirus* in the family *Closteroviridae*. Cross protection has also been reported to control infections by ilarviruses. Cherry rugose mosaic disease is a serious disease of sweet cherry and the causal agent is a virulent strain of PNRSV. Less virulent isolates of PNRSV have been identified for the control of the disease in sweet cherry. One possible mechanism of cross protection is RNA silencing. It has been suggested that virus infection is involved in the formation of double-stranded RNAs which trigger virus induced gene or RNA silencing (VIGS). VIGS explains why infection by an attenuated version of a virus can induce resistance to the virus isolates sharing similar sequences with the attenuated isolate. VIGS can also be used to silence host genes by insertion of a short fragment of a target gene into the attenuated virus. Recently, using a modified PNRSV, we have successfully silenced a gene in plum that is required for plum pox virus infection. The resulting plum plants show resistance to Plum pox virus.

Crop management to minimize risk of infection, and methods of early detection are the effective strategy to combat infections by ilarviruses and other plant viruses. Certification programs are important measures which should be implemented for introduction of viral pathogens. Many countries have programs to certify growing materials (seed, rootstocks, scion cuttings) as being free of viruses after testing. Using certified virus free materials is important when starting both perennial and annual crops. In particular for fruit trees which are propagated over decades in densely populated orchards such as members of the *Prunus* genus, the use of virus free material is a worthwhile long-term investment.

Hygiene practices including cleaning of pruning shears and other implements which contact plants should be routinely performed. When virus infections are found, common practices place an emphasis on sanitation and eradication including the removal and destruction of infected plants and plant material, removal of perennial weeds which may act as alternative hosts near annual growing fields. In response to the spread of viral diseases caused by transport of infected plant materials, some countries have imposed legislative methods to fight plant disease by enforcing strict regulations to restrict the movement of plant material. Control of vectors such as thrips or other insect vectors is another effective approach to minimize spread of insect- and/or pollen-transmitted viruses.

It is worth mentioning that the introduction of resistance using modern genetic engineering technologies has been demonstrated to be highly effective in the control of plant viruses. This type of resistance operates through RNA silencing by generation of transgenic plants that express a double-stranded RNA fragment sharing the sequence of the target virus. However, due to public concerns and regulatory barriers, this technology has not been widely used for the control of ilarviruses.

Symptoms and Pathogenesis

Symptoms induced by ilarviruses vary for different virus-plant combinations. As the name suggests, ringspot is a common symptom found on young leaves of ilarvirus-infected plants (**Fig. 3(A)**). Ringspots often become necrotic and the necrotic tissue dies leaving a shot-hole symptom pattern. Infection by TSV induces various symptoms in tobacco leaves including vein clearing, streaking, ringspots, stunting, leaf deformation and systemic necrosis. Cucumber plants infected with PDV initially show small chlorotic lesions, then a more severe mottling of leaves and stunted growth (**Fig. 3(B)**). Iilarvirus-infected plants often undergo an initial shock phase with the development of disease symptoms such as necrotic ringspots, followed by symptom attenuation and eventual recovery on new leaves. On woody plants, this disease process may take a few years. For example, infection of blueberry and cranberry by Blueberry shock virus, which is pollen-transmitted with the aid of honeybees, induces clear symptoms on the entire plant or some branches in the first 1–2 years. Leaves and flowers of infected plants appear blighted, and gradually drop off. After 1–3 years, the infected plants become asymptomatic. Cranberry bushes infected with TSV also show symptom recovery. However, different from the typical recovery phenomenon in which the abatement of symptoms is associated with antiviral RNA silencing and decreased viral accumulation (termed viral clearance), symptom recovery following infection by some ilarviruses may be caused by different mechanisms.

The molecular determinant for the recovery phenomenon in TSV has been identified. A single nucleotide substitution (A → G) in the intergenic region of RNA3, upstream of the transcription start site of sgRNA4 is responsible for the failure of a TSV isolate to initiate recovery in the host. Plants infected with this recovery deficient isolate develop disease symptoms which never subside. Interestingly, the viral titer in distal plant tissues is lower than that in the corresponding tissues of plants infected with the TSV wild type virus that can induce symptom recovery. These data support the assumption that symptom recovery in plants infected by at least some ilarviruses is not necessary to be related to viral titers. It is possible that ilarvirus-induced RNA silencing downregulates host resistance genes, which may contribute to an infection with abated symptoms where a higher viral titer is maintained.

Determinants of pathogenicity have been identified in some ilarviruses such as Asparagus virus 2 (AV-2) and PNRSV. The 2b protein encoded by AV-2 ORF2b acts as a suppressor of systemic RNA silencing to indirectly function as a determinant by counteracting the host plants defense mechanism. A more direct example of a disease determinant came from a study on PNRSV. In this study, full-length infectious cDNA clones derived from a pathogenically aggressive isolate (Pch12) and a mild isolate (Chr3) of PNRSV (**Fig. 3(C)**) were used to identify determinants of pathogenicity. By reassortment of genomic segments, swapping

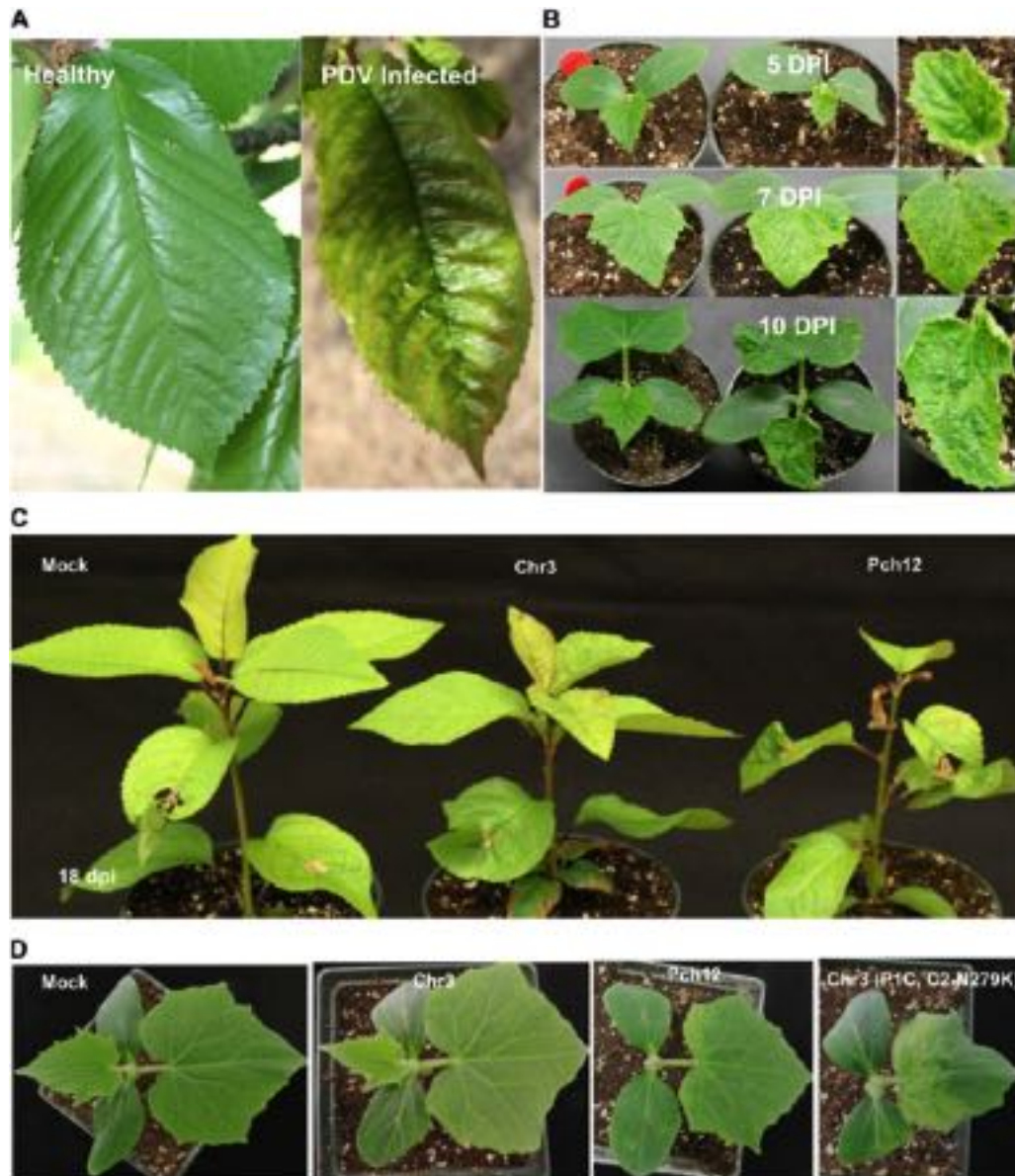


Fig. 3 Ilarvirus-infected herbaceous experimental and natural host plants. (A) Healthy, asymptomatic leaf from sweet cherry (*Prunus avium* cv. Vogue) compared to leaf from an individual infected with PDV showing leaf deformation and yellowing. (B) Experimental herbaceous host cucumber (*Cucumis sativus* cv. Wisconsin) showing viral symptoms after being mechanically inoculated with homogenate prepared from PDV infected foliar tissue shown in A. Newly emerging and fully expanded symptomatic first true leaf and symptomatic second true leaf are magnified to show mottling and vein clearing symptoms at 5, 7, and 10 days post inoculation (DPI). (C) Symptoms induced by two infectious cDNA clones of PNRSV on its natural host cherry (*Prunus avium* cv. Bing) at 18 dpi. The two PNRSV isolates differ in symptom severity Chr3 induces the development of milder symptoms (necrotic spots on young leaves) whereas Pch12 causes severe top wilting and more severe necrosis. (D) Identification of pathogenicity determinants of PNRSV. Herbaceous host cucumber plants (*Cucumis sativus* cv. Straight Eight) were inoculated with two isolates of PNRSV via agrobacterium mediated infiltration of infectious cDNA clones. The mild isolate Chr3 shows minimal disease however the more aggressive isolate, Pch12 causes stronger disease symptoms in cucumber as well. The chimeric virus formed from the Chr3 isolate of PNRSV by the substitution of the 3' region of RNA1 from Pch12, and the substitution of asparagine to lysine at position 279 in RNA2 shows severe symptom development leading to the identification of pathogenicity determinants in PNRSV. Figs. 3(C and D) reproduced with permission from the American Phytopathological Society.

of partial genomic segments, and site-directed mutagenesis, it was shown that the 3' terminal nt sequence in RNA1 and aa K at residue 279 in RNA2-encoded P2 are the severe virulence determinants and are required for severe virulence and high levels of viral accumulation (Fig. 3(D)). PNRSV RNA1 and RNA2 seem to codetermine viral pathogenicity to adapt to alternating natural *Prunus* hosts, likely through mediating viral accumulation.

Diagnosis

Diagnosis of an infection by ilarviruses is temporally sensitive as initial shock symptoms subside, making the distinction of healthy and infected plants difficult. Symptoms are easily seen at the beginning of growing seasons before peak seasonal temperatures are reached; young, newly emergent leaves usually display the clearest symptoms. Common diagnosis techniques for other plant viruses are applicable to ilarviruses. Since ilarviruses usually have a wide experimental host range, indexing on herbaceous and woody indicators is an option for detection. The main advantages of this method are the relatively low costs and minimal reagents required. It should be noted that chemical additives such as 2-mercaptoethanol, ethylenediaminetetraacetic acid, cysteine hydrochloride, sodium diethyl-dithiocarbamate and polyvinyl-pyrrolidone are routinely added to inoculum preparations. These additives protect the virion from degradation while infected host tissue is homogenized which results in the release of endogenous RNA degrading enzymes, oxidizing agents, and tannins, all compounds that are known to degrade virions or inactivate them. There are several disadvantages to indexing including the long duration of experiments, the labor intensive nature of these studies and the low specificity. Generation of antibodies specific to some ilarviruses CPs has led to the commercial availability of serological diagnostic kits based on enzyme-linked immunosorbent assay (ELISA), which is fast, accurate, highly sensitive and specific, compared to other diagnostic methods. Disadvantages of serological diagnostic kits are that they are often highly specific, and are not sensitive to detect low titer of viruses. Nucleic acid-based detection techniques such as reverse transcription polymerase chain reaction (RT-PCR) serve as more powerful diagnostic tools. Multiplex PCR can be used to detect multiple viruses simultaneously with greater sensitivity and specificity compared to serological techniques. Indeed, multiplex PCR has also been shown to allow the detection of several ilarviruses simultaneously. Another powerful molecular technique is next generation sequencing (NGS). This method requires the greatest amount of time for preparation and is the most expensive option. However, NGS is a broad detection tool and additionally provides the user with incredibly extensive sequence data. Such sequencing data from infected hosts have been used to identify causal agents of plant disease including novel viruses, and even to obtain full-length genome sequences of some viruses.

Further Reading

- Bujarski, J.J., Gallitelli, D., García-Arenal, F., *et al.*, 2019. ICTV virus taxonomy profile: Bromoviridae. *Journal of General Virology* 100 (8), 1206–1207. doi:10.1099/jgv.0.001282.
- Cui, H., Hong, N., Wang, G., Wang, A., 2013. Genomic segments RNA1 and RNA2 of *Prunus necrotic ringspot virus* codetermine viral pathogenicity to adapt to alternating natural *Prunus* hosts. *Molecular Plant-Microbe Interactions* 26, 515–527.
- Cui, H., Wang, A., 2017. An efficient viral vector for functional genomic studies of prunus fruit trees and its induced resistance to plum pox virus via silencing of a host factor gene. *Plant Biotechnology Journal* 15 (3), 344–356.
- Koziet, E., Bujarski, J., Otulak, K., 2017. Molecular biology of Prune dwarf virus – A lesser known member of the family *Bromoviridae* but a vital component in the dynamic virus–host cell interaction network. *International Journal of Molecular Sciences* 18 (12), 2733.
- Moreno, A., Fereres, A., 2012. Virus diseases in lettuce in the Mediterranean basin. II *Alfalfa mosaic virus* (*Bromoviridae*, *Alfavirus*). *Advances in Virus Research* 8, 250–253.
- Pallas, V., Aparicio, F., Herranz, M.C., *et al.*, 2012. Iilarviruses of *Prunus* spp.: A continued concern for fruit trees. *Phytopathology* 102, 1108–1120.
- Pallas, V., Aparicio, F., Herranz, M.C., Sanchez-Navarro, J.A., Scott, S.W., 2013. The molecular biology of ilarviruses. *Advances in Virus Research* 87, 139–181.
- Rubio, M., Martínez-Gómez, P., Marais, A., *et al.*, 2017. Recent advances and prospects in *Prunus* virology. *Annals of Applied Biology* 171, 125–138.
- Wang, A., 2015. Dissecting the molecular network of virus-plant interactions: The complex roles of host factors. *Annual Review of Phytopathology* 53, 45–66.
- Wang, A., 2018. Virus and host plant interactions. In: eLS. Chichester: John Wiley & Sons Ltd. Available at: <https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0000758.pub3>.

Luteoviruses (*Luteoviridae*)

Leslie L. Domier, Agricultural Research Service, US Department of Agriculture, Urbana, IL, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of L.L. Domier, C.J. D'Arcy, Luteoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00438-6](https://doi.org/10.1016/B978-012374410-4.00438-6).

Nomenclature

aa	Amino acid(s)	PTGS	Post-transcriptional gene silencing
CP	Coat protein or capsid protein	RdRp	RNA-dependent RNA polymerase
IRES	Internal ribosomal entry site	RTD	Readthrough domain
kb	Kilobases; the size of a ssDNA or ssRNA molecule	RT-PCR	Reverse transcription polymerase chain reaction
kDa	Kilodaltons; the size of a protein	S _{20,w}	Corrected sedimentation coefficient; extrapolated to that in water at 20°C and infinitely diluted
nm	Nanometer(s)	sgRNA	Sub-genomic RNA
nt	Nucleotide(s)	ssRNA	Single-stranded ribonucleic acid
ORF	Open reading frame	Vpg	Genome-linked protein
Poly(A)	Polyadenylated		

Glossary

Hemocoel The primary body cavity of most arthropods that contains most of the major organs and through which the hemolymph circulates.

Hemolymph A circulatory fluid in the body cavities (hemocoels) and tissues of arthropods that is analogous to blood and/or lymph of vertebrates.

Pseudoknot A folded RNA that contains two or more stem-and-loop structures where the loop of one stem-and-loop structure interacts with the stem of a second stem-and-loop structure.

Introduction

Viruses of the family *Luteoviridae* (luteovirids) cause economically important diseases in many monocotyledonous and dicotyledonous crop plants, including cereal grasses, cucurbits, legumes, lettuce, potatoes, sugar beets, and woody perennials. Yield reductions as high as 30% have been reported in epidemic years, although in some cases crops can be totally destroyed. Diseases caused by the viruses were recorded decades and even centuries before they were associated with the causal viruses. In many cases, the stunted, deformed, and discolored plants that result from luteovirid infections were thought to be the result of abiotic factors, such as mineral imbalances or stressful environmental conditions, or other biotic agents. This, along with their inability to be transmitted mechanically, delayed the initial association of the symptoms with plant viruses. For example, curling of potato leaves was first described in Lancashire, UK, in the 1760s, but was not recognized as a specific disease of potato until 1905 and to be caused by an aphid-transmitted virus until the 1920s. The causal agent, Potato leaf roll virus (PLRV), was not purified until the 1960s. Similarly, widespread disease outbreaks in cereals, probably caused by Barley yellow dwarf virus (BYDV), were noted in the United States in 1907 and 1949. In 1951, a virus was proposed as the causal agent. Other diseases caused by luteovirids, like yellowing of Suakwa vegetable sponge, which is caused by Suakwa aphid-borne yellows virus (SABYV), were not described until recently.

Taxonomy and Classification

Members of the family *Luteoviridae* were first grouped because of their common biological properties. These properties included persistent transmission by aphid vectors and the induction of yellowing symptoms in many infected host plants. 'Luteo' comes from the Latin luteus, which translates as yellowish. All luteovirids have small (c. 25 nm diameter) icosahedral particles, composed of one major and one minor protein component and a single molecule of positive-sense single-stranded RNA of approximately 5600 nt in length.

The family *Luteoviridae* is divided into three genera – *Luteovirus*, *Polerovirus* (derived from potato leaf roll), and *Enamovirus* (derived from pea enation mosaic) – based on the arrangements, sizes, and phylogenetic relationships of the predicted aa sequences of the open reading frames (ORFs). In some plant virus families, a single gene can be used to infer taxonomic and

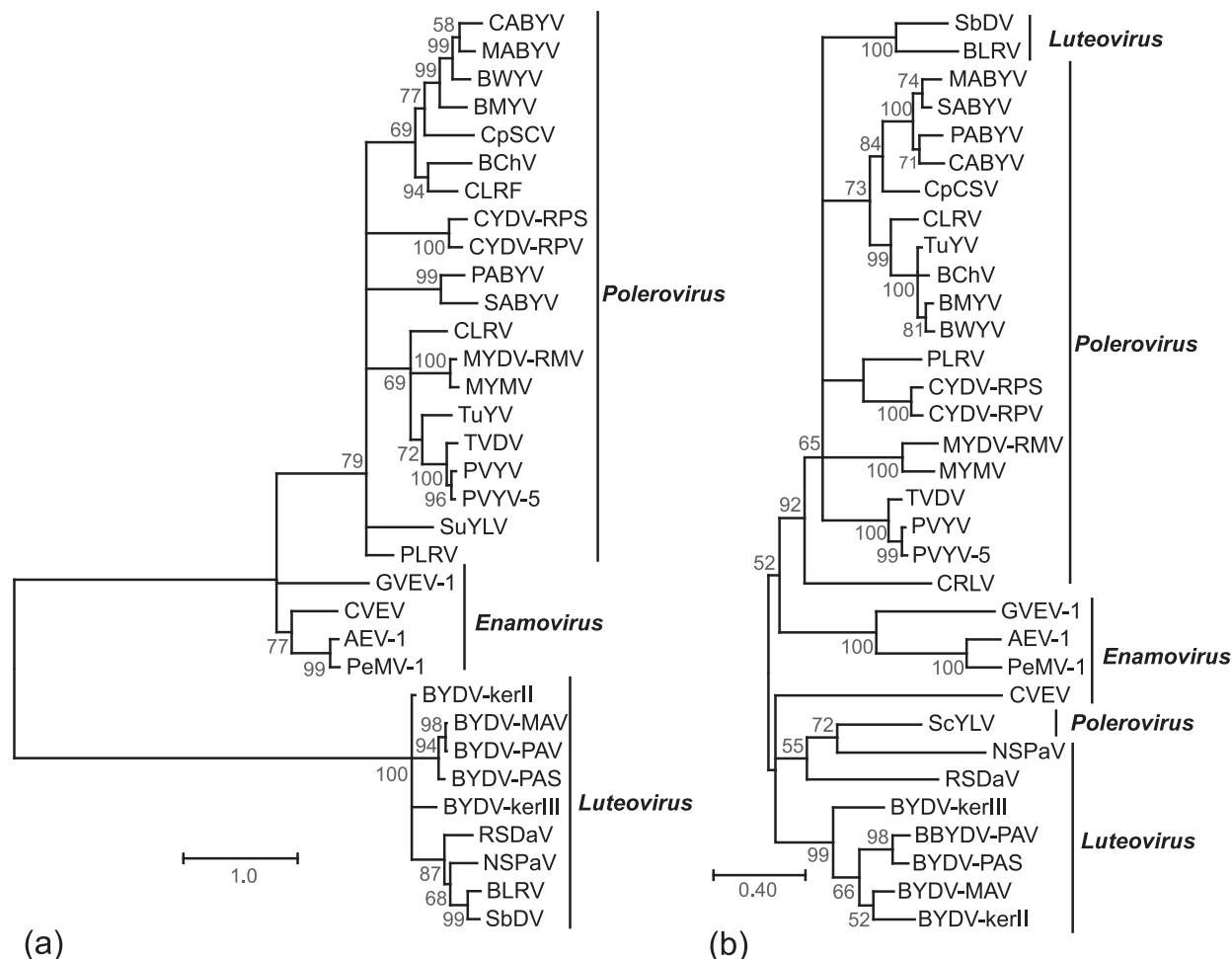


Fig. 1 Phylogenetic relationships of the predicted amino acid sequences of the (a) RNA-dependent RNA polymerase (ORF2) and (b) major capsid protein (ORF3). When predicted amino acid sequences from ORF2 are used to group virus species the genera form three distinct groups. Using predicted amino acid sequences from ORF3, species of the genera *Luteovirus* and *Polerovirus* are intermingled in the tree. The resulting consensus trees from 500 bootstrap replications are shown. The numbers above each node indicate the percentage of bootstrap replicates in which that node was recovered. For virus abbreviations and GenBank accession numbers, see [Table 1](#).

phylogenetic relationships. Within the family *Luteoviridae*, however, different taxonomic relationships can be predicted depending on whether sequences of the replicase (ORF2) or coat protein (CP; ORF3) genes are analyzed ([Fig. 1](#)). ORFs 1 and 2 of the luteoviruses are most closely related to the polymerase genes of viruses of the family *Tombusviridae*, while ORFs 1 and 2 of the poleroviruses and enamoviruses are related to those of the family *Solemoviridae*. These polymerase types are distantly related in evolutionary terms. Consequently, it has been suggested that the family *Luteoviridae* be abolished and the *Luteovirus* genus be moved to the family *Tombusviridae* and the *Enamovirus* and *Polerovirus* genera be moved to the family *Solemoviridae*.

It is hypothesized that luteovirid genomic RNAs arose by recombination between ancestral genomes containing the CP genes characteristic of the family *Luteoviridae* and genomes containing either of the two polymerase types. For taxonomic purposes, the polymerase type has been the primary determinant in assigning a virus to a genus. For this reason, viruses for which only CP sequences have been determined have not been assigned to a genus. The current members of the family are listed in [Table 1](#). The genus *Luteovirus* contains nine species, and the *Polerovirus* genus has 19 species. The genus *Enamovirus* contains four species. In addition, with recent advances in virus detection by next-generation sequencing, over 60 partial or complete genome sequences have been deposited in GeneBank that potentially represent new virus species in the family.

Virion Properties

The sedimentation coefficients $S_{20,w}$ (in Svedberg units) for luteoviruses and poleroviruses range from 106 S to 127 S. Buoyant densities in CsCl are approximately 1.40 g cm^{-3} . The particles formed as result of the mixed infections by PEMV-1

Table 1 Virus members in the family *Luteoviridae*

Genus	Species	Abbreviation	Accession number ^a
<i>Luteovirus</i>	<i>Barley yellow dwarf virus KerII</i>	BYDV-KerII	NC_021481.1
	<i>Barley yellow dwarf virus KerIII</i>	BYDV-KerIII	KC559092.1
	<i>Barley yellow dwarf virus – MAV</i>	BYDV-MAV	NC_003680.1
	<i>Barley yellow dwarf virus – PAS</i>	BYDV-PAS	NC_002160.2
	<i>Barley yellow dwarf virus – PAV</i>	BYDV-PAV	NC_004750.1
	<i>Bean leafroll virus</i>	BLRV	NC_003369.1
	<i>Nectarine stem pitting-associated virus</i>	NSPaV	NC_027211.1
	<i>Rose spring dwarf-associated virus</i>	RSDaV	NC_010806.1
	<i>Soybean dwarf virus</i>	SbDV	NC_003056.1
<i>Polerovirus</i>	<i>Beet chlorosis virus</i>	BChV	NC_002766.1
	<i>Beet mild yellowing virus</i>	BMVY	NC_003491.1
	<i>Beet western yellows virus</i>	BWYV	NC_004756.1
	<i>Carrot red leaf virus</i>	CRLV	NC_006265.1
	<i>Cereal yellow dwarf virus – RPS</i>	CYDV-RPS	NC_002198.2
	<i>Cereal yellow dwarf virus – RPV</i>	CYDV-RPV	NC_004751.1
	<i>Chickpea chlorotic stunt virus</i>	CpCSV	NC_008249.1
	<i>Cotton leafroll dwarf virus</i>	CLRV	NC_014545.1
	<i>Cucurbit aphid-borne yellows virus</i>	CABYV	NC_003688.1
	<i>Maize yellow dwarf virus RMV</i>	MYDV-RMV	NC_021484.1
	<i>Maize yellow mosaic virus</i>	MYMV	KU248489.1
	<i>Melon aphid-borne yellows virus</i>	MABYV	NC_010809.1
	<i>Pepo aphid-borne yellows virus</i>	PABYV	NC_030225.1
	<i>Pepper vein yellows virus</i>	PVYV	NC_015050.1
	<i>Pepper vein yellows virus 5</i>	PVYV-5	NC_036803.1
	<i>Potato leafroll virus</i>	PLRV	NC_001747.1
	<i>Suakwa aphid-borne yellows virus</i>	SABYV	NC_018571.2
	<i>Sugarcane yellow leaf virus</i>	ScYLV	NC_000874.1
	<i>Tobacco vein distorting virus</i>	TVDV	NC_010732.1
	<i>Turnip yellows virus</i>	TuYV	NC_003743.1
<i>Enamovirus</i>	<i>Alfalfa enamovirus 1</i>	AEV-1	NC_029993.1
	<i>Citrus vein enation virus</i>	CVEV	NC_021564.1
	<i>Grapevine enamovirus 1</i>	GVEV-1	NC_034836.1
	<i>Pea enation mosaic virus 1</i>	PEMV-1	NC_003629.1
Unassigned	<i>Barley yellow dwarf virus – GPV</i>	BYDV-GPV	NC_039035.1
	<i>Barley yellow dwarf virus – SGV</i>	BYDV-SGV	AY541039.1
	<i>Chickpea stunt disease associated virus</i>	CpSDaV	Y11530.1
	<i>Groundnut rosette assistor virus</i>	GRAV	NC_038509.1
	<i>Indonesian soybean dwarf virus</i>	ISDV	
	<i>Sweet potato leaf speckling virus</i>	SPLSV	NC_038510.1
	<i>Tobacco necrotic dwarf virus</i>	TNDV	

^aAccession numbers beginning with NC_ represent reference genomic sequences.

and PEMV-2 sediment as two components. The $S_{20,w}$ are 107–122 S for B components (PEMV-1) and 91–106 S for T components (PEMV-2, an umbravirus). Virions are moderately stable and are insensitive to treatment with chloroform or nonionic detergents but are disrupted by prolonged treatment with high concentrations of salts. Luteovirus and polerovirus particles are insensitive to freezing.

Virion Structure and Composition

All members of the *Luteoviridae* have nonenveloped icosahedral particles with diameters of 25–28 nm (Fig. 2). Capsids are composed of major (21–23 kDa) and minor (54–76 kDa) CPs, which contain a C-terminal extension to the major CP called the readthrough domain (RTD). According to X-ray diffraction and molecular mass analysis, virions consist of 180 protein subunits, arranged in $T=3$ icosahedra. Virus particles do not contain lipids or carbohydrates. Using an X-ray crystallography-derived structure of virions of the Rice yellow mottle sobemovirus, which shares a CP aa sequence similarity of 33% with PLRV, it was possible to predict the virion structure of PLRV and other luteovirids.

Virions contain a single molecule of single-stranded positive-sense RNA of 5300–5900 nt. The RNAs do not have a 3' terminal poly(A) tract. A small protein (VPg) is covalently linked to the 5' end of polerovirus and enamovirus genomic RNAs. Cereal yellow dwarf virus - RPV (CYDV-RPV) also encapsidates a 322 nt satellite RNA that accumulates to high levels in the presence of the

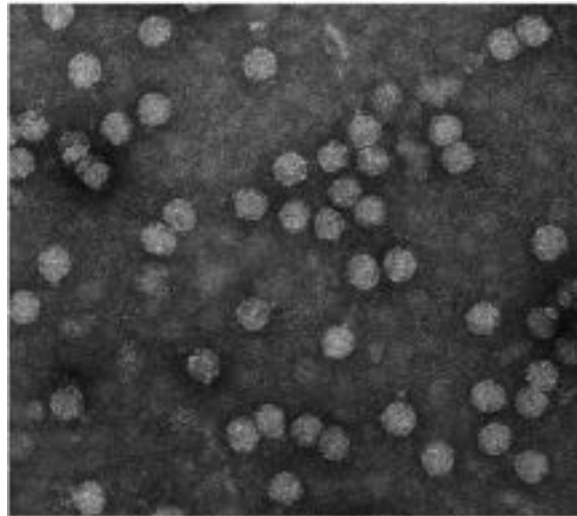


Fig. 2 Transmission electron micrograph of Soybean dwarf virus particles magnified $240,000\times$. Virions (stained with uranyl acetate) are c. 25 nm in diameter, hexagonal in appearance, and have no envelope.

helper virus. Complete genome sequences have been determined for more than 30 species of the family *Luteoviridae* (Table 1). For several viruses, genome sequences have been determined from multiple isolates.

Genome Organization and Expression

Genomic RNAs of luteovirids contain five to seven conserved ORFs (Fig. 3). ORFs 1, 2, 3, and 5 are shared among all members of the *Luteoviridae*. Luteoviruses lack ORF0. Enamoviruses lack ORF4. Luteo- and polerovirus genomes contain a small ORF (ORF3a) upstream of ORF3. Luteovirus genomes contain a small ORF (ORF6) downstream of ORF5. The PLRV genome contains ORFs 6 and 7, within ORF5 and ORF8 within ORF1. In the enamo- and poleroviruses ORF0 overlaps ORF1 by more than 600 nt, which also overlaps ORF2 by more than 600 nt. In the luteoviruses, ORF1 overlaps ORF2 by less than 50 nt. In most luteo- and polerovirus genome sequences, ORF4 is contained completely within ORF3. A single, in-frame amber (UAG) termination codon separates ORF5 from ORF3.

Luteovirids have relatively short 5' and intergenic non-coding sequences. The first ORF is preceded by 21 nt in CABYV RNA and 142 nt in Soybean dwarf virus (SbDV) RNA. ORFs 2 and 3 are separated by 112–200 nt of non-coding RNA. There is considerable variation in the length of sequence downstream of ORF5, which ranges from 167 nt for CYDV-RPV to 650 nt for SbDV.

Luteovirids employ a wide range of strategies to express their compact genomes. ORFs 0, 1, 2, and 8 are expressed directly from genomic RNAs. Downstream ORFs are expressed from sub-genomic RNAs (sgRNAs) that are transcribed from internal initiation sites by virus-encoded RNA-dependent RNA polymerases (RdRps) from negative-strand RNAs and are 3' co-terminal with the genomic RNA. Since the initiation codon for ORF0 of polero- and enamoviruses is upstream of that of ORF1, translation of ORF1 is initiated by 'leaky scanning' in which ribosomes bypass the AUG of ORF0 and continue to scan the genomic RNA until they reach the ORF1 AUG. The protein products of ORF2 are expressed as a translational fusion with the product of ORF1. At a low but significant frequency during the expression of ORF1, translation continues into ORF2 through a –1 frameshift that produces a large protein containing sequences encoded by both ORFs 1 and 2 in a single polypeptide. The frameshift is mediated by a "slippery hepta-nucleotide sequence" (in the form X XXY YYZ) and a downstream RNA secondary structure termed a pseudoknot that causes ribosomes to pause and then shift back one nt before continuing translation in the new reading frame. ORF8, which has only been identified in PLRV, resides entirely within ORF1 in a different reading frame and encodes a 5 kDa replication-associated protein. To express ORF8, sequences within the ORF fold into a structure called an internal ribosome entry site (IRES), which recruits ribosomes to initiate translation about 1600 nt downstream of the 5' terminus of PLRV RNA.

ORFs 3a, 3, 4, and 5 are expressed through a leaky scanning mechanism from the 5' terminus of sgRNA1, which is located about 200 nt upstream of ORF3 at the end of ORF2 and extends to the 3' terminus of the genome. Translation of ORF3a is initiated at a non-AUG codon. ORF4 of most luteo- and poleroviruses is contained within ORF3. In all luteovirids, ORF5 is expressed only as a translational fusion with the products of ORF3 by readthrough of the UAG stop codon at the end of ORF3, which produces a protein with the product of ORF3 at its N-terminus and the product of ORF5 at its C-terminus. Readthrough is regulated by local and long-distance RNA-RNA interactions and in the case of luteoviruses and some poleroviruses requires the presence of CCXXX repeats (where X is any base) downstream of the ORF3 stop codon. Luteo- and poleroviruses produce second smaller sgRNA capable of expressing ORFs 6 and 7. Third sgRNAs, which do not appear to encode proteins, are produced at very high levels in luteoviruses, but at only low levels in PLRV.

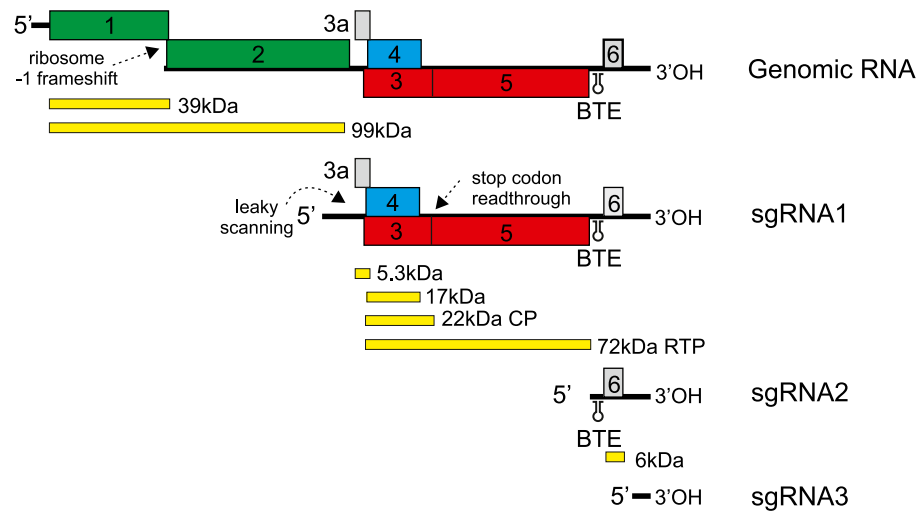
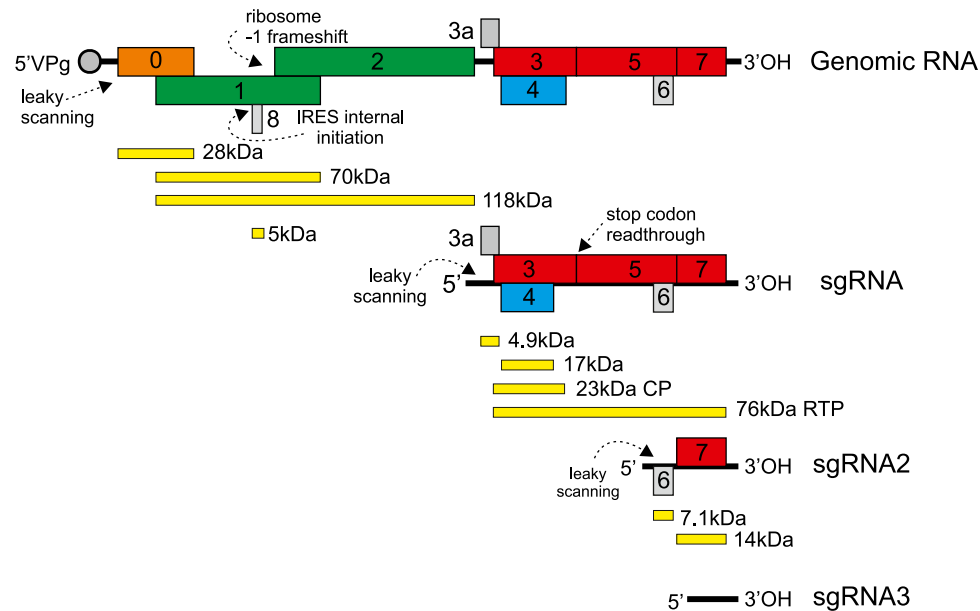
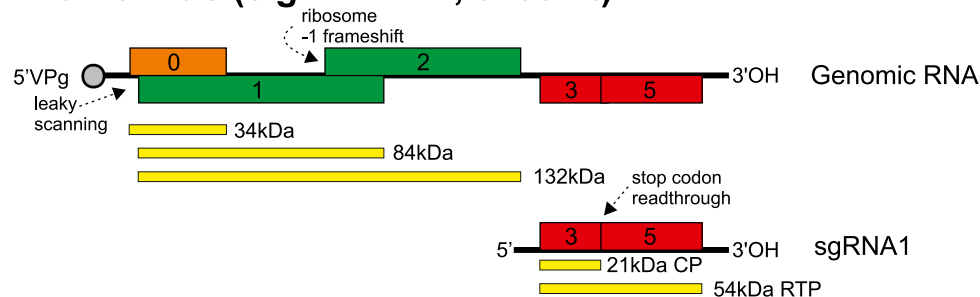
Luteovirus (e.g. BYDV-PAV, 5677 nt)**Polerovirus (e.g. PLRV, 5882 nt)****Enamovirus (e.g. PEMV-1, 5705 nt)**

Fig. 3 Maps of the virus genomes of genera in the family *Luteoviridae*. Individual ORFs are shown with open boxes. The ORFs are staggered vertically to show the different reading frames occupied by each ORF. The yellow boxes indicate protein products with the predicted sizes listed to the right of each. The polyproteins encoded by ORF1 of enamo- and poleroviruses contain the protease and the genome-linked protein (VPg). The predicted amino acid sequences of proteins encoded by ORF2 are similar to RNA-dependent RNA polymerases. ORF3, which encodes the major coat protein, is separated from ORF5 by an amber termination codon. ORF4, when present, is most often contained within ORF3 and encodes a protein required for virus cell-to-cell movement. Luteo- and poleroviruses contain an ORF3a for which translation is initiated at a non-AUG codon. The 3' non-coding regions of luteoviruses contain translation enhancer elements (BTE). In PLRV, ORF7 is in-frame with the C-terminus of ORF5, and translation of ORF8 is mediated by an internal initiation ribosome entry site (IRES). Luteo- and poleroviruses produce three sub-genomic RNAs (sgRNAs), but enamoviruses produce a single sgRNA.

While enomo- and polerovirus RNAs contain 5' VPgs that interact with translation initiation factors, luteovirus RNAs contain only a 5' phosphate. Unmodified 5' termini are recognized poorly for translation initiation. To circumvent this problem, the BYDV-PAV genome contains a short sequence (BYDV translation element; BTE) located in the 3' non-coding region downstream of ORF5 that interacts with sequences near the 5' termini of genomic RNA and sgRNA1 to promote cap-independent translation initiation.

Post-transcriptional gene silencing (PTGS) is an innate and highly adaptive antiviral defense found in all eukaryotes that is activated by double-stranded RNAs (dsRNAs), which are produced during virus replication. Research into the functions of the proteins encoded by luteovirids has shown that the 28–34 kDa proteins encoded by ORF0 are strong suppressors of local and systemic PTGS for polero- and enamoviruses. Luteovirus genomes lack ORF0, but the product of ORF4 in luteoviruses functions to suppress systemic PTGS.

The ORF1-encoded proteins of enomo- and poleroviruses contain the VPg and a chymotrypsin-like serine protease that is responsible for the proteolytic processing of ORF1-encoded polyproteins. The protease cleaves the ORF1 protein internally to liberate the VPg, which is covalently attached to genomic RNA. The protein expressed by ORF8 of PLRV is required for virus replication. Luteovirid ORF2s have a coding capacity of 59–67 kDa for proteins that are very similar to known RdRps and hence likely represent the catalytic portion of the viral replicase.

For luteo- and poleroviruses ORF3a produces highly conserved 4.8–5.3 kDa proteins that are essential for long-distance movement. ORF3 encodes the major CP of the luteovirids, which ranges in size from 21 to 23 kDa. ORF5 has a coding capacity of 29–56 kDa. However, ORF5 is expressed only as a translational fusion with the product of ORF3 when, about 10% of the time, translation does not stop at the end of ORF3 and continues through to the end of ORF5. The ORF5 portion of this readthrough protein has been implicated in aphid transmission and virus stability. Experiments with PLRV and BYDV-PAV have shown that the N-terminal region of the ORF5 readthrough protein determines the ability of virus particles to bind to proteins produced by endosymbiotic bacteria of aphid vectors. Interactions of virus particles with these proteins seem to be essential for persistence of the viruses in aphids. Nucleotide sequence changes within ORF5 of PEMV-1 abolish aphid transmissibility. The N-terminal portions of ORF5 proteins are highly conserved among luteovirids while the C-termini are much more variable.

The luteo- and polerovirus genomes possess an ORF4 that is contained within ORF3 and encodes proteins of 17–21 kDa. Viruses that contain mutations in ORF4 are able to replicate in isolated plant protoplasts, but are deficient or delayed in systemic movement in whole plants. Hence, the product of ORF4 seems to be required for movement of the virus within infected plants. This hypothesis is supported by the observation that enamoviruses lack ORF4. While luteo- and poleroviruses are limited to phloem and associated tissues, the enamovirus PEMV-1 is able to move systemically through other plant tissues in the presence of PEMV-2, which under natural conditions invariably coexists with PEMV-1.

Some luteo- and polerovirus genomes contain small ORFs within and/or downstream of ORF5. In luteoviruses, no protein products have been detected from these ORFs in infected cells. BYDV-PAV genomes that do not express ORF6 are still able to replicate in protoplasts. The predicted sizes of the proteins expressed by ORFs 6 and 7 of PLRV are 7.1 and 14 kDa, respectively. Based on mutational studies, it has been proposed that these genome regions may regulate transcription late in infection.

Evolutionary Relationships

As mentioned above, viruses in the family *Luteoviridae* have replication-related proteins that are similar to those in other plant virus families and genera. The luteovirus replication proteins encoded by ORFs 1 and 2 resemble those of members of the family *Tombusviridae*. In contrast, the serine protease-VPg-polymerase complex of polero- and enamoviruses resemble those of viruses in the family *Solemoviridae*. The structural proteins of some members of the *Sobemovirus* genus in the family *Solemoviridae* also are similar to the major CP of luteovirids.

Host Range and Transmission

Several luteovirids have natural host ranges largely restricted to one plant family. For example, BYDV and CYDV infect many grasses, BLRV infects mainly legumes, and CtRLV infects mainly plants in the family Apiaceae. Other luteovirids infect plants in several or many different families. For example, Beet western yellows virus (BWYV) infects more than 150 species of plants in more than 20 families. As techniques for infecting plants with recombinant viruses have improved, the experimental host ranges of luteovirids have been expanded to include plants on which aphid vectors would not normally feed. For example, BYDV, CYDV, PLRV, and SbDV have been shown to infect *Nicotiana* species when inoculated biolistically with viral RNA or using *Agrobacterium tumefaciens* harboring binary plasmids containing infectious copies of the virus genomes even though those species had not been described previously as experimental hosts for the viruses. These results suggest that feeding preferences of vector aphids play important roles in defining luteovirid host ranges.

Luteovirids are transmitted in a circulative manner with varying efficiencies by at least 25 aphid species. With the exception of the enamovirus PEMV-1, members of the family *Luteoviridae* are transmitted from infected plants to healthy plants in nature only by the feeding activities of specific species of aphids. There is no evidence for replication of the viruses within aphid vectors. *Myzus*

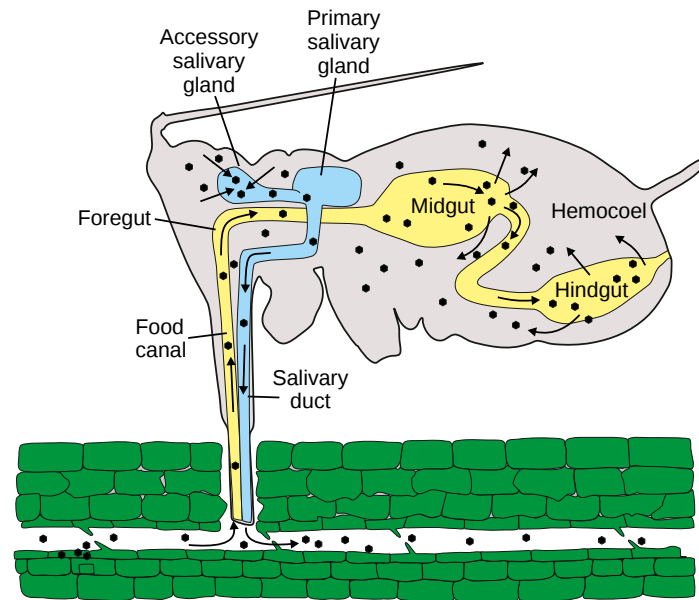


Fig. 4 Circulative transmission of viruses of the family *Luteoviridae* by vector aphids. While feeding from sieve tubes of an infected plant, an aphid (shown in cross section) acquires virus particles, which travel up the stylet, through the food canal, and into the foregut. The virions are actively transported across cells of the posterior midgut and/or hindgut into the hemocoel in a process that involves receptor-mediated endocytosis. Virions then passively migrate through the hemolymph to the accessory salivary gland where they are again transported by a receptor-mediated process to reach the lumen of the gland. Once in the salivary gland lumen, the virions are expelled with the saliva into the vascular tissue of host plants. Aphids can retain the ability to transmit virus for several weeks. Hindgut membranes usually are much less selective than those of the accessory salivary glands, which is why viruses that are not transmitted by a particular species of aphid often accumulate in the hemocoel, but do not traverse the membranes of the accessory salivary gland.

persicae is the most common aphid vector of luteovirids that infect dicots. Several different species of aphids transmit luteovirids that infect monocotyledonous plants (BYDV and CYDV) in a species-specific manner.

Circulative transmission of the viruses is initiated when aphids acquire viruses from sieve tubes of infected plants during feeding. The viruses travel up the stylet, through the food canal, and into the foregut (Fig. 4). The viruses then are actively transported across the cells of the alimentary tract into the hemocoel in a process that involves receptor-mediated endocytosis of the viruses and the formation of tubular vesicles that transport the viruses through epithelial cells and into the hemocoel. Luteovirids are acquired at different sites within the gut of vector aphids. PLRV and BWYV are acquired in the posterior midgut. BYDV, CYDV, and SbDV are acquired in the hindgut. CABYV is taken up at both sites. Viruses then passively migrate through the hemolymph to the accessory salivary gland where the viruses must pass through the membranes of the accessory salivary gland cells in a similar type of receptor-mediated transport process to reach the lumen of the gland. Once in the salivary gland lumen, viruses are expelled with saliva into vascular tissues of host plants. Since large amounts of virus can accumulate in the hemocoel of aphids, they may retain the ability to transmit virus for several weeks. Typically, hindgut membranes are much less selective than those of the accessory salivary glands. Consequently, viruses that are not transmitted by a particular species of aphid often are transported across gut membranes and accumulate in the hemocoel, but do not traverse the membranes of the accessory salivary gland.

The RTD of the minor capsid protein plays a major role in aphid transmission of luteovirids. The RTD interacts with symbionin produced by endosymbiotic aphid-borne bacteria, which may protect virions from degradation by the aphid immune system. The specificity of aphid transmission and gut tropism has been linked to the RTD in multiple luteovirids.

Unlike other luteovirids, PEMV-1 can be transmitted by rubbing sap taken from an infected plant onto a healthy plant, in addition to being transmitted by aphids. This difference in transmissibility is dependent on its multiplication in cells co-infected with PEMV-2, but aphid transmissibility can be lost after several mechanical passages.

Replication

Luteovirids infect and replicate in sieve elements and companion cells of the phloem and occasionally are found in phloem parenchyma cells. PEMV-1 can move systemically into other tissues in the presence of PEMV-2. Virus infections commonly result in cytopathological changes in cells that include formation of vesicles containing filaments and inclusions that contain viral RNA and virions. The subcellular location of viral RNA replication has not been determined unequivocally. However, early in infection, negative-strand RNAs of BYDV-PAV are first detected in the nucleus and later in the cytoplasm, which suggests that at least a

portion of luteovirus replication occurs in the nucleus. Synthesis of negative-strand RNA, which requires tetraloop structures at the 3' end of BYDV-PAV genomic RNAs, is detected in infected cells before the formation of virus particles. Late in infection, BYDV-PAV sgRNA2 inhibits translation from genomic RNA, which may promote a switch from translation to replication and packaging of genomic RNAs.

Virus–Host Relationships

While some infected plants display no obvious symptoms, most luteovirids induce characteristic symptoms that include stunting, leaves that become thickened, curled or brittle, and yellow, orange, or red leaf discoloration, particularly of older leaves of infected plants. These symptoms result from phloem necrosis that spreads from inoculated sieve elements and causes symptoms by inhibiting translocation, slowing plant growth, and inducing the loss of chlorophyll. Symptoms may persist, may vary seasonally, or may disappear soon after infection. Temperature and light intensity often affect symptom severity and development. In addition, symptoms can vary greatly with different virus isolates or strains and with different host cultivars.

Yield losses caused by luteovirids are difficult to estimate because the symptoms often are overlooked or attributed to other agents. US Department of Agriculture specialists estimated that yield losses from BWYV, BYDV, and PLRV infections were over \$65 million during the period 1951–60. Plants infected at early stages of development by luteovirids suffer the most significant yield losses, which often are linearly correlated with the incidence of virus infection.

Epidemiology

Luteovirid infections have been reported from temperate, subtropical, and tropical regions of the world. Some of the viruses are found worldwide, such as BWYV, BYDV, and PLRV. Others have more restricted distribution, such as tobacco necrotic dwarf virus, which has been reported only from Japan, and groundnut rosette assistor virus, which has been reported only in African countries south of the Sahara.

Luteovirids infect annual and perennial, cultivated and wild plants. In annual crops, the viruses must be reintroduced each year by their aphid vectors, sometimes from native perennial hosts. Some viruses are disseminated in infected planting material. For example, infected potato tubers are the principal source of inoculum for new epidemics of PLRV. Consequently, programs to produce clean stock are operated around the world to control these viruses. Alate, that is, winged, aphid vectors may transmit viruses from local cultivated, volunteer, or weed hosts. Alternatively, alate aphids may be transported into crops from distant locations by wind currents. These vectors may bring the virus with them, or they may first have to acquire virus from locally infected hosts. The agronomic impact of luteovirid diseases depends both on meteorological events that favor movement and reproduction of vector aphids and susceptibility of the crop at the time of aphid arrival. Only aphid species that feed on a particular crop plant can transmit virus. Aphids that merely probe briefly to determine a plant's suitability will not transmit the viruses. Secondary spread of the viruses is often primarily by apterous, that is, wingless, aphids. The relative importance of primary introduction of virus by alate aphids and of secondary spread of virus by apterous aphids in disease severity varies with the virus, aphid species, crop, and environmental conditions.

Some members of the family *Luteoviridae* occur in complexes with other members of the family or with other plant viruses. For example, BYDV and CYDV often are found co-infecting cereals; BWYV and SbDV are often found together in legumes; and PLRV is often found co-infecting potatoes with Potato virus Y and/or Potato virus X. Some other plant viruses depend on luteovirids for their aphid transmission, such as the Groundnut rosette virus, Carrot mottle virus, and Bean yellow vein banding virus (all umbraviruses), which depend on Groundnut rosette assistor virus, Carrot red leaf virus, and PEMV-1, respectively.

Diagnosis

An integral part of controlling luteovirid diseases is accurate diagnosis of infection. Because symptoms caused by luteovirids often resemble those induced by other biotic and abiotic factors, visual diagnosis is unreliable and other methods have been developed. Initially, infectivity, or biological, assays were used to diagnose infections. These techniques also have been used to identify species of vector aphids and vector preferences. In bioassays, aphids are allowed to feed on infected plants and then are transferred to indicator plants. These techniques are very sensitive but can require several weeks for symptoms to develop on indicator plants. The strong immunogenicity of luteovirids has facilitated development of very specific and highly sensitive serological tests that can discriminate different luteovirids and sometimes even strains of a single virus species. Poly- and monoclonal antibodies for virus detection are produced by immunizing rabbits and/or mice with virus particles purified from infected plants. Techniques also have been developed to detect viral RNAs from infected plant tissues by reverse transcription-polymerase chain reaction (RT-PCR), which can be more sensitive and discriminatory than serological diagnostic techniques. High-throughput sequencing of plant transcriptomes has become a powerful tool for identification of new luteovirids and characterization of virus complexes infecting plants. Even so, serological tests are the most commonly used techniques for the detection of infections because of their simplicity and speed.

Control

Because methods are not available to cure luteovirid infections after diagnosis, emphasis has been placed on reducing losses through the use of tolerant or resistant plant cultivars and/or on reducing the spread of viruses by controlling aphid populations. Many luteovirids are transmitted by migrating populations of aphids that occur at similar times each year. For those virus–aphid combinations, it is sometimes possible to plant crops so that young, highly susceptible plants are not in the field when the seasonal aphid migrations occur. Insecticides have been used in a prophylactic manner to reduce crop losses. While insecticide treatments do not prevent initial infections, they can greatly limit secondary spread of aphids and therefore of viruses. In some instances biological control agents such as predatory insects and parasites have reduced aphid populations significantly. Genes for resistance or tolerance to infection by luteovirids have been identified in most agronomically important plant species infected by the viruses. For BYDV, PLRV, and SbDV, transgenic plants that express portions of the virus genomes have been produced through DNA-mediated transformation. In some cases, the expression of these virus genes in transgenic plants confers higher levels of virus resistance than naturally occurring plant resistance genes.

Further Reading

- Brault, V., Perigon, S., Reinbold, C., *et al.*, 2005. The polerovirus minor capsid protein determines vector specificity and intestinal tropism in the aphid. *Journal of Virology* 79, 9685–9693.
- Fusaro, A.F., Barton, D.A., Nakasugi, K., *et al.*, 2017. The *Luteovirus* P4 movement protein is a suppressor of systemic RNA silencing. *Viruses* 9, 16.
- Gray, S., Cilia, M., Ghanim, M., 2014. Circulative, “nonpropagative” virus transmission: An orchestra of virus-, insect-, and plant-derived instruments. In: Maramorosch, K., Murphy, F.A. (Eds.), *Advances in Virus Research*, vol. 89. San Diego: Elsevier Academic Press Inc., pp. 141–199.
- Hipper, C., Monsion, B., Bortolamiol-Becet, D., Ziegler-Graff, V., Brault, V., 2014. Formation of virions is strictly required for Turnip yellows virus long-distance movement in plants. *Journal of General Virology* 95, 496–505.
- Hogenhout, S.A., van der Wilk, F., Verbeek, M., Goldbach, R.W., van den Heuvel, J.F., 2000. Identifying the determinants in the equatorial domain of Buchnera GroEL implicated in binding potato leafroll virus. *Journal of Virology* 74, 4541–4548.
- Hulo, C., de Castro, E., Masson, P., *et al.* 2019. *Luteoviridae*. ViralZone. EXPASY Bioinformatics Resource Portal. Switzerland. Available at: <https://viralzone.expasy.org/45> (accessed 26.05.2019).
- Lee, L., Palukaitis, P., Gray, S.M., 2002. Host-dependent requirement for the Potato leafroll virus 17-kDa protein in virus movement. *Molecular Plant–Microbe Interactions* 10, 1086–1094.
- Miller, W.A., Jackson, J., Feng, Y., 2015. *Cis*- and *trans*-regulation of luteovirus gene expression by the 3′ end of the viral genome. *Virus Research* 206, 37–45.
- Miller, W.A., Liu, S.J., Beckett, R., 2002. Barley yellow dwarf virus: *Luteoviridae* or *Tombusviridae*? *Molecular Plant Pathology* 3, 177–183.
- Moonan, F., Molina, J., Mirkov, T.E., 2000. Sugarcane yellow leaf virus: An emerging virus that has evolved by recombination between luteoviral and poleroviral ancestors. *Virology* 269, 156–171.
- NCBI, 2019. Complete Genomes: *Luteoviridae*. Bethesda, MD: National Center for Biotechnology Information, Available at: <https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=119163> (accessed 26.05.2019).
- Pfeffer, S., Dunoyer, P., Heim, F., *et al.*, 2002. P0 of Beet western yellows virus is a suppressor of posttranscriptional gene silencing. *Journal of Virology* 76, 6815–6824.
- Robert, Y., Woodford, J.A., Ducray-Bourdin, D.G., 2000. Some epidemiological approaches to the control of aphid-borne virus diseases in seed potato crops in Northern Europe. *Virus Research* 71, 33–47.
- Siddell, S.G., Walker, P.J., Lefkowitz, E.J., *et al.*, 2019. Virus Taxonomy: 2018b Release. International Committee on Taxonomy of Viruses. Washington, DC, July 2018. Available at: <https://talk.ictvonline.org/taxonomy> (accessed 26.05.2019).
- Smirnova, E., Firth, A.E., Miller, W.A., *et al.*, 2015. Discovery of a small non-AUG-initiated ORF in poleroviruses and luteoviruses that is required for long-distance movement. *PLOS Pathogens* 11, e1004868.
- Taliansky, M., Mayo, M.A., Barker, H., 2003. Potato leafroll virus: A classic pathogen shows some new tricks. *Molecular Plant Pathology* 4, 81–89.
- Terradot, L., Souchet, M., Tran, V., Giblot Ducray-Bourdin, D., 2001. Analysis of a three-dimensional structure of Potato leafroll virus coat protein obtained by homology modeling. *Virology* 286, 72–82.
- Thomas, P.E., Lawson, E.C., Zalewski, J.C., Reed, G.L., Kaniewski, W.K., 2002. Extreme resistance to Potato leafroll virus in potato cv. Russet Burbank mediated by the viral replicase gene. *Virus Research* 71, 49–62.
- Xu, Y., Ju, H.J., DeBlasio, S., Carino, E.J., *et al.*, 2018. A stem-loop structure in Potato leafroll virus open reading frame 5 (ORF5) is essential for readthrough translation of the coat protein ORF stop codon 700 bases upstream. *Journal of Virology* 92. (e01544-17).

Machlomovirus and Panicoviruses (*Tombusviridae*)

Kay Scheets, Oklahoma State University, Stillwater, OK, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)

CITE Cap-independent translation enhancer

CP Coat protein or capsid protein

kb Kilobase

kDa Kilo Dalton

LDRI Long distance RNA–RNA interaction

MP Movement protein

nt Nucleotide(s)

ORF Open reading frame

Poly(A) Polyadenylated

RAP Replicase-associated protein

RdRp RNA-dependent RNA polymerase

satRNA satellite RNA

SDS-PAGE Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

UTR Untranslated region

vRNA virion RNA

xrRNA exoribonuclease-resistant RNA

Glossary

3'CITE An RNA tertiary structure located in the 3'UTR which binds to translation initiation factor(s) and to a structure nearer the 5' end of the RNA which allows assembly with a ribosome and initiation factors for translation.

Cryo-electron microscopy Technique for rapidly cooling dilute aqueous biomolecules to determine structures at the molecular level.

Movement proteins Viral proteins that assist in the movement of vRNA or virions through plasmodesmata, the tube-like structures connecting one plant cell to another.

Phloem Vascular tissue that transports small molecules throughout the plant.

Plasmodesmata Microscopic membrane channels through plant cell walls that allow small molecules to readily move between plant cells. Plasmodesmata are internally coated by the endoplasmic reticulum.

Poaceae A family of monocotyledonous plants commonly called the grass family.

Potyvirus A virus in the family *Potyviridae* which contains the genera *Potyvirus* and *Tritimovirus*.

Protoplast A plant cell (either from an undifferentiated plant cell culture or from a plant tissue such as a leaf) that has been treated in an isotonic solution with a mixture of enzymes that degrades the cell wall, leaving membrane-bound structures similar to single animal cells.

Satellite RNA Non-coding ssRNAs dependent on a helper virus for replication.

Satellite virus ssRNA encoding only a CP that requires a helper virus to replicate.

Silencing suppressor - (suppressor of silencing) Virally encoded proteins from diverse virus families that interfere with host production or effectiveness of siRNAs.

Small interfering RNAs (or silencing RNAs), 20–25 bp RNAs with 2 nt 3' overhanging ends that target invading viral RNAs or cellular RNAs for degradation in eukaryotic cells.

Synergism/synergistic disease When disease symptoms of a double infection are more severe than the sum of each single infection.

Tombusvirid Member of the family *Tombusviridae*.

Classification

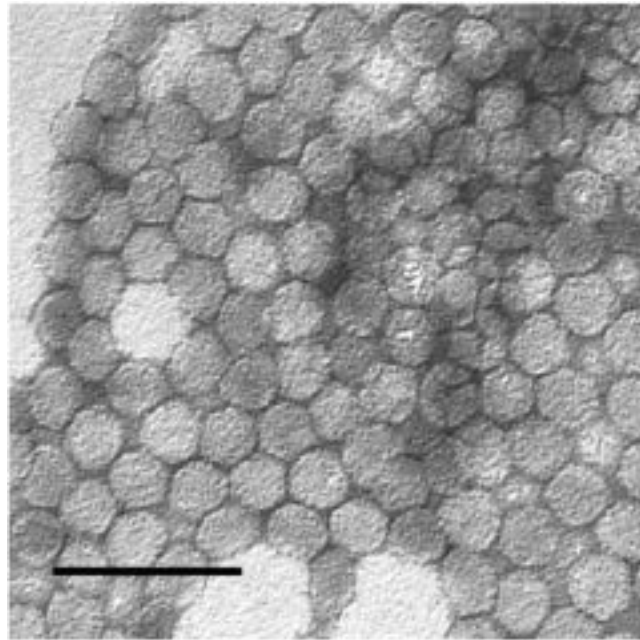
Realm *Riboviria*, family *Tombusviridae*, subfamily *Procedovirinae*, genera *Machlomovirus*, and *Panicovirus*. The genus *Machlomovirus* contains only one species, *Maize chlorotic mottle virus*, while the genus *Panicovirus* contains three species: *Cocksfoot mild mosaic virus*, *Panicum mosaic virus* (the type species) and *Thin paspalum asymptomatic virus* (Table 1).

Virion Structure

Like the majority of plant viruses, Maize chlorotic mottle virus (MCMV) and *Panicum mosaic virus* (PMV) are composed of only the viral genome within a protein shell made of multiple copies of a single coat protein (CP). Similar smooth virion structures have been determined for MCMV (cryo-electron microscopy) and PMV (X-ray diffraction). Both viruses are ~30 nm in diameter, exhibit icosahedral symmetry, and are composed of 180 subunits of a single 25 or 26 kDa CP that encapsidates one copy of their single ssRNA genomes (Fig. 1).

Table 1 List of members of the genera *Machlomovirus* and *Panicovirus* in the family *Tombusviridae*. Type species are in bold

Subfamily/Genus	Species	Acronym	Accession #	Length (nt)	# sgRNAs ^a
Procedovirinae					
Machlomovirus	Maize chlorotic mottle virus	MCMV	NC003627	4437	1
Panicovirus	<i>Cocksfoot mild mosaic virus</i>	CMMV	NC011108	4198	1
	Panicum mosaic virus	PMV	NC002598	4326	1
	<i>Thin paspalum asymptomatic virus</i>	TPAV	NC021705	4195	1

^aprotein-coding sgRNAs only.**Fig. 1** Purified maize chlorotic mottle virus negatively stained with uranyl acetate. The bar represents 100 nm.

Genomes

The genome organizations of the 16 genera in the family *Tombusviridae* fit one of four patterns. MCMV and panicoviruses have a carmovirus-like genome organization in which the cell-to-cell MP ORFs immediately follow the RdRp ORF, and the CP ORF is located at the 3' end of the genome. Like other members of the family *Tombusviridae*, the vRNA for both viruses is uncapped and does not contain a poly(A) tail. MCMV (4437 nt), PMV (4326 nt) and the other two recognized panicoviruses have larger genomes than other tombusvirids that form smooth virions (lacking a protruding domain [24.5–29.9 kDa CPs]), and are longer than some aureusviruses, which encode CPs with protruding domains (35.1–42.2 kDa). The size reflects the presence of larger genes for the pre-readthrough region (RAP) of their replicases, which encode carboxyl regions with high homology to other family members while the amino-terminal extensions are unique to each genus. MCMV encodes two overprinted genes that, 30 years after the first vRNA submission to GenBank, are still completely unique. And unlike all other tombusvirids, the RAP ORF start codon is preceded by the start codon of the unique 32 kDa ORF. P32 and p50, due to the p32 ORF overlap of the p50 ORF, are highly charged. P32 has a pI ~ 4 while p50 has a pI ~ 10 which causes an anomalous (opposite) migration pattern of the two proteins in SDS-PAGE. Initially, it was predicted that the second unique MCMV ORF encoding an extension of the upstream MP (p7a) was expressed via stop codon readthrough to make a 31 kDa protein. But the p7a ORF stop codon is not in a context that is likely to support readthrough, so it is more likely that the second unique gene is expressed from another non-canonical start codon to make a 23 kDa protein. PMV encodes an unrelated unique gene overprinted within its CP ORF that encodes a 15 kDa protein which is only found in viruses in the genus *Panicovirus* (Fig. 2).

Hosts and Distribution

MCMV was initially found in maize (*Zea mays*), also known as corn, in Peru in 1973. MCMV next spread long-distance via contaminated seeds to Kansas and Nebraska in midwestern USA (1976). Testing of native grasses and common cereal crops

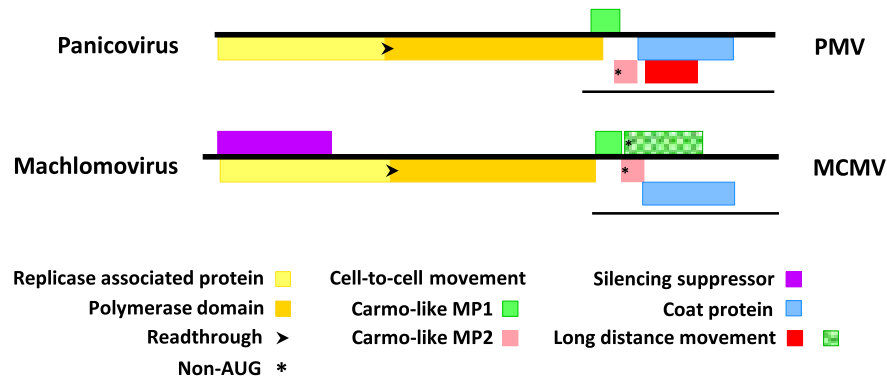


Fig. 2 Genome organizations of MCMV and PMV. Thick black lines mark the gRNAs while the thinner short lines are sgRNAs. ORFs encoding proteins with the same function are the same color while unique ORFs have unique colors (identified below the genomes) along with symbols for readthrough stop codons and noncanonical start codons. ORFs are located relative to the RdRp reading frame (– 1 above and + 1 below).

indicated that MCMV-K and MCMV-P systemically infected 19 and 15 species, respectively, including wheat and barley. Since MCMV seed transmission rate varies widely for different maize cultivars, and maize production has increased worldwide, the virus has spread to most continents. Over the years MCMV has appeared in Mexico, additional South American countries, Hawaii, Thailand, Mainland China, and Taiwan, seven countries in Eastern Africa, and most recently was detected in maize and johnsongrass in Spain. MCMV has infected additional *Poaceae* crops in Africa (sorghum, sugarcane, and finger millet) and was found in johnsongrass, a common agricultural weed which grows nearly worldwide. China has reported MCMV in maize, sugarcane, sorghum, and Job's tears (*Coix lacryma-jobi*).

PMV infects switchgrass (*Panicum virgatum*) which grows in most of the USA and Canada except the farthest Western & Northern States/Provinces. PMV mostly occurs in 12 states: Nebraska-to-Texas, and Eastern States below parallel 36°30'N except Tennessee. PMV also infects the warm-season lawn grasses St. Augustine grass (*Stenotaphrum secundatum*) and centipedegrass (*Eremochloa ophiuroides*), and the model organism purple false brome (*Brachypodium distachyon*). For the two other recognized panicoviruses, CMMV infects a cool season perennial grass (*Dactylis glomerata*) native to northern Europe, temperate Asia, and northern Africa that is grown in hay meadows and pastures, while TPAV was found in a native tallgrass prairie in Oklahoma, USA. Bermuda grass latent virus (BGLV), found in South-Eastern USA, is likely another asymptomatic panicovirus that infects Bermuda grass (*Cynodon dactylon*) which is a warm-season grass used for sports fields and lawns.

Life Cycle

For both MCMV and PMV, upon entering a plant cell the virion disassembles, exposing the vRNA, which, based on work with many tombusvirids, folds into complex secondary and tertiary structures that regulate the various stages of virus replication and gene expression via LDRIs, including translation, replication, and sgRNA synthesis. Both viruses replicate in the cytoplasm, and the vRNA (and later the sgRNA) engages with ribosomes via unique genus-specific 3' CITEs: the PMV-like 3' CITE (PTE) and a PTE variant for MCMV in their 3' UTRs. Base-pairing between a small stem loop in the 5' UTRs of both vRNA and sgRNA with a sequence in the 3' CITE allows the circularization function for each RNA that is typically performed for eukaryotic mRNAs by translation factors that bind poly(A) tails and m⁷G-caps. RAP and lower quantities of RdRp are synthesized from vRNA, and RAP and RdRp are the only viral proteins required for replication in protoplasts for both viruses. Replication, including (–)strand synthesis, most likely occurs in membrane-bound vesicles on mitochondrial or peroxisomal membranes as has been shown for some tombusviruses. During a viral infection (–)RNA accumulates to levels that allow isolation of dsRNA from leaf tissue. For MCMV, p32 is also synthesized from vRNA at low levels from the 5' overprinted ORF. MCMV p32 is not required for replication in protoplasts or plants, but infectious transcripts lacking p32 expression accumulate virus at ~1/3 of the level of wild type in protoplasts, and produce delayed, mild infections in maize plants, suggesting this gene might encode a unique suppressor of silencing. A single large sgRNA is synthesized for MCMV and PMV, which contains all the ORFs necessary for cell-to-cell and long distance movement. For both species, proteins expressed from the two small ORFs (MP1 and MP2) are required for cell-to-cell movement, while MP1, MP2, and CP are required for efficient long-distance movement. The unique ORFs in PMV and MCMV overlapping their CP ORFs encode proteins required for long distance viral movement. Fortuitously, when MCMV gene functions were analyzed, the mutations introduced to “stop” the ability to read through the MCMV p7a stop codon also altered the noncanonical start codon for p23, so the function of the downstream ORF was accurately identified. MCMV infections also produce a small 337 nt 3' coterminal RNA, sgRNA2. This is likely an exoribonuclease-resistant RNA (xrRNA) produced by a cellular exoribonuclease, which enhances the viral infection. xrRNAs have been identified in viruses from five other tombusvirid genera.

Epidemiology

PMV and other panicoviruses are transmitted via mechanical inoculation, usually lawnmowers (St. Augustine grass and centipede grass) or other grass-harvesting/cutting tools such as hay swathers and balers for switchgrass. MCMV is also readily transmitted via mechanical inoculation, but it has also been shown to be transmitted via corn rootworm beetles and thrips. It is likely that transmission via these insects occurs through the maceration of the tissue (beetles) or rasping/sucking mouthparts of the thrips after the insects have fed on infected plants, which would be mechanical inoculation via insects. MCMV also exhibits transmission through the soil since immediately replanting maize in fields that contained MCMV-infected maize results in a high rate of infection of the new crop. Some chytrid fungi in the genus *Olpidium* can transmit tombusviruses via binding of assembled virions to fungal zoospores, so a fungus may be the vector for MCMV transmission via soil. Growing a nonhost crop such as soybeans between maize crops stops soil-transmission.

Pathogenesis

Both MCMV and PMV are part of highly destructive synergistic diseases that are produced by different co-infections. SAD is caused by coinfection of PMV and Satellite panicum mosaic virus (SPMV) or the closely related Satellite St. Augustine decline virus (SSADV) in St. Augustine grass. PMV supports the replication of SPMV (824 nt) and two subviral RNAs, satS (375 nt) and satC (444 nt). SPMV encodes its own CP (17 kDa), and SPMV forms icosahedral virions that exclusively encapsidate its own RNA or PMV satRNA. Encapsidated SPMV accumulates in the cytosol, and SPMV CP also accumulates in the cell wall, nuclear membranes, and nucleolus. The satRNA satS replicates in pearl millet protoplasts or plants when RNA from PMV or PMV + SPMV are present, and satS decreases PMV and SPMV accumulation, thereby decreasing disease symptoms.

While single infections of MCMV produce relatively mild symptoms in maize that vary with the cultivar, double infections with the maize-infecting potyviruses [(Maize dwarf mosaic virus (MDMV), Johnson grass mosaic virus (JGMV), Sugarcane mosaic virus (SCMV)], or the tritivirus Wheat streak mosaic virus (WSMV) causes a severe synergistic infection called CLN or MLN. Potyviruses are transmitted by aphids while WSMV is transmitted by the wheat curl mite. In most synergistic diseases involving potyvirids, the potyvirus accumulates to about the same level in single or double infections, while the other virus (MCMV in CLN) can increase dramatically, and this occurs for MLN caused by potyviruses. This is probably due to the highly effective potyvirus silencing suppressor, HC-Pro. In MCMV + WSMV infections of maize, the level of WSMV also increased in doubly-infected plants, and the increase was temperature-dependent. The recent MLN outbreak in Africa involved SCMV and JGMV, while SCMV is associated with CLN in China.

Diagnosis

Antibody-based detection kits are commercially available for both MCMV and PMV, and both viruses are readily detected from leaf tissue via RT-PCR using primers within the sgRNA region.

Prevention

Scrupulous cleaning of lawn equipment and farming equipment (balers, tillers, combines) decreases the chance of transmitting the viruses to additional fields. There are no treatments for living infected plants. Crop rotation of maize with any dicot crop stops soil transmission of MCMV.

Further Reading

- Chkuaseli, T., White, K.A., 2018. Intra-genomic long-distance RNA–RNA interactions in plus-strand RNA plant viruses. *Frontiers in Microbiology* 9, 529.
- Kraft, J.J., Peterson, M.S., Cho, S.K., *et al.*, 2019. The 3' untranslated region of a plant viral RNA directs efficient cap-independent translation in plant and mammalian systems. *Pathogens* 8, 8010028.
- Miller, W.A., Koev, G., 2000. Synthesis of sub-genomic RNAs by positive-strand RNA viruses. *Virology* 273, 1–8.
- Miller, W.A., Shen, R., Staplin, W., Kanodia, P., 2016. Non-coding RNAs of plant viruses and viroids: sponges of host translation and RNA interference machinery. *Molecular Plant-Microbe Interactions* 29, 156–164.
- Miras, M., Miller, W.A., Truniger, V., Aranda, M.A., 2017. Non-canonical translation in plant RNA viruses. *Frontiers in Plant Science* 8, 494.
- Mwando, N.L., Tamiru, A., Nyasani, J.O., *et al.*, 2018. Maize chlorotic mottle virus induces changes in host plant volatiles that attract vector thrips species. *Journal of Chemical Ecology* 44, 681–689.
- Nagy, P.D., 2016. Tombusvirus-host interactions: Co-opted evolutionarily conserved host factors take center court. *Annual Review of Virology* 3, 491–515.
- Palukaitis, P., 2016. Satellite RNAs and satellite viruses. *Molecular Plant-Microbe Interactions* 29, 181–186.
- Pyle, J.D., Scholthof, K.B.G., 2017. Biology and pathogenesis of satellite viruses. In: Hadidi, A., Flores, R., Randles, J., Palukaitis, P. (Eds.), *Viroids and Satellites*. Academic Press.
- Scheets, K., 1998. Maize chlorotic mottle machlomovirus and Wheat streak mosaic rymovirus concentrations increase in the synergistic disease corn lethal necrosis. *Virology* 242, 28–38.

- Scheets, K., 2000. Maize chlorotic mottle machlomovirus expresses its coat protein from a 1.47-kb sub-genomic RNA and makes a 0.34-kb sub-genomic RNA. *Virology* 267, 90–101.
- Scheets, K., 2016. Analysis of gene functions in maize chlorotic mottle virus. *Virus Research* 222, 71–79.
- Steckelberg, A.L., Vicens, Q., Kieft, J.S., 2018. Exoribonuclease-resistant RNAs exist within both coding and noncoding subgenomic RNAs. *mBio* 9, e02461-18.
- Wang, Z., Parisien, M., Scheets, K., Miller, W.A., 2011. The cap-binding translation initiation factor, eIF4E, binds a pseudoknot in a viral cap-independent translation element. *Structure* 19, 868–880.

Maize Streak Virus (*Geminiviridae*)

Darren P Martin, University of Cape Town, Cape Town, South Africa

Aderito L Monjane, Norwegian Veterinary Institute, Oslo, Finland

© 2021 Elsevier Ltd. All rights reserved.

This is an update of D.P. Martin, D.N. Shepherd, E.P. Rybick, Maize Streak Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00707-X](https://doi.org/10.1016/B978-012374410-4.00707-X).

Abbreviations

C-ori complementary strand (i.e., the half of the DNA duplex that is not packaged into virus particles) origin of replication.

CP coat protein. The only protein component of the virus particle also believed to be involved in nuclear trafficking and cell-to-cell movement of viral DNA.

LIR long or large intergenic region containing the origin of virion strand replication and gene promoters.

MP movement protein. A small (c. 10 kDa) protein believed to be involved in intercellular virus movement via plasmodesmata.

Rep replication-associated protein involved in the initiation of virion strand replication.

RepA a truncated version of Rep with a unique C-terminal domain believed to be involved in regulation of host and/or virus gene expression.

SIR short or small intergenic region containing gene polyadenylation signals and the origin of complementary strand replication.

V-ori virion strand (i.e., the half of the DNA duplex that is packaged into virus particles) origin of replication.

Glossary

Agro-infection Technique used for the infection of host plants with cloned virus genomes involving transfer of virus DNA into the nuclei of host cells by the bacterium *Agrobacterium tumefaciens*.

Agro-inoculation See agro-infection.

Bicistronic Contains two protein-coding regions within a single mRNA transcript.

Capsomer A subunit of the mature virus particle containing an ordered series of polymerized coat protein molecules.

Cicadulina spp. A group of leafhopper species involved in transmission of MSV.

Oviposition Egg-laying. To oviposit means to lay eggs.

Plastochron The time interval between successive leaf primordia, or the attainment of a certain stage of leaf development.

Viruliferous A state in which an MSV vector species is carrying and is capable of transmitting the virus.

Introduction

Maize streak virus (MSV) is the causal agent of maize streak disease (MSD), one of the most serious viral diseases of maize in Africa. It is a major contributor to the continent's food security problems and is endemic throughout Africa south of the Sahara. It is also found on the Indian Ocean islands of Madagascar, Mauritius, and La Réunion. There is no obvious barrier to spread of the virus outside of this region; hence it should be considered as a serious potential problem for other as yet unaffected maize-growing areas.

History and Taxonomy

"The disorder of the mealie plant, locally described as "Mealie Blight", "Mealie Yellows", or "Striped Leaf Disease", belongs to a group of plant troubles arising from obscure causes ..." was how MSD was first described by Claude Fuller in 1901 in Natal, South Africa. Fuller mistakenly attributed the disease to a soil disorder, but in retrospect it is quite clear that the "mealie (a local word for maize or corn) variegation" he described and drew in minute detail can be attributed to MSV.

The first milestone in MSD research was reached in 1924, when Storey determined that a virus transmitted by leafhopper species of the genus *Cicadulina* (Fig. 1) was the causal agent of MSD. Storey named the virus "maize streak virus", and was also the first to determine both the genetic basis of MSV transmission by *Cicadulina mbila*, and the heritability of MSD resistance in maize.

When MSV particles were first purified in 1974 they were found to have a novel twinned quasi-icosahedral (geminiate) shape (Fig. 2), from which the name "geminivirus" was derived. This was followed by the unexpected discovery in 1977 that geminivirus particles contain circular single-stranded DNA (ssDNA); a genome type never before observed in plant viruses. These novel characteristics led to the proposal of a new virus group – the geminiviruses – comprising MSV and other viruses with geminate particles and ssDNA genomes. *Maize streak virus* is now recognized as the type species of the genus *Mastrevirus*, in the family *Geminiviridae*.



Fig. 1 The leafhopper vector of MSV, *Cicadulina mbila* Naudé. Photograph courtesy of Dr. Benjamin Odhiambo, Kenyan Agricultural Research Institute (KARI).



Fig. 2 MSD symptoms on a maize leaf: note characteristic veinal streaks. Photograph courtesy of Dr. Frederik Kloppers, PANNAR (Pty) Ltd., Greytown, KwaZulu-Natal, South Africa.

Host Range and Symptoms

While most notorious for the yield losses it causes when infecting maize, MSV also infects over 80 other grass species including the economically important crops wheat, barley, and rye. In susceptible maize and grass genotypes, the virus first causes symptoms between 3 and 7 days after inoculation. These first appear as almost circular pale spots of 0.5–2 mm diameter in the lowest exposed portions of the youngest leaves. Later, fully emerged symptomatic leaves show veinal streaks from a few millimeters long to the entire length of the leaf and between 0.5 and 3 mm wide. These streaks often fuse laterally and symptomatic leaves may become >95% chlorotic (**Fig. 2**).

Plants are worst affected when infected within a few days of coleoptile emergence; symptoms only develop above the site of inoculation on newly emerging leaves. Susceptible varieties may display severe stunting as well as very severe streaking. Afflicted plants will frequently produce deformed cobs or may fail to produce cobs altogether and per-field yield losses can reach 100% when susceptible maize genotypes are infected early.

Of the eleven major MSV strains so far identified (designated MSV-A through MSV-K; (**Fig. 3**)), only MSV-A produces economically significant infections in maize. The “grass-adapted” MSV strains (MSV-B, -C, -D, -E, -F, -G, -H, -I, -J and -K) differ from MSV-A types by 5%–25% in nucleotide sequence, and produce substantially milder symptoms in maize than do MSV-A viruses. In many cases these grass-adapted strains are incapable of producing symptomatic infections in MSV-resistant maize genotypes.

Diversity and Evolution

MSV is closely related to the other distinct ‘African streak’ mastreviruses, Panicum streak virus, Sugarcane streak virus, Sugarcane streak Egypt virus, and Sugarcane streak Reunion virus, with which it shares ~65% genome sequence identity. It is, however, most similar to, although not necessarily more closely related to, an isolate of *Digitaria streak virus* from the Pacific island of Vanuatu, with which it shares ~67% genome sequence identity (**Fig. 3**).

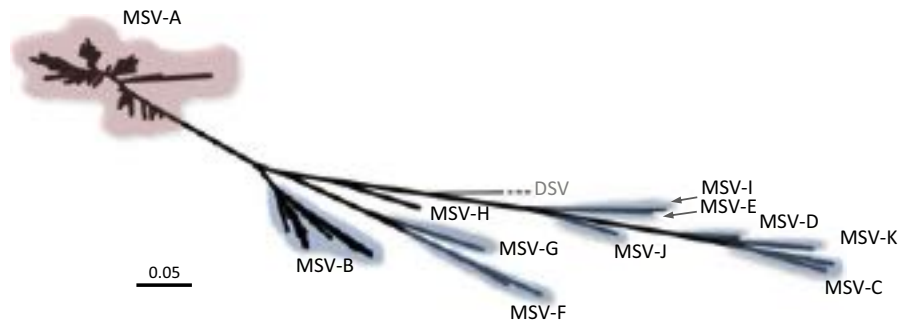


Fig. 3 Phylogenetic relationships between the full genome sequences of different MSV strains. The tree is constructed using the maximum-likelihood method (HKY model transition and transversion weight determined from the data and 100 bootstrap replicates) and numbers associated with branches indicate degrees of bootstrap support for those branches. Branches with less than 70% support have been collapsed and the genome sequence of a *Digitaria* streak virus (DSV) from Vanuatu is included as an outgroup. Only viruses in the MSV-A group have been isolated from maize. All viruses in the other groups have been isolated from wheat, barley, or wild grass species.

The full genome nucleotide sequences of MSV-A isolates display relatively low degrees of diversity, with any two MSV-A isolates obtained from anywhere in Africa invariably having genome sequences that are more than 97% identical. MSV isolates from La Réunion share ~95% identity with mainland isolates.

Although Maize was first introduced at multiple points into Africa and its neighboring islands in the 16th century, it is apparent that the most recent common ancestor of all presently sampled MSV-A genomes only came into existence in the mid 1800s somewhere in South-Eastern Africa. It is also likely that this ancestral virus was the recombinant offspring of grass adapted-adapted parental viruses: one parent an MSV-B variant, and the other parent belonging to a presently unsampled MSV strain that is most closely related to MSV-F and MSV-G.

Transmission

In nature, MSV and other African streak viruses are neither seed nor contact transmissible and rely instead on transmission by cicadellid leafhoppers in the genus *Cicadulina* (including, among others, *C. mbila*, *C. storeyi*, *C. bipunctella zaeae*, *C. latens*, and *C. parazeae*). Of these, *C. mbila* is considered the most important MSV vector as it is the most widely distributed. Also, a greater proportion of *C. mbila* individuals are capable of transmitting the virus than is found in other *Cicadulina* populations. The virus may be acquired by leafhoppers at any developmental stage in less than 1 h of feeding with a minimum acquisition time of 15 s. A latent period within the vector, during which the virus cannot be transmitted, lasts between 12 and 30 h at 30°C. Once this latent period is over (signaled by the appearance of virus within the leafhopper's body fluids) the virus can be transmitted within 5 min of feeding. Once MSV has been acquired by a leafhopper, the insect will transmit the virus for the rest of its life.

Particle Structure

MSV particles (Fig. 4) consist of two incomplete icosohedra with a $T = 1$ surface lattice, comprising 22 pentameric capsomers each containing five coat protein (CP) molecules. Particle dimensions are 38 nm × 22 nm with 110 CP molecules in each virion packaging a ~2690 nt covalently closed mostly single-stranded circular DNA genome. The packaged DNA has annealed to it a complementary ~80 nt sequence believed to act as a primer for complementary strand synthesis following infection and uncoating.

Genome Organization

As with all other mastreviruses discovered to date, the MSV genome encodes four proteins: a movement protein (MP), a coat protein (CP), a replication-associated protein (Rep), and a regulatory protein (RepA). Whereas CP and MP are expressed off alternatively spliced virion sense transcripts, Rep and RepA are expressed off alternatively spliced complementary sense transcripts. MSV genomes also contain two intergenic regions: these are a short or small one (SIR), and a long or large one (LIR), which are the complementary sense strand and virion sense strand origins of replication, respectively (Fig. 5).

The Long Intergenic Region

Besides containing divergent RNA polymerase II-type promoters and other transcriptional regulatory features necessary for the expression of the complementary and virion sense genes, the LIR also contains sequence elements that are essential for replication. The

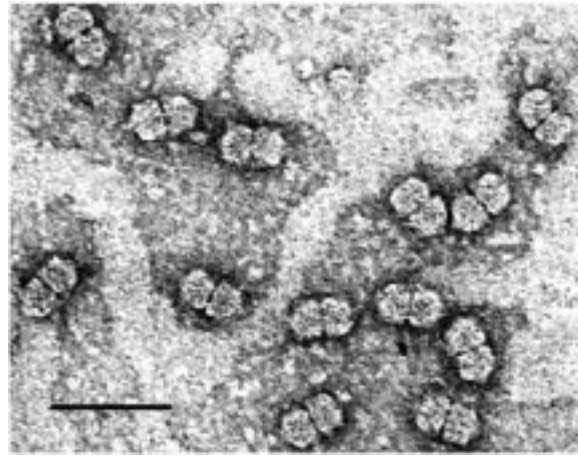


Fig. 4 Electron micrograph of MSV purified from infected maize, showing particles of 18 nm × 30 nm stained with uranyl acetate. Scale = 50 nm. Photograph courtesy of Kassie Kasdorf; copyright EP Rybicki.

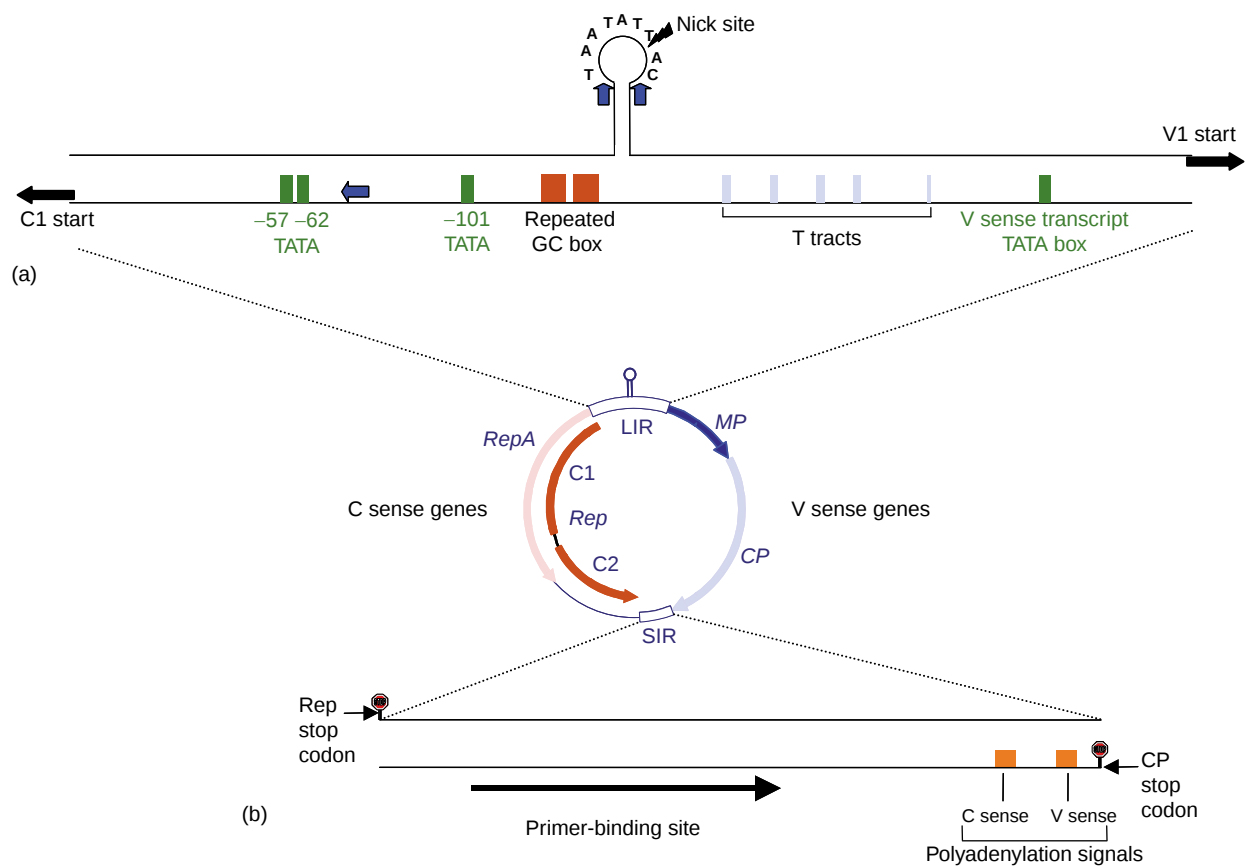


Fig. 5 A schematic representation of the MSV LIR (a) and SIR (b), shown in context with the MSV genome. In (a) the main features of the MSV LIR are shown. These include a stem-loop structure with the loop's nonanucleotide sequence conserved among all geminiviruses and other rolling-circle systems. The site at which Rep introduces an endonucleolytic nick to initiate virion strand replication is shown. Iterated sequences (iterons) are shown in the V sense, with blue arrows indicating their location in the LIR. Iterons are potentially specific Rep-recognition sequences via which Rep may bind to the LIR. 5' of the stem-loop is a repeated GC-box, which binds host transcription factors. A series of T tracts 3' of the stem-loop may be involved in DNA bending of this region of the LIR. TATA boxes 5' and 3' of the stem-loop are potential C sense and V sense transcription initiation sites, respectively. In (b), the main features of the SIR include polyadenylation signals for V and C sense transcripts, and a primer-binding site on the plus strand. An ~80 bp DNA primer-like molecule, encapsidated with the viral genome and annealed to this site, is thought to be involved in initiating complementary strand replication. Both the MSV LIR and SIR are essential for viral replication.

most striking of these is an inverted repeat sequence that is capable of forming a stable hairpin loop structure. All geminiviruses sequenced to date have the highly conserved nonanucleotide sequence TAATATTAC within the loop sequence of similar hairpin structures: this sequence contains the virion sense strand origin of replication (*V-ori*;1).

A sequence 6–12 nt long occurring in all known mastreviruses between the TATA box that directs *rep/repA* transcription and the *repA* initiation codon is directly repeated in the stem near the *V-ori* hairpin, and is probably involved in *Rep* and/or *RepA* binding during replication.

The hairpin and two GC boxes on the 5' side of the stem also forms part of an upstream activator sequence (UAS) required for efficient CP expression. The GC boxes bind nuclear factors to the UAS and are known as the rightward promoter element (*rpe1*).

The Short Intergenic Region

The MSV SIR occurs between the termination codons of the *CP* and *rep* genes (Fig. 5) and contains the polyadenylation and termination signals of the virion and complementary sense transcripts. The SIR also contains the origin of complementary strand synthesis (*C-ori*). A small 80-nt-long primer-like molecule is bound to the SIR of encapsidated MSV DNA and, at the onset of an infection, probably enables synthesis of double-stranded DNA (dsDNA) replicative forms (RFs) of the genome from newly uncoated virion strand DNA.

The Complementary Sense Genes (*Rep* and *RepA*)

Rep is the only MSV gene product that is absolutely required for virus replication. In mastreviruses *Rep* is encoded within two open reading frames (ORFs) referred to as C1 and C2 (Fig. 5). Beginning at the same transcription initiation site, two C sense transcripts (1.5 and 1.2 kbp in size) are produced during MSV infections. Splicing of the larger transcript removes an intron, which permits expression of full-length *Rep* from the two ORFs.

It is very probable, although as yet unproven *in vivo*, that *RepA* is translated from both the unspliced 1.5 kbp transcript and the 1.2 kbp transcript. If expressed in infected cells, MSV *RepA* would have the same N-terminal 214-amino-acid sequence as *Rep*, but would have a different C-terminus. *RepA* is likely a multifunctional protein that modifies the nuclear environment to favor viral replication.

The N-terminal portions of *Rep* and *RepA* contain three conserved amino acid sequence motifs commonly found in replication-associated proteins of many extremely diverse rolling-circle replicons. Other significant landmarks include a plant retinoblastoma-related protein (pRBR) interaction motif (via which *RepA* binds to host pRBR molecules to manipulate the cell cycle), and oligomerization domains (via which *Rep* and *RepA* bind to other *Rep/RepA* molecules to form homo- and heterooligomers). It is also likely that the approximately 100 N-terminal amino acids of both *Rep* and *RepA* are involved in binding these proteins to the viral LIR during replication.

The C-terminal portion of *Rep* contains a dNTP-binding domain with motifs similar to those found in proteins with kinase and helicase activities. The dNTP-binding domain also sits within a region with similarity to the DNA-binding domains of the myb-related class of plant transcription factors: this domain may be functional in the induction of virus and/or host gene transcription.

The C-terminal portion of *RepA*, which is different from that of *Rep*, contains another potential transactivation domain also possibly involved in the regulation of virus and/or host gene expression. A second domain within the C-terminal portion of *RepA* possibly interacts with host proteins involved in developmental regulation.

The Virion Sense Genes (*MP* and *CP*)

Transcription of the MSV virion (V) sense genes is directed by two TATA boxes within the LIR 26 and 214 nucleotides 5' of the MP start codon. Each TATA box directs the production of different-sized transcripts, both of which terminate at the same place. Splicing of an intron within the MP portion of V sense transcripts appears to be an important determinant of relative MP and CP expression levels. Whereas CP is expressed from both long and short, spliced and unspliced V sense transcripts, MP is most likely only expressed from unspliced long transcripts.

The MP is post-translationally modified and contains a hydrophobic domain that may either facilitate its interaction with host cell membranes or be involved in homo- or hetero-oligomerization with the CP.

The N-terminal ~100 amino acids of the CP contain both nuclear localization signals and a sequence-nonspecific dsDNA- and ssDNA-binding domain. The CP and MP interact with one another and it is possible that this interaction is involved in trafficking of naked and/or packaged virus DNA from nuclei through nuclear pores, to the cell periphery, through plasmodesmata, and into the nuclei of neighboring cells.

Molecular Biology of MSV

Replication

As with other geminiviruses, MSV replicates by both a rolling-circle mechanism (rolling-circle replication, or RCR; (Fig. 6)) and a recombination-dependent mechanism (recombination dependent replication or RDR). As with other rolling-circle replicons, MSV

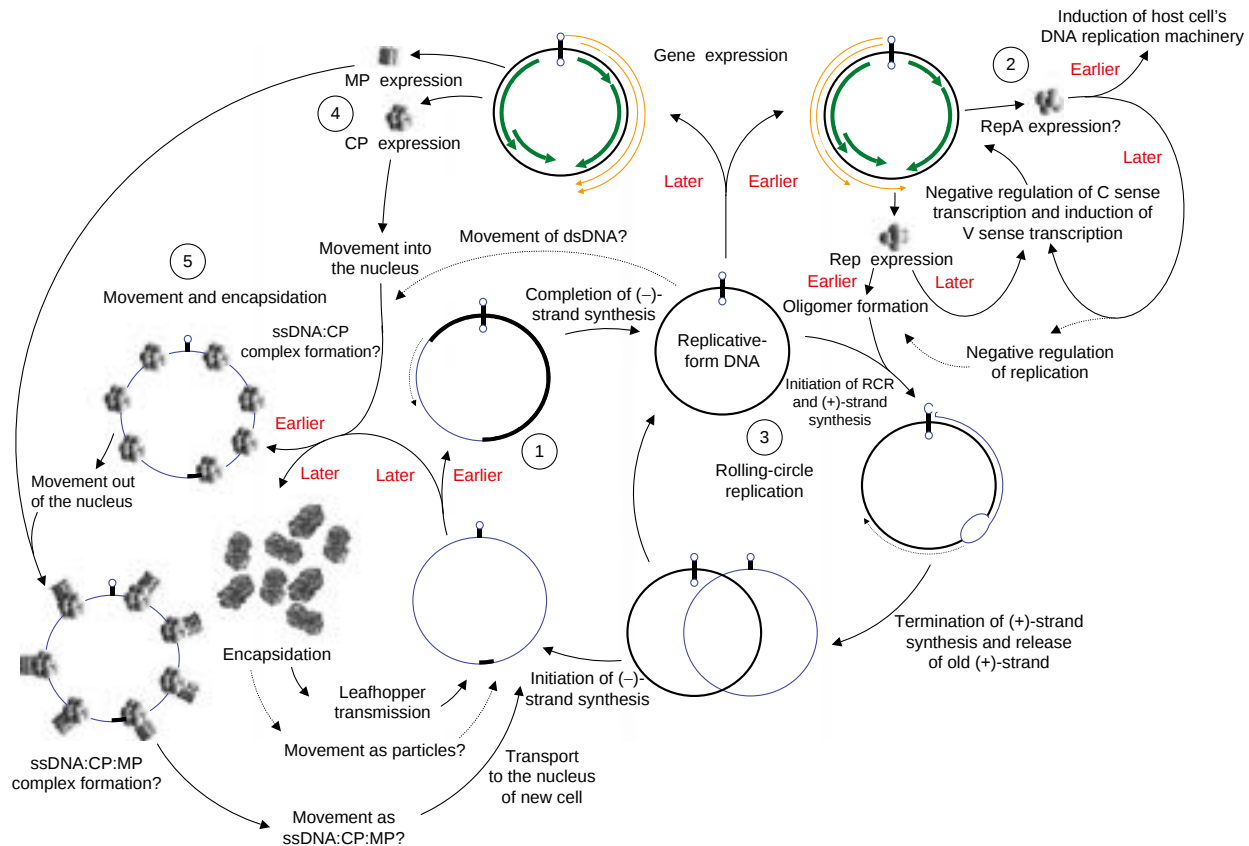


Fig. 6 Summary of the MSV infection process. Early during an infection following the synthesis of a dsDNA replicative form (RF; 1) RepA is most likely expressed and induces a cellular state in which viral DNA replication can occur (2). Rep is also expressed early and RCR begins (3). At a later point in the infection process, following genome amplification and possibly Rep and/or RepA induction of the V sense promoter, MP and CP are expressed (4), and movement and encapsidation occur (5). Represented here is movement of un-encapsidated ssDNA, but it should be noted that it is possible that dsDNA and/or encapsidated ssDNA may also be moved either cell to cell or systemically within the phloem of plants. Whereas the involvement of MSV CP and MP in movement has been demonstrated, the mechanics of the process are obscure. While the probable timing of events is indicated, it is unlikely, for example, that absolutely no MP and CP expression occurs during the earlier stages of the infection process. ssDNA is represented by blue lines, dsDNA by bold black lines, and RNA by orange lines.

replication is discontinuous with virion strand replication being initiated from the hairpin structure in the LIR and complementary strand synthesis being initiated from a short 80 nt primer-like molecule synthesized on the SIR of newly replicated virion strands.

Particle Assembly and Movement

Besides being the primary location of replication, the nucleus is also the site of virus particle assembly. CP molecules in the nucleus nonspecifically bind virion strands released during RCR (there is no known encapsidation signal in mastrevirus genomes), arresting the synthesis of new RF DNAs. Viral ssDNA molecules are packaged into particles that aggregate to form large paracrystalline nuclear inclusions. Crystalline arrays of MSV particles have also been detected outside nuclei within physiologically active phloem companion cells, and inside the vacuoles of dead and dying cells within chlorotic lesions. These lesions are caused by an as yet unexplained degeneration of chloroplasts in infected cells.

The mechanistic details of MSV cell-to-cell movement are still obscure, but it seems to involve an interaction between the CP, MP, and viral DNA. Besides requiring the coordinated interactions of viral gene products and DNA, the successful movement of MSV genomes from infected to uninfected cells is strongly dependent on the extent of plasmodesmatal connections between neighboring cells. Also, in maize it appears as though certain cell types are more sensitive to MSV infection than others. For example, in maize leaves the virus infects all photosynthetic cell types (e.g., mesophyll and bundle sheath cells) but despite abundant plasmodesmatal connections between photosynthetic, epidermal, and parenchyma cells, MSV is only rarely detectable in the latter two cell types.

It is unknown whether systemic movement of geminiviruses within plants simply relies on normal cell-to-cell movement to deliver genomic DNA into the phloem, or whether viral DNA is specifically packaged for long-distance transport. It is possible that

cell-to-cell movement might involve un-encapsidated ss- or dsDNA but that long-distance movement in the phloem might require encapsidation.

Long-distance movement of MSV within infected plants occurs via phloem elements and it is believed that MSV is incapable of invading the root apical, shoot apical, and reproductive meristems due to the absence of developed vasculatures in these tissues. Thus, the virus is not found in tissues that develop into gametes and is therefore not seed-borne. It also does not appear to travel within plants from sites of infection into older uninfected tissues.

Within the shoot apex where most productive MSV replication occurs, MSV first enters developing leaves at approximately plastochron five. While the virus is restricted to the developing leaf vasculature before plastochron 12, it is likely that the development of metaphloem elements at approximately plastochron 12 provides an opportunity for the virus to escape the vasculature into the photosynthetic cells of the leaf. Metaphloem develops with the abundant plasmodesmatal connections required for efficient loading of photoassimilates once the leaf emerges from the whorl. Before emergence, however, the developing photosynthetic tissues are still net importers of photoassimilates and the virus most likely moves into these cells through their plasmodesmatal connections with the metaphloem.

On the leaves, the pattern of chlorotic streak-like lesions that characterizes MSV infections is directly correlated with the pattern of virus accumulation within the leaves and the virus can only be acquired by leafhoppers from these lesions. The degree of chlorosis that occurs within lesions can differ between MSV isolates and is related to the severity of chloroplast malformation that occurs in infected photosynthetic cells.

Control of MSD

Although effective control of MSD in cultivated crops is possible with the use of carbamate insecticides, and it is possible to avoid leafhopper infestations by varying planting dates, the fact that small farmers cannot generally use these options means that the development and use of MSV-resistant crop genotypes is probably the best way to minimize the impact of MSD on African agriculture.

MSV resistance is associated with up to five separate alleles conferring a mixture of both dominant and recessive traits, none of which are sufficient by themselves to prevent MSV infections. Despite great successes achieved in the development of maize genotypes that tolerate MSV infection without significant yield loss, there has been only limited success in the field. For example, severe infections of so-called MSV-tolerant genotypes can occur when they are grown under environmental conditions different from those in which the plants were selected, meaning that distinct geographical growing areas may each require genotypes with MSV resistance that is tailored to those areas.

The problem facing breeders is that natural genetic resistance to MSD is not usually associated with desirable agronomic traits such as good yield. It can therefore be difficult to transfer resistance traits without also transferring undesirable characteristics. Moreover, the number of alleles involved means that successful breeding takes years for each release. Even in the absence of any predictive modeling of sporadic MSD outbreaks, most farmers would still prefer to gamble on the use of higher-yielding MSV-sensitive genotypes.

Efforts are also currently underway to introduce MSV resistance traits into commercial maize genotypes by genetic engineering. This could have the advantage of enabling the direct transfer of single-gene resistance, without linkage to undesirable characteristics, to many different breeding lines suited to different environmental conditions. However, up till now this strategy has been limited by negative public perception of genetically modified organisms, and the expensive and time-consuming risk assessment necessary to ensure a safe feed and food product.

MSD Epidemiology

There are loose correlations between MSD incidence and both environmental conditions and agricultural practices. Environmental influences on MSD epidemiology are mostly driven by a strong correlation between rainfall and leafhopper population densities. For example, drought conditions followed by irregular rains at the beginning of growing seasons tend to be associated with severe MSD outbreaks. Also, maize planted later in the growing season tends to get more severely infected than that planted at the beginning of the season, probably due to steady increases in leafhopper numbers and inoculum sources over the course of the season. As is the case with most insect-borne virus diseases, however, the incidence of MSD is erratic. Whereas MSD can devastate maize production in some years, in others it has only a negligible effect. The reason for this is that, apart from MSD epidemiology being strongly dependent on environmental variables, it is also the product of complex interactions between the various MSV leafhopper vector and host species, and an as yet unknown number of virus strains.

Leafhopper Vectors

Serious MSD outbreaks are absolutely governed by leafhopper acquisition and movement of severe MSV isolates from infected plants (wild grasses or crop plants) to sensitive, uninfected crop plants. The distance that MSV spreads from a source of inoculum

is determined by the movement behavior of leafhoppers. Distinct long- and short-distance flight morphs have been detected among certain *Cicadulina* populations. It is believed that the long-flight morphs are a migratory form and, as such, these may play an important part in the rapid long-distance spread of virulent MSV variants. Migratory movement is more common in certain *Cicadulina* species than in others and it is probably influenced by environmental conditions.

The dynamics of primary infection following leafhopper invasion of a susceptible maize crop are influenced by leafhopper population densities, the proportions of viruliferous individuals in populations, and virus titers within these individuals. Disease spread within individual maize fields is apparently linear when only a few viruliferous leafhoppers are involved in transmission, but becomes exponential once the number of insects exceeds one individual per three plants.

Plant Hosts

Although attempts to understand the dynamics of MSD epidemics have focused primarily on vector population dynamics and behavior, an important component of MSD epidemiology is the population density, turnover, and demographics of the over 80 grass species that are both MSV and vector hosts. Because *Cicadulina* species favor certain annual grass hosts for mating and oviposition, the species composition of grass populations that vary seasonally in any particular area will directly influence leafhopper population densities and feeding behaviors in that area.

The species composition and age distribution of grasses (including cultivated crops) in an area may also affect the amount of MSV inoculum available for transmission in that area. While MSV infects at least 80 of the 138 grass species that leafhoppers feed on, both the susceptibility of these grasses to MSV infection and the severity of symptoms that occur following their infection may be strongly influenced by a number of factors. While sensitivity to infection can vary substantially from species to species, it can also vary within a species with genotype and plant age at the time of inoculation: for example, plants from many species, including maize, generally become more resistant to MSV infection with age, thereby reducing the inoculum available for transmission to other plants.

The Virus

While efforts are underway to promote the widespread cultivation of MSV-resistant maize in Africa, surprisingly little is known about the MSV populations that will confront these new genotypes. Although to date eleven major MSV strain groupings have been discovered, it is unknown whether any other than the maize-adapted MSV-A strain play an important role in the epidemiology of MSD. MSV-B, -C, -D, and -E isolates only produce very mild symptoms in MSV-sensitive maize genotypes and are therefore unlikely to pose any significant direct threat to maize production. Mixed MSV-A and -B infections have, however, been detected in nature and there is also strong evidence of recombination occurring between these strains. It is therefore possible that MSV-B, and possibly other MSV strains, may indirectly influence MSD epidemiology through recombination with MSV-A-type viruses. Recombination has been linked with the emergence of a number of geminivirus diseases and it may also have contributed to the original emergence of MSV-A. It is plausible therefore that it could also eventually contribute to the evolution of MSV-A variants (or even variants of the other MSV strains) with elevated virulence in resistant maize genotypes.

Future Threat

MSV is rightly regarded as a significant potential threat to maize production outside of Africa: while the vectors do not occur outside of Africa, there is no obvious reason that they would not survive in other equatorial or temperate regions of the world such as the southern USA, South America and Eurasia. It is a distinct possibility that they could inadvertently spread, or even be deliberately taken, to these regions. If, following the establishment of a MSV vector species outside of Africa, MSV was then introduced, the virus would inevitably begin spreading within native grasses, cultivated maize and the other cereals that it is known to infect. As none of the maize varieties grown outside of Africa has the slightest resistance to MSV, the probability of severe economic consequences would be very high.

See also: Emerging Geminiviruses (*Geminiviridae*). Geminiviruses (*Geminiviridae*). Plant Resistance to Geminiviruses

Further Reading

- Bosque-Pérez, N.A., 2000. Eight decades of maize streak virus research. *Virus Research* 71, 107–121.
 Boulton, M.I., 2002. Functions and interactions of mastrevirus gene products. *Physiological and Molecular Plant Pathology* 60, 243–255.
 Damsteegt, V., 1983. Maize streak virus. Part I: Host range and vulnerability of maize germ plasm. *Plant Disease* 67, 734–737.
 Efron, Y., Kim, S.K., Fajemisin, J.M., *et al.*, 1989. Breeding for resistance to maize streak virus – A multidisciplinary team-approach. *Plant Breeding* 103, 1–36.
 Fuller, C., 1901. First Report of the Government Entomologist, 1899–1900. P. Davis. <http://www.mcb.uct.ac.za/msv/fuller.htm>. (accessed February 2008).

- Harrison, B.D., Barker, I., Bock, K., *et al.*, 1977. Plant viruses with circular single-stranded DNA. *Nature* 270, 760.
- Martin, D.P., Willment, J., Billharz, R., *et al.*, 2001. Sequence diversity and virulence in zea mays of maize streak virus isolates. *Virology* 288, 247.
- McLean, A.P., 1947. Some forms of streak virus occurring in maize, sugarcane and wild grasses. *Science Bulletin of Department of Agriculture for Union of South Africa* 265, 1–39.
- Palmer, K.E., Rybicki, E.P., 1998. The molecular biology of mastreviruses. *Advances in Virus Research* 50, 183–234.
- Storey, H.H., 1925. The transmission of streak disease of maize by the leafhopper. *Balclutha mbila* naudé. *Annals of Applied Biology* 12, 422–443.
- Varsani, A., Shepherd, D.N., Dent, K., *et al.*, 2009. A highly divergent South African geminivirus species illuminates the ancient evolutionary history of this family. *Virology Journal* 25 (6), 36. doi:10.1186/1743-422X-6-36. (PubMed PMID: 19321000; PubMed Central PMCID: PMC2666655).
- Zhang, W., Olson, N.H., Baker, T.S., *et al.*, 2001. Structure of the maize streak virus germinate particle. *Virology* 279, 471–477.

Relevant Websites

<http://www.mcb.uct.ac.za>
Molecular & Cell Biology.

Nanoviruses (*Nanoviridae*)

Bruno Gronenborn, Institute for Integrative Biology of the Cell, CNRS, University of Paris-Sud, CEA, Gif sur Yvette, France
H Josef Vetten, Julius Kühn Institute, Braunschweig, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of H.J. Vetten, Nanoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00644-0](https://doi.org/10.1016/B978-012374410-4.00644-0).

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

CR-M Common region major

CR-SL Common region stem-loop

DAS-ELISA Double antibody sandwich form of ELISA

ELISA Enzyme-linked immuno-sorbent assay

EM Electron microscopy

kb Kilobases; the size of a ssDNA or ssRNA molecule

kDa Kilodaltons; the size of a protein

MAB Monoclonal antibody

MP Movement protein

nm Nanometer(s)

NSP Nuclear shuttle protein

nt Nucleotide(s)

PCR Polymerase chain reaction

pRB Protein retinoblastoma

RCA Rolling-circle amplification

RCR Rolling-circle replication

Rep Replication initiator protein

ssDNA Single-stranded deoxyribonucleic acid

TAS-ELISA Triple antibody sandwich-ELISA

Glossary

Alphasatellite Alphasatellites are single-stranded DNA molecules of 1.0–1.3 kb, which behave like virus satellites in many respects. Unlike the true non-coding satellites, alphasatellites code for a protein that specifically catalyzes only their own replication initiation. They are encapsidated in particles formed by the capsid protein of the respective helper virus.

Clink The protein named Clink (for “cell cycle link”) interacts with the cell cycle regulator protein retinoblastoma and thereby modulates cell cycle regulation in favor of DNA synthesis (S-phase). Clink enhances nanovirus replication.

Enzyme-linked immunosorbent assay The enzyme-linked immunosorbent assay (ELISA) is a sensitive and versatile serological test with two common variants, the double antibody sandwich (DAS)-ELISA and the triple antibody sandwich (TAS)-ELISA. A double antibody sandwich (DAS)-ELISA consists of four steps (separated by individual washings): (1) the immunoglobulins (IgG) isolated from a virus-specific antiserum are used for coating of microtiter plates; (2) IgG-coated plates are incubated with extracts of samples to capture antigen, if present; (3) incubation with enzyme (usually alkaline phosphatase)-labeled IgG [see (1)]; (4) incubation with substrate of the respective enzyme followed by measuring its activity, for instance by colorimetry. The triple antibody sandwich (TAS)-ELISA differs from DAS-ELISA in step (3), which is

replaced by two steps consisting of (i) incubation with a specific (often monoclonal) antibody from a different animal species than the animal used to produce the coating IgG in step (1), and (ii) incubation with enzyme-labeled IgG raised against IgG of that same different animal species.

Replication initiator protein A replication initiator protein (Rep protein or Rep, also referred to as RC-Rep) initiates rolling-circle replication of DNA by cleaving double-stranded DNA at a defined sequence (replication origin). The Rep protein remains bound to the 5′ phosphate of the cleaved strand while a polymerase uses the liberated 3′ hydroxyl of the cleaved strand as primer for DNA synthesis. After one round of DNA synthesis another subunit of Rep cleaves the newly synthesized DNA at the replication origin sequence. The thus generated 3′ hydroxyl attacks the phosphotyrosyl ester bond of the bound Rep subunit, and by a nucleotidyltransfer a covalently closed circular single-stranded DNA molecule is created.

Rolling-circle replication Rolling-circle replication (RCR) is a process of unidirectional nucleic acid amplification, by which multiple copies of circular DNA or RNA molecules such as plasmids, bacteriophage- or eukaryote virus genomes are reproduced.

Satellites Satellites are subviral nucleic acids that depend for replication, systemic movement, encapsidation, and/or vector transmission on a helper virus, with which they co-infect a host; they do not code for a protein.

Introduction

In 1988 and 1991 a novel type of single-stranded DNA (ssDNA) virus causing yellowing and dwarfing diseases of legumes or banana was described. The viruses were persistently transmitted by aphids and not transmissible by sap. Unusually small, icosahedral particles measuring only 18–20 nm in diameter were isolated from infected plants. Because of the small size of the virions, the small size of individual genome components of only about 1 kb, and the dwarfing symptoms they caused, these

viruses were referred to as nanoviruses. They differ from geminiviruses, the other group of ssDNA plant viruses, in particle morphology, genome size, number and size of DNA components, genomic organization, mode of transcription, and vector species. Common with other ssDNA viruses including geminiviruses is the mode of nanovirus replication. Unlike bipartite geminiviruses, in which the two genomic DNAs are individually encapsidated, the nanovirus genome is multipartite and consists of six or eight ssDNAs each of which is individually encapsidated in a separate particle. Here we provide a brief synopsis of our current knowledge of these viruses.

Classification and Taxonomy

To date, eleven established and three recently identified members of the family *Nanoviridae* have been described. On the basis of differences in biology (host range, aphid vectors) and both genome size and organization, the family members (nanovirids) are subdivided into the genera *Nanovirus* and *Babuvirus*. The respective type species of these two genera are *Subterranean clover stunt virus* and *Banana bunchy top virus*. Three babuvirus and eight nanovirus species have been approved by the International Committee on Taxonomy of Viruses (ICTV). Information on the geographic distribution, important aphid vectors and some natural hosts of the nanovirids as well as the amount and type of genome sequence data available for them is given in Table 1.

Crops in mainly tropical and subtropical countries of the Old World harbor babuviruses, such as Abaca bunchy top virus (ABTV) in the Philippines, Cardamom bushy dwarf virus (CBDV) in India and Banana bunchy top virus (BBTV). The latter occurs widely on Pacific Islands (including Hawaii), in Australia and Indochina but has an erratic geographic distribution in South Asia and Africa. While Subterranean clover stunt virus (SCSV) occurs only on legumes in Australia, Milk vetch dwarf virus (MDV) has been reported

Table 1 Assigned and tentative members of the genera *Babuvirus* and *Nanovirus* of the family *Nanoviridae*

	Geographic distribution	Aphid vector	Natural host	Genome components	Complete sequences
Genus <i>Babuvirus</i>					
<i>Banana bunchy top virus</i> (type species)	Africa, Australia, Pacific Islands, southern Asia,	<i>Pentalonia nigronervosa</i>	<i>Musa</i> spp.	6	> 200
<i>Abaca bunchy top virus</i>	Borneo, Philippines	no data	<i>Musa</i> spp.	6	2
<i>Cardamom bushy dwarf virus</i>	India	<i>Micromyzus kalimpongensis</i>	<i>Elettaria cardamomum</i> , <i>Amomum subulatum</i>	6	> 160
Genus <i>Nanovirus</i>					
<i>Subterranean clover stunt virus</i> (type species)	Australia	<i>Aphis craccivora</i> , <i>Aphis gossypii</i> , <i>Acyrtosiphon pisum</i>	<i>Trifolium subterraneum</i> and other legumes	8	2
<i>Black medic leaf roll virus</i>	Austria, Sweden, Azerbaijan	<i>Aphis craccivora</i> , <i>Acyrtosiphon pisum</i>	<i>Medicago lupulina</i> , <i>Pisum sativum</i>	8	3
<i>Faba bean necrotic yellows virus</i>	Near East, North Africa, Ethiopia, Spain	<i>Aphis craccivora</i> , <i>Aphis fabae</i> , <i>Acyrtosiphon pisum</i>	<i>Vicia faba</i> and other legumes	8	8
<i>Faba bean necrotic stunt virus</i>	Ethiopia, Morocco, Azerbaijan, Iran	<i>Aphis craccivora</i> , <i>Acyrtosiphon pisum</i>	<i>Vicia faba</i> , <i>Lens culinaris</i> , <i>Phaseolus vulgaris</i> and other legumes	8	8
<i>Faba bean yellow leaf virus</i>	Ethiopia	No data	<i>Vicia faba</i>	8	1
<i>Milk vetch dwarf virus</i>	China, Japan, Bangladesh	<i>Aphis craccivora</i> , <i>Acyrtosiphon pisum</i> , <i>Megoura viciae</i> , <i>Myzus persicae</i>	<i>Vicia faba</i> and many legumes and non-legumes i.e., <i>Nicotiana tabacum</i> , <i>Carica papaya</i> , <i>Catharanthus roseus</i>	8	8
<i>Pea necrotic yellow dwarf virus</i>	Germany, Austria, Hungary, Denmark, Serbia, Netherlands	<i>Aphis craccivora</i> , <i>Aphis fabae</i> , <i>Acyrtosiphon pisum</i> , <i>Megoura viciae</i>	<i>Pisum sativum</i> , <i>Vicia faba</i> and other legumes	8	8
<i>Pea yellow stunt virus</i>	Austria	No data	<i>Pisum sativum</i>	8	1
<i>Cow vetch latent virus</i> ^a	France	No data	<i>Vicia cracca</i>	8	1
<i>Parsley severe stunt-associated virus</i> ^a	Germany	No data	<i>Petroselinum crispum</i>	7 ^b	1
<i>Sophora yellow stunt-associated virus</i> ^a	Iran	No data	<i>Sophora alopecuroides</i> et al.	8 ^c	2

^aRecently described viruses not yet assigned to the genus *Nanovirus* by ICTV.

^bNo DNA-U4 found; four different types of DNA-R found but M-Rep not experimentally identified.

^cTwo different DNA-C types found; four different DNA-R types found but M-Rep not experimentally identified.

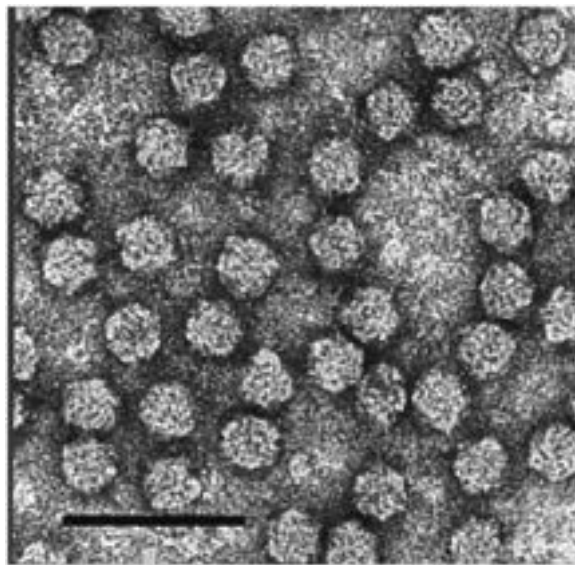


Fig. 1 Faba bean necrotic yellows virus particles. Negative contrast electron micrograph of virions of faba bean necrotic yellows virus. The scale bar corresponds to 50 nm. Courtesy of D.E. Lesemann and L. Katul.

from several legumes and non-legumes in Japan, China, and recently also in Bangladesh, suggesting that it has a wide distribution in several other countries of South-east Asia. Whereas Faba bean yellow leaf virus (FBYLV) was reported only from Ethiopia, Faba bean necrotic yellows virus (FBNYV) appears to have a much wider geographic distribution in West Asia, the Middle East, North- and East Africa and Europe (Spain). Faba bean necrotic stunt virus (FBNSV) has been found in Ethiopia, Morocco, Azerbaijan and Iran. Also from Iran, Sophora yellow stunt associated virus (SYSaV), a nanovirus infecting *Sophora alopecuroides*, a leguminous fodder crop, was reported. During recent years a number of nanoviruses have been described from Central- and Northern Europe. This situation is exemplified by Pea necrotic yellow dwarf virus (PNYDV) from Germany, Denmark, the Netherlands, Austria, Hungary and Serbia, Black medic leaf roll virus (BMLV) from Austria, Sweden and Azerbaijan, Pea yellow stunt virus (PYSV) from Austria, and Cow vetch latent virus (CvLV) from France. The most recently identified nanovirid is Parsley severe stunt associated virus (PSSaV) from Germany, the only nanovirus potentially restricted to a non-leguminous host. No nanovirids have as yet been found in the Americas. All available evidence suggests that nanovirids are quite variable in their biological and molecular properties.

Particle Properties

Virions of the nanovirids are not enveloped, 17–20 nm in diameter, and presumably of an icosahedral $T = 1$ symmetry structure containing 60 subunits. Capsomeres may be evident, producing an angular or hexagonal outline (Fig. 1). Virions are stable in Cs_2SO_4 but may not be stable in CsCl . The buoyant density of virions is about 1.24–1.30 g/cm^3 in Cs_2SO_4 , and 1.34 g/cm^3 in CsCl . They sediment as a single particle component in sucrose rate-zonal and isopycnic Cs_2SO_4 density gradients.

Virions have a single capsid protein (CP) of about 19 kDa. No other proteins have been found associated with virions. Each ssDNA component is encapsidated in a separate particle. Virions are strong immunogens. Most nanovirid species are serologically distinct from one another.

Genome Organization

Based on DNA sequence analyses of a range of geographical isolates there is evidence that six and eight DNAs form the genome of a babu- and nanovirus, respectively (Table 1). Both encapsidated genomic nanovirid DNAs and nanovirid-associated DNAs (*Alphasatellitidae*) are covalently closed ssDNA circles of about 1 kb in size. A characteristic feature of all nanovirid and nanovirid-associated DNAs are two short (8–13 nt) inverted repeat sequences flanking a stretch of 11–13 nt containing a conserved nonanucleotide motif: TATTATTAC in babuviruses and TAGTATTAC in nanoviruses (exception: PSSaV, which has TATTATTAC). These sequences have the potential to form a stem-loop structure in ssDNA or a cruciform extrusion in dsDNA. Comparable DNA-sequence or DNA-structure motifs, also referred to as stem-loop or hairpin-loop, are common in all ssDNA viruses that multiply their genome by rolling-circle replication (see below). Sequences 5' and 3' of the inverted repeats are also conserved among the individual genome components, and their respective extensions in both directions are quite variable depending on the virus. This region has been designated common region stem-loop (CR-SL). In extreme cases sequences of up to 160 nt (PSSaV) or 250 nt (MDV), located 3' of the inverted repeat are conserved among individual genomic DNAs.

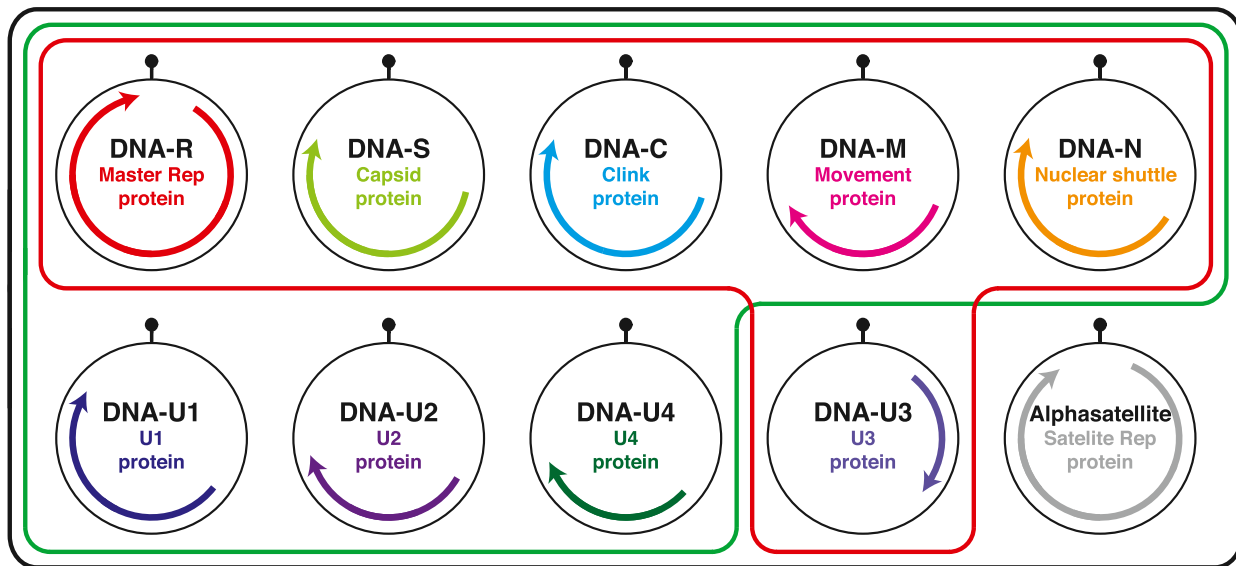


Fig. 2 Genome organization of nanovirids and associated alphasatellites. Individual genomic DNAs are represented as circles. A knob, often also referred to as hairpin, indicates the position of the inverted repeat sequences flanking the conserved nonanucleotide at the origin of viral strand replication. Component designations are displayed inside each circle. Arrows represent coding sequences of the respective encoded proteins. Protein names are given below DNA component names. Constituents of the babuvirus genome are framed in red and those of the nanovirus genome in green. DNAs of the top row have homologs in viruses of both genera; DNAs of the bottom row are unique for either nano- or babuviruses as indicated by the respective frame. Alphasatellites differing in numbers and types are frequently found associated with members of each genus of the *Nanoviridae*, hence included in the black frame. Graphics courtesy of I. Grigoras.

Apart from the CR-SL both babu- and nanoviruses possess a second conserved region that is common to all genomic DNAs. In babuviruses it is located 5' of the inverted repeats at distances that vary from about 50 nt to about 350 nt. The region is referred to as common region major (CR-M) and has a length of around 60 nt in ABTV, 65–92 nt in BBTV and 75–80 nt in CBDV. There is no such rather uniform CR-M in nanoviruses. Instead, short stretches of conserved sequences from 17 to 34 nt constitute a so-called common region II (CR-II), which is located about 70 nt 5' of the CR-SL. In some cases (e.g., FBVLV) there is no CR-II, and in others (e.g., MDV and PNYDV) rather long sequence stretches 5' of the inverted repeats are conserved between all genomic DNAs, except DNA-R. The genomic DNAs of nanoviruses are about 100 nt shorter than those of babuviruses and the coding region of DNA-R terminates closely 5' of the CR-SL. This restricts the possibility for longer common conserved sequences due to constraints dictated by aa conservation. Hence only short CR-II sequences are common to all genomic nanovirus DNAs whereas longer CR-II stretches may occur between nanovirus genomic DNAs other than DNA-R.

The genomic information of each nanovirid is distributed over six or eight different molecules of the aforementioned ssDNA type (Table 1). Upon infection, the circular ssDNAs are converted into circular double-stranded DNAs (mini-chromosomes) that serve as templates for gene expression. Each of these DNA molecules contains a single open reading frame of positive (virion-sense) polarity, preceded by a promoter sequence with a TATA box and followed by a polyadenylation signal (Fig. 2).

The fact that a typical set of respectively six or eight distinct DNAs has been consistently identified from babu- and nanoviruses suggests that the babuvirus genome consists of six DNAs and the nanovirus genome of eight DNAs. DNA-R, -S, -C, -M and -N have been identified from all nanovirids (Fig. 2, Table 1). DNA-U1, -U2 and -U4 homologs are typical for all nanoviruses but are absent from babuvirus genomes. In contrast, DNA-U3 is restricted to babuviruses and is absent from nanovirus genomes.

Transcript analyses of BBTV demonstrated a single virion-sense transcript per genome component. A minor transcript was found internal of the DNA-R ORF. For nanoviruses only transcripts of FBNYV and SCSV DNA-R and FBNYV DNA-C, along with a FBNYV-associated alphasatellite (see below) transcript have been mapped in detail. They are 3'-polyadenylated and their 5' ends map shortly after the TATA sequences preceding the respective ORFs. Consistent with these transcript analyses is the notion that each nanovirid DNA encodes a single gene defined by the respective ORFs and their flanking transcription initiation and termination signals. Interestingly the DNA-R transcripts of the two nanoviruses studied, FBNYV and SCSV, are terminally redundant by 54–161 nt (FBNYV) and 49 nt (SCSV). Considering the positions of TATAA box- and polyadenylation sequences of DNA-R molecules of all nanoviruses, such a terminal redundancy is predicted to be a unique and distinctive feature of nanovirus DNA-R components in general. Its biological significance remains, however, unknown.

Although eight DNAs (DNA-R, -S, -C, -M, -N, -U1, -U2, and -U4) had invariably been found associated with nanovirus infections, the question of 'how many DNAs constitute a nanovirus genome' had remained open for some time until experimental infection using individually cloned FBNSV DNAs became possible. In this way, the type of disease symptoms induced as well as the efficiency of vector transmission of the thus reconstituted virus could be analyzed. Due to this considerable technical accomplishment it was demonstrated that without DNA-R, DNA-S and DNA-M no infection was possible. DNA-C, DNA-U1, and DNA-U2 were required for

efficient disease development and vector transmission but not absolutely essential. Interestingly, infections devoid of DNA-N resulted in a disease phenotype indistinguishable from infections with wild-type virus. However, neither cowpea- nor pea aphids, the most efficient nanovirus vectors, were able to transmit virions produced in DNA-N-less infections. Hence DNA-N encodes the transmission helper factor postulated earlier. Finally, DNA-U4 proved to be non-essential for infection and sustained virus transmission under experimental conditions. Although these results were obtained with FBNSV in faba bean, it is very likely that they are equally valid for other nanoviruses. Experimental infection using cloned genome DNAs has been successfully demonstrated also for BMLRV, FBNYV and PNYDV. To date, no comparable experimental infection with DNAs of any babuvirus has been achieved.

There are indications that the genome organization of nanoviruses can sometimes be even more complex. In infections by each of the recently discovered SYSAV and PSSaV four DNA-R-like components were identified. However, which one of them encodes a Rep protein capable of acting as a master-Rep protein, i.e., catalyzing replication initiation of the other genomic DNAs, has not been determined. These findings illustrate well that the current picture of nanovirus genome organization may still be in flux.

Properties and Functions of Nanovirid Proteins

In addition to the capsid protein (CP) of about 19 kDa encoded by DNA-S, at least 5–7 non-structural proteins are encoded by the other 5 or 7 genomic DNAs of nanovirids (Table 1, Fig. 2). Besides acting as the structural protein required for virion formation, the protein encoded by DNA-S of BBTV has also been identified as a silencing suppressor.

One of the best-studied nanovirid proteins is the ca. 33 kDa replication initiator protein (Rep) encoded by their DNA-R. It belongs to the vast protein family of rolling-circle replication (RCR) initiator proteins of ssDNA viruses, bacteriophages, RCR plasmids and RCR-transposable elements (helitrons). Nanovirids express a unique Rep protein that is essential for replication initiation of all genome segments, six in case of babuviruses and eight in case of nanoviruses. This Rep protein recognizes specific sequence motifs in the so-called origin region, DNA stretches flanking the inverted repeat sequences that are shared by all genome components. By contrast to the Rep proteins encoded by DNA-R, those encoded by the nanovirid-associated alphasatellites are not able to trigger replication initiation of any nanovirid genome component. This led to the name ‘master replication initiator protein’ (M-Rep) for the DNA-R-encoded protein.

Several salient features characteristic of RCR-initiator proteins have been demonstrated also for a nanovirus (FBNYV) M-Rep protein, e.g., origin DNA cleavage and nucleotidyltransferase activity and ATPase activity. Also, the three-dimensional structure of the DNA-binding and endonuclease domain of FBNYV M-Rep has been determined, revealing a striking similarity to those of geminiviruses, circoviruses, parvoviruses, polyomaviruses and papillomaviruses. By analogy to geminivirus Rep proteins, nanovirus M-Rep proteins are probably also the smallest helicases of the superfamily 3 helicases (S3H), experimental proof however pending. Replication initiator and capsid proteins are, in addition to the so-called ‘stem-loop’ motif in the ssDNA, common elements shared by the very large group of so-called CRESS (circular Rep-encoding single-stranded DNA) viruses.

The third protein essential for nanovirid infection, apart from CP and M-Rep, is the 13–14 kDa protein encoded by DNA-M. Experimental evidence obtained for the protein encoded by DNA-M of BBTV indicates that it is involved in cell-to-cell movement. A stretch of 25–30 hydrophobic residues at its N-terminus is conserved in all its nanovirid homologs, consistent with a function as movement protein (MP). As shown for the BBTV CP, the MP of BBTV is also able to act as a suppressor of RNA silencing by reversing systemically established gene silencing.

The protein encoded by DNA-C is 19–20 kDa in size and has an N-terminal F-box sequence and the aa motif LxCxE characteristic for proteins binding to the cell-cycle regulator protein retinoblastoma (pRB). Moreover, the DNA-C encoded protein has been shown to interact with pRB. Therefore, the name “Clink” for ‘cell cycle link’ protein was coined for it.

Based upon evidence obtained for the activity of the BBTV DNA-N encoded 17–18 kDa protein it was proposed to act as a nuclear shuttle protein (NSP). In addition to that it has been shown that the NSP homologs of FBNSV and PNYDV are mandatory for virus transmission by vector aphids. Another biological property of a nanovirus NSP is its interaction with cellular proteins involved in stress granule formation, as shown for PNYDV. It is speculated that this way NSP may inhibit stress granule formation, a protective cellular response to virus infection.

Little is known about the biochemical properties or biological functions of the proteins encoded by DNA-U1, -U2, -U3, and -U4, where “U” stands for “unknown function”. As mentioned above, FBNSV DNA-U1 and DNA-U2 influence disease symptoms and the total amount of virions produced. It is plausible to assume that this is due to the action of the respective encoded proteins U1 (17–18 kDa) and U2 (14–15 kDa). The experimental disruption of a nuclear localization signal in the U1 protein altered its subcellular localization, yet no specific function was thus revealed for the U1 protein.

The U3 protein deduced from an ORF on DNA-U3 of BBTV is ca. 9 kDa in size. There are no data about its potential function. A transcript spanning the BBTV DNA-U3 ORF has been mapped. A homologous ORF is absent in the DNAs termed U3 of ABTV and CBDV. It remains, therefore, questionable whether these DNAs can be considered being true homologs of BBTV DNA-U3.

The deduced U4 protein encoded by DNA-U4 of the nanoviruses has a size of ca. 13 kDa. The fact that DNA-U4 homologs were detected in all nanovirus infections except those by PSSaV underpins the importance of the respective DNA-U4 product in natural infections. This notion appears valid despite the observation that DNA-U4 was not required for development of disease symptoms in experimental infections of faba bean seedlings with FBNSV. Moreover, FBNSV lacking DNA-U4 was serially transmitted to faba bean by aphids for >2 years. To date, only in planta protein-protein interaction data of bi-molecular fluorescence complementation assays employing PNYDV U4 protein are available and revealed auto-interaction and interaction with the PNYDV movement protein (MP).

Nanovirid Replication

The knowledge about geminivirus replication can largely be transposed also to nanovirid replication. Upon inoculation by aphid vectors virions enter host cells, presumably phloem parenchyma cells. As shown for BBTV and geminiviruses the ssDNAs are converted with the help of virion-encapsidated primers into double-stranded transcriptionally active mini-chromosomes. Most probably, this step occurs in the nucleus, but whether virions or only ssDNAs enter the nucleus is not known. Once a sufficient amount of M-Rep protein is available, a subunit of M-Rep cleaves the phosphodiester backbone of the origin DNA between nt T7 and A8 of the nonanucleotide TAG/TTATT-AC, becomes covalently linked to the 5' phosphate of A8 via its active site tyrosine (Y79 in case of FBNYV M-Rep), and a host DNA polymerase utilizes the 3'-OH of T7 as primer for DNA synthesis. After one round of replication, a second subunit of M-Rep cleaves the newly synthesized nonanucleotide sequence between T7 and A8, becomes linked to the 5' phosphate of the newly synthesized A, and the 3'-OH of T7 attacks the phosphotyrosyl ester bond between the first M-Rep subunit and the DNA, which leads to a nucleotidyltransfer of the M-Rep-linked 5' phosphate to the newly available 3'-OH. As a result, a phosphodiester bond is formed resulting in a liberated circular single-stranded DNA molecule, and a free tyrosine of the first M-Rep subunit, a step that requires no energy. These steps are repeated as long as rolling-circle replication continues. This so-called 'nicking and joining' activity by M-Rep proteins of DNA with the conserved nonanucleotide sequence has been experimentally demonstrated in vitro for BBTV and FBNYV as well as for one Rep of a FBNYV-associated alphasatellite. Replication of nanovirus DNAs is enhanced by the action of Clink, the nanovirid-encoded cell cycle modulator protein that is assumed to re-program the cell cycle of non-proliferating plant cells towards S-phase.

One interesting and quite unique characteristic of nanovirus replication has been discovered using FBNSV as a model: multiplication of each individual genomic DNA to a different and component-specific maximum level, the so-called genome formula. This genome formula differs between the two hosts faba bean (*V. faba*) and barrel medic (*Medicago truncatula*) and is also different in its vector, the pea aphid *A. pisum*.

Taxonomy and Phylogeny

Nanovirids share a set of five homologous DNA components, referred to as DNA-R, -S, -C, -M and -N (Fig. 2). Three other DNAs (DNA-U1, -U2 and -U4) encoding proteins of as yet unknown functions have been identified from nanoviruses and one further DNA (DNA-U3) only from babuviruses. Viruses of the individual species typically have narrow host ranges. Under natural and experimental conditions all nanoviruses only infect dicotyledonous plants. Most nanoviruses have largely overlapping host ranges and predominantly infect a range of leguminous species (*Fabaceae*) and a few other dicotyledonous species (Table 1), whereas babuviruses have been reported only from a few monocotyledonous species, such as the *Musaceae* or *Zingiberaceae*. Since these plants are colonized by a restricted number of (often monophagous) aphid species, the nanoviruses studied for their biological properties are transmitted by the cowpea aphid *Aphis craccivora*, the pea aphid *Acyrtosiphon pisum* and some other aphid species predominantly feeding on legumes. The vector of PSSaV, a nanovirus identified from parsley, a non-leguminous host, is unknown. The currently known babuviruses are vectored by the banana aphid *Pentalonia nigronervosa* and *Micromyzus kalimpongensis*.

The other major differences between babu- and nanoviruses are that (1) nanovirus DNA components range in size from 962 to 1045 nt and are thus slightly smaller (by ca. 100 nt) than those of babuviruses (1013–1116 nt), (2) babuviruses are serologically unrelated to members of the genus *Nanovirus*, and (3) nanoviruses share low levels of aa sequence identities ranging from 14% to 58% in individual genes with babuviruses. Moreover, nanovirus DNA-R transcripts are terminally redundant unlike those of the babuviruses.

Phylogenetic analyses of the nt sequences of nano- and babuviruses demonstrated (Fig. 3) that the five babuvirus sequences representing three species form a cluster clearly distinct from the nanoviruses. The latter consist of a group of eight related species from Asia, Europe and North Africa and three more distantly related species each from Australia (SCSV), Iran (SYSaV) and Europe (PSSaV).

Comparison of the deduced aa sequences of the five proteins encoded by DNA-R, -S, -C, -M, and -N by representative isolates of the nanovirid species listed in Table 1 shows that the master Rep protein (M-Rep) is the most conserved protein. It has an overall aa identity of 80%–97% between members of the same genus and 54%–58% between the two genera. The second most conserved protein is the NSP with 55%–91% intra-genus aa identity and 41%–50% inter-genus aa identity. Capsid protein, MP and Clink protein follow with maximum intra-genus aa identities of 88%, 81% and 76%, respectively. The sequence conservation of these proteins between babu- and nanoviruses is lower than that of M-Rep and NSP with maximum sequence identities of 33% (MP), 27% (CP) and 25% (Clink), respectively.

If one considers intra-genus variability of the respective proteins there is no significant difference in variability among homologous nanovirus proteins and among homologous babuvirus proteins. The variability of nanovirus proteins appears slightly higher, but this is probably due to the greater number of nanovirus genomes sequenced, including one from parsley, a non-leguminous species. With respect to the currently used criteria for species discrimination in the family, the reader is referred to the latest ICTV description of the family *Nanoviridae*.

Based on similarities in particle morphology and size and based on the identification of an alphasatellite-like DNA (1.3 kb), Coconut foliar decay virus (CFDV), the causal agent of a severe disease of coconut palms in Vanuatu, is listed as an unassigned species of the family *Nanoviridae*. However, following rolling-circle amplification and deep sequencing analyses of historic and fresh samples twelve distinct circular DNAs have recently been identified. Sequence analysis suggested that nine of the DNAs represent alphasatellites (Fig. 4) and that the CFDV genome is possibly tripartite, consisting of an alphasatellite-like DNA (1.3 kb), a 1.3 kb DNA encoding a capsid protein related to that of grabloviruses (*Geminiviridae*), and a unique DNA-gamma (641 nt) that



Fig. 3 Maximum likelihood phylogenetic tree of nanovirid genomes. Individual DNAs of homologous genome components were concatenated in the order DNA-R, -S, -C, -M, -N and aligned using Muscle. The tree was constructed in MEGA7 using the GTR + G + I nt substitution model and is mid-point rooted. The scale bar corresponds to 0.1 substitutions per site. Percent bootstrap support of 5000 non-parametric replicates is shown. Branches below 50% support are collapsed.

may encode a movement protein. Thus, CFDV differs clearly from established members of the family *Nanoviridae* not only in genomic organization, DNA sequence as well as in the type of insect vector (planthopper versus aphids), implicating that CFDV possibly represents a member of a new family of plant viruses.

Nanovirid Evolution

Virus genomes that consist of single-stranded nucleic acids are known to undergo rapid sequence evolution. This is also true for nanovirids as has been shown for FBNSV. It has a rate of 1.8×10^{-3} substitutions per nt per year, the highest substitution rate among ssDNA viruses and is only outrivalled by those of influenza A viruses and human immunodeficiency viruses. Any potential disadvantage caused by such an elevated nt substitution rate is probably compensated for by the fact that every type of genome component acts individually as a quasi-species-like swarm of molecules.

Also, reassortment of individual genomic DNAs or groups thereof and recombination between them would alleviate the genetic load imposed by such a high nt substitution rate. Indeed, both recombination and reassortment of genome component has been observed in nanovirids, for instance BBTv and a range of nanoviruses. In fact, given the extent of common sequences shared by several genomic DNAs of the hitherto identified nanoviruses, recombination between them is probably frequent during replication.

Interestingly, each type of nanovirus genomic DNA does not need to be present simultaneously in an individual cell of a host plant to trigger a productive infection, as has been shown recently for FBNSV. This finding led to the assumption that nanovirus infection, exemplified by FBNSV, does not necessarily need to operate at the level of individual cells but rather at the level of an organism, here a plant host. Such a behavior, coined “multicellular way of life”, is unique for multipartite viruses and represents a novel concept in virology. A multicellular way of life would alleviate the problems caused by the very high multiplicity of infection required for a virus with an octopartite genome if it were productively operating only at the level of a single cell. This observation also fits with the proposition that the de novo origin of multipartite ssDNA viruses may postdate the emergence of multicellular eukaryotes.

Nanovirid-Associated Alphasatellites

In addition to the genomic DNAs of nanovirids, a large number of additional DNAs encoding Rep proteins have been found associated with almost all nanovirid infections. These Rep-encoding DNAs are quite diverse and phylogenetically distinct from any nanovirid DNA-R, however related to Rep-encoding DNAs associated with begomoviruses and CFDV (Fig. 4). Collectively, these DNAs are now referred to as alphasatellites (*Alphasatellitidae*). Using conventional approaches of sequence analysis, nanovirus-associated alphasatellites were often detected prior to the essential nanovirid DNA-R components. This is explained by the fact that these alphasatellites are much more abundant than any genomic DNA of the nanovirus they associate with, as was revealed by

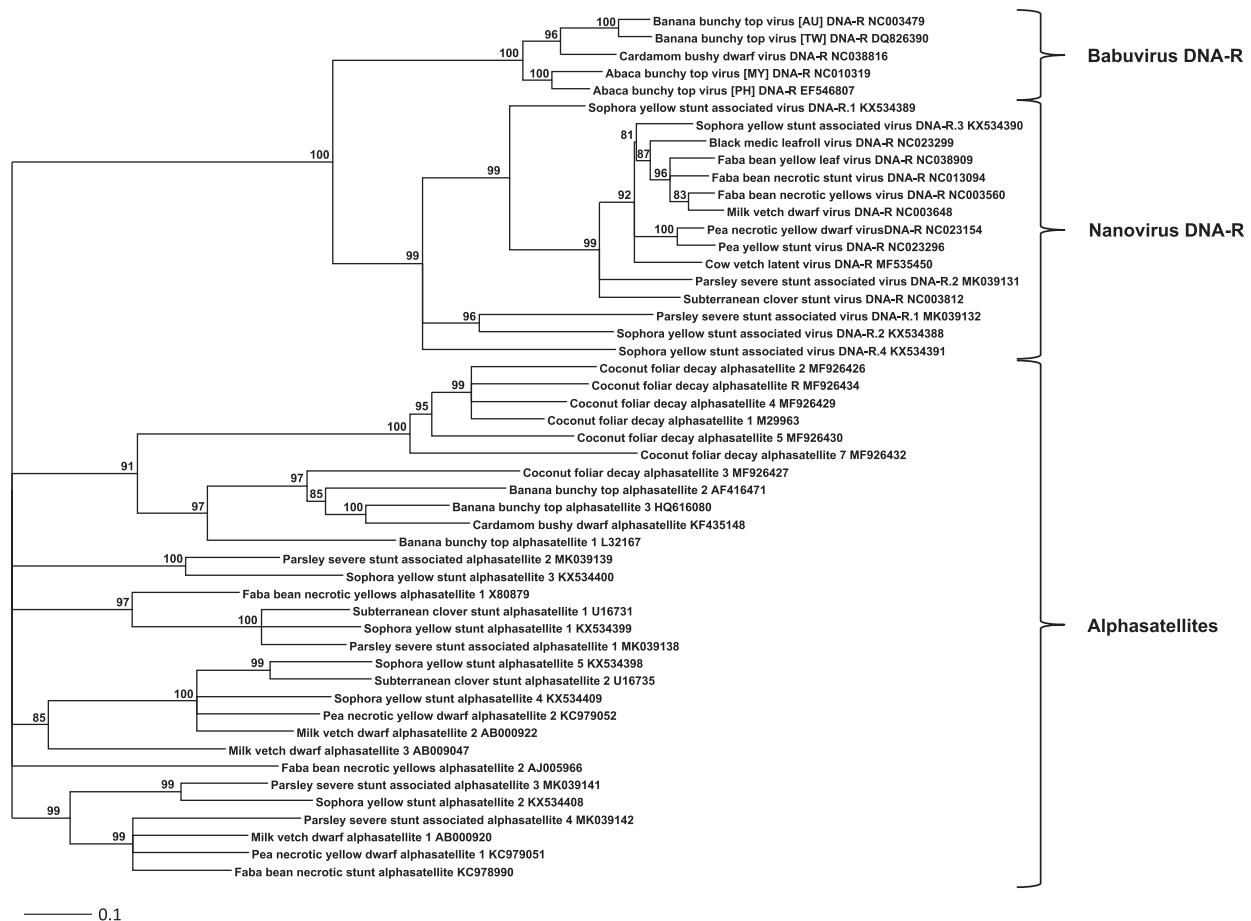


Fig. 4 Maximum likelihood phylogenetic tree of nanovirid DNA-R components and nanovirid-associated alphasatellites. DNAs were aligned using Muscle and a mid-point rooted tree was constructed in MEGA7 using the GTR + G + I nt substitution model. The scale bar corresponds to 0.1 substitutions per site. Percent bootstrap support of 1000 non-parametric replicates is shown. Branches below 80% support are collapsed.

deep sequencing analyses. Deep sequencing also permitted the detection of up to seven or fourteen distinct alphasatellites in single plants infected with PSSaV or SYSaV, respectively. Begomovirus- and CFDV-associated alphasatellites are larger (ca. 1300 nt) than nanovirid-associated alphasatellites (ca. 1000 nt) due to the inclusion of an A-rich sequence within their non-coding region. Nanovirus-associated alphasatellites are individually encapsidated in nanovirid-like particles. Proof of their encapsidation stems from the first sequenced FBNYV-associated Rep-encoding DNA and was recently also provided by sequencing DNA extracted from CFDV virions. In contrast to the activity of the M-Rep proteins encoded by DNA-R, alphasatellite-encoded Rep proteins are only capable of initiating replication of their cognate DNA, as shown for FBNYV- and BBTV-associated alphasatellites. Despite the fact that almost all nanovirid infections are accompanied by varying numbers of alphasatellites, little is known about their influence on the pathogenicity or epidemiology of the respective helper viruses. In this context it is noteworthy that two begomovirus-associated alphasatellites were shown to act as silencing suppressors.

Recently alphasatellites were classified in the newly established family *Alphasatellitidae*, subdivided in the subfamilies *Geminialphasatellitinae* (five genera) and *Nanoalphasatellitinae* (seven genera).

Transmission and Host Range

Tissue Tropism and Means of Transmission

SCSV and FBNYV have been shown to replicate in inoculated protoplasts. All assigned viruses are restricted to the phloem tissue of their host plants and are not transmitted mechanically and through seeds. Apart from graft transmission, vector transmission had been the only means of experimentally infecting plants with nanovirids for a long time, until infectivity of purified FBNYV virions by biolistic bombardment was demonstrated. Meanwhile, vector transmission of virus derived from experimental infections of various hosts by cloned nanovirus genome components is possible for BMLRV, FBNSV, FBNYV, and PNYDV, and has been used to identify the virus-encoded helper factor of aphid transmission.

Transmission by Aphids

Under natural conditions, all nanovirids are transmitted by certain aphid species, in which they can persist for many days or weeks without replicating in their vectors. At least 10 aphid species belonging to the tribes *Macrosiphini* and *Aphidini* were reported as vectors of nanovirids. Whereas only one aphid species, the banana aphid (*Pentalonia nigronervosa*) and *Micromyzus kalimpongensis* has been reported as respective vector of BBTV and CBDV, several aphid species transmit FBNYV, MDV, and SCSV. The cowpea aphid (*Aphis craccivora*) appears to be the major natural vector of these viruses as it is the most abundant aphid species on legume crops in the afflicted areas and was among the most efficient vectors under experimental conditions. Other aphid vectors of FBNYV are the black bean aphid (*Aphis fabae*) and the pea aphid (*Acyrtosiphon pisum*). SCSV has been reported to be vectored also by the cotton aphid (*Aphis gossypii*), the green peach aphid (*Myzus persicae*) and the potato aphid (*Macrosiphum euphorbiae*), and PNYDV by the cowpea- and pea aphid and the vetch aphid (*Megoura viciae*). The vector aphids of ABTV, FBVLV, PYSV, SYSaV, and PSSaV have not been determined (Table 1).

Transmission studies showed that aphids are able to transmit BBTV, FBNYV and SCSV following short acquisition and inoculation access feeding periods of about 30 min each. Although viruliferous aphids often retain transmission ability for life, nanovirids do not multiply in their insect vectors. Together with the observation that longer acquisition and inoculation access feeding periods resulted in higher transmission rates, the long persistence of nanovirids in their insect vectors indicates that they are transmitted in a circulative persistent manner similar to that of luteoviruses. In contrast to observations for luteoviruses, however, immunofluorescence localization of BBTV in the banana aphid *Pentalonia nigronervosa* indicates that BBTV antigens occur in cells of the anterior midgut and the principal salivary glands. This translocation resembles that of the geminiviruses, the other family of plant ssDNA viruses.

For FBNYV it has been demonstrated that purified virions alone are not transmissible by its aphid vector, regardless of whether they are acquired from artificial diets using membrane feeding techniques or directly microinjected into the aphid's hemocoel. However, faba bean seedlings biolistically inoculated with intact virions or viral DNA developed symptoms typical of FBNYV infections and were efficient sources for FBNYV transmission by aphids. These observations together with results from complementation experiments suggest that FBNYV (and other nanovirids) require a virus-encoded helper factor for its vector transmission that is absent from purified virion preparations. Using the individually cloned DNAs of FBNSV and PNYDV, it has recently been demonstrated that the proteins encoded by DNA-N of FBNSV and PNYDV are mandatory for virus transmission by vector aphids. This sets the nanovirids clearly apart from other viruses transmitted in a circulative manner, such as geminiviruses and luteoviruses. The uniqueness of nanovirids is also substantiated by the observation that the so-called genome formula of FBNSV in its vector, the pea aphid *A. pisum*, differs from those in two host plants, faba bean (*V. faba*) and barrel-medick (*Medicago truncatula*).

Host Range

Nanoviruses often have largely overlapping host ranges by infecting over 50 legume species and only a few non-legume species under experimental and natural conditions. FBNYV, MDV and SCSV naturally infect a range of leguminous species, whereas BBTV and ABTV have been reported only from *Musa* species and closely related species within the *Musaceae*, such as abaca (*M. textilis* Née) and ensete (*Ensete ventricosum* Cheesem). There are no confirmed non-*Musa* hosts of BBTV. Symptoms of BBTV include plant stunting, foliar yellowing, and most characteristic dark green streaks on the pseudostem, petioles, and leaves. CBDV has been reported from black cardamom (*Amomum subulatum*) and green cardamom (*Elettaria cardamomum*), two closely related species of the family *Zingiberaceae*. All economically important natural hosts of FBNYV, MDV, and SCSV are legumes, in which these viruses generally cause plant stunting and a range of foliar symptoms, such as leaf deformations and chlorosis or reddening. Although FBNYV infects >50 legume species and only few non-legume species (*Stellaria media*, *Amaranthus* and *Malva* spp.) under experimental and natural conditions, major legume crops naturally infected by FBNYV are faba bean, lentil (*Lens culinaris*), chickpea (*Cicer arietinum*), pea (*Pisum sativum*), French bean (*Phaseolus vulgaris*) and cowpea (*Vigna unguiculata*). Likewise, SCSV experimentally infects numerous legume species, but its economically important natural hosts only include subterranean clover, common bean, faba bean, pea, and medics. In addition to *Sophora alopecuroides*, SYSaV was found in Iran in a few other legumes, such as milk vetch (*Astragalus* sp.), chickpea, licorice (*Glycyrrhiza glabra*), and lentil but also in a non-leguminous plant, wild rue (*Peganum harmala*; *Nitrariaceae*). MDV is known to cause yellowing and dwarfing in Chinese milk vetch (*Astragalus sinicus* L.), a common green manure crop in Japan, as well as in cowpea, faba bean, pea, and soybean. The host range of MDV appears to be larger than that of most other nanoviruses as a number of non-leguminous plants have been reported as experimental hosts of MDV, and natural infections by MDV have recently been recorded from tobacco (*Nicotiana tabacum*; *Solanaceae*) in China and papaya (*Carica papaya*; *Caricaceae*) and periwinkle (*Catharanthus roseus*; *Apocynaceae*) in Bangladesh. This and the recent identification of a novel type of nanovirus from parsley (*Petroselinum crispum*; *Apiaceae*) in Germany may indicate that there are further nanovirids which naturally infect several other non-leguminous plants.

Epidemiology and Control

Based on symptomatology and transmission characteristics, luteoviruses were earlier suspected as the causative agents of several diseases now known to have a nanovirid etiology. Unlike luteoviruses, however, nanovirids generally cause more severe

symptoms. In many nanovirid-host combinations, early infections lead to very severe effects and even premature plant death. The disease caused by BBTV is considered the most serious viral disease of banana worldwide. SCSV and FBNYV are also of considerable economic importance as they have caused serious diseases of subterranean clover in Australia and of faba bean in Egypt, respectively, leading to repeated crop failures. The same applies to PNYDV, which has occurred in certain years (e.g., 2016) at an epidemic scale in central European countries where it caused considerable yield losses in legume crops such as green pea and faba bean. Since PNYDV seems to have a wide geographic distribution in Central Europe, it poses a potential threat to legume production in Europe. This threat might be increased by future changes in climatic conditions that favor survival of aphid vector populations during milder winters in Central Europe and thus facilitate the spread of PNYDV. The economic importance and control of other nanovirids such as ABTV, FBYLTV, PYSV, SYSAV and CvLV cannot be commented on as they were identified from single samples and the biological properties of these viruses have not been studied so far.

Disease control is based on the results from ecological and epidemiological studies. Since nanoviruses and legume luteoviruses are transmitted by almost the same aphid species, are phloem-limited in their host plants, cause nearly identical symptoms and are not seed-transmitted, they appear to have similar ecologies. Because of this similarity and because luteoviruses have been better studied than nanoviruses, information generated in relation to the control of luteoviruses are useful in the development of control strategies for nanoviruses. Control measures can be subdivided into three major categories: (1) those aimed directly at the virus, (2) those aimed directly at the vector, and (3) integrated approaches, which combine all possible components in a way to complement each other and be applied as one control package.

Since virus control is a complex topic, only two examples of virus control, one for a babuvirus (CBDV) in a perennial crop and the other for a nanovirus (FBNYV) in an annual crop are briefly presented. The control of BBTV is dealt with in a separate article (see BBTV (*Nanoviridae*)).

CBDV causes foorkey disease of large cardamom by inducing a remarkable reduction in size of leaves of the infected plants, stimulating proliferation of a large number of stunted shoots arising from the rhizome, transforming the inflorescence into leafy vegetative parts and suppressing fruit formation. CBDV-infected plants remain unproductive and gradually degenerate. The primary spread of disease from one area to another is through infected rhizomes and further spread within the plantation is by aphids. For the management of foorkey disease a regular rogueing of diseased plants by uprooting and immediately destroying them is recommended. As preventive measures, cardamom growers should use CBDV-free planting material preferably seedlings, avoid suckers as planting material from diseased areas and avoid keeping nurseries in the vicinity of infected plantations.

In countries of West Asia and North Africa, FBNYV can occur at a high incidence and cause severe yield losses in cool-season legumes, such as chickpea, faba bean, lentil and pea. A significant decrease in disease incidence can be expected if sources of infection are eliminated from within or near crops. In practice, this is not easy and it is rather impossible to eliminate all weed hosts of FBNYV. Prior to advocating such an approach for FBNYV control, more information on the relative importance of wild legumes and weeds as sources of FBNYV infection is needed. Severe epidemics of FBNYV can be prevented by proper timing of sowing dates, avoiding peak aphid population and by spraying the fields with appropriate insecticides once or twice during the period when viruliferous aphids introduce FBNYV to the crop. Another measure to prevent further spread and to decrease disease incidence is to eliminate or reduce primary infection foci as sources of infection from within fields. Since sources of resistance to FBNYV have been identified in wild legume species a classical breeding approach might be used for incorporating resistance into established cultivars of some of the aforementioned cool-season legumes.

Diagnosis

Nanovirids can be detected and identified by a range of methods, such as electron microscopy (EM), serology, polymerase chain reaction (PCR), isothermal rolling-circle amplification (RCA) and various sequencing techniques. Because of the small size of the nanovirid particles, which can be easily confused with cell constituents, EM (i.e., use of leaf extracts for adsorption preparates) is not the method of choice for nanovirid detection. The virions can only be visualized following particle purification and by immunosorbent EM using nanovirid-specific antisera for coating of EM grids.

Most nanovirid species are serologically distinct from one another. BBTV and the two other babuviruses are serologically unrelated to members of the genus *Nanovirus*. However, antisera and some monoclonal antibodies (MAbs) to BBTV cross-react strongly with ABTV and CBDV. Therefore, BBTV antibodies have been used for the detection of ABTV and CBDV, the two other babuviruses. Of the 10 MAbs raised to BBTV, only two reacted with ABTV. This is consistent with a CP aa sequence difference of about 20% between ABTV and BBTV. On the contrary, antisera to FBNYV and SCSV cross-reacted weakly with SCSV and FBNYV, respectively, in Western blots and immunoelectron microscopy but not at all in double-antibody sandwich enzyme-linked immunosorbent assays (DAS-ELISA), indicating that the serological relationship between these two nanoviruses is remote. However, MDV antigen reacts strongly not only with an antiserum to FBNYV in DAS-ELISA but also with the majority of MAbs to FBNYV in TAS-ELISA, indicating a close serological relationship between FBNYV and MDV (CP aa sequence identity of about 83%). Some of the FBNYV MAbs also allowed detection and differentiation of PNYDV, PYSV, and BMLRV, indicating that these nanoviruses share certain epitopes with FBNYV. However, species-specific MAbs are indispensable for specific detection of all nanovirids by serological means.

Under routine conditions and for the analysis of large numbers of samples, nanovirids are typically tested in TAS-ELISA using a cross-reactive antiserum (or a mixture of antisera) for coating of plates and a mixture of broad-spectrum MAbs for antigen detection.

Samples that give positive reactions can then be used for aphid transmission experiments, PCR analysis and various sequencing techniques. After the initial screening of samples using TAS-ELISA, PCR or immuno-capture PCR using nanovirus component-specific primers have been used for the molecular characterization of nanovirids. Also, isothermal multiple displacement amplification (MDA) or RCA using Phi29 polymerase has become a widely used technique to enrich nanovirus DNAs in DNA preparations from tissue samples or even ancient particle preparations. For instance, the multiple DNAs associated with coconut foliar decay disease were uncovered this way from virion samples stored for about 30 years.

At present, RCA- or PCR-amplified DNA is routinely used for deep sequencing analyses and has yielded considerable insights into the diversity of known nanovirids and has also led to the discovery of PSSaV, a novel and quite different nanovirus.

Concluding Remarks

Apart from graft transmission and biolistic bombardment with purified virions, vector transmission had been the only means of experimentally infecting plants with nanovirids for a long time, until infectivity of cloned nanovirus DNA was demonstrated for FBNSV and fully infectious and aphid transmissible virus could be reconstituted this way. Due to this technical achievement functional analyses of individual nanovirus genes have become possible. Meanwhile infection using cloned genome DNAs has been successfully demonstrated also for BMLRV, FBNYV and PNYDV and allowed the application of reverse genetics to studying also nanoviruses, which was not possible before. As a consequence, the experimental independence of vector transmission has substantially advanced nanovirus research during the last years and yielded new and unexpected findings about nanovirus biology. In particular, the recent discovery of a ‘multicellular lifestyle’ of these multipartite ssDNA plant viruses represents a remarkable feat that would have been impossible without reverse genetics.

Further Reading

- Aronson, M.N., Meyer, A.D., Györgyey, J., *et al.*, 2000. Clink, a nanovirus encoded protein binds both pRB and SKP1. *Journal of Virology* 74, 2967–2972.
- Chu, P.W.G., Helms, K., 1988. Novel virus-like particles containing circular single-stranded DNA associated with subterranean clover stunt disease. *Virology* 167, 38–49.
- Franz, A.W., van der Wilk, F., Verbeek, M., Dulleman, A.M., van den Heuvel, J.F., 1999. Faba bean necrotic yellows virus (genus *Nanovirus*) requires a helper factor for its aphid transmission. *Virology* 262, 210–219.
- Grigoras, I., Timchenko, T., Katul, L., *et al.*, 2009. Reconstitution of authentic nanovirus from multiple cloned DNAs. *Journal of Virology* 83, 10778–10787.
- Grigoras, I., Vetten, H.J., Commandeur, U., *et al.*, 2018. Nanovirus DNA-N encodes a protein mandatory for aphid transmission. *Virology* 522, 281–291.
- Gronenborn, B., Randles, J.W., Knierim, D., *et al.*, 2018. Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded DNA virus representing a new taxon. *Scientific Reports* 8, 5698.
- Koonin, E.V., Dolja, V.V., Krupovic, M., 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 479–480, 2–25.
- Sicard, A., Yvon, M., Timchenko, T., *et al.*, 2013. Gene copy number is differentially regulated in a multipartite virus. *Nature Communications* 4, 2248.
- Sicard, A., Michalakakis, Y., Gutierrez, S., Blanc, S., 2016. The strange lifestyle of multipartite viruses. *PLoS Pathogens* 12, e1005819.
- Sicard, A., Piroles, E., Gallet, R., *et al.*, 2019. A multicellular way of life for a multipartite virus. *eLife* 8, 43599.
- Timchenko, T., de Kouchkovsky, F., Katul, L., *et al.*, 1999. A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. *Journal of Virology* 73, 10173–10182.
- Timchenko, T., Katul, L., Sano, Y., *et al.*, 2000. The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. *Virology* 274, 189–195.
- Vetten, H.J., Dale, J.L., Grigoras, I., *et al.*, 2012. Family *Nanoviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.C., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier/Academic Press, pp. 395–404.
- Vega-Rocha, S., Gronenborn, B., Gronenborn, A.M., Campos-Olivas, R., 2007. Solution structure of the endonuclease domain from the master replication initiator protein of the nanovirus faba bean necrotic yellows virus and comparison with the corresponding geminivirus and circovirus structures. *Biochemistry* 46, 6201–6212.
- Wanitchakorn, R., Hafner, G.J., Harding, R.M., Dale, J.L., 2000. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *Journal of General Virology* 81, 299–306.

Necro-Like Viruses (*Tombusviridae*)

Luisa Rubino, Institute for Sustainable Plant Protection, National Research Council, Bari, Italy

Giovanni P Martelli[†], University of Bari Aldo Moro, Bari, Italy

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)

BTE Barley yellow dwarf virus-like translational enhancer

CITE Cap-independent translational enhancer

CP Coat protein

CsCl Cesium chloride

kb Kilobase

kDa Kilo dalton

MP Movement protein

nt Nucleotide(s)

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

S_{20,w} Sedimentation coefficient

satRNA Satellite RNA

TED Translational enhancer domain

UTR Untranslated region

Glossary

Procedovirinae A subfamily in the family *Tombusviridae*, which RNA-dependent RNA polymerase is translated after the read through of a stop codon (from the Latin “*procedo*”, go ahead).

Read-through Read-through consists of the “translation” of a stop codon, such that the synthesis of a larger protein is

allowed. It is a common phenomenon in viruses, and it has a regulatory function.

Riboviria A higher taxonomic rank including all viruses possessing a RNA genome.

Tombusviridae A family of single-stranded positive sense RNA viruses within realm *Riboviria*.

Taxonomy and Classification

As the representative of a monotypic group, *Tobacco necrosis virus A* (TNV-A) was among the 16 groups of plant viruses described in 1971, and became the type species of the genus *Necrovirus* when it was established in 1995. Since 2012, the genus *Necrovirus* does not exist anymore as such, but phylogenetic analyses of the RNA-dependent RNA polymerase (RdRp) and movement proteins 1 and 2 (MP1 and MP2) supported the division of the former genus *Necrovirus* into two genera, *Alphanecrovirus* and *Betanecrovirus*. Due to the expression of their RNA-dependent RNA polymerase after translational readthrough of a stop codon, both genera belong to the subfamily *Procedovirinae* in the family *Tombusviridae* (realm *Riboviria*).

The genus *Alphanecrovirus* comprises four definitive member species, i.e., the type species *Tobacco necrosis virus A* (TNV-A), *Olive latent virus 1* (OLV-1), *Olive mild mosaic virus* (OMMV), and *Potato necrosis virus* (PoNV). The genus *Betanecrovirus*, besides the type species *Tobacco necrosis virus D* (TNV-D), includes two additional virus species: *Beet black scorch virus* (BBSV), and *Leek white stripe virus* (LWSV). *Chenopodium necrosis virus* (ChNV), formerly a member of the genus *Necrovirus* and for which sequence data are not available, is now an unassigned species in the subfamily *Procedovirinae*.

Virion Properties, Structure and Composition

Virus particles are very stable, resist temperatures in excess of 90°C, and chloroform, ether and non-ionic detergents. TNV-A virions sediment as a single component with a coefficient ($S_{20,w}$) of 118S and have buoyant density of 1.399 g ml⁻¹ at equilibrium in CsCl.

Icosahedral virions are approximately 28 nm in diameter, have angular profile and a capsid made up of 60 copies of a trimer consisting of three chemically identical but independent protein subunits (A, B, and C) stabilized by Ca²⁺ ions, arranged in a $T = 3$ lattice. Subunit size ranges from 24 to 27 kDa (BBSV and LWSV, respectively) to 29–30 kDa (other viral species). The capsid has a smooth appearance as protein subunits have an internal R domain and a shell (S), but lack the protruding domain proper of members of the majority of the other genera in the family *Tombusviridae* (Fig. 1).

Necro-like virions encapsidate one molecule of single-stranded, positive-sense RNA, neither capped at the 5' end, nor polyadenylated at the 3'-terminus, approximately 3.6–3.7 kb in size, constituting c. 19% of the particle weight.

Genome Organization and Expression

The monopartite genome contains four open reading frames (ORFs) which, in the order from the 5' to the 3' terminus, code for replication-associated proteins (ORF1 and ORF1-RT), movement proteins (ORF2 and ORF3), and the coat protein (CP) (ORF4). Some species possess a fifth ORF either located in the 3' terminal region (TNV-A) or in the middle of the genome (TNV-D^H and

[†]Deceased.

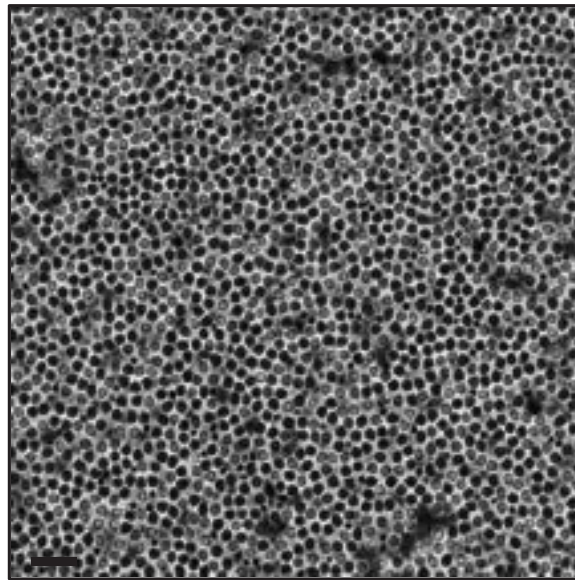


Fig. 1 Negatively stained TNV-D virions. Bar, 100 nm.

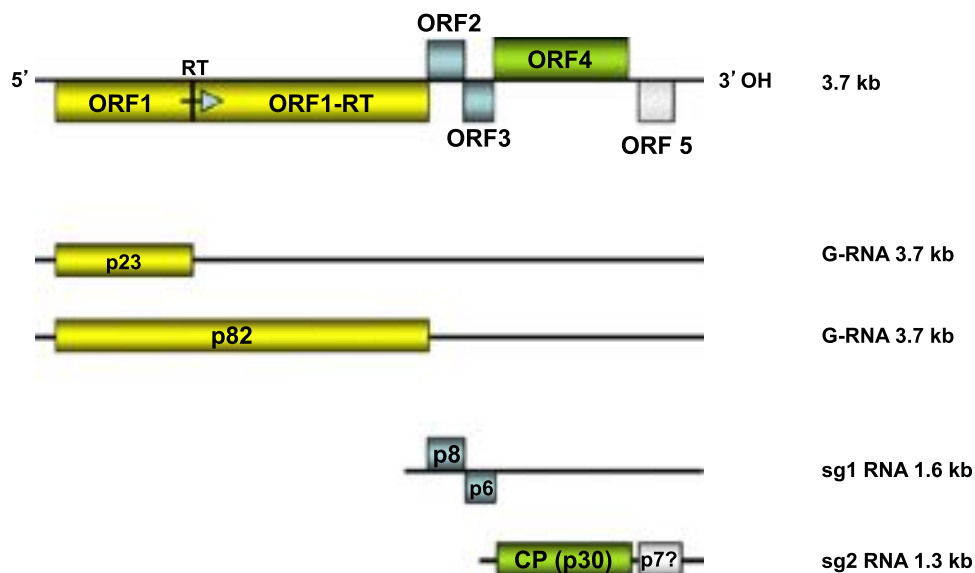


Fig. 2 Genome organization and expression of tobacco necrosis virus A. G-RNA, genomic RNA; sg1 RNA, sub-genomic RNA 1; sg2 RNA, sub-genomic RNA 2; RT, readthrough; CP, coat protein. Molecular weights of the ORF-encoded proteins (p) are indicated.

BBSV), partially overlapping the C-terminus of ORF1-RT and the N-terminal region of ORF2. The genome is very compact, having noncoding regions of limited size (Fig. 2). To compensate for the lack of a 5' cap or a 3' poly(A), RNA *cis*-elements are required for cap-independent translation of the virus genome. A cap-independent translational enhancer (CITE) has been identified in the 3'-end untranslated region (UTR) of all necro-like viruses, belonging to the Barley yellow dwarf virus (BYDV) like-element (BTE) class of CITEs. The BTE has a helical cloverleaf-like structure and contains a highly conserved nucleotide stretch able to form a stem-loop structure, and binds the eukaryotic translation initiation factor 4F (eIF4F) complex.

Genomic RNA acts as messenger for the translation of a protein of 22–24 kDa from ORF1. By translational readthrough of the UAG termination codon of ORF1, a protein of 82–83 kDa is synthesized (Fig. 2), which contains, in the readthrough portion, the conserved GDD motif of RNA-dependent RNA polymerases (RdRp). This protein, together with the expression product of ORF1, is indispensable for virus replication. RNA elements have been identified in TNV-D genome, which are able to modulate the readthrough efficiency.

Genes downstream ORF1-RT are expressed *via* the synthesis of two sub-genomic RNAs, 1.4–1.6 nt and 1.1–1.3 nt in size (Fig. 2). The larger sub-genomic RNA is responsible for the expression of ORFs 2 and 3, whereas the smaller sub-genomic RNA

constitutes the translational template for the expression of CP. ORFs 2 and 3 are two small, centrally located ORFs which, depending on the viral species, encode proteins of 7–8 and 6–7 kDa, respectively. Both of these proteins are involved in cell-to-cell transport but are dispensable for viral replication. LWSV ORF2 differs in size from that of other sequenced necro-like viruses, for it codes for a protein of 11 kDa. The additional central ORFs of TNV-D^H and BBSV code for a 7 kDa and a 5 kDa protein, respectively, which, like the products of ORFs 2 and 3, are necessary for cell-to-cell movement. A possible transmembrane motif has been predicted in OLV-1, OMMV and PoNV ORF3, which may have a membrane anchoring function.

ORF4 is the CP gene encoding a 24–30 kDa protein, the building block of the capsid. Necro-like virus CP was shown to be a multifunctional protein: besides encapsidation, it is required also for symptom modulation and efficient systemic virus spreading in the host and in the environment, since it has been reported to interact with the zoospores of the fungus *Olpidium brassicae*. Moreover, TNV-D CP was shown to modulate virus replication, since its deletion enhances virus RNA accumulation, whereas BBSV CP is able to interact with viral and host factors during virus replication. Silencing suppression ability was shown for OMMV CP, which acts coordinately with the ORF4 p6 product.

No experimental evidence is available for the expression of TNV-A ORF5.

Intraspecific Relationships

Necro-like viruses are serologically distinguishable from one another. Some are unrelated (e.g., LWSV with TNV-A and TNV-D), whereas others share antigenic determinants that result in weak relationships. Serological relationships are independent from whether viruses belong to genera *Alphanecrovirus* or *Betanecrovirus*. This is the case of OLV-1 (*Alphanecrovirus*) which is distantly related (serological differentiation index from 6 to 9) with both TNV-A (*Alphanecrovirus*) and TNV-D (*Betanecrovirus*), and of OMMV and PoNV (*Alphanecrovirus*) which are serologically related with TNV-D (*Betanecrovirus*).

The level of molecular similarity varies very much with the viral species. The RdRp sequence identity is high within the members of the genera *Alphanecrovirus* or *Betanecrovirus*, respectively, but low between alphanecroviruses and betanecroviruses. Thus, TNV-A, OLV-1, OMMV and PoNV share 86%–92% sequence identity with one another in the RdRp at the amino acid level. Likewise, TNV-D, BBSV and LWSV share 57%–69% sequence identity with one another, but sequence identity ranges from only 32%–36% between members of the two genera.

MP1 aa sequences show high identities among each other alphanecrovirus (66%–90%) and betanecrovirus (39%–49%), respectively, but only 6%–23% identity if alphanecroviruses are compared with betanecroviruses. Similarly, alphanecrovirus MP2 amino acid sequences share 95%–100% identity, but they show only 10%–26% identity with betanecrovirus MP2 sequences. Members of the genus *Betanecrovirus* share 39%–47% sequence identity with one another at the MP2 amino acid level.

The CP sequences of both alpha- and betanecroviruses are related and share 26%–86% sequence identity. OMMV (genus *Alphanecrovirus*), a putative recombinant between TNV-D and OLV-1, shows the highest CP sequence identity (86.2%) with that of TNV-D (genus *Betanecrovirus*), which may account for the cross reactivity with this virus, and 91.2% identity at the amino acid level with the RdRp of OLV-1.

Satellites

Satellites are subviral agents depending on the presence of a helper virus for replication, movement and transmission. Satellite nucleic acids are encapsidated by the helper virus encoded coat protein, whereas satellite viruses are able to encode for their own capsid protein.

Three satellite viruses are associated with necro-like viruses: Satellite tobacco necrosis virus (STNV)-1, STNV-2 and STNV-C. They all belong to the genus *Albetovirus* (from alphanecrovirus, betanecrovirus, tobacco) (realm *Riboviria*) and are classified as representative isolates of the species *Tobacco albetovirus 1*, *Tobacco albetovirus 2* and *Tobacco albetovirus 3*, respectively. STNV-1 and -2 are associated with the replication of TNV-A, whereas STNV-C is associated with TNV-D. The association of the satellite virus with its helper virus is specific, since no heterologous helpers have been reported to support STNV replication. STNV-1, -2 and -C are transmitted in association with TNV-A or -D by the fungus *Olpidium brassicae*, using the same mechanism as the helper virus.

The three STNV have isometric particles *c.* 17 nm in diameter, sedimenting as a single component with a coefficient of 50S, for which the crystal structure has been determined. The capsid is constructed with 60 identical subunits *c.* 21–22 kDa in size, arranged in a *T* = 1 lattice, and contains a single-stranded, positive-sense RNA molecule 1221–1245 nt in size. STNV RNA accounts for *c.* 20% of the particle weight, and comprises a single ORF encoding the CP. Similarly to the helper virus genome, STNV RNA is uncapped and lacks a 3′ poly(A) tail. The 5′ and 3′ untranslated regions share limited identity (47% and 36% for STNV-C) with the corresponding regions in the cognate virus genome, but are interchangeable between satellite and helper viruses, thus suggesting that they share also the same replication mechanism. Short stem-loop structures are responsible for the specific packaging of the satellite virus RNA in its own CP. The first 3′-translational enhancer domain (TED) ever characterized is present downstream the CP encoding sequence in the 3′-untranslated region of STNV RNA, which is able to function as a 5′ cap and to interact with the eIF4F complex. Long-distance interactions between the 3′-TED and a 38 nt stretch in the 5′ UTR are important in enhancing and regulating the translation of satellite virus RNA.

The satellite virus genome contains signals for the correct packaging and specific assembly of the virus particles. The CP sequences of the three satellite viruses share 49%–62% sequence identity, but are different from any other viral protein, with the

exception of a limited sequence homology with the CP of Satellite maize white line mosaic virus (SMWLMV). The TNV satellite viruses are serologically distinguishable.

STNV-1 interferes to some extent with helper virus infections, ameliorating the TNV-induced symptomatology, probably because of competition for factors essential for shared replication mechanisms, for its presence in TNV inocula reduces slightly the virus concentration and the size but not the number of local lesions.

Whereas satellite viruses are not associated with any other member of the genus, BBSV supports the replication of a small, linear, single-stranded, noncoding satellite RNA 615 nt in size, which is encapsidated in the virions in monomeric or, more rarely, dimeric form. Monomers are thought to be produced from multimeric templates. BBSV satRNA enhances the aggressiveness of the helper virus for more lesions are produced on infected hosts when mixed inocula are used. Moreover, the effect of BBSV satRNA on the helper virus pathogenicity is further enhanced at low temperatures probably because of some satRNA interference with the expression of the host plant antiviral silencing genes.

Transmission and Host Range

All necro-like viruses are readily transmitted by mechanical inoculation to experimental herbaceous hosts, which usually react with necrotic local lesions, not followed by systemic infection. Under natural conditions, infection is often restricted to the roots. All necro-like viruses are transmitted through the soil, either by the chitrid fungus *Olpidium brassicae* (TNV-A, TNV-D, BBSV), or without the apparent intervention of a vector (OLV-1). Particles of species transmitted by *O. brassicae* are acquired by the vector from the soil where they are released from roots of infected plants through sloughing off epidermal cell layers and/or following decay of plant debris. Virions are bound tightly to the plasmalemma and the axoneme (flagellum) of the fungal zoospores and are transported inside the zoospore cytoplasm when the flagellum is retracted prior to encystment that precedes penetration in the host. The same mechanism is exploited by STNV to gain entrance into both vector and host. Since necro-like viruses and their satellites (STNV and BBSV satRNA) multiply in the plant cells but not in the fungal plasmodium, zoospores released from infected roots are virus-free. Transmission can therefore occur only if they come again in contact with and adsorb virus particles.

Seed transmission of necro-like viruses has not been reported, except for OLV-1, which was detected in the integuments and internal tissues of 82% olive seeds, and transmitted to 35% of the seedlings. Pollen transmission has not been reported for necro-like viruses so far.

The natural host range and geographical distribution of necro-like viruses varies with the species. Thus, TNV-A and TNV-D are ubiquitous and infect a wide range of cultivated and wild plants. OLV-1 was recorded from olive in several Mediterranean countries, citrus in Turkey, tomato in Poland and tulip in Japan; OMMV from olive in Portugal, spinach in Greece and tulip in the Netherlands; PoNV, the most recently characterized alphanecrovirus, was found on potato in UK; BBSV was recorded from beet in China, LWSV from leek in France.

Cytopathology

Necro-like viruses replicate very actively in their hosts, which translates into the production of a large number of virus particles in infected cells of all tissue types, including vessels. Virions are either scattered in the cytoplasm, gathered in bleb-like evaginations of the tonoplast into the vacuole, or arranged in crystalline arrays of various sizes. STNV particles can also give rise to crystals that can be found in the same cells along with those of the helper virus. Cells infected by TNV and OLV-1 also contain two types of inclusions, that is, clumps of electron-dense amorphous material resembling accumulations of excess CP and fibrous bundles made up of thin filaments with a helical structure. In OLV-1 infected cells, these bundles were identified as accumulations of the 8 kDa movement protein expressed by ORF2. The same protein and the 6 kDa movement protein coded for by ORF3 were detected by immunogold labeling near plasmodesmata.

Cytoplasmic clusters of membranous vesicles with fibrillar material, derived from the endoplasmic reticulum or lining the tonoplast, were observed in cells infected by OLV-1 and LWSV.

Further Reading

- Bringloe, D.H., Pleij, C.W., Coutts, R.H., 1999. Mutational analysis of *cis*-elements in the 3'- and 5'-untranslated regions of satellite tobacco necrosis virus strain C RNA. *Virology* 264, 76–84.
- Huang, Y.W., Hu, C.C., Hsu, Y.H., Lin, N.A., 2017. Replication of satellites. In: Hadidi, A., Flores, R., Randles, J.W., Palukaitis, P. (Eds.), *Viroids and Satellites*. London: Elsevier Academic Press, pp. 577–586.
- Kotta-Loizou, I., Peyret, H., Saunders, K., Coutts, R.H., Lomonosoff, G.P., 2019. Investigating the biological relevance of in vitro-identified putative packaging signals at the 5' terminus of satellite tobacco necrosis virus 1 genomic RNA. *Journal of Virology* 93, e02106–e02118. doi:10.1128/JVI.02106-18.
- Miras, M., Miller, W.A., Truniger, V., Aranda, M.A., 2017. Non-canonical translation in plant RNA viruses. *Frontiers in Plant Science* 8, 494. doi:10.3389/fpls.2017.00494.
- Monger, W., Jeffries, C., 2018. A new virus, classifiable in the family *Tombusviridae*, found infecting *Solanum tuberosum* in the UK. *Archives of Virology* 163, 1585–1594.
- Newburn, L.R., White, K.A., 2017. Atypical RNA elements modulate translational readthrough in tobacco necrosis virus D. *Journal of Virology* 91, e02443. doi:10.1128/JVI.02443-16.
- Pu, H., Li, J., Li, D., Han, C., Yu, J., 2013. Identification of an internal RNA element essential for replication and translational enhancement of tobacco necrosis virus A^C. *PLoS One* 8 (2), e57938. doi:10.1371/journal.pone.0057938.

- Pyle, J.D., Scholthof, K.-B.G., 2017. Biology and pathogenesis of satellite viruses. In: Hadidi, A., Flores, R., Randles, J.W., Palukaitis, P. (Eds.), *Viroids and Satellites*. London: Elsevier Academic Press, pp. 627–636.
- Rubino, L., 2017. Biology of satellites. In: Hadidi, A., Flores, R., Randles, J.W., Palukaitis, P. (Eds.), *Viroids and Satellites*. London: Elsevier Academic Press, pp. 567–575.
- Simon, A.E., Miller, W.A., 2013. 3' Cap-independent translation enhancers of plant viruses. *Annual Review of Microbiology* 67, 21–42. doi:10.1146/annurev-micro-092412-155609.
- Truniger, V., Miras, M., Aranda, M.A., 2017. Structural and functional diversity of plant virus 3'-Cap independent translation enhancers (3'-CITEs). *Frontiers in Plant Science* 8 (2047), doi:10.3389/fpls.2017.02047.
- Varanda, C.M., Materatski, P., Campos, M.D., *et al.*, 2018. Olive mild mosaic virus coat protein and p6 are suppressors of RNA silencing, and their silencing confers resistance against OMMV. *Viruses* 10 (416), doi:10.3390/v10080416.
- Vilar, M., Sauri, A., Monné, M., *et al.*, 2002. The insertion and topology of a plant viral movement protein in the endoplasmic reticulum membrane. *Journal of Biochemical Chemistry* 277, 23447–23452.
- Wang, X., Cao, X., Zhang, R., *et al.*, 2018. Hsc70–2 is required for beet black scorch virus infection through interaction with replication and capsid proteins. *Scientific Reports* 8 (4526).
- Xu, J., Liu, D., Zhang, Y., *et al.*, 2016. Improved pathogenicity of beet black scorch virus variant by low temperature and co-infection with its satellite RNA. *Frontiers in Microbiology* 7 (1771), doi:10.3389/fmicb.2016.01771.

Relevant Websites

<https://talk.ictvonline.org/taxonomy/Taxonomy>.

Nepoviruses (*Secoviridae*)

Hélène Sanfaçon, Agriculture and Agri-Food Canada, Summerland, BC, Canada

Crown Copyright © 2021 Published by Elsevier Ltd. All rights reserved.

Glossary

Lectin Protein that has the ability to bind carbohydrates (sugars).

Odonstyle Anterior section of the stylet of a longidorid nematode. The odonstyle consists of a needle-like mouth spear used to penetrate plant root cells.

Odontophore Posterior section of the stylet of a longidorid nematode. The odontophore is located at the

base of the nematode mouth. Stylet protractor muscles are attached to the odontophore.

Triradiate lumen Posterior region of the lumen (or food channel) of the esophagus of a longidorid nematode. The triradiate lumen is located in a muscular bulb at the base of the esophagus. The radial muscles attached to the triradiate lumen are used for food ingestion.

Introduction and Historical Perspective

Nepoviruses were among the first 16 groups of viruses recognized by the International Committee on Taxonomy of Viruses (ICTV). The name stands for nematode-transmitted viruses with polyhedral particles. From its origin as the “nepovirus group”, it became the genus *Nepovirus* within the family *Comoviridae*, which also included the genera *Fabavirus* and *Comovirus*. Taxonomic updates later led to the dissolution of the family *Comoviridae* and the creation of the family *Secoviridae*, to which the genus *Nepovirus* is assigned as part of the sub-family *Comovirinae*. Although nematode transmission was one of the original defining characteristics of the genus *Nepovirus*, the primary criteria for inclusion of viruses in the genus are now the structure of the RNA genome and of the particles and phylogenetic relationships. As a result, not all viruses belonging to the genus *Nepovirus* are nematode transmitted ([Table 1](#)), and some nematode-transmitted polyhedral viruses that were originally classified as nepoviruses, have been reclassified (e.g., *Strawberry latent ringspot virus*, currently an unassigned species within the family *Secoviridae*). As of 2019, there are 40 recognized nepovirus species ([Table 1](#)). Nepoviruses cause a variety of diseases of economic importance, notably on grapevine, fruit trees, small fruits and other horticultural crops.

Taxonomy, Phylogeny, and Evolution

The genus *Nepovirus* belongs to the sub-family *Comovirinae* (also including the genera *Comovirus* and *Fabavirus*), family *Secoviridae* (encompassing the genera *Cheravirus*, *Sadwavirus*, *Torradovirus*, *Sequivirus* and *Waikavirus*, in addition to the sub-family *Comovirinae*), order *Picornavirales* (consisting of six families including *Secoviridae* and *Picornaviridae*), realm *Riboviria*. Nepoviruses share several characteristics in common with other members of the order *Picornavirales*: the small polyhedral viral particles, the positive-strand RNA genomes that are polyadenylated at their 3' end and linked to a small genome-linked protein (VPg) at their 5' end, the production of mature viral proteins by proteolytic cleavage of large polyproteins, and the conserved signature sequences and order of viral replication protein domains within the polyproteins including the nucleotide triphosphate-binding protein (a putative helicase abbreviated either as NTB or as Hel in the literature), the VPg, a chymotrypsin-like protease (Pro), and the RNA-dependent RNA polymerase (Pol). Nepoviruses have a bipartite genome with each RNA encoding one large polyprotein, a property they share with several other members of the family *Secoviridae* (referred to as secovirids from hereon). Similar to other secovirids with a bipartite genome, the RNA1 polyprotein includes the domains for the replication proteins and the RNA2 polyprotein contains the domains for the movement protein (MP) and coat protein (CP). Nepoviruses encode a single large coat protein of 53–60 kDa, a characteristic they only share with a few other secovirids (e.g., *Strawberry mottle virus*). In contrast, other secovirids have either two (*Comovirus*, *Fabavirus*, *Sadwavirus*) or three (*Cheravirus*, *Torradovirus*, *Sequivirus*, *Waikavirus*) coat proteins. In phylogenetic trees based on the Pro-Pol amino acid sequence, which is defined by the sequence between the conserved catalytic cysteine (or serine, see below) of the Pro domain and the conserved GDD motif of the Pol domain, nepoviruses branch together as a sub-branch within a larger branch corresponding to the sub-family *Comovirinae* ([Fig. 1](#)). Nepoviruses are genetically very diverse, more so than most other genera in the family. Phylogenetic trees based on different protein domains are not always congruent ([Fig. 1](#)). This is likely due to reassortment between the two viral RNAs and/or recombination events that have contributed to the evolution of nepoviruses. Nepoviruses have been divided into three subgroups (A, B and C) based on the length and packaging of RNA2, and serological properties ([Table 1](#)). These subdivisions may need to be further refined in the future. Indeed, while subgroup B nepoviruses confidently branch together in the phylogenetic trees based on the Pro-Pol or CP amino acid sequences, subgroup C nepoviruses only cluster together in the CP phylogenetic tree and subgroup A nepoviruses do not consistently group together ([Fig. 1](#)).

Table 1 Recognized nepovirus species

Name	Abbreviation	Vector ^a	RNA1 ^b	RNA2 ^b	Satellite RNAs ^b	
					Type B	Type D
Subgroup A						
<i>Aeonium ringspot virus</i>	AeRSV		NC_038762	NC_038761		
<i>Arabidopsis mosaic virus</i>	ArMV	<i>Xiphinema diversicaudatum</i>	NC_006057	NC_006056	NC_003523	NC_001546
<i>Arracacha virus A</i>	AVA					
<i>Artichoke Aegean ringspot virus</i>	AARSV					
<i>Cassava American latent virus</i>	CsALV					
<i>Grapevine deformation virus</i>	GDefV		NC_017939	NC_017938		
<i>Grapevine fanleaf virus</i>	GFLV	<i>X. index</i>	NC_003615	NC_003623	NC_003203	
<i>Melon mild mottle virus</i>	MMMoV		NC_038765	NC_038766		
<i>Mulberry mosaic leaf roll associated virus</i>	MMLRaV		NC_038767	NC_038768		
<i>Olive latent ringspot virus</i>	OLRSV			NC_038863		
<i>Potato black ringspot virus</i>	PBRV		NC_022798	NC_022799		
<i>Raspberry ringspot virus</i>	RpRSV	<i>Longidorus elongatus</i> ; <i>L. macrosoma</i> <i>Paralongidorus maximus</i>	NC_005266	NC_005267		
<i>Tobacco ringspot virus</i>	TRSV	<i>X. americanum</i>	NC_005097	NC_005096		NC_003889
Subgroup B						
<i>Artichoke Italian latent virus</i>	AILV	<i>L. apulus</i> ; <i>L. fasciatus</i>	LT608395	LT608396		
<i>Beet ringspot virus</i>	BRSV	<i>L. elongatus</i>	NC_003693	NC_003694		
<i>Cocoa necrosis virus</i>	CNV		EU741694 (partial)			
<i>Crimson clover latent virus</i>	CCLV					
<i>Cycas necrotic stunt virus</i>	CNSV		NC_003791	NC_003792		
<i>Grapevine Anatolian ringspot virus</i>	GARSV		NC_018383	NC_018384		
<i>Grapevine chrome mosaic virus</i>	GCMV		NC_003622	NC_003621		
<i>Mulberry ringspot virus</i>	MRSV	<i>L. martini</i>				
<i>Myrobalan latent ringspot virus</i>	MLRSV			+		
<i>Potato virus B</i>	PVB		KX656670	KX656671		
<i>Tomato black ring virus</i>	TBRV	<i>L. attenuatus</i>	NC_004439	NC_004440	NC_003890	
Subgroup C						
<i>Artichoke yellow ringspot virus</i>	AYRSV		NC_038862 (partial)			
<i>Blackcurrant reversion virus</i>	BRV	<i>Cecidophyopsis ribis</i> (mite)	NC_003509	NC_003502	NC_003872	
<i>Blueberry latent spherical virus</i>	BLSV		NC_038764	NC_038763		
<i>Blueberry leaf mottle virus</i>	BLMoV			U20621 (partial)		
<i>Cassava green mottle virus</i>	CsGMV					
<i>Cherry leaf roll virus</i>	CLRV		NC_015414	NC_015415		
<i>Chicory yellow mottle virus</i>	ChYMV				NC_003778	NC_003971
<i>Grapevine Bulgarian latent virus</i>	GBLV		NC_015492	NC_015493	+	
<i>Grapevine Tunisian ringspot virus</i>	GTRSV					
<i>Hibiscus latent ringspot virus</i>	HLRSV					
<i>Lucerne Australian latent virus</i>	LALV					
<i>Peach rosette mosaic virus</i>	PRMV	<i>X. americanum</i>	NC_034214	NC_034215		
<i>Potato virus U</i>	PVU					
<i>Soybean latent spherical virus</i>	SLSV		NC_032270	NC_032271		
<i>Tomato ringspot virus</i>	ToRSV	<i>X. americanum</i> ; <i>X. bricolensis</i> <i>X. californicum</i> ; <i>X. rivesi</i>	NC_003840	NC_003839		

^aOnly cases for which a specific association with a nematode vector has been verified experimentally are listed.^bGenBank accession numbers are listed. A (+) symbol indicates that a type B satellite is known to be associated with the virus but has not been sequenced.

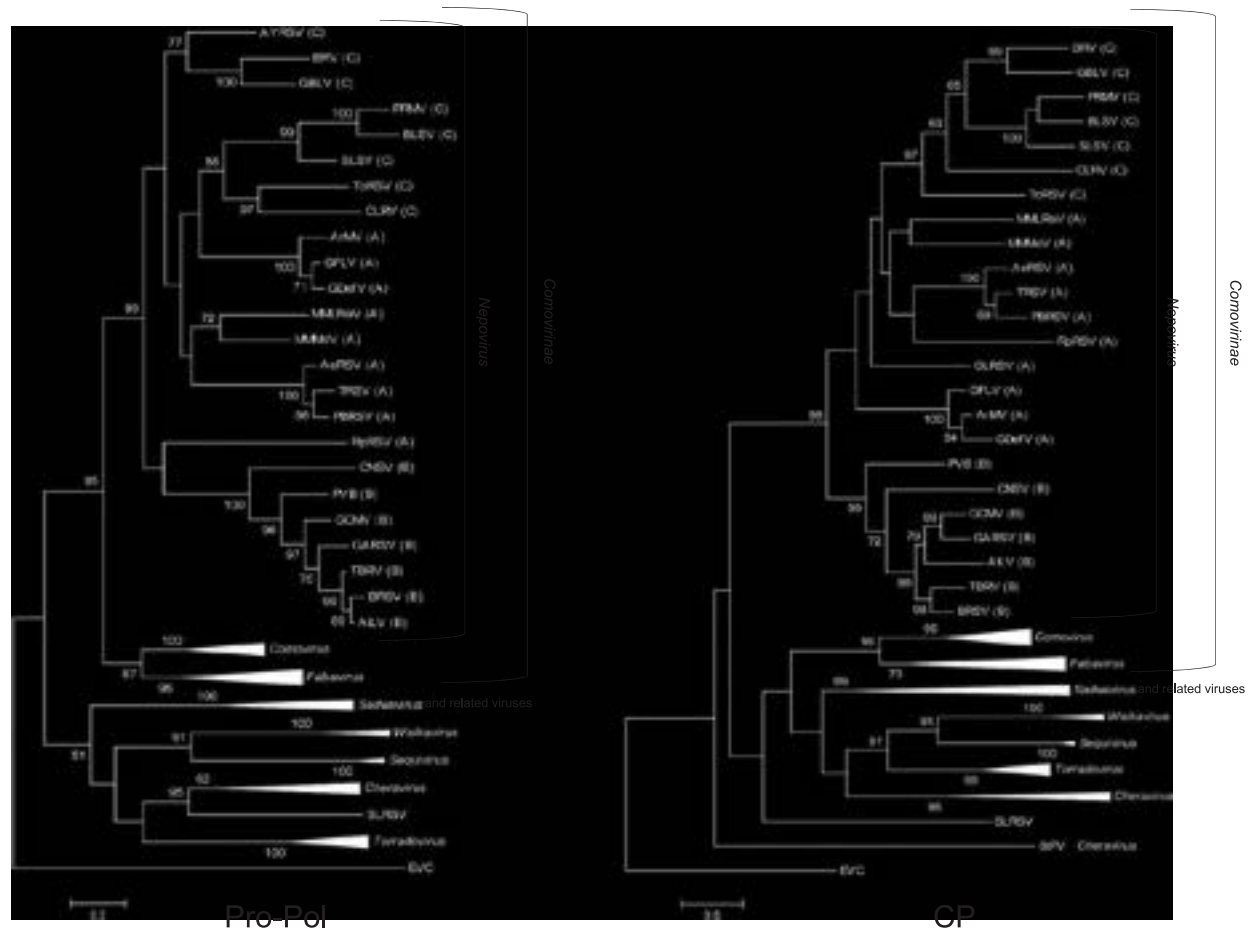


Fig. 1 Phylogenetic analyses of nepoviruses and other members of the family *Secoviridae*. The amino acid sequence of the CP or Pro-Pol domains from recognized species within the family *Secoviridae* were aligned using Clustal W. Phylogenetic analyses were conducted using the maximum likelihood method as implemented in MEGA 6. *Enterovirus C* (EVC, genus *Enterovirus*, family *Picornaviridae*) was used as an outgroup. The results of 1000 bootstraps are shown for each node (only values above 50% are shown). A scale for the genetic distance is shown below the trees. Nepovirus acronyms are as listed in Table 1, the letter in parenthesis after each acronym represents the subgroup. Branches corresponding to genera other than nepoviruses were collapsed. The *Sadwavirus* branch includes *Satsuma dwarf virus* (the only recognized species of the genus) and four related species that are unassigned within the family *Secoviridae* (*Strawberry mottle virus*, *Black raspberry necrosis virus*, *Chocolate lily virus A* and *Dioscorea mosaic associated virus*). SLRSV: *Strawberry latent ringspot virus*, an unassigned species in the family *Secoviridae*. StPV: *Stocky prune virus*.

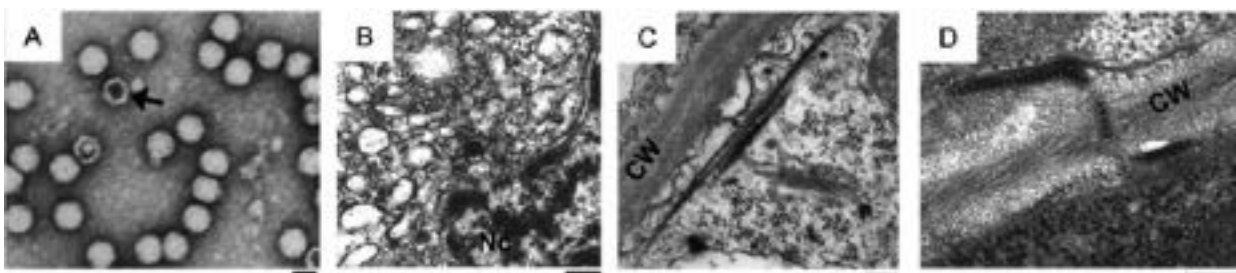


Fig. 2 Electron micrograph depicting purified nepovirus particles and cytopathological structures typical of nepovirus-infected cells. (A) Purified ToRSV particles in negative staining. Note the empty particle (T-particle) which is penetrated by the negative stain (arrow). (B) Proliferation of membrane vesicles observed in the vicinity of the nucleus (Nc) in ToRSV-infected cells. (C) Tubular structures containing virus-like particles accumulating near the cell wall (CW) in PRMV-infected cells. (D) Tubular structure traversing the cell wall in ArMV-infected cells. Bars represent 25 nm in panel A and 200 nm in panels B–D.

Virion Structure

Nepoviruses have isometric particles of 26–30 nm in diameter with sharp hexagonal outlines (Fig. 2(A)). Equilibrium centrifugation in CsCl of purified virus particles typically reveals the presence of three types of particles. T-particles sediment at 50S and do not contain an RNA molecule. In electron microscopy, these empty particles are penetrated by a negative stain. B-particles sediment at 115–134S and contain a single molecule of RNA1. In the case of sub-group A nepoviruses, B-particles can also contain two molecules of RNA2. M-particles sediment at 86–128S and contain a single molecule of RNA2. M- and B-particles of sub-group C nepoviruses are often difficult to separate due to the larger size of RNA2.

Nepovirus particles contain 60 molecules of the single CP. The crystal structure of virus particles was solved for TRSV, GFLV and ArMV (see Table 1 for acronyms). The structure of BRV particles was also inferred by cryo-electron microscopy. The pseudo T=3 structure is similar to that of comoviruses and picornaviruses. The CP contains three functional domains. Each functional domain corresponds to one of the three smaller CPs found in picornaviruses and consists of eight β -strands that form a β -barrel. Comoviruses also share a similar structure with one large CP containing two β -barrels and one small CP with a single β -barrel. The N-terminal region of the CP is buried inside the particle while the C-terminal region of the CP is exposed to the outside of the particle. In many nepoviruses, a small fragment of the C-terminus of the CP is cleaved off from the virions at late stages of infection, although the biological function of this cleavage is not known. Several loops connecting the β -strands are exposed to the surface of the particle and interact with the viral movement protein to support cell-to-cell movement or with vector receptors to enable transmission. The mechanism of encapsidation is not well-characterized, but the presence of empty particles with structures similar to full particles suggests that presence of the viral RNAs is not strictly required to initiate capsid assembly.

Genome Organization and Function of Viral Proteins

The genomic organization of a representative virus from each subgroup is shown in Fig. 3. The C-terminal protein domains of the polyproteins are conserved amongst nepoviruses, while the N-terminal regions of the polyproteins are more variable. The C-terminal portion of the RNA1 polyprotein contains the NTB, VPg, Pro and Pol domains, which are implicated in viral RNA replication and regulation of polyprotein processing. Identification of cleavage sites upstream of NTB defined two protein domains in the N-terminal region of the ToRSV and ArMV RNA1 polyproteins, termed X1 and X2. Putative cleavage sites at corresponding positions were found in the RNA1 polyproteins of other nepoviruses. Thus, the presence of two protein domains upstream of NTB is likely a common feature of nepovirus RNA1 polyproteins, which is not shared by comoviruses or fabaviruses. The X2 protein domain is highly hydrophobic, shares conserved sequence motifs with the 32 kDa protein of comoviruses and has been suggested to play a role in the assembly of replication complexes. The X1 domain shows a high degree of sequence diversity amongst nepoviruses and its function is unknown. Because RNA1 encodes all the essential replication proteins, it is able to replicate in single plant cells without the assistance of RNA2. The RNA2 polyprotein includes the domains for the CP and MP at its C-terminus and one or two additional protein domains at its N-terminus: the single 2a domain for subgroup A and B nepoviruses and the two X3 and X4 domains for ToRSV, a subgroup C nepovirus. The GFLV 2a protein plays a role in the replication of RNA2. The function of X3 and X4 is not known. Both RNA1 and RNA2 are required for cell-to-cell and systemic movement of the virus in the plant. Infectious cDNA clones have been produced for subgroup A (GFLV, ArMV, RpRSV, TRSV) and subgroup B (TBRV) nepoviruses, allowing reverse genetic experiments for these viruses.

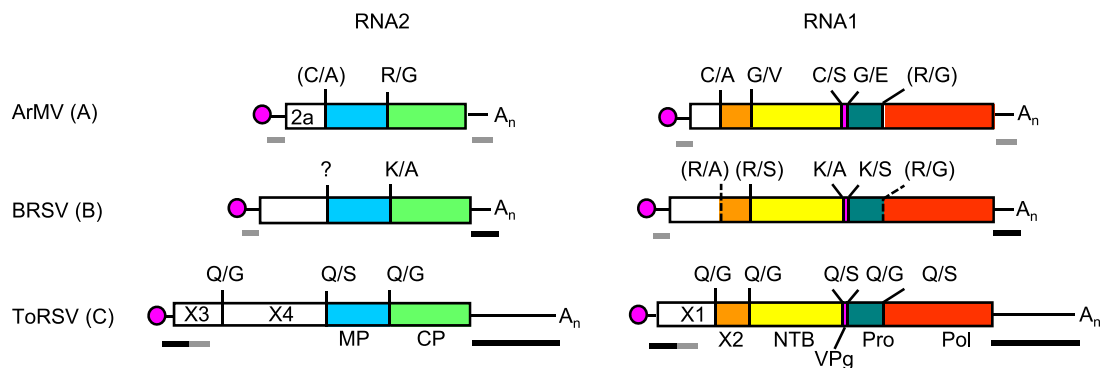


Fig. 3 Genomic organization of representative nepoviruses of subgroup A (*ArMV*), B (*BRSV*) and C (*ToRSV*). Each RNA is represented with the covalently attached VPg (pink circle) and the polyA tail (A_n). The coding regions are represented by the boxes. Cleavage sites confirmed by *in vitro* processing experiments or by the detection of viral proteins in infected plants are indicated by the continuous vertical lines. Putative cleavage sites that have not been confirmed experimentally are shown with dashed lines. Dipeptide sequence of confirmed or putative (in parenthesis) cleavage sites are indicated above each line. Thick bars below each RNA indicate regions with high degree of sequence identity between RNA1 and 2 (black for regions with near 100% sequence identity and gray for regions with 75%–83% sequence identity).

Each RNA includes untranslated regions (UTR) at its 5' and 3' ends. The 5' UTR is 70–300 nucleotides (nts) long while the 3' UTR varies in length from 200 to 400 nts for subgroup A and B nepoviruses to 1300–1600 nts for subgroup C nepoviruses. RNA1 and 2 often share regions of complete or partial sequence identity in the UTRs (Fig. 3). For example, the 5' UTRs of ToRSV RNA1 and 2 share a region of 100% sequence identity which extends into the coding region. On the other hand, the 5' UTRs of the BRV RNAs share only 78% sequence identity and do not extend beyond the UTRs. The 3' UTRs of RNA1 and 2 are often nearly identical, especially those of subgroup C nepoviruses. Conservation of sequence identity in the UTRs has been attributed to recombination between the two RNAs occurring during replication of the viral RNAs and/or to selection pressure.

Production of Viral Proteins: Translation and Polyprotein Processing

The viral RNAs are polyadenylated but are not capped, requiring a cap-independent translation mechanism. Studies on BRV RNAs suggested that long-distance interaction between the 5' and 3' UTRs (possibly directed by sequences in the loops of putative stem-loop structures) is required for efficient translation of the viral RNAs. In addition, the polyA tail enhances translation efficiency. Further work will be necessary to elucidate the exact mechanism of BRV RNAs translation, in particular the requirement for plant translation factors, and to determine whether other nepoviruses utilize similar or distinct mechanisms to translate their RNAs.

Nepoviruses encode a single protease which is responsible for processing the two polyproteins. Nepovirus Pros are related to the 3C-Pro of picornaviruses. The catalytic triad normally consists of a histidine, glutamic acid and cysteine, although the Pros of a few subgroup C nepoviruses have a serine instead of the cysteine as the third component of the triad. Proteases also contain a substrate-binding pocket that determines their cleavage site specificity. Similar to picornaviruses, a conserved histidine is found in the substrate-binding pocket of all subgroup C nepovirus Pros characterized so far and of one subgroup A nepovirus Pro (MMLRaV). The cleavage sites recognized by these Pros normally include a glutamine, asparagine or aspartate at the -1 position. This is similar to the conserved glutamine or glutamate found at the -1 position of picornavirus cleavage sites. However, there are reported exceptions, including a characterized cleavage site recognized by the BLSV Pro with a cysteine at the -1 position and several putative cleavage sites of subgroup C nepoviruses with a histidine at the -1 position. The Pros of subgroup A and B nepoviruses contain a leucine instead of a histidine in their substrate-binding pocket and recognize a variety of cleavage sites that have lysine, cysteine, arginine or glycine at the -1 position. Thus, nepovirus proteases have very diverse specificities, often making the prediction of cleavage sites difficult.

The RNA1 polyprotein is cleaved predominantly intra-molecularly (*cis*-cleavage). The RNA2 polyprotein is cleaved *in trans* by the Pro. The processing cascade results in the release of mature proteins and stable processing intermediates containing two or more protein domains. These intermediate polyproteins accumulate in infected plant cells and may have different activities from the mature proteins. In ToRSV-infected cells, several polyprotein intermediates containing the NTB domain are detected in addition to the mature NTB protein. In BRSV-infected cells, an intermediate containing the Pro and Pol domains accumulates rather than the mature Pro and Pol proteins. This suggests preferential recognition of some cleavage sites by the viral Pro. Slow release of mature proteins by processing of stable intermediates at suboptimal cleavage sites provides a regulatory mechanism to control the accumulation of specific proteins during the replication cycle. In fact, the activity of the Pro itself is regulated by its release from larger polyprotein precursors. Indeed, the mature Pro of GFLV and ToRSV processes the cleavage sites of the RNA2 polyprotein more efficiently than the VPg-Pro precursor.

Viral RNA Replication

Infection of plant cells by nepoviruses results in membrane proliferation and the formation of membrane vesicles (Fig. 2(B)). The use of cerulenin, an inhibitor of *de novo* phospholipid synthesis, has confirmed that membrane proliferation is required for the replication of GFLV RNAs. Brefeldin A also inhibits GFLV replication suggesting a requirement for intact vesicle trafficking between the endoplasmic reticulum (ER) and the Golgi apparatus. The replication complexes of two nepoviruses (GFLV and ToRSV) were shown to co-localize with ER-derived membranes in infected cells. Double-strand RNA replication intermediates, viral replication proteins and replication activity are associated with the membrane-bound complexes. GFLV VPg antibodies were used to isolate membrane vesicles that had a rosette-like structure similar to that associated with picornavirus replication complexes.

As shown for ToRSV, the mature NTB protein and the intermediate NTB-VPg and X2-NTB-VPg polyproteins co-fractionate with the replication complex, suggesting that they act as membrane anchors for the replication complex. The NTB protein contains two membrane-binding domains: a C-terminal transmembrane domain and a putative N-terminal amphipathic helix. The ToRSV X2 protein also contains several transmembrane helices that traverse the ER membrane. Other replication proteins (Pro, Pol or intermediate polyproteins containing these domains) are soluble when expressed individually but are found in association with the membrane-bound replication complex in infected cells. Thus, they are probably brought to the replication complex either as part of a larger polyprotein that includes the NTB domain or through protein-protein interaction with the viral membrane anchors. By analogy with picornaviruses, the VPg protein may act as a primer for viral replication, although this has not been demonstrated experimentally for nepoviruses. The GFLV RNA2-encoded 2a protein is also

associated with the replication complex. The 2a protein is required for the replication of RNA2 but not RNA1 and probably interacts with RNA1-encoded replication proteins.

Cell-to-Cell and Systemic Movement in the Plant

Nepovirus-infected cells are characterized by the presence of tubular structures containing virus-like particles in or near the cell wall (Fig. 2(C) and (D)). These tubules are similar to the ones found in comovirus-infected cells and have been suggested to direct the cell-to-cell movement of intact virus particles. The viral movement protein is a structural component of the tubular structure. Expression of the GFLV MP alone is sufficient to induce the formation of empty tubular structures in intact plant cells or in protoplasts. In the latter case, tubular extensions are found projecting from the surface of the protoplasts, a phenomenon also induced by comovirus MP. The GFLV MP is an integral membrane protein. The secretory pathway and the cytoskeleton are involved in the intracellular targeting of the GFLV MP from its site of synthesis (probably in association with the ER-bound replication complex) to specific foci in the cell wall where it assembles into tubules. GFLV cell-to-cell movement also requires myosin-motor plant proteins, which may assist in transporting other plant proteins to the plasmodesmata to initiate the formation of the MP tubules. By analogy with comoviruses, it is likely that an interaction between MP and CP is necessary to enable nepovirus cell-to-cell movement. A surface-exposed region of the CP and the C-terminal region of the MP are probably involved in this interaction. Long-distance movement of TRSV occurs through the phloem resulting in the invasion of most tissues of the plant including meristematic tissues. The virus probably reaches the phloem through cell-to-cell movement from inoculated cells to phloem sieve tubes.

Host Range, Symptom Determinants and Interaction With Plant Defense Responses

Most nepoviruses have a wide host range that includes woody and herbaceous hosts. In nature, many hosts remain symptomless while others display symptoms such as necrotic or chlorotic rings which can appear as concentric rings (ringspots) or lines. Other symptoms can include leaf flecking and mottling, vein necrosis, plant stunting and in some cases death. Common experimental hosts used for virus propagation are *Chenopodium quinoa* (in which most nepoviruses induce obvious symptoms), *C. amaranticolor*, *C. murale*, *Cucumis sativus*, *Nicotiana clevelandii*, *N. benthamiana*, *Petunia hybrida* and *Phaseolus vulgaris*. The intensity of symptoms produced by nepovirus infection depends on the specific virus-host combination and on environmental conditions.

In many herbaceous hosts and in particular in *Nicotiana* species, symptoms develop on the inoculated leaves and on the first upper systemic leaves. Later in infection, new leaves remain free of symptoms although the virus is present. This phenomenon is termed symptom recovery. Recovered leaves of nepovirus-infected plants often show reduced concentration of viral RNAs and this has been attributed at least in part to the induction of the plant antiviral RNA silencing. Recovered leaves are also resistant to reinfection in a sequence-dependent manner. In the *Nicotiana benthamiana*-ToRSV interaction, the establishment of symptom recovery depends on the specific isolate and also on the growth temperature (Fig. 4). The onset of symptom recovery from a severe ToRSV isolate is not accompanied with a significant reduction of viral RNA levels. However, translation of the viral RNAs is suppressed limiting the accumulation of viral proteins and encapsidation. It has been hypothesized that the low nepovirus titers typically found in late stages of infection contribute to their ability to escape plant surveillance systems and invade meristematic tissues.

When they occur, necrotic symptoms share characteristics of a hypersensitive response (HR) normally induced by activation of plant dominant resistance gene. Although the virus can be restricted to local necrotic lesions developing on inoculated leaves in some hosts, the HR-like response often fails to contain the virus and systemic infection is observed, which may be followed by recovery or not. The plant genetic determinants of this HR-like response have been characterized in the *Arabidopsis thaliana*-TRSV interaction. Most *A. thaliana* ecotypes are tolerant to TRSV and do not display visible symptoms. However, the lethal systemic necrosis induced by TRSV in some ecotypes was linked to a dominant mutation in the TTR1 resistance gene. Viral symptom determinants have also been identified in a few interactions. In the *Nicotiana occidentalis*-GFLV interaction, construction of chimeric clones between two isolates identified the 2a protein as the elicitor for the necrotic HR-like response. Using a similar approach, a single amino acid in the GFLV Pol was found to determine vein clearing symptoms in *N. benthamiana*.

Transmission

Many nepoviruses are transmitted by soil dagger nematodes belonging to the genera *Xiphinema*, *Longidorus* or *Paralongidorus* in the order *Dorylaimida*, family *Londigoridae*. The nematodes feed ectoparasitically on the roots using long mouth stylets. There is no evidence that nepoviruses replicate in the nematodes. The acquired viruses remain transmissible for varying periods of time (9 weeks for *Longidorus* species and up to 4 years for *Xiphinema* species), suggesting different modes of retention and release of the virus. Because of the restricted movement of the nematodes through the soil, the spread of nematode-transmitted nepoviruses through an infected field is slow and often occurs in patches. The interaction between nepoviruses and nematodes is

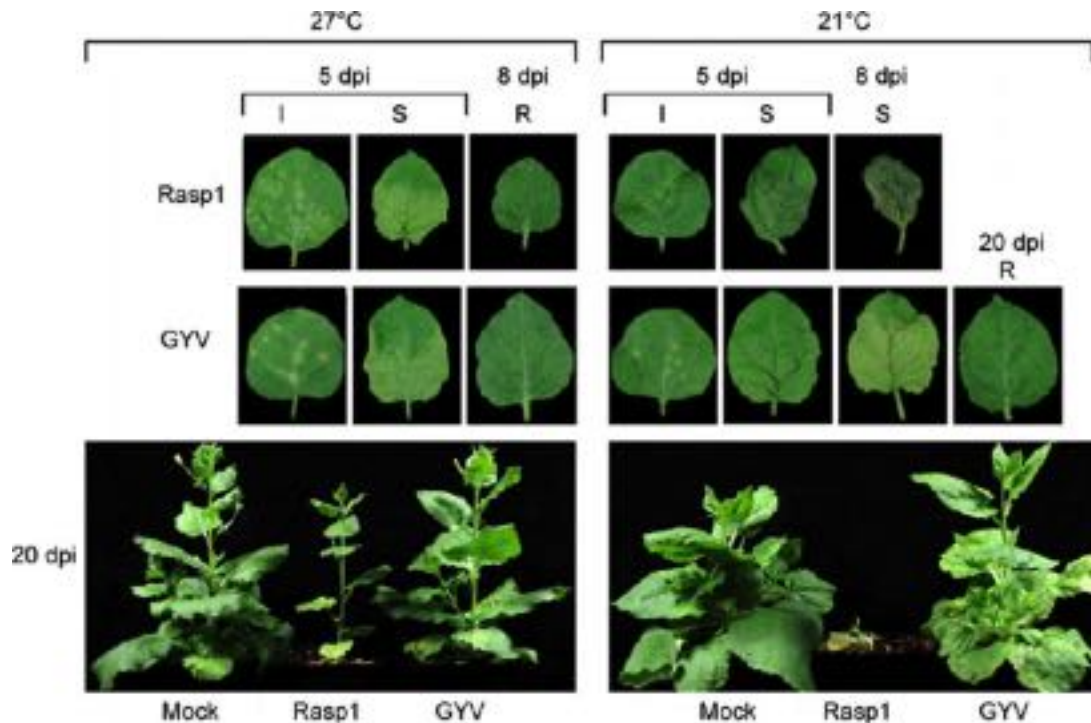


Fig. 4 Symptom recovery in ToRSV-infected *Nicotiana benthamiana* plants. Plants inoculated with a severe (Rasp1) or mild (GYV) isolate of ToRSV were grown at either 27°C or 21°C. GYV-infected plants recover from infection at both temperature, while Rasp1-infected plants only recover from infection at 27°C. Detached inoculated leaves (I), upper symptomatic leaves (S) or recovered leaves (R) are shown at the indicated days post-infection (dpi). The lower panel compares plants that were mock-inoculated to plants infected with the two isolates at 20 dpi after growth at 27°C (left) or 21°C (right). Modified from Ghoshal and Sanfacon, 2014. Temperature-dependent symptom recovery in *Nicotiana benthamiana* plants infected with tomato ringspot virus is associated with reduced translation of viral RNA2 and requires ARGONAUTE 1. *Virology* 456–457, 188–197. Paudel *et al.*, 2018. Expression and antiviral function of ARGONAUTE 2 in *Nicotiana benthamiana* plants infected with two isolates of tomato ringspot virus with varying degrees of virulence. *Virology* 524, 127–139. Both under Crown Copyright.

usually specific with only one or two species of nematodes transmitting a given nepovirus (Table 1). In the case of GFLV, the viral determinant for the specificity of nematode transmission has been mapped to the CP, in particular a positively charged pocket exposed to the surface of virus particles. It is not known whether specific nematode receptors recognize the viral CP. Earlier studies suggested that carbohydrates may be involved in the retention of ArMV particles in its vector. However, further experiments will be required to determine if the viral coat protein has lectin properties. It has been suggested that pH changes may be involved in the release of viral particles from their site of retention within the nematode. Nepoviruses transmitted by *Longidorus* species are usually associated with the odontostyle, while viruses transmitted by *Xiphinema* species are found associated with the cuticle lining the lumen of the odontophore and the esophagus. Interestingly, while TRSV and ToRSV are both transmitted by *X. americanum*, immunofluorescence labeling of these viruses in the nematode vector revealed different sites of retention. TRSV is retained predominantly in the lining of the lumen of the stylet extension and the anterior esophagus, while ToRSV is localized only in the triridiate lumen. Although the predominant vector for transmission of TRSV is a nematode, possible aerial vectors have been suggested including *Thrips tabaci* and *Epitrix hirtipennis* (flea beetle). The importance of these vectors in natural epidemics of TRSV-induced diseases needs to be confirmed. A recent study also reported detection of TRSV double-strand RNA replication intermediates in bees, suggesting viral replication. However, TRSV is not transmitted by bees and the biological relevance of this observation, including its possible association with bee decline, is not clear. Other nematode-transmitted nepoviruses do not have known aerial vectors. BRV is transmitted by the eriophyid gall mite of black currant (*Cecidophyopsis ribis*) and possibly other *Cecidophyopsis* species but not by nematodes. Plant-to-plant transmission can occur rapidly (in only 4 h). Virus particles have not been found inside mites suggesting that the transmission may be non-persistent or semi-persistent. Two surface-exposed amino acid triplets are conserved between the CP of BRV and other mite-transmitted viruses from unrelated genera, suggesting that they may play a role in the interaction between the virus and its vector. However, this remains to be determined experimentally.

Seed transmission has been reported for most but not all nepoviruses. For example, BRV is apparently not seed transmitted. Infection can occur through the ovule or the pollen, although nepovirus-infected pollen may not compete effectively with healthy pollen. An exception to this is BLMoV and CLRV, which are efficiently transmitted by pollen. Honeybees likely contribute to the pollen transmission of BLMoV by increasing the movement of infected pollen. Many nepoviruses are readily transmitted through grafting and mechanical inoculation. The propagation of infected seed and plant stocks plays an important role in the long-

distance movement of nepoviruses. Because many nepoviruses have a wide host range including common weeds, dormant weed seeds may provide a reservoir for the virus in the field.

Population Structures: Quasi-Species, Satellite RNAs and Defective-Interfering RNAs

Examination of nepovirus isolates recovered from different hosts or geographic locations has revealed a degree of sequence diversity of 2%–20% at the nucleotide level. Isolates can also differ in the length of their RNAs with the RNA2 polyprotein N-terminal domains being the most variable, as shown for the GFLV and ArMV 2a protein and the ToRSV X4 protein. Evidence for recombination and reassortment between viral RNAs is well-documented resulting in the emergence of new viral strains or species, although interspecies recombination has so far only been reported among closely related nepovirus species. Mixed infections between viral isolates and/or closely related species frequently occur in the field. In agreement with the quasi-species model, detailed analyses of GFLV population structure highlighted the presence of a few predominant variants along with other minor variants. An ArMV field isolate was shown to contain two species of RNA2, each encoding a distinct polyprotein. It is possible that co-existing variants act synergistically to facilitate infection or determine symptoms but this has not been tested experimentally. Analysis of a number of CLRV isolates revealed that the degree of diversity is defined primarily by the host rather than the geographic location. This is an unusual situation among plant viruses and may reflect the fact that this virus is pollen-transmitted.

Two classes of satellite RNAs (satRNAs) have been found in association with some nepoviruses ([Table 1](#)). SatRNAs depend on the helper virus for their replication and are encapsidated in virus particles. Type B satRNAs are 1100–1500 nts in length. They are linked to a VPg molecule at their 5' end, polyadenylated at their 3' end and encode a non-structural protein which is essential for their replication. The exact function of the encoded protein is not known but it has been suggested that it interacts with the viral replication complex. The 5'-ends of type B satRNAs often have short sequence motifs that are identical or nearly identical to the 5' ends of the viral RNAs and are likely recognized by the viral replication complex. One or several copies of type B satRNAs are packaged in the viral particles, either alone or together with one molecule of RNA2. As a general rule, type B satRNAs are replicated specifically by the virus with which they are associated although there are exceptions (e.g., the replication of GFLV satRNA is supported by ArMV). Type B satRNAs are usually found in low concentration and do not affect significantly the replication of the helper virus or the symptomatology.

Type D satRNAs are less than 500 nts long. They are not linked to a VPg molecule or polyadenylated and do not encode a protein. Type D satRNAs are encapsidated as monomeric or multimeric linear molecules. A circular form of the molecule is present in infected cells and serves as the polymerase template for replication through a rolling circle mechanism. The multimeric linear forms produced during replication are cleaved in an autocatalytic reaction which allows the release of linear monomers. These monomers are circularized to form new templates of positive or negative polarity. Type D satRNAs have been shown to either attenuate (TRSV) or intensify (ArMV) symptoms associated with the helper virus.

TBRV is the only nepovirus reported to be associated with defective-interfering RNAs (diRNAs). These 400–500 nts RNAs are derived exclusively from RNA1 and are probably produced during viral RNA replication. Presence of the diRNAs was associated with attenuated symptoms and reduced virus accumulation in some hosts. However, further work will be necessary to confirm the biological significance of the diRNAs and to determine whether they are also observed in other nepovirus-plant pathosystems.

Diseases, Economic Considerations and Control

Nepoviruses cause a wide range of diseases on a variety of crops including: grapevine, small fruits such as strawberry, raspberry, blueberry, black currant and red currant, fruit trees such as peach, apricot, almond, cherry, plum, walnut and apple, and other horticultural crops including but not limited to hop, soybean, potato, beet and tobacco. Many nepoviruses also induce diseases in ornamental species. Most nepoviruses are restricted geographically by the natural distribution of their nematode vector. An exception to this is GFLV which has been disseminated worldwide along with its vector. GFLV is the most significant nepovirus at the economic level and can reduce yield in grapevine by as much as 80%. Other nepoviruses can cause significant diseases where they occur.

Nepoviruses are mainly controlled through the removal of infected plants and replanting with resistant cultivars (when available) or with virus-free certified plant material. Weed control can reduce potential reservoirs for further nepovirus infection. In the case of nematode-transmitted nepoviruses, soil can be treated with broad-spectrum fumigant nematicides. However, because nematode populations can occur at considerable depths in the soil (one meter or more), fumigation rarely completely eliminates plant parasitic nematodes from a site and population densities typically return to pre-fumigation levels within a few years in the presence of a suitable host. Further, many fumigant nematicides are now banned due to their toxicity.

Few natural sources of resistant cultivars are available for nepoviruses. Transgenic expression of small regions of the viral genome was shown to induce the plant antiviral RNA silencing and provide virus resistance in herbaceous hosts (several nepoviruses) and in grapevine (GFLV). Broad spectrum resistance against a range of GFLV isolates was also obtained in *N. benthamiana* and in grapevine upon transgenic expression of camelid-derived heavy-chain only GFLV-specific antibodies. The characterization of plant factors essential for nepovirus infection combined with emerging gene editing approaches will likely provide new opportunities for generating non-transgenic nepovirus-resistant plants in the future.

Concluding Remarks

Although the infection cycle of nepoviruses is relatively well-understood, at least in herbaceous hosts, there are many outstanding questions. Many plant factors facilitating the translation, replication and movement of nepoviruses remain to be identified and/or characterized. Such plant factors may prove useful new targets for antiviral strategies, based on gene editing techniques. It will also be critical to evaluate the impact of nepoviruses and their complex populations (including the presence of multiple variants) in the field and to gain further insights into the emergence of virulent strains or new species. Additional research is needed to better understand the complex interaction between nepoviruses and various plant defense responses leading to the range of symptoms, including tolerance, symptom recovery and lethal systemic necrosis. Finally, evaluating the contribution of environmental factors on virus titers, disease severity and seed or vector transmission will also likely be a focus of future studies.

Further Reading

- Amari, K., Lerich, A., Schmitt-Keichinger, C., Dolja, V.V., Ritzenthaler, C., 2011. Tubule-guided cell-to-cell movement of a plant virus requires class XI myosin motors. *PLoS Pathogens* 7, e1002327.
- Fuchs, M., Schmitt-Keichinger, C., Sanfacon, H., 2017. A renaissance in nepovirus research provides new insights into their molecular interface with hosts and vectors. *Advance in Virus Research* 97, 61–105.
- Ghoshal, B., Sanfacon, H., 2014. Temperature-dependent symptom recovery in *Nicotiana benthamiana* plants infected with tomato ringspot virus is associated with reduced translation of viral RNA2 and requires ARGONAUTE 1. *Virology* 456–457, 188–197.
- Hemmer, C., Djennane, S., Ackerer, L., *et al.*, 2018. Nanobody-mediated resistance to Grapevine fanleaf virus in plants. *Plant Biotechnology Journal* 16, 660–671.
- Martin, I.R., Vigne, E., Berthold, F., *et al.*, 2018. The 50 distal amino acids of the 2A(HP) homing protein of Grapevine fanleaf virus elicit a hypersensitive reaction on *Nicotiana occidentalis*. *Molecular Plant Pathology* 19, 731–743.
- Nam, M., Koh, S., Kim, S.U., *et al.*, 2011. *Arabidopsis* TTR1 causes LRR-dependent lethal systemic necrosis, rather than systemic acquired resistance, to Tobacco ringspot virus. *Molecules and Cells* 32, 421–429.
- Osterbaan, L.J., Choi, J., Kenney, J., *et al.*, 2019. The identity of a single residue of the RNA-dependent RNA polymerase of grapevine fanleaf virus modulates vein clearing symptoms in *Nicotiana benthamiana*. *Molecular Plant Microbe Interactions* 32 (7), 790–801.
- Paudel, D.B., Ghoshal, B., Jossey, S., *et al.*, 2018. Expression and antiviral function of ARGONAUTE 2 in *Nicotiana benthamiana* plants infected with two isolates of tomato ringspot virus with varying degrees of virulence. *Virology* 524, 127–139.
- Sanfacon, H., 2013. Investigating the role of viral integral membrane proteins in promoting the assembly of nepovirus and comovirus replication factories. *Frontiers in Plant Science* 3, 313.
- Santovito, E., Mascia, T., Siddiqui, S.A., *et al.*, 2014. Infection cycle of Artichoke italian latent virus in tobacco plants: Meristem invasion and recovery from disease symptoms. *PLoS One* 9, e99446.
- Schellenberger, P., Sauter, C., Lorber, B., *et al.*, 2011. Structural insights into viral determinants of nematode mediated Grapevine fanleaf virus transmission. *PLoS Pathogens* 7, e1002034.
- Thompson, J.R., Dasgupta, I., Fuchs, M., *et al.*, 2017. ICTV virus taxonomy profile: Secoviridae. *Journal of General Virology* 98, 529–531.

Ophioviruses (*Aspiviridae*)

Anna M Vaira, Institute for Sustainable Plant Protection, National Research Council of Italy, Torino, Italy

John Hammond, Floral and Nursery Plants Research, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A.M. Vaira, R.G. Milne, Ophiovirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00645-2](https://doi.org/10.1016/B978-012374410-4.00645-2).

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

ELISA Enzyme-linked immunosorbent assay

EVEs Endogenous viral elements

GFP Green fluorescent protein

kb Kilobase

kDa Kilodalton

LAMP Loop mediated isothermal amplification

MP Movement protein

NES Nuclear export signal

NGS Next generation sequencing

NLS Nuclear localization signal

NP Nucleoprotein

nt Nucleotide(s)

ORF Open reading frame

PD Plasmodesmata

RdRp RNA-dependent RNA polymerase

RT-PCR Reverse transcriptase-polymerase chain reaction

SEL Size exclusion limit

TEM Transmission electron microscopy

UTR Un-translated region

vcRNA Viral complementary RNA

VIGS Virus-induced gene silencing

VSR Viral suppressor of RNA silencing

Glossary

Bunyavirales Order of viruses comprising species that have a segmented linear negative-sense RNA genome and a lipid envelope. The order includes 12 families:

Arenaviridae, *Cruliviridae*, *Fimoviridae*, *Hantaviridae*, *Leishbuviridae*, *Mypoviridae*, *Nairoviridae*, *Peribunyaviridae*, *Phasmaviridae*, *Phenuiviridae*, *Tospoviridae*, and *Wupedeviridae*.

Mononegavirales Order of viruses comprising species that have a linear, negative-sense RNA genome and a lipid envelope. The order includes 11 families: *Artoviridae*, *Bornaviridae*, *Filoviridae*, *Lispiviridae*, *Myonnaviridae*, *Nyamiviridae*, *Paramyxoviridae*, *Pneumoviridae*, *Rhabdoviridae*, *Sunviridae*, and *Xinmoviridae*.

Serpentovirales Order of viruses comprising 1 family, *Aspiviridae*, and 1 genus, *Ophiovirus*.

Introduction

The genus *Ophiovirus* is the only genus listed in the family *Aspiviridae* (formerly *Ophioviridae*), order *Serpentovirales*. The genus comprises plant-infecting virus species. In some cases ophioviruses are known to be the cause of well-known and “classical” major plant diseases, but in other cases the presence of the virus has not been linked to any specific symptom, due to almost invariable occurrence in mixed infections with other viruses. Ophioviruses occur in monocots and dicots, in vegetables, ornamentals, trees and shrubs, in the New and Old World, suggesting a well-adapted group of viruses. Where identified, the vectors of ophioviruses have proved to belong to *Olpidium* spp., soil-inhabiting chytrid fungi. Ophioviruses have been slow to emerge mainly because the virions are not easy to see in the electron microscope; indeed, the first ophiovirus particle was observed only in 1988 and its morphology understood in 1994. The genus name derives from the Greek word *ophis*, meaning snake, in reference to the serpentine appearance of the virus particles and the family name derives from the Latin word *aspis*, with the same meaning.

Taxonomy, Phylogeny and Evolution

There are currently seven species in the genus *Ophiovirus* (Table 1). The accepted species in the genus are: *Citrus psorosis ophiovirus* (the type species) (Citrus psorosis virus, CPsV), *Mirafiori lettuce big-vein ophiovirus* (Mirafiori lettuce big-vein virus, MLBVV), *Lettuce ring necrosis ophiovirus* (Lettuce ring necrosis virus, LRNV), *Ranunculus white mottle ophiovirus* (Ranunculus white mottle virus, RWMV), *Tulip mild mottle mosaic ophiovirus* (Tulip mild mottle mosaic virus, TMMMV), *Freesia sneak ophiovirus* (Freesia sneak virus, FreSV) and *Blueberry mosaic associated ophiovirus* (Blueberry mosaic associated virus, BlMaV). To date, the genomes of four species, MLBVV, CPsV, LRNV and BlMaV, are fully sequenced. For species demarcation within the genus, different criteria have been considered: the different coat protein (CP) sizes; absence of, or distant serological relationship between the CPs; differences in natural host range; and different number, organization, and/or size of genome segments. From available data, there is 95%–100%

Table 1 Virus members in the Genus *Ophiovirus* with relevant NCBI database accession numbers

Species (alternative name)	Acronym	Accession numbers ^a
Blueberry mosaic associated ophiovirus	BIMaV	NC036635, NC024476, NC036634
Citrus psorosis ophiovirus (Citrus ring-spot virus)	CPsV	NC006314, NC006315, NC006316
Freesia sneak ophiovirus (Freesia ophiovirus)	FreSV	MN365016, DQ885455
Lettuce ring necrosis ophiovirus	LRNV	NC006051, NC006052, NC006053, NC006054
Mirafiori lettuce big-vein ophiovirus (Mirafiori lettuce virus)	MLBVV	AF525933, AF525934, AF525935, AF525936
Ranunculus white mottle ophiovirus	RWMV	NC043389, AY542957
Tulip mild mottle mosaic ophiovirus	TMMMV	NC043390

^aComplete genome sequences are reported when available.

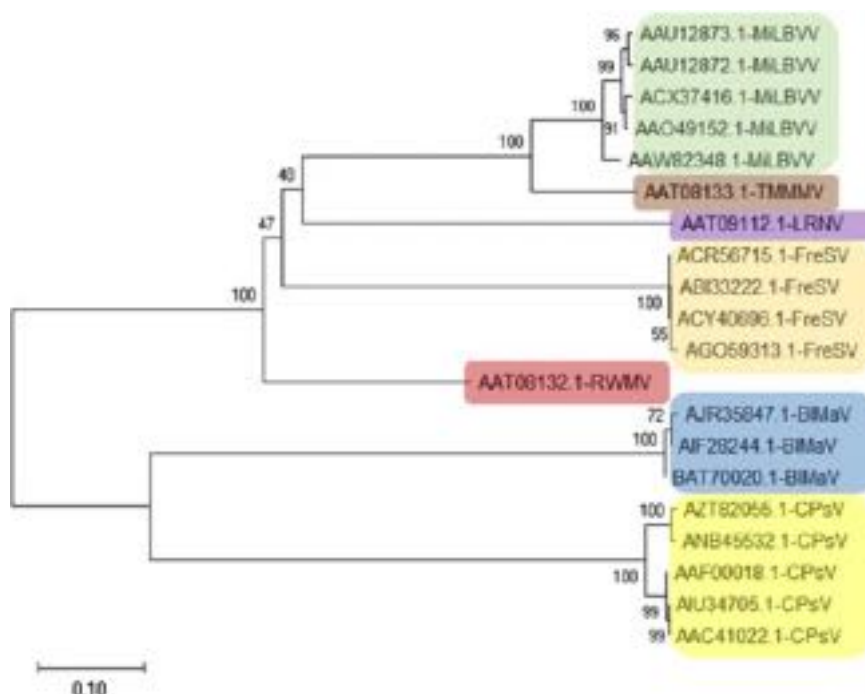


Fig. 1 Unrooted phylogenetic tree of selected ophiovirus isolates from different geographical origin, based on their NP amino acid sequences. The evolutionary history was inferred using the Neighbor-Joining method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The evolutionary distances were computed using the Poisson correction. Evolutionary analyses were conducted in MEGA X. The TMMMV, RWMV and BIMaV AJR35847 nucleoprotein sequences analyzed are partial. Rectangles grouping isolates represent established viral species.

identity among complete CP aa sequences of isolates belonging to the same species detected in different world areas, in particular for MLBVV, CPsV, BIMaV and FreSV. The percentage of identity falls to 28%–59% for interspecies alignments, with the exception of the identity between MLBVV and TMMMV CP sequences, which is slightly above 80%, confirming the closeness of these two viruses. They are currently considered as two different species at least because of different host ranges; more information on TMMMV sequences is needed prior to any revision of their taxonomic position.

The percentage of identity/similarity between CP aa sequences may also be proposed as a further tool for species demarcation as several full sequences are now available. A phylogenetic tree based on alignment of representative CP aa sequences is shown in Fig. 1. CPsV and BIMaV, presently the only ophioviruses infecting perennial hosts with asexual means of propagation, are more correlated. These differences, if confirmed and reinforced, could lead to considering placement of these species in a different genus within the family Aspiviridae. RT-PCR amplification, using a pair of degenerate primers (see below), is currently a good tool for detecting and identifying official and potential new species within the genus, providing also indications of the taxonomic position of newly diagnosed isolates. The analysis of RdRp aa sequences shows the Aspiviridae family as a monophyletic group among negative-stranded RNA virus families of orders Bunyavirales and Mononegavirales, newly assigned to the order Serpentovirales. Recently, due to the wide use

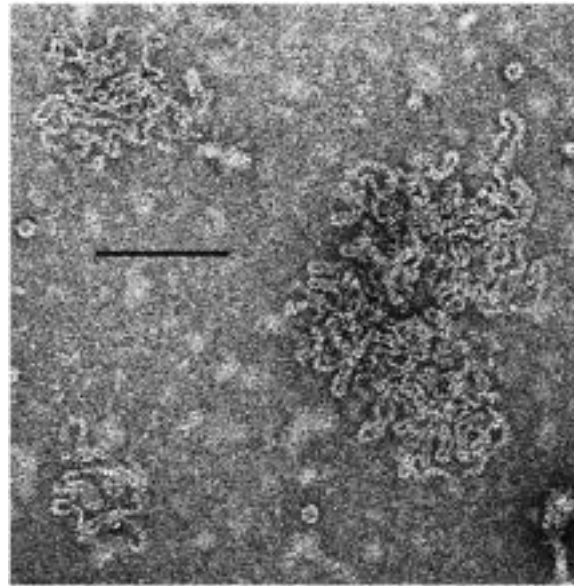


Fig. 2 Negative contrast electron micrograph (uranyl acetate) of partially purified virion preparation from field lettuce showing big-vein symptoms. Note large and small ophiovirus particles. The bar represents 100 nm. Courtesy of RG Milne, IPSP-CNR, Torino, Italy.

of NGS technologies on complex samples, reports regarding aspivirus-like isolates in fungi note similarities of their RdRp protein sequences to those of LRNV and MLBVV, but further characterization is needed to assess their putative taxonomic position in the family *Aspiviridae*. Partial ophiovirus-like sequences have also been reported from NGS studies of *Plantago lanceolata*. Interestingly, a recent report describes the presence of endogenous viral elements (EVEs) putatively derived from an ophiovirus in the genome of the eelgrass *Zostera marina*. EVEs are sequences in eukaryotic genomes that are derived from the ancestral integration of viral sequences into germline cells; in this case the ancestral sequence is represented by the ophiovirus CP ORF; this finding, if confirmed, would extend the *Aspiviridae* host range to seagrasses and expand ophiovirus evolutionary history.

Virion Structure

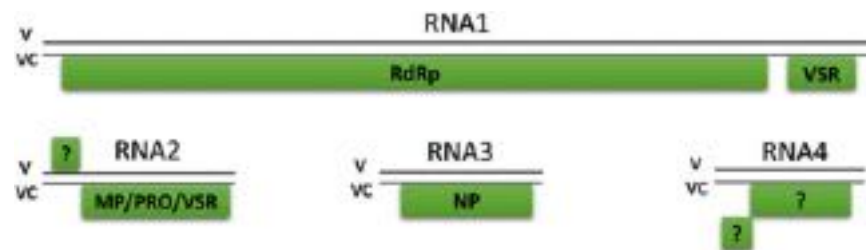
The virions are naked filamentous nucleocapsids about 3 nm in diameter forming circularized structures of different lengths, presumably corresponding to the different lengths of the genomic segments. The virions appear to form internally coiled circles that in some cases can collapse into pseudolinear duplex structures (Fig. 2) and can be difficult to visualize in TEM assays due to a low contrast with the background. There is no evidence of a lipoprotein envelope and the mechanism underlining the apparent circularization of ophiovirus particles remains incompletely explained. Ophiovirus particles can be purified from infected plant tissue using differential centrifugation, clarifying the suspension using PEG6000 and sodium chloride, and density gradient ultracentrifugation. Particles are unstable in cesium chloride but the particle structure remains intact in cesium sulfate gradients (buoyant density is 1.22 g/cm³ for RWMV and MLBVV). Particle structure survives limited treatment with organic solvents and nonionic or zwitterionic detergents.

Genome Organization and Replication

The ophiovirus genome is ssRNA, 11.3–12.5 kb in size, divided into three or four segments (RNAs 1–4) of mainly negative sense. Positive- and negative-sense RNAs are individually encapsidated in ophiovirus particles, and the more abundant negative sense RNA is referred to as viral RNA (vRNA), while the complementary RNA will be described as viral complementary (vcRNA). The 3' termini of the vRNAs show sequences of 6–13 nt conserved within the RNAs of each species: A₇GUAUC for CPsV, A_{4–6}UAAUC for MLBVV, A₇GUAUCA and A₃UA₃GUAUCA for LRNV; A₂UAUC for BImaV; A₄UGUAUC for FreSV (preliminary results); these may be involved in the recognition of the RNAs by the RdRp. The 5' termini of BImaV RNAs were shown to fold into conserved stem-loop structures. These structures may be involved in packaging of genomic RNAs or in the long-distance interactions for transcription and replication. The pseudo-circularized appearance of ophiovirus particles suggests the presence of panhandle structures at genomic RNA ends, but no conclusive results were inferred for the four fully sequenced species. In CPsV, the first 12 nt of 5' end of vcRNAs were found almost identical in the three RNAs, but unexpectedly no identity among the three RNAs is found at their 3' ends, and no self-complementary panhandle structures between the 3' and 5' ends of each RNA could be hypothesized. In the case of MLBVV and LRNV, both 5' and 3' ends are highly or partially conserved among the four

Table 2 Details of genomic organization of fully sequenced viral segments for viruses of the genus Ophiovirus

<i>Virus</i>	<i>RNA1</i>	<i>RNA2</i>	<i>RNA3</i>	<i>RNA4</i>
BIMaV	7963 nt 2 ORFs in vcRNA ORF1: 23 kDa ORF2: 272 kDa	1934 nt 1 ORF in vcRNA ORF1: 58 kDa	1570 nt 1 ORF in vcRNA ORF1: 50 kDa	–
CPsV	8186 nt 2 ORFs in vcRNA ORF1: 24 kDa ORF2: 280 kDa	1645 nt 1 ORF in vcRNA ORF1: 54 kDa	1447 nt 1 ORF in vcRNA ORF1: 49 kDa	–
LRNV	7651 nt 2 ORFs in vcRNA ORF1: 22 kDa ORF2: 261 kDa	1830 nt 1 ORF in vcRNA ORF1: 50 kDa	1527 nt 1 ORF in vcRNA ORF1: 48 kDa	1417 nt 1 ORF in vcRNA ORF1: 37 kDa
MLBVV	7794 nt 2 ORFs in vcRNA ORF1: 25 kDa ORF2: 263 kDa	1788 nt 1 ORF in vRNA ORF1: 10 kDa (?) 1 ORF in vcRNA ORF1: 55 kDa	1515 nt 1 ORF in vcRNA ORF1: 49 kDa	1402 nt 2 ORFs in vcRNA ORF1: 37 kDa ORF2: 11 kDa
FreSV	nd ^a	1722 nt 1 ORF in vRNA ORF1: 54 kDa	nd 1 ORF in vcRNA ORF1: 48.8 kDa	?

^anot fully determined; – not detected.**Fig. 3** Genome organization of ophioviruses. The four genomic RNAs of MLBVV, the first ophiovirus to be full sequenced, are represented to scale. An RNA4 is not described in all species, see text and [Table 2](#) for details; v-RNA and vc-RNA indicate the viral and the viral complementary RNAs, respectively. Boxes indicate ORFs and the function of the encoded protein, if known, is reported.

viral RNAs but are not predicted to form panhandle structures; one hypothesis is that their RNAs ends fold into structures resembling the “corkscrew” conformation of RNA termini of members of the family *Orthomyxoviridae*, but this structure was not found for CPsV. Regarding BIMaV no complementarity was found between the 5′ and 3′ ends of each genomic RNA, as found for CPsV; in conclusion, alternative explanations for the pseudo-circular structure of ophiovirus particles are required.

The genome organization of ophioviruses is summarized in [Table 2](#) and [Fig. 3](#). All RNA segments have 5′ and 3′ untranslated regions. RNA1, the largest genome segment, is negative sense and contains two ORFs separated by an AU-rich intergenic region. The presence of two ORFs, the longest encoding the RdRp, with the same polarity in the largest RNA is a distinct genomic feature of *Aspiviridae* when compared to all other segmented negative-strand viruses. RNA2 contains one ORF in the negative strand in all four fully sequenced viruses. An additional minor ORF, in the virus-sense strand, is present for MLBVV, but to date there is no evidence for expression of the predicted 10 kDa protein. RNA3 contains one ORF in the vcRNA in all viruses, which has been identified as coding for the nucleocapsid protein. Up to now, an RNA4 has been found only in MLBVV and LRNV. Both encode a 37 kDa protein in the vc strand, with only MLBVV having a smaller (10.6K) ORF which lacks an initiation codon and is potentially expressed by a +1 translational frameshift near the 3′ end of the 37K ORF. A slippery sequence, GGGAAAU, can be recognized immediately in front of the UGA stop codon of the 37 kDa ORF. This unique feature of ORF positioning has been shown for two different isolates of MLBVV, and therefore seems not to be caused by cloning artifacts.

RNAs 1, 2, and 3 of the other partially sequenced species are similar in size to those described. No RNA4 has been found in CPsV, RWMV, BIMaV and TMMV, in spite of multiple NGS reactions, or PCRs using primers designed from the NGS data and conserved regions of the MLBVV/LRNV 37K proteins, which have been performed at least for BIMaV; its presence in FreSV is under study. Analysis for the presence of sub-genomic RNAs has not extensively been pursued for all species, even though several studies have used minus-strand and plus-strand probes in Northern blot hybridization using RNA extracted from infected tissues as well as virus particles, to obtain information regarding the polarity of encapsidated viral RNA. No subgenomic RNAs were ever validated for any of RNA1, 2, or 3 of CPsV. Reassortment between genome segments from different isolates may occur in plants infected with more than one isolate, as it has been shown for BIMaV.

The presence of one or more NLS reported for the RdRp protein (see below) suggests that replication may occur in the nucleus; however, gold immunolabeling and electron microscopy of RWMV-infected tissues showed accumulation of CP in the cytoplasm.

Properties and Functions of Gene Products

Two proteins are encoded in the viral complementary (vc) strand of *RNA1*, separated by an intergenic region of about 100 nt; the first of these is a protein of approximately 22–25 kDa. For CPsV and MLBVV this protein has been shown to function as viral suppressor of RNA silencing (VSR), and it is likely that the equivalent proteins of other ophioviruses have the same function; consistent with this function, a CPsV 24K:GFP fusion protein was found to accumulate in the nucleus and probably the nucleolus despite having an identified NES. Nucleolar import of CPsV 24K may be aided by an N-terminal proximal “GW” motif, while the BImaV 23K protein has an identified bipartite NLS not seen in the other ophioviruses for which the small protein from *RNA1* has been characterized. Indeed, the BImaV 23K has low identity to the equivalent proteins of CPsV, MLBVV, and LRNV. The CPsV 24K and MLBVV 25K proteins have both been demonstrated to bind long (114 bp) but not short (21 bp) dsRNAs, a property shared with their 54–55K MPs which have also been identified as VSRs (see below), and which also affect miRNA maturation.

The second, and major protein encoded by *RNA1* is the viral RdRp, which is of 261–280 kDa. The RdRp of each virus has the signature motifs present in RdRp of viruses of the order *Mononegavirales*, having a core polymerase module of about 240 aa residues which includes five conserved motifs; pre-motif A, plus motifs A to D. Of these, motif C is absolutely conserved between the five viruses for which the core polymerase sequence is available (CPsV, MLBVV, LRNV, RWMV, and BImaV), and includes the presumed RdRp active site “SDD” triad found in all negative strand RNA viruses with segmented genomes, including ophioviruses. In motif D, a Gly residue is conserved between MLBVV, LRNV, and RWMV, but is substituted by an Asp residue in CPsV and BImaV. The RdRps of CPsV, MLBVV, RWMV, and BImaV all have identified bipartite NLSs, with CPsV having dual NLSs, and the other viruses having only one, though no NLS has been reported for the RdRp of LRNV. In most viruses with a negative sense RNA genome, the RdRp is also incorporated in the virions, as in the absence of translation from the incoming genomic RNA, the RdRp is required to initiate infection. However, a variable proportion of positive (vc) strand RNA of ophioviruses is also encapsidated, which may overcome the necessity for other viruses with negative sense genomes to incorporate the RdRp into virions.

All of the characterized ophioviruses produce a protein of 50–58 kDa translated from the *RNA2* vc strand, initially identified as a movement protein belonging to the “30K superfamily” of MPs, and later shown to also include VSR activity. The MPs of CPsV and MLBVV interact with PD (presumably increasing its SEL, a characteristic of many different MPs), and with cellular microtubules, and also spread into adjacent cells. CPsV and MLBVV MPs interact with their homologous nucleocapsid protein (NP); interaction of MP with NP/coat protein is a common feature of MPs of other viral genera and families.

The predicted structures of ophiovirus MPs align with each other, and have a similar organization comprising a core domain, common to all MPs of the “30K superfamily”, including one α -helix, seven β -strands, and the conserved Aspartic acid (D) residue associated with increased SEL of PD and full MP function. The C-terminal domain has conserved secondary structure potentially involved in RNA binding and interactions with host proteins, and a conserved DTG tripeptide not observed in any other “30K superfamily” MPs. The 54K MP of CPsV is autocatalytically cleaved by an aspartic protease to generate an N-terminal 34K protein including the “30K core domain”, which forms tubules at PD to allow virus movement between cells, and a 20K product. The 20K includes the protease activity within the MP conserved C-terminal domain, with the absolutely conserved DTG tripeptide near the N-terminus of this domain; the cleavage site occurs about 30 residues upstream. The D residue is critical for cleavage; however, a high proportion of the CPsV 54K protein remains unprocessed even in native infections. Cleavage affects MP subcellular localization, and is required for efficient nuclear localization (presumably of the 20K cleavage product). Uncleaved 54K accumulated to a greater extent at chloroplasts, and although also observed at PD, no tubular structures were observed; whereas 34K expressed alone resulted in tubule production at PD. Chloroplast localization is presumably due to a predicted 21 aa residue N-terminal chloroplast transit peptide. The 34K cleavage product complements a movement-deficient TMV more efficiently than the full-length 54K, confirming movement function is due to the 34K. Both the 34K and 20K cleavage products form homodimers, consistent with 34K forming tubules, and the presumed dimeric catalytic triad for aspartic protease activity of the 20K product. Most viruses with tubule-forming MPs have isometric virions; it has been suggested that interaction between CPsV MP and NP may allow guidance of virions (or a ribonucleoprotein complex of viral RNA with both MP and NP) through tubules linking adjacent cells through PD. To date, only protease activity has been ascribed to the 20K cleavage product. Both (54K) MP and the *RNA1* 24K protein each have VSR activity, which is enhanced when the two proteins are co-expressed; each presumably functions at a different step of the RNAi pathway. As a 20K:GFP fusion protein traffics to the nucleus, the 20K cleavage product may have the VSR activity originally assigned to the 54K product.

It has been proposed that the CPsV 54K protein should be renamed “MP-PRO”, with the 30K cleavage product identified as “MP”, and the 20K cleavage product as “PRO”. This suggested nomenclature may be further modified if the VSR activity first associated with the uncleaved 54K protein is attributed to a specific cleavage product.

A minor protein of about 10 kDa is apparently encoded in the viral (v) strand of MLBVV *RNA2* and this would represent a unique case among the known ophiovirus genome structures, but its significance is still unclear.

RNA3 of all characterized ophioviruses encodes a single protein of 43–50 kDa, which is the viral nucleoprotein, alternatively described as the “coat protein” (CP) forming at least the bulk of the filamentous virions together with the genomic RNAs. The NP protein sizes of different ophioviruses ranges between 43 and 50 kDa (see [Table 2](#)), with FreSV NP of 48.4 kDa, and the sizes of RWMV (~43 kDa) and TMMMV (~47 kDa) NPs estimated by western blotting rather than from partial NP sequence data.

Ophiovirus NPs are relatively poor antigens. The antisera or, in one case, monoclonal antibodies, have been produced using purified virus preparations or recombinant protein. Western blots show that the NPs of RWMV, TMMMV, MLBVV, and LRNV are slightly to moderately related; CPsV NP appears to be unrelated to the others; no clear information is available for FreSV or BImaV. Genetic variability of the NP gene has been extensively studied in CPsV and MLBVV. Variability of the NP gene of CPsV was assessed serologically, and by sequence analysis of two regions located in the 3' and 5' halves of the gene. CPsV serological variability yielded 14 reaction patterns, without correlation between serogroups and specific aa sequences, field location, or citrus cultivar. Sequence analysis showed limited nucleotide diversity in the NP gene within the population, with diversity slightly higher in the 5' region. The ratio between non-synonymous and synonymous substitutions (dN/dS) for the two regions indicated a negative selective pressure for aa changes, with greater selection pressure at the 3' end. When the entire CP sequences were considered, two clusters were identified, one comprising 19 isolates from Italy and an isolate from Spain and a second one containing only the Florida CPsV isolate. Phylogenetic analysis of MLBVV NP genes has revealed two distinct clades, A and B, further divided into subclades A1 and A2, B1 and B2; however, this grouping was not correlated with symptom severity on lettuce or the geographic origin of the isolates, except that all Australian isolates fell into subclades A1 and A2; whereas European isolates were found in all four subclades, only European isolates fell into B1 and B2, including all reported isolates from *Sonchus* spp. in subclade B2. Whether these two clades or subclades show different characteristics with regard to virulence in indicator plants or serological relationships remains to be determined; as with CPsV, a low ratio of dN/dS was estimated for all MLBVV isolates and between the two subgroups, implying negative selection pressure for aa changes. A similar study of BImaV MP and NP sequences of isolates from British Columbia (Canada), Slovenia, and several regions within the United States found a low ratio of dN/dS for both genes. Under natural conditions, genetic stability of ophiovirus NPs seems to be the rule rather than the exception, perhaps due to the necessity both for structural integrity of the virions, and interaction with host or vector proteins.

The NP encapsidates both negative (v) and positive (vc) strands of the genomic RNA. Each NP includes a domain of almost 180 aa residues that is conserved among the NPs of viruses with negative strand ssRNA genomes, and is rich in α -helices. The NP of MLBVV and partial NP sequence of TMMMV share about 80% identity, showing the closest relationship between ophiovirus species; although distant serological relationships between MLBVV and TMMMV, and MLBVV and RWMV indicate the presence of some conserved epitopes between these viruses, no serological relationship was detected between these viruses and CPsV. CPsV and BImaV NPs share only 38% identity, despite their current status as the only ophioviruses infecting woody hosts.

A putative leucine zipper motif of six (CPsV) or four turns (MLBVV, TMMMV, FreSV, LRNV, and RWMV) followed by an SRCK tetrapeptide has been identified in the NPs of the six ophioviruses for which all or most of the NP sequence is available; for the complete NP of BImaV, a five turn leucine zipper motif followed by a TRCK tetrapeptide at the same spacing can be identified. Leucine zippers are associated with protein oligomerization and genome encapsidation, consistent with NP self-interaction and role of the NP in virion formation. Whether the SRCK/TRCK tetrapeptide may be involved in a specific interaction with zoospores of the various *Olpidium* spp. known to be vectors of most ophiovirus species remains to be determined.

An RNA4 has been identified only for MLBVV and LRNV; the 37 kDa proteins of MLBVV and LRNV share just over 40% identity, with a higher proportion of conserved residues in the N-terminal 60 aa residues than any other portion of the protein, suggesting that this is a functional domain; both the MLBVV and LRNV 37K proteins are predicted to have three trans-membrane domains, with the N-terminus 'outside', and the C-terminus "inside". Expression of the putative 10.6K frame-shift ORF present only in RNA4 of MLBVV has not been demonstrated. To date no function has been assigned to either of the ORFs encoded by RNA4.

Pathogenicity and Geographic Distribution

Psorosis is a graft-transmissible disease of citrus characterized primarily by bark scaling of the trunk and main branches of various *Citrus* species (Fig. 4). These symptoms were first reported from Florida and California, USA in the 1890s, and in the great majority of cases psorosis symptoms are associated with infection by CPsV. However, in a small proportion of trees with "psorosis-like" bark-scaling CPsV is not detected, and the idea of a non-CPsV psorosis-like disease of unknown etiology is emerging. In the past, different kinds of symptoms were described: "psorosis A" (PsA), characterized by causing bark-scaling in trunk and limbs of infected field trees, and staining of the wood; "psorosis B" (PsB), causing rampant scaling of thin branches in field trees and chlorotic blotches in old leaves with gummy pustules in the underside; "ringspot", characterized by presence of chlorotic blotches and rings in the old leaves of inoculated seedlings but apparently no specific symptoms on infected field trees or in other cases, chlorotic flecks and ringspots on leaves, and trunk and fruit symptoms. All these appear to be caused by CPsV. CPsV has been documented in many citrus production areas, including the USA, Argentina, Uruguay, Spain, Italy, Turkey, Lebanon, Egypt, Morocco, Tunisia, China, and Australia. The disease caused by CPsV has been brought under control in most advanced citrus-growing countries through implementation of rigorous indexing and quarantine. In Argentina it remains a severe problem, and it is still reported in the Mediterranean areas, while in citrus-growing parts of Asia the disease may well be widespread, although rigorous testing for the presence or absence of CPsV has generally been lacking. Among citrus types, sweet orange, grapefruit, and mandarin are most severely affected, whereas sour orange and lemon usually show foliar symptoms without obvious bark lesions. Infection with PsA isolates offers some cross-protection against the more severe symptoms induced by PsB isolates (also sometimes referred to as "Citrus ringspot" isolates). Disease severity is also affected by temperature, with cool temperatures increasing the probability of a shock reaction in the first growth flush following inoculation, whereas warmer temperatures may result in masking of foliar symptoms. Increased temperatures appear to allow greater RNA silencing by the host, resulting in lower viral accumulation.



Fig. 4 Bark scaling in citrus, typical of severe psorosis.

Lettuce big-vein disease was first reported in the United States in 1934 and occurs in all major lettuce producing areas in the world. The disease becomes more serious during cooler periods of the year. The main symptoms are vein-banding in the leaves, due to zones parallel to the vein cleared of chlorophyll (**Fig. 5**); there is associated leaf distortion, delayed head formation, and decreased head size. The causal agent of the disease, long known to be soil-transmitted, was first identified as "Lettuce big-vein virus" (LBVV) (genus *Varicosavirus*); more recently, re-evaluation of the etiology was necessary as a second less easily detected virus, MLBVV, was found in lettuces with big-vein symptoms. Following experimental inoculation of the two viruses together and separately, MLBVV was shown to be the etiological agent of big-vein, while LBVV, renamed as Lettuce big-vein associated virus (LBVaV), apparently plays no part in the disease, although it is almost always present, and is, like MLBVV, transmitted by *Olpidium*. Studies on symptom development in the field have confirmed that both viruses very commonly occur together in lettuce crops. MLBVV likely occurs worldwide; it is reported in California (USA), Brazil, Chile, France, Germany, Italy, Spain, the Netherlands, England, Denmark, Turkey, South Africa, Saudi Arabia, Japan, Australia, and New Zealand. To date, natural infection has been reported only in cultivars of *Lactuca sativa*, endive (*Cichorium endivia*), and two species of sowthistle (*Sonchus asper* and *S. oleraceus*).

LRNV is closely associated with lettuce ring necrosis disease, first described in the Netherlands and in Belgium as "kring necrosis" and also in France as "*maladie des taches orangées*", in the 1980s. It is an increasingly important disease of butterhead lettuce crops in Europe. Definitive proof that LRNV is the cause of ring necrosis is still awaited. In southern France the disease is observed primarily in winter lettuces (September–January) under plastic or glass, when the crop is maturing, the day length is short, and both light intensity and temperature are low. Symptoms depend on the lettuce type and environmental conditions, and mainly consist of necrotic rings and ring-like patterns on leaves, which may render the product unmarketable. LRNV is often found together with MLBVV (and, of course, LBVaV) with which it shares host and vector. The virus has also been reported in California.



Fig. 5 Leaf of butterhead lettuce showing big-vein symptoms. Close-up showing vein-banding.



Fig. 6 Shoot of blueberry infected with BIMA virus showing part of the range of symptoms often observed, which include chlorotic fleck and mottling, strong chlorotic mosaic, and combinations of yellow mosaic with areas of pink or orange coloration.

BIMA virus was initially described from an infected plant from Arkansas, USA, showing symptoms of mild to brilliant foliar mottle and mosaic (Fig. 6) observed since the 1950s in blueberry production areas across North America, Argentina, Chile, Europe, New Zealand, and South Africa. The disease is associated with late ripening, reduced yield, and poor quality fruit. Some leaves may turn partly or completely yellowish green, yellow, orange, or pink. BIMA virus has been specifically identified by molecular methods from several states in the USA, from British Columbia (Canada), Slovenia, Serbia, Turkey, and Japan, with the Slovenian isolates being most closely related to isolates from New Jersey, USA. Most of the Slovenian isolates were propagated from plants originally imported from New Jersey.

Necrotic disorders of freesia (*Freesia refracta* hyb., family Iridaceae), known as “leaf necrosis” and “severe leaf necrosis”, were first described in freesia crops in the Netherlands before 1970, and a similar disease, named “freesia streak”, was reported in England and Germany in the same years. The ‘severe leaf necrosis’ appeared to be caused by mixed infection with the potyvirus freesia mosaic virus and a virus with varicosavirus morphology and the same mode of transmission, that has been tentatively named freesia leaf necrosis virus. More recently in northwest Italy, a necrotic disease of freesia spread and caused considerable economic losses (Fig. 7), even though it was present since 1989. Both in Italy and in the Netherlands, the disease has been linked to the presence of an ophiovirus, named freesia sneak virus (FreSV). Typical symptoms are chlorotic inter-veinal spots and streaks at the leaf tips that expand downwards and turn necrotic, and may vary according to cultivar and climate. FreSV has since also been identified in freesia in England, Germany, the Netherlands, Bulgaria, New Zealand, Korea, Japan, and the USA, and from *Lachenalia* hybrids (family Asparagaceae) from South Africa.

RWMV was initially reported in two species, ranunculus (*Ranunculus asiaticus* hyb.) and anemone (*Anemone coronaria*) in northwest Italy since the 1990s. The pathogenic impact is uncertain as it was almost always found in mixed infection with other viruses commonly infecting the two species. The symptom description “white mottle” present in its name derives from the bright white mottling symptoms consistently observed on *Nicotiana benthamiana* leaves mechanically infected by RWMV. In some cases the bright mottle has also been observed in naturally infected ranunculus, always in mixed infection. Indexing for the presence of RWMV in ranunculus crops done in 1996 showed an incidence of 2.5% among symptomatic plants; furthermore, the very few plants apparently infected only with RWMV did not show any distinctive symptoms. Infection of ranunculus seedlings by mechanical inoculation results in limited necrosis and



Fig. 7 Symptoms of freesia necrotic disease on freesia leaves. Courtesy of D. Salvi, Studio Ferrari Salvi, Sanremo (IM) Italy.



Fig. 8 TMMMV symptoms on tulip (healthy tulip on the left). Courtesy of T. Morikawa, Toyama Agric. Res. Center, Tonami, Toyama, Japan.

deformation of stems and leaves; however, it is one of the most prevalent viruses in ranunculus crops in Italy, possibly as a result of establishment of the *Olpidium* vector in production fields. Infection of anemone was also typically in mixed infections, leaving the pathogenic potential of RWMV unclear. A report published by Horticulture Australia Ltd notes RWMV infecting pepper showing vein yellowing disease in Australia; the disease showing incidence up to 30% is widespread on the North Adelaide Plain but does not appear to cause significant yield loss or reduction in fruit quality. The symptoms are easily seen in young plants and fading as plants mature. More recently RWMV was detected in *Pittosporum tobira* in Italy, in mixed infection with Eggplant mottled dwarf rhabdovirus.

Tulip mild mottle mosaic disease, caused by TMMMV, is one of the most serious diseases in some bulb-producing areas of Japan (which produces bulbs on a large scale for the Southeast Asian market) and has been reported since 1979. Symptoms on tulips include color-attenuating mottle on flower buds and color-increasing streaks on petals ([Fig. 8](#)); mild chlorotic mottle and mosaic may appear along the leaf veins. TMMMV infection is apparently restricted to *Tulipa* species and cultivars, and has not been reported in other geographical areas.

Diagnosis

A pair of degenerate primers was originally designed from a conserved region of RNA1 (in the RdRp gene, located outside the sequence coding the polymerase domain). These primers, used in RT-PCR experiments, amplify a 136 bp DNA fragment from all recognized ophiovirus species and not from representative plant-infecting species of *Rhabdoviridae*, *Tenuivirus*, *Orthospovirus* and *Varicosavirus*. These genus-specific primers, by a simple RT-PCR assay, may lead to the detection of new ophioviruses, with the establishment, at least provisionally, of their taxonomic position, through phylogenetic analysis of the predicted 45 aa translation of the amplified fragment. When an ophiovirus infection is suspected, RT-PCR and qRT-PCR with specific primers are the most commonly used and sensitive method for specific diagnosis, both in plants and in the natural vector, even if ELISA using polyclonal antisera is still used for mass diagnosis. An RT-LAMP protocol was also developed for MLBVV; compared with DAS-ELISA and RT-PCR, RT-LAMP and IC-RT-LAMP (Immuno-Capture of particles) seem to have higher sensitivity (100-fold) but

similar specificity, with the advantage of a shorter assay time and no need for RNA extraction in IC-RT-LAMP. Recently NGS technology gained high importance in virus diagnosis, and the “small RNA-Seq” technique has been used to characterize the causal viruses of lettuce big vein disease, confirming the crucial involvement of MLBVV. NGS technology has a huge potential and may be easily used to find new hosts for ophiovirus species (as happened for the RWMV infection of *Pittosporum tobira*) and help to find new ones (as happened for the disclosure of blueberry infection by BLMaV, the most recently recognized official ophiovirus).

Transmission, Experimental Host-Range and Control

Most ophiovirus infections in nature are recognized as soil-borne. The natural vector of most ophiovirus species has been identified as *Olpidium* spp., ubiquitous obligate root-infecting fungal parasites (*Chytridiomycota*, chytrids) with cosmopolitan distribution. At least two species, *O. brassicae* and *O. virulentus* seem to be involved in ophiovirus transmission. Studies on MLBVV and *O. virulentus* suggest that the virus is carried internally (*in vivo*) in zoospores and resting sporangia. Virus-free zoospores acquire the virus during the vegetative part of their life cycle in the roots of virus-infected host plants. Zoospores released from infected sporangia carry the virus internally in their protoplasts and transmit it to healthy roots. *In vivo* acquisition is also supported given that resting spores can remain viruliferous in soil for decades in the absence of a growing host plant. The mechanism by which MLBVV and likely other ophioviruses are acquired and released into plant hosts by *Olpidium* is unknown; the possibility of ophioviruses multiplying in the fungal vector has not been reported to date. *Olpidium* spp. transmission has been proved for TMMMV, MLBVV (*O. virulentus*), LRNV, FreSV (*O. brassicae*) and BLMaV (*O. virulentus*). No information on natural vectors is available for RWMV but soil transmission is likely. CPsV is commonly transmitted by vegetative propagation and no natural vectors have been so far identified; in some cases natural tree-to-tree spread of psorosis in limited citrus areas has been reported, but the spatial patterns would suggest a hypothetical aerial vector instead of a soil-borne one. However, CPsV has been detected in the roots of infected grapefruit trees, and in zoospores of an unidentified *Olpidium*-like fungus released from sporangia in infected roots, but not from healthy controls.

The thick cell-walled nature of *Olpidium* spp. resting spores and their persistence in the soil for greater than 20 years makes ophiovirus-transmitted disease control most challenging, especially after the banning (Montreal Protocol) of use of chemicals for soil sterilization due to deleterious effects on the environment. Environmentally more sensitive control methods can be applied in order to reduce the spread and severity of the diseases. For nurseries, for example, disinfection of soil and potting mix, surface sterilization of seed, and the use of filtered/UV-treated water and nutrients, and clean production areas may all be important in the management of MLBVV and LRNV in lettuce. Crop covers, which affect the favorable environmental conditions for the viruses by lowering soil temperature and raising air temperature, were also shown to reduce the disease symptoms in lettuce. Moreover, the use of resistant or tolerant crops may be a good choice. In Japan, the use of resistant tulip cultivars is the most important component of managing TMMMV disease, as it can be highly effective and has no deleterious effect on the environment; resistance assays have allowed researchers to identify highly resistant tulip lines and use them for breeding new resistant cultivars. In the case of CPsV, control of the sanitary status of mother plants for producing propagating material is essential. Shoot-tip grafting *in vitro* associated with thermotherapy or somatic embryogenesis from stigma and style cultures have been successfully used to eliminate CPsV from plant propagating material. Preliminary results of a shoot apical meristem culture procedure performed on RWMV-infected ranunculus was also successfully used for ophiovirus eradication: the use of seeds and *in vitro* virus eradication can also be applied to currently vegetatively propagated ornamentals like freesia, ranunculus and anemone, to control ophiovirus infection. Genetic engineering of citrus and lettuce in recent years has shown promising results for obtaining ophiovirus resistance; several pineapple sweet orange plant lines transformed with a hairpin construct derived from the CPsV nucleoprotein gene showed varying responses from tolerance to full resistance traits toward both psorosis variants, a strategy that seems promising for a biotech product aimed at eradicating psorosis. Stable, marker-free and MLBVV-resistant lettuce lines were also engineered, in which MLBVV nucleoprotein gene is controlled by plant regulators; the new lines can be used as resistant cultivars or parental source for breeding.

All of the characterized ophioviruses are mechanically transmissible to a limited range of test plants, including Chenopodiaceae and Solanaceae, inducing local lesions; in some cases systemic mottle or apical leaf deformation and necrosis can occur. Systemically infected test plants can, in many instances, be advantageous for virus characterization, because of the enhanced virus multiplication and the possibility to obtain a high amount of purified virus. Chilling of source tissue, buffers, and equipment, together with additives such as antioxidants, enhance mechanical transmission of most, if not all ophioviruses. *Nicotiana benthamiana*, *N. clevelandii*, *N. occidentalis*, and *Chenopodium quinoa* are among the bioassay plants most often reported to be susceptible (Table 3). However, mechanical transmission is often difficult, and serial transfer between test plants is not always successful; there are also differences in ability or efficiency of mechanical transmission between isolates of CPsV. Transmission by the natural vectors, *Olpidium* spp. (*O. brassicae*, *O. virulentus*, or *O. bornovanus*) has rarely been used for host range tests, while dodder (*Cuscuta* spp.) has been demonstrated to transmit CPsV between citrus types.

CPsV is efficiently transmitted to susceptible citrus or citrus-related species by grafting, infecting *Microcitrus inodora* and *Fortunella hindsii* symptomlessly but with high ELISA values; other citrus relatives (*Citrus depressa*, Carrizo citrange, *Eremocitrus glauca*, *Microcitrus virgata*, and *Clausena excavata*) produced a severe shock reaction in the first flush of leaves, but did not flush again. Symptom intensity of some susceptible hosts (typically chlorotic flecking in young leaves) is correlated with virus accumulation as measured by ELISA, but mild symptoms were observed in Cleopatra mandarin, sour orange (*C. aurantium* L.), and citrumelo (*C. paradisi* × *Poncirus trifoliata*), which were negative by ELISA. Mechanical inoculation to test plants is also achieved.

MLBVV is reported to be more readily mechanically-transmissible than LBVaV, with which it is commonly found in mixed infection in lettuce; serial transfers from an initial mixed infection have resulted in recovery of only MLBVV after the third transfer to either lettuce or *N. benthamiana*. Continued mechanical transfers may result in loss of transmission by *O. brassicae*, and loss of

Table 3 Symptomatology on test plants commonly used for ophiovirus biological assays

Test plant	CPsV	MLBVV	LRNV	RWMV	TMMMV	FreSV
<i>C. quinoa</i>	CLL ^a	LL ^b	C/NLL	LL	LL	CLL
<i>C. amaranticolor</i>	C/NLL	–	C/NLL	LL	LL	CLL
<i>Gomphrena globosa</i>	NLL ^c , SS ^d	–	–	–	–	–
<i>Datura metel</i>	CLL, SS	–	–	–	–	–
<i>N. benthamiana</i>	–	SSI ^e	SSI	CLL, SS	LL	–
<i>N. rustica</i>	SS	–	–	–	–	–
<i>N. occidentalis</i>	–	LL, SS	C/NLL, SS	–	–	NLL, SS
<i>N. clevelandii</i>	–	SS	SSI	CLL, SS	–	–
<i>N. tabacum</i>	–	LL	–	LL	LL	–
<i>N. hesperis</i>	–	SS	C/NLL, SS	–	–	NLL, SS
<i>N. megalosiphon</i>	–	–	–	CLL, SS	–	–

^aChlorotic local lesions.^bLocal lesions.^cNecrotic local lesions.^dSystemic symptoms.^eSymptomless systemic infection.

– no information available.

ability to cause big vein symptoms in lettuce. Mechanical transmission of LRNV to test plants has been reported, with the most evident systemic symptoms in *N. occidentalis* and *N. hesperis*.

BlMaV is the most recently identified ophiovirus, and to date there are no reports of mechanical transmission, and no host range data are available, beyond blueberry infection. Transmission assays using blueberry seedlings and BlMaV-infected *Olpidium virulentus* were positive.

RWMV is able to infect several test plants; the inoculum source can maintain infectivity only if stored in liquid nitrogen. Local lesions are observed also in *C. capitatum*, *C. murale*, *C. album*, and local infection is also detected in *Spinacia oleracea* and *Vigna unguiculata*. Ranunculus seedlings (*Ranunculus sardous* and *Ranunculus* hybrid) challenged by mechanical inoculation with RWMV became infected, showing local and systemic necrosis.

FreSV can be transmitted to test plants; in some cases serial passage on *N. occidentalis* may yield evident chlorotic mosaic. Mechanical inoculation of purified FreSV to healthy freesia seedlings was never successful, however, freesia seedlings challenged by FreSV-infected *O. brassicae* resting spores became positive for FreSV infection.

Transmission of TMMMV to common test plants and to *Tetragonia expansa*, *Sonchus oleracea*, and *Beta vulgaris* resulted in only local infection.

Cytopathology

Thin sections of *Nicotiana clevelandii* leaves mechanically infected by RWMV have been examined by electron microscopy, but no distinctive inclusions were observed and no virus particles were seen. In classical thin sections, viral nucleic acids stain well but protein coats show low contrast, so it is not surprising that very thin nucleoprotein threads in random orientations should escape detection. After gold immune labeling with RWMV-specific polyclonal antibodies, sections of the cytoplasm of parenchyma cells were seen in the EM to be clearly labeled, but nuclei, chloroplasts, mitochondria, and microbodies were unlabeled. Confocal observation for subcellular localization of CPsV and MLBVV coat proteins linked to the fluorescent GFP molecule show cytoplasmic localization.

Ultrastructural changes have been reported in leaves of grapefruit infected with CPsV, but apparently free of Citrus tristeza virus (CTV) and Citrus exocortis viroid (CEVd), resulting in decreased thickness of both the phloem and the xylem, including reduced vessel diameter. Fewer oil glands, of smaller diameter than normal were observed, together with many abnormal chloroplasts with disorganized grana stacks and thylakoid membrane. Crystal idioblasts (cells containing calcium oxalate crystals) were found to be lacking in the palisade layer below the upper epidermis, in contrast to leaves of healthy plants, while inclusion bodies and tubular multi-membrane bodies were also observed. Although negative for CTV and CEVd, these plants may have been infected with other viruses, and some of the effects reported may result from reduced auxin levels rather than a direct presence of viral components.

Concluding Remarks

Ophioviruses are probably more widespread than currently known and they apparently have a very long evolutionary history. Psorosis of *Citrus* spp. and big-vein disease of lettuce, together with the necrotic disease of freesia and the blueberry syndrome, are relevant in agricultural production. Crop protection against ophiovirus infection is prevention centered; for this reason, there is a

need for resistance genes to be identified, or for introduction of resistance through biotechnological methods, together with application of good agricultural practices.

Further Reading

- Andika, I.B., Kondo, H., Sun, L., 2016. Interplays between soil-borne plant viruses and RNA silencing-mediated antiviral defense in roots. *Frontiers in Microbiology* 7, 1458. doi:10.3389/fmicb.2016.01458.
- García, M.L., Dal Bó, E., da Graça, J.V., *et al.*, 2017. ICTV virus taxonomy profile: *Ophioviridae*. *Journal of General Virology* 98, 1161–1162.
- García, M.L., Dal Bó, E., Grau, O., Milne, R.G., 1994. The closely related citrus ringspot and citrus psorosis viruses have particles of novel filamentous morphology. *Journal of General Virology* 75, 3585–3590.
- García, M.L., Goodin, M., Sasaya, T., Haenni, A.-L., 2011. Negative-strand RNA viruses: The plant-infecting counterparts. *Virus Research* 162, 184–202.
- Robles-Luna, G., Peña, E.J., Borniego, M.B., Heinlein, M., García, M.L., 2018. Citrus psorosis virus movement protein contains an aspartic protease required for autocleavage and the formation of tubule-like structures at plasmodesmata. *Journal of Virology* 92, e00355 doi:10.1128/JVI.00355-18.
- Robles-Luna, G., Reyesa, C.A., Peña, E.J., *et al.*, 2017. Identification and characterization of two RNA silencing suppressors encoded by ophioviruses. *Virus Research* 235, 96–105.
- Thekke-Veetil, T., Ho, T., Keller, K.E., Martin, R.R., Tzanetakis, I.E., 2014. A new ophiovirus is associated with blueberry mosaic disease. *Virus Research* 189, 92–96.
- Vaira, A.M., Accotto, G.P., Costantini, A., Milne, R.G., 2003. The partial sequence of RNA 1 of the ophiovirus *Ranunculus* white mottle virus indicates its relationship to rhabdoviruses and provides candidate primers for an ophio-specific RT-PCR test. *Archives of Virology* 148, 1037–1050.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/w/aspiviridae
Aspiviridae.

Orthotospoviruses (*Tospoviridae*)

Renato O Resende, University of Brasilia, Brasilia, Brazil

Hanu R Pappu, Washington State University, Pullman, WA, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of M. Tsompana, J.W. Moyer, Tospovirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00517-3](https://doi.org/10.1016/B978-012374410-4.00517-3).

Nomenclature

aa Amino acid(s)

AGO Argonaute 1

CAPS Cleaved amplified polymorphic sequence

CP Coat protein or capsid protein

diRNA defective interfering RNA

ELISA Enzyme-linked immunological assays

ER Endoplasmic reticulum

HC-Pro Helper component-protease

IGR Intergenic region

kb Kilobase

kDa Kilo Dalton

LAMP Loop mediated amplification

MP Movement protein

NBS-LRR Nucleotide-Binding site - Leucine-Rich Repeat

NGS Next-generation sequencing

nt Nucleotide(s)

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

RISC RNA-induced silencing complex

RNP Ribonucleoprotein

satRNA satellite RNA

UTR Un-translated region

VIGS Virus-induced gene silencing

VLPs Virus-like particles

VPg Viral protein genome-linked

VRC Virus replication complexes

vRNA virion RNA

Glossary

Ambisense genome Viral RNA genome with open reading frames in both the viral- and viral complementary (vc) sense on the same genome segment.

Envelope Membrane-like structure that packages genome segments.

Intergenic region It is the untranslated, A-U rich region found between the two open reading frames on the S and M RNA segments.

Negative sense genome Viral RNA genome that codes for

proteins in the vc sense. Transcription of vc mRNA is required for translation of viral proteins.

Nucleocapsid Viral RNA encapsidated by nucleoproteins.

Protoplast Plant cell lacking its cell wall.

Ribonucleoprotein complex Consisting of the viral RNA genome segment, nucleoprotein, and a small number of polymerase molecules.

Virion Quasispherical structure containing the viral genome and bounded by a membrane-like envelope.

History

Diseases now known to be caused by Tomato spotted wilt virus (TSWV) were first reported in 1915 and were shown to be of viral etiology by 1930. This taxon of plant viruses was categorized as a monotypic virus group consisting of a single virus (TSWV) until the report of impatiens necrotic spot virus (INSV) in 1991. Thus, most of the characteristics which define the genus *Orthotospovirus* were obtained through investigation of TSWV even after the discovery of additional viruses in the genus (Table 1). Biological investigations beginning in the 1940s revealed a virus that had an unusually large host range and occurred in nature as a complex mixture of phenotypic isolates. However, it was one of the least stable viruses and most difficult plant viruses to mechanically transmit. Although the enveloped virions were observed in the 1960s, molecular characterization and elucidation of the genome organization were not completed until the early 1990s. The virus was shown to be vectored by thrips in the 1930s and later transmitted in a persistent manner. Thrips were demonstrated to be a host for replication of the virus and that replication was required for transmission in the early 1990s. Later, it was recognized that limited, localized replication may occur in thrips that does not result in the thrips becoming viruliferous. Advances in gene function and cellular biology have been limited due to the absence of a robust *in vitro* plant or thrips cell culture system, and lack of an efficient reverse genetics system. However, limited progress has been made utilizing gene expression systems and classical viral genetics.

Table 1 Viruses members and tentative members of the genus *Orthotospovirus* (family *Tospoviridae*)

Virus species names	Acronym	Reported thrips vectors	GenBank accession #
<i>Alstroemeria necrotic streak orthotospovirus</i>	ANSV	<i>Frankliniella occidentalis</i>	L: MG696851; M: MG696852; S: MG696853
<i>Alstroemeria yellow streak orthotospovirus</i>	AYSV	<i>Thrips tabaci</i>	L: MF469033; M: MF469034; S: MF469035
<i>Bean necrotic mosaic orthotospovirus</i>	BeNMV	Unknown	L: JF417980; M: JN587269; S: JN587268
<i>Calla lily chlorotic spot orthotospovirus</i>	CCSV	<i>Thrips palmi</i>	L: FJ822962; M: FJ822961; S: AY867502
<i>Capsicum chlorosis orthotospovirus</i>	CaCV	<i>Ceratothripoides claratris</i> <i>Thrips palmi</i> <i>Frankliniella schultzei</i>	L: DQ256124; M: DQ256125; S: DQ256123
<i>Chrysanthemum stem necrosis orthotospovirus</i>	CSNV	<i>Frankliniella occidentalis</i> <i>Frankliniella intonosa</i> <i>Frankliniella schultzei</i>	L: KM114546; M: KM114547; S: KM114548
<i>Groundnut bud necrosis orthotospovirus</i>	GBNV	<i>Frankliniella schultzei</i> <i>Scirtothrips dorsalis</i> <i>Thrips palmi</i>	L: AF025538; M: U42555; S: U27809
<i>Groundnut chlorotic fan-spot orthotospovirus</i>	GCFSV	<i>Scirtothrips dorsalis</i>	L: KP146140; M: KP146141; S: AF080526
<i>Groundnut ringspot orthotospovirus</i>	GRSV	<i>Frankliniella occidentalis</i> <i>Frankliniella schultzei</i> <i>Frankliniella gemina</i> <i>Scirtothrips dorsalis</i>	L: KT972590; M: KT972592; S: KT972594
<i>Groundnut yellow spot orthotospovirus</i>	GYSV	Unknown	S: HQ402596
<i>Hippeastrum chlorotic ringspot orthotospovirus</i>	HCRV	Unknown	L: HG763861; M: JX833565; S: JX833564
<i>Impatiens necrotic spot orthotospovirus</i>	INSV	<i>Frankliniella occidentalis</i> <i>Frankliniella fusca</i> <i>Frankliniella intonosa</i>	L: X93218; M: M74904; S: X66972
<i>Iris yellow spot orthotospovirus</i>	IYSV	<i>Frankliniella fusca</i> <i>Thrips tabaci</i>	L: FJ623474; M: AF214014; S: AF001387
<i>Melon severe mosaic orthotospovirus</i>	MeSMV	Unknown	L: KX698424; M: KX698423; S: KX698422
<i>Melon yellow spot orthotospovirus</i>	MYSV	<i>Thrips palmi</i>	L: AB061774; M: AB061773; S: AB038343
<i>Mulberry vein banding-associated orthotospovirus</i>	MVBaV	Unknown	L: KM819698; M: KM819699; S: KM819701
<i>Pepper chlorotic spot orthotospovirus</i>	PCSV	Unknown	L: KX247379; M: KX247378; S: KX247377
<i>Polygonum ringspot orthotospovirus</i>	PolRSV	<i>Dictyothrips betae</i>	L: KJ541746; M: KJ541745; S: KJ541744
<i>Soybean vein necrosis orthotospovirus</i>	SVNV	<i>Neohydatothrips variabilis</i> <i>Frankliniella fusca</i> <i>Frankliniella tritici</i>	L: HQ728385; M: HQ728386; S: HQ728387
<i>Tomato chlorotic spot orthotospovirus</i>	TCSV	<i>Frankliniella occidentalis</i> <i>Frankliniella schultzei</i> <i>Frankliniella intonosa</i>	L: MH742959; M: MH742960; S: MH742961
<i>Tomato spotted wilt orthotospovirus</i>	TSWV	<i>Frankliniella occidentalis</i> <i>Frankliniella fusca</i> <i>Frankliniella schultzei</i> <i>Frankliniella intonosa</i> <i>Frankliniella bispinosa</i> <i>Frankliniella cephalica</i> <i>Frankliniella gemina</i> <i>Thrips setosus</i> <i>Thrips tabaci</i>	L: D10066; M: S48091; S: D00645
<i>Tomato yellow ring orthotospovirus</i>	TYRV	<i>Thrips tabaci</i>	L: JN560178; M: JN560177; S: AY686718
<i>Tomato zonate spot orthotospovirus</i>	TZSV	<i>Frankliniella occidentalis</i>	L: EF552435; M: EF552434; S: EF552433
<i>Watermelon bud necrosis orthotospovirus</i>	WBNV	<i>Thrips palmi</i>	L: GU735408; M: GU584185; S: GU584184
<i>Watermelon silver mottle orthotospovirus</i>	WSMoV	<i>Thrips palmi</i>	L: AF133128; M: U75379; S: U78734
<i>Zucchini lethal chlorosis orthotospovirus</i>	ZLCV	<i>Frankliniella zucchini</i>	L: KU641378; M: KU641379; S: KU641380
<i>Tentative species in the genus</i>			
<i>Lisianthus necrotic ringspot virus^a</i>	LNRV	Unknown	S: AB852525
<i>Pepper necrotic spot virus^a</i>	PNSV	Unknown	S: HE584762
<i>Tomato necrotic spot virus^a</i>	TNSV	Unknown	

^aTentative orthotospoviruses that have been described and proposed to the International Committee on Taxonomy of Viruses (ICTV) and waiting for the final approval as definitive species.

Note: Updated table from Rotenberg, D., Jacobson, A.L., Schneeweis, D.J., Whitfield, A.E., 2015. Thrips transmission of tospoviruses. *Current Opinion in Virology* 15, 80–89. doi:10.1016/j.coviro.2015.08.003.

Taxonomy and Classification

In addition to the genus *Tenuivirus* (family *Phenuiviridae*), the genus *Orthotospovirus* (family *Tospoviridae*) encompasses the plant-infecting viruses of the order *Bunyavirales*. All these plant viruses share many molecular characteristics with animal-infecting viruses classified in this order. The orthotospoviruses, for example, have an enveloped virion containing the viral genome which is distributed among three RNA segments that replicate in a manner consistent with that of other vertebrate-infecting bunyaviruses such as hantaviruses, phleboviruses, orthobunyaviruses, etc. All three RNA segments have highly conserved, complementary termini resulting in a pan-handle structure and genes with similar functions. Differently, the small (S) and medium (M) segments of orthotospoviruses encode two genes in opposite or ambisense polarity.

Classification of an orthotospovirus population as a distinct viral species is based upon the similarity of sequence between the nucleocapsid genes of the respective viruses (Fig. 1). This is in contrast to the system used to differentiate animal viruses in other genera which traditionally relied on serological neutralization of infectivity or other biological properties (hemagglutination)

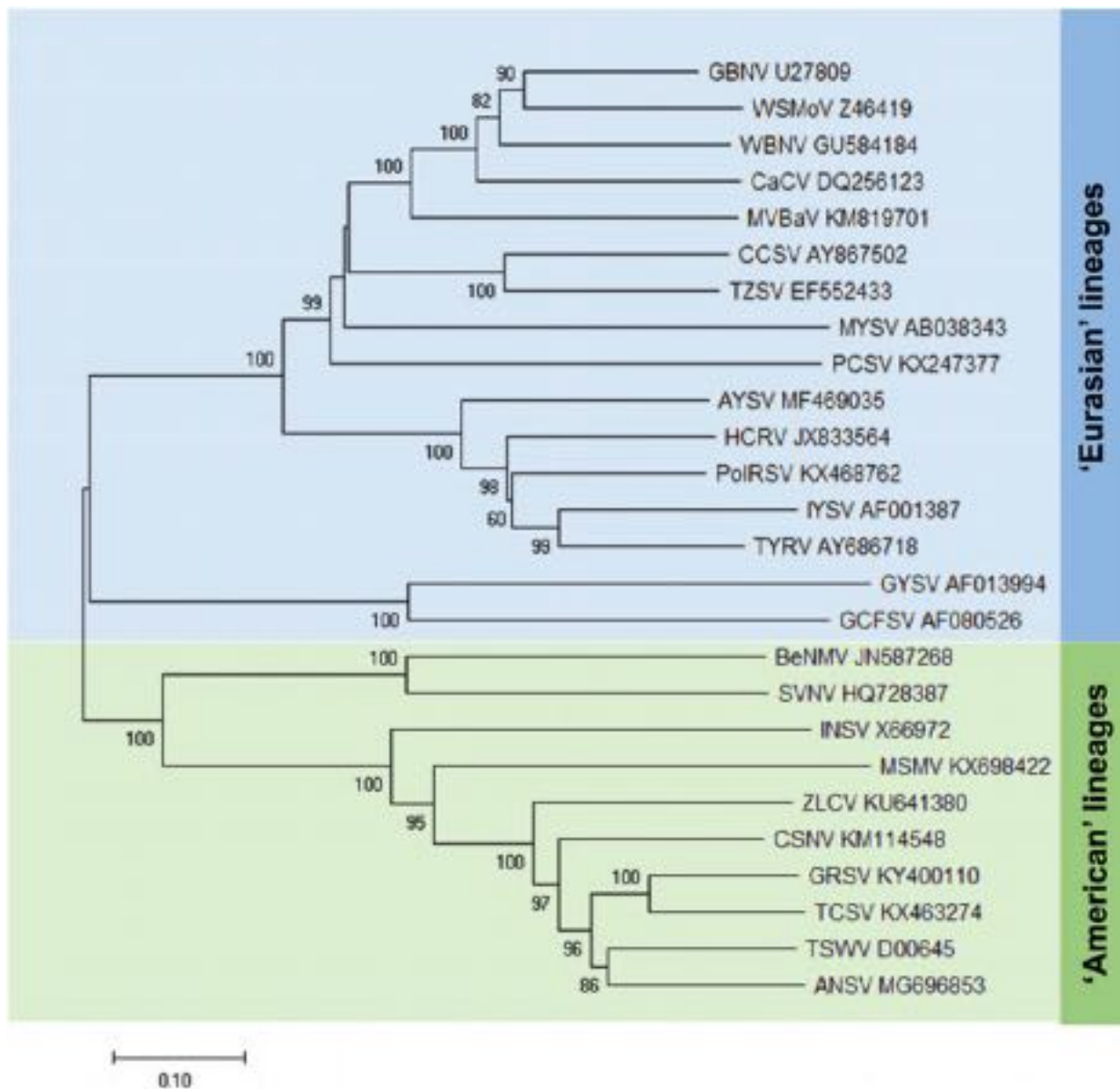


Fig. 1 Phylogeny of the complete S-RNA genome segment sequences of orthotospoviruses belonging to recognized and proposed species within the genus *Orthotospovirus* in the family *Tospoviridae*. The midpoint-rooted tree was deduced in MEGA X v. 10.0.5 after alignment in Muscle, using the neighbor-joining method based on the maximum composite likelihood substitution model with uniform rates and 1000 bootstrap replications. The scale bar indicates the number of substitutions per site. Bootstrap support for branches is shown at the junctions of branches where it was >50%. Accession numbers are given in Table 1.

mediated by the glycoproteins. Orthospovirus isolates with greater than 90% nt similarity in the nucleocapsid gene are classified as isolates of the same viral species.

Geographic Distribution

TSWV, the type species member of the orthospoviruses is found worldwide in association with its thrips vector. The wide host-range of TSWV and its thrips vectors is consistent with the geographic distribution. Other orthospoviruses have more well-defined distribution (**Fig. 1**). For example, Groundnut bud necrosis virus (GBNV), Watermelon bud necrosis virus (WBNV), and Watermelon silver mottle virus (WSMoV), that are transmitted by *Thrips palmi*, a thrips species found only in the subtropics and only known to occur in South-East Asia. Another anomaly is INSV. While INSV is reported to occur around the world, it is almost entirely limited to greenhouse-grown floral crops.

Host-Range and Virus Propagation

TSWV has one of the most diverse host-ranges of any plant-infecting virus. This virus infects over 900 plant species belonging to 90 botanical families, both monocots and dicots. In addition, TSWV is reported to be transmitted by at least nine thrips species. Important economic plants susceptible to TSWV include tomato, potato, tobacco, peanut, pepper, lettuce, papaya, and chrysanthemum. Other orthospoviruses (e.g., IYSV) have much narrower host-ranges and thus the broad host-range of TSWV is not characteristic of the genus. These viruses can be transmitted mechanically or by their thrips vectors. Recently, it has been shown that Soybean vein necrosis virus (SVNV) can be seed-transmitted at a rate of 6% in soybean plants, but this phenomenon has not been observed for any other orthospovirus. RNA preparations are not infectious. There are no robust cell culture systems for orthospoviruses. However, plant protoplasts and insect cells have been successfully inoculated.

Virion Properties

Orthospovirus virions are quasispherical, enveloped particles with 80 – 120 nm in diameter (**Fig. 2**). Two viral coded glycoproteins, GN and GC, are embedded in the viral envelope and form 5–10 nm surface projections. Ribonucleoprotein (RNP) particles consisting of the viral RNA encapsidated by nucleoproteins (ribonucleocapsid), and a small number of polymerase molecules are contained within the envelope. Ribonucleocapsids are pseudocircular due to noncovalent bonding of the complementary RNA termini. Intact

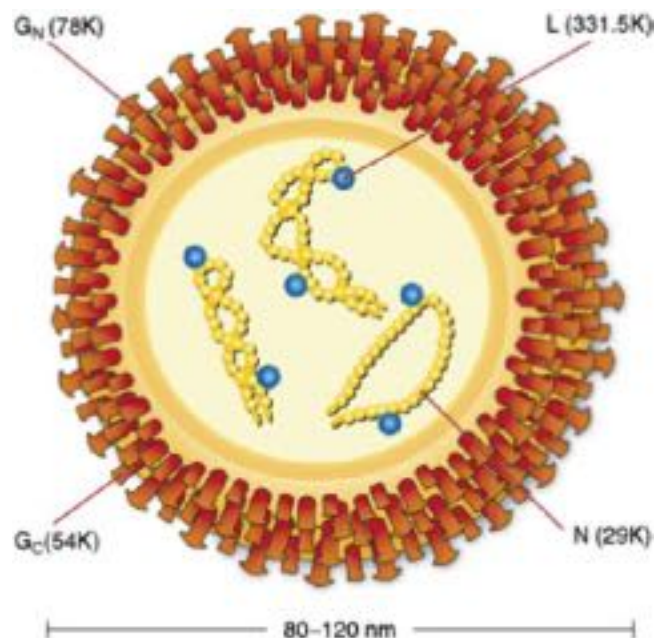


Fig. 2 *Orthospovirus* quasispherical virion particles. Left. Electron micrograph of negatively stained particles of Tomato spotted wilt virus (TSWV). The bar represents 100 nm. Courtesy of J. van Lent. Right. The S, M, and L RNA genomic segments are encapsidated by nucleoproteins, which are in association with L protein (polymerase) molecules, and form pan-handle structures due to the complementarity of their 5'- and 3'-ends. The glycoproteins GN and GC are embedded within the viral envelope.

virions as well as carefully prepared RNPs retrieved from sucrose or CsSO₄ gradients are infectious. There are several reports that TSWV and INSV isolates, while infectious, are defective for virion formation.

Genome Properties

Orthospoviruses have a single-stranded, tripartite RNA genome with segments designated as L, M, and S in order of decreasing size (Fig. 3). The termini of each of the RNA segments consist of an eight nt sequence (5'-AGAGCAAU-3') that is strictly conserved among all orthospoviruses. The remaining untranslated region at the termini also has a high degree of complementarity. Base pairing at the termini between the inverted complementary sequences supports a pan-handle structure that most likely serves as a promoter for replication. For TSWV, the L RNA has 8.9 nt and codes for the L or RNA-dependent RNA polymerase (RdRp) protein in the viral complementary (vc) sense (Fig. 3). The M and S RNAs are in ambisense orientation. The M RNA has 4.8 nt and codes in the viral sense for the non-structural protein NSm and for the GN/GC precursor glycoprotein in the vc sense. The S RNA presents 2.9 nt and codes in the viral sense for the nonstructural protein NSs and the nucleocapsid (N) protein in the vc sense (Fig. 3).

Intergenic regions (IGRs) of TSWV M and S RNA have variable lengths, are A–U rich, and are the most hypervariable regions of the genome. The 5'- and 3'-ends of the IGRs are conserved, separated by variable sequences, deletions, and insertions. In addition, highly conserved sequences are embedded within the S RNA IGR. A 33 nt duplication occurring in the S RNA IGR of some isolates has been correlated with loss of competitiveness in mixed infections of isolates with and without the duplication. A 31 nt conserved sequence, with significantly higher GC-content compared to the remaining S RNA IGR, has also been found in some TSWV isolates. The IGRs of the M and S segments have high inclination for base pairing thought to be involved in initiation and termination of transcription.

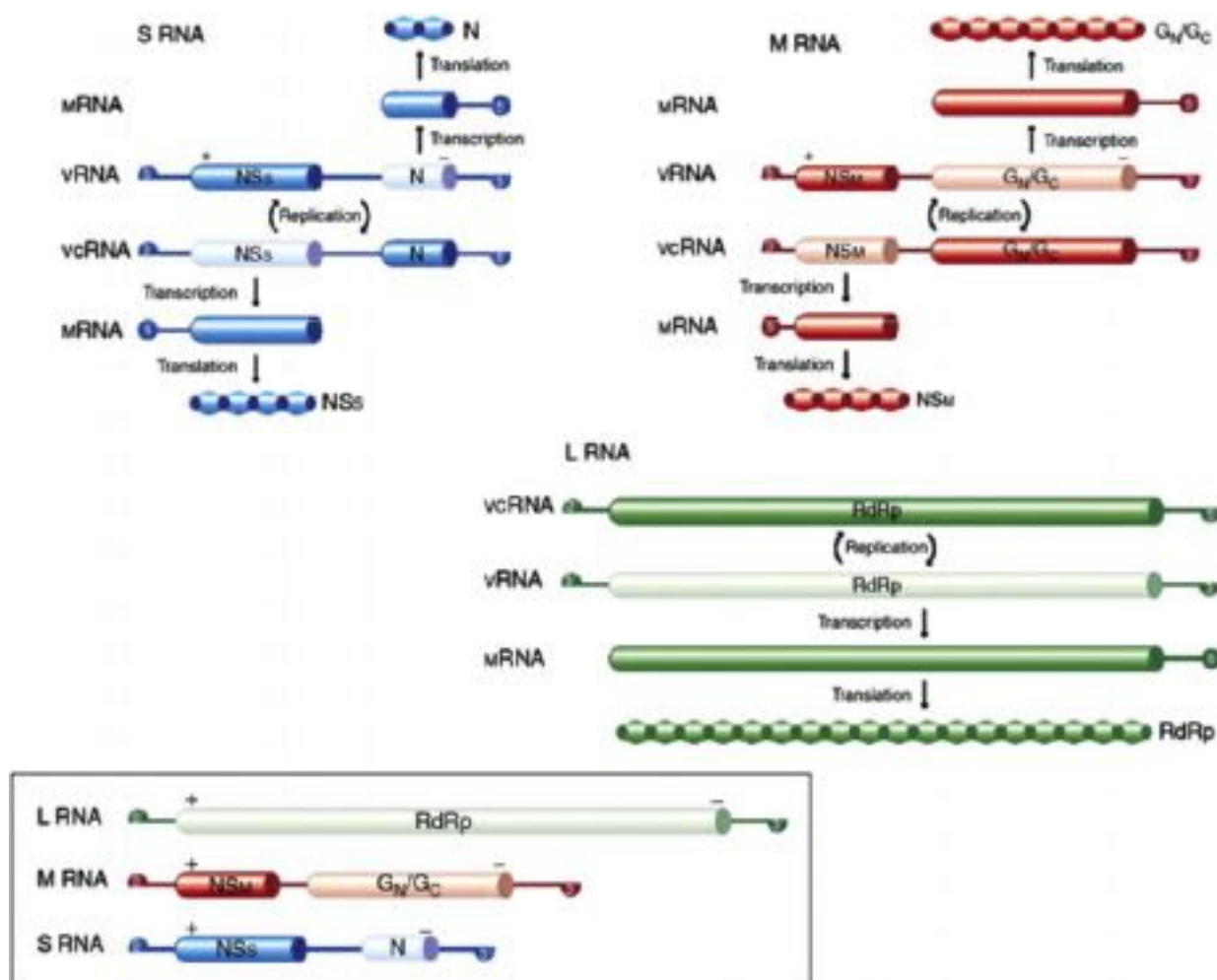


Fig. 3 *Orthospovirus ambisense* genome organization (inset in figure) and expression strategy. Positive (+) and negative (–) sense ORFs are dark and light shaded tubes, respectively. Proteins from the S and M segments are translated from subgenomic mRNAs which are capped with 10–20 nt of host-derived RNA at the 5'-end.

There is speculation that the termination of transcription is dictated by a conserved nt sequence (CAACUUUGG) in the center of the S and M RNA IGR or that it is due to a secondary structure highly stabilized in the 31 nt region referred to above.

Full length molecules of the M and S RNAs are found in infected tissue and purified virions in both the viral and vc sense (approximate ratio of 10:1), consistent with ambisense segments from other viruses. Defective interfering RNAs (DIs) associated with attenuated symptom expression and increased replication rate are also frequently observed. DIs from the L ORF in TSWV infected tissue are the result of and associated with attenuated infectivity. Deletions, frameshift, and nonsense mutations in the GN/GC ORF have been shown to interfere with thrips transmissibility and virion assembly. Frameshift and nonsense mutations with unknown effect have also been identified in the N ORF. The formation of DIs is favored by repeated mechanical passage in certain plant hosts, high inoculum concentration, and low temperatures. Available evidence supports the hypothesis that secondary structure rather than the RNA sequence itself is the primary determinant of the site of deletion. There is also a high frequency of DIs that maintain the original reading frame resulting in translation of truncated proteins whose existence was confirmed in nucleocapsid preparations.

Protein Properties

The 330–336 kDa L protein encoded by the L RNA has been identified as a RdRp through sequence homology with other members of the order *Bunyavirales* and identification of sequence motifs characteristic of polymerases. RdRp activity has been associated with detergent-disrupted TSWV virion preparations.

The 34 kDa NSm protein encoded by the M RNA has been shown to induce tubule structures in plant protoplasts and insect (*Spodoptera* and *Trichoplusia*) cells. Induction of tubules in plants, ability to change the size exclusion limit of plasmodesmata, an early expression profile and complementation of cell-to-cell and systemic movement in a movement-defective Tobacco mosaic virus vector is evidence that NSm is the TSWV movement protein and that it supports long-distance movement of viral RNAs. In thrips, NSm does not aggregate into tubules, indicating that this protein might not have cell-to-cell movement function in the vector's life cycle. It is also known that NSm specifically interacts with the N protein, the At-4/1 intra- and intercellular trafficking plant protein and binds single-stranded RNA in a sequence-nonspecific manner. An NSm homolog is absent in the vertebrate-infecting viruses of the order *Bunyavirales*. The 127 kDa GN/GC precursor glycoprotein also coded by the M RNA contains a signal sequence that allows its translation on the endoplasmic reticulum. Proteolytic cleavage of the polyprotein does not require other viral proteins. The molecular mass of GN and GC is 78 kDa and 54 kDa, respectively, for TSWV. Evidence for the involvement of the glycoproteins in thrips transmission is provided by: (1) their interaction with proteins of the thrips vector, (2) their association with the insect midgut during acquisition, (3) the loss of thrips transmissibility of envelope-deficient mutants, (4) the presence of a glycoprotein sequence motif that is characteristic for cellular attachment domains, and (5) the observation that only reassortants with the M RNA of a thrips-transmissible isolate rescue thrips transmissibility. Specifically, GN is involved in virus binding and/or entry in thrips midgut cells, whereas GC is a possible fusion protein playing a significant role in pH-dependent virus entry. It is also believed that these proteins are implicated in virion assembly.

The NSs protein encoded on the S RNA is 52 kDa and accumulates to high levels as loose aggregates or paracrystalline arrays of filaments. NSs has RNA silencing suppressor activity, affects symptom expression in TSWV-infected plants, and is not present in the mature virus particle. The N protein, also encoded by the S RNA, ranges in size from 28.1 to 30.9 kDa depending on the virus. This protein encapsidates the viral RNA segments, is highly abundant, and is the predominant protein detected in serological assays. A 'head-to-tail' interaction of the nucleoprotein N terminus (aa 1–39) with the C terminus (aa 233–248) results in multimerization.

Replication

Replication of viral RNA and assembly of virions occurs in the cytoplasm of both plant and insect cells. Orthotospovirus replication, however, has mainly been described based on plant infection. Upon entry into the plant cell, the virus loses its membrane and releases infectious nucleocapsids into the cytoplasm. In thrips cells, infection by orthotospoviruses is accommodated by binding of the viral surface glycoproteins to a host cell receptor(s) (possibly a 50 kDa and/or a 94 kDa protein). This is followed by release of infectious nucleocapsids into the cytoplasm, through fusion between the viral and thrips membranes possibly initiated by low pH. Depending on the concentration of N protein, the viral RNA is either transcribed or replicated. At low N concentrations, the polymerase transcribes mRNAs that are translated into the virus proteins. Translation of proteins from the S and M ambisense RNAs occurs from subgenomic mRNAs (Fig. 3). The S and M subgenomic mRNAs are capped at the 5' terminus with 10–20 nt of non-viral origin indicating that orthotospoviruses utilize a cap-snatching mechanism to regulate transcription. Leader sequences of Alfalfa mosaic virus have also been detected as caps of TSWV mRNA in mixed infections of the two viruses. The TSWV transcriptase has a reported preference for caps with multiple base complementarity with the viral template. Upon increase of N protein concentration, the polymerase switches its mode to replication with the viral RNA serving as the template. Replicated viral RNAs form RNPs that can presumably associate with the NSm protein for movement through plasmodesmata to adjacent plant cells through tubular structures. Alternatively, RNPs form new virions by associating with the glycoproteins and budding through the Golgi membranes. Virions are initially double-membraned, but soon coalesce and form groups of virions with a single membrane surrounded by another membrane.

Pathogenicity and Cytopathology

Orthotospoviruses are noted for the severity of the diseases they cause in plants. Symptoms are highly variable, depending on the virus, the virus isolate, the host plant, time of the year, and environment, and are thus of little diagnostic value. Chlorosis, necrosis, ring or line patterns, mottling, silvering, and stunting often appear on inoculated and systemically infected leaves. Systemic invasion of plants is frequently nonuniform. Stems and petioles may exhibit necrotic lesions. Observed symptoms often mimic disease and injury caused by other biotic and abiotic stresses. Infection of younger plants results in severe stunting and high mortality rates. TSWV has been shown to affect more severely *Datura*, *Nicotiana*, and *Physalis* plants under a specific temperature regime (daytime, $29 \pm 2^\circ\text{C}$, nighttime, $24 \pm 3^\circ\text{C}$). The effect of orthotospovirus infection on thrips has been controversial, due to the confounded effects of plant host, virus, and environment on the insect vector and the genetic variability of thrips and virus populations. TSWV infection of *F. occidentalis* provided evidence that thrips exhibit an immune response to the virus. Recent work with TSWV-infected *F. fusca* reared on infected foliage indicated a direct effect of the virus on thrips resulting in reduced fitness. The same study showed that the plant infection status and the TSWV isolate have also an effect on the insect, explaining the variable results obtained from independent studies of virus pathogenicity on the insect vector.

Orthotospoviruses induce characteristic cytopathic structures that are host and virus-isolate dependent. In addition to virions, inclusions of viroplasms consisting of the NSs or N protein may be abundant in the cytoplasm. NSs may aggregate in loose bundles (e.g., TSWV) or in highly ordered paracrystalline arrays (e.g., INSV). Excess N protein occurs in granular electron dense masses. NSs and nucleocapsid protein inclusions have been observed in infected plant and insect cells. NSm protein induces tubule structures in plant protoplasts and insect cells.

Transmission and Epidemiology

Orthotospoviruses are transmitted from plant to plant by at least sixteen thrips species in the genera *Ceratothripoides*, *Dictyothrips*, *Frankliniella*, *Neohydatothrips*, *Scirtothrips*, and *Thrips*. Among the more common vectors are *Frankliniella occidentalis*, *F. fusca*, *F. schultzei*, *F. intonsa*, *F. bispinosa*, *Thrips palmi*, *T. setosus*, and *T. tabaci*. Thrips feed on the cytoplasm of plant cells. The contents of infected cells are ingested and the virus is transported along the lumen of the digestive tract to the midgut, the primary binding and entry site into the insect cells. The brush border of the midgut lumen is the first membrane barrier that the virus encounters. The virus replicates in the midgut and crosses the basement membrane into the visceral muscle cells. The virus subsequently enters the primary salivary glands. It has been hypothesized that the virus moves from the midgut to the salivary glands through infection of ligament-like structures, or when there is direct contact between membranes of the visceral muscles and the primary salivary glands during the larval stages of development. A less plausible hypothesis is that the virus infects the salivary glands after entry and circulation in the hemocoel. Viral inoculum is introduced into plants in the insect saliva coincident with feeding on the plant by adult thrips.

The process of successful acquisition occurs only by larvae and acquisition rates decrease as larvae develop, affecting adult vector competency. Vector competency is also determined by the thrips' feeding preference on a particular host, the uniformity of distribution of virus in plant cells, the rate of virus replication in the midgut, and the extent of virus migration from the midgut to the visceral muscle cells and the salivary glands. In some instances the virus can be acquired by adult thrips and infects midgut cells, but is unable to spread further possibly due to the formation of an age-dependent midgut barrier (e.g., basal lamina). Research has shown the existence of thrips transmitters with detectable levels of virus, nontransmitters with detectable virus, and nontransmitters with no detectable virus, supporting multiple sites for vector specificity between orthotospoviruses and thrips. Evidence for replication of the virus in the insect vector is based on the accumulation of NSs and the visualization of viral inclusions in midgut epithelial cells, muscle cells, and the salivary glands. Although the virus is maintained transtadially throughout the life of the insect, there is no evidence for transovarial transmission. Thus, each generation of thrips must acquire the virus during the larval stages.

The primary dispersal of orthotospoviruses is by adult thrips and dissemination of infected somatic tissue in vegetatively propagated crops. These viruses are thought to move long distances in thrips carried by wind currents. They may also survive in commercial agricultural systems in weeds that serve as a bridge between crops. Infected summer weeds (e.g., in NC *I. purpurea*, *I. hederacea*, *M. verticillata*, *A. palmeri*, *C. obtusifolia*, *R. scabra*, *Ambrosia artemisiifolia* L., *Polygonum pennsylvanicum* L., and *Chenopodium album* L.) are the principal source for spread of TSWV to winter annual weeds, from which the virus is spread to susceptible crops in spring. Secondary spread within a crop can only occur in crops that concomitantly support virus infection and reproduction of the vector as only the larval stage can acquire the virus for transmission. Transmission through plant seed has only been demonstrated for SVN but at a low rate. The emergence of these viruses as serious pathogens in crops has been attributed to the increased prevalence of *F. occidentalis* as an agricultural pest on a worldwide basis.

Genetics and Evolution

The knowledge base for genetics and evolution of orthotospoviruses has been derived almost exclusively from TSWV. TSWV has a characteristic ability to adapt to new or resistant hosts and to lose phenotypic characters following repeated passages in experimental

hosts, especially *Nicotiana benthamiana*. The virus occurs in plants as a heterogeneous mutant population with one or two predominant haplotypes and 9–21 rare haplotypes. Recent research shows that natural TSWV variants evolve in nature through recombination, random genetic drift, and mutation. Inter-genomic recombination is important for the genesis and evolution of ancestral TSWV lineages. Genetic drift during thrips transmission and mutation concurrent with virus population growth, shape the genetic architecture of the most recently evolved lineages. The existence of single viral strains as mutant populations and recombination in ancestral viral lineages arm TSWV with a unique genetic reservoir for causing disease and spreading in epidemic proportions in nature. Additional research at the species level supports a distinct TSWV geographical structure and the occurrence of species-wide population expansions. TSWV is also known to use reassortment of genome segments to adapt to resistant hosts under specific laboratory conditions. The determinants of adaptation to resistance in tomato and pepper have been mapped to the M and S RNAs, respectively. Tomato chlorotic spot virus (TCSV) and Groundnut ringspot virus (GRSV) share the same M segment due to host adaptation reasons. Phylogenetic analyses suggest that TCSV putatively incorporated this genomic segment from GRSV upon a reassortment event since the parental genotype of TCSV remains unknown. Little is known about the thrips–orthospovirus coevolution and the genetic diversity of the thrips vector itself. The altered status of *Thrips tabaci* as a TSWV vector is one of the very few likely examples of coevolution between orthospoviruses and their insect vector.

Detection and Diagnosis

Orthospoviruses have certain unique biological properties that are useful for diagnosis. These viruses can be mechanically transmitted by gently rubbing inoculum on plants dusted with carborundum. *Nicotiana glutinosa* L., *Chenopodium quinoa* Wild., and garden petunia give characteristic lesions that progress as spots or concentric zones, and sometimes as lethal necrosis. Orthospoviruses can also be identified by electron microscopy of leaf-dip preparations on thin sections of infected plants. Additional techniques for identification are based on the enzyme-linked immunosorbent assay (ELISA) using polyclonal and monoclonal antibodies, and detection of viral-specific nucleic acids using ribo- and cDNA-probes. The reverse transcription-polymerase chain reaction (RT-PCR) is the most powerful and commonly used technique for detecting small amounts of orthospovirus RNA. Real time RT-PCR has been successfully used to detect and quantify TSWV in leaf soak and total RNA extracts from infected plants and thrips. RT-PCR with degenerate primers can detect five distinct orthospovirus species. Tissue selection and sampling strategy are critical factors in TSWV diagnosis and detection regardless of the technique. Because, TSWV titer varies throughout the plant and does not spread uniformly throughout plants that are ‘systemic’ hosts, sampling strategies should be validated in each situation. Nowadays, with the availability of high-throughput sequencing techniques, also referred to as next-generation sequencing (NGS), viruses have been identified and had their genomes completely sequenced following a fast and unbiased manner. Thus, reports of new viruses include the genetic characterization of entire genomes, which facilitates evolutionary and biological studies.

Prevention and Control

Orthospoviruses cause significant economic losses annually, due to suppressed growth, yield, and reduced quality. These viruses can be partially managed in well-defined cropping systems such as glasshouses by obtaining uninfected plant propagules, implementing a preventative thrips control program in high risk areas, together with constant monitoring of production areas for thrips and infected plants. However, these strategies are costly and require intensive management. Control in field crops is problematic due to the array of external sources of inoculum. Vector control is generally ineffective against the introduction of virus from external sources, due to thrips’ high fecundity, ability to develop insecticide resistance, and to infest many TSWV-susceptible crops. Some measure of control can be achieved using thrips-proof mesh tunnels in the field and reflective mulches. Cultural practices such as utilization of virus-tested planting stock, careful selection of planting dates, removal of cull piles and weeds, rotation with nonsusceptible crops, prevention of planting TSWV-susceptible crops adjacent to each other, reduced in-field cultivation to avoid movement of thrips from infected sources, can reduce the spread of orthospoviruses. In peanuts, higher plant density, planting from early until late May and application of selected insecticides have reduced the incidence of TSWV. In flue-cured tobacco, early-season treatment with activators of plant defenses and insecticides have also significantly reduced TSWV incidence.

Deployment of resistant cultivars has provided benefits in few crops infected by orthospoviruses. Single-gene resistance is available for orthospovirus in a limited number of tomato (*Sw-5b*) and pepper (*Tsw*) cultivars. Naturally occurring, resistance-breaking isolates of TSWV have been recovered from pepper and tomato cultivars containing their respective resistance genes. Nevertheless, a correct management of *Sw-5b*-resistant cultivars has shown that this resistance can be durable. The main strategy is to decrease the selection pressure on virus populations by alternating resistant and susceptible cultivars or by planting susceptible tomatoes alongside production fields. The *Sw-5b* protein is a Nucleotide-Binding site and Leucine-Rich Repeat (NB-LRR) receptor with nucleocytoplasmic distribution that confers a dominant, high-spectrum resistance against isolates of Alstroemeria necrotic streak virus (ANSV), Chrysanthemum stem necrosis virus (CSNV), GRSV, INSV, TCSV, and TSWV. The *Sw-5b* protein activates resistance by direct recognition of orthospovirus NSm proteins. Co-dominant markers have been developed for screening *Sw-5b*-containing tomato cultivars. Co-dominant cleaved amplified polymorphic sequence (CAPS) marker has also been developed for TSWV marker-assisted selection in pepper. ‘Field’ resistance has been reported for some peanut varieties. Progress has been made in

understanding the genetic basis of the ability of TSWV to overcome single gene resistance by mapping determinants to specific segments of the TSWV genome and characterizing the selection process. Pathogen-derived resistance utilizing the N and NSm genes has been effective in some greenhouse and field tests; however, isolates have been obtained that overcome nucleocapsid mediated resistance. Best suppression of TSWV epidemics has been achieved with the integrated use of resistant cultivars, chemical, and cultural practices.

Future Perspectives

The last few years have been characterized by exciting developments in our understanding of orthotospovirus molecular biology, diversity, evolution, and virus–host relationships. Within this context, several new viruses and thrips vectors have been reported. Advances have also been made in understanding orthotospovirus resistance mechanisms in plants, especially concerning *Sw-5* resistance genes, which have brought some insights in the management of orthotospovirus in the field. Unfortunately, efficient reverse genetic systems of orthotospoviruses still remain to be developed as well as stable insect cell cultures for *in vitro* orthotospovirus propagation. The availability of editing DNA tools as CRISPR Cas9 may allow relevant changes in how to control orthotospovirus infection from dismissing key elements in plant hosts necessary for virus infection.

Further Reading

- Abudurexiti, A., *et al.*, 2019. Taxonomy of the order *Bunyavirales*: Update 2019. *Archives of Virology* 164. doi:10.1007/s00705-019-04253-6.
- de Oliveira, A.S., Boiteux, L.S., Kormelink, R., Resende, R.O., 2019. The *Sw-5* gene cluster: Tomato breeding and research toward orthotospovirus disease control. *Frontiers in Plant Science* 19. doi:10.3389/fpls.2018.01055.
- Oliver, J.E., Whitfield, A.E., 2016. The genus *Tospovirus*: Emerging bunyaviruses that threaten food security. *Annual Review of Virology*, 3, 101–124. doi:10.1146/annurev-virology-100114-055036.
- Pappu, H.R., Jones, R.A.C., Jain, R.K., 2009. Global status of tospovirus epidemics in diverse cropping systems: Successes achieved and challenges ahead. *Virus Research* 141 (2), 219–236. doi:10.1016/j.virusres.2009.01.009.
- Rotenberg, D., Jacobson, A.L., Schneeweis, D.J., Whitfield, A.E., 2015. Thrips transmission of tospoviruses. *Current Opinion in Virology* 15, 80–89. doi:10.1016/j.coviro.2015.08.003.
- Turina, M., Kormelink, R., Resende, R.O., 2016. Resistance to tospoviruses in vegetable crops: Epidemiological and molecular aspects. *Annual Review of Phytopathology* 54, 347–371. doi:10.1146/annurev-phyto-080615-095843.
- Zhu, M., van Grinsven, I.L., Kormelink, R., Tao, X., 2019. Paving the way to tospovirus infection: Multilined Interplays with plant innate immunity. *Annual Review of Phytopathology* 57. doi:10.1146/annurev-phyto-082718-100309.

Ourmiaviruses (*Botourmiaviridae*)

Gian Paolo Accotto, Institute for Sustainable Plant Protection, National Research Council of Italy, Torino, Italy
Cristina Rosa, Pennsylvania State University, University Park, PA, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
CP Coat protein or capsid protein
HVT Horizontal virus transfer.
kb Kilobase

kDa Kilo dalton
MP Movement protein
nt Nucleotide(s)
ORF Open Reading Frame
RdRp RNA-dependent RNA polymerase

Glossary

Monocistronic RNA that encodes for only one polypeptide.

Reassortment Exchange of genomic segments between viruses.

Introduction

The only three virus species of this genus were described a long time ago. Since then, only a handful of articles have been published, mainly concerning Ourmia melon virus (OuMV), the type species. The other two members of the genus are Cassava virus C (CsVC), reported only once in a mixed infection in cassava from Malawi, and Epirus cherry virus (EpCV), found in a single cherry tree. This tree has since died. OuMV has been detected on several occasions in melons in north-western Iran. In general, OuMV has been found in mixed infections with other cucurbit viruses, so the extent of economic losses is unknown. Ourmiaviruses are potentially damaging, as they have been experimentally transmitted to a rather wide range of dicot species, including important crops in which they produce symptoms of mosaic, mottle, and necrosis.

Taxonomy, Phylogeny and Evolution

The genus is not assigned to a specific family. The members have the same structural features but their coat proteins (CPs) share less than 70% aa identity. Phylogenetic analysis of the ourmiavirus RdRp places them in a clade originated from ourmia-like viruses of invertebrates, infecting members of the subphylum Chelicerata and Crustacea, and of the superphylum Lophotrochozoa. This clade was shown to be sister to the one containing ourmia-like mycoviruses. Such a phylogeny suggests that ourmia-like viruses infecting fungi, crustacea, and plants could belong to distinct genera in the same virus family. These genera are distinct from the *Narnavirus* genus. Since most of the ourmia-like viruses have been found in invertebrates, it has been hypothesized that the plant ourmiaviruses evolved by horizontal virus transfer (HVT) from herbivorous invertebrates. In the new host the acquisition of a movement protein (MP) from a plant virus was necessary. Actually, analyses of the ourmiavirus MP indicates similarities with the MPs of viruses in the family *Tombusviridae*. The CP of ourmiaviruses shows distant affinities with the CPs of sobemo-, tombusluteo- and nodaviruses. As of 2017, three species in the genus *Ourmiavirus* were recognized by ICTV: *Cassava virus C* (CsVC), *Epirus Cherry Virus* (EpCV), and *Ourmia melon virus* (OuMV).

Virion Structure

The virions are non-enveloped and made of subunits of a single CP (Fig. 1). They have cylindrical bodies of discrete lengths, 18 nm in diameter and capped by conical (probably hemi-icosahedral) ends. The bodies of the particles are composed by a series of disks, commonly two or three for particles with a length of 30–37 nm, more rarely four or six disks for particles with a length of 45.5–62 nm. Particles with five disks and with a hypothetical length of 54 nm have never been observed. The buoyant density in CsCl of the particles of all sizes is 1.375 g cm⁻³.

Genome Organization

Viruses in this genus have a tripartite single stranded positive sense RNA genome. Each genomic segment is monocistronic. The lengths of the genomic segments are approximately 2800 nt, 1100 nt and 970 nt (Fig. 2).

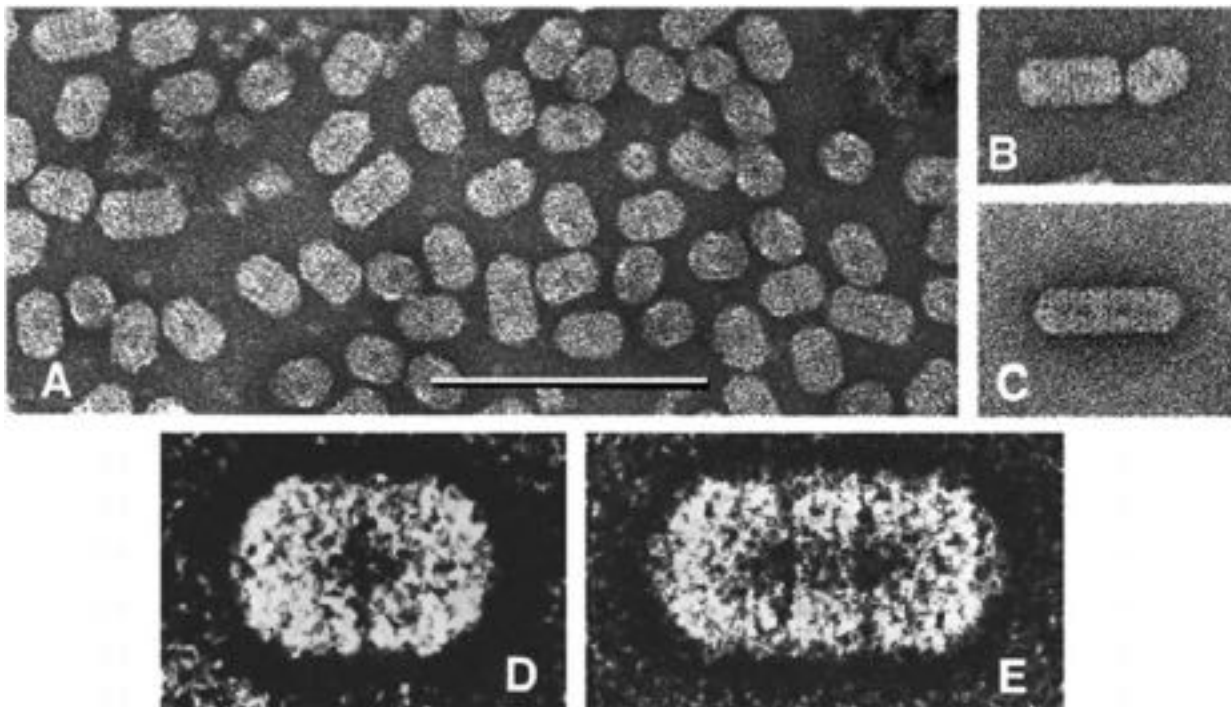


Fig. 1 *Ourmiavirus* virion morphology: (A, B, C) Negative contrast electron micrographs (uranyl acetate) of purified particles of Ourmia melon virus. The bar represents 100 nm. (D, E) Features of the two commonest particle types, enhanced by photographic superimposition. (from ICTV 9th report, page 1178, [Fig. 2](#), with permission from Elsevier).

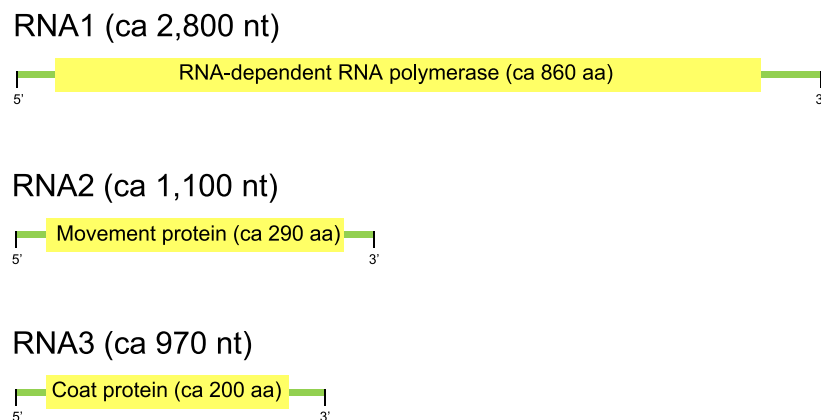


Fig. 2 Diagram summarizing genome organization of members of the genus *Ourmiavirus*. Green bars: genomic RNAs. Yellow bars: ORFs.

Properties and Functions of Gene Products

Each genomic segment contains a single ORF. RNA1 encodes for a putative RdRp of ca. 860 aa and 97 kDa, RNA2 for a putative MP of ca. 290 aa and 32 kDa, and RNA3 for a putative CP of ca. 200 aa and 23 kDa ([Fig. 2](#)). The functions of the RdRp, MP and CP have been experimentally demonstrated for OuMV, the better characterized virus in the genus. OuMV RdRp is necessary for virus replication and contains the polymerase motif GDD.

OuMV MP belongs to the "30K" Superfamily of movement proteins. It is a symptoms determinant, is localized at plasmodesmata of infected plant cells and forms tubular structures inside the cytoplasm and between adjacent cells. The OuMV CP is required for assembly of the virions and systemic movement. It is localized in the nucleolus when expressed as fusion with GFP and without the expression of RNA1 and RNA2. The KR-rich region at the N-terminus of the CP has been shown to induce tissue tropism, to be a host determinant and to interfere with the plant silencing pathway. Virion assembly requires an active replication and RNA3, but not RNA2.

Replication and Propagation

OuMV replication occurs in the cytoplasm of infected plant cells. OuMV MP is required for cell-to-cell movement. While effective systemic infection requires the expression of the OuMV CP, RNA1 and 2 can move up to some vascular tissue exit sites without CP expression.

Transmission, Host Range

Ourmiaviruses are plant infecting viruses. OuMV was isolated from melon, CsVc from cassava and EpCV from cherry. Extensive surveys have not been conducted to reveal other natural hosts of these viruses in the geographic areas where the 3 species were originally isolated, aside from one survey in Iran on cucurbits. Experimentally, OuMV and other ourmiaviruses have been mechanically transmitted to dicots in ca. 15 plant families, including Solanaceae, Brassicaceae and Cucurbitaceae. Transmission studies found that these viruses could not be transmitted by several aphid species, the whiteflies *Trialeurodes vaporariorum* and *Bemisia tabaci*, and the mite *Tetranychus urticae*. Transmission experiments via soil or water were also unsuccessful. Seed-transmission rates in *Nicotiana benthamiana* and *N. megalosiphon* are 1%–2%.

Epidemiology and Control

Not much is known about the epidemiology of these viruses. A survey conducted on melon and squash in 16 regions in Guilan Province (Iran) in 2005 and 2006 revealed that OuMV was present in 4 regions. Based on this survey, OuMV was the most prevalent virus detected in melon (*Cucumis melo* L.) and was frequently found in squash (*Cucurbita* sp.). Based on the limited information available, standard agronomic practices aimed at reducing the economic damage caused by ourmiaviruses could include screening for virus presence in seeds and other plant propagative material, exclusion and eradication in areas where the viruses are not present, and rouging of symptomatic plants.

Virus–Host Relationships

Plants systemically infected with ourmiaviruses show an array of symptoms, such as vein chlorosis, rasp-leaf, necrotic spots and rings, stunting, yellowing and necrosis.

Diagnosis

Virions and tubular structures of OuMV have been used to produce species specific polyclonal antibodies for ELISA and western blotting, but commercial diagnostic kits are not available. PCR primers used in published experiments can be adopted for diagnosis via molecular techniques.

Concluding Remarks

Due to the little effort dedicated so far to this genus, in spite of its very special characteristics, many questions remain unsolved. Evidence from metagenomics suggests that the origin of this genus could have been genomic reassortment of RNA viruses infecting hosts outside the plant kingdom, accompanied by HVT. More studies are necessary to elucidate the missing information about this interesting genus.

Further Reading

- Accotto, G.P., Boccardo, G., Riccioni, L., Barba, M., 1997. Comparison of some molecular properties of Ourmia melon and Epirus cherry viruses, two representatives of a proposed new virus group. *Journal of Plant Pathology* 78, 87–91.
- Accotto, G.P., Milne, R.G., 2008. Ourmiavirus. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*. London, UK: Elsevier, pp. 500–501.
- Aiton, M.M., Lennon, A.M., Roberts, I.M., Harrison, B.D., 1988. Two new cassava viruses from Africa. In: *Proceedings of the 5th International Congress of Plant Pathology*. p. 43. Kyoto, Japan.
- Avgelis, A., Barba, M., Rumbos, I., 1989. Epirus cherry virus, an unusual virus isolated from cherry with rasp-leaf symptoms in Greece. *Journal of Phytopathology* 126, 51–58.
- Crivelli, G., Ciuffo, M., Genre, A., Masenga, V., Turina, M., 2011. Reverse genetic analysis of Ourmiaviruses reveals the nucleolar localization of the coat protein in *Nicotiana benthamiana* and unusual requirements for virion formation. *Journal of Virology* 85, 5091–5104.
- Dolja, V.V., Koonin, E.V., 2018. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Research* 244, 36–52.
- Gholamalizadeh, R., Vahdat, A., Hossein-Nia, S.V., Elahinia, A., Bananej, K., 2008. Occurrence of Ourmia melon virus in the Guilan province of Northern Iran. *Plant Disease* 92 (7), 1135.

- Lisa, V., Milne, R.G., Accotto, G.P., *et al.*, 1988. Ourmia melon virus, A virus from Iran with novel properties. *Annals of Applied Biology* 112, 291–302.
- Margaria, P., Anderson, C.T., Turina, M., Rosa, C., 2016. Identification of Ourmiavirus 30K movement protein amino acid residues involved in symptomatology, viral movement, subcellular localization and tubule formation. *Molecular Plant Pathology* 17, 1063–1079.
- Rastgou, M., Habibi, M.K., Izadpanah, K., *et al.*, 2009. Molecular characterization of the plant virus genus *Ourmiavirus* and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *Journal of General Virology* 90, 2525–2535.
- Rossi, M., Genre, A., Turina, M., 2014. Genetic dissection of a putative nucleolar localization signal in the coat protein of Ourmia melon virus. *Archives of Virology* 159, 1187–1192.
- Rossi, M., Vallino, M., Abba, S., *et al.*, 2015. The importance of the KR-rich region of the coat protein of Ourmia melon virus for host specificity, tissue tropism, and interference with antiviral defense. *Molecular Plant-Microbe Interactions* 28, 30–41.
- Shi, M., Lin, X.D., Tian, J.H., *et al.*, 2016. Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543.
- Turina, M., Hillman, B.I., Izadpanah, K., *et al.*, 2017. ICTV virus taxonomy profile: *Ourmiavirus*. *Journal of General Virology* 98, 129–130.

Papaya Ringspot Virus (*Potyviridae*)

Cécile Desbiez and Hervé Lecoq, Plant Pathology Unit, INRAE – French National Research Institute for Agriculture, Food and Environment, Montfavet, France

© 2021 Elsevier Ltd. All rights reserved.

This is an update of D. Gonsalves, J.Y. Suzuki, S. Tripathi, S.A. Ferreira, Papaya Ringspot Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00731-7](https://doi.org/10.1016/B978-012374410-4.00731-7).

Nomenclature

aa Amino acid

CP Capsid protein

DAS-ELISA Double antibody sandwich enzyme-linked immunosorbent assay

HC-Pro Helper component-protease protein of potyviruses

kDa kilodalton (1 Da = 1 g/mol)

LAMP Loop-mediated isothermal amplification

nt Nucleotide

Glossary

Gene silencing Plant defense mechanism based on the detection and fragmentation of double-stranded RNAs, including the replicative forms of RNA viruses.

Mild-strain cross-protection A plant systemically infected by a mild virus strain will not develop additional symptoms when inoculated by a severe strain of the same virus.

Molecular clock The molecular clock hypothesis considers that mutations between homologous nucleotide or amino-

acid sequences accumulate in a clock-like fashion, i.e., at a constant rate with time. Molecular changes along the branches of a phylogenetic tree can thus be used to estimate the time of divergence from a common ancestor.

Pathogen-derived resistance Engineered resistance to pathogens through the synthesis in transgenic plants of pathogen proteins or nucleic acids.

History and Taxonomy

The term Papaya ringspot virus (PRSV) was first used in the 1940s to describe a viral disease of papaya, initially observed in Hawaii. The name was used primarily to describe the ringspots that appeared on fruits from infected plants. Early investigations showed that the virus was transmitted by several species of aphids in a non-persistent manner. That is, the aphid vector could acquire the virus in a short period of time while feeding on infected plants and likewise transmit the virus in a span of few seconds to less than a minute during subsequent feeding. In the same decade researchers from India and other places like Puerto Rico reported the occurrence of an aphid-transmitted disease of papaya; based on the symptoms on the leaves, it was identified as Papaya mosaic virus. Work in the 1980s showed that the aphid-transmitted papaya mosaic virus and PRSV were the same virus, and the name of PRSV was adopted. PRSV is a member of the family *Potyviridae*, a large and arguably the most economically important group of plant viruses. Today, the term *Papaya mosaic virus* (PapMV) is reserved for a virus that is not aphid transmitted, belonging to the genus *Potexvirus* in the family *Alphaflexiviridae*, and causes the papaya mosaic disease which is seldom observed and not important commercially, even if mixed infections between PRSV and PapMV can increase symptoms and virus accumulation. PRSV should not be confused with another potyvirus, *Papaya leaf distortion mosaic virus* (PLDMV), which occurs in Okinawa and other parts of Asia, such as Taiwan. This virus causes very similar symptoms as PRSV on papaya and cucurbits but is serologically unrelated and its sequence shares only 55%–59% similarity to that of PRSV.

In the 1930s a mosaic disease was described in cucurbits in the USA and named Watermelon mosaic virus (WMV). Later on, biological and serological studies showed that the disease might be caused by two distinct viral entities called WMV-1 and WMV-2. More recently, serological and molecular characterization showed that PRSV and WMV-1 are virtually identical. Based on their close relationship, a single name was adopted to unify both viruses into one species. The name PRSV was chosen due to its being named before WMV-1. To clarify host range, 'P' (PRSV-P) or 'P type' is used to designate virus infecting papaya and cucurbits, while 'W' (PRSV-W) or 'W type' refers to virus infecting cucurbits only.

Classification and Relationship With Other Potyviruses: The “PRSV Cluster”

PRSV was classified as a member of the genus *Potyvirus* in the family *Potyviridae* based on particle morphology, aphid transmissibility, serological relationships, ability to induce pinwheel cytoplasmic inclusions in host cells, genome organization and nucleotide (nt) sequences. Several other potyviruses infect papaya or cucurbits. The main other potyvirus infecting papaya is PLDMV, that shares no particular relationship with PRSV. More than 10 potyviruses infect cucurbits. Some of them are not closely

related to PRSV based on serological and molecular data: Zucchini yellow mosaic virus (ZYMV), Watermelon mosaic virus (WMV), Melon vein banding mosaic virus (MVBV), Telfairia mosaic virus (TeMV), Cucurbita vein-banding virus (CVBV), Turnip mosaic virus (TuMV), Clover yellow vein virus (CIYVV) and Bean yellow mosaic virus (BYMV). However, in the 1990s, several viral isolates, mostly infecting cucurbits, were found to share biological and serological relationships with PRSV, forming what was called a “PRSV cluster” of related members that could not always be unambiguously distinguished. When sequences became available, the different criteria of the International Committee for the Taxonomy of Viruses (ICTV) could be applied, mainly: (1) less than 76% nt identity (80% aa identity) in the coat protein (CP)-coding region or in the whole genome; (2) difference in polyprotein cleavage site; (3) host range and key host reactions; (4) different inclusion body morphology; (5) serological relatedness. Partial or complete sequences have revealed the presence of at least seven species besides PRSV in the “PRSV cluster”: Moroccan watermelon mosaic virus (MWMV), Algerian watermelon mosaic virus (AWMV), Zucchini yellow fleck virus (ZYFV), Zucchini tigre mosaic virus (ZTMV), Zucchini shoestring virus (ZSTV), Sudan watermelon mosaic virus (SuWMV) and Wild melon vein banding virus (WMVBV). Partial sequences suggest that more species are present in Asia, although they remain to be fully characterized (Fig. 1). Most members of the PRSV cluster were found in a limited geographic area and/or presented a low prevalence in the countries where they were present. AWMV, MWMV, SuWMV, WMVBV, ZYFV and ZSTV were found only in Africa and/or the Mediterranean Basin, indicating that this region is most probably their center of origin. ZTMV was found in Europe, Asia, America and Africa but at a rather low prevalence, although it has probably been frequently mis-diagnosed as PRSV because of the frequent cross-reactions observed in serological tests.

Host Range, Symptomatology, and Geographic Distribution

The systemic host range of PRSV is confined to plants in the families *Caricaceae* and *Cucurbitaceae*, with the primary economically important host being papaya and a range of cucurbits such as squash, cucumber, watermelon, and melons. PRSV-P isolates cause generally severe symptoms in papaya and mild symptoms in cucurbits, while PRSV-W do not infect papaya and cause severe symptoms in cucurbits. In addition, some PRSV isolates cause local lesions on plants of the family *Chenopodiaceae* such as *Chenopodium quinoa* and *C. amaranticolor*. In papaya, PRSV infection is characterized by mosaic and chlorosis symptoms on leaves, water-soaked streaks on the petiole, and deformation of leaves that can result in shoestring-like symptoms that resemble mite damage (Fig. 2). The virus can cause deformation and ringspot symptoms on the fruit, hence the name PRSV. PRSV-P is present in all papaya-growing areas. PRSV-W infects cucurbits worldwide, but is more prevalent in tropical, subtropical and Mediterranean conditions than in temperate ones. This may be related to the difficulty for the virus to overwinter in areas where cucurbits are not present all year round. On cucurbits, PRSV-W causes severe mosaic and chlorosis on leaves, which are also often deformed with shoestring-type symptoms. The fruits are often deformed and bumpy. Some PRSV-W isolates induce systemic necrotic spots and top necrosis in some melon cultivars.

Transmission and Epidemiology

PRSV is transmitted in a non-persistent manner by more than 24 aphid species among which *Aphis gossypii*, *Myzus persicae* and *A. craccivora* are the most efficient. PRSV is acquired and transmitted during very short probes, what makes its spread within a field generally rapid. However PRSV persists only few hours on the stylets: therefore long-distance spread by aphids is limited. Virus sources are generally neighboring infected crops, but for PRSV-W several wild cucurbits (*Melothria pendula*, *Momordica* sp.) were shown efficient virus sources in tropical and sub-tropical regions. There is no report of PRSV-P or -W seed transmission. Long-distance spread may occur through the movement of infected material (seedlings, fruits). Indeed, aphids were shown to be able to acquire and transmit efficiently PRSV-W from commercial infected melon fruits.

General Properties and Genome of PRSV

Like other potyviruses, PRSV has flexuous filamentous particles about 760–800 nm × 12 nm encapsidating a single-stranded RNA of about 10,326 nt in length. The genome is translated as a single polyprotein that is further cleaved autocatalytically in 10 functional proteins (P1, HC-Pro, P3, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb and CP), with a smaller 11th protein P3N-PIPO expressed through translational frameshift. Virus particles consist of 94.5% protein and 5.5% nucleic acid by weight. PRSV has a single coat protein (CP) of about 36 kDa. The density of the virion in purified preparations is 1.32 g cm⁻³ in cesium chloride.

A rather interesting feature of the PRSV is that sequence analysis predicts two potential cleavage sites at the N-terminus of the CP. One of the sites (VFHQ/SKNF) predicts a CP of 33 kDa and a polymerase (nuclear inclusion b protein, NIb) of 537 aa about 20 aa larger than those of other potyviruses. The second predicted cleavage site (VYHE/SRGTD) generates a CP of 36 kDa and a NIb of 517 aa. There is no firm evidence to suggest that only one cleavage site is used. If both sites are used in polyprotein processing, one would expect heterogeneous products. This may explain why the analysis of purified CP preparations that are stored frequently shows the major ~36 kDa form and CPs that are 2–5 kDa smaller.

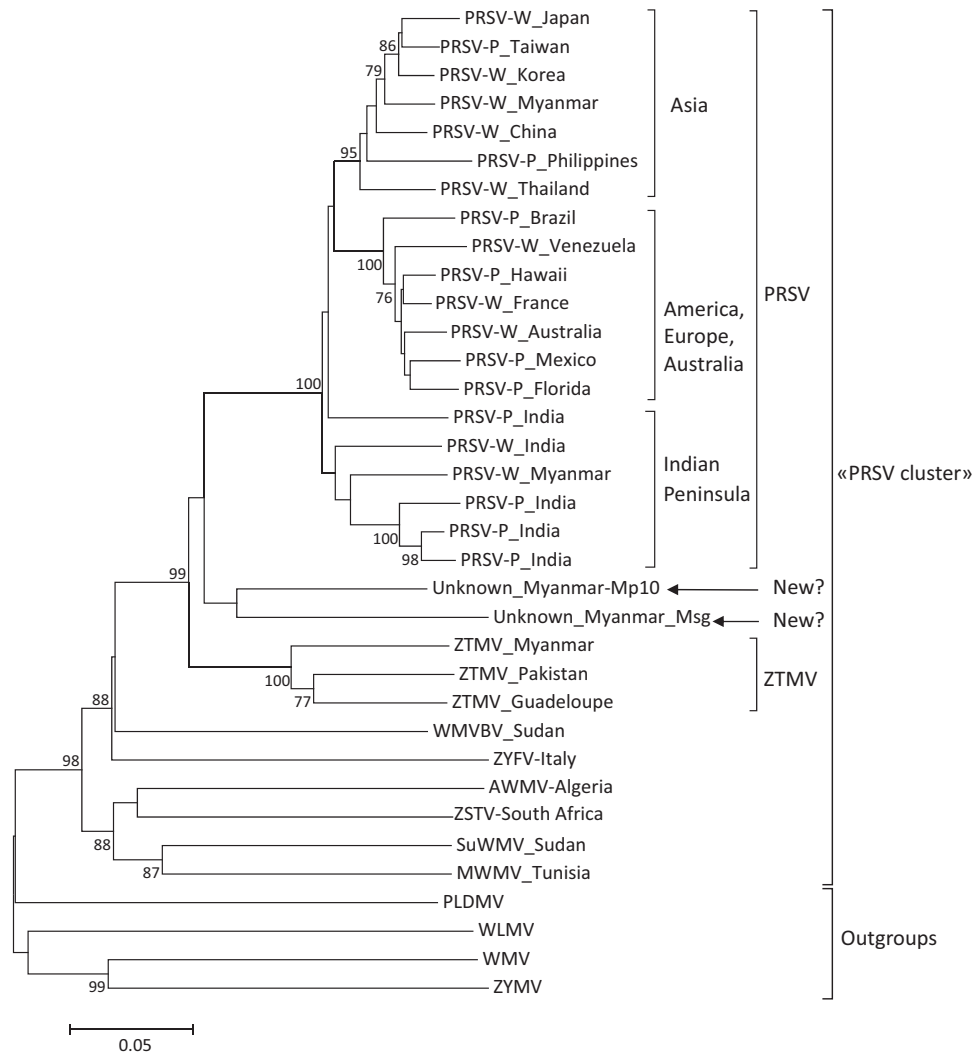


Fig. 1 Taxonomy, based on coat protein nucleotide sequences, of potyviruses in the “PRSV cluster” including PRSV, six acknowledged species and two putative new species. PRSV-P isolates of PRSV infect papaya whereas PRSV-W isolates infect cucurbits. Branch lengths indicate the molecular divergence of viruses (the scale bar represents 0.05 mutations per residue). Figures at some nodes represent bootstrap values (in %), indicating the robustness of each node. Only values above 75% are indicated.

Sequence Diversity and Evolution; Molecular Clock Analyzes

Knowledge of the sequence diversity among isolates of a virus has great implications in developing an effective virus disease management program and in understanding the origin and biology of the virus. Numerous PRSV-P and -W sequences from virus isolated from different parts of the world have been reported. Nucleotide and aa sequences among PRSV isolates differ by as much as 21% and 13% respectively. Heterogeneity in CP length ranging from 837 to 870 nt has been noted, resulting in CPs of between 279 and 290 aa. The first 50 or so aa of the N-terminal region of the PRSV CP gene are highly variable and all differences in CP length were confined to this region. The N-terminal part of PRSV CP is characterized by the presence of several –up to 8-Glutamic acid (E) and Lysine (K) repeats, and the majority of the size differences were due to differences in the number of EK repeats.

Phylogenetic analyzes based on HC-Pro and CP nt sequences indicated the presence of two main clades of PRSV isolates, one from Asia and one from Africa, Australia and America. Isolates from India presented a basal location in the phylogenetic tree, suggesting that the Indian Peninsula could correspond to the center of origin of the virus. Two major radiations appear to have arisen from different branches of the Indian population. The earliest included all the isolates in east and south east Asia (China, Indonesia, Malaysia, Philippines, Taiwan, Thailand and Vietnam). The other includes all the isolates from the Americas and Europe with small separate sub-clusters of isolates from Hawaii, Taiwan, and Australia. A “molecular clock” approach, relying on the hypothesis that mutations accumulate in viral populations at an almost constant yearly rate, was applied to PRSV. Based on the molecular divergence between the Indian basal group and the “American” lineage, PRSV introduction to America presumably happened about 300 years



Fig. 2 Symptoms of PRSV-P on papaya (left) and of PRSV-W on zucchini (right, top) and on melon (right, bottom).

ago, maybe through Europe. Isolates from the “Asian” cluster have been observed recently in Europe and the Caribbean, showing that population exchanges are currently taking place, probably through the worldwide trade of plant materials.

There was no relationship between the biotype (W or P) and the viral phylogeny, suggesting that jumps between W and P happened several times during the evolution of the virus. Evidences from various sources suggest that PRSV is primarily a pathogen of cucurbits, and that PRSV-P originated from PRSV-W. This suggestion is supported by the diversity in cucurbit-infecting potyviruses that are phylogenetically related to PRSV. In Australia, the outbreak of PRSV-P about 25 years ago most probably arose from the population of PRSV-W already present in the country.

Molecular Determinants for Biological Properties

The easy jump between PRSV-W and PRSV-P pathotypes suggested a simple genetic determinism related to a few mutations. Studies utilizing infectious transcripts from recombinant viruses followed by bioassays of the transcripts have demonstrated that the region of PRSV genome encoding the NIa-Protease was critical for papaya infection. In particular, two aa at positions 2309 (K → D) and 2487 (I → V) of PRSV are significantly different between papaya-infecting type P and non-papaya-infecting type W. Further point mutational studies in these sites indicated that the K → D mutation was responsible for conferring the ability to infect papaya.

As noted earlier, PRSV causes local lesions on *C. amaranticolor* and *C. quinoa*. The severe strain PRSV HA strain from Hawaii causes local lesions on *C. quinoa* but a mild nitrous acid mutant of it, PRSV HA 5-1, does not. Recombinant infectious viruses were generated by exchanging genome parts between PRSV HA and PRSV HA 5-1. The study revealed that the pathogenicity-related region is present between nt positions 950 and 3261 of the PRSV HA genome and mutations in the P1 and HC-Pro resulted in the attenuation of PRSV HA symptoms and the loss of ability to produce local lesions on *C. quinoa*. The HC-Pro coding region of PRSV is thus the major determinant factor for local lesion formation.

Contrary to other potyviruses where the N-terminal part of the HC-Pro is dispensable for infectivity, this region was shown to be required in PRSV-W for systemic infection of zucchini. Virus–host interaction studies based on recombinant analyzes between severe and mild strains of PRSV indicated that the HC-Pro gene plays an important role in viral pathogenicity and virulence and acts as a suppressor of the gene-silencing defense mechanism in the papaya host plant. In addition, the comparative reaction of recombinant PRSV with chimeric CP gene sequences showed that heterologous sequences and their position in the CP gene influences their pathogenicity on PRSV-resistant transgenic papaya. As with other potyviruses the HC-Pro/virions interaction is essential for aphid transmission of PRSV.

Diagnostics

As for many other potyviruses, serological methods (DAS-ELISA) are widely used for PRSV diagnostics. Polyclonal antibodies against purified virus particles have been obtained, and commercial DAS-ELISA kits are available. However serological cross-reactivity between viruses in the “PRSV cluster” can be observed, potentially resulting in misdiagnosis. Since many PRSV CP or complete sequences are now available, specific primers have been designed for the molecular diagnostic of PRSV using reverse transcription and polymerase chain reaction (RT-PCR), alone or in multiplex with other papaya- or cucurbit-infecting viruses. More recently, isothermal amplification techniques (LAMP) have been developed for PRSV.

Virus Control

Cultural Practices: Prophylactic Measures

Prophylactic measures are not specific to one virus. They generally tend to prevent or limit the contact of viruliferous aphids with susceptible crops. Insecticide treatments are usually inefficient for non-persistently transmitted viruses. The control or management of PRSV on papaya has been approached through practices such as quarantine, eradication, avoidance by planting crops in areas isolated from the virus, and continual rogueing of infected plants. In Taiwan a successful but costly method used was the cultivation of papaya under insect proof nets.

For cucurbits, the same precautions can be used, particularly through careful weeding near plantings and avoiding overlapping crops in the same areas to reduce virus and aphid sources. Besides, plastic mulches are frequently used, as they limit weeds but they also have a repelling effect on aphids and contribute to delay virus spread, at least at the beginning of the crop before the plants grow and cover most of the mulch surface. Different covers (plastics, unwoven) can also be used to prevent winged aphids from reaching low-growing cucurbit plants, but they must be removed to allow pollination.

Cross-Protection

Cross-protection can be defined as the use of a mild strain of virus to infect plants that are subsequently protected against economic damage caused by a severe strain of the same virus. This practice has been used successfully for many years to minimize damage by Citrus tristeza virus in Brazil, for example. In the early 1980s, a mild strain of PRSV (described above as PRSV HA 5-1) was developed through nitrous acid treatment of a severe strain, PRSV HA isolated from Oahu Island in Hawaii. This mild strain was tested on Oahu Island and showed good protection against damage by severe strains but produced symptoms that were very obvious on certain cultivars, especially in the winter months. This prominent symptom induction on certain cultivars and the logistics of mild strain buildup and inoculation of plants, among others, were factors that caused it not to be consistently used on the island of Oahu. Interestingly, the mild strain was used extensively for several years in Taiwan, but it did not afford sufficient protection against the severe strains from Taiwan and thus its use was abandoned after several years.

Breeding for Resistance

The use of resistant cultivars, when they are available, is the easiest and cheapest way for farmers to control plant viruses. However resistance must be present in the germplasm of the species of interest, and introducing the resistance in commercial cultivars is a time-consuming process. It is important, before starting the breeding process, to check if the resistance is not broken by preexisting virus strains, or if the virus can easily evolve to overcome it.

Research to develop PRSV-resistant papaya varieties started in the 1970s. Since resistance to PRSV has not been identified in *Carica papaya*, researchers have used tolerant germplasm in an attempt to develop papaya cultivars with acceptable PRSV tolerance and horticultural characteristics. However, tolerance to PRSV is apparently governed by a family of genes that is inherited quantitatively, which makes it technically difficult to develop cultivars of acceptable horticultural quality. Furthermore, the tolerant lines do become infected with PRSV, although fruit production continues still at a lower level. Indeed, in Thailand, the Philippines, and Taiwan, a number of tolerant lines have been developed and are used. However, tolerance could not be introduced in all types of papayas and infected tolerant plant may act as virus sources to contaminate susceptible crops.

Resistance to PRSV has been described in the germplasm of different cucurbits, including melon, cucumber, squash and watermelon. The resistance was characterized in most cases as monogenic, either dominant or recessive. Interestingly in melon, a dominant resistant gene was found efficient for PRSV-W as well as for several other virus species of the “PRSV cluster”. Resistance is now present in several commercial cultivars of melon, cucumber and squash.

Pathogen-Derived Resistance for Controlling PRSV: The Hawaii Case

The control or management of PRSV in Hawaii has been approached through cultural practices, cross-protection with mild virus strains, tolerance, and resistance using the “pathogen-derived resistance” technology. Ultimately, the most successful has been the “pathogen-derived resistance” approach.

PRSV was first detected in the 1940s on Oahu Island where Hawaii's papaya industry was located at that time. Efforts to control the virus on Oahu largely consisted of state officials and farmers continually monitoring for infected plants and rogueing them, especially in areas where the virus was not prevalent. However, by the late 1950s, PRSV was causing extensive damage, which caused the papaya industry to relocate to Puna on the island of Hawaii.

The relocation of the industry to Puna was timely and effective because Puna had an abundance of land that was suitable to grow Kapoho, a cultivar of excellent quality that adapted to the volcanic soil base there, allowing excellent drainage, had high rainfall and yet lots of sunshine, and the land there could be bought or leased at reasonable prices. By the 1970s, the Kapoho papaya grown in Puna accounted for 95% of the state's papaya production, making papaya the second most important fruit crop behind pineapple.

Despite strict quarantine on movement of papaya seedlings between islands, PRSV was discovered in Puna in May 1992 ([Fig. 3\(a\)](#)) and the Hawaiian papaya industry would be forever changed. By 1995, a third of the papaya growing area was completely infected and much of the rest of Puna had widespread infection ([Fig. 3\(b\)](#)). By 1998, the production of papaya in Puna had dropped to 27 million pounds of papaya from 52 million pounds in 1992 when PRSV was discovered in Puna. In retrospect, the efforts of quarantine, monitoring and rogueing of infected plants in Hilo, and suppression efforts of PRSV in Puna all played key roles in helping Hawaii's papaya industry, because it gave researchers time to develop control measures for PRSV.

Transgenic Resistance

In the late-1980s, an exciting development on Tobacco mosaic virus (TMV) provided a rationale that resistance to plant viruses could be developed by expressing the viral CP gene in a transgenic plant. This approach was called CP-mediated protection, and, at about the same time, a report introduced the concept of "parasite-derived resistance". The report on transgenic resistance to TMV set off a flurry of work in many laboratories to determine if this approach could be used for developing resistance to other plant viruses. Likewise, work was initiated in 1985 to use this approach for developing PRSV-resistant transgenic papaya for Hawaii.

Key requirements for successful development and commercialization of transgenic virus-resistant plants are the isolation and engineering of the gene of interest, vectors for mobilization into and expression of the gene in the host, transformation and subsequent regeneration of the host cells into plants, effective and timely screening of transformants, testing of transformants, and the ability to deregulate and commercialize the product.

The CP gene of the mild strain of PRSV was chosen as the "resistance" gene because it had been recently cloned and it was of the PRSV P type. The commercial cultivars Kapoho, Sunrise, and Sunset were chosen for transformation. In 1991, tests of the R0 lines identified a transgenic Sunset that expressed the CP gene of PRSV HA 5-1, and showed resistance to PRSV from Hawaii. A field trial of R0 plants was started in April 1992 on as the island of Oahu, and a month later PRSV was discovered at Puna in May 1992, as discussed above.

The Oahu field trial showed that R0 plants of line 55-1 were resistant, and line 55-1 was further developed to obtain the cultivar "SunUp" which is line 55-1 that has the CP gene in a homozygous state, and "Rainbow" which is an F1 hybrid of SunUp and the non-transgenic "Kapoho". SunUp is red-fleshed and Rainbow is yellow-fleshed. In 1995, SunUp and Rainbow were tested in a subsequent field trial in Puna and showed excellent resistance ([Fig. 3\(c\)](#)). Due to its yellow flesh and good shipping qualities, Rainbow was especially preferred by the growers. Line 55-1 was deregulated by the US government and commercialized in May 1998. The deregulation also applied to plants that were derived from line 55-1. The timely commercialization of the transgenic papaya in 1998 was crucial since PRSV had decreased papaya production in Puna by 50% that year compared to 1992 production levels. The transgenic Rainbow papaya was quickly adopted by growers and recovery of papaya production in Hawaii was underway ([Fig. 3\(d\)](#)). The transgenic papaya is sold throughout Hawaii ([Fig. 3\(e\)](#)) and the mainland USA, and to Canada where it was deregulated in 2003. Japan, the main importer of Hawaii-grown papaya, approved transgenic papayas for import in 2011. However, several challenges remain: coexistence, exportation, deregulation, and the adoption of transgenic papaya in other countries that suffer from PRSV. The use of transgenic papayas in Hawaii has now been going on for more than 20 years and the resistance is still highly durable.

One limitation of larger-scale deployment of the transgenic papayas lies in the fact that the efficiency of resistance is related to the percentage of identity between the transgene and the challenging PRSV isolate: the transgenic papayas from Hawaii were found susceptible to PRSV isolates from Hainan. The success of transgenic papaya in Hawaii has encouraged other papaya-cultivating states and countries to develop transgenic papayas resistant to their local PRSV strains. Transgenic varieties have thus been developed in America (Florida, Brazil, Venezuela, Jamaica), Asia (China, Indonesia, Malaysia, the Philippines, Taiwan) and Australia. Depending on the construct and the country, the plants were more or less resistant, ranging from complete and durable resistances to easily broken ones. Resistance-breaking could be related to the emergence of new strains that presented a reduced sequence identity with the transgene. In Taiwan, a transgenic papaya initially conferred resistance to PRSV, but the plants became infected with the emerging PLDMV. Transgenic papayas containing partial CP genes of PRSV and PLDMV conferred resistance to both viruses, but the resistance was later broken by a recombinant "super-strain" of PRSV. This shows that viral diversity and rapid evolution remain challenging for a durable control. Pathogen-derived resistance appears as the most effective way to control papaya ringspot disease, but its use should be associated with a good knowledge of the local viral diversity and the use of cultural practices for an integrated management of PRSV.

Although the transgenic approach was not as successful to control PRSV-W in cucurbits, it is used in the USA to control ZYMV, WMV and Cucumber mosaic virus in squash. Interestingly a commercial squash cultivar possessing transgenic resistance to these three viruses and conventional resistance to PRSV-W shows a good protection against these four viruses.



Fig. 3 (a) Healthy Puna papaya in 1992. (b) Severely infected papaya orchards in Puna in 1994. (c) Field trial of transgenic papaya in Puna in 1994: PRSV-resistant transgenic papaya (center) surrounded by PRSV-infected non-transgenic papaya (d) Commercial planting of transgenic papaya one year after releasing seeds of PRSV-resistant transgenic papaya. (e) Transgenic papaya commonly sold in supermarkets. (f) Risk of growing non-transgenic papaya still exists (photograph from 2005). Foreground is PRSV-infected non-transgenic papaya that are cut, and background shows healthy PRSV-resistant transgenic papaya.

Summary Remarks

PRSV has been thoroughly characterized and is a typical member of the family *Potyviridae*, arguably the largest and economically most important plant virus group. The complete genome sequence has been elucidated and infectious transcripts have provided a means to determine the genetic determinants of some important biological functions such as host range and virulence. Furthermore, pathogen-derived resistance has been used to control PRSV-P in Hawaii through the use of virus-resistant transgenic papaya. The transgenic virus-resistant papaya provides a potential means to test the global acceptance of genetically modified (GM) organisms while presenting a plausible approach to control a disease affecting papaya worldwide. In order to avoid resistance-breaking, GM papayas presenting resistance to local strains should be included in an integrated pest management of PRSV.

Cross Check With Other Chapters

Potyviruses (*Potyviridae*); Plum Pox Virus (*Potyviridae*); Potato Virus Y (*Potyviridae*); Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus (*Potyviridae*); Plant Resistance to Viruses: Engineered Resistance; Plant virus disease: Economic aspects, Vector transmission of plant viruses.

See also: Plant Resistance to Viruses: Engineered Resistance. Plant Viral Diseases: Economic Implications. Plum Pox Virus (*Potyviridae*). Potato Virus Y (*Potyviridae*). Potyviruses (*Potyviridae*). Vector Transmission of Plant Viruses. Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus (*Potyviridae*)

Further Reading

- Chen, K.C., Chiang, C.H., Raja, J.A., *et al.*, 2008. A single amino acid of NlaPro of papaya ringspot virus determines host specificity for infection of papaya. *Molecular Plant Microbe Interactions* 21, 1046–1057.
- Desbiez, C., Wipf-Scheibel, C., Millot, P., *et al.*, 2017. New species in the papaya ringspot virus cluster: Insights into the evolution of the PRSV lineage. *Virus Research* 241, 88–94.
- Fuchs, M., Gonsalves, D., 2007. Safety of virus-resistant transgenic plants two decades after their introductions: Lessons from realistic field risk assessments studies. *Annual Review of Phytopathology* 45, 173–202.
- Gibbs, A.J., Ohshima, K., Phillips, M.J., Gibbs, M.J., 2008. The prehistory of potyviruses: Their initial radiation was during the dawn of agriculture. *PLoS One* 3, 1–11.
- Gonsalves, D., 1998. Control of papaya ringspot virus in papaya: A case study. *Annual Review of Phytopathology* 36, 415–437.
- Gonsalves, D.A., Vegas, A., Prasartsee, V., *et al.*, 2006. Developing papaya to control papaya ringspot virus by transgenic resistance, intergeneric hybridization, and tolerance breeding. *Plant Breeding Reviews* 26, 35–78.
- Hamim, I., Borth, W.B., Marquez, J., *et al.*, 2018. Transgene-mediated resistance to Papaya ringspot virus: Challenges and solutions. *Phytoparasitica* 46, 1–18.
- Olarte-Castillo, X.A., Fermin, G., Tabima, J., *et al.*, 2011. Phylogeography and molecular epidemiology of papaya ringspot virus. *Virus Research* 159, 132–140.
- Quiot-Douine, L., Lecoq, H., Quiot, J.-B., Pitrat, M., Labonne, G., 1990. Serological and biological variability of virus isolates related to strains of papaya ringspot virus. *Phytopathology* 80, 256–263.
- Tripathi, S., Suzuki, J.Y., Ferreira, S.A., Gonsalves, D., 2008. Papaya ringspot virus-P: Characteristics, pathogenicity, sequence variability and control. *Molecular Plant Pathology* 9, 269–280.
- Yeh, S.D., Jan, F.J., Chiang, C.H., *et al.*, 1992. Complete nucleotide sequence and genetic organization of papaya ringspot virus RNA. *Journal of General Virology* 73, 2531–2541.

Pecluviruses (*Virgaviridae*)

Hema Masarapu and Pothur Sreenivasulu, Sri Venkateswara University, Tirupati, India

Philippe Delfosse, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Claude Bragard and Anne Legreve, University of Louvain, Louvain-la-Neuve, Belgium

DVR Reddy, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

© 2021 Elsevier Ltd. All rights reserved.

This is an update of D.V.R. Reddy, C. Bragard, P. Sreenivasulu, P. Delfosse, Pecluvirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-0-12374410-4.00542-2](https://doi.org/10.1016/B978-0-12374410-4.00542-2).

Nomenclature

aa Amino acid(s)

AGO Argonaute 1

CITE Cap-independent translation enhancer

Co-Pro Protease-cofactor

CP Coat protein or capsid protein

CRP Cysteine-rich protein

ELISA Enzyme-linked immunosorbent assay

ER Endoplasmic reticulum

HC-Pro Helper component-proteinase

IRES Internal ribosome entry site

kb Kilobase

kDa Kilodalton

LAMP Loop-mediated isothermal amplification

mAbs monoclonal antibodies

MP Movement protein

NCR Noncoding regions

nt Nucleotide(s)

OAS Origin of assembly

ORF Open reading frame

pAbs Polyclonal antibodies

PCR Polymerase chain reaction

RdRp RNA-dependent RNA polymerase

RISC RNA-induced silencing complex

RT-qPCR Reverse transcription quantitative PCR

satRNA satellite RNA

UTR Untranslated region

VIGS Virus-induced gene silencing

VLPs Virus-like particles

VPg Viral protein genome-linked

VRC Virus replication complex

vRNA virion RNA

Glossary

Forma specialis An informal rank in a classification, allowed by the International Code of Nomenclature for Algae, Fungi, and Plants, that is applied to a parasite (most frequently a fungus), which is characterized from a physiological standpoint (e.g., by the ability to infect a specific host), but scarcely or not at all from a morphological standpoint.

Fortuitous hosts of obligate biotrophic

Plasmodiophorida Host on which some primary infections evolve into resting spores without going through the zoosporangial stage.

Hetero-encapsidation Partial or full coating of the genome of one virus with the coat protein of a differing

virus. Also termed trans-capsidation or heterologous encapsidation.

Leaky scanning Mechanism by which the ribosomes fail to initiate translation at the first AUG start codon and scan downstream for the next AUG codon.

Obligate endoparasite A parasitic organism that cannot complete its life cycle without exploiting a suitable living host, and that lives within that host.

Post-transcriptional gene silencing Mechanism for sequence-specific RNA degradation in plants.

Triple gene block A specialized evolutionarily conserved gene module involved in the cell-to-cell and long-distance movement of plant viruses.

History

Peanut (groundnut, *Arachis hypogaea*) clump disease, characterized by severe stunting and clumping, was first described from India in 1927 and subsequently from West Africa in 1931. The causal agent in West Africa as well as in India was identified as a virus, *Peanut clump virus* (PCV) and *Indian peanut clump virus* (IPCV), respectively. Annual losses to the peanut crop due to PCV and IPCV were estimated to exceed US\$38 million in the late 1990s. Pecluviruses also cause diseases on several dicotyledonous and monocotyledonous crops.

Taxonomy and Classification

The genus *Pecluvirus* has only two species, isolates from West Africa were grouped under the species *Peanut clump virus* (Peanut clump virus; PCV) and those from the Indian subcontinent were grouped under the species *Indian peanut clump virus* (Indian peanut clump

virus; IPCV). The two virus species were initially assigned to the genus *Furovirus* mainly based on virion morphology and vector transmission by *Polymyxa graminis* (Rhizaria, *Phytomyxea*, *Plasmodiophorida*). Subsequently, based on molecular features, the ICTV in 1997 assigned PCV and IPCV to the newly established genus *Pecluvirus* (siglem from Peanut clump virus). They differ in their geographical distribution, host range, antigenic properties and genomic sequences. Later, the genera *Tobamovirus*, *Tobravirus*, *Hordeivirus*, *Furovirus*, *Pecluvirus*, *Pomovirus*, and *Goravirus* were assigned to the newly established family *Virgaviridae* (from the Latin *Virga*, meaning rod).

Geographic and Seasonal Distribution

PCV occurs in Burkina Faso, Benin, Chad, Congo, Côte d'Ivoire (Ivory Coast), Niger, Mali, Gabon, Senegal, and Sudan, and IPCV in the Indian states of Punjab, Andhra Pradesh, Telangana, Tamil Nadu, Rajasthan, Haryana, and Gujarat, and in Pakistan (Sind and Punjab provinces). IPCV and PCV infections are noticed mainly in the peanut crops cultivated during the rainy season (July–November in India and April–October in West Africa). Nevertheless, post-rainy season crops can also be infected.

Host Range and Symptoms

PCV infects peanut, cowpea, sugarcane, maize, sorghum, pearl millet, and finger millet, whereas IPCV infects peanut, cowpea (*Vigna unguiculata*), pigeonpea (*Cajanus cajan*), chili (*Capsicum annuum*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), maize (*Zea mays*), sorghum (*Sorghum bicolor*), pearl millet (*Pennisetum glaucum*), foxtail millet (*Setaria italica*), and finger millet (*Eleusine coracana*). *Cynodon dactylon*, *Cyperus rotundus*, *Oldenlandia aspera*, and *Vigna subterranea* (bambara groundnut) that can play a significant role in both virus survival and dissemination.

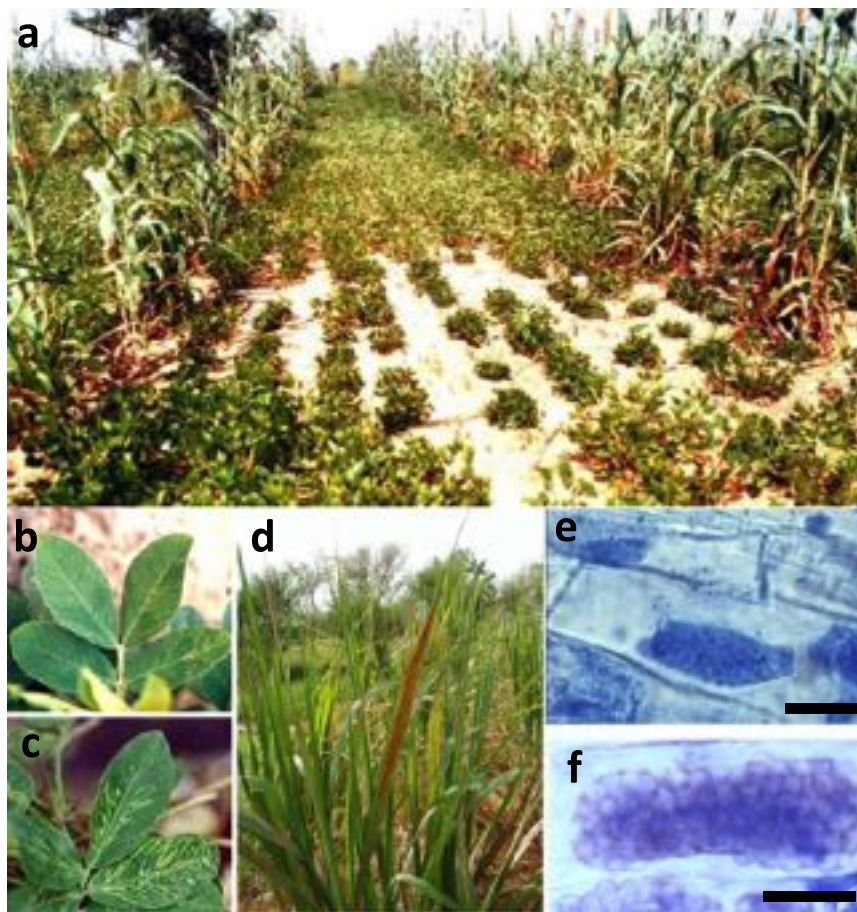


Fig. 1 Symptoms induced by Peanut clump virus (PCV), in West Africa. (a) field-infected peanut plants occurring in patches of various sizes, intercropped with pearl millet in Maradi, Niger Republic, (b) peanut leaflets showing initial symptoms of chlorosis and mild mosaic, (c) chlorosis along the veins resembling oak leaf, (d) PCV-infected sugarcane plants exhibiting red leaf mottle in Burkina Faso, (e) IPCV and PCV vector, *Polymyxa graminis*, in transversal section of sorghum roots, stained with lactophenol blue showing zoosporeangia, (f) sporosori, bars represent 25 µm. Photographs courtesy of P. Delfosse.

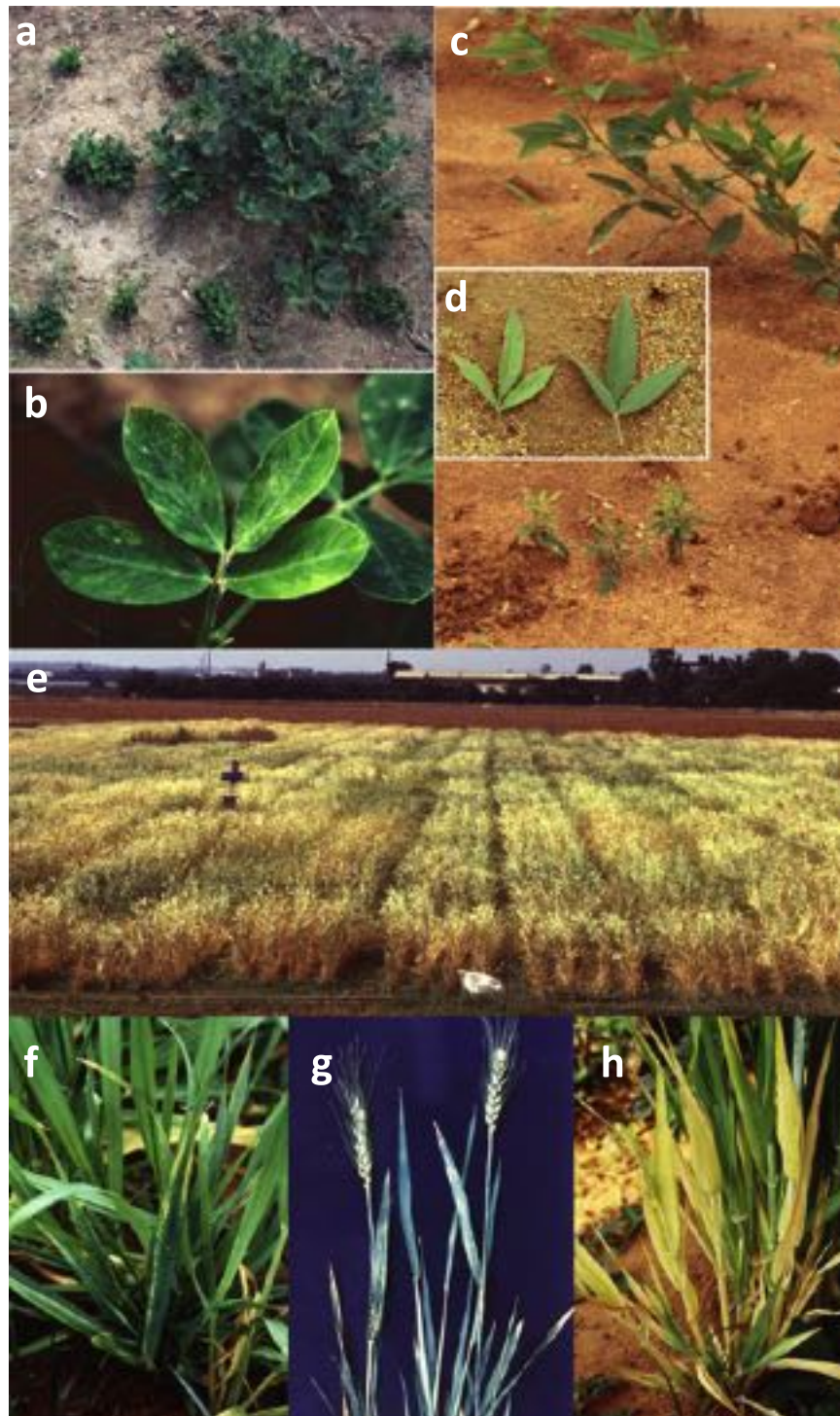


Fig. 2 Symptoms induced by Indian peanut clump virus (IPCV) on various hosts. (a) on peanut plants showing severe stunting, occurring in patches in the field, and an apparently uninfected plant, (b) young peanut leaflets with initial symptoms of chlorosis and mosaic symptoms, (c) early infected pigeonpea plants showing severe stunting with an apparently uninfected plant, (d) inset showing chlorotic symptoms on young pigeonpea leaflets (left) and by its side leaflets from healthy plant, (e) infected wheat plants occurring in patches, (f) early infected wheat plants showing dark green leaves and chlorotic streaks, (g) late infected wheat plants showing panicle formation. Seed derived from them can transmit IPCV up to 1%, (h) early infected barley plants showing stunting and severe yellowing. Photographs courtesy of P. Delfosse.

Diseased peanut plants are conspicuous under field conditions because of their dark green appearance, stunting, and occurrence in patches (Fig. 1(a) and Fig. 2(a)). The disease incidence is high in fields where peanut crops are grown in rotation with wheat, barley, and pearl millet. Early-infected peanut plants often did not yield pods and late-infected crops showed decreased crop yields up to 60%.

In peanut, the symptoms incited by PCV are essentially similar to those induced by IPCV. In West Africa, peanut clump disease symptoms are often confused with green rosette because of the presence of stunted dark green plants occurring in patches in the fields. The peanut clump disease occurs in the same areas year after year with marginal increase in the periphery of the patches in the successive peanut crops. Initial symptoms on young leaflets of peanut are mottling, mosaic and chlorotic veins or the 'oak leaf symptom' (Fig. 1(b) and (c), Fig. 2(b)). When such leaflets mature, they turn dark green with or without faint mottling. Late infected plants may not show conspicuous symptoms and clumping is restricted to a few branches in some plants. PCV causes the red leaf mottle disease in sugarcane (*Saccharum officinarum*) (Fig. 1(d)).

Pigeonpea plants infected with Hyderabad isolate of IPCV (IPCV-H) showing stunting are presented in Fig. 2(c). Fig. 2(d) shows chlorosis on infected pigeonpea leaflets, left, and uninfected leaflets, right. Wheat plants, up to three weeks old, infected by IPCV-H show rosette symptoms similar to those caused by the *Soil-borne wheat mosaic virus* (SBWMV) and grain yield losses up to 58% are recorded. Wheat cultivar RR-21 infected with IPCV-H showing stunting is presented in Fig. 2(e). Early infected wheat plants show chlorotic streaks (Fig. 2(f)) and will produce panicles (Fig. 2(g)). Seed from these plants can transmit IPCV up to 1%. Barley plants infected with IPCV-H are stunted and bushy with chlorotic or necrotic leaves (Fig. 2(h)) and the majority of these plants die. The plants that reached maturity produced poorly developed spikes. IPCV-H also infected pearl millet, finger millet, foxtail or Italian millet, and sorghum plants. In maize IPCV-H caused aerial biomass losses up to 33% and grain loss up to 36%.

Various isolates of PCV and IPCV are readily detected in the cells of roots, stems, and leaves of systemically infected hosts. In wheat cells, PCV particles are found in the cytoplasm, near the nucleus or along the plasmalemma, and arranged in angled-layer aggregates.

In experimental host range studies by sap inoculation, PCV and IPCV infected a wide range of both dicotyledonous and monocotyledonous plants. For IPCV, *Nicotiana clevelandii* x *N. glutinosa* hybrid and *Phaseolus vulgaris* (cv. Top crop) are suitable for virus propagation and as assay/diagnostic hosts, respectively. *Chenopodium amaranticolor* and *C. quinoa* are local lesions hosts. IPCV isolates collected from clump diseased peanut crops from different locations in India differ slightly in their host ranges. *Canavalia ensiformis* and *N. clevelandii* x *N. glutinosa* hybrid are found to be useful for differentiating the isolates of IPCV. In the case of PCV, *C. amaranticolor*, *N. benthamiana*, *N. glutinosa*, *P. vulgaris*, and *Triticum aestivum* are of diagnostic value. *N. benthamiana* and *P. vulgaris* are suitable for virus propagation. The symptoms induced in *C. amaranticolor* by various PCV isolates collected from Burkina Faso, Senegal, and Niger are shown to differ markedly.

Transmission

Seed and Sap Transmission

Pecluviruses can be transmitted by sap inoculation and through peanut seed. The transmission rate through the peanut seed depends on the cultivar, the mode of infection (up to 24% for plants infected through soil and up to 50% for plants infected through seed) and the age at which the plants are infected. IPCV is transmissible up to rates of <2% through the seeds of pearl millet, finger millet, foxtail millet, wheat, and maize. Pearl millet CV GGP-16 did not transmit IPCV through seed.

Soil-Borne Vector Transmission

Both PCV and IPCV are naturally vectored by soil-borne protists, *P. graminis* (Fig. 1(e) and (f)). Two formae speciales of this species are associated with the transmission of PCV and IPCV: *P. graminis* f. sp. *subtropicalis* and *P. graminis* f. sp. *tropicalis*. These obligate root endoparasites complete their entire life cycle, including sporangial and sporogenic (resting spores) phases (Fig. 1(e) and (f); Fig. 3 right panel), only in monocotyledonous hosts in the family *Poaceae* (*P. graminis tropicalis* on sorghum, pearl millet, and maize but rarely on wheat and barley and *P. graminis subtropicalis* on sorghum, pearl millet, wheat, and barley). Nonetheless, they can infect and transmit viruses to dicotyledonous hosts, including peanut, which are considered to be fortuitous hosts of the vector. The presence of *P. graminis* in infested soils can be identified by growing bait plants such as pearl millet, sorghum, or wheat on infested soils or by PCR. *In vitro* transmission studies under controlled conditions is complicated because of the obligate parasitic nature of *P. graminis* and consequent difficulties in maintaining the cultures of both the virus and the vector in the same hosts. Nevertheless, the acquisition and the transmission of PCV by the vector is demonstrated using certain hosts such as sugarcane (CV CP-89-2377) and *Sorghum arundinaceum* supporting abundant zoospore production and virus infection. Presence of *P. graminis* in roots can also be quantified by reverse transcription quantitative PCR (RT-qPCR). The P39 viral protein, encoded by RNA2 and expressed by leaky scanning of P23, is involved in the transmission by *P. graminis* based on its similarity with P75 protein encoded by RNA2 of *Beet necrotic yellow vein virus* (BNYVV).

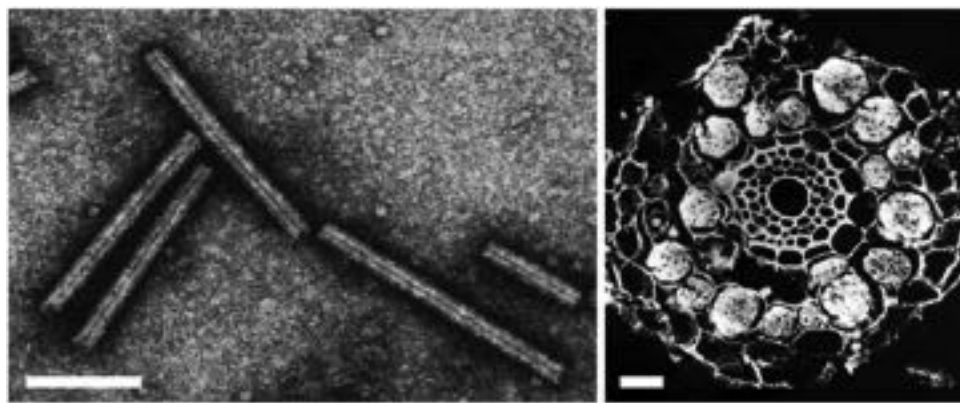


Fig. 3 Electron micrographs of the Indian Peanut clump virus (IPCV; *Pecluvirus*, *Virgaviridae*) and its vector *Polymyxa graminis*. (left) Virus particles with two predominant lengths of 249 and 184 nm are decorated with homologous antiserum for the Hyderabad isolate, bar represents 100 nm. (right) Transversal section of a sorghum root heavily infested by *P. graminis* showing sporosori, bar represents 25 μm . Photographs courtesy of P. Delfosse and Larry Littlefield.

Virion Morphology

The virions of IPCV (Fig. 3 left panel) and PCV are non-enveloped rigid straight rod shaped, 24 nm (IPCV) and 21 nm (PCV) in diameter with two predominant lengths of c. 250 and c. 180 nm. They have helical symmetry with a pitch of 2.6 nm and contain a single coat protein (CP) of 23 kDa.

Physical Properties

Virions sediment as two major components with S_{20w} of 183S and 224S. Buoyant density in CsCl is 1.32 g cm^{-3} and isoelectric point is pH 6.45. Thermal inactivation of virus infectivity occurs at 64°C . Virions are stable in frozen leaves.

Molecular Properties

Genome Structure and Expression Strategies

The genome of pecluviruses is bipartite, with two positive-sense, linear single-stranded RNAs (ssRNAs) (4% in weight). RNA1 is c. 5900 nt long and RNA2 is c. 4500 nt long. The PCV RNAs show little sequence identity with those of IPCV RNAs except for the 3' terminal 273 nt, which are conserved among the two RNAs. Both the RNAs have methylated cap structure (to be confirmed) at their 5' ends. The 3' ends of the RNAs have tRNA-like structure and are not polyadenylated. Both RNAs are flanked by untranslated regions (5'- or 3' UTR) or noncoding regions (5'- or 3' NCR). The genome organization and expression of the PCV is shown in the Fig. 4.

Coding Sequences

RNA1 contains two open reading frames (ORFs) that encode two proteins involved in viral RNA replication (P131 and P191, Fig. 4; Table 1). The proximal ORF encodes a 131 kDa polypeptide and P191 is a C-terminally extended form of P131 produced by translational readthrough of the UGA termination codon. The P15 ORF is downstream of the P191 ORF and separated from it by a noncoding region of about 60 nt. The polypeptides of 131 and 191 kDa contain NTP-binding methyl transferase, helicase and RNA-dependent RNA polymerase (RdRp) domains and are involved in the putative replication complex (Table 1). The 15 kDa polypeptide, a cysteine-rich protein (CRP) is translated from a sub-genomic RNA (Fig. 4). It is a suppressor of post-transcriptional gene silencing (PTGS). The P15 possesses four C-terminal proximal heptad repeats that can generate a coiled-coil interaction and is targeted to peroxisomes via a C-terminal Ser-Lys-Leu (SKL) motif. Such a motif is conserved among pecluviruses from both Africa and India. It has been demonstrated that a coiled-coil motif is necessary for the anti-PTGS activity of P15, but the peroxisomal localization signal is not, although it is required for efficient intercellular movement of the virus. P15 resembles CRPs of *Barley stripe mosaic virus* (BSMV), *Poa semilatifolia virus* (PSLV), and SBWMV. These proteins have been suggested to act as regulatory factor during virus replication as well as for long-distance movement and contribute to the virulence factor. RNA1 can replicate in the absence of RNA2 in protoplasts of tobacco BY-2 cells. However, both RNAs are required for infection and systemic invasion of plants. Experiments using enhanced 5' green fluorescent protein and 5'-bromouridine 5'-triphosphate labels have suggested that PCV replication complexes co-localize with endoplasmic reticulum (ER) bodies accumulating around

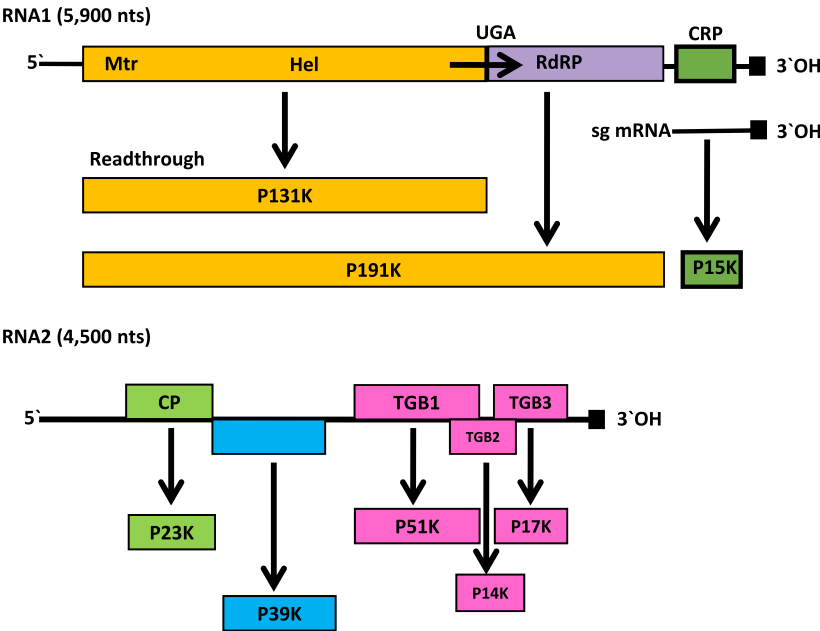


Fig. 4 Genome organization of PCV (genus *Pecluvirus*, family *Virgaviridae*). The tRNA structure motifs at the 3' termini of the RNAs are represented by a dark square. ORFs are indicated by rectangles and a suppressible termination codon (UGA) by an arrow. Mtr, methyl transferase; Hel, helicase; RdRp, RNA-dependent RNA polymerase; CRP, cysteine-rich protein; CP, coat protein; TGB, triple gene block. Modified from Adams, M.J., Adkins, S., Bragard, C., *et al.*, 2017. ICTV virus taxonomy profile: Virgaviridae. Journal of General Virology 98, 1999–2000.

Table 1 Pecluvirus open reading frames (ORFs), polypeptides and their functions

Genomic RNA	ORF	<i>M_r</i> of polypeptide (kDa)	Function
RNA1	P131	131	Methyltransferase, helicase, replicase
	P191	191	
	P15	15	Suppressor of PTGS
RNA2	P23	23	Coat Protein
	P39	39	Putative vector transmission factor
	P51	51	Virus movement (triple gene block)
	P14	14	
	P17	17	

the nucleus during infection. P15 does not act directly at sites of viral replication but intervenes indirectly to control viral accumulation levels.

RNA2 contains five ORFs (Fig. 4). The 5' proximal ORF (23 kDa) encodes the CP, and the adjacent ORF (ORF2) encodes a 39 kDa polypeptide which is expressed by leaky scanning mechanism *in vitro* and thought to be involved in the transmission of PCV by its vector. ORF2 starts 1 nt upstream of the first residue of UGA stop codon of the CP cistron. Further, downstream and separated by 135 nt intergenic region, is a triple gene block (TGB) that codes for proteins of 51, 14, and 17 kDa that are presumed to be involved in the cell to cell movement.

5' and 3' Noncoding Sequences

The 5' and 3' NCRs of RNA1 of pecluviruses are about 130 and 300 nt in length, respectively (Fig. 4), whereas the RNA2 has more diverse NCR of 390–500 nt at 5' end and c. 300 nt at 3' end. Within the 3' NCR of RNA2, approximately 100 terminal nt are conserved among pecluvirus RNAs sequenced so far. The NCRs differ in size among isolates from the different serotypes. The 3' NCR of pecluviruses, as in the case of furoviruses, forms a t-RNA-like structure (TLS) that aids in the replication of both RNAs.

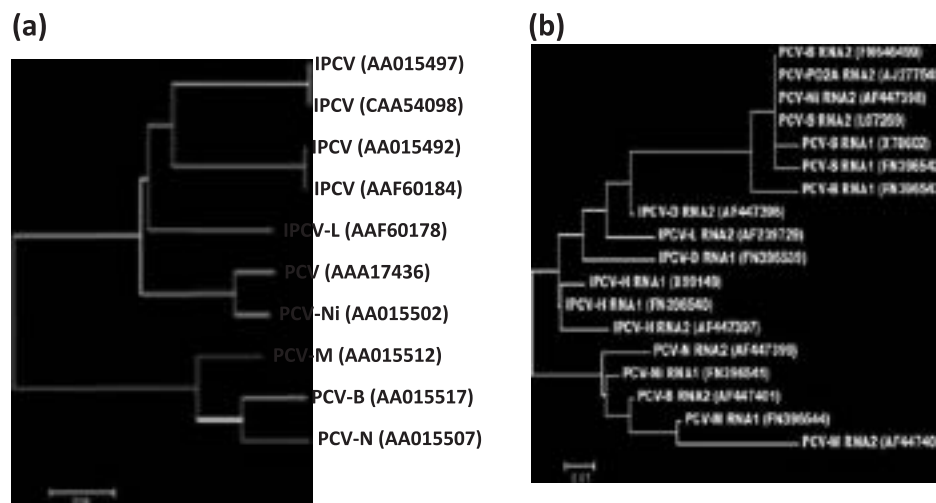


Fig. 5 (a) Phylogenetic tree constructed using coat protein (CP-P23) amino acid sequences from eight PCV/IPC isolates. (b) Phylogenetic tree constructed using 18 sequences of the 3' UTR of RNA1 and RNA2 of PCV/IPC isolates. The evolutionary distances were computed using Maximum Composite Likelihood method by Tamura, K., Dudley, J., Nei, M. Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24, 1596-1599. GenBank Accession numbers are given in parentheses.

Sequence Comparisons and Phylogeny

Only one complete RNA1 genome sequence of each of PCV (X78602) and IPCV (X99149) are available, while five and three complete RNA2 sequences of PCV (L07269, AF447401, AF447400, AF447399, AF447398), and IPCV (AF447397, AF239729, AF447396), respectively are available in the database.

PCV and IPCV exhibited considerable differences in homology depending on the part of the genome sequence compared. The complete nt sequence of RNA1 of PCV is 78% similar to that of IPCV. The RNA1-encoded polypeptides (P191, P131, and P15) of PCV and IPCV share identities ranging from 75% (P15) to 95% (read-through portion of P191) and show significant similarities with furoviruses (e.g., 56% identity with polymerase of SBWMV). Comparison of the P15 gene shows a close relationship between IPCV isolates and a relatively high diversity among PCV isolates. The full-length sequence comparisons between RNA2 of PCV and IPCV isolates revealed a high degree of variability in size (between 58%–79%) and the proteins encoded by RNA2 are 39%–89% identical between species. Among the five ORFs of RNA2, ORF4 encoding P14 of TGB is highly conserved (90%–98%), whereas ORF2 encoded P39 is less conserved (25%–60%). Phylogenetic comparisons, based on complete RNA2 sequences, showed that the eight isolates could be grouped into three distinct clusters with no geographical distinction between PCV and IPCV isolates. Phylogenetic analysis of individual ORFs revealed an overall similarity with that obtained from complete RNA2 sequences, but the relative positions of individual isolates varied within each cluster. These studies indicate that there is substantial divergence among the RNA2's of pecluviruses and suggest that different polypeptides have evolved differently, possibly due to different selection pressures.

In the most members of the family *Virgaviridae*, CP is conserved among the isolates, but in the case of PCV and IPCV isolates, coat proteins are highly diverse (aa sequence identities between 37% and 89%) and have c. 30% similarity with the CP of BSMV (genus *Hordeivirus*). A conserved motif 'F-E-X₆-W' is present near the CP C-terminus of all three IPCV serotypes and PCV, as in the CPs of other rod-shaped viruses [*Tobacco mosaic virus* (TMV) and *Tobacco rattle virus* (TRV)]. Phylogenetic analysis of CP of eight PCV and IPCV isolates (PCV-B, PCV-M, PCV-N, PCV-Ni, PCV-S, IPCV-D, IPCV-H, IPCV-L) revealed that the PCV isolates grouped into two major clusters and IPCV into one cluster (Fig. 5(a)). The TGB proteins resemble those of *Potato mop top virus* (PMTV, genus *Pomovirus*).

Comparison of 3' end nt sequences revealed that they are highly conserved (88% for RNA1 and 85% for RNA2), thus offering a potential target for broad-spectrum detection by RT-PCR. Conservation of 3' UTR is verified by the alignment of eighteen 3' UTR (124 nt long) sequences of RNA1 and RNA2 (collected from NCBI database) of PCV/IPC isolates from different geographical regions and a phylogenetic tree has been constructed using the neighbor-joining method (Fig. 5(b)). The isolates were grouped into one Indian and two African clusters. Indian cluster appeared to be intermediate between the African clusters. Though there is high percentage of identity between 3' UTR of RNA1 and RNA2 of PCV and IPCV isolates from different geographical regions, there is no evidence of reassortment between isolates. Putative recombination events and their frequencies in the complete genome of 168 accessions extracted from the International databases representing several members of six genera Viz. *Furovirus*, *Pecluvirus*, *Comovirus*, *Tobravirus*, *Hardeivirus*, and *Tobamovirus* indicate that exchange of heredity material occurred between non-homologous sequences like those of IPCV and PCV.

Assembly of Virus Particles

Sequences required for assembly of PCV virions have been established by testing the ability of encapsidation of deletion mutants of RNA1 and RNA2 *in vivo*. A putative origin of assembly (OAS) is mapped in the 5' proximal part of the P15 gene of RNA1. Two sequence positions, 5' proximal CP gene and the other in the p14 gene near the RNA2 3' terminus that can drive encapsidation of RNA2, have been identified. No obvious sequence similarities between different assembly initiation sequences are noted.

The possible localization of an OAS in the CP gene of IPCV is realized as a result of the formation of virus-like particles (VLPs) in *Escherichia coli* and *N. benthamiana*, confirming the results observed for PCV. The monomer VLP size corresponded to the length of the encapsidated CP gene transcript RNA. Using immunocapture RT-PCR (IC-RT-PCR), such VLPs have been demonstrated to contain RNA encoding IPCV-H CP. Transgenic *N. benthamiana* expressing IPCV-H CP gene inoculated with the serologically distinct IPCV-L serotype accumulated virus particles that contained both types of CP, indicating that hetero-encapsidation occurs in transgenic plants.

Detection and Diagnosis

Pecluvirus diseased peanut plants occur in patches, mostly restricted to sandy/sandy-loam soils. The diseased peanut plants are stunted and dark green in color with or without chlorotic symptoms on leaves. Pecluvirus affected cereal and millet plants and grass weeds are difficult to recognize because they show mild or no visible symptoms. Diseases caused by pecluviruses reappear nearly in the same location year after year.

Immunoassays

Several immunological tests utilizing polyclonal (pAbs) and monoclonal antibodies (mAbs) have been used for the detection and identification of pecluvirus isolates. This approach revealed a diversity of serotypes and serogroups in Africa and India. The pAbs of PCV did not react with IPCV, BSMV, TMV, BNYVV, SBWMV, and PMTV. Wide serological diversity exists among PCV and IPCV isolates. Thus, serological tests have limitations to detect more than one isolate from a single antibody source. Antisera to different isolates of IPCV facilitated the grouping of isolates into three serotypes, IPCV-H (Hyderabad), IPCV-D (Durgapura), and IPCV-L (Ludhiana). All IPCV isolates are serologically distinct from PCV isolates and vice-versa. Utilizing a panel of mAbs raised against a PCV isolate and four different ELISA formats, several PCV isolates are grouped into five serotypes. Surprisingly, one of the mAbs reacted with IPCV-D in triple antibody sandwich ELISA. ISEM and alkaline phosphate or penicillinase-based ELISA formats have been applied for detection of both PCV and IPCV in various investigations.

Molecular Assays

It is essential to employ diagnostic tests that are highly sensitive and broadly specific for the detection of pecluviruses in disease surveys, to eliminate virus-contaminated sources in quarantine, and to devise strategies for disease control.

A broad-spectrum RT-PCR protocol, based on the conserved sequences of the 3' end region of genomic and sub-genomic RNAs of several virus isolates, has been developed to detect all the currently known IPCV and PCV isolates in various plant samples originated from different countries. Primers are also designed for quantitation of pecluvirus isolates using a TaqMan-based real time RT-PCR (RT-qPCR). RT-PCR is compared with penicillinase-based DAS-ELISA for the detection of pecluvirus isolates in crop and weed plants. The virus is detected in 70% of suspected plants by RT-PCR as compared to 24% detected by DAS-ELISA. The RT-PCR enabled the detection of virus in the plants that did not exhibit overt symptoms. The detection and quantification of PCV and IPCV isolates (IPCV-L, IPCV-H, IPCV-D, PCV-B, PCV-M, PCV-N, PCV-S) is performed by RT-qPCR. These PCR-based techniques are very useful for detecting PCV and IPCV especially in non-symptomatic hosts. RT-qPCR is recognized as an indispensable tool for tracing resistant cultivars and for the maintenance of pecluvirus-free germplasm. Immunocapture RT-PCR has also been used to demonstrate the presence of RNA in PCV VLPs.

Vector

Separation of *P. graminis* resting spores from sorghum root material facilitated pAbs production. In DAC-ELISA, they detected one sporosorus per well of the ELISA plate. In spiked root samples, the procedure detected one sporosorus per mg of dried sorghum roots. The majority of *P. graminis* isolates from Europe, North America, and India reacted strongly with pAbs. However, *P. graminis* isolates from Rajasthan state (North India), from Pakistan, and an isolate from Senegal (West Africa) reacted weakly with pAbs. The DAC-ELISA procedure is applied to detect various stages in the life cycle of *P. graminis* and to detect infection that occurs under natural and controlled environments. Resting spores are detected in root sections by using fluorescein 5-isothiocyanate - labeled pAbs.

Probes specific for nuclear ribosomal DNA (rDNA) detected the presence of *P. graminis* f. sp. *subtropicalis*, *tropicalis*, and *temperata* in roots by RT-PCR. *P. graminis* f. sp. *tropicalis* is quantified in plants inoculated with viruliferous zoospores using a specific RT-qPCR. The ribotypes of *P. graminis* f. sp. *temperata* and *P. graminis* f. sp. *tepida* are identified by restriction fragment length polymorphism analysis of PCR amplified rDNA, nested and multiplex PCR.

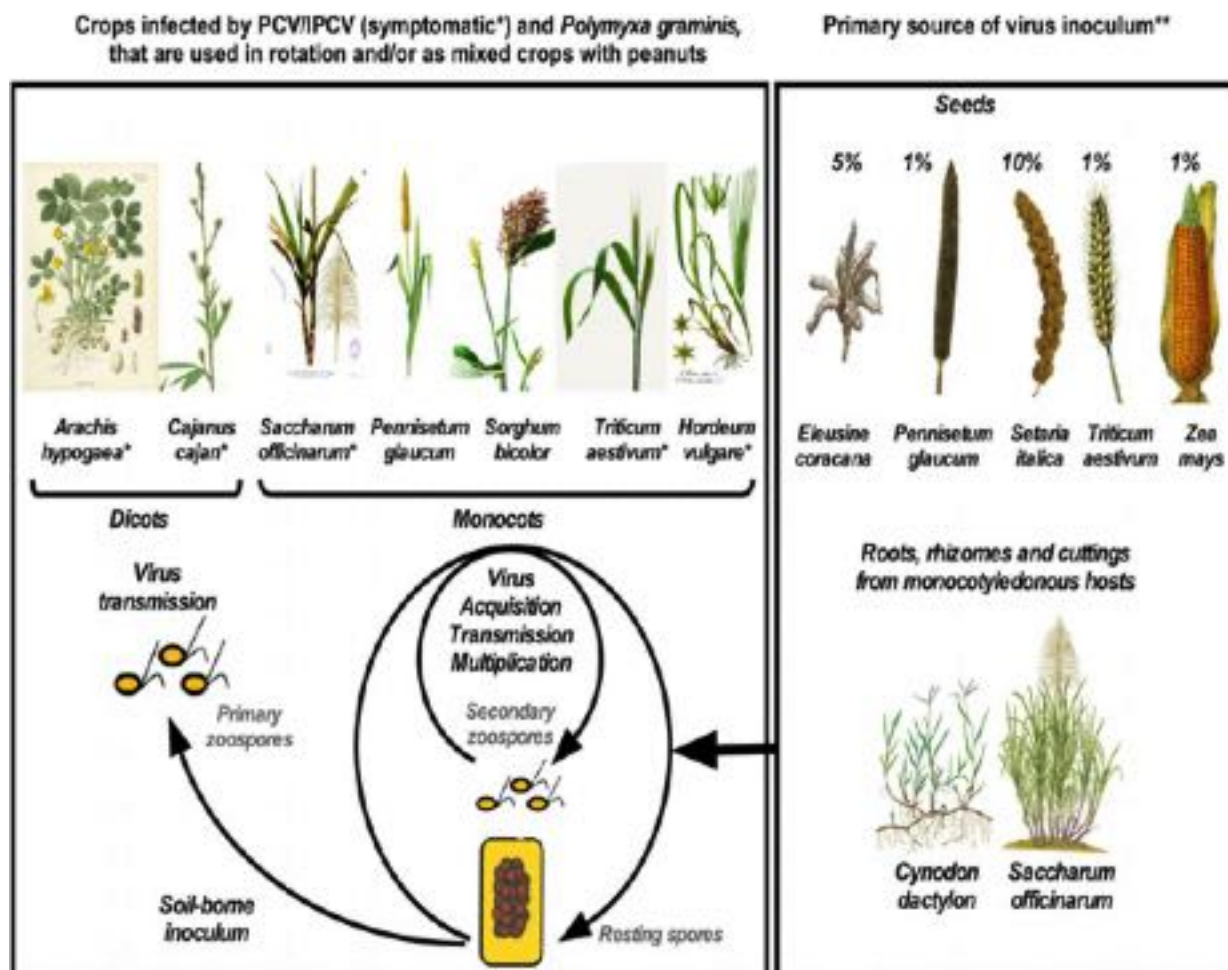


Fig. 6 Disease development cycle of Pecluviruses showing ways of acquisition, multiplication, and transmission by *Polymyxa graminis* through a variety of host plants. Modified from Delfosse P., 2000. Epidemiology and Management of the Indian peanut clump virus. Louvain-la-Neuve, Belgium: Université catholique de Louvain, PhD Thesis. Dieryck, B., Otto, G., Doucet, D., *et al.*, 2009. Seed, soil and vegetative transmission contribute to the spread of pecluviruses in West Africa and the Indian sub-continent. *Virus Research* 141, 184–189. Seed transmission data from Reddy, A.S., Hobbs, H.A., Delfosse, P., Murthy, A.K., Reddy, D.V.R., 1998. Seed transmission of Indian peanut clump virus (IPCV) in peanut and millets. *Plant Disease* 82, 343–346.

Epidemiology and Disease Cycle

The study of epidemiology and formulation of cost-effective disease control practices depend on the precise and sensitive detection of PCV and IPCV and their vector in plants, seeds, and soil. Peanut clump disease is largely confined to sandy soils and loams. Detection is required to identify infested fields, to select virus-free seed lots and vegetative propagules, to identify alternate hosts, and to assess the resistance of peanut, sugarcane and cereal and millet crops breeding lines.

Disease Development Cycle

The presence of viruliferous *P. graminis* vector, adequate soil moisture to facilitate the mobility of the vector zoospores from plant-to-plant in the sandy and sandy-loam fields, the optimum temperatures conducive to vector transmission and planting with seeds of virus-infected cereal crops contribute to primary spread. Presence of cereal hosts, ambient temperatures near 30°C and adequate moisture are vital for secondary spread. The relative contributions of favorable hosts for both the pecluvirus and its vector in the disease cycle are summarized (Fig. 6).

Dicotyledonous hosts limit the multiplication of the plasmodiophorid vector and hence considered them as fortuitous hosts that may not contribute to the perpetuation of the virus inoculum. Indeed, virus-infected peanut roots could not transmit or establish the disease. Monocotyledonous plants such as maize, sorghum, and pearl millet are 'preferred' hosts of the vector and contribute to the build-up of vector population potential in the soil. Seed of millets, maize and wheat, rhizomatous grasses (e.g., *C. dactylon*) and sugarcane vegetative propagules are likely to contribute to the disease establishment in new areas by supporting

the multiplication of the virus and vector. Soils from disease infested areas containing roots/rhizomes harbor resting spores of the vector. Dispersal of such infested soils by wind and displacement during tillering and plowing can contribute to the establishment of the disease in disease-free areas. During the rainy season, prolonged dry spells followed by supplementary irrigation in peanut fields may also contribute to the dispersal of resting spores and disease spread.

Disease Control

Research during the past four decades has led to the accumulation of considerable information on IPCV and PCV. It is now apparent that pecluviruses are not the viruses of peanut but should be regarded as viruses of cereals that opportunistically infect peanut. The difficulty in controlling such viruses is that they are both seed (in peanut, finger millet, foxtail millet, pearl millet, maize, wheat) and soil transmitted (through thick walled resting spores of the vector embedded in roots/rhizomes) and can also be propagated through sugarcane stem cuttings or the rhizomes of grass weeds. Continuous monocropping with fortuitous hosts such as peanut, bambara groundnut and pigeon pea will contribute to a reduction in plasmodiophorid population in soil. Seeds from infected cereal hosts should be avoided for planting. Thus, understanding precisely the sources of virus and vector and transmission mechanisms for each type of host plant in virus endemic agro-ecosystems using sensitive and specific diagnostic tools may lead to formulation of effective disease control practices.

Initial experiments with IPCV in India showed that application of soil biocides, e.g., DD (a mixture of 1,3-dichloropropene and 1,2-dichloropropane), fumigant nematicides that also have fungicidal action and soil solarization are effective on a limited scale in decreasing the disease incidence in peanut fields. Clump affected peanut plants always occur in patches and easy to avoid. Additionally, under field conditions early infected, which is the case, hardly produce any seed.

Cultural Practices

The following crop cultural practices are suggested for reducing disease incidence:

- Early planting of peanut, before the onset of monsoons under judicious irrigation.
- Planting with trap crops such as pearl millet at a high density, plowing of two-week-old seedlings into the soil followed by planting with peanut seed.
- Avoid crop rotation and intercropping with cereal and millet crops.
- Adopt continuous cropping with dicotyledonous crops (peanut, bambara groundnut, cowpea, pigeon pea) for at least three successive growing seasons.

Host Plant Genetic Resistance

No resistance to IPCV is found in over 9000 cultivated *Arachis* germplasm. Unfortunately attempts to incorporate resistance to IPCV by genetic engineering, deploying CP gene, have not produced successful results. However, cultivation with cereal hosts that did not show any seed transmission may help in limiting the disease establishment at *hitherto* disease-free locations.

Future Perspectives

At present the geographical distribution of pecluviruses is restricted to the Indian subcontinent and West Africa. Locations where peanut crops are grown in rotation or as mixed crops with cereal crops on sandy soils can be regarded as targets for establishment of clump disease. Hence measures must be taken to avoid planting with seeds from cereal crops resourced from clump infested fields. They also should not be used towards germplasm exchange because they pose a quarantine risk of introduction of pecluviruses into new geographical regions. Mapping of locations with peanut crops grown in sandy soils in rotation with cereal crops and peanut crops intercropped with cereals, along with the temperature range required for infection, should help in determining the potential for occurrence of pecluviruses in new locations.

RNA interference (RNAi)-mediated resistance has been found to be highly effective in controlling RNA virus infections, and such approach needs to be exploited to produce pecluvirus-resistant crops. The suspected role of P39 in vector transmission of the pecluvirus needs to be confirmed by mutational analysis in order to deploy this ORF in transgenic research. The presence of sub-genomic RNAs encoding P14 and P17 in infected tissues needs to be verified. The cultural control measures tested for IPCV are worth exploiting in West Africa to minimize the impact of PCV. Development of duplex or multiplex RT-PCR is necessary to distinguish peanut clump and groundnut green rosette diseases that visually look alike. Broadly specific nonradioactive probes must be made available to researchers in Africa and India.

Further Reading

- Adams, M.J., Antoniw, J.F., Kreuze, J., 2009. *Virgaviridae*: A new family of rod-shaped plant viruses. *Archives of Virology* 154, 1967–1972.
- Adams, M.J., Adkins, S., Bragard, C., *et al.*, 2017. ICTV virus taxonomy profile: *Virgaviridae*. *Journal of General Virology* 98, 1999–2000.
- Boulila, M., 2012. Molecular evolution and genetic distinctness among rod-shaped plant RNA viruses of the family *Virgaviridae*. In: Chiheb, B., Khalil, B. (Eds.), *Bioinformatics Research: New Developments*. New York: Nova Science Publishers, pp. 107–130.
- CABI/EPPO, 2006. Peanut clump virus. *Distribution Maps of Plant Diseases*. Wallingford: CAB International, (No. 988).
- Delfosse, P., Reddy, A.S., Legrève, A., *et al.*, 1999. Indian peanut clump virus (IPCV) infection on wheat and barley: Symptoms, yield loss and transmission through seed. *Plant Pathology* 48, 273–282.
- Delfosse, P., 2000. *Epidemiology and Management of the Indian peanut clump virus*. Louvain-la-Neuve: Université catholique de Louvain, (PhD Thesis).
- Delfosse, P., Reddy, A.S., Legreve, A., *et al.*, 2000. Serological methods for detection of *Polymyxa graminis*, an obligate root parasite and vector of plant viruses. *Phytopathology* 90, 537–545.
- Dieryck, B., Otto, G., Doucet, D., *et al.*, 2009. Seed, soil and vegetative transmission contribute to the spread of pecluviruses in West Africa and the Indian sub-continent. *Virus Research* 141, 184–189.
- Dieryck, B., Delfosse, P., Reddy, A.S., Bragard, C., 2010. Targeting highly conserved 3'-untranslated region of pecluviruses for sensitive broad-spectrum detection and quantitation by RT-PCR and assessment of phylogenetic relationships. *Journal of Virological Methods* 169, 385–390.
- Dieryck, B., Weyns, J., Doucet, D., Bragard, C., Legreve, A., 2011. Acquisition and transmission of *Peanut clump virus* by *Polymyxa graminis* on cereal species. *Phytopathology* 101, 1149–1158.
- Pandey, V., Mandal, B., Jain, R.K., Roy, A., 2017. Indian peanut clump virus, a fungal transmitted *Pecluvirus* infecting both monocotyledonous and dicotyledonous plants in India. In: Mandal, B., Rao, G.P., Baranwal, V.K., Jain, R.K. (Eds.), *A Century of Plant Virology in India*. Singapore: Springer Nature Singapore Pvt. Ltd, pp. 351–359.
- Reddy, A.S., Hobbs, H.A., Delfosse, P., Murthy, A.K., Reddy, D.V.R., 1998. Seed transmission of Indian peanut clump virus (IPCV) in peanut and millets. *Plant Disease* 82, 343–346.
- Tamada, T., Kondo, H., 2013. Biological and genetic diversity of plasmodiophorid-transmitted viruses and their vectors. *Journal of General Plant Pathology* 79, 307–320.
- Vaïanopoulos, C., Bragard, C., Moreau, V., Maraite, H., Legrève, A., 2007. Identification and quantification of *Polymyxa graminis* f. sp. *temperata* and *P. graminis* f. sp. *tepida* on barley and wheat. *Plant Disease* 91, 857–864.

Pepino Mosaic Virus (*Alphaflexiviridae*)

Rene AA van der Vlugt, Wageningen University and Research Center, Wageningen, The Netherlands
CCMM Stijger, Wageningen University and Research Center, Bleiswijk, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa	Amino acid(s)	kDa	Kilo Dalton
CP	Coat protein or capsid protein	MET	Methyl-transferase domain
DAS-ELISA	Double antibody Sandwich - Enzyme-linked-Immuno-sorbent-Assay	nm	Nanometer
ELISA	Enzyme-linked immunological assays	nt	Nucleotide(s)
EM	Electron microscope	ORF	Open reading frame
HEL	NTPase/helicase domain	RdRp	RNA-dependent RNA polymerase
IEM	Immuno-sorbent electron microscopy	RISC	RNA-induced silencing complex
IR	Intergenic region	RT-PCR	Reverse transcriptase polymerase chain reaction
		RT	Reverse transcriptase
		TGB	Triple gene block

Glossary

Double antibody sandwich - Enzyme-Linked-Immuno-Sorbent-Assay An assay in which by the use of a virus-specific antibody the presence of a virus in plants or plant parts can be established.

Immuno-sorbent electron microscopy Identification of a virus by visualization in the electron microscope of the reaction of the virus coat protein with a specific antiserum.

Reverse transcriptase polymerase chain reaction A PCR reaction whereby first a cDNA copy of the viral RNA has to be generated by Reverse Transcriptase (RT) before PCR-amplification is possible.

Virus symptomatology Typical local or systemic symptoms on natural host plants or experimental test plants that can be used to characterize and differentiate plant viruses and plant virus strains.

Introduction

Pepino mosaic virus (PepMV) was first found in 1974 in field samples of pepino plants (*Solanum muricatum*) located near some potato fields, collected in the Canete valley in coastal Peru. These plants showed some yellow mosaic in young leaves. Electron microscope (EM) investigations showed the presence of typical filamentous potexvirus particles of approximately 500 nm, which did not react with antisera against Potato virus X (PVX). Serologically it appeared most closely related to Narcissus mosaic virus (NaMV) but the host ranges from both viruses differ considerably. It was concluded that Pepino mosaic virus was a new and distinct potexvirus species.

Since its first description in 1980 the virus was never reported and not considered to be of any agricultural significance until it was described in 1999 from commercial protected tomato crops (*S. lycopersicum*) in the UK and The Netherlands. Since then the virus has spread very rapidly through commercial tomato crops and is now reported worldwide (See Relevant Websites Section). It is particularly present in Europe and The America's. In Africa and Asia its presence appears to be restricted and it has not been reported from Australia or New Zealand.

Taxonomy and Classification

Pepino mosaic virus is a species member of the genus *Potexvirus* within the family *Alphaflexiviridae*.

Virion Properties

PepMV has typical filamentous potexvirus particles with a normal length of 510 nm. Particles are comprised of a single capsid protein (CP) of approximately 26 kDa. Ultrathin section of infected leaf material may contain inclusions consisting of arrays of filamentous virus-like particles. The virus particles are less clearly cross-banded as those of PVX.

As a member of the genus *Potexvirus*, PepMV is fairly stable. In endpoint dilution studies, sap from infected *N. glutinosa* was always infectious at dilutions of 10^{-4} , occasionally at 10^{-5} but never at 10^{-6} . Sap lost most of its infectivity after 10 min at 65°C and was no longer infectious at 70°C. Sap stored at 20°C still shows some infectivity after 3 months while leaves of *N. glutinosa* desiccated over silica gel were still infectious after 6 months. Under practical conditions in greenhouses and fields the virus may



Fig. 1 Schematic organization of the RNA genome of Pepino mosaic virus with its five open reading frames (ORFs). M7G = 5' cap, AAAAA_n = poly-A tail.

easily survive for several weeks in plant debris and on surfaces (e.g., tools, clothes, containers) that have come in contact with virus-infected leaves or fruits.

Genome Organization and Expression

The genome of PepMV resembles that of typical potexviruses. For all strains its positive single-stranded RNA is around 6410 nt long, capped at the 5'-end, polyadenylated at the 3'-end and contains 5'- and 3'-untranslated regions. It encodes five putative partly overlapping open reading frames (ORFs 1–5, see Fig. 1). The following data are based on the RNA sequence of the original EU-tomato isolate PD99901066 (FJ940223) as found in 1999 in The Netherlands. Positions for regions and motifs in other isolates and other strains may vary slightly.

The 5'-untranslated region (5'-UTR) of the virus is 86 nt long and like all other potexviruses it starts with the 5'-GAAAA pentanucleotide.

ORF1 (nt 86–4406) encodes a 164 kDa protein of 1439 aa. It contains a putative methyl-transferase domain (MET; aa 59–224) specific for the supergroup of 'Sindbis-like' viruses, a NTPase/helicase domain (HEL; aa 708–934) with the NTP-binding motifs GCGGSGKS and VVIFDD and a RNA dependent RNA polymerase (RdRp) domain (aa 1217–1374) characterized by the SGEPTFDANT-X22-GDD motif.

The stop codon of the first ORF is followed by a short intergenic region (**IR1**) of 25 nt and a set of three partially overlapping ORFs typically known as the triple gene block (TGB).

ORF2 (nt 4432–5136) encodes the first Triple Gene Block protein (TGB1), a 234 aa protein of 26 kDa. This protein contains a typical NTPase/helicase motif (aa 26–233) characterized by 7 conserved motifs, two of which may be involved in NTP binding. TGBp1 belongs to the superfamily I of RNA helicases.

ORF3 (nt 5117–5488) overlaps 19 nt with the 3'-end of ORF2 and extends 148 nt past the start codon of ORF4. It encodes the TGB2 protein, a small 14 kDa protein which contains a potexvirus specific consensus motif: PxxGDxxHxL/FPxGGxYxDGTXxxxY.

ORF4 (nt 5340–5594) encodes the third TGB protein (TGB3), a 85 aa protein of 9 kDa. This protein is the most variable among potexviruses. It contains a CxV/lxxxG consensus motif among potexvirus TGB3 proteins. Substitution of Lysine (K) for glutamic acid (E) at position 67 in TGB3 was described as being involved in increased virus accumulation and the induction of systemic necrosis both in tomato and test plants. The TGB proteins are essential for virus movement, suppression of RNA silencing and the formation of the virus factory.

The second intergenic region (**IR2**) of 38 nt (nt 5595–5632) precedes **ORF5** (nt 5633–6346) which encodes the 238 aa coat protein (CP) of 25 kDa. This CP contains the amphipathic core sequence KFAAFDFFDGV. A similar sequence is also found in the CP of other potexviruses and might be responsible for binding of virus RNA to the CP through hydrophobic interactions. The PepMV CP is also required for cell-to-cell and long-distance movement and is involved in the suppression of RNA silencing.

The 64 nt long 3'-untranslated region (3'-UTR: nt 6347–6410) precedes the poly-A tail and contains the hexamer 5'-ACUAAA sequence which is also present in the 3'-UTR of all potexviruses sequenced so far. This motif is proposed to be a cis-acting element involved in the positive and negative viral RNA synthesis. The 5'-AAUAAA polyadenylation signal terminates the RNA genome whereby the AAA portion form the first A residues of the poly(A) tail.

Host Range and Symptomatology

The host range of PepMV is relatively broad infecting plants from different families including both cultivated and wild hosts. PepMV has not been reported to infect plants from the *Cucurbitaceae* or *Fabaceae*. Most plant species belong to the *Solanaceae* of which many species are infected systemically. Its main hosts appear to be *Solanum* spp. The originally described pepino strain was found to infect wild and commercial tuber-bearing *S. tuberosum* sp. (potato), mostly with a symptomless systemic infection or with mild mosaic symptoms but in two local Peruvian *S. tuberosum* cultivars and *S. stoloniferum* PI 230557 it caused a severe systemic necrosis. The virus was shown to be transmitted through the tubers.

Its best-known natural hosts are pepino (*S. muricatum*) and cultivated tomato (*S. lycopersicum*) but surveys showed infection with PepMV of several related wild *Solanum* spp. These include *S. peruvianum*, *S. parviflorum*, *S. chilense*, *S. chmielewskii*, and *S. pimpinellifolium*. Infections in these wild species were generally symptomless.

PepMV was also reported to infect other solanaceous crop and test plants; *Solanum melongena* (eggplant), *S. nigrum*, *S. tuberosum* (potato), *Datura stramonium*, *D. metel*, *D. innoxia*, *Nicotiana bethaminana*, *N. occidentalis*, *N. glutinosa*, and *N. tabacum* cv Xanthi. Sweet

pepper (*Capsicum annuum*) does not seem a suitable host since most PepMV strains were not able to induce a systemic infection, with the exception of the US1 strain which could induced systemic necrotic lesions.

Several weed hosts from various plant families (e.g., Convolvulaceae, Brassicaceae, Boraginaceae, Asteraceae, Plantaginaceae and Polygonaceae) were found to be hosts for PepMV. The original virus description reports weed hosts in Peru including *Datura stramonium*, *Nicandra physaloides* and *Physalis peruvianum*. Studies on sampled material from mainland Spain and the Canary Islands show a large number of weeds species latently infected (i.e., showing no symptoms)).

The initially described PepMV strain caused a distinct yellow mosaic on young leaves of pepino and most infected plant also showed dark green enations on the lower surface of some leaves. In *Solanum* spp. it caused a symptom less systemic infection as shown by back-inoculation on a sensitive indicator plants like *N. glutinosa*. Only *D. metel*, *D. stramonium*, and a number of *Nicotiana* sp. showed distinct symptoms upon systemic infection.

In contrast to the original Pepino isolate the EU-tomato virus isolate found in 1999 in commercial tomato crops in the UK and The Netherlands did cause distinct although generally mild symptoms in commercial tomato crops. In general, PepMV symptoms depend on the particular virus isolate in combination with the tomato cultivar, the age of the plant when first infected and environmental conditions. Symptoms may include small yellow leaf spots, yellow-green mosaic, mottle on the older leaves and/or slight curling in the top leaves (nettle-head symptom) or a greyish appearance of the top of the plant. Fruit symptoms may range from discoloration in the form of blotchy ripening or flaming, fruit marbling, mild sometimes concentric yellow/orange mottling and uneven ripening to netting and cracking ('open fruits') and shape distortion.

Basically, severity of symptoms in tomato may vary with the particular virus isolate involved. Both mild and severe symptoms inducing PepMV isolates have been described and there appears to be no correlation between symptom severity in tomato and the PepMV virus strain. Mild symptoms may vary from small yellow spots on leaves to a slight reduction in fruit production. Certain isolates can cause much more pronounced plant symptoms: more severe leaf chlorosis, bright yellow mosaic, leaf distortions ('bubbling'), green striation on the stem and sepals up to more or less pronounced leaf and stem necrosis. Because of fruit symptoms, fruit quality can be affected which may result in economic damage. Symptom expression is reported to depend on environmental conditions with low light conditions and low temperatures favouring more pronounced symptoms.

Inoculation studies of three PepMV strains (EU, CH2, and US1) clearly showed a number of differences between these strains on tomato and other solanaceous crops and test plants. Tomato is susceptible to all three strains, while sweet pepper (*C. annuum*) is generally not a systemically susceptible host for the three strains. Eggplant is susceptible to all three strains however local symptoms are seldom seen and systemic leaf symptoms may depend on locality and virus isolate. In potato systemic infections may develop in a sensitive cultivar but local and systemic symptoms are seldom seen. The virus however can infect the tubers and can thus be transmitted.

In test plants *N. occidentalis* 37B proved a useful susceptible host which is able to differentiate between strains based on typical local and systemic symptoms.

Virus Isolates and Strains

The first PepMV isolate was isolated in 1974 from pepino crops in Peru and only in 1999 an isolate was described on commercial tomato. Comparison of this tomato isolate with the original Peruvian pepino isolate (#BBA1137 in the DSMZ plant virus collection in Germany) showed some differences in a number of characteristics. These were most pronounced in the reactions of both isolates on tomato (*S. lycopersicum*) in which the pepino isolate is symptomless and the particular isolate of the tomato strain caused mild but distinct symptoms in the form of small sharp yellow flecks on leaves. Also on a number of test plants the two isolates can be distinguished, most clearly on *N. glutinosa* and *D. stramonium*.

Since 1999 a large number of different PepMV isolates have been described. These were either isolated from commercial tomato crops or tomato seed lots in different countries or from wild *Solanum* sp. Most show the typical characteristics of the tomato strain but some deviant isolates were observed. A number of isolates show more severe symptoms, (including leaf and stem necrosis and sever fruit symptoms) in tomato or wild *Solanum* sp. There is considerable biological variation between isolates and severity of symptoms is not related to the virus strain and in most strains both mild and severe isolates have been described.

Currently available biological data (i.e., host and indicator plant symptoms), serological relationships and multiple nucleotide sequence alignments (Fig. 2) suggest at least five clusters of relationship between the PepMV isolates which are now considered as separate genotypes or strains:

- (1) The Peruvian pepino strain (LP),
- (2) The European tomato strain (EU),
- (3) The Chile-2 strain (CH2),
- (4) The US1 strain (US1),
- (5) The PES strain, identified from wild *Solanum* species in Peru.

A significant number of full genome sequences are now available for PepMV with most sequences derived EU and CH2 isolates. Only a limited number of sequences from the LP, US1, and PES strains are published. Sequence identity between isolates of the same strain is high (i.e., above 98%) and several studies have shown that single strain populations of PepMV shown very low genetic variation over time. Direct full genome nucleotide sequence comparisons between strains showed that LP and EU strains

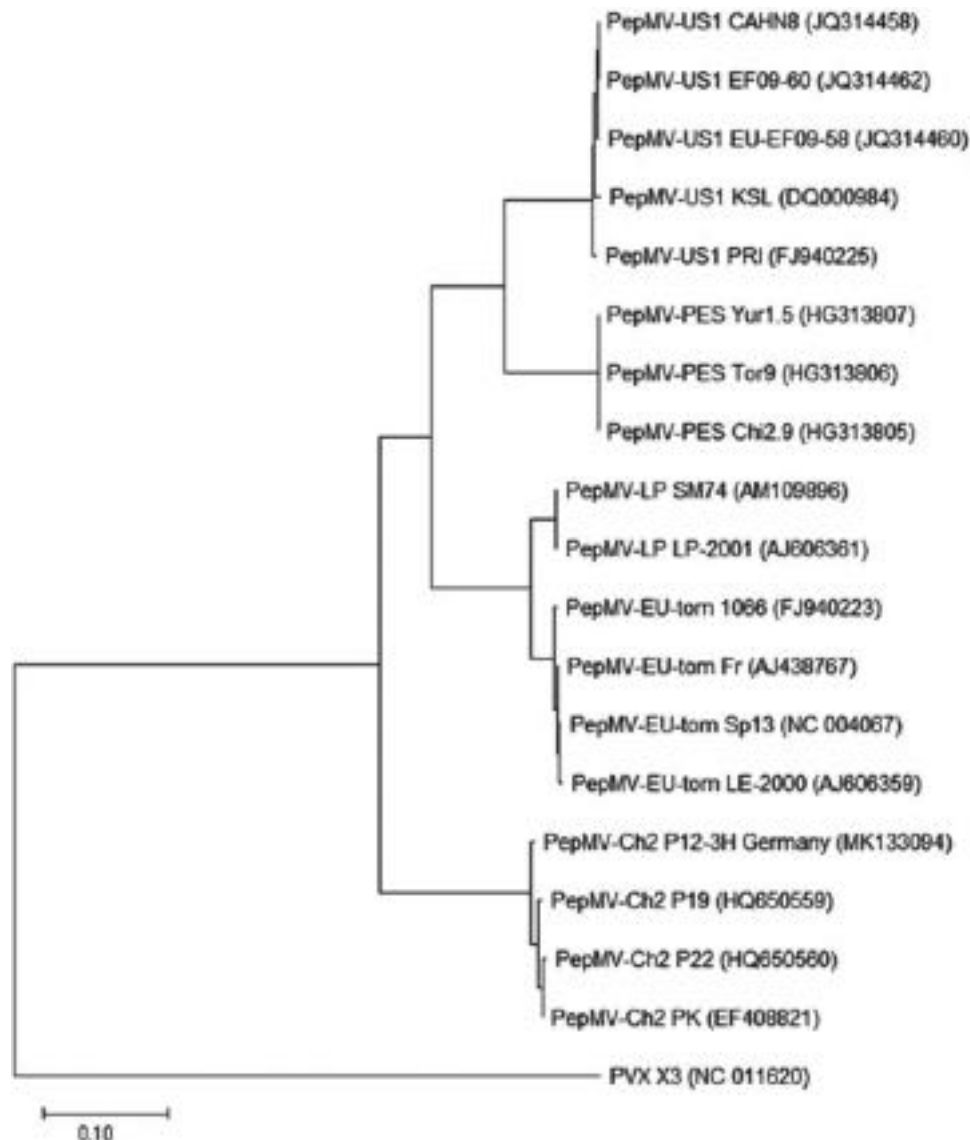


Fig. 2 Phylogenetic Maximum Likelihood tree of full-length RNA sequences of selected Pepino mosaic virus isolates of the five main strains. LP = Peruvian strain, EU = European tomato strain, CH2 = Chile-2 strain, US1 = US1 strain, PES = wild *Solanum* strain.

are most closely related with around 95% sequence identity. The US1 and PES strains share around 86% identity to each other and around 81% identity to the LP and EU strain. The CH2 strain is the most deviant as it generally shares around 78% sequence identity with the four other strains [Fig. 3](#).

Virus Transmission and Epidemiology

PepMV clearly originates from South America. Surveys in Peru showed that the virus was widespread, even in isolated wild *Solanum* sp., and phylogenetic analyses suggested that the virus most likely originated from a wild *Solanum* sp. Interestingly most infected *Solanum* sp. showed no distinct symptoms. Several studies have confirmed that the original Peruvian pepino isolate (LP) is characterized by the absence of, or only very mild symptoms on commercial tomato crops. The EU-strain isolate first found in 1999 in Europe also induces only mild symptoms in commercial tomato crops. This likely contributed to the initial rapid spread of the virus within and outside Europe.

Initial studies between 2000 and 2005 employing comparisons of viral gene sequences revealed that the tomato strain was prevalent in all countries reporting the virus. All tomato isolates found were all highly similar and could not be grouped according to geographic origin, or symptoms on naturally or experimentally infected plants. These data indicated a fairly recent introduction from limited sources and expansion of the tomato strain into commercial tomato crops, as the most likely cause for the virus epidemic.

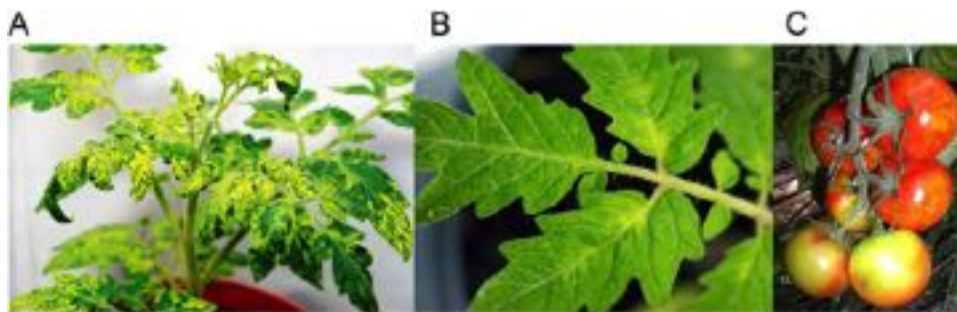


Fig. 3 Symptoms of Pepino mosaic virus on tomato. A. Severe symptom on tomato plant. B. Detail of mild leaf symptom on tomato leaf. C. Severe symptom on tomato fruits. *Source:* Courtesy of R. van der Vlugt.

The first known occurrence of PepMV in North America can be traced back to 2000. Three strains were identified from pooled symptomatic leaf material collected in Arizona; US1, US2, and US3. US1 and US2 clearly differed from the EU and LP strains while US3 was highly homologous to the EU strain. Later studies revealed that US1 is indeed a distinct strain while US2 was a recombinant sequence between US1 and another strain now known as Chile-2 (CH2). US1 and CH2 were later confirmed as distinct strains on tomato seed originating from Chile.

In 2005 the CH2 strain was first found in commercial tomato crops in Europe and soon after that it started to spread rapidly, apparently because it seems to have a biological advantage over the EU strain in tomato.

Detailed population studies from Spain on material collected between 1998 and 2004 confirmed the prevalence and homogeneity of the tomato strain isolates as well as their presence since 1998. However, there were clear indications for several independent introductions of this virus strain, both on the Canary Islands and the Spanish Main land. Interestingly in addition to the tomato strain, also the original LP strain and the CH2 strain were found, but always in mixed infections with the tomato strain. Both strains were shown to be present in Spain since 2000.

Since 2005 the CH2 strain spread rapidly and is now the most prevalent strain in most commercial tomato crops where the virus is found however, it did not replace the EU tomato strain. Isolates for both strains can occur in mixed infections the same plant. The US1 strain is found only occasionally.

PepMV is a mechanically transmitted virus. The most important transmission routes are through contaminated tools, clothes, and surfaces. The virus is relatively stable at room temperature and can survive and stay infectious for several weeks in plant debris and on contaminated surfaces. Fruits from infected plants can contain high concentrations of virus and do not necessarily show symptoms. Long distance transmission of the virus in this way is relatively likely. Implementation of strict hygiene protocols during the growing season and thorough cleaning of greenhouses at the end of the growing season have shown to effectively control the introduction and spread of the virus.

PepMV is also transmitted by tomato seed at low levels whereby the virus is likely located on the seed coat. The virus was also shown to be efficiently transmitted between tomato plants in closed recirculating hydroponic systems as well as by bumblebees. The latter two transmission routes are likely to contribute to the spread of the virus in commercial closed greenhouse production systems whereby bumblebees are used for pollination.

The possible involvement of *Olpidium virulentus* in the transmission of PepMV has been reported as well as low levels of transmission from tomato to tomato under experimental conditions by the greenhouse whitefly (*Trialeurodes vaporariorum*) and the aphids *Myzus persicae*. Transmission of PepMV by these vectors remains to be confirmed.

Diagnostic

A polyclonal antiserum raised against the original pepino isolate shows no reaction with Potato virus X (PVX) and Potato aucuba mosaic virus (PAMV), the only two other potexviruses known to infect tomato. Both in immuno-sorbent electron microscopy (IEM) and in ELISA it clearly reacted with different tomato isolates. Comparisons of polyclonal antisera raised against the pepino and tomato strains showed differences in heterologous titres between them, but the antisera show clear reactions with isolates from all known PepMV strains.

Virus Damage and Control

Initially PepMV infections in commercial tomato crops were reported as relatively insignificant with no apparent plant or fruit symptoms and no or very limited yield reduction. Studies from the UK however, reported significant effects on fruit quality. Smaller sized fruits with different grades of uneven ripening and discoloration and occasionally mis-shaped fruits, led to production unsuited for the fresh UK market and hence economic damage.

Studies now clearly show that symptoms induction by PepMV is highly dependent on environmental conditions and on the particular isolate involved. In Spain, PepMV infected tomato plants showed only symptoms from autumn through spring and symptoms disappeared with higher temperatures in late spring. Similar effects of high light conditions and high temperatures have been reported from other countries. However, symptoms of PepMV infected plants can be highly variable, ranging from very mild leaf symptoms to severe leaf and stem necrosis. Fruit symptoms may vary from mild discolorations to open fruits showing severe cracking. Mild, chlorotic and even necrotic PepMV isolates have been reported and the capacity to induce different symptoms is a property of the isolate. There are indications that severity of symptoms may be associated with increased virus accumulation. Distinct single nucleotide mutations resulting in specific amino acid mutations in the viral CP and TGB3 genes were shown to be involved in the development of yellowing and necrotic symptoms, respectively. Also the highly conserved 'palm' domain in the Polymerase domain of the viral RdRp encoded by ORF1 of plays a role in the induction of plant necrosis which is apparently modulated by the interplay between the TGB3 protein, plant genotype and environmental conditions.

Currently no resistance genes to PepMV are available in commercial tomato cultivars. Extensive screening in wild *Solanum* sp. has been identified only a very limited number of potential resistance sources. Some accessions of *S. lycopersicoides* and *S. chilense* were resistant to isolates from the EU strain. Whether this resistance holds against the other PepMV strains remains to be investigated.

One method currently applied in various countries to control negative effects of PepMV infections is cross-protection i.e., the use of attenuated virus isolates. A mild isolate is inoculated onto plants in an early growth stage and the resulting systemic infection protects the plant against more severe isolates of the virus. Cross-protection is based on the sequence specific activation of the RNA-induced silencing complex (RISC). This general antiviral defense mechanism is sequence homology-dependent and is therefore not virus isolate- but strain-specific. Protection against severe EU-strain isolates requires an attenuated EU isolate while an attenuated CH2 isolate is required to protect against aggressive CH2 isolates. In many instances isolates of the CH2 and EU strains occur in mixed infections so for adequate protection mild isolates of both strains are required.

PepMV is a mechanically transmitted virus. The most important transmission routes are through contaminated tools, clothes and surfaces. The virus is relatively stable at room temperature and can survive and stay infectious for several weeks in plant debris and on contaminated surfaces. Fruits from infected plants can contain high concentrations of virus and do not necessarily show symptoms. Long distance transmission of the virus in this way is relatively likely. Implementation of strict hygiene protocols during the growing season and thorough cleaning of greenhouses at the end of the growing season have shown to effectively control the introduction and spread of the virus.

Effective control of the virus in commercial tomato culture should come from the use of virus-free seeds and planting material. For this reliable and sensitive methods, both serology and molecular based, for the detection of the virus in seed, plants and fruits are available. Since no adequate plant resistance is available yet strict hygiene measures and a constant monitoring for possible infections are required to avoid unwanted spread of the virus.

See also: Alphaflexiviruses (*Alphaflexiviridae*). Potexviruses (*Alphaflexiviridae*)

Further Reading

- Agüero, J., Gómez-Aix, C., Sempere, R.N., *et al.*, 2018. Stable and broad spectrum cross-protection against Pepino mosaic virus attained by mixed infection. *Frontiers in Plant Sciences* 9, 1810. doi:10.3389/fpls.2018.01810.
- Blystad, D.-R., van der Vlugt, R., Alfaro-Fernández, A., *et al.*, 2015. Host range and symptomatology of Pepino mosaic virus strains occurring in Europe. *European Journal of Plant Pathology* 143, 43–56.
- Hanssen, I., Thomma, B., 2010. Pepino mosaic virus: a successful pathogen that rapidly evolved from emerging to endemic in tomato crops. *Molecular Plant Pathology* 11, 179–189. doi:10.1111/j.1364-3703.2009.00600.x.
- Jones, R.A.C., Koenig, R., Lesemann, D.-E., 1980. Pepino mosaic virus, a new potexvirus from pepino (*Solanum muricatum*). *Annals of Applied Biology* 94, 61.
- Moreno-Pérez, M.-G., Pagán, I., Aragón-Caballero, L., *et al.*, 2014. Ecological and genetic determinants of Pepino mosaic virus emergence. *Journal of Virology* 88 (6), 3359.
- Van der Vlugt, R.A.A., Cuperus, C., Vink, J., *et al.*, 2002. Identification and characterization of Pepino mosaic virus in tomato. *EPPO Bulletin* 32, 503.

Relevant Websites

- <https://gd.eppo.int/taxon/PEPMV0/documents>
EPPO Global Database
Pepino mosaic virus.
- <https://gd.eppo.int/taxon/PEPMV0/distribution>
EPPO Global Database
Pepino mosaic virus
Distribution.

Plant Reoviruses (*Reoviridae*)

Yu Huang and Yi Li, Peking University, Beijing, China

© 2021 Elsevier Ltd. All rights reserved.

This article is an update of R.J. Geijskes, R.M. Harding, Plant Reoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00468-4](https://doi.org/10.1016/B978-012374410-4.00468-4).

Nomenclature

aa Amino acid(s)

cDNA Complementary deoxyribonucleic acid

dsDNA Double-stranded deoxyribonucleic acid

dsRNA Double-stranded ribonucleic acid

ELISA Enzyme-linked immuno-sorbent assay

mRNA Messenger RNA

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

RT-LAMP Reverse transcription loop-mediated isothermal amplification

RT-PCR Reverse transcription polymerase chain reaction

VSR Viral suppressor of RNA silencing

Glossary

Capsid The protein shell that surrounds a virus particle.

Cell-to-cell movement protein Virally encoded protein that helps viral virions or genome RNA (s)/protein complexes to move from cell to cell.

Dicotyledon A flowering plant that has two cotyledons in the seed.

Icosahedron Having 20 equal sides or faces.

Monocotyledon A flowering plant that has only one cotyledon, or seed leaf, in the seed.

RNA-dependent RNA polymerase Virally encoded RNA-dependent RNA polymerase that replicates viral genome(s).

Transovarial transmission Transmission from one generation to another through eggs.

Viral suppressor of RNA silencing Virally encoded protein(s) that suppress(es) host RNA silencing pathways.

Viroplasm Cytoplasmic, non-membranous structure in which plant reoviruses are thought to replicate and assemble.

Introduction

The family *Reoviridae* comprises a diverse group of viruses that can infect vertebrates, invertebrates, and plants. Despite their large host range, all members of the family *Reoviridae* share common properties, including an icosahedral-shaped virion and a segmented double-stranded RNA (dsRNA) genome. The family *Reoviridae* consists of 15 genera divided between two subfamilies, *Spinareovirinae* and *Sedoreovirinae*. Plant-infecting reoviruses are members of the genera *Oryzavirus* and *Fijivirus*, which belong to the *Spinareovirinae* subfamily, and the genus *Phytoreovirus*, part of the *Sedoreovirinae* subfamily. These reoviruses generally replicate both in their plant hosts and in their insect vectors. Infection of the insect vector is non-cytopathic and often persists throughout the life of the insect. Infection of the host plant is tissue specific (except for Rice dwarf virus [RDV]) and can cause severe disease. Fiji leaf gall disease, caused by Fiji disease virus (FDV), has caused yield losses of up to 90% in susceptible varieties of sugarcane in Australia. Rice ragged stunt virus (RRSV) is reported to reduce yield of rice by up to 100% in severe infections (although reductions of 10%–20% are more common). Rice common dwarf disease, caused by RDV, can also cause significant losses as infected plants often fail to set seed. Southern rice black-streaked dwarf disease, caused by Southern rice black-streaked dwarf virus (SRBSDV), was first discovered in Yangxi County, Guangdong Province, China, in 2001, and its distribution was limited to southern China until the end of 2008. By 2010, however, it had spread to 29 provinces of Vietnam and 13 provinces of China, and it has also now been found in some areas in Japan. The rapid spread of SRBSDV has caused serious rice losses in East and Southeast Asia.

Taxonomy and Classification

Currently, three genera of the family *Reoviridae* are classified as plant-infecting reoviruses: *Fijivirus*, *Oryzavirus*, and *Phytoreovirus*. The genera of plant-infecting reoviruses are differentiated according to the number of genomic dsRNA segments and their electrophoretic profile, hosts, serological relationships, and capsid morphology (Table 1). These reoviruses replicate both in plant hosts (except for one *fijivirus*: *Nilaparvata lugens reovirus*) and in their insect vectors (Table 1). Infection of the host plant is species specific, although the host range can often be extended under experimental conditions and can produce a variety of symptoms, including severe disease. None genome sequence information is now available for PaSV, GDV and ERSV. Partial genome sequence information is now available for WTV and OSDV. Complete genome sequence information have been obtained for all the other plant reoviruses. This has allowed detailed comparisons to be made within these genera and across all of the family *Reoviridae*,

Table 1 Characteristics of plant reoviruses

Subfamily, Genus Species name	Abbr.	dsRNA segments	Host	Vector (s)
Spinareovirinae Fijivirus				
Fiji disease virus	(FDV)	10	Gramineae	Planthoppers: <i>Perkinsiella saccharicida</i> , <i>P. vastatrix</i> , <i>P. vitiensis</i>
Rice black-streaked dwarf virus	(RBSDV)	10	Gramineae	Planthoppers: <i>Laodelphax striatellus</i> , <i>Ribautodelphax alba</i> fascia, <i>Unkanodes sapporona</i>
Maize rough dwarf virus	(MRDV)	10	Gramineae	Planthopper: <i>Ribautodelphax notabilis</i>
Pangola stunt virus	(PaSV)	10	Gramineae	Planthoppers: <i>Sogatella furcifera</i> , <i>S. kolophon</i>
Mal del Rio Cuarto virus	(MRCV)	10	Gramineae	Planthopper: <i>Delphacodes kuscheli</i>
Southern rice black-streaked dwarf virus	(SRBSDV)	10	Gramineae	Planthopper: <i>Sogatella furcifera</i>
Oat sterile dwarf virus	(OSDV)	10	Gramineae	Planthoppers: <i>Javesella pellucida</i> , <i>J. discolour</i> , <i>J. dubia</i> , <i>J. obscurella</i> , <i>Dicranotropis hamata</i>
Garlic dwarf virus	(GDV)	10	Liliaceae	Planthopper: Unknown
Nilaparvata lugens reovirus	(NLRV)	10	No plant host reported	Planthoppers: <i>Nilaparvata lugens</i> , <i>Laodelphax striatellus</i>
Spinareovirinae Oryzavirus				
Rice ragged stunt virus	(RRSV)	10	Gramineae	Planthopper: <i>Nilaparvata lugens</i>
Echinochloa ragged stunt virus	(ERSV)	10	Gramineae	Planthoppers: <i>Sogatella longifurcifera</i> and <i>S. vibix</i>
Spinareovirinae Phytoreovirus				
Wound tumor virus	(WTV)	12	Dicot	Leafhoppers: <i>Agallia constricta</i> , <i>A. quadripunctata</i> , <i>Agalliopsis novella</i>
Rice dwarf virus	(RDV)	12	Gramineae	Leafhoppers: <i>Nephotettix cincticeps</i> , <i>N. nigropictus</i> , <i>Recillia dorsalis</i>
Rice gall dwarf virus	(RGDV)	12	Gramineae	Leafhoppers: <i>Nephotettix cincticeps</i> , <i>N. nigropictus</i> , <i>N. virescens</i> , <i>N. malayanus</i> , <i>Recillia dorsalis</i>

providing a basis for their classification into species and genera. One general feature is the presence of inverted repeats several bases long adjacent to the conserved terminal sequences in all plant-infecting reoviruses; the sequences of the repeats differ among the three genera.

Within the genus *Fijivirus*, individual species have considerable similarities to FDV, the type species. Their classification into separate species is based on unique characteristics, such as capacity to exchange genome segments, relatively high aa sequence similarity, serological cross-reaction, cross-hybridization of RNA or cDNA probes, host species, and insect vector species. Analysis of the available genome sequences has further assisted the identification of *Fijivirus* species. The gross genome characteristics of fijiviruses includes a genome size of approximately 29 kbp and a characteristically low G + C content of 34%–36%. The 3' terminal trinucleotide in the positive strand is unique and highly conserved in all RNA segments among all species within the genus *Fijivirus* (Table 2). Members of the genera *Oryzavirus* and *Phytoreovirus* have significant similarity to their type species, RRSV and *Wound tumor virus* (WTV), respectively. Demarcation of species within the oryzaviruses and phytoreoviruses is primarily based on the ability to exchange genome segments although other characteristics, as mentioned for fijiviruses above, are also used. When available, genomic sequences are examined to reveal distinguishing features to support the classifications. Oryzaviruses have a total genome size of approximately 26 kbp and specific 5'- and 3' terminal sequences in all RNA segments, resulting in the canonical structure 5'GAUAAA–GUGC3'. Phytoreoviruses have a total genome size of approximately 25 kbp, a G + C content between 38% and 48%, and specific 5'- and 3'-terminal sequences in all RNA segments [canonical structure 5'GG(U/C)A–(U/C)GAU3'].

Virion Structure and Genome Organization

Fijivirus

Fijivirus virions have a complex double icosahedral capsid construction and consist of a capsid, a core, and a nucleoprotein complex. Virions are fragile structures and readily break down *in vitro* to give cores unless pretreated with fixative. The outer capsid is 65–70 nm in diameter with 12 “A”-type spikes located at the vertices of the icosahedron. The inner core is about 55 nm in diameter, with 12 “B”-type spikes located at the vertices. The viral nucleic acid is located at the center of the virus particle, within the inner core capsid (Fig. 1(A)). Each virion contains a single full-length copy of the genome. Fijivirus genomes contain 10 dsRNA genomic segments varying from approximately 1.8 to 4.5 kbp (Fig. 2). The total genome is approximately 29 kbp, with a low G + C content of 34%–36%. Within the genus, the only conserved sequence is found at the 3'-terminus (canonical structure ...GUC-3'; Table 2). Segment-specific inverted repeats are found adjacent to these terminal sequences. Segments 1–6, 8, and 10 are monocistronic, containing one open reading frame (ORF), while segments 7 and 9 each contain two ORFs, although the expression of the second ORF has not been certified *in vivo* in either plant cells or insect vectors. NLRV is the only fijivirus identified to date that differs from this standard structure, with only one ORF on segment 10, which is equivalent to segment

Table 2 Conserved terminal sequences (positive strand) in fijiviruses

Virus	5' conserved sequence	3' conserved sequence ^a
Fiji disease virus	5'-AAGUUUUU-3'	5'-CAGCNNNNGUC ^{3'}
Rice black-streaked dwarf virus	5'-AAGUUUUU-3'	5'-AGCUNN(C/U)GUC ^{3'}
Maize rough dwarf virus	5'-AAGUUUUUU-3'	5'-UGUC ^{3'}
Mal del Rio Cuarto virus	5'-AAGUUUUU-3'	5'-CAGCUNNNNGUC ^{3'}
Oat sterile dwarf virus	5'-AACGAAAAAA-3'	5'-UUUUUUUUUAGUC ^{3'}
Nilaparvata lugens reovirus	5'-AGU-3'	5'-GUUGUC ^{3'}
Southern rice black-streaked dwarf virus	5'-AAGUUUUU-3'	5'-CAGCUGAUGUC ^{3'}

^aItalicized trinucleotides are conserved in all fijivirus sequences reported to date.

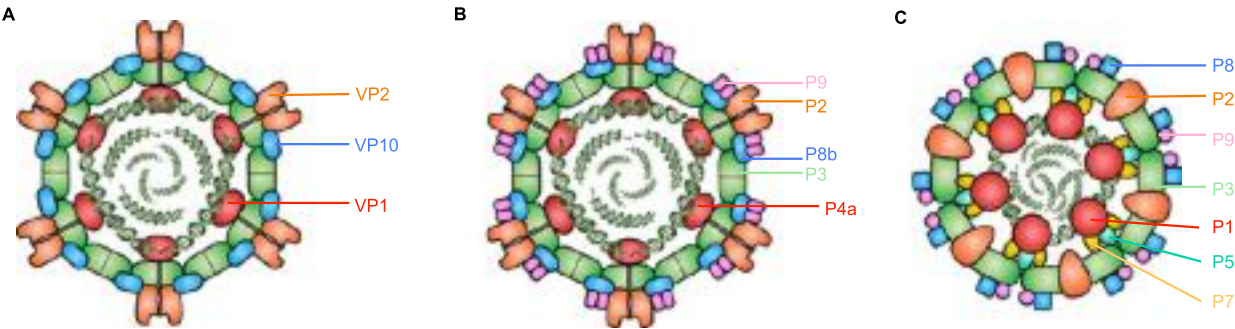


Fig. 1 Diagrammatic representation of the particle structures of Fiji disease virus (A), Rice ragged stunt virus (B), and Rice dwarf virus (C).

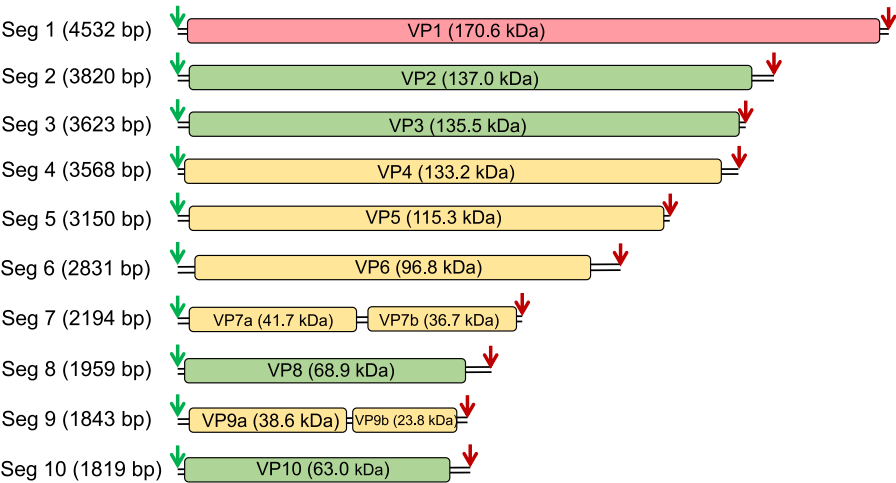


Fig. 2 Genome organization of Fiji disease virus (FDV) containing 10 dsRNA segments. Each segment contains one ORF except for Seg 7 and Seg 9 which contain two ORFs. The green arrows indicate the conserved 5' terminal sequence (5'-AAGUUUUU-3') while the red arrows indicate the conserved 3' terminal sequence (5'-CAGCNNNNGUC^{3'}). The red segment encodes the RdRp. The green segments encode the structural proteins. The yellow segments encode non-structural proteins or proteins with unknown function.

7 of FDV. The functions of the proteins encoded by some of the ORFs are still unconfirmed; the functions of segments 1–3 and 7–10 have been predicted based on protein expression studies or deduced from those of the equivalent segments of other virus species within the genus (Table 3).

Oryzavirus

Oryzavirus virions have a double-shelled icosahedral capsid and consist of an outer capsid, an inner capsid, and a core. Virions are fragile and readily break down *in vitro* to give subviral core particles unless pretreated with fixative. The outer capsid is 75–80 nm in diameter with 12 “A”-type spikes located at the 5-fold axis of the icosahedron. The core capsid is about 57–65 nm in diameter,

Table 3 Genome organization of FDV and predicted gene function

Segment	Size (bp)	Protein nomenclature	Predicted MW (kDa)	Predicted function (location)
S1	4532	VP1	170.6	RdRp (core)
S2	3820	VP2	137.0	Major core (core)
S3	3623	VP3	135.5	Possible B spike (outer capsid)
S4	3568	VP4	133.2	Unknown
S5	3150	VP5	115.3	Unknown
S6	2831	VP6	96.8	Unknown
S7	2194	VP7a	41.7	Possible tubule protein (nonstructural)
		VP7b	36.7	Unknown
S8	1959	VP8	68.9	Possible NTP-binding (core)
S9	1843	VP9a	38.6	Structural protein (viroplasm)
		VP9b	23.8	Unknown
S10	1819	VP10	63.0	Outer capsid protein (major outer capsid)

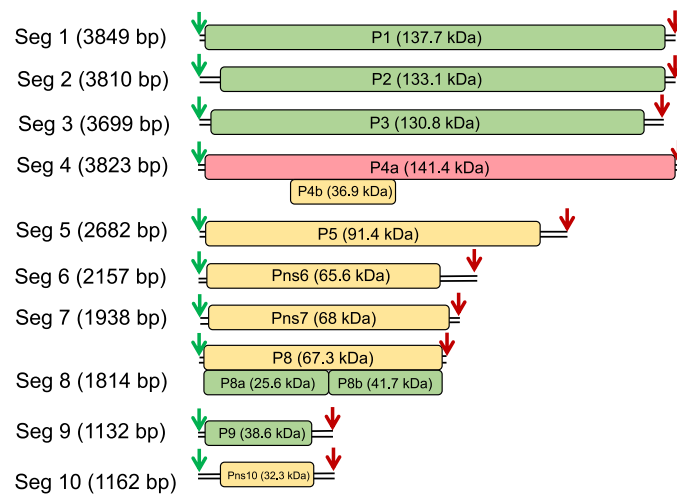


Fig. 3 Genome organization of Rice ragged stunt virus (RRSV) containing 10 dsRNA segments. Each segment contains one ORF except for Seg 4 which contains two ORFs. The green arrows indicate the conserved 5' terminal sequence ($5'$ GAUAAA- $3'$) while the red arrows indicate the conserved 3' terminal sequence ($5'$ -GUGC $3'$). The red segment encodes the RdRp. The green segments encode the structural proteins. The yellow segments encode non-structural proteins or proteins with unknown function.

with 12 “B”-type spikes. The viral nucleic acid is located at the center of the viral particle, within the core capsid (**Fig. 1(B)**). The virus genome consists of 10 linear dsRNA segments ranging in size from 1162 to 3849 bp (RRSV) with a total length of 26 kbp (**Fig. 3**). Genome segments 1–3 and 5–10 of RRSV each contain a single ORF, while segment 4 contains two ORFs. Segment 8 encodes a polyprotein that is autocatalytically cleaved into at least two polypeptides. **Table 4** summarizes the organization of the RRSV dsRNA segments and the predicted function of the encoded proteins.

Phytoreovirus

Phytoreovirus virions have a double-shelled icosahedral capsid construction and consist of an outer capsid, a core capsid, and a smooth core. Virions are approximately 70 nm in diameter, with 12 spikes located at the 5-fold vertices of the icosahedron, and generally remain intact when purified. WTV, the type member, has three protein shells: an outer amorphous layer made up of two proteins, an inner capsid made up of two proteins, and a smooth core made up of three proteins that is about 50 nm in diameter (**Fig. 1(C)**). Each virion contains a single full-length copy of the genome. Phytoreoviruses have 12 segments of linear dsRNA ranging in size from approximately 1 to 4.5 kbp, with a total genome length of approximately 25 kbp (**Fig. 4**) and a G + C content of 41%–48%. Each segment of RDV contains a single ORF except for segments 11 and 12, which contains two ORFs. **Table 5** summarizes the organization of the RDV dsRNA segments and the putative function of the encoded proteins.

Table 4 Genome organization of RRSV and predicted gene function

Segment	Size (bp)	Protein nomenclature	Predicted MW (kDa)	Apparent MW (kDa)	Predicted function (location)
S1	3849	P1	137.7	137	Virion core associated (B spike)
S2	3810	P2	133.1	118	Inner core capsid, putative guanylyl-transferase (core capsid)
S3	3699	P3	130.8	130	Major core capsid (core capsid)
S4	3823	P4a	141.4	145	RdRp
		P4b	36.9		Unknown
S5	2682	P5	91.4	90	Capping enzyme
S6	2157	Pns6	65.6		Non-structural protein, VSR, cell-to-cell movement
S7	1938	Pns7	68	66	Non-structural protein, NTP-binding protein (unknown)
S8	1814	P8	67.3	67	Precursor polyprotein/protease
		P8a	25.6	47	Spike
		P8b	41.7	44	Major capsid protein
S9	1132	P9	38.6	37	Vector transmission, mild silencing suppressor activity (spike)
S10	1162	Pns10	32.3	32	Non-structural protein associating with a mitochondrial membrane component, oligomycin-sensitivity conferral protein (OSCP), to facilitate virus propagation in the vector

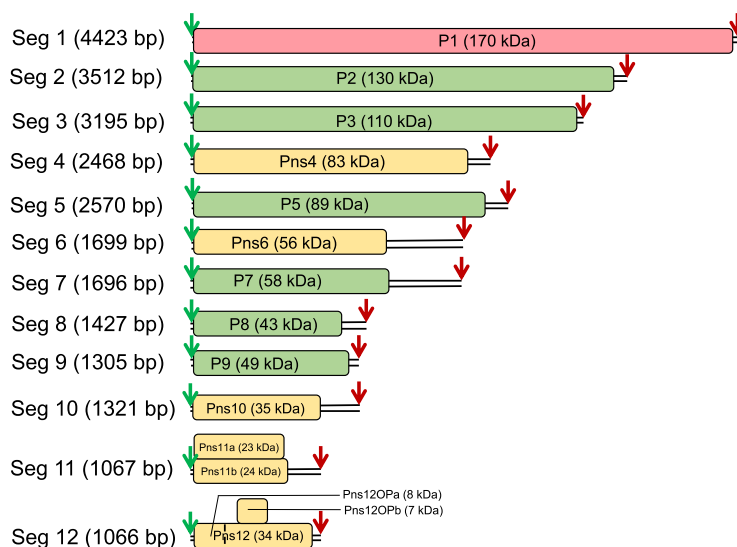


Fig. 4 Genome organization of Rice dwarf virus (RDV) containing 12 dsRNA segments. Each segment contains one ORF except for Seg 11 and Seg 12 which contain two ORFs. The green arrows indicate the conserved 5' terminal sequence ($5'$ GG(U/C)A- $3'$) while the red arrows indicate the conserved 3' terminal sequence ($3'$ -(U/C)GAU- $5'$). The red segment encodes the RdRp. The green segments encode the structural proteins while the yellow segments encode non-structural proteins.

Replication and Gene Expression

The replication and gene expression of plant-infecting reoviruses is thought to be similar to that of other reoviruses. The best described of these is *Bluetongue virus* (BTV), the type member of the genus *Orbivirus*. If the BTV model is accurate for the plant-infecting reoviruses, replication occurs after virions (or viral cores) are delivered into the host cell. Replication is initiated when the viral capsid layer is removed and the transcriptionally active core enters the cytoplasm. The viral genome (10–12 segments) remains packaged in the central cavity of the viral core to ensure that host cell defense responses to dsRNA are not activated. The core is biochemically active with an RNA-dependent RNA polymerase (RdRp), a capping enzyme, and a helicase. The viral core contains a number of channels, the largest of which is at the 5-fold axis of the icosahedral structure. Smaller channels allow the entry of nucleotides, which are required for transcription, into the core. The large channel is located adjacent to the replicase/transcriptase complex, which through its helicase activity executes the unwinding and rewinding of the dsRNA genome during transcription of the negative (antisense) RNA strand. The newly formed positive-strand (sense) mRNA molecules are modified to form a Cap 1 structure [7mGpppG^(2-Om)...] by the guanylyl-transferase, nucleotide phospho-hydrolase, and transmethylase

Table 5 Genome organization of RDV and gene function

Segment	Size (bp)	Protein nomenclature	MW ^a	Function (location)
S1	4423	P1	170	RdRp (core)
S2	3512	P2	130	Capsid structural protein, essential for vector transmission and symptom formation in plant host (outer capsid)
S3	3195	P3	110	Major core protein (core capsid)
S4	2468	Pns4	83	Phosphorylated nonstructural protein, associated with large cytoplasmic fibrils to form novel mini-tubules in infected cultured leafhopper cells, essential for viral infection in its insect vector (viroplasm, at the early stages of infection; amorphous or fibrillar inclusions, at the later stages of infection)
S5	2570	P5	89	Guanylyltransferase (core)
S6	1699	Pns6	56	Nonstructural protein, viroplasm matrix protein in insect vector cells, involved in cell-to-cell movement and nuclear acid binding in plant host
S7	1696	P7	58	RNA-binding protein (core)
S8	1427	P8	43	Major outer capsid protein
S9	1305	P9	49	Structural protein (outer capsid)
S10	1321	Pns10	35	Nonstructural protein, VSR in plant host, involved in intercellular transport in insect vector cells
S11	1067	Pns11a	23	Nonstructural protein, viroplasm matrix protein, involved in nuclear acid binding protein
		Pns11b	24	Non-structural protein, enhancing the enzymatic activity of rice S-adenosyl-L-methionine synthetase, triggering ethylene production in rice in order to achieve efficient infection
S12	1066	Pns12	34	Nonstructural protein, essential for viroplasm formation, viral replication, and infection in insect vector
		Pns12OPa	8	
		Pns12OPb	7	

^aMW determined by SDS-PAGE (kDa).

activities of the capping enzyme and are then extruded through the major pore into the cytoplasm, where they can be translated to produce viral proteins. Non-structural viral proteins form viroplasm. The viroplasm is the site of most of the mRNA synthesis, genome replication, and assembly of core proteins. The various species of mRNAs synthesized in the viroplasm display in non-equimolar ratios. However, the dsRNA genome segments are usually packaged in exactly equimolar ratios (*i.e.*, one copy of each genome segment per particle), which indicates that the selection of viral mRNAs for packaging is highly specific; recognition signals on each mRNA species may be related to this process. Once a copy of an mRNA is inside a new viral core, the corresponding negative strand is synthesized, completing the replication of the dsRNA genome. The progeny viral core containing dsRNA then moves to the periphery of the viroplasm, where capsid proteins are assembled to form the complete new viral particle and are thought to stop further mRNA synthesis. Like the parental-virus-derived smooth cores (without a capsid layer), progeny viral cores also synthesize mRNAs, providing an amplification step during replication.

In recent years, researchers have proposed a new model for the genesis and maturation of viroplasms induced by SRBSDV in insect vector cells. The SRBSDV viroplasm consists of a granular matrix formed by the non-structural protein P9 and a filamentous matrix formed by the non-structural protein P5. The progeny viral cores and mature virions assemble in the filamentous matrix. The new viral mRNAs synthesized by the progeny viral cores and virions within the filamentous matrix are then transported to the granular matrix by association with SRBSDV non-structural protein P6. As a result, viral mRNAs accumulate in the granular matrix, and these viral mRNAs may either migrate to the adjacent cytoplasm to synthesize viral proteins that promote further maturation of viroplasms or migrate back to the filamentous viroplasm matrix and be packaged into empty core particles, which are assembled and aligned along the filaments within the filamentous matrix to form viral-RNA-containing core particles. The subsequent minus-strand RNA synthesis and outer capsid protein assembly processes also take place in the filamentous matrix, leading to the maturation of a new virion. Thus, the P9 formed granular matrix plays important roles in viral RNA and protein production and filamentous matrix maturation. Compared to the globular structures of phytoreovirus and oryzavirus viroplasms, the filamentous matrix of the fijivirus (SRBSDV) viroplasm gives it a much larger surface-area-to-volume ratio, which leads to produce more progeny virions in the insect vectors.

The control of gene expression for a multi-segmented genome is complex and not fully understood. Each genome segment contained within the viral core is associated with a single replicase/transcription complex, located adjacent to the major pore in the vertices of the icosahedron, and is transcribed separately to make full-length positive-strand RNA copies. The location of the replicase/transcription complex also restricts the number of genome segments to a maximum of 12. These 10–12 mRNAs are produced in different molar amounts based largely on segment size, resulting in more copies of the smaller mRNAs. The conserved 5'- and 3' terminal sequences of the RNA segments may function as recognition signals for the replicase/transcription complex and be involved in the efficient initiation of viral mRNAs translation, providing some control over the expression levels of individual virus genes. In addition, the conserved and non-hybridized 3'-terminal sequences have been shown to be required for efficient negative-strand synthesis. A third level of control

results from the presence of multiple or overlapping ORFs on one mRNA strand that are translated at different efficiencies. Lastly, some ORFs encode a polyprotein that must be processed to form functional proteins.

Distribution

Plant-infecting reoviruses are seasonally distributed as a result of plant host/crop cycles and abundance of insect vectors. They can overwinter in certain weed crops, in autumn-sown cereals, and in diapausing insect hosts. With SRBSDV, for example, the distribution of the virus is consistent with the migration and existing ranges of its main host vector, the white back planthopper (WBPH; *Sogatella furcifera*). In Southeast Asia, WBPH is a long-distance migratory rice pest. In most rice-growing areas in China, no winter rice is sown, with the exceptions being the southwestern province, Yunnan and Hainan. These two areas have therefore become the major WBPH overwintering regions in China. In early spring, viruliferous WBPH begin their migration northward with the help of the northeastward air flow, and by late June or early July, they can reach as far as Japan. As a result, viruliferous WBPH transmit the virus throughout its journey, in regions including Guangdong, Guangxi, Hunan, Jiangxi, and Fujian provinces in China. Moreover, viruliferous WBPH also lay eggs in the newly colonized areas, and the next generation of WBPH propagate on the infected rice seedlings, acquire viruses, become viruliferous, disperse, and induce new disease outbreaks. In the autumn, the viruliferous WBPHs then return to their ideal overwintering region, Yunnan and Hainan provinces, when the monsoon winds shift direction.

Plant reoviruses have been identified from every continent, but some genera are more widespread than others. Fijiviruses are the most widely distributed, which is not surprising given that they are the most numerous, although they are absent from North America and India. Fijiviruses have been reported to occur in Africa, Europe, South America, Asia, Australia, and the South Pacific Islands. Oryzaviruses have only been isolated from the Indian subcontinent and Asia, while phytoreoviruses have been isolated from North America, Asia, and Africa.

Host Range and Virus Transmission

Fijiviruses

The genus *Fijivirus* contains nine species whose members infect a range of monocotyledonous plants of the families *Gramineae* and *Liliaceae*. Common natural plant hosts include *Avena sativa*, *Oryza sativa*, *Saccharum officinarum*, and *Zea mays* among the *Gramineae* and *Allium sativum* among the *Liliaceae*, although this host range can be extended significantly by experimental virus infection. Viruses are transmitted by delphacid planthoppers (Table 1). Viruses can be acquired in juvenile plant growth stages, replicate in the vector, and, following a latent period of at most 2 weeks, are transmitted to plants in a persistent manner. No transovarial or seed transmission of viruses has been reported. In addition to transmission by insect vectors, mechanical transmission of the virus to susceptible hosts has been achieved for some fijiviruses with difficulty. In 2017, researchers reported a novel method for transmitting SRBSDV to rice, the bud-cutting method. With this approach, the 1.0-cm pregerminated rice bud was cut at a 45° angle approximately 0.5 cm from the base of the bud. The seedlings were then soaked in a crude SRBSDV extract obtained from SRBSDV-infected rice samples with a phosphate-buffered saline and then incubated in the dark at 28–30°C for 3 days. Lastly, the rice seedlings were planted in a soil matrix and incubated under growing conditions appropriate for rice. After 25 days, the rice seedlings displayed disease symptoms similar to those of rice seedlings infected with WBPH, and intact SRBSDV virions were isolated from the seedlings. The transmission efficiency with this method was more than 80%.

Oryzaviruses

The genus *Oryzavirus* contains two species whose members infect monocotyledonous plants of the family *Gramineae*. Common plant hosts include *O. sativa* and *Echinochloa crus-galli*. However, this natural host range can be extended by experimental virus infection to include other economically important species such as *Hordeum vulgare*, *Triticum aestivum*, and *Zea mays*. Virus is transmitted by delphacid planthoppers (Table 1). An acquisition period of 3 h is required followed by a 9-day latent period prior to transmission at all life stages in an intermittent manner. No transovarial transmission or mechanical transmission of virus has been reported.

Phytoreoviruses

The genus *Phytoreovirus* contains two species whose members infect monocotyledonous plants of the family *Gramineae* and one species that infects dicotyledonous plants. Common plant hosts include the *Gramineae* *O. sativa* and the dicot *Melilotus officinalis*. However, the natural host range of the dicot-infecting WTV can be extended significantly by experimental virus infection. Virus is transmitted by cicadellid leafhoppers (Table 1). Virus can be acquired after a short feeding period, replicates in the vector, and, following a 10- to 20-day latent period, is transmitted to plants throughout the life of the vector. Transovarial transmission of virus has been reported. Attempts to transmit the virus to susceptible hosts by mechanical methods have been unsuccessful.

Pathogenicity

The pathogenicity of plant reoviruses is particularly interesting as most viruses replicate in both insects and plant hosts. Most do not appear to cause any disease in the insect host, so that their pathogenicity is restricted to the plant host. The pathogenicity of fijiviruses varies considerably. Fiji leaf gall disease (caused by FDV) is reported to cause losses of up to 90% in susceptible sugarcane varieties, while NLRV has no known plant host and, therefore, no known pathogenicity. Oryzaviruses can also cause important yield losses. Rice ragged stunt disease (caused by RRSV) is reported to cause losses of generally 10%–20%, but sometimes as high as 100% in severe infections of susceptible varieties. The pathogenicity of phyto-reoviruses is much milder, although rice dwarf disease (caused by RDV) can have severe effects, as infected plants often fail to bear seeds. Furthermore, different strains of RDV–O, D84, and S–cause diseases of varying severity: RDV-S causes the severest symptoms and RDV-O the mildest.

Although many aspects of the molecular bases for pathogenicity remain opaque, we now know a small part of the processes leading to rice dwarf disease symptoms. As mentioned before, RDV-infected plants exhibit disease symptoms that include dwarfing, short crown root, and increased tiller number. Some researchers have found this phenomenon similar to the phenotype observed when plants lose their sensitivity to auxin signaling. After meticulous experimentation, researchers finally determined that the symptoms are caused by the RDV capsid protein P2, which disrupts interactions within the auxin coreceptor complex, leading to the plant's failing to perceive auxin. Besides this, viruses have evolved many other mechanisms to counteract plant antiviral strategies, most notably through the expression of viral suppressors of RNA silencing (VSRs). VSRs take advantage of RNA silencing and plant hormone pathways to destroy plant fortifications. As another example, the RDV nonstructural protein Pns11 can manipulate the plant ethylene biosynthesis process by enhancing the enzymatic activity of S-adenosyl-L-methionine synthetases, resulting in higher ethylene level in plants and making plants more susceptible to RDV infection. It is also has been demonstrated that RRSV infection mediates disease symptoms *via* suppression of jasmonic acid-associated antimicrobe signal through the microRNA319-TCP21 regulatory module.

Furthermore, the pathogenic effects of different plant-infecting reoviruses can influence each other. RRSV is widely distributed in Eastern and Southeastern Asia, but until recently had not caused serious crop losses in China owing to its low incidence. Lately, as SRBSDV has spread, RRSV has become more and more common in China and has been frequently found coinfecting plants with SRBSDV. Researchers propose that the recent increase in RRSV incidence in China is partially due to synergism between SRBSDV and RRSV because they have found that not only do rice plants coinfecting with both viruses exhibit higher virus titer and enhanced and earlier disease symptoms, but also the respective insect vectors of SRBSDV and RRSV acquired the viruses from the coinfecting plants at higher rates than from singly infected plants.

Diagnosis and Control

Diagnosis of plant reovirus infections can be done on the basis of disease symptoms or through the use of molecular tests. Symptoms vary between virus/host complexes but commonly include dark-green coloration, suppression of flowering, plant stunting, increased numbers of side shoots, and vein swelling or gall formation derived from phloem tissue on the backs of leaves. Given the variability in time to symptom expression and symptom severity, alternative tests are often used. Molecular and serological tests have been developed to assist in the diagnosis of viral infection in non-symptomatic plant material and vector insects. Serological tests are usually in enzyme-linked immunosorbent assay (ELISA) format and rely on polyclonal antisera raised against virions or expressed viral proteins. For example, antibodies against glutathione-S-transferase–NS7 fusion protein are useful in ELISAs for the detection of RRSV in infected plants and insect vectors, and a rapid, highly reliable, sensitive, and specific dot-format ELISA has been developed for SRBSDV detection using anti-SRBSDV rabbit antiserum. Recently, molecular tests such as reverse transcription–polymerase chain reaction (RT-PCR), which are faster and more specific than serology, have become the most common diagnostic methods. As increasing numbers of plant reovirus genomes are sequenced, species-specific primers are now becoming commonly available. A simple, rapid, sensitive, so-called multiple (m) RT-PCR method, which uses a mixed set of primers specific for *Rice stripe virus* (RSV), RBDSV, and SRBSDV together, can efficiently detect all three viruses simultaneously, and a one-step mRT-PCR assay has been established for simultaneous detection of RSV, RBDSV, and RDV. Meanwhile, a reverse transcription loop-mediated isothermal amplification (RT-LAMP) assay has been developed for the detection of RBDSV, RDV, RGDV, and RRSV, which can obtain a diagnosis within 2 h when combined with RNA rapid extraction.

Control strategies for plant reoviruses can be focused on either the host plant or the insect vector. Plant-based control through breeding to develop resistant plant species is most commonly utilized in combination with the removal of susceptible varieties and infected plants that provide a source of inoculum. This approach has provided robust control of RDV in rice and FDV in sugarcane. Genetic engineering of plant hosts has also been explored as an alternative control strategy. Pathogen-derived resistance approaches directed at either coat proteins or other viral genes to control RDV, RRSV, and FDV have proved less successful than those used to control other RNA plant viruses.

Control of insect vector populations with insecticide has provided some additional disease control. Unfortunately, chemical control appears to be of limited use in cases of high vector pressure. The current combination of diagnosis and control measures is already relatively effective and has resulted in reduced disease incidence and impact. Control measures based on protecting plants from WBPH, including seedbed coverage, chemical seed treatments, and chemical spraying of seedlings, have been demonstrated to be useful in controlling SRBSDV in China.

Despite breakthroughs over the years, such as the discovery that RDV can use actin-based tubule motility to overcome transmission barriers and exploit the ancient oocyte entry path of the bacterial symbiont in leafhopper, many gaps in our understanding of transmission mechanisms remain. However, this situation may improve in the future as advances in genome sequencing, reverse genetics systems, genome editing, and molecular technologies help provide more information on virus infection and replication and, thus, new resistance targets.

Future

Although our understanding of plant reoviruses is expanding, many of the molecular and biological properties of these viruses are still unknown. The complete sequence information now available for a number of these viruses will allow the production of cDNA probes to further elucidate the viral infection and replication processes in both plant and insect cells. The potential to produce infectious clones also holds promise for detailed studies of both plant and insect host ranges and of the methods of resistance employed by non-host species. This information, combined with knowledge gained from comparisons with animal reoviruses, may assist the further development of control strategies for diseases caused by these viruses in plants.

Further Reading

- Bamford, D.H., 2000. Virus structures: Those magnificent molecular machines. *Current Biology* 10, R558–R561.
- King, A.M., Adams, M.J., Carstens, E.B., Lefkowitz, E.J., 2011. *Virus Taxonomy – Classification and Nomenclature of Viruses: Ninth Report of the International Committee on the Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press.
- Hull, R., 2002. *Matthews' Plant Virology*. London: Academic Press.
- Mertens, P., 2004. The dsRNA viruses. *Virus Research* 101, 3–13.
- Mertens, P.P.C., Diprose, J., 2004. The bluetongue virus core: A nano-scale transcription machine. *Virus Research* 101, 29–43.
- Wei, T., Li, Y., 2016. Rice reoviruses in insect vectors. *Annual Review of Phytopathology* 54, 99–120.
- Miyazaki, N., Nakagawa, A., Iwasaki, K., 2013. Life cycle of phytoreoviruses visualized by electron microscopy and tomography. *Frontiers in Microbiology* 4, 306.

Relevant Websites

- <http://www.dpvweb.net/>
DPVWeb Home Page.
- <https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses (ICTV).
- <https://viralzone.expasy.org/>
ViralZone root - ExPASy.

Plant Resistance to Geminiviruses

Basavaprabhu L Patil, ICAR–Indian Institute of Horticultural Research, Bengaluru, India

Supriya Chakraborty, Jawaharlal Nehru University, New Delhi, India

Henryk Czosnek, The Hebrew University of Jerusalem, Rehovot, Israel

Elvira Fiallo-Olivé, Institute for Mediterranean and Subtropical Horticulture “La Mayora”–Spanish National Research Council–University of Malaga, Algarrobo-Costa, Málaga, Spain

Robert L Gilbertson, University of California, Davis, CA, United States

James Legg, International Institute of Tropical Agriculture, Dar-es-Salaam, Tanzania

Shahid Mansoor, National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Jesús Navas-Castillo, Institute for Mediterranean and Subtropical Horticulture “La Mayora”–Spanish National Research Council–University of Malaga, Algarrobo-Costa, Málaga, Spain

Rubab Z Naqvi and Saleem U Rahman, National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Francisco M Zerbini, Federal University of Viçosa, Viçosa, Brazil

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AGO Argonaute protein

CAS CRISPR-associated proteins

CMD Cassava mosaic disease

CMG Cassava mosaic geminiviruses

CRISPR Clustered regularly interspaced short palindromic repeats

DCL Dicer-like

dsRNA Double-stranded ribonucleic acid

GM Genetically modified crops

gRNA Genomic RNA

HR Hypersensitive response

LRR Leucine-rich repeat

mRNA Messenger RNA

NGS Next generation sequencing

nt Nucleotide(s)

QTL Quantitative trait loci

R Dominant resistance

RDR RNA-dependent RNA directed polymerase

RdRp RNA-dependent RNA polymerase

RGAs Resistance gene analogs

RIL Recombinant inbred line

RISC RNA-induced silencing complex

RNAi RNA interference

SNP Single-nucleotide polymorphism

ssRNA Single-stranded ribonucleic acid

VIGS Virus-induced gene silencing

Glossary

Allele Allele is a distinct form of a gene at a particular locus, that shows different effect on the phenotype from the other forms of alleles.

Betasatellite A type of circular single-stranded satellite DNA associated with begomoviruses.

CRISPR-Cas The acronym for Clustered Regularly Interspaced Short Palindromic Repeat – CRISPR-associated protein, an adaptive immune system used by bacteria and archaea against viruses and mobile genetic elements, which has been adapted for genome editing in eukaryotes.

Dominant allele An allele expressing its phenotypic effect even when having a recessive allele as its counterpart in a heterozygous system.

NB-LRR protein Plant disease resistance proteins containing a central nucleotide-binding domain (NB) and C-terminal leucine-rich repeat (LRR) domain. Also known as NBS-LRR or NB-ARC-LRR proteins.

Quantitative trait loci Corresponds to genomic regions that is associated with the phenotypic variation of a quantitative trait.

Recessive allele An allele whose phenotypic effect is not expressed in a heterozygote.

Resistance gene Plant gene conferring resistance to a specific pathogen, that possess a corresponding avirulence gene.

RNA interference A broad antiviral defense system used by plants, animals, and fungi. It recognizes double-stranded RNA as a cue of viral presence and induces degradation or silencing of viral transcripts or genomes.

Transcriptional gene silencing A mechanism of gene silencing involving decreased RNA synthesis because of promotor methylation.

Virus immunity Genetic mechanism by which plants are unable to support replication of a virus, and consequently where the virus cannot be isolated by any mean from the infected plant.

Virus resistance Genetic mechanism by which plants are able to defend themselves against a viral invasion (by one or the other mechanism) that leads to reduced viral replication levels.

Virus tolerance Genetic mechanism by which plants are infected and viral replication/infection occurs at wild type levels but without any visual symptoms.

Virus-induced gene silencing It is a technology that exploits an RNA-mediated antiviral defense mechanism.

Introduction

Domestication of crops from their wild relatives has resulted in the loss of many resistance genes because these are often genetically linked to unwanted traits (such as low yield, poor flavor, small size). With the commercialization of agriculture and increasing production of crop plants in monoculture, genetic diversity has dwindled, which has made crops susceptible to a variety of pathogens. Therefore, breeding for resistance has consisted in reintroducing these resistance genes in the susceptible domesticated crop by crossing and selection. The wild relatives or ancestors of these crop plants present in their centers of origin or diversity are the best sources of resistance genes (R genes). Pathogen resistance in crop plants can also be generated by either natural or induced mutations. For example, Rice yellow mottle virus (RYMV) resistance in cultivated rice spontaneously appeared under RYMV pressure. Cassava varieties such as TME3 are also probably mutants which were selected under cassava mosaic disease pressure.

Geminiviruses are single-stranded, circular DNA viruses that cause economically important diseases in a wide range of crop plants across the globe. Geminiviruses are characterized by small geminate particles (18×30 nm) containing either one (A component, monopartite) or two (A and B components, bipartite) single-stranded circular DNA molecules of ~ 2.7 kb and sometimes associated with satellite DNA molecules of ~ 1.4 kb. Around 460 distinct species and a larger number of strains belong to the family *Geminiviridae*. In this article, we briefly discuss the resistance genes and the sources of resistance against the most important geminiviral diseases in cassava, tomato, bean, maize, and cotton.

An Overview of Natural Resistance to Viruses in Plants

Depending on the number of genes involved in providing the resistance phenotype, the host resistance can be classified into monogenic, digenic, or polygenic. Many of the monogenic and digenic types of resistance are qualitative in nature, with phenotypes showing complete absence of pathogen in the host plant. Whereas the polygenic resistance is manifested by several genes. A quantitative trait locus (QTL) is a region of DNA that is associated with a specific phenotypic trait, that varies in degree and can be attributed to polygenic effects, i.e., the product of two or more genes, and their interaction with the environment. Several such QTLs could be associated with one single trait. Such QTLs are mapped by genetic markers by using recombinant inbred lines (RILs) that are derived by crosses between parents with two contrasting resistance phenotypes. The inheritance pattern of the resistance trait can be dominant or recessive in nature and half of the virus resistance traits known hitherto are recessive in nature.

Plant R genes confer resistance to plant pathogens including plant viruses and each R gene confers resistance to a specific pathogen triggered by a specific avirulence (*avr*) gene or an effector protein. More than 80% of the R genes studied so far are monogenic in nature and one third of the R genes are tagged by molecular genetic markers. R-genes can be either dominant or recessive in nature. Studies done so far have shown that dominant R genes are of the NB-LRR (Nucleotide-Binding-Leucine-Rich Repeat) type and that recessive ones encode for translation initiation factors. Dominant gene resistance is usually associated with HR while recessive genes function at single cell level. R genes are known to show two unique patterns of clustering: in one case R genes with similar inheritance and resistance phenotypes are clustered at one locus, while in another case the R genes for virus resistance are clustered with unrelated R genes. Most of the recessive R genes identified to date are against potyviruses. Although there are no reports of R genes showing resistance to geminiviruses, there are reports of genetic mapping of the resistance loci through molecular markers. There are also reports demonstrating the interaction of geminivirus proteins with the trans-membrane receptor kinases and components of brassinosteroid signaling. It was recently shown that silencing of a *Permease 1-like protein* gene transmembrane transporter in tomato using the virus-induced gene silencing (VIGS) vector system, rendered it susceptible to Tomato yellow leaf curl virus. In response to the pathogen recognition by the R gene, changes in ion fluxes are observed, leading to activation of signaling pathways, alteration in transcriptional profiles, production of reactive oxygen species (ROS) and generation of nitric oxide (NO).

Gene silencing is another important mechanism which plants have evolved to defend against viruses. In plants, post-transcriptional gene silencing (PTGS) acts as a natural anti-viral defense system and is considered as part of the plant's innate immune system. RNA silencing involves suppression of gene expression by sequence-specific interaction at the transcriptional or post-transcriptional level in diverse eukaryotes. Double stranded RNA (dsRNA) is the trigger for gene silencing, which involves a chain of events to produce small interfering RNA (siRNA) through a diverse range of enzymes and its complexes, including the RISC-complex (RNA induced silencing complex) as well as DCL (Dicer-Like) and RDR (RNA dependent RNA polymerase) proteins. In virus-infected plants, RNA silencing is initiated by the double-stranded (ds) RNA that can be a viral replication intermediate or by 'aberrant' RNAs, or single-stranded RNA (ssRNA) that is converted to dsRNA by host-encoded RNA-dependent RNA polymerase (RdRP). The dsRNAs can also be formed because of bi-directional transcription of these viruses with transcripts occurring from opposite polarity overlapping at their 3'-ends, which partly explains how these geminiviruses induce PTGS in infected plants. Alternatively, the early and abundant transcripts of the AC1 gene of these geminiviruses can serve as the template for the host RdRP to produce dsRNAs. Yet another possibility is the fold-back structures of geminivirus transcripts which can serve as a template for DICERs to cleave at specific locations and produce siRNAs. In plants, some virus-host interactions naturally lead to host recovery that are similar to RNA-mediated virus resistance. This symptom recovery phenomenon is also reported for some of the geminiviruses (Fig. 1). This recovery phenomenon is associated with the production of virus-derived siRNAs, that later become abundant in the newly developed symptom-less recovered leaves. This increase in virus-derived siRNA accumulation is accompanied by a reduction in the levels of both viral DNA and mRNA accumulation.



Fig. 1 Symptom recovery in *Nicotiana benthamiana* plants infected by two distinct cassava mosaic geminiviruses (ACMV and EACMKV), as a result of gene silencing.

Since RNA silencing and R gene signaling are both components of the plants innate immunity, it will be interesting to know if there is any cross-talk between these two pathways through one of the systemic signaling systems such as the salicylic acid pathway. It has been shown that the RdRPs *NbRdRp1m*, *NtRdRp1* and *RDR1* are all inducible by both salicylic acid and certain viruses, although recent studies failed to find any overlap between the DCL endoribonucleases involved in silencing and the SA-induced resistance to positive-sense RNA viruses. More recently it has been shown that the virus resistance induced by NB-LRR proteins involved Argonaute4-dependent translational control. It has also been shown that the viral silencing suppressor HC-Pro enhances R-gene mediated resistance to viruses and this resistance is weak at higher temperatures. Geminiviruses are known to be associated with DNA satellites which modulate the viral symptoms and the betasatellites in particular are known to enhance the symptoms through their PTGS suppressor proteins (RC1).

In this article we discuss in greater detail and on a case by case basis the principal sources of resistance for the major geminiviral diseases of the world (Table 1).

Resistance to Cassava Mosaic Geminiviruses

It was recognized from the earliest period of research on cassava mosaic geminiviruses (CMGs) (Family *Geminiviridae*; Genus *Begomovirus*) that resistance would be the most effective method for controlling the cassava mosaic disease (CMD) caused by these viruses. The team responsible for the first biological characterization of CMGs in the former Tanganyika in the 1930s realised that wild relatives of cassava (*Manihot esculenta* Crantz) would provide an important potential source of resistance genes. Resistant progeny from *Manihot esculenta* × *Manihot glaziovii* (Müll-Arg.) crosses were backcrossed with cultivated cassava to produce the first CMD-resistant cassava varieties. Material from this initiative was subsequently used in the 1970s to extend the development of CMD resistance within the cassava breeding program of the International Institute of Tropical Agriculture (IITA). A recurrent selection approach was taken with the primary source of CMD resistance being the cassava clone 58308. It had been suggested since the 1950s that the resistance source derived from the *M. glaziovii* cross was polygenic, but this was also shown to be recessive in the early 1970s. During this period, evaluations of resistant varieties were undertaken in West, Central and East Africa, as well as in south India. Resistance was shown to be effective in all of these regions, providing a strong demonstration of the broad activity of the resistance source as well as its likely durability. We now know that there is significant virus diversity across these regions (ACMV, EACMV-like viruses, ICMV and SLCMV), which confirms the validity of the earlier conclusions about the broad-based activity of the *M. glaziovii*-derived resistance. This polygenic or quantitative recessive resistance source is now referred to as CMD1.

In the 1980s and 1990s in Africa there were extensive trial-based epidemiology studies of CMGs, several of which compared the responses of varieties with contrasting levels of resistance. Different components of resistance were identified, including: resistance to infection, resistance to virus multiplication, resistance to virus movement (leading to incomplete systemicity), and resistance of normal plant function to the effects of virus infection. These were assessed through measurements of virus titre, incidence of infected plants, severity of symptoms of infection, distribution of symptoms through the plant, and yield of healthy versus infected plants. Strong correlations were demonstrated between each of these variables. Resistant varieties were therefore infected less, had

Table 1 List of geminivirus resistant genes/loci and their characteristics for the major geminivirus diseases

Target virus (es)	Target host	Gene/Locus	Nature/Location of resistance gene/locus	Source of resistance	Type of resistance
CMGs	Cassava	CMD1 CMD2 CMD3		<i>Manihot glaziovii</i> West African Cassava landraces Cassava genotype TMS97/2205	Polygenic recessive Single dominant gene Polygenic or quantitative
TYLCVs	Tomato	Ty-1 Ty-2 Ty-3 Ty-4 ty-5 Ty-6 –	RdRp NB-LRR RdRp – Pelota – –	<i>Solanum chilense</i> <i>S. habrochaites</i> <i>S. chilense</i> <i>S. chilense</i> <i>S. peruvianum?</i> <i>S. chilense</i> <i>S. habrochaites</i>	Partial dominant Dominant Partial dominant Partial dominant Recessive Partial dominant Oligogenic (Complete resistance)
ToLCV		tgr-1	Cell-to-cell movement	<i>S. chilense</i>	
ToCMoV		tcm-1		<i>S. lycopersicum</i>	
ToLCNDV	Cucurbits	CmoCh08G001490	BZIP	<i>Cucurbita moschata</i>	Quantitative
PYMV	Potato	–	–	<i>S. pimpinellifolium</i>	Bigenic
BCTV	Common bean	Bct-1 –		Red Mexican common bean varieties <i>Arabidopsis thaliana</i> (3 ecotypes, Ms-0, Pr-0, and Cen-0)	Single dominant gene Monogenic
CaLCuV and TYLCV		gip-1		<i>Arabidopsis thaliana</i> (Pla-1 ecotype)	
BGYMV	Common Bean	bgm-1 bgm-2 bgm-3 Bgp-2		<i>Phaseolus vulgaris</i> (Red S) <i>Phaseolus coccineus</i> (scarlet runner Bean)	Recessive Dominant
MYMIV	Black gram Soybean	CYR-1 Glyma18g02850	Non-TIR-NBS-LRR LRR-RP (on chromosome 18)	<i>Vigna mungo</i> <i>Glycine max</i>	Dominant Dominant
MSV	Maize	msv1 Q1 - PHM5502_31 Q2 - PZA02616_1 Q3 - PZA02872_1 Q4 - PHM1766_1	Chromosome-1 SNP on: Chromosome-3 Chromosome-3 Chromosome-7 Chromosome-9	<i>Zea mays</i> (CML206 × CML312) <i>Zea mays</i> (TZIL07A01005 × TZIL07A01322)	Partial dominant effect Dominant Additive Dominant Additive & dominant

Abbreviations: CMGs, cassava mosaic geminiviruses; TYLCVs, tomato yellow leaf curl viruses; ToLCNDV, tomato leaf curl New Delhi virus; ToLCV, tomato leaf curl virus; ToCMoV, tomato chlorotic mottle virus; PYMV, potato yellow mosaic virus; BCTV, beet curly top virus; CaLCuV, cabbage leaf curl virus; BGVMV, bean golden yellow mosaic virus; MYMIV, mungbean yellow mosaic India virus; MSV, maize streak virus; NBS-LRR, nucleotide-binding site Leucine-rich repeat; LRR-RP, leucine-rich repeat receptor-like protein kinase.

less severe symptoms which were spatially limited in the plant, had lower virus titers and yielded more, demonstrating that each of these were expressions of the same resistance.

Following the increasing accessibility of PCR and sequencing analyses from the late 1990s onwards, it became possible to examine the molecular characteristics of CMD-resistant cassava varieties. A key development came in 2002 when a novel source of resistance to CMGs was identified from West African cassava landraces. Molecular genetic mapping was used to identify a single dominant gene referred to as CMD2, and this development made it possible for breeding programs to speed up their resistance breeding work through marker-assisted selection (MAS). Several markers were identified which could pick out clones carrying CMD2 with a high degree of reliability. Moreover, similar approaches also led to the identification of a third resistance source – CMD3 – in the genotype TMS97/2205. This clone also carried CMD2, and it was observed that the combination of CMD2 and CMD3 gave the plants very high levels of resistance to CMD. An additional spin-off from the development of MAS was that the technique could be applied in regions where CMGs did not even occur, but where there was an interest in having strategic stocks of CMD-resistant germplasm. This would allow these regions to respond rapidly to any future invasive spread of CMGs. This activity was pioneered by the International Center for Tropical Agriculture (CIAT) in Colombia, to prepare for the potential future introduction of CMGs to Latin America.



Fig. 2 (Top) CMD resistant cassava variety TME3 showing recovery from ACMV (2 weeks post inoculation) and EACMV (4 weeks post inoculation) infections in glasshouse experiments. Reproduced from Patil, B.L., Fauquet, C.M., 2015. Studies on differential behavior of cassava mosaic geminivirus DNA components, symptom recovery patterns, and their siRNA profiles. *Virus Genes*. 50 (3), 474–486. doi:10.1007/s11262-015-1184-y. (Bottom) CMD resistant and susceptible varieties of cassava cultivated in fields of Burundi.

In addition to extensive field evaluations for CMD resistance, targeted approaches such as microparticle bombardment using infectious clones have been used to screen for resistance to CMGs. In comparisons of ACMV and EACMV resistance, putative resistant accessions with CMD1, CMD2, or CMD3 were shown to have high levels of resistance to ACMV yet were severely affected by EACMV. The only exception to this was TMS97/2205 which was known to combine CMD1, CMD2, and CMD3. As a consequence of these and related findings, the primary objective of most cassava geminivirus resistance breeding programs is to combine these different resistance sources into varieties that also have favorable yield, organoleptic, and processing qualities.

There has been limited progress to date in identifying specific genes associated with CMD1, CMD2, and CMD3. However, transcriptome data have revealed differences between CMG-susceptible (T200) and CMG-resistant (TME3; Fig. 2) cassava varieties. Moreover, resistance gene analogs (RGAs) were also shown to be expressed differentially during the process of symptom recovery (from CMD infection), suggesting that they may provide a complementary resistance mechanism to RNA silencing. Methylation of viral DNA has been shown in several systems to be an important resistance strategy employed by plants against geminivirus infection and has been demonstrated specifically for ICMV. However, although resistance sources such as CMD2 have been pinpointed down to chromosome and linkage group level through genome-wide sequencing and mapping studies, further research is required before a more comprehensive understanding can be obtained of the genetics of the responses of cassava plants to virus infection.

While studies continue to expand knowledge about natural resistance to CMGs, there are widespread and highly successful programs being implemented to deploy CMD-resistant varieties (Fig. 2). These were disseminated throughout East and Central Africa to control the severe CMD pandemic of the 1990s/2000s with great success, and similar programs are now being used to address the more recent CMD pandemic in South-East Asia. Africa-derived CMD resistance has proven to be effective in controlling virus infection caused by the South Asian CMGs – ICMV and SLCMV – so they should also provide effective control for the currently spreading South-East Asian pandemic associated with SLCMV.



Fig. 3 Resistant (green plants) and susceptible tomato plants (dead or yellow plants) from a tomato breeding field of a program for begomovirus resistance in Guatemala, tomato experimental field in Sanarate Guatemala, 2005. Funded by USAID-CDR. Courtesy: Luis Mejia, Douglas P Maxwell, Favi Vidavski, Henryk Czosnek.

Geminivirus Resistance in Tomato

The Case of Resistance to Tomato Yellow Leaf Curl Virus

Tomato crops can be infected by a large number of geminiviruses (family *Geminiviridae*) mainly belonging to the genus *Begomovirus*. Tomato yellow leaf curl disease, one of the most devastating virus disease affecting tomato worldwide, is caused by a complex of begomoviruses originated in the Middle East but expanded initially to the Mediterranean Basin and then to the Far East, the Caribbean, North America, and Australia. Tomato yellow leaf curl virus (TYLCV) was the first monopartite begomovirus characterized and the most successful in spreading the disease worldwide.

Breeding for TYLCV Resistance

Breeding programs aimed at producing tomato varieties resistant to TYLCV started in the late 1960s and have expanded since (Fig. 3). These programs are based on the introgression of resistance found in some accessions of wild tomato species into the domesticated tomato (*Solanum esculentum*). Certain accessions of the wild tomato species *S. chesmanii*, *S. chilense*, *S. habrochaites*, *S. peruvianum*, and *S. pimpinellifolium* are resistant (with mild or no symptoms, but containing virus) to whitefly-mediated inoculation. The discovery of loci, and later on of genes, associated with TYLCV resistance, was facilitated by the development of saturated maps based on DNA polymorphism distinguishing resistance from susceptibility (RFLP, AFLP, SSR, SCAR, etc.).

To date, six loci, coined *Ty-1* to *Ty-6*, have been found to be associated with TYLCV resistance (Table 1). They have been mapped on different tomato chromosomes and five of them are independently inherited. The *Ty-1* and *Ty-3* loci originated from *S. chilense* accessions LA1969 and LA2779, respectively, and they were mapped to the long arm of tomato chromosome 6. Later, *Ty-3* was shown to be allelic to *Ty-1*. *Ty-2* was introgressed from *S. habrochaites* accession B6013 and was located on the long arm of chromosome 11. *Ty-4* was introgressed from *S. chilense* accession LA1932 and mapped to the long arm of chromosome 3. Compared to other *Ty* genes, *Ty-4* is less effective against TYLCV. *ty-5*, the only recessively inherited *Ty* gene known to date, was introgressed from *S. peruvianum* and mapped on chromosome 4. *ty-5* has been reported from different sources including line TY172 (derived from four different accessions of *S. peruvianum* which were crossed with *S. arcanum*) and tomato cultivar Tyking. *Ty-6* originated from *S. chilense* accessions LA1938 and LA2779 and was located on the long arm of chromosome 10. It has been shown that *Ty-6* is effective, in addition to TYLCV, against Tomato mottle virus, a New World begomovirus.

Pyramiding resistance from various wild tomato species in a single tomato line often broadens resistance. Today, most commercial tomato varieties resistant to TYLCV contain the *Ty-1* gene. Upon whitefly-mediated infection in the field and greenhouse, the plants remain symptomless and their yield compares with that of non-infected plants. However, reports in 2008 indicated that *Ty-1*-mediated resistance could be broken under high virus pressure. Worse, it was recorded in 2016 in Jordan and Israel that *Ty-1* resistant varieties suddenly became susceptible to TYLCV. Molecular cloning and sequencing indicated that the symptomatic plants were contaminated with the cotton leaf curl Gezira betasatellite, which in association with TYLCV, overcame the resistance based on *Ty-1*. These findings were later confirmed in the laboratory using cloned virus and betasatellite.

Characterization of Ty Genes

Three of the *Ty* loci have also been associated to genes which have been recently cloned and characterized.

Ty-1/Ty-3 were identified as coding for an RNA-dependent RNA polymerase (RdRp) belonging to the RdRp type which has an atypical DFDGD motif in the catalytic domain. It has been shown that *Ty-1* confers resistance to TYLCV and other begomoviruses by enhancing transcriptional gene silencing through an increase in cytosine methylation of viral genomes.

Ty-2 encodes a nucleotide-binding domain and leucine-rich repeat-containing (NB-LRR) gene, *TYNBS₁*. Introduction of genomic fragments containing the *TYNBS₁* gene into susceptible tomato plants resulted in transgenic lines resistant to TYLCV.

ty-5 is a loss-of-function mutant allele of the gene that encodes a messenger RNA surveillance factor Pelota (*Pelo*) located on chromosome 4. Pelota is involved in ribosome recycling phase of protein synthesis. The mutation in *ty-5* is caused by a T-to-G transversion in the coding region.

Ty-6 is located on chromosome 10, that compliments resistance conferred by both *Ty-3* and *ty-5* genes, thus providing resistance to both monopartite and bipartite begomoviruses.

Genetic Engineering for TYLCV Resistance

In some cases, as mentioned above for *Ty-2*, host-derived resistance has been engineered into susceptible tomato plants. Nonetheless, most of the tomato transgenic plants with incorporated begomovirus resistance were produced by expressing viral genes. TYLCV genes or gene segments, whether functional or not, have been used to confer a narrow as well as a broad-range virus-specific resistance (pathogen-derived resistance).

The expression of the TYLCV *CP* gene in an interspecific *S. lycopersicum* x *S. pennellii* hybrid was characterized by the late development of disease symptoms, often followed by their disappearance and recovery. The expression of the TYLCV *Rep* gene in tomato cultivars provided a virus-specific resistance. For example, tomato plants transformed with a truncated TYLCV *Rep* gene exhibited various degrees of resistance to TYLCV infection in the field in Florida. However, these TYLCV-resistant plants were susceptible to the bipartite begomovirus Tomato mottle virus conspicuous in Florida tomato fields. Similarly, tomato plants transformed with a 3' end-lacking *Rep* gene of Tomato yellow leaf curl Sardinia virus (TYLCSV, a monopartite begomovirus closely related to TYLCV) were resistant to TYLCSV from Italy but were susceptible to a cognate TYLCSV isolate from Spain.

Resistance was also achieved by mimicking the plant RNA interference (RNAi) antiviral mechanisms relying on pathogen transcriptional (TGS) and post-transcriptional (PTGS) gene silencing. Small interfering RNAs (siRNA) directed against TYLCV genes were designed, based on hairpin RNAi (hpRNAi), and used to produce transgenic tomato plants. It was found that hpRNAi-based resistance was highly sequence-specific. The first set of experiments using PTGS hpRNAi demonstrated that TYLCV and TYLCSV non-coding regions could trigger the accumulation of virus-specific siRNAs, followed by high levels of resistance against these two viruses. TYLCV *Rep* was also the target of hpRNAi. Homozygous transgenic tomatoes expressing hpRNAi TYLCV *Rep* were resistant to TYLCV in greenhouses and fields and carried large quantities of small RNAs. In addition, resistance was associated with changes in the general plant host transcriptome, mainly genes involved in biological regulation, metabolic and cellular processes such as catalytic and binding activities. The hpRNAi strategy has also been used against the *IR*, *CP*, *V2*, and *Rep* genes of TYLCV from Oman. The transgenic lines obtained exhibited various levels of resistance upon TYLCV agro-inoculation.

The CRISPR-Cas system, which protects prokaryotes from potentially deleterious foreign DNA, is the latest tool in the molecular breeder's panoply aimed at conferring TYLCV resistance. CRISPR-Cas was recently implemented to the making of virus-free crops, demonstrating that this technology can be applied to the molecular breeding of virus-immune tomatoes. Indeed, TYLCV resistance was conferred by editing the tomato genome. Targeting the TYLCV *CP* and *Rep* genes with Cas9 single RNA-guided DNA endonuclease resulted in stable resistance to the virus. The Cas9-treated tomato plants carried reduced TYLCV amounts and their progeny was resistant to TYLCV. Therefore, antiviral strategies based on genome editing have the potential to constitute an additional, maybe preferred, tool in the breeding of TYLCV-resistant tomato varieties.

Resistance to Beet Curly Top Virus

Beet curly top virus (BCTV, genus *Curtovirus*, family *Geminiviridae*), is a phloem-limited virus that causes curly top disease in several economically important crops such as common bean, pepper, sugar beet, and tomato. This disease is more prevalent and is known to cause heavy economic losses in the western USA, north-central Mexico and countries of the Mediterranean Basin and the Middle East (e.g., Iran and Turkey). Under favorable environmental conditions the disease outbreaks are driven by populations of the beet leafhopper vector (*Circulifer tenellus*).

Sources of curly top disease resistance have been identified in germplasm of common bean, flax, squash and sugar beet. However, the availability of commercially available resistant varieties that can be used in a management program for BCTV depends on the crop plant. Resistant varieties have played a major role in management of curly top in sugar beets and somewhat of a role in common bean, whereas there are no commercially available resistant varieties for pepper or tomato. In the early 1900s in the western USA, curly top disease emerged as a major constraint on sugar beet production, in some cases resulting in sugar refineries having to close. To search for resistance, mass selection from heavily infested fields led to the identification and release of the first curly top resistant sugar beet variety, US1, in 1933. This variety became widely grown in the western USA, despite some shortcomings. The identification and selection of resistant sugar beet varieties has been facilitated by the use of breeding plots, known as curly top nurseries, in which high populations of leafhoppers viruliferous for the most prevalent and virulent strains of BCTV are released. As a result of the high selection pressure in these nurseries, combined with the development of commercial hybrids and molecular breeding technologies, curly top resistance in sugar beet has advanced tremendously, despite this being a complex quantitatively inherited trait. Modern curly top resistant varieties have greater uniformity, higher yield and improved

resistance. However, plants of these resistant sugar beet varieties will develop curly top symptoms and may suffer yield losses when infected at an early stage of development (two to four leaf stage). Therefore, resistant varieties need to be part of an IPM program that addresses the susceptibility of young plants, e.g., seed treatment with systemic insecticides.

Common bean is another crop that can be heavily impacted by curly top in the western USA. Curly top resistance was identified in some common bean varieties (e.g., Red Mexican type). Genetic studies identified a single dominant gene (*Bct-1*) associated with curly top resistance in common bean. Breeding efforts led to the release of the commercial curly top-resistant varieties Great Northern UI 15 and Red Mexican UI 34. Subsequently, snap bean varieties with curly top resistance were developed and released. These varieties have provided acceptable yields under high curly top pressure. Now there is a need to develop curly top-resistant varieties for other classes of common bean.

In terms of tomato and pepper, curly top resistance was not found in commercial cultivars. However, high levels of resistance were found in accessions of some wild species of tomato, including *Solanum chilense*, *S. habrochaites*, *S. lycopersicoides* and *S. peruvianum*. Unfortunately, introgression of this resistance into commercial tomato varieties has been difficult. However, as described in the case of resistance to TYLCV, a series of 6 loci (genes) conferring some degree of resistance to the virus, have been introgressed into tomato breeding lines. These lines have been used to generate commercial varieties with effective resistance against monopartite begomoviruses such as TYLCV. The mechanisms of Ty gene products determined to date, such as the Ty-1/Ty-3 enhancing transcriptional gene silencing (TGS) and cysteine methylation of the viral genome, appear to be common to all geminiviruses rather than genus-specific. Therefore, BCTV agro-inoculation screening was used to assess the response of 15 breeding lines, having different combinations of Ty genes (e.g., Ty-1, Ty-2, Ty-3, ty-5, and Ty-6), kindly provided by the World Vegetable Center. The response was assessed based upon a 0–4 rating scale, with 0=no symptoms to 4=severe stunting and leaf curling and vein purpling. In screens with the BCTV-Svr-[US: SVR:Cfh], the susceptible control received a rating of 4.0, whereas 11 of the breeding lines had ratings <2.5, with three lines with rating of 1.5–1.7 (lines with Ty3 only, Ty3 and Ty2 and Ty5 and Ty6). In an equivalent screen with BCTV strain BCTV-LH-71-[US: Cal:10], which was associated with the 2013 curly top outbreak in processing tomatoes in California, only 6 of the Ty breeding lines had ratings <2.5, with two lines having ratings of 0.9 and 1.2 (both having the Ty2 and Ty3 combination). Together, these results indicate that pyramiding Ty genes in tomato can provide resistance to BCTV, consistent with these gene products targeting a general geminivirus property, such as replication or transcription. Importantly, the effectiveness of the resistance was not the same for two strains of BCTV. A similar result was reported in the reverse scenario in common bean, where the *Bct-1* gene that confers qualitative (strong) resistance to BCTV, provided a high level of quantitative resistance to the bipartite begomovirus, Bean dwarf mosaic virus (BDMV). These results suggest the possibility of generating broad spectrum geminivirus resistance.

To assess the potential for using Ty genes to generate a curly top resistant tomato variety, a breeding line was generated by screening progeny of crosses aimed at pyramiding the Ty-1, Ty-2, and Ty-3 genes by agro-inoculation. Plants with a resistance phenotype were then selected and further selection performed by selfing plants and screening progeny by agro-inoculation. The breeding line that was generated, named Line 20#12, carries the Ty-1, Ty-2, and Ty-3 genes and has moderate to strong resistance to curly top (Fig. 4). Moreover, the resistance in this line involves initial development of mild up-curling and vein purpling in the first true leaves, followed by a recovery phenotype in which all subsequently emerging leaves show no symptoms (although virus can be detected in these leaves). This resistance phenotype is consistent with TGS targeting the BCTV genome for methylation, reducing transcription and replication. Thus, these results suggest that pyramiding Ty genes in a commercial tomato background could be a strategy to generate commercial curly top-resistant varieties.

Geminivirus Resistance in Beans

Geminiviruses cause severe diseases in common bean (*Phaseolus vulgaris*), and also in other species of *Phaseolus* such as lima bean (*P. lunatus*). In the Americas, the viral species involved are Bean golden mosaic virus (BGMV), Bean golden yellow mosaic virus (BGYMV) and Macropitium yellow spot virus (MaYSV), all of which cause similar symptoms of severe yellow ("golden") mosaic. BGMV and MaYSV occur in South America (MaYSV has so far been only reported in Brazil), while BGYMV is present throughout Central America and the Caribbean.

Many institutions in the Americas had breeding programs for resistance to golden mosaic disease in *Phaseolus*. The most successful program was conducted at the International Center for Tropical Agriculture (CIAT) in Cali, Colombia, starting in the mid-1970s. This program targeted BGYMV, although most of the relevant literature refers to BGMV, since the distinction between the two viruses was not completely clear at the time. Indeed, the fact that resistance to BGYMV was not successful against BGMV provided additional evidence for the distinction between the two viruses. A first source of resistance to BGYMV was identified in a black bean landrace named Porriño Sintético, originated from El Salvador (Central America). The analysis of segregating populations derived from crosses with this source indicated that resistance was quantitative. Resistance was partial and often broke down under high inoculum pressure. The red-seeded line DOR 364 provided additional quantitative resistance genes, and a number of cultivars containing these two sources, with resistance that was markedly superior to that conferred by either source alone, were released during the 1980s. During that decade, virtually all resistant cultivars that were deployed in Central America and the Caribbean derived from these two sources. These cultivars were largely responsible for the recovery of bean yields in that region. The main limitation with these sources of resistance was that all resistant cultivars produced black seeds, and many of the commercial common bean cultivars grown in disease-prone regions had different seed colors that were changed when crossed with

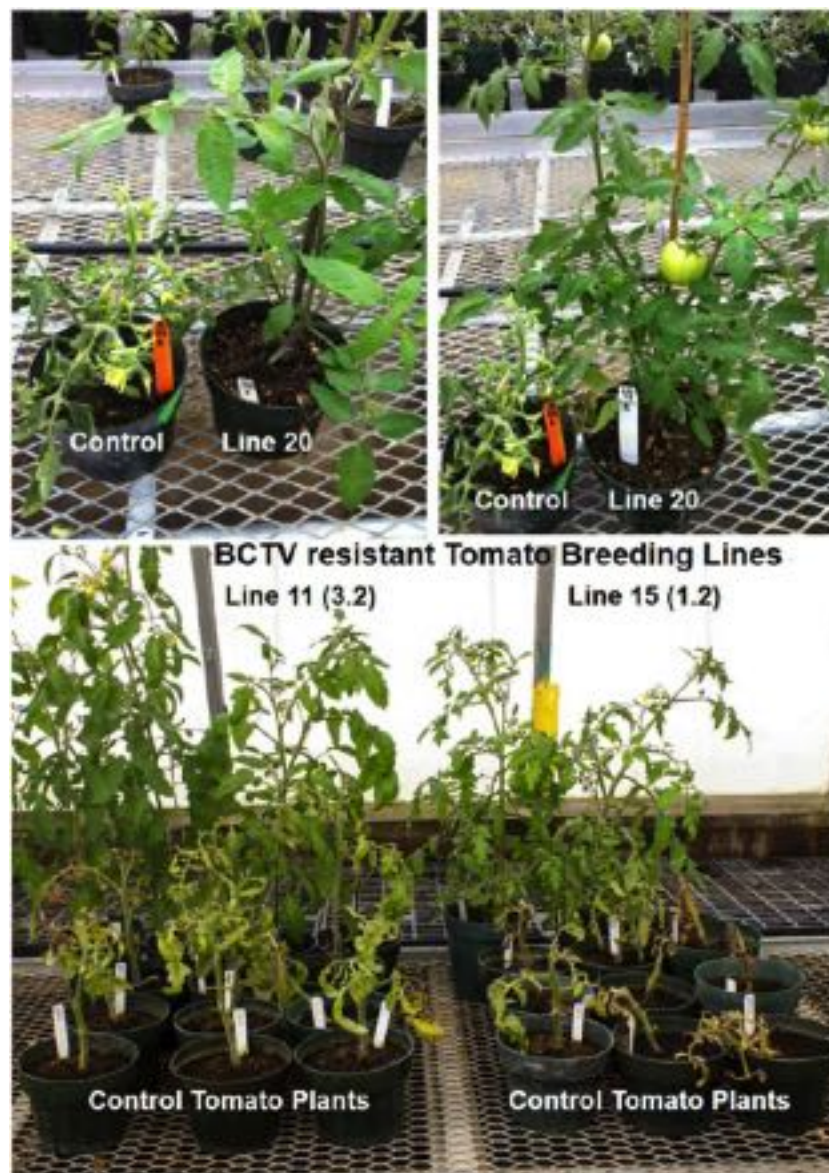


Fig. 4 Advancement and characterization of tomato breeding lines for beet curly top virus resistance, by agro-inoculation of BCTV in glasshouse conditions at the University of California, Davis, CA, USA.

the black-seeded parental sources of resistance. Moreover, resistance was not complete and could still be overcome in years when conditions were highly favorable to the disease.

An improved source of resistance was identified in the early 1990s in Garrapato, a landrace from the Durango group from Mexico. A breeding line derived from Garrapato, named A429, provided the best level of resistance. Inheritance studies indicated that the resistance in A429 was controlled by a single recessive gene, named *bgm-1*. The presence of the *bgm-1* gene strongly reduces the mosaic and yellowing symptoms caused by BGYMV. The *bgm-1* containing DOR lines developed by the breeding program at CIAT remain the best source of durable and stable BGYMV resistance (Fig. 5).

A co-dominant sequence-characterized amplified region (SCAR) marker, SR2, which is tightly linked to *bgm-1*, was developed by genetic mapping. The position of the *bgm-1* gene was estimated to be close to that of *bc-1*, which confers resistance to the potyvirus Bean common mosaic virus (BCMV), on chromosome 5. Interestingly, the mapping study indicated that *bgm-1* may be sub-telomeric. Sub-telomeric regions tend to be highly recombinogenic, a characteristic that is advantageous in the context of host-pathogen co-evolution.

The success at obtaining resistant varieties to BGYMV was not, unfortunately, replicated in South America for BGMV (a different virus with a low degree of sequence identity with BGYMV). Breeding for resistance to BGMV initiated in the 1970s at the Campinas Agronomical Institute (IAC in the Portuguese acronym) in Campinas, Brazil, and also at center for bean research at Embrapa, in Goiânia. Although a number of varieties with delayed incidence of the disease were obtained, inheritance was complex, yields were



Fig. 5 The "Mosaico Dorado Amarillo de Frijol" (Bean golden yellow mosaic virus) nursery established at San Andrés Experimental Station of the Centro Nacional de Tecnología Agropecuaria y Forestal "Enríques Álvarez Córdova" in El Salvador (Central America), as part of the breeding program conducted by CIAT. The nursery contains a large number of landraces and breeding lines of black and red beans with different levels of resistance to Bean golden yellow mosaic virus (BGMV). Courtesy of M.A. Grajales.

unremarkable, and their agronomical traits were not generally favored by growers. The most promising varieties were released during the 1990s by the breeding program at the Paraná Agronomical Institute (IAPAR) in Londrina. However, the quantitative, partial resistance provided by these lines would break down under the high BGMV incidence levels that often occurred in southern Brazil. The lack of success at obtaining cultivars with good natural resistance to BGMV led to the development of a genetically modified common bean, based on RNAi technology, which provided excellent resistance.

Resistance to Mungbean Yellow Mosaic Viruses

In black gram (*Vigna mungo*) Mungbean yellow mosaic India virus (MYMIV) resistance is governed by a single recessive gene having NB-ARC domain with normal functional group. A resistant gene (CYR1) for MYMIV has been characterized, similarly its truncated allele (*cyr1*) was also found in susceptible black gram. Moreover, a SNP in LRR-like protein kinase gene responsible for a G/C transversion was identified in soybean. The yellow mosaic disease (YMD) was found to be monogenic. The comparative transcriptome analysis showed an upregulation in NAC transcription factor, Argonaute, NB-LRR, Ankyrin and WRKY33 genes. Full length cDNA of black gram MAPK (Mitogen-activated protein kinases) designated as *VmMAPK1* was found increased during MYMIV infection, which helps in autophosphorylation and restricting the multiplication of MYMIV by mediating salicylic acid signaling pathway. MAPK16 and MAPK3 genes are down regulated during MYMIV infection. Moreover, the G2/mitotic specific cyclin-1-like and Cyclin-B1 (CYCB1) were found up-regulated. MYMIV interacts with proliferative cell nuclear antigen, recombination-dependent replication (RDR51, RDR54), minichromosome maintenance protein subunit 2 (MCM2) and replication protein A, respectively. Micro RNAs (miRNAs) also confer resistance against MYMIV such as *gma-miR5785* found in soybean, that helps in cleavage of BC1 gene of MYMIV. In Mungbean RAPD markers OPP 07₈₉₅ and OPP 07₉₀₀ are associated with Mungbean yellow mosaic virus (MYMV) resistance, while OPP 07₇₃₀ was found in French bean.

A resistance gene analog (RGA) marker has been associated with MYMV resistance in mungbean and black gram. Moreover, the MYMIV-resistant markers YR4 and CYR1 were found in urdbean and mungbean. In black gram the SSR marker (CEDG180) was found closely linked with YMD resistance. The QTLs (qTMIV1 to qYMIV5), qMYMIV2 and qMYMIV were identified for MYMIV resistance in *Vigna radiata*. The SSR marker CEDG044 was found highly linked with MYMIV resistance. The MYMIV resistance gene was mapped on chr 6 (LG C2), which is present 3.5-cM on genomic region between two SSR markers GMAC7L and Satt322. Similarly, in black gram the ISSR8111357 marker was developed and confirmed for MYMV resistance. **Fig. 6** illustrates the screening of soybean germplasm for resistance to Mungbean yellow mosaic virus.

Resistance to Cotton Leaf Curl Virus in Cotton

Livelihoods of many rely on cotton in many countries across the world. Cotton leaf curl viruses (CLCuVs), members of the genus *Begomovirus*, in the family *Geminiviridae*, and their natural vector whitefly (*Bemisia tabaci*) pose serious threat to cotton production. In the Indian subcontinent, whitefly-mediated cotton leaf curl disease (CLCuD) stands a devastating factor that limits overall production of cotton. Multiple strategies based on pathogen-derived resistance have been tried but huge diversity of geminiviruses has resulted in limited success under field conditions. *Gossypium hirsutum*, tetraploid cotton cultivars produce high quality lint and fiber but are susceptible to CLCuD. On the contrary, one of the diploid progenitors of cotton, *G. arboreum* (A genome) is resistant to CLCuD, making it a valued source for novel genes to improve CLCuD resistance. Whole transcriptome analysis of *G. arboreum* under graft-mediated CLCuD infection revealed the genetic basis and underlying mechanism of resistance. The genes related to oxidative stress, transcription factors, R-gene family, phytohormone signaling, membrane transporters and channel proteins were found differentially expressed and may be involved in resistance to CLCuD. Correlation

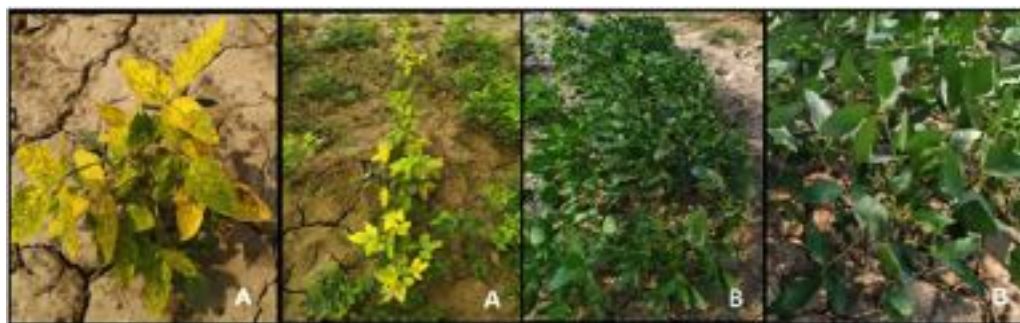


Fig. 6 Mungbean yellow mosaic disease resistant germplasm. The panels (A) show soybean (*Glycine max*) line V-47 showing susceptibility to Mungbean yellow mosaic virus. The panels (B) show soybean line SG-16, which is resistant to Mungbean yellow mosaic virus.



Fig. 7 Cotton germplasm showing resistance to CLCuD. Mac7 plants grown in field condition showing resistance to Cotton leaf curl disease (CLCuD), while all the susceptible varieties grown around are showing typical symptoms of CLCuD. The black panel represents plant of CLCuD resistant Mac7 accession and susceptible varieties grown in glasshouse conditions. On the left side of panel is the closer look of asymptomatic leaf of Mac7 (Upper) and diseased or symptomatic leaf of susceptible (lower) cotton varieties.

of these identified genes with available QTLs (associated with CLCuD resistance) can help in picking the desirable genes further to impart CLCuD resistance in cotton.

Several efforts have been made for the identification of resistance sources among large collections of available cotton germplasm (Fig. 7). Around 5000 cotton USDA germplasm accessions from their different breeding programs were screened against a high pressure of CLCuD. This massive screening revealed a disease resistant accession Mac7.

The understanding of mechanisms underlying the interaction of virus, whitefly and host cotton plant can help in devising robust strategies for controlling the virus and vector (Fig. 8). Therefore, transcriptome analysis of Mac7 under whitefly-mediated CLCuV infestation has shed light on the molecular insights of CLCuD and whitefly resistance in cotton. This study has pinpointed the involvement of heat shock proteins and geminivirus-interacting genes in Mac7 resistance to CLCuD. Virus-induced gene silencing (VIGS) based silencing of HSC80, E3 ligase and serine threonine kinase (STK) genes exhibited increased susceptibility to whitefly and CLCuD in Mac7 cotton. These identified genes in resistant cotton accession could be used as useful sources of resistance to CLCuD and can help in improving the whitefly and virus resistance in cotton germplasm.

Understanding and exploitation of host plant resistance against whitefly can also help in control of geminiviruses. In this context, a study in cotton infested with whitefly showed differential gene expression of WRKY40, GhMPK3 and copper transport protein. Furthermore, VIGS of GhMPK3, validated the suppression of Jasmonic acid (JA) and Ethylene (ET) signaling related genes leading to the whitefly susceptibility. Small RNA-seq and genome-wide miRNA analysis of whitefly resistant and susceptible cultivars of *G. hirsutum* has been done recently, that revealed the involvement of miRNAs related to leucine-rich repeat (LRR) protein, MYB transcription factors and auxin response factor (ARF). VIGS of ARF8 from whitefly resistant cotton targeting miRNA390 enhanced the auxin and JA accumulation that increased the whitefly tolerance. Susceptible cotton variety of *G. hirsutum* when infected with CLCuD viruliferous whiteflies indicated the differential gene expression of genes including transcription factors (NAC, bHLH, MYB), heat shock proteins, methyl-transferases and metabolism related genes like cytochrome p450. Thus the genes identified and highlighted in these studies can be used in developing and enhancing cotton resistance to whitefly.

Now in the era of new plant breeding technologies, CLCuD/whitefly resistant cotton cultivars harboring ample genetic diversity can be achieved using CRISPR/Cas9 system. CRISPR/Cas9 based gene targeting is now becoming a method of choice with high reliability and efficiency for cotton genome editing. It would help in targeting CLCuD related susceptibility genes in cotton to

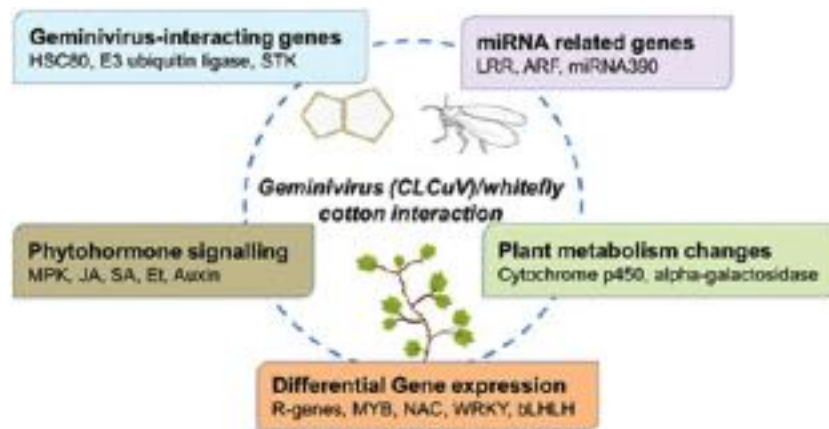


Fig. 8 *Geminivirus (CLCuV)/whitefly – cotton interactions and mechanisms of host plant resistance.* Cotton leaf curl geminivirus disease complex entry into cotton through whitefly mediates different layers of mechanisms in cotton to deal with pest and virus including interaction of geminivirus proteins with plant genes, activation of differential gene expression, miRNA machinery, metabolic changes and phytohormone signaling pathways.



Fig. 9 *Maize streak virus (MSV) induced stunting in a susceptible maize variety Gusau pool 16 (bottom) and resistant variety TZL Composite 4 CL (top).* Source: Taiwo *et al.*, 2006. Plant Disease 90, 199–202.

counter the whitefly and fast evolving viral genomes. Moreover, development of improved resistant crop varieties by bridging the conventional and new genetic approaches along with pest/disease awareness to the farmers, capacity building of researchers would be helpful in controlling the geminiviruses and related diseases effectively.

Resistance to Maize Streak Virus

Maize streak virus (MSV, genus *Mastrevirus*, family *Geminiviridae*) causes maize streak disease (MSD), one of the most important viral diseases of maize in sub-Saharan Africa. MSV largely remains uncontrolled in most of the African continent and during epidemic years it can lead to widespread yield losses and famine. Fig. 9 depicts field performance of a MSV resistant maize variety “TZL composite 4L” along with a MSV susceptible maize variety “Gusau Pool 16” in Nigeria. Resistance to MSV is associated with up to five different QTLs located on chromosome 1, 3, 7, and 9, that are either dominant or additive in nature (Table 1). None of these QTLs in isolation can prevent MSV infections, however combinations of them can resist the MSV infections. Although maize

genotypes that tolerate MSV infection without compromising much with yield losses have been developed, there has been only limited success in the field. The MSV resistance seems to be specific to certain geographic locations, selection made of MSV resistance at one location may fail to tolerate MSV infections in another location. Thus the breeding efforts and screening for MSV resistance have to be specific to different geographical areas or the agro-climatic zones of Africa. One of the major concern is that the improved MSV tolerant maize genotypes don't have the desirable agronomic traits, such as good yields. Thus developing maize genotypes that possess both MSV tolerance and high yields remains a big challenge to the maize breeders. Further involvement of multiple MSV tolerant QTLs makes it difficult to introgress each of them in one single maize genotype and development of such a variety may take several years of breeding. Despite the humungous efforts made in development of MSV resistant maize varieties, most farmers prefer to use high-yielding maize varieties.

Conclusion

The current knowledge on the genes that confer resistance to geminiviruses is scarce when compared to the information available for plant infecting RNA viruses. Although the geminivirus resistant locus/loci are mapped in several cultivated plant species, the actual genes associated with the resistance are yet to be cloned and characterized. The few resistant genes that have been identified, the underlying resistant mechanism is yet to be understood. In this article we have summarized the sources and genes for geminivirus resistance in the most important crop plants that are heavily infested by geminiviruses. The geminivirus resistant genes identified so far are diverse in their characteristics, they are monogenic and polygenic, dominant and recessive, and specific as well as broad spectrum in nature. Introgression of these resistant genes in cultivated crop varieties has not been very successful since some of these genes are associated with poor agronomic traits. Loci or genes conferring resistance to the whitefly vectors are not yet identified and identification of such genes may boost the breeding for begomovirus resistance. Revolution in genome editing technologies and employment of CRISPR-Cas based technologies for precise editing of host genes are promising for manipulating the recessive resistant genes that makes them immune to geminivirus infections. Transgenics by employing CRISPR-Cas technology that destroys the geminiviral DNA genome have been developed, however their durability remains uncertain with reference to the fast mutating geminiviruses which may evade the targeting by sgRNAs that are key for Cas9 based cleavage. However, these are considered as transgenics and their social acceptance is questionable. Thus emphasis on identification and characterization of novel geminiviral resistant genes offers more promise for development of improved geminivirus resistant crop varieties.

Further Reading

- Beebe, S.E., Ochoa, I., Skroch, P., Nienhuis, J., Tivang, J., 1995. Genetic diversity among common bean breeding lines developed for Central America. *Crop Science* 35, 1178–1183.
- Bianchini, A., 1999. Resistance to Bean golden mosaic virus in bean genotypes. *Plant Disease* 83, 615–620.
- Blair, M.M., Rodríguez, L.M., Pedraza, F., Morales, F., Beebe, S., 2007. Genetic mapping of the Bean golden yellow mosaic geminivirus resistance gene *bgm-1* and linkage with potyvirus resistance in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 114, 261–271.
- Chellappan, P., Vanitharani, R., Fauquet, C.M., 2004. Short interfering RNA accumulation correlates with host recovery in DNA virus-infected hosts and gene silencing targets specific viral sequences. *Journal of Virology* 78, 7465–7477.
- Dharajiya, D.T., Ravindrababu, Y., 2019. Identification of molecular marker associated with Mungbean yellow mosaic virus resistance in mungbean [*Vigna radiata* (L.) Wilczek]. *Vegetos* 32, 532–539.
- Ji, Y., Schuster, D.J., Scott, J.W., *et al.*, 2007. Sources of resistance, inheritance, and location of genetic loci conferring resistance to members of the tomato-infecting begomoviruses. In: Czosnek, H. (Ed.), *The Tomato Yellow Leaf Curl Virus Disease: Management, Molecular Biology, Breeding for Resistance*. Dordrecht: Springer, pp. 343–362.
- Kundu, A., Singh, P.K., Dey, A., Ganguli, S., Pal, A., 2019. Complex molecular mechanisms underlying MYMIV-resistance in *Vigna mungo* revealed by comparative transcriptome profiling. *Scientific Reports* 9, 1–13.
- Lapidot, M., Friedmann, M., 2002. Breeding for resistance to whitefly-transmitted geminiviruses. *Annals of Applied Biology* 140, 109–127.
- Li, J., Zhu, L., Hull, J.J., *et al.*, 2016. Transcriptome analysis reveals a comprehensive insect resistance response mechanism in cotton to infestation by the phloem feeding insect *Bemisia tabaci* (whitefly). *Plant Biotechnology Journal* 14, 1956–1975.
- Naqvi, R.Z., Shan-e-Ali Zaidi, S., Akhtar, K.P., *et al.*, 2017. Transcriptomics reveals multiple resistance mechanisms against cotton leaf curl disease in a naturally immune cotton species, *Gossypium arboreum*. *Scientific Reports* 7, 1–15.
- Navas-Castillo, J., Fiallo-Olivé, E., Sánchez-Campos, S., 2011. Emerging virus diseases transmitted by whiteflies. *Annual Review of Phytopathology* 49, 219–248.
- Patil, B.L., 2018. *Genes, Genetics and Transgenics for Virus Resistance in Plants*. Norfolk: Caister Academic Press, doi:10.21775/9781910190814.
- Patil, B.L., Fauquet, C.M., 2011. Chapter 11. Ecology of plant viruses, with special reference to geminiviruses. In: Hurst, C. (Ed.), *Studies in Viral Ecology*. Volume 1. Hoboken, NJ: John Wiley & Sons, Inc, pp. 273–306.
- Patil, B.L., Fauquet, C.M., 2015. Studies on differential behavior of cassava mosaic geminivirus DNA components, symptom recovery patterns, and their siRNA profiles. *Virus Genes* 50 (3), 474–486. doi:10.1007/s11262-015-1184-y.
- Rojas, M.R., Macedo, M.A., Maliano, M.R., *et al.*, 2018. World management of geminiviruses. *Annual Review of Phytopathology* 56, 637–677.
- Singh, N., Mallick, J., Sagolsem, D., Mandal, N., Bhattacharyya, S., 2018. Mapping of molecular markers linked with MYMIV and yield attributing traits in mungbean. *Indian Journal of Genetics* 78, 118–126.
- Strausbaugh, C.A., Eujayl, I.A., Wintermantel, W.M., 2017. Beet curly top virus strains associated with sugar beet in Idaho, Oregon, and a Western U.S. Collection. *Plant Disease* 101, 1373–1382.
- Zaidi, S.S.e.A., Naqvi, R.Z., Asif, M., *et al.*, 2019. Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton (*Gossypium hirsutum*). *Plant Biotechnology Journal* 18, 691–706.

Plant Rhabdoviruses (*Rhabdoviridae*)

Ralf G Dietzgen, The University of Queensland, St. Lucia, QLD, Australia

Michael M Goodin, University of Kentucky, Lexington, KY, United States

Zhenghe Li, Zhejiang University, Hangzhou, China

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A.O. Jackson, R.G. Dietzgen, R.-X. Fang, M.M. Goodin, S.A. Hogenhout, M. Deng, J.N. Bragg, Plant Rhabdoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00492-1](https://doi.org/10.1016/B978-012374410-4.00492-1).

Nomenclature

aa Amino acid(s)

AGO Argonate

ER Endoplasmic reticulum

G Glycoprotein

IGSs Intergenic or “gene-junction” sequences

kb Kilobase

kDa Kilo Dalton

ldr Leader

M Matrix protein

MP Movement protein

N Nucleocapsid protein

NC Condensed nucleocapsid

NE Nuclear envelope

nt Nucleotide(s)

ORF Open reading frame

P Phosphoprotein

RdRp RNA-dependent RNA polymerase

trl Trailer

UTR Untranslated region

Glossary

Chytrid fungi Obligate parasitic root-infecting fungi of the *Chytridiomycetes* that rarely appear to affect the host plants but act as vectors for plant viruses, for example *Olpidium virulentus*.

L protein Main component of the viral polymerase complex, responsible for replication and transcription with RNA-dependent RNA polymerase, mRNA 5′ capping, 3′ polyadenylation and protein kinase activities.

Leader and trailer Regulatory regions at the 3′ and 5′ termini of rhabdovirus genomes that show a high degree of sequence complementarity and are predicted to form a panhandle structure during replication.

Negative-sense RNA viruses Have negative-sense single-stranded RNA as their genetic material. Genomic RNA of negative polarity is complementary to the mRNA that is transcribed from it for translation of encoded proteins. Purified negative-sense RNA by itself is not infectious but requires viral core proteins to commence replication and transcription.

Nucleocapsid core The transcriptionally-competent ribonucleoprotein that represents the minimal infectious unit of rhabdoviruses.

Viroplasm The site of viral RNA replication, transcription and nucleocapsid formation.

Introduction

‘Classic’ plant rhabdoviruses have distinctive enveloped bacilliform or bullet-shaped particles and are distinguished based on whether they replicate and undergo morphogenesis in the cytoplasm or in the nucleus. Consequently, they have been classified in two genera, *Cytorhabdovirus* or *Nucleorhabdovirus*. More than 70 putative plant rhabdoviruses have been described over the years based on their morphology although, in many cases, molecular characterizations necessary for unambiguous classification are incomplete or lacking. More recently, an increasing number of novel plant rhabdovirus genomes have been assembled from high throughput sequencing data of diverse plant and insect species. All sequenced plant rhabdoviruses have the same general genome organization as other members of the *Rhabdoviridae*, and each encodes at least six open reading frames (ORFs), one of which facilitates cell-to-cell movement of the virus. Thus, plant rhabdoviruses have a number of similarities to members of other rhabdovirus genera, but they differ in several respects from rhabdoviruses infecting vertebrates. Two new genera of bisegmented plant viruses were recently classified as rhabdoviruses, the dichorhavirus that infect horticultural and ornamental plant species and are transmitted by *Brevipalpus* mites, and the varicosaviruses that are transmitted by chytrid fungi.

Rhabdoviruses infect plant species from a large number of different families, including numerous weed hosts and several major horticultural and cereal crops. Symptoms of infection vary substantially and range from stunting, vein clearing, mosaic, and mottling of leaf tissue, to tissue necrosis. The most serious pathogens include Maize mosaic virus (MMV), Lettuce necrotic yellows virus (LNYV), Rice yellow stunt virus (RYSV), Rice stripe mosaic virus (RSMV), Eggplant mottled dwarf virus (EMDV), Strawberry crinkle virus (SCV), Potato yellow dwarf virus (PYDV), Barley yellow striate mosaic virus (BYSMV), Northern cereal mosaic virus (NCMV), Orchid fleck virus (OFV), and Coffee ringspot virus (CoRSV). A number of other rhabdoviruses also have disease potential that can be affected by agronomic practices, incorporation of genes for disease resistance, and control of insect vectors.

The transmission of most plant rhabdoviruses is dependent on specific phytophagous insects that support replication of the virus. Therefore, viral prevalence and distribution is influenced to a large extent by the ecology and host preferences of their vectors. Although some rhabdoviruses can be transmitted mechanically by abrasion of leaves, this mode of transmission does not contribute significantly to their natural spread due to the labile nature of the virion. Moreover, seed or pollen transmission of plant rhabdoviruses has not been described; thus, aside from vegetative propagation, direct plant-to-plant transmission is unlikely to be a major factor in the ecology or epidemiology of these pathogens.

This article focuses on recent findings concerning the taxonomy, structure, replication, and vector relationships of the expanded group of plant rhabdoviruses. More extensive aspects of plant rhabdovirus biology, specifically ecology, disease development and control, can be found in earlier reviews.

Taxonomy and Classification

Plant rhabdoviruses are taxonomically classified in the order *Mononegavirales*, family *Rhabdoviridae*. The International Committee on Taxonomy of Viruses (ICTV) in its 9th Report (2011) listed two genera containing plant rhabdovirus species, *Cytorhabdovirus* and *Nucleorhabdovirus*, the members of which have unsegmented genomes and are discriminated by their replication and morphogenesis in the cytoplasm and nucleus, respectively of infected cells. For the 10th Report (2018), the genera *Dichorhavirus* and *Varicosavirus* were added to the family, based on phylogenetic relatedness of its members to the rhabdoviruses. The genus *Dichorhavirus* was newly created, whereas *Varicosavirus* had not previously been assigned to a family. Dichorhaviruses are most closely related to nucleorhabdoviruses, but have a bisegmented genome and virions do not appear to be enveloped. Varicosaviruses will not be discussed further in this chapter.

In 2019, in the family *Rhabdoviridae*, out of the 19 recognized genera 4 contain plant viruses. The ICTV has listed eleven accepted plant virus species in the genus *Cytorhabdovirus*, ten in the genus *Nucleorhabdovirus*, five in the genus *Dichorhavirus* and one in the genus *Varicosavirus*, a total of 27 plant virus species out of 289. Phylogenetic analyses of these rhabdoviruses have confirmed their taxonomic classification, although it has become evident that following inclusion of the dichorhaviruses, the genus *Nucleorhabdovirus* no longer has a monophyletic origin (Fig. 1) and will need to be split. The genomes of more than 20 novel plant rhabdoviruses have been recently identified by high throughput sequencing and await official classification. However, many plant rhabdoviruses that were identified before next generation sequencing technologies were available have not been investigated in much detail beyond cursory infectivity studies, crude physicochemical analyses of virus particles, and electron microscopic observations of morphogenesis. Consequently, more than 70 putative rhabdoviruses await assignment to a species and genus once genome sequences can be obtained. Viruses characterized in greater detail, with coding sequences deposited in GenBank, that have approved or pending taxonomic assignments are listed in Table 1.

Particle Morphology and Composition

In comparison to rhabdoviruses restricted to vertebrate hosts, plant-adapted rhabdoviruses are typically bacilliform in shape (Fig. 2). The modal lengths of particles, measured by electron microscopy of thin sections, range from 45 to 100 nm wide by 130–350 nm in length. Suggestive of packaging genomic RNAs in separate particles, particles of dichorhaviruses and varicosaviruses tend to be approximately half the length of rhabdovirus particles. While dichorhavirus particles lack a lipid envelope, those of the majority of rhabdoviruses consist of an outer layer of host-derived membranes, the lipid composition of which is commensurate with that of the membrane environment where morphogenesis occurs, be it on the nuclear envelope (NE) or endoplasmic reticulum (ER) (Fig. 2(b)). Protruding from the envelope is the spike glycoprotein (G) that is distributed more or less uniformly around the virion (Fig. 2(a)). Serving to both stabilize the particle and connect the G-containing envelope to the condensed nucleocapsid (NC) is the matrix protein (M). However, the biological functions of M may extend far beyond that of its structural role. The NC contained within the virion is a ribonucleoprotein complex composed of single-stranded negative-sense RNA, encapsidated over its entire length by the nucleocapsid protein (N) (Fig. 2(a)). Associated with this complex are small amounts of the viral phosphoprotein (P) and large (L) RNA-dependent RNA polymerase (RdRp). This “core” NC is transcriptionally competent for synthesizing viral mRNAs once it gains entry to cells compatible for replication. Collectively, virions of plant rhabdoviruses have a composition that is approximately 70% protein, 2% RNA, 20%–25% lipid, with a minimal fraction of carbohydrate associated with the G protein. Genomic RNAs of plant rhabdoviruses tend to range from 10 to 15 kb, while those of dichorhaviruses, and varicosaviruses are 5–6 kb for each segment.

Genome Structure and Organization

The general genomic structure of plant rhabdoviruses is conserved in a manner similar to that of the prototypical rhabdoviruses like Vesicular stomatitis virus (VSV) or Rabies virus. As such, the protein-encoding open reading frames (ORFs) are flanked by short untranslated leader (ldr), and untranslated trailer (trl) regions, at the 3' and 5' ends of the genomic RNA. These regions show a high degree of sequence complementarity (Fig. 3(a)) and are thought to potentially form a panhandle structure of the genome during

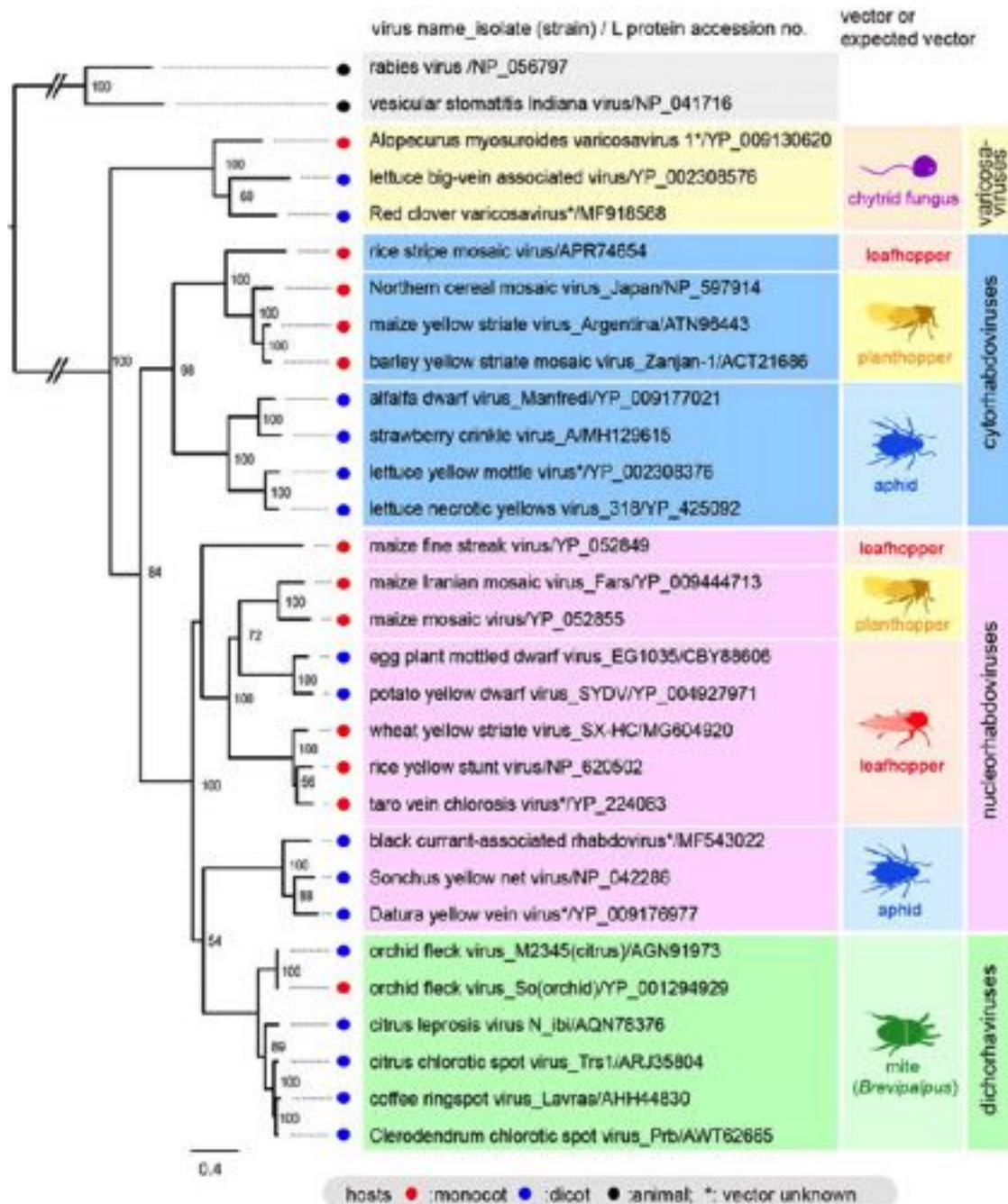


Fig. 1 Maximum likelihood (ML) phylogeny of plant rhabdoviruses. L protein amino acid sequences of selected plant rhabdoviruses (cytorhabdoviruses, nucleorhabdoviruses, dichorhaviruses, and varicosaviruses) and two well-known animal rhabdoviruses (as outgroup) were multiple aligned using MAFFT version 7. Subsequently, ambiguously aligned regions were trimmed with Gblocks. A ML tree was constructed based on the sequence alignment using PhyML 3.0 with the best fit model LG with + I + G + F. The types of plant host and arthropod vectors are differentiated by circles filled with different colors (monocot and dicot) or shaded cartoons, respectively. An asterisk following the virus name indicates that the vector has not yet been determined. Numbers at the nodes indicate bootstrap values > 50%. Reprinted with permission from Whitfield, A.E., Huot, O.B., Martin, K.M., Kondo, H., Dietzgen, R.G., 2018. Plant rhabdoviruses – Their origins and vector interactions. *Current Opinion in Virology* 33, 198–207, Copyright 2018, Elsevier B.V.

replication. Located between each ORF are so-called intergenic or “gene-junction” sequences (IGSs) that provide cis-regulatory elements for transcription and replication. IGSs have a modular structure composed of three elements, (1) Module 1; a poly(U) tract at the 3′ end of each gene located after the un-translated region (UTR) of the preceding ORF, (2) Module 2; a di- to poly-nucleotide spacer typically starting with a guanyl residue that is not templated in mRNAs, (3) Module 3; a transcriptional “start” sequence that is typically a tri- to penta-nucleotide sequence found in each of the viral mRNA transcripts (Fig. 3(b)).

Table 1 Characteristics of plant rhabdoviruses, members and/or tentative members of 4 genera of the family *Rhabdoviridae*. Only accessions of complete or coding-complete genome sequences are provided. Type species are in bold

Genus/Species	Acronym	Genbank accession		Host plant	Arthropod vector
Cytorhabdovirus					
<i>Alfalfa dwarf cytorhabdovirus</i>	ADV	KP205452		Dicot	Aphid?
<i>Barley yellow striate mosaic cytorhabdovirus</i>	BYSMV	KM213865		Monocot	Planthopper
<i>Broccoli necrotic yellows cytorhabdovirus</i>	BNYV			Dicot	Aphid
<i>Colocasia bobone disease-associated cytorhabdovirus</i>	CBDaV	KT381973		Monocot	Planthopper?
<i>Festuca leaf streak cytorhabdovirus</i>	FLSV			Monocot	Planthopper
<i>Lettuce necrotic yellows cytorhabdovirus</i>	LNyV	AJ867584		Dicot	Aphid
<i>Lettuce yellow mottle cytorhabdovirus</i>	LYMoV	EF687738		Dicot	Unknown
<i>Northern cereal mosaic cytorhabdovirus</i>	NCMV	AB030277	GU985153	Monocot	Planthopper
<i>Sonchus cytorhabdovirus</i>	SonV			Dicot	Unknown
<i>Strawberry crinkle cytorhabdovirus</i>	SCV	MH129615		Dicot	Aphid
<i>Wheat American striate mosaic cytorhabdovirus</i>	WASMV			Monocot	Leafhopper
Tentative species					
Bean-associated cytorhabdovirus	BaCV	MK202584		Dicot	Unknown
Cabbage cytorhabdovirus 1	CCyV – 1	KY810772		Dicot	Unknown
Maize-associated cytorhabdovirus	MaCV	KY965147		Monocot	Unknown
Maize yellow striate virus	MYSV	KY884303		Monocot	Unknown
Papaya virus E	PpVE	MH282832		Dicot	Unknown
Persimmon virus A	PeVA	AB735628		Dicot	Unknown
Raspberry vein chlorosis virus	RVCV	MK240091	MK257717	Dicot	Aphid
Rice stripe mosaic virus	RSMV	KX525586		Monocot	Leafhopper
Strawberry cytorhabdovirus 1	StrV – 1	MK211271		Dicot	Aphid
Tomato yellow mottle-associated virus	TYMaV	KY075646		Dicot	Aphid?
Trifolium pratense cytorhabdovirus A	TpVA	MH982250		Dicot	Unknown
Trifolium pretense cytorhabdovirus B	TpVB	MH982249		Dicot	Unknown
Yerba mate chlorosis-associated virus	YmCaV	KY366322		Dicot	Unknown
Nucleorhabdovirus					
<i>Datura yellow vein nucleorhabdovirus</i>	DYVV	KM823531		Dicot	Unknown
<i>Eggplant mottled dwarf nucleorhabdovirus</i>	EMDV	KC905081		Dicot	Leafhopper
<i>Maize fine steak nucleorhabdovirus</i>	MFSV	AY618417		Monocot	Leafhopper
<i>Maize Iranian mosaic nucleorhabdovirus</i>	MIMV	MF102281		Monocot	Planthopper
<i>Maize mosaic nucleorhabdovirus</i>	MMV	MK828539		Monocot	Planthopper
<i>Potato yellow dwarf nucleorhabdovirus</i>	PYDV	GU734660	KY549567	Dicot	Leafhopper
<i>Rice yellow stunt nucleorhabdovirus</i>	RYSV	AB011257	AB516283	Monocot	Leafhopper
<i>Sonchus yellow net nucleorhabdovirus</i>	SYNV	L32603		Dicots	Aphid
<i>Sow thistle yellow vein nucleorhabdovirus</i>	SYVV			Dicot	Aphid
<i>Taro vein chlorosis nucleorhabdovirus</i>	TaVCV	AY674964		Monocot	Unknown
Tentative species					
Alfalfa-associated nucleorhabdovirus	AaNV	MG948563		Dicot	Unknown
Apple rootstock virus A	ApRVA	MH778545		Dicot	Unknown
Bird's-foot trefoil-associated virus 1	BFTV – 1	MH614262		Dicot	Unknown
Black currant-associated rhabdovirus 1	BCaRV – 1	MF543022		Dicot	Unknown
Green Sichuan pepper nucleorhabdovirus 1	GSPNuV	MH323437		Dicot	Unknown
Morogoro maize-associated virus	MMaV	MK112501		Monocot	Unknown
Physostegia chlorotic mottle virus	PCMV	KX636164		Dicot	Unknown
Wheat yellow striate virus	WYSV	MG604920		Monocot	Leafhopper
Dichorhavirus		RNA1	RNA2		
<i>Citrus chlorotic spot dichorhavirus</i>	CiCSV	KY700685	KY700686	Dicot	<i>Brevipalpus</i> mites
<i>Citrus leprosis dichorhavirus N</i>	CiLV-N	KX982176	KX982179	Dicot	<i>Brevipalpus</i> mites
<i>Clerodendrum chlorotic spot dichorhavirus</i>	CiCSV	MG983790	MG983791	Dicot	<i>Brevipalpus</i> mites
<i>Coffee ringspot dichorhavirus</i>	CoRSV	KF812525	KF812526	Dicot	<i>Brevipalpus</i> mites
<i>Orchid fleck dichorhavirus</i>	OFV	AB244417	AB244418	Dicot	<i>Brevipalpus</i> mites
Varicosavirus					
<i>Lettuce big-vein associated varicosavirus</i>	LBVaV	AB075039	AB114138	Dicot	<i>Olpidium</i> fungi

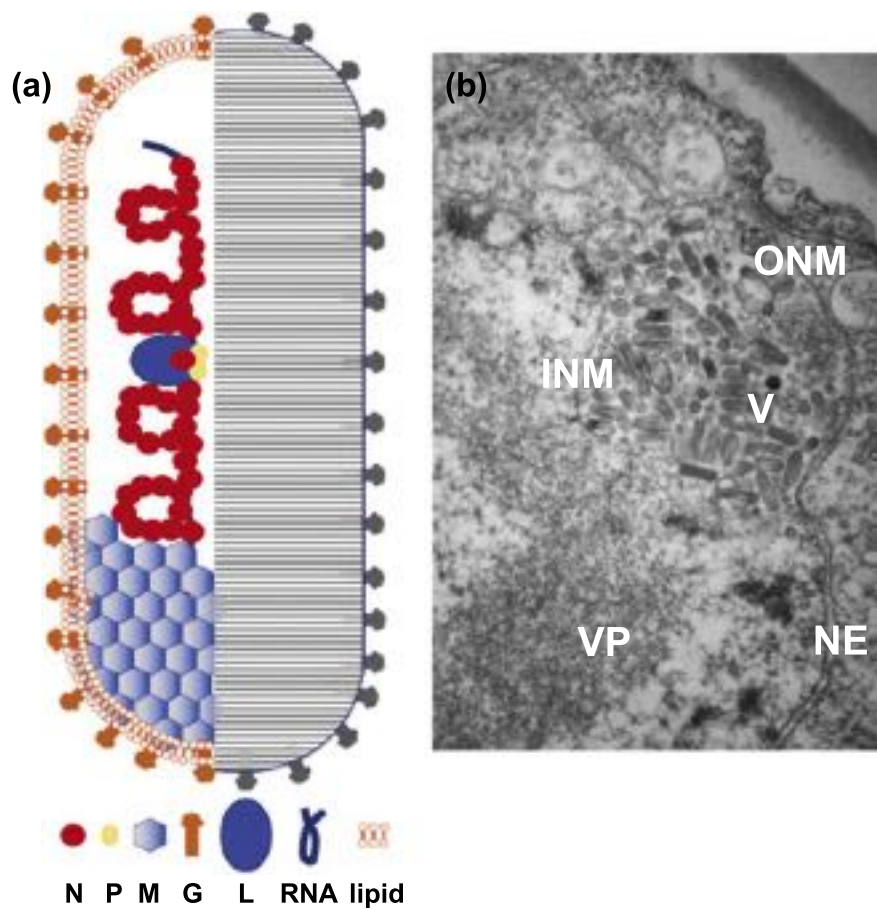


Fig. 2 Virion structure and morphology of enveloped plant rhabdoviruses. (a) Stylized virion structure of an enveloped, unsegmented plant rhabdovirus. The left side of the figure shows the approximate orientation and shape of the glycoprotein (G) and lipid bilayer envelope (orange). The matrix protein (M; blue hexagons) connects the envelope to the condensed nucleocapsid core, which is composed of the genomic-length single-stranded RNA, encapsidated by the nucleocapsid protein (N; red circles). Associated with this RNA-N complex are the phosphoprotein (P; yellow circles), and the large (L) RNA-dependent RNA polymerase (blue circle). (b) Transmission electron micrograph of a thin-section of sonchus yellow net virus-infected leaf tissue of *Nicotiana benthamiana*. Note accumulation of bacilliform shaped virus particles (V) in the perinuclear space between the outer (ONM) and inner (INM) membranes of the nuclear envelope (NE). Nucleocapsids are condensed in the viroplasm (VP) that is highly enriched with P protein prior to migrating to the NE where morphogenesis takes place. EM image courtesy of Dr. Tea Meulia, Molecular and Cellular Imaging Center, The Ohio State University (<https://mcic.osu.edu/home>).

(a)

PYDV 3' -UGGUUGGGUUCUAUGGAACUUAACCCGUAU-5' -ldr
 5' -ACCAACCCAUCGAAAGCUCUGUACUCCUG-3' -trl

LNYV 3' -AAUGCCUGUUUUUAUCUUCUUUUUUAGUU-5' -ldr
 5' -ACGGACGAUAAUAAAUCAAAAAGUCCA-3' -trl

OFV (RNA1) 3' -UGUGUCCUAUUGGGCGAUAAACCAGAAGAUC-5' -ldr
 5' -ACACCGAGCCAUGAGUACUGUUGUAAAUA-3' -trl

(b)

	Module 1	Module 2	Module 3
OFV/RNA1	UAAAAUUUUUU	GUUG	UU
PYDV	UUUUUUUU	GGG	UUG
LNYV	AUUCUUUU	G (N) _n	CUA

Fig. 3 Intergenic and terminal noncoding regions of plant rhabdovirus genomes. (a) Complementary sequences at the 3' and 5' termini of plant rhabdovirus genomic RNAs; complementary nucleotides are shown in bold font. Dichorhavirus OFV, Orchid fleck virus; nucleorhabdovirus PYDV, Potato yellow dwarf virus; cytorhabdovirus LNYV, Lettuce necrotic yellows virus. (b) Conserved modular intergenic sequences separating each gene.

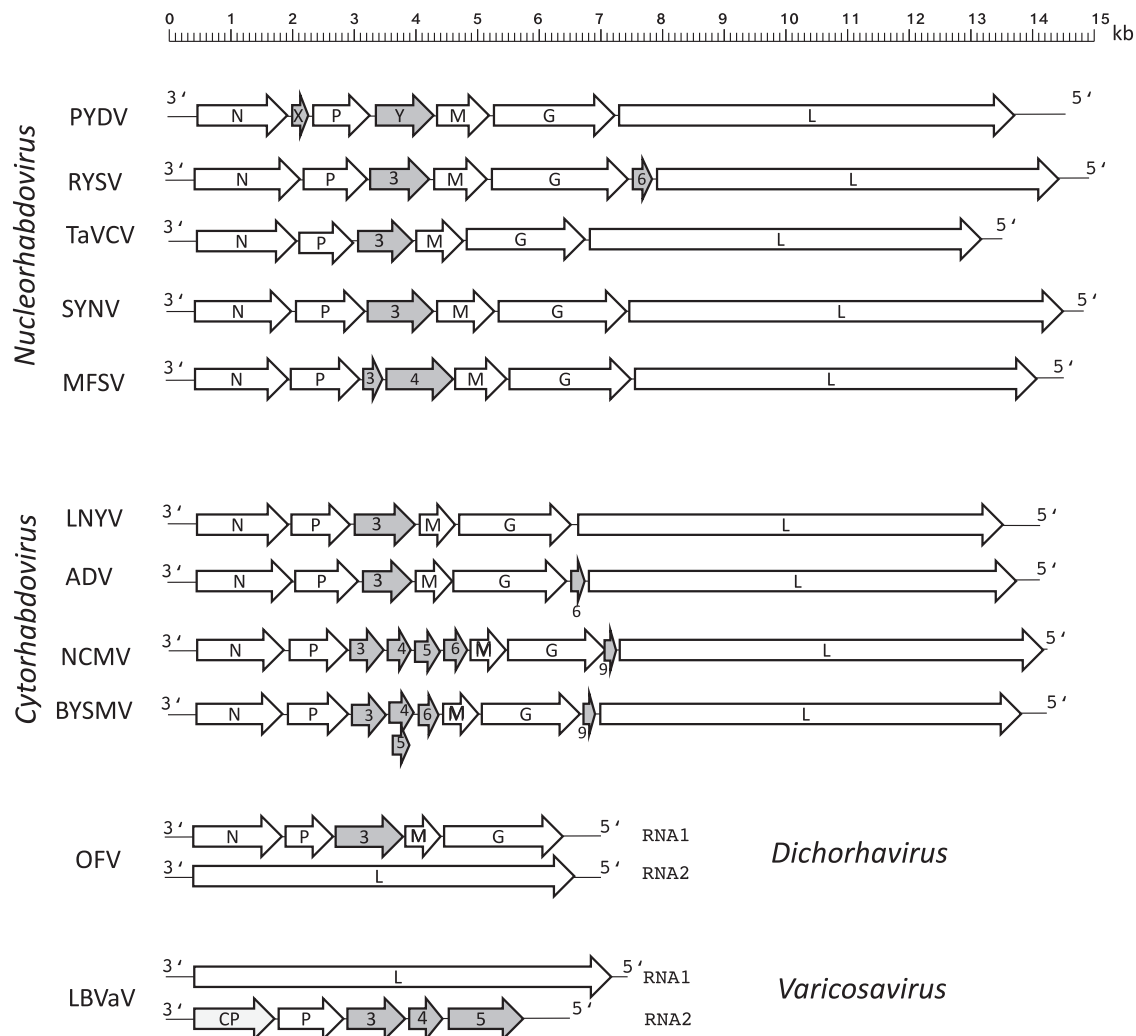


Fig. 4 Schematic diagrams of the genome organization of representative plant rhabdoviruses classified in the genera *Nucleorhabdovirus*, *Cytorhabdovirus*, *Dichorhavirus*, and *Varicosavirus*. See [Table 1](#) for full virus names. Each open reading frame (ORF) is indicated by an arrow labelled with the name of the encoded protein (N = nucleocapsid protein; P = phosphoprotein; M = matrix protein; G = glycoprotein; L = Large polymerase). Accessory genes are shaded in grey and numbered. In some cases, ORF 3 has been shown to encode a cell-to-cell movement protein.

In all genomes characterized to date the N ORF follows immediately after the leader, and the large (L) RdRp ORF precedes the trailer region. Situated between these ORFs are three additional protein coding regions conserved in all rhabdoviruses, namely the P protein, M protein, and G protein. Interspersed between these conserved ORFs are additional coding regions peculiar to plant-adapted viruses generally, or unique to a particular virus. These regions, typically designated as 'P3' or 'P6' are located in genomes in the general order 3'-Ldr-N-P-P3-M-G-P6-L-trl-5' ([Fig. 4](#)). The number of ORFs between the P and M genes may be variable. Overall, the plant rhabdovirus genomes analyzed so far contain between 6 and 10 ORFs ([Fig. 4](#)). Proteins encoded in position P3 ([Fig. 4](#)) have long been predicted, and recently experimentally demonstrated, to be required for the cell-to-cell movement functions essential for all plant viruses.

The general gene order and genome structure of plant rhabdoviruses are also conserved in the dichorhviruses and varicosaviruses. Key distinctions are that the N to G ORFs are encoded on RNA1 in dichorhviruses, and on RNA2 in varicosaviruses, while the cognate L polymerases are encoded on RNA2 and RNA1, respectively ([Fig. 4](#)).

Properties of Viral Proteins

Comprehensive biochemical, molecular, subcellular localization, and protein-protein interaction analyses of plant rhabdovirus-encoded proteins have been carried out mostly with the *Sonchus yellow net virus* (SYNV), PYDV, LNYV, and OFV models. Putative functions have been assigned to proteins encoded by less-well studied plant rhabdoviruses on the basis of their genomic locations and sequence relatedness to plant and animal model rhabdoviruses. Overall, amongst the five structural proteins of plant

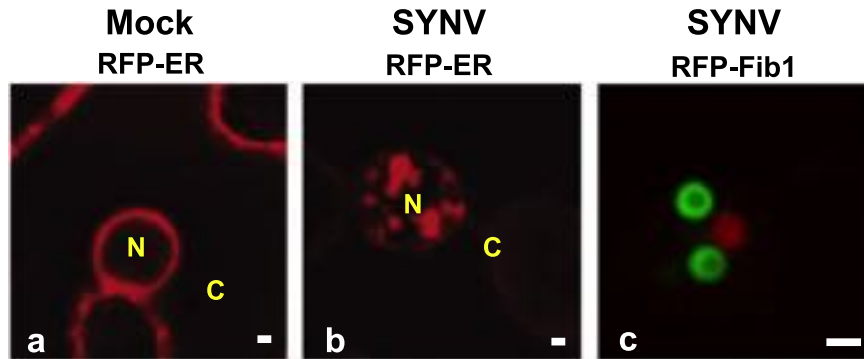


Fig. 5 Confocal micrographs of leaf epidermal cells of transgenic *N. benthamiana* expressing RFP targeted to endomembranes. (a) in the absence of virus, RFP accumulates in the envelope of the nucleus (N). (b) Infection by the nucleorhabdovirus SYN V results in intranuclear accumulation of the inner nuclear membrane. The resulting expansion of perinuclear volume fills with the RFP-ER marker. In the image planes shown here, the reticulate pattern of the ER is not visible in the cytoplasm (C). (c) Within SYN V-infected nuclei, nucleocapsid formation takes place in viroplasm “rings” that are highly enriched in P protein (expressed here as a fusion to green fluorescent protein, in transgenic plants expressing the nucleolar marker protein RFP-Fibrillarin 1). Scale bar = 2 μ m.

rhabdoviruses, the L protein has conserved polymerase motifs common to all mononegaviruses, whereas the N, P, M, and G proteins are less conserved and have only limited sequence relatedness to analogous proteins of animal rhabdoviruses.

The Nucleocapsid Protein (N)

The N protein functions to encapsidate the entire viral RNA genome and tightly packages it to form NC that are RNase-resistant (Fig. 2(a)). VSV N protein has a tendency to form large aggregates and binds cellular RNAs non-specifically. Interactions of P protein with free N protein form a soluble complex, designated N(0)-P, which selectively encapsidates viral RNA. The availability of N(0)-P complexes regulates a critical switch from transcription to replication in the viral infection cycle, since genome replication entails encapsidation of nascent RNAs and only the encapsidated form of the genome can serve as template for viral polymerase. Biochemical experiments have shown that the solubility of SYN V N protein is increased upon co-expression and interactions with the P protein, so the SYN V P protein appears to have a similar chaperone function to keep the N protein in a biologically active form.

The N protein is also a major component of the viroplasm, the site of viral RNA replication, transcription and NC formation (Fig. 5). Confocal microscopic studies have revealed the subcellular localization of N proteins of selected cyto- and nucleorhabdoviruses consistent with their role in viroplasm formation. For instance, the N protein of the cytorhabdovirus LYNV localizes to the cytoplasm, whereas the homologs from the nucleorhabdoviruses SYN V, PYDV, MMV, Maize fine streak virus (MFSV), and RYSV are targeted to subnuclear foci. When the N proteins were co-expressed with their respective cognate P proteins, the N-P complexes co-localized to distinct cytoplasmic (in the case of LYNV) or subnuclear (SYN V, PYDV, OFV, etc.) inclusions characteristic of viroplasms, which are often different from the localization of either protein alone. In SYN V, formation of these subnuclear foci requires self-interactions of the N protein that are mediated by a helix-loop-helix motif near the amino (N)-terminus, as well as interactions of the P protein with the N-terminus of the N protein. However, heterologous combinations of SYN V and MFSV N and P proteins fail to form subnuclear foci. These co-localization studies suggest that cognate N-N and N-P protein interactions are essential for viroplasm formation.

The N proteins of nucleorhabdoviruses and dichorhavirus contain a nuclear localization signal (NLS) for nuclear import. Experiments conducted in plant and yeast cells have shown that SYN V and PYDV N proteins contain a bipartite NLS near the carboxy (C)-terminus that is required for nuclear localization, and biochemical studies have shown that the N proteins interact *in vitro* with importin α homologs. Related bipartite NLS sequences are also present in the C-terminus of MFSV and RYSV N proteins, while the analogous regions of the MMV and Taro vein chlorosis virus N proteins contain a monopartite NLS.

The Phosphoprotein (P)

In addition to the chaperone role in maintaining the N protein in soluble form, rhabdovirus P protein is also a co-factor of the L polymerase and a component of the viral NC core (Fig. 2(a)). The SYN V P protein forms complexes *in vivo* with the N and L proteins that are analogous to N-P and P-L complexes found in VSV-infected cells, and also engages in self-interactions that are mediated by the N-terminus of the P protein. Likewise, pairwise N-P and P-P interactions have also been documented for several other plant rhabdoviruses including LYNV, PYDV, and OFV. Rhabdoviral P proteins are highly phosphorylated as their name suggests, and phosphorylation of specific residues play important roles in regulating the P protein activities in transcription and

replication. Direct evidence showing phosphorylation of plant rhabdovirus P protein is available for SYN. In SYN, the P protein is phosphorylated *in vivo* at threonine residues, which differs from the VSV P protein that is phosphorylated at serine residues. Hence, the P proteins of plant rhabdoviruses probably have replication-related functions similar to their well-characterized vertebrate counterparts.

Sequence analyses show that the P proteins of the nucleorhabdoviruses PYDV, RYSV, and MMV carry a NLS, which is consistent with their pronounced nuclear localization patterns. Although similar NLS sequences have not been identified for the SYN and MFSV P proteins, these proteins localize to the nucleus as well as the cytoplasm when transiently expressed in plant cells. Furthermore, the SYN P protein binds directly to human importin β derivatives *in vitro* and the central third of the P protein is required for nuclear import, but other regions of the protein affect the import efficiency. However, co-expression with cognate N proteins relocates the SYN and MFSV P proteins to subnuclear foci exclusively, suggesting that the P proteins may contain a nuclear export signal whose function is masked by the N protein-binding.

Apart from functioning in replication, the P proteins of two cytorhabdoviruses have been shown to exhibit relative weak, yet evident RNA silencing suppressor activities. LYNV P suppresses local RNA silencing and delays systemic silencing. Further mechanistic studies have shown that LYNV P inhibits microRNA-mediated mRNA slicing and transitive RNA silencing likely by interactions with Argonaute (AGO) proteins and the RdRp 6/ Suppressor of Gene Silencing 3 (RDR6/SGS3) complex. Similarly, Alfalfa dwarf virus (ADV) P suppresses local silencing weakly but systemic silencing strongly. ADV P interacts with AGO1 and AGO4, inhibits miRNA-guided AGO1 cleavage and prevents transitive RNA silencing.

The P3 Movement Protein

Like other plant viruses, plant rhabdoviruses encode proteins to assist in cell-to-cell movement of viral infection entities through the plasmodesmata and their systemic transport through the vascular system. Considerable evidence for a movement protein (MP) function has been accumulated for proteins encoded by the gene located between the P and M genes designated P3, whose name varies with viruses, including SYN sc4, LNYV 4b, MFSV P4, PYDV Y, and P3 in most other viruses (Fig. 4). The predicted secondary structures of several plant rhabdovirus P3 homologs have a distant relatedness to structural motifs of the Tobacco mosaic virus 30K superfamily of plant viral MP. Additional evidence for a role of the P3 homologs in cell-to-cell movement is their localizations to the cell periphery or the plasmodesmata and associations with host membranes during transient expression. The movement function for the P3 proteins has been reinforced by *trans*-complementation experiments, in which P3 homologs encoded by various cyto-, and nucleorhabdoviruses, and dichorhviruses have been shown to support cell-to-cell movement of MP-defective positive-strand RNA viruses. In addition, a recombinant SYN sc4 deletion mutant is unable to move from initially infected cells to adjacent tissues, but the movement defect is rescued by transient expression of the sc4 protein *in trans*. Protein interaction experiments revealed that the P3 homologs from SYN, PYDV, RYSV, and Tomato yellow mottle-associated virus (TYMaV) interact with their respective N and/or P proteins, and in the case of SYN sc4, the specific sc4-N and sc4-P interactions appear to facilitate inter- and intracellular NC trafficking. Together, these lines of evidence collectively support a MP role for the plant rhabdovirus P3 proteins.

Protein localization and cellular fractionation assays indicate that several rhabdovirus P3 proteins are membrane-associated, although membrane protein prediction algorithms generally fail to identify any transmembrane regions in these proteins. In addition, SYN sc4 protein was found to be present in purified virions, and after sucrose gradient fractionation, sc4 co-sedimented with the G protein-containing membrane fractions of the virion preparations. Similarly, immunogold labelling has shown that Rice transitory yellowing virus (a strain of RYSV) P3 is also associated with virus particles, but in this case the P3 appears to localize to the interior of the virion in the NC. Thus, these findings provide evidence that plant rhabdovirus MP can also function as structural proteins, but the actual structural role remains obscure.

The Matrix Protein (M)

Within the rhabdovirus particle, the M protein forms a layer connecting the helical NC core and the transmembrane G protein (Fig. 2(a)). During infections, the M protein plays critical roles in virus budding by interacting with the NC core, membrane lipid and the G protein. Studies with vertebrate rhabdoviruses have shown that late in replication, the M protein interacts with genomic NC and directs the condensed core to membrane budding sites that are enriched in viral G protein. Similar functions for plant rhabdovirus M protein are supported by protein localization and interactions studies. The N-M interactions appears to be common among several plant rhabdoviruses including SYN and PYDV. Consistent with the morphogenesis sites of cyto- and nucleorhabdoviruses, the LYNV M protein localizes to the cytoplasm near the cell periphery, but the SYN and PYDV M proteins are targeted to the nucleus. Notably, when transiently expressed in uninfected cells, the SYN and PYDV M proteins distribute uniformly throughout the nucleoplasm; however, in virus-infected cells, the M proteins appear as subnuclear foci bridging the periphery of the inner nuclear membrane (INM) and the viroplasm. These foci likely represent virion maturation sites where coiled NC cores bud through the nuclear membranes.

As with VSV M, the PYDV and SYN M proteins also interact with and remodel host membranes. Infection by SYN, and to a lesser extent by PYDV, results in the formation of intranuclear spherules derived from invagination of the INM (Fig. 5). The spherule formation is also induced when PYDV M is ectopically expressed by itself, and requires a di-leucine motif located at aa

residues 223 and 224 of the M protein. In contrast, expression of SYN V M is insufficient to trigger nuclear membrane rearrangement, and even infection by a recombinant SYN V M deletion mutant fails to induce intranuclear spherule formation. The SYN V M protein requires interactions with the G protein's C-terminal tail to associate with INM and induce membrane remodelling.

In addition to their roles in viral morphogenesis, preliminary experiments indicate that rhabdovirus M proteins have important roles in host–virus interactions because they appear to be able to inhibit host gene expression. The nucleorhabdovirus M protein may also play a role in NC egress from the nucleus. SYN V M protein forms motile foci liberating from the nuclear envelope and moving towards the cytoplasm, whereas PYDV M contains a putative nuclear export signal that may be involved in transport of NC cores into the cytoplasm.

The Glycoprotein (G)

The G protein forms the glycoprotein spikes of rhabdovirus virions (Fig. 2(a)). The plant rhabdovirus G proteins are typical type I membrane proteins, consisting of a signal peptide sequence of ~20 aa, an N-terminal luminal (or virion surface) domain that contains potential glycosylation sites, a transmembrane region followed by a short C-terminal tail. The signal peptide is cleaved by signal peptidases during co-translational translocation to the ER lumen to yield a “mature” transmembrane G protein. In addition, the SYN V G protein contains a putative NLS near the C-terminus that could be involved in transit to the inner nuclear membrane prior to virion morphogenesis. Several glycosylation inhibitors interfere with N-glycosylation of the SYN V G protein and also block SYN V morphogenesis. A prominent role of G protein in morphogenesis is also supported by SYN V reverse genetics studies. A SYN V G protein deletion mutant is abortive in virion maturation, leading to accumulation of striking arrays of naked NC cores that fail to bud through the INM. The C-terminal tail of SYN V G protein is critical for morphogenesis, since this domain is required for relocalization of the M protein to INM surfaces and subsequently membrane remodelling. The N-terminal domain of SYN V G also contributes to efficient budding, but is dispensable for systemic infections.

Although SYN V G protein is not absolutely required for infections of *Nicotiana benthamiana* plant host, a more critical role can be envisioned during infections of insect vectors by analogy to animal rhabdoviruses. The plant rhabdovirus G protein may be crucial for binding to proposed insect midgut receptors and host cell entry. Indeed, antibody blockage or enzymatic digestion of PYDV G protein drastically reduces virus infectivity in insect vector cells.

The Polymerase Protein (L)

The L proteins are the most closely related of the rhabdovirus-encoded proteins and contain conserved mononegavirus polymerase domains and RNA-binding motifs. Alignment of the L protein sequences of non-segmented negative-strand RNA viruses reveals conservation within 12 motifs. Phylogenetic trees derived from L protein alignments indicate that the nucleorhabdoviruses, dichorhavirus and cytorhabdoviruses cluster together in two major clades separated from the vertebrate rhabdoviruses (Fig. 1).

The L proteins of plant rhabdoviruses are present in low abundance within virus particles and in infected cells (Fig. 2(a)). A viral RdRp is activated after treatment of virions of the cytorhabdoviruses LNYV and Broccoli necrotic yellows virus with mild nonionic detergents, and this activity co-sediments with loosely coiled NC filaments that are released from virions. The transcribed products are complementary to the genome, as expected of mRNAs. Thus, the described polymerases of these plant cytorhabdoviruses appear to be similar to the extensively studied polymerases of the vesiculoviruses. In contrast, no appreciable polymerase activity is evident in dissociated preparations of SYN V or other nucleorhabdovirus virions that have been analyzed. However, an active polymerase can be recovered from the nuclei of plants infected with SYN V. Polymerase activity is associated with a complex consisting of the N, P, and L proteins that co-sediments with SYN V NC cores. The polymerase complex can be precipitated with P protein antibodies, but the activity of the complex is not inhibited by these antibodies. However, antibody inhibition experiments demonstrate that the L protein is required for polymerase activity. Kinetic analysis of transcription products also reveals that the complex is capable of sequentially transcribing a polyadenylated plus-sense leader RNA and polyadenylated mRNAs corresponding to each of the six SYN V-encoded proteins. Potential replication intermediates consisting of short incomplete minus-strand products homologous to the genomic RNA are also transcribed. These results thus provide a model whereby nucleorhabdovirus particles require polymerase activation by host components early in infection. In contrast, the polymerases of the cytorhabdoviruses appear to be present in an active form in virions and the released cores are capable of initiating primary transcription immediately upon uncoating *in vitro*.

Other Accessory Proteins

Besides the canonical genes, plant rhabdovirus genomes may also contain 1–4 additional ORFs encoding putative accessory proteins (Fig. 4). For example, an X ORF is present in PYDV and EMDV between the N and P genes, and within the P and M genes, there are 4 ORFs in the two cereal cytorhabdoviruses, *i.e.*, NCMV and BYSMV, instead of only one ORF (P3) in most of the other plant rhabdoviruses. In addition, a short transcriptional unit was found in several cyto- and nucleorhabdoviruses, designated P6 in RYSV, RSMV, ADV, and SCV, and P9 in NCMV and BYSMV. Small ORFs preceding the L gene were also found in the genomes of some animal rhabdoviruses, such as ephemero-, hapa-, and tibroviruses. These ORFs encode small, highly hydrophobic, and basic proteins named viroporins, which function to modify the cell membrane permeability to ions or other small molecules and are

involved in virion assembly and release. The plant rhabdovirus P6 and P9 proteins do not share obvious sequence identity with each other or with the viroporins of animal rhabdoviruses, but some share a characteristic transmembrane domain.

Only very limited information is available regarding the functions of plant rhabdovirus accessory genes. RYSV P6 protein is present in virion preparations purified from infected rice plants and in protein samples from viruliferous leafhoppers, suggesting that P6 has a structural role in infection. RYSV P6 was also shown to suppress systemic RNA silencing by affecting RDR6-mediated secondary siRNA amplification, but did not affect local RNA silencing. However, ADV P6 does not appear to have similar RNA silencing suppressor activities, suggesting that this function is not conserved amongst plant rhabdoviruses.

Cytopathology and Replication

Entry

Little information is available about early entry and uncoating events during plant infections. It is hypothesized that the rigid plant cell wall is breached through insect vector feeding or experimentally by mechanical abrasion, and that the introduced enveloped virions may fuse to cellular membranes to release NC cores into the cytoplasm, followed by M protein disassembly to liberate the flexuous NC filaments. However, during insect vector infections, the surface G proteins of plant rhabdoviruses presumably play a role in binding to proposed midgut receptors that mediate virion uptake into epithelial cells by endocytosis, although such receptors have so far not been identified.

Transcription and Genome Replication

The fundamental replication and transcription processes of plant rhabdoviruses are similar to their animal-infecting counterparts. Cytorhabdoviruses closely resemble animal rhabdoviruses in that both replicate in the cytoplasm, whereas nucleorhabdoviruses and dichorhavirus have exclusive nuclear replication phases (**Fig. 6**). The latter two groups of viruses utilize the host nucleocytoplasmic transport machinery to recognize karyophylic signals present on the N and P proteins to facilitate nuclear targeting of viral core proteins and NC. The plant rhabdovirus replication cycles described below were developed largely by analogy to the VSV model and supported by biochemical characterization of viral polymerase fractions, analyses of viral mRNAs, electron microscopic observations, protein localization and interaction assays, and recent reverse genetics studies.

After NC release from virions, primary rounds of transcription are initiated by using the negative-stranded genome template to produce capped and polyadenylated leader RNA and mRNAs for each of the viral genes. Transcription occurs progressively and sequentially from the genomic 3' end, yielding discrete mRNA species in a decreasing molar gradient, *i.e.*, the mRNA levels follow the order $N > P > P3 > M > G > L$. This sequential and polar transcription is mediated by interactions of the viral polymerase complex with *cis*-regulatory elements located in each gene junction. As primary transcription proceeds, newly synthesized viral proteins increase in abundance. Accumulation of the N and P proteins triggers a switch of the polymerase function from mRNA transcription to genome replication. Upon replication, the viral polymerase initiates antigenomic RNA synthesis at the exact 3' terminus of the genomic RNA and ignores all internal regulatory signals located in the gene junctions. During genome replication, the nascent RNAs are encapsidated by the N-P complex to form antigenomic NC, which then serve as template of the P-L replicase complexes for synthesis of genomic RNAs that are subsequently encapsidated to form genomic NC. Continuous NC formation depletes free N, P, and L proteins, and lower core protein abundances signal a transition from replication to secondary rounds of mRNA transcription to increase the pool of viral mRNAs for replenishment of viral proteins. This cyclic phase of replication and transcription is coordinated by a feedback mechanism that measures the abundance of the core proteins by the levels of nascent leader RNA encapsidation.

Later in cytorhabdovirus replication, accumulation of viral core proteins and progeny NC result in formation of thread-like viroplasms in the cytoplasm that are located in close proximity to dense networks of the ER. Likewise, replication of nucleorhabdoviruses and dichorhavirus induces formation of subnuclear viroplasms in greatly enlarged nuclei that consist of large masses of granular material containing viral RNA and the N, P, and L proteins (**Fig. 2(b)**).

Morphogenesis

The plant rhabdoviruses vary profoundly in their sites of morphogenesis and differ substantially from vertebrate rhabdoviruses that undergo virion maturation at the host plasma membrane. In plant cells, cytorhabdoviruses bud through ER membranes, and mature particles accumulated in ER-associated cytoplasmic cisternae that are greatly proliferated during virus infections. In contrast, the budding of nucleorhabdoviruses, such as SYN and PYDV, occurs in association with the INM, and enveloped virions accumulate in enlarged perinuclear spaces formed between the inner and outer nuclear envelopes (**Fig. 2(b)**; **Fig. 5**). However, dichorhavirus that also replicate in the nucleus do not have an enveloped form of virions. Rather, rod-like non-enveloped particles are present in nuclei of infected cells, either scattered through the nucleoplasm, forming crystalline arrays or arranged perpendicularly onto the INM. Occasionally, these particles are also present in the cytoplasm and appear to be enwrapped by membranous tubular structures. In this case, the virions are arranged radially and perpendicularly to the membrane and form a "spoke wheel" configuration.

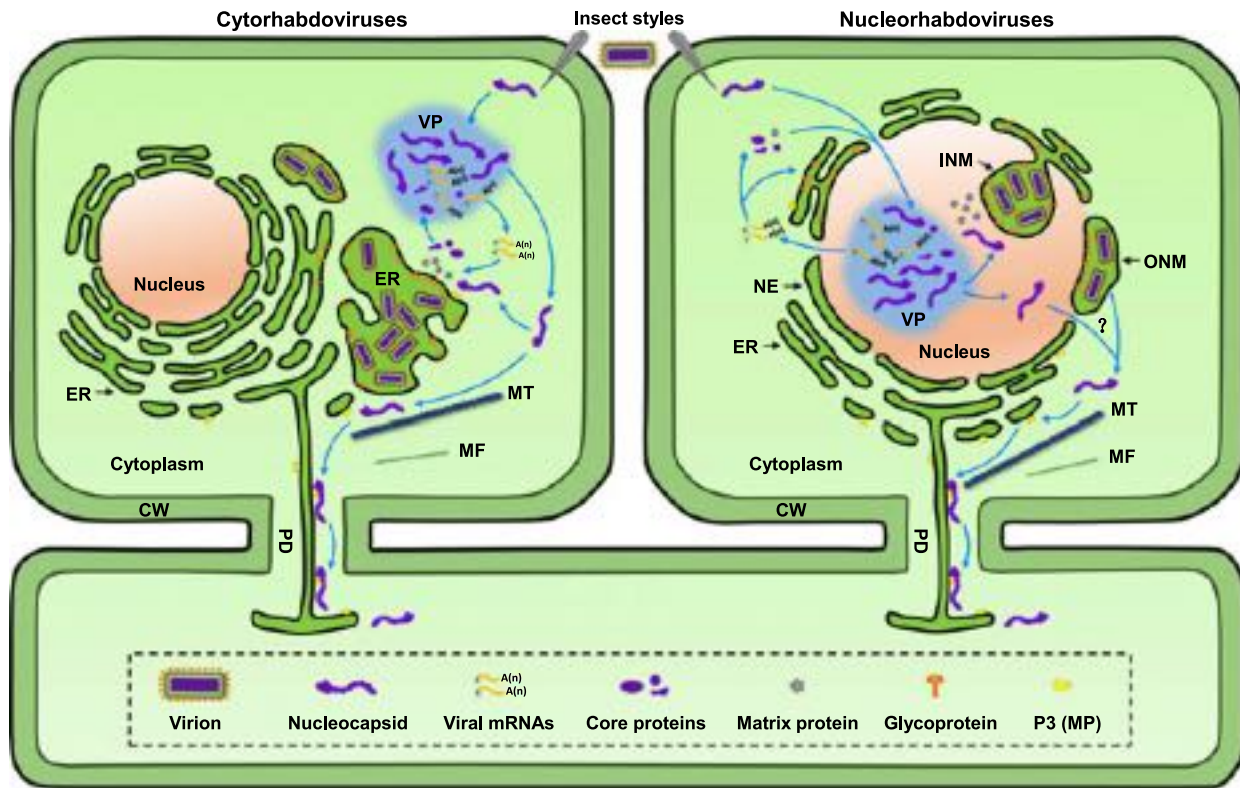


Fig. 6 Models for cyto- and nucleorhabdovirus infection cycles in plant cells. Plant rhabdoviruses are initially introduced into plant cells through mechanical breach of the cell wall (CW) during insect vector feeding. Upon entry into the cell, viral nucleocapsids (NC) are liberated from the virions and replication takes place in the cytoplasm (cytorhabdoviruses) or nucleus (nucleorhabdoviruses). For cytorhabdoviruses, the NC initiate primary rounds of transcription in the cytoplasm to yield capped, polyadenylated mRNAs for translation of each viral protein. Primary transcription is followed by genome replication to produce viral antigenomic and genomic RNAs that are encapsidated by the newly synthesized nucleocapsid proteins to form progeny NC, and then by secondary rounds of transcription to increase the pool of mRNAs. As cyclic waves of transcription and replication proceed, accumulation of viral core proteins and NC form cytoplasmic viroplasms (VP). During the later stages of replication, matrix (M) proteins coil NC and likely interact with ER-anchored viral glycoproteins, and morphogenesis occurs by budding of condensed NC into the proliferated ER membranes. Matured bacilliform virions accumulate in pronounced ER vesicles. The replication cycles of nucleorhabdoviruses are fundamentally similar to those of cytorhabdoviruses, but their NC are transported into the nucleus for transcription and replication cycles. The viral mRNAs are exported to the cytoplasm and translated, and the core proteins and M protein are then transported through the nuclear pore complex into the nucleus. As replication progresses, subnuclear VP appear and the nuclei become greatly enlarged. Morphogenesis occurs by budding of coiled NC into inner nuclear membrane (INM), and enveloped virions accumulate between INM and outer nuclear membrane (ONM) of the nuclear envelope (NE). During cell-to-cell movement, the NC exit the nuclei either by a nuclear export activity of a viral core protein or by a “bud-out” process of the enveloped virions. Once in the cytoplasm, the NC of cyto- and nucleorhabdoviruses interact with membrane-associated P3 movement proteins (MP) in a species-specific manner. The MP then direct intracellular NC trafficking to the cell periphery and intercellular movement through gated plasmodesmata (PD) across the CW by mechanisms that may involve microtubule (MT)- and microfilament (MF)-associated host proteins.

Morphogenesis of cyto- and nucleorhabdoviruses occurs during the late stages of replication in which the M protein accumulates and reaches concentrations sufficient to initiate coiling of the genomic NC. The M protein condenses the NC into a compact bacilliform core, which leads to downregulation of RNA synthesis and a transition from replication to virion assembly. The coiled cores are brought to the vicinity of endomembrane surfaces where the transmembrane G proteins are enriched, and then bud through the membranes to acquire lipid envelope and the G proteins. Infections with SYN or PYDV induce massive INM invagination into the nucleoplasm, forming intranuclear spherule-like structures (Fig. 5). This structural rearrangement may function to bring the INM surfaces in close proximity to viroplasms and to provide readily accessible sites for virion maturation.

Cell-To-Cell Movement

A MP function has been assigned to the P3 accessory gene located between the P and M genes for several cyto- and nucleorhabdoviruses, as well as dichorhavirus, since the P3 homologs of these diverse plant rhabdoviruses display cross-family movement *trans*-complementation with positive-strand RNA viruses. Reverse genetics analyses using recombinant SYN

further reinforce that the sc4 protein, a P3 homolog, is essential for SYNIV cell-to-cell movement. The plant rhabdovirus movement complex has been postulated to involve viral NC, since NC constitutes the minimal infectious unit. This hypothesis has been verified by SYNIV reverse genetics studies, which show that SYNIV minigenomes comprising the N, P, sc4, and L genes are capable of autonomous localized movement. Reverse genetics analyses have also revealed a highly specific cell-to-cell movement mechanism for cyto- and nucleorhabdoviruses. The movement defects of MP-deletion mutants of SYNIV and TYMaV can only be rescued by their respective cognate MP, but not by MP from related rhabdoviruses, distantly related or unrelated plant RNA viruses. The movement specificities of SYNIV and TYMaV are explained by the species-specific interactions between NC core proteins and MP.

Available evidence suggests that during replication, some of the newly synthesized NC in the cytoplasm interact specifically with membrane-anchored MP, which facilitate intracellular NC trafficking to the cell periphery, followed by intercellular transport through the plasmodesmata (Fig. 6). Furthermore, NC transport may also involve host cytoskeleton components. Microtubule-associated factors have been identified to interact with MP of SYNIV and LNYV, and actin filaments are required for BYSMV P protein trafficking (Fig. 6). For nucleorhabdoviruses and dichorhavirus that employ nuclear phases of replication, their NC have to be exported from the nuclear replication sites to the cytoplasm before cell-to-cell movement can occur. How the NC nuclear egress is achieved has not been resolved, but this process may involve nuclear exporting function of a viral NC core protein or development of intraluminal virions.

Vector Relationships, Distribution and Evolution

Phylogenetic studies support the general hypothesis that natural host ranges of plant rhabdoviruses are determined by their arthropod vectors more so than by host plant species (Fig. 1). Moreover, where studied, plant rhabdoviruses replicate in cells of their insect vectors as well as in their plant hosts. Taken together, the adaptive evolutionary trajectory of plant rhabdoviruses is likely to be from insects to plants. Plant rhabdoviruses are most frequently transmitted by aphids (Aphididae), leafhoppers (Cicadellidae), or planthoppers (Delphacidae), while dichorhavirus are transmitted by false spider mites (*Brevipalpus* sp.) (Table 1, Fig. 1). The relationship between virus and vector appears to be tightly regulated, given that closely related rhabdoviruses are transmitted by different insect species. This is exemplified by the sanguinolenta and constricta strains of PYDV that are distinguished by their differential transmission by the leafhopper vectors, *Aceratagallia sanguinolenta* and *Agallia constricta*. This relationship occurs in cytorhabdoviruses as well, where Soybean blotchy mosaic virus, for example, has been shown to replicate and persist in its vector, the leafhopper, *Peragallia caboverdensis*. In some cases, plant rhabdoviruses have no known vector, however, phylogenetic clustering of viruses that group according to vector can inform prediction of what type of insect maybe involved in the epidemiology of such viruses (Fig. 1). Importantly, while the virus-insect relationship is constrained tightly, the same rhabdovirus may occur naturally in multiple plant host species. For example, whereas PYDV was once agronomically important on potato, it still occurs in reservoirs of red clover (*Trifolium incarnatum*). Although hampered by the inability to mechanically transmit many plant rhabdoviruses, experimental host ranges have been shown to be substantially broader than those reported from field data. However, in most cases the reservoir species are unknown for this group of viruses, which limits our understanding of their true ecology. Unbiased detection methods, such as RNA-Seq, in wild plant communities are likely to provide significant insight into the population dynamics of rhabdoviruses, and related plant viruses.

Once acquired by the insect vector, it appears that cyto- and nucleorhabdoviruses are persistently transmitted in a propagative manner and can be transmitted to vector progeny, in those cases where this aspect of rhabdovirus biology has been examined. Typical of such relationships, long latent periods (hours to days) are required for virus acquisition before vectors are competent for transmission. Insects remain viruliferous throughout their lives, and transovarial passage through eggs and nymphs has been determined experimentally in some cases. Evidence for tissue tropism of rhabdoviruses is derived from studies with MMV in its planthopper vector *Peregrinus maidis*. Following acquisition from plant tissues, MMV accumulates in the epithelial cells of the anterior midgut, followed by infection of nerve cells. Subsequent infection of tracheal cells, hemocytes, muscles, salivary glands, fat cells, mycetocytes, and, finally, epidermal tissues follows, thus completing the circulative infection cycle. Delivery of MMV by microinjection of *P. maidis* results in efficient virus infection and transmission, facilitating functional virus-vector interaction studies in combination with RNA silencing.

Akin to cyto- and nucleorhabdoviruses, following long periods of acquisition (up to two weeks), dichorhavirus are transmitted in a circulative-propagative manner by *Brevipalpus* mites. Furthermore, transovarial transmission has been reported for a citrus leprosis-associated dichorhavirus. However, to-date vertical transmission of dichorhavirus has not been confirmed. Cytological studies have demonstrated the presence of both viroplasm and particle formation in mites that have fed on plants infected with several dichorhavirus including OFV, Citrus leprosis virus N, Clerodendrum chlorotic spot virus and CoRSV (Fig. 7), suggesting that the ability to replicate in vector tissues is conserved within this group of viruses. As with cyto- and nucleorhabdoviruses, tropism for specific cell types is evident with dichorhavirus, and cells from midgut epithelium and other cell types were shown to have similar cytopathology to that observed in infected plant cells.

There have only been limited studies on the population structures of rhabdoviruses but certain generalities are evident such as the low degree of sequence divergence within populations. Genetic variability studies have been conducted for LNYV, type member of the genus *Cytorhabdovirus*, using the N gene of isolates sampled from both Australia and New Zealand, as well as analyzing symptom expression on *Nicotiana glutinosa*. Sequence analyses supported separation of LNYV isolates into two

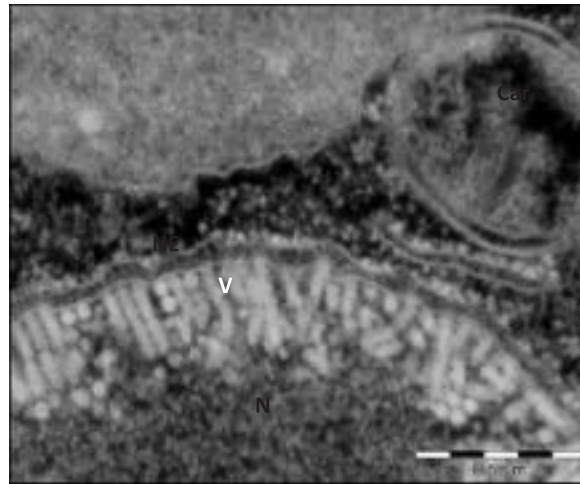


Fig. 7 Transmission electron micrograph of a thin section of the midgut of the mite *Brevipalpus phoenicis* collected from bleeding heart (*Clerodendrum x speciosum*) infected by Clerodendrum chlorotic spot virus. Rod-like particles (V) can be seen in the periphery of the nucleus (N), some arranged perpendicularly onto the nuclear envelope (NE). Nearby, a cell of the symbiotic bacterium *Candidatus Cardinium* (Car) can be seen. Courtesy of Prof. E.W. Kitajima, University of São Paulo, Brazil.

subgroups (I and II), with subgroup I predicted to be an older lineage than subgroup II. Interestingly, these subgroups could not be differentiated based on symptom expression in the experimental host *N. glutinosa*. Sub-group I has apparently gone extinct in Australia, presumably due to its less-efficient dispersal in plant reservoir and/or vectors. Similarly, the dichorhavirus CoRSV was found at one hundred percent of locations ($n=45$) sampled in an area spanning 75% of the coffee growing region of Brazil. All cultivars, regardless of cherry color or growth habit, were found to serve as hosts, suggesting that there is limited resistance in commercially deployed germplasm. Phylogenetic analysis based on the N gene identified a strong geo-spatial relationship among isolates, which clustered into three clades, although the overall sequence diversity between isolates was very low. In contrast to LNYV, and despite low genetic diversity among isolates, variation in symptom expression of CoRSV infection was observed in the experimental host *Chenopodium quinoa*. Based on these studies the hypothesis that the spread of CoRSV is constrained by the clonal expansion of thelytokous populations of *B. papayensis* has been proposed. Further studies, particularly in reservoir host species, once identified, are required to determine if population structure observed in the agronomic hosts holds across a broader range of hosts. While the distribution of most dichorhviruses appears to be limited to South America, OFV has a global distribution.

Further Reading

- Dietzgen, R.G., Freitas-Astúa, J., Chabi-Jesus, C., *et al.*, 2018. Dichorhviruses in their host plants and mite vectors. *Advances in Virus Research* 102, 119–148.
- Dietzgen, R.G., Kondo, H., Goodin, M.M., Kurath, G., Vasilakis, N., 2017. The family *Rhabdoviridae*: Mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Research* 227, 158–170.
- Fang, X.-D., Yan, T., Gao, Q., *et al.*, 2019. A cytorhabdovirus phosphoprotein forms mobile inclusions trafficked on the actin/ER network for viral RNA synthesis. *Journal of Experimental Botany* 70, 4049–4062. doi:10.1093/jxb/erz195.
- Gao, Q., Xu, W.-Y., Yan, T., *et al.*, 2019. Rescue of a plant cytorhabdovirus as versatile expression platforms for planthopper and cereal genomic studies. *New Phytologist* 223, 2120–2133. doi:10.1111/nph.15889.
- Goodin, M.M., Chakrabarty, R., Yelton, S., *et al.*, 2007. Membrane and protein dynamics in live plant nuclei infected with Sonchus yellow net virus, a plant-adapted rhabdovirus. *Journal of General Virology* 88, 1810–1820.
- Jackson, A.O., Dietzgen, R.G., Goodin, M.M., Li, Z., 2018. Development of model systems for plant rhabdovirus research. *Advances in Virus Research* 102, 23–57.
- Jang, C., Wang, R., Wells, J., *et al.*, 2017. Genome sequence variation in the constricta strain dramatically alters the protein interaction and localization map of *Potato yellow dwarf virus*. *Journal of General Virology* 98, 1526–1536.
- Mann, K.S., Beijerman, N., Johnson, K.N., Dietzgen, R.G., 2016. Cytorhabdovirus P3 genes encode 30K-like cell-to-cell movement proteins. *Virology* 489, 20–33.
- Mann, K.S., Johnson, K.N., Carroll, B.J., Dietzgen, R.G., 2016. Cytorhabdovirus P protein suppresses RISC-mediated cleavage and RNA silencing amplification *in planta*. *Virology* 490, 27–40.
- Martin, K.M., Whitfield, A.E., 2018. Cellular localization and interactions of nucleorhabdovirus proteins are conserved between insect and plant cells. *Virology* 523, 6–14.
- Meulia, T., Stewart, L., Goodin, M.M., 2018. Sonchus yellow net virus core particles form on ring-like nuclear structure enriched in viral phosphoprotein. *Virus Research* 258, 64–67.
- Sun, K., Zhou, X., Lin, W., *et al.*, 2018. Matrix-glycoprotein interactions required for budding of a plant nucleorhabdovirus and induction of inner nuclear membrane invagination. *Molecular Plant Pathology* 19, 2288–2301.
- Wang, Q., Ma, X., Qian, S., *et al.*, 2015. Rescue of a plant negative-strand RNA virus from cloned cDNA: Insights into enveloped plant virus movement and morphogenesis. *PLoS Pathogens* 11, e1005223.
- Whitfield, A.E., Huot, O.B., Martin, K.M., Kondo, H., Dietzgen, R.G., 2018. Plant rhabdoviruses—their origins and vector interactions. *Current Opinion in Virology* 33, 198–207.
- Zhou, X., Lin, W., Sun, K., *et al.*, 2019. Specificity of plant rhabdovirus cell-to-cell movement. *Journal of Virology* 93. e00296-19. doi:10.1128/JVI.00296-19.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/rhabdoviridae

Family: Rhabdoviridae

Mononegavirales

ICTV.

Plant Satellite Viruses (*Albetovirus*, *Aumaivirus*, *Papanivirus*, *Virtovirus*)

Mart Krupovic, Archaeal Virology Unit, Institut Pasteur, Paris, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

nt Nucleotide(s)

ORF Open reading frame

satRNA Satellite RNA

ssRNA Single-stranded ribonucleic acid

UTR Untranslated region

Glossary

Albetovirus *Al-* for alphacrovirus [helper virus], *be-* for betanecrovirus [helper virus], *to-* for tobacco.

Aumaivirus *Au-* for aureusvirus [helper virus], *mai-* for maize.

Hyperparasite A parasite whose host (in this article, a helper virus) is also a parasite.

Papanivirus *Pa-* for panicovirus [helper virus], *pani-* for panicum.

Satellite RNA A subviral RNA genome dependent on a helper virus for both replication and encapsidation.

Satellite virus A polyphyletic group of viruses encoding structural components of their virions, but incapable of completing the infection cycle without the assistance of a helper virus.

Virtovirus *Vir-* for virgavirus [helper virus], *to-* for tobacco.

Introduction

Viruses span an entire range of morphological, genomic, and functional complexity. Some viruses are exceedingly complex, surpassing many unicellular organisms in terms of physical dimensions and genome sizes. For instance, members of the family *Mimiviridae* carry megabase-sized genomes and encode over 1000 proteins, including many molecular machineries required for virus multiplication. On the other side of the complexity spectrum are viruses carrying just one or two genes, suggesting a stronger reliance on the host compared to the bigger viruses. Notably, some viruses depend not only on the host cells but also on other viruses for reproduction.

The phenomenon whereby one virus depends for its propagation on another virus has been first described in a plant virus system in the early 1960s. Certain preparations of tobacco necrosis viruses (TNV; genera *Alphanecrovirus* and *Betanecrovirus*, family *Tombusviridae*) contained two types of spherical particles that differed in size and antigenic properties. The larger, TNV, particles could propagate autonomously, whereas the smaller ones were unable to replicate in the absence of TNV. The smaller particles became known as particles of satellite tobacco necrosis virus (STNV), whereas TNV is referred to as STNV's helper virus. Subsequently, several other plant viruses having features similar to those of STNV have been discovered. Furthermore, viruses which depend on another virus for reproduction have been described in protists and animals. Overall, satellite viruses vary considerably in terms of complexity and evolutionary origins. Also, the functions supplemented by the helper viruses vary from one virus system to another. For instance, satellite viruses of plants rely on helper viruses for genome replication, whereas members of the *Lavidaviridae* depend on giant mimiviruses for transcription, clearly indicating that satellite viruses have evolved on several independent occasions.

The tripartite relationship between satellite viruses, helper viruses and host cells is an example of hyperparasitism, which is not uncommon in complex natural ecosystems. Indeed, there are other genetic elements which parasitize viruses, most notably, including satellite nucleic acids. The key difference between satellite nucleic acids and satellite viruses is that the latter necessarily encode components of their virions, whereas the former do not. Instead, satellite nucleic acids commonly encode proteins responsible for the replication of the cognate nucleic acid, whereas the components of the virion for horizontal transfer between the hosts are being provided by the helper viruses.

Due to their dependency on other, quasi-autonomous viruses for reproduction, most satellite viruses, with a notable exception of members of the genera *Dependoparvovirus* and *Deltavirus*, have been for a long time considered as sub-viral agents and not officially classified by the International Committee on Taxonomy of Viruses (ICTV). However, in 2016, the arbitrary distinction between satellite viruses and other viruses has been abandoned and satellite viruses were incorporated into the ICTV classification scheme, resulting in establishment of several genera and families. Plant satellite viruses have been classified into genera *Albetovirus*, *Aumaivirus*, *Papanivirus*, and *Virtovirus*; family *Sarothroviridae* has been created for classification of positive-sense (+)RNA satellite viruses that infect arthropods, and *Lavidaviridae* includes double-stranded DNA viruses replicating in protist cells and parasitizing giant viruses of the family *Mimiviridae*.

This article describes the diversity and properties of plant satellite viruses belonging to the genera *Albetovirus*, *Aumaivirus*, *Papanivirus* and *Virtovirus*.

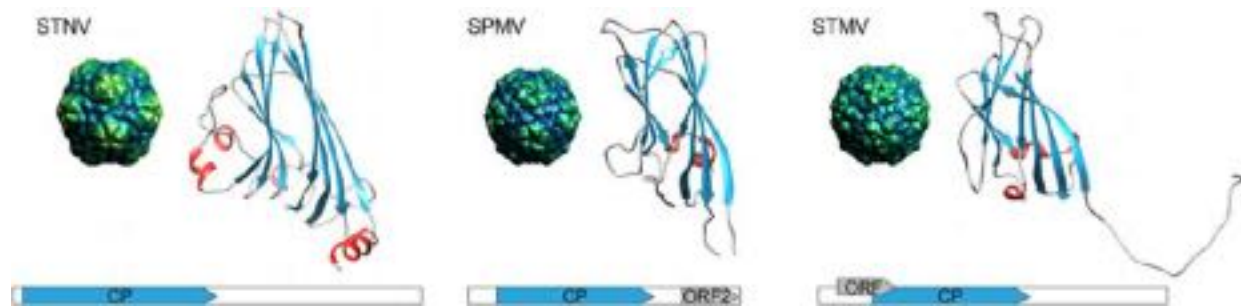


Fig. 1 Plant satellite viruses. Structural similarity between the virions and jelly-roll capsid proteins of satellite tobacco necrosis virus (STNV; PDB ID: 2BUK), satellite panicum mosaic virus (SPMV; PDB ID: 1STM) and satellite tobacco mosaic virus (STMV; PDB ID: 4QQ8). All three virions have $T = 1$ icosahedral symmetry. Images of the depicted virions were downloaded from the VIPERdb database (viperd.b.scripps.edu/). Genome maps of the three viruses are shown at the bottom with the gene encoding the capsid protein shown in blue.

Albetoviruses: Viruses Related to Satellite Tobacco Necrosis Virus

STNV is one of the most extensively studied satellite viruses. Over the years, many properties of this virus, including its genome sequence, virion structure and mechanism of capsid assembly, as well as details of the interaction with the helper virus, have been elucidated. The linear positive-sense (+)RNA genome of STNV consists of 1239 nt and encodes a single jelly-roll-fold capsid protein (CP) of 195 aa, which is necessary and sufficient for formation of icosahedral ($T = 1$) virions (Fig. 1). Virion assembly proceeds cooperatively via interactions between the packaging signals, degenerated stem-loop structures distributed throughout the genome, and multiple CP copies.

In the natural habitat, STNV is transmitted through the soil the same way as its helper virus, i.e., by zoospores of a plant-pathogenic fungus *Olpidium brassicae*, suggesting that both viruses share certain transmission determinants. For both TNV and STNV, the 5' ends of the genomes are phosphorylated and lack a 7-methylguanylate cap or a genome-linked protein, whereas the 3'-termini lack a polyadenylation sequence. Several *cis*-acting elements located within the 5' and 3' untranslated regions (UTRs) are responsible for efficient translation and replication of the STNV genome. The 3' and 5' UTRs of STNV and TNV can be exchanged without abolishing RNA accumulation, although the translation elements at the 3'-terminal regions appear to be unrelated in the two viruses.

Three serotypes of STNV have been described: STNV-1 (or STNV), STNV-2, and STNV-C (Table 1). Different STNV strains are activated by different helper viruses. The replication of STNV-1 and STNV-2 is supported by isolates of TNV-A, the sole member of the *Tobacco necrosis virus A* species (genus *Alphanecrovirus*), whereas TNV-D, the sole member of the *Tobacco necrosis virus D* species (genus *Betanecrovirus*), supports the replication of STNV-C. Genome sequences and organizations for the three STNV strains are overall similar (Table 1). CPs are $\approx 50\%$ – 63% identical in sequence. Whereas the 5' UTRs of STNV-1 and STNV-2 are 30 nt in length and are nearly identical, the corresponding region in STNV-C is considerably longer (101 nt). Similarly, whereas the 3' UTRs of STNV-1 and STNV-2 approximately 64% similar to each other, they share only 40% and 38% identity to that of STNV-C.

STNV-like viruses STNV-1, STNV-2, and STNV-C are classified into three species, *Tobacco albetovirus 1*, 2, and 3, respectively, within a genus *Albetovirus* (sigil: *Al-* for *al*phanecrovirus [helper virus], *be-* for *be*tanecrovirus [helper virus], *to-* for *to*bacco).

Aumavirus: Satellite Maize White Line Mosaic Virus

The fourth, more divergent member of the STNV-like virus group is satellite maize white line mosaic virus (SMWLMV; Table 1). SMWLMV depends on maize white line mosaic virus (MWLMV; species *Maize white line mosaic virus*, genus *Aureusvirus*, family *Tombusviridae*) for multiplication. MWLMV can infect maize in the absence of SMWLMV, whereas the same is not true for SMWLMV, which replicates in maize only when co-inoculated with MWLMV. The ssRNA genome of SMWLMV is 1168 nt in length and encodes one CP (Table 1). Similar to STNV-like viruses, the SMWLMV virion is 17 nm in diameter. SMWLMV CP displays 32% identity to the CP of STNV-1, strongly suggesting that SMWLMV and the STNV-like viruses have diverged from a common ancestor. However, due to lower sequence similarity between the CPs of STNV-like viruses and SMWLMV, the latter virus has been classified into a species, *Maize aumavirus 1*, within a separate genus *Aumavirus* (sigil: *Au-* for *au*reusvirus [helper virus], *mai-* for *mai*ze).

Papaniviruses: Viruses Related to Satellite Panicum Mosaic Virus

Satellite panicum mosaic virus (SPMV) depends on Panicum mosaic virus (PMV; species *Panicum mosaic virus*, genus *Panicovirus*, family *Tombusviridae*) for replication as well as systemic spread in plants. The 5'-terminus of the SPMV genome is phosphorylated and lacks a 7-methylguanylate cap. Several secondary structure elements implicated in the replication of the SPMV genome were

Table 1 General properties of plant satellite viruses

Satellite virus	Species	Helper virus	Accession #	Genome, nt	Capsid Ø, nm	Comments
<i>Albetovirus</i>						
Satellite tobacco necrosis virus (STNV-1)	<i>Tobacco albetovirus 1</i>	tobacco necrosis virus A (<i>Tombusviridae</i>)	V01468	1239	17	STNV suppresses the replication of its helper virus and ameliorates the TNV-induced symptoms in different hosts
Satellite tobacco necrosis virus 2 (STNV-2)	<i>Tobacco albetovirus 2</i>	tobacco necrosis virus A (<i>Tombusviridae</i>)	M64479	1245	17	STNV and STNV-2 coat protein genes share 55% nucleotide sequence identity, whereas the UTRs are more similar
Satellite tobacco necrosis virus C (STNV-C)	<i>Tobacco albetovirus 3</i>	tobacco necrosis virus D (<i>Tombusviridae</i>)	AJ000898	1221	17	STNV and STNV-C coat proteins share 62% sequence identity, whereas the 3'UTRs are only 40% identical
<i>Aumavirus</i>						
Satellite maize white line mosaic virus (SMWLMV)	<i>Maize aumavirus 1</i>	maize white line mosaic virus (<i>Tombusviridae</i>)	M55012	1168	17	CP is 32% identical to that of STNV-1
<i>Papanivirus</i>						
Satellite panicum mosaic virus (SPMV)	<i>Panicum papanivirus 1</i>	Panicum mosaic virus (<i>Tombusviridae</i>)	M17182	826	16	Besides virion formation, CP of SPMV has several other biological functions, including systemic accumulation, maintenance and movement of the SPMV RNA
Satellite St. Augustine decline virus (SSADV)	<i>Panicum papanivirus 1</i>	St. Augustine decline virus strain of PMV (<i>Tombusviridae</i>)	L10083	824	ND	SSADV is a strain of SPMV (36 nt substitutions; 5 aa changes)
Satellite grapevine virus (SGVV)	unclassified	grapevine virus F (<i>Betaflexiviridae</i>)?	KC149510	1060	ND	SGVV and GVF share stem-loop structures at the 5' ends of the genomes
<i>Virtovirus</i>						
Satellite tobacco mosaic virus (STMV)	<i>Tobacco virtovirus 1</i>	tobacco mosaic virus (<i>Virgaviridae</i>)	M25782	1059	17	The 3'UTR is similar to that of tobamoviruses, with a clear sequence similarity between STMV and TMV

predicted in the 5' and 3' UTRs. The 826 nt-long ssRNA genome of SPMV contains two open reading frames (ORF; [Table 1](#)). However, only one of them (the one encoding CP) is expressed in *in vitro* translation assays.

The sequence of SPMV CP is not appreciably similar to those of STNV-like viruses (below 15% identity). However, X-ray structure analysis of the SPMV particle revealed that the protein has a similar jelly-roll fold as the corresponding protein of STNV ([Fig. 1](#)). Beside its structural role, SPMV CP has several other biological functions, most notably, systemic accumulation, maintenance, and movement of the cognate SPMV RNA in plants. Interestingly, the latter activities of SPMV CP apparently extend to the helper virus RNA, assisting in its maintenance or stabilization. It is noteworthy that PMV and SPMV are involved in a peculiar tripartite association with a 350 nt-long satellite RNA (satRNA), whereby PMV provides necessary factors for satRNA replication and SPMV provides CPs for satRNA encapsidation.

Two other viruses encoding SPMV-like CPs have been reported ([Table 1](#)). The first one, satellite St. Augustine decline virus (SSADV), is associated with the St. Augustine decline strain of PMV. SSADV is 95% identical to SPMV over the entire genome length (36 nt changes, 5 aa changes) and is considered as a different strain of SPMV rather than a different species. Notably, SSADV genome is 2 nucleotides shorter than that of SPMV, making it the shortest known satellite virus genome ([Table 1](#)). The second putative satellite virus, satellite grapevine virus (SGVV), has been discovered by deep sequencing of total genomic RNA from grapevine. However, neither the viral particles nor the associated helper virus have been characterized. SGVV CP shares ≈24% sequence identity with SPMV CP.

SPMV and SSADV have been assigned into a genus *Papanivirus* (sigil: *Pa-* for panicovirus [helper virus], *pani-* for panicum), within a species *Panicum papanivirus 1*. By contrast, SGVV remains unclassified due to lack of information on its ability to form virions as well as on the identity of its helper virus.

Interestingly, homologs of SPMV CP are encoded by certain satRNAs. In particular, bamboo mosaic virus satellite RNA (satBaMV; 836 nt) encodes a protein (P20) which is 44% identical to the CP of SPMV. P20 plays a role in the accumulation and movement of the satBaMV but does not participate in satRNA encapsidation. Instead, satBaMV is packaged into rod-shaped particles by the CP of the helper bamboo mosaic virus, a member of the family *Alphaflexiviridae*. In addition, a sequence of olive viral satellite RNA (OVsatRNA) has been deposited to GenBank that encodes a protein 35% identical to P20 of satBaMV. Considering the conservation of the SPMV-like CPs, it appears likely that satBaMV and OVsatRNA have evolved from genuine

satellite viruses, once again emphasizing the apparent ease with which transitions between different types of mobile elements (i.e., parasitic nucleic acids and viruses) occur.

Virtovirus: Satellite Tobacco Mosaic Virus

Satellite tobacco mosaic virus (STMV) has been isolated from *Nicotiana glauca* (tree tobacco) and is naturally associated with and dependent on tobacco mild green mosaic virus (TMGMV; species *Tobacco mild green mosaic virus*, genus *Tobamovirus*, family *Virgaviridae*). However, under experimental settings, STMV can adapt and replicate in many plant hosts (e.g., tobacco, pepper, tomato) in association with other tobamoviruses, including tobacco mosaic virus (TMV). The adaptation of STMV to other helper viruses involves the deletion of a single G residue in a stretch of five G residues (positions 61–65) in the 5' UTR, together with a change in the 5' nucleotide. Thus far, STMV is the only known satellite virus that uses rod-shaped viruses as helpers.

The STMV genome is a linear ssRNA molecule of 1059 nt that contains two ORFs (Fig. 1), both of which are functional in the *in vitro* translation assay. The first ORF encodes a protein of 58 aa which lacks similarity to proteins in the public databases and appears to be dispensable for STMV multiplication. Indeed, certain naturally occurring isolates of STMV contain a deletion within ORF1 and do not produce the corresponding product. The second ORF encodes STMV CP, which also has no identifiable homologs in sequence databases. However, structural analysis shows that the STMV CP has a jelly-roll fold similar to those of STNV and SPMV (Fig. 1), suggesting that the three satellite viruses might be evolutionarily related.

As in the case of STNV and SPMV, but different from the helper TMV virus, the genome of STMV lacks a 7-methylguanylate cap and the first 6 nt of the STMV RNA are identical to those of the STNV genome. However, in contrast to STNV and SPMV, the 5' end of STMV genome is not phosphorylated. The 3' UTR is predicted to contain a series of pseudoknots followed by a tRNA-like structure which can be amino acylated with histidine. The latter features are strikingly similar to those of the genome of helper TMV and other tobamoviruses, with 40–50 nt-long regions of near identity among the STMV and TMV 3' UTRs. These secondary structure elements play critical roles in STMV genome replication, translation, and initiation of virion assembly.

STMV is currently a sole representative of the species *Tobacco virtovirus 1* within the genus *Virtovirus* (sigil: *Vir-* for virgavirus [helper virus], *to-* for tobacco).

Effect of Mixed Infections on the Host and the Helper Viruses

The parasitic lifestyle of satellite viruses is usually associated with negative effects on the fitness of the helper virus, whereas the effect on the host cell during a co-infection might differ drastically between viruses. The negative effect of STNV on the propagation of its helper virus manifests as a decrease in (1) the number of necrotic lesions formed in co-inoculated plant leaves, (2) diameter of the lesions, and (3) the amount of TNV produced in inoculated leaves. However, several factors influence the extent to which TNV multiplication is affected. In particular, these include the relative concentrations of the satellite and helper viruses in the inoculum, the physiological state of the host plants before and during the infection, and the type of plant used for the assay. An excess of STNV present during co-infection reduces the yield of TNV to non-detectable levels. This reduction is associated with a selective decrease in the synthesis of TNV capsid protein, suggesting that STNV affects TNV virion production. By contrast, co-infection of PMV with SPMV exacerbates PMV disease phenotype in millet plants resulting in a severe chlorosis and stunting, reduced seed production, and often death of the infected plant. The synergism between PMV and SPMV is also characterized by the more rapid systemic accumulation of PMV and SPMV CPs and genomic RNAs than during infection by PMV alone. Expression of the SPMV CP from a potato virus X-based vector was shown to induce the formation of chlorotic spots on the host as well as non-host plants, confirming that SPMV CP is a pathogenicity determinant.

The effects of STMV on the multiplication of its helper virus, as well as on the helper virus-induced symptoms, are dependent on the host. In tobacco plants, STMV does not change the mild mosaic symptom caused by TMGMV, whereas in jalapeno pepper severe leaf blistering induced by TMGMV is attenuated by STMV infection. Furthermore, tobamovirus titers are greatly decreased by STMV in pepper compared to other hosts.

Common Ancestry of Plant Satellite Viruses?

Plant satellite viruses from the four groups described above all propagate in flowering plants (angiosperms) and share several genomic and structural characteristics that distinguish them from other known viruses. Namely, representatives of all four satellite virus groups form capsids with $T = 1$ icosahedral symmetry and composed of 60 CP subunits (Fig. 1). Although common among ssDNA viruses (e.g., nanoviruses and circoviruses), $T = 1$ capsids are not used by other known ssRNA viruses, which typically have larger $T = 3$ or $T = 4$ capsids. Furthermore, the CPs of all described plant satellite viruses have the same jelly-roll fold (Fig. 1), despite negligible sequence similarity. The similarities also extend to the genomic characteristics. In particular, in all plant satellite viruses, the linear ssRNA genomes lack the 5' 7-methylguanylate caps and polyadenylation sequences in their 3' UTRs. Furthermore, 5' ends of the STNV and SPMV genomes are phosphorylated, whereas the first six

nucleotides at the 5'-terminus of the STMV genome are identical to those in STNV. Considering these similarities, it is conceivable that all plant satellite viruses have evolved from a common ancestor. Notably, in structural comparison of the satellite virus CPs with the jelly-roll CPs of other viruses, satellite viruses formed a monophyletic clade, indicating that they indeed might share a common ancestor, despite high sequence divergence.

Concluding Remarks

Due to their simplicity, satellite viruses have served (and continue to serve) as useful models for elucidating the molecular details underpinning various processes, including mechanisms of cap-independent protein translation, virion assembly and viral genome encapsidation. For instance, it has been recently demonstrated that assembly of the icosahedral STNV virion proceeds through recognition of the packaging signals on the viral RNA by the CP, which promotes the protein-protein interactions needed to build the capsid. Similar virion assembly mechanism has been also validated for more complex ($T = 3$) bacterial and animal ssRNA viruses, underscoring the relevance of satellite viruses as model systems. Plant satellite viruses also provide a window into the complexity of natural ecosystems, where "one-on-one" virus-host encounters are likely to be rare, but rather involve multiple viruses and other types of genetic parasites interacting in an antagonistic or synergistic manner depending on environmental as well as virus- and host-specific cues. Thus, further studies on tripartite host-virus-satellite virus systems are likely to reveal new facets of virus-host and virus-virus interactions. Finally, despite their minimalism, the ultimate origins of satellite viruses remain enigmatic. On the one hand, it is possible that plant satellite viruses resemble the simplest, potentially ancestral, RNA viruses, which could have existed at the early stages of the evolution of the virosphere. On the other hand, it cannot be excluded that they are a result of reductive evolution from more complex quasi-autonomous RNA viruses. Further exploration of satellite virus diversity in other plants and animals should shed light on the evolutionary origins of this remarkable group of viruses.

Further Reading

- Ban, N., Larson, S.B., McPherson, A., 1995. Structural comparison of the plant satellite viruses. *Virology* 214 (2), 571–583.
- Bringloe, D.H., Pleij, C.W., Coutts, R.H., 1999. Mutation analysis of cis-elements in the 3'- and 5'-untranslated regions of satellite tobacco necrosis virus strain C RNA. *Virology* 264 (1), 76–84.
- Dodds, J.A., 1999. Satellite tobacco mosaic virus. *Current Topics in Microbiology and Immunology* 239, 145–157.
- Duponchel, S., Fischer, M.G., 2019. Viva laidaviruses! Five features of virophages that parasitize giant DNA viruses. *PLoS Pathogens* 15 (3), e1007592.
- Ford, R.J., Barker, A.M., Bakker, S.E., *et al.*, 2013. Sequence-specific, RNA-protein interactions overcome electrostatic barriers preventing assembly of satellite tobacco necrosis virus coat protein. *Journal of Molecular Biology* 425 (6), 1050–1064.
- Francki, R.I., 1985. Plant virus satellites. *Annual Review of Microbiology* 39, 151–174.
- Gnanasekaran, P., Chakraborty, S., 2018. Biology of viral satellites and their role in pathogenesis. *Current Opinion in Virology* 33, 96–105.
- Jones, I.M., Reichmann, M.E., 1973. The proteins synthesized in tobacco leaves infected with tobacco necrosis virus and satellite tobacco necrosis virus. *Virology* 52 (1), 49–56.
- Kassanis, B., 1962. Properties and behaviour of a virus depending on its multiplication on another. *Journal of General Microbiology* 27, 477–488.
- Kassanis, B., Nixon, H.L., 1960. Activation of one plant virus by another. *Nature* 187, 713–714.
- Kneller, E.L., Rakotonirafara, A.M., Miller, W.A., 2006. Cap-independent translation of plant viral RNAs. *Virus Research* 119 (1), 63–75.
- Krupovic, M., Czirke-Krupovic, V., 2011. Virophages or satellite viruses? *Nature Reviews Microbiology* 9 (11), 762–763.
- Krupovic, M., Kuhn, J.H., Fischer, M.G., 2016. A classification system for virophages and satellite viruses. *Archives of Virology* 161 (1), 233–247.
- Larson, S.B., McPherson, A., 2001. Satellite tobacco mosaic virus RNA: Structure and implications for assembly. *Current Opinion in Structural Biology* 11 (1), 59–65.
- Mougari, S., Sahmi-Bounsiar, D., Levasseur, A., Colson, P., La Scola, B., 2019. Virophages of giant viruses: An update at eleven. *Viruses* 11 (8), E733.
- Murant, A.F., Mayo, M.A., 1982. Satellites of plant viruses. *Annual Review of Phytopathology* 20, 49–70.
- Patel, N., Dykeman, E.C., Coutts, R.H.A., *et al.*, 2015. Revealing the density of encoded functions in a viral RNA. *Proceedings of the National Academy of Sciences of the United States of America* 112 (7), 2227–2232.
- Patel, N., Wroblewski, E., Leonov, G., *et al.*, 2017. Rewriting nature's assembly manual for a ssRNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 114 (46), 12255–12260.
- Scholthof, K.B., Jones, R.W., Jackson, A.O., 1999. Biology and structure of plant satellite viruses activated by icosahedral helper viruses. *Current Topics in Microbiology and Immunology* 239, 123–143.
- van Lipzig, R., Gultyaev, A.P., Pleij, C.W.A., *et al.*, 2002. The 5' and 3' extremities of the satellite tobacco necrosis virus translational enhancer domain contribute differentially to stimulation of translation. *RNA* 8 (2), 229–236.

Plum Pox Virus (*Potyviridae*)

Miroslav Glasa, Biomedical Research Center, Slovak Academy of Sciences, Bratislava, Slovakia

Thierry Candresse, The National Research Institute for Agriculture, Food and the Environment, University of Bordeaux, Villenave d'Ornon, France

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Asia Minor The westernmost extent of Western Asia, comprises most of the Asian part of modern Turkey and the Armenian highland.

Cross-protection Protection of plant against the symptoms of a severe virus isolate by the pre-inoculation with a less severe isolate.

Interactomics Systematic survey of all interactions between viral proteins and plant cellular proteins

Polyacrylamide gel electrophoresis (PAGE) An analytical method for separating biological macromolecules (proteins or nucleic acids) by size and charge.

Sharka disease Sharka means pox in Bulgarian, as such symptoms on fruits were first observed in this country.

Introduction

Plum pox virus (PPV), the agent responsible for the Sharka disease, belongs to the genus *Potyvirus* (family *Potyviridae*). The natural host range of this virus is restricted to *Prunus* spp. (stone fruits and ornamental trees). The infection of susceptible genotypes results in characteristic foliar and fruit symptoms and premature fruit drop. The geographical distribution covers the vast majority of *Prunus* producing area (except Australia, New Zealand, South Africa and the USA), although with widely different incidence levels in different countries. PPV is transmitted non-persistently by more than 20 aphid species, by grafting and vegetative multiplication of infected plants, but is not seed-borne. To date, at least 10 strains/molecular groups of PPV have been identified based on biological, serological, and molecular properties (M, D, Rec, EA, T, W, C, CR, CV and An). Although many diagnostic tools are available for the sensitive and/or specific detection of PPV, its huge variability, uneven distribution in infected woody hosts and its low titre outside of the active growth period significantly complicate its detection. In regions free of PPV, strict quarantine measures are usually enforced.

With the limited availability of resistant or tolerant fruit tree varieties, a mix of prophylactic approaches including the use of virus-free propagation material, eradication of diseased trees, and vector control is generally used in an effort to control the virus in regions where it has not reached an endemic status.

Economical Importance

PPV is considered as the most detrimental viral pathogen of stone fruit *Prunus* crops (peach, apricot, plum, Japanese plum). The disease is named according to the characteristic symptoms on fruits (sharka means pox in Bulgarian, as it was first observed in this country). During approximately a century of recognized existence, PPV has had a devastating impact on the global stone-fruit industry and nursery production, mainly in the central and south European countries. The damages are linked not only to the direct fruit yield and quality losses, but include also quarantine, eradication, prevention and compensatory measures. Since the 1970s the global negative impact of PPV has been estimated to exceed 10 billion euros.

In the field, *Prunus* trees infected with PPV cannot be cured and are often eliminated as a consequence of disease eradication or containment efforts. PPV infected trees are usually not stunted and do not die, however, the infection is manifested by pronounced symptoms negatively affecting fruit yield and quality. The fruits of susceptible *Prunus* hosts are in some cases unsuitable for consumption or processing because of decreased weight and sugar content, overall lower gustative quality and visual defects (depressions, flesh browning, necrosis or gumming). The most spectacular symptom is the premature fruit drop, which may reach 80%–100% in the most susceptible cultivars. Consequently, traditional susceptible cultivars have in many places been replaced by less susceptible or tolerant cultivars, which are often of lower gustatory quality.

The economic losses are not only associated with fruit production, but PPV also negatively affect the production of propagation materials (rootstock, budwood, scions) as in case of contamination these cannot be commercialized. Because of its high potential impact on stone fruit crops, PPV has been included in a list of the top ten plant viruses for impact or scientific interest and is listed as a quarantine pathogen in many parts of the world. In Europe for example, PPV is listed in the EC Plant Health Directive (Annex II of the European Union council directive 2000/29/EEC).

History and Geographical Distribution

The sharka disease was observed for the first time around 1917 on plum cv. Kyustendil in the village of Zemen in western Bulgaria, near the Yugoslavian border. The viral character of the disease on plum and on apricot was described in 1933. Later, PPV was recognized on peach (in the 1960's) and on cherries (in the 1980's). Mainly after World War II, sharka was gradually reported in

various *Prunus* species from different regions across the European continent and Mediterranean basin and later from other continents. Although these reports and their timescale do not necessarily represent the actual dissemination events, they suggest a progressive spread of PPV, probably through infected propagation material. Apart from the widespread distribution in Europe, the virus has been reported from South American (Chile, Argentina), the North American continent (USA, Canada), Asia (China, Japan, Asia Minor) and North Africa (Egypt, Tunisia).

However, its prevalence varies from region to region from the endemic occurrence observed in central and eastern European countries, where the virus is well established, to local and more limited incidence observed in other countries where the virus was only introduced recently or where strict phytosanitary control measures have been enforced and have successfully kept the virus under some level of control (e.g., USA, where eradication was recently achieved).

Host Range and Symptomatology

The natural host range of PPV is restricted to species of the genus *Prunus* (family *Rosaceae*), including cultivated stone fruits – plum (*P. domestica*), Japanese plum (*P. salicina*), apricot (*P. armeniaca*), peach and nectarine (*P. persica*), almond (*P. dulcis*) and myrobalan (*P. cerasifera*). Sweet cherry (*P. avium*), sour cherry (*P. cerasus*) and mahaleb cherry (*P. mahaleb*) can also be naturally infected by cherry-adapted PPV isolates.

PPV also infects wild and ornamental species, such as nanking cherry (*P. tomentosa*), Japanese apricot (*P. mume*), Canadian plum (*P. nigra*), American plum (*P. americana*), dwarf flowering almond (*P. glandulosa*), and blackthorn (*P. spinosa*), which may thus act as local reservoirs of the virus.

Symptoms on *Prunus* trees may be observed on leaves, flowers, fruits, and/or stones, however, they vary depending on host/cultivar susceptibility, virus isolate, physiological status and age of the host and environmental conditions (Fig. 1). The presence of other viruses in the trees, such as *Prunus* necrotic ringspot virus (PNRSV), prune dwarf virus (PDV) or apple chlorotic leafspot virus (ACLSV) may further increase the severity of symptoms.

Under field conditions, the symptoms on *Prunus* plants are often masked late in the season or during the warm period of the growing season. Also, early infections are often characterized by symptoms restricted to only some parts of tree canopy and full systemic invasion of a tree may require several years. These factors may complicate in some cases the identification of the disease by visual inspection. Moreover, the irregular distribution and translocation of the virus in the trees and the low titre outside the active growth period may complicate the detection of the virus when methods of insufficient sensitivity are used.

Typical foliar symptoms on plum consist generally of pale green chlorotic rings, spots, or patterns. Susceptible cultivars develop shallow ring or arabesque depressions on fruits, sometimes with brown or reddish necrotic flesh and gumming. Tolerant plum cultivars show no symptoms on fruits.

Infected apricots develop chlorotic or pale-green rings and lines on leaves, light-colored depressed rings on fruits, which may be severely deformed. Stones are marked with typical discolored rings.

Symptoms on susceptible peach genotypes are pronounced vein clearing, small chlorotic blotches, and distortions of the leaves. Color-breaking symptoms on the petals are observed in some varieties. Pale rings or diffuse band are visible on the skin of the fruits. In general, the symptoms on peach tend to be less visible in comparison to those on plums and apricots. Infection of almond is often symptomless or with limited foliar symptoms.

Characteristic symptoms on cherries consist of pale green patterns and rings on leaves. Fruits may be slightly deformed, with chlorotic and necrotic rings and notched marks.

Premature drop of fruits (reaching up to 100% in the most susceptible cultivars) is frequently observed, particularly in plum and apricot.

Virion Structure and Genome Properties

The flexuous filamentous viral particles are approximately 750 × 15 nm. Viral particles are composed of the single-stranded genomic RNA encapsidated by a single type of capsid protein subunit. The genome is of positive polarity and is 9741–9795 nt in length. It has a polyadenylated 3' end and a virus-encoded protein (VPg) covalently bound at its 5' end. The genomic organization is typical of potyviruses, with a single open reading frame encoding a large polyprotein precursor (3125–3143 amino acids, c. 355 kDa) that is proteolytically processed by three virus encoded proteinases (P1, HC-Pro, and NIa-Pro) to yield as many as ten mature functional proteins. As in other potyviruses, an additional product, P3N-PIPO, produced by a polymerase slippage mechanism from an alternative short ORF embedded within the P3 coding sequence, has been predicted (Fig. 2).

Strains/Molecular Groups

Initial attempts to study PPV variability involved the description of symptoms observed following mechanical inoculation of susceptible experimental herbaceous hosts. The most widely used biological classification system resulting from these efforts was based on the reaction of an experimental local-lesion host, *Chenopodium foetidum*, and PPV isolates were assigned as yellow, intermediate, and necrotic types depending on the type of local lesions they induced. It was also observed that different isolates

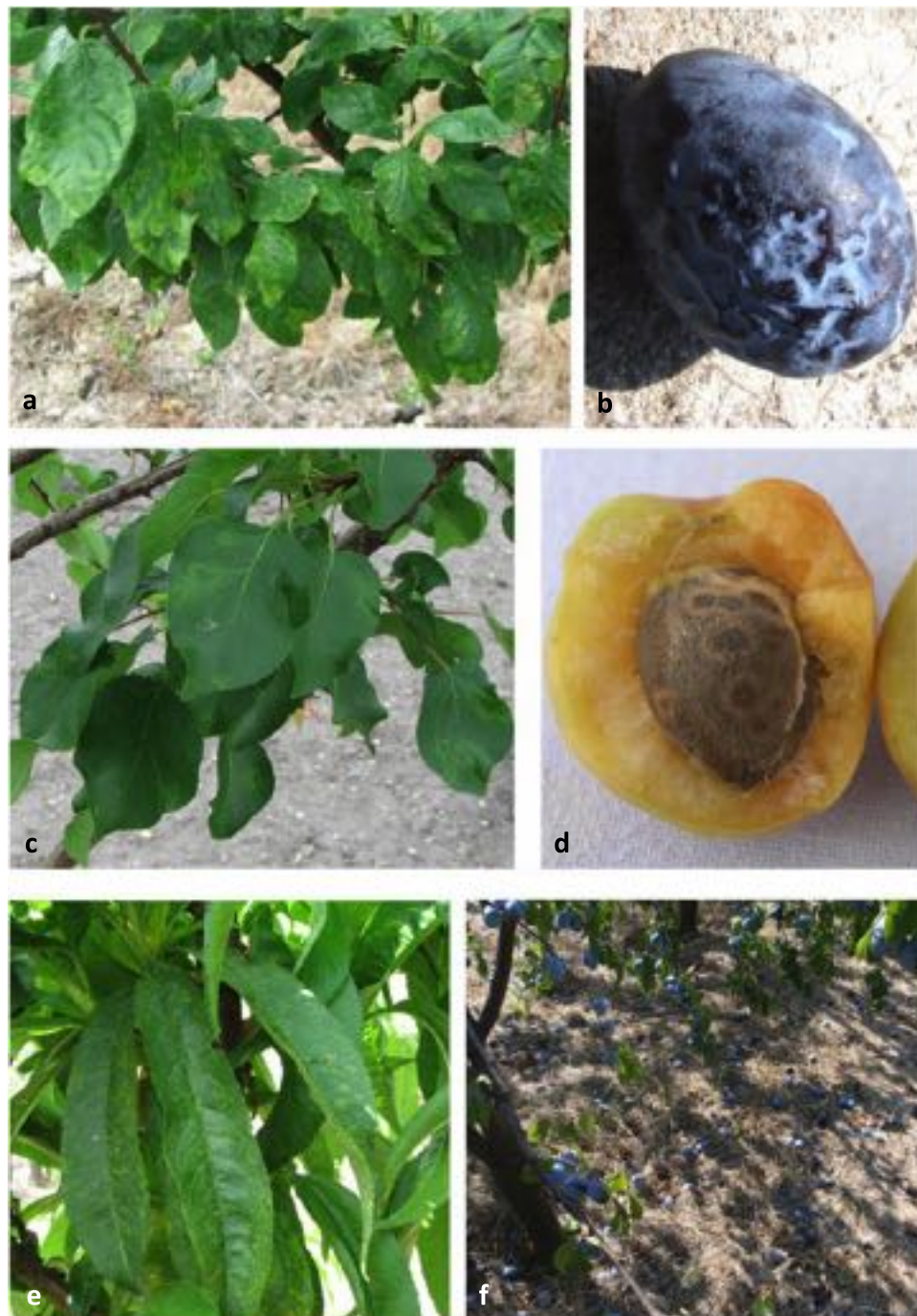


Fig. 1 Typical leaf and fruit symptoms caused by Plum pox virus infection on susceptible *Prunus* genotypes, i.e., plum (*P. domestica*) (a, b, f), apricot (*P. armeniaca*) (c, d) and peach (*P. persica*) (e).

can induce variously severe symptoms on some woody indicators (*Prunus persica* GF305, *P. tomentosa*) after chip budding or controlled aphid transmission.

However, such phenotypic characterization efforts remained often unreproducible because of difficulties to standardize the assays and of the influence, at least to some extent, of environmental conditions. Moreover, these experiments did not establish the link between these biological properties and the serological/molecular properties of PPV isolates.

In the late 1970s, the existence of two serogroups of PPV was demonstrated using agar double diffusion assays employing polyclonal antibodies and a purified, formaldehyde-treated, suspension of undegraded viral particles. These two serogroups were named M and D, after the first inoculum sources used in these experiments (Marcus from peach in Greece and Dideron from apricot in south-eastern France, respectively).

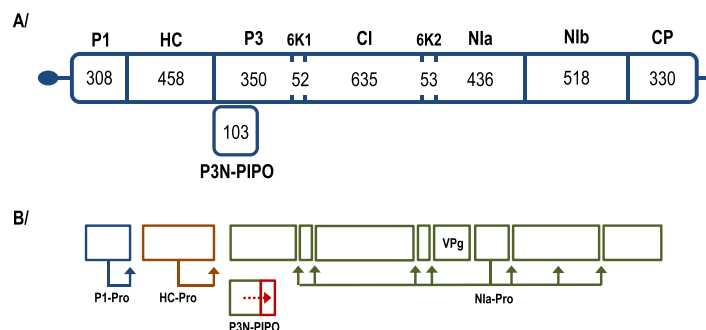


Fig. 2 Schematic representation of the Plum pox virus genome (A). The viral polyprotein and the alternative PIPO small ORF are represented by oval rectangles divided into boxes representing the respective functional products with their deduced length in amino acids (based on the AY028309 isolate). Polyprotein processing (B) by the three viral-encoded proteinases to yield the mature viral proteins. PIPO is expressed by a frameshifting resulting from a polymerase slippage mechanism.

It was later shown that different PPV isolates, belonging to these two serogroups (recognized as strains), show also reproducible coat protein (CP) mobility differences in denaturing polyacrylamide gel electrophoresis (PAGE) as well as a strain-specific *Rsa* I restriction site polymorphism in the region encoding the CP carboxy-terminus (CP C-terminus). However, some rare isolates behaved unusually using this RFLP typing assay, indicating that the serological and/or molecular variability of PPV is wider than that observed in these two D and M strains.

Indeed, on the basis of recent molecular analysis, at least 10 strains of PPV, grouping isolates sharing common molecular (and, partially, serological) properties are now recognized. This classification has been validated by comparisons of complete genomic sequences of representative isolates for each of these strains.

Although isolates of each strains were initially thought to have also in common some biological and epidemiological properties, such shared characteristics are now understood to exist only in some strains, and to be limited by the high intra-strain biological variability of PPV. Most PPV isolates can be assigned to one of the three major strains (PPV-M, D and Rec). Other strains are geographically limited PPV-W, EA and T, cherry adapted strains PPV-C, CR and CV and the atypical PPV-An strain.

PPV-M (named after the type isolate Marcus). The isolates of this strain are present in many European countries, but absent from the Americas and China. They are often associated with rapidly spreading epidemics in peach but are less frequently found on plums. Usually, PPV-M isolates are transmitted efficiently by aphids. A high intra-strain molecular diversity was recently confirmed in PPV-M by the identification of divergent PPV-M-Ist isolates from Turkey.

PPV-D (named after the type isolate Dideron). Based on geographical distribution data, PPV-D seems to be the most successful strain as PPV-D isolates are widespread in all areas where PPV has been reported, including the recent outbreaks in Asia and South and North America. PPV-D isolates naturally infect all the susceptible *Prunus* species excluding cherries, but are less frequently associated with spreading epidemics in peach.

PPV-Rec (from "recombinant"). This group of isolates was recognized through the use of improved strain-typing methods targeting different genome portions. PPV-Rec genomes are characterized by a conserved homologous recombination event between PPV-M and PPV-D-like sequences, with a cross-over located in the 3' terminal part of the NIb coding region. PPV-Rec is therefore unrecognisable from PPV-M when using CP-based serological tests. Therefore, a number of isolates originally described as PPV-M were later retyped as belonging in fact to the PPV-Rec strain. PPV-Rec isolates are widespread in several central and eastern European countries, frequently associated with plums and efficiently transmitted by aphids. Interestingly, natural infections of peach by PPV-Rec are rare and mostly symptomless.

PPV-W (from "Winona"). Although the first PPV-W report is from a few infected plum trees in Canada, the probable origin of this strain can be placed in Eastern Europe. PPV-W isolates were found naturally spreading and disseminated in different *Prunus* in Russia, Latvia and Ukraine. The Canadian PPV-W isolate shows multiple recombination events in its genome, one of which is absent in the genomes of East European PPV-W. PPV-W shows the highest intra-strain variability. PPV-W isolates are efficiently transmitted by aphids. Although not reported from central or western Europe, PPV-W isolates are competitive with the main strains in their regions of occurrence.

PPV-EA (named after the type isolate El Amar) was originally isolated from apricots in the El Amar region of the Nile delta in Egypt in the late 1980s. It has also been observed in peach and in other regions of Egypt but has thus far not been reported outside this country.

PPV-T (from "Turkey"). Found widespread in Turkey and also reported from Albania, PPV-T isolates are characterized by a unique recombination event affecting the 5' part of their genome.

PPV-C (from "Cherry") was first reported on sour cherry in Moldova and on sweet cherry in Italy. This group of cherry-adapted isolates has until now been sporadically reported from some central and eastern European countries, apparently without an epidemiological significance. PPV-C isolates are aphid transmissible and systemically infect sweet and sour cherry and *P. mahaleb*, however, they can also infect other *Prunus* species under experimental conditions.

PPV-CR (from "Cherry Russian") is the second strain naturally infecting cherries, although molecularly distinct from PPV-C. The PPV-CR isolates were found in two geographically distant areas of European Russia (Moscow and Samara regions). Similarly, to PPV-C, the isolates of this strain can be experimentally inoculated to other *Prunus* species and are transmitted by aphids.

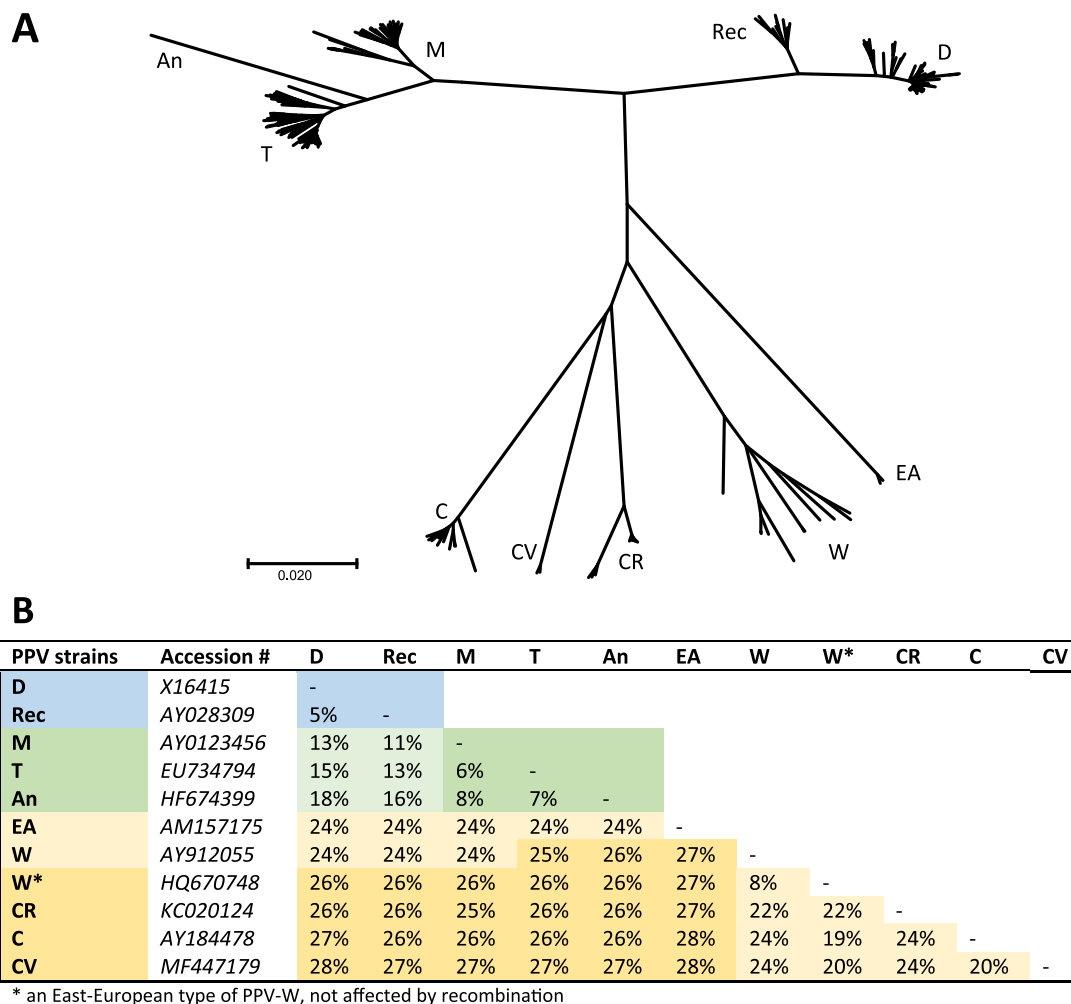


Fig. 3 Phylogenetic relationships of strains of PPV. (A) Unrooted phylogenetic tree computed using full-length genome sequences (available in the GenBank database, accessed on March 2019) showing the relationships between isolates of all ten recognized PPV strains. (B) Pairwise genetic distances calculated on the full-length genomes of representative isolates of all ten recognized PPV strains. The representative isolates are identified by their Genbank accession numbers.

PPV-CV (from “Cherry Volga”) is the third cherry-adapted strain of PPV, identified recently on sour cherries in the Tatarstan region (middle Volga river region of Russia). The sequence identity between PPV-CV isolates and those of other cherry-adapted strains (C, CR) reaches only ca. 83%, clearly separating them in phylogenetic analyzes into distinct monophyletic groups. Due to their restricted geographical dissemination and recent discovery, the epidemiological importance of these strains for cherry production remains to be determined. Cherry-adapted strains calculated intra-strain diversity are low, probably reflecting some epidemiological constraints limiting their genetic divergence, a possible recent emergence or insufficient sampling to date.

PPV-An (from “ancestral”). A putative PPV strain represented by a recently identified isolate from eastern Albania. The full-length genomic sequence of this isolate fulfils the features of an ancestral PPV-M isolate previously hypothesized in the PPV evolutionary scenario.

Full-length genomic sequences for more than 200 PPV isolates, representing each of the ten recognized strains, are available in public databases (Fig. 3(A)). Moreover, partial sequences, focusing mainly on the 3′ terminal part of genome or the P3–6K1 region, are available for a large number of isolates, making PPV one of the genetically best studied potyviruses. The comparison of complete genomic sequences revealed from 4.7% (PPV-D versus Rec) up to 28% (cherry adapted strains versus EA) nucleotide divergence between PPV strains (Fig. 3(B)).

Recombination seems to have played a significant role in the evolutionary history of PPV. The PPV-Rec strain derives from a single recombination event involving isolates belonging to the PPV-D and PPV-M strains, with a recombination breakpoint in the C-terminus of the NIb gene. In addition, an analysis of complete genomic sequences has recently demonstrated that the PPV-D and PPV-M strains themselves share an ancestrally recombined 5′ part of their genome (5′ non-coding region, P1, HC-Pro, and N-terminus of P3). Other recently identified recombination events involve the W3174 isolate of PPV-W, which shows a mosaic structure as a consequence of recombination with PPV-M and PPV-D in the P1-HC region. The genome of PPV-T

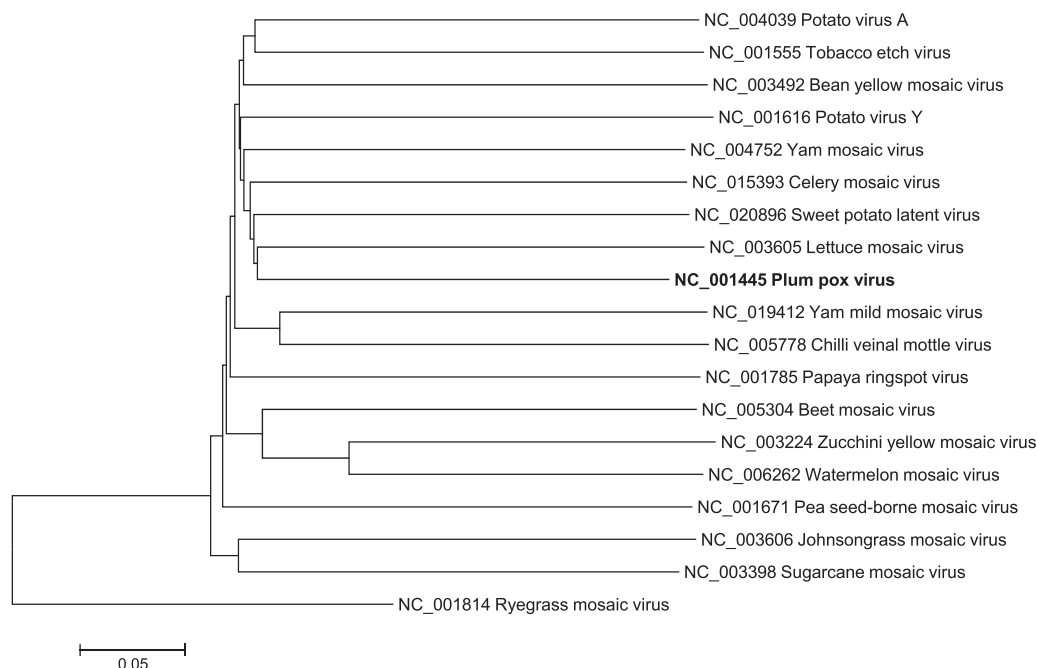


Fig. 4 Neighbor joining phylogenetic tree of the full-length genome sequences of selected potyviruses, including Plum pox virus (in bold). The scale bar indicates branch lengths in substitutions per nucleotide.

isolates shows a recombination in the HC-Pro-P3 region (with PPV-D as a minor parent). In addition, a recent analysis involving a larger set of PPV-T isolates identified other recombination breakpoints in the central part of the genomes of divergent isolates of this strain (Fig. 4).

Thus, a growing number of available full-length genomes and of bioinformatics tools has enabled to show that besides the well-established recombinant strains (PPV-Rec, PPV-T), some other strains have recombination events in their ancient or more recent evolutionary history. Moreover, although less recognisable, it seems that at least for some PPV-W isolates, an intra-strain recombination event has also been documented. On the other hand, several strain (typical PPV-W, PPV-C, PPV-CR, PPV-CV and PPV-EA) appear to represent independent evolutionary lineages not affected by recombination.

Although the complexity of PPV diversity is unlikely to have yet been fully uncovered, phylogenetic relationships based on complete genomes suggest that PPV evolved as two major branches or groups. One group involves the M, D, Rec, T, An and EA strains, that share partially or completely a common 5' genomic region as a result of an ancestral recombination. The second group is represented by the W, C, CR and CV strains. As most isolates of these latter strains were, up to now, detected in Russia and its historically linked East European countries, an influence of such geographical separation could be connected to their limited spread or absence in other countries.

Detection

The recommended detection methods include biological indexing as well as serological and molecular assays. The use of woody biological indicators (such as peach GF305 or *P. tomentosa*) for diagnostic purposes is, however, restricted due to the labor intensive nature of such assays as well as to their length. Moreover, particular PPV strains do not seem to produce symptoms on these indicators with equal effectiveness (as in the case of GF305 and PPV-Rec).

Specific monoclonal antibodies (MAbs) against several PPV serotypes (D, M/Rec, EA, W, C) have been prepared, although some isolates may show abnormal typing properties, escaping also to the recognition by broad-spectrum antibodies. A panel of molecular methods based on RT-PCR, real-time PCR or loop-mediated isothermal amplification can be used for the accurate identification or characterization of PPV strains. Some of these methods are part of the International Plant Protection Convention-Food and Agriculture Organization (IPPC-FAO) protocol for PPV diagnosis.

Proper identification of recombinant isolates requires the use of several techniques targeting different parts of the viral genome or the use of specific primers with binding sites located on both sides of the targeted recombination breakpoint. In addition, whole genome sequencing, or sequencing of several strain-informative genome portions can provide unambiguous PPV strain identification.

An additional level in our understanding of PPV diversity derives from the analysis of PPV-infected plants by high-throughput sequencing, enabling an unbiased identification of new sequence variants. Moreover, the analysis of mixed infections involving different PPV variants/strains and their persistence, until now not well documented, is another challenge in PPV research.

Virus Spread

Natural spread of PPV occurs through aphid vectors from neighboring infected reservoirs. The aphids can acquire the virus from leaves or young shoots of cultivated, wild, or ornamental *Prunus* trees, but also from contaminated fruits, indicating that fruit shipments could also represent a minor pathway for international virus movement.

The experimental host range of PPV is large, with over 60 reported host species (such as *Chenopodium foetidum*, *Nicotiana benthamiana* and *N. clevelandii*, *Pisum sativum*, *Ranunculus arvensis*, *Senecio vulgaris*, *Stellaria media*, etc.) in eight families. However, no epidemiologically significant contribution of weeds and annual herbaceous plants to the spread of PPV has been reported.

The most efficient long-range movement of the virus is linked to international exchange of contaminated *Prunus* propagation materials (rootstocks, budwood, nursery plants).

PPV is thus usually introduced in a new location by infected propagation material. Once established in the plantation, the virus is rapidly spread by aphid vectors when making test probes and migrating between plants. PPV is naturally transmitted by over 20 different aphid species (*Myzus persicae*, *M. varians*, *Aphis gossypii*, *A. craccivora*, *A. hederae*, *A. spiraeicola*, *Brachycaudus cardui*, *B. helichrysi*, *B. persicae*, *Phorodon humuli*, *Rhopalosiphum padi*, etc.).

The virus is transmitted in a non-persistent manner so that aphids may acquire it very rapidly when probing infected plants with their stylets (piercing-sucking mouthparts). However, aphids can only transmit the virus for a short time after virus acquisition as it is retained in a viable state in the aphid's mouthparts for a period of only a few minutes to a few hours. The virus appears to be lost by the aphids the first time they probe on the next plant and aphids do not remain infectious after molting nor do they pass PPV onto their progeny.

The efficiency of transmission depends on the particular aphid species, on the virus isolate and on the host species from which the virus is acquired and to which it is transmitted. As for other potyviruses, two viral proteins are known to be involved in the transmission mechanism, the HC-Pro (helper component) and the CP. The DAG and PTK amino acid motifs are highly conserved in the N-terminus of the CP and HC protein of aphid-transmissible PPV isolates, respectively, and loss of transmissibility is correlated with their deletion or modification.

Although the virus was detected in fruits and seed coat and some conflicting results have been reported in the past concerning seed transmission of PPV in some *Prunus* species, recent studies have failed to demonstrate seed transmission of PPV in any of its woody hosts.

Ecology and Control

In the countries and regions where the virus is not yet present (e.g., Australia), strict quarantine measures need to be established to prevent the introduction of PPV through legal or illegal importation of infected fruit tree propagation material or fruits. In regions where the disease is present but still localized and under control, a mix of prophylactic approaches is usually implemented, including the application of programmes for production and use of virus-free propagation material (rootstocks, budwoods), tight aphid vector control in nurseries (e.g., by application of mineral oil), regular inspection of orchards, and prompt eradication of infected plants.

In regions of endemic PPV occurrence, where eradication brings no solution to the spread of the disease, the practical measures to pursue stone fruit production rely on the use of less susceptible or tolerant genotypes. This practice, however, does not restrict the virus spread and, paradoxically, contributed in the past to the destruction of traditional high-quality genotypes in the central and south-eastern European countries (e.g., Pozegaca or Bistrita plums).

Early attempts to use cross-protection with attenuated virus isolates, a technique successfully used to control some other potyviruses (Zucchini yellow mosaic virus, Papaya ringspot virus), have not met with success in the case of PPV.

In theory, a long term PPV control could be achieved using PPV-resistant varieties. Therefore, many breeding programmes aiming at the development of such varieties have been engaged. They were based on the use of natural resistance sources identified mainly within the *P. armeniaca*, *P. davidiana* and *P. domestica* germplasms. Efficient breeding progress is, however, complicated by the high degree of heterozygosis, the existence of incompatibility barriers between species, the long generation times and the length and non-standardisation of the resistance tests. Some progress has been achieved in plum and apricot, with the deployment of a few resistant varieties in some countries. However, the durability of these resistances remains an open question and so is their stability if confronted to the high diversity of PPV strains.

Development of genotypes with a hypersensitive response as an active defense mechanism against PPV is another promising way to produce resistant fruit trees or rootstocks, as was demonstrated for some plum varieties (e.g., Jojo), but no hypersensitivity sources are so far known in apricot and peach.

The pathogen-derived resistance strategy and the use of sequence-specific gene silencing offer a complementary approach for the development of PPV-resistant stone fruit cultivars. Experimental herbaceous hosts or *Prunus* plants transformed with different regions of the PPV genome (i.e., P1, HCPro, CI, NIa, NIb, or CP) or with post-transcriptional gene silencing (PTGS)-inducing hairpin-constructs containing viral transgenes have been developed and shown to provide partial or complete resistance to PPV. The transgenic plum "HoneySweet", transformed with a sense PPV CP gene construct, has been shown to be highly resistant to PPV infection and was recently approved for commercial cultivation in the USA, although its use is so far local and limited.

The most recently explored strategies rely on interactomics or genetic studies aimed at the identification and inactivation of host susceptibility factors, such as the translation initiation factor eIF4E and its isoforms.

Further Reading

- Cirilli, M., Geuna, F., Babini, A.R., *et al.*, 2016. Fighting sharka in pach: Current limitations and future perspectives. *Frontiers in Plant Science* 7, 1290. doi:10.3389/fpls.2016.01290.
- Clemente-Moreno, M.J., Hernández, J.A., Díaz-Vivancos, P., 2015. Sharka: How do plants respond to plum pox virus infection? *Journal of Experimental Botany* 66, 25–35.
- García, J.A., Glasa, M., Cambra, M., Candresse, T., 2014. Plum pox virus and sharka: A model potyvirus and a major disease. *Molecular Plant Pathology* 15, 226–241.
- Illardi, V., Tavazza, M., 2015. Biotechnological strategies and tools for plum pox virus resistance: Trans-, intra-, cis-genesis, and beyond. *Frontiers in Plant Science* 6, 379. doi:10.3389/fpls.2015.00379.
- Rimbaud, L., Dallot, S., Gottwald, T., *et al.*, 2015. Sharka epidemiology and worldwide management strategies: Learning lessons to optimize disease control in perennial plants. *Annual Review of Phytopathology* 53, 357–378.

Relevant Websites

- <https://www.ippc.int/en/publications/dp-2-2012-plum-pox-virus/>
DP 02: Plum pox virus.
- <https://gd.eppo.int/taxon/PPV000>
Plum pox virus (PPV000).
- <https://www.cabi.org/isc/datasheet/42203>
Plum pox virus (sharka).
- <https://www.sharco.eu/>
SharCo.
Accueil.

Poleroviruses (*Luteoviridae*)

Hernan Garcia-Ruiz, Natalie M Holste, and Katherine LaTourrette, University of Nebraska–Lincoln, Lincoln, NE, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AGO Argonaute proteins
CP Coat protein or capsid protein
ELISA Enzyme-linked immunological assays
ER Endoplasmic reticulum
HC-Pro Helper component-protease
IRES Internal ribosomal entry site
kb Kilobase
kDa KiloDalton
MP Movement protein

mRNA messenger RNA
nt Nucleotide(s)
ORF Open Reading Frame
Rap1 Replication associated protein 1
RdRp RNA-dependent RNA polymerase
RISC RNA-induced silencing complex
satRNA satellite RNA
UTR Untranslated region
VIGS Virus-induced gene silencing
VPg Viral protein genome-linked

Glossary

Aphid-borne Is a term used to describe aphid-transmitted viruses that infect plants.

Argonaute Are the catalytic component of RNA-induced silencing complexes. They associate with cellular or virus-derived small interfering RNAs. Together, this complex targets cellular or viral RNA for degradation or translational repression, resulting in gene silencing.

F-box Is a conserved region of P0 that mediates interaction with Argonaute protein 1, leading to Argonaute protein 1 ubiquitination and degradation.

Gene silencing Also known as RNA interference, is a sequence-specific gene inactivation system mediated by small interfering RNAs (siRNAs). It downregulates RNA accumulation without affecting DNA sequence.

Leaky scanning Happens during initiation of translation when a start codon is skipped over in preference for another start codon located downstream on the mRNA. This process allows translation of several proteins from the same mRNA.

P0 Is the silencing suppressor encoded for by poleroviruses.

Phloem-limited Refers to viruses that can only replicate in the phloem of the host plant. These viruses cannot infect mesophyll cells and are generally not transmitted mechanically.

Poleroviruses Are viruses in the genus *Polerovirus* within the family *Luteoviridae*. They are obligatorily transmitted by aphids, phloem-limited, and encode a strong silencing suppressor (P0).

Ribosomal frameshift Occurs during translation of an mRNA with two overlapping open reading frames. During the shift, the ribosome slips one base backward towards the 5'-end or one base forward towards the 3'-end. The result is the fusion of two proteins encoded for by different open reading frames on the same mRNA.

Ribosomal read-through Occurs when ribosomes overlook a stop codon and continue translation.

Silencing suppressors Are proteins encoded by viruses, which interfere with gene silencing. They prevent degradation of viral RNA, enhance transcription of viral RNA, and alter silencing of cellular genes. Combined, these effects are in part responsible for symptom development in virus-infected plants.

Introduction

The family *Luteoviridae* consists exclusively of plant infecting viruses divided into three genera: *Luteovirus*, *Polerovirus*, and *Enamovirus* (Table 1). *Luteoviruses* diversified into three genera approximately 1500 years ago in correlation with the expansion of agriculture. Members of the genus *Luteovirus* (luteoviruses) contain the standard genome of the entire *Luteoviridae* family: a single, positive single-stranded RNA (ssRNA) encoding the seven proteins P1 through P7 with multiple overlapping open reading frames (ORFs).

Barley yellow dwarf virus (BYDV) is the type species of the genus *Luteovirus*, which encode a weak silencing suppressor (P4). Members of the genus *Polerovirus* (poleroviruses) are similar to luteoviruses and contain an additional protein (P0) that is a strong silencing suppressor. Luteoviruses and poleroviruses shared a common ancestor approximately 900 years ago with P0 deriving separately in the polerovirus lineage. The type species of the genus *Polerovirus* is *Potato leafroll virus* (PLRV). *Pea enation mosaic virus* (PEMV) is the type species of the genus *Enamovirus* (enamoviruses), which encode P0 but not P4. Within the family *Luteoviridae*, BYDV and PLRV infect important staple crops and cause major economic damage. However, poleroviruses are the most damaging and diverse genus, and have a wide host range.

Poleroviruses have a single, positive ssRNA genome of 5.3–5.7 kb, encapsidated in an icosahedral non-enveloped virion (Fig. 1). Unique features of poleroviruses include obligate transmission by aphids in a circulative, non-propagative manner, infection restricted to the phloem, and lack of mechanical transmission. Symptoms induced by poleroviruses include stunting,

Table 1 Taxonomic organization of the family *Luteoviridae*. Species are grouped by genus

Genus	Species	Abbreviation	Accession number ^a
<i>Luteovirus</i>	<i>Barley yellow dwarf virus KerII</i>	BYDV-KerII	NC_021481.1
	<i>Barley yellow dwarf virus KerIII</i>	BYDV-KerIII	KC559092.1
	<i>Barley yellow dwarf virus – MAV</i>	BYDV-MAV	NC_003680.1
	<i>Barley yellow dwarf virus – PAS</i>	BYDV-PAS	NC_002160.2
	<i>Barley yellow dwarf virus – PAV</i>	BYDV-PAV	NC_004750.1
	<i>Bean leafroll virus</i>	BLRV	NC_003369.1
	<i>Nectarine stem pitting-associated virus</i>	NSPaV	NC_027211.1
	<i>Rose spring dwarf-associated virus</i>	RSDaV	NC_010806.1
<i>Polerovirus</i>	<i>Soybean dwarf virus</i>	SbDV	NC_003056.1
	<i>Beet chlorosis virus</i>	BChV	NC_002766.1
	<i>Beet mild yellowing virus</i>	BMV	NC_003491.1
	<i>Beet western yellows virus</i>	BWYV	NC_004756.1
	<i>Carrot red leaf virus</i>	CRLV	NC_006265.1
	<i>Cereal yellow dwarf virus – RPS</i>	CYDV-RPS	NC_002198.2
	<i>Cereal yellow dwarf virus – RPV</i>	CYDV-RPV	NC_004751.1
	<i>Chickpea chlorotic stunt virus</i>	CpCSV	NC_008249.1
	<i>Cotton leafroll dwarf virus</i>	CLRV	NC_014545.1
	<i>Cucurbit aphid-borne yellows virus</i>	CABYV	NC_003688.1
	<i>Maize yellow dwarf virus RMV</i>	MYDV-RMV	NC_021484.1
	<i>Maize yellow mosaic virus</i>	MYMV	KU248489.1
	<i>Melon aphid-borne yellows virus</i>	MABYV	NC_010809.1
	<i>Pepo aphid-borne yellows virus</i>	PABYV	NC_030225.1
	<i>Pepper vein yellows virus</i>	PVYV	NC_015050.1
	<i>Pepper vein yellows virus 5</i>	PVYV – 5	NC_036803.1
	<i>Potato leafroll virus</i>	PLRV	NC_001747.1
	<i>Suakwa aphid-borne yellows virus</i>	SABYV	NC_018571.2
	<i>Sugarcane yellow leaf virus</i>	ScYLV	NC_000874.1
	<i>Tobacco vein distorting virus</i>	TVDV	NC_010732.1
	<i>Turnip yellows virus</i>	TuYV	NC_003743.1
<i>Enamovirus</i>	<i>Alfalfa enamovirus 1</i>	AEV – 1	NC_029993.1
	<i>Citrus vein enation virus</i>	CVEV	NC_021564.1
	<i>Grapevine enamovirus 1</i>	GVEV – 1	NC_034836.1
	<i>Pea enation mosaic virus 1</i>	PEMV – 1	NC_003629.1
Unassigned	<i>Barley yellow dwarf virus – GPV</i>	BYDV-GPV	NC_039035.1
	<i>Barley yellow dwarf virus – SGV</i>	BYDV-SGV	AY541039.1
	<i>Chickpea stunt disease associated virus</i>	CpSDaV	Y11530.1
	<i>Groundnut rosette assistor virus</i>	GRAV	NC_038509.1
	<i>Indonesian soybean dwarf virus</i>	ISDV	
	<i>Sweet potato leaf speckling virus</i>	SPLSV	NC_038510.1
	<i>Tobacco necrotic dwarf virus</i>	TNDV	

^aAccession numbers in GenBank. Accessions beginning with NC_ are the reference for a particular species.

yellowing, a streaking pattern, and stiff leaves. These symptoms are often confused with adverse environmental factors. Most poleroviruses can be present in seed, tubers, and plant parts used for vegetative propagation. Furthermore, some plant-virus combinations remain asymptomatic. Currently, there are 32 species in the genus *Polerovirus* infecting both monocots and dicots (Table 1). These include economically important staple crops including maize, wheat, sugarcane, and potato. Poleroviruses with the most economic importance are PLRV, Sugarcane yellow leaf virus, and three beet-infecting poleroviruses.

Polerovirus Physical Properties

The polerovirus virion has a T3 icosohedral symmetry with an average diameter of 23 nm (Fig. 1). The capsid is formed by 180 monomers that consist mainly of the coat protein (CP) (approximately 23 kDa) and also contain minor amounts of a readthrough protein (approximately 80 kDa). The readthrough protein substitutes one coat protein monomer when assembling the virion. The ratio of CP to readthrough protein varies from 4:1 to 100:1. The thermal inactivation point is between 50°C and 65°C, with a dilution endpoint between 10⁻³ and 10⁻⁴. Polerovirus virions withstand deep-freeze and thaw, and withstand chloroform and detergents. The longevity in sap at 2°C is between 5 and 10 days.

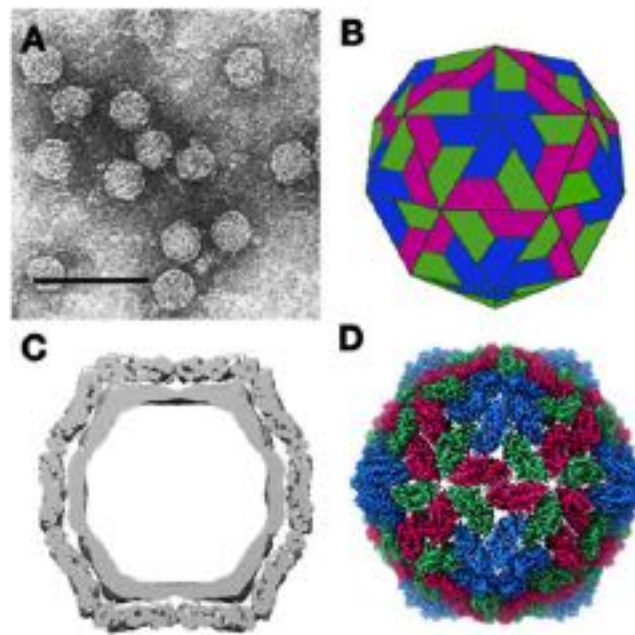


Fig. 1 Illustration of polerovirus virion structure based on potato leafroll virus (PLRV). (A) Transmission electron microscope picture of PLRV virions. The bar represents 100 nm. Photograph courtesy of I.M. Roberts. (B) The coat protein (CP) creates a virion that with $T=3$ icosahedral symmetry composed of 180 capsid proteins organized into 60 asymmetric units. colored according to CP quasi-conformers, where subunit A is blue, subunit B is green, and subunit C is red. There is no envelope and the diameter averages 23 nm. (C) Cryo-electron microscopy of PLRV-like particles. Section of representative density and molecular model, slice through unsharpened maps, depicting density for packaged RNA and/or disordered R domain. (D) cryo-EM maps of whole virus capsid. Structure refinement was carried out with icosahedral symmetry imposed, yielding density maps at a resolution of 3.4 Å. Reproduced with permission from Byrne, M.J., Steele, J.F.C., Hesketh, E.L., *et al.*, 2019. Combining transient expression and cryo-EM to obtain high-resolution structures of luteovirid particles. Structure 27, 1761–1770. doi:10.1016/j.str.2019.09.010.

Polerovirus Genome Organization and Gene Expression

The polerovirus genome consists of a single, positive-strand ssRNA encoding P0 through P7 organized in overlapping ORFs (**Fig. 2(A)**). The 5'-end is protected by the genome-linked protein VPg. The 3'-end contains an -OH group and lacks a poly-A tail. Two sub-genomic RNAs are formed during replication. Translation of polerovirus proteins involves a combination of strategies: leaky scanning, internal ribosomal entry, frameshift, and ribosomal read-through (**Fig. 2(B)**). Additionally, VPg is released from protein P1 by protease processing.

Proteins P0 and P1 are translated from the genomic RNA using leaky scanning and alternate translation initiation codons (**Fig. 2(B)**). P1 can be expressed either individually or fused with P2. When P1 is expressed by itself, it contains two putative domains: VPg and a protease that releases VPg. A ribosomal frameshift produces a P1–P2 fusion protein that generates the RNA-dependent RNA polymerase (RdRp) responsible for viral RNA replication and sub-genomic RNA synthesis. P2 is never expressed by itself. Replication associated protein 1 (Rap1) is translated from genomic RNA through an internal ribosome entry site (IRES).

Protein 3a and the movement protein (MP, P4) are translated by leaky scanning from sub-genomic RNA1 and both are involved in virus movement along with the CP. A ribosomal read-through is required for the translation of the CP read-through (P3–P5), which is less abundant than the CP (P3). An amber stop codon (UAG) separates these ORFs in sub-genomic RNA1 (**Fig. 2(B)**). The CP read-through is not necessary for virion formation, but it is essential for aphid transmission and virus movement in plants. The N-terminal half of the CP read-through determines vector specificity by regulating the efficiency of virus movement through the salivary tissues and gut. Accordingly, mutants lacking the CP read-through accumulate to low levels and are not transmitted by vectors. The C terminal half of the CP read-through is involved in efficient virus movement, tissue tropism, and symptom development in plants.

Several proteins in the genome are currently not well understood. P3a is newly discovered part of the genome. It sits directly upstream of the CP ORF (P3) and is translated by a non-AUG start codon. P3a is required for long-distance movement of poleroviruses. Proteins P6 and P7 are translated by leaky scanning from sub-genomic RNA2. P7 has nucleic acid binding properties. However, the biological role of P6 and P7 remains to be determined.

Polerovirus Phylogenetic Diversity

The evolutionary relationship and phylogenetic diversity of poleroviruses is just beginning to be elucidated. Published studies used the CP and read-through domains based on the assumption that they are highly conserved (**Fig. 3**). Based on the CP

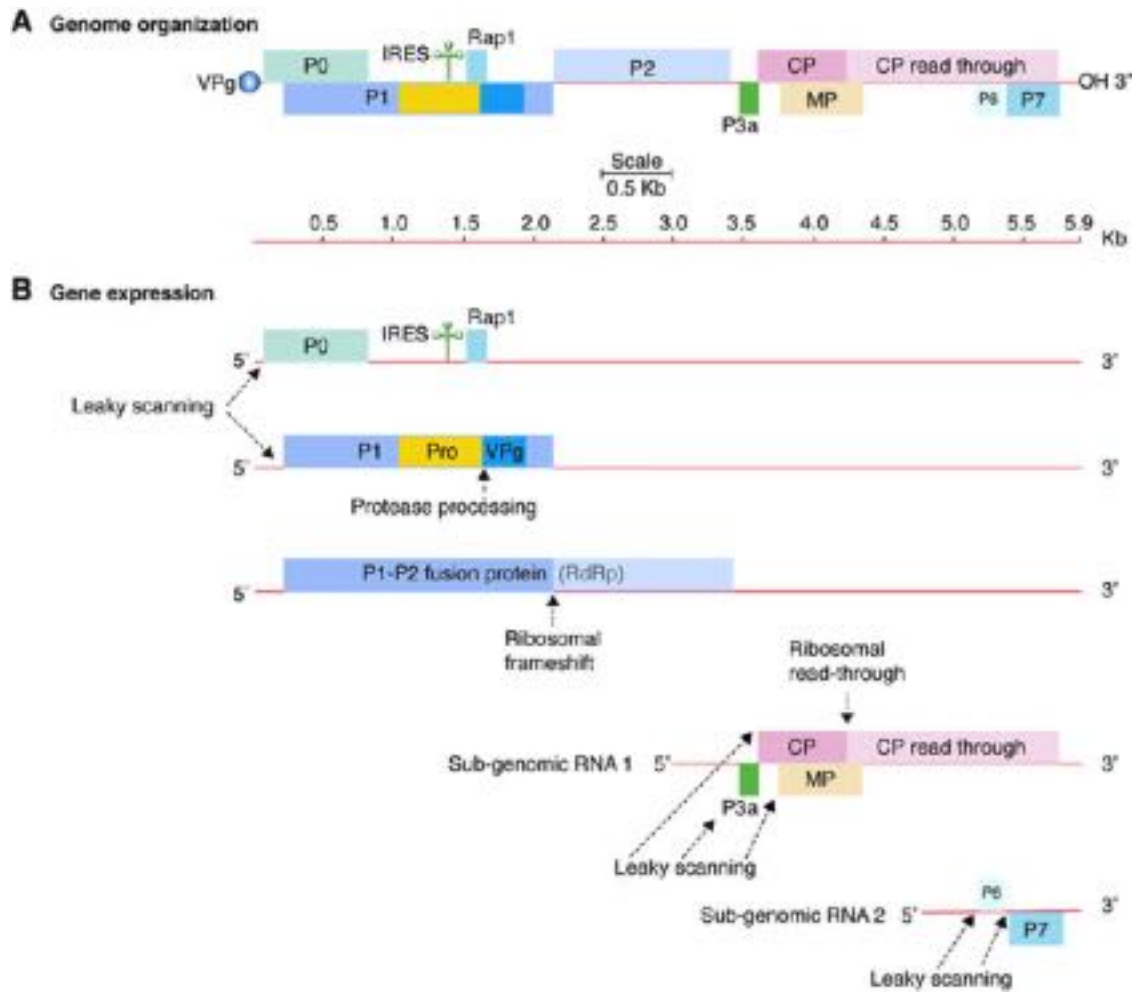


Fig. 2 Schematic representation of polerovirus genome organization and gene expression. Single lines represent non-coding regions and labeled boxes represent cistrons. Sub-genomic RNAs, their formation and proteins translated from them are indicated. (A) Generalized polerovirus genome organization. Coordinates are based on Potato leafroll virus accession number KY856831. (B) Polerovirus gene expression strategies include formation of sub-genomic RNAs, translation by IRES-mediated internal initiation, leaky scanning, ribosomal frameshift, and ribosomal read-through. Protein 1 is processed into mature VPg by proteolysis. Pro: Putative protease. VPg: Viral protein genome-linked. RdRp: RNA-dependent RNA polymerase. Rap1: Replication associated protein. CP: Capsid protein, major and minor. MP: Putative movement protein. p3a: Protein essential for systemic virus movement. IRES: Internal ribosomal entry site.

(Fig. 3(A)), PLRV is an out-group, while Pepper yellow leaf curl virus (PYLCV), Pepper vein yellows virus (PeVYV), and Pepper yellows virus (PepYV) clustered on the same branch, and probably evolved from TVDV. However, the N-terminus of the CP read-through (Fig. 3(B)) separates Tobacco vein-distorting virus (TVDV) and places Cucurbit aphid-borne yellows virus (CABYV) close to pepper-infecting poleroviruses. In contrast, the C-terminus of the CP read-through (Fig. 3(C)) separates pepper-infecting poleroviruses and place TVDV close to PYLCV. Differences in the arrangements of poleroviruses based on CP, N, or C terminal parts of the CP read-through suggest that RNA recombination occurs frequently and is an important contributor to the evolution of poleroviruses.

Polerovirus Transmission

Polerovirus species have evolved to be efficiently transmitted by particular aphid species. PLRV is efficiently transmitted by *Myzus persicae*, while maize poleroviruses are efficiently transmitted by the corn leaf aphid (*Rhopalosiphum maidis*). Other aphids that vector poleroviruses include *R. padi*, *Stiabion avenae*, and *Aphis gossypii*. Aphids vector poleroviruses in a circulative, non-propagative manner.

The cycle begins when an aphid feeds on a polerovirus-infected plant. The virus first will reach the salivary glands of the aphid. It has been found that the read-through domain is not required for the virion to cross the salivary gland, but it does improve the success of the transport. If the species is from a yellow dwarf lineage, the virion then moves through the hindgut. If the species

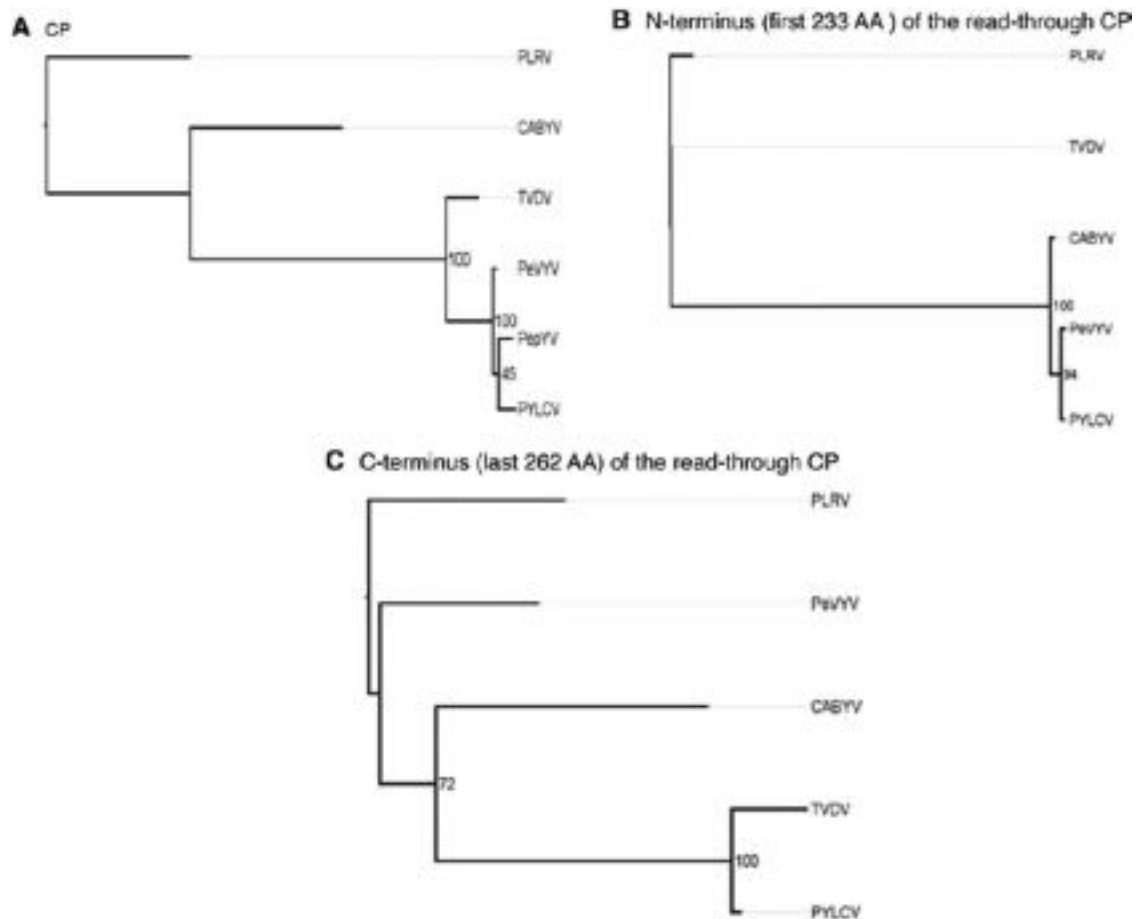


Fig. 3 Phylogenetic analysis of selected poleroviruses based on the following proteins: (A) CP (ORF3), (B) the N-terminus of the RTD (ORF5, first 233 aa), (C) the C-terminus of the RTD (ORF5, last 262 aa) of Potato leaf curl virus [PLRV (Y07496)], Tobacco vein-distorting virus [TVDV (EF529624)], Cucurbit aphid-borne yellows virus [CABYV (X76931)], Pepper vein yellows virus [PeVYV (AB5948280)], Pepper yellow leaf curl virus [PYLCV (HM439608)], and Pepper yellows virus [(PepYV) FN600344]. Adapted with permission from Dombrovsky, A., Glanz, E., Lachman, O., *et al.*, 2013. The complete genomics of Pepper yellow leaf curl virus (PYLCV) and its implications for our understanding of evolution dynamics in the genus Polerovirus. PLoS One 8, e70722. doi:10.1371/journal.pone.0070722.

mainly uses dicots as their host, such as PLRV or Beet western yellows virus (BWYV), the virion instead moves through the midgut. Once inside the gut, the virus normally moves between the cytoplasm and the epithelial cells. It then fuses with the plasmalemma and is released between the basal lamina and the membrane. Aphids can then release the virion into the phloem parenchyma and/or the companion cells to initiate local infection. Cell-to-cell and systemic infection may occur and require the combined activity of the movement protein, capsid protein, capsid protein read-through, and P0. Phloem-limited viruses cannot normally be mechanically transmitted. However, using particle bombardment, infection has been achieved with PLRV and BWYV. The high number of aphid vectors allows poleroviruses to infect a wide range of hosts.

Potatoes, sugarcane, and beets are important species infected by poleroviruses. These all propagate in a vegetative manner. Poleroviruses can spread through infected contaminated plants parts used for propagation, such as tubers and sugarcane cuttings.

Virus-Virus Interactions

Co-infections of PLRV with Potato virus X (PVX) (*Potexvirus*) or Potato virus Y (PVY) (*Potyvirus*) result in an enhancement in symptom severity and yield loss. Similarly, co-infection of Brassica yellows virus (BrYV) (unassigned *Polerovirus*) and Pea enation mosaic virus 2 (PEMV-2) (*Umbravirus*) results in similar synergism. PEMV 2 is an umbravirus that accompanies PEMV as a satellite virus. PEMV-2 can only infect plants when a member of the *Luteoviridae* is present. Co-infection of BrYV and PEMV-2 results in higher accumulation of BrYV, more severe symptoms, and the acquisition of mechanical transmission.

Co-infection of BWYV and the potyvirus Beet mosaic virus (BtMV) causes faster systemic virus movement and earlier, more severe symptoms. The combination of the polerovirus and the potyvirus disrupts photosynthesis and vascular transport, and both

viruses accumulate to high levels when co-infected. Experimentally, the potyvirus helper-component proteinase (HC-Pro), a strong silencing suppressor, increased the accumulation of PLRV. It also allowed the virus to spread into *mesophyll* cells.

Maize lethal necrosis is a re-emerging disease of current epidemic proportions in Sub-Saharan Africa. Maize lethal necrosis disease was discovered in 1976 in Nebraska and Kansas, USA. It was caused by Maize chlorotic mottle virus (MCMV) in combination with a potyvirus, Sugarcane mosaic virus (SCMV). Several recent studies have found poleroviruses in maize in combination with MCMV, SCMV, or both. The poleroviruses detected include the Maize yellow dwarf virus-RMV (MYDV-RMV), Maize yellow mosaic virus (MaYMV), and Barley virus G (BVG). However, MYDV-RMV was the most common. Plants with a combination of a polerovirus, SCMV, and MCMV have atypical symptoms. These and other observations suggest that poleroviruses, in combination with MCMV, could also cause maize lethal necrosis disease.

Silencing Suppression by Poleroviruses

In plants, gene silencing is an essential component of antiviral defense. Virus infection induces antiviral gene silencing. Argonaute (AGO) proteins are the catalytic components of the RNA-induced silencing complex (RISC) and associate with cellular or virus-derived small interfering RNAs (siRNA). Binary complexes formed between argonaute proteins and siRNAs specifically target RNA, including viral RNA, complementary to the siRNA. They have also been implicated in cell-to-cell and systemic movement of gene silencing signals. This results in amplification of gene silencing in areas beyond the initial activation site, thereby conferring virus immunity.

In order to establish infection and move within plants, viruses encode specialized proteins that suppress gene silencing. In poleroviruses, P0 is a silencing suppressor. P0 silencing suppression activity has been demonstrated for 10 of the 32 poleroviruses (Table 1). For Beet chlorosis virus (BChV), and some strains of Beet mild yellowing virus (BMV), no silencing suppression activity was found for P0. For all other poleroviruses, no information is available or the suppression activity of P0 has not been determined. In standard experimental assays, for some poleroviruses, P0 suppresses either local or both local and systemic gene silencing.

P0 suppresses gene silencing by targeting argonaute 1 (AGO1) protein for degradation by ubiquitination. Through the F-box-like motif, P0 interacts with the DUF1785 motif in AGO1 to mark it for degradation. In *Arabidopsis thaliana*, the P0 F-box-like motif interacts with the F-box of the S phase kinase-associated protein 1 (SKP1) and with the ASK1 and ASK2 orthologues. AGO1 is the primary interaction partner of microRNAs and is a crucial for normal plant development. Thus, symptoms induced by polerovirus infection are in part due to the effect on AGO1, and potentially other AGO proteins being tagged for degraded by P0 which in turn affects normal plant development.

Potato Leafroll Virus

PLRV is the first polerovirus discovered, one of the most damaging poleroviruses worldwide, and the most damaging potato virus. It is highly prevalent and has been found on every continent except Antarctica. PLRV was first detected in the 1770s, causes 50%–60% yield loss, and costs the United States 100 million-dollars yearly. PLRV is transmitted by infected tubers and by aphids. When the virus is transmitted by aphids, symptoms begin in young, top leaves that roll and turn pale. When grown from an infected tuber, the plants may be pale or dwarfed, and the leaves may be upright, rolled, yellow, or brittle (Fig. 4(A)). However, the appearance of water-soaked leaves is usually the first symptom. In the stem and the tuber sieve tubes, abnormal amounts of callose accumulates. The carbohydrates in the leaves reach high levels causing the phloem transport to be impaired, which results in tuber reduction. This could occur because photo-assimilation is reduced, sucrose is unable to enter the phloem, or a combination of the two. These factors result in leaves with an upright and rolled appearance. In some cultivars, the margins of the leaves may turn purple or red and develop necrosis in later stages. This necrosis starts in the phloem of the petioles and stems.

Sugarcane Yellow Leaf Virus

Worldwide, damage by poleroviruses in sugarcane is a close second to PLRV. ScYLV is a good representation of the yellow leaf or yellow dwarf viruses amongst poleroviruses. ScYLV was first discovered in Hawaii in 1989 and is distributed world-wide. Currently, it is primarily detected in South America, Asia, and the Pacific islands. ScYLV affects sugarcane production in over 90 countries, which grow sugarcane for sugar, biofuel, and fibers. Crops affected by ScYLV have losses that reach up to 43%, and it is spread by infected seed canes and aphids. ScYLV infection reduces the cane thickness, the number of canes produced, and the rate of photosynthesis in the plant. It causes yellowing at the midrib (Fig. 4(B)), and, at 6–8 months, the yellowing spreads laterally to the leaf lamina and causes necrosis at the tip. Because of the short internode spacing, this causes the plant to be dwarfed. Interestingly, when co-infected with a certain bacterium, *Leifsonia xyli* sub species *xyli*, it increases the severity of the disease. Even when the virus is latent in the plant, the yield is decreased, especially in non-resistant varieties. Unfortunately, all yellow leaf viruses are hard to distinguish from normal environmental damage.

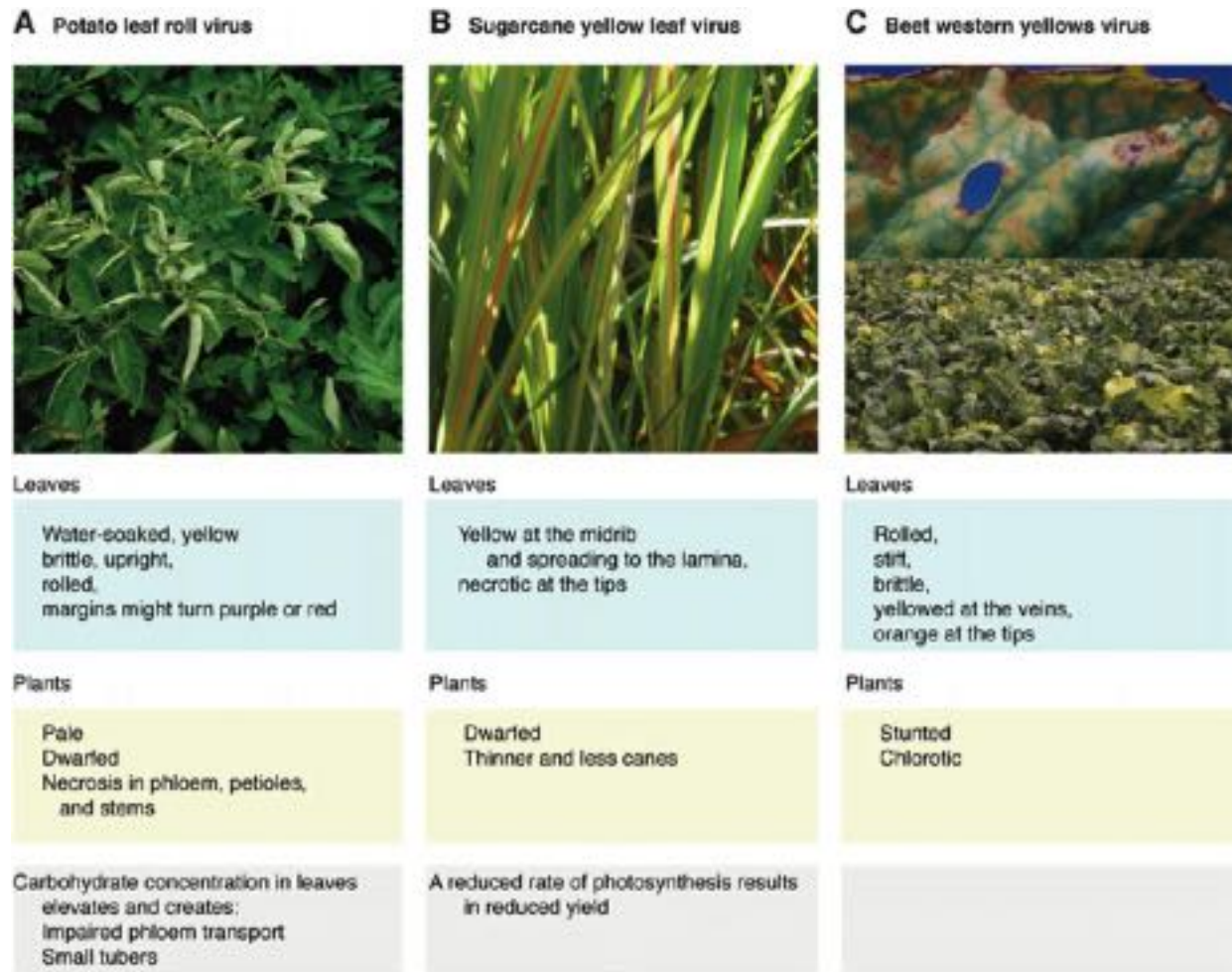


Fig. 4 Representative symptoms, in leaves and whole plants, caused by the top three poleroviruses. Other features of the symptoms are described below the images. (A) Symptoms caused by Potato leafroll virus in potato plants and leaves. (B) Symptoms caused by Sugarcane yellow leaf virus in sugarcane. (C) Symptoms caused by Beet western yellows virus in sugar beet. Reproduced with permission from (A) Jack Kelly Clark, University of California Statewide IPM Program. (B) CIRAD: The French Agricultural Research Organization working for the sustainable development of tropical and Mediterranean regions. (C) G.J. Holmes, California Polytechnic State University at San Luis Obispo.

Beet Poleroviruses

The three main poleroviruses infecting beet are BWYV, BMV, and BChV (**Table 1**). Most of the strains are different isolates of BWYV originating from Europe. Within the last decade, the virus has spread to Australia and resulted in a 26% yield loss in pluses, canola, and various vegetables. In canola, 59% of the 65% yield loss was due to BWYV alone. BMV is also known to result in about 22% crop loss if it appears in June. However, it has been found that these viruses have a much broader host range within the whole *Amaranthaceae* family, temperate legume crops, and brassicas. They also widely infect the weeds that grow around these crops. Of the beet-infecting poleroviruses, BChV has a smaller host range because it only infects sugar beets. These persistent viruses are transmitted by a wide variety of aphids, with the highest being the genus *Myzus*. Normally, sugar beets are durable crops that are tolerant to drought and can withstand intense wilting with no yield loss. However, poleroviruses cause stunting, chlorosis, rolling leaves, stiff followed by brittle leaves, and yellowing in the veins (**Fig. 4(C)**). Some leaves may turn orange rather than the typical yellow, which starts at the tip of the leaves. Beet-infecting poleroviruses remain problematic because of their impacts on crops, wide host range, and difficulty to diagnosis.

Diagnosis

There are several methods of detecting poleroviruses. The first test designed was the Ingel–Lange test. This test stained callose with a resorcin blue dye in the tubers of a potato infected with PLRV. Currently, poleroviruses are detected using a variety of RNA-based

and protein-based approaches. Protein-based approaches use antibodies that recognize the capsid protein. The most common protein-based approach is enzyme-linked immunosorbent assay (ELISA). A variation is the double antibody sandwich ELISA. Other protein-based approaches include immune-electron microscopy, immune-electrophoresis, and double diffusion agar tests.

RNA-based approaches include reverse-transcription polymerase chain reaction (RT-PCR), deep sequencing of the siRNAs, and high-throughput sequencing of transcript RNA. There are universal polerovirus primers for detection by RT-PCR. Universal primers Pol-G-F and Pol-G-R amplify a 1.4 kb PCR product spanning part of the RdRp gene, the intergenic region, and the complete CP gene. Additionally, northern blotting may be used to detect genomic and sub-genomic RNAs.

Electron microscopy is commonly used to visually detect poleroviruses. Using the transmission electron microscope, Cereal yellow dwarf virus (CYDV), BWYV, PLRV, and CABYV have been observed being transported through the gut and epithelial cells into the aphid. Electron microscopy was also used to track the readthrough domain to determine its role inside the aphid.

Disease Management

Polverovirus resistant plants are not common. In crops of economic importance, there are no polerovirus-resistant varieties. Varieties that do exist are only resistant to one or a few strains of the virus. Because resistant cultivars impose selection pressure and viruses mutate quickly, viruses break genetic resistance within a few years.

Since poleroviruses are vectored by aphids, physical barriers have been implemented to prevent polerovirus spread. Plastic reflective mulch can be placed around the crops. UV wavelengths will be reflected, thus repelling the aphids. Floating row covers with a fine mesh can also physically block the aphids from reaching the plant. Topical approaches include mineral oil and aphicides. Mineral oil can be applied to crops to smother aphids. Aphicides can be used on plants similar to chemical pesticides.

Carefully planned planting can reduce virus transmission by aphids. Planting when aphids are low, like after a short rain season, could potentially reduce the number of aphids in the field. This strategy can then be combined with timed pesticide use. Government restricted closed seasons for planting certain crops will also reduce polerovirus transmission. Several countries already restrict crop planting during certain times, but the approach could be further expanded around the globe. Farmers should also practice good crop rotation and diversification. A maize and soybean rotation will likely have different aphid vectors, so there is a decreased likelihood of continual crop infection. A less practiced method is to control the weeds in the field. Weeds that are not removed between crop planting could still be infected, thus leading to crop infection in the next season. It is also good practice to burn any infected plant material because it is the only surefire method to destroy the virus. This method also eliminates the possibility aphids may feed on infected plant material and spread the infection to healthy plants. Aphids are attracted to bare soil, so farmers should work to plant crops closer, include cover crops, and have untilled soil. Preventative measures provide options to limit poleroviruses exposure, but the possibility of resistant plants remains the only option to fight infection directly.

Further Reading

- Byrne, M.J., Steele, J.F.C., Hesketh, E.L., *et al.*, 2019. Combining transient expression and cryo-EM to obtain high-resolution structures of luteovirid particles. *Structure* 27, 1761–1770. doi:10.1016/j.str.2019.09.010.
- Chen, S., Jiang, G., Wu, J., *et al.*, 2016. Characterization of a novel polerovirus infecting maize in China. *Viruses* 8 (5), doi:10.3390/v8050120.
- DeBlasio, S.L., Chavez, J.D., Alexander, M.M., *et al.*, 2015. Visualization of host-polerovirus interaction topologies using protein interaction reporter technology. *Journal of Virology* 90 (4), 1973–1987. doi:10.1128/JVI.01706-15.
- DeBlasio, S.L., Xu, Y., Johnson, R.S., *et al.*, 2018. The interaction dynamics of two Potato leafroll virus movement proteins affects their localization to the outer membranes of mitochondria and plastids. *Viruses* 10. doi:10.3390/v10110585.
- Dombrovsky, A., Glanz, E., Lachman, O., *et al.*, 2013. The complete genomics of Pepper yellow leaf curl virus (PYLCV) and its implications for our understanding of evolution dynamics in the genus *Polverovirus*. *PLoS One* 8, e70722. doi:10.1371/journal.pone.0070722.
- Gray, S., Cilia, M., Ghanim, M., 2014. Circulative, non-propagative virus transmission: An orchestra of virus-, insect-, and plant-derived instruments. In: Maramorosch, K., Murphy, F.A. (Eds.), *Advances in Virus Research* 89. Elsevier, pp. 141–199. doi:10.1016/B978-0-12-800172-1.00004-5.
- Kaplan, I.B., Lee, L., Ripoll, D.R., *et al.*, 2007. Point mutations in the Potato leafroll virus major capsid protein alter virion stability and aphid transmission. *Journal of General Virology* 88 (6), 1821–1830.
- Knierim, D., Deng, T.C., Tsai, W.S., Green, S.K., Kenyon, L., 2010. Molecular identification of three distinct polerovirus species and a recombinant Cucurbit aphid-borne yellows virus strain infecting cucurbit crops in Taiwan. *Plant Pathology* 59, 991–1002.
- Krueger, E.N., Beckett, R.J., Gray, S.M., Miller, W.A., 2013. The complete nucleotide sequence of the genome of Barley yellow dwarf virus-RMV reveals it to be a new polerovirus distantly related to other yellow dwarf viruses. *Frontiers in Microbiology* 4, 205. doi:10.3389/fmicb.2013.00205.
- Lee, L., Palukaitis, P., Gray, S.M., 2002. Host-dependent requirement for the Potato leafroll virus 17kda protein in virus movement. *Molecular Plant-Microbe Interactions* 15, 1086–1094. doi:10.1094/MPMI.2002.15.10.1086.
- Massawe, D.P., Stewart, L.R., Kamatenesi, J., Asiimwe, T., Redinbaugh, M.G., 2018. Complete sequence and diversity of a maize-associated polerovirus in East Africa. *Virus Genes* 54, 432–437. doi:10.1007/s11262-018-1560-5.
- Pagan, I., Holmes, E.C., 2010. Long-term evolution of the *Luteoviridae*: Time scale and mode of virus speciation. *Journal of Virology* 84 (12), 6177–6187. doi:10.1128/JVI.02160-09.
- Peter, K.A., Liang, D., Palukaitis, P., Gray, S.M., 2008. Small deletions in the Potato leafroll virus readthrough protein affect particle morphology, aphid transmission, virus movement and accumulation. *Journal of General Virology* 89, 2037–2045. doi:10.1099/vir.0.83625-0.
- Rashid, M.-O., Zhang, X.-Y., Wang, Y., *et al.*, 2019. The three essential motifs in P0 for suppression of RNA silencing activity of Potato leafroll virus are required for virus systemic infection. *Viruses* 11 (2), 170. doi:10.3390/v11020170.

- Wintermantel, W.M., 2005. Co-infection of Beet mosaic virus with Beet yellowing viruses leads to increased symptom expression on sugar beet. *Plant Disease* 89 (3), 325–331. doi:10.1094/pd-89-0325.
- Xu, Y., Da Silva, W.L., Qian, Y., Gray, S.M., 2018. An aromatic amino acid and associated helix in the C-terminus of the Potato leafroll virus minor capsid protein regulate systemic infection and symptom expression. *PLoS Pathog* 14, e1007451. doi:10.1371/journal.ppat.1007451.
- Zhou, C.-J., Zhang, X.-Y., Liu, S.-Y., *et al.*, 2017. Synergistic infection of BrYV and PEMV-2 increases the accumulations of both BrYV and BrYV-derived siRNAs in *Nicotiana benthamiana*. *Scientific Reports* 7, 45132. doi:10.1038/srep45132.

Pomoviruses (*Virgaviridae*)

Eugene I Savenkov, Swedish University of Agricultural Sciences, Uppsala, Sweden and Linnean Center for Plant Biology, Uppsala, Sweden

© 2021 Elsevier Ltd. All rights reserved.

This is an update of L. Torrance, Pomovirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00543-4](https://doi.org/10.1016/B978-012374410-4.00543-4).

Nomenclature

aa Amino acid(s)

AGO Argonaute

CP Coat protein or capsid protein

ELISA Enzyme-linked immunosorbent assay

ER Endoplasmic reticulum

GFP Green fluorescent protein

ID Internal domain

kb Kilobase

kDa Kilodalton

MP Movement protein

mRFP Monomeric red fluorescent protein

NGS Next generation sequencing

nt Nucleotide(s)

NTD N-terminal domain

ORF Open Reading Frame

RdRp RNA-dependent RNA polymerase

RISC RNA-induced silencing complex

RTD Read-through domain

S Sedimentation coefficient; Svedberg units

SEL Size exclusion limit

siRNA Small interfering RNA

ssRNA Single-stranded ribonucleic acid

TGB Triple gene block

tRNA Transfer RNA

UTR Un-translated region

vRNA Viral RNA

vRNP Viral ribonucleoprotein

VSR Viral suppressor of RNA silencing

Glossary

Spore balls Resting spores (cystosori) of *Spongospora subterranea* found in lesions or pustules (powdery scabs) on potato tubers; a single cystosorus comprises 500–1000

resting spores aggregated to form in a ball that is partially hollow, traversed by irregular channels.

Spraing Virus disease symptoms of internal brown lines and arcs in potato tuber flesh.

Introduction

Pomovirus is a genus of plant viruses assigned to the family *Virgaviridae*. Pomoviruses have tubular rod-shaped particles and tripartite genomes; they are transmitted by soil-borne zoosporic organisms belonging to two genera (*Polymyxa* and *Spongospora*) in the family *Plasmodiophoraceae*. Pomoviruses have limited host ranges, infecting species in a few families of dicotyledonous plants. Agriculturally important hosts include potato and sugar beet. There are six member species in the genus *Pomovirus*: *Potato mop-top virus* (PMTV), *Beet soil-borne virus* (BSBV), *Beet virus Q* (BVQ), *Broad bean necrosis virus* (BBNV), *Colombian potato soil-borne virus* (CPSbV) and a tentative member *Soil-borne virus 2* (SbV2). BVQ and BSBV can occur in mixed infections with the benyvirus *Beet necrotic yellow vein virus* (BNYVV).

Taxonomy and Classification

The classification of tubular rod-shaped viruses transmitted by soil-borne plasmodiophorid vectors was revised in 1998 to establish four genera: *Furovirus*, *Pomovirus*, *Benyvirus*, and *Pecluvirus*. The revision was prompted by new virus sequence information that revealed major differences in genome properties (number of RNA segments, sequence, and genome organization). The genus *Pomovirus* is assigned to the family *Virgaviridae*. *Pomovirus* is a siglum from *Potato mop-top virus*, the type species.

Physical Properties of Particles

Pomovirus particles are hollow, helical rods, 18–20 nm in diameter, comprising multiple copies of a single major coat protein (CP; *c.* 19–20 kDa). The CP gene is terminated by a UAG (or UAA in BBNV) stop codon that is thought to be suppressible, readthrough of which would produce a fusion protein with a read-through domain (RTD) of variable mass (54–104 kDa), CP-RTD. One or a few copies of the CP-RTD fusion protein are present at the extremity of PMTV particles and thought to contain the 5' end of the virus RNA. *Pomovirus* particles are fragile and particle size distribution measurements are variable; PMTV

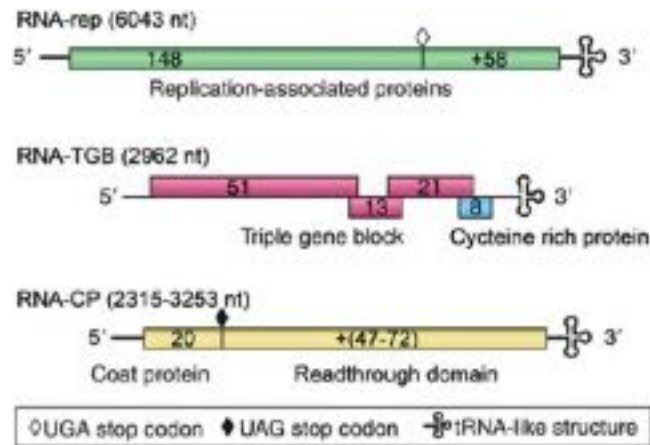


Fig. 1 Diagram of the PMTV genome organization; boxes indicate ORFs with the molecular masses of predicted protein products (kDa) indicated within.

particles have predominant lengths of 125, 137, and 283 nm. PMTV particles sediment as three components with sedimentation coefficients ($S_{20,W}$) of 126, 171, and 236S.

Genome Properties and Functions of the Encoded Proteins

Complete genome sequences are available for four member species. An almost complete sequence (without the 5' and 3' untranslated regions (UTRs)) is available for BBNV. Complete sequences of two genomic components, namely, RNA-rep and RNA-TGB are available for SbV2. Pomovirus genomes comprise three species of positive-sense single-stranded RNA (ssRNA) of c. 5.8–6.2, 2.8–3.4, and 2.3–3.1 kb (Fig. 1). RNA-rep encodes the replicase proteins. It contains a large open reading frame (ORF) that is interrupted by a UGA stop codon (ORF1) (or UAA in BVQ and BSBV; UAG in SbV2); the sequence continues in-phase to encode a readthrough protein (204–207 kDa). The protein encoded by ORF1 (145–149 kDa) contains methyl-transferase and helicase motifs while the RTD contains the GDD RNA-dependent RNA polymerase (RdRp) motif. Phylogenetic analysis reveals that the RdRps of the pomoviruses and the furovirus Soil-borne wheat mosaic virus share between 50% and 60% sequence identity. Phylogenetic analysis of available PMTV RNA-rep sequences reveals two major clades. The isolates from Peru, Colombia, Europe, China, Canada and USA group together in clade I, whereas clade II is exclusively represented by Colombian isolates.

The 5' UTR of pomovirus RNAs contains the starting sequence GU(A)_{1–4}(U)_n (except BVQ RNA-rep which begins with AUA, CPSbV RNA-rep and RNA-CP which begin with ACAT(G)_{4–5}TATT, and CPSbV RNA-TGB as well as RNA-rep and RNA-TGB of SbV2 have a different sequence at the 5' end). The RNAs are probably capped at the 5' end since the RNA-rep-encoded viral replicase contains methyl-transferase motifs associated with capping activity. The terminal 80 nt of the 3' UTR can be folded into a tRNA-like structure that contains an anticodon for valine. Both BSBV and PMTV RNAs were shown to be valylated experimentally. Different from other five pomoviruses, a tRNA-like structure of CPSbV genomic components contains an anticodone for leucine.

The virus movement proteins (MPs) are encoded on RNA-TGB of pomoviruses. Three 5'-proximal overlapping ORFs encode a conserved module of movement proteins known as the triple gene block (TGB). TGB movement proteins are found in the genomes of other rod-shaped viruses in the families *Virgaviridae* (hordei-, gora- and pecluviruses) and *Benyviridae* (benyvirus) and in some genera of monopartite filamentous viruses in the families *Alphaflexiviridae* and *Betaflexiviridae*. The TGB proteins (TGB1, TGB2, and TGB3) are named according to the position of their genes on the RNA and have molecular masses of 48–54, 13, and 20–22 kDa, respectively. The TGB1 contains three structurally distinct domains: an unstructured intrinsically disordered N-terminal domain (NTD), an ordered internal domain (ID) and a helicase domain (belonging to the viral superfamily 1 RNA helicases). TGB1 NTDs of different pomoviruses do not share obvious sequence similarity. TGB1 binds RNA and is thought to interact with genomic RNAs to facilitate movement. The NTD contains two nucleolar localization signals, interacts with nuclear transport receptors of Importin- α family and accumulates in the nucleoplasm and in the nucleolus. The nucleolar passage of PMTV TGB1 is needed for the virus systemic movement. The half of ID of PMTV TGB1 is predicted to form an α -helix that seems to mediate oligomerisation of TGB1 through self-interaction. The ability for self-interaction (TGB1 oligomerisation) is crucial for the cell-to-cell movement of PMTV. PMTV TGB1 also interacts (through its helicase domain) with a read-through domain (RTD) of CP-RTD at one extremity of the virions assisting in the systemic movement of the virus particles.

The sequence of the second TGB protein is the most conserved with BSBV, BVQ, and PMTV sharing 63%–75% nt sequence identity and 49% identity with that of BBNV; there is little sequence identity among the TGB3 sequences. Analysis shows that TGB2 and TGB3 proteins contain two hydrophobic regions (predicted transmembrane domains) separated by a hydrophilic domain and these proteins are associated with intracellular membranes in infected plants. PMTV TGB2 associates with the membranes of endoplasmic reticulum (ER), mobile granules in the cytoplasm, components of endosomal network, and the outer

membrane of chloroplasts. PMTV TGB2 also binds RNA in a sequence non-specific manner, but is not required for RNA replication, although it was suggested that TGB2 might play a role in delivering the viral RNA to the chloroplast membranes for replication. PMTV TGB2 can increase permeability of plasmodesmata (so-called the size exclusion limit, SEL) and together with TGB3 (which also increases SEL of plasmodesmata) is thought to facilitate movement of the viral ribonucleoprotein (vRNP) complexes through plasmodesmata to the neighboring cells (cell-to-cell movement).

At the subcellular level, TGB3 associates with the ER, motile granules, stationary aggregates at the periphery of the cell, termed “peripheral bodies”, and plasmodesmata. Similar to TGB2, the TGB3 protein of pomoviruses contains two trans-membrane domains and embedded into the membranes in U-shaped orientation with its central hydrophilic domain protruding into the ER lumen and N-terminus located in the cytosol. TGB3 contains a conserved tyrosine-based motif YQDLN involved in targeting TGB3 to plasmodesmata.

In PMTV and BBNV, a fourth small ORF of RNA-TGB is predicted to encode an 8 kDa cysteine-rich (8K) or 6 kDa glycine-rich protein, respectively, whereas no such ORF is present in BSBV, BVQ, CPSbV or SbV2. The 8K protein is classified as a zinc-finger protein with a putative SWIM zinc-finger motif characterized by a consensus sequence CxC_xCxC, in which C stands for cysteine residues and x refers to any aa residue. The 8K protein of PMTV is not needed for the virus movement or infection of *Nicotiana benthamiana*. However, 8K is essential for the efficient virus accumulation and the appearance of the viral symptoms is delayed in the absence of the 8K expression. The 8K protein functions as a viral suppressor of RNA silencing (VSR). Mutations resulting in substitution of cysteine residues with alanines within a putative SWIM zinc-finger motif abolish 8K VSR activity. Comparison of the 8K sequences from more than 80 isolates reveals that the 8K protein sequences show an extraordinary variability with 23 variable sites per 68-aa-residues-long protein. Moreover, the 8K ORF is under positive selection, whereas other PMTV ORFs are under neutral or negative selection. The extraordinary variability and positive selection seem to drive efficient RNA silencing suppression activity of the 8K protein as the 8K protein encoded by certain PMTV isolates has stronger RNA silencing suppression activity compared those encoded by other isolates. Site-directed mutagenesis identified a serine residue (Ser-50) as being essential for enhanced VSR activity of one natural variant of 8K. Next generation sequencing (NGS) of small RNA populations in the presence of the 8K proteins with contrasting abilities to suppress RNA silencing (strong versus weak) reveals a lower abundance of 21 nt and 22 nt small interfering RNAs (siRNAs) with uracil (U) residue at the 5' end after expression of relatively strong 8K natural variant compared to that of a weak 8K allele. As accurate sorting of siRNAs into ARGONAUTE (AGO)-containing antiviral RNA-induced silencing complex (RISC) – with AGOs preferentially binding to siRNA with certain residue at the 5' terminus – is important for antiviral defense, it has been speculated that the natural variant of 8K with relatively strong RNA silencing suppression activity may destabilize either siRNAs with U residue at the 5' end or complexes into which such siRNAs are sorted, i.e., AGO1-containing RISC that play a primary role (along with AGO2) in antiviral RNA silencing. The sequence of the 8K ORF along with the beginning of the sequence of TGB1 ORF is involved in biogenesis of a defective interfering (DI) RNA. DI RNA is readily detectable, accumulates at high levels in the virus-infected plants, and interferes with efficient virus accumulation. The disruption of a predicted minus-strand stem-loop secondary structure comprising complementary sequences of the 5' TGB1 ORF and the 3' 8K ORF results in inhibition of DI RNA production suggesting intramolecular template switching during positive-strand synthesis as a possible mechanism for the DI RNA biogenesis.

Phylogenetic analysis of available PMTV RNA-TGB sequences reveals two major clades. The isolates from Peru, Colombia, Europe, China, Canada and USA group together in clade I, whereas clade II is represented by a Peruvian isolate.

The CP and CP-RTD proteins are encoded on RNA-CP. The PMTV RNA-CP is of variable size, from 2315 to 3252 nt. The longest RNA-CP (3252 nt) is found in PMTV isolate J20 from Cajamarca, Peru. Compared to other PMTV isolates, isolate J20 has additional sequences in 3'-UTR. In most PMTV isolates, RNA-CP is 3134 nt in length. Analysis shows that RTD of PMTV and BVQ contains two transmembrane domains, TM1 and TM2, suggesting that CP-RTD is a membrane protein integrated into the lipid bilayer in a U-shaped orientation. Indeed, PMTV CP-RTD localizes to cell periphery probably through association with plasma membrane. Analysis of naturally occurring and glasshouse-propagated isolates revealed that deletions of c. 500–1000 nt occur in both field and laboratory isolates and that the deletions occur predominantly in RTD, particularly in the region toward the 3' end of the RTD ORF. Deletions in this region are correlated with loss of transmission by the natural vector *Spongospora subterranea*. Hydrophobicity analysis shows that these in-frame deletions result in the loss of the transmembrane domain TM2.

The occurrence of encapsidated deleted forms of RNA-CP and RNA-TGB in natural and laboratory isolates of PMTV has been described and sequence analysis as well as mutagenesis (for RNA-TGB) suggests that these PMTV RNAs contain sites that are susceptible to recombination possibly through a template switching mechanism. Variable base composition is found in natural isolates of BSBV in the sequence at the 3' end of RNA-TGB between the stop codon of the third TGB and the terminal tRNA-like structure.

Phylogenetic analysis of available PMTV RNA-CP sequences reveals three major clades. The isolates from Peru, Colombia, Europe, China, Canada and USA group together in clade I, whereas clade II and clade III are exclusively represented by Peruvian isolates.

Virus–Host Interactions and Movement

Cytoplasmic inclusions of enlarged ER and the accumulation of distorted membranes and small virion bundles can be seen by electron microscope examination of thin sections of BSBV- and BVQ-infected leaves. In PMTV-infected potato leaves, abnormal chloroplasts with cytoplasmic invaginations were seen in thin sections as well as tubular structures in the cytoplasm associated with the ER and tonoplast, and in the vacuole.

PMTV does not require CP for movement and it is thought that TGB1 interacts with viral RNA forming a movement competent ribonucleoprotein (RNP) complex. Studies of transiently expressed PMTV TGB proteins fused to marker proteins such as green fluorescent protein (GFP) or monomeric red fluorescent protein (mRFP) in epidermal cells of *N. benthamiana* have helped to elucidate events in intracellular trafficking and indicate that PMTV interacts with the cellular membrane recycling system. PMTV TGB2 and TGB3 were shown to co-localize on the ER (Fig. 2(a)–(c)) and in small motile granules that utilize the actin–ER network to reach the cell periphery and plasmodesmata and TGB3 contains a putative tyrosine sorting signal (YQDLN), mutation of which inhibits TGB3 localization to plasmodesmata. TGB2 and TGB3 act together to transport GFP-TGB1 to the plasmodesmata for movement into neighboring cells and TGB2 and TGB3 have the capacity to gate the plasmodesmata pore. TGB2 co-localizes in endocytic vesicles with the Rab 5 ortholog Ara 7 (AtRabF2b) that marks the early endosomal compartment. Also, protein–interaction analysis revealed that recombinant TGB2 interacted with a tobacco protein belonging to the highly conserved RME-8 (receptor mediated endocytosis-8) family of J-domain chaperones, essential for endocytic trafficking.

The systemic movement of many plant viruses occurs as virions that require expression and accumulation of CP in the infected cells. Yet, a small group of viruses including PMTV can move systemically in plants in the absence of CP as an RNP complex formed by interaction of the TGB1 protein with viral RNA. Moreover, mutational analysis revealed an essential role of CP-RTD in the long-distance movement of RNA-CP because in the absence of the CP-RTD expression PMTV CP inhibits systemic movement of RNA-CP, whereas the movement of RNA-TGB and RNA-rep is not affected. CP-RTD is needed for PMTV transmission in soil by its soil-borne vector *Spongospora subterranea* and minor amounts of CP-RTD molecules are incorporated into one terminus of the virus particles forming polar virions. Mutational analysis identified a CP-RTD domain consisting of 140 aa residues near the C-terminus as required for RNA-CP long-distance transport. Moreover, this particular domain is involved in the interaction of CP-RTD with TGB1 suggesting long-distance movement of polar virions that contain CP-RTD and TGB1 at the tip. Thus, the structure of movement-competent PMTV virions appears to be quite complex and involves at least two supplementary proteins at the particle tip.

Since PMTV TGB1 plays indispensable role in the movement of both viral RNPs (vRNPs) and polar virions identification of the host proteins as well as sub-cellular domains that interact with TGB1 has been crucial to understanding the mechanism of PMTV movement in plants. At early stages of virus infection (at the very front of infection foci) PMTV TGB1 is localized to plasmodesmata consistent with the central role of the protein in the virus movement. In addition to localization in plasmodesmata TGB1 is observed in the nucleoplasm, nucleolus and in association with microtubules (Fig. 2(d)) at later stages of virus infection (two-four cell behind the infection front). The nuclear/nucleolar localization of PMTV TGB1 is mediated by interaction of the protein N-terminus with a nuclear transport receptor Importin- α . Disruption of TGB1 nucleolar localization (through mutagenesis) or knock down of Importin- α expression or disruption of microtubules by chemical inhibitors does not affect the virus cell-to-cell movement but inhibits its systemic spread.

As the N terminus of TGB1 appears to be important for the virus systemic movement and such movement correlates with the association of TGB1 with the nucleolus and microtubules, *Nicotiana benthamiana* cDNA library was panned using a yeast two-hybrid screen to identify host proteins that interact with the TGB1 protein. The screen yielded HIPP26, a metallochaperone, as TGB1 interacting partner. The HIPP26 gene belongs to the large gene family that is unique in vascular plants. The HIPP proteins act in the heavy metal homeostasis (hence the name: Heavy metal-associated Isoprenylated Plant Protein, HIPP), involved in the regulation of transcriptional response to the biotic and abiotic stress and acts in a signal transduction as a plasma membrane-to-nucleus relay during abiotic stress. HIPP26 is post-translationally modified through lipidation (S-acylation and prenylation). Knockdown of HIPP26 expression inhibits virus long-distance movement of both virions (when all three RNA genomic components are used as inoculum) and vRNPs (produced by inoculation of RNA-rep and RNA-TGB only) but does not affect cell-to-cell movement of the virus. HIPP26 is upregulated in plants subjected to drought as well as in PMTV-infected plants. Moreover, PMTV infection protects plants from drought stress. Analysis of the spatial expression pattern of HIPP26 using HIPP26 promoter-reporter fusions reveals its vascular tissue-specific expression consistent with the role of HIPP26 in systemic movement of the virus. Biochemical and mutational analyses show that HIPP26 subcellular localization at the plasmodesmata and plasma membrane is mediated by post-translational modifications through lipidation (S-acylation and prenylation), as nonlipidated HIPP26 is mostly found in the nucleus. Furthermore, HIPP26 is predominantly accumulates in the nucleus when co-expressed with TGB1. TGB1 interacts with the C-terminal ¹⁵²CVVM¹⁵⁵ (prenyl) domain of HIPP26, and the TGB1-HIPP26 complex decorates microtubules and accumulates in the nucleolus. Overall, these data support a mechanism in which interaction TGB1 with HIPP26 results in the inhibition of the HIPP26 protein lipidation or/and de-lipidation of already modified HIPP26, thus releasing membrane-associated HIPP26 from plasma membrane into the cytosol. The HIPP26/TGB1 complex, than, localizes to microtubules and subsequently directed to the nucleus and nucleolus in Importin- α -dependent manner. In the nucleus HIPP26 activates the drought stress response to facilitate virus long-distance movement.

Host Range, Geographical Distribution, Transmission by Vector and Variability

Pomoviruses have a limited host range and are transmitted in soil by zoosporic plasmodiophorid vectors that have been classified as protists (Fig. 3). Viruses that are transmitted by plasmodiophorid vectors include pomo-, peclu-, furo-, beny- and bymoviruses and they are thought to be carried within the zoospores. The vector life cycle includes production of environmentally resistant thick-walled resting spores. The viruliferous resting spores can survive in soil for decades.

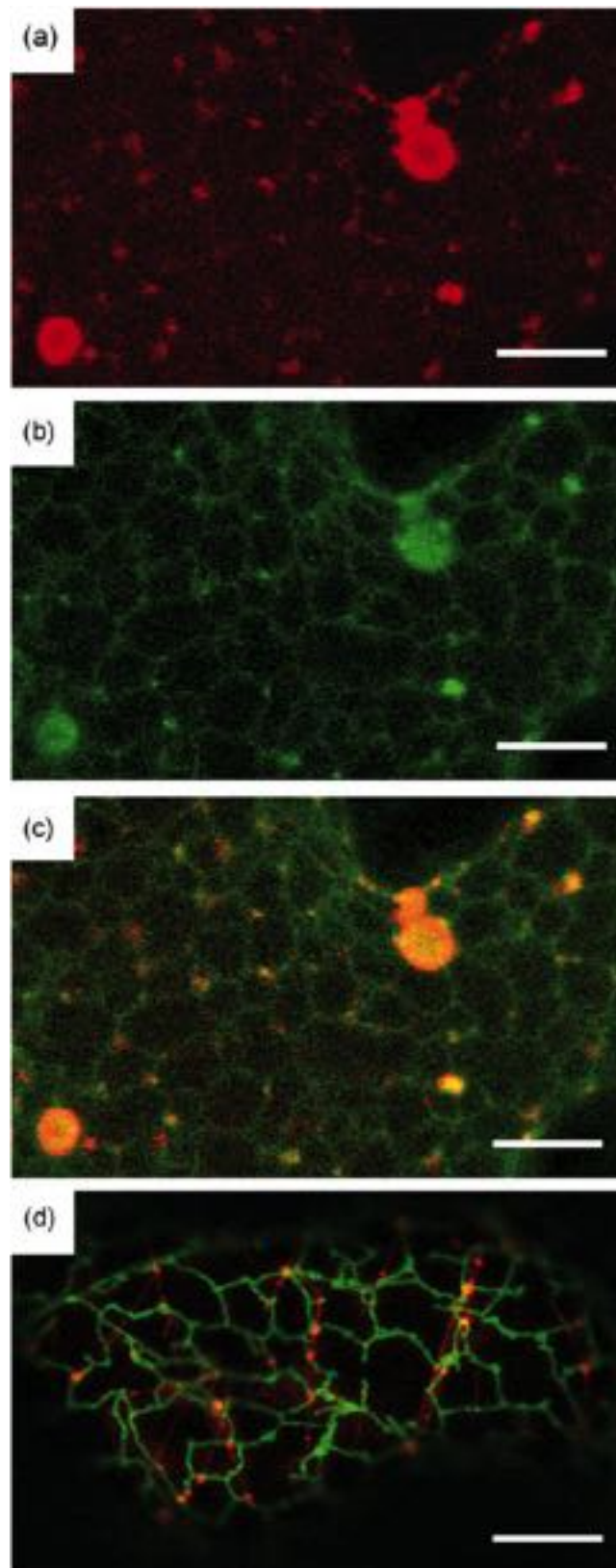


Fig. 2 Confocal laser scanning microscope images of PMTV TGB2 and TGB3 (a–c) and TGB1 (d) fluorescent fusion proteins in epidermal cells of *Nicotiana benthamiana*. Monomeric red fluorescent protein (mRFP) tagged PMTV-TGB2 was transiently expressed with green fluorescent protein (GFP) tagged PMTV-TGB3, the proteins co-localized on membranes of the ER and in motile granules seen moving on the ER network. (a) Expression of mRFP-TGB2 (red channel); (b) expression of GFP-TGB3 (green channel); (c) merged image of (a) and (b). (d) Merged image of the expression of yellow fluorescent (YFP) tagged TGB1 (red channel) from modified PMTV in *N. benthamiana* plants expressing ER-tagged GFP (green channel). At later stages of virus infection TGB1 does not associate with ER (green channel) anymore and mostly localizes on microtubules (red channel). Scale = 10 nm. (a–c) Courtesy Lesley Torrance, The James Hutton Institute, UK.

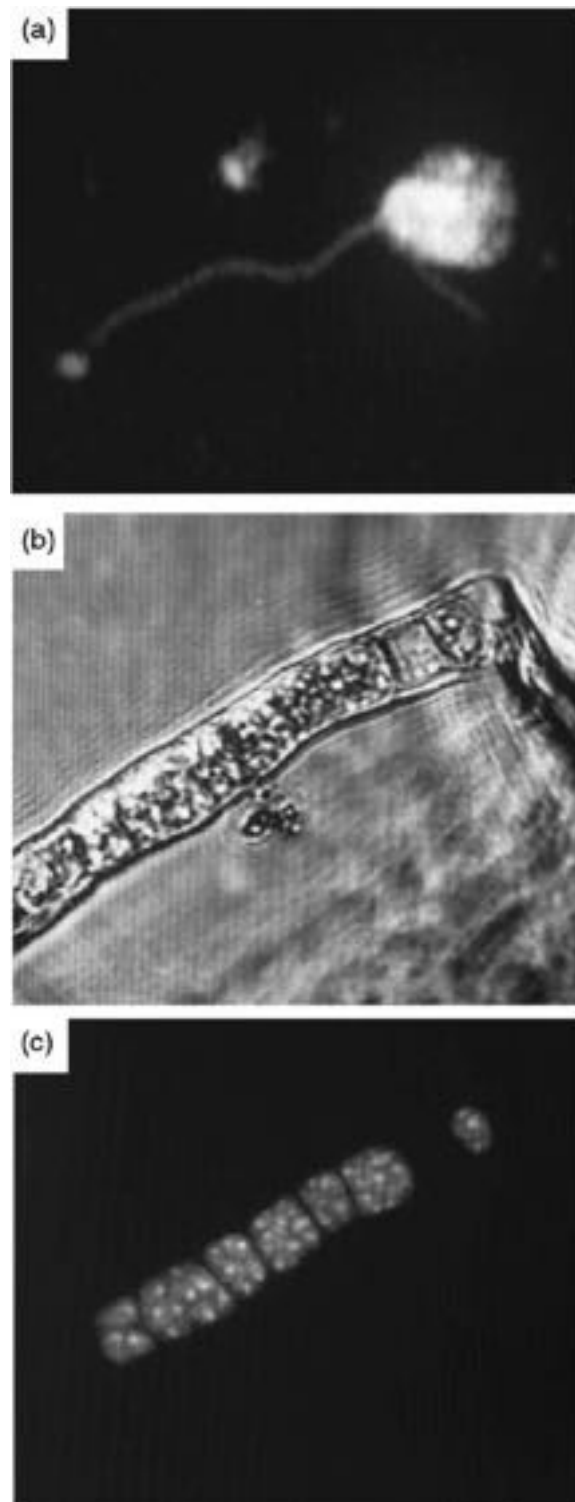


Fig. 3 (a) Biflagellate zoospore of *Spongospora subterranea*. (b) Bright field and (c) fluorescence microscope images of zoosporangia in tomato root hair. Courtesy Lesley Torrance, The James Hutton Institute, UK.

BBNV has been reported only from Japan; it causes necrosis and stunting in broad beans and peas. It is mechanically transmitted by inoculation of sap to a few species including *Vicia faba*, *Pisum sativum*, and *Chenopodium quinoa*.

BVQ and BSBV are often found associated with BNYVV; BVQ is reported only from Europe whereas BSBV is found in sugar-beet-growing areas worldwide. No symptoms have been attributed to BSBV or BVQ alone in sugar beet. BSBV is ubiquitous virus and is often

found in the soils in sugar-beet growing areas, but the virus alone does not cause any disease in sugar beet. The viruses are thought to be transmitted in soil by *Polymyxa betae*. BVQ is mechanically transmissible only to *C. quinoa* and BSBV to members of the *Chenopodiaceae*.

PMTV is reported to have originated in the center of potato domestication – the Andean region of South America – and has now found in potato-growing regions of Europe, North and South America, Asia and recently in New Zealand; virus incidence is favored by cool, wet growing conditions. It is transmitted in soil by *S. subterranea*, also a potato pathogen which causes the tuber blemish disease powdery scab. PMTV can be transmitted mechanically to members of the *Solanaceae* and *Chenopodiaceae*.

CPSbV and SbV2 has been reported only from Colombia; the viruses seem to cause symptomless infections in potato. The viruses are thought to be transmitted in soil by *S. subterranea*. CPSbV and SbV2 are mechanically transmitted by inoculation of sap to members of the *Solanaceae* and *Chenopodiaceae*.

The complete and partial genomic sequences are available for numerous PMTV isolates from around the globe. Based on the CP-RTD sequence, two major strains of PMTV are recognized and coined as S (for “severe”) and M (for “mild”). Seven signature aa residues differ between the S (RRSIGAV) and the M (KKNVREM) strains in the CP-RTD protein encoded by RNA-CP. The M strain is only reported from Peru, whereas the S strain is found worldwide including Peru. Phylogenetic analysis identified two major genotypes (clades) for RNA-rep (I and II), two major genotypes for RNA-TGB (I and II), and three genotypes for RNA-CP (I, II and III). As PMTV has a tripartite genome, a classification of PMTV isolates based on genotype constellations has been suggested. In this classification, the genotype of each genomic RNA is taken into consideration. Four genetic constellation has been found so far. Two genetic constellations – I + I + II and UN + II + III (RNA-rep + RNA-TGB + RNA-CP, “UN”, indicates unknown genotype of the RNA-rep segment, for which sequence is not available) – are exclusively found in Peru. Another genetic constellation (II + I + I) is found in Colombia, whereas the constellation I + I + I is found worldwide including Peru and Colombia. These observations suggest that this particular constellation (I + I + I) originates from the Andean region of Peru, the center of potato domestication, and was spread to the other parts of the world. In general, greatest diversity of PMTV is found in the Andean region of South America, whereas the virus has very little variability in the rest of the world.

Serological Relationships, Diagnosis, and Control

The viruses are serologically distinct. Distant relationships have been reported between BSBV and BVQ; PMTV and Soil-borne wheat mosaic virus (*Furovirus*); and PMTV, BBNV, and tobamoviruses. PMTV, BSBV, and BVQ CPs contain a conserved sequence (SALNVAHQ) that reacts in Western blots with a monoclonal antibody (SCR70) produced against PMTV, but that is not exposed on intact particles. PMTV particles contain an immunodominant epitope at the N-terminus of the CP that is exposed at the surface along the sides of the particles and can be detected by monoclonal antibody SCR69 (Fig. 4).

The viruses can be detected by serological tests (immunosorbent electron microscopy, enzyme-linked immunosorbent assay), by assays based on the reverse transcriptase-polymerase chain reaction (RT-PCR) and by quantitative real-time RT-PCR (RT-qPCR) in leaves, roots, or tubers from naturally infected plants. PMTV is known to be erratically distributed in potato leaves and tubers and can move systemically in the absence of CP in potato leaves which raises a risk of false negative diagnosis. The problem can be overcome by using RT-PCR or RT-qPCR for detection of RNA-TGB and/or RNA-rep as RNA-CP might be absent in those samples.

PMTV causes an economically important disease affecting the quality of tubers grown for the fresh and processing markets. Tubers of susceptible potato cultivars that are infected from soil during the growing season develop “spraing” symptoms that include unsightly brown lines, arcs, or marks in the flesh sometimes accompanied by slightly raised external lines and rings (Fig. 5(a)). Potato plants grown from infected tubers display V-shaped yellow markings or chevrons on the leaves (Fig. 5(b)) and may have shortened internodes (mop-top) producing cracked or malformed tubers; both tuber quality and yield can be affected. However, the virus does not infect all plants grown from infected tubers. Haulm and tuber symptoms vary markedly with cultivar, and some cultivars are symptomlessly infected. Environmental conditions also affect disease incidence and severity, and PMTV

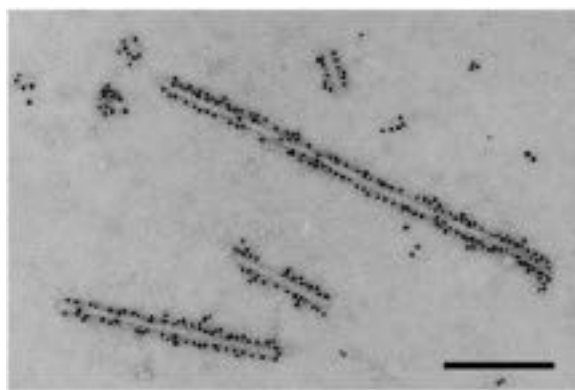


Fig. 4 Electron micrograph of PMTV particles labeled with monoclonal antibody SCR69/gold conjugate. Courtesy Lesley Torrance, The James Hutton Institute, UK.

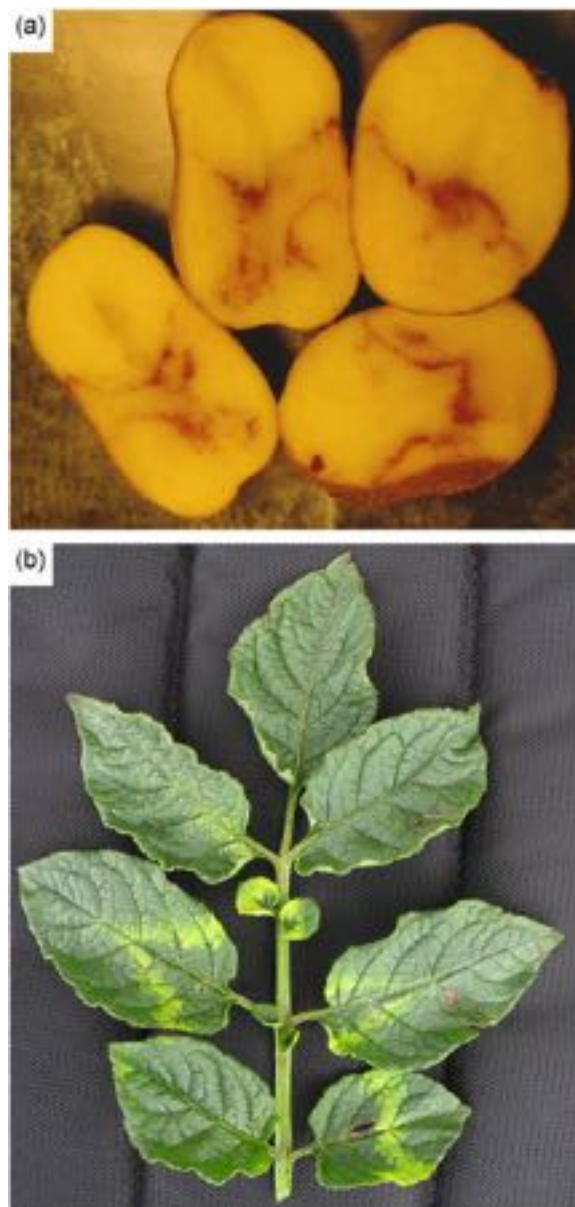


Fig. 5 (a) Symptoms of PMTV in potato tubers cv. Saturna. (b) Chevron-like symptoms of PMTV on leaves of potato cv. Qompis.

incidence was shown to increase with annual rainfall. Soil temperatures of 12–17°C and high soil moisture at tuber initiation favor powdery scab incidence.

PMTV can be established at new sites by planting infected tubers and once established, PMTV is a persistent problem, as the resting spore balls of the vector *S. subterranea* are long-lived and resistant to drought and agrochemicals. Viruliferous spore balls are spread readily to new sites by farm vehicles, contaminated seed tubers, wind-blown surface soil; motile zoospores can be spread through contaminated irrigation or drainage water. In certain areas, potatoes have become infected by PMTV 18 years after potatoes were last grown (the longest period recorded). There is no effective practicable means to control *S. subterranea* but decreased severity of powdery scab can be achieved by application of fluazinam to soil. In addition, disinfection of tubers with chemicals such as formaldehyde decreases virus incidence, although the efficacy of this treatment depends on the level of soil infestation where the tubers are planted and a combination of tuber and soil treatments may be more effective.

The best prospect for virus disease control is development of resistant cultivars but there are no known sources of PMTV resistance in commercial potato cultivars and most of the commercially grown cultivars are also susceptible to *S. subterranea*. However, plants transformed with virus-derived transgenes (*CP* and a mutated form of *TGB2*) have exhibited resistance to PMTV with decreased virus accumulation and incidence in tubers of plants grown in infested soil.

Further Reading

- Cowan, G.H., Roberts, A.G., Jones, S., *et al.*, 2018. Potato mop-top virus co-opts the stress sensor HIP26 for long-distance movement. *Plant Physiology* 176, 2052–2070.
- Gil, J.F., Adams, I., Boonham, N., Nielsen, S.L., Nicolaisen, M., 2016. Molecular and biological characterization of two novel pomovirus-like viruses associated with potato (*Solanum tuberosum*) fields in Colombia. *Archives of Virology* 161, 1601–1610.
- Haupt, S., Cowan, G.H., Ziegler, A., *et al.*, 2005. Two plant-viral movement proteins traffic in the endocytic recycling pathway. *Plant Cell* 17, 164–181.
- Kalyandurg, P., Gil, J.F., Lukhovitskaya, N.I., *et al.*, 2017. Molecular and pathobiological characterization of 61 Potato mop-top virus full-length cDNAs reveals great variability of the virus in the center of potato domestication, novel genotypes and evidence for recombination. *Molecular Plant Pathology* 18, 864–877.
- Kalyandurg, P.B., Tahmasebi, A., Vetukuri, R.R., *et al.*, 2019. Efficient RNA silencing suppression activity of Potato Mop-Top Virus 8K protein is driven by variability and positive selection. *Virology* 535, 111–121.
- Lukhovitskaya, N.I., Cowan, G.H., Vetukuri, R.R., *et al.*, 2015. Importin- α mediated nucleolar localization of Potato mop-top virus TRIPLE GENE BLOCK1 (TGB1) protein facilitates virus systemic movement, whereas TGB1 self-interaction is required for cell-to-cell movement in *Nicotiana benthamiana*. *Plant Physiology* 167, 738–752.
- Lukhovitskaya, N.I., Thaduri, S., Garushyants, S.K., Torrance, L., Savenkov, E.I., 2013. Deciphering the mechanism of defective interfering RNA (DI RNA) biogenesis reveals that a viral protein and the DI RNA act antagonistically in virus infection. *Journal of Virology* 87, 6091–6103.
- Pereira, L.G., Torrance, L., Roberts, I.M., Harrison, B.D., 1994. Antigenic structure of the coat protein of Potato mop-top virus. *Virology* 203, 277–285.
- Reavy, B., Arif, M., Cowan, G.H., Torrance, L., 1998. Association of sequences in the coat protein/read-through domain of Potato mop-top virus with transmission by *Spongospora subterranea*. *Journal of General Virology* 79, 2343–2347.
- Rochon, D., Kakani, K., Robbins, M., Reade, R., 2004. Molecular aspects of plant virus transmission by oleriid and plasmodiophorid vectors. *Annual Review of Phytopathology* 42, 211–241.
- Sandgren, M., Savenkov, E.I., Valkonen, J.P.T., 2001. The readthrough region of Potato mop-top virus (PMTV) coat protein encoding RNA, the second largest RNA of PMTV genome, undergoes structural changes in naturally infected and experimentally inoculated plants. *Archives of Virology* 146, 467–477.
- Solovyev, A.G., Savenkov, E.I., 2014. Factors involved in systemic transport of plant RNA viruses, the emerging role of the nucleus. *Journal of Experimental Botany* 65, 1689–1697.
- Torrance, L., Lukhovitskaya, N.I., Schepetilnikov, M.V., *et al.*, 2009. Unusual long-distance movement strategies of Potato mop-top virus RNAs in *Nicotiana benthamiana*. *Molecular Plant-Microbe Interactions* 22, 381–390.
- Torrance, L., Wright, K.M., Crutzen, F., *et al.*, 2011. Unusual features of pomovirus RNA movement. *Frontiers in Microbiology* 2, 1–7. (article 259).
- Zamyatnin, A.A., Solovyev, A.G., Savenkov, E.I., *et al.*, 2004. Transient co-expression of individual genes encoded by the triple gene block of Potato mop-top virus reveals requirements for TGBp1 trafficking. *Molecular Plant Microbe Interactions* 17, 921–930.

Relevant Websites

- <https://www.cabi.org/isc/datasheet/42827#toDistributionMaps>
Invasive Species Compendium.
- <https://pestdisplace.org/>
PestDisPlace.

Potato Virus Y (*Potyviridae*)

Laurent Glais, French Federation of Seed Potato Growers/Research, Development, Promotion of Seed Potato, Paris, France and Institute for Genetics, Environment and Plant Protection, Agrocampus West, French National Institute for Agriculture, Food, and Environment, University of Rennes 1, Le Rheu, France

Benoît Moury, Plant Pathology Unit, INRAE – French National Institute for Agriculture, Food and Environment, Montfavet, France

© 2021 Elsevier Ltd. All rights reserved.

This is an update of C. Kerlan, B. Moury, Potato Virus Y, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00737-8.

Nomenclature

aa Amino acid(s)

AGO Argonaute 1

CI Cylindrical inclusion

CITE Cap-independent translation enhancer

Co-Pro Protease cofactor

CP Coat protein or capsid protein

ELISA Enzyme-linked immunological assays

ER Endoplasmic reticulum

HC-Pro Helper component-protease

HR Hypersensitive reaction

IRES Internal ribosomal entry site

kb Kilobase

kDa Kilo dalton

LAMP Loop mediated amplification

MAB Monoclonal antibody

MP Movement protein

NLR Nucleotide-binding and leucine-rich

nt Nucleotide(s)

NTR Non-translated region

ORF Open reading frame

PIPO Pretty interesting Potyvirus ORF

PTNRD Potato tuber necrosis ringspot disease

RdRp RNA-dependent RNA polymerase

RISC RNA-induced silencing complex

satRNA Satellite RNA

UTR Untranslated region

VIGS Virus-induced gene silencing

VLP Virus-like particles

VPg Viral protein genome-linked

VRC Virus replication complexes

vRNA virion RNA

Glossary

Hypersensitive Reaction A specific virus strain plant defense reaction mediated by *N* genes associated with an

infection limited to tissue surrounding the initially infected cells with visible local or systemic necrosis.

Introduction

Potato virus Y (PVY) was first recognized in 1931 as an aphid-transmitted member within a group of viruses associated with potato degeneration, a disorder known since the eighteenth century. PVY is the type species of the genus *Potyvirus*, one of the six genera in the family *Potyviridae*. PVY is naturally spread by vegetatively propagated material and by aphids in non-persistent and non-circulative manners. Horizontal transmission by contact between plants and vertical transmission by true seeds have also been reported. PVY has a wide host range and is highly variable with some host specificity. Genome sequences and reliable tools for detection and strain differentiation are available. Bioassays and serology have been largely developed. PVY is one of the most damaging plant pathogens causing significant losses in four main crops around the world: potato, pepper, tomato, and tobacco. In surveys of viruses with worldwide economic importance, PVY was listed in the top-ten viruses affecting field-grown vegetables. PVY was also found responsible for damages in petunias in Europe, in eggplant crops in India and poha (or Cape gooseberry; *Physalis peruviana*) in Hawaii. Efficient control strategies depending on the crop have been developed. However, none of them seems capable to contain entirely the risks of PVY epidemics linked to virus evolution or changes of epidemiological conditions.

Viral Particle, Genome and Cytopathology

PVY particles are non-enveloped, filamentous, flexuous rods, 730–740 nm long, 11–12 nm in diameter, with an axial canal 2–3 nm in diameter and helical symmetry (Fig. 1). Each virion is composed of circa 2000 units of viral coat protein (CP) of 30 kDa, encapsidating a single-stranded, linear, positive-sense RNA molecule ($3.1\text{--}3.2 \times 10^6$ Da) of approximately 10 kb long.

The PVY genome has two distal non-translated regions (NTR) at the 5'- (from 183 to 195 nt) and the 3'- (from 326 to 407 nt) ends, with a poly-A tail at the 3'-terminus and a viral protein covalently linked (VPg) at the 5'-end. The genomic RNA contains one single open reading frame (ORF), and is translated into a large polyprotein (3061–3062 aa) cleaved in *cis* or *trans* by three virus-

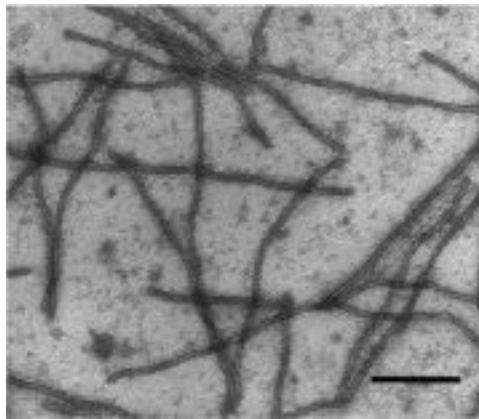


Fig. 1 Purified suspension of PVY. Magnification 22,000 $\times$. Photograph: D. Thomas, CNRS-Rennes 1 University, France. The bar represents 250 nm.

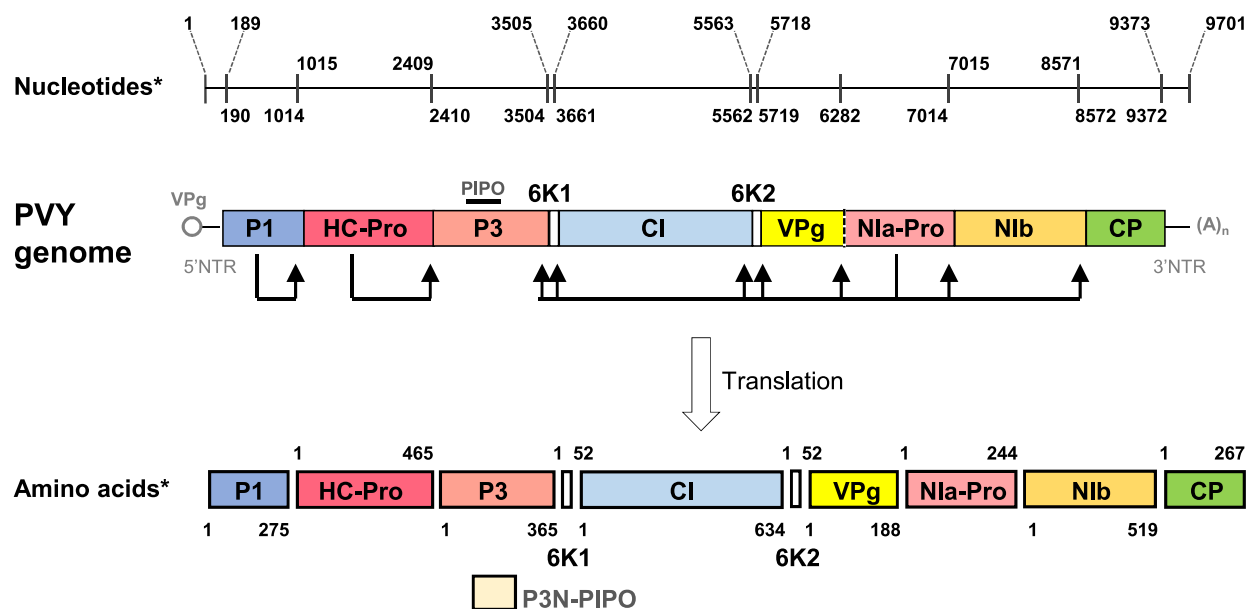


Fig. 2 Schematic representation of genomic organization of PVY and polyprotein processing. The viral genome-linked protein (VPg) is illustrated by a circle at the 5'-end of the genome. The poly-A tail is presented at the 3'-end. The large open reading frame encoded by the viral genome is presented with the names of corresponding proteins (P1, HC-Pro, P3, P3N-PIPO, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, CP). The short PIPO ORF overlapping the P3 protein is presented. The nucleotide and amino acids are numbered according to isolate N605 (accession number X97895). NTR: non-translated region.

encoded proteases (P1, HC-Pro, NIa-Pro) into ten functional proteins: P1 (275 aa), HC-Pro (465 aa), P3 (364–365 aa), 6K1 (52 aa), CI (634 aa), 6K2 (52 aa), VPg (188 aa), NIa-Pro (244 aa), NIb (519 aa), and CP (267 aa). An eleventh protein, P3N-PIPO, is produced by slippage of the viral polymerase at a specific motif embedded in the P3-coding region called PIPO (Pretty Interesting *Potyviridae* ORF) (Fig. 2).

These 11 viral proteins are mainly multi-functional. Some of them interact with some other viral proteins and/or with some host proteins to determine plant infection or virus plant-to-plant transmission. In the process of PVY infection, viral proteins are often associated with specific subcellular compartments (chloroplasts, endoplasmic reticulum and Golgi apparatus) where they perform their functions (Fig. 3). After PVY particles are introduced into the cytoplasm of plant cells by mechanical or vector-mediated inoculation, the PVY infection cycle is initiated by uncoating of virions and release of the viral RNA genome which is then recruited by the host translation machinery (eukaryotic initiation factors, ribosomes, etc.) to synthesize the polyprotein. After its processing, the 11 viral proteins are released. Nuclear inclusion proteins (NIa, NIb) migrate to the nucleus and cylindrical inclusion (CI) proteins accumulate in the cytoplasm by forming pinwheel-shaped inclusions. Viral replication complexes associate to the membrane of host organelles (endoplasmic reticulum, Golgi apparatus and vesicles) through the 6K2 protein. During replication, the positive-sense single-stranded RNA (ssRNA) is copied into a complementary negative-sense RNA and vice versa by

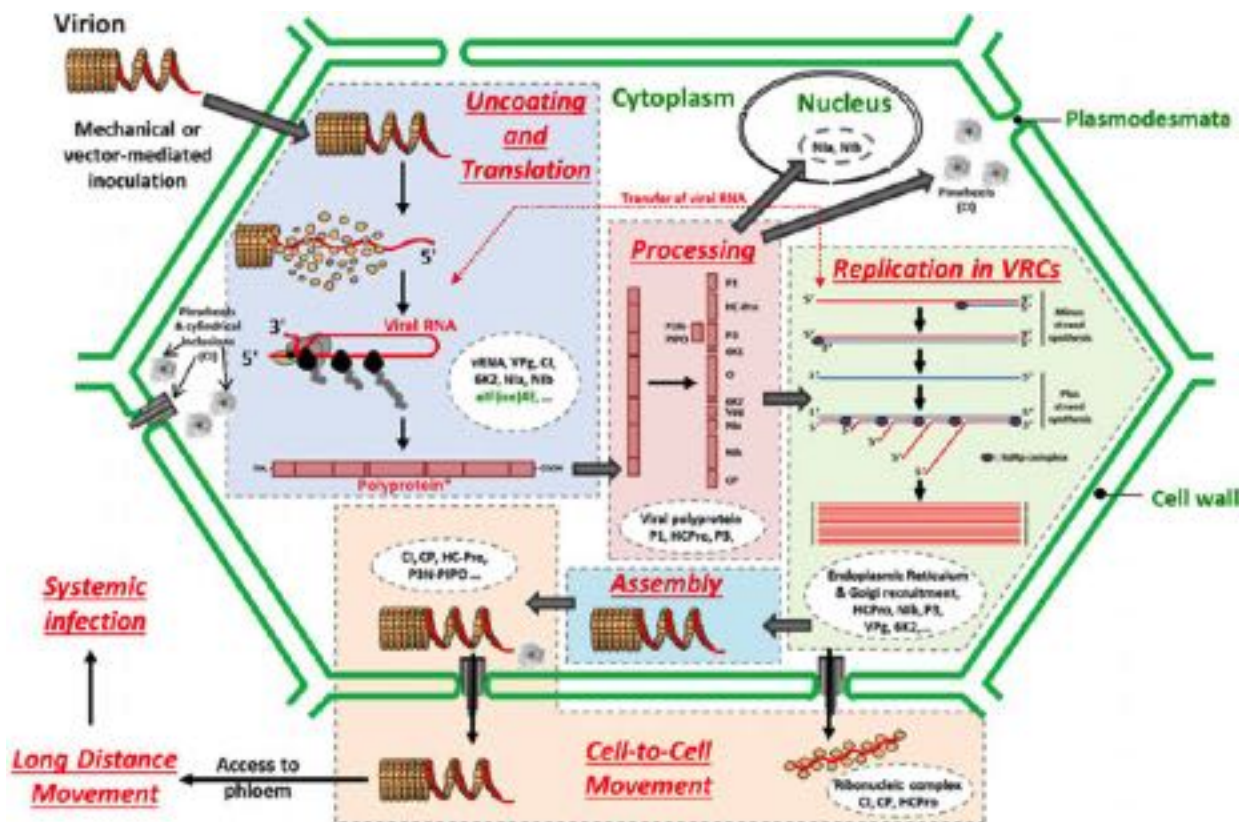


Fig. 3 Schematic representation of PVY infection cycle in infected plant cell, with the location of viral proteins and their mutual interactions. Extracted from Lacomme, C., Jacquot, E., 2017. General characteristics of potato virus Y (PVY) and its impact on potato production: An overview. In: Lacomme, C., Glais, L., Bellstedt, D.U., *et al.* (Eds.), *Potato Virus Y: Biodiversity, Pathogenicity, Epidemiology and Management*. Switzerland: Springer, pp. 1–19.

the RNA-dependent RNA polymerase (RdRp) (NIB protein) and the RNA helicase (CI protein). The obtained positive-sense ssRNA are either encapsidated to produce new virions or recruited by viral proteins (CI, CP, HC-Pro, P3N-PIPO) and host proteins to form ribonucleic complexes that move from cell to cell through plasmodesmata and systemically in phloem vessels.

Host Range

The natural host range of PVY comprises plants in nine families. It includes potato (*Solanum tuberosum* ssp. *tuberosum*), several species of native potatoes in the Andes including *Solanum andigena* and numerous *Solanum* species, pepper (*Capsicum* spp.), tobacco (*Nicotiana* spp.), tomato (*Solanum lycopersicum*), eggplant (*Solanum melongena*), Cape gooseberry (*Physalis peruviana*), ornamentals (*Petunia* spp., *Dahlia* spp.), perennial plants (*Physalis virginiana*, *Physalis heterophylla*), and a number of annual plants. Many region-specific hosts were recorded such as *Cotula australis* in New Zealand or *Sorbaria tomentosa* in Himalaya.

Weeds in the family Solanaceae are often potential inoculum sources for PVY infections in tomato and pepper crops: *Solanum nigrum* and *Solanum dulcamara* in many countries; *Solanum chacoense* and *Physalis viscosa* in Southern America; *Solanum gracile*, *Solanum aculeatissimum*, *Physalis angulata*, *Physalis ciliata*, and *Physalis floridana* in Florida. *Physalis virginiana* and *Physalis heterophylla* (perennial ground cherries) are overwintering hosts in Northern America. *Solanum dulcamara* in Western Europe or *Datura* spp. in Mediterranean countries are potential inoculum sources in potato crops as well as volunteer potatoes. Infected pepper and tomato plants, and seed and volunteer potatoes were also identified as potential sources of inoculum in tobacco crops in the USA, Canada, and Italy.

PVY is easily sap-transmitted and can be transmitted by stem- and tuber-grafting. Its experimental host range comprises plants in 495 species and in 72 genera of 31 families. It includes 287 species in the family Solanaceae (among which 141 *Solanum* species and 70 *Nicotiana* species), 28 species of Amaranthaceae, 25 species of Fabaceae, 20 species of Chenopodiaceae, and 11 species of Asteraceae. A large part of these plants show only local lesions upon PVY inoculation. *Datura stramonium*, formerly reported as a host plant, was demonstrated to be totally immune to all strains tested in 1980–1990. Some PVY isolates can infect the model plant *Arabidopsis thaliana*.

Nicotiana tabacum, *Nicotiana benthamiana*, *Nicotiana occidentalis* can be used as diagnostic species susceptible to all PVY strains. *N. tabacum* plantlets are often used as source and test plants for aphid transmission experiments. *N. tabacum* is a suitable host for



Fig. 4 Winged form of *Myzus persicae*, an insect vector of PVY. Photograph: B. Chaubet, INRAE, France.

virus purification. Yields of purification including a final step of cesium chloride gradient centrifugation usually vary from 10 to 25 mg kg⁻¹ of tobacco leaves. Leaves of *N. tabacum* are the best virus-infected material to store. Antigenic properties can be retained for 1 year in freeze-dried crude extracts. Infectivity was reported to be preserved in freeze-dried material stored over calcium chloride at 4°C for 15 years. However from our own experience numerous long-term stored isolates of PVY available in international collections are no longer infectious.

Solanum demissum Y, *S. demissum* PI 230579, and the hybrid *S. demissum* A6 are local lesion hosts. The 'A6 test' on detached leaves was commonly used in the past as a diagnostic tool for differentiating PVY from Potato virus A (PVA), another potyvirus. Cultivars of *N. tabacum*, potato and pepper, *Chenopodium amaranticolor*, *P. floridana*, some accessions of *Solanum brachycarpum* and *Solanum sparsipilum* are useful for distinguishing among PVY strains and pathotypes. Potato cultivars such as Bintje or Saco can be used to separate PVY from plants co-infected with PVA or Potato virus X (PVX). The capacity of PVY isolates from the C1 group to infect pepper (*C. annuum*) was shown to require mutations in both the CI and the P3 and/or P3N-PIPO coding regions, and seem to have evolved several times independently.

Transmission

In natural conditions, different transmission routes are used by PVY. The first one is vertical transmission from infected mother plants to their progeny. Potato plants are propagated vegetatively by planting tubers from one generation to the next. When PVY is inoculated to a potato plant, after a multiplication phase in the first infected cells, the virus migrates from cell to cell, cross different cell layers of the inoculated organ and reach the phloem. Once in the phloem, PVY is rapidly translocated to the whole plant including to daughter tubers, where it multiplies. If these infected tubers ('seed' potatoes) are planted the year after, the germinated seedlings will also be infected and usually initiate a rapid PVY dissemination in fields through secondary aphid transmission. Transmission of PVY from mother plants to its 'true' seeds has been reported from *Solanum nigrum*, *Nicandra physaloides* and pepper (*Capsicum annuum*).

The second transmission route corresponds to horizontal transmission from plant to plant by aphids (Fig. 4). PVY displays a large range of aphid vectors. Aphids in 65 species, all in the family Aphidinae, colonizing or not the source plants, were demonstrated to be able to transmit PVY. *Myzus persicae* is the most efficient PVY vector and most of the other aphid species have much lower transmission efficiencies comparatively. Both alate and wingless aphids are vectors of PVY. PVY is transmitted by aphids in a non-persistent manner, meaning that acquisition and inoculation periods are brief (a few seconds or minutes) and that the virus is not retained on the long term by the aphids. Aphid stylets penetrate into the epidermal cell layer of the plants and puncture plant cell membranes, allowing insects to suck the cellular content. In the case of infected cell, this probe or feeding is accompanied by the sucking of virus particles that bind to specific receptors located in the extreme tip of the aphid's stylet, called the acrostyle. Only particles attached to these receptors are involved in aphid transmission. PVY transmission by aphids is also noncirculative. There is no discernible latent period and the virus does not need to circulate into the aphid body to allow transmission. PVY does not pass through insect moults and retention of the virus in the aphid usually lasts not more than 1 or 2 h. However, longer retention periods (up to 17 h in *Aphis nasturtii*) were reported. Prior starvation of the aphids increases the efficiency of transmission though it does not affect the number of electrically-recorded membrane punctures during acquisition periods.

PVY transmissibility is determined by both HC-Pro and CP proteins. The non-structural HC-Pro protein corresponds to the helper component of the aphid transmission, which interacts in its homodimer active form as a bridge between the virion and the aphid stylet. It has been recently proved with other potyviruses that this HC-Pro dimerization is a key step for an efficient transmission. Studies performed on LMV (Lettuce mosaic virus) and TuMV (Turnip mosaic virus) demonstrated that viruses activate their own transmission by dimerization of HC-Pro protein from the moment they perceive an unknown plant signal announcing the penetration of infected cells by aphid vector. Without the presence of this active form of HC-Pro protein, the transmission is impossible. This phenomenon is called the transmission activation (AT). Otherwise, two domains in HC-Pro protein are crucial for PVY transmission: the KITC motif (Lysine/Isoleucine/Threonine/Cysteine) located in its N-terminal region,

which is involved in the binding to a specific receptor in aphid's mouthparts, and the PTK motif (Proline/Threonine/Lysine) in its C-terminal region, which interacts with the DAG motif (Aspartic acid/Alanine/Glycine) at the CP N-terminus. With these two motifs, HC-pro provides the link between the aphid's stylet and the virion.

PVY transmission by pollen has never been proved for any host plant. In experimental conditions, other methods are also used to transmit PVY from plant to plant. The most commonly used is mechanical inoculation. A healthy plant becomes infected by manually rubbing leaves or cotyledons with a mixture containing PVY-infected plant sap and a fine abrasive powder. Transmission by grafting a scion from a virus-infected plant to a healthy stock plant (or vice versa) is also efficient. Transmission of PVY by woundings caused by squashing, squeezing or bouncing of healthy plants in the vicinity of infected plants can happen but is not frequent in field conditions. Transmission of PVY from an infected to a healthy tuber by seed cutting was not reported.

Serology

PVY is considered to be strongly immunogenic. The PVY antigenic structures are carried by the CP. Its C-terminal region contains epitopes specific to PVY, whereas its highly variable N-terminal region carries epitopes allowing the differentiation of PVY in two main serogroups: serotype O (PVY^O, PVY^C, PVY^{N-Wi} strain groups), and serotype N (PVY^N, PVY^{N^{NTN}}, PVY^Z, PVY^E). Antisera and monoclonal antibodies (MAbs) have been produced in rabbits or mice immunized with purified virus preparations or synthetic peptides, or by using DNA-based immunization, or phage display antibody technology. Numerous sources of antibodies and serological detection kits are available. Polyclonal antibodies do not discriminate among PVY strain groups, whereas monoclonal antibodies, triggering a single epitope, allow the serotyping of PVY isolates. These serological reagents have been mainly used in ELISA (Enzyme-Linked Immunosorbent Assay) and related protocols (dot-blot immunobinding assay, immunosorbent electron microscopy, latex and viro-bacterial agglutination tests, immunodiffusion in agar gels) for the detection and the identification of PVY from plant material.

Phylogeny and Evolution

Outside recombinant isolates, three major PVY phylogenetic groups with worldwide distributions, namely C, N, and O, can be distinguished based on genome sequences (Fig. 5). Group C can be further divided into two subgroups, C1 and C2. These phylogenetic groups partly correlate with PVY host specificity. In potato, most isolates belonged to groups C2, N, and O, whereas in pepper (*Capsicum* spp.), most isolates belonged to group C1. All kinds of isolates could be found in tomato or tobacco. A large number of isolates are recombinants between these main phylogenetic groups. Thirty-six different recombination patterns have been identified, mainly between the N and O groups. Such recombinants are now becoming predominant in potato crops worldwide. N × O recombinants are also becoming widespread in pepper crops, whereas isolates from their parental groups are poorly infectious in this crop plant.

Two additional PVY phylogenetic groups were recently identified and have more restricted distributions: a "Chilean" group on tobacco and pepper (*Capsicum baccatum*) and a Brazilian U group on tobacco (Fig. 5). Phylogenetic reconstruction of host specificity in PVY suggests that infection in pepper (*Capsicum* spp.) is a recently-acquired trait and that it was acquired at least twice independently. In contrast, it is uncertain if infection in potato is an ancestral or derived trait in PVY.

PVY belongs to a cluster of 19 potyvirus species, the closest to PVY being Pepper severe mosaic virus. Most of these potyviruses infect plants in the family Solanaceae but others infect plants of the Amaranthaceae, Asteraceae, Liliaceae, Amaryllidaceae, or Verbenaceae.

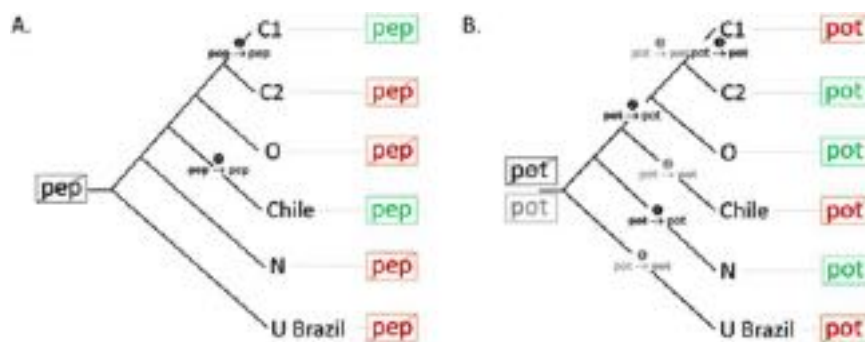


Fig. 5 Phylogenetic relationships between PVY groups and inference of adaptation to pepper (pep) (A) and potato (pot) (B). (A) The most parsimonious scenario infers that the most recent common ancestor (MRCA) of PVY was not infectious in pepper and get adapted to pepper at least two times independently. (B) There are two equally parsimonious scenarios for PVY adaptation to potato, considering that the MRCA was either infectious (in gray) or not infectious (in black) in potato. Both scenarios involve at least three changes in PVY adaptation to potato. In green: extant PVY clades that are infectious in the respective crop plant. In red: extant PVY clades that are not infectious in the respective crop plant.

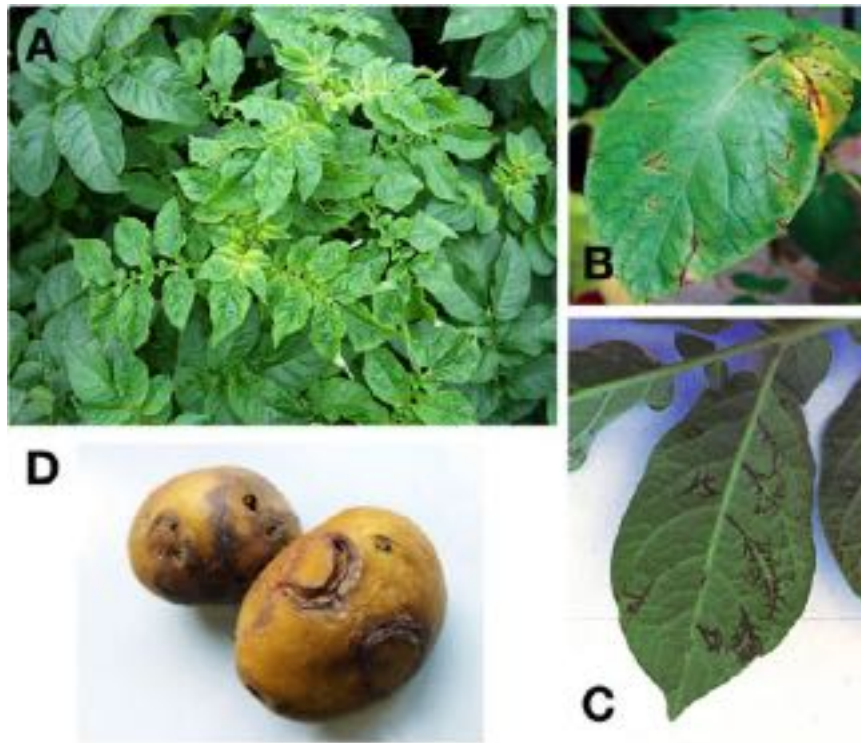


Fig. 6 (A) Natural secondary infection by PVY in a potato field: crinkling, yellowing, and growth reduction of the leaflets. Symptoms associated with the potato tuber necrotic ringspot disease (PTNRD). (B) Yellowing and necrosis on a basal leaf. (C) Necrotic oak-leaf pattern on an intermediate leaf. (D) Typical symptoms on tubers of the cv. Monalisa. Photographs: Y. Le Hingrat, FN3PT, France (A), K. Charlet-Ramage, GNIS, France (B) and (C), L. Glais, FN3PT, France (D).

PVY in Potato

For almost 50 years, PVY has become the most important virus in most growing areas for seed, ware, and processed potatoes. Serious outbreaks were reported in the 2000s. Tuber quality can be severely affected due to necrosis or defects for processing potatoes. Reduction in size and number of harvested tubers can result in losses up to 80% if no control of PVY spreading was used. Symptoms consist of mild to severe mottle, often associated with crinkling of the leaves (**Fig. 6(A)**). Yellowing and necrosis (vein necrosis and necrotic spots) frequently occur in the lower leaves (**Fig. 6(B)**). Symptoms also include collapse and dropping of intermediate leaves, which remain clinging to the stem (leaf drop). Secondarily-infected plants (when infected tubers are planted) are stunted with crinkled and smaller leaflets.

Necrosis on and around leaf veins, on petioles, stems, and tubers may occur in numerous cultivars. Some necroses on stems and petioles are called stipple-streak disease. The potato tuber necrotic ringspot disease (PTNRD) is characterized by particular patterns on leaves (sometimes oak-leaf necrosis) (**Fig. 6(C)**) and stems, and superficial annular and arched necroses on tubers, first slightly protruding from tuber skin, then becoming dark brown with sometimes crackings of tuber skin (**Fig. 6(D)**).

Potato PVY isolates are classified according to three criteria: their capacity to induce necrotic symptoms on *N. tabacum*, to overcome different resistance genes in *Solanum tuberosum* and to produce necrosis on potato tubers. Based on these biological features, PVY isolates are divided into seven strain groups: PVY^O, PVY^C, PVY^N, PVY^{NTN}, PVY^{N-Wi}, PVY^Z, PVY^E. Besides these strain groups, several individual isolates sharing particular biological properties are till now unclassified. PVY^O and PVY^C induce mosaic symptoms on tobacco leaves. They are mainly separated on the basis of hypersensitive reactions (HR) in potato cultivars. PVY^O induces a HR reaction in cultivars bearing the resistance genes *Ny_{1br}*, *Ny-1* and *Ny-2* (**Table 1**), whereas PVY^C elicits HR response in cultivars carrying the *Nc* gene. PVY^N differs from PVY^O and PVY^C in causing a veinal necrosis in *N. tabacum* cv. Samsun or cv. Xanthi and eliciting a HR in cultivars carrying the *Ny-1* and *Ny-2* resistance genes (**Table 1**). Two aa at positions 400 and 419 in the HC-Pro protein are involved in the induction of the necrotic reaction in tobacco.

Isolates of the PVY^{NTN} group express the same biological feature as PVY^N, except that they are able to induce tuber necrosis on susceptible cultivars. One of the viral molecular determinants inducing PTNRD was identified at aa position 419 in the HC-Pro protein, but some others are still unknown. Whole genome analyses of PVY^{NTN} isolates revealed that most of them result from three to four recombination events between PVY^O and PVY^N isolates, with breakpoints being not exactly at the same nucleotide position. Some PVY^{NTN} isolates however are not recombinants (**Fig. 7**).

Table 1 Molecular determinants of pathogenicity and transmission identified in PVY

Phenotype	Plant	Genome region	Mutation
<i>Pathogenicity</i>			
Veinal necrosis	<i>Nicotiana tabacum</i>	Hc-Pro	N339D, D400K, D419E
Veinal necrosis	<i>Nicotiana tabacum</i> with the <i>Rk</i> nematode resistance gene	Nib	–
Tuber necrosis	<i>Solanum tuberosum</i>	HC-Pro	D419E
Leaf necrosis	<i>Physalis floridana</i>	CP	–
<i>Host range</i>			
Systemic infection	<i>Capsicum annuum</i>	P3 and/or P3N-PIPO	N131Y, N131F
Systemic infection	<i>Capsicum annuum</i>	CI	–
<i>Interaction with plant resistance</i>			
Resistance induction	<i>Solanum tuberosum</i> with the <i>Nc</i> resistance gene	Hc-Pro	K269R and/or R270K (local necrosis induction)
Resistance induction	<i>Solanum tuberosum</i> with the <i>Ny</i> resistance gene	Hc-Pro	K269R and/or R270K (necrosis induction)
Resistance breakdown	<i>Capsicum annuum</i> with the <i>pvr2</i> resistance gene	VPg	Different mutations in codons 101–123
Resistance breakdown	<i>Capsicum annuum</i> with the <i>pvr2³</i> resistance allele	VPg and CI	–
Resistance breakdown	<i>Capsicum annuum</i> with the <i>Pvr4</i> resistance gene	Nib	K472E
Resistance breakdown	<i>Solanum lycopersicum</i> with the <i>pot – 1</i> resistance gene	VPg	R119H
Resistance breakdown	<i>Nicotiana tabacum</i> with the <i>va</i> resistance gene	VPg	105E, 105T, 105M, 119A
Loss of resistance induction	<i>Solanum tuberosum</i> with the <i>Ry</i> resistance gene	Nla-Pro	H46Y, D81A, C151F or H167P
<i>Aphid transmission</i>			
Loss of transmissibility		Hc-Pro	K59E
Increase of transmissibility		CP	E68K

Note: “–”: unavailable data.

Isolates of the PVY^{N-Wi} group, similar to PVY^{N-O} in North America, are differentiated from PVY^N isolates by their O/C (instead of N) serotype, and by their capacity to induce PTNRD on specific cultivars. The fact that these isolates share both PVY^N and PVY^O properties is due to the presence of two to four recombination breakpoints in their genome leading to a succession of O- and N-type nucleotide sequences (Fig. 7). PVY^Z and PVY^E isolates are closely related to PVY^{NTN} isolates because they cause PTNRD in tubers and display a similar recombinant genome, but they differ by their capacity to induce mosaic symptoms in tobacco leaves, as PVY^O isolates. Moreover, PVY^Z isolates elicit a HR reaction in potato cultivars carrying the *Nz* resistance gene, which is not the case of PVY^E isolates.

The first PVY groups identified in the early 1930s were PVY^C and PVY^O, and PVY^N a few years later. In contrast to PVY^C which became less frequent in potato, PVY^O remained the dominant group in potato fields up to the 2000s. However, from the 1960s, infections caused by PVY^N isolates increased progressively in Europe and became dominant 40 years later. This change was associated to the emergence of PVY^{NTN} isolates, which were first described in Hungary in 1982 and soon spread in most potato growing countries, except in North America where they were classified as quarantine pathogens. So far, PVY^{NTN} isolates have a worldwide distribution and are predominant notably in Europe. In Switzerland and Estonia, a significant decrease in the occurrence of PVY^{NTN} isolates was recently observed concurrently with an increase of PVY^{N-Wi} infections, probably due to a change in potato varieties. PVY^{N-Wi} isolates were identified for the first time in Poland in 1984 and they are currently present throughout the world with varying frequencies. Some surveys performed these last two decades reported that PVY^{NTN} and PVY^{N-Wi} groups have almost replaced the PVY^O and PVY^N ancestral groups in many countries. PVY^Z and PVY^E groups are not frequent.

Differences in aphid transmission between strains were reported. PVY^N isolates seem better transmitted than other PVY isolates with longer retention periods in aphids. Whatever the strain, *Myzus persicae* is clearly the most important vector in most potato-growing areas. However, despite their low transmission efficiency, other species colonizing potatoes such as *Aphis nasturtii*, or visiting potatoes can also contribute to PVY spread. Some of these species, notably cereal aphids or the pea aphid, seem to be involved in PVY epidemics in potato crops in Europe and the USA due to their high abundance.

Control Methods

Control methods are first based on sanitary control of seed potatoes and breeding for resistance, especially regarding PVY^N. Sophisticated schemes of seed production include monitoring of aphid vectors, treatments by mineral oils against aphid transmission, eradication of weeds, and post-harvest detection tests using large-scale ELISA. Numerous molecular assays were also developed for PVY detection in leaves, potato tubers, and aphids, including numerous RT-PCR protocols, microarray technology, and real-time RT-PCR methods. Resistance based on hypersensitive reaction (HR) and extreme resistance (ER) against PVY are

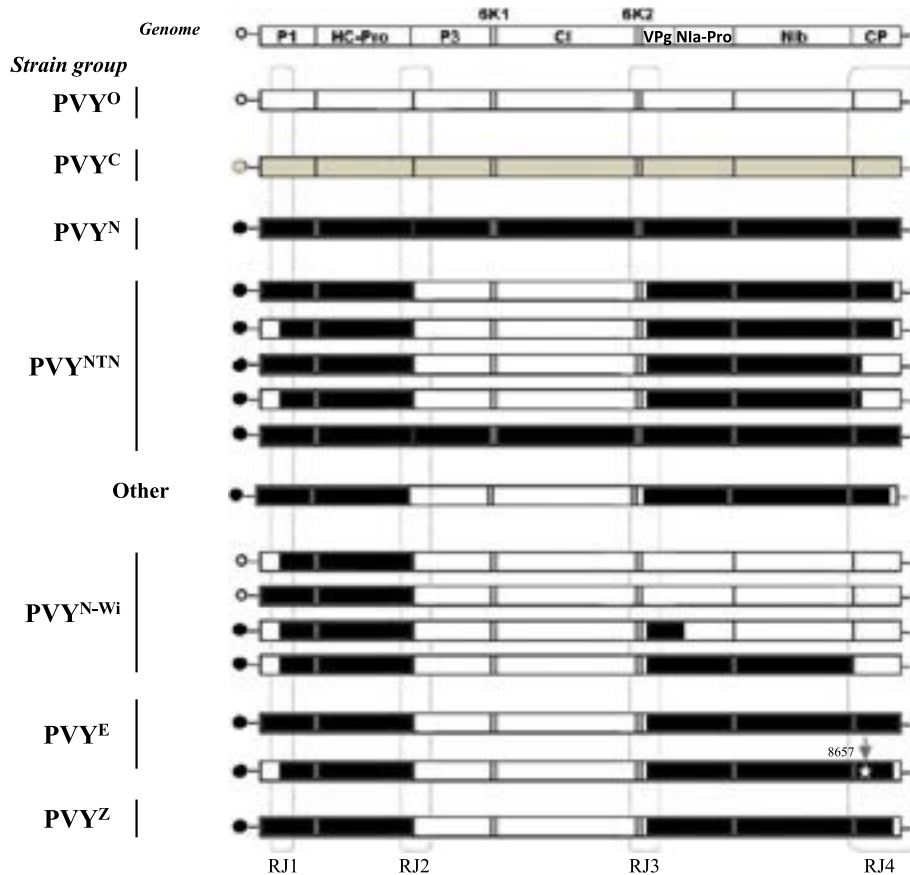


Fig. 7 Schematic representation of the main molecular diversity within potato PVY strain groups. Boxes in white, brown and black correspond to PVY^O, PVY^C and PVY^N nucleotide sequences, respectively. RJ1, RJ2, RJ3 and RJ4 correspond to the position of the hot spots of recombination described in PVY genome. “Other” refers to some PVY isolates displaying particular biological properties. The arrow indicates the presence of single nucleotide mutation leading to a serological misidentification. Adapted from Glais, L., Bellstedt, D.U., Lacomme, C., 2017. Diversity, characterization and classification. In: Lacomme, C., Glais, L., Bellstedt, D.U., *et al.* (Eds.), Potato Virus Y: Biodiversity, Pathogenicity, Epidemiology and Management. Switzerland: Springer, pp. 43–76.

largely used. HR is based on the single dominant genes *Nc*, *Nz* and *Ny* and ER is based on the *Ry* genes. The *Ry* genes (*Ry_{adg}*, *Ry_{chr}*, *Ry(o)phu*, *Ry_{sto}*), from wild *Solanum* species, map on chromosomes 11, 9, 9, and 12, respectively, and confer broad-spectrum resistance, whereas HR genes *Nc*, *Ny_{thr}* (which map on chromosome 4) and *Nz* found in old potato cultivars, protect against PVY^C, PVY^O, and PVY^Z, respectively. More recently, *Ny-1* and *Ny-2* genes eliciting hypersensitive response against both PVY^C and PVY^N isolates were mapped on potato chromosomes 9 and 11, respectively. Beside these well documented resistance genes, new putative resistance genes were reported according to local symptoms observed on leaves inoculated with a set of differential PVY strains. *Nw*, *Ne*, *Nn*, *Nna* seem to confer resistance against PVY^{N-Wi} (or PVY^{N-O}), PVY^E, PVY^N, and PVY^N from North America, respectively (Table 2). Sources of resistance against aphids were identified in wild *Solanum* species but not introgressed in potato cultivars yet.

PVY in Pepper

PVY is the causal agent of major diseases and production losses in pepper crops. PVY affects pepper production worldwide, while other pepper-infecting potyviruses, like Tobacco etch virus, Pepper severe mosaic virus, Pepper yellow mosaic virus, Pepper vein mottle virus or Chilli vein mottle virus are mostly restricted to particular continents and are not present in Europe. Symptoms are mostly visible on the vegetative parts of the plants but rarely on fruits. Depending on the pepper genotype and on the virus isolate, they consist of mosaic and vein banding on leaves or necrotic symptoms on leaves, petioles, and stems (**Fig. 8(A)** and **(B)**).

The most widespread and efficient way to control PVY in pepper is through the growing of cultivars carrying resistance genes. The first resistance genes used to control PVY were the *pvr2*¹ and *pvr2*² alleles, which control high-level resistance since no virus can be detected in inoculated organs of the plants. These genes have been exploited for decades in a large number of pepper cultivars and proved to be durable. However, PVY isolates virulent to the *pvr2*¹ allele were occasionally observed, especially in the Mediterranean basin and in tropical regions, while isolates virulent to the *pvr2*² allele are exceptional.

Table 2 List of identified PVY resistance genes in various hosts

Carrier chromosome	Gene or allele name	Nature of encoded protein	Source of resistance
<i>Potato</i>			
Chr. 11	<i>Ry_{adg}</i>	–	<i>S. andigena</i>
Chr. 12	<i>Ry_{sto}</i>	NLR	<i>S. stoloniferum</i>
Chr. 9	<i>Ry_{chc}</i>	–	<i>S. chacoense</i>
Chr. 9	<i>Ry(o)phu</i>	NLR	<i>S. phureja</i>
Chr. 4	<i>Ny_{thr}</i>	–	<i>S. tuberosum</i>
Chr. 4	<i>Nc</i>	–	<i>S. tuberosum</i>
Chr. 9	<i>Ny – 1</i>	–	<i>S. tuberosum</i>
Chr. 11	<i>Ny – 2</i>	–	<i>S. tuberosum</i>
–	<i>Nz</i>	–	<i>S. tuberosum</i>
–	<i>Nw</i>	–	<i>S. tuberosum</i>
–	<i>Ne</i>	–	<i>S. tuberosum</i>
–	<i>Nn</i>	–	<i>S. tuberosum</i>
<i>Tobacco</i>			
Chr. 21	<i>va</i>	eIF4E	<i>N. tabacum</i> 'Virgin A Mutant'
<i>Pepper</i>			
P4	<i>pvr2</i> (>30 resistance alleles)	eIF4E	>30 accessions in <i>C. annuum</i> , <i>C. frutescens</i> , <i>C. chinense</i> , <i>C. baccatum</i> ,
P10	<i>Pvr4</i>	NLR	<i>C. annuum</i> 'CM334'
P10	<i>Pvr7</i>	–	<i>C. chinense</i> 'PI159236'
<i>Tomato</i>			
Chr. 3	<i>pot – 1</i>	eIF4E	<i>Solanum habrochaites</i> 'PI247087'
Chr. 3	<i>pot – 1²</i>	eIF4E	<i>Solanum pimpinellifolium</i> 'LA0411'

Note: eIF4E: eukaryotic initiation factor 4E; NLR: nucleotide-binding leucine-rich repeats.

“–” corresponds to unknown data.

These two alleles and the wild-type susceptibility allele *pvr2*⁺ encode different copies of the eukaryotic translation initiation factor 4E1 (eIF4E1) that differ by a few amino acid substitutions. Reciprocally, amino acid substitutions in the central part of the VPg of PVY determine the virulence towards *pvr2* alleles. In addition to *pvr2*¹ and *pvr2*², more than 30 *pvr2* resistance alleles are found in locally-grown pepper accessions of almost all *Capsicum* species analyzed. Most of these alleles have a specific resistance spectrum that match PVY VPg diversity. The large amino acid diversity in both pepper eIF4E1 and PVY VPg and evidence of positive selection during the evolution of these two proteins suggest a co-evolution between pepper and PVY. Pepper infection and resistance based on eIF4E1/VPg interaction obey a lock-key model: direct binding between eIF4E1 and VPg proteins allows PVY to replicate and translate its genome and to infect the plant. In contrast, amino acid substitutions in either the eIF4E1 or VPg that abolish binding are responsible for plant resistance. There are great differences in durability for the resistance conferred by different *pvr2* alleles.

From the 1990s, the dominant *Pvr4* gene, originating from a Mexican *C. annuum* population, was introgressed into bell-pepper cultivars. This gene confers a high-level resistance to all tested isolates of PVY in addition to five other potyviruses belonging to the PVY phylogenetic group: EcRV (Ecuadorian rocoto virus), PepMoV (Pepper mottle virus), PepSMV (Pepper severe mosaic virus), PepYMV (Pepper yellow mosaic virus), and PTV (Peru tomato mosaic virus). This resistance shares common properties with hypersensitivity-based resistances. Only a few laboratory-selected PVY isolates are able to break down the *Pvr4* resistance, indicating a high durability of the resistance. The *Pvr4* gene was shown to encode a nucleotide-binding and leucine-rich (NLR) domain protein. It shares common structure and sequence with the *Tsw* gene that confers resistance to Tomato spotted wilt virus, suggesting a common origin. The ability of laboratory isolates to break the *Pvr4* resistance was shown to be due to a single substitution in the NlB coding region. This mutation induces a high competitiveness cost to PVY in plants devoid of the *Pvr4* gene, which may explain the high durability of this resistance.

Other pepper resistances to PVY have been described but not used in commercial cultivars. A polygenic resistance from the *C. annuum* accession Perennial was analyzed and several QTLs (quantitative trait loci) have been mapped in the plant genome. These QTLs can be used in combination to a *pvr2* resistance allele to improve its durability. The *Pvr7* resistance gene from the *Capsicum chinense* accession PI159236 confers a high level of resistance to PVY and PepMoV and is allelic to *Pvr4*.

Other control measures such as the use of oil sprays or physical barriers (polyethylene sheets and coarse nets) were also used for controlling spread of PVY in pepper crops in Israel and Florida.

PVY in Tomato

PVY can induce severe diseases in tomato crops but has long been considered a pathogen of secondary importance, inducing mild mosaic on the foliage only. Since the 1980s, new strains of PVY have arisen in Mediterranean countries, causing serious

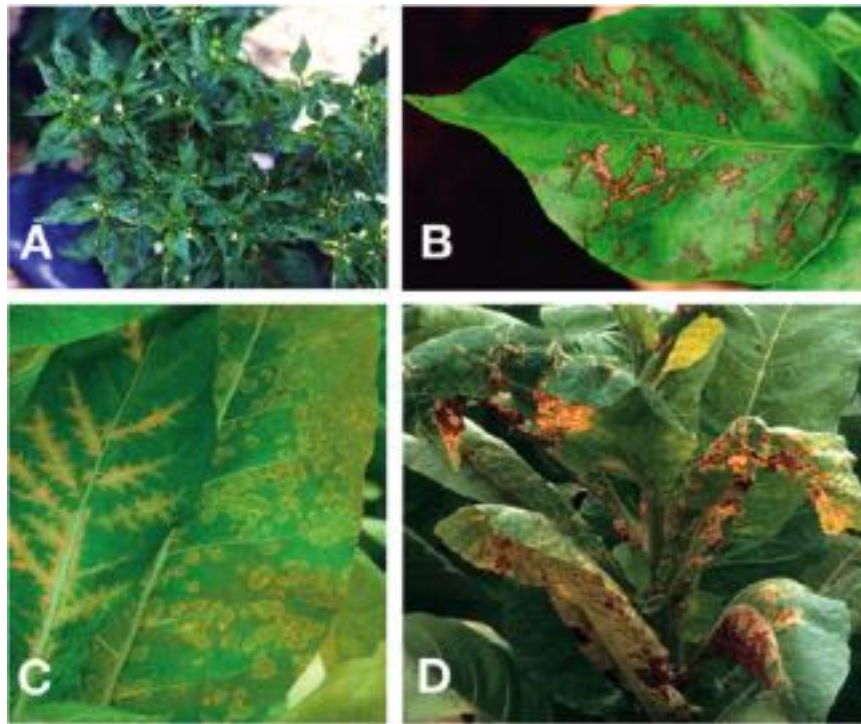


Fig. 8 PVY infection in a pepper field: (A) severe crinkling and distortion of the leaves. (B) Necrotic patterns on a leaf. Most pepper isolates of PVY belong to the C1 phylogenetic group, but more and more N/O recombinants are observed. An O/C recombinant strain inducing veinal necrosis was also isolated in pepper in Italy. PVY infection on tobacco cv. Burley in field trials in Southwestern France. (C) Chlorotic oak-leaf and halo patterns. (D) Severe systemic necrosis. Photographs: P. Gognalons, (A) and M. Pepelnjak Slovenia (B), Blancard (C) and (D), INRAE, France.

yield and quality loss of tomato fruits. These strains induce necrotic lesions on leaves and necrotic streaks on stems in all tomato varieties and frequently affected 100% of the plants in greenhouses as well as in open fields. Importantly, combination of PVY infections with other viruses such as Cucumber mosaic virus can induce very severe diseases in tomato crops. Data about the genetic diversity of PVY isolates from tomato are scarce. Isolates belonging to the C1, N, O, and N/O groups have been observed in tomato crops. In laboratory tests, most of the isolates belonging to all PVY phylogenetic groups induce systemic infections in tomato.

There are few reports about PVY resistance in tomato and related species. Accession PI247087 of *Solanum habrochaites* was described highly resistant to PVY. The resistance is efficient toward all tested PVY isolates, inhibits the multiplication of the virus in the inoculated organs and is controlled by a single recessive gene (named *pot-1*) which is homologous to the *pvr2* gene in pepper and encodes the eIF4E1. Strains of PVY virulent to the *pot-1* resistance can be selected in laboratory tests, a property that is controlled by a single aa substitution in the VPg. The *pot-1* gene was introduced into genotypes of *S. lycopersicum*, but varieties carrying this gene are not widespread, possibly because PI247087 is an accession of the wild species *S. habrochaites* and introgression of the resistance into cultivated tomato is difficult or is deleterious for yield and/or quality.

Accession LA0411 of *Solanum pimpinellifolium* carries also a *pot-1* mutant allele (*pot-1*²) that confers resistance to PVY, but the spectrum of the resistance is narrow (only a few PVY isolates) and its durability potential is low.

Engineered resistance was also developed in tomato (*S. lycopersicum*) using a TILLING (targeting induced local lesions in genomes) approach, which combines ethyl methane sulfonate mutagenesis and gene-specific detection of mutations. An eIF4E1 KO (knock out) mutant showed a narrow spectrum of resistance against PVY. This spectrum was enlarged to all tested PVY isolates (and to Tobacco etch virus isolates) when the eIF4E1 KO mutation was combined with a KO mutation in another gene of the eIF4E family, eIF4E2. However, this double mutant showed a stunted growth.

PVY in Tobacco

PVY was reported to be the most damaging virus in tobacco crops. It causes height reductions and modifies the chemical composition of cured leaves, especially the nicotine content. Yield losses of 14%–59% were reported with incidence of up to 100%. Symptoms on leaves are usually mild mottling but particular chlorotic patterns and necroses, notably the veinal necrosis disease, may also occur (Fig. 8(C) and (D)). PVY isolates are split into two groups according to their capacity to induce (PVY^N) or

not (PVY^O and PVY^C) veinal necrosis on tobacco leaves. As said previously, the HC-Pro was shown to carry molecular determinants involved in this phenotype. Recently, the *RPP8-like* gene in tobacco, called *NtTPN1* (*Nicotiana tabacum* Tolerance to PVY-induced Necrosis1), was shown to determine the expression of the necrotic symptoms, presumably through direct or indirect interaction with viral HC-Pro protein, because the Glycine to Arginine mutation at codon position 497 in *NtTPN1* is sufficient to suppress PVYN-induced veinal necrosis. Genome analyses of PVY isolates collected in tobacco in different countries reported that O, N, and C1 isolates as well as N/O recombinants (PVY^{NTN}, PVY^{N-wi}) are found on tobacco.

Breeding for resistance is the main control method against PVY in tobacco production. Many European and American commercial cultivars carry recessive resistances of varying efficiency and durability. The *va* gene of the Virgin A Mutant (VAM) line, corresponding to a large deletion around the *elF4E-1* gene on chromosome 21 after X-ray irradiation, has been extensively used, though its introgression is often associated with a decrease of yield and cured leaf quality. Nevertheless, some *va*-breaking necrotic PVY isolates were observed in open fields. Recently, it was reported that the durability of this resistance is enhanced by a complex genetic locus on chromosome 14 containing three other *elF4E* genes. The viral VPg protein is involved in the breakdown of the resistance conferred by *va* (Table 1).

In countries where early infection occurs frequently, additional measures consist in protecting seedbeds by fleece, eradication of weeds and isolation from potato, tomato and pepper fields. Pathogen-derived resistance has been extensively studied in tobacco, although more for investigating the mechanisms of protection than for use in commercial tobacco cultivars.

Prospects

Since the first identification of PVY, a significant amount of information was collected by researchers. However, some gray areas remain in understanding the complexity of the interaction between PVY and its hosts and its vectors: How to explain the emergence and prevalence of PVY recombinants in crops? How is PVY aphid transmission activated? What are the plant and virus molecular determinants involved in pathogenicity (symptoms, resistance breakdown or induction)? How to better understand and exploit resistance genes against PVY? Are aphids the only insects to transmit PVY? A better understanding of the plant/PVY/vector pathosystem would allow us to control better PVY epidemics.

See also: Bean Common Mosaic Virus and Bean Common Mosaic Necrosis Virus (*Potyviridae*). Papaya Ringspot Virus (*Potyviridae*). Plum Pox Virus (*Potyviridae*). Potyviruses (*Potyviridae*)

Further Reading

- Faurez, F., Baldwin, T., Tribodet, M., Jacquot, E., 2012. Identification of new Potato virus Y (PVY) molecular determinants for the induction of vein necrosis in tobacco. *Molecular Plant Pathology* 13, 948–959.
- Glais, L., Bellstedt, D.U., Lacomme, C., 2017. Diversity, characterization and classification. In: Lacomme, C., Glais, L., Bellstedt, D.U., *et al.* (Eds.), *Potato Virus Y: Biodiversity, Pathogenicity, Epidemiology and Management*. Switzerland: Springer, pp. 43–76.
- Glais, L., Faurez, F., Tribodet, M., Boulard, F., Jacquot, E., 2015. The amino acid 419 in HC-Pro is involved in the ability of PVY isolate N605 to induce necrotic symptoms on potato tubers. *Virus Research* 208, 110–119.
- Janzac, B., Tribodet, M., Lacroix, C., *et al.*, 2014. Evolutionary pathways to break down the resistance of allelic versions of the PVY resistance gene *va*. *Plant Disease* 98, 1521–1529.
- Janzac, B., Willemsen, A., Cuevas, J.M., *et al.*, 2015. Brazilian Potato virus Y isolates identified as members of a new clade facilitate the reconstruction of evolutionary traits within this species. *Plant Pathology* 64, 799–807.
- Kim, S.B., Kang, W.H., Huy, H.N., *et al.*, 2017. Divergent evolution of multiple virus-resistance genes from a progenitor in *Capsicum* spp. *New Phytologist* 213, 886–899.
- Lacomme, C., Jacquot, E., 2017. General characteristics of potato virus Y (PVY) and its impact on potato production: An overview. In: Lacomme, C., Glais, L., Bellstedt, D.U., *et al.* (Eds.), *Potato virus Y: Biodiversity, Pathogenicity, Epidemiology and Management*. Switzerland: Springer, pp. 1–19.
- Michel, V., Julio, E., Candresse, T., *et al.*, 2018. NTPN1: A *RPP8-like* R gene required for Potato virus Y-induced veinal necrosis in tobacco. *Plant Journal* 95, 700–714.
- Michel, V., Julio, E., Candresse, T., *et al.*, 2019. A complex *elF4E* locus impacts the durability of *va* resistance to Potato virus Y in tobacco. *Molecular Plant Pathology* 20, 1051–1066.
- Moury, B., 2010. A new lineage sheds light on the evolutionary history of Potato virus Y. *Molecular Plant Pathology* 11, 161–168.
- Quenouille, J., Vassilakos, N., Moury, B., 2013. Potato virus Y: A major crop pathogen that has provided major insights into the evolution of viral pathogenicity. *Molecular Plant Pathology* 14, 439–452.
- Rowley, J.S., Gray, S.M., Karasev, A.V., 2015. Screening potato cultivars for new sources of resistance to Potato virus Y. *American Journal of Potato Research* 92, 38–48.
- Ruffel, S., Gallois, J.L., Lesage, M.L., Caranta, C., 2005. The recessive potyvirus resistance gene *pot-1* is the tomato orthologue of the pepper *pvr2-elF4E* gene. *Molecular Genetics and Genomics* 274 (4), 346–353.
- Uzest, M., Gargani, D., Drucker, M., *et al.*, 2007. A protein key to plant virus transmission at the tip of the insect vector style. *Proceedings of the National Academy of Sciences of the United States of America* 104, 17959–17964.
- White, K.A., 2015. The polymerase slips and PIP0 exists. *EMBO Reports* 16, 885–886.

Potexviruses (*Alphaflexiviridae*)

Ki H Ryu and Eun G Song, Seoul Women's University, Seoul, South Korea

Jin S Hong, Kangwon National University, Chunchon, South Korea

© 2021 Elsevier Ltd. All rights reserved.

This is an update of K.H. Ryu, J.S. Hong, Potexvirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00738-X](https://doi.org/10.1016/B978-012374410-4.00738-X).

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

EM Electron microscopy

FD1 Ferredoxin 1

Hel Helicase

ISH In situ hybridization

JAX1 Jacalin-type lectin gene

kb Kilobase

kDa Kilo Dalton

LAMP Loop mediated amplification

LFIA Lateral flow immunoassay

Met Methyl-transferase

MP Movement protein

nt Nucleotide(s)

NTR Non-translated region

ORF Open reading frame

PCR Polymerase chain reaction

Pol RNA-dependent RNA polymerase

Poly(A) Polyadenylated

RdRp RNA-dependent RNA polymerase

RT-PCR Reverse transcription polymerase chain reaction

SL Stem loop

ssRNA Single-stranded RNA

TEM Transmission electron microscopy

TGB Triple gene block

Glossary

Jacalin-type lectin 1 (JAX1) Antiviral resistance gene to Potato virus X related to lectin – mediated resistance (LMR) confers plant innate immunity against viruses.

RNA silencing suppressor Plant viruses have encoding protein to counteract the plant antiviral system known as RNA silencing.

Rx protein A highly similar CC-NBS-LRR protein, which prevents virus accumulation, mediates extreme resistance against PVX.

Triple gene block The protein comprises of a substantially similar element of three partially overlapping ORFs and is contained within their genomes. The TGB proteins are correlated with cell-to-cell movement and potentially influence on host antiviral defenses.

Introduction

A genus *Potexvirus* is one of seven genera in the family *Alphaflexiviridae*. Most potexviruses are found wherever their hosts are grown. The genus *Potexvirus* contains 38 definitive species including 35 species determined complete genome sequences. The genomes of Cassava virus X (CsVX) and Lagenaria mild mosaic virus (LaMMoV) have been partially sequenced and that of Plantain virus X (PIVX) has yet been reported. There are also 5 tentative species for which genomic sequences are available, but they have not yet been accepted by ICTV (**Table 1**). The members of the genus *Potexvirus* can be divided into three main phylogenetic groups based on their genome sequences. The phylogenetic relationships of the genus *Potexvirus* are not associated with their natural host other than cactus-infecting viruses. The virions of the genus *Potexvirus* have non-enveloped and flexible particles of 470–1000 nm in length and 12–14 nm in diameter. Their genomes of 5.8–7.2 kb in length consist of a positive-sense single-stranded RNA (ssRNA) with 5' terminal cap structure and a 3' terminal poly(A) tail. The genomes encode for an RNA-dependent RNA polymerase (RdRp) for viral replication, a triple gene block (TGBp1, 2, and 3) for viral movement, and a coat protein (CP). The RdRp is associated with jacalin-type lectin required for potexvirus resistance 1 (JAX1)-mediated resistance of host plant. The TGBp1, known as viral suppressors of RNA silencing, acts an avr protein associated with an HR-like response. Host plants naturally infected with potexviruses may show mosaic, mottle, necrosis, chlorosis, spots, or dwarf symptoms or may be symptomless. The natural host range of individual viruses is usually restricted to a few plant species, although a few of the viruses can infect a wide range of plant species. Potexviruses are transmitted mechanically through contact with virus-infected plant sap, or with contaminated agricultural equipment. Most potexviruses are not transmitted by vertebrate, invertebrate, or fungal vectors; Strawberry mild yellow edge virus (SMYEV) and Potato aucuba mosaic virus (PAMV) are transmitted by aphids. The type species of the genus *Potexvirus* is *Potato virus X* (PVX). During the last few decades, PVX has greatly contributed to our understanding of host resistance and gene silencing mechanisms, in particular because of the use of viral vectors in studying gene expression and RNA silencing in plants.

Table 1 Virus species in the genus *Potexvirus* of the family *Alphaflexiviridae*. Type species in the genus is written in bold. Complete sequences accession number and length (nt) of the genomes are indicated (ICTV, 2019)

Species name	Acronym	Accession number	Genome length (nt)
<i>Actinidia virus X</i>	AVX	NC028649	6888
<i>Allium virus X</i>	AlVX	NC012211	7118
<i>Alstroemeria virus X</i>	AlsVX	NC007408	7009
<i>Alternanthera mosaic virus</i>	AltMV	NC007731	6607
<i>Asparagus virus 3</i>	AV3	NC003400	6985
<i>Babaco mosaic virus</i>	BabMV	NC036587	6692
<i>Bamboo mosaic virus</i>	BaMV	NC001642	6366
<i>Cactus virus X</i>	CVX	NC002815	6614
<i>Cassava common mosaic virus</i>	CsCMV	NC001658	6376
<i>Cassava virus X</i>	CsVX	NC034375 ^a	5879
<i>Clover yellow mosaic virus</i>	CIYMV	NC001753	7015
<i>Cymbidium mosaic virus</i>	CymMV	NC001812	6227
<i>Foxtail mosaic virus</i>	FoMV	NC001483	6151
<i>Hosta virus X</i>	HVX	NC011544	6528
<i>Hydrangea ringspot virus</i>	HdRSV	NC006943	6185
<i>Lagenaria mild mosaic virus</i>	LaMMoV	NC043079 ^a	3600
<i>Lettuce virus X</i>	LeVX	NC010832	7212
<i>Lily virus X</i>	LVX	NC007192	5823
<i>Malva mosaic virus</i>	MalMV	NC008251	6858
<i>Mint virus X</i>	MVX	NC006948	5914
<i>Narcissus mosaic virus</i>	NMV	NC001441	6955
<i>Nerine virus X</i>	NVX	NC007679	6582
<i>Opuntia virus X</i>	OpVX	NC006060	6653
<i>Papaya mosaic virus</i>	PapMV	NC001748	6656
<i>Pepino mosaic virus</i>	PepMV	NC004067	6450
<i>Phaius virus X</i>	PhVX	NC010295	5816
<i>Pitaya virus X</i>	PiVX	NC024458	6677
<i>Plantago asiatica mosaic virus</i>	PIAMV	NC003849	6128
<i>Plantain virus X</i>	PIVX	ND	
<i>Potato aucuba mosaic virus</i>	PAMV	NC003632	7059
<i>Potato virus X</i>	PVX	NC011620	6435
<i>Schlumbergera virus X</i>	SchVX	NC011659	6633
<i>Strawberry mild yellow edge virus</i>	SMYEV	NC003794	5966
<i>Tamus red mosaic virus</i>	TRMV	NC016003	6495
<i>Tulip virus X</i>	TVX	NC004322	6056
<i>Vanilla virus X</i>	VVX	NC035205	6295
<i>White clover mosaic virus</i>	WCIMV	NC003820	5845
<i>Yam virus X</i>	YVX	NC025252	6158
<i>Zygocactus virus X</i>	ZyVX	NC006059	6624
Tentative species			
<i>Euonymus yellow vein virus</i>	EuYVV	NC035190	7279
<i>Potexvirus</i> sp.		NC040842	5839
<i>Senna mosaic virus</i>	SeMV	NC030746	6775
<i>Turtle grass virus X</i>	TuGVX	NC040644	6278

^aIncomplete genome sequence.

Taxonomy, Phylogeny, and Evolution

The genus *Potexvirus* is one of seven genera in the family *Alphaflexiviridae*. Its name is derived from *Potato virus X* (the type species of potexviruses). The genus contains 39 definitive species, of which 36 have complete genome sequences (Table 1). The complete genome sequences of some members have recently been determined. The genomes of Cassava virus X (CsVX) and LaMMoV have been partially sequenced. The genome sequences of PIVX has yet been reported. Another 4 tentative species have genomic sequences available, but they have not yet been accepted by ICTV (Table 1). Phylogenetic analysis using the complete genome sequences of the 36 potexviruses of their RNA-dependent RNA polymerase (RdRp), has revealed that they can be divided in seven major branches from I to VII (Fig. 1). Phylogenetic analysis of the complete sequences of the same 36 potexviruses of their coat protein (CP), (Fig. 1) and their triple gene block (TGB) proteins (Fig. 2), are showing that the seven clusters, with only a few exceptions, can be identified for all the proteins with the except for the TGBp3, which is showing a major reorganization of the potexvirus diversity. However it is also important to note that the TGBp3 phylogenetic tree is not very stable as only 5 branches have a bootstrap value above 50%. This result seems to indicate that there has been a co-evolution of the genes/proteins composing the potexvirus genome, with perhaps the exception of TGBp3. These

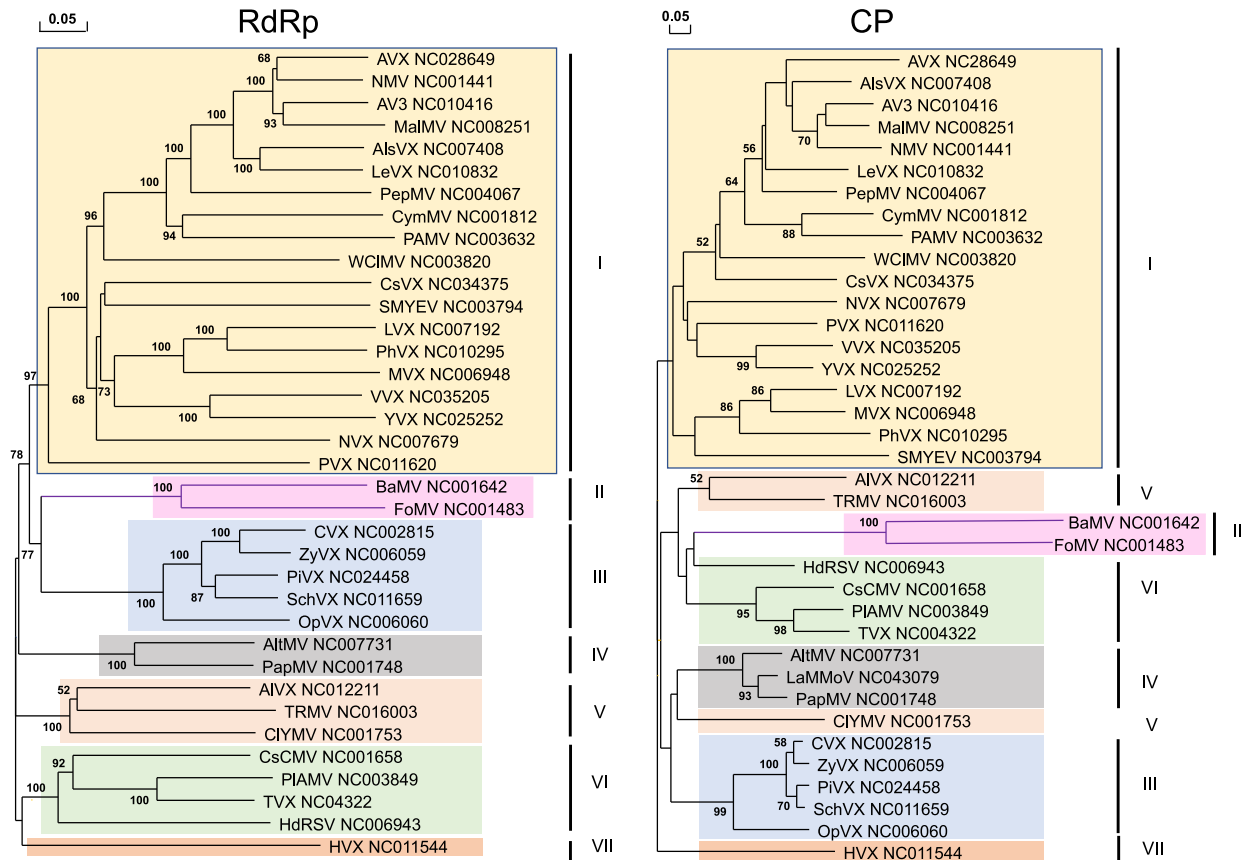


Fig. 1 Phylogenetic analysis of genus *Potexvirus* based on the amino acid sequences of the RdRp and CP. Multiple sequence alignments were generated using DNAMAN software version 5.1 (Lynnon Biosoft, San Ramon, Ca, USA) and phylogenetic trees were constructed by the neighbor-joining algorithm, based on calculations from pairwise amino acid sequence distances for protein analyses derived from the multiple alignment format. The horizontal branch lengths are proportional to the genetic distance, and the numbers at each point indicate bootstrap values. The dataset was subjected to 10,000 bootstrap replicates and values above 50% are indicated on the branches. Virus names are indicated by their abbreviation and their full names are in [Table 1](#). NCBI database accession numbers follow the virus abbreviation name. The black vertical bars identify the seven major clusters (from I to VII) as defined per the RdRp phylogenetic tree and differential colored boxes are shown in each tree.

phylogenetic trees do not show a correlation of the viruses with their type of host with the exception of the cactus-infecting viruses such as Cactus virus X (CVX), Pitaya virus X (PiVX), Opuntia virus X (OpVX), Schlumbergera virus X (SchVX), and Zygocactus virus X (ZyVX) that are grouped together in each tree (cluster III).

The species demarcation criteria in the genus *Potexvirus* are the followings: (1) members of distinct species have less than 78% nt or 61% aa identity of their RdRp, less than 73% nt or 73% aa identity of their TGBp1, less than 65% nt or 64% aa identity of their TGBp2, less than 66% nt or 59% aa identity of their TGBp3, and less than 57% nt or 49% aa identity of their CP; (2) members of distinct species are readily differentiated by serology, and some virus strains can be differentiated with monoclonal antibodies; and (3) members of distinct species fail to cross-protect in common host plant species, and they usually have distinguishable experimental host ranges.

Virion Structure

Virions are non-enveloped and flexible filaments of 470–1000 nm in length and 12–14 nm in diameter ([Fig. 3](#)). They have helical symmetry with a pitch of 3.3–3.6 nm. The number of protein subunits per turn of the primary and final symmetries is slightly less than 9.0. The viral RNA is at a radius of 3.0–3.5 nm. Virion Mr is about 3.5×10^6 with 5%–7% nucleic acid content. PVX virion is a filamentous particle of about 515 nm in length and 13.5 nm in diameter and forms a helical array (3.5 nm pitch) of about 1300 identical CP subunits with the viral RNA packed between turns of the helix. The virion has 8.9 CP subunits, each consisting of 236 aa residues, per primary helix and five nucleotides are related to each protein subunit. The viral RNA molecule have about 6% by weight of the virion. Narcissus mosaic virus (NWV) helix has eight subunits per turn and an outer radius of 5.5 nm. The assembly of potexvirus ribonucleoproteins does not depend on the RNA nucleotide sequences.

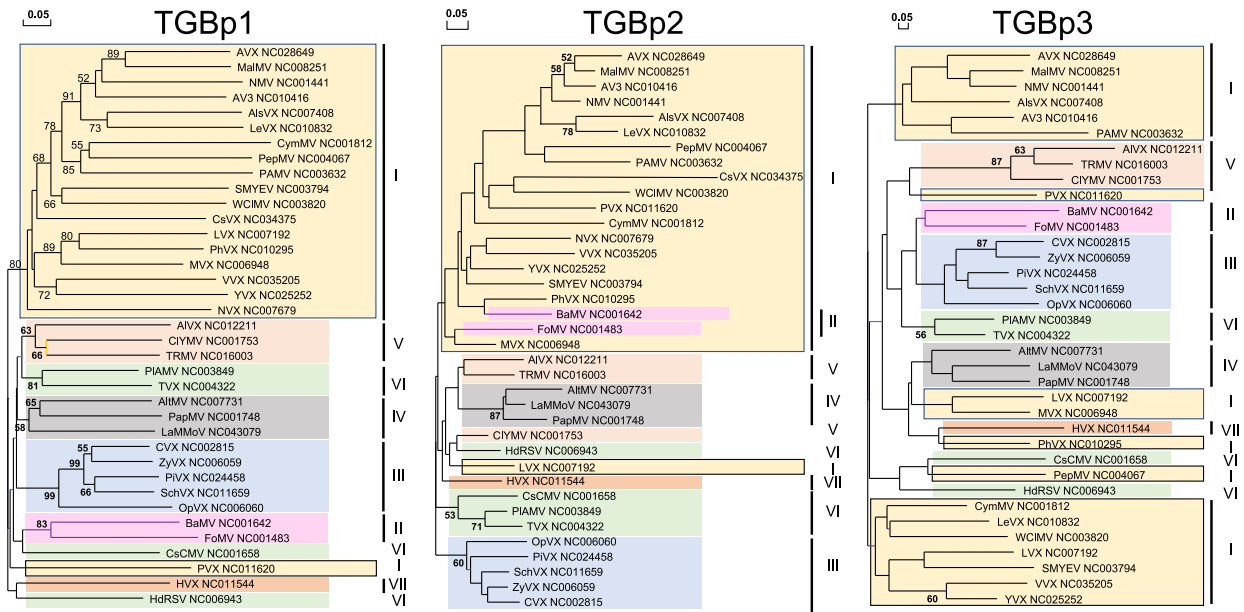


Fig. 2 Phylogenetic analysis of genus *Potexvirus* based on the amino acid sequences of the TGBp1, TGBp2, and TGBp3. Multiple sequence alignments were generated using DNAMAN software version 5.1 (Lynnon Biosoft, San Ramon, Ca, USA) and phylogenetic trees were constructed by the neighbor-joining algorithm, based on calculations from pairwise amino acid sequence distances for protein analyses derived from the multiple alignment format. The horizontal branch lengths are proportional to the genetic distance, and the numbers at each point indicate bootstrap values. The dataset was subjected to 10,000 bootstrap replicates and values above 50% are indicated on the branches. Virus names are indicated by their abbreviation and their full names are in [Table 1](#). NCBI database accession numbers follow the virus abbreviation name. The black vertical bars identify the seven major clusters (from I to VII) as defined per the RdRp phylogenetic tree and differential colored boxes are shown in each tree.

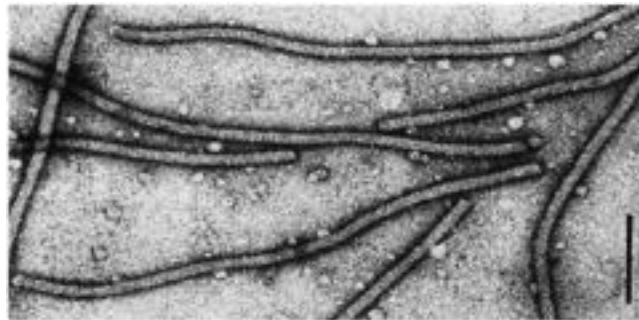


Fig. 3 Electron microscopy of Potato virus X (PVX) particles negatively stained. The scale bar is 100 nm. Picture courtesy of Lesemann, D.E.

Genome Organization

The genome of potexviruses is a sense single-stranded positive-sense RNA (ssRNA) with 5' terminal cap structure and a 3' terminal poly (A) tail. The genome of 5.8–7.2 kb in size contains five ORFs (encoding an RdRp), three movement proteins called TGB, and the CP from the 5' to the 3' end ([Fig. 5](#)). The genome organization resembles that of genera *Foveavirus* and *Robigovirus* in the family *Betaflexiviridae*, but is distinguishable from them by the size of genome (8–9.3 kb). The complete genomic sequences of almost all potexviruses has been determined ([Table 1](#)). The genome structure of PVX is showed in [Fig. 4](#). The genome consists of 6435 nt and contained five ORFs coding for the proteins of an RdRp of 166 kDa (ORF1; 4371 nt), a TGBp1 of 25 kDa (ORF2; 681 nt), a TGBp2 of 12 kDa (ORF3; 348 nt), a TGBp3 of 8 kDa (ORF4; 213 nt), and a CP of 25 kDa (ORF5; 714 nt) from the 5' to 3' end. Both terminal regions of ORF3 (12 kDa) shows overlap with ORF2 and ORF4, respectively. The 5' terminal region of ORF5 shows an overlap with ORF3. The lengths of the 5' non-translated region (NTR) and the 3' NTR have 84 and 73 nt. The 3' NTR region contains the poly(A) tail. Some potexviruses eg, Actinidia virus X (AVX); Alstroemeria virus X (AlsVX); Cymbidium mosaic virus (CymMV); NMV; PVX; SMYEV; White clover mosaic virus (WClMV), have a putative ORF6 located within the ORF5 or over the ORF5 and 3' NTR. Bamboo mosaic virus (BaMV) is frequently associated with a satellite RNA (satRNA). The satRNA is a linear RNA of 836 nt and comprises a single ORF encoding a 20 kDa protein.

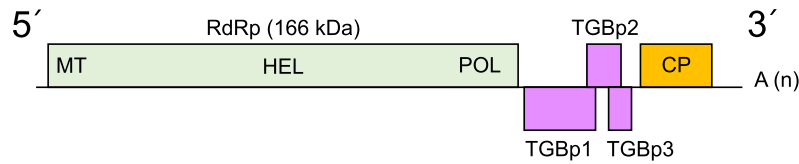


Fig. 4 Genome organization of Potato virus X, the type species of the genus *Potexvirus*. RNA-dependent RNA polymerase (166 kDa) (green box) includes a methyl-transferase (Met), a helicase domain (Hel), a polymerase (Pol) domains. Purple boxes indicate the triple gene block corresponding to three overlapping ORFs (TGBp1, TGBp2, and TGBp3), and ORF5 encodes the CP (orange box).

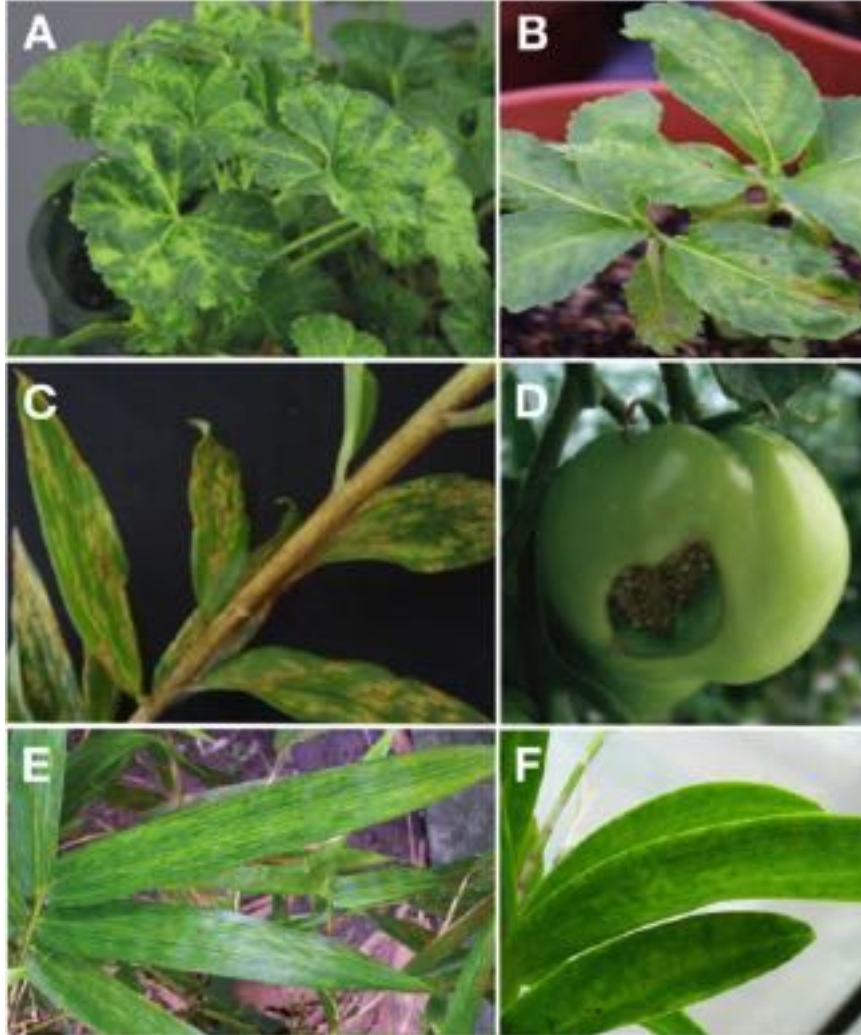


Fig. 5 Symptoms of potexvirus-infected plants. (A) *Malva* spp. infected with Malva mosaic virus (MalMV) exhibiting symptom of mosaic on leaves. (B) *Hydrangea* spp. infected with Hydrangea ringspot virus (HdRSV) exhibiting symptoms of chlorotic and necrotic ringspots on leaves. (C) *Lilium* spp. infected with Plantago asiatica mosaic virus (PIAMV) exhibiting symptoms of brown necrotic streaking and irregular chlorosis on leaves. (D) *Solanum lycopersicum* infected with Pepino mosaic virus (PepMV) exhibiting symptoms of open fruit. (E) *Bambusa vulgaris* infected with Bamboo mosaic virus (BaMV) exhibiting symptoms of interveinal chlorotic mosaic and stripes on leaves. (F) *Dendrobium phalaenopsis* infected with Cymbidium mosaic virus (CymMV) exhibiting symptoms of chlorotic and necrotic ringspots on leaves.

Properties and Functions of Gene Products

The RdRP of potexviruses contains three conserved domains: a methyl-transferase (Met) domain; an RNA helicase (Hel) domain; a polymerase (Pol) domain. The RdRP is related to jacalin-type lectin required for potexvirus resistance 1 (JAX1)-mediated resistance of host plant. The cell-to-cell movement of potexviruses requires the potexviruses TGB and CP. TGBp1 and CP form a complex with viral RNA that transfers to the next cell via the plasmodesmata. The TGBp1 also acts as viral suppressors of RNA silencing and an avr protein

associated with an HR-like response. It has been recently demonstrated that PVX TGBp1 interacts with the chloroplast protein ferredoxin 1 (FD1) related to its role in the chloroplast electron transport chain. The TGBp1 of Pepino mosaic virus (PepMV) is associated with a thioredoxin domain-containing protein that is related to virus movement. TGBp2 and TGBp3 are endoplasmic reticulum-binding proteins for virus movement. The TGBp2 is also associated with a functional role in PVX replication. BaMV TGBp2 interacts with a *Nicotiana benthamiana*-expressed thioredoxin type h protein (*NbTRXh2*), which acts as a negative regulator of viral movement. CP is required for virion assembly. PVX CP is the elicitor for Rx-mediated resistance that impairs virus accumulation in host plant. The CP of Potato aucuba mosaic virus (PAMV) has a DAG motif associated with aphid transmissibility, located in the aa residues 14–16 from the N-terminus of the CP, which is also found in the CP of potyviruses. Satellite RNA (satRNA), with crucial roles in modulating viral replication and symptom development, is associated with BaMV and it was found to be the only satRNA of the genus *Potexvirus*. The RNA-binding domain of PAMV CP, located between 90 and 130 aa, is associated with virion assembly.

Replication and Propagation

The 5' NTR regulates plus-strand RNA synthesis, sub-genomic RNA synthesis, cell-to-cell movement, and virion assembly. Two RNA stem-loop (SL) structures within the 5' NTR, named 5' SL1 and 5' SL2, are required for viral replication. The 5' SL1 is a multifunctional element associated with virus replication, gene expression, cell-to-cell movement, and virion assembly. RdRP can be translated directed from the genomic RNA and then three sub-genomic RNAs are generated for TGB and CP synthesis. The 3' NTR including three RNA stem-loop structures, named 3' SL1, 3' SL2, and 3' SL3, controls minus-strand and plus-strand RNA accumulations. The 3' NTR has multifunctional activities, related to viral replication and host-factor interaction. The RNA replication and accumulation of PVX interact synergistically with potyviruses (Pepper mottle virus, Plum pox virus, Potato virus Y, Tobacco vein mottling virus, and Tobacco etch virus). The PVX/potyviral-associated synergism is related to the 5' proximal potyviral sequence and microRNAs derived from co-infected plants. The 5' NTR of satRNA, which associated with BaMV, acts as a regulator of BaMV replication. The transgene induced from potexvirus-based vector can be stably expressed as a N-terminal translational CP fusion linked by the autocatalytic 2A peptide, which is derived from foot-and-mouth disease virus.

Transmission, Host Range

Potexviruses can be naturally transmitted to various horticultural crops. Most potexviruses can be transmitted mechanically via contact with virus-infected plant sap, or with contaminated agricultural equipment. PAMV and PVX cause common symptoms in *Solanum tuberosum* grown worldwide. PAMV exhibiting slight leaf mottling on *S. jasminoides* causes yellow leaf flecking, deformation, and stunting on susceptible potato plants. PVX cause mild mosaic, necrosis, or is symptomless on potato plants and can be co-infected with Potato virus Y and *Candidatus phytoplasma trifolii* on *S. melongena*. Five viruses, CVX, OpVX, PiVX, SchVX, and ZyVX, cause mild mosaic, mosaic, necrosis, ringspots or are symptomless on *Cactaceae*. The CVX, OpVX, SchVX, and ZyVX can be transmitted by mechanical inoculation. LVX, PIAMV, and Tulip virus X (TVX) can be naturally transmitted to *Lilium* spp. The PIAMV can cause mosaic and necrosis on *Plantago asiatica* and *Solanaceae*, as well as *Lilium* spp. The TVX can cause chlorotic and necrotic spots of leaves and streaks of petals in *Lilium* spp. and *Tulipa* spp. The LVX appears to be asymptomatic in *Lilium* spp. Cassava common mosaic virus (CsCMV) and CsVX can be naturally transmitted to *Manihot esculenta* and these viruses can be mechanically transmitted to *N. benthamiana*. The CsCMV causes mosaic and chlorosis on *M. esculenta* while the CsVX exhibits no symptoms. Clover yellow mosaic virus (CIYMV) and WCIMV can be naturally transmitted to *Trifolium* spp. The CIYMV causes chlorotic streaks of leaves and color breaking of flowers on *T. gesneriana* and it causes yellow mosaic or is symptomless on *T. repens*. CymMV causes chlorotic or necrotic spots on the leaves and flowers of orchid plants. The flowers infected with the CymMV exhibit deformation and color breaking. CymMV has also been associated with flecking on leaves, necrotic spots on the stem, and vine decline of vanilla plants. Papaya mosaic virus (PapMV) causes mosaic and stunting on *Carica papaya* while it exhibits no symptoms on *Ullucus tuberosus*. PepMV can be naturally transmitted to greenhouse tomato species (*S. lycopersicum*) and wild tomato species (*S. chilense*, *S. chmielewskii*, *S. parviflorum*, and *S. peruvianum*) and it causes fruit marbling and discoloration, open fruit, nettle-heads, leaf blistering or bubbling, leaf chlorosis and yellow angular leaf spots, leaf mosaic and leaf or stem necrosis in susceptible tomato plants. SMYEV causes marginal yellowing symptoms on leaves of commonly cultivated strawberry crops in worldwide. LVX, SMYEV, PAMV, and PepMV can be transmitted by aphids. Hosta virus X (HVX) can be naturally transmitted to *Hosta* spp. and cannot be transmitted mechanically to *N. benthamiana* and *N. clevelandii*. HVX and PepMV can be transmitted by seeds. BaMV can be transmitted by dipterans (*Gastrozona fasciventris* and *Atherigona orientalis*). Some viruses are not transmitted by mechanical inoculation. Allium virus X (AIVX) causes diffuse yellow stripes or mottle on leaves of *Allium* spp. Yam virus X (YVX) causes mild chlorosis on leaves of *Dioscorea trifida*. The methods for transmission of three potexviruses, such as AIVX, PiVX, and YVX, are presently not known.

Epidemiology and Control

Most potexviruses can cause usually mild mosaic, mosaic, chlorosis, spots, or symptomless in their host plants. Some food and ornamental crops sustain economic damage caused by potexviruses that require suitable disease management. Transgenic cactus,

orchid, potato, and tobacco plants, which are resistant to BaMV, CymMV, CVX, PVX, and SMYEV respectively, have been developed. PVX can cause yield losses reaching 20% of productivity depending upon the virus strain. New potato varieties carrying the *Rx1* and/ or *Rx2* genes derived from the wild species *Solanum tuberosum* ssp. *andigena* and *S. acaule*, respectively, have been developed to control the virus. Salicylic acid has been demonstrated to mitigate physiological and proteomic changes caused by PVX (SPCP1 strain) in tomato. CsCMV has been found in cassava, which grows in the Americas. The CsCMV can cause a reduction of up to 60% of root productivity. CymMV has been reported in most cultivated orchid species grown in worldwide since it was first described on *Cymbidium* in California. The virus has also been found in vanilla from most growing areas since it was first detected in vanilla (*Vanilla* spp.) in the Society Islands (1987). Skimmed milk represented inactivation of CymMV on local host, but has not been evaluated on systemic host. RNA silencing is an important defense mechanism against plant viruses and it has been applied to control potexviruses. The potexvirus TGBp1, is known as RNA silencing suppressor, and it has been used frequently for producing virus-derived small interfering RNAs. PLAMV has been isolated from various species worldwide (eg, *Achyranthes bidentate*; *Lilium* spp.; *Nandina domestica*; *Primula sieboldii*; *Rehmannia glutinosa*; *Stellaria media*; *Urtica urens*; *Viola grypoceras*), since it was first reported from *Plantago asiatica* grown in Russia (1976). This viral disease is especially known to reduce considerably the commercial value of lilies. Early bulb test of the viral infection should be carried out in order to control the virus in greenhouse. PepMV causes a severe disease in greenhouse tomato crops worldwide, since it was found in the Netherlands. The factor associated with the epidemiology of the PepMV can be dependent on the properties of the different virus strains. The virus is currently controlled using virus-free seeds. SMYEV has often been found in co-infection with other viruses, such as Arabis mosaic virus (ArMV, genus *Nepovirus*), Strawberry crinkle virus (unclassified within the family *Rhabdoviridae*), Strawberry mottle virus (unclassified within the family *Secoviridae*), Strawberry polerovirus 1 (unclassified within family *Luteoviridae*), Strawberry pallidosis-associated virus (unclassified within the family *Closteroviridae*), or Tomato ringspot virus (ToRSV, genus *Nepovirus*), and the viral co-infection can significantly reduce the yield of cultivated strawberry crops grown in greenhouses. Chemotherapy with ribavirin has recently been tested to control of the ArMV, SMYEV, and ToRSV.

Diagnosis

Potexviruses can be detected by polymerase chain reaction (PCR)-based methods, such as reverse transcription (RT)-PCR, immunocapture-RT-PCR, loop-mediated isothermal amplification (RT-LAMP) assay, multiplex RT-PCR, and real-time PCR. Some commercial antisera are used for immunological detection of potexviruses. Lateral flow immunoassay (LFIA) used with alkaline phosphatase has been developed for PVX detection and it has a lower limit of detection (0.3 ng/mL^{-1}) than that of conventional LFIA. Electron microscopy (EM) is an essential tool for virus detection and transmission electron microscopy (TEM) is especially used to observe the virus particles in virus-infected plants. Amorphous inclusion bodies, formed by virus particles, can be observed in PVX-infected plant cells. *In situ* hybridization (ISH) technique detecting the mixed-infection of two PepMV isolates has been recently developed. Digoxigenin-labeled RNA probes, which targeted PepMoV genomic regions exhibiting a highly specificity and sensitivity for the two isolates, was used in the ISH technique.

Concluding Remarks

The complete genome sequences of AVX, YVX, and VVX have been determined recently. Novel potexviruses have been isolated from various plants. The genome of three definitive species, CsVX, LaMMoV, and PIVX, have not yet been completely sequenced. The phylogenetic relationship based on their genome sequences are not significantly associated with their natural host other than cactus-infecting viruses containing CVX, PiVX, OpVX, SchVX, and ZyVX. Some potexviruses have a putative ORF 6 within CP coding region or CP-3' NTR regions. Potexviruses are commonly known to be transmitted by aphids. Additionally, BaMV can be transmitted by diptera insects. *Rx* and *JAX1* genes regulate plant innate immunity against PVX. TGBp1 of PVX is a suppressor of RNA silencing that protects host cells against the pathogen. Potexviruses are currently mostly controlled by using virus-free seed production and early test of viral infection. Immunological and PCR-based techniques are most commonly used for detection of potexviruses.

Further Reading

- Aguilar, E., del Toro, F.J., Brosseau, C., *et al.*, 2019. Cell death triggered by the P25 protein in Potato virus X-associated synergisms results from endoplasmic reticulum stress in *Nicotiana benthamiana*. *Molecular Plant Pathology* 20, 194–210.
- Baulcombe, D.C., Lloyd, J., Manoussopoulos, I.N., Robert, I.M., Harrison, B.D., 1993. Signal for potyvirus-dependent aphid transmission of Potato aucuba mosaic virus and the effect of its transfer to Potato virus X. *Journal of General Virology* 74, 1245–1253.
- Chang, K., Chang, L., Huang, Y., *et al.*, 2017. Transmission of Bamboo mosaic virus in bamboo mediated by insects in the Order *Diptera*. *Frontier in Microbiology Virology* 8, 870.
- Cruz, S.Z., Chapman, S., Roberts, A.G., *et al.*, 1996. Assembly and movement of a plant virus carrying a green fluorescent protein overcoat. *Proceedings of the National Academy of Sciences of the United States of America* 93, 6286–6290.
- Hanssen, I.M., Thomma, B.P.H.J., 2010. Pepino mosaic virus: A successful pathogen that rapidly evolved from emerging to endemic in tomato crops. *Molecular Plant Pathology* 11, 179–198.

- Pacheco, R., Garcia-Marcos, A., Barajas, D., Martiane, J., Tenllado, F., 2012. PVX-potyvirus synergistic infections differentially alter microRNA accumulation in *Nicotiana benthamiana*. *Virus Research* 165, 231–235.
- Petrova, E.K., Trifonova, E.A., Nikitin, N.A., *et al.*, 2016. Study of the potexvirus ribonucleoproteins signal of assembly. *Virology* 71, 45–49.
- Ruiz-Ramon, F., Sempere, R.N., Mendez-Lopez, E., Sanchez-Pina, M.A., Aranda, M.A., 2019. Second generation of Pepino mosaic virus vectors: Improved stability in tomato and wide range of reporter genes. *Plant Methods* 15, 58.
- Sugawara, K., Shiraishi, T., Yoshida, T., *et al.*, 2013. A replicase of Potato virus X acts as the resistance-breaking determinant for JAX1-mediated resistance. *Molecular Plant-Microbe Interactions* 26, 1106–1112.
- Verchot-Lubicz, J., Ye, C., Bamunusinghe, D., 2007. Molecular biology of potexviruses: Recent advances. *Journal of General Virology* 88, 1643–1655.

Potyriviruses (*Potyriviridae*)

Adrián Valli and Juan A García, National Center for Biotechnology-Spanish National Research Council, Madrid, Spain

Juan J López-Moya, Center for Research in Agricultural Genomics and Spanish National Research Council, Barcelona, Spain

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J.J. López-Moya, J.A. García, Potyriviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00475-1](https://doi.org/10.1016/B978-012374410-4.00475-1).

Glossary

Pinwheel Characteristically shaped cylindrical inclusions present in the cytoplasm of cells infected by potyvirids.

PIPO (Pretty Interesting Potyviral ORF) Conserved ORF (Open Reading Frame) embedded within the genome of potyvirids in a frame different of the polyprotein, and expressed through polymerase slippage occurring in a conserved motif.

Polyprotein Protein, generally big in size, made of several smaller proteins. It is the gene expression strategy of potyvirids, as they are translated from large ORFs further processed by virus-encoded proteases.

Potyvirus Member of the family *Potyriviridae*.

Virion Viral particle.

Introduction

The existence of potyvirids as a specific group of plant pathogens with numerous members was recognized soon after the onset of virology as a scientific discipline. Historical evidence of symptoms caused by potyvirids in plants includes the color-breaking of infected tulip flowers that were reproduced in old Dutch paintings. Some members of this family were among the first plant viruses to be identified due to the importance of the diseases they cause, and it happened relatively soon after potyvirids were grouped according to common characteristics. Electron microscopy provided a clear taxonomic criterion to allocate viruses into the group by showing flexuous rod-shaped viral particles, and the consistent presence of pinwheel-shaped cytoplasmic inclusions in infected plant cells. Biological and serological studies, including vector organisms, genome composition and the deciphering of their full-length genome sequences, have contributed to taxonomically differentiate potyvirids into several genera. Currently, a still evolving classification recognizes at least ten genera and a few unclassified viruses.

The large number of viral species within the family *Potyriviridae* reveals its evolutionary success. More than 210 species are included in the family, with approximately 175 in the aphid-transmitted genus *Potyvirus*. Members of the family are distributed throughout the world, with each particular virus being geographically limited to the area potentially occupied by susceptible host plants. Generally speaking, host ranges are constrained to a limited number of natural and experimental hosts, although some potyvirids can infect a considerable number of plant species in many botanical families. Members of the family can infect economically important crops, including grain, legumes, forages, vegetables, fruits and ornamentals, causing severe losses and consequently fostering the interest of researchers, plant breeders and agronomists seeking practical solutions to the diseases that they cause.

The present entry focuses on current knowledge about viruses in the family *Potyriviridae*, dealing with taxonomy, evolution, diagnostics, molecular biology, functional and structural aspects of viral proteins, vector transmission and host-virus interactions, including characterization of resistance genes and defense responses, as well as biotechnological applications.

Taxonomy, Phylogeny and Evolution

Taxonomic standards for classification into the *Potyriviridae* family include: (i) properties of virus particles: long flexuous filaments (Figs. 1(a) and 2); (ii) cytopathological manifestations: presence of pinwheel or scroll-shaped cylindrical cytoplasmic inclusions in infected plant cells (Fig. 1(b)), and (iii) genome structure and expression strategy: positive sense ssRNA genomes, with 5' terminal proteins and 3' poly-A tails, translated as large polyprotein precursors, and at least one embedded out-of-frame ORF (see Fig. 4).

The family *Potyriviridae* is related to the picornavirus supergroup, having an equivalent main strategy for genome expression through polyprotein synthesis and further processing, and a well conserved set of replication-related proteins being present at equivalent positions in the different genomes, in particular a viral RNA-dependent RNA polymerase (RdRp).

Within the family, ten different genera are differentiated so far according to biological and molecular criteria. Potyviruses, for instance, are aphid-transmissible and possess a monopartite genome, characteristics that they share with macluraviruses, although particles of the latter are shorter in length and their first gene product in the polyprotein is missing. Other monopartite genera include whitefly-transmitted ipomoviruses, and eryophid mite-transmitted poacevirus, rymoviruses and tritoviruses, with vector organisms belonging to genera *Abacarus* and *Aceria*. The remaining genera (*Bevemovirus*, *Brambyvirus* and *Roymovirus*) include few species and, along with unclassified viruses, they still have no assigned vector. Bymoviruses, in turn, have bipartite genomes being encapsidated separately, and plasmodiophoraceous fungus-like organisms of the genus *Polymyxa* act as their transmission vectors. Phylogenetic analysis based on the alignment of genomic regions has established standard thresholds of around 50% and 75% of

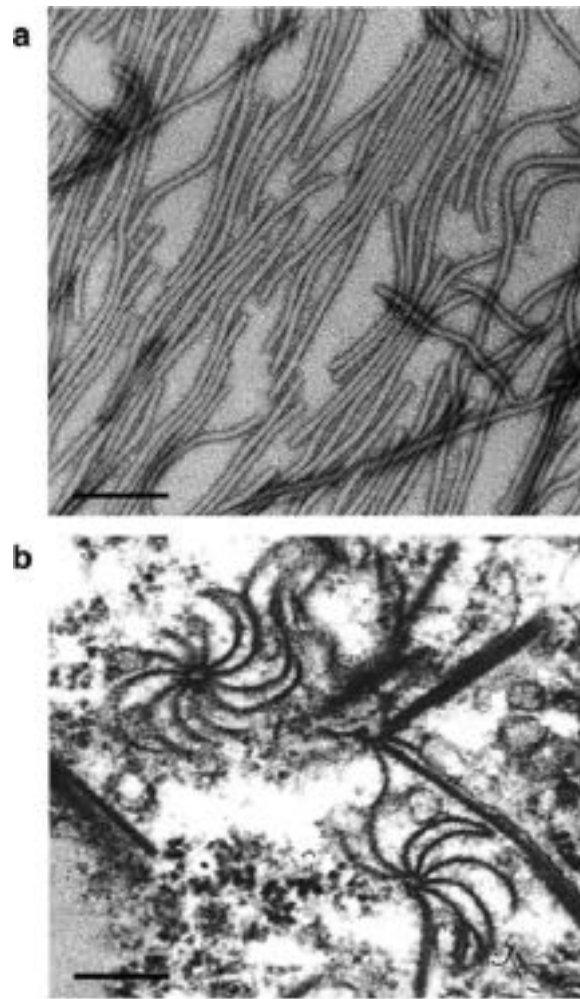


Fig. 1 (a) Negative stain preparation of purified TEV particles. (b) Thin section showing the typical pinwheel shaped cytoplasmic inclusions present in cells of a *Nicotiana benthamiana* plant infected with PPV (bars equal 200 nm). Pictures are courtesy of D. López-Abella, CIB, CSIC.

nucleotide identities (variable depending on the considered segments) as demarcation criteria for assigning viruses to genera and species, respectively. The current classification is shown in [Table 1](#).

An example of phylogenetic analysis of potyvirus genomes is shown in [Fig. 3](#). All genera are differentiated in well-defined branches, with rymoviruses being close to potyviruses, although their distinctive biological characteristics justify their placement in different genera. Viral subgroups, particularly in members of the large genus *Potyvirus*, are also used for classification, following for instance host range restrictions as biological demarcation criteria. Information about taxonomy and sequence of members of the family *Potyviridae* is available at several web-based databases ([Table 2](#)).

Recombination, switching events, radiation, as well as host and geographical adaptation are proposed as drivers of evolution. Sub-populations able to differentiate and evolve independently within infected perennial plants illustrate the versatility and capacity of adaptation, and similar phenomena can happen in epidemics of viruses affecting annual hosts. Recombination of virus genomes can give rise to new virus entities with novel phenotypes in virulence, induction of symptoms, and even host range. Therefore, recombination events, partial duplications, indels and point mutations conferring different responses in infected host, as well as other factors, might help to explain the extraordinary variability observed among potyvirids.

Virion Structure

Virions of viruses belonging to the family *Potyviridae* are flexuous rods made by protein (95%) and RNA (5%) ([Figs. 1\(a\)](#) and [2](#)). They are about 11–15 nm wide, with lengths ranging from less than 700 nm, as in the case of macluraviruses, to up to 900 nm, as in ipomoviruses ([Table 1](#)). Available data on particle structure show a helical assembly of identical coat protein (CP) subunits

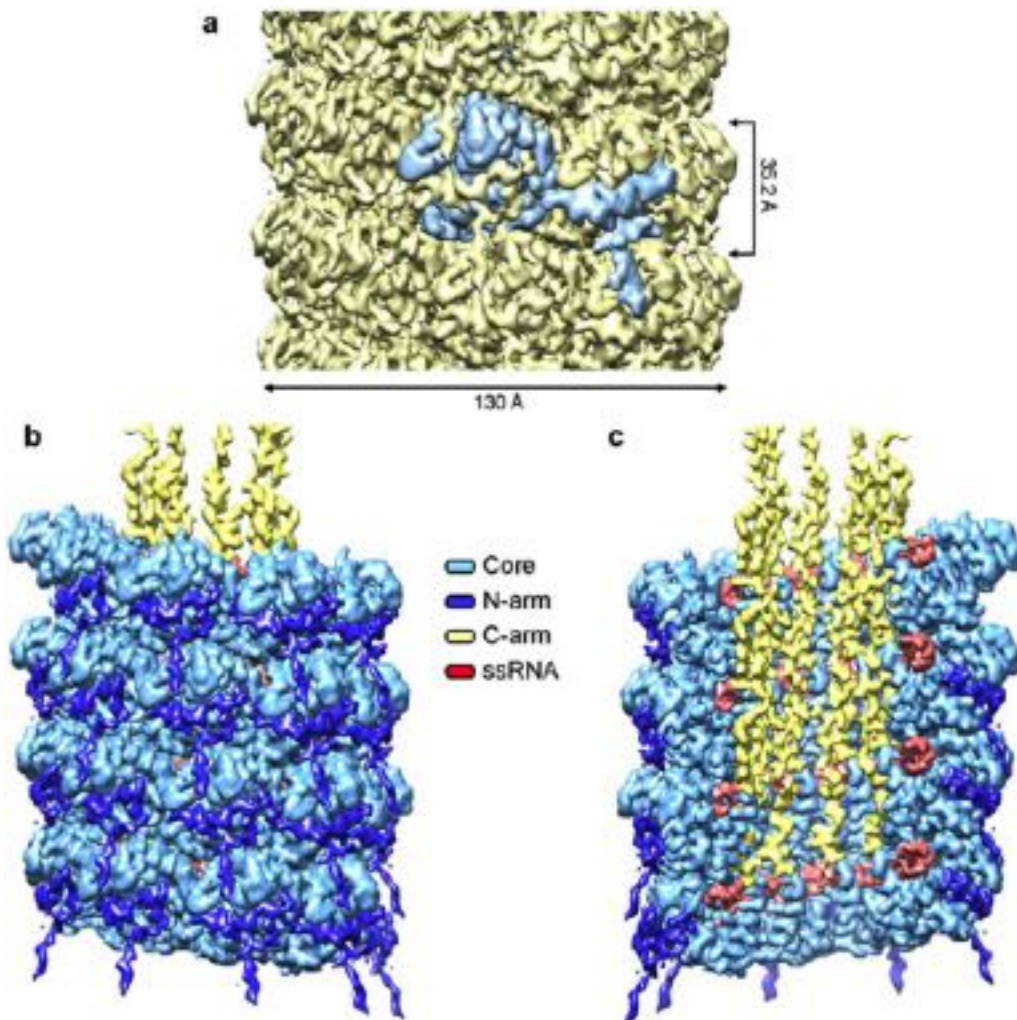


Fig. 2 Structure of potyvirus virions as determined at near-atomic resolution by cryoEM for WMV. (a) The diameter and pitch dimensions corresponding to the helical arrangement of 8.8 CP subunits per turn is shown in 3D density map (resolution 4.0 Å), with one of the subunit highlighted in blue. (b and c) The different parts are indicated according to the color code in the model and a cutaway section shows the inner side of the helix that forms a 70 Å channel. Images are courtesy of M. Valle, CIC bioGUNE, Spain. Reproduced from Zamora, M., Méndez-López, E., Agirrezabala, X., *et al.*, 2017. Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. *Science Advances* 3. (eaao2182).

(about 2000) surrounding the nucleic acid, and a distribution of 8, 8 subunits per turn in the case of *Watermelon mosaic virus* (WMV).

Molecular weight of CP subunits ranges from 30 to 40 kDa, with differences mainly due to the variable length of the N-terminal region. A more conserved internal CP core (about 220 amino acids) defines particle architecture. The N-terminus appears to be surface-exposed, while the C-terminal part is located in the interior. Superficially located residues might interact with other proteins during essential processes of the virus life cycle. As an example, the conserved DAG motif near the N-terminus of the CP is required for aphid transmission. Posttranslational modifications, including phosphorylations and glycosylations, have been found in the CP of potyviruses, having potential regulatory functions during the viral cycle. Also, peculiar structures at one of the virion ends might represent the presence of VPg, HCpro and CI proteins, as observed in particles of purified Potato virus Y (PVY) and Potato virus A (PVA).

Properties of the Genome

The genome of monopartite potyvirids is an ssRNA molecule of + (messenger) sense. In all members, genomic RNA presents a 5' terminal protein (VPg) and a 3' poly-A tail of variable length. The genome size ranges from 8.2 to 11.5 kb, and it comprises an ORF that spans the whole genome and codes for a long polyprotein (340–370 kDa), which generates different mature products after ongoing an autoproteolytic processing cascade. An extra ORF, embedded in the P3 cistron and transcribed by viral RNA polymerase slippage events,

Table 1 Classification of virus species of the family *Potyviridae* into genera, with indication of type species, number of virus species, number of genomic RNAs, and taxonomically relevant characteristics like transmission vectors and particle morphology

Genus	Type species	Abbreviation	Number ^a	Genome	Particle ^b	Vector ^c
<i>Bevemovirus</i>	<i>Bellflower veinal mottle virus</i>	(BVMoV)	1	Monopartite	760 nm	Unknown
<i>Brambyvirus</i>	<i>Blackberry virus Y</i>	(BVY)	1	Monopartite	800 nm	Unknown
<i>Bymovirus</i>	<i>Barley yellow mosaic virus</i>	(BYMV)	6	Bipartite	250–300 and 500–600 nm	Fungus-like
<i>Ipomovirus</i>	<i>Sweet potato mild mottle virus</i>	(SPMMV)	7	Monopartite	900 nm	Whiteflies
<i>Macluravirus</i>	<i>Maclura mosaic virus</i>	(MacMV)	8	Monopartite	650–675 nm	Aphids
<i>Poacevirus</i>	<i>Triticum mosaic virus</i>	(TriMV)	3	Monopartite	680–750 nm	Mites
<i>Potyvirus</i>	<i>Potato virus Y</i>	(PVY)	175	Monopartite	700–900 nm	Aphids
<i>Roymovirus</i>	<i>Rose yellow mosaic virus</i>	(RoYMV)	1	Monopartite	Unreported	Unknown
<i>Rymovirus</i>	<i>Ryegrass mosaic virus</i>	(RyMV)	3	Monopartite	700 nm	Mites
<i>Tritimovirus</i>	<i>Wheat streak mosaic virus</i>	(WSMV)	6	Monopartite	690–700 nm	Mites
Unclassified ^d	–		(5)	–	–	–

^aNumber of species according to the International Committee on Taxonomy of Viruses (Master Species List 2018b.v1).^bApproximate length of virus particles.^cVector organisms responsible for transmission.^dAmong viruses still unclassified into genera there are members of two potentially new genera *Arepavirus* and *Celavirus*.

codes for the Pretty Interesting Potyviral QRF (PIPO) and generates the P3N-PIPO product. The bipartite genome of bymoviruses is divided into two RNAs: a long RNA-1 homologous to the 3' three-quarters portion of potyvirus monopartite genomes, and a short RNA-2 with partial similarity to the 5' region of the genome of other potyvirids. The proteolytic processing of the viral polyprotein depends on the action of either two (P1 and NIap_{ro}, or HCpro and NIap_{ro}) or three (P1, HCpro and NIap_{ro}) virus-encoded proteinases, depending on the species. In some cases, two divergent copies of P1 or HCpro could be part of the proteinase set of a single potyvirus. P1 serine- and HCpro cysteine-proteinases cleave in *cis* at their respective C-ends, whereas the NIap_{ro} serine-proteinase cleaves all the other sites, both in *cis* and *trans*. Information on processing cleavage sites for all known viruses of the family is available on a web-based database (Table 2). The dynamics of the proteolytic processes have been suggested to have regulatory implications for the sequential appearance and accumulation of intermediate and final products.

A typical potyvirus genome starts with a 5' non-coding region (NCR), with a variable length (less than 300 bp). This leader might act as a translational enhancer, and although the general mechanism is not fully understood, regulatory elements were found in this region. Studies in Plum pox virus (PPV), for instance, showed that (i) most of this region is dispensable for infectivity, although it contributes to viral competitiveness and pathogenesis, and (ii) translation takes place by a cap-independent leaky scanning mechanism. In the case of the 5'-leader of Tobacco etch virus (TEV), the existence of an internal initiation site has been suggested, with the presence of an RNA pseudoknot that provides cap-independent translation.

Another NCR variable in size (generally up to 200 bp) is located at the 3' end of the genome, just upstream of the poly-A tail. Putative RNA structures in this region confer pathogenic properties in Tobacco vein mottling virus (TVMV), and studies with TEV and Clover yellow vein virus (CIYVV) demonstrated the existence of *cis*-acting elements necessary for infectivity.

Properties and Functions of Gene Products Derived From the Polyprotein

The polyprotein encoded in the genome of most potyvirids comprises the following gene products (from N-terminus to C-terminus): P1, HCpro, P3 and P3N-PIPO (with PIPO in a different reading frame), 6K1, CI, 6K2, VPg, NIap_{ro}, NIb and CP. For bymoviruses, RNA-2 encodes two proteins P2-1 and P2-2, while RNA-1 encodes the remaining products starting with P3. Besides the information provided in Fig. 4, some characteristics of these products are given in this section.

The P1 protein is a serine protease, extremely variable in size (from 30 to 60 kDa), considered as an accessory factor for genome amplification. This region contains important sequence differences among potyvirids, and the whole cistron is even absent in macluraviruses and some unclassified potyvirids. In some ipomoviruses, such as Cucumber vein yellowing virus (CVYV), two P1-like serine proteinases are found in tandem, with the second one acting as a suppressor of gene silencing, likely compensating the absence of the HCpro cistron in its genome.

The helper component HCpro is a multifunctional protein named after its participation as a helper during aphid transmission. HCpro of tritimoviruses is also implicated in mite transmission. HCpro is smaller in macluraviruses, and sequences encoding a tandem of similar small HCpros are found in the genome of two potyvirids infecting areca palm, which has been proposed to become a novel genus tentatively named *Arepavirus*. A modular distribution of functions has been proposed for the HCpro of potyviruses: the N-terminus is essential for vector transmission, the central region is involved in suppression of RNA silencing, and the C-terminal part contains the papain-like cysteine protease activity. HCpro can interact with RNA, including double-stranded small interfering RNA molecules, a common feature of many other viral-derived RNA silencing suppressors. Structural studies with *Lettuce mosaic virus* (LMV) and TEV showed that HCpro form pairwise oligomeric structures. Relevant interactions of HCpro with several host factors have been described.

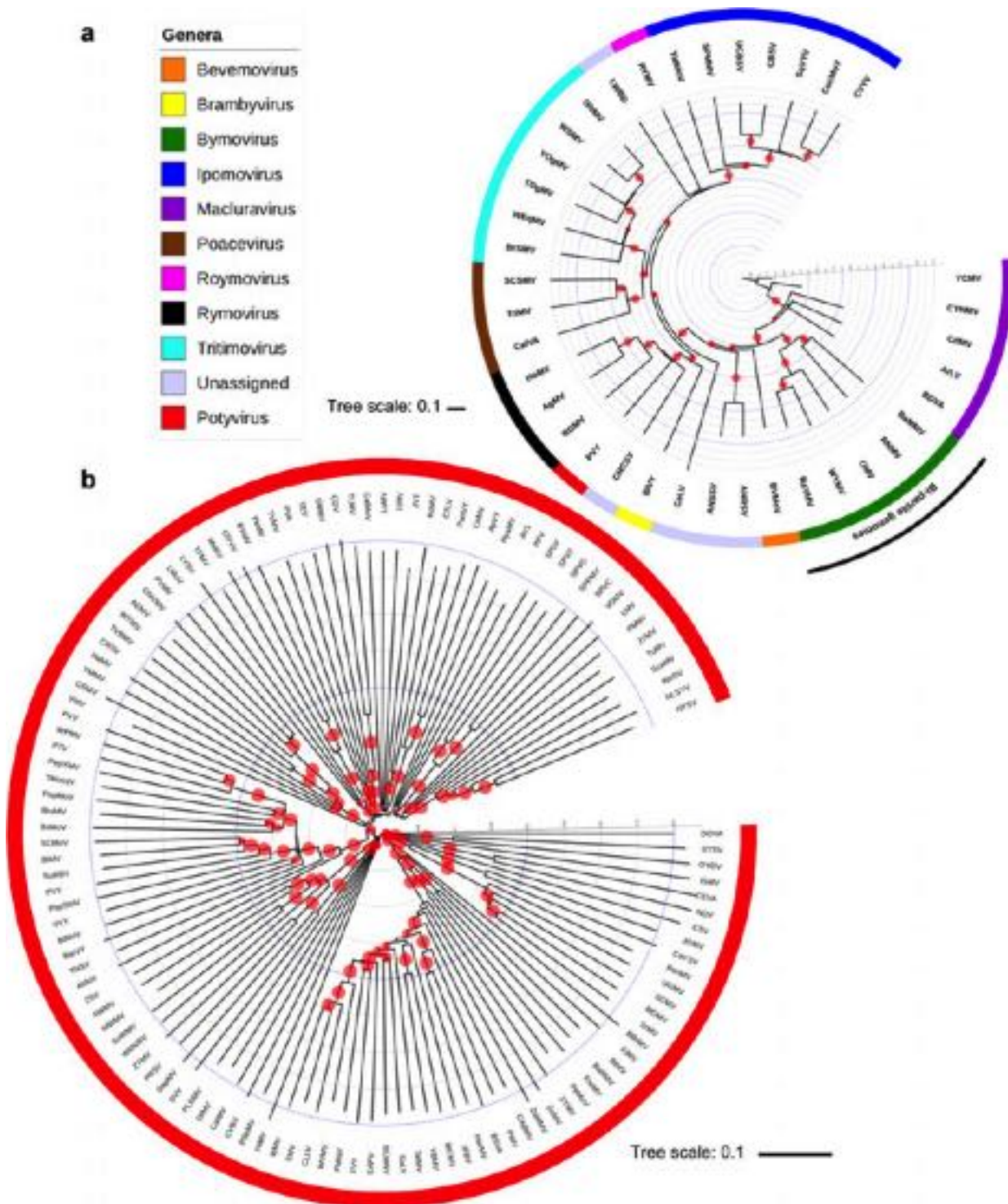


Fig. 3 Phylogenetic analysis derived from the comparison of full-length viral genome sequences of 38 members of the family *Potyviridae* belonging to different genera (a), and 122 members of the genus *Potyvirus* (b). Figures depict neighbour joining trees derived from Mega v6.06 after alignment by Muscle with 1000 bootstrap replicates, with branch lengths being proportional to genetic distance (substitutions per site). The bootstrap support (>0.7) for branches is indicated by proportional dots. For bymoviruses, the comparison was performed with a concatamer of RNA2 followed by RNA1. Viruses are identified by their official acronyms available at the ICTV webpage, and the sequences were obtained using the GenBank accession numbers of the corresponding type isolates, as found in the latest release (MSL34, from March 2019) of the Virus Metadata Repository (see Table 2). Grouping by genera and by number of genome segments are indicated as sectors in (a), where only the type member of potyviruses (PVY) is included, while the rest of selected members in the genus are shown in (b).

Table 2 Selection of resources available on the Internet (World Wide Web) about the family *Potyviridae*

Web Page Address (URL)	Contents and characteristics
https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/271/potyviridae	ICTV (International Committee on Taxonomy of Viruses). Taxonomy structure, list of species
https://talk.ictvonline.org/taxonomy/vmr/	Virus Metadata Resource (VMR) spreadsheet, with the latest release containing data on new virus species ratified by ICTV
https://viralzone.expasy.org/48?outline=all_by_species	Web-resource providing general molecular and epidemiological information, along with virion and genome figures, and access to UniProtKB/Swiss-Prot viral protein entries.
http://bio-mirror.im.ac.cn/mirrors/pvo/viderefs.htm	VIDE (Virus Identification Data Exchange) project. Nomenclature, host range, virion properties
http://www.dpvweb.net/potycleavage/index.html	Analysis of the polyprotein cleavage sites

In the case of bymoviruses, very limited information is available about the role of P2-1 and P2-2. It has been postulated that P2-2 might participate in vector transmission, since non-transmissible variants exhibited deletions in this region. Intriguingly, sequence alignments indicate that P2-1, a putative cysteine proteinase, is indeed more closely related to HCpro than P2-2.

With the exception of P1 and HCpro, potyvirus gene products are excised from the viral polyprotein by the action of the NIpro proteinase. The third gene product is P3, a protein of unknown functions presumably implicated in RNA replication. An intrinsic difficulty to establish the actual role of P3 lays on the presence of the overlapping P3N-PIPO gene, which is directly involved in cell-to-cell movement. P3 is followed by the 6K1 peptide. Remarkably, separation of 6K1 from P3 in TEV does not occur *in vitro* and, moreover, this excision seems to be not required for the viability of PPV. Immunology studies determined the presence of both P3 and P3 + 6K1 in infected cells, and the latter is perhaps the functional product. Proteins encoded in the P3-6K1 region have been shown to play roles in host range definition and pathogenicity of several potyvirids.

The CI protein is the largest potyvirus gene product. It forms very distinctive pinwheel-shaped cylindrical inclusions in the cytoplasm of infected cells, with high taxonomic value because they are unique to members of the *Potyviridae* family (Fig. 1(b)). CI exhibits RNA helicase activity, and is supposed to act during RNA replication. Along with P3N-PIPO, CI forms conical structures at the cell plasmodesmata to allow viral cell-to-cell movement. 6K2 is a trans-membrane protein that induces the membrane remodeling needed for viral RNA replication and anchors the replication complex to virus-induced membrane structures. Studies with PVA 6K2 revealed that this protein might be also involved in movement and symptom induction.

The whole NIa (nuclear inclusion a) protein has about 49 kDa, but it is subjected to an internal suboptimal cleavage that gives rise to VPg (21 kDa), the protein covalently attached to the 5' end of viral RNA, and a proteinase fragment (NIpro, 28 kDa). VPg can be uridylylated *in vitro* by the RNA replicase NIb, suggesting that it is used as a primer during RNA replication, as in the case of picornaviruses. The VPg from diverse potyvirids, such as the one from Turnip mosaic virus (TuMV), can also counteract antiviral RNA silencing, a function involving direct interaction with, and promoting the degradation of, components of the host silencing machinery. In PVA, VPg is translocated as a 'phloem protein' that specifically acts in companion cells to facilitate virus unloading. ATPase activity was found in the VPg of Pepper vein banding virus (PVBV), being functionally autoinhibited by its N-terminal region, and stimulated by NIpro. The main role of NIpro is to act in *cis* and *trans* as the protease responsible for most cleavages in the polyprotein, processing sites mainly defined by specific heptapeptide sequences. NIpro shares structural motifs with host serine proteases, but it has a cysteine residue at the catalytic site. NIpro has non-specific RNA binding activity, and was shown to exhibit non-specific double-stranded DNA degradation activity in PVBV. The complete NIa can form inclusion bodies in the nucleus, where it is translocated responding to a bipartite nuclear localization signal. The crystal structure of TEV NIpro has been solved, providing clues on its mode of action.

NIb (nuclear inclusion b) is the RdRp responsible for viral replication. The recruitment of this protein towards the replication complex is postulated to occur via interaction with NIa. All NIb functions essential for RNA amplification can be provided in *trans*, as knockout TEV mutants lacking NIb infect transgenic plants expressing this protein. NIb is the second component of nuclear inclusions, being directed to the nucleus by specific signals in its sequence. Recent results with TuMV have shown that sumoylation of NIb occurs in the nucleus, and this modification retargets it to the cytoplasm where viral replication takes place. Moreover, sumoylation of NIb contributes to the suppression of the host NPR1-mediated immune response. Phosphorylation of TuMV NIb has been described, although it is still not known the significance of this post-translational modification.

The last product of the viral ORF is CP, a protein with many roles apart from coating and protecting the RNA genome. It is implicated in aphid transmission through interaction with HCpro, and in both cell-to-cell and long-distance movement. The importance of maintaining the net charge of the CP N-terminus for infectivity has been shown in TMV and Zucchini yellow mosaic virus (ZYMV).

Expression of Overlapping Gene Products

The presence of PIPO within the P3 cistron was predicted by bioinformatics analyses, including alignments of potyvirus genomes and synonymous sequence variability. This ORF is preceded by a conserved GA6 nucleotide sequence, a motif not present in other viral regions except in the case of a subgroup of sweet potato-infecting potyviruses, which harbor an additional GA6 within the P1 region that precedes a larger ORF named PISPO. Interestingly, the predicted P1N-PISPO gene product was found to be a suppressor of gene

silencing. Experimental evidence deriving from high-throughput sequencing of different potyvirus progenies support the idea that the changes in the reading frame required to produce these overlapping products occur during positive or negative strand (it is not yet defined) RNA synthesis by polymerase slippage. Interestingly, in another set of RNA species, an apparently equivalent mechanism causes the deletion of one nucleotide at the same GA6 motif in Clover yellow vein virus (CIYVV), with the consequent change in the reading frame. This deletion introduces a premature stop codon, and therefore a truncated P3, termed P3N-ALT, is produced. As P3N-PIPO, CIYVV P3N-ALT seems to participate in cell-to-cell movement.

Replication and Propagation

Potyvirus replicates in the cytoplasm of infected cells, as schematically shown in Fig. 5.

After entering the cell, the viral genomic RNA must be first translated in the cytoplasm. In general, the early events during infection are still poorly understood. Disassembly of particles might occur co-translationally, and after translation the replication complex must be formed with participation of several viral products and probably host factors. This complex uses the genomic RNA (+ sense) as template to generate a complementary chain (- sense) through a dsRNA intermediate, and proceed with the asymmetric synthesis of numerous molecules of genomic RNA. Viral-derived siRNAs have been shown to accumulate during potyvirus infections, revealing the action of host RNA silencing-mediated antiviral defenses. Therefore, potyviruses need to overcome this barrier, and for that they express RNA silencing suppressors. In the case of potyviruses, this role is mainly exerted by HCpro, although this function might be shared or displaced to other products in different genera. Studies performed with *Pea seed-borne mosaic virus* (PSbMV) served to identify a translation shut-off affecting many host proteins during infection. In the case of PVA, the shift for translation to replication appears to be controlled by phosphorylation-dependent RNA binding of non-assembled viral CP. The potyvirus expression strategy implies production of equimolar amounts of all gene products, giving rise to large excesses of some proteins, which might remain soluble, be degraded or secreted, or end up in inclusion bodies. It is intriguing that, whereas CI pinwheels are always formed, HCpro amorphous or NIa/NIb crystalline inclusions are present in some, but not all, viral infections.

Potyvirus RNA replication takes place in membranous vesicles derived from the endoplasmic reticulum that target chloroplasts and aggregate into large perinuclear globular structures. NIb forms the core of the replication machinery, with involvement of CI, VPg, 6K2, and NIaPro. Other factors, viral- and host-derived, might also be involved. For instance, the VPg of TuMV can interact with initiation factors eIF4E and eIF(iso)4E, and also with the Poly-A binding protein. The fact that several resistance genes, identified in host plants as potyvirus-specific, are related to translation initiation factors points to the importance of these interactions for the infection cycle.

Another critical process for a successful virus infection is the genomic RNA transport from the initially infected cell towards adjacent cells and distant tissues. In principle, the virus can invade the whole plant, with the probable exception of meristems, although a few seed-transmissible potyviruses might be capable of invading these particular tissues. The CP, but probably not virions, is essential for virus movement. RNA replication and virus movement appear to be closely linked. It has been proposed that replication vesicles are able to move through the cytoplasm along microfilaments to reach plasmodesmata and cross them in order to invade the neighboring uninfected cell, and that these vesicles are even able to traffic through the vascular system. In the process of local and systemic spread, host defenses must be confronted by the virus in a race to avoid its blockage by the action of both general and specific long-distance defensive signals.

The final step of the infective cycle is the virus spread to new plants, which rely on the acquisition of virions by the corresponding vector organisms and further transmission. Two viral proteins, CP and HCpro, are involved in the non-persistent aphid transmission of potyviruses. Available data supports a "bridge" hypothesis, in which HCpro serves as a reversible linker to retain virus particles in aphid mouthparts. A conserved PTK motif in the central portion of HCpro could participate in binding to the DAG motif at the N-terminus of CP, while a KTC domain at the N-terminal region of HCpro would be involved in binding to unknown structures on the aphid stylet. HCpro is also implicated in the semipersistent transmission of tritoviruses by eriophid mites. Less information is available about the transmission processes of macluraviruses by aphids, bymoviruses by plasmodiophorids, rymoviruses and poaceviruses by eriophid mites, and ipomoviruses by whiteflies.

Multifunctionality is observed in most potyviral proteins, illustrated for instance by HCpro. This capacity to participate in multiple processes suggests that potyvirus infection is not a consecutive succession of independent events, but a tightly regulated network of complex interactions, still to be elucidated, between viral and host factors.

Pathogenicity

The information about how potyviruses cause diseases in their host plants is very limited. Several studies have served to identify sequences and/or products in the genome of potyviruses directly implicated in the production of symptoms. As mentioned above, symptom determinants are even present in the 5'- and 3'-NCR. In addition, HCpro, P3, CI, 6K2 and VPg of different viruses have been described as determinants of pathogenicity, although single products were not always responsible for the different pathogenic responses.

As already mentioned, the P1/HCpro region was found to be responsible for the synergistic effect in mixed infections of potyviruses with unrelated viruses, perhaps reflecting the capacity of HCpro to counteract the RNA silencing-mediated defense. Interestingly, HCpro also interferes with the action of microRNAs, a particular type of regulatory host-derived small RNAs,

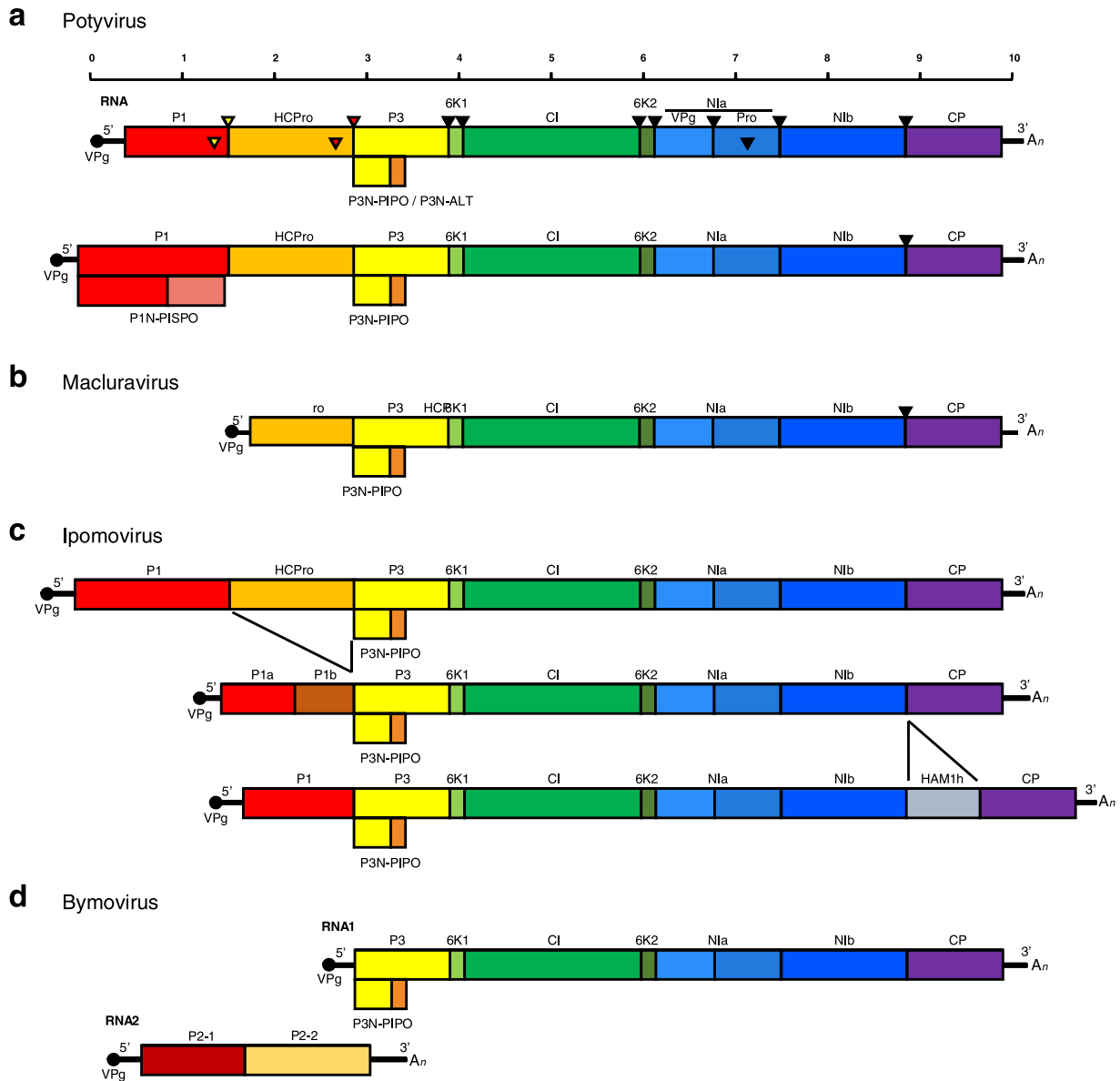


Fig. 4 Genome maps of representative viruses in the family *Potyviridae*. A monopartite virus characteristic of a potyvirus such as the type member of the genus PVY is shown in (a), together with SPFMV as representative of the sweet-potato infecting subgroup. Maps of representatives of other genera are shown for the macluravirus CdMV (b), three different ipomoviruses SPMMV, CVYV and CBSV (c), and the two genomic RNAs of the bipartite bymovirus BYMV (d). ssRNA genomes are shown as solid horizontal lines with VPg represented by solid circles at 5' end, and poly-A tails at 3' end. Viral ORFs are depicted as boxes divided in the different gene products, with their names indicated. The partially out-of-frame products deriving from polymerase slippage mechanisms are shown below. The protease-specific cleavages sites are indicated by arrows above the first map, with the matching symbol shown in the gene products responsible of the proteolytic process. Additional genome structures of other viruses (not shown) include the potyvirus EuRSV, containing a HAM1-like gene between Nib and CP coding sequences, and the still unclassified viruses of the proposed new genus *Arepavirus*, with presence of two HCPro-related cistrons, or the proposed new genus *Celavirus*, with a signal peptide preceding P1 and a very long 3' NCR.

providing a molecular explanation for some virus-induced symptoms. However, although it is tempting to regard interference with small RNAs metabolism in susceptible plants as a major element of pathogenicity, many other processes may also be affected. For example, specific features of the soybean *Rsv1*, a Soybean mosaic virus (SMV) resistance gene, appear to be responsible for the induction of either systemic mosaic or lethal systemic hypersensitive response in infected plants. The fact that the HCpro protein of LMV targets, and affects the functions of, the 26S proteasome might exemplify another way by which these viruses can cause disease symptoms. Despite the fact that in most cases the mechanisms of pathogenicity remain uncertain, it is reasonable to conclude that the final macroscopic effects of potyvirids are caused by a combination of additive interferences with several host functions and processes.

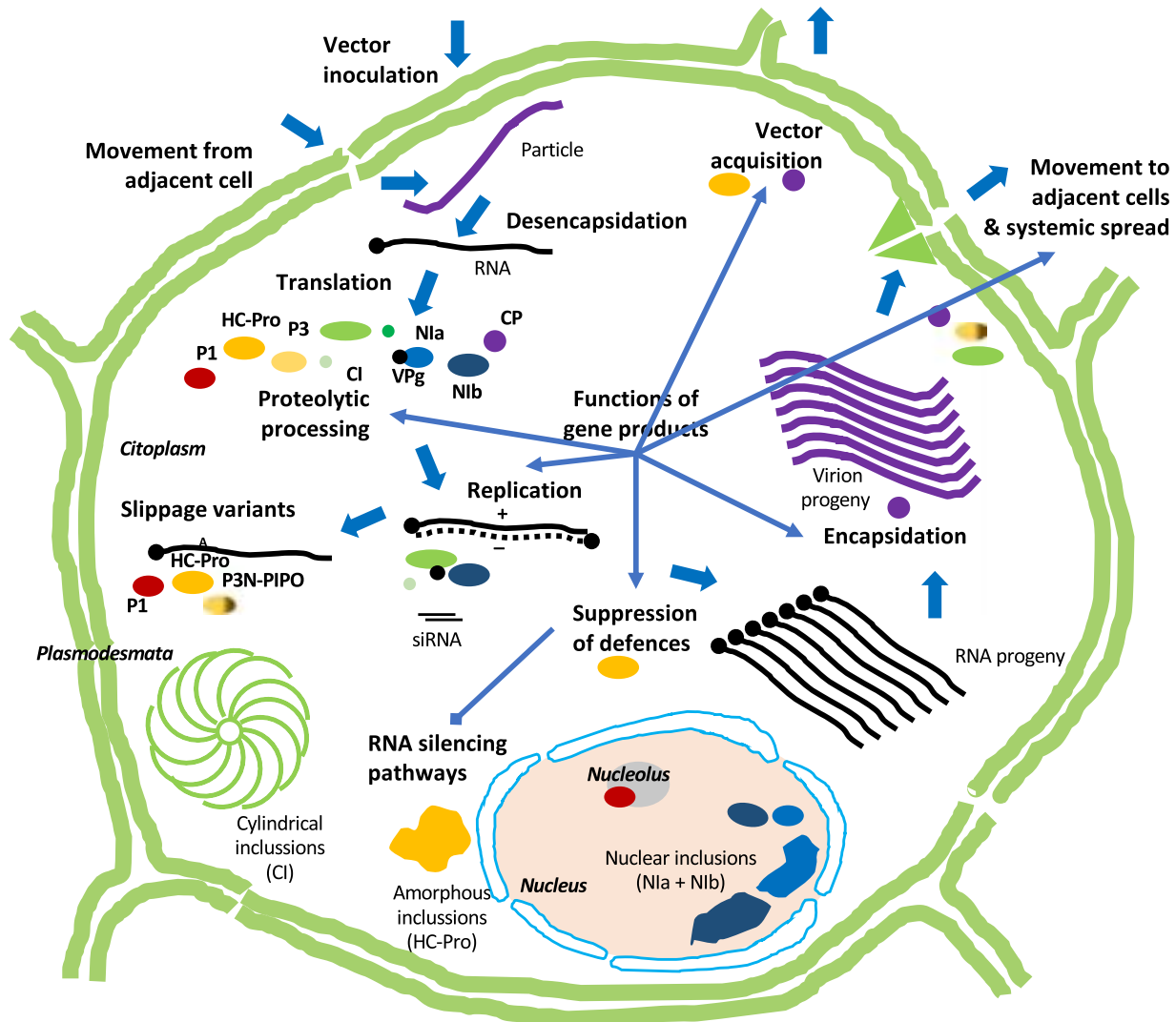


Fig. 5 Schematic representation of events during potyvirus infection of a plant cell. Colors of gene products match those displayed in Fig. 4. For simplicity, most organelles have been omitted, as well as the participation of host factors during the cycle. The accumulation of viral proteins as inclusions in different compartments is indicated. The cycle begins (left up) when the virus enters the cell, either from an adjacent infected cell or by inoculation from a vector organism. Then, the genomic RNA undergoes des-encapsidation, translation and processing to originate mature products. The replication complex involves participation of at least Nib, CI, VPg, 6K2, and Nla, together with host-derived factors, and uses the genomic RNA (+ sense) to generate a complementary chain (- sense), serving as template for the synthesis of numerous genomic RNAs. Also, during polymerization, variants with slippage at GA6 motifs could be generated and translated to yield partially out-of-frame products. dsRNA replicative intermediates or regions with secondary structure produce viral siRNAs by action of the host silencing machinery, which is counterattacked by suppressors of RNA silencing, such as HCpro or other viral products. After replication, the RNA progeny can move to adjacent cells through plasmodesmata, with involvement of P3N-PIPO, CI, VPg and CP, or it can be encapsidated and acquired to be transmitted again by a vector organism like aphids, in a process requiring HC-Pro. See text for further details.

Epidemiology and Control

Strategies currently applicable for the control of potyvirids are diverse, from cultural practices to the use of genetic resistance. Insecticide or other treatments against vectors are frequently considered unsatisfactory and have limited use because of the transmission type.

Severity of outbreaks is commonly related to abundance of initial foci of infection, dynamics of vector populations and other factors, such as the presence of weeds acting as reservoirs of viruses. The human intervention is responsible in many cases for the introduction of emerging diseases into new territories, while vector organisms are mainly involved in propagation within particular regions. For a few potyviruses, seed transmission is also an important means of dissemination. A typical example of well documented spread of a potyvirus over a territory and over time is the progressive emergence of Sharka disease, caused by PPV, in European countries during the XX century, and its recent diffusion to other continents to become nowadays a global pandemic.

The exploitation of pathogen-derived resistance and RNA silencing is yielding promising results for virus control. Transgenic plants incorporating viral sequences were found to be resistant to several potyvirids. An example of success is provided by the engineered papaya varieties, which are resistant to *Papaya ringspot virus* due to the expression of the cognate viral CP as a transgene. Nowadays, RNA silencing is being further exploited for resistance by designing specific transgenes with hairpin structures as sources of viral-derived siRNAs, or even by direct application of dsRNAs. Other approaches to achieve resistance include the expression of ribozymes, plantibodies or proteinase inhibitors. Cross-protection has also been related to RNA silencing mechanisms, although other processes could be operating, as suggested by the involvement of specific proteins, such as NIap and CP, in superinfection exclusion. Stability, durability and biosafety issues of the different strategies are under constant evaluation.

The identification of potyvirus-specific resistance genes in host plant is also leading to important advances. These studies indicate that canonical resistance genes might be operating, for instance TIR-NBS-LRR class in PVY, or NBS-LRR class in SMV. Other examples of dominant resistance genes acting against different potyviruses are the three *Restricted TEV Movement (RTM)* genes, which code for a jacalin-type protein, a small heat shock protein and a MATH domain-containing protein. As mentioned, mutations in eukaryotic initiation factors are frequently found as responsible for recessive resistance against potyvirids, but particular features of other host factors, as the products of *PROTEIN DISULFIDE ISOMERASE LIKE 5-1* of barley or two MATH domain-containing genes of apricot, has been also found to be associated with natural recessive resistance against potyvirids. Therefore, the generation of variability in candidate genes, using TILLING or equivalent platforms, must be a promising strategy for the generation of resistant plants.

Diagnosis

Serological tools have been the preferred diagnostic system for potyvirids. While the non-structural viral proteins have limited use in serological diagnosis, the virus particles are in general strongly immunogenic. However, serological relationships among potyvirids are complex, with unexpected and inconsistent cross-reactivities hindering their application in taxonomy and specific virus identification. The cause of this phenomenon is the presence of common epitopes in the conserved internal CP core, while specific epitopes map in the variable N-terminus, a surface-exposed region prone to degradation. New species-specific antibodies targeting the immunodominant N-terminal region of CP provide excellent tools for the serological detection of many potyvirids.

RT-PCR is another way to deal with diagnosis. Degenerated primers matching conserved genomic regions might allow the amplification of virtually any potyvirus. For instance, primers flanking the variable CP N-terminal region have been developed and successfully used for both detection and identification of highly divergent potyvirids. Interestingly, the combination of serological capture of virions followed by highly sensitive RT-PCR has resulted in robust detection systems. Together, these molecular tools have been extensively used to identify new viruses infecting new hosts. The application of new technologies and modern molecular methodologies to diagnostics is being continually updated, which results in more sensitive and more specific approaches.

Concluding Remarks

The abundant information on potyvirids have reached a level that allows to develop biotechnological applications, including the production of foreign proteins in plants using virus-based expression vectors, and also the exploration of viruses as direct sources of genetic elements and functionally active products, such as the NIap proteases used to remove affinity tags from fusion proteins. Another example is the use of HCpro to block the host RNA silencing machinery with the consequent boost in the expression of products in plants.

Several potyvirids have been engineered and tested as expression vectors. Small peptides can be expressed as CP-fusion products, allowing the protein chimeras to be used as antigen presentation systems for immunization or diagnosis. Complete foreign genes have been expressed in vectors based on a quite large list of potyvirids, which covers an important range of potential hosts. The seminal work with TEV pointed to the P1/HCpro junction as an adequate insertion site for foreign sequences, but additional sites were exploited later, such as the NIb/CP junction or the P1 region. Moreover, more recent reports demonstrated the huge versatility of potyvirids for the expression of foreign genes, as they can even express different products simultaneously from independent insertion sites.

To summarize, it can be concluded that the family *Potyriviridae* is a fascinating group of viruses. Keeping key genomic features, potyvirids have diverged in a multitude of lineages, which we have yet to finish to discover, and one of them, the genus *Potyvirus*, can be considered one of the most successful in the world of viruses. The currently available knowledge about potyvirus replication, movement and transmission is allowing the development of new control strategies aiming to interfere with key-steps in the virus life cycle. Structure resolution of viral components might also serve to explore new resistance strategies. Diagnostic tools are continually under improvement, and future developments will certainly supply specific and sensitive means of virus identification. Our current insight about the selective participation of different host factors in resistance and pathogenesis is fueling works aiming to understand the complex virus-plant crosstalk that occurs during the infection. Besides, peculiarities of certain viruses are being unraveled, and these findings will help to discover new generic features and specificities of members of the *Potyriviridae* family. All in all, the study of potyvirus infections at the molecular level has been extraordinarily successful over the last years, and is likely to continue providing relevant data on the viruses, as well as in terms of understanding many plant processes.

Further Reading

- Adams, M.J., Antoniw, J.F., Fauquet, C.M., 2005. Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Archives of Virology* 150, 459–479.
- Cheng, X., Xiong, R., Li, Y., *et al.*, 2017. Sumoylation of *Turnip mosaic virus* RNA polymerase promotes viral infection by counteracting the host NPR1-mediated immune response. *Plant Cell* 29, 508–525.
- Elena, S.F., Rodrigo, G., 2012. Towards an integrated molecular model of plant-virus interactions. *Current Opinion in Virology* 2, 719–724.
- Ivanov, K.I., Eskelin, K., Lohmus, A., Mäkinen, K., 2014. Molecular and cellular mechanisms underlying potyvirus infection. *Journal of General Virology* 95, 1415–1429.
- Jiang, J., Laliberté, J.F., 2011. The genome-linked protein VPg of plant viruses—a protein with many partners. *Current Opinion in Virology* 1, 347–354.
- Lauber, C., Seifert, M., Bartenschlager, R., Seitz, S., 2019. Discovery of highly divergent lineages of plant-associated astro-like viruses sheds light on the emergence of potyviruses. *Virus Research* 260, 38–48.
- Movahed, N., Patarroyo, C., Sun, J., *et al.*, 2017. Cylindrical Inclusion protein of Turnip mosaic virus serves as a docking point for the intercellular movement of viral replication vesicles. *Plant Physiology* 175, 1732–1744.
- Olsper, A., Chung, B.Y.-W., Atkins, J.F., Carr, J.P., Firth, A.E., 2015. Transcriptional slippage in the positive-sense RNA virus family *Potyviridae*. *EMBO Reports* 16, 995–1004.
- Revers, F., García, J.A., 2015. Molecular biology of potyviruses. *Advances in Virus Research* 92, 101–199.
- Rodamilans, B., Shan, H., Pasin, F., García, J.A., 2018. Plant viral proteases: Beyond the role of peptide cutters. *Frontiers in Plant Science* 9, 666.
- Rodamilans, B., Valli, A., Mingot, A., *et al.*, 2015. RNA polymerase slippage as a mechanism for the production of frameshift gene products in plant viruses of the *Potyviridae* family. *Journal of Virology* 89, 6965–6967.
- Sorel, M., García, J.A., German-Retana, S., 2014. The *Potyviridae* cylindrical inclusion helicase: A key multipartner and multifunctional protein. *Molecular Plant-Microbe Interactions* 27, 215–226.
- Truniger, V., Aranda, M.A., 2009. Recessive resistance to plant viruses. *Advances in Virus Research* 75, 119–159.
- Valli, A.A., Gallo, A., Rodamilans, B., López-Moya, J.J., García, J.A., 2018. The HCPro from the *Potyviridae* family: An enviable multitasking Helper Component that every virus would like to have. *Molecular Plant Pathology* 19, 744–763.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9, e02329. 18.
- Wylie, S.J., Adams, M., Chalam, C., *et al.*, 2017. ICTV Virus taxonomy profile: *Potyviridae*. *Journal of General Virology* 98, 352–354.
- Zamora, M., Méndez-López, E., Aguirrezabala, X., *et al.*, 2017. Potyvirus virion structure shows conserved protein fold and RNA binding site in ssRNA viruses. *Science Advances* 3. (eaao2182).

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/271/potyviridae
ICTV 9th Report (2011).
- <http://www.dpvweb.net/potycleavage/index.html>
Potyviridae cleavage sites
Descriptions of Plant Viruses.
- <http://bio-mirror.im.ac.cn/mirrors/pvo/vide/refs.htm>
VIDE (Virus Identification Data Exchange) project. Nomenclature, host range, virion properties.
- https://viralzone.expasy.org/48?outline=all_by_species
ViralZone.
- <https://talk.ictvonline.org/taxonomy/vmr/>
VMR - International Committee on Taxonomy of Viruses (ICTV).

Quinviruses (*Betaflexiviridae*)

Ki H Ryu and Eun G Song, Seoul Women's University, Seoul, South Korea

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
AlkB Alpha-ketoglutarate-dependent dioxygenase
CP Coat protein or capsid protein
CRP Cysteine-rich protein
cSSR Compound simple sequence repeat
Hel Helicase
kb Kilobase
kDa Kilo dalton
LAMP Loop mediated amplification
MP Movement protein
NLS Nuclear localization signal
nt Nucleotide(s)
NTR Non-translated region

ORF Open Reading Frame
PCR Polymerase chain reaction
Pol RNA-dependent RNA polymerase
Poly(A) Polyadenylated
P-Pro Papain-like protease
RdRp RNA-dependent RNA polymerase
RT-PCR Reverse transcription polymerase chain reaction
S_{20,w} Corrected sedimentation coefficient; extrapolated in water at 20°C and infinitely diluted
SSR Simple sequence repeats
ssRNA Single-stranded RNA
TGB Triple gene block
ZF Zinc finger motif

Glossary

AlkB protein The protein found in bacteria, fungi, and mammalian genomes is iron (II)- and 2-oxoglutarate-dependent dioxygenases that remove alkyl adducts from oxidative dealkylation DNA repair reverse methylation damage, such as 1-methyl-adenine and 3-methyl-cytosine. Viral AlkB proteins can repair alkylated bases in DNA or RNA.

Cysteine-rich protein The protein related to viral pathogenicity determinants. The CRPs can be divided into two groups according to their domain structures. The CRPs with a conserved CGxxH motif (C, cysteine; G, glycine;

H, histidine; x, any aa residue). The CRPs with an arginine-rich nuclear localization signal (NLS) in the central part of the protein. Carlavirus CRPs contain a zinc finger motif (ZF) located adjacent to the NLS.

RNA silencing suppressor Plant viruses are encoding proteins to counteract the plant antiviral system known as RNA silencing.

Triple gene block The protein consists of a strikingly similar element of three partially overlapping ORFs and is included in their genomes. The TGB proteins are related to cell-to-cell movement and potentially influence on host antiviral defenses.

Introduction

The subfamily *Quinvirinae*, family *Betaflexiviridae*, contains three genera (*Carlavirus*, *Foveavirus*, and *Robigovirus*) and three unassigned species. The complete genome sequences of most of the members of the subfamily *Quinvirinae* have been determined recently. The members of the genus *Carlavirus* contain 58 species including *Carnation latent virus* (CLV) as the type species for the genus *Carlavirus*. The members of the genus *Foveavirus* contain 8 species including *Apple stem pitting virus* (APSV) as the type species for the genus *Foveavirus*. The members of the genus *Robigovirus* contain 5 species including *Cherry necrotic rusty mottle virus* (CNRMV) as the type species for the genus *Robigovirus*. The genus *Robigovirus* has been recently established as a new member of the subfamily *Quinvirinae* by the International Committee on Taxonomy of Viruses (ICTV). The members of the subfamily *Quinvirinae* can be divided into two or three main phylogenetic groups based on their genome sequences. The two genera, *Foveavirus* and *Robigovirus*, and unassigned species can be clustered in one branch. The phylogenetic relationships of the subfamily *Quinvirinae* are not associated with their natural host. The virions of the subfamily *Quinvirinae* have non-enveloped and flexible filaments of 640–1000 nm in length and 12–15 nm in diameter. Their genomes are composed of a positive-sense single-stranded RNA (ssRNA) with 5' terminal cap structure and a 3' terminal poly(A) tail, and contain RNA-dependent RNA polymerase (RdRp) for viral replication, triple gene block (TGBp1, 2, and 3) for viral movement (MP), and coat protein (CP). The RdRp contains five conserved protein domains as follows: an AlkB domain; a papain-like protease (p-Pro) domain; an RNA helicase (HEL) domain; a polymerase (POL) domain. The sizes of the genomes have 7.4–9.3 kbp. The genomes of the carlaviruses contain mostly a cysteine-rich protein (CRP) including a nuclear localization signal (NLS) and a zinc finger motif (ZF). Most viruses have a relatively restricted host range, which may exhibit mosaic, mottle, ringspot, necrosis, or pitting symptoms or may be symptomless. Most viruses are spread by mechanical transmission and some viruses can be transmitted by insect vectors, such as aphid and whitefly. Some viruses are often transmitted by seed and grafting. Virus diseases caused by some viruses can seriously damage major crops and the viruses require suitable disease managements.

Taxonomy and Phylogeny

Three genera, *Carlavirus*, *Foveavirus*, and *Robigovirus*, are members of the subfamily *Quinvirinae* within family *Betaflexiviridae*. The genus *Carlavirus* contains 53 definitive species, of which 41 are determined completed genome sequences (Table 1). Its name is derived from *Carnation latent virus* (CLV, the type species of *Carlavirus*). The CLV genome has not been completely sequenced. The genus *Foveavirus* contains 8 definitive species, these viruses have complete genome sequences (see Table 1). Its name is derived from Latin *fovea* for “pit” or “hole”, which is a type of symptom induced by the type species (Apple stem pitting virus, ASPV). The genus *Robigovirus*, a newly created genus within the subfamily *Quinvirinae*, contains 5 definitive species, these viruses have complete genome sequences (Table 1). Its name is derived from Latin *robigo* for “rust”, which is a type of symptoms induced by the type species (Cherry necrotic rusty mottle virus, CNRMV). Additionally, the subfamily *Quinvirinae* contains three unassigned species *Banana mild mosaic virus* (BanMMV), *Banana virus X* (BanVX), and *Sugarcane striate mosaic-associated virus* (SCSMaV). Phylogenetic tree of members of the subfamily *Quinvirinae* has revealed a clustering into two major groups of their RNA-dependent RNA polymerase (RdRp) aa sequences (Fig. 1(A)). The major group I consists of members of the genus *Carlavirus* whereas the major group II can be divided into three distinct subgroups (carlavirus, foveavirus, and robigovirus groups). The carlavirus group within the major group II contains Sweet potato chlorotic fleck virus (SPCFV), Melon yellowing-associated virus (MYaV), and Nerine latent virus (NeLV). The foveavirus group has divided into three branches and ASPV, the type species of *Foveavirus*, is closely grouped with Apricot latent virus (ApLV). The two unassigned species, BanMMV and SCSMaV, are supported as members of the robigovirus group within the major group II. The robigovirus group is divided into two branches and CNRMV, the type species of *Robigovirus*, is more closely related to Cherry twisted leaf associated virus (CTLAV). In the phylogenetic tree of the coat protein (CP), *Foveavirus*, *Robigovirus*, and unassigned species are clustered in one branch (Fig. 1(B)). The foveavirus group could be divided into three branches and the ASPV is more closely grouped with Apple green crinkle associated virus (AGCaV) than the ApLV. The robigovirus group is divided into two branches and the CNRMV is more closely grouped to Cherry rusty mottle associated virus (CRMAV) than the CTLAV. Of three sambucus-infecting viruses in the phylogenetic trees of the RdRp and CP, *Sambucus virus C* and *Sambucus virus E* that are grouped together are more closely related than *Sambucus virus D*. Four cherry-infecting robigoviruses in the phylogenetic trees, Cherry green ring mottle virus (CGRMV), CNRMV, CTLAV, and CRMAV, are clustered in one branch. The aa sequence identities of their RdRp and CP proteins of the subfamily *Quinvirinae* are less than 67.6% and 85.5%.

The phylogenetic trees of the TGBp1, TGBp2, and TGBp3 of the subfamily *Quinvirinae* could be divided into three major groups (Fig. 2). The two genera, *Foveavirus* and *Robigovirus*, and unassigned species in the phylogenetic tree of the TGBp1 are clustered in one branch, similar to those of the RdRp and CP. The two genera in the TGBp2 and TGBp3 are clustered in distinct branches. In the tree of the TGBp3, the members of the genus *Foveavirus* are clustered in distinct branches. Two foveaviruses, Grapevine rupestris stem pitting-associated virus (GRSPaV) and Grapevine virus T (GVT), clustered together are more closely related to three carlaviruses, Potato latent virus (PotLV), Hippeastrum latent virus (HLV), and Cowpea mild mottle virus (CPMMV), than to other foveaviruses. The cheery-infecting viruses except for CGRMV, which is missing the TGB aa sequences, are clustered in one branch. The aa sequence identities of their TGBp1, TGBp2, and TGBp3 proteins of the subfamily *Quinvirinae* are less than 72.2%, 79.0%, and 63.9%, respectively. Phylogenetic tree based on a cysteine-rich protein (CRP) aa sequences of members of the genus *Carlavirus* is revealed that divided into five groups. Of three sambucus-infecting viruses, SVC and SVE are grouped together, whereas SVD is grouped with Pea streak virus (PSV) in a distinct branch. The aa sequence identities of their CRP proteins is less than 77.1%.

Virion Structure

Virions are non-enveloped and flexible filaments of 640–1000 nm in length and 12–15 nm in diameter. Carlavirus and foveavirus particles are 610–700 nm and 800–1000 nm in length, respectively. Carlaviruses have helical symmetry with a pitch of 3.4–3.6 nm. Carlaviruses virion Mr is about 60×10^6 with 5.4%–6% nucleic acid content. Carlaviruses have the sedimentation coefficient ($S_{20, w}$) of 147–176S, the buoyant density in CsCl of 1.3 g cm^{-3} . Potato virus M (PVM), Shallot latent virus (SLV), and Daphne virus S (DVS) have the sedimentation coefficient ($S_{20, w}$) of 157S, 147.5S, and 174 S, respectively. A carlavirus isolated from elderberry (*Sambucus* spp.) has the sedimentation coefficient ($S_{20, w}$) of 155S and molecular weight of CP subunits of 31 kDa. The virus is an average normal length of 678 nm in length and 12 nm in diameter. The virus particles can be observed in the cytoplasm of their infected plant. The virus particles of *Helloborus net necrosis virus* (HeNNV) is approximately 800 nm in length and 17 nm in diameter. The virions of ASPV, Grapevine rupestris stem pitting-associated virus (GRSPaV), and CGRMV apparently do not contain lipids or carbohydrates.

Genome Organization

The genome is a positive-sense single-stranded RNA (ssRNA) with 5' terminal cap structure and a 3' terminal poly(A) tail. The genome is 7.4–9.3 kb in size and contains five or more ORFs. ORF1 encodes a large polypeptide that is RdRp including domains of methyl-transferase, helicase, and polymerase. ORFs 2, 3, and 4 constitute the TGB that facilitate virus movement. ORF5 encodes the CP and ORF6 encodes a cysteine-rich protein (CRP). The genome structures of the distinguished genera in the subfamily *Quinvirinae* is showed in Fig. 3. The genome of PVM, a member of the genus *Carlavirus*, consists of 8533 nt and contains six ORFs

Table 1 Virus species of three genera (*Carlavirus*, *Foveavirus*, and *Robigovirus*) of the subfamily *Quinivirinae*, family *Betaflexiviridae*. Type species in the genera are written in bold. Complete sequences of the genomes are indicated

Genus/Virus name	FYI	Particle size (nm)	Natural Hosts	Symptoms	Transmission	Distribution	Access #
Carlavirus							
<i>Aconitum latent virus</i>	AcLV	640	<i>Aconitum napellus</i> / <i>Delphinium</i> sp.	Mosaic/Spots on petals/Symptomless/ Mottle	MT Exp Hosts/Aphid-(<i>A. napellus</i>)	Japan/Israel	NC002795
<i>American hop latent virus</i>	AHLV		<i>Humulus lupulus</i>	Symptomless	Aphid (<i>Phorodon humuli</i>)	North America/Europe	NC017859
<i>Atractylodes mottle virus</i>	AtrMoV		<i>Atractylodes macrocephala</i>	Silver mottle		South Korea	NC038966
<i>Blueberry scorch virus</i>	BIScV	690 × 14	<i>Vaccinium corymbosum</i>	Blossom blight/Necrosis/Chlorosis	MT	North America/ Europe	NC003499
<i>Butterbur mosaic virus</i>	ButMV	670 × 13	<i>Petasites japonicus</i>	Necrotic ringspot	MT (<i>C. quinoa</i>)/Aphid	Japan	NC013527
<i>Cactus virus 2</i>	CV – 2	650 × 12	<i>Echinopsis</i> sp./ <i>Mammillaria elongata</i>			Ukraine	ND
<i>Caper latent virus</i>	CapLV	662	<i>Capparis spinosa</i>	Symptomless	MT Exp Hosts	Italy	NC043080
<i>Carnation latent virus</i>	CLV	650 × 14	<i>Dianthus caryophyllus</i>	Chlorotic spots/Mosaic/Mottle	MT Exp Hosts/Aphid (<i>Myzus persicae</i>)	Europe	NC038865
<i>Chrysanthemum virus B</i>	CVB	685 × 12	<i>Chrysanthemum</i> spp.	Mosaic/Vein banding/Mottle/Vein chlorosis	Aphid (<i>M. persicae</i>)	Worldwide	NC009087
<i>Cole latent virus</i>	CoLV	650	<i>Brassica oleracea</i>	Symptomless	MT/Aphid (<i>M. persicae</i>)	Brazil	NC038322
<i>Coleus vein necrosis virus</i>	CVNV	650	<i>Verena</i> x <i>hybrida</i>	Mottle/Necrosis	MT Exp Hosts	USA	NC009764
<i>Cowpea mild mottle virus</i>	CPMV		<i>Arachis hypogaea</i> / <i>Glycine max</i> / <i>Phaseolus vulgaris</i> / <i>Solanum lycopersium</i>	Mottle/Leaf malformation/Vein chlorosis/ Chlorotic ringspot and blotch/Leaf rolling	Whitefly (<i>Bemisia tabaci</i>) MT Exp Hosts	Worldwide	NC014730
<i>Cucumber vein-clearing virus</i>	CuVCV		<i>Citrullus lanatus</i>	Severe mosaic/Mottle/Deformation	MT/Whitefly (<i>B. tabaci</i>)	Tanzania	NC038322
<i>Daphne virus S</i>	DVS		<i>Daphne</i> spp.	Chlorotic mosaic/Distortion	Aphid (<i>M. persicae</i>)	New Zealand/South Korea	NC008020
<i>Gaillardia latent virus</i>	GallLV		<i>Gaillardia aristata</i>			Germany	NC023892
<i>Garlic common latent virus</i>	GarCLV		<i>Allium sativum</i>	Chlorotic spots/Yellow stripes/Symptomless	Aphid/MT (<i>C. murale</i>)	Worldwide	NC016440
<i>Helenium virus S</i>	HVS	676–683	<i>Impatiens holstii</i>	Spots/Color breaking/Flower distortion	MT/Aphid (<i>M. persicae</i>)	USA	NC038323
<i>Helleborus mosaic virus</i>	HeMV		<i>Helleborus niger</i>			Germany	NC043082
<i>Helleborus net necrosis virus</i>	HeNNV	800 × 17	<i>Helleborus</i> spp.	Black streaks/Black spots (sepals)		USA/Japan/New Zealand	NC012038
<i>Hippeastrum latent virus</i>	HLV	706	<i>Hippeastrum hybridum</i>	Mosaic/Chlorotic spots	MT (<i>G. globosa</i> , <i>H. niger</i>)	Netherlands/Taiwan	NC011540
<i>Hop latent virus</i>	HpLV	675 × 14	<i>Humulus lupulus</i>	Symptomless	Aphid (<i>Phorodon humuli</i>)/ MT	Europe/USA/Australia/ China	NC002552
<i>Hop mosaic virus</i>	HpMV	675 × 14	<i>Humulus lupulus</i>	Chlorotic vein-banding/Leaf distortion/Stunting	Aphid (<i>M. persicae</i> , <i>Marcrosiphum euphorbiae</i> , <i>P. humuli</i>)	Europe/Australia/ North America/China/Japan	NC010538
<i>Hydrangea chlorotic mottle virus</i>	HdCMV	650	<i>Hydrangea macrophylla</i>	Leaf blistering/Reddening/Chlorosis/Chlorotic mottle/Leaf distortion/Ringspot	MT Exp Hosts	USA/New Zealand/ South Korea	NC012869

<i>Kalanchoe latent virus</i>	KLV	688	<i>Kalanchoë blossfeldiana</i>	Symptomless	MT (<i>C. quinoa</i>)	Denmark	NC013006
<i>Ligustrum necrotic ringspot virus</i>	LNRSV	674	<i>Ligustrum japonicum</i> / <i>Mazus reptans</i>	Necrotic ringspot/Line patterns	MT	USA	NC010305
<i>Ligustrum virus A</i>	LVA		<i>Ligustrum obtusifolium</i>	Stunting/Leaf malformation		South Korea	NC031089
<i>Lily symptomless virus</i>	LSV		<i>Liliaceae</i>	Symptomless/Very mild mosaic	Aphid	Netherlands/South Korea/China/USA/India/Brazil	NC005138
<i>Melon yellowing-associated virus</i>	MYaV		<i>Cucumis melo</i>	Yellowing/Mottle	Whitefly/GT	Brazil	NC038324
<i>Mirabilis jalapa mottle virus</i>	MiMV		<i>Mirabilis jalapa</i>	Slight leaf mottle/Leaf wrinkling	MT	USA	NC016080
<i>Narcissus common latent virus</i>	NCLV		<i>Narcissus tazetta</i>	Yellow stripes/Mosaic		China	NC008266
<i>Nerine latent virus</i>	NeLV-		<i>Hippeastrum</i> spp.	Chlorotic streaking		Australia	NC028111
<i>Passiflora latent virus</i>	PLV	660	<i>Passiflora</i> spp.	Mosaic	MT	Germany/Australia/USA/Israel	NC008292
<i>Pea streak virus</i>	PeSV	619	<i>Pisum sativum</i> / <i>Trifolium</i> spp.	Severe necrosis	MT	North America/Europe	NC027527
<i>Phlox virus B</i>	PhIVB		<i>Phlox</i> spp.	Virus-like symptoms	MT	USA	NC009991
<i>Phlox virus M</i>	PhIVM		<i>Phlox</i> spp.	Virus-like symptoms		USA	
<i>Phlox virus S</i>	PhIVS		<i>Phlox</i> spp.	Virus-like symptoms	MT	USA	NC009383
<i>Poplar mosaic virus</i>	PopMV	685	<i>Populus trichocarpa</i>	Severe mosaic/Mild mosaic	MT (<i>N. megalosiphon</i>)	Germany	NC005343
<i>Potato latent virus</i>	PotLV	670	<i>Solanum tuberosum</i>	Symptomless/Slight vein clearing/Chlorotic spots	MT/Aphid (<i>M. persicae</i>)	Canada/USA/Netherlands	NC011525
<i>Potato virus H</i>	PVH	570	<i>Solanum tuberosum</i>	Mild symptoms	MT/Aphid (<i>M. persicae</i>)	Netherlands/China	NC018175
<i>Potato virus M</i>	PVM		<i>Solanum tuberosum</i>	Mottle/Mosaic/Crinkling/Rolling/Stunting	MT/Aphid (<i>M. persicae</i>)	Worldwide	NC001361
<i>Potato virus P</i>	PVP		<i>Solanum tuberosum</i>	Symptomless	MT/Aphid (<i>M. persicae</i>)	Brazil/Russia/Argentina	NC009759
<i>Potato virus S</i>	PVS	610–700	<i>Solanum tuberosum</i>	Symptomless	MT/Aphid (<i>M. persicae</i>)	Worldwide	NC007289
<i>Red clover vein mosaic virus</i>	RCVMV	645	<i>Pisum sativum</i> / <i>Cicer arietinum</i> / <i>Lens culinaris</i>	Chlorosis/Stunting/Rosetting/Vein clearing/Plant death	MT/Aphid (<i>Acyrtosiphon pisum</i>)	USA/New Zealand/Netherlands	NC012210
<i>Sambucus virus C</i>	SVC		<i>Sambucus</i> spp.	Mosaic/Leaf yellowing/Malformation/Dwarfing		USA	NC029087
<i>Sambucus virus D</i>	SVD		<i>Sambucus</i> spp.	Mosaic/Leaf yellowing/Malformation/Dwarfing		USA	NC029088
<i>Sambucus virus E</i>	SVE		<i>Sambucus</i> spp.	Mosaic/Leaf yellowing/Malformation/Dwarfing		USA	NC029089
<i>Shallot latent virus</i>	SLV	650–652	<i>Allium</i> spp.	Symptomless	MT/Aphid (<i>M. ascalonicus</i>)	Worldwide	MG657357
<i>Sint-Jan onion latent virus</i>	SJOLV		<i>Allium</i> spp.	Symptomless	MT/Aphid	Netherlands	NC043084
<i>Strawberry pseudo mild yellow edge virus</i>	SPMYEV	625 × 12	<i>Fragaria</i> spp.	Chlorotic mottle/Reddening/Necrosis	Aphid (<i>Chaetosiphon frageaefolii</i> , <i>Aphis gossypii</i>)	Japan	ND
<i>Sweet potato C6 virus</i>	SPC6V	750–800	<i>Ipomoea batatas</i>	Chlorotic spots/Vein clearing	MT/GT	Dominican Republic	NC018448
<i>Sweet potato chlorotic fleck virus</i>	SPCFV	800	<i>Ipomoea batatas</i>	Chlorotic flecks/Symptomless	MT	Worldwide	NC006550
<i>Verbena latent virus</i>	VeLV	650	<i>Verbena</i> spp./ <i>Tropaeolum majus</i>	Symptomless/Mild mottle	MT	Israel/New Zealand	NC043085
<i>Yam latent virus</i>	YLV		<i>Dioscorea opposita</i>			China	NC026248
Foveavirus							
Apple stem pitting virus	ASPV	800 × 15	<i>Malus</i> spp./ <i>Pyrus</i> spp./ <i>Prunus</i> spp.	Yellow banding/Red mottling/ flecking	MT/GT	Europe/Asia	NC003462

(Continued)

Table 1 Continued

Genus/Virus name	FYI	Particle size (nm)	Natural Hosts	Symptoms	Transmission	Distribution	Access #
<i>Apricot latent virus</i>	ApLV		<i>Prunus</i> spp.	Symptomless/Chlorotic spots	GT	Europe/Africa	NC014821
<i>Asian prunus virus 1</i>	APV1		<i>Prunus numel</i> / <i>Prunus persica</i>	Chlorotic spots/Fruit Deformation/Vein enlargement & discoloration/Necrotic spots	MT (<i>N. occidentalis</i>)/GT	USA	NC025388
<i>Asian prunus virus 2</i>	APV2		<i>Prunus mume</i>	Chlorotic spots/Fruit deformation/Vein enlargement & discoloration	GT	USA	NC028868
<i>Grapevine rupestris stem pitting-associated virus</i>	GRSPaV		<i>Vitis</i> spp.	Pitting/Grooving/Distortion	GT	USA/Uruguay/China	NC001948
<i>Grapevine virus T</i>	GVT		<i>Vitis vinifera</i>	Fanleaf symptoms/Symptomless		Europe	NC035203
<i>Peach chlorotic mottle virus</i>	PCMV		<i>Prunus persica</i>	Chlorotic mottle/Chlorotic ringspots	MT	Canada	NC009892
<i>Rubus canadensis virus 1</i>	RuCV1		<i>Rubus canadensis</i>	Mild chlorosis/Vein clearing	MT	USA	NC019025
Robigovirus							
<i>African oil palm ringspot virus</i>	AOPRV	800	<i>Elaeis guineensis</i>	Ringspots/Yellowing/Necrosis		South America/South Africa	NC012519
<i>Cherry green ring mottle virus</i>	CGRMV		<i>Prunus</i> spp.	Yellow mottle/Ringspots	GT	North America/Europe	NC001946
Cherry necrotic rusty mottle virus	CNRMV		<i>Prunus avium</i> / <i>Prunus persica</i>	Leaves: Brown necrotic spots/Rusty Chlorosis/ Shot holes Bark: Blisters/Gum pockets/ Necrosis	GT	North America/Europe/ New Zealand/South Korea	NC002468
<i>Cherry rusty mottle associated virus</i>	CRMaV		<i>Prunus avium</i>	Rusty mottle	GT	North America	NC020996
<i>Cherry twisted leaf associated virus</i>	CTLaV		<i>Prunus avium</i>	Twisting/Ring pox/Pit pox	MT (<i>N. occidentalis</i>)/GT	North America	NC024449
Unassigned species							
<i>Banana mild mosaic virus</i>	BaMMV	800	<i>Musa</i> spp.	Mild chlorotic streaks	Seed	India	NC002729
<i>Banana virus X</i>	BVX		<i>Musa</i> spp.		MT	France	NC04308
<i>Sugarcane striate mosaic-associated virus</i>	SCSMaV	950 × 15	<i>Saccharum</i> spp	Short chlorotic streaks		Australia	NC003870

Footnotes: MT; mechanical transmission, GT, graft transmission.

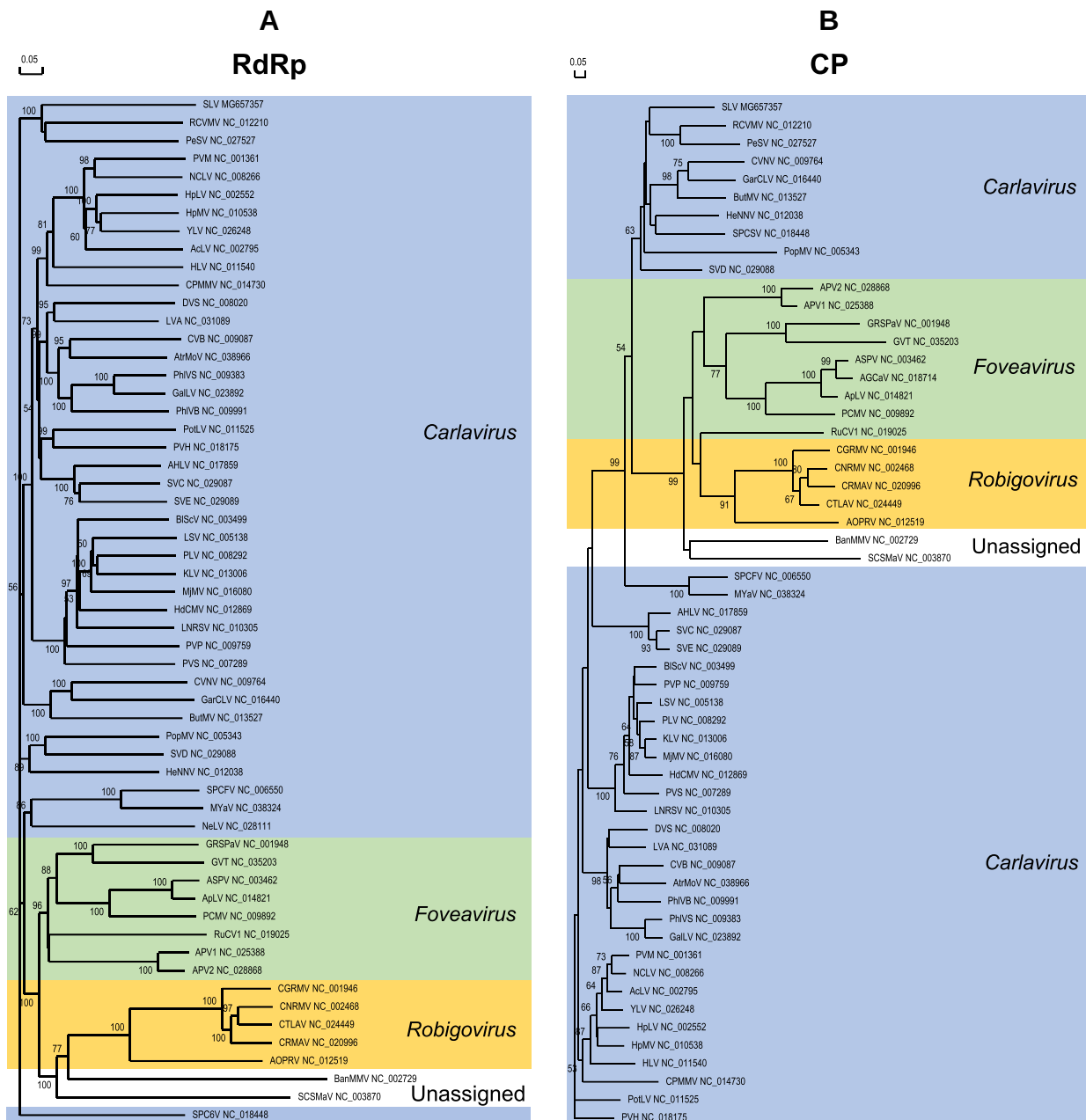


Fig. 1 Phylogenetic analysis of subfamily *Quinvirinae* based on the amino acid sequences of the RdRp (A) and CP (B). Multiple sequence alignments were generated using DNAMAN software version 5.1 (Lynnon Biosoft, San Ramon, CA, USA) and phylogenetic trees were constructed by the neighbor-joining algorithm, based on calculations from pairwise amino acid sequence distances for protein analyses derived from the multiple alignment format. The horizontal branch lengths are proportional to the genetic distance, and the numbers at each point indicate bootstrap values. The dataset was subjected to 10,000 bootstrap replicates. Square brackets denote accession numbers in the NCBI database.

coding for proteins of 223 kDa RdRp (5907 nt), 25 kDa TGBp1 (690 nt), 12 kDa TGBp2 (330 nt), 7 kDa TGBp3 (192 nt), 34 kDa CP (915 nt), and 11 kDa CRP (327 nt) from the 5' to 3'-end. The both terminal regions of TGBp2 (12 kDa) shows overlap with TGBp1 and TGBp3, respectively. The 5'-terminal region of ORF6 shows an overlap with ORF5 (CP). The lengths of the 5' non-translated region (NTR) and the 3'-NTR have 73 and 72 nt. The 3'-NTR region contains the poly (A) tail. The genome of ASPV, a member of the genus *Foveavirus*, consists of 9332 nt and contains five ORFs encoding 247 kDa RdRp (6552 nt), 25 kDa TGBp1 (672 nt), 13 kDa TGBp2 (363 nt), 7 kDa TGBp3 (213 nt), and 44 kDa CP (1245 nt) from the 5' to 3'-end. The 3'-terminal region of TGBp2 shows an overlap with TGBp3. The lengths of the 5'-NTR and the 3'-NTR have 59 and 133 nt. The 3'-NTR region contains the poly(A) tail. The genome of CNRMV, a member of the genus *Robigovirus*, has 8432 nt and contains five ORFs encoding 232 kDa RdRp (6117 nt), 25 kDa TGBp1 (669 nt), 12 kDa TGBp2 (348 nt), 7 kDa TGBp3 (204 nt), and 30 kDa CP (804 nt) from

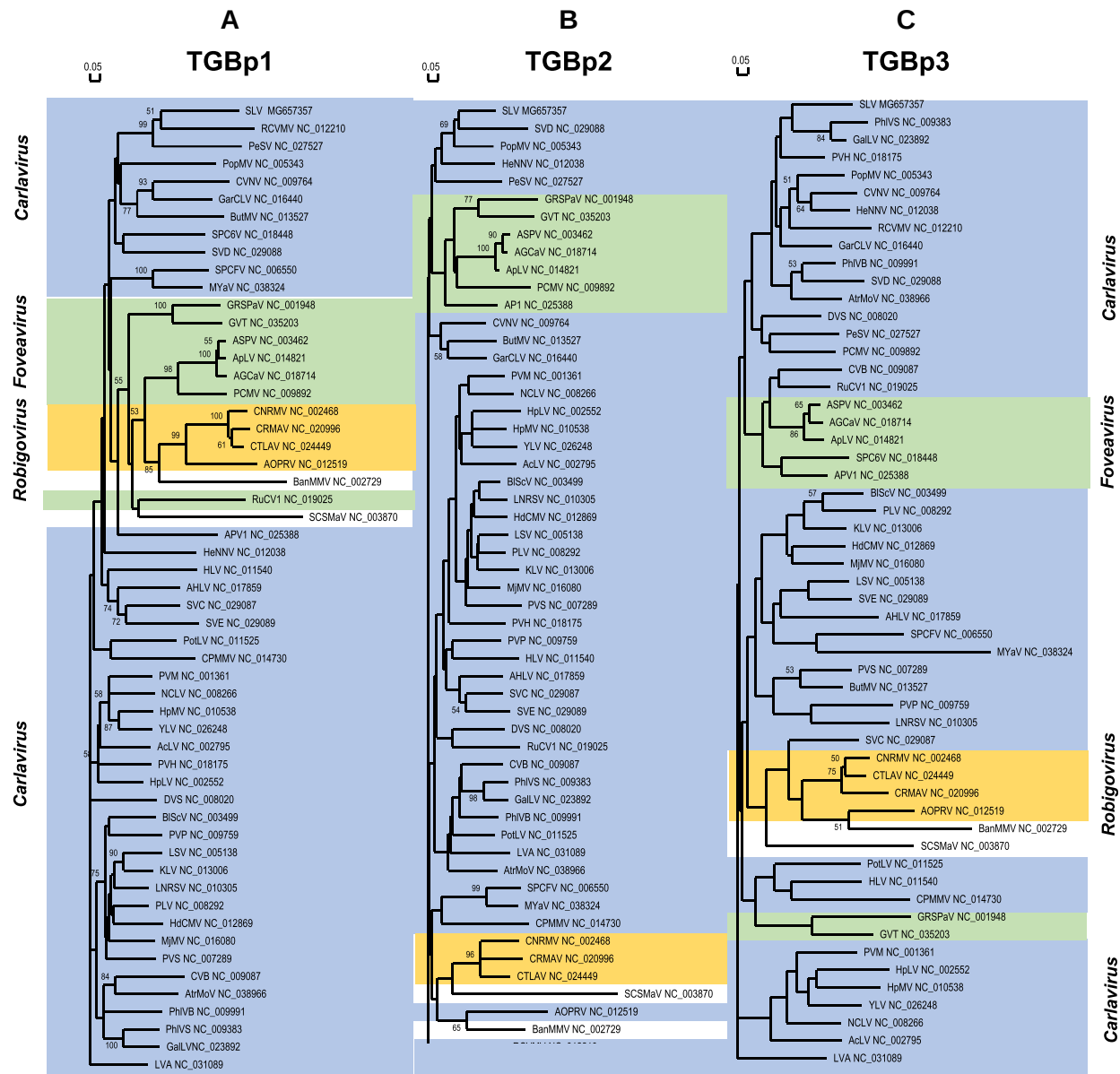


Fig. 2 Phylogenetic analysis of subfamily *Quinivirinae* based on the amino acid sequences of the TGBp1 (A), TGBp2 (B), and TGBp3 (C). Green and yellow boxes indicate respectively members of the genus *Foveavirus* and members of the genus *Robigovirus*. Multiple sequence alignments were generated using DNAMAN software version 5.1 (Lynnon Biosoft, San Ramon, CA, USA) and phylogenetic trees were constructed by the neighbor-joining algorithm, based on calculations from pairwise amino acid sequence distances for protein analyses derived from the multiple alignment format. The horizontal branch lengths are proportional to the genetic distance, and the numbers at each point indicate bootstrap values. The dataset was subjected to 10,000 bootstrap replicates. Square brackets denote accession numbers in the NCBI database.

the 5' to 3'-end. The 3'-terminal region of TGBp2 shows an overlap with TGBp3. The genome has putative 13 kDa and 18 kDa within the TGBp1 and the CP, respectively. The lengths of the 5'-NTR and the 3'-NTR have 117 and 192 nt. The 3'-NTR region contains the poly(A) tail.

Properties and Functions of Gene Products

The RdRp of quiniviruses contains five conserved domains: a methyl-transferase (MT) domain; an AlkB domain; a papain-like protease (*p*-Pro) domain; a RNA helicase (HEL) domain; a polymerase (POL) domain. The AlkB domain that repairs alkylated RNA is located in the RdRp of quiniviruses, potexviruses, and potyviruses. The papain-like protease (*p*-Pro) is related to polyprotein processing in viral replication. TGB resembling potex-like TGB that is essential for their cell-to-cell and long-distance movement.

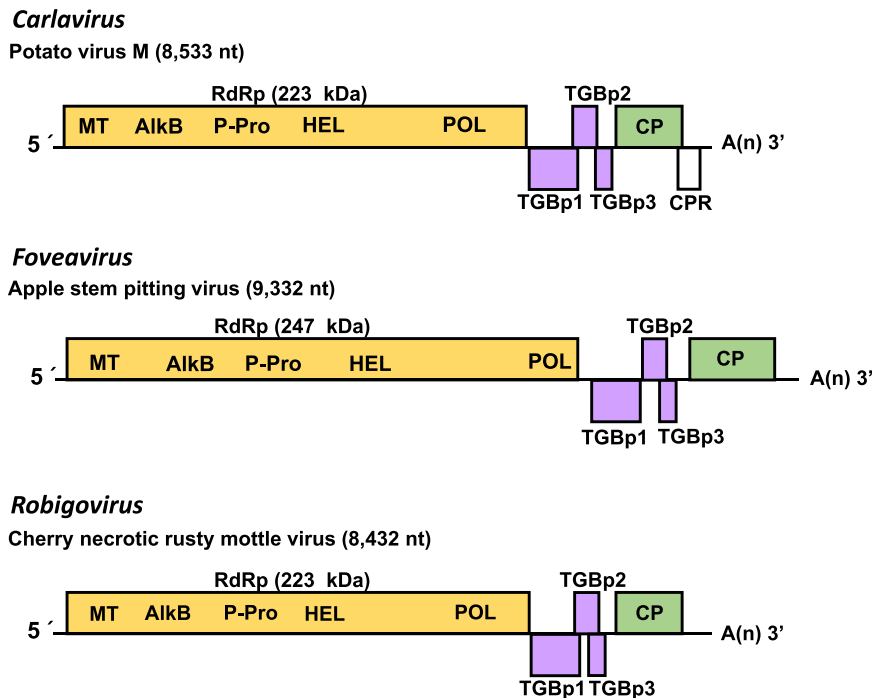


Fig. 3 Comparison of the genome organizations of the different genera (*Carlavirus*, *Foveavirus*, *Robigovirus*) in the subfamily *Quinvirinae*. RNA-dependent RNA polymerase (RdRp) includes a methyl-transferase (MT), an AlkB domain, a papain-like protease (P-Pro) domain, a helicase (HEL) domain, and a polymerase (POL) domain. Gray box indicates triple gene block corresponding to three overlapping ORFs (TGBp1, TGBp2, and TGBp3). ORF5 encodes coat protein (CP). The ORF6 of PVM, genus *Carlavirus*, encodes a cysteine-rich protein (CRP).

TGBp1 is associated with ATPase, RNA-binding, and RNA-helicase activities. TGBp2 and TGBp3 are integral membrane proteins required for virus movement. The TGBp1 and CP form a complex with viral RNA that transfers via the plasmodesmata. TGBp1 acts as viral suppressors of RNA silencing. The TGBp1 of PVM demonstrates suppressor activity on systemic silencing. The CP is required for virion assembly. The CRP (ORF6) of carlaviruses contains two motifs: a nuclear localization signal (NLS) and a zinc finger motif (ZF). The CRP is implicated as silencing suppressors for a pathogenicity determinant. The CRPs of PVM, DVS, Chrysanthemum virus B (CVB), Lily symptomless virus (LSV), HeNNV, and Narcissus common latent virus (NLV) have the ability to enhance the pathogenicity of a heterologous virus (Potato virus X) and effects on virus accumulation levels. The CRPs of genera *Benyvirus*, *Tobravirus*, *Furovirus*, and *Hordeiviruses* have a similar role to that of the carlavirus CRP. The analysis of simple sequence repeats (SSRs) and compound simple sequence repeat (cSSR) have been demonstrated in 32 species of carlaviruses. Mononucleotide A/T is most commonly followed by dinucleotide GT/TG and trinucleotide AAG/GAA in their genome. The SSR and cSSR are mainly located in the RdRp and CRP independent of their genome size.

Replication and Propagation

ORF1 (RdRp) is translated directed from the genomic RNA. The ORF1 of Blueberry scorch virus (BlScV), a member of the *Carlavirus* genus, is proteolytically processed by a papain-like proteinase. The largest product of 190 kDa can be synthesized *in vitro* using the viral RNA of Potato virus S (PVS) isolated from infected potato plant. The proteolytic processing of the 190 kDa protein is associated with accumulation of 150 kDa peptide after 60 min incubation. L-canavanine is related to specifically reduce the quantity of 34 kDa for PVS CP. *Carlavirus* ORF1 and TGBp3 contain a heptanucleotide motif (C/UUUUAGGU) for sub-genomic RNA promoter. Two sub-genomic RNAs for TGB, CP, and CRP synthesis have been demonstrated in some carlaviruses. These sub-genomic RNAs might be encapsidated into their viral particles. Virus particles are usually found in the cytoplasm in infected plant.

Transmission and Host Range

Most carlaviruses are spread by mechanical transmission in horticultural crops and herbaceous plants. Some carlaviruses, American hop latent virus (AHLV), BlScV, DVS, LSV, Hop latent virus (HPLV), Hop mosaic virus (HpmV), Potato virus S (PVS), PVM, and PotLV, can be transmitted non-persistently by aphid. AHLV, HPLV, and HpmV are known to infect to hop plants and known to occur in mixed infection. The three viruses are transmitted by hop aphid (*Phorodon humuli*). HPLV and HpmV are

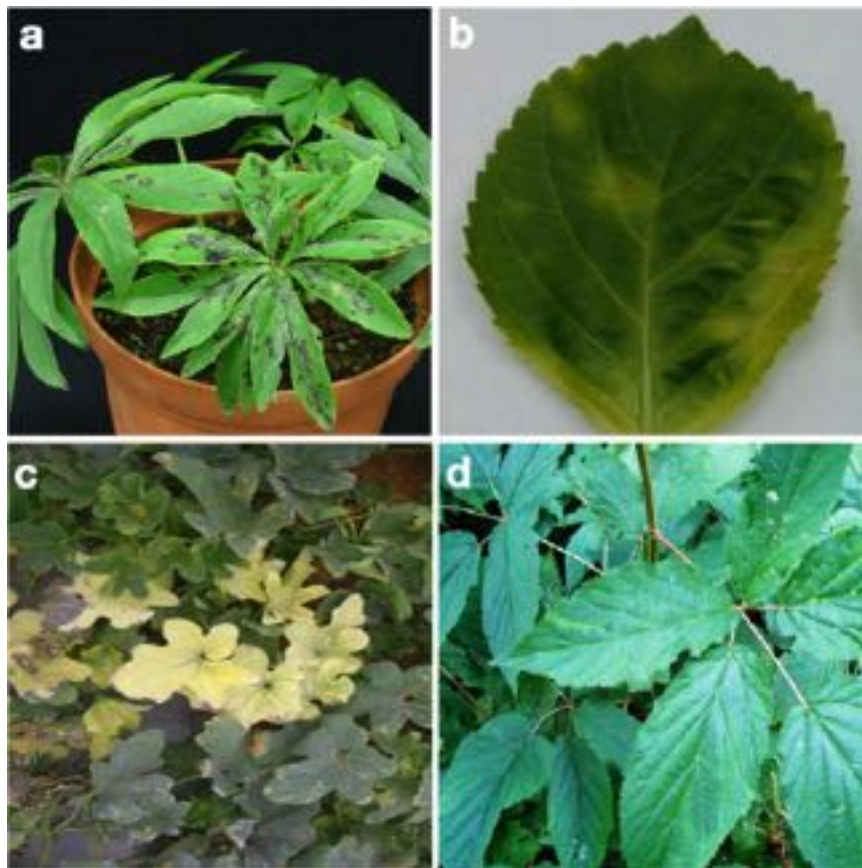


Fig. 4 Symptoms of diseases induced by carlaviruses (a). Helleborus net necrosis virus (HeNNV; *Carlavirus*, *Quinvirinae*, *Betaflexiviridae*) on *Helleborus* spp. (b). Hydrangea chlorotic mottle virus (HdCMoV; *Carlavirus*, *Quinvirinae*, *Betaflexiviridae*) on *Hydrangea macrophylla* (c). Melon yellowing associated virus (MYaV; *Carlavirus*, *Quinvirinae*, *Betaflexiviridae*) on watermelon and on cucumber (d). Source of photographs: (a) Shiraishi, T., Hoshi, H., Eimori, K., *et al.*, 2011. First report of Helleborus net necrosis virus isolated from hellebores with black death syndrome in Japan. *Journal of General Plant Pathology* 77, 269–272. doi:[10.1007/s10327-011-0321-2](https://doi.org/10.1007/s10327-011-0321-2). (b) Tang, J., Harper, S.J., Wei, T., *et al.*, 2010. Characterization of hydrangea chlorotic mottle virus, a new member of the genus *Carlavirus*. *Archives of Virology* 155, 7–12. doi:[10.1007/s00705-009-0541-3](https://doi.org/10.1007/s00705-009-0541-3). (c) & (d) Nagata, T., Dutra, L.S., Oliveira, P.A., *et al.*, 2010. Analysis of the triple gene block sequence in an important melon pathogen, Melon yellowing-associated virus. *Journal of General Plant Pathology* 76, 268–272. doi:[10.1007/s10327-010-0241-6](https://doi.org/10.1007/s10327-010-0241-6).

transmitted by potato aphid (*Macrosiphum euphorbiae*) and green peach aphid (*Myzus persicae*), and can also be transmitted by mechanical inoculation. AHLV and HpLV cause commonly symptomless on any commercial hop plants. HpMV causes chlorotic mosaic mottling on susceptible hop plants. BLSv can cause necrosis of flower and young leaves, shoot blight, and chlorosis in highbush blueberry (*Vaccinium corymbosum* L.) whereas it appears to be asymptomatic in cranberry (*Vaccinium macrocarpon* L.). PVS is mechanically transmissible and causes severe disease in potato and localized infection of *Chenopodium* spp. The virus recently has been detected from arracacha (*Arracacia xanthorrhiza*) in Peru. Two purified carlaviruses, Pea streak virus and Red clover vein mosaic virus (RCVMV) causing severe streaking symptoms on pea, can be transmitted by *Acyrthosiphon pisum* (pea aphid). The RCVMV can be mechanically transmitted to chickpea (*Cicer arietinum* L.), faba bean (*Vicia faba* L.) and lentil (*Lens culinaris* Medik.). Ligustrum necrotic ringspot virus (LNRSV) isolated from *Ligustrum japonicum* Thunb showing necrotic ringspots and line patterns is mechanically transmissible to *Chenopodium amaranticolor*. CPMV is transmitted non-persistently by whitefly and is the possibility of seed transmission. The virus can cause systemic mottling, chlorotic blotches, and leaf malformations in cowpea (*Vigna unguiculata* L.) and can be naturally transmitted to peanut (*Arachis hypogaea* L.), soybean (*Glycine max* L.), and common bean (*Phaseolus vulgaris* L.). Cucumber vein-clearing virus (CuVCV) can be transmitted by mechanical inoculation and by whitefly. Melon yellowing-associated virus also is transmitted by whitefly but is not transmitted by mechanical inoculation. Rose virus A has been detected recently in a memorial rose (*Rosa wichuraiana* Cresp.) showing leaf deformation and mosaic symptoms. Foveaviruses and robogoviruses naturally infect woody plants (Fig. 5). These viruses can be transmitted by grafting and sap inoculation. ASPV causes systemic infection on *Nicotiana occidentalis* and local infection on *Chenopodium murale*. ApLV, a member of the genus *Foveavirus*, causes chlorotic spots on *Prunus persica* and *P. cerasifera*, red to purple rings and mottling on *P. avium*, and chlorotic blotching and malformation on *P. armeniaca* cv. Tirynthos and Haward but not on flowers and fruits. The ApLV can cause asymptomatic infection on cherry and plum cultivars. CGRMV can be sap transmitted from cherry to cherry. Novel foveavirus has been detected in stems and leaves of ginseng (*Panax ginseng*) exhibiting mosaic and ringspot symptoms. GRSPaV can be transmitted by grafting but not

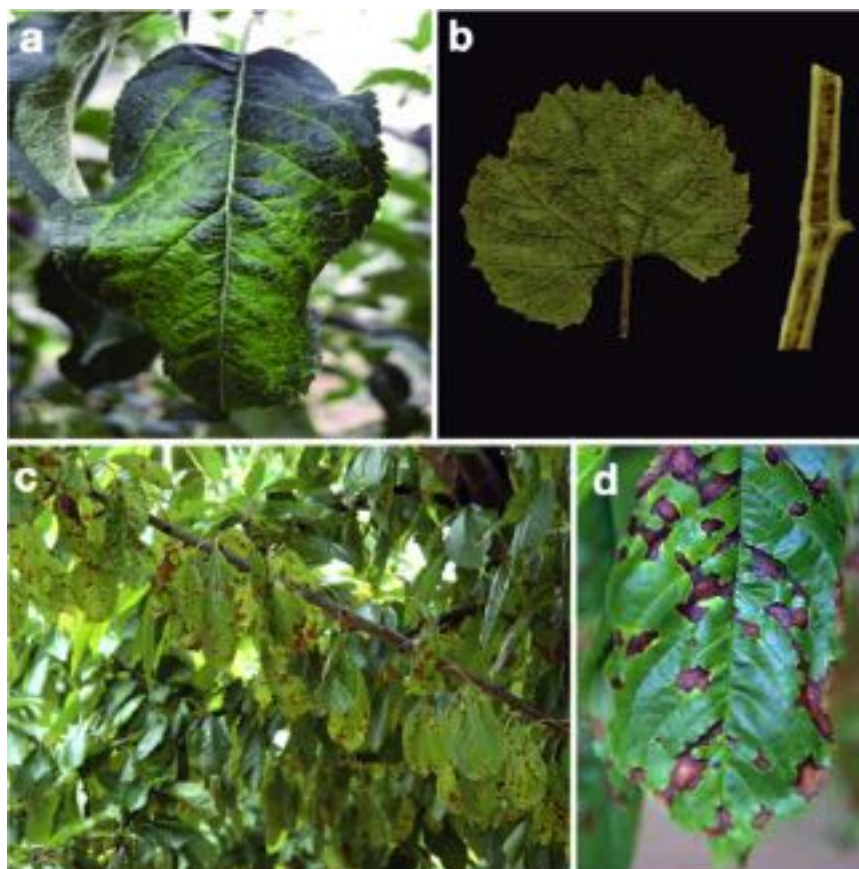


Fig. 5 (a) Apple cultivar 'Hongro' with mixed infection of Apple stem pitting virus (ASPV; *Foveavirus*, *Quinvirinae*, *Betaflexiviridae*), Apple chlorotic leaf spot virus (ACLSV, *Trichovirus*, *Trivirinae*, *Betaflexiviridae*), and Apple stem grooving virus (ASGV; *Capillovirus*, *Trivirinae*, *Betaflexiviridae*), showing chlorosis along the leaf veins. (b) Symptoms of Grapevine rupestris stem pitting-associated virus (GRSPaV; *Foveavirus*, *Betaflexiviridae*) of vein necrosis in the grapevine indicator 110 R: necrosis of leaf veinlets and petioles (left) and of the pith (right). (Photograph courtesy of G. Martelli). (c) Necrotic symptoms from Cherry necrotic rusty mottle virus (CNRMoV; *Robigovirus*, *Quinvirinae*, *Betaflexiviridae*) on Sweetheart cherries tree (Photo from Iain MacSwann) and (d) detail of necrosis on Corum cherry leaves (Photograph from Jay W. Pscheidt). Source of photographs: (a) Yoon, J.Y., Joa, J.H., Choi, K.S., *et al.*, 2014. Genetic diversity of a natural population of Apple stem pitting virus isolated from apple in Korea. *The Plant Pathology Journal* 30 (2), 195–199. doi:10.5423/PPJ.NT.02.2014.0015. (b) Bouyahia, H., Boscia, D., Savino, V.N., *et al.*, 2015. Grapevine rupestris stem pitting-associated virus is linked with grapevine vein necrosis. *Vitis* 44 (3), 133–137. (d) Pacific Northwest Plant Disease Management Handbook. <https://pnwhandbooks.org/node/2463>.

transmitted by mechanical inoculation. The five members of the genus *Robigovirus* can be transmitted to cherry and other stone fruit tree and cause serious diseases (Figs. 4 and 5; Table 1).

Epidemiology and Control

Quinviruses can cause usually mosaic, mottle, and ringspot symptoms in their host plants. PVS and PVM, members of the genus *Carlavirus*, are essentially symptomless in most potato cultivars. The two viruses are easily transmitted during the cutting of tubers to make seed pieces. The PVS can cause yield losses of up to 20% in potato field. The primary method for producing virus-free potato tuber seeds has been used to control the viruses. The *Ns* gene derived from potato (*Solanum tuberosum* spp. *andigena*) induces hypersensitive resistance to PVS and have been used in potato breeding programs. Orchards and vegetables sustain economic damage caused by some quinviruses that require suitable disease management. MYaV, a member of the genus *Carlavirus*, causes severe yellowing disease in melon crops and could transmitted by whitefly. New melon plants containing a dominant MYaV resistance gene recently have been developed for the virus disease management. A whitefly-transmitted CPMMV, a member of the genus *Carlavirus*, causes substantial losses in the yield and quality of cowpea. The whitefly-transmitted CPMMV can be controlled with insecticides such as Nuprid 2F and Dimethoate. The SSR markers associated with CPMMV resistance in cowpea have been recently identified and will be used for new cowpea variety resistance to CPMMV. HeNNV, a member of the genus *Carlavirus*, associated with black death of *Helleborus* spp. has been reported in New Zealand, Japan, UK, and the USA. LSV, a member of the genus *Carlavirus*, which is generally symptomless in field-grown plants may cause problems under the insufficient light and

growing conditions of year-round cut-flower production in greenhouses. In the Netherlands, an aphid-transmitted LSV spreads mainly in June–July, considerably less in May, and least in August/September. It is curtailed by a mixture of mineral oil and pyrethroid insecticide. Three hop viruses, AHpLV, HpLV, and HpMV, can be controlled by the use of certified virus-free planting stock and insecticides for reducing aphids. HpMV has been found infecting hops in Europe, Australia, the United States, New Zealand, China, Japan, and South Africa. Blueberry scorch disease caused by aphid-transmitted BLSV is a severe disease of highbush blueberry (*Vaccinium corymbosum* L.) in Canada, Europe, and the USA. The viral disease can be controlled by heat therapy and apical meristem culture. Transgenic chrysanthemum, tobacco, and potato, which are resistant to CPMMV, Chrysanthemum virus B, PVM, and PVS have been developed.

Diagnosis

Quinviruses can be detected by polymerase chain reaction (PCR)-based methods, such as reverse transcription (RT)-PCR, immunocapture-RT-PCR, multiplex RT-PCR, and real-time PCR. Universal primers for detection of potato carlaviruses (PVS, PVM, and PotLV) have been developed based on the conserved aa sequences “SNNMA” and “GLGVPTE” in the CP. Other universal primer, a Carla-CP, has been designed based on the conserved nt sequences of the GLGVPTE motif in the CPs derived from 17 carlavirus species (Aconitum latent virus, BLSV, Carnation latent virus, Cole latent virus, CPMMV, CVB, DVS, Helenium virus S, HpMV, LSV, Narcissus common latent virus, Passiflora latent virus, PotLV, PVM, Potato virus P, PVS). Carlaviruses can be detected by RT-PCR using the Carla-CP and oligo-d(T21) primers. Some commercial and non-commercial antisera are used for immunological detection of quinviruses. Four carlaviruses, CLV, HpLV, HpMV, and PVS are antigenically similar.

Concluding Remarks

Two genera, *Carlavirus* and *Foveavirus*, belong to subfamily *Quinvirinae* and a newly created genus *Robigovirus* has been recently added to that subfamily. The completed genome sequences of novel carlaviruses (LAV, MYaV, and SPCFV) and foveavirus (GVT) have been determined recently. African oil palm ringspot virus, Cherry green ring mottle virus, CNRMV, Cherry rusty mottle associated virus, and Cherry twisted leaf associated virus belong to newly established genus *Robigovirus*. In phylogenetic trees based on their RdRp and CP proteins of members of the subfamily *Quinvirinae*, the genera *Carlavirus* and *Foveavirus* have not been grouped according to their natural host species. Cowpea, cherry, grapevine, melon, and potato cause severe symptoms by some quinviruses. Some carlaviruses can be transmitted by aphid and whitefly. Quinviruses have TGB resembling potex-like TGB for viral movement. Carlaviruses have two proteins (TGBp1 and CRP) identified as RNA silencing suppressor. Immunological and PCR-based techniques are most commonly used for detection of quinviruses.

Further Reading

- Alam, C.M., Singh, A.K., Sharfuddin, C., Ali, S., 2014. Genome-wide scan for analysis of simple and imperfect microsatellites in diverse carlaviruses. *Infection, Genetics and Evolution* 21, 287–294.
- Belay, D.K., Huckaba, R.M., Ramirez, A.M., Rodrigues, C.V., Foster, J.E., 2012. Insecticidal control of *Bemisia tabaci* (Hemiptera: Aleyrodidae) transmitting Carlavirus on soybeans and detection of the virus in alternate hosts. *Crop Protection* 35, 53–57.
- Bhattarai, G., Shi, A., Qin, J., *et al.*, 2017. Association analysis of Cowpea mosaic virus (CPMV) resistance in the USDA cowpea germplasm collection. *Euphytica* 213, 230.
- Costa, T.M., Blawid, R., da Costa Junior, A.C., *et al.*, 2017. Complete genome sequence of Melon yellowing-associated virus from melon plants with the severe yellowing disease in Brazil. *Archives of Virology* 162, 3899–3901.
- De Souza, J., Gamarra, H., Müller, G., Kreuze, J., 2018. First report of Potato virus S naturally infecting arracacha (*Arracacia xanthorrhiza*) in Peru. *Plant Disease* 102, 460.
- Diaz-Lara, A., Molloy, D., Golino, D., Rwahni, M.A., 2020. Complete genome sequence of Rose virus A, the first carlavirus identified in rose. *Archives of Virology* 165, 241–244.
- Fujita, N., Komatsu, K., Ayukawa, Y., *et al.*, 2016. N-terminal region of cysteine-rich protein (CRP) in carlaviruses is involved in the determination of symptom types. *Molecular Plant Pathology* 19, 180–190.
- Ho, T., Quito-Avila, D., Keller, K.E., *et al.*, 2016. Evidence of sympatric speciation of elderberry carlaviruses. *Virus Research* 215, 72–75.
- Park, D., Zhang, M., Hahn, Y., 2019. Novel foveavirus (family *Betaflexiviridae*) species identified in geinseng. *Acta Virologica* 63, 155–161.
- Rodamilans, B., Shan, H., Pasin, F., Garcia, J.A., 2018. Plant viral protease: Beyond the role of peptide cutters. *Frontiers in Plant Science* 9, 666.
- Williamson, M.A., Scott, S.W., 2018. First report of hellebore black death of Lenten rose, caused by *Helleborus net necrosis virus*, in South Carolina. *Plant Disease* 102, 2667.
- Wu, L., Liu, H., Bateman, H., Komorowska, B., Li, R., 2019. First identification and molecular characterization of a novel cherry robiovirus. *Archives of Virology* 164, 3103–3106.

Reverse-Transcribing Viruses (*Belpaoviridae*, *Metaviridae*, and *Pseudoviridae*)

Carlos Llorens, Biotechvana, Scientific Park University of Valencia, Valencia, Spain

Beatriz Soriano, Biotechvana, Scientific Park University of Valencia and Institute for Integrative Systems Biology (I2SysBio), University of Valencia–Spanish National Research Council, Valencia, Spain

Maria A Navarrete-Muñoz, Biotechvana, Madrid, Spain; Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid; and Rey Juan Carlos University Hospital, Móstoles, Spain

Ahmed Hafez, Biotechvana, Valencia, Spain; Pompeu Fabra University, Barcelona, Spain; and Minia University, Minya, Egypt

Vicente Arnau, Institute for Integrative Systems Biology (I2SysBio), University of Valencia–Spanish National Research Council, Valencia, Spain

Jose Miguel Benito, Health Research Institute of the Jiménez Díaz Foundation, Autonomous University of Madrid and Rey Juan Carlos University Hospital, Móstoles, Spain

Toni Gabaldon, Barcelona Supercomputing Center-National Center for Supercomputing, Institute of Research in Biomedicine, and Catalan Institution for Research and Advanced Studies, Barcelona, Spain

Norma Rallon, Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid and Rey Juan Carlos University Hospital, Móstoles, Spain

Jaume Pérez-Sánchez, Institute of Aquaculture Torre de la Sal, Spanish National Research Council, Castellon, Spain

Mart Krupovic, Archaeal Virology Unit, Institut Pasteur, Paris, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

CP Capsid protein

ENV Envelope

ERV Endogenous retrovirus

Gag Group-specific antigen

GYA Billion (giga) years ago

HERV Human endogenous retrovirus

ICTV International Committee on Taxonomy of Viruses

INT Integrase

Int-Chr Chromodomain-containing Integrase

kb Kilobase

LTR Long terminal repeat

MYA Million (mega) years ago

NC Nucleocapsid

ORF Open reading frame

OTU Operative taxonomical unit

PBS Primer binding site

PPT Polypurine track

PR Protease

RH Ribonuclease H

RNAi RNA interference

RNP Ribonucleoprotein

RT Reverse transcriptase

ssRNA single-stranded ribonucleic acid

SU Surface

TM Transmembrane

VLP Virus like particle

Glossary

Chromoviruses In reference to metaviruses coding for chrodomain-like integrases.

Errantivirus Erranti comes from the Latin errans, meaning to wander.

LTR-retroelement Term used to collectively refer to retroviruses and LTR-retrotransposons, regardless of whether they possess env or env-like proteins.

Metavirus Meta comes from the Greek metathesis for transposition.

Monophyletic Phylogenetic group collecting the full set of operational taxonomic units descending from a single ancestor.

Operational taxonomic unit Termed used to classify groups of closely related individuals.

Ortervirales Orter is an inversion of retro, which was derived from reverse transcription; and virales - which is the suffix for an order in virology.

Paraphyletic clade Phylogenetic group collecting some but not all operational taxonomic units descending from a single ancestor.

Polyphyletic Phylogenetic group composed of OTUs descending from more than one ancestor.

Pro-retroviruses Term used here to describe LTR retrotransposons that have acquired a third ORF, *env*, and are thus potential retrovirus but with no evidence of infectivity.

Pseudovirus Pseudo comes from Greek pseudo, “false” to connote some uncertainty as to whether these are true viruses.

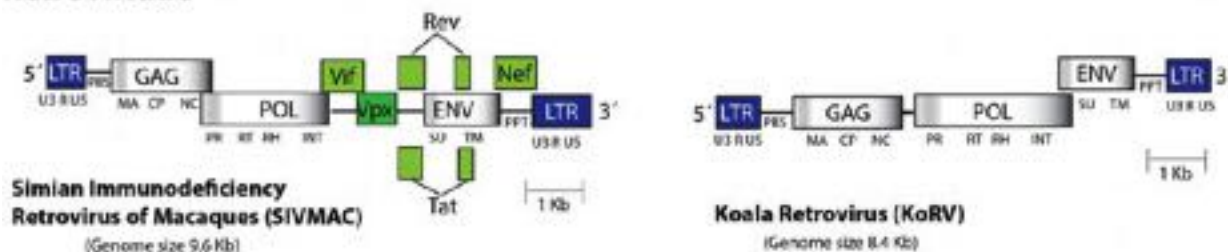
Semotivirus Semoti comes from the Latin semotus, meaning distant or removed (it was formerly used to classify Belpao elements).

Introduction: The Paradigm of LTR Retrotransposons and Reverse-Transcribing Viruses

Metaviridae, *Pseudoviridae*, and *Belpaoviridae* are the three families officially introduced by the International Committee on Taxonomy of Viruses (ICTV) for classification of the eukaryotic Ty3/Gypsy, Ty1/Copia, and Bel/Pao LTR retroelements. The members of the families *Metaviridae*, *Pseudoviridae*, and *Belpaoviridae* are widely (but differentially) distributed in plants, fungi and animals where they may exist as LTR retrotransposons or as potential or true reverse-transcribing viruses. Due to multiple distinct similarities in the life cycles, sequence and architecture of genes, and functional features, it is commonly accepted that members of the families *Metaviridae*, *Pseudoviridae*, and *Belpaoviridae* share origins and evolutionary history with viruses of the *Retroviridae* family, which is thus considered as a member of the same phylogenetic lineage of reverse-transcribing viruses and LTR retrotransposons. It is worth for the reader to distinguish simple reverse-transcribing viruses from the complex ones. In the family *Retroviridae*, simple retroviruses are those that present a core genomic organization of *gag*, *pol*, and *env* genes flanked by LTRs, with no accessory genes (i.e., LTR-*gag-pol-env*-LTR), while complex retroviruses incorporate one or more accessory genes (*vif*, *vpx*, *rev*, *nef*, *tat*, or others) needed to complete the life cycle and infect other cells (Fig. 1). In their most current form, the members of the families *Metaviridae*, *Pseudoviridae*, and *Belpaoviridae* are LTR retrotransposons presenting a “LTR-*gag-pol*-LTR” genomic architecture (Fig. 1). Thus, the major difference between LTR retrotransposons and simple retroviruses is that the later include a third gene (*env*) for an envelope protein which confers infectious abilities to the retroviruses. However, some members of the families *Metaviridae*, *Pseudoviridae*, and *Belpaoviridae* also carry the *env* gene, suggesting that an LTR retrotransposon might become a retrovirus upon acquiring a third *env* gene. Among metaviruses, pseudoviruses and belpaoviruses with *env*, only those from the *Errantivirus* genus in the family *Metaviridae* have been demonstrated to generate infectious virions. Interestingly, the LTR retrotransposon-virus paradigm gets support also in the reverse path because a reverse-transcribing virus may become or act as an LTR retrotransposon when the *env* gene is lost and/or the machinery required for infecting other cells is neutralized. This is the case for most endogenous retroviruses (ERVs) of vertebrates, which usually derive from exogenous retroviruses that after colonizing the germ cells became unable to infect other cells due to accumulation of multiple mutations. Interestingly, the replication system of LTR retroelements (i.e., reverse-transcribing viruses and LTR retrotransposons) constituted by the members of the families *Metaviridae*, *Belpaoviridae*, *Pseudoviridae*, and *Retroviridae* is also related to that of reverse-transcribing viruses of the family *Caulimoviridae* which infect plants. Caulimoviruses are dsDNA viruses whose genome usually harbors two ORFs encoding for coat (*gag*) and *pol* polyproteins (Fig. 1), with domain features closely similar to those of metaviruses, pseudoviruses and belpaoviruses. These similarities suggest that caulimoviruses share a common evolutionary history with LTR retrotransposons and that caulimoviruses would be chimeric viruses resulting from recombination between a metavirus and another RNA virus or LTR retrotransposon that evolved to the status of a virus by recruiting new genes through horizontal transfer. In recognition of all the structural, functional and evolutionary relationships among the members of the families *Caulimoviridae*, *Belpaoviridae*, *Metaviridae*, *Pseudoviridae*, and *Retroviridae*, the ICTV has recently unified the five families into a single viral order, the *Ortervirales*.

Genome Organization and Functions of Gene Products and Other Structural Features

The full-length genome sequence of canonical LTR retroelements of the *Metaviridae*, *Pseudoviridae*, and *Belpaoviridae* families normally present two LTRs that flank an internal region of variable size (from 3 to 15 kb) that contains open reading frames (ORFs) for the encoded *Gag*, *Pol*, and/or *Env* protein domains (Fig. 1). The LTRs are two normally homologous non-coding sequences that usually begin and end in dinucleotide (5'-TG ... CA-3') inverted repeats. A typical LTR displays three regions: a U3 region of 200–1200 nt that contains the promoters; a region called “R” because it is repeated on each end of the transcript; and a U5 region of 75–250 nt that constitutes the first portion of the reverse-transcribed genome. Normally, the LTRs do not contain coding genes but present regulatory elements such as “enhancers” and “promoters” that regulate the expression of the genes in the internal region of the genome and also, in certain cases, the host genes. The internal region is usually delimited by two small motifs: a 18 nt sequence downstream to the 5' LTR called Primer Binding Site (PBS) and by another small region of ~10 A/G located upstream of the 3' LTR called “Polypurine Tract” (PPT). The internal region normally includes ORFs encoding for two polyproteins, *Gag* and *Pol*. The *Gag* polyprotein presents domains for the structural proteins including the capsid protein (CP), which forms the core of the Virus-Like Particle (VLP), and a nucleocapsid (NC) involved in RNA packaging through recognition of a specific region in the viral genome called Ψ . Curiously, Ψ is a frequent site of deletions in the proviral genome, giving defective VLPs which could generate non-infectious virions in some reverse-transcribing viruses with the entire LTR-*gag-pol-env*-LTR structure. On the other hand, *Pol* presents domains for distinct enzymes such as a protease (PR) belonging to the clan AA of aspartic peptidases that is required for proteolytic processing of the two *Gag* and *Pol* polyproteins; a reverse transcriptase (RT) involved in the reverse transcription synthesis of cDNA from the single-stranded RNA (ssRNA) template; a ribonuclease H (RH) responsible for the hydrolysis of the original RNA template that is part of the RNA/DNA hybrid generated during reverse transcription; an integrase (INT) needed for the integration of the proviral cDNA into the host genome. As also shown in Fig. 1, viruses in the *Metaviridae* and *Belpaoviridae* families usually present a common organization of *gag-pol* domains (CP-NC-PR-RT-RH-INT), which are also observed in retroviruses. The only difference between the two groups is that retroviruses additionally contain the matrix domain (MA) in the *Gag* polyprotein, upstream of the CP, which is not present (or at least not preserved at the domain level) in metaviruses and belpaoviruses (nor in pseudoviruses). The order of the RT and INT domains in the internal region of members of the *Pseudoviridae* family is different from that of metaviruses, belpaoviruses and retroviruses. While

Retroviridae**Metaviridae****Caulimoviridae**

Cauliflower mosaic virus (CaMV)
(Genome size 8.6 Kb)

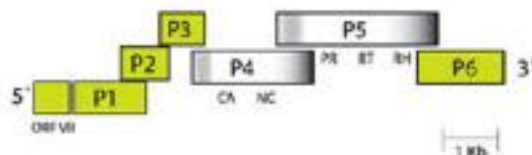
**Belpaoviridae****Pseudoviridae**

Fig. 1 Representative elements of each family within the order *Ortervirales*. Simian immunodeficiency virus (SIV) of *Macaca mulatta* is taken as a complex retrovirus example of the family *Retroviridae*, where the typical accessory genes are colored green. Koala retrovirus (KoRV) is used as an example of a simple retrovirus in the family *Retroviridae*. Legolas and Gypsy are used as representatives for the family *Metaviridae*. Copia and Sire are shown as representatives of the family *Pseudoviridae*. Pao and Tas represent the family *Belpaoviridae*. The Cauliflower mosaic virus (CaMV) is provided as a representative of the family *Caulimoviridae*. Domain features of each *gag*, *pol*, and *env* region are indicated below each represented genome description. PBS and PPT acronyms mean Primer Binding Site and Polypurine Tract, respectively. MA, CP and NC acronyms in *gag* refer to Matrix, Capsid and Nucleocapsid domains, respectively. PR, RT, RH and INT acronyms in *pol* refer to Protease, Reverse Transcriptase, Ribonuclease H and Integrase domains, respectively. SU and TM acronyms in *env* refer to Surface and Transmembrane domains, respectively. P1, P2, P3 and P6 in caulimoviruses refer to Movement protein, Aphid transmission factor, Virion associated protein, and Transactivator/Viroplasm, respectively. For more examples of orterviral representative full-length genomes, visit the GyDB (<http://gydb.org>). In addition, Table 1 provides a summary of all members of the *Metaviridae*, *Belpaoviridae*, and *Pseudoviridae* families currently classified by the ICTV.

metaviruses and belpaoviruses usually present RT-RH domains upstream of INT, pseudoviruses present a distinct organization with the INT domain being interspersed between PR and RT domains. Besides the “LTR-gag-pol-LTR” domain architecture, many lineages of the members of the *Metaviridae*, *Pseudoviridae* and *Belpaoviridae* families additionally contain a third *env* ORF which potentially endows them with infectivity. In retroviruses, Env protein binds to a receptor on the target cell surface and this event triggers a conformational change which leads to the fusion of viral and cellular membranes. Furthermore, at least some *Metaviridae* members, such as those belonging to the Tat clade in plants, encode accessory proteins (similar to complex retroviruses), whose function, however, remains unclear.

Taxonomy, Phylogeny and Members of the Families

Classification of the distinct LTR retroelements or viruses in the order *Ortervirales* is performed by phylogenetic methods. In Fig. 2, we show the phylogeny of the *Ortervirales* inferred based on the Pol polypeptide domains. For simplicity, branches of caulimoviruses and retroviruses are collapsed. Each family splits into distinct phylogenetic lineages, herein referred to as clades or branches (they are bigger in terms of both OTUs - Operative Taxonomical Units - and host distribution). We use popular nomenclature to refer to each clade or branch in the tree. In the following section the taxonomic structure of each family is discussed in detail. As an additional information, Table 1 provides a summary of all members of the *Metaviridae*, *Belpaoviridae*, and *Pseudoviridae* families currently classified by the ICTV.

The *Metaviridae* family is composed of metaviruses (LTR retrotransposons) and pro-retroviruses (LTR retroelements with *env* gene but with no evidence of an infective behavior). Originally, the distinct elements belonging to the *Metaviridae* family were classified into three major genera: *Errantivirus* (Ty3/Gypsy elements with *env* gene), *Metavirus* (without *env*), and *Semotivirus* (Bel/Pao elements). This classification derives from the early discovery of Ty3-like LTR retrotransposons of fungi and their Gypsy-like Env-encoding relatives in drosophilids, followed by the characterization of other elements (including Bel/Pao elements) in multiple other eukaryotes. More recently, Bel/Pao elements (genus *Semotivirus*) were removed from the family *Metaviridae* because they do not form a monophyletic group with other *Metaviridae* members. The family *Metaviridae* needs further revision because the diversity of Ty3/Gypsy elements is much greater and more complex than previously thought. The genus *Errantivirus* includes Gypsy-like viruses of drosophila, which as shown in Fig. 2, forms a monophyletic clade splitting into at least two sub-clades (denoted as Gypsy and 17.6). Other Ty3/Gypsy elements form distinct clades, including Athila, Osvaldo and Tor4a, and encode for a third ORF showing structural features typical of Env polypeptides, such as the large Surface (SU) domain providing antigenic sites rich in cysteines, the transmembrane (TM) domain near their carboxyl terminus, multiple putative N-glycosylation sites, and putative protease cleavage sites at conserved positions. These lineages are pro-retroviruses because they carry an *env* gene, albeit they do not cluster phylogenetically with errantiviruses but more properly with other Ty3/Gypsy lineages of LTR retrotransposons (thus, contradicting the errantivirus-metavirus classification). This is for instance the case of Athila clade of pro-retroviruses of plants that are a sister clade of LTR retrotransposons called Tat. Athila and Tat are supported as a monophyletic branch in all Ty3/Gypsy phylogenies (although Athila is a pro-retrovirus and Tat a metavirus). On the other hand, the phylogenetic relationships of all the Ty3/Gypsy lineages originally considered to constitute the genus *Metavirus* cannot be resolved based on current data due to the high sequence divergence in that family. Moreover, members of one of the largest and best supported branches of Ty3/Gypsy LTR retrotransposons, labeled *Chromovirus* in Fig. 2, differ from other Ty3/Gypsy elements in that (with very few exceptions, such as Ty3, TF1 and TF2) their INTs have a conserved chromodomain at the C-terminus. This branch of chromodomain-Ty3/Gypsy LTR retrotransposons are formally classified as representative members of the genus *Metavirus* but is widely referred to as the genus “*Chromovirus*” (because of their chromodomain INTs) and splits into at least three major branches widely distributed in plants, fungi and vertebrates herein referred to as “Plants”, “Algae” and “Fungi/Vertebrates” according to the host distribution of the corresponding elements. Another lineage traditionally classified within the genus *Metavirus* includes Mag – a large branch collecting different clades of Ty3/Gypsy LTR retrotransposons of metazoans. The remaining diversity of elements originally assigned to the genus *Metavirus* is distributed across a number of smaller clades (Cer-like, Mdg3/Micropia, 412/Mdg1, Osvaldo/Gmr1, CsRN1, Tor-like and others) (Table 1).

The *Belpaoviridae* family groups all the Bel/Pao LTR retrotransposons and pro-retroviruses of the belpao family formerly classified by the ICTV as the genus *Semotivirus* within the family *Metaviridae* because they present the same *gag-pol* organization as metaviruses (Fig. 1). However, belpaoviruses are significantly divergent on the sequence level from metaviruses and they constitute a well supported monophyletic branch that falls outside of the *Metaviridae* family. Taking this into consideration, the ICTV has recently moved the *Semotivirus* genus from the family *Metaviridae* into a separate family, the *Belpaoviridae*, within the order *Ortervirales*. Based on current data, belpaoviruses are less diverse than their metavirus counterparts, albeit their currently known diversity is also greater than previously supposed. Based on the Pol tree shown in Fig. 2, five clades can be defined: Pao, Sinbad, Bel, Tas, and Suzu. Of these, Pao, Sinbad, and Suzu clades group together in a well supported branch. The relationship of this branch with the two other clades (Bel and Tas) remains unresolved. Bel and Tas clades mainly include LTR retrotransposons. However, within Tas clade, some elements, such as Tas (the element that gives the clade its name), carry an ORF encoding for a putative Env polypeptide, suggesting that at least some belpaoviruses might be able to form infectious virions.

The *Pseudoviridae* family is also highly diverse and includes all Ty1/Copia LTR retrotransposons and pro-retroviruses. At present, the ICTV recognizes three genera within the *Pseudoviridae* – *Pseudovirus*, *Hemivirus*, and *Sirevirus*. Although these three genera are well supported in the current phylogeny, they are not sufficient to cover the whole diversity of Ty1/Copia LTR retroelements. Based

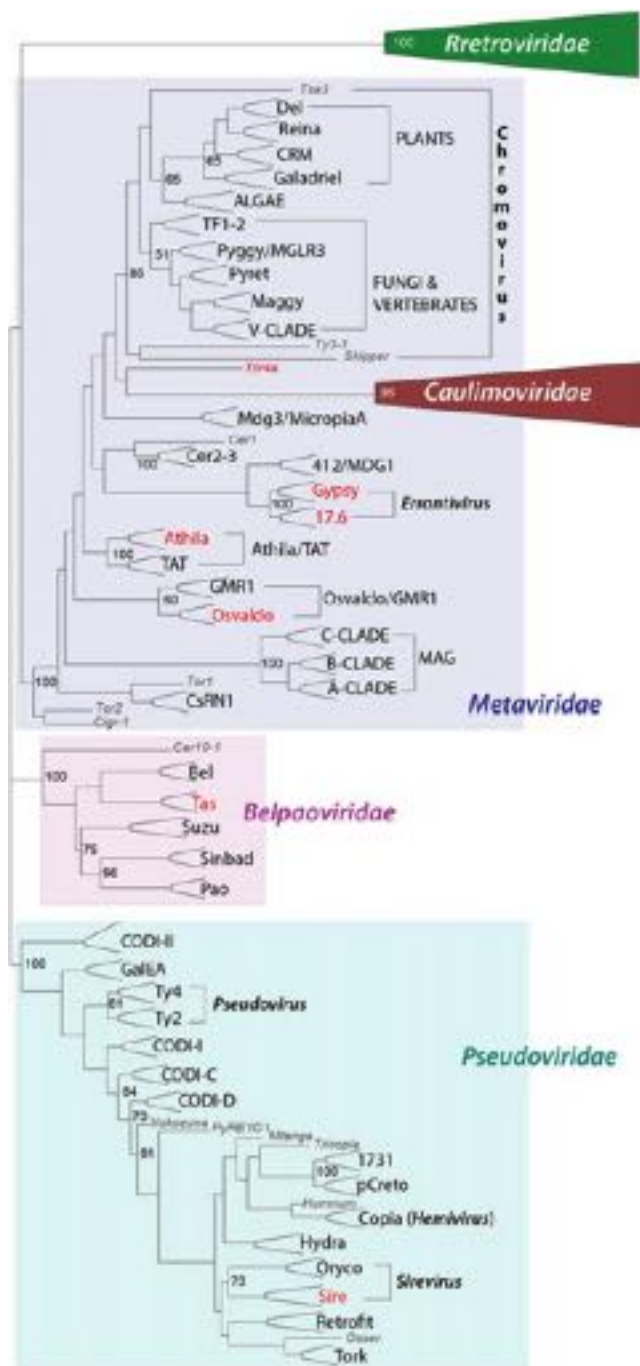


Fig. 2 Phylogeny of members of the order *Ortervirales* based on an alignment concatenation of the most conserved part or cores of the four pol polyprotein domains (PR-RT-RH-INT) of 268 OTUs belonging to the *Retroviridae*, *Caulimoviridae*, *Metaviridae*, *Pseudoviridae*, and *Belpaoviridae* families. OTUs are collapsed in clades except for those elements providing clusters of a single OTU, which are represented with their name in gray and italicized. Families and genera currently recognized by the ICTV are provided in bold italics. The absence of an INT domain in the alignment is represented with a gap in the pol alignment. For simplicity, bootstrap information is provided only for the most internal nodes of the tree and only when bootstrap values are higher than 50. Families collecting true or potential retroviruses or LTR retrotransposons are highlighted in red. The figure is adapted from Fig. 1 in Llorens C., *et al.* 2009. The full expanded version of this tree in web format is available as Supplemental file 1 in Llorens C., *et al.*, 2009. The web version of this tree provides information about names, Genbank accessions and hosts of all LTR retroelement taxa used. By clicking on each OTU in the pol tree, the reader can access the accession of the requested element from GyDB or NCBI, respectively. Although our current knowledge about the diversity of *Ortervirales* is still incomplete, this Pol tree provides an acceptable and realistic perspective on their evolutionary history, mainly because of the strong phylogenetic signal of the RT, which is the pol domain most frequently used to infer retroelement phylogenies in general. With reasonable variations due to differences in rates of evolution, this evolutionary history is also supported by gag (the other protein domain common to all *Ortervirales*) and by other proteins encoded by these LTR retroelements. Readers interested in more specific details about the evolutionary history of the order *Ortervirales* can visit GyDB and access a comprehensive collection of phylogenetic trees inferred based on all encoded products of these LTR retroelements. The figure is adapted from Fig. 1 in Llorens, C., Muñoz-Pomer, A., Bernad, L., Botella, H., Moya, A., 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biology Direct* 4, 41. The full expanded version of this tree in web format is also provided as Supplemental file 1 in that article.

Table 1 Members in the families *Belpaoviridae*, *Metaviridae*, and *Pseudoviridae* (order *Ortervirales*). Genera type species are written in bold

Family/Genus /Phylogenetic GyDB clade ^a	Species	Exemplar virus	Acronym	Literature name	GenBank accession
<i>Belpaoviridae</i>					
<i>Semotivirus</i>					
Tas	<i>Ascaris lumbricoides</i> Tas virus	Ascaris lumbricoides Tas virus	AluTasV	Tas	Z29712
Pao	<i>Antheraea</i> Tamy virus	Antheraea mylitta Tamy virus	TamyV	Tamy	AF530470
	<i>Bombyx mori</i> Pao virus	Bombyx mori Pao virus	BmoPaoV	Pao	L09635
	<i>Drosophila simulans</i> Ninja virus	Drosophila simulans Ninja virus	DsiNinV	Ninja	D83207
Bel	<i>Drosophila melanogaster</i> Bel virus	Drosophila melanogaster Bel virus	DmeBelV	Bel	U23420
	<i>Drosophila melanogaster</i> Roo virus	Drosophila melanogaster Roo virus	DmeRooV	Roo	AY180917
	<i>Drosophila melanogaster</i> Max virus	Drosophila melanogaster Max virus	DmeMaxV	Max	AJ487856
	<i>Anopheles gambiae</i> Moose virus	Anopheles gambiae Moose virus	AgaMooV	Moose	AF060859
Sinbad	<i>Schistosoma</i> Sinbad virus	Schistosoma mansoni Sinbad virus		Sinbad	AY506538
Suzu	<i>Takifugu rubripes</i> Suzu virus	Fugu rubripes Suzu virus	FruSuzV	Suzu	AF537216
Undefined at GyDB	<i>Caenorhabditis elegans</i> Cer13 virus	Caenorhabditis elegans Cer13 virus	CelCer13V	Cer13	Z81510
<i>Metaviridae</i>					
<i>Errantivirus</i>					
Gypsy	<i>Ceratitis capitata</i> Yoyo virus	Ceratitis capitata Yoyo virus	CcaYoyV	Yoyo	U60529
	<i>Drosophila melanogaster</i> Gypsy virus	Drosophila melanogaster Gypsy virus	DmeGypV	Gypsy	M12927
17.6	<i>Drosophila melanogaster</i> 17–6 virus	Drosophila melanogaster 17.6 virus	Dme176V	17.6	X01472
	<i>Drosophila melanogaster</i> 297 virus	Drosophila melanogaster 297 virus	Dme297V	297	X03431
	<i>Drosophila melanogaster</i> Idefix virus	Drosophila melanogaster Idefix virus	DmeldeV	Idefix	AJ009736
	<i>Drosophila ananassae</i> Tom virus	Drosophila ananassae Tom virus	DanTomV	Tom	Z24451
	<i>Drosophila melanogaster</i> Zam virus	Drosophila melanogaster Zam virus	DmeZamV	Zam	AJ000387
	<i>Drosophila virilis</i> Tv1 virus	Drosophila virilis Tv1 virus	DviTv1V	Tv1	AF056940
	<i>Trichoplusia ni</i> TED virus	Trichoplusia ni TED virus	TniTedV	Ted	M32662
Undefined At GyDB	<i>Drosophila melanogaster</i> Tirant virus	Drosophila melanogaster Tirant virus	DmeTirV	Tirant	X93507
<i>Metavirus</i>					
Athila/Tat	<i>Arabidopsis thaliana</i> Athila virus	Arabidopsis thaliana Athila virus	AthAthV	Athila	AC007209
	<i>Arabidopsis thaliana</i> Tat4 virus	Arabidopsis thaliana Tat4 virus	AthTatV	Tat	AB005247
Mag	<i>Bombyx mori</i> Mag virus	Bombyx mori Mag virus	BmoMagV	Mag	X17219
	<i>Tripeustis gratilla</i> SURL virus	Tripeustis gratilla SURL virus	TgrSurV	SURL	M75723
Cer1	<i>Caenorhabditis elegans</i> Cer1 virus	Caenorhabditis elegans Cer1 virus	CelCer1V	Cer1	U15406
Oswaldo	<i>Drosophila virilis</i> Ulysses virus	Drosophila virilis Ulysses virus	DviUliV	Ulysses	X56645
	<i>Tribolium castaneum</i> Woot virus	Tribolium castaneum Woot virus	TcaWooV	Woot	U09586
	<i>Drosophila buzzatii</i> Oswaldo virus	Drosophila buzzatii Oswaldo virus	DbuOsvV	Oswaldo	AJ133521
412/Mdg1	<i>Drosophila melanogaster</i> 412 virus	Drosophila melanogaster 412 virus	Dme412V	412	X04132
	<i>Drosophila melanogaster</i> Mdg1 virus	Drosophila melanogaster Mdg1 virus	DmeMdg1V	Mdg1	X59545
Micropia/Mdg3	<i>Drosophila melanogaster</i> Mdg3 virus	Drosophila melanogaster Mdg3 virus	DmeMdg3V	Mdg3	X95908
	<i>Drosophila melanogaster</i> Micropia virus	Drosophila melanogaster Micropia virus	DmeMicV	Micropia	X14037
	<i>Drosophila melanogaster</i> Blastopia virus	Drosophila melanogaster Blastopia virus	DmeBlaV	Blastopia	Z27119
Chromovirus ^b	<i>Fusarium oxysporum</i> Skippy virus	Fusarium oxysporum Skippy virus	FoxSkiV	Skippy	L34658
	<i>Lilium henryi</i> Del1 virus	Lilium henryi Del1 virus	LheDel1V	Del	X13886
	<i>Saccharomyces cerevisiae</i> Ty3 virus	Saccharomyces cerevisiae Ty3 virus	ScerTy3V	Ty3	M34549
	<i>Schizosaccharomyces pombe</i> Tf1 virus	Schizosaccharomyces pombe Tf1 virus	SpoTf1V	Tf1	M38526
	<i>Schizosaccharomyces pombe</i> Tf2 virus	Schizosaccharomyces pombe Tf2 virus	SpoTf2V	Tf2	L10324
	<i>Takifugu rubripes</i> Sushi virus	Takifugu rubripes Sushi virus	TruSusV	Sushi	AF030881
	<i>Dictyostelium discoideum</i> Skipper virus	Dictyostelium discoideum Skipper virus	DdiSkiV	Skipper	AF049230
Undefined At GyDB	<i>Cladosporium fulvum</i> T – 1 virus	Cladosporium fulvum T – 1 virus	CfuT1V	T – 1	Z11866

Table 1 Continued

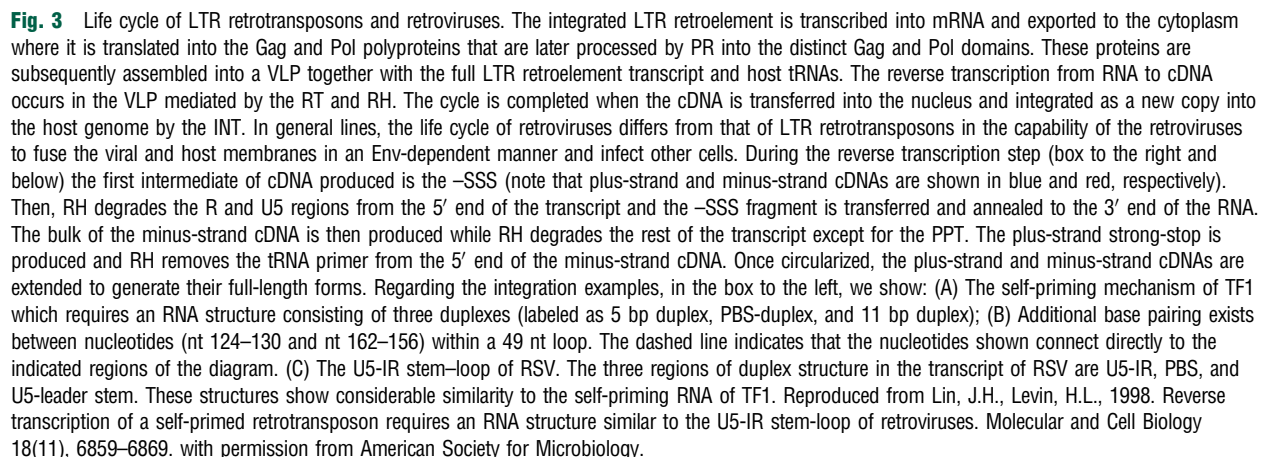
Family/Genus /Phylogenetic GyDB clade ^a	Species	Exemplar virus	Acronym	Literature name	GenBank accession
Pseudoviridae					
Hemivirus					
1731	<i>Drosophila melanogaster</i> 1731 virus	<i>Drosophila melanogaster</i> 1731 virus	Dme1731V	1731	X07656
Copia	<i>Drosophila melanogaster copia</i> virus	<i>Drosophila melanogaster copia</i> virus	DmecopiaV	Copia	X04456
Osser	<i>Volvox carteri</i> Osser virus	<i>Volvox carteri</i> Osser virus	VcaOssV	Osser	X69552
Undefined At GyDB	<i>Aedes aegypti</i> Mosqcopia virus	<i>Aedes aegypti</i> Mosqcopia virus	AaeMosV	Mosqcopia	AF134899
	<i>Candida albicans</i> Tca2 virus	<i>Candida albicans</i> Tca2 virus	CalTca2V	Tca2	AF050215
	<i>Saccharomyces cerevisiae</i> Ty5 virus	<i>Saccharomyces cerevisiae</i> Ty5 virus	SceTy5V	Ty5	U19263
	<i>Volvox carteri</i> Lueckenbuesser virus	<i>Volvox carteri</i> Lueckenbuesser virus	VcaLeuV	Lueckenbuesser	U90320
	<i>Candida albicans</i> Tca5 virus	<i>Candida albicans</i> Tca5 virus	CalTca5V	Tca5	AF065434
Pseudovirus					
Tork	<i>Zea mays</i> Sto4 virus	<i>Zea mays</i> Sto — 4 virus	ZmaSto4V	Sto — 4	AF082133
	<i>Nicotiana tabacum</i> Tto1 virus	<i>Nicotiana tabacum</i> Tto1 virus	NtaTto1V	Tto1	D83003
Pseudovirus	<i>Saccharomyces cerevisiae</i> Ty1 virus	<i>Saccharomyces cerevisiae</i> Ty1 virus	SceTy1V	Ty1	M18706
	<i>Saccharomyces cerevisiae</i> Ty2 virus	<i>Saccharomyces cerevisiae</i> Ty2 virus	SceTy2V	Ty2	X03840
	<i>Saccharomyces cerevisiae</i> Ty4 virus	<i>Saccharomyces cerevisiae</i> Ty4 virus	SceTy4V	Ty4	M94164
Retrofit	<i>Brassica oleracea</i> Melmoth virus	<i>Brassica oleracea</i> Melmoth virus	BolMelV	Melmoth	Y12321
	<i>Oryza longistaminata</i> Retrofit virus	<i>Oryza longistaminata</i> Retrofit virus	OloRetrofitV	REetrofit	AH005614
	<i>Zea mays</i> Hopscotch virus	<i>Zea mays</i> Hopscotch virus	ZmaHopV	Hopscotch	U12626
Undefined At GyDB	<i>Cajanus cajan</i> Panzee virus	<i>Cajanus cajan</i> Panzee virus	CcaPanV	Panzee	AJ000893
	<i>Glycine max</i> Tgmr virus	<i>Glycine max</i> Tgmr virus	GmaTgmrV	Tgmr	U96748
	<i>Hordeum vulgare</i> BARE — 1 virus	<i>Hordeum vulgare</i> BARE — 1 virus	HvuBar1V	BARE — 1	Z17327
	<i>Nicotiana tabacum</i> Tnt1 virus	<i>Nicotiana tabacum</i> Tnt1 virus	NtaTnt1V	Tnt1	X13777
	<i>Arabidopsis thaliana</i> Art1 virus	<i>Arabidopsis thaliana</i> Art1 virus	AthArt1V	Art1	Y08010
	<i>Arabidopsis thaliana</i> AtRE1 virus	<i>Arabidopsis thaliana</i> AtRE1 virus	AthAtRE1V	atRE1	AB021265
	<i>Arabidopsis thaliana</i> evelknievel virus	<i>Arabidopsis thaliana</i> evelknievel virus	AthEveV	Evelknievel	AF039373
	<i>Arabidopsis thaliana</i> Ta1 virus	<i>Arabidopsis thaliana</i> Ta1 virus	AthTA1V	Ta1	X13291
	<i>Oryza australiensis</i> RIRE1 virus	<i>Oryza australiensis</i> RIRE1 virus	OauRIRE1V	RIRE1	D85597
	<i>Physarum polycephalum</i> Tp1 virus	<i>Physarum polycephalum</i> Tp1 virus	PpoTp1V	Tp1	X53558
	<i>Solanum tuberosum</i> Tst1 virus	<i>Solanum tuberosum</i> Tst1 virus	StuTst1V	Tst1	X52387
	<i>Triticum aestivum</i> WIS2 virus	<i>Triticum aestivum</i> WIS — 2 virus	TaeWis2V	Wis-2-1A	X63184
Sirevirus					
Sire	<i>Arabidopsis thaliana</i> Endovir virus	<i>Arabidopsis thaliana</i> Endovir virus	AthEndV	Endovir	AY016208
	<i>Glycine max</i> SIRE1 virus	<i>Glycine max</i> SIRE1 virus	GmaSIRV	SIRE1	AF053008
	<i>Lycopersicon esculentum</i> ToRTL1 virus	<i>Lycopersicon esculentum</i> ToRTL1 virus	LesToRV	TorRTL1	U68072
	<i>Zea mays</i> Opie2 virus	<i>Zea mays</i> Opie — 2 virus	ZmaOp2V	Opie — 2	U68408
Undefined At GyDB	<i>Zea mays</i> Prem2 virus	<i>Zea mays</i> Prem — 2 virus	ZmaPr2V	Prem — 2	U41000
Unassigned					
Undefined At GyDB	<i>Phaseolus vulgaris</i> Tpv2-6 virus	<i>Phaseolus vulgaris</i> Tpv2-6 virus	PvuTpvV	Tpv2-6	AJ005762

^aGyDB [<http://gydb.org>].^bThe clade "Chromovirus" is often considered a different genus.

on the Pol tree shown in Fig. 2, Ty4 and Ty2 clades form a branch which is representative of the *Pseudovirus* genus, Copia clade represents the genus *Hemivirus*, whereas Oryco and Sire clades form a group constituting the *Sirevirus* genus. The remaining lineages constitute small clades whose relationships remain unresolved. Finally, the clades designated as CoDi-like include four different Ty1/Copia lineages described in the diatom *Phaeodactylum tricornutum*. It is worth mentioning that Ty1/Copia elements in the CoDi-I clade, similar to Ty3/Gypsy chromoviruses, encode INTs containing a chromodomain at their C-terminus.

Life Cycle and Replication

The life cycle of all LTR retroelements consists of four steps (Fig. 3): (1) transcription and protein synthesis, (2) RNA packaging and VLP formation, (3) reverse transcription, and (4) integration. In the first step, integrated copies of the LTR retroelement are transcribed by the cellular RNA polymerase II. The transcription initiates from a promoter located in the 5' LTR and then, the full-



retroelement mRNA is translated into the Gag and Pol polyproteins that are proteolytically processed by PR to release mature CP, NC, INT, and RT-RH proteins (Pol can be expressed as a translational fusion with Gag or through one or more ribosomal frameshifting).

During the second step of RNA packaging and VLP formation, translated Pol peptides and two copies of full-length mRNA are typically encapsidated together with a specific cellular tRNA involved in priming into a VLP assembled using Gag peptides. The CP oligomerizes to form the immature capsid, while NC packages the genome. The initial VLP assembly is driven by the uncleaved Gag polyprotein, ensuring that the full-length retroelement mRNA is recruited and encapsidated, allowing the Gag-RNA-Gag interactions, contributing to the assembly of an immature Gag/ribonucleoprotein (Gag-RNP) particle. The immature VLP is subsequently matured by proteolytic cleavage of PR between CP and NC domains. Current evidence suggests that the VLPs assembled by metaviruses, pseudoviruses and belpaoviruses might not be infectious when applied extracellularly (except for errantiviruses) but we cannot dismiss further evidence to change this interpretation. Note for instance that the vast majority of fungal RNA viruses are not infectious when applied extracellularly. Yet, they are still considered viruses. The capsids of LTR retroelements play a protective role against defense systems of the host cell during the subsequent steps of the life cycle. Note that in contrast to retroviruses which are released from the host cell to the environment, LTR retrotransposons do not have an extracellular stage in their life cycle, similar to most fungal and some plant viruses. However, high-resolution structures of recently determined immature and mature Ty3 VLPs have shown a close structural similarity between the capsids of metaviruses and retroviruses. Based on the sequence conservation of the CP and NC domains, this similarity can be also extended to all other members of the order *Ortervirales*.

In the third step of reverse transcription, the full-length retroelement mRNA serves as a template for the synthesis of cDNA within the VLP. As also shown in Fig. 3, the cellular tRNA (packaged in the VLP) anneals to the retroelement mRNA at the PBS region that is complementary to the 3' end of the tRNA and is used by RT as a primer to synthesize the (–) DNA chain complementary to the R-U5 zone of 5' LTR. The tRNA primes the reverse transcription of a short minus-strand species of DNA referred to as the minus-strand-strong-stop (–SSS). As this cDNA is being synthesized, RH degrades the RNA template once it is reverse transcribed. Degradation of the 5' end of the RNA releases the –SSS from the 5' end of the RNA and allows it to anneal to the R sequence at the 3' end of the same RNA or other co-packaged genome. This positions the –SSS to be extended by RT to generate the minus-strand of the LTR-retroelement. The remainder of the RNA is degraded by RH with the exception of the PPT adjacent to the 3' LTR, sequence that confers to RNA fragment resistance to RH cleavage. This 3' PPT RNA serves as the plus-strand primer for reverse transcription. Plus strand cDNA is synthesized until 5'-end of the minus strand, and the PBS sequence of tRNA is employed as a template before producing a plus strand strong-stop DNA (+SSS). The +SSS is transferred to the 3' end of the minus strand previously generated and then, the tRNA primer is removed by the RH activity. The complementary sequences of PBS at the 3' ends of the +SSS and of the minus strand generated mate with each other and a circular structure is formed. Each strand serves as a template for RT extension until the double strand DNA is full-length synthesized and LTR ends are duplicated.

During minus strand synthesis most recombination events take place, when the –SSS fragment is transferred from 5' end of an RNA to R sequence at 3' end of a genetically distinct RNA. This process, together with several mutations accumulated by the low fidelity of RT, promote viral evolution and confer to the virus a better capacity to escape from host system. In contrast, other non beneficial mutations for the virus are also produced. For example, in retrovirus such as HIV several mutations in PPT sequence have been observed, what has been correlated to low infectious levels.

It is worth returning to the priming mechanism to highlight that the tRNA complementarity of the PBS varies within each family and depending on the element. For instance, the PBS for errantiviruses is predicted to normally be complementary to a tRNA^{Ser}, with a notable exception of the Gypsy element that presents a PBS with complementarity to a tRNA^{Lys}. In contrast, plant chromoviruses normally use a tRNA^{Met} as a primer, whereas chromoviruses of fungi and vertebrates use a selfpriming mechanism (also depicted in Fig. 3), with some exceptions such as Ty3 which is though to be primed by the tRNA^{Met}. The self-priming mechanism of fungal and vertebrate chromoviruses has been extensively studied in elements such as TF1. According to this model, the first 11 nt of the mRNA are complementary to its PBS and act as a primer of the –SSS cDNA synthesis.

The fourth step of integration takes place when the reverse transcription is completed (as also illustrated in Fig. 3(A)). During this step, the double-stranded cDNA is imported from the VLP into the nucleus and is inserted into a chromosomal target site, usually generating a 5 bp duplication of host target DNA. Integration is catalyzed by the retroelement INT (which belong to the superfamily of DDE INTs) in two reactions: 3' processing and strand transfer. During the 3' processing reaction, the INT removes two or three nucleotides from the 3' ends of the cDNA to position the highly conserved CP dinucleotide at the 3' end of the cDNA. In retroviruses and some LTR-retrotransposons, the tRNA primers initiate reverse transcription two or three nucleotides from the LTR and this results in the addition of deoxynucleotides at the cDNA termini that the INT subsequently removes during this reaction. In other cases, such as the TF1 LTR retrotransposons, the minus-strand primer initiates reverse transcription with no extra nucleotides added before the LTR.

With respect to the strand transfer, reverse-transcribing viruses seem to direct the integration of their genome in specific locations, suggesting that it is not a random process. This event could be in benefit of viruses, when cDNA integration takes place into particular active genes. This leads to viral clonal expansion and persistence in the infected cells. Active genes are usually located in the nuclear periphery where chromatin regions are in an open conformation. Activity of integrated LTR promoters depends on the open chromatin to make it accessible for the binding of the different transcription factors. On the other hand, integration of cDNA is able to disrupt or alter gene functions, with potential detrimental effects on the viability of host cells and, by extension, viability of the inserted LTR retroelement (which depends entirely on the well-being of the host cell). However, in

the course of evolution, LTR retroelements have developed mechanisms to specifically target integration without altering the gene functions. This is primarily achieved by integration into noncoding regions, preferential targeting of heterochromatin regions (not permissive for transcription) or by association with centromeric regions. For example, Ty1 and Ty3 retrotransposons of *S. cerevisiae* integrate near sites of RNA polymerase III (pol III) transcription by recognizing pol III transcription complexes or chromatin states associated with pol III transcription. By contrast, TF1 recognizes certain RNA polymerase II (pol II) promoters. The great majority of Gypsy elements are located in the pericentromeric regions of the *D. melanogaster* genome and this location is conserved in strains from different geographic origins thus suggesting that a high number of Gypsy elements are non-mobile and were already present in the pericentromeric sequences of the common ancestor of the *D. melanogaster* strains. The accumulation of metaviruses such as Athila/Tat and chromoviruses in centromeric sequences is also observed in plants. Recent studies have shown that the chromodomain specifies the target site preference of chromoviral LTR retrotransposons by the recognition of characteristic chromatin modifications. These mechanisms for silencing of the viral genomes are also used to contribute to viral latency and to escape from the immune system. For instance, human immunodeficiency virus (HIV), a retrovirus that targets integration into heterochromatin regions, makes association with alphoid pericentromeric regions in different chromosomes, or integrates its genome in an orientation opposite to the host gene transcriptional readthrough.

Host Range and Evolutionary Dynamics

Phylogenies tend to separate metaviruses, pseudoviruses, belpaoviruses, caulimoviruses, and retroviruses into five well defined phylogenetic clusters to which we refer as families within the order *Ortervirales*. Such perspective is consistent with the idea of gradual evolution and, although accurate, it is based solely on the phylogenetic signal coming from the gag and Pol polyproteins and, inherently, does not consider other biological and evolutionary phenomena that probably played a relevant role in shaping the diversity of the *Ortervirales*. These include: (1) modularity, the distinct genes and/or genomic regions of a virus/LTR retrotransposon usually present different rates of evolution, (2) chimerism, different events of gene recruitment, genome rearrangement and recombination that can be traced in the evolution of the *Ortervirales* at different evolutionary times, (3) alternance of horizontal and vertical means of transmission; while eukaryotes usually transmit their LTR retroelement communities via germ lines (i.e., by vertical means), exogenous retroviruses and caulimoviruses spread via infection (i.e., horizontally) and there is evidence that several LTR retrotransposons might have been able to protagonize some episodes of horizontal transmission, (4) clonal interference and competition with other viruses and other self genetic elements, and (5) co-evolution with eukaryotes; different cellular and molecular host environments that arose at distinct evolutionary times in plants, fungi and animals have forced the distinct orterviral communities to adapt depending on the changes in host environment. All these considerations lead to a scenario of reticulate evolution suggesting that our comprehension of the history of the *Ortervirales* is far from complete and could be more accurately traced as a phylogenetic network shaped by a range of changes in the cellular and molecular environments provided by eukaryotes to their retrotransposable parasites that reciprocally were also key for eukaryotic evolution. In Fig. 4, we superimpose the Pol tree previously shown in Fig. 2 over the evolutionary history of eukaryotes. For simplicity, we divide the resulting tree into three major temporal ranges to provide the reader with a system biology macroevolutionary perspective of the origin and evolution of members of the order *Ortervirales* undergoing three major evolutionary transitions in their eukarotic hosts.

The first transition covers events from the earliest eubacterial organisms around 3500 million years ago (MYA) and the first unicellular algae to the diversification of eukaryotes into the major supergroups (approximately 2170–1950 MYA) between the Archean and the Proterozoic periods (as shown in Fig. 4). Current sequencing projects show that *Metaviridae* chromoviruses are present in multiple distinct algae, amoebae, chromalveolates, fungi, plants and animals. The most parsimonious hypothesis for such a wide distribution is thus that early members of the family *Metaviridae* have already existed before the split between plants and unikonts (1550 MYA) and that chromoviruses might constitute the oldest branch within this family. Sequencing projects also show pseudoviruses to be widespread among algae, plants, fungi and animals, suggesting that pseudoviruses co-existed with metaviruses during this transition. Thus, the ultimate origins of the ortervirads likely took place at the onset of eukaryogenesis and involved combination of gag and pol genes with the LTR features to raise an ancestral LTR retrotransposons from which pseudoviruses and metaviruses diverged into two independent but stable lineages following the genomic relocation of the *int* gene that distinguishes the two families.

The second transition covers a period of 1330–380 MYA from the Proterozoic to the Phanerozoic, during which eukaryotes underwent key events of diversification, such as the divergence between primitive animal phyla (e.g., Porifera, Cnidaria, Ctenophora) and the more complex animals, including the emergence of eumetazoans, the Cambrian explosion, the divergence between arthropods and priapulids, the rise of vertebrates, the emergence of land plants and the establishment of plant-fungal interactions. Respecting the *Metaviridae*, although chromoviruses are present in vertebrates, such as fish, in amphibians and in amniotes, they are absent in the known genomes of cnidarians, mollusks, protostomes and basal deuterostomes. However, we find others metavirus lineages. For example, Mag and other elements, from which we have apparently emerged before the split of bilaterians into protostomes and deuterostomes (761–531 MYA). It is currently unclear whether chromoviruses were driven to extinction in protostomes, echinoderms and urochordates or were horizontally transmitted from plants or fungi to vertebrates. Horizontal transmission seems to be a good explanation for vertebrate chromoviruses, taking into account the strong similarity to their fungal relatives, but if so, such an event might be as ancient as the split of osteichthyes into actinopterygians and sarcopterygians

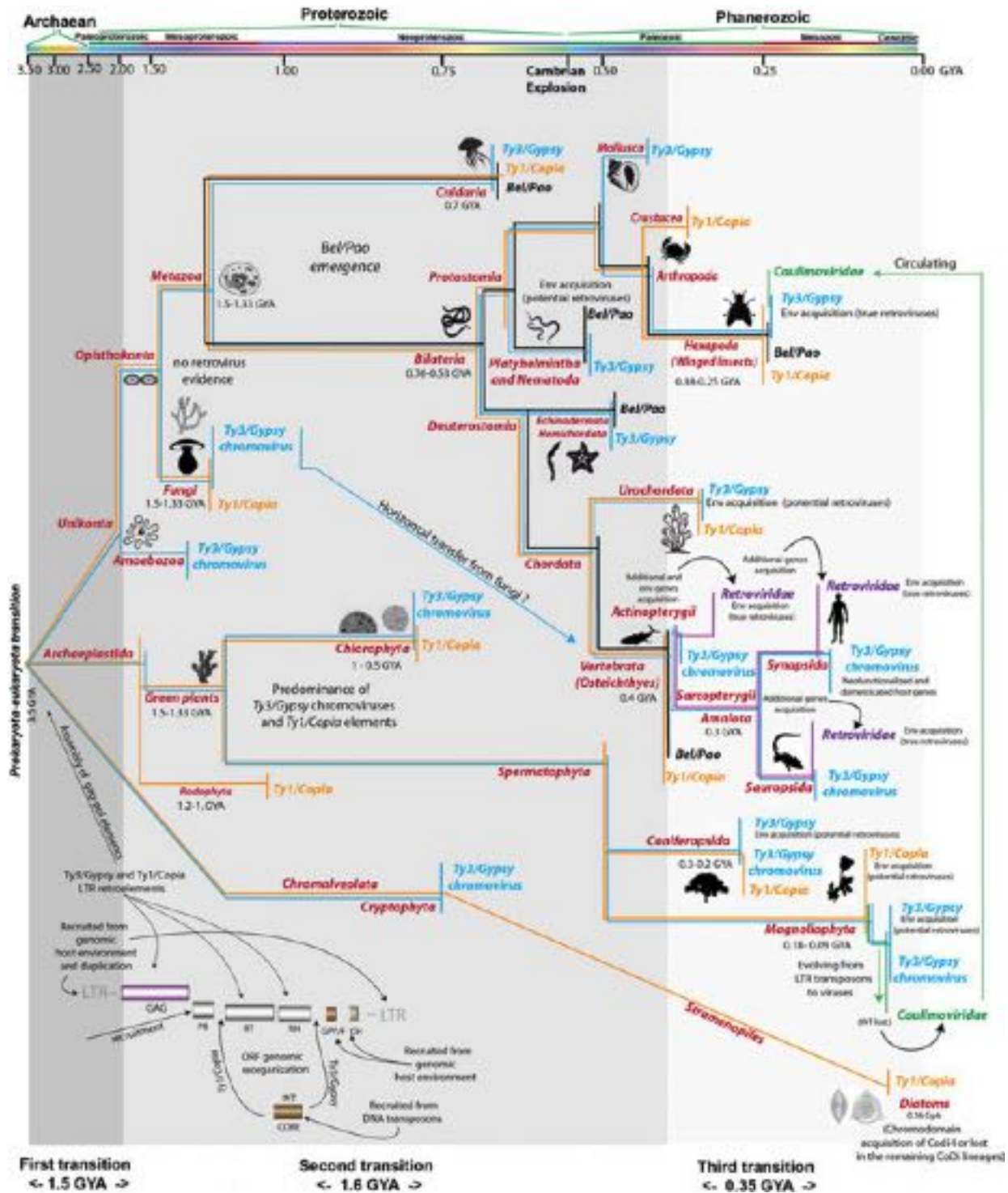


Fig. 4 Evolutionary history of LTR retroelements superimposed over the evolutionary history of eukaryotes. The tree of Ty3/Gypsy LTR retroelements is traced in blue; Ty1/Copia elements yellow; Bel/Pao black; *Retroviridae* violet; and *Caulimoviridae* green. This representation defines the natural origin and distribution of the *Ortervirales* constrained by the evolutionary history of eukaryotes and also includes information about diverse reticulate events in the evolutionary history of LTR retroelements. Host ranges are indicated in red. The framework has been divided into three major eukaryotic transitions (colored in gradient of gray) to interpret the distribution of the distinct LTR retroelement families and communities on the basis of the most likely ages of their hosts. GYA refers to billions (giga) years ago. The distinct evolutionary paths traced in this figure only reflect the host distributions of the distinct orterviral families and their likely times in those hosts, not extinction events. For example, if a host range is represented in the second transition but not in the third one, it indicates that that orterviral family is inhabiting the genomes of those hosts from that age to the present time. Adapted from Llorens, C., Muñoz-Pomer, A., Bernad, L., Botella, H., Moya, A., 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biology Direct* 4, 41. with permission from BioMed Central, part of Springer Nature.

(410–415 MYA). Regarding the *Pseudoviridae*, sequencing projects reveal pseudoviruses to be also widespread in cnidarians, arthropods and deuterostomes. However, to our knowledge, there is no evidence of pseudoviruses in platyhelminthes and nematodes, suggesting that metaviruses underwent more explosive diversification in metazoans compared to pseudoviruses. As for the *Belpaoviridae*, their wide distribution in metazoans and their absence in fungi and plants suggests that the origin of belpaoviruses probably predates the split between primitive animals and higher eumetazoans (1300–380 MYA). With respect to the emergence of *bona fide* retroviruses (family *Retroviridae*), sequencing projects reveal three *Retroviridae* lineages including gammaretroviruses, epsilonretroviruses and spumaretroviruses to overlap with chromoviruses in fishes, amphibians and sauropsids thus suggesting that the first *Retroviridae* emerged in marine vertebrates before the Osteichthyes split into the actinopterygians and the sarcopterygians. We can thus assume that metaviruses and pseudoviruses have radiated into a variety of lineages, with belpaoviruses and the *Retroviridae* emerging during this second transition.

The third eukaryotic transition (range in light gray within Fig. 4) covers a period from the Paleozoic era (330 MYA) to the present, a timeframe which included the origin of first gymnosperms (360–248 MYA), the split of amniotes into sauropsids and synapsids (330–312 MYA), the diversification of winged insects (380–325 MYA) and the origin of flowering plants (160–90 MYA). New levels of complexity and ecosystems probably accompanied the aforementioned events in the evolutionary history of eukaryotes, activating parallel radiations of LTR retrotransposons as well as the emergence of new viral and retroviral forms to colonize the new host environments such as those of synapsids (mammals and their extinct relatives). Consistent with this perspective, sequencing projects revealed that metaviruses and retroviruses overlap in sauropsids but there is no trace of functional metaviruses, pseudoviruses or belpaoviruses in mammals. The only orterviral family known to spread in mammals is the family *Retroviridae*, albeit several host genes, most notably including the neuronal gene *Arc* which evolved from the CP domain of domesticated metaviruses, can also be detected in the mammalian genomes. This suggests that while metaviruses were co-opted and driven to extinction by synapsids, the retroviruses co-evolved with their hosts, recruiting new accessory genes to adapt their infective means of transmission to the new and more complex mammalian host environments. In parallel, mammals evolved more sophisticated mechanisms of innate immunity to combat retroviruses. Retrovirus accessory genes and mammalian mechanisms of innate immunity will be best understood when considered as the joint products of macro-evolutionary conflicts played out over a geological scale. During this period, other retroviral forms, such as errantiviruses, also emerged from the *Metaviridae* family in winged insects such as *Drosophila*, albeit the lower biological complexity of insects (in comparison to that of mammals) allowed the diversification of the metaviruses in other lineages as well as the expansion of pseudoviruses and belpaoviruses in these and other insects. A very similar perspective is found in plants. In this particular, gymnosperms and angiosperms show massive radiations of *Athila/Tat*-like metaviruses and LTR retrotransposons to overlap with their chromoviral counterparts and with pseudoviruses which also show some examples of protoretroviral forms, such as some elements belonging to Sire clade. Thus, pro-retroviruses might have emerged in plants before the divergence of gymnosperms and angiosperms. On the other hand, phylogenetic and comparative genomic analyses of *gag-pol* regions also support the evolution of caulimoviruses from the *Metaviridae* (note in Fig. 2 how the *Caulimoviridae* are rooted within the *Metaviridae*). However, the genomes of caulimoviruses additionally present other genes, such as the one encoding for the movement protein, which are necessary for the viral life cycle and transmission of these viruses in plants. While it is not yet clear whether the ancestors of caulimoviruses were inhabitants of plants or insects (the latter act as mechanical vectors), the origin of caulimoviruses was probably concomitant with the emergence of insects and flowering plants (130–90 MYA). These two events probably gave an opportunity to metaviruses to explore other gene combinations and viral strategies from which caulimoviruses probably emerged (recruiting new genes or recombining with other viruses).

Impact and Relationships of LTR retrotransposons and Retroviruses With Their Hosts

Until recently, LTR retrotransposons were considered to be merely molecular parasites whose sole aim is to self-replicate in the host genomes and be transmitted from one host generation to another. Current evidence suggests, however, that the co-evolutionary history of LTR retrotransposons and their hosts is much more complex than previously thought. Retroelements with and without LTRs are mutagens, whose activities might lead to diverse genetic alterations, including: (1) gene knockouts, (2) modifications of gene expression, (3) epigenetic reprogramming, (4) insertions and deletions, (5) structural rearrangements, and (6) segmental duplications. To avoid potential damages to their hosts, LTR retroelements have developed specific mechanisms to integrate the retrotranscribed cDNAs into non-coding genomic regions, as discussed above. The mechanisms ensuring specific insertion of LTR retrotransposons are the basis of a co-evolutionary history that has permitted these elements to successfully proliferate in their eukaryotic hosts. In sharp contrast, the host may benefit from the new insertions of retroelements because they are a constant source of new sequence substrate. In plants, for instance, the metaviruses and the pseudoviruses constitute the major fraction of all plant transposons. This fraction is, in fact, correlated with the genome size in some groups of plants, such as angiosperms, thus providing a potential explanation why the amount of DNA in a haploid genome does not match with the complexity of these taxa (i.e., the C-value paradox). The situation is similar in other plant, fungal and animal taxa across the Tree of Life. In humans, for example, endogenous retroviruses (HERVs) constitute approximately 8% of the human genome. If we also consider the rest of retrotranscribing elements (with and without LTRs) and related sequences, the fraction reaches up to 45% of the human genome.

The contribution of new sequence substrate as a consequence of the mobilization of retrotransposable elements can help the host organism to explore new levels of complexity. Integration of multiple copies of LTR retroelements and other mobile genetic elements into a particular chromosomal region may result in repeat clusters, which are hotspots for chromosomal rearrangements and segmental duplications due to erroneous homologous recombinations during mitosis and meiosis. Understanding these clusters is still lacking but they appear to provide for genome plasticity to their hosts. Notably, several fungal pathogens, such as *Leptosphaeria maculans* or *Verticillium dahlia*, have been able to take advantage of these plastic loci for evolving increased virulence and pathogenesis, and adapt better to changing environments than members of these species lacking the corresponding clusters.

The sequences contributed by retroelements as a consequence of their mobilization is not only a source of new substrate but also of variability from which new genetic combinations and functions may emerge and be co-opted by the host, giving rise to new genes and regulatory elements, able to rewire the gene expression networks of eukaryotes. In parallel, a functional equilibrium must exist to regulate the activities of retrotransposable elements, to avoid the disruption of the normal regulatory mechanisms of gene expression and pathways. The loss of this equilibrium may have severe consequences, leading to developmental disorders and diseases. In humans, for example, the activation and mobilization of HERVs and other reverse-transcribing elements have been associated with cancer, autoimmune diseases, genomic instability and neurodegenerative diseases, including Tau pathology and Alzheimer. In *D. melanogaster*, the mobilization of *Gypsy* and *Copia* elements (*Metaviridae* and *Pseudoviridae*, respectively) has also been correlated with the activity of Tau, rendering *D. melanogaster* an excellent model to explore the molecular mechanisms behind diseases associated not only with neurodegenerative processes but also with tumorigenesis and tumor immunity. The relationship between cancer and retroelements is, in fact, object of current research both in humans and in other taxa. For instance, some reverse-transcribing viruses, such as retroviruses, have shown strong preference to integrate into genes encoding cellular transcription factors involved in cell growth. BACH2 and MKL2 are two of the proteins encoded by the genes where retrovirus integration has been observed. Integration into these genes alters their functions and increases the life-time and survival of the host cell. In particular, rearrangements of these genes have been associated to proto-oncogenic activities. In other taxa, like *Mya arenaria*, a soft-shelled clam characteristic of many areas in the North Atlantic, a fatal leukemia-like disease of unknown origin has been reported to be catalyzed by the expression of *Steamer*, an LTR retrotransposon of the *Metaviridae* family. Therefore, the proper regulation of gene activation and silencing has an extreme importance for the co-existence of LTR retroelements and the hosts they inhabit. To this end, host genomes and transposable elements may both exist in variable DNA methylation states whose differential transcription is modulated by distinct types of regulatory elements, such as enhancers, silencers and chromatin insulators. Of these, chromatin insulators play a central role in the proper regulation of gene expression and organization of chromosome architecture. A well studied example of chromatin insulator in *D. melanogaster* is the insulator of *Gypsy*-like errantiviruses, that is able to modulate enhancers and promoters. Uncontrolled expression of *Gypsy* may result in insertions of this retrovirus with negative consequences on the normal expression patterns of the host genes. To keep *Gypsy* under control, *Drosophila* implements an RNA interference (RNAi) system constituted by Piwi and *flamenco*, a locus that is the source of antisense small RNAs interacting with Piwi to repress the expression and infective properties of *Gypsy*. The repressive role of *flamenco* in *Drosophila* also extends to other metaviral and pseudoviral elements whose expression is also modulated by this system.

Concluding Remarks

Viruses and mobile genetic elements oftentimes represent different faces of the same coin. This is the case of the members of the families *Metaviridae*, *Pseudoviridae*, and *Belpaoviridae* which may exist as LTR retrotransposons and as potential or even true viruses that share origins and evolutionary history with other reverse-transcribing viruses, such as vertebrate retroviruses and plant caulimoviruses. In recognition of this evolutionary connection, the ICTV has recently grouped the *Belpaoviridae*, *Metaviridae*, *Pseudoviridae*, *Retroviridae*, and *Caulimoviridae* families into a single viral order, the *Ortervirales*. In this article, we have reviewed what is currently known about metaviruses, pseudoviruses and belpaoviruses, with the aim to provide the reader with a systems biology perspective on these viruses and discuss the deep evolutionary history of the *Ortervirales*. Although retroviruses and caulimoviruses have been extensively studied, our understanding of the real impact of reverse-transcribing viruses in eukaryotes is still in its infancy. Due to this reason, the members of the families *Metaviridae*, *Pseudoviridae*, and *Belpaoviridae* present excellent opportunities not only to investigate the origin and evolution of reverse-transcribing viruses in plants, fungi and animals but also to investigate their roles as inductors of disease, regulatory elements and vectors of complexity and evolution in plants, fungi and animals.

Further Reading

- Burns, K.H., 2017. Transposable elements in cancer. *Nature Reviews Cancer* 17 (7), 415–424.
- de la Chaux, N., Wagner, A., 2011. BEL/Pao retrotransposons in metazoan genomes. *BMC Evolutionary Biology* 11, 154.
- Dodonova, S.O., Prinz, S., Bilanchone, V., Sandmeyer, S., Briggs, J., 2019. Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. *Proceedings of the National Academy of Sciences of the United States of America* 116 (20), 10048–10057.
- Eickbush, T., Boeke, J.D., Sandmeyer, S.B., Voytas, D.F., 2011. *Metaviridae*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. San Diego: Elsevier Academic Press, pp. 457–466.
- Gao, X., Hou, Y., Ebina, H., Levin, H.L., Voytas, D.F., 2008. Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Research* 18 (3), 359–369.

- Geering, A.D.W., 2014. Caulimoviridae (plant pararetroviruses). eLS. Chichester, United Kingdom: John Wiley & Sons, Ltd.
- Greenig, M., 2019. HERVs, immunity, and autoimmunity: Understanding the connection. *PeerJ* 7, e6711.
- Grzebelus, D., 2018. The functional impact of transposable elements on the diversity of plant genomes. *Diversity* 10 (2), 18.
- Guo, C., Jeong, H.H., Hsieh, Y.C., *et al.*, 2018. Tau activates transposable elements in Alzheimer's disease. *Cell Reports* 23 (10), 2874–2880.
- Havecker, E.R., Gao, X., Voytas, D.F., 2004. The diversity of LTR retrotransposons. *Genome biology* 5 (6), 225.
- Howlett, B.L., Lowe, R.G.T., Marcroft, S.J., van de Wouw, A.P., 2015. Evolution of virulence in fungal plant pathogens: Exploiting fungal genomics to control plant disease. *Mycologia* 107 (3), 441–451.
- Katzourakis, A., Gifford, R.J., Tristem, M., Gilbert, M.T., Pybus, O.G., 2009. Macroevolution of complex retroviruses. *Science* 325 (5947), 1512.
- Krupovic, M., Blomberg, J., Coffin, J.M., *et al.*, 2018. *Ortervirales*: New Virus Order Unifying Five Families of Reverse-Transcribing Viruses. *Journal of Virology* 92 (12), e00515–e00518.
- Krupovic, M., Dolja, V.V., Koonin, E.V., 2019. Origin of viruses: Primordial replicators recruiting capsids from hosts. *Nature Reviews Microbiology* 17 (7), 449–458.
- Krupovic, M., Koonin, E.V., 2017. Homologous capsid proteins testify to the common ancestry of retroviruses, caulimoviruses, pseudoviruses, and metaviruses. *Journal of Virology* 91 (12), e00210–e00217.
- Lin, J.H., Levin, H.L., 1998. Reverse transcription of a self-primed retrotransposon requires an RNA structure similar to the U5-IR stem-loop of retroviruses. *Molecular and Cell Biology* 18 (11), 6859–6869.
- Llorens, C., Muñoz-Pomer, A., Bernad, L., Botella, H., Moya, A., 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biology Direct* 4, 41.
- Metzger, M.J., Reinisch, C., Sherry, J., Goff, S.P., 2015. Horizontal transmission of clonal cancer cells causes leukemia in soft-shell clams. *Cell* 161 (2), 255–263.
- Pelisson, A., Mejlumian, L., Robert, V., Terzian, C., Bucheton, A., 2002. *Drosophila* germline invasion by the endogenous retrovirus gypsy: Involvement of the viral env gene. *Insect Biochemistry and Molecular Biology* 32 (10), 1249–1256.
- Sarot, E., Payen-Groschène, G., Bucheton, A., Péliisson, A., 2004. Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the *Drosophila melanogaster* flamenco gene. *Genetics* 166 (3), 1313–1321.
- Wei, W., Brennan, M.D., 2001. The gypsy insulator can act as a promoter-specific transcriptional stimulator. *Molecular and Cellular Biology* 21 (22), 7714–7720.

Relevant Websites

- <http://gydb.org>
Gypsy Database 2.0.
- <https://www.girinst.org/replibase>
Replibase.GIRI.Genetic Information Research Institute.
- <http://urgi.versailles.inra.fr/repdb>
RepetDB: Home.
INRA-URGI.
- <http://repeatexplorer.org>
RepeatExplorer.
- https://talk.ictvonline.org/ictv-reports/ictv_9th_report/reverse-transcribing-dna-and-rna-viruses-2011/w/rt_viruses
Reverse Transcribing DNA and RNA Viruses.
ICTV 9th Report (2011).

Rice Tungro Disease (*Secoviridae*, *Caulimoviridae*)

Gaurav Kumar, Fauzia Zarreen, and Indranil Dasgupta, University of Delhi, New Delhi, India

© 2021 Elsevier Ltd. All rights reserved.

This is an update of R. Hull, Rice Tungro Disease, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00714-7.

Glossary

Agro-inoculation Infection of a plant by injecting Agrobacterium containing a modified Ti plasmid into which an infectious copy of the viral genome has been inserted. In the case of pararetroviruses, this infectious copy comprises the sequence that can express the more-than-full-length RNA transcript required for the reverse transcription stage of viral replication.

Pararetrovirus A virus that replicates by reverse transcription but differs from retroviruses in that it encapsidates the DNA phase of replication and that virus replication does not require integration of the viral genome into the host chromosome.

RNA Interference (RNAi) RNAi is an evolutionarily conserved sequence-specific natural defense mechanism in plants against transposable elements and intracellular

pathogens such as viruses, targeting the cellular or viral double-stranded RNA (dsRNA). In this process non-coding small RNAs degrade viral and cellular transcripts, resulting in post-transcriptional gene silencing. These small RNAs are formed as a result of cleavage or slicing of dsRNA or RNA having strong secondary structures by host ribonucleases and are called small interfering RNA and microRNA.

Semi-persistent transmission Virus-vector interaction in which the vector acquires the virus rapidly upon feeding and the acquisition period is relatively short, typically less than a week. The virus is believed to interact with components of the vectors' mouthparts or foregut.

Silencing suppressor Viral proteins that counteract host RNAi defense by inhibiting key steps of cellular RNAi pathways.

Introduction

Rice tungro disease is caused by a complex of two viruses, Rice tungro bacilliform virus (RTBV), which is a member of the family *Caulimoviridae* and Rice tungro spherical virus (RTSV), a member of the family *Secoviridae*. The viral complex is transmitted by several species of green leafhoppers in a semi-persistent manner. The genomes of RTBV and RTSV are double-stranded DNA and single-stranded RNA respectively and are replicated using very different mechanisms. Genetic resistance against RTBV has not been characterized, but for RTSV, recessive resistance has been mapped to a gene encoding a translational initiation factor. We are reviewing here the status of knowledge for the disease and for both viruses.

Rice Tungro Disease

The characteristic symptoms of rice tungro disease had been reported from several South-East Asian countries for many years before they were recognized in 1963 to be of viral etiology. During and since the 1960s, there have been several severe outbreaks of the disease mainly associated with the major agronomic improvements in rice cultivation resulting from the 'green revolution'. The disease is found in all countries of South and South-East Asia and in the South and South-East of China (Fig. 1(A)). The importance and distribution of the disease are evidenced by the variety of local names for it: *accep na pula* (Philippines), *mentek* (Indonesia), *penyakit merah* (Malaysia), yellow-orange leaf (Thailand), leaf yellowing (India). In the early 1990s, it was estimated that tungro caused annual losses in excess of \$1.5 billion without taking account of the cost of insecticide to control the vector. In parts of India, the disease causes losses of up to 50% of the expected crop yield.

In most of the rice cultivars, tungro infection is manifested by stunting and reduced tillering, yellow to orange-yellow leaf discoloration beginning from the leaf tip and extending towards leaf base, delay in flowering, failure of panicle emergence and rust-colored patches or twisted leaves in some cases. Infection to a young rice plant is more severe with mottled appearance and interveinal chlorosis as compared to an old rice plant having rust-colored specks of varying sizes. A typical view of a paddy field showing tungro symptoms is shown in Fig. 2. The disease symptoms are often mild or almost absent in tolerant rice varieties. The disease is transmitted in a semi-persistent manner by several leafhopper species with the rice green leafhopper (GLH), *Nephotettix virescens* (Fig. 3), being the most important vector. It is also transmitted by *Nephotettix cincticeps*, *Nephotettix nigropictus*, *Nephotettix malayanus* and *Recilia dorsalis*; these other vectors may be of relative importance in certain situations. The distribution of the disease is circumscribed by the distribution of *N. virescens* (Fig. 1(A)).

It was not until 1978 that it was recognized that tungro is caused by a complex of two viruses, Rice tungro bacilliform virus (RTBV) and Rice tungro spherical virus (RTSV). RTBV alone causes severe disease symptoms in the host but it cannot be transmitted by GLH in the absence of RTSV. RTSV, on the other hand, can be transmitted independently by GLH (Fig. 4) but causes only mild stunting

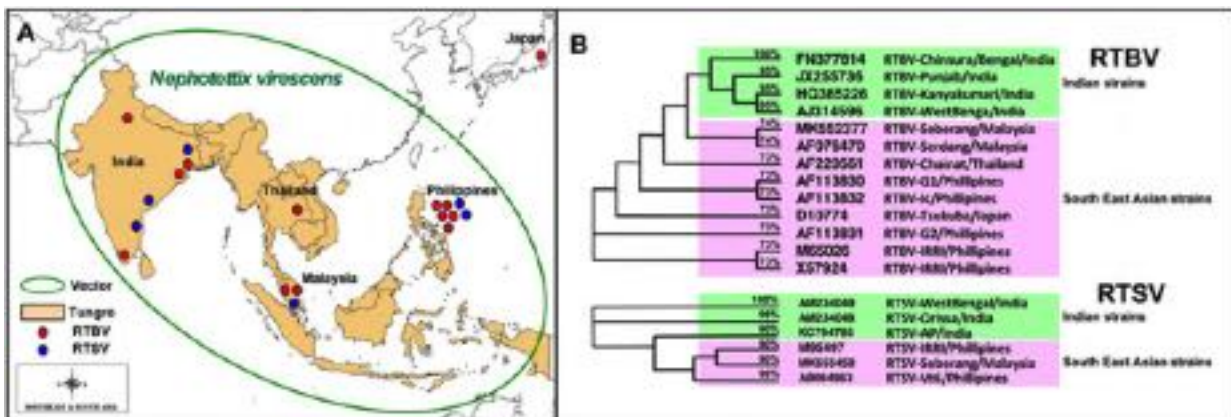


Fig. 1 (A) Distribution of rice tungro disease in Asia and the circle shows the distribution of the main rice tungro disease leafhopper vector. (B) Phylogenetic trees of complete sequences of RTBV (top) and RTSV (bottom). The percentage of identity relative to the first sequence is indicated.



Fig. 2 Field symptoms of rice tungro disease.



Fig. 3 The rice green leafhopper, *Nephrotettix virescens*, major vector of the rice tungro disease.

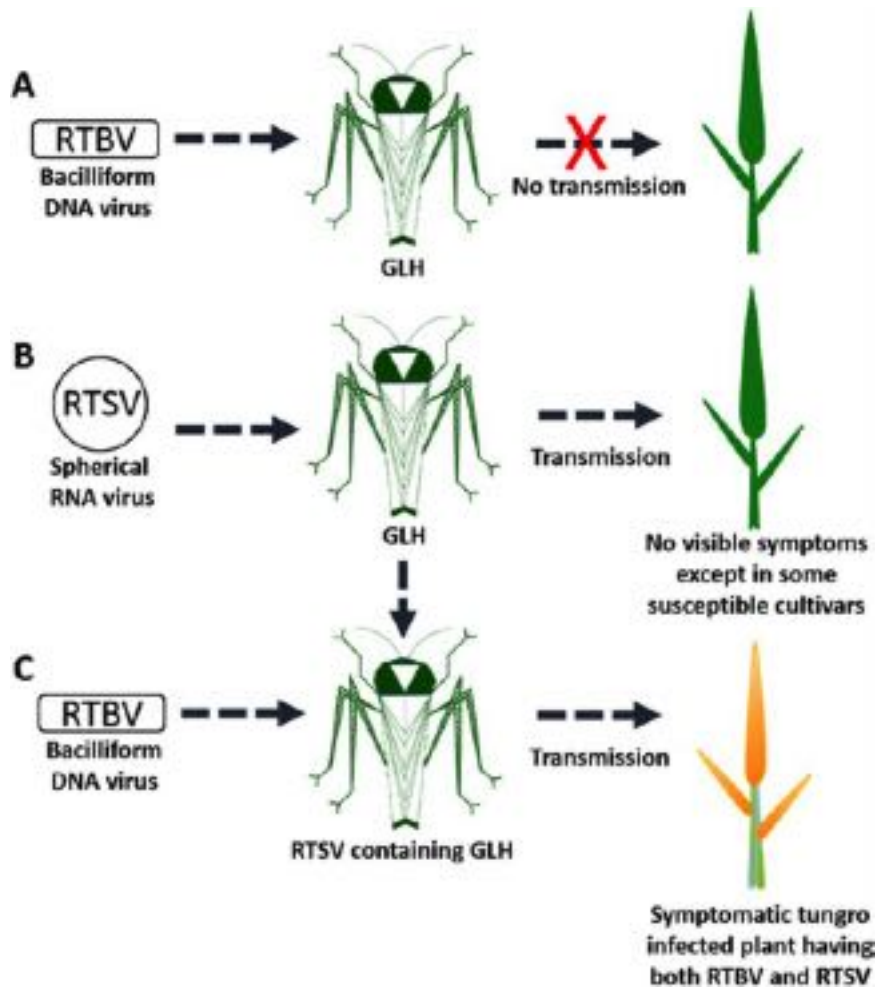


Fig. 4 Rice tungro virus transmission. (A) RTBV cannot be transmitted on its own by GLH; (B) RTSV can be transmitted on its own by GLH but cause very less or negligible symptoms to appear; (C) An RTSV harboring GLH is able to transfer RTBV alone and the presence of both RTBV and RTSV causes visible symptoms.

symptoms. The dependence of RTBV upon RTSV for transmission suggests the presence of a “helper component” in RTSV or RTSV-infected cells. The origin, nature and functional properties of the above “helper component” remains elusive.

Rice Tungro Bacilliform Virus

Taxonomy

RTBV is a dsDNA pararetrovirus belonging to a mono-specific genus, *Tungrovirus*, in the family *Caulimoviridae*.

Transmission and Host Range

GLH transmits RTBV in a semi-persistent manner in rice only in the presence of RTSV. However, cloned RTBV DNA can be transmitted to rice plants if inoculated using agrobacterium, even in the absence of RTSV. This process is known as agro-inoculation. Till now, this method has not been used to investigate the host range of RTBV.

Structure

RTBV consists of bacilliform particles of about 130 nm length (though there can be longer ones in some isolates) and 30 nm diameter (Fig. 5). The particle structure is based on a $T = 3$ icosahedron cut across its threefold axes, the tubular portion being made up of rings of hexamer subunits with a repeat distance of about 10 nm.

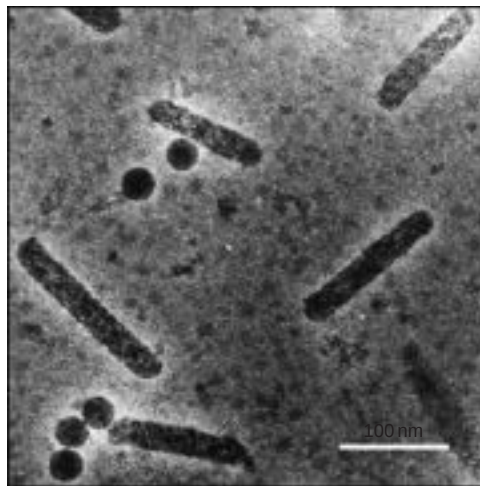


Fig. 5 Particles of RTBV (long particles) and RTSV (spherical particles) negatively stained in uranyl acetate. The bar indicates 100 nm.

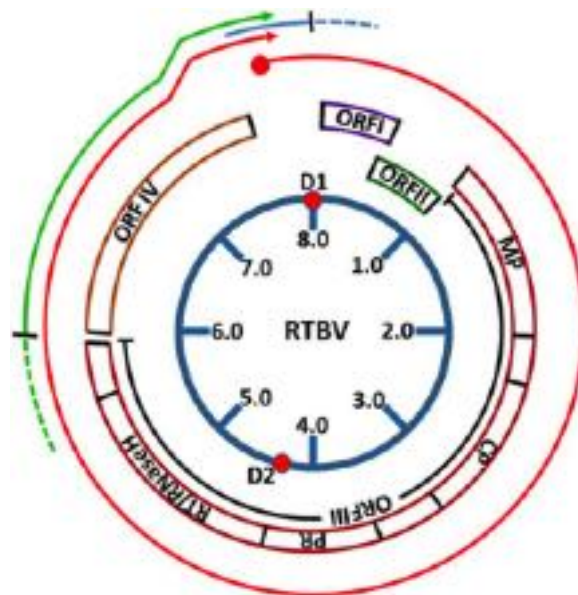


Fig. 6 The genome organization of RTBV and the different transcripts associated with it. The innermost circle represents the double-stranded 8 kb DNA. The numbers inside the circle represent the position of the nucleotide ($\times 1000$) on the genome. The approximate sites of the two discontinuities D1 and D2 are marked. The position of the four open reading frames, ORF I, ORF II, ORF III and ORF IV, are represented as arcs. The proteins encoded by ORF III polyprotein have been marked. MP: Movement protein; CP: Coat protein; PR: Protease and RT/RNase H; Reverse transcriptase/ribonuclease H. The arc in orange represents pre-genomic RNA. The blue-green arc represents the spliced ORF IV RNA. The dotted arcs represent the intronic region of the transcript.

Genome

The RTBV genome is a double-stranded circular DNA of an approximate size of 8 kb (**Fig. 6**). It has high resemblance with badnaviruses and caulimoviruses, in having site-specific discontinuities one on each strand of the double stranded circular genome, a poly-protein analogous to retroviral gag-pol and involvement of an RNA intermediate in replication.

The RTBV genome potentially has four Open Reading Frames (ORFs) viz., ORF I (P24), ORF II (P12), ORF III (P194) and ORF IV (P46) named according to the sizes of the proteins they encode. The transcription of RTBV genome is asymmetric with one strand (containing the minus-strand priming site) possessing all the coding capacity. The ORF I, II and III are compactly arranged and have a coinciding start-stop signal (ATGA). The ATG of this signal acts as a start codon for downstream ORFs and TGA of this signal codes for stop codon for the upstream ORF. The third ORF encodes for a poly-protein, which is cleaved subsequently to produce four different proteins. A short non-coding intergenic region separates the ORF III from ORF IV, which is expressed from a

spliced RNA. ORF I and ORF IV are also separated by a long intergenic region which includes the single RTBV promoter, the signal for poly-adenylation, various short ORFs and the splice donor site.

Functions of Proteins

The 24 kDa protein (P24) encoded by ORF I is poorly characterized till date although it is reported to play some role in viral particle assembly. P24 interacts with the host D1 protein, associated with photosystem II, which may affect the host photosynthetic machinery, leading to the appearance of tungro-associated symptoms.

The ORF II encodes for a 12 kDa protein (P12), which is shown to interact both *in vivo* and *in vitro* with the coat protein domain of the ORF III encoded poly-protein and plays a role as a protein scaffold in the assembly of the virus capsid.

The longest ORF of RTBV is the ORF III, which encodes for a polyprotein having four clear-cut domains for movement protein (MP), coat protein (CP), aspartate protease (PR) and reverse transcriptase/ribonuclease H (RT/RNase H) arranged clockwise on the genome map.

The presence of an ORF IV downstream of the gag-pol analog of retroviruses is the single feature that differentiates RTBV from other badnaviruses. ORF IV possesses an unconventional isoleucine (ATT) start codon rather than the canonical methionine ATG codon and is expressed as a spliced product in association with small ORF I. ORF IV encodes for a 46.2 kDa (P46) protein, which acts as a viral suppressor of RNA silencing, affecting the cell-to-cell spread of the RNA silencing signal.

The viral coat protein comprises one major species of Mr 37 kDa; this may be processed to give a product of Mr about 33 kDa and possibly others. The molecular mass of the coat protein has been measured by mass spectrometry and the N- and C-terminal sequences have been determined. This protein contains two specific amino acid sequence motifs, CXCX₂CX₄HX₄C characteristic of reverse transcribing virus gag proteins and CX₂CX₁₁CX₂CX₄CX₂C unique to badnaviruses and tungroviruses.

Genome Variability

There are 13 complete nucleotide sequences of RTBV available in databases till date, four from India and nine from South-East Asia. The genome sizes of the South-East Asian isolates are slightly more (~100 bp) as compared to the Indian isolates. The molecular as well as the nucleotide and amino acid sequence based phylogenetic analyses of these RTBV isolates have now grouped them into several clusters; those from India forming a distinct cluster (Indian strains), whereas the rest are more diverse (Fig. 1(B)). The Indian strains of RTBV showed divergence from the rest due to several insertions, deletions and substitutions, mostly in the intergenic regions. A replication mechanism with the absence of proof-reading may be a plausible reason for the slight variations among these isolates. The sequence variability of RTBV genomes is similar to those of badnaviruses, such as banana streak virus, dioscorea bacilliform virus and cacao swollen shoot virus. For badnaviruses, virologists consider 80% as the similarity threshold to differentiate species. If this threshold is applied to RTBV, 7 species would exist among the 13 complete genome sequences (Fig. 1(B)), including three in the Philippines alone. Therefore, we would have a single disease, the Rice Tungro Disease, associated with several different viruses, a situation very similar to badnaviruses. The situation is completely different for RTSV, where all the six complete sequences from three countries (Philippines, Malaysia and India) are within 90% similarity or above (Fig. 1(B)).

Virus-Host Relationships

Host plants, upon infection with viruses, launch an RNA-based defense response against the virus, which leads to degradation of the viral transcript into 21–24 nucleotide fragments, which are called viral si (small-interfering) RNAs. This degradation is sequence-specific and viral siRNAs initially formed by the recognition of RTBV transcript for degradation, leads to an amplification of further degradation in a sequence-specific manner, resulting in the accumulation of copious amounts of siRNAs from RTBV in infected rice plants. This finally results in a gradual fall in RTBV levels in a few weeks' time. Certain regions of the RTBV genome, namely the portion immediately prior to ORF I, the 5'-untranslated region (5' UTR), gives rise to a large amount of siRNA, as compared to other regions of the genome. Despite the production of large quantities of siRNAs, the RTBV transcripts largely escape degradation because the 5' UTR folds into a strong degradation-resistant secondary structure. This could be a new type of escape employed by RTBV against host defense, not generally seen in other viruses.

Control

There are three major approaches to controlling tungro; by the manipulation of various cultural practices such as planting time, having a crop-free period, and seedbed protection; invasion of young susceptible rice plants by viruliferous GLH can be avoided. Spraying insecticide can control the leafhoppers but this has not proved to be particularly effective. There have been various programs to breed resistance to tungro but most have proved of limited success. The International Rice Research Institute Philippines, has screened more than 40,000 accessions of rice for resistance to tungro viruses and have found two basic types of resistance, that to the vector and that to the virus. Since RTBV and RTSV occur together in the plant, breeding efforts have not been directed against exclusively RTBV or RTSV, but both together. Resistance to the vector has also been used, but has proved to be not very durable.

In addition to the breeding-based approach, transgenic strategies have also been employed, targeting RTBV. With the development of the concept of pathogen derived resistance (PDR), in 1980s, there has been several attempts to confer resistance against RTBV through generation of transgenic plants containing virus-derived genome fragments or full-length genes. Both protein-mediated and RNA-mediated strategies of PDR have been used for generating virus resistant transgenic plants. The movement protein gene, the coat protein (CP) genes, protease or the viral replicase genes have been exploited for initiating successful PDR. The proteins or their parts act as a dominant negative competitor during the infection of virus. RNA-based approaches have been used, using the concept of RNA-interference (RNAi). Rice plants expressing an RNAi construct against RTBV ORF IV have been produced, which show only mild symptoms upon GLH-mediated inoculation of both RTBV and RTSV. This resistance has been recently diversified into several popular rice varieties of southern India by marker-assisted breeding and many of the progenies have acquired the resistance along with the transgene, which is characterized by almost 1000-fold fall in the titers of RTBV upon GLH-mediated inoculation, when compared to the titers seen in non-transgenic rice plants of the same variety. The above plants do not show any fall in RTSV titers. A similar approach has also been used to generate transgenic rice plants expressing RNAi constructs targeting both RTBV and RTSV simultaneously, which, as expected, show resistance against both the viruses. However, it needs to be emphasized that all the work involving transgenic rice plants described above, have been conducted under controlled laboratory or green house conditions and are yet to be validated in the field.

Biotechnological Applications

Viral promoters, such as CaMV 35S, find wide application as expression modules for heterologous genes in plants. The potential of RTBV promoter, which has a similar overall location in the viral genome as CaMV 35S, has been explored. RTBV promoter shows tissue-specific activity, specifically in the phloem, in transgenic rice plants generated with a construct wherein a reporter gene is driven by the upstream sequences of the RTBV ORF I taken from a Philippine isolate. Further work with an Indian isolate of RTBV identified a complex arrangement of positive and negative regulatory elements in the promoter region, some of which were located downstream to the transcription start site. These constructs have potential use in driving heterologous gene expression in transgenic rice plants, and possibly in other monocot species.

Transient gene silencing in plants can be achieved by the use of viral vectors, utilizing the phenomenon of RNAi. This phenomenon is termed Virus Induced Gene Silencing or VIGS. RTBV genome has been modified suitably to develop a VIGS system for rice. Using this system, any rice gene can be silenced in a transient manner and its effect can be studied, giving the investigator a glimpse of its function in the plant. This approach has important applications in revealing gene functions in the rice plant.

Rice Tungro Spherical Virus

Taxonomy

RTSV belongs to the genus *Waikavirus*, in the family *Secoviridae*, in the order *Picornavirales*.

Transmission and Host Range

RTSV appears to have a restricted host range limited to members of the Poaceae and Cyperaceae (*Echinochloa crus-galli*, *E. glabrescens*, *E. colona*, *Leptochloa chinensis*, *Leersia hexandra*, *Oryza sativa*, *Panicum repens*, *Cyperus rotundus*); RTSV causes few, if any, symptoms in these hosts.

Structure

RTSV has isometric particles having a diameter of 30 nm (Fig. 5). The virus is limited to vascular tissues where it is restricted to the phloem cells. Virus particles are scattered in the cytoplasm or are embedded in non-enveloped electron-dense granular inclusions. Particles also occur as crystalline aggregates often in vacuoles. Small vesicles containing fibers are found in the cytoplasm, usually along the wall of infected cells.

Gene Functions

The capsid is made up of three coat protein species (CP1, CP2, and CP3) that are apparently in equimolar amounts, and is thought to have a similar T = 1 icosahedral architecture to that of picornaviruses. Identification of cleavage sites by N-terminal sequencing shows that CP1 is 22.9 kDa and CP2 is 22.3 kDa; the C-terminus of CP3 has not been identified. In Western blots of crude extracts from infected plants, an antiserum to CP3 identifies a band at 33 kDa and several in the range 40–42 kDa; the nature of these larger bands is unknown. The genome is translated into a poly-protein, which is processed to give various products including the three coat protein species, a putative type III helicase, a cysteine or 3C-type protease, and an RNA-dependent RNA polymerase at its 3' end. Functions of other products have not yet been determined. No functions have been ascribed to the products of the 30 short ORFs and there is no evidence that these are expressed.

Genome Diversity

Unfortunately, there is not a lot of sequence information available for RTSV, there are however 6 complete genome sequences originating from three countries and all sequences are 90% similar or greater (Fig. 1(B)), indicating that they most probably all belong to the same species.

Genome

Each RTSV particle contains one molecule of single stranded positive-sense RNA of about 12.2 kbp. The RNA encodes a large poly-protein of about 393 kDa and has two short ORFs at the 3' end (Fig. 7). A total of six nucleotide sequences, representing the full RTSV genome are available in the NCBI Nucleotide database (June 2019) of which three are from India, two from Philippines and one from Malaysia (Fig. 1(B)). RTSV strain Vt6 overcomes the resistance to the type strain in rice cultivar TKM6. Several serological variants of RTSV have been reported. Molecular techniques reveal much micro-variation in the sequences encoding CP1 and CP2 but there is no evidence for major geographic strains that reflect those of RTBV. Comparatively speaking, RTSV is much more conserved than RTBV, when sequences available from various regions of South and South-East Asia are compared.

Virus-Host Relationships

Recently, the virus-derived siRNAs have been studied in rice infected with both RTBV and RTSV using high throughput sequencing. Despite a reasonably large quantity of such siRNAs accumulating in the infected plants against RTBV, the accumulation of the corresponding siRNAs against RTSV is extremely low. This study points towards possibly an uncharacterized mechanism of escape from the siRNA-mediated resistance response of the plant by RTSV.

Control

Natural resistance in rice against RTSV is conferred as a recessive trait controlled by the translation initiation factor 4-gamma gene (eIF4G). Three residues of eIF4G, Y¹⁰⁵⁹ V¹⁰⁶⁰ V¹⁰⁶¹ are known to be associated with resistance against RTSV. Using CRISPR/Cas9 system in the RTSV-susceptible variety IR64, mutations in eIF4G, have been generated to confer resistance against RTSV. These mutations were located adjacent to the YVV residues and were successfully transmitted to the next generations. However, these approaches have yet to be used at the field level to effectively control RTSV.

Concluding Remarks

RTBV and RTSV are a unique combination of unrelated viruses, which have come together to cause the Tungro Disease, one of the most important crop diseases. Tungro has the potential to cause major disruptions in the production of rice in south and South-East Asia, thereby posing a threat to food security in this part of the World. The functions of the proteins encoded by the two viruses are yet to be worked out in detail. Recent research has revealed a novel role for RTBV P24 as an RNAi suppressor, hence possibly acting as a virulence factor for the virus. Despite the above RNAi suppressor being active, RNAi-mediated resistance has been demonstrated against RTBV and RTSV in rice under experimental conditions. Such studies can have

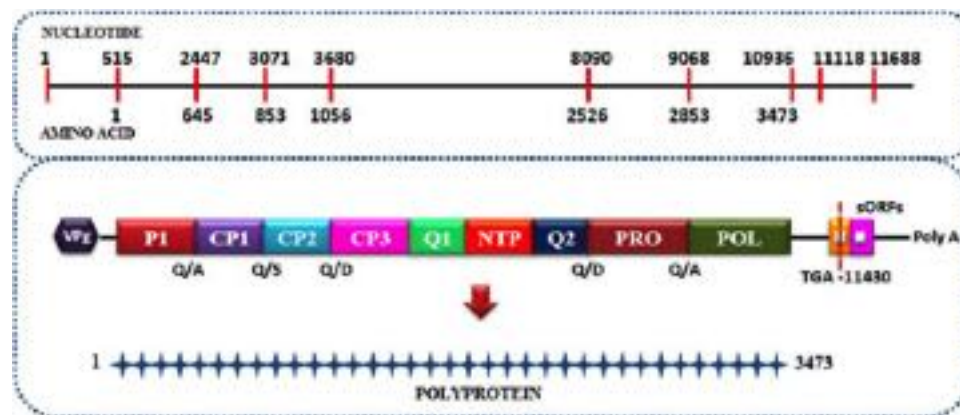


Fig. 7 The genome organization of RTSV. The polyadenylated RNA is indicated as a single line and the coding regions as boxes. The nucleotide and amino acid positions of the different coding regions are indicated above. The amino acids present at the cleavage sites are indicated below the boxes. The position of the TGA stop codon in sORF II has been indicated by red dotted line. The 3473 amino acids long polyprotein is indicated by spiraled line below the RTSV genome. VPg: Viral protein of the genome; P1: Leader protein 1; CP1–3: Coat protein 1–3; NTP: Nucleotide triphosphate binding protein; Pro: Protease; Pol: RNA dependent RNA polymerase; sORF: short ORF.

important applications in controlling the viruses and hence the Tungro Disease. Other important aspects of the viruses, namely the semi-persistent transmission of the viral complex by the vector GLH and the potential helper function mediated by RTSV for the transmission of RTBV are, however, still shrouded in mystery.

See also: Caulimoviruses (*Caulimoviridae*). Plant Antiviral Defense: Gene-Silencing Pathways. Plant Resistance to Viruses: Engineered Resistance. Secoviruses (*Secoviridae*)

Further Reading

- Anjaneyulu, A., Satapathy, M.K., Shukla, V.D., 1994. Rice Tungro. New Delhi: Oxford and IBH Publishing Co. Pvt. Ltd.
- Hull, R., 1996. Molecular biology of rice tungro viruses. *Annual Review of Phytopathology* 34, 275–297.
- Hull, R., 2014. *Matthews' Plant Virology*, fifth ed. London: Academic Press.
- Kumar, G., Dasgupta, I., 2017. Molecular biology of rice tungro viruses and strategies for their control. In: Shamim, M., Singh, K.N. (Eds.), *Biotic Stress Management in Rice*. Apple Academic Press, pp. 1–15.

Rice Yellow Mottle Virus (*Solemoviridae*)

Eugénie Hébrard and Nils Poulicard, Interactions Plantes Microorganismes Environnement, Institut de Recherche pour le Développement, Centre de coopération internationale en recherche agronomique pour le développement, University of Montpellier, Montpellier, France

Mbolarinsoy Rakotomalala, FOFIFA, Antananarivo, Madagascar

© 2021 Elsevier Ltd. All rights reserved.

This is an update of E. Hebrard, D. Fargette, G. Konate, Rice Yellow Mottle Virus. In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00715-9](https://doi.org/10.1016/B978-012374410-4.00715-9).

Introduction

Rice yellow mottle virus (RYMV) is a member of the genus *Sobemovirus* in the family *Solemoviridae*. RYMV is present only on the African continent. It was first reported in Kenya in 1966 and in Côte d'Ivoire in 1974 in irrigated rice fields. Since the early 1990s, RYMV is present everywhere where rice is grown in sub-saharan Africa and in Madagascar. It affected all types of rice cultivations, including lowland, upland rainfed, floating and mangrove rice. RYMV regularly induces severe yield losses to rice production ranging from 20% to 100%. In some regions, when epidemics are recurrently very severe, farmers abandoned their fields and planted new rice fields. Highly susceptible cultivars have been eliminated by the disease. RYMV is now ranked as the main biotic threat to rice production in Africa.

Virion Properties

RYMV has icosahedral particles of 25 nm in diameter ([Fig. 1](#)). The virion particles are composed of a single coat protein (CP) of 29 kDa, a genomic RNA (gRNA), and one subgenomic RNA (sgRNA) molecule. The capsid contains 180 copies of the coat protein subunit arranged with $T = 3$ symmetry. The structure of RYMV was determined by X-ray crystallography at 2.8 Å resolution and compared to the structure of Southern cowpea mosaic virus (SCPMV) and Sesbania mosaic virus (SeMV). As for other sobemoviruses, RYMV CP subunits are chemically identical but structurally not equivalent. Three types of CP subunits termed A, B, C are related by quasi three-fold axes of symmetry and are involved in different inter-subunit contacts. In the C-type subunits, a longer part of the N-terminus is ordered (residues 27–49), forming an additional β -strand named β A arm. Sobemovirus particles are stabilized by divalent cations, pH-dependent protein-protein interactions, and salt bridges between protein and RNA. Upon alkali treatment in presence of chelators, the capsid shell swells and becomes sensitive to enzymes and denaturants. In these conditions, RYMV is more stable than SCPMV. This property is likely due to the 3D swapping of the RYMV β A arm around the distal quasi-six-fold axes as found with SeMV whereas the SCPMV β A arm makes a U-turn and is localized around the nearby quasi-six-fold axes. Another reason for the better stability of RYMV is likely to be the strong RNA-protein interactions resulting from the presence of ordered RNA within the capsid shell. RYMV particles exist in three forms with different stability (1) an unstable swollen form dependent on basic pH but lacking Ca^{2+} , (2) a transitional compact form dependent on acidic pH, but also lacking Ca^{2+} and (3) a stable compact form that is pH independent and contains Ca^{2+} . The compact form is highly infectious and probably required for virus movement and transport by vector.

Molecular diversity of the CP was analysed in relation to the capsid three-dimensional structure in order to identify which amino acids are involved in the differential recognition by certain monoclonal antibodies. The residues in position 178 and 180, despite their internal localisation in the capsid, can modify the antigenic reactivity, with Mabs G and E allowing serotypes Sr3 and Sr5 to be differentiated.

Organization of the Genome

The genomic RNA is one single-stranded messenger-sense molecule, ca. 4450 nt in size. The 5'-terminus of the RNAs has a genome-linked protein (VPg), and lacks a poly(A) tail. RYMV often encapsidates, in addition to its genomic and subgenomic RNA, a small circular satellite RNA (satRNA), that is dependent on a helper virus for replication. RYMV satRNA has 220 nt and it is a viroid-like RNA. No link between the RYMV satRNA and its pathogenicity has been demonstrated.

Like other sobemoviruses, the RYMV genome is organized in 5 overlapping open reading frames (ORF) ([Fig. 2](#)). Two non-coding regions at the RNA extremities contain 80 and 289 nt, respectively. The ORF1 and the ORF3 are localized at the 5' and 3'-extremities, respectively. An intergenic region of 12 nt separates ORF1 and ORF3. All the following coding regions overlapped two by two (ORF3 and ORF2a, ORF2a and ORF2b, ORF2b and ORF3).

The ORF1 encodes a protein P1 which is dispensable for replication but is required for systemic movement in plants. Moreover, RYMV P1 displays multiple functions in silencing by locally suppressing transgenic silencing but enhancing the local and systemic spread of silencing in the non-host plant *Nicotiana benthamiana* and inhibiting the DCL4-dependent endogenous siRNA pathway when transgenically expressed in rice.

The putative protein encoded by ORF3 is predicted to be translated from a leaky scanning and a non-AUG initiation. As for all the sobemoviruses, the protein Px has not been detected in infected plants and its function is unknown.

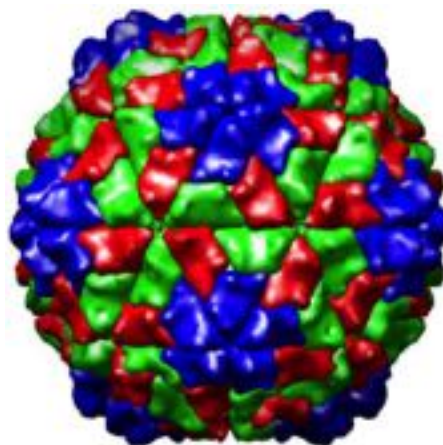


Fig. 1 RYMV capsid structure. The capsid comprised 180 copies of one single type of polypeptide arranged in $T = 3$ quasi-equivalent symmetry. The icosahedral asymmetric unit contains three subunits: A (in blue), B (in red) and C (in green). Each subunit is involved in different inter-subunit contacts. This image was automatically generated from VIPER virus capsid PDB file 1F2N using the MultiScale extension to the Chimera interactive molecular graphics package. Reproduce from Goddard, T.D., Huang, C.C., Ferrin, T.E., 2005. Software Extensions to UCSF Chimera for Interactive Visualization of Large Molecular Assemblies, Structure 13 (3), 473–482. Shepherd, C.M., Borelli, I.A., Lander, G., *et al.*, 2006. 3rd, Reddy VS., VIPERdb: A relational database for structural virology. Nucleic Acids Research 34, (Database issue), D386–9.

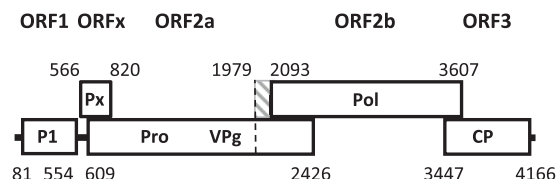


Fig. 2 RYMV genomic organization. Positions of the open reading frames (ORFs) are indicated in nucleotides. P1, protein X (Px), proteinase (Pro), VPg, RNA-dependent RNA-polymerase (Pol) and coat protein (CP) are labelled. The dotted line at nucleotide 1979 represents the frameshifting signal. The fusion point of the polyprotein P2a + b is unknown, and an AUG codon present at the beginning of the ORF2b (nucleotide 2093) is indicated by the vertical line.

The polyproteins encoded by the RYMV ORF2 are presumably translated from leaky translation. The start codon of ORF2 appeared to be in a more favorable context than ORF1. The ORF2b is translated as a polyprotein fused with ORF2a after a -1 programmed ribosomal frameshifting mechanism. The presence of a heptanucleotide slippery sequence UUUAAC and a predicted stem-loop structure are necessary. The ORF2a encodes the proteinase, the VPg and two putative proteins P10 and P8, the ORF2b encodes the RNA-dependent RNA-polymerase. RYMV proteinase contain the consensus sequence $H(X_{32-35})[D/E](X_{61-62})TXXGXSG$ characteristic of serine proteinases, they cleave between E-T and E-S amino acid residues. RYMV VPg is covalently linked to a serine, is characterized by the presence of a conserved $W[A/G]D$ sequence followed by a D- and E-rich region and can be phosphorylated. The interaction between the VPg and the rice eukaryotic translation initiation factor eIF4(iso)G1 is needed for a successful infection. The VPg and the P8 protein are intrinsically disordered. The localisation of the polymerase was predicted from the presence of the GDD motif and surrounding conserved motifs characteristics of RdRp.

The ORF3 is translated from the subgenomic RNA and encodes the coat protein (CP). RYMV CP is dispensable for replication but required for movement to long-distance likely through encapsidation. This movement was suggested to occur during the cell differentiation to vessels. The N-terminal sequence contains a putative bipartite nuclear localisation signal (NLS). This highly basic region is thought to interact with viral RNA and to stabilize the virion but virus like-particles (VLP) can be formed.

Relationships of the Species With Other Taxa

The species *Rice yellow mottle virus* belongs to the genus *Sobemovirus* in the family *Solemoviridae*. The genus contains 19 definitive species including *Southern bean mosaic virus*, the type species. The sobemovirus CPs are related to those of necroviruses (*Tombusviridae*) whereas the proteinase, VPg and polymerase are related to those of poleroviruses and enamoviruses (*Luteoviridae*).

Biological Properties

The natural host range of RYMV is restricted to a few members of the Poaceae family, principally *Eragrostidae* and *Oryzae* sp. The most commonly cited species are *Oryza sativa*, *O. glaberrima*, *O. longistaminata*, *O. barthii* and *Echinochloa colona*. A few additional Poaceae plants have been infected experimentally.

RYMV particles are present in all plant parts including roots and seeds. Particles are present in large amount in mesophyll cells, and in vascular tissues mainly in xylem and associated parenchyma cells. Particles were found individually, as aggregates, or as crystalline forms in the cytoplasm and the vacuoles of infected cells. In addition, the cytoplasm of infected cells contains fibrillar material located either in vesicles or distributed in diffused patches. Virus particles are also found in mature xylem, as well as inside the primary wall.

Cytological changes were found in the chloroplasts of infected mesophyll cells where the starch grains decreased in size and number. The chloroplasts form invaginations containing mitochondria, peroxisomes and virus particles. Electron dense materials were observed in the nuclei and associated with fibrillar elements in cytoplasmic vesicles. The most dramatic changes induced by RYMV occurred in the cell walls of parenchyma and mature xylem cells causing disorganisation of the middle lamellae wall. From infected cells, a viral ribonucleoprotein complex moves from cell-to-cell to reach the vascular bundle sheath. Viral replication, encapsidation and storage in vacuoles occur mainly in vascular parenchyma cells. The calcium linkage from pit membranes to virus particles during xylem differentiation could contribute to disruption of these membranes and facilitate systemic virus transport. In the upper leaves, virions are suspected to cross the pit membrane to infect new vascular cells and virions spread out by cell-to-cell movement through plasmodesmata. At this stage of infection, most replication occurs in mesophyll and vascular cells, and most of the virions accumulate in large crystalline patches.

Vector transmission is mainly due to coleoptera vectors such as *Chaetocnema* spp., *Sesselia pusilla* et *Trichispa sericea*, but other biting insects including the grasshopper *Conocephalus* have been reported to occasionally transmit the virus. Transmission is not persistent. The virus is present within the seed, although RYMV is not seed-transmissible. Other transmission modes have been also reported through animals like rats, common in rice fields, or cows and donkeys, used for field work or during grazing. Abiotic transmissions through plant residues, irrigation water, guttation water, wind, direct contact between plants and contamination by infected agricultural tools have been observed. Experimentally, mechanical transmission is easily achieved.

Diagnostic and Identification

RYMV diagnosis was first based on symptoms. However, many factors such as mineral deficiencies can induce symptoms of yellow mottle on rice. Therefore, several tools for specific detection of RYMV were developed. First, polyclonal antibodies (PABs) directed against different isolates were used in double-diffusion tests and in direct antibody sandwich enzyme linked immunosorbent assay (DAS-ELISA). No cross reactions with other sobemoviruses was observed, demonstrating the high specificity of these antibodies. Western-blot analysis can be used to detect RYMV with PABs. These antibodies were also used in immunolocalisation techniques. Immuno-printing tests derived from the direct tissue blotting technique were developed to analyse the distribution of the virus in entire leaves. Moreover, cytological detection of RYMV CP in rice tissues can be performed by immunofluorescence microscopy after glutaraldehyde/paraformaldehyde fixation and labelling with secondary antibodies linked to fluorescein. Polyclonal antibodies directed against P1 produced in *E. coli* are also available, and were used in Western-blot analyses. Monoclonal antibodies were also developed to study RYMV diversity. Indirect triple antibody ELISA (TAS-ELISA) with strain-specific MABs allows the major serotypes to be differentiated.

In parallel to the serological methods, molecular tools were developed. The first full-length sequence was obtained after RNA extraction from purified virus. Several specific primers were defined to perform RT-PCR amplification and to produce infectious transcripts. Northern and Southern blot methods are available using a DNA probe obtained by PCR amplification of the CP gene (ORF4) from the infectious clone CIa. *In situ* hybridisation tests of the RNA in rice tissues were also performed. In this case, the probe was DNA, obtained from the full-length genome of the clone CIa labelled randomly with digoxigenin-UTP. Real time-RT-PCR assay is also available using primers in the conserved ORF2a. RT-PCR amplification of the CP gene followed by direct DNA sequencing progressively has replaced other molecular methods of analyses.

Epidemiological Aspects

Typical symptoms of RYMV are a mottle and a yellowing of the leaves. However, streaks, necrosis, and whitening are sometimes observed with some cultivars and under specific growing conditions. Infection resulted in stunting of the plant, reduced tillering, poor panicle exertion and sterility. Death of the plants of susceptible cultivars occurs after early infection. Severe yield losses ranging from 20% to 100% after RYMV infection have been reported in several countries. Most cultivars currently grown are susceptible to RYMV. Heavy infection is associated with reduced fertility.

RYMV is an emergent disease. It is thought that the virus was originally harboured in wild Poaceae and transferred only recently to cultivated rice. The rapid and intense spread of the virus was associated to the change of agricultural practices in response to increasing food demand. Very productive but susceptible *O. sativa* cultivars have been introduced from Asia which allows two crop

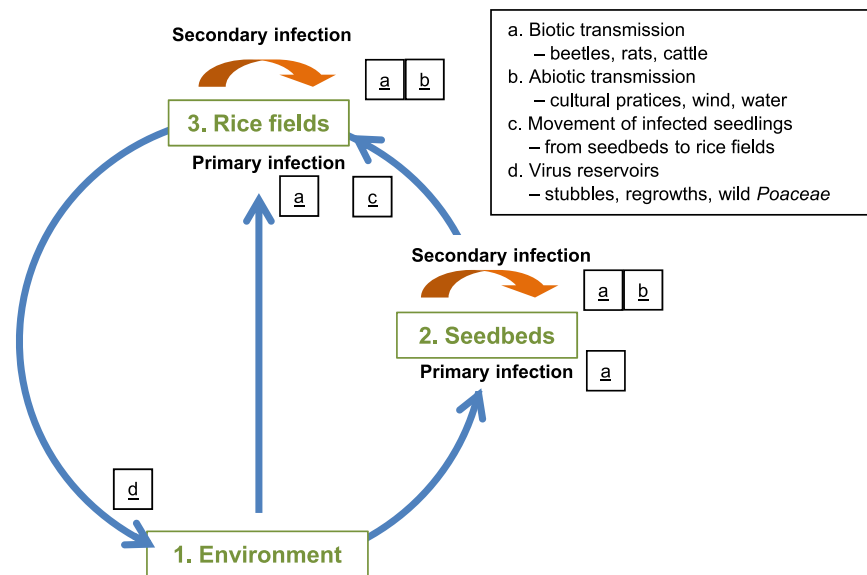


Fig. 3 Descriptive model of RYMV epidemiology. (1). Environment. RYMV is present in the environment in rice stubble and regrowths and in wild *Poaceae*. (2). Rice seedbeds. Primary infection occurred via biotic transmission (beetles, rats, cows) whereas cultural practices contribute to secondary infection in seedbeds. (3). Rice fields. Primary infection occurred via biotic transmission from the environment, and also when transplanting infected seedlings from seedbeds. Both biotic and abiotic transmissions contribute to secondary spread in rice fields. After harvesting, RYMV is perpetuated in contra-season in the environment through rice stubbles and regrowths and transmitted to wild *Poaceae*. From Hébrard, E., Fargette, D., Konaté, G., 2008. Rice Yellow Mottle Virus. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology* third ed. Oxford: Academic Press, pp. 485–490.

cycles a year resulting in the maintenance of the inoculum all over the year. RYMV propagation probably occurs as follows: (1) a virus source made of volunteers, of wild *Poaceae* or plant residues act as primary inoculum; (2) a direct contamination of the new rice fields occurs after mechanical transmission by man or biological infection by insects; (3) rice beds can also be infected by insect vector; (4) the new rice field originating from these infected seed-beds are contaminated by man during replanting; (5) further propagation of the virus is done by man animals, wind and irrigation water.

The epidemiology has been little studied and there are even uncertainties about the respective importance of biotic and abiotic transmission (Fig. 3). Under these conditions, precise forecasting of RYMV infection is not possible although growing intensification of rice cultivation in Africa will no doubt favour RYMV spread, unless durable resistant cultivars are introduced. Phytosanitary measures are sometimes advised although they are often economically not practicable and their impact to reduce virus spread is unknown. They include protection of seed-beds by nets, disinfection of tools used at replanting, destruction of volunteers and rice residues. Chemical control of the vectors is not economically feasible, is ecologically dangerous, and is most unlikely to be successful considering the large number of species involved in transmission.

Several breeding programs have been conducted in order to select and breed resistant cultivars. The genotype and the phenotype of two kinds of natural resistance have been characterized. Partial resistance is encountered in cultivars of *Oryza japonica* species with a delayed virus multiplication and symptom expression. Partial resistance is polygenic and a major quantitative trait locus has been identified on chromosome 12. High resistance has been identified in a limited number of accessions of *Oryza glaberrima* and two *O. sativa indica* species. No symptoms are apparent and virus content is most often undetectable. Three resistance genes have been identified. The recessive gene *RYMV1* that maps on chromosome 4 encodes the translation initiation factor eIF(iso)4G1. Compared to susceptible varieties, the resistance alleles are characterized by a point mutation and/or a small deletion in the conserved domain of the gene. The VPg was identified as the virulence factor, it interacts directly with eIF(iso)4G1 in susceptible plants. The *RYMV1* resistance was efficient against some of the representative isolates of the main virus strains. However, resistance-breakdown was reported in experimental conditions depending on the viral strain and the resistance allele. A single point mutation was sufficient to break the resistance but negative epistasis can inhibit the virulence emergence. The amino acid located at position 49 of the VPg reflects inter-species adaptation of the viral strains and has indirect effects on resistance durability. The recessive resistance gene *RYMV2* is homologous to the *Arabidopsis thaliana* gene for Constitutive expressor of Pathogenesis-Related genes-5 (CPR5), a nucleoporin controlling defense mechanisms. *RYMV2*-mediated resistance is conferred by a null allele. The virulence factor is the polyprotein P2a. The dominant resistance gene *RYMV3* was mapped in a 15 kb interval in which a CC-NBS-LRR domain-containing protein was annotated. Recently, a hypervirulent pathotype able to overcome all natural sources of high resistance has been identified in West-Central Africa. Transgenic plants with portions of the ORF2 expressing partial resistance have also been produced. This transgenic resistance is thought to involve a gene silencing mechanism. However, the transgenic lines showed a less effective, partial and temporary resistance compared to natural resistances.

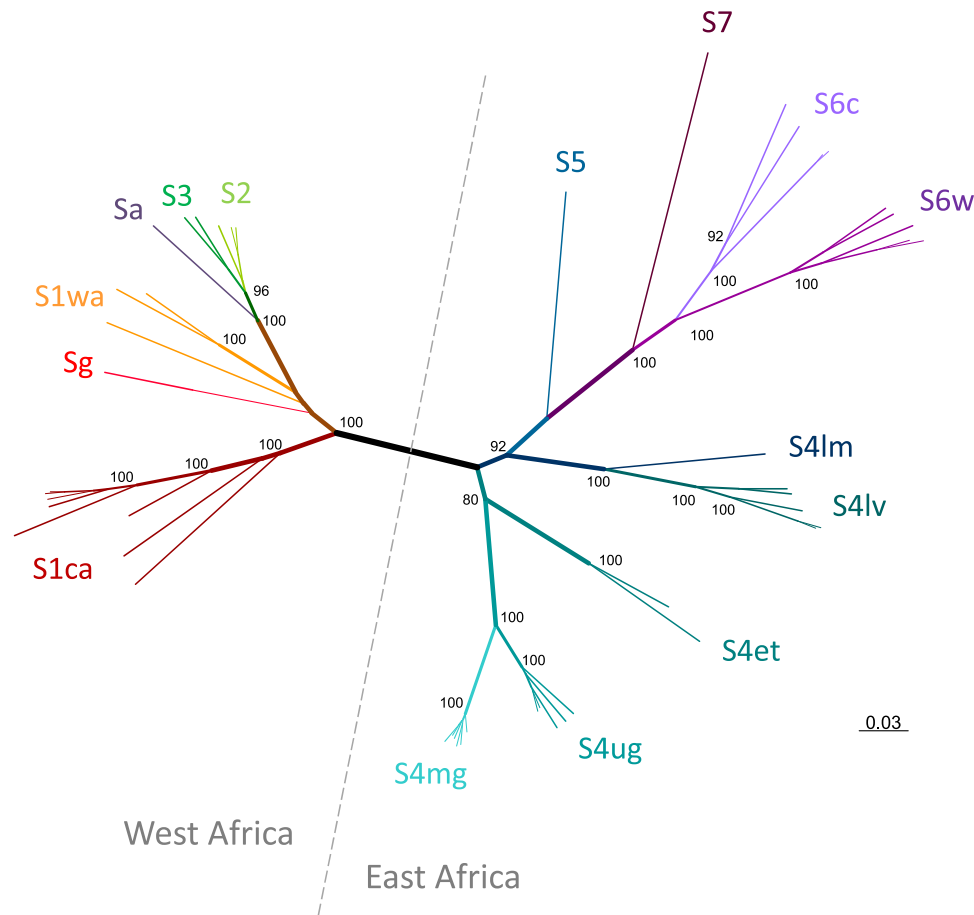


Fig. 4 Phylogeny of RYMV. Phylogenetic tree inferred by maximum likelihood from the full sequences of 48 isolates representative of the genetic diversity and geographic distribution of RYMV in Africa. The different strains identified in West Africa (S1ca, S1wa, S2, S3, Sa, Sg) and East Africa (S4et, S4lm, S4lv, S4mg, S4ug) are differentiated by color. Strong bootstrap support of the nodes (>80%) is indicated.

Diversity and Evolution

The diversity of RYMV was assessed by studying isolates from several countries where the disease has been observed. The sequences of more than 500 CP genes are now available. Forty-eight isolates representative of the geographic and molecular diversity were fully sequenced. RYMV is a highly variable virus and analyses of the geographic distribution of the genetic diversity elucidated the process of evolution and of dispersal of the virus. RYMV showed a high level of population structure marked at the continental scale with two subdivisions: East-Africa and West-Africa (**Fig. 4**). The highest diversity was observed in East-Africa, with a pronounced peak in Eastern Tanzania, and a decrease from the east to the west of the continent. This pattern suggests a westward expansion with a succession of founder effects and subsequent diversification phases. Accordingly, genetic diversity would be adversely affected by recurrent bottlenecks occurring along the route of colonization, resulting in the lowest diversity in the extreme west. This, together with accumulation of *de novo* mutations postdating population separation, provides an explanation for the genetic differences among strains across Africa. The geographical overlaps between several strains would explain the recent detection of some recombination strains.

Further Reading

- Albar, L., Bangratz-Reyser, M., Hébrard, E., *et al.*, 2006. Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to Rice yellow mottle virus. *The Plant Journal* 47, 417.
- Bonneau, C., Brugidou, C., Chen, L., *et al.*, 1998. Expression of the Rice yellow mottle virus P1 protein in vitro and in vivo and its involvement in virus spread. *Virology* 244, 79.
- Brugidou, C., Holt, C., Yassi, *et al.*, 1995. Synthesis of an infectious full-length cDNA clone of Rice yellow mottle virus and mutagenesis of the coat protein. *Virology* 206, 108.
- Brugidou, C., Opalka, N., Yeager, M., *et al.*, 2002. Stability of Rice yellow mottle virus and cellular compartmentalization during the infection process in *Oryza sativa* (L.). *Virology* 297, 98.
- Hébrard, E., Pinel-Galzi, A., Oludare, A., *et al.*, 2018. Identification of a hypervirulent pathotype of Rice yellow mottle virus: A threat to genetic resistance deployment in West-Central Africa. *Phytopathology* 108 (2), 299–307.

- Pidon, H., Ghesquière, A., Chéron, S., *et al.*, 2017. Fine mapping of RYMV3: A new resistance gene to Rice yellow mottle virus from *Oryza glaberrima*. *Theoretical and Applied Genetics* 130 (4), 807–818.
- Pinel-Galzi, A., Dubreuil-Tranchant, C., Hébrard, E., *et al.*, 2016. Mutations in Rice yellow mottle virus polyprotein P2a involved in RYMV2 gene resistance breakdown. *Frontiers in Plant Science* 7, 1779.
- Pinel-Galzi, A., Rakotomalala, M., Sangu, E., *et al.*, 2007. Theme and variations in the evolutionary pathways to virulence of an RNA plant virus species. *PLOS Pathogens* 3, e180.
- Pinel-Galzi, A., Traoré, O., Séré, Y., Hébrard, E., Fargette, D., 2015. The biogeography of viral emergence: Rice yellow mottle virus as a case study. *Current Opinion in Virology* 10, 7–13.
- Pinel, A., Abubakar, Z., Traore, O., *et al.*, 2003. Molecular epidemiology of the RNA satellite of Rice yellow mottle virus in Africa. *Archives of Virology* 148, 1721.
- Poulicard, N., Pinel-Galzi, A., Traoré, O., *et al.*, 2012. Historical contingencies modulate the adaptability of Rice yellow mottle virus. *PLoS Pathogens* 8 (1), e1002482.
- Qu, C., Liljas, L., Opalka, N., *et al.*, 2000. 3D domain swapping modulates the stability of members of an icosahedral virus group. *Structure* 8, 1095.
- Sorho, F., Pinel, A., Traoré, O., *et al.*, 2005. Durability of natural and transgenic resistances in rice to Rice yellow mottle virus. *European Journal of Plant Pathology* 112, 349.
- Traoré, O., Sorho, F., Pinel, A., *et al.*, 2005. Processes of diversification and dispersion of Rice yellow mottle virus inferred from large-scale and high resolution phylogeographic studies. *Molecular Ecology* 14, 2097.
- Traoré, O., Traoré, M., Fargette, D., Konaté, G., 2006. Rice seedbed as a source of primary infection by Rice yellow mottle virus. *European Journal of Plant Pathology* 115, 181.

Satellite Nucleic Acids and Viruses

Olufemi J Alabi, Texas A&M AgriLife Research and Extension Center, Weslaco, TX, United States

Alfredo Diaz-Lara and Maher Al Rwahnih, University of California, Davis, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of P. Palukaitis, A. Rezaian, F. Garcia-Arenal, Satellite Nucleic Acids and Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00500-8](https://doi.org/10.1016/B978-012374410-4.00500-8).

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

diRNAs Defective interfering RNAs

dsDNA Double-stranded deoxyribonucleic acid

gRNA Genomic RNA

HEL Helicase

HV Helper virus

kb Kilobases; the size of a ssDNA or ssRNA molecule

kbp Kilobase pairs; the size of a dsDNA or dsRNA molecule

kDa Kilodaltons; the size of a protein

MCP Major capsid protein

mCP Minor capsid protein

NTR Non-translated region

ORF Open reading frame

Poly(A) Polyadenylated

PRO Protease

sgRNA Sub-genomic RNA

ssDNA Single-stranded deoxyribonucleic acid

ssRNA Single-stranded ribonucleic acid

tRNA Transfer RNA

Vpg Genome-linked protein

Glossary

Satellite RNA or satellite DNA A subviral genome dependent on a helper virus for both replication and encapsidation.

Satellite virus A subviral genome dependent on a helper virus for its replication but encoding its own capsid protein.

Satellite-like RNA A subviral genome dependent upon another virus for replication and encapsidation but is required for vector transmission of the helper virus.

Virophages Double-stranded DNA that depend on giant viruses for their replicate in co-infected eukaryotic cells.

Virus-associated nucleic acid A subviral genome that depends on another virus for encapsidation and transmission, but not for its replication.

Introduction

Satellites are a diverse collection of subviral agents that are capable of attenuating or exacerbating disease symptoms caused by their cognate helper viruses (HVs). Satellites can be differentiated from other subviral nucleic acids, such as sub-genomic (sg) RNAs, defective (d) or defective interfering (di) RNAs, and viroids, by their molecular, biological, and genetic nature. Unlike sgRNAs, dRNAs, and diRNAs, satellites and viroids have little or no sequence similarity to any known virus, except for the terminal regions. Whereas viroids are replicated by host polymerases, RNA satellites rely on the HV's polymerase for their replication. Satellites have been found associated with DNA and RNA viruses, and in the latter case, with both ssRNA and dsRNA viral genomes, the satellites being of the same nucleic acid type as the HV. While, in general, the HV can exist independent of the satellite, there are exceptions where a satellite contributes to the transmission of the HV and thus is referred to as being satellite-like. In addition, some viral RNAs and DNAs are found in association with other viral genomes on which they depend for their encapsidation and transmission, but not for replication. These viral RNAs and DNAs are referred to here as virus-associated nucleic acids. True satellites are divided basically into two main groups; those that encode their own capsid protein (CP) are called satellite viruses, while those that rely on their HV for both replication and encapsidation are referred to as satellite RNAs or DNAs (satellite nucleic acids). There are also satellites of satellites, in which some satellite RNAs are replicated by the HV but are encapsidated by the CP of a satellite virus. In addition, in the case of turnip crinkle virus (TCV, genus *Betacarmovirus*, family *Tombusviridae*), which replicates both satellite RNAs and diRNAs, it also produces a chimeric RNA (referred to as satC) that is in part a diRNA and in part a satellite RNA (satD). While most satellites are found in association with plant viruses, a few have been found in association with animal, fungal or protist viral genomes. Most plant viruses are not associated with satellites. However, the few that does have attracted significant attention due to the disease modulating roles of the associated satellites. With the advent of metagenomics, additional satellites have been identified prompting a revisit of their taxonomic classification to align them with that of *bona fide* viruses.

History

The use of the term satellite as a subviral agent was first conceptualized in 1962, to describe the relationship of a 17 nm diameter viral particle found in association with some isolates of the 26 nm diameter tobacco necrosis virus (TNV; genus *Alphanecrovirus*, family *Tombusviridae*). The smaller particle was dependent on TNV for its accumulation but was serologically unrelated to TNV. A few other satellite viruses have been described since then, but, in general, they are rare. By contrast, satellite nucleic acids are more common. The first satellite RNA was described in 1969, in association with tobacco ringspot virus (TRSV, genus *Nepovirus*, family *Secoviridae*), the satellite RNA being encapsidated by the CP of TRSV. The symptoms induced by TRSV were attenuated dramatically by the presence of the satellite RNA; but this is not always the case. Some satellite RNAs have no effect on either the accumulation or pathogenesis of the HV while a few may exacerbate the HV-induced disease. Most of these satellite RNAs contain ssRNA genomes, while a few contain genomes of dsRNA. Some of the satellite RNAs with ssRNA genomes are translated to produce proteins, which may or may not be required for their replication, depending on the specific satellite RNA. The first demonstration that satellite molecules are not limited to RNA viral systems was made in 1997 when a tomato leaf curl virus satellite (TLCV-sat-DNA) was characterized from disease-affected tomato plants in Australia. TLCV-sat-DNA is dependent on its TLCV HV for replication and encapsidation. In addition, it was shown that TLCV-sat-DNA replication was also supported by other taxonomically distinct geminiviruses, including the begomoviruses tomato yellow leaf curl virus (TYLCV) and African cassava mosaic virus (ACMV), and the curtovirus beet curly top virus (BCTV). Since then, additional classes of satellite DNAs have been characterized and described.

Geographical Distribution

The host range and geographical distribution of satellites mirror those of their HVs owing to the dependence of the former on the latter for genome replication and encapsidation. Similarly, satellite nucleic acids depend on the HV's transmission mechanism (including the dependence on arthropod vectors) for short and long-distance spread. Thus, the worldwide prevalence and distribution of satellite molecules draws from the global distribution of their HVs. Interestingly, alphasatellites and betasatellites are mostly associated with Old World monopartite geminiviruses, except for some alphasatellites that have been recently reported to occur in association with New World bipartite begomovirus infections.

Classification

Satellite viruses, satellite nucleic acids, virus-associated nucleic acids, and other satellite-like agents are very diverse in terms of their nucleic acid type, size, sequence, structure, or translatability. Satellite genomes may be composed of single or double stranded RNA or DNA molecules, which may be linear or circular. Whereas satellite viruses encode a CP for genome encapsidation, satellite nucleic acids depend on the CP of their HV for genome encapsidation. True satellites such as satellite viruses and satellite nucleic acids are dependent on the HV for their multiplication in co-infected host cells since they lack genes that encode functions needed for replication. Most or all portions of the genomes of true satellite, except for the terminal regions, are distinct from the genomes of their cognate HV.

Satellite-like nucleic acids are another class of subviral agents that differ from true satellites in that they encode a function necessary for the biological success of the associated virus. However, like the true satellites, they also are incapable of autonomous replication because they do not encode a replicase, with the exception of alphasatellites. Thus, they function primarily to remedy a deficiency in a defective virus and have sometimes been classified as part of the genome of the virus they assist. For instance, satellite-like RNAs associated with groundnut rosette virus (genus *Umbravirus*, family *Tombusviridae*) or beet necrotic yellow vein virus (BNYVV, genus *Benyvirus*, family *Benyviridae*) contribute to vector transmissibility of their associated viruses. Other classes of autonomously replicating nucleic acids (virus-associated nucleic acids) such as alphasatellites are dependent on their HV for encapsidation, cell-to-cell and long-distance movement and vector transmission. Thus, satellites do not constitute a homogeneous group and are traditionally not formally classified into species and higher taxa by the International Committee on Taxonomy of Viruses (ICTV). However, because several satellite molecules encode functional proteins, their taxonomy has been revisited to bring it in alignment with that of *bona fide* viruses. A list of currently known satellites categories is provided in [Table 1](#).

General Properties and Effects of Satellites

Satellite Viruses

Category 1: Chronic bee-paralysis associated satellite virus

Chronic bee-paralysis associated satellite virus (CBPVA; unassigned) is the only member of this group characterized to date ([Table 2](#)). The isometric particles of this satellite virus are found in bees infected with the HV, chronic bee-paralysis virus (CBPV), with CBPVA capable of interfering with CBPV replication.

Table 1 Categories of satellites

<i>Satellite viruses:</i>	
1	Chronic bee-paralysis virus-associated satellite virus
2	Satellites that resemble tobacco necrosis satellite virus
3	Nodavirus-associated satellite virus
4	Adenovirus-associated satellite virus (<i>Dependovirus</i>)
5	Mimivirus-associated satellite virus (Sputnik, virophage)
<i>Virus-dependent nucleic acids:</i>	
<i>Single stranded DNAs</i>	
6a	Alphasatellites
6b	Betasatellites and deltasatellites
<i>Double stranded RNAs</i>	
7	Satellites dsRNAs
<i>Single stranded RNAs</i>	
8a	Large linear single stranded satellite RNAs
8b	Small linear single stranded satellite and satellite-like RNAs
8c	Small circular single stranded satellite RNAs
8d	Hepadnavirus-associated satellite-like RNAs (<i>Deltavirus</i>)
8e	Polerovirus-associated RNAs

Table 2 Classification of satellites viruses

Category ^a	Family	Subfamily	Genus	Species	Type species ^b accession #	Genome/Host types ^c	Helper virus genus
<i>RNA Satellite viruses</i>							
1				Chronic bee-paralysis associated satellite virus (CBPVA)	N/A	ssRNA/ Arthropod	Unclassified
2			<i>Albetovirus</i>	<i>Tobacco albetovirus 1</i>	V01468	ssRNA/Plant	<i>Alphanecrovirus</i>
			<i>Albetovirus</i>	<i>Tobacco albetovirus 2</i>	M64479	ssRNA/Plant	<i>Alphanecrovirus</i>
			<i>Albetovirus</i>	<i>Tobacco albetovirus 3</i>	AJ000898	ssRNA/Plant	<i>Alphanecrovirus</i>
			<i>Aumaivirus</i>	<i>Maize aumaivirus 1</i>	M55012	ssRNA/Plant	<i>Aureusvirus</i>
			<i>Papanivirus</i>	<i>Panicum papanivirus 1</i>	M17182	ssRNA/Plant	<i>Panicovirus</i>
			<i>Virtovirus</i>	<i>Tobacco virtovirus 1</i>	M25782	ssRNA/Plant	<i>Tobamovirus</i>
3	<i>Sarothroviridae</i>		<i>Macronovirus</i>	<i>Macrobrachium satellite virus 1</i>	AY247793	ssRNA/ Arthropod	<i>Nodavirus</i>
<i>DNA Satellite viruses</i>							
4	<i>Parvoviridae</i>	<i>Parvovirinae</i>	<i>Dependoparvovirus</i>	<i>Adeno-associated dependoparvovirus A</i>	AF063497	ssDNA/Humans, Vertebrates	<i>Adenovirus</i>
	<i>Parvoviridae</i>	<i>Parvovirinae</i>	<i>Dependoparvovirus</i>	<i>Adeno-associated dependoparvovirus B</i>	AF085716	ssDNA/Humans, Vertebrates	<i>Adenovirus</i>
	<i>Parvoviridae</i>	<i>Parvovirinae</i>	<i>Dependoparvovirus</i>	<i>Anseriform dependoparvovirus 1</i>	U22967	ssDNA/Humans, Vertebrates	<i>Adenovirus</i>
	<i>Parvoviridae</i>	<i>Parvovirinae</i>	<i>Dependoparvovirus</i>	<i>Avian dependoparvovirus 1</i>	AY186198	ssDNA/Humans, Vertebrates	<i>Adenovirus</i>
	<i>Parvoviridae</i>	<i>Parvovirinae</i>	<i>Dependoparvovirus</i>	<i>Chiropteran dependoparvovirus 1</i>	GU226971	ssDNA/Humans, Vertebrates	<i>Adenovirus</i>
	<i>Parvoviridae</i>	<i>Parvovirinae</i>	<i>Dependoparvovirus</i>	<i>Pinniped dependoparvovirus 1</i>	JN420372	ssDNA/Humans, Vertebrates	<i>Adenovirus</i>
	<i>Parvoviridae</i>	<i>Parvovirinae</i>	<i>Dependoparvovirus</i>	<i>Squamate dependoparvovirus 1</i>	AY349010	ssDNA/Humans, Vertebrates	<i>Adenovirus</i>
	<i>Parvoviridae</i>	<i>Parvovirinae</i>	<i>Dependoparvovirus</i>	<i>Squamate dependoparvovirus 2</i>	KP733794	ssDNA/Humans, Vertebrates	<i>Adenovirus</i>
5	<i>Lavidaviridae</i>		<i>Mavirus</i>	<i>Cafeteriavirus-dependent mavirus</i>	HQ712116	dsDNA/Protists	<i>Mimivirus</i>
	<i>Lavidaviridae</i>		<i>Sputnikvirus</i>	<i>Mimivirus-dependent virus Sputnik</i>	EU606015	dsDNA/Protists	<i>Mimivirus</i>
	<i>Lavidaviridae</i>		<i>Sputnikvirus</i>	<i>Mimivirus-dependent virus Zamilon</i>	HG531932	dsDNA/Protists	<i>Mimivirus</i>

^aCategory adopted from the 9th ICTV Report.^bN/A, unavailable.^cssRNA, single-stranded RNA; ssDNA, single-stranded DNA; dsRNA, double-stranded RNA; dsDNA, double-stranded DNA.

Category 2: Satellites that resemble tobacco necrosis satellite virus (genera *Albetovirus*, *Aumavirus*, *Papanivirus* and *Virtovirus*)

The group consists of several plant-infecting satellite viruses that are found in association with taxonomically distinct HVs that used to be collectively known as the tobacco necrosis satellite virus or satellite tobacco necrosis virus (STNV) group. Recently, the ICTV recognized species belonging to this group and assigned them to the genera *Albetovirus*, *Aumavirus*, *Papanivirus*, and *Virtovirus*. Member belonging to these unassigned genera are found in association with ss(+)RNA plant HVs belonging to the genera *Alphanecrovirus*, *Aureusvirus*, *Panicovirus* (*Tombusviridae*), and *Tobamovirus*, (*Virgaviridae*) respectively (Table 2). These satellite viruses depend on their HVs for replication. With the exception of the rod-shaped *Tobacco virtovirus 1* (unassigned genus *Virtovirus*), particles of the plant satellite viruses are isometric, 17 nm in diameter, with a $T = 1$ symmetry, built of 60 protein subunits, as determined at high resolution by X-ray diffraction for *Tobacco albetovirus 1* (genus *Albetovirus*) and *Panicum papanivirus 1* (genus *Papanivirus*). Thus, the particle morphology of most satellite viruses differs from that of their respective HVs. In spite of the different structure of CP and virus particle, the particles of both the satellite and HV may share important properties; for example, the particles of *Tobacco albetovirus 1* can bind specifically to the zoospores of the vector fungus *Olpidium brassicae*. The ca. 1200 nt RNA of *Tobacco albetovirus 1* does not have a methylated cap structure or a genome-linked protein (VPg) in its 5' end. Unlike most plant virus RNAs, it has a phosphorylated 5' end. The 3' termini may share a structure with that of the HV genomic RNA (gRNA) as is the case for both tobacco mosaic virus and *Tobacco virtovirus 1* where their 3' termini form tRNA-like structures that are aminoacylable with histidine. Cis-acting sequences necessary for RNA amplification have been mapped at the 5' and 3' non-translated regions (NTRs) of *Panicum papanivirus 1* RNA as well as within the CP open reading frame (ORF), and are conserved in a defective RNA, which is maintained by this satellite virus. A translational enhancer domain has been mapped in the 3' NTR of albetoviruses. Like other satellite viruses, the isometric particle of *Maize aumavirus 1* (genus *Aumavirus*, unassigned) is 17 nm in diameter, but its CP has limited sequence similarity with orthologous proteins of other satellite viruses. Albetoviruses have been reported to interfere with each other's replication and with the accumulation of the HV. Satellite viruses can also modify the symptoms induced by the HV, as is the case with *Panicum papanivirus 1*, which in co-infection with Panicum mosaic virus (PMV) enhances the mild symptoms of the HV to cause severe mosaic and chlorosis in millet plants. Symptom induction is due to the *Panicum papanivirus 1* CP, and a chlorosis-inducing domain has been mapped. In addition to its structural and symptom-inducing functions, the CP of *Panicum papanivirus 1* is involved in systemic movement, binds PMV particles, and counters the effects of post-transcriptional gene silencing suppressors. Furthermore, the particles of albetoviruses may contain non-coding satellite RNA of about 620 nt, which depends on TNV for replication and on the satellite virus for encapsidation. A satellite RNA with similar dependence relationships has been described for PMV and *Panicum papanivirus 1*. These systems are good examples of the complexity of dependence relationships in satellitism.

Category 3: Nodavirus-associated satellite virus (genus *Macronovirus*, family *Sarothroviridae*)

Macrobrachium satellite virus 1 (genus *Macronovirus*; family *Sarothroviridae*) is the only currently recognized species of this group (Table 2), with extra small virus (XSV) as its representative member. XSV and its HV *Macrobrachium rosenbergii* nodavirus (MrNV) infect giant freshwater prawns and cause the debilitating and fatal white tail disease or white muscle disease in several countries. XSV is consistently found in co-infection with MrNV on which it depends for its replication. The linear, positive sense ssRNA genome of XSV is 796 nt in length and contains a short poly(A) tail of 15–20 nt at the 3' end. The non-enveloped virions of XSV are spherical (ca. 15 nm in diameter) and they are serologically and structurally different to those of MrNV. It encodes a single ORF that is translated into two CPs, one an N-terminally truncated version of the other. XSV and MrNV have also been detected in several species of aquatic insects that were sampled from infected freshwater prawn nursery ponds. In addition, the satellite and HVs can replicate in mosquito cell lines, suggesting that aquatic insects serve as vectors for their spread.

Category 4: Adenovirus-associated satellite virus (family *Parvoviridae*; genus *Dependoparvovirus*)

The genus *Dependoparvovirus* (family *Parvoviridae*) currently consists of eight ICTV approved species that are collectively known as adenovirus-associated satellite viruses (AAVs; Table 2). They are: *Adeno-associated dependoparvovirus A* (the type species), *Adeno-associated dependoparvovirus B*, *Anseriform dependoparvovirus 1*, *Avian dependoparvovirus 1*, *Chiropteran dependoparvovirus 1*, *Pinniped dependoparvovirus 1*, *Squamate dependoparvovirus 1*, and *Squamate dependoparvovirus 2*. Their non-enveloped icosahedral virion is 18–26 nm in diameter with the capsid consisting of 60 copies of CPs, encapsidating a linear, ssDNA(+/–) genome of about 4.7 kb in size. As their genus name implies, the AAVs are mostly (or thought to be likely) dependent on helper DNA viruses, most commonly an adenovirus or herpesvirus, for replication in co-infected cells of their mammalian, avian or reptilian hosts. Though ubiquitous in humans and most other vertebrates, AAVs are by themselves not associated with any known pathology.

Category 5: Mimivirus-associated satellite virus (genera *Mavirus*, *Sputnikvirus*, family *Lavidaviridae*)

The discovery of giant viruses (extremely large genomes) infecting protists also resulted in the identification of new satellite viruses. The family *Lavidaviridae* includes the Sputnik and Zamilon viruses and the Maverick-related virus (genus *Mavirus*); also called giant satellite viruses (17–30 kbp linear or circular dsDNA genomes). “*Lavida*” stands for large virus-dependent or -associated virus and refers to the property of Sputnik and other satellite viruses of depending on or associating with large dsDNA viruses. Sputnik virus was identified in an amoebal host in co-infection with its HV, *Acanthamoeba polyphaga mimivirus* (genus *Mimivirus*, family *Mimiviridae*). Though the icosahedral virus particle of Sputnik virus is only a tenth of the size of *Acanthamoeba polyphaga mimivirus*, the satellite virus may target the cytoplasmic replication factory of the giant virus and cause aberrant capsid phenotypes. As a result, the Sputnik virus was considered a viral parasite of a virus and named as “virophage”. This terminology has also been used to describe other satellites associated with giant viruses. Another typical feature of virophages is the presence of core genes in their genomes, a conserved set of six proteins or domains: major capsid protein (MCP), minor capsid protein (mCP),

ATPase, protease (PRO), helicase (HEL), and a zinc-ribbon-containing domain. Currently, the family *Lavidaviridae* is divided in the genera *Sputnikvirus* and *Mavirus* with two and one species, respectively (Table 2).

Satellite Nucleic Acids and Satellite-Related Nucleic Acids

Category 6a: Satellite ssDNA (family *Alphasatellitidae*; subfamilies *Nanoalphasatellitinae* and *Geminialphasatellitinae*)

The recently created family *Alphasatellitidae* consists of circular single stranded DNA (ssDNA) molecules of approximately 1–1.4 kb that are associated with nanoviruses (subfamily *Nanoalphasatellitinae*) and some geminiviruses that belong to the genera *Begomovirus* and *Mastrevirus* (subfamily *Geminialphasatellitinae*). Currently, the subfamily *Geminialphasatellitinae* consists of four genera (*Ageyiasatellite*, *Clecrusatellite*, *Colecusatellite*, and *Gosmusatellite*) and 43 established species (Table 3) based on genus and species demarcation thresholds of $\approx 70\%$ and $\approx 88\%$, respectively of their complete genomes. The subfamily *Nanoalphasatellitinae* is more variable with at least seven genera (*Babusatellite*, *Clostunsatellite*, *Fabenesatellite*, *Milvetsatellite*, *Mivedwarsatellite*, *Sophoyesatellite*, and *Subclovsatellite*) and 19 species (Table 3) using complete genome genus and species demarcation thresholds of $\approx 67\%$ and $\approx 80\%$, respectively. Collectively referred to as alphasatellites, these plant-infecting molecules resemble the DNA-R component of nanoviruses in that they encode a replication-associated protein (Rep) and are capable of autonomous replication in host cells. They also possess a stem-loop region containing the nonanucleotide TAA/GTATTAC. However, they depend on their ssDNA HVs for encapsidation, movement, and insect transmission. The specific role of alphasatellites in the pathogenesis of their HV is unclear but some published studies have implicated their co-infections in reduced or exacerbated symptoms of the HV, reduced titer of the HV or betasatellite, or suppression of host defenses based on RNA interference.

Category 6b: Satellite ssDNA (family *Tolecusatellitidae*; genera *Betasatellite* and *Deltasatellite*)

Different satellite ssDNAs have been classified under the family *Tolecusatellitidae* with two genera (*Betasatellite* and *Deltasatellite*) based on their nt sequences. Like members of the *Geminialphasatellitinae* subfamily, the circular ssDNA molecules of betasatellites are about half the sizes (≈ 1.3 kb) of their helper Old World monopartite begomoviruses, possess a stem-loop region containing the nonanucleotide TAATATTAC, and code for a single ORF (β C1), whose product has been shown to function as a pathogenicity determinant and suppressor of host gene silencing. Deltasatellites, however, are usually about one quarter (≈ 0.7 kb) of the size of the genome of their monopartite Old World begomoviruses and bipartite New World begomoviruses. Unlike other geminivirus-associated satellite ssDNA molecules, deltasatellites do not encode for any protein. Consequently, like betasatellites, they also rely on their HV for encapsidation, replication, movement, and vector transmission. Whereas betasatellites are known to enhance the virulence of the helper begomovirus and may be essential for the maintenance of disease in the field, deltasatellites may cause a reduction on begomovirus accumulation and/or symptom severity in certain satellites-virus-host interactions. Current species demarcation thresholds within the family stands at $<91\%$ complete nt sequence identity for both genera, leading to the recognition of 61 betasatellite and 11 deltasatellite species (Table 3).

Category 7: Satellite dsRNAs

Most of the satellite dsRNAs (Table 4) occur in association with viruses in the families *Totiviridae* and *Partitiviridae*, on which they depend for the encapsidation and replication of their 0.5–1.8 kbp genomes. The totivirus-associated satellite dsRNAs encode a secreted protein toxin (the killer toxin) that is lethal to virus-free sensitive cells and to the HV that lacks the satellite dsRNA molecule. In turn, the HV that carry the toxin-encoding satellite dsRNA has immunity to that toxin; thus, the satellite dsRNA confers ecological advantage to its HV by killing competing virus- or satellite-free fungi. Unlike their *Totiviridae* counterparts, the satellite dsRNAs associated with partitiviruses do not code for functional proteins and their biological significance is not known.

Category 8a: Large linear satellite ssRNAs

Members of the large linear satellite ssRNAs have messenger RNA properties. They range in size between 0.8 and 1.5 kb and encode a nonstructural protein that may be essential for satellite RNA multiplication. All but two of the currently listed members are satellites of viruses in the family *Secoviridae*, the exceptions being bamboo mosaic virus (BaMV) satellite RNA and BNYVV RNA5 that are associated with a potyvirus and a benyvirus, respectively (Table 4). Several members of the *Secoviridae*-associated large linear satellite ssRNAs have been studied. They share 5' and 3' structural features of the gRNAs of the HV (a 3' terminal poly(A) sequence and a 5' terminal VPg). The encoded non-structural protein of different satellite RNAs are basic proteins of 36–48 kDa. For the satellite RNAs of tomato black ring virus, grapevine fanleaf virus, and Arabis mosaic virus (ArMV) (all three: genus *Nepovirus*, family *Secoviridae*), it has been shown that the encoded proteins are needed for the replication of the satellite RNA. Most large satellite RNAs of nepoviruses do not seem to have an effect on the accumulation or pathogenicity of the HV. However, this may depend on the experimental system, as the large satellite RNA of ArMV was shown to modulate the symptoms of the HV depending on the species of host plant. The BaMV satellite RNA is the only member that is encapsidated into rod-shaped particles. It encodes a 20 kDa protein that is expressed *in vivo* but is not needed for satellite RNA replication. The presence of BaMV satellite RNA significantly reduces the accumulation of the HV's gRNA. Cis-acting sequences required for BaMV satellite RNA replication have been mapped at the 5' NTR and comprise a stem-loop structure that is conserved among BaMV satellite RNA variants. The BNYVV RNA5 encode a protein of 26 kDa, which is responsible for intensification of the rhizomania disease induced by infection of BNYVV RNAs 1 and 2, plus its two satellite-like RNAs 3 and 4.

Category 8b: Small linear satellite ssRNAs

The group consists of small (<0.7 kb), linear satellite RNA molecules that are associated with plant-infecting HVs in the families *Tombusviridae*, *Bromoviridae*, and the genus *Umbravirus* (Table 4). Some of the *Umbravirus*-associated members such as groundnut rosette virus satellite RNA and tobacco bushy top virus satellite RNA may be regarded as satellite-like RNAs due to their

Table 3 Classification of ssDNA nucleic acid satellites

Family/Subfamily	Genus	Species	Type species accession #	Genome/Host types ^a	Helper virus genus
<i>Alphasatellitidae</i>					
<i>Geminialphasatellitinae</i>	<i>Ageyesisatellite</i>	<i>Ageratum yellow vein Singapore alphasatellite</i>	AJ416153	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Ageyesisatellite</i>	<i>Cotton leaf curl Saudi Arabia alphasatellite</i>	HG530543	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Clecrusatellite</i>	<i>Cleome leaf crumple alphasatellite</i>	FN436007	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Clecrusatellite</i>	<i>Croton yellow vein mosaic alphasatellite</i>	FN658711	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Clecrusatellite</i>	<i>Euphorbia yellow mosaic alphasatellite</i>	FN436008	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Clecrusatellite</i>	<i>Melon chlorotic mosaic alphasatellite</i>	HM163578	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Clecrusatellite</i>	<i>Sida Cuba alphasatellite</i>	HE806451	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Clecrusatellite</i>	<i>Tomato yellow spot alphasatellite</i>	KX348228	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Clecrusatellite</i>	<i>Whitefly associated Guatemala alphasatellite 2</i>	KT099170	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Clecrusatellite</i>	<i>Whitefly associated Puerto Rico alphasatellite 1</i>	KT099173	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Ageratum enation alphasatellite</i>	JX913532	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Ageratum yellow vein alphasatellite</i>	AJ238493	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Ageratum yellow vein China alphasatellite</i>	KF785752	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Ageratum yellow vein India alphasatellite</i>	JX570736	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Bhendi yellow vein alphasatellite</i>	FN658716	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Cassava mosaic Madagascar alphasatellite</i>	HE984148	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Chili leaf curl alphasatellite</i>	KF471043	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Cotton leaf curl Egypt alphasatellite</i>	AJ512960	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Cotton leaf curl Gezira alphasatellite</i>	EU589450	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Cotton leaf curl Lucknow alphasatellite</i>	AJ132344	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Cotton leaf curl Multan alphasatellite</i>	HQ343234	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Gossypium darwinii symptomless alphasatellite</i>	EU384623	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Malvastrum yellow mosaic alphasatellite</i>	FN675297	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Malvastrum yellow mosaic Cameroon alphasatellite</i>	AM236765	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Pedilanthus leaf curl alphasatellite</i>	KX168428	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Sida leaf curl alphasatellite</i>	FR772088	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Sida yellow vein Vietnam alphasatellite</i>	DQ641718	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Sunflower leaf curl Karnataka alphasatellite</i>	JX569789	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Synedrella leaf curl alphasatellite</i>	KJ939346	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Tobacco curly shoot alphasatellite</i>	HQ407396	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Tomato leaf curl Buea alphasatellite</i>	FN675299	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Tomato leaf curl Cameroon alphasatellite</i>	FN675296	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Tomato yellow leaf curl China alphasatellite</i>	AJ579359	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Tomato yellow leaf curl Thailand alphasatellite</i>	AM749493	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Colecusatellite</i>	<i>Tomato yellow leaf curl Yunnan alphasatellite</i>	KX759649	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Gosmusatellite</i>	<i>Gossypium mustelinum symptomless alphasatellite</i>	EU384656	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Gosmusatellite</i>	<i>Hollyhock yellow vein alphasatellite</i>	FR772086	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Gosmusatellite</i>	<i>Mesta yellow vein mosaic alphasatellite</i>	JX183090	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Gosmusatellite</i>	<i>Okra enation leaf curl alphasatellite</i>	HF546575	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Gosmusatellite</i>	<i>Okra yellow crinkle Cameroon alphasatellite</i>	FN675285	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	<i>Gosmusatellite</i>	<i>Vernonia yellow vein Fujian alphasatellite</i>	KC959931	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	Unassigned	<i>Dragonfly associated alphasatellite</i>	JX458742	ssDNA/Plant	<i>Begomovirus</i>
<i>Geminialphasatellitinae</i>	Unassigned	<i>Whitefly associated Guatemala alphasatellite 1</i>	KT099172	ssDNA/Plant	<i>Begomovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Babusatellite</i>	<i>Banana bunchy top alphasatellite 1</i>	HQ616080	ssDNA/Plant	<i>Babuvirus</i>
<i>Nanoalphasatellitinae</i>	<i>Babusatellite</i>	<i>Banana bunchy top alphasatellite 2</i>	L32167	ssDNA/Plant	<i>Babuvirus</i>
<i>Nanoalphasatellitinae</i>	<i>Babusatellite</i>	<i>Banana bunchy top alphasatellite 3</i>	AF416471	ssDNA/Plant	<i>Babuvirus</i>
<i>Nanoalphasatellitinae</i>	<i>Babusatellite</i>	<i>Cardamom bushy dwarf alphasatellite</i>	KF435148	ssDNA/Plant	<i>Babuvirus</i>
<i>Nanoalphasatellitinae</i>	<i>Clostunsatellite</i>	<i>Milk vetch dwarf alphasatellite 2</i>	AB000922	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Clostunsatellite</i>	<i>Pea necrotic yellow dwarf alphasatellite 2</i>	KC979052	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Clostunsatellite</i>	<i>Sophora yellow stunt alphasatellite 4</i>	KX534409	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Clostunsatellite</i>	<i>Sophora yellow stunt alphasatellite 5</i>	KX534398	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Clostunsatellite</i>	<i>Subterranean clover stunt alphasatellite 2</i>	U16735	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Fabenesatellite</i>	<i>Faba bean necrotic yellows alphasatellite 2</i>	AJ005966	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Milvetsatellite</i>	<i>Milk vetch dwarf alphasatellite 3</i>	AB009047	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Mivedwarsatellite</i>	<i>Faba bean necrotic stunt alphasatellite</i>	KC978990	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Mivedwarsatellite</i>	<i>Milk vetch dwarf alphasatellite 1</i>	AB000920	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Mivedwarsatellite</i>	<i>Pea necrotic yellow dwarf alphasatellite 1</i>	KC979051	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Mivedwarsatellite</i>	<i>Sophora yellow stunt alphasatellite 2</i>	KX534408	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Sophyesatellite</i>	<i>Sophora yellow stunt alphasatellite 3</i>	KX534400	ssDNA/Plant	<i>Nanovirus</i>

Table 3 Continued

Family/Subfamily	Genus	Species	Type species accession #	Genome/Host types ^a	Helper virus genus
<i>Nanoalphasatellitinae</i>	<i>Subclovsatellite</i>	<i>Faba bean necrotic yellows alphasatellite 1</i>	X80879	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Subclovsatellite</i>	<i>Sophora yellow stunt alphasatellite 1</i>	KX534399	ssDNA/Plant	<i>Nanovirus</i>
<i>Nanoalphasatellitinae</i>	<i>Subclovsatellite</i>	<i>Subterranean clover stunt alphasatellite 1</i>	U16731	ssDNA/Plant	<i>Nanovirus</i>
Unassigned	Unassigned	<i>Coconut foliar decay alphasatellite</i>	M29963	ssDNA/Plant	<i>Nanovirus</i>
<i>Tolecusatellitidae</i>					
	<i>Betasatellite</i>	<i>Ageratum leaf curl Buea betasatellite</i>	FR717140	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Ageratum leaf curl Cameroon betasatellite</i>	FM164737	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Ageratum yellow leaf curl betasatellite</i>	AJ316026	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Ageratum yellow vein betasatellite</i>	AJ252072	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Ageratum yellow vein India betasatellite</i>	AJ557441	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Ageratum yellow vein Sri Lanka betasatellite</i>	AJ542498	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Alternanthera yellow vein betasatellite</i>	DQ641716	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Andrographis yellow vein leaf curl betasatellite</i>	KC967282	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Bhendi yellow vein mosaic betasatellite</i>	AJ308425	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Cardiospermum yellow leaf curl betasatellite</i>	AM933578	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Chili leaf curl betasatellite</i>	AJ316032	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Chili leaf curl Jaunpur betasatellite</i>	HM007103	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Chili leaf curl Sri Lanka betasatellite</i>	JN638445	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Cotton leaf curl Gezira betasatellite</i>	DQ644564	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Cotton leaf curl Multan betasatellite</i>	AJ298903	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Croton yellow vein mosaic betasatellite</i>	AM410551	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Eupatorium yellow vein betasatellite</i>	AJ438938	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Eupatorium yellow vein mosaic betasatellite</i>	AB300464	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>French bean leaf curl betasatellite</i>	JQ866298	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Hedyotis yellow mosaic betasatellite</i>	KF641186	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Honeysuckle yellow vein betasatellite</i>	AJ316040	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Honeysuckle yellow vein mosaic betasatellite</i>	AB182263	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Malvastrum leaf curl betasatellite</i>	AM072289	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Malvastrum leaf curl Guangdong betasatellite</i>	KF912951	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Mirabilis leaf curl betasatellite</i>	KT454829	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Momordica yellow mosaic betasatellite</i>	LK054803	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Mungbean yellow mosaic betasatellite</i>	JX443646	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Okra leaf curl Oman betasatellite</i>	KF267444	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Papaya leaf curl betasatellite</i>	AY244706	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Papaya leaf curl China betasatellite</i>	KJ642219	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Papaya leaf curl India betasatellite</i>	HM143906	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Rhynchosia yellow mosaic betasatellite</i>	GQ478344	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Rose leaf curl betasatellite</i>	KP752092	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Siegesbeckia yellow vein betasatellite</i>	KF499590	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tobacco curly shoot betasatellite</i>	AJ421484	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tobacco leaf curl betasatellite</i>	AM260465	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tobacco leaf curl Japan betasatellite</i>	HQ180394	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tobacco leaf curl Patna betasatellite</i>	AJ316036	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Bangalore betasatellite</i>	AY428768	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Bangladesh betasatellite</i>	AB236324	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl betasatellite</i>	AJ542489	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl China betasatellite</i>	AJ704609	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Gandhinagar betasatellite</i>	KC952006	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Java betasatellite</i>	KC282642	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Joydebpur betasatellite</i>	AJ966244	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Laguna betasatellite</i>	AB307732	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Laos betasatellite</i>	AJ542491	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Malaysia betasatellite</i>	KM051528	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Nepal betasatellite</i>	AJ542492	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Patna betasatellite</i>	EU862324	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Philippine betasatellite</i>	AB308071	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Sri Lanka betasatellite</i>	AJ542493	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato leaf curl Yemen betasatellite</i>	JF919717	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato yellow leaf curl China betasatellite</i>	AJ420313	ssDNA/Plant	<i>Begomovirus</i>

(Continued)

Table 3 Continued

Family/Subfamily	Genus	Species	Type species accession #	Genome/Host types ^a	Helper virus genus
	<i>Betasatellite</i>	<i>Tomato yellow leaf curl Rajasthan betasatellite</i>	AY438558	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato yellow leaf curl Shandong betasatellite</i>	KP322555	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato yellow leaf curl Thailand betasatellite</i>	AJ566746	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato yellow leaf curl Vietnam betasatellite</i>	DQ641714	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Tomato yellow leaf curl Yunnan betasatellite</i>	KF640694	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Vernonia yellow vein betasatellite</i>	JF733779	ssDNA/Plant	<i>Begomovirus</i>
	<i>Betasatellite</i>	<i>Vernonia yellow vein Fujian betasatellite</i>	FN435836	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Croton yellow vein deltasatellite</i>	AJ968684	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Malvastrum leaf curl deltasatellite</i>	KF433066	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Sida golden yellow vein deltasatellite 1</i>	JN986808	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Sida golden yellow vein deltasatellite 2</i>	JN819490	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Sida golden yellow vein deltasatellite 3</i>	JN819498	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Sweet potato leaf curl deltasatellite 1</i>	FJ914390	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Sweet potato leaf curl deltasatellite 2</i>	KF716173	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Sweet potato leaf curl deltasatellite 3</i>	KT099179	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Tomato leaf curl deltasatellite</i>	U74627	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Tomato yellow leaf distortion deltasatellite 1</i>	JN819495	ssDNA/Plant	<i>Begomovirus</i>
	<i>Deltasatellite</i>	<i>Tomato yellow leaf distortion deltasatellite 2</i>	KU232893	ssDNA/Plant	<i>Begomovirus</i>

^assDNA, single-stranded DNA.

essential roles as components of disease complexes. Although they do not encode functional proteins, some members can attenuate or exacerbate the symptoms induced by HV infection.

Category 8c: Small circular satellite ssRNAs

Members have small (0.35 kb) genomes that may be present in both circular and linear forms in infected cells. They include satellite RNAs associated with viruses in the plant virus families *Secoviridae*, *Luteoviridae*, and *Solemoviridae*; and possibly a viroid-like satellite RNA (Table 4). There is no mRNA activity associated with these various satellite RNAs, and they all have a high degree of secondary structure. The satellite RNAs either have no effect on (*Solemoviridae*-associated satellites) or reduce the accumulation of (*Secoviridae*- and *Luteoviridae*-associated satellites) their HV.

Category 8d: Hepadnavirus-associated satellite-like RNAs (genus *Deltavirus*)

Hepatitis delta virus (HDV, unassigned genus *Deltavirus*) is a subviral agent of the human pathogen, hepatitis B virus (HBV, genus *Orthohepadnavirus*, family *Hepadnaviridae*). HDV is replicated in the nucleus by the host DNA-dependent RNA polymerase II, but is dependent upon the envelope proteins of HBV for its encapsidation and transmission, and is thus not a satellite. The HDV RNA (1679 nt) folds into an unbranched rod-like structure similar to that of viroids, and in common with the circular ssRNA satellites, contains ribozymes that are involved in the processing and maturation of this RNA during replication. The HDV RNA encodes two forms of a protein designated the delta antigen (δ Ag). The smaller, 22 kDa form (δ Ag-S), which is a nuclear phosphoprotein, is required for replication of HDV RNA, while the 24 kDa larger form (δ Ag-L), which contains an additional 19 aa at the C-terminus and is produced late in infection, is required along with δ Ag-S for assembly of the HDV RNA into HBV envelope particles. δ Ag-L also acts as an inhibitor of HDV replication. HDV intensifies the severity of liver disease caused by HBV.

Category 8e: Polerovirus-associated RNAs

Some isolates of three poleroviruses (genus *Polerovirus*, family *Luteoviridae*), *Beet western yellows virus*, carrot red leaf virus (CRLV), and *Tobacco vein distorting virus* were found to have associated with them additional RNAs of size 2.8–3 kb. These associated RNAs encoded their own viral replicases and are thus capable of autonomous replication. However, they each depend on their respective helper poleroviruses for movement, encapsidation, and transmission. Some members, such as CRLV-associated RNA, contribute to the disease caused by their HV.

Replication and Structure

Satellite RNAs

A general property of satellite RNAs is that their replication depends on the replication machinery of the HV, which involves virus- and host-encoded factors. Hence, replication of satellite RNAs depends on interactions with both the HV and the host plant. Satellite RNAs may or may not share structural features at their 5' and 3' termini in common with the HV RNA. For instance, the large satellite RNAs of the nepoviruses have an HV-encoded VPg at the 5' end and a poly(A) tail at the 3' end, as do their HV RNAs, but the satellite RNAs of cucumber mosaic virus (CMV) have a 5' cap structure, as do the CMV RNAs. But unlike the HV, the satellite RNAs do not have a tRNA-like structure that can be validated at the 3' end. Structural differences and, often, differences in the replication process between the satellite RNA and the HV RNA, indicate that the replication machinery of the HV may need to

Table 4 Classification of satellites RNA nucleic acids

Category/Species and Tentative Species	Type isolate ^a accession #	Genome/Host types ^b	Helper virus genus
7. Satellite dsRNAs			
Bombyx mori cypovirus 1 satellite RNA	AB183384	dsRNA/Fungi	Cypovirus
Satellite of Atkinsonella hypoxylon virus	L39127	dsRNA/Fungi	Betapartitivirus
Satellite of Gremmeniella abietina virus MS1	AY089995	dsRNA/Fungi	Gammapartitivirus
Satellite of Penicillium stoloniferum virus F	AY738338	dsRNA/Fungi	Gammapartitivirus
Satellite of Zygosaccharomyces bailii virus	AF515592	dsRNA/Fungi	Zyobavirus
Satellites of Amasya cherry disease-associated virus: Sat A, Sat B	AM085138, AM085139	dsRNA/Fungi	Alphachrysovirus
Satellites of cherry chlorotic rusty spot -associated virus: Sat A, Sat B, Sat C	N/A	dsRNA/Fungi	Alphapartitivirus
Satellites of Discula destructiva virus 1: dsRNA 3, dsRNA4	AF316994, AF316995	dsRNA/Fungi	Gammapartitivirus
Satellites of Saccharomyces cerevisiae L-A virus: M1, M2, M28	U78817, X56987, N/A	dsRNA/Fungi	Totivirus
Satellites of Trichomonas vaginalis T1 virus	U15991	dsRNA/Fungi	Trichomonasvirus
Satellites of Ustilago maydis virus H: M-P1, M-P4, M-P6	M63149, L12226, P16948	dsRNA/Fungi	Totivirus
8a. Large linear satellite ssRNAs			
Arabis mosaic virus large satellite RNA	D00664	ssRNA/Plant	Nepovirus
Bamboo mosaic virus satellite RNA	L22762	ssRNA/Plant	Potexvirus
Beet necrotic yellow vein virus RNA5*	D63759	ssRNA/Plant	Benyvirus
Beet ringspot virus satellite RNA	KX033801	ssRNA/Plant	Nepovirus
Blackcurrant reversion virus satellite RNA	AF112119	ssRNA/Plant	Nepovirus
Chicory yellow mottle virus large satellite RNA	D00686	ssRNA/Plant	Nepovirus
Grapevine Bulgarian latent virus satellite RNA	N/A	ssRNA/Plant	Nepovirus
Grapevine fanleaf virus satellite RNA	D00442	ssRNA/Plant	Nepovirus
Myrobalan latent ringspot virus satellite RNA	N/A	ssRNA/Plant	Nepovirus
Strawberry latent ringspot virus satellite RNA	X69826	ssRNA/Plant	Nepovirus
Tomato black ring virus satellite RNA	X00978	ssRNA/Plant	Nepovirus
8b. Small linear satellite ssRNAs			
Artichoke mottled crinkle virus satellite RNA	N/A	ssRNA/Plant	Tombusvirus
Beet black scorch virus satellite RNA	AY394497	ssRNA/Plant	Betanecrovirus
Carnation Italian ringspot virus satellite RNA	N/A	ssRNA/Plant	Tombusvirus
Carrot mottle mimic virus satellite RNA	EU914919	ssRNA/Plant	Umbravirus
Cucumber mosaic virus satellite RNA (several types)	X69136	ssRNA/Plant	Cucumovirus
Cymbidium ringspot virus satellite RNA	D00720	ssRNA/Plant	Tombusvirus
Groundnut rosette virus satellite RNA*	Z29702	ssRNA/Plant	Umbravirus
Panicum mosaic virus satellite RNA	M17182	ssRNA/Plant	Panicovirus
Pea enation mosaic virus satellite RNA	U03564	ssRNA/Plant	Umbravirus
Peanut stunt virus satellite RNA	Z98197	ssRNA/Plant	Cucumovirus
Pelargonium leaf curl virus satellite RNA	N/A	ssRNA/Plant	Tombusvirus
Petunia asteroid mosaic virus satellite RNA	N/A	ssRNA/Plant	Tombusvirus
Tobacco bushy top virus satellite RNA*	N/A	ssRNA/Plant	Umbravirus
Tobacco necrosis virus small satellite RNA	E03054	ssRNA/Plant	Alphanecrovirus
Tomato bushy stunt virus satellite RNA (several types)	AF022787–8, FJ666076	ssRNA/Plant	Tombusvirus
8c. Small circular satellite ssRNAs			
Arabis mosaic virus small satellite RNA	M21212	ssRNA/Plant	Nepovirus
Cereal yellow dwarf virus-RPV satellite RNA	M63666	ssRNA/Plant	Polerovirus
Cherry small circular viroid-like RNA	Y12833	ssRNA/Plant	
Chicory yellow mottle virus satellite RNA	D00721	ssRNA/Plant	Nepovirus
Lucerne transient streak virus satellite RNA	X01984	ssRNA/Plant	Sobemovirus
Rice yellow mottle virus satellite	AF039909	ssRNA/Plant	Sobemovirus
Solanum nodiflorum mottle virus satellite RNA	J02386	ssRNA/Plant	Sobemovirus
Subterranean clover mottle virus satellite RNA (2 types)	M33000, M33001	ssRNA/Plant	Sobemovirus
Tobacco ringspot virus satellite RNA	M14879	ssRNA/Plant	Nepovirus
Velvet tobacco mottle virus satellite RNA	J02439	ssRNA/Plant	Sobemovirus
8d. Hepadnavirus-associated RNAs			
Hepatitis delta virus	M21012	ssRNA/Animal	Orthohepadnavirus
8e. Polerovirus-associated RNAs			
Beet western yellows virus ST9-associated RNA	L04281	ssRNA/Plant	Polerovirus
Carrot red leaf virus-associated RNA	AF020617	ssRNA/Plant	Polerovirus
Tobacco vein distorting virus-associated RNA	EF529625	ssRNA/Plant	Polerovirus

^aN/A, unavailable.^bssRNA, single-stranded RNA; dsRNA, double-stranded RNA.

be adapted to replicate the satellite RNA, in ways which are not completely understood. The HV replication complex could be modified by satellite-encoded factors, as has been hypothesized for the large satellite RNAs of nepoviruses, or by unidentified host factors. In the best characterized systems, the efficiency of replication depends on the host plant. For CMV satellite RNAs, replication efficiency also depends on the strain of HV. Similarly, while the expression of the HV RNA-dependent RNA polymerase was enough for the replication of cereal yellow dwarf virus (CYDV) satellite RNA in the homologous host, this was not the case for Cymbidium mosaic virus satellite RNA in a heterologous host system, in which, however, a diRNA was amplified.

Replication has been studied best in the small noncoding satellite RNAs. For the small linear satellite RNAs of TCV and CMV, multimeric forms of both positive (arbitrarily defined as the encapsidated sense) and negative sense are found in infected tissues. The junction between monomers can be perfect or have deletions. Circular forms are not found, and replication does not proceed through a rolling-circle mechanism. Regulatory sequences for RNA replication have been mapped in detail in the satellite RNAs of TCV. In the hybrid satC molecule, a hairpin structure is a replication enhancer and has a role in depressing the accumulation of the HV. Also, structural motifs favoring recombination have been analyzed extensively in satD and satC of TCV.

The replication of small, circular satellite RNAs is by a rolling-circle mechanism; upon infection, linear, multimeric forms of the negative strand are synthesized. The multimeric plus strand is then synthesized on this template or on negative strand circular monomers, depending on the satellite RNA. The positive strand multimer is cleaved autocatalytically and ligation occurs for those satellite RNAs that are encapsidated as a circular form. Hammerhead and hairpin ribozyme structures are found in these satellite RNAs monomers or dimers, which catalyze hydrolysis or hydrolysis and ligation, respectively. Other structures required for replication and encapsidation have been mapped in CYDV satellite RNA.

Structural analyses have been done mostly with small satellite RNAs. Models for the *in vitro* secondary structure have been proposed for several satellite RNAs (CMV satellite RNA, TCV satellite RNA, TRSV satellite RNA, and CYDV satellite RNA) based on nuclease sensitivity and chemical modification of bases data. For CMV satellite RNA, *in vivo* models have also been proposed. All analyzed satellite RNAs have a high degree of secondary structure, with more than 50% and up to 70% of bases involved in pairing. The high degree of secondary structure could explain the high stability of these molecules as well as the very high infectivity reported for some of them, such as CMV satellite RNA and TCV satellite RNA. The high degree of secondary structure may also be related to their biological activity, which for these noncoding molecules must depend on structural features. As has been detailed in other sections, sequences, structural domains, and tertiary interactions involved in pathogenicity and replication have been well characterized for some small satellite RNAs.

Satellite DNAs

The replication of satellite ssDNA mimics that of the HV. Unlike RNA viruses, geminiviruses do not encode a polymerase and depend on host DNA polymerases for replication. The only geminivirus-encoded protein essential for replication is Rep. Satellite DNAs utilize the HV Rep and contain the conserved sequence for DNA nicking by the Rep. Replication takes place by a combination of rolling circle and recombination dependent mechanisms and the DNA is encapsidated in the HV virion. The replication is independent of β C1 expression. Sequence comparisons at the nucleotide and aa levels reveal several conserved structural features, despite a significant level of sequence divergence. These include an A-rich region, a sequence of about 80 nt referred to as the satellite conserved region, and putative stem-loop structure containing a mononucleotide motif, also present in geminivirus genomes.

Sequence Variation and Evolution

Sequence variation and evolution has been analyzed most for CMV satellite RNA. Early experiments under controlled conditions showed a high variability and genetic plasticity. Populations started from cDNA clones rapidly evolved to a swarm of sequences whose master sequence changed upon passage on different hosts. Analysis of genetic variants from natural populations showed again a high diversity for CMV satellite RNA. In natural populations, satellite RNA diversification and evolution proceeded by mutation accumulation and by recombination. Constraints to genetic variation were analyzed, and were related to the maintenance of base pairing in secondary structure elements. Population diversity of CMV satellite RNA was higher than for the HV, and the population structures of CMV and the satellite RNA were not correlated. These analyses also showed that CMV satellite RNA behaved in nature as a molecular hyperparasite spreading epidemically on the CMV populations. The incidence of CMV satellite RNA in CMV populations has been shown to be low, except during episodes of epidemics of tomato necrosis. High diversity of natural populations has also been shown for panicum mosaic satellite virus RNA and for BaMV satellite RNA. For BaMV satellite RNA, the diversity was highest in the NTRs. Conversely, the population of rice yellow mottle virus satellite RNAs across sub-Saharan Africa showed little sequence diversity. The incidence of these three satellite RNAs in the analyzed populations of the respective HV was, in all cases, high.

Phylogenetic analyses divide geminivirus-related satellites broadly into two categories, those isolated from host plants of the family Malvaceae and the rest isolated from other plants, mostly in the Solanaceae. Compared to satellite RNAs, satellite DNAs do not exhibit a wide diversity. They are structurally similar and lack strict specificity for HV species. Geminivirus-related satellite DNA sequences analyzed exhibit a minimum overall similarity of about 47% and 37%, at the nt level and aa level, respectively.

There has been much speculation on the origin of satellite RNAs. For CMV satellite RNA and TCV satellite RNA, it has been suggested that they could have been generated out of small sequences synthesized upon the HV RNA as a template by the HV

polymerase. This hypothesis is no longer sustained since attempts to locate the sequences with significant similarity to satellite RNA in the genomes of hosts, HVs and vectors have in general been unsuccessful. Whatever the mechanism of generation of the satellite RNAs, it seems that the satellitism as a phenomenon, that is, the dependence on a certain virus for replication, has evolved independently several times. This is suggested by the lack of correlation between satellite and HV taxonomy, and by the existence of subviral nucleic acids with different degrees of dependence on an HV. For instance, an evolutionary line from nondependent viruses to satellites such as BaMV satellite RNA through satellite viruses could have proceeded by size and information content reduction. A phylogenetic relationship between viroids and small circular satellite RNAs has also been proposed.

Expression of Foreign Sequences From Satellite Vectors

The satellite RNAs of BaMV and BNYVV, as well as the satellite DNA of tomato yellow leaf curl China virus isolate Y10 (TYLCCNV-Y10), all of which express proteins not required for the replication or spread of the satellite or HV, have been used as expression vectors of foreign sequences. In the case of the satellite RNA of BaMV, the expression of such sequences and the accumulation of the satellite RNA were reduced considerably in systemically infected leaves. Nevertheless, this satellite expression system has been useful for studying cis- and trans-acting replication signals. The 26 kDa protein encoded by RNA5 of BNYVV has been replaced with the sequence encoding the green fluorescent protein, which was expressed in both the inoculated leaves and systemically infected leaves. The small size of the satellite DNA of TYLCCNV-Y10 precludes use of this system for expression of most genes, but the system has been used to express plant gene segments inducing RNA silencing in several plants species. The STMSV system also has been used to express plant gene sequences, in place of its CP, resulting in RNA silencing of those plant genes, but not of the satellite virus or its HV. The modified STMSV expression vector was able to spread through the plant in the absence of its CP.

Satellite-Mediated Control of Viruses

The ability of some satellite RNAs to interfere with the replication of, and attenuate the symptoms induced by, their HV has been exploited to develop strategies for disease control. Two prophylactic strategies have been employed for this biocontrol approach. In the first strategy (cross-protection), the target plants are pre-inoculated with a combination of the satellite RNA and a mild strain of its HV to “vaccinate” the plant against a virulent strain of the HV. The second strategy utilizes the transgenic approach wherein the satellite RNA have been expressed in transgenic plants to provide either resistance to infection by the respective HV, or tolerance to the disease induced by the HV and/or related pathogenic satellites. Multiple mechanisms like competition for viral replicase, gene silencing and possibly other immune actions of the host may be involved in resistance. A recent study evaluated the effects of post-inoculation with CMV satellite RNA on symptoms derived from pre-infection by its HV and demonstrated that this new therapeutic strategy could be useful for combating plant viruses using properly timed treatment with attenuation satellite RNAs.

See also: Alphasatellites (*Alphasatellitidae*). Beet Necrotic Yellow Vein Virus (*Benyviridae*). Betasatellites and Deltasatellites (*Tolecusatellitidae*). Geminiviruses (*Geminiviridae*). Luteoviruses (*Luteoviridae*). Machlomovirus and Panicoviruses (*Tombusviridae*). Nanoviruses (*Nanoviridae*). Plant Satellite Viruses (*Albetovirus*, *Aumavirus*, *Papanivirus*, *Virtovirus*). Secoviruses (*Secoviridae*). Tombusviruses (*Tombusviridae*).

Further Reading

- Bonami, J.-R., Sri Widadha, J., 2011. Viral diseases of the giant freshwater prawn *Macrobrachium rosenbergii*: A review. *Journal of Invertebrate Pathology* 106, 131–142.
- Briddon, R.W., Martin, D.P., Roumagnac, P., *et al.*, 2018. Alphasatellitidae: A new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Archives of Virology* 163, 2587–2600.
- Cao, X., Liu, S., Yu, C., Li, X., Yuan, X., 2019. A new strategy of using satellite RNA to control viral plant diseases: Post-inoculation with satellite RNA attenuates symptoms derived from pre-infection with its helper virus. *Plant Biotechnology Journal* 17, 1856–1858.
- Cui, H.G., Liu, H.Z., Chen, J., *et al.*, 2015. Genetic diversity of prunus necrotic ringspot virus infecting stone fruit trees grown at seven regions in China and differentiation of three phylogroups by multiplex RT-PCR. *Crop Protection* 74, 30–36.
- Dodds, J.A., 1998. Satellite tobacco mosaic virus. *Annual Review of Phytopathology* 36, 295–310.
- Dry, I.B., Krake, L.R., Rigden, J.E., Rezaian, M.A., 1997. A novel subviral agent associated with a geminivirus: The first report of a DNA satellite. *PNAS* 94, 7088–7093.
- Gosselé, V., Metzlaiff, M., 2006. Using satellite Tobacco mosaic virus vectors for gene silencing. *Current Protocols in Microbiology* 00 (1), 161.5.1–161.5.17.
- Idris, A.M., Shahid, M.S., Briddon, R.W., *et al.*, 2011. An unusual alphasatellite associated with monopartite begomoviruses attenuates symptoms and reduces betasatellite accumulation. *Journal of General Virology* 92, 706–717.
- Kassanis, B., 1962. Properties and behaviour of a virus depending for its multiplication on another. *Microbiology* 27, 477–488.
- Kassanis, B., White, R.F., 1972. Interference between two satellite viruses of tobacco necrosis virus. *Journal of General Virology* 17, 177–183.
- Krupovic, M., Kuhn, J.H., Fischer, M.G., 2016. A classification system for virophages and satellite viruses. *Archives of Virology* 161, 233–247.
- Rivière, M., Olivier, V., Blanchard, P., 2010. Chronic bee paralysis: A disease and a virus like no other? *Journal of Invertebrate Pathology* 103, S120–S131.
- Schneider, I.R., 1969. Satellite-like particle of Tobacco ringspot virus that resembles Tobacco ringspot virus. *Science* 166, 1627–1629.
- Zhou, X., 2013. Advances in understanding begomovirus satellites. *Annual Review of Phytopathology* 51, 357–381.
- Zinn, E., Vandenberghe, L.H., 2014. Adeno-associated virus: Fit to serve. *Current Opinion in Virology* 8, 90–97.

Secoviruses (*Secoviridae*)

Jeremy R Thompson, Cornell University, Ithaca, NY, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
AGO Argonaute 1
CITE Cap-independent translation enhancer
Co-Pro Protease cofactor
CP Coat protein or capsid protein
ELISA Enzyme-linked immunological assays
ER Endoplasmic reticulum
HC-Pro Helper component-protease
IRES Internal ribosomal entry site
kb Kilobase
kDa Kilo dalton
LAMP Loop mediated amplification

MP Movement protein
nt Nucleotide(s)
ORF Open reading frame
RdRp RNA-dependent RNA polymerase
RISC RNA-induced silencing complex
satRNA Satellite RNA
UTR Un-translated region
VIGS Virus-induced gene silencing
VLPs Virus-like particles
VPg Viral protein genome-linked
VRC Virus replication complexes
vRNA Virion RNA

Glossary

Affimers Small proteins that mimic antibodies.
Argonaute 1 Member of the Argonaute family of proteins involved in RNA silencing and a component of the RNA-induced silencing complex.
Glutamic protease A proteolytic enzyme that contains a glutamic acid residue within the active site.
Internal ribosomal entry site An RNA element that facilitates cap-independent translation.
Satellite RNA RNA that depends on the helper virus for replication and encapsidation, and can in some cases modulate symptoms.

Viral protein genome-linked A protein covalently attached to the 5' end of positive strand viral RNA and acting as a primer during RNA synthesis.
Virus re-assortment The mixing of the genome segments of a virus into new combinations.
Virus replication complexes Intracellular complexes formed by association with intracellular membranes that are sites for stages in the replication cycle of a virus.
 β -barrel Beta-sheet composed of tandem repeats forming a closed structure.

Introduction

Secoviruses are the plant-infecting members of the taxonomic order *Picornavirales*, a group of viruses that includes more broadly known members such as poliovirus, rhinovirus and foot-and-mouth disease virus. Symptoms of Grapevine fanleaf virus (GFLV) were some of the first to be described independently across a number of European countries at the end of the nineteenth century. In the early 1960s the plant virus group to which GFLV belongs, the nematode-transmissible nepoviruses, was one of the first to be recognized, followed by, in subsequent decades, the related comovirus and fabavirus groups. Unlike all other members of the *Picornavirales* order which contain their genetic information on one positive sense strand of RNA (monopartite), the comoviruses, fabaviruses and nepoviruses are segmented, meaning they have their genome split in two. This organization is typical for the vast majority of secoviruses that have been characterized so far. However, members of the sequivirus and waikavirus groups – the latter represented by the type member Rice tungro spherical virus (RTSV), a causal component of the eponymous disease that devastated South East Asia in the 1960s and 70s – are the exceptions with an unsegmented monopartite genome.

The secoviruses, like all members of the *Picornavirales* have a number of common features. Their genomic segments produce polyproteins (contiguous clusters of functional domains) which are proteolytically cleaved into smaller functional proteins by a virus protease which is encoded within a commonly conserved helicase-protease-polymerase replication block. They are all predicted to encode a small protein (VPg) that binds to the 5' end of their genome. Virus particles are non-enveloped and approximately 30 nm in diameter. The structure of Cowpea mosaic virus, one of the frontrunners in the development of nanobiotechnological tools, is particularly well-characterized.

Since the 2000s, exceptions to the general rule that genomic RNA segments encode single polyproteins within the *Picornavirales* order have been identified. For secoviruses, evidence has come with the emergence and characterization of the torradoviruses which are apparently bicistronic for one of their genome segments, encoding a predicted protein of as yet unknown function. Another exception to assumed commonalities has been the recent discovery that Strawberry mottle virus (SMoV) encodes for a glutamic protease, in addition to the 3C-like protease conserved across all other secoviruses. The former protease is the first of its kind reported for any positive single stranded RNA virus.

Taxonomy, Phylogeny, and Evolution

The International Committee for the Taxonomy of Viruses (ICTV) (see “Relevant Websites section”) lists specific demarcation criteria for the recognition of taxa which are to be used as guidelines on a case-by-case basis for determining taxonomies.

The family *Secoviridae*, is classified within the *Picornavirales* order (a name derived from fusing pico meaning ‘small’ prefixed to ‘RNA’), a group of viruses which encode their genome on single stranded positive sense RNA. The *Secoviridae* is the only plant-infecting member of this group. Other family members (along with host and typical species) include; *Dicistroviridae* (arthropods – Cricket paralysis virus), *Iflaviridae* (arthropods – Infectious flacherie virus), *Marnaviridae* (algae – *Heterosigma akashiwo* RNA virus), *Picornaviridae* (vertebrates – Enterovirus C), *Polycipiviridae* (arthropods – *Solenopsis invicta* virus 2). There are also two floating genera; *Bacillarnavirus* and *Labrynavirus* infecting algae and protists, respectively.

At present, the family *Secoviridae* contains eighty-six species, eighty-one of which are grouped into eight genera. The vast majority (sixty-two) of species fall into three related genera (*Comovirus*, *Fabavirus* and *Nepovirus*) that form the subfamily *Comovirinae*. Nineteen species fall within the remaining five genera; *Cheravirus*, *Sadwavirus*, *Sequivirus*, *Torradovirus* and *Waikavirus*. There are five unassigned species *Chocolate lily virus A* (CLVA), *Black raspberry necrosis virus* (BRNV), *Dioscorea mosaic associated virus* (DMAV), *Strawberry mottle virus* (SMoV), and *Strawberry latent ringspot virus* (SLRSV). Recent taxonomical changes, not yet ratified by the ICTV, will place four of these species within the *Sadwavirus* genus leaving only SLRSV as unassigned. Absorption of these four species to the *Sadwavirus* genus was achieved through the generation of three new subgenera, *Cholivirus*, *Satsumavirus* and *Stramovirus*. The *Nepovirus* genus is usually informally split into three subgroups A, B and C depending on phylogenetic relationships (Fig. 1).

These relationships for all members of the order *Picornavirales* have been traditionally defined by a contiguous region (Pro-Pol) spanning the functional domains of the protease (Pro) and polymerase (Pol), characterized at each extreme by the amino acids CG and GDD, respectively (Fig. 2). Trees produced from alignments of other functional domains on the whole are less structured but are not incongruent with the relationships defined by Pro-Pol trees. Notably, monophyly of the genera *Sequivirus*, *Torradovirus* and *Waikavirus* has been detected for the helicase and coat protein domains.

There is substantial evidence for the modular nature of virus genomes and despite the overall lack of phylogenetic discordance across their functional domains the secoviruses possess genetic characteristics that are modular in nature as is suggested by the detection of distant, yet significantly closer affinities with functional domains of invertebrate-infecting members of the *Picornavirales* than with fellow secoviruses. The experimentally demonstrated existence of a completely novel glutamic protease encoded by SMoV, is in line with this paradigm.

For most secoviruses, functional domains have been ascribed based on their similarity in sequence to proteins of known function. Because of the differences in their hosts and modes of infection, within the order *Picornavirales*, MPs are unique to the secoviruses. For plant viruses in general MPs are believed to be ultimately host in origin, for secoviruses these origins are possibly multiple as illustrated in tomato-infecting and non-tomato-infecting torradoviruses where putative MPs have their closest similarities to either umbraviruses and bromoviruses, respectively.

Current models using comparative genomics suggest secoviruses are likely invertebrate in origin. For this to be the case there must have been one or many host-switching events (invertebrate to plant) leading to a period of *in planta* adaptation followed by virus emergence. The intimate relationship between virus, host and vector provides ample opportunity for host-switching in either direction (invertebrate to plant or plant to invertebrate) to have occurred. The mechanisms necessary for such a significant (interkingdom) evolutionary step to occur are still to be discovered, but unlike members of other virus families (e.g., *Reoviridae*, *Bunyaviridae*) which replicate in both plant and vector, the secoviruses have apparently lost their capacity to infect the latter.

Attempts to establish when members of the *Secoviridae* might have evolved from ancestral viruses has been done using coalescent analyses on ninety-four coat protein sequences of the subfamily *Comovirinae*. An estimated time of emergence of less than a 1000 years



Fig. 1 Schematic of the taxonomic organization of the family *Secoviridae* showing typical features for each taxon. Horizontal numbers – present number of species per taxon. Vertical numbers – year taxon was recognized (irrespective of name given). Rounded rectangles (gray) – subfamily; (colored) – genera. Ellipses – subgroup groupings. Note – listed features may not apply to all members of group. * – presented subgenera organization of the *Sadwavirus* genus has yet to be ratified by the ICTV.

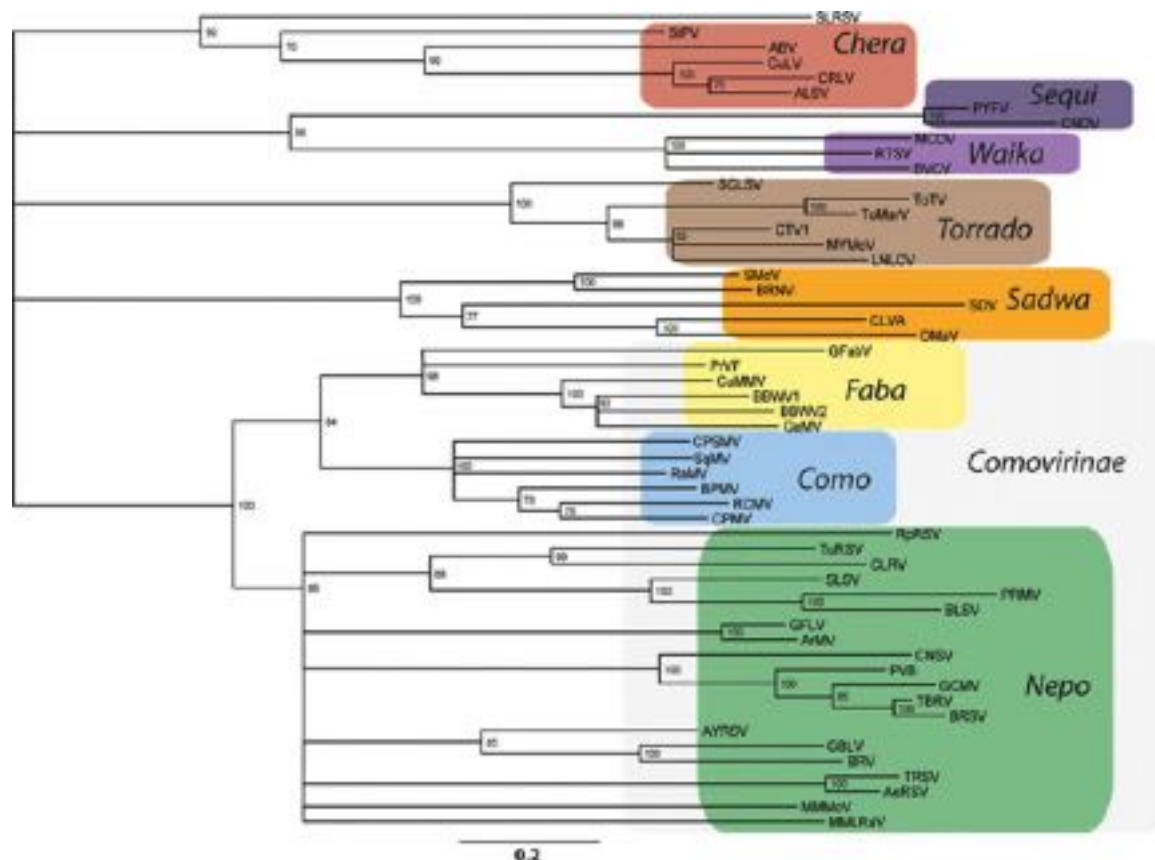


Fig. 2 Maximum likelihood inferred phylogenetic unrooted tree of members of the family *Secoviridae* based on an alignment of amino acid sequences of the conserved domains between the “CG” motif of the 3C-proteinase and the “GDD” motif of the polymerase (Pro-Pol region). The bar represents the genetic distance. Rounded rectangles (gray) – subfamily; (colored) – genera. For each species, the sequence of the exemplar isolate was used for the alignments (see species tables for the sequence accession numbers). Percent bootstrap values (1000 replicates) are shown at each branch node. Branches with bootstrap values below 70% were collapsed. Virus names and abbreviations are listed in [Table 1](#).

ago was calculated for the whole group with individual species emerging at various times between 50 and 250 years ago a period coinciding with an intensification in agricultural practices. Such estimates overlook possible variations in mutation rates and the birth and death of virus lineages over time. Nevertheless, this information provides a foundation to work from as computational phylogenetic methods become more sophisticated and the number of secovirus sequences deposited in the database grows.

The bipartite nature of most secoviruses is a likely adaptation from monopartite ancestors. In the virosphere those segmented viral genomes so far discovered are predominantly plant-infecting. Segmentation results in an uncoupling of the structural block of proteins from the replication-associated protein block allowing a more refined control of translation. It also facilitates genetic shuffling of each RNA to quickly produce novel combinations (reassortants) that may or may not have a selective advantage. Genome splitting might also allow for more efficient virion packaging. A bipartite version of foot and mouth disease picornavirus, produced spontaneously by multiple passages, was demonstrated to produce more stable virions than its monopartite counterpart, with presumably a longer viability outside of the cell. Nevertheless, a bipartite genome strategy means that both segments (and virions containing each segment) have to infect the same cell for successful infection, a theoretically more probable scenario for plants viruses, as their entry into the first cells of the host will be mechanically-mediated rather than receptor-mediated as for animal viruses.

Genetic mixing, whether reassortment or recombination, an important driver of evolution, requires a physical proximity of divergent viruses achieved by mixed infections which will be prolonged in perennial hosts. Evidence for the reassortment of secovirus RNA1 and 2 has been found for a raspberry isolate of Tomato ringspot virus (ToRSV) and a number of Korean isolates of Broad bean wilt virus 2 (BBWV2). Recombination is more frequently detected, at times leading to the putative emergence of new species as in the case of grapevine chrome mosaic and grapevine deformation nepoviruses.

Members of the Family

There are presently eighty-six recognized species of secoviruses distributed across eight genera, with one unassigned species SLRSV. For the majority (62) full-length genomic sequences are available including a NCBI RefSeq assembly number. For others (24) there is still no sequence information available ([Fig. 1](#); [Table 1](#)).

Table 1 List of all ICTV classified members of the *Secoviridae*, their taxonomic order, acronyms and GenBank sequence information. # Seq – number of sequences for each virus registered in NCBI Taxonomy Browser data from which a complete genome has been assembled to give a RefSeq assembly number (GCF_number). The last column to the right provides explanation if no RefSeq assembly number is available in the form of accession numbers for incomplete viral sequences or failing that, when no sequence information is available at all, the year, (in parentheses) of the species ratification by the International Committee for the Taxonomy of Viruses (ICTV)

Subfamily	Genus	Subgenus	Species	Abbreviation	# Seq	RefSeq assembly	GenBank accession (ICTV entry)
Comovirinae	Comovirus	–	Andean potato mottle virus	APMV	8	GCF_002833585.1	
			Bean pod mottle virus	BPMV	46	GCF_000862925.1	
			Bean rugose mosaic virus	BRMV	10	GCF_001430295.1	
			Broad bean stain virus	BBSV ^a	2	–	FJ028650.2, KJ746622.1
			Broad bean true mosaic virus	BBTMV ^b	7	GCF_000910755.1	
			Cowpea mosaic virus	CPMV	33	GCF_000860385.1	
			Cowpea severe mosaic virus	CPSMV	27	GCF_000861225.1	
			Glycine mosaic virus	GMV	0	–	(1982)
			Pea green mottle virus	PGMV	0	–	(1995)
			Pea mild mosaic virus	PMiMV	0	–	(1991)
			Quail pea mosaic virus	QPMV	0	–	(1976)
			Radish mosaic virus	RaMV	18	GCF_000879335.1	
			Red clover mottle virus	RCMV	13	GCF_000860425.1	
			Squash mosaic virus	SqMV	33	GCF_000861625.1	
			Ullucus virus C	UVC	0	–	(1995)
	Fabavirus	–	Broad bean wilt virus 1	BBWV1	123	GCF_000853465.1	
			Broad bean wilt virus 2	BBWV2	187	GCF_000861605.1	
			Cucurbit mild mosaic virus	CuMMV	5	GCF_002867105.1	
			Gentian mosaic virus	GeMV	12	GCF_000880435.1	
			Grapevine fabavirus	GFabV	18	GCF_003033815.1	
			Lamium mild mosaic virus	LMMV	4	GCF_000916715.1	
	Nepovirus	–	Prunus virus F	PrVF	18	GCF_003033845.1	
			Aeonium ringspot virus	AeRSV	4	GCF_002867145.1	
			Apricot latent ringspot virus	ALRSV	2	–	AJ278875
			Arabis mosaic virus	ArMV	150	GCF_000855205.1	
			Arracacha virus A	AVA	2	–	KY569301.1, KY569302.1
			Artichoke Aegean ringspot virus	AARSV	0	–	(1998)
			Artichoke Italian latent virus	AILV	9	GCF_005410585.1	
			Artichoke yellow ringspot virus	AYRSV	6	GCF_002986065.1	
			Beet ringspot virus	BRSV	12	GCF_000860405.1	
			Blackcurrant reversion virus	BRV	113	GCF_000849565.1	
			Blueberry latent spherical virus	BLSV	4	GCF_002867165.1	
			Blueberry leaf mottle virus	BLMV	5	–	U20622.1, U20621.1
			Cassava American latent virus	CsALV	0	–	(1991)
			Cassava green mottle virus	CGMV	2	–	MG581962.1, MG581963.1
			Cherry leaf roll virus	CLRV	395	GCF_000893515.1	
			Chicory yellow mottle virus	ChYMV	0	–	D00685.1, D00686.1 ^c
			Cocoa necrosis virus	CoNV	2	–	EU741694
			Crimson clover latent virus	CCLV	0	–	(1982)
			Cycas necrotic stunt virus	CNSV	13	GCF_000860925.1	
			Grapevine Anatolian ringspot virus	GARSV	5	GCF_000897495.1	
			Grapevine Bulgarian latent virus	GBLV	5	GCF_000892035.1	
			Grapevine chrome mosaic virus	GCMV	8	GCF_000862025.1	
			Grapevine deformation virus	GDeV	5	GCF_000898115.1	
			Grapevine fanleaf virus	GFLV	788	GCF_000860305.1	
			Grapevine Tunisian ringspot virus	GTuRSV	0	–	(1991)
			Hibiscus latent ringspot virus	HLRSV	0	–	(1982)
			Lucerne Australian latent virus	LALV	0	–	(1982)
			Melon mild mottle virus	MMMoV	4	GCF_002867185.1	
			Mulberry mosaic leaf roll associated virus	MMLRaV	4	GCF_002867205.1	
			Mulberry ringspot virus	MRSV	0	–	(1979)
			Myrobalan latent ringspot virus	MLRSV	0	–	(1979)
			Olive latent ringspot virus	OLRSV	2	GCF_002986085.1	
			Peach rosette mosaic virus	PRMV	7	GCF_002029615.1	
			Potato black ringspot virus	PBRSV	16	GCF_000913055.1	

(Continued)

Table 1 Continued

n/a	Cheravirus	–	Potato virus B	PVB	5	–	KX656670.1. KX656671.1
			Potato virus U	PVU	4	GCF_004117715.1	
			Raspberry ringspot virus	RpRSV	35	GCF_000854425.1	
			Soybean latent spherical virus	SLSV	4	GCF_001923735.1	
			Tobacco ringspot virus	TRSV	115	GCF_000856105.1	
			Tomato black ring virus	TBRV	86	GCF_000853325.1	
			Tomato ringspot virus	ToRSV	138	GCF_000860465.1	
			Apple latent spherical virus	ALSV	4	GCF_000861545.1	
			Arracacha virus B	AVB	14	GCF_000907115.1	
			Cherry rasp leaf virus	CRLV	10	GCF_000859565.1	
			Currant latent virus	CuLV	6	GCF_001503195.1	
			Stocky prune virus	StPV	4	–	DQ143874.1, DQ143875.1
	Sadwavirus ^d	Satsumavirus	Satsuma dwarf virus	SDV	33	GCF_000860985.1	
			Strawberry mottle virus	SMoV	83	GCF_000850445.1	
	Stramovirus	Black raspberry necrosis virus	BRNV	26	GCF_000867325.1		
			Choliavirus	Cholavirus	4	GCF_000895075.1	
	Sequivirus	–	Dioscorea mosaic associated virus	DMAV	4	GCF_001876835.1	
			Carrot necrotic dieback virus	CNDV	2	GCF_002817415.1	
	Dandelion yellow mosaic virus	–	DaYMV	2	–	DQ675189	
			Parsnip yellow fleck virus	PYFV	2	GCF_000862945.1	
	Torradovirus	–	Carrot torradovirus 1	CaTV1	10	GCF_000927615.2	
			Lettuce necrotic leaf curl virus	LNLCV	4	GCF_002219685.1	
	Motherwort yellow mottle virus	–	MYMoV	4	GCF_002220005.1		
			Squash chlorotic leaf spot virus	SCLSV	35	GCF_002219665.1	
	Tomato marchitez virus	–	ToMarV	25	GCF_000879575.1		
			Tomato torrado virus	ToTV	148	GCF_000872965.1	
	Waikavirus	–	Anthriscus yellows virus	AYV	0	–	(1995)
			Bellflower vein chlorosis virus	BVCV	2	GCF_001308735.1	
	Maize chlorotic dwarf virus	–	MCDV	4	GCF_000861445.1		
			Rice tungro spherical virus	RTSV	114	GCF_000860625.1	
	Unassigned	–	Strawberry latent ringspot virus	SLRSV	32	GCF_000857165.1	

^aBBSV acronym has been confused with BBTMV. BBSV has highest identity (65%) with RNA2 of RCMV.

^bOriginal acronym of BBTMV was TBBMV when ratified by the ICTV in 1972.

^cAccession numbers listed for chicory yellow mottle virus are satellite sequences.

^dSubgenera organization of the *Sadwavirus* genus has yet to be ratified by the ICTV.

Virion Structure

Virions are non-enveloped particles (ca. 30 nm diameter) with icosahedral symmetry ($T = 1$, pseudo $T = 3$, Fig. 3). Virions can either be RNA containing or empty. In the case of those viruses with segmented (bipartite) genomes virions can contain either RNA1 or RNA2, but not both. Thus, for successful completion of the virus infection cycle inoculum must consist of both RNA-containing particles. For bipartite viruses, virions can be grouped into three components based on their buoyant densities; the top (T) component which is empty, and the middle (M) and bottom (B) components which contain RNA2 and RNA1, respectively.

Irrespective of the number of CPs proteolytically cleaved the CP unit as a whole comprises three β -barrel domains each one containing two anti-parallel wedge-shaped β -sheets in a jelly roll fold conformation. Each sheet is made up of four β strands. The three β -barrel domains are either contained within one CP (genus *Nepovirus*) or cleaved to generate either three individual CPs each with one domain (genera *Waikavirus*, *Cheravirus*, *Sequivirus* and *Torradovirus*), or unevenly to produce a large (L) and small (S) CP with two and one domain, respectively (genera *Sadwavirus*, *Fabavirus* and *Comovirus*). The virion is formed by the assembly of 60 CP units (whether comprised of one, two or three cleaved products), and as such is said to have pseudo $T = 3$ symmetry as compared to viruses that use 180 identical CP copies to assemble virus particles with canonical $T = 3$ symmetry. Pseudo $T = 3$ symmetry is typical of all members of the *Picornavirales* and has been suggested in the case of secoviruses to be an evolutionary vestige that in animal-infecting counterparts permits improved evasion from the immune system when compared to true $T = 3$ capsids.

High resolution structures have been determined for a number of secoviruses, including Cowpea mosaic virus (CPMV), Bean pod mosaic virus (BPMV), Red clover mottle virus (RCMV) (*Comovirus*); Arabis mosaic virus (ArMV), Blackcurrant reversion virus (BRV), GFLV and Tobacco ringspot virus (TRSV) (*Nepovirus*). The existence of empty (T) particles demonstrates RNA is not required for the maintenance of virus particles, but does not preclude its involvement in assembly. Nevertheless, there appears to be some specificity to the interactions of viral RNA in the assembled particles, electron densities indicating a consensus base composition of the interacting RNA. At the three-fold axis of symmetry of the BPMV virion, viral RNA folds into a trefoil shape implying a potential function for this structure in packaging and particle assembly. More defined interacting amino acids have been identified for the CPs; in CPMV M-virions the amino acid N174 in the large CP subunit is intimately associated with and shown to be essential for genomic RNA encapsidation. Attempts to develop an *in vitro*

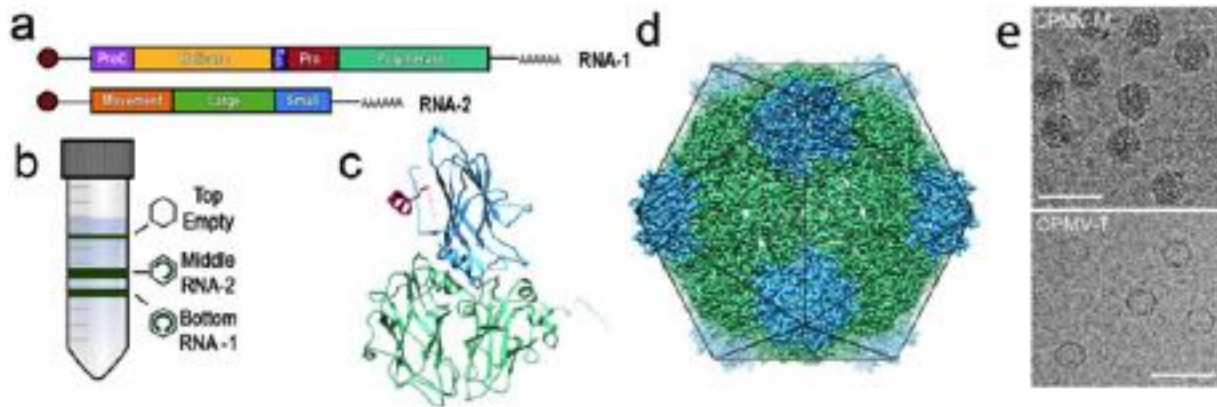


Fig. 3 Structure of Cowpea mosaic virus (CPMV) (a) Genome organization of CPMV. (b) Gradient centrifugation of wild-type CPMV permits separation into three components; empty CPMV particles sediment at the top (CPMV-T), CPMV containing RNA2 in the middle (CPMV-M) and RNA1 containing CPMV particles at the bottom (CPMV-B). (c) An asymmetric unit of CPMV empty virus-like particle (CPMV-T). The large CP subunit (L subunit, green) and the small coat protein subunit (S subunit, blue). The C terminal extension, only visualized in CPMV-T particles is colored pink. (d) Icosahedral organization of CPMV using an EM derived map of CPMV-T particles. Each of the 60 asymmetric units comprises one copy of the L subunit and the S subunit (colored as in 3c). A view down the two-fold axis is shown. (e) Electron micrographs of CPMV-M (top) and CPMV-T (bottom) particles. Courtesy of Hesketh, E.L., Meshcheriakova, Y., Thompson, R.F., Lomonosoff, G.P., Ranson, N.A., 2017. The structures of a naturally empty cowpea mosaic virus particle and its genome-containing counterpart by cryo-electron microscopy. *Scientific Reports* 7.

assembly assay, similar to other icosahedral viruses such as Cowpea chlorotic mottle virus, for CPMV which could help shed light on the mechanisms and requirements for assembly have so far proved unsuccessful due to the insolubility of the coat proteins.

Genome Organization

Whether monopartite or bipartite, the genome of secoviruses shows an overall similar organization, with two distinct blocks for replication-associated and structural functional domains. For bipartite viruses in which these blocks are segregated the adopted standard nomenclature is RNA1 and RNA2, respectively. Typically, viral proteins are expressed as large polyproteins, one per RNA, that are eventually cleaved into smaller functional mature proteins by viral encoded proteases (Fig. 4). However, comoviruses have been demonstrated to produce two different sized polyproteins from the same RNA2 ORF derived from two different initiation codons. And unlike any other genus in the family, torradoviruses are predicted to encode a protein 5' and overlapping to the longer more conserved long polyprotein containing the predicted CPs. The lengths of the 5' and 3' untranslated region vary considerably from species to species. For some viruses (nepoviruses and sadwaviruses), there is high sequence identity between RNA1 and/or RNA2 in the 5' and/or 3' untranslated regions (UTRs). For example, the 3'UTRs of Strawberry mottle virus are ca. 1150 nt long and 96% identical.

Protein domain organization, within the polyprotein(s) is similar to that of other members of the order *Picornavirales*. The structural block, which in monopartite genomes is located upstream of the replication block encodes the CP domain, upstream of which is the MP domain. The number of CP domain proteolytic derivatives that assemble to form the capsid varies from one to three depending on the genus (see Section "Virion Structure"). Upstream of the MP, nepoviruses encode a domain that is cleaved into one (2a protein in GFLV) or two (X3 and X4 proteins in ToRSV) proteins. The 2a protein fused to the green fluorescent protein of *Aequorea victoria* has been shown to associate with the replication complex assembled from RNA1-encoded proteins. The replication block contains domains with signature motifs typical of Helicases (NTP-binding domain proteins) (Hel), 3C-like protease (Pro) and RNA-dependent RNA polymerase (Pol). Between the Hel and Pro domains lies a small domain predicted to encode the VPg (viral protein genome-linked) (Fig. 4). Upstream of the Hel domain there exists for many secoviruses a divergent region encoding a predicted hydrophobic domain of unknown function. In CPMV the domain is encompassed by a single cleaved protein derivative that has been shown to affect protease activity, and thus was assigned the name of 'protease cofactor' (Co-Pro). In Tomato ringspot virus, there are two protein cleavage derivatives upstream of the Hel domain one (X2) sharing a conserved motif with the Co-Pro of CPMV.

Genome sizes of secoviruses is variable; the smallest are the como- and fabaviruses (RNA1 ~5.9 kb and RNA2 3.3–4 kb), while the related nepoviruses are some of the biggest (e.g., ToRSV, RNA1 ~8.2 kb and RNA2 7.3 kb). In the latter, this size includes a long almost identical 3'UTR in both RNAs. In monopartite secoviruses the genome sizes range from 9.9 kb (sequiviruses) to 12.2 kb (RTSV). All genome segments are assumed to be 3' polyadenylated although experiments carried out on the sequivirus Parsnip yellow fleck virus (PYFV) found no evidence of a poly(A) tail. The only other fully sequenced sequivirus, Carrot necrotic dieback virus, is polyadenylated. Based on the genome organization of secoviruses and comparisons with the related animal-infecting genera (*Picornaviridae*, *Discistroviridae* and *Iflaviridae*) translation is plausibly mediated through a 5' internal ribosomal entry site (IRES), a highly structured RNA element that facilitates cap-independent translation. Experimental evidence is limited to date but the presence

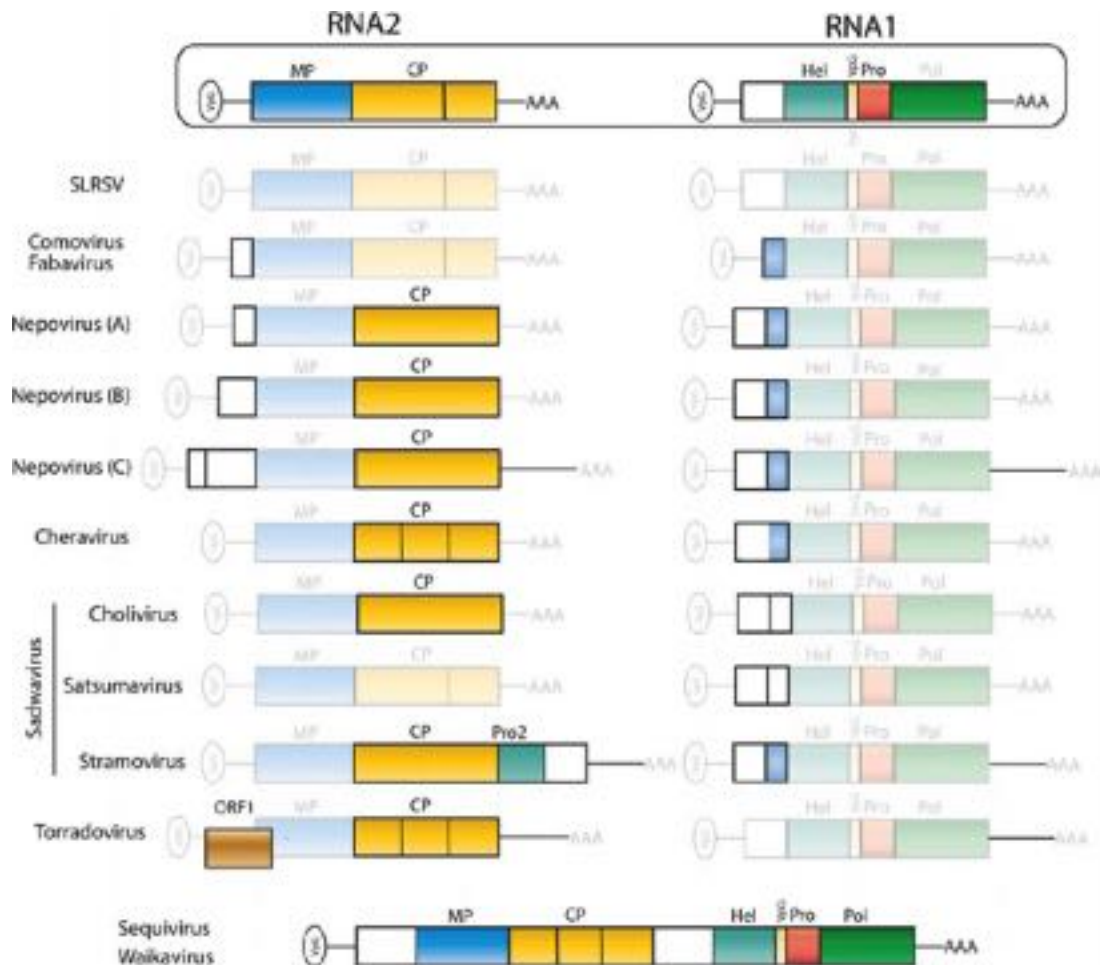


Fig. 4 Genome organization of representative members of the family *Secoviridae*. Boxed depiction at top is the generic secovirus organization. Only regions that notably deviate from this for each taxon are highlighted in the structures below. Each RNA is shown with the open reading frames (ORFs) represented by boxes (not to scale). Ovals depict VPg molecules covalently attached at the 5' end of the RNAs. Poly(A) tails are represented at the 3' end of the RNAs by AAA. Black vertical lines indicate evidence of cleavage sites. Movement protein – MP, coat protein(s) – CP, Helicase NTP-binding protein – Hel, protease – Pro, RNA-dependent RNA polymerase – Pol. Note – indicated subgenera organization of *Sadwavirus* genus not yet approved by the ICTV.

of a functional 5' IRES has been identified in the 5'UTR of Blackcurrant reversion virus, although sequences in the 3'UTR were also shown to be required for translation.

Properties and Functions of Gene Products

Once inside the cell the viral RNA can function as a messenger RNA resulting in the translation of the viral polypeptide(s). IRES have been identified for a number of members of the *Picornavirales*. The only experimental evidence for a putative IRES in secoviruses comes from Blackcurrant reversion virus for which both a candidate stem-loop (SL) in the 5' leader sequence, a cap-independent translation enhancer (CITE) at the 3'UTR and a potential kissing-loop interaction between them have been identified. Concordant with this translational model, mutational analyses of nucleotide stretches in the 5'UTR of BRV complementary to the 18S rRNA imply a potential recruitment of the 40S ribosomal subunit (Fig. 5) at the leader sequence. Similar conserved 5'UTR sequence stretches complementary to the 18S rRNA have been found for a number of other nepoviruses.

After translation, cleavage by the 3C protease (encoded on RNA1 in bipartite viruses) of the polyprotein into mature functional proteins occurs. Although cleavage site specificity of the protease varies across genera, the mechanism involved is assumed to be the same; an amino acid in the substrate-binding pocket of the protease interacts directly with the amino acid directly upstream and flanking the cleavage site. Although structurally similar to serine proteases the catalytic amino acid of secovirus proteases is a cysteine (C) – except in the Pro of Blueberry latent spherical virus for which it is a serine (S). Specificity of 3C proteases depends on the substrate binding pocket which for the majority of secoviruses is a histidine (H), but can be a leucine (nepovirus subgroups A

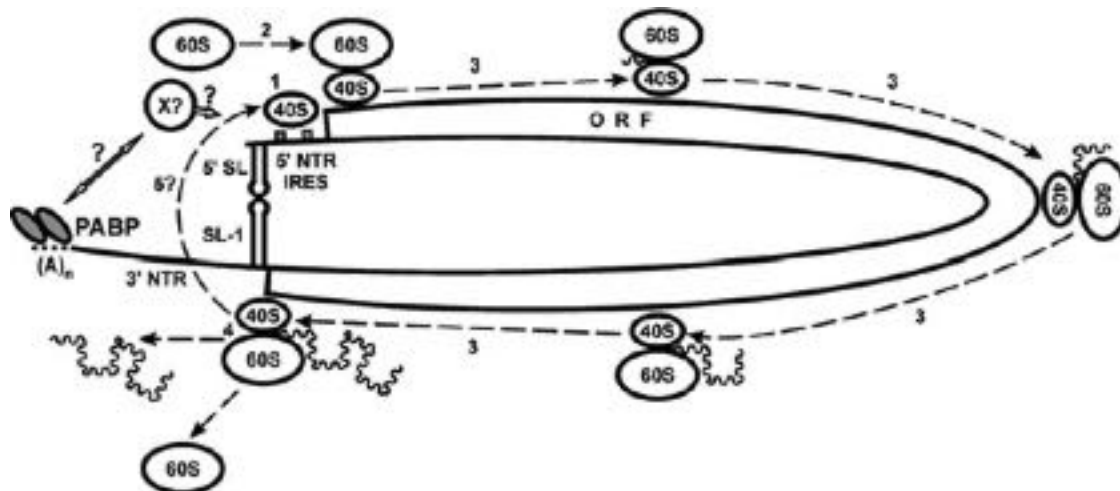


Fig. 5 Proposed model describing the possible mechanisms of translation for BRV RNAs. Step 1, the 40S ribosomal subunit (small oval) is recruited to the 5' untranslated region (UTR) internal ribosomal entry site (IRES), with base pairing between plant 18S rRNA and the corresponding complementary sequences of the BRV 5' leader (two small gray boxes). Step 2, after the start codon recognition, the 60S ribosomal subunit (large oval) joins to form the 80S initiation complex. Step 3, during translation elongation, the polypeptide (thin curly line) is synthesized. Step 4, after translation termination, the 80S complex is disassembled and the polypeptide is released. Step 5, long distance RNA-RNA base pairing between the 5' SL of the BRV 5' leader and the SL-1 of the 3'UTR holds two mRNA ends in close proximity, facilitating recycling of the ribosomal subunits again to the 5'UTR. An alternative mode of mRNA circularization could involve an interaction (double-headed arrow) of the poly(A)-binding protein (PABP; gray ovals) with a yet unidentified factor X, which in turn interacts (directly or through other factors) with the 5'UTR. Curved box – open reading frame (ORF); thick lines, UTRs; (A)_n – poly(A) tail. From Karetnikov, A., Lehto, K., 2008. Translation mechanisms involving long-distance base pairing interactions between the 5' and 3' non-translated regions and internal ribosomal entry are conserved for both genomic RNAs of Blackcurrant reversion nepovirus. *Virology* 371, 292–308.

and B and PYFV) or valine (Stocky prune virus, TRSV). When the binding pocket contains an H the cleavage site is typically, though not always, a glutamine (Q). The efficiency of cleavage depends on the cleavage site composition itself, the context of the amino acids flanking the cleavage site and the overall protein conformation, thus allowing for the existence of proteolytically intermediate products with functions different to that of their individual components. For example, in nepoviruses the structural block polypeptide derived from RNA2 is less efficiently cleaved by the VPg-Pro intermediate than by the fully released and mature Pro implying a possible mechanism for sequential processing and maturation of the individual functional domains in line with the infection cycle. In addition to viral protein maturation the secovirus Pro may also interact with and cleave host translation factors as has been demonstrated for the Pro of picornaviruses.

Once the main replication associated proteins (Pol, Hel and VPg) are functional, replication of the virus genome can occur within virus replication complexes (VRC) (see next section). The Hel domain, because of homology to known RNA helicases is assumed to be involved in binding and unwinding double stranded RNA structures that may form during replication. By analogy with picornaviruses the VPg is essential for viral replication acting as a primer for RNA synthesis but it may be removed from virion RNA by a host component (similar to the poliovirus unlinking enzyme). VPg is also absent on viral mRNA undergoing translation and has no effect on the infectivity of naked RNA. In nepoviruses full maturation of the Pro would shift proteolysis of the viral polypeptide and its intermediates to releasing the structural functional domains (CP and MP) of RNA2 both of which are required for virus movement. In the unique case of the newly proposed subgenus *Stramovirus* full release of the CP would also require proteolytic removal of the downstream glutamic protease.

There is plenty of evidence to suggest secoviruses move cell-to-cell in the form of virions (rather than ribonucleoproteins), como- and nepoviral MPs having been shown to be a component of tubular structures that act as a conduit for virions passing through the cell wall. Besides having an ultimate role in RNA encapsidation, some CP(s) also function as suppressors of gene silencing. Although not as potent as other classic plant virus silencing repressors (VSR) in that it does not prevent targeted mRNA degradation by the plant silencing machinery the CP of ToRSV has been shown to interact with AGO1 and relieve a translation repression mechanism. VSR activity has also been demonstrated for the small CP (CP-S) of CPMV and the smallest CP (Vp20) of Apple latent spherical virus (ALS).

Replication and Propagation

Purified total viral RNA is highly infectious but for viruses with a bipartite genome, neither RNA species by itself can infect plants systemically. RNA1 encodes for the replication-associated protein block and this is able to replicate in individual cells, but can neither move cell-to-cell or assemble into virus particles without the complementary functions of those proteins present on RNA2.

Replication occurs in VRCs, intracellular complexes formed by association with intracellular membranes. For como- and nepoviruses the Hel domain and upstream encoded proteins have been shown to interact with the endoplasmic reticulum (ER). Replication is reduced upon treatment with an inhibitor of type II fatty acid synthase, an observation that suggests VRC proliferation is not a simple reorganization of existing intracellular membranes. The Hel protein and the protein directly upstream of the Hel (Co-Pro in CPMV and X2 in ToRSV) are implicated in the accumulation of VRCs. In CPMV, VRCs in the form of irregularly shaped vesicles develop first in the cytoplasm and then coalesce around the nucleus. Electron tomography of picornavirus VRCs shows structures that during the exponential phase of infection are predominantly closed smooth single-membrane tubules that as infection progresses develop into double-membrane multilamellar structures. The ToRSV X2, Hel and Hel-VPg fusion protein and the CPMV Co-Pro and Hel-VPg fusion protein associate with the ER while other RNA1 derived mature proteins (VPg, Pro, and Pol) are soluble when expressed ectopically and are only peripherally associated with the VRCs in infected cells. This localization can be explained by the predicted hydrophobicity of the X2 and Co-Pro and equivalent domains in other secoviruses. Both molecules have predicted transmembrane helices and an amphipathic helix suggesting they may act as membrane anchors in the VRC. How viral RNA is recruited to the VRC in the first place is still unclear but replication and translation of RNA1 are tightly linked. Consistent with other virus replication-associated proteins, BPMV infection RNA1 derived proteins are exclusively *cis*-acting supporting the notion of a tight spatial molecular relationship. The RNA2 derived N-terminal proteins of BPMV (p58) and GFLV (2A) also function in *cis* and potentially serve to recruit RNA2 to the VRC.

Transmission, Host Range

In nature the majority of secoviruses are transmitted in a specific manner by insects or nematodes. Some (e.g., nepoviruses, cheraviruses, Satsuma dwarf virus (SDV), SLRSV) can be transmitted by seed and/or pollen. For members of the subgroup *Comovirinae* there is some predictability between the virus species within a genus and the vector (e.g., nepoviruses-nematodes, comoviruses-beetles, fabaviruses-aphids), while for other genera vector-virus relationships are more variable. For example, the torradoviruses; Tomato torrado virus (ToTV) and Tomato marchitez virus (ToMarV) are transmitted by whitefly species in a semi-persistent manner. However, detectable spread (horizontal transmission) of Carrot torradovirus 1 (CaTV1) and Lettuce necrotic leaf curl virus (LNLCV) in regions where whitefly numbers are limited (United Kingdom and Netherlands, respectively) has led researchers to identify aphid species as the most likely vectors. The type species of the *Cheravirus* genus Cherry rasp leaf virus (CRLV), is transmitted by the nematode *Xiphinema americanum* in the field and also through seeds, whereas ALSV has no known vector; transmission through seed and via pollen from infected apple trees occurring at a rate of 4.4% and 0.4%, respectively.

Such is the relationship sometimes between virus and their vector that their names have become intertwined, as in the case of nepoviruses and nematodes – nematode-transmitted viruses with polyhedral particles. This relationship can be very specific, exemplified by GFLV, for which a single amino acid change (Gly297Asp) on a surface loop of the CP is sufficient to significantly reduce efficiency of transmission by its only known vector *Xiphinema index*. This mutation does not seem to alter virion structure but has been proposed to inhibit the physical interaction between virus and the nematode vector receptor(s) necessary for successful transmission. However, broader nepovirus-vector relationships do exist: TRSV and ToRSV are known to be vectored by at least five nematode species; and BRV, consistent with its serological uniqueness, is transmitted by the eriophyid gall mite *Cecidophyes ribis*.

In some rare cases, secoviruses assist in the transmission of other viruses. Rice tungro, a disease that has periodically hit rice production in South East Asia is found associated with two distinct viruses; RTSV and Rice tungro bacilliform virus (RTBV) from the family *Caulimoviridae*. RTSV is transmitted by the leafhopper *Nephotettix virescens* but causes mild to no symptoms. RTBV is the main causal agent of the disease but depends on RTSV for transmission. Similarly, another waikavirus, Anthriscus yellows acts a helper virus in the aphid transmission of PYFV.

Secoviruses can either have broad or narrow host ranges, infecting predominantly dicotyledonous plants. Most nepoviruses and fabaviruses have wide host ranges, while comovirus infection is limited to a few species in the family *Leguminosae*. The whitefly transmissible ToTV, initially considered limited in host range to tomato, has been detected in over twenty weed species from eight families. SDV and related viruses have no known vector and natural infection appears restricted to species of citrus via grafting and mechanical transmission. Nevertheless, experimentally SDV has been shown to have a potentially broad host range infecting cowpea, cucumber, sesame and tobacco in a strain specific manner.

A small number of secoviruses (some nepoviruses and Strawberry latent ringspot virus) are found associated with a satellite RNA (satRNA); RNA that depends on the helper virus for replication and encapsidation, and can in some cases modulate symptoms. TRSV satellite RNA was the first satRNA ever described, and consists like other satRNAs in its class of a small (200–500 nt) RNA molecule lacking a functional open reading frame (ORF), an attached VPg and a poly(A)tail. In contrast, the other class of secoviral satRNAs are larger (1.1–1.5 kb), polyadenylated and associated with a 5' terminus VPg, and encode a protein essential for replication. Sequence similarities between different satRNAs is very limited although the predicted ORF of the large satRNAs show some structural similarities indicative of a shared function.

Epidemiology and Control

Dispersal of viruses occurs either through the infected host or the vector. In the past, plant virus spread would be local and limited to where virus and vector coincided geographically. In the wake of human globalization, virus movement across continents has

accelerated; for example, human beings are likely responsible for the present global distribution of GFLV and its nematode vector *X. Index*. And despite each country's legal measures to minimize the introduction of new plant materials the implementation of up-to-date certification schemes is essential in the monitoring and mitigation of disease outbreaks. Nematodes are slow moving, as will be the spread of any virus they vector in the field. Nevertheless, in the case of woody plants like grapevine that are cropped for multiple years, if phytosanitation practices are inadequate, virus-infected budwood can be inadvertently introduced at the very beginning of the vineyards lifetime.

Unlike nematodes; whiteflies and aphids are highly mobile and can move long distances rapidly extending the range and spread of the viruses they vector. Surveys made in the 2000s in Spain showed that ToTV was often found in mixed infections in tomato with another emerging non-whitefly-vectored potexvirus, Pepino mosaic virus (PepMV). The reason for the co-existence of both viruses has been hypothesized to be in part due to the conspicuous presence of whitefly populations for the former and a lack of genetic resistance against the latter. No increase in severity of torrado symptoms was observed in mixed infections.

In 1970s Japan a disease in rice termed *waiika* caused by infection of RTSV emerged to induce stunting and a reduction in grain quality. Generally RTSV alone causes mild symptoms but its ability to complement the transmission by leafhoppers of RTBV allowed for the emergence of tungro disease in the late 60s to mid-70s. Both viruses are transmitted in a semi-persistent manner by at least six leafhopper species that can remain viruliferous for up to a week. Inoculated plants can become sources of inoculum within one week. These conditions meant rapid spread of tungro in susceptible varieties. Attempts to control the disease has mainly alternated between the application of pesticides to reduce vector populations and the introduction of resistant varieties. Resistance principally took the form of resistance to the vector which over time broke down forcing growers to return to chemical control. In more recent years transgenic approaches that exploit RNA interference (RNAi) against both viruses have been explored along with RTSV resistance generated by CRISPR/Cas9 targeted mutagenesis of the eukaryotic translation initiation factor 4G (eIF4G).

RNAi technologies have also been applied to other nepoviruses (GFLV, ArMV, TRSV and ToRSV) mainly in herbaceous hosts, although a number of grapevine rootstocks transgenic for the CP have been developed. Constitutive expression of specific camelid-derived heavy-chain-only antibodies, or nanobodies, in *Nicotiana benthamiana* conferred strong resistance to GFLV, neutralizing the virus before cell-to-cell movement.

What control methods are employed and deployed against any plant pathogen will eventually depend on public perception of the associated risks and benefits.

Biotechnological Applications

CPMV has been at the vanguard nanobiotechnological developments and applications in part because of its high yield, simple purification and virion stability. Like other "nanotech" plant virus systems CPMV has the advantages of scalability, low-cost, low-toxicity and biodegradability. These factors, in combination with an early knowledge of its particle structure, resulted in it becoming the first plant virus to be successfully engineered to present foreign peptides on its surface with a 25 aa insertion of part of the VP1 domain of Foot and mouth disease virus (FMDV) into the β B- β C loop of the CP-S. Expression of other epitopes inserted in the β B- β C and other surface loops have also been effective, provided the insert is no greater than 40 aa with an isoelectric point (pI) of less than 9. Employing these virus-like particles (VLPs) to elicit an immunogenic response to the inserted epitope has been shown for Human rhinovirus 14, Mink enteritis virus, Human immunodeficiency virus, the bacteria *Pseudomonas aeruginosa* and *Staphylococcus aureus*, and for a number of human cancers (e.g., lung, melanoma, ovarian). The main motivation for these studies is the development of vaccines, and their stable expression in plants offers the appealing possibility of their delivery as oral vaccines provided conditions can be optimized so that CPMV VLPs tolerate the extreme physio-chemical conditions of the gastrointestinal tract long enough to be absorbed into the bloodstream. Uptake of CPMV particles into mammalian endothelial cells has been shown to be mediated by binding of the virus to the cell surface protein vimentin, a known receptor for some mammalian picornaviruses. This unexpected property could be put to use in the development of CPMV as a delivery system for medical therapeutics and imaging - RNA-containing virions of CPMV being readily fusible with a range of functional molecules.

Aside from nanotechnological applications CPMV containing vectors exploiting the hyper-translatability of a modified RNA2 5'UTR sequence have been engineered for the rapid expression of large amounts (1 g per kilogram of tissue) of pharmaceutical proteins in plants via agroinfiltration. Other secoviruses too have been cloned for biotechnological applications. In particular ALSV vectors have been used for functional genomic studies in a number of plants by means of virus-induced gene silencing (VIGS). One of the benefits of the ALSV system over other VIGS vectors is the virus' broad experimental host range which includes nightshades, cucurbits and legumes. Additionally its asymptomatic infection of most plants also means that the influence of virus infection on phenotype, as encountered with other VIGS vectors, is minimized. The ALSV system has been effectively used to "vaccinate" plants against infection from agronomically important viruses like Cucumber mosaic virus and potyviruses.

Diagnosis

One of the classical characteristics of nepovirus infection is symptom recovery, where the youngest emerging leaves of an infected plant are symptomless. Although not typical (or unique) for all nepoviruses, the cause for this recovery phenomenon has its origins in the gene silencing mechanism of the plant.

Minimizing the introduction of secoviruses into new planting material is best achieved by effective diagnostic screens and certification of seeds and propagative materials. Since the advent of molecular and serological methods diagnosis is commonly achieved by enzyme-linked immunological assays (ELISA) and/or the polymerase chain reaction. Despite its robustness, relative low cost and high throughput, ELISA or any serological method is sometimes impractical or intractable due to the technical challenges of virion purification or alternative methods of heterologous protein expression in the development of antibodies. These can be specific to each virus.

Commercial ELISA kits available for the following secovirus species – Andean potato mottle virus (APMV), ArMV, Arracacha virus B (ABV), Broad bean true mosaic virus (BBTMV), BBWV1, BBWV2, Carrot necrotic dieback virus (CNDV), CPMV, Cowpea severe mosaic virus (CPSMV), GFLV, Potato black ringspot virus (PBRV), Radish mosaic virus (RaMV), Raspberry ringspot virus (RpRSV), Squash mosaic virus (SqMV), SLRSV, Tomato black ring virus (TBRV), TRSV and ToRSV. A subset of these viruses can also be detected using commercial lateral flow devices. The utility of an interesting new technology that employs small proteins, called affimers, that mimic antibodies, has been shown to detect CPMV in crude extracts. Affimers are synthetically produced on a cystatin consensus scaffold and identified by phage display screening. Their alleged benefits over antibodies are their low cost and thermostability.

Multiplex detection methods where multiple viruses are detected in a single test are normally crop specific. Multiplex RT-PCR assays have been developed for citrus that include Satsuma dwarf viruses, and for grapevine that include ArMV and GFLV. Micro- and macro-arrays too have been employed for the multiplex detection of multiple grapevine infecting secoviruses.

With the advent of high throughput sequencing methods and their continuing drop in cost and turnover time the ability to detect multiple pathogens in multiple samples simultaneously is becoming routine and is rapidly superseding previous methodologies. Nevertheless there is a burgeoning need for “quick and dirty” tests that can be carried out by anyone anywhere. Lateral flow assays have addressed some of this need but more sensitive isothermal nucleic acid based methods like loop-mediated amplification (LAMP) and recombinase polymerase amplification (RPA) are rapidly becoming the methods of choice for point-of-use diagnostics.

Concluding Remarks

Research into secoviruses has so far provided a fascinating insight into a diverse group of viruses that share many features with their animal- and algal-infecting relatives. As human agricultural activities continue expanding and globalizing, the threat of novel disease outbreaks from existing and emergent secoviruses will remain. A clearer understanding of the biology and evolution of this group of viruses will undoubtedly better prepare us for combating their negative impact, and enlighten us as to how they and their picornavirus relatives adapt and exploit new environments and hosts. Which methods are used to control secovirus outbreaks will ultimately depend on the public's level of acceptance and perception of the associated risks and benefits. It is certainly plausible that the promise CPMV holds as a biotechnological and biomedical tool will help the secovirus name reach a broader more informed audience.

Further Reading

- Azzam, O., Chancellor, T.C.B., 2002. The biology, epidemiology, and management of rice tungro disease in Asia. *Plant Disease* 86, 88–100.
- Fuchs, M., Schmitt-Keichinger, C., Sanjaon, H., 2017. A Renaissance in nepovirus research provides new insights into their molecular interface with hosts and vectors. *Advances in Virus Research* 97, 61–105.
- Lin, T.W., Johnson, J.E., 2003. Structures of picorna-like plant viruses: Implications and applications. *Advances in Virus Research* 62, 167–239.
- Sainsbury, F., Canizares, M.C., Lomonosoff, G.P., 2010. Cowpea mosaic virus: The plant virus-based biotechnology workhorse. *Annual Review of Phytopathology* 48, 437–455.
- Sanjaon, H., 2015. *Secoviridae*: A family of plant picorna-like viruses with monopartite or bipartite genomes. In: eLS. Chichester: John Wiley & Sons, Ltd. doi:10.1002/9780470015902.a0000764.pub3.
- Sanjaon, H., Wellink, J., Le Gall, O., *et al.*, 2009. *Secoviridae*: A proposed family of plant viruses within the order *Picornavirales* that combines the families *Sequiviridae* and *Comoviridae*, the unassigned genera *Cheravirus* and *Sadwavirus*, and the proposed genus *Torradovirus*. *Archives of Virology* 154, 899–907.
- Thompson, J.R., Dasgupta, I., Fuchs, M., *et al.*, 2017. ICTV virus taxonomy profile: *Secoviridae*. *Journal of General Virology* 98, 529–531.
- Thompson, J.R., Kamath, N., Perry, K.L., 2014. An evolutionary analysis of the *Secoviridae* family of viruses. *PLoS One* 9 (9), e106305. doi:10.1371/journal.pone.0106305.
- van der Vlugt, R.A.A., Verbeek, M., Dulleman, A.M., *et al.*, 2015. *Torradoviruses*. *Annual Review of Phytopathology* 53, 485–512.

Relevant Websites

<https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses.

Sequiviruses and Waikaviruses (*Secoviridae*)

Lucy Rae Stewart, Agricultural Research Service, US Department of Agriculture, Wooster, OH, United States

Published by Elsevier Inc.

This is an update of I.-R. Choi, Sequiviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00462-3](https://doi.org/10.1016/B978-012374410-4.00462-3).

Nomenclature

aa Amino acid(s)
CP Coat protein or capsid protein
kb Kilobase
kDa Kilo dalton
nm Nanometer

nt Nucleotide(s)
ORF Open reading frame
RdRp RNA-dependent RNA polymerase
UTR Untranslated region
VPg Viral protein genome-linked

Glossary

3C-like protease Viral enzymes for polyprotein cleavage with similarity to the picornavirus 3C proteases, with similarity to chymotrypsin.

Helper component A viral-encoded protein not part of the capsid that mediates the transmission of viruses by vectors.

Monopartite Adjective describing a virus with a single particle containing a single complete genomic nucleic acid.

Polyprotein Large precursor polypeptide that is post- or co-translationally cleaved into mature protein components,

often with dynamics reflecting changing functions over time.

Semipersistent transmission A mode of vector-mediated transmission of plant viruses in which viruses are usually acquired within few minutes, and retained up to several days by vectors. Viruses do not multiply or circulate in vectors, and require no latent period between acquisition and transmission.

Synergy Co-infection of two or more viruses where disease is greater than additive and/or titer of one or more virus is enhanced.

Introduction

Viruses in the genera *Sequivirus* and *Waikavirus* are monopartite members of the family *Secoviridae*, comprising the former family *Sequiviridae*. Viruses belonging to these groups are often referred to as ‘picorna-like’ due to their similarities to picornaviruses in virion morphology and genome structure, and polyprotein expression strategy. They have monopartite single-stranded RNA (ssRNA) genomes encapsidated in isometric particles composed of three capsid proteins. They infect plants and are usually transmitted by insect vectors in nature. The genus *Waikavirus* includes pathogens of important cereal crops: Rice tungro spherical virus (RTSV), and Maize chlorotic dwarf virus (MCDV). Rice tungro disease is caused by a synergistic co-infection of RTSV with an unrelated caulimovirus, Rice tungro bacilliform virus (RTBV), and impacts rice production in South and South-East Asia. Maize chlorotic dwarf disease has been problematic in the Midwest and Southeastern United States. Members of the genus *Waikavirus* with described symptomology are phloem-limited and cause stunting, yellowing, and sometimes veinal chlorosis (**Fig. 1**). The genus *Sequivirus* shares similar genome organization but is sequence-divergent from the waikaviruses and fewer members are described, with the type species being *Parsnip yellow fleck virus* (PYFV). Members of the genus *Sequivirus* cause vein yellowing, yellow flecks, mosaic, and necrosis in infected hosts, can be mechanically transmitted experimentally, and are not phloem-limited. A unique biological feature of viruses in these genera is the ability of some members to act as (waikaviruses) or require (sequiviruses) a helper virus for transmission by insect vectors.

Viruses belonging to the genus *Sequivirus* infect mesophyll and epidermal cells and are able to be transmitted by mechanical inoculation and insect vectors, while those described in the genus *Waikavirus* are limited in phloem tissue and transmitted only by insect vectors or direct mechanical inoculation to phloem, as by vascular puncture inoculation in maize. Sequiviruses are dependent on a helper virus for their transmission by insect vectors. Waikaviruses are independently transmitted by insect vectors and presumably encode a helper component in their genomes. The viruses in each genus are primarily classified on the basis of their biological and physical characteristics, but conspicuous differences are also found in their genome features. The genomes of MCDV and RTSV are approximately 12 kb and polyadenylated at the 3′ end, whereas that of PYFV is about 10 kb and lacks 3′ polyadenylation.

Classification: Taxonomy, Phylogeny, and Evolution

The family *Secoviridae* is a combination of former families *Comoviridae* and *Sequiviridae*. The family is composed of eight genera: *Sequivirus*, *Waikavirus*, *Torradovirus*, *Sadwavirus*, *Cheravirus*, *Nepovirus*, *Fabavirus*, and *Comovirus*, as well as some unassigned members

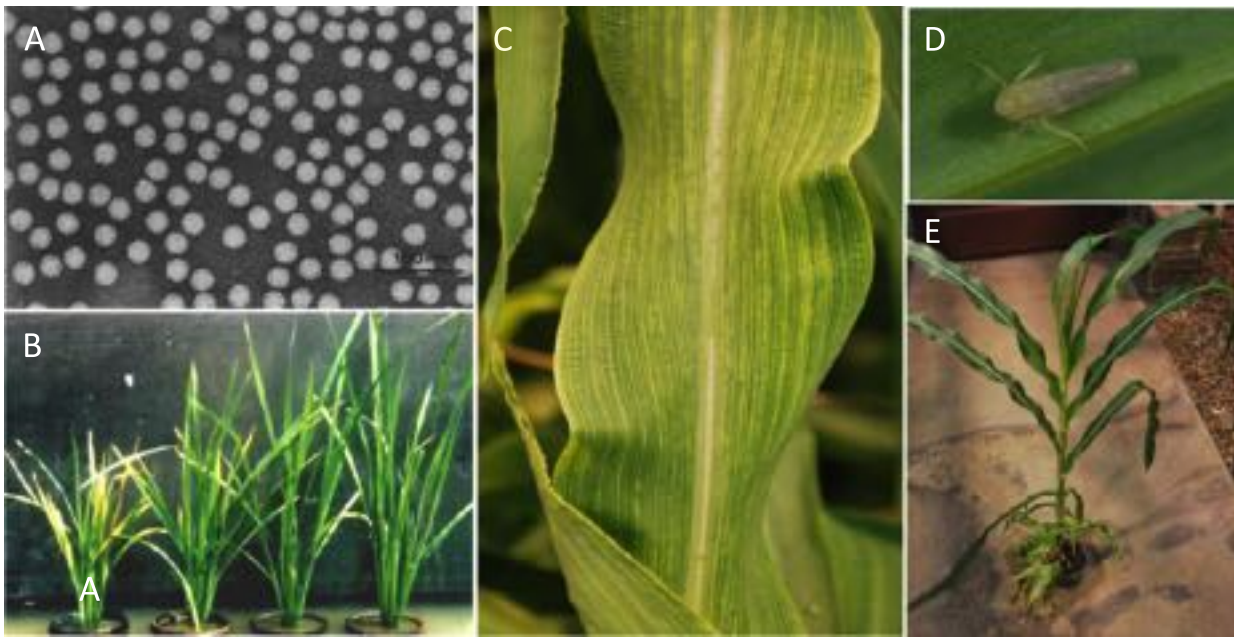


Fig. 1 Waikavirus symptoms and particles. (A) RTSV virions, ca. 30 nm diameter. Bar = 100 nm. (Photograph courtesy of Dr. Filomena Sta Cruz, University of the Philippines-Los Banos). (B) Rice plants with rice tungro disease synergistic disease from coinfection with RTSV and RTBV (left) compared to single infections with RTBV (center left), RTSV (center right), and uninfected plant (right). (Photograph courtesy of Dr. Filomena Sta Cruz, University of the Philippines-Los Banos). (C) MCDV-S vein chlorosis in corn. (D) The black-faced leafhopper, *Graminella nigrifrons*, vector of MCDV. (Photograph courtesy of Jane Todd, USDA-ARS). (E) MCDV-S-infected stunted plants in a pot with a single uninfected plant. Reproduced, by permission, from Stewart, L.R., 2011. Waikaviruses: Studied but not understood. APS Features. doi:10.1094/APSFeature-2011-11. <https://www.apsnet.org/edcenter/apsnetfeatures/Pages/waikavirus.aspx>.

(Fig. 2). Most of these are bipartite viruses, but the genera *Waikavirus* and *Sequivirus* are monopartite. The 5' end of the monopartite viruses corresponds to RNA2 and the 3' end to RNA1 of bipartite family members (Fig. 3). Currently there are three species described in the genus *Sequivirus*: *Parsnip yellow fleck virus* (the type species); *Dandelion yellow mosaic virus*; and *Carrot necrotic dieback virus*, formerly named *Parsnip yellow fleck virus*, Anthriscus strain (Table 1). In the genus *Waikavirus*, four species are described: *Rice tungro spherical virus* (type species); *Maize chlorotic dwarf virus* (MCDV); *Anthriscus yellows virus* (AYV); and *Bellflower vein chlorosis virus* (BVCV; Table 1). Other putative waikaviruses have been found in deep sequencing studies, including Blackcurrant waikavirus A, Red clover associated virus 1, and Brassica napus RNA virus 1.

Sequiviruses and waikaviruses, like other members of the family *Secoviridae*, share properties of plant picorna-like viruses, and are classified in the order *Picornavirales*. Comparison of amino acid sequences revealed that the sequence identity of the MCDV (isolate TN) polyprotein to those of RTSV (strain A) and PYFV (isolate P-121) were 51% and 35%, respectively. The NTP-binding domain in the polyprotein of MCDV (TN isolate) showed significant sequence identity to those of viruses such as RTSV (strain A, 79%), PYFV (isolate P-121, 55%), CPMV (46%), Hepatitis A virus (HAV, 47%), and Poliovirus (45%). The RdRp also shows significant similarities among members of the family *Secoviridae*. For instance, the sequence similarities in the RdRp region of MCDV (TN isolate) to those of related viruses are 75% for RTSV, 50% for PYFV, 48% for CPMV, 44% for TBRV, and 40% for HAV. Phylogenetic analysis based on the sequences of NTP-binding domains indicates that PYFV and RTSV are not noticeably similar to each other compared to their relatedness to picornaviruses and comoviruses. However, phylogenetic analysis based on the regions of RdRp suggests that PYFV and RTSV are more closely related to each other than to picornaviruses and comoviruses. The 26 kDa CP of PYFV and the 22 kDa CP of RTSV contain amino acid sequences resembling those in VP3 of Encephalomyocarditis virus and Human rhinovirus 14. Like other members of the *Secoviridae*, members of the genera *Waikavirus* and *Sequivirus* encode CPs with three jelly-roll domain in each of the 20 faces of the icosahedral virus particle. For the waika- and sequiviruses, each jellyroll domain is on a separate CP, in contrast to family members with one jellyroll domain on a small CP and two on a large CP.

Variation of Isolates and Strains

Complete sequences of sequiviruses PYFV and CNDV but not DaYMV are available. Although CNDV was formerly considered the Anthriscus strain of PYFV, strain variation in the currently classified species is not described for members of the genus *Sequivirus*. Complete sequences of the waikaviruses RTSV, MCDV, and BVCV, and putative waikaviruses Red clover associated virus 1 and Brassica napus RNA virus 1 are available, as are partial sequences of the putative member Blackcurrant virus A, but no sequences

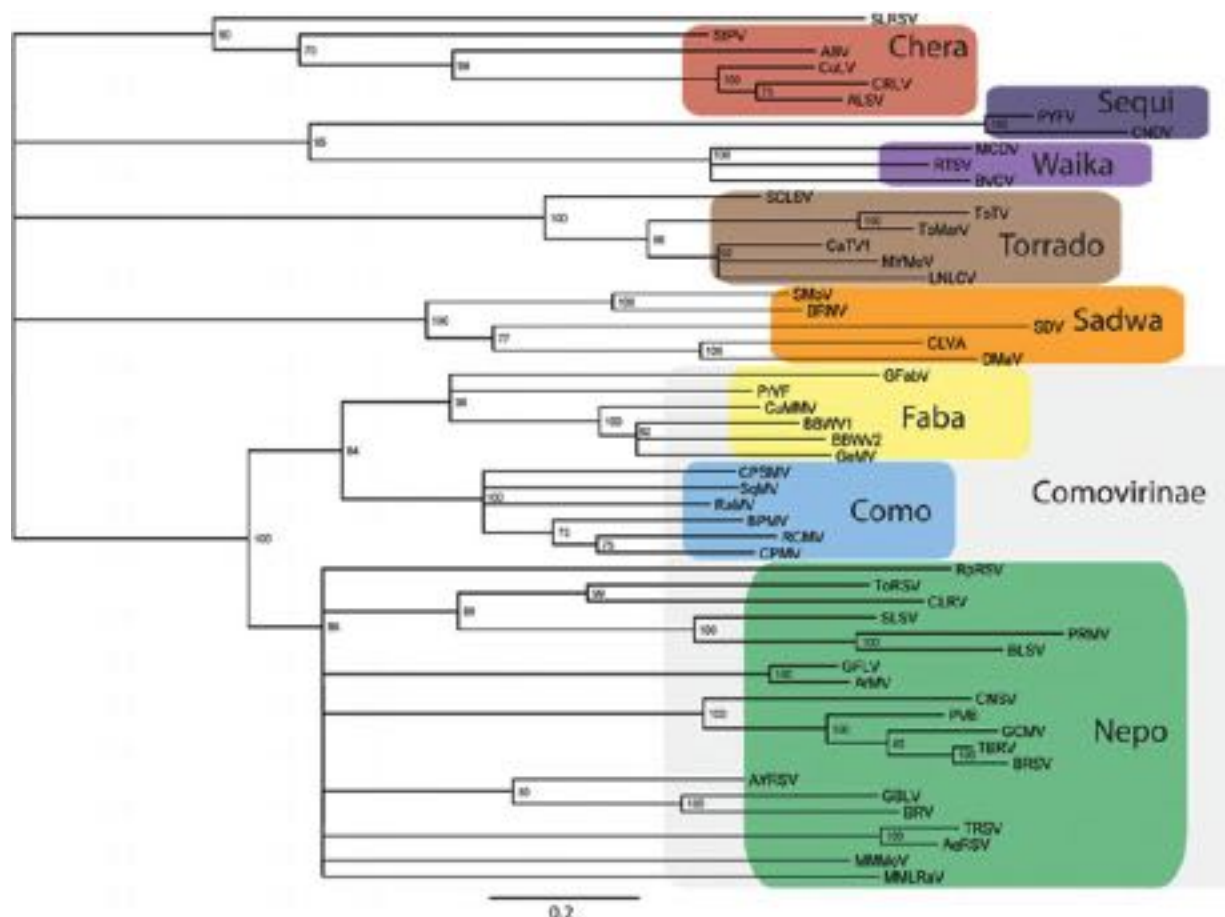


Fig. 2 Maximum likelihood inferred phylogenetic unrooted tree of members of the family *Secoviridae* based on an alignment of amino acid sequences of the conserved domains between the “CG” motif of the 3C-proteinase and the “GDD” motif of the polymerase (Pro-Pol region). The bar represents the genetic distance. Rounded rectangles (gray) – subfamily; (colored) – genera. For each species, the sequence of the exemplar isolate was used for the alignments (see species tables for the sequence accession numbers). Percent bootstrap values (1000 replicates) are shown at each branch node. Branches with bootstrap values below 70% were collapsed. Virus names, abbreviations, and GenBank accession numbers of sequences used to create the tree are: Strawberry latent ringspot virus (SLRSV), GCF_000857165; Stocky prune virus (StPV), DQ143874 and DQ143875; Arracacha virus B (AVB), GCF_000907115; Currant latent virus (CuLV), GCF_001503195; Cherry rasp leaf virus (CRLV), GCF_000859565; Apple latent spherical virus (ALSV), GCF_000861545; Parsnip yellow fleck virus (PYFV), GCF_000862945; Carrot necrotic dieback virus (CNDV), GCF_002817415; Maize chlorotic dwarf virus (MCDV), GCF_000861445; Rice tungro spherical virus (RTSV), GCF_000860625; Bellflower vein chlorosis virus (BVCV), GCF_001308735; Squash chlorotic leaf spot virus (SCLSV), GCF_002219665; Tomato torrado virus (ToTV), GCF_000872965; Tomato marchitez virus (ToMarV), GCF_000879575; Carrot torradovirus 1 (CaTV1), GCF_000927615; Motherwort yellow mottle virus (MYMoV), GCF_002220005; Lettuce necrotic leaf curl virus (LNLGV), GCF_002219685; Strawberry mottle virus (SMoV), GCF_000850445; Black raspberry necrosis virus (BRNV), GCF_000867325; Satsuma dwarf virus (SDV), GCF_000860985; Chocolate lily virus A (CLVA), GCF_000895075; Dioscorea mosaic associated virus (DMaV), GCF_001876835; Grapevine fabavirus (GFabV), GCF_003033815; Prunus virus F (PrVF), GCF_003033845; Cucurbit mild mosaic virus (CuMMV), GCF_002867105; Broad bean wilt virus 1 (BBWV1), GCF_000853465; Broad bean wilt virus 2 (BBWV2), GCF_000861605; Gientian mosaic virus (GeMV), GCF_000880435; Cowpea severe mosaic virus (CPSMV), GCF_000861225; Squash mosaic virus (SqMV), GCF_000861625; Radish mosaic virus (RaMV), GCF_000879335; Bean pod mottle virus (BPV), GCF_000862925; Red clover mottle virus (RCMV), GCF_000860425; Cowpea mosaic virus (CPMV), GCF_000860385; Raspberry ringspot virus (RrRSV), GCF_000854425; Tomato ringspot virus (ToRSV), GCF_000860465; Cherry leaf roll virus (CLRV), GCF_000893515; Soybean latent spherical virus (SLSV), GCF_001923735; Peach rosette mosaic virus (PRMV), GCF_002029615; Blueberry latent spherical virus (BLSV), GCF_002867165; Grapevine fanleaf virus (GFLV), GCF_000860305; Arabis mosaic virus (ArMV), GCF_000855205; Cycas necrotic stunt virus (CNSV), GCF_000860925; Potato virus B (PVB), KX656670.1 and KX656671; Grapevine chrome mosaic virus (GCMV), GCF_000862025; Tomato black ring virus (TBRV), GCF_000853325; Beet ringspot virus (BRSV), GCF_000849565; Artichoke yellow ringspot virus (AYRSV), GCF_002986065; Grapevine Bulgarian latent virus (GBLV), GCF_000892035; Blackcurrant reversion virus (BRV), GCF_000860405; Tobacco ringspot virus (TRSV), GCF_000856105; Aeonium ringspot virus (AeRSV), GCF_002867145; Melon mild mottle virus (MMMoV), GCF_002867185; Mulberry mosaic leaf roll associated virus (MMLRaV), GCF_002867205. Created by Thompson, J.R. for Encyclopedia of Virology and reproduced with permission.

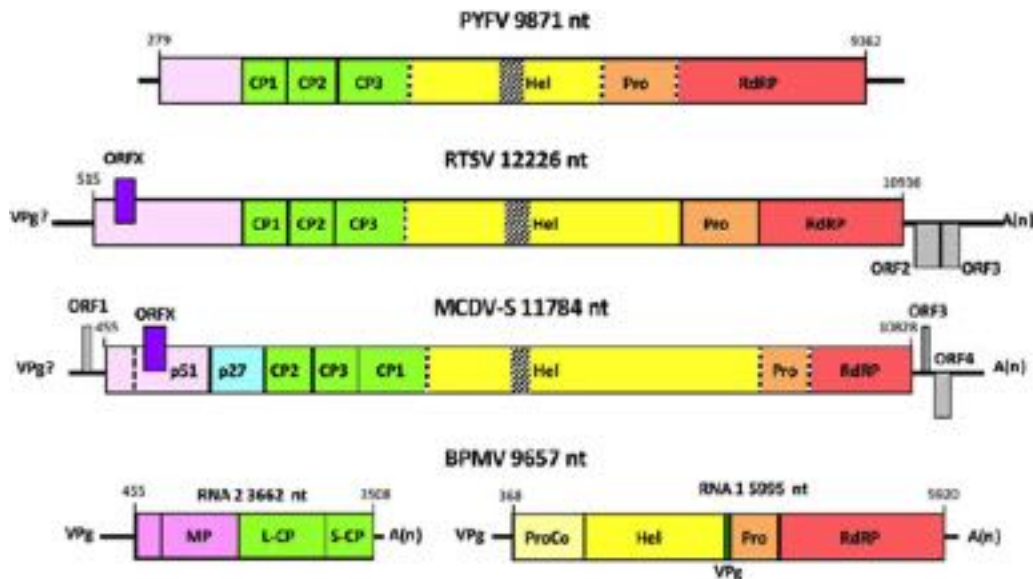


Fig. 3 Comparative genome organizations of PYFV, RTSV-A, MCDV-S, and BPMV. ORFs are indicated with rectangles. Horizontal lines at both ends of the ORFs represent the 5'- and 3'-UTRs of the genome. Solid vertical lines dividing the ORF correspond to the positions of verified proteolytic cleavage sites, while the positions corresponding to predicted cleavage sites are indicated with dotted lines. Nucleotide coordinates of translation start and stop codons for the polyproteins and the length of the entire genomes are indicated with numbers above the genomes. Numbers below the regions encoding CP1–CP3 are the predicted molecular weight (kDa) for the respective CP. Shaded areas represent the approximate region encoding NTP-binding domain. The regions in the ORF are shown with Hel for putative NTPase/helicase, Pro for 3C-like protease, and RdRp for RNA-dependent RNA polymerase. A(n) indicates polyadenylation.

Table 1 Virus members in the genera *Sequivirus* and *Waikavirus* in the family *Secoviridae*. Type species are written in bold fonts

Genus/virus	Acronym	Accession#	Genome size (kb)	Geographical distribution	Major natural host	Transmission vector
<i>Sequivirus</i>						
<i>Parsnip yellow fleck virus</i>	PYFV	GCF_000862945.19.9		Europe	Parsnip, hogweed, cow parsley	Aphids (<i>Cavariella aegopodii</i> , <i>C. pastinacae</i>)
<i>Carrot necrotic dieback virus</i>	CNDV	GCF_002817415.19.9		Europe	Carrot, cow parsley	Aphids (<i>Cavariella aegopodii</i>)
<i>Dandelion yellow mosaic virus</i>	DaYMV	GCF_002817435.110.0 ^a		Europe	Lettuce, dandelion	Aphids (<i>Acyrtosiphon solani</i> , <i>Myzus ornatus</i> , <i>M. ascalonicus</i> , <i>M. persicae</i>)
<i>Waikavirus</i>						
<i>Rice tungro spherical virus</i>	RTSV	GCF_000860625.112.2		Asia	Rice, <i>Oryza</i> spp.	Green leafhoppers (<i>Nephotettix virescens</i> and four other species)
<i>Anthriscus yellows virus</i>	AYV	—	10.6 ^a	Eurasia, UK	Cow parsley	Aphids (<i>Cavariella aegopodii</i>)
<i>Bellflower vein chlorosis virus</i>	BVCV	GCF_001308735.111.6		Korea	Bellflower	Unknown
<i>Maize chlorotic dwarf virus</i>	MCDV	GCF_000861445.111.8		USA	Maize, Johnsongrass	Deltoccephaline leafhopper (<i>Graminella nigrifrons</i>)
<i>Tentative members</i>						
Blackcurrant virus	ABcVA	—	11.0 ^a	USA	Blackcurrant	Unknown
Brassica napus RNA virus 1	BnRV1	GCF_00413465.1	12.3	Korea	Rapeseed	Unknown
Red clover associated virus 1	RCaV1	MH325329	11.9	Czech Republic	Red clover	Unknown

^aEstimated size (nucleotide sequence not reported).

are available for AYV. Multiple isolates of RTSV and MCDV have been completely sequenced and strains are described based on sequence and biological differences. Complete sequences are available for RTSV isolates RTSV-A (Philippines), Seberang Perai (Malaysia), and Vt6, a resistance-breaking Philippine isolate, and Andhra Pradesh, West Bengal, and Orissa collected in India. Nucleotide sequences of these isolates share 90%–97% identity, with Indian isolates clustering separately from isolates from the Philippines and Malaysia. Most diversity analyses for RTSV focus on the CP. Examination of nucleotide sequences of the RTSV CP revealed broad genotypic variation among and within geographic isolates, although the relationship of genotypic variation with the pathogenicity is not understood yet. CPs of isolates of RTSV from the Philippines and Malaysia show about 95% sequence similarity, while those from Bangladesh and India differ from the Philippine isolate by about 15%. The CP3 of the Indian isolate was also distinguishable from that of the Philippine isolate in electrophoretic mobility and the response to cellulolytic enzyme. Genotypic surveys on the CP sequences of RTSV field isolates collected from various sites in the Philippines and Indonesia indicated that a high degree of genetic diversity exists among the field isolates, and that infections with mixed genotypes in single sites are not uncommon. Phylogenetic analysis based on the CP sequences of the RTSV field isolate suggested that the clustering of genotypes found in Philippine sites was significantly different from that found in Indonesia sites, indicating some geographic isolation of RTSV populations. Strain Vt6 of RTSV was found to possess infectivity to some rice cultivars which the type strain A may not be able to infect. Amino acid sequence of strain Vt6 was approximately 96% identical to that of strain A, with greater dissimilarity in the leader protein region and the putative small ORF found near the 3' end.

Four isolates of MCDV have been reported. The S isolate of MCDV shares 99% nucleotide and polyprotein sequence identity to the type (T) isolate and both have bright veinal chlorosis and strong symptoms, particularly the S isolate. The sequences and origin of isolates described as S and T in earliest pre-sequencing literature are somewhat unclear, but S was probably a lab-selected subset of T collected in Ohio, and both are considered the same strain. The mild (M1) isolate exhibits very mild or no obvious symptoms by itself including no measurable stunting, but leaf texture changes and infected plant weight is reduced. M1 may develop transient severe symptoms in co-infection with S. M1 has only 61% polyprotein amino acid identity with isolate T. Antisera raised against isolate T react strongly with isolate S, but not with isolate M1. Isolate Tennessee (TN) of MCDV is also significantly divergent from isolate T, showing only 60% of amino acid sequence identity, but comparative TN symptom severity is unreported. In a survey of Ohio, where MCDV-M1 and MCDV-T isolates were first collected and identified, M1 and T sequences but no intermediate, recombinant, or divergent sequences were found. A more divergent partial MCDV sequence was identified in an Oklahoma tallgrass survey for viruses, but other MCDV isolates have yet to be identified. The low levels of amino acid sequence identity among the isolates of MCDV has raised the suggestion that they may represent distinct virus species. Within the family, species are demarcated by CP aa sequences <75% identical, Pro-Pol aa sequences <80% identical, antigenic differences, host range differences, vector specificity differences, and absence of cross-protection. MCDV-M1 CP polyprotein amino acid sequence is 85% identical to that of MCDV-TN, but only 65% identical to that of MCDV-S or MCDV-T. MCDV Pro-Pol amino acid sequences (based on RTSV-A protease start annotation) from MCDV-M1 are 80% identical to MCDV-TN, but each is only 67%–68% identical to MCDV-S or MCDV-T. Thus, MCDV-TN and MCDV-M1 meet sequence criteria for classification as a distinct species from MCDV-S or MCDV-T, but biological differences other than symptom severity have not been described.

Properties of Virions

Sequiviruses and waikaviruses have non-enveloped isometric particles approximately 30 nm in diameter (Fig. 1). The sedimentation coefficients of the virions are 153–159S for sequiviruses and 175–183S for waikaviruses. The buoyant density of PYFV virions in CsCl is 1.49 g ml⁻¹, and that of waikaviruses is 1.51–1.55 g ml⁻¹. Genome sequences and immunodetection of the viruses indicate that the virions consist of three capsid proteins (CPs). The sizes of CPs range from 22 to 35 kDa, depending on virus species and isolates. The virion of RTSV (Philippine-type strain A) consists of three proteins of 22.5 (CP1), 22 (CP2), and 33 (CP3) kDa. The predicted molecular masses for the CP of MCDV TN are 22, 23, and 31 kDa, and those of PYFV (P-121 isolate) are 22.5, 26, and 31 kDa.

Genome Organization

The genomes of sequiviruses and waikaviruses are positive-sense, monopartite ssRNA. The genome sizes range from approximately 10 kb for sequiviruses to 12 kb for waikaviruses. The genomes contain a large open reading frame (ORF) encoding a polyprotein presumably proteolytically processed by viral-encoded protease(s) during translation (Fig. 3). The large ORF in the genome of PYFV putatively encodes a polyprotein consisting of 3027 aa residues with a predicted molecular mass of about 336 kb. The large ORF in the genomes of waikaviruses encodes a polyprotein of approximately 400 kb, which has about 3440–3470 aa residues. In addition to the ORF for the polyprotein, short ORFs were also identified near the 3' and the 5' ends of MCDV and RTSV genomes. Two RNA species which seemingly correspond to sub-genomic transcripts from the short ORFs locating near the 3' end of the RTSV genome (strain A) were detected from infected plants. However, the lengths and locations of the short ORF in the genomes of waikaviruses vary considerably among viruses and isolates, and the presence of products translated from these ORFs has not been confirmed in plants. Another small ORF overlapping the N-terminal leader protein, ORFX, was identified by bioinformatic analysis. ORFX sequence is conserved across waikaviruses, but is not found in sequiviruses. ORFX expression is not experimentally

confirmed and mode of expression is not known. An AUG is present, but upstream sequence is also in frame. Waikaviruses and sequiviruses are likely to have a 5' virus-linked protein (VPg) and not a 5-methyl guanosine cap, but covalent linkage of a 5' VPg has not been demonstrated.

The 5'-untranslated region (UTR) in the genome of PYFV is about 0.28 kb in length, while those of waikaviruses are longer, about 0.43 kb in MCDV and 0.52 kb in RTSV. The 5'-UTR in the genomes of MCDV and RTSV has several AUG sequences upstream of the putative polyprotein start site, and appear to form extensive secondary structures. It is possible that such secondary structures may serve as the internal ribosomal entry site, as observed in the genomes of picornaviruses, to avoid the interference on translation by upstream AUG sequences, but no experimental evidence for this is yet available. No stable secondary structures are recognized in the 5'UTR of the PYFV genome; however, several short stretches of pyrimidines (UCUCUY) are present in the region. The 3'UTR in the genomes of waikaviruses are polyadenylated and unusually long. The length of 3'UTR in the genomes of MCDV and RTSV are approximately 1.0 and 1.2 kb, respectively, provided that the short ORFs near the 3' end of genomes are not translated. The 3'UTR in the genome of PYFV is approximately 0.5 kb in length. Unlike the genomes of waikaviruses, that of PYFV is not polyadenylated, but a part of 3'UTR is likely to form a stem-loop structure which is similar to that found in the genomes of flaviviruses.

Properties of Viral Proteins

Polyproteins of sequiviruses and waikaviruses are predicted to be cleaved into seven or more mature proteins through proteolytic maturation. Polyproteins of sequiviruses and waikaviruses contain a protease region similar to the 3C cysteine protease of picornaviruses and other viral proteases resembling the 3C protease such as the 24 kb protease of family member cowpea mosaic virus (CPMV) and the Nla protease of the potyvirus tobacco etch virus. For example, considerable similarity was found between the region delimited by the aa residues 2643 and 2853 in the polyprotein of RTSV (strain A), and a region in the 24 kb protease of CPMV. RTSV protease cleavage sites are not clearly conserved in MCDV, and the sequence context is not well-defined. 3C-like protease cleavage sites are not reliably predicted, but usually cleavage occurs between a glutamine (Q) residue and a small amino acid residue (e.g., M, S, V, A, L, I, G). Possible cleavage sites within polyproteins have only been validated for a subset of mature proteins. Whether other protease activity besides the 3C-like protease is encoded by sequiviruses or waikaviruses is also not experimentally determined.

The N-terminal part of the polyprotein is highly divergent between waikaviruses and sequiviruses, and among members of the family *Secoviridae*. Within the family, proteins encoded in this position function in virus movement (e.g., Bean pod mottle virus; [Fig. 3](#)) or have unknown function. Although the predicted sizes of the putative leader proteins are about 40 kDa in PYFV and 70–78 kDa in waikaviruses, the sizes of proteins detected with the antisera specific to the putative leader proteins are significantly smaller than predicted. An antiserum specific to the putative RTSV leader protein detected a protein of about 32 kDa in extracts from infected plants. An antiserum specific to the ca. 78 kDa MCDV leader protein detected proteins with of ca. 50, 35, and 25 kDa in extracts from plants infected with MCDV-S. The MCDV 78 kDa polyprotein was demonstrated to be cleaved by the virus-encoded 3C-like protease into p51 and p27 proteins. The precursor 78 kDa protein and p51 have silencing suppressor activity. Silencing suppressor function is conserved in BVCV and other waikaviruses but appears to have variable strength in the agroinfiltration assay among the orthologs. No evidence for silencing suppressor activity or *in vitro* cleavage by the cognate viral protease was observed for the corresponding N-terminal ca. 43 kDa PYFV protein, which is sequence divergent from similarly-positioned waikavirus proteins and may not be proteolytically processed further.

A leader protein found in the polyprotein of aphthoviruses in the family *Picornaviridae* has a protease activity which cleaves itself autocatalytically from the polyprotein. It was proposed that the ca. 35 kDa protein of MCDV is generated from the leader protein region through autoproteolysis of p51 at the putative cleavage site ALVRLFHGSAE (aa residues 150–160), while p27 results from the cleavage at Q445/S446 and Q686/S687 by the viral 3C-like protease. Autocatalytic cleavage has not been experimentally demonstrated, however. The p27 product of the MCDV 78 kDa protein has been implicated as a helper component since a protein of this size was detectable in leafhoppers fed on infected plant using antiserum against the MCDV leader polyprotein, but function has not been definitively demonstrated. This postulated role of p27 is consistent with the result from a serological blocking experiment showing that the helper component is not the virion or CPs.

Three consecutive CPs are encoded after the N-terminal protein region of the sequi- and waikaviruses. Protease cleavage sites of the N-termini of CPs are defined by amino acid sequencing of purified virion proteins for PYFV, MCDV-S and MCDV-TN, and RTSV. The proteolytic cleavage sites for the RTSV CPs determined by N-terminal amino acid sequencing were mapped to Q644/A645, Q852/S853, and Q1055/D1056 for the junctions between leader protein/CP1, CP1/CP2, and CP2/CP3, respectively. Similarly, MCDV-S CP junctions were mapped to Q686/S687, Q896/L897, and Q1098/V1099 from N to C-terminus. The context of these cleavage sites suggests that the CPs are processed *in trans* by the viral 3C-like protease. However, cleavage at these sites with the RTSV protease was not detected *in vitro*. Based on N-terminal sequencing of CPs, PYFV sequenced CP junctions were mapped to S394/P395, N588/A589, and Q810/A811.

The difference in the reactivity among the CPs of RTSV with the antibody raised against the virus particles indicated that the largest 33 kDa CP (CP3) is the major antigenic determinant on the surface of particles, although the structural details of virus particles have not been elucidated yet. The size of CP3 of RTSV (Philippine isolate) in the crude extract from infected plants detected by the antibody specific to CP3 appeared to be 40–42 kDa, markedly larger than that detected from purified virus

preparation or that expected from the genome sequence. It is likely that the size of CP3 in infected cells is larger due to post-translational modification, but the modified moiety appeared to be cleaved off probably by the treatment with cellulolytic enzymes during virion purification. For MCDV, the antisera against the purified virion detects the largest two coat proteins, CP1 and CP2, most strongly, and CP3 weakly or not at all, suggesting that this coat protein is internal or less antigenic. (For MCDV, CPs are numbered by size rather than ORF location as for RTSV, and the most antigenic CPs correspond to the largest ones in both.) A slower-migrating protein was also observed with MCDV virion antisera, which was also hypothesized to be a post-translationally modified largest CP.

The central region flanked by the CPs and the protease regions in the polyproteins of viruses belonging to the family *Secoviridae* contains sequences with similarities to helicase, protease, and RNA-dependent RNA polymerase (RdRp) of picornaviruses and comoviruses. The NTP-binding domain of the central polyprotein regions shows extensive similarity to the corresponding domains in the 58 kDa protein of CPMV and the 2C proteins of picornaviruses and contains motifs GX4GKS and DD, which are conserved among proteins with NTP-binding domains. The 58 kDa protein and the 2C protein presumably function as a nucleoside triphosphatase (NTPase)/helicase required for the initiation of negative-strand RNA synthesis. Antibodies specific to protein segments from the central region of the MCDV polyprotein reacted with three protein species in extracts from infected plants, indicating that the central region containing NTP-binding and helicase domains is likely also processed into smaller proteins.

The 3C-like protease of sequi- and waikaviruses follows the central region and precedes the polymerase. Immunodetection using an antiserum raised against a protein containing the putative protease region of RTSV indicated that the size of mature protease in infected plants is approximately 35 kDa. The sequence context of cleavage sites for the 3C-like cysteine protease and the results from *in vitro* translation and autocatalytic cleavage assays from partial genomic RNA templates defined the region of the protease in the RTSV polyprotein to be from aa 2527–2852. An internal cleavage site of the RTSV protease was also identified *in vitro* in a later study, but its biological significance, if any, is not known. The RTSV protease presumably acts in *cis* to cleave itself from the adjacent protein regions, but *trans*-cleavage by the RTSV protease was also observed to occur between the protease and the putative helicase regions *in vitro*. Based on the sequence alignment among the 3C-type proteases, aa Glu2717 and Cys2811 in the RTSV polyprotein were predicted to constitute part of the catalytic triad with His2680 and His2830 as contributing residues. Substitutions of these aa abolished or drastically reduced the proteolytic activity, substantiating their important roles in the protease. These aa are conserved in the protease regions of other waikaviruses.

The C-terminal region in the polyproteins of sequi- and waikaviruses is a predicted RNA-dependent RNA polymerase (RdRp). The RdRp region of RTSV (strain A) was defined in the region between aa 2853 and 3473 with the conserved YGDD motif at aa 3270–3273. The RdRp region of MCDV (isolate TN) has the conserved YGDD motif at aa 3238–3241. In addition, motifs such as DYSXFDG (aa 3129–3135 in the MCDV polyprotein) and PSGX3TX3NS (aa 3189–3200) were identified to be conserved among the RdRp regions of MCDV, CPMV, and Tomato black ring virus (TBRV). Multiple alignment of the RdRp region of PYFV (isolate P-121) with those of CPMV, TBRV, and Poliovirus showed that they share several conserved motifs including sequences YGDD (aa 2629–2632 in the PYFV polyprotein), PSGX3TX3NS (aa 2580–2591), and FLKR (aa 2684–2687).

Transmission

Several waikaviruses exhibit transmission ‘helper’ phenomena, as observed in cow parsley infected with PYFV and the waikavirus Anthriscus yellows virus (AYV), and rice plants infected with Rice tungro bacilliform tungrovirus (RTBV) and the waikavirus RTSV. In these cases mixed infections are the outcome of a waikavirus ‘helping’ insect transmission of another virus (AYV helping PYFV or RTSV helping RTBV). Due to transmission helping, both the waikavirus and the virus it helps are also transmitted by aphid or leafhopper vectors in a semipersistent manner, and in both cases the non-waikavirus cannot otherwise be transmitted by the insect vector.

PYFV is transmitted by aphids from plants infected with both viruses or by aphids previously fed on plants infected with AYV. Aphids seem to retain PYFV and AYV for up to 4 days. Even though AYV cannot infect parsnip, PYFV can be transmitted to parsnip by aphids fed on other plants infected with AYV and PYFV. Therefore, AYV apparently acts only as the helper virus for aphids to acquire PYFV, and is not necessary for the infection process of PYFV. The name of the genus *Sequivirus* comes from the Latin word ‘*sequi*’, which means to follow or accompany, in reference to the dependent insect transmission of PYFV. Incidences of mixed infection in lettuce with dandelion yellow mosaic virus and lettuce mosaic virus were also reported, but their relationships in aphid-mediated transmission are still unclear.

Green leafhoppers (*Nephotettix* spp.) transmit RTSV and RTBV simultaneously or singly from source plants infected with both viruses. RTSV is independently transmitted by leafhoppers, but RTBV can only be transmitted by leafhoppers that have fed on plants infected with RTSV. Although leafhoppers retain RTSV for only 3–4 days, their ability to acquire and transmit RTBV may persist for up to 7 days. Neutralization of RTSV-viruliferous green leafhoppers with anti-RTSV immunoglobulin markedly reduced the ability to transmit RTSV but still retained the ability to acquire and transmit RTBV. These observations indicate that while RTSV is the helper virus for the leafhopper-mediated transmission of RTBV, the helper function is associated with factors other than RTSV virions or CP.

Evidence for a helper component in the waikavirus MCDV is based on experiments showing requirements for its own transmission, which include a component not purified with virions in membrane feeding experiments. Purified MCDV-S virions fed to

blackfaced leafhopper vectors by membrane feeding were not transmitted unless leafhoppers were previously fed on plants infected with the MCDV mild strain MCDV-M1. Leafhoppers fed on MCDV show an electron-dense “matrix” in the foreguts observed by transmission electron microscopy, which has been proposed to be composed of helper component protein. The ca. 25 kDa proteolysis product of MCDV R78 (likely p27) was postulated to be the helper component protein based on detected retention in virus-exposed leafhoppers. However, the helper component for MCDV or other waikaviruses has not yet been definitively demonstrated. For most more recently discovered members of the genera *Sequivirus* and *Waikavirus* discovered by deep sequencing, the vectors and transmission modes are not yet described.

Diseases and Management

Important diseases are described for the most-studied waika- and sequiviruses, but not known for some of the viruses identified by sequences. Major diseases caused by waikaviruses include rice tungro disease and maize chlorotic dwarf disease. Rice tungro disease is caused by the synergistic coinfection of RTSV and RTBV. Both viruses are transmitted by green leafhoppers in a semipersistent manner. RTSV alone does not cause conspicuous symptoms except occasional slight stunted growth, but severe disease results when plants are co-infected with RTSV and RTBV (Fig. 1). Such synergistic effects of RTSV on symptom development are also observed when RTSV co-infects plants with viruses such as Rice grassy stunt virus and Rice ragged stunt virus. Rice plants infected with RTBV alone exhibit symptoms such as stunted growth, yellow to yellow-orange discoloration of leaves, and reduced tillering. Severity of rice tungro disease outbreaks vary from year to year with vector populations. It is managed with vector control using insecticides or crop rotation as well as genetic resistance in rice to leafhoppers and RTSV. MCDV is present in the U.S. and is managed by removal of its overwintering host, Johnsongrass (*Sorghum halepense*) in and around fields. Other unidentified changes in cultivation practices may have contributed to its incidence reduction since the 1970s. Serological and amplification-based assays are effective for diagnostics of both RTSV and MCDV. PYFV and CNDV are pathogens of vegetable crops, with PYFV causing yellowing and mosaic of parsnip and celery, and CNDV causing mosaic, necrosis and dieback in carrot, chervil, coriander, and dill. These are managed by planting seed in areas with no virus reservoir and limited vector aphid populations.

Concluding Remarks

Waikaviruses and Sequiviruses are unique monopartite viruses that include important pathogens and have unusual molecular biology and transmission features. Only a few viruses are well-described from each genus, although additional viruses with sequences similarity continue to be identified by deep sequencing, and tools continue to add information about the biology and molecular biology of these viruses in the family *Secoviridae*.

Further Reading

- Ammar, E.D., Nault, L.R., 1991. Maize chlorotic dwarf viruslike particles associated with the foregut in vector and nonvector leafhopper species. *Phytopathology* 81, 444–448.
- Azzam, O., Chancellor, T.C.B., 2002. The biology, epidemiology, and management of rice tungro disease in Asia. *Plant Disease* 86, 88–100.
- Azzam, O., Yambao, M.L.M., Muhsin, M., McNally, K.L., Umadhay, K.M.L., 2000. Genetic diversity of Rice tungro spherical virus in tungro-endemic provinces of the Philippines and Indonesia. *Archives of Virology* 145, 1183–1197.
- Chaouch-Hamada, R., Redinbaugh, M.G., Gingery, R.E., Willie, K., Hogenhout, S.A., 2004. Accumulation of maize chlorotic dwarf virus proteins in its plant host and leafhopper vector. *Virology* 325, 379–388.
- Elnagar, S., Murrant, A.F., 1976. Relations of the semi-persistent viruses, parsnip yellow fleck and anthriscus yellows, with their vector, *Cavariella aegopodii*. *Annals of Applied Biology* 84, 153–167.
- Firth, A., Atkins, J., 2008. Bioinformatic analysis suggests that a conserved ORF in the waikaviruses encodes an overlapping genes. *Archives of Virology* 153, 1379–1383.
- Galvez, G.E., 1971. Rice tungro virus. *Description of Plant Viruses* 67. Available at: <http://www.dpvweb.net/dpv/showadpv.php?dpvno=67>.
- Gingery, R.E., Bradfute, O.E., Gordon, D.T., Nault, L.R., 1978. Maize chlorotic dwarf virus. *Description of Plant Viruses* 67. Available at: <http://www.dpvweb.net/dpv/showdpv.php?dpvno=194>.
- Hibino, H., 1983. Transmission of two Rice tungro-associated viruses and Rice waika virus from doubly or singly infected source plants by leafhopper vectors. *Plant Disease* 67, 774–777.
- Hull, R., 1996. Molecular biology of rice tungro viruses. *Annual Review of Phytopathology* 34, 275–297.
- Hunt, R.E., Nault, L.R., Gingery, R.E., 1988. Evidence for infectivity of Maize chlorotic dwarf virus and for a helper component in its leafhopper transmission. *Phytopathology* 78, 499–504.
- Isogai, M., Cabauatan, P.Q., Masuta, C., Uyeda, I., Azzam, O., 2000. Complete nucleotide sequence of the Rice tungro spherical virus genome of the highly virulent strain V16. *Virus Genes* 20, 79–85.
- Murrant, A.F., 2003. Parsnip yellow fleck virus. *Description of Plant Viruses* 394. Available at: <http://www.dpvweb.net/dpv/showadpv.php?dpvno=394>.
- Reddick, B.B., Habera, L.F., Law, M.D., 1997. Nucleotide sequence and taxonomy of Maize chlorotic dwarf virus within the family *Sequiviridae*. *Journal of General Virology* 78, 1165–1174.
- Shen, P., Kaniewska, M., Smith, C., Beachy, R.N., 1993. Nucleotide sequence and genomic organization of Rice tungro spherical virus. *Virology* 193, 621–630.
- Stewart, L.R., 2011. Waikaviruses: Studied but not understood. *APS Features*. doi:10.1094/APSFeature-2011-11.
- Stewart, L.R., Jarugula, S., Zhao, Y., Qu, F., Marty, D., 2016. Identification of a Maize chlorotic dwarf virus silencing suppressor protein. *Virology* 504, 88–95.
- Thole, V., Hull, R., 1998. Rice tungro spherical virus polyprotein processing: Identification of a virus-encoded protease and mutational analysis of putative cleavage sites. *Virology* 247, 106–114.
- Turnbull-Ross, A.D., Mayo, M.A., Reavy, B., Murrant, A.F., 1993. Sequence analysis of the Parsnip yellow fleck virus polyprotein: Evidence of affinities with picornaviruses. *Journal of General Virology* 74, 555–561.

Relevant Websites

https://viralzone.expasy.org/660?outline=all_by_species

Sequivirus ~ ViralZone page.

https://viralzone.expasy.org/661?outline=all_by_species

Waikavirus ~ ViralZone page.

Solemoviruses (*Solemoviridae*)

Cecilia Sarmiento, Merike Sõmera, and Erkki Truve[†], Tallinn University of Technology, Tallinn, Estonia

© 2021 Published by Elsevier Ltd.

This is an update of M. Meier, A. Olspert, C. Sarmiento, E. Truve, Sobemovirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00507-0](https://doi.org/10.1016/B978-012374410-4.00507-0).

Nomenclature

aa Amino acid(s)

ARM Arginine-rich motif

CP Coat protein or capsid protein

diRNA Defective interfering RNA

kb Kilobase

kDa Kilo dalton

MP Movement protein

NLS Nuclear localization signal

nt Nucleotide(s)

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

satRNA Satellite RNA

sgRNA Sub-genomic RNA

TM Trans-membrane anchor

UTR Untranslated region

VPg Viral protein genome-linked

Glossary

Icosahedral particle Spherical viral particle that is a polyhedron having 20 faces.

Leaky scanning of ribosomes Mechanism during translation initiation for escaping the first start codon, which occurs when the first AUG resides in a very poor context and therefore only some ribosomes initiate translation at that point.

Ribosomal frameshifting – 1 Event occurring during translation elongation when the ribosome shifts its frame for reading the mRNA exactly one position in the upstream direction.

Satellite RNA Subviral agent consisting of RNA that becomes packaged in protein shells made from coat protein of the helper virus and whose replication is dependent on that virus.

Introduction

Solemoviruses are plant viruses with single-stranded RNA (ssRNA) genome of positive polarity with small icosahedral virion. Some solemoviruses are distributed throughout the world; others are limited to one continent or even to one endemic region. A number of solemoviruses are the causative agents of viral diseases of important crops. Rice yellow mottle virus (RYMV) is economically the most important virus causing 20%–100% yield loss of rice production in Africa. Subterranean clover mottle virus (SCMoV) damages seriously Australian grasslands, and Papaya lethal yellowing virus (PLYV) is responsible for a severe disease in Brazilian papaya plantations.

The first isolated solemovirus was Southern bean mosaic virus (SBMV), described in 1943. It was proposed in 1969 that single-component-RNA beetle-transmitted viruses be placed into a Southern bean mosaic virus group. Since 1995, this group was recognized by the International Committee on Taxonomy of Viruses (ICTV) as the genus *Sobemovirus*, unassigned to any family. The family *Solemoviridae* was created in 2017 to group together the genera *Sobemovirus* (including 19 species) and *Polemovirus* (with only one member; [Table 1](#)). During the last years, the list of members has increased and is expected to keep growing, especially due to new available metagenomic data.

From the name Southern bean mosaic virus derives the genus name *Sobemovirus*. The name of the genus *Polemovirus* comes from polerovirus and sobemovirus to indicate the relationship of polemoviruses to those genera, as explained in the genome section.

Virion Structure

Solemovirus particles are very stable icosahedral virions of 25–34 nm in diameter displaying T=3 symmetry. The particle is built of 180 coat protein (CP) molecules (~30 kDa) that are translated from sub-genomic RNA (sgRNA). The single-stranded genomic RNA and sgRNA together with a viral genome-linked protein (VPg) are packaged inside the virion.

The three-dimensional structures of *Cocksfoot mottle virus* (CfMV), *Sesbania mosaic virus* (SeMV), *Southern cowpea mosaic virus* (SCPMV), RYMV, and *Ryegrass mottle virus* (RGMoV) virions have been determined utilizing X-ray crystallography. CP monomers are chemically identical and can adopt three different conformations (A, B, and C). The A subunits group at 5-fold axes to form 12 pentamers, while pairs of B and C subunits meet at 3-fold axes to form 20 hexamers. The combination of hexamers and pentamers gives the particle its characteristic shape. The atomic structure of SBMV particle drawn with computer graphics is shown in [Fig. 1](#) (A subunits in red, B and C in blue and green).

[†]Deceased.

Table 1 Members of the family *Solemoviridae*

Genus/Species	Acronym	Accession #	Vector
<i>Sobemovirus</i>			
<i>Artemisia virus A</i>	ArtVA	NC_017914.1	
<i>Blueberry shoestring virus</i>	BSSV	NC_029578.1	aphid
<i>Cocksfoot mottle virus</i>	CfMV	NC_002618.2	beetle
<i>Cymbidium chlorotic mosaic virus</i>	CyCMV	NC_027123.1	
<i>Imperata yellow mottle virus</i>	IYMV	NC_011536.1	
<i>Lucerne transient streak virus</i>	LTSV	NC_001696.2	
<i>Papaya lethal yellowing virus</i>	PLYV	NC_018449.1	
<i>Rice yellow mottle virus</i>	RYMV	NC_001575.2	beetle
<i>Rottboellia yellow mottle virus</i>	RoMoV	NC_027198.1	
<i>Ryegrass mottle virus</i>	RGMoV	NC_003747.2	
<i>Sesbania mosaic virus</i>	SeMV	NC_002568.2	
<i>Solanum nodiflorum mottle virus</i>	SNMoV	NC_033706.1	beetle, mirid
<i>Southern bean mosaic virus</i>	SBMV	NC_004060.2	beetle
<i>Southern cowpea mosaic virus</i>	SCPMV	NC_001625.2	beetle
<i>Sowbane mosaic virus</i>	SoMV	NC_011187.1	aphid, leafminer, leafhopper, mirid
<i>Soybean yellow common mosaic virus</i>	SYCMV	NC_016033.1	
<i>Subterranean clover mottle virus</i>	SCMoV	NC_004346.1	beetle
<i>Turnip rosette virus</i>	TRoV	NC_004553.3	beetle, mirid
<i>Velvet tobacco mottle virus</i>	VTMoV	NC_014509.2	beetle, mirid
<i>Polemovirus</i>			
<i>Poinsettia latent virus</i>	PnLV	NC_011543.1	

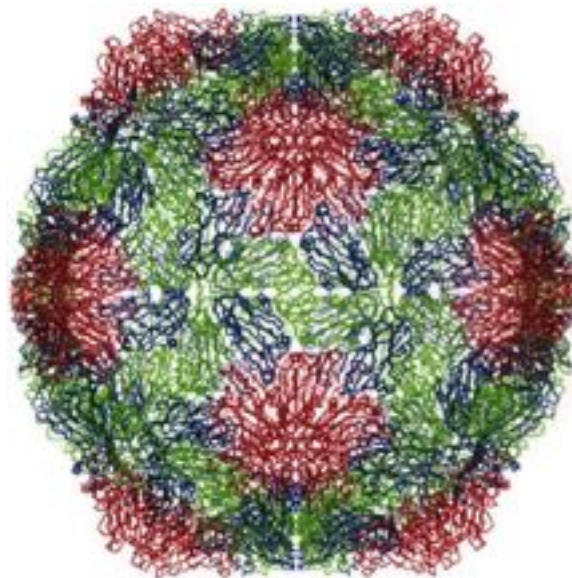


Fig. 1 Atomic structure of the entire virus particle of SBMV shown with computer graphics. Each subunit is shown as the C α tracing in different colors to depict the different geometric environments within the icosahedral shell. Red: Icosahedral 5-fold; Blue-Green: icosahedral 3-fold or quasi-6-fold clusters; Green at the center of the viral particle: icosahedral 2-fold. From Abad-Zapatero., G., 2002. *Crystals and Life: A Personal Journey*. International University Line, ISBN: 0-9720774-0-5. plate XVIII, with permission.

The root mean square deviations between superimposed coordinates of C α atoms of solemoviral CP residues are in general 1.4–1.5 Å. Interestingly, the virion of RGMoV is smaller and slightly more similar to the virion of Tobacco necrosis virus (family *Tombusviridae*) than to other solemovirus virions. Homology modeling of the 3D structures of the Rottboellia yellow mottle virus (RoMoV) and Artemisia virus A (ArtVA) CPs confirm with 100% confidence that they adopt the overall fold shown for the RGMoV CP and tombusvirus-like viral CPs. In addition to protein–protein and protein–RNA interactions, each icosahedral unit is stabilized by three calcium-binding sites located between subunits AB, BC, and CA. Studies with SCPMV, SeMV, and RYMV particles

demonstrate that the stability of the virions depends greatly on pH and the availability of calcium ions. Upon alkaline pH or removal of the cations, the virus particles swell and become less stable.

The primary sequences of CPs among solemoviruses are quite different. However, the three-dimensional structures are nearly identical. The CP is divided into two domains: C-terminal S (shell) domain, which has an eight-strand jellyroll β -sandwich topology, common to non-enveloped icosahedral viruses, and N-terminal R (random) domain, which is disordered in subunits A and B, but is partially structured in subunit C. The disordered beginning of R domain seems to interact with the genomic RNA in the interior of the virus particle. The S domain is the building block of the virion. The R domain is rich in basic amino acid residues and contains at its first half an arginine-rich motif (ARM). The ARM comprises a bipartite nuclear targeting signal (NLS). In the case of CfMV, abolishment of the NLS does not affect virus infectivity. Studies with SCPMV and SeMV CPs demonstrate that ARM is essential for RNA encapsidation but not for particle formation. However, the presence of RNA enhances the overall stability of capsids.

Packing of viral nucleic acid is supposed to happen when CP recognizes a specific region inside the virus genome. In the case of SCPMV, a putative stem-loop sequence within a conserved region encoding serine protease has been reported to bind CP. Nevertheless, this sequence has not been demonstrated to nucleate SCPMV assembly.

According to a number of studies, solemovirus particles assemble as follows: CP subunits with disordered amino termini form a pentamer of AB dimers. Then, ARM interacts with RNA and CC dimers are formed. More CC dimers are added and a swollen T=3 particle is built that becomes compacted when calcium ions enter the subunit interfaces.

Virion disassembly occurs when cations are removed. In the case of SBMV virions, pentamers have been suggested as exit ports for viral RNA. Particles of this virus disassemble completely after initiation of RNA translation.

In addition to the genomic and sgRNA, some solemoviruses (Lucerne transient streak virus (LTSV), *Solanum nodiflorum* mottle virus (SNMoV), SCMoV, Velvet tobacco mottle virus (VTMoV), and RYMV) encapsidate a circular viroid-like satellite RNA (satRNA). The sizes of these circular satRNAs range from 220 to 390 nt, RYMV satRNA being the smallest (220 nt) known naturally occurring circular RNA.

Genome

Solemoviral genomes are 4.0–4.6 kb in length. The polycistronic positive-sense ssRNA genome consists of five open reading frames (ORFs) in the case of sobemoviruses and four in the case of polemoviruses (Fig. 2). All ORFs, except the most 3' proximal, are expressed from the genomic template. The last ORF, encoding CP, is translated from the sgRNA.

The genomic and sub-genomic RNAs have a VPg covalently bound to its 5' end. The 3' terminus of the genomic RNA is non-polyadenylated. An approximately 100 nt long 5' UTR and up to 289 nt long 3' UTR are supposed to compensate the absence of cap and poly(A) tail. VPg may attach to the genome via phosphodiester bond formed through tyrosine, serine, or threonine residues, depending on the viral species. The 5' UTR of sobemoviruses comprise a polypurine tract and 3' UTR are more diverse. In the case of CfMV, 5' UTR acts as an enhancer of translation and 3' UTR seems to contain sequences or structures important for the transport of viral RNA. For this virus, as well as for RYMV, a potential tRNA-like structure has been predicted at the 3' UTR. SeMV 3' UTR has a stem-loop 28–55 nt upstream of the 3' end that is important for RNA-dependent RNA polymerase (RdRp) template recognition. Interestingly, also the polemovirus Poinsettia latent virus (PnLV) has a stable hairpin at its 3' UTR, which buries the last nucleotide. Not all, but most of the solemoviral genomes – including PnLV – have a conserved 5' sequence ACAA(AA), considered to play a role in viral RNA replication. This motif is also present upstream of the translation initiation codon of CPs, as a possible 5' terminus of sgRNA.

In addition to the approximately 1 kb sgRNA encoding the CP, PnLV has a larger putative sgRNA of 2.60 kb. Further studies are needed to find out whether this is a real sgRNA or a defective RNA. In fact, sobemovirus CfMV particles may contain different defective interfering RNAs (diRNA).

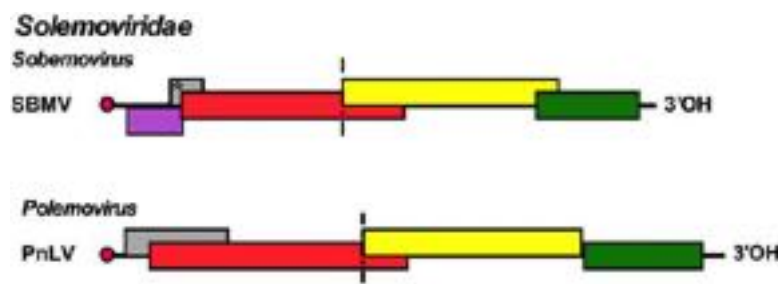
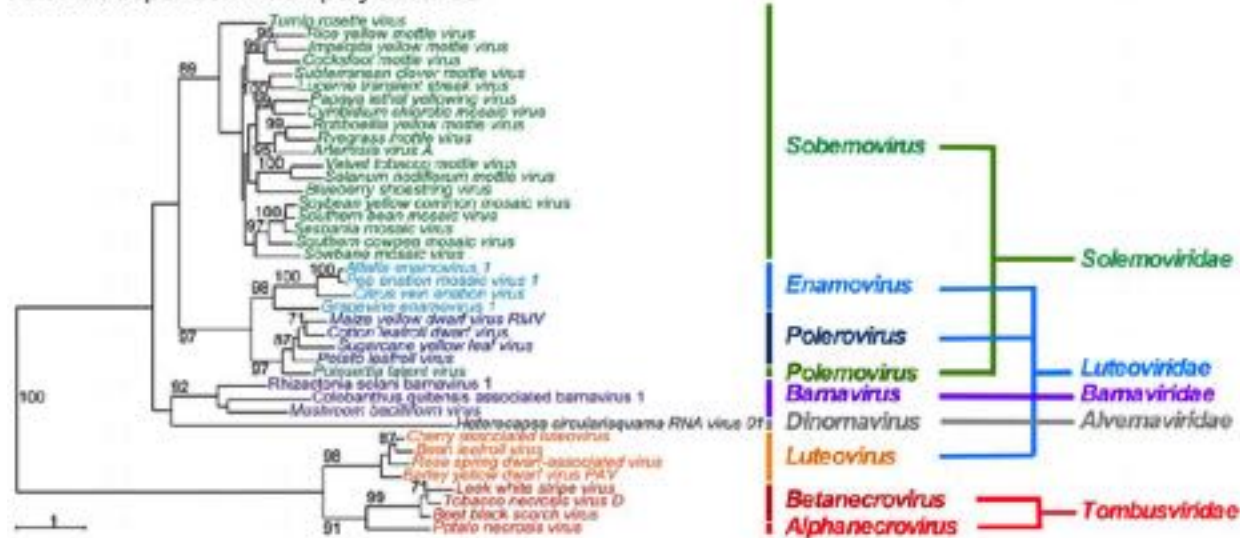


Fig. 2 Genome organization of the genera *Solemovirus* and *Polemovirus* in the family *Solemoviridae*. Comparison of the genome organization of Southern bean mosaic virus (SBMV) and Poinsettia latent virus (PnLV), type species of the corresponding genera. ORFs encoding proteins sharing sequence homology are indicated in similar colors: silencing suppressor (violet), serine protease-VPg-s (red), RNA-dependent RNA polymerases (yellow), coat proteins (green). VPg-s are depicted as red dots at the 5'-end of the genomic RNA. Proteins with unknown function and no sequence homology are shown in gray. Non-canonical AUG start codon is marked with asterisk (*). – 1 ribosomal frameshifting is marked by long vertical dashed line.

A. RNA-dependent RNA polymerases



B. Coat proteins

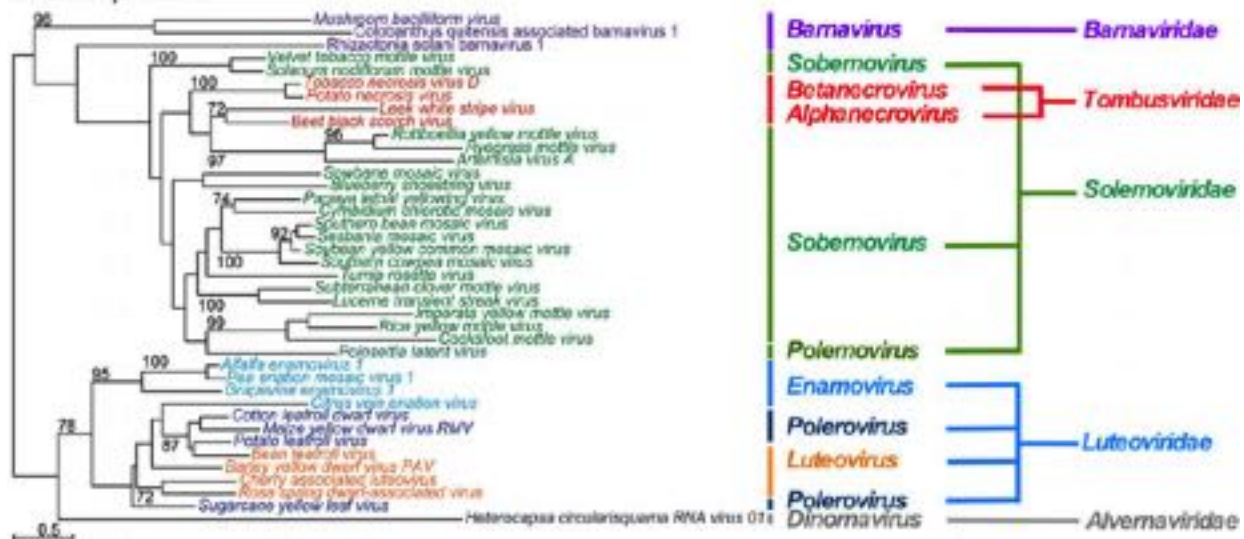


Fig. 3 Maximum-likelihood trees showing phylogenetic relationships between solemoviruses and viruses belonging to related families. Phylogenetic relationships between members of the genera *Sobemovirus* (family *Solemoviridae*, shown in green), *Polemovirus* (family *Solemoviridae*, shown in teal), *Polemovirus* (family *Luteoviridae*, shown in blue), *Enamovirus* (family *Luteoviridae*, shown in turquoise), *Luteovirus* (family *Luteoviridae*, shown in orange), *Alphaneecrovirus* and *Betanecrovirus* (family *Tombusviridae*, shown in red), *Barnavirus* (family *Barnaviridae*, shown in violet), and *Dinornavirus* (family *Alverniviridae*, shown in gray) demonstrate the discordant placement of these genera on the unrooted maximum-likelihood (ML) trees. ML trees based on amino acid sequence alignments were predicted for RNA-dependent RNA polymerases (RdRps, panel A) and for coat proteins (CPs, panel B). Genera *Polemovirus* and *Luteovirus* are represented by a reduced number of species. Species officially recognized by ICTV are shown in italics. The ML trees were calculated with 1000 bootstrap replicates using the Whelan and Goldman (WAG) model for RdRps and Le Gascuel (LG) model for CPs. The percentage of trees in which the associated taxa clustered together is shown next to the branches if >70. Branch lengths are proportional to genetic divergence. Scale indicates the number of amino acid substitutions per site.

Solemoviruses differ in the 5' terminal part of their genomes. This is part of the differences between the members of the genera *Polemovirus* and *Sobemovirus*. Indeed, the name *Polemovirus* comes from *polemovirus* and *sobemovirus* to indicate the chimeric nature of this genus: the 5'-three quarters of the genome are similar in organization and sequence to polemoviruses (polemoviral RdRp groups together with polemoviral RdRps, Fig. 3(A)) and the CP shares the highest homology with sobemoviruses (Fig. 3(B)).

The most 5' proximal ORF in the case of polemovirus is ORF0 that encodes a protein of 30 kDa of unknown function. Sobemoviruses first ORF is called ORF1 and encodes the protein P1, a suppressor of RNA silencing. P1 is dispensable for replication and is suspected to be the viral movement protein, as it has been shown to be required for systemic infection. Moreover, in the case of SeMV, P1 has been shown to interact with its genome-bound VPg and with P10, an ATPase that could

supply the needed energy for an active transport. It remains to be elucidated, if P1 RNA silencing suppressor function is uncoupled from its role in viral movement. The only conserved element among sobemoviral P1 is a Zn-finger motif that is required for suppressing RNA silencing. ORF1 partially overlaps the small ORF_x from which the P_x protein is translated. The translation initiation codon of ORF1 is located in an unfavorable context and therefore a leaky scanning of ribosomes occurs. Since P_x translation starts from a non-AUG codon, protein synthesis rates are very low. P_x is a protein of unknown function and it has been shown to be needed for viral infection in the case of Turnip rosette virus (TRoV). ORF_x and ORF1 are the most variable regions in sobemoviral genomes. No small ORF has been detected for polemovirus PnLV.

All solemoviruses comprise at the central part of their genome two overlapping ORFs. The second ORF always encodes the RdRp that is translated via -1 ribosomal frameshifting. Polemoviral ORF1 contains putative helicase and protease domains as well as a predicted VPg sequence in-between (altogether the product is 72 kDa), while ORF2 encodes the RdRp (122 kDa). In the case of sobemoviruses, the ineffective initiation of translation of P1 and P_x proteins allows the expression of ORF2a that encodes the most abundant polyprotein P2a. ORF2a encodes an N-terminal trans-membrane anchor (TM), a virus serine protease (Pro) and the VPg. The exact functions of the proteins encoded by the C-terminal portion of ORF2a are unclear, but this part has been shown to have ATPase and NTP-binding properties and is further processed to P10 and P8, at least in the case of SeMV. Sobemoviral ORF2b is expressed as a fusion protein with ORF2a through -1 ribosomal frameshifting shortly after VPg domain. For CfMV it is known that in 10%–20% of the cases the longer version polyprotein P2ab (TM–Pro–VPg–RdRp) is translated instead of the regular polyprotein P2a (TM–Pro–VPg–P10–P8). The -1 ribosomal frameshift is needed to regulate the production of sobemoviral RdRp, which is encoded by ORF2b and involves two elements: a slippery sequence conserved among sobemoviruses (UUUAAAC) and a stem-loop located 7 or 8 nt downstream of it. PnLV slippery sequence for RdRp translation seems to be GGGAAAC, shared by some poleroviruses.

Proteolytic cleavages in the polyproteins are carried out by the viral serine protease that cuts at E/T residues for SBMV, at E/N and E/T residues for CfMV, at E/S and E/T residues for RYMV and at E/S residues in the case of RGMoV. The C-terminal cleavage of VPg takes place upstream of the slippery sequence. SeMV is the only solemovirus for which all four cleavage sites have been characterized and its protease domain has been crystallized. The structure exhibits the characteristic features of trypsin fold. Mutations of glutamate-binding site residues H298, T279, and N308 render the protease inactive. Pro–VPg is crucial for protease activity and this precursor protein accumulates in membranes of infected cells. Notably, Pro–VPg is released from P2a, while Pro is detected when P2ab is expressed. TM putative function is to anchor the polyprotein into cellular membranes to facilitate proteolytic processing and probably minus strand synthesis.

If no VPg processing occurs, the virus is not viable. VPg is an intrinsically disordered protein that is able to bind different partners. SeMV VPg has been shown to interact with P1 but not with RdRp. Importantly, RYMV VPg interacts with eukaryotic translation initiation factor eIF(iso)4G and it seems that also P8 is involved in this binding. SeMV P8 contains an RNA binding region and is needed for the ATPase activity of P10 in the precursor P10–P8. SeMV P10 partners are P1, as mentioned before, and RdRp, meaning that this protein may be related to virus movement and replication.

According to their sequences, RdRp-s of solemoviruses are grouped together with the RdRps of polero-, enamo-, and barnaviruses (Fig. 3(A)). This is the so-called “sobemo-lineage” in the supergroup I of plus-sense plant RNA viruses. RdRps of this group seem to lack the NTP-binding region conserved in viral helicases. The GDD motif (SGSYCTSSSTNX_{19–35}GDD) that characterizes RdRps of RNA viruses with ssRNA and positive polarity genomes is essential. Studies on SeMV RdRp have shown that it can synthesize RNA in a primer-independent manner. This polymerase interacts at its C-terminal region with P10 that apparently catalyzes RdRp activity.

Some solemoviral CPs are related to processes different to capsid formation. CP of SeMV has been reported to interact with P1, protein related to viral movement, as explained before. Interestingly, CP of CfMV has been shown to act as suppressor of RNA silencing, in addition to P1.

Phylogenetic Relationships

The viruses belonging to the families *Solemoviridae*, *Luteoviridae*, and *Tombusviridae* seem to have evolved via recombinational shuffling of the genes encoding the protease, RdRp and CP. The sequences of solemoviral long polyproteins are similar to the polyproteins of polero- and enamoviruses (genera belonging to the family *Luteoviridae*), whereas solemoviral CPs share more similarity to CPs of alpha- and betanecroviruses (belonging to the family *Tombusviridae*).

The only known member of the genus *Polemovirus* is the product of a recombination between a polerovirus and a sobemovirus. As mentioned before and as shown in Fig. 3, RdRp of polemovirus PnLV is closely related to poleroviral RdRps, but its CP sequence is highly similar to sobemoviral CPs. It has been proposed that the recombination happened in the presence of a co-infection by switching templates close to the sgRNA start site during replication. Although PnLV shares similar sequences with poleroviruses and sobemoviruses, it is most closely related to the genus *Sobemovirus* than to any other genus and therefore, polemo- and sobemoviruses are grouped into one family.

The phylogenetic trees based on RdRp and on CP sequences show discordant placement of the analyzed species. This corroborates the supposition of possible recombinational events between ancestral solemo-, luteo-, and tombusviruses (Fig. 3). In addition, representatives of other genera (*Barnavirus* and *Dinornavirus*) seem to be involved in this evolutionary process.

Among sobemoviral CPs, sequences from RGMoV, RoMoV, and ArtVA are slightly more similar to CP sequences of several tombusviruses than to other sobemoviral CPs (Fig. 3(B)). As explained above, also the 3D structures of the CPs of these three sobemoviruses are predicted to be closely related to tombusviral CPs.

Studies comparing related viruses at the level of their RdRp sequences have concluded that sobemo-, barna-, polero-, and luteoviruses divergence time is around 9000 years. If luteoviruses are not considered, then the divergence is reduced to 5000 years, and leaving apart poleroviruses means going down to 4000 years. Based on this kind of calculations, a hypothesis correlating the development of agriculture to the appearance of different viruses has been raised.

Replication and Movement

The replication of solemoviruses is poorly understood. The genomic RNA of incoming virus particles is probably uncoated by the co-translational disassembly mechanism that is followed by RNA replication once the RdRp has been synthesized.

Little is known about the signals needed for viral replication. As mentioned before, solemoviral genomes have a conserved 5' sequence ACAA(AA) that is also characteristic for 5'-termini of polero-, diantho-, and barnaviruses. This motif is also present at the 5'-terminus of sgRNA. Due to the conservation of this region, the complementary sequence in minus-strand has been predicted to function in replication by promoting or enhancing the binding of RdRp to both genomic and sgRNA. The stem-loop structure found at the 3'-end of some solemoviral genomes is important for template recognition by RdRp and thus, for the synthesis of minus-strand. It is not yet clear, whether solemoviral VPg primes the synthesis of viral RNA. SeMV VPg has been shown to be dispensable for *in vitro* synthesis of minus-strand RNA.

At least five diRNA molecules have been cloned from CfMV-infected plants. Their sequences consist of the 5'-terminal 35–40 nt linked to the 850–950 nt of the 3'-terminus. This indicates the possibility that RdRp is able to switch templates and create recombinant RNA molecules.

In addition to the viral genome, RdRps of some solemoviruses also replicate satRNAs. These are 220–390 nt circular molecules that are replicated by a helper virus. satRNAs have several interesting interactions with the replication of the helper viruses. For example, LTSV supports the replication of satRNA of SNMoV but SNMoV does not replicate LTSV satRNA. At the same time, LTSV satRNA replication is supported by CfMV, SBMV, Sowbane mosaic virus (SoMV), and TRoV, whereas this support is host dependent. Importantly, satRNAs influence sometimes the symptoms caused by the helper virus. satRNAs are predicted to have viroid-like secondary structures and replicate via rolling circle mechanism. After replication, monomers are self-cut by endogenous hammerhead ribozyme structures. RYMV satRNA is the smallest described so far (220 nt long). Remarkably, LTSV satRNA encodes a protein of 16 kDa with unknown function. Since the length of the circular genome is not a number divisible by three, translations happen one after another continuing from different codons in each round.

Solemoviral particles are detected mainly in mesophyll and vascular tissues, but also in epidermal, bundle sheath, guard cells, and even in meristem cells, depending on the viral species. Concerning vascular tissues, virus particles have been reported in both xylem and phloem: CfMV, SCPMV, and SBMV particles have been found mainly in the phloem, whereas RYMV, SoMV, and Blueberry shoestring virus (BSSV) virions have been detected predominantly in the xylem. CfMV particles have been reported in phloem parenchyma and bundle sheath cells, when infection starts. In later stages of infection, CfMV spreads in mesophyll cells surrounding vascular bundles and is only seldom present in xylem parenchyma.

At the subcellular level, virus particles are found in cytoplasm, vacuoles, and nuclei. Some sobemoviruses have NLS signals in their CP sequences that make a nuclear localization of the virion feasible. At late stages of infection, particles accumulate forming crystalline structures in the cytoplasm or vacuoles. No particles have been detected in mitochondria and chloroplasts, but the latter are noted to form finger-like extrusions in infected cells.

Studies with solemoviruses emphasize that cell-to-cell and long-distance (systemic) movement are two distinct processes. For the latter movement, correct particle formation may be essential. This is the case for SCPMV and RYMV. Indeed, studies with full-length clones of these two viruses have shown that CP is dispensable for virus replication but systemic virus movement is completely abolished in the absence of CP. In the case of SeMV, it has been proven that its virion and CP itself interact with P1, the protein related to virus movement. On the contrary, experiments with CfMV lacking CP demonstrated that this virus does not need CP for cell-to-cell or for systemic movement, implying that it moves as a ribonucleoprotein complex and not as viral particle. However, CP-deficient CfMV was not able to be transmitted mechanically using sap-inoculation.

An efficient transmission of solemoviruses by vectors is facilitated by a correct virion formation that is resistant to RNases present in the regurgitant or in the hemolymph and even in feces of insects.

Host Range and Transmission

The host range of individual solemoviruses is narrow – limited to one plant family – with the exception of SoMV that naturally infects plants from the families *Chenopodiaceae*, *Vitaceae*, and *Rosaceae*. Genetic relationships are stronger between virus species infecting plants from the same family. Most of the so far sequenced solemoviruses infect dicotyledonous plants. According to phylogenetic data, commelinid (*Poales*)-infecting sobemoviruses have emerged at least twice during diversification of sobemovirus species.

The main route of transmission for solemoviruses is mechanical wounding of plants. In addition, transmission from soil has been suggested for many solemoviruses. The polemovirus PnLV is distributed by grafting and vegetative propagation of the host plant; however, the natural means of transmission remains unknown. Viruses in the genus *Sobemovirus* are transmitted by vectors like beetles, aphids, bugs, and thrips. Interestingly, phylogenetically related sobemoviruses may have vectors that belong to different orders. In the case of RYMV, it has been shown that the virus is transmitted over short distances by herbivore mammals and trampling livestock. Importantly, several sobemoviruses are seed-transmissible.

Pathological and Epidemiological Aspects

The polemovirus PnLV is spread worldwide in cultivated poinsettia (*Euphorbia pulcherrima*) plants without causing symptoms. Equally, some sobemoviruses do not provoke symptoms in their host plants. However, in most cases, solemoviral infections induce symptoms like chlorosis at different levels, mosaic or mottling patterns on leaves. In addition, stunting, necrosis, vein clearing, and even sterility have been reported. The outcomes depend on the specific virus, the host and other factors like environmental conditions and the number of viral particle present in the infected plant.

As said, solemoviruses form crystalline arrays and tubules in the cytoplasm. Some of these structures are enveloped in endoplasmic reticulum-derived vesicles. Viral particles from some solemoviruses are associated with chloroplast membranes and thus, induce changes in the structure of this organelle. In CfMV- and SoMV-infected cells, a proliferation of tonoplast membranes that bulge into the vacuole takes place. Many cytological changes have been observed in RYMV-infected cells. Probably the most dramatic change induced by this virus occurs in the cell walls of parenchyma and mature xylem cells, where middle lamellae of the wall are disorganized.

RYMV infection also causes important changes in the abundance of host proteins. For instance, the expression levels of several defense- and stress-related proteins like superoxide dismutase and different heat shock proteins increase several times. The changes in the levels of antioxidant enzymes and production of reactive oxygen species has been analyzed in cocksfoot plants infected with CfMV. Susceptible plants and plants with acquired resistance showed completely different regulation of the levels of antioxidant enzymes and reactive oxygen species after inoculation with CfMV.

Natural resistance to solemoviruses has been detected at least for CfMV in cocksfoot, for SBMV in beans, for SCMoV in subterranean clover, for SCPMV in cowpea and for RYMV in different *Oryza* species including rice. The molecular mechanisms conferring resistance have only been described for RYMV in *Oryza* species. Pathogen-derived transgenic resistance against RYMV has been achieved by transforming plants with constructs expressing parts of ORF2. However, this RNA silencing-based resistance is low compared to natural resistance that is being exploited to obtain RYMV-resistant rice varieties.

Further Reading

- Adams, M.J., Jeske, H., 2012. Genus *polemovirus*. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus taxonomy: Classification and Nomenclature of Viruses*. United States: Elsevier Academic Press, pp. 1181–1184.
- Albar, L., Bangratz-Reyser, M., Hébrard, E., *et al.*, 2006. Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to Rice yellow mottle virus. *The Plant Journal* 47, 417–426.
- aus dem Siepen, M., Pohl, J.O., Koo, B.J., Wege, C., Jeske, H., 2005. Poinsettia latent virus is not a cryptic virus, but a natural polerovirus-sobemovirus hybrid. *Virology* 336, 240–250.
- Gayathri, P., Satesh Kumar, P.S., Prasad, K., *et al.*, 2006. Crystal structure of the serine protease domain of Sesbania mosaic virus polyprotein and mutational analysis of residues forming the S1-binding pocket. *Virology* 346, 440–451.
- Ling, R., Pate, A.E., Carr, J.P., Firth, A.E., 2013. An essential fifth coding ORF in the sobemoviruses. *Virology* 446, 397–408.
- Mäkeläinen, K., Mäkinen, K., 2005. Factors affecting translation at the programmed – 1 ribosomal frameshifting site of Cocksfoot mottle virus RNA *in vivo*. *Nucleic Acids Research* 33, 2239–2247.
- Meier, M., Truve, E., 2007. Sobemoviruses possess a common CfMV-like genomic organization. *Archives of Virology* 152, 635–640.
- Nair, S., Savithri, H.S., 2010. Natively unfolded nucleic acid binding P8 domain of SeMV polyprotein 2a affects the novel ATPase activity of the preceding P10 domain. *FEBS Letters* 584, 571–576.
- Olsper, A., Peil, L., Hébrard, E., Fargette, D., Truve, E., 2011. Protein-RNA linkage and post-translational modifications of two sobemovirus VPgs. *Journal of General Virology* 92, 445–452.
- Plevka, P., Tars, K., Zeltins, A., *et al.*, 2007. The three-dimensional structure of Ryegrass mottle virus at 2.9 Å resolution. *Virology* 369, 364–374.
- Sömera, M., Truve, E., 2013. The genome organization of lucerne transient streak and turnip rosette sobemoviruses revisited. *Archives of Virology* 158, 673–678.
- Sömera, M., Truve, E., 2015. Rottboellia yellow mottle virus is a distinct species in the genus *Sobemovirus*. *Archives of Virology* 160, 857–863.
- Sömera, M., Sarmiento, C., Truve, E., 2015. Overview on sobemoviruses and a proposal for the creation of the family Sobemoviridae. *Viruses* 7, 3076–3115.
- Tamm, T., Suurväli, J., Lucchesi, Y., Olsper, A., Truve, E., 2009. Stem-loop structure of Cocksfoot mottle virus RNA is indispensable for programmed – 1 ribosomal frameshifting. *Virus Research* 146, 73–80.
- Tars, K., Zeltins, A., Liljas, L., 2003. The three-dimensional structure of Cocksfoot mottle virus at 2.7 Å resolution. *Virology* 310, 288–289.

Tenuiviruses (*Phenuiviridae*)

Bertha Cecilia Ramirez, The Institute for Integrative Biology of the Cell, The French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

Anne-Lise Haenni, Institut Jacques Monod, French National Center for Scientific Research, Paris Diderot University, Paris, France

© 2021 Elsevier Ltd. All rights reserved.

This is an update of B.C. Ramirez, Tenuivirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00510-0.

Nomenclature

aa Amino acid(s)
IR Intergenic region.
kb Kilo base
kDa Kilo Dalton
MP Movement protein
NC Nucleocapsid protein
ORF Open reading frame

RdRp RNA-dependent RNA polymerase
RNP Ribonucleoprotein
sgRNA sub-genomic RNA
ssRNA single-stranded RNA
vcRNA viral complementary RNA
vRNA virion RNA
VSR Viral suppressor of RNA silencing

Glossary

Ambisense polarity The presence of open reading frames (ORFs) in both the vRNA and its vcRNA, means that both strands encode proteins. This expression strategy requires strong transcription termination signals to avoid the synthesis of complementary RNA and the formation of double stranded RNA that may induce RNA silencing.

Cap-snatching A mechanism of viral mRNA synthesis by the viral RdRp that cleaves off several nucleotides from the 5' terminus of host capped mRNAs and used them as primers to initiate transcription.

Movement protein This refers to a viral protein that in plants assists in cell-to-cell movement of viral complexes through the plasmodesmata before systemic transport through the plant vascular system.

Persistent viral transmission This is a mode of transmission of a plant virus by an insect vector in which the virus after entering the body of the vector, replicates and circulates for systemic invasion of vector insect tissues during days to weeks. The virus passes through the insect's midgut and salivary gland barriers for transmission to new host plants.

RNA silencing An evolutionarily conserved mechanism that is active in many eukaryotic organisms (plants and insects) as antiviral defense. It is a sequence-specific gene-silencing process induced by double-stranded RNA that is either targeted for degradation by the RNA-induced silencing complex or for translational repression.

RNA silencing suppressor A viral factor or protein that counteracts the host plant RNA silencing defense response.

Introduction

The tenuiviruses were first described in the Fifth Report of the International Committee on Taxonomy of Viruses in 1982 as non-enveloped plant viruses possibly possessing a negative ssRNA genome. Their name comes from the Latin 'tenuis', (thin, fine, weak) that refers to the structure of the viral particle as seen by electron microscopy (Fig. 2). Epidemics of Rice stripe virus (RSV) and Rice hoja blanca virus (RHBV), two members of the genus *Tenuivirus*, cause important yield losses in rice-growing areas of Asia and the former USSR, and of tropical South America, respectively. Tenuiviruses exhibit unique properties that make them different from other plant viruses.

Some properties of tenuiviruses are the following:

- (1) The peculiar flexuous viral particles have a thread-like morphology and can adopt circular forms (Fig. 2 Left).
- (2) A non-structural protein (NS4) also designated disease specific (DS) protein (16–22 kDa) accumulates in large amounts in infected plants, forming large inclusions (Fig. 2 Right).
- (3) The viruses are persistently transmitted by a species of planthopper (Fig. 5 Bottom) in a propagative manner. For some of the members of the genus it has been demonstrated that the virus multiplies both in the host plant and in the insect vector. Multiplication of the virus in the vector may have deleterious effects on the insect. The viruses can be transmitted transovarially by viruliferous female planthoppers to their offspring, and through sperm from viruliferous males. The ability to disseminate and to infect reproductive tissues is a prerequisite for the transovarial transmission of tenuiviruses.

- (4) The genome of tenuiviruses is multi-segmented and composed of ssRNAs that have either a negative or an ambisense polarity (Fig. 3).
- (5) An RdRp is associated with the viral particle.
- (6) It has been observed for some tenuiviruses that the mRNAs are synthesized via cap-snatching (Fig. 4).
- (7) The presence of a highly conserved octanucleotide at the 5' (5'-ACACAAAG) and 3' (CUUUGUGU-3') termini of their genomic segments.
- (8) Tenuiviruses infect crop plants of the family *Poaceae*.

Taxonomy

In 2016 the ICTV grouped under the order *Bunyavirales* all the negative stranded RNA viruses previously classified in the family *Bunyaviridae* and the unassigned “bunyavirus-like” viruses such as the tenuiviruses, in order to better reflect their evolutionary relationships. A total of 12 families were created in the order *Bunyavirales* and the family *Phenuiviridae* was created to group the related phleboviruses and tenuiviruses. Among these 12 families, 3 of them comprise viruses infecting plants; *Fimoviridae*, *Tospoviridae*, and *Phenuiviridae*. The *Phenuiviridae* family clusters 15 genera, and only the genus *Tenuivirus* hosts plant viruses (Table 1). *Rice stripe tenuivirus* (RStV) is the type species of the genus *Tenuivirus*. Other species in the genus are *Echinochloa hoja blanca tenuivirus* (EHBV), *Iranian wheat stripe tenuivirus* (IWSV), *Maize stripe tenuivirus* (MStV), *Rice grassy stunt tenuivirus* (RGSV), *Rice hoja blanca tenuivirus* (RHBV), and *Urochloa hoja blanca tenuivirus* (UHBV). Tentative species are European wheat striate mosaic virus (EWStMV), Maize yellow stripe virus (MYSV), and Melon chlorotic spot virus (MChSV) (Table 2).

Classification

The criteria for species demarcation are given as follows:

- (1) Vector specificity: transmission by different vector species.
- (2) Different sizes and/or numbers of RNA segments.
- (3) Host range: ability to infect different key plant species.
- (4) Amino acid (aa) sequence identity of less than 85% between corresponding gene products.
- (5) Nucleotide (nt) sequence identity of less than 60% between corresponding non-coding intergenic regions.

An example of species discrimination is that between RStV and MStV. RStV is transmitted by *L. striatellus* and infects 37 graminaceous species including rice and wheat. MStV is transmitted by *P. maidis* and infects maize, occasionally sorghum, and a few other graminaceous plants but not rice or wheat. Finally, RStV isolates have genomes of four RNA segments and the MStV genome consists of five segments.

An example of difficult species demarcation is the group of the hoja blanca viruses (RHBV, EHBV, and UHBV). They have different vectors, different sizes and numbers of RNA segments, different hosts, and the nt sequence identity of their intergenic regions is less than 60%. However, the aa sequences of the four proteins on RNAs 3 and 4 are about 90% identical between RHBV, EHBV, and UHBV. Since four of five criteria are met, these viruses could be considered as distinct species that possibly separated recently and are now diverging with little field contact between them. Phylogenetic analysis of the sequence data from RNA3 and RNA4 of the tenuiviruses shows that RHBV, EHBV, and UHBV are related and form a group distinct from RStV and MStV (Fig. 1). IWSV was recently included as a member of the tenuivirus genus, it is related but distinct from RHBV and EHBV based on the phylogenetic analysis of RNA 3 and RNA 4 (Fig. 1) as well as on serology.

Virion Structure

Morphology

The particles also referred to as ribonucleoproteins (RNPs) are thin filaments that appear circular (Fig. 2 Left) or spiral shaped. The RNPs are 3–10 nm in diameter, with lengths proportional to the sizes of the RNAs they contain. No envelope has been observed.

Physical and Physicochemical Properties

RNP preparations can be separated into four to six components by sucrose density gradient centrifugation; they correspond to the four to six RNA segments coated in the nucleocapsid protein (NC). The buoyant density of the RNP in CsCl when centrifuged to equilibrium is 1.282–1.288 g cm⁻³. The RNA constitutes 5%–12% of the particle weight.

Table 1 List of families and genera in the order *Bunyavirales*. Viruses infecting plants are surlined in light gray, and type species are written in bold

Family	Genus	Species
<i>Arenaviridae</i>		
<i>Cruliviridae</i>		
<i>Fimoviridae</i>	<i>Emaravirus</i>	<i>Actinidia chlorotic ringspot-associated emaravirus</i> <i>European mountain ash ringspot-associated emaravirus</i> <i>Fig mosaic emaravirus</i> <i>High Plains wheat mosaic emaravirus</i> <i>Pigeonpea sterility mosaic emaravirus 1</i> <i>Pigeonpea sterility mosaic emaravirus 2</i> <i>Raspberry leaf blotch emaravirus</i> <i>Redbud yellow ringspot-associated emaravirus</i> <i>Rose rosette emaravirus</i>
<i>Hantaviridae</i>		
<i>Leishbuviridae</i>		
<i>Mypoviridae</i>		
<i>Nairoviridae</i>		
<i>Peribunyaviridae</i>		
<i>Phasmaviridae</i>		
<i>Phenuiviridae</i>	<i>Banyangvirus</i> <i>Beidivirus</i> <i>Goukovirus</i> <i>Horwuvirus</i> <i>Hudivirus</i> <i>Hudovirus</i> <i>Kabutovirus</i> <i>Laulavirus</i> <i>Mobuvirus</i> <i>Phasivirus</i> <i>Phlebovirus</i> <i>Pidchovirus</i> <i>Tenuivirus</i>	<i>Echinochloa hoja blanca tenuivirus</i> <i>Iranian wheat stripe tenuivirus</i> <i>Maize stripe tenuivirus</i> <i>Rice grassy stunt tenuivirus</i> <i>Rice hoja blanca tenuivirus</i> <i>Rice stripe tenuivirus</i> <i>Urochloa hoja blanca tenuivirus</i>
	<i>Wenrivirus</i> <i>Wubeivirus</i>	
<i>Tospoviridae</i>	<i>Orthotospovirus</i>	<i>Bean necrotic mosaic orthotospovirus</i> <i>Calla lily chlorotic spot orthotospovirus</i> <i>Capsicum chlorosis orthotospovirus</i> <i>Chrysanthemum stem necrosis orthotospovirus</i> <i>Groundnut bud necrosis orthotospovirus</i> <i>Groundnut ringspot orthotospovirus</i> <i>Groundnut yellow spot orthotospovirus</i> <i>Impatiens necrotic spot orthotospovirus</i> <i>Iris yellow spot orthotospovirus</i> <i>Melon severe mosaic orthotospovirus</i> <i>Melon yellow spot orthotospovirus</i> <i>Polygonum ringspot orthotospovirus</i> <i>Soybean vein necrosis orthotospovirus</i> <i>Tomato chlorotic spot orthotospovirus</i> <i>Tomato spotted wilt orthotospovirus</i> <i>Watermelon bud necrosis orthotospovirus</i> <i>Watermelon silver mottle orthotospovirus</i> <i>Zucchini lethal chlorosis orthotospovirus</i>
<i>Wupedeviridae</i>	<i>Wumivirus</i>	
Unassigned	<i>Coguvirus</i>	<i>Citrus coguvirus</i>

Table 2 List of member species of the genus *Tenuivirus*, family *Phenuiviridae*. The type species is written in bold

<i>Species/Virus name</i>	<i>Acronym</i>	<i>RNA #</i>	<i>RefSeq</i>	<i>Size (Kb)</i>
<i>Echinochloa hoja blanca tenuivirus</i>				
Echinochloa hoja blanca virus	(EHBV)	RNA 3	NC_038934.1	2.34
		RNA 4	NC_038935.1	1.87
		RNA 5	NC_038936.1	1.33
<i>Iranian wheat stripe tenuivirus</i>				
Iranian wheat stripe virus	(IWStV)	RNA 2	NC_038748.1	3.47
		RNA 3	NC_038750.1	2.34
		RNA 4	NC_038749.1	1.83
<i>Maize stripe tenuivirus</i>				
Maize stripe virus	(MStV)	RNA 2	NC_038751.1	3.34
		RNA 3	NC_038754.1	2.36
		RNA 4	NC_038752.1	2.23
		RNA 5	NC_038753.1	1.32
<i>Rice grassy stunt tenuivirus</i>				
Rice grassy stunt virus	(RGSV)	RNA 1	AB009656.1	9.76
		RNA 2	AB010376.1	4.06
		RNA 3	AB010377.1	3.12
		RNA 4	AB010378.1	2.92
		RNA 5	AB000403.1	2.70
		RNA 6	AB000404.1	2.58
<i>Rice hoja blanca tenuivirus</i>				
Rice hoja blanca virus	(RHBV)	RNA 1	NC_036597.1	9.00
		RNA 2	NC_036598.1	3.62
		RNA 3	NC_036602.1	2.30
		RNA 4	NC_036599.1	2.00
<i>Rice stripe tenuivirus</i>				
Rice stripe virus	(RStV)	RNA 1	NC_003755.1	8.97
		RNA 2	NC_003754.1	3.51
		RNA 3	NC_003776.1	2.50
		RNA 4	NC_003753.1	2.16
<i>Urochloa hoja blanca tenuivirus</i>				
Urochloa hoja blanca virus	(UHBV)	RNA 3	NC_038757.1	2.27
		RNA 4	NC_038758.1	1.86
Unassigned				
European wheat striate mosaic virus	(EWStMV)	RNA 1	MN160329.1	9.50
		RNA 2	MN160345.1	3.45
		RNA 3	MN160360.1	2.06
		RNA 4	MN160376.1	1.89
Maize yellow stripe virus	(MYStV)	RNA 1	AJ969412.1 ^a	1.15
		RNA 2	AJ969413.1 ^a	1.09
		RNA 3	AJ969414.1 ^a	1.21
		RNA 5	AJ969416.1 ^a	1.56
Melon chlorotic spot virus	(MChSV)	RNA 1	NC_040450.1	9.10
		RNA 2	NC_040451.1	1.85
		RNA 3	NC_040448.1	1.60
		RNA 4	NC_040454.1	1.59
		RNA 5	NC_040449.1	1.55
		RNA 6	NC_040452.1	1.51
		RNA 7	NC_040453.1	1.49
		RNA 8	NC_040455.1	1.41

^aPartial sequence.

Components

The RNPs contain RNA coated with the NC protein of 34–35 kDa, and small amounts of the RdRp of 230–330 kDa. The RdRp associated with the RNPs of RStV and RHBV is capable of replicating and transcribing the RNA segments *in vitro*. The tenuivirus genome is composed of four to six non-capped ssRNA segments. The approximate size of the genome of the type member RStV is 17 kb (**Fig. 3**). The largest RNA segment (9.0 kb) of RStV, MStV, and RHBV is of negative polarity and encodes the RdRp. RNA2 (3.3–3.6 kb), RNA3 (2.2–2.5 kb) and RNA4 (1.9–2.2 kb) of RStV, MStV, IWSV, EHBV, UHBV, and RHBV are ambisense. RNA5 (1.3 kb), detected in virions of MStV and of EHBV is of negative polarity. A fifth RNA segment has also been reported for some

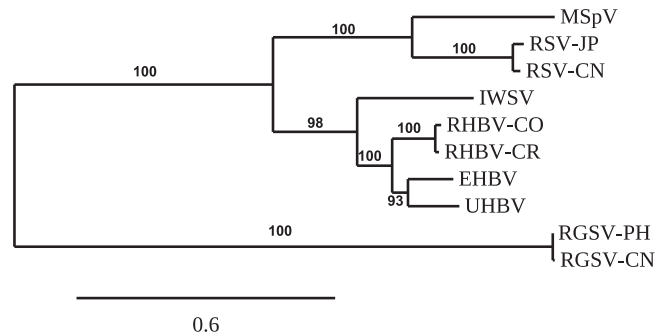


Fig. 1 *Phylogenetic relationships of tenuiviruses.* Phylogenetic tree showing the relationships between tenuiviruses on the basis of the nucleotide sequences of the coding regions of RNAs 3 and 4 for all tenuiviruses except RGSV for which the coding regions of RNAs 5 and 6 were used. Shown is the percentage bootstrap support at each node. Echinochloa hoja blanca virus (EHBV), Iranian wheat stripe virus (IWSV), Maize stripe virus (MSpV), Rice grassy stunt virus (RGSV), Rice hoja blanca virus (RHBV), Rice stripe virus (RStV) and Urochloa hoja blanca virus (UHBV). The following sequences were used: EHBV (L48441, L75930), IWSV (AY312436, AY312435), MSpV-US (L13438, M57426), RGSV-PH (AB000404, AB000403), RGSV-CN (AF287949, AF290947), RHBV-CO (L14952, AF004658), RHBV-CR (AF004657, L07940), RSV-JP (D10979, X53563), RSV-CN (Y11096, Y11095) and UHBV (U82446, U82447). The MUSCLE multiple alignments and the phylogenetic tree were performed using Phylogeny. fr (<http://www.phylogeny.fr/index.cgi>), as described in Dereeper, A., Guignon, V., Blanc, G., *et al.*, 2008. Phylogeny. fr: Robust phylogenetic analysis for the non-specialist. Nucleic Acids Research. 36 (Web Server issue), W465–W469.

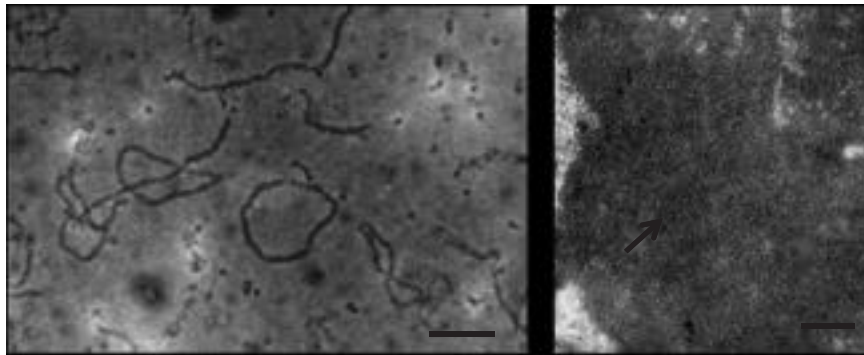


Fig. 2 *Structure of tenuiviruses.* (Left) Electron micrograph of sucrose density gradient purified RNP of Rice hoja blanca virus (RHBV) The bar represents 100 nm. A circular RNP is indicated by the arrow. (Right) Ultrathin section of an RHBV-infected rice leaf cell. The arrow indicates viral inclusion bodies inside a nucleus showing cytopathic effects. The bar represents 500 nm. Courtesy of Francisco Morales.

isolates of RStV. RGSV RNA1, 2, 5, and 6 are homologous to RNA1, 2, 3, and 4, respectively, of other tenuiviruses. RGSV RNA3 (3.1 kb) and RNA4 (2.9 kb) are ambisense and unique to RGSV.

Genome Organization

The 5' and 3' terminal sequences (about 20 nt) are complementary to each other and can base-pair to give rise to circular RNPs. The terminal 8 nt (5' ACACAAAG) and their complement are conserved between tenuiviruses and viruses of the genus *Phlebovirus*. Several RNA segments encode two proteins in an ambisense arrangement (Fig. 3).

The RStV gene products of each RNA segment are: - The RdRp (337 kDa) is encoded by the viral complementary vcRNA1, - The p2 (23 kDa) is present in the RNA2 and the pC2 (94 kDa) in the vcRNA2, - The p3 (24 kDa), also designated NS3, is encoded by the vRNA3. NS3 is a viral suppressor of RNA silencing (VSR) as demonstrated for RStV and for RHBV in both plant and insect hosts. The NC protein (35 kDa), also designated N or pC3, is encoded by the vcRNA3. The major non-structural protein NS4 (21 kDa), also designated p4 or disease-specific protein, is encoded by vRNA4. NS4 accumulates in infected plant tissues in large inclusion bodies. RStV NS4 facilitates tissue tropism in the insect vector. Finally, the pC4 (32 kDa), also designated movement protein (MP) or NSvc4, is encoded by the vcRNA4. The MP is implicated in the passage of viruses from cell-to-cell.

Replication and Transcription

The RdRp synthesizes full-length vRNA and vcRNA that are not capped and sgRNAs of different sizes and of either polarity that serve as mRNA for the synthesis of the corresponding viral proteins. For MStV, RHBV, and RStV, it has been shown that

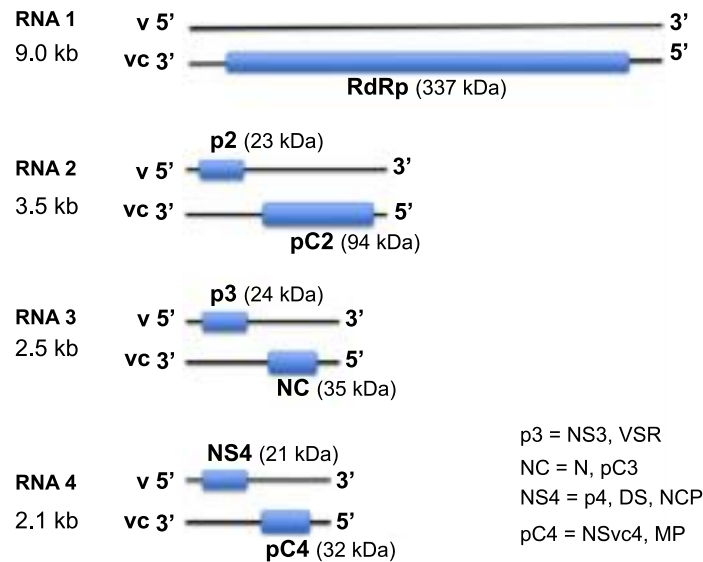


Fig. 3 Schematic representation of the genome organization of *RStV*. v, viral sense RNA. vc, viral complementary RNA. RdRp, RNA-dependent RNA polymerase. NS3, non-structural protein 3 that acts as viral suppressor of RNA silencing (VSR). NC, nucleocapsid protein. NS4, non-structural protein that accumulates in viral inclusion bodies in infected plant tissues; it is also designated disease-specific (DS) or non-capsid protein (NCP). MP, movement protein.

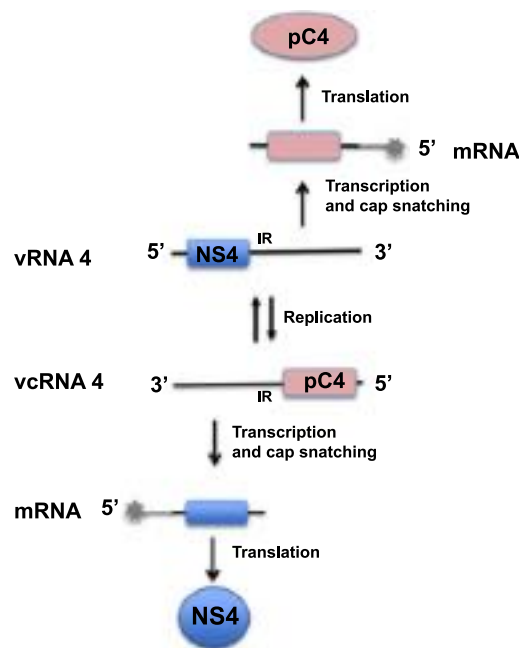


Fig. 4 Schematic representation of the replication, transcription, and translation strategies of the ambisense RNA segments of tenuiviruses. The cap and the non-viral nucleotides of host origin used as primers by the RdRp as a result of cap-snatching are represented at the 5' end of the mRNA. IR, intergenic region.

these mRNAs are synthesized via cap-snatching (**Fig. 4**). The 5' end of the mRNAs contains 10–17 non-viral nt and are capped; these extra sequences are derived from host cell mRNAs that are taken or snatched by the RdRp and are used as primers to initiate mRNA synthesis. Intergenic non-coding regions located between the ORFs can in certain cases adopt hairpin structures. The cap-snatching mechanism has been observed for mRNA synthesis of influenza viruses and viruses of the order *Bunyavirales*.

Biological Properties

Host Range and Geographic Distribution

Plant hosts of tenuiviruses all belong to the family *Poaceae*. Members of the genus *Tenuivirus* cause severe disease in important food cereal crops such as rice and maize and their planthopper vectors themselves are also serious crops pests. Tenuiviruses-induced diseases are economically important resulting in serious yield losses. RStV was estimated in the 1960s to affect around 19% of the total Japan's rice area. RHBV was reported to be responsible of 25%–75% yield losses in Latin American rice, and MSpV-associated losses in maize of 80% were reported from Mauritius.

The geographic distribution of tenuiviruses is determined by the presence of their vector. RStV and RGSV are limited to the Far East, especially in China, Japan, and Korea; RHBV and EHBV to tropical and subtropical regions of the Western Hemisphere, Central and South America, the Caribbean, and the Southern United States; and MSpV has the most widespread geographic distribution around the world but only in subtropical and tropical regions.

Cytopathology

In infected plants large inclusion bodies (Fig. 2 Right) are observed that contain the major non-structural protein NS4.

Disease Symptoms and Control

Typical disease symptoms induced by tenuivirus on host plants are chlorotic stripes and yellow stipplings on the leaf blade between the veins, as well as plant dwarfing and panicle sterility. Young infected plants often exhibit complete chlorosis on the emerging whorl leaf and symptom severity and subsequent yield losses vary with age of the plant at the time of infection. RStV-infected rice leaves show yellow strips and dead tissue streaks. RHBV (Fig. 5 Top) and EHBV induce chlorotic leaf striping and mottling - called white leaf or *hoja blanca* -, plant dwarfing, panicle sterility, premature wilting, and necrosis, causing serious yield losses. MSpV-infected maize plants show early leaf chlorosis between the veins that later develop into continuous stripes, often with a “brush-out” appearance toward the tip of stripes.

RStV-induced disease has been successfully controlled in Japan and China during the last few decades by using naturally resistant rice. RGSV causes significant economic losses in rice production in South, Southeast, and East Asian countries and

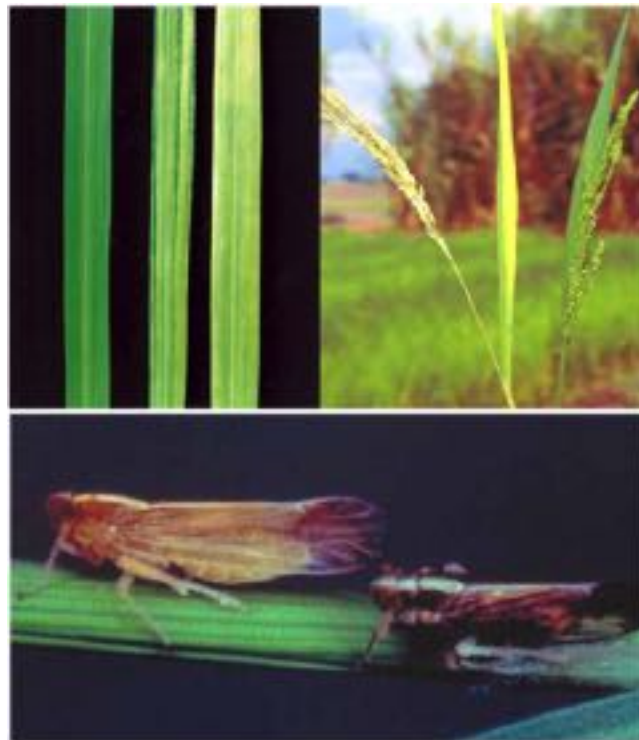


Fig. 5 Disease symptoms and insect vector. (Top Left) RHBV-infected chlorotic rice leaf with *hoja blanca* striping and mottling. (Top Right) RHBV-infected rice panicle. Healthy leaves and panicle are shown for comparison. (Bottom) *Tagosodes orizicolus* vector of RHBV.

suitable resistance genes against RGSV have not yet been found in natural rice resources. A promising method to confer resistance against tenuiviruses is the use of RNA interference to target viral genes NC and MP that play important roles in viral infection and proliferation at an early stage of viral replication. Transgenic plants have been reported to show resistance against RStV and RGSV infection and inhibited the proliferation of the virus.

In the 1980s, the identification of RHBV as the causal agent of the hoja blanca disease of rice together with genetic studies on the interaction of this virus with its planthopper vector, confirmed the pathogenic nature of the virus to its vector and helped to explain the cyclical nature of the epidemics. This disease was first controlled by hybridization of susceptible indica and resistant japonica rice genotypes and the adoption of integrated disease control practices. Recently, new RHBV-resistance sources belonging to the indica type have been identified; making it feasible to use these potential sources in crosses targeting the tropics.

Insect Vector and Type of Transmission

Each virus species is persistently transmitted by a particular species of planthopper in a propagative manner and can be transmitted transovarially by viruliferous female planthoppers to their offspring, and through sperm from viruliferous males. The principal vectors of the species are *Laodelphax striatellus* for RStV, *Tagosodes cubanus* for EHBV, *Peregrinus maidis* for MStV, *Nilaparvata lugens* for RGSV, *Tagosodes orizicolus* for RHBV (Fig. 5 Bottom), *Ukanoes tanasijevici* for IWSV and *Caenodelphax teapae* for UHBV. Known vector of the tentative species is *Javesella pellucida* for EWStMV.

Relation to Other Taxa

Viruses of the genus *Tenuivirus* share several properties with those of the genus *Phlebovirus*. The multi-segmented genomes of tenuiviruses and phleboviruses contain negative-sense and ambisense components. The presence of the highly conserved octanucleotide at 5' (5'-ACACAAAG) and 3' (CUUUGUGU-3') termini of their genomic segments, these terminal complementary sequences can base-pair and may give rise to circular RNPs. The synthesis of mRNAs by the viral RdRp follows a cap-snatching mechanism for viruses of the two genera. In addition, they encode functionally related proteins at equivalent genome positions: the tenuivirus RdRp and 94 kDa proteins are homologous to the RdRp and glycoproteins of other members of the family *Phenuiviridae*. Finally, their infection cycle includes arthropod transmission and the ability to replicate in the arthropod vector.

The different number of genome components and the apparent lack of an enveloped viral particle distinguish tenuiviruses from other viruses of the family *Phenuiviridae*. Tenuiviruses are more complex in terms of the number of genome segments and generally encode more proteins than phleboviruses. However, differences in number of genome segments are not uncommon traits among genera of viruses belonging to the same family (e.g., *Reoviridae*, *Closteroviridae*, *Secoviridae*, *Potyviridae*, *Rhabdoviridae*), or even between members of the same genus (e.g., *Emaravirus*). Although enveloped tenuivirions have not been found, tenuiviruses do encode a glycoprotein closely related to those of phleboviruses (a similar situation can be found in the family *Rhabdoviridae* among members of the genus *Dichorhavirus*).

Further Reading

- de Miranda, J.R., Muñoz, M., Wu, R., Espinoza, A.M., 2001. Phylogenetic placement of a novel tenuivirus from the grass *Urochloa plantaginea*. *Virus Genes* 22, 329–333.
- de Miranda, J.R., Ramirez, B.C., Muñoz, M., *et al.*, 1997. Comparison of Colombian and Costa Rican strains of Rice hoja blanca tenuivirus. *Virus Genes* 15, 191–193.
- Estabrook, E.M., Suyenaga, K., Tsai, J.H., Falk, B.W., 1996. Maize stripe tenuivirus RNA 2 transcripts in plant and insect hosts and analysis of pvc2, a protein similar to the phlebovirus virion membrane glycoproteins. *Virus Genes* 12, 239–247.
- Falk, B.W., Tsai, J.H., 1998. Biology and molecular biology of viruses in the genus tenuivirus. *Annual Review of Phytopathology* 36, 139–163.
- Huiet, L., Feldstein, P.A., Tsai, J.H., Falk, B.W., 1993. The Maize stripe virus major non-capsid protein messenger RNA transcripts contain heterogeneous leader sequences at their 5' termini. *Virology* 197, 808–812.
- Kormelink, R., Garcia, M.L., Goodin, M., Sasaya, T., Haenni, A.-L., 2011. Negative-strand RNA viruses: The plant-infecting counterparts. *Virus Research* 162, 184–202.
- Liu, X., Jin, J., Qiu, P., *et al.*, 2018. Rice stripe tenuivirus has a greater tendency to use the prime-and-realign mechanism in transcription of genomic than in transcription of antigenomic template RNAs. *Journal of Virology* 92 (92), e01414–e01417.
- Nguyen, M., Ramirez, B.C., Golbach, R., Haenni, A.-L., 1997. Characterization of the in vitro activity of the RNA-dependent RNA polymerase associated with the ribonucleoproteins of Rice hoja blanca tenuivirus. *Journal of Virology* 71, 2621–2627.
- Ramirez, B.C., Garcin, D., Calvert, L.A., Kolakofsky, D., Haenni, A.-L., 1995. Capped non-viral sequences at the 5' end of the mRNA of Rice hoja blanca virus RNA4. *Journal of Virology* 69, 1951–1954.
- Ramirez, B.C., Haenni, A.-L., 1994. Molecular biology of tenuiviruses, A remarkable group of plant viruses. *Journal of General Virology* 75, 467–475.
- Ramirez, B.C., Lozano, I., Constantino, L.-M., Haenni, A.-L., Calvert, L.A., 1993. Complete nucleotide sequence and coding strategy of Rice hoja blanca virus RNA4. *Journal of General Virology* 74, 2463–2468.
- Ramirez, B.C., Macaya, G., Calvert, L.A., Haenni, A.-L., 1992. Rice hoja blanca virus genome characterization and expression in vitro. *Journal of General Virology* 73, 1457–1464.
- Shimizu, T., Toriyama, S., Takahashi, M., Akutsu, K., Yoneyama, K., 1996. Non-viral sequences at the 5'-termini of mRNAs derived from virus-sense and virus-complementary sequences of the ambisense RNA segments of Rice stripe tenuivirus. *Journal of General Virology* 77, 541–546.
- Toriyama, S., Kimishima, T., Takahashi, M., 1997. The proteins encoded by Rice grassy stunt virus RNA5 and RNA6 are only distantly related to the corresponding proteins of other members of the genus tenuivirus. *Journal of General Virology* 78, 2355–2363.
- Toriyama, S., Kimishima, T., Takahashi, M., *et al.*, 1998. The complete nucleotide sequence of the Rice grassy stunt virus genome and genomic comparisons with viruses of the genus *Tenuivirus*. *Journal of General Virology* 79, 2051–2058.

Tobacco Mosaic Virus (*Virgaviridae*)

Marc HV Van Regenmortel, University of Strasbourg, Strasbourg, France

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Co-translational disassembly A biological phenomenon by which translation of the tobacco mosaic virus (TMV) occurs simultaneously as the disassembly of the TMV particle.

Cryptotope A determinant (or epitope) of an immunological antigen or immunogen which is initially hidden and becomes functional only when the molecule is broken or degraded.

Heterospecific antibodies Antibodies that are unable to react with the immunogen used to produce them but are capable to recognize immunogens from other species or mutants.

Mapping of epitopes Epitope mapping is the process of experimentally identifying the binding sites, or "epitopes", of antibodies on their target antigens.

Movement protein A movement protein is a non-structural protein which is encoded by some plant viruses to enable their movement from one infected cell to neighboring cells.

Neotope Types of epitopes which depend on the protein quaternary structure and are absent in dissociated protein subunits.

Segmental mobility in proteins The mobility of an antigenic determinant may make it easier to adjust to a pre-existing antibody site not fashioned to fit the exact geometry of a protein.

Serological differentiation index Serological cross-reactivity between two viruses expressed as the number of two-fold dilutions steps separating homologous and heterologous titers.

Viral mutants The term mutant is also applied to a virus with an alteration in its nucleotide sequence.

Virus disassembly The disassembly of TMV particles is initiated when the end of the particles containing the 5' terminus of the RNA becomes associated with ribosomes.

Virus self-assembly Biological phenomenon by which the different components of a virus particles are capable, under certain conditions, to auto-assemble to form complete and functional virus particles.

The Beginnings of Virology

Tobacco mosaic virus occupies a unique place in the history of virology and was in the forefront of virus research since the end of the nineteenth century. It was the German Adolf Mayer, working in the Netherlands, who in 1882 first described a severe disease of tobacco which he called tobacco mosaic disease. He showed that the disease was infectious and could be transmitted to healthy tobacco plants by inoculation with capillary glass tubes containing sap from diseased plants. Although Mayer was not able to isolate a germ as the cause of the disease, he did not question the then prevailing view that all infectious diseases were caused by microbes and he remained convinced that he was dealing with a bacterial disease.

About the same time, in St. Petersburg, Dmitri Ivanovsky was studying the same disease and he reported in 1892 that when sap from a diseased tobacco plant was passed through a bacteria-retaining Chamberland filter, the filtrate remained infectious and could be used to infect healthy tobacco plants. Ivanovsky was the first person to show that the agent causing the tobacco mosaic disease passed through a sterilizing filter and this gave rise to the subsequent characterization of viruses as filterable agents. A virology conference was held in 1992 in St. Petersburg to celebrate the centenary of this discovery. Although Ivanovsky is often considered one of the fathers of virology, the significance of his work for the development of virology remains somewhat controversial because all his publications show that he did not really grasp the significance of his filtration experiments. He remained convinced that he was dealing with either a small bacterium or with bacterial spores and never appreciated that he had discovered a new type of infectious agent.

Following in the footsteps of Mayer, Beijerinck in Delft, Holland, showed in 1898 that sap from tobacco plants infected with the mosaic disease was still infectious after filtration through porcelain filters. He also demonstrated that the causative agent was able to diffuse through several millimeters of an agar gel and he concluded that the infection was not caused by a microbe.

Beijerinck called the agent causing the tobacco mosaic disease a *contagium vivum fluidum* (a contagious living liquid), in opposition to a *contagium vivum fixum*. In those days the term *contagium* was used to refer to any contagious, disease-causing agent, while the term *fixum* meant that the agent was a solid particle or a cellular microbe. On the basis of his filtration and agar diffusion experiments, Beijerinck was convinced that the agent causing tobacco mosaic was neither a microbe nor a small particle or corpuscle (meaning a small body or particle from the Latin *corpus* for body). He proposed instead that the disease-causing agent, which he called a virus, was a living liquid containing a dissolved, non-particular and non-corpuscular entity.

Lute Bos has claimed that Beijerinck's introduction in 1898 of the unorthodox and rather odd concept of a *contagium vivum fluidum* marked the historic moment when virology was conceived conceptually. However, it is clear that his definition of a virus as a soluble, living agent not consisting of particles certainly does not correspond to our modern view of what a virus is. It is Beijerinck's insistence that a virus is not a microbe and his willingness to challenge the then widely held view that all infectious diseases are caused by germs that make many people regard his contribution to the development of virology as more important than that of Ivanovsky. In 1998, a meeting held in the Netherlands commemorated the centenary of the work of Beijerinck on TMV. About the same time a meeting was held in Germany to honor the contribution of Loeffler and Frosch who, also in 1898, had shown that the agent causing the foot-and-mouth disease of cattle was able to pass through a Chamberland-type filter, in the

same way as TMV. In addition the German workers also established that their disease-causing agent was not able to go through a finer grain Kitasato filter, from which they concluded that the agent, which was multiplying within the host, was corpuscular and not soluble as claimed by Beijerinck. Loeffler, however, continued to believe that his pathogen was a very small germ or spore, invisible in the light microscope, that was unable to cross the small pores of a Kitasato filter.

There is no doubt that none of these 'fathers' of virology realized the nature of the filterable pathogenic agents they were investigating and that the term virus which they used did not have the meaning it has today. It took another 30 years before chemical analysis eventually revealed what viruses are actually made of protein and RNA.

Physical and Chemical Properties of TMV

A major change in our perception of what viruses are occurred in 1935 when Stanley, working at the Rockefeller Institute in Princeton, obtained needle-like crystals of TMV that were infectious and consisted of protein. He used methods that were being developed at the time for purifying enzymes and was greatly helped by a bioassay developed by Holmes in 1929 that made it possible to quantify the amount of TMV present in plant sap. In this assay, extracts containing the virus are rubbed on the leaves of *Nicotiana glutinosa* tobacco plants which leads to a number of necrotic lesions proportional to the amount of virus present.

Stanley's demonstration that TMV was a crystallizable chemical substance rather than a microorganism, a discovery that earned him the Nobel Prize in 1946 and had a profound impact on the thinking of biologists because it suggested that viruses were actually living molecules able to reproduce themselves. For a while, it seemed that viruses closed the gap between chemistry and biology and might even hold the key to the origin of life.

Stanley had originally described TMV as a pure protein, but in 1936 Bawden and Pirie showed that the virus contained phosphorus and carbohydrate and actually consisted of 95% protein and 5% RNA. TMV thus became the first virus to be purified and shown to be a complex of protein and nucleic acid. Following the development of ultracentrifuges in the 1940s, ultracentrifugation became the standard method for purifying TMV and many other viruses.

In 1939, TMV became the first virus to be visualized in the electron microscope and in 1941, TMV particles were shown to be rods about 280 nm long and 15 nm wide. Subsequently, X-ray analysis established that the rods were hollow tubes consisting of a helical array of 2130 identical protein subunits with $16\frac{1}{3}$ subunits per turn and containing one molecule of RNA deeply embedded in the protein subunits at a radius of 4 nm (Fig. 1). The length of the TMV particle is controlled by the length of the RNA molecule which becomes fully coated with protein and is thereby protected from nuclease attack.

The viral protein subunits can also aggregate on their own, without incorporating RNA, to form rods very similar to TMV particles but in this case the rods are of variable length (Fig. 2). It is possible to degrade the virus particles with acetic acid or weak alkali and to obtain in this way dissociated protein subunits and viral RNA, the latter being rapidly degraded by nucleases. In 1956 Gierer and Schramm in Tübingen, Germany, and Fraenkel-Conrat in Berkeley, California, showed that if the viral RNA was obtained by degrading the particles with phenol or detergent, intact RNA molecules were obtained that were infectious and could

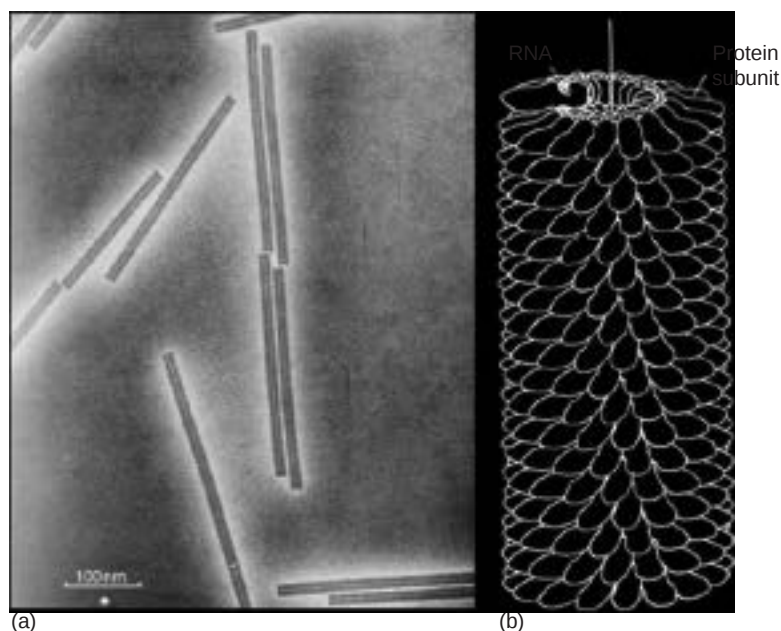


Fig. 1 (a) Electron micrograph of negatively stained TMV particles. (b) Diagram of a TMV particle showing about one-sixth of the length of a complete particle. The protein subunits form a helical array with $16\frac{1}{3}$ units per turn and the RNA is packed at a radius of about 4 nm from the helix axis. From Klug, A., 1999. The tobacco mosaic virus particle: Structure and assembly. *Philosophical Transactions of the Royal Society: Biological Sciences* 354, 531–535.

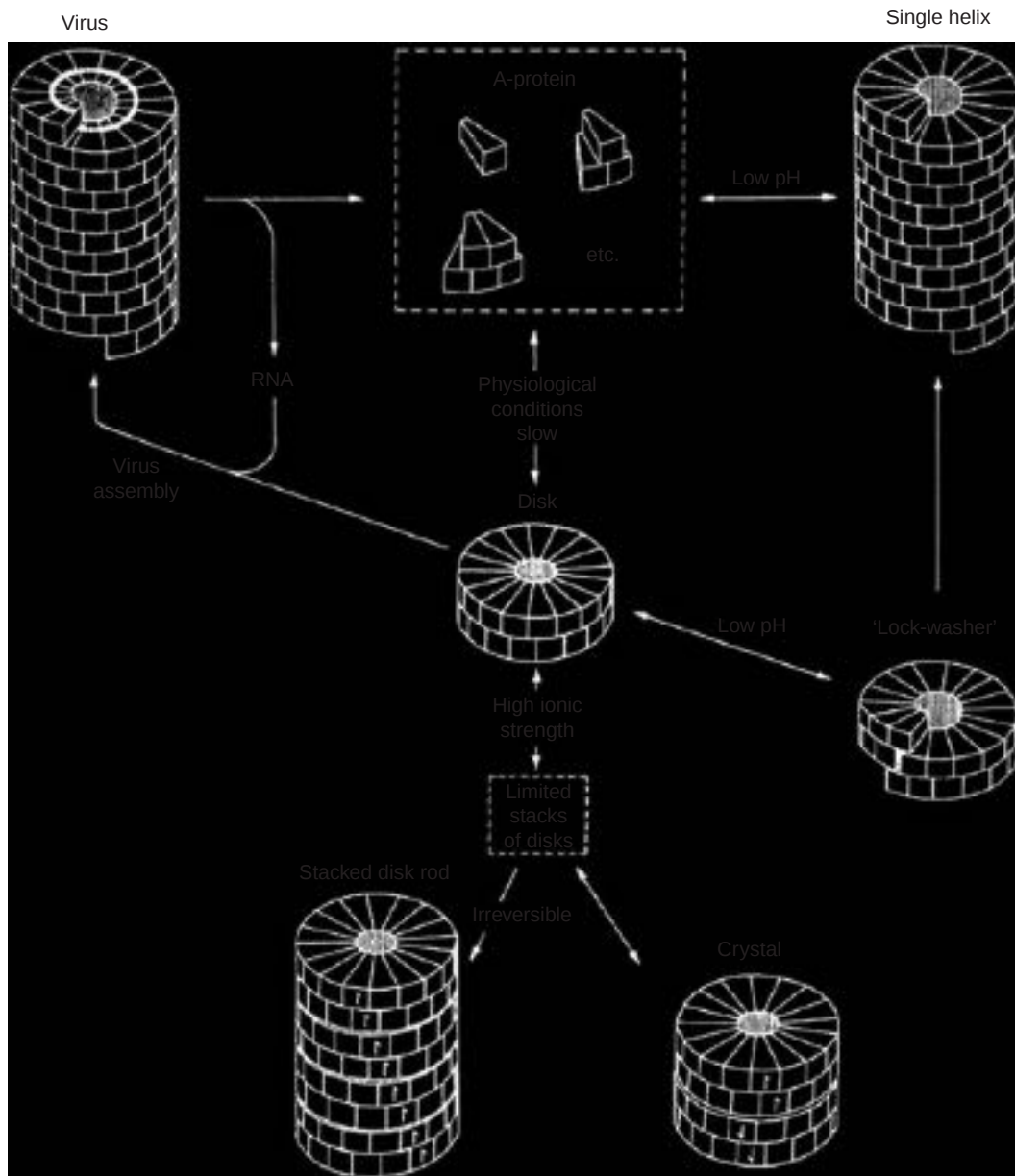


Fig. 2 Diagram showing the different polymorphic aggregates of TMV coat protein obtained at different pH values and ionic strengths. Note that both polar and nonpolar stacked disks can be obtained. The “lock-washer” is not well defined and represents a metastable transitory state. Adapted from Durham, A.C.H., Klug, A., 1972. Structures and roles of the polymorphic forms of tobacco mosaic virus protein. *Journal of Molecular Biology* 67, 315–332, with permission from Elsevier.

produce the same disease as intact virus. This was the first demonstration that the nucleic acid component of a virus was the carrier of viral infectivity and that it possessed the genetic information that coded for the viral coat protein.

In 1960 the TMV coat protein was the first viral protein to have its primary structure elucidated when its sequence was determined simultaneously in Tübingen and in Berkeley and found to consist of 158 amino acids. Subsequently, the coat protein sequence of numerous TMV mutants obtained by treating the virus with the mutagenic agent nitrous acid was also determined and these sequence data helped to establish the validity of the genetic code that was being elucidated in the early 1960s in Nirenberg's and Ochoa's laboratories. One of the changes induced by the action of nitrous acid on viral RNA is the nucleotide conversion of cytosine (C) to uracil (U). This changes the codon CCC for the amino acid proline to the codons UCC or CUC for serine and leucine, respectively. In addition, these two codons can be further converted to UUU which codes for phenylalanine. When the coat protein sequences of various TMV mutants were determined, it was found that most of the exchanged amino acids could be attributed to C→U nucleotide conversions, which confirmed the validity and universality of the proposed genetic code.

A particularly interesting mutation present in mutant Ni 1927 was the exchange proline to leucine at position 156 in the sequence which was found to greatly decrease the chemical stability of the virus. It had been known for many years that the virus was resistant to degradation by the enzyme carboxypeptidase and that this enzyme released only about 2000 threonine residues from each virus particle. The C-terminal residue in the coat protein is in fact threonine and the enzyme degradation data were interpreted at the time to mean that the virus particle contained 2000 subunits. However, when the proline at position 156 was replaced by leucine in the mutated protein, the enzyme was able to degrade the protein far beyond the C-terminal residue. The location of a proline near the C-terminus together with the presence of an acetyl group on the N-terminus are actually responsible for the remarkable resistance of the virus to exopeptidase degradation. The same exchange at position 156 in mutant Ni 1927 also led to an interesting discovery in immunology (see below).

Self-Assembly of TMV Particles

Dissociated coat protein molecules of TMV are able to assemble into different types of aggregates depending on the pH and ionic strength (Fig. 2). The disk aggregate is a two-layer cylindrical structure with a sedimentation coefficient of 20S, where each layer consists of a ring of 17 subunits compared with the 16(1/3) molecules present in each turn of the assembled helix. The disks are able to form stacks which can be either polar or nonpolar, depending on the relative orientation of adjacent disks (Fig. 2). Short stacks of disks were shown to be nonpolar by immunoelectron microscopy, using monoclonal antibodies that reacted with only one end of the viral helix but were able to bind to both ends of the stacked disks. Whereas short polar stacks of disks can be transformed via lock-washer intermediates into helices (Fig. 2), this is not possible for nonpolar stacks which cannot be incorporated as such during the elongation process that leads to the formation of RNA-containing virus particles.

It is generally accepted that initiation of TMV assembly from its RNA and protein components requires a 20S protein aggregate, which could be either a disk or a short, lock-washer type proto-helix. The surface of this aggregate constitutes a template which is recognized by a specific viral RNA sequence. The nucleation of the assembly reaction was found to occur by a rather complex process involving a hairpin structure on the viral RNA, known as the origin of assembly sequence (OAS) located about 1000 nt from the 3' end, which is inserted through the central hole of the 20 S aggregate. The nucleotide sequence of the stem-loop OAS hairpin is responsible for the preferential incorporation of viral RNA rather than foreign RNA. Elongation occurs by addition of more protein subunits which pull up more RNA through the central hole. This process requires that the 5' tail of the RNA loops back down the central hole and is slowly "swallowed up" while the 3' tail continues to protrude freely from the other end. This unexpected mechanism was visualized by electron micrographs obtained in Hirth's laboratory in Strasbourg which revealed how particles were growing with both 3' and 5' RNA tails protruding initially from the same end of the particle. The elongation is thus bidirectional and occurs faster along the longer 5' tail by incorporation of 20S aggregates and more slowly toward the 3' end through incorporation of 4S protein. The elucidation of the assembly process was greatly facilitated by the sequencing of the viral RNA which was completed in 1982. TMV RNA consists of 6395 nt and was the first genome of a plant virus to be sequenced completely.

Virus Disassembly

Since TMV particles are extremely stable *in vitro*, it was not at all obvious how the RNA managed to be released from the particles in order to start the virus replication cycle. Using a cell-free translation system, Wilson discovered that the disassembly of TMV particles is initiated when the end of the particles containing the 5' terminus of the RNA becomes associated with ribosomes. This leads to viral subunits becoming dislodged from the particles while the 5' terminal open reading frame in the RNA is being translated by the ribosomes, a process known as co-translational disassembly. This mechanism allows the coat protein subunits to protect the RNA from enzymatic degradation until the particle has reached a site in the infected cell where translation can be initiated.

Antigenicity of TMV

The antigenic properties of TMV have been studied extensively for more than 60 years and these studies have given us much information on how antibodies recognize proteins and viruses. TMV is an excellent immunogen and antibodies to the virus are readily obtained by immunization of experimental animals. When the sequence of TMV coat protein became available, it was possible for the first time to locate the antigenic sites or epitopes of a viral protein at the molecular level. Initial studies focused on two antigenic regions of the coat protein, the C-terminal region (residues 153–158) situated at the surface of virus particles and the disordered loop region (residues 103–112) located in the central hole of the particles which is accessible to antibodies only in dissociated protein subunits (Fig. 3). The C-terminal hexapeptide coupled to bovine serum albumin was used to raise antibodies and the resulting antiserum was found to precipitate the virus and neutralize its infectivity. Since both natural peptide fragments and synthetic peptides were used in this work, Anderer and his colleagues should be credited with the discovery that synthetic peptides can elicit antibodies that neutralize the infectivity of a virus. Only when similar results were obtained with animal viruses 15 years later, did the potential of peptides for developing synthetic vaccines become clear.

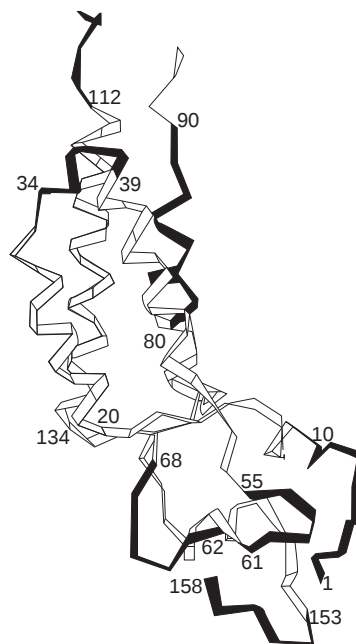


Fig. 3 Backbone of the TMV coat protein subunit based on crystallographic data. Residues 94–106 have been omitted because this region, located on the inside of the particle, is disordered. The N and C termini of the protein are located on the outer surface of TMV particles. The position of 7 continuous epitopes (residues 1–10, 34–39, 55–61, 62–68, 80–90, 108–112, and 153–158) is indicated by solid lines. Reproduced from Al Moudallal, Z., Briand, J.P., Van Regenmortel, M.H.V., 1985. A major part of the polypeptide chain of tobacco mosaic virus is antigenic. *EMBO Journal* 4, 1231–1235, with permission from Nature Publishing Group.

It has been known since the 1950s that intact TMV particles and dissociated coat protein subunits harbor different types of epitopes recognized by specific antibodies. Certain epitopes present only on virions are constituted by residues from neighboring subunits that are recognized as a single entity by certain antibodies; other epitopes arise from conformational changes in the protein that result from inter-subunit bonds. Both these types of epitopes which depend on the protein quaternary structure and are absent in dissociated protein subunits have been called neotopes. Another type of epitope known as cryptotope occurs on the portion of the protein surface that is buried in the polymerized rod and becomes accessible to antibodies only in the dissociated subunits.

The mapping of neotopes and cryptotopes on the surface of TMV coat protein was greatly simplified once monoclonal antibodies (Mabs) became available. Many continuous epitopes were identified at the surface of dissociated coat protein subunits by measuring the ability of peptides to react with Mabs or with antisera (Fig. 3). These epitopes were found to correspond to regions of the protein shown by X-ray crystallography to possess a high segmental mobility. This correlation between antigenicity and mobility along the peptide chain was also found to exist in other proteins and was used to develop algorithms for predicting the location of certain epitopes in proteins from their primary structure.

The surface of the protein subunits accessible at either end of the virus particle is different and the one located near the 5' terminus end of the RNA harbors the two helices corresponding to residues 73–89 and 115–135 (Fig. 3). Many Mabs specific for this surface have been obtained and in addition to reacting with both ends of nonpolar stacked disks, they were found to block TMV disassembly by sterically preventing the interaction between RNA and ribosomes. It has been suggested that if such antibodies could be expressed in plants, they might be able to control viral infection by preventing the disassembly of virions.

Another interesting immunological phenomenon was discovered when TMV antibodies were analyzed for their ability to react with certain TMV mutants. It was found that all rabbits immunized with TMV induced the formation of heterospecific antibodies, that is, antibodies that were unable to react with the TMV immunogen but recognized the mutant Ni 1927 quite well. When all the antibodies in a TMV antiserum capable of reacting with TMV were removed by cross-absorption with the virus, it was found that the depleted antiserum still reacted strongly with this mutant which had a single proline → leucine exchange. Apparently the removal of this proline at position 156 exposes binding sites for both peptidases and antibodies that are normally out of reach in the wild-type protein structure. The induction of heterospecific antibodies by immunization with TMV is another illustration of the difference between the antigenicity and immunogenicity of proteins. TMV particles possess the immunogenic capacity of eliciting heterospecific antibodies that react with the Ni 1927 mutant although they do not have the antigenic capacity of reacting with these antibodies. The reverse situation where an antigenic peptide or protein is able to react with a particular antibody but is unable to induce the same type of antibody when used as immunogen is a more commonly observed phenomenon which greatly hampers the development of synthetic peptide vaccines.

Taxonomy and Classification

Many of the viruses that are currently classified in the genus *Tobamovirus* were initially considered to be strains of TMV on the basis of similar particle morphology and ability to cross-react with TMV antibodies. Antigenic relationships between different tobamoviruses have been quantified using a parameter known as the serological differentiation index (SDI) which is the average number of twofold dilution steps separating homologous from heterologous antiserum titers. A close correlation exists between the antigenic distance of two viruses expressed as SDI values and the degree of sequence difference in their coat proteins. It is often accepted that when two viruses differ antigenically by an SDI value larger than 4, they are considered to belong to separate species. This is valid only if relatedness is measured with polyclonal antisera containing antibodies to a range of different epitopes since comparisons made with Mabs specific for a single epitope will emphasize antigenic similarities or differences depending on which region of the protein is recognized by the particular Mab that is used.

Relationships between individual tobamoviruses are nowadays usually assessed by comparisons between viral genomes. Phylogenetic studies have shown that tobamoviruses are very ancient and co-evolved with their angiosperm hosts, which means that they are at least 120 million years old. The *Tobamovirus* genus is one of seven recognized genera in the family *Virgaviridae* and it comprises 37 species.

Replication and Cell-to-Cell Movement of TMV

TMV-infected tissues contain four viral proteins: the 126 kDa and 183 kDa proteins of the replicase complex, the 30 kDa movement protein, and the 17.6 kDa coat protein. TMV replication is initiated by the translation of the viral RNA to produce the replicase proteins and this leads to the synthesis of minus-sense and plus-sense copies of the RNA. Translation of the viral coat protein gene then occurs leading to the assembly of progeny virus particles from full-length genomic RNA and coat protein. The translation of the coat protein and movement protein is controlled by the production of two separate subgenomic mRNAs, a mechanism later found to occur in many other viruses but first demonstrated with TMV.

The mechanism that allows plant viruses to move from cell to cell in their hosts has remained a mystery for many years. The rigid cellulose-rich walls of plant cells impede intercellular communication which occurs only through tubular connections known as plasmodesmata. However, plasmodesmata are too narrow to allow the passage of virus particles. Studies with the 30 kDa TMV movement protein showed that this protein changed the size exclusion limit of the plasmodesmata, allowing the virus to move through them in the form of a thin, less than 2 nm wide, ribonucleoprotein complex consisting of the genomic RNA and the movement protein. Movement proteins that possess RNA-binding properties have subsequently been found in many other plant viruses. Studies with TMV also revealed how plant viruses encode a suppressor to combat the post-transcriptional gene silencing reaction that plants use to fight virus infection.

Biotechnology Applications of TMV

TMV-resistant plants have been obtained by transforming them with a DNA copy of the TMV coat protein gene. This coat protein-mediated resistance is due to the ability of transgenically expressed coat protein to interfere in transgenic cells with the disassembly of TMV particles. Coat protein-mediated resistance has subsequently been obtained with many positive-sense RNA plant viruses.

A portion of the TMV RNA leader sequence, called omega, was shown to enhance the translation of foreign genes introduced into transgenic plants. This translational enhancer has been incorporated successfully in many gene vectors for a variety of applications.

In another type of application, TMV particles have been used as surface carriers of foreign peptide epitopes for constructing recombinant vaccines and producing them in plants. It was found that the N- and C-termini of the TMV coat protein as well as the surface loop area corresponding to residues 59–65 were able to accept foreign peptide fusions without impairing the ability of the resulting chimeric virus to infect plants systemically. Several experimental vaccines against viral and parasitic infections that are based on genetically engineered TMV particles produced in tobacco are currently under evaluation. In conclusion, a series of contingent discoveries made at the beginning of the 20th century focused on TMV as a new type of pathogen which was extremely stable and obtainable in large quantities from infected plants, and this allowed this virus to play a major role in the development of virology. Furthermore, studies of TMV also contributed significantly to the advancement of molecular biology and to our understanding of the physicochemical and antigenic properties of macromolecules.

Bionanotechnology Applications of TMV

Our considerable knowledge of the three-dimensional structure of TMV particles combined with our understanding of the mechanism of virion self-assembly gave rise to numerous applications in the new field of bionanotechnology. This has led, for instance, to the development of efficient vectors for virus-induced gene silencing in plants, the use of virus-derived sequences for the creation of virus-resistant lines of plants and to the fusion of peptides and various metals on the exterior surface of the virus

for creating a variety of TMV -derived nanostructures. It is also possible to encapsidate any RNA molecule provided it contains the OAS of TMV which then leads to the protection of labile RNA against degradation.

TMV -like particles can also be coupled to various organic molecules, including chromophores or polymers, in order to create classes of effective colloids that have potential biomedical applications inside living organisms. TMV-based nanorods of various sizes have been fabricated *in vitro* and some of them have been administered via the bloodstream for developing tumor-targeting agents. It is quite remarkable that a simple viral structure consisting of a single type of protein subunit together with a single molecule of RNA has led to such an enormous number of possible biotechnological applications.

There is of course always another side to a coin, which in this case is the finding that the stability of TMV particles allows the virus to retain its infectivity for several years during the manufacturing process of cigarettes and cigars. TMV RNA has been shown to be present in the saliva of cigarette smokers (but not of non-smokers) and these smokers were also found to have a higher level of anti-TMV IgG antibodies in their blood than non-smokers. TMV antibodies were also found to cross react with the human mitochondrial protein TOMM40L which shares a six residue sequence (FGTGGA) with residues 36–41 of the TMV coat protein which is also a continuous epitope of the viral protein. It cannot be excluded that these antibodies resulting from an antigenic mimicry phenomenon between TOMM40L and TMV may play role in mitochondrial dysfunction implicated in the pathogenesis of Parkinson disease, Alzheimer disease and other degenerative disorders (Liu *et al.*, 2013).

See also: Tobamoviruses (*Virgaviridae*). Virgaviruses (*Virgaviridae*)

Further Reading

- Al Moudallal, Z., Briand, J.P., Van Regenmortel, M.H.V., 1985. A major part of the polypeptide chain of tobacco mosaic virus is antigenic. *EMBO Journal* 4, 1231–1235.
- Balique, F., Colson, P., Raoult, D., 2012. Tobacco mosaic virus in cigarettes and saliva of smokers. *Journal of Clinical Virology* 55, 374.
- Bos, L., 1999. Beijerinck's work on tobacco mosaic virus: Historical context and legacy. *Philosophical Transactions of the Royal Society, London, Series B* 534, 675–685.
- Calisher, C.H., Horzinek, M.C., 1999. 100 Years of Virology. The Birth and Growth of a Discipline. Vienna: Springer, pp. 1-220.
- Durham, A.C.H., Klug, A., 1972. Structures and roles of the polymorphic forms of tobacco mosaic virus protein. *Journal of Molecular Biology* 67, 315–332.
- Harrison, B.D., Wilson, T.M.A., 1999. Tobacco mosaic virus: Pioneering research for a century. *Philosophical Transactions of the Royal Society, London, Series B* 354, 517–685.
- Hirth, L., Richards, K.E., 1981. Tobacco mosaic virus: Model for structure and function of a simple virus. *Advances in Virus Research* 26, 145–199.
- Liu, R., Vaishnav, R.A., Roberts, A.M., Friedland, R.P., 2013. Humans have antibodies against a plant virus: Evidence from tobacco mosaic virus. *PLoS One* 8 (4), e60621. doi:10.1371/journal.pone.0060621.
- Lomonosoff, G.P., 2018. So what have plant viruses ever done for virology and molecular biology? *Advances in Virus Research* 100, 145–162.
- Lomonosoff, G.P., Wege, C., 2018. TMV particles: The journey from fundamental studies to bionanotechnology applications. *Advances in Virus Research* 102, 149–176.
- Mahy, B.W.J., Lvov, D.K., 1993. Concepts in Virology: From Ivanovsky to the Present. Langhorne, PA: Harwood Academic Publishers.
- Scholthof, K.-B.G., 2004. Tobacco mosaic virus: A model system for plant biology. *Annual Review of Phytopathology* 42, 13–34.
- Scholthof, K.-B.G., Shaw, J.G., Zaitlin, M., 1999. Tobacco Mosaic Virus: One Hundred Years of Contributions to Virology. St. Paul, MN: American Phytopathological Society Press, pp. 1–256.
- The life of a virus. In: Creager, A.N.H. (Ed.), Tobacco Mosaic Virus as an Experimental Model. Chicago, IL: University of Chicago Press.
- Van Helvoort, T., 1991. What is a virus? The case of tobacco mosaic disease. *Studies in History and Philosophy of Science* 22, 557–588.
- Van Regenmortel, M.H.V., 1999. The antigenicity of TMV. *Philosophical Transactions of the Royal Society, London, Series B* 354, 559–568.
- Van Regenmortel, M.H.V., Fraenkel-Conrat, H., 1986. The Plant Viruses: The Rod-Shaped Plant Viruses. vol. 2. New York, NY: Plenum, pp. 1–180.

Tobamoviruses (*Virgaviridae*)

Ulrich Melcher, Oklahoma State University, Stillwater, OK, United States

Dennis J Lewandowski, University of Florida, Lake Alfred, FL, United States

William O Dawson, Citrus Research and Education Center, Lake Alfred, FL, United States and University of Florida, Lake Alfred, FL, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of D.J. Lewandowski, Tobamovirus, In Encyclopedia of Virology (Third Edition), Edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00514-8.

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

ELISA Enzyme-linked immunological assays

ER Endoplasmic reticulum

kb Kilobase

kDa Kilo dalton

LAMP Loop mediated amplification

MET Methyl-transferase

miRNA MicroRNA

MP Movement protein

mRNA Messenger RNA

nt Nucleotide(s)

OAS Origin of assembly

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

RT-PCR Reverse transcription-polymerase chain reaction

satRNA Satellite RNA

sgRNA Sub-genomic RNA

siRNAs Small interfering RNAs

TILLING Targeting induced local lesions in genomes

tRNA Transfer RNA

UTR Un-translated region

Glossary

Origin of assembly Stem-loop structure that is the site of initiation of virion assembly.

Protohelix Contain approximately 40 CP TMV subunits arranged in a spiral around a central hollow core, similar to the arrangement within the virion.

Pseudoknot RNA structure with base pairing between a loop and other regions of the RNA.

History of Tobamovirus Research

Research in the late 1800s on the causal agent of the mosaic disease of tobacco led to the discovery of viruses as new infectious agents. Thus *Tobacco mosaic virus* (TMV), the type species of the genus *Tobamovirus*, became the first virus to be discovered, and subsequently has had a significant role in many fundamental discoveries in virology. The first quantitative biological assay for plant viruses was the use of *Nicotiana glutinosa* plants, which produce necrotic local lesions when inoculated with TMV or other tobamoviruses. The resistance gene, N, that confers this hypersensitive response-type resistance to TMV was the first resistance gene against a plant virus to be cloned and characterized. TMV was also the first virus to be purified and crystallized, which led to the discovery of the nucleoprotein nature of viruses and determination of the atomic structure of the coat protein (CP) and the virion. TMV was the first virus to be visualized in the electron microscope, confirming the predicted rigid rod-shaped virions. The genetic material of TMV was shown to be RNA, a property previously thought to be restricted to DNA. The first viral protein for which an aa sequence was determined was the TMV CP. TMV was also the first virus to be mutagenized. The subsequent determination of the CP sequences from a number of strains and mutants helped establish the universality of the genetic code. Methods of infecting plant protoplasts with viruses were developed with the tobacco-TMV system, creating a synchronous system to study events in the infection cycle. The TMV 30 kDa protein was the first viral protein shown to be required for virus movement. Finally, TMV was the first plant virus genome to be completely sequenced in 1975.

Taxonomy and Classification

The genus *Tobamovirus* has been assigned to the family *Virgaviridae*. Currently, there are 38 recognized species within the genus (Table 1). Several recently sequenced viruses are tentative members of new species in the genus (Table 1). Although tobamoviruses comprise one of the more intensively studied plant virus genera, taxonomy has often been confusing. Historically, plant viruses with rigid virions of approximately 18 × 300 nm and causing various diseases were all designated strains of TMV. Thus, many viruses originally referred to

Table 1 Definitive and tentative species members of the genus *Tobamovirus*, family *Virgaviridae*

Species name	Acronym	Accession #	Length (nt)	Status	Natural host family
<i>Bell pepper mottle virus</i>	BPMoV	NC009642	6375	Complete	Solanaceae
<i>Brugmansia mild mottle virus</i>	BrMMoV	NC010944	6381	Complete	Solanaceae
<i>Cactus mild mottle virus</i>	CMMoV	NC011803	6449	Complete	Cactaceae
<i>Clitoria yellow mottle virus</i>	CIYMoV	NC016519	6514	Complete	Fabaceae
<i>Cucumber fruit mottle mosaic virus</i>	CGMoMV	NC002633	6562	Complete	Cucurbitaceae
<i>Cucumber green mottle mosaic virus</i>	CFMoMV	NC001801	6424	Complete	Cucurbitaceae
<i>Cucumber mottle virus</i>	CuMoV	NC008614	6485	Complete	Cucurbitaceae
<i>Frangipani mosaic virus</i>	FrMV	NC014546	6643	Complete	Apocynaceae
<i>Hibiscus latent Fort Pierce virus</i>	HLFPV	NC025381	6431	Complete	Malvaceae
<i>Hibiscus latent Singapore virus</i>	HLSGV	NC008310	6485	Complete	Malvaceae
<i>Kyuri green mottle mosaic virus</i>	KGMoMV	NC003610	6514	Complete	Cucurbitaceae
<i>Maracuja mosaic virus</i>	MaMV	NC008716	6794	Complete	Passifloraceae
<i>Obuda pepper virus</i>	ObPV	NC003852	6507	Complete	Solanaceae
<i>Odontoglossum ringspot virus</i>	ORSV	NC001728	6618	Complete	Orchidaceae
<i>Opuntia virus 2</i>	OV2	NC040685	6453	Complete	Solanaceae
<i>Paprika mild mottle virus</i>	PaMMoV	NC004106	6524	Complete	Solanaceae
<i>Passion fruit mosaic virus</i>	PfMV	NC015552	6791	Complete	Passifloraceae
<i>Pepper mild mottle virus</i>	PeMMoV	NC003630	6357	Complete	Solanaceae
<i>Plumeria mosaic virus</i>	PIMV	NC026816	6688	Complete	Apocynaceae
<i>Rattail cactus necrosis-associated virus</i>	RCNaV	NC016442	6506	Complete	Cactaceae
<i>Rehmannia mosaic virus</i>	ReMV	NC009041	6395	Complete	Orobanchaceae
<i>Ribgrass mosaic virus</i>	RMV	NC002792	6311	Complete	Plantaginaceae
<i>Streptocarpus flower break virus</i>	SFBV	NC008365	6279	Complete	Gesneriaceae
<i>Sunn-hemp mosaic virus</i>	SHMV	NC043383	4683	Incomplete	Leguminosae
<i>Tobacco latent virus</i>	TLV	NC038703	1415	Incomplete	Solanaceae
<i>Tobacco mild green mosaic virus</i>	TMGMV	NC001556	6355	Complete	Solanaceae
<i>Tobacco mosaic virus</i>	TMV	NC001367	6395	Complete	Solanaceae
<i>Tomato brown rugose fruit virus</i>	ToBRFV	NC028478	6393	Complete	Solanaceae
<i>Tomato mosaic virus</i>	ToMV	NC002692	6383	Complete	Solanaceae
<i>Tomato mottle mosaic virus</i>	ToMoMV	NC022230	6398	Complete	Solanaceae
<i>Tropical soda apple mosaic virus</i>	TSAMV	NC030229	6350	Complete	Solanaceae
<i>Turnip vein-clearing virus</i>	TuVCV	NC001873	6311	Complete	Brassicaceae
<i>Ullucus mild mottle virus</i>	UMMoV	ND			Brassicaceae
<i>Wasabi mottle virus</i>	WMoV	NC003355	6298	Complete	Brassicaceae
<i>Yellow tailflower mild mottle virus</i>	YTMMoV	NC022801	6379	Complete	Taxaceae
<i>Youcai mosaic virus</i>	YMoV	NC004422	6303	Complete	Brassicaceae
<i>Zucchini green mottle mosaic virus</i>	ZGMoV	NC003878	6513	Complete	Cucurbitaceae
Tentative species					
<i>Chara corallina Canada virus</i>	CCCAV	MK521928	9593	Complete	Characeae
<i>Chara corallina Australia virus</i>	CCAUV	AEJ33768	9065	complete	Characeae
<i>Hoya chlorotic spot virus</i>	HCSV	NC034509	6386	complete	Apocynaceae

as TMV are now recognized as belonging to separate species. For example, the tobamovirus that was referred to as the tomato strain of TMV, and is approximately 80% identical to TMV at the nt sequence level, is now recognized as a different species, *Tomato mosaic virus* (ToMV). One major criterion for distinguishing the members of separate tobamovirus species is a nt sequence difference of at least 10%.

Evolution

Both nt and aa sequences have been used to suggest evolutionary relationships of these viruses. Phylogenetic trees of the MP, CP, and RdRp replicase ORFs have similar topologies, indicating that this block has remained more or less constant as the viruses evolved. Additionally, the trees appear to identify clusters of viruses based on the family of the hosts from which they were identified. Comparison of phylogenetic trees for different coding regions of the genome suggest the clustering of tobamoviruses based on the host family. For example, viruses from solanaceous plants separate from those of cruciferous plant species in all coding regions. Similar congruence of phylogenetic positions is also seen for viruses whose principal hosts are in the *Cucurbitaceae* family. The observations suggest that the viruses were present in plant ancestors to both plant families. The observation that an orchid virus, ORSV, clustered with viruses of solanaceous plants in all but the methyl-transferase region of the replicase, where it branches with viruses from cruciferous plants, challenges that view. Such placement is suggestive that RNA–RNA recombination played a role in evolution of some tobamoviruses (Fig. 1).

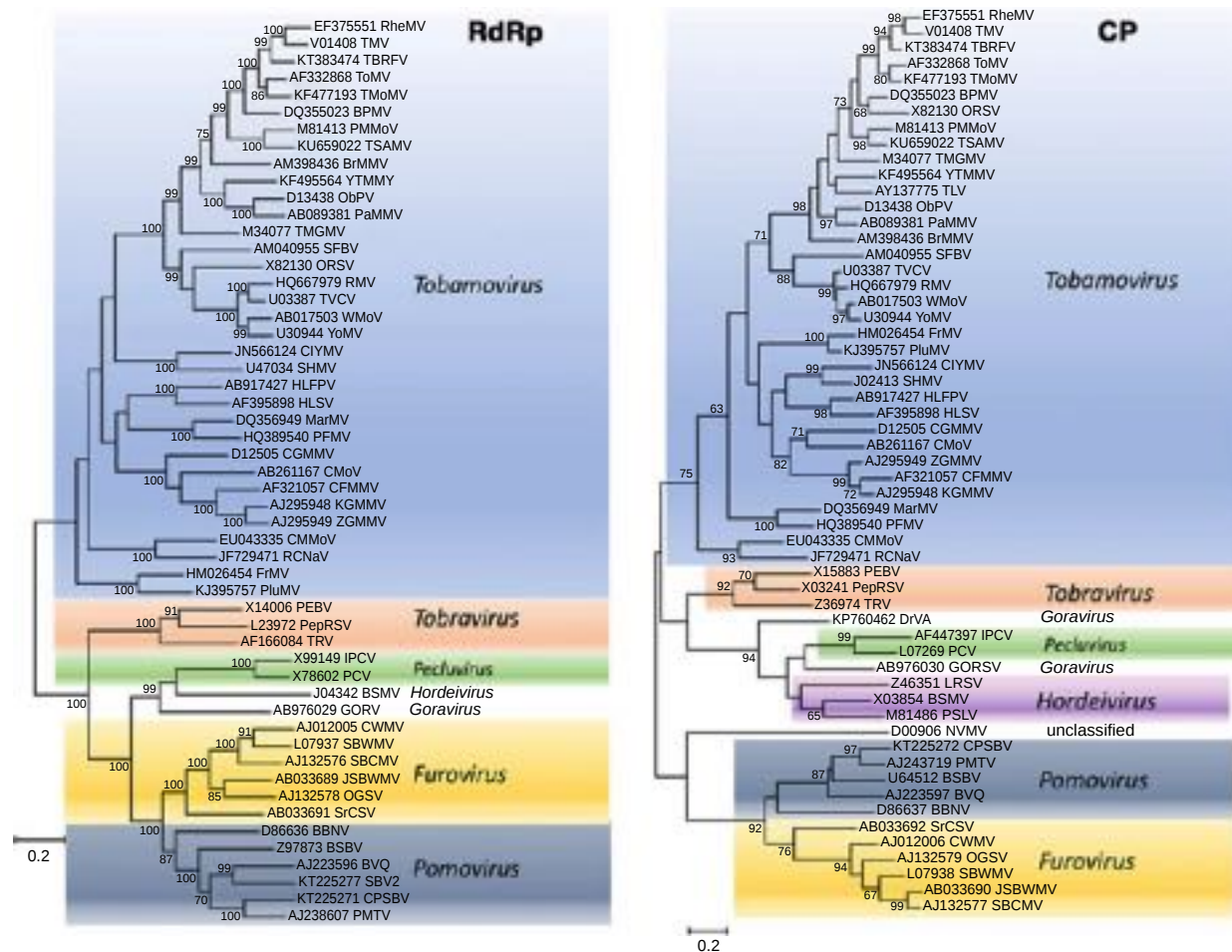


Fig. 1 Phylogenetic trees of tobamoviruses based on RdRp (left) and CP (right) gene sequences. The analyses were conducted in MEGA7 using the Maximum Likelihood method based on the Tamura-Nei model and 1000 bootstrap replicates. The tree with the highest log likelihood is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches (where > 60%). The scale indicates the number of substitutions per site. All positions containing gaps and missing data were eliminated. Virus abbreviations are explained in the tables of member species and other viruses. Reproduced from: https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-ssRNA-viruses/w/virgaviridae.

The tobamovirus members are divided into those having the origin of assembly in the MP coding region (subgroup 1) and those for which the origin of assembly is within the CP coding region (subgroup 2). This division is strongly supported by phylogenetic analysis that shows the tree with all available sequences to consist of two major branches, one corresponding exactly to subgroup 1 and the other to subgroup 2 (Fig. 1).

In these phylogenetic analyses, the closest neighbors that were not designated members of the tobamoviruses were a group of viruses typified by *Beet necrotic yellow vein virus*, thus defining the closest family of the *Tombusviridae* as the *Benyvirus* genus.

The assembly of evolutionary relationships in phylogenetic trees allows some thoughts on the evolution of virus genera. The tobamoviruses have sequence neighbors in other genera of ssRNA viruses. They include genera *Benyvirus* (as mentioned above and consisting of plant viruses) and *Orthohepevirus* (which are animal viruses). Most closely related to the genus *Tobamovirus* are the charaviruses which have been isolated from the algal genus *Chara* (Fig. 2).

Virus Structure and Composition

Tobamovirus virions are straight particles of approximately 18×300 nm with a central hollow core 4 nm in diameter (Fig. 3(A) and (B)). Virion composition is approximately 95% protein and 5% RNA. For TMV, approximately 2100 subunits of a single CP are arranged in a right-handed helix around a single genomic RNA molecule, with each subunit associated with three adjacent nucleotides. Protein-protein associations are the essential first event of virion assembly. Coat protein subunits assemble into several types of aggregates. Coat protein monomers and small heterogeneous aggregates of a few subunits are collectively referred to as 'A-protein'. The equilibrium between A-protein and larger aggregates is primarily dependent upon pH and ionic strength. Purified CP and viral RNA can assemble into

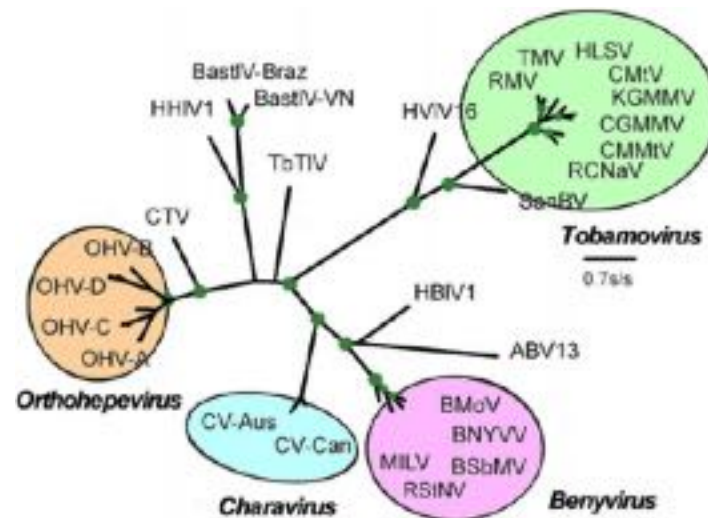


Fig. 2 Maximum-likelihood phylogeny of the amino acid sequences of the RdRp-2 regions of the replicase proteins of the charaviruses, benyviruses and selected tobamoviruses and relatives. Acronyms: ABV13, agaricus bisporus virus 13 (AQM49942); BastV-VN, bastrovirus-like virus-VietNam Bat (YP009333174); BastV-Braz, bastrovirus Brazil/sewage (ASM79505); BMoV, burdock mottle virus (YP008219063); BNYVV, beet necrotic yellow vein virus (NP612615); BSbMV, beet soil borne mosaic virus (NP612601); CGMMV, cucumber green mottle mosaic virus (NP044577); CMMV, cactus mild mottle virus (YP002455590); CTV, cutthroat trout piscihevirus (YP004464917); CuMtV, cucumber mottle virus (YP908760); CV-Aus, charavirus australis (AEJ33768); CV-Can, charavirus canadensis (MK521928); HBIV1, hubei Beny-like virus 1 (APG77690); HHV1, hubei hepe-like virus 1 (YP009336840); HLSV, hibiscus latent Singapore virus 178 (YP719997); HvIV16, hubei virga-like virus 16 (YP009336677); KGMMV, kyuri green mottle mosaic 179 virus (YP908760); MILV, mangifera indica latent virus (AMQ23297); OHV-A, orthohepevirus A (ABB88699, AGE83293, AGE83340, AGT38396, ANW09725, BAE86910); OHV-B, orthohepevirus B (AEX93357, CAQ16023, YP_009001465; OHV-C, orthohepevirus C (ADB96199, AFL69932, ANJ02843, BAO47898, BAT70058); OHV-D, orthohepevirus D (AIF74285, YP006576507); RCNaV, rattail cactus necrosis-associated virus (YP0044936166); RMV, ribgrass mosaic virus (YP005476600); RStNV, rice stripe necrosis virus (ABU94739); SanBV, san bernardo virus (AQM55436); TbTIV, tick borne tetravirus-like virus (AII01815); TMV, tobacco mosaic virus (NP597746). The green disks mark nodes with >0.9 SH support; two thirds of the nodes in the Orthohepevirus cluster have >0.9 SH support but, for clarity, are not marked.

infectious particles *in vitro*. Larger aggregates are disks composed of two individual stacked rings of CP subunits, and protohelicies. Protohelicies contain approximately 40 CP subunits arranged in a spiral around a central hollow core, similar to the arrangement within the virion. A sequence-specific stem-loop structure in the RNA, the origin of assembly (OAS), initiates encapsidation and prevents defective packaging that could result from multiple independent initiation events on a single RNA molecule. Virion assembly initiates as the primary loop of the OAS is threaded through a CP disk or protohelix with both ends of the RNA trailing from one side. The conformation of the CP protohelix changes as the RNA becomes embedded within the groove between the two layers of subunits. Elongation is bidirectional, proceeding rapidly toward the 5' end of the RNA as the RNA loop is extruded through the elongating virion and additional CP disks are added. There is disagreement about the mechanism of elongation toward the 3' terminus of the RNA, but it appears that this slower process involves the addition of smaller protein aggregates (Fig. 3(C)).

Sub-genomic mRNAs containing the OAS are encapsidated into shorter particles that are not required for infectivity. The OAS is located within the open reading frame (ORF) for the movement protein (MP) of most tobamoviruses. The level of accumulation of a particular sub-genomic mRNA containing the OAS determines the relative proportion of that particular virion species. Thus, all tobamovirus virion populations contain a small percentage of MP sub-genomic mRNAs (sgRNA). In some tobamoviruses, including Cucumber green mottle mosaic virus (CGMMV), Hibiscus latent Singapore virus (HLSV), Kyuri green mottle mosaic virus (KGMMV), Maracuja mosaic virus (MaMV), Sunn-hemp mosaic virus (SHMV), Zucchini green mottle mosaic virus (ZMoMV), and Cactus mild mottle virus (CMMoV), the OAS is located within the CP ORF. Thus, these so-called subgroup 2 tobamoviruses produce a significant proportion of small virions that contain CP sgRNA. Hybrid non-viral RNAs containing an OAS will also assemble with CP into virus-like particles of length proportional to that of the RNA.

Genome Organization

The genomes of tobamoviruses consist of one single-stranded positive-sense 6.3–6.8 kb RNA (Fig. 4). A methyl-guanosine cap at the 5' terminus, precedes a 55–75 nt AU-rich leader. The 3' end of the RNA can be folded into a series of pseudoknot structures including a tRNA-like terminus. The hibiscus-infecting tobamoviruses Hibiscus Latent Singapore virus (HLSV) and Hibiscus latent Fort Pierce virus (HLFPV) contain, in addition, a poly(A) stretch between the 3' end of the CP ORF and the tRNA-like structure. The tRNA-like termini can be amino-acylated *in vitro*, in most cases specifically accepting histidine. The 3' terminus of SHMV is an

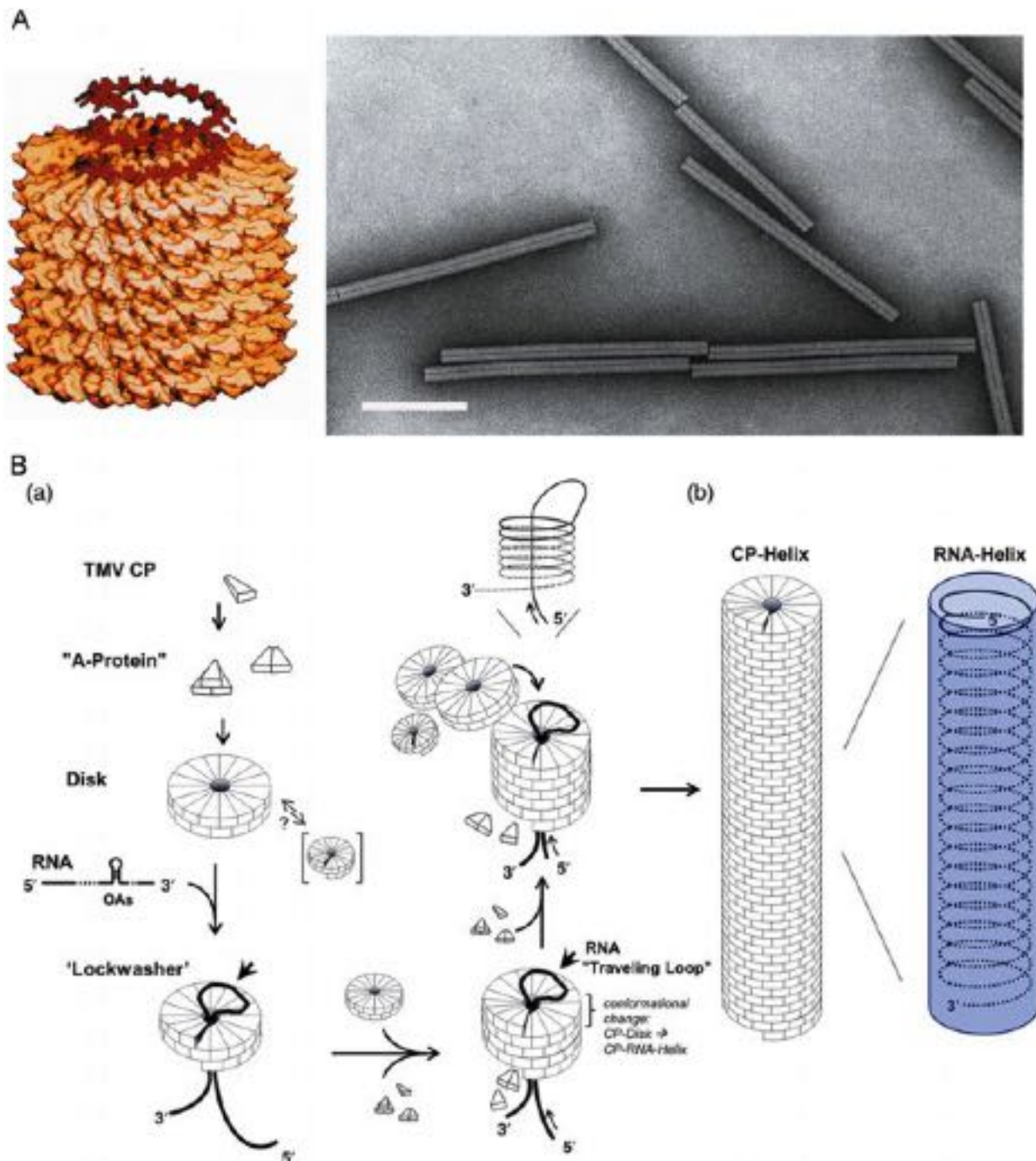


Fig. 3 Structure of the Tobacco mosaic virus (*Tobamovirus*, *Virgaviridae*). (A) Model of three-dimensional structure of a hollow tube that is assembled from TMV CP subunits. (B) Electron micrograph of purified TMV virions showing tubular structures. (C) Diagram of proposed steps in the assembly of TMV virions from CP subunits and RNA. Scheme of the assembly stages of TMV nanoparticles from OAs-containing RNA and CP subunits. Refer to the text for a thorough description. (a) Shows the successive assembly intermediates and the overall movement of the RNA scaffold upon its encapsidation. (b) Visualizes the organization of the final particle and its representation as applied in subsequent figures, with the RNA helix sandwiched between the helically arranged CP subunits. Images are not to scale. CP, coat protein; OAs, origin of assembly; TMV, tobacco mosaic virus. (A) Courtesy of J.Y. Sgro, VirusWorld, Institute for Molecular Virology, University of Wisconsin-Madison, USA. (B) Photograph of J.T. Finch reproduced from Fauquet, C.M., *et al.*, 2005. VIIIth ICTV Report, Academic Press. (C) Adapted according to the Creative Commons Attribution 2.0 International Public License from Koch, C., Eber, F.J., Azucena, C., *et al.*, 2016. *Beilstein J. Nanotechnol.* 7, 613–629. doi:10.3762/bjnano.7.54.

exception, accepting valine and appears to have arisen by a recombination event between a tobamovirus and a tymovirus. RNA extracted from infected plant material can be used to make transcription plasmids, whose transcription will produce transcribable RNA molecules used to allow production of specific proteins in recipient plants.

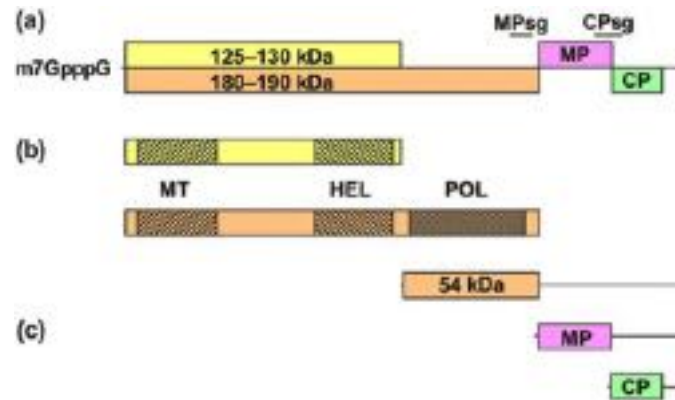


Fig. 4 Tobamovirus genome organization and gene expression strategy. (a) Tobamovirus genome organization. ORFs designated as open boxes. Non-translated sequences designated as lines; positions of sub-genomic promoters are marked. (b) Non-structural proteins involved in tobamovirus replication. Functional domains shared with other viruses within the 'alphavirus supergroup' are designated as hatched boxes. (c) Sub-genomic mRNAs with 5' proximal ORF labeled. MP, movement protein; CP, coat protein; MT, methyl-transferase; HEL, helicase; POL, polymerase; MPsg, MP sub-genomic mRNA; CPsg, CP sub-genomic mRNA.

The four ORFs that are contained within all tobamovirus genomes (Fig. 4) correspond to the proteins found in infected tissue. Two overlapping ORFs begin at the 5' proximal start codon. Termination at the first in-frame stop codon produces a 125–130 kDa protein. A 180–190 kDa protein is produced by readthrough of this leaky termination codon approximately 5–10% of the time. The remaining proteins are expressed from individual 3' co-terminal sub-genomic mRNAs, from which only the 5' proximal ORF is expressed (Fig. 4). The next ORF encodes the 28–34 kDa movement protein, which has RNA-binding activity and is required for cell-to-cell movement. The 3' proximal ORF encodes a 17–18 kDa CP. A sub-genomic mRNA containing an ORF for a 54 kDa protein that encompasses the readthrough domain of the 180–190 kDa ORF has been isolated from infected tissue, although no protein has been detected.

Within the protein-coding regions of the genome, there are nt sequences that also function as cis-acting elements for sub-genomic mRNA synthesis, virion assembly, and replication. Gene expression from sub-genomic mRNAs is regulated both temporally and quantitatively. The MP is produced early and accumulates to low levels, whereas the CP is produced late and accumulates to high levels. The regulatory elements for sub-genomic mRNA synthesis are located on the genome-length complementary RNA overlapping the upstream ORF (Fig. 4). There is limited (40%) sequence identity between the TMV MP and CP sub-genomic promoters. The TMV MP sub-genomic promoter is located upstream of the MP ORF, flanking the transcription initiation site. Unlike the MP sub-genomic promoter, full activity of the CP sub-genomic promoter requires sequences within the CP ORF.

Viral Proteins

The tobamovirus 125–130/180–190 kDa proteins are involved in viral replication, gene expression, and movement. Both are contained in crude replicase preparations, and temperature-sensitive replication-deficient mutants map to these ORFs. The 125–130/180–190 kDa proteins contain two functional domains common to replicase proteins of many positive-stranded RNA plant and animal viruses (Fig. 4). The N-terminal domain has methyl-transferase and guanylyl-transferase activities associated with capping of viral RNA. The second common domain is a proposed helicase, based upon conserved sequence motifs. The readthrough domain of the 180–190 kDa protein has sequence motifs characteristic of RNA-dependent RNA polymerases (RdRp). Both proteins are necessary for efficient replication, although the TMV 126 kDa protein is dispensable for replication and gene expression in protoplasts. The 125–130 kDa protein (or sequences within this region) of the 180–190 kDa protein are required for cell-to-cell movement. Additionally, these multifunctional proteins are symptom determinants, as mutations in mild strains map to these ORFs.

The 28–34 kDa MP has a plasmodesmatal binding function associated with its C-terminus and a single-stranded nucleic acid-binding domain associated with the N-terminus. The MP–host interaction determines whether the virus can systemically infect some plant species. Although principally a structural protein, the CP is also involved in other host interactions. Coat protein is required for efficient long-distance movement of the virus. Coat protein is also a symptom determinant in some susceptible plant species and an elicitor of plant defense mechanisms in other plant species.

Interactions Between Viral and Host Proteins

Available evidence suggests that the interactions of viral proteins with host factors are important determinants of viral movement and host ranges. Amino acid substitutions in the movement protein and 125–130/180–190 kDa proteins can alter the movement function in different hosts. Some viruses, including tobamoviruses, can assist movement of other viruses that are incapable of

movement in a particular plant species. These interactions suggest that there are more precise associations of viral proteins with host factor(s) than with viral RNA. Additionally, precise CP–plant interactions are required for movement to distal positions within the plant. The helicase domain of the 130–190 kDa proteins elicits the N gene-mediated resistance in *N. glutinosa*.

Virus Replication

Virions or free viral RNA will infect plants or protoplasts. Because tobamoviruses have a genome consisting of messenger-sense RNA that is infectious, one of the first events is translation of the 5′ proximal ORFs to produce the proteins required for replication of the genomic RNA and transcription of sgRNAs. When virions are the infecting agent, the first event is thought to be co-translational disassembly, in which the CP subunits at the end of the virion surrounding the 5′ end of the RNA loosen, making the RNA available for translation. Ribosomes then associate with the RNA, and translation of the 126/183 kDa ORFs is thought to displace CP subunits from the viral RNA. After the formation of an active replicase complex, a complementary negative-strand RNA is synthesized from the genomic positive-strand RNA template. Negative-strand RNA serves as template for both genomic and sgRNAs. Negative-strand RNA synthesis ceases early in infection, while positive-strand RNA synthesis continues. This results in an asymmetric positive- to negative-strand RNA ratio. Early in infection, genomic RNA functions as template for negative-strand RNA synthesis and as mRNA for production of the 126/183 kDa proteins. Later in the infection cycle, most of the newly synthesized genomic RNA is encapsidated into virions. Sub-genomic mRNAs transcribed during infection function as mRNA for the 3′ ORFs. Within cells of an infected leaf, replication proceeds rapidly between approximately 16 and 96 h post infection within a cell, then ceases. Even though the infected cells become packed with virions, these cells remain metabolically active for long periods. During the early stages of infection of an individual cell, the infection spreads through plasmodesmatal connections to adjacent cells. This event requires the viral MP that modifies plasmodesmata to accommodate larger molecules and the 126/183 kDa proteins. Movement through plasmodesmata does not require the CP. A second function of the MP appears to be binding to the viral RNA to assist its movement through the small plasmodesmatal openings. The MP also appears to associate with the cytoskeleton. As the virus spreads from cell to cell throughout a leaf, it enters the phloem for rapid long-distance movement to other leaves and organs of the plant, a complex process that requires the CP.

The Cis-Acting Sequences

The 5′ non-translated region contains sequences that are required for replication. This region is an efficient translational leader. The 3′ non-translated region contains cis-acting sequences that are involved in replication. Certain deletions within the pseudoknots are not lethal, but result in reduced levels of replication. Exchange of 3′ non-translated elements between cloned tobamovirus species has resulted in some lethal and nonlethal hybrids, suggesting a requirement for sequence specificity and/or secondary structure. The 3′ non-translated region appears to be a translational enhancer, both in the viral genome and when fused to heterologous reporter mRNAs. Sequences encoding the internal ORFs for the MP and CP are dispensable for replication. Duplication of the sub-genomic promoters results in transcription of an additional new sub-genomic mRNA. Heterologous tobamovirus sub-genomic promoters inserted into the viral genome are recognized by the replicase complex and transcribed. Foreign sequences inserted behind tobamovirus sub-genomic mRNA promoters have been expressed to high levels in plants and protoplasts. Placement of the RNA sequence required for initiation of RNA encapsidation determines the size of the resulting virion accounting for the major differences in particle sizes within the tobamoviruses.

Satellite Tobacco Mosaic Virus

Satellite tobacco mosaic virus (STMV), a tobamovirus-dependent satellite virus, was isolated from *Nicotiana glauca* plants infected with Tobacco mild green mosaic virus (TMGMV). The STMV genome consists of one single-stranded positive-sense RNA of 1059 nt. The 240 3′ nt share approximately 65% sequence identity with TMGMV and TMV, contain pseudoknot structures, and have a tRNA-like terminus. No sequence similarity with any tobamovirus exists over the remainder of the genome. Two overlapping ORFs that are expressed in *in vitro* translation reactions are present in the genomic RNA of most STMV isolates. The 5′ proximal ORF encodes a 6.8 kDa protein that has not been detected *in vivo* and is not present in all STMV isolates. The second ORF encodes a 17.5 kDa CP that is not serologically related to any tobamovirus CP. The 17 nm icosahedral virions are composed of a single STMV genomic RNA encapsidated within 60 STMV CP subunits. Replication of natural populations of STMV is supported by other tobamoviruses, but at lower levels than with the natural helper virus, TMGMV. The host range of STMV parallels that of the helper virus.

Host Range and Symptomology

Tobamoviruses have a fairly large host range. All-in-all, dicotyledonous and monocotyledonous have been identified as hosts for the members of the 36 tobamovirus species and the 3 tentative members. TMV alone is known to infect members of nine plant

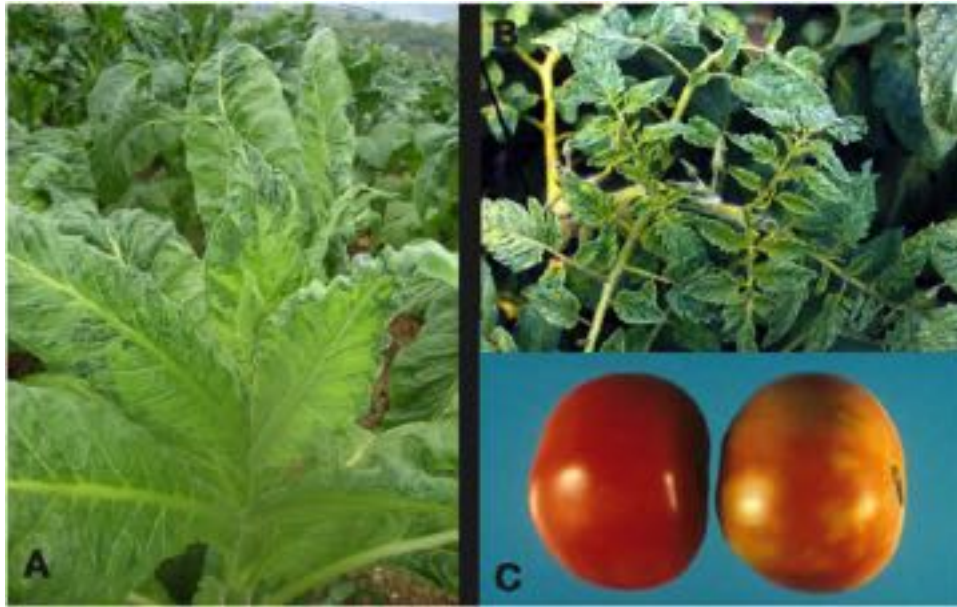


Fig. 5 Symptoms of Tobacco mosaic virus on tobacco White Burley (A. left) (Courtesy of Dominique Blancart, INRA, France) on tomato leaves (B. top right) and fruits (C. bottom right) (Courtesy T.A. Zitter, Vegetablemdonline, Cornell University, Ithaca, NY).

families, and at least 125 individual species, including tobacco, tomato, pepper (many members of the family Solanaceae), cucumbers, and a number of ornamental flowers, including orchids.

Many tobamoviruses infect plants producing a variety of virus-specific leaf pigment patterns, described by terms such as mosaics, mottles and yellows, depending on the virus tested. The first symptom of this virus disease is a light green coloration between the veins of young leaves. This is followed quickly by the development of a mosaic symptom of light and dark green areas on the leaves. "Rugosity" (crinkling of leaves) is observed in several cases of plant infection by tobamoviruses, as have other leaf deformations. These symptoms are more pronounced on younger leaves. TMV infection does not result in plant death, but if infection occurs early in the season, plants could be stunted. On tobacco, lower older leaves could be subject to "mosaic burn" especially during periods of hot and dry weather. In these cases, large dead areas develop in the leaves. This constitutes one of the most destructive phases of TMV infection. However, if TMV infects crops like grape and apple, it is almost symptomless. TMV over the last century has been known to cause a relatively stable production loss for tobacco of up to two percent per year in North Carolina, which is financially important for the tobacco industry.

TMV and ToMV are worldwide considered as major diseases for vegetable production (Fig. 5) and cause tremendous losses in many solanaceous crops, especially in greenhouses where a lot of movement and passages of workers happen.

Transmission

TMV and related tobamoviruses are transmissible to new plants only by contact of a plant with an inoculum. Therefore plants may become infected by rub inoculation of virions containing substances like Insect, animals, and humans visiting the fields. Plant infection begins when healthy plants contact virus contaminated sources. Particularly effective are rubbing with an aqueous suspension of virions. Animals moving through a field of infectious plants can, by brushing, pick up virions and transfer them to new host plants. Insects feeding on infected plants can also transmit tobamoviruses into their next host. Bee hives from an infected field can initiate an infection in another field. It is not clear whether the hive contains infectious particles and must itself be considered as the vector or not.

TMV is known as one of the most stable viruses. It has a very wide survival range. As long as the surrounding temperature remains below approximately 40°C, TMV can maintain its infectious capacity and all it needs is a host to infect. Greenhouses and botanical gardens provide the most favorable condition for TMV spreading, due to the high population density of possible hosts, the number of workers going through the plants and the constant temperature throughout the year.

One of the common and useful control methods for TMV is sanitation, which includes removing infected plants and washing hands and tools in between each planting. Crop rotation is also employed to avoid infected soil/seed beds for some time. As for any plant disease, looking for resistant varieties against TMV has been done for tobacco, tomato and other plants. Furthermore, the cross protection method has been used, whereby a more severe strain of TMV is inhibited by previously infecting the host plant with a mild strain of TMV. This method has been used for decades in the vegetable industry in Europe. Breeders have selected genes of resistance in tobacco, tomato and other crops, and those genes have been used in the industry. Precautionary measures taken in the use of tools and other objects has also been proven to decrease TMV spread in crops.

Further Reading

- Fraile, A., Garcia-Arenal, F., 2013. Tobamoviruses as models for the study of virus evolution. *Advances in Virus Research* 102, 117.
- Gibbs, A.J., 1977. Tobamovirus Group. CMI/AAB Descriptions of Plant Viruses, No. 184. Available at: <http://www.dpvweb.net/dpv/showadvp.php?Dpvno184>.
- Gibbs, A.J., Torronen, M., Mackenzie, A.M., *et al.*, 2011. The enigmatic genome of *Chara australis* virus. *Journal of General Virology* 92, 2679.
- Hull, R., 2002. *Matthews' Plant Virology*, fourth ed. New York: Academic Press.
- Ishibashi, K., Ishikawa, M., 2016. Replication of tobamovirus RNA. *Annual Review of Phytopathology* 54, 55–78.
- Lewandowski, D.J., 2005. Tobamovirus genus. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselburger, U., Ball, L.A. (Eds.), *Virus Taxonomy, Classification and Nomenclature of Viruses: Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press, pp. 1009–1014.
- Malpica-López, N., Rajeswaran, R., Beknazariants, D., *et al.*, 2018. Revisiting the roles of tobamovirus replicase complex proteins in viral replication and silencing suppression. *Molecular Plant-Microbe Interactions* 31, 125.
- Pogue, G.P., Lindbo, J.A., Gargner, S.J., Fitzmaurice, W.P., 2002. Making an ally from an enemy: Plant virology and the new agriculture. *Annual Review of Phytopathology* 40, 45–74.
- Scholthof, K.B., 2005. Tobacco mosaic virus: A model system for plant virology. *Annual Review of Phytopathology* 42, 13–34.
- Scholthof, K.B., Shaw, J.G., Zaitlin, M., 1999. *Tobacco Mosaic Virus: One Hundred Years of Contributions to Virology*. St. Paul, MN: APS Press.
- van Regenmortel, M.H.V., Fraenkel-Conrat, H., 1986. *The Plant Viruses: The Rod-Shaped Plant Viruses*. 2. New York: Plenum.
- Vlok, M., Gibbs, A.J., Suttle, C.A., 2019. Metagenomes of a freshwater Charavirus from British Columbia provide a window into ancient lineages of viruses. *Viruses* 11, 299.

Tobraviruses (*Virgaviridae*)

Stuart A MacFarlane, The James Hutton Institute, Invergowrie, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S.A. MacFarlane, Tobravirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00515-X](https://doi.org/10.1016/B978-012374410-4.00515-X).

Glossary

Agroinfiltration Infiltration into air spaces within the plant leaves of cultured *Agrobacterium tumefaciens* bacteria leading to (usually) transient expression of functional genes or virus infectious clones in plants by transcription from a binary plasmid carried within the bacteria.

NM infection Infection of some plant species (primarily potato) by only the RNA1 molecule of TRV.

Initially erroneously thought to be “non-multiplying” hence NM.

Spraing Symptoms of dark flecking and arc-formation in tubers of some potato cultivars resulting from TRV infection.

Systemic infection Movement of virus from the point of inoculation to more distantly located leaves or roots via the plant vascular system (phloem).

Taxonomy and Characteristics

The genus *Tobravirus* is comprised of three species, the type species *Tobacco rattle virus* (TRV) together with *Pea early-browning virus* (PEBV) and *Pepper ringspot virus* (PepRSV), which was initially referred to as the CAM strain of TRV. The genus has been assigned to the *Virgaviridae* family. Tobraviruses have a genome of two, positive-sense, single-strand RNAs that are packaged separately into rod-shaped particles. In some situations, the larger genomic RNA (RNA1) can cause a systemic infection in the absence of the second, smaller RNA (RNA2) and without the formation of virus particles. Tobraviruses are transmitted between plants by root-feeding nematodes of the genera *Trichodorus* and *Paratrichodorus*, and in some plant species are also seed transmitted.

Virus Particle Production and Structure

Tobravirus RNA1 is encapsidated into the L (long) particle with a length of between 180 and 215 nm, depending on virus species, and RNA2 is encapsidated into S (short) particles which range in length from 46 to 115 nm, depending on virus isolate (Fig. 1). Both L and S particles have an apparent diameter of 20–23 nm, depending on the technique used to examine them. *In vitro* translation experiments using RNA extracted from purified virus preparations, as well as studies with the TRV SYM isolate, which has an unusual genome structure, showed that some, if not all, tobnavirus sub-genomic (sg)RNAs are also encapsidated, in particles of various lengths. The tobnaviruses encode a single coat protein (CP), molecules of which assemble in a helical arrangement around a central cavity with a diameter of 4–5 nm, and with a distance of 2.5 nm between successive turns of the helix. *In vitro* reconstitution experiments suggested that virus particle formation initiates at the 5′ end of the viral RNA, although the encapsidated sgRNAs do not carry the 5′ terminal part of the virus genomic RNAs. No specific “origin-of-assembly” sequence has been identified for any of the different tobnavirus RNAs. Peptide mapping showed that the major antigenic regions of the CP are, in descending order of strength, the C-terminal 20 amino acids (aa), 5 aa in the central region of the protein and 5 aa at the N-terminus. This and other spectroscopic analysis suggest that the N- and C-termini are exposed on the outer surface of the particle, while the central region is exposed in the central canal (where interactions of the CP with the viral RNA take place). The C-terminal domain appears to be unstructured and plays a role in interactions with other virus proteins that are involved in nematode transmission of the virus. Particularly with TRV, there is significant amino acid sequence difference between the CP of different virus isolates, resulting in many different serotypes of this virus.

M-Type and NM-Type Infections

In early studies with tobnaviruses (usually with TRV) the infectivity of fractionated and purified L and S particles was examined, showing that L particles were infectious (producing local and systemic symptoms on particular hosts) but that S particles were not. However, plants infected with L particles did not contain virus particles, whereas, plants infected with both L and S particles did contain (both) virus particles. This was explained by sequencing, which revealed that RNA1 encodes proteins for RNA replication and movement, whereas, RNA2 encodes the CP. Because infections with RNA1 do not produce virus particles they were very difficult to maintain by repeat inoculation of sap extracts, and were referred to as non-multiplying (NM-type) infections. Infections derived from both RNA1 and RNA2 and producing virus particles were easily passaged and were referred to as multiplying (M-type) infections. Subsequently it was found that extraction with phenol of RNA from NM-infected plants in fact produced highly infectious preparations.

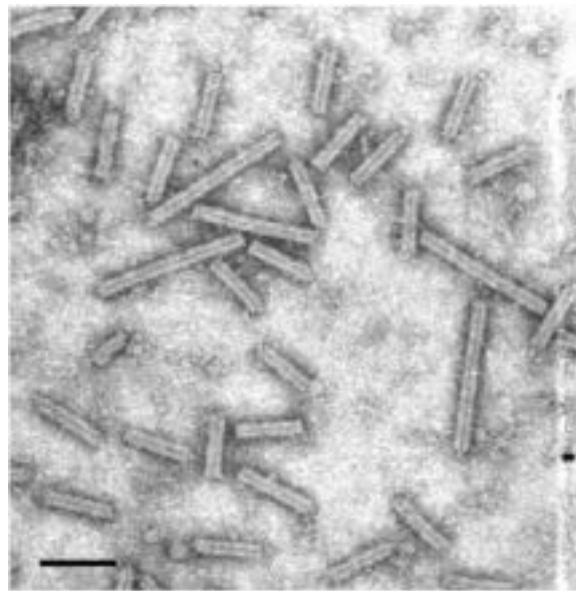


Fig. 1 Electron micrograph of long and short particles of TRV isolate RH. Bar is 100 nm.

Early observations also suggested that NM-type infections caused more severe symptoms than M-type infections and moved only slowly up the plant, giving rise to the theory that un-encapsidated RNA1 could spread systemically only from cell to cell via plasmodesmata. However, experiments with TRV and PEBV carrying defined mutations in the CP gene demonstrated unequivocally that un-encapsidated virus RNAs moved very rapidly via the vascular system in both *Nicotiana benthamiana* and *N. clevelandii*. In addition, with wild type (encapsidated) virus systemic infection always included both RNA1 and RNA2 but with the (un-encapsidated) CP mutants RNA2 occasionally became separated from RNA1 and was not detected in some systemic leaves. One PEBV mutant carried only a small (28 aa) deletion at the C-terminus of the CP but did not produce virus particles or indeed any detectable CP in infected plants. Nevertheless, this mutant moved systemically as quickly as did wild type virus and in this case RNA2 did not become separated from RNA1. This suggests, perhaps, that some form of CP, which can be present at very low levels but is not incorporated into particles, enables the co-ordinated transport of RNA1 and RNA2 through the phloem.

In another study, plants were infected with TRV RNA1 and two types of RNA2, one wild type and the second encoding a CP in which the 15 aa at the C-terminus were replaced with 3 non-viral residues. Both RNA2 species were encapsidated by the CP that they encoded but neither appeared to be encapsidated by the CP from the other RNA2. Apparently, although RNA1 is encapsidated *in trans*, RNA2 is only encapsidated *in cis*. The mechanism for this is not known but may be linked to the role of CP in the co-ordinated systemic movement of RNA1 and RNA2.

Genome Structure and Expression

The viral RNAs have a 5' methylated guanosine cap. The 3' end of the RNAs is not polyadenylated but folds into a tRNA-like pseudoknot structure that, in contrast to some other virus RNAs, cannot be aminoacylated. RNA1 is from 6.8 to 7 kb in size, and RNA2 varies between isolates from 1.8 to 4.2 kb (**Fig. 2**). By February 2019 GenBank contained complete sequences for 17 TRV RNA1 and 36 TRV RNA2 molecules. However, for both PEBV and PepRSV only a single RNA1 and three RNA2 sequences were reported.

RNA Sequences

The larger RNA (RNA1) is highly conserved in nucleotide sequence between different isolates of the same virus (*e.g.*, 99% identity between RNA1 of TRV isolates SYM and ORY) but there is much less identity between isolates of the different tobnavirus species (*e.g.*, 62% identity between RNA1 of TRV isolate SYM and PEBV isolate SP5). The smaller RNA (RNA2) varies considerably both in terms of overall nucleotide sequence identity as well as protein coding capacity between isolates of the same virus, although phylogenetic analysis shows that the coat proteins (CPs) of the different tobnaviruses, which is encoded by RNA2, are more closely related to each other than to CPs from viruses in other genera. The tobnaviruses are most closely related to tobamoviruses, in terms of particle structure, overall gene organization and viral protein sequence homologies.

Depending on the specific virus and isolate, RNA1 has a 5' non-coding region (NCR) of between 126 and 202 nt, and a 3' NCR of 459 to 255 nt. RNA2 has a much larger 5' non-coding region (NCR) of between 470 and 710 nt, and a 3' NCR of 780 to 392 nt. There is almost no sequence homology between the 5' NCRs of the tobnaviruses, however, the 25 nt at the 3' terminus of the TRV

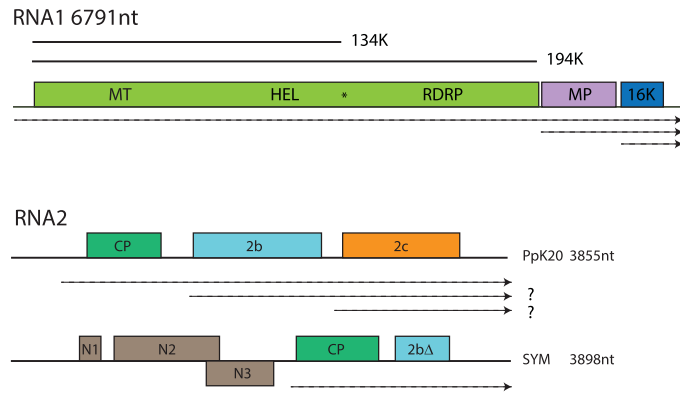


Fig. 2 Genome diagram and expression strategy of TRV isolates PpK20 and SYM. Open boxes denote virus genes. The solid lines above the RNA1 genes show the location of the two, overlapping replicase genes. The 134K protein contains methyltransferase (MT) and helicase (HEL) motifs. The C-terminal part of the 194K protein contains an RNA-dependent RNA polymerase (RDRP) motif. The asterisk denotes the 134K gene termination codon where readthrough translation to produce the 194K protein occurs. The movement protein (MP) gene is also known as the 29K or 1a gene. The cysteine-rich 16K gene is also known as the 1b gene. For RNA2, CP denotes the coat protein gene. N1 to N3 denote novel open reading frames present in isolate SYM RNA2. 2bΔ is a truncated 2b gene. The dashed lines beneath each RNA denote the 3' co-terminal sub-genomic RNAs that are known or suspected (?) to exist.

and PEBV RNAs are identical and there is 70% identity over the next 140 nt. Consequently, when recombinants were made between the RNA2 of TRV PpK20 and PEBV SP5, the RNA1-encoded replicase complex of both viruses could replicate RNA2 molecules carrying the 3' NCR from either virus. However, RNA2 carrying the TRV 5' NCR could only be replicated by a TRV RNA1, and RNA2 carrying the PEBV 5' NCR could only be replicated by a PEBV RNA1.

Expression of Virus Genes

The 5' proximal gene of RNA1, encoding viral replicase proteins, is expressed by direct translation of the genomic RNA. The two other RNA1-encoded genes (1a/29K and 1b/16K) are expressed from sgRNAs that start with the sequence AUA (within a conserved motif GCAUA) and that are co-terminal with the 3' end of the genomic RNA.

The 5' proximal gene of RNA2 of almost all tobnavirus isolates encodes the CP. However, the CP gene is located at least 470 nt downstream of the 5' terminus of RNA2, and this region contains numerous (at least six) AUG codons upstream of the translation initiation codon of the CP gene, that are inhibitory to CP expression. Experiments to delete parts of the 5' NCR of PEBV RNA2 showed that only a very small amount of CP is translated *in vitro* from full-length (genomic) RNA2 but that translation is increased 25-fold when the 5' NCR is removed. Thus, even though the CP gene is usually the 5' proximal gene on RNA2, it (and the other genes on RNA2), is expressed from a sgRNA. However, the CP gene of TRV SYM is located near to the 3' end of RNA2, downstream of other genes. The sgRNA for this CP is much shorter than genomic RNA2 and is encapsidated into a clearly identifiable VS (very short) virus particle.

A stem-loop sequence is found upstream of tobnavirus sgRNA start sites and is particularly conserved for the CP sgRNA. The introduction of mutations into this structure (8 bp stem, 4 nt loop) upstream of the PEBV CP gene showed that the stem-loop forms an essential part of the sgRNA promoter, and that the structure of the stem rather than its actual sequence is important for sgRNA synthesis *in planta*. The promoter regions for the TRV and PEBV CP genes could be used as a cassette and moved to other locations in RNA2 to express non-viral genes.

Recombination and Sequence Deletions in Tobnavirus RNAs

The large variation in the overall length of RNA2 from different tobnavirus isolates reflects the fact that most are recombinant molecules, where the 3' part of RNA2 has been replaced by sequences from the 3' part of RNA1. The recombination junction can occur at any position in RNA2 that is downstream of the CP gene, and the recombinant RNA2 molecule may retain none, some or even both of the other genes (2b and 2c) that might be considered RNA2-specific. The region of RNA1 that is transferred to RNA2 always includes the 3' NCR, often includes the 1b/16K/12K gene immediately upstream of the 3' NCR and may occasionally include part of the 1a/29K gene that is located upstream of the 1b gene. This means that many of these recombinant isolates carry two, probably both functional, copies of the 1b gene, one on RNA1 and the second on RNA2. In one study, a single *Alstroemeria* plant was found to contain seven different recombinant TRV RNA2 molecules, three with PEBV RNA1-like 3' NCR sequences and four with different TRV RNA1 3' coding and NCR sequences. The mechanism for such recombination is not known but is speculated to be caused by template switching by viral replicase from RNA1 to RNA2 during minus strand RNA synthesis. Sequences that resemble the 5' terminus of tobnavirus genomic RNAs (rich in A and U residues) are often found at or near the recombination junction, and conceivably could facilitate the recombination event. It is also not completely clear when tobnavirus recombination occurs, though the feeling is that repeated passage of virus by mechanical inoculation in glasshouse-grown plants, bypassing the nematode

transmission process, either encourages recombination or reduces selection against the survival of recombinants. Nevertheless, at least one known recombinant isolate, TRV PaY4, which was cloned after only a very limited period of multiplication in the glasshouse, did retain its ability to be nematode transmitted. When a TRV-infected potato plant was propagated by bud cuttings, a small (80 bp) deletion in RNA2 was detected only 5 days after planting, and by 33 days post planting an RNA2 was found lacking nearly 50% of its original sequence. By 55 days post planting no RNA2 could be detected in the plant, even though an RNA1 infection was maintained.

A different recombination process also occurs in which the CP and/or 2b gene in RNA2 of some TRV isolates appears to have been derived from PEBV. For this reason, serological analysis is not always successful in discriminating between the different tobaviruses. Attempts to reproduce this recombination by co-inoculating TRV and PEBV to plants in the glasshouse were not successful, suggesting that this may be only a rare occurrence, or may be stimulated by particular environmental conditions. Nevertheless, one report showed that 30% of the TRV isolates recovered from fields in the coastal bulb growing region in the Netherlands were TRV/PEBV recombinants.

For PepRSV RNA2, the three sequenced molecules all have RNA1-derived sequences at their 3' end and do not retain any 2b or 2c sequences. However, it is assumed in nature that this virus is nematode transmitted and, thus, that isolates carrying the 2b and 2c genes must exist.

Viral Proteins

RNA1

RNA1 encodes proteins for virus RNA replication and intra-plant movement (local and systemic). The 5' half of RNA1 encodes a large (134–141 kDa) protein that contains methyltransferase and helicase domains and is expected to be part of the viral replicase complex. Readthrough translation of the stop codon of this protein produces a larger protein (194–201 kDa) that contains motifs associated with RNA-dependent RNA polymerase proteins. *In vitro* translation experiments showed that the opal (UGA) translation termination codon of the TRV 134 kDa protein is suppressed by tRNAs that incorporate tryptophan or cysteine at this position. Downstream of the replicase genes is the 1a gene encoding a 29–30 kDa cell-to-cell movement protein (MP) belonging to the “30K” superfamily of plant virus movement proteins. Disruption of this gene prevents accumulation of TRV in inoculated leaves but can be overcome by co-inoculation with tobacco mosaic virus (TMV) or in transgenic plants expressing the TMV 30K movement protein. The C-terminal 1b gene encodes a 12 kDa (PEBV; PepRSV) or 16 kDa (TRV) cysteine-rich protein that is involved in seed transmission of PEBV in pea, and pathogenicity of TRV and PEBV. The TRV 16K protein is a suppressor of RNA silencing that is predominantly cytoplasmic but can also be found in the nucleus and nucleolus. The nuclear/nucleolar localization is important for the interaction of 16K with the plant protein coilin and the modulation of host plant anti-viral defense pathways. The TRV 16K protein is required to enable the virus to transiently enter the plant meristem. In *N. benthamiana* plants, a TRV 16K mutant virus does not enter the meristem, causes more severe infection symptoms and persists in the plant, whereas, with wild type TRV the plant recovers as the infection subsides and symptoms disappear. One study reported that the TRV 29K protein could also suppress RNA silencing when expressed during virus replication but not when expressed transiently via agroinfiltration. The PepRSV 12K protein also is a silencing suppressor and was found to enable potato virus X (PVX) to infect Arabidopsis when co-inoculated with PepRSV.

RNA2

RNA2 encodes proteins for virus particle formation and transmission by nematodes. Generally, the CP gene is the most 5' proximal gene on RNA2, even though it is expressed from a sgRNA rather than from the full-length, genomic RNA. As mentioned above TRV SYM has an unusual gene organization, with three potential genes located upstream of the CP gene. Several other TRV isolates from potato have also been identified in which the CP gene is preceded on RNA2 by novel genes that have no known function. Several studies have identified the CP as being involved in the nematode transmission process, and deletion of part of the C-terminal domain of the CP prevented the transmission of PEBV without affecting virus particle formation. Infectious, nematode-transmissible clones of three TRV isolates and one PEBV isolate have been constructed. In addition to the CP, these RNAs all carry two other genes encoding the 2b and 2c proteins, both of which (for PEBV) have been detected by western blotting in extracts of virus-infected plants. Mutation studies showed that the 2b gene is necessary for transmission of TRV and PEBV, whereas the 2c protein was only required for transmission of PEBV. This difference may reflect the different species of nematode that transmitted the particular virus isolates used in these studies, rather than a clear mechanistic difference between TRV and PEBV transmission. A further, small open reading frame, encoding a putative 9 kDa protein immediately downstream of and in-frame with the CP gene, is present in PEBV-TpA56 and TRV TpO1. For PEBV, mutation of this gene greatly reduced nematode transmission frequency, although the protein was not detected by western blotting.

The 2b proteins of the different tobavirus isolates share some amino acid sequence homology with each other, ranging from only 11% identity (TRV Umt1 v. TRV PaY4) to 99% identity (TRV Umt1 v. TRV OR2). They also range in size from 238 aa

(TRV PaY4) to 354 aa (TRV PpK20). The TRV PaY4 2b protein has been shown to act *in trans*, with nematode transmission of a 2b-mutant virus being complemented by co-infection with wild type virus. However, this appears to be isolate specific, as the TRV PaY4 2b mutant was complemented only by wild type TRV PaY4 and not by wild type TRV PpK20.

The 2b protein may influence nematode transmission by more than one mechanism. In a microscopy study of the roots of *N. benthamiana* plants infected by PEBV, large numbers of particles of the wild type virus were found in all regions of the root tip, which is where the vector nematodes preferentially feed, whereas virus particles of a mutant deleted for the 2b and 2c genes were present in roots only in much lower numbers. A similar enhancement was found in the efficiency of invasion of roots in *N. benthamiana*, as well as in leaves and roots of *Arabidopsis thaliana*, when TRV carrying the 2b gene was compared with TRV lacking the 2b gene.

Chemical analysis studies indicate that TRV infection increases the production of specific root volatiles and that TRV-infected plants are more attractive to vector nematodes. Furthermore, it seems that the TRV 2b gene plays some, as yet unclear, role in this process.

The 2b protein also physically interacts with the TRV CP (as examined by yeast two-hybrid (Y2H) analysis) and the virus particle (as examined by immuno-electron microscopy). One suggestion is that the 2b protein acts as a bridge to trap virus particles to specific sites of retention on the nematode esophageal cuticle. Aggregates of virus particles have been observed using the electron microscope to collect in this part of the nematode but the co-location of the 2b protein has not yet been demonstrated. It appears that interaction between the CP and 2b protein stabilizes the 2b protein, as in plants infected with wild type TRV PaY4, carrying both PaY4 CP and PaY42b genes, the 2b protein was detected by western blotting. However, with a recombinant TRV carrying the TRV PpK20 CP gene and the TRV PaY4 2b gene, the 2b protein could not be detected, possibly because these proteins from two different isolates cannot interact with one another.

Although the 2c protein is involved in transmission of PEBV, very little else is known about this protein. Amino acid sequence homologies between the 2c proteins of different TRV isolates range from almost none to over 95% identity. In a Y2H assay the TRV PpK20 2c protein interacted with the PpK20CP, and removal of the CP C-terminal flexible domain did not affect the interaction.

Tobraviruses as Gene Expression/Silencing Vectors

As RNA1 encodes all the proteins necessary for tobnavirus replication and movement, the RNA2 can be modified to carry non-viral sequences without greatly affecting virus infection. Expression vectors have been constructed from all three tobnaviruses, in which a duplicate CP promoter sequence is inserted downstream of the native CP gene followed by restriction sites to allow the cloning of other sequences. Together, the tobnaviruses can infect a wide range of plant species, often without causing particularly severe symptoms, features which increase their utility as expression vectors.

Plants infected with tobnaviruses often undergo a rapid recovery in which infection symptoms and virus levels fall dramatically, although the plants do not become free of virus, most particularly in the meristem regions. However, these plants have developed a strong resistance to further infection by the same virus, most likely by an RNA silencing-based mechanism. Although not well understood, it seems that tobnaviruses are potent triggers of RNA silencing but do not encode a strong silencing suppressor protein to counteract this host defense activity. A consequence of this is that when a host plant sequence is inserted into the virus genome, very strong silencing is initiated that targets expression of the host gene itself, a process known as virus-induced gene silencing (VIGS). TRV has become one of the most widely used VIGS vectors for studies of plant gene function in many different plant species, and PEBV is also a very effective VIGS vector for studies in pea.

Diseases Caused by Tobraviruses

TRV is found in many countries (Europe, North America, Japan, and Brazil) and has a particularly wide host range, infecting more than 100 species in nature and more than 400 species when tested in the glasshouse, although not all of these infections are systemic. As tobnaviruses are transmitted by soil-inhabiting nematodes, in the field infection may be limited to the roots. Weeds may play an important role in the maintenance and spread of tobnaviruses, with *Capsella bursa-pastoris*, *Senecio vulgaris*, *Stellaria media*, and *Viola arvensis* being the most commonly found (in the UK) weed hosts of TRV. Virus transmission in seed of these plants was also reported. Many crop plants are infected by TRV, the major diseases being those of potato and ornamental bulbs (narcissus, gladiolus, tulip, lily, and crocus). The symptoms of TRV infection in potato are the formation of arcs and flecks of brown corky tissue in the tuber which is referred to as spraing, and which can make the tuber unfit for sale. Both M-type and NM-type infections can produce spraing symptoms, which have been shown to be a hypersensitive (defense) response by the plant to the virus. It was thought that potato cultivars that did not show spraing symptoms were resistant to TRV, however, some infections can be symptomless in particular cultivars. Nevertheless, even these symptomless infections lead to significant reductions in tuber yield and tuber quality. Inclusion of tubers carrying symptomless infection could have major consequences for the production and distribution of seed potatoes (Fig. 3).



Fig. 3 Cross-section of a potato tuber showing spraing symptoms caused by TRV infection.

PEBV has been reported in several European countries (UK, Netherlands, Italy, Belgium, Sweden) as well as Algeria and Morocco. In the field it infects mainly legumes, including pea, faba bean, French bean, lupin, and alfalfa. Several weeds and other crop plants may be infected, although often only in the roots. Early studies referred to Broad bean yellow band virus (BBYBV) as a defined virus infecting various legumes. This is now known to be a distinct serotype of PEBV.

PepRSV has only been reported from Brazil, where it infects pepper, tomato, and artichoke, as well as local weed species.

Further Reading

- Harrison, B.D., Robinson, D.J., 1978. The Tobraviruses. *Advances in Virus Research* 23, 25–77.
- MacFarlane, S.A., 1999. Molecular biology of the Tobraviruses. *Journal of General Virology* 80, 2799–2807.
- MacFarlane, S.A., 2010. Tobraviruses – Plant pathogens and tools for biotechnology. *Molecular Plant Pathology* 11, 577–583.
- MacFarlane, S.A., Robinson, D.J.R., 2004. Transmission of plant viruses by nematodes. In: Gillespie, S.H., Smith, G.L., Osbourne, A. (Eds.), *Proceedings of the SGM Symposium 63 on Microbe-Vector Interactions in Vector-Borne Diseases*. Cambridge University Press.

Tomato Leaf Curl New Delhi Virus (*Geminiviridae*)

Supriya Chakraborty and Manish Kumar, Jawaharlal Nehru University, New Delhi, India

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa	Amino acid(s)	NASH	Nucleic acids spot hybridization
AGO	Argonaute protein	NBS-LRR	Nucleotide-binding site–leucine-rich repeat
amiRNA	Artificial microRNA	nm	Nanometer(s)
CAS	CRISPR-associated protein	NSP	Nuclear shuttle protein
CP	Coat protein or capsid protein	nt	Nucleotide(s)
CRISPR	Clustered regularly interspaced short palindromic repeats	NtRDR1	RNA-Dependent RNA Polymerase 1 from <i>N. tabacum</i>
dsRNA	Double-stranded ribonucleic acid	PCR	Polymerase chain reaction
ERF4	Ethylene response factor-4	PTGS	Post-transcriptional gene silencing
HR	Hypersensitive response	RDR6	RNA-dependent RNA polymerase 6
kb	Kilobases	SCR	Satellite conserved region
kDa	Kilo Daltons	SGS3	Suppressor of gene silencing 3
LAMP	Loop mediated amplification	siRNA	Small interfering RNA
LTP	Lipid transfer protein	SRAP	Sequence-related amplified polymorphism
MEAM1	Middle-East Asia minor I	sRNA	Small RNA
MED1	Mediterranean 1	ssDNA	Single-stranded deoxyribonucleic acid
miRNA	microRNA	TGS	Transcriptional gene silencing
MP	Movement protein	TrAP	Transcriptional activator protein
		VSR	Viral suppressor of RNA silencing

Glossary

Agro-inoculation A method enabling the transfer of DNA (such as cloned begomoviral molecules) into a host plant using *Agrobacterium tumefaciens*.

Rolling circle amplification A technique that allows amplification of circular DNA molecules (such as geminivirus genomes) with $\phi 29$ DNA polymerase and random primers under isothermal conditions.

Tolerance Defined as an interaction in which viruses accumulate to some degree without causing significant loss of vigor or fitness to their hosts.

Viral titer detection Radiolabelled nucleic acid molecules (begomoviral DNA) detection by using virus gene-specific labeled probe through Southern blot experiments.

Viruliferous whitefly A whitefly insect-vector population carrying an infectious begomoviral molecule.

Introduction

Tomato (*Solanum lycopersicum*) belongs to the family *Solanaceae* and is one of the most important vegetable crops grown in India. India is the second-largest tomato producing country in the world. Tomato is originating from America, probably in Mexico, and the center of diversification was in South America, probably in Peru.

Tomato leaf curl New Delhi virus (ToLCNDV) is a bipartite begomovirus species (genus *Begomovirus*, family *Geminiviridae*) and is a well-known plant DNA virus responsible for the disease called Tomato leaf curl disease (ToLCD). The epidemic might be linked to the increased population of viruliferous whitefly (*Bemisia tabaci*), the insect vector accountable for the spread of the pathogen of ToLCD, around the world including India. The first reported incidence of this disease was from northern India in 1948. The occurrence was also documented from central India during the 1950s, followed by disease incidence in tomato-growing regions of South India. The disease has since begun to emerge as a major threat to the production of the tomato crop in India. The total yield loss in tomato production by begomoviruses ranges from 18% to 100%. Factors such as age of the plant, viruliferous whitefly populations, aggressivity of the virus isolate and susceptibility of the tomato variety influence the production and may cause up to 100% yield loss. Tomato plants affected by the leaf curl disease exhibit typical symptoms of upward and downward leaf curling, leaf wrinkling, leaf blistering, vein clearing, distortion of leaves, and stunted plant growth that might result in a complete yield loss (Fig. 1).

Tomato leaf curl New Delhi virus (ToLCNDV) has been recorded in recent decades to cause devastating damage to tomato yield and other vegetable crop productions in the Indian agriculture sector, notably the disease in the northern part of India is more widespread. Using infected tissue samples from two separate locations, ToLCNDV was first sequenced and characterized in India in 1992. Until recently, the epidemic of ToLCNDV was confined to South-East Asian countries like India, Bangladesh, Thailand, Indonesia, Sri-Lanka and Pakistan. Recent reports are indicative of its presence in other countries such as Spain, Morocco, Italy, and

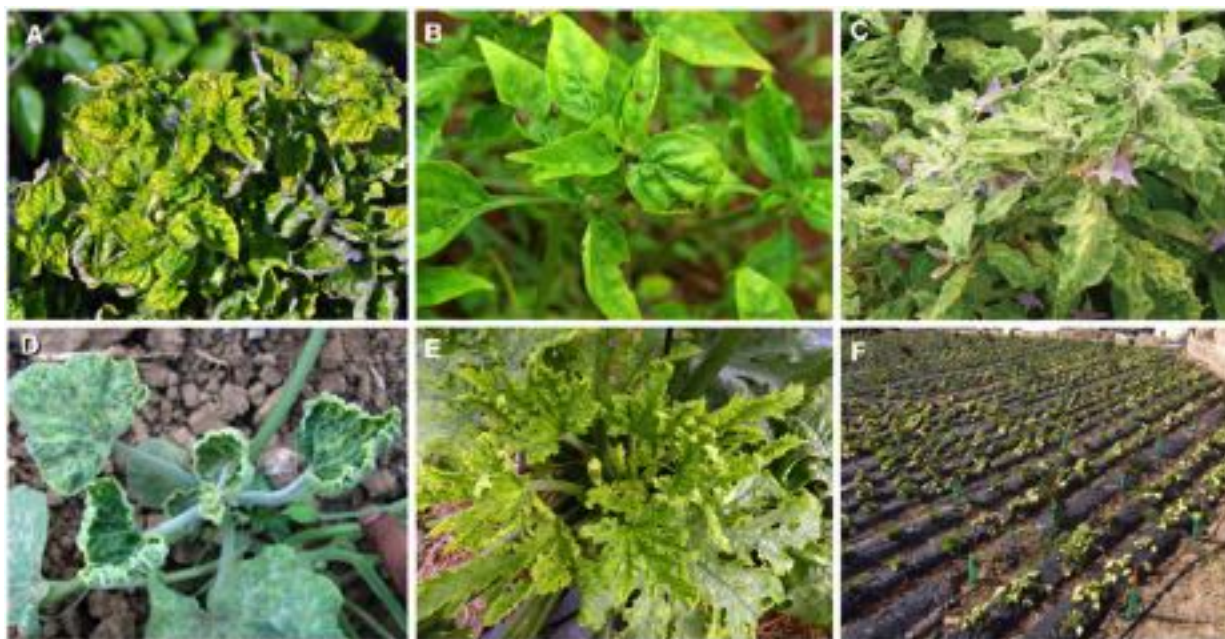


Fig. 1 Plants infected with ToLCNDV displaying typical yellow mosaic symptoms on leaves (A) tomato, (B) chilli, (C) egg plant (D) pumpkin, (E) zucchini plants, (F) An infected zucchini plants at the field. Zucchini plants pictures courtesy of E. Moriones, Institute of Subtropical and Mediterranean Horticulture La Mayora, Algarrobo-Costa, Málaga, Spain. Picture of pumpkin crop courtesy of B. Mandal, Indian Agricultural Research Institute, New Delhi, India and picture of eggplant reproduced from the research article by Pratap, D., Kashikar, A.R., Mukherjee, S.K., 2011. Molecular characterization and infectivity of a Tomato leaf curl New Delhi virus variant associated with newly emerging yellow mosaic disease of eggplant in India. *Virology Journal* 8 (1), 305.

Israel. However, ToLCNDV infectivity studies have not been performed yet on tomato, or in any other solanaceous or cucurbitaceous crops, in other countries and therefore the virus may be more widespread than anticipated.

Tomato Leaf Curl Disease and Virus Biodiversity on Tomato in the Indian Subcontinent

Members of the *Geminiviridae* family infect various monocots and dicots plants across the continents causing severe losses. These insect-transmissible geminiviruses are encapsulated in 2.7–5.2 kb twinned icosahedral particles with one or two (mono or bipartite) circular, non-enveloped single-stranded DNA (ssDNA) genomes. The members of the family *Geminiviridae* are further classified into nine genera such as *Becurtovirus*, *Begomovirus*, *Capulavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Mastrevirus*, *Topocuvirus*, and *Turncurtovirus* based on their genome organization, nucleotide sequence identities, and mode of insect transmission. Furthermore, *Begomovirus*, the largest genus of the family, comprises approximately 450 species of geminiviruses that infect many economically important crops, ornamental plants as well as weed plants, around the world, and they are all transmitted by whiteflies (*Bemisia tabaci* Genn).

As per the current International Committee on Virus Taxonomy (ICTV) recommendation and revision of taxonomy from the Xth ICTV Report, based on pairwise nucleotide alignment of the begomovirus genome sequences, the name of ToLCNDV covers in fact four geminivirus species. These four virus species are *Tomato leaf curl New Delhi virus* (ToLCNDV DNA-A and DNA-B, U15015 and U15017), *Tomato leaf curl New Delhi virus 2* (ToLCVNDV2, JQ897969), *Tomato leaf curl New Delhi virus 4* (ToLCVNDV4, KF551592) and *Tomato leaf curl New Delhi virus 5* (ToLCVNDV5, EF450316). The taxonomic classification also depends upon other factors such as relatedness with the major or minor parental sequences, the frequency and extent of recombination and the biological properties of the viruses. More than 280 ToLCNDV's complete full-length sequences have been reported till date at NCBI GenBank public database and out of that number at least 80 of them are reported from India infecting different crop plants but tomato. This virus seems to be very successful at invading space and new host niches, in the Indian subcontinent and also in several regions of the world. This information suggests that the emergence of ToLCNDV, which is already underway, could be the major agricultural threat for vegetable crops to India and other countries in the very near future.

Other begomovirus species such as *Tomato leaf curl Gujarat virus*, *Tomato leaf curl Bangalore virus*, *Tomato leaf curl Bangladesh virus*, *Tomato leaf curl Pune virus*, *Chilli leaf curl virus*, *Chilli leaf curl India virus*, *Pepper leaf curl Lahore virus*, *Ageratum enation virus*, and *Tobacco curly shoot virus* are also known to be associated with ToLCD, affecting the tomato production in the Indian subcontinent. The fact that ToLCD remains yet to be fully characterized even after 70 years of its first report, might be due to the diversity of the viruses causing the disease, broader host range, and high level of host-virus adaptability.

The full-length DNA sequence of an Indian isolate of ToLCNDV DNA-A (U15015) and DNA-B (U15017) was first elucidated in the early 1990s from a begomovirus infected tomato (*Solanum lycopersicum*) from New Delhi (hence the name of the virus species). The size of the ToLCNDV-[India: New Delhi: Severe:1992] DNA-A and DNA-B are 2739 and 2696 bp, respectively. Similarly, ToLCNDV2-[India: IANDS1:2011] (JQ897969) isolated from tomato in New Delhi, is 2735 bp, ToLCNDV4-[India: Junagadh: TC306:2011] (KF551592) isolated from tomato in Gujarat is 2739 bp and ToLCNDV5-[Bangladesh: cucumber:2006] (EF450316) isolated from *Cucumis sativus* (cucumber) infected leaves, from Bangladesh is 2738 bp. Although all of the above-mentioned ToLCNDV species are nearly similar in size, they share sequence-similarities of less than 90% and have been assigned so far into the four separate ToLCNDV species listed above.

Phylogeny and Molecular Diversity

Based on phylogeny, genome organization, and geographical distribution, begomoviruses are divided into two super groups namely the Old-World (OW) and New-World (NW) begomoviruses. Among the Old World viruses more than 57 begomovirus species have been identified on tomato and there are many other begomoviruses capable of infecting tomato as well. In the Indian subcontinent alone more than 17 species of begomoviruses have been identified from tomato. However, among all these viruses ToLCNDV has emerged in the recent years as the major virus infecting tomato in the subcontinent, as already mentioned four species bear the name of ToLCNDV and as of 2019, there was a total of 282 records at NCBI under the name of ToLCNDV. In other words there is a lot of begomovirus variability at large associated with the general syndrome of tomato leaf curl disease (ToLCD) in the world and in India particularly, but ToLCNDV has prevailed in the recent years on tomato and other crops in the Indian subcontinent.

When a detailed study comparison of all the 282 ToLCNDV isolates is performed, it is obvious that there are no strictly defined demarcation strains among all these isolates, with a few exception (see below) indicating that there have been a lot of recombination events among the individuals of this population. The pairwise multiple sequence alignment of all records bearing the name of ToLCNDV spans a great variability between 72% and 100%. Officially, four viruses, ToLCNDV (U15015), ToLCNDV2 (JQ897969), ToLCNDV4 (KF551592) and ToLCNDV5 (EF450316), have been identified by their sequence as representatives of four species. The original reference of ToLCNDV (U15015) scores 74% identity with ToLCNDV2, 87% with ToLCNDV5 and 89% with ToLCNDV4, however ToLCNDV2 is sharing 94% identity with the members of the species PeaLDV (identified since ToLCNDV2 was sequenced) indicating that ToLCNDV2 is an isolate of this species now called PeaLDV. For simplicity and clarity it is therefore proposed to drop the name ToLCNDV2. ToLCNDV5 is borderline and ToLCNDV4 is even more border line for species demarcation in the ToLCNDV population and it is believed it is best for the time being to consider them as unassigned isolates of the species ToLCNDV. In addition another isolate from *Capsicum annuum*, from New Delhi, is showing less than 75% identity with the original ToLCNDV reference (U15015), and 91% with members of the species ToLCNDV making it an isolate of that species. Finally a last isolate from a weed from Pakistan named ToLCNDV-[PK: Chiniot:ZI27: G11V4:12] (HG937523), is in fact 94–98% identical to PaLCV isolates and should be renamed as PaLCV-[PK: Chiniot:ZI27: G11V4:12].

Within the ToLCNDV cluster of the remaining 276 isolates spanning pairwise identities between 86% and 100%, one can recognize several sub-clusters that we recorded for simplicity from Strain A to P plus an additional 13 unassigned isolates that could each become a node for new strains when more samples will be sequenced. Among these 15 Strains some are very well defined with no overlap with others like the strain M and G for the Thai-Indonesian isolates, the strain C for the Spain-Italy isolates, or some Indian-Pakistan strains such as Strains L, A, B. The strains D, E, F, J, K, L and H are more overlapped and consequently more difficult to define. The original first isolate of ToLCNDV (U15015) belongs to the Strain H and it is the largest cluster within the ToLCNDV population with 86 isolates (Fig. 2).

Apparently for all the isolates collected so far there is some level of correlation with the geographical distribution for Europe and South East Asia, a correlation with the hosts has not been found. The same virus can infect a range of different host in the same region. Recently it is of interest to note that ToLCNDV has been identified on cotton (*Gossypium hirsutum*) in Pakistan along with the betasatellite Cotton leaf curl Multan betasatellite. These 8 cotton isolates cluster in the middle of the Strain H which is the cluster of the original ToLCNDV isolate (U15015).

Genome Organization and Function of Gene Products

Four families of ssDNA viruses (*Circoviridae*, *Parvoviridae*, *Spiraviridae* and *Geminiviridae*) infect a range of hosts such as plants, hyper-thermophilic archaeobacteria, vertebrates (mammals, birds, reptiles, fish) and invertebrates (arthropods, insects, crustacean, echinoderms). Among these four families, only the members of the family *Geminiviridae* are known to infect plants and the members are further categorized into nine genera and among them, only the genus *Begomovirus* contains either mono or bipartite genome.

ToLCNDV DNA-A comprises six open reading frames (ORFs), two in virion or sense strand (AV1/CP and AV2/Pre-coat) and four in complementary or antisense strand (AC1/Rep, AC2/Trap, AC3/Ren, and AC4). AC1 encodes a multifunctional and most conserved protein, Rep or replication-associated protein that initiates the viral replication. AC2 encodes Trap or transcriptional-activator protein and is involved in the modulation of transcriptional gene silencing (TGS) as well as in the post-transcriptional

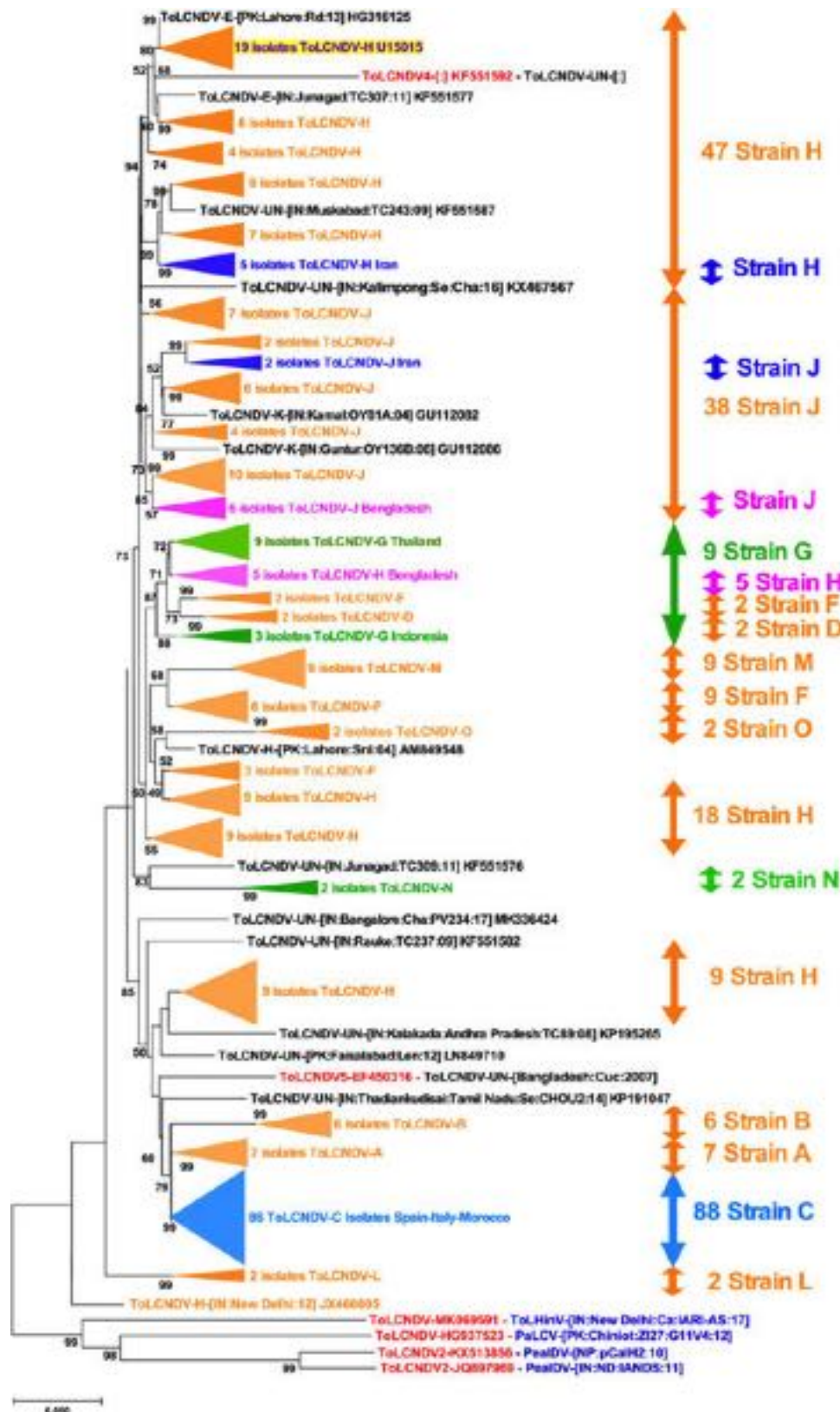


Fig. 2 Maximum-likelihood dendrogram of ToLCNDV isolates, available at the NCBI GenBank database. Accession numbers for all these sequences are mentioned in the phylogenetic tree. The nucleotide sequences are aligned using Muscle algorithm with bootstrap values of 1000 replicates. The scale bar indicates the nucleotide substitution rate per site, estimated by the Jukes-Cantor model. The branches of the individuals belonging to the same cluster/strain have been collapsed for simplification. The orange color indicate isolates from India-Pakistan, dark blue indicates isolates from Iran, Green from Thailand, Indonesia, pink from Bangladesh, pale blue from Spain-Italy-Morocco. The Unassigned isolates are written in black. The putative other ToLCNDV species are written in red and their corrected name in dark blue (other species) or black (unassigned ToLCNDV isolates).

gene silencing (PTGS) pathways. The viral DNA replication is facilitated by another early transcriptional gene, AC3 or REn or replication enhancer protein. AC4 is the least conserved ORF and structurally overlapped with the AC1 ORF. The AC4 functions as an RNA-silencing suppressor and also the suppressor of PTGS pathway. AC4 encoded protein of ToLCNDV interacts with tomato AGO4. Non-functional mutant of ToLCNDV AC4 remained however infectious on solanaceous crops. The AV1 ORF encodes coat protein (CP) which is a highly conserved protein facilitating the assembly and packaging of the viral genome. ToLCNDVAV2 encoded pre-coat protein acts as host silencing suppressor and pathogenicity determinant.

In nature a severe and a mild strain of ToLCNDV (U15015 and U15016, respectively) isolated from tomato in India, the viruses were cloned and sequenced, and infectious clones were produced. It was demonstrated on *Nicotiana benthamiana* and *S. lycopersicum* plants, that the severe symptom was associated with the higher capacity of the Rep of the severe isolate to replicate components A and B. The infectivity data suggested that the asparagine at the 10th position (Asn10) of the Rep protein (AC1) of the severe strain specifically recognizes the 3rd base of the conserved iteron sequences although both severe and mild strain shares 94% pairwise nt identity on their Rep genes. It is therefore possible that in nature such mutation may be able to trans-replicate more molecules that could benefit the ToLCNDV DNA-A and DNA-B to impact its host range and ultimately its epidemiology.

Bipartite begomoviruses encode two additional ORFs (BV1 and BC1) in their DNA-B component. BV1 ORF encodes nuclear shuttle protein (NSP) in the virion or sense strand and BC1 encodes movement protein (MP) in the complementary or antisense strand of the viral genome. According to recent studies, transient overexpression of ToLCNDV-NSP induces the hypersensitive response (HR) or necrosis in *N. tabacum* plant. Similarly, ToLCNDV-BV1 N- or C-terminal mutant failed to show HR in *S. lycopersicum*. It is therefore postulated that the NSP may act as a symptom determinant and could serve as a silencing suppressor. Movement protein encoded by BC1 plays an important role in the intercellular movement of the viral DNA through the plasmodesmata.

In the context of the Indian subcontinent, one characteristic feature of ToLCNDV is its frequent association with either alpha- or betasatellites, despite being a bipartite begomovirus. Betasatellites are approximate ~1.5 kb in size and are half of the size of their helper DNA-A viruses, which are necessary for their replication and transportation. These molecules contain a satellite conserved region (SCR), an adenine-rich (A-rich) region and a single ORF in complementary sense strand named β C1. β C1 is a multi-functional protein that acts as a symptom-determinant in the host plant during begomoviral pathogenesis. Another category of ssDNA satellite molecules of ~1.5 kb was also identified and named as alphasatellites, which require a helper begomoviruses for inter- and intra-cellular movements, but is independent for replication. Alphasatellites were shown to be responsible for the reduced accumulation of betasatellite levels in begomovirus-beta satellite complexes and to act as a suppressor of RNA silencing.

Under laboratory conditions, maintenance of betasatellites (Cotton leaf curl Multan betasatellite, CLCuMB and Radish leaf curl betasatellite, RaLCuB) by ToLCNDV has been well documented. *N. benthamiana* plants agro-inoculated with ToLCNDV DNA-A contained lesser viral titer as compare to when co-inoculated with either CLCuMB or RaLCuB. Co-inoculated plants also exhibited severe symptoms indicating role of betasatellites in leaf curl disease development. However, it is important to note that DNA-A level was higher in plants co-inoculated with DNA-B as compared with betasatellites co-inoculated plants, which emphasized the important role of DNA-B over betasatellite in facilitating the movement of DNA-A. Furthermore, reduced level of CLCuMB in plants co-inoculated along with DNA-A and DNA-B as compared to DNA-A + CLCuMB indicates the role of DNA-B in modulating betasatellite levels in infected plants, precise mechanism of which is not yet known. Likewise, three alphasatellites were found to be associated with ToLCNDV in chilli, cotton and cucurbits plants during mixed infection (Table 1). However, the role of alphasatellites on ToLCNDV pathogenesis is yet to be ascertained (Fig. 3).

Synergism and Genetic Changes Leading to Begomovirus Emergence and Evolution

There are plenty of non-genetic based factors such as mixed infection, synergism, vector and vegetative dispersal, cultural practices and above all human movement of geminiviruses in new environments, which are promoting geminivirus infection in non-cultivated plants and/or other new cultivated crops. However, among all plant viruses, geminiviruses have a particularly high emergence potential because they can use each one or the combination of 4 important genetic-based mechanisms to rapidly change their genetic make-up and capacity, and thereby evolve within a concurrently changing environment to thrive and survive. In addition to having access to these molecular mechanisms, geminiviruses also enjoy a very high rate of multiplication, producing billions of new genomic molecules in a very short span of time, therefore they can afford a large fraction of non-functional molecules to explore many avenues at the same time. These four mechanisms are: the mutations, the recombination process, the pseudo-recombination and the capture of satellites.

Recent studies suggest that geminiviruses and nanoviruses, do show a high level of genetic variation and nucleotide substitution rate (10^{-3} – 10^{-4} substitution/site/year), and the mechanism is quite similar to RNA viruses. It is believed that these plant DNA viruses may use low-fidelity DNA polymerases for their genome replication which might lead to sequence variation in their genome fairly rapidly. Although, other factors which cause spontaneous mutation in the begomoviral genome (oxidation, transition, transversion, deamination and methylation of nitrogenous bases) might lead to their genetic variability and variation within the species as well.

Recombination plays an important role during geminivirus evolution and it can be achieved by two mechanisms, the rolling circle replication and the recombination dependent replication, which leads to the formation of complete new viral genomes. Studies suggest that inter and intra-species recombination events are very frequent in the geminivirus genomes. The impact of recombination has been studied and documented for several viruses such as the Tomato yellow leaf curl viruses, the East African cassava mosaic

Table 1 Association of ToLCNDV components with other begomoviruses and DNA satellites, in natural and experimental hosts

ToLCNDV genome	Accession # viral genomic components and satellites associated with ToLCNDV			Associated BGs	Associated betasatellite	Access# Betasatellite	Natural host	Experimental host
	DNA-A	DNA-B	β -satellite					
ToLCNDV-A, B and β -sat	HM345979	HM803117	EF095958		CLCuMuB	LN845926, HG934394	<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-A, B and β -sat	EF068246	EF408038	EF068245 AY438562		LuLDB	KR957354	<i>S. lycopersicum</i>	
ToLCNDV-A, B	LN845962	LN845963		CLCKoV	CLCuMuB	AY083590	<i>G. hirsutum</i>	
ToLCNDV-A, B	EF043231	EF043232			PaLCuB	EF043234	<i>S. lycopersicum</i> <i>S. tuberosum</i>	
ToLCNDV-A, B	HM989845	HM989846			LuLDB	AY728262	<i>L. cylindrica</i> <i>M. charantia</i> <i>C. sativus</i>	<i>Nb</i>
	AY9399 26	AY9399 24				DQ020491	<i>L. siceraria</i>	
ToLCNDV-A, B	U15015	U15017			RaLCB	EF175734	<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-A, B	HQ141673	HQ141674		ToLCPaIV-A ToLCPaIV-B			<i>S. lycopersicum</i>	<i>SI</i>
ToLCNDV-A, B	U15015	U15017		ToLCPaIV-A			<i>C. melo</i>	<i>Nb</i>
ToLCNDV-A, B	HQ141673	HQ141674		ToLCPaIV-A			<i>S. lycopersicum</i>	<i>SI, Cs</i>
ToLCNDV-A, B	HQ141673	HQ141674		ToLCPaIV-B			<i>S. lycopersicum</i>	<i>SI, Cs</i>
ToLCNDV-A, B	U15015	U15017		RaLCV			<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-A, B	U15015	U15017		CYVMV			<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-A	U15015	U15017		CYVMV	CYVMB	FJ593630	<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-A, B	U15015	U15017		ToLCRnV			<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-A, B	U15015	U15017			ToLCRnB	GQ994096	<i>S. lycopersicum</i>	<i>Nb</i>
ToLCNDV-A	KU196750			ChLCV-A ToLCGV-B ToLCGV-B	ToLCBDB	KR957354	<i>C. annuum</i> <i>C. chinense</i> <i>S. lycopersicum</i>	<i>Nb, Ca</i> <i>Nb, SI, Nt</i>
ToLCNDV-A	U15015						<i>C. annuum</i>	
ToLCNDV-A	KP235541				CroYVMB	FJ593630	<i>A. esculentus</i> <i>G. hirsutum</i>	
ToLCNDV-B		GU112083		BYVMV			<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-B		LN845955		ToLCGV-A			<i>S. lycopersicum</i>	<i>Nb, Nt, Ca</i>
ToLCNDV-B		HQ141674		ToLCGV-A			<i>S. lycopersicum</i>	<i>Nb, Nt, SI</i>
ToLCNDV-B		U15017		PepLCLV			<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-B		U15017		ToLCGV-A			<i>S. lycopersicum</i>	<i>Nb, SI</i>
ToLCNDV-B		HM803117		ToLCGV-A	TYLCTHB	HM989847	<i>C. annuum</i>	<i>Nb, Nt, Ca</i>
ToLCNDV-B		U15017		PepLCLV	ChLCB	AM279673	<i>C. annuum</i>	
ToLCNDV-B		JN663871		ChLCV-PK	ToLCRnB	JN663872	<i>C. annuum</i>	
ToLCNDV-B		JN663848		ToLCV-Ban	ToLCBDB	JN663849	<i>C. annuum</i>	
ToLCNDV-B		KP235543		PePLCV-Pal	CroYVMB	JN663878	<i>C. annuum</i>	
ToLCNDV-B		JN663867		ToLCV-Kar	ToLCBDB	JN663868	<i>C. annuum</i>	
ToLCNDV-B		KF471060		ChILCV	RaLCB	JN663873	<i>C. annuum</i>	

Three alphasatellites have been reported to be associated with ToLCNDV DNA-A and other DNA-B in chili leaf curl disease (ChLCD) infected samples. These are: ChILCA (KF471051) with ToLCNDV (DNA-A, KU196750), associated with ChLCD from New Delhi; ChILCA (KF471052) with ToLCNDV (DNA-A and B, respectively, KP235540 and JN663871), associated with ChLCD, Himachal Pradesh and ChILCA (KF471046) with ToLCNDV (DNA-B, JN663848) associated with ChLCD, from Karnataka. In cucurbit plants, association of cucurbit yellow mosaic alphasatellite (CYMA, KT948075), ToLCNDV (DNA-A, HM98984; DNA-B, AM286435) and Papaya leaf curl betasatellite, KT948074). In addition, CLCuMA (LN845922, LN845923, LN845924, HG934391, HG934392, and HG934393) was also found to be associated with ToLCNDV A in cotton leaf curl disease (CuLCD) plants.

viruses and the Cotton leaf curl viruses. Although there has not been such detailed recombination analysis done for ToLCNDV, there is no reason to believe that ToLCNDV, does not benefit from the recombination strategy that many begomoviruses explored extensively. Furthermore, there are reports suggesting that ToLCNDV serves as a major parent donor helping in the production of other recombinant begomoviruses, such as Radish leaf curl virus (RaLCuV), Tomato leaf curl Rajasthan virus (ToLCRaV), Papaya leaf crumple virus (PaLCrV), Tomato leaf curl Patna virus (ToLCPaV) and Tomato leaf curl Palampur virus (ToLCPaV).

Interestingly pseudo-recombination or, re-assortment of genomic components of DNA-A and DNA-B for the bipartite begomoviruses results in exchange of the genetic materials similar but different to what's happening during intra-molecular recombination. It is similar because the bipartite begomovirus unit will acquire new proteins with new functions, but it is different because it is the whole set of two proteins acquired instantly and together with the DNA-B re-assortment. There is no creation of de novo chimeric proteins. This event has been reported to be more frequent for Old World begomoviruses, like that of between African cassava mosaic virus (ACMV) and East African Cassava mosaic virus (EACMV) but it has also been reported for New World begomoviruses such as the Pepper golden mosaic virus (PePGMV) and Pepper Huasteco yellow vein virus (PHYVV).

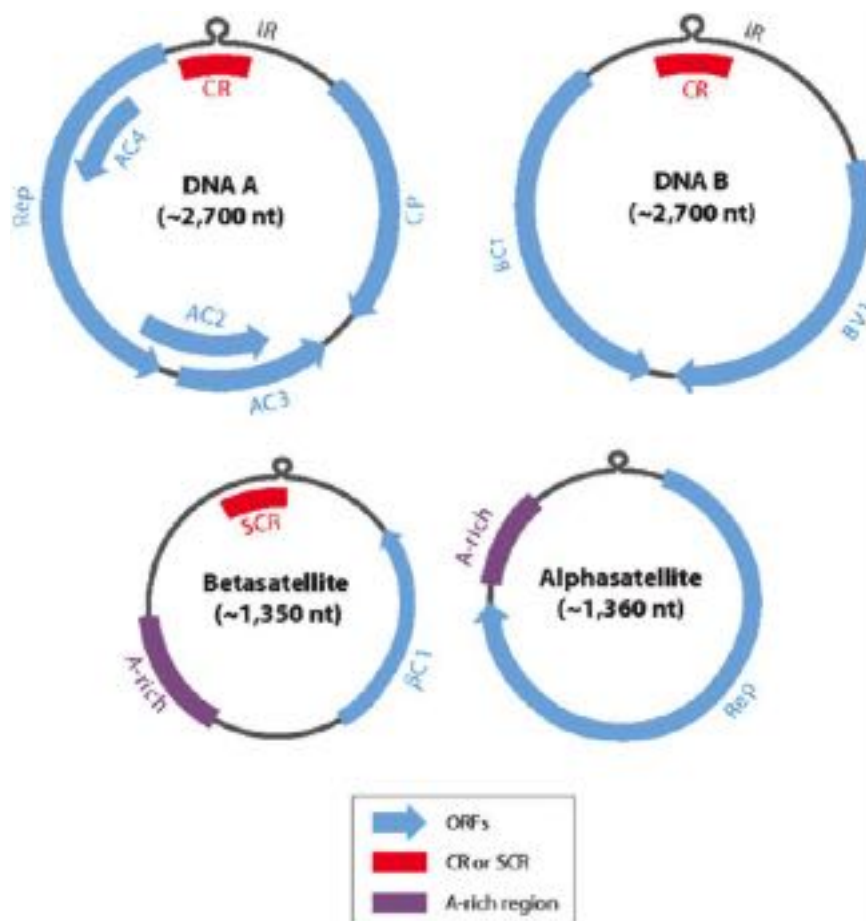


Fig. 3 Genome organization of ToLCNDV (DNA-A and DNA-B), a betasatellite and an alphasatellite. ToLCNDV is a bipartite geminivirus with components known as DNA-A and DNA-B. The proteins encoded by the components are indicated as the replication associated protein (Rep), the transcriptional activator protein (TrAP), the replication enhancer protein (REn), the C4 protein (C4), the coat protein (CP) and the V2 protein (V2) on the genomes of the DNA A component. The DNA-B component encodes the nuclear shuttle protein (NSP) and the movement protein (MP). A sequence conserved between the DNA-A and DNA-B components of bipartite begomoviruses is known as the conserved region (CR). The betasatellites have a single gene ($\beta C1$) in the complementary-sense and a region of sequence highly conserved between all betasatellites [known as the satellite conserved region (SCR)] and a region of sequence rich in adenine (A-rich). The alphasatellites have a single large protein (Rep) and also contain an A-rich sequence. For each component the conserved hairpin structure, containing the nonanucleotide sequence TAATATTAC (TAGTATTAC for most alphasatellites) within the loop structure, is shown at position zero.

Besides, in a laboratory-based experiment done with a pair of highly virulent pseudo-recombinant viruses between the DNA-A from Tomato leaf curl New Delhi virus (ToLCNDV, # U15015) and the DNA-B from Tomato leaf curl Gujarat virus DNA-B (ToLCGuV, # AY190291) it was shown that re-assortment was perfectly functional. This genetic re-assortment resulted in causing a very severe disease on tomato and other solanaceous crops. In those pseudo-recombinant infected plants, the levels of DNA-A and DNA-B were several-fold enhanced. Subsequently, these two molecules were observed under natural condition in the diverse agro-climatic regions of India in chilli, tomato and other solanaceous crops (Table 1). Furthermore, this association also resulted in the breakdown of geminivirus resistance in chilli indicating the relevance of fitness of viral molecules in the host plant. This super compatible integration of two different molecules in an innate condition may affect the host range and the epidemiology of the new virus disease.

The fourth possibility for genetic changes for begomoviruses is the capture of ssDNA satellites that are particularly abundant and variable (Table 1).

Host Range of ToLCNDV

The members of the ToLCNDV species share today a broad spectrum of host range, and they are known to infect 13 families of vegetable crops, flowering plants and weeds (Table 2). It is not known if the host range of ToLCNDV has effectively changed over

Table 2 List of plants, and their families, infected with ToLCNDV

Family	Crops
<i>Acanthaceae</i>	<i>Crossandra infundibuliformis</i>
<i>Amaranthaceae</i>	<i>Chenopodium album</i>
	<i>Convolvulus arvensis</i>
<i>Apiaceae</i>	<i>Daucus carota</i>
<i>Apocynaceae</i>	<i>Calotropisprocera</i>
	<i>Catharanthus roseus</i>
<i>Asteraceae</i>	<i>Calypocarpus vialis</i>
	<i>Chrysanthemum indicum</i>
	<i>Parthenium hysterophorus</i>
<i>Caricaceae</i>	<i>Carica papaya</i>
<i>Cucurbitaceae</i>	<i>Benincasa hispida</i>
	<i>Citrullus lanatus</i>
	<i>Coccinia grandis</i>
	<i>Cucumis melo</i>
	<i>Cucumis anguinus</i>
	<i>Cucumis sativus</i>
	<i>Cucurbita moschata</i>
	<i>Cucurbita pepo</i>
	<i>Lagenaria siceraria</i>
	<i>Luffa aegyptiaca</i>
	<i>Luffa acutangula</i>
	<i>Luffa cylindrical</i>
	<i>Momordica charantia</i>
	<i>Momordica dioica</i>
	<i>Sechiumedule</i>
<i>Fabaceae</i>	<i>Glycine max</i>
	<i>Lens culinaris</i>
<i>Malvaceae</i>	<i>Abelmoschusesculentus</i>
	<i>Gossypium hirsutum</i>
<i>Papaveraceae</i>	<i>Papaver somniferum</i>
<i>Phyllanthaceae</i>	<i>Sauropus androgynus</i>
<i>Polygonaceae</i>	<i>Rumex dentatus</i>
<i>Solanaceae</i>	<i>Capsicum annuum</i>
	<i>Lycopersicon esculentum</i>
	<i>Nicotiana tabacum</i>
	<i>Solanum lycopersicum</i>
	<i>Solanum nigrum</i>
	<i>Solanum tuberosum</i>

time, but we can only take note that this virus has been isolated over time from a growing number of natural hosts, cotton being the last one. Because of the great capacity for this virus to capture other B components in mixed infections, acquiring thereby the capacity to move more efficiently both components in these hosts, or capturing satellites that could play an important role to suppress gene silencing of the hosts, or a combination of both strategies, it is possible that ToLCNDV has *de facto* expanded its host range.

It is also noticeable, that the new epidemic of ToLCNDV in Europe is mostly on cucurbitaceae, where the symptoms are very severe compared to tomato (Fig. 1). The sequence of both DNA-A and DNA-B components, although having both a large fraction of the typical ToLCNDV components, are novel and unique in the whole ToLCNDV molecular diversity for both components. A detail recombination study is needed to better characterize these molecules and understand the epidemiology of this new ToLCNDV strain.

Interestingly, co-infection of ToLCNDV is not limited to bipartite begomoviruses but it is also observed in association with monopartite begomoviruses such as Bhendi yellow vein mosaic virus (BYVMV), Chilli leaf curl virus (ChiLCV), Croton yellow vein mosaic virus (CYVMV), Squash leaf curl China virus (SLCCNV), Tomato leaf curl Ranchi virus (ToLCRnV) and Cotton leaf curl Kokhran virus (CLCKoV). A recent study demonstrates the presence of the monopartite begomovirus (ChiLCV) along with ToLCNDV (a bipartite begomovirus) and a betasatellite leading to a severe chilli leaf curl disease (ChiLCD). This interaction between the betasatellite and the geminiviruses led to a synergistic interaction benefitting the geminiviruses, and as a consequence enhancing the disease severity on chilli. Most surprisingly, in Pakistan, ToLCNDV (LN845962) was recently detected in cotton (*Gossypium hirsutum*) plants showing a typical cotton leaf curl disease. The association of ToLCNDV together with Cotton leaf curl Kokhran virus strain Burewala (CLCKoV-Bur) and Cotton leaf curl Multan betasatellite (CLCMuB) also indicate the possibility of synergistic interactions that might lead to the breakdown of CLCuD resistance in cotton.

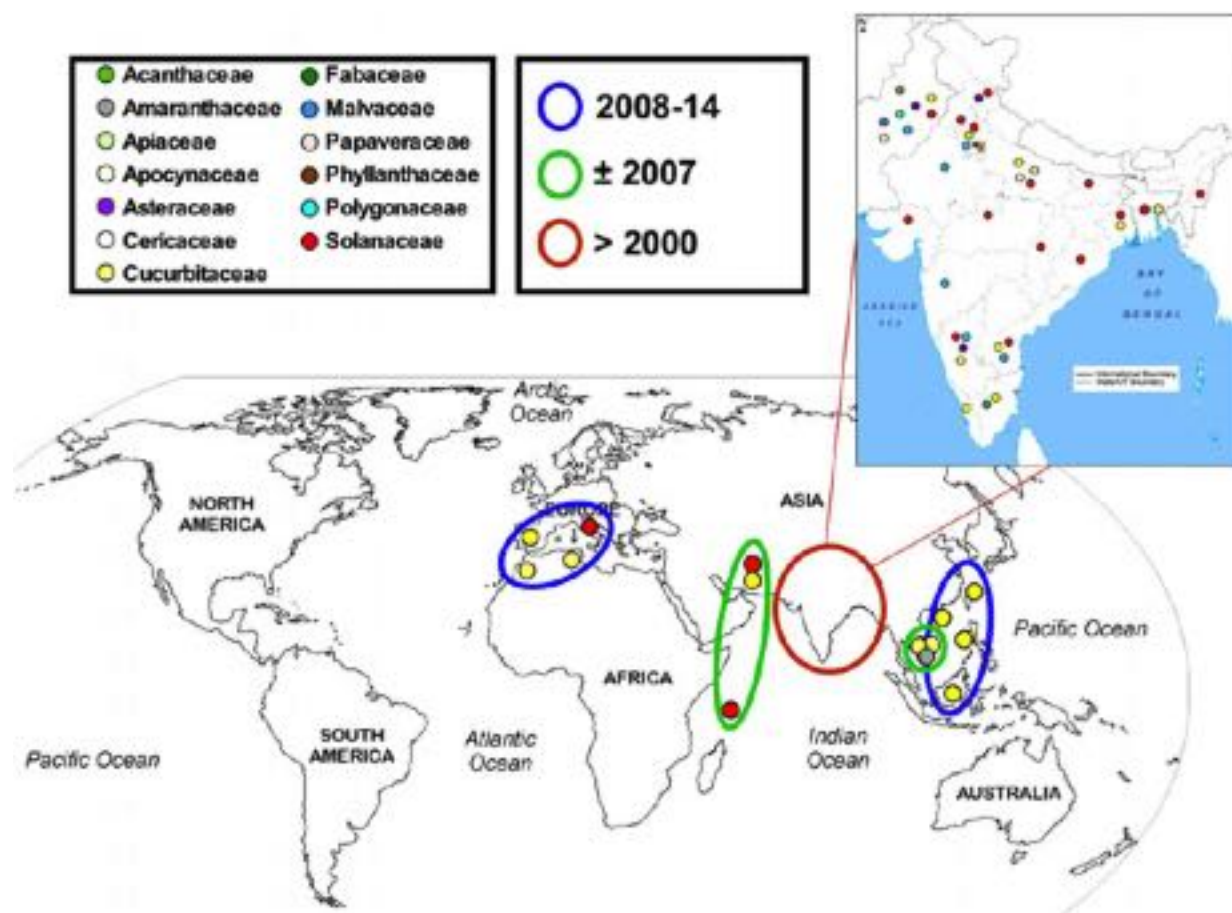


Fig. 4 Geographical distribution of ToLCNDV isolates in the world. ToLCNDV has been isolated from thirteen different plant families that are shown in different colors. The dates for the world spread are roughly indicated with the color circles.

Transmission and Epidemiology

The occurrence of ToLCNDV in several countries in Asia, South East Asia, Middle East or Europe indicates that it is not limited by any geographical barrier and it must be efficiently transmitted by a range of whiteflies wherever the virus is present (Fig. 4).

ToLCNDV is transmitted by whiteflies (*Bemisia tabaci*, Genn., *Aleyrodidae*) under natural conditions and the role of seed transmission has not been evidenced yet but is most probably not relevant. Several biotypes/species of *B. tabaci* are reported to be transmitting begomoviruses to the Solanaceae, Cucurbitaceae, Malvaceae and many other crops. Middle-East Asia Minor I (MEAM1) (formally known as biotype B) is the most predominant whitefly population in India, especially in the southern part. Initially, the role of MEAM1 whiteflies as vector for transmission of begomoviruses onto solanaceous crops was not considered to be common, despite its presence along with other whitefly biotypes. There is a possibility that the invasive species population favors the emergence of a particular breed because of climate change. The other invasive species of whitefly known as Mediterranean 1 (MED1) (formerly known as biotype Q) was recently identified in Spain to cause ToLCNDV epidemics. This could be an alarming situation and a strategic plan to combat the disease needs to be developed. ToLCNDV has emerged as an Old World bipartite begomovirus that threatens several economically important crops in the world. The presence of ToLCNDV has been reported from thirteen different families of crop plants (Table 2) affected worldwide.

The epidemic outbreaks of ToLCNDV in India might be dependent on several factors such as the use of susceptible varieties for crop cultivation, the co-existence of different agro-climatic zones favoring the vector population survival, average rainfall, and the almost constant availability of the whitefly vector population that are swapping from one crop to another. Surprisingly, ToLCNDV is now more widespread in members of the Cucurbitaceae family as compared to Solanaceae or any other crops. In India, farmers generally cultivate cucurbits (bottle gourd, bitter gourd, muskmelon, squash, pumpkin and water melon) in almost every climatic season. The population of whitefly vector remains available throughout the year and the alternate hosts could act as a reservoir for ToLCNDV. In addition, the solanaceous crops such as tomato and chilli, are mostly grown in summer seasons. The higher temperature, relative humidity and availability of various whitefly biotypes/species may therefore provide an advantage for the insect vector to grow more and transmit the virus to several plant species.

Virus-Host Interactions

The geminiviruses utilize the host machinery to effectively function within the host. The plant-virus interplay differs along the spectrum of antagonism and mutualism. Begomoviruses have relatively small genomes that encode multitasking proteins required for the viral replication, encapsidation, transport, viral suppressor of RNA silencing and vector transmission. Many studies suggest plant viruses interact with many host factors and also manipulate several basic mechanisms of the hosts to perform their essential functions. As a counter defense mechanism, plants have developed several antiviral approaches. In a broader perspective, plants utilize resistance gene (R-gene) mediated defense and small non-coding RNA based defense mechanisms against the viral pathogen. In the next few paragraphs these two defense approaches have been discussed.

Small Non-Coding RNAs

Geminivirus infection triggers the production of two major classes of small non-coding RNA (sRNA): -small interfering RNAs (siRNAs) and -micro RNAs (miRNAs). These sRNA plays vital roles in the RNA silencing pathways and also control the expression of host genes. As an example, tomato plants infected with ToLCNDV showed upregulation of conserved miR319 and miR172 sRNA expression. These 21 nt long miRNAs play crucial role in regulating antiviral mechanism. Microarray data from ToLCNDV's agro-inoculated tomato plants suggested the differential regulation of miR398, miR399, miR162 and miR168. The miR162 and miR168 are known to regulate the expression of Dicer-like protein-1 (DCL1) and Argonaute 1 (AGO1) protein, respectively. The miR398 is involved in the abiotic-stress related response in plants, and it was observed that the level was higher when infected with the ToLCNDV. These data suggest ToLCNDV infection results in the transcriptional activation and suppression of specific miRNAs in host plant. Therefore, begomoviruses use this distressed gene expression to their purpose, resulting in building a suitable environment for survival inside the host cells. The functional characterization of siRNAs in susceptible, tolerant, and resistant tomato varieties during ToLCNDV infection also has been studied. In addition, the virus-specific siRNAs (Rep specific siRNAs) accumulation at 21 days post inoculation was higher in the tolerant varieties as compared to the susceptible varieties, as showed by northern blot hybridization experiments. The presence of virus specific siRNA were also noted in the systemic tomato leaves. These data further suggest that the virus specific siRNA might be inducing the gene silencing pathways in the tolerant varieties and it failed or is compromised to do so in the susceptible varieties. The higher ToLCNDV DNA-A level was effectively detected in the susceptible variety by Southern experiment.

Counter Defense RNA Silencing

Plants have an RNA silencing mechanism to control virus infection and as counter defense, viruses have viral suppressor of RNA silencing (VSRs) which regulate pathogenesis in the host cell. A bipartite begomovirus, such as ToLCNDV, encodes different proteins such as AC2/TraP, AC4, AV2/Pre-coat, and BV1/NSP to attenuate the gene silencing defense mechanism of the host by different means. ToLCNDV-AV2 inhibits the symptom remission on *N. tabacum* plants by blocking the RNA-Dependent RNA Polymerase 1 from *N. tabacum* (NtRDR1). NtRDR1 mediated antiviral silencing and the spread of virus specific siRNAs. Mutation in the AC4 ORF of ToLCNDV N50H (Asn position 50) led to the changes in β -sheet structure in SH3 domain resulting in affecting its ability to suppress the RNA silencing pathway. Although it was observed that these mutations do not affect the physical interaction of AC4 with AGO4. Hence, ToLCNDV-AC4 might have a role in targeting SH3 domain containing proteins to further modulate the begomoviral pathogenesis in a host cell.

Basal Level of Defense

It has been shown that ToLCNDV infections (in tomato and potato) can potentially induce the expression of plant-innate immunity and hormone signaling genes including ethylene-signaling genes such as ethylene response factor-4 (ERF4), WRKY, and NAC transcription factors. The ToLCNDV infection prompts the hypersensitivity response (HR) in the infected cells (in tomato plants) as a result of accumulation of higher reactive oxygen species (ROS) level. Oxidative burst at the infected sites eventually kills both the host and the invaded pathogen. Fortunately, during the course of evolution, a very efficient ROS-scavenging mechanism has also evolved in the plants.

Resistance Gene or R-Gene Mediated Defense

Plants have different strategies to combat the pathogens. One of the mechanisms of active plant defense is the resistance gene or R-gene mediated response. The R-gene can identify a specific effector molecule which is produced by the pathogen, and then initiate the HR, resulting in programmed cell death to inhibit the further amplification or spread of the pathogen. Over many decades, breeders have manipulated a likely host-target-loci that is liable with natural sources of resistance for viral infection. In some solanaceous (tomato) and cucurbitaceous plant genotypes, different resistance or tolerant genes have been identified to develop resistance against the ToLCNDV. Particularly, in tomato (wild type species) six TYLCV resistance loci (*Ty-1*, *Ty-2*, *Ty-3*, *Ty-4*, *Ty-5*, and *Ty-6*) have been identified, in which *Ty-2* alone failed to show resistance against ToLCNDV. Interestingly, the combination of *Ty-2*

(in homozygous or heterozygous condition) and *Ty-3* (heterozygous) through marker-assisted selection and breeding, is showing enhanced resistance against ToLCNDV and associated betasatellite. The loci *Ty-1* and *Ty-3* encode RNA-dependent RNA polymerase (RDR6). RDR6 is involved in the amplification of virus-specific siRNA which leads to the activation of the RNA silencing pathway. Similarly, ToLCNDV infection in solanaceous crop (*Capsicum annuum*, *N. tabacum*, *N. benthamiana* and *S. lycopersicum*) showed differential expression of defense-related genes such as Suppressor of Gene Silencing 3 (SGS3), RNA-Dependent RNA Polymerase 6 (RDR6) (genes which are involved in PTGS pathways), and the resistance gene (R-gene) such as nt-binding site and leucine-rich repeat (NBS-LRR), Lipid transfer protein (LTP) and pathogenesis-related (PR) genes.

A recent study suggests that some of the cultivated species of cucurbits (genus, *C. pepo*), wild type germplasm lines and F1-hybrid (subspecies *C. pepo* x *C. pepo* and subspecies *C. pepo* x *C. ovifera*) accessions exhibit resistance against ToLCNDV. Furthermore, the cucurbits accessions used for this study showed very mild symptoms, resulting in low viral titer even after several rounds of mechanical and whitefly inoculations. Similarly, inheritance study of *Luffa cylindrica* (sponge gourd) conducted during both rainy and spring seasons revealed presence of two sequence-related amplified polymorphism (SRAP) markers closely associated with dominant monogenic inheritance of ToLCNDV-resistance. Therefore these studies further indicate the existence of a genetic basis of resistance in cucurbit plants against ToLCNDV. The continuous evolution of geminiviruses and augmented host range, call for the need of resistance gene that should be investigated continuously to confer useful resistance against ToLCD to benefit farmers.

Diagnosis and Management

It is essential to confirm that a given disease is actually caused by a geminivirus, before developing a management program. Currently, the ToLCNDV detection is done by using polymerase chain reaction (PCR)-based amplification of the viral genome and loop mediated isothermal amplification assay (LAMP). Another recent technique is developed is the use of gold nanoparticle conjugated to a probe as biosensors. Furthermore, there are other methods of ToLCNDV detection such as coat protein specific antibody and non-radioactive nucleic acids spot hybridization (NASH) assay.

Because of its ability to infect and adapt in various plants, ToLCNDV is exceptionally tricky to manage and control. Some progress on the identification of resistance genes both in tomato and cucurbits is providing hopes that ToLCNDV could be controlled efficiently. Genetic engineering technique has been a key promising approach against the pathogen in a wide variety of crops around the world. One study suggests targeting of AV1/AV2 common overlapping regions by artificial miRNA (amiRNA), conferring tolerance against the ToLCNDV in transgenic tomato crops. However, as begomoviruses are continuously evolving it is very difficult to generate stable resistant or tolerant transgenic lines. The siRNA profiling in tomato and other solanaceous crop gives an idea about the novel resistance host genes. Their characterization can further provide a potential target to develop virus resistance/tolerance plants. Identifying the immune sources and implementing the specific gene in a targeted location to address the challenges of the specific virus, might lead to the sustainable management of the disease.

Various other approaches such as sense/antisense technology, RNAi techniques and application of double-stranded RNA (dsRNA) can also be used as potential method to develop resistance against the plant affecting begomoviruses. Recent advancement in the genetic manipulation of the DNA leads to the development of another system termed as clustered regularly interspaced short palindromic repeats-CRISPR associated 9 (CRISPER-Cas9) and is used to engineer broad-spectrum geminivirus resistance in plants. Other management practices like whitefly vector control, breeding of resistance varieties and micro-propagation of virus-free plants could be the alternative and complementary approaches to control ToLCNDV.

See also: Betasatellites and Deltasatellites (*Tolecusatellitidae*). Emerging Geminiviruses (*Geminiviridae*). Geminiviruses (*Geminiviridae*). Plant Resistance to Geminiviruses. Tomato Yellow Leaf Curl Viruses (*Geminiviridae*)

Further Reading

- Basu, S., Kushwaha, N., Singh, A.K., *et al.*, 2018. Dynamics of a geminivirus-encoded pre-coat protein and host RNA-dependent RNA polymerase 1 in regulating symptom recovery in tobacco. *Journal of Experimental Botany* 69 (8), 2085–2102.
- Chatterji, A., Padidam, M., Beachy, R.N., Fauquet, C.M., 1999. Identification of replication specificity determinations in two strains of tomato leaf curl virus from New Delhi. *Journal of Virology* 73 (7), 5481–5489.
- Fortes, I.M., Sánchez-Campos, S., Fiallo-Olivé, E., *et al.*, 2016. A novel strain of Tomato leaf curl New Delhi virus has spread to the Mediterranean basin. *Viruses* 8 (11), 307.
- García-Arenal, F., Zerbini, F.M., 2019. Life on the edge: Geminiviruses at the interface between crops and wild plant hosts. *Annual Review of Virology* 6, 411–433.
- Hussain, M., Mansoor, S., Iram, S., Fatima, A.N., Zafar, Y., 2005. The nuclear shuttle protein of Tomato leaf curl New Delhi virus is a pathogenicity determinant. *Journal of Virology* 79 (7), 4434–4439.
- Islam, S., Anilabh, D.M., Verma, M., *et al.*, 2011. Screening of *Luffa cylindrica* Roem. for resistance against Tomato leaf curl New Delhi virus, inheritance of resistance, and identification of SRAP markers linked to the single dominant resistance gene. *The Journal of Horticultural Science and Biotechnology* 86 (6), 661–667.
- Kumar, R.V., Singh, A.K., Singh, A.K., *et al.*, 2015. Complexity of begomovirus and betasatellite populations associated with chilli leaf curl disease in India. *Journal of General Virology* 96 (10), 3143–3158.
- Kushwaha, N., Singh, A., Basu, S., Chakraborty, S., 2015. Differential response of diverse solanaceous hosts to Tomato leaf curl New Delhi virus infection indicates coordinated action of NBS-LRR and RNAi-mediated host defense. *Archives of Virology* 160 (6), 1499–1509.

- Malik, A.H., Briddon, R.W., Mansoor, S., 2011. Infectious clones of Tomato leaf curl Palampur virus with a defective DNA-B and their pseudo-recombination with Tomato leaf curl New Delhi virus. *Virology Journal* 8 (1), 173.
- Moriones, E., Praveen, S., Chakraborty, S., 2017. Tomato leaf curl New Delhi virus: An emerging virus complex threatening vegetable and fiber crops. *Viruses* 9 (10), 264.
- Naqvi, A.R., Haq, Q.M., Mukherjee, S.K., 2010. MicroRNA profiling of Tomato leaf curl New Delhi virus (ToLCNDV) infected tomato leaves indicates that deregulation of mir159/319 and mir172 might be linked with leaf curl disease. *Virology journal* 7 (1), 281.
- Padidam, M., Beachy, R.N., Fauquet, C.M., 1995. Tomato leaf curl geminivirus from India has a bipartite genome and coat protein is not essential for infectivity. *Journal of General Virology* 76, 25–35.
- Pratap, D., Kashikar, A.R., Mukherjee, S.K., 2011. Molecular characterization and infectivity of a Tomato leaf curl New Delhi virus variant associated with newly emerging yellow mosaic disease of eggplant in India. *Virology Journal* 8 (1), 305.
- Sáez, C., Martínez, C., Ferriol, M., *et al.*, 2019. Resistance to Tomato leaf curl New Delhi virus in *Cucurbita* spp. *Annals of Applied Biology* 169 (1), 91–105.
- Sahu, P.P., Rai, N.K., Chakraborty, S., *et al.*, 2010. Tomato cultivar tolerant to Tomato leaf curl New Delhi virus infection induces virus-specific siRNA accumulation and defense associated host gene expression. *Molecular Plant Pathology* 11 (4), 531–544.
- Singh, A.K., Kushwaha, N., Chakraborty, S., 2016. Synergistic interaction among begomoviruses leads to the suppression of host defense-related gene expression and breakdown of resistance in chilli. *Applied Microbiology and Biotechnology* 100 (9), 4035–4049.
- Van, Vu T., Choudhury, N.R., Mukherjee, S.K., 2013. Transgenic tomato plants expressing artificial microRNAs for silencing the pre-coat and coat proteins of a begomovirus, Tomato leaf curl New Delhi virus, show tolerance to virus infection. *Virus Research* 172 (1–2), 35–45.

Tomato Spotted Wilt Virus (*Tospoviridae*)

Hanu R Pappu, Washington State University, Pullman, WA, United States

Anna E Whitfield, North Carolina State University, Raleigh, NC, United States

Athos S de Oliveira, University of Brasília, Brasília, Brazil

© 2021 Elsevier Ltd. All rights reserved.

This is an update of H.R. Pappu, Tomato Spotted Wilt Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00648-8](https://doi.org/10.1016/B978-012374410-4.00648-8).

Nomenclature

aa Amino acid(s)

Avr Avirulence

CC Coiled coil

ELISA Enzyme-linked immunological assays

GRSV Groundnut ringspot virus

HR Hypersensitive response

IDM Integrated disease management

JA Jasmonate

LRR Leucine-rich repeat

MP Movement protein

N Nucleocapsid

NB Nucleotide-binding

NLR Nucleotide-binding, leucine-rich repeat receptor

ORF Open reading frame

PB Processing body

RB Resistance-breaking

RdRp RNA-dependent RNA polymerase

RI Resistance-inducing

RNAi RNA interference

RNP Ribonucleoprotein

SA Salicylic acid

SD Solanaceae domain

TEM Transmission electron microscopy

TSG Tubular salivary glands

TSWV Tomato spotted wilt virus

Glossary

Avirulence determinant Microorganism molecular patterns, e.g., a viral protein, that activate plant resistance proteins, inhibiting infection/virulence.

Circulative, propagative transmission Virus transmission process in which the virus circulates

through insect body and replicates in various tissue systems.

Processing bodies RNA-protein complexes located in the cytoplasm where RNA decay pathways occur.

Classification

Tomato spotted wilt orthotospovirus is the founding virus species in the genus *Orthotospovirus*, family *Tospoviridae*, order *Bunyavirales*. Tomato spotted wilt virus (TSWV) was first described in 1915 and was recognized as a bunyavirus in the 1990s based on virion morphology, genome sequence, and replication in the insect vector, thrips. TSWV was the first phytovirus classified as a bunyavirus. At present, there are 26 accepted species and 5 tentative species in the genus *Orthotospovirus* and there are three other phytovirus genera (*Tenuivirus*, *Emaravirus*, and *Coguvirus*) within different families in the order *Bunyavirales*.

Virion Structure

TSWV has the classic bunyavirus particle structure that is composed of a membrane that encapsulates the three genome segments (L, M, and S) that are coated in nucleocapsid (N) proteins. The genome is composed of negative-sense RNAs and the 5' and 3' ends of the viral RNAs are highly conserved and complementary. The complementary ends of the genome segments enable the formation of panhandle structures and these sequences serve as promoters for transcription of genes and genome replication. The N proteins arrange as asymmetric trimeric rings, creating an inner cavity where the viral RNAs bind. The virus particle also contains several copies of the L protein, an RNA-dependent RNA polymerase (RdRp), that is essential for infectivity of the virus upon delivery into the host plant or vector. The two glycoproteins, Gn and Gc, decorate the surface of the virion and are the attachment and entry proteins involved in the infection cycle in the insect vector (**Fig. 1**).

Among tospoviral proteins, the glycoproteins, and the N protein of TSWV and groundnut ringspot virus (GRSV) and the NSs have been predicted by folding prediction, and the native structure of the N protein was determined by crystallization, and the RNA binding sites were mapped in the simulated 3D structure. The N protein structure is composed of three main parts: N-terminal arm, C-terminal arm, and core domain. To form the trimeric ring, the N- and C-terminal arms of each N protein

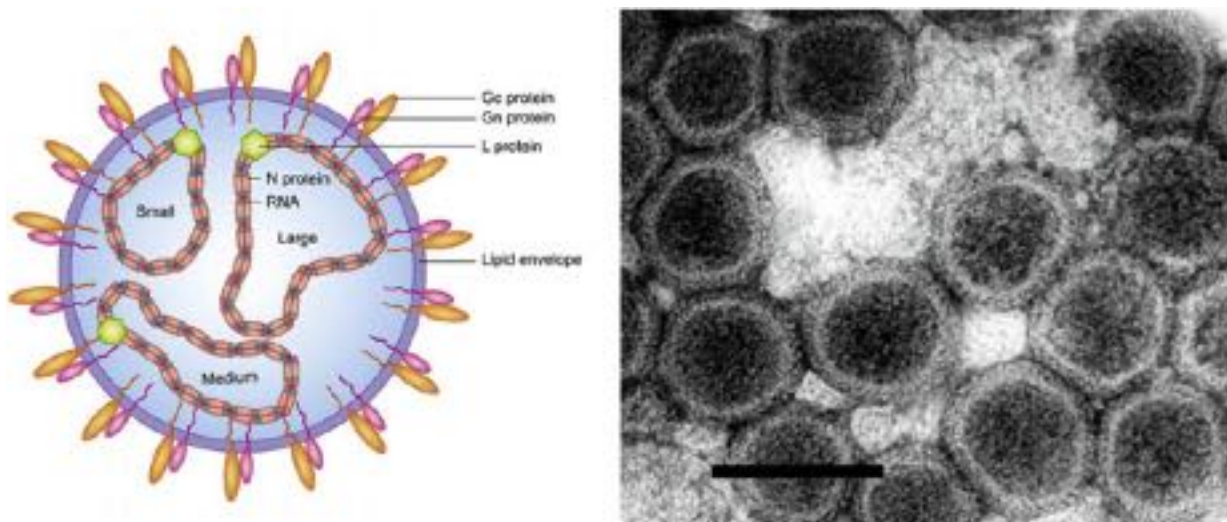


Fig. 1 (Left) Virus particle diagram of TSWV. (Right) Transmission electron microscope picture of TSWV particles. The bar represents 100 nm. Virion cartoon used with permission from Rotenberg, D., Jacobson, A.L., Schneeweis, D.J., Whitfield, A.E., 2015. Thrips transmission of tospoviruses. *Current Opinion in Virology* 15, 80–89. Photo courtesy of Dr Jan Van Lent.

interact with the core domain of the other two adjacent proteins. The folding pattern consisted of a globular core shape containing two β -sheets and 12–17 α -helix, and two terminal chains corresponding to the N and the C termini exposed on the surface of the protein. The N protein structure has important aa for binding to RNA and protein folding such as R94/R95 and K183/Y184, which are mapped onto a charged surface cleft of the three-dimensional structure. Interestingly R95 is conserved among all orthotospoviruses and is consistently located in a α -helix.

Genome

The three negative-sense RNAs of TSWV encode six proteins on five genes, and some functions of each TSWV protein have been elucidated. The L protein (RdRp) works as a transcriptase, synthesizing mRNAs to be translated into functional proteins, and as a replicase, synthesizing new genomic RNAs which are then covered by N proteins. Interestingly, the L protein also snatches 5' capped ends from host mRNAs to cap-protect the viral mRNAs, the so-called 'cap-snatching' phenomenon. The L protein prefers cap-donors with multiple base complementarity to the viral RNA templates and seems to select host mRNAs which are destined for decapping in processing bodies (PBs). It has been shown that alterations in protein components of PBs, where mRNA decay pathways occur, interfere on TSWV cap-snatching and infection in *Arabidopsis* mutants.

The nucleoprotein (N) functions as a protective layer encapsidating the three viral genomic RNA segments, but also plays an important role in viral RNA transcription and replication. The N in the RNP complex interacts with the cytoplasmic tails of the glycoproteins which collectively form the mature virion or can interact with the NSm protein for cell to cell movement through the plasmodesmata.

The glycoproteins of TSWV are the viral determinants of insect transmission. The glycoproteins are encoded on single ORF and post-translationally cleaved into mature independent Gn and Gc proteins. The glycoproteins associate together and move from the ER to the Golgi where virus budding occurs. The glycoproteins in the Golgi form the external core that encapsidates the RNPs through interactions with the N protein and the cytoplasmic tails of the glycoproteins. The resulting double enveloped virion interacts with the endoplasmic reticulum cisternae, clustering several virions on vesicles. Gn also functions as an attachment protein for virus acquisition and is a determinant for virus transmissibility by thrips, while Gc is a viral fusion-like protein. The upstream sequence of Gn/Gc was found to contain promoter elements that are light and auxin inducible when expressed in *Arabidopsis*.

The NSm, a non-structural protein, serves as the cell to cell and long-distance movement protein (MP), through interaction with the N protein in the RNP complex. For cell to cell movement, the NSm forms tubular structures and modifies the size exclusion of the plasmodesmata with neighboring cells. Additionally, the NSm is associated with symptom expression and was shown to be an avirulence (avr) determinant in resistant plant varieties harboring the gene *Su-5b*.

The NSs, another non-structural protein, has been shown to serve as suppressor of antiviral RNA silencing in host plants and insect vector cells. The NSs also plays a major role in TSWV transmission by suppressing the biosynthesis of volatile monoterpenes, plant repellents of thrips. The NSs targets MYC transcription factors, regulators of the jasmonate(JA)-signaling pathway that controls the expression of *terpene synthase* genes. Therefore, TSWV-infected plants become more attractive to thrips, which favors virus acquisition by these insect vectors. The NSs is also associated with symptom development and supports systemic viral movement by interfering with antiviral RNA silencing in plants.

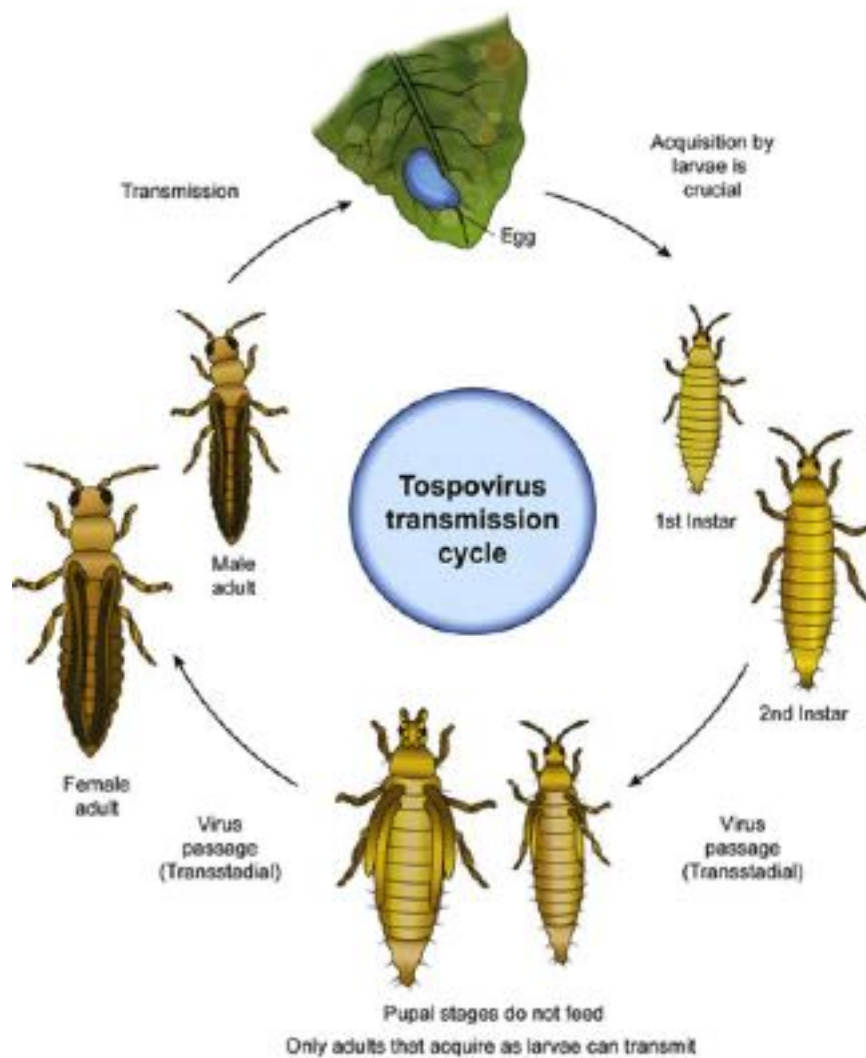


Fig. 2 The tomato spotted wilt virus transmission cycle. Thrips eggs are oviposited into plant tissue and within a few days the first instar larvae emerge. Virus acquisition occurs during the larval stages after which the virus persists in the vector and is passed transstadially to the adult. The pupal insects may occur in the soil or leaf tissue and these stages are nonfeeding. Adults emerge and may disperse widely. Only adult thrips (male and female) that acquired the virus during their larval stages can transmit tospoviruses. Used with permission from Rotenberg, D., Jacobson, A.L., Schneweis, D.J., Whitfield, A.E., 2015. Thrips transmission of tospoviruses. *Current Opinion in Virology* 15, 80–89.

Transmission

Insects in the order Thysanoptera are commonly called ‘thrips’ and they transmit TSWV and other orthotospoviruses in a circulative, propagative manner. There are over 7000 thrips species, but only nine species have been shown to transmit TSWV (*Frankliniella bispinosa*, *F. cephalica*, *F. fusca*, *F. gemina*, *F. intonsa*, *F. occidentalis*, *F. schultzei*, *Thrips setosus*, *T. tabaci*). All thrips species that transmit plant viruses are classified in the family Thripidae and they are phytophagous insects that feed in a piercing-sucking manner. Thrips feeding behaviors and high reproductive capacity make them efficient vectors of TSWV.

TSWV has a unique insect developmental-stage-dependent transmission cycle (Fig. 2). Virus is acquired when larval thrips feed on TSWV-infected tissue. As thrips mature, they lose their ability to become efficient vectors and adult thrips cannot effectively acquire virus and become virus transmitters. In the larval thrips that acquire TSWV, the virus disseminates from the initial site of infection, the midgut, to the salivary gland tissues. The pupal stages do not feed and thus do not acquire or inoculate virus. Both male and female adults are efficient virus transmitters and they can effectively disperse the virus over short and long distances. The small size of thrips makes them an efficient transport mechanism for viruses and thrips are often found in shipments of diverse types of plant materials.

The TSWV dissemination pathway in thrips has been described in detail. During larval thrips acquisition, the virus is ingested and enters the anterior midgut epithelial cells. This is the primary site of virus replication and virus infection appears to spread to surrounding cells. From the gut epithelial cells, virions traverse the basal membrane and basal labyrinth to then infect the muscle

cells surrounding the gut. Confocal and TEM immunolabeling studies have shown that the cells of the tubular salivary glands (TSG) contain detectable amounts of TSWV. The TSGs connect the midgut to the principal salivary glands and may serve as a conduit for virus movement between insect tissue systems. Alternatively, virus may move between gut tissues and principal salivary glands during the larval stage when these tissues are in direct contact. The principal salivary glands are the final site for virus replication prior to delivery of virus with feeding and salivation into the plant host. The extent of principal salivary gland infection has been shown to be associated with transmission efficiency. The virus abundance in larval thrips is positively correlated with infection status as adults and transmission efficiency. Interestingly, TSWV has not been found in the hemolymph of infected insects but this could be due to technical difficulties associated with virus detection in this small insect.

Epidemiology

Numerous experiments have confirmed that TSWV has a beneficial effect on *F. occidentalis*. The infection of several different plant hosts has resulted in faster developmental times for this species of vector. In addition, female thrips are attracted to the virus infected plants and this synergistic interaction facilitates transmission of the virus because larval thrips acquire TSWV. *F. occidentalis* is the most widespread vector of TSWV, but other vectors have regional importance and more varied interactions with the virus. For example, in the South-Eastern U.S. *F. fusca* is an important vector of TSWV to peanut but insect exposure to virus-infected peanuts negatively impacted vector development time. In addition to variation between TSWV-vector species interactions, there can be virus and insect isolate differences in interactions. Examination of TSWV-infected plants shows that they have higher levels of free amino acids and may have modified volatile organic compound profiles.

Pathogenesis

TSWV infects over a thousand plant species and infection results in significant yield losses for many agronomically important crops. Typical virus symptoms include stunting, wilting, and necrotic or chlorotic concentric spots on susceptible plants (Fig. 3). Many plants have an age-dependent susceptibility to TSWV. Physiological measurements of TSWV-infected peanut revealed that photosynthesis, transpiration, and water-use efficiency were significantly decreased in virus-affected plants. In TSWV-infected plants, the phytohormone salicylic acid is present at higher levels and SA-related genes are up-regulated indicating a plant defense response to virus infection. Several TSWV resistance genes have been identified and studied at the ecological and molecular level. In plants and insects, the TSWV NSs protein is a suppressor of RNA silencing and therefore a counter defense to one of the major antiviral responses in eukaryotic hosts. Also, the NSs of TSWV has been reported as an avr determinant in pepper (*Capsicum annuum*). This suggests an additional role for the NSs besides the well-defined RNAi suppressor activity. Likewise, it was recently suggested that the NSs of TSWV has a role in translation, and persistent infection and transmission by *F. occidentalis*. Furthermore, accumulation of high levels of NSs in the salivary glands of thrips could be indicative of NSs protein being co-injected into plants during thrips feeding. The NSs protein interferes with the RNA silencing response in plants by affinity for different types of dsRNA molecules.

Progress has been made in characterizing the thrips response to virus infection. There are diverse thrips-TSWV interactions that have been documented ranging from symbiotic to pathogenic and for the best characterized system, TSWV and *F. occidentalis*, the interaction is beneficial to the insect in most experiments. Gene expression analysis of western flower thrips, *F. occidentalis*, response to TSWV revealed that there is a developmental-stage-dependent transcriptome response to virus infection. The number and identity of responsive transcripts was unique to each thrips stage of development (larval, pupal, and adult). Only three sequences were differential in all life stages tested. The diverse response to virus infection over thrips development is likely due to the unique tissue tropism over the lifespan of the insect and dynamics of the virus replication cycle at the distinct infection stages. Similar experiments with TSWV-infected *F. fusca* yielded a greater number of differential transcripts but again a unique suite of responsive genes in larval, pupal, and adult thrips with only eight being responsive at all developmental stages.

Diagnosis

TSWV causes distinct foliar symptoms on some susceptible crops such as groundnut (peanut), potato, tobacco, and tomato that are of diagnostic value. Confirmatory, laboratory-based assays include ELISA-based serological test using antibodies specific to the N protein. The genome of several isolates of TSWV was sequenced and RT-PCR-based molecular detection methods are available.

Prevention

In case of perennials and vegetatively propagated plants and crops, virus can spread from infected mother plants through cuttings and tubers. Use of virus-free material is very effective in preventing disease outbreaks. TSWV is not seed-transmitted, and the primary source of infection in annual crops is mainly due to feeding by viruliferous thrips. Reducing yield losses due to TSWV infection is mainly through an integrated disease management (IDM) strategy that involves vector management by biological and

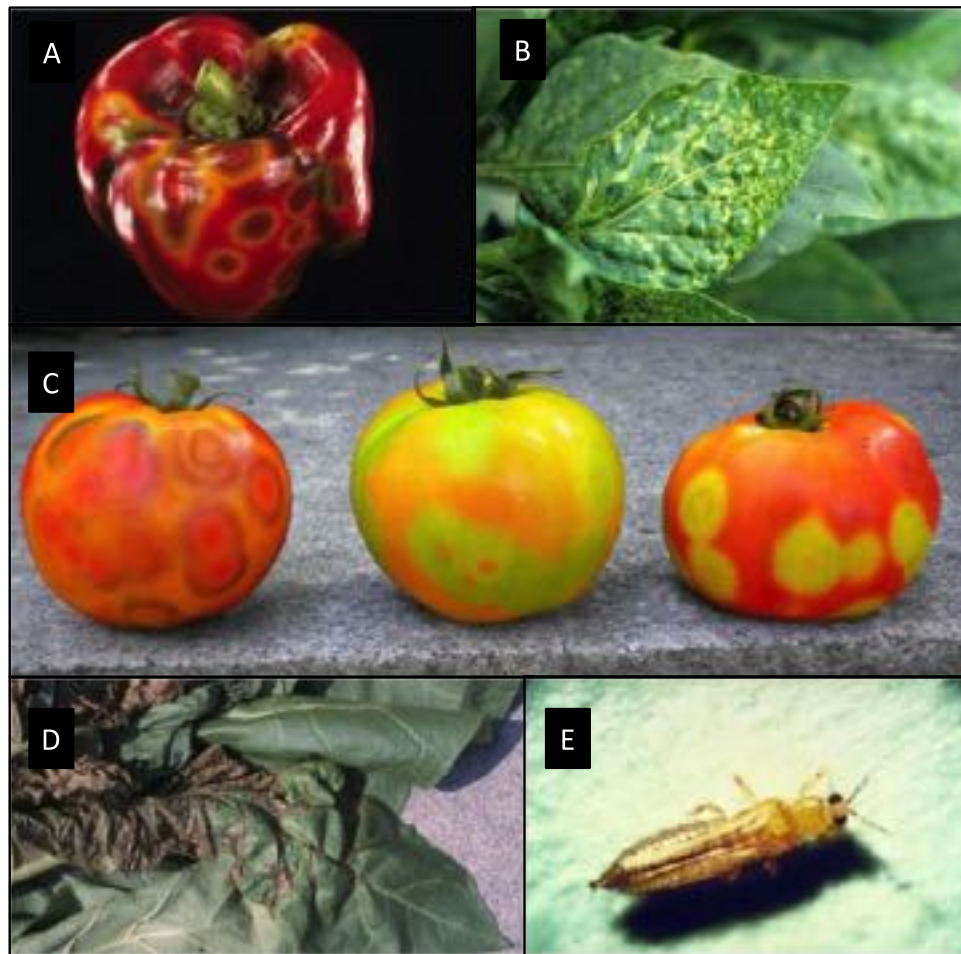


Fig. 3 Symptoms caused by TSWV in plants and the thrips *Frankliniella occidentalis*. Pepper fruit (A) and (B) leaf photographed by Gerald Holmes, California Polytechnic State University at San Luis Obispo. (C) Tomatoes photographed by Elizabeth Bush, Virginia Polytechnic Institute and State University (VT). (D) Tobacco leaves photographed by Mary Ann Hansen, VT. (E) *Frankliniella occidentalis* photographed by Jack T. Reed, Mississippi State University.

chemical means, cultural practices, and growing TSWV resistant cultivars. A set of IDM tactics was developed in peanut based on the relative importance of risk factors (TSWV Risk Index) that result in disease epidemics.

Plant Resistance

Several natural resistance sources have been reported to halt TSWV infection. Among them, only the tomato *Sw-5b* gene-mediated resistance has been deeply studied. Originally, the *Sw-5b* gene is from *Solanum peruvianum*, a wild species of tomato naturally occurring in northern Chile and southern Peru. The *Sw-5b* shares homology with other four *Sw-5* paralogs, named *Sw-5a*, *Sw-5c*, *Sw-5d*, and *Sw-5e*, encoding nucleotide-binding (N) leucine-rich repeat (L) receptors (R). The NLR proteins are usually activated by direct or indirect interaction with avr determinants, which eventually result in the appearance of necrotic spots on inoculated leaves, phenomenon known as hypersensitive cell death response (HR). Although many *Sw-5* paralogs and orthologs have been reported, only the *Sw-5b* protein has been proved to activate resistance to orthotospoviruses (Fig. 4). Due to its durability and broad-spectrum resistance, the *Sw-5b* gene has been introgressed into many commercial tomato cultivars, becoming the main natural resistance source against TSWV for this crop. This broad-spectrum resistance is not just observed for TSWV isolates, but also for orthotospoviruses classified in other species phylogenetically-related to TSWV.

The NSm protein is the cognate avr determinant of *Sw-5b*. Single expression of NSm in *Sw-5b*-transformed *N. benthamiana* and in *Sw-5b*-resistant tomato lines triggers HR in agro-infiltrated leaves. The recognition of NSm as avr occurs by its direct association with *Sw-5b*. Two domains of *Sw-5b*, NB-ARC and LRR, directly interact with NSm by recognizing an epitope of 21 aa, which is conserved among TSWV-related orthotospoviruses. Most NLR proteins, however, are thought to indirectly recognize pathogens since they monitor host proteins that are targeted and modified by avr determinants.

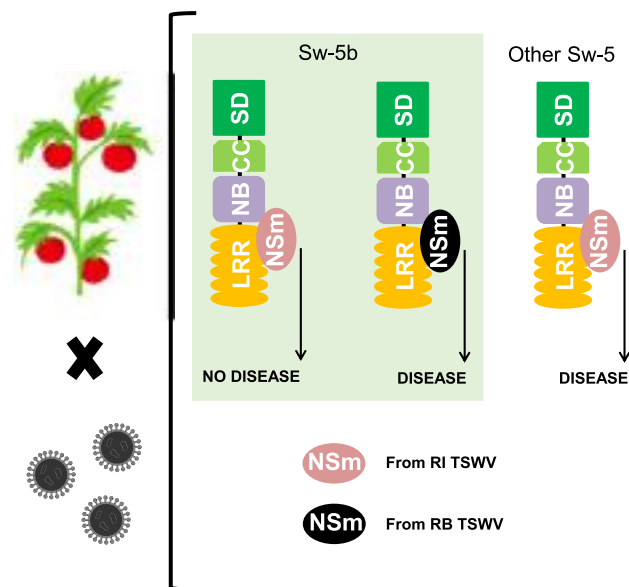


Fig. 4 *Sw-5b*-mediated resistance to TSWV in tomatoes. Although there are several *Sw-5* paralogs in tomato plants bred with *S. peruvianum*, only *Sw-5b* senses an epitope of 21 aa in NSm. This protein-protein interaction activates resistance to TSWV. Some point mutations in NSm, however, can turn resistance-inducing (RI) TSWV isolates into resistance-breaking (RB) isolates. The *Sw-5* proteins contain the domains SD (Solanaceae Domain), CC (Coiled-Coil), NB (Nucleotide-Binding), and LRR (Leucine-Rich-Repeat).

Other resistance sources to TSWV (*Sw-1a*, *Sw-1b*, *Sw-2*, *Sw-3*, *Sw-4*, *Sw-6*, and *Sw-7*), either recessive or dominant, have been reported and introgressed into commercial tomato cultivars. In comparison to *Sw-5b*, they commonly have a narrower-spectrum resistance against TSWV isolates and have not been characterized molecularly. In pepper, the *Tsw* gene activates resistance against TSWV isolates, inducing a clear HR on inoculated leaves. Besides some similarities with *Sw-5b*, the *Tsw*-mediated HR is triggered by the NSs protein of TSWV. Many other plant species and accessions have been reported to be resistant to TSWV isolates but, contrasting with plants harboring *Sw-5b* or *Tsw*, there is none or little genetic and molecular information available.

Overtime, the constant use of a monogenic resistance source on a virus population selects virus mutants that are not sensed by the resistance gene product or that suppress the resistance mechanism. In this context, TSWV resistance-breaking (RB) isolates have been reported for both *Sw-5b* and *Tsw* genes in tomato and pepper growing fields, respectively. Two aa substitutions in NSm proteins, C118Y or T120N, overcome *Sw-5b*-mediated resistance by TSWV isolates. A single mutation in the NSs protein, T104A, overcomes *Tsw*-mediated resistance. Evidences of reassortment events have also been reported for TSWV. This evolutionary mechanism is an additional advantage of TSWV in adapting to new plant hosts and thrips vectors, and from escaping selection pressures as resistance plant genes. The durability of a resistance gene in the field, nevertheless, can be spanned with proper management of resistant cultivars. Actions such as alternating resistant and susceptible cultivars or planting susceptible cultivars alongside production fields may reduce the selection and arising of RB isolates.

Further Reading

- Abudurexiti, A., Adkins, S., Alioto, D., *et al.*, 2019. Taxonomy of the order *Bunyavirales*: Update 2019. *Archives of Virology* 164. doi:10.1007/s00705-019-04253-6.
- Brown, Steve L., Culbreath, A.K., Todd, J.W., *et al.*, 2005. Development of a method of risk assessment to facilitate integrated management of spotted wilt of peanut. *Plant Disease* 89, 348–356.
- Culbreath, A.K., Todd, J.W., Brown, S.L., 2003. Epidemiology and management of Tomato spotted wilt virus in peanut. *Annual Review of Phytopathology* 2003 (41), 53–75.
- de Oliveira, A.S., Boiteux, L.S., Kormelink, R., Resende, R.O., 2018. The *Sw-5* gene cluster: tomato breeding and research toward orthotospovirus disease control. *Frontiers Plant Science* 9, 1055. doi:10.3389/fpls.2018.01055.
- Ma, X., Zhou, Y., Moffett, P., 2019. Alterations in cellular RNA decapping dynamics affect Tomato spotted wilt virus (TSWV) cap snatching and infection in *Arabidopsis*. *New Phytologist*. doi:10.1111/nph.16049.
- Montero-Astúa, M., Rotenberg, D., Leach-Kieffaber, A., *et al.*, 2014. Disruption of vector transmission by a plant-expressed viral glycoprotein. *Molecular Plant Microbe Interactions* 27 (3), 296–304. doi:10.1094/MPMI-09-13-0287-FI.
- Montero-Astúa, M., Stafford-Banks, C.A., Badillo-Vargas, I.E., *et al.*, 2016. Chapter 20: Tospovirus–thrips biology. In: Brown, J.K. (Ed.), *Vector-Mediated Transmission of Plant Pathogens*. APS Press. doi:10.1094/9780890545355.020.
- Montero-Astúa, Mauricio, Ullman, Diane E., Whitfield, Anna E., 2016. Salivary gland morphology, tissue tropism and the progression of tospovirus infection in *Frankliniella occidentalis*. *Virology* 493, 39–51. doi:10.1016/j.virol.2016.03.003.
- Olaya, C., Adhikari, B., Raikhy, G., Cheng, J., Pappu, H.R., 2019. Identification and localization of *Tospoviridae* family-wide conserved residues in 3D models of the nucleocapsid and the silencing suppressor proteins. *Virology Journal* 16, 7. doi:10.1186/s12985-018-1106-4.
- Rothenberg, D., Whitfield, A.E., 2018. Molecular interactions between tospoviruses and thrips vectors. *Current Opinion in Virology* 33, 191–197. doi:10.1016/j.coviro.2018.11.007.

- Schneweis, D.J., Whitfield, A.E., Rotenberg, D., 2017. Thrips developmental stage-specific transcriptome response to Tomato spotted wilt virus during the virus infection cycle in *Frankliniella occidentalis*, the primary vector. *Virology* 500, 226–237. doi:10.1016/j.virol.2016.10.009.
- Srinivasan, R., Abney, M., Culbreath, A., *et al.*, 2017. Three decades of managing Tomato spotted wilt virus in peanut in southeastern United States. *Virus Research* 241, 203–212.
- Zhai, Y., Peng, H., Neff, M.M., Pappu, H.R., 2019. Putative auxin and light responsive promoter elements from the Tomato spotted wilt virus genome, when expressed as cDNA, are functional in *Arabidopsis*. *Frontiers in Plant Science* 10. doi:10.3389/fpls.2019.00804.

Relevant Websites

- <https://site.extension.uga.edu/mitchellag/files/2016/06/Peanut-Rx-and-Various-Fungicide-Programs.pdf>
Peanut Rx and Various Fungicide Programs.
University of Georgia.
- <http://www.ento.csiro.au/thysanoptera/worldthrips.php>
Thysanoptera (Thrips) World Checklist.
CSIRO Entomology.
- <http://climate.ncsu.edu/thrips>
Tobacco Thrips Flight and TSWV Intensity Predictor
North Carolina Climate Office.
- <https://www.hgsc.bcm.edu/arthropods/western-flower-thrips-genome-project>
Western Flower Thrips Genome Project.
BCM-HGSC.

Tomato Yellow Leaf Curl Viruses (*Geminiviridae*)

Henryk Czosnek, The Hebrew University of Jerusalem, Rehovot, Israel

© 2021 Elsevier Ltd. All rights reserved.

This is an update of H. Czosnek, Tomato Yellow Leaf Curl Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00717-2](https://doi.org/10.1016/B978-012374410-4.00717-2).

Glossary

Agroinoculation, agroinfection An alternative route for viral infection; the viral genome is cloned, usually as a head-to-tail dimer, in the T-DNA of *Agrobacterium tumefaciens* and is delivered to plants by inoculation.

Introgression Incorporation of chromosomal segments of wild tomato species in the domesticated tomato by crosses and selection; introgression lines are used to localize genes to the tomato chromosomes

Rolling circle Mechanism of replication followed by many viral DNA and by begomoviruses in particular.

Viruliferous whitefly Insect that has acquired virus from an infected plant and is ready to infect other host plants.

Whitefly Insect that belongs to the order *Homoptera*; they cause damage to plants by feeding and by vectoring plant viruses.

Introduction

Tomato yellow leaf curl virus (TYLCV) was coined in the late 1950s to describe a virus infecting tomato plants in the Middle East. TYLCV causes one of the most devastating diseases of tomato worldwide. The virus (genus *Begomovirus*, family *Geminiviridae*) is transmitted by the whitefly *Bemisia tabaci* (family *Aleyrodidae*, order *Hemiptera*) in circulative manner. The virus was isolated, cloned and sequenced in the late 1980s. It has a circular single-stranded DNA genomic molecule of 2787 nt encapsidated in a 20 × 30 nm geminate particle. The viral strand encodes four proteins while the viral complementary strand encodes two. The TYLCV genome replicates in host cell nuclei according to the rolling circle model. From the early 1960s, TYLCV has quickly spread from the Eastern Mediterranean Basin to the entire Middle East, Central and Southeast Asia, China, North and West Africa, Southeast Europe, the Caribbean islands, Southeast USA, the Southern Indian Ocean islands, Australia and Japan. Global warming may be instrumental to the expansion of the whitefly-transmitted virus to presently temperate regions. Sequence comparisons of TYLCV isolates from different regions revealed that the name TYLCV encompasses a complex of closely, as well as distantly, related begomovirus species affecting tomato. Over the last 40 years, TYLCVs have been the subject of numerous investigations, including epidemiology, phylogenetics, plant-virus-vector interactions, diagnosis, and breeding for resistance.

Classification and Taxonomy

TYLCV is a member of the genus *Begomovirus* of the family *Geminiviridae*. Begomoviruses are transmitted by the whitefly *Bemisia tabaci* (family *Aleyrodidae*, order *Hemiptera*). *B. tabaci* comprises many species (or biotypes), distinguished by their DNA sequence. Among them, the Middle East – Asia Minor 1 (MEAM1, or B) and the Mediterranean (MED, or Q) species are efficient vectors of TYLCV. TYLCV was isolated and its genome sequenced in the late 1980s. Begomoviruses have a genome either split between two circular single-stranded DNA (ssDNA) molecules of approximately 2700 nt each, named DNA A and DNA B (bipartite), or with a single genomic DNA A-like molecule (monopartite). TYLCVs are monopartite, except for two viruses from Thailand.

Sequence comparisons revealed that the name TYLCV encompasses a complex of closely, as well as distantly, related begomoviruses affecting tomato. Guidelines for the classification of begomoviruses state that genome-wide pairwise identities of 91% and 94% constitute the demarcation threshold between species and strains, respectively. Accordingly, TYLCVs are classified into several species and strains, listed in [Table 1](#) (updated as of January 2015). This classification is rendered more complicated by the discovery that recombination between different TYLCVs happens relatively frequently (see below). Additional begomoviruses that infect tomato cultures are described in other chapters. These viruses clearly differ from TYLCVs in the symptoms they induce on tomato, in their host range, and in their nucleotide sequence.

Virion Structure

Like all geminiviruses, TYLCV has a particle of twinned morphology of approximately 20 × 30 nm in size ([Fig. 1\(a\)](#)). The virus capsid consists of two joined, incomplete icosahedra, with a $T = 1$ surface lattice containing a total of 22 capsomeres each containing five units of a 260 amino acid coat protein (30.3 kDa). The geminate particle contains a single 2787 nt covalently closed genomic circular ssDNA molecule. A model of the structure of Ageratum yellow vein virus, obtained by cryo-EM at 3.3 Å, indicates that the DNA fills the entire available space within the particle.

Table 1 List of Tomato yellow leaf curl virus species, as of January 2015. Species names are shown in bold; isolate names are given in regular font. For species that do not have any known strains, only one isolate is listed, and that isolate is recognized as the “type” isolate. For species that have known strains, one isolate from each strain is shown, and the type isolate is the first one listed. Abbreviations are in the second column and the sequence accession numbers in the third

<i>TYLCV species and isolates</i>	<i>Abbreviation of name of virus isolates</i>	<i>Accession number</i>
<i>Tomato yellow leaf curl Axarquia virus</i>		
Tomato yellow leaf curl Axarquia virus – [Spain-Algarrobo-ES-mh800-2000]	TYLCAxV-[ES-Alg-ES-mh800-00]	AY227892
<i>Tomato yellow leaf curl China virus</i>		
Tomato yellow leaf curl China virus – Baoshan1 [China-Yunnan 10-Tobacco-2000]	TYLCCNV-BS1[CN-Yn10-Tob-00]	AJ319675
Tomato yellow leaf curl China virus – [China-Yunnan 25-Tomato-2000]	TYLCCNV-[CN-Yn25-Tom-00]	AJ457985
Tomato yellow leaf curl China virus – Baoshan3 [China-Yunnan 278-Malvastrum-2007]	TYLCCNV-BS3[CN-Yn278-Mal07]	AM980509
Tomato yellow leaf curl China virus – Bean [China-Yunnan-Bean-2004]	TYLCCNV-Bea[CN-Yn-Bea-04]	DQ256460
Tomato yellow leaf curl China virus – Dali [China-Yunnan 5-Tobacco-1999]	TYLCCNV-DL[CN-Yn5-Tob-99]	AJ319674
Tomato yellow leaf curl China virus – Datura [China-Yunnan 72- Datura-2005]	TYLCCNV-Dat[CN-Yn72-Dat-05]	EF011559
Tomato yellow leaf curl China virus – Honghe [China-Guangxi]	TYLCCNV-HH[CN-Gx]	AF311734
<i>Tomato yellow leaf curl Indonesia virus</i>		
Tomato yellow leaf curl Indonesia virus – [Indonesia-Lembang-2005]	TYLCIDV-[ID-Lem-05]	AF189018
<i>Tomato yellow leaf curl Kanchanaburi virus</i>		
Tomato yellow leaf curl Kanchanaburi virus – [Thailand-Kanchanaburi 1-2001]	TYLCKaV-[TH-Kan1-01]	AF511528 AF511529
<i>Tomato yellow leaf curl Malaga virus</i>		
Tomato yellow leaf curl Malaga virus – [Spain-421-1999]	TYLCMaV-[ES-421-99]	AF271234
<i>Tomato yellow leaf curl Mali virus</i>		
Tomato yellow leaf curl Mali virus – Ethiopia [Ethiopia-Melkassa-2005]	TYLCMLV-ET[ET-Mel-05]	DQ358913
Tomato yellow leaf curl Mali virus – Mali [Mali-2003]	TYLCMLV-ML[ML-03]	AY502934
<i>Tomato yellow leaf curl Sardinia virus</i>		
Tomato yellow leaf curl Sardinia virus – [Italy-Sardinia-1988]	TYLCSV-[IT-Sar-88]	X61153
<i>Tomato yellow leaf curl Thailand virus</i>		
Tomato yellow leaf curl Thailand virus – A [Thailand-1]	TYLCTHV-A[TH-1]	X63015 × 63016
Tomato yellow leaf curl Thailand virus – B [Thailand-Chiang Mai]	TYLCTHV-B[TH-ChMai]	AY514630 AY514633
Tomato yellow leaf curl Thailand virus – C [China-Yunnan 72-2002]	TYLCTHV-C[CN-Yn72-02]	AJ495812
Tomato yellow leaf curl Thailand virus – D [Myanmar-Yangon-1999]	TYLCTHV-D[MY-Yan-99]	AF206674
Tomato yellow leaf curl Thailand virus – E [Thailand-Sakon Nakhon]	TYLCTHV-E[TH-SaNa]	AY514632 AY514635
<i>Tomato yellow leaf curl virus</i>		
Tomato yellow leaf curl virus – [Israel-Rehovot-1986]	TYLCV-[IL-Reo-86]	X15656
Tomato yellow leaf curl virus – Boushehr [Iran-Genaveh 29-2006]	TYLCV-Bou[IR-Gen29-06]	GU076454
Tomato yellow leaf curl virus – Iran [Iran-Irانشahr-1998]	TYLCV-IR[IR-Ira-98]	AJ132711
Tomato yellow leaf curl virus – Kahnoo [Iran-Kahnooj-2007]	TYLCV-Kah[IR-Kah-07]	EU635776
Tomato yellow leaf curl virus – Kerman [Iran-Hormozgan 32-2006]	TYLCV-Ker[IR-Hor32-06]	GU076442
Tomato yellow leaf curl virus – Mild [Israel-1993]	TYLCV-Mld[IL-93]	X76319
Tomato yellow leaf curl virus – Oman [Oman-Al-batinah 22-2005]	TYLCV-OM[OM-Alb22-05]	FJ956700
<i>Tomato yellow leaf curl Yunnan virus</i>		
Tomato yellow leaf curl Yunnan virus – [China-YN2013-2011]	TYLCYnV-[CN-YN2013-11]	KC686705

Note: Adapted from Brown, J.K., Zerbini, F.M., Navas-Castillo, J., *et al.*, 2015. Revision of Begomovirus taxonomy based on pairwise sequence comparisons. Archives of Virology 160, 1593–1619.

Genome

TYLCVs have a monopartite genome, except for two viruses from Thailand (Tomato yellow leaf curl Thailand virus – TYLCTHV, and Tomato yellow leaf curl Kanchanaburi virus – TYLCKaV), which have bipartite genomes.

Monopartite TYLCVs

A schematic drawing of the monopartite TYLCV genome is shown in Fig. 2. The TYLCV DNA-A-like viral genome encodes two large open reading frames (ORF) on the viral strand (V1 and V2), and four on the complementary strand (C1-C4). A 313 nt long intergenic region (IR) contains a 29 nt-long stem-loop structure with the conserved nanonucleotide TAAATATTAC, which is the origin of replication (Ori) of the virus. The IR also contains the promoters of the V1, V2, C1 and C4 genes.

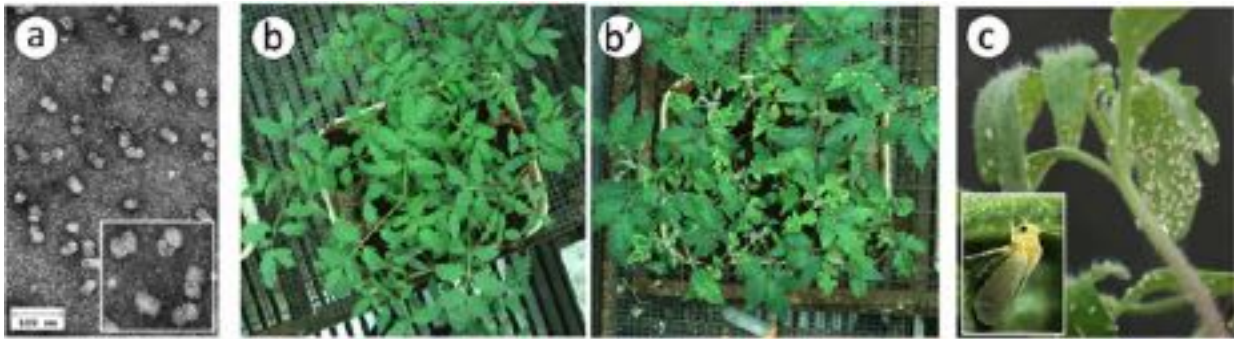


Fig. 1 (a) Electron micrograph of TYLCV particles with twinned morphology of approximately 20×30 nm in size. (b) Non-infected and (b') TYLCV-infected tomato plant with young leaves showing typical yellowing and curling. (c) Tomato leaflet infested with *B. tabaci* whiteflies; insert: close up of adult whitefly.

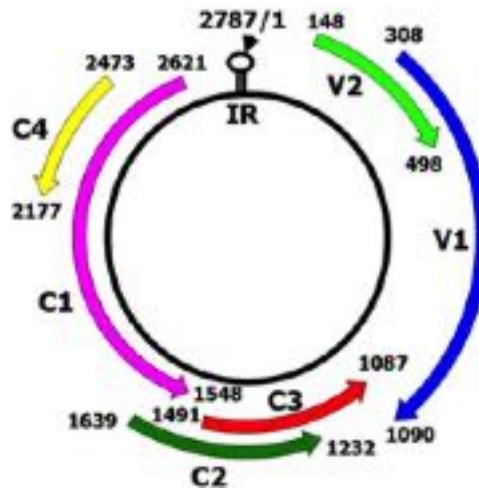


Fig. 2 Genome organization of Tomato yellow leaf curl virus. The single-stranded virion DNA comprises 2787 nt. Open reading frames (ORFs) of virion-sense and complementary-sense strand polarity are designated (V) and (C), respectively. ORFs are represented by an arrow; numbers indicate first and last nucleotide of each ORF. V1 encodes the capsid protein (CP), V2 a movement protein, C1 the replication initiator protein (Rep), C2 a transcriptional activator protein (TrAp), C3 a replication enhancer protein (REn), and C4 a symptom and movement determinant. IR: intergenic region. The conserved inverted repeat flanking the conserved nanonucleotide sequence TAATATTAC is symbolized by a stem-loop; an arrow head indicates the cleaving position of Rep in the TAATATT/AC loop; A at the cutting site (/) is nucleotide number one, by definition.

- V1 encodes the capsid protein (CP). The CP is multifunctional: monomers interact to form the capsid. It has a nuclear localization signal (NLS) and it is able to shuttle the viral genomic DNA in and out of the nucleus. It is essential for infectivity, and it is the only viral protein recognized by the insect vector.
- V2 has properties analogous to those of a movement protein.
- C1 encodes the replication initiator protein (Rep). The functional protein is an oligomer. Rep recognizes the Ori and specifically cleaves the nanonucleotide TAATATTAC between nucleotides 7 and 8. Together with plant host polymerase(s) it initiates viral DNA replication according to the rolling circle model (RCR) (see paragraph "Life cycle").
- C2 encodes a transcription activator protein (TrAp). TrAp enhances transcription of the viral strand. It is able to bind to viral genomic DNA and possesses an NLS. In addition, it may act as silencing suppressors.
- C3 encodes the replication enhancer protein (REn). This protein interacts with itself to form oligomers. REn interacts with Rep and with cell-cycle associated host proteins to increase viral DNA amounts (genomic and double-stranded) in the infected plant.
- C4 encodes a protein not essential for infectivity but contributes to the spread of the virus in the plant and induction of symptoms. The protein may also act as a silencing suppressor.

Bipartite TYLCVs

Bipartite TYLCVs have a second genomic 2.7 kb component (coined DNA-B) with an ORF on the viral strand (BV1, encoding a nuclear shuttle protein) and another on the complementary strand (BC1, encoding a movement protein). Usually the DNA-A is enough to produce disease symptoms.

Satellite DNAs Associated With TYLCVs

Monopartite begomoviruses, among them TYLCVs, are often associated with alphasatellites or betasatellites, two types of 1.3 kb circular ssDNA molecules. The only sequence homology between begomovirus and satellite DNAs is the nanonucleotide within the conserved stem-loop structure of begomovirus genomes required for RCR. Alphasatellites encode a RCR initiator protein (Rep). Betasatellite encodes the multifunctional protein β C1, which may be involved in virus movement, and suppression of gene silencing. β C1 may also enhance symptoms, virus amounts and whitefly transmission. Begomoviruses assist satellites for replication, and encapsidation.

The introduction of betasatellites into western Mediterranean countries (*e.g.*, Jordan, Israel, Egypt, and Arabian Peninsula) where satellites have never been reported before is a potential threat to tomato cultivation. Indeed, it was reported that TYLCV and an associate betasatellite overcome resistance provided by the *Ty-1* gene.

TYLCV Genome Evolution

Epidemiological recordings and phylogenetic analyses suggest that TYLCVs most probably arose in the Middle East between the 1930s and 1950s. The global spread of TYLCV only began in the 1980s after the evolution of the TYLCV and TYLCV-Mild strains. Other centers of TYLCV evolution have been shown to occur, for example in Iran; however because this region is geographically isolated, these TYLCV variants have not spread.

TYLCV species often recombine. For example, TYLCV is a recombinant of TYLCV-Mild and Tomato leaf curl Karnataka virus (Table 1), while other begomoviruses from the Mediterranean basin are recombinants of TYLCV and TYLCV-Mild, and of TYLCV and Tomato yellow leaf curl Sardinia virus (TYLCSV). In Italy, Morocco and Spain, recombinations between TYLCV and TYLCSV occurred in tomato-growing areas. Recombinants were obtained in the laboratory by inoculating tomato plants with infective clones of two TYLCV species. Indeed, upon mixed infections, TYLCV and TYLCSV were detected in the same nucleus, a prerequisite for recombination.

Life Cycle

During feeding on the host plant, whiteflies introduce TYLCV in the phloem with their stylets. The particles penetrate phloem-associated cells. The viral ssDNA genome is uncoated in the cell cytoplasm. Then, it is imported into the nucleus where it is converted to dsDNA (or replicative form RF) thanks to the host cell replication machinery. The dsDNA acts as a substrate for transcription of the viral mRNAs, which are exported to the cytoplasm for translation of the viral proteins required in the nucleus to carry out further replication and transcription. In the nucleus, the RF is duplicated several times according to the RCR model. Several host factors are required for progression and termination of the RCR fork. The IR has a hairpin structure with repeat sequences known as 'iterons'.

Replication is initiated by Rep binding the iterons followed by the specific nicking of the nanonucleotide TAATATT/AC (/: site of nicking). Rep binds covalently at the 5' end of nick, while the 3' end is used for extension of RCR. REN oligomers interact with Rep and enhance its ATPase activity. TYLCV Rep also acts as DNA helicase and as a site-specific type I topoisomerase, unwinding the DNA double helix in the IR for progression of the RCR fork. One round of RCR gives rise to a full viral dsDNA RF along with a genome-long viral ssDNA, which is released following Rep nicking. The new viral strand ssDNA is circularized using the closing activity of Rep and can be further converted to several dsDNA RF that serve as template for transcription/replication. From each RF molecule, several viral ssDNAs are produced. The replication process stops, perhaps with the increasing concentration of viral CP. The ssDNAs are either packaged in virions, or are transported out of the cell as nucleo-protein complexes. With the help of MP encoded by the virus, the transport occurs through the plasmodesmata into the neighboring cells, eventually leading to the long-distance spread within the host plant using its vascular system.

Geminiviruses activate the host cell replication machinery for their own multiplication. Several plant factors interact with viral Rep and Ren to modulate viral DNA replication. The retinoblastoma related (RBR) protein, a plant homolog of the tumor suppressor retinoblastoma (pRb), controls the plant cells entry into S phase. RBR interacts with Rep, inducing the accumulation of both viral and host DNA. Rep is also associated with other host factors such as the Replication factor C (RFC) that helps generate a primer with 3'-OH terminus during initiation of DNA replication.

Posttranslational modification processes like acetylation and sumoylation regulate various cellular processes, such as nuclear-cytosolic transport, transcriptional regulation, apoptosis, protein stability, response to stress, and progression through the cell cycle. As such, they are targets of viruses, which aim at modifying the host environment. The viral Rep interacts with the host sumoylation enzyme, alters sumoylation of the host Proliferative cell nuclear antigen (PCNA) that acts as a cofactor for DNA polymerase δ , which

enhances processing of the leading strand DNA synthesis during replication. Rep activity is reduced following the interaction with PCNA. In contrast, other host proteins may enhance the replication by Rep. For example, the host ssDNA-binding protein, Replication protein A (RPA) 32 kDa subunit (RPA32) directly interacts with the Rep C-terminus. RPA32 modulates the functions of Rep by enhancing its ATPase and decreasing its nicking and closing activities.

The 3-D configuration of the genomic viral DNA may by itself guide the assembly of the particle from a pool of CP monomers. Once in the phloem, TYLCV virions can be picked up by whiteflies, circulate in the insect and may be used to infect new plants during feeding.

Hosts

The domesticated tomato *Solanum lycopersicum* is the primary host of TYLCV. Most of the wild tomato species include accessions that either are symptomless carriers or are immune to the virus. In addition to tomato, several cultivated plants are hosts of TYLCV and present severe symptoms upon whitefly-mediated inoculation: bean (*Phaseolus vulgaris*), petunia (*Petunia hybrida*), and lisianthus (*Eustoma grandiflorum*). Pepper (*Capsicum annuum*) is a symptomless host of the virus. Weeds, such as *Datura stramonium* and *Cynanchum acutum*, present distinct symptoms while others, such as *Malva parviflora*, are symptomless carriers. Plants used to rear whiteflies in the laboratory, such as cotton (*Gossypium hirsutum*) and eggplant (*Solanum melongena*), are immune to TYLCV. Experimental hosts of the virus include jimsonweed (*Datura stramonium*) and Arabidopsis (*Arabidopsis thaliana*). Some plants, such as *Nicotiana benthamiana* and *N. tabacum*, which are recalcitrant to whitefly-mediated inoculation, may be inoculated with infectious TYLCV DNA clones using agroinoculation or DNA-coated micro-projectile bombardment.

Epidemiology

Geographical Distribution and Spread

The name Tomato yellow leaf curl virus (TYLCV) was coined in the early 1960s to describe a virus transmitted by the whitefly *B. tabaci* that affected tomato cultures in Israel. At the time, diagnosis of TYLCV was based essentially on symptom observation: border curling, yellowing and thickening of leaves, and plant growth stunting.

The whitefly *B. tabaci*, has been instrumental in the worldwide spread of TYLCV. Whiteflies can fly for several km. They are very resilient and their eggs can sustain low temperatures. Global warming may facilitate the expansion of *B. tabaci* to presently cold regions devoted of whiteflies.

Perennial weeds such as *Cynanchum acutum* and the annual weed *Malva parviflora* constitute TYLCV reservoirs in fields, especially outside crop growing seasons. A recent report claimed that TYLCV from South Korea is also transmitted through tomato seeds. This report needs to be confirmed.

DNA sequencing has become the tool of choice allowing accurate identification of TYLCV isolates and their relationship with others. The virus first reported in Israel in the 1960s was isolated and sequenced in the late 1980s. Another, different, TYLCV species from Israel was coined TYLCV Mild and was sequenced in the early 1990s. TYLCV was reported in the mid and late 1970s in Cyprus, Jordan, and Lebanon. It was identified in Egypt and Turkey in the early 1980s. The virus has spread to Turkey, Iran and the Asian republics of the former USSR, and to the Arabian Peninsula during the mid-late 1990s. Two TYLCV isolates closely related to the Middle Eastern virus were described in Japan in 1996, from where they spread to South Korea where TYLCV was reported in 2008. In China, TYLCV isolates were identified in Shanghai in 2006 and have rapidly spread since to most of the country, thanks to invasive whitefly species. A local species was coined Tomato yellow leaf curl China virus (TYLCCNV). It was shown to possess a betasatellite. In the early 1990s two virus isolates belonging to a new species related to the Middle Eastern TYLCV, named Tomato yellow leaf curl Sardinia virus (TYLCSV) were identified in Sardinia and in Sicily, Italy. The Sardinian isolate was discovered in Spain in the early 1990s. By the mid 1990s the Middle Eastern TYLCV was found in Portugal and in Spain. In the latter country, it tended to displace the previously established Italian virus. Recombinants between the two viruses are not rare. TYLCV has invaded North Africa, possibly from Spain and Italy. In Morocco and Tunisia, both the Italian and the Middle Eastern strain were discovered in the early 2000s. TYLCV appeared in Southern France in 1999. In 2000, the Middle Eastern strain of TYLCV was identified in Greece. In East Africa, TYLCV was present in Sudan as early as the late 1970s. In Tanzania, a virus distinct from the Mediterranean isolates was identified. In the Réunion Island, TYLCV was detected in the late 1990s. The Middle Eastern strain of TYLCV has appeared in the Western Hemisphere in the mid 1990s in the Caribbean Islands, first in the Dominican Republic, then Cuba, Jamaica, Puerto Rico and the Bahamas. From there, the virus has reached the USA, identified first in Virginia in the late 1990s, then in Florida, Georgia, Louisiana, North Carolina and Mississippi. TYLCV was reported in the late 2000s in several regions of Mexico. In 2006, TYLCV was reported in the Brisbane area, Australia, likely imported from Japan.

Hence, TYLCV has spread extremely fast and at present, constitutes a major limiting factor to tomato cultivation, worldwide. In many regions, the invading Mediterranean strain of TYLCV co-exists with local TYLCV strains but in several cases, it has almost replaced them, as in Southern Spain. Several introduction of TYLCV seemed to have occurred to the American/Caribbean region from the East-Asian, Australian, and Western Mediterranean regions, and several introductions of TYLCV to the East-Asian region

from the Eastern Mediterranean and Western Mediterranean regions. The detection of TYLCSV in Jordan and in Israel, indicate that TYLCV spread may occur from the Western to the Eastern Mediterranean region and to other parts of the world.

TYLCV-Whitefly Vector Relationship

Path of the virus in the insect

Like all begomoviruses, TYLCV follows a well-defined path in its *B. tabaci* vector (**Fig. 1(c)**). Whiteflies feed when their stylets reach the virus-rich phloem of infected leaves. TYLCV can be detected in the insect head approximately 10 min after the beginning of feeding (acquisition access period AAP) on infected tomato. From the esophagus, the virus reaches the midgut where it can be detected about 30 min after it was first found in the head. At this point, some of the virus might be excreted through the hindgut and rectum (honeydew). The virus reaches the hemolymph 30 min after it was first detected in the midgut, 90 min after the beginning of the AAP. The crossing of the gut is likely to be an active process involving specialized, unknown, receptors. To avoid degradation in the hemolymph, TYLCV interacts with a GroEL-like chaperonin produced by the insect endosymbiotic bacteria and excreted in the hemolymph. TYLCV can be detected in the salivary glands approximately 5.5 h after it was first detected in the hemolymph, 7 h after the beginning of the AAP, approximately 1 h before the insects are able to infect tomato plants. Once the virus reaches the salivary gland, crossing several cell walls that may constitute selective receptor-mediated barriers, it is almost immediately excreted into the salivary pump and from there into the plant, together with the saliva. TYLCV transits with the same velocity in males and females. The virus capsid is the only structure recognized by the would-be receptors in the gut and salivary gland barriers, and by endosymbiotic GroEL.

Retention of the virus in the insect vector

Once TYLCV is acquired, viral DNA remains associated with *B. tabaci* for the entire adult life of the insect, although infectivity decreases with time, but not entirely. TYLCV concentrates mainly in the filter chamber and the midgut. Various TYLCV isolates may present different parameters of interactions with their whitefly host.

TYLCV, TYLCSV and TYLCCNV were detected with very low incidence in eggs and in the adult progeny of viruliferous whiteflies. The CP was shown to be essential for virus invasion of the female reproductive system of *B. tabaci* B and Q. Entry of TYLCV into eggs developing in ovaries is mediated by the interaction of CP with vitellogenin. The CP region necessary for crossing the gut into the hemolymph is also necessary for salivary glands and ovaries penetration. TYLCV and TYLCCNV were shown to be transmitted with very low frequency among whiteflies during mating.

Replication of the virus in the insect vector

Replication of begomoviruses in their whitefly vector is a controversial issue. To prove, or disprove replication, viral DNA (genomic and RF) and CP amounts were usually appraised in whiteflies following a short AAP on TYLCV-infected plants and transfer to a non-host such as cotton, or on an artificial diet.

In one set of experiments involving TYLCV, TYLCCNV and TYLCV-Mild, the amount of viral genomic DNA and its complementary strand, and of viral transcripts (especially from the viral genome complementary strand) increased with time, and then decreased, as measured by qPCR. These molecules were visualized by FISH in the midgut epithelial cell nuclei and in cells of the salivary glands. *In situ* fluorescent immunodetection and western blots showed increases in CP amounts. These results supported TYLCV replication in whiteflies.

In another set of experiments, TYLCV and TYLCV-Mild did not accumulate following the end of the AAP. Moreover, whiteflies fed with purified TYLCV and TYLCV-Mild virions through membranes did not show any evidence of replication. These results led to the conclusion that TYLCV did not replicate in its vector.

Effects of TYLCV on life expectancy and fertility of the whitefly vector

Several studies indicated direct and/or indirect influence of begomoviruses on *B. tabaci*, which can be detrimental, neutral or beneficial, depending on the virus-whitefly-plant combinations. In experiments performed in China, TYLCV infection of tomato indirectly improved the overall performance (higher survival, greater fecundity, shorter development time) of the *B. tabaci* Q biotype, whereas it reduced the overall performance of the B biotype. Other studies performed in China indicated TYLCV has little effect on the reproduction and development of Q whiteflies. In Israel, 3-days-old B whiteflies raised on eggplants (a TYLCV non-host) following a 48 h AAP on TYLCV-infected tomato plants showed a significant reduction in their life expectancy and a ~50% decrease in the number of eggs laid. TYLCCNV also reduces female longevity and fecundity of B whiteflies. The negative impact of TYLCCNV was explained by TYLCCNV's suppression of the whitefly immune responses by down-regulation genes in the Toll-like signaling and mitogen-activated protein kinase (MAPK) pathways. The deleterious effects on its whitefly vector, in addition to the invasion of the reproductive system, suggest that TYLCV has some features reminiscent of an insect.

Pathogenesis

In the field, it is common that tomato seedlings are infected immediately after planting. Infection is always associated with large amounts of whiteflies (**Fig. 1(c)**). Young leaves and shoot apices are the best targets for whitefly-mediated virus inoculation. In the field as well as the greenhouse, symptoms of TYLCV infection include severe stunting, reduction of leaf size, upward cupping/curling of leaves, chlorosis on leaves and flowers, and reduction of fruit production (**Fig. 1(b)** and **(b')**). TYLCV can cause up to 90%–100% yield loss. In nethouses and greenhouses, infected plants are found near the (double) doors and close to ventilation openings.

Diagnosis

Symptom observation

The first step of diagnosis is examining symptoms, and inspecting the lower leaf surface for the presence of whiteflies. Early symptoms include yellowing, curling and thickening of young leaves. Later on, the plant becomes stunted and presents a bushy-like growth.

Microscopy

Light microscope observation of leaves infected with geminiviruses, including TYLCV, reveals characteristic large blue-violet nuclear inclusions following azure-A staining. Typical geminate particles can be observed with the transmission electron microscope (TEM) following virion purification. The virions can be decorated with anti-CP antibodies (immune electron microscopy) prior to observation under TEM. Particles can be TEM-observed in thin sections of infected leaves, using CP antisera and immuno-gold labeling. These methods are not of practical use for the detection of the virus in large numbers of plant samples.

Immunodetection

Immunodetection of TYLCV includes ELISA and western blots. Particles are also visualized in preparations of plant and insect tissues using labeled antibodies. However, the antibodies raised against the viral CP (overexpressed in bacteria or from purified virions) are usually unable to distinguish TYLCV from closely related begomoviruses, and unable to discriminate between TYLCV species and strains.

Nucleic acid hybridization

TYLCV genomic DNA and replicative dsDNA are readily detectable in DNA isolated from plants subjected to Southern blot hybridization, using a labeled DNA probe (preferably not radioactive). Similarly, TYLCV genomic DNA is detectable in viruliferous whiteflies.

TYLCV DNA sequences are detected by hybridization of infected plant tissues and viruliferous whiteflies squashed onto a nylon membrane (squash-blot, or tissue-blot) with a labeled specific DNA probe. No treatment of the sample is necessary prior to squashing and hybridization. Large numbers of plant and insect samples can be collected from a field and analyzed in the laboratory.

Viral RNAs are detectable in plant and insect preparations using fluorescently labeled (usually by Cy3 or Cy5) synthetic primers complementary to the RNAs. The samples are observed with a fluorescence microscope or a laser confocal microscope.

A microarray containing 279 40-mer oligonucleotide probes was constructed to detect ten important tomato viruses. It included sequences derived from TYLCV CP DNA. It allowed detecting TYLCV but was unable to discriminate between TYLCV and TYLCVSV.

Nucleic acid amplification

PCR is widely used for the diagnosis of TYLCV strains. The virus is detectable in very small amounts of infected plant and vectors. Subsequently, the amplicon can be sequenced to identify the TYLCV strain.

Whitefly GroEL, whether extracted from *B. tabaci* or overexpressed in *E. coli*, binds begomoviruses with strong affinity. This property was exploited to trap TYLCV by incubating infected plant and viruliferous whitefly extracts with GroEL-coated PCR tubes. TYLCV was then detected by PCR.

RCA allows amplifying full-length viral genomes, including TYLCVs, in a single-step and without the use of specific primers. The amplicons can be sequenced for identification and phylogenetic analyses. RCA was implemented to generate infectious genomic clones directly from plant and whitefly DNA, and from specimens sometimes stored for decades, including TYLCV.

Sequencing

Mass sequencing (millions of 100–700 nt reads) using platforms such as Illumina combined with bioinformatic softwares was used to study the effects of TYLCV on the transcriptome, and on virus-related RNAi and microRNA populations of plants and whiteflies.

Metagenomic analyses of whiteflies and infected plants allowed delivering fast and inexpensive information on the presence of TYLCV isolates and other viruses (some previously unknown). This is an exquisite tool to survey the diversity and movement of TYLCV in small as well as large environmental settings.

Treatment

Prevention

Planting in whitefly-free period, eradication of inoculum

Disease management strategies include the elimination of possible inoculum (residual infected plants and weeds), and planting when whitefly population pressure is low. In all regions where TYLCV is endemic, seedlings to be planted in the field or greenhouse need to be grown under cover to prevent early infection.

Cultures under cover

Growing tomato in net or plastic houses covered with insect-proof 50-mesh screens and with double-doors, constitutes the optimal solution for crop protection against the TYLCV whitefly vector. Sticky yellow polyethylene sheets are used to decrease the pressure of the whitefly population in greenhouses, with limited success. Ultraviolet (UV)-absorbing plastic sheets or UV-absorbing 50-mesh screens used as nethouse covers, greatly repel whiteflies and reduce TYLCV incidence.

Chemical control of whiteflies

TYLCV is usually controlled by spraying insecticides. Many chemicals used in the past, such as organophosphates, carbamates and pyrethroids, effectively reduced whitefly population, until the insects developed resistance. More potent insecticides, with novel modes of action, such as imidacloprid, have been introduced in the 1990s. As anticipated, whitefly populations resistant to imidacloprid have emerged in many places.

Biological control of whiteflies

Whitefly IPM control in greenhouse tomato production of Europe, and to a lesser extent of the USA, includes biological control with natural enemies such as the parasitic wasp *Eretmocerus* spp. and the predatory bug *Nesidiocorus tenuis*, used alone or in combination. These practices are not yet common in the open-field cropping systems.

Breeding for Resistance

Certain accessions of the wild tomato species *S. chesmanii*, *S. chilense*, *S. habrochaites*, *S. peruvianum*, and *S. pimpinellifolium* are resistant/tolerant to whitefly-mediated inoculation (show mild or no symptoms, but contain virus). Breeding programs aimed at producing tomato cultivars resistant to TYLCV started in the late 1960s. They were based on the introgression of resistance from wild tomato species into the domesticated tomato *Solanum esculentum*.

The discovery of loci, and later on of genes, associated with TYLCV resistance, was facilitated by the development of saturated maps based on DNA polymorphism distinguishing resistance from susceptibility (RFLP, AFLP, SSR, SCAR, etc.). Altogether six loci, coined *Ty-1* to *Ty-6*, were found to be associated with TYLCV resistance. The *Ty-1* locus originating from *S. chilense* was mapped to tomato chromosome 6. Ultimately, *Ty-1* was identified as coding for an RNA-dependent RNA polymerase. Another TYLCV resistance locus, *Ty-3*, was shown to be allelic to *Ty-1*. *Ty-2* was introgressed from *S. habrochaites* and was located on chromosome 11. *Ty-2* encodes a nucleotide-binding domain and leucine-rich repeat-containing (NB-LRR) gene. *Ty-4* was introgressed from *S. chilense* and mapped to chromosome 3. *Ty-5* was introgressed from *S. peruvianum* and mapped on chromosome 4. *Ty-5* encodes a messenger RNA surveillance factor Pelota, which is involved in ribosome recycling phase of protein synthesis. *Ty-6* originated from *S. chilense* and was located on chromosome 10. Combination of resistance loci from different sources in a single tomato line often enhances resistance.

Today, most commercial tomato varieties are resistant/tolerant to TYLCV; in most cases, they have incorporated *Ty-1*. They remain symptomless (or with mild symptoms) upon infection, and yield nearly as uninfected plants (Fig. 3). However, *Ty-1*-mediated resistance collapses under high disease pressure. It was recently reported in Jordan and in Israel that a newly found betasatellite associated with TYLCV overcomes the resistance based on *Ty-1*.

Genetic Engineering

Expression of viral genes

A variety of strategies have been devised based on the pathogen-derived resistance concept, which involves the expression of functional as well as dysfunctional viral genes. A transgenic interspecific F1 hybrid *S. lycopersicum* x *S. pennellii* expressing the TYLCV CP under the Cauliflower mosaic virus (CaMV) 35S promoter showed delay and remission of symptoms upon whitefly-mediated inoculation.



Fig. 3 TYLCV-infected tomato fields planted with susceptible and resistant tomato genotypes. Note that the susceptible plants are stunted; they will not produce fruits. For comparison, the resistant plants will remain symptomless, or with mild symptoms and will yield.

Tomato and tobacco plants transformed with a truncated (2/5) TYLCV *Rep* gene under the control of the 35S CaMV promoter. Following whitefly-mediated inoculation with TYLCV in the laboratory, T2 generation lines evaluated in the field in Florida showed variable rates of resistance. However, the TYLCV-resistant plants were susceptible to the bipartite begomovirus Tomato mottle virus (ToMoV) from Florida. Similarly, tomato plants transformed with the *Rep* gene of TYLCSV truncated in its 3' end exhibited resistance to TYLCSV but were susceptible to the related strain from Spain.

Expression of antisense sequence of the TYLCSV *Rep* gene in transgenic *Nicotiana benthamiana* resulted in lesser virus replication. The incorporation of a ribozyme structure into the antisense construct did not increase the efficiency of the system.

RNAi

The silencing-based RNA interference (RNAi) antiviral defense mechanism was exploited to express siRNA directed against TYLCV genes. Post-translational gene silencing (PTGS) was activated through the transgenic expression of hairpin dsRNA cognate to the target viral genome. PTGS was efficient but was highly-sequence dependent.

In the early experiments, non-coding conserved regions from the genome of TYLCV, TYLCV-Mild, and several TYLCSVs from Spain, were used to design constructs aimed at triggering broad resistance against different strains of TYLCV. Upon whitefly-mediated inoculation, a high level of resistance was obtained against all three viruses, which correlated with the accumulation of TYLCV-specific siRNAs, indicating that PTGS can be used to engineer geminivirus-resistant plants.

A hairpin RNAi (hpRNAi) aimed at expressing dsRNA homologous to sequences of the IR, the CP, V2 and *Rep* genes of TYLCV from Oman was used to transform tomato plants. Nine transgenic lines were obtained, which showed various levels of resistance upon TYLCV agroinoculation.

RNAi directed at TYLCV *Rep* gene was expressed in transgenic tomato plants. The homozygous F4 and F6 generations exhibited resistance to TYLCV in the greenhouse and in whitefly-infested fields, and contained high amounts of small RNAs.

CRISPR/Cas9

The clustered regularly interspaced short palindromic repeats system (CRISPR/Cas), which protects prokaryotes from potentially deleterious foreign DNA, was used to engineer the genome of tomato and *N. benthamiana* to confer immunity to TYLCV. Targeting the TYLCV CP and *Rep* sequences with Cas9-single guide RNA resulted in efficient virus interference. The treated plants had low TYLCV load and their T3 progeny was tolerant to TYLCV. Therefore, genome editing might be the tool of the future to obtain TYLCV-resistant tomato genotypes.

TYLCV Transmission to Tomato and Host Response

Transmission by whiteflies

In nature, TYLCV is transmitted by the whitefly *B. tabaci* in a persistent circulative manner. Transmission of TYLCV to tomato by *B. tabaci* B biotype has been studied thoroughly. The virus is transmitted to tomato plants after vector feeding on infected tomato plants or alternative hosts. The incidence of the disease is directly correlated with the pressure of the whitefly population. One to three viruliferous insect are able to infect a tomato plant. The efficient AAP is 15–30 min, the latent period is 8–24 h, and the efficient inoculation access period (IAP) is at least 15 min. Female *B. tabaci* are better vectors than males. The ability of the whiteflies to transmit TYLCV to tomato test plants decreases with age, from 100% to 10%–20% during their 4–5-week adult lifetime.

Response of plants to TYLCV infection

RNAi

RNAi is a natural defense mechanism against viruses. It is triggered by the processing of virus-derived dsRNA by host Dicer-like (DCL) enzymes into 21–24 nt small interfering RNA (siRNAs). Viral siRNAs bind Argonaute (AGO) proteins and guide the resulting RNA-induced silencing complexes toward their cognate RNA molecules, leading to post-transcriptional or transcriptional silencing in the case of RNA and DNA viruses, respectively. Viral siRNAs can move to other cells and into the phloem. As a counter-defense, geminiviruses encode one or more silencing suppressor proteins. TYLCV C2 and V2 proteins were reported to suppress RNA silencing by inhibiting the siRNA amplification stage. Infection depends on the balance between silencing and suppression. This balance can be shifted to virus tolerance by overexpressing antiviral dsRNA in the form of RNA hairpins in transgenic plants.

Aggregates

Upon tomato infection, TYLCV CP-DNA complexes accumulate in the phloem associated cells of tomato leaves. At the early stages of infection, punctate aggregates are detected in the cytoplasm; in the later stages aggregates of increasing size are localized in cytoplasm and nuclei. Nuclear aggregates contained infectious particles transmissible to test plants by whiteflies. The formation of large aggregates was delayed in TYLCV-resistant tomatoes, perhaps as part of the resistance mechanism.

Transcriptome and metabolome

TYLCV induces significant changes already during the first days after inoculation, changes that exacerbate as infection progresses. They include the activation of genes involved in general stress-response, hormone biosynthesis, signal transduction, RNA regulation and processing, induction of the ubiquitination pathway, and initiation of autophagy. Photosynthesis decline, is accompanied by release of hexose that activates defense response or, if failing, promote pathogen replication and disease expression.

Upon TYLCV infections, tomato plants emit reactive oxygen species (ROS), pathogenesis-related (PR), and wound-induced proteins. Sources of carbon and nitrogen are highly affected, accompanied by changes in the abundance of various classes of metabolites such as amino acids and polyamines, phenolic and indolic metabolites. Hence, upon TYLCV infection, there is a coordinated reprogramming of phenylpropanoid, tryptophan/nicotinate, urea/polyamine, and salicylic acid biosynthesis pathways leading to the production of defense compounds.

Involvement of heat shock and quality control proteins in the establishment of TYLCV infection

TYLCV does not induce a hypersensitive response and cell death upon whitefly-mediated infection of susceptible tomato plants. Cell death can be induced by the inactivation of heat shock protein 90 (HSP90) as well as by silencing the genes *Hsp90* and *Sgt1* (HSP90 co-chaperone), causing the dissociation of the 26S proteasome and a decrease of its peptidase activity, which led to the accumulation of damaged ubiquitinated proteins. TYLCV alleviates cell death and these accompanying effects to create an environment suitable to its replication.

The interactions between TYLCV and HSP70 and HSP90 chaperons in plants and vectors are essential for virus infection to proceed. In infected host cells, HSP70 and HSP90 are redistributed from a soluble to an aggregated state. These aggregates contain, together with viral DNA/proteins and virions, HSPs and components of the protein quality control system such as ubiquitin, 26S proteasome subunits, and the autophagy protein ATG8. TYLCV CP can form complexes with HSPs in tomato and whitefly. Nonetheless, HSP70 and HSP90 play different roles in the viral cell cycle in the plant host. In the infected host cell, HSP70, but not HSP90, participates in the translocation of CP from the cytoplasm into the nucleus. Viral amounts decrease when HSP70 is inhibited, but increase when HSP90 is downregulated. In the whitefly vector, HSP70 impairs the circulative transmission of TYLCV; its inhibition increases transmission.

Further Reading

- Brown, J.K., Zerbini, F.M., Navas-Castillo, J., *et al.*, 2015. Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Archives of Virology* 160, 1593–1619.
- CABI, 2019. Invasion Species Compendium. Tomato Yellow Leaf Curl Virus (Leaf Curl). Available at: <https://www.cabi.org/isc/datasheet/55402>.
- Czosnek, H., 2007. Tomato Yellow Leaf Curl Virus Disease: Management, Molecular Biology, Breeding for Resistance. Dordrecht: Springer.
- Czosnek, H., Hariton-Shalev, A., Sobol, I., Gorovits, R., Ghanim, M., 2017. The incredible journey of begomoviruses in their whitefly vector. *Viruses* 9, 273.
- Gorovits, R., Czosnek, H., 2017. The involvement of heat shock proteins in the establishment of *Tomato yellow leaf curl virus* infection. *Frontiers in Plant Science* 8, 355.
- Hanley-Bowdoin, L., Bejarano, E.R., Robertson, D., Mansoor, S., 2013. Geminiviruses: Masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology* 11, 777–788.
- Mabvakure, B., Martin, D.P., Krabberger, S., *et al.*, 2016. Ongoing geographical spread of Tomato yellow leaf curl virus. *Virology* 498, 257–264.
- Sneh, S.K., Raj, S.K., Prasad, V., Singh, V., 2015. Recent research findings related to management strategies of Begomoviruses. *Journal of Plant Pathology and Microbiology* 6, 273.
- Zhou, X., 2013. Advances in understanding begomovirus satellites. *Annual Review of Phytopathology* 51, 357–381.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/geminiviridae/392/genus-begomovirus
Genus: Begomovirus.

Tombusvirus-Like Viruses (*Tombusviridae*)

K Andrew White, York University, Toronto, ON, Canada

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

', Prime symbol

γ Lowercase gamma (Greek)

Å Angstrom

Glossary

Frameshifting A ribosome shifting its reading frame during protein translation.

Higher-order RNA structures RNA secondary and tertiary structures, such stem-loops or pseudoknots, respectively.

Monocot A class of angiosperm plant with a single embryonic leaf and parallel veins.

Nanoparticle A microscopic particle in the approximate range of 1–100 nm.

Open reading frame A segment of RNA encoding a protein.

T Triangulation number, which describes the icosahedral geometry of spherical viruses.

Tombusvirid A member of the virus family *Tombusviridae*.

Transcription Herein, describes the process by which viral sg mRNAs are synthesized.

Type A designated member that represents a group.

Introduction

Tombusvirus is the type genus in the family *Tombusviridae*, which currently includes a total of sixteen genera. The members of this extensive family possess linear, single-stranded, positive-sense RNA genomes that are packaged into spherical particles and, collectively, infect a wide variety of plant hosts. The overall grouping of these viruses is based on the presence of a conserved RNA-dependent RNA polymerase (RdRp). Within the family, members are further subdivided into subfamilies, genera, and species based on similarities in gene expression strategies and encoded proteins. In this article, the genus *Tombusvirus* and those genera that are most closely related to this genus, *i.e.*, *Aureusvirus* and *Zeavirus*, are described.

Taxonomy, Phylogeny, and Evolution

The family *Tombusviridae* resides within the realm *Riboviria*, and includes three subfamilies (*Procedovirinae*, *Regressovirinae* and *Calvusvirinae*) and sixteen genera (*Alphacarmovirus*, *Alphanecrovirus*, *Aureusvirus*, *Avenavirus*, *Betacarmovirus*, *Betanecrovirus*, *Dianthovirus*, *Gallantivirus*, *Gammacarmovirus*, *Macanavirus*, *Machlomovirus*, *Panicovirus*, *Pelarspovirus*, *Tombusvirus*, *Umbravirus*, and *Zeavirus*), each of which contain member species that range in number from one to seventeen. Most genera belong to the subfamily *Procedovirinae*, all of which express their RdRps via translational readthrough of a stop codon. In contrast, dianthoviruses are in *Regressovirinae*, because they translate their RdRps via frameshifting. Umbraviruses, which do not encode a capsid gene, occupy *Calvusvirinae*. The current taxonomy is the result of recent reorganization within the family, and additional modifications and refinements are anticipated in the near future. Consequently, updates should be monitored by visiting the International Committee on Taxonomy of Viruses website (see "Relevant Websites section").

Taxonomic classification of tombusvirids is based primarily on the structural relationship of their viral RdRps. All contain supergroup-II RdRps that share high levels of amino acid sequence identity, including several tombusvirid-specific motifs located in C-terminal regions. Indeed, taxonomic organization down to the level of species can be guided entirely by sequence comparisons of RdRps. Nonetheless, additional features distinct to specific genera, such as the type of movement protein or silencing suppressor encoded, also aid in appropriate grouping. The notable level of diversity observed within these two protein types suggests that they were acquired via recombination-mediated modular transfer of gene cassettes.

Phylogenetic analysis of the RdRps of tombusvirids have revealed an evolutionary tree that is comprised of distinct lineages (Fig. 1). The largest two of these lineages can be defined roughly as either tombusvirus-like or carmovirus-like, with the latter further classified into numerous genera. Only four genera reside within the tombusvirus-like clade, *Tombusvirus*, *Aureusvirus*, *Zeavirus* and *Betanecrovirus*. However, the latter genus has movement and capsid proteins typical of carmoviruses. Consequently, betanecroviruses are not discussed here, and instead are presented in the article covering necrovirus-like viruses. Of the remaining three tombusvirus-like genera, all share high degrees of similarity with respect to gene order, protein relatedness, and expression strategies. Based on RdRp phylogenetic analysis, zeaviruses are predicted to be more closely related to tombusviruses than aureusviruses. However, the zeavirus host range of monocot plants (*e.g.*, grasses) is more akin to certain aureusviruses. Thus, although these genera are closely related, they also possess features that distinguish them from one another. Currently, the genus



Fig. 1 Phylogenetic tree depicting evolutionary relationships of genera in the family *Tombusviridae* based on the RdRp amino acid sequences of tombusvirids. The tombusvirus-like lineage is shaded in yellow and the genera discussed in this article are shown in red. Adapted from Gunawardene, C.D., Donaldson, L.W., White, K.A., 2017. Tombusvirus polymerase: Structure and function. *Virus Research* 234, 74–86.

Tombusvirus is composed of seventeen species, including *Tomato bushy stunt virus* (TBSV), which is the type virus of both the genus and family (Table 1). The less crowded *Aureusvirus* genus has five confirmed member, with *Pothos latent virus* (PoLV) designated as the type species. Presently, the genus *Zeavirus* contains only one member, its type species, *Maize necrotic streak virus* (MNeSV).

Virion Structure, Assembly, and Disassembly

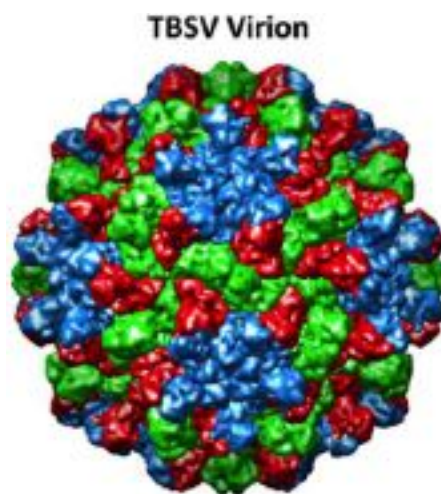
Members of the tombusvirus-like group form non-enveloped, spherical virions that encase a single viral RNA genome. The capsids are ~30 nm in diameter, have T=3 icosahedral symmetry, and are composed of 180 subunits of a capsid protein (Fig. 2). The capsid proteins of tombus- and aureusviruses are of similar molecular masses, 40 and 41 kDa, respectively. In contrast, that for zeaviruses is substantially smaller, 27 kDa, due to the absence of a C-terminal domain present in the other two capsid proteins.

Among the three tombusvirus-like genera, the particles of tombusviruses have been studied most extensively. Going from N- to C-terminus, these capsids are composed of three distinct structural domains, RNA-binding (R), shell (S), and protruding (P). Orientation of the subunits within a capsid is such that C-terminal P-domains extend out from the particle surface formed by the S-domain, while the N-terminal R-domain, a highly basic region, is positioned internally and neutralizes the negatively charged phosphate groups in the packaged RNA genome. The single capsid subunit is able to assume three different conformations that allow for its packing into the 5- and 3-fold centers of particle symmetry. Only conformation A subunits reside at the 5-fold axes (blue), while alternating B and C subunits (red and green, respectively) fill the 3-fold centers (Fig. 2). The P-domains of two adjacent subunits interact in AB or CC dimers to form ninety projections that can be observed in electron micrographs. This latter feature is not present in zeaviruses, as their capsid proteins lack the C-terminal P-domain that forms these protrusions.

Details of the process of virion assembly in these viruses remain largely unknown. For many viruses, key regions of their RNA genomes have been identified as packaging signals that bind to capsid subunits and initiate virion assembly. Currently, no corresponding RNA assembly signals in the genomes of tombusvirus-like viruses have been identified. However, studies on the capsid protein of the tombusvirus Cucumber necrosis virus (CuNV) have identified a determinant in the R-domain, a basic KGKKGK motif, that dictates the efficiency of capsid assembly, as well as particle size, i.e., T=1 (60 subunits) versus T=3 (180 subunits). Furthermore, a second basic

Table 1 List of current tombusvirus-like viruses grouped by genera. The type species of each genus is indicated in bold and the corresponding abbreviation for each virus is provided

<i>Genus</i>	<i>Species</i>	<i>Abbreviation</i>
<i>Aureusvirus</i>	<i>Pothos latent virus</i>	PolV
	<i>Cucumber leaf spot virus</i>	CLSV
	<i>Johnsongrass chlorotic stripe mosaic virus</i>	JCSMV
	<i>Maize white line mosaic virus</i>	MWLMV
	<i>Yam spherical virus</i>	YSV
<i>Tombusvirus</i>	<i>Tomato bushy stunt virus</i>	TBSV
	<i>Artichoke mottled crinkle virus</i>	AMCV
	<i>Carnation Italian ringspot virus</i>	CIRV
	<i>Cucumber Bulgarian latent virus</i>	CBLV
	<i>Cucumber necrosis virus</i>	CuNV
	<i>Cymbidium ringspot virus</i>	CymRSV
	<i>Eggplant mottled crinkle virus</i>	EMCV
	<i>Grapevine Algerian latent virus</i>	GALV
	<i>Havel River virus</i>	HaRV
	<i>Lato River virus</i>	LRV
	<i>Limonium flower distortion virus</i>	LFDV
	<i>Moroccan pepper virus</i>	MPV
	<i>Neckar River virus</i>	NRV
	<i>Pelargonium leaf curl virus</i>	PLCV
	<i>Pelargonium necrotic spot virus</i>	PENSV
	<i>Petunia asteroid mosaic virus</i>	PeAMV
	<i>Sitke waterborne virus</i>	SWBV
<i>Zeavirus</i>	<i>Maize necrotic streak virus</i>	MNeSV

**Fig. 2** TBSV virion structure. Image of TBSV particle at 2.9 Å, which is typical of tombusvirus and aureusvirus virions. Subunit conformations are color coded: A (blue), B (red) and C (green). Image reproduced from <http://viperdbscripps.edu>. Ho, P.T., Montiel-Garcia, D.J., Wong, J.J., *et al.*, 2018. VIPERdb: A tool for virus research Annual Review of Virology 5 (1), 477–488.

motif in the arm region connecting the R- and S-domains is important for determining the amount of full-length RNA that gets incorporated into virions and may contribute to the specificity of packaging viral genomes. Accordingly, interactions involving these positively-charged interior regions of the capsid and the negatively-charged viral RNA are proposed to mediate virion formation; the detail of which remain to be determined.

An equally important aspect of virion function is its ability to disassemble and release its viral genome in newly infected cells. Studies with the tombusvirus CuNV suggest that plant Hsp70 protein chaperone homologs, Hsc70 or Hsc70-2, assist in CuNV disassembly. Hsc70-2 copurifies with CuNV particles, overexpression of Hsc70 in plants facilitates infections, and incubation of Hsc70 with particles destabilizes them. These observations are consistent with the involvement of Hsc70 or its homologs in

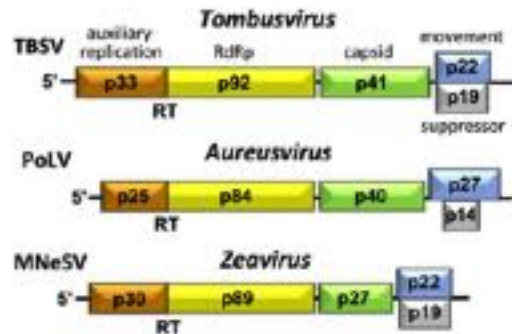


Fig. 3 Cartoon representation of the plus-strand RNA genomes of the type species for tombusvirus-like genera: *Tomato bushy stunt virus* (TBSV), *Pothos latent virus* (PoLV) and *Maize necrotic streak virus* (MNeSV). Encoded proteins are represented by color-coded boxes along with the molecular masses of corresponding proteins (p), in kDa. The functions of proteins are labeled for the TBSV genome, with corresponding proteins in the other two genomes having equivalent functions. RT indicates the position of the stop codon for accessory replication proteins in the genomes where translation readthrough occurs to produce their RNA-dependent RNA polymerases (RdRps).

mediating disassembly of this tombusvirus. Moreover, since Hsp70 homologs also associate with virions of the aureusvirus *Cucumber leafspot virus*, and overexpression of Hsc70-2 in host plants mediates more robust infections, this proposed mode for unpacking may also apply to this related genus.

Genome Structure

The linear, plus-sense, single-stranded RNA genomes of tombusvirus-like viruses are neither 5'-capped nor 3'-polyadenylated. The sizes of these genomes are ~4.1, ~4.2, and ~4.8 kb in length, respectively, for ze-, aureus-, and tombusviruses (**Fig. 3**). These genomes encode five proteins, and share a similar gene composition and organization that is unique among tombusvirids. Encoded 5'-proximally is an auxiliary RNA replication protein, homologs of which are present at comparable genomic positions in all members of the family. Translational readthrough of the stop codon of this protein leads to the production of a supergroup-II RdRp, also a common product in all tombusvirids. Encoded downstream from the polymerase is a single capsid protein, which contains a P-domain in tombus- and aureusviruses that is absent in zeaviruses. The feature that clearly distinguishes the tombusvirus-like group from other genera is their movement and suppressor of gene silencing proteins. These two proteins are encoded as overlapping open reading frames (ORFs) and are positioned 3'-proximally in viral genomes (**Fig. 3**). The coding regions of the suppressor proteins are smaller than, and are located entirely within, corresponding movement protein ORFs. However, because the two ORFs reside in different reading frames, each encodes a structurally unique protein. Homologs of these two viral products are not found in other tombusvirids, thereby defining these genera as a distinctly-related trio.

In addition to their role in encoding viral proteins, these plus-strand RNA genomes also house multiple RNA elements that are involved in regulating different viral processes. Some of these elements are local, such as the promoter sequences located at the 3'-termini of these genomes. Interestingly, tombusviruses also utilize at least six different sets of intragenomic, long-range, base pairing RNA-RNA interactions for different viral functions. Some of these span almost the entire length of these 4.8 kb genomes, whereas others traverse shorter distances, while still exceeding a kilobase in length. The presence of these interactive RNA-RNA networks suggests that there must be a significant level of structural organization in these viral genomes to assist bringing distal partner regions together and to coordinate the different interactions and their corresponding functions. Indeed, analysis of the RNA secondary structure of the complete TBSV genome revealed that much of it is ordered and forms discrete RNA domains (**Fig. 4**). Notably, the predicted structure included only two of the six long-range interactions needed during viral infections (*i.e.*, AS1-RS1 and DE-CE) (**Fig. 4**). This suggests that dynamic reorganization of the genome is required to achieve the six different long-range interactions (**Fig. 5(A)**). Such structural rearrangements could be facilitated by viral or host factors that are able to direct particular RNA folding pathways. In this regard, the auxiliary RNA replication protein p33 of TBSV has been found to possess RNA chaperone activities, and thus could participate in such genome reorganizing events.

Gene Function

Studies exploring protein function in the group have been carried out largely in tombusviruses (*e.g.*, TBSV), however the corresponding proteins in the other two genera are predicted to operate similarly. Both of the 5'-proximally encoded proteins in these viral genomes are essential for viral RNA synthesis. The auxiliary RNA replication protein of TBSV, p33, is a multifunctional master regulator of viral RNA replication that interacts with a myriad of host factors. This protein is also a major structural component of virus replication complexes (VRCs), the sites of viral RNA synthesis. VRCs also contain an RdRp, p92 for TBSV, that catalyzes both replication of the genome and transcription of viral subgenomic (sg) mRNAs. Additional information on these two processes are provided in subsequent sections.

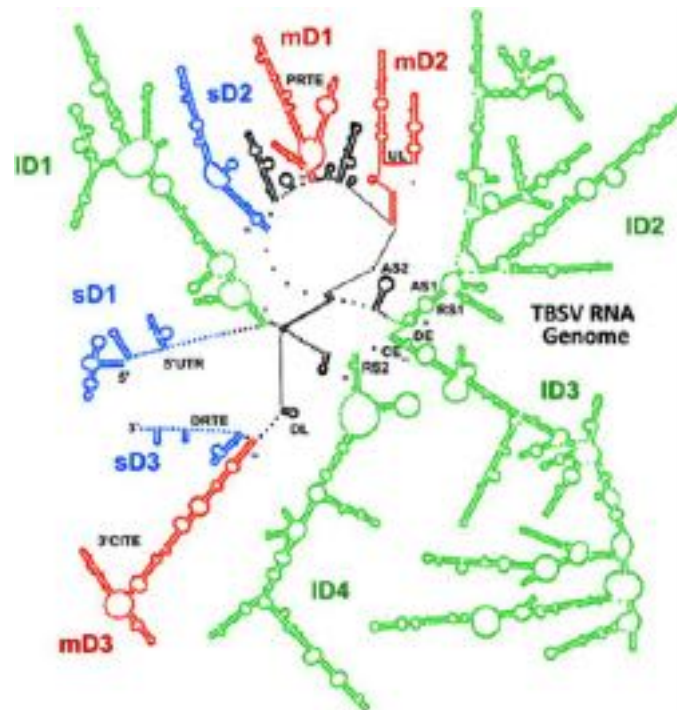


Fig. 4 RNA secondary structure model of the TBSV genome. RNA domains within the genome are labeled and color-coded according to their relative sizes: large (green), medium (red) or small (blue). Inter-domain regions are depicted in black. The RNA segments that participate in known functional long-range RNA–RNA interactions are labeled (refer to Fig. 5 for details). Only the AS1–RS1 and DE–CE interactions involved in sg mRNA transcription are base paired in this genome structure. Adapted from Wu, B., Grigull, J., Ore, M.O., Morin, S., White, K.A., 2013. Global organization of a positive-strand RNA virus genome. *PLoS Pathogens* 9 (5), e1003363.

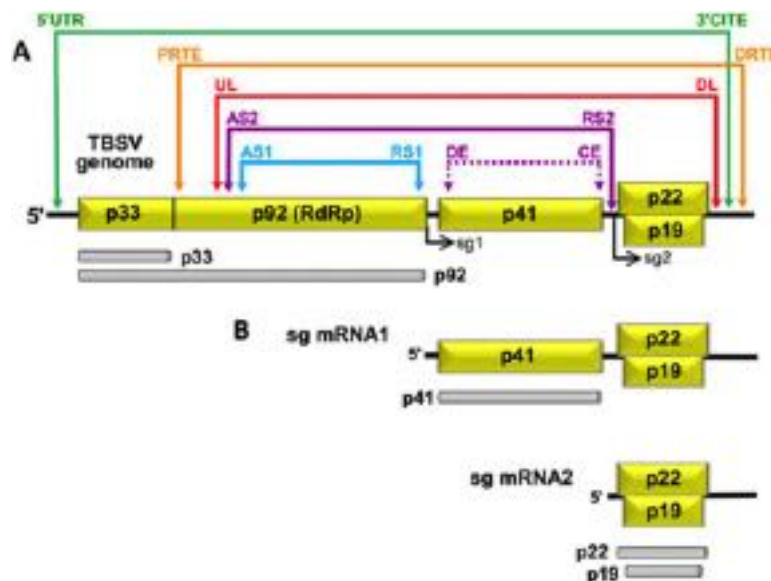


Fig. 5 TBSV RNA genome and sg mRNAs. (A) TBSV genomic RNA, with encoded proteins represented by boxes. Proteins translated directly from the genome, p33 and p92, are represented by grey bars and sg mRNA transcription start sites within the genome are labeled sg1 and sg2. Long-range RNA–RNA interactions involved in different viral processes are indicated by double-headed arrows above the genome and are color coded: 5'UTR–3'CITE, translation initiation; PRTE–DRTE, ribosome readthrough; UL–DL, genome replication; AS1–RS1, sg mRNA1 transcription; AS2–RS2, sg mRNA2 transcription. DE–CE is a second long-range interaction needed for sg mRNA2 transcription. (B) TBSV sg mRNAs. The two sg mRNAs transcribed during infections are shown along with the proteins that are translated from them.

Movement proteins are necessary for cell-to-cell movement of viral infections in plants. The local spread of an infection occurs through plasmodesmata, which are channels that connect adjacent plant cells. This process requires the assistance of a virally-encoded movement protein that facilitates transport of viral RNA to, and through, these intercellular junctions. This event requires modification of the structure of plasmodesmata to allow transit and, in tombusviruses, this action is mediated by p22. Properties of p22, such as binding to RNA and localization to cell membranes and walls, are consistent with this role. The viral genome is likely transported intra- and intercellularly as a p22-RNA complex, and may be shuttled by host factor HFi22, a homeodomain-containing protein, that interacts with p22. This concept is supported by the observation that some movement-deficient p22 mutants are no longer able to interact with HFi22.

Gene silencing is an antiviral mechanism utilized by plants. The first step in this defence mechanism occurs when long double-stranded (ds) viral RNAs generated during genome replication are bound and cleaved by the host protein dicer into small dsRNAs, termed small interfering RNAs (siRNA). Subsequently, single-strands of these siRNAs are loaded into a RNA-induced silencing complex (RISC) that uses them as guides to base pair with and cleave viral genomes. To counteract this attack, plant viruses encode suppressors of gene silencing that are able to inhibit different steps in the process. The p19 suppressor of tombusviruses functions as a homodimer that binds size-specifically to siRNAs generated by dicer. In doing so, it prevents them from being loaded into RISC, thereby preventing virus genome-specific RNA targeting and cleavage by this complex. The smaller p14 homologue in aureusvirus functions in a similar manner but, due to its lack of size-specificity, it also binds to longer dsRNAs.

Gene Expression

Gene expression in these viruses involves cap-independent translation, ribosome readthrough, sg mRNA transcription, and leaky scanning. Intriguingly, for function, all but the latter activity require RNA-RNA interactions that span large distances within the viral genomes (**Fig. 5(A)**). All tombusvirid RNA genomes lack a 5'-cap and a 3'-poly(A) tail. Instead, they contain one of several types of 3'-cap independent translational enhancer (3'-CITE) in their 3'-untranslated regions (3'-UTRs). 3'-CITEs in the tombusvirus-like group are higher-order RNA structures that bind to translation initiation factor 4F (eIF4F). These 3'-CITEs, while bound to eIF4F, also interact with the 5'-UTRs of their genomes, via a long-range RNA-RNA base pairing interactions (**Fig. 5(A)**). This positions the bound eIF4F near the 5'-end of the genomes, allowing for 43S subunit recruitment that mediates translation of the auxiliary replication proteins, p33 in the case of TBSV (and the p92 RdRp via readthrough). Though closely related, the three genera contain different types of 3'-CITEs; tombusviruses have Y-shaped versions (except for, Cucumber Bulgarian latent virus, CBLV, which has an I-shaped type), aureusviruses contain either a PTE or I-shaped class, while zeaviruses contain only I-shaped versions. Accordingly, modularity in these viral genomes also exists with respect to translational regulatory RNA elements.

Viral RdRps, such as p92 in TBSV, are expressed by translational readthrough of the stop codon for the auxiliary replication protein. Programmed ribosome readthrough occurs when RNA elements in a message promote the use of near-cognate tRNAs for decoding of a termination codon. In tombusviruses, this process requires an extended stem-loop RNA structure, termed RTSL, positioned just 3' to the stop codon. However, in order for readthrough to occur, a sequence in RTSL (termed PRTE) must base pair, via a long-range RNA-RNA interaction, with a segment in the 3'-UTR of the genome (DRTE) (**Fig. 5(A)**). This interaction shuts down minus-strand RNA synthesis during readthrough, thereby preventing a conflict between these processes that occur in opposite directions on the viral genome. Similar local and long-distance RNA structural requirements for readthrough production of RdRp are also predicted for aureusviruses and zeaviruses.

Within the context of tombusvirus-like genomes, the three ORFs encoded in their latter halves are translationally silent. Consequently, expression of these gene products requires the transcription of viral sg mRNAs. During infections, two sg mRNAs are synthesized that encode at their 5'-ends either the capsid protein ORF or the two overlapping ORFs. The capsid protein is translated from the larger sg mRNA1, while both the movement and silencing suppressor are translated from sg mRNA2 (**Fig. 5(B)**). The latter protein is translated when 43S subunits scan past the more 5'-proximally positioned movement protein start codon and initiate at the downstream suppressor protein start codon, a mechanism referred to as leaky scanning.

Transcription of sg mRNAs involves premature termination of the RdRp while synthesizing complementary genomic minus-strands. RdRp termination occurs when the viral polymerase encounters certain higher-order RNA structures in a genome, called attenuation structures. This results in truncated minus-strands that include transcription initiation sites, for either sg mRNA1 or 2, at their 3'-ends. These sg mRNA-sized minus-strands are then used as templates to transcribe the two different size classes of sg mRNAs. Interestingly, in tombusviruses and zeaviruses, the attenuation RNA structures for both sg mRNA1 and 2 are formed by long-range, base pairing, RNA-RNA interactions (AS1-RS1 and AS2-RS2, respectively) (**Fig. 5(A)**). These genome-level interactions may provide an RNA-based mechanism to coordinate transcription with other viral processes. Aureusviruses have comparable long-distance AS2-RS2 interactions for sg mRNA2 production, but, for sg mRNA1, the attenuation structure is formed by a locally-folded section of RNA. Although variations in attenuation structures exist, the sg mRNAs generated are structurally similar in that all are 3'-coterminal with their cognate genomes. This feature ensures that these smaller viral messages contain 3'-CITEs, which enhance their translation via cognate 5'UTR-3'CITE interactions.

Genome Replication

Tombusvirus genome replication has been investigated extensively in different systems, including in vitro cell extracts, plant protoplasts, whole plants, and surrogate yeast cells. These studies, particularly those carried out in yeast-based systems, have

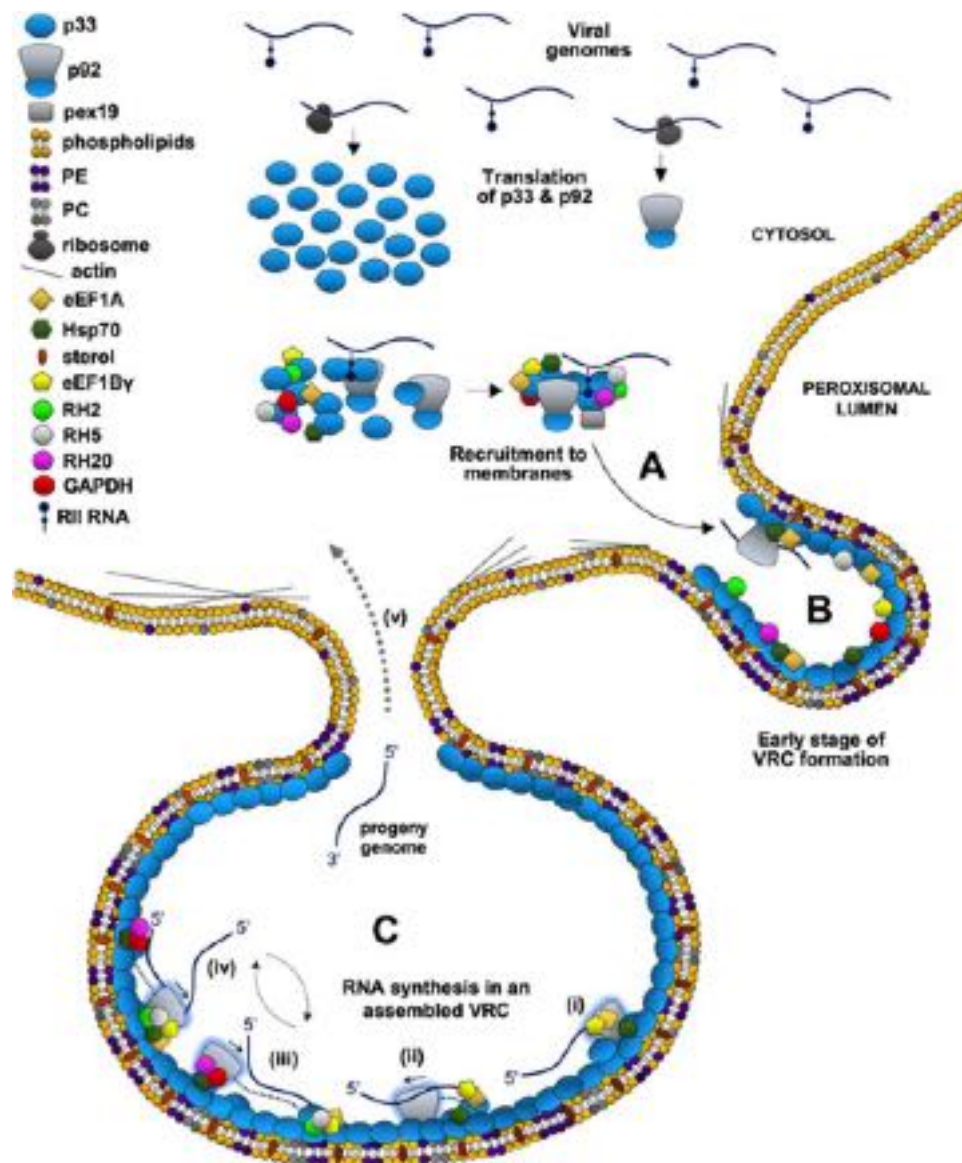


Fig. 6 Steps in VRC assembly and RNA synthesis in TBSV. (A) Peroxisome membrane targeting, (B) VRC assembly and (C) Viral RNA synthesis. Plus- and minus-strand genomes are represented as a solid and dotted lines, respectively. RdRps are shown at different steps in viral genome replication in (i) through (iv). See key for the identity of molecular components and refer to text for details. Diagram is not to scale. Adapted from Gunawardene, C.D., Donaldson, L.W., White, K.A., 2017. Tombusvirus polymerase: Structure and function. *Virus Research* 234, 74–86.

generated a large body of information that has allowed for the development of detailed models for various steps in this process. In general, the overall progression during infections can be roughly divided into three phases (1) membrane targeting, (2) VRC assembly, and (3) viral RNA synthesis. Here, these fundamental steps are summarized along with descriptions of the key viral and cellular factors involved, using TBSV as the example (Fig. 6).

At the centre of this process are the viral proteins p33 and p92, with the former being responsible for orchestrating the majority of events required for productive RNA genome replication. p33 is a membrane-anchored protein that possesses multiple domains and motifs that allow it to interact with diverse host factors that are required at different steps in the process. Also important for p33's activity is its ability to multimerize and interact with p92, and an early stage in RNA replication involves the targeting of these two viral proteins to specific intracellular membranes which, for TBSV, are peroxisomal (Fig. 6(A)). Membrane localization is dependent on N-proximal targeting sequences present in both of these proteins as well as interaction with peroxisomal biogenesis factor 19 (pex19). The viral genome also has to be shuttled to this membrane location, and this is accomplished by p92/p33 binding to an RNA structure, RII, located in the p92 coding region (Fig. 6(A)).

Assembly of VRCs rely heavily upon p33 activity. VRCs are invaginations of intracellular membranes that maintain a narrow neck with the cytosol. p33 multimers are a key structural component of these so-called spherules, and their internal compartments house the

RdRp and other factors necessary for viral RNA synthesis (Fig. 6(B)). A key step in forming these curved compartments is the hijacking of the host endosomal sorting complexes required for transport (ESCRT). p33 redirects the ESCRT system to form VRCs by interacting with different key components, such as Vps23p. The nature of the lipids within VRC membranes is also important for function. Phosphatidylethanolamine and phosphatidylcholine have stimulatory effects on VRC activity (Fig. 6(B)). To this end, p33 upregulates phospholipid production by interacting with and modulating regulatory proteins Opi1p and Scs2p. Other membrane components, such as sterols, which decrease membrane fluidity and directly bind to p33, also increase VRC function. Here again, p33 is involved in seizing control of these molecules by binding to and modifying the activity of oxysterol-binding proteins (ORPs), which are involved in controlling sterol distribution between closely associated membranes. Actin filaments also contribute to this process by stabilizing membrane contact sites, which allows for exchange of sterols from the endoplasmic reticulum (ER) to peroxisomal membranes that are destined for VRC formation. Stable actin filaments are required for this to occur, and an inhibitory interaction of p33 with cofilin, an actin depolymerization factor, ensures that a polymerized state is maintained. ER-peroxisomal intermembrane junctions thus represent critical cellular hubs for preparing appropriately composed lipid bilayers for VRC assembly.

The p92 RdRp inside assembled VRCs synthesizes a full-length complementary negative-strand of a genome that is then used repeatedly as a template to produce multiple copies of progeny genomes (Fig. 6(C)). However, prior to any RNA synthesis, p92, in association with a phospholipid-enriched membrane, has to be activated by cellular protein chaperone Hsp70. Other host proteins are also needed during genome replication. Two translation elongation factors, eEF1A and eEF1B γ , are involved in the early stages of minus-strand synthesis (Fig. 6(C)(i)). Both bind to a promoter region in 3'-end of the genome and facilitate RdRp positioning and promoter access, respectively. Additionally, p92 that is bound to the RII RNA structure in RdRp-coding region of the viral genome is repositioned to the 3'-end by a long-range RNA-RNA interaction, UL-DL (Fig. 5(A)). p92 initiates RNA synthesis at a promoter sequence located at the 3'-end of the viral genome (Fig. 6(C)(ii)). Upon completion of minus-strand genome synthesis, the RNA helicase RH20 mediates release of p92 from the dsRNA intermediate, and facilitates its rebinding to the minus-strand template. This latter step is also assisted by two other RNA helicases, RH2 and RH5, as well as glyceraldehyde 3-phosphate dehydrogenase (GAPDH), all of which interact with the minus-strand (Fig. 6(C)(iii)). Notably, the activities of these helicases, as well as Hsp70, require ATP. This requirement is met by the recruitment of glycolytic enzymes, namely ATP-generating phosphoglycerate kinase and pyruvate kinase, to the vicinity of VRCs. The resulting activities of these enzymes leads to high local levels of ATP for components of the VRCs, thereby allowing for robust virus genome replication. Newly synthesized progeny viral genomes leave VRCs and enter the cytosol via the neck of spherules (Fig. 6(C)(iv) and (v)).

Subviral RNAs

Tombusviruses are known to associate with two distinct types of subviral RNAs, satellite RNAs (satRNAs) and defective interfering RNAs (DI-RNAs). SatRNAs are naturally-occurring, noncoding, small linear RNAs that require coinfection with a standard virus, termed helper virus, for their replication and packaging. They are also dependent on their helper virus for gene silencing suppression and movement functions. Structurally, satRNAs do not share sequence identity with their helper virus. The discrete sequence of these RNAs is a paradoxical feature, because they utilize the same RdRp and capsid protein as their helpers. This suggests that they may engage viral and host factors in unique ways. SatRNAs have been found only in tombusvirus infections. Four distinct satRNAs have been identified, with the first being satRNA-Cym, which was discovered in association with Cymbidium ringspot virus (CymRSV). Subsequently, three others were identified in different TBSV infections, satRNA-B1, B10, and L. These satRNAs have lengths in the range of 600–800 nt and maintain 5'-ends that form higher-order RNA structures similar to those found in the 5'-UTRs of their helper genomes. When present in infections, satRNAs can increase, decrease, or have no effect on symptoms. Of the four tombusvirus satRNAs, only satRNA-B10 modulates infections; *i.e.*, it attenuates symptoms and prevents death of infected plants.

The other class of subviral RNA associated with tombusvirus infections are DI-RNAs. Similar to satRNAs, DI-RNAs are small, parasitic, noncoding, linear RNAs that depend on their helper virus for replication and packaging. However, unlike satRNAs, DI-RNAs are not normally found in natural infections, and they contain sequences that correspond directly to those in their helper virus genomes. DI-RNAs associated with tombusviruses were the first to be discovered for any plant virus. These 400–600 nt long RNAs contain sort genomic segments from both termini and two different internal regions (RI through RIV), and these segments are either essential (RI, RII and RIV) or greatly enhance (RII) replication of the DI RNA. This knowledge of DI-RNA structure and function also aided the identification of corresponding RNA replication elements in full-length viral genomes. Moreover, due to their small size, efficient accumulation in coinfections, and absence of coding functions, DI-RNAs have been excellent tools for studying virus RNA replication; particularly in the yeast-based system. They have also provided insights into viral RNA recombination, a process mediated by template hopping or switching by the viral RdRp during RNA synthesis. DI-RNA formation, via recombination, and their subsequent accumulation is promoted when infections are passaged at high virus titer; an artificial condition that favors coinfection of cells with helper genomes and nascently generated DI-RNAs. Although they are not found in infections in the wild, adding DI-RNAs to tombusvirus infections in laboratory experiments revealed their ability to suppress helper virus levels and the symptoms induced in plant hosts. This interference was initially believed to be due primarily to competition with the helper virus for RNA replication factors. However, tombusvirus DI-RNAs have also been shown to strongly activate gene silencing in coinfections, while being highly resistant to its effects. Thus, the disease attenuation observed could be caused by enhanced targeting and degradation of the helper genome.

A third class of subviral RNA, a satellite virus (SV), is present in some infections of the aureusvirus *Maize white line mosaic virus* (MWLMV). SV have properties similar to those of satRNAs, except that they encode their own unique capsid protein, in which their RNA is packaged. The linear RNA genome of SV-MWLMV is ~1.2 kb in length and encodes a 24 kDa capsid protein. Together, these components form spherical virions of ~17 nm in diameter, which are about half the size of the ~35 nm virions of the helper MWLMV. These particle sizes are consistent with corresponding T=1 and T=3 icosahedral symmetries. Although some SVs are known to modulate symptoms induced by their helper virus, the possible effects of SV-MWLMV on MWLMV infections have not been reported.

Symptoms and Transmission

Symptoms induced by tombusviruses range from mild to severe, depending on the host. Local effects at sites of infection include chlorotic or necrotic lesions, or mosaic patterns. TBSV infections of tomato plants cause mottling and crinkling of infected leaves, along with stunted growth. Infections that go systemic result in necrosis of apical leaves and cause deformation and necrosis of fruit. Aureusvirus infections range from symptomless, such as infections of pothos plants by PoLV, to systemic mottling and fruit streaking in cucumber infections by Cucumber leaf spot virus (CLSV). Maize infected with the MNeSV zeavirus initially exhibit pale green or yellow spots and streaks on leaves. But, as the infection progresses, the chlorotic bandings expand and merge on leaves and stalks, with some areas becoming necrotic.

Both tombusviruses and aureusviruses accumulate to high levels in infected plants and, as soil-borne viruses, are readily transmitted to healthy plants grown in virion-containing soil. Other means of natural transmission for certain members also include seed and pollen transmission. A biological vector has been identified that actively spreads certain members of both of these genera. The root-associated fungus *Olpidium bornovanus* is able to transmit the tombusvirus CuNV and the aureusvirus CLSV to healthy plants. This process has been studied extensively in CuNV and results revealed that the capsid protein is the viral determinant of transmissibility. Outer capsid residues in CuNV particles are able to interact with glycoproteins on the surface of the zoospore. Subsequent plant invasion, mediated by encystment of virus-tethered zoospores, results in virus transfer into root tissue.

No biological vectors for zeaviruses have been identified to date, although different potential candidates have been tested; including select aphids, leafhopper, planthopper, and beetles. Mechanical laboratory infection of maize plants is accomplished by performing vascular puncture inoculation, because traditional rub-inoculation of leaves that works for most tombusvirids is not effective.

Host Range and Geographic Distribution

Natural infections of different tombusvirus species are generally limited to a few hosts. Examples of such natural hosts can be gleaned from the names given to species which, for most viruses, correspond to the plant from which they were first isolated (Table 1). However, not evident from this list is the ability of some species, such as TBSV, to infect cherry and other woody hosts. Aureusviruses infect a range of herbaceous hosts, ranging from vegetable fruits, such as cucumber, to monocots, like wild grasses and cultivated maize. Zeaviruses also infect maize, however as a relatively newly discovered genus, the existence of additional natural hosts remains to be explored.

The distribution of tombusviruses is wide and includes locations in North and South America, different regions in Europe, Algeria, as well as Mediterranean areas. Reports of aureusvirus infections include countries such as Italy, Great Britain, Greece, Jordan, Bulgaria, and India. Zeaviruses were first identified in infected maize in Arizona USA, and since then there have been no additional reports. Collectively, members of the tombusvirus-like group can be considered to be dispersed worldwide.

Applied Aspects

Tombusviruses have been used as expression vectors for the efficient production of foreign proteins in plants. For example, antigens of HIV have been produced using different strategies. A complete protein was expressed by replacing the capsid protein ORF of TBSV with that of HIV p24 capsid protein. In an alternative approach, a peptide epitope of gp120 was displayed on the surface of TBSV particles by fusing its corresponding nucleotide sequence to the 3'-end of the capsid protein ORF. In both of these cases, the foreign protein and modified particle were produced in plants. A related tactic, using insect cells, was used to explore developing a potential vaccine for the poison ricin. Virus-like particles (VLPs) were produced using a baculoviral vector encoding the TBSV capsid protein containing a C-terminally fused ricin peptide. This position in the capsid, *i.e.*, an extension of the P-domain, allowed for optimal presentation of 180 copies of this epitope on the surface of particles. For all cases described above, the proteins or particles were found to be immunogenic and elicited corresponding antibody production. Though still in the proof of concept stage, such products could potentially be useful for vaccine production and diagnostics.

Whether protein expression occurs via virus vectors or agrobacterium-mediated schemes, in whole plants or plant cell suspension cultures, a common obstacle faced in producing optimal levels of a foreign protein is the plant gene silencing system. Overexpression of exogenous mRNAs often leads to targeting and degradation of these messages by this antiviral system. It has been shown, in different production schemes, that these negative effects can be mitigated by co-expressing p19. Depending on the system, doing this can increase yields of the foreign protein by a modest ~20% to over fivefold. Accordingly, further exploration of the utility of this gene silencing suppressor in helping to maximize yield in different large-scale protein production schemes appears warranted.

Another emerging field in which tombusviruses are being employed is nanoparticle technology. The existing nanoscale of self-ordering virus particles makes them attractive candidates for further manipulation. Reengineering of their structural features, such as size and physicochemical properties, can convert them into nanoscaffolds, nanocontainers, or nanobuilding blocks. Genetic and chemical manipulations of TBSV capsids have generated virus-based nanoparticles capable of being chemically derivatized or encapsulating cargo. Additionally, skeletonized TBSV particles have been assembled using multiple copies of a peptide that is a key structural feature that interconnects full-sized capsid subunits in wild type particles. Such designer nanoparticles could be utilized for a variety of purposes, ranging from molecular medicine to biomaterial science. For human applications, safety is an obvious concern, and the effects of TBSV-based nanoparticles on different biological systems is currently being assessed.

Concluding Remarks

Tombusvirus-like viruses are distributed throughout the world and can cause damaging diseases in important crops. Tombusviruses, in particular, have served as excellent model systems to understand different steps in their reproductive cycle. Indeed, members of this genus are among the best characterized plus-strand RNA viruses at the molecular level. Virion analysis has yielded novel information on particle structure and function. Studies of other viral proteins have revealed diverse and critical roles for these viral gene products. Functional analyses of viral RNA genomes have uncovered a remarkable network of intragenomic RNA-RNA interactions that regulate a variety of viral processes. This knowledge, combined with the identification of a multitude of host factors, both protein and lipid-based, has provided an extraordinarily detailed understanding of many steps in virus reproduction. Some infections have been found to contain subviral RNAs that can modulate helper virus levels and disease progression. Among these, DI-RNAs have been particularly useful tools for identifying and studying RNA elements and protein factors involved in viral RNA replication. At an applied level, tombusviruses are also being explored for use in different areas of biotechnology, such as vaccines and nanoparticle development. Future studies of this group of viruses, at both the basic and applied levels, will undoubtedly reveal other fundamental properties and uncover additional useful traits.

Further Reading

- Alam, S.B., Reade, R., Theilmann, J., Rochon, D., 2017. Evidence for the role of basic amino acids in the coat protein arm region of Cucumber necrosis virus in particle assembly and selective encapsidation of viral RNA. *Virology* 512, 83–94.
- Chkuaseli, T., White, K.A., 2018. Intragenomic long-distance RNA–RNA interactions in plus-strand RNA plant viruses. *Frontiers in Microbiology* 9, 529.
- Danielson, D.C., Pezacki, J.P., 2013. Studying the RNA silencing pathway with the p19 protein. *FEBS Letters* 587 (8), 1198–1205.
- Gunawardene, C.D., Donaldson, L.W., White, K.A., 2017. Tombusvirus polymerase: Structure and function. *Virus Research* 234, 74–86.
- Ho, P.T., Montiel-Garcia, D.J., Wong, J.J., *et al.*, 2018. VIPERdb: A tool for virus research. *Annual Review of Virology* 5 (1), 477–488.
- Jiwan, S.D., White, K.A., 2011. Subgenomic mRNA transcription in Tombusviridae. *RNA Biology* 8 (2), 287–294.
- Matsuura, K., 2018. Synthetic approaches to construct viral capsid-like spherical nanomaterials. *Chemistry Communication* 54 (65), 8944–8959.
- Nagy, P.D., 2016. Tombusvirus-host interactions: Co-opted evolutionarily conserved host factors take center court. *Annual Review of Virology* 3 (1), 491–515.
- Nagy, P.D., Pogany, J., 2006. Yeast as a model host to dissect functions of viral and host factors in tombusvirus replication. *Virology* 344 (1), 211–220.
- Nagy, P.D., Pogany, J., Xu, K., 2016. Cell-free and cell-based approaches to explore the roles of host membranes and lipids in the formation of viral replication compartment induced by tombusviruses. *Viruses* 8 (3), 68.
- Palukaitis, P., 2016. Satellite RNAs and satellite viruses. *Molecular Plant Microbe Interaction* 29 (3), 181–186.
- Scholthof, H.B., 2006. The tombusvirus-encoded P19: From irrelevance to elegance. *Nature Reviews in Microbiology* 4 (5), 405–411.
- Truniger, V., Miras, M., Aranda, M.A., 2017. Structural and functional diversity of plant virus 3′-cap-independent translation enhancers (3′-CITEs). *Frontiers in Plant Science* 8, 2047.
- White, K.A., Nagy, P.D., 2004. Advances in the molecular biology of tombusviruses: Gene expression, genome replication, and recombination. *Progress in Nucleic Acid Research and Molecular Biology* 78, 187–226.
- Wu, B., Grigull, J., Ore, M.O., Morin, S., White, K.A., 2013. Global organization of a positive-strand RNA virus genome. *PLoS Pathogens* 9 (5), e1003363.

Relevant Websites

<https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses.

Tombusviruses (*Tombusviridae*)

Luisa Rubino, Institute for Sustainable Plant Protection, National Research Council, Bari, Italy

Kay Scheets, Oklahoma State University, Stillwater, OK, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S.A. Lommel, T.L. Sit, Tombusviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Inc, 2008, doi:[10.1016/B978-012374410-4.00382-4](https://doi.org/10.1016/B978-012374410-4.00382-4).

Nomenclature

– 1FS Minus one frameshift

aa Amino acid(s)

CP Coat protein

DI RNA Defective interfering RNA

kb Kilobase(s)

MP Movement protein

nt Nucleotide(s)

PTGS Post-transcriptional gene silencing

RAP Replicase associated protein

RdRp RNA dependent RNA polymerase

S_{20,w} Sedimentation coefficient of a biomolecule in a solution normalized to acceleration in pure water at 20°C

Glossary

Citopathology The study of the cellular alterations induced by infectious agents.

Dicer Eukaryotic ribonuclease that cuts long dsRNAs and structured ssRNAs into ~20–25 bp fragments with 2-nt 5′ extensions that activate RISC (see below).

Frameshift A change of the reading frame, which results in a different translation and a different protein product.

Plasmodesmata (Singular: plasmodesma) Channels traversing the plant cell walls, allowing for the cell-to-cell transport of substances. Plasmodesmata are internally coated by the endoplasmic reticulum.

PTSG Post-transcriptional gene silencing; rapid degradation of cellular mRNAs or vRNAs triggered by siRNAs or cellular microRNAs (miRNAs).

Readthrough Readthrough consists of the “translation” of a stop codon, such that the synthesis of a larger protein is

allowed. It is a common phenomenon in viruses, and it has a regulatory function.

Riboviria A higher taxonomic rank including all viruses possessing an RNA genome.

RISC RNA-induced silencing complex; a eukaryotic multiprotein complex that recognizes single stranded microRNAs or double-stranded siRNAs to initiate the destruction of cellular mRNAs or double-stranded viral RNAs.

Silencing suppressor (Suppressor of silencing): Virally encoded proteins from diverse families that interfere with host production or effectiveness of siRNAs.

siRNAs Small interfering RNAs or silencing RNAs, 20–25 bp dsRNAs with two-nt 3′ overhanging ends that target invading viral RNAs or cellular RNAs for degradation in eukaryotic cells.

Tombusvirid Member of the family *Tombusviridae*.

Classification

Realm Riboviria, Kingdom Orthornavirae, Phylum Kitrinoviricota, Class Tolucaviricetes, Order Tolivirales, Family Tombusviridae, Subfamilies Calvirusirinae, Procedovirinae, Regressovirinae.

Taxonomy, Phylogeny and Evolution

The members of this family all have uncapped positive sense, single-stranded RNA [(+)ssRNA] genomes that are encapsidated in icosahedral virions and are related by the similarities of their RNA-dependent RNA polymerases (RdRps). The number of genera doubled between 1998 and 2016 and currently consists of 16 genera assigned to three subfamilies ([Table 1](#)). The increase occurred by subdivision of two genera (*Carmovirus* and *Necrovirus*), incorporation of genus *Umbravirus*, discovery of new unique species, and the availability of complete genome sequences of earlier members. Genus *Umbravirus* is the only genus in the subfamily *Calvirusirinae* (from the Latin “*calvus*”, naked). Umbraviruses have monopartite genomes that use a – 1FS mechanism to express their RdRps but lack a coat protein (CP) gene. The subfamily *Procedovirinae* (from the Latin “*procedo*”, go ahead) contains 14 genera and six species unassigned within the subfamily. These viruses possess a monopartite genome and use a stop codon readthrough (RT) mechanism for the expression of the RdRp. The subfamily *Regressovirinae* (from the Latin “*regressus*”, step back) contains the genus *Dianthovirus*, the members of which have a bipartite genome and use a – 1FS mechanism for RdRp expression. The continued expansion of genera in the family is partly due to the apparent ability of unrelated tombusvirids infecting the same plant to recombine via exchange of coding regions and/or untranslated regions (UTRs) that may change the host range, leading to

Table 1 List of members of the family *Tombusviridae* organized by subfamilies and genera

Subfamily/Genus	Species	Acronym	Accession No.	Length (nt)	No. sgRNAs ^a
<i>Calusvirinae</i>					
<i>Umbravirus</i>	<i>Carrot mottle mimic virus</i>	CMoMV	NC001726	4201	1
	<i>Carrot mottle virus</i>	CMoV	NC011515	4193	
	<i>Ethiopian tobacco bushy top virus</i>	ETBTv	NC024808	4236	
	<i>Groundnut rosette virus</i>	GRV	NC003603	4019	
	<i>Lettuce speckles mottle virus</i>	LSMV	ND		
	<i>Opium poppy mosaic virus</i>	OPMV	NC027710	4230	
	<i>Pea enation mosaic virus 2</i>	PEMV2	NC003853	4253	
	<i>Tobacco bushy top virus</i>	TBTv	NC004366	4152	
	<i>Tobacco mottle virus</i>	TMoV	NC043206	2159 ^a	
Unassigned umbravirus	<i>Ixeridium yellow mottle virus 2</i>	IYMoV2	NC034243	4196	
<i>Procedovirinae</i>					
<i>Alphacarmovirus</i>	<i>Angelonia flower break virus</i>	AFBV	NC007733	3962	2
	<i>Calibrachoa mottle virus</i>	CbMV	NC021926	3919	
	<i>Carnation mottle virus</i>	CarMV	NC001265	4003	
	<i>Honeysuckle ringspot virus</i>	HoRSV	NC014967	3956	
	<i>Nootka lupine vein clearing virus</i>	NLVCV	NC009017	4172	
	<i>Pelargonium flower break virus</i>	PFBV	NC005286	3923	
	<i>Saguaro cactus virus</i>	SgCV	NC001780	3879	
unassigned alphacarmovirus	<i>Adonis mosaic virus</i>	AdMV	LC171345	3991	
<i>Alphanecrovirus</i>	<i>Olive latent virus 1</i>	OLV1	NC001721	3699	2
	<i>Olive mild mosaic virus</i>	OMMV	NC006939	3683	
	<i>Potato necrosis virus</i>	PoNV	NC029900	3674	
	<i>Tobacco necrosis virus A</i>	TNVA	NC001777	3684	
<i>Aureusvirus</i>	<i>Cucumber leaf spot virus</i>	CLSV	NC007816	4431	2
	<i>Johnsongrass chlorotic stripe mosaic virus</i>	JCSMV	NC005287	4421	
	<i>Maize white line mosaic virus</i>	MWLMV	NC009533	4293	
	<i>Pothos latent virus</i>	PoLV	NC000939	4415	
	<i>Yam spherical virus</i>	YSV	NC022895	4464	
Unassigned aureusvirus	<i>Elderberry aureusvirus 1</i>	EIAV1	NC040713	1134 ^a	
<i>Avenavirus</i>	<i>Oat chlorotic stunt virus</i>	OCSV	NC003633	4114	1
<i>Betacarmovirus</i>	<i>Cardamine chlorotic fleck virus</i>	CCFV	NC001600	4041	2
	<i>Hibiscus chlorotic ringspot virus</i>	HCRSV	NC003608	3911	
	<i>Japanese iris necrotic ring virus</i>	JINRV	NC002187	4014	
	<i>Turnip crinkle virus</i>	TCV	NC003821	4054	
<i>Betanecrovirus</i>	<i>Beet black scorch virus</i>	BBSV	NC004452	3644	2
	<i>Leek white stripe virus</i>	LWSV	NC001822	3662	
	<i>Tobacco necrosis virus D</i>	TNVD	NC003487	3762	
<i>Gallantivirus</i>	<i>Galinsoga mosaic virus</i>	GaMV	NC001818	3803	2
<i>Gammacarmovirus</i>	<i>Cowpea mottle virus</i>	CPMV	NC003535	4029	2
	<i>Melon necrotic spot virus</i>	MNSV	NC001504	4266	
	<i>Pea stem necrosis virus</i>	PSNV	NC004995	4048	
	<i>Soybean yellow mottle mosaic virus</i>	SYMMV	NC011643	4009	
<i>Macanavirus</i>	<i>Furcraea necrotic streak virus</i>	FNSV	NC020469	3966	1
<i>Machlomovirus</i>	<i>Maize chlorotic mottle virus</i>	MCMV	NC003627	4437	1
<i>Panicovirus</i>	<i>Cocksfoot mild mosaic virus</i>	CMMV	NC011108	4198	1
	<i>Panicum mosaic virus</i>	PMV	NC002598	4326	
	<i>Thin paspalum asymptomatic virus</i>	TPAV	NC021705	4195	
Unassigned panicovirus	<i>Bermuda grass latent virus</i>	BGLV	NC032405	4044	
<i>Pelarspovirus</i>	<i>Clematis chlorotic mottle virus</i>	CCMoV	NC033777	3880	1
	<i>Elderberry latent virus</i>	ELV	NC026239	3892	
	<i>Pelargonium chlorotic ring pattern virus</i>	PCRpv	NC005985	3904	
	<i>Pelargonium line pattern virus</i>	PLPV	NC007017	3883	
	<i>Pelargonium ringspot virus</i>	PeRSV	NC026240	3865	
	<i>Rosa rugosa leaf distortion virus</i>	RrLDV	NC020415	3971	

(Continued)

Table 1 Continued

Subfamily/Genus	Species	Acronym	Accession No.	Length (nt)	No. sgRNAs ^a
unassigned pelarspovirus	Jasmine virus H	JaVH1	KX897157	3867	
	Jasmine mosaic associated virus 1	JMaV	MG958505	3859	
<i>Tombusvirus</i>	<i>Artichoke mottled crinkle virus</i>	AMCV	NC001339	4789	2
	<i>Carnation Italian ringspot virus</i>	CIRV	NC003500	4763	
	<i>Cucumber Bulgarian latent virus</i>	CBLV	NC004725	4576	
	<i>Cucumber necrosis virus</i>	CNV	NC001469	4701	
	<i>Cymbidium ringspot virus</i>	CyRV	NC003532	4733	
	<i>Eggplant mottled crinkle virus</i>	EMCV	NC023339	4767	
	<i>Grapevine Algerian latent virus</i>	GALV	NC011535	4731	
	<i>Havel River virus</i>	HRV	NC038690	2088 ^a	
	<i>Lato River virus</i>	LRV	ND		
	<i>Limonium flower distortion virus</i>	LFDV	NC038691	1231 ^a	
	<i>Moroccan pepper virus</i>	MPV	NC020073	4772	
	<i>Neckar River virus</i>	NRV	NC038927	1305 ^a	
	<i>Pelargonium leaf curl virus</i>	PLCV	NC030452	4789	
	<i>Pelargonium necrotic spot virus</i>	PNSV	NC005285	4744	
	<i>Petunia asteroid mosaic virus</i>	PAMV	NC038692	1238 ^a	
	<i>Sitke waterborne virus</i>	SWBV	NC038693	1194 ^a	
	<i>Tomato bushy stunt virus</i>	TBSV	NC001554	4776	
unassigned tombusvirus	Gentian virus A	GVA	LC373507	4742	
<i>Zeavirus</i>	<i>Maize necrotic streak virus</i>	MNeSV	NC007729	4094	1
unassigned <i>Procedovirinae</i>	<i>Trailing lespedeza virus 1</i>	TLV1	NC015227	3929	
	Gompholobium virus A	GomVA	NC030742	3935	
	Rice virus A	RVA	NC035452	4832	
<i>Regressovirinae</i>					
<i>Dianthovirus</i>	<i>Carnation ringspot virus</i>	CRSV	NC003531	1403	1
			NC003530	3840	
	<i>Red clover necrotic mosaic virus</i>	RCNMV	NC003775	1448	
			NC003756	3890	
	<i>Sweet clover necrotic mosaic virus</i>	SCNMV	NC003806	3876	
			NC003807	1449	
unassigned <i>Regressovirinae</i>	Rice virus X	RVX	AB033715	4486	

^aNo. sgRNAs; genus demarcation criteria uses number of protein-coding sgRNAs only; protein-coding sgRNAs only for genus demarcation.

speciation. The members of the genus *Luteovirus* (current family *Luteoviridae*) have viruses that encode frameshifting replicases that are phylogenetically similar to dianthoviruses and are likely to be moved into the family *Tombusviridae* in the near future.

This family serves as an excellent example supporting the concept of modular evolution of viruses. The RpRp is the sole shared module binding the family (Fig. 1). Within the limitations of the RdRp, the family is highly variable in the arrangement and expression of the various genes on its polycistronic RNA. While all members of the family have a T = 3 icosahedral virion, members of different genera achieve this structure using two phylogenetically distinct, but related, CP modules. The family has acquired at least three phylogenetically different cell-to-cell movement modules. The sources of the virus suppressor of host gene silencing appear to be equally diverse. In addition, some species within a particular genus have acquired additional unique accessory modules, possibly for transmission by fungal vectors or for host-specific viral movement.

Virion Structure

CP-coding tombusvirids produce non-enveloped virions, consisting of 180 CP subunits with a T = 3 icosahedral symmetry. Virus infections produce virions containing a single copy of their genome ranging in size from 3.6 to 4.8 kb for monopartite species and 5.3 kb for dianthoviruses. Dianthoviruses package one each of the genomic RNAs into one virion via a *trans*-activator (TA) loop on RNA1 binding to a complementary *trans*-activator binding site (TABS) just upstream of the subgenomic RNA (sgRNA) start site of RNA2. Genomic RNA constitutes about 17% of the particle weight.

Although the underlying size and structure of the virions are similar, the members can be broadly subdivided into two groups based on the presence or absence of a protruding (P) domain at the C-terminus of the CP. Viruses in genera containing P domains (*Alphacarmovirus*, *Aureusvirus*, *Avenavirus*, *Betacarmovirus*, *Dianthovirus*, *Gallantivirus*, *Gammacarmovirus*, *Macanavirus*, *Pelarspovirus*, and *Tombusvirus*) produce virions that are 32–38 nm in diameter and display a granular surface. Most of these CP subunits range in size from 35 to 42 kDa, while the avenavirus oat chlorotic stunt virus (OCSV) encodes a 48 kDa CP. The X-ray crystal structures of tomato bushy

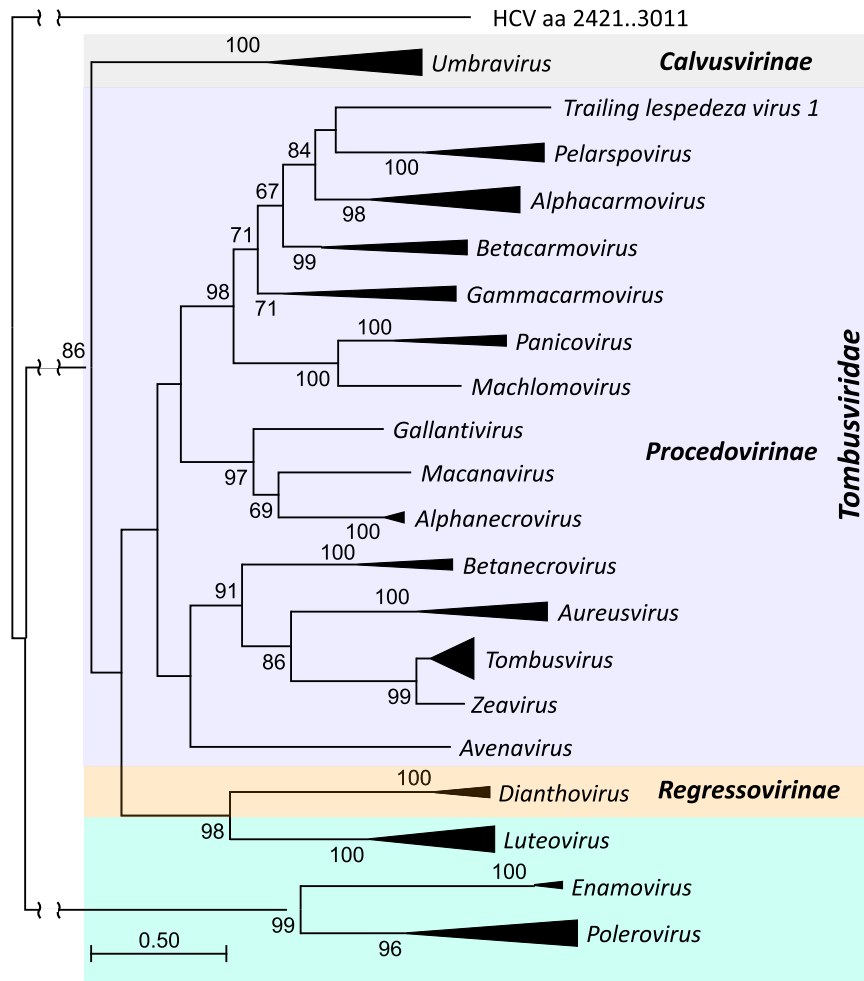


Fig. 1 Phylogenetic (distance) analysis of the complete RdRps for all sequenced tombusvirids plus selected viruses from the family *Luteoviridae* (turquoise box). Alignments of the 78 sequences were made using MUSCLE while trees were generated with the Maximum Likelihood (ML) algorithm of Mega7 using 1000 bootstrap replicates (showing values >50%). Positions with <50% site coverage were eliminated, leaving 809 positions in the final dataset. Condensed triangles mark monophyletic lineages. Brackets connecting tombusvirid branching, enamovirus plus polerovirus branching, and HCV RdRp (aa 2421–3011) were proportionately shortened. *Tombusviridae* subfamily names are on the right, and colored boxes enclose member genera. Species names are in [Table 1](#).

stunt virus (TBSV) and turnip crinkle virus (TCV) revealed the discrete organization of each CP subunit, and virion structures for many additional species have confirmed the structures for both types of CPs. Generally, the CP subunit can be divided into four domains: the RNA-binding (R) domain is located at the N-terminus followed by the arm (a), shell (S), and P domains (66, 35, 67, and 110 aa, respectively, in TBSV). The R domain contains many basic residues and is found in the interior of the virion. Cryo-electron microscopy reconstructions have further revealed the presence of internal ordered cages of RNA intertwined with CP residues from the R domains beneath the virion surface. This internal scaffold may play a role in directing specific packaging of viral RNA and formation of the icosahedral virions. The virion structure is primarily formed by the globular S domains (stabilized by a pair of Ca^{2+} per subunit) which are composed of two sets of four-stranded antiparallel β -sheets. The S domain is also the most highly conserved region of the CP subunit. There is a flexible hinge of five residues between the S and P domains that allows the CP subunit to adopt different configurations by varying the angle between the domains. This feature of the CP subunit overcomes the structural constraints imposed by the icosahedral morphology. P domains (containing antiparallel β -sheet structures in a jellyroll conformation, with one six-stranded β -sheet and one four-stranded β -sheet) of adjacent CP subunits dimerize to produce 90 projections leading to the granular surface texture. Genera lacking the P domain (*Alphanecrovirus*, *Betanecrovirus*, *Machlomovirus*, *Panicovirus* and *Zeavirus*) produce virions that are 28–32 nm in diameter with a smooth surface similar to viruses in the genus *Sobemovirus*. These smaller CP subunits range from 25 to 30 kDa in size. In natural infections, umbravirus RNA is encapsidated separately within their helper-virus CPs to produce virions of ~30 nm, similar in structure and size to tombusvirids encoding small CPs.

Due to the quasi-equivalent nature of the CP subunits when arranged in an icosahedron, each subunit can take on one of three distinct conformations, termed A, B, and C (respectively, blue, red, and green in [Fig. 2](#)). The A and B conformations differ in the angle between the S and P domains and their arrangement on the virion surface (A: fivefold axes, B: threefold axes). The C conformation

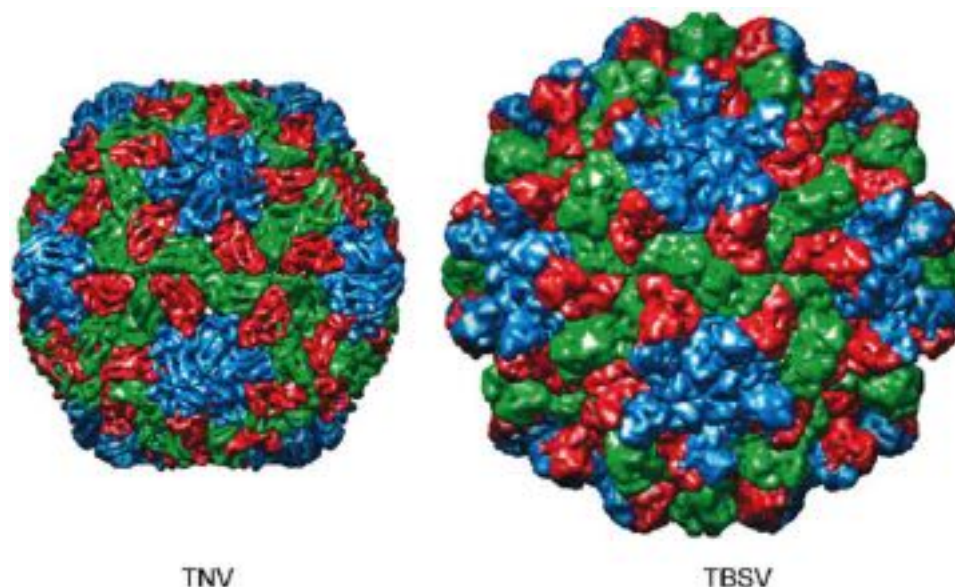


Fig. 2 Virion images based on X-ray crystal structures. (Left) Representation of a particle of tobacco necrosis virus (TNV) at 2.25 Å. The capsid protein of this virus does not contain a protruding domain and is representative of the species in the genera *Alphanecrovirus*, *Betanecrovirus*, *Machlomovirus*, *Panicovirus*, *Umbravirus*, and *Zeavirus*. (Right) Representation of a particle of tomato bushy stunt virus (TBSV) at 2.9 Å. The capsid protein of this virus contains a protruding domain and is representative of species in the remaining genera. For both particles, the individual subunits are colored according to their various conformations: A (blue), B (red), and C (green). From Shepherd, C.M., Borelli, I.A., Lander, G., *et al.*, 2006. VIPERdb: A relational database for structural virology. *Nucleic Acids Research* 34 (Database Issue), D386–D389.

differs from A and B in that their R/a domains intertwine to form an ordered internal structure around threefold axes of symmetry termed the β -annulus. The three different CP subunit conformations pack together as either AB or CC dimers in the virion particle. For TCV, virion assembly initiates with three CP dimers and viral RNA followed by formation of the virion shell according to structural constraints imposed by CP/RNA interactions. In TCV, the origin of assembly RNA sequence (which specifically initiates the interaction with CP subunits) is contained within a 186 nt region at the 3' end of the CP ORF. Assembly studies with TCV also showed that only RNAs equal to or smaller than 4.35 kb in size could be packaged, suggesting very strict size limitations for icosahedral virions.

Virions of tombusvirids have a Mr of $\sim 5.7\text{--}9.1 \times 10^6$ and produce a single, well-defined band upon centrifugation with sedimentation coefficients ranging from 118 to 140 $S_{20,w}$. Virion densities range from 1.34 to 1.36 g cm^{-3} in CsCl gradients. Virus particles are stable at acidic pH, but expand above pH 7 and in the presence of ethylenediaminetetraacetic acid. They are resistant to elevated temperatures, although thermal inactivation usually occurs above 80°C. Due to the lack of a lipid membrane, virions are insensitive to organic solvents and nonionic detergents with the exception of umbraviruses, which lack a CP ORF. In the absence of the assistant virus, plants manually inoculated with the umbravirus genome produce lipid-containing ribonucleoprotein structures, which are sensitive to organic solvents.

Genome Organization and Replication Strategy

The unifying genomic feature is the presence of an interrupted ORF for the virally encoded replicase at the 5' end of the genome with identical aminoacyl termini: a high copy replicase-associated protein (RAP) and the larger RdRp, containing the GDD (glycine aspartate) active site motif of supergroup-II RdRps. Both proteins are indispensable for virus RNA (vRNA) replication. In all members but maize chlorotic mottle virus (MCMV), the replicase gene is the most 5' ORF. Thus, incoming vRNA serves as the mRNA for RAP and RdRp. All genes downstream of the RdRp ORF are expressed from 1 or 2 sgRNAs (see [Table 1](#) and [Fig. 3](#)), which is one of the discriminating characteristics that define genera. These polymerase subunits share a high degree of sequence similarity within the family. Notably, they do not contain a helicase-type motif and the genomes do not encode any helicase-like ORF. ORFs downstream of the polymerase are expressed via 3' co-terminal sgRNAs with the resident ORFs arranged with an assortment of translational strategies such as ribosomal scanning and frameshifting. The single CP ORF is located either internally (tombusviruses, aureusviruses, and zeavirus) or in the 3'-terminal region of the genomic RNA ([Fig. 3](#)). Tombusvirid genomes also encode one or more movement proteins (MPs) that are involved in cell-to-cell spread of the virus. These range in size from the 34–35 kDa version produced by dianthoviruses to the 21–22 kDa proteins expressed by tombusviruses, aureusviruses and maize necrotic streak virus (MNeSV), and down to the 6–9 kDa pair of peptides found in carmo-like and necro-like viruses. Although no mutational analyses of OCSV have been reported, it is likely that the single 8 kDa protein gene overprinted in the CP ORF encodes an MP. In most cases, the MPs can potentiate the local movement (cell-to-cell) of unencapsidated viral RNA but systemic infection

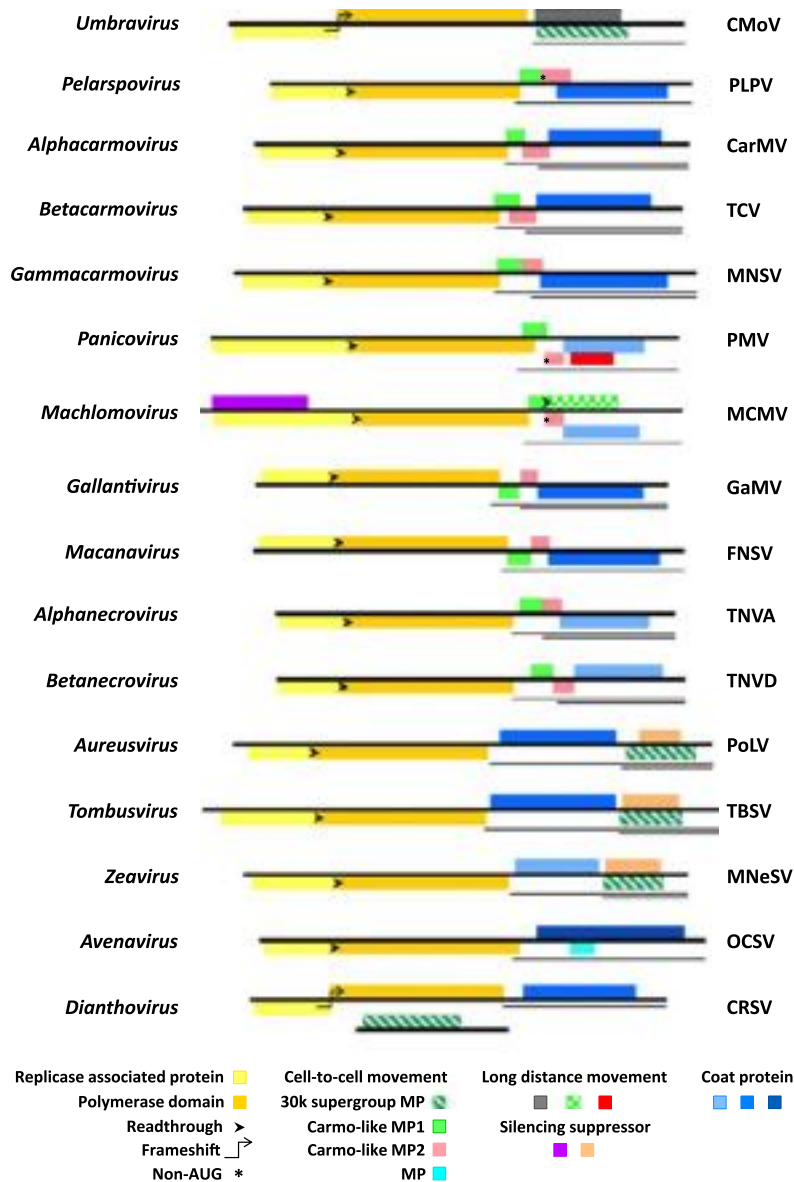


Fig. 3 Genome organizations for type species of family *Tombusviridae*. Genomes are arranged in the same order as in Fig. 2. Genus name (left) and acronym (right) of type species are shown. Genome lengths (heavy black lines) are proportional, and ORFs are color-coded for protein homology. Subgenomic RNAs are thin black lines below the genomes. Stop codon readthrough (arrowhead), frameshift (bent arrow), and noncanonical start codon (asterisk) locations are black symbols. Below the genomes, functions of gene products are identified by color-coded boxes. 30k supergroup MPs are related to tobacco mosaic virus 30 kDa movement protein.

generally requires encapsidation. The exception is for tombusviruses which can infect systemically without CP, albeit at a reduced rate. The other prominent protein product with a clearly defined function is the tombusvirus p19 protein which is a potent suppressor of PTGS. p19 binds siRNAs that are generated by a Dicer-like RNase as part of the host RNA-silencing response to viral infection. This binding prevents incorporation of the siRNAs into the RNA-induced silencing complex (RISC) to prevent further cleavage of viral RNAs. Similar to the function, structure, and mechanism of p19, aureusvirus p14 binds siRNAs, albeit less specifically. Although other members of the family may not encode a unique suppressor protein, it is known that other proteins can have suppressor activity aside from their primary functions. This is the case for TCV where the CP is the viral suppressor of PTGS. Various additional ORFs are also encoded in some tombusvirids with undefined functions. One other method utilized by tombusvirids for gene expression is genome segmentation which is only employed by dianthoviruses where RNA2 is monocistronic and encodes the MP.

Additional ORFs have been identified in the genomes of panicoviruses and MCMV. For the panicovirus PMV, the unique p15 ORF overprinted within the CP ORF is not necessary for replication but is required for systemic infection along with the CP. For the closely related MCMV, the unrelated ORF overprinted within its CP ORF is required for efficient systemic infection.

Since the genomic RNAs are positive sense, they are directly infectious as unencapsidated RNAs. vRNAs lack a 5' cap structure and the 3' ends are not polyadenylated nor do they form tRNA-like structures. Instead, tombusvirids and the closely related viruses in the luteovirus genus form several different types of complex structures at their 3' ends called cap-independent translational enhancers (3'CITEs) that interact with small or large stem loops (SLs) usually in the 5'UTRs of vRNA or sgRNAs to allow the circularization that is typically accomplished by recruitment of eukaryotic translation initiation factors (eIFs) that bind to poly(A), 5' caps, and ribosome subunits to initiate translation.

Various 3'CITEs have been identified in tombusvirid genomes; the barley yellow dwarf virus (BYDV)-like 3'CITE (BTE) of tobacco necrosis virus (TNVD), the panicum mosaic virus (PMV)-like 3'CITE (PTE) of panicoviruses and cucumber leaf spot virus (CLSV), the Y-shaped 3'CITE of TBSV and carnation Italian ringspot virus (CIRV), the PTE variant of MCMV, and the smaller I-shaped 3'CITE of the zeavirus MNeSV which is also present in some aureusviruses and carmo-like viruses. The 3'CITEs have different structures and binding properties. The direct translation readthrough or frame-shift strategies ensure that the RdRp active site is present in only 5%–10% of the vRNA translation products since readthrough and frameshifting are inefficient. Translation of both RAP and RdRp subunits is followed by localization to and proliferation of cellular membranes. The location of viral replication varies depending on the particular virus and is specified by the pre-readthrough portion of the polymerase (see Cytopathology). In all cases studied to date, the polymerase anchors on membranes, causing proliferation. The advent of yeast-based replication systems for four tombusviruses i.e. CIRV, TBSV, cucumber necrosis virus (CNV) and cymbidium ringspot virus (CymRSV) has provided valuable insights into the host components utilized by the viral replication complex.

Once the polymerase has been assembled, it binds to the 3' terminus to initiate minus-strand synthesis. If full-length copies are generated, these serve as templates for synthesis of progeny genomic RNAs. This synthesis is highly asymmetrical with (+) RNA accumulating at much higher levels than (–) RNA. This is controlled by a small base-pairing interaction between a structure at the extreme 3' end of vRNAs and sgRNAs. These are the replication silencer element (RSE) and a 3'-complementary silencer sequence (3'CSS) present in a bulge ~50–60 nt upstream which sequesters the 3' end from the RdRp. Occasionally, positive-strand synthesis will be terminated prematurely which leads to the formation of templates for plus-strand sgRNA synthesis. This premature termination is controlled by long-distance RNA–RNA interactions which occur *in cis* for monopartite tombusvirids such as TBSV, or *in trans* for dianthoviruses via the same TA/TABS base-pairing interaction used to ensure encapsidation of both RNAs inside one virion. Generally, sgRNAs are produced later in the infection cycle and behave as monocistronic mRNA or polycistronic RNAs, depending on the genus. The sgRNAs demonstrate various translational strategies such as ribosomal scanning to express the p19 ORF which is nested within the p22 ORF in TBSV. In this case, the p22 start codon is suboptimal and is occasionally read through by ribosomes before initiation at the optimal p19 start codon. MNeSV and aureusviruses express their silencing suppressor ORFs using the same mechanism. Similar ribosomal scanning past suboptimal start codons express one or more of the downstream ORFs of carmo-like viruses and necro-like viruses.

Very small noncoding viral RNAs that are coterminal with the parent virus 3' end have been identified in infections of MCMV, TCV, CNV, TNVD, and the three dianthoviruses red clover necrotic mosaic virus (RCNMV), carnation ringspot virus (CRSV), and sweet clover necrotic mosaic virus (SCNMV). RNase analyses of RNA fragments from the 3' UTRs of RCNMV, MCMV and the umbravirus opium poppy mosaic virus (OPMV) using Xrn1 *in vitro* identified a conserved pseudoknot in the 3'UTR that protects the RNA from further 5'-3' degradation. For RCNMV and TNVD, these were shown to be generated by the host 5'-3' exoribonuclease Xrn4, which is stalled by RNA secondary structure. The function of these small RNAs is not yet fully clarified. Sequence searches for program-predicted exoribonuclease-resistant RNAs (xrRNAs) followed by *in vitro* analyses identified additional Xrn-blocking structures in the 5'UTR of most umbravirus sgRNAs.

Life Cycle

For most genera, infection is initiated either by mechanical inoculation (piercing of leaf or root tissues with virus-contaminated material) or via vector transmission by a fungal vector in roots. For umbraviruses natural infections occur via viruliferous aphids in the presence of the umbravirus helper virus. After the virus enters an epidermal cell, the virion disassembles, and vRNA engages ribosomes and translation factors to synthesize RAP and RdRp. When enough newly-synthesized vRNA accumulates in initially infected cells, vRNA is transmitted via their MP(s) through plasmodesmata to neighboring cells for further rounds of replication and spread cell-to-cell. At some point newly assembled virions can move throughout the plant via vascular tissue, to produce systemic infections. Viruses can survive in debris and soil for a long time before a new infection cycle starts.

Cytopathology

Plants infected with tombusvirids display a distinctive cytopathology, observed by electron microscopy as dense-staining features. Depending on the genus, different virus species can replicate on, remodel, and proliferate membranes of peroxisomes, mitochondria and/or the cortical ER. The hallmark of infections by members of the genus *Tombusvirus* are the cytopathic structures known as multivesicular bodies, which are generated after the virus-induced progressive vesiculation and invagination either of the limiting membrane of peroxisomes or the outer membrane of mitochondria. The vesicles are believed to be sites of viral replication. Exceedingly high concentrations of progeny virions can accumulate in the multivesicular bodies, but are also arranged to form crystals in the cytoplasm.

Subviral Molecules

Three different types of subviral entities are associated with the replication of members in the family *Tombusviridae*: satellite viruses, satellite RNAs (satRNAs) and defective interfering (DI) RNAs. All depend on the presence of a helper virus for their replication, movement, and transmission and may affect helper virus infections. Whereas DI RNA sequences are identical to those of the helper genomic RNAs, satellite viruses and satRNAs share no or limited sequence identity with their cognate viruses. Satellite viruses code for their own CP. In most cases satRNAs are encapsidated by the helper virus-encoded CP, but satC and satS, which are only found in plants infected with PMV and satPMV, are encapsidated by satPMV CP.

Satellite viruses described so far are supported by viruses in the genera *Alpha*- and *Betanecrovirus* (Satellite tobacco necrosis virus, STNV), *Panicovirus* (Satellite panicum mosaic virus, SPMV) and *Aureusvirus* (Satellite maize white line mosaic virus, SMWLMV), and are classified in their own genera *Albetovirus*, *Papanivirus* and *Aumaivirus*, respectively. Satellite viruses associated with tombusvirids possess an uncapped (+) ssRNA genome of ~0.8–1.2 kb, lack a poly(A) tail, and harbor a single ORF that encodes their own CP. The subunits are arranged in a T = 1 lattice, and some satellite viruses (SPMV) harbor signals for specific packaging of the satellite virus genome, while STNV1 packaging appears to be mostly related to the abundance of appropriate-sized RNAs. As mentioned above, satellite virus particles are readily distinguished from the helper virions, due to their different size.

The STNV genome can form secondary structures for specific recognition by the helper virus RdRp, and contains a 3'-translational enhancer domain (TED) which enhances translation by interacting with a short nucleotide stretch in the 5'UTR. STNV can attenuate the symptoms induced by its helper virus, probably by competing for factors essential for replication. Conversely, SPMV results in a severe exacerbation of PMV-induced symptoms, often leading to plant death, probably due to synergy with the helper virus and to a cytotoxic effect exerted by the satellite CP. Whereas SPMV has the smallest known satellite virus genome, SMWLMV is characterized by the largest satellite virus CP. No effect on the helper virus symptomatology for SMWLMV has been reported.

The origin of nonhomologous satRNAs is unclear at present, but they can be generated from recombination events with other satRNAs, the viral genome, DI RNAs, or host sequences. Generally, satRNAs associated with tombusviruses do not share sequence homology with their helper viruses, except for the four satRNAs associated with tombusviruses, which have short conserved nucleotide stretches at the 5' and 3' ends, in addition to a ~50 nt sequence in the 5'-terminal region, in common with genomic and DI RNAs. Also the satRNAs (C, D and F) associated with the betacarrovirus TCV have a conserved heptanucleotide at the 3' end. In addition, whereas satRNA D and satRNA F are true satRNAs, satRNA C is a recombinant between satRNA D (5' region) and TCV (3' region). Other tombusvirids are also known to harbor satRNAs, like the 615-nt satRNA associated with the betanecrovirus beet black scorch virus (BBSV) and the satRNAs reported in association with umbraviruses. All the satRNAs described so far have high levels of secondary structure, which account for their stability and survival. Some predicted structures are conserved and are important for the recognition and replication of these molecules by the helper virus. The effect on virus-induced symptomatology is variable. SatRNA C enhances TCV symptoms, whereas satRNA D and F do not interfere with TCV symptomatology. BBSV satRNA exacerbates its cognate virus symptoms. The four 600–800 nt satRNAs associated with tombusviruses differently influence the viral symptoms: the satRNA associated with CyRSV and TBSV satRNAs B1 and L do not modulate the helper virus infection, whereas TBSV satRNA B10 attenuates symptom expression. The satRNAs associated with the umbraviruses groundnut rosette virus (GRV) and tobacco bushy top virus (TBTv) are essential for the encapsidation and transmission of their helpers by an assistant virus.

Many tombusvirids generate DI RNAs that can affect pathogenesis. DI RNAs are essentially small, non-coding deletion mutants of the viral genome which are generated by errors in replication such as rearrangement or recombination. Recombination can occur between both homologous and nonhomologous sequences, probably by a copy-choice mechanism in which the viral RdRp jumps from one template to another during the replication process. They usually consist of genomic-derived 5' and 3' non-coding sequences and some internal regions, which are essential for their replication by the virus' RdRp. DI RNAs are more readily generated at high multiplicity of infection. Several species such as TBSV and TCV produce DI RNAs during viral infection. For TBSV the presence of DI RNAs results in attenuation of symptoms, whereas for TCV the presence of DI RNAs intensifies symptoms. The attenuation of symptoms is not only due to competition for host factors, but largely resides in the highly structured DI molecules, which can trigger post transcriptional gene silencing against the helper virus genome.

Both DI RNAs and satRNAs have been useful molecular tools for studying viral replication and recombination. Since both satRNAs and DI RNAs are dependent on the host or parent virus for replication, they must retain any sequence or structural signals required for recognition by the viral polymerase, and therefore can help determine those signals.

Recombination can facilitate viral evolution as well as repair of viral genomes, so it is an important factor in the virus life cycle, and it can have profound effects on pathogenesis. In the tombusvirus family, recombination in the carmo-like and tombus-like viruses has been shown to repair damaged or deleted 3' ends of virus-associated RNAs, as well as generate new satRNAs or DI RNAs.

Pathogenesis

Depending on the host and on the genus, tombusvirid infections can be symptomless or consist of mottling, crinkling, mosaic, distortion, chlorotic and/or necrotic lesions on the leaves, and stunting of the plants. The natural host range of most species is relatively narrow; however, the experimental host range tends to be broad. Members can naturally infect either monocotyledonous or dicotyledonous plants. Fewer tombusvirids infect monocots: eight infect grasses (Family Poaceae) while four infect members of the Order *Asparagales*. Natural infections can be limited to or are often concentrated in the root system. Many species are found in the soil, where

they survive for a long time. All species are readily transmitted by host vegetative propagation and mechanical inoculation (although some species cannot be inoculated by leaf-rub inoculation), and some may be transmitted by contact and through seeds. A number of species are readily detected in surface waters, rivers, lakes, and even the ocean. Transmission by chytrid fungi in the genus *Olpidium* and by beetles has also been reported for members of several genera. For viruses that can be transmitted by fungi, there is a considerable amount of specificity required between the virus and the fungal vector. For example, the tombusvirus CNV, the gammacarmovirus melon necrotic spot virus (MNSV), and the aureusvirus CLSV are transmitted by the root-inhabiting chytrid fungus *Olpidium bornovanus*, whereas the alphacarmovirus TNVA, and betanecroviruses TNVD and BBSV are transmitted exclusively by *Olpidium brassicae*. Umbraviruses are aphid-transmitted, but only when plants are co-infected with an “assistant” luteovirid. Many members can be transmitted through the soil either dependent on, or independent of, a biological vector. This rather unusual soil and water mode of transmission is unique to this plant virus family and is based on the unusually robust constitution of the virion.

Geographical distribution of particular species varies from wide to restricted and is strongly linked to the presence of the specific host. Most species occur in temperate regions although legume-infecting gammacarmoviruses and furcraea necrotic spot virus (FNSV) have been found in tropical areas. Two alphacarmoviruses occupy more extreme environments: Nootka lupine vein clearing virus (NLVCV) grows in Alaska, USA (northern temperate climate above the arctic circle), and saguaro cactus virus (SgCV) was found in a hot desert in Arizona, USA. MCMV was initially found in South America, but it is now one of the most widespread family members due to the extreme increase in maize (*Zea mays*)-producing nations throughout tropical and temperate zones.

Further Reading

- Chkvaseli, T., White, K.A., 2018. Intragenomic long-distance RNA-RNA interactions in plus-strand RNA plant viruses. *Frontiers in Microbiology* 9, 529. doi:10.3389/fmicb.2018.00529.
- Gunawardene, C.D., Newburn, L.R., White, K.A., 2019. A 212-nt long RNA structure in the Tobacco necrosis virus-D RNA genome is resistant to Xrn degradation. *Nucleic Acids Research* 47, 3929–3942.
- Kim, M., Roossinck, M.J., 2017. Small linear satellite RNAs. In: Hadidi, A., Flores, R., Randles, J.W., Palukaitis, P. (Eds.), *Viroids and Satellites*. London, UK: Elsevier Academic Press, pp. 567–575.
- Kotta-Loizou, I., Peyret, H., Saunders, K., Coutts, R.H.A., Lomonosoff, G.P., 2019. Investigating the biological relevance of in vitro-identified putative packaging signals at the 5′ terminus of Satellite tobacco necrosis virus 1 genomic RNA. *Journal of Virology* 93, e02106–e02118.
- May, J.P., Yuan, X., Sawicki, E., Simon, A.E., 2018. RNA evasion of nonsense-mediated decay. *PLoS Pathogens* 14, e1007459.
- Miller, W.A., Shen, R., Staplin, W., Kanodia, P., 2016. Noncoding RNAs of plant viruses and viroids: Sponges of host translation and RNA interference machinery. *Molecular Plant-Microbe Interactions* 29, 156–164.
- Nagy, P.D., 2017. Exploitation of a surrogate host, *Saccharomyces cerevisiae*, to identify cellular targets and develop novel antiviral approaches. *Current Opinion in Virology* 26, 132–140.
- Nagy, P.D., Strating, J.R., van Kuppeveld, F.J., 2016. Building viral replication organelles: Close encounters of the membrane types. *PLoS Pathogens* 12, e1995912.
- Navarro, J.A., Pallas, V., 2017. An update on the intracellular and intercellular trafficking of carmoviruses. *Frontiers in Plant Sciences* 8, #1801.
- Pyle, J.D., Scholthof, K.-B.G., 2017. Biology and pathogenesis of satellite viruses. In: Hadidi, A., Flores, R., Randles, J.W., Palukaitis, P. (Eds.), *Viroids and Satellites*. London, UK: Elsevier Academic Press, pp. 627–636.
- Rubino, L., 2017. Biology of satellites. In: Hadidi, A., Flores, R., Randles, J.W., Palukaitis, P. (Eds.), *Viroids and Satellites*. London, UK: Elsevier Academic Press, pp. 649–658.
- Rubino, L., Russo, M., Martelli, G.P., 2014. Tombusvirus-induced multivesicular bodies: Origin and role in virus-host interaction. In: Gaur, R.K., Hohn, T., Sharma, P. (Eds.), *Plant Virus-Host Interaction. Molecular Approaches and Evolution*. Oxford, UK: Elsevier Academic Press, pp. 163–175.
- Scheets, K., Jordan, R., White, K.A., Hernandez, C., 2015. *Pelarspovirus*, a proposed new genus in the family *Tombusviridae*. *Archives of Virology* 160, 2385–2393.
- Serra-Soriano, M., Navarro, J.A., Pallas, V., 2017. Dissecting the multi-functional role of the N-terminal domain of the Melon necrotic spot virus coat protein in RNA packaging, viral movement and interference with antiviral plant defence. *Molecular Plant Pathology* 18, 837–849.
- Steckelberg, A.L., Vicens, Q., Kieft, J.S., 2018. Exoribonuclease-resistant RNAs exist within both coding and non-coding subgenomic RNAs. *mBio*. 9, e02461.

Relevant Websites

- <https://talk.ictvonline.org/taxonomy/>
 Virus Taxonomy: 2019 Release.
 International Committee on Taxonomy of Viruses (ICTV).

Tritimoviruses and Rymoviruses (*Potyviridae*)

Satyanarayana Tatineni, Agricultural Research Service, US Department of Agriculture, Lincoln, NE, United States and University of Nebraska–Lincoln, Lincoln, NE, United States

Gary L Hein, University of Nebraska–Lincoln, Lincoln, NE, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)

CP Coat protein or capsid protein

ELISA Enzyme-linked immunosorbent assays

HC-Pro Helper component-protease

kb Kilobase

kDa Kilo dalton

Mr Relative molecular mass

nm Nanometer(s)

nt Nucleotide(s)

NTR Non-translated region

ORF Open Reading Frame

PAGE Polyacrylamide gel electrophoresis

PCR Polymerase chain reaction

qRT-PCR Quantitative reverse transcription polymerase chain reaction

RdRp RNA dependent RNA polymerase

SDS-PAGE Sodium dodecyl sulfate polyacrylamide gel electrophoresis

ssRNA single-stranded RNA

VPg Viral protein genome-linked

Glossary

Cleavage Processing of large proteins into smaller peptides at specific amino acid sequences by proteases.

Motif A short amino acid or nucleotide sequence that is presumed to have a biological function through sequence-specific binding.

Polyprotein A large protein with a chain of covalently joined smaller peptides that can be cleaved by virus-encoded proteases into smaller proteins with different biological functions.

Virion The protective shell of genomic RNA/DNAs of viruses covered mostly with virus-encoded capsid and often with other virus-encoded accessory proteins.

Introduction

The members of the genus *Tritimovirus* have previously been included in the genus *Rymovirus* based on biological properties such as monocotyledonous hosts and transmission by eriophyid mites. However, phylogenetic studies based on partial or complete genome sequences of *Rymovirus* members revealed that the eriophyid mite transmission characteristic was a paraphyletic trait within the genus *Rymovirus*. Hence, Wheat streak mosaic virus (WSMV) and Brome streak mosaic virus (BrSMV) were separated from the genus *Rymovirus* and classified into the new genus *Tritimovirus* with WSMV as the type species. Subsequently, four other definitive species were assigned to the genus *Tritimovirus*. Among the members of genera *Tritimovirus* and *Rymovirus*, WSMV is the most economically important virus on wheat in most wheat growing regions of the world, causing significant yield losses. WSMV was first reported from Nebraska, USA as wheat mosaic disease, and disease caused by WSMV was referred as wheat streak mosaic disease (Fig. 1).

Taxonomy and Classification

The family *Potyviridae* contains a total of ten genera, plus three unassigned species, and includes *Tritimovirus* and *Rymovirus* as distinct genera. Some members of the *Rymovirus* and *Tritimovirus* genera are transmitted by eriophyid mites *Abacarus hystrix* (Nalepa) and *Aceria tosichella* (Keifer) (Fig. 2(A)), respectively; however, members of *Rymovirus* are phylogenetically more related to aphid-transmitted members of the genus *Potyvirus*. Recently, there was an unsuccessful proposal to assimilate the genus *Rymovirus* into the genus *Potyvirus* as a rymovirus subgroup based on phylogenetic relatedness. Currently, the genus *Rymovirus* consists of three known definitive members (*Agropyron mosaic virus*, *Hordeum mosaic virus*, and *Ryegrass mosaic virus*) with *Ryegrass mosaic virus* (RGMV) as the type species, while the genus *Tritimovirus* consists of six recognized species (*Brome streak mosaic virus*, *Oat necrotic mottle virus*, *Tall oatgrass mosaic virus*, *Wheat eglid mosaic virus*, WSMV, *Yellow oat grass mosaic virus*) with *Wheat streak mosaic virus* as the type species (Table 1).

Virion Structure

Virions of *Tritimovirus* and *Rymovirus* members are flexuous filaments of 690–720 nm in length and 11–15 nm in diameter (Fig. 2(B)) made up of 95% protein and 5% RNA. The molecular weight (MW) of rymo- and tritimoviral CP subunits range between 27 and

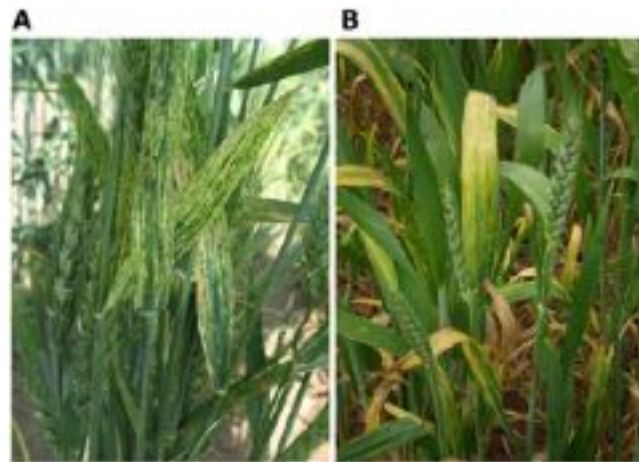


Fig. 1 Wheat streak mosaic virus-infected wheat plants showing chlorotic streaks (A) and mosaic symptoms (B) under field conditions. Photographs are courtesy of Mary Burrows, Montana State University, USA and the Government of Western Australia, <https://www.agric.wa.gov.au>.

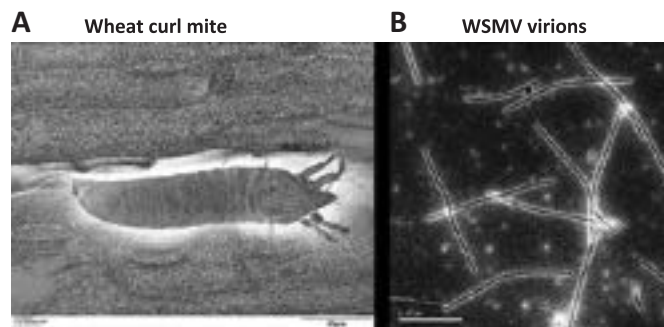


Fig. 2 Scanning electron micrograph of an adult wheat curl mite (*Aceria tosichella* Keifer) on wheat leaf (A). Wheat curl mite micrograph picture credit: G. Bauchan and R. Ochoa, USDA-ARS. Electron micrograph picture showing flexuous filamentous virion particles of Wheat streak mosaic virus. Bar indicates 500 nm (B).

Table 1 Genome features of members of *Rymovirus* and *Tritimovirus* genera in the family *Potyviridae*

Genus/Species	Acronym	Vector	Genome size (nt)	5'-NTR ^a (nt)	Polyprotein size; amino acids # and MW (kDa)	3'-NTR (nt)	Coat protein MW (kDa)	Accession #
<i>Rymovirus</i>								
<i>Agropyron mosaic virus</i>	AgMV	<i>A. hystrix</i>	9540	131	3079; 349	173	37.8	NC_00590
<i>Hordeum mosaic virus</i>	HoMV	NVR ^b	9463	131	3051; 345	180	34.2	NC_005904
<i>Ryegrass mosaic virus</i> *	RGMV	<i>A. hystrix</i>	9522	112	3087; 348	163	35.5	NC_001814
<i>Tritimovirus</i>								
<i>Brome streak mosaic virus</i>	BrSMV	<i>A. tosichella</i>	9672	145	3094; 348	244	34.1	NC_003501
<i>Oat necrotic mottle virus</i>	ONMoV	NVR	9346	130	3024; 344	144	36.0	NC_005136
<i>Tall oatgrass mosaic virus</i>	TOGMV	NVR	9359	124	3030; 345	146	35.9	NC_022745
<i>Wheat eglid mosaic virus</i>	WEMV	NVR	9636	137	3110; 352	170	38.6	NC_009805
<i>Wheat streak mosaic virus</i> *	WSMV	<i>A. tosichella</i>	9384	130	3036; 344	147	36.7	NC_001886
<i>Yellow oat grass mosaic virus</i>	YOGMV	NVR	9292	128	3007; 341	144	34.8	NC_024471

^aNon-translated region.

^bNo vector reported.

Source: ^cexcluding polyA tail.

39 kDa (Table 1) with differences mainly due to variable lengths at the N-terminal region. The internal core of the CP residues is likely involved in virion architecture. The N- and C-terminal aa are exposed on the surface of the viral particles, and these exposed aa most likely interact with other viral, host, and vector proteins to successfully complete the virus life-cycle. The conserved N-terminal DAG-motif found in potyviral CPs that is required for aphid transmission is not present in most rymo- and tritimoviral CPs.

The helical virion assembly of WSMV suggests that each virion particle possesses ~2000 CP subunits surrounding the RNA with 7–8 subunits per turn. The aa sequence suggests the MW of WSMV CP to be 37 kDa; however, SDS-PAGE predicts the molecular weight to be 45 kDa. Partially purified virions of WSMV resolved into a 45 kDa major CP, plus 32 and 29 kDa minor CP bands. These minor proteins are part of the 45 kDa CP. The exact nature of the production of these additional CP bands is not known. However, these bands are thought to be produced due to proteolytic activity in senescence leaves.

Genome Organization

The genomic organization of *Rymovirus* and *Tritimovirus* members is similar to those in the genus *Potyvirus*. *Rymovirus* and *Tritimovirus* members possess a positive-sense single-stranded RNA (ssRNA) with sizes ranging from 9292 to 9672 nt (Table 1). The genomic RNAs of these viruses likely contain a viral protein genome-linked (VPg) at the 5' terminus and a poly-A-tail at the 3' terminus. The genome encodes a single ORF coding for a polyprotein with 3007 to 3110 aa with a predicted MW of 341–352 kDa protein (Table 1). The polyprotein is processed into mature proteins by three virus-encoded proteinases: P1, HC-Pro and NIa-Pro. The serine proteinase P1 and cysteine proteinase HC-Pro cleave the polyprotein *in cis* at their C-terminal ends, while the serine proteinase NIa-Pro cleaves the polyprotein *in trans* at other heptapeptide cleavage sites. Thus, cleavage of the polyprotein results in 10 mature proteins. An additional small ORF, PIPO, is expressed by the polymerase slippage mechanism from the P3 cistron as P3N-PIPO. The 5' and 3' non-translated regions (NTRs) in all members range from 112–145 and 144–244 nt, respectively (Table 1). The NTRs are predicted to contain the *cis*-acting elements for replication and translation enhancers. The mechanism of genomic RNA translation in rymo- and tritimoviruses is not known. However, translation of genomic RNA potentially takes place through the cap-independent leaky scanning mechanism similar to members of potyviruses.

Functions of Viral Proteins

The mature polyproteins encoded by members of the genera *Rymovirus* and *Tritimovirus* consist of the following proteins from the N-terminus to C-terminus: P1, HC-Pro, P3, P3N-PIPO, 6K1, CI, 6K2, NIa-VPg, NIa-Pro, NIb, and CP (Fig. 3). WSMV is the most extensively studied virus outside the genus *Potyvirus* in the family *Potyviridae*. Hence, functions described for viral proteins in this section are mostly based on studies on WSMV.

The P1 protein is a serine proteinase and is the most divergent among WSMV isolates. The P1 protein of WSMV and Oat necrotic mottle virus (ONMoV) but not Agropyron mosaic virus (AgMV) and Hordeum mosaic virus (HoMV) is a suppressor of RNA silencing and a pathogenicity enhancer (Fig. 3). In contrast to the multi-functional helper component proteinase (HC-Pro) of potyviruses, the HC-Pro from WSMV is dispensable for systemic infection, but it is required for wheat curl mite transmission. WSMV HC-Pro is predicted to be a virus-encoded accessory protein of virion component similar to members of the genus *Potyvirus*. The HC-Pro proteins of AgMV and HoMV of the genus *Rymovirus* are suppressors of RNA silencing. The P3 protein is required for replication, while P3N-PIPO of WSMV is required for cell-to-cell movement. The P3N-PIPO is highly conserved among WSMV isolates and mutating P3N-PIPO without affecting the P3 coding sequence severely debilitates cell-to-cell movement of WSMV.

The exact role of 6K1 in virus biology is not known though deletion of this cistron is lethal to WSMV. The CI protein is the largest of all virus-encoded proteins. CI proteins of rymo- and tritimoviruses possess RNA helicase activity and an ATP binding domain. These domains are predicted to form the viral replication complex. The CI is also predicted to be involved in cell-to-cell movement similar to potyviral CI proteins. The 6K2 is another small peptide present in all rymo- and tritimoviruses and is thought to be involved in viral replication, movement and pathogenicity.

The ~49 kDa NIa protein is cleaved into an N-terminal VPg (~23 kDa) and a C-terminal proteinase (Pro; ~26 kDa) (Fig. 3). The VPg protein is covalently attached to the 5' end of viral genomic RNA and is predicted to act as a primer during replication. VPg is also predicted to interact with host factors for translation, long-distance movement, and symptom development. NIa-Pro, a serine protease, is responsible for the cleavage and release of mature proteins from the polypeptide except for P1 and HC-Pro. The NIa-Pro cleaves *in cis* at its C-terminus for its own release from the polypeptide, followed by *in trans* cleavage at specific heptapeptide cleavage sites. The NIa-Pro of WSMV was identified as an elicitor of superinfection exclusion. NIa and NIb proteins form inclusion bodies in the nucleus, hence the name 'nuclear inclusion' for these proteins. The NIb protein is an RNA-dependent RNA polymerase (RdRp) and is responsible for viral replication through interactions with other viral and host proteins.

The C-terminal product of the polyprotein is the CP with multifunctional roles. The CP, in addition to virion encapsidation, is also involved in wheat curl mite transmission, cell-to-cell and long-distance movement, pathogenicity, and superinfection exclusion (Fig. 3). The C-terminal region of WSMV CP is dispensable for virion assembly, but it is required for cell-to-cell movement. The CP aa 36–84 are dispensable for virion assembly and systemic infection, but deletions comprising aa 57–84 elicited more severe symptoms compared with wild type virus. The CP aa 57–100 are identified as viral determinants along with HC-Pro for wheat curl mite transmission.

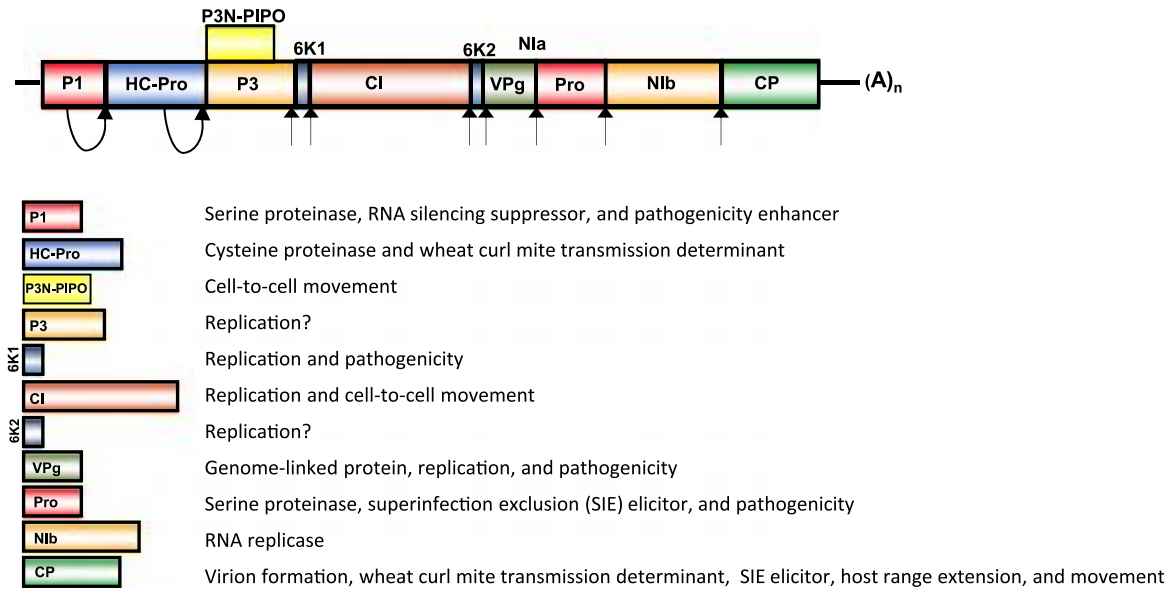


Fig. 3 Schematic representation of genomic organization of Wheat streak mosaic virus (WSMV). The large rectangle indicates an open reading frame encoding a large polyprotein that is cleaved *in cis* by P1 and HC-Pro (indicated with curved arrows). The remainder of the polyprotein is predicted to be cleaved by Nla-Pro (indicated with upward arrows) *in cis* at its C-terminal end and *in trans* between P3/6K1, 6K1/CI, CI/6K2, 6K2/Nla-VPg, Nla-VPg/Nla-Pro, Nla-Pro/Nib, and Nib/CP at heptapeptide cleavage sites. Boxes below the genomic organization indicate proteins encoded by WSMV and their putative functions in virus biology.

Life Cycle and Epidemiology

Host Range

The known host range of rymo- and tritimoviruses is restricted to monocotyledonous plants including several cereals and other grass species. Among members of rymo- and tritimoviruses, the host range of WSMV has been most extensively examined. Wheat (*Triticum aestivum*) and corn (*Zea mays*) are the most important hosts for WSMV, but natural WSMV infection of corn has been mostly controlled with the introduction of hybrid maize cultivars resistant to WSMV. The host range of WSMV also includes oats (*Avena sativa*), barley (*Hordeum vulgare*), rye (*Secale cereale*), foxtail millet (*Setaria italica*), and proso millet (*Panicum miliaceum*). In addition to cereals, a number of annual and perennial grasses serve as hosts for WSMV. Many of these plants also serve as hosts for wheat curl mite, the vector of WSMV.

Vector Transmission

Some members of both the rymo- and tritimoviruses are transmitted by eriophyid mites. RGMV is efficiently transmitted by the cereal rust mite (*A. hystris*), but this vector also transmits AgMV inefficiently. No vector has been found for HoMV, the third member of the genus *Rymovirus*. The mechanisms and viral determinants of rymoviral transmission by *A. hystris* are not known.

Among members of the genus *Tritimovirus*, WSMV and BrSMV are the only viruses with known vectors, and these two are transmitted by the wheat curl mite *Aceria tosichella* (Keifer) (Fig. 2(A)). Transmission of WSMV has been extensively studied. The HC-Pro and CP of WSMV are identified as viral determinants of mite transmission. Only immature mites can acquire the virus in as little as 15 min with improved transmission efficiencies as acquisition time increases up to 16 h. Mites remain viruliferous through molting and as adults for at least seven days. Thus, WSMV transmission by wheat curl mites is not stylet or foregut borne but exhibits several characteristic features of persistent transmission. However, the short acquisition time and lack of a latent period for WSMV transmission by wheat curl mites are not typical of persistent transmission. This incongruence could be due to the microscopic morphologies of these mites that could impact feeding and internal mite-virus interactions. Overall, WSMV transmission by wheat curl mites does appear to fit the category of persistent transmission with some differences.

The remaining viruses in these two genera have not been associated with a particular vector. Recent work has demonstrated that both *A. tosichella* and *A. hystris* are each a complex of several genetic variants. Some variants within these two species have distinct, sometimes unique, host relationships within the grasses. Additionally, some of these variants have been shown to have differing virus transmission capacity with WSMV, but little is known about the virus capabilities of many of these variants. Further research is needed to delineate the vectors of several of the viruses in these two genera, and perhaps the genetic variants of these two mite species should be considered as potential vectors for these remaining viruses.

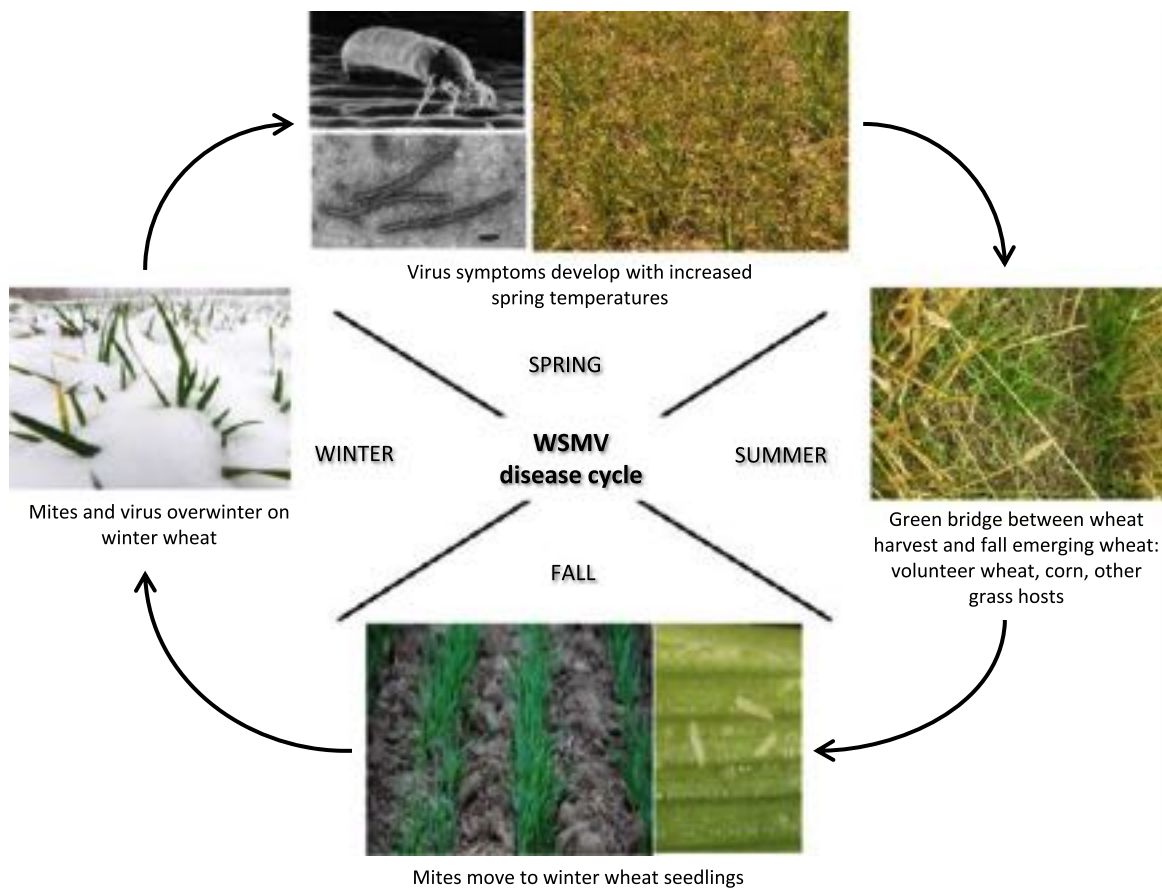


Fig. 4 Wheat streak mosaic virus (WSMV) disease cycle in winter wheat. Wheat curl mites disperse from maturing wheat onto summer green bridge hosts; wheat curl mites move from green bridge hosts onto newly emerged wheat seedlings during the fall and transmit WSMV; mites and virus overwinter in winter wheat; as temperature warms up in the spring, virus symptoms develop and impact wheat yield.

Seed Transmission

No information on seed transmission of rymo- and tritimoviruses is available except for WSMV. Seed transmission of WSMV was first reported at 0.1% level in maize in seed production lots from Iowa. WSMV transmission through seed in eight wheat genotypes was reported at 0.2%–0.5%. However, seed transmission of up to 1.5% was reported in individual wheat genotypes. Low seed transmission of WSMV may not be significant in establishing epidemics in areas where the virus is prevalent, but wheat plants infected through seed transmission can serve as a source of virus inoculum for the secondary spread through wheat curl mites. Seed transmission of WSMV is epidemiologically most significant with the possibility of virus spread through global trading and movement of infected seed.

Disease Cycle

In eriophyid mite-transmitted diseases, the bridging of hosts between two cropping seasons plays an important role in continuing the disease cycle. The epidemiology and disease cycle for wheat curl mite-transmitted WSMV have been extensively studied. Knowledge and management of the WSMV disease cycle (Fig. 4) likely apply to other viruses in the genera *Rymovirus* and *Tritimovirus*.

Due to the lack of well-developed legs, wheat curl mites rely on wind currents to disperse within and between the fields. Mites move from the leaves and stems to wheat spikes in headed wheat fields for further propagation and protection. As the wheat crop matures with the drying of heads and foliar tissue, the mites must find new hosts to survive during the summer. This is particularly critical as the mites cannot survive for more than a day or two off a living host. These hosts consist of volunteer wheat, corn and other cereal crops, and summer grasses that form the green bridge for mites and viruses between the summer wheat harvest and emergence of the fall-planted crop. Mites from WSMV-infected volunteer wheat and other cereal and grass hosts can move onto wheat seedlings during the fall season via wind. These mites propagate rapidly due to their rigorous reproductive capacity and transmit WSMV to wheat seedlings. In general, warmer fall conditions, early wheat planting, and presence of a green bridge to host the mites and virus between summer harvest and emergence of the fall crop increase virus epidemics in susceptible wheat cultivars. Overwintering of mites will occur as eggs, larvae, nymphs, and adults in wheat and other plant hosts, and WSMV will survive in

live wheat leaves and crown of virus-infected plants. Serious epidemics result from extensive mite presence and virus transmission occurring through the fall with the most serious symptoms and impact developing with the advent of warm temperatures in the spring. As temperatures warm up in the spring season, mites become active, multiply rapidly, and eventually move to the summer green bridge just before harvest to continue the disease cycle (Fig. 4).

Pathogenicity

Knowledge of how rymo- and tritimoviruses elicit disease is scant except for WSMV. The availability of reverse genetics system for WSMV facilitated the identification of viral determinants involved in disease development. The P1 of WSMV was identified as a suppressor of RNA silencing and a pathogenicity enhancer. Potato virus X infection of transgenic *N. benthamiana* plants expressing WSMV P1 protein resulted in more severe symptoms compared to non-transgenic *N. benthamiana* plants. Wheat plants infected by WSMV with complete deletion of HC-Pro elicited mild symptoms compared with wild-type virus. Surprisingly, WSMV harboring deletions comprising aa 57–84 in CP elicited more severe symptoms in multiple cereal hosts with increased accumulation of viral genomic RNA copies. In members of the genus *Potyvirus*, the 5'- and 3' NTRs have been identified as symptom determinants as mutations in these regions affected symptom severity. However, it is possible that these mutations might have affected viral *cis*-acting elements located in the 5' and 3' NTRs for replication. Additionally, P3, CI, 6K2, and NIa-VPg of different potyviruses have been identified as pathogenicity enhancers. However, no information is available on the role of these proteins in disease development of rymo- and tritimoviruses. Co-infection of wheat by unrelated WSMV and Triticum mosaic virus (TriMV), a *Poacevirus* in the family *Potyviridae*, exacerbates the disease phenotype with enhanced accumulation of both interacting viruses. Interactions between WSMV and TriMV during early stages are asymmetrical because prior infection of wheat by TriMV caused accelerated cell-to-cell movement and genomic RNA accumulation of WSMV but not vice-versa. The mechanisms and viral and host factors responsible for synergistic interaction between WSMV and TriMV are not known.

Diagnosis

The signature pinwheel-shaped inclusion bodies and flexuous filamentous virion particles are diagnostic features of potyviral infections. However, the requirement of sophisticated methodologies and equipment limit the use of these characteristic features as diagnostic methods for potyviruses. Most of the purified virions of rymo- and tritimoviruses are immunogenic in rabbits for the production of polyclonal antibodies. The polyclonal antibodies raised against purified virions have been widely used for the detection of rymo- and tritimoviruses in direct antigen coating, double antibody sandwich, and triple antibody sandwich forms of enzyme-linked immunosorbent assays (ELISAs). ELISA is less expensive and can be used for large-scale detection of samples, but the sensitivities of ELISAs depend on the quality of antibodies. The availability of genome sequences of rymo- and tritimoviruses facilitated the development of RT-PCR and real-time RT-PCR (RT-qPCR) methods for sensitive detection of viruses to overcome the limitations of ELISA-based detection methods. Additionally, multiplex RT-PCR methods allow for the simultaneous detection of multiple viruses. RT-qPCR has been used for absolute quantification of WSMV genomic RNA copies to study virus biology and virus-host, virus-vector and virus-virus interactions.

Disease Management

WSMV is the most economically important virus among eriophyid mite-transmitted rymo- and tritimoviruses. Hence, extensive research on the development of management practices for WSMV in wheat has been performed. Viruliferous mites from over-summering hosts are the primary source of infection. Hence, cultural practices targeted at managing mite populations and deployment of plants resistant to eriophyid mites, viruses, or both are needed for effective management of rymo- and tritimoviruses.

Cultural Practice

Cultural practices developed for the management of WSMV are most likely applicable to other viruses in these two genera, though some fine tuning might be required. Due to the secluded nature of mite feeding, management of wheat curl mites through pesticide application is ineffective. The main cultural management practice for WSMV is the elimination of the green bridge by removing volunteer wheat and other cereals and grasses that host both virus and vector through the summer, thus preventing the disease cycle from continuing into the fall and the next wheat crop. Mite populations are at their peak just before wheat harvest, so any volunteer wheat growing at this time is rapidly infested with mites and if not controlled will serve as an optimal green bridge host for mites and virus. Severe WSMV outbreaks are likely if hailstorms occur prior to wheat harvest and increase volunteer wheat growth. The overlap of emerged winter wheat in the fall with adjacent late-maturing crops such as maize, sorghum and small grain cover crops should also be avoided. These late-maturing crops act as reservoirs for virus and vector. Additionally, avoiding early planting of winter wheat will reduce infestation potential and the period of favorable fall temperatures for mite and virus development. Thus, effective management of WSMV depends on an integrated disease management approach by optimizing cultural practices in combination with genetic resistance.

Genetic Resistance

Except for WSMV, little or no information is available on genetic resistance against rymo- and tritimoviruses. The deployment of genetically resistant wheat cultivars with nonallelic *Wsm1*, *Wsm2*, or *Wsm3* genes has been used for WSMV disease management. Wheat cultivar Mace with the *Wsm1* gene, a single dominant R gene, associated with chromosome 4D from intermediate wheatgrass [*Thinopyrum intermedium* (Host) Barkworth & D.R. Dewey] provides resistance to WSMV and TriMV. The *Wsm2* gene, a single dominant gene, present in wheat cultivars RonL, Snowmass, Clara CL, and Oakley CL is resistant to WSMV but not to TriMV. The *Wsm2* gene is associated with chromosome 3BS with unknown origin. Wheat cultivars with *Wsm1* or *Wsm2* genes provide resistance through temperature-dependent impairment of viral long-distance movement; however, for both of these genes, resistance begins to break down as temperatures increase above 20°C. *Wsm3* is a true resistance gene that is transferred from wheatgrass (*T. intermedium*) to provide resistance against WSMV at higher temperatures. However, commercial wheat cultivars with *Wsm3* gene are not yet available.

Eriophyid mites not only transmit rymo- and tritimoviruses, but they can also cause yield losses in crop plants due to their great reproductive and feeding capabilities. Thus, the development of crop plants resistant to eriophyid mites could minimize losses from viruses as well as feeding damage by eriophyid mites. Thus far, genetic resistance to wheat curl mite has been reported for four curl mite colonization (*Cmc*) genes. The *Cmc1/Cmc4*, *Cmc2*, or *Cmc3* genes in wheat have resulted from DNA transfer from *Aegilops tauschii*, *A. elongatum*, and *S. cereal*, respectively. However, with the extensive deployment of wheat cv. TAM107 (*Cmc3* gene), wheat curl mite populations were able to adapt to this resistance gene. Thus, the durability of genetic resistance against wheat curl mites has been questioned. Wheat cultivars with multiple lines of resistance through gene pyramiding may provide more durable resistance to wheat curl mites. Additionally, pyramiding mite-resistance with virus-resistance could provide an effective management tool.

Mutation of recessive genes encoding the eukaryotic translation initiation factor 4E or its isomers to prevent interaction with viral proteins without being lethal to host plants provides resistance against several members in the genus *Potyvirus*. However, there is no information on recessive resistance against rymo- and tritimoviruses. With the advent of CRISPR-Cas9 technology, mutation of candidate recessive genes in crop plants can be achieved to obtain resistance against rymo- and tritimoviruses.

Biotechnological Application

Plant viruses can be used for the expression of specialty products and for functional genomics of host genes through virus-induced gene silencing. WSMV has an excellent reverse genetics system that can be exploited for recombinant protein expression, vaccine production, and drug delivery as a nano-agent. WSMV has been used to express GUS, GFP, and RFP proteins in wheat to examine virus movement, virus-host, and virus-virus interactions. Lastly, small peptides can be fused to CP or HC-Pro to integrate into virions that can be used for biotechnological purposes.

A major limitation of using plant viruses as expression vectors is the instability of inserted sequences since many plant viruses recombine by deleting all or parts of inserted sequences. However, inserted GFP or RFP in the WSMV genome was stably expressed for more than 120 days with no indication of deletions or reassortment in inserted sequences. Additionally, WSMV stably maintained and simultaneously expressed two foreign genes inserted between the P1/HC-Pro and N1b/CP cistrons for more than 30 days in wheat.

WSMV has been used as an expression vector to express TriMV cistrons in wheat to examine virus-virus interactions. The P1 protein of WSMV is a strong suppressor of RNA silencing. Consequently, even though many biotechnological applications are impaired by the host defense mechanism, WSMV P1 can be used to attain consistent high-level expression of transgenes in monocotyledonous plants. Additionally, transgenic plants expressing WSMV P1 can be used to enhance the levels of expression of viral vectors.

Concluding Remarks

Viral gene functions, viral proteins interacting with host and vector proteins, and mechanisms of suppression of host defense have been extensively studied for members of the genus *Potyvirus*. Among rymo- and tritimoviruses, extensive research has been performed particularly on WSMV regarding genetic diversity and the roles of HC-Pro and CP in vector transmission and disease development. However, knowledge of host and vector proteins required for disease development and transmission, respectively, is lacking. Knowledge of viral gene functions and interaction of viral proteins with host and vector factors could eventually facilitate the development of new management strategies. Identification of candidate host factors involved in disease development could be targeted via gene-editing technology to interrupt virus-host interactions. Ultimately, disruption of virus-host interactions could mitigate the viruses' ability to infect plants. This could be an important development especially for WSMV, the most important virus found in these two genera.

Further Reading

- French, R., Stenger, D.C., 2003. Evolution of Wheat streak mosaic virus: Dynamics of population growth within plants may explain limited variation. *Annual Review of Phytopathology* 41, 199–214.
- French, R., Stenger, D.C., 2005. Genome sequences of Agropyron mosaic virus and Hordeum mosaic virus support reciprocal monophyly of the genera *Potyvirus* and *Rymovirus* in the family *Potyviridae*. *Archives of Virology* 150, 299–312.

- Gibbs, A.J., Gibbs, M.J., 2018. Rymovirus: A cautionary tale. *Archives of Virology* 163, 815–817.
- Rabenstein, F., Seifers, D.L., Schubert, J., French, R., Stenger, D.C., 2002. Phylogenetic relationships, strain diversity and biogeography of tritimoviruses. *Journal of General Virology* 83, 895–906.
- Revers, F., Garcia, J.A., 2015. Molecular biology of potyviruses. *Advances in Virus Research* 92, 101–199.
- Sing, K., Wegulo, S.N., Skoracka, A., Kundu, J.K., 2018. Wheat streak mosaic virus: A century old virus with rising importance worldwide. *Molecular Plant Pathology* 19, 2193–2206.
- Skoracka, A., Dabert, M., 2010. The cereal rust mite *Abacarus hystrix* (Acari: Eriophyoidea) is a complex of species: Evidence from mitochondrial and nuclear DNA sequences. *Bulletin of Entomological Research* 100, 263–272.
- Skoracka, A., Kuczynski, L., de Mendonca, R.S., *et al.*, 2012. Cryptic species within the wheat curl mite *Aceria tosichella* (Keofer) (Avari: Eriophyoidea), revealed by mitochondrial, nuclear and morphometric data. *Invertebrate Systematics* 26, 417–433.
- Stenger, D.C., French, R., 2004. Complete nucleotide sequence of Oat necrotic mottle virus: A distinct *Tritimovirus* species (family *Potyviridae*) most closely related to Wheat streak mosaic virus. *Archives of Virology* 149, 633–640.
- Stenger, D.C., Hall, J.S., Choi, I., French, R., 1998. Phylogenetic relationships within the family *Potyviridae*: Wheat streak mosaic virus and Brome streak mosaic virus are not members of the genus *Rymovirus*. *Phytopathology* 88, 782–787.
- Stenger, D.C., Hein, G.L., Gildow, F.E., Horken, K.M., French, R., 2005a. Plant virus HC-Pro is a determinant of eriophyid mite transmission. *Journal of Virology* 79, 9054–9061.
- Szydło, W., Hein, G.L., Denizhan, E., Skoracka, A., 2015. Exceptionally high levels of genetic diversity in wheat curl mite (Acari: Eriophyoidea) populations from Turkey. *Journal of Economic Entomology* 108, 2030–2039.
- Tatineni, S., Graybosch, R.A., Hein, G.L., Wegulo, S.N., French, R., 2010. Wheat cultivar-specific disease synergism and alteration of virus accumulation during co-infection with Wheat streak mosaic virus and Triticum mosaic virus. *Phytopathology* 100, 230–238.
- Tatineni, S., Hein, G.H., 2018. Genetics and mechanisms underlying transmission of Wheat streak mosaic virus by the wheat curl mite. *Current Opinion in Virology* 33, 47–54.
- Valli, A.A., Gallo, A., Rodamilans, B., Lopez-Moya, J.J., Garcia, J.A., 2018. The HC-Pro from *Potyviridae* family: An enviable multitasking helper component that every virus would like to have. *Molecular Plant Pathology* 19, 744–776.
- Wang, A.M., Krishnaswamy, S., 2012. Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *Molecular Plant Pathology* 13, 795–803.
- Ward, C.W., 2017. Is it time to retire the genus *Rymovirus* from the family *Potyviridae*? *Archives of Virology* 162, 2175–2179.
- Wosula, E.N., McMechan, A.J., Oliveira-Hofman, C., Wegulo, S.N., Hein, G.L., 2016. Differential transmission of two isolates of Wheat streak mosaic virus by five wheat curl mite populations. *Plant Disease* 100, 154–158.

Relevant Websites

<https://talk.ictvonline.org/taxonomy/>

Taxonomy

International Committee on Taxonomy of Viruses (ICTV).

Triviruses (*Betaflexiviridae*)

Yahya ZA Gaafar and Heiko Ziebell, Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa	Amino acid(s)	nm	Nanometer
AlkB	Alpha-ketoglutarate-dependent dioxygenase	nt	Nucleotide(s)
CP	Coat protein	ORF	Open reading frame
dsRNA	Double stranded ribonucleic acid	OTu-Pro	Ovarian tumor-like protease
Hel	Helicase	Pol	RNA-dependent RNA polymerase
HTS	High-throughput sequencing	Poly(A)	Polyadenylated
IC-RT-PCR	Immunocapture reverse transcription polymerase chain reaction	P-Pro	Papain-like protease
kb	Kilobase	qRT-PCR	Quantitative reverse transcription polymerase chain reaction
kbp	Kilobase pair	RP	Replicase
kDa	Kilo Dalton	RT-PCR	Reverse transcription polymerase chain reaction
LAMP	Loop-mediated isothermal amplification	sgRNA	Sub-genomic RNA
ML	Maximum-likelihood	ssRNA	Single-stranded ribonucleic acid
MP	Movement protein	TGB	Triple gene block
Mtr	Methyl-transferase	V-region	Variable region

Glossary

Capillovirus	From Latin <i>capillus</i> (hair).	Prunevirus	From <i>Prunus</i> , genus name of the apricot (<i>Prunus armeniaca</i>), the natural host of the type species.
Chordovirus	From chord, the name of the strings on a harp or stringed instruments.	Tepovirus	From the type species <i>Potato virus T</i> .
Citivirus	From the type species <i>Citrus leaf blotch virus</i> .	Trichovirus	From Greek <i>thrix</i> (hair).
Divavirus	From the type species <i>Diuris virus</i> .	Trivirus	A virus member of the <i>Trivirinae</i> subfamily.
Fivivirus	From <i>Ficus</i> – <i>Vitis</i> , the genera name of the natural hosts of the type species in this newly proposed virus genus.	Vitivirus	From <i>Vitis</i> , generic name of the grapevine (<i>Vitis vinifera</i>) the host of the type species.
		Wamavirus	From the type species <i>Watermelon virus A</i> .

Introduction

The subfamily *Trivirinae* of the family *Betaflexiviridae* consists of viruses infecting predominantly woody plants such as fruit trees and grapevine but infection of herbaceous plants (in particular tuber crops) is also possible. Transmission of triviruses occurs through insect vectors, in particular pseudococcid mealybugs, aphids and erineum mites; transmission by grafting, seeds or mechanical means are also common. Triviruses are distributed worldwide; they often occur in mixed infections with other viruses (e.g., closteroviruses) so that it is difficult to correlate the symptomatology with the identity of the virus causing these symptoms. Triviruses have flexuous, non-enveloped particles (ranging in size from 640 to 1500 nm) that contain a positive-sense single-stranded RNA (ssRNA) molecule. The genome lengths range from approximately 6.45 to 8.75 kb. The name of the subfamily reflects the three conserved proteins that all members' genomes commonly encode, i.e., the replicase (RP), movement (MP) and coat proteins (CP). Currently, nine genera are recognized within the *Trivirinae* subfamily as well as thirty-eight assigned species although a range of newly proposed species is pending approval with constantly increasing numbers. Moreover, a new genus tentatively named "*Fivivirus*" was proposed recently.

Classification

Trivirinae is a subfamily within the family *Betaflexiviridae* (order: *Tymovirales*, realm: *Riboviria*). Nine genera are currently assigned to this subfamily i.e., *Capillovirus*, *Chordovirus*, *Citivirus*, *Divavirus*, *Prunevirus*, *Tepovirus*, *Trichovirus*, *Vitivirus*, and *Wamavirus*, with thirty-eight assigned virus species (Table 1). Several new members have been proposed recently, including Agave tequilana leaf virus (ATLV), Grapevine virus M (GVM) and Peach virus M (PeVM) (Table 1). Additionally, a new genus "*Fivivirus*" to include the tentative species i.e., Fig latent virus 1 (FLV1) and Grapevine Kizil Sapak virus (GKSV) was proposed recently. The name of the subfamily reflects the three conserved proteins that all members' genomes encode in common i.e., the RP, MP, and CP (Fig. 1). This distinguishes them from the members of the *Quinvirinae*, which have five conserved genes encoding the RP, triple gene block

Table 1 Overview of the genera assigned of the subfamily *Trivirinae* (*Betaflexiviridae*) including type species and biological properties (ICTV, 2019)

Genus	Particle size	Virus name	Acronym	Natural Hosts	Symptoms	Transmission	Distribution	Accession ^a
<i>Capillovirus</i>	670–700 × 12 nm	<i>Apple stem grooving virus^b</i>	ASGV	<i>Actinidia chinensis</i> , <i>A. deliciosa</i> , <i>Bambusa nana</i> , <i>Bauhinia variegata</i> , <i>Citrus</i> spp., <i>Cnidium officinale</i> , <i>Dendrocalamus dianxiensis</i> , <i>D. hamiltonii</i> , <i>Fargesia somnigensis</i> , <i>Ficus palmata</i> , <i>Malus</i> spp., <i>Nandina domestica</i> , <i>Phyllostachys pubescens</i> , <i>Prunus</i> spp., <i>Pyrus communis</i> , <i>P. pyrifolia</i> , <i>Rosa</i> spp. & <i>Rubus ellipticus</i>	Stem grooving/Graft Necrosis & Pitting/Tatter Leaf/Leaf Deformation/Chlorotic mottle/Leaf Epinasty	Graft- & seed-transmitted (lily, <i>Chenopodium quinoa</i>)	Worldwide in apple & citrus producing countries	NC001749
		<i>Cherry virus A</i>	CVA	<i>Prunus</i> spp.	Symptomless/Mixed Infection	Graft-transmitted	Worldwide	NC003689
		<i>Currant virus A</i>	CuVA	<i>Ribes rubrum</i>	Unknown	Unknown	Czech Republic	NC029301
		<i>Mume virus A</i>	MuVA	<i>Prunus mume</i>	Chlorotic Spots	Graft-transmitted	Japan	NC040568
		<i>Birch capillovirus^c</i>	BCV	<i>Betula pubescens</i> & <i>B. pendula</i>	Symptomless/Mixed Infection	Graft-transmitted	Finland & Germany	MK402233
		<i>Camellia ringspot associated virus 3^c</i>	CRSaV3	<i>Camellia</i> spp.	Foliar Chlorotic & Necrotic Ringspots/Rough Vein/Flower Variegation/Mixed Infection	Seed-transmitted	USA	MK050795
		<i>Loquat virus A^c</i>	LoVA	<i>Eriobotrya japonica</i>	Unknown/Mixed Infection	Unknown	China	MK936045
		<i>Rubber tree virus 1^c</i>	RTV1	<i>Hevea brasiliensis</i>	Unknown	Unknown	China	MN047299
		<i>Yacon virus A^c</i>	YVA	<i>Smallanthus sonchifolius</i>	Unknown	Unknown	Poland	NC030657
<i>Chordovirus</i>	640–760 × 10–12 nm	<i>Carrot Ch virus 1^b</i>	CtChV1	<i>Daucus carota</i>	Unknown	Unknown	UK	NC025469
		<i>Carrot Ch virus 2</i>	CtChV2	<i>Daucus carota</i>	Unknown	Unknown	UK	NC025468
		<i>Lettuce chordovirus 1^c</i>	LeCV1	<i>Lactuca sativa</i> & <i>L. serriola</i>	Mosaic/Symptomless	Unknown	France	NC040627
<i>Citrivirus</i>	960 × 12–15 nm	<i>Citrus leaf blotch virus^b</i>	CLBV	<i>Actinidia chinensis</i> , <i>Citrus</i> spp., <i>Nagami kumquat</i> , <i>Paeonia lactiflora</i> & <i>Prunus avium</i>	Abnormal Graft/Leaf Blotching	Graft-transmitted	Worldwide	NC003877
		<i>Citrus leaf blotch virus 2^c</i>	CLBV2	<i>Citrus tamuranua</i>	Leaf Chlorotic Blotching	Unknown	China	MH144344
<i>Divavirus</i>	Unknown	<i>Diuris virus A^b</i>	DVA	<i>Diuris pendunculata</i>	Unknown	Unknown	Australia	NC019029
		<i>Diuris virus B</i>	DVB	<i>Diuris pendunculata</i>	Unknown	Unknown	Australia	NC019030
		<i>Hardenbergia virus A</i>	HarVA	<i>Hardenbergia comptoniana</i>	Unknown	Mechanically-transmitted	Australia	NC015395
		<i>Ocimum basilicum RNA virus 1^c</i>	ObRV1	<i>Ocimum basilicum</i>	Unknown	Unknown	Italy	NC035462

<i>Fivivirus</i> ^a ~ 700 nm	<i>Fig latent virus</i> _{f,b,d}	FLV1	<i>Ficus carica</i>	Mosaic/Stunting/Symptomless	Mechanically & seed-transmitted	Worldwide	FN377573
	Grapevine Kizil Sapak virus ^c	GKSV	<i>Vitis vinifera</i>	Symptomless	Unknown	Turkmenistan	MN172165
<i>Prunivirus</i> 950–1500 × 13 nm	<i>Actinidia seed borne latent virus</i>	ASbLV	<i>Actinidia</i> spp.	Vein Chlorosis/Mottle/Symptomless	Seed-transmitted	China, New Zealand & South Korea	NC040800
	<i>Apricot vein clearing associated virus</i> ^b	AVCaV	<i>Prunus</i> spp.	Unknown	Unknown	Australia, China, Europe	NC023295
	<i>Caucasus prunus virus</i>	CPrV	<i>Prunus amygdalus</i>	Vein Chlorotic Spots/Leaf Reddening	Unknown	Azerbaijan	NC038325
	Camellia ringspot associated virus 1 ^c	CRSaV1	<i>Camellia</i> spp.	Foliar Chlorotic & Necrotic Ringspots/ Rough Vein/Symptomless/Mixed Infection	Seed-transmitted	USA	MK050792
	Camellia ringspot associated virus 2 ^c	CRSaV2	<i>Camellia</i> spp.	Foliar Chlorotic & Necrotic Ringspots/ Rough Vein/Symptomless/Mixed Infection	Seed-transmitted	USA	MK050793
<i>Tepovirus</i> 640 × 12 nm	<i>Potato virus</i> ^{Tb}	PVT	<i>Chenopodiaceae, Fabaceae, Solanaceae</i> & other herbaceous hosts	Vein Necrosis/Chlorotic Spots/Top Necrosis/Symptomless	Mechanically- & seed-transmitted	South America	NC011062
	<i>Prunus virus</i> T	PrVT	<i>Prunus</i> spp.	Unknown	Unknown	Azerbaijan & Italy	NC024686
	<i>Zostera virus</i> T ^c	ZVT	<i>Zostera muelleri</i>	Unknown	Unknown	Australia	MK514426
<i>Trichovirus</i> 680–780 × 9.5–12 nm	<i>Apple chlorotic leaf spot virus</i> ^b	ACLSV	<i>Chanenomeles japonica, Cydonia oblonga, Malus domestica, Prunus domestica, P. persica, Pyrus communis & Sorbus aucuparia</i>	Mottle/Ringspot/Line Patterns/Fruit Deformation/Symptomless	Mechanically- & graft-transmitted	Worldwide	NC001409
	<i>Apricot pseudo-chlorotic leaf spot virus</i>	APsCLSV	<i>Prunus</i> spp.	Stem Pitting/Stem Grooving/Symptomless	Mechanically- & graft-transmitted	Asia, Australia & Europe	NC006946
	<i>Cherry mottle leaf virus</i>	CMLV	<i>Prunus</i> spp.	Chlorotic Mottle/Distortion & Puckering/ Late Ripening/Fruit Wart-like Outgrowths/Symptomless	Mechanically-, graft-transmitted & by <i>Eriophyes inaequalis</i> eriophyid mites	China, Canada & USA	NC002500
	<i>Grapevine berry inner necrosis virus</i>	GINV	<i>Vitis</i> spp.	Leaf Mottle/Ring Spot/Mosaic/Necrotic Spots	Mechanically-, graft-transmitted & by erineum mite (<i>Colomerus vitis</i>)	China & Japan	NC015220
	<i>Grapevine Pinot gris virus</i>	GPGV	<i>Vitis riparia</i> & <i>V. vinifera</i>	Mottle/Leaf Chlorosis/Leaf Deformation/ Symptomless	Erineum mite (<i>Colomerus vitis</i>)	Worldwide	NC015782
	<i>Peach mosaic virus</i>	PcMV	<i>Prunus</i> spp.	Retarded Foliation/Leaf & Fruit Deformation/Stunting /Flower Breaking	Mechanically-, graft-transmitted & by peach bud mite (<i>Eriophyes insidiosus</i>)	Canada, Mexico & USA	NC011552

(Continued)

Table 1 Continued

Genus	Particle size	Virus name	Acronym	Natural Hosts	Symptoms	Transmission	Distribution	Accession ^a
		<i>Phlomis mottle virus</i>	PhMV	<i>Phlomis</i> spp.	Mosaic/Yellow Flecking/Stunting	Mechanically-transmitted	Greece & Italy	NC043412
		Cherry latent virus 1 ^c	CLV1	<i>Prunus</i> spp.	Symptomless	Graft-transmitted	Georgia	MK770441
		Peach chlorotic leaf spot virus ^c	PCLSV	<i>Prunus persica</i>	Unknown	Unknown	China	MH084695
		Peach virus M ^c	PeVM	<i>Prunus</i> spp.	Yellow Mottle/Chlorotic Ringspot	Unknown	Mexico, USA	MK012336
<i>Vitivirus</i>	725–825 × 12 nm	<i>Actinidia virus A</i>	AcVA	<i>Actinidia</i> spp.	Vein Chlorosis/Flecking/Ringspots/ Symptomless	Mechanically- & graft-transmitted	China, New Zealand & South Korea	NC043087
		<i>Actinidia virus B</i>	AcVB	<i>Actinidia</i> spp.	Leaf Vein Chlorosis/Flecking/Ringspots/ Vein Clearing/Mottle/Symptomless	Mechanically- & graft-transmitted	China, New Zealand & South Korea	NC016404
		<i>Arracacha virus V</i>	AVV	<i>Arracacia xanthorrhiza</i>	Mild Mosaic Chlorosis/Leaf Deformation	Unknown	Brazil	NC034264
		<i>Blackberry virus A</i>	BVA	<i>Rubus</i> spp.	Unknown	Unknown	USA	NC040630
		<i>Grapevine virus A^b</i>	GVA	<i>Vitis vinifera</i>	Stem Pitting & Grooving/Mixed infections	Mechanically-, graft-transmitted & by mealybugs	Worldwide	NC003604
		<i>Grapevine virus B</i>	GVB	<i>Vitis</i> spp.	Corky Bark/Symptomless	Mechanically-, graft-transmitted & by mealybugs	Worldwide	NC003602
		<i>Grapevine virus D</i>	GVD	<i>Vitis vinifera</i>	Corky Rugose Bark	Mechanically- & graft-transmitted	Brazil & Europe	NC038326
		<i>Grapevine virus E</i>	GVE	<i>Vitis</i> spp.	Unknown	Mechanically- & graft-transmitted	Worldwide	NC011106
		<i>Grapevine virus F</i>	GVF	<i>Vitis vinifera</i>	Unknown	Unknown	Worldwide	NC018458
		<i>Grapevine virus G</i>	GVG	<i>Vitis vinifera</i>	Unknown	Unknown	New Zealand & USA	NC040554
		<i>Grapevine virus H</i>	GVH	<i>Vitis vinifera</i>	Symptomless	Unknown	Worldwide	NC040545
		<i>Grapevine virus I</i>	GVI	<i>Vitis vinifera</i>	Symptomless	Unknown	New Zealand & USA	NC037058
		<i>Grapevine virus J</i>	GVJ	<i>Vitis vinifera</i>	Symptomless	Unknown	Turkmenistan	NC040564
		<i>Heracleum latent virus</i>	HLV	<i>Apiaceae</i> spp.	Symptomless	Mechanically- & aphid transmitted	Europe	NC039087
		<i>Mint virus 2</i>	MV2	<i>Mentha x gracilis</i>	Symptomless	Aphid transmitted	USA	NC043088
		<i>Agave tequilana leaf virus^c</i>	ATLV	<i>Agave tequilana</i>	Unknown	Unknown	Mexico	NC034833
		<i>Arracacha latent virus V^c</i>	ALVV	<i>Arracacia xanthorrhiza</i>	Unknown	Unknown	Peru	KY451036
		<i>Blueberry green mosaic-associated virus^c</i>	BGMaV	<i>Vaccinium</i> spp.	Green Mosaic	Unknown	USA	MK460433
		<i>Grapevine virus K^c</i>	GVK	<i>Vitis vinifera</i>	Unknown	Unknown	Croatia & Italy	NC035202

Grapevine virus L ^c	GVL	<i>Vitis</i>	Symptomless	Unknown	China, Croatia, New Zealand & USA	MH248020	
Grapevine virus M ^c	GVM	<i>vinifera</i>	Unknown	Unknown	USA	MK492703	
<i>Wamavirus</i>		<i>Vitis</i> spp.					
Unknown	<i>Watermelon virus</i> A ^b	WVA	<i>Citrullus lanatus</i>	Crinkling/Mosaic	Unknown	China	NC034377

^aproposed genus.

^btype species.

^ctentative species.

^dit has been proposed to remove Fig latent virus 1 from the genus *Trichovirus* to the newly proposed genus *Fivivirus*.

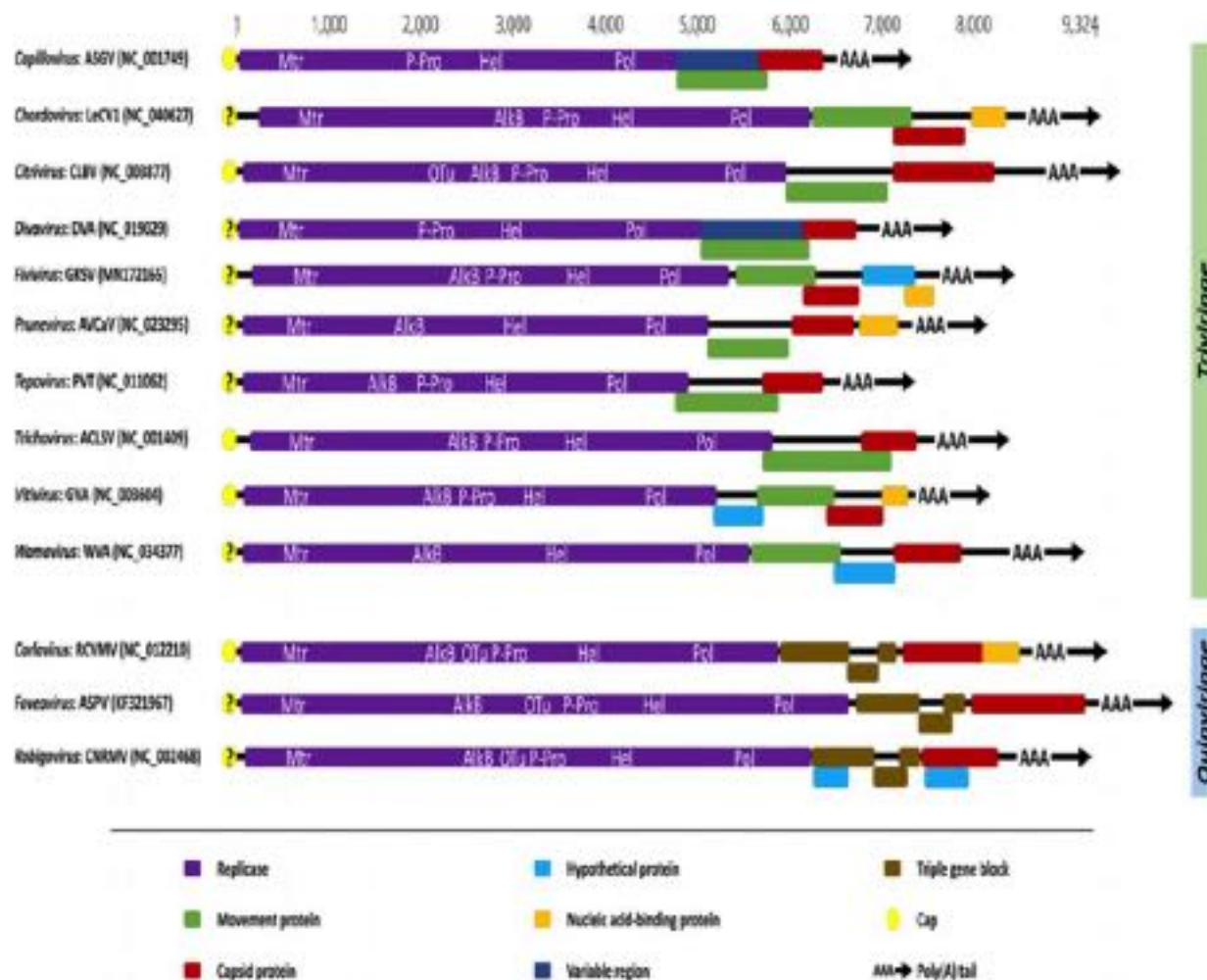


Fig. 1 Graphical representation of the genomic organization of selective members of the family *Betaflexiviridae* showing the relative positions of the ORFs and their gene products. The motifs are Mtr: methyltransferase; OTu-Pro: ovarian tumor-like protease; AlkB: alpha-ketoglutarate-dependent dioxygenase; P-Pro: papain-like protease; Hel: helicase; Pol: RNA-dependent RNA polymerase; ?: cap structure has not been demonstrated yet. *Trivirinae*: ACLSV: Apple chlorotic leaf spot virus, ASGV: Apple stem grooving virus, AVCaV: Apricot vein clearing associated virus, CLBV: Citrus leaf blotch virus, DVA: Diuris virus A, GKSU: Grapevine Kizil Sapak virus, GVA: Grapevine virus A, LeCV1: Lettuce chordovirus 1, PVT: Potato virus T, WVA: Watermelon virus A. *Quinvirinae*: ASPV: Apple stem pitting virus, CNRMV: Cherry necrotic rusty mottle virus and RCMV: Red clover vein mosaic virus.

(TGB) and CP. The RP of triviruses belongs to the “alphavirus-like” supergroup of RNA viruses. The RP proteins are most closely related to those of other virus families in the order *Tymovirales* whereas the MPs of the “30K” superfamily are more closely related to those in the family *Virgaviridae*. For taxonomic assignment, host range, serological specificity, nt (less than 72%) or aa (less than 80%) identity of CP or RP as well as vector specificity have been used to distinguish species.

Virion Structure

The virions of triviruses are non-enveloped flexuous filaments with helical symmetry (with a pitch of approximately 3.3–3.8 nm). Depending on the negative contrast material used for electron microscopy, criss-cross or rope-like features can be visible. The virions are composed of a ssRNA molecule (representing approximately 5% of the virion weight) and a polypeptide (ranging in size from 18 to approximately 41 kDa). The length of the virions is genus-dependent and can vary from 640 to 1500 nm (Table 1) with a diameter ranging from 10 to 15 nm. Fig. 2 shows the virions of Potato virus T (PVT), a tepovirus. To maintain the integrity of the particles, divalent cations are required.

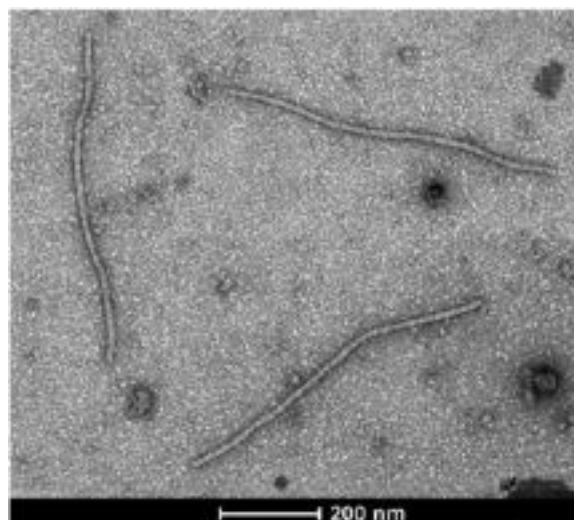


Fig. 2 Electron micrograph of Potato virus T (PVT; *Tepovirus*) particles. The micrograph shows non-enveloped flexuous filaments virions with helical symmetry. Courtesy: Katja R. Richert-Pöggeler, Julius Kühn Institute, Germany.

Genomic Organization and Properties of Encoded Proteins

The genomes of triviruses are encoded by a linear positive-sense ssRNA with two to five open reading frames (ORF) (**Fig. 1**). The 5' terminus has a 7-methyl-guanylate cap (m^7G) but this has not been confirmed for all genera. The 3' terminus is polyadenylated (poly(A)). The lengths range from approximately 6.45 to 8.75 kb excluding the poly(A) tail.

Genomes of triviruses have three conical genes encoding the RP, MP, and CP. ORF1 encodes the RP with a viral methyl-transferase (Mtr) motif at the N-terminal region, and viral helicase 1 (Hel) and RNA-dependent RNA polymerases (Pol) core motifs near the C-terminal region (**Fig. 1**). Additional motifs were reported in most members i.e., alpha-ketoglutarate-dependent dioxygenase (AlkB) and papain-like protease (*p*-Pro) motifs between the Mtr and Hel motifs. The ovarian tumor-like protease motif (OTu-Pro) could only be found on ORF1 of citriviruses (**Fig. 1**). A second ORF encodes a cell-to-cell MP of the "30K" superfamily (~ 31 to ~ 50 kDa). The third conserved ORF of triviruses encodes a CP of approximately 21–40.5 kDa.

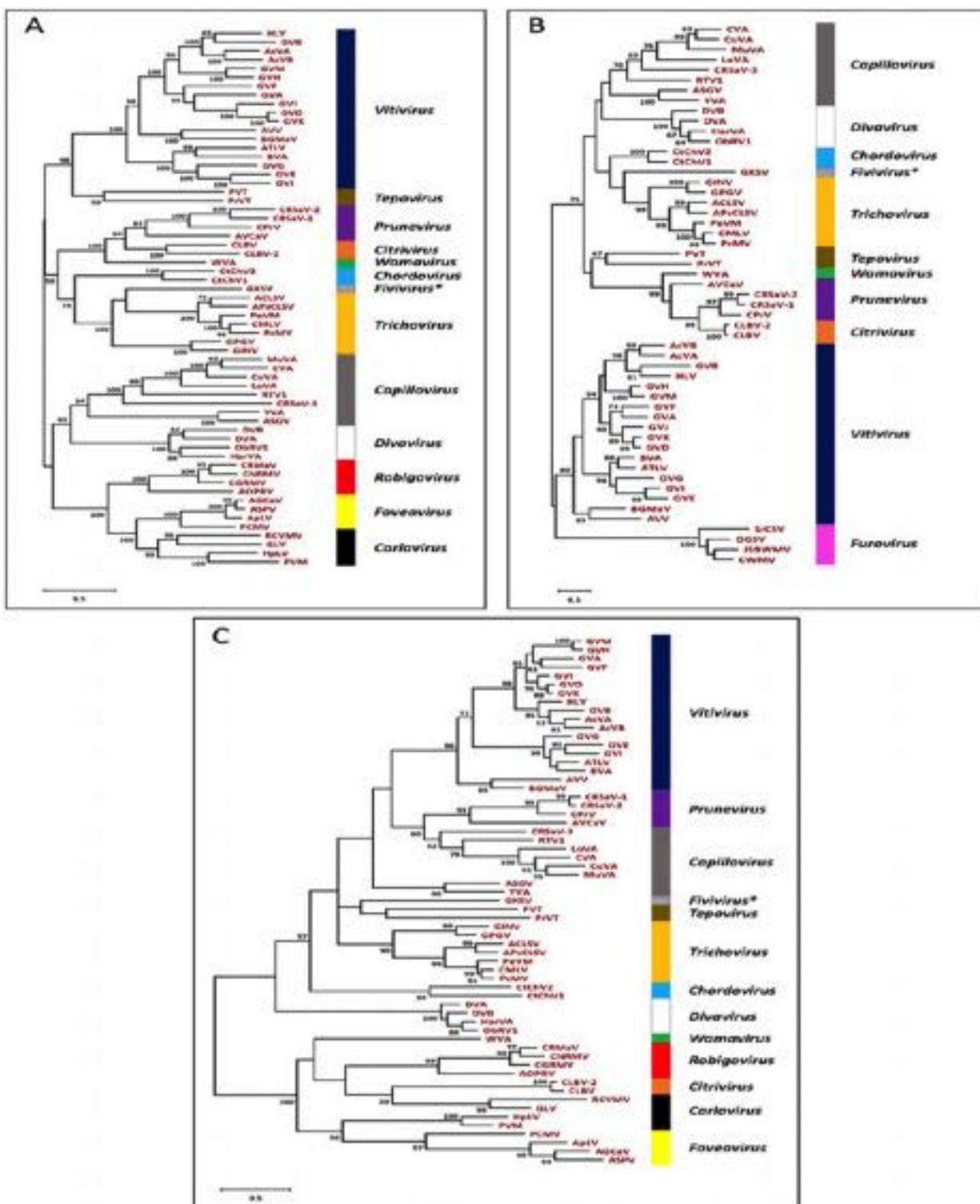
The genomic organization of capilloviruses and divaviruses differs from the other genera in the subfamily *Trivirinae*. They have only two overlapping ORFs (**Fig. 1**). Their ORF1 encodes a polyprotein containing the RP and the CP. A "variable region (V-region)" can be found between the Pol motif (within the RP) and the CP where the aa sequence shows high variability among different virus isolates (**Fig. 1**). A second ORF, encoding the MP, is nested within the ORF1. An exception of capilloviruses is the recently identified Loquat virus A (LoVA). It possesses a genome composed of three ORFs, where the CP coding ORF3 is separated from the RP coding ORF1.

Chordoviruses, pruneviruses, vitiviruses, and some trichoviruses have, in addition to the RP, MP, and CP, a 3'-terminal ORF encoding a putative nucleic acid binding protein (NABP) with a zinc binding finger motif (approximately 10.2–17.1 kDa) (**Fig. 1**). Citriviruses, tepoviruses, and some trichoviruses have only the three conical genes. In vitiviruses, there is an additional ORF encoding a hypothetical protein (~ 17.2 to ~ 24.8 kDa) between the RP and the MP (**Fig. 1**). In wamaviruses, another ORF is located between the MP and the CP encoding a hypothetical protein of approximately 25 kDa.

The genome organization of the two putative members of the proposed genus *Fivivirus* FLV1 and GKSV are different. Four predicted, slightly overlapping ORFs were identified in the FLV1 genome, the putative RP, MP (43 kDa), CP (46 kDa), and NABP (12 kDa). GKSV has five predicted ORFs where the RP (196.7 kDa) is followed by the MP (31.7 kDa) which is slightly overlapped by the CP (22 kDa) and a hypothetical protein (21.4 kDa) is overlapped with a predicted NABP which overlap (11.7 kDa).

Phylogenetic Relationships of the Different Triviruses

Phylogenetic analysis based on RPs alignment of betaflexivirids shows that the triviruses of each genus are grouped together, mostly within well-supported branches (**Fig. 3(A)**). A slightly different picture emerges when using aa alignments of the MP: The MP of the citriviruses i.e., Citrus leaf blotch virus (CLBV) and Citrus leaf blotch virus 2 (CLBV2) form a group with pruneviruses. As triviruses have a "30K"-type MP while quinviruses have the TGB, quinviruses are not included in the tree. The closely related *Furovirus* genus (family: *Virgaviridae*) was used instead for the Maximum-likelihood (ML) tree (**Fig. 3(B)**). **Fig. 3(C)** shows an ML tree representing the phylogenetic relationship between the CP aa sequence of selective members of the subfamily *Trivirinae* and *Quinvirinae*. Interestingly, citriviruses and the wamavirus WVA are grouped with quinviruses (**Fig. 3(C)**).



Cellular Localization of Trivirus Particles and Viral Replication

Trivirus replication and assembly occur in the cytoplasm. Virions can be found in the cytoplasm of infected mesophyll, parenchyma, and phloem cells of leaves and roots. The cellular localization however is virus- and host-dependent. For example, virions of Grapevine virus A (GVA), Grapevine virus B (GVB) and Heracleum latent virus (HLV) (vitiviruses), and Apple chlorotic leaf spot virus (ACLSV; a trichovirus) are strictly phloem-limited, but in herbaceous hosts, they can be found in the parenchyma. Moreover, virions of ACLSV could be located in the nucleus. Trivirus particles accumulate in discrete bundles or paracrystalline aggregates in the cytoplasm where multiplication occurs. Trivirus-infected cells were found to be damaged to a varying extent. Some triviruses were found to cause tonoplast vesicular evagination where fine fibrillar material (suggested to be the replicative forms of the viral RNA) can be observed. Other cellular damages include enlarged starch grains, swollen chloroplasts and damaged mitochondria. Moreover, a study showed that ASGV infection of apple trees extensively alters the gene expression patterns even with no disease symptoms observed.

The expression of trivirus proteins relies on production of sub-genomic RNA (sgRNA) as suggested by the analyses of dsRNA patterns from infected plants. As an example, five dsRNA species were detected in plant tissues infected with apple stem grooving virus (ASGV; a capillovirus) i.e., 6.5, 5.5, 4.5, 2.0, and 1.0 kbp. The 6.5 kbp species represents the dsRNA of the full genome of ASGV. The 5.5 and 4.5 species are the 5' co-terminal with ASGV genomic RNA while the 2.0 and 1.0 kbp are suggested to be dsRNA of sgRNA of the MP and CP, respectively. Six dsRNAs species with sizes of about 7.5, 6.4, 5.4, 2.2, 1.1, and 1.0 kbp were detected in ACLSV-infected plant tissues. The largest dsRNA species here represent the full genome of the virus, the 6.4 and 5.4 dsRNA species are the 5' co-terminal with ACLSV genomic RNA. The dsRNA 2.2 and 1.1 kbp species are 3' co-terminal with ACLSV genomic RNA which could be representing the sgRNA of the MP and the CP, respectively. The 1.0 kbp species is internal to the genome.

It is suggested that the MP of triviruses acts as a suppressor of plant RNA silencing pathways. For example, ACLSV-MP suppresses the systemic movement of the silencing signal whereas CLB-V-MP suppresses the intracellular silencing without affecting the systemic spread of the signals.

Host Range and Distribution of Triviruses

Triviruses infect predominantly grapevine, citrus fruit and fruit trees/shrubs and therefore they can be found worldwide. The host range of triviruses appears to be quite narrow under natural conditions, infecting only one or few related plants species. However, under experimental conditions it is possible to transmit many triviruses mechanically to herbaceous indicator plants such as *Nicotiana* or *Chenopodium* spp. (see Table 1 for details on natural hosts). In recent years, many new triviruses have been described that infect herbaceous plants. For a long time, the tepovirus PVT (identified from potato crops in South America and classified in many countries as quarantine pest) and Heracleum latent virus (HLV) that was described from hogweed (*Heracleum sphondylium*)

Fig. 3 Maximum-likelihood (ML) trees of members of the subfamily *Trivirinae* (*Betaflexiviridae*) based on alignments of the amino acid sequences of (A) the replicase, (B) the movement, (C) the capsid proteins, and their homologs from representative members of subfamily *Quinvirinae* (*Betaflexiviridae*) and the genus *Furovirus* (*Virgaviridae*). MEGAX was used to perform MUSCLE alignments and for producing the ML trees (based on Jones-Taylor-Thornton (JTT) matrix-based model). The numbers on branches indicating the percentage of bootstraps (1000 bootstrap replications; only >50% are shown). The viruses and their accession numbers are: *Betaflexiviridae*: *Trivirinae*: *Capillovirus*: Apple stem grooving virus (ASGV; NC001749), Camellia ringspot associated virus 3 (CRSaV-3; MK050795), Cherry virus A (CVA; NC003689), Currant virus A (CuVA; NC029301), Loquat virus A (LoVA; MK936045), Mume virus A (MuVA; NC040568), Rubber tree virus 1 (RTV1; MN047299), Yacon virus A (YVA; NC030657), *Chordovirus*: Carrot Ch virus 1 (CtChV1; NC025469), Carrot Ch virus 2 (CtChV2; NC025468), *Citivirus*: Citrus leaf blotch virus (CLBV; NC003877), Citrus leaf blotch virus 2 (CLBV2; MH144344), *Divavirus*: Diuris virus A (DVA; NC019029), Diuris virus B (DVB; NC019030), Hardenbergia virus A (HarVA; NC015395), Ocimum basilicum RNA virus 1 (ObRV1; NC035462), *Fivirus* (*proposed genus): Grapevine Kizil Sapak virus (GKSV; MN172165), *Prunavirus*: Apricot vein clearing associated virus (AVCaV; NC023295), Camellia ringspot associated virus 1 (CRSaV-1; MK050792), Camellia ringspot associated virus 2 (CRSaV2; MK050793), Caucasus prunus virus (CPrV; NC038325), *Tepovirus*: Potato virus T (PVT; NC011062), Prunus virus T (PrVT; NC024686), *Trichovirus*: Apple chlorotic leaf spot virus (ACLSV; NC001409), Apricot pseudo-chlorotic leaf spot virus (APsCLSV; NC006946), Cherry mottle leaf virus (CMLV; NC002500), Grapevine berry inner necrosis virus (GINV; NC015220), Grapevine Pinot gris virus (GPGV; NC015782), Peach mosaic virus (PcMV; NC011552), Peach virus M (PeVM; MK012336), *Vitivirus*: Actinidia virus A (AcVA; LC491285), Actinidia virus B (AcVB; NC016404), Agave tequilana leaf virus (ATLV; NC034833), Arracacha virus V (AVV; NC034264), Blackberry virus A (BVA; NC040630), Blueberry green mosaic associated virus (BGMAV; MK460433), Grapevine virus A (GVA; NC003604), Grapevine virus B (GVB; NC003602), Grapevine virus D (GVD; MF774336), Grapevine virus E (GVE; NC011106), Grapevine virus F (GVF; NC018458), Grapevine virus G (GVG; NC040554), Grapevine virus H (GVH; NC040545), Grapevine virus I (GVI; NC037058), Grapevine virus J (GVJ; NC040564), Grapevine virus K (GVK; NC035202), Grapevine virus M (GVM; MK492703), Heracleum latent virus (HLV; MN314973), *Wamavirus*: Watermelon virus A (WVA; NC034377). *Betaflexiviridae*: *Quinvirinae*: *Carlavirus*: Potato virus M (PVM; NC001361), Hop latent virus (HpLV; NC002552), Garlic latent virus (GLV; NC003557), Red clover vein mosaic virus (RCVMV; NC012210), *Foveavirus*: Apple stem pitting virus (ASPV; NC003462), Peach chlorotic mottle virus (PCMV; NC009892), Apricot latent virus (ApLV; NC014821), Apple green crinkle associated virus (AGCaV; NC018714), *Robigovirus*: Cherry green ring mottle virus (CGRMV; NC001946), Cherry necrotic rusty mottle virus (CNRMV; NC002468), African oil palm ringspot virus (AOPRV; NC012519), Cherry rusty mottle associated virus (CRMaV; NC020996). *Virgaviridae*: *Furovirus*: Oat golden stripe virus (OGSV; NC002358), Chinese wheat mosaic virus (CWMV; NC002359), Sorghum chlorotic spot virus (SrCSV; NC004014), Japanese soil-borne wheat mosaic virus (JSBWMV; NC038850).

in Europe were the only triviruses found naturally on herbaceous plants. New triviruses found on non-woody plants include Yacon virus A (YVA; a capillovirus) that infects tuberous plants (*Smallanthus sonchifolius*); Carrot chordovirus 1 and 2 (CtChV1 and CtChV2) as well as Lettuce chordovirus 1 (LeCV1; *Chordovirus*) that were identified from carrots (in the UK) and lettuce (from France), respectively; Diuris virus A and B (DVA and DVB) from donkey orchids (*Diuris penduculata*) and Hardenbergia virus A (HVA; *Divavirus*) from the legume *Hardenbergia comptoniana*, from Australia; Zosteria virus T (ZVT; *Tepovirus*) that infects *Zostera muelleri*, a seagrass species that can be found in regions of Australia, New Zealand, and Papua New Guinea; Phlomis mottle virus (PhMV; *Trichovirus*) infecting *Phlomis* spp. in the Mediterranean region; Agave tequilana leaf virus (ATLV; *Vitivirus*) infecting *Agave tequilana* in Mexico, Arracacha virus V (AVV) and Arracacha latent virus (ALVV; *Vitivirus*) infecting the South American root crop *Arracacia xanthorrhiza*; Mint virus 2 (MV2; *Vitivirus*) from mint in USA and Watermelon virus A (WVA) infecting watermelon (*Citrullus lanatus*) in China. Recently discovered triviruses include also LoVA from loquat (*Eriobotrya japonica*) in China, Rubber tree virus 1 (RTV1; *Capillovirus*) infecting rubber trees (*Hevea brasiliensis*) in China, Camellia ringspot associated virus 1, 2 and 3 (CRSaV1, CRSaV2; *Prunivirus*, and CRSaV3; *Capillovirus*) from *Camellia* spp. in the USA, and Blueberry green mosaic-associated virus" (BGMaV; *Vitivirus*) from blueberries in the USA. Only limited information is available on alternative hosts or distribution of these viruses in other regions.

Symptomology, Impact and Vectors

Common symptoms in fruit trees include graft incompatibilities, necrosis and pitting of the graft union and stem grooving of wood; leaf symptoms may include chlorotic spots, line patterns or ring spots, mottling, mosaic or necrotic spots (Fig. 4). For example, ACLSV causes symptoms like bark split and dark green sunken mottle (Fig. 4(A) and (B)), AVCaV causes symptoms of

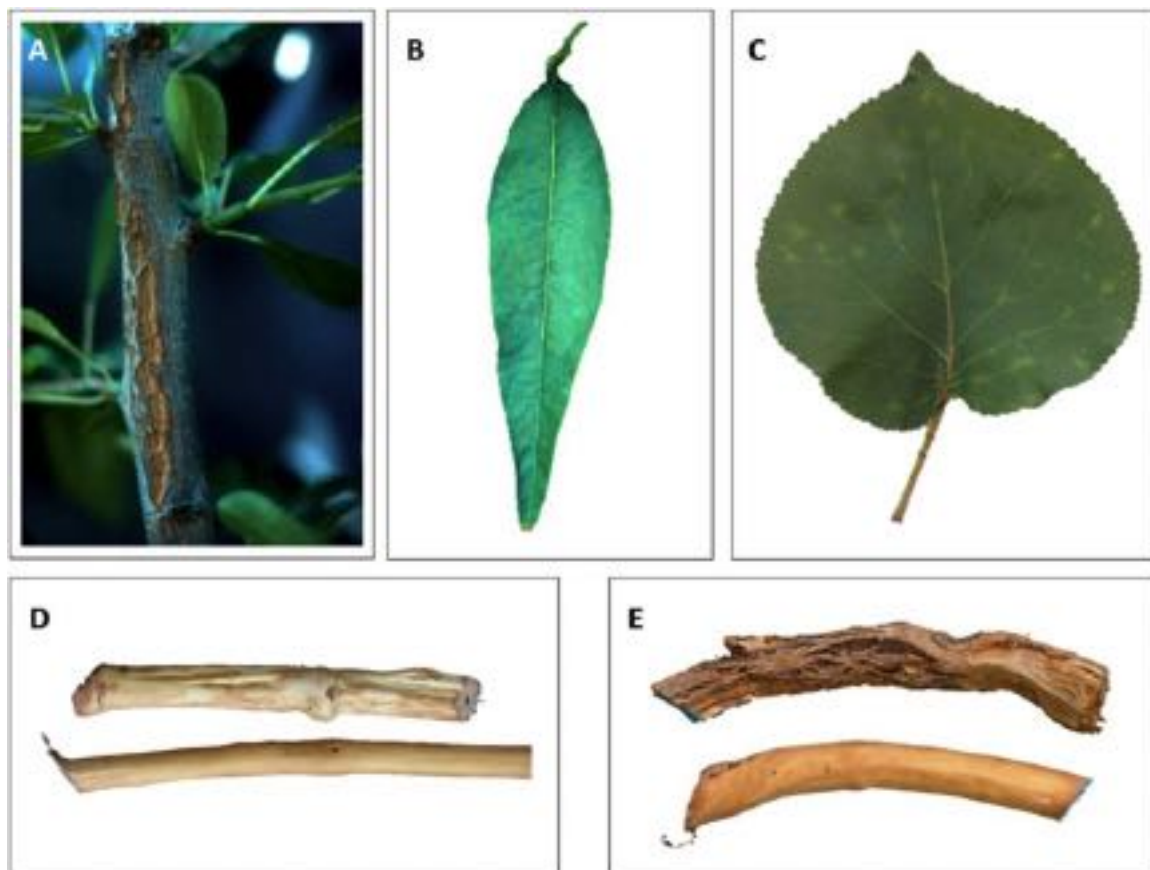


Fig. 4 Symptoms induced by triviruses. (A) Symptoms of bark split caused by Apple chlorotic leaf spot virus (ACLSV) in plum (cv. Prune d'Ente), (B) Typical dark green sunken mottle caused by an ACLSV isolates in the GF305 peach seedling indicator, (C) Symptoms of vein clearing and blotching observed on apricot leaf infected with Apricot vein clearing-associated virus (AVCaV), (D) *Vitis berlandieri* × *V. riparia* (indicator plants) expressed stem grooving disease associated with infection by Grapevine virus A (GVA); Bottom is healthy control, (E) Couderc 1613 × *V. berlandieri* (indicator plants), developed corky-bark symptoms associated with infection by Grapevine virus B (GVB); Bottom is healthy control. Courtesy: (A) Jean Dunez, INRAE, France, (B) Thierry Candresse, INRAE, France, (C) Toufic Elbeaino, Istituto Agronomico Mediterraneo di Bari, Italy, (D) and (E) Maher Al Rwahnih, Foundation plant Services – University of California – Davis, USA.

vein clearing and blotching on apricot leaves (Fig. 4(C)), and color breaking of petals was observed for PcMV infection. Fruits may be smaller, ripen late or may have wart-like outgrowths or injuries. The majority of grapevine-infecting triviruses does not cause visible symptoms on the plants; some infections may lead to necrotic spots, mosaic, mottle or ring spots on leaf but more dramatic is the effect on yield quantity and quality e.g., of Grapevine Pinot gris virus (GPGV) on berry quantity and quality that may be both reduced in these economically important crops. As another example, GVA is associated with pitting and grooving of wood but occurs very often in mixed infections with other viruses (Fig. 4(D)); corky bark disorder was described for GVB infection (Fig. 4(E)) whereas Grapevine virus D (GVD) is associated with corky rugose wood disease. Both GVA and GVB are associated with the Kober stem grooving syndrome.

However, for many triviruses, no specific symptoms were observed, e.g., HLV may cause no symptoms at all. Many of the recently added triviruses were identified during virus ecology studies using high-throughput sequencing technologies (HTS) of pooled samples and therefore symptoms cannot be attributed to the identified viruses. Mixed infections of triviruses and in particular with closteroviruses (family: *Closteroviridae*) are quite common in nature; multiple infections of one plant with several triviruses have also been reported. An example is the occurrence of Grapevine leafroll-associated viruses (GLRaVs; closteroviruses) and the vitiviruses GVA, GVB, GVD, GVE, and GVF in mixed infections of vineyards. As another example, the Plum bark necrosis stem pitting-associated virus (PBNPaV, *Closterovirus*) and Apricot vein clearing associated virus (AVCaV, *Prunivirus*) can be found in dual/mixed infections. Therefore, it is difficult to assign a certain symptomology to infection with a specific virus.

Most, if not all, fruit tree- and grapevine-infecting triviruses are graft-transmissible; many triviruses could be transferred mechanically under experimental conditions to herbaceous indicator plants. Transmission via seeds was reported in few studies e.g., PVT, CRSaV1, CRSaV2, CRSaV3, and ASGV (for ASGV, seed-transmission was shown only for indicator plants e.g., lily and *Malus platycarpa*, but it is unclear if this dissemination route is of major importance under natural conditions). Insect vectors appear to have a more prominent role in transmission of triviruses; HLV and Mint virus 2 (MV2) can be transmitted by aphids; fruit tree and grapevine-infecting triviruses can be transmitted by eriophyid mites (*Eriophyes inaequalis*: transmitting CMLV; *Eriophyes insidiosus*: transmitting PcMV), erineum mites (*Colomerus vitis*: transmitting GINV and GPGV) and various mealybug species (*Pseudococcus longispinus*, *Ps. affinis*, *Planococcus ficus*, *Pl. citri*, *Phenacoccus aceris*) and scale insects (*Neopulvinaria innumerabilis*) that are involved in the transmission of GVA, predominantly. It appears that L1 stages of *Ph. aceris* are more efficient vectors than L2, L3 or adult females although the reasons are unknown. They are transmitted in semi-persistent manner. Moreover, HLV was found to require Heracleum virus 6 (HV6; synonym of Carrot yellow leaf virus [CYLV], a closterovirus) as a helper for vector transmission. For many of the novel triviruses, information on vectors or alternative routes of transmission is missing (Table 1).

Diagnosis

Many of the early investigations on triviruses used test plants as bioassays for induction of symptoms and propagation of viruses. However, these bioassays cannot be used for identification and characterization of triviruses and are laborious. It is possible to identify virions by various electron microscopy techniques, i.e., transmission electron microscopy, immunogold labeling, or immunosorbent electron microscopy. Depending on the genus, trivirus virions are poor to moderate antigens. Several polyclonal antisera and monoclonal antibodies have been developed for the detection and identification of several triviruses. Cross-reaction of antibodies can occur; e.g., antibodies raised against CMLV cross-react with PcMV. Other monoclonal antibodies have been produced for the specific identification of GVD; those antibodies did not cross-react with other vitiviruses.

Various PCR-based assays have been developed for the sensitive and reliable detection of triviruses. Depending on primer design, these assays were successfully implemented to detect specifically individual viral species but the use of degenerate primers such as dPR1 and dPR2 allowed the detection of several vitiviruses (GVA, GVB, GVD, HLV), a tepovirus (PVT) and a divergent ACLSV isolate (*Trichovirus*).

Improved sensitivity compared to a conventional RT-PCR assay was achieved using an immunocapture RT-PCR (IC-RT-PCR) assay for the detection of ASGV. In this assay, polyclonal antibodies were used to trap the virus particles followed by RT-PCR using ASGV specific primers. However, these assays have mainly been replaced with multiplex assays targeting multiple viruses and/or quantitative RT-PCR (qRT-PCR) protocols.

As many triviruses are present in mixed infections, several multiplex detection assays have been developed, in particular for grapevine and fruit trees. For example, multiplex RT-PCR and qRT-PCR assays were designed for the simultaneous detection of several grapevine viruses including GVA and GVB. For a Canadian survey of grapevines, a multiplex RT-PCR assay targeting seventeen grapevine viruses (including GVA, GVB, GPGV) was generated and successfully applied in the detection of the triviruses in the Ontario region. For fruit trees, RT-PCR protocols for the simultaneous detection and identification of eight stone fruit viruses (including ACLSV), four cherry viruses (including CVA), and pear viruses (including ASGV or ACLSV) have been reported.

A single qRT-PCR assay for simultaneous detection of GVA, GVB, and GVD was more sensitive and faster than a conventional RT-PCR assay that used degenerate primers to detect these viruses. Similar qRT-PCR protocols were for example developed for the detection of citrus viruses such as Citrus tristeza virus (*Closteroviridae*), Citrus psorosis virus (genus *Ophiovirus*, family *Aspiviridae*) and CLBV or the detection of ASGV, Apple stem pitting virus (*Foveavirus*) and Apple mosaic virus (*Ilarvirus*).

Recently, loop-mediated isothermal amplification (LAMP) assays have also been established for fast and sensitive detection of triviruses and have been applied for the detection of ACLSV.

The biggest challenge, however, for the development of specific and sensitive detection assays lies in the development of HTS technologies and the subsequent description of novel virus species and divergent isolates of known viruses. Many new triviruses have been detected recently using HTS (e.g., Grapevine virus M (GVM), WVA, AVCaV, etc.); diagnostic assays need to either be adapted to reflect the detection of these new species or have to be developed from scratch.

Prevention and Treatment

Grapevine and fruit trees, the major host groups of viruses belonging to the *Trivirinae*, are crops of high economic value. They are also perennial plants, and often the highest yield or quality of fruits is only achieved after a few years of planting. The effect of infection with triviruses on these crops may not only be seen in quantitative yield reduction but also in qualitative damage to fruits or secondary metabolites influencing, e.g., the fruit aroma. Most of the preventative and curative measurements apply generally to viruses and viroids. Foremost, preventative measures such as quarantine programs and certification schemes have been adapted to provide healthy root stocks and bud wood as well as other planting material. Rules and recommendations for the provision of “virus-free” material are set by numerous organizations and networks world-wide for grapevine, stone and citrus fruits (e.g., International Council for the Study of Virus and Virus-like Disease of the Grapevine (ICVG), National Clean Plant Network (NCPN), Canadian Plant Protection Export Certification Program (PPECP), Commission of the European Union, North American Plant Protection Organization (NAPPO), European and Mediterranean Plant Protection Organization (EPPO), etc.). However, correct identification of (latent) infection is often hampered by the lack of sensitive detection methods or the presence of unknown viruses or divergent isolates for which a specific test has not been developed yet. Nevertheless, the prophylactic measures can provide healthy propagation material including seeds from which new orchards/vineyards can be planted.

Eradication of infected plants might be successful if detected early and limited to a small geographic region. Secondary spread of the disease may be slowed down but in practice, it has been difficult to eradicate the pathogens completely, in particular if vectored by (fast moving) insects. Vector control using insecticides has become increasingly difficult to apply due to the detrimental effects on non-target organisms and the environment, and the development of insecticide-resistant vectors themselves. Additionally, in certain areas of the world there is a political desire to reduce the use of synthetic pesticides. Biological antagonists (such as predatory insects) of trivirus vectors (mites, mealy bugs and scale insects) are poorly studied, especially under field conditions. This knowledge gap whether antagonists can reduce vector populations thus reducing virus transmission needs to be closed in the future.

Only few curative measurements for virus elimination of plant material have been described. Some of these have been successfully applied to eradicate trivirus infection; in particular, elimination of GVA seems to be quite successful. These methods include thermotherapy, meristem tip/shoot tip cultures, somatic embryogenesis, chemotherapy (using ombuin, glycyrrhizin/quercetin, ribavirin, and quercetin/ribavirin), and electrotherapy, applied either alone or in combination. However, these methods may induce undesired by-effects such as phytotoxicity, putative mutagenic effects and the induction of somaclonal variation.

To date, no natural resistance against triviruses has been described. Transgenic approaches showed some limited effects in the control of triviruses: In model plants (such as *N. benthamiana*) reduction of virus titer and absence of virus symptoms was observed in plants expressing GVA-CP or GVA/GVB-MPs. *N. benthamiana* plants expressing a GVA mini-replicon were extremely resistant to GVA infection but susceptible to GVB. Interestingly, *N. occidentalis* plants that were expressing ACLSV-CP were resistant to GINV infection but not to ASGV or ASPV infection. Furthermore, partially functional deletion mutants of ACLSV also conferred resistance to GINV, probably due to interference with viral long-distance movement. An artificial microRNA derived from grapevine was engineered to target GVA genomic sequences in *N. benthamiana* thus inducing various degrees of resistance in that host.

Although it was possible to transform grapevine embryonic tissue with MP genes from GVA or GVB, the resulting plants showed phenotypic abnormalities. Attempts to use the CP genes from GVA or GVB for transformation grapevine seems to be not successful. It therefore seems that due to practical and political reasons, virus-resistance introduced by transgenic approaches is currently not a short-term option.

Further Reading

- Al Rwahnih, M., Alabi, O.J., Hwang, M.S., Stevens, K., Golino, D., 2019. Identification and genomic characterization of grapevine Kizil Sapak virus, a novel grapevine-infecting member of the family *Betaflexiviridae*. *Archives of Virology* 164 (12), 3145–3149. doi:10.1007/s00705-019-04434-3.
- Barba, M., Ilardi, V., Paquini, G., 2015. Control of pome and stone fruit virus diseases. *Advances in Virus Research* 91, 47–83. doi:10.1016/bs.aivir.2014.11.001.
- Lee, R.F., 2015. Control of virus diseases of citrus. *Advances in Virus Research* 91, 143–173. doi:10.1016/bs.aivir.2014.10.002.
- Maliogkka, V.I., Martelli, G.P., Fuchs, M., Katis, N.I., 2015. Control of viruses infecting grapevine. *Advances in Virus Research* 91, 175–227. doi:10.1016/bs.aivir.2014.11.002.
- Marais, A., Faure, C., Mustafayev, E., Candresse, T., 2015. Characterization of new isolates of *Apricot vein clearing-associated virus* and of a new *Prunus*-infecting virus: Evidence for recombination as a driving force in *Betaflexiviridae* evolution. *PLOS One* 10 (6), e0129469. doi:10.1371/journal.pone.0129469.
- Maree, H.J., Blouin, A.G., Diaz-Lara, A., *et al.*, 2020. Status of the current vitivirus taxonomy. *Archives of Virology* 165, 451–458. doi:10.1007/s00705-019-04500-w.
- Zherdev, A., Vinogradova, S., Byzova, N., *et al.*, 2018. Methods for the diagnosis of grapevine viral infections: A review. *Agriculture* 8 (12), 195. doi:10.3390/agriculture8120195.

Relevant Websites

<https://inspection.gc.ca/plant-health/plant-pests-invasive-species/directives/date/d-97-06/eng/1312330811581/1312331075782>

Directive D-97-06: Plant Protection Export Certification Program for Grapevine Nursery Stock, *Vitis* spp.

https://ec.europa.eu/info/index_en

European Commission, official website.

<https://www.eppo.int/>

European and Mediterranean Plant Protection Organization (EPPO).

<http://icvg.org/>

International Council for the Study of Viruses and Virus-like Diseases of the Grapevine (ICVG).

<http://nationalcleanplantnetwork.org/>

National Clean Plant Network.

<https://www.nappo.org/>

North American Plant Protection Organization (NAPPO).

Tymoviruses (*Tymoviridae*)

Rosemarie W Hammond and Peter Abrahamian, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A.-L. Haenni, T.W. Dreher, Tymoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00523-9](https://doi.org/10.1016/B978-012374410-4.00523-9).

Nomenclature

aa Amino acid(s)

ATC Artificial top component

CP Coat protein or capsid protein

EM Electron microscopy

kb Kilobase

kDa Kilo dalton

nt Nucleotide(s)

ORF Open reading frame

RdRp RNA-dependent RNA polymerase

RP Replication protein

sgRNA Sub-genomic RNA

ssRNA Single stranded RNA

TLS tRNA-like structure

UTR Un-translated region

Glossary

Icosahedron A solid having 20 faces and 12 vertices.

Phylogeny The complete evolutionary history of group of animals, plants, bacteria, viruses, etc.

Pseudoknot RNA structure formed by base-pairing between nucleotides within the loop subtending a stem and nucleotides outside of this loop.

Sub-genomic RNA Derived from a viral RNA genome during replication, serves as mRNA for the expression of genes that are translationally silent in the genomic RNA.

The Family and Its Distinguishing Features

The members of the family *Tymoviridae*, in the order *Tymovirales*, are presented in [Table 1](#). Like the family itself, the genus *Tymovirus* derives its name from the type species, *Turnip yellow mosaic virus*. Turnip yellow mosaic virus (TYMV) was first isolated in 1946 and is by far the most intensively studied member of the family. Indeed, TYMV is one of the best-characterized plant viruses. The family was created in recognition of the close relationships between the genera *Tymovirus*, *Marafivirus*, and the founding member of newly created genus *Maculavirus*, Grapevine fleck virus (GFkV). Poinsettia mosaic virus (PnMV) and Bombyx mori latent virus (BmLV) are currently family members unassigned to a genus. Complete genome sequences are available for the type species of each of the three genera of the *Tymoviridae* and for PnMV and BmLV. Newly described tentative tymoviruses having sequence homologies and similar genome expression strategies, but not yet assigned to the family are listed at the end of [Table 1](#).

The members of the family *Tymoviridae* are characterized by their icosahedral, non-enveloped, ~29 nm virions, that can readily be visualized by negative-staining electron microscopy (EM). Infections produce a characteristic mixture of filled, infectious virions, and empty or near-empty capsids. Regular surface features, representing prominent peaks formed by pentamers and hexamers of coat protein (CP) molecules, are evident by EM and by molecular structure and computer rendering ([Fig. 1](#)).

The genomes of members of the *Tymoviridae* are composed of a single positive ssRNA generally 6.0–6.5 kb long, although the genome of GFkV is 7.5 kb long. Sub-genomic RNAs of 1 kb or less are associated with CP expression. All *Tymoviridae* genomes have a distinctive skewed nt composition that is rich in C residues (32%–50%). The unifying characteristic of the genome design of all *Tymoviridae* members is the presence of a long open reading frame (ORF) that covers most of the genome and encodes the replication polyprotein with identifiable domains: methyl-transferase, papain-like proteinase, helicase, and RNA-dependent RNA polymerase (RdRp) in order N- to C-terminal ([Fig. 2](#)). The CP ORF is situated downstream of the polyprotein ORF, to which it is fused in the marafivirus and PnMV genomes. Close familial relationships are easily discerned from alignments of the sequences of the RdRp and CP genes.

Most viruses in the family have narrow host ranges. Many of the tymoviruses have been isolated from non-crop hosts and have thus far not presented major disease threats to crops. Marafiviruses are associated with significant crop losses, perhaps resulting from their more effective transmission by flying insects. Disease symptoms include bright yellow mosaic or mottling (tymoviruses and PnMV), chlorotic stripes, vein clearing, etched lines or dwarfing (marafiviruses), and leaf flecking (maculaviruses).

Table 1 Members of the family Tymoviridae

Species name	Virus acronym	Complete sequence accession no.
<i>Genus Tymovirus</i>		
<i>Anagyris vein yellowing virus</i>	AVYV	AY75180
<i>Andean potato latent virus</i>	APLV	JX508291
<i>Andean potato mild mosaic virus</i>	APMMV	JX508290
<i>Belladonna mottle virus</i>	BeMV	
<i>Cacao yellow mosaic virus</i>	CYMV	
<i>Calopogonium yellow vein virus</i>	CalYVV	
<i>Chayote mosaic virus</i>	ChMV	AF195000
<i>Clitoria yellow vein virus</i>	CYVV	
<i>Desmodium yellow mottle virus</i>	DYMoV	
<i>Dulcamara mottle virus</i>	DuMV	AY789137
<i>Eggplant mosaic virus</i>	EMV	J04374
<i>Erysimum latent virus</i>	ErLV	AF098523
<i>Kennedya yellow mosaic virus</i>	KYMV	D00637
<i>Melon rugose mosaic virus</i>	MRMV	
<i>Nemesia ring necrosis virus</i>	NeRNV	AY751778
<i>Okra mosaic virus</i>	OkMV	EF554577
<i>Ononis yellow mosaic virus</i>	OYMV	J04375
<i>Passion fruit yellow mosaic virus</i>	PFYMV	KY832429
<i>Peanut yellow mosaic virus</i>	PeYMV	
<i>Petunia vein banding virus</i>	PetVBV	
<i>Physalis mottle virus</i>	PhyMV	Y16104
<i>Plantago mottle virus</i>	PIMoV	AY751779
<i>Scrophularia mottle virus</i>	SrMV	AY751777
<i>Tomato blistering mosaic virus</i>	ToBMV	KC840043
<i>Turnip yellow mosaic virus</i>	TYMV	J04373, X16378, X07441
<i>Voandzeia necrotic mosaic virus</i>	VNMV	
<i>Wild cucumber mosaic virus</i>	WCMV	
<i>Genus Marafivirus</i>		
<i>Bermuda grass etched-line virus</i>	BELV	
<i>Blackberry virus S</i>	BVS	FJ915122
<i>Citrus sudden death-associated virus</i>	CSDaV	AY884005, DQ185573
<i>Grapevine Syrah virus 1</i>	GSyV-1	FJ436028
<i>Maize rayado fino virus</i>	MRFV	AF265566
<i>Nectarine virus M</i>	NeVM	KT273411
<i>Oat blue dwarf virus</i>	OBdV	U87832
<i>Olive latent virus 3</i>	OLV-3	FJ444852
<i>Peach virus D</i>	PeVD	KY084481
<i>Tentative species</i>		
<i>Grapevine asteroid mosaic-associated virus</i>	GAMaV	KX354202
<i>Grapevine rupestris vein feathering virus</i>	GRVfV	AY706994
<i>Genus Maculavirus</i>		
<i>Grapevine fleck virus</i>	GfKV	AJ309022
<i>Tentative member</i>		
<i>Grapevine red globe virus</i>	GRGV	KX171166
<i>Unassigned viruses in the family</i>		
<i>Bombyx mori latent virus</i>	BmMLV	AB186123
<i>Poinsettia mosaic virus</i>	PnMV	AJ271595
<i>Tentative Tymoviruses</i>		
<i>Alfalfa virus F</i>	AVF	MG676465
<i>Grapevine virus Q</i>	GVQ	FJ977041
<i>Medicago sativa marafivirus 1</i>	MsMV1	MF443260
<i>Narajilla chlorotic mosaic virus</i>	NarCMV	MG323924

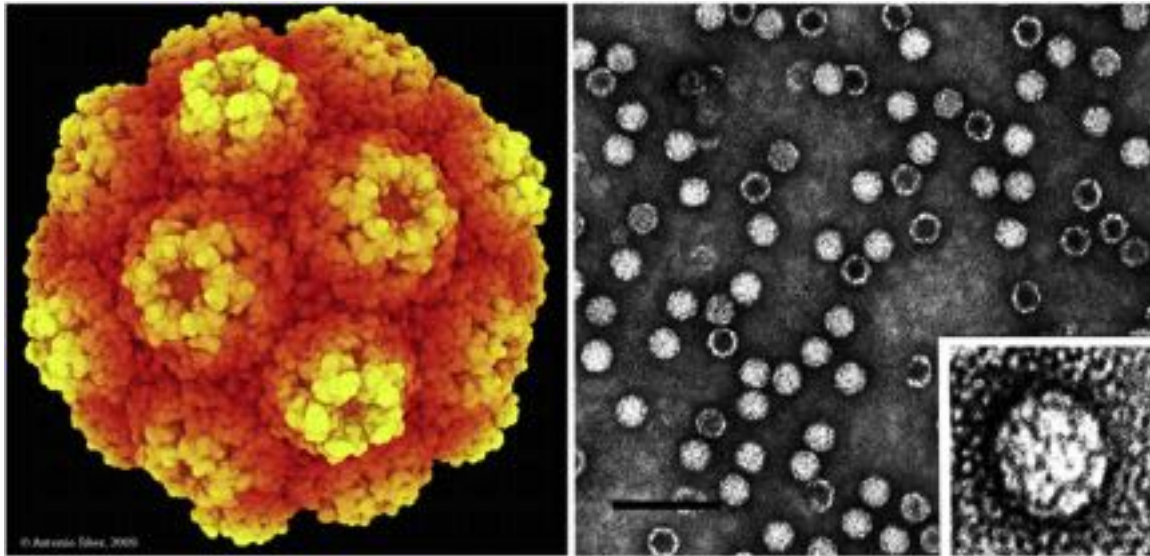


Fig. 1 Tymovirus particle structure. Rendering of a tymovirus particle based on experimental data deposited in the ViperDB database is shown on the left. The different colors denote the distance from the center of the virus, the brighter regions being more distant. Molecular structure courtesy of Antonio Šiber, available at: http://www.antoniosiber.org/publikacije_en.html. A negative-contrast electron micrograph of virions and 'empty' particles of belladonna mosaic tymovirus is shown at right, with the high-magnification inset showing the prominent surface structure. Scale = 100 nm. Image courtesy Dr. D.-E. Lesemann.

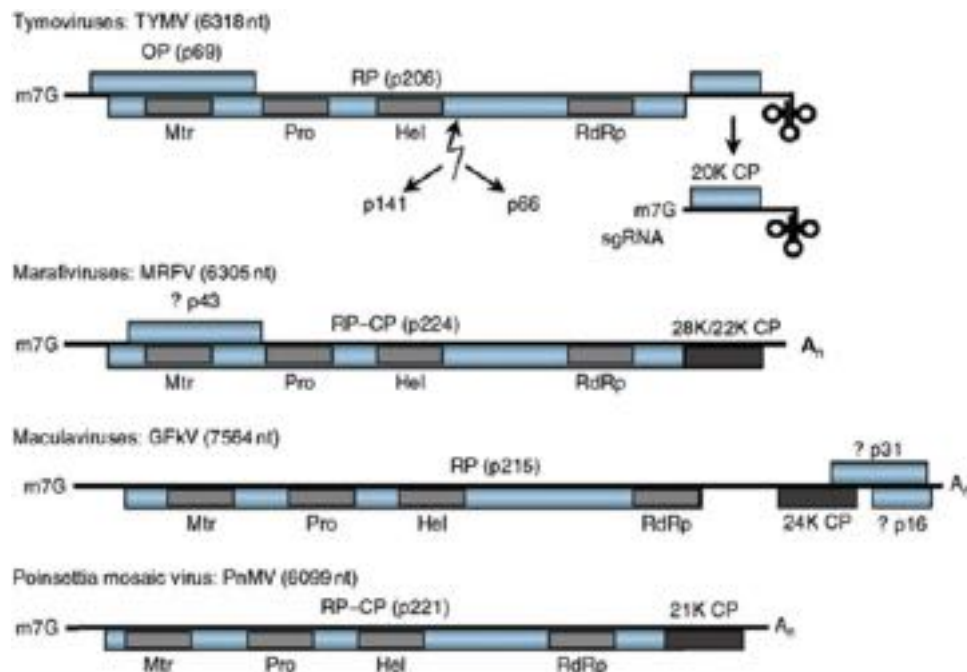


Fig. 2 Genomes of the type members of the genera of the family *Tymoviridae* and of PnMV. All genomes have a 5'-m7GpppG cap, but the 3' terminal structures vary. TYMV has a valine-specific tRNA-like structure (cloverleaf), MRFV, GFKV and PnMV have a poly(A) tail. The known or predicted (?) expressed ORFs are indicated with the molecular weight (K, kDa) of the predicted protein. The replicases (RPs) all possess methyltransferase (Mtr), papain-like proteinase (Pro), helicase (Hel), and RdRp domains. The TYMV RP is cleaved as indicated, and similar cleavages are expected for the other viruses. The TYMV overlapping protein (OP) is the viral movement/RNAi suppressor protein. The TYMV CP is expressed from a sgRNA, and all other CPs are probably also expressed from sgRNAs (not shown). The 28 kDa MRFV CP (whose true size is likely to be closer to 25 kDa) may be produced by proteolytic cleavage from the RP-CP fusion protein, and an analogous event may occur with PnMV. Expression of the MRFV p43 and GFKV p31 and p16 has not yet been validated.

Tymoviruses

Properties and Distinguishing Characteristics

Although virtually all information about tymoviruses has been derived from studies on TYMV, it is considered to be generally applicable to the other members of the genus. Virions of tymoviruses contain genomic RNAs 6.0–6.7 kb long and produce a single 3′ colinear sgRNA less than 1 kb in length encoding the ~20 kDa CP. Their genomes include a distinctive 16-nt-long ‘tymobox’ sequence, and the genomes of most species possess a 3′ tRNA-like structure (TLS) that can be efficiently esterified with valine. The genomes encode three ORFs (Fig. 2), two of which almost completely overlap: the OP and RP ORFs, which begin at AUG codons separated by only 4 nt. Tymoviruses replicate in all major tissues of their host plants, and accumulate to high levels (more than 0.1 mg g⁻¹ leaf tissue). Both filled and empty or near-empty particles accumulate. They are readily transmitted mechanically under laboratory conditions, and are spread over limited distances by beetle vectors in nature. Infection produces mosaic symptoms and a distinctive cytopathy that is evident upon EM observation by the appearance of small vesicles on the surface of chloroplasts, together with vacuolation and clumping of chloroplasts.

Capsid Structure

Virions of tymoviruses are highly stable ~29 nm, *T*=3 icosahedra formed by 180 copies of the single CP, arranged as 12 pentamers and 20 hexamers. These groupings form the vertices with fivefold and sixfold symmetry that constitute the surface peaks visible upon high-resolution EM observation (Fig. 1). Inter-subunit stabilization is provided primarily by hydrophobic protein–protein contacts, allowing the formation of stable shells that appear to be devoid of RNA. These empty or near-empty capsids can account for about one-third of the particles present in infected tissues and are readily identifiable by internal staining in negative-contrast EM (Fig. 1). They sediment as the ‘top component’ at 45–55S in CsCl density gradients, and readily separate from the ‘bottom component’, the infectious 110–120S virions that contain the genomic RNA. Minor components of intermediate density contain a range of sub-genomic-size RNAs that can be translated to yield CP. Roughly equimolar amounts of genomic and sgRNAs are encapsidated, but the precise disposition of sgRNA in the various particles has not been determined. Like many other viral CPs, tymoviral CPs fold into eight-stranded, β -barrel, jelly-roll structures. However, unlike some other CPs, there is no positively charged domain at the N-terminus for interaction with RNA and charge neutralization during packaging. Charge neutralization is thought to be provided by polyamine molecules (mostly spermidine) associated with the RNA. The crystal structures of empty and infectious particles have been determined for TYMV, Desmodium yellow mottle virus (DYMV), and Physalis mottle virus. Consistent with the dominance of protein–protein interactions in particle stabilization, empty and infectious particles have very similar structures.

Packaging signals in tymoviral RNAs have yet to be identified, though RNA recruitment may involve conserved hairpins in the 5′ untranslated region (UTR) and interaction of C-rich segments of the RNA with CP at low pH. Localized areas of low pH (5–6) have been postulated to arise at the surfaces of photosynthesizing chloroplasts. Tymoviral replication occurs in characteristic vesicles that form at the chloroplast surfaces (see below), and capsid formation is most active near these vesicles, suggesting a possible coupling between replication and encapsidation during infection. Despite much effort, there has been no success in developing a cell-free packaging system. These efforts have focused on the possible role of ‘artificial top component’ (ATC) capsids as decapsidation and encapsidation intermediates. ATCs are protein shells devoid of RNA that can be made from infectious virions by treatments such as freeze–thawing and exposure to high pH and pressure. They are similar in structure to infectious virions, but lack a capsomere of six CP molecules that is ejected during the treatment, allowing RNA escape.

Genome Organization

Tymoviral genomic RNAs possess a 5′ m⁷GpppG cap and terminate at the 3′ end in -CC(A) except for Dulcamara mottle virus (DuMV) that is in most cases part of a TLS. The typical tymoviral TLS is just over 80 nt long, has a distinctive pseudoknot close to the 3′ terminus, and is a close structural mimic of cellular tRNA^{Val}. A valine-specific anticodon is present, and the 3′ terminus can be specifically amino-acylated with valine. The valylated RNAs of some, though not all tymoviruses, can form a tight complex in vitro with the GTP-bound form of translation elongation factor eEF1A. Most molecules of encapsidated RNA lack the terminal A residue in the -CC(A) end, which is thought to be added by the host tRNA-specific CCA-nucleotidyl-transferase at the beginning of infection. The main roles of the TYMV TLS are thought to be (1) as a 3′ enhancer of translation initiation, (2) as a regulator of the onset of RNA replication by modulating access by the polymerase to the 3′ end, and (3) in maintaining an intact-CCA 3′ end. A minority of tymoviral genomic RNAs have 3′ UTRs that lack the typical valine-specific TLS (DuMV, Erysimum latent virus, and Nemesia ring necrosis virus, NeRNV).

Tymoviral RNAs act as cap-dependent messenger RNAs. Of the three ORFs, two (OP and RP) are expressed directly from the genomic RNA, whereas the third (CP) is expressed from the 5′ capped sgRNA (Fig. 2). Although the sequence contexts around the OP and RP initiation codons vary, 4 nt always separates these closely spaced AUGs. The RP ORF covers most of the viral genome, and encodes the ~200 kDa RP, the only viral protein essential for supporting viral RNA replication in protoplasts. This ORF encodes discrete domains (Fig. 2) that are discernable by sequence relationships to similar domains encoded by a wide

variety of positive-stranded RNA viral genomes. Particularly the RdRp domain is well conserved and appropriate for phylogenetic comparisons. Closely related, yet distinct, tymoviral genomes have RdRp domains with about 70% nt sequence identity. The RP is translated as a precursor protein that subsequently undergoes maturation cleavage catalyzed by the papain-like proteinase domain. The single known cleavage separates the RdRp domain from the remainder of the RP; thus, for TYMV, the precursor p206 is processed to yield p141 and p66.

The OP ORF almost entirely overlaps the RP ORF. Its length and the sequence of the encoded protein are highly variable. OPs range between 49 and 82 kDa, and have only 25%–40% aa sequence identity between the most closely related pairs of viruses. OP expression is needed for establishment of infection and for spread of the virus in plants. The OP is a suppressor of the host RNA silencing antiviral response and is also believed to be the viral movement protein. It is a substrate for ubiquitin-dependent degradation by the proteasome. Long-distance movement of the virus in plants requires expression of the CP. The CP ORF initiates close to the end of the RP ORF, sometimes even a little upstream. CP sequences are variably conserved among tymoviruses, with 36%–86% aa sequence identity between the most closely related pairs of viruses.

Tymoviral genomes possess a highly conserved sequence, the 16 nt tymobox (–GAGUCUGAAUUGCUUC) with small variations in some viruses, especially wild Cucumber mosaic virus, just upstream of the CP ORF. The tymobox and its associated ‘initiation box’ sequence CAA(U/G) positioned 8 or 9 nt downstream of the tymobox are believed to serve as the core elements of the sgRNA promoter. The tymobox overlaps with the 3′ end of the RP ORF, resulting in the presence of the tripeptide – ELL – near the C terminus of all RPs.

Replication Cycle

The ultrastructural changes that reflect viral replication activity and that result in the distinctive pathology of the chloroplasts have been particularly well described for TYMV. As is typical of positive-stranded RNA viruses, the replication cycle is completed entirely in the cytoplasm, and RNA replication occurs in association with membranes, specifically the chloroplast outer membrane. Vesicles 50–80 nm in diameter form as invaginations of the two outer membranes of chloroplasts. They form before the appearance of virions, which are initially found close to vesicle clusters and later throughout the cytoplasm; in some cases, empty capsids accumulate in nuclei. The RdRp has been localized to zones on chloroplasts that are rich in vesicles, and this localization depends on protein–protein interaction between the proteinase and RdRp domains of the mature RPs p141 and p66, respectively.

RNA replication occurs via the production of full-length minus strands, whose synthesis in TYMV is directed by the 3′ terminal –CCA serving as promoter and initiation site. No other minus-strand promoter elements have been identified in the 3′ UTR. As mentioned above, the tymobox appears to function with the initiation box as the core of the promoter directing sgRNA synthesis by internal initiation on the minus strand. The sgRNA of TYMV has a m7GpppA 5′ terminus.

Studies on the functions of viral proteins and of cis-acting sequences in the genomic RNA have been greatly facilitated by the use of ‘infectious clones’, that is, molecularly cloned cDNA versions of the genome from which infectious RNA can be derived.

Infection and Transmission

Tymoviruses cause chlorotic mosaic, vein-clearing, and mottling symptoms, generally without strong stunting. Host ranges are mostly narrow, and to date, no tymoviruses infecting monocot plants have been isolated. Tymoviruses are transmitted over limited distances by chrysomelid beetles, and some are weakly seed-transmissible. Tymoviruses and PnMV are readily transmitted mechanically and accumulate to high titers in infected plants.

Phylogenetic Relationships and Species Demarcation

In the past, serology using antisera raised against intact virus was the main criterion in classifying tymoviruses and in distinguishing between species. This is no longer the most convenient approach, however, and it has in fact been shown that distinct viruses whose genomic and CP sequences have identities less than 80% and 90%, respectively, can appear to be serologically identical. This seems to be due to similar or identical dominant epitopes within otherwise distinct CPs.

The study of relationships among tymoviruses and to other viruses currently primarily relies on the interpretation of sequences derived from the genomic RNA. The phylogenetic trees based on the sequences of the CP and the RP (**Fig. 3**) show similar relationships, both among the tymoviruses and to the other members of the family *Tymoviridae*. Genomes of the *Tymoviridae* are not under positive selection and recombination might be the major driving force of diversity within the family. For instance, the genome of NeRV, which has a tobamovirus-type TLS capable of amino-acylation with histidine, indicates that recombination can occur and shape tymoviral genome evolution. Also, family-wide recombination analysis showed that *Physalis* mottle virus can be a donor for other marafivirus members. The complete genome sequences are available for 35 tymoviruses or tentative tymoviruses (**Table 1**).

Alignments based on the replicase protein sequences indicate that the capilloviruses, carlaviruses, trichoviruses, and potexviruses, all members of the *Flexiviridae*, are the next most closely related virus groups. More distant sequence relationships indicate that members of the family *Tymoviridae* belong to the alpha-like virus group in the *Tymovirales*, the newly established order of viruses reported by the International Committee on Taxonomy of Viruses.

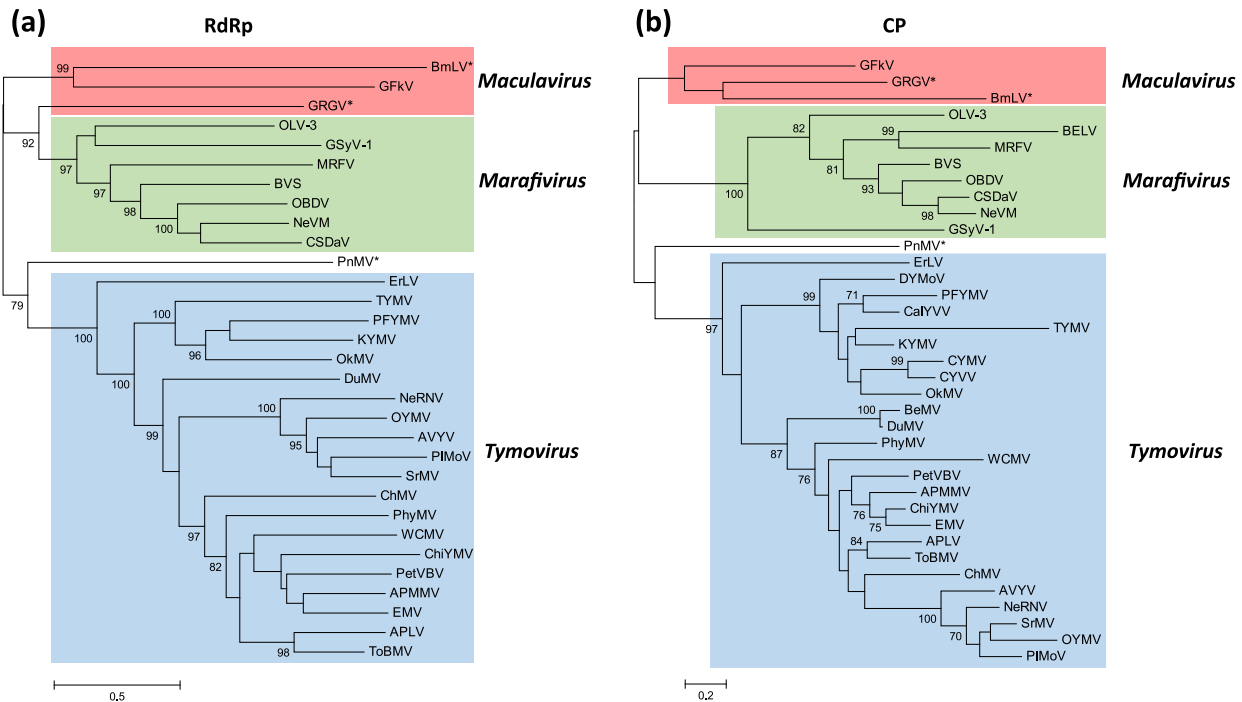


Fig. 3 Maximum-likelihood phylogenetic trees showing the relationships between members of the family *Tymoviridae* based on amino acid sequence alignments of (a) RdRp and (b) coat protein (CP) sequences. Bootstrap support values (%) above 70% are indicated at supported nodes. Branch lengths reflect the difference between amino acid sequences. Viruses from each of the three genera are separately shaded as marked. Viruses with tentative taxonomic status are marked with an asterisk.

Marafiviruses

Properties and Distinguishing Characteristics

Three current members of the genus *Marafivirus* are viruses principally infect monocots (grasses), while others infect citrus, tree fruits, grapevine, and blackberry. Another tentative two marafiviruses were isolated from grapevine. Plant host ranges are narrow. The genomes are 6.3–6.8 kb long, with a single ORF encoding an RP/CP fusion protein covering most of the genome. Two forms of the same CP (~25 and ~21 kDa) are found in virions. A single 3'-colinear sgRNA (< 1 kb) encoding the smaller form of the CP is produced during infection. A slightly modified form of the tymobox sequence, termed 'marafibox', is present in the genome near the junction between the RdRp- and capsid-coding sequences. Marafiviruses are phloem-limited and not mechanically transmissible. The grass-infecting marafiviruses are transmitted by cicadellid leafhoppers in a persistent-propagative manner, involving virus replication in the insect.

A distinctive feature of the marafiviruses is the presence of two CPs in the virions, a major CP ~21 kDa form and a minor N-terminally extended 22–28 kDa form. The latter is thought to arise by proteolytic release from the polypeptide, whereas the former is expressed from the sgRNA. The production of this sgRNA is probably under the control of the marafibox, a 16 nt sequence (–CA(A/G)GGUGAAUUGCUUC–) very closely related to the tymobox, and the –CA(A/U)-initiation box located 8 nt downstream. The roles of the two forms of CP are unclear.

Virion Structure, Genome Organization, Replication Cycle, and Phylogenetic Relationships

Marafiviruses produce virions and empty capsids that are similar to those produced by tymoviruses when observed by negative-staining EM. However, due to the phloem-limited replication of marafiviruses, the yield of virus is low, limiting the usefulness of EM for viral identification. Complete genome sequences are available for eight confirmed marafiviruses (Table 1).

Marafivirus genomic RNAs appear to be 5' capped but they lack the 3' TLS that is distinctive of tymoviruses, and marafivirus genomes possess a 3' poly(A) tail. Almost the entire genome of marafiviruses is devoted to encoding a large 224–240 kDa polypeptide, which includes replication-associated and CP domains (Fig. 2). The relationships between the CP and RP genes of the marafiviruses and other members of the family *Tymoviridae* are indicated in Fig. 3. The polypeptide is believed to be proteolytically processed by the viral proteinase at two sites, between the helicase and RdRp domains as in tymoviruses, and between the RdRp and CP domains. Polypeptide cleavage sites were experimentally mapped in TYMV and OBDV, and aa sequence alignments have

discerned probable cleavage sites occurring immediately downstream of Gly-Gly or Gly-Ala dipeptides at the helicase–RdRp and RdRp–CP junctions in marafivirus polyproteins.

As the number of complete marafivirus genome sequences increases, the taxonomic implications of evident variation in genome organization may be clarified. MRFV possesses a predicted significant overlapping ORF, encoding a putative 43 kDa protein, while Grapevine syrah virus 1 has a predicted 26 kDa protein that shares 43% identity with the central portion of the 43 kDa of MRFV; expression *in vivo* has not been verified for either protein. Related but strongly interrupted coding sequences have been identified overlapping the methyl-transferase domains of other marafiviral genomes and, based on sequence similarities to tymoviral OP, it has been postulated that these sequences represent variously degenerate versions of the tymoviral OP. The loss of OP during evolution may explain the phloem limitation of marafiviruses. Citrus sudden death-associated virus (CSDaV) is the only marafivirus with an additional ORF near the 3' end of the genome. This ORF almost completely overlaps the CP ORF and could encode a 16 kDa protein.

The three grass-infecting marafiviruses, MRFV, OBDV, and Bermuda grass etched-line marafivirus, are serologically related, but are not serologically cross-reactive with tymoviruses or more distantly related viruses. Despite the indications of genome variability mentioned above, the marafiviruses do form a phylogenetic group with interspecies relationships similarly close to those between the various tymoviruses (Fig. 3). Sequence relationships among the RP and CP ORFs clearly link the marafiviruses to the other members of the *Tymoviridae*. Like the tymoviral CPs, marafivirus CPs lack a cluster of basic aa at the N terminus.

Infection and Transmission

Marafi- and maculaviruses are phloem limited. The best-studied marafiviruses are MRFV and OBDV. They cause stunting and chlorotic leaf spots or streaks, general leaf discoloration, or enations on leaf and stem veins. Cytological symptoms are restricted to the phloem and adjacent parenchyma cells, with clearly discernible hyperplasia and hypertrophy. These viruses are not mechanically transmissible or transmissible through seed, but are vectored in nature by leafhoppers. Significantly, MRFV and OBDV are maintained for long periods and through molts in leafhopper vectors, and transmission can occur from viruliferous vectors over a period of several weeks. OBDV replicates in the aster leafhopper (*Macrostelus fascifrons*), and the transmission behavior of MRFV is also consistent with replication in the cicadellid leafhopper (*Dalbulus maidis*). MRFV is commonly associated and co-transmitted with maize stunt spiroplasma or maize bushy stunt phytoplasma in causing economically significant outbreaks of corn stunt syndrome. OBDV can also be co-transmitted with a phytoplasma. The presence of virus in the leafhopper vectors is not associated with detectable symptoms or loss of reproductive fitness.

The most economically significant marafivirus is probably CSDaV. This virus is believed to be the causative agent of the citrus sudden death syndrome, which has had devastating effects on orange trees in Brazil. Preliminary evidence suggests transmission by aphids. The grapevine-infecting tentative marafiviruses, Grapevine asteroid mosaic-associated virus and Grapevine rupestris vein-feathering virus, have no known insect vector.

Maculaviruses

Properties and Distinguishing Characteristics

Recent sequencing of the complete genome of GfKV supported the creation of the new genus *Maculavirus*. The distinguishing characteristics were the length of the GfKV genome, the absence of a tymobox/marafibox, the lack of a fused RP/CP ORF as present in marafivirus genomes, and the lack of phylogenetic clustering with either tymoviruses or marafiviruses. Grapevine red globe virus (GRGV) is a tentative member of the genus. Both viruses infect only *Vitis* species (grapevine and relatives). In common with marafiviruses, these maculaviruses are restricted to the phloem and are not sap transmissible. No insect vectors have been identified. As for all other members of the family *Tymoviridae*, maculaviruses produce ~29 nm particles with prominent surface features observable by EM. Top component and infectious bottom component particles are produced. GfKV infections are cytologically distinct from marafivirus infections through the presence of severely modified mitochondria ('multivesiculate bodies').

The genome of GfKV (7564 nt) is the longest among the members of the family *Tymoviridae*, and has the highest cytosine content (49.8%). The arrangement of RP and CP ORFs is like that of tymoviruses (Fig. 2). The RP ORF encodes a 215 kDa protein with replication-associated domains that present sequence relatedness to the tymoviral and marafiviral RPs. The aa sequence of the RdRp domain shows the closest relationship to sister genera (59% and 67% identities with the RdRps of a tymovirus and marafivirus, respectively).

The 24.5 kDa GfKV CP has 23%–31% sequence identity with CPs of tymoviruses and marafiviruses, and like other CPs of members of the family *Tymoviridae* it lacks a cluster of basic aa at the N-terminus. The GfKV genome encodes two additional proteins from ORFs that partially overlap the CP ORF in the 3' region of the genome. No functions for these putative proteins are predicted from sequence comparisons. Close relationship between GfKV and GRGV is rather weakly supported by sequence comparisons (Fig. 3).

The GfKV genome is probably capped at the 5' end and has a poly(A) tail at the 3' end, in common with most marafiviruses and PnMV. It is the only known member of the family *Tymoviridae* to lack a tymobox- or marafibox-like sequence. Nevertheless, two or more sgRNAs, ~1.3 and ~1.0 kb in length, are present in infected tissues. These RNAs appear to be packaged in both top- and bottom-component particles, in the latter case apparently together with the genomic RNA.

Poinsettia Mosaic Virus

Properties and Distinguishing Characteristics

PnMV possesses properties characteristic of both the tymoviruses and marafiviruses. The ~28 nm virus particles have the EM structure typical of the *Tymoviridae* and separate into typical top components containing the sgRNA and infectious bottom components. Like tymoviruses, PnMV is produced in high yield during infection, is not phloem-limited, and can be mechanically transmitted in the laboratory. Replication is associated with chloroplast cytopathy, though not with the appearance of the replication-associated vesicles typical of tymovirus infection. The properties and coding arrangement of the 6.1 kb RNA are more closely aligned with those of the marafiviruses (Fig. 2). The genomic RNA is polyadenylated, contains a marafibox-related putative sgRNA promoter, and possesses a single long ORF that encodes an RP/CP fusion protein (p221). No OP ORF is present. The 21–24 kDa PnMV CP is presumably expressed from the 0.65 kbp sgRNA; the involvement of potential cleavage of the RP/CP fusion protein to produce a second CP variant is not known. Although PnMV sequences are more closely related to the marafiviruses than tymoviruses, the relationships appear to be insufficiently close to warrant inclusion in the genus *Marafivirus* (Fig. 3).

In commercial poinsettia cultivation, PnMV is transmitted by vegetative propagation; the virus is not transmissible via seed or pollen. The natural host range is restricted to *Euphorbia* sp., especially poinsettia, *E. pulcherrima*. Symptoms appear seasonally, varying from unapparent to light mottling. PnMV is often associated with poinsettia cryptic virus (family *Partitiviridae*) and with a phytoplasma that is responsible for the desirable free-branching phenotype.

Bombyx Mori Latent Virus

Properties and Distinguishing Characteristics

Bombyx mori latent virus (BmLV) was initially identified in a cultured cell line (BmN) of the silk moth (*Bombyx mori*) by cDNA microarray. The full-length sequence is 6153 nt excluding the poly(A) tail. Putative isometric virions of ~28–30 nm were observed and were found to be infections to insect Sf-9 cells and tissues of *B. mori* larvae. The genome encodes a protein of 1747 aa residues containing the conserved MTR, PRO, HEL, and POL regions of tymovirus replication polyprotein. It shares 44% aa identity to GFkV. The single CP is encoded in a second ORF and is expressed from a 1.25 kb sub-genomic RNA. A third putative ORF is located at the 3' terminus of the genome and encodes a putative 15.5 kDa protein. The genome lacks the 'tymobox' or 'marafibox' sub-genomic promoter. Phylogenetic analysis of the replicase and CP sequences suggests that BmLV is a putative member of the family *Tymoviridae*, distinct from tymo- and marafiviruses, but closely related to GFkV. In the absence of additional data, it has been designated as an unassigned species in the family.

Further Reading

- Bradel, B.G., W. Preil, W., Jeske, H., 2000. Sequence analysis and genome organization of Poinsettia mosaic virus (PnMV) reveal closer relationship to marafiviruses than to tymoviruses. *Virology* 271, 289–297.
- Canady, M.A., Larson, S.B., Day, J., McPherson, A., 1996. Crystal structure of Turnip yellow mosaic virus. *Nature Structural & Molecular Biology* 3, 771–781.
- Ding, S.W., Howe, J., Keese, P., *et al.*, 1990. The tymobox, a sequence shared by most tymoviruses: Its use in molecular studies of tymoviruses. *Nucleic Acids Research* 18, 1181–1187.
- Dreher, T.W., 2004. Turnip yellow mosaic virus: Transfer RNA mimicry, chloroplasts and a C-rich genome. *Molecular Plant Pathology* 5, 367–375.
- Edwards, M.C., Zhang, Z., Weiland, J.J., 1997. Oat blue dwarf marafivirus resembles the tymoviruses in sequence, genome organization, and expression strategy. *Virology* 232, 217–229.
- Hirth, L., Givord, L., 1988. Tymoviruses. In: Koenig, R. (Ed.), *The Plant Viruses*. New York: Plenum, pp. 163–212.
- Jakubiec, A., Notaise, J., Tournier, V., *et al.*, 2004. Assembly of Turnip yellow mosaic virus replication complexes: Interaction between the proteinase and polymerase domains of the replication proteins. *Journal of Virology* 78, 7945–7957.
- Kadaré, G., Rozanov, M., Haenni, A.-L., 1995. Expression of the Turnip yellow mosaic virus proteinase in *Escherichia coli* and determination of the cleavage site within the 206 kDa protein. *Journal of General Virology* 76, 2853–2857.
- Katsuma, S., Tanaka, S., Omuro, N., *et al.*, 2005. Novel macula-like virus identified in *Bombyx mori* cultured cells. *Journal of Virology* 79, 5577–5584.
- Koenig, R., Pleij, C.W., Lesemann, D.E., Loss, S., Vetten, H.J., 2005. Molecular characterization of isolates of Anagris vein yellowing virus, Plantago mottle virus and Scrophularia mottle virus: Comparison of various approaches for tymovirus classification. *Archives of Virology* 150, 2325–2338.
- Maccheroni, W., Alegria, M.C., Greggio, C.C., *et al.*, 2005. Identification and genomic characterization of a new virus (*Tymoviridae* family) associated with citrus sudden death disease. *Journal of Virology* 79, 3028–3037.
- Martelli, G.P., Sabanadzovic, S., Abou-Ghanem Sabanadzovic, N., Edwards, M.C., Dreher, T., 2002. The family *Tymoviridae*. *Archives of Virology* 147, 1837–1846.
- Mayo, M.A., Dreher, T.W., Haenni, A.-L., 2000. Genus tymovirus. In: van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., *et al.* (Eds.), *Seventh Report of the International Committee on Taxonomy of Viruses*. New York: Academic Press.
- Morch, M.D., Boyer, J.C., Haenni, A.-L., 1988. Overlapping open reading frames revealed by complete nucleotide sequencing of Turnip yellow mosaic virus genomic RNA. *Nucleic Acids Research* 16, 6157–6173.

- Sabanadzovic, S., Ghanem-Sabanadzovic, N.A., Saldarelli, P., Martelli, G.P., 2001. Complete nucleotide sequence and genome organization of Grapevine fleck virus. *Journal of General Virology* 82, 2009–2015.
- Weiland, J.J., Dreher, T.W., 1989. Infectious TYMV RNA from cloned cDNA: Effects in vitro and in vivo of point substitutions in the initiation codons of two extensively overlapping ORFs. *Nucleic Acids Research* 17, 4675–4687.

Relevant Websites

- <https://viralzone.expasy.org/54>
Chordopoxvirinae
ViralZone.
- https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/245/tymoviridae
Tymoviridae
Positive Sense RNA Viruses.
- <https://link.springer.com/protocol/10.1385/0-89603-385-6.219>
Tymovirus Isolation and Genomic RNA Extraction.

Umbraviruses (*Tombusviridae*)

Eugene V Ryabov, USDA, Agricultural Research Service, Beltsville, MD, United States

Michael E Taliany, The James Hutton Institute, Dundee, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

This is an update of M. Taliany, E. Ryabov, Umbravirus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:[10.1016/B978-012374410-4.00524-0](https://doi.org/10.1016/B978-012374410-4.00524-0).

Glossary

Fibrillarin An abundant nucleolar protein that participates in the processing and modification of rRNAs and a methyltransferase.

Nucleolus A prominent subnuclear domain and the site of transcription of ribosomal RNA (rRNA), processing of the pre-rRNAs and biogenesis of pre-ribosomal particles, and

also participates in other aspects of RNA processing and cell function.

Phloem Plant vascular system used for rapid long-distance transport of assimilates and macromolecules.

Plasmodesmata Plant unique intercellular membranous channels that span plant cell walls linking the cytoplasm of contiguous cells.

Introduction

Umbraviruses differ from other plant viruses in that they do not encode a coat protein (CP), and thus no conventional virus particles are formed in the infected plants. The name of the genus *Umbravirus* is derived from the Latin *umbra*, which means a shadow, or an uninvited guest who comes with an invited one. This name reflects the way in which umbraviruses depend for survival in nature on an assistor virus, which is always a member of the *Enamovirus* or *Polerovirus* genera of the family *Luteoviridae*. For transmission between plants, the CP of the assistor virus forms aphid-transmissible hybrid virus particles encapsidating umbraviral RNA. In nature, each umbravirus is associated with one particular member of the family *Luteoviridae*, although in experiments trans-capsidation is not needed for umbravirus infection accumulation within infected plants because functions such as protection and movement of the virus RNA, both cell-to-cell and long distance, do not require the presence of the assistor virus or its CP. Moreover, under experimental conditions, mechanical transmission of umbraviruses can take place without the aid of an assistor virus. This implies that umbraviruses encode some product(s) that functionally compensate for the lack of a CP (see below).

Classification

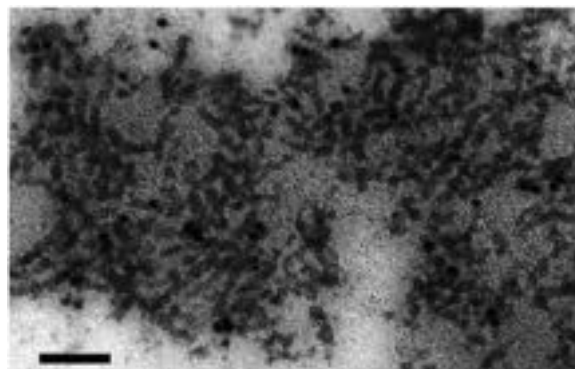
Currently the genus *Umbravirus* includes nine distinct virus species which are accepted by ICTV taxonomy: *Carrot mottle virus* (CMoV) – the type species, *Carrot mottle mimic virus* (CMoMV), *Groundnut rosette virus* (GRV), *Ethiopian tobacco bushy top virus* (ETBTIV), *Lettuce speckles mottle virus* (LSMV), *Opium poppy mosaic virus* (OPMV), *Pea enation mosaic virus-2* (PEMV-2), *Tobacco mottle virus* (TMoV), and *Tobacco bushy top virus* (TBTIV). These viruses together with the tentative members of the genus *Umbravirus* are listed in [Table 1](#). Some of these viruses have been known since the early days of plant virology. The first to be described was TMoV, reported from Zimbabwe and Malawi in 1945. The most economically important umbravirus is GRV, which is endemic throughout sub-Saharan Africa. Sporadic, unpredictable outbreaks of groundnut rosette disease (actually caused by a satellite RNA of GRV) cause severe crop losses. Pea enation mosaic disease outbreaks are also sporadic and localized, but losses of nearly 90% in peas and up to 50% in field beans have been reported. One or both of the carrot-infecting umbraviruses probably occurs worldwide wherever carrots are grown, but they are uncommon in commercial crops because the insecticides used to control carrot fly also control the aphid vectors of CMoV and CMoMV. LSMV is reported only from California, USA.

Virion Structure

Although in plants infected with umbraviruses unaccompanied by the assistor virus no virions are formed, the infectivity of CMoV and GRV in buffer extracts of infected leaves is surprisingly stable: survives for several hours at room temperature and is resistant to treatment with ribonuclease. It is however, sensitive to a treatment with organic solvents, suggesting that lipid-containing structures are involved in protection of umbravirus RNA. Indeed, in plants infected with a number of umbraviruses, including CMoV, enveloped structures ~50 nm in diameter were observed. An infective fraction from GRV-infected tissue contained complexes with a buoyant density of 1.34–1.45 g/cm³ consisting of filamentous ribonucleoprotein particles, composed of the umbraviral ORF3 protein (see below) and virus RNA, embedded in a matrix ([Fig. 1](#)). The relationship between these two kinds of structure is unclear.

Table 1 Virus members of the genus *Umbravirus*

<i>Virus (isolate)</i>	<i>Abbreviation</i>	<i>Accession no.</i>	<i>Assistor virus</i>
<i>Carrot mottle mimic virus</i> Carrot mottle mimic virus-Australia	(CMoMV-AU)	U57305	<i>Carrot red leaf virus</i>
<i>Carrot mottle virus</i> Carrot mottle virus-Weddel	(CmoV-We)	FJ188473	<i>Carrot red leaf virus</i>
<i>Ethiopian tobacco bushy top virus</i> Ethiopian tobacco bushy top virus 18–2 Satellite RNA	(ETBTv18–2)	KJ918748 KJ918747	<i>Potato leafroll virus</i>
<i>Groundnut rosette virus</i> Groundnut rosette virus-MC1 Satellite RNAs	(GRV-MS1)	Z69910 Z29702-Z29711	<i>Groundnut rosette assistor virus</i>
<i>Lettuce speckles mottle virus</i> Lettuce speckles mottle virus-California Opium poppy mosaic virus Opium poppy mosaic virus-PHEL5235	(LSMV-CA) (OPMV-PHEL5235)	MG182693	<i>Beet western yellows virus</i> Opium poppy mosaic-associated virus
<i>Pea enation mosaic virus-2</i> Pea enation mosaic virus-2-WGS Bean yellow vein-banding virus Satellite RNA	(PEMV-2-WGS) (BYVbV)	U03563 U03564	<i>Pea enation mosaic virus-1</i>
<i>Tobacco bushy top virus</i> Tobacco bushy top virus-Baoshan	(TBTv-Bao)	AF402620	<i>Tobacco vein distorting virus</i>
<i>Tobacco mottle virus</i> Tobacco mottle virus-Zimbabwe	(TMoV-ZW)	AY007231	<i>Tobacco vein distorting virus</i>
<i>Tentative members in the genus</i> <i>Sunflower crinkle virus</i> <i>Sunflower yellow blotch virus</i> <i>Tobacco yellow vein virus</i>	(SuCV) (SuYbV) (TYVV)		

**Fig. 1** Electron microphotograph of a section of a *Nicotiana benthamiana* cell expressing GRV ORF3 protein. The section shows a complex of filamentous RNP particles embedded in an electron dense matrix. The section was labeled by *in situ* hybridization with an RNA probe specific for viral RNA. Bar, 100 μ m.

Genome Organization and Expression

The genome of umbraviruses consists of single linear segment of positive-sense single stranded RNA (ssRNA) (**Fig. 2**). The complete genomic RNA sequences of umbraviruses are of 4019–4253 nt in length. There is no polyadenylation at their 3' termini and there is no information about modifications of their 5' termini. At the 5' end, a very short non-coding sequence precedes ORF1, which encodes a putative 31–37 kDa protein. The ORF2, which overlaps the 3' end of ORF1, potentially encodes a

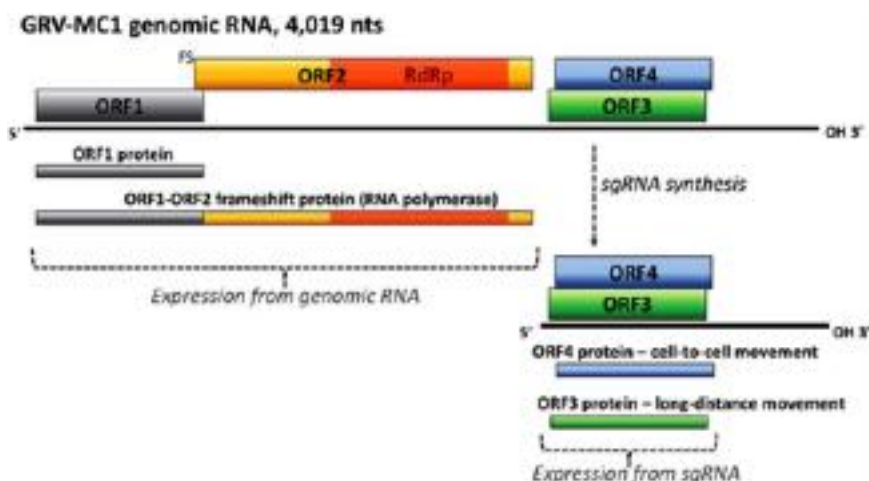


Fig. 2 Organization and expression strategy of GRV genome. Translation of the ORF1 and the ORF2 with a frameshift event results in production of the ORF1-ORF2 frameshift protein; subgenomic RNA (sgRNA) synthesis is required for expression of the ORF3 and ORF4. Solid lines represent RNA molecules, boxes represent open reading frames (ORFs), corresponding translation products are shown below. RdRp, RNA-dependent RNA polymerase; FS, – 1 frameshift signal.

63–65 kDa protein, but lacks an AUG initiation codon. It is likely that ORF1 and ORF2 are translated as a single 94–98 kDa protein by a –1 translational frameshifting mechanism; the sequence associated with frameshifting events in genomes of several animal and plant RNA viruses is found in the region at the 3' end of the ORF1. The predicted amino acid sequence of the ORF2-encoded product includes an RNA-dependent RNA polymerase (RdRp) domain typical to that of other members of the family *Tombusviridae*, containing all eight conserved motifs of RdRps of positive-strand RNA viruses.

A short untranslated stretch separated ORF2 from ORF3 typical to that of and ORF4 which are almost completely overlap in different reading frames and each encode a 26–29 kDa products. The ORF4 contains motifs characteristic of the cell-to-cell movement proteins (MPs) of plant viruses, in particular with those of cucumoviruses. The ORF3 products possess up to 50% homology to each other but shows no significant similarity to any other viral or non-viral product. The ORF3 and ORF4 are translated from subgenomic RNA, the 3' terminal RN

A of the appropriate size was detected in the GRV infected plant tissue. It should be pointed out that the sequenced genomes of umbraviruses do not encode a potential CP(s). The essential role of the ORF3-ORF4 block in movement of umbraviruses within plants has been experimentally demonstrated and is discussed below.

The 3' untranslated region of PEMV2 contains three 3' proximal, cap-independent translation enhancers (3'CITEs), which are required for proper virus translation. Efficient translation of the ORF3 and ORF4 proteins from the sgRNA requires all three (3'CITEs), whereas translation of the genomic requires only two 3'CITEs.

Similarity With Other Taxa

The only gene conserved across all RNA viruses is the RdRp which is cenral to virus replication and existence, therefore RdRp was used to investigate global phylogenetic relationships between viral taxonomic groups. Based on the RdRp phylogeny, umbraviruses were included in the family *Tombusviridae*, alongside with the members of the genera *Alphacarmovirus*, *Alphanecrovirus*, *Aureusvirus*, *Avenavirus*, *Betacarmovirus*, *Dianthovirus*, *Gammacarmovirus*, *Macanavirus*, *Machlomovirus*, *Panicovirus*, *Pelarspovirus*, *Tombusvirus*, and *Zeavirus*, all of which infect plant hosts.

A recent analysis of the RdRp sequences of all known RNA viruses showed that the family *Tombusviridae*, together with the genus *Luteovirus* and the family *Carmotetraviridae*, with which it has highest RdRp homology, belongs so-called "Branch 3" group of global RdRp and reverse transcriptase phylogeny, which also includes subset of positive-sence RNA viruses of viruses, such as "alphavirus supergroup", "flavivirus supergroup" and nodaviruses. Members of the Branch 3 infect hosts across all eukaryotic kingdoms.

Replication

Replication of umbravirus RNA presumably involves the ORF2-encoded RdRp. Leaves of plants infected with umbraviruses contain abundant dsRNA including a major species of about 4.4–4.8 kbp corresponding in size to that expected for double stranded form of the viral genomic RNA which may be umbravirus RNA replication intermediate. No other details of the replication mechanism have been elucidated.

Satellite RNA

Satellite RNAs are associated with some umbraviruses. In the case of GRV, satellite RNA is found in all naturally occurring isolates, and is primarily responsible for the symptoms of groundnut rosette disease. GRV satellite RNA is a ssRNA of about 900 nt which relies on GRV for its replication and, more unusually, is also required for the *Groundnut rosette assistor virus* (GRAV)-dependent aphid transmission of GRV. Thus, unlike most virus satellite RNAs, it is essential for the biological survival (though not the replication) of its helper virus. The role of the satellite RNA in the transmission process is to mediate transcapsidation of GRV RNA by GRAV protein to form stable aphid-transmissible hybrid virus particles. Although different GRV satellite RNA variants contain up to five potential ORFs, none of the ORFs is essential for any of the functions and biological properties that have been ascribed to GRV satellites. In contrast, the satellite RNA that is associated with some isolates of PEMV-2 is not required for transcapsidation of PEMV-2 RNA by the CP of its assistor virus PEMV-1 or for aphid transmission of the hybrid particles, and other umbraviruses, such as CMoV, do not have satellite RNAs, yet are transcapsidated by their assistors and thus transmitted by aphids. The reasons for these differences have not been explained.

Cell-to-Cell Movement Function

The highly conserved ORF4-encoded proteins of umbraviruses exhibit significant sequence similarity with cell-to-cell movement proteins (MPs) of other plant viruses, in particular the 3a proteins of cucumoviruses. Therefore it has been suggested that this protein involved in cell-to-cell trafficking of umbravirus RNA through plasmodesmata. This suggestion has been confirmed by a number of genetic, cytological and biochemical approaches. By using the gene replacement strategy, it was demonstrated that the GRV ORF4 protein could functionally replace the MPs of unrelated viruses, *Potato virus X* (PVX) (all the products encoded by the triple gene block and the CP) and *Cucumber mosaic virus* (CMV) (the 3a MP and the CP). Localization of the GRV ORF4 protein has much in common with localization of MPs of other plant virus groups. The green fluorescent protein (GFP)-tagged GRV ORF4 protein targeted to plasmodesmata ([Fig. 3\(A\)](#)). Also, this protein formed extending tubular structures on the surface of protoplasts infected either with GRV or with the heterologous virus expressing the GFP-tagged GRV ORF4. The GRV ORF4 protein binds to RNA *in vitro* in non-cooperative manner which may make the viral RNA of the ribonucleoprotein (RNP) complex accessible for translation and replication.

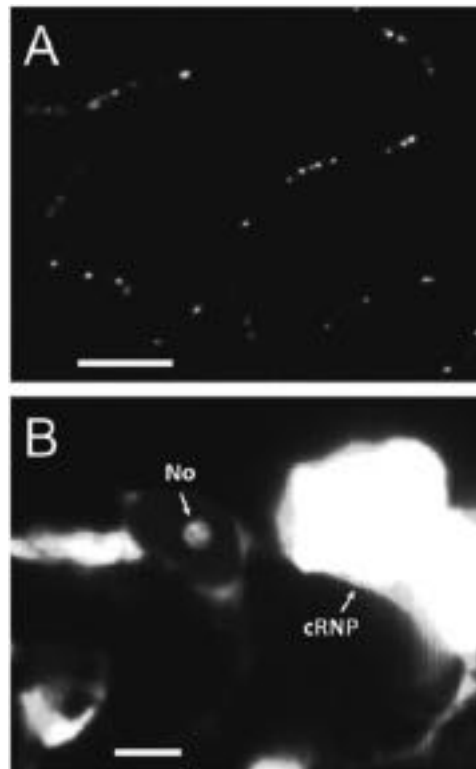


Fig. 3 Confocal laser scanning images showing the localization of (A) GRV ORF4 protein fused to GFP to plasmodesmata (shown by arrows) (bar, 25 μ m) and (B) GRV ORF3 protein fused to GFP to the nucleolus (No) and cytoplasmic inclusions containing ORF3 protein RNP particles (cRNP) (bar, 10 μ m).

Phloem-Dependent Long-Distance Movement Function

One of the striking features of umbraviruses is their ability to move long distance within a plant without having a conventional CP. Involvement of CPs in long distance movement of viral infection was shown for a number of plant viruses, including *Tobacco mosaic virus* (TMV) which utilizes the CP exclusively for its long-distance but not cell-to-cell movement function. By using the gene replacement strategy it was demonstrated that the ORF3 proteins of GRV, PEMV-2 and ToMV were able to functionally substitute the TMV CP in the long-distance movement process. The hybrid TMV mutants expressing the umbraviral ORF3 proteins were able to move rapidly within through the phloem. It was also shown that specific mutations in the ORF3 protein of PEMV-2 abolish ability of this virus to move long distance without affecting its cell-to cell spread within inoculated leaves.

It should be noted that mechanisms of the long-distance movement facilitated by umbraviral ORF3 proteins is different from those mediated by suppressors of RNA silencing such as the 2b proteins of cucumber mosaic virus and the HC-Pro protein of potyviruses. Rather than supporting virus infection by suppressing the host RNA-mediated response, the ORF3 protein seems to protect RNA by binding to viral RNA. It was shown that the ORF3 proteins of GRV, PEMV-2 and ToMV increase stability of viral RNA and protect it from RNases. Immunogold labeling and *in situ* hybridization experiments showed that the GRV ORF3 protein accumulated in cytoplasmic granules consisted of filamentous RNP particles composed of the ORF3 protein and viral RNA (Fig. 1). It has been suggested that these particles may be a form in which viral RNA moves long-distance through the phloem. Also the ORF3-containing RNP complex may protect viral RNA from nucleases and RNA silencing machinery.

Involvement of Nucleolus in Umbravirus Systemic Infection

In addition to cytoplasmic granules, the GRV ORF3 protein was also found in nuclei, preferably targeting nucleoli (Fig. 3(B)). Sequence analysis of the ORF3 proteins of umbraviruses revealed the presence of both nuclear localization signal (NLS) and nuclear export signal (NES), functional role of which in nuclear import and export of the GRV ORF3 has been confirmed experimentally by the genetic analysis. Functional analysis of ORF3 protein mutants revealed a correlation between the ORF3 protein nucleolar localization and its ability to form the RNP particles and transport viral RNA long distances. It was also shown that the ORF3 protein directly interacts with a nucleolar protein, fibrillarin, re-distributing it from the nucleolus to cytoplasm and such an interaction is absolutely essential for umbravirus long-distance movement through the phloem. The umbraviral ORF3 protein utilizes trafficking pathways involving Cajal bodies to enter the nucleolus and, along with fibrillarin, exit the nucleus to form viral 'transport-competent' RNP particles in the cytoplasm.

Interaction With Assistor Virus

Although umbraviruses accumulate and spread very efficiently within infected plants, they depend on assistance of luteoviruses for their survival in nature as they require encapsidation by luteovirus CP for horizontal transmission by aphids. In its turn, umbraviruses facilitate movement of phloem limited luteoviruses to and from the phloem, as well as cell-to-cell movement of luteoviruses between mesophyll and epidermal cells. It was shown that such an ability to promote cell-to-cell movement of luteoviruses is a unique feature of the umbraviral ORF4 MP. Moreover, the ORF4 MP of GRV can facilitate cell-to-cell movement of *Potato leafroll luteovirus* even when expressed from heterologous PVX or CMV genomes. In most instances, the luteovirus partner does not depend on the umbravirus infection for survival in nature. A complex consisting of PEMV-1 (the genus *Enamovirus*, family *Luteoviridae*) and PEMV-2 is a notable exception. Unlike other members of the family *Luteoviridae*, PEMV-1 on its own lacks the ability to move, even through the phloem; both long-distance and cell-to-cell movement functions of PEMV-1 are provided by the umbraviral component of the complex, PEMV-2. Such a strong mutual dependence and adaptation between umbraviruses and luteoviruses suggest a long co-evolution, which had resulted in establishing a range of complexes from facultative co-existence to complete dependence of PEMV-1.

Host Range

Individual umbraviruses are confined in nature to one or two host plant species. For example, groundnut is the only known natural host of GRV and entire rosette disease complex (GRV, its satellite RNA and GRAV). Experimental host ranges of umbraviruses are broader but still restricted. The symptoms induced in infected plants are usually mottles and mosaics.

Transmission

In nature, umbraviruses are transmitted by aphids, but only from plants that also infected with an assistor member of the family *Luteoviridae*. The mechanism of the dependent transmission is the encapsidation of the umbraviral RNA by the CP provided by the assistor virus. Hence, the transmission of the dependent umbravirus occurs in the same persistent (circulative, nonpropagative) manner as that of the luteovirus-assistor.

Prevention and Control

For the avoidance of pea enation mosaic disease, it is recommended that pea or faba bean crops should be sited away from alfalfa and clover fields. Early sowing and close spacing can reduce groundnut rosette disease incidence, probably by inhibiting the landing response of the vector. This approach is also effective against tobacco rosette disease. However, the best control approach is to use resistant cultivars if they are available. Resistance to GRV controlled by two independent recessive genes has been found and groundnut lines possessing this resistance have been developed. A possibility for the future is the deployment of transgenic forms of resistance. Strategies for engineering transgenic resistance against umbraviruses include the transformation of plants with translatable or non-translatable sequences from the umbraviruses themselves or their satellite RNAs.

Further Reading

- Demler, S.A., Borkhsenius, O.N., Rucker, D.G., de Zoeten, G.A., 1994. Assessment of the autonomy of replicative and structural functions encoded by the luteo-phase of pea enation mosaic virus. *Journal of General Virology* 75, 997–1007.
- Gao, F., Alekhina, O.M., Vassilenko, K.S., Simon, A.E., 2018. Unusual dicistronic expression from closely spaced initiation codons in an umbravirus subgenomic RNA. *Nucleic Acids Research* 46, 11726–11742.
- Gao, F., Kasprzak, W.K., Szarko, C., Shapiro, B.A., Simon, A.S., 2014. The 3' Untranslated region of Pea enation mosaic virus contains two T-shaped, ribosome-binding, cap-independent translation enhancers. *Journal of Virology* 88, 11696–11712.
- Kim, S.H., Ryabov, E.V., Brown, J.W., Taliany, M., 2004. Involvement of the nucleolus in plant virus systemic infection. *Biochemical Society Transactions* 32, 557–560.
- Kim, S.H., Ryabov, E.V., Kalinina, N.O., *et al.*, 2007. Cajal bodies and the nucleolus are required for a plant virus systemic infection. *The EMBO Journal* 26, 2169–2179.
- Nurkiyanova, K.N., Ryabov, E.V., Kalinina, N.O., *et al.*, 2001. Umbravirus-encoded movement protein induces tubule formation on the surface of protoplasts and binds RNA incompletely and non-cooperatively. *Journal of General Virology* 82, 2579–2588.
- Ryabov, E.V., Robinson, D.J., Taliany, M., 2001. Umbravirus-encoded proteins that both stabilize heterologous viral RNA *in vivo* and mediate its systemic movement in some plant species. *Virology* 288, 391–400.
- Ryabov, E.V., Robinson, D.J., Taliany, M.E., 1999. A plant virus-encoded protein facilitates long distance movement of heterologous viral RNA. *Proceedings of the National Academy of Sciences of the United States of America* 96, 1212–1217.
- Taliany, M., Roberts, I.M., Kalinina, N., *et al.*, 2003. An umbraviral protein, involved in long-distance RNA movement, binds viral RNA and forms unique, protective ribonucleoprotein complexes. *Journal of Virology* 77, 3031–3040.
- Taliany, M.E., Robinson, D.J., 2003. Molecular biology of umbraviruses: Phantom warriors. *Journal of General Virology* 84, 1951–1960.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9. (e02329-18).

Varicosaviruses (*Rhabdoviridae*)

Takahide Sasaya, National Agriculture and Food Research Organization, Fukuyama, Japan

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Big-vein disease of lettuce A lettuce disease that induces abnormal dilation or enlargement of a vein or the artery in a lettuce leaf, decreasing market value and reducing the proportion of harvestable lettuce. The name 'big-vein disease' refers to the appearance of the major symptoms.

L protein The main subunit of the polymerase complex, responsible for most of the functions required for transcription and replication: RNA-dependent RNA polymerase, mRNA 5'-capping, 3'-poly(A) synthesis, and protein kinase activities.

Necrotic syndrome of lettuce A lettuce disease that induces necrotic rings and spots in lettuce leaf and root

under a hydroponic culture of a nutrient film technique, decreasing market value and reducing the yield of lettuce.

Olpidium virulentus A obligate parasitic root-infecting fungus and a member of the *Chytridiomycetes* that seldom appears to be deleterious to the host plants but functions as vector of plant viruses.

Stop-and-start model A model for RNA synthesis of nonsegmented negative-sense RNA viruses that the viral polymerase complex transcribes a gene, polyadenylates and releases mRNA, and reinitiates transcription of the next gene under the direction of the gene-junction sequence.

Introduction

The big-vein disease of lettuce (*Lactuca sativa* L.) was first reported in 1934. The causal agent was postulated to be a root-infecting virus, but the virus particles have remained unidentified for more than half a century. In 1983, a rod-shaped virus was found in field-grown lettuce showing big vein-band chlorosis (Fig. 1). Because the rod-shaped virus is always associated with big-vein-affected lettuce, the virus was regarded as the agent that induced big-vein disease, without it being confirmed that the virus alone actually induced big-vein symptoms in lettuce. The rod-shaped virus was named Lettuce big-vein virus and was classified as the type species of a new genus *Varicosavirus*, which was established by the International Committee on the Taxonomy of Viruses in 2000.

A second virus, Mirafiori lettuce big-vein virus (MLBVV, genus *Ophiovirus*), has been found in commonly coinfecting field-grown lettuce showing big-vein symptoms. Mechanical inoculation of partially purified MLBVV-preparation and transmission experiments using MLBVV- or Lettuce big-vein virus-carrying fungal vectors independently showed that MLBVV, but not Lettuce big-vein virus, was the causal agent of big-vein disease. For these reasons, lettuce big-vein virus was renamed lettuce big-vein associated virus (LBVaV) in 2005.

LBVaV does not seem to induce vein-band chlorosis in lettuce, which is the characteristic symptom of the big-vein disease. However, LBVaV induces necrotic rings and spots in lettuce under a hydroponic culture of a nutrient film technique and it has become clear that the virus is the causal agent of the necrotic syndrome of lettuce.

Stunt disease of tobacco (*Nicotiana tabacum* L.) reported so far only in Japan was first recorded in 1931 as a seed-bed disorder. Stunt disease has seriously damaged tobacco leaf production in Japan until the 1960s, when the disease was found to be transmitted by an obligate parasitic soil-borne fungus and tobacco growers paid more attention to tobacco seed-bed sanitation.

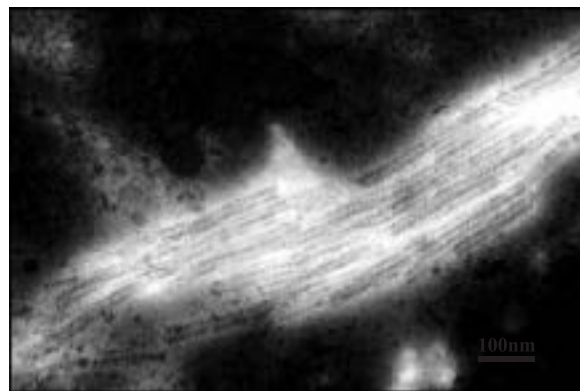


Fig. 1 Electron micrograph of Lettuce big-vein associated virus. Transmission electron micrograph of a negative-stained purified virus preparation. The bar represents 100 nm.

Tobacco with stunt symptoms invariably has a varicosavirus, and mechanical inoculation of tobacco seedlings with partially purified virus-preparation leads to typical tobacco stunting; Tobacco stunt virus (TSStV) is regarded as the causal agent of tobacco stunt disease. Because of marked differences in host reactions, LBVaV and TSStV were considered to be distinct viruses. However, a comparison of partial sequences of LBVaV and TSStV suggests now that TSStV is a strain of LBVaV.

Taxonomy and Classification

The floating genus *Varicosavirus* was first established by the International Committee on the Taxonomy of Viruses in 2000. Even though LBVaV is a two-segmented negative-sense RNA virus with a fragile non-enveloped rod shape, LBVaV shares several features with plant rhabdoviruses, which generally have linear non-segmented negative-sense RNA genomes and large enveloped particles with a prominent fringe of peplomers. On the base of genome structures and amino acid sequence comparisons between LBVaV and other plant rhabdoviruses, the taxonomic position was re-evaluated by the current committee in 2018, and the genus *Varicosavirus* is now classified in the family *Rhabdoviridae*. *Lettuce big-vein associated varicosavirus* is the type species of this genus, and LBVaV and three tentative viruses, have been tentatively placed into the genus; *Alopecurus myosuroides varicosavirus* 1, Tobacco stunt virus, and Red clover associated varicosavirus.

LBVaV particles are fragile non-enveloped rods about 18 nm in diameter, with modal lengths of 320–360 nm. Virus particles have a central canal about 3 nm in diameter and an obvious helix with a pitch of about 5 nm, appearing similar to the inner striated nucleocapsid core of rhabdoviruses. Virus particles are very unstable *in vitro*, so the helix of particles, especially those in purified preparations, tends to loosen and become partially uncoiled, even in preparations fixed with glutaraldehyde.

LBVaV was initially believed to have a divided genome consisting of two segments of double-stranded RNA, approximately 7.0 and 6.5 kbp. Once LBVaV purification was established and the complete nucleotide sequence of LBVaV determined, a precise reinvestigation of LBVaV genome components showed that LBVaV is not a double-stranded RNA virus but a single-stranded RNA virus with a bipartite genome. Negative-sense and positive-sense RNAs are separately encapsidated in the virions and the negative-sense RNA is predominant.

Criteria for membership in the genus are non-enveloped rod-shaped particles, a two-segmented negative-sense RNA genome, and transmission through moist soil by motile zoospores of the obligate parasitic soil-borne fungus, *Olpidium virulentus*.

Genetics and Evolution

Nucleotide sequence and genome mapping studies showed that the LBVaV genome consists of 12878 nucleotides, divided into two segments (Fig. 2). The larger segment (RNA1) contains antisense information for a small hypothetical protein and a large gene that encodes the large protein (L) of 2040 amino acids. The smaller LBVaV genome segment (RNA2) contains antisense information for five major genes, which have coding capacities for 397, 333, 290, 164, and 368 amino acids, respectively.

L Protein Encoded on RNA1

As with other negative-sense RNA viruses, the L protein of LBVaV is positively charged and contains RNA-directed RNA polymerase (RdRp) and RNA binding domains. Alignment of the L protein sequence with those of several other negative-sense RNA viruses shows the conservation of functional RdRp motifs (Fig. 2). The L protein of LBVaV contains conserved premotif A, which presumably plays an important role in RNA template binding and positioning, and four motifs that correspond to motifs A, B, C, and D, thought to comprise the palm and finger regions of the polymerase active site in negative-sense RNA viruses. A putative ATP binding motif, Kx₁₇GxGxG, proposed to be associated with polyadenylation or protein kinase activity, is present at position 1643–1665. LBVaV has been experimentally shown to transcribe 3'-polyadenylated mRNAs.

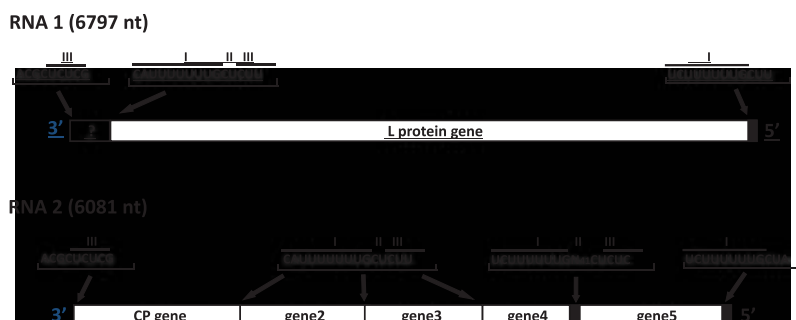


Fig. 2 Schematic representation of the genome organization of a negative-sense arrangement of genes encoded in the genome of Lettuce big-vein associated virus. Gene-junction regions, including the gene-end sequence (I), the intergenic sequence (II), and the gene-start sequence (III), are shown above negative-sense genomic RNAs.

		Premotif A			Motif A
LBVaV	518	K <u>ERE</u> IKVAARMYSLMTERMRYYFVLTEGL	<39>	NINIDFSKWNNTNMR	<54>
LNyV	490	K <u>ERE</u> MNPTPRMFALMSHLMRVYVVITESM	<48>	CMSLDFEKWNGHMR	<55>
NCMV	544	K <u>ERE</u> MNPVARMFALMTLKMRSYVVITENM	<42>	CINMDFEKWNLNMR	<56>
SYNV	567	K <u>ERE</u> MKTKARFFSLMSYKLRLMYVTSTEEL	<41>	SMNIDFSKWNQNMR	<54>
RYSV	537	K <u>ERE</u> LKIMARFFALLSFKMRLYFTATEEL	<41>	VINMDFVKWNQQMR	<55>
VSIV	530	K <u>ERE</u> LKLAGRFFSLMSWKLREYFVITEYL	<42>	ANHIDYEKWNNHQR	<56>
RABV	543	K <u>ERE</u> LKIEGRFFALMSWNLRLYFVITEKL	<42>	AFHLDYEKWNNHQR	<58>
BEFV	556	K <u>ERE</u> LKEEGRFFSLMSYELRDYFVSTEYL	<42>	ANNIDYEKWNNYQR	<56>
IHNv	494	K <u>EME</u> LKIKGRGFLMTFMPRLQLVRESI	<40>	NKSLDINKFCTSQR	<67>
TSWV	1286	KMQRTKT-DREIYLMMSKVKMMLYFIEHT	<49>	FLSADQSKWSASGL	<69>
RSV	1416	KNQHGG--LREIYVLNIFERIMQKTVEF	<41>	STSDDASKWNQGHY	<75>

		Motif B			Motif C			Motif D
LBVaV		MCYRGHLGG <u>F</u> ELRQKGWTVATVCLL	<12>	LMGQGD <u>N</u> QII	<64>	LDGRQLPQWYKKT		
LNyV		YSFTGHKGG <u>Q</u> ELRQKGWTIFTVVCL	<12>	IMGMGD <u>N</u> QVL	<64>	LKGVPLSMDLKKI		
NCMV		KSYEGHIRG <u>F</u> ELRQKGWTVFTVLI	<12>	LMGQGD <u>N</u> QVL	<64>	YKGVPLCSSLKRI		
SYNV		WSRTGDES <u>G</u> ELRQKGWTITTVCDI	<12>	LIGGGD <u>N</u> QVL	<63>	YSGVPLRGRLKVI		
RYSV		VCWIDDGAG <u>K</u> ELRQKAWTIMTVCDI	<12>	LVGGGD <u>N</u> QVL	<64>	YKGVPLRSPLKQV		
VSIV		VCWQGGEG <u>L</u> ELRQKGWTILNLLVI	<12>	VLAQGD <u>N</u> QVI	<63>	FRGVIRGLETKRW		
RABV		TCWNGQDG <u>L</u> ELRQKGWSLVSLMI	<12>	VLAQGD <u>N</u> QVL	<63>	FRGNILVPESKRW		
BEFV		VCWEGQKG <u>L</u> ELRQKGWSILNYLMI	<12>	ILAQGD <u>N</u> QTI	<63>	IEGTIKGLPTKRW		
IHNv		GVFSGLKGG <u>I</u> ELCQYVWTICLLLRV	<12>	ILAQGD <u>N</u> VII	<62>	HCPQHLLTLAIKKA		
TSWV		NTYPVSMNLQGNLNYLSVYHSCAM	<18>	WIVHSD <u>D</u> NAT	<32>	KSFCITLNPCKSY		
RSV		SYIETETGMMQGILHYTSSLFHAIFL	<31>	NMESSD <u>D</u> SSF	<31>	GYTLGIYKSPKST		

Fig. 3 Comparison of conserved motifs of the L protein among LBVaV, eight rhabdoviruses, and two plant segmented negative-sense RNA viruses. The amino acid sequence of the L protein of LBVaV is shown at the top followed by amino acid sequences of the cytorhabdoviruses Lettuce necrotic yellows virus (LNyV) and Northern cereal mosaic virus (NCMV), the nucleorhabdoviruses Sonchus yellow net virus (SYNV) and Rice yellow stunt virus (RYSV), and the vertebrate rhabdoviruses Vesicular stomatitis Indiana virus (VSIV)(now Indiana vesiculovirus), Rabies virus (RABV)(now Rabies lyssavirus), Bovine ephemeral fever virus (BEFV)(now Bovine fever ephemerovirus), Infectious hematopoietic necrosis virus (IHNv)(now Salmonid novirhabdovirus), the orthotospovirus Tomato spotted wilt virus (TSWV) and the tenuivirus Rice stripe tenuivirus (RSV). Numbers at the beginning of lines indicate the position of the first displayed amino acid. Numbers within brackets indicate the numbers of amino acids not represented in the figure. Conserved residues recognized previously for L proteins of negative-sense RNA viruses are shown in bold letters and residues characteristic of the division between nonsegmented and segmented negative-sense RNA viruses are underlined.

The L protein is the only region sufficiently conserved to enable evolutionary relationships to be established among negative-sense RNA viruses. Even though LBVaV is segmented, most of the conserved motifs previously identified in the L proteins of non-segmented negative-sense RNA viruses are conserved in the L protein of LBVaV (Fig. 3), particularly the conserved sequence GDN in motif C, which likely constitutes the evolutionary and functional equivalent of the sequence GDD in polymerases of positive-sense RNA viruses. This is in striking contrast to segmented negative-sense RNA viruses whose polymerase has the conserved sequence SDD. Amino acids G at position 662 and W at position 671 in motif B, specific to the nonsegmented viruses, are also conserved in the L protein of LBVaV. Tetrapeptide E(F/Y)xS, located downstream from motif D, which is specifically conserved in polymerases of segmented negative-sense RNA viruses, is not found in the L protein of LBVaV.

Phylogenetic analysis of the partial L proteins between LBVaV and rhabdoviruses shows that the L protein of LBVaV clusters within the family *Rhabdoviridae*, and that the L protein of LBVaV is more distantly related to those of plant rhabdoviruses than they are to each other, but is more closely related to those of plant rhabdoviruses than those of vertebrate rhabdoviruses (Fig. 4).

Proteins Encoded on RNA2

The first gene at the 3' end of genomic RNA encodes the coat protein (CP), which shows low but significant similarity to the nucleocapsid (N) proteins of plant rhabdoviruses. Most other genes encode hypothetical proteins with unknown functions, while the third encoded protein has several properties related to a viral cell-to-cell movement protein (MP).

Gene Bank database homology searches using the amino acid sequence of the CP of LBVaV consistently retrieve N proteins of rhabdoviruses. Phylogenetic analysis between the CP of LBVaV and N proteins of rhabdoviruses shows that the CP clusters within the family *Rhabdoviridae*, and that the CP is most closely related to the N protein of a fish rhabdovirus, infectious hematopoietic necrosis virus (IHNv), and to a lesser extent to that of plant rhabdoviruses. In a PSI-BLAST search against the ProDom protein domain database, the carboxy-terminal part of the CP shows similarity to two domains, one is the homologous domain among N proteins of plant rhabdoviruses and another is the homologous domain among those of fish rhabdoviruses.

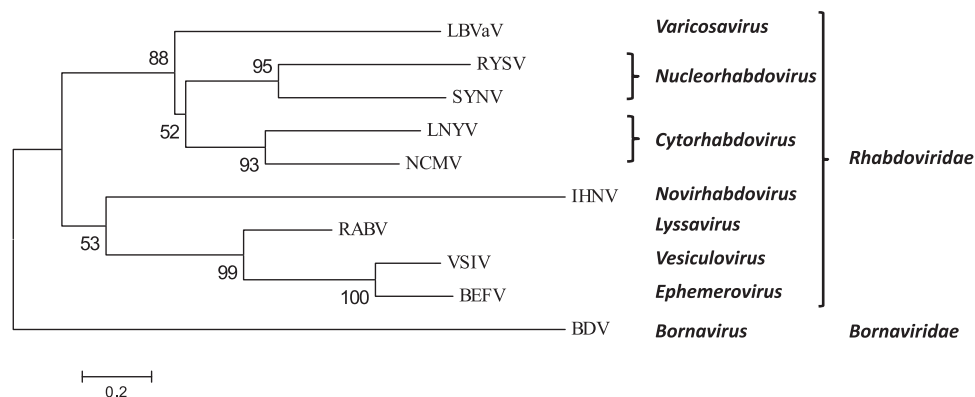


Fig. 4 Computer-generated tree illustrating phylogenetic relationships between LBVaV and selected rhabdoviruses derived from aligned amino acid sequences of the partial L protein between premotif A and motif D. The tree was generated using the neighbor-joining method of the MEGA6 software. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown. The evolutionary distances were computed using the Poisson correction method. Borna disease virus (BDV)(now Mammalian 1 orthobornavirus) is used as an out-group. See Fig. 3 for abbreviations of virus names.

The protein 2 of LBVaV is coded in the second gene that corresponds in position to the gene coding for the phosphoprotein (P) of rhabdoviruses, which interacts with the N protein and the L protein and constitutes the active polymerase complex. The predicted overall structure of the protein 2 and P proteins, assessed by amino acid composition and by relative hydropathicity distribution, is similar. Coiled-coil regions of P proteins of rhabdoviruses have been used to predict the regions for their oligomerization and interaction. Analysis using the COILS2 program predicts that the protein 2 contains three possible coiled-coil regions at the amino-terminal (aa 29-48), central (aa 122-135), and carboxy-terminal (aa 239-253) domains that are similar in position, to those in the P protein of Sonchus yellow net virus (SYNV). A detailed analysis of phosphorylation sites for cellular casein kinase II in the P protein of vesicular stomatitis Indiana virus (VSIV) confirmed five of ten potential phosphorylation sites of the P protein, which may regulate transcription and replication, are required to form the active polymerase complex. The protein 2 of LBVaV also has ten potential phosphorylation sites, and the five essential phosphorylation sites for the P protein of VSIV are conserved in positions similar to those of LBVaV. A predominantly basic 20-amino-acid-long region near the carboxy-terminus of the P protein of VSIV, which may serve as an attachment domain for combination with the L protein, also occurs in the carboxy-terminal region of the protein 2 of LBVaV.

The protein 3 of LBVaV is encoded in the third gene that corresponds in position to the gene coding for the MP of plant rhabdoviruses, which is essential for the plant viruses to assist in cell-to-cell movement through the host plant plasmodesmata. The protein 3 had a predicted structure similar to the 30K superfamily MP consensus structure and consisted of a series of β -elements flanked by an α -helix on each end. Our unpublished data of a transient expression using the GFP-fused P3 protein in epidermal cells of *Nicotiana benthamiana* showed the localization of the P3 protein to the plasmodesmata. The movement hypothesis for the protein 3 has been reinforced by our *trans*-complementation experiment with movement-defective tomato mosaic virus in *N. benthamiana* leaves: the protein 3 facilitated the spread of the movement-defective mutant from initially infected cells through two or more cell layers.

The protein 5 of LBVaV, which corresponds in position to the gene coding for the glycoprotein (G) of plant rhabdoviruses, has also little direct relatedness to the envelope-associated glycoproteins of rhabdoviruses. The protein 5 sequence does not contain any canonical sites for N-linked glycosylation, and does not contain the N-terminal signal peptide or carboxy-terminal hydrophobic transmembrane anchor domains, which are common among G proteins of rhabdoviruses. Conservation of cysteine residues is another common feature of G proteins of rhabdoviruses. Alignment using the Clustal W program for G proteins with plant rhabdoviruses shows that eleven cysteine residues are situated at conserved positions. In the case of LBVaV, six of ten cysteine residues are located at similar conserved positions in G proteins of plant rhabdoviruses. The protein 5 of LBVaV may thus result from the degeneration of G proteins of plant rhabdoviruses. The six cysteine residues in the protein 5 of LBVaV that may play an important role in activities, such as forming the secondary structure of G proteins of plant rhabdoviruses appear to have been stringently conserved among the protein 5 of LBVaV and G proteins of plant rhabdoviruses. The other five cysteine residues, however, appear to have evolved extensively to accommodate diverse host requirements and vector adaptations.

Transcription Termination/Initiation Strategy

LBVaV transcribes capped and polyadenylated mono-cistronic mRNAs. In the gene-junction regions of LBVaV, transcription termination/polyadenylation and initiation signal sequences comparable to those of rhabdoviruses are recognized (Fig. 5), suggesting that LBVaV and rhabdoviruses may use a similar mechanism of transcription, a stop-and-start model, to differentially express individual genes from a contiguous virus genome.

		I	II	III
LBVaV	TypeA	UNCAUUUUUUU	G (N) _x	CUCU
	TypeB	AAUCUUUUUU		
LNyV		AUUCUUUUU	G (N) _x	CUCU
NCMV		AUUCUUUUU	GA	CUCU
SYNV		AUUCUUUUU	GG	UUGW
RYSV		AUUAUUUUU	GGG	UUGU
VSIV		AUACUUUUUUU	GA	UUGU
RABV		GYACUUUUUUU	G (N) _x	UUGU
IHNV		URUCUUUUUUU	G/A	CCRW

Fig. 5 Comparison of the gene-junction regions between LBVaV and selected rhabdoviruses. Gene-junction regions are separated into three sequences: I is the gene-end sequence that constitutes the poly(U) tract at the 3' end of each gene on negative-sense genomic RNAs and plays the role of a transcription termination/polyadenylation signal; II is the short intergenic sequence that is not transcribed during mRNA synthesis; and III is the gene-start sequence that constitutes the initiation site for transcription of each mRNA. Bold type in viral sequences indicates consensus nucleotides, R and W indicate A/G and A/U, and (N)_x corresponds to a variable number of nucleotides. See Fig. 3 for abbreviations of virus names.

The gene-junction regions of LBVaV are grouped into three elements: a gene-end sequence (I) at the 3' end of each gene on the genomic template, which is required for transcription termination/polyadenylation; a short non-transcribed intergenic sequence (II), which is not represented in mRNAs and separates each gene; and a gene-start sequence (III) located at the beginning of each subsequent gene, which plays an important role in transcription initiation. Gene-end sequences of rhabdoviruses consist of an A-U rich region, a cytidylate, and a poly(U) tract. The poly(U) tract is thought to be a template for the poly(A) tail of mRNA due to reiterative transcription or slippage. The gene-end sequences of LBVaV are either 3'-UNCAUUUUUUU-5' (Type A) or 3'-AAUCUUUUUUU-5' (Type B), reminiscent of those observed in other rhabdoviruses. When the gene-end sequence of LBVaV is Type A, transcription of the next gene is initiated immediately after a single nucleotide G, while for Type B, the two genes (genes 5 and L) are located at the 5'-most ends of each RNA (negative-sense) or transcription is reinitiated after a long 42 nt intergenic sequence located between genes 4 and 5, suggesting that the two variations of the gene-end sequences may influence signaling initiation of downstream mRNA synthesis as reported for VSIV in which the gene-end sequence plays an important role both in transcription termination/polyadenylation and in reinitiating transcription of downstream genes.

In contrast to other rhabdoviruses, including nucleorhabdoviruses whose gene-start sequence in the gene-junction regions is 3'-UUUGU-5', the conserved gene-start sequence of LBVaV is 3'-CUCU-5'. Sequence 3'-CUCU-5' is also conserved in the gene-junction regions of the cytorhabdoviruses lettuce necrotic yellows virus and northern cereal mosaic virus, suggesting that sequence 3'-CUCU-5' may be important in acting as a transcription initiation signal for both LBVaV and these cytorhabdoviruses, which replicate in the cytoplasm of plant cells.

Host Range and Geographic Distribution

Fungal vector transmission experiments in the greenhouse have shown that LBVaV systemically infects seven plant species in four families, and the infected plants including lettuce do not show any symptoms. However, the virus induced necrotic rings and spots in lettuce in hydroponic culture.

In contrast, the experimental host range of TSIV encompasses 41 species in nine dicotyledonous families by mechanical inoculation and 35 plant species in 13 families through fungal vector transmission. The virus does not infect lettuce, but induces vein clearing, vein necrosis, mottling, and stunting in many *Nicotiana* species.

LBVaV is widespread in cool to temperate regions and several parts of subtropical regions in Europe, Asia, Australia, and North and South America, but TSIV has only been reported in Japan.

Transmission and Vectors

LBVaV is transmitted through motile zoospores of an obligate parasitic soil-inhabiting fungus. LBVaV is very stable in soil within long-lived resting fungal spores for 20 years or more, but it is remains unclear whether the virus replicates inside fungal spores.

In 1878, the natural vector of LBVaV was reported to be *Olpidium brassicae* (Wor.) Dang. *O. brassicae* was first found in a cabbage (*Brassica oleracea*) root and the specific name of the fungus is derived from the genus name of the cabbage. Several strains of *O. brassicae* that differ significantly in host specificity, morphology, and virus transmissibility have been recognized, particularly a crucifer strain, which readily multiplies in roots of crucifer plants such as cabbage and mustard, and a noncrucifer strain that can not multiply in roots of crucifer plants. The crucifer strain is host-specific and the non-crucifer strain is highly polyxenous. Morphologically, zoosporangia and exit tubes of the crucifer strain differ slightly from those of the noncrucifer strain. The non-crucifer strain is a vector of the soil-borne viruses MiLV and tobacco necrosis virus, but the crucifer strain fails to transmit these viruses. Furthermore, the crucifer strain shows sexual reproductivity, although the noncrucifer strain does not reproduce sexually. Zoospores from a single-sporangium isolate of the crucifer strain do not develop a resting spore, but after mating between different mating (sexual) types, a resting spore form. In contrast, zoospores from the noncrucifer strain develop a resting spore without mating. Based on these differences, the crucifer and noncrucifer strains have been divided into two distinct species, but this

proposal has long been neglected. Recent rDNA-ITS analysis of *O. brassicae* strongly supported the hypothesis that crucifer and non-crucifer strains belong to separate species rather than being strains of the same *Olpidium* species, and a new species name, *O. virulentus*, has been proposed for the non-crucifer strain. The non-crucifer strain was confirmed to be a vector of LBVaV, although there is no reliable evidence that the crucifer strain (*O. brassicae*) is capable of transmitting LBVaV.

Further Reading

- Banerjee, A.K., Barik, S., De, B.P., 1991. Gene expression of non-segmented negative strand RNA viruses. *Pharmacology & Therapeutics* 51, 47.
- Jackson, A.O., Dietzgen, R.G., Goodin, M.M., Bragg, J.N., Deng, M., 2005. Biology of plant rhabdoviruses. *Annual Review of Phytopathology* 43, 623.
- Lot, H., Campbell, R.N., Souche, S., Milne, R.G., Roggero, P., 2002. Transmission by *Olpidium brassicae* of Mirafiori lettuce virus and Lettuce big-vein virus, and their roles in lettuce big-vein etiology. *Phytopathology* 92, 288.
- Maccarone, L.D., 2013. Relationships between the pathogen *Olpidium virulentus* and viruses associated with lettuce big-vein disease. *Plant Disease* 97, 700–707.
- Poch, O., Sauvaget, I., Delarue, M., Tordo, N., 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *The EMBO Journal* 8, 3867.
- Rose, J.K., Whitt, M.A., 2001. *Rhabdoviridae*. The viruses and their replication. In: Knipe, M., Howley, P.M. (Eds.), *Field Virology*, fourth ed. New York: Raven Press, p. 1221.
- Sasaya, T., Ishikawa, K., Koganezawa, H., 2002. The nucleotide sequence of RNA1 of Lettuce big-vein virus, genus *Varicosavirus*, reveals its relation to nonsegmented negative-strand RNA viruses. *Virology* 297, 289.
- Sasaya, T., Koganezawa, H., 2006. Molecular analysis and virus transmission tests place *Olpidium virulentus*, a vector of Mirafiori lettuce big-vein virus and Tobacco stunt virus, as a distinct species rather than a strain of *Olpidium brassicae*. *Journal of General Plant Pathology* 72, 20.
- Sasaya, T., Kusaba, S., Ishikawa, K., Koganezawa, H., 2004. Nucleotide sequence of RNA2 of *Lettuce big-vein virus* and evidence for a possible transcription termination/initiation strategy similar to that of rhabdoviruses. *Journal of General Virology* 85, 2709.
- Verbeek, M., Dulleman, A.M., van Bekkum, P.J., van der Vlugt, R.A.A., 2013. Evidence for Lettuce big-vein associated virus as the causal agent of a syndrome of necrotic rings and spots in lettuce. *Plant Pathology* 62, 444–451.

Virgaviruses (*Virgaviridae*)

Eugene I Savenkov, Swedish University of Agricultural Sciences, Uppsala, Sweden and Linnean Center for Plant Biology, Uppsala, Sweden

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa Amino acid(s)
CP Coat protein or capsid protein
CRP Cysteine-rich protein
diRNA Defective interfering RNA
EM Electron microscopy
ER Endoplasmic reticulum
Hel Helicase
IPM Integrated pest management
kb Kilobase
kDa Kilo Dalton
MET Methyl-transferase

MP Movement protein
NGS Next generation sequencing
nt Nucleotide(s)
ORF Open reading frame
PCR Polymerase chain reaction
RdRp RNA-dependent RNA polymerase
RTD Read-through domain
sgRNA Sub-genomic RNA
ssRNA single-stranded RNA
TGB Triple gene block
UTR Untranslated region
VSR Viral suppressor of RNA silencing

Glossary

126-kDa protein The first protein expressed during tobamovirus infection, it has methyl-transferase and helicase domains, but lacks a polymerase domain. Read-through of the stop-codon of the open reading frame (ORF) encoding 126 kDa protein results in translation of the larger the 183 kDa protein, which has a polymerase domain.

Alpha-like viruses A group of viruses in the proposed classification of positive-stranded RNA viruses based on the sequence similarities and arrangement of domains of RNA-dependent RNA polymerase. Three groups have been recognized: the picorna-like group, the flavi-like group and the alpha-like group. Alpha-like group is named after the genus *Alphavirus* of the family *Togaviridae*.

Cross-protection A phenomenon whereby prior infection by one virus strain (the primary strain) can prevent or limit multiplication of another virus strain (the secondary strain) or a closely related virus. For this purpose, protecting strains (the primary strains) should cause attenuated symptoms and should reduce marketable yields only slightly. A few attenuated virus strains have been commercially used as a “vaccine”.

Cross-protection has been applied in the fields to control poty-, tobamo- and closteroviruses. The phenomenon was first reported with Tobacco mosaic virus (TMV) in 1929.

Defective interfering RNAs Subviral agents produced during replication of RNA viruses. Defective interfering (di) RNAs are parasitic RNAs because they take advantage of the parent virus-coded protein for their multiplication. These RNAs are called “defective” because they have lost the

capacity to code for the proteins needed for independent replication and thus are defective in the absence of the helper (parent) virus. diRNAs are referred as “interfering” because they can attenuate the symptoms caused by the helper virus, they multiply rapidly and eventually lead to slowing down of the helper virus multiplication.

Lateral flow detection A chromatography method based on lateral flow using specific antibodies applied on strips. Once the strip is submerged with the “sample” side into the sample extract, the liquid migrates upwards and initiates the antigen-antibody reaction which results in visible lines. Both test and control lines become visible with positive extracts, whereas negative samples produce the upper control line only. Lines start developing after 1–2 mn and reach maximum intensity after 10–15 mn. Dried test strips can be kept as permanent records.

Leaky scanning A mechanism used during the initiation steps of eukaryotic translation in which a “weak” initiation codon on mRNA is sometimes skipped by ribosome. During translation initiation, the small 40S ribosomal subunit “scans” or moves in a 5′ → 3′ direction along the 5′ untranslated region to locate a start codon to commence elongation. Sometimes, the scanning ribosome may encounter an “unfavorable nucleotide context” around the start codon, bypasses this start codon and begins translation at further downstream AUG start codons. This way an mRNA can encode several different proteins if the AUG are not in frame or code for proteins with different N-termini if the AUG start codons are in the same frame.

Introduction

The family *Virgaviridae* contains seven genera (*Furovirus*, *Goravirus*, *Hordeivirus*, *Pecluvirus*, *Pomovirus*, *Tobamovirus*, and *Tobravirus*) and twenty-seven species unassigned to a genus (Table 1). Virgaviruses possess positive-sense single-stranded RNA genomes encapsidated in rod-shaped particles. Virions are helically constructed from one type of coat protein (CP) with exception in members of the *Furovirus* and *Pomovirus* genera, which incorporate an additional larger minor CP into one extremity of the virus

Table 1 List of members in the genera of the family *Virgaviridae*. Type species are in bold

<i>Genus/Species</i>	<i>Acronym</i>	<i># Segments</i>	<i>Access. #</i>	<i>Length (nt)</i>
Furovirus				
<i>Chinese wheat mosaic virus</i>	CWMV	2	NC002359 NC002356	7147 3569
<i>Japanese soil-borne wheat mosaic virus</i>	JSBWMV	2	NC038850 NC038851	7226 3574
<i>Oat golden stripe virus</i>	OGSV	2	NC002357 NC002358	3232 7111
<i>Soil-borne cereal mosaic virus</i>	SBCMV	2	NC002351 NC002330	7025 3683
Soil-borne wheat mosaic virus	SBWMV	2	NC002041 NC002042	7099 3593
<i>Sorghum chlorotic spot virus</i>	SCSV	2	NC004015 NC004014	3418 6878
Unassigned Furoviruses				
<i>French barley mosaic virus</i>	FBMV	2	AJ749658 AJ749657	1507 ^a 3291 ^a
Goravirus				
<i>Drakaea virus A</i>	DraVA	2	NC043398 NC043399	4490 ^a 2905 ^a
Gentian ovary ringspot virus	GORSV	2	NC024501 NC024502	5519 3869
Hordeivirus				
<i>Anthoxanthum latent blanching virus</i>	ALBV	—	ND	
Barley stripe mosaic virus	BSMV	3	NC003469 NC003478 NC003481	3768 3164 3289
<i>Lychnis ringspot virus</i>	LRSV	3	NC038933 NC038932 ND	1468 ^a 3065 ^a —
<i>Poa semilatifolia virus</i>	PSLV	3	MK377386 MK377387 MK377388	3873 3608 3169
Pecluvirus				
<i>Indian peanut clump virus</i>	IPCV	2	NC004730 NC004729	4507 5841
Peanut clump virus	PCV	2	NC003668 NC003672	4504 5897

Table 1 Continued

Genus/Species	Acronym	# Segments	Access. #	Length (nt)
Pomovirus				
<i>Beet soil-borne virus</i>	BSBV	3	NC003520 NC003519 NC003518	5834 3005 3454
<i>Beet virus Q</i>	BVQ	3	NC003512 NC003511 NC003510	2529 2913 6003
<i>Broad bean necrosis virus</i>	BBNV	3	NC004424 NC004423 NC004425	2831 5600 2417
<i>Colombian potato soil-borne virus</i>	CoPSNV	3	NC029037 NC029035 NC029034	3028 3164 6170
Potato mop-top virus	PMTV	3	NC003723 NC003725 NC003724	6043 2964 3134
Unassigned Pomoviruses				
<i>Soil-borne virus 2</i>	SBV2	2	KT225277 KT225278	5961 2532
Tobamovirus				
<i>Bell pepper mottle virus</i>	BPMV	1	NC009642	6375
<i>Brugmansia mild mottle virus</i>	BrMMoV	1	NC010944	6381
<i>Cactus mild mottle virus</i>	CMMoV	1	NC011803	6449
<i>Clitoria yellow mottle virus</i>	CIYMoV	1	NC016519	6514
<i>Cucumber fruit mottle mosaic virus</i>	CGMoMV	1	NC002633	6562
<i>Cucumber green mottle mosaic virus</i>	CFMoMV	1	NC001801	6424
<i>Cucumber mottle virus</i>	CuMoV	1	NC008614	6485
<i>Frangipani mosaic virus</i>	FrMV	1	NC014546	6643
<i>Hibiscus latent Fort Pierce virus</i>	HLFPV	1	NC025381	6431
<i>Hibiscus latent Singapore virus</i>	HLSV	1	NC008310	6485
<i>Kyuri green mottle mosaic virus</i>	KGMoMV	1	NC003610	6514
<i>Maracuja mosaic virus</i>	MaMV	1	NC008716	6794
<i>Obuda pepper virus</i>	ObPV	1	NC003852	6507
<i>Odontoglossum ringspot virus</i>	ORSV	1	NC001728	6618
<i>Opuntia virus 2</i>	OV2	1	NC040685	6453
<i>Paprika mild mottle virus</i>	PaMMoV	1	NC004106	6524
<i>Passion fruit mosaic virus</i>	PfMV	1	NC015552	6791
<i>Pepper mild mottle virus</i>	PeMMoV	1	NC003630	6357
<i>Plumeria mosaic virus</i>	PIMV	1	NC026816	6688
<i>Rattail cactus necrosis-associated virus</i>	RCNaV	1	NC016442	6506
<i>Rehmannia mosaic virus</i>	ReMV	1	NC009041	6395
<i>Ribgrass mosaic virus</i>	RMV	1	NC002792	6311
<i>Streptocarpus flower break virus</i>	SFBV	1	NC008365	6279
<i>Sunn-hemp mosaic virus</i>	SHMV	1	ND	
<i>Tobacco latent virus</i>	TLV	1	NC038703	1415 ^a
<i>Tobacco mild green mosaic virus</i>	TMGMV	1	NC001556	6355
Tobacco mosaic virus	TMV	1	NC001367	6395
<i>Tomato brown rugose fruit virus</i>	ToBRFV	1	NC028478	6393
<i>Tomato mosaic virus</i>	ToMV	1	NC002692	6383
<i>Tomato mottle mosaic virus</i>	ToMoMV	1	NC022230	6398
<i>Tropical soda apple mosaic virus</i>	TSAMV	1	NC030229	6350

(Continued)

Table 1 Continued

<i>Genus/Species</i>	<i>Acronym</i>	<i># Segments</i>	<i>Access. #</i>	<i>Length (nt)</i>
<i>Turnip vein-clearing virus</i>	TuVCV	1	NC001873	6311
<i>Ullucus mild mottle virus</i>	UMMoV	1	ND	
<i>Wasabi mottle virus</i>	WMoV	1	NC003355	6298
<i>Yellow tailflower mild mottle virus</i>	YTMMoV	1	NC022801	6379
<i>Youcai mosaic virus</i>	YMoV	1	NC004422	6303
<i>Zucchini green mottle mosaic virus</i>	ZGMoV	1	NC003878	6513
Unassigned Tobamoviruses				
<i>Actinidia tobamovirus</i>	AcTV	1	ND	
<i>Abutilon yellow mosaic virus</i>	AbYMV	1	EU559678	5903
<i>Acidomyces richmondensis tobamovirus 1</i>	ArTV1	1	MK279511	10291
<i>Bottle gourd mottle virus</i>	BGMoV	1	MF960828	6497
<i>Brugmansia latent virus</i>	BrLV	1	NC010944	6381
<i>Crucifer tobamovirus</i>	CTV	1	AB003936	2378 ^a
<i>Hoya chlorotic spot virus</i>	HCSV	1	NC034509	6386
<i>Nigerian tobacco latent virus</i>	NTLV	1	ND	
<i>Penstemon ringspot virus</i>	PRSV	1	ND	
<i>Piper chlorosis virus</i>	PiCV	1	KX683424	5686 ^a
<i>Podosphaera prunicola tobamovirus</i>	PpTV	1	ND	
<i>Watermelon green mottle mosaic virus</i>	WGMoV	1	MH837097	6482
Tobravirus				
<i>Pea early-browning virus</i>	PEBV	2		
			NC001368	3374
			NC002036	7073
<i>Pepper ringspot virus</i>	PeRSV	2		
			NC003670	1799
			NC003669	6828
<i>Tobacco rattle virus</i>	TRV	2		
			NC003811	3855
			NC003805	6791
Unassigned Virgaviruses				
<i>Atrato virga-like virus 1</i>	AVLV1		MN661103	9006
<i>Atrato virga-like virus 2</i>	AVLV2		MN661104	8775
<i>Atrato virga-like virus 3</i>	AVLV3		MN661112	10309
<i>Atrato virga-like virus 4</i>	AVLV4		MN661113	9638
<i>Atrato virga-like virus 5</i>	AVLV5		MN661116	10862
<i>Atrato virga-like virus 6</i>	AVLV6		MN661118	10224
<i>Atrato virga-like virus 7</i>	AVLV7		MN661120	10266
<i>Bombus-associated virus 1</i>	BaV1		MH614308	1681 ^a
<i>Bombus-associated virus 2</i>	BaV2		MH614309	962 ^a
<i>Bombus-associated virus 3</i>	BaV3		MH614318	2265
<i>Bombus-associated virus 4</i>	BaV4		MH614319	1840 ^a
<i>Botryosphaeria dothidea tobamo-like virus</i>	BdTLV		MK189194	3687 ^a
<i>Cucumis melo virga-like virus</i>	CmVLV		MG195286	2463 ^a
<i>Culex virga-like virus</i>	CuVLV		MH188019	10917
<i>Hammariskog virga-like virus</i>	HVLV		MN513382	10200
<i>Macrophomina phaseolina tobamo-like virus</i>	MPTLV		NC025674	9524
<i>Macrophomina phaseolina tobamo-like virus 1a</i>	MpTLV1a		KP900897	7499
<i>Nephila clavipes virus 3</i>	NcV3		NC040613	8138
<i>Nephila clavipes virus 4</i>	NcV4		NC040614	11919
<i>Plasmopara viticola associated virga-like virus 1</i>	PvaVLV1		MN551100	6782
<i>Plasmopara viticola associated virga-like virus 2</i>	PvaVLV2		MN551103	3486
<i>Plasmopara viticola associated virga-like virus 3</i>	PvaVLV3		MN551104	3594
<i>Plasmopara viticola associated virga-like virus 5</i>	PvaVLV5		MN551118	3407
<i>Plasmopara viticola associated virga-like virus 6</i>	PvaVLV6		MN551125	3528
<i>Plasmopara viticola associated virga-like virus 7</i>	PvaVLV7		MN551126	3560
<i>Rosellinia necatrix virga-like-virus 1</i>	RnVLV		LC333735	877 ^a
<i>Soil-borne barley mosaic virus</i>	SBBMV		KU236377	3574

^aIncomplete genomic sequence.

Table 2 Classification of members of the family *Virgaviridae* into genera, with indication of type species, number of virus species, number of RNA genome components, type of movement protein, transmission vector and host type

Genus	Type species	Acronym	Number ^a	Genome	MP	3' structure	Transmission	Host type
<i>Furovirus</i>	<i>Soil-borne wheat mosaic virus</i>	SBWMV	6 + 1	Bipartite	'30K'	tRNA ^{Val}	<i>P. graminis</i>	Monocots
<i>Goravirus</i>	<i>Gentian ovary ringspot virus</i>	GORSV	2	Bipartite	TGB	tRNA [–]	Pollen	Dicots
<i>Hordeivirus</i>	<i>Barley stripe mosaic virus</i>	BSMV	4	Tripartite	TGB	tRNA ^{Tyr}	Pollen/seed	Monocots/dicots
<i>Pecluvirus</i>	<i>Peanut clump virus</i>	PCV	2	Bipartite	TGB	tRNA ^{Val}	<i>P. graminis</i>	Monocots/dicots
<i>Pomovirus</i>	<i>Potato mop-top virus</i>	PMTV	5 + 1	Tripartite	TGB	tRNA ^{Val/Leu}	Plasmodiophorids	Dicots
<i>Tobamovirus</i>	<i>Tobacco mosaic virus</i>	TMV	37 + 12	Monopartite	'30K'	tRNA ^{His}	Mechanical	Dicots
<i>Tobravirus</i>	<i>Tobacco rattle virus</i>	TRV	3	Bipartite	'30K'	tRNA [–]	Nematodes	Dicots

^aNumber of members according to the International Committee on Taxonomy of Viruses (ICTV) as of January 2020.

particles. This protein is produced through translational read-through of the stop codon of CP-encoding open reading frame (ORF) and is needed for acquisition and transmission of the viruses of these two genera by plasmodiophorid protist vectors. Many virgaviruses cause economically significant diseases in vegetable, fruit, cereal and fiber crops, and in ornamental plants worldwide.

The members of *Virgaviridae* infect both monocot and dicots species and the viruses belonging to the *Furovirus*, *Pecluvirus*, *Pomovirus* and *Tobravirus* genera are soil-borne and transmitted in soil by plasmodiophorid and nematode vectors. Most Tobamoviruses are transmitted mechanically and seems not to have natural vectors. Tobacco mosaic virus (TMV; genus *Tobamovirus*) was the first virus to be discovered (in 1886) and crystallized (in 1935). The virus accumulates in extremely high concentrations in infected plants, extremely stable and is easily transmitted from plant to plant just through the contact of an infected plant with a healthy plant, e.g., through contacts facilitated by wind or air flow. This particular property of the virus, its various strains, and related viruses is used in agriculture for cross-protection.

Taxonomy, Phylogeny and Evolution

Genus demarcation criteria in the family *Virgaviridae* are based on number of genome components (mono-, bi- or tripartite), genome organization (e.g., type of the encoded movement protein, MP), mode of transmission (pollen, seed, mechanical or by vector), type of vector (plasmodiophorid or nematode), and phylogenetic relationships. Currently, virgaviruses are classified into seven genera: *Furovirus*, *Goravirus*, *Hordeivirus*, *Pecluvirus*, *Pomovirus*, *Tobamovirus*, and *Tobravirus*, briefly summarized below (Table 2).

The genus *Furovirus* comprises six recognized species: *Chinese wheat mosaic virus* (CWMV), *Japanese soil-borne wheat mosaic virus* (JSBWMV), *Oat golden stripe virus* (OGSV), *Soil-borne cereal mosaic virus* (SBCMV), *Soil-borne wheat mosaic virus* (SBWMV), and *Sorghum chlorotic spot virus* (SCSV); and one unassigned member: *French barley mosaic virus* (FBMV). The members of the genus are all bipartite soil-borne viruses transmitted by the plasmodiophorid *Polymyxa graminis* to graminaceous plants (monocots).

The genus *Goravirus* comprises two recognized species, *Drakaea virus A* (DraVA) and *Gentian ovary ringspot virus* (GORSV). The members of the genus are bipartite pollen-transmitted viruses, which infect ornamental plants (an orchid and a clustered gentian).

The genus *Hordeivirus* comprises four recognized species, *Anthoxanthum latent blanching virus* (ALBV), *Barley stripe mosaic virus* (BSMV), *Lychnis ringspot virus* (LRSV), and *Poa semilatifolia virus* (PSLV). The members of the genus are tripartite pollen- and seed-transmitted viruses, which infect monocot and dicot plants.

The genus *Pecluvirus* includes two recognized species, *Indian peanut clump virus* (IPCV) and *Peanut clump virus* (PCV). The members of the genus are bipartite soil-borne viruses transmitted naturally by *Polymyxa graminis* or by seed (in groundnuts). The viruses are also mechanically transmissible and infect monocots and dicots.

The genus *Pomovirus* comprises five recognized species and one unassigned member: *Beet soil-borne virus* (BSBV), *Beet virus Q* (BVQ), *Broad bean necrosis virus* (BBNV), *Colombian potato soil-borne virus* (CPSbV), *Potato mop-top virus* (PMTV) and *Soil-borne virus 2* (a tentative member). The members of the genus are tripartite soil-borne viruses transmitted by the plasmodiophorids to dicot plants.

The genus *Tobamovirus*, with 37 accepted species and 12 unassigned species, is the largest genus in the family *Virgaviridae*. Tobamoviruses are the only members of the family to have a non-segmented (monopartite) genome, are not vector-transmissible and when seed transmitted, the embryo is not affected. Tobamoviruses are efficiently transmitted mechanically and include TMV, the first virus to be discovered and crystallized, thus, extensively studied.

The genus *Tobravirus* comprises three recognized species, *Pea early-browning virus* (PEBV), *Pepper ringspot virus* (PRSV), and *Tobacco rattle virus* (TRV). The members of the genus are bipartite soil-borne viruses transmitted by nematodes to dicot plants.

Additionally, twenty-seven unassigned species in the family *Virgaviridae* have been recognized and include only two plant-infecting viruses: *Cucumis melo virga-like virus* and *Soil-borne barley mosaic virus*. The remaining twenty-five species were identified in metagenomics libraries and include viruses found in fungi, oomycetes, insects (mosquitoes and bumble bees), and spiders (Table 1, Table 3).

Phylogenetic analysis of the sequences of replication protein of plant-infecting *Virgaviridae* shows that virgaviruses group in clusters corresponding to the seven existing genera (Fig. 1). Similar clustering is obtained when phylogenetic analysis is performed for the sequences of movement proteins, although '30K'-type and TGB1-type MPs are analyzed separately. There are also close relationships between the cysteine-rich proteins (CRPs) encoded by the viruses of the *Furovirus*, *Hordeivirus*, *Pecluvirus*, and

Table 3 Unassigned virgaviruses, with indication of virus species, number of genes, host and country of origin

Species	Acronym	Number of genes	Host	Host species	Country
Atrato virga-like virus 1	AVLV1	2 (PolyPr ^a , CP)	mosquito	<i>Culex</i> sp.	Colombia
Atrato virga-like virus 2	AVLV2	2 (PolyPr, CP)	mosquito	<i>Coquillettidia venezuelensis</i>	Colombia
Atrato virga-like virus 3	AVLV3	2 (PolyPr, CP)	mosquito	<i>Mansonia titillans</i>	Colombia
Atrato virga-like virus 4	AVLV4	2 (PolyPr, CP)	mosquito	<i>Psorophora albipes</i>	Colombia
Atrato virga-like virus 5	AVLV5	4 (PolyPr, CPs, UL36 ^b)	mosquito	<i>Mansonia titillans</i>	Colombia
Atrato virga-like virus 6	AVLV6	4 (PolyPr, HPs ^c)	mosquito	<i>Psorophora</i> sp.	Colombia
Atrato virga-like virus 7	AVLV7	4 (PolyPr, HPs)	mosquito	<i>Anopheles darlingi</i>	Colombia
Bombus-associated virus 1	BaV1	?	bumble bee	<i>Bombus</i> sp.	UK: Scotland
Bombus-associated virus 2	BaV2	?	bumble bee	<i>Bombus</i> sp.	UK: Scotland
Bombus-associated virus 3	BaV3	3 (?)	bumble bee	<i>Bombus</i> sp.	UK: Scotland
Bombus-associated virus 4	BaV4	?	bumble bee	<i>Bombus</i> sp.	UK: Scotland
Botryosphaeria dothidea tobamo-like virus	BdTLV	?	fungus	<i>Botryosphaeria dothidea</i>	China
Culex virga-like virus	CuVLV	3 (Rep ^d , HPs)	mosquito	<i>Culex</i> sp.	USA
Hammariskog virga-like virus	HVLV	4 (HPs)	mosquito	<i>Culex pipiens</i>	Sweden
Macrophomina phaseolina tobamo-like virus	MPTLV	4 (Rep, MP, CP)	fungus	<i>Macrophomina phaseolina</i>	USA
Macrophomina phaseolina tobamo-like virus 1a	MpTLV1a	4 (Rep, MP, CP)	fungus	<i>Macrophomina phaseolina</i>	USA
Nephila clavipes virus 3	NcV3	2 (Rep, gp2 ^e)	spider	<i>Nephila clavipes</i>	USA
Nephila clavipes virus 4	NcV4	3 (Rep, gp1, gp2)	spider	<i>Nephila clavipes</i>	USA
Plasmopara viticola associated virga-like virus 1	PvaVLV1	1 (Rep)	oomycete	<i>Plasmopara viticola</i>	Spain
Plasmopara viticola associated virga-like virus 2	PvaVLV2	3 (Rep, HPs)	oomycete	<i>Plasmopara viticola</i>	Italy
Plasmopara viticola associated virga-like virus 3	PvaVLV3	2 (HP, Rep)	oomycete	<i>Plasmopara viticola</i>	Italy
Plasmopara viticola associated virga-like virus 5	PvaVLV5	1 (Rep)	oomycete	<i>Plasmopara viticola</i>	Spain
Plasmopara viticola associated virga-like virus 6	PvaVLV6	1 (Rep)	oomycete	<i>Plasmopara viticola</i>	Italy
Plasmopara viticola associated virga-like virus 7	PvaVLV7	2 (Rep, HP)	oomycete	<i>Plasmopara viticola</i>	Italy
Rosellinia necatrix virga-like-virus 1	RnVLV	?	fungus	<i>Rosellinia necatrix</i>	Spain

^aPolyPr, polyprotein.^bUL36, Large tegument protein UL36.^cHP, hypothetical protein.^dRep, replicase.^egp, glycoprotein.

Tobravirus genera, although CRP of a pomovirus, Potato mop-top virus (PMTV), do not align with them and has an entirely different sequence.

In addition, within genera, closely-related tobamoviruses are grouped according to host they infect. Among 37 tobamovirus species, there are clearly groupings of closely-related viruses infecting similar hosts, namely those infecting cruciferous plants (*Ribgrass mosaic virus*, *Turnip vein-clearing virus*, *Wasabi mottle virus*, *Youcai mosaic virus*), cucurbits (*Cucumber fruit mottle mosaic virus*, *Cucumber green mottle mosaic virus*, *Cucumber mottle virus*, *Kyuri green mottle mosaic virus*, *Zucchini green mottle mosaic virus*), and solanaceous species (*Brugmansia mild mottle virus*, *Obuda pepper virus*, *Paprika mild mottle virus*, *Pepper mild mottle virus*, *Rehmannia mosaic virus*, *Tobacco mild green mosaic virus*, TMV, *Tomato brown rugose fruit virus*, *Tomato mosaic virus*, *Tropical soda apple mosaic virus*, *Yellow tailflower mild mottle virus*). Six other genera in the family Virgaviridae do not have enough diversity to distinguish particular sub-groupings because of the limited number virus species (two to six species depending on the genus) within the genus and very narrow host range (e.g., furoviruses, which infect graminaceous plants).

Genetic diversity and evolution of virgaviruses seem to be driven by mutation, recombination and reassortment of genome components (for multipartite viruses) that significantly contribute to the emergence of new viral variants, increasing their potential of adaptation to different crop varieties, hosts and environmental conditions. These is nicely illustrated by the analysis of numerous isolates of PMTV. Phylogenetic analysis of available PMTV sequences for RNA-rep, RNA-TGB, and RNA-CP indicates that PMTV isolates around the world can be divided into two (RNA-rep and RNA-TGB) to three (RNA-CP) lineages depending on genomic segment. In all lineages, there are virus isolates from Andean region of South America demonstrating the greater variability of the virus in the Andes relative to the other parts of the world. Moreover, three of these lineages are exclusively represented by isolates from Andean region of Peru, the center of potato domestication; indicating that in Peru, PMTV has undergone great evolutionary divergence, resulting in the establishment of distinct phylogenetic clades, genotypes and genetic constellations (as PMTV has a tripartite genome).

Virion Structure

Virgaviruses have non-enveloped, rod-shaped particles with helical symmetry with a pitch of 2.3–2.5 nm and an axial canal (Fig. 2). The virions are ca. 20 nm in diameter and up to 310 nm in length. No lipids have been reported in the virions of

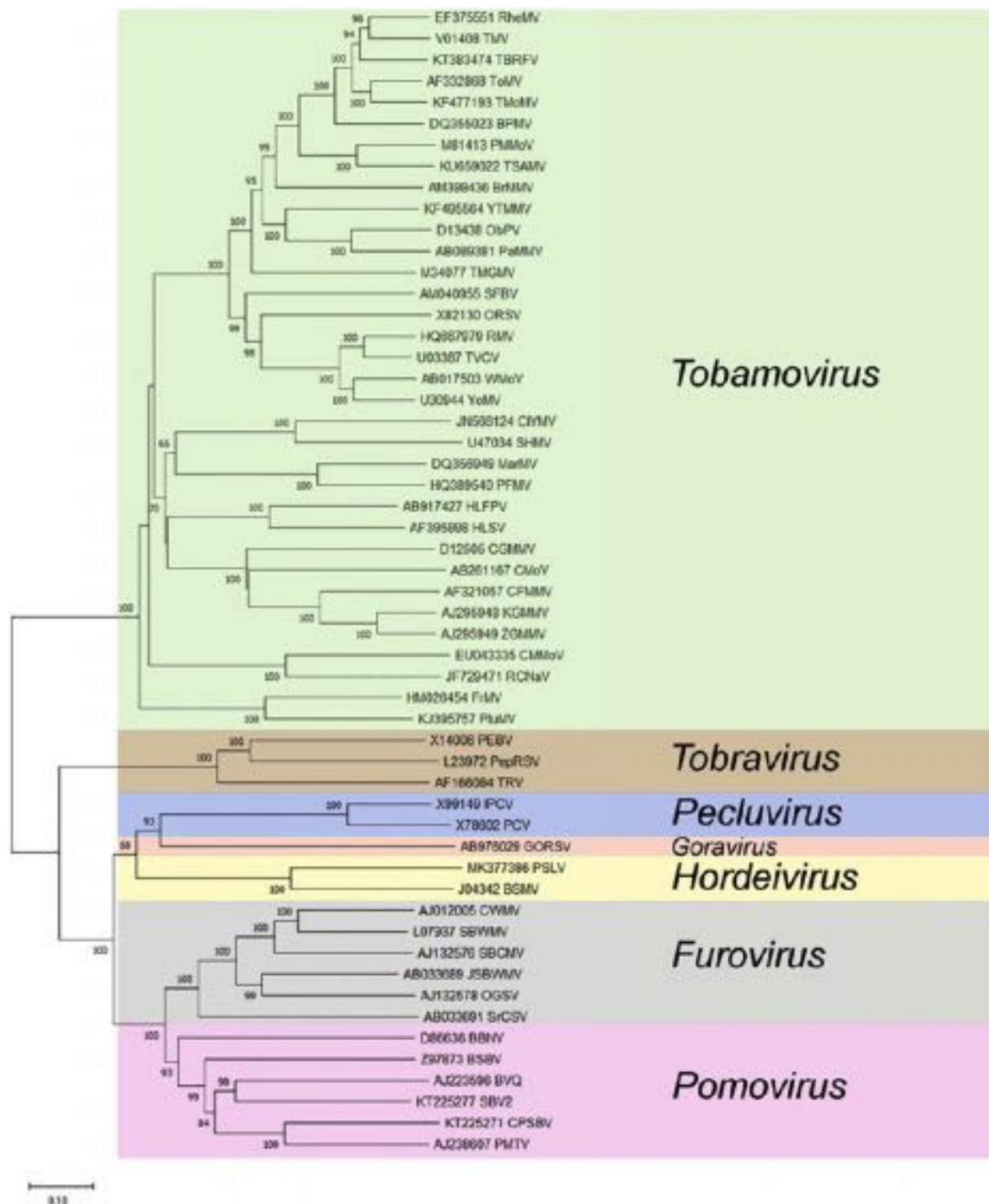


Fig. 1 Neighbor-joining phylogenetic tree of the codon-aligned nucleotide sequences of the replication proteins of plant-infecting viruses in the family *Virgaviridae*. The analyses were conducted in MEGA X using the Maximum Likelihood method, Tamura-Nei model and 1000 bootstrap replicates. The tree with the highest log likelihood (−185021.20) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 54 nucleotide sequences. Virus abbreviations are explained in the [Table 1](#). The bar below the tree indicates nucleotide substitutions per site.

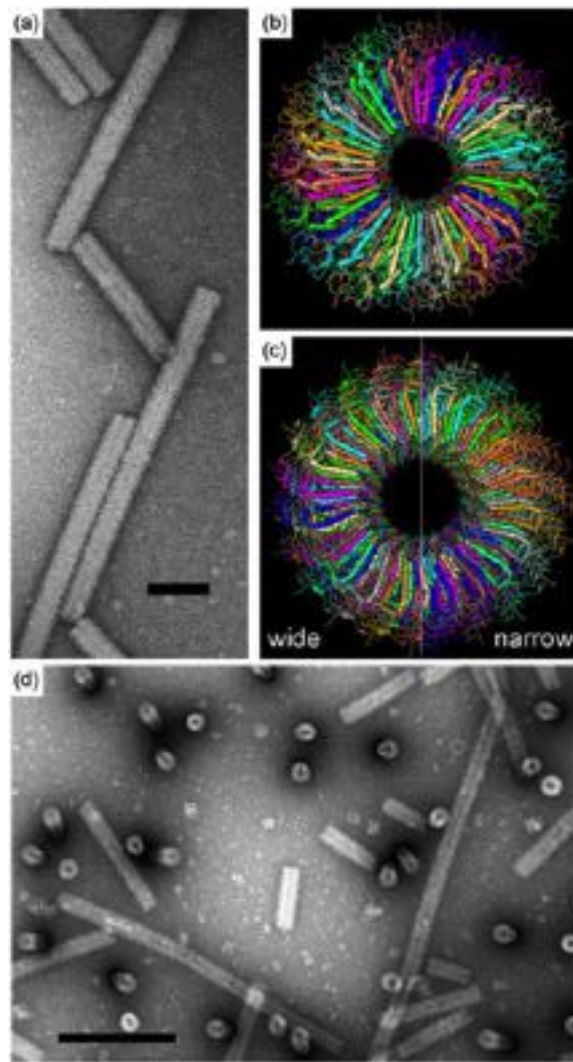


Fig. 2 (A) Negative stain electron micrograph of Barley stripe mosaic virus virions. Scale bar 500 Å. (B) Reconstruction of the Tobacco mosaic virus capsid by cryo-electron microscopy at 4.6 Å resolution. Molecular graphics were generated using data deposited in Protein Data Bank under the accession number 2XEA. (C) Reconstruction of the Barley stripe mosaic virus capsid by cryo-electron microscopy at 4.1 Å resolution. Molecular graphics were generated using data deposited in Protein Data Bank under the accession numbers: 5a79 for the wide structure (left panel) and 5a7a for the narrow structure (right panel). (D) Negative stained electron micrograph of Tobacco rattle virus virions. Scale bar 250 nm. Courtesy Elena V. Orlova, Institute of Structural and Molecular Biology, London, UK. (A) Reproduced from Kendall, A., Williams, D., Bian, W., Stewart, P.L., Stubbs, G., 2013. Barley stripe mosaic virus: Structure and relationship to the tobamoviruses. *Virology* 443, 265–270. Elsevier with permission. (D) Courtesy Elena V. Orlova, Institute of Structural and Molecular Biology, London, UK.

virgaviruses. The virions of all members of the family (except for members of the *Furovirus* and *Pomovirus* genera) are composed of identical CP subunits ranging in size from 17 to 24 kDa. In furoviruses and pomoviruses, the CP-RTD (coat protein – read-through domain) protein – produced by translational read-through of the CP ORF stop codon – is a minor virion component attached to one extremity of the virus particles.

Currently, high-resolution capsid structures are available for five tobamoviruses including TMV, *Cucumber green mottle mosaic virus* (CGMMV), *Tobacco mild green mosaic virus* (TMGMV), *Odontoglossum ringspot virus* (ORSV), *Ribgrass mosaic virus* (RMV), and one hordeivirus – BSMV. The capsid structures of tobamoviruses are very similar to each other, whereas compared to the tobamovirus capsids, the BSMV virions are considerably more labile and, in general, the physico-chemical properties of BSMV virions resemble to certain extent those of flexuous filamentous viruses rather than tobamoviruses.

The X-ray fiber diffraction analysis at 2.9 Å resolution and cryoelectron microscopy (EM) at 4.6 Å resolution has shown that the virions of TMV are built from one distinct conformation of a single CP. The inter-subunit contacts in tobamovirus virions include strong lateral and axial electrostatic interactions, and the hydrophobic girdle formed by lateral hydrophobic contacts between adjacent subunits at a high virion radius (**Fig. 2(B)**). These types of interactions are believed to predispose the exceptionally high stability of tobamovirus virions. Surprisingly, Cryo-EM structure at 4.1 Å resolution has shown that the capsids of the hordeivirus BSMV are built from two

distinct conformations of a single CP, and that these conformational differences facilitate the formation two structurally distinct forms of BSMV virions, 'wide' and 'narrow' virions with diameters of 224 Å and 216 Å, respectively (Fig. 2(C)). Wide virions incorporate 111 subunits per helix period, while narrow virions have 106 subunits. Both virion forms co-exist in the populations of BSMV particles. The functional differences between the two forms of BSMV virions are not known. Structural data show that the narrow BSMV virions are more stable than the wide virions because of the stronger CP contacts with RNA and additional contacts between CP subunits.

Each virion of virgaviruses contains a single copy of ssRNA. Thus, for viruses with bipartite or tripartite genomes, two or three virions, respectively, containing different genomic components are required for establishing infection. The differences in virion size are particularly evident for bipartite tobamoviruses as they have virus particles of two predominant lengths, (L) 180–215 nm and (S) ranging from 46 to 115 nm, depending on the isolate (Fig. 2(D)). RNA1 is encapsidated into the L-particles, whereas the S-particles contain RNA2. Defective genome components (defective interfering RNA, diRNA) associated with pomoviruses have also been found to be encapsidated.

Genome Organization

The 5' untranslated region (UTR) of many virgaviruses contain the starting sequence GU(A)_{1–4}(U)_n. The RNAs are probably capped at the 5' end since the viral replicases contains methyl-transferase motifs associated with capping activity. The terminal ca. 80–240 nt of the 3' UTR can be folded into a tRNA-like structure that contains an anticodon for valine (furo-, peclu- and pomoviruses), tyrosine (hordeiviruses) or histidine (tobamoviruses; Table 2). For some hordei-, pomo- and tobamovirus the genomic RNAs were shown to be amino-acylated (tyrosylated, valylated and histidinated, respectively) experimentally. Different from other virgaviruses, tRNA-like structure of a pomovirus CPSPV contains an anticodone for leucine. Although genomic RNAs of virgaviruses are not polyadenylated, hordeiviruses BSMV and LRSV contain internal poly(A) tracts, which precede the tRNA-like structure.

Genomic RNAs of virgaviruses typically have three to seven ORFs. The genome expression strategy is based on: (i) read-through of leaky stop codons - e.g., expression of replicases and CP-RTD proteins; (ii) expression of the downstream ORFs via the synthesis of a nested set of 3' co-terminal sub-genomic mRNAs (sgRNAs) - e.g., expression of MPs, cysteine-rich proteins and CP (in tobamoviruses); and (iii) leaky scanning - e.g., expression of TGB2/TGB3 (gora-, hordei- peclu- and pomoviruses), and CP/P39 (pecluviruses).

The genome organization is conserved among species within the same genus, but not in the family in general (Fig. 3). However, the 5' proximal ORF of all largest genomic components invariably encodes the replication-associated proteins (replicases) with conserved domains for a methyltransferase (MET), a helicase (HEL), and an RNA-dependent RNA polymerase (RdRp), a domain arrangement typical of alpha-like viruses. The replicases are translated directly from the genomic RNAs. In viruses of all genera except the genus *Hordeivirus*, RdRp is expressed as the C-terminal part of the replicase by read-through of a leaky stop codon (Fig. 3). In the genus *Hordeivirus* RdRp is translated from a smaller genomic component RNA_γ. Other viral proteins are translated either directly from the smaller genomic RNAs or from sgRNAs, some of which may be bicistronic and are translated by leaky scanning mechanism. In viruses of all genera except the genus *Tobamovirus*, CP is translated directly from the smaller genomic RNAs. In the genus *Tobamovirus*, CP is expressed via formation of sgRNA.

The viruses of the genera *Furovirus*, *Tobamovirus* and *Tobravirus* encode a single MP of the "30K" superfamily, whereas in other genera there is a triple gene block (TGB) of MPs (Fig. 3; Table 2). The coordinated expression and action of the proteins encoded by TGB mediates virus movement in the infected plant.

The viruses of the genera *Furovirus*, *Goravirus*, *Hordeivirus*, *Pecluvirus*, *Tobravirus*, and PMTV, but not other pomoviruses, encode cysteine-rich proteins (CRPs, Fig. 3). These CRPs, except the one encoded by PMTV, share well-conserved C3H1-type (C, cysteine residue; H, histidine residue) zinc-finger domain. CRP encoded by PMTV has a SWIM-type zinc-finger characterized by a consensus sequence CxC_nCxC, in which C stands for cysteine residue and x refers to any aa residue. CRPs of virgaviruses are expressed from sgRNAs, function as suppressors of RNA silencing and are essential for virus infectivity or/and efficient virus accumulation in the infected plants. However, in tobamoviruses soluble replication proteins (not associated with membranes) are involved in suppression of RNA silencing, but do not participate in RNA replication (Fig. 3).

Replication

After virus entry, the genomic RNA is translated to produce replication proteins. Replication occurs on virus-induced spherular invaginations of cytoplasmic membranes similar to RNA replication-linked spherules induced by many positive-sense RNA viruses. Full length negative-sense strand synthesis and subsequent positive-sense strand synthesis takes place inside the spherules. Replication proteins bind to the 3' terminal tRNA-like structure to initiate negative-strand RNA synthesis. In the spherule, the interior space of which is connected with the cytoplasm via a thin neck, viral positive-sense RNA strands are synthesized using negative-sense strand as a template. Subsequently, multiple copies of positive-sense RNA strands and sgRNAs are released into the cytoplasm through the neck. The replication proteins of virgaviruses, as those of other positive-sense RNA viruses, are tightly associated with membranes. Attempts to purify viral replicases from infected plants cells and to solubilize them resulted in drastic changes in their properties.

sgRNAs are transcribed from negative-sense RNA through recognition of sub-genomic promoters by replication proteins. For some virgaviruses, comparison of the sequences of sub-genomic promoters reveals conservative consensus motifs.

The membrane used to form the replication compartments varies and depends on the virus, for example, TMV, ToMV and Youcai mosaic virus (and probably other tobamoviruses) may forms active replication machinery on the vacuolar and endoplasmic

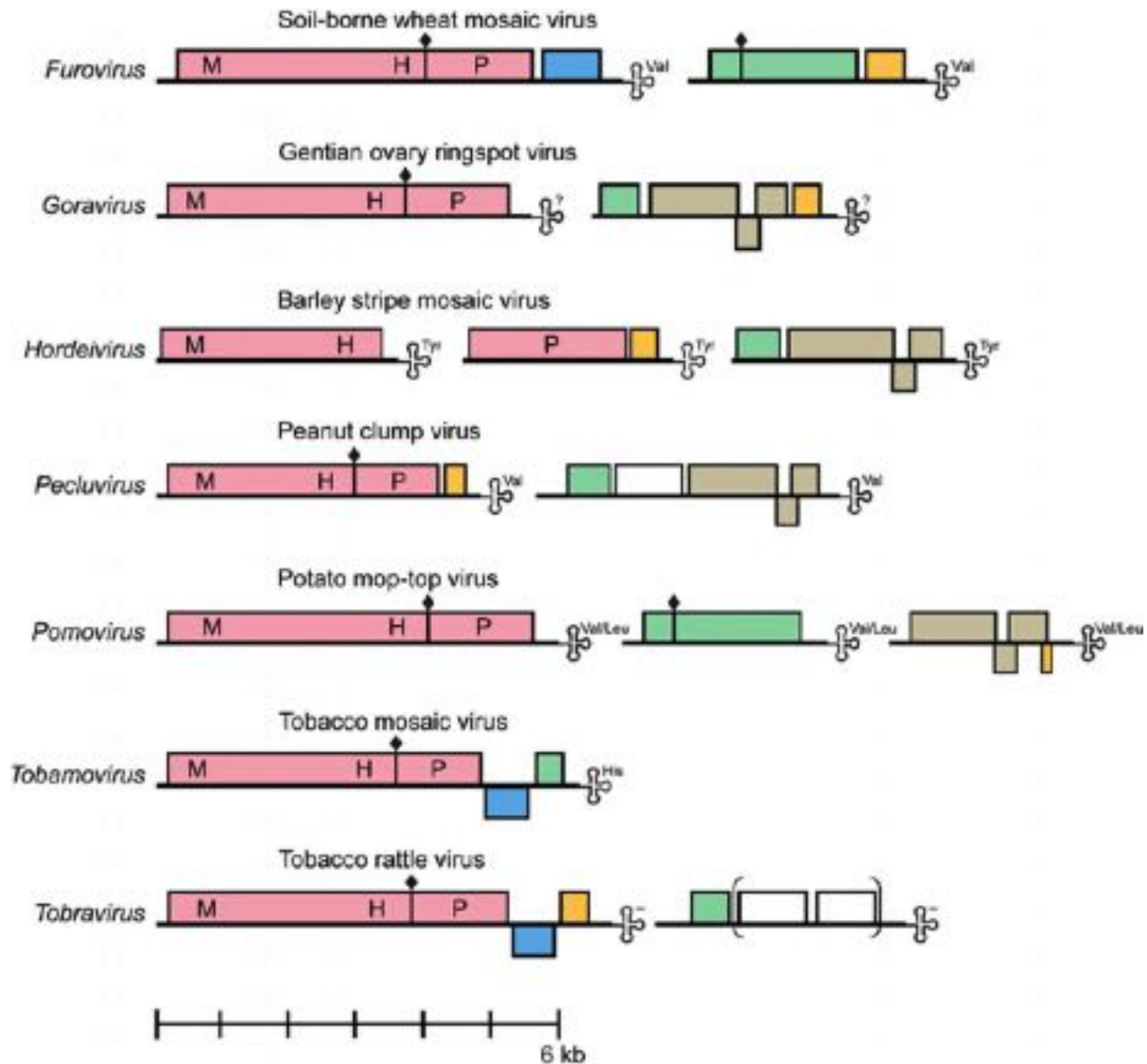


Fig. 3 Diagram showing the genome organization of representative members of the seven genera in the family *Virgaviridae*. ORFs are color-coded according to the function of their protein products: replication proteins are in pink with the methyltransferase (M), helicase (H), and RdRp (P) domains; coat proteins are in green; '30K'-type movement proteins are in blue; triple gene block movement proteins are in light brown; cysteine-rich proteins are in yellow; and other ORFs are in white. Positions of 'leaky' stop codons are shown by black diamonds and tRNA-like structures with clover leaves. Val/Tyr/Leu/His/?/-: amino-acylation of tRNA-like structures, respectively. Brackets indicate ORFs that are missing from some strains.

reticulum (ER) membranes, whereas BSMV (a hordeivirus) uses the outer membranes of chloroplasts for replication. Evidence suggests that PMTV TGB2, a two-pass transmembrane protein, targets membranes of chloroplasts to establish the virus replication compartment in there. TRV viral factories are enriched in mitochondria-derived membranes.

Tobamovirus replication proteins co-translationally bind to AC-rich region of the 5' UTR of the genomic RNA to form so-called premembrane targeting complex. This represents the first step in tobamovirus RNA replication. Host factors involved in the tobamovirus replication were identified by forward genetic approaches and co-immunoprecipitation, and include TOM1, TOM2A, TOM2B, and ARL8 proteins. Tobamovirus replication complexes are anchored into membranes through interaction of the helicase domain of the replicase with TOM1, a seven-pass transmembrane protein. In turn, TOM2A, a four-pass transmembrane protein, interacts with TOM1. ARL8, an ADP-ribosylation factor-like small GTP-binding protein, interacts with both TOM1 and an N-terminal alpha-helix of the helicase domain of the replication proteins. The replication proteins first bind to membranes and then interact and form a complex with TOM1, TOM2A, and ARL8. Interaction with membranes and TOM1 and ARL8 proteins activates the RNA 5' capping activity of the 126 kDa protein (a tobamovirus protein harboring methyl-transferase and helicase domains). The timing and the trigger of the activation of the tobamovirus RdRp activity and the mechanism of its template strand switch are unknown so far.

When significant amount of positive-sense genomic RNA builds up in the infected cell, progeny rod-shaped virions are assembled in the cytoplasm. Mature virions accumulate in the infected cells and might be acquired by the vectors.

Infectious Cycle

Infection with virgaviruses is initiated when a virus is delivered into the plant cell. Depending on the virus genus this is achieved either by rub-inoculation (mechanical transmission) – the mode of virus transmission possible for most virgaviruses and the only way of transmission for the viruses of the genus *Tobamovirus* – or by the vector (plasmodiophorid or nematode). In the plant cell, ssRNA of virgaviruses gets uncoated from virions and is translated to produce replication-associated proteins. The replication complex then initiates rearrangement and remodeling of host intracellular membranes to establish viral replication organelles or virus factories. Such compartments secure efficient RNA replication and synthesis of sgRNA, and protect viral RNA from nucleases and pattern recognition receptors that upon activation can trigger host innate immunity responses, *e.g.*, in response to dsRNA accumulation. Later in infection, viral MPs are expressed from sgRNA (for most virgaviruses). MPs facilitate cell-to-cell and systemic movement of the virus in the infected plant. Some members of the family *Virgaviridae* require expression of CP for systemic movement. When sufficient amount of CP is expressed, ssRNA can then be encapsidated by CP into virions, which become available for vector acquisition and subsequent transmission. Virions of virgaviruses can aggregate forming inclusion bodies, virion arrays and other paracrystalline structures in the plant cell.

An important step of the life cycle of the soil-borne viruses of the *Furovirus*, *Pecluvirus* and *Pomovirus* genera is the transmission between plants by a plasmodiophorid vector. The dormant resting spores of plasmodiophorids carrying a virgavirus can persist in soil for more than fifteen years. Under high levels of soil moisture and cool temperatures, the resting spores germinate and release primary virus-carrying zoospores that attach to and penetrate root tissue of the host, mostly root hairs. Zoospores are responsible for transmission of virgaviruses to host plants and the viruses reside inside the zoospores. After penetration, the zoospore develops into a multinucleate plasmodium, which after growth, increasing in size and proliferating into segments forms zoosporangia, each harboring numerous secondary zoospores that can infect new roots. Although the exact mechanism of the virus acquisition by plasmodiophorids is not known, it has been speculated that it might occur at the plasmodium stage, probably, through fusion of membranes induced by CP-RTD proteins located at one extremity of virus particles. Intriguingly, CP-RTD proteins of virgaviruses have two transmembrane hydrophobic segments - the arrangement resembles membrane fusion proteins of enveloped animal viruses. Deletions in CP-RTD proteins resulting in the loss of at least one transmembrane segment prevent transmission of the virus by the vector. Although, whether virgaviruses replicate in the plasmodiophorid vector is still unknown.

Diagnosis

The members of the family *Virgaviridae* incite various leaf symptoms including yellow mosaic, ring spots, stripes, streaks, necrotic spots, leaf malformations *etc.* However, identification of the virgaviruses based these symptoms might be misleading because similar symptoms may be induced by other viruses. Thus, diagnosis based on appearance of symptoms is not sufficient and specific diagnostic tests must be employed to confirm virgavirus infection.

Brown arcs and rings in the potato tuber flesh are typical symptoms of ‘spraing’, a Scottish word referring to a dark, variegated streak or stripe. ‘Spraing’ can be caused by PMTV (genus *Pomovirus*) or TRV (genus *Tobravirus*). Reliable methods are needed to determine whether symptoms are caused by PMTV or TRV.

Several techniques have been developed and successfully used to detect the presence of virgaviruses, including serological assays, nucleic-acid-based methodologies, and next generation sequencing (NGS). A number of serological diagnostic techniques such as an enzyme-linked immunosorbent assay (ELISA) have been developed for fast, simple, cost-effective and high-throughput testing. Serological reagents are available for many viruses in the family *Virgaviridae*, including reagents for ELISA (*e.g.*, for PMTV, TMV, ToMV, TRV *etc.*) and strips for lateral flow detection (*e.g.*, for detection of TMV and ToMV). Lateral flow assays provide very rapid diagnosis within minutes, are easily portable and highly robust, and have been adopted by agronomists and farm advisors as means of quickly confirming the presence of a specific virgavirus within a crop in the field.

Nucleic acid-based diagnostic methods have become increasingly popular for detection and characterization of virgaviruses. Such techniques include various modifications of reverse transcription (RT) polymerase chain reaction (PCR): classical RT-PCR, immunocapture-RT-PCR, quantitative real-time RT-PCR (RT-qPCR), loop-mediated isothermal amplification, and oligonucleotide microarrays. These assays are often more robust and sensitive than serological assays. Recently, NGS, including RNA-seq and sequencing of small RNAs, has become increasingly accessible for discovery of new viruses, including virgaviruses. These holistic approaches provide an unbiased identification and characterization of a whole virome in a sample without any prior knowledge of the pathogens present.

Pathogenicity

Despite significant advances in studies of the mechanisms of pathogenicity (the ability of a virus to cause disease on a particular host), the information about how virgaviruses cause disease in their host is surprisingly scant. Pathogenicity determinants include

CRPs encoded by the viruses of the *Furovirus*, *Goravirus*, *Hordeivirus*, *Pecluvirus*, *Pomovirus*, and *Tobravirus* genera. For example, natural variants of CRPs from various BSMV strains have a number of specific effects on the virus pathogenesis including the ability of certain strains to establish systemic infection in the absence of the CRP expression. However, the failure of certain strains to establish systemic infection can be reversed by mutations (deletions of N-terminal duplication of the polymerase domain) that increase the abundance of viral polymerase, and, thus, increase the virus replication rate.

The function of CP in pathogenicity and symptom modulation was shown for pomo- and tobamoviruses. PMTV causes symptomless infections in the absence of the CP expression. On the contrary, TMV mutants defective in CP expression still display mild mosaic symptoms. However, the appearance of bright chlorosis symptoms is linked to specific aa substitutions within the TMV CP and results from disruption of chloroplast development and homeostasis. TMV CP associates with and disrupts photosystem II complexes in the thylakoid membranes of chloroplasts.

Epidemiology and Control

The viruses in the genera *Furovirus*, *Pecluvirus* and *Pomovirus* are transmitted in soil by plasmodiophorid vectors (Table 2). *Polymyxa graminis* and *Spongospora subterranea* are known vectors of the viruses in these three genera. Viruses transmitted by plasmodiophorids reside inside the zoospores or dormant resting spores of their vector and, thus, well protected from the environment. The virus-carrying resting spores also remain viruliferous for many years. These features make eradication of the diseases caused by plasmodiophorid-transmitted viruses very difficult, and infestations are usually permanent, once the virus is introduced into the field.

The viruses in the genus *Tobravirus* are transmitted in soil by nematode vectors. In general, management of nematodes is difficult. The existing infestations can be reduced through fallowing, crop rotation, and soil solarization.

Transmission through seed and pollen has been documented for the viruses in the genera *Goravirus* and *Hordeivirus*. No vector is known for viruses in the genus *Tobamovirus*. Tobamoviruses are readily transmittable by mechanical inoculation. Most of the viruses in the family *Virgaviridae* are mechanically transmittable in lab conditions.

The family *Virgaviridae* includes viruses (BSMV, PCV, PMTV, TMV, ToMV, TRV etc.) that are responsible for significant quantitative and qualitative yield losses in various crops, whereas the economic significance of other family members is limited, not well studied or not important. Many factors affect the occurrence and incidence of diseases caused by virgaviruses. Application of integrated pest management (IPM) strategies would be beneficial to control of virgavirus-induced diseases as IPM takes into account all those factors. For plasmodiophorid- and nematode-transmitted viruses disease management is mainly preventive through the use of virus-free seed material (e.g., virus-free seed tubers) and non-infested fields. Cross-protection with mild viral strains has been used successfully to control some virus diseases caused by tobamoviruses (e.g., TMV and ToMV). Mild strain protection of tomato against ToMV is probably one of the most successful applications of cross-protection. It is successfully and routinely used in Europe, Canada, New Zealand, Israel, and Japan.

One of the best ways to manage virgaviruses is to use varieties that are resistant to the viruses. Four dominant genes conferring resistance to virgaviruses have been cloned and sequenced. Three of these genes encode members of the nucleotide-binding site-leucine-rich repeat (NBS-LRR) class of R proteins, whereas *Tm-1* gene encodes a TIM barrel structure protein. The R proteins localize to the cytoplasm consistent with the lifecycle of virgaviruses. ToMV was once one of the most important pathogens of tomato crop, but effective control has been established using resistant cultivars with three ToMV resistance genes: *Tm-2/Tm-2²* and *Tm-1*. *N*-gene-mediated resistance to TMV is another example of solving a practical problem and putting disease resistance in the field.

Unassigned Virgaviruses

Recent advances in low-cost high-throughput NGS led to identification of a number of virga-like viruses in metagenomics data generated for invertebrates, fungi and oomycetes (Table 3). Most of these viruses seem to be monopartite with genomes ranging 6.8–11.9 kb. However, one cannot exclude the possibility that some of the viruses in this group might have multipartite genomes as currently only partial sequences are available.

Further Reading

- Adams, M.J., Adkins, S., Bragard, C., *et al.*, 2017. ICTV virus taxonomy profile: Virgaviridae. *Journal of General Virology* 98, 1999–2000.
- Chujo, T., Ishibashi, K., Miyashita, S., Ishikawa, M., 2015. Functions of the 5'- and 3'-untranslated regions of tobamovirus RNA. *Virus Research* 206, 82–89.
- Gibbs, A.J., Wood, J., Garcia-Arenal, F., Ohshima, K., Armstrong, J.S., 2015. Tobamoviruses have probably co-diverged with their eudicotyledonous hosts for at least 110 million years. *Virus evolution* 1, vev019.
- Gil, J.F., Adams, I., Boonham, N., Nielsen, S.L., Nicolaisen, M., 2016. Molecular and biological characterisation of two novel pomo-like viruses associated with potato (*Solanum tuberosum*) fields in Colombia. *Archives of Virology* 161, 1601–1610.
- Ishibashi, K., Ishikawa, M., 2016. Replication of tobamovirus RNA. *Annual Review of Phytopathology* 54, 55–78.
- Jackson, A.O., Lim, H.S., Bragg, J., Ganesan, U., Lee, M.Y., 2009. Hordeivirus replication, movement, and pathogenesis. *Annual Review of Phytopathology* 47, 385–422.

- Kalyandurg, P., Gil, J.F., Lukhovitskaya, N.I., *et al.*, 2017. Molecular and pathobiological characterization of 61 Potato mop-top virus full-length cDNAs reveals great variability of the virus in the centre of potato domestication, novel genotypes and evidence for recombination. *Molecular Plant Pathology* 18, 864–877.
- Koonin, E.V., Dolja, V.V., 1993. Evolution and taxonomy of positive-strand RNA viruses. *Crit Rev Biochem Mol Bio* 28, 375–430.
- Lecoq, H., 1998. Control of plant virus diseases by cross-protection. In: Hadidi, A., Khetarpal, R.K., Koganezawa, H. (Eds.), *Plant Virus Disease Control*. St. Paul, MN: APS Press, pp. 33–40.
- Melcher, U., 2000. The '30K' superfamily of viral movement proteins. *Journal of General Virology* 81, 257–266.
- Rochon, D.'A., Kakani, K., Robbins, M., Reade, R., 2004. Molecular aspects of plant virus transmission by oïdium and plasmodiophorid vectors. *Annual Review of Phytopathology* 42, 211–241.
- Solovyev, A.G., Makarov, V.V., 2016. Helical capsids of plant viruses: Architecture with structural lability. *Journal of General Virology* 97, 1739–1754.
- Solovyev, A.G., Savenkov, E.I., 2014. Factors involved in systemic transport of plant RNA viruses, the emerging role of the nucleus. *Journal of Experimental Botany* 65, 1689–1697.
- Zhang, K., Zhang, Y., Yang, M., *et al.*, 2017. The Barley stripe mosaic virus gammab protein promotes chloroplast-targeted replication by enhancing unwinding of RNA duplexes. *PLoS Pathogens* 13, e1006319.

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/virgaviridae
Virgaviridae-Positive-sense RNA Viruses-ICTV.
- <https://talk.ictvonline.org/taxonomy/>
Taxonomy-International Committee on Taxonomy of Viruses.

Viroids (*Pospiviroidae* and *Avsunviroidae*)

Ricardo Flores, Polytechnic University of Valencia, Higher Council of Scientific Research, Valencia, Spain

Robert A Owens, Beltsville Agricultural Research Center, Beltsville, MD 20705, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Catalytic RNA RNA molecules that are able to catalyze, in a protein-free medium, specific reactions involving the formation or breakage of covalent bonds. In nature, these reactions are usually transesterifications (self-cleavage and ligation) affecting the catalytic RNA itself.

Hammerhead structure The conserved secondary/tertiary structure shared by the smallest class of natural ribozymes. Most were initially found in one or both strands of certain

viroid and viroid-like satellite RNAs where they mediate self-cleavage of multimeric intermediates arising from replication through a rolling-circle mechanism. More recently, the hammerhead ribozyme motif has been reported along the tree of life.

Ribozyme RNA motif responsible for the catalytic activity of certain RNA molecules. In nature, they are found embedded within catalytic RNAs.

Taxonomy

According with the presence of conserved sequence/structure motifs, and mode of replication and subcellular accumulation sites, viroids are allocated to the family *Pospiviroidae* (composed of genera *Pospiviroid*, *Hostuviroid*, *Cocadviroid*, *Apsacaviroid* and *Coleviroid*), type species *Potato spindle tuber viroid*, and family *Avsunviroidae* (composed of genera *Avsunviroid*, *Pelamoviroid* and *Elaviroid*), type species *Avocado sunblotch viroid*.

Introduction

Viroids are the smallest known agents of infectious disease – small (246–401 nt), highly structured, single-stranded, circular RNAs – that lack detectable messenger RNA activity. Whereas viruses have been described as “obligate parasites of the cell’s translational system” and supply some or most of the genetic information required for their replication and movement, viroids can be regarded as “obligate parasites of the cell’s transcriptional machinery”. Thus far, viroids are known to infect only plants.

The first viroid disease to be studied by plant pathologists was potato spindle tuber. In 1923, its infectious nature and ability to spread in the field led Schultz and Folsom to group potato spindle tuber disease with several other “degeneration diseases” of potatoes. Nearly 50 years were to elapse before Diener’s 1971 demonstration that the molecular properties of its causal agent, Potato spindle tuber viroid (PSTVd), were fundamentally different than those of conventional plant viruses.

Genome Structure

Efforts to understand how viroids replicate and cause disease without the assistance of any viroid-encoded polypeptides has prompted detailed analysis of their structure. Viroids possess rather unusual properties for single-stranded RNAs (*e.g.*, a pronounced resistance to digestion by ribonucleases and a highly cooperative thermal denaturation profile), leading to an early realization that they might have an unusual higher-order structure.

To date, the complete sequences of 32 distinct viroid species plus a large number of sequence variants have been determined (Table 1). All are single-stranded circular RNAs with unmodified nucleotides. Theoretical calculations and physicochemical studies indicate that PSTVd and related viroids assume a highly base-paired, rod-like conformation *in vitro* (Fig. 1), with recent data supporting a similar secondary structure *in vivo* for PSTVd and Avocado sunblotch viroid (ASBVd), which appear to accumulate *in planta* as free RNAs unprotected by tightly bound proteins. Pair-wise sequence comparisons suggest that the series of short double helices and small internal loops that comprise this so-called “native” structure are organized into five domains whose boundaries are defined by sharp changes in sequence identity.

The “central domain” is the most highly conserved viroid domain and contains the site where multimeric PSTVd RNAs are cleaved and ligated to form circular progeny. The “pathogenicity domain” contains one or more structural elements that modulate symptom expression, and the relatively small “variable domain” exhibits the greatest sequence variability between otherwise closely-related viroids. The two “terminal domains” appear to play an important role in viroid replication and evolution. Although these five domains were first identified in PSTVd, Apple scar skin viroid (ASSVd) and related viroids also contain a similar domain arrangement. Certain viroids, such as Columnea latent viroid (CLVd), Australian grapevine viroid (AGVd) and *Coleus blumei* viroid 2 (CbVd 2), appear to be “mosaic molecules” formed by exchange of domains between two or more viroids infecting the same cell. RNA rearrangement/

Table 1 Classification of viroids of known nucleotide sequence

Family ^a	Genus ^a	Name	Abbreviation	Nucleotides ^b
<i>Pospiviroidae</i>	<i>Pospiviroid</i>	<i>Chrysanthemum stunt viroid</i>	CSVd	354–356
		<i>Citrus exocortis viroid</i>	CEVd	368–375 (463–467)
		<i>Columnnea latent viroid</i>	CLVd	370–373
		<i>Iresine viroid</i>	IrVd	370
		<i>Pepper chat fruit viroid</i>	PCFVd	348
		<i>Potato spindle tuber viroid</i>	PSTVd	356–361 (341)
		<i>Tomato apical stunt viroid</i>	TASVd	360–363
		<i>Tomato chlorotic dwarf viroid</i>	TCDVd	360
		<i>Tomato planta macho viroid</i>	TPMVd	359–360
		<i>Citrus bark cracking viroid</i>	CBCVd	282–286 (273–277)
		<i>Coconut cadang-cadang viroid</i>	CCCVd	246–247 (287–301)
		<i>Coconut tinangaja viroid</i>	CtiVd	254
		<i>Hop latent viroid</i>	HLVd	256
	<i>Cocadviroid</i>	<i>Dahlia latent viroid</i>	DLVd	342
		<i>Hop stunt viroid</i> ^c	HSVd	294–303
	<i>Apscaviroid</i>	<i>Apple dimple fruit viroid</i>	ADFVd	306,307
		<i>Apple scar skin viroid</i> ^d	ASSVd	329–334
		<i>Australian grapevine viroid</i>	AGVd	369
		<i>Citrus bent leaf viroid</i>	CBLVd	315,318
		<i>Citrus dwarfing viroid</i>	CDVd	294,297
		<i>Citrus viroid V</i>	CVd-V	284
		<i>Citrus viroid VI</i>	CVd-VI	330
		<i>Grapevine yellow speckle viroid 1</i>	GVYSVd 1	366–368
		<i>Grapevine yellow speckle viroid 2</i> ^e	GVYSVd 2	363
		<i>Pear blister canker viroid</i>	PBCVd	315,316
	<i>Coleviroid</i>	<i>Coleus blumei viroid 1</i>	CbVd 1	248–251
		<i>Coleus blumei viroid 2</i>	CbVd 2	301,302
		<i>Coleus blumei viroid 3</i>	CbVd 3	361–364
<i>Avsunviroidae</i>	<i>Avsunviroid</i>	<i>Avocado sunblotch viroid</i>	ASBVd	246–251
	<i>Pelamoviroid</i>	<i>Chrysanthemum chlorotic mottle viroid</i>	CchMVd	398–401
		<i>Peach latent mosaic viroid</i>	PLMVd	335–339 (348–351)
	<i>Elaviroid</i>	<i>Eggplant latent viroid</i>	ELVd	332–335

^aClassification follows scheme proposed by Owens *et al.* 2012 (see Virus Taxonomy, IX Report of the International Committee on Taxonomy of Viruses) with minor modifications. Whether Apple fruit crinkle viroid, Apple hammerhead viroid, Coleus blumei viroids 4, 5 and 6, Grapevine hammerhead viroid-like RNA, Grapevine latent viroid, Grapevine yellow speckle viroid 3, Persimmon latent viroid, Persimmon viroid 2, and Portulaca latent viroid should be considered variants or new viroid species of a different genus is pending.

^bSizes of variants containing insertions or deletions arising *in vivo* are shown in parentheses.

^cIncludes Cucumber pale fruit viroid, Citrus cachexia viroid, Peach dapple viroid, and Plum dapple viroid.

^dIncludes Pear rusty skin viroid and Dapple apple viroid.

^eFormerly termed Grapevine viroid 1B.

recombination can also occur within individual domains, leading, in Coconut cadang-cadang viroid (CCCVd) and Citrus exocortis viroid (CEVd), to duplications of the right terminal domain plus part of the variable domain. This domain model is not shared by ASBVd and related viroids.

Much less is known about the tertiary structure of viroids, especially *in vivo*. UV-induced cross-linking of two nucleotides within a loop E motif in the central domain of PSTVd provided the first definitive evidence for such tertiary interactions. Similar UV-sensitive structural elements have also been discovered in a number of other RNAs including 5S eukaryotic rRNA, adenovirus VAI RNA, and the viroid-like domain of the hepatitis delta virus genome. Loop E forms during the conversion of multimeric PSTVd RNAs into monomers. The ability of ASBVd-related RNAs to undergo spontaneous self-cleavage mediated by hammerhead ribozymes, as well as the presence of kissing-loop interactions critical for infectivity in some other members of the *Avsunviroidae* (Fig. 1), provide additional evidence for the functional importance of viroid tertiary structure.

Classification

Based upon differences in their structural and functional properties, viroids are assigned to one of two taxonomic families (Table 1). Members of the family *Pospiviroidae* (type species PSTVd) fold into a rod-like secondary structure with five structural-

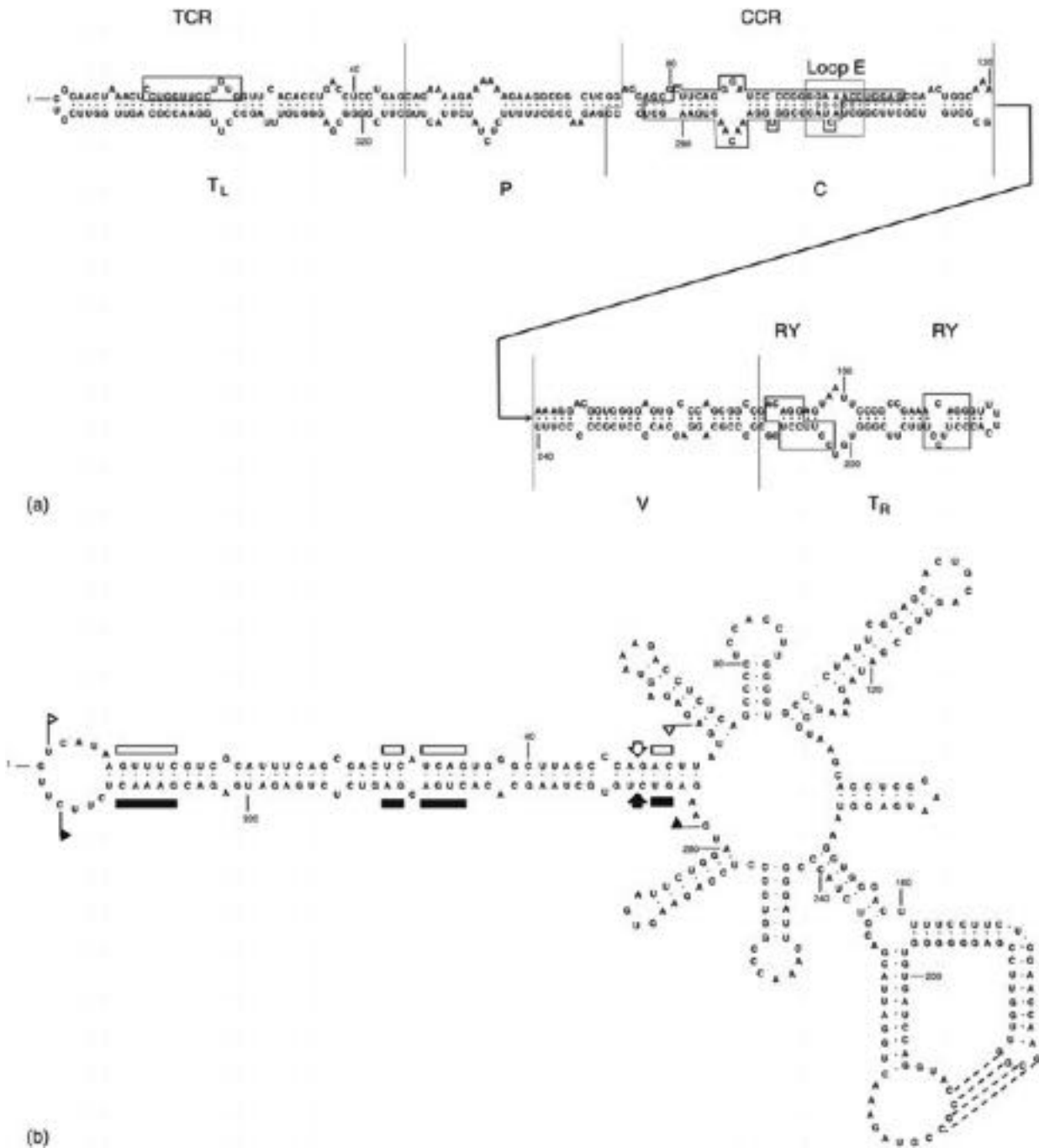


Fig. 1 (a) The rod-like secondary structure of PSTVd (intermediate strain) showing the five domains characteristic of members of the family *Pospiviroidae*: the Terminal Left (T_L), Pathogenicity (P), Central (C), Variable (V), and Terminal Right (T_R). The Central Conserved Region (CCR) is located within the C domain and contains a UV-sensitive loop E motif with non-canonical base-pairs (denoted by circles). The T_L domains of genera *Pospiviroidae* and *Apscaviroid* contain a Terminal Conserved Region (TCR), while those of genera *Hostuviroid* and *Cocadviroid* contain a Terminal Conserved Hairpin (not shown). The T_R may also contain 1–2 copies of a protein-binding RY motif. (b) The branched secondary structure of PLMVd (reference variant). Plus and minus self-cleavage domains are indicated by flags, nucleotides conserved in most natural hammerhead structures by bars, and the self-cleavage sites by arrows. Black and white symbols refer to plus and minus polarities, respectively. Nucleotides involved in a kissing-loop interaction are indicated by broken lines. Redrawn with modifications from Gross, H.J., Domdey, H., Lossow, C., *et al.*, 1978. Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature* 273, 203–208. Hernández, C., Flores, R., 1992. Plus and minus RNAs of peach latent mosaic viroid self-cleave in vitro via hammerhead structures. *Proceedings of the National Academy of Sciences of the United States of America* 89, 3711–3715. Bussiére, F., Ouellet, J., Côté, F., Lévesque, D., Perreault, J.P., 2000. Mapping in solution shows the peach latent mosaic viroid to possess a new pseudoknot in a complex, branched secondary structure. *Journal of Virology* 74, 2647–2654.

functional domains and several conserved motifs. Most members of the family *Avsunviroidae* (type species ASBVd), in contrast, appear to adopt a branched conformation, and multimeric RNAs of all family species behave as catalytic RNAs and undergo spontaneous self-cleavage (Fig. 1). Differences in their sites of replication also support this classification scheme; *i.e.*, PSTVd and ASBVd replicate in the nucleus and plastids (mostly chloroplasts), respectively, and the same appears to occur for other family species. Each family is subdivided into genera according to certain demarcating criteria. Groups of sequence variants that show > 90% sequence identity in pair-wise comparisons and share some common biological property are arbitrarily defined as viroid species. *In vivo*, each viroid species is actually a “quasi-species”; *i.e.*, a collection of closely-related sequence variants subject to a continuous process of mutation, competition, and selection. Phylogenetic evidence supports an evolutionary link between viroids and other viroid-like subviral RNAs (Fig. 2).

Host Range and Transmission

All viroids are mechanically transmissible, and most are naturally transmitted from plant to plant by man and his tools. Individual viroids vary greatly in their ability to infect different plant species. PSTVd can replicate in about 160 primarily solanaceous hosts, while only two members of the family *Lauraceae* are known to support ASBVd replication. Hop stunt viroid (HSVd) has a particularly wide host range that includes herbaceous species as well as woody perennials. Many natural hosts are either vegetatively propagated or crops that are subjected to repeated grafting or pruning operations. PSTVd, ASBVd, and *Coleus blumei* viroid 1 (CbVd 1) are vertically transmitted through pollen and/or true seed, but the significance of this mode of transmission in their natural spread is unclear. PSTVd can be encapsidated by the coat protein of Potato leafroll virus (PLRV, a luteovirus) as well as Velvet tobacco mottle virus (VTMoV, a sobemovirus), and epidemiological surveys suggest that PLRV facilitates viroid spread under field conditions.

Commonly used techniques for the experimental transmission of viroids include the standard leaf abrasion methods developed for conventional viruses, “razor slashing” methods in which phloem tissue in the stem or petiole is inoculated via cuts made with a razor blade previously dipped into the inoculum, and, in the case of CCCVd, high-pressure injection into folded apical leaves. Viroids can also be transmitted by either plant transformation or “agro-inoculation” during which a modified *Agrobacterium tumefaciens* Ti plasmid is used to introduce full-length viroid-complementary DNA into the potential host cell. Either technique can overcome the marked resistance of some hosts to mechanical inoculation. Identification of the molecular mechanism(s) that determine viroid host range remains an important research goal.

Symptomatology

Viroids and conventional plant viruses induce a very similar range of macroscopic symptoms. Symptom expression is usually optimal at the same relatively high temperatures (30–33°C) that promote viroid replication/accumulation. Stunting and leaf epinasty (a downward curling of the leaf lamina resulting from unbalanced growth within the various cell layers) are considered the classic symptoms of viroid infection. Other commonly observed symptoms include vein clearing, veinal discoloration or necrosis, and the appearance of localized chlorotic/necrotic spots or mottling in the foliage. Symptoms may also be expressed in flowers and bark, and fruits or tubers from viroid-infected plants may be abnormally shaped or discolored. Viroid infection of certain citrus rootstock/scion combinations may result in tree dwarfing (Fig. 3). Viroid infections are often latent and rarely kill the host.

Viroid infections are also accompanied by a number of cytopathic effects – chloroplast and cell wall abnormalities, the formation of membranous structures in the cytoplasm, and the accumulation of electron-dense deposits in both chloroplasts and cytoplasm. Metabolic changes include dramatic alterations in growth regulator levels.

Geographic Distribution

Although PSTVd, CEVd, HSVd and ASBVd are widely distributed throughout the world, other viroids have never been detected outside the areas where they were first reported. Several factors may contribute to this variation in distribution pattern. Among the crops most affected by viroid diseases are a number of valuable woody perennials such as grapes, citrus, various pome and stone fruits, and hops. Propagation and distribution of improved cultivars is highly commercialized, with the result that many cultivars are now grown worldwide. The international exchange of plant germplasm also continues to increase at a rapid rate. In both instances, the large number of latent (asymptomatic) hosts facilitates viroid spread.

Epidemiology and Control

Viroid diseases represent a potential threat to agriculture, and several are of considerable economic importance. Ready transmission of PSTVd by vegetative propagation, foliar contact, and true seed or pollen continues to pose a serious threat to potato germplasm and breeding programs. Coconut cadang-cadang has killed over 30 million palm trees in the Philippines since it was first recognized in the early 1930s. While many viroids were first detected in ornamental or crop plants, most viroid diseases are

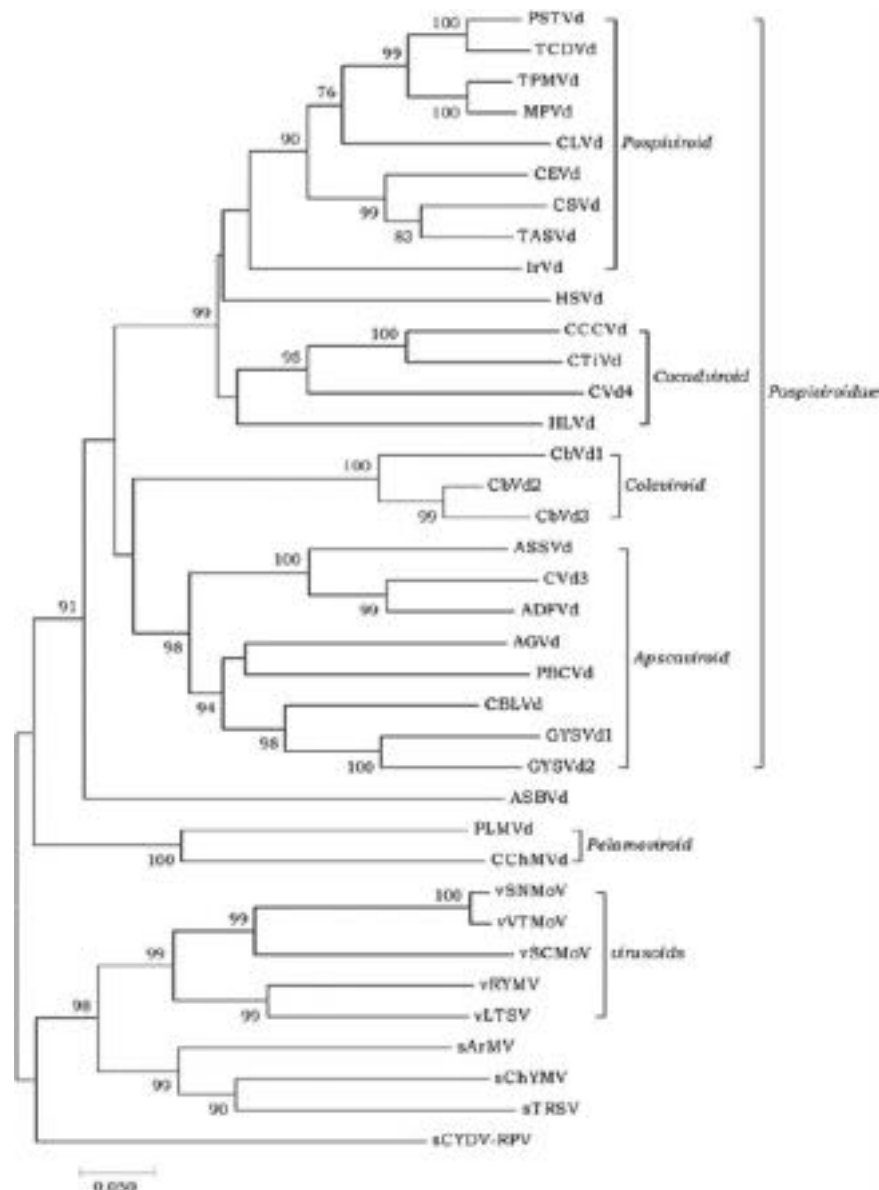


Fig. 2 Neighbor-joining phylogenetic tree obtained from an alignment manually adjusted to take into account local similarities, insertions/deletions, and duplications/rearrangements described in the literature for viroid and viroid-like satellite RNAs. Bootstrap values were based on 1000 random replicates (only values > 70% are reported). Viroid abbreviations as in [Table 1](#). Viroid-like satellite RNAs: sLTSV (lucerne transient streak virus); sRYMV (rice yellow mottle virus); sSCMoV (subterranean clover mottle virus); sSNMoV (*Solanum nodiflorum* mottle virus); sVTMoV (velvet tobacco mottle virus); sTRSV (tobacco ringspot virus); sArMV (*Arabidopsis* mosaic virus); sChYMV (chicory yellow mottle virus); sCYDV-RPV (cereal yellow dwarf virus-RPV). Adapted from Elena, S.F., Dopazo, J., de la Peña, M., *et al.*, 2001. Phylogenetic analysis of viroid and viroid-like satellite RNAs from plants: A reassessment. *Journal of Molecular Evolution* 53, 155–159, with permission.

thought to result from chance transfer from endemically-infected wild species to susceptible cultivars ([Diener, 1979](#)). Several lines of circumstantial evidence are consistent with this hypothesis:

- (1) The experimental host ranges of several viroids include many wild species, and these wild species often tolerate viroid replication without the appearance of recognizable disease symptoms.
- (2) Although co-evolution of host and pathogen is often accompanied by appearance of gene-for-gene vertical resistance, no useful sources of resistance to PSTVd infection have been identified in the cultivated potato.
- (3) Viroids and/or viroid-related RNAs closely related to Tomato planta macho viroid (TPMVd) and CCCVd have been detected in weeds and other wild vegetation growing near fields containing viroid-infected plants.

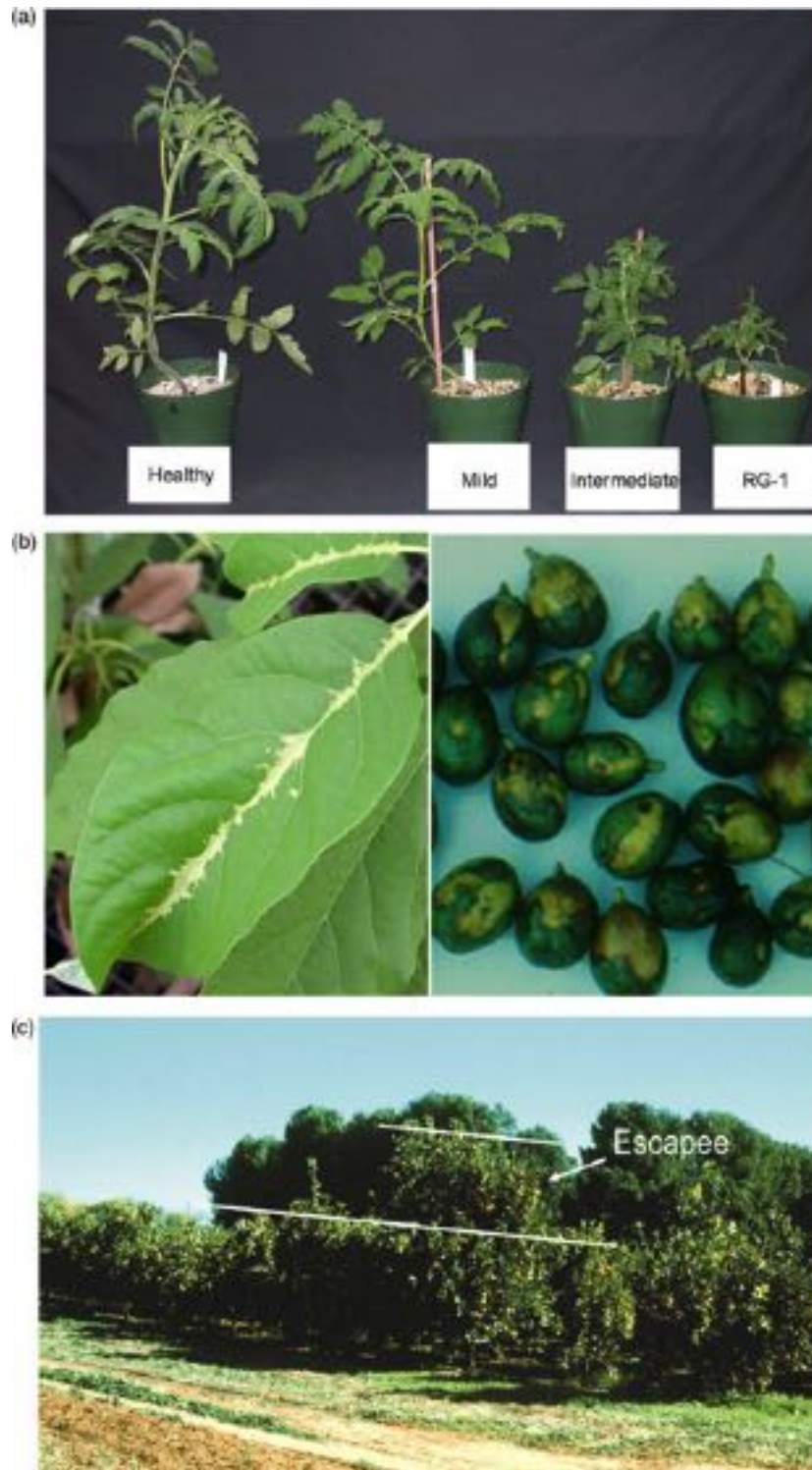


Fig. 3 (a) Symptoms of PSTVd infection in Rutgers tomato approximately four weeks after inoculation of cotyledons with PSTVd strains causing mild, intermediate, and severe symptoms. (b) Symptoms of ASBVd infection in avocado fruits and leaves. (c) CDVd-induced dwarfing of citrus growing on susceptible rootstocks. All trees in the block were graft-inoculated with CDVd shortly after transfer to the field; only one tree (right foreground) escaped infection. Note difference in height.

Growers and plant pathologists are unlikely to have simply overlooked diseases with symptoms as severe as those of chrysanthemum stunt or cucumber pale fruit, two diseases first reported after World War II. Large-scale monoculture of genetically identical crops and the commercial propagation/distribution of many cultivars are two comparatively modern developments which would facilitate the development of serious disease problems following the chance transfer of viroids from wild hosts to cultivated plants. Viroid diseases may also arise by transfer between cultivated crop species. For example, pears provide a latent reservoir for ASSVd; likewise, while HSVd infections of grapes are often symptomless, this viroid causes severe disease in hops. In both instances, the two crops are often grown in close proximity. In other instances, however, such proximity does not apparently exist, as in a hop disease induced by Citrus bark cracking viroid (CBCVd) recently reported in Slovenia, where citrus is not grown.

Because no useful sources of natural resistance to viroid diseases are known, diagnostic tests continue to play a key role in efforts to control these diseases. Since viroids lack a protein capsid, the antibody-based techniques used to detect many plant viruses are not applicable. Tests based upon their unique molecular properties have largely supplanted biological assays, which require extended periods (weeks to years) of time for completion and have difficulties in detecting mild or latent strains. Bioassays remain indispensable for assessing the autonomous replication of a candidate viroid-like RNA and, eventually, for fulfilling Koch's postulates. Several rapid (1–2 day) protocols involving polyacrylamide gel electrophoresis under denaturing conditions take advantage of the circular nature of viroids. Using these protocols, nanogram amounts of viroid can be unambiguously detected without the use of radioactive isotopes. In recent years, diagnostic procedures based upon nucleic acid hybridization or the polymerase chain reaction (PCR) have become widely used. The simplest methods involve the hybridization of a non-radioactively labeled viroid-complementary DNA or RNA probe to viroid samples that have been bound to a solid support followed by colorimetric or chemiluminescent detection of the resulting DNA-RNA or RNA-RNA hybrids. Such conventional "dot blot" assays can detect picogram amounts of viroids using clarified plant sap or tissue prints rather than purified nucleic acid as the viroid source. PCR-based protocols are finding increasing acceptance in those cases where either this level of sensitivity is not reached or a number of closely-related viroids are present in the same sample.

Molecular Biology

Although devoid of messenger RNA activity, viroids replicate autonomously and cause disease in a wide variety of plants. Much has been learned about the molecular biology of viroids and viroid-host interaction over the past 40 years, but the precise nature of the molecular signals involved remains elusive. A series of questions first posed by Diener summarizes the many gaps in our current understanding of the biological properties of these unusual molecules:

- (1) What molecular signals do viroids possess (and cellular RNAs evidently lack) that induce certain DNA-dependent RNA polymerases to accept them as templates for the synthesis of complementary RNA molecules?
- (2) What are the molecular mechanisms responsible for viroid replication? Are these mechanisms operative in uninfected cells? If so, what are their functions?
- (3) How do viroids induce disease? In the absence of viroid-specified proteins, disease must arise from direct interaction(s) of viroids (or viroid-derived RNA molecules) with host cell constituents. Infections by PSTVd or ASBVd induces RNA silencing (see below), but does this mechanism mediate pathogenesis for all viroids?
- (4) What determines viroid host range? Are viroids restricted to higher plants, or do they have counterparts in animals?
- (5) How did viroids originate?

Replication

A variety of multimeric plus and minus strand RNAs have been detected by nucleic acid hybridization in viroid-infected tissues. Based on their analysis, viroid replication has been proposed to proceed via a "rolling circle" mechanism that involves reiterative transcription of the incoming plus circular RNA to produce a minus strand RNA template. ASBVd and related viroids utilize a symmetric replication cycle in which the multimeric minus strand is cleaved to unit-length molecules and circularized before serving as template for the synthesis of multimeric plus strands. PSTVd and related viroids utilize an asymmetric cycle in which the multimeric minus strand is directly transcribed into multimeric plus strands. In both cases, the multimeric plus strands are cleaved to unit-length molecules and circularized (Fig. 4).

A diversity of host-encoded enzymes have been implicated in viroid replication. Low concentrations of α -amanitin specifically inhibit the synthesis of both PSTVd plus and minus strands in nuclei isolated from infected tomato, strongly suggesting the involvement of DNA-dependent RNA polymerase II, transcribing an RNA template, in the replication of PSTVd and related viroids. In nuclear extracts, synthesis of the PSTVd minus strand by RNA polymerase II starts in the left terminal loop of the plus strand template; furthermore, incubation of active replication complexes containing CEVd with a monoclonal antibody directed against the carboxy-terminal domain of RNA polymerase II results in the immunoprecipitation of both CEVd plus and minus strand RNAs. Mature PSTVd plus strands accumulate in the nucleolus and the nucleoplasm, while *in situ* hybridization indicates that minus strand RNAs are confined to the nucleoplasm. After some debate, the identity of the polymerase(s) responsible for replication of members of the family *Avsunviroidae* in the chloroplast appears now more certain. ASBVd synthesis is resistant to

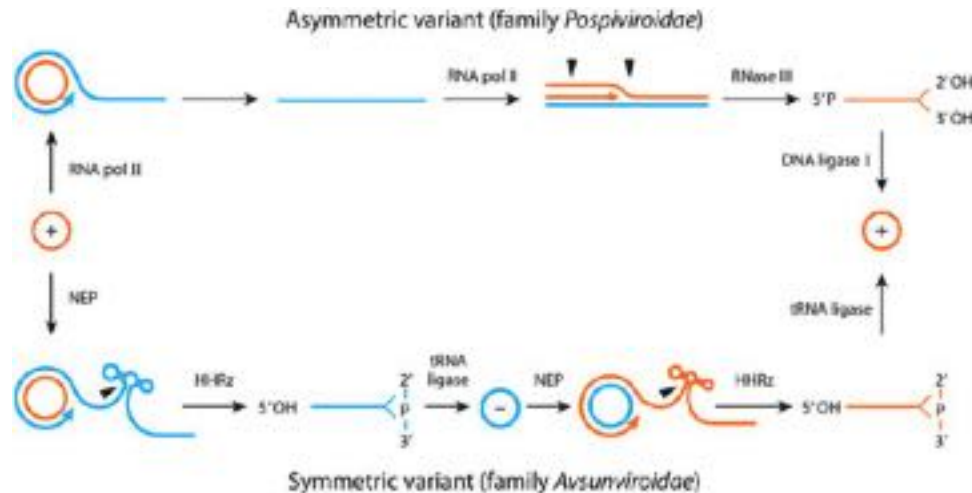


Fig. 4 Asymmetric and symmetric variants of the rolling-circle mechanism proposed for replication of members of the families *Pospiviroidae* and *Avsunviroidae*, respectively. Orange and blue colors refer to plus and minus polarities, respectively, with cleavage sites denoted by arrowheads. The enzymes and ribozymes that presumably catalyze the replication steps are indicated. Notice that RNA polymerase II (and NEP) is redirected to transcribe RNA templates, and DNA ligase 1 to circularize RNA substrates. Abbreviations: HHRz, hammerhead ribozyme; NEP, nuclear-encoded polymerase. Adapted from Flores, R., Gago-Zachert, S., Serra, P., Sanjuán, R., Elena, S.F., 2014. Viroids: survivors from the RNA world? *Annual Review of Microbiology* 68, 395–414, with permission.

tagetitoxin, strongly indicating the involvement of a nuclear-encoded chloroplastic DNA-dependent RNA polymerase (NEP) again transcribing an RNA template. Initiation sites for both ASBVd plus and minus strand synthesis have been mapped to the AU-rich terminal loops of their respective native structures. Furthermore, Peach latent mosaic viroid (PLMVd) replicates actively in albino leaf areas (where only NEP is active), with the initiation sites of plus and minus strands mapping at a short double-stranded RNA motif that also contains the self-cleavage sites of both polarity strands.

In vitro and *in vivo* evidence indicates that specific cleavage of multimeric plus strand RNAs of PSTVd and related viroids requires: (1) rearrangement of the conserved central region, and (2) the action of one or more host-encoded nucleases. Other less efficient processing sites may also be used *in vivo*. Multimeric plus and minus strand RNAs of ASBVd and related viroids, in contrast, undergo spontaneous self-cleavage co-transcriptionally through hammerhead ribozymes to form linear monomers. Addition of certain chloroplast proteins acting as RNA chaperones facilitates this hammerhead ribozyme-mediated self-cleavage reaction. The final step in viroid replication, the ligation of linear monomers to form mature circular progeny, is catalyzed by DNA ligase 1 (redirected to accept RNA substrates) in the case of PSTVd, and by a plastid isoform of the tRNA ligase in members of the *Avsunviroidae* (Fig. 4).

Movement

Upon entering a potential host cell, viroids must move to either the nucleus (*Pospiviroidae*) or chloroplast (*Avsunviroidae*) before beginning replication. Available data suggest that PSTVd enters the nucleus as a ribonucleoprotein complex formed by the interaction of cellular proteins with specific viroid sequence or structural motifs. VirP1, a bromodomain-containing protein isolated from tomato, has a nuclear localization signal and binds to the terminal right domain of PSTVd. Proteins such as TFIIA and ribosomal protein L5 that bind to the loop E motif may also be involved in viroid transport into the nucleus. How ASBVd or other members of the family *Avsunviroidae* enter and exit the chloroplast is currently unknown.

To establish a systemic infection, viroids must leave the initially infected cell – moving first from cell-to-cell through plasmodesmata and then long distances through the host vasculature. Upon injection into symplastically isolated guard cells in a mature tomato leaf, fluorescently-labeled PSTVd RNA does not move. Injection into interconnected mesophyll cells, in contrast, is followed by rapid cell-to-cell movement through the plasmodesmata. Long distance movement of viroids, like that of nearly all plant viruses, occurs in the phloem where it follows the typical source-to-sink pattern of photoassimilate transport. Viroid movement in the phloem almost certainly requires formation of a ribonucleoprotein complex, possibly involving a dimeric lectin known as Phloem Protein 2 (PP2), the most abundant protein in phloem exudate. Movement of PSTVd in the phloem appears to be sustained by replication in supporting cells and is tightly regulated by developmental and cellular factors. For example, *in situ* hybridization reveals the presence of PSTVd in vascular tissues underlying the shoot apical meristem of infected tomato, but entry into the shoot apical meristem itself appears to be blocked. Other important control points for PSTVd trafficking are the bundle sheath-mesophyll boundary and that between palisade and spongy mesophyll in the leaf. By disrupting normal pattern of viroid movement it may be possible to create a plant that is resistant/immune to viroid infection.

Pathogenicity

Sequence comparisons of naturally-occurring PSTVd and CEVd variants, as well as infectivity studies with chimeric viroids constructed by exchanging the pathogenicity domains of mild and severe strains of CEVd, have clearly shown that the pathogenicity domain in the family *Pospiviroidae* contains important determinants of symptom expression. In addition, symptom expression is also affected by the rate of viroid accumulation, and sequence changes in the variable domain have been shown to regulate progeny titers in infected plants. Studies with Tomato apical stunt viroid (TASVd) revealed the presence of a third pathogenicity determinant in the left terminal loop. Also, a single U/A change position 257 in the central domain of PSTVd results in the appearance of severe stunting and a “flat top” phenotype in tomato, and a single base pair in the right terminal domain of TPMVd is a virulence determinant factor. In the family *Avsunviroidae*, determinants of pathogenicity have been mapped to either a tetraloop capping a hairpin stem in Chrysanthemum chlorotic mottle viroid (CChMVd) or an insertion that folds into a hairpin also capped by a tetraloop in PLMVd.

The ability of novel viroid chimeras to replicate and move normally from cell-to-cell implies certain basic similarities between their structures *in vitro* and *in vivo* but provides no information about the nature of the molecular interactions responsible for symptom development. Until recently, it was widely assumed that the mature viroid RNA was the direct pathogenic effector. Just like viruses, however, viroid replication is also accompanied by the appearance of a variety of small (21–24 nt) RNA molecules produced by the RNA silencing defensive response of their hosts. The role of these viroid-derived small RNAs (vd-sRNAs) in viroid pathogenicity is not yet clear, but the inverse relationship between accumulation levels of the mature viroid RNAs and the corresponding vd-sRNAs for members of the family *Avsunviroidae* suggests that the latter may regulate the titer of the former. In addition, vd-sRNAs containing the pathogenicity determinants of specific PLMVd variants that incite albinism or an intense yellow mosaic target and direct the cleavage (as predicted by RNA silencing) of host mRNAs coding for proteins involved in chloroplast biogenesis. The corresponding PLMVd-sRNAs have 5'-terminal Us, implicating Argonaute 1 in what likely are the initial alterations eliciting distinct chloroses. Also, recovery of tomato plants from the symptoms of severe PSTVd infections is preceded by the accumulation of vd-sRNAs (Fig. 5).

Viroid infections are accompanied by quantitative changes in a variety of host-encoded proteins. Certain of these are “pathogenesis-related” proteins whose synthesis or activation is part of a general host reaction to biotic or abiotic stress, but others appear to be more specific. In tobacco, PSTVd infection results in the preferential phosphorylation of a host-encoded 68 kDa protein that is immunologically related to an interferon-inducible, dsRNA-dependent mammalian protein kinase of similar size. The human kinase is differentially activated by PSTVd strains of varying pathogenicity *in vitro*, while infection of tomato by intermediate or severe strains of PSTVd induces the synthesis of PKV, a dual-specificity, serine/threonine protein kinase. Broad changes in host gene expression following infection by PSTVd and CEVd have been detected by complementary DNA macroarray and proteomic analyses, respectively.

Implications for Host Range

Possibly as a result of its involvement in the cleavage/ligation of progeny RNA, nucleotides in the central domain of PSTVd and related viroids appears to play an important role in determining host range. For example, a single nucleotide substitution in the

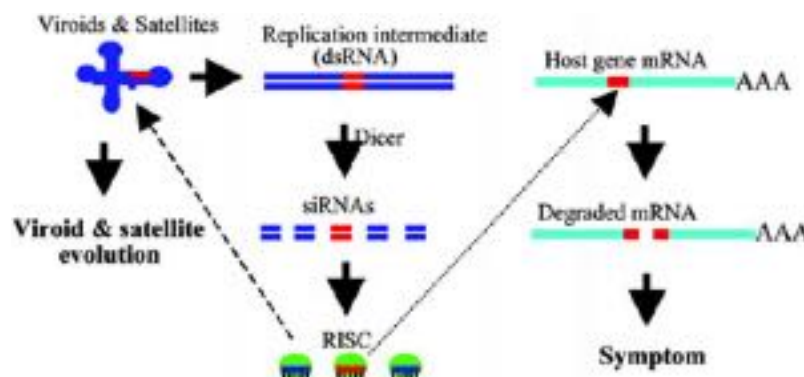


Fig. 5 A model for the role of RNA silencing in the pathogenicity and evolution of viroids and viral satellite RNAs. Replication of the subviral RNAs generates dsRNA intermediates that are processed by Dicer into 21- to 24-nt siRNAs (or vd-sRNAs), one strand of which are then incorporated into the RNA-induced silencing complex (RISC). If significant sequence identity exists between a region in the subviral RNA genome and a region in host gene mRNA (shown in red), RISC will target the host mRNA for cleavage leading to symptoms development. RISC can also target the subviral genome for degradation, forcing the subviral RNA to evolve and to adopt and maintain an RNA silencing-resistant secondary structure. From Wang, M.B., Bian, X.Y., Wu, L.M., *et al.*, 2004. On the role of RNA silencing in the pathogenicity and evolution of viroids and viral satellites. *Proceedings of the National Academy of Sciences of the United States of America* 101, 3275–3280, with permission, Copyright (2004) National Academy of Sciences, USA.

loop E motif results in a dramatic increase in the rate of PSTVd accumulation (and possibly replication) in tobacco. The biological properties of CLVd also suggest that this domain contains one or more host range determinants. CLVd appears to be a natural mosaic of sequences present in other viroids; phylogenetic analysis (Fig. 2) suggests that it can be considered to be a PSTVd-related viroid whose conserved central domain has been replaced by that of HSVd. Like HSVd (but not PSTVd or related viroids), CLVd can replicate and cause disease in cucumber.

Origin and Evolution

Much of the early speculation about viroid origin involved their possible origin as “escaped introns” (*i.e.*, descent from normal host RNAs). More recently, however, viroids have been proposed to represent “living fossils” of a precellular RNA world that assumed an intracellular mode of existence sometime after the evolution of cellular organisms. The presence of ribozymes in members of the *Avsunviroidae* strongly supports this view, although the recent characterization of circular long non-coding RNAs transcribed from DNA templates has rekindled the idea that viroids might have a host origin. No viroid is known to code for protein, a feature that is consistent with the possibility that viroids are phylogenetically older than introns.

The inherent stability of viroids and viroid-like satellite RNAs (structurally similar to viroids but functionally dependent on helper viruses) which arises from their small size and circularity would have enhanced the probability of their survival in primitive, error-prone RNA self-replicating systems and assured their complete replication without the need for initiation/termination signals. Some viroids (but not satellite RNAs or random sequences of the same base composition) also display structural periodicities with repeat units of 12, 60, or 80 nt. The high error rate of prebiotic replication systems may have favored the evolution of polyploid genomes, and the mechanism of viroid replication (*i.e.*, rolling-circle transcription of a circular template) provides an effective means of genome duplication.

Viroids and viroid-like satellite RNAs all possess efficient mechanisms for the precise cleavage of their oligomeric replication intermediates to form monomeric progeny. PSTVd and related viroids appear to require proteinaceous host factor(s) for cleavage, but others (members of the family *Avsunviroidae* and viroid-like satellite RNAs) contain ribozymes far smaller and simpler than those derived from introns. Thus, ASBVd and the other self-cleaving viroids may represent an evolutionary link between viroids and viroid-like satellite RNAs.

Phylogenetic evidence for an evolutionary link between viroids and other viroid-like subviral RNAs has been presented (Fig. 2). Among several subviral RNAs possibly related to viroids is carnation small viroid-like RNA, a 275 nt circular molecule with self-cleaving hammerhead structures in both its plus and minus strands that has a DNA counterpart. This retroviroid-like element shares certain features with both viroids and viroid-like satellite RNAs on the one hand, and with small RNA transcripts from newt and many other species on the other hand.

References

Diener, T.O., 1979. *Viroids and Viroid Diseases*. New York: Wiley Interscience.

Further Reading

- Diener, T.O., 2003. Discovering viroids: A personal perspective. *Nature Reviews Microbiology* 1, 75–80.
- Ding, B., 2009. The biology of viroid-host interactions. *Annual Review of Phytopathology* 47, 105–131.
- Ding, B., 2010. Viroids: Self-replicating, mobile, and fast-evolving noncoding regulatory RNAs. *Wiley Interdisciplinary Reviews-RNA* 1, 362–375.
- Flores, R., Gago-Zachert, S., Serra, P., Sanjuán, R., Elena, S.F., 2014. Viroids: Survivors from the RNA world? *Annual Review of Microbiology* 68, 395–414.
- Flores, R., Hernandez, C., Martinez de Alba, A.E., Daros, J.A., Di Serio, F., 2005. Viroids and viroid-host interactions. *Annual Review of Phytopathology* 43, 117–139.
- Hadidi, A., Flores, R., Randles, J.W., Palukaitis, P. (Eds.), 2017. *Viroids and Satellites*. London: Academic Press.
- Hull, R., 2013. *Plant Virology*, fifth ed. New York: Academic Press.
- Owens, R.A., Flores, R., Di Serio, F., *et al.*, 2011. Viroids. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy*. IX Report of the International Committee on Taxonomy of Viruses. London: Elsevier/Academic Press, pp. 1221–1234.

Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus (*Potyviridae*)

Cécile Desbiez and Hervé Lecoq, Plant Pathology Unit, INRAE – French National Research Institute for Agriculture, Food and Environment, Montfavet, France

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Cross-protection A plant systemically infected by a mild virus strain will not develop additional symptoms when inoculated by a severe strain of the same virus. Most often, the severe strain does not multiply in the cross-protected plant.

Filiform symptom A leaf deformation symptom in which the limb is drastically reduced but not the veins, giving a shoe strings aspect to the leaf.

Transmission propensity A measure of vector importance quantifying the natural ability of a species to inoculate a plant with a virus under conditions that allow vectors to move and feed freely.

History and Taxonomy

Watermelon Mosaic Virus

Mosaic diseases of cucurbit crops were first reported in the 1920s and the early literature contains a diversity of names for viruses or virus diseases that were only partially characterized. In 1940, a severe watermelon (*Citrullus lanatus*) mosaic disease was reported in California, but the nomenclature of the causal agent remained controversial until 1979. In 1965, ten isolates from Southern USA of what was then called the watermelon mosaic virus (WMV) complex were compared. Based on cross-protection experiments, serological relationships and host range reactions they were divided into two groups: WMV1 and WMV2. WMV1 and WMV2 were considered as different viruses, also different from a watermelon mosaic isolate reported from South Africa in 1960. Further work in 1969 increased the confusion by concluding that WMV1 and WMV2 were strains of a same virus. In 1979, the situation was definitively clarified by demonstrating that WMV1 and WMV2 were indeed serologically distinct entities, and that WMV1 was closely related to *Papaya ringspot virus* (PRSV). Now, WMV1 is considered as the W strain of PRSV (PRSV-W), while WMV2 is referred to as *Watermelon mosaic virus* (WMV). In addition, the same authors showed that a watermelon mosaic virus isolate from Morocco was a third serological entity. This isolate is considered as a distinct virus species, *Moroccan watermelon mosaic virus* (MWMV), to which probably also belongs the South African isolate.

So, from the initial watermelon mosaic virus complex emerged three different virus species: *Watermelon mosaic virus*, *Papaya ringspot virus* and *Moroccan watermelon mosaic virus*.

Zucchini Yellow Mosaic Virus

An apparently new cucurbit virus was isolated in 1973 from a zucchini squash (*Cucurbita pepo*) plant in Northern Italy, this virus was described as a new potyvirus species, *Zucchini yellow mosaic virus* (ZYMV). In 1979, many melon (*Cucumis melo*) crops were devastated in Southwestern France by an apparently new virus disease, whose causal agent was tentatively named muskmelon yellow stunt virus. Very rapidly it appeared that it was a strain of ZYMV. Within a few years, ZYMV was reported in many countries on the five continents. In this regards, ZYMV appears as a typical example of an emerging plant virus.

Classification

Based on particle morphology, aphid transmissibility, serological relationships, ability to induce pinwheel cytoplasmic inclusions in host cells, genome organization and nucleotide sequences, WMV and ZYMV were identified as member species of the genus *Potyvirus*, family *Potyviridae*. Potyviruses belonging to several other species can infect cucurbit crops including PRSV-W and related viruses: MWMV, *Zucchini yellow fleck virus* (ZYFV), *Zucchini tigre mosaic virus* (ZTMV), *Zucchini shoestring virus* (ZSTV), *Algerian watermelon mosaic virus* (AWMV), *Sudan watermelon mosaic virus* (SuWMV), *Wild melon vein banding virus* (WMVBV), *Melon vein banding mosaic virus* (MVBV), *Telfairia mosaic virus* (TeMV), *Cucurbita vein-banding virus* (CVBV), *Turnip mosaic virus* (TuMV), *Clover yellow vein virus* (CIYVV) and *Bean yellow mosaic virus* (BYMV). Except for PRSV-W, MWMV and ZTMV, these viruses have either only limited geographic distribution or minor economical incidence in cucurbit crops.

Molecular analyses based on the coat protein (CP) coding sequence revealed that cucurbit-infecting potyviruses belong to several “clusters” of closely related species: ZYMV, WMV and MVBV belong to a cluster that contains mostly legume-infecting potyviruses, whereas PRSV, MWMV and 6 other viruses are grouped in a “PRSV-like” cluster containing mostly cucurbit-restricted viruses (Fig. 1).

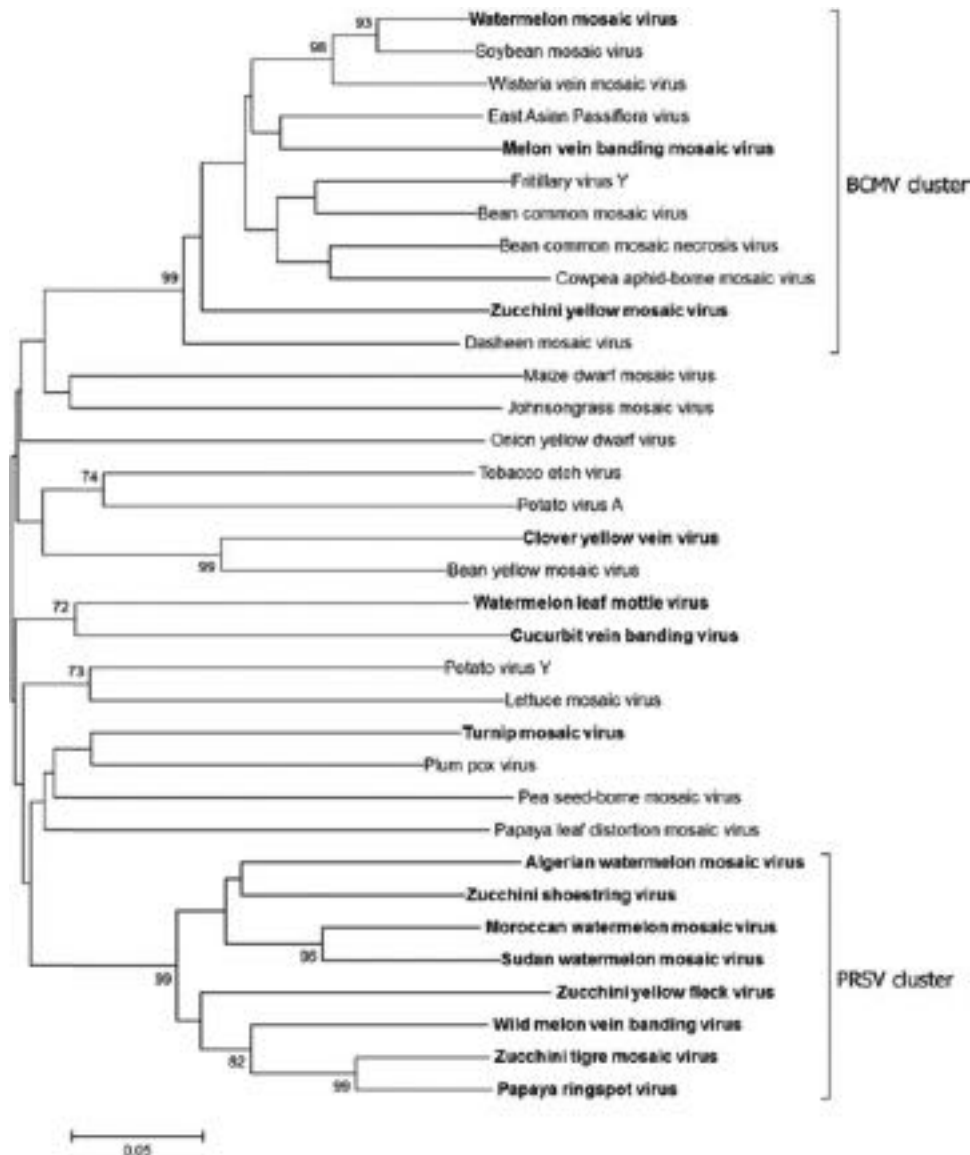


Fig. 1 Taxonomy, based on coat protein amino-acid sequences, of potyviruses including cucurbit-infecting ones (in bold). Branch lengths indicate the molecular divergence of viruses (the scale bar represents 0.05 mutations per residue). Figures at some nodes represent bootstrap values (in %), indicating the robustness of each node. Only values above 75% are indicated.

Symptomatology

Watermelon Mosaic Virus

WMV induces a diversity of symptoms depending on the isolate and the host cultivar. On leaves, symptoms are mosaics, vein banding, more or less severe leaf deformations and filiform symptom. On fruits, some isolates induce discoloration and greening on yellow fruit zucchini squash cultivars, while they will not affect fruit yield and quality in green fruit cultivars. Mosaic and discoloration are also observed on leaves and fruits of some melon cultivars (Fig. 2).

Zucchini Yellow Mosaic Virus

Since its first detection, ZYMV was recognized as a virus causing extremely severe symptoms leading to complete yield losses in the case of early contamination. This severe symptomatology was probably an important factor for the rapid identification of ZYMV soon after its first outbreaks in many countries. In melon, leaf symptoms include vein clearing, yellow mosaic, leaf deformation occasionally with blisters and enations, often with severe plant stunting. Some ZYMV isolates induce a rapid and complete wilt in



Fig. 2 Mosaic symptoms on a leaf and fruit of a melon plant infected by a moderately severe WMV isolate.



Fig. 3 Severe mosaic and deformations on leaves and fruits of a zucchini squash plant infected by ZYMV.

cultivars possessing the *Fn* gene. On fruits, a diversity of symptoms is observed: external mosaic or necrotic cracks, internal marbling and hardening of the flesh. Seeds are occasionally severely deformed and have poor germination rates. In zucchini squash, symptoms are very severe on leaves with mosaic, yellowing, leaf distortion and sometimes very severe filiform symptom. Fruits are generally severely misshaped with prominent knobs and are of course unmarketable (**Fig. 3**). In cucumber (*Cucumis sativus*) and watermelon, mosaic and deformations are generally observed on leaves and fruits.

Synergism and Antagonism

Synergism has been observed between WMV or ZYMV and other cucurbit-infecting viruses. This synergism can be expressed either by increase in virus multiplication rates or by more severe symptoms. Significant increase in Cucumber mosaic virus (CMV) multiplication rate was observed in cucumber plants co-infected by ZYMV. CMV could also partially overcome a resistance in cucumber when in mixed infection with ZYMV. In grafted cucumbers, double infection by CMV and ZYMV induces a severe and rapid wilting reaction, not observed in single infection by CMV or ZYMV. ZYMV was shown to facilitate the entry of CMV in cucumber xylem in co-infection. When WMV or ZYMV are in mixed infection with the polerovirus Cucurbit aphid-borne yellows virus (CABYV), CABYV multiplication rate and symptom intensity are increased. As for other potyviruses, these synergetic effects could be related to the strong silencing suppressor activity of WMV and ZYMV HC-Pro.

In mixed infections of ZYMV and WMV, the multiplication rate of ZYMV is not affected whereas WMV accumulation is significantly reduced, but the virus is still readily transmitted by aphids. In field conditions, early epidemics of ZYMV seem to hinder WMV spread.

Infection of ZYMV limits establishment and severity of powdery mildew infection in wild squash populations, as well as the incidence of bacterial wilt caused by *Erwinia tracheiphila*. The lower bacterial infection appears related to a reduction of floral volatile emissions in ZYMV-infected plants, resulting in a lower attractiveness for the cucumber beetles that transmit the bacteria.

Geographic Distribution

Both WMV and ZYMV are now widely distributed in the major cucurbit production areas worldwide. Their geographic distributions are broadly overlapping and frequent mixed infections are observed in the fields.

For unknown reasons, WMV appear to be rare or absent in cucurbits in subtropical or tropical areas. For instance, in exhaustive surveys conducted in Nepal, Sudan and French Western Indies no WMV was detected, although ZYMV was relatively abundant. In Florida, WMV is frequent in Northern and Central counties but is not detected in Southern counties while ZYMV can be found throughout the state. This cannot be related to a lack of potential WMV vectors or reservoirs because both are abundant in tropical or subtropical areas.

ZYMV is present worldwide in almost all countries where cucurbits are grown, under temperate, Mediterranean, subtropical and tropical climatic conditions. It affects highly mechanized cropping systems – such a glasshouse crop production in Northern Europe – as well as more traditional agroecosystems – such as flood irrigated crops on the Nile banks. ZYMV has been reported in very remote areas including semi-desertic regions or islands.

Host Range

WMV has a relatively wide experimental host range for a potyvirus. It infects over 170 species in 26 mono- or dicotyledonous families. Besides cucurbits, WMV causes mosaic diseases in legumes (pea, bean) and orchids (vanilla, *Habenaria radiata*). It has been described on many cultivated and ornamental hosts, including *Panax ginseng* and *Lagerstroemia indica*. It also infects many weeds, either annual or perennial, that can serve as alternative hosts. Generally, naturally infected weeds do not present evident symptoms of viral infection.

ZYMV has a relatively narrow host range. In natural conditions, it infects mostly cultivated or wild cucurbits but also a few flower species (*Consolida ajacis*, *Althea rosea*, *Begonia sempervirens*) or weeds.

Diagnostic Methods

The confusion that was prevalent in the early descriptions of watermelon mosaic diseases was mainly due to the convergence of symptoms caused by WMV, PRSV and MWMV in cucurbits and to the lack of proper diagnostic tools. Symptomatology and virus particle morphology were clearly insufficient to differentiate these viruses. The production of specific polyclonal antisera, and the development of simple serological tests, such as the gel double-diffusion test in agar containing sodium dodecyl sulphate and sodium azide (SDS-ID), brought a major contribution to the proper diagnosis of cucurbit potyviruses. In particular, this method contributed to the rapid and unequivocal identification of ZYMV in several countries soon after its first observation. Now, DAS-ELISA is generalized and commercial kits are available for WMV and ZYMV. Commercial dip-stick serological tests based on the lateral flow technique have also been developed that allow an easy and rapid diagnosis of ZYMV in the fields.

Monoclonal antibodies (MAbs) have been produced for ZYMV and WMV. They proved to be very useful to study the serological variability and to differentiate ZYMV and WMV subgroups. They have also been used successfully to analyze virus interactions, and in particular cross protection efficiency and specificity. Nanobodies, i.e., camel-derived variable domains of the heavy chain antibodies, have also been obtained against ZYMV CP.

Many nucleotide sequences are now available for ZYMV and WMV, particularly in the coat protein (CP) coding region, what allowed the development of specific primers for each virus. Loop-mediated isothermal amplification (LAMP) has also been developed for the diagnostic of WMV and ZYMV.

Vector Relationships

WMV is transmitted by at least 35 aphid species in 19 genera. Fewer aphid species were tested for their ability to transmit ZYMV, and 11 were identified as ZYMV vectors. *Aphis craccivora*, *Aphis gossypii*, *Macrosiphum euphorbiae* and *Myzus persicae* are efficient WMV and ZYMV vectors. Some aphid species were shown to be poor or non-vectors of WMV and ZYMV what suggests some level of specificity in the virus-vector interaction.

WMV and ZYMV are transmitted on the non-persistent mode: they are acquired and transmitted during very short probes (a few seconds to minutes), and their retention period in the vector is relatively short (a few hours).

WMV and ZYMV as typical potyviruses require the presence of a virus-encoded helper component (HC-Pro) protein for transmission. HC-Pro from WMV and ZYMV are interchangeable and both mediate efficiently the transmission of purified virions of both species.

Several ZYMV isolates that have lost aphid transmissibility have been characterized, and a unique feature for this virus is that single amino acid mutants have been identified in the three domains important for transmission. ZYMV-NAT has a A to T substitution in the DAG motif in the CP, ZYMV-PAT a T to A substitution in the PTK motif and ZYMV-R1A a K to E substitution in

the KLSC motif, both in the HC-Pro. These mutants led to the identification of an interaction between the HC-Pro and CP through their PTK and DAG domains.

The non-transmissible isolate ZYMV-NAT (having the DTG motif in the CP) could be transmitted by aphids from plants infected concomitantly by a transmissible isolate of PRSV. This occurred through hetero-encapsidation, a phenomenon by which ZYMV RNA is completely or partially encapsidated by the functional PRSV CP. An aphid non-transmissible isolate deficient for the HC-Pro can also be transmitted by aphids when in mixed infection with an isolate that has a functional HC-Pro. The transmissible isolate provides its functional HC-Pro to mediate the transmission of the deficient isolate. These two mechanisms can contribute to the maintenance, in natural conditions, of variants which have lost their vector transmissibility.

An interesting interaction has been observed between ZYMV and *A. gossypii*, an aphid vector colonizing cucurbit crops. *A. gossypii* lives longer and produce more offspring on ZYMV infected than on non-infected plants. In addition, more alatae are produced on infected plants, which may stimulate ZYMV spread. These phenomena might be related to the observed changes in phloem exudates composition (free amino acids, sugars) in virus infected plants.

Epidemiology

Virus Sources

WMV has not been described as seed borne in cucurbits or other crops, but may be transmitted through vegetative propagation in vanilla. ZYMV seed transmission has long remained controversial, with estimated rates ranging from 0% to 18%. Seed transmission was observed in naked seed pumpkin. Using RT-PCR detection, a vertical transmission rate of 1.6% was observed in *C. pepo* subsp. *texana*. The vertically-infected plants were symptomless and had low virus titers, but they could constitute sources for further horizontal transmission, either mechanically or by aphids. Pollen transmission occurred at a rate of only 0.13%, indicating that most seed transmission is *via* the ovule.

Another possible way for long distance dissemination of ZYMV is through the globalization of vegetable production and trade. It has been shown that ZYMV infected fruits imported from Central America into Europe could be very efficient virus sources for aphids.

Virus sources from which epidemics will initiate could be infected overwintering weeds or crops. In tropical and sub-tropical regions, cucurbit crops or weeds are growing all year round, and viruses could easily move from an old infected crop or cucurbit weed to a young planting. In more temperate regions non-cucurbit weeds were found to be efficient reservoirs for WMV but not for ZYMV. However, a few ornamental plants such as *Begonia sempervirens* were shown to be involved in ZYMV overwintering in Southern France. Winter protected crops could contribute to ZYMV overwintering in Mediterranean regions and residential gardens were found important sources of ZYMV in California.

Efficient Vectors

Many potential aphid vector species have been identified for WMV and ZYMV. A study was conducted to compare the vector capacity of two aphid species, one colonizing cucurbits (*A. gossypii*), the other not (*A. craccivora*). Two parameters were used. Transmission efficiency was measured in the laboratory with single aphids exposed in sequence to an infected plant and then to four healthy plants. Transmission propensity was measured by arena tests (more representative of natural conditions) in which aphids could move between plants and feed without interference. It was shown that *A. craccivora* had both a higher efficiency and propensity to disseminate ZYMV than *A. gossypii*. This highlights the importance of non-colonizing transient vector species in the epidemiology of ZYMV. In addition to vector transmission, ZYMV transmission has been shown to occur in zucchini squash either through direct contact or trampling on leaves of neighboring plants or through tools used for collecting fruit.

Pattern of Spread

The same general pattern is observed for WMV and ZYMV. The first contaminations occur generally shortly after planting, depending upon the availability of virus sources in the environment. Then, aphid flights, particularly of non-colonizing species, spread the virus from the primary infection foci to the rest of the crop, what corresponds to the secondary virus spread. The diseased plants are often distributed in large patches that rapidly extend and join each others, leading to the complete contamination of the crop within a few weeks. Simultaneously, the infected crop will serve as a source of virus to contaminate weeds or nearby young plantings. Epidemic development curves have an overall S shape, fitting generally well to the logistic model.

Control Methods

Prophylactic Measures

Prophylactic measures are intended to prevent or limit the contact of viruliferous aphids with cultivated plants. They are not specific for a particular virus and are generally efficient for all aphid-borne viruses. These include careful weeding near plantings

and avoiding overlapping crops in the same area to reduce virus and aphid sources near new plantings. When possible, manipulation of planting dates can avoid exposing young plants to peaks of vector populations. Plastic mulches have a repelling action on aphids and significantly delay WMV and ZYMV spread. However, they confer only a temporary protection that is limited to the early stages of the crop, because their efficiency decreases when plant growth covers their surface. Row covers of different types (unwoven, perforated plastics...) can also be used; they physically prevent winged aphids from reaching the plants, but they must be removed to allow insect pollination necessary for cucurbits. Both methods have a major drawback: they require a lot of plastic material that farmers must dispose in an ecologically sound way after the crop cycle. Tall non-host barriers (millet, *Pennisetum glaucum*) planted upwind of potential infection sources also contribute to reducing virus infection. Insecticides applications have been generally found inefficient in limiting WMV and ZYMV spread. This is to relate to the large number of winged aphids that land on the plants and to the rapidity of the transmission process. Oil applications can delay virus spread when inoculum pressure is moderate. When applicable, a one-month crop-free period has been shown to be efficient in limiting ZYMV spread. On-farm hygiene practices such as decontamination of tools are also recommended to limit virus spread through contact transmission.

Cross-Protection

Cross-protection has been developed at a commercial level to protect cucurbit crops against ZYMV. The principle is simple: when a mild virus isolate (i.e., that has no significant impact on commercial yield) is inoculated to young seedlings, it protects the plant from subsequent contaminations by severe isolates of the same virus. The mechanism appears related to virus-induced gene silencing. The mild strain ZYMV-WK is a natural variant of a severe aphid non-transmissible isolate. Although efficient against most ZYMV isolates (Fig. 4), ZYMV-WK does not protect against very divergent isolates such as those from Réunion Island indicating some specificity in the protection. A single amino acid change (R to I) in the FRNK conserved domain of the HC-Pro is responsible for symptom attenuation of ZYMV-WK. A complete technological package (mild strain production, quality control protocols, and inoculation machines) has been developed to implement commercially ZYMV cross-protection. Mild WMV isolates have been reported that could also have a potential for cross-protection but they have not been used commercially.

Resistant Cultivars

The use of virus-resistant cultivars is probably the easiest and cheapest way to control plant viral diseases at the farmer's level. Breeding for resistance still mainly relies upon search for resistance characters in germplasm collections and introgression of the resistance gene(s) into commercially acceptable cultivars. Considerable efforts have been made to look for resistance to WMV and ZYMV in genetic resources and some WMV- or ZYMV-resistant commercial cultivars are now available. Some resistance genes confer complete and durable resistance (such as the recessive *zym* gene in cucumber) while others confer only partial resistances or may be overcome by virus evolution (such as the dominant *Zym* gene in melon). An interesting situation is observed for ZYMV resistance in squash. Although the resistance level was high in the original accession of *Cucurbita moschata* in which the resistance was identified, when transferred through interspecific crosses to zucchini squash (*C. pepo*), the resistance phenotype was different: ZYMV multiplies but the plants present only very mild symptoms (this is called tolerance). However, tolerance appears not to be stable since aggressive variants of the virus (i.e., causing severe symptoms in tolerant plants) may emerge in these plants. A single amino acid change in the P3 gene is sufficient to confer this aggressive phenotype. However, the aggressive variants are counter-selected when in competition with common ZYMV isolates in susceptible cucurbits. This genetic load associated with aggressiveness could be a factor that will make the tolerance durable. In melon, a resistance to WMV and ZYMV transmission by



Fig. 4 Cross-protected zucchini crop in greenhouse conditions. Left: unprotected plants infected with a severe strain of ZYMV; right: plants preventively inoculated with the mild strain ZYMV-WK, protected against infection with severe strains.

A. gossypii was found to be governed by the single dominant gene *Vat*. This gene is present in many commercial cultivars, but confers only limited protection in the fields, probably because WMV and ZYMV are also transmitted by many other aphid species in natural conditions.

In the last three decades, attempts were made to obtain WMV and ZYMV transgenic resistant plants using the pathogen derived resistance approach. Different constructs were tested to obtain resistant WMV or ZYMV plants: untranslatable or full-length CP gene, ribozymes. Freedom II, a transgenic squash hybrid containing the WMV and ZYMV CP genes was released in the USA in 1995, as the first virus-resistant transgenic crop to be commercially cultivated in the world. It proved to have a very efficient resistance to WMV and ZYMV in field conditions. Similar hybrids are presently grown mainly in South Eastern USA, particularly during late summer and fall when WMV and ZYMV inoculum pressure are high. Surprisingly, protection against ZYMV infection in transgenic squash presented indirect costs: higher predation by beetles and higher incidence of bacterial wilt.

Exogenously applied dsRNA molecules deriving from the ZYMV genome conferred variable levels of resistance to ZYMV in experimental conditions: protection up to 82% was achieved in cucumber using a 588 bp fragment of the HC-pro. Inoculation of cucumber with an apple latent spherical virus vector harboring part of ZYMV genome prevented subsequent experimental infection by ZYMV, and also had a curative effect on ZYMV-infected plants. Recently, melons and cucumbers either silenced or mutated in the eukaryotic translation initiation factor 4E (eIF4E) that is required for potyvirus infection, were found to be resistant to several viruses including ZYMV.

Variability

Only limited biological variability has been reported for WMV. This concerns mainly differences in symptom intensity, host range or aphid transmissibility. Among the few host range specificities, the ability or not to infect systemically *Chenopodium quinoa* is related to a point mutation in WMV CP.

In contrast, from its first description, ZYMV appeared to have a very important biological diversity in host range and symptomatology on susceptible hosts, with isolates producing mild or atypical mosaic symptoms, necrosis or wilting reactions. Different pathotypes could be differentiated on melon or squash varieties possessing resistance genes. Important variability has also been observed in aphid transmissibility and several aphid non- or poorly-transmissible isolates have been described.

Limited serological variability was observed in WMV and ZYMV when polyclonal antibodies were used. However, the development of monoclonal antibodies against WMV and ZYMV allowed the characterization of serotypes closely correlated to the molecular variability.

The molecular variability of ZYMV and WMV was assessed first on the CP region, particularly the N-terminal part of the CP that is known for potyviruses to be highly variable and is frequently used for molecular studies. Besides, more than 50 complete sequences are now available for both ZYMV and WMV. For ZYMV, 3 major molecular groups were defined. Groups B and C contain isolates from specific geographic origins (Reunion Island, Singapore, Vietnam, Australia and East Timor for group B; China, Vietnam and Poland for group C) whereas group A contains worldwide isolates, including the first ZYMV isolates described in the 1970s in Europe. Within group A, 6–7 subgroups were defined without strong geographic structure (Fig. 5). Molecular analyses have allowed, in a few cases, tracking the putative origins of ZYMV strains emerging in a new area.

For WMV, the first studies indicated the presence of 3 major molecular groups more or less correlated with the geographic origin of the isolates: Group 1 in Europe, the Mediterranean Basin and part of Asia, Group 2 in South America and the Pacific area and Group 3 in Eastern Asia. With the increase of the number of sequences available, the situation has become more complex (Fig. 5), and evidence for long-distance exchanges of genetic diversity was observed, sometimes followed by strain replacement as observed in several European countries in the 2000s.

Genetics and Evolution

WMV and ZYMV genome organization is very similar to that of other potyviruses sequenced so far. The single-stranded, positive-sense RNA genome (9.6 kb for ZYMV, 10 kb for WMV excluding the polyadenylated 3' extremity) is translated as a single polyprotein that is self-cleaved in 10 functional proteins, with an 11th protein expressed through replication slippage in the P3 coding region.

The full-length sequence revealed that WMV is very closely related to the legume-infecting soybean mosaic virus (SMV); however, the P1 protein is 135 amino-acids longer than that of SMV, and the N-terminal half of the P1 shows no relation to SMV but is 85% identical to another legume-infecting potyvirus, bean common mosaic virus (BCMV). This suggests that WMV has emerged through an ancestral recombination event between a SMV-like and a BCMV-like potyvirus. SMV and BCMV have a narrow host range, mostly restricted to legumes, while WMV is one of the potyviruses with the broadest host range, including monocots and dicots. The impact of its recombinant nature on WMV biological properties remains unknown. Some SMV isolates from China have the same recombination with BCMV in the 5' part of the genome as WMV but they did not infect watermelon even though, contrary to regular SMV isolates, they could infect common bean.

Partial and complete sequence data indicate the frequent presence of intraspecific recombinants in WMV. By sequencing the full-length genome of putative recombinants, recombination points were characterized in different parts of the genome, with a

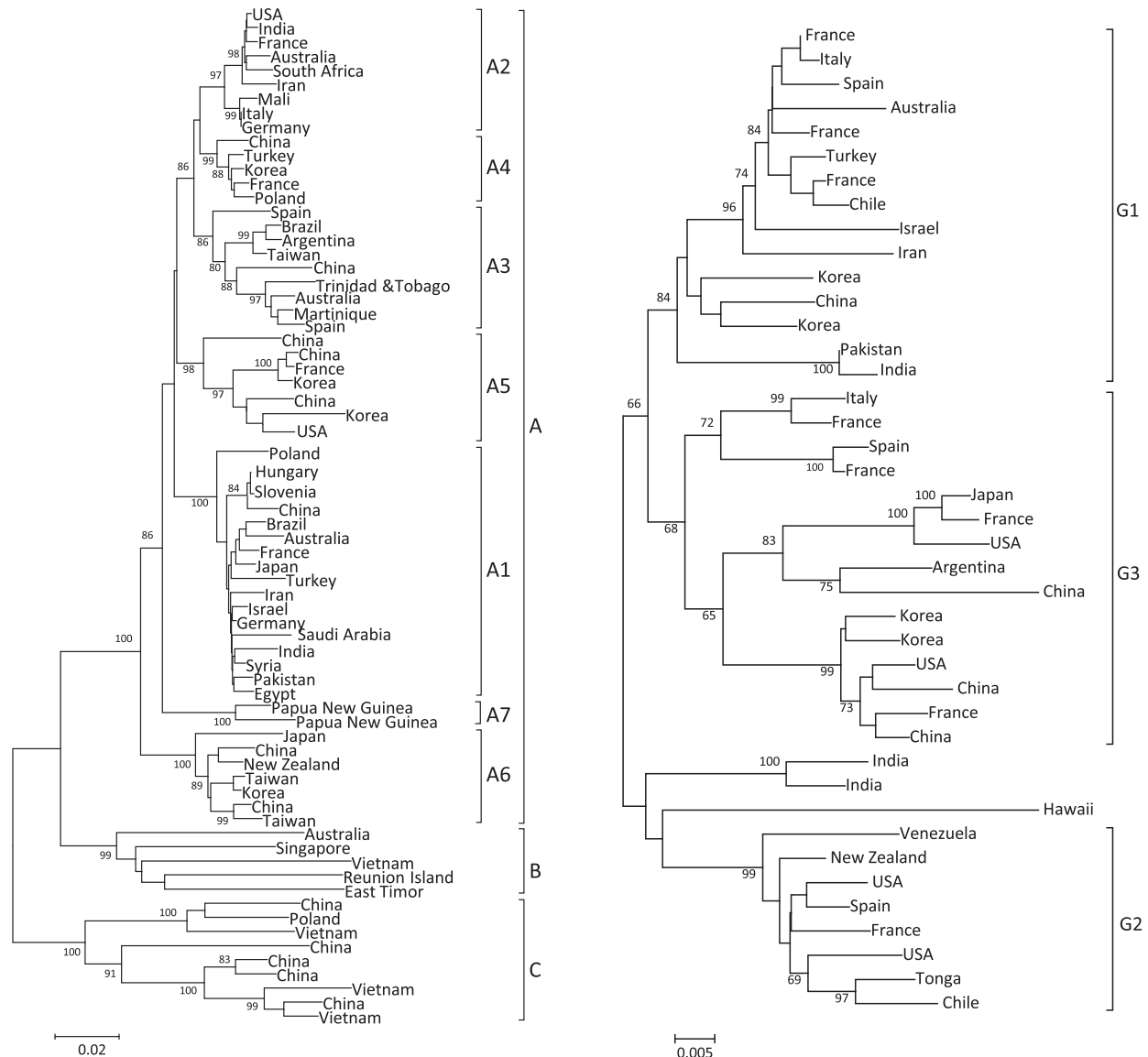


Fig. 5 Distance tree based on the nucleotide sequence of the coat protein coding region of ZYMV (left) and WMV (right), and showing the main molecular groups currently acknowledged for these viruses. Branch lengths indicate the molecular divergence between sequences (the scale bar represents 0.02 and 0.005 mutations per site for ZYMV and WMV respectively). Figures at some nodes represent bootstrap values (in %), indicating the robustness of each node.

hotspot in the P3-CI region. In France, recombinants between Groups 1 and 3 were observed in the few years following the introduction of Group 3, but they did not become prevalent in populations.

In the case of ZYMV, sequence analysis software suggest that recombination might have taken place, but the situation is not as clear cut as for WMV.

Besides recombination, viral genomes also evolve by mutations that take place during replication. In the case of ZYMV, the biological consequences of several point mutations that emerged in natural conditions were assessed by molecular studies: sequences of closely related strains with different biological properties were compared, and the mutations observed – particularly non-silent mutations located in conserved domains – were introduced by site-directed mutagenesis in an infectious cDNA of ZYMV in order to check if the mutation alone is sufficient to induce the difference in biological properties.

Molecular studies have thus shown that single point mutations in HC-Pro or P3 protein of ZYMV were important for symptomatology. Similarly, single mutations in HC-Pro or CP greatly affect aphid transmissibility. For unknown reasons, some ZYMV isolates such as ZYMV-E15 seems to be particularly prone to mutations and many variants (mild, aphid non-transmissible, aggressive or virulent) have been derived from this isolate.

Biotechnological Application

The development of infectious cDNA clones of ZYMV brought the possibility of using ZYMV as a biotechnological tool to produce proteins of pharmaceutical or crop protection interest. The gene coding for the protein of interest can be inserted in the ZYMV genome either between the P1 and HC-Pro genes or between the NIb and CP genes. For optimal protein production, it is possible to use clones that have the mutation in the FRNK domain so that they produce mild symptoms and do not affect plant growth. The protein may be produced in the edible cucurbit fruits that could be used directly for oral administration. Several molecules of pharmaceutical interest have been efficiently produced through this technology: the human interferon-alpha 2, antiviral and antitumor proteins MAP30 and GAP31 and a mite allergen. This technology also provided a way to produce large amounts of nucleocapsid proteins of 5 tospoviruses that could be used for immunization and production of specific antibodies. Finally, the *bar* gene coding for a phosphinothricin acetyltransferase that confer resistance to glufosinate ammonium based herbicides has been inserted into the mild ZYMV expression vector. In weed infested plots, this construct efficiently protected inoculated zucchini squash plants from damage caused by an herbicide treatment that completely destroyed the weeds. In addition, these plants were protected against severe ZYMV isolates. It is interesting to see that ZYMV, which emerged as the most damaging cucurbit virus in the last decades, can be manipulated and used for the benefit of human health and agriculture.

Further Reading

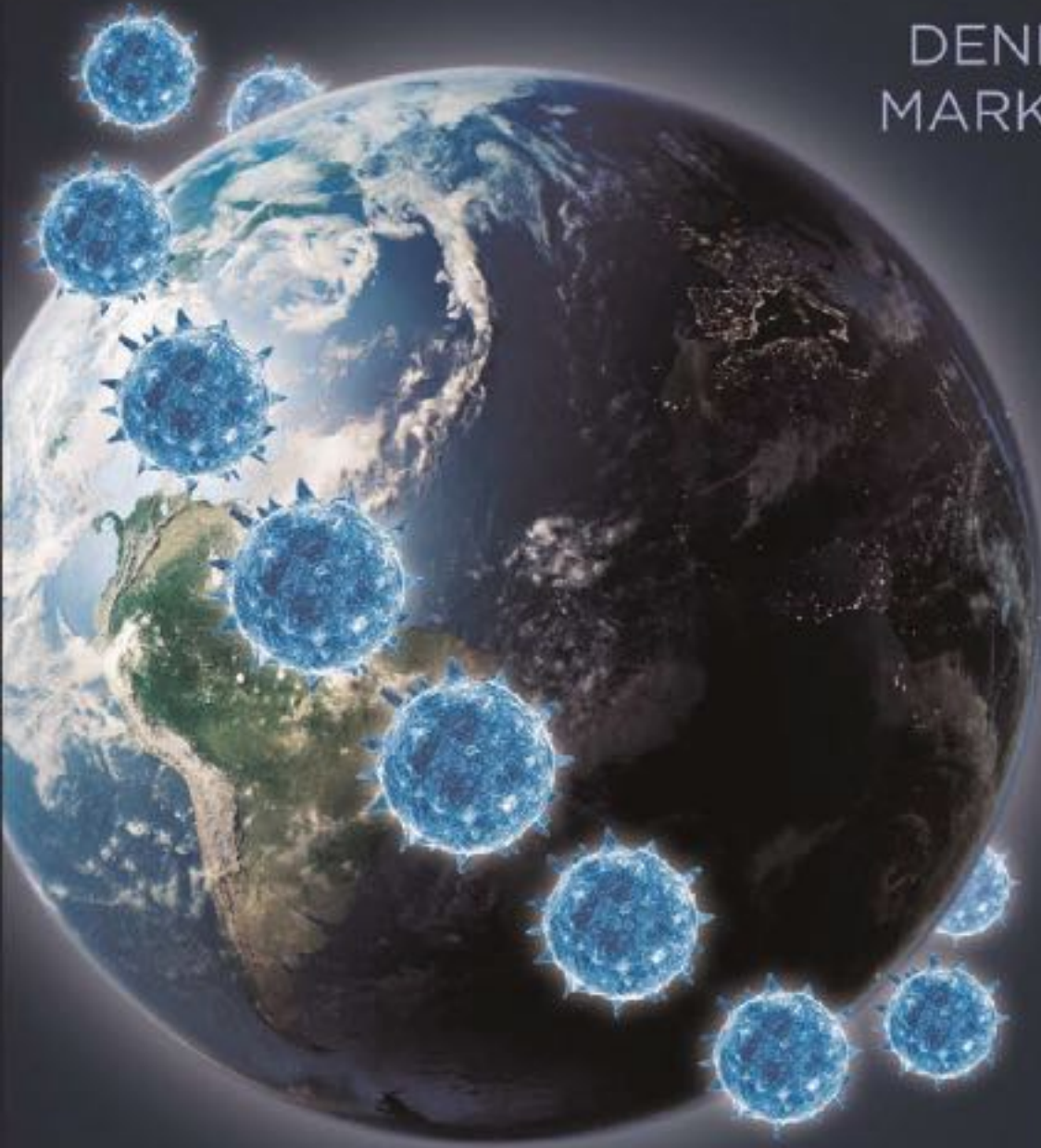
- Arazi, T., Slutsky, S.G., Shibolet, Y.M., *et al.*, 2001. Engineering Zucchini yellow mosaic potyvirus as a non-pathogenic vector for expression of heterologous proteins in cucurbits. *Journal of Biotechnology* 87, 67–82.
- Coutts, B.A., Kehoe, M.A., Jones, R.A.C., 2011. Minimizing losses caused by Zucchini yellow mosaic virus in vegetable cucurbit crops in tropical, sub-tropical and Mediterranean environments through cultural methods and host resistance. *Virus Research* 159, 141–160.
- Coutts, B.A., Kehoe, M.A., Jones, R.A.C., 2013. Zucchini yellow mosaic virus: Contact transmission, stability on surfaces, and inactivation with disinfectants. *Plant Disease* 97, 765–771.
- Coutts, B.A., Kehoe, M.A., Webster, C.G., Wylie, S.J., Jones, R.A.C., 2011. Zucchini yellow mosaic virus: Biological properties, detection procedures and comparison of coat protein gene sequences. *Archives of Virology* 156, 2119–2131.
- Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C., Lecoq, H., 2011. Recombination in natural populations of watermelon mosaic virus: New agronomic threat or damp squib? *Journal of General Virology* 92, 1939–1948.
- Desbiez, C., Lecoq, H., 1997. Zucchini yellow mosaic virus. *Plant Pathology* 46, 809–829.
- Desbiez, C., Lecoq, H., 2004. The nucleotide sequence of watermelon mosaic virus (WMV, *Potyvirus*) reveals interspecific recombination between two related potyviruses in the 5' part of the genome. *Archives of Virology* 149, 1619–1632.
- Jones, R.A.C., 2009. Plant virus emergence and evolution: Origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Research* 141, 113–130.
- Lecoq, H., 1998. Control of plant virus diseases by cross protection. In: Hadidi, A., Khetarpal, R.K., Koganezawa, H. (Eds.), *Plant Virus Disease Control*. St Paul, Minnesota: APS Press, pp. 33–40.
- Lecoq, H., Desbiez, C., 2012. Viruses of cucurbit crops in the Mediterranean region: An ever-changing picture. *Advances in Virus Research* 84, 67–126.
- Lecoq, H., Wipf-Scheibel, C., Nozeran, K., Millot, P., Desbiez, C., 2014. Comparative molecular epidemiology provides new insights into Zucchini yellow mosaic virus occurrence in France. *Virus Research* 186, 135–143.
- Lisa, V., Lecoq, H., 1984. Zucchini yellow mosaic virus. CMI/AAB Descriptions of Plant Viruses No. 282. Kew, Surrey: Commonwealth Mycological Institute.
- Purcifull, D., Hiebert, E., Edwardson, J., 1984. Watermelon mosaic virus 2. CMI/AAB Descriptions of Plant Viruses No. 293. Kew, Surrey: Commonwealth Mycological Institute.

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Editors in Chief

DENNIS BAMFORD
MARK ZUCKERMAN



ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Volume 4

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

EDITORS IN CHIEF

Dennis H. Bamford

*Molecular and Integrative Biosciences Research Programme
Faculty of Biological and Environmental Sciences
University of Helsinki, Helsinki, Finland*

Mark Zuckerman

*South London Specialist Virology Centre
King's College Hospital NHS Foundation Trust
London, United Kingdom*

and

*Department of Infectious Diseases
School of Immunology and Microbial Sciences, King's College London Medical School
London, United Kingdom*

Volume 4



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2021 Elsevier Ltd. unless otherwise stated. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-12-814515-9

For information on all publications visit our
website at <http://store.elsevier.com>

		Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org		

Publisher: Oliver Walter

Acquisitions Editor: Priscilla Braglia

Content Project Manager: Katarzyna Miklaszewska

Associate Content Project Manager: Gayathri S.

Designer: Matthew Limbert

EDITORS IN CHIEF



Dennis H. Bamford, PhD, is Professor Emeritus of Virology at the Faculty of Biological and Environmental Sciences, University of Helsinki, Finland. He obtained his PhD in 1980 from the Department of Genetics, University of Helsinki. During 1981–1982 he was an EMBO postdoctoral fellow at the Public Health Research Institute of the City of New York, United States, and during 1983–1992 he worked as a Senior Scientist at the Academy of Finland. In 1993 he was appointed Professor of General Microbiology at the University of Helsinki. He was awarded the esteemed Academy Professorship twice, in 2002–2007 and 2012–2016, and he also served twice as the Director of the Finnish Center of Excellence (in Structural Virology, 2000–2005, and in Virus Research, 2006–2011). Prof. Bamford has had continuous external research funding (e.g., from several European Union, Academy of Finland, TEKES and Jusélius Foundation funds, as well as the Human Frontier Science Program). He is an EMBO member and has held several positions of trust in scientific and administrative organizations.

Prof. Bamford has published approx. 400 articles in international peer-reviewed journals in virology, microbiology, biochemistry, and molecular biology (36 of them in high impact journals). About half of the primary articles have been published with international collaborators showing high international integration. He has also been invited to give 56 keynote and plenary presentations in major international meetings. Prof. Bamford has supervised over 35 Master's and

over 40 PhD theses. Seven of his graduate students or post docs have obtained a professorship and a similar number have a principal investigator status. Prof. Bamford has studied virus evolution from a structure-centered perspective, showing that seemingly unrelated viruses, such as bacteriophage PRD1 and human adenovirus have similar virion architecture. When the corona virion architecture was gradually revealed, it was observed that its structural elements were close to those seen in RNA bacteriophage phi6 so that phi6 has been actively used as surrogate for pathogenic viruses - quite a surprise!



Dr. Mark Zuckerman is Head of Virology, Consultant Medical Virologist, and Honorary Senior Lecturer at South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust and King's College London Medical School, Department of Infectious Diseases, School of Immunology and Microbial Sciences in London, United Kingdom. His interests include the clinical interface between developing molecular diagnostic tests relevant to the local population of patients, respiratory virus infections, herpesvirus infections in immunocompromised patients and blood-borne virus transmission incidents in the healthcare setting. He has chaired the UK Clinical Virology Network, Royal College of Pathologists Virology Specialty Advisory Committee and Virology Examiners Panel and is a member of the Specialty Advisory Committee on Transfusion Transmitted Viruses. He is a co-author on four editions of the "Mims' Medical Microbiology" textbook, has written chapters in a number of other textbooks and has over 100 publications in international peer-reviewed journals and is an associate editor for two journals.

EDITORIAL BOARD

Editors in Chief

Dennis H. Bamford

Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Mark Zuckerman

South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust, London, United Kingdom and Department of Infectious Diseases, School of Immunology and Microbial Sciences, King's College London Medical School, London, United Kingdom

Section Editors

Claude M. Fauquet

St Louis, MO, United States

Michael Feiss

Department of Microbiology and Immunology, Carver College of Medicine, University of Iowa, Iowa City, IA, United States

Elizabeth E. Fry

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Said A. Ghabrial[†]

Department of Plant Pathology, University of Kentucky, Lexington, KY, United States

Eric Hunter

Department of Pathology and Laboratory Medicine, Emory University School of Medicine and Emory Vaccine Center, Emory University, Atlanta, GA, United States

Ilkka Julkunen

Institute of Biomedicine, University of Turku, Turku, Finland

Peter J. Krell

Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Mart Krupovic

Archaeal Virology Unit, Institut Pasteur, Paris, France

Maija Lappalainen

HUS Diagnostic Center, HUSLAB, Clinical Microbiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Hubert G.M. Niesters

Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Massimo Palmarini

MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

David Prangishvili

Institut Pasteur, Paris, France and Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

David I. Stuart

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom and Diamond Light Source, Didcot, United Kingdom

Nobuhiro Suzuki

Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

[†]Deceased.

SECTION EDITORS



Claude Fauquet received his PhD in biochemistry from University Louis Pasteur in Strasbourg, France in 1974. Dr. Fauquet joined the Institut de Recherche pour le Développement (IRD) and worked there as a plant virologist for 28 years, and served in Ivory Coast, West Africa for 14 years. In 1991, he founded the International Laboratory for Tropical Agricultural Biotechnology (ILTAB) at The Scripps Research Institute, CA, United States. ILTAB was then hosted by the Donald Danforth Plant Science Center, St. Louis, MO, from 1999 to 2012. In 2003, he co-founded the Global Cassava Partnership for the 21st Century (GCP21), which he directed until 2019 and which goal is to improve the cassava crop worldwide.

Dr. Fauquet is an international leader in plant virology including taxonomy, epidemiology, molecular virology, and in gene-silencing as an antiviral strategy. He was Secretary of the International Committee on Taxonomy of Viruses (ICTV) for 18 years and the editor of several ICTV Reports including the VIIIth ICTV Report in 2005.

He has published more than 300 research papers in reviewed journals and books. He is a fellow of the American Association for the Advancement of Science, of the American Phytopathological Society and a member of the St. Louis Academy of Sciences. In 2007, Dr. Fauquet was knighted “Chevalier de l’Ordre des Palmes Académiques” by the French Minister of High Education and Research.



Dr. Michael Feiss is Professor Emeritus in the Department of Microbiology and Immunology of the Carver College of Medicine at the University of Iowa, IA, United States. Dr. Feiss received his PhD in Genetics at the University of Washington followed by a postdoctoral traineeship in the laboratory of Dr. Allan Campbell at Stanford. Dr. Feiss is a microbial geneticist who studies virus assembly with an emphasis on how a DNA virus, bacteriophage lambda, packages viral DNA into the empty prohead shell. The lab investigates how sites in the viral DNA orchestrate the initiation and termination of the DNA packaging process. This work includes comprehensive examination of the DNA recognition sites. A related interest is study of terminase, the viral DNA packaging enzyme, including the functional domains for protein–DNA and protein–protein interactions. A second focus has been the roles of the bacterial host’s IHF and DnaJ proteins in the lytic life cycle of the virus. More recent work has involved a genetic dissection of the role of terminase’s ATPase center that powers translocation of viral DNA into the prohead. This interest in the ATP hydrolysis-driven packaging motor involves a multidisciplinary collaboration examining the kinetics of DNA packaging during individual packaging events. Finally, recent studies have also looked at how the packaging process has diverged among several lambda-like phages, including phages 21, N15, and Gifsy-1.



Elizabeth E. Fry is a senior postdoctoral scientist in structural biology at the University of Oxford, Oxford, United Kingdom, where she received her DPhil. for studies relating to the structure determination of Foot-and-Mouth Disease Virus. Specializing in structural virology, Dr. Fry has studied many virus/viral protein structures but her primary focus is on picornavirus structure and function, in particular receptor interactions and virus uncoating. She is particularly interested in rationally designing virus-like-particles as next generation vaccines to reduce the inherent risks in handling live viruses.



Said A. Ghabrial[†] received his BSc in 1959 from Cairo University, Cairo, Egypt, and his PhD from Louisiana State University, Baton Rouge, LA, United States, in 1965. Dr. Ghabrial did postdoctoral research at the University of California, Davis, CA, United States, before returning to Cairo, where he served as a plant virologist in the Ministry of Agriculture. He returned to the United States in 1970 to do postdoctoral research at Purdue University, West Lafayette, IN. In 1972, he joined the Plant Pathology Department at the University of Kentucky, Lexington, KY, United States, where he rose to the rank of professor in 1986 and worked until 2013.

Dr. Ghabrial has served as an associate and senior editor of *Phytopathology*. He served on the editorial boards of the *Encyclopedia of Virology*, 3rd edition and *Encyclopedia of Plant Pathology*, and edited a thematic issue of *Advances in Virus Research* on "Mycoviruses". He was a member of the American Phytopathological Society (APS) and the American Society for Virology (ASV); in July 2002 he was elected as a Fellow of the American Phytopathological Society. He also acted as Chair of the ICTV Subcommittee on Fungal Viruses in 1987–1993 and 2011–2014.

His long professional career allowed him to make many scientific achievements in phytopathology and virology. Among them are molecular dissection of a legume-infecting RNA virus, bean pod mottle virus (BPMV), development of BPMV-based vectors, discovery of a transmissible debilitation disease of the phytopathogenic ascomycete, *Helminthosporium victoriae* (*Cochliobolus victoriae*), establishment of a viral etiology of the *H. victoriae* disease, and advancement of structural biology of diverse fungal viruses.



Eric Hunter, PhD, is Professor of Pathology and Laboratory Medicine at Emory University, Atlanta, GA, United States. He serves as Co-Director of the Emory Center for AIDS Research and is a Georgia Research Alliance Eminent Scholar.

Dr. Hunter's research focus has been the molecular virology and pathogenesis of retroviruses, including human immunodeficiency virus. He has made significant contributions to the understanding of the role of retroviral glycoprotein structural features during viral entry and providing unique insights into the assembly and replication of this virus family. In recent years the emphasis of his research has been on HIV transmission and pathogenesis, defining the extreme genetic bottleneck and selection of viruses with unique traits during HIV heterosexual transmission. He has described the selection of fitter viruses at the target mucosa, a gender difference in the extent of selection bias, and a role for genital inflammation in reducing selection. His research has defined the impact of HIV adaptation to the cellular immune response on immune recognition and control of HIV after

transmission, as well as on virus replicative fitness in vitro and in vivo. Recent work highlights the roles that virus replicative fitness and sex of the host play in defining disease progression in a newly infected individual. His bibliography includes over 300 peer-reviewed articles, reviews, and book chapters. He has also been the recipient of four NIH merit awards for his work on retrovirus and HIV molecular biology.

Dr. Hunter served as the Editor in Chief of the journal *AIDS Research and Human Retroviruses* for 10 years. He was Chair of the AIDS Vaccine Research Subcommittee which is charged with providing advice and consultation on AIDS vaccine research to the National Institute of Allergy and Infectious Diseases and continues to serve on editorial boards for several academic journals and on external advisory committees for several government, academic, and commercial institutions.



Ilkka Julkunen graduated as an MD/PhD in 1984 from the Department of Virology, University of Helsinki, Helsinki, Finland. He worked as a postdoctoral research fellow at Memorial Sloan-Kettering Cancer Center in New York, United States, in 1986–1989, followed by positions as a senior scientist, group leader and research professor at Finnish Institute for Health and Welfare in 1989–2013. In 2013 he became a Professor of Virology at the University of Turku, Turku, Finland. The research interests of Dr. Julkunen have concentrated on innate and adaptive humoral immunity in viral and microbial infections. He has studied intracellular signaling and RIG-I and TLR-mediated activation of interferon system in human macrophages and dendritic cells and stable cell lines in response to human and avian influenza, Sendai, Zika and coronavirus infections. In addition, he has analyzed the downregulation of innate immunity by viral regulatory proteins from influenza, HCV, flavi-, filo- and coronaviruses. He has expertise in vaccinology, biotechnology and development of methods to analyze antiviral immunity, he has also been actively involved in research training and collaborations with biotechnological industry.

[†]Deceased.



Peter Krell started his career in virology early as a summer high school student working for the Canadian Forestry Service studying the resistance of nuclear polyhedrosis viruses (now called baculoviruses) to environmental exposure with Dr. Fred T. Bird at the Insect Pathology Research Institute in Sault Ste. Marie, ON, Canada. He received his BSc and MSc in biology from Carleton University studying the iridovirus *Tipula Iridescent Virus* with Dr Peter Lee, in Ottawa, the Canadian capital. For his PhD he headed east to Dalhousie University in Halifax, Nova Scotia on the Atlantic coast. In addition to enjoying the salt sea air, fresh cod, lobster and mussels, he studied the molecular biology of polydnaviruses under the guidance of Dr Don Stoltz. Heading south to Texas A&M University in College Station, TX, United States, as a Postdoctoral Fellow he worked with Dr. Max Summer, of baculovirus fame, and Dr. Brad Vinson continuing to study polydnaviruses, but also became steeped in the early days of molecular baculovirology. He then accepted a faculty position in the Department of Microbiology and Immunology at the University of Guelph in Guelph, ON, Canada. There he switched to baculovirus research, which was more tractable, due in part to available cell cultures and focused on viral DNA replication and functional genomics,

particularly on chitinase, cathepsin and ME53. In collaboration with Dr. Eva Nagy he studied molecular biology of different animal viruses, notably Fowl Avian adenoviruses and their development as vaccine vectors, but also on the birnavirus infectious pancreatic necrosis virus, the coronavirus porcine endemic diarrhea virus, fowlpox virus and the paramyxovirus Newcastle disease virus. He has been involved extensively with virus taxonomy, being active in the International Committee on Taxonomy of Viruses (ICTV) as member of the *Polydnaviridae* and *Baculoviridae* study groups, national representative of Canada on the ICTV, member of the Executive Committee for the ICTV and Chair of the ICTV Invertebrate Virus Subcommittee. In terms of governance, Peter Krell was President of the Canadian Society of Microbiology, Secretary and later President of the Society for Invertebrate Pathology, as well as being on the Editorial Boards of the *Canadian Journal of Microbiology* and the *ASM Journal of Virology*. While at the University of Guelph, he rose through the ranks to Professor and is currently University Professor Emeritus.



Mart Krupovic is the Head of the Archaeal Virology Unit in the Department of Microbiology at the Institut Pasteur of Paris, France. He received his MSc in Biochemistry in 2005 from the Vilnius University, Vilnius, Lithuania and PhD in 2010 in general microbiology from the University of Helsinki, Helsinki, Finland. His current research focuses on the diversity, origin, and evolution of viruses, as well as molecular mechanisms of virus–host interactions in archaea. He has published over 170 journal articles and serves as an editor or on the editorial boards of *Biology Direct*, *Research in Microbiology*, *Scientific Reports*, *Virology*, and *Virus Evolution*. He is also a member of the Executive Committee of the International Committee on Taxonomy of Viruses (ICTV) and chairs the Archaeal Viruses Subcommittee of the ICTV.



Maija Lappalainen, MD, PhD, Associate Professor of Clinical Microbiology, is the Head of Clinical Microbiology in the HUS Diagnostic Center, HUSLAB, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. In her thesis during the years 1987–1992 she studied the incidence and diagnostics of congenital toxoplasmosis. After PhD, her research interest has been in diagnostic clinical virology, viral hepatitis, respiratory infections, viral infections in the immunocompromised patients and viral infections during pregnancy.



Hubert G.M. Niesters (1958) studied biology and chemistry in Nijmegen, the Netherlands. After obtaining his PhD in Utrecht (Prof. dr. M. Horzinek and Prof. dr. B. van der Zeijst, 1987) on the molecular epidemiology of infectious bronchitis virus, he worked as a post-doctoral fellow with Prof. dr. Jim Strauss at the California Institute of Technology (Pasadena, United States) on the replication of Alphaviruses. He received a Niels Stensen fellowship (The Netherlands) and an E.S. Gosney fellowship (Caltech) during this period.

After returning to the Netherlands (1989), he became a research associate in medical microbiology at the Diagnostic Medical Center (Delft) but moved back to clinical virology as a senior research associate in 1991 at the Erasmus University Medical Center Rotterdam (Head Prof. dr. Ab Osterhaus). From 1993 to 2007, he was responsible for the molecular diagnostics unit. During this period, he was involved in the discovery and characterization of several new viruses and variants. In 2007, he became full professor and director of the Laboratory of Clinical Virology within the Department of Medical Microbiology at the University Medical Center Groningen and University of Groningen.

He has been actively involved in the implementation and development of new technologies like real-time amplification and automation within clinical virology. He has been focusing on molecular diagnostics and its use and the clinical value in a transplant setting, as well as in monitoring treatment of hepatitis viruses. Recently, his interest focuses on rapid regional epidemiology, automation including MiddleWare solutions for molecular diagnostics, as well as the cost-benefit of rapid point-of-impact molecular testing. Special interest is focused on raising awareness for the detection of enteroviruses (enterovirus D68) and its relationship with acute flaccid myelitis (AFM).

Since 2017, he is the Chair of the executive board of QCMD (Quality Control of Molecular Diagnostics, Glasgow). He is an auditor and team leader for the Dutch Council of Accreditation and Co-Editor in Chief of the *Journal of Clinical Virology*. He is an (co)-author of more than 250 peer-reviewed papers, chapters and reviews including emerging viruses, such as enterovirus D68 and hepatitis E virus (H-index 80).

For his entire work, he received in 2016 the “Ed Nowakowski Senior Memorial Clinical Virology Award” from the Pan American Society for Clinical Virology.



Massimo Palmarini is the Director of the MRC-University of Glasgow Centre for Virus Research and Chair of Virology at the University of Glasgow, Glasgow, United Kingdom. A veterinarian by training, his research programs focus on the biology, evolution and pathogenesis of arboviruses and the mechanisms of virus cross-species transmission. His work is funded by the MRC and the Wellcome Trust. Massimo Palmarini has been elected Fellow of the Academy of Medical Sciences, of the Royal Society of Edinburgh and of the Royal Society of Biology and he was a Wolfson-Royal Society Research Merit Awardee. He is a Wellcome Trust Investigator.



David Prangishvili, PhD, Honorary Professor at the Institut Pasteur, Paris, France, and Professor at Tbilisi State University, Tbilisi, Georgia, is one of the pioneers in studies on the biology of Archaea and their viruses. His scientific career spans ex-USSR (Institute of Molecular Biology, Moscow; 1970–1976), Georgia (Georgian National Academy of Sciences, Tbilisi; 1976–1991), Germany (Max-Planck Institute for Biochemistry, Munich; University of Regensburg; 1991–2004) and France (Institut Pasteur, Paris, 2004–2020). In the research groups headed by him, several dozens of new species and eight new families of archaeal viruses have been discovered and characterized, which display remarkable diversity of unique morphotypes and exceptional genome contents. The results of his research contribute to the knowledge on viral diversity on our planet and change the field of prokaryotic virology, leading to the notion that viruses of

hyperthermophilic Archaea form a particular group in the viral world, distinctive from viruses of Bacteria and Eukarya, and to the recognition of the virosphere of Archaea as one of the distinct features of this Domain of Life. David Prangishvili is a member of the Academia Europaea, the European Academy of Microbiology, and the Georgian National Academy of Sciences.



David I. Stuart is MRC Professor of Structural Biology in the Nuffield Department of Medicine, Oxford University, Oxford, United Kingdom, Life Science Director at Diamond Light Source and Director of Instruct-ERIC (pan-European organisation providing shared access to infrastructure and methods for structural biology). He has diverse interests in structural virology from picornaviruses, double-stranded RNA viruses and enveloped RNA viruses. His drive to develop structural techniques led to the determination of the structure of Bluetongue virus (1995) and then the first membrane containing virus, PRD1. More recently, he has been at the fore-front of bringing Cryo-EM technology to bear on virus structure determination and its future role in visualizing virus function in cellulose. In addition to basic science he has a strong commitment to structural vaccinology and the development of antiviral drugs.



Dr. Nobuhiro Suzuki, PhD, received his MSc (1985) in phytopathology and PhD (1989) in virology from Tohoku University in Sendai, Japan. Dr. Suzuki currently serves as a full Professor of the Institute of Plant Stress and Resources, formerly Institute of Plant Sciences and Bioresources at Okayama University and as an Editor of *Virus Research*, *Frontiers in Virology*, *Journal of General Plant Pathology*, *Virology Journal*, and *Biology*. He has also been Guest Editor to *PLoS Pathogens*, *PNAS*, and *mBio*, and an Editorial Board member of *Virology* and *Journal of Virology*.

Suzuki Laboratory focuses on characterization of diverse viruses infecting phytopathogenic fungi and exploration of their interplays. Recent achievements include the discovery of a neo-virus lifestyle exhibited by a (+)ssRNA virus and an unrelated dsRNA virus in a plant pathogenic fungus and of multilayer antiviral defense in fungi involving Dicer. Prior to coming to Kurashiki, Okayama Prefecture, he was a visiting fellow of the Center for Agricultural Biotechnology at the

University of Maryland Biotechnology Institute (UMBI), College Park, MA, United States, for 4 years (1997–2001) to study molecular biology of hypoviruses in the laboratory of Professor Donald L. Nuss. Before visiting UMBI, he served as an assistant professor and a lecturer of the Biotechnology Institute at the Akita Prefectural College of Agriculture, Japan, for 11 years (1988–1998) where he was engaged in a project on molecular characterization of rice dwarf phytoecovirus, a member of the family *Reoviridae*. He received awards from the Japanese Phytopathological Society of Japan and Japanese Society for Virology for his outstanding achievements in plant and fungal virology.

FOREWORD

I am delighted to write the foreword to this wonderful Fourth Edition of the *Encyclopedia of Virology*. The Third Edition was published in 2008, how the world has changed in the intervening years. The release of the updated fourth edition could not be more timely or more prescient. It is superb and a huge tribute to the authors, Elsevier the publisher, and to the brilliant editors, Dennis Bamford and Mark Zuckerman.

SARS-CoV-2 has dominated the world since it emerged in 2019 and affected every continent and every aspect of life. A reminder, if it were needed, of the impact of infectious diseases, the importance of virology and the vulnerability and interconnectivity of our world.

There is no doubt that with rapidly changing ecology, urbanization, climate change, increased travel, and fragile public health systems, epidemics and pandemics will become more frequent, more complex and harder to prevent and contain. Most of these epidemics will be caused by viruses, those we know about and maybe able to predict and some we do not know of that will emerge from animals, plants or the environment. Our changing climate will change the epidemiology of viruses, their vectors and the infections they cause, hence the critical importance of this totally revised Fourth Edition of the *Encyclopedia of Virology* which brings together research and an understanding of viruses in animals, plants, bacteria and fungi, the environment, and among humans. Never has a holistic, one-health understanding been more important.

That starts with an understanding of the fundamentals of virology, a field of science that has been transformed in the years since the Third Edition. An understanding transformed by embracing traditional fields of molecular and structural biology, genomics, and influenced by immunology, genetics, pharmacology and increasingly by epidemiology and mathematics. Events of 2020 and 2021 also show why it is so important to integrate within traditional virology an understanding of the animal and human health and behavior, of climate change and its impact on the ecology of viruses, plant sciences and vectors. And why we must understand the viruses we think we know well, and those viruses less extensively studied. Research is critical to this, research that pushes the boundaries of what we know, has the humility to seek answers to things we do not understand and shares that knowledge with the widest possible community. That research will be most exciting at the interface between disciplines, most impactful when dynamic, open, inclusive, global, and collaborative.

This is what the Fourth Edition of the *Encyclopedia of Virology*, the largest reference source of research in virology sets out to achieve. It is a wonderful contribution to a critical field of knowledge. It contains new chapters, every chapter revised and updated by a dedicated global community who have come together to provide what is a brilliant and inspiring reference. It is an honor to contribute in a very small way to the timely release of the Fourth Edition of the *Encyclopedia of Virology*.

Jeremy Farrar

PREFACE

The fourth edition of the Encyclopedia of Virology is encyclopedic, but we wanted to move away from an alphabetical list, apart from where it was more logical, to a vision that encompassed a different structure. Articles describing novel trends as well as original discoveries in specific subfields of virology have been distributed into a set of five volumes, namely Fundamentals of Virology, Human and Animal Viruses, Plant Viruses, Bacterial, Archaeal, Fungal, Algal and Invertebrate Viruses, and Diagnosis, Treatment and Prevention of Virus Infections.

We had hoped that the new edition would 'go viral' but it was ironic that the time to publication 12 years after the previous edition had been made a bit longer due to a virus infection. The world encountered a devastating global pandemic, COVID-19, caused by a new type of a coronavirus, SARS-CoV-2. Scientists in many disciplines all over the world started immediate efforts to discover solutions as to how to mitigate and stop the spread of the pandemic. Virology moved from being a highly specialized subject to one in which everyone became a virologist, proving just how significant the different aspects of virology are in terms of understanding the nature of viral infection. Since the previous edition, the growth in the field of general virology has been enormous, including huge advances in basic science, identification of novel viruses, diagnostic methods, treatment and prevention. Taking this into account, the introduction of the articles within the Encyclopedia are very timely and crucial for providing a wealth of knowledge of the latest findings in the field of virology to a vast range of people, whether school students, undergraduates, postgraduates, teachers, scientists, researchers, journalists and others interested in infections and the conflict between the host and the pathogen.

Pandemic viruses have become a serious public concern in the changing world. We can ask ourselves whether we have reached the point in which nature can no longer cope with the consequences of increased population density and human activities that are harmful to the environment. Although several pandemics have threatened mankind before, this COVID-19 pandemic has highlighted the massive adverse economic consequences towards the wellbeing of society and the importance of research in virology.

We aimed to produce a Major Reference Work that differs in approach to others and binds all the virology disciplines together. Chapters have been included on origin, evolution and emergence of viruses, environmental virology and ecology, epidemiology, techniques for studying viruses, viral life cycles, structure, entry, genome and replication, assembly and packaging and taxonomy and viral–host interactions. Information has been included on all known species of viruses infecting bacteria, fungi, plants, vertebrates and invertebrates. Additional topics include antiviral classification and examples of their use in management of infection, diagnostic assays and vaccines, as well as the economic importance of viral diseases of crops and their control.

This edition used viral classification according to the 9th Report of the International Committee on Taxonomy of Viruses published in 2012. Updating it to the 10th Report in 2020 was affected by the pandemic and can be found online at <http://ictv.global/report/>.

We wish to acknowledge the hard work, interest, flexibility and patience, during such difficult times both socially and professionally, of everybody involved in the process of writing this edition of the Encyclopedia of Virology, especially Katarzyna Miklaszewska, Priscilla Braglia, Sam Crowe and colleagues at Elsevier. We sincerely thank all the authors and section editors for their excellent contributions to this edition.

Book Cover Image: Viruses are obligate parasites and all cells have their own viruses increasing the total number of viruses to the estimated astronomical number of 10^{31} that extends the number of stars in the universe. The viral string illustrates how pandemic viruses surround the globe. The original picture was created by Dr. Nina Atanasova (Finnish Meteorological Institute and University of Helsinki) and amended by Matthew Limbert at Elsevier.

*Dennis H. Bamford
Mark Zuckerman*

HOW TO USE THE ENCYCLOPEDIA

Structure of the Encyclopedia

All articles in the encyclopedia are arranged thematically as a series of entries within subjects/sections, apart from volume 2 where there it was more logical to have articles arranged alphabetically.

There are three features to help you easily find the topic you are interested in: a thematic contents list, a full subject index, and contributors.

1. Thematic contents list: The alphabetical contents list, which appears at the front of each volume, lists the entries in the order that they appear in the encyclopedia.
2. Index: The index appears at the end of volume 5 and includes page numbers for quick reference to the information you are looking for. The index entries differentiate between references to a whole entry, a part of an entry, and a table or figure.
3. Contributors: At the start of each volume there is a list of the authors who contributed to all volumes.

LIST OF CONTRIBUTORS

Stephen T. Abedon

*The Ohio State University, Mansfield, OH,
United States*

Peter Abrahamian

*Agricultural Research Service, US Department of
Agriculture, Beltsville, MD, United States*

Jônatas S. Abrahão

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Florence Abravanel

*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Nicola G.A. Abrescia

*Center for Cooperative Research in Biosciences, Basque
Research and Technology Alliance, Derio, Spain;
Ikerbasque, Basque Foundation for Science, Bilbao,
Spain; and Center for Biomedical Research in the Liver
and Digestive Diseases Network, Carlos III Health
Institute, Madrid, Spain*

Gian Paolo Accotto

*Institute for Sustainable Plant Protection, National
Research Council of Italy, Torino, Italy*

Elisabeth Adderson

*St. Jude Children's Research Hospital, Memphis, TN,
United States and University of Tennessee Health
Sciences Center, Memphis, TN, United States*

Mustafa Adhab

University of Baghdad, Baghdad, Iraq

Alexey A. Agranovsky

Lomonosov Moscow State University, Moscow, Russia

Nasim Ahmed

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Maher Al Rwahnih

University of California, Davis, CA, United States

Olufemi J. Alabi

*Texas A&M AgriLife Research and Extension Center,
Weslaco, TX, United States*

Aurélien A. Albertini

*Institute for Integrative Biology of the Cell (I2BC),
French Alternative Energies and Atomic Energy
Commission, French National Center for Scientific
Research, Paris-Sud University, University of
Paris-Saclay, Gif-sur-Yvette, France*

Aleksandra Alimova

*The City University of New York (CUNY), School of
Medicine, The City College of New York, New York,
NY, United States*

Juan C. Alonso

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Imran Amin

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Stephanie E. Ander

*University of Colorado School of Medicine, Aurora, CO,
United States*

Danielle E. Anderson

Duke-NUS Medical School, Singapore, Singapore

Ida Bagus Andika

Qingdao Agricultural University, Qingdao, China

Ana C.d.S.P. Andrade

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Juana Angel

Pontifical Javeriana University, Bogota, Colombia

Vanesa Anton-Vazquez

King's College Hospital, London, United Kingdom

Guido Antonelli

Sapienza University of Rome, Rome, Italy

Josefa Antón

University of Alicante, Alicante, Spain

Nanako Aoki

*Tokyo University of Agriculture and Technology, Fuchu,
Japan*

Timothy D. Appleby

King's College Hospital, London, United Kingdom

Miguel Arenas

*Department of Biochemistry, Genetics and Immunology,
University of Vigo, Vigo, Spain and CINBIO
(Biomedical Research Center), University of Vigo, Vigo,
Spain*

Basil Arif

*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Vicente Arnau

*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain*

Gaurav Arya

Duke University, Durham, NC, United States

Leyla Asadi

University of Alberta, Edmonton, AB, Canada

Sassan Asgari

The University of Queensland, Brisbane, QLD, Australia

Nina S. Atanasova

*Finnish Meteorological Institute, Helsinki, Finland and
University of Helsinki, Helsinki, Finland*

Houssam Attoui

*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Silvia Ayora

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Walid Azab

Free University of Berlin, Berlin, Germany

Sasha R. Azar

*The University of Texas Medical Branch, Galveston, TX,
United States*

Fengwei Bai

*The University of Southern Mississippi, Hattiesburg, MS,
United States*

Dalan Bailey

The Pirbright Institute, Pirbright, United Kingdom

S.C. Baker

*Loyola University of Chicago, Maywood, IL,
United States*

Fausto Baldanti

*University of Pavia, Pavia, Italy and Scientific Institute
for Research, Hospitalization and Healthcare, San
Matteo Polyclinic Foundation, Pavia, Italy*

Logan Banadyga

*Public Health Agency of Canada, Winnipeg, MB,
Canada*

Ashley C. Banyard

*Animal and Plant Health Agency, Addlestone,
United Kingdom; University of West Sussex, Falmer,
United Kingdom; and St. George's Medical School,
University of London, London, United Kingdom*

Xiaoyong Bao

*The University of Texas Medical Branch, Galveston, TX,
United States*

Yiming Bao

*Beijing Institute of Genomics, Chinese Academy of
Sciences, University of Chinese Academy of Sciences,
Beijing, China*

Alan D.T. Barrett

*The University of Texas Medical Branch, Galveston, TX,
United States*

Diana P. Baquero

*Archaeal Virology Unit, Institut Pasteur, Paris, France
and Sorbonne University, Paris, France*

Moshe Bar-Joseph

*Agricultural Research Organization, Volcani Center, Bet
Dagan, Israel*

Rachael S. Barr

*Bristol Royal Hospital for Children, Bristol,
United Kingdom*

Ralf Bartenschlager

Heidelberg University, Heidelberg, Germany

David L.V. Bauer

Francis Crick Institute, London, United Kingdom

Oliver W. Bayfield

University of York, York, United Kingdom

Sally A. Baylis

Paul-Ehrlich-Institute, Langen, Germany

Philippa M. Beard

*The Pirbright Institute, Pirbright, United Kingdom and The
Roslin Institute, University of Edinburgh, United Kingdom*

Paul Becher

University of Veterinary Medicine, Hannover, Germany

Björn Becker

Saarland University, Saarbrücken, Germany

Karen L. Beemon

Johns Hopkins University, Baltimore, MD, United States

Martin Beer

Friedrich-Loeffler-Institute, Insel Riems, Germany

Jose Miguel Benito

*Health Research Institute of the Jiménez Díaz
Foundation, Autonomous University of Madrid and Rey
Juan Carlos University Hospital, Móstoles, Spain*

Mária Benkő

*Institute for Veterinary Medical Research, Center for
Agricultural Research, Budapest, Hungary*

Max Bergoin
National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada

Sabrina Bertin
Council for Agricultural Research and Economics,
Research Center for Plant Protection and Certification,
Rome, Italy

Shweta Bhatt
University of Copenhagen, Copenhagen, Denmark

Dennis K. Bideshi
California Baptist University, Riverside, CA,
United States and University of California, Riverside,
CA, United States

Yves Bigot
INRAE – French National Research Institute for
Agriculture, Food and Environment, Nouzilly, France

Richard J. Bingham
University of York, York, United Kingdom

Vera Bischoff
Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany

Kate N. Bishop
Francis Crick Institute, London, United Kingdom

Lindsay W. Black
The University of Maryland School of Medicine,
Baltimore, MD, United States

Romain Blanc-Mathieu
Institute for Chemical Research, Kyoto University, Kyoto,
Japan

Soile Blomqvist
National Institute for Health and Welfare, Helsinki,
Finland

Bryony C. Bonning
University of Florida, Gainesville, FL, United States

Lisa M. Bono
Rutgers, The State University of New Jersey,
New Brunswick, NJ, United States

Alexia Bordigoni
Aix-Marseille University, CNRS, IRD, Mediterranean
Institute of Oceanography, Marseille, France and
Aix-Marseille University, IRD257, Assistance-Publique
des Hôpitaux de Marseille, UMR Microbes, Evolution,
Phylogeny and Infections (MEPHI), IHU Méditerranée
Infection, Marseille, France

Mihnea Bostina
University of Otago, Dunedin, New Zealand

Maxime Boutier
University of Liège, Liège, Belgium

P.R. Bowser
Cornell University, Ithaca, NY, United States

Daniel Bradshaw
Public Health England, London, United Kingdom

Claude Bragard
University of Louvain, Louvain-la-Neuve, Belgium

Aaron C. Brault
Centers for Disease Control and Prevention, Fort Collins,
CO, United States

Nicolas Bravo-Vasquez
St. Jude Children's Research Hospital, Memphis, TN,
United States

Rob W. Briddon
University of Agriculture, Faisalabad, Pakistan

Thomas Briebe
Columbia University, New York, NY, United States

Paul Britton
The Pirbright Institute, Pirbright, United Kingdom

Thomas J. Brouwers
Athena Institute, VU Amsterdam, Amsterdam, The
Netherlands

Kevin E. Brown
Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom

Corina P.D. Brussaard
NIOZ Royal Netherlands Institute for Sea Research,
Den Burg, Texel, The Netherlands and Utrecht
University, Utrecht, The Netherlands

Harald Brüssow
Laboratory of Gene Technology, Department of
Biosystems, KU Leuven, Leuven, Belgium

Joachim J. Bugert
Bundeswehr Institute of Microbiology, Munich, Germany

Jozef J. Bujarski
Northern Illinois University, DeKalb, IL, United States
and Polish Academy of Sciences, Poznan, Poland

Laura Burga
University of Otago, Dunedin, New Zealand

Sara H. Burkhard
University Hospital of Zurich, Zurich, Switzerland

Cara C. Burns
Centers for Disease Control and Prevention, Atlanta,
GA, United States

Felicity Burt
University of the Free State, Bloemfontein, South Africa

Kerry S. Burton
Leamington Spa, United Kingdom

Sarah J. Butcher
University of Helsinki, Helsinki, Finland

Mathias Büttner
Leipzig University, Leipzig, Germany

Jesse Cahill
Sandia National Labs, Albuquerque, NM, United States

Marianna Calabretto
Sapienza University of Rome, Rome, Italy

Thierry Candresse
The National Research Institute for Agriculture, Food and the Environment, University of Bordeaux, Villenave d'Ornon, France

Alan J. Cann
University of Leicester, Leicester, United Kingdom

Lorenzo Capucci
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Irene Carlon-Andres
University of Oxford, Oxford, United Kingdom

José M. Casasnovas
National Center for Biotechnology, Spanish National Research Council (CSIC), Madrid, Spain

J.W. Casey
Cornell University, Ithaca, NY, United States

R.N. Casey
Cornell University, Ithaca, NY, United States

Sherwood R. Casjens
University of Utah, Salt Lake City, UT, United States

Antonella Casola
The University of Texas Medical Branch, Galveston, TX, United States

José R. Castón
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Carlos E. Catalano
University of Colorado Anschutz Medical Campus, Skaggs School of Pharmacy and Pharmaceutical Sciences, Aurora, CO, United States

Roberto Cattaneo
Mayo Clinic, Rochester, MN, United States

Patrizia Cavadini
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Supranee Chaiwatpongsakorn
Nationwide Children's Hospital, Columbus, OH, United States

Supriya Chakraborty
Jawaharlal Nehru University, New Delhi, India

Yu-Chan Chao
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Tyler P. Chavers
Centers for Disease Control and Prevention, Atlanta, GA, United States

Keping Chen
Jiangsu University, Zhenjiang, China

Xiaorui Chen
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Yanping Chen
Bee Research Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

Dayna Cheng
National Cheng Kung University, Tainan, Taiwan

Quentin Chesnais
University of Strasbourg, Colmar, France

Sotaro Chiba
Nagoya University, Nagoya, Japan

Wah Chiu
Stanford University, Stanford, CA, United States

David Chmielewski
Stanford University, Stanford, CA, United States

Irma E. Cisneros
The University of Texas Medical Branch, Galveston, TX, United States

Lark L. Coffey
University of California, Davis, CA, United States

Alanna B. Cohen
Rutgers University, New Brunswick, NJ, United States

Jeffrey I. Cohen
Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Seth Coleman
Rice University, Houston, TX, United States

Miquel Coll
*Institute for Research in Biomedicine, Barcelona, Spain
 and Institute for Molecular Biology of Barcelona,
 Barcelona, Spain*

John Collinge
*UCL Institute of Prion Diseases, London,
 United Kingdom*

Carina Conceicao
The Pirbright Institute, Pirbright, United Kingdom

Gabriela N. Condezo
*National Center for Biotechnology, Spanish National
 Research Council, Madrid, Spain*

Michaela J. Conley
*MRC-University of Glasgow Centre for Virus Research,
 Glasgow, United Kingdom*

Charles A Coomer
University of Oxford, Oxford, United Kingdom

Anne K. Cordes
*Hannover Medical School, Institute of Virology, Hannover,
 Germany*

Mauricio Cortes Jr.
*Department of Chemistry, College of Arts and Sciences,
 Fort Wayne, IN, United States*

Robert H.A. Coutts
University of Hertfordshire, Hatfield, United Kingdom

Jeff A. Cowley
CSIRO Livestock Industries, Brisbane, QLD, Australia

Robert W. Cross
*The University of Texas Medical Branch, Galveston, TX,
 United States*

Henryk Czosnek
The Hebrew University of Jerusalem, Rehovot, Israel

Håkon Dahle
*Department of Biological Sciences, University of Bergen,
 Bergen, Norway*

Janet M. Daly
*University of Nottingham, Sutton Bonington,
 United Kingdom*

Subha Das
Okayama University, Kurashiki, Japan

Indranil Dasgupta
University of Delhi, New Delhi, India

Sibnarayan Datta
*Defence Research Laboratory, Defence Research and
 Development Organisation (DRDO), Tezpur, Assam,
 India*

Amy Davis
*St Jude Children's Research Hospital, Memphis, TN,
 United States*

William O. Dawson
*Citrus Research and Education Center, Lake Alfred, FL,
 United States and University of Florida, Lake Alfred, FL,
 United States*

Erik De Clercq
*Rega Institute for Medical Research, KU Leuven,
 Leuven, Belgium*

Raoul J. de Groot
Utrecht University, Utrecht, The Netherlands

Juan C. de la Torre
*The Scripps Research Institute, La Jolla, CA,
 United States*

Marcelo De las Heras
University of Zaragoza, Zaragoza, Spain

Juliana Gabriela Silva de Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Athos S. de Oliveira
University of Brasília, Brasília, Brazil

Nicole T. de Stefano
University of Florida, Gainesville, FL, United States

Greg Deakin
NIAB-EMR, East Malling, United Kingdom

Philippe Delfosse
University of Luxembourg, Esch-sur-Alzette, Luxembourg

Natacha Delrez
University of Liège, Liège, Belgium

Tatiana A. Demina
*Molecular and Integrative Biosciences Research Program,
 Faculty of Biological and Environmental Sciences,
 University of Helsinki, Helsinki, Finland*

Ismail Demir
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

Zihni Demirbağ
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

X. Deng
*Loyola University of Chicago, Maywood, IL,
 United States*

Cécile Desbiez
*Plant Pathology Unit, INRAE – French National
 Research Institute for Agriculture, Food and
 Environment, Montfavet, France*

Christelle Desnues

Aix-Marseille University, CNRS, IRD, Mediterranean Institute of Oceanography, Marseille, France and Aix-Marseille University, IRD 257, Assistance-Publique des Hôpitaux de Marseille, UMR Microbes, Evolution, Phylogeny and Infections (MEPHI), IHU Méditerranée Infection, Marseille, France

Samantha J. DeWerff

University of Illinois at Urbana-Champaign, Urbana, IL, United States

Daniele Di Carlo

Sapienza University of Rome, Rome, Italy

Arturo Diaz

La Sierra University, Riverside, CA, United States

Alfredo Diaz-Lara

University of California, Davis, CA, United States

Ralf G. Dietzgen

The University of Queensland, St. Lucia, QLD, Australia

Michele Digiario

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Michael Dills

Montana State University, Bozeman, MT, United States

Wayne Dimech

National Serology Reference Laboratory, Fitzroy, VIC, Australia

Savithramma P. Dinesh-Kumar

University of California, Davis, CA, United States

Linda K. Dixon

The Pirbright Institute, Pirbright, United Kingdom

Valerian V. Dolja

Oregon State University, Corvallis, OR, United States

Aušra Domanska

University of Helsinki, Helsinki, Finland

Leslie L. Domier

Agricultural Research Service, US Department of Agriculture, Urbana, IL, United States

Pilar Domingo-Calap

Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, Valencia, Spain

Tatiana Domitrovic

Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Sarah M. Doore

Michigan State University, East Lansing, MI, United States

Lucy Dorrell

University of Oxford, Oxford, United Kingdom

Rosemary A. Dorrington

Rhodes University, Grahamstown, South Africa

Andor Doszpoly

Hungarian Academy of Sciences, Budapest, Hungary

Andreas Dotzauer

University of Bremen, Bremen, Germany

Simon B. Drysdale

St George's University Hospitals NHS Foundation Trust, London, United Kingdom and St George's, University of London, London, United Kingdom

Robert L. Duda

University of Pittsburgh, Pittsburgh, PA, United States

Carol Duffy

University of Alabama, Tuscaloosa, AL, United States

Siobain Duffy

Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

David D. Dunigan

University of Nebraska–Lincoln, Lincoln, NE, United States

Stéphane Duquerroy

University of Paris-Saclay, Orsay, France and Institut Pasteur, Paris, France

Bas E. Dutilh

Utrecht University, Utrecht, The Netherlands and Radboud University Medical Center, Nijmegen, The Netherlands

Michael Edelstein

Faculty of Medicine, Bar Ilan University, Ramat Gan, Israel

Herman K. Edskes

National Institutes of Health, Bethesda, MD, United States

Rosina Ehmann

Bundeswehr Institute of Microbiology, Munich, Germany

Toufic Elbeaino

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Joanne B. Emerson

University of California, Davis, CA, United States

Ann Emery

University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Christine E. Engeland
University Hospital Heidelberg and German Cancer Research Center, Heidelberg, Germany and Witten/Herdecke University, Witten, Germany

Luis Enjuanes
National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Katri Eskelin
University of Helsinki, Helsinki, Finland

Rosa Esteban
Institute of Biology and Functional Genomics, CSIC/University of Salamanca, Salamanca, Spain

Mary K. Estes
Baylor College of Medicine, Houston, TX, United States

Cassia F. Estofolete
São José do Rio Preto School of Medicine, São José do Rio Preto, Brazil

Alyssa B. Evans
National Institutes of Health, Hamilton, MT, United States

Øystein Evensen
Norwegian University of Life Sciences, Oslo, Norway

Alex Evilevitch
Department of Experimental Medical Science, Lund University, Lund, Sweden

Montserrat Fàbrega-Ferrer
Institute for Research in Biomedicine, Barcelona, Spain and Institute for Molecular Biology of Barcelona, Barcelona, Spain

Francesco Faggioli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Bentley A. Fane
University of Arizona, Tucson, AZ, United States

Brian A. Federici
University of California, Riverside, CA, United States

F. Fenner
Australian National University, Canberra, ACT, Australia

Isabel Fernández de Castro
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Giovanni Ferrara
University of Alberta, Edmonton, AB, Canada

Elvira Fiallo-Olivé
Institute for Mediterranean and Subtropical Horticulture “La Mayora”–Spanish National Research Council–University of Malaga, Algarrobo-Costa, Málaga, Spain

Andrew E. Firth
University of Cambridge, Cambridge, United Kingdom

Roland A. Fleck
King’s College London, London, United Kingdom

Ricardo Flores
Polytechnic University of Valencia, Higher Council of Scientific Research, Valencia, Spain

Ervin Fodor
University of Oxford, Oxford, United Kingdom

Anthony R. Fooks
Animal and Plant Health Agency, Addlestone, United Kingdom; University of Liverpool, Liverpool, United Kingdom; and St. George’s Medical School, University of London, London, United Kingdom

Patrick Forterre
Archeal Virology Unit, Institut Pasteur, Paris, France and French National Center for Scientific Research, Institute of Integrative Biology of the Cell, University of Paris-Saclay, Gif sur Yvette, France

Rennos Fragkoudis
University of Nottingham, Sutton Bonington, United Kingdom and University of Edinburgh, Edinburgh, United Kingdom

Manuel A. Franco
Pontifical Javeriana University, Bogota, Colombia

Giovanni Franzo
Department of Animal Medicine, Production and Health (MAPS), Padua University, Padua, Italy

Graham L. Freimanis
The Pirbright Institute, Pirbright, United Kingdom

Juliana Freitas-Astúa
Brazilian Agricultural Research Corporation (Embrapa) Cassava and Fruits, Cruz das Almas, Brazil

Elizabeth E. Fry
Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Marc Fuchs
Cornell University, Geneva, NY, United States

Tsutomu Fujimura
Institute of Biology and Functional Genomics, CSIC/University of Salamanca, Salamanca, Spain

- Kuko Fuke
Tokyo University of Agriculture and Technology, Fuchu, Japan
- Toshiyuki Fukuhara
Tokyo University of Agriculture and Technology, Fuchu, Japan
- To S. Fung
South China Agricultural University, Guangzhou, China
- Yahya Z.A. Gaafar
Julius Kuehn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany
- Toni Gabaldon
Barcelona Supercomputing Center-National Center for Supercomputing, Institute of Research in Biomedicine, and Catalan Institution for Research and Advanced Studies, Barcelona, Spain
- Morgan Gaïa
University of Paris-Saclay, Evry, France
- José Gallardo
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain
- Hernan Garcia-Ruiz
University of Nebraska–Lincoln, Lincoln, NE, United States
- Juan A. García
National Center for Biotechnology-Spanish National Research Council, Madrid, Spain
- Matteo P. Garofalo
The University of Texas Medical Branch, Galveston, TX, United States
- Yves Gaudin
Institute for Integrative Biology of the Cell (I2BC), French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, Paris-Sud University, University of Paris-Saclay, Gif-sur-Yvette, France
- Andrew D.W. Geering
The University of Queensland, St. Lucia, QLD, Australia
- Thomas W. Geisbert
The University of Texas Medical Branch, Galveston, TX, United States
- Andrea Gentili
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy
- Volker Gerdts
University of Saskatchewan, Saskatoon, SK, Canada
- Said A. Ghabrial[†]
Department of Plant Pathology, University of Kentucky, Lexington, KY, United States
- Clément Gilbert
Evolution, Genomes, Behavior and Ecology Laboratory, CNRS University of Paris-Sud UMR 9191, IRD UMR 247, Gif-sur-Yvette, France
- Robert L. Gilbertson
University of California, Davis, CA, United States
- Efstathios S. Giotis
Imperial College London, London, United Kingdom and University of Essex, Colchester, United Kingdom
- Laurent Glais
French Federation of Seed Potato Growers/Research, Development, Promotion of Seed Potato, Paris, France and Institute for Genetics, Environment and Plant Protection, Agrocampus West, French National Institute for Agriculture, Food and Environment, University of Rennes 1, Le Rheu, France
- Miroslav Glasa
Biomedical Research Center, Slovak Academy of Sciences, Bratislava, Slovakia
- Ido Golding
University of Illinois at Urbana-Champaign, Urbana, IL, United States
- Esperanza Gomez-Lucia
Complutense University of Madrid, Madrid, Spain
- Zheng Gong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
- Andrea González-González
University of Florida, Gainesville, FL, United States
- Michael M. Goodin
University of Kentucky, Lexington, KY, United States
- Alexander E. Gorbalenya
Leiden University Medical Center, Leiden, The Netherlands
- Paul Gottlieb
The City University of New York (CUNY), School of Medicine, The City College of New York, New York, NY, United States
- M.-A. Grandbastien
INRAE – French National Research Institute for Agriculture, Food and Environment, Versailles, France

[†]Deceased.

Meritxell Granell
*National Center for Biotechnology, Madrid, Spain and
 Institute of Chemical Research of Catalonia (ICIQ),
 Tarragona, Spain*

Patrick L. Green
*The Ohio State University, Columbus, OH,
 United States*

Sandra J. Greive
University of York, York, United Kingdom

Diane E. Griffin
*Johns Hopkins Bloomberg School of Public Health,
 Baltimore, MD, United States*

Jonathan M. Grimes
University of Oxford, Oxford, United Kingdom

Nigel Grimsley
*Integrative Biology of Marine Organisms Laboratory,
 Banuyls-sur-Mer, France and Sorbonne University,
 Banuyls-sur-Mer, France*

Bruno Gronenborn
*Institute for Integrative Biology of the Cell, CNRS,
 University of Paris-Sud, CEA, Gif sur Yvette, France*

Julianne H. Grose
Brigham Young University, Provo, UT, United States

Scott Grytdal
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

Duane J. Gubler
Duke-NUS Medical School, Singapore, Singapore

Peixuan Guo
*College of Pharmacy, The Ohio State University,
 Columbus, OH, United States*

Tongkun Guo
*Beijing Institute of Genomics, Chinese Academy of
 Sciences, Beijing, China*

Anne-Lise Haenni
*Institut Jacques Monod, French National Center for
 Scientific Research, Paris Diderot University, Paris,
 France*

Susan L. Hafenstein
*Pennsylvania State University, Hershey, PA,
 United States*

Ahmed Hafez
*Biotechvana, Valencia, Spain; Pompeu Fabra University,
 Barcelona, Spain; and Minia University, Minya, Egypt*

Marie Hagbom
Linköping University, Linköping, Sweden

Sébastien Halary
*National Museum of Natural History, UMR 7245
 CNRS/MNHN Molécule de Communication et
 Adaptation des Micro-organismes, Paris, France*

Aron J. Hall
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

John Hammond
*Floral and Nursery Plants Research, Agricultural
 Research Service, US Department of Agriculture,
 Beltsville, MD, United States*

Rosemarie W. Hammond
*Agricultural Research Service, US Department of
 Agriculture, Beltsville, MD, United States*

Virginia Hargest
*St Jude Children's Research Hospital, Memphis, TN,
 United States and University of Tennessee Health
 Science Center, Memphis, TN, United States*

Scott J. Harper
Washington State University, Prosser, WA, United States

Balázs Harrach
*Institute for Veterinary Medical Research, Center for
 Agricultural Research, Budapest, Hungary*

Masayoshi Hashimoto
The University of Tokyo, Tokyo, Japan

Muhammad Hassan
University of Agriculture, Faisalabad, Pakistan

Asma Hatoum-Aslan
University of Alabama, Tuscaloosa, AL, United States

Philippa C. Hawes
The Pirbright Institute, Pirbright, United Kingdom

Janelle A. Hayes
*University of Massachusetts Medical School, Worcester,
 MA, United States*

Guijuan He
Virginia Tech, Blacksburg, VA, United States

Klaus Hedman
*University of Helsinki, Helsinki, Finland and Helsinki
 University Hospital, Helsinki, Finland*

Albert Heim
Hannover Medical School, Hanover, Germany

Gary L. Hein
*University of Nebraska–Lincoln, Lincoln, NE,
 United States*

Manfred Heinlein
*IBMP-CNRS, University of Strasbourg, Strasbourg,
 France*

Mercedes Hernando-Pérez
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Carmen Hernández
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Etienne Herrbach
University of Strasbourg, Colmar, France

Stephen Higgs
*Biosecurity Research Institute, Kansas State University,
Manhattan, KS, United States*

Bradley I. Hillman
Rutgers University, New Brunswick, NJ, United States

Deborah M. Hinton
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Judith Hirsch
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Jody Hobson-Peters
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Natalie M. Holste
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Jin S. Hong
Kangwon National University, Chunchon, South Korea

Margaret J. Hosie
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Olivia G. Howell
University of Alabama, Tuscaloosa, AL, United States

Liya Hu
Baylor College of Medicine, Houston, TX, United States

Zhaoyang Hu
Jiangsu University, Zhenjiang, China

Kuan-Ying A. Huang
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Yu Huang
Peking University, Beijing, China

Natalia B. Hubbs
Hanover College, Hanover, IN, United States

Elisabeth Huguet
*Research Institute on Insect Biology, French National
Center for Scientific Research, University of Tours,
Tours, France*

Roger Hull
John Innes Centre, Norwich, United Kingdom

Kiwamu Hyodo
Okayama University, Kurashiki, Japan

Eugénie Hébrard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

Martin Hölzer
University of Jena, Jena, Germany

Tetsuro Ikegami
*The University of Texas Medical Branch at Galveston,
Galveston, TX, United States*

Niina Ikonen
*Finnish Institute for Health and Welfare, Helsinki,
Finland*

Cihan İnan
*Department of Molecular Biology and Genetics,
Karadeniz Technical University, Trabzon, Turkey*

İkbal Agah İnce
*Department of Medical Microbiology, Acibadem
University School of Medicine, Istanbul, Turkey*

Katsuaki Inoue
Diamond Light Source, Didcot, United Kingdom

Toru Iwanami
Tokyo University of Agriculture, Tokyo, Japan

Jacques Izopet
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Fauziah Mohd Jaafar
*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Andrew O. Jackson
China Agricultural University, Beijing, China

Daral J. Jackwood
*The Ohio State University/OARDC, Wooster, OH,
United States*

Jean-Rock Jacques
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Tiffany Jenkins
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University, Columbus,
OH, United States*

Jeffrey D. Jensen
Arizona State University, Tempe, AZ, United States

Daohong Jiang
Huazhong Agricultural University, Wuhan, China

Zhihao Jiang
China Agricultural University, Beijing, China

Allison R. Jilbert
The University of Adelaide, Adelaide, SA, Australia

Peng Jing
*Department of Chemistry, College of Arts and Sciences,
Fort Wayne, IN, United States*

Xixi Jing
Central South University, Changsha, China

Meesbah Jiwaji
Rhodes University, Grahamstown, South Africa

Kyle L. Johnson
*The University of Texas at El Paso, El Paso, TX,
United States and CQuentia, Memphis, TN, United
States*

Welkin E. Johnson
Boston College, Chestnut Hill, MA, United States

Ian M. Jones
*University of Reading, Reading, United Kingdom and
London School of Hygiene and Tropical Medicine,
London, United Kingdom*

Ramon Jordan
*Agricultural Research Service, US Department of
Agriculture, Beltsville, MD, United States*

Thomas Joris
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Ilkka Julkunen
*Institute of Biomedicine, University of Turku, Turku,
Finland*

Sandra Junglen
Charité - University Medicine Berlin, Berlin, Germany

Masanori Kaido
Kyoto University, Kyoto, Japan

Laura Kakkola
University of Turku, Turku, Finland

Hannimari Kallio-Kokko
*University of Helsinki and Helsinki University Hospital,
Helsinki, Finland*

Nassim Kamar
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Phyllis J. Kanki
*Harvard T.H. Chan School of Public Health, Boston,
MA, United States*

Peter Karayiannis
University of Nicosia, Nicosia, Cyprus

Henry M. Kariithi
*Kenya Agricultural and Livestock Research Organization,
Nairobi, Kenya*

Brian A. Kelch
*University of Massachusetts Medical School, Worcester,
MA, United States*

Karen E. Keller
*Horticultural Crops Research Unit, Agricultural Research
Service, US Department of Agriculture, Corvallis, OR,
United States*

Japhette E. Kembou-Ringert
University of Tel Aviv, Tel Aviv, Israel

Peter J. Kerr
*University of Sydney, Sydney, NSW, Australia and
CSIRO Health and Biosecurity, Black Mountain
Laboratories, Canberra, ACT, Australia*

Tiffany King
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University College
of Medicine, Columbus, OH, United States*

Andrea Kirmaier
Boston College, Chestnut Hill, MA, United States

Thomas Klose
Purdue University, West Lafayette, IN, United States

Barbara G. Klupp
*Friedrich-Loeffler-Institute, Greifswald-Insel Riems,
Germany*

David M. Knipe
Harvard Medical School, Boston, MA, United States

Nick J. Knowles
The Pirbright Institute, Pirbright, United Kingdom

Guus Koch
*Wageningen Bioveterinary Research, Lelystad,
The Netherlands*

Renate Koenig
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

Susanne E. Kohalmi
*The University of Western Ontario, London, ON,
Canada*

Hideki Kondo
Okayama University, Kurashiki, Japan

Jennifer L. Konopka-Anstadt
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Eugene V. Koonin
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States and
National Institutes of Health, Bethesda, MD, United States*

Marion P.G. Koopmans
Erasmus Medical Center, Rotterdam, The Netherlands

Richard Kormelink
*Wageningen University and Research, Wageningen,
The Netherlands*

Ioly Kotta-Loizou
Imperial College London, London, United Kingdom

Peter J. Krell
*Department of Molecular and Cellular Biology,
University of Guelph, Guelph, ON, Canada*

Mart Krupovic
Archaeal Virology Unit, Institut Pasteur, Paris, France

Andreas Kuhn
University of Hohenheim, Stuttgart, Germany

Jens H. Kuhn
*National Institutes of Health, Frederick, MD,
United States*

Richard J. Kuhn
Purdue University, West Lafayette, IN, United States

Suvi Kuivanen
University of Helsinki, Helsinki, Finland

Ranjababu Kulasegaram
*Guy's and St Thomas' NHS Foundation Trust, London,
United Kingdom*

Raghavendran Kulasegaran-Shylini
*Department of Pathogen Infection, Faculty of Infectious
and Tropical Diseases, London School of Hygiene and
Tropical Medicine, London, United Kingdom*

Gaurav Kumar
University of Delhi, New Delhi, India

Manish Kumar
Jawaharlal Nehru University, New Delhi, India

Gael Kurath
*US Geological Survey, Western Fisheries Research
Center, Seattle, WA, United States*

Satu Kurkela
*University of Helsinki and Helsinki University Hospital,
Helsinki, Finland*

Wan-Chun Lai
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Kevin Lamkiewicz
University of Jena, Jena, Germany

Rebecca K. Lane
*University of Texas Health Science Center at San
Antonio, San Antonio, TX, United States*

Andrew S. Lang
*Memorial University of Newfoundland, St. John's, NL,
Canada*

Daniel Carlos Ferreira Lanza
Federal University of Rio Grande do Norte, Natal, Brazil

Maija Lappalainen
*HUS Diagnostic Center, HUSLAB, Clinical
Microbiology, University of Helsinki and Helsinki
University Hospital, Helsinki, Finland*

Katherine LaTourrette
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Chris Lauber
*TWINCORE – Center for Experimental and Clinical
Infection Research, Hannover, Germany*

Antonio Lavazza
*The Lombardy and Emilia Romagna Experimental
Zootechnic Institute, Brescia, Italy*

C. Martin Lawrence
Montana State University, Bozeman, MT, United States

Hervé Lecoq
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Young-Min Lee
Utah State University, Logan, UT, United States

Kristen N. LeGault
University of California, Berkeley, CA, United States

James Legg
*International Institute of Tropical Agriculture, Dar es
Salaam, Tanzania*

Anne Legreve
University of Louvain, Louvain-la-Neuve, Belgium

Petr G. Leiman
The University of Texas Medical Branch, Galveston, TX, United States

Stanley M. Lemon
Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Sebastian Leptihn
Zhejiang University-Edinburgh University Institute, Zhejiang University, Haining, China

Dennis J. Lewandowski
University of Florida, Lake Alfred, FL, United States

Sébastien Lhomme
Toulouse University Hospital, Toulouse, France and Toulouse University Paul Sabatier, Toulouse, France

Dawei Li
China Agricultural University, Beijing, China

Guangdi Li
Central South University, Changsha, China

Guoqing Li
Huazhong Agricultural University, Wuhan, China

Yi Li
Peking University, Beijing, China

Zhefeng Li
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Zhenghe Li
Zhejiang University, Hangzhou, China

Jia Q. Liang
South China Agricultural University, Guangzhou, China

Sebastian Liebe
Institute of Sugar Beet Research, Göttingen, Germany

João Paulo Matos Santos Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Bruno Lina
HCL Department of Virology, National Reference Center for Respiratory Viruses, Institute of Infectious Agents, Croix-Rousse Hospital, Lyon, France and Virpath Laboratory, International Center of Research in Infectiology (CIRI), INSERM U1111, CNRS—UMR 5308, École Normale Supérieure de Lyon, University Claude Bernard Lyon, Lyon University, Lyon, France

Walter Ian Lipkin
Columbia University, New York, NY, United States

Jan G. Lisby
Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

Ding X. Liu
South China Agricultural University, Guangzhou, China

Qiang Liu
University of Saskatchewan, Saskatoon, SK, Canada

Sijun Liu
Iowa State University, Ames, IA, United States

Carlos Llorens
Biotechvana, Scientific Park University of Valencia, Valencia, Spain

L. Sue Loesch-Fries
Purdue University, West Lafayette, IN, United States

George P. Lomonosoff
John Innes Centre, Norwich, United Kingdom

L. Letti Lopez
The University of Texas at Austin, Austin, TX, United States

Alan T. Loynachan
University of Kentucky, Lexington, KY, United States

Garry A. Luke
University of St. Andrews, St. Andrews, United Kingdom

M. Luo
University of Alabama at Birmingham, Birmingham, AL, United States

Juan J. López-Moya
Center for Research in Agricultural Genomics and Spanish National Research Council, Barcelona, Spain

Che Ma
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Stuart A. MacFarlane
The James Hutton Institute, Invergowrie, United Kingdom

Saichetana Macherla
J. Craig Venter Institute, La Jolla, CA, United States

Kensaku Maejima
The University of Tokyo, Tokyo, Japan

Fabrizio Maggi
University of Pisa, Pisa, Italy and University of Insubria, Varese, Italy

Melissa S. Maginnis
The University of Maine, Orono, ME, United States

Edgar Maiss
Leibniz University Hannover, Hannover, Germany

Kira S. Makarova
National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, United States

Ariana Manglli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Annette Mankertz
Robert Koch-Institute, Berlin, Germany

Pilar Manrique
The Ohio State University, Wexner Medical Center, Columbus, OH, United States

Shahid Mansoor
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Marco Marklewitz
Institute of Virology, Charité – University Medicine Berlin, Berlin, Germany

Giovanni P. Martelli[†]
University of Bari Aldo Moro, Bari, Italy

Darren P. Martin
University of Cape Town, Cape Town, South Africa

Robert R. Martin
Horticultural Crops Research Unit, Agricultural Research Service, US Department of Agriculture, Corvallis, OR, United States

Manuel Martinez-Garcia
University of Alicante, Alicante, Spain

Francisco Martinez-Hernandez
University of Alicante, Alicante, Spain

Natalia Martín-González
Autonomous University of Madrid, Madrid, Spain

Joaquín Martínez Martínez
Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Manja Marz
University of Jena, Jena, Germany

Andrea Marzi
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, United States

Hema Masarapu
Sri Venkateswara University, Tirupati, India

Chikara Masuta
Hokkaido University, Sapporo, Japan

Carlos P. Mata
University of Leeds, Leeds, United Kingdom

Jelle Matthijnsens
Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

Claire P. Mattison
Centers for Disease Control and Prevention, Atlanta, GA, United States and Cherokee Nation Assurance, Arlington, VA, United States

William McAllister
Rowan University School of Osteopathic Medicine, Stratford, NJ, United States

Alison A. McBride
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Michael McChesney
University of California, Davis, CA, United States

Elaine McCulloch
Quality Control for Molecular Diagnostics (QCMD), Glasgow, United Kingdom

Andrew J. McMichael
University of Oxford, Oxford, United Kingdom

Alexander McPherson
University of California, Irvine, CA, United States

Irene K. Meki
French National Center for Scientific Research, Montpellier, France

Ulrich Melcher
Oklahoma State University, Stillwater, OK, United States

Tomas A. Melgarejo
University of California, Davis, CA, United States

Michael J. Melzer
Department of Plant and Environmental Protection Sciences, University of Hawaii, Honolulu, HI, United States

Luiza Mendonça
University of Oxford, Oxford, United Kingdom

Xiang-Jin Meng
Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Peter P.C. Mertens
University of Nottingham, Sutton Bonington, United Kingdom

[†]Deceased.

Thomas C. Mettenleiter
Friedrich-Loeffler-Institute, Greifswald-Insel Riems, Germany

Philip D. Minor
St Albans, United Kingdom

Ali Mirazimi
National Veterinary Institute, Uppsala, Sweden and Karolinska Hospital University, Huddinge, Sweden

Nischay Mishra
Columbia University, New York, NY, United States

Edward S. Mocarski
Emory University School of Medicine, Atlanta, GA, United States

Florian Mock
University of Jena, Jena, Germany

Volker Moennig
University of Veterinary Medicine, Hannover, Germany

Ian J. Molineux
The University of Texas at Austin, Austin, TX, United States

Aderito L. Monjane
Norwegian Veterinary Institute, Oslo, Finland

Jacen S. Moore
University of Tennessee Health Science Center, Memphis, TN, United States

Marc C. Morais
The University of Texas Medical Branch, Galveston, TX, United States

Cristina Moraru
Institute for Chemistry and Biology of the Marine Environment, Oldenburg, Germany

Hiromitsu Moriyama
Tokyo University of Agriculture and Technology, Tokyo, Japan

Sergey Y. Morozov
Lomonosov Moscow State University, Moscow, Russia

Thomas E. Morrison
University of Colorado School of Medicine, Aurora, CO, United States

Léa Morvan
University of Liège, Liège, Belgium

Benoît Moury
Plant Pathology Unit, INRAE – French National Institute for Agriculture, Food and Environment, Montfavet, France

Muhammad Mubin
University of Agriculture, Faisalabad, Pakistan

Nicolas J. Mueller
University Hospital of Zurich, Zurich, Switzerland

Emmanuelle Muller
The French Agricultural Research Center for International Development, Joint Research Units–Biology and Genetics of Plant-Pathogen Interactions, Montpellier, France and Biology and Genetics of Plant-Pathogen Interactions, University of Montpellier, The French Agricultural Research Center for International Development, French National Institute for Agricultural Research, Montpellier SupAgro, Montpellier, France

John S. Munday
Massey University, Palmerston North, New Zealand

Jacob H. Munson-McGee
Montana State University, Bozeman, MT, United States and Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Hacer Muratoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Kenan C. Murphy
University of Massachusetts Medical School, Worcester, MA, United States

Ugrappa Nagalakshmi
University of California, Davis, CA, United States

Keizo Nagasaki
Kochi University, Nankoku, Japan

Nazia Nahid
GC University, Faisalabad, Pakistan and University of Agriculture, Faisalabad, Pakistan

Venugopal Nair
The Pirbright Institute, Pirbright, United Kingdom

Remziye Nalçacıoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Shigetou Namba
The University of Tokyo, Tokyo, Japan

Rubab Z. Naqvi
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Rachel Nash
The Pirbright Institute, Surrey, United Kingdom

C.K. Navaratnarajah
Purdue University, West Lafayette, IN, United States

Maria A. Navarrete-Muñoz
Biotechvana, Madrid, Spain; Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid; and Rey Juan Carlos University Hospital, Móstoles, Spain

Jesús Navas-Castillo
Institute for Mediterranean and Subtropical Horticulture "La Mayora"—Spanish National Research Council—University of Malaga, Algarrobo-Costa, Málaga, Spain

Muhammad S. Nawaz-ul-Rehman
University of Agriculture, Faisalabad, Pakistan

Christopher L. Netherton
The Pirbright Institute, Pirbright, United Kingdom

Thu V.P. Nguyen
Baylor College of Medicine, Houston, TX, United States

Annette Niehl
Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

Hubert G.M. Niesters
Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Jozef I. Nissimov
University of Waterloo, Waterloo, ON, Canada

Norman Noah
London School of Hygiene and Tropical Medicine, London, United Kingdom

Mauricio L. Nogueira
São José do Rio Preto School of Medicine, São José do Rio Preto, São Paulo, Brazil

Johan Nordgren
Linköping University, Linköping, Sweden

C. Micha Nübling
Paul-Ehrlich-Institute, Langen, Germany

Visa Nurmi
University of Helsinki, Helsinki, Finland

Donald L. Nuss
University of Maryland, Rockville, MD, United States

M. Steven Oberste
Centers for Disease Control and Prevention, Atlanta, GA, United States

Hiroyuki Ogata
Institute for Chemical Research, Kyoto University, Kyoto, Japan

Ane Ogbe
University of Oxford, Oxford, United Kingdom

Hanna M. Oksanen
Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Graziele P. Oliveira
Federal University of Minas Gerais, Belo Horizonte, Brazil

Francesco Origi
University of Bern, Bern, Switzerland

Nikolaus Osterrieder
Free University of Berlin, Berlin, Germany

Robert A. Owens
Beltsville Agricultural Research Center, Beltsville, MD 20705, United States

Emine Özsahin
University of Guelph, Guelph, ON, Canada

Sergi Padilla-Parra
University of Oxford, Oxford, United Kingdom; Department of Infectious Diseases, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom; and Randall Division of Cell and Molecular Biophysics, King's College London, London, United Kingdom

Joshua Pajak
Duke University, Durham, NC, United States

Massimo Palmarini
MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Amanda R. Panfil
The Ohio State University, Columbus, OH, United States

Marcus Panning
Institute of Virology, Freiburg University Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Vitantonio Pantaleo
National Research Council, Research Unit of Bari, Bari, Italy

Anna Papa
Aristotle University of Thessaloniki, Thessaloniki, Greece

Nikolaos Pappas
Utrecht University, Utrecht, The Netherlands

Hanu R. Pappu
Washington State University, Pullman, WA, United States

Kristin N. Parent
*Michigan State University, East Lansing, MI,
United States*

Colin R. Parrish
Cornell University, Ithaca, NY, United States

A. Lorena Passarelli
Kansas State University, Manhattan, KS, United States

Basavaprabhu L. Patil
*ICAR–Indian Institute of Horticultural Research,
Bengaluru, India*

Jade Pattyn
University of Antwerp, Antwerp, Belgium

T.A. Paul
Cornell University, Ithaca, NY, United States

Lillian Pavlik
*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Susan L. Payne
Texas A&M University, College Station, TX, United States

Michael N. Pearson
The University of Auckland, Auckland, New Zealand

Mark E. Peeples
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University College of
Medicine, Columbus, OH, United States*

Ben Peeters
*Wageningen Bioveterinary Research, Lelystad,
The Netherlands*

Joseph S.M. Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Malik Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Judit J. Péntzes
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Miryam Pérez-Cañamás
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Marta Pérez-Illana
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Jaume Pérez-Sánchez
*Institute of Aquaculture Torre de la Sal, Spanish
National Research Council, Castellon, Spain*

Jean-Marie Peron
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Karin E. Peterson
National Institutes of Health, Hamilton, MT, United States

Karel Petrzik
*Biology Center CAS, Institute of Plant Molecular
Biology, České Budějovice, Czech Republic*

Mahtab Peyambari
*Pennsylvania State University, State College, PA,
United States*

Sujal Phadke
J. Craig Venter Institute, La Jolla, CA, United States

Hanh T. Pham
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Mauro Pistello
University of Pisa, Pisa, Italy

Daniel Ponndorf
John Innes Centre, Norwich, United Kingdom

Leo L.M. Poon
The University of Hong Kong, Pok Fu Lam, Hong Kong

Welkin H. Pope
University of Pittsburgh, Pittsburgh, PA, United States

Minna M. Poranen
University of Helsinki, Helsinki, Finland

Claudine Porta
*The Pirbright Institute, Pirbright, United Kingdom and
University of Oxford, Oxford, United Kingdom*

Samuel S. Porter
*National Institute of Allergy and Infectious Diseases,
National Institutes of Health, Bethesda, MD,
United States and University of Maryland, College Park,
MD, United States*

Frank A. Post
*King's College Hospital NHS Foundation Trust, London,
United Kingdom*

Nils Poulicard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

David Prangishvili
*Institut Pasteur, Paris, France and Ivane Javakhishvili
Tbilisi State University, Tbilisi, Georgia*

B. V. Venkataram Prasad
Baylor College of Medicine, Houston, TX, United States

Lalita Priyamvada
Centers for Disease Control and Prevention, Atlanta, GA, United States

Simone Prospero
Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

Elisabeth Puchhammer-Stöckl
Medical University of Vienna, Vienna, Austria

Jianming Qiu
University of Kansas Medical Center, Kansas City, KS, United States

S.L. Quackenbush
Colorado State University, Fort Collins, CO, United States

Killian J. Quinn
King's College Hospital, London, United Kingdom

Diego F. Quito-Avila
Department of Life Sciences, ESPOL Polytechnic University, Guayaquil, Ecuador

Frank Rabenstein
Julius Kühn Institute, Quedlinburg, Germany

Sheli R. Radoshitzky
United States Army Medical Research Institute of Infectious Diseases, Frederick, MD, United States

Saleem U. Rahman
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Mbolarinosa Rakotomalala
FOFIFA, Antananarivo, Madagascar

Norma Rallon
Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid and Rey Juan Carlos University Hospital, Móstoles, Spain

Robert P. Rambo
Diamond Light Source, Didcot, United Kingdom

Bertha Cecilia Ramirez
The Institute for Integrative Biology of the Cell, The French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

María D. Ramos-Barbero
University of Alicante, Alicante, Spain

Chris M. Rands
University of Geneva Medical School and Swiss Institute of Bioinformatics, Geneva, Switzerland

Venigalla B. Rao
The Catholic University of America, Washington, DC, United States

Janne J. Ravantti
University of Helsinki, Helsinki, Finland

Mandy Ravensbergen
Wageningen University and Research, Wageningen, The Netherlands

Georget Y. Reaiche-Miller
The University of Adelaide, Adelaide, SA, Australia

D.V.R. Reddy
International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

Vishwanatha R.A.P. Reddy
The Pirbright Institute, Pirbright, United Kingdom

Juan Reguera
Aix-Marseille University, French National Center for Scientific Research, Marseille, France and French National Institute of Health and Medical Research, Marseille, France

William K. Reisen
University of California, Davis, CA, United States

Jingshan Ren
University of Oxford, Oxford, United Kingdom

Renato O. Resende
University of Brasilia, Brasilia, Brazil

Peter A. Revill
The Peter Doherty Institute of Infection and Immunity, Royal Melbourne Hospital, Melbourne, VIC, Australia

Félix A. Rey
Institut Pasteur, Paris, France

Lara Rheinemann
University of Utah, Salt Lake City, UT, United States

Simone G. Ribeiro
Embrapa Genetic Resources and Biotechnology, Brasília, Brazil

Daniel Rigling
Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Cristina Risco
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Efraín E. Rivera-Serrano

Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Cécile Robin

INRAE – French National Research Institute for Agriculture, Food and Environment, UMR BIOGECO, Cestas, France

Rodrigo A.L. Rodrigues

Federal University of Minas Gerais, Belo Horizonte, Brazil

Elina Roine

University of Helsinki, Helsinki, Finland

Maria R. Rojas

University of California, Davis, CA, United States

Marilyn J. Roossinck

Pennsylvania State University, State College, PA, United States

Vera I.D. Ros

Wageningen University and Research, Wageningen, The Netherlands

Cristina Rosa

Pennsylvania State University, University Park, PA, United States

Hanna Rose

Leibniz University Hannover, Hannover, Germany

David A. Rosenbaum

University of Florida, Gainesville, FL, United States

Shannan L. Rossi

The University of Texas Medical Branch, Galveston, TX, United States

Michael G. Rossmann[†]

Purdue University, West Lafayette, IN, United States

L. Roux

University of Geneva Medical School, Geneva, Switzerland

Simon Roux

US Department of Energy Joint Genome Institute, Walnut Creek, CA, United States

J. Rovnak

Colorado State University, Fort Collins, CO, United States

David J. Rowlands

University of Leeds, Leeds, United Kingdom

Polly Roy

Department of Pathogen Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom and University of Reading, Reading, United Kingdom

Aaron P. Roznowski

The University of Texas at Austin, Austin, TX, United States and University of Arizona, Tucson, AZ, United States

Luisa Rubino

Institute for Sustainable Plant Protection, National Research Council, Bari, Italy

Olli Ruuskanen

Turku University Hospital, Turku, Finland

Eugene V. Ryabov

USDA, Agricultural Research Service, Beltsville, MD, United States

Martin D. Ryan

University of St. Andrews, St. Andrews, United Kingdom

Ki H. Ryu

Seoul Women's University, Seoul, South Korea

Hanns-Joachim Rziha

Eberhard Karls University of Tübingen, Tübingen, Germany

Sead Sabanadzovic

Mississippi State University, Starkville, MS, United States

Roghayeh Safari

Cellular and Molecular Epigenetics (GIGA), Liège, Belgium and Molecular Biology (TERRA), Gembloux, Belgium

Azeez Sait Sahul Hameed

C. Abdul Hakeem College, Melvisharam, India

Nicole Samies

University of Alabama at Birmingham, Birmingham, AL, United States

Carmen San Martín

National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Ruth-Anne Sandaa

Department of Biological Sciences, University of Bergen, Bergen, Norway

Hélène Sanfaçon

Agriculture and Agri-Food Canada, Summerland, BC, Canada

[†]Deceased.

Rafael Sanjuán
*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia-CSIC, Valencia, Spain*

Neeraja Sankaran
Utrecht University, Utrecht, The Netherlands

Fernando Santos
University of Alicante, Alicante, Spain

Cecilia Sarmiento
Tallinn University of Technology, Tallinn, Estonia

Takahide Sasaya
*National Agriculture and Food Research Organization,
Fukuyama, Japan*

Preethi Sathanantham
Virginia Tech, Blacksburg, VA, United States

Panayampalli S. Satheshkumar
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Yukiyo Sato
Okayama University, Kurashiki, Japan

Andreas Sauerbrei
Jena University Hospital, Jena, Germany

Eugene I. Savenkov
*Swedish University of Agricultural Sciences, Uppsala,
Sweden and Linnean Center for Plant Biology, Uppsala,
Sweden*

Carita Savolainen-Kopra
*National Institute for Health and Welfare, Helsinki,
Finland*

Kay Scheets
*Oklahoma State University, Stillwater, OK,
United States*

Uffe V. Schenider
*Copenhagen University Hospital Hvidovre, Hvidovre,
Denmark*

Richard H. Scheuermann
*J. Craig Venter Institute, La Jolla, CA, United States;
University of California, San Diego, CA, United States;
La Jolla Institute for Immunology, La Jolla, CA,
United States; and Global Virus Network, Baltimore,
MD, United States*

Manfred J. Schmitt
Saarland University, Saarbrücken, Germany

James E. Schoelz
University of Missouri, Columbia, MO, United States

Jason R. Schrad
*Michigan State University, East Lansing, MI,
United States*

Declan C. Schroeder
*University of Reading, Reading, United Kingdom and
University of Minnesota, St. Paul, MN, United States*

Stacey Schultz-Cherry
*St. Jude Children's Research Hospital, Memphis, TN,
United States*

Thomas F. Schulz
*Hannover Medical School, Institute of Virology,
Hannover, Germany and German Center for Infection
Research, Hannover-Braunschweig Site, Braunschweig,
Germany*

Catherine A. Scougall
The University of Adelaide, Adelaide, SA, Australia

Kimberley D. Seed
University of California, Berkeley, CA, United States

Joaquim Segalés
*Departament of Animal Health and Anatomy, Faculty of
Veterinary Medicine, Autonomous University of
Barcelona, Barcelona, Spain; Animal Health Research
Center (CReSA) – Institute of Agrifood Research and
Technology (IRTA), Campus UAB, Barcelona, Spain;
and OIE Collaborating Center for the Research and
Control of Emerging and Re-emerging Swine Diseases in
Europe (IRTA-CReSA), Barcelona, Spain*

Mateo Seoane-Blanco
National Center for Biotechnology, Madrid, Spain

Madhumati Sevvana
Purdue University, West Lafayette, IN, United States

Kazım Sezen
*Department of Biology, Karadeniz Technical University,
Trabzon, Turkey*

Arvind Sharma
Institut Pasteur, Paris, France

Sumit Sharma
Linköping University, Linköping, Sweden

James M. Sharp
*University of Zaragoza, Zaragoza, Spain and Edinburgh,
United Kingdom*

Qunxin She
Shandong University, Qingdao, China

Keith E. Shearwin
The University of Adelaide, Adelaide, SA, Australia

Hanako Shimura
Hokkaido University, Sapporo, Japan

Reina S. Sikkema
Erasmus Medical Center, Rotterdam, The Netherlands

Aaron Simkovich
Agriculture and Agri-Food Canada, London, ON,
Canada and The University of Western Ontario,
London, ON, Canada

Peter Simmonds
University of Oxford, Oxford, United Kingdom

Tarja Sironen
University of Helsinki, Helsinki, Finland

Susanna Sissonen
Finnish Institute for Health and Welfare, Helsinki,
Finland

Michael A. Skinner
Imperial College London, London, United Kingdom

Douglas E. Smith
University of California, San Diego, La Jolla, CA,
United States

Melvyn Smith
Viapath Analytics, Specialist Virology Centre, King's
College NHS Foundation Trust, London, United
Kingdom

Thomas J. Smith
The University of Texas Medical Branch, Galveston, TX,
United States

Teemu Smura
Helsinki University Hospital and University of Helsinki,
Helsinki, Finland

Eric J. Snijder
Leiden University Medical Center, Leiden,
The Netherlands

Gisela Soboll Hussey
Michigan State University, East Lansing, MI,
United States

Maria Söderlund-Venermo
University of Helsinki, Helsinki, Finland

Merike Sõmera
Tallinn University of Technology, Tallinn, Estonia

Eun G. Song
Seoul Women's University, Seoul, South Korea

Milan J. Sonneveld
Erasmus University Medical Center, Rotterdam,
The Netherlands

Beatriz Soriano
Biotechvana, Scientific Park University of Valencia and
Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain

Thomas E. Spencer
University of Missouri, Columbia, MO, United States

Pothur Sreenivasulu
Sri Venkateswara University, Tirupati, India

Ashley L. St. John
Duke-NUS Medical School, Singapore, Singapore

David K. Stammers
University of Oxford, Oxford, United Kingdom

John Stanley
John Innes Centre, Colney, United Kingdom

Glyn Stanway
University of Essex, Colchester, United Kingdom

Thilo Stehle
University of Tuebingen, Tuebingen, Germany and
Vanderbilt University School of Medicine, Nashville, TN,
United States

Gregory W. Stevenson
Iowa State University, Ames, IA, United States

Lucy Rae Stewart
Agricultural Research Service, US Department of
Agriculture, Wooster, OH, United States

C.C.M.M. Stijger
Wageningen University and Research Center, Bleiswijk,
The Netherlands

Peter G. Stockley
University of Leeds, Leeds, United Kingdom

David Stone
Weymouth Laboratory, Centre for Environment, Fisheries
and Aquaculture Science, Weymouth, United Kingdom

Ashley E Strother
The University of Texas Medical Branch, Galveston, TX,
United States

Sundharraman Subramanian
Michigan State University, East Lansing, MI,
United States

William C. Summers
Yale University, New Haven, CT, United States

Liyang Sun
Northwest A&F University, Yangling, China

Wesley I. Sundquist
University of Utah, Salt Lake City, UT, United States

Petri Susi
University of Turku, Turku, Finland

Curtis A. Suttle
University of British Columbia, Vancouver, BC, Canada

Nobuhiro Suzuki
Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

Lennart Svensson
Linköping University, Linköping, Sweden and Karolinska Institute, Stockholm, Sweden

Ronald Swansson
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Daniele M. Swetnam
University of California, Davis, CA, United States

Moriah L. Szpara
Pennsylvania State University, University Park, PA, United States

Keisuke Tabata
Heidelberg University, Heidelberg, Germany

Anna Taglienti
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Naoki Takeshita
Tokyo University of Agriculture and Technology, Fuchu, Japan

Kana Takeshita Urayama
Tokyo University of Agriculture and Technology, Fuchu, Japan

Michael E. Taliany
The James Hutton Institute, Dundee, United Kingdom

Pan Tao
The Catholic University of America, Washington, DC, United States

Jacqueline E. Tate
Centers for Disease Control and Prevention, Atlanta, GA, United States

Satyanarayana Tatineni
Agricultural Research Service, US Department of Agriculture, Lincoln, NE, United States and University of Nebraska–Lincoln, Lincoln, NE, United States

Sisko Tauriainen
University of Turku, Turku, Finland

Norbert Tautz
University of Luebeck, Luebeck, Germany

Paulo Tavares
Institute for Integrative Biology of the Cell, CEA, CNRS, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

Nicholas M.I. Taylor
University of Copenhagen, Copenhagen, Denmark

Xu Tengzhi
University of California, Davis, CA, United States

Raquel Tenorio
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Robert B. Tesh
The University of Texas Medical Branch, Galveston, TX, United States

Vaskar Thapa
Pennsylvania State University, State College, PA, United States

John E. Thomas
The University of Queensland, Brisbane, QLD, Australia

Julie A. Thomas
Rochester Institute of Technology, Rochester, NY, United States

Lynn C. Thomason
Frederick National Laboratory for Cancer Research, Frederick, MD, United States

Elizabeth Ashley Thompson
The University of Southern Mississippi, Hattiesburg, MS, United States

Jeremy R. Thompson
Cornell University, Ithaca, NY, United States

Antonio Tiberini
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Peter Tijssen
National Institute of Scientific Research – Armand-Frappier Health Research Centre, Microbiology and Immunology, Laval, QC, Canada

Yuji Tomaru
Japan Fisheries Research and Education Agency, Kanagawa, Japan

Laura Tomassoli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Ruben Torres
National Biotechnology Center–Spanish National Research Council, Madrid, Spain

Jia Q. Truong
The University of Adelaide, Adelaide, SA, Australia

Erkki Truve[†]
Tallinn University of Technology, Tallinn, Estonia

Chih-Hsuan Tsai
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Roman Tuma
University of Leeds, Leeds, United Kingdom and University of South Bohemia, České Budějovice, Czech Republic

Topi Turunen
Infectious Disease Unit, Espoo, Finland and Finnish Institute for Health and Welfare, Helsinki, Finland

Reidun Twarock
University of York, York, United Kingdom

Ioannis E. Tzanetakis
University of Arkansas, Fayetteville, United States

Antti Vaheri
University of Helsinki, Helsinki, Finland

Eeva J. Vainio
Natural Resources Institute Finland (Luke), Helsinki, Finland

Anna M. Vaira
Institute for Sustainable Plant Protection, National Research Council of Italy, Torino, Italy

Steven M. Valles
Center for Medical, Agricultural and Veterinary Entomology, Agricultural Research Service, US Department of Agriculture, Gainesville, FL, United States

Adrián Valli
National Center for Biotechnology-Spanish National Research Council, Madrid, Spain

Rodrigo A. Valverde
Louisiana State University Agricultural Center, Baton Rouge, United States

Pierre Van Damme
University of Antwerp, Antwerp, Belgium

Rene A.A. van der Vlugt
Wageningen University and Research Center, Wageningen, The Netherlands

Bernard A.M. Van der Zeijst
Leiden University Medical Center, Leiden, The Netherlands

Koenraad Van Doorslaer
University of Arizona, Tucson, AZ, United States

James L. Van Etten
University of Nebraska–Lincoln, Lincoln, NE, United States

Suzanne van Meer
University Medical Center Utrecht, Utrecht, The Netherlands

Monique M. van Oers
Wageningen University and Research, Wageningen, The Netherlands

Mark J. van Raaij
National Center for Biotechnology, Madrid, Spain

Marc H.V. Van Regenmortel
University of Strasbourg, Strasbourg, France

Piet A. van Rijn
Wageningen Bioveterinary Research, Lelystad, The Netherlands and North-West University, Potchefstroom, South Africa

Alain Vanderplasschen
University of Liège, Liège, Belgium

Dana L. Vanlandingham
College of Veterinary Medicine, Kansas State University, Manhattan, KS, United States

Olli Vapalahti
Helsinki University Hospital and University of Helsinki, Helsinki, Finland

Mark Varrelmann
Institute of Sugar Beet Research, Göttingen, Germany

Nikos Vasilakis
The University of Texas Medical Branch, Galveston, TX, United States

Michael Veit
Free University of Berlin, Berlin, Germany

Česlovas Venclovas
Vilnius University, Vilnius, Lithuania

H. Josef Vetten
Julius Kühn Institute, Braunschweig, Germany

Marli Vlok
University of British Columbia, Vancouver, BC, Canada

Anne-Nathalie Volkoff
Diversity, Genomes and Insects-Microorganisms Interactions, National Institute of Agricultural Research, University of Montpellier, Montpellier, France

Ian E.H. Voorhees
Cornell University, Ithaca, NY, United States

Alex Vorsters
University of Antwerp, Antwerp, Belgium

[†]Deceased.

Jonathan D.F. Wadsworth
UCL Institute of Prion Diseases, London, United Kingdom

Peter J. Walker
The University of Queensland, St. Lucia, QLD, Australia

Paul Wallace
*Quality Control for Molecular Diagnostics (QCMD),
Glasgow, United Kingdom*

Aiming Wang
*Agriculture and Agri-Food Canada, London, ON,
Canada*

Jen-Ren Wang
National Cheng Kung University, Tainan, Taiwan

Lin-Fa Wang
Duke-NUS Medical School, Singapore, Singapore

Nan Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiangxi Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiaofeng Wang
Virginia Tech, Blacksburg, VA, United States

Katherine N. Ward
University College London, London, United Kingdom

Matti Waris
University of Turku, Turku, Finland

Ranjit Warrier
Purdue University, West Lafayette, IN, United States

Daniel Watterson
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Marta L. Wayne
University of Florida, Gainesville, FL, United States

Friedemann Weber
*FB 10 – Institute for Virology, Justus Liebig University
Giessen, Giessen, Germany*

Sung-Chan Wei
*Institute of Molecular Biology, Academia Sinica, Taipei,
Taiwan*

Robin A. Weiss
University College London, London, United Kingdom

Tao Weitao
*Southwest Baptist University, Bolivar, MO,
United States*

Kerstin Wernike
Friedrich-Loeffler-Institute, Insel Riems, Germany

Rachel J. Whitaker
*University of Illinois at Urbana-Champaign, Urbana, IL,
United States*

K. Andrew White
York University, Toronto, ON, Canada

Anna E. Whitfield
*North Carolina State University, Raleigh, NC,
United States*

Richard Whitley
*University of Alabama at Birmingham, Birmingham,
AL, United States*

Reed B. Wickner
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Luc Willems
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Brian J. Willett
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Alexis Williams
*The University of Texas Medical Branch, Galveston, TX,
United States*

Stephen A. Winchester
*Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom*

Clayton W. Winkler
*National Institutes of Health, Hamilton, MT,
United States*

Stephan Winter
*Leibniz Institute – DSMZ – German Collection of
Microorganisms and Cell Cultures GmbH, Braunschweig,
Germany*

William M. Wintermantel
*Agricultural Research Service, US Department of
Agriculture, Salinas, CA, United States*

Jennifer Wirth
Montana State University, Bozeman, MT, United States

Yuri I. Wolf
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States*

Thorsten Wolff
Robert Koch Institute, Berlin, Germany

Blaide Woodburn
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Michael E. Woodson
The University of Texas Medical Branch, Galveston, TX, United States

Courtney Woolsey
The University of Texas Medical Branch, Galveston, TX, United States

Chien-Fu Wu
Tokyo University of Agriculture and Technology, Fuchu, Japan

Mingde Wu
Huazhong Agricultural University, Wuhan, China

Songsong Wu
Huazhong Agricultural University, Wuhan, China

Yan Xiang
University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

Jiatao Xie
Huazhong Agricultural University, Wuhan, China

Zhuang Xiong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Hajime. Yaegashi
Institute of Fruit Tree and Tea Science, NARO, Morioka, Japan

Mehtap Yakupoğlu
Trabzon University, Trabzon, Turkey

Yasuyuki Yamaji
The University of Tokyo, Tokyo, Japan

Meng Yang
China Agricultural University, Beijing, China

Teng-Chieh Yang
Scarsdale, NY, United States

Qin Yao
Jiangsu University, Zhenjiang, China

Tianyou Yao
Baylor College of Medicine, Houston, TX, United States

Nobuyuki Yoshikawa
Iwate University, Morioka, Japan

George R. Young
Francis Crick Institute, London, United Kingdom

Lawrence S. Young
University of Warwick, Coventry, United Kingdom

Mark J. Young
Montana State University, Bozeman, MT, United States

Ry Young
Texas A&M University, College Station, TX, United States

Isaac T. Younker
University of Alabama, Tuscaloosa, AL, United States

Qian Yu
School of Life Sciences, Jiangsu University, Zhenjiang, China

Sang-Im Yun
Utah State University, Logan, UT, United States

Fauzia Zarreen
University of Delhi, New Delhi, India

Francisco M. Zerbini
Federal University of Viçosa, Viçosa, Brazil

Dong-Xiu Zhang
University of Maryland, Rockville, MD, United States

Jianqiang Zhang
Iowa State University, Ames, IA, United States

Junjie Zhang
Texas A&M University, College Station, TX, United States

Long Zhang
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Pan Zhang
Central South University, Changsha, China

Peijun Zhang
University of Oxford, Oxford, United Kingdom and Electron Bio-Imaging Centre, Diamond Light Source, Didcot, United Kingdom

Tao Zhang
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Yongliang Zhang
China Agricultural University, Beijing, China

Zhenlu Zhang
Shandong Agricultural University, Tai'an, China

Lixia Zhou
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Ling Zhu
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Heiko Ziebell
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

John Ziebuhr
*The Queen's University of Belfast, Belfast,
United Kingdom*

Jeffrey J. Zimmerman
Iowa State University, Ames, IA, United States

Falk Zucker
*Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany*

CONTENT OF ALL VOLUMES

Editors in Chief	v
Editorial Board	vii
Section Editors	ix
Foreword	xv
Preface	xvii
Guide to Use	xix
List of Contributors	xxi

VOLUME 1

The Virus as a Concept – Fundamentals of Virology

A Brief History of Virology <i>David J Rowlands</i>	3
The Origin of Viruses <i>Patrick Forterre and Morgan Gaïa</i>	14
The Virocell Concept <i>Patrick Forterre</i>	23
Virus Taxonomy <i>Jens H Kuhn</i>	28
The Greater Virus World and Its Evolution <i>Eugene V Koonin and Valerian V Dolja</i>	38
The Virus Species Concept <i>Peter Simmonds</i>	47
Genetic Diversity and Evolution of Viral Populations <i>Rafael Sanjuán and Pilar Domingo-Calap</i>	53
Mechanisms of RNA Virus Evolution <i>Lisa M Bono and Siobain Duffy</i>	62
Mechanisms of DNA Virus Evolution <i>Moriah L Szpara and Koenraad Van Doorslaer</i>	71
Paleovirology <i>Clément Gilbert</i>	79
Evolution Steered by Structure <i>Nicola GA Abrescia</i>	87

Pairwise Sequence Comparison in Virology <i>Tao Zhang, Zheng Gong, Tongkun Guo, Zhuang Xiong, and Yiming Bao</i>	100
Computational Analysis of Recombination in Viral Nucleotide Sequences <i>Miguel Arenas</i>	108
Phylogeny of Viruses <i>Alexander E Gorbalenya and Chris Lauber</i>	116
Virus Bioinformatics <i>Nikolaos Pappas, Simon Roux, Martin Hölzer, Kevin Lamkiewicz, Florian Mock, Manja Marz, and Bas E Dutilh</i>	124
Metagenomics in Virology <i>Simon Roux, Jelle Matthijnsens, and Bas E Dutilh</i>	133
Database and Analytical Resources for Viral Research Community <i>Sujal Phadke, Saichetana Macherla, and Richard H Scheuermann</i>	141
Classification of the Viral World Based on Atomic Level Structures <i>Janne J Ravantti and Nicola GA Abrescia</i>	153
Isolating, Culturing, and Purifying Viruses With a Focus on Bacterial and Archaeal Viruses <i>Katri Eskelin and Hanna M Oksanen</i>	162
High Throughput Sequencing and Virology <i>Graham L Freimanis and Nick J Knowles</i>	175
Single-Virus Genomics: Studying Uncultured Viruses, One at a Time <i>Manuel Martinez-Garcia, Francisco Martinez-Hernandez, and Joaquín Martínez Martínez</i>	184
Biophysical Characterizations in the Solution State <i>Robert P Rambo and Katsuaki Inoue</i>	191
Virus Crystallography <i>Jonathan M Grimes</i>	199
Advanced Light and Correlative Microscopy in Virology <i>Sergi Padilla-Parra, Charles A Coomer, and Irene Carlon-Andres</i>	208
Atomic Force Microscopy (AFM) Investigation of Viruses <i>Alexander McPherson</i>	218
Cryo-Electron Microscopy (CEM) Structures of Viruses <i>David Chmielewski and Wah Chiu</i>	233
Analysis of Viruses in the Cellular Context by Electron Tomography <i>Peijun Zhang and Luiza Mendonça</i>	242
Mathematical Modeling of Virus Architecture <i>Reidun Twarock</i>	248
Principles of Virus Structure <i>Madhumati Sevvana, Thomas Klose, and Michael G Rossmann[†]</i>	257
Structures of Small Icosahedral Viruses <i>Elizabeth E Fry, Jingshan Ren, and Claudine Porta</i>	278
Structural Principles of the Flavivirus Particle Organization and of Its Conformational Changes <i>Stéphane Duquerroy, Arvind Sharma, and Félix A Rey</i>	290
Reoviruses (Reoviridae) and Their Structural Relatives <i>Liya Hu, Mary K Estes, and B V Venkataram Prasad</i>	303

[†]Deceased.

Structures of Tailed Phages and Herpesviruses (<i>Herpesviridae</i>) <i>Montserrat Fàbrega-Ferrer and Miquel Coll</i>	318
Adenoviruses (<i>Adenoviridae</i>) and Their Structural Relatives <i>Gabriela N Condezo, Natalia Martín-González, Marta Pérez-Illana, Mercedes Hernando-Pérez, José Gallardo, and Carmen San Martín</i>	329
Negative Single-Stranded RNA Viruses (<i>Mononegavirales</i>): A Structural View <i>Juan Reguera</i>	345
Structure of Retrovirus Particles (<i>Retroviridae</i>) <i>David K Stammers and Jingshan Ren</i>	352
Structure of Helical Viruses <i>C Martin Lawrence</i>	362
Giant Viruses and Their Virophage Parasites <i>Rodrigo AL Rodrigues, Ana CdSP Andrade, Grazielle P Oliveira, and Jônatas S Abrahão</i>	372
Viral Replication Cycle <i>AJ Cann</i>	382
Viral Receptors <i>José M Casasnovas and Thilo Stehle</i>	388
Bacterial and Archeal Virus Entry <i>Minna M Poranen and Aušra Domanska</i>	402
Nonenveloped Eukaryotic Virus Entry <i>Ian M Jones and Polly Roy</i>	409
Enveloped Virus Membrane Fusion <i>Aurélié A Albertini and Yves Gaudin</i>	417
Genome Replication of Bacterial and Archaeal Viruses <i>Česlovas Venclovas</i>	429
Viral Transcription <i>David LV Bauer and Ervin Fodor</i>	439
Translation of Viral Proteins <i>Martin D Ryan and Garry A Luke</i>	444
Recombination <i>Jozef J Bujarski</i>	460
Assembly of Viruses: Enveloped Particles <i>CK Navaratnarajah, R Warriar, and RJ Kuhn</i>	468
Assembly of Viruses: Nonenveloped Particles <i>M Luo</i>	475
Virion Assembly: From Small Picornaviruses (<i>Picornaviridae</i>) to Large Herpesviruses (<i>Herpesviridae</i>) <i>Ling Zhu, Nan Wang, and Xiangxi Wang</i>	480
Genome Packaging <i>Richard J Bingham, Reidun Twarock, Carlos P Mata, and Peter G Stockley</i>	488
Virus Factories <i>Isabel Fernández de Castro, Raquel Tenorio, and Cristina Risco</i>	495
Release of Phages From Prokaryotic Cells <i>Jesse Cahill and Ry Young</i>	501
Virus Budding <i>Lara Rheinemann and Wesley I Sundquist</i>	519

Vesicle-Mediated Transcytosis and Export of Viruses <i>Efraín E Rivera-Serrano and Stanley M Lemon</i>	529
Vector Transmission of Animal Viruses <i>Houssam Attoui, Fauziah Mohd Jaafar, Rennos Fragkoudis, and Peter PC Mertens</i>	542
The Human Virome <i>Alexia Bordigoni, Sébastien Halary, and Christelle Desnues</i>	552
Epidemiology of Human and Animal Viral Diseases <i>Michael Edelstein</i>	559
Zoonosis, Emerging and Re-Emerging Viral Diseases <i>Janet M Daly</i>	569
Antiviral Innate Immunity: Introduction <i>Friedemann Weber</i>	577
Humoral and T Cell-Mediated Immunity to Viruses <i>Ane Ogbe and Lucy Dorrell</i>	584
Antigenicity and Antigenic Variation <i>Kuan-Ying A Huang, Xiaorui Chen, Che Ma, Dayna Cheng, Jen-Ren Wang, and Wan-Chun Lai</i>	597
Antigen Presentation <i>Andrew J McMichael</i>	601
Defense Against Viruses and Other Genetic Parasites in Prokaryotes <i>Kira S Makarova, Yuri I Wolf, and Eugene V Koonin</i>	606
Defective-Interfering Viruses <i>L Roux</i>	617
Ecology and Global Impacts of Viruses <i>Joanne B Emerson</i>	621
The Role of Retroviruses in Cellular Evolution <i>Andrea Kirmaier and Welkin E Johnson</i>	627
The Role of Bacteriophages in Bacterial Evolution <i>Chris M Rands and Harald Brüssow</i>	633
Viruses and Their Potential for Bioterrorism <i>Dana L Vanlandingham and Stephen Higgs</i>	644
The Use of Viral Promoters in Expression Vectors <i>Ian M Jones</i>	652
Oncolytic Viruses <i>Laura Burga and Mihnea Bostina</i>	658
Biotechnology Approaches to Modern Vaccine Design <i>George P Lomonosoff and Daniel Ponndorf</i>	662
Viruses: Impact on Science and Society <i>Neeraja Sankaran and Robin A Weiss</i>	671

VOLUME 2

Viruses as Infectious Agents: Human and Animal Viruses

Adenoviruses (<i>Adenoviridae</i>) <i>Balázs Harrach and Mária Benkő</i>	3
---	---

African Horse Sickness Virus (<i>Reoviridae</i>) <i>Piet A van Rijn</i>	17
African Swine Fever Virus (<i>Asfarviridae</i>) <i>Linda K Dixon, Rachel Nash, Philippa C Hawes, and Christopher L Netherton</i>	22
Akabane Virus and Schmallenberg Virus (<i>Peribunyaviridae</i>) <i>Martin Beer and Kerstin Wernike</i>	34
Alphaviruses Causing Encephalitis (<i>Togaviridae</i>) <i>Diane E Griffin</i>	40
Anelloviruses (<i>Anelloviridae</i>) <i>Fabrizio Maggi and Mauro Pistello</i>	48
Animal Lentiviruses (<i>Retroviridae</i>) <i>Esperanza Gomez-Lucia</i>	56
Animal Morbilliviruses (<i>Paramyxoviridae</i>) <i>Carina Conceicao and Dalan Bailey</i>	68
Animal Papillomaviruses (<i>Papillomaviridae</i>) <i>John S Munday</i>	79
Astroviruses (<i>Astroviridae</i>) <i>Virginia Hargest, Amy Davis, and Stacey Schultz-Cherry</i>	92
Avian Hepadnaviruses (<i>Hepadnaviridae</i>) <i>Allison R Jilbert, Georget Y Reaiche-Miller, and Catherine A Scougall</i>	100
Avian Herpesviruses (<i>Herpesviridae</i>) <i>Vishwanatha RAP Reddy and Venugopal Nair</i>	112
Avian Influenza Viruses (<i>Orthomyxoviridae</i>) <i>Nicolas Bravo-Vasquez and Stacey Schultz-Cherry</i>	117
Avian Leukosis and Sarcoma Viruses (<i>Retroviridae</i>) <i>Karen L Beemon</i>	122
Bluetongue Virus (<i>Reoviridae</i>) <i>Raghavendran Kulasegaran-Shylini and Polly Roy</i>	127
Borna Disease Virus and Related Bornaviruses (<i>Bornaviridae</i>) <i>Susan L Payne</i>	137
Bovine Leukemia Virus (<i>Retroviridae</i>) <i>Thomas Joris, Roghaiyeh Safari, Jean-Rock Jacques, and Luc Willems</i>	144
Bovine Viral Diarrhea, Border Disease, and Classical Swine Fever Viruses (<i>Flaviviridae</i>) <i>Paul Becher, Volker Moennig, and Norbert Tautz</i>	153
Capripoxviruses, Parapoxviruses, and Other Poxviruses of Ruminants (<i>Poxviridae</i>) <i>Philippa M Beard</i>	165
Chikungunya Virus (<i>Togaviridae</i>) <i>Thomas E Morrison and Stephanie E Ander</i>	173
Circoviruses (<i>Circoviridae</i>) <i>Giovanni Franzo and Joaquim Segalés</i>	182
Coronaviruses: General Features (<i>Coronaviridae</i>) <i>Paul Britton</i>	193
Coronaviruses: Molecular Biology (<i>Coronaviridae</i>) <i>X Deng and SC Baker</i>	198

Crimean-Congo Hemorrhagic Fever Virus and Nairoviruses of Medical Importance (<i>Nairoviridae</i>) <i>Ali Mirazimi, Felicity Burt, and Anna Papa</i>	208
Dengue Viruses (<i>Flaviviridae</i>) <i>Ashley L. St. John and Duane J. Gubler</i>	218
Ebola Virus (<i>Filoviridae</i>) <i>Andrea Marzi and Logan Banadyga</i>	232
Enteroviruses (<i>Picornaviridae</i>) <i>Carita Savolainen-Kopra, Soile Blomqvist, and Petri Susi</i>	245
Enveloped, Positive-Strand RNA Viruses (<i>Nidovirales</i>) <i>L. Enjuanes, AE Gorbalenya, RJ de Groot, JA Cowley, J Ziebuhr, and EJ Snijder</i>	256
Epstein-Barr Virus (<i>Herpesviridae</i>) <i>Lawrence S. Young</i>	267
Equine Herpesviruses (<i>Herpesviridae</i>) <i>Gisela Soboll-Hussey, Nikolaus Osterrieder, and Walid Azab</i>	278
Equine, Canine, and Swine Influenza (<i>Orthomyxoviridae</i>) <i>Janet M. Daly and Japhette E. Kembou-Ringert</i>	287
Feline Calicivirus (<i>Caliciviridae</i>) <i>Margaret J. Hosie and Michaela J. Conley</i>	294
Feline Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>Brian J. Willett and Margaret J. Hosie</i>	300
Fish and Amphibian Alloherpesviruses (<i>Herpesviridae</i>) <i>Maxime Boutier, Léa Morvan, Natacha Delrez, Francesco Origi, Andor Doszpoly, and Alain Vanderplasschen</i>	306
Fish Retroviruses (<i>Retroviridae</i>) <i>TA Paul, RN Casey, PR Bowser, JW Casey, J Rovnak, and SL Quackenbush</i>	316
Fish Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Gael Kurath and David Stone</i>	324
Foot-and-Mouth Disease Viruses (<i>Picornaviridae</i>) <i>David J. Rowlands</i>	332
Fowlpox Virus and Other Avipoxviruses (<i>Poxviridae</i>) <i>Efstathios S. Giotis and Michael A. Skinner</i>	343
Hantaviruses (<i>Hantaviridae</i>) <i>Tarja Sironen and Antti Vaheri</i>	349
Henipaviruses (<i>Paramyxoviridae</i>) <i>Lin-Fa Wang and Danielle E. Anderson</i>	355
Hepatitis A Virus (<i>Picornaviridae</i>) <i>Andreas Dotzauer</i>	362
Hepatitis B Virus (<i>Hepadnaviridae</i>) <i>Peter Karayiannis</i>	373
Hepatitis C Virus (<i>Flaviviridae</i>) <i>Ralf Bartenschlager and Keisuke Tabata</i>	386
Hepeviruses (<i>Hepeviridae</i>) <i>Xiang-Jin Meng</i>	397
Herpes Simplex Virus 1 and 2 (<i>Herpesviridae</i>) <i>David M. Knipe and Richard Whitley</i>	404

History of Virology: Vertebrate Viruses <i>F Fenner</i>	414
Human Boca- and Protoparvoviruses (<i>Parvoviridae</i>) <i>Maria Söderlund-Venermo and Jianming Qiu</i>	419
Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (<i>Coronaviridae</i>) <i>Ding X Liu, Jia Q Liang, and To S Fung</i>	428
Human Cytomegalovirus (<i>Herpesviridae</i>) <i>Edward S Mocarski</i>	441
Human Immunodeficiency Virus (<i>Retroviridae</i>) <i>Blaide Woodburn, Ann Emery, and Ronald Swanstrom</i>	460
Human Metapneumovirus (<i>Pneumoviridae</i>) <i>Antonella Casola, Matteo P Garofalo, and Xiaoyong Bao</i>	475
Human Norovirus and Sapovirus (<i>Caliciviridae</i>) <i>Sumit Sharma, Marie Hagbom, Lennart Svensson, and Johan Nordgren</i>	483
Human Papillomaviruses (<i>Papillomaviridae</i>) <i>Alison A McBride and Samuel S Porter</i>	493
Human Parainfluenza Viruses (<i>Paramyxoviridae</i>) <i>Elisabeth Adderson</i>	502
Human Pathogenic Arenaviruses (<i>Arenaviridae</i>) <i>Sheli R Radoshitzky and Juan C de la Torre</i>	507
Human Polyomaviruses (<i>Papillomaviridae</i>) <i>Melissa S Maginnis</i>	518
Human T-Cell Leukemia Virus-1 and -2 (<i>Retroviridae</i>) <i>Amanda R Panfil and Patrick L Green</i>	528
Infectious Bursal Disease Virus (<i>Birnaviridae</i>) <i>Daral J Jackwood</i>	540
Infectious Pancreatic Necrosis Virus (<i>Birnaviridae</i>) <i>Øystein Evensen</i>	544
Influenza A Viruses (<i>Orthomyxoviridae</i>) <i>Laura Kakkola, Niina Ikonen, and Ilkka Julkunen</i>	551
Influenza B, C and D Viruses (<i>Orthomyxoviridae</i>) <i>Thorsten Wolff and Michael Veit</i>	561
Jaagsiekte Sheep Retrovirus (<i>Retroviridae</i>) <i>James M Sharp, Marcelo De las Heras, Massimo Palmarini, and Thomas E Spencer</i>	575
Japanese Encephalitis Virus (<i>Flaviviridae</i>) <i>Sang-Im Yun and Young-Min Lee</i>	583
Kaposi's Sarcoma-Associated Herpesvirus (<i>Herpesviridae</i>) <i>Anne K Cordes and Thomas F Schulz</i>	598
Marburg and Ravn Viruses (<i>Filoviridae</i>) <i>Courtney Woolsey, Thomas W Geisbert, and Robert W Cross</i>	608
Measles Virus (<i>Paramyxoviridae</i>) <i>Roberto Cattaneo and Michael McChesney</i>	619
Molluscum Contagiosum Virus (<i>Poxviridae</i>) <i>Joachim J Bugert and Rosina Ehmann</i>	629

Mumps Virus (<i>Paramyxoviridae</i>) <i>Stephen A Winchester and Kevin E Brown</i>	634
Murine Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>George R Young and Kate N Bishop</i>	643
Newcastle Disease Virus (<i>Paramyxoviridae</i>) <i>Ben Peeters and Guus Koch</i>	648
Orthobunyaviruses (<i>Peribunyaviridae</i>) <i>Alyssa B Evans, Clayton W Winkler, and Karin E Peterson</i>	654
Parapoxviruses (<i>Poxviridae</i>) <i>Hanns-Joachim Rziha and Mathias Büttner</i>	666
Parechoviruses (<i>Picornaviridae</i>) <i>Sisko Tauriainen and Glyn Stanway</i>	675
Parvoviruses of Carnivores, and the Emergence of Canine Parvovirus (<i>Parvoviridae</i>) <i>Colin R Parrish, Ian EH Voorhees, and Susan L Hafenstein</i>	683
Polioviruses (<i>Picornaviridae</i>) <i>Philip D Minor</i>	688
Porcine Reproductive and Respiratory Syndrome Virus and Equine Arteritis Virus (<i>Arteriviridae</i>) <i>Jianqiang Zhang, Alan T Loynachan, Gregory W Stevenson, and Jeffrey J Zimmerman</i>	697
Prions of Vertebrates <i>Jonathan DF Wadsworth and John Collinge</i>	707
Pseudorabies Virus (<i>Herpesviridae</i>) <i>Thomas C Mettenleiter and Barbara G Klupp</i>	714
Rabbit Hemorrhagic Disease Virus and European Brown Hare Syndrome Virus (<i>Caliciviridae</i>) <i>Lorenzo Capucci, Patrizia Cavadini, and Antonio Lavazza</i>	724
Rabbit Myxoma Virus and the Fibroma Viruses (<i>Poxviridae</i>) <i>Peter J Kerr</i>	730
Rabies and Other Lyssaviruses (<i>Rhabdoviridae</i>) <i>Ashley C Banyard and Anthony R Fooks</i>	738
Respiratory Syncytial Virus (<i>Pneumoviridae</i>) <i>Tiffany King, Tiffany Jenkins, Supranee Chaiwatpongsakorn, and Mark E Peebles</i>	747
Rhinoviruses (<i>Picornaviridae</i>) <i>Matti Waris and Olli Ruuskanen</i>	757
Rift Valley Fever Virus and Other Phleboviruses (<i>Phenuiviridae</i>) <i>Tetsuro Ikegami</i>	765
Roseoloviruses: Human Herpesviruses 6A, 6B and 7 (<i>Herpesviridae</i>) <i>Katherine N Ward</i>	778
Rotaviruses (<i>Reoviridae</i>) <i>Juana Angel and Manuel A Franco</i>	789
Rubella Virus (<i>Picornaviridae</i>) <i>Annette Mankertz</i>	797
Saint Louis Encephalitis Virus (<i>Flaviviridae</i>) <i>William K Reisen, Lark L Coffey, Daniele M Swetnam, and Aaron C Brault</i>	805
Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (<i>Coronaviridae</i>) <i>Joseph SM Peiris and Leo LM Poon</i>	814

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (<i>Coronaviridae</i>) <i>Malik Peiris</i>	825
Simian Immunodeficiency Virus (SIV) and HIV-2 (<i>Retroviridae</i>) <i>Phyllis J Kanki</i>	827
Sindbis Virus (<i>Togaviridae</i>) <i>Satu Kurkela</i>	837
Tick-Borne Encephalitis Virus (<i>Flaviviridae</i>) <i>Teemu Smura, Suvi Kuivanen, and Olli Vapalahti</i>	843
Transmissible Gastroenteritis Virus of Pigs and Porcine Epidemic Diarrhea Virus (<i>Coronaviridae</i>) <i>Qiang Liu and Volker Gerds</i>	850
Vaccinia Virus (<i>Poxviridae</i>) <i>Yan Xiang and Rebecca K Lane</i>	854
Varicella-Zoster Virus (<i>Herpesviridae</i>) <i>Jeffrey I Cohen</i>	860
Variola and Monkeypox Viruses (<i>Poxviridae</i>) <i>Lalita Priyamvada and Panayampalli S Satheshkumar</i>	868
Vesicular Stomatitis Virus and Bovine Ephemeral Fever Virus (<i>Rhabdoviridae</i>) <i>Peter J Walker and Robert B Tesh</i>	875
West Nile Virus (<i>Flaviviridae</i>) <i>Fengwei Bai and Elizabeth Ashley Thompson</i>	884
Yellow Fever Virus (<i>Flaviviridae</i>) <i>Ashley E Strother and Alan DT Barrett</i>	891
Zika Virus (<i>Flaviviridae</i>) <i>Nikos Vasilakis, Shannan L Rossi, Sasha R Azar, Irma E Cisneros, Cassia F Estofolete, and Mauricio L Nogueira</i>	899

VOLUME 3

Viruses as Infectious Agents: Plant Viruses

An Introduction to Plant Viruses <i>Roger Hull</i>	3
Emerging and Re-Emerging Plant Viruses <i>Sabrina Bertin, Francesco Faggioli, Andrea Gentili, Ariana Manglli, Anna Taglienti, Antonio Tiberini, and Laura Tomassoli</i>	8
Emerging Geminiviruses (<i>Geminiviridae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, and Muhammad Mubin</i>	21
Movement of Viruses in Plants <i>Manfred Heinlein</i>	32
Plant Antiviral Defense: Gene-Silencing Pathways <i>Vitantonio Pantaleo, Chikara Masuta, and Hanako Shimura</i>	43
Plant Resistance to Viruses: Engineered Resistance <i>Marc Fuchs</i>	52
Plant Resistance to Viruses: Natural Resistance Associated With Dominant Genes <i>Mandy Ravensbergen and Richard Kormelink</i>	60

Plant Resistance to Viruses: Natural Resistance Associated With Recessive Genes <i>Masayoshi Hashimoto, Kensaku Maejima, Yasuyuki Yamaji, and Shigetou Namba</i>	69
Plant Viral Diseases: Economic Implications <i>Basavaprabhu L Patil</i>	81
Retrotransposons of Plants <i>M-A Grandbastien</i>	98
Vector Transmission of Plant Viruses <i>Etienne Herrbach and Quentin Chesnais</i>	106
Viral Suppressors of Gene Silencing <i>Hernan Garcia-Ruiz</i>	116
Virus-Induced Gene Silencing (VIGS) <i>Xu Tengzhi, Ugrappa Nagalakshmi, and Savithramma P Dinesh-Kumar</i>	123
Alfalfa Mosaic Virus (<i>Bromoviridae</i>) <i>L Sue Loesch-Fries</i>	132
Alphaflexiviruses (<i>Alphaflexiviridae</i>) <i>Sergey Y Morozov and Alexey A Agranovsky</i>	140
Alphasatellites (<i>Alphasatellitidae</i>) <i>Rob W Briddon and Muhammad S Nawaz-ul-Rehman</i>	149
Amalgaviruses (<i>Amalgaviridae</i>) <i>Ioannis E Tzanetakis, Sead Sabanadzovic, and Rodrigo A Valverde</i>	154
Badnaviruses (<i>Caulimoviridae</i>) <i>Andrew DW Geering</i>	158
Banana Bunchy Top Virus (<i>Nanoviridae</i>) <i>John E Thomas</i>	169
Barley Yellow Dwarf Viruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	176
Bean Common Mosaic Virus and Bean Common Mosaic Necrosis Virus (<i>Potyviridae</i>) <i>Ramon Jordan and John Hammond</i>	184
Bean Golden Mosaic Virus and Bean Golden Yellow Mosaic Virus (<i>Geminiviridae</i>) <i>Francisco M Zerbini and Simone G Ribeiro</i>	192
Beet Curly Top Virus (<i>Geminiviridae</i>) <i>Robert L Gilbertson, Tomas A Melgarejo, Maria R Rojas, William M Wintermantel, and John Stanley</i>	200
Beet Necrotic Yellow Vein Virus (<i>Benyviridae</i>) <i>Sebastian Liebe, Annette Niehl, Renate Koenig, and Mark Varrelmann</i>	213
Benyviruses (<i>Benyviridae</i>) <i>Annette Niehl, Sebastian Liebe, Mark Varrelmann, and Renate Koenig</i>	219
Betaflexiviruses (<i>Betaflexiviridae</i>) <i>Nobuyuki Yoshikawa and Hajime Yaegashi</i>	229
Betasatellites and Deltasatellites (<i>Tolecasatellitidae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, Muhammad Hassan, and Muhammad Mubin</i>	239
Bluner-, Cile-, and Higreviruses (<i>Kitaviridae</i>) <i>Diego F Quito-Avila, Juliana Freitas-Astúa, and Michael J Melzer</i>	247
Brome Mosaic Virus (<i>Bromoviridae</i>) <i>Guijuan He, Zhenlu Zhang, Preethi Sathanantham, Arturo Diaz, and Xiaofeng Wang</i>	252

Bromoviruses (<i>Bromoviridae</i>) <i>Jozef J Bujarski</i>	260
Bymoviruses (<i>Potyviridae</i>) <i>Annette Niehl and Frank Rabenstein</i>	268
Cacao Swollen Shoot Virus (<i>Caulimoviridae</i>) <i>Emmanuelle Muller</i>	274
Carmo-Like Viruses (<i>Tombusviridae</i>) <i>Miryam Pérez-Cañamás and Carmen Hernández</i>	285
Cassava Brown Streak Viruses (<i>Potyviridae</i>) <i>Basavaprabhu L Patil</i>	293
Cassava Mosaic Viruses (<i>Geminiviridae</i>) <i>James Legg and Stephan Winter</i>	301
Caulimoviruses (<i>Caulimoviridae</i>) <i>James E Schoelz and Mustafa Adhab</i>	313
Cheraviruses, Sadwaviruses and Torradoviruses (<i>Secoviridae</i>) <i>Toru Iwanami and René AA van der Vlugt</i>	322
Citrus Tristeza Virus (<i>Closteroviridae</i>) <i>Moshe Bar-Joseph, Scott J Harper, and William O Dawson</i>	327
Closteroviruses (<i>Closteroviridae</i>) <i>Marc Fuchs</i>	336
Comoviruses and Fabaviruses (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	348
Cotton Leaf Curl Disease (<i>Geminiviridae</i>) <i>Nasim Ahmed, Imran Amin, and Shahid Mansoor</i>	355
Cowpea Mosaic Virus (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	364
Cucumber Mosaic Virus (<i>Bromoviridae</i>) <i>Judith Hirsch and Benoît Moury</i>	371
Dianthovirus (<i>Tombusviridae</i>) <i>Kiwamu Hyodo and Masanori Kaido</i>	383
Endornaviruses (<i>Endornaviridae</i>) <i>Toshiyuki Fukuhara</i>	388
Fimoviruses (<i>Fimoviridae</i>) <i>Toufic Elbeaino and Michele Digiaro</i>	396
Furoviruses (<i>Virgaviridae</i>) <i>Annette Niehl and Renate Koenig</i>	405
Geminiviruses (<i>Geminiviridae</i>) <i>Jesús Navas-Castillo and Elvira Fiallo-Olivé</i>	411
Hordeiviruses (<i>Virgaviridae</i>) <i>Zhihao Jiang, Meng Yang, Yongliang Zhang, Andrew O Jackson, and Dawei Li</i>	420
Idaeoviruses (<i>Mayoviridae</i>) <i>Robert R Martin and Karen E Keller</i>	430
Ilarviruses (<i>Bromoviridae</i>) <i>Aaron Simkovich, Susanne E Kohalmi, and Aiming Wang</i>	439

Luteoviruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	447
Machlomovirus and Panicoviruses (<i>Tombusviridae</i>) <i>Kay Scheets</i>	456
Maize Streak Virus (<i>Geminiviridae</i>) <i>Darren P Martin and Aderito L Monjane</i>	461
Nanoviruses (<i>Nanoviridae</i>) <i>Bruno Gronenborn and H Josef Vetten</i>	470
Necro-Like Viruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Giovanni P Martelli[†]</i>	481
Nepoviruses (<i>Secoviridae</i>) <i>Hélène Sanfaçon</i>	486
Ophioviruses (<i>Aspiviridae</i>) <i>Anna M Vaira and John Hammond</i>	495
Orthospoviruses (<i>Tospoviridae</i>) <i>Renato O Resende and Hanu R Pappu</i>	507
Ourmiaviruses (<i>Botourmiaviridae</i>) <i>Gian Paolo Accotto and Cristina Rosa</i>	516
Papaya Ringspot Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	520
Pecluviruses (<i>Virgaviridae</i>) <i>Hema Masarapu, Pothur Sreenivasulu, Philippe Delfosse, Claude Bragard, Anne Legreve, and DVR Reddy</i>	528
Pepino Mosaic Virus (<i>Alphaflexiviridae</i>) <i>Rene AA van der Vlugt and CCMM Stijger</i>	539
Plant Reoviruses (<i>Reoviridae</i>) <i>Yu Huang and Yi Li</i>	545
Plant Resistance to Geminiviruses <i>Basavaprabhu L Patil, Supriya Chakraborty, Henryk Czosnek, Elvira Fiallo-Olivé, Robert L Gilbertson, James Legg, Shahid Mansoor, Jesús Navas-Castillo, Rubab Z Naqvi, Saleem U Rahman, and Francisco M Zerbini</i>	554
Plant Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Ralf G Dietzgen, Michael M Goodin, and Zhenghe Li</i>	567
Plant Satellite Viruses (<i>Albetovirus, Aumaivirus, Papanivirus, Virtovirus</i>) <i>Mart Krupovic</i>	581
Plum Pox Virus (<i>Potyviridae</i>) <i>Miroslav Glasa and Thierry Candresse</i>	586
Poleroviruses (<i>Luteoviridae</i>) <i>Hernan Garcia-Ruiz, Natalie M Holste, and Katherine LaTourrette</i>	594
Pomoviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	603
Potato Virus Y (<i>Potyviridae</i>) <i>Laurent Glais and Benoît Moury</i>	612
Potexviruses (<i>Alphaflexiviridae</i>) <i>Ki H Ryu, Eun G Song, and Jin S Hong</i>	623

[†]Deceased.

Potyviruses (<i>Potyviridae</i>) <i>Adrián Valli, Juan A García, and Juan J López-Moya</i>	631
Quinviruses (<i>Betaflexiviridae</i>) <i>Ki H Ryu and Eun G Song</i>	642
Reverse-Transcribing Viruses (<i>Belpaoviridae, Metaviridae, and Pseudoviridae</i>) <i>Carlos Llorens, Beatriz Soriano, Maria A Navarrete-Muñoz, Ahmed Hafez, Vicente Arnau, Jose Miguel Benito, Toni Gabaldon, Norma Rallon, Jaume Pérez-Sánchez, and Mart Krupovic</i>	653
Rice Tungro Disease (<i>Secoviridae, Caulimoviridae</i>) <i>Gaurav Kumar, Fauzia Zarreen, and Indranil Dasgupta</i>	667
Rice Yellow Mottle Virus (<i>Solemoviridae</i>) <i>Eugénie Hébrard, Nils Poulicard, and Mbolarinosy Rakotomalala</i>	675
Satellite Nucleic Acids and Viruses <i>Olufemi J Alabi, Alfredo Diaz-Lara, and Maher Al Rwahnih</i>	681
Secoviruses (<i>Secoviridae</i>) <i>Jeremy R Thompson</i>	692
Sequiviruses and Waikaviruses (<i>Secoviridae</i>) <i>Lucy Rae Stewart</i>	703
Solemoviruses (<i>Solemoviridae</i>) <i>Cecilia Sarmiento, Merike Sõmera, and Erkki Truve[†]</i>	712
Tenuiviruses (<i>Phenuiviridae</i>) <i>Bertha Cecilia Ramirez and Anne-Lise Haenni</i>	719
Tobacco Mosaic Virus (<i>Virgaviridae</i>) <i>Marc HV Van Regenmortel</i>	727
Tobamoviruses (<i>Virgaviridae</i>) <i>Ulrich Melcher, Dennis J Lewandowski, and William O Dawson</i>	734
Tobraviruses (<i>Virgaviridae</i>) <i>Stuart A MacFarlane</i>	743
Tomato Leaf Curl New Delhi Virus (<i>Geminiviridae</i>) <i>Supriya Chakraborty and Manish Kumar</i>	749
Tomato Spotted Wilt Virus (<i>Tospoviridae</i>) <i>Hanu R Pappu, Anna E Whitfield, and Athos S de Oliveira</i>	761
Tomato Yellow Leaf Curl Viruses (<i>Geminiviridae</i>) <i>Henryk Czosnek</i>	768
Tombusvirus-Like Viruses (<i>Tombusviridae</i>) <i>K Andrew White</i>	778
Tombusviruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Kay Scheets</i>	788
Tritimoviruses and Rymoviruses (<i>Potyviridae</i>) <i>Satyanarayana Tatineni and Gary L Hein</i>	797
Triviruses (<i>Betaflexiviridae</i>) <i>Yahya ZA Gaafar and Heiko Ziebell</i>	805
Tymoviruses (<i>Tymoviridae</i>) <i>Rosemarie W Hammond and Peter Abrahamian</i>	818

[†]Deceased.

Umbraviruses (<i>Tombusviridae</i>) <i>Eugene V Ryabov and Michael E Taliansky</i>	827
Varicosaviruses (<i>Rhabdoviridae</i>) <i>Takahide Sasaya</i>	833
Virgaviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	839
Viroids (<i>Pospiviroidae</i> and <i>Avsunviroidae</i>) <i>Ricardo Flores and Robert A Owens</i>	852
Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	862

VOLUME 4

Viruses as Infectious Agents: Bacterial, Archaeal, Fungal, Algal, and Invertebrate Viruses

Bacterial Viruses

History of Virology: Bacteriophages <i>William C Summers</i>	3
Icosahedral Phages – Single-Stranded DNA (ϕ X174) <i>Bentley A Fane and Aaron P Roznowski</i>	10
Single-Stranded RNA Bacterial Viruses <i>Peter G Stockley and Junjie Zhang</i>	21
Enveloped Icosahedral Phages – Double-Stranded RNA (ϕ 6) <i>Paul Gottlieb and Aleksandra Alimova</i>	26
Membrane-Containing Icosahedral DNA Bacteriophages <i>Roman Tuma, Sarah J Butcher, and Hanna M Oksanen</i>	36
Tailed Double-Stranded DNA Phages <i>Robert L Duda</i>	45
Helical and Filamentous Phages <i>Andreas Kuhn and Sebastian Leptihn</i>	53
Replication of <i>Bacillus</i> Double-Stranded DNA Bacteriophages <i>Silvia Ayora, Paulo Tavares, Ruben Torres, and Juan C Alonso</i>	61
Lytic Transcription <i>William McAllister and Deborah M Hinton</i>	69
Lysogeny <i>Keith E Shearwin and Jia Q Truong</i>	77
Decision Making by Temperate Phages <i>Ido Golding, Seth Coleman, Thu VP Nguyen, and Tianyou Yao</i>	88
Mobilization of Phage Satellites <i>Kristen N LeGault and Kimberley D Seed</i>	98
Portal Vertex <i>Peng Jing and Mauricio Cortes Jr.</i>	105

Prohead, the Head Shell Pre-Cursor <i>Marc C Morais and Michael E Woodson</i>	115
Enzymology of Viral DNA Packaging Machines <i>Carlos E Catalano</i>	124
DNA Packaging: DNA Recognition <i>Sandra J Greive and Oliver W Bayfield</i>	136
DNA Packaging: The Translocation Motor <i>Janelle A Hayes and Brian A Kelch</i>	148
Biophysics of DNA Packaging <i>Joshua Pajak, Gaurav Arya, and Douglas E Smith</i>	160
Energetics of the DNA-Filled Head <i>Alex Evilevitch</i>	167
Bacteriophage Receptor Proteins of Gram-Negative Bacteria <i>Sarah M Doore, Kristin N Parent, Sundharraman Subramanian, Jason R Schrad, and Natalia B Hubbs</i>	175
Tail Structure and Dynamics <i>Shweta Bhatt, Petr G Leiman, and Nicholas MI Taylor</i>	186
Bacteriophage Tail Fibres, Tailspikes, and Bacterial Receptor Interaction <i>Mateo Seoane-Blanco, Mark J van Raaij, and Meritxell Granell</i>	194
Phage Genome and Protein Ejection In Vivo <i>Ian J Molineux, L Letti Lopez, and Aaron P Roznowski</i>	206
Dealing With the Whole Head: Diversity and Function of Capsid Ejection Proteins in Tailed Phages <i>Lindsay W Black and Julie A Thomas</i>	219
Jumbo Phages <i>Isaac T Younker and Carol Duffy</i>	229
CRISPR-Cas Systems and Anti-CRISPR Proteins: Adaptive Defense and Counter-Defense in Prokaryotes and Their Viruses <i>Asma Hatoum-Aslan and Olivia G Howell</i>	242
Bacteriophage: Therapeutics and Diagnostics Development <i>Teng-Chieh Yang</i>	252
Bacteriophage Vaccines <i>Pan Tao and Venigalla B Rao</i>	259
Bacteriophage Diversity <i>Julianne H Grose and Sherwood R Casjens</i>	265
Genetic Mosaicism in the Tailed Double-Stranded DNA Phages <i>Welkin H Pope</i>	276
Bacteriophages of the Human Microbiome <i>Pilar Manrique, Michael Dills, and Mark J Young</i>	283
Bacteriophage: Red Recombination System and the Development of Recombineering Technologies <i>Lynn C Thomason and Kenan C Murphy</i>	291
Nanotechnology Application of Bacteriophage DNA Packaging Nanomotors <i>Tao Weitao, Lixia Zhou, Zhefeng Li, Long Zhang, and Peixuan Guo</i>	302
General Ecology of Bacteriophages <i>Stephen T Abedon</i>	314
Marine Bacteriophages <i>Vera Bischoff, Falk Zucker, and Cristina Moraru</i>	322

Ecology of Phages in Extreme Environments <i>Tatiana A Demina and Nina S Atanasova</i>	342
---	-----

Archaeal Viruses

Diversity of Hyperthermophilic Archaeal Viruses <i>David Prangishvili, Mart Krupovic, and Diana P Baquero</i>	359
Euryarchaeal Viruses <i>Tatiana A Demina and Hanna M Oksanen</i>	368
Vesicle-Like Archaeal Viruses <i>Elina Roine and Nina S Atanasova</i>	380
Virus–Host Interactions in Archaea <i>Diana P Baquero, David Prangishvili, and Mart Krupovic</i>	387
Antiviral Defense Mechanisms in Archaea <i>Qunxin She</i>	400
Discovery of Archaeal Viruses in Hot Spring Environments Using Viral Metagenomics <i>Jennifer Wirth, Jacob H Munson-McGee, and Mark J Young</i>	407
Metagenomes of Archaeal Viruses in Hypersaline Environments <i>Fernando Santos, María D Ramos-Barbero, and Josefa Antón</i>	414
Extreme Environments as a Model System to Study How Virus–Host Interactions Evolve Along the Symbiosis Continuum <i>Samantha J DeWerff and Rachel J Whitaker</i>	419

Fungal Viruses

An Introduction to Fungal Viruses <i>Nobuhiro Suzuki</i>	431
Cross-Kingdom Virus Infection <i>Liyang Sun, Hideki Kondo, and Ida Bagus Andika</i>	443
Diversity of Mycoviruses in Aspergilli <i>Ioly Kotta-Loizou</i>	450
Evolution of Mycoviruses <i>Mahtab Peyambari, Vaskar Thapa, and Marilyn J Roossinck</i>	457
Mixed Infections of Mycoviruses in Phytopathogenic Fungus <i>Sclerotinia sclerotiorum</i> <i>Jiatao Xie and Daohong Jiang</i>	461
Mycovirus-Mediated Biological Control <i>Daniel Rigling, Cécile Robin, and Simone Prospero</i>	468
Mycoviruses With Filamentous Particles <i>Michael N Pearson</i>	478
Prions of Yeast and Fungi <i>Reed B Wickner and Herman K Edskes</i>	487
Single-Stranded DNA Mycoviruses <i>Daohong Jiang</i>	493
Structure of Double-Stranded RNA Mycoviruses <i>José R Castón, Nobuhiro Suzuki, and Said A Ghabrial[†]</i>	504

[†]Deceased.

<i>Ustilago maydis</i> Viruses and Their Killer Toxins <i>Alexis Williams and Thomas J Smith</i>	513
Vegetative Incompatibility in Filamentous Fungi <i>Songsong Wu, Daohong Jiang, and Jiatao Xie</i>	520
Viral Diseases of <i>Agaricus bisporus</i> , the Button Mushroom <i>Kerry S Burton and Greg Deakin</i>	528
Viral Killer Toxins <i>Manfred J Schmitt and Björn Becker</i>	534
Alternaviruses (Unassigned) <i>Hiromitsu Moriyama, Nanako Aoki, Kuko Fuke, Kana Takeshita Urayama, Naoki Takeshita, and Chien-Fu Wu</i>	544
Barnaviruses (Barnaviridae) <i>Peter A Revill</i>	549
Botybirnaviruses (Botybirnavirus) <i>Mingde Wu, Guoqing Li, Daohong Jiang, and Jiatao Xie</i>	552
Chrysoviruses (Chrysoviridae) - General Features and Chrysovirus-Related Viruses <i>Ioly Kotta-Loizou, Robert HA Coutts, José R Castón, Hiromitsu Moriyama, and Said A Ghabrial[†]</i>	557
Fungal Partitiviruses (Partitiviridae) <i>Eeva J Vainio</i>	568
Fusariviruses (Unassigned) <i>Sotaro Chiba</i>	577
Giardiavirus (Totiviridae) <i>Juliana Gabriela Silva de Lima, João Paulo Matos Santos Lima, and Daniel Carlos Ferreira Lanza</i>	582
Hypoviruses (Hypoviridae) <i>Dong-Xiu Zhang and Donald L Nuss</i>	589
Megabirnaviruses (Megabirnaviridae) <i>Yukiyo Sato and Nobuhiro Suzuki</i>	594
Mitoviruses (Mitoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	601
Mycoreoviruses (Reoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	607
Mymonaviruses (Mymonaviridae) <i>Daohong Jiang</i>	615
Narnaviruses (Narnaviridae) <i>Rosa Esteban and Tsutomu Fujimura</i>	621
Phlegiviruses (Unassigned) <i>Karel Petrzik</i>	627
Plant and Protozoal Partitiviruses (Partitiviridae) <i>Hanna Rose and Edgar Maiss</i>	632
Quadriviruses (Quadriviridae) <i>Hideki Kondo, José R Castón, and Nobuhiro Suzuki</i>	642
Totiviruses (Totiviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	648

[†]Deceased.

Yado-kari Virus 1 and Yado-nushi Virus 1 (Unassigned) <i>Subha Das and Nobuhiro Suzuki</i>	658
Yeast L-A Virus (<i>Totiviridae</i>) <i>Reed B Wickner, Tsutomu Fujimura, and Rosa Esteban</i>	664
Algal Viruses	
Algal Marnaviruses (<i>Marnaviridae</i>) <i>Marli Vlok, Curtis A Suttle, and Andrew S Lang</i>	671
Algal Mimiviruses (<i>Mimiviridae</i>) <i>Ruth-Anne Sandaa, Håkon Dahle, Corina PD Brussaard, Hiroyuki Ogata, and Romain Blanc-Mathieu</i>	677
Miscellaneous Algal Viruses (<i>Alvernnaviridae, Bacilladnaviridae, Dinodnavirus, Reoviridae</i>) <i>Keizo Nagasaki, Yuji Tomaru, and Corina PD Brussaard</i>	684
Phycodnaviruses (<i>Phycodnaviridae</i>) <i>James L Van Etten, David D Dunigan, Keizo Nagasaki, Declan C Schroeder, Nigel Grimsley, Corina PD Brussaard, and Jozef I Nissimov</i>	687
Invertebrate Viruses	
An Introduction to Viruses of Invertebrates <i>Peter Krell</i>	699
Ascoviruses (<i>Ascoviridae</i>) <i>Sassan Asgari, Dennis K Bideshi, Yves Bigot, and Brian A Federici</i>	724
Baculovirus–Host Interactions: Repurposing Host-Acquired Genes (<i>Baculoviridae</i>) <i>A Lorena Passarelli</i>	732
Baculoviruses: General Features (<i>Baculoviridae</i>) <i>Vera ID Ros</i>	739
Baculoviruses: Molecular Biology and Replication (<i>Baculoviridae</i>) <i>Monique M van Oers</i>	747
Bidensovirus (<i>Bidnaviridae</i>) <i>Qin Yao, Zhaoyang Hu, and Keping Chen</i>	759
Bunyaviruses of Arthropods (<i>Myxoviridae, Nairoviridae, Peribunyaviridae, Phasmaviridae, Phuniviridae, Wupedeviridae</i>) <i>Sandra Junglen</i>	764
Dicistroviruses (<i>Dicistroviridae</i>) <i>Yanping Chen and Steven M Valles</i>	768
Entomobirnaviruses (<i>Birnaviridae</i>) <i>Marco Marklewitz</i>	776
Hytroviruses (<i>Hytroviridae</i>) <i>Henry M Kariithi and Irene K Meki</i>	780
Iflaviruses (<i>Iflaviridae</i>) <i>Bryony C Bonning and Sijun Liu</i>	792
Iridoviruses of Invertebrates (<i>Iridoviridae</i>) <i>İkbal Agah İnce</i>	797
Mesoniviruses (<i>Mesoniviridae</i>) <i>Jody Hobson-Peters and Daniel Watterson</i>	804

Nimaviruses (<i>Nimaviridae</i>) <i>Peter Krell and Emine Ozsahin</i>	808
Nodaviruses of Invertebrates and Fish (<i>Nodaviridae</i>) <i>Kyle L Johnson and Jacen S Moore</i>	819
Nudiviruses (<i>Nudiviridae</i>) <i>Yu-Chan Chao, Chih-Hsuan Tsai, and Sung-Chan Wei</i>	827
Parvoviruses of Invertebrates (<i>Parvoviridae</i>) <i>Judit J Péntzes, Hanh T Pham, Qian Yu, Max Bergoin, and Peter Tijssen</i>	835
Polydnaviruses (<i>Polydnaviridae</i>) <i>Anne-Nathalie Volkoff and Elisabeth Huguet</i>	849
Poxviruses of Insects (<i>Poxviridae</i>) <i>Basil Arif, Lillian Pavlik, Remziye Nalçacıoğlu, Hacer Muratoğlu, Cihan İnan, Mehtap Yakupoğlu, Emine Özşahin, İsmail Demir, Kazım Sezen, and Zihni Demirbağ</i>	858
Reoviruses of Invertebrates (<i>Reoviridae</i>) <i>Peter Krell</i>	867
Rhabdoviruses of Insects (<i>Rhabdoviridae</i>) <i>Andrea González-González, Nicole T de Stefano, David A Rosenbaum, and Marta L Wayne</i>	883
Sarothroviruses (<i>Sarothroviridae</i>) <i>Azeez Sait Sahul Hameed</i>	888
Solinviviruses (<i>Solinviviridae</i>) <i>Steven M Valles and Andrew E Firth</i>	892
Tetraviruses (<i>Alphatetraviridae, Carmotetraviridae, Permutotetraviridae</i>) <i>Rosemary A Dorrington, Tatiana Domitrovic, and Meesbah Jiwaji</i>	897

VOLUME 5

Diagnosis, Treatment and Prevention of Virus Infections

Diagnosis

Introduction to Virus Diagnosis and Treatment <i>Maija Lappalainen and Hubert GM Niesters</i>	3
Electron Microscopy for Viral Diagnosis <i>Roland A Fleck</i>	5
Serological Approaches for Viral Diagnosis <i>Klaus Hedman and Visa Nurmi</i>	15
A Brief History of the Development of Diagnostic Molecular-Based Assays <i>Hubert GM Niesters</i>	22
Sequencing Strategies <i>Sibnarayan Datta</i>	27
Validating Real-Time Polymerase Chain Reaction (PCR) Assays <i>Melvyn Smith</i>	35
Rapid Point-of-Care Assays <i>Jan G Lisby and Uffe V Schenider</i>	45

Standardization of Diagnostic Assays <i>Sally A Baylis, C Micha Nübling, and Wayne Dimech</i>	52
Quality Assurance in the Clinical Virology Laboratory <i>Paul Wallace and Elaine McCulloch</i>	64
Biosafety and Biosecurity in Diagnostic Laboratories <i>Hannimari Kallio-Kokko and Susanna Sissonen</i>	82
Screening for Viral Infections <i>Walter Ian Lipkin, Nischay Mishra, and Thomas Briebe</i>	91
Clinical Diagnostic Virology <i>Marcus Panning</i>	98
Virus Diagnosis in Immunosuppressed Individuals <i>Elisabeth Puchhammer-Stöckl and Fausto Baldanti</i>	105
Diagnosis; Future Prospects on Direct Diagnosis <i>Marianna Calabretto, Daniele Di Carlo, Fabrizio Maggi, and Guido Antonelli</i>	112

Treatment

Antiviral Classification <i>Guangdi Li, Xixi Jing, Pan Zhang, and Erik De Clercq</i>	121
Antiretroviral Therapy – Nucleoside/Nucleotide and Non-Nucleoside Reverse Transcriptase Inhibitors <i>Timothy D Appleby and Killian J Quinn</i>	131
Protease Inhibitors <i>Vanesa Anton-Vazquez and Frank A Post</i>	139
HIV Integrase Inhibitors and Entry Inhibitors <i>Daniel Bradshaw and Ranjababu Kulasegaram</i>	145
Management of Respiratory Syncytial Virus Infections (<i>Pneumoviridae</i>) <i>Rachael S Barr and Simon B Drysdale</i>	155
Management of Influenza Virus Infections (<i>Orthomyxoviridae</i>) <i>Bruno Lina</i>	160
Management of Herpes Simplex Virus Infections (<i>Herpesviridae</i>) <i>Nicole Samies and Richard Whitley</i>	175
Management of Varicella-Zoster Virus Infections (<i>Herpesviridae</i>) <i>Andreas Sauerbrei</i>	181
Treatment and Prevention of Herpesvirus Infections in the Immunocompromised Host <i>Sara H Burkhard and Nicolas J Mueller</i>	190
Management of Adenovirus Infections (<i>Adenoviridae</i>) <i>Albert Heim</i>	197
Management of Hepatitis A and E Virus Infection <i>Sébastien Lhomme, Florence Abravanel, Jean-Marie Peron, Nassim Kamar, and Jacques Izopet</i>	206
Management of Patients With Chronic Hepatitis B (<i>Hepadnaviridae</i>) and Chronic Hepatitis D Infection (<i>Deltavirus</i>) <i>Milan J Sonneveld and Suzanne van Meer</i>	217
Studying Population Genetic Processes in Viruses: From Drug-Resistance Evolution to Patient Infection Dynamics <i>Jeffrey D Jensen</i>	227

Virus-Based Cancer Therapeutics <i>Roberto Cattaneo and Christine E Engeland</i>	233
Prevention	
Surveillance of Infectious Diseases <i>Norman Noah</i>	247
Preparing for Emerging Zoonotic Viruses <i>Reina S Sikkema and Marion PG Koopmans</i>	256
Use of Immunoglobulins in the Prevention of Viral Infections <i>Leyla Asadi and Giovanni Ferrara</i>	267
Vaccine Production, Safety, and Efficacy <i>Thomas J Brouwers and Bernard AM Van der Zeijst</i>	281
Vaccines Against Viral Gastroenteritis <i>Scott Grytdal, Tyler P Chavers, Claire P Mattison, Jacqueline E Tate, and Aron J Hall</i>	289
Human Papillomavirus (HPV) Vaccines and Their Impact <i>Jade Pattyn, Pierre Van Damme, and Alex Vorsters</i>	295
Influenza Vaccination <i>Topi Turunen</i>	300
Polio Eradication <i>M Steven Oberste, Cara C Burns, and Jennifer L Konopka-Anstadt</i>	310
Subject Index	315

BACTERIAL VIRUSES

History of Virology: Bacteriophages

William C Summers, Yale University, New Haven, CT, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of H.-W. Ackermann, History of Virology: Bacteriophages, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00590-2.

Discovery of Phages

Although bacteriophages are the most abundant organisms on our planet, and they often exert drastic effects (lysis) on their bacterial hosts, it was not until the second decade of the 20th century that they were recognized and their basic biological nature was studied. While the literature contains several dozen reports that may be, only in retrospect, of phenomena associated with bacteriophages, the first clear recognition and description of what was at the time called “the bacteriophage phenomenon” was by a rather mysterious French-Canadian autodidact, Félix Hubert d’Herelle. D’Herelle, working at the Pasteur Institute during the first world war, was investigating the problem of virulence of bacterial strains. The basis for differing virulence of different isolates of the same bacterial species was of both theoretical and practical interest. Basing his reasoning on earlier (later shown to be incorrect) ideas that microbial virulence was modulated by “associated” organisms, d’Herelle fractionated a bacterial culture from a particularly virulent outbreak of dysentery by filtration to obtain a bacteria-free filtrate that he added back to a pure culture of the original organism. To his surprise, rather than increase the virulence of this culture, this filtrate caused lysis of the bacterial culture and, on solid media, the bacterial growth was inhibited in small spots. Material recovered from the spots was sterile with respect to bacteria, but could cause sequential lysis of fresh cultures. This “lytic principle” could be serially propagated indefinitely. D’Herelle reasoned that this “lytic principle” was a filter-passing, particulate agent that could multiply at the expense of the bacterial cell. He called it “bacteriophage” and classed as an “ultravirus”, meaning that it was smaller than the usual “virus”, a term used at the time to indicate any agent of transmissible disease. In his words, bacteriophage was an invisible obligate parasite.

D’Herelle was more expansive than just describing a new type of microbe, however. Since the bacteriophage was able to kill bacteria, and it appeared to him that the quantity of bacteriophage in patient samples increased as the patient was recovering, he reasoned that the recovery itself was caused by the takeover by the bacteriophage. Again, in his words, it was the agent of recovery and immunity. In the pre-antibiotic era this concept promised here-to-fore unavailable opportunities to treat infectious diseases.

Almost immediately after d’Herelle’s initial report in 1917 and his assertion that phages were involved in recovery from infectious diseases as well as the agent of what he called natural immunity, his discovery was explored by others and controversy ensued. The initial controversy concerned the biological nature of phage, that is, was it an ultravirus or a cellular product? This controversy became acute when d’Herelle challenged the theory of immunity recently advocated by the eminent director of the Pasteur Institute in Brussels, and recent Nobel prize winner, Jules Bordet. In his 1923 monograph on phage and disease, d’Herelle referred to Bordet’s work on bacteriolytic antibodies as “an error”. Bordet and his young protégé, Andre Gratia, took up the challenge, initially not defending the antibody work, but by attacking d’Herelle’s interpretation of the nature of phage. This approach was of little effect, so when they discovered Twort’s rather obscure paper on ultra-microscopic viruses from 1915, Bordet and Gratia launched an attack on d’Herelle’s priority of discovery. This dispute was intense for almost a decade, and probably was a deterrent and distraction to phage research for several decades. The priority dispute survives to the present.

D’Herelle’s response to the priority challenge was to try to demonstrate that what he had discovered was not the same phenomenon as observed by Twort. This work did help to clarify the nature of phage but did not convince most observers for the simple reasons that Twort’s paper was vague and he published nothing later in the way of clarifying experiments.

Two key problems were at the center of d’Herelle’s work on the nature of phage in the 1920s: is phage a unique organism that shows various manifestations, or are there many species of phage? D’Herelle argued for the “unicity” of phage, and claimed that under different conditions, for example, when supplied with different host bacteria, it adapted to such changes. This notion was not so strange as it seems today. In the 19th century, bacteriologists argued the same way about bacterial species, one species with variable forms (polymorphism) or many species, each with a stable form (monomorphism). Also, in the 1920s, a theory called “cyclogeny” was popular that suggested microbial cultures could change form and behavior depending on cultural conditions. Examples include bacterial shape changes observed in exponentially growing cultures versus stationary phase cultures and changes observed upon bacterial sporulation and germination.

The second key problem involved the physical nature of phage. D’Herelle argued that the fact that phage produced plaques, and that the number of plaques produced upon dilution followed a Poisson distribution, proved that phage were particulate in nature. He even enlisted Einstein, with whom he discussed the matter, on his side in the argument: “I was very glad to see how this deservedly-famous mathematician evaluated my experimental demonstration, for I do not believe that there are a great many biological experiments whose nature satisfies a mathematician”. Still, most bacteriologists of the day thought of phage as some

sort of cellular product with lytic properties, such as the recently discovered enzyme, lysozyme. The “corpuscular nature of phage” was not definitively settled until they were visualized by electron microscopy in 1939.

A third controversy that smoldered for many years surrounded the status of phages (and filterable viruses in general) was their status as “living” beings. Microbiologists in particular struggled with various definitions of “life”, probably a legacy of earlier vitalism. While viruses seemed to “reproduce”, which was one key property of living things, it was not clear that they could “assimilate” material from their environment in order to grow. Their dependence on a full-fledged cellular host was confusing in terms of the existing criteria for “life”. At this period, autocatalytic enzymes such as pepsinogen and trypsinogen, which could convert themselves from inactive to active pepsin and trypsin, were discovered. Could phage “growth” be such an autocatalytic phenomenon? Nobelist John Northrop and his colleagues in the Rockefeller Virus Unit at Princeton thought so.

Biologists were intrigued by things “at the borderline of life”. D’Herelle even developed an extensive theory of evolution based on phage as the chemical basis for life, asserting that phage were a colloidal form of matter that coalesced into protoplasm and cells. He undertook a search for phages in sulfur hot springs and other places where the “origin of life” might have occurred. He termed the study of phage “protobiology” and assigned the Linnaean binomial *Bacteriophagus protobius* early in his work on “protobes”. Indeed, during his tenure at Yale in the late 1920s and early 1930s, he referred to himself as Professor of Protobiology. The question of the organismal nature of viruses led scientists such as Jacques Bronfenbrenner to examine the respiration of phages using techniques such as the newly devised Warburg apparatus. To add to the mystery of viruses, in 1935 Wendell Stanley managed to obtain viruses in crystalline form, the *sine qua non* of chemical purity. Stanley’s chemical analysis of the crystals of tobacco mosaic virus found them to be made of protein (he missed the 7% phosphorus content, an indication of the presence of nucleic acid, later found by Bawden and Pirie in 1936). The question of whether viruses are “living or dead” has by now been relegated to late night undergraduate debating, and in 1957, Andre Lwoff simply stated “viruses are viruses”.

Phage Therapy

One of d’Herelle’s earliest observations of phage involved dysentery patients and he noted the quantity and virulence of the phages in stool samples seemed to increase as the patients were recovering from the disease. This, as noted above, suggested that the phage was contributing to the recovery. It was no big leap to suggest that phages, isolated in the laboratory, might be used as treatment of such diseases. In the era before antibiotics, such therapeutic potential generated great interest, and the interest in phage in the decades of the 1920s and 1930s focused mainly on various attempts to exploit this therapeutic promise.

The first trials of phage therapy were carried out in the veterinary setting. D’Herelle carried out a trial of phages in an epidemic of avian typhosis, a bacterial disease of chickens, that was becoming endemic in rural France in 1918. He isolated a phage that was virulent against the pathogen as isolated from the diseased birds, and in both field trials and laboratory tests, found that phage had both therapeutic and prophylactic usefulness in actual disease outbreaks. He continued his search for opportunities to field-test his ideas and became a veritable globe trotter in this endeavor, spreading phages in North Africa, Argentina, French Indochina, India, and Egypt, as well as the hospitals of Paris. In the pre-antibiotic era, his work attracted much attention and many attempts to employ phage therapy. After over a decade of clinical experience, the American Medical Association’s Council on Pharmacy and Chemistry reviewed the literature on phage therapy and the 1934 report written by Monroe Eaton and Stanhope Bayne-Jones concluded that the efficacy of phage therapy was still not established because of the conflicting reports, ambiguous results, and lack of understanding of the biological nature of phage. This report coincided with d’Herelle’s departure from the US to set up a phage institute in Tbilisi in the USSR, an ill-fated venture that ended with his return to France in 1937 after the execution of his host and fellow phage worker, Georgi Eliava, who was caught up in the Stalinist purges of that period. This tragedy, shortly followed by the events leading up to World War 2, curtailed the influence of the most ardent and effective advocate of phage therapy. The discovery and effectiveness of antibiotics (and chemotherapy such as the sulfa drugs) in the late 1930s and on into the next decade led to the eclipse of interest in phage therapy, especially in the West.

In the Eastern Bloc of nations, partly because of the difficulty in developing the capacity to produce antibiotics and partly because of the lingering interest in phage therapy, research on phage therapy continued in several places, especially in Tbilisi and Poland. As understanding of phage biology increased, these centers profited from new knowledge as applied to phage therapy and have emerged in recent years as leaders in the renewed interest in phage therapy in this era of multi-drug antibiotic resistance.

By the mid-1980s, especially with the widespread recognition of MRSA and the ubiquity of horizontal gene transmission of drug-resistance, the search for solutions to this problem became acute. New antibiotic discovery was recognized as both economically non-viable as well as likely to be futile. Phage therapy was again seriously considered. In the West phage therapy seemed to have been discarded for two reasons, one theoretical and one political. Since phages were known to be antigenic, the development of immune resistance was considered a fatal drawback as was the well-known ability of bacteria to mutate to phage resistance. Further, cold-war ideology militated against any recognition of scientific work from the Soviets. A key study leading to the resurgence of interest in phage therapy was published in 1996 by Carl Merrill and his colleagues, who showed that the problem of immune clearance of therapeutic phages was not insurmountable. They showed that by repeated cycles of selection *in vivo*, phages strains (of Lambda bacteriophage) could be derived that were invisible to the host (mouse) immune recognition system. Detailed understanding of the nature of phage receptors and phage resistance mechanisms suggested that these were not such daunting impediments. By 2000, phage therapy had become respectable again.

Phage Typing

Soon after the discovery of bacteriophages by d'Herelle in 1917, it became clear that there was some sort of specificity of phages for particular host strains of bacteria. While the issue of "unicity" of phage was unresolved, it was still observed that phage isolates could be obtained (whether by adaptation or by isolation of pre-existing strains) that could distinguish between different bacterial hosts by their "host range". D'Herelle and his colleagues noted some correlation between phage strains and the immunological properties of the host bacteria. Other phage researchers, especially James Craigie in Canada, inverted this notion to devise phage strains that could be used to classify bacterial based on their susceptibility to known tester phage strains. Classification of the *Salmonellae* was a particularly fraught problem, and in 1938, Yen and Craigie published a landmark classification of *S. typhi* (*B. typhosus*) by a set of "typing phages" that had host ranges that correlated precisely with the Vi surface antigen types. Phage typing became a widespread technique in diagnostic bacteriology and epidemiological investigations, yet it also remained a distinct branch of phage research. Phages were isolated and characterized for their utility in discriminating one bacterial variant from another, and standard sets of phages were developed for microbiological reference laboratories. Phage typing continues as a useful and reliable tool in diagnostic bacteriology even today. Phage typing is relatively rapid, easy, and low tech, and the diversity of phages has made it possible to find phages with the needed fine discriminatory properties. Interestingly, phage typing was, and continues to be, successful even though its practitioners and developers have been rather separate from the basic phage researchers who were attacking the fundamental biology of bacteriophages.

Phage Genetics

The very first defining property of phages was their serial transmissibility, a property that implied some sort of heredity mechanism. Indeed, as H.J. Muller famously noted in 1922, "if these d'Herelle bodies were really genes, fundamentally like our chromosome genes, they would give us an utterly new angle from which to attack the gene problem". The hereditary properties of phage were understood only slowly, partly because the concepts of adaptation, mutation, and speciation were just being extended to the microbial world. By the mid-1930s it was becoming clear that certain observable characters of viruses, such as host range and lesion morphology, behaved as stable, heritable properties. This was as true for phages as for animal and plant viruses.

One oddity that phage presented was the ability of some phages to associate with their host bacteria in mysterious ways that at times appeared to be a kind of latency (later termed the lysogenic state) and at times appeared to be a kind of genetic modifier that changed one of more bacterial property, for example, toxin production or surface antigens (later termed phage conversion or phage transduction). The debates over the biologic nature of phage (microbe or self-activating enzyme) hampered the clarification of the genetic nature of phage biology.

In the 1930s, Eugene and Elizabeth Wollman at the Pasteur Institute advanced the understanding of the lysogenic state and the relationship of phage infection to the introduction of new genes into the host cells. In the immediate post-war period with a renewed interest in the fundamental biology of phage (as opposed to their potential as anti-infectious agents for therapy) several groups saw the utility of phages as simple genetic organisms analogous to other microbes such as bacteria and fungi. Different "strains" of phage had been isolated from the earliest research that differed in their size, host range, and plaque appearance on a uniform "lawn" of a given host. Further, it was noted that rare variants arose which had the characteristics of mutant forms of the same strain. In 1929 Vladimir Sertic was the first to report that it was possible to obtain phages that apparently adapted to be able to grow on strains that were normally resistant to such phages. Salvador Luria made a systematic study of host range mutants in 1945 and Alfred Hershey described plaque type mutations in the same year. Phage biologists quickly adopted the obvious analogies to higher organisms and spoke of bacteriophage "strains", "races", "species", "variants", and "mutants".

As phage biologists sought to understand phage reproduction by testing for intracellular interference and competition between phages for the cellular processes involved in reproduction, they also noted that phages of differing properties but of the same strain could recombine or "cross" in the same way that higher organisms behaved. A kind of mating between strains with different genetic "markers" was developed to construct recombination frequency maps and to devise models to explain various oddities of phage biology (such as multiplicity reactivation, lysogeny, and lateral gene transfer or transduction).

The technical simplicity of phage genetics together with the statistical power of the large numbers of progeny phage and effective selection methods led to very high resolution genetic mapping experiments as well as detailed characterization of certain kinds of mutations. The ability to detect very rare recombinants meant that very close markers could be studied. In the famous T4rII gene, mutations that were in adjacent base pairs could be observed as separate markers in the mapping experiments of Seymour Benzer, thus clarifying and reconceiving the very concept of the gene. The use of mutagens with known mechanisms (insertion mutations produced by acridines) allowed Crick and Brenner to cross phage mutations in such a way as to establish, by genetic means, the triplet nature of the genetic code. With phages, the genes under study (in the phage) were separable from the environment in which they were expressed (the bacterial host), it became possible to study the rather mysterious phenomenon of suppressor mutations. Suppressors are, in general, genes that change (usually revert) other mutations but do not map in the gene of the function that was originally mutated. They are sometimes called extragenic suppressors. Bacteria that carried certain kinds of suppressor genes were used to isolate phage variants that required suppression for growth. These "sus" (suppressor sensitive, later called amber, ochre, and umber mutants for trivial reasons) mutants of phage turned out to be important in understanding the role of transfer RNA and the stop codons of the genetic code in protein synthesis.

The early observation that phage infection sometimes modified the properties of the few bacterial survivors of infection suggested that phages might be mutagens of some sort. This turned out to be true in the broad sense that they caused genetic changes, but by acting as vectors to bring in new genes rather than by direct changes to existing genes. This mode of what is now called “lateral gene transfer” turns out to be of major significance in evolution and gene flow in unicellular organisms in general. Many strains of pathogenic bacteria owe their virulence to toxin genes that are associated with temperate phages (phages that can become latent without lysing their hosts) that have carried these genes into an otherwise benign organism. Common examples include the toxin genes in cholera carried on the temperate phage CTX ϕ , and the toxin genes in the diphtheria bacteria. In addition to toxins, temperate phages can control cell adhesion, invasion, and colonization properties of the bacterial host.

For molecular biologists, the detailed genetic study of a single phage, Lambda, a temperate phage of *E. coli*, provided a trove of information on both the mechanism of lateral gene transfers (such as the mechanism of genome integration and recombination) as well as the molecular processes of gene expression and regulation. These discoveries with bacteriophage led directly to the current understanding of retrovirus biology (HIV integration and viral oncogenesis, for example) as well as the entire field of transcriptional control of gene expression.

At the time phage workers were exploring host range mutants and the strain specificity of various phages, two groups observed a rather strange sort of host-range phenomenon. A phage grown on one particular host strain sometimes could grow efficiently on that host but fail to grow well on a closely related host (measured by the “efficiency of plating” e.o.p., a theoretical e.o.p. of 1 was taken to mean, without much evidence, that every physical phage particle was infective and produced one plaque on a lawn of sensitive bacteria). While the e.o.p. on the “restrictive host” was not zero, it was often in the range of 10^{-4} . The phages that escaped this restriction, however, were now somehow adapted to grow with an e.o.p. of near 1 again on the so-called restricting host. Sometimes it grew as well on the earlier host, too. It was as if the phage somehow remembered the last host in which it grew. This phenomenon seemed like some sort of unstable mutation or adaptation to the growth requirements of the new host. This phenomenon was observed in both T4 phage by Luria and Human and in phage Lambda by Weigle and Bertani in 1952 and was termed host-induced variation (later called modification) to distinguish it from stable hereditary mutation.

Host-induced modification of phages was a rather specialized and esoteric aspect of phage biology that did not attract much attention at first. The modification involved seemed to reside in the genome of the phage, but its chemical nature and physiological significance was obscure. By the mid 1960s, however, it became clear that some chemical modifications of DNA were involved, methylation of certain bases, hydroxymethylation and glycosylation in other cases. It appeared, then, that some, but not all, bacterial hosts could modify phage DNA in ways that could be recognized by other host bacteria. How this mechanism worked, however, was unclear. In the mid 1960s, however, experiments on the fate of phage DNA upon entering the bacteria showed that the phage DNA that was restricted entered the bacterial cell but was degraded rapidly after entry.

Two lines of research soon clarified this phenomenon. First, in 1969 Herbert Boyer and Daisy Roulland-Dussoix dissected the genetics of bacteria that could modify and restrict phages, and identified a three gene system (ramABC) that controlled this process: one gene controlled modification, one gene controlled restriction, and a third gene controlled both. The interpretation was that there was one gene product each for marking and degrading the DNA, and a third gene product which directed the other two gene production to specific sequences in the phage DNA, some sort of sequence recognition system. Second, starting in 1968, numerous investigators demonstrated nucleases in various systems that showed specificity for DNA which was destined for restriction *in vivo*. These various enzymes, surprisingly ubiquitous, came to be called “restriction endonucleases” (later shortened to “restriction enzymes”). In 1970, Thomas Kelly, Jr., and Hamilton Smith found that one of these enzymes, a nuclease from *H. influenza*, cleaved DNA at a specific palindromic sequence of 6 base pairs. These nucleases, with their specificity and diversity, soon were recognized as powerful new tools for gene manipulation and formed the basis for many rapid developments in biotechnology as well as basic scientific research. From an esoteric problem in phage biology of interest to a tiny handful of researchers in the 1960s, this work led to a multibillion dollar global enterprise. For anyone who thinks that support for science can and should be predicated on prospects of future payoff, the story of phage restriction and modification is certainly a cautionary tale.

The small size of phage genomes coupled with their homogeneity and ease of preparation made them the organisms of choice for the application of nucleotide sequence studies starting in the 1960s when the hypothesis of a genetic code was being explored. Phages provided a ready source of “genes” of limited complexity compared to cellular DNAs. Except for the indirect approach in which the frequency of pairs of neighboring nucleotides were compared, the first attempts at sequence comparisons were based on the formation of stable hybrid DNA duplexes between two DNAs to be tested. Ingenious variations of this method were devised including observing DNA hybrids in the electron microscope (“heteroduplex mapping”) and later hybridizations with one strand immobilized on a solid matrix (filter hybridization and “Southern” blotting). With the discovery of the sequence-specific restriction endonucleases in the 1970s, the mapping of cleavage sites on phage and small virus genomes provided a significant advance in sequence knowledge and comparisons that became a way to characterize and manipulate phage genes and genomes. When methods were developed for complete DNA sequence analysis, phage genomes were the obvious first choices for application of these powerful new techniques, and the DNA of phage ϕ X174 became the first DNA genome to be completely sequenced in 1977 by Fred Sanger and his colleagues at Cambridge University.

Phage Biochemistry

The chemical study of bacteriophage seems much simpler than the genetic and biological work on phage, probably because phage are relatively simple entities, and the techniques needed were perfected rather later than other methods. As a consequence of their small size yet delicate composition, it was hard to isolate and prepare phage preparations that were suitable for the chemists until the availability of the ultracentrifuge in the post World War 2 period.

Early study of phages relied on rather crude preparations that were probably mostly inactive because of the harsh methods of organic chemistry were unsuitable for most complex biological materials. Modern nucleoprotein chemistry was essentially unknown before World War 2. That phages contained protein and nucleic acid was clearly shown in the 1930s by Martin Schlesinger, first in Germany and later, when he was a refugee in England. His phage preps contained amino acid polymers by chemical analysis, and nucleic acids by certain diagnostic reactions used by the histochemists (the Feulgen reaction). They were semiquantitative at best. The physical chemist, William J. Elford, working with the virologist Christopher Andrewes in England used calibrated dialysis membranes as filters to get rough estimates of the sizes of different phages.

The widespread introduction in 1949 of the model L preparative ultracentrifuge by Spinco (Specialty Instrument Company, later acquired by Beckman) the spinoff company of E.G. Pickles of the University of Virginia greatly advanced the chemical and biochemical study of phages as well as other biological materials. With the ability to sediment virus particles and separate them from other cellular material ("debris"), biochemists could analyze the whole and subviral components of specific phages. Soon it was found that there was great diversity among different phages. The DNA in phages was found to be a single polymer chain, not occurring in segments such as the chromosomes of higher cells. The DNA in some small phages was found to be a single strand of DNA in the form of a covalently closed circle (ϕ X174), and some phages even contained RNA rather than DNA (f2, MS2, R17, Q β). Cases of unusual nucleic acid components, for example hydroxymethyl cytosine and deoxyuridine were found in some phages. The finding of methylation and glycosylation of DNA in phages was crucial to understanding the important discovery of phage restriction and modification.

As the biochemists elucidated new metabolic events that occurred after infection, new phage-induced enzymes were discovered, and later shown to be encoded by phage genes (rather than phage-activated cell genes). These enzymes included new DNA polymerases, phage-specific RNA polymerases and transcription factors, polynucleotide kinase, DNA ligase, and several enzymes of nucleotide metabolism. Several of these activities were unknown as cellular functions and became useful tools in nucleic acid biochemistry.

Phage biology was central to the problem of protein biosynthesis from the early 1940s. Since phage reproduction yields a burst of new extracellular material in a rapid and synchronous fashion, it was an ideal model to study the production of new protein. Early experiments with radiolabeled substrates settled the question as to the origin of the material in the new phages: it came from precursors in the cell by *de novo* synthesis after infection, rather than from repurposing pre-existing proteins. In the case of DNA, some phages produced enzymes that degraded the host DNA and reutilized the precursor nucleotides for phage DNA synthesis. These pathways became the first such steps in the central dogma to be dissected by *in vitro* study of the replication machinery with small phage ϕ X174 DNA as the template. The discovery of phages containing RNA genomes provided an ideal source of RNA that could be used to test the new ideas about messenger RNA and ribosome function in the 1960s.

Phage and Biotechnology

The first inkling that phages might be exploited for technological applications (excluding phage for medical therapies) was the demonstration that a phage could be used to isolate and purify a cellular gene. In 1969 Jon Beckwith and his colleagues isolated pure lac operon DNA from *E. coli* using the technology afforded by Lambda phages that had incorporated that cellular DNA to become a "transducing phage". While this method was complicated and of apparent limited application, the program to isolate and purify specific genes for biotechnological applications had been launched. The subsequent discovery that restriction endonucleases could be used to prepare genes and gene fragments at will, along with the use of phages and plasmids as vectors to introduce genes into cells meant that this new technology had come of age. Phage biologists had developed a good understanding of the process of transduction (finding phages that had incorporated cell genes into their genomes) meant that phages such as Lambda were rapidly exploited as vectors for introducing a specific DNA sequence into a single host cell that could then be grown up into a mass culture from a single, genetically defined progenitor, a process that came to be called "molecular cloning" or simply cloning (derived from the original meaning from plant biology, "a shoot" or "offspring").

The understanding of gene regulation in phages, obtained from years of basic phage research, provided the new generation of biotechnologists with tools to control the expression of inserted genes at will. A particularly important technology depended on the discovery that one *E. coli* phage, T7, encoded an entirely new RNA polymerase that transcribed only certain genes expressed late in the phage infectious cycle. The transcription signals for this polymerase were linked to the genes to be cloned and thus their expression would only occur if the T7 RNA polymerase was present and active. By controlling this RNA polymerase in turn, the genetic engineer could turn the cloned gene on and off as needed. This system was particularly useful since many genes made products that were detrimental to the cell growth when fully expressed, making it hard to grow large cultures with high expression levels. The T7 system, based on rather esoteric phage research, turned out to make possible the overproduction of cloned gene products in vast excess for other biotechnological uses. In some cases, up to half of the intracellular protein could be forced from

the cloned gene, in effect, making the crude cell extract 50% pure to begin with. Much of the current progress in structural biology owes its success to these phage expression systems which produce abundant quantities of proteins normally found in miniscule amounts in the cell.

Another ingenious application of phage biology to biotechnological problems was the development of “phage display” methods, pioneered by George Smith in 1985. Since phages have multiple copies of identical coat proteins on their surface and since these coat proteins are encoded by the phage genome, it was possible to insert a mixture of foreign gene sequences into the coat protein genes of individual phages in such a way that the phages produced fusion proteins that could still function as a coat protein molecule yet at the same time “display” some of the foreign protein domains on the surface of the phage. A mixture of phages, each containing a different introduced foreign sequence attached to its coat protein gene, was called a “phage display library”. The power of this method was based on various ways of selecting for the properties of the foreign, “displayed” protein. Affinity methods with a chemical ligand or a protein target linked to a solid matrix could then be used to adsorb only those phages which “displayed” a specific functional cloned gene. Antibody affinity methods were sometimes used as well. The beauty of this method, is that an individual phage, selected by virtue of that particular phage carrying the specific selected gene, could be plaque-purified and grown into a pure stock of phage with only the desired cloned gene.

Another recent tool for genetic engineers and biotechnologists that has only a peripheral relation to the history of bacteriophage is the gene editing techniques based on the CRISPR Cas 9 system. CRISPR is an acronym for “clustered regularly interspaced short palindromic repeats”, a description of DNA sequence regularities noted in bacterial DNA sequences by several research groups between 1987 and 2000. The ubiquity of these regularities and their relation to small RNAs led researchers to seek some function for these sequences. Eventually, they were found to be related to a DNA cutting function that was connected to one of the many ways that prokaryotic cells seem to defend against lateral gene transfer, in particular, bacteriophages. The exploitation of this system for simple alteration of gene sequences represents a major advance in gene manipulation technology. It is historically incorrect, however, to present the history of CRISPR as part of the history of bacteriophages, since this mechanism was applied to phage biology only *post hoc*, and phage research had essentially no role in its discovery or elucidation, serving only as a subsequent biological justification for its presence in cells.

Phage Ecology and Evolution

Phages are ubiquitous in the biosphere, but only relatively recently has the diversity and ecological importance of bacteriophages attracted much attention. Roger Hendrix noted that, given that the average phage titer of the world’s sea waters is about 10^8 per ml., there are about 75 million blue whale units of phage on earth. That is, the combined global mass of phages equals that of the mass of about 75 million blue whales, clearly an impressive number. Placed end-to-end, these tiny phages would stretch about 100 million light years, about 100 times the diameter of our own galaxy. Why are phages so abundant? What is their function in diverse ecosystems? How many kinds of phages are there? Only recently have these questions been studied.

Early phage researchers were troubled by the differing results they obtained with phages isolated from nature. D’Herelle and his colleagues tabulated some of their isolates and noted that they differed (in no obvious or systematic way) in size, immunological properties, host range, grow rates, and so on. Later, when phages were examined under the electron microscope, diversity but no clear systematics emerged. For years, since its first report in 1971 under Peter Wildy, the International Committee on the Nomenclature of Viruses (ICNV) has struggled (with essentially no success) to systematize the classification of viruses including phages. Every year, the diversity of phages continues to surprise phage biologists, especially with more study of the archaea and their phages.

Microbes are now recognized as more than just passive parasites, rather they are essential symbionts or key components of complex ecosystems. Phages are now recognized as playing an important role in many biological communities. For example, the interplay of phages and various cyanobacteria in the oceans are crucial to the global carbon cycle. Without massive phage lysis of cyanobacteria every day, atmospheric carbon would be sequestered in these bacteria, sink to the ocean floor (eventually becoming oil) but in the meantime, the depletion of atmospheric CO_2 would reduce the greenhouse effect so much that the entire planet could freeze.

Another example of phage-host equilibrium is found in the Ganges River in India. For millennia Hindus have used this sacred river for interment of the remains of the dead. In a traditionally endemic cholera region, it is surprising that the Ganges water is not a source of major cholera infections. Over centuries, apparently, there had been an evolution of anti-cholera phages in the Ganges that effectively suppresses the cholera vibrio populations despite the episodic contaminations during lethal disease outbreaks. This “Ganges paradox” was noted long before phages were discovered, but this observation cannot be considered as an earlier “discovery” of phages.

Conclusion

Bacteriophages, discovered just over a century ago, are the most prevalent biological agents on our planet. They are also likely the most diverse. Their small size, simple genetic and physical structure, and rapid reproduction seem to account for their adaptability, their multiple ecological roles and impact, and their utility for diverse scientific purposes. Some phages were exploited from their

first discovery as anti-infectious disease agents in human and animal medicine. Others, because of their seemingly simplicity, were useful biological models that led to deep insights into basic biological problems such as heredity and cellular physiology. Knowledge of phage biology and biochemistry together with their ease of laboratory manipulation led to diverse applications in biotechnology far removed from their natural biological activities. The gradual recognition of the diversity and complexities of microbial ecology has recently brought phages once again into a central role in the global biological economy, extending even as far as having a part to play in the eventual attack on the problem of global climate change.

Further Reading

- Cragie, J., Yen, C.H., 1938. The demonstration of types of *B. typhosus* by means of preparations of type II Vi phage: I. Principles and technique. *Canadian Journal of Public Health* 29, 448–463.
- d'Herelle, F., 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. *Comptes rendus de l'Académie des Sciences* 165, 373–375.
- Ellis, E.L., Delbrück, M., 1939. The growth of bacteriophage. *The Journal of General Physiology* 22, 365–384.
- Holmes, F.L., 2006. In: Summers, W.C. (Ed.), *Reconceiving the Gene: Seymour Benzer's Adventures in Phage Genetics*. New Haven: Yale University Press.
- Lederberg, J., 1998. Plasmid (1952–1997). *Plasmid* 39, 1–9.
- Luria, S.E., 1945. Mutations of bacterial viruses affecting their host range. *Genetics* 30, 84–99.
- Muller, H.J., 1922. Variation due to change in the individual gene. *The American Naturalist* 56, 32–50.
- Smith, G.P., Petrenko, V.A., 1997. Phage display. *Chemical Reviews* 97, 391–410.
- Stent, G.S., 1963. *Molecular Biology of Bacterial Viruses*. San Francisco: W.H. Freeman.
- Summers, W.C., 2001. Bacteriophage therapy. *Annual Review of Microbiology* 55, 437–451.
- Summers, W.C., 1999. *Félix d'Herelle and the Origins of Molecular Biology*. New Haven: Yale University Press.
- Suttle, C.A., 2005. Viruses in the sea. *Nature* 437, 356–361.
- Twort, F.W., 1915. An investigation on the nature of ultra-microscopic viruses. *The Lancet* 186, 1241–1243.
- Wagner, P.L., Waldor, M.K., 2002. Bacteriophage control of bacterial virulence. *Infection and Immunity* 70, 3985–3993.

Icosahedral Phages – Single-Stranded DNA (ϕ X174)

Bentley A Fane, University of Arizona, Tucson, AZ, United States

Aaron P Roznowski, The University of Texas at Austin, Austin, TX, United States and University of Arizona, Tucson, AZ, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of B.A. Fane, M. Chen, J.E. Cherwa, A. Uchiyama, Icosahedral ssDNA Bacterial Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00456-8.

Glossary

Ångström (Å) 10^{-10} m. The unit used when describing distances within protein structures.

T = 1 icosahedron A geometric shape consisting of 20 faces and 12 vertices. A T = 1 virion is composed of 60 viral coat proteins.

π - π stacking Attractive, noncovalent interactions between aromatic rings.

Procapsid A viral assembly intermediate containing a full complement of scaffolding proteins but devoid of a genome.

Scaffolding protein A protein that directs the assembly of the virus, found in the procapsid assembly intermediate but not in the mature virion.

S value or Svedberg A unit of speed to describe the rate a particle moves through a gradient.

History

Sometime in the 1920s, Les Années Folles, Sertic and Bulgakov isolated bacteriophage ϕ X174, the most well-known virus of the *Microviridae* (micro: Greek for small). From the onset it was an *enfant-terrible*, so different from the other bacteriophages isolated during this period of extensive phage discovery. It was unusually tiny, readily passing through the smallest of ultrafilters. This biophysical characteristic defined its “race,” race X (Roman numeral ten) and the isolate was placed in a vial labeled #174, from which the phage derived its name, the race X phage (ϕ) in test tube #174. And there it sat, relatively unperturbed, for decades. The first electron micrographs revealed small, vague isometric particles, vastly different from the tailed morphologies, typically associated with bacteriophages.

The atypical capsid encapsulated an equally atypical genome, which shared qualities with both single-stranded (ss) RNA and double-stranded (ds) DNA. In 1959, Robert Sinsheimer determined that ϕ X174 genome was a single-stranded DNA molecule, the first one defined. While genetic maps of other phages consisted of orderly linear gene progressions, the ϕ X174 map was most peculiar with genes located within genes. When Sanger and colleagues sequenced the genome, the complex arrangement of overlapping reading frames was confirmed. It was so beguiling that many suspected that ϕ X174 had extraterrestrial origin. As the New York Times reported the theory, an advanced race engineered ϕ X174 and disseminated it into the cosmos where it would “persist until the evolution of intelligent life and finally of investigators interested in the genetics of phage.” Although attempts were actually made to decipher the hypothesized hidden message, they all failed. However, rumors persist that the code has been broken: it reads, “Behold this marvelous little thing.”

This small virus would continue to have a large impact on molecular biology. Arthur Kornberg and colleagues used it to elucidate the molecular mechanism of prokaryotic DNA replication. Masaki Hayashi and colleagues defined the ϕ X174 assembly pathway and were the first to demonstrate that viral DNA packaging could be achieved *in vitro*. As the genome sequence initially defied imagination in 1978, the atomic structure of the viral procapsid, the first such structure solved at this high resolution, beguiled structural virologists, with its unusual external scaffolding protein lattice. And when events were at their most bizarre, the skeletons in the family closet emerged: the gokushoviruses (Japanese for very small). These viruses escaped detection for decades, hiding within obligate intracellular parasitic bacterial hosts, Chlamydia and Bdellovibrio, or mollicutes, which are difficult to propagate in a laboratory setting. With the advent of environmental sequencing, these distant ϕ X174 relatives appear to be quite common, especially in marine environments. However, their hosts have yet to be identified.

The ϕ X174-like phages’ most distinguishing features have always been their extensive overlapping reading frames, in which new genes have been discovered as recently as 2014, and their tailless structure. Double-stranded phages use their tails to deliver their genomes. Lacking that structure, the mechanism of microvirus DNA penetration was entirely vague. In the 1970s, Jazwinski and colleagues recognized the critical involvement of protein H, a minor capsid protein, calling it the DNA pilot protein. They speculated that “protein H facilitates the transit of nucleic acid, a polyelectrolyte, through the membrane lipid bilayers, by the creation of a pore.” Thirty-five years later, their hypothesis was shown to be exceptionally insightful. In 2014, the atomic structure of the H protein’s central domain was determined. Ten H monomers form a tube that emerges from the capsid while on the host cell’s surface. Its unique biophysical features suggest energy utilization mechanisms that may be common during DNA transport through conduits.

Virion Morphology and Genome Content

All members of the *Microviridae* package circular, ssDNA genomes of positive polarity within T = 1 icosahedral capsids. High resolution atomic structures of several microvirus capsids, including ϕ X174, have been determined. The bulk of the capsid is composed of 60 copies of the coat protein (Fig. 1 and Table 1), which exhibits the common β -barrel fold, or Swiss jellyroll

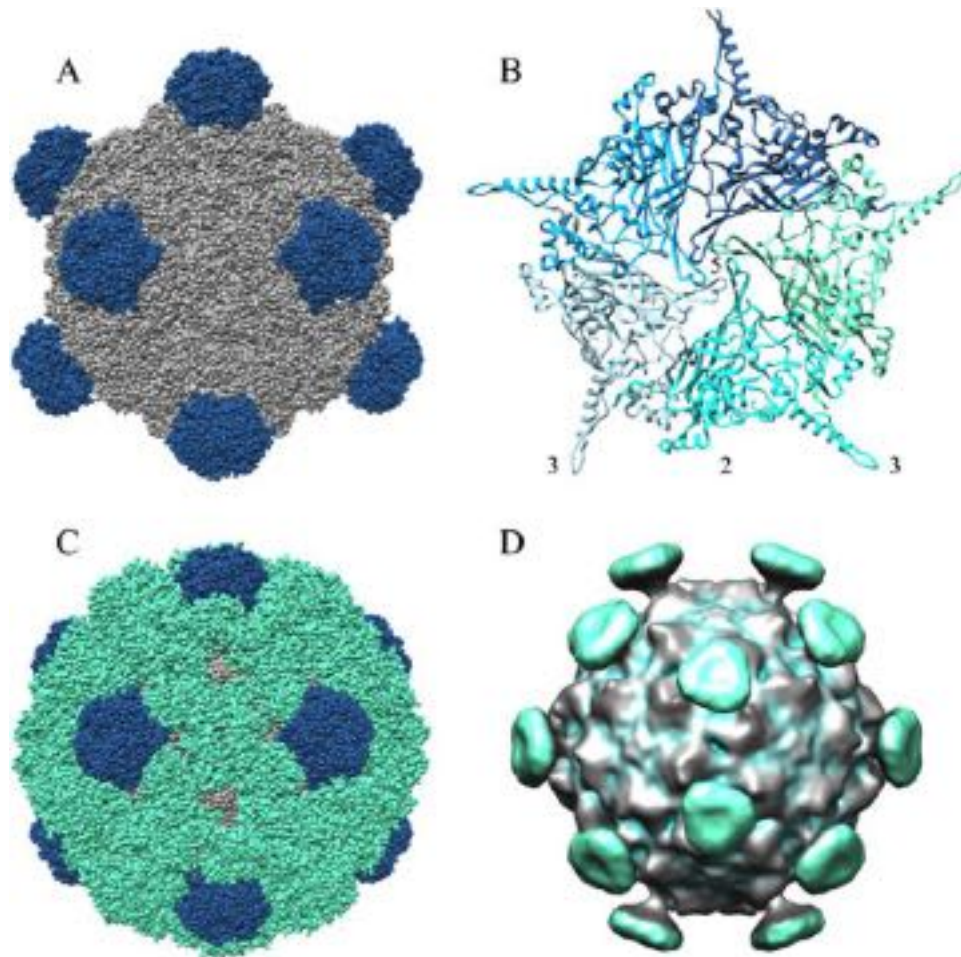


Fig. 1 Virion structures. (A) A space filling rendering of the ϕ X174 virion. The major coat protein is depicted in grey. The major spike protein is depicted in blue. (B) The arrangement of five ϕ X174 coat proteins around a five-fold axis of symmetry. Numbers indicate the axis of symmetry and the number of coat proteins interacting at that axis. (C) A space filling rendering of the ϕ X174 procapsid. The major coat, major spike, and external scaffolding proteins are respectively depicted in grey, blue, and aquamarine. (D) A cryo-EM image reconstruction of gokushovirus Chp2, courtesy of S. Kailasan and M. Agbandje-McKenna.

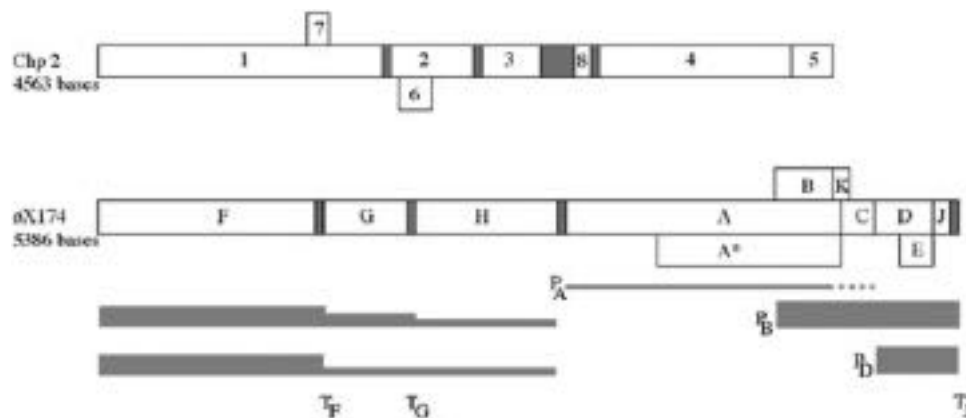
motif, found in many icosahedral virion capsid proteins (**Fig. 1(B)**). In addition to the coat protein, ϕ X174-like viruses also contain 60 copies each of the major spike protein G and DNA binding J proteins. There are also 12 copies of a fourth structural protein, the DNA pilot protein H, which is not visible within the capsid's overall icosahedral symmetry. Although the exact location of the H proteins are not known, 10 of them form a tube that emerges from the capsid once attached to the host cell's surface.

There are several major differences between microvirus and gokushovirus capsid superstructure and protein content. In the 250 Å diameter ϕ X174-like virions (**Fig. 1(A)**), 70 Å wide spikes, composed of five G proteins, rise 30 Å above the capsid surface. By contrast, the gokushoviruses lack major spike proteins: hence, the fivefold axes of symmetry are not decorated. However, the cryoelectron microscopy image reconstruction of gokushovirus Chp2 shows mushroom-shaped protrusions at the threefold axes of symmetry. These structures, which rise approximately 55 Å above the 270 Å diameter capsid (**Fig. 1(D)**), are composed of amino acid sequences from three interacting coat proteins (VP1). The gokushovirus VP2 is also a structural protein. Although its function has not been experimentally defined, bioinformatic analysis strongly suggests it is analogous to the DNA pilot protein H. Finally, gokushoviruses likely contain another protein, VP8, which is similar to the microvirus DNA binding protein. Both are short, highly basic peptides. However, it has yet to be detected in virions. This may be due to the protein's small size and the difficulties associated with propagating large quantities of gokushoviruses for biochemical characterization.

The genetic maps of ϕ X174 and Chp2 are depicted in **Fig. 2**. Gokushoviral genomes encode neither an external scaffolding nor a major spike protein. The absence of these genes accounts for their smaller genomes. The external scaffolding protein has at least three known functions in ϕ X174 morphogenesis. It nucleates procapsid assembly, mediates the subsequent elongation reaction, and stabilizes the assembled procapsid at the two- and threefold axes of symmetry (**Fig. 1(C)**). These functions are either not required or performed by different proteins in the gokushoviruses. Gokushovirus procapsid

Table 1 Microviridae gene products

ϕ X174-like protein	Gokushovirus
	ProteinFunction
A	VP4 Stage II and Stage III DNA replication.
A*	UD ^a An unessential protein for viral propagation. It may play a role in the inhibition of host cell DNA replication and superinfection exclusion.
B	VP3 Internal scaffolding protein, required for procapsid morphogenesis and the assembly of early morphogenetic intermediates. 60 copies present in the procapsid.
C	VP5? ^b Facilitates the switch from Stage II to Stage III DNA replication. Required for Stage III DNA synthesis.
D	NP ^c External scaffolding protein, required for procapsid morphogenesis. 240 copies present in the procapsid.
E	UD Host cell lysis.
F	VP1 Major coat protein. 60 copies present in the virion and procapsid.
G	NP Major spike protein. 60 copies present in the virion and procapsid.
H	VP2? DNA pilot protein need for DNA injection, also called the minor spike protein. 12 copies in the procapsid and virion.
J	VP8? DNA binding protein, needed for DNA packaging, 60 copies present in the virion.
K	UD An unessential protein for viral propagation. It may play a role optimizing burst sizes in various hosts.

^aUD, undetermined.^bA “?” indicates a hypothesized function based on bioinformatic data.^cNP, not present in Gokushoviruses.**Fig. 2** Genetic maps. (Top) gokushovirus Chp2. (Middle) ϕ X174. (Bottom) transcripts found in ϕ X174 infected cells. The promoters and transcription terminators are indicated on the linear map of ϕ X174. Line thickness indicates relative transcript abundance. The gene A transcript is very unstable; the terminator for this transcript is unknown. For protein functions, see Table 1.

assembly is most likely mediated by the internal scaffolding protein VP3, which is analogous to the ϕ X174 internal scaffolding protein B.

In the ϕ X174-like viruses, genes A and C encode proteins mediating DNA replication. Gokushovirus gene 4 and possibly gene 5 encode their respective homologues. Lastly, gene E encodes a lysis protein. A gokushoviral equivalent has yet to be identified. All ϕ X174-like genomes contain two additional genes, A* and K. They encode proteins with unknown or poorly defined function. Furthermore, neither is required for growth under laboratory conditions. Both genes are contained within larger genes. Thus, maintaining them does not increase genome size.

The ϕ X174-like phages fall into three major clades, each one represented by a canonical virus: ϕ X174, G4 and $\alpha 3$. Genome size corresponds to the three clades: ~ 5.3 kb, ~ 5.7 kb, and ~ 6.1 kb, for the ϕ X174, G4 and $\alpha 3$ clades, respectively. Although genomes vary in size, the capsids' internal volumes are nearly identical. The $\alpha 3$ -clade genomes are more complex. They contain two versions of gene C; C_{big} and C_{small} . Both proteins are synthesized in infected cells. Although both are not required for viability, they are required for optimal fitness. In addition, three small ORFs are found in the $\alpha 3$ -clade H-A intercistronic regions, which are significantly larger than those found in the other clades. The functions of the encoded proteins, or if they are synthesized, are unknown.

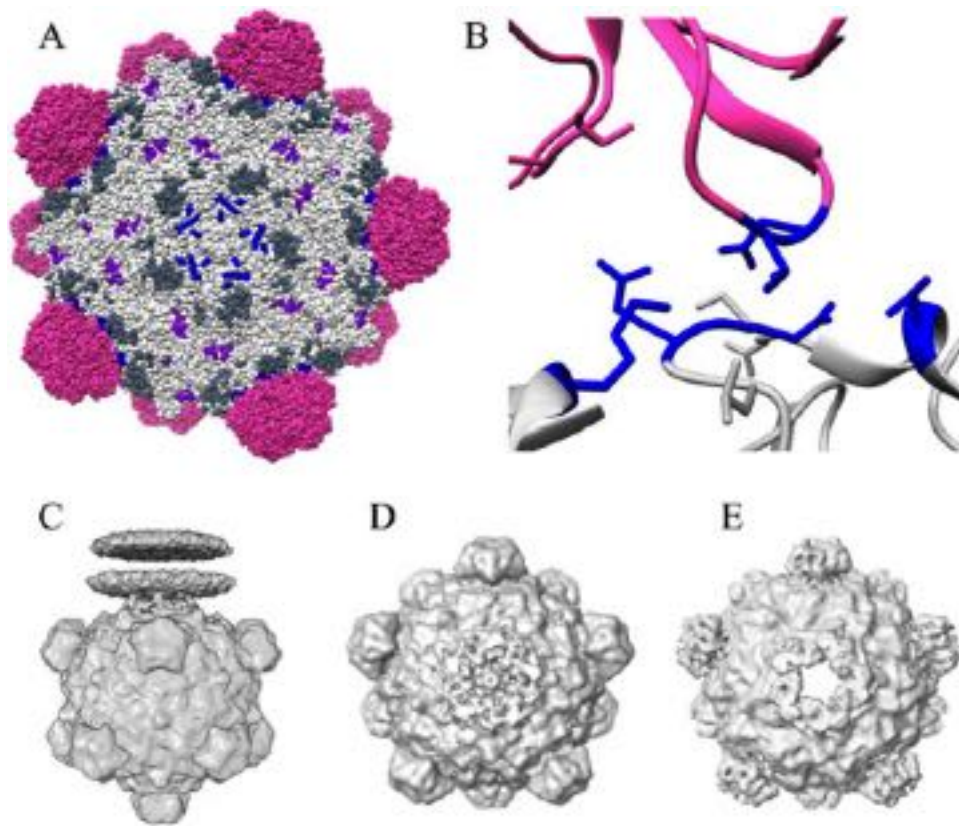


Fig. 3 Surface residues and conformational changes associated with host range, attachment, and eclipse. (A) Space filling rendering of the $\alpha 3$ X-ray structure, with one G protein spike removed. The coat protein is depicted in white, the spike protein in magenta. The 6-carbon sugar binding site, which likely mediates initial attachment, is depicted in purple. Dark grey residues denote the regions within the coat protein that may merge with the lipid bilayer (see panel C). Mutations that alter host range are depicted in blue. One G protein spike has been removed to illustrate the location of these substitutions within the coat–spike protein interface. (B) The clustering of mutations affecting host range within the coat–spike interface. Proteins are depicted as described above. Mutations at blue residues alter host range. (C) Cryo-EM image reconstruction of ϕ X174 interacting with a lipid bilayer. (D and E) Interactions with the lipid bilayer at the point of cell-virus contact removes the spike protein pentamer. Before DNA ejection the coat protein pentamer is closed (panel D), whereas a pore later forms for the release of the H-tube and DNA (panel E).

Host Cell Recognition, Attachment, Eclipse and Penetration

ϕ X174-like phages recognize terminal glucose and galactose moieties on host lipopolysaccharides (LPSs). The structure of the LPS is both necessary and sufficient to mediate both attachment and penetration. In X-ray structures of ϕ X174, six coat protein residues constitute six carbon sugar binding sites when Ca^{2+} is present (Fig. 3(A)). Due to the icosahedral symmetry of the capsids, each virion has 60 such sites, which likely mediate attachment to host LPS. This is a reversible reaction: infectious virions can be dissociated from host cell surfaces.

During eclipse, also called irreversible attachment, DNA-filled virions can still be visualized on the surface of host cells. However, particles are no longer infectious when removed from membranes. A cryo-EM structures of a DNA-filled particle attached to a lipid bilayer elucidate the conformational switches associated with eclipse (Fig. 3(C, D, and E)). In this structure, the major spike pentamer at the lipid-interacting vertex has been removed, and an annulus of coat protein residues merges with the lipid bilayer, anchoring the infecting virion to the membrane. An additional conformational change was observed in a post-DNA release structure: a pore formed at the attached vertex (Fig. 3(E)). The DNA translocating conduit, or H-tube, most likely emerged through this opening during penetration (see below).

The results of host range studies underscore the importance of coat-spike protein interactions during early infection (Fig. 3(A) and (B)). Despite ~90% amino acid identity, the natural host of microvirus $\alpha 3$ is *E. coli* C; whereas microvirus ST-1 is an *E. coli* K12-specific phage. Virions attached to and eclipsed on both native and unsusceptible hosts; however, they breached only their native host's cell wall. Thus, unsusceptible host-phage interactions promote off-pathway reactions, inactivating particles without penetration. Single amino acid substitutions in either the $\alpha 3$ coat or spike proteins expand host range. The substitutions cluster around the outer circumference of the coat-spike interface (Fig. 3(A) and (B)).

During DNA penetration, bacteriophages must propel their genomes across a formidable barrier: the cell wall. The Gram-negative prokaryotic cell wall is composed of two lipid membranes that surround a peptidoglycan layer. The emblematic tail structures found

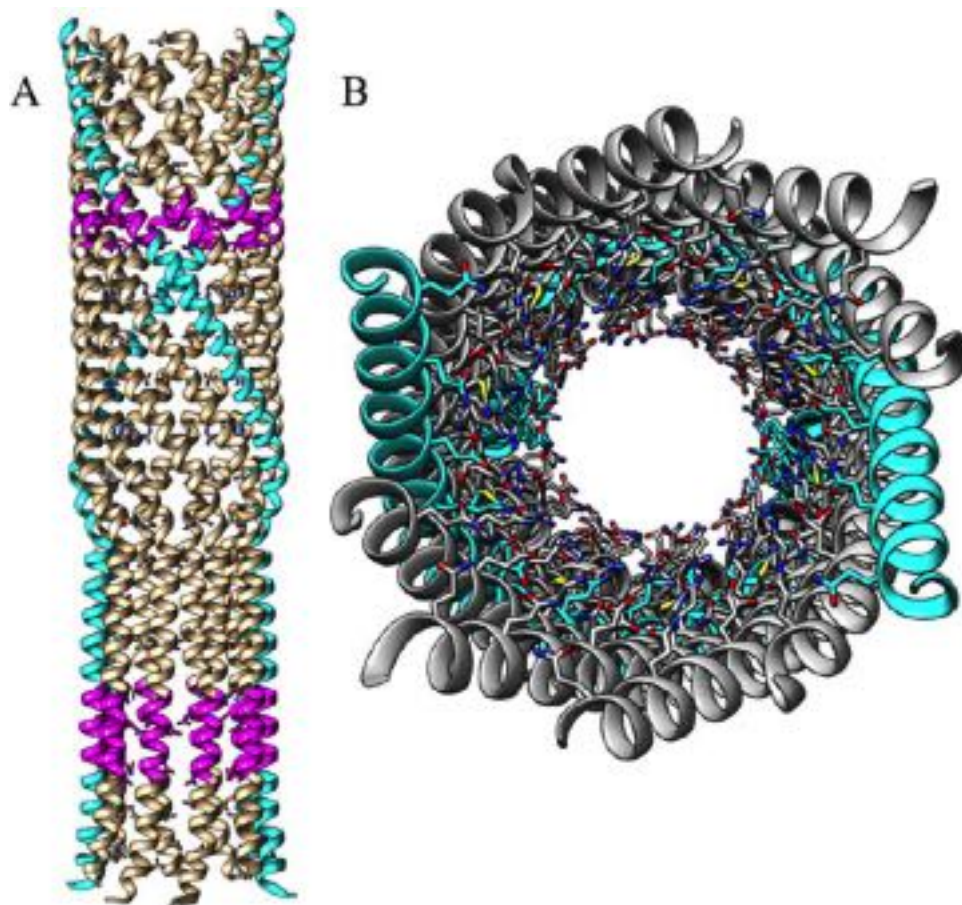


Fig. 4 The atomic structure of the ϕ X174 DNA translocating conduit, also called the H-tube. (A) Two of the ten H protein monomers are depicted in cyan. The structure is divided into two domains, which are distinguished by repeating heptad or hendecad motifs. A heptad is a group of seven, whereas a hendecad is a group of eleven. An example of each motif is highlighted in magenta. (B) Looking through the H-tube from its C-terminal opening. The inward facing amide and guanidinium containing side chains are depicted as stick models. As above, two monomers are depicted in cyan.

on double-stranded (ds) phages often perform this function. However, the ϕ X174-like phages are strictly icosahedral and lack and external tail structure. Although early studies (circa 1975) indicated that the DNA pilot protein H mediated penetration, efforts to visualize the protein within the virion structure have been dismally unsuccessful. Thus, the penetration mechanism remained unknown for almost 40 years. Then, a 13-year effort to both purify and generate diffracting H protein crystals finally yielded results.

Although protein H is a monomer in all viral assembly intermediates (see below), it crystalized as an oligomer (**Fig. 4(A)**). Ten α -helical monomers formed a 170 Å-long, coiled-coil, helical barrel. This barrel, or tube, is long enough to span membrane adhesion sites, regions of the cell wall where the inner and outer membranes make contact. Infecting particles tend to congregate at these sites. The H-tube's internal diameter is 22 Å, which can accommodate two single-stranded (ss), un-base paired DNA strands. Since the genome is circular, two ssDNA strands must simultaneously pass through the conduit. The helices run parallel to each other and a structural kink divides the tube into two domains. Domain A has a 11,3 coiled-coil structure (11 residues per 3 helical turns), whereas domain B has a 7,2 coiled-coil structure. The tube has been visualized *in situ* by cryo-EM tomography. Upon cell contact, a fully formed tube emerges from the capsid and descends into the host cell wall. After DNA delivery, the tube dissociates within the cell wall, which is necessary to preserve the cell's membrane potential.

The tube's inner surface is lined with glutamine, asparagine, and arginine amino acid side chains (**Fig. 4(B)**). Containing amide and guanidinium groups, these side chains can form hydrogen bonds with nucleic acids. The side chains point toward the tube's center and are angled towards the host cell's cytoplasm. This orientation may prevent the genome from moving backwards during transit, thus creating a one-way passage. In addition, these side chains may play two other roles during DNA transport. Although the tube's internal diameter can accommodate two ssDNA strands, it is not wide enough to accommodate the hair-pin structures formed by ss nucleic acids. By competition, these side chains, which interact with nitrogenous bases, may eliminate hair-pin structures during translocation. These tube-DNA interactions may also modulate the release of potential energy used when transporting DNA.

Mature virions; containing a highly compacted, volumetrically constrained genome; may represent the particle's highest energy state. This inherent potential energy may be used to initiate genome delivery and begin moving the genome into the host cell. In most biological systems, potential energy performs work more efficiently if released in small, usable increments. For example, during oxidative phosphorylation, for example, electrons are not directly passed from NADH_2 to O_2 . The released energy from a direct transfer would be too great for the cell to effectively harness. Instead, the electrons flow through the electron transport system, in which energy is liberated in a series of exploitable reactions. The amide and guanidinium groups may regulate genome delivery by slowing the transport process, creating a frictional force that counters the pressure driving the genome into the cell. The system can be compared to a compressed gas driven machine. In this analogy, the gas cylinder is the capsid, the packaged genome is the compressed gas, the amide and guanidinium groups are the regulatory gauge, and the H-tube is the pipe connected to the cylinder. If too much air (energy) is released at one time, the pipe could fail.

The frictional force created by the amide and guanidinium groups may prevent the transport of the entire genome, which appears to require an additional energy source. This source could be linked to DNA replication. If cells are not actively synthesizing DNA, the entire genome is not efficiently transferred. Thus, the first stage of DNA synthesis and penetration may be concurrent processes. As the ssDNA enters the cell it is converted into double-stranded molecule. The phosphate bond hydrolyzed within the deoxynucleotide triphosphate (dNTP) contains more energy than the newly formed phosphate bond in the DNA backbone. Using this small energy differential, the host cell DNA polymerase may help pull the ssDNA genome into the cell.

Due to their recent discovery, the details of gokushovirus attachment and penetration have yet to be defined. However, unlike the ϕ X174-like viruses, gokushovirus Chp2 appears to utilize a host cell protein as a receptor molecule, as opposed to LPS. The results of bioinformatic approaches combined with chlamydia phage host range studies suggest that the threefold related protrusion may govern host cell specific attachment. However, there is no direct experimental evidence to support this hypothesis.

DNA Replication

The studies of ϕ X174 DNA replication are historically significant. Positive polarity ssDNA replication strategies are complex, occurring in three separate stages, which are described below (Fig. 5). Kornberg and colleagues reconstituted the first two stages *in vitro* while Hayashi and colleagues reconstituted the third stage of DNA replication, which includes concurrent packaging of the single-stranded genome. Collectively, these studies established the first defined viral genome replication process on the biochemical level.

As single-stranded (ss) viral DNA is delivered into the infected cell via the H-tube (Fig. 5), Stage I DNA replication commences. During this stage, the ss genome is converted into a covalently closed, double-stranded (ds), circular molecule, called replicative form one (RFI) DNA. Since purified ssDNA produces progeny when transfected, host cell proteins are both necessary and sufficient for stage I replication *in vivo*. *In vitro*, a stem-loop structure in the FG intercistronic region serves as the host cell protein recognition site, which initiates the assembly of the primosome. Whether other regions of the genome can perform this function *in vivo* has yet to be determined. After primosome assembly, the complex migrates along the ssDNA in a 5'→3' direction synthesizing the requisite RNA primers for DNA replication. Addition of the holoenzyme leads to chain elongation. The 13 host cell proteins required for this stage of synthesis are the same proteins involved in replicating the host cell chromosome.

During stage II DNA synthesis, RFI DNA is amplified. In addition to the host cell proteins required for stage I replication, stage II replication is dependent on the viral A protein and the host cell rep protein, which functions as a helicase. The viral A protein binds to the stage II/III origin of replication and nicks it to initiate (+) strand synthesis, which occurs via a rolling circle mechanism. After nicking, protein A forms a covalent ester bond with the DNA, forming a relaxed circular molecule called RFII. The host cell rep protein unwinds the helix and the host ssDNA binding protein (Ssb) stabilizes the separated strands. After one round of rolling circle synthesis, protein A cuts the newly generated origin and acts as a ligase, generating a covalently closed circular molecule. Minus strand synthesis is mechanistically similar to stage I DNA synthesis.

Stage III DNA synthesis involves the concurrent synthesis and packaging of the ssDNA genome. Procapsids and viral protein C are required for this reaction along with all the proteins involved in the previous stages of replication with the exception of the Ssb. Thus, a single-stranded genome is not synthesized unless there is a procapsid to accept it. As the dsDNA is unwound at the stage II/III origin of replication, competition between Ssb and protein C for binding determines whether the dsDNA or ssDNA will be synthesized. This has been demonstrated both *in vitro* and *in vivo*. *In vitro*, excess Ssb promotes another round of stage II DNA synthesis occurs. By contrast excess protein C inhibits stage II DNA replication, and single-stranded genomes are concurrently synthesized and packaged. *In vivo*, altering the Ssb/protein C ratio by exogenously over-expressing cloned genes is detrimental to plaque formation. If the ratio increases, ssDNA synthesis is delayed. Thus, fewer mature virions are formed before programmed cell lysis. Ratio decreases result in fewer RF molecules, which is the template for both genome biosynthesis and transcription. Thus, the detrimental effects are more global.

When packaging occurs, proteins A, C, and rep form a complex on the dsDNA that docks to procapsids, presumably in a groove that spans one of the twofold axes of symmetry. Procapsid binding does not occur in the absence of protein C. Mechanistically, stage III DNA synthesis is similar to the stage II (+) rolling circle synthesis. As the new (+) strand is synthesized, the displaced strand is translocated into the procapsid. After one round of synthesis, protein A, which is covalently attached to the origin of replication, cuts the newly synthesized origin and acts as a ligase to generate a covalently closed circular molecule on the displaced strand.

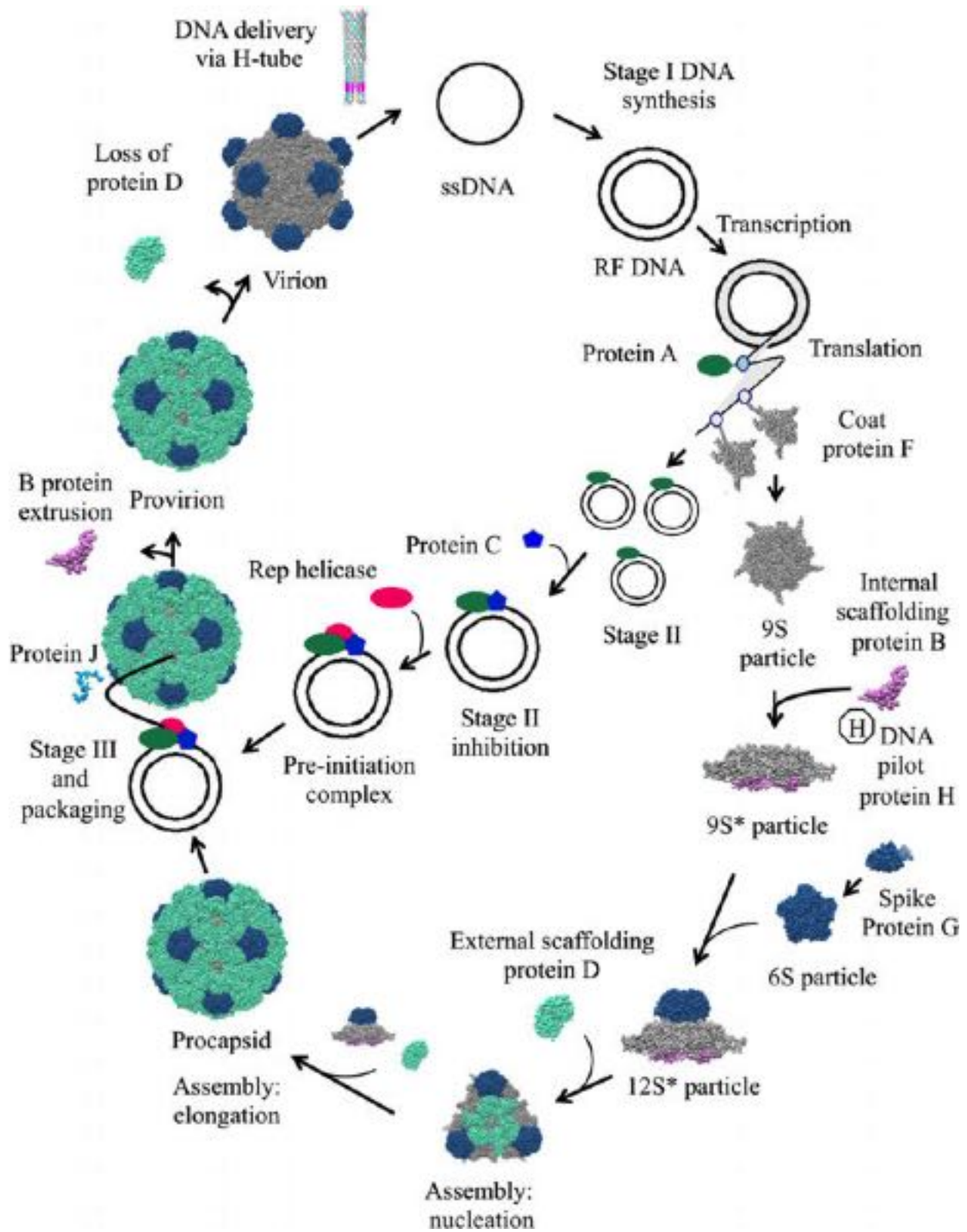


Fig. 5 Morphogenesis and DNA replication of ϕ X174.

Gene Expression

Since *Microviridae* contain single-stranded genomes of positive polarity, Stage I DNA synthesis, which generates the negative strand, must occur before transcription. Unlike large bacteriophages, *Microviridae* gene expression is neither temporal nor mediated by *trans*-acting mechanisms. Thus, the timing and relative production of viral proteins is entirely dependent on *cis*-acting regulation signals: promoters, transcription terminators, mRNA stability sequences, and ribosome binding sites. In the ϕ X174-like viruses, promoters are found upstream of genes A, B, and D. Terminators are found after genes J, F, G, and H (Fig. 2). The terminators are not 100% efficient; thus, a wide variety of transcripts are produced. There is a rough correlation between gene transcript abundance and the amount of encoded protein required for the viral life cycle. For example, the virus requires more copies of protein D, the external scaffolding protein, than any other viral protein. Gene D transcripts are the most abundant in the cell. Similarly, there are more transcripts of genes F, J, and G than transcripts of gene H. The relative stoichiometry of these structural proteins are 5:5:5:1, respectively. Protein expression is also affected by mRNA stability. Each mRNA species decays with a characteristic rate. Transcripts of gene A decay very rapidly, ensuring that this non-structural protein is not over expressed. And finally, regulation can also be achieved on the translational level. Despite the gene E's location within gene D on the most abundant transcript, very few E proteins, which mediate cell lysis, are translated due to an extremely pathetic ribosome binding site.

Morphogenesis

Despite extensive searches, a dependence on host molecular chaperones, such as GroEL and GroES, has never been documented. Thus, chaperone-independence is another characteristic that distinguishes the ϕ X174-like viruses from dsDNA bacteriophages. The first virally encoded assembly intermediates are the 9S and 6S particles, respective pentamers of the major coat F and spike G proteins (Fig. 5).

Unlike most viruses that utilize a scaffolding protein to mediate morphogenesis, two scaffolding proteins orchestrate ϕ X174 assembly, temporally dividing the pathway into two discernable phases. The early phase is mediated by the internal scaffolding protein B. After or during 9S particle formation, five B proteins bind to the pentamer's underside, which becomes the capsid's internal surface. The B proteins also recruits one copy of the DNA pilot protein H. However, the DNA pilot protein is not required for assembly, as H-less particles can be assembled into uninfected, virus-like particles. The finished product is the 9S* particle. Early assembly then concludes when a major spike protein G pentamer, the 6S particle, binds to the top of the 9S* particle, forming the 12S* assembly intermediate. As previously discussed, the DNA pilot H protein forms a decameric tube for DNA delivery. However, 9S* and 12S* assembly intermediate contain one copy of the H protein. Thus, the H-tube's discovery was not expected and formation must occur after capsid assembly.

Coat-internal scaffolding protein contacts are mediated primarily by aromatic interactions, creating an elaborate π - π stacking network (Fig. 6(D)). While the X-ray structure elucidates the molecular details of the coat-scaffolding protein interactions, genetic and biochemical analyses elucidate the functional consequences. The π - π stacking network has been mutationally modified to both eliminate or alter B protein binding. The former leads to 9S particles aggregation, whereas the latter kinetically traps 9S* or 12S* assembly intermediates. Thus, the B protein temporally guides the coat protein through assembly. It suppresses interactions between coat protein pentamers until late assembly and ensures that 9S* and 12S* particles are primed to respectively associate with 6S pentamers and the external scaffolding protein D, which orchestrates late assembly. Mutations within the B protein's N-terminus, which is not part of the π - π stacking network, affect H protein incorporation.

During the second phase of assembly, 240 copies of the external scaffolding protein D, arranged as dimers, organize twelve 12S* particles into the procapsid. External scaffolding proteins are rare. The only other known external scaffolding proteins belong to parasitic satellite virus systems such as bacteriophage P4. The satellite virus encodes an external scaffolding protein that hijacks the helper virus's capsid proteins, forcing them to construct a smaller capsid, which is only suitable for the satellite virus's smaller genome. Thus, the ϕ X174-like viruses are the only non-satellite viruses known to utilize an external scaffolding protein. This protein performs many of the functions typically associated with internal scaffolding proteins in single scaffolding protein systems: it nucleates procapsid assembly, organizes procapsid assembly precursors during elongation, and stabilizes the procapsid during DNA packaging.

Unique in nature, the ϕ X174 external scaffolding protein is also biophysically unique, having the inherent ability to achieve at least six different structures. The D proteins operate as dimers. 120 D protein dimers organize twelve, 12S* particles into a procapsid. There are four D subunits (D₁, D₂, D₃ and D₄) per coat protein in the procapsid (Fig. 6(A) and (C)). They are arranged as two similar, but not identical, asymmetric dimers (D₁D₂ and D₃D₄). Each subunit makes a unique set of contacts with the underlying coat protein, the spike protein, and neighboring D protein subunits. Accordingly, the structure of each subunit is unique. The atomic structure of the assembly naïve D protein dimer has also been determined. The subunits within that dimer, D_AD_B, appear poised to achieve the four structures found in the procapsid. D_A has a structure somewhat between D₁ and D₃, while D_B has a structure midway between D₂ and D₄.

The 12S* + D protein $\rightarrow$ procapsid reaction has been analyzed both *in vitro* and *in vivo*. Capsid assembly reactions are typically divided into two phases: nucleation and elongation. Nucleation reactions, which yield a short-lived nucleation complex, are generally higher order reactions and/or require higher critical concentrations than elongation reactions. Therefore, once a reaction is nucleated, a slower process, the elongation reactions quickly go to completion. This can be regarded as an evolutionary mandate.

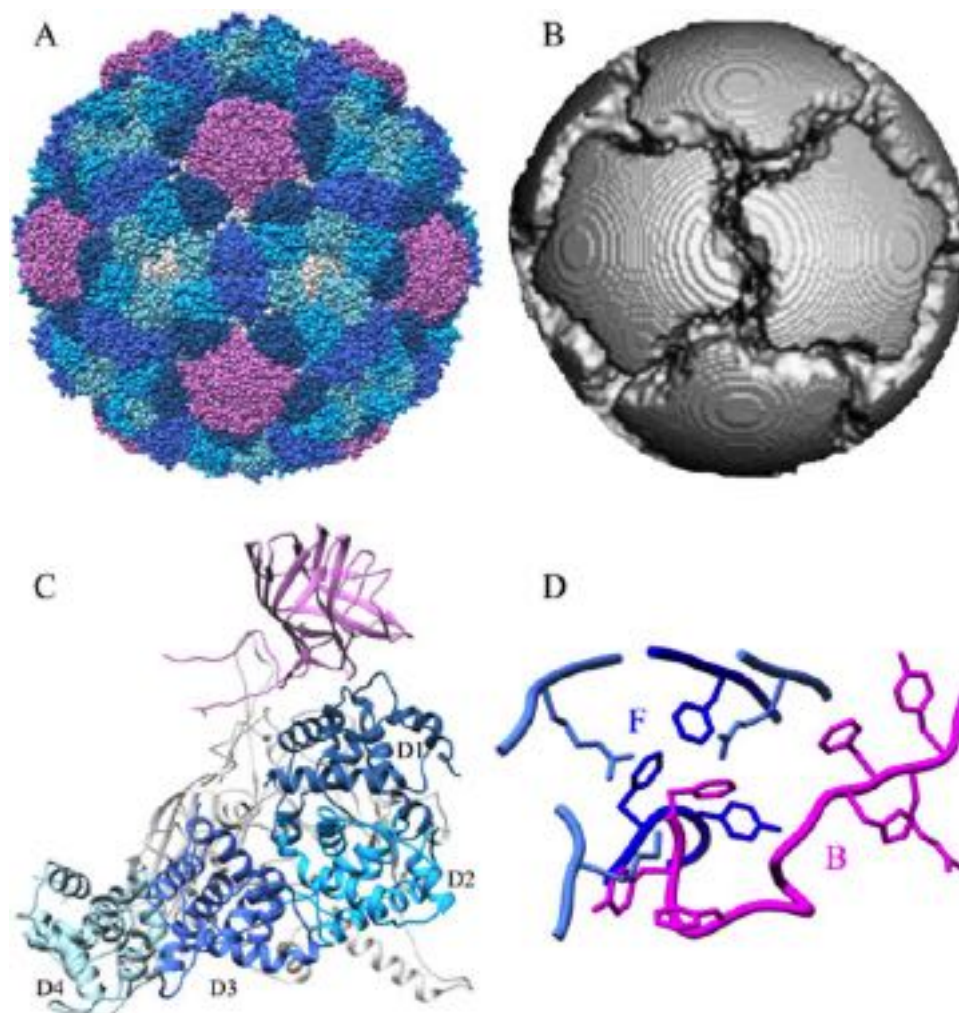


Fig. 6 Cryo-EM and atomic renderings of the ϕ X174 procapsid. (A) Space filling rendering of the ϕ X174 procapsid. The four structurally unique external scaffolding proteins (D₁-D₄) are depicted using the colors indicated in panel C. (B) Cryo-EM rendering of the procapsid with the removal of the external scaffolding protein. (C) The four unique D proteins, depicted in shades of blue, are associated with the coat and spike proteins, respectively colored white and periwinkle. (D) Aromatic interactions between the ϕ X174 coat and internal scaffolding protein. The internal scaffolding protein is depicted in magenta. The coat protein is depicted in two shades of blue. Dark blue residues contain aromatic side chains, whereas residues with positively charged side chains are light blue. The most extensive π - π stacking interactions; the attractive, noncovalent interactions between aromatic rings; occur between the last amino acid in protein B and three aromatic coat protein residues.

If nucleation reactions occurred more readily than elongation reactions, subunits would be divided among too many nucleation complexes. This would quickly exhaust the assembly subunit pool and result in partially formed capsids. In this regard, the 12S* + external scaffolding protein D $\rightarrow$ procapsid transition is very similar to single scaffolding protein-mediated assembly systems. However, ϕ X174 has two scaffolding proteins, which has resulted in an interesting role reversal among the reactants.

Compared to other systems, the external scaffolding protein behaves more like a viral coat protein, whereas the coat protein, a component of the 12S* particle, acts more like a scaffolding protein. Typically, viral coat proteins form aberrant assemblies in the absence of scaffolding proteins, whereas scaffolding proteins alone are relatively inert. Nucleation reactions require higher coat protein critical concentrations than scaffolding protein concentrations. During elongation, the scaffolding proteins control fidelity. By contrast, the ϕ X174 external scaffolding protein D self-associates to form large heterogeneous spherical complexes. The coat proteins in the 12S* particles do not self-associate *in vitro*, *in vivo*, or in the assembled procapsid, which is held together almost exclusively by D-D protein contacts across the two-fold axes of symmetry. Remarkably, there appears to be no associations between coat protein pentamers (**Fig. 6(B)**). Lastly, nucleation dependent a high scaffolding:coat protein ratios, and 12* particles in reaction control elongation fidelity.

Gokushovirus genomes do not encode an external scaffolding protein and their assembly pathway remains to be elucidated. A gokushovirus particle containing VP3 in addition to the structural coat and DNA pilot proteins has been isolated. Unlike virions, these particles are devoid of DNA. This particle most likely represents a gokushovirus procapsid and indicates that VP3, which is not found in virions, functions as an internal scaffolding protein.

DNA Packaging and the DNA Binding Protein

Genome biosynthesis and packaging are concurrent processes in ϕ X174. The pre-initiation complex; consisting of the host cell Rep viral A and C proteins; associates with the procapsid forming the 50S complex. As described above, the viral A protein binds the origin of replication in replicative form DNA. The results of genetic studies indicate that the pre-initiation docking site resides along a two-fold axis of symmetry. The DNA binding protein J enters the procapsid during packaging and is absolutely required for genome encapsidation. The N-termini of microvirus J proteins are extremely basic and bind the genome via simple charge-charge interactions. Once in the procapsid, the J protein's C-termini, which are very hydrophobic and aromatic, compete with the internal scaffolding protein for binding to a cleft in the viral coat protein. This competition results in the extrusion of the internal scaffolding protein during the packaging reaction.

Although the majority of the packaged genome appears to exist as a dense core, the J protein's basic amino acids, along with a small cluster of basic capsid residues, tether the genome to the capsid's inner surface. In the ϕ X174 X-ray structure, the J protein forms an S shaped polypeptide chain devoid of secondary structure. The C-terminus is tightly associated with the cleft located near the center of the coat protein. Each of the 60 J proteins traces a path toward the five-fold axis of symmetry, crosses over to the adjacent capsid protein, and veers toward the C-terminus of the adjacent J protein. Thus, the DNA binding protein guides the incoming genome into a somewhat icosahedrally ordered conformation and a portion of the genome is ordered in the X-ray structure.

The biophysical characterization of ϕ X174 particles packaged with foreign genome-length DNA or mutant DNA binding proteins suggests that protein-DNA interactions influence the final stages of morphogenesis. Morphogenesis terminates with the provision to virion transition: the dissociation of the external scaffolding protein and an 8.5 Å radial collapse of the capsid pentamers around the genome. The tethered genome constrains the spatial orientation and secondary structure of the remaining nucleotides. Therefore, altering the tether or the base composition of the packaged nucleic acid may affect the magnitude or integrity of the collapse. However, the role of DNA-capsid interactions in ϕ X174 is not as dramatic as those seen in other viral systems in which abrogating genome-capsid interactions leads to severely aberrant particles.

Lysis

Unlike large ds bacteriophages, ssDNA microviruses do not have the genetic capacity to encode a two-component holin-endolysin system. Instead, they have evolved a small protein, lysis protein E, that inhibits peptidoglycan biosynthesis. Thus, lysis is dependent on host cell division, during which the cell becomes sensitive to osmotic pressure. The results of several elegant genetic studies elucidated protein E function. By selecting for lysis resistant cells, Ryland Young and colleagues first uncovered the *slyD* (sensitivity to lysis) gene, which encodes a peptidyl-prolyl cis-transferase-isomerase, or PPIase. However, it is unlikely that the *slyD* gene product is the E protein target. Considering the function of PPIase's in protein folding and the E protein's five prolyl bonds, it seemed more likely that the E protein was a substrate for the host cell enzyme. In fact, gene E mutants, *Epos* (plates on *slyD*), can be readily isolated. In order to determine the target of protein E, *Epos* proteins were used in a second round of selection with *slyD* cells. The surviving colonies contained mutations in the *mraY* gene, which encodes translocase I. This enzyme catalyzes the formation of the first lipid-linked intermediate in cell wall biosynthesis. In the presence of protein E, its activity is greatly reduced. Thus, it is likely the E protein's direct target.

Evolution and Evolutionary Studies

The evolutionary relationship between the gokushoviruses and the ϕ X174-like viruses remains somewhat obscure. There is a deep evolutionary rift between the two groups, with no known intermediate species. The rift is likely a function of their host's biology and not the hosts' evolutionary relationship. The gokushoviruses have been isolated from obligated intracellular parasitic bacteria and mollicutes. For example, the *Bdellovibrio* host for the gokushovirus ϕ MH2K is a proteobacterium, like the ϕ X174-like phage hosts, but ϕ MH2K is much more closely related to the phages of chlamydia, which is in an unrelated bacterial phylum. Two of the primary differences between the two groups are the external scaffolding and spike proteins.

Many gokusho or gokusho-like viral genomes have been detected in environmental samples, primarily oceanic, and microbiomes. However, their hosts have yet to be identified. Nonetheless, these viruses likely play a major role in the shape and functioning of ecosystems.

Evolution of a Two Scaffolding Protein System

While the existence of two scaffolding proteins is unique in any system, it is particularly peculiar for T = 1 capsids. No other T = 1 virus requires a scaffolding protein, let alone two. The results of an exhaustive search for new *E. coli* ϕ X174-like phages indicate that the 47 known species can be divided into three separate clades, typified by bacteriophages ϕ X174, G4 and α 3. Although there is evidence for some horizontal gene transfer between the species, the extent is considerably lower than that observed for dsDNA

viruses. The one gene that seems to be the most recent acquisition, or at least its present form, is the external scaffolding protein gene, which appears to have originated in the ϕ X174 clade and spread to the two the others.

An accumulation of both genetic, biochemical, and structural data indicates that the external scaffolding protein is more essential for morphogenesis. The internal scaffolding protein, on the other hand, may be best viewed as an efficiency protein, facilitating several morphogenetic reactions, but not absolutely essential for any one in particular. This hypothesis was tested by evolving a sextuple mutant strain that no longer requires the internal scaffolding protein. Although mutations in structural and external scaffolding proteins did arise, two mutations were found in promoters upstream of the external scaffolding gene, leading to its over-expression. These two regulatory mutations, along with one conferring a substitution in the external scaffolding protein, have a kinetic effect on virion assembly. This indicates that one function of the internal scaffolding protein is to lower the critical concentration of the external scaffolding protein needed to nucleate procapsid assembly.

The morphogenesis of wild-type ϕ X174 is extremely rapid, progeny virions can be detected as quickly as 5.0 min post infection, which may be critical for a small phage without the genome content to encode super infection exclusion functions. At five minutes post infection, most other phages are just concluding early gene expression. Rapid morphogenesis may be a consequence of having two scaffolding proteins, allowing the ϕ X174-like viruses to compete with the larger and vastly more prevalent dsDNA phages. In this evolutionary model, the external scaffolding protein is a recent acquisition. Those phages that did not acquire the gene, the gokushoviruses, may have persisted by finding a niche free of competition: obligate intracellular parasitic hosts like chlamydia.

ϕ X174-Like Viruses as a Model System for Experimental Evolution

Due to the small genomes and the ability to interpret amino acid substitutions within the context of the virion and procapsid atomic structures, the ϕ X174-like viruses have become one of the leading systems for molecular evolution analyses. In studies pioneered by Jim Bull, Holly Wichman and colleagues, viruses are placed under selective conditions, and grown for numerous generations either in chemostats or by passaging viruses between fresh cell cultures. Individual genomes are sequenced at various time intervals, which monitors the appearance and disappearance of mutations. The obtained mutations differ depending on the selective conditions. While these studies identify beneficial changes in both structural and nonstructural proteins, many mutations are in genetic regulatory sequences, which most likely optimize the relative level of viral proteins synthesized under the experimental conditions.

Further Reading

- Aoyama, A., Hamatake, R.K., Hayashi, M., 1983. *In vitro* synthesis of bacteriophage ϕ X174 by purified components. *Proceedings of the National Academy of Sciences of the United States of America* 80, 4195–4199.
- Bernhardt, T.G., Struck, D.K., Young, R., 2001. The lysis protein E of ϕ X174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis. *Journal of Biological Chemistry* 276, 6093–6097.
- Cherwa Jr., J.E., Tyson, J., Bedwell, G.J., *et al.*, 2017. ϕ X174 procapsid assembly: The effects of an inhibitory external scaffolding protein and resistant coat proteins *in vitro*. *Journal of Virology* 91. (pii: e01878-01816).
- Clarke, I.N., Cutcliffe, L.T., Everson, J.S., *et al.*, 2004. Chlamydia phage Chp2, a skeleton in the ϕ X174 closet: Scaffolding protein and procapsid identification. *Journal of Bacteriology* 186, 7571–7574.
- Dokland, T., McKenna, R., Ilag, L.L., *et al.*, 1997. Structure of a viral procapsid with molecular scaffolding. *Nature* 389, 308–313.
- Doore, S.M., Baird, C.D., Roznowski, A.P., Fane, B.A., 2014. The evolution of genes within genes and the control of DNA replication in microviruses. *Molecular Biology and Evolution* 31, 1421–1431.
- Doore, S.M., Fane, B.A., 2016. The microviridae: Diversity, assembly, and experimental evolution. *Virology* 491, 45–55.
- Hayashi, M., Aoyama, A., Richardson, D.L., Hayashi, N.M., 1988. Biology of the bacteriophage ϕ X174. In: Calendar, R. (Ed.), *The Bacteriophages*, vol. 2. New York: Plenum Press, pp. 1–71.
- Kornberg, A., 1980. *DNA Replication*. San Francisco: Freeman.
- Rokyta, D.R., Burch, C.L., Caudle, S.B., Wichman, H.A., 2006. Horizontal gene transfer and the evolution of microvirid coliphage genomes. *Journal of Bacteriology* 188, 1134–1142.
- Roznowski, A.P., Fane, B.A., 2016. Structure-function analysis of the ϕ X174 DNA-piloting protein using length-altering mutations. *Journal of Virology* 90, 7956–7966.
- Sun, L., Young, L.N., Zhang, X., *et al.*, 2014. Icosahedral ϕ X174 forms a tail for DNA transport. *Nature* 505, 432–435.
- Sun, Y., Roznowski, A.P., Tokuda, J.M., *et al.*, 2017. Structural changes of tailless bacteriophage ϕ X174 during penetration of bacterial cell walls. *Proceedings of the National Academy of Sciences of the United States of America* 114, 13708–13713.
- Szekely, A.J., Breitbart, M., 2016. Single-stranded DNA phages: From early molecular biology tools to recent revolutions in environmental microbiology. *FEMS Microbiology Letters* 363, fnw027.
- Wichman, H.A., Brown, C.J., 2010. Experimental evolution of viruses: Microviridae as a model system. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, 2495–2501.

Single-Stranded RNA Bacterial Viruses

Peter G Stockley, University of Leeds, Leeds, United Kingdom

Junjie Zhang, Texas A&M University, College Station, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of P.G. Stockley, Icosahedral ssRNA Bacterial Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00760-3.

Glossary

Pfu Stands for plaque-forming unit; that is the number of phage particles per host cell required to generate productive infections. Theoretically this value could be 1 but very few of the phage that attach to the edge of a pilus will get access to the inside of a cell, so Pfu values in the 100s are not uncommon.

Polycistronic RNA The term cistron was coined to mean any section of an mRNA that gets translated. The RNA phage genomes are “polycistronic” because they encode multiple protein products.

Quasi-equivalent conformers These are different conformations of the identical polypeptide sequence that form at the differing symmetry-related positions of viral capsids.

Translational repression Refers to the process whereby a protein product interacts at a defined site within a mRNA molecule preventing translation of the downstream cistron.

Triangulation number, T Is defined mathematically as $T = (h^2 + hk + k^2)$, where h and k are any pair of positive integers, including zero. For instance giving rise to triangulation numbers of 1 for h, k = 1,0, and 3 for h, k = 1,1. These h and k values can be determined by analysis of real viruses by analyzing them in terms of networks of equilateral triangles that cover the surface of a sphere. The details of this analysis can be found in the article by Caspar and Klug listed under further reading.

UTR Untranslated region, found at the 5' and 3' ends of phage RNA genome where they form specific secondary structures recognized by the replication enzyme.

Icosahedral ssRNA Bacterial Viruses

Background

Some *E. coli* cells contain a separate genetic element, the F factor plasmid, in addition to their chromosomal DNA. Such F+ bacteria produce a pilus protein that self-assembles a hair-like appendage, the pilus, on the outside of the cell that is an essential part of conjunctive DNA transfer between bacterial cells. This primitive version of sexual genetic transfer is known as conjugation and the term “male” is used for bacteria carrying the F factor. A group of bacterial viruses, the RNA phages exploit this feature binding to the sides of the pilus and transferring their ssRNA genomes into the bacterium as the pilus retracts, the latter being part of its natural function. Tim Loeb working in Norton Zinder's laboratory at the Rockefeller University in New York was the first to realize that there might be such male-specific bacterial viruses, bacteriophages. Together they isolated and began to characterise such phages from an initial sample of raw sewage. One intensively studied version known as MS2 is believed to have been named because it was male-specific factor 2, although there are unconfirmed stories that the MS also stands for Metropolitan Sewer, the source of some of the samples they screened for phage!

Similar phages, all now grouped as *Leviviruses*, were rapidly isolated world-wide after this initial discovery and we now know that such phages are extremely common in the environment with up to 10^7 Pfu/mL in sewage. Both human and animal hosts appear to harbor *E. coli* populations able to support such phages which are thus termed coliphage. Virtually identical phages have been found that are specific to *Acinetobacter* (AP205) and other gram-negative bacteria. In such phages the infection process still occurs via pili but these are distinct from the sex pili of *E. coli*. Phage PRR1, for instance, infects hosts via pili encoded by incompatibility type P plasmids which are widely dispersed allowing the phage to infect a range of different bacterial cells. On the basis of immunological cross-reactivity and genome organization the *Leviviridae* are classified into two genera. These have been further sub-divided into two groups. All the phages are members of the *Leviviridae* or the *Alleloleviridae*, the latter being distinguished by having slightly longer genomes and because of the presence of read-through products from the CP gene (Fig. 1). Sub-groups I and II are typified by phages MS2 and GA, respectively, whilst typical Group III and IV phages are Q β and SP, respectively.

They have a single copy, positive-sense, single-stranded genomic RNA (gRNA) of ~3–4.5 kb in length, that encodes a maturation protein (MP, also called A or A2, depending on the phage), a coat protein subunit (CP), a replicase subunit and a lysis protein (Fig. 1). The two most studied are canonical genetic systems specific for the *E. coli* F pilus: MS2, the first organism to have its genome sequenced, and Q β , which provided the genome and replicase for the early studies of RNA-dependent RNA replication and in vitro evolution. Until recently only ~30 *Leviviridae* genomes had been deposited in the NCBI database, more than half of which are F-specific phages resembling MS2 or Q β in genome organization. However, orders of magnitude more novel ssRNA phages have been identified in meta-transcriptomes, which may infect more diverse groups of hosts and represent a vast “dark matter” that is under-characterized.

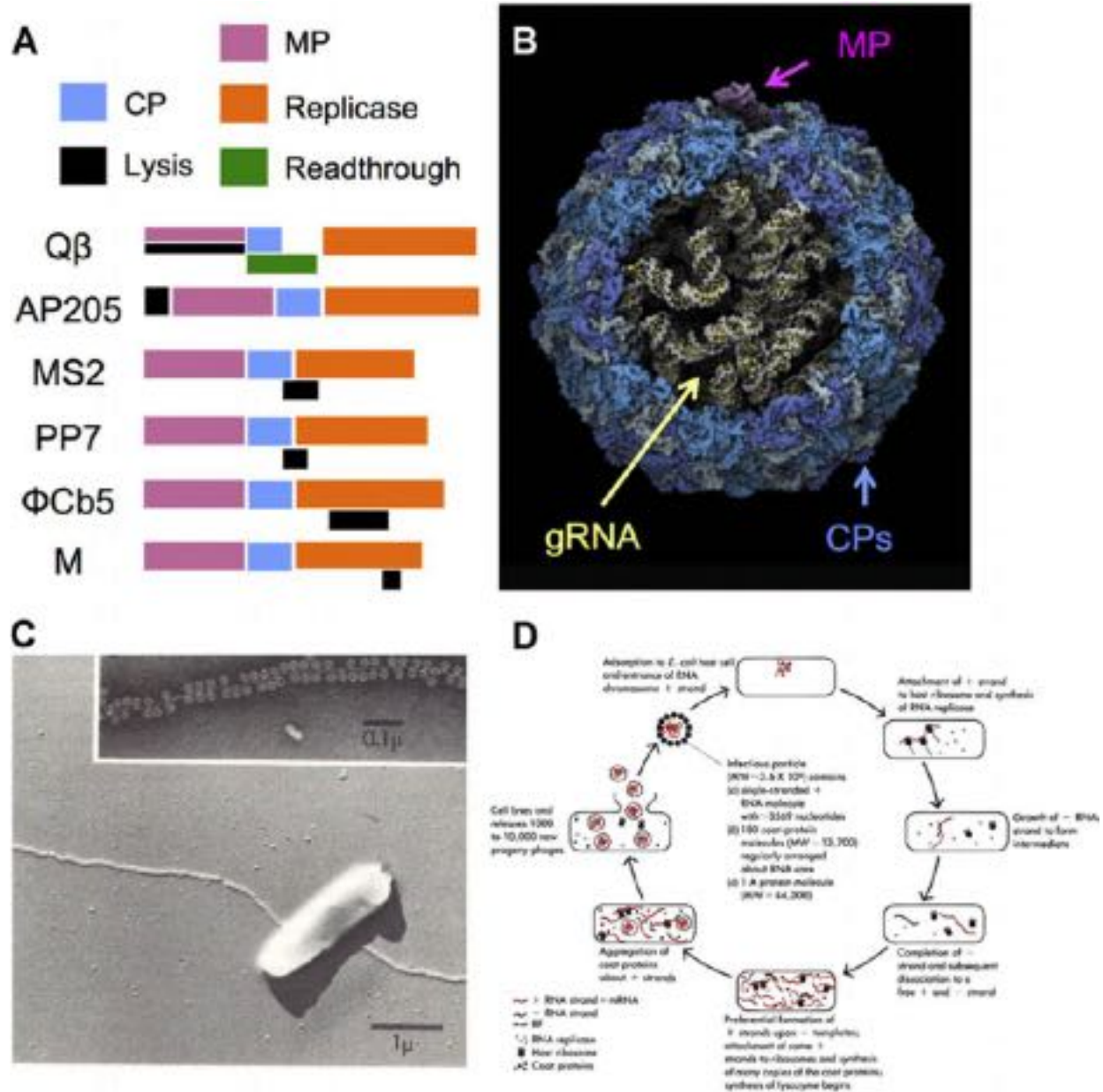


Fig. 1 Genetic maps and structural organization of ssRNA phages. (A) Color-coded genomic maps of the ssRNA phages Q β , AP205, MS2, PP7, ϕ Cb5, M, from top to bottom. The gene for the lysis protein (black) is the most variable among different ssRNA phages. Note that for Q β , the lysis protein also acts as the maturation protein; and it has an additional minor coat protein due to the read-through of the gene for the major coat protein (Read-through: green). (B) Cross-sectional view of an ssRNA phage. The front part of the capsid has been removed to show the interior of the virion and the various components are labeled. (C) Shows an electron micrograph of an *E. coli* cell with two pili, each of which is decorated with MS2 phage, i.e., the initial complex with the host receptor, see inset for details. The magnification is $\sim 16,000\times$ and reveals the relative sizes of the host and its virus particles. Pili occasionally contract toward their host and it is believed that during this process the bound phage get internalized but for MS2 only the MP-RNA complex gets into the cell. Once there a protease cleaves the MP allowing the RNA to act as an mRNA for production of phage proteins. (D) Life-cycle post genome entry including translation to produce phage proteins, genomic RNA replication, self-assembly of progeny phage particles and subsequent lysis of the host. Reproduced from Paranchych, W., 1975. Attachment, ejection and penetration stages of the RNA phage infectious process. In: Zinder, N.D. (Ed.), RNA Phages. Cold Spring Harbor, New York: Cold Spring Harbor Press, pp. 85–112. with permission from Cold Spring Harbor Laboratory Press (NY). Watson, J.D., Hopkins, N.H., Roberts, J.W., Steltz, J.A., Winer, A.M., 1993. Molecular Biology of the Gene : General Principle, vol. I. fifth ed. New Delhi: Pearson Education, Inc., with permission.

Phage lifecycle

Infectious phage must contain at least one copy of the MP as well as a CP shell to protect the genome (Fig. 1). Infection is initiated by the interaction of MP with the edge of a bacterial pilus (Fig. 1C), the first observation that this component is both structural and at least partially exposed to the surface. Loss of MP results in particles that are no longer infectious and renders part of the

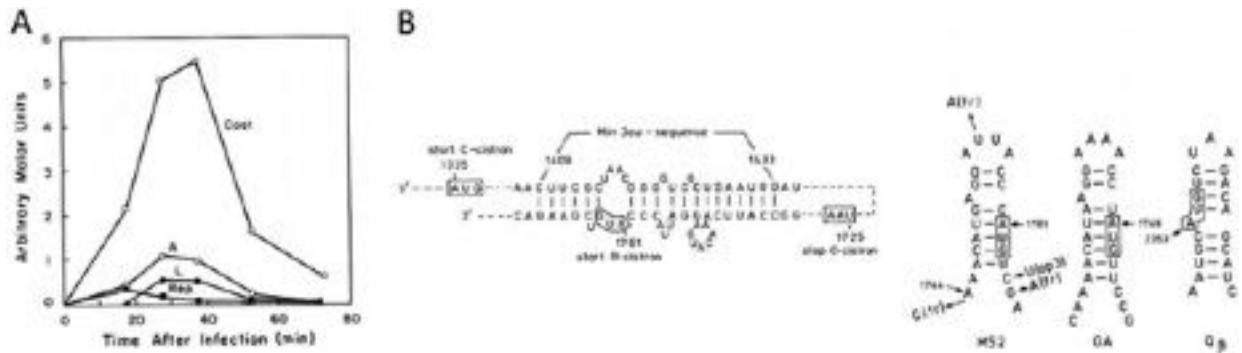


Fig. 2 The Regulation of Phage Gene Expression. (A) The amounts of each of the encoded proteins produced during an infection cycle. Although each gene is present in equal amounts along each copy of every genome there is a dramatic difference in the amounts of each protein in the cell, reflecting the requirements for different numbers of these protein products for replication, self-assembly of progeny phage and lysis of the host. The implication is that the expression of these protein products is tightly controlled. Editor: use images from the 3rd edition (B) RNA-RNA and Translational Repression - At times in the phage lifecycle the stem-loop structures (right) encompassing the start codon for the phage encoded replicase subunit are sequestered by long range Watson-Crick base-pairing interactions to a segment of the CP gene, the Min Jou sequence (left). Translation of the CP gene releases this sequence allowing the expression of replicase. However, as the CP concentration rises it binds sequence-specifically to a cognate stem-loop sequestering the replicase start codon leading to translational repression. Editor: use images from the 3rd edition. Reproduced from Van Duin, J., 1988. Single-stranded RNA bacteriophages. In: Calendar, R., (Ed.), *The Bacteriophages*, vol. 1. New York & London: Plenum Press, pp. 117–167 with permission.

genome RNase accessible, implying a second function namely MP-RNA interaction. Careful work by Paranchych and others shows that infection occurs when just the MP-gRNA complex enters the cell, the CP subunits of the capsid being left outside. The precise mechanism of how the MP-RNA complex is internalized has not been worked out in any detail. Once inside however the MP undergoes a single proteolytic cleavage event which initiates the phage lifecycle in the cell (Fig. 1(D)).

Expression of phage proteins is tightly controlled, both in the amounts and timing of the appearance of each protein product, reflecting the need for differing amounts of phage proteins, despite their genes all being present equally on the same RNA (Fig. 2). An essential early gene product is the RNA-dependent RNA polymerase. This contains only one phage-encoded subunit, from the 3' replicase gene, which forms an active complex with host proteins S1 and elongation factors Tu and Ts. This enzyme then copies the entire positive-sense strand into its negative-sense complement, and then repeats the process on the negative-strand to create progeny genomes. The complementary strands are only base-paired in the region being copied preventing formation of dsRNA. These events require the enzyme to recognize distinct signals in the 3' UTRs of each strand. The positive-sense progeny produced serve as mRNAs for the production of phage proteins and subsequently get repurposed as assembly substrates during the formation of progeny phage. The replication enzyme from Q β was used to establish *in vitro* RNA transcription. Subsequent evolution studies identified a non-viral replicator that outcompeted the genome as a replication substrate revealing the compromises required to encompass the multiple functions needed to create an infectious particle.

CP expression occurs from the start of infection and its level increases throughout an infection, reflecting the need to assemble large numbers (> 1000) progeny phage. There seems little regulation of this expression other than the sequence and stability of the initiator hairpin structure that binds the ribosome. Upstream flanking sequences may also positively promote translation, perhaps by forming structures that favor binding of the ribosome to this site. The phage makes more MP than is strictly necessary but its level is tightly controlled. In part this regulation is due to poor translational initiation, the start codon in MS2 being GUG not AUG. Additionally, the ribosome binding site of the MP gene may base pair with an internal region of its gene preventing translation. This creates a window for expression in which as the RNA is replicated the MP start site is free to bind ribosomes but these must compete with internal base-pairing once the full MP gene sequence is present.

The replicase is made at a level of ~ 5 copies/genome and its levels are also tightly controlled. This regulation is achieved in two distinct ways. As with the MP gene the ribosome replicase start site can be sequestered in another long-range base pairing interaction, this time with a section of the CP gene. Translation of the CP gene sequesters this complementary sequence, known as the Min Jou sequence after its discoverer, freeing the region around the start of the replicase gene so that it can fold to form a ribosome binding site (Fig. 2(B)).

However replicase expression is further controlled via a translational repression complex that forms with the phage CP. A stem-loop translational operator hairpin, TR (Fig. 2(B)) can form that is bound sequence-specifically by a CP dimer in a concentration-dependent fashion. This operator encompasses the start codon for replicase and formation of this complex effectively switches off replicase translation. The complex has become a paradigm for studies of sequence-specific RNA-protein recognition. It also appears to be the site of assembly initiation of the phage CP shell around the genome (see below). Regulating phage gene expression is vital because of the importance of timing the switch in the viral life-cycle from replication to assembly of progeny phage. During that process the viral mRNAs are repurposed as assembly substrates.

Lysis – cell death – is the ultimate result of the infection ensuring release of progeny phage. In MS2, this is achieved by the action of a phage-encoded lysis protein expressed inefficiently from the +1 reading-frame within the CP gene (Fig. 1). The lysis

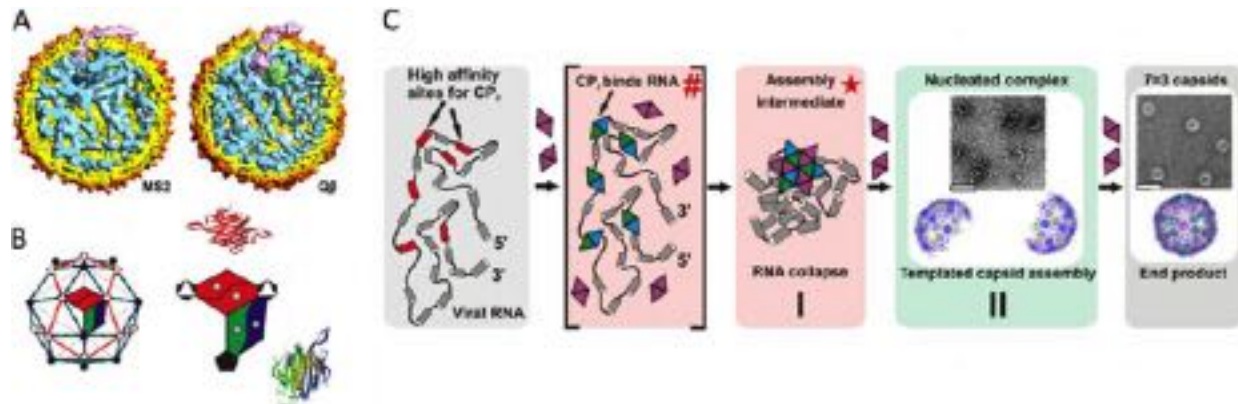


Fig. 3 Structure & Assembly of RNA Phages. (A) Asymmetric cryo-EM reconstructions of MS2 and Q β shown with the front half of their CP shells removed to reveal the density for the genome (light blue) and the MP (purple). The encapsidated CP dimer in Q β is shown in green. (B) CP quasi-conformations required to assemble the $T = 3$ capsids. CP dimers are either symmetric (C/C conformation) or asymmetric (A/B conformations). These different conformers interdigitate at particle 3-fold vertices (Δ), but 5-folds ($\blacklozenge$) are surrounded only by B-type conformers. (C) Proposed model for assembly. Multiple TR-like stem-loops across the MS2 genome (red cylinders) trigger CP dimer conformational change in favor of the A/B conformer. Multiple such cognate RNA-CP interactions collectively make assembly highly efficient, initially collapsing the CP-free RNA conformation and promoting the assembly pathway illustrated.

protein is the last gene product to begin expression, consistent with its functional role in the life-cycle. The delayed production of this gene product appears again to be a consequence of translational control, the ribosome binding site allowing initiation of translation being inaccessible until the ribosome translating the CP gene terminates at a defined site. Experiments where the CP gene stop codons are moved further 3' prevent lysis expression implying that correctly regulated expression is due to a defined RNA folding event similar to those controlling replicase translational repression. Variations of the details of this mechanism may partially account for the frequency of read-through products in some of the CPs in some phages.

Phage Structure and Assembly

Simple spherical viruses like the RNA phages were important at the beginning of the study of molecular biology since they posed the important question of how a gene product could encapsidate the RNA that encoded it. The solution of course is that the RNA genome is used multiple times as an mRNA producing many small CP subunits that subsequently associate to form a closed container into which the nucleic acid co-assembled. The simplest form of such a container would be highly symmetrical, allowing each CP to make identical intermolecular contacts with its neighbors. The container with the highest symmetry is an icosahedron, implying that viral shells should contain 60 CP subunits. Caspar & Klug extended this idea to include higher CP stoichiometries that are multiples of 60 arguing that CPs can occupy quasi-equivalent symmetry-related positions in the capsid. This concept has been central to virology for over 50 years. It solves the problem of how viruses can encode a container large enough for their genetic material. Very many simple viral capsids correspond to this arrangement in which icosahedral facets are sub-triangulated by multiple CP subunits, whose numbers and arrangements can be predicted to derive a Triangulation Number (see Glossary) for a viral particle.

The earliest viral structures, starting with Harrison's pioneering structure of the $T = 3$ Tomato Bushy Stunt Virus, seemed to confirm this concept. They used symmetry-averaging to calculate atomic resolution electron density maps. However, whilst such structures produce high resolution maps of viral CP shells major non-symmetrical components such as the viral genome and any minor structural proteins are not seen. Indeed flexible sections of the CPs, which are often involved in contacts to the genome, are also lost. This anomaly is being addressed by using modern cryo-electron microscopy (cryo-EM) which unlike crystallography can calculate atomic resolution reconstructions of virions without symmetry averaging, resulting in electron density maps that reveal the previously unseen asymmetric components.

Asymmetric cryo-EM structures at close to atomic resolution are now available for both MS2 and Q β (Fig. 3). They reveal shells of CP dimers that roughly occupy the lattice positions expected for the A/B and C/C dimers needed to create an idealized $T = 3$ particle (see above). However, this lattice is interrupted by the presence of a single copy of the MP, which replaces a CP dimer at a C/C position. Also visible is $\sim 80\%$ of the electron density for the encapsidated genomes. Instead of an esthetically pleasing icosahedrally-symmetric protein shell of CPs we can now see the viral molecular machinery poised for host attachment and infection.

Given that the encapsidated genome appears well-ordered a reasonable assumption is that it consists of only a single conformer. Protein-free RNA molecules of that size (3569 nt for MS2) would be expected to sample multiple conformational states in an ensemble of structures. Thus the structures of these particles begs an important question of how such a quasi-symmetrical structure can be assembled around a highly asymmetric genome, incorporating the MP as it occurs. This problem has been

addressed using a variety of biophysical techniques. Mass spectrometry and NMR spectroscopy showed that TR operator-CP dimer interaction biases the conformational preference of the CP dimer towards the A/B state. The sequence-specifically bound RNA acts as an allosteric effector for this change. Sixty A/B dimers are required to make the $T = 3$ capsid of the virion but there is only one copy of the TR sequence within the genome. Single molecule fluorescence measurements of the hydrodynamic radius of dye-labeled genomes in assembly reactions, however, have revealed that the RNA must be compacted during encapsidation. This compaction is driven by multiple CP dimers binding to sites across the genome, consistent with experiments triggering assembly *in vitro* with stem-loops other than TR. The multiple genomic sites being bound have been termed Packaging Signals (PSs), and assembly using such sites is PS-mediated assembly. They act to make the assembly process highly efficient in the biological environment restricting the molecular components from exploring all possible assembly pathways.

The asymmetric structure of MS2 confirms these predictions about its assembly mechanism. There are > 50 stem-loops that point at CP dimers and could act as PSs, and 15 of these remain bound by their CP partners in the virion. Not all the PS sites need to stay bound following assembly, since they are only required to act as allosteric effectors during the assembly process itself. Whilst this mechanism accounts for assembly of a symmetric $T = 3$ particle it does not explain how the unique MP gets incorporated, creating the biologically functional asymmetric protein shell. Cryo-EM and earlier biochemistry reveal that MP makes a specific genome contact in an additional stem-loop adjacent to the 3' end of the genome. This may occur before or shortly after translational repression of the replicase cistron by binding of a CP dimer at TR (see above). A fascinating insight into the assembly mechanism, and potentially virion evolution, comes from the fact that in the infectious phage particle the CP dimer at TR, which includes the start codon of the replicase is in contact with the MP bound to the untranslated region at the end of that gene. Thus both structural proteins act to sequester the replicase gene, preventing the gene from being translated and any existing replicase beginning to make the minus-strand. The importance of this arrangement is highlighted by a variation seen in the structure of Q β , where instead of replacing a CP dimer it appears that the MP protein may have forced the CP dimer to become internalized within the phage protein shell.

Biotechnological Applications

RNA phage biology has been exploited in many differing areas of biotechnology. The sequence-specific interaction between the phage CP dimer and its RNA translational operator has been used to monitor RNA-protein interactions and RNA/RNA-binding protein localization in cells using fluorescently modified components/antibodies. In addition the ease by which virus-like particles can be assembled has led to their use to create bespoke VLPs carrying artificial cargoes, such as toxins or siRNAs that can then be used therapeutically. Bacteriophage MS2 CP has become the first viral protein to have every one of its natural amino acids mutated creating a detailed structure-function map. It is therefore very likely that this list of applications of this system will continue to expand.

Further Reading

- Borodavka, A., Tuma, R., Stockley, P.G., 2012. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *PNAS* 109, 15769–15774.
- Caspar, D.L.D., Klug, A., 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposium Quantitative Biology* 27, 1–24.
- Cui, Z., Gorzelnik, K.V., Chang, J., *et al.*, 2017. Structures of Q β virions, virus-like particles, and the Q β –MurA complex reveal internal coat proteins and the mechanism of host lysis. *PNAS* 114, 11697–11702.
- Dai, X., Li, Z., Lai, M., *et al.*, 2017. *In situ* structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* 541, 112–116.
- Gorzelnik, K.V., Cui, Z., Reed, C.A., *et al.*, 2016. Asymmetric cryo-EM structure of the canonical Allevivirus Q β reveals a single maturation protein and the genomic ssRNA *in situ*. *PNAS* 113, 11519–11524.
- Meng, R., Jiang, M., Cui, Z., *et al.*, 2019. Structural basis for the adsorption of a single-stranded RNA bacteriophage. *Nature Communications* 10, 3130.
- Paranchych, W., 1975. Attachment, ejection and penetration stages of the RNA phage infectious process. In: Zinder, N.D. (Ed.), *RNA Phages*. Cold Spring Harbor, New York: Cold Spring Harbor Press, pp. 85–112.
- Twarock, R., Stockley, P.G., 2019. RNA-mediated virus assembly: Mechanisms and consequences for viral evolution and therapy. *Annual Review of Biophysics* 48, 495–514.
- Valegård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J., Liljas, L., 1994. Crystal structure of a bacteriophage-RNA coat protein-operator complex. *Nature* 371, 623–626.
- Witherell, G.W., Gott, J.M., Uhlenbeck, O.C., 1991. Specific interaction between RNA phage coat proteins and RNA. *Progress in Nucleic Acid Research and Molecular Biology* 40, 185–220.

Enveloped Icosahedral Phages – Double-Stranded RNA ($\phi 6$)

Paul Gottlieb and Aleksandra Alimova, The City University of New York (CUNY), School of Medicine, The City College of New York, New York, NY, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ATP Adenosine triphosphate
Cryo-EM Cryo electron microscopy
Da Dalton, 1.67×10^{-24} g
dsRNA Double stranded RNA
LPS Lipopolysaccharide
mRNA Messenger RNA
NC Nucleocapsid
NMR Nuclear magnetic resonance
NTP Nucleoside triphosphate
NTPase Nucleoside tri-phosphatase

OM Outer membrane
PC Procapsid
RdRP RNA-dependent RNA polymerase
RLPS Rough lipopolysaccharide
S_{20,w} Sedimentation coefficient, standardized to conditions corresponding to pure water at 20°C and extrapolated to zero protein concentration, 10^{-13} s, and it's expressed in Svedberg value S
ssRNA Single stranded RNA
T Triangulation number

Glossary

Capsid The protein shell of the virus containing the nucleic acid genome.
Carrier state The complete or incomplete viral genome maintained in the host cell as a stable episome without cell lysis.
Genomic precursor Viral mRNA that is subsequently packaged in the capsid.
Genomic segment One of the 3 double stranded RNA molecules in the virus core.
Nucleocapsid Expanded, RNA filled capsid, covered by the outer shell matrix and capable of transcription.
Packaging Insertion of the mRNA genome precursors into the virus core

Polymerase complex The viral protein complex that contains the RNA polymerase and the dsRNA genome segments. The structure is capable of replication and transcription of the viral genome.
Procapsid The viral capsid prior to acquisition of the genome precursors.
Reverse genetics Rescue of an RNA virus genome from cDNA recombinant clones.
Self-assembly Mechanism with which protein precursors organize into a defined structure utilizing local interactions and without the intervention of external direction.

Introduction

The *Cystoviridae* Family

The cystoviruses (family *Cystoviridae*) are a unique group of viruses that have proven of great utility in the study of many facets of molecular virology. First discovered in 1973 as a bacteriophage of phytopathogenic pseudomonads, $\phi 6$ proved unique in that it contained lipid and possessed a genome consisting of three strands of double-stranded RNA (dsRNA). The replicative mechanism and structure make cystoviruses analogous to members of the *Reoviridae* family, therefore much of the research interest was driven to shed light on analogous systems in the pathological reoviruses. Until 1999, $\phi 6$ had been the only member identified in the genus and the extensive investigation of its genomic sequence, replication cycle, structure, and RNA packaging system formed the foundation of this field. Additional viruses have since been isolated that share the physical and structural characteristics of $\phi 6$. All are chloroform-sensitive, indicating they contain lipids, derived from the host cell. Their genome consists of three dsRNA segments of different sizes denoted 'S', 'M' and 'L' for 'small', 'medium', and 'large', respectively. (Transcripts from the three genome segments are designated in lower case; *i.e.*, 's', 'm', and 'l', respectively). The overall viral organization is shown schematically in Fig. 1. All the cystoviruses are virulent and infect strains of the plant pathogen, *Pseudomonas phaseolicola*. Based on genetic sequence and host range analysis, the well-classified cystovirus bacteriophages were divided into two major groups in which $\phi 7$, $\phi 9$, $\phi 10$, and $\phi 11$ are closely related to $\phi 6$, while $\phi 8$, $\phi 12$, and $\phi 13$ are distantly related to $\phi 6$. Additional cystovirus family members have subsequently been isolated from environmental samples throughout the United States, Finland and China indicating the genus is extremely widespread. After the synthesis and isolation of a recombinant $\phi 6$ procapsid (PC) in 1990, seminal studies of the structure, replication, genome packaging, and transcriptional regulation of $\phi 6$ were performed. This research established the first *in vitro* RNA packaging and replication assay for this class of viruses and subsequently utilized it to rescue infectious particles.

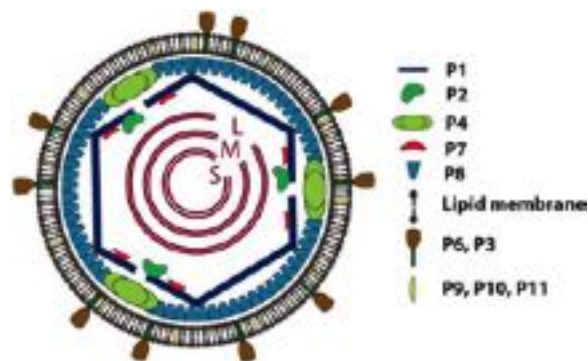


Fig. 1 Schematic organization of layered structure of the $\phi 6$ virion. The inner layer, procapsid PC, includes the shell, composed of 120 copies of P1 protein or 60 non-symmetrical dimers of P1. Inside the shell, are the P2 and P7 proteins, which are localized at the 5-fold axis of symmetry portal and three ds-RNA segments, which have low symmetry quasi-concentric shell organization. The P4 proteins form a hexameric ring around the 5-fold axis; the NC includes the PC and protein P8 matrix layer, and the complete virion has envelope from the cellular bi-lipid membrane with randomly distributed P3, P6, P9, P10 and P11 proteins.

Classification and Host Range

$\phi 6$ was originally isolated from a bean straw plant that was infested with *Pseudomonas syringae* pv. *Phaseolicola*, strain HB10Y and until 1999 was the lone type in the family *Cystoviridae*. The isolation of eight additional lipid-containing, three-segmented dsRNA bacteriophages expanded the group with additional types some of which were closely related and others distantly related to $\phi 6$. The host range of cystoviruses is governed by the structure of the attachment apparatus and which cell surface receptor it utilizes. Of the $\phi 6$ and its close relatives, host infection requires binding to type IV pili. Therefore the $\phi 6$ group has a host range limited to *Pseudomonas syringae* (and for several laboratory selected $\phi 6$ mutants *Pseudomonas pseudoalcaligenes*). In species $\phi 8$, $\phi 12$, and $\phi 13$, the attachment apparatus is a heteromeric protein assembly that utilizes the rough lipopolysaccharide (RLPS) as a receptor. These types also have the ability to infect other gram-negative bacteria containing an RLPS – although without plaque formation.

In 2010 the *Cystoviridae* family was again expanded with the isolation of $\phi 2954$ from radish leaves. The base sequences for many of the genes and for the segment termini were similar but not identical to those of bacteriophage $\phi 12$. However, the host specificity was for the type IV pili of *Pseudomonas syringae* HB10Y rather than for the RLPS to which $\phi 12$ attaches. Next ϕNN was isolated from a freshwater habitat demonstrating that the virus group is quite diverse and widely distributed. This phage utilizes the type IV pilus of *Pseudomonas syringae* for attachment and phylogenetic trees of the genome segments indicate a close relationship to $\phi 6$. Of great interest was the 2016 isolation in China of ϕYY from hospital sewage, a cystovirus that uses the opportunistic pathogen *Pseudomonas aeruginosa* strain, PAO38, as a host.

Virion Structure and Properties

Physico-Chemical Properties

The virus particle molecular weight is approximately 9.9×10^7 Da, sedimentation coefficient $S_{20,w}$ approximately 405S (or 4.05×10^{-11} s) and density 1.24 g cm^{-3} in sucrose. The molecular weight of the nucleocapsid (NC) is approximately 4.0×10^7 Da. The composition is 70% (w/w) protein, 14% dsRNA, and 16% phospholipids.

Architecture of the Polymerase Complex and the Nucleocapsid

All the cystovirus members are enveloped bacteriophages of multilayered structure (Fig. 1). The first species isolated, $\phi 6$, is the best characterized in regard to its structure and assembly mechanisms. The inner layer, the PC, is composed of 60 asymmetric dimers of protein P1 that form a dodecahedral $T = 1$ (triangulation number) shell. At each of the twelve dodecahedron faces protein P4 hexamers (the packaging nucleoside triphosphatase, NTPase) localized at the 5-fold axis, protrude from the PC. The P2 and P7 proteins localized at the portal at the 5-fold axis of symmetry are within the P1 shell. The $\phi 6$ NC outer shell has icosahedral symmetry and is formed of 200 P8 trimers, a small membrane-associated protein, in a $T = 13$ matrix that partially covers the PC. The NC is enveloped by a lipid membrane with the lipids derived from the host cell cytoplasmic membrane and the viral envelope surface proteins which have an asymmetric distribution. The individual species of the cystoviruses are not entirely isomorphic but similar in overall design. The NC structures have similarity in P1, P2, P4 and, P7 organization but the P8 layer is quite different among the species. In $\phi 8$ the NC cover is virtually non-existent consisting of only 60 copies of P8, a small membrane associated protein. In contrast to $\phi 6$ and $\phi 8$, NC architecture based on analysis of cryo-electron microscope single particle reconstruction analysis placed $\phi 12$ in an intermediate category showing an incomplete $T = 13$, the P8 organization with significant vacancies occupied by protein P7 dimers.

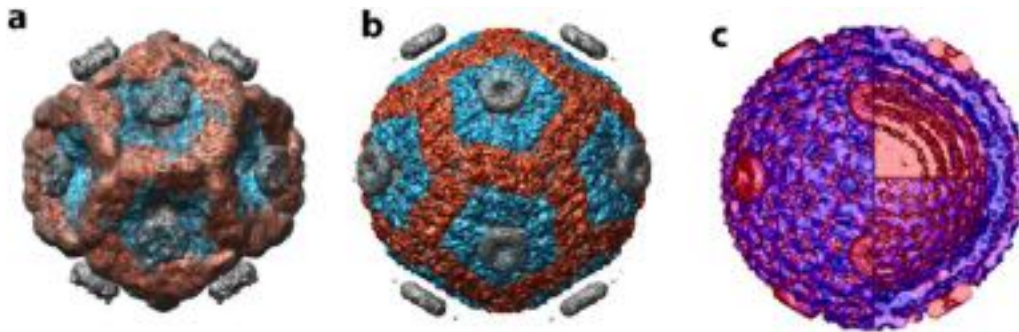


Fig. 2 The cystovirus PC 3D structure before (unexpanded) RNA packaging and after (expanded) RNA packaging, replication and the structure of the mature NC. Data reproduced from Huiskonen, J.T., de Haas, F., Bubeck, D., *et al.*, 2006. Structure of the bacteriophage $\phi 6$ nucleocapsid suggests a mechanism for sequential RNA packaging. Structure 14, 1039–1048.

Cystoviruses package their plus sense messenger RNA (mRNA) sequentially into a preformed PC in the order s, m and l and the PC expands to accommodate the genome (Fig. 2). In $\phi 6$, packaging relies upon signals at the RNA 5' end called "pac" containing about 200 nucleotides (with extensive secondary structure) while specific sequences at the 3' end are RNA replication signals. The PC expansion is facilitated by hinge like movement of the P1 subunits approximately parallel to the two fold edges of the dodecahedral PC. The expansion and packaging activates the P2 RNA-dependent RNA polymerase (RdRP) which replicates the mRNA to dsRNA. The expanded genome-filled PC is finally enclosed within the 200 P8 trimer matrix completing the assembly of the mature NC. The packaged dsRNA genome segments appear to be located in lower symmetry quasi-concentric shells based upon electron cryo-microscopy reconstructions.

The Protein Components of the Cystovirus Portal Complex and Nucleocapsid

The cystovirus RNA portal complexes, consisting of protein elements P2, P4, and P7, are located on each of the apex of the assembled PC. There is a potential total of twelve RNA portal complexes upon the dodecahedral PC however not all the sites are occupied. The dynamic interaction of these three portal proteins governs the packaging, replication, and transcription of the viral RNA genome. This intricate molecular machine can be described in regard to each of the separate proteins and below is reviewed what is currently understood about each individual component.

P1

The PC organizing protein is P1 and it forms the framework of the entire complex. P1 interacts with the other the three portal proteins to properly position them within PC structure. The PC is built from 120 identical P1 proteins organized into a $T = 1$ shell containing 60 non-symmetric dimers where each dimer is composed of subunits A and B of the same 3-dimensional fold. The P1 crystal structure has been determined for viruses $\phi 6$ and $\phi 8$ and is a trapezoidal shape that accommodates the significant conformational changes that occur during RNA packaging (Fig. 3(A)). Intriguingly different conclusions describing the evolutionary history of the $\phi 6$ and $\phi 8$ P1's were reached by the respective research groups. The $\phi 6$ P1 was reported to show no P1 fold similarity with the known folds noted in other dsRNA viruses. Therefore the $\phi 6$ P1 structure analysis led to the suggestion that cystoviruses derived from a different lineage from other dsRNA viruses. The second interpretation that was derived from the $\phi 8$ P1 structural determination reasoned that the protein was indeed similar to the corresponding reovirus coat proteins. This conclusion was reached notwithstanding the fact that reoviruses do not undergo particle expansion when the segmented dsRNA genome is packaged, as their coat proteins are already highly elongated.

P2

The P2 of cystoviruses exhibits RNA-dependent RNA polymerase (RdRP) activity catalyzing *de novo* initiation and elongation of replication and producing double-stranded RNA from single-stranded RNA templates. The crystal structure of the $\phi 6$ RdRP revealed that it is highly similar to that of Hepatitis C Virus (HCV) suggesting an evolutionary link even between double-stranded RNA viruses and flaviviruses. The basic architecture demonstrated that the $\phi 6$ polymerase resembles the hand-like features model in that the 664 amino acid residues of the monomeric protein form 25 α -helices and 21 β -strands whose 3-dimensional geometry is reminiscent of fingers, palm, and thumb in a human hand (Fig. 3(B)). The key aspartate residues characteristic of RNA polymerases are positioned analogous to that of the HCV polymerase. Co-crystallization of the $\phi 6$ RdRP with oligonucleotides resembling the 3' ends of the viral RNA templates showed that they bind within a tunnel lined with mostly basic amino acids that leads to an active site. A binding site in the substrate pore orders triphosphate moieties by their attachment to key basic residues. The observations suggested a mechanism whereby the incoming

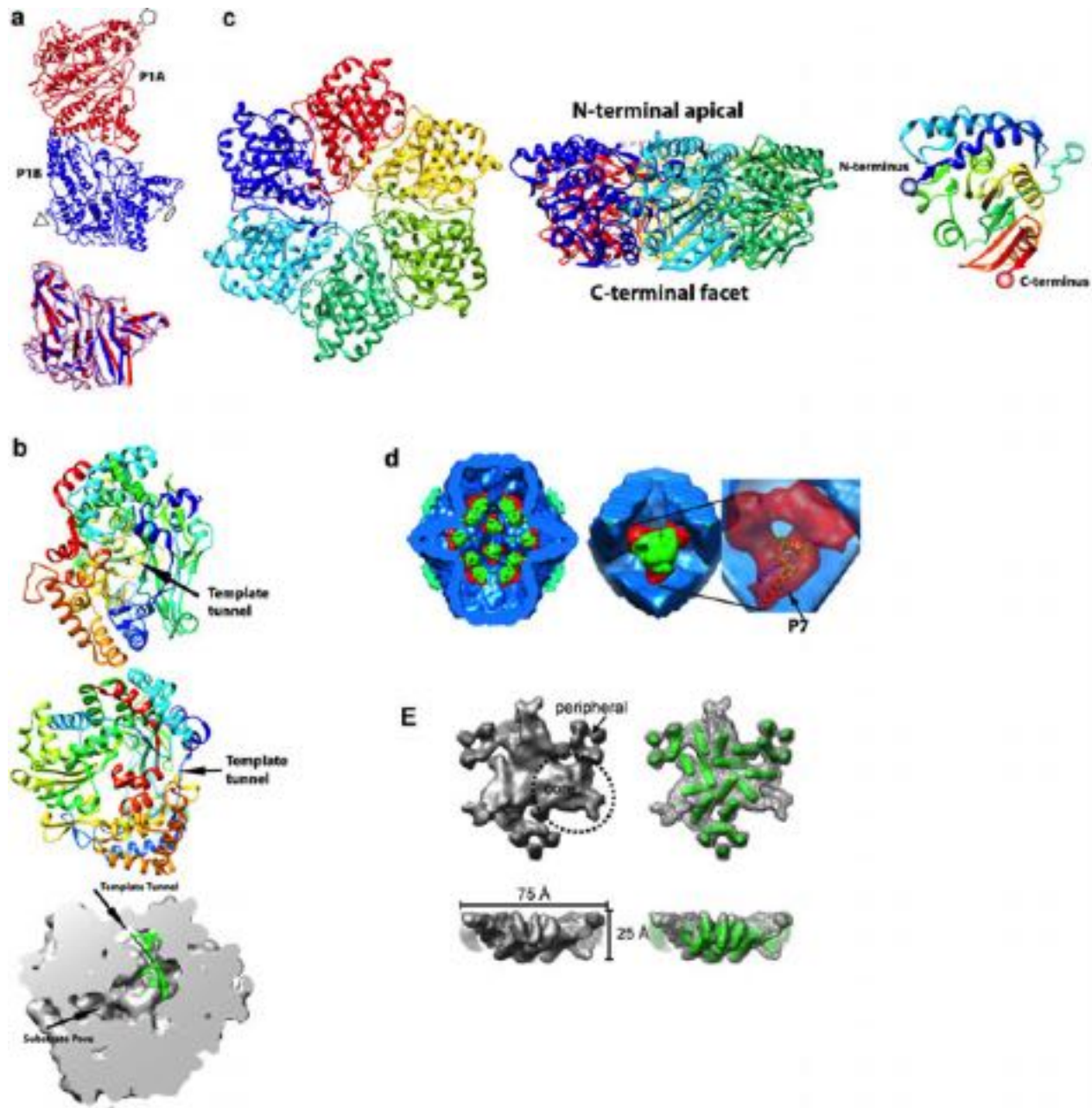


Fig. 3 Structure of individual proteins and the assembly of $\phi 6$. The structure reconstructions are derived from cryo-electron microscopy and x-ray crystallography. (A) The $\phi 6$ P1 protein organized showing non-symmetrical dimers, P1A (red) and P1b (blue) (PDB #). The conformational mismatch of superimposed P1A and P1B is shown on lower part of the image. Symmetry mismatch of trapezoidal isomers allow for significant conformational change during packaging. (B) The $\phi 6$ P2 protein (PDB #1HHT, data were provided by Butcher, S.J., Grimes, J.M., Makeyev, E.V., *et al.*, 2001. A mechanism for initiating RNA-dependent RNA polymerization. *Nature* 410, 235). The arrow indicates the RNA template tunnel entrance. The cut-through rendering of the 3D structure shows the internal organization of RNA template tunnel and substrate pore. (C) The $\phi 6$ P4 protein (PDB #5MUV, data provided Sun, Z., El Omari, K., Sun, X., *et al.*, 2017. Double-stranded RNA virus outer shell assembly by bona fide domain-swapping. *Nature Communication* 8, 14814–14814). A hexameric toroid built of six identical subunits is localized around 5-fold axis of symmetry, which creates a symmetry mismatch. The top and side views are shown. The monomer of the $\phi 6$ P4 is shown with the N- and C-terminus indicated. (D) The $\phi 6$ P7 protein localization within PC based on difference map between PCs mutants lacking different structural proteins. The central cut of the PC complex shows the localization P2-P7 complexes at 5-fold axis proximity. P2 protein shown in green, P7 – in red. P2 density is blue and P4 toroid are in cyan. Magnified section showing P2 surrounded by three P7 monomers and an insert showing docked atomic model of P7. Data reproduced from Katz, G., Wei, H., Alimova, A., *et al.*, 2012. Protein P7 of the cystovirus $\phi 6$ is located at the three-fold axis of the unexpanded procapsid. *PLoS ONE* 7 (10). (E) The average $\phi 6$ P8 trimers, rendered as solid isosurface or mesh. Green rods are α -helices. The top and side views are shown. Data reproduced from Huiskonen, J.T., de Haas, F., Bubeck, D., *et al.*, 2006. Structure of the bacteriophage $\phi 6$ nucleocapsid suggests a mechanism for sequential RNA packaging. *Structure* 14, 1039–1048.

double-stranded RNA is unwound and fed through a template site to an active site, while nucleotides enter through a substrate pore. P2 is found closer to the 3-fold axis of symmetry in the unfilled PC. Cryo-EM data strongly suggests that the RdRP moves to the 5-fold axis upon RNA packaging and PC expansion.

P4

The cystoviruses package their own mRNA into the PC and the energy for this translocation activity is provided by the P4 NTPase. This protein is a hexamer, and when assembled on each of the twelve 5-fold axes on the PC, creates a symmetry mismatch (Fig. 3(C)). The initial x-ray diffraction high-resolution structure was first determined for the ϕ 12 P4 and subsequently for ϕ 6, ϕ 8, and ϕ 13. The six identical P4 subunits are arranged in a ring and resemble a domed toroid (Fig. 3(C)). Six ATP binding sites are located near the external perimeter of the ring at interfaces between adjacent subunits. The quaternary structure of the P4 toroid is similar to other hexameric helicases and the ϕ 8 and ϕ 13 examples both have 5'-3' RNA-helicase activity. The hexameric conformation contains a central region through which the mRNA being packaged is translocated. In addition there is a loop in each monomer found to be central within the hexamer structure. This loop is involved in the RNA translocation and is known to bind NTPs. Only one P4 – based portal is thought to sequentially package the three RNA segments. This could be the statistical consequence of a low occupancy of P4 on the unexpanded procapsid. The tenuous connection between P4 C-terminal and P1 is likely stabilized when all three genome segments are packaged, replicated and the P8 matrix assembles on the mature NC. Thus the fully assembled NC contains all of the twelve P4 hexamers.

P7

The P7 protein is the least characterized of the PC proteins in regard to precise function but *in vivo* studies demonstrate that it is essential for virus viability. P7 has been referred to as an accessory packaging factor but it is conceivable that this protein plays a direct role in RNA translocation and RNA packaging. The ϕ 12 P7 crystal structure was resolved at 1.83 angstroms for the N-terminal portion of the protein. P7 forms a novel α/β fold and measurement of solution NMR relaxation rates, chemical cross-linking, and static light scattering studies show that it exists in the dimeric form in solution (Fig. 3(D)). P7 has a C-terminal tail that is extremely flexible and *in vitro* assembly studies showed that incorporation of the protein stabilizes, and in ϕ 6, greatly enhances the rate of the PC assembly. As the PC is filled with the viral genome the PC undergoes expansion and as noted above the hinge of this conformational change lies along the two-fold axis, where P7 was postulated to exist as an elongated dimer at the interface between two P1 dimers. It was considered that the flexible C-terminal tail of P7 plays a role in inducing the P1 conformational change acting as the molecular hinge. Yet the position and function of the P7 packaging factor remain controversial. Cryo-EM studies of the unfilled ϕ 6 PC suggest that the P7 density overlaps the P2 density implying mutual exclusion of each protein at an inner 3-fold axis. However, difference maps generated between PC reconstructions and mutants lacking either P2 or P7 strongly suggest that P7 stabilizes P2 at the inner three-fold axis prior to RNA packaging. This implies P2 and P7 can interact and that they are located near each other in the vicinity of the inner 3-fold axis prior to packaging. Single particle reconstructions of the mature ϕ 12 NC strongly suggest that the P7 dimer is subsequently located at the portal 5-fold axis of symmetry surrounding the P4 NTPase hexamer.

P8

The addition of the P8 matrix to the genome filled PC completes the viral core maturation process and the arrangement of trimers of P8 determine the architecture of the NC. The NC matrix contains two types of holes, designated I and II with 60 of each type based on the arrangement of trimers of P8. In the ϕ 6 T = 13 lattice there are four classes of interdigitating P8 trimers termed Q, R, S, and T with Q closest to the 5-fold axis of symmetry (Fig. 3(E)). The five-fold axis of the P8 layer of the NC is occupied by the hexameric P4 packaging protein. In ϕ 12 the T = 13 lattice is considered to be incomplete in that a missing P8 density is occupied by densities corresponding to P7 dimers that surround the P4 hexamer. In ϕ 6 two states of the P8 trimer were identified as either open or closed. Shifting between the two states of the P8 trimers swap domains (or interdigitate) in order to stabilize the outer shell.

Membrane Envelope

The third, outermost, layer of the cystoviruses is a lipid bilayer. The viruses are 20% lipid by weight and constitute enough to cover approximately 50% of surface of the particle; the vacancies composed of the envelope protein. The ϕ 6 lipid bilayer contains phospholipids derived from the host plasma membrane and includes four virally encoded integral membrane proteins, P6, P9, P10, and P13. P3 forms an external spike that binds the host receptor and is anchored to the bilayer by P6. In the viruses that bind the host RLPS, the envelope also originates from the host cell membrane but the surface attachment proteins differ substantially from ϕ 6, reflecting the different host-cell attachment mechanism. Nevertheless, in all cases, P6 appears to anchor the P3 complex to the membrane and is also able to mediate membrane fusion releasing the uncovered NC into the periplasmic space. P12 is a non-structural protein that mediates membrane acquisition during assembly of the ϕ 6 viral particle and presumably does the same for other cystoviruses.

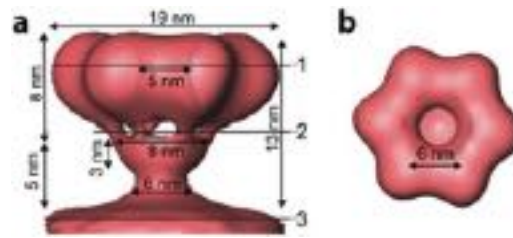


Fig. 4 The $\phi 12$ attachment P3 protein complex. The complex consist of six globular copies of P3a and three copies of P3c. Data reproduced from Leo-Macias, A., Katz, G., Wei, H., *et al.*, 2011. The toroidal surface complexes of bacteriophage $\phi 12$ are responsible for host-cell attachment. *Virology* 414 (2), 103–109.

Host Cell Attachment Complex

The $\phi 6$ spike protein, P3, extends from the viral outer envelope where it is anchored to the integral membrane protein P6. The P3-P6 complex constitutes the viral attachment apparatus that binds to type IV pili of the host cell to initiate infection and the pilus contracts to bring the virus particle to the outer cell membrane (OM) (Fig. 5). The $\phi 8$, $\phi 12$, and $\phi 13$ outer surfaces differ radically from that of $\phi 6$, $\phi 8$, $\phi 12$ and $\phi 13$ which bind to a truncated LPS O chain to initiate host cell infection. The $\phi 12$ attachment protein complex has been analysed in detail and is of a toroid shape consisting of six globular domains with six-fold symmetry (Fig. 4). Each of the globular domains is likely composed one copy of protein P3a and three copies of protein P3c. In all the species, once delivered to the host cell OM, the viral envelope fuses with that of the host using the fusogenic protein P6.

Genome Organization and Sequence Similarity

The *Cystoviridae* have three dsRNA segments that are designated according to the size as L (large, with 6374 base pairs in $\phi 6$), M (medium, with 4063 base pairs in $\phi 6$), and, S (small, with 2948 base pairs in $\phi 6$) (Fig. 6). Each virion particle contains only one copy of each segment and the genes are clustered into functional groups. Each genomic segment is transcribed into a polycistronic mRNA that is designated by a corresponding lower case letter. The genes encoding the PC are located on the L segment in the order *gene 7*, *gene 2*, *gene 4*, and *gene 1*. *Gene 14* precedes *gene 7* and the function of its protein product remains incompletely described. While in most members of the *Cystoviridae* the order of the L segment genes is 14, 7, 2, 4, and, 1; in $\phi 8$ the order is 14, H, 2, 4, 1, and 7. The H region is composed of a pair of genes, *Ha* and *Hb* in which the latter is required for temporal control of the 1 transcript (see below). In $\phi 6$ the production of the RNA polymerase P2 is down-regulated to approximately 10% the level of the other three PC proteins by translational coupling to the P7 gene, as *gene 2* lacks a ribosome binding site. The medium segment carries the genes 10, 6, 3, and, 13. *Genes 6* and 3 are in a polar relationship with each other as mutants lacking production of one suppresses the expression of the other. *Gene 3* encodes a single polypeptide in $\phi 6$, P3, whereas in $\phi 12$, *gene 3* encodes at least two proteins, P3a and P3c. The presence of two ORFs encoding P3 is also observed in $\phi 8$ and $\phi 13$, which like $\phi 12$ recognize RLPS as a host-cell receptor. This observation suggests that the oligomeric P3 complex extending from the viral surface uniquely encodes host cell specificity (see below). For each virus member, each of the three segments has a short, highly conserved sequence at the 5'-end (*i.e.*, designated from the plus sense strand of the dsRNA). The highly conserved sequence is followed by a unique packaging sequence that identifies each segment and ensures that it is packaged in the PC.

In 2018, a detailed phylogenetic analysis of each of the genomic segments for the characterized cystovirus types was conducted and the degree of relatedness established. The analysis allowed the diagramming of phylogenetic trees showing the relationships among each of the three genomic segments for the characterized virus types. Notably the phages $\phi 6$ and ϕNN were found to have significant nucleotide sequence similarity in the S and L segments and the proteins encoded are nearly identical. Greater diversity was seen in the M segment possibly reflecting the host range adaptation due to distant habitats. Overall a taxonomic description for the *Cystoviridae* was achieved and in regard to genome characteristics (such as GC content, gene organization, and gene synteny) there is a great similarity. It was proposed that based upon criteria of 95% nucleotide sequence identity for the demarcation of species in the *Cystoviridae* family, phages $\phi 8$, $\phi 12$, $\phi 13$, $\phi 2954$, ϕNN , and ϕYY can be classified as distinct species.

Replication Cycle

The entire cystovirus replication cycle is represented schematically in Fig. 5. While the host cell-viral P3 attachment mechanism (Fig. 5(A) and (B)) is either via the pili or RLPS (depending on the specific cystovirus) the remaining steps are for the most part the same in all the classified types.

Plasma Membrane Penetration and NC Entrance into the Cytoplasm

The integral membrane protein P6 facilitates the membrane fusion of the viral envelope with that of the host OM. The NC is then able to enter the periplasmic space and the peptidoglycan layer is digested by endopeptidase protein P5 that is loosely associated

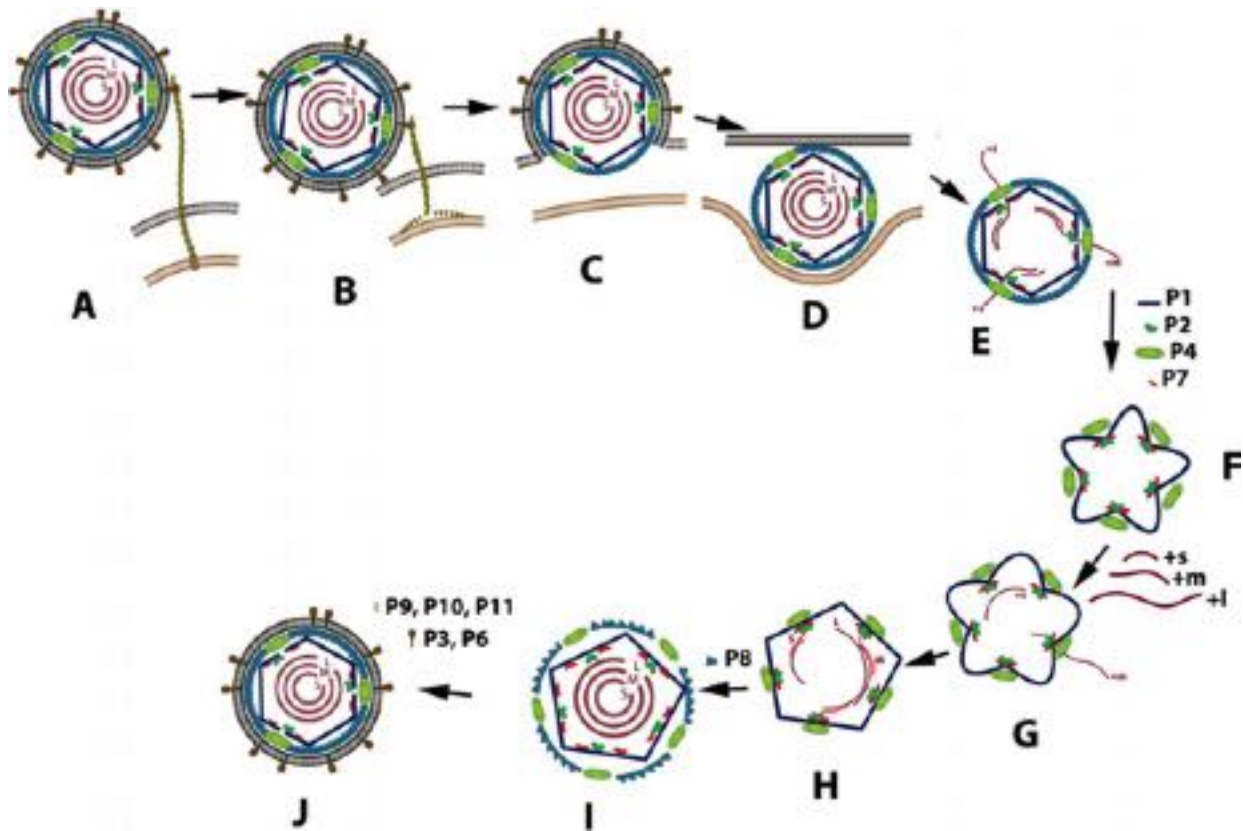


Fig. 5 The $\phi 6$ replication cycle (in details of each step are in the text). (A, B) P3 attached to the pili, followed by pili retraction; (C) viral envelope fusion with OM; (D) $\phi 6$ NC entry into intracellular vesicle; (E) viral core release into the cytoplasm, P8 disassembly and initiation of transcription; (F) procapsid self-assembly; (G) packaging initiation in the order +s, +m, +l and PC expansion; (H) + strand RNA replication to the dsRNA genome; (I) P8 matrix loosely assembled around PC; (J) cell-derived bi-lipid membrane incorporated into the virion structure, followed by cell lysis and virion release.

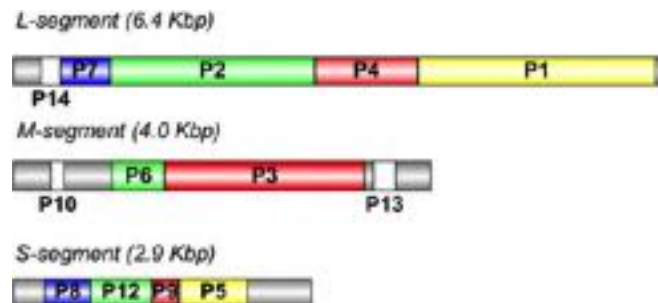


Fig. 6 The three RNA segments of $\phi 6$ and genes encoding the proteins.

on the NC (**Fig. 5(C)**). The NC penetration is mediated by the P8 lattice which interacts with the host cytoplasmic membrane resulting in containment of the NC in an intracellular vesicle (**Fig. 5(D)**). The viral core is released into the host cytoplasm and the P8 lattice disassembly initiates viral genome transcription (**Fig. 5(E)**).

Transcription

Once within the host cell cytoplasm, the NC transcribes the three dsRNA genomic segments to produce the message sense l, m, and s strands that are expelled from the particle (**Fig. 5(E)**). Transcription functions by a semiconservative strand displacement mechanism in which the nascent plus strand that is displaced replaces the original one which is released from

the NC. In $\phi 6$, transcription initially produces equimolar concentrations of all of the three segments. However translation of the l transcript is favored which allows assembly of new unfilled PC particles early in the infection. Later in the replication cycle, the l transcription level is reduced to 5%–10% of the initial concentration and translation of m and s predominates. The mechanism that regulates the temporal control of cystovirus transcription is incompletely understood but has been partially elucidated for $\phi 6$ and $\phi 8$. In $\phi 6$ the 5' 18-base sequences of the three-plus sense strands are identical with the exception that the m and s begin with GG while l begins with GU. The GU acts to signal the control the altered transcription profile of the l message RNA. Isolated of PC particles containing dsRNA and primed for transcription were found to contain a host cell-derived protein, YajQ, that influenced temporal control of transcription by increasing l production. YajQ does not directly activate the P2 polymerase but was found to bind the PC framework protein P1. Phage $\phi 2954$ has adapted to utilize host cell protein glutaredoxin 3 (GrxC) for the activation of L segment transcription and mutations that resulted in GrxC transcription independence were found in the gene for protein P1. In $\phi 2954$ (and $\phi 12$) the 5' end of the L segment begins with ACAA as an identifier while the S and M segments begin with GCAA (GAA in $\phi 12$). In $\phi 8$ the level of l transcription is independently controlled, without the need of a host cell protein. The three $\phi 8$ 5' plus sense RNA ends have short consensus sequences in that only the first seven nucleotides are identical. Yet in spite of the absence of a 5' L sequence identifier late in the infection cycle the l message transcription is considerably lower than that of the m and s messages. A gene found on the 5' end L segment (see above) encoding product Hb acts with host cell RNase R to degrade excess l transcript and bring it down to a concentration required for the late infection phase.

PC Assembly and Packaging

The $\phi 6$ replicative mechanism starts with the assembly of the empty PC (Fig. 5(F)). This is followed by the specific selection and packaging of three segments of viral mRNA within the PC. The dodecahedral-shaped PC (Fig. 5(G) and (H)) contains the RNA packaging and replication apparatus at each of twelve portals constituting polymerase complex. The polymerase complex consists of P2, P4 and, based on cryo-electron microscopy reconstruction of the $\phi 12$ virus, P7. The viral mRNA strands are packaged in a specific size order and replicated to the double-stranded format utilizing P2 RdRP only after the packaging is completed (Fig. 5(H)). The P2 RdRPs are located at each of the 12 portals and appear to undergo a position shift from the three-fold axis to the five-fold axis after PC expansion. The PC undergoes a specific and sequential expansion during RNA packaging that selectively exposes RNA binding sites for each of the three RNA segments. The noncoding sequences at the genome 3' ends direct RNA replication and are discussed below.

NC maturation

After packaging and RNA replication, a matrix composed of 200 trimeric copies of protein P8, is assembled around the filled PC (Fig. 5(I)). The P8 matrix is an open structure which can allow passage of water, cations and other smaller molecules. The PC, P8, and genome together form the NC.

Membrane acquisition and host cell lysis

A lipid-bilayer-membrane envelope, derived from the *Pseudomonas* host-cell, is assembled around the NC (Fig. 5(J)). The mechanism of viral envelope assembly is not entirely understood. It involves the non-structural P12 protein which is not present in the completely assembled virion. The envelope contains four membrane proteins, P6, P9, P10, and P13, within a lipid bilayer. Based on cryo-electron microscopy of the $\phi 12$ virus, protein P5 is likely positioned between the NC and the envelope. P3 proteins form external spikes on the envelope whose function is to bind to the host cell receptor. After the entire virus is assembled host cell lysis releases approximately 150–200 virus particles. This process utilizes two lytic proteins, the P5 murein peptidase working in conjunction with a holin lytic factor P10 that destabilizes the host envelope.

Recombination and Reassortment

RNA strand recombination in cystoviruses is due to template switching and is not overly dependent on sequence similarity at the crossover point. Recombination occurs during minus strand synthesis of a plus sense template and is dependent on the conformation on the RNA 3' end. Replication recognition signals of the plus sense RNA consist of a secondary structure composed of stem-loops near the 3' ends and removing these structures facilitates recombination. The replication of the genome segment after deletion of the RNA 3' structure can only be restored if the template is rescued by recombination.

In addition to genetic exchange by recombination reassortment of the three genome segments occur among the cystovirus species. This mechanism can occur naturally or in the laboratory setting. Investigation of the rates of reassortment in both natural and laboratory types allows analysis of viral evolution in multi-stranded RNA viruses. Experimentally genetic exchanges have been visualized as altering whole virus electrophoretic mobility demonstrating a useful selectable phenotype in studies of viral evolution and fitness.

Carrier State

Under certain conditions, a ϕ 6 infection can produce a carrier state in which the virus replicates but does not cause lysis of the host cells and the phenomenon is often seen in viruses with mutant viral lytic factors. In the carrier state, the phage does not need to reinfect cells and can dispense with select genes. Essentially the viral genome replicates as an episomal element in the host cell cytoplasm. The carrier state can be extremely stable as when drug resistance selection is maintained. For example, a kanamycin resistance gene incorporated into the M segment- with host cell selection for the antibiotic- allowed the loss of the S segment. This carrier mutant clearly had altered the packaging order by dispensing with s that is normally the first viral mRNA packaged. These mutants had altered the P1 protein suggesting that RNA interaction and packaging order is dependent on the PC framework protein.

Reverse Genetics and Self Assembly

The advent of reverse genetics methods for both ϕ 6 and ϕ 8 represented the first example of the production of recombinant genetic methods for segmented dsRNA viruses. Initially, the ϕ 6 rescue system required the packaging of *in vitro* synthesized transcripts into a PC assembled in *E. coli* to which was added the P8 matrix. This particle was then transfected into *pseudomonad* spheroplasts and viable recombinant progeny recovered. This system demonstrated the feasibility of producing viable engineered segmented dsRNA virus particles, however, the *in vitro* assembly was not efficient. Subsequently, methods utilized electroporation of ϕ 6 and ϕ 8 cDNA plasmids carrying three genomic segments that were introduced directly into *pseudomonad* hosts. The host cells expressed either SP6 or T7 RNA polymerase that directed recombinant viral ssRNA transcription. The latter methodology has contributed to the understanding of the precise nature of the RNA packaging and replication signals and mechanisms for RNA segment recognition by the PC. The *in vitro* self-assembly of complete and partial PC particles has been achieved as well by condensation of the purified protein components. The particles were found to be capable of packaging the viral ssRNA via NTP hydrolysis, replication the dsRNA and subsequent transcription of nascent ssRNA. The self-assembly assay has facilitated the study of PC assembly kinetics and the precise protein stoichiometry in the particles.

Applications and Intellectual Property

Since the time of its discovery, ϕ 6 has been the subject of intellectual property applications. The observation that the phage contained dsRNA stimulated the filing and award of a patent in 1974 for the preparation of high yields of dsRNA for interferon production. In 2006 a patent was awarded for the rescue of recombinant ϕ 6 that carried cDNA for the expression of therapeutic genes encoding vaccine antigens, catalytic or antisense RNAs and immunoregulatory proteins. The claims of the invention suggested that the isolated recombinant phage NC could be induced to enter eukaryotic tissue where the bioreactive gene would be expressed by a cap-independent translation enhancer after transcription by the ϕ 6 P2 RdRP. While these two applications have not been brought to practice recombinant ϕ 6 P2 has been made commercially available. The RdRP, with its capacity for primer independent replication and transcription, have made it a useful tool for the synthesis of large RNA molecules and RNA-based sequencing.

Further Reading

- Díaz-Muñoz, S.L., Tenaillon, O., Goldhill, D., *et al.*, 2013. Electrophoretic mobility confirms reassortment bias among geographic isolates of segmented RNA phages. *BMC Evolutionary Biology* 13, 206.
- El Omari, K., Sutton, G., Ravantti, J.J., *et al.*, 2013. Plate tectonics of virus shell assembly and reorganization in phage ϕ 8, a distant relative of mammalian reoviruses. *Structure* 21 (8), 1384–1395.
- Ford, B.E., Sun, B., Carpino, J., *et al.*, 2014. Frequency and fitness consequences of bacteriophage ϕ 6 host range mutations. *PLoS One* 9 (11).
- Gottlieb, P., Wei, H., Potgieter, C., Toporovsky, I., 2002. Characterization of ϕ 12, a bacteriophage related to ϕ 12: Nucleotide sequence of the small and middle double-stranded RNA. *Virology* 293 (1), 118–124.
- Gottlieb, P., Wei, H., Potgieter, C., Toporovsky, I., 2002. Characterization of ϕ 12, a bacteriophage related to ϕ 12: Nucleotide sequence of the large double-stranded RNA. *Virology* 295 (2), 266–271.
- Katz, G., Wei, H., Alimova, A., *et al.*, 2012. Protein P7 of the cystovirus ϕ 6 is located at the three-fold axis of the unexpanded procapsid. *PLoS One* 7 (10).
- Leo-Macias, A., Katz, G., Wei, H., *et al.*, 2011. Toroidal surface complexes of bacteriophage ϕ 12 are responsible for host-cell attachment. *Virology* 414 (2), 103–109.
- Mäntynen, S., Sundberg, L.R., Poranen, M.M., 2018. Recognition of six additional cystoviruses: *Pseudomonas* virus ϕ 6 is no longer the sole species of the family *Cystoviridae*. *Archive of Virology* 163 (4), 1117–1124.
- McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T., 2016. Reassortment in segmented RNA viruses: Mechanisms and outcomes. *Nature Reviews Microbiology* 14 (7), 448–460.
- Mindich, L., 2012. Packaging in dsRNA viruses. *Advances in Experimental Medicine and Biology* 726, 601–608.
- Nemecek, D., Boura, E., Wu, W., *et al.*, 2013. Subunit folds and maturation pathway of a dsRNA virus capsid. *Structure* 21 (8), 1374–1383.
- Nemecek, D., Qiao, J., Mindich, L., Steven, A.C., Heymann, J.B., 2012. Packaging accessory protein P7 and polymerase P2 have mutually occluding binding sites inside the bacteriophage ϕ 6 procapsid. *Journal of Virology* 86 (21), 11616–11624.
- Qiao, J., Mindich, L., 2013. The template specificity of bacteriophage ϕ 6 RNA polymerase. *Journal of Virology* 87 (18), 10190–10194.
- Sun, Y., Qiao, X., Qiao, J., Onodera, S., Mindich, L., 2003. Unique properties of the inner core of bacteriophage ϕ 8, a virus with a segmented dsRNA genome. *Virology* 308 (2), 354–361.
- Vidaver, A.K., Koski, R.K., Van Etten, J.L., 1973. Bacteriophage ϕ 6: A Lipid-Containing Virus of *Pseudomonas phaseolicola*. *Journal of Virology* 11 (5), 799–805.

Relevant Websites

https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsrna-viruses/w/cystoviridae
Cystoviridae.

<https://viralzone.expasy.org/750>
dsRNA virion.

<https://emedicine.medscape.com/article/227348-overview>
Medscape (Reoviruses).

http://www.virology.net/Big_Virology
The Big Picture Book of Viruses.

<http://www.marinespecies.org/aphia.php?p=taxdetails&id=600254>
WoRMS - World Register of Marine Species.

Membrane-Containing Icosahedral DNA Bacteriophages

Roman Tuma, University of Leeds, Leeds, United Kingdom and University of South Bohemia, České Budějovice, Czech Republic

Sarah J Butcher, University of Helsinki, Helsinki, Finland

Hanna M Oksanen, Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

dsDNA Double-stranded DNA

ssDNA Single-stranded DNA

T Triangulation number

Glossary

DNA packaging Energy-requiring process where the empty virus capsid is filled by the virus genome.

Protein-primed DNA replication Duplication of a linear DNA genome utilizing a protein covalently linked to the DNA terminus to initiate the reaction.

Temperate Bacteriophage can either lyse the cell or enter a resting state called lysogeny, staying dormant within the cell until induced into lytic cycle.

Triangulation number, *T* number (*T*) A parameter defined for icosahedral viruses that describes the geometrical arrangement of the protein subunits within the capsid shell.

Virulent Bacteriophages lyse the cell releasing new viral progeny.

Introduction

The icosahedral bacterial viruses with an internal membrane can have double-stranded (ds) or single-stranded (ss) DNA genomes. Currently these viruses are classified into four families: *Tectiviridae*, *Corticoviridae*, *Matsushitaviridae* (previously in *Spherolipoviridae*), and *Finnlakeviridae* (Table 1). In addition, there are virus isolates not yet assigned to any family and a newly proposed family *Autolykiviridae*. The dsDNA genomes are either linear or circular, and the only known ssDNA phage with an internal membrane has a circular genome. They encompass both virulent and temperate viruses, their virions range from 49 nm to 80 nm in diameter and lack a tail but may exhibit a special vertex containing a DNA packaging machinery.

Tectiviruses form the most numerous group, and they are classified into three genera: *Alphatectivirus*, *Betatectivirus*, and *Gammatectivirus* (and a newly proposed *Deltatectivirus*) (Fig. 1). Members of the *Alphatectivirus* (PRD1 and PR4) are virulent phages of Gram-negative hosts (e.g., *Pseudomonas* and *Salmonella*). These viruses along with the four extremely similar phages (PR3, PR5, PR772, and L17) have linear dsDNA genomes. Their sequence similarity is between 91.9% and 99.8%, which is surprising since they have been isolated from different parts of the world. The viruses of the genus *Betatectivirus* are temperate phages (AP50, Bam35, GIL16, and Wip1) replicating in Gram-positive *Bacillus* species. The third genus *Gammatectivirus* has a single member, the first temperate tectivirus, GC1, with a Gram-negative host (*Gluconobacter*). The type virus of the *Tectiviridae* is PRD1, which infects a wide variety of Gram-negative bacteria. Its host range is limited to bacteria that contain a conjugative antibiotic resistance plasmid, since it utilizes the plasmid-encoded cell surface DNA-transfer complex as a receptor. Bam35 serves as a model virus for phages infecting *Bacillus* species. The length of the genomes (about 15 kbp) and the order of the genes is conserved in all the tectiviruses, but there is no significant sequence similarity between the two major groups (*Alphatectivirus* and *Betatectivirus*). The genomes of the tectiviruses encode about 35 proteins as exemplified by the PRD1 genes (Table 2).

The type virus of the family *Corticoviridae* is PM2. PM2 was isolated in 1968 from seawater off the coast of Chile along with its host, which is a marine bacterium *Pseudoalteromonas espejiana*. PM2 was the first membrane-containing bacteriophage isolated. PM2 has a negatively-supercoiled, circular dsDNA of about 10 kbp. It encodes about 17 proteins (Table 3). Recently, *Pseudoalteromonas* phage Cr39582 was induced from *Pseudoalteromonas* sp. strain Cr6751 and assigned to the family *Corticoviridae*. PM2 and Cr39582 have 85% nucleotide identity but their spike protein genes are non-homologous. The entry and the DNA-packaging mechanisms differ from those of tectiviruses with linear dsDNA genomes.

Thermus bacteriophages P23–77 and IN93 (species *Hukuchivirus* P23-77 and *Hukuchivirus* IN93) are members of the family *Matsushitaviridae* and were recently moved from the family *Sphaerolipoviridae* (Table 1). Flavobacterium phage FLiP with a circular ssDNA genome has been recently assigned to the new family *Finnlakeviridae*.

In addition, isolates of *Salisaeta* phage SSIP-1 and *Helicobacter pylori* phage KHP30 with circular dsDNA genomes are yet to be assigned families. These viruses share no significant sequence homology to any previously identified virus group. A group of 18, marine, virulent, non-tailed viruses with linear, dsDNA, 10 kbp genomes infecting members of the *Vibrionaceae* were recently identified and the family *Autolykiviridae* was proposed after phylogenetic analyzes (Table 1). The proposed *Autolykiviridae* members have 21%–25% amino acid identity to PM2 in their major capsid protein but the protein-primed DNA polymerase has 36%–37%

Table 1 Icosahedral internal membrane-containing DNA bacteriophages with complete genome sequence

Genome	Taxonomy, family	Taxonomy, virus species	Putative members
<i>dsDNA, linear</i>	<i>Tectiviridae</i>	<i>Pseudomonas virus PRD1</i> ; <i>Pseudomonas virus PR4</i> ; <i>Bacillus virus AP50</i> ; <i>Bacillus virus Bam35</i> ; <i>Bacillus virus GIL16</i> ; <i>Bacillus virus Wip1</i> ; <i>Gluconobacter virus GC1</i> ;	Enterobacteria phage PR3, PR5, L17, and PR722; Thermus phage P37–14; Bacillus phage phiNS11, GIL01, and GIL16; Rhodococcus phage Toi1; Streptomyces phage WheeHeim, and Forthebois
	Proposed <i>Autolykiviridae</i>		Vibrio phage 1.008. O, 1.040. O, 1.011. O, 1.069. O, 1.062. O, 1.125. O, 2.092. O, 1.020. O, 1.080. O, 1.141. A, 1.043. O, 1.044. O, 1.057. O, 1.095. O, 1.107. A, 1.102. O, 1.048. O, 1.249. A
<i>dsDNA, circular</i>	<i>Corticoviridae</i>	<i>Pseudoalteromonas virus PM2</i> ; <i>Pseudoalteromonas virus Cr39582</i> ;	
	Proposed <i>Matsushitaviridae</i>	<i>Hukuchi virus P23–77</i> ; <i>Hukuchi virus IN93</i>	
	Unassigned Unassigned		Salisaeta phage SSIP-1 Helicobacter pylori phage KHP30
<i>ssDNA, circular</i>	<i>Finnlakeviridae</i>	<i>Flavobacterium virus FLiP</i>	Flavobacterium phage FLiP

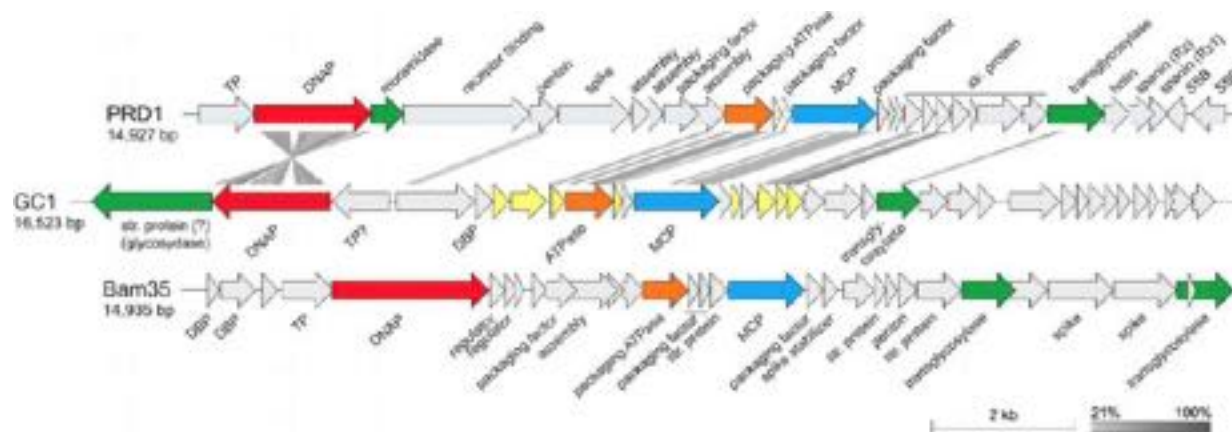


Fig. 1 Genomes of tectiviruses PRD1, Bam35, and GC1 representing *Alphatectivirus*, *Betatectivirus*, and *Gammatectivirus* genera, respectively. Similarity between translated nucleotide sequences is shown. Abbreviations and color codes: TP, terminal protein; DNAP, family B DNA polymerase (red); MCP, major capsid protein (blue); str. protein, structural protein; SSB, single-stranded DNA-binding protein; DBP, DNA-binding protein; muramidase (green), packaging ATPase (orange), transglycosylase (green). GC1 ORFs shared with alphatectiviruses are highlighted in light yellow. Gray ORFs do not have homologs in either of the other two genomes. Reprinted from Philippe, C., Krupovic, M., Jaomanjaka, F., *et al.*, 2018. Bacteriophage GC1, a novel tectivirus infecting *Gluconobacter cerinus*, an acetic acid bacterium associated with wine-making. *Viruses* 10 (1), 39, doi:10.3390/v10010039. Copyright © 2018 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license <http://creativecommons.org/licenses/by/4.0/>.

amino acid identity to members of the *Tectiviridae*. Large numbers of viruses with double jelly-roll major capsid protein have been detected in the genomes of diverse major bacterial and archaeal phyla, and in marine water column and sediment metagenomes, indicating that the diversity of this virus group greatly exceeds the diversity that is currently recognized (Fig. 2; Table 1).

The most studied model system in this group of viruses is bacteriophage PRD1, which is used here as an example to illustrate some of the most common characteristics of this diverse virus group.

Virion Structure and Properties

Overall Structure

The virion of PRD1 is an icosahedrally-symmetric particle approximately 65 nm in diameter (**Fig. 3**). It is composed of about 70% (mass fraction) protein, 15% lipid, and 15% DNA and has a mass of about 66 MDa. The structure of the virion has been studied

Table 2 PRD1 genes, corresponding proteins, and protein functions

<i>Gene</i>	<i>Protein</i>	<i>Mass (kDa)</i>	<i>Description^a</i>
<i>I</i>	P1	63.3	DNA polymerase (N)
<i>II</i>	P2	63.7	Receptor binding protein (S)
<i>III</i>	P3	43.1	Major capsid protein (C)
<i>V</i>	P5	34.2	Trimeric spike protein (S)
<i>VI</i>	P6	17.6	Unique vertex protein, DNA packaging accessory protein (C, P)
<i>VII</i>	P7	27.1	DNA delivery, holin, transglycosylase (L, M)
<i>VIII</i>	P8	29.5	Genome terminal protein (N)
<i>IX</i>	P9	25.8	Unique vertex protein, DNA packaging ATPase (C, P)
<i>X</i>	P10	20.6	Assembly (A, N)
<i>XI</i>	P11	22.2	DNA delivery (M)
<i>XII</i>	P12	16.6	ssDNA binding protein (N)
<i>XIV</i>	P14	15.0	DNA delivery (M)
<i>XV</i>	P15	17.3	Endolysin (L)
<i>XVI</i>	P16	12.6	Infectivity (M)
<i>XVII</i>	P17	9.5	Assembly (A, N)
<i>XVIII</i>	P18	9.8	DNA delivery (M)
<i>XIX</i>	P19	10.5	ssDNA binding protein (N)
<i>XX</i>	P20	4.7	Unique vertex protein (M, P)
<i>XXII</i>	P22	5.5	Unique vertex protein (M, P)
<i>XXX</i>	P30	9.0	Minor capsid protein (C)
<i>XXXI</i>	P31	13.7	Pentameric penton protein, base of spike (S)
<i>XXXII</i>	P32	5.4	DNA delivery (M)
<i>XXXIII</i>	P33	7.5	Assembly (A, N)
<i>XXXIV</i>	P34	6.7	(M)
<i>XXXV</i>	P35	12.8	Holin (L)
<i>XXXVI</i>	P36	13.2	Accessory lysis protein, spanin (L)
<i>XXXVII</i>	P37	9.9	Accessory lysis protein, spanin (L)

^a(N) non-structural early protein; (M) integral membrane protein; (S) spike complex protein; (A) assembly protein; (P) packaging protein; (C) capsid protein; (L) lysis protein.

Table 3 PM2 genes, corresponding proteins, and protein functions

<i>Gene</i>	<i>Protein</i>	<i>Mass (kDa)</i>	<i>Description</i>
<i>I</i>	P1	37.5	Spike protein
<i>II</i>	P2	30.2	Major capsid protein
<i>III</i>	P3	10.8	Major membrane protein
<i>IV</i>	P4	4.4	Membrane protein
<i>V</i>	P5	17.9	Membrane protein
<i>VI</i>	P6	14.3	Major membrane protein
<i>VII</i>	P7	3.6	Membrane protein
<i>VIII</i>	P8	7.3	Membrane protein
<i>IX</i>	P9	24.7	Potential packaging ATPase
<i>X</i>	P10	29.0	Membrane protein
<i>XII</i>	P12	73.4	Replication initiation protein
<i>XIII</i>	P13	7.2	Transcription factor
<i>XIV</i>	P14	11.0	Transcription factor
<i>XV</i>	P15	18.1	Transcription factor
<i>XVI</i>	P16	10.3	Transcription factor
<i>XVII</i>	P17	6.0	Lysis
<i>XVIII</i>	P18	5.7	Lysis

extensively using cryo electron microscopy and X-ray crystallography (Fig. 3(a)). X-ray crystallography results have indicated the roles of four proteins in controlling virus assembly. There are 240 hexagonally shaped trimers of the major capsid protein, P3 (Fig. 4(a)), occupying the surface of the capsid on a pseudo $T = 25$ lattice (Fig. 3(a)), an arrangement that is also found in the human adenovirus capsid. The capsid protein fold and arrangement in the shell belong to the vertical β -barrel structural lineage (Fig. 4(c)) is common among viruses with an internal membrane not only bacterial viruses but also archaeal (STIV, STIV2) and eukaryotic viruses (PBCV-1) having one double β -barrel MCP. In addition, phages (P23-77, the family *Matshushitaviridae*) and archaeal viruses (SH1, HCIV-1, HHIV-2, the family *Sphaerolipoviridae*) have two major capsid proteins obeying vertical single β -barrel fold, thus belong also to the vertical β -barrel structural lineage. In PRD1, a dimer of the linear glue protein P30 extends

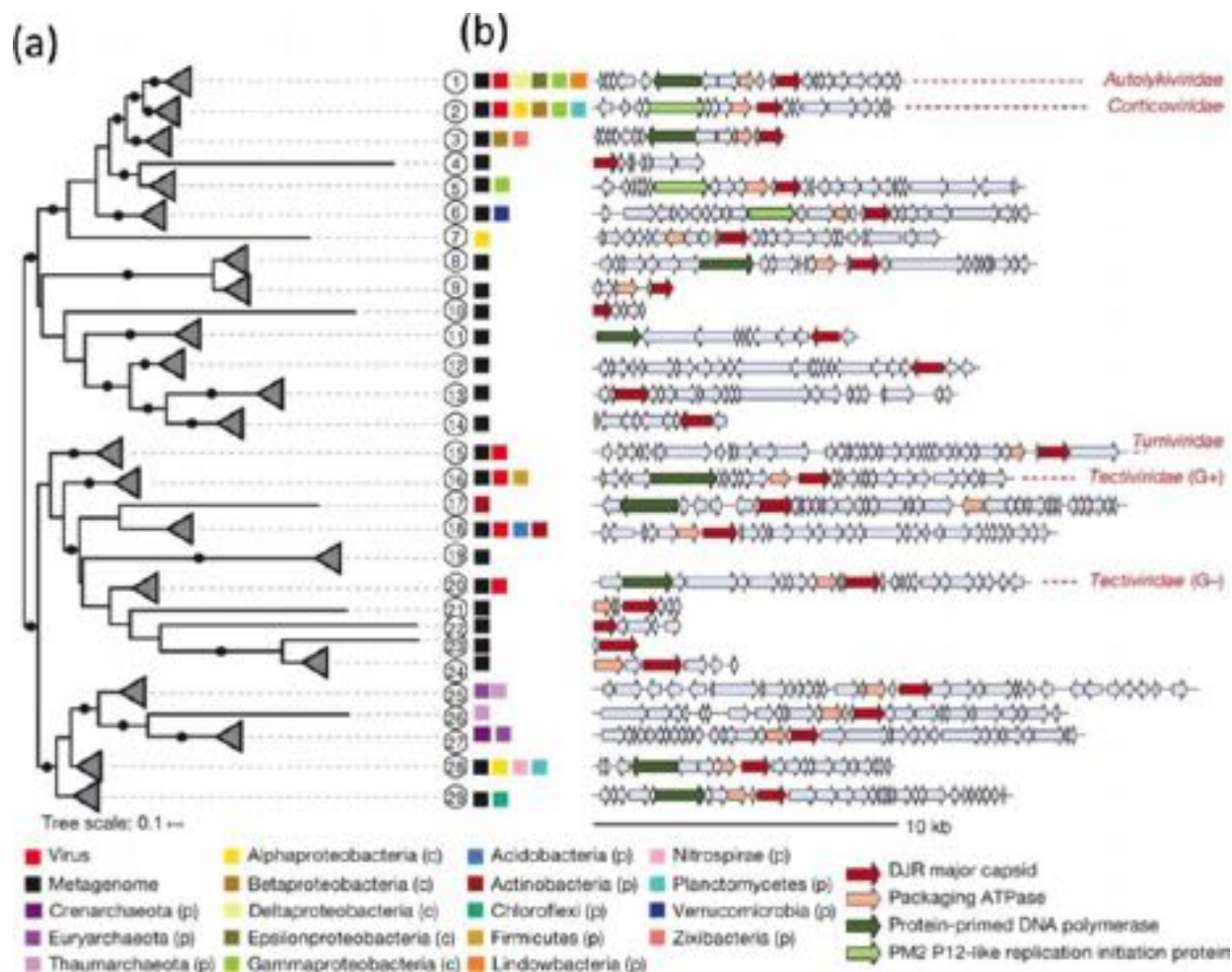


Fig. 2 Diversity of viruses exhibiting the double jelly-roll (DJR) major capsid protein. (a) Phylogeny of bacterial and archaeal DJR virus capsid proteins. Group numbers are assigned to each branch for reference, colored blocks indicate hosts, black circles on branches indicate approximate likelihood-ratio test branch support ≥ 0.9 . (b) Element gene diagrams from each group show prophage host genome neighborhoods and metagenome contigs often contain additional genes common to DJR elements. (G +, Gram-positive; G –, Gram-negative hosts). Reprinted with permission from Kauffman, K., Hussain, F.A., Yang, J., *et al.*, 2018. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* 554, 118–122. © 2018 Springer Nature.

from one vertex to the next, cementing the P3 facets together (Fig. 4(b)) and determining the overall size of the shell. Eleven of the vertices are composed of the pentamer penton protein P31, which interlock with P3 and the transmembrane protein P16 forming the base for the spike complex (Fig. 3(b)).

The receptor binding vertices have two proteins attached to the penton: one is a trimer of P5 (Fig. 5(a)) attached by the N-terminus, the other is a monomer of P2 (Fig. 5(b)), the receptor-binding protein. Both P5 and P2 are elongated molecules (P5 is 17 nm long, P2 is 15.5 nm long). P2 is a club-shaped molecule with a pseudo β -propeller head and a long tail formed from extended β -sheet. The head is proposed to be the site of receptor binding, lying distal to the virus.

The twelfth PRD1 vertex is the unique or special vertex used for genome packaging and DNA delivery (Figs. 3(b) and 6). At the unique vertex, the capsid and the internal membrane are interconnected by a hexamer composed of heterodimers of proteins P20 and P22 and serves as a membrane conduit for the genome and as a nucleation site for the special vertex assembly. The outer part of the unique vertex is formed of a complex of packaging ATPase P9 and packaging efficiency factor P6 (Fig. 6(c)). When the packaging starts, protein P9 is recruited to the procapsid and the PRD1 genome with the terminal protein P8 is packaged into the procapsid powered by ATP hydrolysis. The packaging ATPase P9 is part of the mature virion in contrast to the packaging ATPases of tailed dsDNA bacteriophages.

The overall size and structure of the phage Bam35 is very similar to that of PRD1. However, the exact counterparts of many of the PRD1 structural proteins have not yet been clearly identified. Two of the major differences between PRD1 and Bam35 are firstly, the presence of a large transmembrane protein complex in Bam35 that modulates the curvature of the membrane under the capsid facets, and secondly, the receptor-binding proteins of the attachment vertices.

The corticovirus PM2 particle is icosahedrally-symmetric and measures 57 nm in diameter from vertex to vertex. The mass of the virion is ~ 45 MDa and it is composed of protein (72%), lipid (14%), and DNA (14%). The capsid is composed of 200 trimers

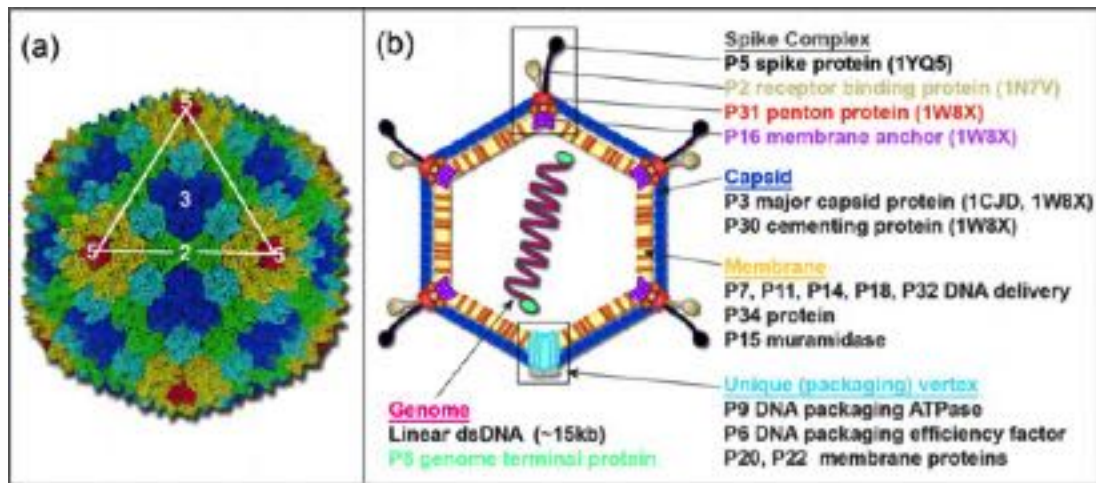


Fig. 3 (a) Cryo-electron microscopy reconstruction of PRD1 virion. One facet of the $T = 25$ capsid is delineated together with position of the symmetry axes. Pentons are red, peripentonal hexons are yellow, three-fold hexon clusters are blue. (b) Schematics of protein, lipid and DNA location within the PRD1 virion. Reprinted from Peralta, B., *et al.*, 2013. Mechanism of Membranous Tunneling Nanotube Formation in Viral Genome Delivery. PLOS Biology 11 (9), e1001667. © 2018 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license <http://creativecommons.org/licenses/by/4.0/>.

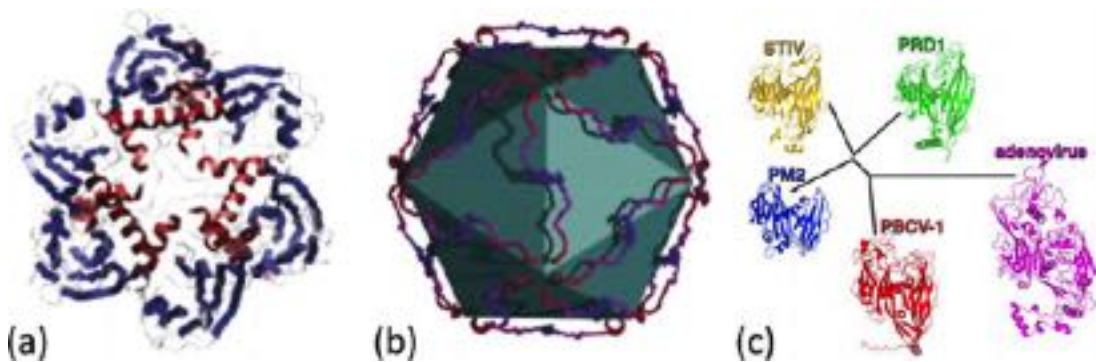


Fig. 4 (a) Structure of the PRD1 capsid P3 protein trimer with a double β -barrel fold (blue), resulting in a hexameric shape, based on X-ray crystallography. (b) Virion structure revealed a dimer of the cementing protein P30 (also known as the tape measure protein) running lying underneath the capsid shell along the twofold positions. The icosahedron is representing the viral membrane. Dimers of P30 are stabilizing the virion structure. Reprinted from Bamford and Butcher (2008) Encyclopedia of Virology (Third Edition) Icosahedral dsDNA Bacterial Viruses with an Internal Membrane. pp 1–6. © 2008 Elsevier. (c) Structural clade tree for the vertical double barrel lineage. Reprinted from Abrescia, N.G., Grimes, J.M., Kivela, H.M., *et al.*, 2008. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. Molecular Cell 31, 749–761. © 2008 Elsevier.

of the MCP P2 arranged on a pseudo $T = 21$ lattice. Pentameric receptor-binding spikes protrude from the vertices (Fig. 5(c)). The receptor binding domain of the spike is structurally related to the PRD1 P5 head domain (Fig. 5(a) and (c)).

Membrane and DNA

In PRD1, about half of the virion proteins are associated with the membrane (Table 2). The lipid headgroups are predominantly phosphatidylethanolamine (53%) and phosphatidylglycerol (43%) with 4% of cardiolipin. The membrane is well-ordered, following the icosahedral outline of the capsid. Many interactions occur between the membrane, the capsid and the DNA. The average separation of the concentric layers of DNA is approximately 2.5 nm, similar to that found in other bacteriophages and animal viruses. Removal of the capsid and spike proteins by heat or guanidinium hydrochloride treatment results in aggregation of the membrane vesicle.

In PM2, the membrane, lying underneath the capsid, follows the shape of the capsid as in PRD1, and there are many interactions mediated by the major membrane proteins P3 and P6 (Table 3). The lipid composition is approximately 64% phosphatidylglycerol, 27% phosphatidylethanolamine, and 8% neutral lipids and a small amount of acyl phosphatidylglycerol. Release of the capsid and spike proteins from the virion by freeze–thawing or by chelation of calcium ions with ethylene glycol tetraacetic acid (EGTA) results in a soluble vesicle called the lipid core.

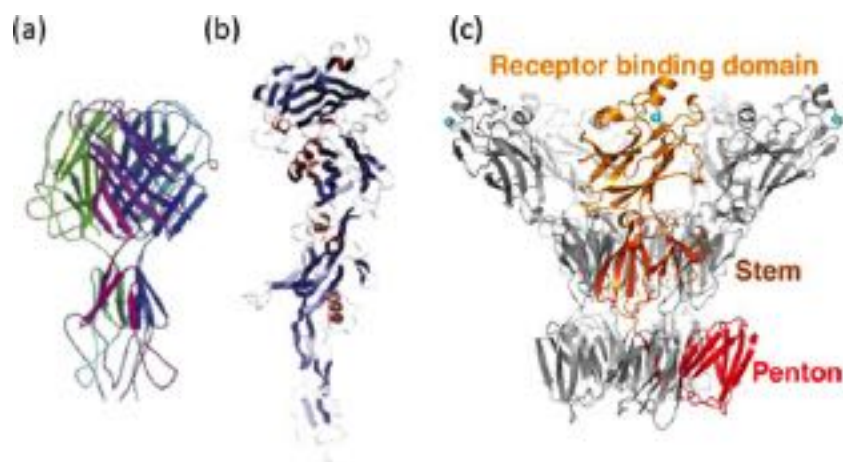


Fig. 5 Structure of the vertex proteins determined by X-ray crystallography. (a) The trimeric PRD1 spike protein P5. (b) The monomeric PRD1 receptor-binding protein P2. Reprinted from Bamford and Butcher (2008) *Encyclopedia of Virology* (Third Edition) Icosahedral dsDNA Bacterial Viruses with an Internal Membrane, pp. 1–6. © 2008 Elsevier. (c) Pentameric vertex structure of PM2 virion. Reprinted from Abrescia, N.G., Grimes, J.M., Kivela, H.M., *et al.*, 2008. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Molecular Cell* 31, 749–761. © 2008 Elsevier.

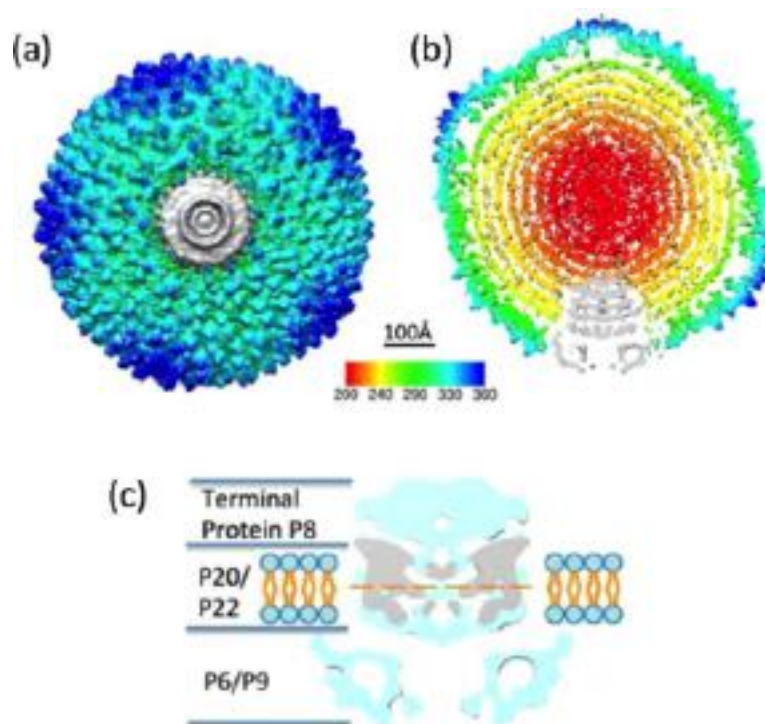


Fig. 6 Structure of the PRD1 packaging vertex obtained by asymmetric cryo-EM reconstruction. (a) View along the five-fold axis. (b) Cross-section through the packaging vertex and internal DNA density. (c) Schematic of protein localization within the vertex density with the packaging ATPase P9 capping the outside of the vertex. Reprinted from Hong, C., Oksanen, H.M., Liu, X., *et al.*, 2014. A structural model of the genome packaging process in a membrane-containing double stranded DNA virus. *PLoS Biology* 12 (12), e1002024. © 2014 by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license <http://creativecommons.org/licenses/by/4.0/>.

Life Cycle

PRD1 is a virulent phage that exploits the transcription functions of its host. The host cell is selected by specific recognition of a plasmid-encoded cell surface receptor with its primary function in bacterial conjugation. PRD1 belongs to the class of broad-host-range, donor-specific phages, which infect cells harboring one of the IncP-, IncN-, or IncW-type multiple-drug resistance

conjugative plasmids. Among the hosts are several opportunistic human pathogens such as *Escherichia coli*, *Salmonella enterica*, and *Pseudomonas aeruginosa*. After adsorption, the genome is injected into the host cell cytosol with the help of the inner membrane, leaving the protein capsid outside. After the replication of the phage genome and production of the phage components, both virus- and host-encoded factors assist in particle assembly. Host cell lysis releases some 500 progeny viruses. Bam35 is a temperate phage, that can have a lytic life cycle like PRD1, or exist in a dormant state within the host (lysogenic life cycle). In contrast to PRD1, the host range of Bam35, and its relatives GIL01 and GIL16, is limited to one species, *Bacillus thuringiensis*. AP50 infects *Bacillus anthracis*, and a defect phage replicating as a linear plasmid (pBClin15) has been described for *Bacillus cereus*. The corticovirus PM2, P23–77 phage of *Thermus thermophilus*, SSIP-1 phage of *Salisaeta* species, the *Vibrio* phages are virulent viruses. *Thermus* phage IN93 and *Streptomyces* phages WheelHeim and Forthebois are temperate phages. Of all the DNA phages with an internal membrane, the life cycle of PRD1 is best understood and is described below.

Receptor Recognition and DNA Delivery

A single phage structural protein, P2, is responsible for PRD1 attachment to its host. Each of the eleven receptor recognition vertices is a metastable structure and possibly capable of recognizing the receptor. After binding to the receptor, the particle most probably re-orientes (possibly guided by secondary binders e.g., protein P5) so that the unique vertex is perpendicular to the cell surface. The association of P2 with the receptor activates, possibly by receptor binding complex removal, the injection process. This leads to irreversible binding. Both empty and DNA-containing particles are bound equally tightly to cells, indicating that DNA injection is not a prerequisite for this tight interaction.

Isolation and analysis of PRD1 mutants have resulted in the identification of eight, phage-specific, structural proteins essential for infectivity and DNA delivery. In addition to the spike complex proteins (P2, P5, and P31 especially P2 needed for adsorption), protein P11 starts the DNA delivery process and membrane proteins P14, P18, and P32 are involved in later stages of the DNA delivery. Mutant particles missing protein P7 (a lytic transglycosylase) are infectious but the DNA entry process is delayed. The PRD1 membrane undergoes a structural transformation from a spherical vesicle to a tubular form which penetrates the host cell envelope and acts as a conduit for DNA delivery. DNA ejection tube protrudes from the unique vertex used for DNA packaging. The packaged genome (DNA-P8 complex) does not have a role in the triggering or in the formation of the DNA delivery tube, since also procapsids devoid of the packaged genome can form tubes. Similar tube formation is typical for all members of the *Tectiviridae* and also seen in the *Vibrio* phages, but not recognized in other virus isolates representing internal membrane-containing DNA phages.

Genome Replication

The genome of PRD1 is a linear dsDNA molecule with proteins covalently (protein P8) attached to both 5' termini and having 110 bp terminal repeats. PRD1 replicates its DNA by a protein-primed replication mechanism similarly to other viruses with linear dsDNA genomes, including adenovirus and the ϕ 29-type phages. PRD1 DNA replication starts with the formation of a covalent bond between the genome terminal protein, P8, and the 5' terminal nucleotide, dGMP, in a reaction catalyzed by the phage DNA polymerase (protein P1). The minimal origin of replication resides in the 20 first-terminal base pairs of both genome ends, and the fourth base from the 3' end of the template directs, by base complementation, the linking of deoxyribonucleoside monophosphate (dNMP) to the terminal protein. The 3' end DNA sequence is maintained by sliding back of the polymerase complex.

After initiation, elongation of the initiation complex by the same DNA polymerase takes place resulting in the formation of full-length daughter DNA molecules. Two phage-encoded ssDNA binding proteins, P12 and P19, are involved in replication in vivo.

Particle Assembly

Approximately 15 min post infection, the major capsid protein P3 and the spike complex proteins P2, P5, and P31 are found soluble in the host cell cytosol, whereas the phage-encoded membrane proteins (e.g., P7, P11, P14, and P18) are associated with or inserted into the host cell cytoplasmic membrane (CM). Correct folding and assembly of several viral proteins are dependent on the host GroEL/ES chaperonins. Upon assembly, a virus-specific patch from the host CM is pinched off by the forming procapsid using the membrane-bound scaffolding protein P10. In addition, two small phage-encoded non-structural proteins P17 and P33 function in the assembly process, possibly in association with the host chaperonin complex GroEL/GroES.

Correct assembly results in an empty capsid enclosing a membrane rich in phage-specific proteins (procapsid). The linear dsDNA genome is packaged into the prohead by the packaging ATPase P9. Unlike packaging ATPases of most other icosahedral dsDNA bacteriophages, P9 is part of the mature virus structure. It resides at a single vertex that also contains proteins P6, P20, and P22. P6 is a soluble protein needed for efficient DNA packaging and the latter two are integral membrane proteins connecting the portal structure to the viral membrane.

Cell Lysis

At the end of the infection cycle, the newly synthesized progeny virions are released via host cell lysis. A two-component, holin–endolysin system operates in phage PRD1. The endolysin protein, P15, is a soluble β -1,4-N-acetylmuramidase that degrades the peptidoglycan of the Gram-negative cell causing host cell lysis. The holin protein, P35, is a helper factor that makes pores in the inner membrane to facilitate the access of the endolysin to the peptidoglycan in the cell wall, controlling the timing of lysis. In addition, PRD1 has accessory lysis proteins P36 and P37 that function as a pair (spanins) which is an interchangeable analog of the Rz/Rz1 of bacteriophage lambda. Spanins act after cell-wall degradation, aggregating and causing the disruption of the outer membrane.

Genomes and Genomics

The length of the PRD1 genome is 14,927 bp and the guanine–cytosine (GC) content is 48.1%. It has 110 bp-long inverted terminal repeat (ITR) sequences at the ends, which are 100% identical. The majority of the PRD1 ORFs products have been identified and currently there are 27 known genes (Table 2). The genomes of the other Gram-negative bacteria-infecting tectiviruses are similar, varying between 14,935 and 14,954 bp, and exhibit remarkably conserved nucleotide sequences, the overall identity being 91.9%–99.8%. The Bam35 genome is 14,935 bp long with a GC content of 39.7%. The ITRs are 74 bp long with only 81% identity between the ends. PRD1, Bam35, and GC1 (representatives of the three genera of *Tectiviridae*) do not share much sequence similarity (Fig. 1). The discovery of the almost invariant genomes within the tectivirus groups contrasts sharply with the situation observed for the tailed dsDNA bacteriophages. The nucleotide identity is 100% between Bam35 and the related GIL01 and 83% between GIL01 and GIL16.

PRD1 genome is organized into two early and three late operons. Although there are no major sequence similarities between PRD1, Bam35, and GC1 genomes, common genes for DNA polymerase, packaging ATPase, muramidase, transglycosylase lytic enzyme, and major capsid protein can be recognized by corresponding conserved amino acid sequence motifs in the databases or identified by comparing the N-terminal amino acid sequence of the major virion protein to the DNA sequence. Typically, the genes responsible for the host cell recognition (the spike protein) are the most diverse even between closest viruses.

The circular PM2 genome is 10,079 bp long with a GC content of 42.2%. It contains 21 putative genes, of which 17 have so far been shown to be functional (Table 3). Promoter mapping by primer extension has revealed three operons, which are expressed in a timely fashion during infection. The first early operon is highly similar to the maintenance region of the *Pseudoalteromonas* plasmid pAS28. The second early operon contains genes for DNA replication and regulation of late phage functions.

The PM2 genome replicates via a rolling-circle mechanism. Protein P12 has conserved sequence motifs common to superfamily I replication initiation proteins. This superfamily consists of the A proteins of certain bacteriophages, such as ϕ X174 and G4, and the initiation proteins of cyanobacterial and archaeal plasmids. The function of the late operon is activated by two phage-encoded transcription factors, P13 and P14. P14 has sequence similarity to the TFIIIS-type general eukaryotic transcription factors most closely resembling those of the archaeal organisms *Thermococcus celer* and *Sulfolobus acidocaldarius*.

The structural proteins, encoded by the late genes, are similar to the corresponding proteins of tectiviruses, either membrane associated or soluble. Based on the conserved amino acid sequence motifs deduced from the nucleotide sequence, one of the structural virion proteins is a putative packaging ATPase. The packaging process of the circular supercoiled PM2 DNA is coupled with the formation of the particle.

Acknowledgment

Supported in part by ERDF CZ.02.1.01/0.0/0.0/15_003/0000441 to R.T.

Further Reading

- Aalto, A.P., Bitto, D., Ravantti, J.J., *et al.*, 2012. Snapshot of virus evolution in hypersaline environments from the characterization of a membrane-containing Salisaeta icosahedral phage 1. *Proceedings of the National Academy of Sciences of the United States of America* 109, 7079–7084. doi:10.1073/pnas.1120174109.
- Abrescia, N.G., Grimes, J.M., Kivela, H.M., *et al.*, 2008. Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. *Molecular Cell* 31, 749–761.
- Abrescia, N.G., Cockburn, J.J., Grimes, J.M., *et al.*, 2004. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* 432, 68–74.
- Atanasova, N.S., Sencilo, A., Pietilä, M.K., *et al.*, 2015. Comparison of lipid-containing bacterial and archaeal viruses. *Advances in Virus Research* 92, 1–61.
- Azinas, S., Bano, F., Torca, I., *et al.*, 2018. Membrane-containing virus particles exhibit the mechanics of a composite material for genome protection. *Nanoscale* 10, 7769–7779.
- Cockburn, J.J., Abrescia, N.G., Grimes, J.M., *et al.*, 2004. Membrane structure and interactions with protein and DNA in bacteriophage PRD1. *Nature* 432, 122–125.
- Demina, T.A., Pietilä, M.K., Svirskaitė, J., *et al.*, 2017. HCIV-1 and other tailless icosahedral internal membrane-containing viruses of the family *Sphaerolipoviridae*. *Viruses* 9, 32.
- Gill, J.J., Wang, B., Sestak, E., Young, R., Chu, K.H., 2018. Characterization of a novel tectivirus phage Tail and its potential as an agent for biolipid extraction. *Scientific Reports* 8, 1062.
- Hong, C., Oksanen, H.M., Liu, X., *et al.*, 2014. A structural model of the genome packaging process in a membrane-containing double stranded DNA virus. *PLOS Biology* 12, e1002024.
- Kauffman, K.M., Hussain, F.A., Yang, J., *et al.*, 2018. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature* 554, 118–122.

- Laanto, E., Mäntynen, S., De Colibus, L., *et al.*, 2017. Virus found in a boreal lake links ssDNA and dsDNA viruses. *Proceedings of the National Academy of Sciences of the United States of America* 114, 8378–8383.
- Laurinmäki, P.A., Huiskonen, J.T., Bamford, D.H., Butcher, S.J., 2005. Membrane proteins modulate the bilayer curvature in the bacterial virus Bam35. *Structure* 13, 1819–1828.
- Mäntynen, S., Sundberg, L.R., Oksanen, H.M., Poranen, M.M., 2019. Half a century of research on membrane-containing bacteriophages: Bringing new concepts to modern virology. *Viruses* 11, 76.
- Oksanen, H.M., Abrescia, N.G.A., 2019. Membrane-containing icosahedral bacteriophage PRD1: The dawn of viral lineages. *Advances in Experimental Medicine and Biology* 1215, 85–109.
- Peralta, B., Gil-Carton, D., Castano-Diez, D., *et al.*, 2013. Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLOS Biology* 11, e1001667.
- Rissanen, I., Grimes, J.M., Pawlowski, A., *et al.*, 2013. Bacteriophage P23-77 capsid protein structures reveal the archetype of an ancient branch from a major virus lineage. *Structure* 21, 718–726.

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsdna-viruses/w/corticoviridae
Corticoviridae.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/finnlakeviridae
Finnlakeviridae.
- https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/133/tectiviridae
Tectiviridae.

Tailed Double-Stranded DNA Phages

Robert L Duda, University of Pittsburgh, Pittsburgh, PA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Capsid or Head The protein container for the Caudovirales phage genome, usually icosahedral in shape, that protects it from the environment until it is delivered to a new host.

Contractile tail The tail type that defines the Myoviridae, composed of a baseplate that contacts the host surface, an outer protein cylinder or sheath and an inner tail tube or core. When a host is bound the sheath contracts driving the tube into the cell to deliver the genome to the host.

DNA packaging The energy-driven process by which a procapsid is filled with phage DNA to become a mature capsid.

GenBank The United States National Institutes of Health genetic sequence database, an annotated collection of all publicly available DNA sequences.

Horizontal gene exchange Transfer of genetic material between different species.

Portal A grommet-like ring composed of 12 copies of the portal protein that occupies one capsid vertex, and

through which DNA is packaged and to which the tail is attached.

Procapsid or Prohead An immature capsid that has a spherical shape and contains no DNA.

T number The Triangulation number (T), a formal descriptor for complex icosahedral structures that describes the geometric arrangement of subunits for a viral capsid. The number of capsid protein subunits is $\sim 60 \times T$.

Tail The part of the virion responsible for attachment to the host and genome injection that is attached to one vertex of the capsid, at the portal.

Tail fiber An elongated protein fiber attached to a phage tail that helps recognize a host by binding to the host by the tip of the fiber.

Virion The entire Caudovirales viral particle composed of a capsid containing the viral genome, tail, tail fibers and other appendages.

Introduction

The icosahedral tailed double-stranded DNA bacterial viruses or bacteriophages of order *Caudovirales* are among the most widely recognized icons of modern molecular biology. Anyone who has participated in an advanced biology, genetics, or molecular biology class is likely to have been exposed to images of the lunar-lander-shaped bacteriophage T4 virion (family *Myoviridae*) with its elongated icosahedral capsid and machine-like contractile tail (Fig. 1(A)) or the plain icosahedral capsid and elegant, gently curving tail of bacteriophage λ (family *Siphoviridae*, Fig. 1(B)). Both T4 and λ are bacteriophages that infect the enteric bacterial host *Escherichia coli* and have been studied extensively. Bacteriophages were originally discovered in the early 20th century (1915–17) by English microbiologist Frederick W. Twort, and French-Canadian microbiologist Felix d'Herelle. Although there was much early interest in bacteriophages because of their potential for use in treating bacterial diseases (phage therapy), research on bacteriophages for disease therapy was largely abandoned after the discovery of antibiotics. Bacteriophage research was revived in the 1940s by Max Delbrück and colleagues (the well-known Phage Group that often met at the Cold Spring Harbor Laboratories on Long Island, NY) and the focused phage research started by this group of scientists led to many fundamental biological discoveries during the birth of molecular biology. The modern emergence of multiply antibiotic resistant strains of human pathogens has rekindled an interest in developing phage therapy as a means of treating human, fish, insect (bees) and livestock diseases.

The Structure of Tailed dsDNA Bacteriophages

The invention and commercialization of the electron microscope led to the first images of phages as sperm-like particles with a head and tail using metal shadowing, in which images were formed by heavy metal atoms such as uranium that were evaporated in a vacuum and allowed to strike a dried phage specimen at an angle. The later introduction of negative stains (salts of heavy metals that dry as a thin layer without forming crystals and in which small particles such as phages could be embedded) for electron microscopy resulted in images that were far richer in detail than earlier methods and revealed the complexity and variety of phage morphology. Fig. 1 shows electron micrographs of phages T4 and λ made using the negative stain technique. Cryo-electron microscopy of frozen hydrated phage samples is the latest innovation and is capable of revealing near-atomic details of phages and their structural components. Electron microscopy has remained a major tool in the study of bacteriophages and, in fact, the current taxonomic system for phages relies heavily on phage morphology as determined by electron microscopy as a major discriminating factor. Table 1 shows a taxonomic table for order *Caudovirales* and some of their characteristics. *Myoviridae* have contractile tails and include bacteriophages T4 (Figs. 1(A) and 2) and P1. *Siphoviridae* have long non-contractile tails and include phages λ (Fig. 1(B)) and T5. *Podoviridae* have shorter stubby or stumpy tails and include phages P22 and T7. The utility of morphological classification as a basis for phage taxonomy has more recently come into question as new insights

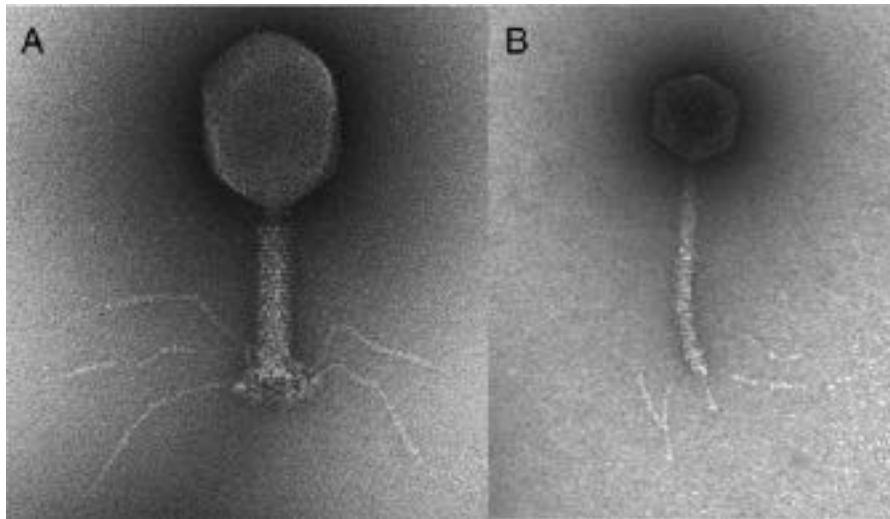


Fig. 1 Electron micrographs of two bacteriophages. (A) Bacteriophage T4. (B) Bacteriophage Ur- λ . Ur- λ is the original phage λ isolate and has the side tail fibers that are missing in the λ strains that are widely used. The micrographs were made using the negative stain Uryanyl acetate to embed the phage particles and provide contrast. The samples appear lighter than the enclosing stain.

Table 1 Tailed dsDNA bacteriophages: Order *Caudovirales*

Family	Genus	Defining example	Capsid T-number	Genome size bp
<i>Myoviridae</i> (phages with contractile tails)	"T4-like viruses"	Enterobacteria phage T4	T = 13; Q = 20	168,903
		Bacillus phage G	T = 52	497,513
	"P2-like viruses"	Enterobacteria phage P2	T = 7	33,593
	"PhiKZ-like viruses"	PhiKZ	T = 27	280,334
	"Mu-like viruses"	Enterobacteria phage Mu	T = 7	36,717
	"SP01-like viruses"	Bacillus phage SP01	T = 16	~132,500
<i>Siphoviridae</i> (phages with thin non-contractile tails)	" ϕ H-like viruses"	Halobacterium virus ϕ H	T = 7?	~57,000
	" λ -like viruses"	Enterobacteria phage λ	T = 7	48,502
	"T1-like viruses"	Enterobacteria phage T1	T = 7?	48,836
	"T5-like viruses"	Enterobacteria phage T5	T = 13	121,750
	"c2-like viruses"	Lactococcus phage c2	T = 4?; Q = 7?	22,172
	"L5-like viruses"	Mycobacterium phage L5	T = 7	52,297
	" ψ M1-like viruses"	Methanobacterium ψ M1	?	~23,246
	"T7-like viruses"	Enterobacteria phage T7	T = 7	39,937
<i>Podoviridae</i> (phages with short stubby tails)	" ϕ 29-like viruses"	Bacillus phage ϕ 29	T = 3; Q = 5	19,366
	"P22-like viruses"	Enterobacteria phage P22	T = 7	41,724

into phage evolution have revealed that horizontal exchange of genes is widespread among phages, and further, that some structural genes of bacteriophages and viruses that infect organisms from other domains of life appear to share common ancestors.

Bacteriophage Structure and Function

Capsids

In the dsDNA bacteriophages, the capsid serves as the container for the DNA chromosome and protects it from the environment until it is delivered to a new host by the phage tail, the organelle of attachment and genome delivery. In many other types of viruses, the capsid is taken up by cells with the genome still inside, and the virus particle uncoats or disassembles within the new host to initiate viral replication. The dsDNA bacteriophages capsids do not have to disassemble in this manner to initiate infection, so they can be constructed to be highly stable – resistant to a wide variety of chemical and physical assaults from the environment and digestive tracts – as they travel from one host to another. The dsDNA genome in bacteriophage capsids is packed to a very high density (~ 0.5 g/ml), so the capsid shell has to be strong enough to contain this tightly packed and highly concentrated DNA. Most of the *Caudovirales* have symmetric icosahedral capsids, but some have capsids with an elongated icosahedral shape, like T4 (Figs. 1(A) and 2) that have icosahedral ends with an elongated tubular middle section in between.

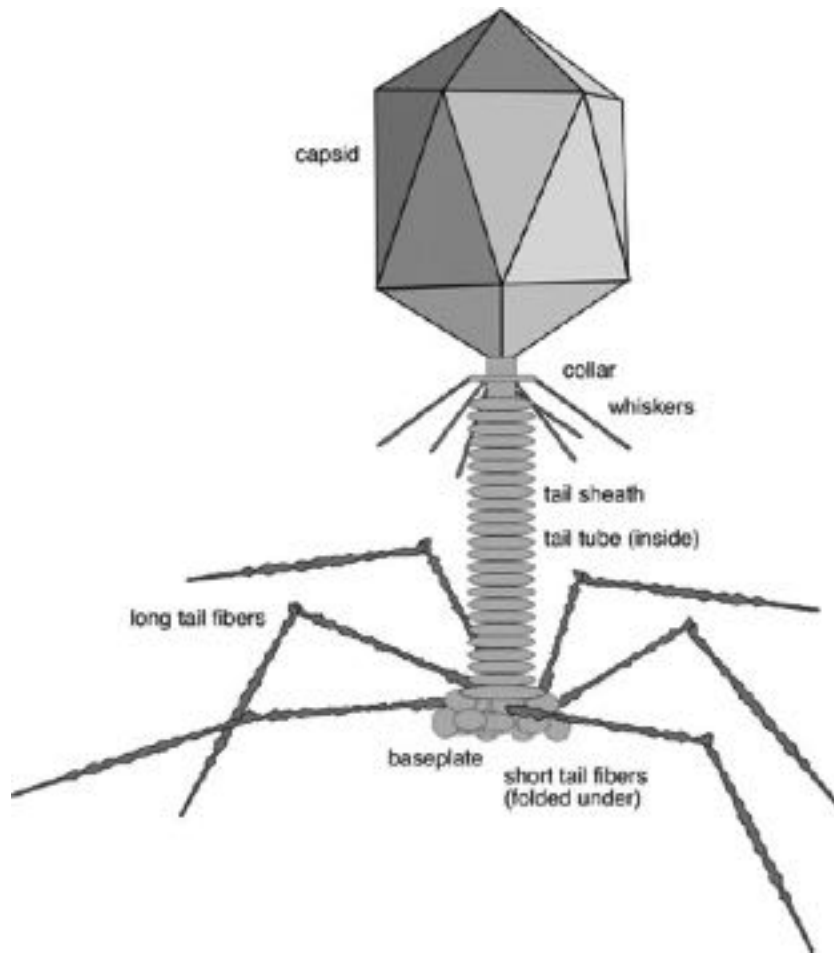


Fig. 2 Structural features of a typical bacteriophage. A drawing of a bacteriophage with a contractile tail, like phage T4. Many of the structural components that are discussed in the article are labeled.

Tails and Tail Fibers

Phage tails function to identify and bind to the correct host and to deliver the phage genome into that new host to initiate replication. Phage tails have a large variety of morphological forms with many parts and appendages, such as those labeled on the diagram of T4 in [Fig. 2](#) and visible in the electron micrograph in [Fig. 1](#). We know little about the functions of some tail parts, but others are understood in considerable detail. For example, in bacteriophage T4, both the long tail fibers and the short tail fibers (which are folded under the baseplate until after the long tail fibers are bound to the host – see [Fig. 2](#)) recognize specific host receptors. The long tail fibers of phage T4 are the primary determinants of host range, and mutations that change the T4 host range cause alterations in the fibers near their distal tips. The whiskers that are attached to the top of the T4 tail have multiple functions. The whiskers act as assembly jigs for adding the long tail fibers to the virion, and they also act as environmental sensors that sequester the long tail fibers under unfavorable conditions and thus prevent attachment to a host.

The long tail fibers of T4 have an analog in bacteriophage λ , but in the ordinary laboratory strains of λ , these fibers are not made because of a mutation in the side tail fiber gene. The electron micrograph in [Fig. 2](#) shows the long side tail fibers present on Ur- λ , a primordial λ that lacks this mutation. These long side tail fibers of Ur- λ speed up the adsorption of this phage to its host, but are not required for infection, because the primary interaction of λ with its host is mediated by the central tail fiber (a trimer of the tail protein gpI) protruding from the end of the lambda tail. The central tail fiber of λ binds to a host maltose transport protein (called LamB or MalB) in the outer membrane of the host. Binding of phage λ to its receptor under suitable conditions acts as a signal that triggers the injection of the phage DNA genome into the host to initiate replication.

The structure and functions of many parts of the contractile tails of Myoviridae have been revealed in depth, especially for the phage T4 tail. After the T4 long tail fibers attach to the host, the short tail fibers unfold from the baseplate and bind to a second receptor on the host surface. The baseplate remains attached to the cell and undergoes a dramatic reorganization in which it changes from a hexagon to a star shape, causing the sheath to contract and drive the internal tail tube into the cell. As the

iron-containing nail-shaped tip of the inner tail tube passes through the outer membrane, tail lysozyme molecules are released to create a small hole in the cell wall peptidoglycan layer. These combined actions allow the tip of the tail tube to reach the inner membrane, where a channel is created that allows the dsDNA genome to enter the cell and initiate replication.

Temperate Versus Virulent Bacteriophages

Virulent bacteriophages, once they have infected a host, are committed to go through their entire growth cycle including replication of the bacteriophage chromosome, production of progeny viral particles and the lysis of the host cell. Temperate phages, on the other hand, are able to grow lytically like their virulent counterparts, but they are also capable of going into a dormant state within the host. This dormant state is called lysogeny, and often involves integration of the phage chromosome into the host chromosome, as occurs with phage λ , or conversion of the phage chromosome into an autonomously replicating plasmid, as occur with phage P1. The seemingly normal host containing a dormant phage chromosome is called a lysogen. The dormant phage chromosome is maintained in a mostly silent state, until some change in the condition of the host (often related to stress, such as damage to the host DNA) acts as a signal to induce the phage to begin a lytic growth cycle, all the way through progeny release. The temperate phage in the lysogenic state makes a repressor protein which can repress nearly all phage functions, except for the synthesis of the repressor itself. If the repressor is subsequently destroyed, as it may be after DNA damage occurs within the cell, the phage initiates a normal lytic growth cycle, usually excising the phage chromosome from the host chromosome as an early step. How a temperate phage decides between a virulent and lysogenic mode of infection is a complex process that appears to depend on the conditions within the host cell at the time of infection. In a simplified version of the process, it is a balance between the production of sufficient repressor protein to shut down phage development and the antagonistic action of other factors that prevent repressor synthesis and function.

Injection

DNA Injection

As described above, all *Caudovirales* bacteriophage particles have tails with highly specialized parts with specific roles to mediate attachment and infection. The fundamental goal of attachment is the introduction of the viral DNA into the host to begin production of progeny virions. The detailed mechanism of DNA entry varies widely, from phage to phage. Although this process is called DNA injection, the phage particles do not act like syringes, as the name implies. In some cases it appears that chromosome entry is largely passive, a result of the DNA being very densely packed into the virion at a high concentration. The high concentration of DNA packaged in the phage capsid appears to play a role in the passive transfer of DNA out of the capsid into the host; this appears to be true for many phages, including phage T4 and phage λ . Other *Caudovirales*, such as phage T7 (family *Podoviridae*) and phage T5 (family *Siphoviridae*) have more elaborate programs of DNA injection with an initial stage that appears to be largely passive, followed by pause after only a fraction of the chromosome has been injected. Such paused DNA injection resumes only after proteins (encoded by the genes injected in the initial stage) act to restart chromosome entry or to actively pull the rest of the chromosome into the host in a controlled manner.

Protein Injection

Many *Caudovirales* inject both DNA and proteins into their hosts. The function of a few of these injected proteins is known, but in many cases remains mysterious. For example, phage T4 injects multiple copies of several proteins into the host. These include one protein that modifies the host RNA-polymerase, another that inhibits a host nuclease that targets and digests T4's highly modified DNA, and other proteins of unknown function. Surprisingly, most of phage T4's injected proteins are not absolutely required for infection and mutants that lack them are able to grow on many ordinary laboratory hosts. Phage N4 injects a very large and essential RNA polymerase protein that transcribes this phage's early genes. Phage T7 was thought to inject several large proteins into its host, but these have now been shown to be extruded from the capsid interior and subsequently form a structure which is the channel for the passage of DNA and other proteins into the host.

Host Interactions and Regulation of Gene Expression

Classes of Genes

A regulated program of gene expression begins soon after any *Caudovirales* has injected its DNA. The expression of phage genes can be divided by function, such as host takeover, replication, virion assembly, or lysis, but more often they are sorted by their temporal order of expression. In most cases gene expression is divided into early genes and late genes, although for some phage there may be a cascade of gene expression modes that occur sequentially. In a cascading series, early phage gene products control the expression of a set middle gene products, which control the expression of late genes.

Early Genes

Phage early genes are usually transcribed by the normal host RNA polymerase, while the expression of late genes is dependant on the expression of the early genes. The earliest genes expressed might include ones that produce proteins to counteract the host's anti-bacteriophage defensive systems, as occurs in phage T7. Early genes of temperate phages are those required to make the decision to become dormant or to grow normally and carry out the first steps of these processes, including, for example, the repressor gene and integration genes, or DNA replication and recombination genes. Early genes of virulent bacteriophages, such as phage T4, include genes for genome replication and genes that shut down unneeded host functions, such as host transcription and host protein synthesis, and may include other genes required to recycle host resources, for example, by specifically degrading the host chromosome to intermediates that can be re-used during the replication of the phage genome.

Late Genes

The proteins encoded by the late genes include the structural proteins required for the assembly of the bacteriophage virion, as well as those needed for packaging the genome into the virion and for cell lysis. Late gene expression is often dependant on early (or middle) gene expression because the late genes have transcriptional promoters that are not recognized by the normal host polymerase. The early genes may direct the synthesis of transcription factors that either (a) change the specificity of the host RNA polymerase or (b) modify the polymerase in other ways, in order to express the whole set of late genes at the appropriate time with high efficiency. Alternatively, one of the early genes may encode for an entirely new RNA polymerase with a new specificity that recognizes late gene promoters. Expression of the structural proteins to high levels allows the production of a large number of progeny phages.

Virion Assembly

Assembly Pathways

The assembly of *Caudovirales* virions is a highly ordered process in which separate subassemblies of the virion are built and then joined to form the mature virion. The assembly of capsids, tails and tail fibers each follows an independent assembly pathway in which individual protein components are added, usually sequentially, to a growing structure until it is complete.

Tail Assembly

The major components of a *Caudovirales* tail are a baseplate or tail tip, major tail proteins (one for a non-contractile tail; two for contractile tails, an inner tube protein and outer sheath protein), and termination or capping proteins to stabilize the completed tail and connect it to the head. Tail assembly begins with the formation of an initiator complex, which may be either a complete tail tip or a complete baseplate, that is assembled via a separate pathway. The initiator complex includes a template that specifies the length of the tail, the tape measure protein complex in a compact form and bound by tape measure-specific chaperone proteins. The major tail proteins bind to the initiator complex and assemble into a tube around the tape measure protein until the full length of the tape measure protein is enclosed and the termination proteins can bind and stabilize the tail.

Capsid Assembly

Capsids of *Caudovirales* are built as precursor procapsids, into which the phage genome is later packed. Procapsids are initially assembled from several types of components: a portal complex, which will connect the capsid to the tail, a major capsid protein, a scaffolding protein (which may be a separate protein or may be a disposable (removable) part of the major capsid protein). Phages may also utilize decoration or stabilization proteins, but these are not part of the procapsid and add to and stabilize the capsid only after the genome is packed inside. The number of capsid protein subunits needed for assembly is equal to $\sim 60 \times T$ subunits, where T is the triangulation number listed in [Table 1](#) (or, for phages with elongated capsids, $\sim 30 \times (T + Q)$ subunits, where T and Q are specified in [Table 1](#)). One vertex of a phage capsid is occupied by the portal complex, so the number of capsid proteins will be 5 less than calculated by the simplified formulae. So for a $T = 7$ virus, such as λ , $420 - 5 = 415$ copies are needed. Some phages, including phage T4 and relatives, also make a specialized minor capsid protein that forms the pentameric vertices of the icosahedral capsids, but most do not. Capsid assembly in *Caudovirales* begins with the completion of the assembly initiator complex, the portal or connector, composed of 12 copies of the portal protein, which occupies one of the 12 icosahedral vertices of the capsid. The major capsid protein(s), together with hundreds of copies of the scaffolding protein co-assemble onto the initiator to produce a complete shell with the scaffolding protein on the inside. After assembly is complete, the scaffolding protein is expelled intact or digested by a special protease that is also incorporated into the procapsid. Procapsids appear spherical and usually have a smaller diameter than mature capsids. When DNA is packaged into the procapsid, the capsid usually expands and changes shape, transforming into the typical angular, icosahedral shape of mature capsids. If present, decoration proteins add to the outside of the capsid after it expands and help

to stabilize the structure. Once DNA is packaged, the capsid is made ready to join to a tail by the addition of proteins that bind to the portal.

Lysis

Lysis of a typical *Caudovirales* host usually requires the breaching of the bacterial envelope, which usually requires separate factors for each of the envelope's layers. For example, in gram negative hosts, holins are used to breach the inner membrane, a lysozyme or endolysin is used to break down the peptidoglycan intermediate layer, and spannins are used to breach the outer membrane. The endolysin is a soluble enzyme with the capacity to break the bonds holding the host cell wall together. The endolysin molecules accumulate within the host cytoplasm during the late stages of infection, but are unable to attack the cell wall because they are sequestered within the cytoplasmic membrane. Holin proteins allow the endolysin to get across the cytoplasmic membrane by forming holes in the inner membrane. Holins are synthesized and inserted into the membrane in a form that does not form holes at first. As the bacteriophage infection proceeds, the holins accumulate in the membrane until a predetermined time at which they suddenly and catastrophically induce membrane breakdown. The membrane holes produced by holins allow the endolysin molecules to attack the bacterial cell wall, causing rapid cell lysis and releasing the progeny *Caudovirales* from the cell. Spannins usually work as a pair (one each for the inner and outer membranes) and appear to disrupt the outer membrane by mediating fusions between the inner and outer membranes.

Genomes and Genomics

Chromosome Diversity and Replication

The chromosomes that are packaged into the capsids of the *Caudovirales* are linear dsDNA. *Caudovirales* virions always contain an entire genome or, alternatively more than an entire genome to ensure that a complete genome is packaged into each and that every particle can initiate an infection. To achieve this, a site-specific mechanism may be used, in which DNA packaging begins and ends at defined sites that are recognized by the packaging machinery to create exactly unit genome length DNA chromosomes. Alternatively, the amount of virion DNA may be regulated by a head-full packaging mechanism which fills a capsid that has the capacity to hold slightly more than one genome's worth of DNA. The head-full-packaged chromosomes have the same sequence at each end and are said to be terminally redundant.

After injection, the phage genomes often rearrange to form a circular chromosome that is the primary replicative form of the genome of many dsDNA phages, however, many other phages replicate without forming such DNA circles. *Caudovirales* chromosome replication often takes place bi-directionally from a single origin, but more complex replication schemes and multiple replication origins have also been observed. Late in infection, most phages switch to a mode of replication that produces DNA concatemers, or long strings of genome copies joined end to end, either by a rolling circle mode of replication from circular chromosomes, or by recombination between multiple overlapping genome copies. Such DNA concatemers, whether linear or branched, are the forms of the genome that are the usual substrate for DNA packaging for both head-full and site specific mechanisms. When it does occur, circularization of the phage chromosome takes place via one of two mechanisms. The first is by the annealing of complementary single-stranded DNA sticky ends left by the packaging enzymes, in the cases where the packaged chromosomes have defined endpoints. The second is by a recombination-like mechanism that joins the complementary regions of the overlapping, terminally redundant chromosome ends to form circular DNA molecules. Circular chromosomes are the most common form for bacterial chromosomes, but some bacteria also have linear chromosomes, and a small subset of the *Caudovirales* also form linear chromosomes that replicate as linear plasmids when in their lysogenic form.

Diversity in Genome Size and Organization

The genomes of *Caudovirales* range in size from ~20,000 base pairs in phage ϕ 29, to ~170,000 base pairs in phage T4, and up to ~500,000 base pairs in other known examples. The smaller phages have proportionally fewer genes than the large phages, but are functional and successful phages none the less. Given this wide range of sizes and diversity of the *Caudovirales*, it is not surprising that there is not a common conserved genome structure for this order of viruses. Within a genus of phages there is often a common genetic structure, but within and across families there is often little resemblance in genome organization. There are notable exceptions to these generalizations, for example, the genus "P22-like viruses" and the genus " λ -like viruses" have rather closely related genomes.

Common Themes in Genome Structure

Despite the differences mentioned, there are many common general features in *Caudovirales* genomes. The genes for the structural proteins, such as those for capsids, tails, or tail fibers, tend to be found clustered together, and within these clusters, the genes that encode proteins that physically interact with each other also tend to be grouped together. For example, the capsid protein genes for

the portal, the scaffolding protein and the major capsid protein usually occur together and in the order listed. Sets of late genes are often grouped together in clusters and, in some cases, the entire set of structural genes are grouped together in clusters and transcribed together from a single promoter, as is the case for phage λ .

Horizontal Exchange of Genes is Widespread

A large number of complete bacteriophage genomes have been fully sequenced, and thousands of these have been deposited in GenBank and other databases. In addition, there are also many temperate phages residing in the genomes of bacterial chromosomes deposited in GenBank, some defective (or cryptic), some complete and able to form viable phage. Analyses of the sequences of these genomes have shown that many of the genes of the *Caudovirales* have highly similar counterparts in other genomes and in many cases the similarity can be inferred to be truly homologous. An important conclusion of these analyses is that there is extensive horizontal exchange of genes between phages that are not within the same species or genus or order. It appears that genetic exchange between phages by both homologous and non-homologous recombination mechanisms is widespread and that many phages are mosaic combinations of genes found elsewhere. Many of the homologous genes are within the *Caudovirales*, but some are from outside this group. In some cases, large modules of genes in a pair of phages are closely related, for example, the head genes of phage HK97 (*Siphoviridae*) and phage SfV (*Myoviridae*), despite belonging to different taxonomic groups. In other cases, a single phage tail fiber encoding gene in one phage may contain several distinct regions of homology with several other different phages.

Common Ancestry

The exact nature of the evolutionary relationships among phages and how it relates to phage taxonomy is a controversial subject, but at the level of individual protein coding genes, it is fairly certain that proteins with a high degree of sequence similarity in analogous proteins almost certainly share a common evolutionary ancestor. The power and utility of bioinformatics to tease out weak (sequence) similarities between distantly related proteins gives the evolutionary scientist powerful tools to detect relationships that escape casual examination. However, the reliance on sequence similarity matches to detect homologous relationships breaks down when homologous proteins have diverged so far that sequence similarity is undetectable. A notable case is that of the major capsid proteins of bacteriophages and other viruses. The 3-dimensional structure and protein fold of the major capsid protein of phage HK97 was determined to high resolution by x-ray crystallography. Sensitive bioinformatic techniques were able to detect weak sequence similarities between the HK97 capsid protein sequence and the capsid proteins of a large number of other *Caudovirales*, suggesting that the HK97 capsid protein fold is quite common. Subsequently, the determination of the protein folds of phages T4 (*Myoviridae*), P22, ϕ 29, and epsilon 15 (*Podoviridae*) and many other phages, by structural methods has shown that all of these phages also have the same protein fold as HK97, despite the lack of any detectable sequence similarity. This structural similarity also extends to the capsid proteins of the Herpesviruses, which use a variant of the HK97 fold for their the floor domains of their much larger major capsid proteins and also have a portal protein with fold similar to their bacteriophage homologs.

Further Reading

- Cairns, J., Stent, G.S., Watson, J.D. (Eds.), 1992. *Phage and the Origins of Molecular Biology*, expanded ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Calendar, R. (Ed.), 2006. *The Bacteriophages*, second ed. Oxford: Oxford University Press.
- Casjens, S.R., 2011. The DNA-packaging nanomotor of tailed bacteriophages. *Nature Reviews Microbiology* 9 (9), 647–657.
- Caspar, D.L., Klug, A., 1962. Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposium on Quantitative Biology* 27, 1–24.
- Chaikerasitak, V., Nguyen, K., Khanna, K., *et al.*, 2017. Assembly of a nucleus-like structure during viral replication in bacteria. *Science* 355 (6321), 194–197.
- Duda, R.L., Teschke, C.M., 2019. The amazing HK97 fold: Versatile results of modest differences. *Current Opinion in Virology* 36, 9–15.
- Hatfull, G.F., Hendrix, R.W., 2011. Bacteriophages and their genomes. *Current Opinion in Virology* 1 (4), 298–303.
- Hendrix, R.W., Casjens, S., 1988. Control mechanisms in dsDNA bacteriophage assembly. In: Calendar, R. (Ed.), *The Bacteriophages*. NY: Plenum Press.
- Hendrix, R.W., Roberts, J.W., Stahl, F.W., Weisberg, R.A., 1983. *Lambda II*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2015. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proceedings of the National Academy of Sciences of the United States of America* 112 (35), E4919–E4928.
- Hua, J., Huet, A., Lopez, C.A., *et al.*, 2017. Capsids and genomes of jumbo-sized bacteriophages reveal the evolutionary reach of the HK97 fold. *mBio* 8 (5).
- Karam, J.D., Drake, J.W., Kreuzer, K.N., *et al.* (Eds.), 1994. *Molecular Biology of Bacteriophage T4*. Washington, DC: ASM Press.
- Katsura, I., 1990. Mechanism of length determination in bacteriophage lambda tails. *Advances in Biophysics* 26, 1–18.
- Leiman, P.G., Kanamaru, S., Mesyanzhinov, V.V., Arisaka, F., Rossmann, M.G., 2003. Structure and morphogenesis of bacteriophage T4. *Cellular and Molecular Life Sciences* 60 (11), 2356–2370.
- Rossmann, M.G., Rao, V.B. (Eds.), 2012. *Viral Molecular Machines*. NY: Springer.
- Xu, J., Hendrix, R.W., Duda, R.L., 2014. Chaperone-protein interactions that mediate assembly of the bacteriophage lambda tail to the correct length. *Journal of Molecular Biology* 426 (5), 1004–1018.
- Young, R., 2014. Phage lysis: Three steps, three choices, one outcome. *Journal of Microbiology* 52 (3), 243–258.

Relevant Websites

<http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=28883>

Complete genomes: Caudovirales
NCBI.

<https://viralzone.expasy.org/256>

Phage index ~ ViralZone page.

<https://phagesdb.org/phages/>

The Actinobacteriophage Database.

<http://www.virology.ws>

Virology blog.

Helical and Filamentous Phages

Andreas Kuhn, University of Hohenheim, Stuttgart, Germany

Sebastian Leptihn, Zhejiang University-Edinburgh University Institute, Zhejiang University, Haining, China

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Adsorption Binding of phage particles to the receptor on the bacterial cell surface.

Atomic force microscopy (AFM) High resolution scanning microscope with a laser-controlled cantilever.

Conjugation Sexual contact of bacteria to exchange genetic information from the donor to the recipient cell involving a F-pilus of the donor cell.

Extremophilic bacteria Bacteria that live at high (thermophilic) or very low (psychrophilic) temperatures.

Filamentous Morphology of the phage that defines the *Inoviridae*. At the proximal end 5 copies of the gp3 adsorb to the tip of a F-pilus of the host cell.

Membrane insertion Newly synthesized proteins that are incorporated into the membrane before their assembly into the phage particle.

Morphogenetic signal Sequence on the viral DNA that is responsible to initiate DNA the packaging reaction.

Pathogenicity Factors encoded by the viral DNA that are toxic for the human organism the bacterial host colonizes.

Rolling circle replication Continuous replication mode to generate the single stranded phage DNA to be packaged to the progeny particles.

Introduction

The first filamentous phage was isolated 60 years ago in Germany, when Hartmut Hoffmann-Berling took a sample from the backyard of a cattle farm close to Heidelberg. In the lab, he tested the sample for plaque formation on *Escherichia coli* Hfr. Since the selected phage contained DNA and appeared male-specific it was named “fd”. Nearly at the same time, Tim Loeb at the Rockefeller institute isolated the male-specific “f1” phage and Hans-Peter Hofschneider in Munich phage M13, who also showed its filamentous morphology with the electron microscope. The “M” in M13 stands for Munich, and it was Hofschneider’s 13th isolate. Today we know that the three phages, fd, f1 and M13, are nearly identical in their DNA sequence and therefore can be considered as an identical phage (termed Pf phage). The fd genome was one of the first to get determined with 6408 bases. The advantage of single-stranded DNA (ssDNA) for the in vitro replication reaction helped to determine the first sequences in general. Because of this, any DNA was ligated into the M13 genome and after transfection the phage was isolated to purify the ssDNA for sequencing. At this time, M13 was the common vehicle for DNA sequencing.

In the early work of Hoffmann-Berling it was recognized that the fd phage was “liberated” from cells without lysis, in an unknown “new phage liberation mechanism”. In addition, the first studies of their morphogenesis revealed no visible cytoplasmic intermediate structures as had been observed with many other phages. Experiments by Rolf Knippers to assemble the phage *in vitro* by mixing the ssDNA with the major coat protein gp8 were unsuccessful. Instead, as we know today, the phage is assembled in the membrane during its secretion and requires the membrane to coat the ssDNA in an ordered secretion process and involving a motor protein complex in the membrane. The assembly and secretion of the phage requires the catalytic proteins gp1, gp4, gp11 and thioredoxin of the host.

The filamentous phage again became famous when Johnny Mekalanos discovered that the cholera toxin gene was embedded in an integrated phage CTX-(ϕ) genome. Following this observation, many more genome-integrated cryptic filamentous phages were detected in *Vibrio cholerae*. In fact, there might be a huge number of integrated filamentous phages in the metagenome which are not yet identified. Virulent filamentous phage have been found in *Pseudomonas* (Pf1 to Pf7), *Xanthomonas* (Xf, Cf, Lf, Xv2), *Salmonella* (Lineavirus lke), *Ralstonia* (*Habenivirus* RSM1 to 3, RSS1, RS603), *Vibrio* (*Fibrovirus* fs1, *Saetivirus* fs2), and in *Spiroplasma* (*Plectovirus* 1-C74, *Vesperitiliovirus* 1-R8A2B) and recently also in *Thermus thermophilus*.

Phylogenetic studies show that gene 1 is the most conserved among the known filamentous phage genes and can be taken as a basis for a phylogenetic tree. This analysis suggests that there are several distinct clades, where the *Inoviridae* group together with *Lineaviridae* and *Saetiviridae* are distinct from the known *Vibrio* and *Pseudomonas* phages.

Similar in structure but genetically unrelated to the filamentous phage are the helical phages that were found in archaea. The *Lipothrixviridae* and *Rudiviridae* infect *Crenarchaea* and are 400–2200 nm long tubular particles with a diameter of about 20 nm. They contain linear dsDNA of 16–60 kbp that is packed with a helical capsid protein. Another helical archaeovirus, termed *Clavavirus* has recently been discovered that contains a circular dsDNA of 5.3 kbp. No sequence homology has been found in the current data base. The diameter of the virus is 16 nm with a length of only 143 nm.

Biotopes: Prolific, Temperate and Pathogenic

The discovery of filamentous phages was delayed and still is hampered by a crucial technical step to remove larger debris including bacterial cells from aqueous samples by employing porous membranes or filters. Due to this step, spheroidal particles are more likely obtained than elongated ones, thus leading to a removal of filamentous phages from most samples. Therefore, it was estimated that filamentous phage representatives are by numbers only about 5% of those of spherical or head-and-tail phages. In contrast to this estimate, a recent publication in the field of (meta-) genomic sequencing indicated that there are many more filamentous phages than previously thought. According to the study, the currently described 56 members of the *Inoviridae* family represent only a small fraction of a highly diverse group that is proposed to even extend to the archaea kingdom. Astonishingly, *Inovirus* sequences have been identified in 35% of the metagenome sequences available until 2019. About 6% of the annotated genomes from bacteria and archaea contain filamentous phage genes. These sequences cover only the ones that are integrated into the host genome. Therefore, it can be assumed that many more filamentous phages exist often in a plasmid like state, as bacterial viruses are found as an episome or chromosomally integrated.

Not only their DNA, but the filamentous phage particles themselves have been discovered in all biotopes on the planet. Most of the ecosystems where filamentous phages can be found are moderate in nature and are inhabited by mesophiles, i.e., environments with “normal” temperatures and pH, such as the gut of mammals including humans, where they infect *proteobacteria*, including *enterobacteria* (an estimated 90%). Surprising is the partnership that some filamentous phages have entered. To the benefit of the phage which uses the cellular replication machinery, the bacterial host itself has been shown in some cases to obtain an evolutionary advantage by being infected with the virus. Many of the filamentous viruses have a parasitic lifestyle, living on the expense of the host cell. However, thus in some cases a mutualistic relationship is discussed as filamentous virus was found to influence certain host characteristics to the benefit of both, the bacterium and the phage. The most prominent example is the bacteriophage CTX- ϕ that infects *Vibrio cholerae*, and delivers the genes for the cholera toxin to a comparably harmless bacterium. Once integrated into the genome, cholera toxin A and B are produced, leading to a destruction of the mucosal membranes resulting in a high influx of water, causing diarrhea. Release of the phage (excision) is triggered by a drop in pH, as found in the intestines of a human host, thus allowing other bacteria to be infected by the phage, turning the infected bacteria into highly virulent ones, in a process called phage conversion, resulting in a so-called toxigenic *Vibrio* strain. It is thought that the cholera toxin genes have been acquired by an evolutionary predecessor of the CTX- ϕ phage, giving the virus a unique advantage as the spreading of a highly virulent bacterial host is also beneficial for the phage. CTX- ϕ is not the only *Vibrio* phage; there is a vast variety of phages most of which are genome-integrated such as VGJ- ϕ which integrates into the same sites as CTX- ϕ . Interesting is the occurrence of prophages with incomplete genomes, also known as satellite phages. These phages piggyback on some of the genes of their relatives, leading to the formation of so-called mosaic phages. An example is RS1 which has been reported to use the morphogenesis protein gp1 of CTX- ϕ in order to form phage particles.

Yersinia pestis is another example where a filamentous phage contributes to the pathogenicity of a bacterium. *Y. pestis* evolved from the enteropathogen *Yersinia pseudotuberculosis* about three millennia ago and became infected by the filamentous phage Ypf-Phi, which remained an extrachromosomal element in many lineages. The *Y. pestis* strain that contained the phage integrated into the host genome, caused the third plague epidemic. This major bubonic plague pandemic began in China in 1855 and spread to all inhabited continents, killing more than 12 million in China and India alone. Interestingly, the phage Ypf- ϕ infects *Y. pestis* but also *Y. pseudotuberculosis* and *Yersinia enterocolitica*. In addition, the transduction is also expanded to *E. coli* strains. In contrast to being genome integrated in *Y. enterocolitica*, the phage exists as an extrachromosomal plasmid in *Y. pseudotuberculosis* and in *E. coli*. The presence of the phage confers better colonization abilities of the host in mammalian tissues. While Ypf- ϕ is not a major virulence factor, it confers a higher fitness and an increase in pathogenicity to the bacterium, albeit modest.

Some commensal bacteria become opportunistic pathogens upon the genetic acquisition of virulence factors. *E. coli* is no exception to this phenomenon where the enterobacterium causes diarrhea, urinary tract infections but also sepsis, and meningitis. For example, the *E. coli* strain O18:K1:H7 is one of the leading causes of neonatal meningitis and cystitis in North America. The serotype CO92 carries a genetic element called CUS-1 which was identified to be a filamentous prophage. The genome of the phage is almost identical to CUS-2 found in *Y. pestis* biovar orientalis. Both contain a gene called *puvA* which, when destroyed, affects the ability of *E. coli* K1 strains to systemically propagate or even survive *in vivo*, indicating a role in pathogenesis. The gene is flanked by several other genes that show high similarity to those coding for coat proteins and proteins crucial for morphogenesis of other filamentous phages. Filamentous phage particles much like M13 were detected in the supernatant of cultures from both, *E. coli* K1 as well as *Y. pestis* biovar orientalis CO92 cells, later called CUS- ϕ .

Another filamentous phage, MDA Φ , which infects *Neisseria meningitidis*, has been implicated to have a contribution to the virulence of the host when the virus is in a prophage state. While it was shown that the presence of the phage has no influence on virulence during the bloodstream stage of *N. meningitidis*, it has been demonstrated that phage-infected bacteria are more efficient in colonizing host epithelial cells during the progression of the disease.

While being directly responsible for the pathogenicity of their host, filamentous phages can affect the immune system of a human directly, when infected by bacteria harboring such a virus. In case of the *Pseudomonas aeruginosa* phage Pf, the co-existence of the phage in their host impacted the healing process negatively, leading to chronic, badly and slow-healing wounds. In the mouse model, 50 times fewer *Pseudomonas* cells were required to cause a wound infection, when comparing phage-infected to non-infected cells. Also, the rate of morbidity was increased when mice were infected with *Pseudomonas* harboring phage. Surprisingly, the effects were caused not by increased virulence of the bacteria but by the immune response of the mammalian host. The uptake

of phage particles by macrophages caused among other cellular reactions, a Toll-like receptor response, mediated by the mRNA which is transcribed from the phage DNA, as well as to the inhibition of phagocytosis, suggesting that phage can play a direct pathogenic role in the infection process.

Filamentous phages can be found on plants, in the soil, on food products and in lakes, rivers and the sea. Also in extreme environments where extremophilic bacteria are present, filamentous phages have been discovered. Examples are volcanic hot springs that are found all over the world, including in Japan, where the phage (ϕ)-OH3 has been discovered together with its host *Thermus thermophilus*. The conditions where host and virus thrive are around 70°C with an almost neutral pH and low osmolarity. The stability and infectivity of the filamentous phage reflect this environment, as a higher temperature, high salt or low and high pH are detrimental and lead to denaturation and inactivation of the virus. While it is not known whether the phage confers any benefit to the host, or is purely parasitic, the existence of a thermostable phage might possibly allow nanotechnological applications under elevated temperatures.

On the other temperature extreme, cold-adapted microorganisms were also found to harbor filamentous phages. A psychrophilic bacterium from the *Pseudoalteromonas* group which thrives in the cold waters and in the ice of the arctic, can be infected by a filamentous phage called f327. Prokaryotes in the arctic ice and water are exposed to very low temperatures and large fluctuations in nutrient levels, salt and gas concentrations, as well as in pH. Interestingly, approximately 50% of the species living in permanent arctic ice belong to the genus *Pseudoalteromonas* of which more than 30% were found to contain filamentous phage genes. As one would expect from a parasite, phage infection resulted in slower growth rates and decreased final cell density. In addition, infected bacteria were also less tolerant towards changes in salinity or higher hydrogen peroxide concentrations compared to uninfected cells. Surprisingly, infected bacteria showed a much higher mobility, contrasting the negative impacts on cell viability mentioned above. Possibly, the increased mobility allows to reach nutrient enriched environments faster, e.g., when released from the ice at higher temperatures. While being psychrophilic, *Shewanella piezophila* is also a piezophile, i.e., a pressure-loving (or -tolerant) bacterium which lives on deep ocean floor, with optimal growth at a pressure of 20 MPa and a temperature of 20°C. The *S. piezophila* strain WP3 has been found to contain a genome-integrated filamentous phage, called SW1. SW1 is significantly induced at cold temperatures around 4°C, the typical temperature in the deep sea. The presence of the phage strongly impacts the expression of almost 50 host genes, many of them involved in flagella production. Interestingly, the phage does not influence the swimming motility of the host, while uninfected cells exhibit an increase in swarming motility at 4°C. Because motility is energetically expensive for cells, it is thought that the phage reduces this energetic cost by suppressing flagella genes and thus increases chance of survival of the host.

While the above-mentioned hosts are all Gram-negative bacteria, filamentous phage-like particles have also been observed to infect *Clostridium acetobutylicum*, a Gram-positive. So far, only one filamentous phage which infects a Gram-positive bacterium has been characterized in detail with regards to genome and proteins. The phage B5 uses *Propionibacterium freudenreichii* as a host to multiply, which is an important microbial component during the maturation of certain cheeses, in particular Emmental, where the carbon dioxide is responsible for forming the characteristic large holes. In contrast to contaminations, especially lytic phages can cause problems in industrial food processes where microorganisms are involved. However, B5 infection does not negatively influence the product. Whether or not this phage changes any pathways in the bacterial host, is currently not known.

Until today, no filamentous phages have been discovered that infect the third kingdom next to eukarya and bacteria, the archaea. However, a recently published metagenomic study revealed sequences of Inovirus-like genes with archaeal genomes that show highly conserved sequences of filamentous phages. In addition, a circular plasmid like episome of a filamentous virus has been isolated from *Methanobolus profundus*. The discovery of filamentous viruses in archaea is fascinating from an evolutionary perspective, as none of the currently characterized viruses infecting archaea are related to those of bacteria. Finding *Inovirus* genomes in both, archaea and bacteria, could be explained by the idea that this family existed very early on with a virus infecting a common ancestor. This conclusion is supported by the fact that the archaeal viruses display a high genetic diversity and lack strong similarity to bacterial phages, thus are likely not to result from a recent genomic exchange.

Structure of Filamentous Phages

Filamentous phages are structurally very different from other phages that usually have head and tail. Instead, they are long flexible filaments with two different ends, the base end which adsorbs to the host and the tip which is the end that leaves the host cell first during the secretion event. Based on early fiber diffraction methods, the filamentous phages were divided into two structural classes, where M13, f1 and fd are members of class 1 on which this review is focused. The fiber diffraction and other spectroscopic techniques showed the dimensions of the phage as a rod of about 890 Å in length and with a 6 Å diameter. The phage coat consists of an interdigitated helical arrangement of about 2750 copies of the major coat protein gp8. The slightly curved α -helical proteins cover the ssDNA with an about 20° tilt along the helical axis of the phage particle. The ssDNA is packed as two individual antiparallel strands within the protein tube where the phosphodiester backbone of the DNA may be electrostatically connected to the positively charged residues at the C-terminal of gp8. The helical arrangement of the coat proteins describe a C₅S₂ symmetry which states that 5 proteins start at the same time leading to a five-fold rotation axis. The coat proteins overlap such that the negatively charged N-terminal regions are exposed on the outside of the particle and the coat actually consists of two α -helical layers.

The two ends of the phage consist of the minor coat proteins. The base of the phage which is used to adsorb on the host cell is made of five copies each of gp3 and gp6. The two proteins form a defined structure visible as knobs with a diameter of about 50 Å

on the phage particles. The tip of the phage is formed by two small hydrophobic proteins gp7 and gp9, also with about five copies each. These two proteins are bound to the intergenic region of the ssDNA and initiate the packaging event.

Infection Process

The filamentous phages M13, fd and f1 are male-specific and infect only bacteria that have a F plasmid encoding the F-pilus. The central D2 domain of the “attachment protein” gp3 interacts with the tip of the F-pilus whereas the C-terminal D3 domain of gp3 is anchored to the phage capsid. The binding of the phage induces a retraction of the pilus similar as the binding of a recipient female cell does. The retraction brings the adsorbed phage close to the outer membrane surface and possibly also into the periplasm where it interacts with the TolQRA, an inner membrane protein complex with two large periplasmic domains. However, it is not known how the infecting phage actually crosses the outer membrane, but this might occur together with the F-pilus retraction process. In the periplasm, the C-terminal domain of TolA interacts with the N-terminal D1 domain of gp3. This binding is thought to trigger a conformational change in the C-terminal domain of gp3 and supports the insertion of gp3 into the inner membrane, possibly with help from TolQRA. Since the Tol complex is anchored in the inner membrane and spans the periplasm, it might offer a path for the infecting phage from the outer to the inner membrane. There, the phage particle is dissociated and the coat proteins partition into the inner membrane. In the inner membrane, it is discussed that the 5 copies of gp3 form a kind of DNA channel, possibly together with TolR, that allows the single stranded DNA to enter the bacterial cytoplasm.

Replication by the Rolling Circle Mechanism

Soon after the infecting phage DNA has entered the cytoplasm of the host it is a target of the DNA polymerase III which converts the single stranded DNA into a double strand. First, *E. coli* RNA polymerase binds to the intragenic region between *gene 4* and *gene 2* (b 5484–5987) to synthesize a short 30 base primer starting from base 5756. The primer is then extended with deoxyribonucleotides by DNA polymerase III and the RNA primer is replaced by deoxynucleotides catalyzed by DNA polymerase I of *E. coli*. DNA ligase of the host seals the ends of the double stranded DNA into a closed circle, termed replicative form RF I. This first stage involves no viral protein and the generation of the parental RF I is independent on the translation and expression of phage proteins.

DNA replication is then initiated by gp2, a soluble globular protein of 46 kDa. It has an endonucleolytic activity, attacks the parental RF DNA at a specific site and cleaves the (+) strand between the bases 5781 and 5782 in the intragenic region. Most likely, gp2 recognizes two hairpin structures (5692–5791) in the (+) strand and provides a free 5' end (RF II). In the early phase of infection *E. coli* polymerase III replicates the circular RF to about 100 copies per cell. Later in infection, the polymerization of the (–) strand is prevented by binding gp5, a single strand binding protein. As a result of this “rolling circle” only (+) strands decorated with gp5 are generated in the cells. Besides generating the endonucleolytic cut, gp2 has a second function in producing unit length ssDNA and sealing the 5' and 3' ends of the ssDNA-gp5 complex together to a closed circular DNA molecule. It is unknown whether also gp10 is involved in this process, since in the absence of gp10 the proliferation of phage is greatly reduced.

The ssDNA-gp5 complex has been thoroughly studied by many laboratories. gp5 is a 9.7 kDa protein and is mainly present as a dimer. The globular protein consists of numerous β sheets that fold into anti-parallel loops and expose positively charged residues. These residues are involved in the binding two strands of ssDNA building a 3-dimensional helical structure which has been modeled. In this model, two strands of the circular ssDNA are packed into a left-handed helix that is surrounded by gp5 dimers. The binding of gp5 to ssDNA occurs cooperatively to the phosphate backbone, but a phenylalanine and tyrosine residues are involved in contacting bases of the bound ssDNA.

Biosynthesis of M13 (fd, f1)

The Coat Proteins

Transcription of the filamentous DNA occurs early after the replicative RF I form has been generated. Early studies have shown that M13 is transcribed from the non-viral strand in 8 RNA species ranging from 8S to 20S. They are mostly generated from individual promoters that are recognized by the *E. coli* RNA polymerase holoenzyme. A strong rho-independent termination site is located shortly after the end of *gene 8* forming a tight hairpin loop structure and a rho-dependent termination site proximal to *gene 1*.

Translation of the 11 viral genes occurs shortly after infection by the host ribosomes. The ribosome binding sites (RBS) are located in a short distance before the start codon of *genes 1* to *9*, where the RBS for gp7 is located within *gene 5*, the RBS for gp6 is within *gene 3* and the RBS for gp1 is within *gene 6*. For gp9 the start codon overlaps with the stop codon of gp7. Interestingly, the coding sequences for gp1 and gp4 overlap with 7 codons and differ by their reading frame. *Gene 10* is located within *gene 2* in the same reading frame and also *gene 11* starts within *gene 1* in the same frame. Therefore, gp10 includes the same amino acid sequence as gp2 from residue 166–276 and gp11 is identical with gp1 from residue 241–348. This shows that the coding sequences of M13 are tightly packed and only the 503 bp intergenic region lacks any protein-coding information.

The Major Coat Protein gp8

Essential for the assembly of the filamentous phage is the expression of the 5 coat proteins gp3, 6, 7, 8, 9. They all are inserted into the membrane where the assembly of the phage particles occurs. gp8 is the major coat protein with more than 2500 copies per particle. It is first synthesized with a signal sequence of 23 amino acids that is cleaved off after membrane insertion. Leader peptidase (LepB) recognizes the transmembrane procoat proteins at the residues -1, -3 and -6 in the signal sequence following the “*von Heijne rules*”. The membrane insertion of procoat requires an essential membrane protein of the host, the membrane insertase YidC. YidC binds to the hydrophobic sequences of procoat with its hydrophobic slide structure leading to the integration of the procoat protein across the membrane. After cleavage by leader peptidase, the 50 amino acid long M13 coat protein exposes the N-terminal first 20 residues into the periplasmic space and the C-terminal 10 residues into the cytoplasm.

The exact mechanism of membrane insertion has been studied with the major coat protein of Pf3 *in vivo* and *in vitro*. This protein lacks a signal sequence but also requires YidC for membrane insertion. The Pf3 coat protein has been purified and added to YidC containing proteoliposomes which inserted the coat protein efficiently. Atto-labeled mutants of Pf3 and YidC allowed the calculation of the speed the membrane insertion reaction takes by single molecule Förster resonance energy transfer (smFRET) and is in the range of 10 ms. Disulfide cross-linking showed that the coat protein enters YidC step-by-step up the hydrophobic slide of YidC and the membrane potential dependent translocation step then moves the N-terminal domain of Pf3 coat from the hydrophilic groove inside YidC into the periplasm.

In the membrane, the M13 coat proteins oligomerize as shown by disulfide cross-linking. Possibly, these oligomeric forms interact with the gp1/11 packaging machine and cover the single stranded DNA. The four positively charged amino acid residues in the C-terminal tail of the coat protein might electrostatically interact with the negatively charged phospholipids of the DNA. When the number of the lysines was reduced the length of the phage particles increased correlating with the altered charge density. On the phage particle, gp8 interacts with neighboring gp8 by hydrophobic interaction in a tilted staggered arrangement (see section on the phage particle structure). The N-terminal region of the coat protein is negatively charged but is quite flexible for alterations. Antigenic peptides and display peptides have been incorporated into this region and were exposed on the phage.

The Minor Coat Proteins gp7 and gp9

The assembly reaction is initiated with the capping proteins gp7 and gp9. Both proteins are synthesized without a signal sequence but fractionate with the inner membrane. gp7 is a small 33 amino acid long single-spanning membrane protein. Its N-terminal region of 8 residues includes 3 negatively charged residues, whereas the C-terminal residue is an arginine residue. For the display of foreign peptides the PelB signal sequence was added to support the membrane insertion of gp7. Even in the absence of an added signal sequence gp7 was useful for phage display.

Gp9 has only 32 residues and is the smallest coat protein. The C-terminal region of 15 residues is positively charged and in the cytoplasm. It is unclear whether the N-terminus is exposed in the periplasm since there is no hydrophilic residue present. When the N-terminus was extended by 20 polar residues encompassing a T7 tag the protein was fully functional and complemented an M13am9 infection. The membrane insertion of this extended gp9 was dependent on the membrane insertase YidC and after insertion the N-terminal extension was digestible by proteinase K. It is possible that the non-extended gp9 protein inserts also spontaneously as was shown with artificial membranes. Therefore, gp9 can be used for phage display and if it is applied with a second display site at the base end of the phage particles (e.g., gp3) bi-functional phage can be designed.

Gp7 and gp9 presumably interact with the hairpin packaging signal to initiate DNA packaging at the membrane surface. Mutants which had the packaging signal deleted were compensated by plaque-forming phage mutations within genes 7, 9 and 1. gp7 can be co-immunoprecipitated with gp8 which suggests that once the formation of packaging complex is initiated the major coat protein is fed into the packaging machinery for assembly.

The Minor Coat Proteins gp3 and gp6

As the packaging process is approaching the end of the single stranded DNA circle, gp3 and 6 bind and terminate the phage assembly. gp3 is first synthesized with a signal sequence of 18 amino acid residues and inserted by the Sec translocase involving SecA. After cleavage it is a single-spanning membrane protein and has a large N-terminal domain of 378 amino acids in the periplasm and a cytoplasmic region of 5 residues. The membrane anchoring of gp3 has been studied in detail. When the length of the membrane anchor was reduced from 23 residues to 17 the topology was unchanged. Larger deletions, however, resulted in a periplasmic location. The periplasmic region of gp3 contains two β structured domains, termed D1 and D2. The two domains (later named as N1 and N2) are connected by a flexible hinge region that is functional during the phage infection process. Whereas N1 (residues 1–68) folds and unfolds rapidly N2 (residues 124–217) folding is slow followed by the folding of the hinge region and the docking of the two domains tightly together. This tight folding ensures that the phage particle is stable in changing environments. Following the two N-domains the periplasmic portion of gp3 contains a C1 and a C2 domain. Whereas C1 is important for phage stability, C2 is involved in the release of the secreted nascent phage progeny.

Gp3 has been extensively used for phage display technology. The random sequences encoding the display peptides are introduced between the signal peptide and the mature sequence of gp3. Therefore, the peptides are displayed on the phage at the very N-terminus not affecting the two β structured domains N1 and N2.

At the base of the phage filament gp3 is connected to gp6, a small very hydrophobic protein of 112 amino acid residues. From its sequence it is predicted to span the membrane three times with the C-terminus in the cytoplasm. In the phage, gp6 is required for the termination of the assembly reaction. Amber mutations cause the generation of polyphage. Based on this observation and the stability of the polyphage it has been concluded that gp6 first binds to the end of the nascent phage and gp3 then follows. Also gp6 has been used for phage display with extensions at its C-terminus.

Assembly and Secretion

The assembly of filamentous phage is initiated at the “morphogenetic signal” on the replicated M13 single stranded DNA (bases 5550–5577) which is a double-stranded hairpin structure close to the 5' end of the replicated ssDNA. The tip of the hairpin a GCA-loop might be the actual structure of the DNA that interacts with the membrane inserted gp7/gp9 initiating the assembly reaction. In the assembled phage, gp7 and gp9 that are located at the tip of the particle and are present with about 5 copies each, together forming the phage cap structure.

The membrane based assembly process is controlled by gp1 and its shorter version gp11 which starts at an internal gp1 codon for the methionine residue 241. The two proteins form a common oligomeric “preinitiation complex” in the membrane together with gp4 that is localized in the outer membrane. gp1 has a large N-terminal domain in the cytoplasm that contains an ATPase binding site with the typical Walker box elements. Mutations within these elements inhibit phage propagation suggesting that ATP hydrolysis is required for phage extrusion. Close to the membrane spanning regions of gp1 and gp11 are a characteristic pattern of positively charged residues similar to the pattern found in gp8. These are presumably involved in the transient binding of the extruding phage DNA molecule. The morphogenetic signal of the phage DNA contacts the gp1/11 complex in the inner membrane together with thioredoxin reductase TrxA of the host converting the preinitiation complex to an “initiation complex”. Since the enzymatic activity of TrxA is not required for phage propagation the contribution of TrxA is likely only structural, similar as it is for phage T7 replication.

The single stranded DNA that interacts with the gp1/11 complex is covered with multiple copies of gp5. The gp5 proteins are continuously displaced from the DNA as it enters the assembly machinery and within the machinery gp5 is replaced by the major coat protein gp8. gp8 exists as oligomers in the membrane and presumably, these are fed into the membrane spanning assembly machine laterally for assembly (Fig. 1). Since in the M13 phage particle the coat proteins are arranged as pentamers in so called five-start helices it is possible that the gp1/11 machinery has five lateral entrances for the gp8 proteins that all five simultaneously bind to the DNA. A model of a cooperative pentameric assembly of gp8 generating the phage has been proposed early on. According to this model, each of the added coat proteins would be followed by the next coat protein in a certain distance leading to the staggered and interdigitated subunit packing of gp8. In the phage, this results in a shingle-like arrangement of the coat proteins covering the particle. Thereby, the nascent filamentous phage particle is drilling itself out of the host cell (Fig. 1).

During the assembly process the growing nascent phage filament passes through the outer membrane with support of gp4. gp4 is a protein of 406 amino acid residues and is synthesized as a precursor of 426 residues. In the outer membrane gp4 is assembled into a 14-meric secretin structure with a central cavity of 6–8 nm forming a gated pore. This diameter allows the passage of the emerging phage particles that have a diameter of about 6 nm. The channel gating of the gp4 secretin is of functional importance since mutants in the respective regions of gp4 lead to a leaky phenotype of the host cells that also leads to an enhanced sensitivity towards antibiotics. In addition, the N-terminal portion of gp4 that is located in the periplasm makes contact to the gp1/11 assembly machine located in the inner membrane possibly forming a continuous structure. In accordance with this, a mutant in gp1 was found to be partially compensated by a mutation in gp4. However, the interaction between gp4 and gp1/11 might be weak since the two proteins are not co-purifying when extracted from cells.

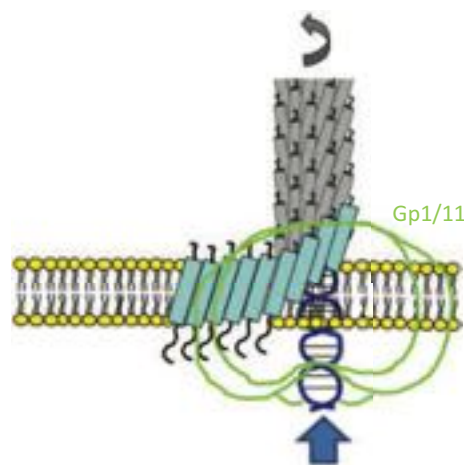


Fig. 1 Model of M13 packaging. The membrane-inserted major coat proteins oligomerize and bind to the ssDNA in a shingle-like fashion catalyzed by the packaging motor gp1/11 and ATP.

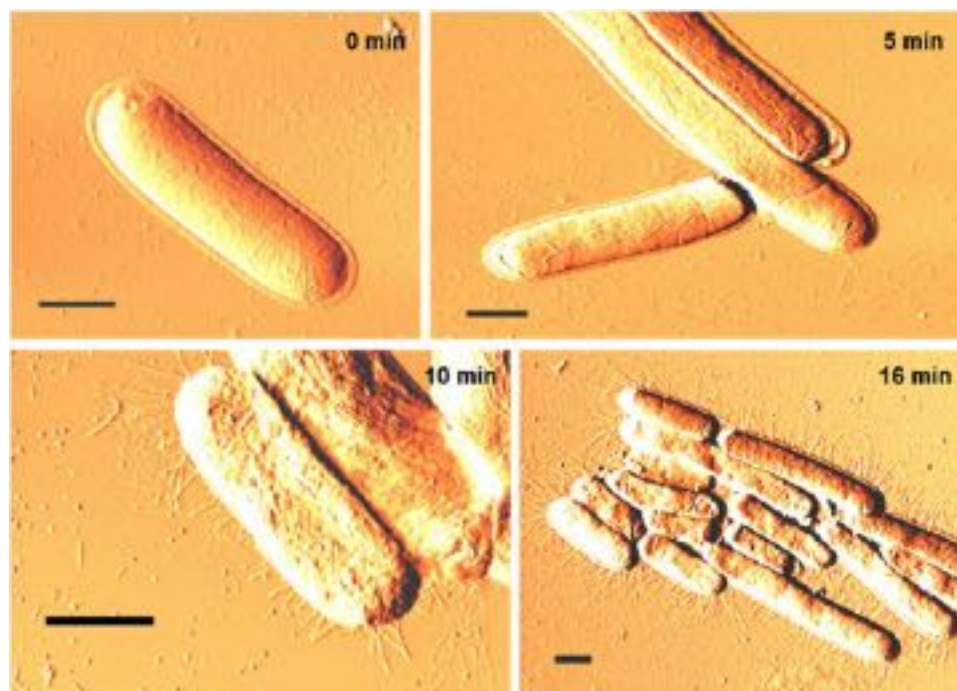


Fig. 2 Infected cells extruding M13 phage particles visualized by the AFM. Samples were taken at 0, 5, 10 and 16 min post infection and recorded with the atomic force microscope. The bar is 1 μm . Courtesy of Martin Ploss.

The extrusion of the filamentous phage was visualized by atomic force microscopy (AFM). Synchronized by the infecting phage, the generation and extrusion of progeny phage from the host cells was observed already after 5 min and could be followed time-dependently reaching a maximum of extruding particles all over the cell surface after 16 min (Fig. 2). Judging from the length of the extruding particles the assembly and secretion is a fast process.

The phage assembly process is terminated by the two proteins gp6 and gp3. Presumably, they both enter the gp1/11 assembly machine when the extruding DNA molecule is entirely covered by gp8. As mentioned above, gp6 is most likely first assembled to the nascent phage followed by the binding of gp3 with its C-terminal region. It has been shown that gp3 and gp6 co-purified when extracted from phage. Mutations in gp3 and gp6 can lead to multiple length particles, also called polyphage. This clearly shows that the two minor coat proteins are involved in the termination step of the assembly. Normally, the size of the phage DNA determines the length of the filamentous phage.

Biotechnological Applications of Filamentous Phage

One of the first applications of M13 phage was its use as a DNA sequencing vector. Since the sequencing reaction *in vitro* is very efficient with single stranded DNA, foreign genes were cloned into M13. The ssDNA was extracted from the phage particles and sequenced with the “universal” oligonucleotide using radiolabeled nucleotides. The universal primer anneals shortly before the multiple cloning sites where the foreign genes were ligated.

Many applications are based on the phage display technology. The coat proteins (namely gp3, 6, 7, 8, 9) can be modified with protein domains that are exposed on the phage surface. These exposed domains can be used as binding components to a plethora of different targets and surfaces. The use of randomized oligonucleotides allows the generation of “display libraries” from which the optimal binders can be selected. Details are described in the recommended “further readings” section.

Due to their unique morphology with a length to diameter ratio of ≈ 130 , filamentous phages can be employed as scaffolds for nanotechnological applications. Many fascinating examples exist where the long and thin structure of the bacterial virus has been made use of to create wire-like structures for various applications. Many groups, especially the research group of MIT scientist Angela Belcher has shown the vast potential of creating virus-based hybrid materials. One such hybrid material consists of a genetically engineered multifunctional M13 phage that binds to carbon nanotubes via the coat protein gp8, which allowed to deposit amorphous FePO_4 in order to achieve high-efficiency electric conductivity. The phage-hybrid material resulted in the reduction of the electron transport from the highly resistive cathode to the conductive elements of the battery. The material displayed a superior performance to any previously engineered amorphous FePO_4 electrode, in terms of power performance and capacity. Filamentous bacteriophage can also be used as nucleation templates for the ambient temperature synthesis of nanowires with phase-changing properties (material capable of changing between two physical states). This is a particularly interesting

application as faster computers can be constructed by drastically reducing the time delays that occur in current architectures from the transfer and storage of information between silicon- and magnetic-based memories (usually milliseconds). Nanosecond time scale information storage can be achieved by specialized *segregating-binary-alloy-type phase-change-materials* (e.g., Ga-Sb-based systems), by employing a single memory structure. In purely inorganic materials, the disadvantages of such materials are the high energy consumption and high temperature-induced elemental segregation. This can be overcome by using filamentous phages as templates: By forming a filamentous phage germanium–tin oxide hybrid material, achieved via an electrostatic binding on the M13 virus coat proteins, wire-like *phase-change-materials* can be obtained. This wire-like structures displayed reliable and controllable phase-changing signatures, with tens of nanoseconds switching times and a great potential for building new and highly advanced computers. Filamentous phage particles have also been used to build a biological metal solar cell arrays that can convert the energy from light with high efficiency into electricity. The M13 phage was used to construct a 3D structure which proved to be highly efficient in light harvesting in specialized dye-sensitized solar cells (DSSCs). Again, the filamentous structure was used to deposit metals onto the phage coat proteins. In this case, gold nanoparticles were used, which in turn were encapsulated with titanium dioxide. This resulted in a plasmon-enhanced nano-wire photoanode. Another fascinating example is the use of M13 to prepare low density 3D aerogels, which are ultralight porous structures with excellent elastic behavior, allowing up to 90% material compression. By using peptide fusions on capsid proteins, inorganic materials can be attached creating inorganic aerogels. Such materials offer a plethora of possible applications, e.g., as bio-scaffolds, for the storage of hydrogen or energy, thermoelectrics, and catalysis. Not using filamentous phages as a wire or wire-network, a novel strategy for biocatalysis for the production of semisynthetic enzymes was developed by the incorporation of an enzymic motif into thermostable M13 phage coat proteins. Examples are phage particles that contain a site modeled after the active site of the enzyme carbonic anhydrase. Filamentous phages can also be employed as nanoparticles for imaging: For near-IR fluorescence imaging of tumors in deep tissues, genetically engineered multifunctional M13 phages were assembled with fluorescent single-walled carbon nanotubes as targeted probes for the specific uptake in prostate tumors.

Further Reading

- Duché, D., Houot, L., 2019. Similarities and differences between colicin and filamentous phage uptake by bacterial cells. *EcoSal Plus*. (ESP-0030-2018).
- Hay, D.H., Lithgow, T., 2019. Filamentous phages: Masters of a microbial sharing economy. *EMBO Reports*. e47427.
- Loh, B., Kuhn, A., Leptihn, S., 2019. The fascinating biology behind phage display: Filamentous Phage Assembly. *Molecular Microbiology* 111, 1132–1138.
- Marvin, D.A., Symmons, M.F., Straus, S.K., 2014. Structure and assembly of filamentous bacteriophages. *Progress in Biophysics and Molecular Biology* 114, 80–122.
- Ptchelkine, D., Gillum, A., Mochizuki, T., *et al.*, 2017. Unique architecture of thermophilic archaeal virus APBV1 and its genome packaging. *Nature Communications* 8, 1436.
- Rakonjac, J., Bennett, N.J., Spagnuolo, J., Gagic, D., Russel, M., 2011. Filamentous bacteriophage: Biology, phage display and nanotechnology applications. *Current Issues of Molecular Biology* 13, 51–76.
- Scott, J.K., Smith, G.P., 1990. Searching for peptide ligands with an epitope library. *Science* 249, 386–390.

Replication of *Bacillus* Double-Stranded DNA Bacteriophages

Silvia Ayora, National Biotechnology Center–Spanish National Research Council, Madrid, Spain

Paulo Tavares, Institute for Integrative Biology of the Cell, CEA, CNRS, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

Ruben Torres and Juan C Alonso, National Biotechnology Center–Spanish National Research Council, Madrid, Spain

© 2021 Elsevier Ltd. All rights reserved.

Glossary

θ replication Circle-to-circle or θ -type DNA synthesis. It occurs when the replisome is assembled on a circular chromosome at the *ori* region, leading to its uni- or bidirectional replication. Its product is two catenated DNA molecules.

σ replication It is also referred as rolling circle replication. Its product is a concatemer generated by recombination-dependent DNA replication (RDR).

Concatemer A head-to-tail linear DNA molecule generated either by a replicative or by a recombination mechanism.

cos A DNA sequence recognized by the viral terminase to initiate and terminate encapsidation of a unit-length genome.

DNA synthesis To synthesize both DNA strands in a concerted and semiconservative fashion in the 5'→3' direction. One strand is synthesized mainly continuously (the leading strand) and the other is synthesized discontinuously (the lagging strand).

Headful packaging Processive viral DNA packaging along a phage DNA concatemer whose first cycle is initiated by an endonucleolytic cleavage at *pac*, and terminated by a sequence-independent cut that occurs when a threshold amount of packaged DNA is reached inside the phage capsid. The subsequent encapsidation cycle begins at the DNA end created by the headful cut, and processive headful

encapsidation cycles ensue. The DNA molecules encapsidated are terminally redundant (usually 1.02–1.1 longer than the genome size) and the order of genes is partially circularly permuted to an extent that depends on the concatemer size.

pac A DNA structure and sequence recognized by the viral terminase to initiate encapsidation by the headful packaging mechanism.

Preprimosome A protein complex that loads the replicative DNA helicase into DNA, and it is subsequently disassembled.

Primosome A protein complex composed of the replicative DNA helicase and the DNA primase, or a single helicase-primase protein. In some cases, a hybrid RNA-DNA primer is synthesized in concert with a DNA polymerase (DnaE) to initiate DNA synthesis.

Replication origin (*ori*) A locus at which a replication initiation protein binds or an RNA polymerase transcript melts the duplex at an adjacent AT rich region (DUE, for DNA-unwinding element). The resulting ssDNA regions act as loading site of the replicative helicase.

Replisome A protein machinery devoted to replicate DNA. In *Bacilli* it is composed of two replicative DNA polymerase holoenzymes (PolC and DnaE), the clamp loader DnaX-HolA-HolB (also termed $\tau\delta\delta'$), the processivity factor DnaN (also termed β clamp), and the primosome.

Introduction

Bacteriophages (or phages) provide the largest unexplored genetic reservoir and are the most common biological entities on earth. The large majority of phages with double-stranded (ds) DNA belong to the ancient *Caudovirales* Order (~97%), and their division is probably older than the separation of the three domains of life (Bacteria, Archaea and Eukarya). The *Caudovirales* Order is composed of three major Families: *Myoviridae* (~20% of total species of the Order), *Siphoviridae* (~68%), and *Podoviridae* (~11%) (Ackermann, 2003).

Bacteria of the Firmicutes phylum (Gram-positive cells with low dC + dG content in their DNA), which are placed at the earliest division of Bacteria, are predominant in a healthy gut microbiota. Among Firmicutes, the Class *Bacilli* includes bacteria with round- (termed coccus) or rod-like form (bacillus) that represent a major health burden in the community as well as in hospitalized patients as e.g., *Streptococcus spp.*, *Enterococcus spp.*, *Staphylococcus spp.*, *Listeria spp.* and several Genera from the *Bacillaceae* family. Thus, the inherent ability of phages to infect those bacteria and control their biomass turnover makes them ideal candidates to develop antimicrobial agents for pathogen-specific remediation (e.g., phage therapy). On the other hand, *Bacilli* phages infecting *Lactococcus spp.*, *Lactobacillus spp.*, *Leuconostoc spp.*, *Oenococcus spp.* as well as certain *Bacillaceae* are a major concern to dairy, wine, and meat bioprocesses. Understanding how these phages multiply is an essential step to harness the full potential of phage-based technologies and to generate novel strategies to combat industrial phage contaminations. Here, we describe present knowledge on the different strategies that *Bacilli* phages use to replicate their DNA. A representative phage from each of the major families of the Order *Caudovirales* that infects the model bacterium *Bacillus subtilis* was chosen (Table 1), and compared with other *Bacilli* phages.

Table 1 Phage- and Host-encoded proteins involved in DNA replication

Function	SPO1	T4 ^a	SPP1	ϕ 29	Host
ssDNA binding	?	gp32	gp36	gp5	SsbA
Replisome organizer	–	–	gp38	–	DnaA/PriA
Protein priming	NA	NA	NA	gp3/TP	NA
Replication mediator (RMP)	?	gp59	gp38-gp39	gp6	PriA-DnaB-D-I
Replicative helicase	gp21.1	gp41	gp40	–	DnaC
Helicase loader	–	–	gp39	–	DnaB-D-I
DNA primase	gp21.95	gp61	host	–	DnaG-DnaE
DNA polymerase ^b	gp31 (A)	gp43 (B)	host	gp2 (B)	PolC, DnaE (C)
Clamp loader	?	gp44/62	host	–	DnaX-HolAB
Sliding clamp	?	gp45	host	–	DnaN
Topoisomerase	J1 mutation	gp39/52/60	host	–	GyrAB, ParCE
Rad50-Mre11nuclease	gp21.3/21.9	gp46/47	?	–	SbcCD
5'→3' exonucleases	?	–	gp34.1	–	–
Recombinase mediator	–	UvsY	–	–	RecOR
Recombinase	–	UvsX	gp35	–	RecA
Translocase	gp19.5	UvsW	?	–	RecG
DNA helicase	gp.34.33	?	?	–	PcrA
DNA helicase	–	Dda	?	–	RecD2
HJ resolvase	gp24.1	gp49	gp44	–	RecU

^aEarly and late replication of large dsDNA viruses of the Myoviridae family was reconstituted *in vitro* only for phage T4 that infects *Escherichia coli*. Therefore, T4 is depicted for its comparison with SPO1.

^bThe family to which the replicative DNA polymerase belongs is denoted within parentheses.

Note: NA, not applied; ?, unknown.

The majority of cells and their viruses initiate DNA replication by building an *orisome*. In the host cell an ATP-dependent replisome organizer protein(s) binds to multiple sites at the *ori* region. This interaction leads to a conformational change that exerts a mechanical force for strand opening of an adjacent AT-rich region termed DUE (for DNA-unwinding element) and recruits, directly or indirectly, the replicative helicase to establish bidirectional replication forks. Most viruses of the Order *Caudovirales* use their own replication machinery or a combined action of viral and host functions, while a minor number of them encodes only for one function (the *ori* region) that is essential for initiation of DNA replication, and relies on the host machinery for replication elongation, as lactococcal phage c2 (and perhaps sk1) of the *Siphoviridae* Family.

The dsDNA phages have evolved an efficient and economical process for amplification of over 100 virus genome copies within minutes. They use three different mechanisms to initiate DNA synthesis. The majority of those phages use two different replication modes (early and late) to replicate their linear or circular genomes to generate the concatemeric DNA substrate for genome packaging into preformed viral procapsids (or proheads). A less common approach is the replication of linear unit-length genomes by a protein priming mechanism (Salas, 1991). These different strategies to initiate DNA replication and the replication mode used are not restricted to a specific tailed phages Family. For example, within the *Podoviridae* Family, phage T7 uses RNA polymerase to initiate DNA replication as earlier described for plasmid ColE1 (Masukata *et al.*, 1987), phage P22 uses an ATP-independent replisome organizer protein to start DNA synthesis, whereas phage ϕ 29 initiates DNA synthesis by a protein priming mechanism.

Viruses, in general, and the different Families of dsDNA phages that infect bacteria of the *Bacilli* class in particular, do not have a universal set of genes. They have a core of ancient virus hallmark genes shared within viral families, combined with non-conserved genes that build genomes with a mosaic organization. Viral infection and horizontal gene transfer strongly contribute to spread and swap genetic information among viruses leading to this mosaicism that obscures their phylogenetic relationship. Such effect is also observed in the genome module encoding the DNA replication and recombination functions of *Caudovirales*. In this article we summarize current knowledge on the replication strategies of the large genome size SPO1-like viruses of *Myoviridae* (with a contractile tail); the middle-size SPP1-like viruses of *Siphoviridae* (with flexible, non-contractile tail); and the small-size ϕ 29-like viruses (with a short, non-contractile tail) of *Podoviridae* Family.

With the exception of ϕ 29-like viruses, the replication of tailed phage genomes starts by an early replication mode that generates a template substrate for late replication. Late replication uses a recombination-dependent replication (RDR) mode to yield the head-to-tail concatemer that is cut at specific site(s) *cos* (to render unit-length genome) or *pac* (redundant ends) for packaging into viral procapsids. Therefore, recombination is an essential step in phage DNA metabolism. Phages can use a *pac* site and a headful packaging mechanism to generate mature terminally redundant DNA molecules with or without circular permutation (Casjens, 2011; Rao and Feiss, 2015). The *pac*-containing phages constitute major vehicles for horizontal gene transfer, including virulence and antibiotic resistance genes. Alternatively, tailed phages use *cos* sites in the replicated genome concatemer to initiate and terminate DNA encapsidation, which generates unit-length packaged DNA molecules (Casjens, 2011; Rao and Feiss, 2015). Other phages like SPO1 package DNA molecules that have a non-permuted, several kb-long, terminal redundancy at their ends. Since the substrate used for processive packaging is a concatemer, their terminal redundancy must be generated by *de novo* DNA synthesis in concert with DNA packaging (Stewart *et al.*, 2009). The underlying mechanism remains unknown.

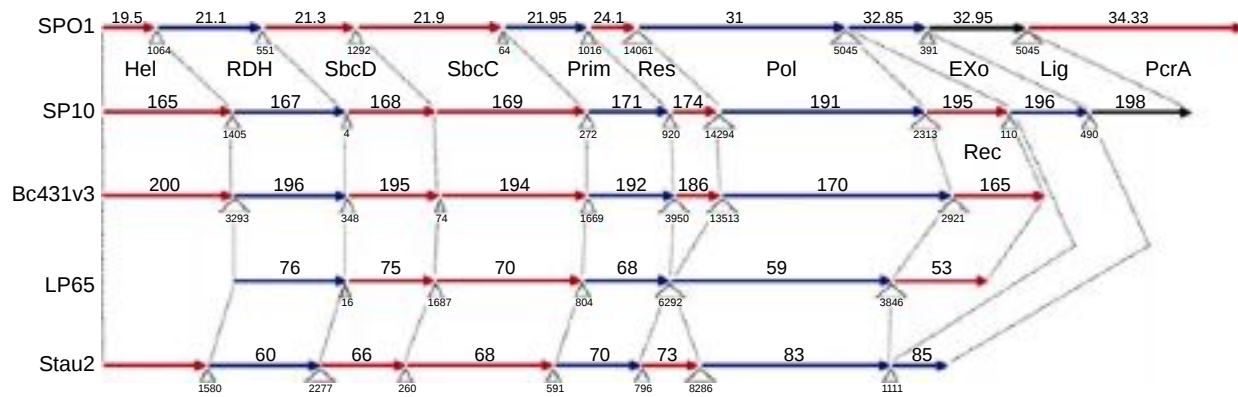


Fig. 1 Genome overview of the replication module of SPO1-like phages. The arrows indicate the direction of gene transcription. Relatedness is derived from P-BLAST comparison of phage-encoded products. The color of each gene identifies its classification in functional groups: DNA recombination (red), DNA replication (blue), and related functions (black). The lengths of the interspersed regions are indicated. The names of the genes are labeled above the arrows and their putative function is annotated. Hel: helicase; RDH: replicative DNA helicase; SbcD and SbcC: nuclease; SbcCD; Prim: primase; Res: holliday junction resolvase; Pol: DNA polymerase; EXo: 5'→3' exonuclease; Lig: DNA ligase; PcrA: helicase; Rec: RecA-like recombinase. The genome sequences accession numbers are NC 011421, MF422185, NC 020873, AY682195 and KP881332.

Finally, ϕ 29-like viruses initiate DNA replication by protein priming that is extended by the DNA polymerase. The synthesized unit-length genome is subsequently packaged with a terminal protein (TP) covalently linked at the 5'-end into an empty prohead (Salas, 1991).

Phages with homology to the SPO1, SPP1 or ϕ 29 prototype genomes can show three degrees of evolutionary relatedness: (i) those whose complete genome has strong nucleotide sequence identity to the phage prototype; (ii) those with a mosaic genome organization, having discrete modules with significant degree of nucleotide sequence homology to the phage prototype; and (iii) those whose sequence identity between the different replication proteins is low, but their coding genes order is conserved. Therefore, our analysis of the replication machinery of SPO1-, SPP1-, and ϕ 29-like phages will not be restricted to those with close phylogenetic relatedness. We will use a *sensu lato* classification to include viruses with conserved synteny of DNA replication gene modules with respect to the prototype virus for the analysis of their replication machineries.

DNA Replication of SPO1-Like Viruses

The genus SPO1-like virus, including the prototype phage SPO1 that infects *B. subtilis*, is in the process of being renamed genus Okubovirus in honor of Shunko Okubo, who isolated SPO1 and carried out the first genetic studies of this myovirus. It includes large dsDNA viruses with a genome size > 100 kb. *B. subtilis* spp SP82, 2C, SP8, ϕ 25, H1, ϕ e and SP5C phages show a significant degree of nucleotide sequence identity to phage SPO1 (group i). A comparative analysis of available phage genome sequences identified two other groups of phages based on their level of relatedness to SPO1: *Bacillus* phages SP10 and Bc431v3, *Listeria* spp P100 and A511; *Staphylococcus* spp phages K, G1, P1, S3K, ϕ 812, Twort, Stau2; and *Lactobacillus* spp phage LP65 and Lb338-1; *Enterococcus* spp phage ϕ EF24C. Phages SPO1 (representative of group i), SP10 and Bc431v3 (group ii) and LP65 and Stau2 (group iii) are depicted in Fig. 1. Replication and recombination genes are clustered in a region of their core genome. They are spaced either by genes encoding for a function optimizing viral DNA amplification or for proteins of unrelated/unknown function.

The genome of SPO1-like viruses is large (130–160 kb-long) and thymine is replaced by 5'-hydroxymethyluracil (hmU). DNA replication starts from at least two replication origins. Although SPO1 is the best-characterized member of this group, there is limited information on its DNA replication mechanism. Here, SPO1 protein candidates participating in the process will be discussed, and potential mechanistic similarities with the well-studied DNA replication of myovirus T4 will be addressed.

At early times of infection, SPO1 DNA replication initiates at specific origin sequences by RNA polymerase transcription to generate a stable untranslated transcript. This RNA remains attached to its template to form a strand displacement RNA (R-loop). This structure maintains the DNA strands melted at the DUE region in a process that is independent of recombination proteins, and DNA replication starts. As infection progresses, a RDR mechanism becomes predominant and the origin-dependent replication mode is turned off. The SPO1-encoded DNA primase (Prim, gene product (gp) 21.95), replicative DNA helicase (RDH, gp21.1), DNA polymerase (Pol, gp31) and 5'→3' exonuclease (Exo, gp32.85) (Fig. 1) may have their functional counterparts in the host DnaG, DnaC, DNA PolII and YcpP proteins, respectively, but these host functions are unable to substitute for the SPO1-encoded proteins (Table 1). When the RDR mechanism is turned on, phage SPO1 probably relies on phage- and host-encoded recombination machineries. Although phage T4 codes proteins with single-stranded DNA-binding and recombinase activities essential for its replication, none of the SPO1-encoded products shows significant similarity to proteins with related functions (Table 1). However, other phages of this Family encode an ATP-dependent recombinase (Rec, counterpart of host RecA)

downstream of the viral-encoded DNA polymerase (Fig. 1), suggesting that in these phages initiation of late replication is controlled by a viral protein. At present, it is unknown if host RecA is able to substitute for the viral encoded counterpart.

Some of the SPO1 DNA replication functions for which no viral gene was identified could be performed by uncharacterized phage proteins. For example, no gene product shows significant sequence similarity to host DNA gyrase. However, SPO1 infection requires a DNA topoisomerase function and SPO1 mutants can replicate in the presence of DNA gyrase inhibitors, suggesting the presence of a viral topoisomerase (Table 1). In late replication, the host-encoded recombination machinery might initiate RDR. Then, SPO1-encoded DNA helicases Hel and PcrA-like (gp19.5 and gp34.33) could prevent over-replication at displacement loops (D-loops) formed by homologous recombination or are involved in the replication of branched intermediates. Branched replication intermediates could be resolved by the phage putative HJ-specific resolvase Res (gp24.1) (Table 1), although some viruses might have lost it, and rely on a host function (Fig. 1). The SPO1 gp21.1 and gp21.9 proteins share homology with the host SbcCD complex, which cleaves DNA ends, hairpins, abnormal replication intermediates, and terminal DNA structures preparing DNA ends for RDR, in a way similar to the eukaryotic Rad50-Mre11 complex. Finally, the gp32.95 LigD-like DNA ligase could seal hmUra containing DNA ends. The majority of SPO1 genes are co-oriented with the DNA replication RDR direction. One strategy to avoid deleterious head-on replication-transcription conflicts is to use the helicase gp34.33 together with the gp21.1-gp21.9 complex to promote replisome progression and to reduce head-on conflicts.

The detailed mechanism of DNA replication of SPO1-like viruses (*Spounavirinae* subfamily) is unknown. Their genome core is reminiscent of the one from *E. coli* T4-like phages (*Tevenvirinae* subfamily). Since T4 DNA replication was extensively studied and reconstituted *in vitro* (Nossal, 1983; Mosig, 1998; Kreuzer, 2000; Benkovic and Spiering, 2017; Barry *et al.*, 2018), we revisit its molecular mechanism as a model for the large viruses of the *Myoviridae* Family. In phage T4, cytosine is replaced by hydroxymethylcytosine (hmC). All the proteins necessary for replication of its linear and/or circular genome are encoded by the phage, with the exception of the RNA polymerase that is provided by the host cell.

Early after infection, T4 DNA replication starts from the *oriE* locus close to an early promoter (major origin) and from four secondary origins (*oriA*, *oriC*, *oriF* and *oriG*) located near middle promoters. The primary and secondary origins are located just upstream of a DUE. RNA polymerase synthesizes a RNA that binds stably to its template, forming an R-loop. Following processing by T4-encoded RNaseH, this untranslated transcript serves as a primer to initiate DNA synthesis. In the R-loop the nascent RNA anneals to the coding DNA strand, leaving the displaced noncoding DNA coated by the single-stranded DNA binding protein (Ssb) gp32. The acidic C-terminal tail of gp32 interacts with the replication mediator protein (RMP) gp59, the family B DNA polymerase gp43, the recombinase UvsX, recombination-mediator protein UvsY, the DNA helicase Dda, and the primase gp61 (Table 1). The interaction of gp32 bound to ssDNA with gp59 and the replicative gp41 helicase loads both proteins onto the viral linear and/or circular genome (Table 1). Gp59 binds to the R-loop-dependent replication origin. Gp41 then interacts with the gp61 whose recruitment promotes assembly of the T4 primosome. The primosome loads the T4 DNA polymerase holoenzyme, composed of gp43, the sliding clamp gp45, and the clamp loader gp44/gp62.

Leading strand synthesis is initiated at the 3'-OH end of the RNA primer by the gp43-gp4-gp44/gp62 complex. After the gp59 mediator frees gp41, the primosome allows gp41 unwinding of the substrate in the 5'→3' direction, gp61-mediated synthesis of the lagging strand primer, and gp32 binding to the ssDNA regions (Table 1) (Benkovic and Spiering, 2017; Barry *et al.*, 2018). Gp43 synthesizing the lagging strand remains in contact with the leading strand polymerase at the replication fork. This organization folds the lagging strand into a 700–1500-nucleotide loop coupling leading and lagging strand synthesis, according to the *trombone model* of replication. The encounter of the replisome with the 5'-end of the previous synthesized Okazaki fragment triggers dissociation of the lagging-strand replisome that disengages from gp45, leaving the clamp protein bound to the lagging-strand. Synthesis of a new Okazaki fragment ensues (Barry *et al.*, 2018).

In the large Myoviruses of the T4 and SPO1 groups, early DNA replication starts from an internal origin of replication in a linear molecule. This implies that when viral replication reaches the chromosome ends, the synthesized DNA molecules are shorter than unit length. The leading strand is replicated to the 5'-end yielding a 3'-tailed duplex while lagging strand replication leaves a ssDNA region at both ends resulting in a 5'- and 3'-tailed duplex replicated DNA molecule. The 3'-ssDNA ends coated by gp32 are competent for recombination by strand invasion. At late stages of phage T4 infection, the UvsW DNA helicase inactivates initiation from R-loops by unwinding potential primer transcripts that may have persisted or were synthesized during late infection, ensuring that RDR predominates. The ATP-dependent UvsX recombinase, loaded by the UvsY recombinase mediator, polymerizes on the 3'-tailed substrate generated by early replication. UvsX-mediated DNA pairing forms a displacement loop (D-loop) at a homologous region on the linear DNA template to initiate DNA synthesis by RDR (Table 1) (Kreuzer, 2000). Viral-mediated RDR resembles eukaryotic break-induced replication, but the latter is mutagenic and prone to chromosome rearrangements (Llorente *et al.*, 2008). Unidirectional replication, initiated from this invading 3'-end that is the primer for leading strand replication was proposed to duplicate most of the T4 DNA molecule. Multiple strand invasion events generate a complex branched network as the infection progresses. This replication mode promotes the appearance of replicating DNA intermediates containing multiple covalently linked copies of the genome. The Holliday junction (HJ) resolvase gp49 (also termed Endo VII) cleaves the network of chromosomes and such debranching generates linear concatemeric DNA molecules (Mosig, 1998). The viral terminase initiates packaging at *pac* within these concatemers into empty T4 phage proheads by a processive headful mechanism. DNA molecules slightly longer than the unit-length of a T4 genome are encapsidated, resulting in a population of terminally redundant and circularly permuted packaged T4 chromosomes. This differs from SPO1 that uses a substrate concatemer to package, by an unknown mechanism, DNA molecules that have a several kb-long terminal redundancy (Stewart *et al.*, 2009).

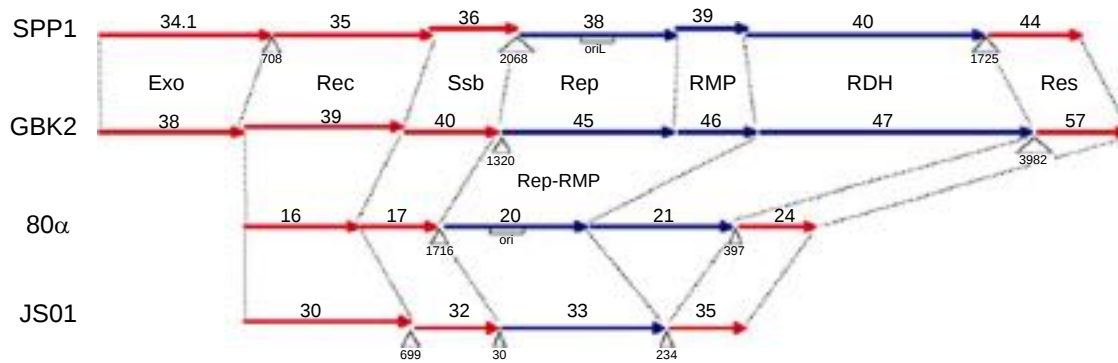


Fig. 2 Genome overview of the replication module of SPP1-like phages. The figure organization and symbology are as in Fig. 1. The protein functional annotation is: Exo: 5'→3' exonuclease; Rec: Recombinase; Ssb: single-stranded DNA binding protein; Rep: replisome organizer; RMP: replisome mediator protein; RDH: replicative DNA helicase; Res: Holliday junction resolvase. oriL and ori are origins of replication. The genome sequences accession numbers are X97918, KJ159566, DQ517338 and KC342645.

Another mechanism proposed to generate the phage T4 DNA concatemer is that 3'-ssDNA ends of the terminal redundant region pair, leading to a circular genome that becomes supercoiled by the action of the T4 DNA topoisomerase (the gp39-gp52-gp60 complex). Then, most of T4 DNA replication forks are initiated by UvsX DNA pairing and D-loop formation at a homologous region on a supercoiled circular DNA template. The gp59 mediator loads the T4 primosome onto the displaced ssDNA coated by gp32. The primosome, in concert with gp32, recruits the T4 DNA polymerase holoenzyme. In the presence of the T4 DNA topoisomerase, the DNA polymerase holoenzyme catalyzes DNA synthesis that begins at the 3'-OH end of the invading strand. DNA is synthesized continuously on one template strand, as in the leading strand of a replication fork, and discontinuously on the other template strand (Okazaki fragments) by a semiconservative mechanism. In one of the models, the circular DNA is then nicked to generate linear dsDNA concatemers by σ replication (Barry *et al.*, 2018) that are the substrate for DNA packaging. If the replication machinery stalls by an unrepaired DNA lesion, the phage RDR system rescues the stalled or collapsed fork. The UvsW DNA helicase can catalyze the regression of a stalled replication fork into a HJ-like structure that was postulated to be an intermediate in an error-free DNA damage tolerance pathway. The T4 gp46/gp47 exo/endonuclease and the Dda DNA helicase play a critical, although still poorly defined, role in replication restart. Finally, the gp30 DNA ligase could seal hmC containing DNA ends.

DNA Replication of SPP1-Like Viruses

B. subtilis phage SPP1 is the best-characterized member of middle-sized viruses from the *Siphoviridae* Family that infects bacteria of the *Bacilli* Class. Currently, phage SPP1 is included in the unassigned viruses of the Family, in spite of the fact that its DNA replication and packaging machineries as well as the viral particle assembly mechanisms are among the best characterized within the *Siphoviridae*. This calls for definition of a SPP1-like genus. A comparative analysis of the complete SPP1 genome shows that *B. subtilis* phages SF6, 41c, ρ 15, Lurz2 and Lurz3 have strong nucleotide sequence identity to phage SPP1, defining group (i). The *Bacillus spp* phage PM1 has discrete modules with significant nucleotide sequence identity to phage SPP1. Phage GBK2 has no homology at the DNA sequence level but has extensive similarity at the protein level, particularly in case of the DNA replication proteins whose coding genes order is co-linear with the SPP1 genome (Fig. 2). Finally, the *Staphylococcus spp* phages 80 α , ϕ 11 and JS01; *Lactobacillus spp* phage LL-H; and *Streptococcus spp* phage Sfi11 viruses show a conserved gene order of the DNA replication module with phage SPP1. Phages SPP1 (representative of group i), GBK2 (group ii) and 80 α and JS01 (group iii) are depicted in Fig. 2. All these phages package their DNA by a headful mechanism, generating encapsidated chromosomes with terminal redundancy and partial permutation.

Phages of this group replicate, most likely, their 40–50 kb genome using a mechanism similar to phage SPP1. At early times of infection, the SPP1 genome circularizes and replication is initiated at a specific and discrete replication origin sequence (*oriL*) in a process that resembles host chromosome replication. The early (θ -type) and late (σ -type) replication modes rely on early transcribed SPP1 genes. Early replication requires the essential SPP1-encoded replisome organizer (Rep, gp38), replication mediator protein (RMP, gp39) and replicative DNA helicase (RDH, gp40). As the infection progresses, σ -type concatemeric DNA replication becomes predominant and θ -type *ori*-dependent replication is turned off. The switch of replication mechanism relies on recombination proteins: the SPP1-encoded ATP-independent recombinase (Rec, gp35), a 5'-3' exonuclease (Exo, gp34.1), a single-stranded DNA binding protein (Ssb, gp36), and a Holliday junction resolvase (Res, gp44). In certain phages the initiator and mediator functions might be fused in a single polypeptide (Rep-RMP: phage 80 α gp20 and phage JS01 gp33) (Fig. 2) while the single-stranded DNA binding protein, the 5'-3' exonuclease and the HJ resolvase can be substituted by host encoded functions. The essential recombinase (Rec) can belong either to the Rad52 superfamily of recombinases (defining the Res β [SPP1 gp35, GBK2 gp39], Sak [80 α gp16] and Erf [C2 gp8] subfamilies of recombinases) or to the RecA-like superfamily as the ATP-dependent Sak4 recombinase (ϕ 11 gp12). The HJ resolvase (Res), like gp44 of SPP1, belongs to the RusA-like superfamily (Fig. 2).

SPP1 relies on numerous host proteins (the SPP1 *missing* functions) that are essential for its DNA replication, namely PolC, DnaE, DnaN (β -clamp), the DnaX-HolA-HolB or $\tau\delta\delta'$ complex and DNA gyrase. Cellular functions can also replace the non-essential gp34.1, gp36 and gp44 proteins. Although the lack of gp34.1 and gp36 is detrimental, the absence of gp44 has only a marginal effect on SPP1 multiplication. Initiation of circle-to-circle θ DNA replication and the late σ RDR (rolling circle replication) have been reconstituted *in vitro* (Seco and Ayora, 2017; Seco *et al.*, 2013).

After entry of the SPP1 linear chromosome in the *B. subtilis* cytosol, viral or host homologous recombination functions use the viral chromosome terminal redundancy to circularize the SPP1 genome, and the host DNA gyrase supercoils the circular DNA. The viral genome contains two replication origins (*oriL* and *oriR*) located ~ 13 kb apart in the circularized genome. The *oriL* sequence lays within the gene of the ATP-independent replisome organizer gp38 and *oriR* in an intergenic region transcribed late. Both *oriL* and *oriR* contain a different number of gp38-binding boxes (boxes A and B).

DNA replication is initiated by the preferential binding of gp38 to *oriL* that induces a localized unwinding of the adjacent AT-rich region on the supercoiled DNA template. The replication mediator gp39, which is an intrinsically disordered protein, interacts with the replicative helicase gp40 and with the replisome organizer gp38. Gp40 promotes gp39 protein folding to form a double hexamer structure together with gp40. The gp39-gp40 complex has neither ATPase nor helicase activity. Gp38 bound to *oriL* interacts with gp39 and loads the gp39-gp40 complex to the DUE region created when gp38 is bound to *oriL*. Gp39 loads and locks gp40 on the *oriL* lagging strand region until the gp38-gp39 complex is disassembled. Concomitantly with gp38-gp39 dissociation, gp36 binds to the ssDNA, and gp40 becomes free of the inhibition exerted by gp39. The hexameric replicative gp40 DNA helicase hub interacts with DnaX (or τ subunit) that is a component of the pentameric DnaX₃-HolA-HolB (also termed $\tau_3\delta\delta'$) complex, with the DnaG primase, and with the DnaE polymerase (Table 1). Then, the $\tau_3\delta\delta'$ complex loads both family C DNA polymerases (PolC and DnaE) and two sliding clamps (DnaN) to assemble the functional replisome (Seco and Ayora, 2017). Unidirectional leading and lagging strand synthesis is initiated at *oriL*. Priming of both strands is carried out by the DnaG-DnaE complex (like eukaryotic DNA polymerase α) that synthesizes a RNA-DNA hybrid. This hybrid is further extended by PolC in the context of the full replisome (i.e., in a complex with DnaN-DnaX-HolA-HolB) and catalyzes unidirectional replication of both strands (Table 1). DNA gyrase is required to remove positive supercoiling produced by the advance of the replication fork on the circular DNA template. The DnaG primase actively contributes to DNA polymerase switch at the leading strand, by inhibiting DnaE and stimulating PolC-mediated leading strand synthesis. On the other hand, gp36 and gp40 bound to the lagging strand recruit the DnaG-DnaE complex. This complex synthesizes the RNA-DNA hybrid while gp36 inhibits DnaE and stimulates PolC (Seco and Ayora, 2017). The replisome catalyzes discontinuous synthesis using the lagging strand template (Okazaki fragments) by a leading and lagging strand coupling via a semiconservative mechanism of DNA synthesis (trombone model). The encounter of the replisome with the 5'-end of previous synthesized Okazaki fragment triggers the dissociation of the lagging-strand replisome that disengages from the DnaN clamp leaving it bound to the lagging-strand.

After one or a few rounds of circle-to-circle DNA synthesis this θ mode of replication is turned off and SPP1 replication switches to linear concatemeric DNA replication (σ mode). The switch was proposed to be triggered by stalling replication fork progression (*road blocking*) by gp38 bound to *oriR*. Seven SPP1-encoded proteins are involved in the process of resuming replication at the stalled fork: four essential proteins (gp38 and gp39, working as a preprimosomal proteins, gp40 and gp35) and three partially dispensable proteins (gp34.1, gp36 and gp44). A branch migration translocase might regress the stalled fork leading to a Holliday junction (HJ) like structure. This extruded HJ can be cleaved by the gp44 HJ resolvase to generate a one-ended double strand break. The 5'→3' exonuclease activity of gp34.1 generates the 3'-tailed substrate required for gp35 activity (Table 1). Gp35, in concert with gp36, catalyzes DNA pairing and D-loop formation at a homologous region on an intact DNA template to initiate homology-directed RDR. Host RecA recombinase cannot replace for the gp35 activity. Gp39 in the preprimosome gp38-gp39 complex loads the replicative gp40 helicase that orchestrates assembly of the active replisome, as described above for the θ -type replication mode, initiating σ -type DNA replication. Two different mechanisms were proposed to achieve concatemeric DNA synthesis (Lo Piano *et al.*, 2011). The first is a bubble migration mechanism that invokes conservative DNA synthesis. The second invokes the action of the gp44 resolvase, or its cellular equivalent, to cleave and rejoin the D-loop structure such that a "rolling-circle"-like mechanism of DNA replication (σ -mode) occurs.

The σ replication mode generates a long linear concatemer. The SPP1-encoded terminase initiates DNA packaging by cleavage at the *pac* sequence that occurs only once along a substrate concatemer. The packaging cycle is terminated by a non-specific sequence cut determined by the amount of DNA inside the capsid, yielding packaged molecules longer than the phage genome (headful packaging mechanism) (Oliveira *et al.*, 2013). Processive headfuls are then sequentially packaged from the concatemer, generating a population of mature chromosomes with terminally redundant and partially circularly permuted sequence.

DNA Replication of ϕ 29-Like Viruses

B. subtilis ϕ 29 phage is the best-characterized member of small-sized viruses from the *Podoviridae* Family that infect bacteria of the *Bacilli* Class. Currently the Genus ϕ 29-like virus is in the transition to be renamed Salasvirus in honor of Margarita Salas, who has worked on ϕ 29 for over five decades. A comparative analysis of the complete genomes of *Bacillus* spp phages Nf, Goe1, ϕ 15 and PZA show strong nucleotide sequence identity to the ϕ 29 genome. All members of the ϕ 29-like genus seem to have a similar core genome and a conserved organization, with phage PZA exhibiting the particularity of an inversion of the left early genome region (Fig. 3). In the ϕ 29-like Genus there are phages with a mosaic genome organization, which have discrete modules with significant degree of

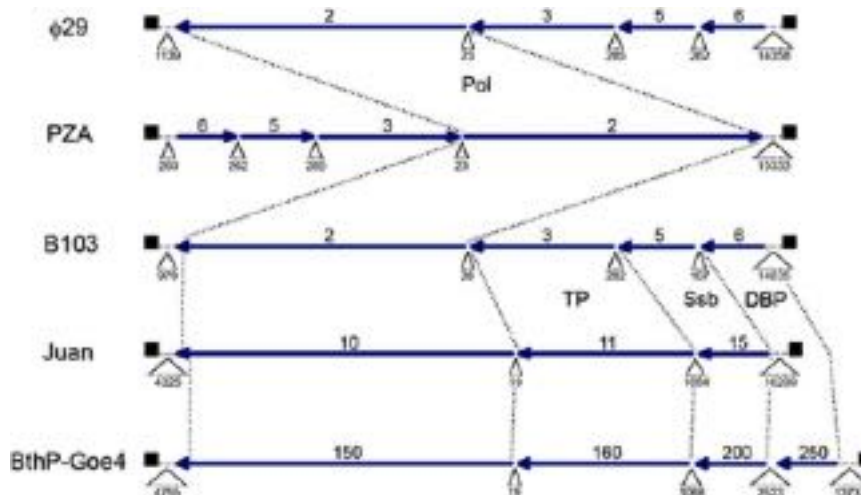


Fig. 3 Genome overview of the replication module of $\phi 29$ -like phages. The figure organization and symbology are as in Fig. 1. The small filled square denotes the TP bound to the 5'-end of the genome. The protein functional annotation is: Pol: DNA polymerase; TP: terminal protein; Ssb: single-stranded DNA binding protein; DBP: double-stranded DNA binding protein. The genome sequences accession numbers are NC 011048, PZACG, X99260, MF156577 and MH817022.

nucleotide sequence identity, as *Bacillus* spp phages GA-1, B103, Juan, BthP-Goe4, and Goe6; *Staphylococcus* spp phages P68 and ϕ IPLA7; and *Streptococcus* spp Cp-1, Cp-5, Cp-7, and Cp-9. Phages $\phi 29$ and PZA (representatives of group i), B103 (group ii) and Juan and BthP-Goe4 (group iii) are depicted in Fig. 3. The protein-priming DNA replication mode of $\phi 29$ -like viruses generates a linear unit-length molecule with a terminal protein (TP, also termed gp3 in $\phi 29$) at the 5'-end. The TP and a RNA molecule are essential for packaging of a unit-length viral genome into empty proheads.

All $\phi 29$ -like viruses have a linear dsDNA with a TP covalently linked to the DNA ends through a phosphoester bond between the OH group of a serine (e.g., $\phi 29$) or threonine (Cp-1) residue in the TP and the 5'-dAMP embedded within a DNA short inverted terminal repeat (ITR). The ITR length varies among phages of the $\phi 29$ Genus. The $\phi 29$ DNA replication system was reconstituted *in vitro* and will be described as the model for this group of phages (Salas, 1991; Salas *et al.*, 2016). In contrast to the complexity of other reconstituted replication systems, efficient synthesis of the $\phi 29$ genome can be accomplished *in vitro* with only four phage proteins. They are the double-stranded $\phi 29$ DNA-binding protein (DBP) gp6, the single-stranded DNA binding protein (Ssb) gp5, TP gp3, and the family B DNA polymerase (Pol) gp2. Gp2 couples DNA polymerization to strand displacement. This monomeric DNA polymerase also shows 3'→5' exonucleolytic activity, enabling proofreading. Its helix destabilizing activity, makes strand displacement by an accessory DNA helicase unnecessary.

Upon delivery of the $\phi 29$ DNA into the host cytoplasm, the parental TP covalently linked to each 5'-end of the genome associates with the bacterial nucleoid (Muñoz-Espin *et al.*, 2010). The abundant replication mediator gp6 binds along the $\phi 29$ DNA to form a nucleoprotein complex on viral DNA. This complex unwinds the ends of the DNA helix to facilitate loading of the heterodimer formed between a free TP and gp2 at a TP-DNA end. This TP-DNA end acts as the origin for replication. The DNA polymerase then catalyzes the formation of a phosphoester bond between a dAMP and the hydroxyl group of a Serine residue of the free TP, giving rise to the TP-dAMP initiation complex. The initiation reaction is directed by the second T at the 3'-end of the six nucleotides long ITR (3'-TTTCAT-5') present at the extremity of the $\phi 29$ DNA molecule. After, the TP-dAMP complex translocates one position backwards to recover the information corresponding to the first T on the template strand. The second T serves again as template for the incorporation of the following nucleotide (*sliding-back mechanism*), followed by addition of five nucleotides. Any mis-incorporated nucleotide in the TP-dNMPs covalent complex is not a substrate for the 3'→5' exonuclease proofreading activity of the $\phi 29$ DNA polymerase. Replication fidelity during the initiation reaction thus relies on the sliding-back mechanism in phages that initiate replication by protein priming. After initiation and sliding back, the $\phi 29$ DNA polymerase/primer TP heterodimer undergoes structural changes during incorporation of 6–9 nucleotides (transition stage). The DNA polymerase dissociates from the primer TP when nucleotide 10 is incorporated into the nascent DNA chain, shifting to the elongation mode.

The polymerase dissociated from the TP-gp2 heterodimer catalyzes highly processive DNA synthesis coupled to strand displacement of the non-template strand that is coated by gp5. The replication of both strands proceeds continuously from each terminal priming event (type I replicative intermediates). When the two replication forks moving in opposite directions meet, the type I replicative intermediate gives rise to two physically separated type II replicative intermediates. These molecules consist of full-length DNA in which a portion of the DNA, starting from one end, is dsDNA and the portion spanning to the other end is ssDNA. Once replication of both strands is accomplished, the two DNA polymerase molecules dissociate from the viral DNA to start initiation of replication in a new $\phi 29$ DNA molecule. The transcription machinery could be an obstacle to type I replicative intermediates and head-on and co-directional collisions could stall the replisome, leading to reassembly of the replication machinery. The small-size $\phi 29$ -like viruses do not encode for replication restart functions, suggesting that if replisome stalling occurs, the re-initiation relies on host functions.

References

- Ackermann, H.W., 2003. Bacteriophage observations and evolution. *Research in Microbiology* 154, 245–251.
- Barry, J., Wong, M.L., Alberts, B., 2018. *In vitro* reconstitution of DNA replication initiated by genetic recombination: A T4 bacteriophage model for a type of DNA synthesis important for all cells. *Molecular Biology of the Cell* 30 (1), (mbc.E18-06-0386).
- Benkovic, S.J., Spiering, M.M., 2017. Understanding DNA replication by the bacteriophage T4 replisome. *Journal of Biological Chemistry* 292, 18434–18442.
- Casjens, S.R., 2011. The DNA-packaging nanomotor of tailed bacteriophages. *Nature Reviews Microbiology* 9, 647–657.
- Kreuzer, K.N., 2000. Recombination-dependent DNA replication in phage T4. *Trends in Biochemical Sciences* 25, 165–173.
- Llorente, B., Smith, C.E., Symington, L.S., 2008. Break-induced replication: What is it and what is it for? *Cell Cycle* 7, 859–864.
- Lo Piano, A., Martinez-Jimenez, M.I., Zecchi, L., Ayora, S., 2011. Recombination-dependent concatemeric viral DNA replication. *Virus Research* 160, 1–14.
- Masukata, H., Dasgupta, S., Tomizawa, J., 1987. Transcriptional activation of ColE1 DNA synthesis by displacement of the nontranscribed strand. *Cell* 51, 1123–1130.
- Mosig, G., 1998. Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annual Review of Genetics* 32, 379–413.
- Nossal, N.G., 1983. Prokaryotic DNA replication systems. *Annual Review of Biochemistry* 52, 581–615.
- Oliveira, L., Tavares, P., Alonso, J.C., 2013. Headful DNA packaging: Bacteriophage SPP1 as a model system. *Virus Research* 173, 247–259.
- Rao, V.B., Feiss, M., 2015. Mechanisms of DNA packaging by large double-stranded DNA viruses. *Annual Review of Virology* 2, 351–378.
- Salas, M., 1991. Protein-priming of DNA replication. *Annual Review of Biochemistry* 60, 39–71.
- Salas, M., Holguera, I., Redrejo-Rodríguez, M., De Vega, M., 2016. DNA-binding proteins essential for protein-primed bacteriophage ϕ 29 DNA replication. *Frontiers in Molecular Biosciences* 3, 37.
- Seco, E.M., Ayora, S., 2017. *Bacillus subtilis* DNA polymerases, PolC and DnaE, are required for both leading and lagging strand synthesis in SPP1 origin-dependent DNA replication. *Nucleic Acids Research* 45, 8302–8313.
- Seco, E.M., Zinder, J.C., Manhart, C.M., *et al.*, 2013. Bacteriophage SPP1 DNA replication strategies promote viral and disable host replication *in vitro*. *Nucleic Acids Research* 41, 1711–1721.

Relevant Websites

- <http://millardlab.org/bioinformatics/bacteriophage-genomes/>
Bacteriophage Genomes – millardlab.
- <http://genome.jouy.inra.fr/phagonaute/>
Phagonaute.
- <http://biodev.extra.cea.fr/virfam/>
VIRFAM – CEA/DSV.

Lytic Transcription

William McAllister, Rowan University School of Osteopathic Medicine, Stratford, NJ, United States

Deborah M Hinton, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, United States

Published by Elsevier Ltd.

Nomenclature

CTD C-terminal domain of a protein

EC Transcription elongation complex

gp Gene product

IC Transcription initiation complex

NTD N-terminal domain of a protein

Glossary

Lytic phage Phage that infect, takeover the host, make progeny, and lyse the cell.

Non-template strand The DNA sequence that will be the same as the RNA, 'the coding strand'.

Promoter DNA sequence that sets the start site for transcription initiation by RNA polymerase.

Template strand The DNA sequence that is copied by RNA polymerase during transcription. It will be the complement of the RNA sequence.

Introduction

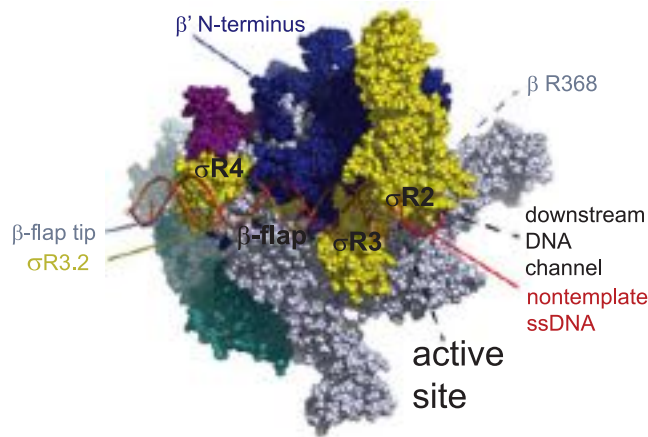
Lytic phage, which infect, make progeny, and lyse the cell, need to rapidly takeover host resources in order to reprogram the host for phage development. Regulation of phage and host gene expression through transcriptional control is the primary mechanism for this appropriation of the host machinery. To accomplish this, phage-encoded proteins can inhibit host gene expression, redirect the host RNA polymerase to phage genes, and/or function as a new RNA polymerase. Here we discuss mechanisms used by the model bacteriophages T4, T7, and N4, as well as phages within those families.

Transcription Initiation and Elongation by Host RNA Polymerase

When a lytic phage enters the cell, it must first manage the host RNA polymerase (RNAP), by redirecting it to phage promoters and/or by inhibiting some or all of its host activity. Bacterial RNAPs belong to the family of multi-subunit polymerases that are highly conserved throughout biology. They consist of a multisubunit core that contains the RNA synthesizing activity and a specificity subunit, called σ factor, that recognizes the promoter sequence and selects the start site for transcription. In bacteria, the core is typically composed of 5 proteins: β , β' , two α 's, and ω . σ factors include a primary σ , such as σ^{70} in *E. coli*, which is used during exponential growth, and alternate σ 's, which are used under different growth conditions or times of stress. The σ^{70} family, the largest group of σ factors, shares regions of sequence and structural homology. For the primary σ 's, these are divided into Regions (R) 1, 2, 3, and 4, which can be further subdivided based on function (R1.1, R1.2, etc.).

Structures of multi-subunit RNAPs, together with decades of biochemistry and genetics, have revealed molecular details about how bacterial RNAPs initiate, elongate, and terminate transcription. We use the well-studied *E. coli* RNAP as the example here (Figs. 1 and 2(A)). The β and β' subunits of the core create a crab-claw like structure, with the active site for RNA synthesis at the center of the claw. The N-terminal domain (NTD) of one α ($\alpha 1^{NTD}$) interacts with β , while $\alpha 2^{NTD}$ interacts with β' . Each α^{NTD} is connected to its C-terminal domain (CTD) via a flexible linker (Fig. 2(A), top; α CTDs are not seen in the structure in Fig. 1). The small ω subunit interacts with β' . σ^{70} interacts extensively with core, primarily through a portion of R2 that interacts with a coiled-coil motif within β' , but in addition, σ R1.1 lies within what will be the downstream DNA channel, σ R3.2 follows the channel through which the synthesized RNA will extrude, and σ R4 interacts with a structure within β , called the β -flap. The very C-terminus of σ^{70} (σ^{CT}) interacts with the β -flap tip (Fig. 2(A)).

A competent transcription complex is formed when *E. coli* σ^{70} -RNAP interacts stably with promoter DNA (Fig. 2(A)). Promoter recognition arises through interactions with double-stranded (ds) portions of the promoter: σ R4.2 with the -35 element (consensus sequence -35TTGACA^{-30}) and σ R3.0/2.4 with the -15TGnT^{-12} motif, which includes the extended -10 sequence -15TG^{-14} and the first position of the -10 element (consensus -12TATAAT^{-7}). The α CTDs can also interact with ds UP elements, AT-rich sequences between -40 and -60 . However, not all of these elements are needed for recognition, and there are two main promoter classes: $-10/-35$ promoters (with matches to the -35 and -10 elements), and extended -10 promoters (with a good match to the -15TGnTATAAT^{-7} motif, but no recognizable -35 element). A single-stranded (ss) transcription bubble (from -11 to $+3$) is bound by σ R2.3, which interacts with specific sequences within the nontemplate strand (consensus -11ATAAT^{-7}), while the template strand is within the active site. Transcription initiation begins as ribonucleoside triphosphates enter the active site. The synthesized RNA exits through the RNA exit channel, a region that is filled by σ R3.2 within the holoenzyme (Fig. 1). Thus, RNA



RNAP: core ($\beta \beta' \alpha 1^{\text{NTD}} \alpha 2^{\text{NTD}} \omega$) + σ

Fig. 1 Structure of *E. coli* RNAP [PDB 4IGC] modeled with promoter DNA. RNAP core subunits β (light gray-blue), β' (dark blue), ω (magenta), $\alpha 1^{\text{NTD}}$ (teal), $\alpha 2^{\text{NTD}}$ (light cyan), the σ specificity subunit $\sigma 70$ (yellow), and ds promoter DNA (template strand, red; non-template strand, orange) are shown. (The α^{CTD} s are not observed in the structure). Features discussed in the text are indicated.

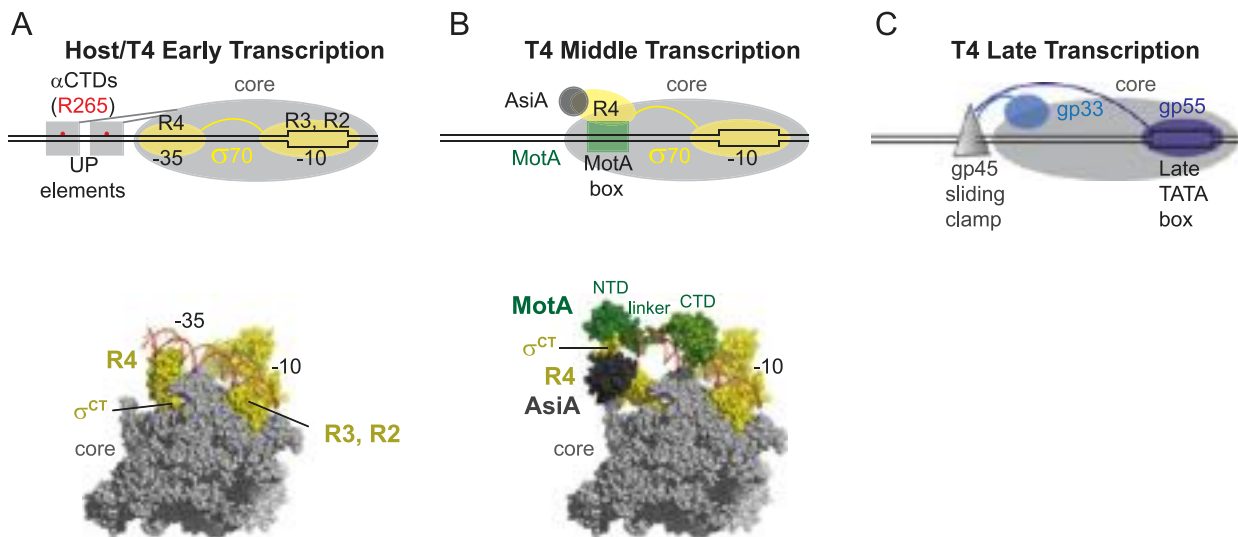


Fig. 2 Changes to *E. coli* RNAP as T4 infection proceeds. (A) Host RNAP used for T4 early transcription. (B) Host RNAP with T4 MotA and AsiA used for middle transcription. (C) Host RNAP with T4 gp33, gp55, and gp45 used for late transcription. In each case the schematic is shown above; for (A) and (B), structural models are shown below. RNAP core is designated in light gray, $\sigma 70$ in yellow, MotA in green, AsiA in dark gray, gp33 in Carolina blue, gp55 in purple and gp45 as the gray triangle. In the bottom panels, the positions of $\sigma 70$ R2, R3, and R4, α^{CTD} s, σ^{CT} , MotA (NTD, linker, and CTD), AsiA, and the promoter elements (-10 , -35) are shown.

extrusion destabilizes the interaction of σ with core, facilitating σ release. Typically, highly processive elongation continues until RNAP encounters an intrinsic termination signal, a stem-loop hairpin followed by a U-stretch, or a factor-dependent termination site, where the RNAP/DNA complex terminates, and RNAP dissociates from the DNA.

Developmental Pathways and Transcription Control by T7, T4, and N4

T4

Bacteriophage T4, a member of the Myoviridae family, does not encode its own RNAP. Instead it encodes factors that change the specificity of the host polymerase as infection proceeds, allowing it to program a temporal pattern of phage development from

early to middle to late promoters. In general, early proteins are needed for host takeover and middle transcription; middle proteins are needed for replication, for factors involved in late transcription, and for additional host takeover functions; and late proteins make up the virion structure. However, the functions of dozens of T4 genes, especially early, are still unknown.

Early promoters are active immediately after injection of the phage DNA. Host σ^{70} -RNAP is diverted to these promoters in two major ways: (1) early T4 promoters, which have near ideal matches to the consensus -35 and extended -10 sequence motifs, compete exceptionally well with host promoters for the available RNAP, and (2) modification of the RNAP by the T4 Alt protein, which is present in the phage head and is injected with the DNA, further increases the activity of early promoters relative to those of the host. Alt ADP-ribosylates one of the α subunits at R265, a residue that is needed for α CTD recognition of the UP element promoter motif (Fig. 2(A)). As approximately $\sim 70\%$ of transcription arises from the very strong host ribosomal promoters and UP element recognition contributes to ribosomal promoter strength, it is thought that this modification serves to immediately suppress a substantial amount of host transcription after T4 infection. It is not known which α is modified by Alt.

T4 early transcription leads to the expression of phage early genes. Two of these encode MotA, a DNA-binding transcription activator, and AsiA, a transcription co-activator, which are required for T4 middle promoter activation through a process called σ appropriation (Fig. 2(B)). This process, which starts about 1 min after infection, redirects RNAP to T4 middle promoters, which contain the σ^{70} -10 element and a MotA box motif [consensus sequence 5' (a/t)(a/t)(a/t)TGCTTA3'] centered at -30 . AsiA first binds to R4 of free σ^{70} to form a complex in which this portion of σ is completely restructured. Consequently, upon binding to core, σ is unable to bind the β -flap or to interact with the -35 element of σ^{70} -dependent promoters. (Curiously, the binding of AsiA does not appear to inhibit T4 early promoter activity). Importantly, σ^{CT} , which would normally be held by the β -flap tip (Fig. 2(A)), is now available for interaction with its binding partner, MotA^{NTD} (Fig. 2(B)). The MotA box motif within the middle promoter is engaged by the MotA^{CTD}, which binds to the downstream portion of the MotA box in the major groove while the MotA^{linker} (between the NTD and CTD) interacts with the upstream portion in the minor groove. The DNA binding motif of MotA^{CTD} represents a previously unrecognized way to interact with DNA – a double wing saddle structure in which both wings lie within the major groove. Overall the binding of AsiA and MotA serves to switch promoter recognition from a standard $-35/-10$ promoter to the middle MotA box/ -10 promoter.

In addition to activation of middle promoters, T4 middle genes are also expressed by the extension of early transcription from early genes into downstream middle genes. Although this extension requires T4 protein synthesis, no specific T4 product, such as an anti-terminator, has been identified. Rather it is thought that the process of translation itself serves to promote either transcription elongation and/or stabilizes the extended RNA.

All cytosines within the T4 genome are modified by the addition of a 5-hydroxymethylated, α - or β -glucosylated moiety, whose presence protects the phage DNA from *E. coli* and T4 nucleases. Although this modification is not required for T4 transcription, it does provide significant advantages for phage gene expression. In the case of σ appropriation, modeling of the modification within the MotA^{linker-CTD}/DNA crystal structure indicates that the glucosyl sugar will make significant contact with the MotA^{CTD} saddle. Binding analyses demonstrate that the presence of the 5'-hydroxymethyl, glucosyl moiety increases MotA affinity for the MotA box by > 100 -fold.

The T4 Alc protein also uses this modification to distinguish between host and phage DNA. Alc, a T4 early protein of 16.7 kDa, specifically promotes transcription termination of unmodified DNA. This function is thought to arise through an interaction of Alc with a nonconserved region of the β subunit of *E. coli* RNAP (Dispensable Region 1) near a residue (R368) that lies close to the RNAP active site (Fig. 1). Interestingly, Alc function requires rapidly elongating RNAP, consistent with the optimal conditions for T4 infection in rapidly growing cells.

The activation of middle promoters also results in middle products that severely disrupt host transcription. First, T4 nucleases, encoded by middle genes *denA* and *denB*, specifically degrade unmodified DNA, resulting in $\sim 50\%$ decrease in the host genome and effectively destroying the ability of the host to express its DNA. In addition, the T4 Mod protein completes the ADP-ribosylation of RNAP by modifying the other α subunit (again at residue 265). This modification then prevents any interaction of the α CTDs with UP elements, presumably further inhibiting host promoter activity.

About 5 min after infection at 37°C , T4 late promoters become active (Fig. 2(C)). These promoters are recognized by *E. coli* RNAP core together with two T4 middle products that together constitute a new σ factor: gp55 and gp33. Gp55, a distant member of the σ^{70} family, retains an R2 motif that interacts with the T4 late -10 promoter element, $^{-13}\text{TATAATA}^{-6}$. Gp33 bears no sequence resemblance to σ factors, but nevertheless interacts with the β -flap of core, just like σ^{70} R4. However, gp33 does not interact with the DNA, and late promoters do not have an upstream element like the -35 element of σ^{70} -dependent promoters. Thus, both middle and late T4 transcription involves employing a 'new' σ factor, either by reconfiguring σ^{70} with MotA and AsiA (middle) or replacing σ^{70} with gp55/gp33 (late).

Besides gp55/gp33, late promoter activation is also dependent on DNA replication. This connection stems from the requirement for gp45, a replication protein, called the 'sliding clamp' that can interact with either T4 DNA polymerase, gp33, or gp55. The structure of gp45 is a triangular donut, which encircles the dsDNA (Fig. 2(C)). Gp45 is loaded onto the DNA by the T4 clamp loader complex (gp44/gp62) and then moves along the DNA. During replication the interaction between gp45 and DNA polymerase 'clamps' the replication machinery onto the DNA, insuring processivity. However, the interaction between gp45 and gp33/gp55 allows RNAP to also move along the DNA, scanning for a late promoter. Consequently, gp45 behaves as a transcriptional enhancer that lacks a specific enhancer binding motif. The connection between late transcription and DNA replication allows the phage to produce an amount of virion assembly proteins that is consistent with the level of replicated phage DNA.

Besides gp33/gp55, other T4 middle products, RpbA and DsbA, bind to RNAP during late transcription. *In vitro* the presence of RpbA together with ADP-ribosylation of the α CTDs improves late promoter activity while diminishing middle promoter function. However, RpbA is not essential, and whether this is its function *in vivo* is not known. DsbA is a ~ 10 kDa protein that binds weakly upstream of some late promoters. Despite an early report that the presence of DsbA enhances transcription at these promoters, it is now reported that it has no effect on late transcription *in vitro*. However, the *dsbA* ORF overlaps the 3' end of gene 33, suggesting that translational coupling of these two genes could influence late transcription *in vivo*.

T7

Bacteriophage T7 is the prototype of the *Podoviridae*. The hallmark of this group is that most encode an RNAP that is responsible for transcription of most of the phage DNA.

The 40 kb T7 genome encodes Class I, Class II, and Class III genes (Fig. 3(A)). Class I genes are responsible for synthesis of the phage RNA polymerase and proteins required to protect the DNA from degradation (among other functions). Class II genes are largely involved in DNA replication and recombination. Class III genes encode proteins required for virion assembly, packaging of phage DNA, and lysis of the cell. All transcription proceeds unidirectionally, and only one strand of the DNA is transcribed. Initially, only the first ~ 850 bp of DNA is injected into the cell; the remainder is transported by transcription, first by the host RNAP and subsequently by the phage RNAP.

Early transcription

Transcription of the Class I genes by the host RNAP is initiated near the left end of the genome from a cluster of three strong promoters that have evolved to be optimal under different intracellular conditions (Fig. 3(A)). Termination occurs at the end of the early region at TE, which is similar to intrinsic terminators for *E. coli* RNAP and encodes an RNA with a potential stem-loop structure followed by a run of Us. The polycistronic early region transcript is processed by RNase III into individual mRNAs. Inhibition of transcription by the host RNAP involves the actions of gp0.7, a protein kinase that phosphorylates β' (as well as other host proteins), resulting in increased factor-dependent transcription termination, and by gp2, which binds to σ^{70} R1.1, preventing its movement from the downstream DNA channel (Fig. 1), thereby blocking initiation. In addition, nucleases encoded by gp3 and gp6 efficiently degrade the host genome; over 80% of phage DNA is derived from host DNA.

Phage RNAP and late gene transcription

T7 RNAP is related structurally and by sequence homology to other members of the Pol I family of nucleotide polymerases, which include DNA Pol I, reverse transcriptase, and mitochondrial RNA polymerase. A catalytic core domain is conserved within this class, while other regions are specific to the transcription functions of the enzyme. The preservation of the core among these varied polymerases, which utilize different templates and substrates, is remarkable; the relationship to mitochondrial RNAP is thought to reflect the origin of this organelle as an early prokaryotic symbiont of eukaryotic cells. The T7 RNAP consists of 883 amino acids (Figs. 3(B) and 4(A)). The 'cupped right hand' finger, thumb and palm subdomains, containing the catalytic core, are within the conserved C-terminal region while the N-terminal domain is involved in promoter recognition, elongation, and termination.

In addition to its role in transcription, the phage RNAP is involved in other essential functions, including transport of the phage DNA into the cell, priming of DNA replication, and maturation and packaging of the DNA.

Promoters for the phage RNAP are related to a consensus sequence that extends from -17 to $+6$, where initiation is at $+1$ (Fig. 3(C)). Whereas the class II promoters differ from the consensus at a number of positions, class III promoters preserve this sequence. Control of late transcription involves T7 lysozyme (Lys, gp3.5), a dual function protein that includes an amidase that is involved in cell lysis (in its C-terminal portion) and a separable regulatory function that is involved in modulating activity of the RNAP during initiation and pausing (in the N-terminal portion). The mechanism of action of Lys is unclear. Binding of Lys to the RNAP alters the configuration of the C-terminus of the T7 RNAP, which is in close proximity to the active site, and increases K_m for the initiating NTPs, suggesting that it may exert its effect by preferentially inhibiting initiation at the weaker class II promoters.

The transcription cycle

T7 RNAP goes through the same steps in the transcription cycle as the multi-subunit RNAPs. These include promoter binding and initiation, promoter clearance, elongation, and termination. Extensive genetic, biochemical, and structural analyses have provided insight as to how these steps occur in exquisite detail.

Promoter binding involves interaction of the upstream region of a promoter with elements in the N-terminal domain of the RNAP (e.g., the AT-rich recognition loop and the intercalating loop (β -IH)) and with the "specificity loop" (a beta hairpin that projects from the fingers subdomain in the C-terminal core). Meanwhile, the downstream region of the promoter, which includes the start site for initiation ($+1$), interacts with elements in the catalytic core domain. Binding of the polymerase is accompanied by bending of the promoter that results in strand separation ("melting") of the promoter from -4 to $+3$, exposing the template strand bases in the initiation region and allowing initiation and the incorporation of the first few nt into the nascent transcript.

The initiation complex (IC) can only accommodate up to three base pairs of the growing RNA:DNA hybrid. As the transcript is elongated from 3 to 7 bp, steric clash with the N-terminal domain results in a rotation of this domain as a rigid body by about 40 degrees, without promoter release. The resulting strain in the complex can lead to release of abortive initiation products before the

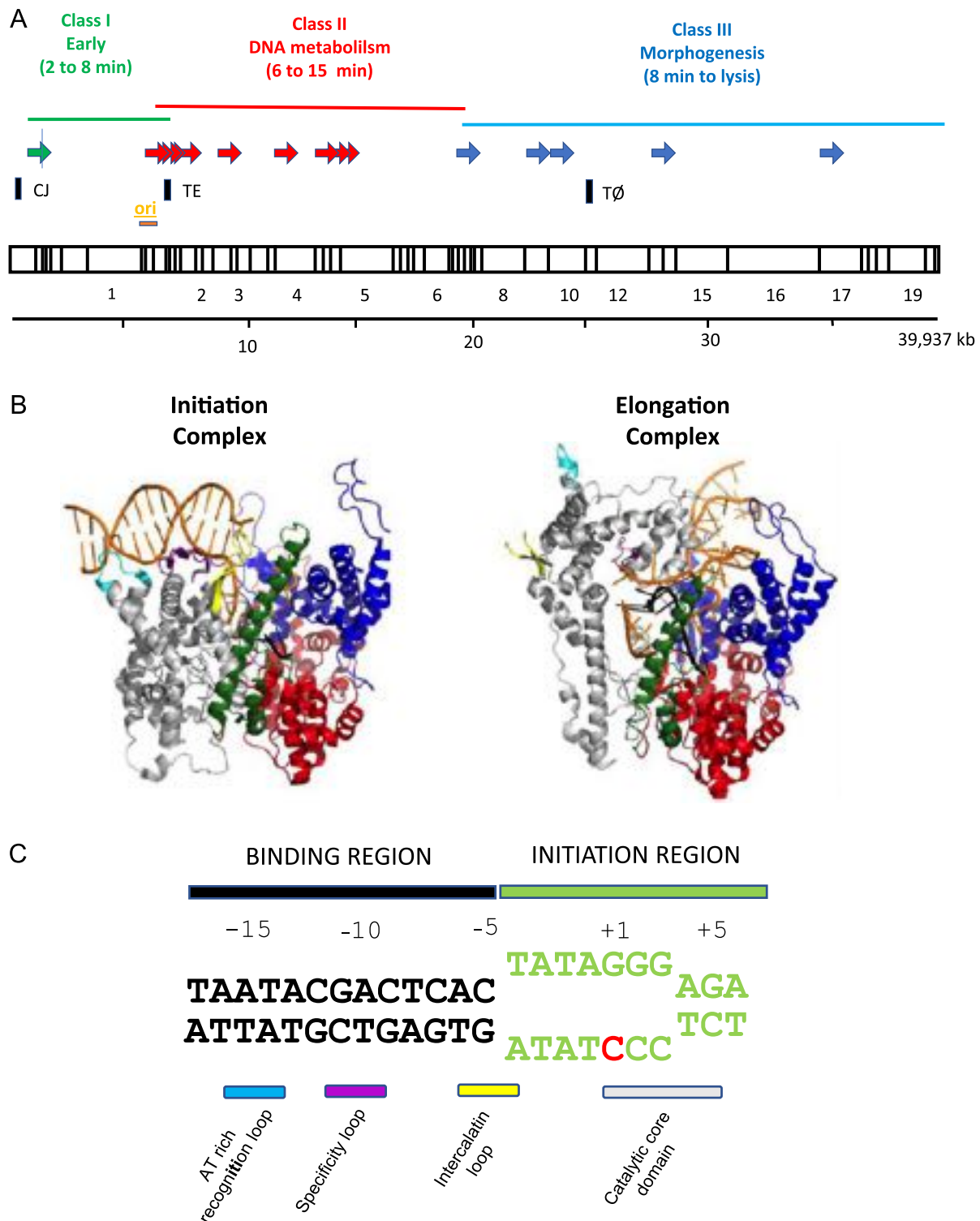


Fig. 3 T7 transcription. (A) Transcription map of T7. All transcription is from left to right. Locations of promoters and the direction of transcription for host (green) and phage (red, blue) RNAPs are indicated by arrows. Termination/pause signals are denoted by vertical bars. The primary origin of DNA replication (ori) is indicated. The genetic and physical maps are at the bottom; not all genes are identified in this figure. (B) Structural changes that occur during the transition from IC (PDB 1QLN) to an EC (PDB 1MSW). The two structures are oriented similarly with regard to subdomains in the catalytic C-terminal core: palm (red), fingers (blue), and thumb (green). The N-terminal domain is in gray. Elements involved in upstream promoter recognition include the specificity loop (purple), the AT-rich recognition loop (cyan), and β -IH, the intercalation loop (yellow). (C) Consensus promoter sequence for T7 RNAP. Key elements of the RNAP that interact with the binding and initiation regions of the promoter are indicated. The region that is melted open during promoter binding is indicated by displaced template (bottom) and non-template (top) strands of the DNA. The initiation site is at +1.

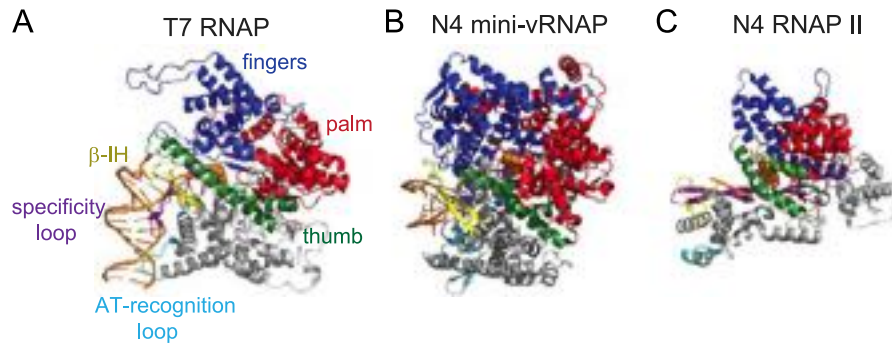


Fig. 4 Comparison of the structures of the single-subunit RNAPs encoded by T7 and N4 complexed with DNA. (A) T7 RNAP (PDB 1CEZ). (B) N4 mini-vRNAP (PDB 3C3L). (C) N4 RNAP II (PDB 6DT7). In each case, the right ‘cupped hand’ composed of fingers (dark blue), palm (red), and thumb (green), the AT-recognition loop (cyan), the specificity loop (purple), and the β -IH (“intercalation loop”, yellow) are shown. DNA is shown in orange with position -1 indicated as a sphere.

transition to a stable elongation complex (EC) occurs. Notably, as the length of the transcript approaches ~ 8 nt (the length of the hybrid in the EC), the 5' end of the RNA comes into contact with the specificity loop, which is involved in upstream promoter contacts. This is reminiscent of the situation with *E. coli* RNAP in which the lengthening transcript displaces σ R3.2 (see above), suggesting that such interactions may be a common feature of promoter release. Beyond this point (after synthesis of 9–12 nt), release of promoter contacts occurs, and is accompanied by major conformational changes that result in the formation of an RNA exit tunnel, and a stable EC (**Fig. 3(B)**).

Elongation

Once the complex has isomerized into the EC, it is highly stable and will transcribe many kb of template without dissociation until it encounters a pause or a termination signal. Each cycle of nucleotide addition is accompanied by opening and closing of the active site in the catalytic core, and translocation of the EC along the template. Recognition of the next incoming nucleotide occurs in the “open” state by the formation of a nascent base-pair with the exposed template strand base. Closing of the active site places the incoming nucleotide and template base in a position that is suitable for catalysis. Formation of the phosphodiester bond results in the release of PP_i, which destabilizes the closed complex, triggering a return to the open complex and translocation of the EC to the next position along the template.

Termination

T7 RNAP pauses or terminates at two types of signals. The termination signal $T\phi$, located downstream of gene 10, resembles intrinsic terminators for *E. coli* RNAP such as TE, but has a longer stem-loop structure and a longer run of U residues downstream. Termination at this site is not complete, and read through transcription is required for expression of genes 11 and 12. The mechanism of termination at $T\phi$ is not clear, but may involve steric clash of the stem-loop structure with the RNA exit pore, and/or induced slippage of the transcript at the run of U residues, resulting in misalignment in the active site.

A second type of signal is present in the junction of replicating T7 DNA concatemers (CJ). This signal does not encode an RNA with secondary structure but involves recognition of a specific sequence (TATCTGTT) and results in pausing rather than termination. The efficiency and duration of the pause is dramatically increased in the presence of T7 Lys, and the halted complex is thought to recruit other phage proteins that are involved in processing and packaging of the DNA. The mechanism of pausing is not clear, but is likely to involve recognition and binding of the signal in the nontemplate strand. This feature may be a common mechanism for pausing/termination, as *E. coli* RNAP has also been found to recognize pause sites in this strand.

N4

The lytic podoviridae phage N4 employs the host multi-subunit RNAP and 2 phage-encoded RNAPs, both of which are distantly related to the T7 family of single-subunit RNAPs. Expression of early N4 genes depends upon a virion-encapsulated RNAP (vRNAP) that is injected into the host upon infection. This injection results in the entry of the first 500 bp of the left end of the linear N4 DNA. The DNA is then constrained/supercoiled by the action of host gyrase to generate a template that extrudes hairpins on the template and non-template strands at three vRNAP promoter sequences. Binding of the *E. coli* single-stranded binding protein (Eco SSB) generates a single-stranded (ss) non-template strand and stabilizes the template strand hairpin, which is recognized through the interaction of vRNAP with bases -11 , -10 , -8 , and $+1$. Thus, Eco SSB has been described as an ‘architectural transcription factor’. Once transcription initiates, Eco SSB also binds the exiting transcript. vRNAP lacks a domain present in T7 RNAP, which binds RNA, separating it from the template. Consequently, it is thought that the Eco SSB/transcript interaction serves to replace this function.

Despite the large size of vRNAP (3500 aa), only 1/3 of the protein (1106 aa) is required for transcription. Structural analyses indicate that this mini-vRNAP shares common structural features of the ‘cupped right hand’ single-subunit RNAPs (**Fig. 4(B)**). This structural similarity exists despite practically no sequence homology. In both vRNAP and T7 RNAP (**Fig. 4(A)**), the DNA is recognized similarly: a specificity loop interacts with the major groove of the ds DNA (the hairpin in the case of vRNAP promoters), the β -IH (intercalating loop) recognizes the ds/ss DNA junction present at the upstream edge of the transcription bubble, and the upstream recognition element recognizes position –11 in the N4 promoter or the upstream AT-rich region of a T7 promoter (**Figs. 3(C)** and **4**).

Once the N4 early genes are transcribed, the remaining portion of the N4 genome is injected and middle gene transcription commences. Middle transcription requires a second N4 RNAP, RNAP II, encoded by genes 15 and 16. Like vRNAP, RNAP II is a minimal, T7-like RNAP (**Fig. 4(C)**). It bears much structural and some sequence homology to the T7 protein. RNAP II associates tightly with the cytoplasmic membrane, and transcription requires either denatured DNA or a membrane-bound DNA containing the N4 ssDNA binding protein, gp2. While either substrate is active, only the gp2/membrane/DNA substrate results in specific transcription. It has been proposed that another N4 protein, perhaps gp1, denatures the promoter region, which then allows gp2 to bind and recruit RNAP II.

In a perverse twist, N4 late transcription, which generates the RNAs needed for the production of morphological proteins, requires the host RNAP. Besides host σ^{70} -RNAP, late transcription also requires a middle gene product, the N4 SSB protein. Although N4 SSB is required for N4 replication, a distinct region of the protein at the C-terminus interacts with the β' subunit of RNAP. Thus, as in T4 late transcription, a part of the N4 late transcriptional machinery is directly connected to DNA replication.

Inhibition of Host RNAP by T4, T7, and N4, and Other Members of Those Families

As detailed above for T4 and T7, phages produce products to eliminate or down-regulate transcription by the host RNAP. Recent work has identified novel phage inhibitors. Xp10, a Siphoviridae phage of the plant pathogen *Xanthomonas oryzae*, encodes a protein (p7), which both inhibits recognition of host –10/–35 promoters and functions as an antiterminator at intrinsic (factor-independent) termination sites. This results in efficient transcription from the phage extended –10 promoters while host transcription is inhibited. The primary binding site of p7 is the N-terminus of the β' subunit (**Fig. 1**), but the β -flap domain and the ω subunit, have also been implicated in transcription inhibition and anti-termination, respectively.

The *Thermus thermophilus* phage P23–45 produces two inhibitors of *T. thermophilus* RNAP, gp76 (an early gene product) and gp39 (a middle gene product). Gp76 interacts with the downstream DNA channel of RNAP (**Fig. 1**) and the path that would be occupied by the upstream portion of the ss template strand (–11 to –4) within the transcription bubble. Gp39 binds to the β -flap and σ R4, thus interfering with the σ R4/ β -flap interaction needed for –10/–35 promoter activity. This facilitates phage promoter activity since P23–45 middle and late promoters belong to the extended –10 class.

Finally, gp67 of G1, a Myoviridae phage that infects *Staphylococcus aureus*, inhibits the host by yet another mechanism. In this case, gp67 binds to σ R4, but this binding does not interfere with its ability to recognize the –35 element. Instead this binding allows gp67 to interfere with α CTD binding to UP elements between –40 and –60. Since the strong ribosomal promoters rely on α CTD/UP element interaction and ribosomal transcription accounts for the majority of overall bacterial transcription, this inhibition leads to a major down-regulation of transcription in the host.

Summary

Despite their relatively small genomes, lytic phages employ highly sophisticated and varied strategies to maximize transcription of their own DNA while limiting host gene transcription. Understanding these processes at a molecular level has deepened our understanding of transcription mechanisms used by both multi-subunit and single subunit (Pol I-like) RNAPs. In particular, phages are providing a rich resource for studies of RNAP inhibitors. Interest in the detailed mechanisms of these inhibitors has surged in recent years as researchers search for ways to combat emerging and reemerging bacterial pathogens. It is likely that a vast array of other mechanisms will be discovered as new phage functions are identified.

Further Reading

- De Smet, J., Hendrix, H., Blasdel, B.G., Danis-Wlodarczyk, K., Lavigne, R., 2017. Pseudomonas predators: Understanding and exploiting phage-host interactions. *Nature Reviews Microbiology* 15, 517–530.
- Decker, K.B., Hinton, D.M., 2013. Transcription regulation at the core: Similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annual Review of Microbiology* 67, 113–139.
- Geiduschek, E.P., Kassavetis, G.A., 2010. Transcription of the T4 late genes. *Virology Journal* 7, 288.
- Hinton, D.M., 2010. Transcriptional control in the prereplicative phase of T4 development. *Virology Journal* 7, 289.
- James, T.D., Cardozo, T., Abell, L.E., et al., 2016. Visualizing the phage T4 activated transcription complex of DNA and E. coli RNA polymerase. *Nucleic Acids Research* 44, 7974–7988.
- Lenneman, B.R., Rothman-Denes, L.B., 2015. Structural and biochemical investigation of bacteriophage N4-encoded RNA polymerases. *Biomolecules* 5, 647–667.

- McAllister, W.T., Raskin, C.A., 1993. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Molecular Microbiology* 10, 1–6.
- Molineux, I., 2005. *The T7 Group*. Oxford: Oxford University Press.
- Molodtsov, V., Murakami, K.S., 2018. Minimalism and functionality: Structural lessons from the heterodimeric N4 bacteriophage RNA polymerase II. *Journal of Biological Chemistry* 293, 13616–13625.
- Saecker, R.M., Record Jr., M.T., Dehaseth, P.L., 2011. Mechanism of bacterial transcription initiation: RNA polymerase – Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *Journal of Molecular Biology* 412, 754–771.
- Steitz, T.A., 2009. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Current Opinion in Structural Biology* 19, 683–690.
- Sutherland, C., Murakami, K.S., 2018. An introduction to the structure and function of the catalytic core enzyme of *escherichia coli* RNA polymerase. *EcoSal Plus* 8.
- Washburn, R.S., Gottesman, M.E., 2015. Regulation of transcription elongation and termination. *Biomolecules* 5, 1063–1078.

Lysogeny

Keith E Shearwin and Jia Q Truong, The University of Adelaide, Adelaide, SA, Australia

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Lysogen A bacterial cell carrying one or more bacteriophage genomes, either integrated into the host chromosome or existing as independently replicating extra-chromosomal elements. A true lysogen is able to exit lysogeny and re-enter lytic development.

Lysogenic conversion The situation where a bacterial host acquires a new trait as a direct result of the expression of a gene or genes encoded by a prophage.

Prophage The lysogenic form of a bacteriophage.

Prophage induction The process of a bacteriophage exiting lysogeny and entering lytic development, resulting in lysis of the host cell.

Temperate bacteriophage A bacteriophage capable of entering either lytic development or lysogeny upon host cell infection.

Terminally redundant DNA DNA that contains repeated sequences at each end called terminal repeats. These ends are used to join the ends of the linear DNA to form circular DNA.

Transcriptional interference The suppressive influence of one transcriptional process, directly and *in cis* on a second transcriptional process.

Introduction

Bacteriophages (phages) are obligate bacterial parasites. Many phages exhibit a purely lytic lifecycle, where following infection of a susceptible host, the phage DNA is replicated and the phage hijacks the bacterium's cellular machinery to produce new virion particles, which are released upon lysis of the host cell. Other phages, the so-called temperate phages, are able to make a developmental 'decision' between two developmental regimes, the lytic and lysogenic cycles. In lysogeny, the phage persists indefinitely inside the bacterial host as a prophage, with the phage genome either integrated into the bacterial chromosome or existing extra-chromosomally (Fig. 1). Lysogeny is a non-bacteriocidal state where the bacteriophage genome replicates without virion production. In this state, most of the phage genome is transcriptionally inactive, as the expression of lytic functions would be lethal to the host cell. A bacterial host cell lysogenic for a particular phage is also immune to further infection by the same phage.

Filamentous phage, which have ssDNA genomes packaged into filament-like virions, are able to replicate without killing the host. Among the filamentous phage, some integrate in the host chromosome, while non-integrative filamentous phage replicate exclusively as extrachromosomal elements or episomes. Both classes of filamentous phages continually shed viral particles without host cell death, even while inserted into the bacterial genome as a prophage. Thus, bacteria infected permanently with filamentous phage represent a form of lysogeny, but don't meet the definition of true temperate phage, in that there is no stage which brings about host cell lysis.

In this article, we discuss specific examples of how true temperate phage establish and maintain lysogeny. The phenotypic effects of lysogeny on the host bacterium and the evolutionary impacts of lysogeny are also discussed.

Why Lysogeny?

For a phage to successfully propagate, it must coax its host into manufacturing new virions. These new virions are released through host cell lysis, and need to infect another host bacterium in order for the phage to continue to propagate. From the phage perspective, during times of abundant resources and the presence of many host cells, lytic development allows for large numbers of phages to be produced, to subsequently infect other host cells and thus continue the cycle. However, when resources and/or susceptible host cells are scarce, the phage can benefit from entering lysogeny until conditions improve. The process of making a choice between the lytic and lysogenic pathways following host cell infection, is covered in more detail in another article (Golding). While there must be some (small) fitness cost in replicating extra genomic material, many phages have evolved such that lysogeny can provide a selective advantage to the host cell in a number of ways. Such benefits may include carrying genes which confer some growth advantage to the host, or by providing protection from subsequent infection, and potential lysis, by phage from the same or different families.

Persistence of DNA

Integration Into the Chromosome

The long term persistence of phage DNA inside their bacterial host cell is the key feature of a lysogen. Phage have evolved many mechanisms to achieve this long term persistence. In many cases, such as the paradigm bacteriophage λ , the phage DNA is

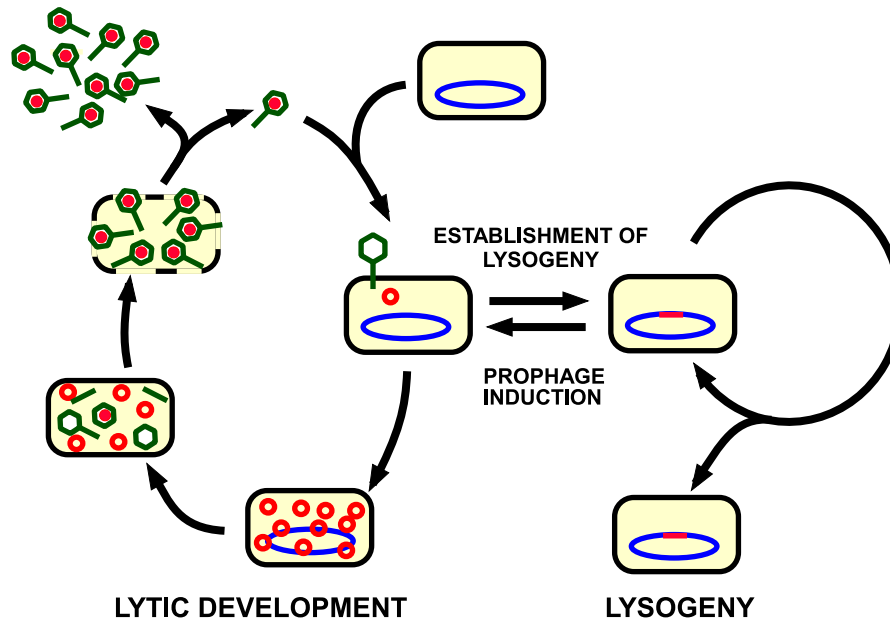


Fig. 1 Lytic and lysogenic developmental cycles of a temperate bacteriophage. A bacteriophage infects a bacterial host by injecting its DNA (red) into the host cell. In the lytic lifecycle, phages produce new virions, lysing the cell to release them. The lysogenic lifecycle has the phage DNA integrated into the bacterial chromosome or persisting extra-chromosomally. This state is stable, with prophage DNA maintained in daughter cells over subsequent generations. Most temperate phage can efficiently exit lysogeny to reenter the lytic phase in response to specific environmental cues.

integrated into the bacterial chromosome via site-specific recombination (**Fig. 2**). The phage-encoded integrase protein is a site-specific recombinase that recognizes attachment (*att*) sites on both the phage genome (*attP*) and the bacterial chromosome (*attB*). The integrase catalyses recombination between the *attP* and *attB* sites. This process generates two new *att* sites, *attL* and *attR* that flank the phage genome. When the phage exits lysogeny, the excision machinery, consisting of at least the integrase and a phage-encoded recombination directionality factor (RDF), and often with additional host factors, recognizes the *attL* and *attR* sequences to catalyse the reverse reaction to excise the prophage genome.

In general, integrase proteins belong to one of two families, the serine integrases or the tyrosine integrases. These proteins use a catalytic serine or tyrosine, respectively, to carry out the recombination reaction. Serine integrases create a double stranded break in both the phage and bacterial chromosome, and mediate strand rearrangement to bring them into the recombinant configuration, followed by strand ligation. Tyrosine recombinases, on the other hand, mediate recombination via a Holliday junction intermediate. Single stranded cuts are made in DNA and rejoined to single stranded DNA from the recombinant partner. Both types of integrases have been used extensively for biotechnology and molecular biology applications.

Some temperate phage integrate their DNA less specifically into the bacterial genome. Upon infection of a host cell, phage Mu initially integrates at random into the host chromosome. Only after integration is a decision on lytic or lysogenic development made. If the lytic pathway is followed, the phage genome undergoes successive rounds of replicative transposition. In each round, Mu DNA is duplicated, and a range of host genome rearrangements (deletions, inversions, translocations) are introduced. Viral genomes, along with variably sized segments of the host genome, are packaged directly from the integrated copies dispersed around the scrambled host genome. Similar to other temperate phages, the Mu lysogenic pathway (termed the latent pathway in Mu) is followed if the Repc repressor protein accumulates to a sufficiently high concentration. Repc competes for overlapping operators with DDE-recombinase A, which binds to an internal activation sequence (IAS), to initiate the lytic pathway.

In Mu, the initial integration process occurs by random transposition (**Fig. 3**). This process utilizes transposition machinery consisting of phage-encoded and host-supplied components to cleave the Mu DNA and insert it into target sites in the host genome with little or no target specificity. When Mu injects its DNA into a new host cell, its DNA is linear and flanked by DNA that was acquired from the previous host from which it was assembled. The phage supplied N protein, injected alongside DNA, non-covalently circularizes the DNA. To integrate Mu DNA into the chromosome, the phage DNA is nicked by MuA transposase to generate two 3' hydroxyl (-OH) ends on both strands at the Mu-Host DNA junctions. The 3'-OH ends attack two phosphodiester bonds spaced 5 bps apart, which join the 3'-OH to the 5' phosphate of the host DNA at the attack site. The flanking segments of DNA from the previous host are removed and the DNA gap is repaired.

A natural consequence of phage DNA integration into the host chromosome is that host DNA replication will also replicate the prophage DNA. Thus, all daughter cells of a lysogen will inherit the prophage. This form of lysogeny can be extremely stable, with very low rates of 'spontaneous' induction.

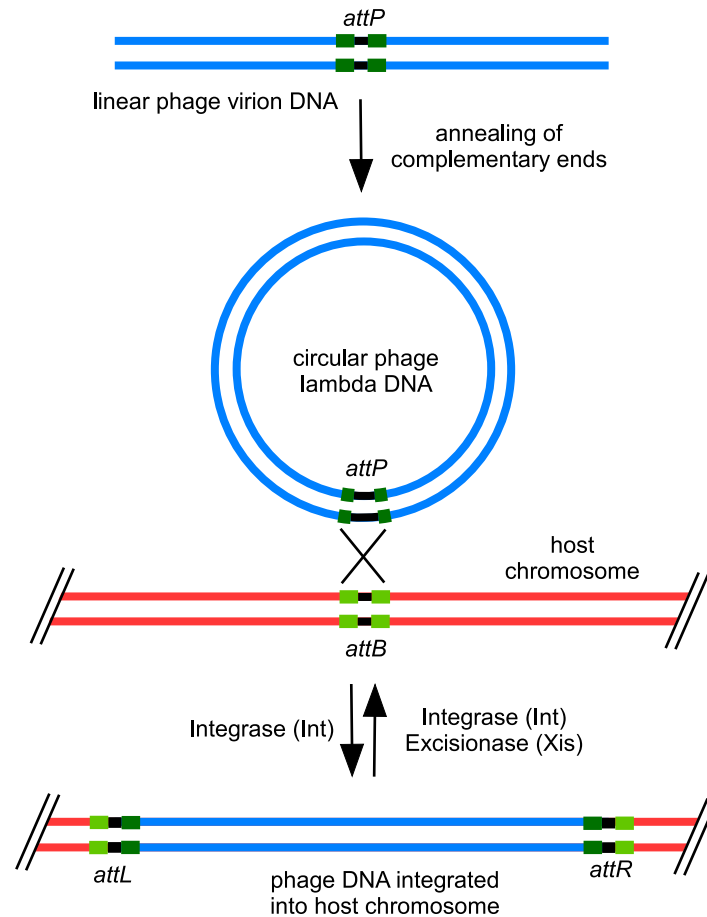


Fig. 2 Prophage DNA can be maintained by site-specific integration into the bacterial chromosome. This reaction is mediated by a phage-encoded integrase protein (Int). Initially upon injection into the host, the phage DNA is linear. Circularization of the viral DNA occurs through annealing of complementary ends, followed by ligation. Integrases recognize the attachment sites *attP* on the viral DNA (dark green) and *attB* on the host chromosome (light green) to catalyse the site-specific recombination. This generates two new *att* sites, *attL* and *attR* which flank the viral DNA in the chromosome. The attachment sites have a conserved core, shown in black. To return to the lytic cycle, the viral DNA must be excised from the chromosome through a reversal of the site-specific recombination between *attL* and *attR*, mediated by the excision machinery, which include the Int and excisionase (or recombination directionality factor RDF) proteins.

Extrachromosomal Persistence of DNA

Some prophages can also maintain their DNA separate from the host chromosome as a plasmid, either circular or linear. A well-characterized example of a circular extrachromosomal lysogen is bacteriophage P1. Infectious particles of P1 consist of a large (93 kb) linear double stranded genome with terminal redundancy of 10–15 kb. Upon being injected into the P1 host *E. coli*, the ends of the DNA undergo homologous recombination to circularize into an autonomous plasmid. In lysogeny, the prophage/plasmid copy number is maintained to be the same as the number of bacterial chromosomes. This is achieved through the expression of P1 proteins involved in origin-specific initiation and partitioning. The phage also uses a toxin-antitoxin system to kill daughter cells that have failed to receive a P1 molecule upon cell division. The antitoxin is degraded rapidly and needs to be continually expressed to counteract the toxin's activity. Loss of the P1 molecule in daughter cells results in a rapid depletion of antitoxin to a level where the more stable toxin will kill the host cell.

Some prophages, such as *E. coli* bacteriophage N15 maintain their DNA as a linear plasmid. Similar to linear eukaryotic chromosomes, this presents a problem as DNA polymerase fails to replicate all the way to the ends of the DNA. Bacteriophage N15 belongs to a group of phages that express a protelomerase to create covalently closed hairpins. After entering the cell, N15 DNA circularizes through its cohesive termini. The phage encoded protelomerase cuts an inverted sequence in the phage genome and ligates the phosphodiester bonds of each strand to form two closed hairpin ends (Fig. 4).

The replication of such as linear plasmid also presents mechanistic challenges and requires prophage encoded proteins. In the case of bacteriophage N15, the phage repA protein is necessary for replication. This multi-domain protein contains domains that resemble prokaryotic primases and helicases required for DNA replication.

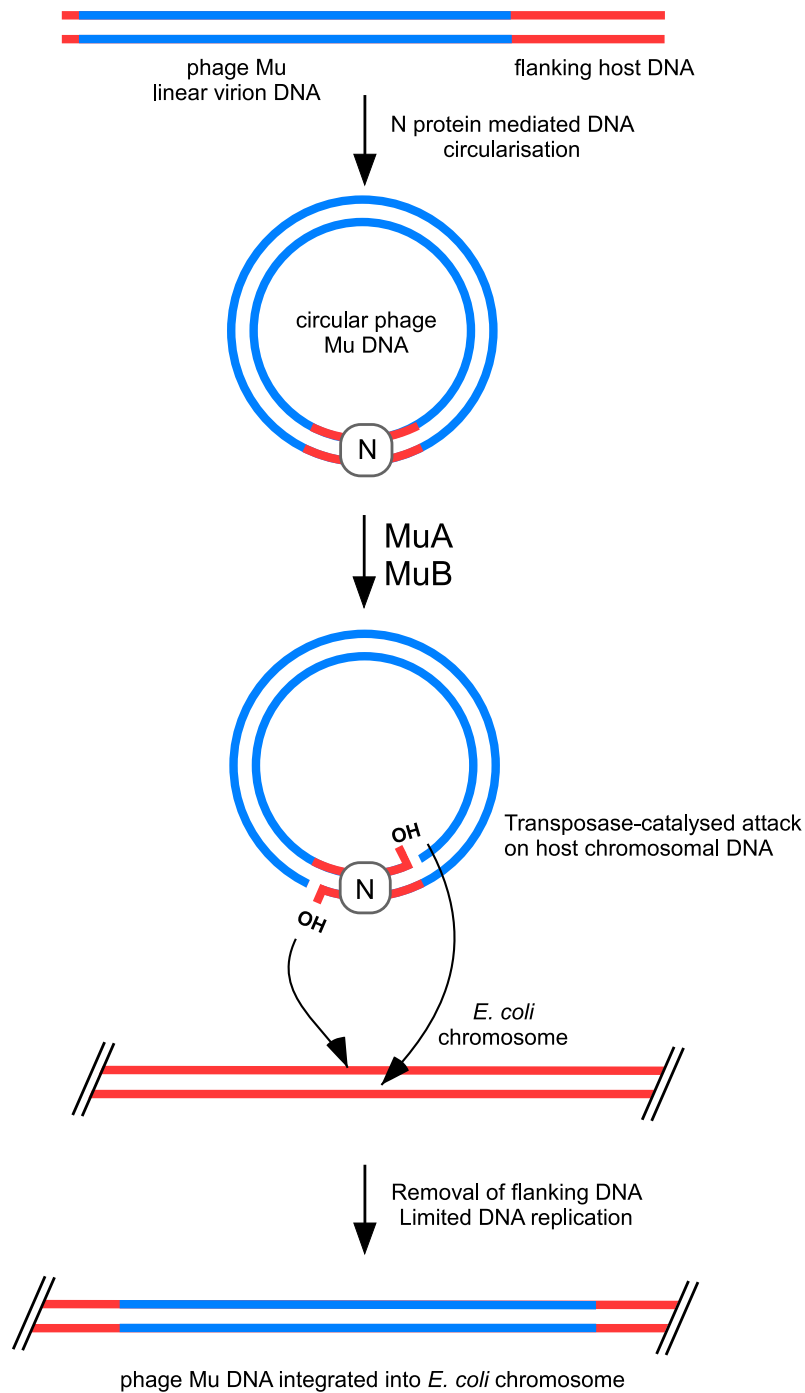


Fig. 3 Bacteriophage Mu integrates into the *E. coli* chromosome via random transposition. Linear Mu DNA is non-covalently circularized via the N protein. The Mu B transposase generates single stranded nicks at the junctions of the Mu (blue)-prior host DNA (red) sequence, revealing two 3'-OH ends. These ends attack a target site on the *E. coli* chromosome to join the 3'-OH of Mu DNA to the 5' phosphate of the host DNA. Removal of the prior host DNA and subsequent repair of the DNA gap at the host-Mu DNA junction via limited replication results in successful integration of Mu into the chromosome.

Maintenance of the Lysogenic State

Temperate phages often contain natural examples of genetic switches which enable the developmental choice between the lytic or lysogenic lifecycles. The biochemical basis of this decision-making process will be covered in detail in a separate chapter of the encyclopedia. After entering lysogeny, true lysogens are able to exit back into the lytic cycle, either spontaneously or in response to environmental cues. One strategy to generate such a bistable decision making circuit is through mutually repressing transcription factors. The most well characterized bacteriophage employing this strategy is bacteriophage λ .

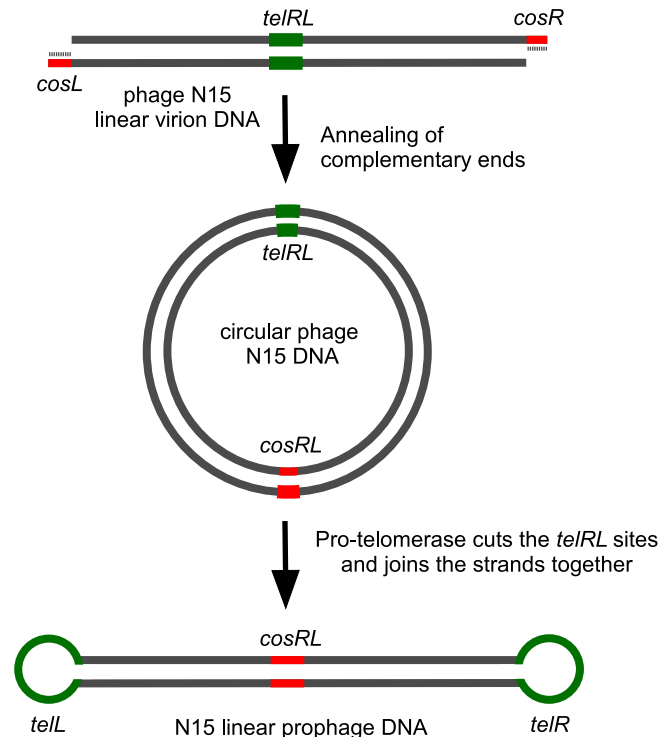


Fig. 4 A prophage of bacteriophage N15 maintains its DNA as a linear plasmid with closed hairpin loops. Upon injecting its DNA into *E. coli*, the phage initially circularizes through the annealing of complementary ends. A phage encoded protelomerase binds and cleaves an inverted repeat sequence (*telRL*) within the genome. The phosphodiester backbone of single strands are then ligated together to form two hairpin loops, creating one covalently closed DNA molecule. Linear plasmid replication requires a phage encoded protein, *repA*.

Bacteriophage λ

The bistable regulatory circuit in λ is generated by the CI and Cro repressors, which block each other's transcription. To maintain lysogeny, the promoters that drive lytic functions must be repressed, whilst maintaining production of the CI repressor. In general, the repressor in temperate phages responsible for turning off lytic function is termed the immunity repressor, as it prevents secondary infections by phage of the same family, by blocking the lytic functions of incoming phage.

λ has a lysogenic promoter, P_{RM} , positioned back to back with the adjacent lytic promoter P_R , with a second lytic promoter, P_L located 2.3 kb away (Fig. 5). The transcriptional cascade in λ is such that repression of the early lytic promoters P_L and P_R will stop all of the lytic genes from being expressed, pushing the phage into lysogeny. This repression is achieved by the λ CI protein, which binds to operators at O_R and O_L operators that overlap the P_R and P_L promoters, respectively. In the lysogenic state, CI simultaneously represses transcription from P_L and P_R , while activating its own expression from P_{RM} .

The lambda CI protein is a two domain protein consisting of an N-terminal DNA binding domain connected by a cleavable linker to an oligomerization domain. The repressor homodimerizes and these dimers can bind to individual operators. However, repressor binding is cooperative such that pairs of dimers bound to adjacent operators (primarily O_{R1} - O_{R2} and O_{L1} - O_{R2}) interact to improve overall binding affinity. In this state, P_R and P_L are repressed, whilst P_{RM} is activated by a contact between the CI dimer bound at O_{R2} and the σ subunit of RNA polymerase. In this manner, λ CI produces positive feedback to increase its own expression. A further level of cooperativity also occurs, when a pair of dimers occupying two operators at O_R interact with a pair of dimers bound at O_L to form a CI octamer, looping out the intervening ~ 2.3 kb of DNA (Fig. 5). Octamer formation in turn positions the O_{L3} and O_{R3} operators to allow λ CI dimers bound at each of these sites to interact and repress P_{RM} at high levels of λ CI. This negative autoregulation is thought to limit the repressor levels to allow for efficient prophage induction back into lytic cycle (Fig. 5(D)).

Cro is the second transcriptional regulator within the lambda genetic switch. The *cro* gene is the first of the early genes expressed from P_R following phage infection, and is required to enforce prophage induction – the decision point when the phage switches from lysogeny towards lysis. Upon activation of the host SOS response, the CI repressor is degraded by a mechanism involving activated RecA improving an intrinsic but normally weak CI self-cleavage activity. Thus the lytic repressors are at least partially de-repressed and Cro is made. Cro binds as a dimer to the same O_L and O_R operators as CI, but with different affinity, binding most strongly to O_{R3} at the O_R operator. This preferential and non-cooperative binding to O_{R3} dampens CI expression from the P_{RM} promoter, while leaving the P_L and P_R promoters active. Thus, Cro repression of P_{RM} is required to prevent re-synthesis of new CI during prophage induction, which would otherwise re-establish the prophage state.

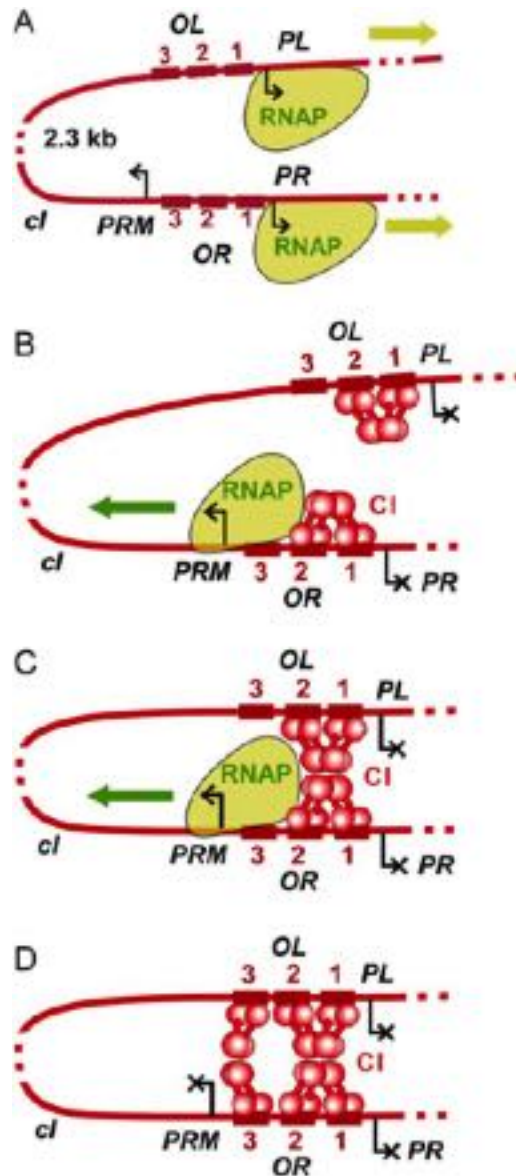
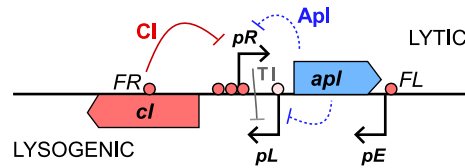


Fig. 5 Regulation of the λ lytic and lysogenic promoters by λ CI involves cooperative interactions at multiple levels. A. The arrangement of early lytic and lysogenic promoters in bacteriophage λ . The regulation of these promoters is crucial in the lytic-lysogeny decision. Promoters are depicted as bent arrows. RNA transcripts from P_L and P_R are shown in yellow. OL and OR operators are depicted as red bars on the DNA. B. λ CI (depicted as red dumbbells) binds to operators at OL and OR. λ CI dimers at adjacent operators can further interact to form tetramers, mediated through the CI C-terminal domain. Binding of λ CI blocks transcription from P_R and P_L , by blocking RNA polymerase (RNAP) binding. Contacts between the N-terminal domain of CI at O_{R2} and the sigma subunit of RNAP activates the P_{RM} promoter to drive expression of λ CI and other lysogenic functions. The lysogenic transcript from P_{RM} is depicted in green. C. The repression of P_R and P_L is further enhanced by interactions between CI tetramers at OR and OL, assisted by DNA looping. D. Formation of the CI octamer allows CI dimers bound at OL3 and OR3 interact cooperatively, to repress P_{RM} and downregulate CI's own expression.

Bacteriophage 186

Bacteriophage 186 is a temperate bacteriophage of the P2 related family, a group of phages evolutionarily distinct from the lambdoid phages. 186 and lambda infect the same host (*E. coli*) and have very similar temperate lifecycles, yet the mechanisms by which these two phages enter and maintain lysogeny are quite different. In 186, early lytic p_R and lysogenic promoters p_L are in a convergent arrangement (Fig. 6), rather than divergent as is the case for the λ P_R and P_{RM} promoters. A consequence of this convergent layout is that RNA polymerases either bound at, or elongating from, the switch promoters can potentially collide with each other. In the absence of the immunity repressor (CI), transcription from the stronger lytic p_R promoter represses transcription from the weaker lysogenic promoter p_L by this mechanism of transcriptional interference (TI), favouring the lytic lifecycle. Conversely, in the presence of CI, p_R is efficiently repressed, relieving transcriptional interference, and providing a form of positive feedback on CI expression.

A. Bacteriophage 186 switch region



B. DNA wrapping and looping by 186 CI

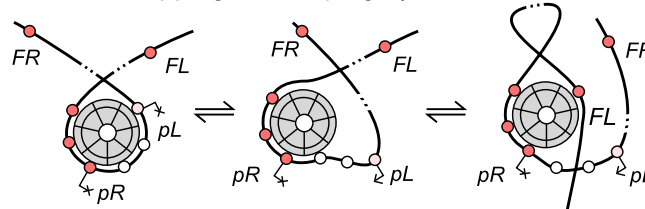


Fig. 6 Relief of transcriptional interference is used to activate transcription of lysogenic promoters in bacteriophage 186. A. The early lytic promoter pR and the lysogenic promoter pL are arranged face-to-face. CI is the 186 immunity repressor. Apl acts as both a repressor of the pR and pL promoters, and also as the excisionase for phage 186. The 186 CI repressor binding sites are drawn as circles – there are three strong sites (red) over the pR promoter, a weak operator located over pL (pink) and two distant sites (FL and FR), located ~300 bp from pR. B. Model of transcriptional activation of 186 pL by 186 CI. CI is shown as a wheel-like structure where DNA wraps around the circumference of the wheel. CI operators are shown as circles. The weak operator at pL is sometimes wrapped to the wheel (left) thus repressing pL, in an equilibrium with a form (middle) where the pL operator is not bound, allowing pL to be active. The flanking sites FL and FR can compete with pL for binding to the CI wheel (right), fine tuning the response of the system to CI levels.

The 186 CI protein is the immunity repressor responsible for maintenance of lysogeny. Similarly to λ CI, it contains an N-terminal DNA binding domain and a C-terminal oligomerization domain and binds to its operators as a dimer. Crystal structures suggest 186 CI forms a wheel consisting of a heptamer of dimers. The repressor's DNA binding domains face outwards from this wheel, allowing DNA to wrap around the protein (**Fig. 6(B)**) to regulate transcription of pR and pL. The CI wheel binds cooperatively to the three strong operators to repress pR and relieve transcriptional interference on pL. Binding of the weaker CI operator at pL to the CI wheel is in an equilibrium between wrapped (repressed) and unwrapped (active) forms. There is further fine tuning of pL promoter activity by the two flanking CI operator sites (FL and FR, located ~300 bp away) which can loop to the wheel and displace pL from the wheel. When the CI operator at pL is displaced by either of the stronger flanking site operators, pL is transcriptionally active, maintaining CI levels. Interestingly, in phage 186, integrase is also expressed on the lysogenic transcript, such that integrase is continually present in a lysogen. When prophage induction is triggered, excision of the phage genome requires only expression of the excisionase (Apl), the first gene of the lytic transcript.

Thus, phage 186 CI repressor maintains an active lysogenic promoter by relieving transcriptional interference on pL, to drive its own expression and to maintain a stable, self-correcting lysogenic level of CI. The lytic pR promoter is tightly repressed in this state, leading to low levels of spontaneous induction.

Integration-Dependent Bacteriophage Immunity

A class of temperate mycobacteriophage have been described which use site-specific integration as the key decision point of the genetic switch. In these phage, the immunity repressor and integrase are expressed from the same promoter P_{rep} . To the right of the repressor gene, under the control of a separate divergent promoter, P_R , is a Cro-like protein. Both the immunity repressor and Cro-like protein are predicted to have a DNA-binding helix-turn-helix (HTH) motif. Crucially, the bacteriophage attachment site *attP* is located within the open reading frame of the immunity repressor (**Fig. 7**).

Unlike bacteriophages λ and 186, the promoter driving the immunity repressor function is not regulated by the immunity repressor. Rather, the key step needed to establish lysogeny is the integration of the phage DNA into the host chromosome. Integration of the phage genome through recombination of *attP* and *attB* sites results in a truncated immunity repressor without a short destabilizing C-terminal tag. This truncated repressor supplies sufficient repression activity to establish lysogeny. The tag, thought to target the viral repressor for degradation by host proteases, reduces its effective activity below the level required for lysogeny.

The determining factor of the phage's developmental fate is the stability of the integrase protein. If conditions favor high integrase activity, such as high multiplicity of infection (MOI) or when host protease levels are low, integration of the phage genome is favored. Integration subsequently establishes and maintains lysogeny via immunity repressor stabilization.

Use of a Host Protein as an Immunity Repressor

In most phages, a phage-encoded immunity repressor binds to promoters that drive lytic functions to repress them. GIL01, a phage that infects *Bacillus thuringiensis*, is an interesting example of a phage that does not have encode an immunity repressor at all. It instead uses the *B. thuringiensis* host LexA repressor as its immunity repressor.

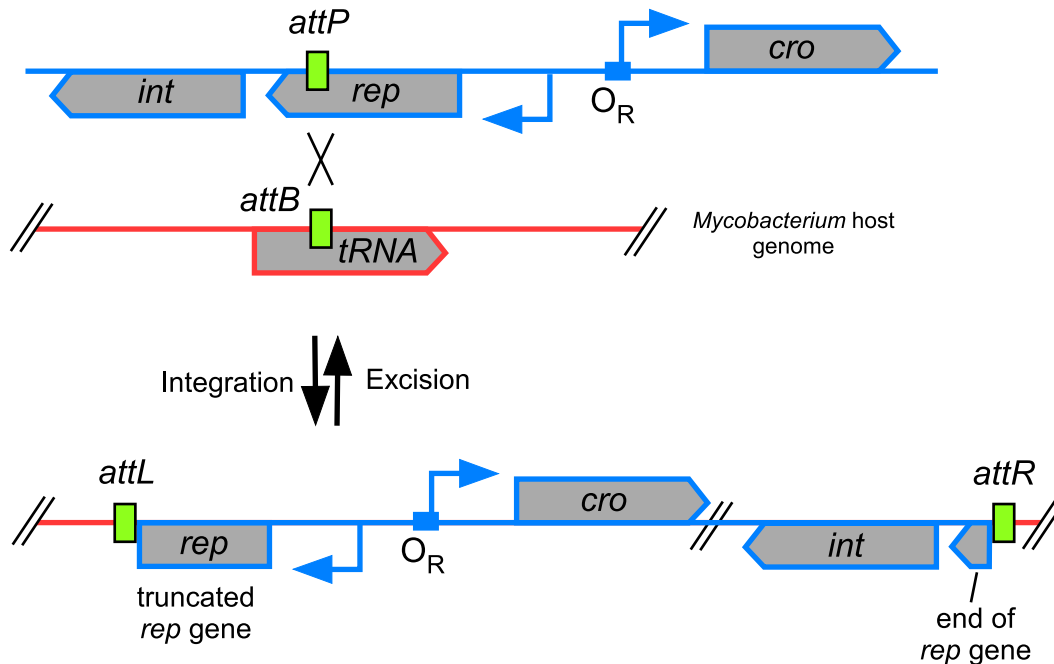


Fig. 7 The organization of integrase (*int*), immunity repressor (*rep*) and *cro*-like protein (*cro*) genes in integration-dependent immunity systems. Promoters are shown as bent arrows. Site-specific attachment sites are shown as green boxes. The *attP* attachment site lies within the *rep* open reading frame. Site specific recombination of *attP* with *attB* results in a truncated repressor gene within the bacterial chromosome. This truncated gene encodes for a more stable immunity repressor, driving the phage towards lysogeny.

It has been shown that the DNA binding activity of LexA in *B. thuringiensis* is regulated by GIL01 phage encoded proteins. When complexed with specific phage proteins, LexA represses lytic promoters by binding to a conserved LexA site within the promoter region. If these additional phage proteins are not expressed, LexA binding alone is too weak to repress the lytic promoter and GIL01 is unable to form stable lysogens.

RNA to Maintain Lysogeny

The mechanisms of maintenance of lysogeny that have been described thus far have been based on proteins, responsible for mediating either repression or integration. However, there are also examples of phage which use RNA-based mechanisms to maintain a stable lysogenic state.

Lactobacillus casei phage A2 has a similar genetic arrangement to that of the *O_R* region of lambda, with back-to-back promoters driving expression of CI and Cro repressors. However, in phage A2, the CI and Cro repressors have very similar affinities for their operators, leading to the expectation that the much stronger lytic pR promoter should dominate, leading to expression of Cro, and favoring the lytic cycle. However, A2 is able to effectively establish and maintain stable lysogens. It has recently been shown that the second gene of the A2 lytic transcript encodes an RNA binding protein (gp25) that is able to specifically bind to a region between the ribosome binding site and *cro* start codon on the lytic transcript. RNA binding by gp25 thus prevents translation of Cro, allowing establishment and maintenance of stable lysogens.

P4 is a so-called satellite phage that relies on an unrelated helper phage, such as P2 or 186, to supply the structural genes needed for its propagation. P4 has the ability to redirect the genetic network of the helper phage in order to redirect the capsid assembly process for its own purposes. The P4 replicon can exist as a plasmid or can be stably integrated into the host chromosome as a lysogen. The integrated lysogenic form is maintained by an RNA-based mechanism. In a P4 lysogen, transcription from a constitutive early promoter generates a ~300 nucleotide transcript that is processed to form the CI RNA. The CI RNA interacts with two specific regions in the untranslated leader sequence of the nascent transcript to mediate transcriptional termination, prevent expression of the P4 replication functions and thus, maintain lysogeny.

Prophage Induction

Established lysogens are generally stable with low probabilities of reverting back to the lytic state. Certain environmental stresses can stimulate the prophage to revert to the lytic state. This process is called prophage induction. The immunity repressor must be degraded or otherwise inactivated during prophage induction in order to establish lytic gene expression.

Many temperate phages are inducible by DNA damage, such as through mitomycin C treatment or UV irradiation. Such DNA damage results in activation of the host SOS response, a global cascade of gene expression centered on DNA repair. As explained previously, the stress response genes are under the control of the host LexA protein. LexA contains a labile linker region that can be cleaved by a C-terminal protease domain within LexA itself. In the SOS response, the activation of co-protease RecA as a result of the accumulation of single stranded DNA, promotes the auto-proteolysis of the host LexA repressor. Phage GIL01, which relies on LexA to directly repress lytic genes, will undergo prophage induction simply due to the removal of LexA. Some phage repressors, such as λ CI repressor, have structural similarity to LexA, and undergo similar RecA-stimulated self-cleavage. Inactivation of CI allows the expression of lytic genes, including the genes for excision and replication.

Other phages, such as bacteriophage 186, express specific anti-repressor proteins, which complex with their cognate repressor to relieve repression of the lytic genes. For example, bacteriophage 186 expresses the anti-repressor Tum, which is in turn under the control of a LexA repressible promoter. Upon activation of the SOS response, RecA mediated proteolysis of LexA results in expression of the Tum anti-repressor, which inactivates 186 CI, to drive prophage induction. The related coliphage P2 does not code for an anti-repressor, and does not induce in response to DNA damage. However, it has a somewhat higher level of spontaneous induction than 186, which must be sufficient for it to sustain a viable number of free phage.

The satellite phage P4, which uses an RNA based transcriptional termination mechanism for maintaining lysogeny, is induced by transcription from an alternative promoter activated by a protein from the helper phage. Translation of nested genes within this longer transcript leads to suppression of the CI RNA mediated termination, and expression of the P4 genes needed for replication.

Evolutionary and Phenotypical Effects of Lysogeny

It is estimated that the number of phage outnumber bacteria in the biosphere by at least an order of magnitude. Thus, it is unsurprising that phages have a large effect on the evolution of their hosts and play a significant role in shaping bacterial communities in almost all ecosystems. Here, we discuss how lysogeny has altered the phenotype and evolution of its hosts.

Lysogenic Conversion

Lysogenic conversion describes the situation where a bacterial host acquires a new trait as a direct result of the expression a gene encoded by a lysogen. Clinically relevant examples include the acquisition of virulence factors by bacterial pathogens. Many of the toxins contributing to virulence in bacteria such as *Staphylococcus aureus*, *Shigella* spp., *Salmonella enterica* and pathogenic *E. coli* have been acquired through lysogenic conversion. *E. coli* O157:H7, a causative agent of food poisoning, harbors Sp5 and Sp15 – two Shiga toxin expressing prophages. *Vibrio cholera*, some strains of which cause Asiatic cholera, acquired the cholera toxin from its CTX ϕ prophage.

Besides toxins, other phage-encoded virulence factors include adhesion factors, superantigens and factors that aid immune system invasion. Various examples of phage-encoded virulence factors are summarized in [Table 1](#).

In many cases, the virulence genes are arranged as a cassette flanked by a σ^{70} promoter on one side and a transcription terminator on the other. Such a configuration reduces interference with the prophage genes adjacent to the cassette. It has been postulated that virulence genes confer some fitness advantage to their host, either offensive (e.g., toxin production, invasins), defensive (e.g., detoxification) or survival (such as improved nutrient uptake).

Very recently, a type of temperate filamentous bacteriophage that infects and integrates into *Pseudomonas aeruginosa* (*Pa*) was found to be associated with chronic human wound infections. In mice, phage-infected *Pseudomonas aeruginosa* led to more severe and longer-lasting wounds compared to wounds colonized by *Pa* alone. The authors showed that uptake of phage-infected *Pseudomonas aeruginosa* by cells of the immune system resulted in phage RNA production, inappropriate antiviral immune responses and suppression of bacterial clearance.

Gene Disruption

Prophage integration into the host chromosome can alter bacterial phenotype by disruption of an open reading frame. The integration of phage may also disrupt sequences with regulatory roles within an intergenic region, altering how downstream genes are regulated. Prominent loci for *attB* phage integration sites are within tRNA open reading frames. The loss of a functional tRNA could potentially lead to a reduction in host fitness, so many phage have evolved such that integration will reconstitute a functional tRNA gene. This holds true for many other *attB* sites within important genes.

There are cases where prophage integration does not reconstitute a functional gene and leads to loss of protein function. For example, the integration of L54a and ϕ 13 phages into the *S. aureus* chromosome leads to a loss of a functional lipase and β -toxin, respectively.

Genomic Rearrangement

Bacteria harboring multiple prophages with similar DNA sequences can give rise to homologous recombination events to cause large genomic rearrangements. The location of rearrangements in the chromosome are often correlated with the loci of prophages. For example, from a sequence alignment, it was observed that the genomic differences between an American and Japanese

Table 1 Non-exhaustive list of phage-encoded virulence factors

<i>Protein</i>	<i>Phage</i>	<i>Bacterial host</i>	<i>Gene</i>
Extracellular toxins			
Diphtheria toxin	β -Phage	<i>C. diphtheriae</i>	<i>tox</i>
Neurotoxin	Phage C1	<i>C. botulinum</i>	<i>C1</i>
Shiga toxins	H-19B	<i>E. coli</i>	<i>stx1, stx2</i>
Enterohaemolysin	ϕ FC3208	<i>E. coli</i>	<i>hly2</i>
Cytotoxin	ϕ CTX	<i>P. aeruginosa</i>	<i>ctx</i>
Enterotoxin	NA	<i>S. aureus</i>	<i>see, sel</i>
Enterotoxin P	ϕ N315	<i>S. aureus</i>	<i>sep</i>
Enterotoxin A	ϕ 13	<i>S. aureus</i>	<i>entA</i>
Enterotoxin A	ϕ Mu50A	<i>S. aureus</i>	<i>sea</i>
Exfoliative toxin A	ϕ ETA	<i>S. aureus</i>	<i>eta</i>
Toxin type A	T12	<i>S. pyogenes</i>	<i>speA</i>
Toxin type C	CS112	<i>S. pyogenes</i>	<i>speC</i>
Cholera toxin	CTX ϕ	<i>V. cholerae</i>	<i>ctxAB</i>
Enterotoxin	CTX ϕ	<i>V. cholerae</i>	<i>ace, zot</i>
Leukocidin	fPVL	<i>S. aureus</i>	<i>pvl</i>
Superantigens	8232.1	<i>S. pyogenes</i>	<i>speA1, speA3, speC, spel, speH, speM, speL, speK, ssa</i>
Cytolethal distending toxin	Unnamed	<i>E. coli</i>	<i>cdt</i>
Proteins altering antigenicity			
Membrane proteins	Pnm1	<i>N. meningitidis</i>	<i>Mu-like</i>
Glucosylation	ϵ^{34}	<i>S. enterica</i>	<i>rfb</i>
Glucosylation	P22	<i>S. enterica</i>	<i>gtr</i>
O-antigen acetylase	Sf6	<i>S. flexneri</i>	<i>oac</i>
Glucosyl transferase	Sfil, Sfv, Sfx	<i>S. flexneri</i>	<i>gtrII</i>
Effector proteins involved in invasion			
Type III effector	SopE ϕ	<i>S. enterica</i>	<i>sopE</i>
Type III effector	GIFSY-2	<i>S. enterica</i>	<i>ssel (gtgB)</i>
Type III effector	GIFSY-3	<i>S. enterica</i>	<i>sspH1</i>
Enzymes			
Superoxide dismutase	Sp4, 10	<i>E. coli</i> O157	<i>sodC</i>
Superoxide dismutase	GIFSY-2	<i>S. enterica</i>	<i>sodC-I</i>
Superoxide dismutase	Fels-1	<i>S. enterica</i>	<i>sodC-III</i>
Neuraminidase	Fels-1	<i>S. enterica</i>	<i>nanH</i>
Hyaluronidase	H4489A	<i>S. pyogenes</i>	<i>hylP</i>
Leukocidin	ϕ PVL	<i>S. aureus</i>	<i>pvl</i>
Staphylokinase	ϕ 13	<i>S. aureus</i>	<i>sak</i>
Phospholipase	315.4	<i>S. pyogenes</i>	<i>sla</i>
DNase/streptodornase	315.6, 8232.5	<i>S. pyogenes</i>	<i>sdn, sda</i>
Serum resistance			
OMP	λ	<i>E. coli</i>	<i>bor</i>
OMP	λ -like	<i>E. coli</i>	<i>eib</i>
Adhesions for bacterial host attachment			
Vir	MAV1	<i>M. arthritis</i>	<i>vir</i>
Phage coat proteins	SM1	<i>S. mitis</i>	<i>pblA, pblB</i>
Others			
Mitogenic factors	370.1, 370.3, 315.3	<i>S. pyogenes</i>	<i>mf2, mf3, mf4</i>
Mitogenic factor	Unnamed	<i>P. multocida</i>	<i>toxA</i>
Mitogenic factor	phisc 1	<i>S. canis</i>	<i>Unnamed</i>
Virulence	GIFSY-2	<i>S. enterica</i>	<i>gtgE</i>
Antivirulence	GIFSY-2, Fels-1	<i>S. enterica</i>	<i>grvA</i>

Note: Table adapted and extended from Brüssow, H., Canchaya, C., Hardt, W., 2004. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews* 68 (3), 560–602.

S. pyrogenes M3 isolate could be largely explained by two sequential inversions. One of these inversions occurred between the lysis modules of two prophages.

Genomic rearrangements may also aid phage evolution. In the example of *S. pyrogenes*, the lysogenic conversion genes were outside of the inversion region. The inversion event shuffled virulence genes between the prophages. This could lead to new prophages with altered host specificity and behaviors.

Conclusion

Lysogeny can change the phenotype, fitness and evolution of the bacterial host cell. In the case of bacterial pathogens, prophages allow the acquisition of new traits such as virulence factors which in turn may improve the host's fitness. This is emphasized by the observation that prophages have played a vital role in the emergence of several clinically relevant human pathogens, such as the food-borne pathogen *E. coli* O157:H7 and cholera causing *V. cholerae*. Prophages also drive bacterial evolution through the introduction of new genes, the disruption of existing genes and mediating large genomic rearrangements within the host.

In the laboratory, temperate bacteriophages have provided numerous tools and reagents for molecular biology, and have been used as models of regulatory networks and biological switches. The study of bacteriophages λ , P2 and its relatives, Mu, P1, N15 amongst others have provided many seminal contributions to these fields. This article has touched on several of these mechanisms and simultaneously highlighted the diversity in means for achieving the persistence of DNA, maintaining the prophage in a quiescent state and for exiting lysogeny. As the number of bacterial genomes sequenced continues to grow, new prophages using novel mechanisms to achieve these lysogenic functions will no doubt be discovered.

Acknowledgments

The Shearwin lab is supported by grants from the Australian Research Council (DP160101450) and the National Health and Medical Research Council (APP1100653).

Further Reading

- Brüssow, H., Canchaya, C., Hardt, W., 2004. Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews* 68 (3), 560–602.
- Carrasco, B., *et al.*, 2016. Modulation of *Lactobacillus casei* bacteriophage A2 lytic/lysogenic cycles by binding of Gp25 to the early lytic mRNA. *Molecular Microbiology* 99 (2), 328–337.
- Chandra, B., Ramisetty, M., Sudhakari, P.A., 2019. Bacterial 'grounded' prophages: Hotspots for genetic renovation and innovation. *Frontiers in Genetics* 10 (February), 1–17.
- Christie, G.E., Calendar, R., 2016. Bacteriophage P2. *Bacteriophage* 7081 (November), e1145782.
- Christie, G.E., Dokland, T., 2012. Pirates of the Caudovirales. *Virology* 434 (2), 210–221.
- Dodd, I.B., Shearwin, K.E., Egan, J.B., 2005. Revisited gene regulation in bacteriophage lambda. *Current Opinion in Genetics and Development* 15 (2), 145–152.
- Golding, I., 2016. Single-cell studies of phage λ : Hidden treasures under occam's rug. *Annual Review of Virology* 3, 7.1–7.20.
- Harshey, R.M., 2014. Transposable phage Mu. *Microbiology Spectrum* 2 (5), 1–22.
- Łobocka, M.B., Rose, D.J., Plunkett, G., *et al.*, 2004. Genome of bacteriophage P1. *Journal of Bacteriology* 186 (21), 7032–7068.
- Mai-Prochnow, A., Gee, J., Hui, K., *et al.*, 2015. Big things in small packages: The genetics of filamentous phage and effects on fitness of their host. *FEMS Microbiology Reviews* 39 (February), 465–487.
- Merrick, C.A., Zhao, J., Rosser, S.J., 2018. Serine integrases: Advancing synthetic biology. *ACS Synthetic Biology* 7, 299–310.
- Ptashne, M., 2004. *A Genetic Switch – Phage Lambda Revisited*. New York: Cold Spring Harbor Laboratory Press.
- Ravin, N.V., 2015. Replication and maintenance of linear phage-plasmid N15. *Microbiology Spectrum* 3 (1), 1–12.
- Schubert, R.A., Dodd, I.B., Egan, J.B., Shearwin, K.E., 2007. Cro's role in the CI-Cro bistable switch is critical for lambda's transition from lysogeny to lytic development. *Genes and Development* 21 (19), 2461–2472.
- Shao, Q., Trinh, J.T., Zeng, L., 2019. High-resolution studies of lysis – Lysogeny decision-making in bacteriophage lambda. *Journal of Biological Chemistry* 294 (10), 3343–3349.

Relevant Websites

<https://ecocyc.org/>
EcoCyc.
<https://viralzone.expasy.org/>
ViralZone root – ExPASy.

Decision Making by Temperate Phages

Ido Golding, University of Illinois at Urbana-Champaign, Urbana, IL, United States

Seth Coleman, Rice University, Houston, TX, United States

Thu VP Nguyen and Tianyou Yao, Baylor College of Medicine, Houston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Induction A process by which a temperate phage switches from the lysogenic state to the lytic pathway, typically in response to a perturbation such as damage to the host cell's DNA.

Lysogenic pathway One of the two possible pathways following temperate phage infection, characterized by suppression of viral replication and lytic functions. The bacterial cell carrying the dormant phage (e.g., in the form of a prophage) is termed a lysogen.

Lytic pathway The other possible pathway following temperate phage infection, in which phages replicate rapidly

and eventually kill (lyse) the host cell, releasing progeny viruses.

Multiplicity of infection (MOI) The number of viruses co-infecting the same cell.

Prophage A phage in the lysogenic state, which has integrated into the host genome. Replication of the viral DNA occurs passively during replication of the bacterial chromosome.

Temperate phage A bacteriophage that is capable of the dormant (lysogenic) pathway, in addition to the active (lytic) one.

Introduction

The defining lifestyle of most bacteriophages, analogous to the behavior of higher viruses, is one in which infection is followed by rampant viral replication and the release of numerous mature progeny, often accompanied by death of the host cell (lysis). However, a subset of phages, denoted as *temperate*, are capable of an alternative lifestyle called lysogeny, where the violent cataclysm is replaced by viral dormancy: Following infection, the phage genome is maintained inside the bacterial host – either integrated (as a *prophage*) into the bacterial chromosome, or replicating extra-chromosomally – with all virulent functions shut off. The dormant phage, now an integral part of the bacterial cell, is inherited from generation to generation. However, while dormant, the phage typically maintains the potential for lytic induction: a switch back to the virulent pathway in response to specific signals indicating stress to the host cells.

Multiple aspects of the lysogenic lifestyle are the subject of current studies. These include the physiological costs and benefits, to both phage and host, of viral dormancy, as well as the ecological, medical, and evolutionary consequences of lysogeny. Here we focus on a single element: The decision between lysis and lysogeny, made by temperate phages upon infection of the host. In particular, we will devote most of our attention to the decision by bacteriophage lambda, which infects *Escherichia coli* (Fig. 1). Through more than half a century of genetic, biochemical, and biophysical studies, lambda has become arguably the best-characterized biological system, albeit one that still presents many open questions. As we review the lambda lysis/lysogeny decision, we will also provide examples for some of the ways that other temperate phages make this choice. In light of the overwhelming diversity of phage lifestyles, it is certain that many more ways in which phages pursue their lysis/lysogeny decisions remain to be discovered.

Owing to the relative compactness and tractability of bacteriophages, many of their functions have been studied not only for their intrinsic value, but also as simplified models for the behavior of higher biological systems. This is also the case for the lysis/lysogeny decision, which serves as a paradigm for the way genetic circuits, receiving external inputs, make binary cell-fate choices. In that role, for example, the decision by phage lambda has provided insights about the transition in and out of dormancy by HIV, and, beyond viruses, regarding the process of cellular differentiation during metazoan development. At the heart of lambda's ability to inform us on higher organisms are common features of binary decision circuits across biological systems, notably, the utilization of auto-regulating, fate-determining genes to provide high cell-state stability while simultaneously allowing efficient fate switching in response to external signals.

As part of its paradigmatic role, lambda has also served as a testbed for the idea of creating a quantitative description, formulated in mathematical terms, for the function of the genetic circuit, with the goal of predicting the decision outcome for given initial conditions. While efforts in this direction have yielded significant progress, they have also been limited by the phenomenon of cellular individuality, whereby genetically identical cells, in a uniform environment, nevertheless end up pursuing different paths. To explain the indeterminacy of single-cell choice, researchers invoked the effect of stochastic biochemical fluctuations ("noise") on the decision circuit. With this concept, too, lambda has paved the way for the elucidation of cellular stochasticity and its consequences across the fields of microbiology, development, ecology and medicine.

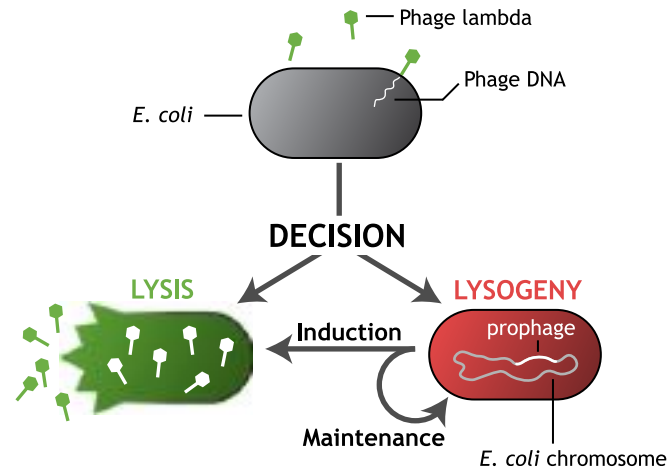


Fig. 1 The post-infection decision of bacteriophage lambda. Following infection of an *E. coli* cell, a binary choice is made between lysis, defined by rampant viral replication and cell death, and lysogeny, in which the viral DNA is integrated into the host's genome to become a dormant prophage. The lysogenic state is stably maintained, but a switch back to lysis can be induced by cellular damage. Adapted with permission from Golding, I., 2016. Single-cell studies of phage λ : Hidden treasures under Occam's rug. *Annual Review of Virology* 3 (1), 453–472, permission conveyed through Copyright Clearance Center, Inc.

The Post-Infection Decision of Phage Lambda

A decision between lysis and lysogeny first takes place following lambda infection, upon entry of the viral genome into the *E. coli* cell (Fig. 1). A genetic circuit encoded by the phage, integrating inputs from the host (and, indirectly, from the extracellular environment), converges on one of the two developmental pathways. If lysis is chosen, the decision is irreversible and, hence, the role of the decision circuit is complete. If, however, the choice is lysogeny, that fate needs to be actively maintained over the long term, using a small subset of the decision gene network. In addition to repressing all virulent functions, this maintenance circuit (known as a genetic “switch”) perpetuates a continuous version of the lysis/lysogeny decision: at any given time, whether to maintain stable dormancy or, if cellular conditions change, undergo lytic induction, switching back to the lytic pathway by relieving the repression of virulent functions.

The choice between lysis and lysogeny can be seen as one of timescale, namely, whether to reproduce (and kill the host) immediately, or at a later, yet to be determined, time. An optimal choice, one which maximizes – in the long term – the expected number of progeny, requires knowledge of the state of the infected cell, e.g., how likely it is to support successful viral reproduction, as well as its environment, e.g., what the chances are that newly-released phages will find additional targets to infect. Below we discuss how lambda and a few other phages measure these critical parameters. From a theoretical point of view, the optimization problem that underlies decision making by temperate phages continues to be an area of considerable interest.

The genetic circuit governing the lambda decision is depicted in Fig. 2. The decision involves a dense network of regulatory interactions between phage genes, as well as multiple inputs from various host functions. The regulatory interactions involve diverse molecular mechanisms, modulating all stages of gene expression: transcription initiation (through transcription factors, alternative promoters, and interference between neighboring promoters) and elongation (anti-terminators), post-transcription (antisense RNA), translation, and post-translation (protein lifetime). The role of some of these interactions in the decision process has been elucidated through genetics, biochemistry, and mathematical modeling of the circuit. The utility of other interactions, however, is less clear. It is possible that those come into play under infection conditions that are not emulated by standard laboratory conditions, serving to increase the robustness of the decision to environmental and biochemical fluctuations.

To gain insight into the decision process, it is helpful to follow the cascade of transcriptional events taking place on the lambda genome following infection, and how these events diverge en route to each of the two possible outcomes (Fig. 3). Upon entry of the viral genome into the *E. coli* cell, transcription begins from the left (P_L) and right (P_R) early promoters. Transcription is initially attenuated at the t_{L1} and t_{R1} terminators, such that only a single protein is expressed from each promoter: N from P_L and Cro from P_R . One of these proteins, N, is an anti-terminator that allows readthrough at t_{L1} and t_{R1} , as well as at another terminator, t_{R2} , further downstream of P_R . This leads to the expression of additional lambda genes. These “delayed early” genes include those that allow progression of the lytic pathway: O and P, required for phage genome replication, and Q, which controls the expression of multiple lytic genes. However, CII and CIII, whose products drive the establishment of lysogeny, are also produced from those same left and right transcripts.

If the Q protein accumulates to a sufficient level, it abrogates termination at the t_R site, to enable expression from the late lytic promoter P_R of genes responsible for the production of new viral particles and lysis of the host cell, and thus completion of the lytic cycle. For the alternative route of lysogeny to be taken, Q production needs to be reduced. This is achieved by CII, which activates the P_{aQ} promoter to produce an antisense transcript to Q, resulting in inhibition of Q expression and thus preventing the expression of the late lytic genes. In addition, CII promotes the lysogenic pathway by activating the P_1 promoter, whose product Int

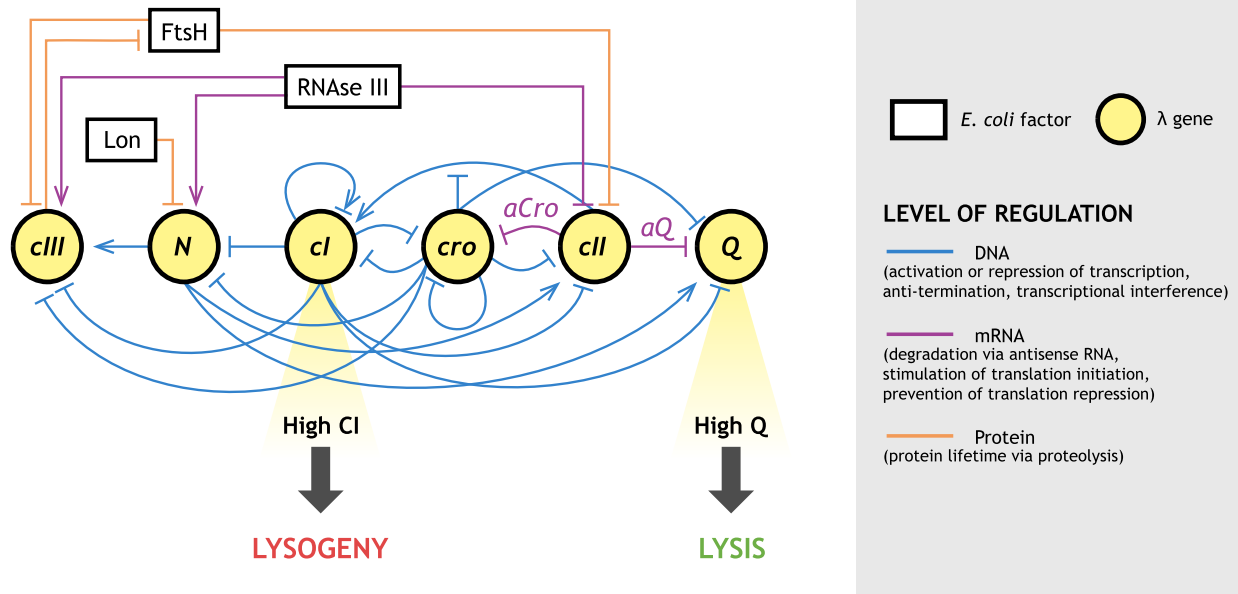


Fig. 2 Key lambda genes and host factors involved in the lysis/lysogeny decision. Regulatory interactions between the various nodes rely on diverse molecular mechanisms, at the level of phage DNA, mRNA and proteins.

catalyzes integration of the phage genome into the bacterial chromosome. Finally, CII activates the repression establishment promoter P_{RE} to produce CI. CI shuts off transcription from P_L and P_R and regulates its own transcription, now from the repression maintenance promoter, P_{RM} . In the lysogenic state, P_{RM} expression of CI is the only transcriptional activity from the decision network of the dormant prophage.

What makes the post-infection decision elusive, even after decades of meticulous interrogation, is that, qualitatively, the expression patterns of early genes are similar in lytic and lysogenic cells, and therefore appear unpredictable of the eventual outcome. Consider *cII*, the key driver of lysogenic choice. Its expression from P_R takes place in a limited time window of ~ 15 min, regardless of the eventual choice. Whether this transient expression will result in lysogeny depends on the exact timing and on the maximal CII level obtained. CII kinetics are regulated at multiple levels – at the promoter (P_R) and terminator (t_{R1}), as well as the mRNA and protein lifetimes. This multi-layer regulation provides the means by which the lysis/lysogeny decision is modulated by host and phage parameters, as discussed later.

Whereas the decision by lambda is by far the one best characterized, progress has also been made towards understanding the decision circuits of other phages. In multiple cases, the decision network is quite reminiscent of lambda's, with the *cl-cII-cIII* module conserved, and an orchestrated gene expression cascade taking place. Even in phages that are otherwise very different from lambda, a dual repressor motif similar to the *cl/cro* pair (discussed in more detail below) is often present, and an auto-regulatory viral “repressor”, which both establishes and maintains the lysogenic state, appears to be near universal. Among temperate phages whose decision process has been studied in some detail, a few offer slight variations on lambda, for example, P22, a “lambdoid” phage (i.e., one with a similar genome architecture to lambda), where the Cro-analog (*ant*) directly inactivates the CI-analog (*c2*) by binding to it and preventing it from effectively binding to DNA. Others exhibit dramatically different behavior. P4, for example, is a “satellite phage”, a genetic element that requires other viruses for its own propagation. Rather than having a global repressor, which establishes and maintains lysogeny by regulation of transcription initiation, the outcome and maintenance of the decision in P4 are driven by control of transcription termination, mediated by RNA-RNA interactions. The default pathway in P4 appears to be lysogeny, with the alternative state, in the absence of a co-infection by a helper phage, being that of a multi-copy plasmid. In that state, infection by a helper phage induces the transition into lysis.

Counting by Infecting Phages

The perceived role of the decision circuit, as stated above, is to choose between lysis and lysogeny based on the conditions of infection. A key question is, thus, how these conditions are sensed by the infecting phage and processed by the decision circuit to yield an optimal outcome. This question is far from settled. To begin with, which aspects of the infection event are pertinent to the outcome? The parameter best characterized in terms of its effect on the lambda lysis/lysogeny choice is the multiplicity of infection (MOI), namely, the number of phages co-infecting the cell. It was found long ago that, the higher the MOI, the higher the probability of lysogeny. Fig. 4 depicts the results of an experiment measuring the relation between the two observables. A known number of bacteria is mixed with varying concentrations of phages, and, once infection is allowed to proceed, the number of

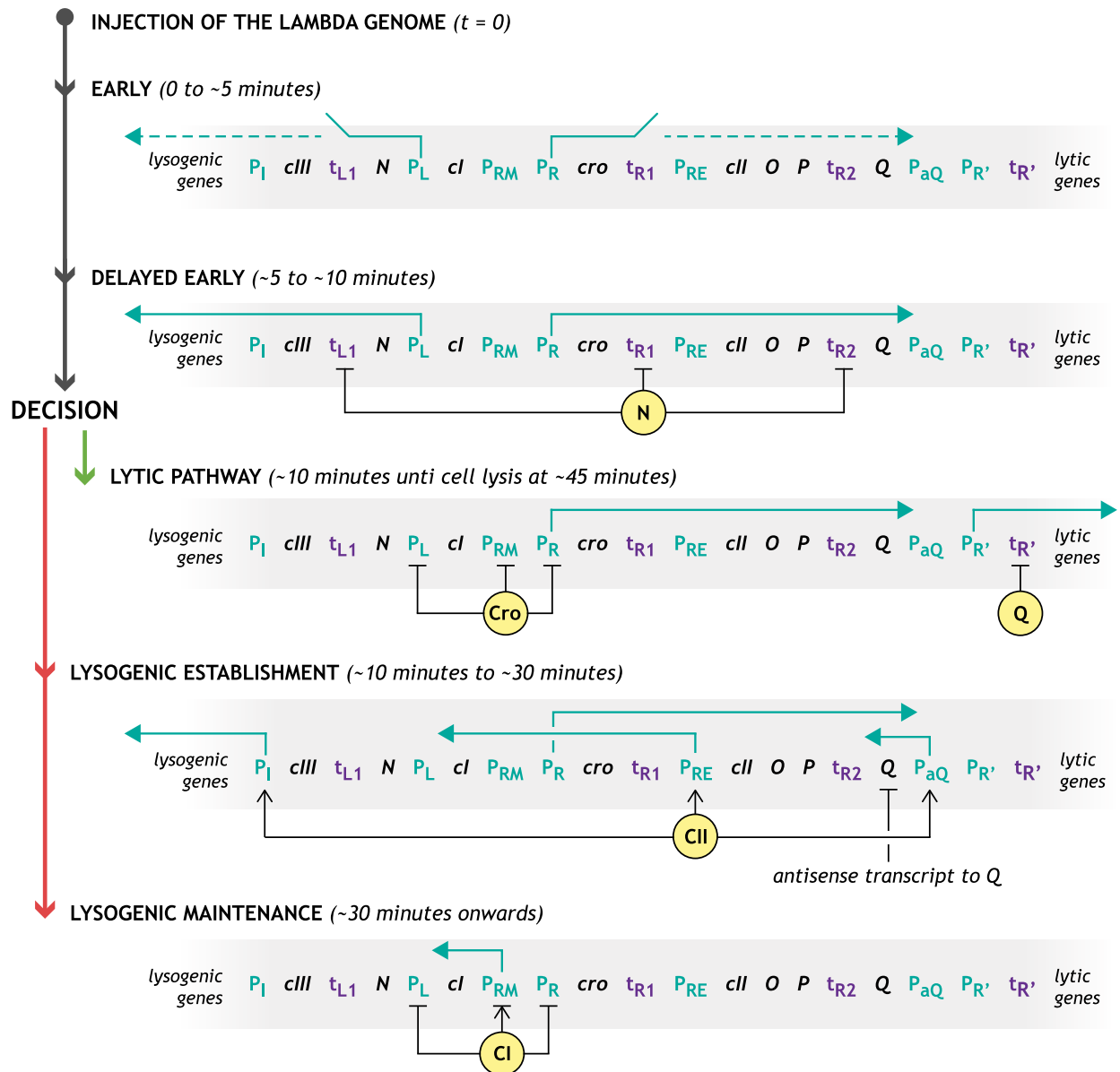


Fig. 3 The cascade of transcriptional events during the lambda lysis/lysogeny decision. Regulatory elements and their interactions are depicted on the relevant region of the lambda genome. The expression pattern of early genes is qualitatively similar irrespective of the eventual fate. After a decision has been reached, different sets of genes are expressed to execute the lysis/lysogeny choice.

resulting lysogens is measured using selection for an antibiotic marker that was engineered into the viral genome. To interpret the experimental results, the measured values are compared to a simple mathematical model, in which random phage-bacteria encounters (following mass-action probability and Poisson statistics) result in lysogeny if the number of phages co-infecting a single cell reaches some number m^* . As seen in Fig. 4, the measured data is consistent with a value of $m^* = 2$, i.e., a scenario where infection by a single phage leads to lysis, whereas simultaneous infection by two or more phages results in lysogeny.

The notion that infecting phages are able to count their numbers in the cell, and then decide on the mode of action based on that number, is intriguing both mechanistically – how do viruses count? – and in terms of its utility – why do they do so? In terms of mechanism, it is commonly held that counting is mediated through the level of CII reached during infection. The MOI affects both CII production (through increased gene dosage) and degradation (through production of CIII, which protects CII from degradation by FtsH, see below). As described earlier, high CII level triggers lysogeny by driving the expression of the integrase and CI, and inhibiting late-lytic gene expression from $P_{R'}$. However, an alternative hypothesis, motivated by recent single-cell experiments and mathematical modeling, posits that phage counting simply reflects the number of cI gene copies, whereas CII levels respond only weakly to MOI, possibly due to the auto-repression of P_R by Cro. If enough CI accumulates, it will then activate its own transcription from P_{RM} and shut down expression of the lytic genes. Elucidating phage counting is complicated

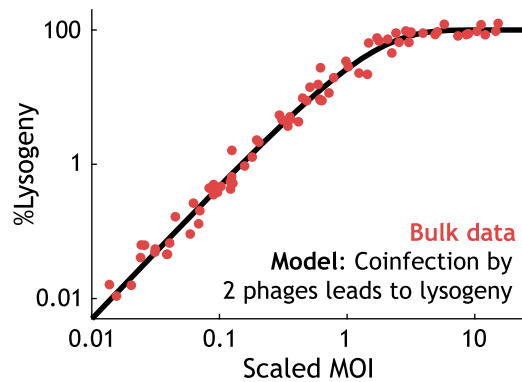


Fig. 4 The dependence of lysogenization on the multiplicity of infection: bulk data. A known number of *E. coli* bacteria is infected with varying concentrations of lambda phage, and the number of resulting lysogens is measured using selection for an antibiotic marker that was engineered into the viral genome. The experimental trend is reproduced by a simple mathematical model, where infection by a single phage leads to lysis, whereas simultaneous infection by two or more phages results in lysogeny.

considerably by the fact that the lambda genome begins replicating shortly after infection, i.e., during, not after, the lysis/lysogeny decision. That early replication plays a role in the decision is evidenced by the fact that replication-deficient lambda mutants require a higher number of initial viruses to achieve lysogeny, compared to wild type phages.

As for the biological utility of choosing lysogeny at higher MOI, the common interpretation is that the number of co-infecting phages is used by lambda as a proxy for the ratio of phage-to-bacteria abundance in the surrounding environment. That ratio, in turn, serves to assess the chances of successful infection by the next generation of phages, should the lytic pathway be chosen. Specifically, high MOI indicates that phages outnumber bacteria and that, therefore, releasing more phages into the environment is futile, since they are unlikely to find new bacterial targets. Hence, high MOI advocates lysogeny. Conversely, low MOI indicates the availability in the environment of yet-uninfected bacteria, thus promoting the release of more viral particles through lysis.

Consistent with the idea that lambda measures the multiplicity of infection in order to assess the abundance of uninfected bacteria in the environment, recent studies found additional ways by which phages can infer this kind of information. During initial rounds of infection of *Bacillus subtilis*, phage phi3T expresses the genes *aimR* and *aimP*. AimR binds to the phage DNA and activates the transcription of *aimX*, which blocks the lysogenic pathway in a mechanism not yet elucidated. The lytic pathway is thus favored in the initial infections. Meanwhile, AimP, a short peptide, is secreted into the extracellular medium, where it accumulates and is taken up by uninfected cells. When some of these cells are later infected, the intracellular AimP, now at high concentration, binds to and inactivates AimR, resulting in repression of *aimX* expression. Consequently, lysogeny is favored in later rounds of infection. Using this intercellular communication system (termed “arbitrium”), infecting phi3T phages are able to record past infections of other cells and tune their lysis/lysogeny decision based on this knowledge.

The ability to hijack the host’s quorum sensing system is also found in VP882, a temperate phage that infects *Vibrio cholerae* and other *Vibrio* species. The outcome of infection by VP882 depends on a repressor (Gp59) that inhibits the expression of the lytic regulator (Gp62), thus promoting the lysogenic pathway. During infection of *Vibrio cholerae*, the phage also expresses VqmA_{phage}, a homolog of the host’s endogenous quorum sensing receptor, VqmA. When bound by *Vibrio*’s autoinducer, DPO, VqmA_{phage} promotes the expression of an antirepressor (Qtip), which sequesters the repressor Gp59, allowing genes involved in the lytic pathways to be expressed. Consequently, when the bacterial density is high, the increased DPO level in the medium promotes the lytic pathway of infecting VP882. Note that, in contrast to phi3T above, which relies on a phage-specific secreted molecule, here, VP882 assesses the bacterial density via an autoinducer encoded by the host.

The cases described above, in lambda and other phages, as well as the ecological argument mentioned, all support the idea that increased phage-to-bacteria ratio promotes lysogeny. However, whether this rule applies universally is still unclear. The propensity for lysogenization by bacteriophage P1, for example, is reported to be insensitive to the number of viruses infecting the host cell. In infections by phage Mu, the probability of lysogeny appears to decrease with the multiplicity of infection, although this trend may reflect the virus’ toxicity to the host at high MOI.

There is also an ongoing debate whether, outside the artificial lab environment, bacterial density is correlated – positively or negatively – with the occurrence of lysogeny, and how to interpret the observed trends. Ecological studies of lytic infections support a “kill the winner” model, in which viral infection increases host diversity by preventing overabundance. However, the dynamics of temperate viruses are much harder to interpret, and the relationship between host density and outcome frequencies is unclear. Previous studies of prophage induction found that lysogeny is more prevalent at low host density, consistent with the picture above of increased lysogenization at high MOI. A more recent work, using metagenomics analysis, reported an inverse trend, but the interpretation of these newer findings is a subject of some controversy.

Whereas the multiplicity of infection is the best-characterized driver of the lambda lysis/lysogeny decision, it is definitely not the only one. Recall that multiple host factors interact with the decision circuit (Fig. 2 above). One such factor is FtsH, a membrane-bound ATP-dependent protease. During lambda infection, FtsH degrades CII and thus impacts the choice between lysis and lysogeny. Part of FtsH’s influence comes about through the MOI, specifically, the dosage-dependent production of CIII, which

is believed to protect CII by serving itself as a target for FtsH. However, beyond the response to MOI, the level and activity of FtsH are regulated by the physiological state of the cell, thus providing a means for the condition of the host cell to inform the lysis/lysogeny decision. For example, the increase in lysogenization at low temperature can be attributed to a decrease in FtsH levels, as well as an increase in the thermodynamic stability of CII, to which FtsH is highly sensitive. Temperature also impacts the decision circuit through its effect on another bacterial protease, Lon, which is expressed in a temperature-dependent manner and targets the N anti-terminator. Two other cellular sensors, cyclic adenosine monophosphate (cAMP) and guanosine tetraphosphate (ppGpp), have also been reported to affect the lysis/lysogeny decision, possibly by inhibiting FtsH. Cellular state also influences the lambda decision through RNase III, whose levels are modulated by the *E. coli* growth rate. RNase III promotes degradation of *cII* transcripts, stimulates ribosome binding and translation initiation of *cIII*, and blocks auto-repression of *N* translation during early infection, thus affecting multiple nodes of the decision network. The ways in which the decision circuit assesses the state of the cell via these and other physiological sensors remain a promising direction for future interrogation.

The View From the Single Cell

Most of what we know about lambda's post-infection decision comes from studies that used traditional genetic and biochemical assays, performed in bulk cultures, and thus involving the averaging of all measured observables over millions of cells. But these individual cells may, in fact, exhibit very different phenotypes. Over the last decade, traditional bulk assays have begun to be supplemented by microscopy-based studies, in which the infection process is followed in real time, at the level of individual cells and phages. Fig. 5(A) shows an example of such an experiment. Here, the lambda capsid was labeled using multiple copies of a fluorescent protein, such that each phage particle appears under the microscope as a diffraction-limited spot. The infected cells were simultaneously imaged using phase contrast microscopy. Time-lapse images show the infection and its outcome for two individual cells. The first cell, infected by a single phage, proceeds to produce more viral proteins (green) and lyse within two hours. The second cell, co-infected by three phages, survives to grow and divide. That cell has chosen the lysogenic pathway, as indicated by the production of a red fluorescent protein, here expressed from the lysogeny establishment promoter, P_{RE} .

Following many infection events in this manner allows one to determine how the decision outcome – lysis or lysogeny – depends on the infection parameters, such as the MOI. Fig. 5(B) shows that the fraction of cells choosing lysogeny increases with MOI, a trend consistent with the observations in bulk. However, in contrast to our original interpretation of the bulk data (Fig. 4 above), the single-cell data suggests that the MOI dependence is probabilistic rather than deterministic: At MOI of 2, for example, an infected cell has about a 50% chance of going either lytic or lysogenic. In other words, when we observe a cell co-infected by two lambda phages, we have no way of telling what route will be chosen!

We are thus confronted with the indeterminacy of single-cell behavior: genetically identical cells, all subject to the same environment, exhibiting different phenotypes from each other. This phenomenon is observed throughout biology, and its origins are a subject of intensive interrogation. According to the prevailing picture, cellular individuality reflects the inherent randomness of biochemical reactions in the cell. In this view, the unavoidable fluctuations in molecular copy number and in the timing of events render the lambda decision “noisy” and unpredictable, rather than precise and deterministic. The plausibility of this argument was first demonstrated using a computational simulation of the lambda decision circuit, showing that fluctuations in biochemical reactions can result in diverging cell fates among infected cells. The concept of noise-driven decisions then evolved and was used to explain cell-fate indeterminacy in higher systems, including the transition in and out of HIV latency, as well as the differentiation and reprogramming of metazoan cells.

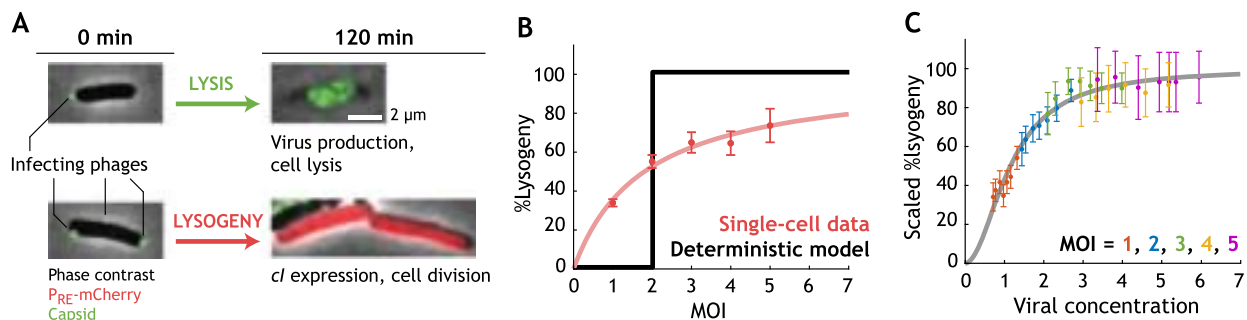


Fig. 5 The lysis/lysogeny decision at the single-cell level. (A) Images from a live-cell movie following the fate of two *E. coli* cells, infected by fluorescently-labeled lambda phages. The upper cell, infected by a single phage, proceeds to produce new viral particles and undergo lysis. The lower cell, co-infected by three phages, enters lysogeny, as indicated by a fluorescent reporter for P_{RE} activity, and proceeds to divide normally. (B) The fraction of cells undergoing lysogeny as a function of the multiplicity of infection, as measured from >1000 infection events. In contrast to the original modeling of the bulk data (Fig. 4 above), the single-cell curve rises gradually, suggesting that the MOI dependence is probabilistic rather than deterministic. (C) Incorporating the effect of intracellular viral concentration (MOI divided by cell volume) captures the experimental data and yields a decision curve that is markedly more step-like. Adapted from Zeng, L., Skinner, S.O., Zong, C., *et al.*, 2010. Decision making at a subcellular level determines the outcome of bacteriophage infection. Cell 141 (4), 682–691, Copyright 2010, with permission from Elsevier.

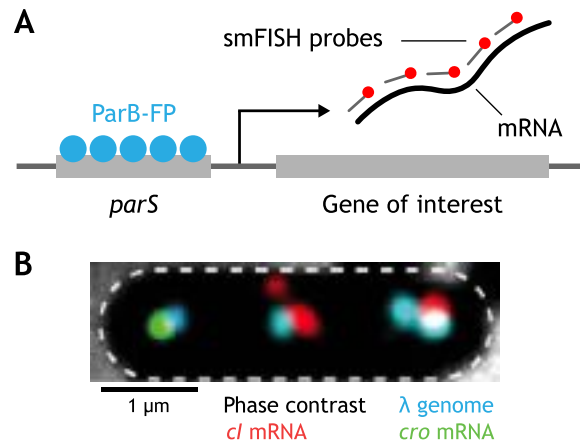


Fig. 6 Detecting the transcriptional activity of individual lambda phages. (A) Each phage is detected through the binding of fluorescently-tagged ParB proteins to the *parS* sequence, engineered into the lambda genome. mRNA molecules transcribed by the phage are simultaneously detected using single-molecule fluorescence *in situ* hybridization (smFISH). (B) Four lambda genomes (cyan) inside a single infected cell, at 10 min after infection. Individual phages vary in their transcriptional activity, with one transcribing *cro* (green) and two others producing *cl* (red).

But the fact that we can describe cell fate decisions probabilistically does not necessarily mean that we should settle for such a narrative and give up seeking a deterministic description of the decision process. While biochemical stochasticity is undisputedly present, automatically attributing all cellular indeterminacy to unknowable “noise” may be taking the easy path. One must consider the alternative hypothesis, which is, that our inability to predict the decision outcome reflects a failure to account for additional cellular variables that have a deterministic effect on the decision. So long as these “hidden variables” remain unknown to us, the decision will appear more random than it truly is, and our understanding of it remain limited.

And, in fact, a number of lambda studies suggest that incorporating additional variables can reveal a more precise decision at the single-cell level. It was first found that, for a given MOI, smaller cells are more likely to be lysogenized than larger ones. This should not surprise us, since decreasing the size of the infected cell is, to a first approximation, the same as infecting with a larger number of phages: both result in an increased concentration of viral gene products in the cell. Detailed analysis revealed that a unique arithmetic combination of the MOI and cell size yields a more step-like (and therefore, more deterministic) probability of lysogenization (Fig. 5(C)). The way in which the infection parameters combine to yield a sharp decision curve points to a nonlinear interaction between the co-infecting phages as they converge on the cell’s fate. Elucidating the nature of this interaction will require characterizing the spatiotemporal dynamics and genetic activity of individual phages within the infected cell. Fluorescent reporters for phage capsid, genome, RNA and protein products, needed for such an investigation, are now becoming available (Fig. 6).

The Decision to Remain Dormant

If, following infection, the lysogenic route is chosen, control of cell fate is then handed over from the post-infection decision circuit to a smaller circuit, whose role is to maintain dormancy by repressing all virulent functions. The maintenance circuit must also be able to trigger a switch back to lysis (induction) when cellular conditions change. Thus, the dormant virus continuously re-evaluates its lysis/lysogeny decision. Some elements of lysogenic maintenance in lambda and other phages are covered by a separate article in this encyclopedia (Shearwin). Here we focus on the decision-making aspect of the maintenance circuit.

In lambda, lysogenic maintenance is handled by a subset of the post-infection decision network discussed above. The smaller maintenance circuit, known as the lambda “switch”, consists of two phage genes, *cl* and *cro*, transcribed respectively from two diverging promoters, P_{RM} and P_R (Fig. 7). The two gene products, CI and Cro, compete for binding to six operator sites (O_{R1-3} and O_{L1-3}) that regulate P_{RM} and P_R transcription, resulting in mutual repression by the two proteins. In the prophage state, high cellular level of CI represses transcription from P_R and P_L , thus maintaining viral dormancy. The lysogenic state is further stabilized by the formation of a DNA loop between O_{R1-3} and O_{L1-3} , secured by oligomerization of CI dimers bound at the two loci. Perturbations that reduce the level of CI (such as activation of the bacterial SOS response, discussed below), can lead to lytic induction. In this process, the inhibition of P_R and P_L is relieved, leading to transcription of early lytic genes, including *cro*. Cro then represses P_{RM} , leading to further reduction of CI level and allowing the lytic cascade to proceed. Despite decades of meticulous studies, recent experiments continue to reveal new features of the lambda maintenance circuit, such as the role of mechanical coupling between transcription, DNA supercoiling, and looping, and how this coupling may affect the stability of the lysogenic state.

As in the case of infection, the phage (now, prophage)-encoded circuit requires input from the bacterial host in order to sense the state of the infected (now, lysogenized) cell and use that information to choose optimally between lysis and lysogeny. Specifically, during lysogenic maintenance, the role of host input is to alert the prophage when the bacterial cell is in danger, indicating that it is time to escape the host through the lytic pathway. Lambda receives this information via *E. coli*’s SOS system,

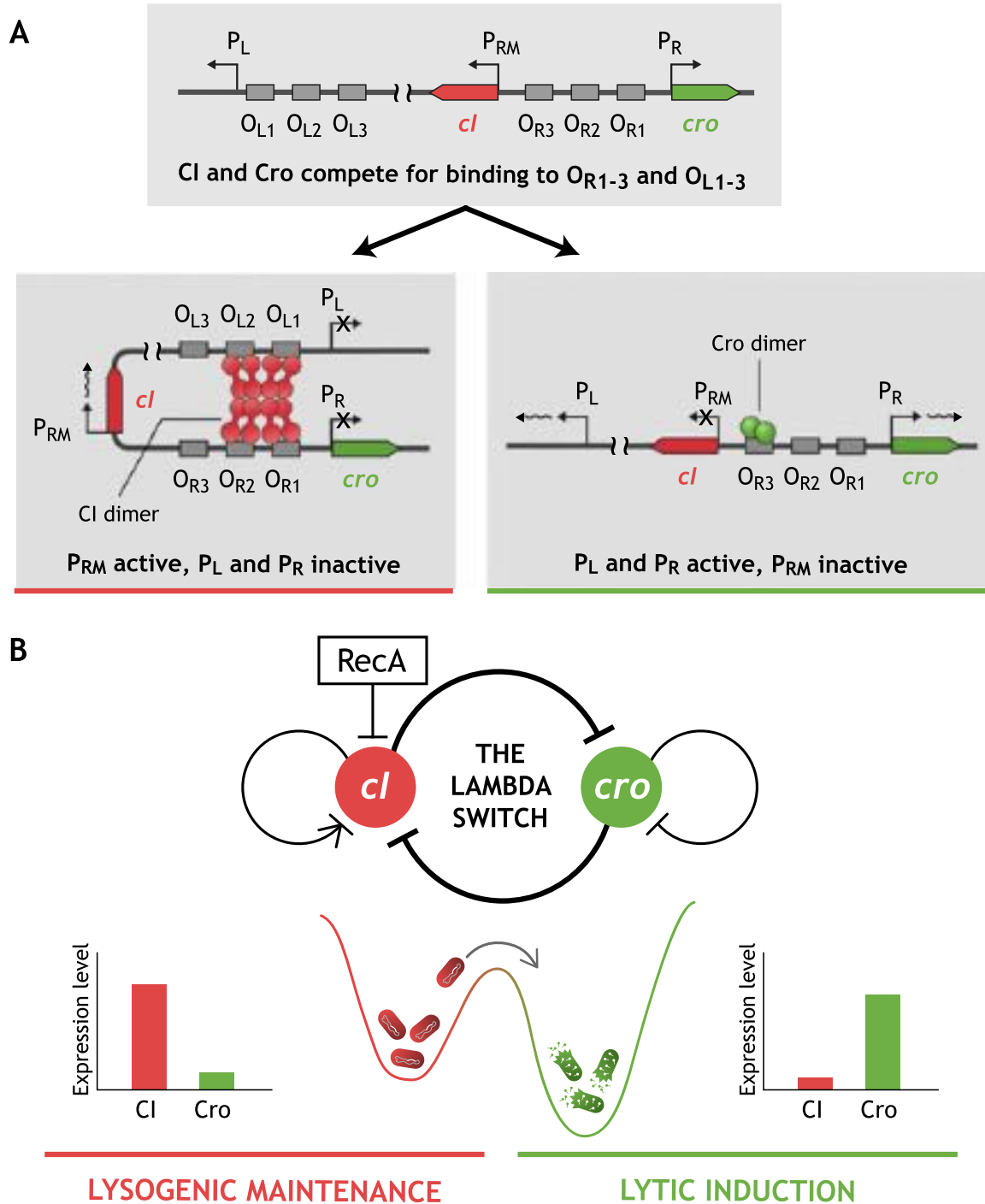


Fig. 7 The maintenance of lambda lysogeny. (A) The lysogenic state is maintained by a regulatory circuit consisting of CI and Cro, expressed from the P_{RM} and P_R promoters, respectively. CI and Cro compete for binding at six operator sites (O_{R1-3} and O_{L1-3}), to determine which promoters (P_{RM} , or P_R and P_L) are active, and thus decide whether lysogeny is maintained or, instead, lytic genes induction takes place. Two examples of binding configurations are shown. On the left, CI dimers bind to four operator sites, resulting in DNA looping that ensures repression of P_R and P_L during lysogeny. On the right, binding of Cro to O_{R3} represses transcription of CI from P_{RM} and allows lytic genes to be expressed. (B) The regulatory interactions between CI and Cro form a bistable switch. The system can alter its state in response to a large perturbation, such as depletion of CI by RecA, leading to lytic induction, but is immune to small perturbations. Adapted with permission from Golding, I., 2016. Single-cell studies of phage λ : Hidden treasures under Occam's rug. Annual Review of Virology 3 (1), 453–472, permission conveyed through Copyright Clearance Center, Inc.

which, in response to cellular DNA damage, halts progression of the cell cycle and triggers DNA repair and mutagenesis. Under normal growth, expression of the SOS response genes is repressed by LexA. However, in the presence of DNA damage (due to, e.g., UV radiation), regions of single-stranded DNA accumulate, leading to recruitment and activation of RecA. Activated RecA facilitates LexA self-cleavage, de-repressing SOS genes, including RecA itself. Phage lambda is coupled to the SOS response through the self-cleavage of CI by activated RecA. The consequent drop in cellular CI concentration leads to the relief of *cro* repression, activation of the lytic pathway, and prophage induction. Many other temperate bacteriophages are also induced following treatment with UV radiation or DNA-damaging agents. However, some, like P2, appear to be non-inducible, and immune to the bacterial SOS system.

The *ci/cro* pair serves as a canonical example for a so-called “toggle switch”, a genetic module exhibiting two stable states, here corresponding to lysogeny and to the onset of lysis. Despite consisting of only two genes, the lysogeny maintenance circuit of lambda captures key features of cell-fate choice, as observed across the spectrum of biological complexity. Through the use of feedback (P_{RM} autoregulation by CI), the system achieves extremely high stability in the absence of external perturbations, with fewer than one spontaneous switching event per 10^6 cell doublings. At the same time, almost 100% of lysogenic cells switch to lysis in response to an inducing signal. These properties have made the lambda maintenance circuit an attractive starting point for understanding cellular differentiation and reprogramming in metazoans. In particular, lambda has served as a fertile test ground for the attempt to formulate a detailed biophysical description of cellular behavior, in the form of a mathematical model that uses the known molecular interactions to predict the resultant cellular phenotype. This effort has been, at least partly, successful. For example, a thermodynamic model can be written, describing the different binding configurations of CI at the operator sites that control transcription from P_{RM} , and this model used to predict the regulatory curve relating CI concentration in the cell to P_{RM} activity. The theoretically predicted curve shows good agreement with experimental measurements of this regulatory relation (Fig. 8). The theoretical calculations can

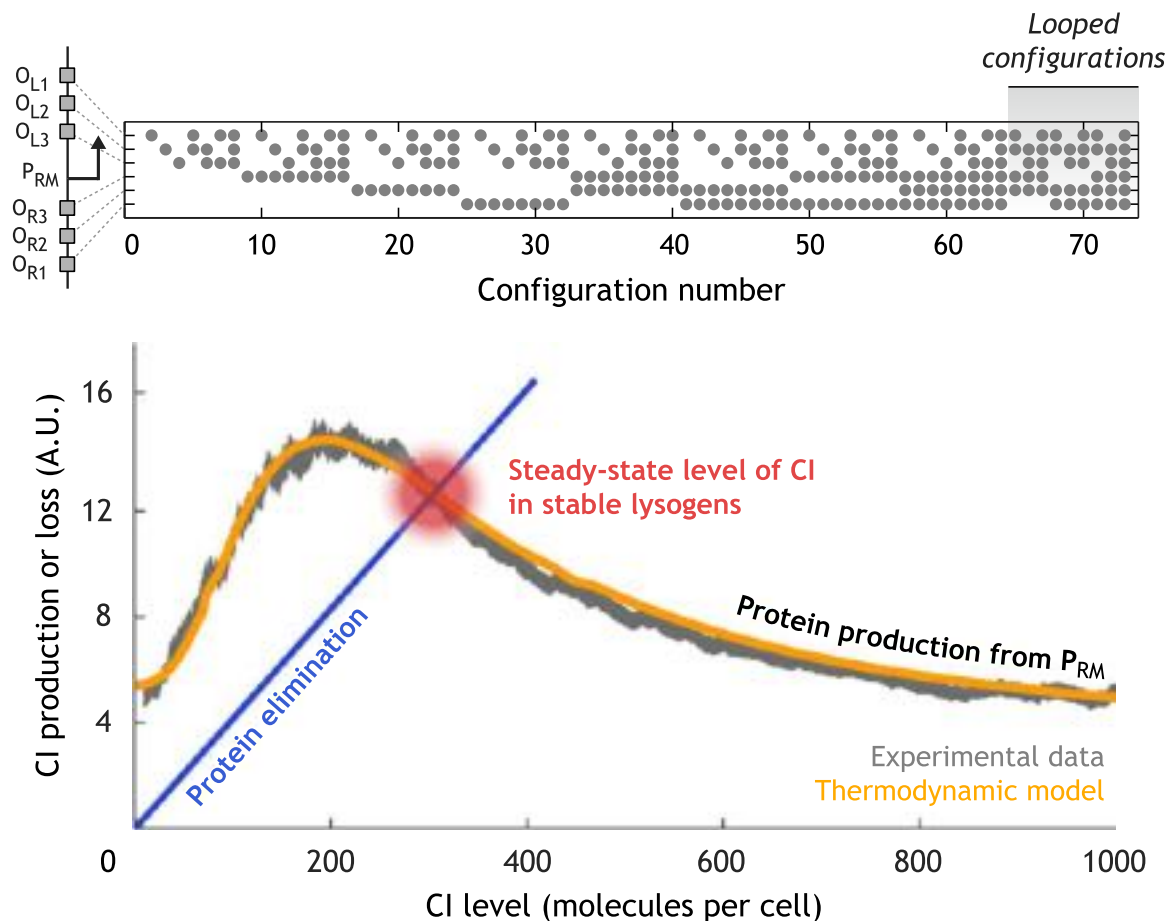


Fig. 8 A theoretical biophysical model captures P_{RM} regulation by CI. The regulatory curve relating CI concentration in the cell to P_{RM} activity can be predicted from a thermodynamic description of the possible binding configurations of CI at the O_{R1-3} and O_{L1-3} operator sites. The theoretically predicted curve agrees with the experimental measurement of the regulatory relation. The average amount of CI protein present in each lysogenic cell can be estimated by requiring that CI production is exactly balanced by CI elimination (via dilution, due to cell growth and division). Adapted from Sepúlveda, L.A., Xu, H., Zhang, J., Wang, M., Golding, I., 2016. Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science* 351, 1218–1222, reprinted with permission from AAAS; and from Golding, I., 2011. Decision making in living cells: Lessons from a simple system. *Annual Review of Biophysics* 40 (1), 63–80 with permission conveyed through Copyright Clearance Center, Inc.

next be utilized to estimate the amount of CI protein present in a lysogenic cell, by requiring that CI production is exactly balanced by CI elimination (via dilution, due to cell growth and division) (Fig. 8).

The eventual test for a theory of the lambda switch is to successfully predict the key phenotype, namely, whether a given cell will remain in the lysogenic state or, instead, switch to lysis. In attempting to answer this question, we are again confronted with the challenge of cellular individuality: In the absence of an external signal, only one out of a million lysogens in a growing culture will spontaneously switch. Can we predict which cell this will be? It is widely believed that the answer is negative, and that we may only aspire to predict the probability of induction, not the actual fate of an individual cell. This is because spontaneous induction is considered a stochastic process, driven by random fluctuations of CI levels in the cell. Small drops in CI number will be corrected by the negative feedback in the P_{RM} -CI circuit, raising CI level and thus reverting to the mean. However, rare, larger drops will overcome the feedback and lead to de-repression of P_R , Cro production and onset of the lytic pathway. In this picture, spontaneous lytic induction is analogous to the way random thermal motion drives a physical system to transition from one stable state to another (Fig. 7 above). The challenge of theoretically predicting the behavior of the *ci/cro* switch dwarfs in comparison to the larger goal of predicting cell fate following infection, when the full decision network (Fig. 2 above) comes into play.

Conclusion

The prevailing narrative for the lambda lysis/lysogeny decision, both following infection and during lysogenic maintenance, offers only a probabilistic prediction, rather a deterministic one, as to which path an individual cell will choose. This probabilistic point of view is similarly applied, further afield, to the choice of latency by mammalian viruses and to cellular differentiation and reprogramming. These decisions are all held to be indeterminate, noise-driven processes. Bacteriophage lambda, where this picture originally emerged, also bears the potential to challenge the probabilistic view by revealing previously-hidden variables that bias the decision outcome, or even determine it in full. Future studies, describing the lysis/lysogeny decision in individual phages and cells, in real time, will be key to delineating true randomness from the hidden precision of cellular decision-making.

Acknowledgments

We are grateful to Ian Dodd and Keith Shearwin for commenting on an earlier draft of this article. Work in the Golding lab is supported by grants from the National Institutes of Health (R01 GM082837), the National Science Foundation (PHY 1147498, PHY 1430124 and PHY 1427654), the Welch Foundation (Q-1759) and the John S. Dunn Foundation (Collaborative Research Award). We gratefully acknowledge the computing resources provided by the CIBR Center of Baylor College of Medicine.

Further Reading

- Arkin, A., Ross, J., McAdams, H.H., 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Annual Review of Genetics* 149 (4), 1633–1648.
- Casjens, S.R., Hendrix, R.W., 2015. Bacteriophage lambda: Early pioneer and still relevant. *Virology* 479–480, 310–330.
- Dodd, I.B., Shearwin, K.E., Perkins, A.J., *et al.*, 2004. Cooperativity in long-range gene regulation by the λ CI repressor. *Genes & Development* 18 (3), 344–354.
- Erez, Z., Steinberger-Levy, I., Shamir, M., *et al.*, 2017. Communication between viruses guides lysis-lysogeny decisions. *Nature* 541 (7638), 488–493.
- Golding, I., 2011. Decision making in living cells: Lessons from a simple system. *Annual Review of Biophysics* 40 (1), 63–80.
- Golding, I., 2016. Single-cell studies of phage λ : Hidden treasures under Occam's rug. *Annual Review of Virology* 3 (1), 453–472.
- Knowles, B., Silveira, C., Bailey, B., *et al.*, 2016. Lytic to temperate switching of viral communities. *Nature* 531 (7595), 466–470.
- Oppenheim, A.B., Kobiler, O., Stavans, J., *et al.*, 2005. Switches in bacteriophage lambda development. *Annual Review of Genetics* 39 (1), 409–429.
- Ptashne, M., 2004. *A Genetic Switch: Phage Lambda Revisited*, third ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Silpe, J.E., Bassler, B.L., 2019. A host-produced quorum-sensing autoinducer controls a phage lysis-lysogeny decision. *Cell* 176 (1–2), 268–280.
- St-Pierre, F., Endy, D., 2008. Determination of cell fate selection during phage lambda infection. *Proceedings of the National Academy of Sciences of the United States of America* 105 (52), 20705–20710.
- Tal, A., Arbel-Goren, R., Costantino, N., Stavans, J., 2014. Location of the unique integration site on an *Escherichia coli* chromosome by bacteriophage lambda DNA in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 111 (20), 7308–7312.
- Weitz, J.S., Beckett, S.J., Brum, J.R., *et al.*, 2017. Lysis, lysogeny and virus-microbe ratios. *Nature* 549 (7672), E1–E3.
- Zeng, L., Skinner, S.O., Zong, C., *et al.*, 2010. Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell* 141 (4), 682–691.

Mobilization of Phage Satellites

Kristen N LeGault and Kimberley D Seed, University of California, Berkeley, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Accessory gene A gene that is not required for the lifecycle of the mobile genetic element, but often confers a fitness advantage to the bacterial host, for example, by encoding genes involved in virulence.

Burst assay Quantification of the number of bacteriophage progeny produced by one infected host cell.

Capsid Protein shell in which the phage genome is packaged.

Helper phage A phage that induces transfer of a satellite phage.

Lytic (virulent) phage A phage that takes over the host to produce progeny phage culminating in lysis of the host cell, and is unable to integrate into the host chromosome or otherwise lysogenize the bacterial host.

Mobile genetic element (MGE) DNA capable of moving within or between genomes.

Phage inducible chromosomal island (PICI) A genetic element that is induced to excise, replicate autonomously and package itself into helper phage structural components.

Phage-inducible chromosomal island-like elements (PLEs)

Anti-phage islands found in clinical isolates of *Vibrio cholerae*, specifically induced by, and inhibitory toward, the lytic phage ICP1.

Satellite phage Subviral element lacking genes encoding functions needed to complete the phage lifecycle. Thus for their multiplication, satellites parasitize a helper phage during co-infection of a host cell.

Staphylococcus aureus pathogenicity island (SaPIs)

Prototypical PICI elements that are ubiquitous in *S. aureus* and carry virulence genes.

Temperate phage (prophage) A phage that can undergo either a lytic cycle or integrate into the host genome to be transmitted vertically until induced by environmental cues to re-enter the lytic cycle.

Transduction Packaging and transmission of bacterial DNA inside phage virions to new host cells.

Introduction

Phages profoundly impact the evolutionary landscape of their bacterial hosts, both through predation, which selects for hosts with defenses to overcome phage killing, and through mobilization and dissemination of genetic material. Phages are the most abundant biological entities on the planet, making phage-mediated genetic transfer a critically important mechanism of gene mobilization between bacteria. Phages with strictly lytic lifecycles infect susceptible bacterial cells and redirect cellular metabolism to promote phage replication, resulting in the release of infectious phage progeny through cell lysis. In contrast to lytic phages, temperate phages are able to integrate into their host genome where they can be transmitted vertically. It is well-established that integrated phages can dramatically alter bacterial phenotypes by providing their host with novel genes. One of the most clinically relevant examples is the CTX prophage integrated into the genome of epidemic *Vibrio cholerae*; CTX encodes cholera toxin, the virulence factor responsible for the profuse watery diarrhea that is the hallmark of the disease cholera. For many temperate phages, signals that trigger the prophage to re-enter into the lytic cycle culminate in the lysis of the host cell and the release of progeny phage. Both lytic and temperate phages can occasionally mis-package regions of the host genome and move it to new recipient cells in a process called generalized transduction. Additionally, temperate phage can excise along with adjacent areas in the host genome, and move these regions at a higher frequency in a process called specialized transduction.

In sharp contrast to these forms of “passive” transduction, there are mobile genetic elements, the phage satellites, that have evolved to precisely exploit phages for high-frequency transfer. These subcellular elements are consequential for bacterial and phage evolution as both providers of novel genetic content and as parasites of phages. Their impact can often go overlooked or be misinterpreted, as observing the transfer of a satellite phage requires a biological relevant tripartite system: the satellite phage, a compatible helper phage and a susceptible bacterium. It was initially thought that the genetic element encoding the toxic shock syndrome toxin was contained on a prophage in *Staphylococcus aureus*, though in the late 1990s it was discovered that this toxin containing island was able to excise, replicate and package itself not as an independent temperate phage, but rather through parasitizing a temperate phage for mobilization. This was demonstrated to be an autonomously replicating pathogenicity island, and was designated SaPI1, for “*Staphylococcus aureus* pathogenicity island 1”. Since this discovery, much has been learned about the SaPI lifecycle, including how SaPIs are induced by specific phages, how they replicate and efficiently package their genomes into virions leading to the spread of virulence genes across large phylogenetic distances. Most *S. aureus* virulence factors are found on these phage satellites, highlighting the phage-dependent lifestyle as a successful means to efficiently spread genetic material. Importantly, SaPIs are the archetype of a larger group of MGEs known as phage-inducible chromosomal islands (PICIs) which have only recently been found to be widespread among Gram-positive bacteria, and prevalent among Gram-negative Gamma-proteobacteria. PICIs have several features in common, which will be discussed below. Other newly discovered PICI-like elements

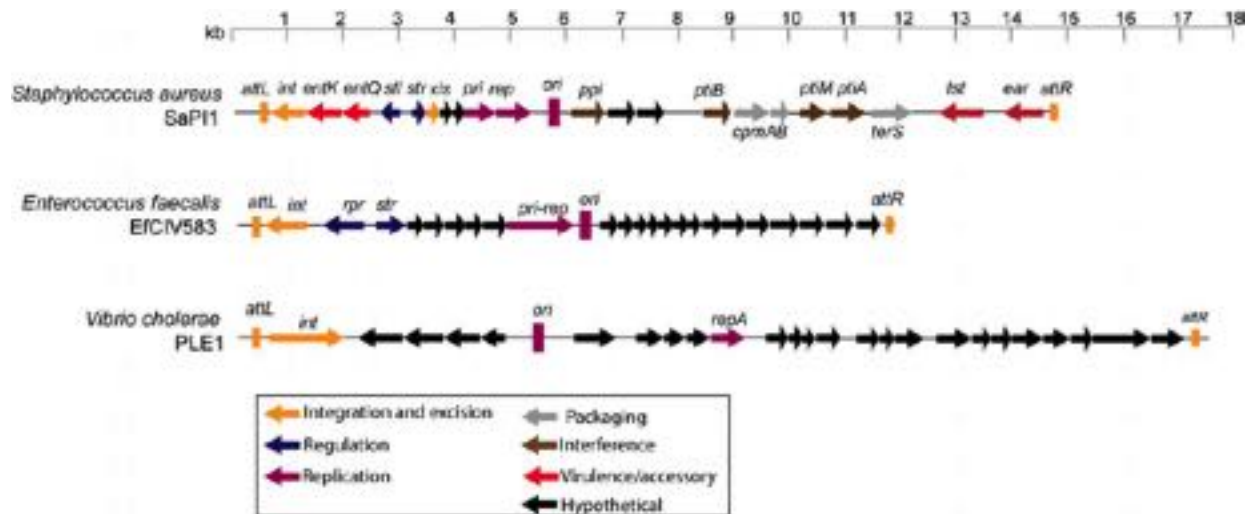


Fig. 1 Genetic organization of representative phage-inducible chromosomal islands. Representative genomes of Gram-positive (SaPI1 and EfCIV583) PICIs and *V. cholerae*'s PLEs (PLE1). All of these satellites share several features, including gene(s) involved in integration and excision from the bacterial chromosome (Int and Xis as well as the left and right attachment sites) in yellow, genes involved in regulation of gene expression in blue (although genes involved in PLE gene regulation have not yet been identified), a replication module consisting of an origin of replication (Ori) and replication genes in purple required for PICI replication, as well as genes involved in packaging and interference with their helper phage in grey. SaPIs additionally encode several genes involved in virulence, including production of toxic shock syndrome toxin (*tsf*), enterotoxins K and Q (*entK* and *entQ*), and *ear* conferring ampicillin resistance. In black are genes encoding hypothetical proteins.

(PLEs) do not share obvious homology to known PICIs, yet are also induced by specific phages, suggesting that the known repertoire of genomic islands mobilized by phages only represent the tip of the iceberg when it comes to phage-mediated genome flux.

Where there is experimental evidence, we will compare genetic elements that are mobilized by phages, highlighting what seem to be generalizable and convergent strategies that these elements use to effectively co-opt their inducing phage, as well as what is unique about these different islands. As we will see, the ability of phages to specifically mobilize regions of their host's chromosome involved in pathogenesis and phage resistance cautions against the use of phage in treating certain bacterial infections, as such mobility has the potential to stimulate evolutionary innovation and undermine the apparent benefits of harnessing phage as antibacterial agents.

Genetic Organization

PICIs and PLEs share many conserved features, which underscores several key elements of their lifecycle as phage satellites and also allows for identification of novel PICIs through sequence analysis. All PICI genomes are characterized by (1) specific attachment sites in their hosts genome, (2) a lack of phage structural or lytic genes and (3) a size of around 15 kb. PICIs have analogous genetic organization to phages, with genes clustered according to function. SaPIs have genes arranged in four modules: regulation, integration and excision, autonomous replication, and packaging (Fig. 1). Functionally, all PICIs discovered thus far also have strategies to interfere with helper phage propagation. In addition to the modules that allow for the propagation of PICIs, many contain accessory genes that result in the phenotypic conversion of a host carrying these mobile genetic elements to a more virulent state. PLEs appear to also have their genes arranged in operons, though PLE-encoded ORFs are not homologous to known PICI genes and few PLE-encoded gene products have known functions. Like the PICIs, PLEs lack obvious phage structural or lytic genes and they have unique attachment sites, though with genomes ranging from 18 to 19 kb, PLEs are larger than the archetypal SaPIs, and, as of now, have not been shown to be capable of autonomous replication in the absence of their inducing phage. Below we will discuss how PICIs and PLEs progress through their lifecycle starting with integration and induction.

Integration and Excision, Gene Induction, and Replication

Both PICIs and PLEs are integrated in a site-specific manner in their bacterial host genome. When integration occurs, the attachment site in the PICI, known as the *attP* site (*attS* for SaPIs), recombines with the bacterial chromosomal attachment site, *attC*, resulting in the creation of unique left and right attachment sites when integrated (*attL* and *attR*, respectively). An integrase (Int), a hallmark of MGEs such as satellite phage, drives the recombination between the bacterial chromosome and the incoming island. For SaPIs, Int is a tyrosine integrase and for PLEs, Int is a large serine recombinase. In both systems, integration of the

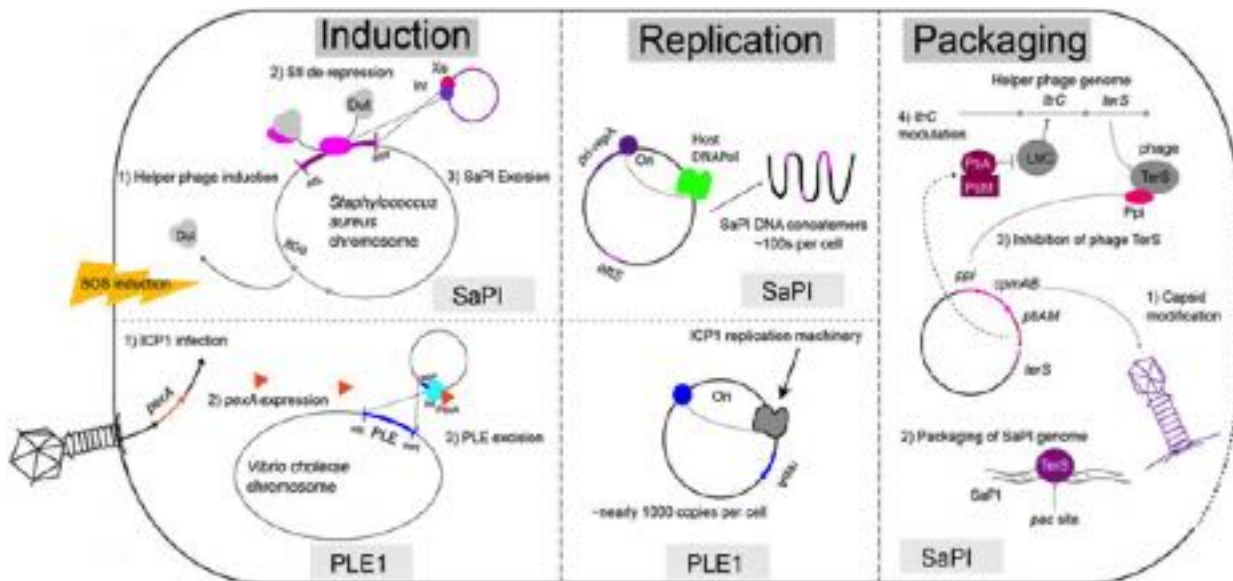


Fig. 2 The lifecycle of SaPI and PLE showing induction, replication and packaging. A schematic of the steps of a PICI lifecycle, contrasting an archetypal SaPI on the top and PLE on the bottom. SaPIs and PLEs are induced by specific helper phage, yet only SaPIs are induced by a temperate phage, which can be activated by the SOS response. SaPIs and PLEs replicate upon induction by their helper phage to several copies to nearly 1000 copies in the case of PLEs. Both PLEs and SaPIs transmit horizontally, but since it is unknown how PLE transduces, only the SaPI packaging is shown. SaPIs encode several mechanisms which interfere with helper phage, ensuring specific packaging and horizontal transmission of only the SaPI genome.

satellite phage is catalyzed by *Int* recognition of the *attC* site, and additional proteins are not required for acquisition. SaPIs integrate into one of six primary *attC* sites, while two *attC* sites are known for PLEs: PLEs 1, 3, 4 and 5 integrate into the *V. cholerae* repeat (VCR) while PLE 2 integrates into the hypothetical gene VCA581. The use of the VCRs as an attachment site for most PLEs provides immense flexibility for PLE integration, as VCRs are present in > 100 copies in *V. cholerae*.

Int is also required for excision of the satellite phage from the chromosome, but it is not sufficient. SaPIs encode an excisionase, *Xis*, which is required for excision and is only expressed upon helper phage mediated de-repression of SaPI genes. PLEs also require the presence of their inducing phage to excise, though the reason for this dependence differs. In contrast to SaPIs, which encode their own excisionase, PLEs utilize a phage-encoded recombination directionality factor to trigger excision. PLEs' *Int* is constitutively expressed, and upon infection by the helper phage ICP1, *Int* binds to the ICP1-encoded recombination directionality factor to catalyze excision. For PLEs 1,3,4 and 5 the recombination directionality factor has been shown to be *PexA*, which is uniquely encoded by ICP1 and underpins the specificity of PLE induction by ICP1. The recombination directionality factor required for PLE 2 *Int* is not yet known, though it is not *PexA*, and evidence points to it also being uniquely encoded by ICP1.

PICI and PLE gene expression is also responsive to the presence of their respective helper phages. Although it is unknown how exactly PLE gene expression is induced by ICP1 infection, induction of the Gram-positive PICI genetic program is well understood. Prior to infection or induction by a helper phage, Gram-positive PICIs, like SaPIs, are thought to be transcriptionally quiescent in the bacterial hosts' genome. A SaPI-encoded master repressor *Stl* prevents gene expression from the left and right SaPI operons. Upon infection by a specific helper phage, a phage-encoded protein disrupts the *Stl*-DNA complex, resulting in SaPI de-repression (Fig. 2). It is thought that Gram-negative PICIs use a different induction strategy than that of the Gram-positive PICIs. Rather than rely on a master repressor which is then de-repressed by a phage-encoded inducer protein, PICIs in Gram-negative bacteria use an *AlpA* activator to induce expression, though it is unknown how *AlpA* responds to infection by helper phage. The mechanism through which ICP1 triggers PLE gene expression is not yet known, but PLE gene expression appears to be dependent on the presence of ICP1.

The responsiveness of gene expression to the SOS response differentiates Gram-positive PICIs from Gram-negative PICIs and PLEs described thus far. The SOS response is a bacterial stress response, triggered by the presence of ssDNA that can be induced in the laboratory by bombardment with UV irradiation or treatment with certain drugs like Mitomycin C or antibiotics like fluoroquinolone. *LexA* is a global repressor of the SOS regulon, which comprises a large number of genes that are repressed by *LexA* in the absence of SOS induction. Once de-repressed, these genes change the state of the cell, resulting in the cessation of translation and cell division. For many prophages, a homolog of *LexA* directly represses the lytic pathway, maintaining the prophage in a quiescent state in un-stressed cells. *S. aureus* hosts that contain an integrated helper prophage that are sensitive to the SOS response can trigger the excision and de-repression of SaPIs upon SOS induction. SaPIs are therefore indirectly triggered by the SOS response to mobilize and horizontally transmit to naïve cells, creating a link between antibiotic treatment and increased spread of virulence factors. Expression of the Gram-negative PICI activator, *AlpA*, is not directly impacted by the SOS response,

however its expression is likewise dependent on a yet to be identified phage-encoded inducer, and so it is possible that AlpA-activated PICIs are also indirectly induced by the SOS response. *V. cholerae*'s PLEs are not thought to be induced by the SOS response, as their induction is specific to ICP1, a lytic phage, rather than temperate phage.

Upon phage-triggered induction and excision from the chromosome, PICIs and PLEs begin replicating. SaPIs utilize their own primase (Pri), replication initiation protein (Rep) and origin (*ori*) in conjunction with their host's bacterial DNA polymerase to replicate as DNA concatemers. PLE replication is likewise dependent on the PLE-encoded replicon comprised of an *ori* and a replication initiator protein RepA. A striking difference between PLEs and PICIs is in their dependence on phage-encoded products for replication: PICIs replicate autonomously, with their helper phage playing a primary role in transcriptional activation. For example, expression of the *pri-rep-ori* module on a plasmid is sufficient to drive replication of that plasmid in several Gram-positive hosts, indicating that host machinery is capable of replicating PICIs from Gram-positive and some experimentally validated Gram-negatives. In contrast, in PLE 1 (for which replication has been characterized), replication itself is dependent on ICP1-encoded products, indicating that PLEs have evolved to parasitize phage-encoded replication proteins. In support of the model of PLE-mediated replication hijacking, PLE replication is associated with decreased ICP1 DNA replication, whereas that has not been reported to be the case for PICIs.

Packaging and Horizontal Transmission

All satellite phage known to date use their helper phage's structural components to package and move their genomes to new bacteria in a form of transduction. Many SaPIs rely on modification, specifically re-sizing, of their helper phage's capsid. These smaller capsids are about one-third the size of helper phage capsids, and therefore are only able to accommodate the SaPI genome, which for many SaPIs is about one-third the size of the helper phage genome. Morphogenesis of a smaller capsid for SaPIs is accomplished by SaPI-encoded proteins, which can act as capsid scaffolds. Capsid modification is one mechanism SaPIs use to interfere with their helper phage (discussed further below).

PICIs, like all phages, use two strategies for packaging: a headful mechanism dependent on a terminase complex and *pac* site, or a *cos*-site dependent mechanism. Many sequenced PICIs of both Gram-positive and Gram-negative bacteria encode their own small terminase subunit, TerS, which recognizes the PICI encoded *pac* site, bringing it to the phage-encoded TerL subunit to preferentially package the PICI genome. However, several PICIs lack a TerS homolog, yet package and transmit at a high frequency. It was first shown that some SaPIs, such as SaPIbov5, utilizes both a headful mechanism for packaging as well as a *cos*-site mechanism. Indeed, PICIs that encode a phage *cos*-site are more common than initially appreciated, and are found in lactococci as well as in Gram-negative PICIs from *Escherichia coli*. PICIs that are packaged via *cos*-sites use an unrelated protein Ccm (for *cos* capsid morphogenesis) which interacts with the phage capsid protein to direct formation of smaller phage capsids to only accommodate the smaller SaPI genome, reducing the titers of helper *cos* phage. By bioinformatics analyses, several Gram-negative PICIs appear to encode for their own capsid monomer, potentially explaining how these PICIs are packaged into smaller capsids despite not possessing the genes *cpmAB*. *V. cholerae*'s PLE has been shown to transduce to naïve cells, but since few PLE-encoded gene products possess homology to any known proteins, there are no clear homologs of Ccm or CpmAB and the structural makeup and morphology of PLE transducing particles remains to be elucidated. Interestingly, the disparity in genome size between PLEs and their only known helper phage ICP1 is much greater than other PICI systems: the genomes of PLEs are one sixth the size of the ICP1 genome, so it remains to be determined if the same tactics used by SaPIs to redirect capsid assembly are utilized by PLE. Nonetheless, horizontal transfer of PLEs requires that the recipient cells possess ICP1's surface receptor, a component of the lipopolysaccharide O-antigen, suggesting that PLE transduction relies on ICP1 structural components. Regardless of how PLEs transmit themselves horizontally, they likely interfere with ICP1 particle assembly, since no infectious phage are produced following ICP1 infection of PLE-containing *V. cholerae*. It is likely a generalizable phage satellite strategy to steal packaging resources to promote horizontal transmission, as this ploy accomplishes both interference with the helper phage and promotes dispersal of the satellite phage genome.

Strategies PICIs Use to Interfere With Their Helper Phages

All PICIs discovered thus far confer a fitness benefit to their host bacterium by interfering with the propagation of their helper phages. As we have discussed, the co-dependence of PICIs on their helper phage for transmission means that natural selection would favor PICIs that only target vulnerable nodes in the phage lifecycle which are not self-defeating. Since most phage proceed through their lifecycle in a temporally regulated manner, where events late in infection depend on the success of those early in infection, PICIs have evolved not to interfere with early stages of their helper phage lifecycle, as they rely on the production of late gene products for their own transmission. PICIs therefore have finely tuned mechanisms that target assembly of their helper phage virions. In well-studied SaPIs three interference mechanisms have been described, and are used to varying extents by different SaPIs. First, as discussed above, SaPIs alter their helper phage capsids to produce smaller particles, restricting the packaging of the larger helper phage genome so that only the smaller SaPI genome can be packaged. Additional mechanisms of interference are found in the same gene cluster as the *cpmAB* operon (Fig. 1). A second mechanism further ensures that only the SaPI genome is packaged by targeting the helper phage terminase. The Ppi protein binds to the phage TerS, directly blocking the packaging of

helper phage DNA. Ppi does not interact with the SaPI TerS protein, ensuring that SaPI genome packaging proceeds uninhibited. A homolog of *ppi* is found in all sequenced SaPIs, as well as in several Gram-positive bacteria that have not been functionally shown to carry a satellite phage, indicating that additional phage satellites are waiting discovery. A third mechanism helps SaPIs interfere with certain helper phages through the PtiA, PtiB and PtiM proteins which interact with the helper phage late gene cluster. The late gene cluster in the helper phage genome encodes for structural components of the phage particles and host cell lysis. PICI helper phage regulate this late gene cluster by a member of the later transcription regulator" (Ltr) superfamily of proteins. The SaPI protein PtiA directly binds to LtrC to inhibit its activation of late gene expression. The gene immediately upstream of PtiA, PtiM, is co-translated with PtiA and acts as a modulator of the LtrC-PtiA interaction (Fig. 2). PtiM binds to PtiA, decreasing its ability to bind LtrC and allowing production of some structural components for SaPI packaging, as well as the lysis genes required for release of transducing particles. A third protein, PtiB, is known to interact with PtiAM, but is toxic to the host cell when expressed and its role in modulating late gene expression is not known.

As discussed above, available evidence suggests that PLEs also transmit horizontally using ICP1 structural components, although the molecular basis for PLE-mediated phage inhibition remains to be determined. PLEs may encode proteins with functions analogous to satellite phage that divert structural components to promote horizontal transmission, and those functions would likely also inhibit ICP1 assembly. However, there are several interesting differences between PICIs and PLEs that operate at the level of phage interference. The most striking of these differences is that PLEs completely block ICP1 production such that no new ICP1 progeny are produced during a burst assay. In contrast, SaPI1 reduces, but does not block, the reproduction of its helper phage 80 α in a comparable burst assay. This could indicate that all PLEs target several distinct assembly hubs, such that absolutely no ICP1 particles can assemble, and/or that the PLE-mediated decrease in ICP1 DNA replication compounds the effects on assembly to ensure that no ICP1 can overcome interference. It has also been observed that PLEs can accelerate cell lysis following ICP1 infection, which again may act in concert with other strategies to ensure complete inhibition of ICP1. PLEs appear to only provide a benefit to their *V. cholerae* host at the population level, as the infected cell always dies, but since no infectious ICP1 are produced, uninfected neighboring cells survive. In contrast to PLEs, PICIs known to target phage DNA replication or phage-mediated lysis have not been reported.

Prevalence of Phage Satellites

Satellite phage that use a helper phage to trigger excision and spread themselves through parasitizing phage structural components appear to be widespread in many Gram-positive bacteria that are important for human health and disease, such as staphylococci, streptococci and lactococci. In addition, a mobile genetic element in *Enterococcus faecalis*, EfCIV583, was recently shown to be functional and use Φ 1 as an inducing and packaging helper phage. EfCIV583 possesses many of the defining features of SaPIs (int/xis function, replication mediated by the elements Pri-Rep-ori cluster, repression by a master regulator Rpr), except that it lacks the TerS subunit that SaPIs use to direct packaging of their own genome.

In addition to PLEs in *V. cholerae*, a recent study has identified PICIs in several Gram-negative bacteria. Putative PICIs have been identified in several strains of *E. coli* and *Pasteurella multocida*, with their functionality demonstrated in *E. coli*, EcCICFT073 and EcCIO42, and from *P. multocida*, PmCIATCC43137. Several differences between Gram-negative and Gram-positive phage satellites suggest that they evolved multiple times in the different lineages. These studies likely underestimate the prevalence of comparable islands however, since bioinformatic approaches have relied on conserved genetic organization to identify PICIs in Gram-negative bacteria, and PLEs that have no known homology to any PICIs and they possess a distinct genetic organization (Fig. 2). Perhaps the earliest characterized phage parasite is the P4 satellite phage found in *E. coli*, which co-opts the phage P2 for key stages of its lifecycle. P4 relies on induction or infection by its helper phage P2 for activation and packaging, while encoding its own replicon. Familiarly, P4 interferes with P2 particle formation by altering capsid morphology to only allow successful packaging of the smaller P4 genome. The similarities between the P2-P4 system and the PICI-helper phage systems we have discussed here were once overlooked because of its distinct genomic structure from the Gram-positive SaPIs. However, as we have seen, the discovery of additional PICIs with alternative arrangements suggests that there is not one formula for a PICI, and the functional similarities between the P2-P4 system and known PICIs suggests these genetic elements should all uniformly be considered phage-satellites.

The convergent evolution of PICIs suggests that life as a phage satellite is a successful strategy to allow for genome reduction, and therefore, would be selected for in multiple bacterial hosts where infection by a specific phage is a reliable feature of the host's biology over time. Unbiased approaches to study genome flux following phage infection are needed to discover the true breadth of MGEs that harness phage for their own spread.

Counter-Evolution by Phages to Avoid Parasitism by Satellites

In the known cases of helper phage parasitism by satellite phages, the excised elements limit the spread of the helper phage – partially in the case of PICIs, or completely inhibiting ICP1 in the case of PLEs. This partial or complete inhibition of helper phage propagation exerts strong purifying selection on the phage to overcome parasitism by its satellite phage. As discussed above, expression of the majority of SaPI genes is under the control of the global repressor Stl. Helper phages de-repress SaPIs by producing a non-essential phage protein that binds to Stl to disrupt the Stl-DNA complex leading to induction of the SaPI

program. Interestingly, these inducing proteins are structurally unrelated, yet all interact with the *Stl* repressor to perform the same function. Helper phage can therefore overcome SaPI interference by acquiring mutations in the inducing protein. Sequence analyses of known SaPI inducing proteins have shown that they are under purifying selection, which, interestingly increases the allelic variance at this locus. Experimentally, *80α* encodes at least four anti-repressor proteins, one of which induces three different SaPIs that all have the same *Stl*: SaPI_{bov1}, SaPI_{bov2} and SaPI₁. Within three passages of *80α* on each of these SaPIs, phages that do not induce SaPIs are selected for, demonstrating that phage readily evolve mutations in the inducing proteins. Interestingly, although these proteins are not essential, phage that possess a mutant version of the SaPI-inducing protein are less fit than wild type phage in the absence of SaPI pressure. *Dut* is a phage-encoded dUTPase, which maintains low cellular dUTP relative to dTTP in order to prevent the introduction of uracil into DNA, and serves as the inducer for SaPI_{bov1} and SaPI_{bov5}. Between phages *Duts* show striking allelic variation and ability to induce SaPIs, suggesting the pressure to avoid inducing SaPIs has driven this diversification. Continued passaging of SaPIs with their mutant non-inducing helper phage results in selection for SaPIs that typically have a mutation in *Stl*, resulting in their constitutive expression which can be detrimental to their bacterial host. Phage-encoded inducer proteins and SaPIs are in a dynamic arms race, where a helper phage evolves resistance to SaPIs by mutating or losing the inducer gene (assuming the fitness costs are not too high relative to benefits). SaPIs then acquire mutations in their repressor so they are constitutively expressed in the absence of their helper phage, meaning they are unable to be packaged and spread by their helper phage. This removes the adaptive advantage that the mutant SaPI allele has and they are lost in the population, favoring again the SaPIs with compatible regulation which can then spread using their helper phage.

Despite many parallels between PLEs and PICIs in terms of mobilization by specific inducing phage and phage interference, induction of PLEs appears to show little functional similarity to the canonical SaPIs. PLE excision requires the previously described, ICP1-encoded gene product *PexA*. It is possible to generate $\Delta pexA$ ICP1 mutants in the laboratory that fail to catalyze PLE excision, however mutations in *PexA* are not seen in natural isolates of ICP1. *PexA* mutations are likely not selected for in nature however, because, surprisingly, PLEs still inhibit ICP1 without excision and circularization. In keeping with this observation, it has not been possible to successfully evolve ICP1 to avoid inducing PLEs *in vitro* with sequential passaging as has been done to study the SaPI-helper phage paradigm. These findings indicate that PLEs may have evolved to respond to several distinct ICP1-encoded inducers, such that ICP1 may not be able to overcome PLE-mediated inhibition through mutation and succeed in nature. However, PLEs do impose a significant evolutionary burden on ICP1 in nature as evidenced by some isolates of ICP1 acquiring an anti-PLE defense strategy in the form of a Type 1-F CRISPR-Cas system. Much in the way that CRISPR-Cas functions as an adaptive immune system to target foreign DNA for destruction in bacteria, the ICP1-encoded CRISPR-Cas system deploys CRISPR-RNAs transcribed from spacers that are complementary to PLE sequence to guide an effector nuclease to limit PLE activity. ICP1 is the first phage demonstrated to encode a functional CRISPR-Cas system, and such an unexpected evolutionary innovation highlights the significance of competition between a helper phage and its satellite in nature. The increasing prevalence of CRISPR-Cas⁺ ICP1 isolates in Bangladesh underscores the constant pressure ICP1 faces against PLEs in its *V. cholerae* host. Intriguingly, even though the ICP1 CRISPR-Cas system limits PLE replication, transmission and ICP1 interference, PLEs are still increasing in frequency in epidemic *V. cholerae* isolates, suggesting that although ICP1 has acquired a formidable weapon against PLEs, PLEs may still confer a fitness advantage to *V. cholerae*.

Consequences of Phage Satellites in Human Health and Disease

The first characterized SaPI demonstrated that phages spread the gene for toxic shock syndrome toxin to naïve *S. aureus* cells through mobilization of a genetic island that has come to be appreciated as a phage satellite. The majority of *S. aureus* virulence factors are found on SaPIs, including toxic shock toxin and other superantigens, in leftward and rightward accessory regions with their own regulatory regions (Fig. 1). Accessory genes can make the host bacterium more pathogenic by encoding for toxin production (such as the *tst* gene or enterotoxins EntQ and EntK). Other accessory genes can confer antibiotic resistance to aminoglycosides and Fosfomycin, or encode for multidrug transporters. Of other relevance in host-pathogen interactions are genes involved in iron transport and biofilm formation that have been found on SaPIs, which are important for successful colonization of human hosts and evasion of the immune response. In addition to SaPI encoded virulence factors, these phage satellites have been shown to transfer portions of the host chromosome between bacteria through generalized transduction. This can greatly accelerate the speed at which bacteria evolve within their human hosts.

With the rise of antibiotic resistance among human pathogens, there is renewed interest in phage therapy as a way of combating antibiotic resistant infections. Current interest in phage therapy has focused on lytic phages, which do not integrate into the host chromosome. Lytic phages are often considered safe because they are not likely to transduce genetic material from host to host through specialized transduction. This simplified view of the potential impact of lytic phages on promoting genomic flux is quite misleading. As highlighted in this article, lytic phages are capable of inducing and horizontally transferring mobile genetic elements. Although all known SaPIs and PICIs (for which helper phages have been identified) have evolved to parasitize temperate phages, PLEs in *V. cholerae* are mobilized by ICP1, a lytic myovirus. The dominance of temperate phages in mediating PICI mobilization is also likely an artefact of the ease at which temperate phages can be identified (*i.e.*, their genomes are present in single colony isolates) and not likely a biological constraint that prohibits MGEs from parasitizing lytic phages. When the lytic phage ICP1 infects a PLE containing *V. cholerae* cell, only PLE transducing particles are produced when the cell lyses, and PLEs can transduce to naïve *V. cholerae* transforming phage sensitive hosts to dead ends for future ICP1 attack. Phage therapy employing

ICP1 has been suggested as a viable means to limit cholera spread in endemic regions, though mechanistic insights demand caution as phage resistance and perhaps other beneficial phenotypes are likely to become widespread. The notion that employing CRISPR-Cas⁺ ICP1 to overcome PLE-mediated resistance in *V. cholerae* is likewise oversimplified, as CRISPR targeting of PLE can permit ICP1 replication while still allowing for some PLE transmission.

Conclusion

PICIs are widespread within both Gram-positive and Gram-negative bacterial hosts, playing a key role in host virulence and phage resistance. These satellite phages do not encode their own structural and lysis genes, but rather parasitize a helper phage for excision, induction, and packaging, and in the case of PLEs, also for genome replication. The reliance on a helper phage allows the satellite to reduce its genome size, suggesting that perhaps phage satellites evolve from defunct temperate phages. However, it remains possible that diverse MGEs evolve *de novo* to parasitize specific phages to promote their own dissemination. PICIs are functionally analogous to the so-called integrative plasmid P4, which acts as a phage satellite of the inducing phage P2 infecting *E. coli*. This strategy of genome loss and parasitism of a helper phage has evolved independently multiple times, and these phage satellites are likely present in diverse bacterial hosts that have not yet been characterized. Once more is known about how PLEs are induced and packaged, they may be included within the larger umbrella of PICIs. PLEs are not induced by SOS induction of their cognate helper phage, and it is unknown whether PLE encodes virulence factors.

With the discovery of CRISPR-Cas and the advent of a public health crisis caused by antibiotic resistant bacterial infections, there has been renewed interest in phage, as both tools for biotechnology and in treatment of bacterial infections through phage therapy. There are many unknowns in how bacteria respond to the constant pressure of phage predation, and even less known about how competing phage satellites battle for a bacterial host. The prevalence of PICIs indicates that phage infection can lead to unexpected outcomes, such as the spread of virulence, anti-phage and antibiotic resistance genes. With the increase in the number of sequenced genomes and improved bioinformatic pipelines, we will likely continue to find new genetic elements that parasitize phage through induction and co-option of replication and packaging materials. Discovery and investigation of PICIs and PICI-like elements will reveal how bacterial evolution can proceed at a pace unexpected by mutation alone – the acquisition of mobile genetic elements can dramatically re-shape the phenotype of their host in one generation. One consequence of acquiring a PICI is restriction of helper phage production when infecting that host. PICI interference can select for helper phage mutants or, in the case of ICP1, selection of phage that have acquired CRISPR-Cas to overcome PLE. This evolutionary arms-race between phage satellites and their inducing phage can yield unexpected outcomes, urging caution before applying the use of phage in treating bacterial infections.

Further Reading

- Donderis, J., Bowring, J., Maiques, E., *et al.*, 2017. Convergent evolution involving dimeric and trimeric dUTPases in pathogenicity island mobilization. *PLOS Pathogens* 13 (9), e1006581. doi:10.1371/journal.ppat.1006581.
- Fillol-Salom, A., Martínez-Rubio, R., Abdulrahman, R.F., *et al.*, 2018. Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *The ISME Journal* 12, 2114–2128. doi:10.1038/s41396-018-0156-3.
- Frígols, B., Quiles-Puchalt, N., Mir-Sanchis, I., *et al.*, 2015. Virus satellites drive viral evolution and ecology. *PLOS Genetics* 11 (10), e1005609. doi:10.1371/journal.pgen.1005609.
- Lindqvist, B.H., Deho, G., Calendar, R., 1993. Mechanisms of genome propagation and helper exploitation by satellite phage P4. *Microbiological Reviews* 57 (3), <https://mmbr.asm.org/content/57/3/683.long>.
- Lindsay, J.A., Ruzin, A., Ross, H.F., Kurepina, N., Novick, R.P., 1998. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Molecular Microbiology* 29 (2), 527–543. <http://www.ncbi.nlm.nih.gov/pubmed/9720870>.
- Martínez-Rubio, R., Quiles-Puchalt, N., Martí, M., *et al.*, 2017. Phage-inducible islands in the gram-positive cocci. *The ISME Journal* 11 (4), 1029–1042. <https://www.nature.com/articles/ismej2016163>.
- McKitterick, A.C., Seed, K.D., 2018. Anti-phage islands force their target phage to directly mediate island excision and spread. *Nature Communications* 9. doi:10.1038/s41467-018-04786-5.
- Novick, R.P., Ram, G., 2017. Staphylococcal pathogenicity islands – Movers and shakers in the genomic firmament. *Current Opinion in Microbiology* 38, 197–204. doi:10.1016/j.mib.2017.08.001.
- O'Hara, B.J., Barth, Z.K., McKitterick, A.C., *et al.*, 2017. A highly specific phage defense system is a conserved feature of the *Vibrio cholerae* mobilome. *PLOS Genetics* 13 (6), e1006838. doi:10.1371/journal.pgen.1006838.
- Penadés, J.R., Christie, G.E., 2015. The phage-inducible chromosomal islands: A family of highly evolved molecular parasites. *Annual Review of Virology* 2, 181–201. doi:10.1146/annurev-virology-031413-085446.
- Ram, G., Chen, J., Kumar, K., *et al.*, 2012. Staphylococcal pathogenicity island interference with helper phage reproduction is a paradigm of molecular parasitism. *Proceedings of the National Academy of Sciences of the United States of America* 109 (40), 16300–16305. doi:10.1073/pnas.1204615109.
- Ruzin, A., Lindsay, J., Novick, R.P., 2001. Molecular genetics of SaP1 – A mobile pathogenicity island in *Staphylococcus aureus*. *Molecular Microbiology* 41 (2), 365–377. doi:10.1046/j.1365-2958.2001.02488.x.
- Tallent, S.M., Langston, T.B., Moran, R.G., Christie, G.E., 2007. Transducing particles of *Staphylococcus aureus* pathogenicity island SaP1 are comprised of helper phage-encoded proteins. *Journal of Bacteriology* 189 (20), 7520–7524. doi:10.1128/JB.00738-07.
- Tormo-Más, M.Á., Mir, I., Shrestha, A., *et al.*, 2010. Moonlighting bacteriophage proteins derepress staphylococcal pathogenicity islands. *Nature* 465. doi:10.1038/nature09065.
- Úbeda, C., Maiques, E., Barry, P., *et al.*, 2008. SaPI mutations affecting replication and transfer and enabling autonomous replication in the absence of helper phage. *Molecular Microbiology* 67 (3), 493–503. doi:10.1111/j.1365-2958.2007.06027.x.

Portal Vertex

Peng Jing and Mauricio Cortes Jr., Department of Chemistry, College of Arts and Sciences, Fort Wayne, IN, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

ATPase ATPase is an enzyme that catalyze the hydrolysis of the phosphate bond in adenosine triphosphate (ATP) to form adenosine diphosphate (ADP). The free energy harnessed from this reaction can be used to drive other cellular reactions.

Brownian motor The Brownian motor is a device that can do mechanical or electrical work at a nanometer scale. In the device, externally sourced energy is needed to generate fluctuating anisotropic energetic potentials. The anisotropic potentials provide forces that would bias the motion of a particle in one direction. Translocation of the particle would be a consequence of a unidirectional Brownian motion.

Capsid A capsid is the protein shell of a virus. It consists of several oligomeric structural subunits made of protein called protomers.

Coat protein A predominant protein in the capsid that can form tiny channels in the major capsid proteins so that only ions or water can pass through.

Concatemeric DNA The concatemeric DNA is a multimeric head-to-tail polymer of viral DNA that contains multiple copies of an entire genome linked in series. It serves as the substrate for the production of the relevant viral genome in a bacteriophage terminase during the DNA assembly.

Portal protein The portal protein is a dodecameric protein, located at the portal vertex. It serves as a conduit through which dsDNA is translocated during DNA packaging in tailed bacteriophages. In Phi29 bacteriophage, the portal protein is also named as connector.

Prohead A prohead, or procapsid, is an immature viral capsid structure formed in the early stages of self-assembly of some bacteriophages including tailed bacteriophages.

Scaffolding protein Scaffolding protein is the protein that directly reacts with the portal protein to help to form the empty prohead along with other proteins. After the formation of the prohead, this protein usually is expelled, leaving space for genomic DNA.

Terminase Terminase is a viral enzyme that usually consists of two subunits with different sizes. The small subunit recognizes viral DNA by binding a specific/nonspecific site on the concatemer. The large subunit, docking on the portal vertex, contains the ATPase activity that powers translocation of the viral genome into the prohead and the endonuclease activity that cuts concatemeric DNA to a single genome length.

The Donnan's effect The effect (also known as the Donnan law, Donnan equilibrium, or Gibbs–Donnan equilibrium) is used to describe uneven distribution of charged particles near a semi-permeable membrane. The usual cause is the presence of a different charged substance, e.g., DNA, that is unable to pass through the membrane and thus creates uneven electrical potentials across the membrane.

The head-tail connector In a mature virion particle, the head-tail connector usually refers to a protein complex composed of the portal protein, the adapter protein and the stopper protein, which connects the filled head to the tail.

The Tail (The tail protein complex) The bacteriophage tail is a molecular machine used during infection to recognize the host cell and ensure efficient genome delivery into the cytoplasm.

Introduction

The *Caudovirales*, an order of viruses, are known as the double-stranded tailed bacteriophages. Based on their tail morphology, they are divided into three families known as *Myoviridae* which have long contractile tails (e.g., phages T4 and P1), *Siphoviridae* which have long noncontractile tails (e.g., λ and SPP1), and *Podoviridae* which have short noncontractile tails (e.g., Phi29 and P22).

A major strategy that bacteriophages use in order to protect their genomic chromosomes from the environment is to pre-assemble icosahedral proheads (procapsids). By means of a powerful DNA packaging motor, viral genomes are subsequently pumped into the preassembled icosahedral heads through a viral portal protein located at a unique axis of 5-fold symmetry known as the portal vertex. The portal provides a unique site through which the viral genome enters and exits the prohead. Because both tailed bacteriophages and certain human pathogens in the herpesvirus family share the similar structural features, their DNA packaging process is assumed to originate from an evolutionarily conserved mechanism. Owing to the ease of operation in research, the tailed bacteriophage is typically used as a model system for studying this viral genome encapsidation mechanism.

In the past 50 years, the assembly mechanisms of several dsDNA bacteriophages have been investigated in depth where the experimental results have shown that the portal vertex is formed by a structurally conserved dodecameric portal protein which plays key roles in the viral chromosome packaging. Fig. 1 shows a typical, simplified viral assembly pathway where the portal protein is directly involved in the following major stages in a tailed bacteriophage:

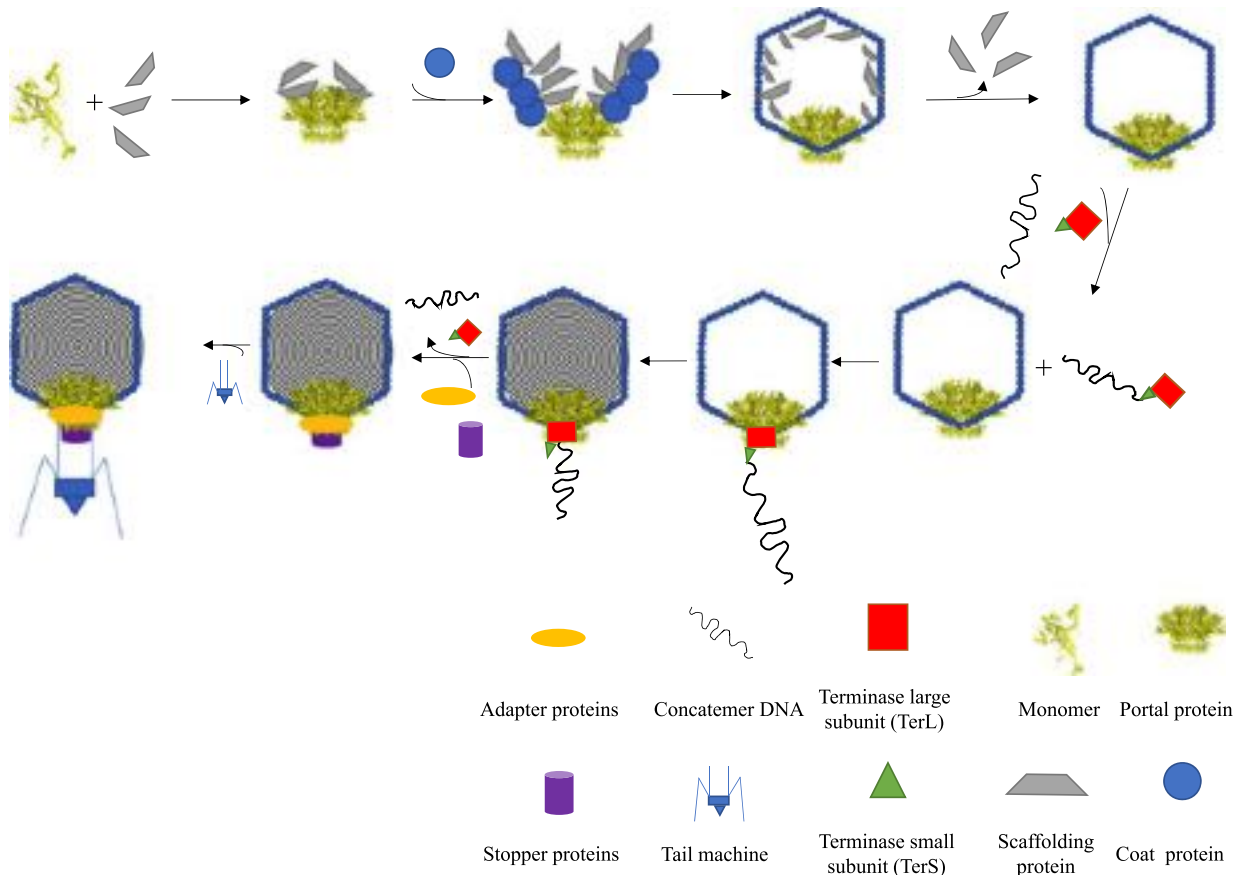


Fig. 1 A Typical and Simplified Pathway for the Viral Assembly of a Mature Virion Particle.

Formation of a Prohead

In order to start the DNA packaging, an icosahedral prohead needs to be preassembled. Aided by scaffolding proteins, the monomer of the portal protein may polymerize and form a dodecameric portal ring. The portal ring may work as a reaction center where both coat proteins and scaffolding proteins associate to form a preassembled prohead. When the assembly is finished, the scaffolding proteins trapped inside the prohead may diffuse out or be degraded and will finally not be present in a completed virion.

DNA Packaging into the Prohead

In the majority of the tailed phages, a protein complex, terminase, is used to begin DNA packaging. The protein complex usually consists of two subunits: the large subunit (TerL) and small subunit (TerS). In the first step, TerS recognizes a concatemeric DNA substrate in which many copies of the viral genome are linked end-to-end. After the formation of the DNA-TerS protein complex, the TerL subsequently docks on the portal protein of a prohead. Under the action of the TerL ATPase, the concatemeric DNA is packaged into the prohead. When a genome length of DNA is completely encapsidated, the portal protein may signal to the TerL and activate the inherent endonuclease to cut the concatemeric DNA substrate at a specific site (e.g., λ) or nonspecific sites (e.g., T4, P22, SPP1) to stop the packaging process. The package mode with non-specific cleavages is named as the headful packaging.

Some tailed bacteriophages use different packaging modes. For example, in Phi29, the ATPase known as GP16, is equivalent to the ATPase on the TerL in other phages, but GP16 does not have endonuclease activity. The portal protein in the phage, however, is bound to a packaging RNA (pRNA), which can form a pentameric ring on the portal protein. The pRNA provides a binding site for the ATPase and is therefore considered to be evolutionarily associated with a part of the TerL. The DNA substrate for the ATPase contains one unit of the viral genome covalently bound by a protein (GP3) at both ends. The ATPase will stop the packaging process when the full genome is completely packaged into the prohead.

Virion Maturation

As the prohead is filled with DNA, the icosahedral lattice may expand to mature the capsid shell. Under the assistance of the portal protein, the packaged DNA is spooling-like and highly ordered. As the terminase is dismantled, the portal protein may close its

channel. Adapter proteins, stopper proteins, as well as the tail are subsequently assembled at the portal vertex, completing the maturation of a virion particle.

The Structural Features of the Portal Proteins

The subunits of the portal proteins can vary in molecular mass and amino acid sequence among the tailed bacteriophages shown in [Table 1](#). However, they share highly conserved structural features in the tailed bacteriophages, such as that the channel of the portal protein in the bacteriophages is always composed of 12 subunits. Whereas the portal rings may consist of 11–14 subunits (e.g., 12 mers and 13 mers can be found in the Phi29 and SPP1 portal proteins) when the portal proteins are purified and self-assembled in vitro. The inconsistency in the oligomerization state can also be found in the pRNA in Phage Phi29 where it is observed to have five subunits in the virion particles by cryogenic electron microscopy (Cryo-EM), while the in vitro assembly experiments showed that the pRNA is a hexameric RNA complex. The results indicate that the in vivo assembly of the ring structure in the portal proteins may need assistance from other components such as the scaffolding proteins ([Fig. 1](#)).

In addition to the 12-fold symmetry, the central channel of the portal proteins has a minimum 2.8 nm diameter at its narrowest constriction so that dsDNA has sufficient space to pass through. The channel in the portal looks similar to a truncated cone ([Fig. 2](#)) where the diameter of the channel for DNA entry is always smaller than the channel end for DNA exit in the packaging process. The amino acid residues in the interior wall of the channel are predominately negatively charged which may generate strong electrostatic repulsion against the DNA.

The subunit in the portal proteins in the tailed bacteriophages have similar core organization in its appearance, which can be divided into four domains ([Fig. 3](#)): the clip, stem, wing, and crown domains. A brief summary of the main functions of the conserved structural motifs is provided in [Table 2](#).

The clip domain: The domain consists of an α/β fold and may interact with adjacent portal subunits by hydrophobic interactions to stabilize the oligomerization state of the portal protein. The domain is exposed to the exterior of the capsid providing an interface when in contact with the coat proteins, provides the binding site for components in the DNA packaging motor, e.g., TerL, and also provides a site where phage tail attachment occurs.

The stem domain: This domain is sandwiched by two domains known as the clip and wing, which is featured with two well-conserved α -helices. The central channel in the dodecamer is surrounded by the 24 stem helices, which may impart structural plasticity and induce conformational changes during the DNA packaging. A conserved structure known as the channel loops is present in the stem domain. The channel loops have been shown to generate a constriction in the channel that prevents leakage of genomic DNA near the end of DNA packaging.

The wing domain: The domain is internal to the prohead and is above the stem domain. It is a globular domain whose size differs among tailed bacteriophages. The hydrophobic amino acid residues such as tryptophan, found in the domain can form a conserved hydrophobic belt around the exterior surface of the portal protein. This region is assumed to interact with scaffolding proteins to form the ring structure and initiate the assembly of the prohead. Owing to the presence of the hydrophobic belt, it is noted that the Phi29 portal protein can be readily inserted into a bilayer lipid membrane for channel property measurements and maybe potentially used for nanopore sensing technology.

The crown domain: The crown domain in the portal proteins of the tailed bacteriophages is located at the C-terminal in the primary structure of the subunit. The crown domain in the portal proteins of phages SPP1, T4 and P22 orients towards the center of the capsid, which is directly in contact with the packaged DNA inside the capsid. The structure of the crown domains has some variations among certain tailed bacteriophages. Due to the flexibility, the crown domain of the portal proteins in some phages, e.g., Phi29, cannot be directly visualized in crystal structures or in Cryo-EM images. On the other hand, in the portal protein of P22, a 20 nm left-handed coiled-coil helical barrel is found to extend from its crown domain. Site-directed mutation studies have shown the barrel may even be involved in the organization of the spool-like DNA structure inside the capsid.

Table 1 A Comparison of the Primary Structures of Several Portal Proteins in the Tailed Bacteriophages

Phage Family	The Virus Name	The number of Amino Acids in a Protomer	The I.D. in the PDB Database
Podoviridae	Phi29	309	1FOU, 1JNB, 1IJG, 1H5W
	P68	327	6Q3G
	T7	490	6QWP, 6QX5, 6QXM, 3J4A
	P22	725	3LJ5, 4V4K, 5JJ3, 5JJ1
Myoviridae	T4	455	3JA7
Siphoviridae	SPP1	503	2JES
	G20C	426	4ZJN, 5NGD, 6IBG

Note: Only the portal proteins whose 3-D structures are available in the RCSB Protein Data Bank (PDB) are included in the table.

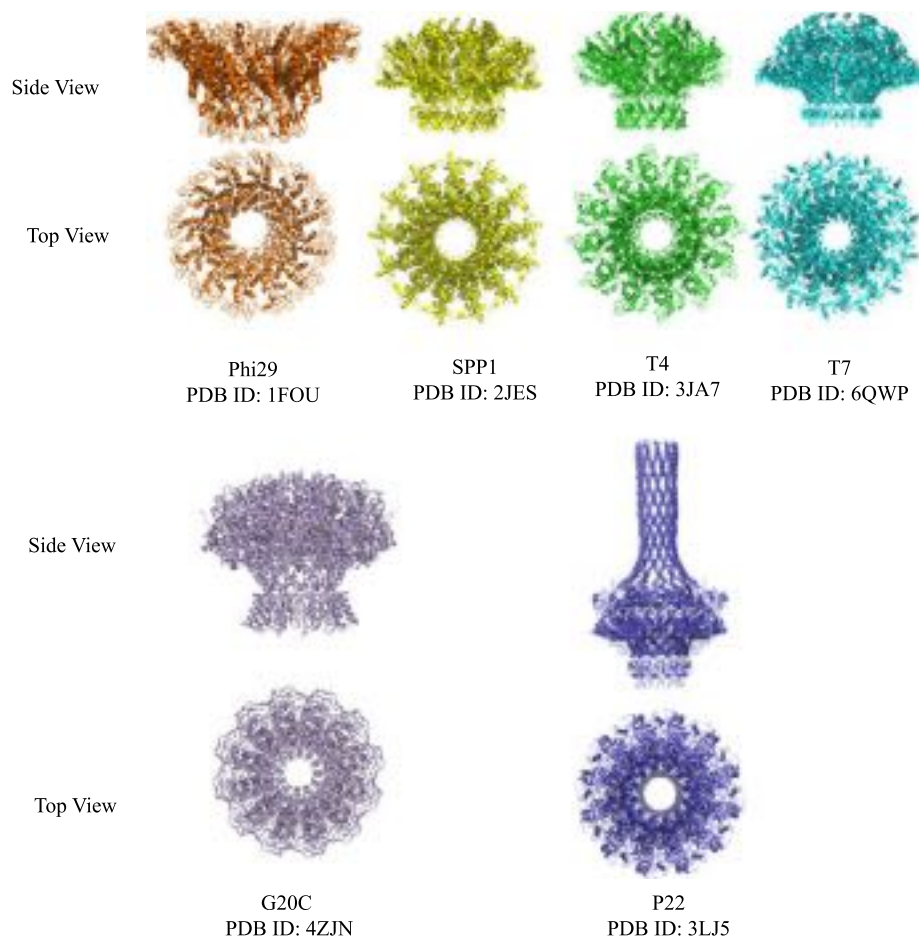


Fig. 2 The Structure of the Portal Proteins Available from the PDB Database.

Functions of the Portal Proteins

The Viral Portal Protein as a Nucleator in the Assembly of the Prohead With Proper Morphology

The prohead assembly occurs at the earliest stage in the DNA assembly of the tailed bacteriophages. Three major proteins are involved in the process: the portal protein, the scaffolding, and the coat proteins. The portal protein functions as a nucleator and the scaffolding proteins may work as assembly chaperone to bind the subunit of the portal protein to assist the oligomerization of the subunits forming the ring structure. The conserved structure, a hydrophobic belt encircling the portal ring at the wing domain, may provide a site where scaffolding proteins bind to the portal protein. After the formation of the portal protein-scaffolding protein complexes, coat proteins then copolymerize with the scaffolding proteins. As copolymerization proceeds, the shell of the prohead grows around the portal ring. Relevant experiments show that more portal rings results in higher yield as well as the rate of prohead assembly. In addition, the portal ring was also shown to be capable of correcting the morphology of viral particles even in the presence of a high ratio of scaffolding protein vs coat proteins.

When the copolymerization ends, the scaffolding proteins may diffuse out of the prohead or may be degraded and finally exit out of the prohead making room for incoming genomic DNA. During the process, the portal protein may act as an exit for scaffolding proteins or their degraded products. The above mechanism can be used to explain why only one portal ring is found at a unique portal vertex in the capsid of the tailed bacteriophages.

Portal Proteins Provide Binding Sites for Terminases

The clip domain is a conserved structure, protruding out of the capsid in the exterior and is the site where the large subunit of the terminase protein (TerL) can bind. For example, in phage T4 the clip domain of the T4 portal binds to the C-terminus of the TerL. After the terminase is bound, genomic DNA will translocate into the prohead in an ATP-dependent process driven by the TerL ATPase. Some tailed bacteriophages differ in the binding process, for example, in Phi29 the clip domain on the portal protein

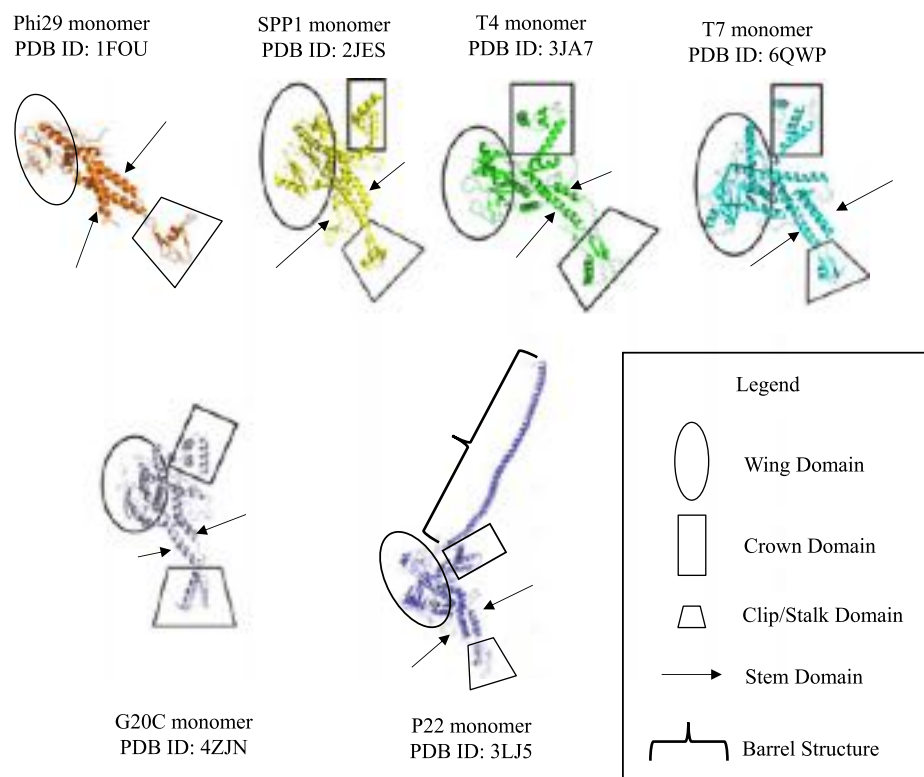


Fig. 3 The Conserved Structural Motifs in the Monomers of the Portal Proteins Available at the PDB Database.

provides a binding site for a pentameric pRNA. The ATPase known as GP16 (equivalent to the TerL in other bacteriophages) docks on the pRNA and self-assembles a ring structure around the pRNA to package the DNA into the prohead.

Portal Proteins Bound With Adapter Proteins Facilitates Tail Attachment

At the end of DNA packaging, after terminase protein dissociates from the portal protein, the clip domain of the portal protein would be converted to be an interface where the adapter proteins dock and subsequently the stopper proteins are attached in order to plug the channel to prevent the genomic DNA from accidental leakage.

The protein complex, composed of three proteins (the dodecameric portal protein, adapter proteins and stopper proteins), forms a knob-like structure at the portal vertex, which is used to connect prohead and the viral tail. Because of this function, the protein complex is also named as the head-tail connector which provides a site for attachment of the tail. It should be noted that, in the literatures about Phi29, the term connector only refers to its portal protein, also known as GP10.

Portal Proteins Help to Spool Condensed Genomic DNA in the Capsid to Form a Highly Ordered Structure

The packaged DNA in the capsid can be condensed to a highly dense concentration, about 500 mg/mL, and images from Cryo-EM experiments have shown that the chromosomal DNA in the capsids are highly orderly packed like liquid crystals. The highly ordered structure would prevent the chromosomal DNA from the entangling between the DNA segments, and subsequent DNA ejection would benefit from it.

The arrangement of the trapped DNA, however, varies among different tailed bacteriophages. For example, T4 DNA is arrayed longitudinally to the head-tail axis and both packed ends are located in the portal protein, whereas in T7 the packed DNA is spooled horizontally around the head-tail axis of the portal protein. The formation of the highly ordered structures of the packed chromosomal DNA in the capsid is assumed to be associated with the crown domain in the portal protein. However, how the domain is involved to spool the DNA into different structures is still unknown.

In P22, the portal protein is unique as both the crown domain and its extension, a coiled-coil alpha helical barrel structure, can be visualized clearly in the protein 3-D structure. The barrel is assumed to interact with genomic DNA and help the genome spool onto the interior surface of the capsid during genome packaging, working like a rotary sprinkler. To facilitate DNA ejection, the domain can also provide a docking site for ejection proteins and assure DNA ejection without tangling. In the phages that do not have the barrel structure, ejection proteins can assemble to a proteinaceous core bound onto the portal crown domain. The

Table 2 The Main Functions of the Conserved Structures in the Portal Proteins Among the Tailed Bacteriophages

<i>Highly Conserved Structural Motifs</i>	<i>The Functions of the Structural Motifs</i>	<i>The Portal Proteins Studied</i>
The Crown Domain as well as its extension (the barrel structure)	Sensing the pressure of chromosome in the capsid to terminate the DNA packaging.	P22 Portal
	The barrel structure: Helps the DNA form spool-like structure. May hold the trailing end of the packaged genome to keep the DNA primed for ejection. May serve as a packaging sensor for the pressure, the length, and the density of the packaged DNA, and as a valve to retain the DNA in the capsid at the end of the DNA Packaging.	
The Wing Domain	May help to spool the DNA inside the capsid.	T4 Portal
	The hydrophobic residues encircling the wing domain are the most likely region that interact with the scaffolding protein to form a dodecamer ring structure and provide binding region for coat proteins.	T4, Phi29, P22 portal proteins
	Helps stabilize the interactions of the adjacent subunits in the portal ring structure.	Phi29 and P22 portal proteins
The Stem Domain	The exterior in the domain: May bind the terminase for DNA packaging in the prohead and the adapter proteins in the mature virion for tail assembly.	P22, Lambda, T3, SPP1, and T4 ortal proteins.
	The channel in the domain help to grip and release of DNA to assist DNA packaging.	T4 portal protein
	The tunnel loops help to translocate DNA. The motion of the helix $\alpha 5$ on the loops may provide a path for signal and force transmission between the loops, DNA and the viral ATPase.	SPP1 portal protein
	The channel loops in the channel play a critical role in the retention of the DNA under a high pressure.	Phi29 Portal
The Clip Doman	May bind to TerL to assist DNA packaging and play a role in the cross talk between the portal and the terminase.	SPP1, P22 Portal proteins
	May bind to coat proteins.	P22 portal
	The negatively charged amino acid residues at the channel entrance at the clip domains can keep translocated DNA centered in the channel.	SPP1, HK97, Phi29, and T4 Portal proteins
	Interlock the adjacent subunits by hydrophobic interactions to impart additional stability to the oligomerization states.	Phi29 portal P22 portal proteins
	Provides a site for the docking of pRNA, that is subsequently bound to the ATPase.	Phi29 portal protein

resultant tunnel created by the stacked proteins, similar to the barrel structure in P22, is assumed to spool the DNA in a highly ordered structure inside the capsid.

Portal Proteins as a Check-Valve Preventing DNA From Slipping Out

A DNA packaging motor transports the viral DNA into the procapsid against a pressure of up to 20–60 atm, and thus a check-valve mechanism is needed to prevent the genomic DNA from slipping out of the prohead. Using the molecular dynamics simulation for the Phi29 portal protein as well as the relevant in vitro viral assembly experiments, three positively charged residues (K234, K235, and R237) on the structurally conserved channel loop are shown to play a crucial role in the check-valve mechanism. As the pressure is increasingly built up inside the prohead, the positively charged amino acids may intensively interact with genomic DNA to prevent it from slipping out. Based upon the structural observation, it is noted that another two rings of positively charged lysine residues (K200 and K209) located on the clip domain in the channel of the Phi29 portal protein are also assumed to be involved in the check-valve mechanism.

The existence of the check-valve mechanism was evidenced in an in vitro T4 DNA packaging experiment measured by optical tweezer. The experiments have found that the terminase proteins did not grip the portal protein strongly when ATP was hydrolyzed into ADP. Under the situation, the T4 portal tunnel loops are assumed to strongly “grip” the DNA to prevent it from slipping out. Similarly, the lysine residues (K331 and K342) in the SPP1 portal protein are also proposed to have similar gripping functions during the DNA packaging.

Portal Proteins as a Sensor to Control the DNA Packaging

In bacteriophages T4, P22, T1, P1, and SPP1, all follow the headful packaging mechanism in which the DNA packaging proceeds until a threshold amount of DNA is reached inside the viral capsid. When the prohead is filled in full, a high internal pressure is maintained. As a result, the high pressure may induce the conformational changes at the clip domain in the portal protein where TerL is docked. The TerL endonuclease is subsequently activated to cleave the DNA concatemer and generate a segment of DNA that contains one unit of the phage genome. In the process, the portal protein serves as a packaging sensor that may use the pressure in the capsid to regulate the maturation process of the capsid. The relevant mutation experiments on the phage P22 portal protein further suggest that the sensor could be located on the barrel of the portal protein.

In Phi29, near the end of DNA packaging, the portal protein may act as a pressure sensor, meaning it can change its conformation to detach the bound pentameric pRNA associated with ATPase. The portal protein is believed to use the conserved structure, the channel loops (residues 229–246), located in the channel interior as a clamp to retain the genomic DNA within the prohead before the tail components are attached. In an optical tweezer experiment on the Phi29 DNA packaging motor, the packaging sensor in the portal protein was assumed to sense the changes in the density and conformation of the packaged DNA at the late stage of DNA packaging. As a result, the activity of the ATPase would be allosterically regulated.

The Portal Proteins Actively Assists DNA Packaging Coupling With ATP Hydrolysis

The terminase-driven mechanism plays a key role at the initial stage of DNA packaging

When DNA enters an empty prohead or a prohead with a low pressure, terminase plays a central role in DNA packaging. A terminase-driven model is proposed to explain the process. The terminase is a hetero-oligomeric complex, composed of two subunits, TerS and TerL. TerS selectively binds the viral DNA concatemer from a pool that includes the host DNA. TerL uses its endonuclease to make a cut and binding itself on a portal vertex of the head to initiate DNA packaging.

The TerL is assumed to form a pentameric ring around the portal protein. The ATPase on each of the five subunits in the TerL may also have a domain that is responsible for DNA gripping. A section of genomic DNA is gripped in the center of the TerL ring when ATP is bound to all five subunits. ATP hydrolysis occurs at one subunit at a time. When the ATPase on one subunit is fired to hydrolyze ATP, the subunit alters its conformation to release DNA and then initiate motion that is transmitted to the adjacent ATP bound subunit, which is still gripping the DNA. The conformational changes of the ATP-hydrolyzing subunit results in an upward translocation motion of DNA at the adjacent subunit. The sequential interlaced ATP hydrolysis/binding on each of the TerL subunits may lead to unidirectional DNA translocation, which requires the activities of the individual subunits are highly coordinated. The terminase-driven model can be used to explain the 2.5 bp translocation per ATP measured for the Phi29 motor at low packaging force by optical tweezers.

The portal proteins actively assists DNA packaging at the late stage of DNA packaging

DNA packaging is an energetically unfavorable process. At the late stage of DNA packaging, a large amount of free energy is used to package DNA to overcome the high pressure in the prohead. In addition, energy is also consumed to assist the genomic DNA to form a highly ordered structure similar to a liquid crystal state in the capsid (an entropy-decreasing process). Images from Cryo-EM have shown that the portal proteins in all bacteriophages undergo conformational changes during the DNA packaging.

These conformational changes may have three possible origins:

- (1) They may be induced by the interior pressure in the capsid as discussed above;
- (2) They may be associated with the mechanical cycles of the ATPase during ATP hydrolysis;
- (3) The conformational changes may cyclically occur during packaging, and could be induced by the electrochemical potentials across the capsid shells.

On how the portal protein uses conformational changes to assist packaging, a handful of models have been proposed, which can be roughly categorized into two modes: portal directly-mediated packaging mode vs portal indirectly-mediated packaging mode.

The portal directly-mediated packaging mode

The ATPase in the terminase uses energy from ATP dephosphorylation to generate movement by mechanical cycles. The moving and pushing by the ATPase on the terminase may induce conformational changes in the portal proteins because TerL is directly docked on the portal protein. In the portal directly-mediated packaging mode, the resultant conformational changes are assumed to directly apply forces to the genomic DNA and push it into the procapsid.

Different models have alternate explanations to the conformational changes, thus the description about the DNA packaging process assisted by the portal protein varies among models. Several models categorized under the mode are highlighted below:

Rotation of the portal protein may apply forces to DNA

Because the portal protein has a 12-fold symmetry and it is located at a vertex with 5-fold symmetry, the model assumes that the asymmetry may facilitate the rotation of the portal protein at the viral vertex. In this model, the energy from ATP hydrolysis was assumed to rotate the portal protein which may screw the helical DNA into the capsid like a bolt moves on a nut. However, this model has been discarded currently due to recent experiments proving that the portal protein does not rotate during DNA packaging.

Lengthwise channel expansion and contraction in the portal protein may apply forces to DNA

The connector expansion and contraction model is based on the structural observation of the portal protein in Phi29. The portal protein is assumed to have plasticity and its channel may experience lengthwise expansion and contraction after the portal protein is slightly rotated by 12 degree when ATPase is activated. The movement in the model is described like a helical spring pumping the chromosomal DNA into prohead.

Sequential movement of tunnel loops may apply force to DNA

The model is proposed on the structure of the portal protein in SPP1. The SPP1 portal is found to have 12 long flexible loops that project into the central channel of the portal. These loops are a conserved structure among other portal proteins. They are assumed to tightly fit the DNA during the packaging. When the portal protein slightly rotates in coupling with ATP hydrolysis, the loops may undergo sequential movement, like an undulating belt. The DNA then moves sequentially along this belt and is packaged into the prohead.

In summary, the models following the portal directly-mediated packaging mode are proposed based on the structural observation of the portal proteins from the Cryo-EM or X-Ray Crystallography. However, the descriptions in the models about how the portal protein apply the forces to assist the DNA packaging as well as how the ATPase uses its mechanical cycles to induce the channel conformational changes have not been visualized experimentally.

The portal indirectly-mediated packaging mode

In this mode, the portal protein is proposed to work like an energy transducer. The energy from cyclic and random conformational changes in the portal is converted to a different energy form that is subsequently used to pack genomic DNA into the capsid. Two physical models that use the indirectly-mediated packaging mode are highlighted here:

DNA translocation driven by the energy from the DNA deformation in the portal protein

This model, named as the scrunchworm model, is based upon molecular dynamic simulation on the portal structures of P22, Phi29, and T4. The simulation assumes that sufficiently large conformational changes in the portal proteins can cause the DNA in the portal channel to undergo conformational changes between stretched and relaxed states. When DNA is in its extended form, the portal protein needs to grip the DNA at the bottom of the channel through electrostatic interactions and release the rest of the DNA at the top. When DNA is packaged in its shortened form, the top of the channel may grip the DNA and the bottom grip would open. The portal protein cyclically changes its conformations and applies force to store the energy of the DNA deformation. Like a spring, the DNA would be compressed and released back and forth, propelling the DNA into the prohead. Apparently, the process simulated in silico requires that the ATPase alternately changes the conformations of the portal protein in a well-controlled manner and this hypothesis needs to be further studied experimentally.

DNA translocation driven by a brownian motor inherent in the portal protein

In this model, the truncated cone structure of the portal protein ensures that the electrostatic interactions between the DNA and the negatively charged residues in the interior of the channel along the direction of DNA movement are asymmetrical. The ATPase continuously packages the DNA into the capsid and results in an increase in the electric potentials across the semipermeable capsid shell membrane called the Donnan's effect. This is the direct energy source for the Brownian motor where the electrochemical potentials can alternately change the channel conformation. Consequently, the portal protein works as a power rectifier, by which the asymmetrical and cyclic interactions between the portal channel and DNA would ratchet the DNA Brownian diffusion back to the channel bottom, and ensure the unidirectional Brownian motion towards the prohead.

The Brownian motor model is proposed based upon planar lipid bilayer membrane (BLM) experiments, in which the channel of the portal protein of Phi29 was shown to randomly and cyclically change its conformations under membrane potentials when the protein was embedded into a lipid membrane. Although the BLM results come from those measured by indirect experimental methods, they strongly suggest that the high probability of the existence of channel fluctuation under the membrane potentials. The Brownian motor model has successfully explained many experimental results from optical tweezers. For example, why the DNA packaging motor operates in a dwell-burst fashion and how it is affected at the late stage of DNA packaging. More importantly, the BLM experiments indicate that, besides the mechanical cycles of the ATPase in the ATP hydrolysis, the membrane potentials on the prohead shell may also play a role in inducing the conformational changes cyclically and alternately in the portal proteins.

Main Techniques Used to Study Portal Protein Structure and Function

Biochemical methods

Biochemical methods are traditional ensemble methods where a mutant component, e.g., the prohead containing a mutant portal protein, is usually designed and prepared using molecular biological methods such as site-directed mutagenesis. The mutant proheads are then mixed with other components for in vitro DNA packaging experiments. By means of enzymes such as DNase I and proteinase K, the size of the DNA packaged inside the capsid can be quantified. The conversion efficiency from filled prohead to phage can also be determined by mixing packaging solutions with bacteria for virus quantification. The methods are often used for investigating the functions of specific domains, e.g., the channel loops, in the portal proteins of the tailed bacteriophages.

Techniques for structural analysis

Two useful structural analysis methods are usually used in Structural Biology: X-Ray Crystallography and Cryo-EM.

In X-Ray Crystallography, the target protein needs to be dissolved in an aqueous environment. When the resultant sample solution reaches its supersaturated state, it may form a single-crystal form under precisely controlled conditions. In the process,

many factors may affect the growth of the protein crystal such as pH, temperature, ionic strength, and even gravity. Once properly developed, these crystals can be used in structural biology to study the molecular structure of the protein. Although X-ray crystallography can provide 3-D protein structure at atomic resolution, the process for protein crystallization may pose a great challenge to the study of portal proteins since many of them are known to be hydrophobic and may aggregate in the aqueous solutions. Another problem with the X-ray crystallography is that the structure of the portal proteins in its crystal form is not necessarily the same as the conformation during DNA packaging.

Cryo-EM can overcome some of the limitations found in X-ray crystallography. The method uses rapid freezing of solutions containing virions or virion components in which the intact structures of the biological macromolecule complexes can be preserved for electron imaging. The 3-D structure of a virus particle in its native conformation can be reconstructed by means of the combined images of virion particles using an image-processing software. The Cryo-EM can provide virus structures at sub-nanometer resolution and therefore is a very useful imaging tool to infer the information about how different motor components move and interact with each other.

However, Cryo-EM only provides snapshots representing the average of large ensembles of individual virion particles, thus upon freezing, the DNA packaging motors in different virion particles would be stalled in different stages of the enzymatic or mechanical cycle. The conformational heterogeneity would be an obstacle for Cryo-EM to directly visualize the conformational change of the target components as well as the relevant interactions at different stages, e.g., how the terminase proteins interact with the portal protein to induce its conformational changes during ATP hydrolysis.

The optical tweezers: a technique to observe DNA packaging in real time

Optical Tweezer is a technique that uses a focused laser beam to trap a micron-sized plastic microsphere in an optical microscope. In the single-molecule studies of viral DNA packaging, a single DNA molecule is tethered to the dielectric microsphere. When the tethered DNA is packaged into a capsid, the position of the trapped microsphere is also changed, which can be detected with great precision by imaging the light scattered by the microsphere. The light sensed by a position-sensitive photodetector can be used to provide sub-nanometer resolution of displacements and subpiconewton resolution of forces. The viral DNA packaging in phages, such as Phi29, T4 and lambda, have been studied in real time and the forces exerted by the motor on the DNA and the translocation dynamics can be measured. The measurement is performed at the single-molecule level, and therefore can provide the information that the ensemble methods cannot provide, for example the dwell-burst phenomena found in the DNA-packaging.

Molecular dynamics simulation

Molecular dynamics simulates the natural motion of molecular systems. The energy provided in a molecular dynamics procedure allows the atoms or molecules to interact for a definite period of time, providing a view of dynamic evolution of the molecular system e.g., the patterns, strength, and the behavior of the target components under various conditions. In the study of viral portal proteins, the methods were usually applied to the interactions between DNA and the amino acid residues on the interior surface of the channel in the portal proteins including Phi29, T4, and P22. It has been shown that, under the influence of sufficient and cyclic conformational changes in the channel of the portal proteins, the DNA can be unidirectionally transported into the capsid.

Planar bilayer membrane (BLM) technology

According to the Brownian motor model, the portal protein in the bacteriophage can fluctuate under the influence of the membrane potentials and high pressure in the capsid (the Donnan's effect). The resultant conformational change may play a critical role to ratchet the Brownian motion of DNA. Because the virion particles are too small (nanometer sized), the patch clamp cannot be used to measure the channel fluctuation under the in vivo potentials across the capsid shell. Therefore, BLM technology would be the only available experimental method that is used to indirectly visualizes the conformational changes on the portal protein under membrane potentials. In the method, a portal protein needs to be inserted into a bilayer lipid membrane. Under the influence of the Donnan's effect, the protein may fluctuate in response to the applied membrane potentials. The BLM technology can be potentially used to look for the domains in the portal protein that are involved in the sensing of the membrane potentials as well as in the channel fluctuation.

Concluding Remarks

The portal protein plays a core role in the packaging of the DNA of tailed bacteriophages. At the beginning stage, the portal protein is a nucleator to initiate the copolymerization of coat and scaffolding proteins to form an empty prohead with a unique portal vertex containing a dodecameric portal ring. The portal protein provides binding sites for terminase protein for packaging chromosomal DNA and for the adapter proteins for the installation of the tail. To prevent the DNA from tangling in the capsid, the portal protein may help DNA form a highly ordered structure. Because of the proximate contact with the DNA in the capsid, the portal protein may serve as a packaging sensor to participate the regulation of ATPase activity and transmit a signal to the terminase to dissociate from the capsid when DNA packaging is completed. As the pressure inside the shell increases, the portal can also serve as a clamp to prevent DNA from leakage near the end of DNA packaging. Because it has a flexible structure, the

portal protein may change its conformation cyclically and assist the DNA packaging in the portal directly-mediated packaging mode or the portal indirectly-mediated packaging mode.

The conformational changes in the portal proteins are assumed to play critical roles in the control, regulation and assistance of DNA packaging. How the conformational changes occur *in vivo* and what energy source is used to induce the conformational changes still lack direct experimental evidence. Therefore, the *in vivo* conformational changes are described differently in many models. In addition, the current experimental techniques are not perfect in providing the relevant experimental evidence. For example, the methods for the structural analysis, e.g., Cryo-EM, cannot provide details about how the terminase hydrolyze ATP in its mechanical cycles to induce the conformational changes in the portal proteins. The molecular dynamic simulations, in addition, cannot answer whether the sufficient conformational changes that cause the deformation of the DNA in a precisely controlled manner can still occur when the portal protein is embedded in the capsid shell composed of coat proteins. BLM technology is only an indirect method to measure the channel's conformational changes in a non-native environment. Optical tweezer measurements have never been used for the force measurement of DNA packaging when the viral assembly happens in a prohead containing a mutant portal protein. The integration of the available experimental techniques therefore is necessary in exploring how the portal protein changes its conformation in the tailed bacteriophages under influence of the pressure, ATPase and the membrane potentials as well as the roles that the resultant conformational changes play.

It has been known that mammalian dsDNA viruses, such as human herpesviruses, package their genome through their portal proteins using the similar mechanism that the tailed bacteriophages use. The structure of the portal proteins in human herpesviruses also share the common features with the bacteriophage portal proteins. The relevant study on the roles of the portal proteins in viral assembly in the tailed bacteriophages would provide very useful information for the preclinical studies that target the portal proteins in the human herpesviruses.

Further Reading

- Bayfield, O.W., Klimuk, E., Winkler, D.C., *et al.*, 2019. Cryo-EM structure and *in vitro* DNA packaging of a thermophilic virus with supersized T=7 capsids. *Proceedings of the National Academy of Sciences of the United States of America* 116, 3556–3561.
- Cuervo, A., Fabrega-Ferrer, M., Machon, C., *et al.*, 2019. Structures of T7 bacteriophage portal and tail suggest a viral DNA retention and ejection mechanism. *Nature Communication* 10, 3746.
- Dedeo, C.L., Cingolani, G., Teschke, C.M., 2019. Portal protein: The orchestrator of capsid assembly for the dsDNA tailed bacteriophages and herpesviruses. *Annual Review of Virology* 6, 141–160.
- Grimes, S., Ma, S., Gao, J., Atz, R., Jardine, P.J., 2011. Role of ϕ 29 connector channel loops in late-stage DNA packaging. *Journal of Molecular Biology* 410, 50–59.
- Hendrix, R.W., 1978. Symmetry mismatch and DNA packaging in large bacteriophages. *Proceedings of the National Academy of Sciences of the United States of America* 75, 4779–4783.
- Jing, P., Burris, B., Zhang, R., 2016. Forces from the portal govern the late-stage DNA transport in a viral DNA packaging nanomotor. *Biophysical Journal* 111, 162–177.
- Kornfeind, E.M., Visalli, R.J., 2018. Human herpesvirus portal proteins: Structure, function, and antiviral prospects. *Reviews in Medical Virology* 28, e1972.
- Lebedev, A.A., Krause, M.H., Isidro, A.L., *et al.*, 2007. Structural framework for DNA translocation via the viral portal protein. *The Embo Journal* 26, 1984–1994.
- Liu, S., Chistol, G., Hetherington, C.L., *et al.*, 2014. A viral packaging motor varies its DNA rotation and step size to preserve subunit coordination as the capsid fills. *Cell* 157, 702–713.
- Olia, A.S., Prevelige, P.E., Johnson, J.E., Cingolani, G., 2011. Three-dimensional structure of a viral genome-delivery portal vertex. *Nature Structural Molecular Biology* 18, 597–603.
- Prevelige, P.E., Cortines, J.R., 2018. Phage assembly and the special role of the portal protein. *Current Opinion in Virology* 31, 66–73.
- Sharp, K.A., Lu, X.-J., Cingolani, G., 2019. DNA conformational changes play a force-generating role during bacteriophage genome packaging. *Biophysical Journal* 116, 2172–2180.
- Simpson, A.A., Tao, Y., Leiman, P.G., *et al.*, 2000. Structure of the bacteriophage Φ 29 DNA packaging motor. *Nature* 408, 745–750.
- Smith, D.E., Tans, S.J., Smith, S.B., *et al.*, 2001. The bacteriophage ϕ 29 portal motor can package DNA against a large internal force. *Nature* 413, 748–752.
- Sun, L., Zhang, X., Gao, S., *et al.*, 2015. Cryo-EM structure of the bacteriophage T4 portal protein assembly at near-atomic resolution. *Nature Communication* 6, 7548.

Prohead, the Head Shell Pre-Cursor

Marc C Morais and Michael E Woodson, The University of Texas Medical Branch, Galveston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Capsid The protein shell that contains and protects the genetic material of a virus.

Capsomer A pentameric or hexameric sub-assembly unit of the capsid.

Icosahedron A regular polyhedron with 20 equilateral triangular faces. Each face consists of 3 equivalent smaller triangles such that an icosahedron is made from 60 identical asymmetric subunits.

Maturation The process by which a prohead transforms into its more stable mature form.

MCP The Major Capsid Protein, multiple copies of which are arranged as capsomers to make up the shell that protects the phage genome.

Quasi-equivalence The slight difference in bonding patterns and conformations assumed by different individual MCPs in the asymmetric units of icosahedra with T-numbers greater than one.

Triangulation number (T-number) The number of additional triangles each of the 60 icosahedral asymmetric subunits can be divided into. Thus, the total number of MCPs will be $T \times 60$.

Introduction

The need for viruses to convert the limited resources of their host cells into the greatest number of progeny in the shortest time represents an optimization problem. Billions of years of evolution have produced a surprisingly small set of general solutions to this problem in the form of ‘assembly pathways’, i.e., the sequences of reactions and products that result in infectious virions. The most successful pathway, in terms of numbers of infectious particles, is that of dsDNA bacteriophages. A notable feature of this pathway, shown in Fig. 1, is that it goes through a structurally distinct intermediate en route to the assembly of mature infectious virions. This extra ‘prohead’ step allows efficient construction of environmentally stable protein capsules, or capsids that contain and protect the viral genome. The prohead is formed when hundreds to thousands of copies of the protein building block of the capsid, the major capsid protein (MCP), self-assemble into a fully enclosed shell. After its completion, the prohead transforms into the next intermediate in the assembly pathway, the mature head. To successfully complete the self-assembly pathway, MCP has conflicting requirements; weak binding is initially needed such that MCPs can detach and re-attach as necessary to correct for errors that may occur during prohead self-assembly, yet strong binding interactions are ultimately required to construct a robust mature capsid that can survive the extracellular environment. This conflict is resolved by an MCP that has evolved to adopt distinct conformations appropriate for the two separate requirements. The prohead state coordinates the transformation between these two conformations so it only occurs after all MCPs have assembled into the correct geometry. Understanding how these complex molecular rearrangements are regulated and coordinated at the atomic level will inform antiviral drug development, engineering of gene therapy vectors, and rational design of nanomaterials and catalysts, among other applications.

Understanding of the viral lifecycle requires a consideration of the evolutionary pressures that influence it. Competition for resources between a virus, its host, and its rivals favors genetic efficiency. Hence viruses generally have small genomes that produce a limited number of multi-functional proteins that can rapidly assemble the structures needed to construct an infectious virus.

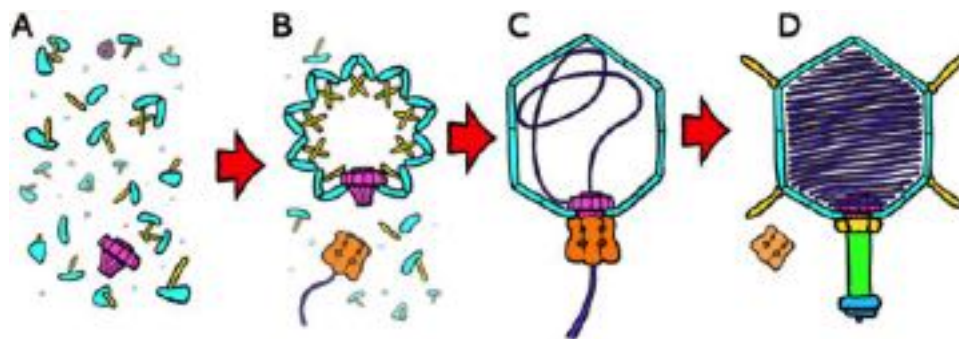


Fig. 1 Assembly Pathway of tailed dsDNA bacteriophages. A simplified assembly pathway for the tailed-phages is shown. In A, soluble capsid protein (blue wedges), scaffold protein (yellow rods) and portal complexes (pink cone) begin to associate. In B, these components have assembled into a closed shell, the prohead. In C, the scaffolding protein has decoupled from the capsid protein, which expands into the mature head. The packaging motor (orange shape) has attached to the portal and begun to feed DNA (purple coil) into the capsid. In D, the capsid is filled with DNA, the motor has detached, and the tail and accessory proteins have bound to the capsid, resulting in the mature, infection-competent virion.

Viral capsids are therefore made of many copies of one or a few types of proteins that bind together to produce large MDa-scale macromolecular assemblies. Viruses generally, and bacteriophage especially, have limited access to sophisticated methods of component segregation, including a means of controlling the extent and timing of component movements. Typically, the protein components that assemble the virion must all coexist in solution simultaneously and perform their various tasks in the proper sequence, without interfering with one another. Double-stranded DNA bacteriophages avoid interference by assembling different structures as separate functional units called the ‘head’ and the ‘tail’, which contain/protect the viral genome and deliver that genome into the host cell, respectively. Because of these ‘tail’ structures, dsDNA bacteriophages are often referred to as tailed-phages. The ‘head’ structure is thus a context-specific term for capsid and is not typically used for viruses that lack tails. In the tailed-phage assembly pathway an empty prohead, or procapsid, is assembled first. A phage-encoded molecular motor then assembles at a unique vertex of the prohead and actively pumps the phage’s genomic DNA into the head. In many phages, DNA packaging is believed to trigger a series of structural changes that transform the prohead to mature head. After packaging is complete, the motor disassembles, and then the tail and related structures replace the motor at the unique vertex to complete assembly of the infectious virion.

All tailed-phage heads are icosahedral in shape. A standard icosahedron has 20 equilateral triangular faces and twelve vertices where 5 triangular faces meet. Some capsids are prolate, meaning that they are elongated along one axis and 10 of the 20 triangular faces are elongated and no longer equilateral. Icosahedral shapes are common among many different types of viruses because icosahedra have the greatest symmetry of any regular polyhedron; of the shapes that can be built from identical subunits, icosahedra use the largest number of subunits (60). Since a large number of subunits are used to construct an icosahedral capsid, the size of individual MCP subunits necessary to encapsidate a given volume can be smaller. Hence, an icosahedral capsid built from 60 copies of the MCP results in a large ratio of internal volume to encoding sequence length. Despite the genetic economy provided by high symmetry, an icosahedral capsid built from 60 copies of the MCP is not large enough to enclose the full genome of dsDNA bacteriophages, which must also code for proteins for the tail, DNA replication, DNA packaging, and other genes. Rather than simply coding for a larger MCP, tailed phages have evolved to build larger icosahedral capsids by using more than 60 copies of the MCP. To accomplish this while still maintaining overall icosahedral symmetry, phages have evolved to have more than one copy of the MCP in each of the 60 identical asymmetric units (ASU) in the icosahedron. The number of MCP copies in each asymmetric unit is called the triangulation number of the icosahedron. An icosahedral capsid with only one copy of the MCP per ASU has a triangulation number of $T = 1$ and a total of 60 MCPs per capsid, whereas a $T = 7$ capsid has $7 \times 60 = 420$ MCPs in total. While each individual copy of the MCP in the ASU has the same amino acid sequence, their structures are only quasi-equivalent, that is, similar but non-identical. It is important to note that the increased volume-to-sequence ratio gained by adopting quasi-equivalence comes at a cost. Because the quasi-equivalent structures are slightly different, they are capable of interacting differently with their neighbors. Thus, while an MCP capable of binding together in different ways can produce larger icosahedra, it is also capable of producing an infinite variety of other shapes, very few of which will allow the virus to infect new hosts, and none of which are optimal. Further challenges arise when we consider that infectious virions must attach the tail to the capsid. As discussed above, icosahedra have 12 vertices, and in tailed-phage capsids 11 of these are occupied by pentamers of MCP while the 12th is a unique vertex occupied by a ‘portal’ complex which allows DNA to enter and exit the capsid. Viral particles that incorporate more than one portal vertex, or none at all, will have reduced or no infectivity.

Despite the simplistic description we have given of the phage assembly pathway, in practice a large number of sequential steps and intermediate states are involved. While the overall pathway is consistent among tailed phage species, differences in host environment, infection strategy, genome size, and other factors result in variations in individual states and sequences in assembly. Only the most stable intermediates from these have been isolated and characterized. Further, the specific intermediate used to represent the ‘prohead’ state in one phage may not be a direct analog of the ‘prohead’ state of another. The term refers collectively to states in which MCP monomers have been assembled and arranged to form a fully enclosed shell, but have not yet transformed into the characteristic, highly stable mature conformation. All studied prohead states share a few common characteristics, and the “variations” observed may in fact be present in other phages but too short-lived to have been characterized. Common features of known prohead states include (1) the presence of a “scaffolding” protein that will not be present in the mature virion, (2) a smaller, more spherical shape than the mature heads, (3) a rougher or more wrinkled surface than mature state, and (4) the ability to bind a packaging motor at the unique portal vertex. Because of the large morphological differences among individual phages and the prohead’s nature as an intermediate state, understanding it is best served by describing features in relation to the mature virion. Hence, these and subsequent features of proheads are described relative to the mature head.

The more angular icosahedral shape of the mature capsid results from the need for tight inter-MCP binding and a high volume-to-sequence ratio. Similarly, the different shape of the prohead is the result of constraints associated with a different set of requirements, i.e., self-assembly of a closed icosahedral shell and incorporation of a portal vertex that, by its very nature, is inconsistent with the otherwise icosahedral symmetry. The sets of requirements of thermodynamic stability in the mature head and kinetically accessible meta-stability in the prohead are conflicting, yet both must be satisfied by the same protein, the phage MCP. The capacity to resolve this conflict is provided by an MCP protein fold capable of adjusting its three-dimensional structure according to the requirements of both quasi-equivalence and the particular step in the phage’s assembly pathway. While significant variations of the MCP occur across all known tailed phages and herpesviruses, the canonical MCP fold, shown in [Fig. 2](#), is always recognizable despite very little sequence conservation. The fold was first described in the *E. coli* phage HK97, which, coincidentally, also best represents the essential structural core. Extra features and embellishments are present in all subsequently determined structures of phage MCPs, and in herpesviruses account for the majority of the MCP fold. To describe the general structure and function of the prohead state, it is necessary to refer to features of this conserved MCP structural core. Briefly, the HK97 protein

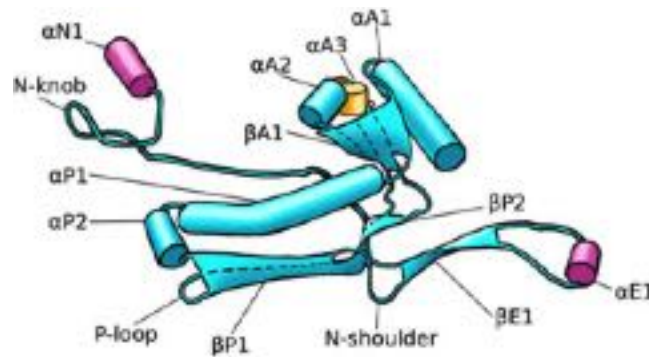


Fig. 2 The HK97-like fold of the MCP. This illustration of an idealized MCP shows the secondary structural elements and other features typical of all phage MCP that have the HK97 fold as a structural core. Blue features are present in all known structures, yellow features are present in some conformations but not others, and pink features are present in some phage species but not others. Features are named according to the type of secondary structure present if any, the domain in which they occur, and a number to differentiate similar elements. Number values are smaller for features that are present in more known structures. For example, the wedge-shaped A domain is composed of the five-stranded β sheet β B1, flanked by α A1 on the left and α A2 on the right. The position and conformation of α A2 varies more over different states and species than does α A1, while α A3 only exists in some states of some phages.

fold is divided into four domains: a wedge-shaped A-domain, a plank-shaped P-domain, an extended E-loop, and an N-terminal arm. All four domains change their structures as needed to meet the requirements of quasi-equivalence and maturational transformation. However, the A- and P-domains that make up the bulk of the shell of the capsid are less conformationally variable than the E-loop and N-arm which form bridges to bind neighboring MCP monomers together. Fig. 2 illustrates the HK97 fold and its conserved features and serves as a reference to the MCP features described below.

Major Capsid Protein Before Capsid Assembly

The structure of the prohead is a consequence of its ability to self-assemble from MCP in solution in the cytoplasm of host cells. As such, it is useful to understand the structure of tailed phage MCP before it has polymerized into a prohead as well as the evolutionary pressures that have dictated its structural properties. The canonical HK97 fold of the MCP appears to be incapable of folding properly on its own, as it precipitates out in random aggregates. This is not surprising; like any protein capable of self-polymerizing, MCPs are in danger of misfolding. Phage MCP sequences must encode complementary stretches of residues that can bind tightly to their intended partner in another monomer to form a robust capsid. However, these stretches could also lead to mis-folding if binding occurs intra- rather than inter-molecularly. HK97 itself avoids aggregation and assembles proheads only in the presence of the host GroEL/ES molecular chaperone system, as do lambda and other phages of *E. coli*. T4 codes for a particularly large MCP (521 AA), which is too large to utilize the host GroEL/ES system and thus encodes its own more capacious co-chaperonin analogous to GroES. In P22 and related phage, assembly-competent folding is provided without the apparent need for host factors by the addition of an “insertion” I-domain that folds rapidly and nucleates folding in the rest of the MCP. Point mutations in the I-domain lead to MCP aggregation and precipitation, presumably due to misfolding of the domain and thus the protein as a whole. P22’s I-domain is similar to the IG-like protein telokin. Other phages also include IG-like domains in their MCPs, but a role for these in promoting proper folding has yet to be demonstrated. However, proper folding requires more than just avoiding aggregation; initial inter-molecular interactions must be relatively weak so that MCPs in the prohead can freely dissociate and re-associate to find the lowest-energy arrangement of MCP subunits. This means that the MCP folds such that the stretches of polypeptide that bind strongly to each other in the final assembled structure are inaccessible until transformation into the mature head begins.

There is likely a soluble intermediate state of the MCP after it has folded but before it has assembled into a prohead. However, that intermediate, i.e., its conformation and oligomerization state, remains unknown for any phage. In HK97, pentamers and hexamers appear to be the soluble units. This is important because MCP is arranged in the finished capsid as pentamers and hexamers, collectively termed capsomers. It seems logical for the capsid to be constructed from pre-assembled capsomers, but this appears to not always be the case. Investigations into P22 assembly indicate that its prohead forms monomer-by-monomer, and non-denaturing gel electrophoresis studies in T7 suggest that MCPs are present as tetramers prior to prohead assembly. Clearly, more definitive biochemical, biophysical, and structural characterization of these intermediate species would significantly illuminate the mechanism underlying capsid assembly.

Prohead Assembly

Assembly of an icosahedral prohead with a Triangularization number (T number) greater than one requires MCP to adopt different conformations and bind together in different ways according to the quasi-equivalent positions of the lattice. However, this capacity for different conformations and binding also allows MCP to link together in a large number of incorrect arrangements

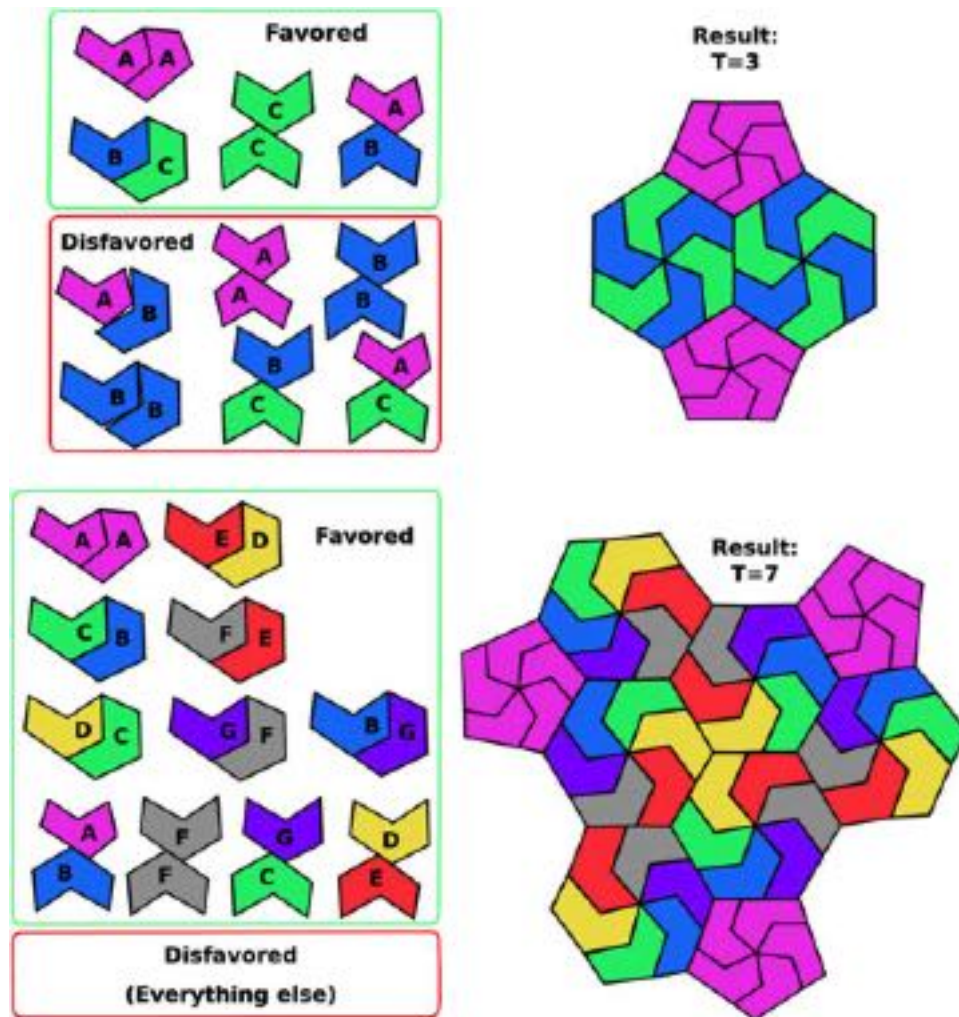


Fig. 3 Local Rules and Prohead Assembly. Two examples of ‘local-rules’ and their resulting assemblies are shown. On the top, monomer pair interactions that are more thermodynamically favorable are in a green box while pairs that do not quite fit together have less favorable interactions and are in the red box. These pairs with different interaction energies can be thought of as ‘rules’ which only allow interactions with more negative free energies of formation. The result of continuously adding monomers in compliance with the rules in the green and red boxes is a $T = 3$ icosahedron. A $T = 7$ icosahedron can be specified by a much larger set of rules, as shown in the lower half of the figure.

in addition to the correct one. Two levels of regulation are utilized to promote the assembly of the single MCP arrangement that yields functional proheads. In the first, termed ‘local rules’, interaction between MCP monomers is favored for specific pairs of different conformations and disfavored for others, as shown in Fig. 3. These rules may be satisfied by monomers with consistent conformations sampling many potential binding sites until one matches or they may move to binding sites randomly and switch conformations to match. The second level of regulation requires the participation of an additional protein known as the scaffold, which is encoded by all known tailed phages. As their name suggests, scaffolding proteins are necessary for proper prohead assembly but are not present in the mature phage. The scaffolding proteins act to protect against errors that would accumulate into fatal defects in an assembly process driven entirely by ‘local-rules’ MCP interactions. The combination of specific binding relationships dictated by the local rules of MCP and more global dynamic interactions of the scaffold ensures correct assembly of the prohead. Point mutations in either component can shift the equilibrium toward more misformed, nonfunctional proheads relative to the correct form.

Scaffolding Protein

The function of scaffolding proteins varies somewhat from phage to phage and has not been determined in full for any species. It has been investigated via *in vivo* studies of scaffold-deletion mutants and *in vitro* assembly experiments in which the scaffold was not included at all, or was included only as a truncated version. These experiments produce a wide range of phenotypes. T7 MCP produces no structures at all without scaffold, and the MCP remains in solution. The prolate phages T4 and phi29, in which

wild-type (WT) capsids consist of both icosahedral and cylindrical regions, produce elongated cylindrical and icosahedral-only structures, respectively. Long cylindrical “polyheads” are also produced by lambda, although its WT capsid is not prolate and thus has no cylindrical regions. P22 deletion mutants are able to produce some complete proheads, although at a greatly reduced yield and rate compared to WT. P2 MCP precipitates as random aggregates when scaffold is not present, as does HK97 MCP mutants from which the scaffold domain has been deleted. These different phenotypes suggest that scaffolding proteins perform several different functions, and the importance of each function varies among different phages. Some functions of scaffolding proteins that have been proposed include: (1) accelerating assembly, (2) promoting the correct overall curvature and shape of the prohead, (3) occupying the interior volume of the prohead to prevent host molecules from being trapped inside, (4) stabilizing MCP in its prohead conformation to prevent premature maturation, and (5) incorporating the portal protein at the unique vertex.

Accelerating assembly

Wild-type scaffolding-protein's accelerating function during prohead assembly is demonstrated by the observed slower assembly time of P22 MCP when scaffolding is absent. The mechanism of acceleration has been proposed to result from scaffold-bound MCPs being brought into close proximity via inherent oligomerization interactions of the scaffold. Specifically, scaffolding protein sequences in all phages have long stretches of sequence that promote alpha-helix formation and coiled-coil interactions. Indeed, the only two scaffolding proteins whose structures have been determined both form dimeric coiled-coils. Hence, if each monomer in a multimer can also bind MCP, then the inherent oligomerization of scaffold would bring MCPs into close proximity, increasing the local concentration and thus accelerating binding interactions. Experiments with truncated P22 scaffolding proteins show that acceleration of assembly occurs if and only if the region that promotes dimerization is present. In addition to the highly structured coiled-coil regions, most scaffolds also code for C-terminal regions of unstructured polypeptide. These partially unfolded regions would permit the polypeptide to sample a larger region of space, increasing the chance of finding its binding partners.

Promote correct geometry

In addition to the coiled-coil driven dimerization described above, the only crystal structures of scaffolding proteins indicate that two dimers associate to form tetramers. While, the mechanisms by which scaffolding protein promote correct assembly are not fully known, they may depend on this ability to form higher order oligomers. As described above, P22 MCP in the presence of mutant scaffold in which the dimerization region was deleted not only assembled proheads with much less accuracy than when WT scaffold was present, but also at a decreased assembly rate. In contrast, a scaffold mutant missing the putative tetramer-promoting region accelerated polymerization of MCP relative to WT scaffold, but produced a greater number of aberrant particles. In many phages, the most commonly seen aberrant particle is a ‘spiral’ which can be imagined as a particle wherein the curvature of the shell is not correct, and two leading edges have continued to grow past each other after failing to close. This defect illustrates the inadequacy of relying on local MCP/MCP interactions to define the geometry of the large number of subunits in the prohead; even small errors in local curvature can lead to a fatal misalignment when added over 360° . The larger volumes sampled by the scaffolding protein could bridge misaligned fronts and prevent larger-scale defects like the spiral shells (Fig. 4). By assuming this error correction function, the scaffold likely allowed the MCP more freedom to evolve, resulting in the wide range of observed capsid forms and likely contributing to the evolutionary success of tailed phages.

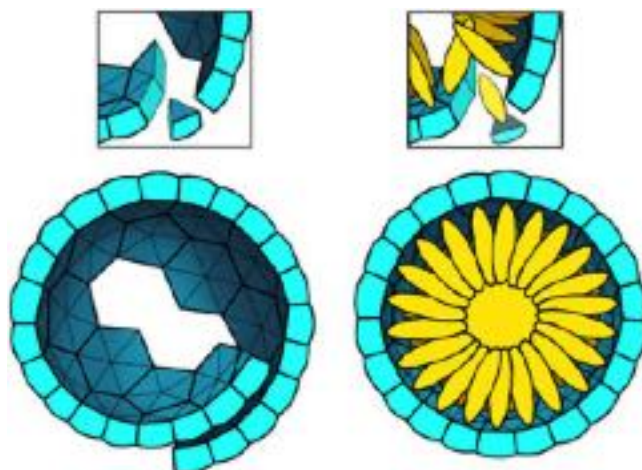


Fig. 4 The Role of the Scaffolding Protein in Correcting Errors During Prohead Assembly. Schematics of a commonly seen ‘spiral’ aberrant particle (left) and a closed prohead, assembled correctly with the aid of scaffolding protein (right) are pictured. As subunits are added to the growing assembly, differences in the curvature of the shell lead to two fronts failing to join and overlapping. The local interactions between MCP monomers cannot ‘detect’ the mismatch. Scaffolding protein monomers explore a large volume of space, so as the growth fronts approach each other they are brought into alignment by dimerization and tetramerization reactions. Proper joining is ensured and a high yield of closed shells results.

Delay MCP transformation

MCP monomers in partially assembled proheads must not transition to their tight-binding mature conformation before the subunits have assembled a closed shell with the thermodynamically favored arrangement. Scaffolding protein interactions with the N-arm of the MCP are involved in preventing premature expansion. In HK97-like phages, scaffold and major capsid protein are initially translated as a single polypeptide where the approximately 100 N-terminal amino acids correspond to scaffold and the rest of the protein to MCP. In phage that have this covalently attached scaffold (δ -domain), it must be proteolytically cleaved before maturation can occur, thus providing a mechanism to control the timing of scaffold departure. In phages that employ a separate scaffolding protein, maturation does not occur until negatively charged DNA and its counter-ions enter the prohead. High salt concentrations have been found to reduce or completely prevent binding of scaffolding protein to MCP; similarly, DNA entry may disrupt critical electrostatic interactions between MCP and scaffold.

Experimentally determined estimates of the stoichiometric ratios of scaffold to MCP in proheads are imprecise but are less than one in all cases where the scaffold is not a covalently-bound δ -domain. This implies that either not all MCP monomers are directly affected by scaffolding interactions or that the scaffold dynamically dissociates and associates with different MCPs to affect the entire prohead. If the latter, the scaffold must rebind after unbinding from an MCP monomer significantly faster than the monomer transforms to its mature state when unbound.

Exclude host factors

Crowding out host cellular components from the prohead interior is necessary to reserve the full volume of the capsid for storage of nucleic acid. A loss of volume to contaminants would prevent complete packaging of the phage genome, with an accompanying loss of infectivity. Scaffolding protein could prevent this scenario by simply occupying space inside the prohead. In experiments that reduce the length of a scaffolding protein, larger copy numbers compensate in the resulting proheads, resulting in a similar final scaffold-to-MCP mass fraction. Of course, scaffolding protein must also be able to exit the complete prohead to make room for the phage genome. Some phages have holes in their proheads' capsomers and their scaffolding proteins exit intact and can participate in assembling new proheads, while other phages co-assemble a protease inside the prohead which digests the scaffold which then exits as smaller fragments. In some cases, the scaffold itself provides the proteolytic function.

Scaffold involvement in portal incorporation

Inclusion of the portal into proheads assembled *in vitro* has been demonstrated to be promoted by the presence of scaffolding protein for many phages. Density near the portal in asymmetric cryo-EM maps of proheads has been attributed to scaffolding protein in HSV-1 and others. Mutations in the scaffolding protein of P22 have resulted in *in vitro* assembly of prohead particles that did not incorporate the portal. However, the mechanisms by which scaffolding proteins and portals interact has not been explored as thoroughly as the interactions that promote MCP assembly into the prohead.

All these functions depend on the scaffold's ability to bind phage MCP and to self-associate in a manner that promotes a particular internal volume. In phages where structural or mutagenic information is available, MCP binding typically involves a helix-turn-helix motif approximately 100 amino acids from C-terminal of the scaffolding protein binding via non-covalent interactions to the N-arm of the MCP. The fact that scaffolding proteins seem to immediately precede MCP in the phage genome suggests an evolutionary relationship between the N-terminal delta-domain of some phages and the separate but sequence-adjacent scaffolding proteins of others. Either the scaffolding protein and MCP genes fused in HK97 and some other phages, or the combined protein is ancestral and split into the two separate genes we now find in most phages.

The Portal

In addition to the MCP and the SP, viable proheads must include a portal protein. This protein assembles into a dodecameric funnel-shaped 'portal' structure, sometimes also referred to as the connector since it will 'connect' the head to the tail in the mature phage. The portal occupies the unique vertex of the icosahedral capsid and provides the pore through which DNA can enter and exit the phage during genome packaging and ejection, respectively. The outermost narrow-end of the funnel-shaped portal interacts with the DNA-packaging ATPase motor in the prohead and with the tail in the mature virion. It has been hypothesized that the wide-end of the portal funnel, which is located in the interior of the prohead, senses DNA filling and initiates an allosteric conformational change that (1) causes the motor to detach at the end of genome packaging and (2) constricts the central channel of the portal such that DNA will be blocked from exiting the prohead until the tail can bind and serve as a plug to prevent premature DNA ejection. The portal is therefore important for maturation. Proheads of T4 showed delayed expansion when the portal protein was modified. The portal also has an important role in prohead assembly, though the details are not consistent for all species of phage. In T4, phi29, HSV1, and T7 the portal has been shown to be necessary for assembly of MCP into capsids with the correct geometry, and is proposed to serve as nucleation site for procapsid assembly. This hypothesis is attractive since it provides a simple explanation as to why only a single portal is incorporated in the prohead structure. However, early experiments in phage P22 suggested that portal incorporation occurs as the last step of prohead assembly rather than the first. Specifically, *in vitro* assembly of icosahedral P22 shells can proceed from MCP and scaffolding protein alone, while inclusion of an excess of portal protein results in aberrant spiral particles with multiple portals incorporated. However, recent studies indicate that P22 assembly proceeds similarly to other phages, with the portal nucleating assembly. The structure of the portal itself suggests it must

be incorporated early in prohead assembly since the radius of the wide end of the funnel-structure of the portal is positioned in the prohead interior and is wider than the hole where the portal sits. Hence, it is difficult to imagine how this wide funnel-end of the portal could pass through this hole if it were incorporated into proheads as the last step in assembly.

Structure of the Prohead

The prohead's structure differs from that of the mature head in its overall shape and size, in the local arrangement of monomers, and in the conformation of the individual monomers themselves (Fig. 5). Compared to the mature head, the prohead is smaller in diameter, has rougher and thicker capsid walls, and more closely resembles a sphere than an angular icosahedron. The spherical shape of the prohead suggests that while the MCP monomers interact quasi-equivalently, the scaffolding proteins' dynamic oligomerization interactions are equivalent and not sensitive to different positions in the icosahedral shell. The capsomers are cup-shaped as opposed to the flat capsomers in mature heads, which is what causes the prohead shell to have a rougher appearance. Cup-shaped capsomers result in shorter distances across capsomers and therefore smaller prohead diameters and internal volumes. The previously described functions of scaffolding proteins can be achieved more easily with a smaller internal volume. The magnitude of expansion varies significantly among phages (e.g., a 100% increase in diameter for phage lambda versus a 30% change for P22). Capsomers in proheads also tend to have lower symmetry, as typified by the skewed, two-fold symmetric hexons in $T = 7$ icosahedral phage which transition into nearly six-fold symmetry during maturation. Reduced symmetry in prohead capsomers increases the ability of 'local rules' to regulate assembly since the energy difference between 'correct' and 'incorrect' hexamer/hexamer interactions will be greater. Many phages encode accessory proteins that bind to the mature capsid and provide further stabilization. For example, hoc and soc in T4, and gpD in lambda stabilize the mature head to the extent that maturation is an essentially irreversible event. These proteins are unable to bind to the MCP in the prohead state, and the detection of their binding has been used to assay prohead maturation. The differences between the prohead and mature capsids are illustrated in Fig. 5, using phage HK97, as a representative of tailed dsDNA phage generally.

All of the differences in size and shape between proheads and mature capsids are the result of conformational transitions in all the individual MCPs. These transformations are complex, and to simplify their description, we focus on the changes that occur at different inter-monomer junctions, categorized based on the number and parts of capsomers involved. We will examine the

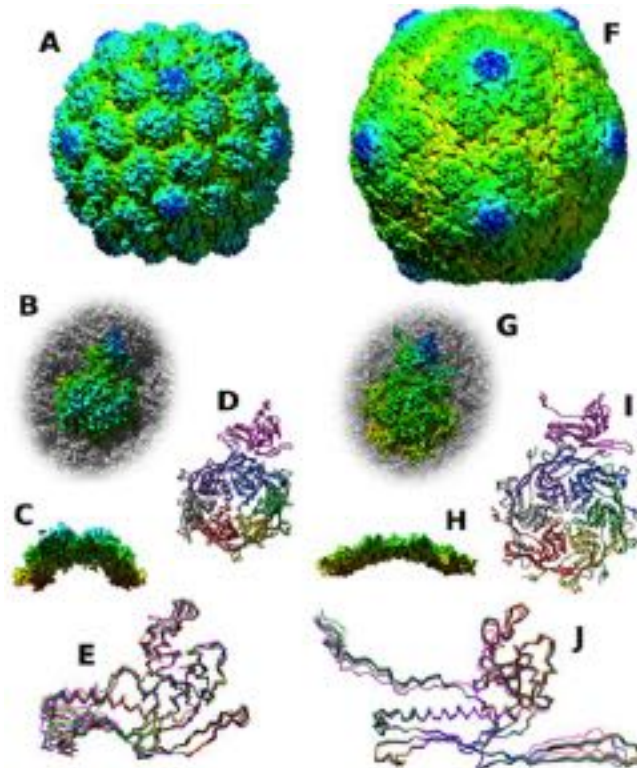


Fig. 5 Transformation from Prohead to Mature Head. The transformation of the HK97 prohead into a mature head is shown at several scales. In A and F, the entire capsid changes from a small bumpy spheroid into a larger, smoother icosahedron. In B and G, density simulated from atomic coordinates of the $T = 7$ asymmetric unit includes an entire hexon, which changes from a 2-fold symmetric skewed form into a more 6-fold symmetric hexagon. C and H show cross-sections of the hexons in which the change from a cup-shape to a more flat shape can be observed. D and I show ribbon diagrams of the atomic models of the asymmetric unit. In E and J, each monomer in the asymmetric unit has been superimposed for each state, showing the greater variability of the monomers in the prohead, especially at the tips of the A-domain and the P-domain.

conformational changes that occur upon maturation at (1) capsomer vertices where three MCPs meet, (2) across capsomer edges where two MCPs interact, and (3) at capsomer centers. In addition to the structural rearrangements at monomer-monomer interfaces, we will also describe that changes in the basic fold of the MCP that accompany maturation, (Fig. 6). Note that only five high-resolution prohead structures have been published at the time of this writing, so we should be cautious in drawing conclusions; features common to these structures may not be present in all subsequent structures.

The interfaces where three capsomers meet at their vertices remain largely the same before and after maturation. In most phages, the P-loops of three monomers contact each other, and these contacts do not change during maturation. MCPs at the five- or six-fold symmetry axis at the center of a capsomer undergo significant refolding and rearrangements during maturation, and this seems to drive the transition from skewed 2-fold hexamers to the more symmetric 6-fold state observed in mature capsids. In the prohead state, the interactions between neighboring monomers at the centers of capsomers mostly involve helices α 1 and α 2. When skewed prohead hexons mature into the more 6-fold symmetric state, this is accompanied by an increase in the length of α 1 and, in some of the monomers, movement of α 1 results in a different set of interactions with α 2. The most substantial structural changes occur at the interfaces joining the edges of two capsomers, resulting in entirely different interaction networks. These maturation changes are consistent in all available prohead/mature head structure pairs, and likely drive the transition from prohead to expanded mature head.

As discussed above, changes at the twofold interface during maturation are the most extensive and give rise to a completely different set of paired interactions. Most of these changes result from movement and refolding of the N-arm, which is uninvolved in the prohead 2-fold interface but prominently involved in the interface in the mature head state. The change in binding is most obvious at the position of the local twofold axis halfway along the interface. In the mature state neighboring N-arm strands cross each other, while in the prohead the 'N-shoulders' of opposite monomers meet. The term 'N-shoulder' refers to the loop connecting the lower strand of the E-loop to the bottom strand of β P2 which leads directly to the N-arm (Fig. 2). The thermophilic archaeal phage P23-45 is an exception in that the tips of the long spine helices meet at the 2-fold axes rather than the N-shoulders.

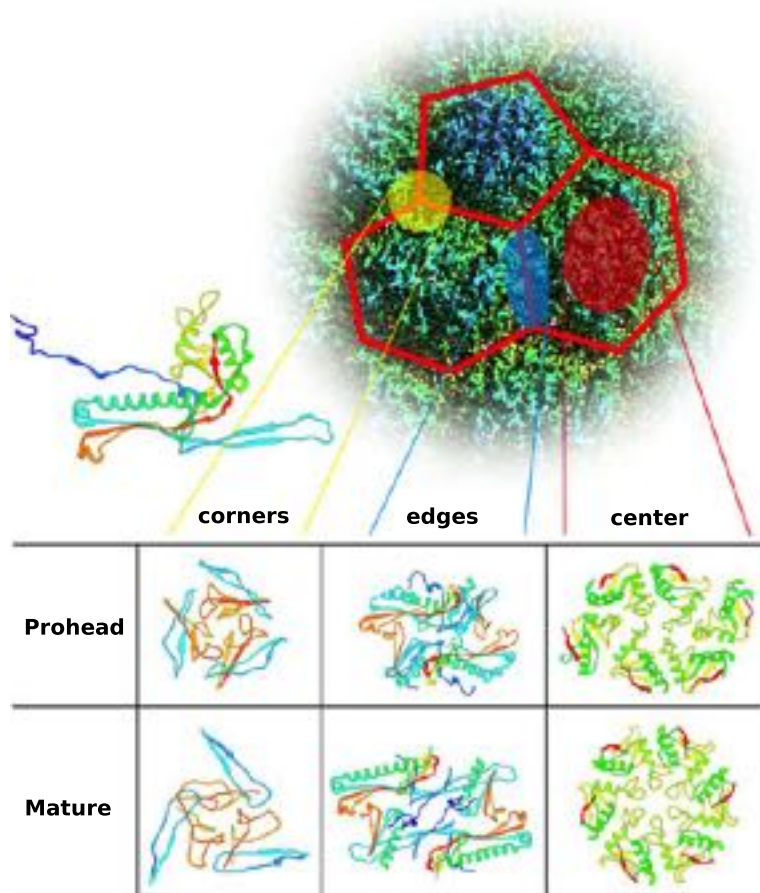


Fig. 6 Inter-MCP interfaces before and after maturation. Conformational changes in the HK97-fold during the prohead/mature head transformation are categorized by their participation in different interfaces in the capsid. Threefold junctions, where the corners of three capsomers meet, remain mostly unchanged. Twofold junctions, where the edges of two capsomers meet, undergo significant unbinding, movement, refolding, and rebinding. Junctions at the centers of capsomers undergo some refolding in all monomers. In skewed hexons, α 1 of some monomers unbind from their prior locations and rebind in positions that match those of the other monomers.

The capsomers of P23–45 (and P74–26) are much larger than those of all known mesophilic phages (~ 170 versus ~ 120 – 140 Å from capsomer center to capsomer center), which is the result of large additions and extensions to the canonical HK97 fold. Features are thus shifted into a different alignment. Despite this, the same overall logic prevails at the twofold interface, with different pairs of features binding across the interface in the prohead and mature state. Non-conserved features that supplement binding across the prohead 2-fold interface are present in most phages. Prominent examples are the G-loop in HK97 and the D-loop of the I-domain in P22. In these cases, the complementary binding site is on the outer strand of $\beta E1$ and $\beta P1$. This complementary site is closely associated with a short helix on the N-arm in the mature head state. In summary, despite differences in the details, it seems that all interactions which stabilize the two-fold interfaces of the prohead are disrupted by the N-arm at some point during maturation. The residue pairs that interact in the prohead state are in proximity with one another when the monomers are in their more convex shape, and farther apart after capsomers have flattened out in the mature state. This change in curvature is especially pronounced in the long ‘spine’ helix of the P-domain. In proheads, a second helix formed by the N-arm lies alongside the spine helix and seems to stabilize its greater curvature. Hence the N-arm is important for both stabilizing the prohead state during assembly and destabilizing it during maturation.

Stability of the Prohead

Proheads are metastable, that is, they are not in the lowest possible thermodynamic state, but the rate at which they transition to the more thermodynamically favored mature state is slow enough that they are ‘stable’ in the conversational sense of the word. The MCP in assembled proheads is in dynamic equilibrium with subunits in the solution phase. Indeed, purified proheads of P22 have been shown to partially disassemble if diluted below their dissociation constant of $5\text{ }\mu\text{M}$. Proheads of several phage have also been found to expand spontaneously (i.e., before the scaffold δ -domain has been cleaved or DNA has been packaged) under certain environmental conditions. Elevated temperature, or the addition of 4 M urea or surfactant can trigger this process. In P22 and HK97, manipulation of the environment has produced ‘wiffle-ball’ capsids, which have all their hexons present and in the flattened expanded state, but which have lost their pentons leaving large holes that give the resulting particle the appearance of a wiffle ball. This may be due to the scaffold remaining bound to the N-arms such that the N-arms from pentameric MCPs are unable to stabilize the 2-fold interfaces as in fully matured capsids. As a result pentameric capsomers are bound too weakly to remain attached to the capsid and dissociate. In Phi-29, structures of partially matured capsids have been determined where there are large gaps between the edges of pentamers and adjacent hexamers. In these particles, scaffold remains attached to the N-arms of pentons after hexons have matured, resulting in weakly bound pentamers with very few contacts joining them to neighboring hexamers. Both these particles and the HK97 wiffle balls likely represent particles where hexamers transitioned to the mature state but pentamers did not.

Assembly Parasites

Scaffold-assisted assembly of a prohead followed by maturation has been an enormously successful assembly strategy for tailed phages, as evidenced by their prevalence throughout the biosphere. However, this strategy has a drawback in that it allows parasitism by virophages that encode scaffold proteins that can compete for another phage’s MCP. The *Staph. aureus* Pathogenicity Island (SAPI) and *E. coli* satellite phage P4 both encode a scaffolding protein that induces the MCP of another ‘helper’ phage to form smaller capsids that cannot contain the genome of the helper phage and preferentially package the virophage genome. In the case of SAPI, the scaffolding protein is internal and analogous to those encoded by the helper phage, 80- α , but has a higher affinity for the 80- α MCP and outcompetes the 80- α helper scaffold for MCP binding, leading to smaller $T = 4$ capsids rather than the $T = 7$ of WT 80- α . Another SAPI-encoded protein may block MCP/helper scaffold binding by binding the helix-turn-helix region of the scaffold that binds to MCP. P4 encodes an external scaffold, sid, which directs the P2 MCP into $T = 4$ capsids and can do so when the helper phage’s internal scaffold is not present. Given the high degree of gene exchange between host cells, viruses, and virophage, we might expect to eventually find HK97-like phages that use an external scaffold similar to P4’s sid for capsid size determination instead of the ‘traditional’ internal scaffold protein.

Further Reading

- Dokland, T., 1999. Scaffolding proteins and their role in viral assembly. *Cellular and Molecular Life Sciences* 56, 580–603.
- Johnson, J., 2010. Virus particle maturation: Insights into elegantly programmed nanomachines. *Current Opinion in Structural Biology* 20, 210–216.
- Moody, M.F., 1999. Geometry of phage head construction. *Journal of Molecular Biology* 293 (2), 401–433.
- Shin, H., 2013. Viruses as self-assembled nanocontainers for encapsulation of functional cargoes. *Korean Journal of Chemical Engineering* 30, 1359–1367.
- Suhanovsky, M.M., 2015. Nature’s favorite building block: Deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* 479–480, 487–497.
- Teschke, C., 2019. The amazing HK97 fold: Versatile results of modest differences. *Current Opinion in Virology* 36, 9–16.
- Zlotnick, A., 2005. Theoretical aspects of virus capsid assembly. *Journal of Molecular Recognition* 18, 479–490.

Enzymology of Viral DNA Packaging Machines

Carlos E Catalano, University of Colorado Anschutz Medical Campus, Skaggs School of Pharmacy and Pharmaceutical Sciences, Aurora, CO, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

AXP Denotes an ATP/ADP binding site.

bZIP Conserved basic leucine zipper DNA-binding motif.

IHF The *Escherichia coli* integration host factor protein.

NTP Denotes an ATP/ADP and GTP/GDP binding site.

TerL The large terminase subunit.

Terminase holoenzyme The heterooligomeric, catalytically competent terminase enzyme complex composed of TerL and TerS subunits.

Terminase protomer The stable, protomeric TerL•TerS₂ heterotrimer of phage λ terminase.

TerS The small terminase subunit.

WA Conserved Walker A ATPase motif.

WB Conserved Walker B ATPase motif.

wHTH Winged helix-turn-helix DNA binding motif.

Introduction

Viruses are obligate intracellular parasites whose developmental pathways are initiated upon insertion of their genetic material into a host cell. Subsequently, viral genomes and structural proteins are synthesized and infectious particles are assembled utilizing a number of viral and host-encoded biomolecules. The assembly pathways follow sequential steps generating transient intermediates that transition to the next, typically in a non-reversible manner. The pathways are generally conserved within broad virus classes such as the large dsDNA viruses, which includes the *Caudovirales* (tailed bacteriophages) and the *Herpesviridae* groups. In these cases, viral DNA is typically replicated via a rolling circle mechanism resulting in linear, head-to-tail concatemers of the viral genome (*immature DNA*). Expression of late viral genes produces structural proteins that self-assemble into procapsid shells into which the newly replicated DNA is actively packaged. *Genome-packaging* represents the intersection between the DNA replication and capsid assembly pathways and involves processive excision of a single genomes from the concatemer and simultaneous packaging of the “mature” DNA into a preassembled procapsid shell; both reactions are catalyzed by a virus-encoded *terminase enzyme*.

Genome packaging pathways are strongly conserved in the dsDNA viruses and can be summarized as outlined in [Fig. 1](#); the terminase enzymes assemble at a specific packaging initiation site in the genome concatemer (*pac* or *cos*) to afford a stable, site-specifically bound *maturation complex*. This activates an endonuclease activity that cuts (matures) the duplex to generate the first end of the genome to be packaged. The post-cleavage complex binds to a portal situated at a unique vertex of a pre-assembled procapsid; the portal is a ring-like structure that provides a conduit for DNA entry into the capsid during packaging and for exit during infection. Terminase binding to the portal affords the *packaging motor complex*, which triggers a transition from a stable, site-specifically bound maturation complex to a dynamic motor that translocates DNA into the shell in a sequence-independent manner, fueled by ATP hydrolysis.

Terminase motors are extremely processive and very powerful; they package the entire genome length in a single binding event and to liquid crystalline density, which generates a remarkable > 25 atmospheres of internal shell pressure. Upon packaging a genome-length of DNA, the enzyme transitions back to a static maturation complex that again cuts the duplex to terminate the packaging process (from which the enzymes derive their name). The binary terminase•DNA complex, formally equivalent to the maturation complex that initiated packaging, is displaced from the genome-filled capsid by “finishing” proteins and binds another empty procapsid to initiate a second round of processive genome packaging ([Fig. 1](#)). Addition of a pre-assembled tail to the nucleocapsid (*Caudovirales*) or a complex lipid envelopment process (*Herpesviridae*) ultimately affords an infectious virus.

The essential features described above are strongly conserved in the large dsDNA viruses, both prokaryotic and eukaryotic, and the terminase enzymes perform two essential functions in virus assembly; (1) nucleolytic excision of individual genomes from a concatemeric precursor (*maturation reaction*) and (2) translocation of viral DNA into the procapsid (*packaging reaction*). These two reactions are catalyzed by two distinct terminase complexes that sequentially alternate during processive genome packaging. Notable exceptions to this general paradigm are represented by the phage $\phi 29$ and adenovirus groups, which replicate monomeric genomes in a protein-primed manner and thus have no maturation requirement. Notwithstanding, these viruses utilize an analogous “packaging ATPase” enzyme to package viral DNA into a pre-assembled procapsid shell. Herein I focus on the enzymology of the dsDNA viruses that package genomes from a concatemeric DNA precursor, which include phages λ , HK97, T4, P74-26, P22 and SPP1, among others, and the eukaryotic herpesviruses.

Unit Length Versus Headful Packaging Mechanisms

Within the broad class of dsDNA viruses, terminases carry out two basic strategies for genome packaging. The “unit-length” phages both initiate and terminate packaging at a “*cos*” sequence and thus package exactly 100% genome length ([Fig. 1](#)); the herpesviruses

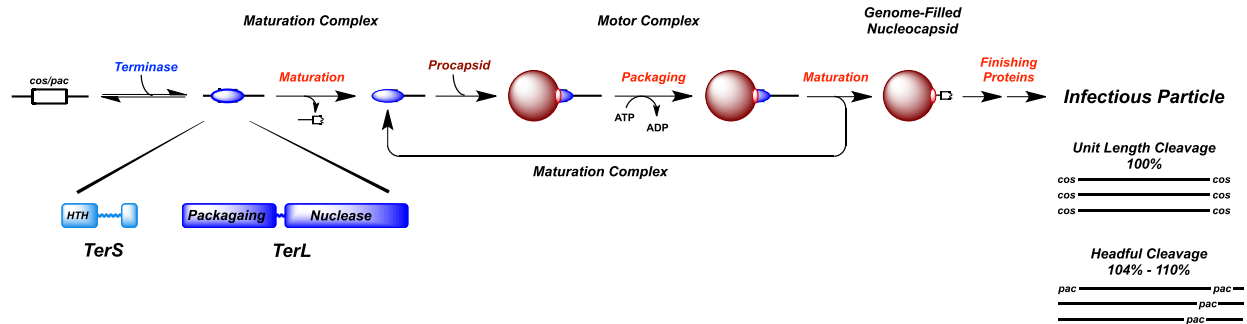


Fig. 1 Processive genome packaging in the dsDNA viruses. Terminase enzymes are responsible for processive excision of an individual genome from a concatemeric packaging substrate (*genome maturation*) and for translocation of the duplex into a pre-formed procapsid shell (*genome packaging*). The functional enzymes are composed of large catalytic (*TerL*) and small DNA recognition (*TerS*) subunits, both of which possess conserved functional domains, as depicted at bottom, left. Two basic strategies for genome packaging, unit-length and headful, are summarized at bottom, right. Processive packaging requires that terminase cycles between a stable maturation complex and a dynamic motor complex. Details are provided in the text.

similarly initiate and terminate packaging at repeated “a” sequences in the concatemer. In contrast, while the “headful phages” initiate packaging at a “*pac*” sequence, the translocating motors bypass the downstream *pac* site and continue packaging until a “head full” of DNA has been inserted into the shell. This affords virions that contain DNA 104%–110% larger than the unit-length genomes and that possess “circularly permuted, terminally-redundant” ends (Fig. 1). In both cases, processive genome packaging from the concatemer requires that the terminase enzymes catalyze both genome maturation and DNA translocation reactions, alternating between distinct, but tightly coupled nucleoprotein complexes.

The Viral Terminase Enzymes

Terminase enzymes function as heterooligomeric complexes consisting of a large catalytic subunit that contains both maturation and packaging activities (*TerL*) and a small DNA recognition subunit that is required for site-specific assembly at the packaging initiation site (*TerS*; Fig. 1); both subunits are essential for virus development *in vivo*. In some cases, such as phages λ and P22, the holoenzymes can be isolated as stable complexes of both subunits, while in others, such as phages T4 and SPP1, the two subunits do not strongly interact until they assemble the maturation complex on DNA.

The *TerS* Subunits

The small DNA-recognition subunits (~17–23 kDa) are generally composed of conserved functional domains (see Fig. 1); (1) an N-terminal DNA binding domain containing a helix-turn-helix DNA binding motif, (2) a highly helical central self-association domain and (iii) a C-terminal domain that also plays a role in DNA binding and that is required for binding to the *TerL* subunit in the holoenzyme. The isolated *TerS* subunits assemble into ring-like structures ranging from 8 to 11 subunits of varying dimensions, and in some cases double rings. Due to this structural diversity, there is debate as to whether *TerS* binding to DNA threads the duplex through the central channel of the rings or if it wraps the duplex around the perimeter of the rings. This is discussed further below.

The *TerL* Subunits

The large catalytic subunits (~50–80 kDa) are typically monomers in solution whose overall fold is conserved despite their divergent primary sequence; they are similarly composed of conserved domains (see Fig. 1). An N-terminal domain contains the packaging ATPase site that fuels motor movement; sequence, structural and biochemical data suggest that the packaging ATPase belongs to the Additional Strand, Conserved Glutamate (ASCE) family of ATPase enzymes. As such, they possess the conserved Walker A (WA) and Walker B (WB) motifs that participate in ATP binding and positioning the γ -phosphate for hydrolysis. A conserved “catalytic glutamate” lies immediately downstream of the classical WB motif and is proposed to position and activate a water molecule for nucleophilic attack on the γ -phosphate in the hydrolysis transition state. The terminase family also possess a *trans*-acting “arginine finger” residue that is conserved in ring ATPase enzymes; these residues are proposed to (1) catalyze ATP hydrolysis in an adjacent subunit and (2) to communicate the nucleotide bound state of one subunit in the motor to the next. These features serve to ensure inter-subunit communication and coordination of ATP hydrolytic events during motor movement. Finally, a “lid” domain is also conserved in the terminase enzymes that is likely involved in coupling ATP hydrolysis to duplex movement.

A short linker connects the packaging ATPase domain to a conserved C-terminal nuclease domain responsible for genome maturation. Structural interrogation suggest that the nuclease domains are related to the RNaseH family of nuclease enzymes. Non-specific nuclease activity has been demonstrated in several headful TerL subunits and their nuclease domains, including phages T4, SPP1 and P74-26, among others, and the nuclease domain of herpes simplex and cytomegaloviruses. Biochemical and kinetic characterization of nuclease activity of these proteins has generally been modest; nevertheless, similar features are apparent in both prokaryotic and eukaryotic systems, as follows. Duplex nicking requires divalent metal and structural studies suggest a two-metal nuclease mechanism. Nuclease activity is non-specific, it is often weak and where examined, it is inhibited by the TerS subunit. Kinetic analysis of the nuclease activity of P74-26 TerL indicates that the enzyme possesses both single-strand nicking and double-strand duplex cleavage activities.

In contrast to the headful phages discussed above, the nuclease activity of unit-length genome packaging terminases, such as those from phages λ , 21 and HK97, is site specific and strictly requires the TerS subunit. This is described in detail below.

The Terminase Holoenzymes

Genetic and biochemical data indicate that holoenzyme complexes composed of both TerL and TerS subunits are required to assemble the maturation complex at *pac* or *cos* to initiate genome packaging. Where characterized, C-terminal residues in the TerS subunits are required for binding to N-terminal residues in TerL to assemble functional holoenzymes. Unfortunately, structural studies performed to date have characterized the TerL and TerS subunits in isolation and while they have provided significant insight into structure-function relationships of the isolated terminase subunits, there is little information on the holoenzyme complexes assembled from both subunits. Elegant single molecule tweezer studies have revealed mechanistic details of DNA translocation by the translocating motors, again most often in the absence of the cognate TerS subunit. Moreover, there is limited information regarding the conserved genome *maturation* complexes and a mechanistic interrogation of the tightly coupled maturation and packaging catalytic activities remains largely unexplored. An exception is phage λ terminase, which is isolated as a stable heterotrimer of TerL and TerS subunits and has been extensively studied genetically, biochemically and by single-molecule techniques.

This article focuses on the enzymology of phage λ terminase as a model system for interrogating the conserved nucleoprotein complexes that catalyze genome maturation, that drive DNA translocation and the cyclic interconversion between the two required for processive genome packaging. Biochemical and kinetic characterization of λ terminase has revealed complex allosteric interactions between the nucleotide and DNA binding sites that bear directly on these conserved transitions. While the focus will be on λ , comparisons to other viral systems are discussed in order to provide a broad discussion of the enzymology and biochemical features of virus DNA packaging.

The Lambda System

Bacteriophage λ is composed of a linear 48.5 kb dsDNA genome tightly packaged within an icosahedral capsid shell that contains a tail structure situated at a unique vertex of the icosahedron (the portal vertex). Injection of viral DNA into an *Escherichia coli* cell initiates the infection cycle, which ultimately affords linear concatemers of the λ genome and packaging-competent procapsids. Phage λ assembly has been fully reconstituted *in vitro* and an infectious virus may be assembled using purified proteins and commercially available λ DNA; this allows mechanistic interrogation of each step along the virus assembly pathway.

The Terminase Enzyme of Phage Lambda

λ terminase purifies as a stable, homogeneous heterotrimer composed of two TerS subunits tightly associated with a single TerL subunit [TerL•TerS₂, the *terminase protomer*]. The two subunits possess the conserved functional domain organization described above and a number of structural motifs and catalytic sites have been characterized in each subunit. The protomer is catalytically silent and is activated only when assembled into a higher-order ring-like *holoenzyme complex*. Detailed biochemical interrogation of the enzyme has revealed complex allosteric interactions between the multiple substrate binding and catalytic sites identified in the enzyme, as outlined in Fig. 2.

The λ TerS Subunit

Genetic, biochemical and structural studies identified an N-terminal DNA binding domain (DBD) in λ TerS that contains a winged helix-turn-helix (*wHTH*) motif and that assembles into a stable dimer. Based on structural modeling data, we proposed that the HTH motif interacts with the major groove of DNA and that the wing acts as a nucleotide modulated switch that regulates *cos*-specific versus non-specific DNA binding interactions that are important for maturation and translocation activities, respectively. Consistent with this model, a low-affinity nucleotide binding site has also been identified in this domain, which binds both ATP/ADP and GTP/GDP nucleotides (NXP, $K_{D,app} \sim 0.5\text{--}1\text{ mM}$). While early studies reported ATP/GTP hydrolysis at this site, more recent studies with the purified and homogenous protomer suggest that turnover at this site is very low. Rather, this nucleotide binding site appears to act as an allosteric regulator of terminase function; NXP binding to this site (1) modulates *cos*-specific versus non-specific DNA binding

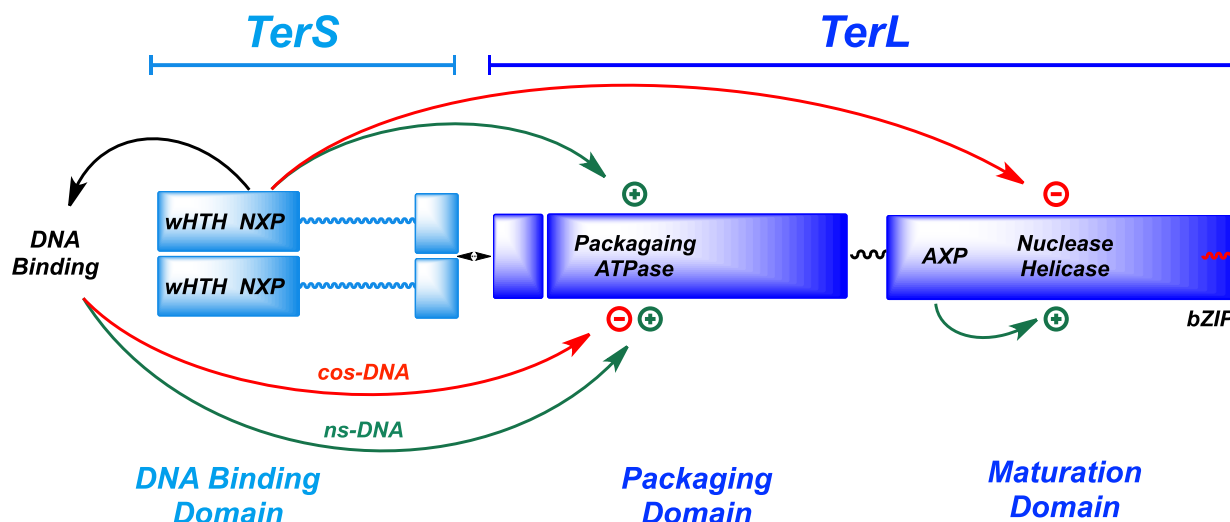


Fig. 2 Domain organization and allosteric communication between the binding and catalytic sites in λ terminase holoenzyme. TerS and TerL are shown in cyan and blue, respectively. The TerS DNA binding domain contains a winged helix-turn-helix DNA binding motif (*wHTH*) and an allosteric NXP binding site (*NXP* denotes ATP/ADP or GTP/GDP). The TerL packaging ATPase and genome maturation domains are depicted, connected by a short linker domain. *AXP*, denotes ATP/ADP and *bZIP* denotes a basic leucine-zipper motif. Characterized allosteric interactions between the nucleotide, DNA and catalytic sites are indicated. Details are provided in the text.

interactions, (2) inhibits the nuclease activity of TerL and (3) stimulates ATP hydrolysis by the packaging ATPase site in TerL (Fig. 2). These features have been incorporated into a “nucleotide switch” model that is responsible for alternation between the maturation and packaging complexes during processive genome packaging (*vide infra*).

Following the DBD is a long (~ 100 residues) hydrophobic coiled-coil helix that plays an important role in TerS self-assembly interactions and that is likely involved in protomer and motor assembly. The helix connects to a C-terminal domain that similarly plays a role in higher-order self-association interactions, that is required for high-affinity DNA binding interactions and that is also responsible for binding to the TerL subunit in the λ protomer (**Fig. 2**). As noted above, the domain organization and structural features described for λ are generally conserved in all of the TerS subunits characterized thus far (*vide supra*).

The TerL DNA Packaging Domain

The genome packaging activity of λ terminase is located in an N-terminal domain of TerL that contains the highly conserved ASCE ATPase catalytic site described in detail above. Kinetic studies reveal that the λ domain binds ATP with high-affinity ($K_M = 5 \mu\text{M}$) and that turnover at this site is stimulated by NXP binding to the TerS subunit in the holoenzyme (see [Fig. 2](#)). Further, non-specific DNA duplexes stimulate ATP hydrolysis at this site, while binding to specific *cos*-containing DNA sequences inhibits ATPase activity. These features are likely involved in regulating terminase ATPase activity in the maturation versus motor complexes.

The λ TerL Maturation Domain

As a unit-length packaging enzyme, λ terminase introduces site specific nicks into the *cos* sequence 12 bases apart. This catalytic activity is contained within a C-terminal RNaseH-like nuclease domain of TerL that is conserved within the terminase family, as discussed above. In contrast to the TerL subunits of the headful phages, λ TerL does not possess non-specific nuclease activity. Moreover, site-specific cleavage at *cos* requires the TerS subunit and further assembly into a multimeric nuclease complex. A so-called “helicase” strand-separation activity is also centered in this domain which separates the nicked, annealed strands generated by the *cos*-cleavage reaction; the upstream D_R fragment is ejected from the complex and the D_L strand, which is to be packaged, remains tightly bound to the enzyme ($T_{1/2} \sim 8$ h).

A second high-affinity ATP/ADP (AXP) binding site is situated in the nuclease domain ($K_{D,app} \sim 1 \mu\text{M}$) and AXP binding to this site strongly stimulates both *cos*-cleavage and strand-separation activities of the enzyme (see Fig. 2). Turnover at this site is not observed and AXP appears to act solely as an allosteric effector of TerL maturation activity. It is noteworthy that ATP stimulates the non-specific nuclease activity of TerL subunits in the P22, Sf6 and P74-26 phage systems, but this has not been studied in detail. Further, a second ATP binding site, distinct from the packaging ATPase site was also proposed in the phage T4 TerL subunit based on sequence homology; however, genetic studies have not revealed a defined function for the putative site.

While the domain organization of λ terminase is strongly conserved among all of the terminase enzymes, catalytic variations exist. An interesting variation is the terminase enzyme from phage HK97, which also possesses a *cos*-specific nuclease activity that is dependent on TerL and TerS subunits; however, efficient duplex nicking further requires a nuclease-associated "HNH" protein, that appears to be common among the long-tailed phages. The mechanism by which this HNH protein stimulates the *cos*-cleavage

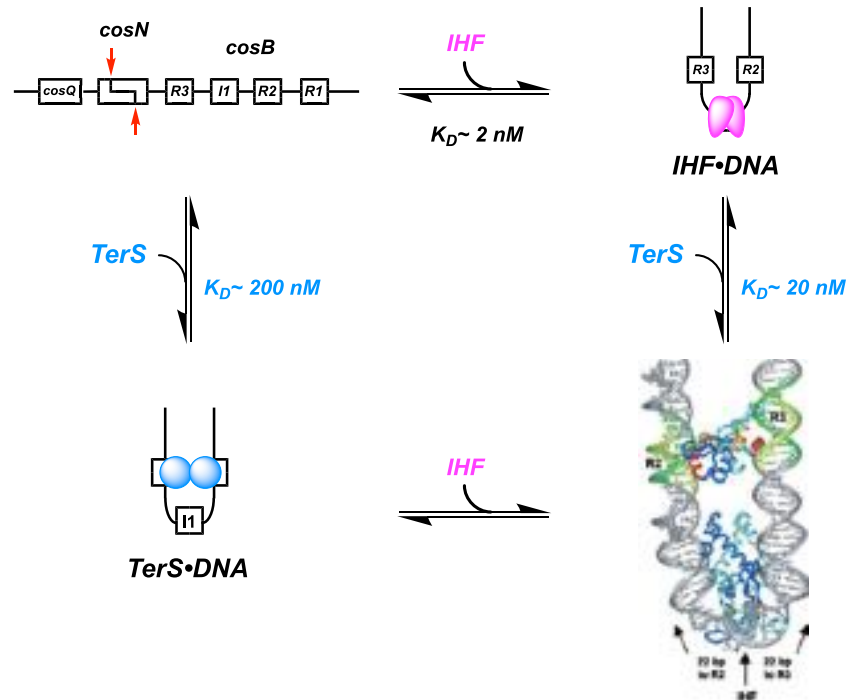


Fig. 3 Phage λ TerS and *E. coli* IHF cooperatively bind and bend *cos*-DNA. The λ *cos* sequence is bi-partite consisting of *cosB* terminase binding and *cosN* duplex nicking subsites. *cosN* possesses a pseudo-dyad symmetry and terminase introduces nicks into the duplex, 12 bases apart (red arrows). IHF binds to the *I1* element while a TerS dimer binds to the *R3* and *R2* elements, both of which introduce a strong bend in the duplex. Characterized thermodynamic binding constants are indicated and a structural model for the ternary IHF•TerS•*cosB* complex is shown. Similar binding constants are obtained with the λ terminase protomer [TerL•TerS₂] indicating that terminase assembly at *cos* is primarily mediated by the TerS subunit. The bent and likely wrapped duplex presumably positions a head-to-head TerL dimer symmetrically disposed at *cosN* as a catalytically competent duplex nicking complex. Structural details remain speculative at this time.

reaction remains obscure. Of note, the herpesviruses also appear to encode a three subunit terminase, but the correspondence between the prokaryotic and eukaryotic terminase subunits is unclear.

Escherichia Coli Integration Host Factor

IHF was first identified as a host protein required for λ lysogeny. This 20 kDa α/β heterodimer plays an important role in host DNA replication, transcription and site-specific recombination. IHF is also required for lytic λ development and strongly stimulates the genome maturation activities of λ terminase *in vitro*. In all characterized cases, IHF binds to a specific recognition element and introduces a strong bend in the DNA (160° – 180°); this provides a duplex architecture conducive to assembly of additional regulatory/catalytic proteins at that site.

The Genome Maturation Complex

The Lambda Cohesive End Site and Assembly of the Genome Maturation Complex

The λ *cos* sequence spans ~ 260 bp and is bi-partite consisting of a binding subsite (*cosB*) that is required for site maturation complex assembly and a nicking subsite (*cosN*) into which terminase introduces symmetric nicks in the duplex (Fig. 3). Genetic and biochemical data indicate that the isolated TerS subunit specifically binds to three repeating “R-elements” in *cosB* and similar to IHF, TerS binding introduces a strong bend in the duplex (160°). Cooperative binding and bending of *cos*-DNA by IHF and λ TerS has been demonstrated ($K_{D,app} = 22 \text{ nM}$, $\Delta\Delta G = 1 \text{ kcal/mol}$). A high-resolution NMR structure of the λ DNA-binding domain has provided insight into the molecular details of the IHF•TerS•*cosB* nucleoprotein complex, as depicted in Fig. 3; IHF binds to the *I1* element in *cosB*, bending the duplex and juxtaposing the *R3* and *R2* TerS binding elements. The wHTH motifs in the TerS DBD are presented on opposite sides of the dimer such that each can appropriately bind to an R-element in the bent duplex.

Ostensibly, the TerS•IHF•*cosB* bent duplex serves to appropriately position two TerL subunits symmetrically disposed at the *cosN* subsite to enable high-fidelity *cos*-cleavage activity (red arrows, Fig. 3). Similar to what is observed with the isolated TerS subunit, biophysical studies have demonstrated that the terminase protomer [TerL₁•TerS₂] and IHF also bind to *cos*-DNA with high affinity ($K_{D,app} = 20 \text{ nM}$) and cooperativity ($\Delta\Delta G = 1 \text{ kcal/mol}$). Comparison of the TerS and protomer thermodynamic binding data suggest

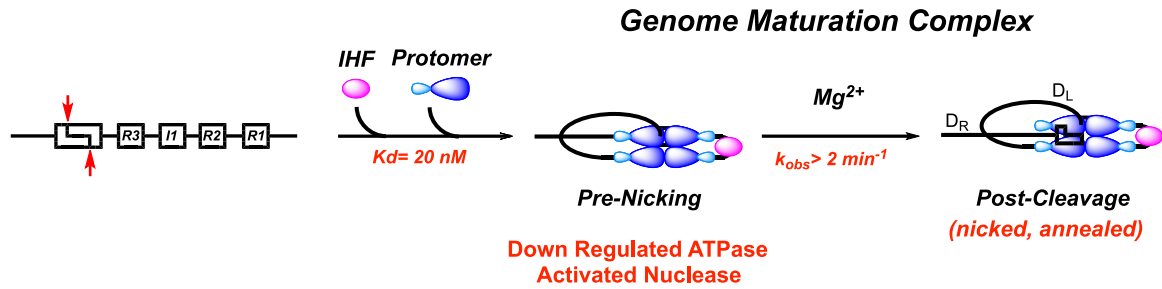


Fig. 4 Model for λ maturation complex assembly at *cos* and duplex nicking. Four λ terminase protomers and IHF cooperatively bind to the *cos* sequence to assemble a dimer of dimers nuclease complex and position a head-to-head TerL dimer symmetrically bound to the *cosN* subsite; the nucleoprotein complex strongly bends and wraps the duplex. Assembly down-regulates the packaging ATPase site and activates nuclease activity, which introduces symmetric nicks into *cosN*, 12 bases apart (red arrows).

that recognition and assembly of the terminase protomer at *cos* is primarily mediated by the TerS subunit. Consistent with these observations, while terminase and IHF cooperatively assemble at the *cos* sequence to engender a catalytically competent nuclease complex, the protomer alone does not discriminate between non-specific and *cos*-containing duplexes and is devoid of nuclease activity.

The physical nature of the duplex in the maturation complex is of interest. As discussed above, isolated TerS subunits and subdomains assemble into ring-like complexes with central channels of varying dimensions, which has led to a controversy as to whether DNA passes through the central channels or whether the complexes wrap the duplex around the TerS ring exterior; however, the role of TerS is to site-specifically assemble TerL into a catalytically-competent holoenzyme complex at *cos/pac*. Thus, interpretation of the structural data is complicated by the fact that (1) the TerS structures were obtained in the absence of the cognate TerL subunit and (2) they were all examined in the absence of DNA, both of which are required for the assembly of a biologically-relevant complex. It is further unclear whether the structurally characterized complexes are representative of TerS subunits assembled into a maturation complex or into the packaging motor complex. Thus, the duplex structure in the maturation complex remains speculative in all systems.

Model for the Assembly of the Genome Maturation Complex

The ensemble of structural, kinetic and biophysical data can be incorporated into a model for maturation complex assembly at *cos* and the activation of maturation activities in phage λ terminase, as depicted in Fig. 4. The terminase protomer, which is the relevant species during a productive infection *in vivo*, is essentially devoid of catalytic activity; however, the protomer and IHF at physiological concentrations cooperatively assemble at a *cos*-sequence of a genome concatemer to afford a catalytically-competent maturation complex. This provides a mechanism to avoid potentially damaging effects of non-specific nuclease; activity is activated only upon assembly at the packaging initiating site. Biophysical interrogation of the λ maturation complex suggests that four protomers and one IHF dimer assemble at the *cos* sequence to engender the catalytically competent ring-like nuclease complex [TerL₄•TerS₈•IHF] in which the duplex is highly bent and/or wrapped. Importantly, kinetic studies demonstrate that the packaging ATPase site in TerL is downregulated in the assembled maturation complex (see Fig. 2); this allosteric interaction serves to ensure a stable, site-specifically bound nucleoprotein complex and prevent pre-mature translocation from the *cos*-sequence.

Given the conservation of structure and function in the terminase enzymes, this model is likely recapitulated, albeit with some variation, in the terminase enzyme family. Consistent with this hypothesis, TerS-mediated duplex bending at the *pac* site has been demonstrated in the SPP1 system, among others. In contrast, the TerS subunits *inhibit* the nuclease activity and *stimulate* the ATPase activity of TerL in the T4, SPP1 and P22 headful phage systems *in vitro*. These observations appear contrary to the role of TerS in assembling a stable, site-specifically bound nuclease complex to initiate genome packaging and it is clear that further biochemical characterization of the headful terminase enzymes is necessary to clarify this discrepancy.

Genome Maturation: The *cos*-Cleavage Reaction

Once assembled at *cos*, λ terminase cuts the duplex within the *cosN* subsite. This site displays two-fold rotational symmetry and the nicks are symmetrically introduced into the upper and lower strands 12 bases apart (Fig. 4). Based on this symmetry and the presence of a basic leucine zipper (bZIP) DNA-binding/dimerization motif in the C-terminus of TerL (Fig. 2), it is not surprising that early models predicted that two TerL subunits are directly bound to the *cosN* half-sites, “head-to-head” in the maturation complex. This model bears analogy to the classical type II restriction endonuclease enzymes and similar head-to-head nuclease complexes have been proposed in the T4 and SPP1 systems; however, biophysical data suggest that the catalytically competent λ complex is composed of *four* terminase protomers. How can this be rationalized?

The *cos*-cleavage endonuclease reaction and the presumed stoichiometry of terminase protomers in the maturation complex is reminiscent of the Type IIE and Type IIF restriction endonucleases. In these enzymes, a dimeric head-to-head complex can assemble at a specific duplex nicking site, but this intermediate possesses weak nuclease activity. The catalytic activity of the dimer is strongly activated by direct association with a *second* DNA-bound dimer to yield a “dimer of dimers” nuclease complex through

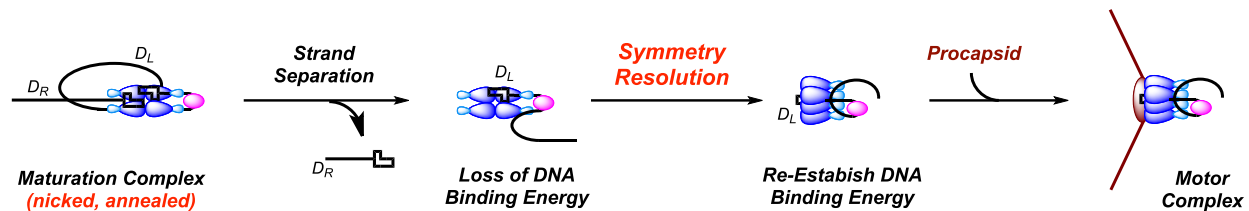


Fig. 5 Model for symmetry resolution and motor assembly. The head-to-head dimer of dimers nuclease complex must re-orient the protomers to engender a parallel ring-like motor complex for binding to the portal vertex (*symmetry resolution*). TerL and TerS subunits are depicted in blue and cyan, respectively. Details are provided in the text.

DNA looping. In the case of the IIE enzymes the second DNA-bound dimer acts as an allosteric effector for cleavage at only one of the cognate sites. In analogy, we propose that two terminase protomers assemble as a head-to-head dimer of TerL subunits bound symmetrically at *cosN*. Bending of the duplex by IHF and TerS positions DNA such that it can interact with a second dimer to afford an activated dimer-of-dimers nuclease complex, as depicted in Fig. 4.

In contrast, the type IIF enzymes similarly assemble as a dimer of dimers, each bound to a cognate recognition sequence and *both* sites are cut by the activated enzyme. Within this context, “synapse” initiation models have been proposed in the phage T4 and T7 systems where it has been suggested that the maturation complex binds to two sequential *pac* sequences in the DNA concatemer, separated by one genome length. Thus, it is feasible that in the context of a multi-genome concatemer, a λ terminase dimer of dimers interacts with two sequential *cos* sequences (separated by 48.5 base pairs) to engender an activated synapse maturation complex as proposed in other systems. Indeed, this model has been previously considered in phage λ .

Genome Maturation Summary

A central dogma is that all dsDNA viruses that package DNA from multi-genome concatemers, from phages to herpesviruses, initiate the process by assembling a maturation complex at a specific initiation site; however, these complexes remain poorly studied and ill-characterized in most systems. All of the characterized viruses encode a TerS subunit that is essential to virus development *in vivo*. Presumably, this reflects the role of TerS in recognition of a specific packaging sequence present only in the viral genome and the assembly of TerL into a catalytically competent nuclease complex, the first and essential step in genome packaging. In this regard, it is noteworthy that phage T4 destroys host DNA and synthesizes viral concatemers that contain 5-hydroxymethyl-cytosine in place of cytosine. In this case, a *pac*-like sequence is not formally required to specifically recognize viral DNA since host DNA does not exist. Nevertheless, T4 TerS is required for packaging concatemeric DNA both *in vitro* and *in vivo*, despite the fact that a defined *pac* sequence is not apparently required to initiate the process. This likely reflects a TerS requirement for maturation complex assembly and duplex cleavage in a concatemeric substrate and/or subsequent motor assembly. Whatever the case, given the conservation of structure and function of the terminase enzymes, the λ terminase maturation complex, which has been extensively characterized, provides a wealth of information that can be generalized to all of the dsDNA viruses.

The Genome Packaging Complex

Assembly of the Packaging Motor

The next step in the packaging pathway is binding of the maturation complex to the portal vertex of an empty procapsid to afford the packaging motor complex (see Fig. 1). While there is some controversy with the motor orientation in phage T4, genetic, biochemical and structural studies have localized procapsid binding residues to the C-terminus of λ , P22 and T3 TerL subunits, and also in the ϕ 29 packaging ATPase. Thus, it has long been presumed that the packaging motors are composed of TerL subunits assembled in a ring-like complex oriented in a parallel manner such the C-terminal portal interaction residues directly bind to the portal ring. In contrast, it is generally presumed that the *maturation* complex orients the TerL subunits in a head-to-head arrangement for symmetric nicking of *cos*, as depicted in Fig. 4. This begs the question as to how a head-to-head maturation complex resolves this symmetry issue in the transition to a parallel packaging motor complex. Based on kinetic and biochemical data, we propose the following plausible mechanism for “*symmetry resolution*” in phage λ (Fig. 5).

Subsequent to duplex nicking, the strand-separation activity of the maturation complex ejects the upstream D_R strand; this results in the loss of DNA binding energy by the *upstream* protomers. Next, the upstream protomers flip into a parallel orientation with all of the TerL subunits bound to the newly formed D_L end. This conformational reorganization is driven by cooperative binding of the TerL subunits to the downstream DNA, plus TerS self-association interactions, mediated by the hydrophobic helix as discussed above (see Fig. 2). This resolved conformation (1) projects the C-terminal procapsid binding residues of TerL in a parallel orientation for direct and cooperative binding to the dodecameric portal complex and (2) affords an oligomeric TerS ring-like structure that is conserved among the published structures of TerS subunits (*vide supra*).

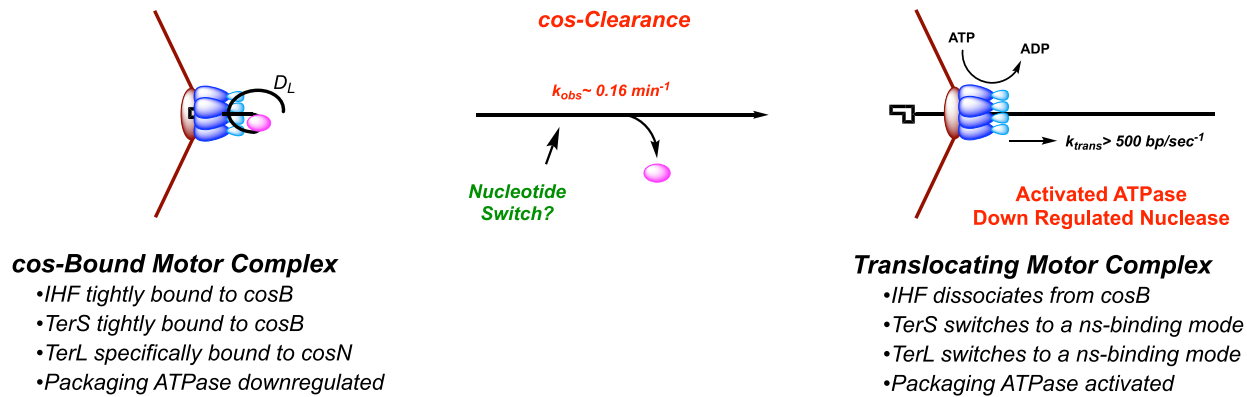


Fig. 6 Model for *cos*-clearance – the transition to a translocating motor complex. Several changes must occur for *cos*-clearance, as summarized in the Figure. The allosteric NXP binding site in TerS may serve as a “nucleotide switch” that modulates DNA binding, nuclease and packaging ATPase catalytic activities to promote the transition. Details are described in the text.

cos-Clearance: Transition to a Translocating Motor

Once the motor complex has been correctly assembled at the portal, a series of events must occur to transition from a stable complex specifically bound at the genome end, to a dynamic translocating motor complex. We refer to this transition as “*cos*-clearance” in analogy to RNA polymerase clearance from the promoter and the transition to processive RNA synthesis. Again, the λ system provides insight into this essential process as depicted in Fig. 6; (1) IHF specifically bound at the *I1* element of *cosB* must release from the duplex and (2) TerS, site-specifically bound to the *R3* and *R2* elements of *cosB* must switch to a non-specific DNA binding mode. This must occur to “un-bend” the DNA to provide a linear duplex for motor translocation from *cos*; (3) the TerL subunits specifically bound at *cosN* must also switch to a non-specific DNA binding mode to release the duplex end; (4) the packaging ATPase activity of TerL, which is downregulated in the maturation complex, must be activated to fuel translocation. In addition, the nuclease activity of the motor must be down-regulated to prevent damage to the duplex during translocation.

A Nucleotide Switch for *cos*-Clearance?

The molecular events leading to *cos*-clearance remain obscure, but it is clear that this requires coordinated allosteric interactions between the DNA and nucleotide binding sites in both the TerL and TerS subunits. As discussed above, several relevant kinetic interactions have been identified and characterized in λ terminase as summarized in Fig. 2. Most relevant to *cos*-clearance is the observation that nucleotide binding to the low affinity NXP binding site in TerS (1) modulates *cos*-specific vs. non-specific DNA binding interactions, (2) inhibits the nuclease activity in TerL and (3) stimulates ATP hydrolysis by the packaging ATPase site in TerL. These are the precise features required for the *cos* clearance transition and we posit that the NXP binding site in TerS serves as a “nucleotide switch” that regulates site specific DNA binding and nuclease activation versus non-specific DNA binding and motor translocation functions. Further, non-specific DNA also activates the packaging ATPase site in TerL, ostensibly by cooperative mechanochemical coupling events in the translocating motor. The mechanistic details linking this complex enzymology and the structural alterations mediating the transition to packaging remain a fertile area of investigation.

The Translocating Motor

In contrast to the maturation complexes discussed above, translocating packaging motors from several systems, including phages $\phi 29$, λ , T4 and P74-26, among others, have been examined in great detail. Several excellent reviews have described the structural and functional features of the motors and only salient points will be discussed here.

The activated packaging motors translocate along the duplex, fueled by ATP hydrolysis, inserting the DNA through the portal and into the procapsid shell. Ensemble biochemical and single molecule studies have shown that the translocating motors are highly processive and very fast. ATP hydrolysis is tightly coupled to motor translocation (mechanochemical coupling) and it is commonly accepted that 2 bp DNA are packaged per ATP hydrolyzed by the terminase enzymes. The motor subunits are tightly coupled and incorporation of a single defective subunit into the motor poisons translocation activity. The motors can generate an astounding ~ 60 pN force and package DNA to near crystalline density within the capsid shell.

Structure of the Translocating Motors

Despite the strict requirement of TerS *in vivo*, structural studies in the T4 and T7 systems suggest that functional headful packaging motors can be assembled as pentamers of TerL subunits assembled at the portal vertex, in the absence of TerS. Pentameric TerL

motors have also been proposed in the P74-26 and D6E systems. Similarly, the phage $\phi 29$ “ATPase” can package DNA without the requirement of a TerS-like subunit, though there remains debate as to the stoichiometry of the $\phi 29$ motor complex with pentamer and hexamer motors proposed based on structural and biochemical data, respectively. In contrast, we have proposed that the functional λ motor is composed of four protomers, which includes the TerS subunit ([TerL•TerS₂]₄). This discrepancy may reflect inherent differences between headful and unit-length genome packaging motors. Alternatively, it is feasible that a fifth protomer is recruited to the tetrameric λ maturation complex at some point during symmetry resolution, procapsid binding or *cos*-clearance steps to afford a pentameric motor (see Figs. 5 and 6). Then again, it is feasible that the λ maturation complex is, in fact, a pentamer of protomers contrary to published biophysical data. We disfavor this model because while there is precedence for tetrameric endonuclease enzymes, there is to our knowledge no precedent for a nuclease complex composed of five catalytic subunits. Notwithstanding, we cannot rigorously exclude these possibilities and the physical nature of the λ maturation and motor complexes must await further structural interrogation.

Coordinated ATP Hydrolysis by the Terminase Motors

Single molecule studies of phage $\phi 29$ and λ motors suggest that non-cooperative ATP binding appears to be associated with motor translocation (Hill coefficient, $n \sim 1$). This observation is a central feature of a mechanistic model describing mechanochemical coupling by the motors, based high resolution data in the $\phi 29$ system, as follows. A pentameric motor is assumed and ATP binds to each of the subunits in a dwell phase. This is followed by a sequential but tightly coordinated hydrolysis of four ATP's, which is proposed to “mask” cooperativity in ATP binding and/or hydrolysis *when measuring motor translocation*. Each hydrolytic event is coupled to the translocation of 2.5 bp into the capsid, for a total of 10 bp per motor cycle; the fifth motor subunit is postulated to “maintain a grip” on the duplex during each catalytic cycle.

In contrast to the single molecule studies, kinetic studies of solution-based steady-state ATP *hydrolysis* by the phage P74-26 TerL subunit (under non-packaging conditions) exhibits modest cooperativity (Hill coefficient, $n = 1.7$). Similarly, solution-based ensemble studies have characterized ATP hydrolysis by the translocating λ motor, composed of both TerL and TerS subunits, which reveal strong cooperativity in ATP hydrolysis ($n = 3.8$). How can these observations be reconciled with the apparent non-cooperative ATP binding events observed in the single molecule studies?

A critical distinction between the single molecule and solution-based ensemble biochemical studies is that the former does not actually measure ATP *hydrolysis*. Rather, the single molecule studies examine *motor translocation* as a function of ATP concentration. Thus, the discrepancy may reflect what is actually observed by the two approaches. Alternatively, it is feasible that the $\phi 29$ motor, which packages genome monomers, is mechanistically distinct from those that package DNA from a genome concatemer. Resolution of this question must await high-resolution single molecule studies of additional terminase enzymes, such as λ , T4 and P74-26, among others, to verify the results observed in the $\phi 29$ system.

Mechanochemical Coupling and DNA Translocation

While extensively studied, the molecular mechanism coupling ATP hydrolysis to duplex translocation remains speculative and several models have been proposed. These include an “inchworm” translocation model, similar to those describing translocation by monomeric and dimeric helicase enzymes. In this model, the TerL nuclease and ATPase domains both bind DNA and conformational switching between a “tense” and “relaxed” state, driven by ATP hydrolysis; this mediates sequential steps of DNA binding and release by the two domains, which drives translocation. A second “lever” hypothesis proposes that duplex translocation is mediated solely by the ATPase domain, which grips and translocates DNA in the central channel of a ring-like motor complex. In this model, ATP-hydrolysis alters the conformation of the lid subdomain, which is propagated with a “lever-like” motion to the adjacent subunit in the motor, driving duplex movement. Within this context, recent studies in the λ system suggest that TerL undergoes a “tight binding transition” upon ATP binding that is required for both tight ATP binding and tight gripping of the DNA. Further, molecular dynamics studies suggest a general terminase mechanism wherein the conserved ASCE catalytic glutamate and a conserved WA arginine act in concert to facilitate this “open” to “closed” transition, which juxtaposes the catalytic glutamate and a water molecule proximate to the γ -phosphate of ATP for catalysis. After hydrolysis and P_i release, the arginine “toggles” to interact with a different glutamate in the lid subdomain to communicate the nucleotide-bound state, which has been implicated in mechanochemical coupling. Analogous models have been proposed for the $\phi 29$, T4 and P74-26 packaging motors based on single-molecule and biochemical data.

The above models presume that the DNA is an “inert” substrate during translocation; however, other models suggest that the DNA itself plays a direct energetic role in duplex translocation. A “crunching” model proposes the motor compresses B-form DNA into a shorter A-conformation, mediated by ATP hydrolysis. Subsequently decompression back to a longer, B-form duplex propels the DNA into the shell. A related “scrunching” model proposes that the motor proteins repeatedly dehydrate and rehydrate sections of DNA, which undergo a cyclic shortening and lengthening transition. In both these models, the motor, which includes both terminase and portal proteins, capture the shortening-lengthening transitions to bias movement of the duplex into the shell interior. Finally, a novel model invokes an “osmotic pump” mechanism wherein expansion/contraction cycle of the capsid, mediated by ATP hydrolysis, “sucks” the duplex into the shell.

Given that the mechanism of coupled ATP hydrolysis and DNA translocation in terminases remains vague, few of the above models are mutually exclusive.

Termination of Packaging

Upon insertion of a full-length genome into the capsid, the translocating motor stops, cuts the duplex and ejects the stable terminase•concatemer complex, shortened by one genome length, from the genome-filled nucleocapsid. Minimally, this requires downregulation of the packaging ATPase site (to avoid motor movement) and re-activating the nuclease activity of TerL. Further, the parallel orientation of TerL subunits in the translocation motor must switch back to a head-to-head conformation (symmetry resolution), which is the presumed conformation for a catalytically competent nuclease complex (*vide supra*). As with all ill-characterized processes, several models have been proposed for this conserved process.

The first, based on structural data in phage T4, is that the nuclease catalytic site of TerL is separate and distinct from the DNA binding site involved in duplex translocation. Upon packaging the genome, the duplex is transferred from the translocation binding site to the nuclease active site. Another model, based on structural data in the P74-26 system, proposes that the TerL nuclease domain is bound to the portal ring during packaging and is sequestered from the DNA. Upon insertion of the full genome length, the nuclease domains release the portal and bind to the duplex to catalyze cleavage. Both of these models suggest that nuclease activity is “regulated” by sequestration of the catalytic site from the DNA during translocation. This begs the question as to what regulates the conformational changes that allow duplex access to the nuclease site.

Single molecule studies have demonstrated a dramatic slowing of the packaging motors at high packaging densities. The simplest model is that motor slowing provides sufficient time for a slow conformational change that allows duplex transfer to the nuclease site. This kinetic model is analogous to DNA strand switching from the polymerase site to the exonuclease editing site in DNA polymerase enzymes. The model predicts that duplex cutting would occur whenever the translocating motor pauses; however, a number of studies do not show evidence for cutting by a stalled motor. Thus, the termination signals appear more complicated and alternative models propose that shell filling density is directly communicated from the nucleocapsid to the translocating motor. These models suggest that the portal serves as a “packaging sensor” that allosterically activates TerL nuclease activity to allow for headful cleavage.

Unit Length Packaging Motors

The situation is more complicated with viruses that package unit-length genomes, such as λ , HK97 and $\phi 80$, among others. In these cases, upon arrival at *cos* and with attenuated motor velocity, the motor subunits must specifically engage the *cos*-sequence prior to nuclease activation (*cos*-capture). Again, phage λ provides a useful model to understand *cos*-capture and the terminal *cos*-cleavage reaction, as depicted in Fig. 7.

Upon arrival at the terminal *cos*, ATPase activity must be downregulated and nuclease activity must be activated in the switch from a dynamic motor to a stable nicking complex. Nuclease activation is likely concomitant with resolution of the motor symmetry to again bind to a symmetric *cosN* nicking site. Specifically, the “parallel” motor complex must switch back to a “head-to-head” complex symmetrically disposed at *cosN* for duplex nicking. We presume that the regulatory ATP binding site in λ TerS (Fig. 2) is involved in this transition, but data is limited.

Feiss and co-workers have proposed that specific elements within the *cos* sequence also play important roles in terminal symmetry resolution and *cos*-capture steps, as follows. Genetic and biochemical studies have identified a *cosQ* element upstream of *cosN* and an I2 IHF binding element between *cosN* and *cosB*, that are essential for terminal *cos*-capture (Fig. 7). When mutations are introduced into *cosQ*, the translocating motor nicks the top strand of *cosN*, but not the bottom strand and the motor continues to package DNA until the capsid is filled to capacity. These mutations are lethal because uncut DNA protrudes from the portal and tails cannot be attached. The observation that only the top strand is nicked in the presence of *cosQ* mutations suggest that this element is required for symmetry resolution steps in the transition from a parallel motor complex to a head-to-head nuclease complex capable of symmetric nicking of the duplex.

Genetic characterization of *cosQ* mutations reveal three modes of suppression; (1) compensating mutations in *cosQ*, (2) an increase in genome length (4–5 kb), which requires IHF and (3) missense mutations within the portal gene. It has been proposed that genome length and portal mutations serve to slow the translocating motor and increase the efficiency of *cosQ*-*cosN*-I2 capture. In this model, *cosQ* serves as a molecular “speed bump”. This is consistent with the observation that unit-length terminase enzymes retain a “headful” component to packaging termination; the efficiency of cleavage at the terminal *cos* sequence in λ decreases from 100% to 75% as the genome is shortened to ~80%. Further, insertion of additional *cos* sequences upstream of the natural termination site do not induce premature nuclease activity and the aggregate data indicate that *cos*-capture requires specific signals in the DNA plus sufficient capsid filling densities. Genetic studies further suggest that the portal may serve as a packaging sensor, transmitting information on packaging density to the TerL subunits in the translocating motor, as has been proposed for the headful packaging motors. Thus, the termination signals include *cosQ*-*cosN*-I2 elements that are likely recognized by the translocating portal-terminase complex and IHF and by attenuated motor velocity.

Terminase Ejection and Virion Completion

Once terminase has nicked the strands within the terminal *cosN* element, the strand separation activity of terminase is poised to release the DNA-filled nucleocapsid and regenerate the maturation complex, as depicted in Fig. 7. This begs the question as

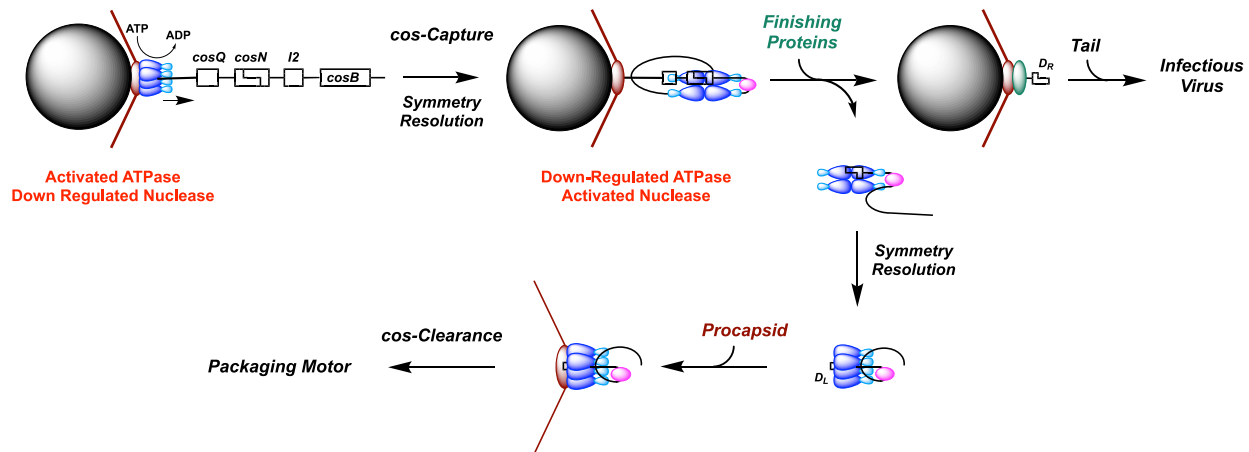


Fig. 7 *cos*-capture and the terminal cleavage reaction. The translocating λ motor slows at high packaging densities to allow terminase binding to *cosQ-cosN* elements, presumably enhanced by IHF binding to the *I2* element. This requires symmetry resolution of the TerL subunits from a parallel orientation in the motor to a head-to-head orientation in the activated nuclease complex. Finishing proteins likely displace the terminase-DNA complex from the nucleocapsid, which affords an infectious virus upon addition of a pre-assembled tail. Note that the matured genome end (D_R) extends from the capsid into the tail tube of the assembled viral particle. The ejected terminase•DNA concatemer complex, formally equivalent to the maturation complex that initiated genome packaging, again undergoes a symmetry resolution transition and binds another procapsid to initiate a second round of DNA packaging.

to how the highly-pressurized packaged DNA is retained within the shell subsequent to terminase release. One model proposes that the portal ring binds tightly to the duplex as a “one-way valve” to prevent release. Additionally, it is feasible that “finishing proteins” that add to the portal of a DNA-filled shell actually play a direct role in ejection of the terminase•concatemer complex, simultaneously stabilizing the DNA-filled shell. Whatever the case, the ejected terminase•concatemer complex, formally equivalent to the complex that initiated genome packaging, again binds to an empty procapsid shell to initiate the next round of processive packaging. In this manner, terminase alternates between a stable maturation complex and a dynamic motor complex.

Conclusion

Herein I have proposed specific molecular events in phage λ required for the site-specific assembly and activation of a stable maturation complex and its transition to a dynamic packaging motor including “symmetry resolution”, procapsid binding and *cos*-clearance steps. We discuss current models that describe motor translocation and mechanochemical coupling of ATP hydrolysis and complex movement. Finally, we discuss the features and mechanistic complications involved in the terminal *cos*-cleavage reaction, including the signals required for *cos*-capture, symmetry resolution and activation of nuclease activity. The terminase enzymes must sequentially cycle between these two complexes, stable nuclease and dynamic motor, to allow processive genome packaging from a concatemeric DNA substrate. The proposed models derive from extensive genetic, kinetic, biochemical and biophysical data obtained over decades of study in our lab and by others. Kinetic and thermodynamic interrogation of λ terminase figure prominently in the development of the models and complement the structural and single-molecule data published in λ and many other systems. Analogous assembly steps and catalytic transitions must occur in all of the dsDNA viruses that package monomeric genomes from concatemeric precursors and while details of these processes may differ, it is likely that the general features proposed for λ are recapitulated in all of the terminase family members. Thus, continued mechanistic interrogation of the packaging complexes using integrated genetic, structural, biophysical, enzymological and computational approaches is warranted. Indeed, the availability of multiple packaging systems including λ , T4, SPP1, P74-26, SF6, among others, provides an attractive pallet from which to choose and define conserved features and system-specific mechanisms of packaging amongst the terminase enzymes.

Further Reading

- Black, L.W., 2015. Old, new, and widely true: The bacteriophage T4 DNA packaging mechanism. *Virology* 479–480, 650–656.
- Casjens, S.R., 2011. The DNA-packaging nanomotor of tailed bacteriophages. *Nature Reviews Microbiology* 9 (9), 647–657.
- Casjens, S.R., Hendrix, R.W., 2015. Bacteriophage lambda: Early pioneer and still relevant. *Virology* 0, 310–330.
- Catalano, C.E., 2005. Viral genome packaging machines: An overview. In: Catalano, C.E. (Ed.), *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. New York, NY: Kluwer Academic/Plenum Publishers, pp. 1–4.
- Cuervo, A., Daudén, M.I., Carrascosa, J.L., 2013. Nucleic acid packaging in viruses. In: *Subcellular Biochemistry* 68, pp. 361–394.

- Oliveira, L., Tavares, P., Alonso, J.C., 2013. Headful DNA packaging: Bacteriophage SPP1 as a model system. *Virus Research* 173 (2), 247–259.
- Ortiz, D., delToro, D., Ordyan, M., *et al.*, 2018. Concerted Action of ASCE glutamate and arginine residues in catalysis of ATP hydrolysis in the phage lambda DNA packaging motor. *Nucleic Acids Research* 47 (3), 1404–1415.
- Rao, V.B., Feiss, M., 2015. Mechanisms of DNA packaging by large double-stranded DNA viruses. *Annual Review of Virology* 2 (1), 351–378.
- Sharp, K.A., Lu, X.J., Cingolani, G., Harvey, S.C., 2019. DNA conformational changes play a force-generating role during bacteriophage genome packaging. *Biophysical Journal* 116 (11), 2172–2180.

DNA Packaging: DNA Recognition

Sandra J Greive and Oliver W Bayfield, University of York, York, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

A Adenine

aa Amino acid

ATP Adenine triphosphate

bp Base pairs

C Cytosine

Cos Cohesive end sequence

DBD DNA binding domain

DNA Deoxy ribonucleic acid

ds Double stranded

E. coli *Escherichia coli*

G Guanosine

gp Gene product

HTH Helix-turn-helix

Pac Packaging sequence

RNA Ribonucleic acid

ss Single stranded

T Thymidine

TerL Large terminase protein

TerS Small terminase protein

TP Terminal protein

TR Terminally redundant

wHTH Winged Helix-turn-helix

Glossary

Affinity Description of the interaction between two components of a binary complex that can be qualitatively referred to as either strong/tight or weak. Usually this is defined quantitatively as the equilibrium dissociation constant.

Avidity The increase in affinity that occurs as each new binary interaction is made between additional components of the final ternary complex.

Capsid or phage head The protein shell that is preassembled from the major capsid protein into which the viral genome is packaged. In non-enveloped viruses this forms the outside of the virus and eventually acts as the protective container for transfer of the viral genome to a new host.

Concatemer The nascent dsDNA product from rolling circle replication of viral genomic DNA that is a continuous dsDNA strand with multiple copies of the genome fused head to tail.

Conformational recognition The situation where assembly of a protein-DNA complex is facilitated by electrostatic attraction between oppositely charged surfaces that have complementary shapes, and the ability of the target DNA sequence to form the matching shape. Since contact between the protein and DNA base pairs are not

directly sequence identifying, this type of binding is an indirect readout of the underlying sequence.

Genome maturation The process where the concatemeric viral DNA is recognized and cut to create the free DNA end that is key for insertion into the capsid pore at the start of packaging.

Headful termination The DNA sequence independent termination of the DNA translocation reaction that occurs once the capsid reaches maximum DNA capacity. The DNA stalled within the molecular motor is cut by an endonuclease domain of the motor protein. The entire motor assembly (DNA bound to ATPase) is then transferred to a new empty capsid and the packaging process is repeated.

Packaging The process where the nascent genomic DNA is pumped into an empty capsid shell to form a new infectious virus particle or virion.

Replication The process in which new copies of the viral genome are synthesized for the creation of new virus particles.

Sequence specific binding The binding interaction between a protein and a defined sequence of DNA, mediated by specific hydrogen bonding contacts between the side chains of the recognition amino acids on the binding surface and the edges of the individual base pairs in the DNA sequence.

Introduction

Bacteriophages repurpose the host cell's machinery to make the specialized molecular machines and macromolecular complexes required to produce new virus particles: a copy of the viral genome encased in a protective protein container known as a capsid or phage head. These new viruses are then released and can infect other bacteria. The nascent copies of the bacteriophage genome produced during replication need to be pumped, by a powerful DNA translocation motor, into pre-assembled capsids through a specialized pore. The packaged DNA forms concentric loops of decreasing diameter around the inside of the capsid until the strands of DNA are tightly packed against each other. As the phage head fills, the translocation motor continues to force DNA into the capsid, driving the ordered bending and stacking of the DNA against the massive physical forces that normally prevent such compaction. Indeed, the pressure of the restrained DNA inside the packaged virus is 10 times greater than the pressure contained

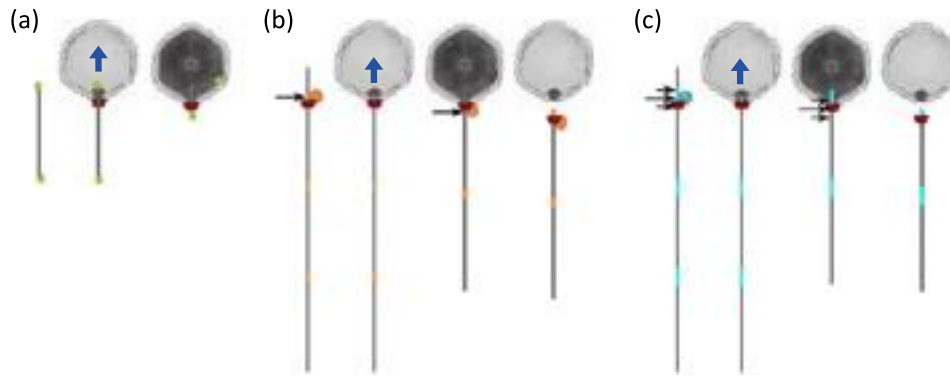


Fig. 1 Overview of dsDNA bacteriophage genome packaging systems. The phage genome is assembled with a powerful DNA translocation motor (red) on a unique pore (portal or connector protein; gray) in the preformed protein capsid. ATP hydrolysis powers the macromolecular machine that pumps the viral genomic DNA into the protective viral shell in a process known as packaging. (A) Protein primed replication of linear dsDNA genomes produce (e.g., *Bacillus* phage ϕ 29 and *Tectiviridae* e.g., PRD1) create monomeric dsDNA products fused to terminal protein (green). (B, C) Rolling circle replication of a circularized template intermediate yield long polymeric concatemers of dsDNA genomes, fused head to tail, that need to be cut (arrows) to create the free DNA end required for packaging into empty capsids. Once the capsid is filled, the DNA is cleaved again and the motor-DNA complex is transferred to a new empty capsid and the process is repeated. (B) Packaging of unit length genomes mediated by *cos* sequences (*E. coli* phage λ or HK97; orange) or terminal repeat sequences (*E. coli* phage T3, T5, T7). (C) The headful packaging mechanism is initiated by a cleavage at the *pac* site (cyan) and terminated at non-specific sequences after ~104% of the genome has been packaged (arrows), producing virions containing slightly different circularly permuted genomes.

in a champagne bottle. This DNA packaging process is started when one end of the viral genome is assembled into the ATP hydrolysis powered DNA translocation motor located on the capsid pore (known as either portal or connector, Fig. 1).

Like all cells, the inside of a bacteria cell is a densely crowded dynamic environment, in which the genomic DNA is stored in ordered structures. The macromolecular machinery that rearranges, copies, repairs and transcribes the information coded within the genome are integrated within this superstructure. Additionally, bacteria also contain other large biomachines, such as ribosomes, proteasomes, and the complexes that drive metabolism, all of which are mixed together with individual protein molecules, nutrients and metabolites into a molecular stew with the consistency of a gummy bear candy. Virus replication and packaging occurs within this complex environment. Consider the case of an average *E. coli* cell in exponential growth, containing: ~2 copies of its own DNA genome; a high copy number average sized plasmid DNA ~100 copies of 5 thousand base pairs (bp); and up to 300 new copies of a replicating phage DNA genome (~50 thousand bp); equating to a total DNA composition of ~5 million base pairs. How does an infecting bacteriophage recognize and selectively package its own genomic DNA?

Genome Replication and Selection Mechanisms

The problem of preferential packaging of phage genomic DNA, rather than host genomic or plasmid DNA, can be overcome through several different methods. While some bacteriophages (e.g., *Pseudomonas chlororaphis* phage 201 ϕ 2-1) keep their replication machinery and the newly synthesized genomes separate from the host cell genome within a virally encoded protein compartment, most rely on some form of selective recognition between host or viral DNA and virally encoded proteins. For example, *E. coli* phages T4, T3, and T7 produce nuclease enzymes that target and degrade the host cell DNA, leaving only the nascent viral DNA to be packaged. However, the most common mechanism uses dedicated viral proteins that selectively bind to bacteriophage genomic DNA, and then direct the assembly of the packing motor-DNA complex, onto the portal of the capsid, to initiate DNA packaging. Since double stranded (ds) DNA bacteriophages use different replication systems to make new copies of their genomic DNA and thus produce diverse dsDNA products, specialized mechanisms are required for specific recognition of these replication products in order to initiate packaging.

Monomeric replication products

Bacteriophages that use protein-primed replication (*Bacillus* phage ϕ 29 and *Tectiviridae* e.g., PRD1) produce linear monomeric dsDNA genomes. The products of this type of replication are most obviously defined by the presence of a protein covalently attached to the terminal 5' phosphate at each end of the genomic dsDNA. This terminal protein (TP) remains attached to the nascent DNA strand after acting as the protein primer for initiation of strand synthesis by the bacteriophage DNA polymerase (Fig. 1(A)).

Concatemeric replication products

New copies of most linear bacteriophage dsDNA genomes are produced through rolling circle replication of a circularized template intermediate (e.g., *E. coli* λ and T3 bacteriophages). This results in a concatemeric genome product: multiple copies of the genome, joined head to tail in a single long dsDNA strand. In a process known as genome maturation, these fused copies must be separated

at the start of each individual genomic unit, in order to create the nascent DNA end that is key for loading into the packaging motor. Since many of these viral packaging systems show no genome preference when provided with a free dsDNA end *in vitro*, the protein participants that generate the nascent DNA end, play an important role in genome selectivity. Several different methods for achieving this genome maturation process have been noted.

Some phages package exactly one genome unit length defined by cohesive end site (*cos*) sequences (*E. coli* phage λ or HK97) or packaging (*pac*) sequence encoded direct terminal repeat (*E. coli* phage T3 and T7) sequences. Other phages package circularly permuted genomes (e.g., *Bacillus* phages SF6 and SPP1, and *Salmonella* phage P22), with the initial cleavage defined by a *pac* sequence site, resulting in packaging of greater than 1 genome unit length. After the initiation event at the *pac* site, when at least one genome length of the concatemeric DNA has been pumped into the phage head, DNA translocation is terminated by another cleavage event and the motor-DNA end complex is transferred to a new empty capsid to begin the packaging process again (Fig. 1(B)). This secondary, 'motor transfer' mode of packaging initiation accounts for at least two thirds of all packaging events in bacteriophages that use rolling circle replication.

Bacteriophages that package unit length genomes, such as λ or HK97, terminate DNA translocation and cut the DNA at the same specific sequence motif (*cos*) in the next adjacent genome. However, viruses that package more than one genome length (~104%), terminate packaging at non-sequence specific sites using a 'headful' mechanism (Fig. 1(C)). Headful cleavage is presumably mediated by the deceleration of the packaging rate that results from the mounting internal pressures that are induced as the capsid approaches maximal DNA capacity. Indeed, this headful signal for termination of translocation appears to be an underlying physicochemical property of dsDNA bacteriophage packaging systems, as it has also been observed in mutant versions of viruses that usually package unit length genomes.

Despite these differences, in each case, initial genome end processing and motor assembly requires specific recognition of the viral genome by virally encoded proteins. This selective binding relies on the physical properties of the targeted DNA: the average base composition and/or modifications, local and global variations in shape, and the presence of specific sequence motifs.

The Shape of DNA

The helical architecture of dsDNA is stabilized by both the hydrogen bonds formed between the complementary Thymidine: Adenine (A:T) and Guanosine:Cytosine (G:C) base pairs at the center of the two sugar-phosphate backbone chains, and the helical stacking interactions between adjacent bases along each strand (Fig. 2(A)). Double stranded DNA is a conformationally dynamic molecule that can exist in many different shapes or forms. The most commonly adopted shape is the 'B-form' that has an average of 10.5 bp per helical turn, creating wide major and narrow minor grooves of equal depths. Since the minor groove is on the sugar-phosphate side of the bases where access to the bases is more sterically hindered, sequence specific interactions occur in the major groove. Additionally, the minor groove of ideal B-form DNA is also comparatively more electrostatically negative than the major groove. While B-form DNA is the most prevalent form within cells, DNA conformation is both dynamic, fluctuating locally between conformational variants, and dependent on sequence, so that some sequence patterns are more likely to enter different conformations. For instance, A-rich sequences form non-ideal B-form DNA with more narrow minor grooves, while GC-rich DNA sequences are often found in 'A-form' shapes (Fig. 2(A)). A-form DNA is shorter and wider, with a looser helical turn (11 bp per turn) creating a deep narrow major groove and a wide shallow minor groove. The extent of hydration of DNA, and the resulting hydrogen bonds formed with water molecules along the phosphate backbone and within the major and minor grooves, is crucial in determining these conformations. Indeed, dehydration of B-form DNA sees a transition to A-form. Further conformational variants are formed by weakening of the hydrogen bonds between base pairs (breathing or melting) resulting in the formation of single stranded (ss) regions such as found in bulges, bubbles or loops. Perturbation of the base stacking interactions and rotation of the backbone strand allows bases to 'flip out' of the helical structure. These variants can be further stabilised and/or induced by local DNA base sequence or protein binding interactions. Over longer distances, such sequence induced local conformational changes in the DNA helix can contribute to curvature of the DNA helix over a single basepair (kinking) or several base pairs (bending) (Fig. 2(B)), conformations which can be sharply bent by the binding of histone-like proteins in bacteria (Fig. 2(C)). When multiple repeats of these sequences occur in defined patterns, the DNA can be curved into circular or superhelical structures, as observed in nucleosome core particles (Fig. 2(D)). These conformations are also somewhat dependent on the base sequence composition, as well as the binding of counter-ions or proteins. Globally, the action of macromolecular machines that separate the DNA strands during RNA transcription or DNA replication, cause the helical dsDNA to become supercoiled (Fig. 2(E)). Positive supercoils, where the DNA is overwound and less likely to melt into individual strands, are formed in front of the advancing enzyme complex, while negative supercoils (underwound DNA with a higher propensity for melting) are formed behind it. Negatively supercoiled DNA is thought to exist as Z-form DNA (Fig. 2(A)), where the phosphate backbone has a zigzag conformation and the stacked bases are exposed towards the outer edge of the fiber. These differently shaped DNA structures can also be specifically bound by specialized proteins using conformational recognition.

Double stranded DNA is a surprisingly versatile information storage device. At its most basic, it contains the ordered sequence of bases that comprise the code, or blue print, for each of the RNA and protein components of the molecular machinery that perform the processes necessary for organism survival. Overlaid around, and sometimes over, this basic code are the sequence patterns, or motifs, that act as signals that are preferentially bound by the protein complexes that copy or transcribe the basic code, or control these processes. Redundancy in the code leads to variation in the average base composition between organisms, and the

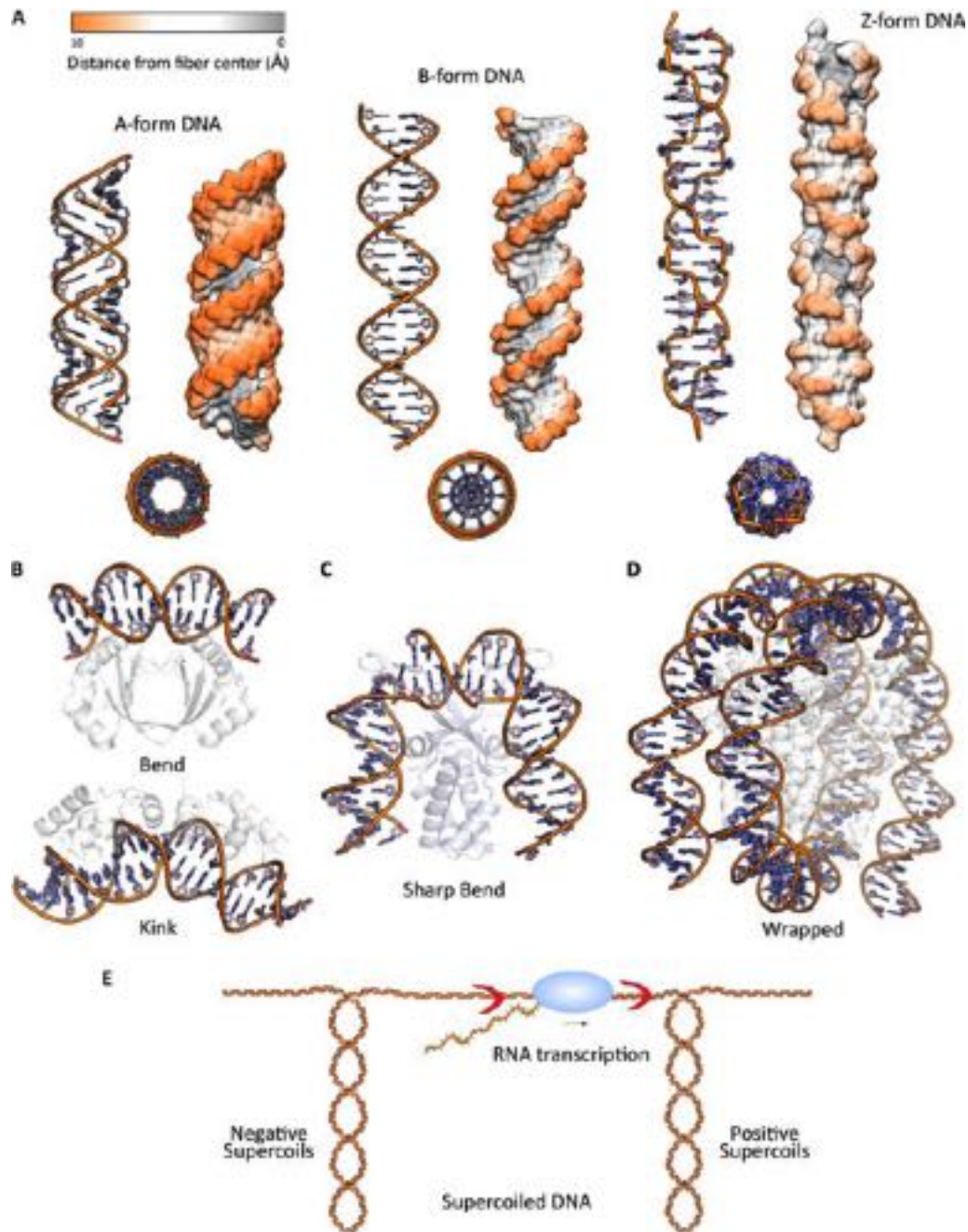


Fig. 2 The conformational variation of DNA. (A) Cartoon depiction of the DNA double helix in A-, B- and Z-form conformations (top panel), showing the sugar-phosphate backbone (orange ribbon; and gray rings), base pairing between complementary bases (blue/white/red shaded rings) and the helical stacking interactions between bases along the same strand (bottom panel). Surface cartoons show the distance from the center of the DNA helix in a gradient colored from dark gray (center) to orange (10 Å from helix center). Models were created from fiber diffraction data using 3DNA (x3DNA.org). Cartoon depiction of: (B) bending (PDB 1LMB) or kinking (PDB 1JJ4) of dsDNA stabilized by proteins; (C) sharply bent DNA bound to Integration Host Factor (PDB 1IHF); and (D) tightly coiled circular DNA in nucleosome core particles (PDB 3LZ0). (E) Representation of positive and negative DNA (orange) supercoiling that occurs during transcription by RNA polymerase (blue).

over- or under- representation of certain patterns, the bases of which are sometimes specifically targeted for chemical modification (e.g., methylation) by dedicated proteins. The conformational repertoire of helical dsDNA introduced by particular sequences of base pairs, as well as the contribution to larger shape variations by repeating patterns of such sequences, can be considered as an additional 'conformational code' that can be specifically acted upon by protein cofactors with complementary shaped binding surfaces. Combination of sequence specific binding with conformational recognition can significantly expand the protein binding code available to a DNA sequence.

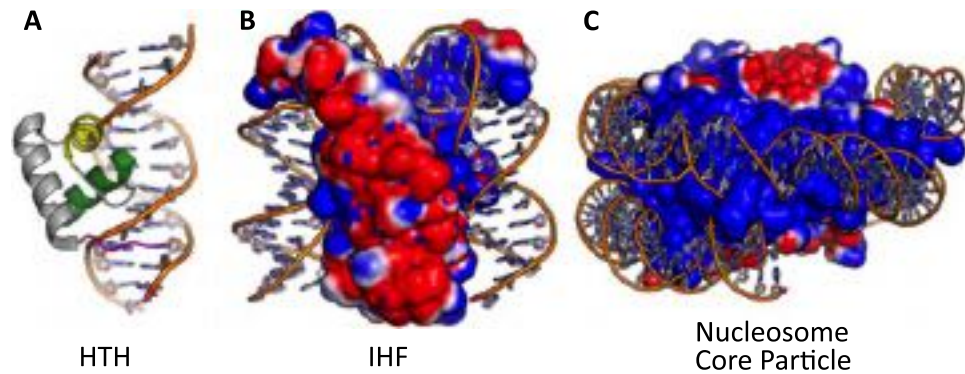


Fig. 3 Examples of preferential DNA-protein interactions. (A) Cartoon image of sequence specific binding to B-form DNA by a Helix-Turn-Helix (HTH) motif (PDB 1WOT), showing the recognition helix (green), the support helix (yellow) and an example of an arginine rich N-terminal arm (magenta). (B,C) Image showing modes of conformational recognition using positively charged protein surfaces with complementary shapes to the bound DNA. The electrostatic surface potential for each protein is depicted (positive, blue; and negative, red). (B) *E. coli* Integration Host Factor (IHF) bound to H-box DNA, and (C) a nucleosome core particle bound to bent DNA. In all frames the phosphate backbone (orange strands), ribose sugars (gray rings) and nucleotide bases (gray, red, blue rings) of the DNA are shown.

Protein-DNA Binding

A Question of Fidelity: Sequence Specific and Non-Specific Binding

Protein binding interactions with nucleic acid molecules are considered sequence-specific when they are stabilised by hydrogen bonding networks or hydrophobic interactions between the amino acid side chains of the protein and the individual base pairs of the DNA sequence motif. This is also known as direct readout. Additional non-sequence specific binding interactions occur around the sequence specific contacts. These largely electrostatic connections between the positively charged amino acids forming the binding surface of the protein and the negatively charged sugar-phosphate backbone of the DNA, serve to stabilize the bound complex. Indeed, the comparatively weak non-sequence specific binding interactions can play an important role in localizing the protein to the DNA, allowing the protein to slide randomly along the dsDNA in a linear search for a sequence specific binding motif. Conformational changes in both the protein and the DNA allow direct contacts to be made with the bases of the binding sequence motif, increasing the affinity of the protein for the DNA at this sequence. Despite not making identifying contacts with the base pairs in the major groove, the overall shape of the positively charged surface of the protein can also play a role in selective binding to segments of DNA with a potential for matching this shape. This indirect readout mode has been noted in the sharply bent or superhelical protein-DNA complexes of genome architectural proteins.

DNA Binding Domains

Many proteins interact with DNA through discrete domains, known as DNA binding domains (DBDs). These are often folded independently of the main protein body and can be swapped between proteins or studied in isolation. DBDs are found in different shapes, also known as folds or structural motifs, depending on the functional requirements of the core protein or complex. DBDs that interact with specific DNA sequence patterns have folds that allow placement of the binding surface so that direct contacts can be made between the residues of this surface and the individual bases that comprise the binding sequence. These include: Helix-Turn-Helix (HTH); Helix-Loop-Helix; Zinc fingers and Leucine zippers, among others. Additional non-sequence specific contacts are often made by extensions to the DBD that are not always considered part of the core DBD motif. Proteins that use a conformational recognition mode for binding to targeted DNA sequences have core structures that place the DNA binding features in a defined geometry to allow non-sequence specific contacts with the sugar-phosphate backbone or the minor groove of DNA of a distinct shape, such as the histone complexes of eukaryotes and the histone-like proteins of bacteria.

HTH motifs are highly conserved and widespread throughout nature, including in bacteriophages. The core domain is usually presented on a small helical bundle and consists of 2 α -helices, connected by a turn, and held tightly together in a cross shape by conserved hydrophobic contacts between the helices. Since an α -helix is just wide enough to lie along the major groove of B-form DNA when aligned roughly parallel to the backbone strands, the recognition helix (usually the second in the HTH) affords a compact scaffold from which to project the combination of amino acid side chains that will contact the base pairs within the major groove and bind specifically to the target sequence. The amino acid residues around this helix assist with correctly positioning the base specific contacts through additional non-sequence specific contacts with the dsDNA backbone. Common extensions to this core fold, such as N-terminal arms and/or an anti-parallel β -sheet wings, such as in the winged HTH (wHTH), make electrostatically mediated contacts with the sugar-phosphate backbone and/or the minor groove (Fig. 3(A)).

Conformational recognition, such as used by genome architecture proteins, uses larger protein assemblies with surfaces of defined geometry and patterns of positive charge to capture and stabilize transient DNA shapes. For instance, histone-like proteins

(e.g., Integration host factor, IHF) of bacteria bind as dimers in a non-sequence specific manner to DNA that has a propensity for forming sharp bends, stabilizing the bend and assisting in wrapping and condensing the DNA chromatid. The protein dimer creates a positively charged, saddle-shaped β -sheet surface that is suspended between two flexible extended β -ribbon arms that lie along the minor groove when bound to DNA. The tips of the arms display conserved proline residues that stabilize the kinked DNA conformation by projecting between two stacked T bases and distorting the helical assembly of base pairs. IHF binds preferentially to DNA sequences in a conformational recognition mode where the spacing between short sequences of conserved base pairs is required for placing narrowed minor grooves at the correct position for interaction with distinct features of the protein (arginine residues along the arms). Electrostatic interactions between the bent DNA strands and the saddle groove over the top of the protein and down the side of the protein further stabilize the bent conformation (Fig. 3(B)). In Eukaryotes, the histone octamers in nucleosome core particles form a somewhat asymmetric disk-like shape that is highly positively charged around the circumference so that ~ 140 bp of DNA wraps twice around the disk to form a superhelical structure. The diameter of the core is such that it selectively binds to DNA with repeating patterns of sequences (A-tracts followed by CA bases) that are kinked, stretched and bent into a tightly wrapped helical coil (Fig. 3(C)). These shapes are further stabilised by precise placement of electrostatic arginine contacts in regions of narrowed minor grooves.

The general properties of these DBDs that modulate interaction with their target DNA sequences are exploited and refined, both in phage encoded proteins and the target DNA sequences, to ensure selective packaging of the viral genomic DNA. Examples from bacteriophage packaging initiation systems are discussed below.

Strength in Numbers: Multiple Binding Events Regulate Complex Assembly

The macromolecular complexes that function on DNA assemble in a cooperative, or synergistic manner, through a network of interactions that form between the different protein components and their individual contacts with the DNA. Consequently, the DNA can be considered as acting both as a substrate, being the functional target of the macromolecular machine, and, as the scaffold on which the functional protein complex is assembled. These multiple interaction events serve to increase the overall affinity, or avidity, of the complex for DNA, above the initial binding level observed for the individual components. Such avidity stems from the apparent increase in the local concentration of the binding components that are tethered together through mutual contacts with other members of the complex. A systematic build-up of weak binding interactions in this modular fashion, not only increases the number of different specific interactions that are possible for a particular genome, but also introduces more flexibility into the range of mechanisms for regulating the formation and disassembly of the functional complexes. For instance, ligand binding and/or chemical processing can trigger conformational changes that either positively or negatively modify the binding affinity of one component for DNA or for other complex components. This modulation of affinity allows the assembly, function and dissociation of the different macromolecular machines to be coordinated efficiently during replication and packaging.

Recognition of Viral Genomic DNA

The combinatorial power of DNA sequence and conformational recognition by specialized proteins is used to ensure selective packaging of bacteriophage genomic dsDNA. This process exploits the basic properties of the different genomic replication products, leading to the variety of specific recognition mechanisms described below.

Monomeric Genomes and Terminal Proteins

In $\phi 29$, the terminal protein (TP or gp3), in complex the viral DNA polymerase, binds preferentially to the end of the dsDNA genome through recognition of the combination of the 6 bp inverted terminal repeat sequence (5'-AAAGTA-3') and the TP covalently attached at the free 5' end during the previous round of replication. The N-terminal 74 amino acids of TP are predicted to form α -helices that have been shown to non-specifically interact with either ss or dsDNA. Presumably this non-sequence specific interaction mediates formation of a loop or lariat structure where the 5'-end bonded TP loops back and binds to the same dsDNA genome molecule. Binding of the gp16 motor assembly (equivalent to the ATPase domain of large terminase) to the TP-lariat loop junction may induce or stabilize supercoiling of the DNA within the lariat loop. The nascent bacteriophage dsDNA genome is negatively supercoiled compared with the host genomic DNA by the action of replication complexes, and the formation of the lariat loop structure. This topology contributes to the conformationally selective binding of the viral genomic DNA to the capsid located connector assembly to make a complex where the DNA is wrapped around the connector. In a motion similar to a rope moving around a pulley, this wrapped DNA may slide around the connector complex until the gp16 bound TP-lariat loop junction reaches the connector and the complete motor assembly is formed at the connector/pRNA pore on the capsid (Fig. 4).

Concatemeric DNA and Terminase Proteins

Packaging of the concatemeric product from rolling circle replication is initiated when the nascent genomic DNA end is produced by a virally encoded endonuclease. This enzyme is recruited to the cleavage site by at least one other viral protein that bind(s)

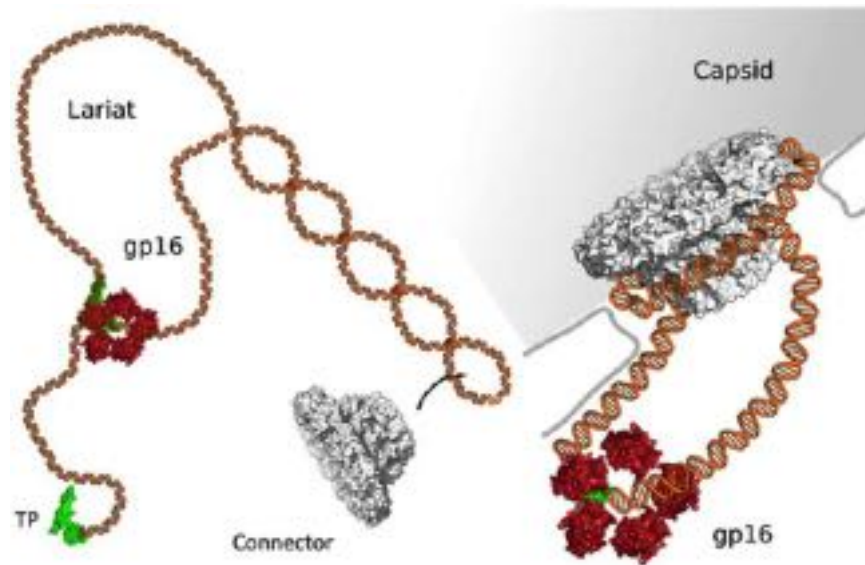


Fig. 4 Model for packaging of monomeric genomic dsDNA. Cartoon model of the features required for specific packaging of ϕ 29 genomic dsDNA (orange). The 5' covalently attached Terminal protein (TP, PDB 2EX3, green) binds non-specifically to an internal segment of the genomic DNA forming a looped lariat structure. The gp16 ATPase motor (red, PDB 5HD9) binds to TP-DNA bound end of the lariat, forming or stabilizing negative supercoils in the DNA within the lariat. The connector protein (gray, PDB 1H5W; assembled into a unique vertex of the capsid, shaded background) binds the negative supercoiled DNA to form a wrapped complex.

specifically to recognition signals in the genomic DNA. These proteins are usually referred to as the terminase, or Ter, proteins, as they help define the termini of packaged DNA. The small terminase protein (TerS) binds specifically to recognition signals that mark the genome end within the concatemeric replication product. This protein-DNA complex is bound and stabilised by the large terminase protein (TerL), which contains the nuclease domain that cleaves the end of the genome as well as the ATPase domain that powers the packaging motor. The terms 'small' and 'large' are in reference to their respective monomeric molecular weights. The nascent DNA end is assembled within a pentameric ring of large terminase subunits at the portal/connector pore complex of the empty capsid, forming the powerful DNA translocation motor that pumps the DNA into the capsid shell. The mechanisms by which the initiating terminase protein complexes of bacteriophages specifically recognize, and assemble, onto their cognate genomes are described below.

Direct terminal repeats: T3 and T7

The packaged linear genomes of T3 and T7 bacteriophages are bounded at each end by terminally redundant (TR) direct repeat sequences of 230 and 160 bp, respectively (Fig. 5(A)). In concatemeric DNA, a single copy of the TR sequence is preceded by the *pac* binding (*pacB*) sequence and bounded at each end by *pac* cleavage (*pacCL* and *pacCR*) sites. The *pacB* site defines phage specific promoter sequences for the cognate viral RNA polymerase, rather than the site of TerS binding, hence coupling packaging initiation to RNA transcription of the TR sequence. Genome end maturation is triggered when an RNA polymerase transcription complex becomes stalled at the concatemer junction (CJ) pause signal downstream of *pacCR*. In a process that is still poorly understood, the corresponding viral terminase proteins (gp18 and gp19) interact directly with the paused RNA polymerase complex in order to assemble at the TR sequence and subsequently nick the dsDNA on opposite strands at *pacCL* and *pacCR*, creating long complementary 5' extensions over TR. However, given that the action of the transcription complex would also produce supercoiling in the dsDNA template, and that the T7 TerS forms an oligomeric complex of ~8 monomers, it is possible that the terminase protein complex also recognizes the global conformation of the DNA surrounding the paused transcription complex to form a stable and functional terminase assembly. The direct repeat sequences of the TR are subsequently generated through an end filling extension reaction, by a DNA polymerase from the 3' end of the nicked strand along each newly formed 5' strand until the second nicked site is reached. The nascent dsDNA end at *pacCR* is then assembled into the macromolecular motor complex for the DNA packaging process.

Cohesive ends: λ -like

In phages λ and 21, the end of the genome is defined by the *cos* sequence; a complex 200 bp region comprised of the *cosQ*, *cosN* and *cosB* sites (Fig. 5(B)). The *cosB* (binding) site is bound specifically by the terminase complex, thought to be a tetramer of protomers, where each protomer is comprised of a heterotrimer of 2 gpNu1 proteins and gpA (TerS and TerL respectively). This proteo-DNA complex directs the nuclease domain of TerL to introduce staggered cuts in the *cosN* (nicking) sequence creating complementary 12 base ss 5' overhangs at the end of each adjacent genome. The *cosQ* containing DNA is freed from the complex, and the remaining DNA end (with the *cosB* site) is assembled into the packaging motor. DNA translocation begins and continues

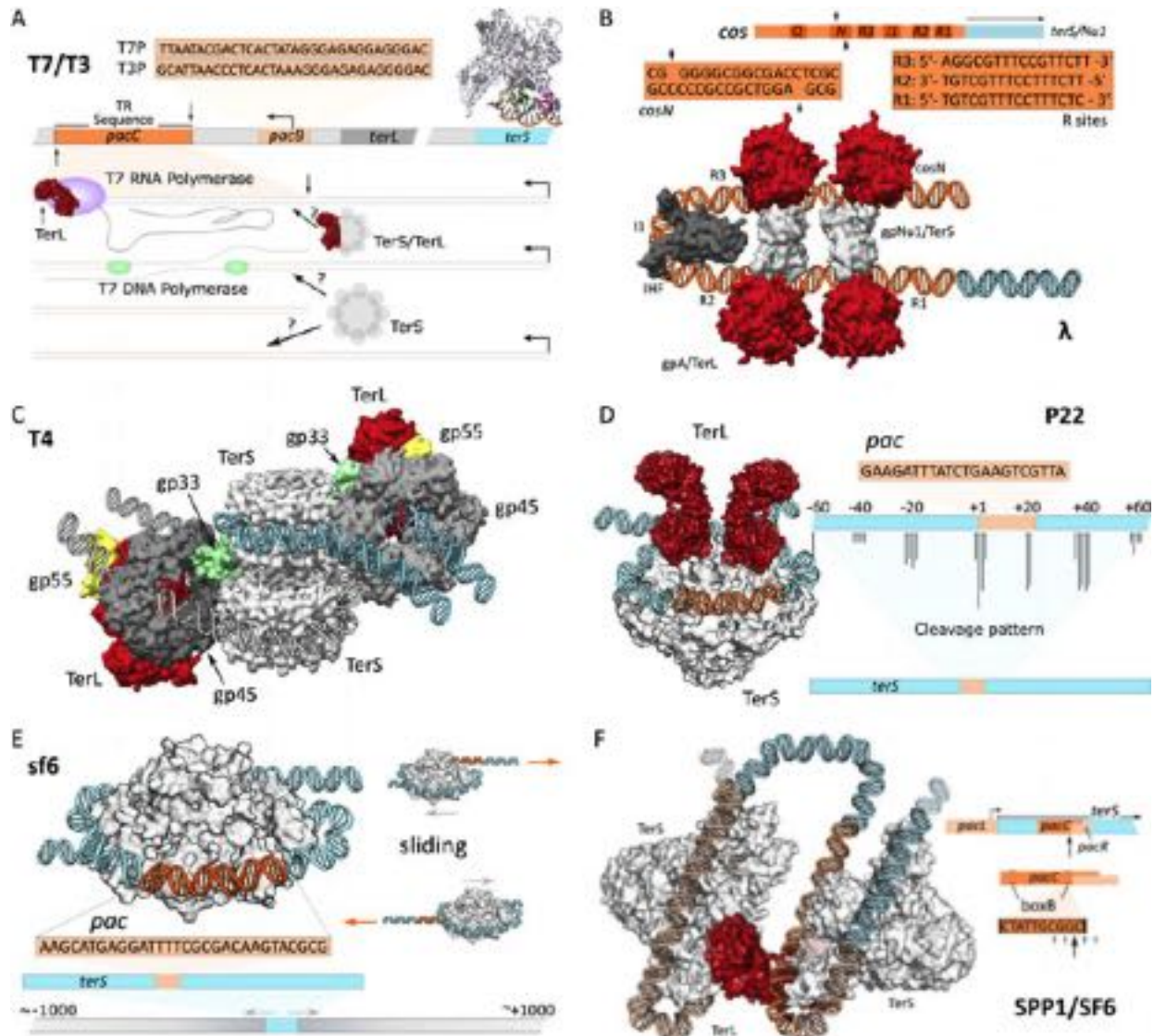


Fig. 5 Models of the packaging initiation processes for concatemeric genomes. In all images packaging proceeds from the cleavage site towards the right hand end of the genome as depicted. (A) Depiction of the genome recognition and packaging initiation events in T3 and T7 phages; inset: surface cartoon showing specific contacts between T7 RNA polymerase (PDB 1CEZ, light purple) and the T7 promoter *pacB* DNA (light orange) with the transcription start site depicted by the bent arrow. RNA polymerase pauses after transcription of the *pacC* Terminally Redundant (TR) sequences (dark orange) and is bound by TerL (red) which cuts the end of the *pacC* sequence. T7 DNA polymerase (green) extends the cleaved 3' ends of the DNA to create the terminal repeat sequences. It is unknown how the cleavage of the opposite end of the TR sequence occurs, or what role the octomeric TerS (gray circles) may play in this process (indicated by the arrows and question marks). (B) Cartoon image of bent *cos* DNA (orange) with bound IHF dimer (dark gray) and the tetramer of terminase protomers binding to I1 and R sites respectively (TerS₂:TerL, white and red respectively); the sequence of both strands of the *cosN* site are shown with the arrows denoting the cleavage sites. *cosN* and R site sequences are shaded in orange. The sequence coding for TerS/gpNu1 is shown in blue. (C) Diagram of T4 packaging initiation showing a model for the synapsis complex formed by two stacked TerS oligomers (white) loaded on to two different genomic segments of the concatemer (gray and blue). TerL (red) interacts with gp55 (late sigma factor, yellow) bound to the T4 DNA sliding clamp (gp45, dark gray) and a transcriptional coactivator (gp33, green). (D) Cartoon representation of genome recognition and maturation events in P22 phage. A complex of 1–3 TerL proteins (red) assembled on the TerS nonamer (white) binds to the *pac* sequence (orange) located within the *terS* gene (blue). Potential binding site of *pac* DNA (orange) located near the C-terminal end of TerS is shown by a segment of DNA. Cleavage occurs in a defined pattern with a 120 bp segment around the *pac* sequence, (arrows, length indicates cleavage frequency). (E) Representation of the recognition of Sf6 genomic DNA recognition by the HTH domains of TerS (white) of the *pac* sequence (orange) located in the *terS* gene (blue). Cleavage occurs non-specifically over a ~2000 bp region of the genome centered on the *pac* sequence, with decreasing frequency depicted by the gradient from gray (high frequency) to white (low frequency). This non-specific cleavage may be mediated by sliding of TerS wrapped DNA along the DNA strand. (F) Depiction of SPP1/SF6 *pac* site showing the *pacL* (light orange) sequence which has a propensity for bending and covers the promoter (bent arrow) regulating *terS* (blue) expression. Two TerS assemblies (white) bind to *pacL* and *pacR* containing DNA in a wrapped complex that recruits TerL (red) to cleave at *pacC*. Cleavage (arrows) occurs within a few bases of the preferred cleavage site (large arrow).

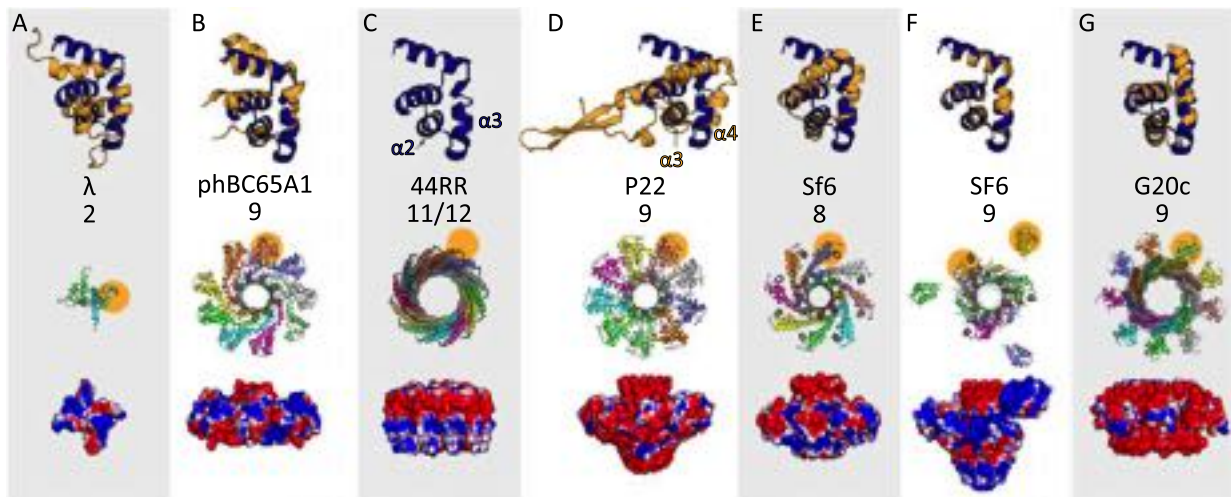


Fig. 6 HTH domains and oligomeric TerS assemblies. Top Panel: Cartoon image of the HTH domains (yellow-orange) from TerS proteins of different phages (A: λ (PDB 1J9I); B: phBC65A1 (PDB 2A09); C: T4-like 44RR (3XTS); D: P22 (PDB 3P9A); E: Sf6 (PDB 4YDQ); F: SF6 (PDB 3ZQ); G: G20c (PDB 4XVN)) aligned with the SF6 DBD (PDB 4ZC3, blue). The support and recognition helices for the SF6 DBD ($\alpha 2$ and $\alpha 3$ respectively) and P22 TerS ($\alpha 3$ and $\alpha 4$, respectively) are labeled. Middle Panel: Ribbon diagram of the oligomeric assemblies of the TerS with the DBDs highlighted by yellow circle and the number of monomers in the assembly listed above. Bottom panel: Image of the electrostatic surface potentials of circular TerS oligomers, showing areas of positive (blue) and negative (red) charge.

until the *cos* sequence in the next adjacent genome reaches the motor. At this point, the *cosQ* motif assists to terminate the packaging process, promoting cleavage at the subsequent *cosN* site and the DNA/motor complex is transferred to an empty capsid for the ensuing round of packaging. Assembly of the terminase complex onto the *cosB* sequence during the initiation process utilizes both DNA sequence specific and conformational recognition.

Crucial to the assembly of the active terminase complex at the *cosB* site is the specific interaction of the IHF dimer, with the I1 subsite (27 bp) located equidistant between the essential TerS binding sites (inverted repeats R3 and R2). As expected, IHF binding introduces a sharp hairpin-like bend ($\sim 160^\circ$) in the I1 segment of dsDNA which places the R2 and R3 motifs directly opposite each other on the inside surfaces of the bent dsDNA strands. This IHF stabilized conformation promotes the cooperative assembly of the final tetrameric protomer (TerS₂-TerL) assembly that is tightly bound to the R2 and R3 sequences, possibly through one TerS subunit from each of two protomers (Fig. 5B). The remaining two protomers in the tetrameric complex are likely bound to DNA at the non-canonical R3 and *cosN* or *cosQ* (also known as R4) sites by weak DNA binding interactions and the increased local concentration generated by the tetrameric complex bound with high affinity to the canonical R2/R3 sequences on the same dsDNA hairpin scaffold. Cleavage at *cosN* by TerL creates an intermediate complex where the *cosQ* containing DNA fragment is only weakly bound to a single protomer and no longer benefits from the cooperative affinity mediated by a continuous uncleaved strand of dsDNA. This results in the release of the *cosQ* containing DNA from the assembled tetramer of TerS₂-TerL protomers bound to the nascent end of the genome.

Molecular detail of the contacts between components of the terminase complex and the *cos* DNA scaffold are limited to the N-terminal DBD of TerS, which forms a winged HTH domain (wHTH). This domain forms a dimer in solution with the helices of the HTH domain on the opposite protein surface to the dimerization interface (Fig. 6). While this model may not represent the natural oligomeric status in either the protomer or the tetrameric protomer assembly, it has proven to be a useful model system for analysis of λ TerS wHTH interaction with its cognate R2 or R3 DNA sequences. Residues on the recognition helix (helix 2) are proposed to form hydrogen bonds with base pairs in the major groove, while non-specific electrostatic contacts are likely made between the negatively charged phosphate backbone of the DNA and the positively charged amino acids on the N-terminal arm, wing and C-terminal end of helix 4 (Fig. 5B). Differences between the closely related λ and 21 phages can be used to infer the interactions that are crucial to genome selectivity. Within the 16 bp sequence of the conserved R motifs, two base positions determine packaging preference, where T7/C10 is bound specifically by λ TerS and G7/G10 is recognized by 21 TerS. However, precise mapping of the contacts requires more biophysical and structural experiments. Cleavage at *cosN* and release of the *cosQ* DNA end allows the reordering of the terminase proteins around the nascent end of the *cosB* DNA and the subsequent assembly of this motor-DNA complex to start the packaging process.

The mature packaged genomes of *E. coli* phage HK97 and related viruses (e.g., HK022) have complementary 10 base 3' ssDNA overhangs (3' GCGGCCGGTTT 5'), defined by a conserved *cos* sequence around the maturation cleavage site. While some proteins are predicted to share >90% amino acid (aa) similarity with λ phage proteins, no similarities are predicted in the terminase proteins. However, weak homology has been noted with the terminase proteins from the *Bacillus* phage phBC6A51 (40% aa similarity), the TerS of which forms a circular oligomeric complex (Fig. 6). Other than the fact that *cos* cleavage requires the terminase proteins (gp1 and gp2) and a virally encoded endonuclease containing an HNH motif (gp74), little is known about how the viral genome is recognized and the nascent ssDNA 3' overhang end is generated.

Genome recognition in headful bacteriophages

Bacteriophages that use the headful packaging system to pump slightly more than one genome length of the concatemeric DNA, result in the packaged DNA being circularly permuted, where the end of the genome that is packaged last overlaps somewhat into the adjacent genome copy, duplicating a short segment of the end of the genome that was packaged first. The iterative secondary transfer events after this initial 1+ packaging event cause the ends to the subsequently packaged genomes to migrate along the genome in the direction of packaging. This degeneracy in the genome composition relative to the ends of the packaged dsDNA is reflected in the variable requirement for sequence specific genome binding and initial cleavage precision used by different headful phages described below.

Since the *E. coli* T4 bacteriophage produces nuclease enzymes that degrade the host genomic DNA during viral replication, specific binding of the terminase proteins (gp16 and gp17) to a *pac* sequence appears to be less critical for packaging of the correct genome. Despite this, weak preferential binding has been observed between TerS (gp 16) and a GC rich sequence at the end of its coding region, subsequently denoted as the *pac* sequence (Fig. 5(C)). However, in the absence of this sequence, cleavage of the concatemeric DNA, prior to packaging, can occur at alternative terminase binding sites. T4 phage TerS forms single circular oligomeric assemblies of 11 monomers, while TerS from a related phage, 44RR, is found as oligomeric assemblies of 11 or 12 monomers (Fig. 6). Single T4 TerS rings may also stack, and two such rings can bind to different sites on the concatemeric T4 genomic DNA creating a 'synapse' that is thought to be important for terminase gene amplifications and packaging initiation. No DBD has been identified in these TerS proteins, and while the central oligomerisation domain does not bind to DNA, it is essential for genome packaging. This suggests a role for conformation recognition of genomic DNA. The T4 phage TerS oligomers support the stable assembly of higher order assemblies of TerL which exhibit enhanced ATPase activity. *In vivo*, genome end processing and packaging initiation of nascent concatemer DNA is coupled to genome replication and RNA transcription of the late genes required for viral survival. This occurs *via* the interaction of TerL with a large regulatory protein complex comprised of the DNA sliding clamp and two transcription cofactors (gp45/gp55/gp33), which assist in localizing the terminase complex to newly synthesized genomic DNA for end maturation (likely at the *pac* site in gp16) and subsequent packaging into the phage heads.

Selective packaging of the *Salmonella* phage P22 genome relies on sequence specific binding of the terminase proteins to the 22 bp *pac* sequence in order to direct the cleavage event that generates the free dsDNA end required for translocation of the genome into the capsid (Fig. 5(D)). A single cleavage event occurs most often close to the first base of the *pac* sequence. However, this initial cleavage event can also occur at any one of several positions, approximately spaced by 20 bp, within the 120 bp region spanning the *pac* site, suggesting that endonuclease activity is not directly sequence dependent. The P22 TerS protein (gp3) forms a circular assembly of 9 monomers (Fig. 6) that is known to associate with up to 3 TerL (gp2) monomers (Fig. 5(D)). The N-terminal domain of TerS does not form a classical HTH motif as it contains a β -hairpin insertion, however, the most similar structure to this domain is the DBD from the *Bacillus* phage SF6 TerS (Fig. 6), an HTH motif. The first two α -helices after the P22 TerS β -hairpin ($\alpha 3$ and $\alpha 4$) are in a similar orientation to the support and recognition helices ($\alpha 2$ and $\alpha 3$, respectively) of the SF6 HTH (Fig. 6). Indeed, mutations in P22 TerS amino acid residues, Leucine 80 and Glutamic acid 81, shown to alter *pac* specificity, map to helix $\alpha 4$ of the HTH-like motif, suggesting that this helix may also play a direct role in *pac* sequence recognition and binding (Fig. 5(D)). The flexible C-terminal segment of the TerS monomer is highly positively charged and can bind to the N-terminal domain of TerL to form the full Terminase complex (TerS₉:TerL₂₋₃). Although this C-terminal TerS segment can interact nonspecifically with DNA in the absence of TerL, stable formation of TerS-TerL complexes, both *in vivo* and *in vitro*, preclude the requirement of this interaction for specific recognition of the P22 viral genome. While a threading model for DNA interaction has been proposed, where the DNA passes through the central channel of the nonamer, the 120 bp cleavage zone around the *pac* site suggests a wrapping model reminiscent of the nucleosome core particle (Fig. 3(C)) and is consistent with the involvement of $\alpha 4$ in *pac* binding (Fig. 5(D)). Assuming a roughly B-form DNA conformation with regular bends, as in the nucleosome wrapped DNA, with the *pac* site centered within the bound DNA segment, evenly spaced sequences (where 20 bp is approximately equal to two turns of B-form DNA) would be presented facing outwards for cleavage by a TerL nuclease domain. This spacing, rather than the 10 bp of a single turn of B-form indicates that additional interactions between the DNA and TerS or TerL may stabilize the complex and occlude access to the DNA by TerL nuclease active site.

The packaging initiation endonuclease reaction of the *Shigella* phage Sf6 concatemer can happen at any one place over a 2000 bp range, centered around a 30 bp *pac* site identified in the TerS (gp1) gene (Fig. 5(E)). This sequence is recognized by the N-terminal HTH domains of the Sf6 TerS protein that are displayed externally around the central stabilising core of the circular octamer (Fig. 6), suggesting that both sequence specific and conformational recognition modes play a role in assembly of the TerS-*pac* DNA complex. While a *pac* sequence for preferential binding has been identified, Sf6 TerS shows high non-specific binding activity to both linear and supercoiled dsDNA, through positively charged residues located on the external surface of the circular array. The C-terminal domain of the DNA bound TerS assembly interacts with TerL (gp2) bringing the nuclease domain into close proximity to the DNA for the maturation cleavage event. Since cleavage can occur anywhere up to 1000 bp from the *pac* site in either direction, it is possible that once assembled on the *pac* sequence, the DNA wrapped TerS complex can slide randomly along the DNA, in a similar mechanism to nucleosome core particle sliding. A random sliding activity is consistent with the observation that the cleavage event takes place more often at *pac* proximal sites than at sites located at the extreme ends of the 2000 bp cleavage zone.

The concatemer DNA from the closely related *Bacillus* phages, SF6 and SPP1, is initially processed to produce DNA that is cleaved within a short 6bp segment of the *pacC* site to produce single base 3' overhangs (Fig. 5(F)). The 270 bp *pac* sequence is comprised of two main binding regions, *pacL* and *pacR*, on either side of the *pacC* site. The *pacL* sequence has a strong propensity

for bending and is bound by the nonameric TerS ring-like assembly through conformation recognition of the bent DNA assisted by weak sequence specific contacts. Nine HTH domains project a positively charged surface from the central oligomeric core (Fig. 6) that presumably allows the protein to form a complex with the DNA wrapped around it. Two TerS oligomers each bind to individual *pacL* and *pacR* sites placing the associated TerL nuclease domain so that it cleaves accurately at *pacC* regardless of the local nucleotide sequence at the cut site.

Atomic structures have been determined for TerS assemblies from understudied bacteriophage (*Thermus* phage G20c and *Bacillus* phBC65A1) where there is little information on the genome maturation and packaging processes. These proteins also form circular assemblies of 9 monomers that display classical HTH domains around the central oligomerisation domain (Fig. 6), however the mechanism of genome recognition and the packaging initiation cleavage event are yet to be defined.

DNA Wrapping/Bending is a Common Feature in Selective Packaging

Bacteriophages have long served as relatively simple model systems in which to study the fundamental mechanisms used by the macromolecular machines that act on nucleic acids. This longstanding focus towards understanding these processes has yet to yield much of the molecular detail of protein-nucleic acid complex assembly and function during packaging initiation. Conspicuous among these is the absence of atomic structural information of Terminase-DNA complexes. However, the puzzle pieces of information that are available for the different phage systems highlight some general features of the mechanisms for selective binding and packaging of cognate genomic DNA.

The N-terminal DBDs of TerS and $\phi 29$ TP are predicted to be α -helical. Indeed, many TerS DBDs form HTH or HTH-like motifs (Fig. 6(A)) supporting a model where sequence specific binding might occur through the placement of a recognition helix along the major groove of the dsDNA at the preferred sequence motif. However, the affinity of the sequence specific contacts between the cognate DNA targets and HTH motifs are generally weak, consistent with the observation that TerS proteins from many phage systems form circular oligomeric arrays of 8–11 monomers (Fig. 6(B)). Oligomeric assemblies of TerS have also been reported for other viruses (8 for T7 and 11 for T4). The positioning of the α -helical DBDs, displayed in a ring around a core oligomerisation domain, creates a ring of positive charges (Fig. 6(C)) around which the genomic DNA is wrapped. Additional non-sequence specific contacts between the DNA and TerL protein(s) bound to the TerS oligomer may act like DBDs, to assist in stabilising the ternary complex in some viruses, such as P22. Flexible positioning of the DBDs relative to the oligomeric core, and each other, provide a pliable DNA binding surface that can accommodate variation in the shape of the target DNA. Similar wrapping of negatively supercoiled DNA around the cylindrical connector protein has been proposed for initiation of $\phi 29$ packaging. This serves two purposes: providing a mechanism for recognizing the general topology of viral genomic DNA compared to host cell DNA (for both monomeric and concatemeric genome replication products); and in concatemeric DNA, allowing the circular oligomeric TerS to form a stable complex with the continuous strand of nascent genomic DNA and to present a more accessible bent strand as the substrate for the cleavage event required for packaging initiation.

Despite the variation in approaches used by different dsDNA bacteriophages to generate new copies of their genomes and to couple packaging to this replication process, specific genome binding and selective packaging generally combines both sequence specific and conformational recognition modes, along with the avidity afforded by multiple contacts between the protein subunits of the oligomeric terminase assembly and DNA. This commonality exploits the basic topological properties of phage genomic DNA that, along with the most powerful molecular motor known in nature, allow the DNA to be packed tightly into the capsid creating an efficient method for packaging and transferring these large viral genomes between host cells.

Further Reading

- Black, L.W., 2015. Old, new, and widely true: The bacteriophage T4 DNA packaging mechanism. *Virology* 479–480, 650–656.
- Djacek, K., Tavares, P., Oliveira, L., 2017. Bacteriophage SPP1 *pac* cleavage: A precise cut without sequence specificity requirement. *Journal of Molecular Biology* 429 (9), 1381–1395.
- Feiss, M., Reynolds, M., Schrock, M., Sippy, J., 2010. DNA packaging by λ -like bacteriophages: Mutations broadening the packaging specificity of terminase, the λ -packaging enzyme. *Genetics* 184 (1), 43–52.
- Grose, J., Jensen, G., Burnett, S., Breakwell, D.P., 2014. Genomic comparison of 93 *Bacillus* phages reveals 12 clusters, 14 singletons and remarkable diversity. *BioMed Central Genomics* 15 (1), 855–875.
- Jardine, P.J., Anderson, D.L., 2006. DNA packaging in dsDNA phages. In: Calendar, R.L. (Ed.), *The Bacteriophages*. Oxford University Press, pp. 49–65.
- Juhala, R.J., Ford, M.E., Duda, R.L., *et al.*, 2000. Genomic sequences of bacteriophages HK97 and HK022: Pervasive genetic mosaicism in the lambdoid bacteriophages. *Journal of Molecular Biology* 299 (1), 27–51.
- Kala, S., Cumby, N., Sadowski, P.D., *et al.*, 2014. HNH proteins are a widespread component of phage DNA packaging machines. *Proceedings of the National Academy of Sciences of the United States of America* 111 (16), 6022–6027.
- Leavitt, J.C., Gilcrease, E.B., Wilson, K., Casjens, S.R., 2013. Function and horizontal transfer of the small terminase subunit of the tailed bacteriophage Sf6 DNA packaging nanomotor. *Virology* 440 (2), 117–133.
- McNulty, R., Lokareddy, R.K., Roy, A., *et al.*, 2015. Architecture of the complex formed by large and small terminase subunits from bacteriophage P22. *Journal of Molecular Biology* 427 (20), 3285–3299.
- Oliveira, L., Tavares, P., Alonso, J., 2013. Headful DNA packaging: Bacteriophage SPP1 as a model system. *Virus Research* 173 (2), 247–259.

- Oram, M., Black, L.W., 2011. Mechanisms of genome packaging. In: Agbandje-McKenna, M., McKenna, R. (Eds.), *Structural Virology*. Royal Society of Chemistry, pp. 205–211.
- Rao, V., Feiss, M., 2008. The bacteriophage packaging motor. *Annual Review of Genetics* 42, 647–681.
- Rohs, R., Jin, X., West, S.M., *et al.*, 2010. Origins of specificity in protein-DNA recognition. *Annual Review of Biochemistry* 79, 233–269.
- Salas, M., Redrejo-Rodríguez, M., de Vega, M., 2016. DNA-binding proteins essential for protein-primed bacteriophage ϕ 29 DNA replication. *Frontiers in Molecular Biosciences* 3, 1–21.
- Yang, T., Ortiz, D., Yang, Q., *et al.*, 2017. Physical and functional characterization of a viral genome maturation complex. *Biophysical Journal* 112 (8), 1551–1560.

Relevant Websites

x3DNA.org

3DNA Homepage – Nucleic Acid Structures.

pdb.org

wwPDB: Worldwide Protein Data Bank.

DNA Packaging: The Translocation Motor

Janelle A Hayes and Brian A Kelch, University of Massachusetts Medical School, Worcester, MA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Arginine finger A basic residue, most commonly an arginine, that works *in trans* during ATP hydrolysis within an oligomeric complex. In other words, the arginine finger of a subunit within a multimeric ATPase will participate in hydrolyzing the ATP bound in a neighboring subunit's active site.

ASCE A member of the Additional Strand Conserved Glutamate family of ATPases.

Concatemer A single DNA strand that contains multiple identical copies of a DNA sequence in series.

Power stroke The force-generating step of a motor.

Rossmann fold A highly conserved nucleotide binding motif which is composed of a β -sheet sandwiched between multiple α -helices.

Introduction

All viruses must encapsulate their genome into a capsid shell during the replication cycle. Many viruses use an ATPase motor to actively pump their DNA into the small volume inside the viral particle, strictly confining the genome. Viruses that use this method of maturation typically have large genomes and the ATPase motor overcomes the entropic and enthalpic penalties for encapsulating this large amount of genetic material. In many of these motor-driven viruses, the genome is packaged to very high density, often approaching a quasi-crystalline state. For example, the ~ 48 kb genome of phage lambda is ~ 16 μ m end-to-end, but is packaged into a head with a diameter of only ~ 60 nm. The motor needs to be extremely powerful in order to pump the genome against the internal pressure that builds during packaging, and also highly regulated to ensure complete and efficient encapsulation. Therefore, viral genome packaging motors are a particularly exciting model for understanding motor force-generation and regulation.

Genome packaging motors are found in many viruses that infect organisms from all three domains of life. This includes viruses as large as the giant mimiviruses to small viruses such as adeno-associated viruses. For many of these motors, the molecular mechanisms and even the identities of motor components remain unknown. Here, we will focus on the best-understood classes of viral genome packaging motors: the terminases and phi29-type motors. These motors catalyze genome packaging in tailed bacteriophage, as well as herpesviruses. Because phage are the most numerous biological entities, these motors likely perform the bulk of viral genome packaging on Earth. Terminases are also important drug targets for combating herpesvirus infections, further highlighting the importance of these motors.

Both terminase and phi29-type motors pump DNA through a dodecameric ring complex that sits at one of the five-fold vertices of the icosahedral shell. In terminase motors, this complex is called the 'portal', whereas it is called 'connector' in phi29-type motors. Not only does DNA enter through the portal ring during packaging, but it also exits through the same channel. The portal also serves as the binding site for the neck and tail proteins that assemble upon maturation of the virus particle. Numerous crystal structures of portal complexes from diverse phages reveal that these proteins share a common core structure. This indicates that all portals are evolutionarily related, despite low sequence homology in their primary sequence. The portal ring has a central channel that is just wide enough for double-stranded DNA to pass through. This finding has led to the proposal that the portal strips the incoming genome of any proteins bound to the DNA. Moreover, this tight channel is suggested to play a critical role in the packaging process itself, perhaps by acting as a one-way valve or even as a primary component of the force-generation step. In all examined cases, the dodecameric portal complex creates a symmetry mismatch with the capsid by replacing a pentameric coat protein complex at the five-fold vertex of the icosahedron. Whether this symmetry mismatch plays a functional role in packaging remains unknown, although recent work suggests a possible role in genome packaging termination.

Terminase and Phi29 Motors are Branches Within the ASCE Family of ATPases

Phi29 motors primarily consist of the connector assembly, ATPase motor, and a specialized prohead RNA molecule, known as 'pRNA' (Fig. 1). During packaging, the connector protein binds to the ATPase motor through interactions with the pRNA, and the motor translocates the DNA through the connector into the capsid. In terminase motors, the portal assembly binds to the ATPase subunit (the large terminase, or TerL) directly without pRNA. Terminase motors have an additional endonuclease domain, which is used to initiate and terminate packaging (discussed below). Despite the differences in motor assembly, both phi29 and terminase motors rely on a similar catalytic core to power DNA translocation: the ASCE ATPase fold.

Viral genome packaging motors are derived from the ASCE (additional strand conserved glutamate) division of ATPases. The ASCE division contains numerous superfamilies of ATPases, including widely studied AAA+ ATPases, ABC transporters, RecA, and

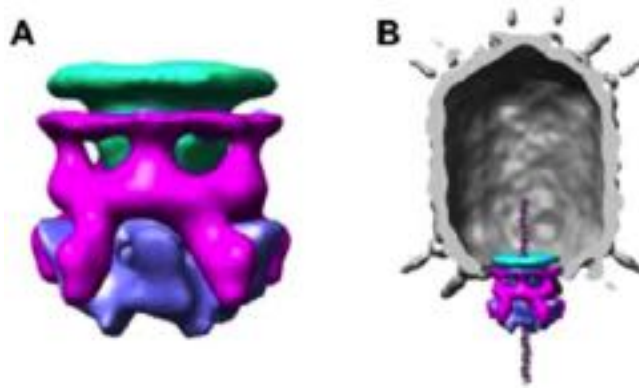


Fig. 1 CryoEM structure of the Phi29 Motor. Density from difference maps isolating the connector (green), pRNA (magenta) and ATPase (blue) were combined to visualize the DNA packaging motor in $\phi 29$. (A) Close-up view of the phi29 motor. (B) The motor is shown in the context of the entire prohead. The front half of the prohead density has been removed so all the motor components can be seen. This figure has been adapted with permission from Morais, M.C., Koti, J.S., Bowman, V.D., *et al.*, 2008. Defining molecular and domain boundaries in the bacteriophage $\phi 29$ DNA packaging motor. *Structure*, 16 (8), 1267–1274.

SF1/SF2 helicases (Fig. 2). ASCE ATPases are ancient machines that function across all cellular life, as well as many viruses. Although their molecular mechanisms and roles are remarkably diverse, ASCE ATPases usually function as multimers that convert the chemical energy of ATP into mechanical work. ASCE ATPases also generally use a similar ATP hydrolysis mechanism. Thus, studies of the packaging motor structure and mechanism have broad relevance towards understanding the general principles of molecular machines.

The ASCE nucleotide binding motif uses a Rossmann ' $\beta\alpha\beta$ ' fold which is comprised of a β -sheet sandwiched between multiple α -helices. Different clades within the ASCE tree may add variable elements to this base structure, however there are four features that are common to all ASCE ATPases: the Walker A and B motifs, a catalytic base that is usually a glutamate residue, and a *trans*-acting motif (Fig. 3). The Walker A motif (also known as the P-loop; consensus sequence [G/A]xxxGK[T/S]), is a nucleotide binding motif that forms hydrogen bonds with the ATP phosphates in the active site. The threonine/serine residue in the Walker A motif is critical, as it coordinates a metal ion, often a magnesium, for ATP hydrolysis. Downstream in the protein sequence, the Walker B motif (consensus sequence $\phi\phi\phi\phi$ [D/E]), further assists in binding this metal, which coordinates the β and γ phosphates of ATP. Immediately downstream of the Walker B motif, a glutamate residue acts as an essential catalytic base, activating water for nucleophilic attack on the γ -phosphate of ATP. (Often this 'catalytic glutamate' is included in the Walker B motif, which yields a consensus sequence $\phi\phi\phi\phi$ [D/E]E.) The fourth feature, a positively charged *trans*-acting element from a neighboring subunit (known as the arginine finger) interacts with the γ phosphate of ATP to stabilize the hydrolysis transition state. This *trans*-acting element also links nucleotide sensing and ATP hydrolysis between subunits, a necessity for regulating activity within oligomeric ATPases.

Terminase and phi29-like motors have numerous features that are similar, suggesting a relatively close phylogenetic relationship. Previous phylogenetic analysis classified the terminases and the phi29-type motors as disparate families within the ASCE division of ATPases. The phi29-type motors were placed within the HerA/FtsK family of dsDNA translocases and the terminases as an offshoot of RecA ATPases. This classification was based purely on sequence analysis and secondary structure prediction. However, recent structural and mechanistic discoveries led to the conclusion that terminase and phi29-type motors are much more closely related and are two distinct clades within the same family (Fig. 2). The ATPase subunits of phi29 and terminase motors have a core Rossmann fold with additional elements appended. For example, there is an additional α -helix and three antiparallel β -strands inserted between strand 2 and helix B of the core Rossmann fold, differentiating them from other ASCE ATPases. An N-terminal helix has also been added, where it packs atop helices A, B, and C of the core Rossmann fold. The phi29 ATPase domain also contains a short antiparallel β -strand inserted between helix D and β -strand 5. Additionally, terminase motors have a C-terminal nuclease domain, while phi29 motor proteins lack nuclease activity. This is not surprising, because the phi29 motor does not cleave DNA during packaging (see below for more details). Finally, the terminase motor has a three-helix bundle subdomain in between the core ATPase domain and the nuclease domain. We refer to this subdomain as the 'lid' because it caps the ATPase active site (others have referred to this substructure as a 'linker' domain or ATPase subdomain II). The three-helix bundle lid subdomain is distinguished from the four-helix bundle lid subdomains of AAA+ and STAND ATPases. It is unclear whether phi29 has a similar lid subdomain, as the protein used for crystallization only contained the core Rossmann fold of the ATPase.

In addition to similar ATPase domain structure, both terminase and phi29 proteins form oligomers when bound to the portal/connector. In the absence of the portal/connector proteins, the subunits of terminase and phi29 motors are generally monomeric. CryoEM and biochemical studies of the terminase motor show a pentameric assembly of subunits. However the oligomeric state of the phi29 motor remains contentious, as biochemical studies suggest a hexameric assembly.

Assembly of the oligomeric motor onto the portal/connector facilitates ATPase and packaging activity. Both terminase and phi29 motors use a *trans*-activated ATPase mechanism through an arginine finger motif, similar to most other ASCE ATPases. The

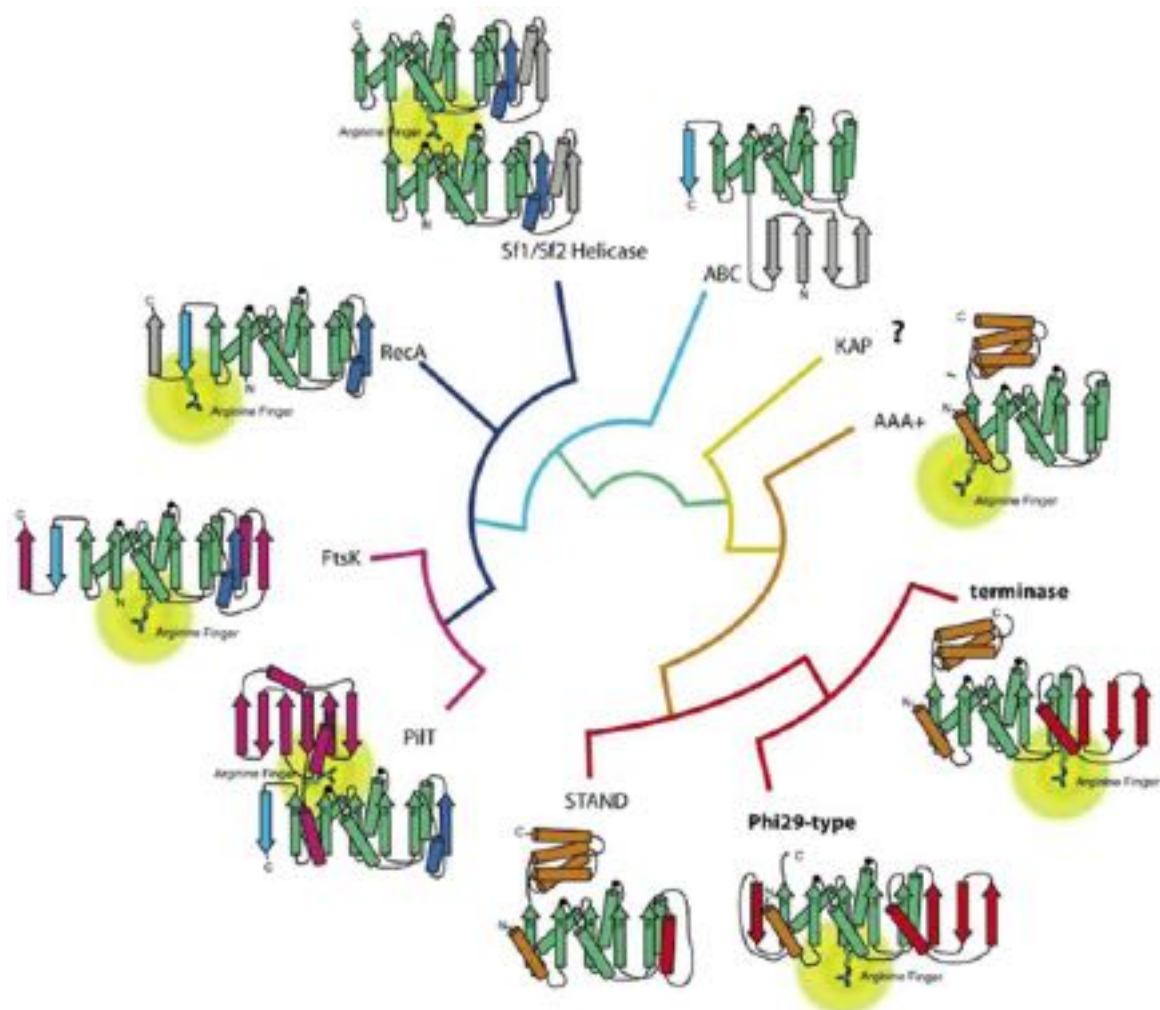


Fig. 2 Phylogenetic tree of ASCE ATPases. The core Rossmann fold is shown in green, while branch-specific structural features are shown in various colors. The terminases and phi29-type ATPase families likely evolved from AAA+ ATPases.

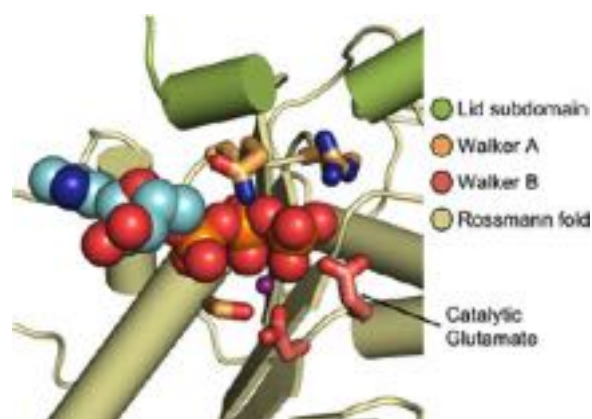


Fig. 3 The terminase ATPase active site. The P74-26 bacteriophage ATPase domain is shown. ADP-BeF₃, a non-hydrolyzable ATP analog, is bound in the active site. The Walker A motif hydrogen bonds with the ATP phosphates, and the conserved serine residue coordinates a magnesium ion (purple sphere). The Walker B motif further assists in binding the magnesium, with the catalytic glutamate coordinating a water (not shown) for nucleophilic attack.

arginine finger is positioned within the interface between two subunits, directly contacting the ATP γ -phosphate to stabilize the developing negative charge during the ATP hydrolysis transition state. In phi29, the arginine finger also causes ejection of ADP from the post-hydrolysis state. Thus, the arginine finger residue not only plays a catalytic role, but also ‘communicates’ nucleotide status between subunits within the ring. Because the arginine finger is central to motor function, its position on the protein surface dictates the relative orientation of neighboring subunits within the ring. In terminase and phi29 motors, the arginine finger is located in different regions of the protein, suggesting that the overall orientation of ATPase subunits may be altered between the two classes.

Phi29 and terminase motors are further distinguished by the accessory subunits required for efficient packaging. Terminase motors use the small terminase protein (TerS), which specifically recognizes the viral genome and modulates large terminase ATPase and nuclease activity. Phi29 lacks a clear homolog of TerS, and instead uses the gp3 protein which is covalently bound to the 5′ ends of the viral genome. The gp3-DNA complex is suspected to be the recognition element for packaging initiation. Additionally, phi29 uses the 174 base pair long pRNA for binding the motor to the connector. The pRNA forms a pentameric assembly on the connector, binding to the phi29 motor through a spoke-like RNA secondary structure. Terminase motors do not require a pRNA to bind portal, and bind through direct protein-protein interactions.

Phi29 Motors

The viral genome packaging machinery in phage phi29 is arguably the best understood system. The phi29 packaging motor was discovered in the 1970s and an *in vitro* packaging system was established in 1986. The ability to reconstitute genome packaging from purified components opened the door for years of incisive study. Furthermore, the phi29 motor chemomechanical cycle has been elucidated to high detail using an elegant single-molecule optical trapping assay for motor activity. Early optical trap experiments established phi29 as an exceptionally powerful motor, working against forces averaging 57 pN, a force close to that at which dsDNA fundamentally alters its conformation through over-stretching. Subsequent experiments identified individual subunit stepping, which provided incredible detail into the mechanisms of force generation. Finally, recent structural analysis has visualized a functioning holoenzyme at low resolution and various individual components at high resolution.

Phi29 genome replication and packaging begins with the gp3 protein. At the beginning of viral genome replication, gp3 covalently binds to the 3′ end of the genome, where it acts as a primer for replicating the 5′ strand. The gp3 priming domain dimensions and charge mimic DNA, allowing the polymerase to initiate replication by covalently linking a nucleotide to the hydroxyl of a serine within gp3. After DNA replication is complete, both ends of the DNA, known as left and right, remain bound to gp3. DNA-gp3 forms lariat loops, with the left end junction preferentially interacting with the packaging motor bound to the prohead. Binding to the motor causes DNA to supercoil, with some evidence suggesting that the supercoil wraps around the connector protein outside of the capsid. In an unknown mechanism, the motor begins packaging the gp3-DNA, possibly through packaging the loop in its entirety, or through gyrase activity that has been speculated but never directly observed, allowing the loop to resolve.

The translocation of the genome into the empty phi29 capsid occurs in a complicated, ATP-dependent mechanism. Each translocation cycle can be separated into two phases: a ‘burst’ phase in which DNA physically pushes into the capsid, and a ‘dwell’ phase in which the motor resets (Fig. 4). At the beginning of the ATP hydrolysis cycle, each subunit of the phi29 pentameric motor binds ATP. Once the ring is fully loaded with ATP, ATP hydrolysis is stimulated in a sequential order around the ring. During the burst phase, DNA is translocated by a total of 10 bp. By slowing the rate of translocation, Bustamante and colleagues showed that the 10 bp burst actually consists of four discrete 2.5 bp substeps. Thus, the pentameric phi29 motor uses 4 translocation events per cycle, which suggests that not every subunit directly translocates DNA per cycle.

One of the five motor subunits hydrolyzes ATP, but does not translocate DNA, and is designated the ‘special’ subunit. The exact role of the special subunit is still unclear; however, it is critical for regulating the motor during the dwell phase, which occurs in between each of the 10 bp translocation increments. It is thought that the special subunit grips DNA during the dwell phase, and releases ADP from its active site, allowing ATP to bind. The arginine finger of the special subunit and each subsequent subunit triggers ADP exchange for ATP sequentially within the ring, resulting in a fully ATP-loaded motor. The DNA-bound special subunit then initiates ATP hydrolysis, breaking its contact with DNA, and triggering the burst translocation phase. During the four subsequent ATP hydrolysis events, ATP is hydrolyzed and P_i released. P_i release induces the conformational change that drives the translocation step, packaging 2.5 bp of DNA before handing it to a neighboring subunit. Upon the power stroke conformational change, the arginine finger of the currently active subunit positions into the neighboring subunit’s active site, priming the next subunit for an ATP hydrolysis event. In this way, hydrolysis in the next subunit is triggered, and the DNA is processively packaged.

During the various packaging steps, DNA needs to undergo multiple ‘grip-and-release’ cycles. DNA is held by each subunit tightly during the translocation substeps as well as during the dwell; however, each subunit must relinquish its hold on DNA to easily pass off the DNA to a neighboring subunit within the ring. DNA is gripped mostly through the 5′-3′ strand, with less contact made with the 3′-5′ strand and minor contacts with sugars and nucleotide bases. Upon passing to the next subunit during translocation, the DNA undergoes rotation. As sequential hydrolysis and DNA translocation occurs during the packaging reaction, translocation step size and DNA rotation per cycle starts to vary. As more DNA is encapsulated, the step decreases from 2.5 to 2.3 bp, and the rotation of DNA increases from -14° to -48° per increment. This is in accord with increased packaging density within the capsid and subsequent elevated backpressure, which the motor must work harder against in order to continue the

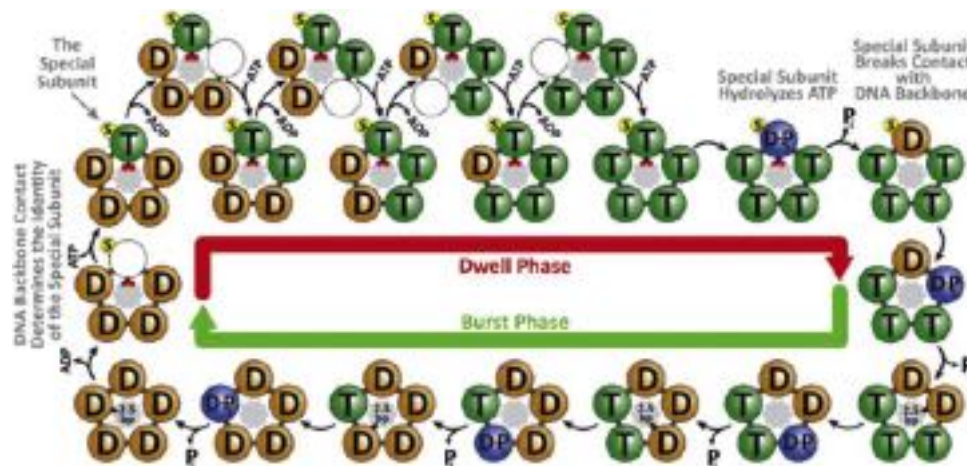


Fig. 4 The Phi29 chemomechanical ATPase cycle. At the end of the burst, all subunits are ADP-bound (“D” label). At the beginning of the dwell, the motor makes an electrostatic contact with two backbone phosphates (small red circles) on the dsDNA substrate (inside the ATPase ring). This unique contact determines the identity of the special subunit (yellow label “s”). The formation of the electrostatic contact facilitates ADP release by the special subunit. Subsequent ATP (“T” label) binding and ADP release events are interlaced, with ATP binding to one subunit enabling ADP release from its neighbor. After all five subunits have bound ATP, the special subunit hydrolyzes ATP (“D•Pi” label), releases Pi, and uses the hydrolysis free-energy to break the electrostatic contact with DNA, triggering the burst phase. During the burst, the remaining four ATP-bound subunits sequentially hydrolyze ATP, release Pi, and translocate DNA by 2.5 bp. The motor-DNA geometry (10.0 bp burst size versus 10.4 bp dsDNA helical pitch) favors a mechanism in which the same subunit is special in consecutive cycles. This figure is reprinted with permission from Chistol, G., Liu, S., Hetherington, C.L., *et al.*, 2012. High degree of coordination and division of labor among subunits in a homomeric ring ATPase. *Cell* 151 (5), 1017–1028.

packaging reaction. Recent work has shown that this backpressure can be relieved by a temporary pause in packaging, which allows the DNA to relax in the capsid. Because this relaxation time is far longer than the packaging time, this result suggests that the fast speed of packaging causes DNA to adopt a metastable conformation. These findings illustrate that motor speed has an important role in determining capsid internal pressure and genome dynamics.

The structure of the phi29 motor is beginning to come into focus. The connector assembly structure was first determined to high resolution by x-ray crystallography, along with a low-resolution cryoEM structure of the motor actively packaging. Subsequent cryoEM studies show that the portal undergoes large conformational changes in structure upon packaging and suggest a pentameric arrangement of the pRNA and ATPase components of the motor. The pentameric arrangement has been challenged, with the more commonly seen hexameric ATPase arrangement proposed. However, high-resolution asymmetric reconstructions of the phi29 motor structure exhibit a clear five-fold or pseudo-five-fold arrangement of subunits. Over the past decade, high-resolution structures of most of the motor components have been determined: a majority of the pRNA structure was determined by crystallography and NMR, as well as the ATPase domain of the gp16 motor protein. By docking the high-resolution structures into the cryoEM density, we are nearing a complete picture of the phi29 motor. High-resolution structures of the motor in action are still needed to understand the motor’s mechanism in detail. However, the rapidly developing field of cryoEM is nearly certain to profoundly impact our understanding of the phi29 motor in the future.

Terminase Motors

In comparison to phi29 motors, less is known about terminase motor mechanism. Because terminases have been studied from multiple phages, here we synthesize the results from many model systems, such as phages T4, lambda, SPP1, P22, Sf6, and P74–26. Terminases initiate genome packaging using the small terminase protein. The small terminase binds to the viral genome and modifies the enzymatic activities of the large terminase protein. In some cases, such as found in phage lambda, host factors augment the small terminase’s role as a regulator of large terminase function. Additionally, terminase activity is modulated by the procapsid, which is required for terminase translocation. Other factors have been proposed to regulate terminase functions, such as the gpFI protein of phage lambda. However, more recent evidence suggests that regulation by gpFI is either unnecessary for phage replication or does not occur *in vitro*. Terminases are also controlled at the level of protein translation, as the terminase protein is often expressed late in the infection cycle. Finally, large terminase genes often contain self-splicing intein sequences in their ATPase active sites that are hypothesized to modulate terminase activity.

Terminase motors translocate DNA at a faster rate than their phi29 counterpart, with the T4 motor taking 5–10 min to package its 171 kb genome. While terminases are faster translocases, they exhibit more variable speed. *In vitro*, the T4 motor translocates DNA at an average rate of 700 bp/s, peaking at 2000 bp/s with an ATP turnover rate of 300/s. The terminase of phage lambda peaks at ~600 bp/s at low capsid density, but this rate drops to ~200 bp/s towards the end of packaging. Meanwhile, phi29 packaging peaks at around 165 bp/s *in vitro*, making it significantly slower. Therefore, the motor speed seems to be correlated with

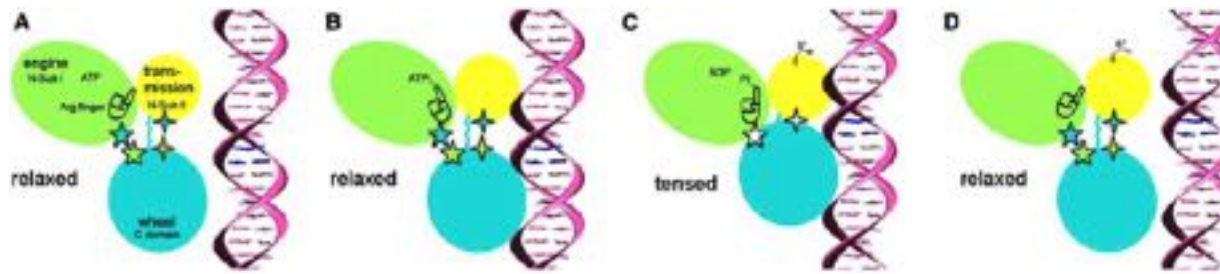


Fig. 5 Proposed ‘inchworm’ model for the mechanism of DNA translocation by the large terminase. Panels (A)–(D) relate to the sequence of events that occur in a single large terminase molecule. The large terminase N-terminal subdomain I, subdomain II, and C-terminal domain are represented as green, yellow, and cyan ovals, respectively. The five-pointed stars show the charge interactions between the N-terminal subdomain I and the C-terminal domain. The four-pointed stars show the charge interaction between the N-terminal subdomain II and the C-terminal domain. The flexible linker between N- and C-terminal domain is represented by a wiggly cyan line. (A) The large terminase C-terminal domain is ready to bind DNA. (B) The C-terminal domain, when bound to the DNA, brings the DNA closer to the N-terminal domain of the same subunit. Conformational change in the N-terminal domain causes Arg162 to be placed into the ATPase active center in preparation for hydrolysis. (C) Hydrolysis of ATP has rotated the N-terminal subdomain II by about 6° , thereby aligning the charge pairs resulting in an electrostatic attraction that moves the C-terminal domain and the DNA 6.8 Å (equivalent to the distance between two base pairs) closer to the N-terminal domain and into the capsid. (D) ADP and P_i are released and the C-terminal domain returns to its original position. DNA is released and is aligned to bind the C-terminal domain of the neighboring subunit. This figure is adapted with permission from Sun, S., Kondabagil, K., Draper, B., *et al.*, 2008. The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* 135 (7), 1251–1262.

genome size: faster motors are found in phages with larger genomes such that the total packaging time is similar across multiple phages. In comparison to other ASCE ATPases, the T4 large terminase moves along DNA considerably quicker than the fastest known helicase RecBCD, which unwinds DNA at a rate of ~ 1000 bp/s. However, it is slower than the FtsK and SpoIIIE dsDNA translocases, which move at a rate of 5–17 kb/s and ~ 4 kb/s respectively.

The speed of the motor coupled with increased forces inside the capsid during packaging requires an incredibly powerful ATPase to complete the packaging reaction. It is estimated that the power density of the T4 motor is approximately 5000 kW/m^3 , which is twice that of a typical automobile engine. How terminases generate this tremendous power is one of the central questions that has been debated within the field for decades, and several models for terminase mechanism have been proposed to explain this phenomenon (see below).

Terminases are generally composed of an N-terminal ATPase domain and a C-terminal nuclease domain. The two domains are connected by a short linker, allowing for some degree of flexibility between the two halves of the protein. In addition to the ASCE ATPase family features listed in the above section, the large terminase ATPase domain contains a subdomain known as the ‘lid subdomain’ immediately upstream of the linker. The lid subdomain forms the upper portion of the nucleotide binding pocket, and is essential for ATPase activity (Fig. 3). In related AAA+ proteins, the lid plays an important role in propagating nucleotide-dependent conformational changes to adjacent subunits within the ring.

Although the mechanism of translocation in terminases remains unknown, many models have been proposed. Some of these models have been proven incorrect. For example, an early hypothesis that the portal ring rotates within the capsid to drive translocation has been disproven by observations that the portal does not rotate during packaging or in the mature virion. Here we examine several of the current leading models in detail.

The “Inchworm” Translocation Model

One of the primary models for DNA translocation by terminases is an “inchworm” model in which DNA is pulled into the capsid via a spring-like motion (Fig. 5). This model was originally derived from biochemical and structural analysis of the T4 phage large terminase protein, including high-resolution crystal structures of the T4-TerL protein, and a low-resolution cryoEM structure of the T4 procapsid bound to TerL. A modified version of the inchworm hypothesis was proposed later from high-resolution crystal structures of Sf6 phage large terminase. In this model, the ATPase domains bind to the capsid and portal while the nuclease domains grip DNA in the center of the pore. In the inchworm mechanism, a key catalytic residue (Arg162) is repositioned into the active conformation from both DNA binding to the nuclease domain and ATP binding to the active site in the core ATPase domain (Fig. 5(A) and (B)). This induces ATP hydrolysis, which in turn is predicted to rotate the lid (also known as subdomain II) by $\sim 6^\circ$ and move the short flexible linker region between the ATPase and nuclease domains by about 3 Å (Fig. 5(C)). This conformational change is proposed to align a set of ion pairs between the ATPase and the nuclease domains, resulting in a stronger electrostatic attractive force between the two domains. The electrostatic interactions pull the domains towards each other, and this ‘relaxed’ to ‘tensed’ conformational change translocates 2 base pairs of DNA upward into the capsid. During the reset phase of translocation, ADP and/or P_i release from the active site reorganizes the terminase subunit back into the relaxed state via the loss of negative charges from ADP and P_i release (Fig. 5(D)). ATPase subdomain II rotates back, attenuating the electrostatic force between the

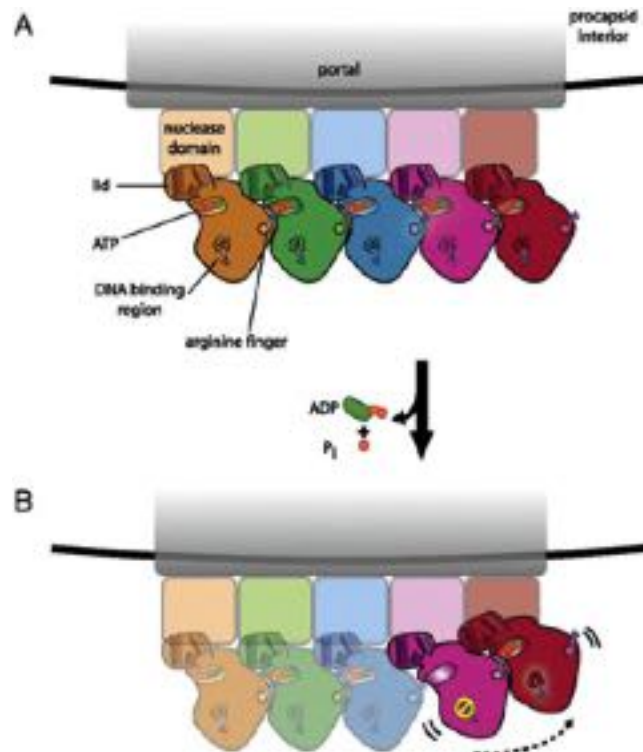


Fig. 6 Proposed ‘lever’ model for the mechanism of DNA translocation by the large terminase. (A) Large terminase ring is shown. The nuclease domains interact with the portal complex (translucent gray rectangle) and possibly the procapsid (black curve). DNA is not shown but interacts with the large terminase through the DNA interaction motif. Each subunit’s lid is bound tightly to the Rossmann fold of the adjacent subunit. (B) Upon ATP hydrolysis and release by the magenta subunit, the lid stays bound to the blue subunit and the Rossmann fold rotates 13° upward. To allow for this movement, the adjacent red subunit must also move in concert with the magenta subunit. To represent that the second site of symmetry breaking is unknown, the other three ATPase domains are faded. After hydrolyzing ATP, the magenta subunit releases DNA to the red subunit to translocate DNA upward through the pore; into the pore of the portal complex; and, ultimately, inside the procapsid. The release of DNA at each cycle by the ATP-hydrolyzing subunit allows for unidirectional DNA translocation. This figure is reprinted with permission from Hilbert, B.J., Hayes, J.A., Stone, N.P., *et al.*, 2015. Structure and mechanism of the ATPase that powers viral genome packaging. *Proceedings of the National Academy of Sciences of the United States of America* 112 (29), E3792–E3799.

ATPase and nuclease domains. The nuclease domain releases its grip on DNA, which is presumably passed to a neighboring subunit preceding or during the release of ADP and/or P_i .

While this model paints a comprehensive picture of the terminase packaging reaction, the proposed mechanism does not explain several observations. First, previous studies found the motor binds the portal in the reverse orientation, with the nuclease domains contacting the portal rather than the ATPase domains. Second, with the ATPase domains splayed radially from each other and bound to portal, the *trans*-activation mechanism required for ringed ASCE ATPase activity cannot occur. Third, the regions predicted to mediate interactions with DNA show particularly low conservation and are only found in T4-like phage. Finally, the observed conformational changes do not fully explain DNA translocation. The small changes in conformation do not match a ~2 bp step size that has been predicted for terminases. No significant conformational changes were observed in the structures of T4-TerL. This may be due to crystal contacts within the active site or the mutation of Walker B motif in the crystallization constructs. Moreover, there is a discrepancy in the Sf6 structures wherein the ATP γ S structure, which should be locked in a ‘ATP-like state’ is identical to the ADP-bound and apo states. This may be due to the fact that the Sf6 TerL was crystallized in the apo state, and then nucleotide ligands were soaked into the pre-assembled crystals. In many cases, the crystal lattice in the apo state does not allow for a ligand-induced conformational change in soaked crystals.

The “Lever” Translocation Model

Another model for viral motor DNA translocation is the lever model (Fig. 6). In this model, a lever-like conformational change in the ATPase domain driven by ATP hydrolysis forms the power stroke of DNA translocation. In contrast to the aforementioned “inchworm” model, the “lever” model positions the ATPase domains as the central hub of the ring, while the nuclease domains protrude radially and attach to portal and capsid. Evidence of this assembly model comes from several crystal structures, molecular docking, and various biochemical analysis to determine the critical residues for motor function.

The identification of a *trans*-acting arginine finger was a critical step in developing the lever model. Identifying this residue led to reinterpretation of the “inchworm” structural model in which the ATPase domains do not contact neighboring subunits. The “lever” model places the ATPase domains into a ring to facilitate *trans*-activated ATP hydrolysis, placing the nuclease domains in contact with portal. Further unbiased pentamer docking experiments of a monomeric crystal structure of the P74–26 ATPase domain positions the predicted arginine fingers precisely for *trans*-activated ATP hydrolysis, providing a second layer of evidence for the inverted structural model. Interestingly, the pentameric model also fits well into the T4 packaging motor electron density map, indicating its potential biological relevance.

Furthermore, the pore of the pentameric docked model is lined with several conserved basic residues, indicating a role in DNA binding. Mutating these residues abrogates TerL binding to DNA. Interestingly, mutation of residues in the nuclease domain has no substantial effect on DNA binding affinity. In fact, the TerL ATPase domain can bind DNA with nearly the same affinity as full-length TerL. Assimilating these findings with the structural model yields a mechanistic model where the ATPase domains grip and translocate DNA in the central pore of the ring, while the nuclease domains bind to the portal and do not interact with the DNA during packaging (Fig. 6(A)).

The lever model predicts that DNA is translocated by a lever-like motion of the ATPase domain while gripping to DNA. These motions were identified by comparing crystal structures of P74–26 phage TerL, which shows significant conformational changes in the apo state or bound to a non-hydrolyzable ATP analog. Comparing the apo and ATP analog bound states, the lid subdomain undergoes a 13° rigid-body rotation. Because modeling and mutational analysis suggests that the lid subdomain forms the primary interaction surface between adjacent ATPase domains, this rotation has a substantial effect on the conformation of the ring. With the lid subdomain stabilized through neighboring contacts, the 13° rotation is transmitted to the Rossmann fold, which is the large domain of the ATPase module. This results in a lever-like movement of the ATPase domain (Fig. 6(B)). This movement is calculated to shift the DNA binding region 8 Å upward towards the capsid and rotate DNA an estimated 2.3°. This is very similar to the 2.5 bp DNA translocation step measured in phi29 motors. It is predicted that upon either P_i or ADP release, the gripping subunit loses affinity for DNA and passes it off to the adjacent subunit. The adjacent subunit will now grip DNA tightly because it is in an ATP-bound state, which has been shown to have the highest affinity for DNA. This handoff would then allow the subunit that hydrolyzed ATP to release ADP and P_i so that it can bind ATP, thereby resetting for the next cycle. By acting in concert with two subunits simultaneously, the DNA can be translocated efficiently. Moreover, the motor produces very high force because the lever arm (the entire Rossmann fold) is quite long.

The DNA “Crunching” or “Compression” Translocation Model

A third DNA packaging model is the DNA “crunching” or “compression” model (Fig. 7). In this mechanistic model, the motor uses ‘torsion’ caused by DNA compression inside the terminase assembly to propel DNA into the capsid. In other words, the DNA itself plays an active role in the force generation process. Specifically, with the motor attached to portal and DNA bound in the center, both portal and the terminase subunits grip the DNA and hydrolyze ATP (Fig. 7(A)). ATP catalysis leads to a brief DNA compression which disrupts the B-form DNA structure, transiently adopting A-form DNA in what is referred to as a “crunched” state (Fig. 7(B)). The return of this compressed state into the B-form state is assumed to cause the DNA translocation power stroke during the packaging cycle (Fig. 7(C)). Evidence for this mechanism lies in several observations. A FRET experiment using a double-labeled fluorescent Y-DNA designed to stall the T4 packaging motor estimated 22%–24% DNA compression within the terminase motor when packaging was stalled. Additionally, subsequent packaging experiments using an intercalating dye, YOYO-1, showed dye release from the DNA during the packaging reaction. This observation is attributed to DNA compression inside the motor, as covalently attached DNA-binding labels are more readily packaged. Observations that the motor does not readily package RNA-DNA hybrids also suggest this mechanism, as RNA-DNA hybrid complexes assume a mostly A-form structure and therefore are not compressible, preventing the scrunched to relaxed transition from driving the translocation power stroke. Two slightly different variants on the scrunching model have been proposed, with one model hypothesizing that the scrunching occurs within the terminase central pore, while the other model hypothesizing that the compression occurs within the portal channel. We note that the compression model is not mutually exclusive with either of the two other models we discuss here; it is possible that force generation is through a combination of the compression model and either the inchworm or lever models.

Termination of Packaging

Phi29 and related bacteriophages replicate their genome as monomers rather than concatemers as most phages do. Thus, packaging is terminated when the entirety of the DNA strand is packaged and no DNA cleavage is required. A 16 Å resolution cryoEM structure of a mature phi29 viral particle shows the DNA packaged inside the prohead with density in the channel of the connector and within the tailtube, both of which were assigned to the right-end gp3. A second cryoEM structure of fiberless phi29 particles at 7.8 Å resolution confirmed the presence of gp3 within the tailtube, indicating that the right-end gp3 does not get fully packaged, while the connector-tailtube channel contains DNA bent in a toroid-like structure. This novel 60 Å-in-diameter highly-bent DNA structure appears to be the result of DNA under extreme pressure. While the function of the toroid structure is unknown, the authors speculate it plays a role in holding the DNA inside the tail during infection. Further studies are necessary to explore this idea.

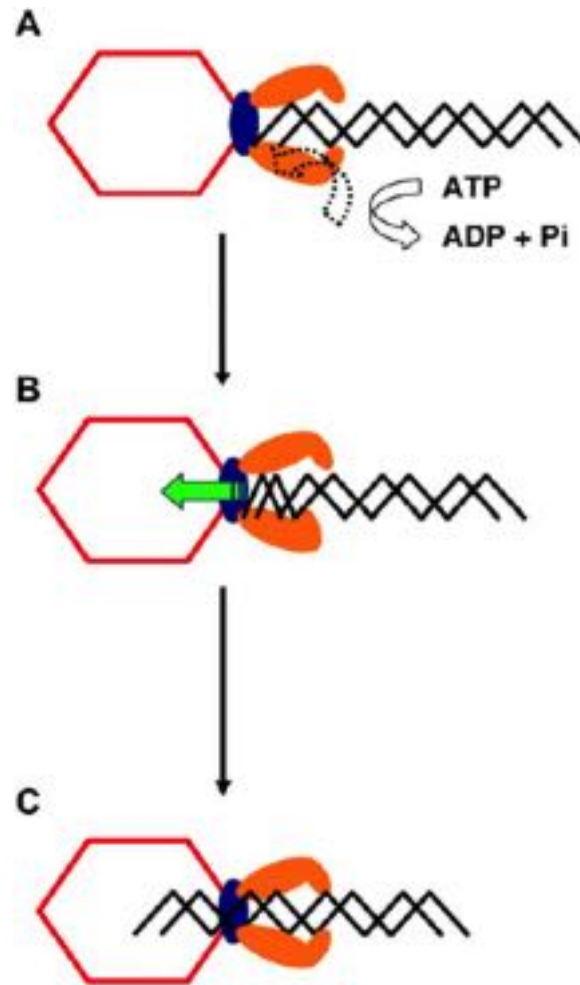


Fig. 7 Phage DNA packaging via a torsional compression mechanism. (A) A fully assembled packaging complex is shown, with the empty prohead (red) carrying the portal ring (blue) shown at one vertex, and the DNA duplex drawn in black. Only two subunits of the large terminase (orange) are shown for clarity. The terminase subunits are sketched with a minor lobe, representing a flexible region that undergoes a conformational change (black dashed arrow) coupled with ATP binding and hydrolysis (white solid arrow). The exact steps and temporal order by which the binding and hydrolysis of ATP is coupled to movement of the large terminase during the reaction cycle is not specified. (B) Directed linear motion of the flexible arm of the large terminase subunit engages the DNA substrate and translocates this towards the prohead. This movement coupled with interaction of the DNA with the portal region causes induced changes to the helical pitch and temporarily stores energy. (C) The stored energy is released by translocation into the prohead (green arrow), restoring the B-form helical repeat. This figure is adapted with permission from Oram, M., Sabanayagam, C., Black, L.W., 2008. Modulation of the packaging reaction of bacteriophage t4 terminase by DNA structure. *Journal of Molecular Biology* 381 (1), 61–72.

Terminases package concatemeric genomes and therefore must cut the DNA at the end of packaging. For viruses in this class, the genomes are typically synthesized by rolling circle DNA replication, resulting in multiple copies of the genome linked in a series (*i.e.*, concatemers). Packaging is accomplished in one of two ways. In the first mechanism, typified by the lambda phage, the terminase cuts a specific site in the genome, resulting in exactly one genome length being packaged into the head (referred to here as ‘unit-length packaging’). In the second mechanism (known as ‘headful packaging’), the volume of packaged DNA triggers the nuclease cleavage mechanism, rather than a specific DNA sequence. The result of a headful packaging mechanism is that each capsid is filled with slightly more than one genome of DNA, creating terminally redundant ends that can be circularly permuted.

For ‘unit-length’ bacteriophage such as lambda phage, the motor nuclease domain cuts at a specific sequence, known as *cosN* in lambda phage. Nicks are made on either side of the DNA strand, leaving a 5′ twelve nucleotide overhang. For the initial cleavage event, a complex with the small terminase protein (gpNu1) and the *E. coli* protein Integration Host Factor (IHF) increase specificity and affinity for the large terminase to the *cosN* site. IHF bends the DNA nearly 180°, positioning two gpNu1 binding sites adjacent to each and presumably facilitating gpNu1 DNA binding, which in turn promotes large terminase nicking at *cosN*. During termination, an upstream site, *cosQ*, and a downstream site, *I2*, are necessary for accurate nicking, and mutations in *cosQ* lead to aberrant termination and formation of particles with uncut DNA protruding from the head. It is proposed that *cosQ* plays a role in properly positioning the terminase/nuclease domains for antiparallel DNA cleavage at the *cosN* site, and that the volume of packaged DNA may trigger its recognition.

In comparison to 'unit-length' packaging bacteriophage, 'headful' bacteriophage do not terminate packaging at specific DNA sequences; instead, termination is determined by the volume of the head. Because headful phages generally package slightly more than one genome into the capsid, the genome is terminally redundant for circular permutation. This process can result in the phenomenon of generalized transduction, in which phages can carry segments of bacterial genome from one cell to another. There are several strategies for headful packaging. For example, phage P22 cleaves after ~2%–10% of the next genome in the concatemer, which results in the next packaging cycle starting with a region downstream from the initial packaging recognition sequence. Other cleavage strategies are seen across different phage, suggesting that termination (and initiation) are controlled idiosyncratically by different phage.

Headful terminases must trigger nuclease activity through a mechanism that senses how full the capsid is. One hypothesis for how this is accomplished relies on the conformational state of the portal protein. In a recent study of the P22 portal ring, the authors found that during packaging the portal adopts an asymmetric assembly that binds tightly to the terminase motor (it is of note that the discovery of this asymmetric state also addressed how a pentameric motor could bind a dodecameric portal, a question that has eluded the field for several decades). However, once the head nears complete fullness, DNA binds around the capsid-enclosed portion of portal. DNA binding in this region triggers the asymmetric portal assembly to adopt a symmetric ring, thus losing affinity for the pentameric motor and presumably discharging it from the capsid head. After portal releases the motor, the motor nuclease domains cleave DNA, severing the packaged DNA from the rest of the concatemer, and terminating the packaging reaction.

Headful packaging mechanisms necessitate strict control to prevent premature cleavage of DNA. The small terminase subunit may play a regulatory role in this process, as it suppresses large terminase nuclease activity (while increasing ATPase activity). How this mechanism works is unclear, and the role of the small terminase during DNA translocation has yet to be fully resolved.

One model of DNA cleavage relies on separate DNA 'binding' and 'cleavage' interfaces of the nuclease domain. Terminases have nuclease domains of the RuvC/RNase-H family that use three acidic active site residues to coordinate two divalent metals (either Mg^{2+} or Mn^{2+}) to catalyze the cleavage reaction. Crystal structures of the T4 nuclease domain show a patch of basic residues far from the active site that are presumed to make electrostatic contacts with the DNA backbone during 'translocation mode'. Once endonuclease activity is triggered, the DNA is transferred through an unknown mechanism to the nuclease active site for cleavage. This putative DNA translocation interface is proposed to regulate the endonuclease activity by holding DNA away from the active site during translocation. However, it was later noted that no positively charged residues from this region are conserved throughout other bacteriophage nuclease domains, and that the nuclease domain has very poor affinity for DNA.

Further structural studies attempt to explain nuclease regulation through nuclease domain conformational changes. Analysis of the SPP1 bacteriophage nuclease domain crystal structure predicts that a conserved extended β -hairpin clashes with DNA bound in an RNase-H-like conformation, and that the loop must reorient upon DNA binding. Heterogeneity of loop positioning in different bacteriophage species and normal mode analysis calculations support this idea, suggesting that loop flexibility may limit active site accessibility for DNA, and thus acts as a nuclease regulation mechanism. Later studies using thermophilic terminases suggest that movement of this loop is not required if the terminase uses a different DNA binding mode. It was suggested that the nuclease domain cleaves DNA in a fashion similar to the thermophilic RuvC resolvase rather than RNaseH. In this model, the β -hairpin and residues surrounding the active site cradle DNA within a groove, positioning it in a desirable conformation for cleavage.

Another suggested mechanism of cleavage regulation is that DNA packaging occurs too quickly for nucleolytic activity to take place. While this is formally possible, motor stalls for extended time periods (seconds to hours) do not result in DNA cleavage, indicating this 'kinetic competition' model is not a primary mechanism for nuclease regulation.

Recently, the 'steric hindrance model' was proposed to explain regulation of terminase nucleolytic activity (Fig. 8). In this mechanism, the nuclease domains bind to portal and capsid during packaging, preventing the nuclease active site from engaging with DNA. Evidence suggests that the ATPase domains are solely responsible for gripping DNA, while the nuclease domains have no measurable affinity, supporting this arrangement. After ejection from portal, two nuclease domains bind and cleave DNA in an antiparallel arrangement, forming a double-stranded break. This rearrangement of the nuclease domains is made possible by a flexible linker connecting to the ATPase domain, allowing the nuclease domains to adopt the correct orientation for DNA cleavage. Once the DNA strands are cleaved, the tail proteins assemble on portal to create a mature virion.

Once packaging is finished, the motor must be replaced with the virus neck and tail. It is not clear how the DNA does not get prematurely released during the transfer of motor for tail. Various mechanisms for maintaining the pressurized DNA have been proposed. For example, the portal could act as a one-way valve. Additionally, the conformation of the DNA itself may help maintain packaged genome. Cryo-electron microscopy studies show the DNA tightly spooled in the capsid, typically with the central axis of the spool coincident with the long axis of the tail. The arrangement of DNA appears to be most homogenous at the periphery of the shell where the genome contacts the capsid directly, with the 'layers' of DNA becoming progressively more heterogeneous towards the center. DNA is often found to be ordered in the portal complex and even within the tail, as if it were locked down but poised for ejection. Typically, the DNA sequence that enters the head first is the last to leave the head upon infection, suggesting a simple unidirectional model for spooling of DNA in the head (T4-like phage are an exception; the first sequence in during packaging is the first sequence out during infection). Moreover, there are often proteins that also reside in or nearby this channel that are ejected soon after infection; these 'ejection proteins' presumably play an important role in early stages of infection, including the process of crossing the periplasm. It remains a mystery whether these proteins hold DNA in the capsid until infection or the motor plays a role in packaging these ejection proteins inside the capsid.

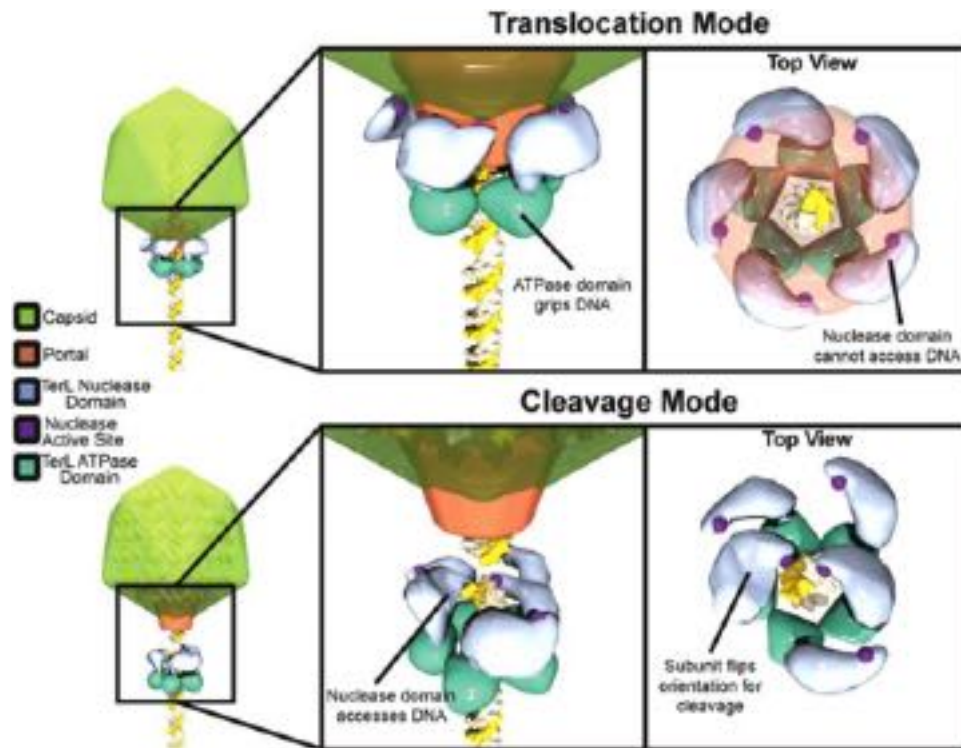


Fig. 8 Proposed 'steric hindrance' model for nuclease regulation. During 'translocation mode' the nuclease domain active site is sequestered from DNA by interactions of the large terminase with portal and capsid, preventing premature cleavage. The ATPase domain serves as the sole surface for gripping DNA during packaging. Upon completion of packaging the large terminase enters 'cleavage mode'. The large terminase loosens its attachment to the portal and capsid, releasing the inhibition of the nuclease domains. The ATPase domains remain tightly bound to DNA. The nuclease domains rearrange to cleave each of the antiparallel DNA strands. Although depicted as a blunt cut, cleavage could also leave overhangs depending on how both nuclease domains engage DNA. This figure is reprinted with permission from Hilbert, B.J., Hayes, J.H., Stone, N.P., Xu, R.-G., Kelch, B.A., 2017. The large terminase DNA packaging motor grips DNA with its ATPase domain for cleavage by the flexible nuclease domain. *Nucleic Acids Research* 45 (6), 3591–3605.

Outlook and Future Studies

Despite decades of research, there remain numerous questions as to how viral genome packaging motors processively translocate DNA. High-resolution structures of actively packaging motors are necessary to elucidate the DNA translocation mechanism. Visualizing the conformational transitions, intersubunit interactions, and DNA gripping contacts will shed light on how the motor generates the power to package DNA against the immense back pressure in the capsid. Structures of both phi29 and terminase motors will determine their similarities and differences, as well as address how force is generated.

While phi29 genome packaging has been investigated to fine detail, less is known about the terminase motor DNA translocation cycle. For instance, do terminases also hydrolyze ATP in a sequential order? Do they follow a similar chemomechanical cycle? Given that there are considerable differences in DNA translocation speeds of phi29 and terminases, it is possible that the two types of motors use different regulation strategies. Furthermore, it appears that the *trans*-acting arginine finger is different in terminases and phi29 motors. If so, how do two proteins with similar ATPase domain topologies employ different mechanisms? Having a clearer understanding of the terminase DNA translocation cycle will provide insight into these questions.

There are also many unknown regulation steps within the endonuclease reaction catalyzed by terminases. For example, how the motor senses when to cleave for both headful and unit-length filling mechanisms is unknown. For headful terminases, the portal has been implicated as a sensor, but whether the large terminase plays an active sensing role is still inconclusive. In unit-length bacteriophage such as lambda, the initial cleavage event has been elucidated in great detail, but how the termination cleavage occurs in the absence of gpNu1 and IHF is less clear. Some evidence supports the idea that lambda may use a headful-like sensing mechanism to trigger *cosN* cleavage, but whether this is regulated by the portal, the terminase itself, or the rate of DNA translocation at the end of packaging is undetermined. Finally, once the motor is released from portal, what prevents the genome from immediately ejecting from the capsid before the pre-assembled tail binds? Answering these questions will undoubtedly require synthesizing structural studies with careful biochemical and single-molecule assays.

Moving forward, the study of motor structure and mechanism will likely lead to new biotechnological applications. For example, the prNA of phi29 has already been used as a therapeutic delivery device. Are there applications for the motor itself? One could imagine

roles in nanomaterials and targeted delivery of biological cargo. As an example, the portal was recently engineered as a nanopore for sensing various biomolecules. Future studies of the motor may lead to a plethora of applications in biotechnology.

Further Reading

- Burroughs, A.M., Iyer, L.M., Aravind, L., 2007. Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dynamics* 3, 48–65.
- Chistol, G., Liu, S., Hetherington, C.L., *et al.*, 2012. High degree of coordination and division of labor among subunits in a homomeric ring ATPase. *Cell* 151 (5), 1017–1028.
- Erzberger, J.P., Berger, J.M., 2006. Evolutionary relationships and structural mechanisms of AAA+ proteins. *Annual Review of Biophysics and Biomolecular Structure* 35, 93–114.
- Hilbert, B.J., Hayes, J.A., Stone, N.P., *et al.*, 2015. Structure and mechanism of the ATPase that powers viral genome packaging. *Proceedings of the National Academy of Sciences of the United States of America* 112 (29), E3792–E3799.
- Lokareddy, R.K., Sankhala, R.S., Roy, A., *et al.*, 2017. Portal protein functions akin to a DNA-sensor that couples genome-packaging to icosahedral capsid maturation. *Nature Communications* 8, 14310.
- Mao, H., Saha, M., Reyes-Aldrete, E., *et al.*, 2016. Structural and molecular basis for coordination in a viral DNA packaging motor. *Cell Reports* 14 (8), 2017–2029.
- Moffitt, J.R., Chemla, Y.R., Athavan, K., *et al.*, 2009. Intersubunit coordination in a homomeric ring ATPase. *Nature* 457 (7228), 446–450.
- Oram, M., Sabanayagam, C., Black, L.W., 2008. Modulation of the packaging reaction of bacteriophage t4 terminase by DNA structure. *Journal of Molecular Biology* 381 (1), 61–72.
- Smits, C., Chechik, M., Kovalevskiy, O.V., *et al.*, 2009. Structural basis for the nuclease activity of a bacteriophage large terminase. *EMBO Reports* 10 (6), 592–598.
- Sun, S., Kondabagil, K., Draper, B., *et al.*, 2008. The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. *Cell* 135 (7), 1251–1262.
- Tafoya, S., Liu, S., Castillo, J.P., *et al.*, 2018. Molecular switch-like regulation enables global subunit coordination in a viral ring ATPase. *Proceedings of the National Academy of Sciences of the United States of America* 115 (31), 7961–7966.
- Zhao, Z., De-Donatis, G.M., Schwartz, C., *et al.*, 2016. An arginine finger regulates the sequential action of asymmetrical hexameric ATPase in the double-stranded DNA translocation motor. *Molecular and Cellular Biology* 36 (19), 2514–2523.

Biophysics of DNA Packaging

Joshua Pajak and Gaurav Arya, Duke University, Durham, NC, United States

Douglas E Smith, University of California, San Diego, La Jolla, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

ASCE The additional strand, conserved E (glutamate) superfamily of ATPases. A classification of proteins based on sequence, structure, and function. Terminases and other viral DNA packaging ATPases are members of this superfamily.

ATPase A protein that catalyzes ATP hydrolysis.

Procapsid The viral capsid prior to complete maturation of the virion.

Terminase A viral protein which packages DNA and performs endonuclease activity.

Introduction

A crucial step in virus assembly is the encapsulation of viral genome into the procapsid. Different viruses have evolved different strategies by which they package their genome. In many double-stranded DNA (dsDNA) viruses, such as some families of bacteriophages, DNA packaging is accomplished by packaging ATPases, powerful molecular motors that convert energy from ATP hydrolysis into mechanical work. This energy from ATP hydrolysis is used to overcome large internal forces arising from electrostatic self-repulsion, entropy loss, and DNA bending rigidity, which oppose DNA confinement within the capsid.

The packaging ATPase is a dimer of homo-oligomeric rings. One ring is comprised of large packaging proteins, and the second ring is comprised of small packaging proteins. The small packaging complex is responsible for recognizing the viral genome at a sequence-specific “pac site” to ensure that the virus only packages its own genome and not stray genome from the host bacteria. In some cases, such as the $\phi 29$ phage, the small packaging proteins are covalently attached to the end of the viral DNA, and recognition is attained via the assembly of the large and small packaging complexes. The large packaging motor protein is responsible for the ATPase activity which provides the necessary energy to translocate the genome into the viral procapsid. The packaging motors have evolved various mechanisms for initiating packaging, translocating DNA, and terminating packaging. The mechanisms involved in DNA translocation are especially elaborate, where mechanical actions such as DNA gripping and un-gripping (and/or steric pushing of the DNA), force generation, and coordination of the multiple motor subunits must be coupled to the ATP binding, bond cleavage, and product release steps of the ATP-hydrolysis cycle.

While there is currently no unified model that explains all aspects of DNA packaging for all phages, our understanding of packaging ATPases, especially their molecular structure, packaging characteristics, and packaging mechanism, has improved significantly through knowledge primarily gained from the three well-studied systems of the $\phi 29$, T4, and λ phages, which will mainly be highlighted here.

Molecular Structure

Viral DNA packaging proteins are members of the Additional Strand, Conserved Glutamate (ASCE) superfamily of ATPases, and are related to other ASCE ATPases such as ATP Binding Cassette (ABC) and ATPases Associated with diverse cellular Activities (AAA+) proteins. ASCE ATPases contain a conserved P-loop/Walker A (WA) motif involved in binding ATP, and a Walker B (WB) motif which helps catalyze ATP hydrolysis. Based on mechanisms deduced for other ATPases, catalysis is thought to be specifically dependent on a conserved glutamate residue located immediately downstream of the WB motif. Many packaging proteins also have endonuclease activity necessary to initiate and terminate the packaging process in the case where viral DNA replication results in catenated strings of multiple genomes; As such, these proteins are called “terminases”. Terminases have two distinct domains: the ATPase and nuclease domains, which are connected by a flexible linker. The T4 and λ phages use terminase proteins. The $\phi 29$ motor does not perform endonuclease activity and its motor enzymes are referred to by the more general term “packaging ATPase.”

The molecular structures of the monomer subunits from several packaging ATPases have been solved either in part or in whole via X-ray crystallography. The first near-complete monomer structure to be solved was from the T4 phage in 2008, followed by those from the Sf6 and D6E phages in 2013 and 2017. The structures of the ATPase domain of the P74-26 phage terminase protein and a partial ATPase from the $\phi 29$ phage were also solved, and so have the structures of the nuclease domains of the P74-26, SPP1, P22, G20c, and RB49 phage terminase proteins. There is significant structural homology across all of the ATPase and nuclease domains, although the relative positioning of the nuclease domain with respect to the ATPase domain is drastically different in the solved T4, Sf6, and D6E terminase structures (Fig. 1). The homology of independent domains suggests that some aspects of the packaging mechanisms may be conserved across all bacteriophages; however, the reason for the observed differences in domain positions across the three proteins is unclear.

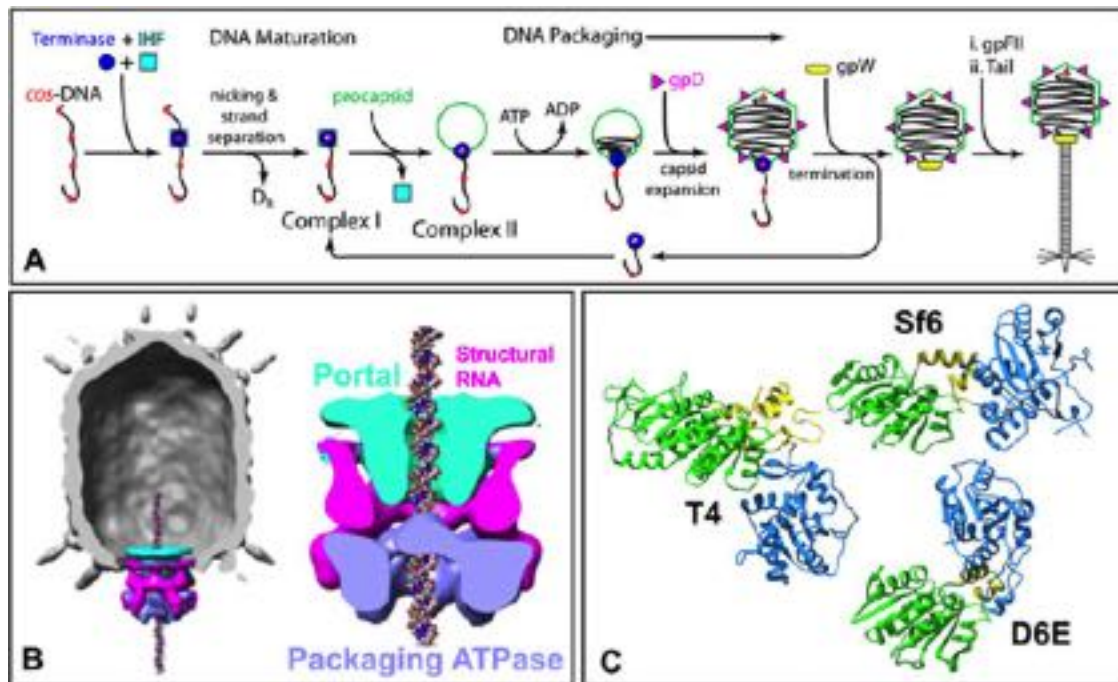


Fig. 1 (A) The pathway of virion maturation for the λ phage. (B) Cryo-EM data of the ϕ 29 packaging motor with modeled DNA. The portal (or connector) density is colored green, the structural RNA density is colored magenta, and the packaging ATPase density is colored purple. (C) Three large terminase structures from the T4 (PDB: 3CPE), Sf6 (PDB: 5IDH), and D6E (PDB: 5OE8) phages are depicted to highlight the unique domain arrangements found in each crystal structure. The ATPase domain is colored green, the lid subdomain is colored yellow, and the nuclease domain is colored blue. Reproduced from (A) Andrews, B.T., Catalano, C.E., 2013. Strong subunit coordination drives a powerful viral DNA packaging motor. *Proceedings of the National Academy of Sciences* 110 (15), 5909–5914. doi:10.1073/pnas.1222820110. (B) Morais, M.C., Koti, J.S., Bowman, V.D., *et al.*, 2008. Defining molecular and domain boundaries in the bacteriophage ϕ 29 DNA packaging motor. *Structure* 16 (8), 1267–1274.

The packaging ATPase complex assembles at a unique vertex of the viral procapsid, which houses the portal protein complex, a 12-mer ring structure that forms the pore through which DNA passes during translocation. It is known that the large packaging ATPases dock to the portal complex. However, it is not known which domain of terminase proteins binds to the portal. Cryo-EM data from the T4 and T7 bacteriophages suggest that the ATPase domain binds to the portal, while Fluorescence Resonance Energy Transfer (FRET) and some genetic and biochemical data suggest that the nuclease domain binds to the portal. This is an important issue to resolve, as the two orientations imply different mechanisms for certain aspects of motor function, such as DNA-gripping residues and force-generation mechanisms.

DNA Packaging Characteristics

Biophysical experiments, particularly single DNA molecule manipulation experiments with optical tweezers, have produced a wealth of quantitative data on the packaging process (Fig. 2). These experiments involve trapping and manipulating dielectric microspheres using focused laser beams. In one approach, used to study ϕ 29 and T4, an empty phage procapsid, with its motor complex attached, is tethered via antibodies to one of the microspheres while its genome is attached by one end via biotin-streptavidin labels to a second microsphere. Packaging is initiated when the laser brings these two spheres close enough for the packaging motor on one sphere to grab the free end of the DNA on the other sphere. In a second approach, used to study λ , a DNA-motor complex attached to one sphere docks onto an empty procapsid bound to a second sphere to initiate packaging. Subsequent packaging of DNA exerts a force on the DNA and the attached microspheres, and the magnitude of this force is determined by measuring the deflections (change in momentum) of the trapping laser beams. A commonly used measurement technique is the “force-clamp” in which a feedback control system is used to change the separation between the two trapped beads so as to apply a small constant stretching force to the DNA as it is being packaged, enabling the length of the DNA packaged to be tracked continuously as a function of time. These measurements have high sensitivity and can resolve force changes < 1 pN and displacements as small as ~ 1 bp in some experiments. Motor characteristics that have been determined include instantaneous and average packaging rates, motor pausing and slipping, and in some cases incremental translocation step sizes.

Packaging Speed

Viral DNA packaging is highly processive (Table 1). Single-molecule experiments showed that the ϕ 29 phage packages at a maximum rate (at $\sim 23^\circ\text{C}$) of ~ 200 base pairs per second (bp/s), the λ phage packages as fast as ~ 800 bp/s, and the T4 phage can package up to

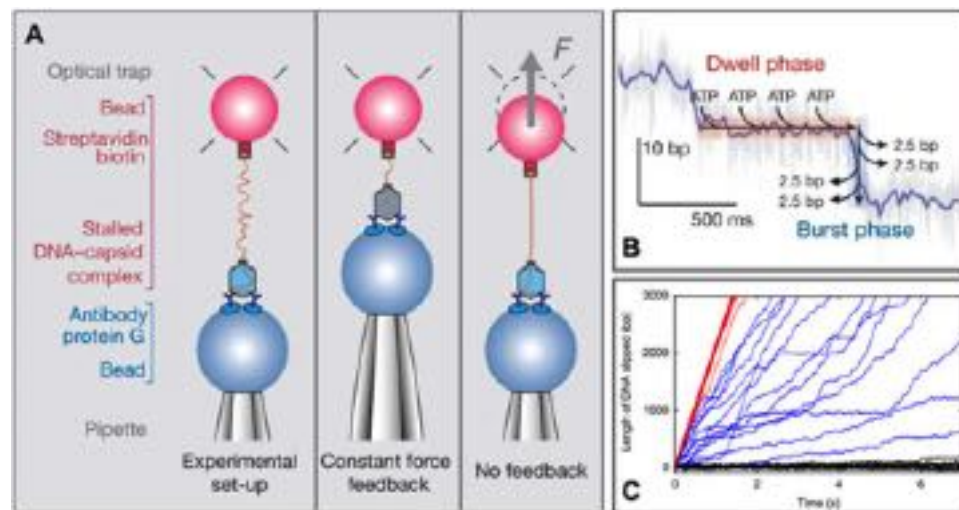


Fig. 2 Single molecule experimental set-up and examples of data collected. (A) A schematic of the single molecule experimental set-up using optical tweezers to study the $\phi 29$ phage packaging process. In these early studies the bottom microsphere was held with a narrow glass pipette. In more recent studies both microspheres were trapped with optical tweezers. (B) DNA translocation distance vs. time data gathered on the $\phi 29$ system displaying dwell and burst phase behavior. During the dwell phase the motor does not package DNA while it waits for all subunits to be loaded with ATP. During the burst phase, four subunits hydrolyze ATP in quick succession to translocate 10 bp of DNA. (C) Nucleotide-dependent slipping of the T4 motor; the black lines are with a high concentration of non- (or slowly) hydrolyzable ATP analog, blue lines are with high ADP concentration, and red lines are with no nucleotides. Reproduced from (A) Smith, D.E., Tans, S.J., Smith, S.B., *et al.*, 2001. The bacteriophage $\phi 29$ portal motor can package DNA against a large internal force. *Nature*, 413 (6857), 748–752. doi:10.1038/35099581 with permission from Springer publishing. (B) Moffitt, J.R., Chemla, Y.R., Aathavan, K., *et al.*, 2009. Intersubunit coordination in a homomeric ring ATPase. *Nature*, 457 (7228), 446–450. doi:10.1038/nature07637 with permission from Springer publishing. (C) Ordyan, M., Alam, I., Mahalingam, M., Rao, V.B., Smith, D.E., 2018. Nucleotide-dependent DNA gripping and an end-clamp mechanism regulate the bacteriophage T4 viral packaging motor. *Nature Communications*, 9 (1). doi:10.1038/s41467-018-07834-2 with permission from Springer publishing.

Table 1 Packaging speed and genome length

Bacteriophage	Maximum packaging speed (bp s ⁻¹)	Genome length (kbp)
$\phi 29$	200	19
λ	800	48.5
T4	2000	170

2000 bp/s. These rates appear to be roughly proportional to the length of the virus' genome: the $\phi 29$ phage genome is ~ 19 kilo base pairs (kbp), and the λ and T4 phage genomes are ~ 48.5 and ~ 170 kbp long, respectively. This proportionality ensures that each virus can package its genome within a relatively short window of time during the infection of the host cell.

The experiments also showed that the packaging speeds of the $\phi 29$, T4, and λ phages decrease with increasing applied load force. Similarly, packaging speed was found to decrease during packaging with the amount of DNA packaged in the capsid, suggesting that internal forces which oppose DNA confinement also slow down the motor. Recent studies of $\phi 29$ indicate that both internal load force and allosteric regulation of ATP binding regulate the motor speed in response to the amount and conformation of packaged DNA. Single-molecule studies of the λ phage revealed a sharp dip in packaging speed when $\sim 30\%$ of the genome was packaged. The interpretation of this feature is that it is due to capsid expansion which occurs in that system induced by a build-up of internal pressure due to the confined DNA, which temporarily reduces the load force acting on the motor.

The $\phi 29$ motor has been best characterized in terms of the kinetics of the mechano-chemical cycle, and because the motor translocates relatively slowly very high-resolution measurements of DNA translocation were possible. High stability optical tweezers measurements at high load forces revealed that $\phi 29$ translocates DNA in incremental steps of 2.5 bp per ATP hydrolyzed. These 2.5 bp steps are further grouped into 10 bp "bursts", during which four subunits package rapidly in sequence (Fig. 3). This sequence of rapid firing is termed the "burst phase". After the burst phase, the motor continues to grip DNA but does not package it. During this "dwell phase", the motor waits for ATP to bind to every subunit that had catalyzed ATP to package DNA during the burst phase. Interestingly, this implies that the multiple motor subunits are tightly coordinated and that one of the five subunits does not participate in DNA translocation and is understood to regulate the occurrence of bursts. It is currently unknown whether other phages also use a strictly coordinated dwell-burst packaging mechanism, as those phages package too quickly to resolve the dwell phase if it were to be present.

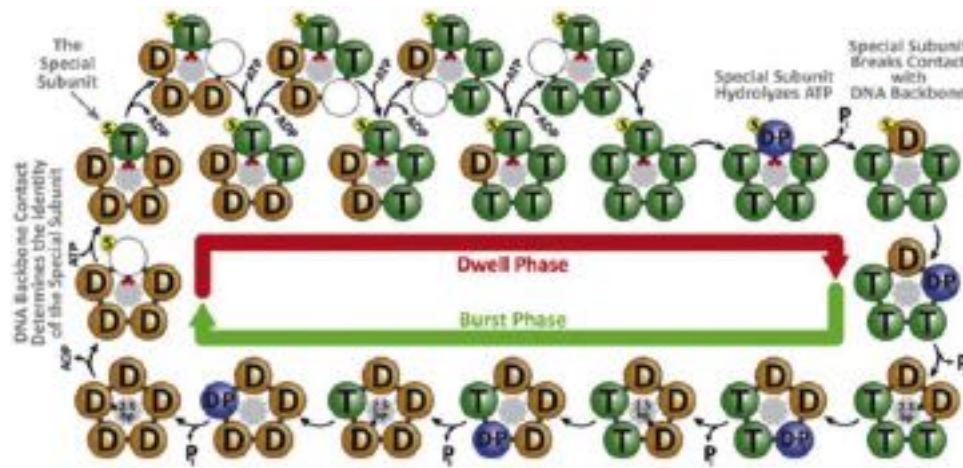


Fig. 3 Kinetic model for the dwell phase/burst phase cycle of the $\phi 29$ packaging process. ATP-bound subunits are labeled as T, ADP/Pi-bound subunits are labeled as D-P, and ADP-bound subunits are labeled as D. Electrostatic contacts with the DNA backbone are indicated in red. The “Special Subunit” grips DNA, but does not translocate. Reproduced from Chistol, G., Liu, S., Hetherington, C.L., *et al.*, 2012. High degree of coordination and division of labor among subunits in a homomeric ring ATPase. *Cell*, 151 (5), 1017–1028. doi:10.1016/j.cell.2012.10.031 with permission from Elsevier.

Force Generation

Viral DNA packaging motors are some of the strongest molecular motors identified in biology. Single-molecule studies on the $\phi 29$, λ , and T4 phages revealed that all three phages could overcome applied load forces in excess of 50–60 pN. For comparison, the skeletal muscle protein myosin can generate ~ 2 pN of force. In fact, the 50–60 pN measurements represent a lower bound on the generated force; at higher forces the procapsids were pulled off the microspheres and at ~ 65 pN the DNA helix sharply deforms. At high prohead filling levels internal forces contribute an estimated ~ 20 –30 pN and, combined with applied external forces ranging up to ~ 50 pN, suggest that these motors can produce forces of at least ~ 80 pN. ATP hydrolysis supplies ~ 18 kcal mol $^{-1}$ (~ 130 pN nm) of free energy. If a motor were to convert this energy into useful work with 100% efficiency, almost 155 pN of force would be available to translocate DNA by the 2.5 bp (~ 8.5 Angstrom) step size observed experimentally. Thus, the predicted packaging force magnitude of ~ 80 pN is within the realm of possibility, provided the motor achieves $\sim 50\%$ efficiency.

Pausing and Slipping

Although these motors are highly processive and capable of producing large forces, the packaging process is not without defects. Pausing and slipping (DNA leaving the procapsid) are induced at high external loads and if ATP concentration is lowered. Interestingly, pausing and slipping events are not proportionally induced in each packaging motor. For example, under high load force, the λ phage packaging motor pauses more but slips less than the $\phi 29$ motor. While packaging the fastest overall, the T4 phage has the most variable packaging speed.

Additionally, pausing and slipping is highly dependent upon the availability of nucleotides in the bulk environment. For instance, slippage is more frequent in solutions with low ATP concentration, while pausing is more frequent in conditions with small amounts of non-hydrolyzable ATP analog. Recent single-molecule experiments on the T4 phage in an ADP-rich environment showed that the motor appears to fluctuate between apo-like, ATP-like, and intermediate states.

DNA Packaging Mechanisms

The molecular mechanisms by which these enzymes package DNA are becoming better understood thanks to a combination of increased availability of high-resolution crystal structures, biochemical analysis of mutant enzymes, and computational modeling. These mechanisms include mechanochemical coupling, motor-substrate interactions, force generation, and motor coordination. However, these mechanisms are not mutually exclusive, and often perturbing one aspect has a cascade effect which perturbs other mechanisms.

Mechanochemical Coupling

Single molecule experiments in which $\phi 29$ translocation was studied with varying ATP concentration, varying applied force, and with mixtures of ATP and non-hydrolyzable ATP analog demonstrates that ATP binding does not induce the DNA translocation

step. Instead, it is likely coincident with product (inorganic phosphate) release. Because the nonhydrolyzable analog induces pausing and not slipping, ATP binding likely “cocks” the motor by causing the enzyme to grip DNA before translocation. Rapid solution exchange experiments with T4 also showed that the motor usually tightly grips the DNA when a non-hydrolyzable ATP analog is bound, while the DNA slips out rapidly when there is no nucleotide. Thus, it is evident that small changes within the ATP binding pocket must be sensed by the protein and then somehow converted into mechanical motion. The connection between the chemical steps of ATP hydrolysis and the mechanical motion of the motor is termed “mechanochemical coupling.” The mechanisms involved are generally expected to have features in common with related ATPases, such as AAA+ motors.

One such recently identified mechanism has been termed the “arginine toggle.” A unique structural element of packaging ATPase enzymes relative to other ASCE proteins is the strong conservation of an arginine residue at the 3 or 4 position of the canonical WA motif. Mutational studies of the λ , T4, P74-26, and D6E phage terminase proteins have provided evidence suggesting that the WA arginine is crucial for ATPase catalytic activity as well as mechanochemical coupling. Further computational studies of the T4, P74-26, and Sf6 terminase proteins predicted that the WA arginine toggles in and out of the binding pocket in response to ATP binding, and this toggling motion is part of the signal transduction pathway which informs the rest of the protein that ATP has bound. In this sense, it appears that the “arginine toggle” is conceptually analogous to the “sensor II motif” arginine found in related AAA+ motors.

A related family of ATPase proteins, helicases, were identified to use a “Q-motif” which also couples ATP binding to conformational change. This motif typically consists of a “YQ” amino acid pairing upstream of the WA motif responsible for binding the adenine ring of ATP. Sequence alignment suggested that the λ phage terminase enzyme contained a Q-motif, namely 46-YQ-47. Single molecule experiments of a Y46F variant displayed slower packaging rate and increased slipping, as well as higher sensitivity to applied force. This suggests that the Q-motif is responsible in part for mechanochemical coupling as well as force generation.

Motor-substrate Interactions

Many proteins that bind to DNA do so at specific sites dictated by short DNA sequence motifs. This allows for increased specificity and increased binding affinity. However, since these motors must package entire genomes, they must grip onto a wide variety of sequence motifs. Thus, sequence specific interactions are not employed by these motors, but instead they rely on non-specific interactions. Such interactions can be classified into two broad categories: electrostatic and steric. Although each base of DNA varies in hydrogen-bonding potential, each base is attached to an identical negatively-charged phosphate backbone. Thus, positively-charged residues located in the pore of the motor can non-specifically bind with these negatively-charged groups via electrostatic interactions. Further, the double-helix structure of DNA creates a major-groove and minor-groove which bulky protrusions could fit into, such that the motor grips DNA via steric interactions.

To test whether the motor primarily grips via electrostatic or steric interactions, single molecule experiments have been performed with structural modifications introduced to the substrate DNA. 10 bp sections of methylated DNA, which neutralizes the negative charge of the phosphate backbone, were able to be packaged with little change in kinetics compared to packaging regular DNA. However, DNA with 11 bp sections of methylated DNA often caused stalling or lengthy pauses in packaging. This suggests that the motor makes crucial electrostatic contacts every 10 bp of DNA. This behavior fits with known structural features as follows: the motor has pentameric stoichiometry and packages in 2.5 bp steps. One helix turn of DNA is 10.5 bp. Thus, every 10 bp of DNA approximately aligns the charged phosphate backbone with the same subunit.

Although the 11 bp insertions strongly affected packaging dynamics, the motor was able to accommodate insertions up to 30 bp long. This indicates that while electrostatic interactions are important, non-specific steric interactions are sufficient for DNA translocation. To test whether electrostatic interactions alone are sufficient for DNA translocation, DNA with 10 bp abasic phosphate backbone insertions were introduced to the system. This modification was also able to be packaged under low applied load, indicating that electrostatic interactions are sufficient at low forces. However, other modifications such as single-strand insertions, unhybridized bulges, and polymer insertions were also able to be packaged. Taken together, this evidence demonstrates that both non-specific steric interactions and attractive electrostatic interactions play roles in the translocation process.

There remains debate whether the ATPase domain or the nuclease domain of the terminase proteins interact with DNA during translocation. The ϕ 29 motor does not contain a functional nuclease domain, which suggests that the ATPase domain may primarily interact with the DNA. However, it is unclear if this is consistent across all packaging proteins. A model based on T4 cryo-electron microscopy data implicates the nuclease domain as gripping the DNA during translocation steps. However, biochemical data from mutant P74-26 terminase proteins identified that crucial DNA gripping residues reside in the ATPase domain and not the nuclease domain. Thus, it is unclear whether one model can explain protein-DNA interactions for all packaging motors.

Force Generation

Several models have been proposed to explain how these molecular machines generate the incredible forces necessary for complete genome packaging. Most models suggest that the force is generated by conformational changes within the ATPase packaging proteins that actuate a “lever” motion (Fig. 4), but one model suggests that the DNA segment threaded through the motor channel itself could be involved in force generation.

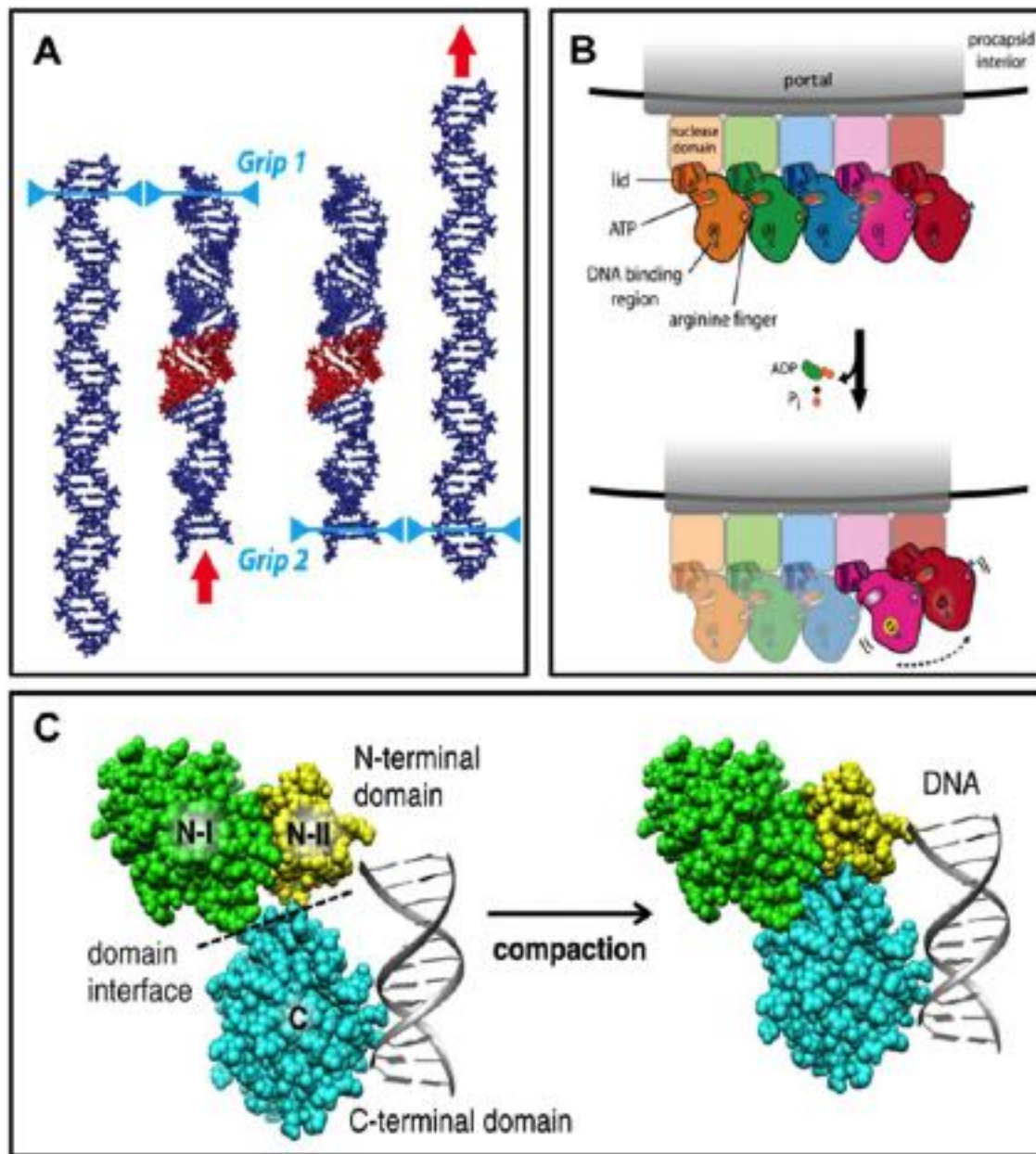


Fig. 4 Three proposed models for the motor force generation mechanism. (A) The scrunchworm model, indicating the motor gripping events needed to facilitate unidirectional motion of the substrate DNA. (B) A model proposed based on crystal structures of the P74-26 terminase enzymes in apo and ATP-bound states. The ATPase domain (N-terminal) grips DNA, and the nuclease domain (C-terminal) attaches to the portal. Rotation about the lid subdomain drives translocation. (C) The extended-to-compact transition model is depicted, proposed based on cryo-EM and X-ray data for the T4 motor. The nuclease domain (C-terminal) grips DNA and the ATPase domain (N-terminal) attaches to the portal. Alignment of charged residues across the domain interface drives the compaction process and DNA translocation. Reproduced from (A) Waters, J.T., Kim, H.D., Gumbart, J.C., Lu, X.J., Harvey, S.C., 2016. DNA Scrunching in the packaging of viral genomes. *Journal of Physical Chemistry B* 120 (26), 6200–6207. doi:10.1021/acs.jpcb.6b02149. (B) Hilbert, B.J., Hayes, J.A., Stone, N.P., *et al.*, 2015. Structure and mechanism of the ATPase that powers viral genome packaging. *Proceedings of the National Academy of Sciences*, 112 (29), E3792–E3799. doi:10.1073/pnas.1506951112. (C) Migliori, A.D., Smith, D.E., Arya, G., 2015. Molecular interactions and residues involved in force generation in the T4 viral DNA packaging motor. *Journal of Molecular Biology*, 426 (24), 4002–4017. doi:10.1016/j.jmb.2014.09.023.

The “scrunchworm” model suggests that the translocation machinery primarily grips and releases DNA while the DNA is compressed and expanded. Coordination between gripping/releasing and compression/expansion could result in directional motion of the DNA. In this model, the force generated is a restorative force stored within the DNA itself and the motor is speculated to induce the DNA to go into the compressed conformation. Support for this model comes from molecular simulations demonstrating that DNA is expanded/compressed within the portal complex. Further, dsRNA and RNA-DNA heteroduplexes could not be packaged by the T4 motor, presumably because they cannot undergo the same compression/expansion as dsDNA.

On the other hand, more conventional models assume that the force to drive translocation is generated inside the motor proteins. One such model proposed for the T4 motor posits that the alignment and misalignment of charged residue pairs at the ATPase/nuclease domain interface drives translocation. Alignment of the charged residue pairs induces compaction, and misalignment induces expansion. If compaction occurred while the protein gripped DNA and expansion occurred while the protein did not grip DNA, this would result in unidirectional movement, and thus packaging. In this model, mechanochemical coupling of the ATP hydrolysis cycle coordinates the alignment and misalignment of the charged residue pairs. This model is supported by single molecule studies of motor protein mutants in which charged amino acids at the interface were substituted for residues of opposite charge. It was found that these residue changes lower translocation velocity and force generated by the motor.

Alternative models have been proposed for the P74-26 and Sf6 packaging motors based on solved crystal structures of the monomeric terminase subunits. In both cases the lid subdomain (a small motif within the ATPase domain) of the ATP-bound structure is rotated relative to the apo state. In these models, DNA is gripped by either the ATPase domain or a combination of the ATPase and nuclease domains. The rotation of the lid subdomain powers DNA translocation and passes DNA from one subunit to the next subunit, and mechanochemical coupling of the ATP hydrolysis cycle coordinates the lid subdomain rotation.

Motor Coordination

As mentioned above, the ϕ 29 motor has been shown to be highly coordinated overall, exhibiting a “dwell” phase during which no ATP is hydrolyzed, and a “burst” phase where four subunits consecutively hydrolyze ATP and translocate DNA. Coordination within the burst phase is beneficial for motor function. For instance, if two subunits attempted to translocate DNA simultaneously, they might spend two ATP hydrolysis events to translocate DNA 2.5 bp. If those two subunits instead were to translocate DNA consecutively, they would spend the same two ATP hydrolysis events to translocate DNA 5 bp. Thus, it is in the motor’s best interest to coordinate its subunits’ actions.

One identified mechanism for inter-subunit communication is a *trans*-acting “arginine finger.” The arginine finger protrudes from one subunit into a neighboring subunit’s ATP-binding pocket. This allows information about an ATP binding event in one subunit to be communicated to its neighbor, so that their actions can be coordinated. Biochemical studies have assigned the *trans*-acting arginine finger in several packaging motors, such those from the ϕ 29, P74-26, and D6E phages. Although sequence alignments suggest that the T4 packaging motor could also contain a *trans*-acting arginine finger, a pentameric model of the T4 packaging motor based on cryo-EM reconstruction places each subunit’s ATP-binding pocket distal from its neighbors, arguing against such a mechanism. Additionally, the Sf6 packaging motor does not appear to contain an arginine at the sequence alignment predicted site but it is possible that it utilizes a deviant lysine to accomplish the same goal.

Further Reading

- Black, L.W., 2015. Old, new, and widely true: The bacteriophage T4 DNA packaging mechanism. *Virology* 479, 650–656.
- Chemla, Y.R., Smith, D.E., 2012. Single-molecule studies of viral DNA packaging. In: Rossmann, M., Rao, V.B. (Eds.), *Viral Molecular Machines*. Advances in Experimental Medicine and Biology, vol 726. Boston, MA: Springer.
- Rao, V.B., Feiss, M., 2015. Mechanisms of DNA packaging by large double-stranded DNA viruses. *Annual Review of Virology* 2, 351–378.
- Tafuya, S., Bustamante, C., 2018. Molecular switch-like regulation in motor proteins. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373 (1749), (20170181).

Energetics of the DNA-Filled Head

Alex Evilevitch, Department of Experimental Medical Science, Lund University, Lund, Sweden

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Capsid The protein shell of a virus. It consists of oligomeric structural subunits made of proteins called protomers.

Enthalpy A property of a thermodynamic system, is equal to the system's internal energy plus the product of its pressure and volume. In a system enclosed so as to prevent mass transfer, for processes at constant pressure, the heat absorbed or released equals the change in enthalpy.

Entropy Entropy is an extensive property of a thermodynamic system. It is closely related to the number of microscopic configurations that are consistent with the macroscopic quantities that characterize the system.

Metastability A stable state of a dynamical system other than the system's state of least energy.

Osmolyte Compounds that influence the properties of biological fluids. Their primary role is to maintain the integrity of cells by affecting the viscosity, freezing point, and ionic strength of the aqueous solution.

Osmotic pressure The minimum pressure which needs to be applied to a solution to prevent the inward flow of its pure solvent across a semipermeable membrane. It is also defined as the measure of the tendency of a solution to take in pure solvent by osmosis.

Persistence length A basic mechanical property quantifying the stiffness of a polymer.

Introduction

Different virus traits determine virus-host interaction with resulting dynamics and heterogeneity of viral populations. Investigating these traits is of fundamental importance for understanding viral replication pathways, viral diversity and fitness, as well as for improving the potency of treatments and viral vaccines. Structural virology has been instrumental in providing a detailed morphological description and classification of viruses and their receptors required for studies of traits influencing virus-host interaction dynamics. However, most of the structural studies have been focused on viral protein shell (termed the capsid) due to its symmetry, rather than on the encapsidated genome where the lack of symmetry prevents high-resolution analysis. Recent findings of pressure-driven viral DNA ejection into cells from phage, archaeal viruses and herpesviruses, have raised strong interest in understanding the role that packaged double-stranded (ds) DNA structure plays in dynamics of genome ejection and viral replication. Knowledge of the structure and directly coupled to its energetics of the encapsidated viral DNA reveals the mechanism of DNA packaging and release during viral replication. Furthermore, it provides insight into the dynamics of viral genome delivery into a host cell and subsequent replication dynamics.

A significant amount of work has been done to describe the structural arrangement and resulting energy of dsDNA packaged in bacteriophage capsids. As mentioned above, while the exact arrangement of dsDNA in the capsid has not been established, the current models combined with the experiments provide a reasonable estimate of the intracapsid DNA equilibrium energy. This energy, dominated by strong DNA-DNA repulsions and bending stress, results in pressure of tens of atmospheres that powers DNA ejection into a cell. We recently found that DNA packaged in phage λ and human herpes simplex virus type 1 undergoes a structural disordering transition from a solid-like to a fluid-like state. In a parallel study, it was shown that DNA packaged in a phage capsid does not have an equilibrium structure and is instead trapped in non-equilibrium conformational states since genome relaxation in the capsid occurs on a slower time scale than processivity of the packaging motor. These findings suggest that DNA packaged in a capsid is in a structural and energetic *metastable* state. Therefore, knowing the intra-capsid DNA equilibrium energy alone is insufficient for understanding of mechanism and dynamics of viral DNA ejection. This article describes the non-equilibrium changes in energy, structure and associated mechanics of encapsidated viral genome facilitating its ejection during viral replication. As will be reviewed here, the metastable state of the packaged viral DNA is an important trait affecting virulence and course of infection (where the term "infection" denotes the introduction of viral nucleic acid into a host cell by a virus). Phage and herpesvirus model systems are used interchangeably below due to major similarities in their DNA ejection and packaging mechanisms, as well as similar DNA packing densities resulting in essentially identical mechanical properties and structure of the encapsidated genome.

Equilibrium Energy and Structure of Intracapsid DNA

As mentioned above, many families of viruses, infecting all three domains of life, have stressed packaged genomes resulting in high internal capsid pressure. This pressure is generated by an ATP-driven packaging motor located at a unique capsid vertex, shown to be the strongest molecular motor known. Structural features of packaging motor components are shared by bacterial and archaeal dsDNA viruses and eukaryotic herpesviruses. The viral genome organization resulting from intra-capsid confinement is closely

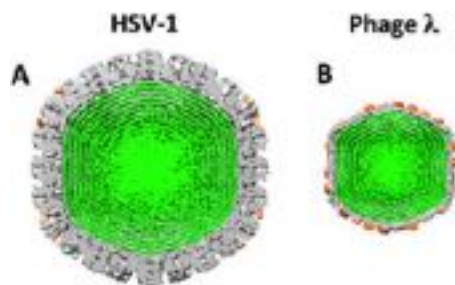


Fig. 1 Cutaway views from cryo-EM reconstructions of (A) HSV-1 C-capsid and (B) phage λ capsid. HSV-1 and λ capsids are shown to scale. Owing to the icosahedral symmetry imposed during the reconstruction, concentrically packed DNA within the capsid becomes shells of density (green). The density maps are downloaded from the Macromolecular Structure Database of the European Bioinformatics Institute with accession codes EMD-1354 (HSV-1) and EMD-5012 (phage λ). Modified from Sae-Ueng, U., Liu, T., Catalano, C.E., *et al.*, 2014. Major capsid reinforcement by a minor protein in herpesviruses and phage lambda. *Nucleic Acids Research*, 42 (14), 9096–107. doi:10.1093/nar/gku634.

associated with its energetic state. The cryo-EM single particle reconstruction of DNA phage λ and human herpes simplex virus type 1 (HSV-1) shown in **Fig. 1** reveals that the entire capsid volume is filled with DNA extending all the way to the center of the capsid. Starting from the capsid walls there are well ordered, multiple concentric DNA layers. The layers are evenly spaced indicating that DNA has adapted an ordered repetitive structure characteristic for a liquid crystalline state. However, toward the center of the capsid the ordered layers disappear suggesting a less ordered DNA structure with lower packing density than in the periphery of the capsid. Similar dsDNA distributions within the capsids were also observed for other viruses. The DNA structure in viral capsids is mainly determined by DNA–DNA interactions, bending stress and packing defects. DNA–DNA interaction and DNA bending are two main energy terms determining the structure of the encapsidated DNA. At high packaging densities, DNA–DNA interactions are mediated mainly by electrostatic and hydration forces. Hydration forces arise from surface structuring of water on the macromolecular surface with a range of interactions of $\sim 15\text{--}20\text{ \AA}$, corresponding to 3–4 water layers on each DNA surface. It was shown that DNA confined in a phage capsid behaves very similar, in terms of interactions and structural ordering, to the multiple DNA arrays confined by a polymer in a bulk solution at equal inter-strand separations. However, the encapsidated DNA, unlike the condensed DNA arrays in the bulk, is strongly bent by the confining capsid walls. This is due to the fact that the dsDNA persistence length is of the same order of magnitude as the diameter of many viral capsids ($\sim 50\text{ nm}$). This leads to an extra bending energy term affecting the structure of the DNA in the capsid. (Persistence length defines the stiffness of a polymer, describing the minimum radius of curvature it can adopt by the available thermal energy. Bending it to a smaller radius requires additional work). To relieve the bending stress, helices are packed closer to the capsid wall, decreasing the bending radius (providing lowest curvature) and also decreasing the spacing, therefore increasing the interaction energy. At the same time the repulsive DNA–DNA interactions will push DNA strands as far from each other as possible, filling the entire capsid volume and maximizing the inter-strand separations. There is a trade-off between bending and interaction energies. Furthermore, when DNA is bending inside the capsid, the initial correlation between two helices that have slightly different radii of curvature is lost, and the mutual orientation between helices must be re-established. This leads to the packing defects that are absent for linear packaging of DNA in solution. Packing defects are required in order to re-establish a favorable phosphate-phosphate “phasing” of helices, reducing the repulsive interactions due to bending.

In order to experimentally dissect the thermodynamic determinants of DNA packaging and ejection energetics, isothermal titration calorimetry (ITC) was used to directly measure the heat released by DNA ejection from a bacteriophage. This approach allows one to address experimentally the relative weight and origin of enthalpic and entropic contributions to the free energy of DNA packing inside a phage capsid. The values of free energy, ΔG , obtained for packaged DNA were one order of magnitude smaller than the values of the enthalpy measured with ITC; the calculated free energy of ejection (ΔG_{ej}) was on the order of 10^{-17} J/virion while the measured enthalpy (ΔH_{ej}) was 10^{-16} J/virion for the wild-type DNA length (WT) phage λ at 25°C . This is due to the fact that the entropy associated with DNA packaging in the capsid is positive, reflecting an increase in the disorder of the system. This may appear surprising, mainly because when DNA is packaged it is becoming more ordered, leading to a reduction in conformational entropy. However, as the DNA helices come closer together in the capsid, the hydration properties of DNA change so that the ordered water molecules directly surrounding the DNA are released, increasing the net disorder of the system, thus increasing its entropy. This observation from ITC measurements suggests that the entropy of packaged DNA is strongly related to hydration.

The observations above describe the equilibrium properties of virally packaged DNA. However, it was found that intra-capsid DNA is not in its equilibrium state but has a metastable behavior. This metastability affects the mobility and related to it dynamics of DNA ejection from the capsid during viral infection. While DNA is always condensed inside the cell, it is not condensed to the same extent as inside a viral capsid. Other than in sperm nuclei, in vivo packaging densities range from $\sim 5\%\text{--}10\%$ by volume. DNA confined in viral capsids, on the other hand, is at the extreme end of the packaging scale where it is confined to $\sim 55\%$ by volume, forming a hexagonally ordered structure. At only a few angstroms of DNA–DNA surface separation (e.g., $7\text{ \AA}$ surface separation in WT DNA length of 48,502 bp packaged in phage λ), hexagonally ordered DNA has been shown to have very restricted mobility. It has been proposed that so-called Coulomb sliding friction between neighboring DNA helices plays a

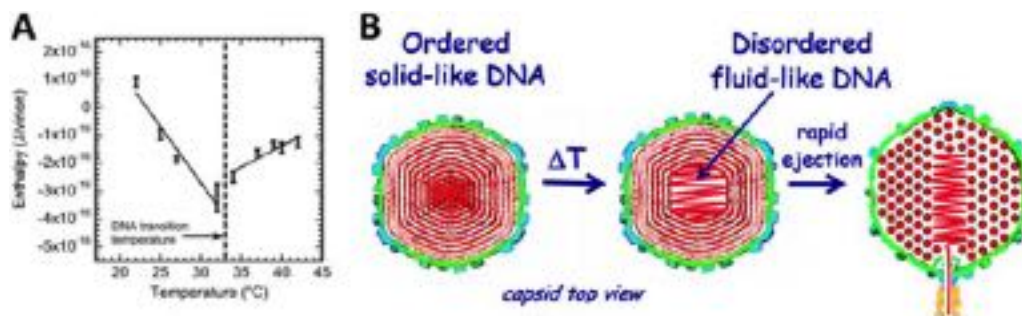


Fig. 2 (A) Enthalpy of DNA ejection per virion (J) versus temperature for wt-DNA phage λ titrated into LamB solution. Dashed lines are drawn to guide the eye. ΔH_{ej} values were obtained as an average of six independent measurements for each sample. Vertical error bars are SEs. (B) DNA disordering transition, which occurs when the temperature is increased, originates close to the center of the capsid, where DNA is more stressed due to stronger bending and larger packing defects. (Left) Cross-sections of the top view of the capsid. (Right) A side view cross-section of the capsid, which illustrates that DNA closer to the center of the capsid is likely to be ejected first since, due to dsDNA bending stress constraints, it is the last DNA portion to be packaged in the capsid during phage assembly. The schematic illustration of DNA inside the capsid shows the ordering of an averaged DNA structure and not the arrangement of individual DNA strands. Modified from Evilevitch, A., 2018. The mobility of packaged phage genome controls ejection dynamics. *eLife* 7. doi:10.7554/eLife.37345.002.

significant role in mobility of DNA at high packing densities in the viral capsids. Indeed, it was found that interhelical sliding friction leads to a kinetically trapped, glassy DNA state inside the capsid. This high friction genome state was shown to significantly affect the rates of DNA packaging and ejection in vitro. This occurs from dragging closely packed negatively charged DNA helices past other helices. The external environmental factors (e.g., ionic conditions and temperature) along with DNA packing density in the capsid strongly affect the metastable state of intra-capsid DNA and facilitate the required mobility of the hexagonally ordered viral genome during the initiation of its ultra-fast ejection reaching 60,000 bp/s. Transition in the metastable states of encapsidated viral genome structure and its energetics associated with the ejection process are reviewed below.

Metastability of Intra-Capsid DNA Facilitating or Inhibiting Viral DNA Ejection

Virion metastability is one of the central concepts in virology. It implies that the virus, in order to successfully replicate, must be sufficiently *stable* to prevent spontaneous release of its genome outside the cell between infection events, and at the same time be *unstable* enough to release its genome during infection. Viral particles are therefore not inert structures and have not attained the minimum free energy conformation, separated by an energetic or kinetic barrier, prior to cell attachment and entry. Thus, viral structure plays an active role in genome delivery to the host cell. Viral metastability is mostly associated with structural transformations in the nucleocapsid and/or surrounding lipid envelope in response to changes in the virion's environment. However, this does not apply to motor-packaged dsDNA viruses (e.g., dsDNA phages and herpesviruses) whose capsid remains intact after the genome is released into a cell through a portal opening in the capsid structure. In these viruses, it is the encapsidated DNA rather than the capsid itself in these viruses that is metastable, where it can undergo a transition from a more ordered to a less ordered state. Atomic force microscopy (AFM) nano-indentation experiments have shown that these two states are characterized by a solid-like (ordered DNA) to a fluid-like (disordered DNA) mechanical transition of intra-capsid genome. This provides a mechanical shift affecting genome mobility. This mechano-structural transition is influenced by temperature, ionic conditions and/or DNA packing density in the capsid which either facilitate or inhibit genome ejection from a capsid into a host cell during viral replication. By mapping both energy (using ITC) and structure (using solution Small Angle X-ray Scattering, SAXS) of the packaged λ -genome a remarkable paradigm of physical adaptation of viruses to their physiologic environment of the host was found. It was shown that mobility of the packaged DNA in λ -capsid is "switched on" to a fluid-like state facilitating its release from the capsid at the most favorable ionic and temperature (37°C) conditions for *E. coli* infection. Correspondingly, the intra-capsid DNA is in a solid-like state inhibiting its released and infection under less favorable conditions for infection, preserving the infectious viral particle. Furthermore, the transition occurs only at the WT λ -DNA length (48,500 kbp) density packaged in the phage λ capsid, as opposed to the shorter DNA lengths of phage λ mutants.

The ITC provides the most direct method to measure the internal energy of the confined viral genome. Using ITC, the enthalpy change (ΔH) associated with DNA ejection from phage λ was measured as heat released when concentrated phage particles are titrated into a LamB receptor solution, which triggers DNA ejection in vitro. (LamB is an outer membrane maltoporin protein in Gram-negative bacteria which also serves as the phage λ receptor). Since the total volume of the system does not change during the DNA ejection in the ITC measurement, and the pressure is constant, the internal energy and the enthalpy (provided by ITC) are approximately equal. The enthalpy change associated with DNA ejection from phage is $\Delta H_{ej}(T) = \Delta H(T)_{DNA \text{ ejected}} - \Delta H(T)_{DNA \text{ inside phage}}$. $\Delta H_{ej}(T)$ can be determined versus temperature. Fig. 2(A) reveals a discontinuity in the approximately linear dependence of ΔH_{ej} on temperature for WT DNA length (48.5 kbp) ejection from phage λ occurring close to the physiologic temperature of infection and optimum ionic conditions for infection (10 mM MgCl₂ Tris-buffer). The discontinuity demonstrates an abrupt transition that is attributed to the DNA

inside the capsid. Prior to the transition, the absolute value of the ejection enthalpy change $|\Delta H_{ej}(T)|$ shows a strong linear increase with increasing temperature. This increase in the internal energy indicates an increase in the stress of the confined genome as temperature is being raised. At the transition temperature, T^* , the internal energy is reduced by almost half, suggesting partial relief of the stressed state. After the transition, $|\Delta H_{ej}(T)|$ shows only weak temperature dependence when the temperature is further increased to 42°C, see Fig. 2(A). This observation demonstrates that the DNA inside λ capsid can exist in two energy states. Thus, the critical genome stress is reached at temperature T^* and is required for the structural transition to occur. There is a strong dependence between the internal genome stress and the structural DNA transition in phage λ . The DNA stress in the capsid is regulated by both packaged DNA length and by the concentration and nature of the DNA counterions freely diffusing through the capsid wall of most viruses. E.g., lower DNA density in the capsid has larger spacings between packaged DNA helices, resulting in weaker repulsive interactions and lower genome stress. As a result, critical genome stress required for transition to occur is not achieved at lower DNA packing density in the capsid than intra-capsid density corresponding to WT λ -DNA length.

Solution small angle X-ray scattering (SAXS) is a powerful technique that provides structural information about the encapsidated genome. In parallel with the increasing inter-strand repulsions, the increase in temperature will decrease the DNA persistence length, leading to less bending stress. If the bending stress decreases and if there is room to expand in the capsid, then DNA-DNA interstrand spacings would increase and interaction energy decrease. However, with SAXS it was shown that the DNA-DNA spacing remains constant, suggesting that there is no room for DNA to expand in the λ -capsid. Instead, as a result of increasing repulsive interactions (due to packing defects) and decreasing bending stress with increasing temperature, the disordering DNA transition occurs, as confirmed by SAXS. This transition is suggested to take place closer to the center of the capsid, where DNA bending stress is stronger and packing defects are larger than for DNA closer to the capsid wall, making the DNA in the center more destabilized and therefore more sensitive to the increasing temperature. At the transition temperature, the DNA bending stress becomes sufficiently small, allowing a fraction of the ordered DNA layers closest to the capsid's center to undergo a disordering transition. The disordered DNA will have a lower packing density than the ordered DNA, which maximizes DNA-DNA spacings and simultaneously reduces the repulsive interactions and DNA-DNA sliding friction. This yields an overall lower energy state of the encapsidated genome and increases DNA mobility in the center of the capsid which is required for rapid DNA release during infection. DNA closer to the center of the capsid is likely to be ejected first since it is the last DNA portion to be packaged in the capsid during the phage assembly, due to the dsDNA bending stress constraints. The schematic representation of this transition behavior versus temperature is shown in Fig. 2(B).

While variation in the monovalent ion concentration has a small influence on intra-capsid DNA stress, polyvalent cations present in the bacterial cytoplasm, such as polyamines and Mg^{2+} , have been shown to have a strong effect on the repulsive interactions between packaged DNA helices. Since the free polyamine concentration in cells is very low as most polyamines are bound to cellular DNA and RNA, there is a particular interest in the effect of the Mg-ion concentration on the DNA transition temperature in phage λ at concentrations similar to those of free Mg^{2+} in vivo. Mg-ions are essential for both cellular metabolism, e.g., enzyme activity, protein synthesis, preservation of ribosome and nucleic acid structures, as well as the phage λ infectious cycle. It has been shown that a concentration of 10–20 mM of Mg^{2+} ions in the extracellular solution is critically important for optimum adsorption and infection of *E. coli* by phage λ . Mg-concentrations below or above 10–20 mM had a strong negative effect on the number of bacterial cells that could be infected. Interestingly, a similar Mg-concentration (~ 10 mM Mg^{2+}) has also been shown to provide optimum conditions for DNA packaging in phage λ and also corresponds to the free Mg-concentration in *E. coli* cytoplasm. Since the Mg-concentration affects DNA stress in the capsid (by freely diffusing through the capsid walls and screening the negative charges in the packaged DNA), it also strongly influences the transition temperature of the encapsidated genome. While T^* varies significantly with Mg^{2+} concentration, the most favorable Mg-concentration for phage λ infection of *E. coli* (~ 10 –20 mM) triggers DNA transition in the capsid precisely at the physiologic temperature of infection ($\sim 37^\circ\text{C}$). This striking correlation shows that the intra-capsid DNA transition mechanism in λ is evolutionarily adapted to both the ionic environment and temperature of its host (phage λ replicates in *E. coli* in the human gut at 37°C), suggesting its significance for viral replication. An analogous metastable DNA state inside the viral capsid is also present in HSV-1, where solid-to-fluid like DNA transition occurs close to 37°C . This suggests that the DNA transition mechanism regulating infectivity may be universal for many pressurized DNA viruses.

These observations suggest that dynamics of DNA ejection can be affected by a transition in mobility of the encapsidated genome, which can also influence viral replication dynamics and course of infection. This is discussed in the next section.

The Mobility of Packaged Genome Controls Ejection Dynamics

Gaining insight into the dynamics of viral gene delivery in a host cell during infection is critical for understanding the virus replication dynamics and viral fitness. We recently found a direct link between the structure (and its associated mobility) of the encapsidated dsDNA in phage λ and dynamics of initiation of its release from the capsid. Time-resolved ITC measurements provide ultra-sensitive detection of ejection dynamics from a phage population triggered by instant mixing with LamB λ -receptor in vitro, revealing the effect of mechano-structural DNA transition in a viral capsid on DNA ejection dynamics from a phage λ population. ITC is mainly used for thermodynamic analysis. However, it was recently demonstrated that it can be successfully used to measure reaction kinetics reflected by heat flow (so-called kinITC) due to the very short response time of modern ITC instruments.

It was shown that the solid-to-fluid-like intra-capsid DNA transition, is a determining factor for either rapid synchronized or slow desynchronized ejection dynamics from a phage λ population. At optimum temperature for infection (i.e., $\sim 37^\circ\text{C}$) and Mg-concentration ≤ 10 mM, the encapsidated DNA is in a fluid-like state and ejection events from the majority of the phage population are synchronized with DNA ejections occurring within ~ 10 s directly after phage adsorption to *E. coli* receptors (this time corresponds to DNA translocation time from a capsid). However, deviations from these conditions for infection (e.g., low temperature, high Mg-concentration) result in intra-capsid DNA being in a solid-like state where DNA is arrested in different conformations due to high interstrand sliding friction. This leads to a striking heterogeneity in phage ejection dynamics, displaying coexistence of populations with synchronized ejection events (where DNA is in a fluid-like state) and desynchronized stochastically occurring ejection events (where DNA is in a solid-like state). The latter presents one to two orders-of-magnitude slower ejection dynamics, ranging from minutes to tens of minutes, depending on salt conditions and temperature, compared to DNA translocation time from a capsid. This in vitro ejection dynamics behavior is in good agreement with earlier observations of temperature's effect on DNA injection dynamics from phage λ after preadsorption to *E. coli* culture. These findings also dismiss an earlier claim that stochasticity of DNA ejections from phage is solely controlled by capsid portal opening dynamics. These assumptions were based on earlier light scattering measurements of phage ejection dynamics that could not distinguish between synchronized and desynchronized ejection events.

The lytic-lysogenic decision is one of the most crucial factors determining virus population dynamics. Lytic course of infection is where a virus replicates immediately, leading to cell lysis and progeny phage release, while lysogenic infection is a latent state where phage DNA is integrated into a cell's chromosome without killing the cell. There are three early viral genes (CI, Cro and CII) which have been previously found to be involved in the decision-making process. CII appears to provide positive and negative feedback in the gene regulatory network. High levels of CII promote production of CI which leads to lysogeny, while low levels of CII promote production of Cro leading to bacterial cell lysis. It was recently proposed that the lysogeny-lytic cell decision initially occurs at the phage level through the timing of DNA delivery from multiple phages into a cell during infection. The timing of DNA injection from phage is what determines the onset of viral genome replication and gene expression because both phage DNA injection and DNA replication processes occur on comparable timescales. In turn, replication and expression dynamics of injected phage genomes determines how phages interact with one another when multiple phage particles are infecting a single cell. It was shown that lysogeny is more likely to occur when infecting phages interact cooperatively during cell infection. Synchronized injections result in immediate presence of DNA from multiple phages in a cell that can be integrated into cell's chromosome without competition for the progeny. Lysogeny results in the absence of selection, which allows various mutants to integrate and persist in the lysogenic state, leading to diversification of phage genes and adaptation to poor or new growth conditions. On the contrary, when DNA injections are desynchronized, few phage particles are delivering their genes faster than the rest of the phage population, resulting in competitive interactions. If the first phage to deliver its genome is lytic, it will override lysogeny, rapidly replicate itself using the cell's limited resources and therefore take a larger share of phage progeny. The asynchronous infection delays decrease the chance of lysogeny by also lowering CII viral transcription factor levels in the cell. It was shown that the cutoff time, where delayed phage infections of one cell no longer affect the lytic-lysogenic cell decision varies from few minutes to tens of minutes, depending on the multiplicity of infection, MOI (higher MOI results in shorter cutoff time), comparable to timescales for overall ejection delays observed in ITC measurements. This suggests that ejection delays resulting from solid-like DNA state in the capsid will increase cell lysis probability.

These in vitro findings suggest a new paradigm affecting virus population dynamics – mechano-signaling mediated by packaged DNA structure in phage λ . The environmental factors regulating the mechanical transition between solid- and fluid-like intra-capsid DNA states strongly influence the timing of genome ejections. In vivo, desynchronized ejections from a phage population lead to competition between phage genomes, affecting the rate of phage gene replication and transcription as well as the lysogeny-lytic cell decision. Given similarities between phage λ and herpesviruses in their mechanism of pressure-driven DNA ejection (where a temperature-induced solid-to-fluid like DNA transition was observed in both viral systems), the new insights into the lysogeny-lytic decision process for phage provide intriguing prospects for understanding of latency mechanism in herpesviruses. Indeed, analogies between phage and herpesviruses in other regulatory factors that affect viral latency have been previously observed.

Many viruses have been shown to have pressurized genomes inside their capsid. However, in order to place this important observation in the biological context of viral infection, it is critical to demonstrate the role that capsid pressure plays in viral genome delivery into a cell. Next Section describes the definitive demonstration of a pressure-dependent infection mechanism in viruses by showing that pressure-driven DNA ejection from HSV-1 capsids into a host cell nucleus is inhibited when intra-capsid DNA pressure is “turned off”.

Pressure-Driven Release of Viral Genome Into a Host Cell is a Mechanism Leading to Infection

All eukaryotic DNA viruses, with the exception of poxviruses, deliver and replicate their genomes within the nucleus. In addition, retro-viruses transport their genomes across nuclear membranes in order to replicate after they have reversely transcribed their single-stranded (ss) RNA to dsDNA. The mechanism of this most significant step of infection, viral DNA entry into the host nucleus, remains poorly understood for the majority of viruses. Despite previous measurements of intra-capsid DNA pressure in phage and herpesvirus, the role that capsid pressure plays for intra-cellular viral genome delivery has not been previously

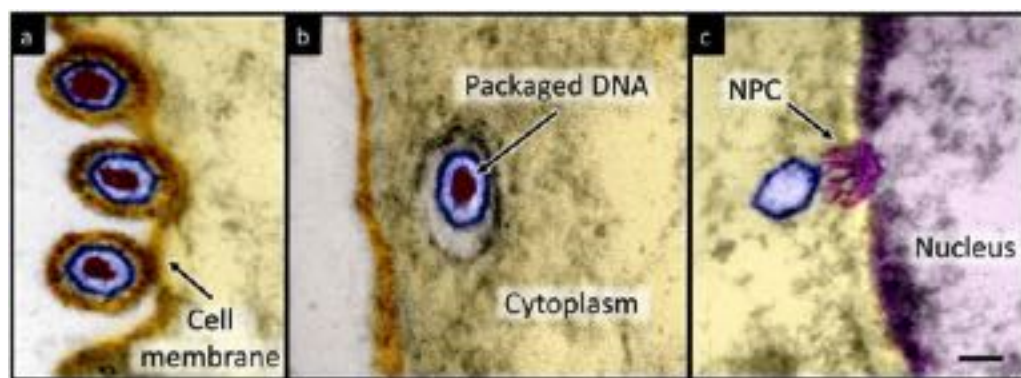


Fig. 3 After binding at the outer membrane (Fig. 3(a)), viruses enter the cell cytoplasm and are transported toward the nucleus (Fig. 3(b)). The viral capsid ejects its genome upon docking to a nuclear pore complex (Fig. 3(c)). From Bauer, D.W., Huffman, J.B., Homa, F.L., Evilevitch, A., 2013. Herpes virus genome, the pressure is on. *Journal of the American Chemical Society* 135 (30), 11216–11221. doi:10.1021/ja404008r.

demonstrated. The reconstituted capsid-nuclei experiments below provide a demonstration of a nuclear entry mechanism of DNA from HSV-1, driven by high mechanical pressure of the encapsidated viral genome.

Herpesviruses consist of a dsDNA genome packaged within an icosahedral capsid that is surrounded by an unstructured protein layer, the tegument, and a lipid envelope. Fig. 3 illustrates the HSV-1 infection process as observed by ultrathin-sectioning transmission electron microscopy (TEM). After binding at the outer membrane (Fig. 3(a)), viruses enter the cell cytoplasm and are transported toward the nucleus (Fig. 3(b)). The viral capsid ejects its genome upon docking to a nuclear pore complex (NPC), which forms a passageway for molecular traffic into the nucleus (Fig. 3(c)).

To investigate the effect of capsid pressure on the single step of viral DNA ejection into a nucleus, cell nuclei were isolated and reconstituted with cytosol supplemented with an ATP-regeneration system. This reconstituted-nuclei assay avoids interference from other processes occurring within the cell during viral replication. The reconstituted nuclei system also accurately reproduces capsids-nuclei binding and nuclear transport of the herpes genome into living cells and it was previously shown that purified HSV-1 capsids bind to NPCs on isolated nuclei and eject their DNA into nuclei. To provide the evidence that intra-capsid DNA pressure is responsible for DNA release from a herpesvirus capsid into a cell nucleus, it was demonstrated that when the capsid pressure is “turned off” with addition of an external osmolyte, herpes capsids bound to NPCs do not eject DNA into a nucleus, while the ejection is completed successfully without osmolyte addition. The external osmotic pressure in solution surrounding the capsids effectively eliminates genome pressure in the capsid when it matches the pressure of the packaged DNA. As shown in Fig. 4, the viral genome ejection from HSV-1 through the NPCs into a cell nucleus can be completely suppressed in the presence of the biologically inert osmolyte polyethylene glycol (PEG) 8 kDa at 30% w/w corresponding to 20 atm of osmotic pressure matching that of DNA pressure in the capsid. While the term “infection” usually refers to both viral genome transport into the cell and subsequent replication of the virus, the primary infection by several types of herpesviruses (including HSV-1) is latent (i.e., the herpes genome is translocated into the host nucleus, without subsequent genome replication). Thus, the osmotic suppression assay, combined with the reconstituted nucleus system, provides a demonstration of a pressure-dependent mechanism of herpesvirus infection focused on the viral genome translocation step.

Concluding Remarks

Correlating structure and energetics of the encapsidated genome with the efficiency of viral replication provides a unique approach to explore the connection between *physical* and *genetic* aspects of viral evolution, and creates novel opportunities for control of viral infections. Various aspects of *genetic evolution* have been investigated in the field of virology. The *physical aspects* of viral evolution, however, are less understood. Recently, an intimate relationship was found between the physical and genetic evolution of dsDNA viruses. Specifically, it was shown that the physical limit of DNA length imposed by the capsid volume has led to gene overlap evolving as a mechanism for producing more proteins from the same genome length. This demonstrates how a *genetic mechanism* has evolved from a *physical* constraint by the capsid on the packaged genome density. Similarly with this finding, the data reviewed in this article suggest that the genome packing density in a viral capsid is unique and highly conserved. This precise packing density not only determines the number of genes required for replication, but also creates an internal stress in the capsid required for the solid-to-fluid like DNA transition to occur at the temperature of infection. Such intracapsid DNA transition facilitates initiation of DNA delivery from viral capsid into a cell. E.g., there is evidence that the packaged genome density in phage λ corresponding to WT λ -DNA length (as opposed to the shorter λ -DNA length mutants), presents the most energetically and structurally optimized balance between genome mobility and its internal stress, both of which are required for efficient DNA ejection from the capsid. DNA transition occurs once the critical DNA stress level in the capsid is reached, which are in turn controlled by the temperature and the ionic conditions of the surrounding solution. The DNA transition in the capsid occurs precisely at the physiologic

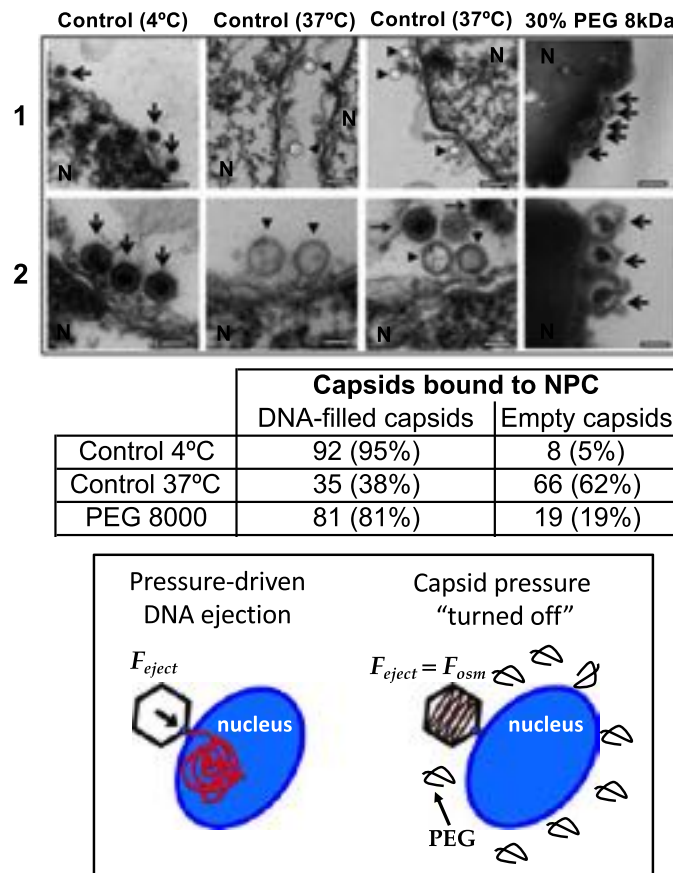


Fig. 4 Ultrathin-sectioning EM visualization of the osmotic suppression of DNA ejection from HSV-1 capsids into reconstituted cell nuclei [Isolated cell nuclei with added cytosol with ATP regeneration system. Cytosol contains importin- β required for efficient HSV-1 capsid binding to NPCs.] Negative control at 4°C without added PEG and without ATP-regenerating system, shows that no ejection from nuclei bound C-capsids occurs. Positive control at 37°C shows complete DNA ejection from C-capsids bound to isolated cell nuclei supplemented with cytosol ATP-regenerating system. EM shows that the addition of 30% PEG 8 kDa (corresponding to 20 atm of osmotic pressure) to reconstituted capsid-nuclei system, inhibits DNA ejection from HSV-1C-capsids into host nuclei through the NPC. Thin arrows show DNA-filled capsids, and bold arrows show empty capsids that ejected DNA. 1. Bar 500 nm. 2. Bar 90 nm. Modified from Brandariz-Nunez, A., Liu, T., Du, T., Evilevitch, A., 2019. Pressure-driven release of viral genome into a host nucleus is a mechanism leading to herpes infection. *eLife* 8. doi:[10.7554/eLife.47212.001](https://doi.org/10.7554/eLife.47212.001).

temperature of infection (37°C) and at the most favorable external Mg^{2+} -concentration for phage λ adsorption to *E. coli* and subsequent replication. This suggests an evolutionary adaptation of the intracapsid DNA transition mechanism to the temperature and ionic environment of its host.

The remarkable observation of mechano-regulation of phage genome ejection dynamics can be compared to gene expression regulation in meters-long, tightly packaged chromosomal DNA in eukaryotic cells, where gene expression is controlled not only by genetic regulation but also by mechanical properties of DNA in nucleosomes. Environmental stress factors, such as an abrupt change in temperature or in ionic conditions, were demonstrated to play a major role in the mechano-regulated gene activation process, turning certain genes on and off, in a process termed epigenetics. Similarly, temperature and ionic conditions control the mechanical properties of virally encapsidated DNA and act as an on-off switch between synchronized and desynchronized ejection dynamics. This can influence timing of viral replication and the lytic-lysogenic cell decision process.

Analogously to DNA pressurized phage capsids, high intra-capsid DNA packing density resulting in tens of atmospheres of pressure is a distinctive trait of all nine human herpesviruses. This strongly suggests that pressure-driven entry of viral DNA into the host nucleus during infection is universal to all herpesviruses. Other types of viruses also involve replication steps dependent on the pressurized state of the intra-capsid genome. For instance, during genome packaging, reoviruses replicate ssRNA to dsRNA inside the capsid, which results in genome packaging densities similar to that of herpesviruses. Such intra-capsid replication could be impacted, at least in part, by generation of internal pressure resulting from the increasing genome packaging density as newly synthesized dsRNA continues to fill the internal capsid volume. Another example is HIV, where similar to herpesviruses, HIV capsids dock to NPCs at the nucleus and release transcribed dsDNA through the NPC channel. It was recently shown that the reverse transcription process from ssRNA to dsDNA inside the HIV capsid is associated with increasing internal DNA pressure. The experiments summarized in this article provide the necessary tools for exploration of pressure-regulated replication in these viruses.

Further Reading

- Bauer, D.W., Huffman, J.B., Homa, F.L., Evilevitch, A., 2013. Herpes virus genome, the pressure is on. *Journal of the American Chemical Society* 135 (30), 11216–11221. doi:10.1021/ja404008r.
- Berndsen, Z.T., Keller, N., Grimes, S., Jardine, P.J., Smith, D.E., 2014. Nonequilibrium dynamics and ultraslow relaxation of confined DNA during viral packaging. *Proceedings of the National Academy of Sciences of the United States of America* 111 (23), 8345–8350. doi:10.1073/pnas.1405109111.
- Booy, F.P., Newcomb, W.W., Trus, B.L., *et al.*, 1991. Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus. *Cell* 64 (5), 1007–1015. doi:10.1016/0092-8674(91)90324-R.
- Brandariz-Nunez, A., Liu, T., Du, T., Evilevitch, A., 2019. Pressure-driven release of viral genome into a host nucleus is a mechanism leading to herpes infection. *eLife* 8. doi:10.7554/eLife.47212.001.
- Catalano, C.E., 2005. *Viral Genome Packaging Machines: Genetics, Structure, and Mechanism*. New York: Bioscience/Eurekah.com.
- Earnshaw, W.C., Harrison, S.C., 1977. DNA arrangement in isometric phage heads. *Nature* 268, 598–602.
- Evilevitch, A., 2018. The mobility of packaged phage genome controls ejection dynamics. *eLife* 7. doi:10.7554/eLife.37345.002.
- Evilevitch, A., Lavelle, L., Knobler, C.M., Raspaud, E., Gelbart, W.M., 2003. Osmotic pressure inhibition of DNA ejection from phage. *Proceedings of the National Academy of Sciences of the United States of America* 100 (16), 9292–9295. doi:10.1073/pnas.1233721100.
- Goldhill, D.H., Turner, P.E., 2014. The evolution of life history trade-offs in viruses. *Current Opinion in Virology* 8, 79–84. doi:10.1016/j.coviro.2014.07.005.
- Hanhijarvi, K.J., Ziedaite, G., Pietila, M.K., Haeggstrom, E., Bamford, D.H., 2013. DNA ejection from an archaeal virus – A single-molecule approach. *Biophysical Journal* 104 (10), 2264–2272. doi:10.1016/j.bpj.2013.03.061.
- Kindt, J., Tzili, S., Ben-Shaul, A., Gelbart, W.M., 2001. DNA packaging and ejection forces in bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America* 98 (24), 13671–13674.
- Lander, G.C., Evilevitch, A., Jeembaeva, M., *et al.*, 2008. Bacteriophage lambda stabilization by auxiliary protein gpD: Timing, location, and mechanism of attachment determined by cryo-EM. *Structure* 16 (9), 1399–1406. doi:10.1016/j.str.2008.05.016.
- Leikin, S., Parsegian, V.A., Rau, D.C., Rand, R.P., 1993. Hydration forces. *Annual Review of Physical Chemistry* 44, 369–395. doi:10.1146/annurev.pc.44.100193.002101.
- Liu, T., Sae-Ueng, U., Li, D., *et al.*, 2014. Solid-to-fluid-like DNA transition in viruses facilitates infection. *Proceedings of the National Academy of Sciences of the United States of America* 111 (41), 14675–14680. doi:10.1073/pnas.1321637111.
- Mackay, D.J., Bode, V.C., 1976. Events in lambda injection between phage adsorption and DNA entry. *Virology* 72 (1), 154–166.
- Petrov, A.S., Harvey, S.C., 2008. Packaging double-helical DNA into viral capsids: Structures, forces, and energetics. *Biophysical Journal* 95 (2), 497–502.
- Purohit, P.K., Kondev, J., Phillips, R., 2003. Mechanics of DNA packaging in viruses. *Proceedings of the National Academy of Sciences of the United States of America* 100 (6), 3173–3178.
- Riener, S.C., Bloomfield, V.A., 1978. Packaging of DNA in bacteriophage heads: Some considerations on energetics. *Biopolymers* 17 (3), 785–794.
- Smith, D.E., Tans, S.J., Smith, S.B., *et al.*, 2001. The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* 413 (6857), 748–752.
- Zeng, L., Skinner, S.O., Zong, C., *et al.*, 2010. Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell* 141 (4), 682–691. doi:10.1016/j.cell.2010.03.034.

Bacteriophage Receptor Proteins of Gram-Negative Bacteria

Sarah M Doore, Kristin N Parent, Sundharraman Subramanian, and Jason R Schrad, Michigan State University, East Lansing, MI, United States

Natalia B Hubbs, Hanover College, Hanover, IN, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Assay An experiment used to assess or measure a property of interest.

Genome The entirety of genetic material present in a cell or organism.

LPS Lipopolysaccharide; a molecule comprised of fatty acid and repeating sugar units that decorates the outer membrane of Gram-negative bacteria and varies widely between species and genera.

Omp Outer membrane protein; a protein found in the outermost membrane of Gram-negative bacteria.

Tail Phage structure comprised of several proteins through which many bacteriophages identify and bind to their host; the tail is bound to the phage capsid, which carries genomic information.

Transposon A sequence of DNA that can insert itself into or change its location in the genome.

History and Overview

Shortly after the discovery of bacteriophages, scientists became interested in multiple aspects of their infection cycle. However, the initial stages of infection – host recognition and attachment – received relatively little attention until three decades later in 1951, when Garen and Puck described two distinct steps during bacteriophage T2 attachment to its host *Escherichia coli*. This two-step process has since been described for many well-characterized bacteriophages. The first step is reversible, meaning the phage can unbind and has not committed to infection. The second step is irreversible and the particle commits to infecting the host. Subsequent experiments determined that the second step involves proteins in the bacterial membrane, but the identity of these proteins was not determined until the 1970s.

Since those initial studies, our knowledge of phage receptors has grown exponentially. We now know that for many Gram-negative bacteria and their associated phages, the primary receptor is typically the lipopolysaccharide (LPS), capsular layer, flagellum, or pilus. At this stage, if the phage has attached to an incompatible host cell, it is still able to detach and search again for a suitable host. If the host is compatible, the phage reaches its secondary receptor by either degrading the LPS or the capsular layer, or by moving closer to the cell *via* rotary action of the host flagellum or retraction of the pilus. If the secondary receptor is compatible, the particle binds irreversibly to this host protein and its genetic material is ejected into the host (**Fig. 1**). Developing better techniques, as discussed later in this article, has been critical to our understanding of membrane proteins and how bacteriophages interact with them.

While bacteriophages as a whole employ vastly different strategies for primary receptor recognition, secondary receptors tend to fall into a few main classes of membrane proteins. This article will focus on secondary receptors by covering a few common techniques and methods to study membrane proteins, the characteristics of outer membrane proteins (Omps), and some of the known interactions between phages and host Omps. Within this context, the ecology and evolution of these interactions will be discussed, along with the importance of studying these interactions for use in phage applications.

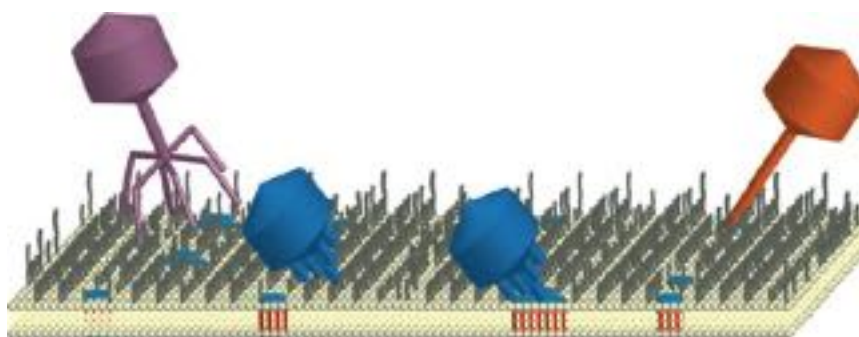


Fig. 1 Representation of tailed bacteriophages interacting with the outer membrane of a Gram-negative bacterium. Membrane proteins are colored using the scheme in **Fig. 2**. Representative myovirus (purple), siphovirus (orange), and podoviruses (blue) are shown interacting with LPS and/or Omps *via* their tail proteins.

Techniques and Methods for Studying Phage Receptors

Since the initial studies from the 1950s, scientists have developed dozens of methods for identifying and studying the membrane proteins that bacteriophages use as secondary receptors. Some of these have been used for decades, including forward genetic screens and gel electrophoresis. Others are still being developed, such as single-gene deletion libraries and methods using advanced microscopy. [Table 1](#) presents a summary of these methods along with strengths and weaknesses. A more in-depth discussion of each is presented at the end of this article.

Once the receptor is identified from genetic or protein screens, and once its structure and/or function has been determined by atomic or low-resolution techniques, scientists can begin to visualize how the phage interacts with its receptor in the context of the native host. This question can be investigated by *in vivo* approaches such as time-lapse fluorescence microscopy and cryo-electron tomography (cryo-ET). These techniques visualize an entire population of phages interacting with their hosts or intact membranes.

Time-lapse fluorescence microscopy uses fluorescent markers, which can selectively dye either the host or the phage. A series of images taken during the course of the infection are then stitched together to form a movie of the infection process. This method can show the timing of host binding and genome ejection, as well as the distribution of entry sites on hosts.

Another direct visualization technique is cryo-ET. Similar to cryo-electron microscopy (cryo-EM), this involves flash-freezing populations of cells being infected with phage in vitreous ice and imaged in the transmission electron microscope. Host mutants called “mini cells” or outer membrane vesicles are useful to replace bona fide cells, as these limit the thickness of the specimen. The sample is then tilted to capture images from many different angles, and image reconstruction techniques are employed to create a more complete 3D picture. With cryo-ET, the image reconstruction results in being able to visualize the whole volume of the sample. From cryo-ET experiments, researchers can see details about the global structure of the phage tails and channels into the host membrane. Since flash freezing stops biological processes within milliseconds, this technique can capture phages at various stages of infection, letting scientists see the sequence of events more clearly.

Properties of Receptor Proteins

Through both classical and modern techniques, scientists have identified many common properties of phage receptor proteins of Gram-negative bacteria. Numerous well-studied phages infect intestinal – or enteric – bacteria. These include species of *Escherichia*, *Klebsiella*, *Salmonella*, *Shigella*, and *Yersinia*. The Omps used by many phages infecting these species are listed in [Table 2](#).

The vast majority of proteins found in the outer membrane are β -barrels, which are proteins rich in β -sheets that twist to form a barrel shape. These barrels are then embedded in the membrane, oriented as channels through which molecules can move into or out of a cell. Many of these proteins have α -helical or disordered loops connecting the β -sheets. Additionally, the majority have larger extracellular domains than periplasmic, or between-membrane, domains. Some of these proteins are involved in the transport of small molecules or nutrients, while others are able to sense membrane stress or contribute to membrane biogenesis. Many have unknown functions. Regardless of their role in bacteria, these proteins can be co-opted by bacteriophages to identify and gain access to their host. Representative structures from this list are shown in [Fig. 2](#).

Bacteriophages use their aptly-named receptor binding proteins to bind their specific host receptor. These receptor binding proteins, which are located on the surface of the phage particle, can recognize a suitable host and trigger a conformational change, resulting in the phage genome moving across the membrane. Many bacteriophages have specialized protein machinery called tails to accomplish this, with receptor binding proteins located on tail fibers or tail spikes. Non-tailed phages can identify their host *via* spikes directly attached to the capsid. Tailed phages, which are classified according to morphology, use different mechanisms of host binding and genome ejection depending on the tail. The three most common families of bacteriophage are the *Myoviridae* (long, contractile tails), *Siphoviridae* (long, non-contractile tails), and *Podoviridae* (short tails).

Several phage-host pairs have been studied in detail to determine the interactions between the host receptor and the phage's receptor-binding proteins. Some of these include the siphovirus lambda (λ) and its receptor LamB; siphovirus T5 and its receptor FhuA; and the podovirus Sf6 and its receptors OmpA and OmpC. These model systems have given us a wealth of information about genetics, protein-protein interactions, and predator-prey dynamics.

Bacteriophage λ and its receptor LamB is a classic system for studying phage-host interactions. The LamB outer membrane protein forms a homotrimer consisting of three β -barrels and functions as a maltose and maltodextrin transporter. The initial model of λ and LamB binding was based on the isolation and characterization of λ -resistant bacteria by Thirion and Hofnung in 1972. These researchers observed that the bacteria were also unable to use maltose as a carbon source, but the exact mechanism for this was unknown. The following year, Randall-Hazelbauer and Schwartz fractionated bacteria, separating the membranes and other components of the bacterial cell *via* ultracentrifugation. The end result is “fractions” of the sample that each contain a specific biomolecule: only membranes, only DNA, only ribosomes, *etc.* By doing this, they learned where this “ λ -resistance factor” was located in the cell: the outer membrane.

Once the specific protein was known, researchers could focus on a more in-depth characterization. Initially, they used directed genetic screens to determine which mutations affected either antibody binding or protease digestion of LamB. This information told researchers which regions of the protein were likely exposed to the extracellular environment and available for antibody binding or protease digestion. Mutations in LamB that made the bacteria resistant to λ could then be compared to these putative

Table 1 Techniques for studying bacteriophage receptor proteins

	<i>Pros</i>	<i>Cons</i>
Techniques for receptor identification:		
Genetics		
Resistance mutants	● Fast (<1 day) ● Simple and inexpensive ● Can process large populations (> 10⁸ cells) in a single experiment	● Mutants can arise from many steps in life cycle, not just receptors ● Resistant host may have multiple mutations complicating interpretation ● Binary results only: resistant or susceptible ● Redundant results if multiple genes contribute to a pathway
Single gene deletion library	● Systematic: checks all non-essential genes ● “Clean”: isogenic populations and a single, sequenced mutation per host	● Genetic tools are not available for all hosts ● Expensive and time consuming to generate
Transposon libraries	● Libraries can be generated from a single starting population ● Non-binary: can also identify genes that delay and accelerate infection	● Multiple genes contributing to the same pathway can make interpretation difficult
Protein-based		
Binding assays	● Detects direct binding	● Non-specific host components can co-purify
2D Gels	● Narrows down a few possible candidates instead of screening an entire library	● Receptor/phage interactions may be too transient to effectively capture
Chemical crosslinking		● The receptor might be present but not detected
Mass spectrometry		
Techniques for receptor characterization:		
Atomic/near atomic resolution		
Crystallography	● Provides detailed information about the receptor, including the position of all important atoms (amino acids or chemical bonds)	● Requires purification: not possible or practical for all cell components
Nuclear Magnetic Resonance (NMR)	● Shows regions that are flexible	● Not all substances readily crystallize.
Cryo-electron microscopy (cryo-EM)		● Size may be a limiting factor: either too large for NMR or too small for cryo-EM.
Lower resolution		
Oligomeric state: Gel shift, analytical ultracentrifugation, native mass spectrometry	● Provides basic information about the overall degree of folding at the level of tertiary or secondary structure	● Resolution is too low for specific details
Optical: Circular dichroism, fluorescence	● Can be used to determine the number of copies of a protein within a complex ● Can be used to assay how dynamic or stable a receptor is	● Membrane proteins often are purified in detergents, which may interfere with some of these methods ● Requires protein purification and extensive optimization
Techniques for characterization of phage-host interactions:		
Interaction location		
Site directed mutagenesis experimental evolution	● Identifies specific amino acids or regions of a protein that contribute to binding	● Some structural information is needed to guide the experiments ● Individual changes may result in subtle phenotypes ● Requires genetic tools such as engineering or sequencing approaches
Chemical crosslinking	● Identifies specific amino acids or regions of a protein that contribute to binding	● Directed approaches may be limited if the structure of the receptor is not known ● Non-specific crosslinking often occurs
Interaction dynamics		
Time-lapse fluorescence microscopy	● Native conditions	● Expensive and time consuming
Cryo-electron tomography (cryo-ET)	● Can directly visualize many distinct infection events	● Whole cells are often too thick for cryo-ET, so mini cells or outer membrane vesicles are needed (less-native)

(Continued)

Table 1 Continued

	<i>Pros</i>	<i>Cons</i>
Computation	● Looks at very complex landscapes including lipids, proteins, LPS, <i>etc.</i> ● Accesses conditions not easily achieved in laboratory settings ● Easily repeatable	● Currently limited to small time scales ● Some structural information is needed to guide the experiments
<i>In vitro</i> kinetics	● Can be used to determine strength of binding interactions ● Rates of binding can be determined ● If the receptor has enzymatic activity this can be studied as well	● Requires protein purification; not all receptors remain active when purified

extracellular sites to see where the phage might be binding. Most of these resistant bacteria contained mutations that mapped to the disordered protein strands (loops) 4, 5, 6, and 8 of LamB. Although a majority of these mutations affected simple binding of λ to host, a single amino acid immediately adjacent to loop 8 was also identified as critical for infection. This mutation, which resulted in glycine 401 being changed to aspartate, allowed the phage to bind but prevented genome ejection.

To examine how λ receptor binding proteins worked, researchers used a combination of methods similar to those described for LamB – examining antibody binding sites and genetic studies – as well as some early techniques in electron microscopy. They determined that λ binds to its host *via* protein J, which forms the needle-like tip of the tail. Images from the electron microscope showed gold-labeled J proteins interacting directly with the cell membrane of wild-type bacteria, but not with a mutant cell lacking *lamB*.

Although λ was one of the earliest phages to be studied in detail, it is still highly relevant in modern techniques. This includes dynamics experiments regarding λ “searching” for receptors using single-cell fluorescence microscopy and utilizing the extensive genetic screens made possible by the KEIO collection – a library of single gene deletion mutants including all non-essential genes in *E. coli* – to determine which genes are important for infection. Besides receptor interactions, this bacteriophage is also used extensively to investigate ecological and evolutionary questions.

After λ , one of the next rising stars in phage-host interactions was bacteriophage T5 and its receptor FhuA. From the late 1970s through the 1980s, several groups determined that the tail fiber, denoted pb1, binds to the O-antigen of LPS during the first step of attachment. They also characterized two other proteins involved in host attachment and entry: pb5, which we now know binds specifically to FhuA; and pb2, which is involved in translocating the DNA into the host cell. The pb5 and FhuA binding interaction was first proposed by Heller and Bryniok in 1984. These researchers isolated a T5 mutant that infected wild-type cells more slowly than wild-type T5 and were completely unable to infect cells lacking the O-antigen: a specific arrangement of sugars on LPS that are often recognized by phages. Instead of a mutation in pb1, which they knew bound to the O-antigen, they mapped the mutation to a separate protein, pb5. They hypothesized that this mutant depended on pb1 sticking to O-antigen for a longer amount of time. This would increase the probability that the phage could bump into and bind to its protein receptor. Since the mutant pb5 could not bind as efficiently to the protein receptor, the phage relied on O-antigen to stick long enough to find its receptor. The researchers then treated cells with ferrichrome, which binds selectively to and inhibits FhuA transport of iron, and saw that this treatment caused the mutant phage to bind even less. From these experiments, they determined that FhuA, which the bacteria uses for iron transport, was also the protein receptor for phage T5.

The role of pb5 as a receptor-binding protein was later confirmed by additional biochemical and structural experiments. Initially, this was investigated much like the phage λ and LamB interaction, using inhibitory peptides to determine which regions of the receptor were recognized by the phage. In 1995, Killmann *et al.* investigated the binding regions of FhuA-dependent phages T1, T5, and ϕ 80. They determined that all three phages use the same large extracellular loop of FhuA, with some minor variation in the exact regions of binding. In 2001, Böhm *et al.* used cryo-EM to look at T5 bound to FhuA-containing vesicles. The following year, Plancon *et al.* purified pb5 and FhuA to look at their binding interaction *in vitro*. They performed a battery of assays to measure DNA ejection rates, the stoichiometry of binding, and the stability of the complex in different denaturants and at a range of temperatures. The combination of these studies has given us a much better picture of where, when, and how T5 recognizes this protein.

A number of bacteriophages can use more than one secondary receptor to infect its host, though these are less well-studied than bacteriophage that rely on a single receptor because of the complicated nature of screening for phages that use more than one protein. In contrast to the systematic genetic approaches taken to identify receptors for λ and T5, the identification of the host receptors used by bacteriophage Sf6 was more serendipitous. When purified phage particles were analyzed by separating the proteins in a gel matrix, two mystery proteins appeared on the gel along with the expected structural proteins. These proteins were later identified by mass spectrometry as OmpA and OmpC by Parent *et al.* (2012). Since these proteins were caught in the act of associating with bacteriophage particles, the researchers tested this further *via in vitro* genome ejection assays and mutational analyzes. To confirm that OmpA was a true secondary receptor, all the components for genome ejection – LPS, OmpA, and phage – were purified individually,

Table 2 Common bacteriophage receptors in Gram-negative bacteria

Outer membrane protein	Species	Bacteriophage(s)	
OmpA	<i>Escherichia coli</i>	TuII	K3
		K4	K5
		Ox2	Ox3
		Ox4	Ox5
		M1	Ac3
OmpC	<i>Shigella flexneri</i>	Sf6	
	<i>Escherichia coli</i>	T4	Tulb
		Me1	PA2
		Hy2	SS4
		434	SS1
		PP01	TP2
	<i>Salmonella enterica</i>	Gifsy-1	Gifsy-2
OmpF	<i>Shigella flexneri</i>	Sf6	
	<i>Klebsiella pneumoniae</i>	GH-K3	
	<i>Escherichia coli</i>	T2	Tula
		TP1	K20
		TP2	
	<i>Yersinia pestis</i>	Yep-phi	
	<i>Yersinia enterocolitica</i>	TG1	φR1-RT
LamB	<i>Escherichia coli</i>	K10	TP1
		λ	SS1
		Stx2φ-II	
Tsx	<i>Escherichia coli</i>	T6	H3
		H8	Ox1
		K9	
FhuA (TonA)	<i>Escherichia coli</i>	T5	D
		E21	T1
		UC-1	φ80
		mEp167	mEp237
		mEp410	mEp213
		mEp003	mEp023
		mEp043	mEp174
		mEp390	mEp416
		mEp506	HK022
	<i>Salmonella enterica</i>	ES18	
TolA	<i>Escherichia coli</i>	fd	f1
		M13	
BtuB	<i>Escherichia coli</i>	BF23	E15
		K6	K8
		K11	M3
		Ac4	EPS7
		SPC35	SPN7C
	<i>Salmonella enterica</i>	SPN9C	SPN10H
		SPN12C	SPN14
		SPN17T	SPN18
		BSP22a	EPS7
		T2	Stx2φ-I
FadL	<i>Escherichia coli</i>	Stx2φ-II	
FepA	<i>Escherichia coli</i>	H8	
	<i>Salmonella enterica</i>	H8	
TolC	<i>Escherichia coli</i>	TLS	
	<i>Salmonella enterica</i>	ST27	ST29
		ST35	
PhoE	<i>Escherichia coli</i>	TC45	TC23
Ail	<i>Yersinia pestis</i>	Yep-phi	
NfrA, NfrB	<i>Escherichia coli</i>	N4	
YncD	<i>Escherichia coli</i>	vB_EcoS_IME347	

Note: Modified from Hubbs, N.B., 2017. Hijacking the Cell: How Bacteriophage Sf6 uses *Shigella flexneri* Outer Membrane Proteins for Infection. Michigan State University. doi:10.25335/M5J19S.

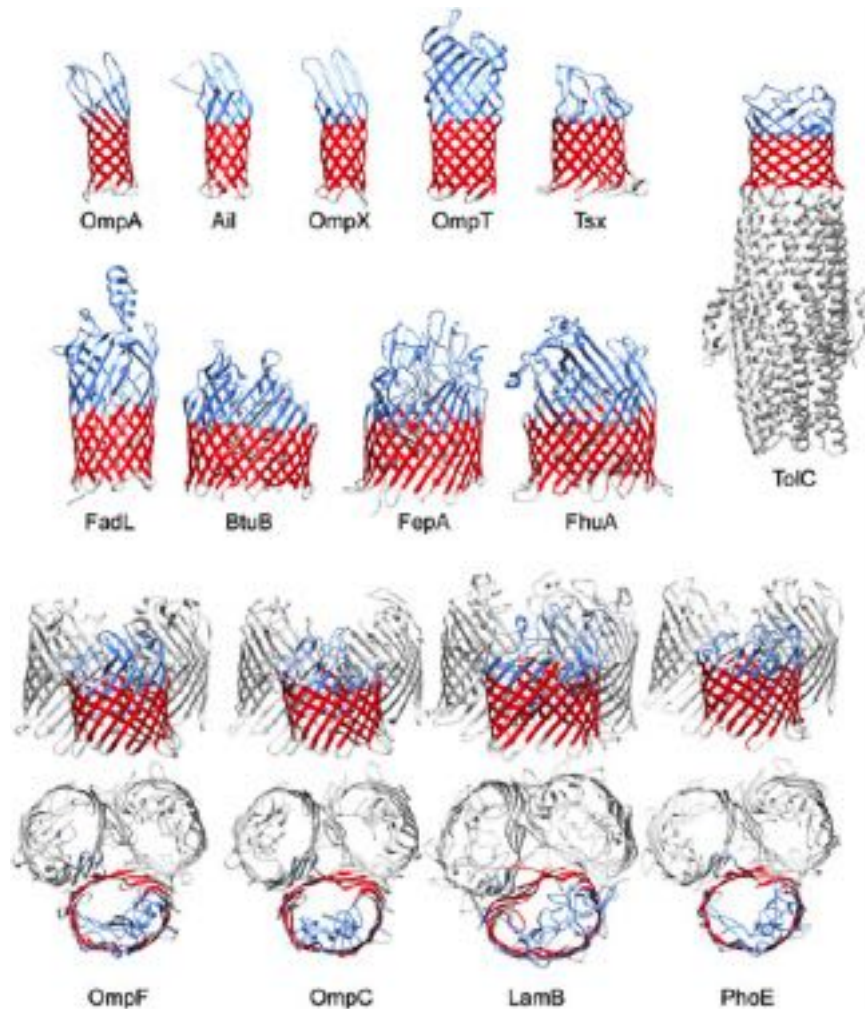


Fig. 2 Structures of representative outer membrane proteins, with extracellular domains shown in blue, transmembrane domains in red, and periplasmic regions in gray. Structures were obtained from the Protein Data Bank (PDB) with accession numbers as follows: OmpA (1BXW), Ail (3QRA), OmpX (1QJ9), OmpT (1I78), Tsx (1TLY), FadL (1T16), BtuB (1NQE), FepA (1FEP), FhuA (1QFG), OmpF (3POX), OmpC (3UPG), LamB (1AF6), PhoE (1PHO).

mixed in a test tube, and incubated. No genome ejection was observed in tubes with only phage, with phage and LPS, or with phage and OmpA alone; however, when a tube containing all three components was incubated, the phage genome was ejected. This indicates that both LPS and OmpA are necessary and sufficient for the phage to infect its host.

In vivo, Sf6 plating efficiency was tested on wild type cells, cells lacking either OmpA or OmpC, or cells lacking both OmpA and OmpC. Parent *et al.* determined that cell survival was much higher for the double *ompA/ompC* mutant compared to either the single mutant or the wild-type bacteria. This suggested that Sf6 can use either of two receptors to infect its host. In addition to plating experiments, cryo-EM studies showed that Sf6 can still bind to outer membrane vesicles produced from cells lacking either OmpA or OmpC but not from cells lacking both. Using plaque assays, they then screened for mutagenized *ompA* clones that would not restore plaque formation and determined that for Sf6 to recognize and bind to OmpA, the critical regions were extracellular loops extending from the β -barrel. With a 10-fold reduced plating efficiency, Sf6 is probably able to use more than two secondary receptors, as cells lacking both these receptors are not fully immune to Sf6. The identity of this mystery third receptor, or if there are even more than three, is currently unknown.

Another system that has been characterized is the *E. coli* phage T2 and its two receptors, OmpF and FadL. Only bacteria lacking both membrane proteins OmpF and FadL are immune to T2. OmpF was the first protein receptor identified for T2, but the ability of T2 to still infect OmpF-negative cells led to the discovery of FadL by Morona and Henning in 1986. To do this, Morona and Henning generated a transposon library in a strain lacking OmpF, then screened the strains of their library to see which mutants could survive T2 infection. Subsequent mapping of these T2-resistant strains indicated that the *ttr* locus was important for T2 infection. This locus encodes FadL, an outer membrane protein involved in fatty acid transport. Given the genetic data and location of the protein in the outer membrane, this was determined to be the other secondary receptor of T2. These researchers also determined the relative preference of T2 for its receptors. Outer membranes isolated from cells lacking OmpF were much more

effective in neutralizing T2 when compared to outer membranes isolated from cells lacking FadL, suggesting a much stronger interaction between T2 and FadL.

While many phages may not initially use more than one receptor, they can adapt to use a different or additional secondary receptor when propagated on mutant bacterial strains that lack the phage's preferred receptor. These mutants, known as host range mutants, can use a different outer membrane protein for infection and may or may not lose the ability to use the original. These types of studies have mostly been done for the bacteriophages Tula, M1, Ox2, all of which infect *E. coli*. Bacteriophage Ox2 uses the outer membrane protein OmpA for infection and has been shown to develop mutations in its genome to use OmpC or OmpX. When propagated on a strain lacking OmpA or a modified version of OmpA, Ox2 mutants that could use OmpC for infection were isolated. One mutant, Ox2h5, could not infect a strain lacking OmpA, while the two other mutants Ox2h10 and Ox2h12 could still infect this strain. These two mutants were further propagated on a strain lacking both OmpA and OmpF, which also expressed a modified version of OmpC. After a series of passages, five new host range mutants were isolated. These mutants were now able to use the outer membrane protein OmpX for infection.

Some phages may require the assistance of other membrane proteins for productive infection, which can complicate interpretation of results. One well-established example is the TonB dependence on FhuA to support infection by T1. While T1 appears to be "TonB-dependent," the activity of TonB is required for FhuA function rather than T1 binding to TonB directly. Though these confounding interactions appear to be limited, they can nevertheless make it difficult to understand the exact interactions between some membrane proteins and their phages.

With the renewed interest in phage therapy, phages able to use more than one receptor (e.g., OmpA and OmpC) or more than one type of receptor (e.g., TonB-dependent vs. non-TonB-dependent) are particularly useful. It is less likely that a host will become immune *via* single mutations or deletion of a single gene, thus rendering the phage more "flexible" in terms of its infectivity. Despite this, relatively few phages are known to use more than one receptor – or, if they are able to use more than one receptor, the additional receptor(s) have not yet been identified. This dual-receptor feature is particularly useful for phages and for our use of them in phage therapy, given the ability for phage and bacteria to co-evolve as discussed in the next section.

Ecology and Evolution of Phage-Host Interactions

Bacteria are capable of blocking a successful phage infection at multiple steps of the phage's replication cycle. One key mechanism of phage immunity is to ensure the phage does not gain access to the host at all. To do this, the host can use four main strategies: (1) receptor loss, (2) receptor mutation, (3) receptor blocking or masking, or (4) producing decoys.

The first strategy is an extreme but effective way to eliminate possible phage infection: to reduce the abundance of whatever protein the phage is recognizing, or to stop producing it altogether. This is most often seen in laboratory experiments, where bacteria are already grown in nutrient-rich media at optimal conditions. A classic example is the phage λ receptor LamB, which is required for maltose uptake in the *E. coli* host. In typical laboratory growth media, plenty of other nutrients are available, so the loss of maltose transport does not significantly hinder the growth of the host. In nature, however, the simple loss of a phage receptor may come with tradeoffs. For some pathogenic bacteria, the loss of a phage receptor actually reduces pathogenicity. This has been observed for *Vibrio cholerae* phage ICP2, which utilizes OmpU. Expression of *ompU* is regulated by the transcription factor ToxR, and deletion of the *toxR* gene renders the bacterium avirulent. In 2014, Seed and coworkers isolated strains of *V. cholerae* that were resistant to phage ICP2. Some of these strains contained mutations in either the *ompU* or *toxR* genes. When tested in an animal model, the *ompU* mutant strains showed slightly decreased fitness compared to wild-type *V. cholerae*. Strains containing the mutant *toxR*, however, were severely attenuated and unable to cause disease. The authors showed that these strains also did not express the *ompU* gene. These latter strains were indistinguishable from a strain completely lacking *toxR*. Thus, although the *toxR* mutant strains were resistant to phage ICP2, they were unable to colonize their mammalian host. This type of interaction is of particular interest for phage therapy applications, as a tradeoff between phage resistance and virulence is a "win-win" scenario.

Another example of receptor loss or reduction is that of "phase variation", where the levels of protein vary depending on environmental signals. An example of this is the phage KVP40, which infects *Vibrio anguillarum* *via* OmpK. Depending on the population density of *V. anguillarum*, individuals may up- or down-regulate expression of *ompK*. In dense populations, where many susceptible hosts are growing together, *ompK* is not expressed; however, as the population becomes less dense, the expression of *ompK* increases. While OmpK serves a function for the bacteria, this added "herd immunity" has a beneficial effect for the entire local population of *V. anguillarum*. Similarly, experiments done by Kim and Ryu in 2012 demonstrated that *Salmonella enterica* serovar Typhimurium will modify its LPS if the culture is shaken vigorously during growth. This modification can completely block phage infection, but the loss of this modification during non-shaking conditions will result in the bacteria becoming susceptible to phage infection once more.

Since the interaction between phage and host can be limited to a small region of the receptor protein, it is also possible for hosts to become immune to phage infection by simple mutations. Returning to the phage Sf6 and its host receptor OmpA, Porcek *et al.* demonstrated in 2015 that a change at just one amino acid in the attachment site could reduce phage infection by 80%. Similarly, mutations occurring in the LPS biosynthesis pathway often render the host immune to specific phages. Simple changes in LPS structure, such as an additional sugar group, can reduce phage infection by several orders of magnitude.

Even if the phage receptors themselves do not change, they can be blocked, masked, or sent into the extracellular space as decoys. Receptor access can be reduced by molecules that block them, such as some sugar molecules produced by species of

Pseudomonas. These produce a mucoid layer around the bacteria, preventing the phage from encountering any underlying receptors. Finally, many species of Gram-negative bacteria produce outer membrane vesicles, which are spheres that bleb off the membrane of the bacteria and are released into the surrounding environment. While these vesicles may not carry anything, they do have all components of the bacteria's outer membrane. Thus, phage can bind to the LPS and receptors and initiate an infection process. Instead of infecting a live cell, however, their genomes remain unused inside this decoy host.

If bacteria were always successful in preventing phage infection, this article would not exist. To overcome bacterial strategies of receptor loss, change, blocking, or decoys, the bacteriophages have a similar array to regain access to their hosts. These include: (1) being able to use multiple receptors, as discussed above; (2) targeting conserved regions of receptors; (3) mutations of the receptor-binding proteins, either by small changes or dramatic rearrangements; or (4) using a tail protein that changes conformation depending on which host protein(s) are available. In addition, phages can also use bacterial defense strategies to their benefit and prevent subsequent infection by other phages.

Phages that target conserved regions of receptors generally have a broad host range, as these regions in the membrane protein tend to be conserved due to functional requirements. Thus, they are conserved across many species and genera of bacteria. If the interaction can be modified by a single mutation in the phage's tail protein, this is another way for phages to regain access to their host. Meyer *et al.* showed in 2012 that bacteriophage λ was capable of recognizing a completely new receptor after only eight days of co-culturing the phage with its host in minimal media. The high number of replicates in this study showed that traversing this evolutionary landscape is highly parallel in laboratory settings, meaning that this evolutionary adaptation is highly favorable for the phage. The ability to recognize this new receptor, OmpF, was conferred by as few as four mutations in the phage's J protein, which forms the needle-like tip of the tail. These mutations mapped to the region known to interact with LamB. Since OmpF and LamB have similar global structures, these few point mutations were sufficient to confer favorable interactions between the phage and this new receptor. A more dramatic method is employed by T4 and related phages, which have hypervariable regions of their tail proteins. These can be rearranged or swapped between phages relatively easily, ensuring that at least some phages are not able to access a host if the host receptor protein becomes mutated. These tail fiber domains can allow the phage to use a different receptor or even a related receptor in a new host.

Even more innovative are the phages that can switch receptors depending on what's available in the environment. Some examples are the *Salmonella* phage SP6 and *E. coli* phages DT57C and DT571/2. These phages have V-shaped tailspikes that swivel to bind to whichever host is more prevalent in the environment. Similarly, phi92 has a radial array of at least 3 receptor-binding regions on its tail fibers. This allows the phage to infect both *Salmonella* and *E. coli* – two separate genera of enteric bacteria. Based on our current understanding, this is an unusual feature for most bacteriophages.

Besides recognizing alternate receptors, phage can actually affect receptor structures by producing their own proteins to modify their host's exterior surface. These modifications can occur in either primary or secondary receptor structures. For example, several *Shigella* phages and some *Salmonella* phages produce enzymes that add sugar moieties to the bacterial O-antigen. This O-antigen modification, known as *seroconversion*, can prevent other phages of the same or related species from double-infecting a host cell. This strategy requires the host to produce an O-antigen, which is not the case for all strains of bacteria.

In addition to LPS modifications, a few bacteriophages can encode "shielding proteins", which bind to secondary protein receptors. The related *E. coli* phages EPS7 and T5 each produce a shielding protein that targets a separate Omp. Phage EPS7 utilizes protein BtuB as a secondary receptor, whereas T5 uses FhuA as discussed above. The shielding proteins produced by these phages bind specifically to their own secondary receptors: the EPS7 protein BF23 binds to BtuB and the T5 protein Llp binds to FhuA. Shielding their own receptors specifically likely serves as a superinfection exclusion mechanism to prevent secondary infection by the same phage species. The effects of phage-encoded proteins on the exterior surface of their hosts will likely need to be studied further, especially if phage therapy starts being used to replace or supplement antibiotics.

Relevance of Phage-Receptor Interactions to Phage Therapy

Phage therapy involves the use of bacteriophages to treat bacterial infections and could be a viable strategy to combat drug resistant strains. Even though the use of phages for therapy was explored as early as 1917, very few countries have approved this treatment for infections. One of the key factors contributing to the under-development of phage therapy is the lack of extensive characterization of bacteriophages and their infection process. The ability of bacteria to develop resistance against phages by evolution is another key challenge, and a successful treatment would require more than one kind of bacteriophage. The use of phage cocktails to treat bacterial infections has gained importance recently, as it could be more effective against a diverse set of bacteria and utilize a suite of different receptors. The challenge is to develop a phage treatment protocol with different sets of phages that can be effectively utilized against an evolving population of bacteria.

One of the strategies that bacteria employ to combat phages is to suppress the production of the Omp that a particular phage utilizes, especially if the protein is not critical for growth in the current environment. Knowing which Omp a particular phage uses is critical to address the issue of bacterial resistance to phages. In addition to isolating novel phages that target different receptors, previously-characterized phages that utilize a set of Omps can be obtained by experimental evolution in the laboratory. Phage Ox2, which utilizes OmpA as its secondary receptor, can be evolved in the lab to utilize OmpC or OmpP/X instead of OmpA. Alternatively, screening for phages that can still infect a bacterial strain lacking its receptor can be a mechanism for generating a diverse population and for investigating co-evolutionary dynamics.

A unique approach developed by Turner *et al.* simultaneously utilizes both antibiotic resistance and phage resistance in *Pseudomonas aeruginosa* to neutralize infections. Certain bacteria utilize membrane-based pumps to expel toxic antibiotics – for example, TolC in *E. coli* or the Mex system of *P. aeruginosa*. Turner *et al.* screened for phages that are dependent on the outer membrane-porin M (OprM) for infection, a component of the drug efflux system, and isolated the phage OMKO1. If exposed to a combination of OMKO1 and antibiotics, the bacteria is doomed: it must downregulate or lose the receptor OprM to become resistant to phage OMKO1. Losing OprM will make the bacteria susceptible to antibiotics by rendering the drug efflux pump ineffective. In this study, researchers demonstrated that both laboratory strains and clinical isolates of *P. aeruginosa* that were resistant to OMKO1 phage were sensitive to antibiotic treatment. Confirming this deadly combination, the addition of both phage OMKO1 and tetracycline was more effective in inhibiting growth than the use of tetracycline alone. This combination treatment has been successfully utilized to clear *P. aeruginosa* infection in a patient's aortic graft. Continuing to identify phage outer membrane receptors that also serve as antibiotic efflux pumps in bacteria will greatly advance phage therapeutics and likely enhance the longevity of both types of treatments.

Summary

Our understanding of phage-host interactions has come a long way since the 1950s and is increasingly becoming more direct and precise. Although receptors are now known and characterized for many phages, we are far from saturated. In addition to not having a phage-receptor pair for all – or even most – types of phages or species of bacteria, several confounding factors may complicate our progress. First, although phages that utilize more than one secondary receptor may be most beneficial for phage therapy, it can be difficult to identify which receptors are used and in which order of preference. Second, Omps that require an additional membrane protein, such as the TonB-dependent Omps, can complicate the interpretation of results. Third, since some phages are able to modify their own receptors – and possibly the receptors of other phages – this may have additional effects on the host that we do not yet realize. Investigating these and other relationships involving phage-receptor interactions is critical for understanding bacteriophage biology, phage-host evolution, and for future development of phages for treatments.

Appendix: Methods for Receptor Identification or Characterization

To identify which receptor a phage recognizes, scientists can use genetic or protein-based screens. Genetic screens can be forward or reverse.

In a forward genetic approach, scientists find naturally-occurring mutants based on experimental conditions. For example, lytic phages and hosts can be co-cultured, sometimes in the presence of a mutagen, and any bacteria that survive phage attack would be isolated afterwards. If the surviving colonies “bred true” and were still resistant to the input phage, they would be analyzed further. Any mutations found in these survivors could indicate which genes and proteins are involved in phage susceptibility. This is a classic method, as it is possible to conduct screens for resistant mutants without immediately knowing an organism's genome sequence.

In reverse genetic screens, the sequence of the bacterial genome is known, and scientists can therefore create mutants by making specific bacterial genes nonfunctional and testing the effect of each gene product in the process of phage infection. These screens make use of single gene deletion (SGD) or transposon libraries. These libraries are constructed in populations of bacteria, where each individual in the population contains a deletion or disruption of only one gene. The population is large enough to ensure that every gene in the genome is disrupted. To determine which genes are important for bacteriophage attachment, researchers can systematically test the growth of phage on the thousands of strains within the collection and identify which disrupted gene(s) prevent and/or inhibit phage production. For transposon libraries, each strain has one gene that has been disrupted by the insertion of a transposon, usually covering every gene not required for bacterial survival. A powerful aspect of using a transposon library is that it can identify non-binary effects; for example, it is possible to identify genes that either delay or accelerate the infection process when disrupted by a transposon. The KEIO collection in *E. coli* and SGD library in *S. enterica* sv. Typhimurium have been used to identify many phage receptors and other host factors that are important for phage infection.

Identifying a bacteriophage's secondary receptor can also be done at the protein level. One very direct way to do this is to isolate phage-receptor protein complexes. To do this, researchers let the phage bind to its host, lysing the host, then re-isolating the phage with any host proteins still attached. The separation of phage and whichever host proteins, glycans, or even whole membrane vesicles are still bound can be achieved by several methods: ultracentrifugation (typically using a sucrose or cesium chloride gradient), precipitating the phage by crowding it out of solution (*e.g.*, via polyethylene glycol), or size exclusion filtration. Once the phage-receptor complex is separated from the rest of the sample, the associated proteins or surface glycans can be identified through a variety of biochemical means. If the interactions are transient, chemical treatment can covalently link the phage to its receptor(s). Otherwise, it can be difficult to capture complexes that are stable enough to persist throughout the preparation and treatment of the sample. Another complicating factor is that phages can bind to very complex structures like intact membrane vesicles. Binding to these vesicles means an entire blob of membrane will be associated with the phage. In these vesicles, there will be a number of proteins and surface glycans, so follow-up studies are needed to confirm which of the host products are the true or critical receptor(s).

Once the receptor's identity is known, it is useful to determine the structure of the protein and which regions are important for interaction with the virus. Structural studies can illustrate the function of a given protein and how it might interact with a phage. Both classic and modern structural studies require the receptor to be reasonably pure and stable during the experiment. Many phage receptors are transmembrane or membrane-associated proteins, which can be challenging to work with as discussed in [Table 1](#). There are two main levels of structural studies: atomic/near-atomic resolution, which is useful for looking at fine details; or low resolution, which can indicate the oligomeric state and general secondary or tertiary structure of a protein.

Structures obtained from atomic resolution techniques are extremely detailed and contain information about most, if not all, atoms in the structure.

A classic method to obtain this information is X-ray crystallography. Here, a purified sample is prepared in a way that causes the proteins to become highly ordered and densely packed, forming a solid crystal lattice of repeating subunits. Using X-ray beams, the angles and intensities of the resulting diffraction pattern can produce a three-dimensional picture of the sample's electron density. This describes the positions of all (or most) of the atoms and chemical bonds within the protein. Omp receptors – especially those with a β -barrel fold, as discussed below – are excellent proteins for this method due to their inherent stability. In most cases, stable proteins can be resolved at a high resolution, meaning more detailed analyses can follow.

Crystallography is not the only method to gain information about a protein's structure and/or function. A newer technique to do this is cryo-electron microscopy (cryo-EM). Cryo-EM is a form of transmission electron microscopy where samples are first flash frozen in amorphous, or vitreous ice. These samples are then imaged under cryo temperatures (below -180°C) in an electron microscope to collect 2D images of the proteins. A common "single particle" experiment requires collection of thousands to millions of images of individual proteins. Ideally, the proteins are sitting in the ice in many different orientations so the structure is visible from every angle. Computational algorithms applied to a subset of individual particles are then used to combine the 2D images and reconstruct a 3D model of the proteins. Recent advances in instrumentation, camera design, and the ability to handle big data have led to the cryo-EM "resolution revolution", where single particle cryo-EM studies are approaching the high resolution given by X-ray crystallography. Cryo-EM is especially useful for large, multimeric complexes, which are often a challenge in X-ray crystallography. It is also useful for less stable proteins or for proteins that do not crystallize easily. Despite these advantages, cryo-EM is currently limited by the size of the complex, so it is not possible for all samples. Smaller Omps, especially if they are monomers, are too small to be effectively imaged by cryo-EM. On the other hand, several of the giant viruses and jumbo phages are too big to effectively image while bound to cells or their purified receptors. With the cryo-EM field still evolving rapidly, this may not remain a barrier for long.

Not all Omps or phage-host complexes can be prepared or used for the atomic resolution techniques listed above. Instead, there are lower-resolution experiments that still provide information about the basic structure of a given phage receptor. For example, many phage receptors form a variety of oligomeric states that can be calculated by the molecular mass of the complexes. Different Omps can be found as monomers, trimers, or larger complexes. The oligomeric state of receptor proteins are commonly studied by native gels or chromatography, analytical ultracentrifugation, and native mass spectrometry. Native gels or chromatography can be used to sort protein complexes by mass in their native state, then compared to size standards to determine how many copies are found in each complex. Analytical ultracentrifugation combined with sucrose or cesium chloride gradients will separate complexes of different mass, shape, and/or density. Modifications of this technique can also quantitatively determine the affinity of proteins within multimeric complexes. Finally, native mass spectrometry – a newer and very powerful method – can actually be used to calculate the distribution of each complex type within a population. This allows researchers to determine whether protein complexes are highly dynamic and form many different oligomeric states, or if they are more uniform and predominantly have only one form.

Even if these specialized techniques are not available or practical, there are other methods that use optical measurements like fluorescence. Measuring the intrinsic fluorescence from aromatic amino acids or targeted fluorescence from the addition of certain chemicals are relatively quick and easy methods to determine a protein's tertiary structure. Circular dichroism, which measures differences in absorption of polarized light, is also a common technique to determine whether a protein is largely comprised of α -helices or β -sheets. Both of these optical techniques can be combined with chemical or thermal denaturants to test conditions that promote unfolding of the protein, providing information about the receptor stability. From these types of experiments, scientists can learn the secondary or tertiary structural elements of a protein or its inherent stability alone or in a complex.

Some properties of the phage-receptor binding interactions can be measured *in vitro*, including: binding affinity, kinetics of binding, and genome ejection after receptor docking. These approaches rely on the ability to purify the receptor protein and for the receptor to stay active in test tubes. To measure binding affinity and kinetics, increasingly popular platforms use optical measurements like fluorescence to measure how fast receptors bind to phage particles at varying conditions. For other types of phage receptors such as the primary LPS receptor, hydrolysis by the phage tail draws the phage closer to the cell membrane. Enzyme assays are therefore a popular method that can measure how quickly the phage degrades LPS and how quickly it can reach its secondary receptor. For secondary protein receptors which have no associated enzymatic activity, measuring association and disassociation rates can indicate binding strength between the phage and host protein. Techniques such as surface plasmon resonance (SPR), biolayer interferometry (BLI), and fluorescence anisotropy can all be used for these measurements. All of the aforementioned experiments can be done with whole phage or with purified phage tail proteins that interact with the phage's specific receptor. When used in combination, these techniques can paint a relatively complete picture of how the phage identifies and binds to its host.

Finally, with advances in computational capabilities, *in silico* methods are becoming common and powerful techniques to study biological processes. Computational approaches such as molecular dynamics (MD) can model the complex landscapes of membranes, membrane protein receptors, and complex structures such as LPS. As a virtual microscope, MD simulations describe structural biophysics of and biochemical interactions between all these components which can be used predictively to determine how membrane biophysics changes in different environments.

Which is the best method to use? Each has strengths and weaknesses, as discussed here and summarized in [Table 1](#). Like all science, a combination of experimental approaches more accurately captures the complexities of biology.

Acknowledgments

This material is based upon work supported by the National Institutes of Health R01 GM110185, the National Science Foundation CAREER 1750125 and Cooperative Agreement No. DBI-0939454, and the JK Billman, Jr., MD Endowed Research Professorship.

Further Reading

- Bohm, K., Porwollik, S., Chu, W., *et al.*, 2018. Genes affecting progression of bacteriophage P22 infection in *Salmonella* identified by transposon and single gene deletion screens. *Molecular Microbiology* 108 (3), 288–305.
- Casjens, S., Hendrix, R., 2015. Bacteriophage lambda: Early pioneer and still relevant. *Virology* 479–480, 310–330.
- Chan, B.K., Siström, M., Wertz, J.E., *et al.*, 2016. Phage selection restores antibiotic sensitivity in MDR *Pseudomonas aeruginosa*. *Scientific Reports* 6 (1), 26717.
- de Jonge, P.A., Nobrega, F.L., Brouns, S.J.J., Dutilh, B.E., 2019. Molecular and evolutionary determinants of bacteriophage host range. *Trends Microbiology* 27 (1), 51–63.
- Hancock, R.E., Reeves, P., 1975. Bacteriophage resistance in *Escherichia coli* K-12: General pattern of resistance. *Journal of Bacteriology* 121 (3), 983–993.
- Heller, K.J., 1992. Molecular interaction between bacteriophage and the Gram-negative cell envelope. *Archives of Microbiology* 158 (4), 235–248.
- Hubbs, N.B., 2017. Hijacking the Cell: How Bacteriophage Sf6 uses *Shigella flexneri* Outer Membrane Proteins for Infection. Michigan State University. doi:10.25335/M5J19S.
- Konovalova, A., Kahne, D.E., Silhavy, T.J., 2017. Outer membrane biogenesis. *Annual Review of Microbiology* 71, 539–556.
- Maynard, N.D., Birch, E.W., Sanghvi, J.C., *et al.*, 2010. A forward-genetic screen and dynamic analysis of lambda phage host-dependencies reveals an extensive interaction network and a new anti-viral strategy. *PLoS Genetics* 6 (7), e1001017.
- Morona, R., Henning, U., 1984. Host range mutants of bacteriophage O_{x2} can use two different outer membrane proteins of *Escherichia coli* K-12 as receptors. *Journal of Bacteriology* 159 (2), 579–582.
- Parent, K.N., Erb, M.L., Cardone, G., *et al.*, 2014. OmpA and OmpC are critical host factors for bacteriophage Sf6 entry in *Shigella*. *Molecular Microbiology* 92 (1), 47–60.
- Rothenberg, E., Sepulveda, L.A., Skinner, S.O., *et al.*, 2011. Single-virus tracking reveals a spatial receptor-dependent search mechanism. *Biophysical Journal* 100 (12), 2875–2882.
- Silva, J.B., Storms, Z., Sauvageau, D., 2016. Host receptors for bacteriophage adsorption. *FEMS Microbiology Letters* 363 (4), fnw002.
- van Houte, S., Buckling, A., Westra, E.R., 2016. Evolutionary ecology of prokaryotic immune systems. *Microbiology and Molecular Biology Reviews* 80 (3), 745–763.
- Wang, J., Hofnung, M., Charbit, A., 2000. The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of *Escherichia coli* K-12. *Journal of Bacteriology* 182 (2), 508–512.

Relevant Websites

- <https://cge.cbs.dtu.dk/services/HostPhinder/>
HostPhinder.
- <http://blanco.biomol.uci.edu/mpstruc/>
Membrane proteins of known structure.
- <https://opm.phar.umich.edu/>
Orientations of Proteins in Membranes (OPM) database.
- <https://phred.herokuapp.com/>
Phage Receptor Database (PhReD).
- <https://www.genome.jp/virushostdb/view/?VirusHostLineage=OtherVirus%7CBacteria>
Virus-Host Database.

Tail Structure and Dynamics

Shweta Bhatt, University of Copenhagen, Copenhagen, Denmark

Petr G Leiman, The University of Texas Medical Branch, Galveston, TX, United States

Nicholas MI Taylor, University of Copenhagen, Copenhagen, Denmark

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ATP Adenosine triphosphate

DNA Deoxyribonucleic acid

dsDNA Double-stranded DNA

gp Gene product (protein expressed by a particular gene)

ssDNA Single-stranded DNA

TMP Tape measure protein

Glossary

Caudovirales Order of dsDNA bacteriophages with an external tail structure. Includes the *Podoviridae*, *Siphoviridae* and *Myoviridae*.

Contractile injection system Protein complex related to (and including) the contractile tail apparatus of *Myoviridae*. Also includes the type VI secretion system, R-type pyocins, the antifeeding prophage/*Photorhabdus* virulence cassette and other tailocins.

Myoviridae Bacteriophages with a long, contractile tail.

N-fold symmetric assembly A protein complex comprised of N parts in which these parts are spatially related by a rotational axis of symmetry of order N is called an N-fold symmetric complex (e.g., a complex composed of three polypeptide chains that are related by a threefold is a threefold symmetric complex).

Podoviridae Bacteriophages with a short, non-contractile tail.

Siphoviridae Bacteriophages with a long, non-contractile tail.

Introduction

To replicate, bacterial viruses or (bacterio)phages have to infect their microbial hosts. Unlike eukaryotic viruses, which are usually taken up through endocytosis or membrane fusion, bacteriophages are required to translocate their genome and certain proteins across the host cell envelope. To address this challenge, an overwhelming number of phages use a structure known as a tail. The tail creates a conduit between the phage capsid and host cytoplasm and allows the phage particle to remain attached to the cell surface. Such phages are known as the tailed bacteriophages or *Caudovirales* (Fig. 1). The genome of all *Caudovirales* is a double-stranded DNA molecule, which is packaged into an icosahedral capsid. The proteinaceous tail structure is attached to a special vertex of the capsid through which the DNA is packaged during capsid assembly.

The tail is crucial in the infection process of *Caudovirales*: it is responsible for the initial host recognition and attachment that may include digestion or modification of the cell surface polysaccharides, for the irreversible attachment to the cell, for the degradation of the peptidoglycan, for sensing the “injection signal”, and for the translocation of DNA and proteins into the host. Because of the multitude of challenges that need to be overcome, it is not surprising that these tails are large, multiprotein assemblies of extreme complexity. For example, the tail of bacteriophage T4, one of best studied and more complicated phages, consists of a total of at least twenty different gene products present in varying stoichiometries.

In gram-positive bacteria and archaea, the genome needs to be injected across the only membrane of the cell. In gram-negative bacteria however, both the inner and outer membranes (as well as the peptidoglycan layer) need to be crossed, generating some additional requirements for the tail structures of phages infecting these organisms.

It should be noted that other viruses outside the order of the *Caudovirales*, have been found to have structures that are reminiscent of tails, and are used to inject the genome: e.g., the ssRNA virus MS2, the ssDNA virus ϕ X174, the dsDNA phage PRD1, the algal virus PBCV-1, certain archaeal viruses such as *Acidianus* two-tailed virus (ATV) and some eukaryotic viruses such as Herpes simplex. These viruses are discussed in other chapters of this collection.

Here, we will discuss the function and dynamics of the tail of the *Caudovirales*. We will examine the similarities and differences of all three families belonging to this order and point out specific differences between tails of bacteriophages targeting gram-positive and gram-negative bacteria.

Caudovirales

The order of the *Caudovirales* represents the largest group of bacteriophages, and therefore, of viruses. *Caudovirales* can infect either bacterial or archaeal hosts, suggesting that they share a common ancestor that is as old as the emergence of the bacteria circa

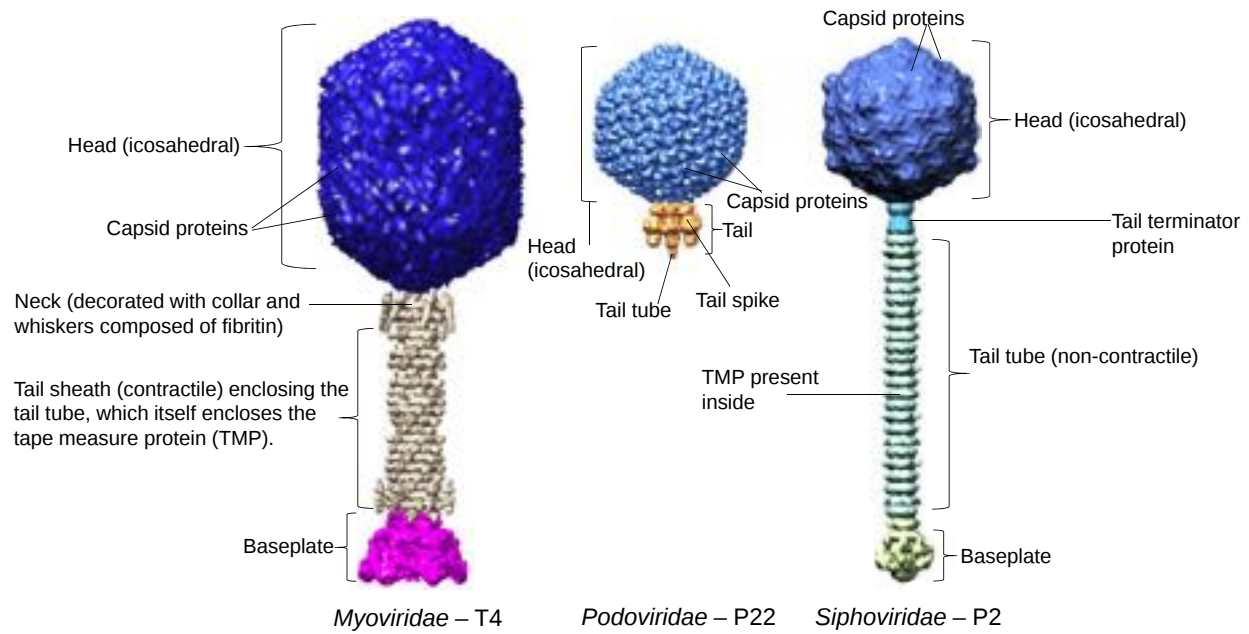


Fig. 1 Cryo-EM reconstructions of archetypal phages representing the three families in the order *Caudovirales*. From left to right: T4, which is characterized by the presence of a long tail surrounded by a contractile sheath and a terminal baseplate. P22, with a short tail and appendages. The lactococcal siphophage P2 with a long, flexible tail encompassed by a non-contractile sheath and a distal baseplate. Figure adapted from Veesler, D., Cambillau, C., 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiology and Molecular Biology Reviews* 75, 423–433.

3.5 billion years ago. The common architecture and sheer complexity of these viruses makes it near impossible to have arisen from convergent evolution.

The tail is attached to a special vertex of the capsid that is occupied not by a capsid protein but by a 12-fold symmetric mushroom- or cone-shaped protein called the head-to-tail connector or portal (**Fig. 1**). Most tails have (mostly) sixfold symmetry, although pseudo-sixfold/true threefold symmetric tails exist. The portal protein is responsible for symmetry adjustment between the tail and the five-fold symmetric vertex of the capsid which houses the portal. During phage particle morphogenesis and prior to tail attachment, the portal protein is part of the phage-encoded machinery that uses ATP to package the DNA genome into the preformed capsid.

Tails can be divided into three different types, and tail morphology has been one of the obvious and widely used characteristics to classify phages. There are bacteriophages with a short, non-contractile tail (*Podoviridae*); with a long, non-contractile tail (*Siphoviridae*) and with a long, contractile tail (*Myoviridae*) (**Fig. 1**). This classification emphasizes host recognition, binding, attachment and genome delivery mechanisms – all the processes that precede host takeover by the phage and production of new phage particles. In that aspect, post-DNA delivery behavior of phages that have different tail morphologies, e.g., P22 and lambda (*Podoviridae* and *Siphoviridae*, respectively), could be more similar to each other than those of the same type, e.g., P2 and T4 (both *Myoviridae*).

Siphoviridae are the most prevalent family of tailed phages (**Fig. 2**). They are characterized by their long tail which is tube-like in appearance. The length of the tail is regulated by a tape measure protein (TMP). *Myoviridae* are also a very common family (**Fig. 3**). They contain a tube that is structurally related to the tube of the *Siphoviridae* and which is also of defined length, controlled by a TMP. However, *Myoviridae* tails are more complex as they also contain a sheath structure wrapped around the tail tube. The high complexity of the sheath, the conservation of the tail sheath protein structure and the fact that *Myoviridae* can be found both in bacteria as well as archaea strongly suggests that *Myoviridae* only evolved once. The *Podoviridae* on the other hand are less prevalent (**Fig. 4**). They only target certain types of bacteria, which suggests that they might have evolved multiple times from the *Siphoviridae* and/or *Myoviridae*.

Organization and Assembly of Long Tails

Certain features of *Siphoviridae* and *Myoviridae* are similar, which is why we discuss them here together. In these phages, the capsid and the tail assemble independently from each other.

Upon completion of DNA packaging, one or two head completion proteins bind to the portal vertex of the capsid making the latter competent for tail attachment. In the absence of tail completion proteins, the DNA can leak from the capsid.

Tail assembly starts from the assembly of its capsid-distant part, which is called the tail tip complex in *Siphoviridae* or the central hub-spike complex in *Myoviridae*. In *Myoviridae*, the central hub-spike complex forms the centerpiece of a roughly planar structure called the baseplate. In both systems, this complex interacts with a “ruler” protein called the TMP, its chaperone, and other

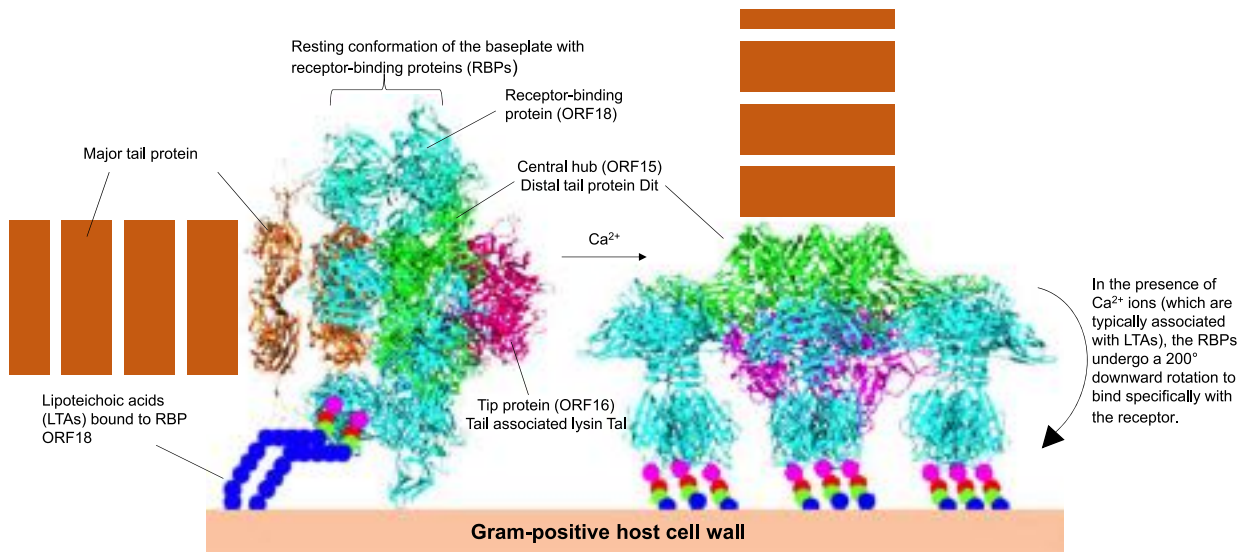


Fig. 2 Schematic representation of the conformational changes in the lactococcal phage P2 baseplate upon adsorption to the host cell wall components. The RBPs interact with the lipoteichoic acids through the host recognition domains and six RBP trimers undergo a 200° downward rotation, enabling them to interact with the host-specific “pellicle” phosphopolysaccharides. Adapted from Sciarra, G., Bebeacua, C., Bron, P., *et al.*, 2010. Structure of lactococcal phage p2 baseplate and its mechanism of activation. *Proceedings of the National Academy of Sciences of the United States of America* 107 (15), 6852–6857.

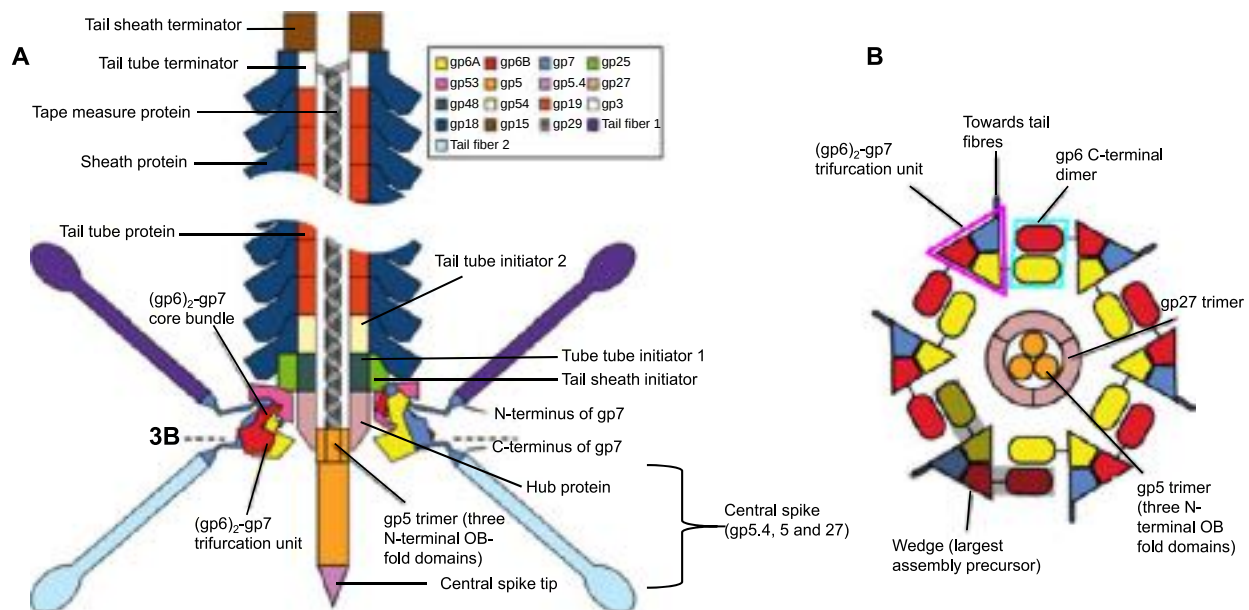


Fig. 3 (A) Organization and architecture of the *Myoviridae* tail. Schematic showing the components of the contractile tail, which have been labeled according to the gene products in bacteriophage T4. (B) Organization of the contractile tail baseplate. Figures have been adapted from Taylor, N.M.I., van Raaij, M.J., Leiman, P.G., 2018. Contractile injection systems of bacteriophages and related systems. *Molecular Microbiology* 108, 6–15.

chaperones whose conservation and functions are unclear. The TMP determines the eventual length of the tail. The tube is then assembled onto the tail tip/central hub-spike complex, first by binding one or two special tube initiator proteins and then by subsequent binding of multiple copies of the tube protein around the TMP (Figs. 2 and 3). The tube initiator proteins and the tube proteins are hexamers, so the trimeric tail tip/central hub-spike complex contains three-to-six-fold adapter domains. Binding of the tail tube terminator protein to the end of the growing tube completes tube assembly. The central hub protein, tube initiator proteins, tube protein and terminator protein are all structurally and evolutionary related.

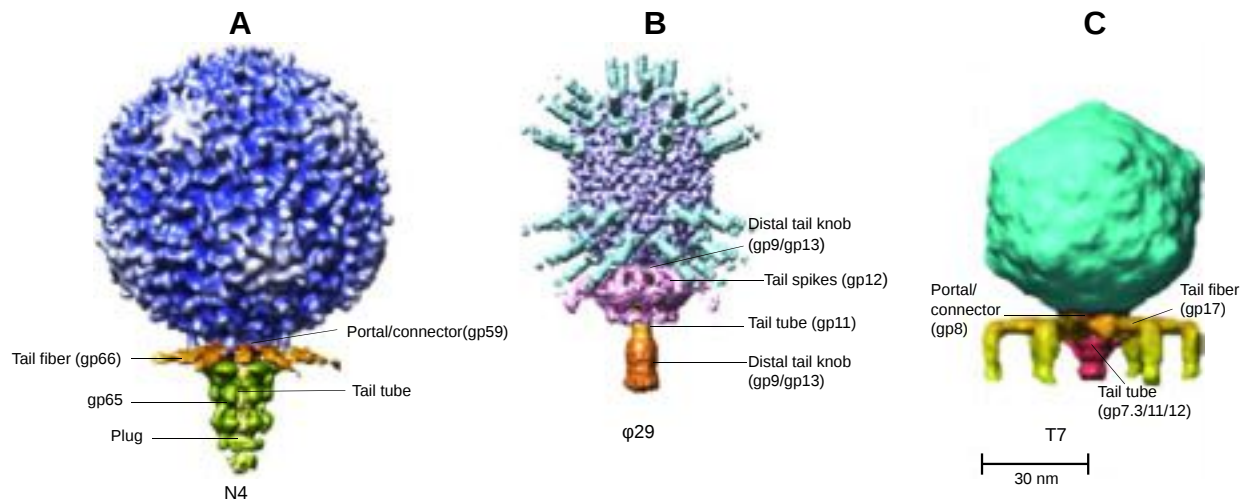


Fig. 4 Cryo-EM reconstructions of different phages of the *Podoviridae* family showcasing the diversity in size, tail structure and components. All the phages share a common structural feature in the form of fibers/appendages attached generally to the proximal part of the tail. Fig. A is adapted from Choi, K.H., McPartland, J., Kaganman, I., *et al.*, 2008. Insight into DNA and protein transport in double-stranded DNA viruses: The structure of bacteriophage N4. *Journal of Molecular Biology* 378, 726–736. Fig. B is adapted from Xiang, Y., Morais, M.C., Cohen, D.N., Bowman, V.D., Anderson, D.L., Rossmann, M.G., 2008. Crystal and Cryo-EM structural studies of a cell wall degrading enzyme in the bacteriophage ϕ 29 tail. *Proceedings of the National Academy of Sciences of the United States of America* 105 (28), 9552–9557. Fig. C is adapted from Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* 339 (6119), 576–579

In the *Siphoviridae*, binding of the tail tube terminator marks the end of tail synthesis. However, in the *Myoviridae*, tube completion is followed by the attachment of the tail sheath. One of the baseplate proteins (gp25 in bacteriophage T4) is homologous to one of the domains of the sheath protein (Fig. 3). It forms an integral part of the baseplate-proximal layer of the sheath and can therefore be thought of as the tail sheath initiator protein. The sheath itself is assembled as a six-start helix with a symmetry matching that of the tube. In the very last step of contractile tail assembly, the tail sheath completion protein binds to the tube terminator and to the last disk of the sheath.

In *Siphoviridae*, the head–tail interface is formed by the head completion proteins and the tail tube terminator. In *Myoviridae*, the head completion proteins instead bind to the tail sheath terminator. In bacteriophage T4, the final step in the assembly is the attachment of the long tail fibers to the tail. However, in some other systems, such as e.g., the R-type pyocin, the tail fiber is required for assembly of the baseplate showing that it acts early in the assembly process.

The Tube

The tube is organized around the TMP. The TMP is mainly α -helical in structure and assembles as a multimer of unknown stoichiometry. Given the long size of the tail, the TMP is likely crucial to transmit the “injection signal” from the tail tip complex to the capsid. It has been shown that the length (in number of amino acids) of the TMP correlates with the length of the tail across different phages. Furthermore, deletion of fragments of the tape-measure protein results in a concomitant shortening of the tail. The TMP likely interacts with both the tail tip complex and the tube terminator protein (Fig. 3). As mentioned previously, structural homology of these proteins positioned at both ends of the tube, as well as for the major tube protein, has been established for both *Myoviridae* and *Siphoviridae*. For bacteriophage T4, a myovirus, the structure of the pseudo-hexameric trimeric hub protein gp27, the two tube initiator proteins gp48 and gp54 (each forming a hexamer), and the tube protein gp19 (forming a tube of hexameric rings) have been determined experimentally. Even though the experimental structure of the tail tube terminator gp3 is unknown, its sequence is similar to that of the tube protein. For the *Siphoviridae* there are also several structures of the tube protein, exemplified by p2 ORF18 (pseudohexameric trimeric hub), p2 ORF15 (hexameric tube initiator protein), the N-terminal domain of bacteriophage λ gpV (major tail protein) and bacteriophage λ gpU (tail terminator protein).

The tube of most *Siphoviridae* phages is rather flexible. In contrast, the tube of bacteriophage T4 (and probably most myoviruses) is rigid, even in the absence of the tail sheath. One exception that has been described however is bacteriophage A511 which has a flexible tail although it is a myovirus.

The Sheath

The tail sheath of *Myoviridae* is a complex and intricate molecular machine. Its initial extended state is a high energy conformation whereas the contracted state found in the particle in the post-injection state is a low energy conformation. The contraction powers a drill-like

motion of the tube that results in breaching the integrity of the cell envelope of the host. The contraction is an irreversible process: this makes it especially important that it is only initiated at the right time, when the bacteriophage is correctly positioned on the cell.

The sheath is assembled starting from the baseplate, and the tube–baseplate complex is a necessary primer for sheath assembly. In T4, the sheath consists of 23 hexameric layers of sheath protein gp18 (six per layer). The extended sheath is a six-start helix, with a rise of 40.6 Å and a rotation of 17.2°, is 240 Å wide and 925 Å long. The symmetry of the sheath repeats the symmetry of the tube. This is an important fact, as it suggests the mechanistic basis for sheath assembly. The tube acts as a template for the assembly of the sheath in the extended, high-energy state: indeed, when gp18 is overexpressed in the absence of the baseplate–tail tube complex, the sheath is assembled into a low-energy state similar to the contracted tail sheath, which is only 420 Å long and 330 Å wide. The rise of the contracted sheath is 16.4 Å, and the rotation is 32.9°.

No atomic structure of an assembled bacteriophage sheath is available, but the structures of the sheath in the extended and contracted state for the related R-type pyocin and the type VI secretion system are known. Sequence similarity shows that the structure of the sheath subunit is conserved in all contractile injection systems, although the size of the subunit varies thanks to additional surface exposed domains in some systems. Also of note is that the sheath subunit of the type VI secretion system is encoded by two sequential genes. Each sheath subunit donates and accepts two long linker arms that interconnect all the subunits into a cylindrical mesh resembling a protective bottle sleeve. The connectivity of this mesh is the same in the extended and contracted state. In the contracted state, the entire mesh is simply widened and twisted. The linker arms of the first layer of the sheath subunits extend into the baseplate (where they interact with a T4 gp25-like protein) whereas the tail sheath terminator protein donates its linker arms to the last layer of the sheath subunits.

Baseplates of Long Contractile Tails

The baseplate is by far the most complex part of the contractile bacteriophage (Fig. 3). The best-characterized baseplate is the one from bacteriophage T4. It is believed that all contractile bacteriophages share a common architecture similar to the one of T4. However, each myophage will have very specific adaptations which are important to infect its host. These especially include enzymes to degrade cell walls and tail fibers to specifically bind cells of choice.

All known contractile tail baseplates are hexameric. They are built up of six “wedges”. In T4, the wedge consists of seven gene products, in different stoichiometries, which assemble in a specific order: gp11, gp10, gp7, gp8, gp6, gp53, and gp25. Once wedge synthesis is complete, the wedges can assemble around the central hub complex, consisting of gp27 (the central hub protein), gp29 (the TMP), gp5 and gp5.4. Gp26 and gp28 likely perform a chaperone function as they are not found in the fully assembled baseplate–tube complex. The formed, high energy baseplate has a “dome” shape. In vitro, in the absence of the central hub complex, the baseplates will assemble in a lower energy “star-shaped” form (see next section). After assembly of the dome-shape baseplate, the tube initiators (gp48 and gp54) can bind, followed by assembly of the tail tube and incorporation of the tail tube terminator as described in the previous section.

A minimal baseplate wedge has been proposed to consist of homologs of the following T4 proteins: gp6, gp7, gp25, gp53, and at least one type of tail fiber protein (usually trimeric). Two chemically identical copies of gp6 interact with gp7 to form a heterotrimer that comprises the main part of the wedge. Starting from the center and moving towards the periphery, the gp6–gp7 heterotrimer consists of a part where the three chains interdigitate or interact extensively (the core bundle and the trifurcation unit) and the part where they go their separate ways (two opposing dimerization domains of gp6 and fiber attachment domain of gp7). Gp25 is positioned at the tip of the core bundle and gp53 is wrapped around the bundle. The trifurcation unit sends the three chains comprising it in three different directions. The dimerization domains of two neighboring wedges interact to link the baseplate into a ring. The fiber attachment domain of gp7 forms a radial protrusion on this ring. The N-terminal part of T4 gp7 (which is upstream of the core bundle domain) also interacts with the tail fibers. However, in simpler baseplates (e.g., in phage P2, phage Mu, the R-type pyocin) but even in some complex baseplates (e.g., of SPO1-like phages), orthologs of gp7 do not contain equivalent N-terminal extensions. To summarize, gp6 has an important role in the circularization of the baseplate and gp7 is crucial for the interaction with the tail fiber network (Fig. 3(B)).

Apart from these conserved baseplate proteins, each bacteriophage has a specific set of tail fibers and other structural proteins and/or adaptations. In the case of T4, gp7 is an extended protein, and together with gp8 and gp9 (a dimer and a trimer, respectively), it makes up the intermediate or less conserved part of the baseplate, which is contracting the tail fiber network. In T4, the tail fiber network consists of two different types of tail fibers: the short and the long tail fibers. The short tail fibers (trimers of gp12) are “curled up” around the periphery of the baseplate and form part of the short tail fiber network, which also consists of two other trimeric proteins (gp10 and gp11). The long tail fibers (which have a stoichiometry $(gp34)_3(gp35)_1(gp36)_3(gp37)_3$) bind to the gp9 trimer. These viral adhesins have a tremendous length (~ 1600 Å) and recognize both LPS as well as the outer membrane protein (OmpC). In the assembled T4 particle, the long tail fibers can interact with the neck protein fibrin (gp wac) when the former are in the retracted state. Upon infection, the long tail fibers need to become extended. This regulation might prevent the inadvertent binding of the tail fibers to the host in conditions that are unfavorable. In vitro, retraction and extension can be controlled by pH, ionic strength, temperature and polyethylene glycol concentration.

During the infection process, it is thought that the binding of the long tail fibers to the cell surface induces a conformational change in the baseplate (going from the high energy, dome-shaped state to the low energy, star-shaped state) which also allows binding of the short tail fibers. Although the precise sequence of events is not clear, the “end state” of this process is structurally well understood from the

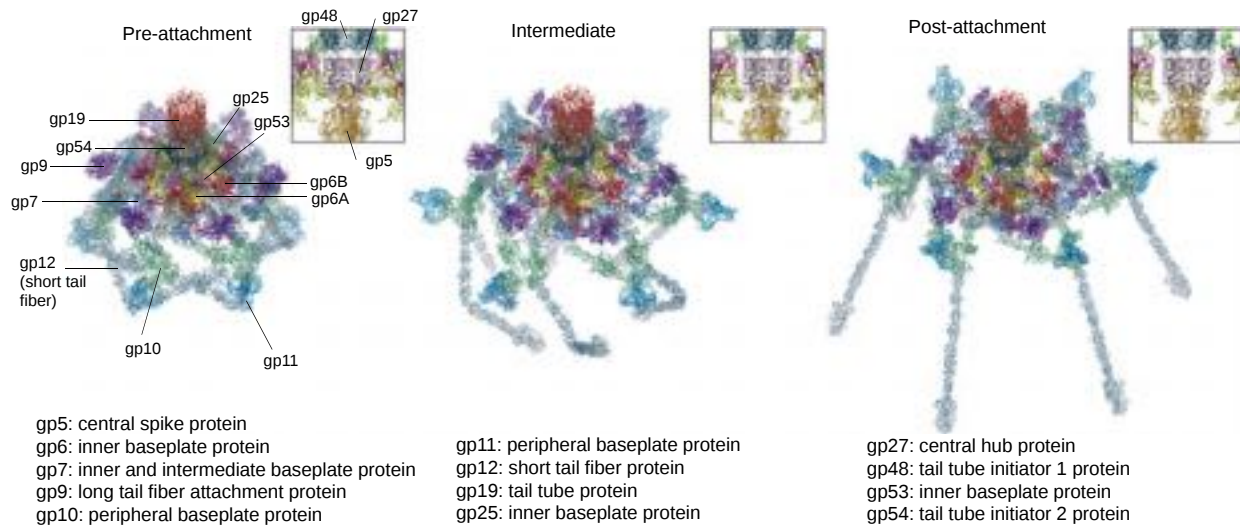


Fig. 5 Structure of the complex baseplate of phage T4 in a pre-attachment (from cryo-EM), intermediate (modeled) and a post-host-attachment state (based on cryo-EM). The insets are focused on the central part of the baseplate and demonstrate a release of the tail tube–central spike complex, which remains in a fixed position throughout the transformation. Figure reproduced from Taylor, N.M.I., Prokhorov, N.S., Guerrero-Ferreira, R.C., *et al.*, 2016. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* 533, 346–352.

structure of baseplates in the “star-shaped” state (Fig. 5). In this post-attachment-mimicking state, the long tail fibers make extended interactions with the baseplate, especially gp11. The short tail fibers are then extended and can interact with the LPS. This is accompanied by a slight opening of the iris formed by gp6 and gp7 and results in an outward movement of the gp6-gp7 core bundles. As gp25 is attached to the tip of the bundles, it also moves outward, pulling on the sheath, initiating its contraction.

As already alluded to, even though the architecture of the baseplate is universally conserved, adaptation to different hosts results in tail fibers and receptor binding proteins (RBPs) of extreme complexity. In fact, the size and complexity of RBPs dwarf the actual baseplate in many large bacteriophages, e.g., in Twort-like phages (e.g., *Listeria monocytogenes* phage A511, *Staphylococcus aureus* phage ϕ 812) or Vil-like phages (e.g., *Escherichia coli* phage CBA120).

Tail Tip Complex of Long Non-Contractile Tails

The tail tip complex and the adjacent part of the tail in siphophages is similar to the basal part of the tail tube and the baseplate hub of myophages. The tip protein, called Tal (tail-associated lysin) in siphophages, is homologous to T4 gp27 (Fig. 2). The protein that forms the interface between Tal and the rest of the tail is called Dit (distal tail protein). Dit is homologous to the T4 tube initiator protein gp48. The tail fibers are attached directly to this complex. In certain phages infecting *Lactococcus lactis* the RBP network is so extensive that it forms a large planar structure that by analogy with myophages is called a ‘baseplate’. However, this ‘baseplate’ appears to be structurally and evolutionary unrelated to myophage baseplate. In some of these phages (e.g., lactophage p2) the RBPs perform an acrobatic move during attachment to the host cell surface – they rotate by 200° upon binding to the cell surface polysaccharides (Fig. 2). This conformational change is accompanied by opening of the Tal trimer, which allows the TMP to exit the tail. The TMP forms a channel across the cell envelope through which the DNA is translocated into the host cytoplasm. Of note, whether the TMP participates in forming a membrane channel in myophages is unclear.

In the case of the coliphage T5, three L-shaped fibers (pb1) are attached to sides of the baseplate-hub protein (BHP) pb3 and one straight fiber containing a single RBP is attached to its center. Pb3 is the Tal protein of T5, as it adopts the same protein fold as that seen with the BHP (gp27) of the T4 myophage. At the end of the tail tube, a two-domain tail tip protein (TTP) pb9 is located in the upper region of the tail tip cone. X-ray crystallography studies have revealed structural similarities between domain A of pb9 and the N-terminal domains of known Dit proteins, such as those present in the lactophage p2. It has hence been proposed that Dit protein modules are likely to be a conserved structural motif amongst both Gram positive and gram-negative host siphophages. It has also been proposed that pb2 forms the TMP and possesses fusogenic and muralytic activity.

In T5, host recognition occurs via the reversible binding of three L-shaped fibers (pb1) to the O-antigen of the lipopolysaccharide. The irreversible interaction of the RBP (pb5) that forms the tip of the central tail fiber with an outer membrane iron-siderophore receptor FhuA initiates a cascade of conformational changes, leading to the expulsion of TMP which digests the peptidoglycan and perforates the cell membrane. TMP plays a role in signal transduction by initiating the opening of the head-tail connector, which leads to DNA release from the capsid. The TMP likely forms a channel that extends the tail through the periplasm and allows the DNA to reach the host cytoplasm.

Short Non-Contractile Tails: Tube-Like Tail Structure

Bacteriophages belonging to the family *Podoviridae* are characterized by the presence of short non-contractile tails varying in lengths in the range of 10–30 nm (Fig. 4). They are the smallest family in the *Caudovirales*, considering the number of different phages discovered. In podophages, the tail is assembled directly onto the portal protein of the capsid after the DNA packaging is complete. The *Podoviridae* tail is generally comprised of two different proteins – a dodecameric protein that interacts with the portal protein (also called, confusingly, the head-to-tail connector in some phages), and a second hexameric protein which forms the rest of the tail tube.

One of the most extensively studied podophages is the gram-positive *Bacillus* phage ϕ 29 (Fig. 4). In ϕ 29, the tail is approximately 380 Å long. Its portal-proximal part consists of dodecameric gp11 and is called the lower collar. It is a funnel-like structure to which twelve receptor-binding tailspikes (called ‘appendages’ in the ϕ 29 literature) are connected. In another podophage, P22, X-ray crystallography has revealed a dodecameric toroidal gp4 complex (called the ‘head-tail connector’ in the P22 literature) to which six P22 tailspike complexes are attached. Phage RBPs that carry enzymatic domains and thus are stockier in their appearance than fibers in electron microscopy images own their ‘tailspike’ name to the P22 tailspikes as they looked like small spikes attached to the P22 tail. A combination of the location, the stoichiometry and the predicted high α -helical content suggests that gp11 of ϕ 29, gp4 of P22 and the head completion proteins of the long-tailed phages could be evolutionarily related.

The distal part of the ϕ 29 tail tube is thicker than the portal-proximal part. It is called the tail “knob” and is formed by a hexamer of gp9. It plays a key role in the infection process because the distal end of the ϕ 29 tube is blocked by a long loop of gp9 prior to DNA ejection. Interestingly, not only the structure of the gp12 tail knob of the *Streptococcus* podophage C1 is similar, but the tubular domain of ϕ 29 gp9 and C1 gp12 is found in the tail tube, neck and baseplate proteins of myophages and siphophages – such as in the gp27 baseplate hub protein of T4 and in the λ tail tube protein gpV. This is an indication that the tails of all three families of phages (*Myoviridae*, *Siphoviridae* and *Podoviridae*) have evolved from a common ancestor.

The tail tip or distal part of the tail tube typically contains proteins that aid either in digesting the host cell wall, or in host recognition or both. ϕ 29 degrades the *Bacillus* cell wall with the help of an enzyme located at the distal end of its tail knob (gp13), capable of cleaving the polysaccharide backbone and the peptide cross-links of the peptidoglycan layer. In phage P22, a long needle extending outwards from the center of the tail base is constructed by a single trimer of gp26. It has been observed that in addition to making the first contact with the host surface, the N-terminal end of the needle acts as a “plug” and prevents DNA leakage.

The Tailspikes, Tail Fibers and Phage-Host Interaction

As the host receptors continuously evolve, the genes present in the tail fibers, tailspikes or tail appendages are selectively pressured to adapt to the ever-changing target, which can range from peptide sequences to polysaccharide moieties. The thin, long tail fibers are rigid and are flexibly attached to the tail. They scout the cell surface for a specific binding target and initiate adsorption. Tailspike proteins on the other hand are shorter, stubbier and possess enzymatic activity (although the substrates are known only for a subset of tailspikes). Tailspikes digest or modify cell surface polysaccharides, thus irreversibly binding the phage particle to the bacterial surface and allowing the tail tube to reach the cell membrane. This event triggers conformational changes in the phage, leading to the formation of a channel or a pore through which genetic material and proteins located in the capsid can traverse across the bacterial cell wall. The tailspikes or fibers hence play a crucial role in the initial host recognition and subsequently in triggering the transfer of the genome.

There is considerable diversity in terms of length, shape and specific functionality not only in fibers and tailspikes of myo-, siph-, and podophages, but even within each group. The majority of these spikes and fibers are homotrimeric – such as gp12 of ϕ 29 and gp17 of T7. In phages such as ϕ 29, the tailspike appendages recognize and cleave the glucosylated poly-glycerol phosphate teichoic acid by a mechanism similar to that seen in some ribonucleases. In phage C1, the appendages are more tail fiber-like and form a skirt around the tail. These appendages carry a flexible globular domain at their distal end.

The dynamics of phage-host interaction depend on the nature of the bacterial host. Gram-positive and gram-negative bacteria differ in cell wall thickness and components, which have led phages to the adoption of different infection strategies. Gram-positive bacteria lack an outer membrane and phages typically interact with cell wall-associated teichoic acids and other glycopolymers embedded in the thick peptidoglycan layer. They must overcome this physical barrier to reach the cell membrane and deliver their genome into the bacterial host cytoplasm. After the phage has reached the cell membrane, it is thought that there needs to be some protein–membrane fusion event. The archetypal phage ϕ 29 relies on the formation of a cone-shaped structure constructed out of the long loop of gp9 that exits from the tail knob upon genome release. This structure functions as a membrane pore, similar to a hydrophobic fusion peptide, which is found in enveloped eukaryotic viruses and of a membrane-active peptide that is present in non-enveloped viruses. This shared mechanism employed by eukaryotic and prokaryotic viruses is likely a consequence of convergent evolution.

For phages of gram-negative hosts, the DNA needs to cross the peptidoglycan and the inner membrane layers. As discussed above, the tails of podo- and siphophages have to be extended by phage and, possibly, host proteins that connect the tail with the cytoplasm. In myophages, the tail tube crosses the outer membrane and the periplasmic space to contact the cytoplasmic membrane, which is invaginated near the tip of the tube. The involvement of other phage proteins (e.g., the TMP) or host proteins in creating the channel spanning the cytoplasmic membrane is unclear and the tip of the tube might be capable of crossing the membrane.

The outer membrane of gram-negative bacteria is dotted with highly conserved porin proteins that serve as reversible binding sites for phage tail fibers. These fibers are also known to interact with the inner (e.g., fibers of bacteriophage T4) or outer (e.g., fibers of bacteriophage ϕ CTX from the *Myoviridae* family) core of LPS. Phage tail spikes on the other hand recognize and degrade the O-antigen (the polysaccharide part of LPS), which brings the particle closer to the bacterial cell surface, thus enabling the recognition of a second receptor in the outer-membrane, which paves the way for irreversible binding and triggers DNA release upon the proper orientation of the phage particle on the cell surface. A number of phages carry multiple sets of tailspike proteins on the particle which allow them to expand their host range from different bacterial strains to different bacterial species (e.g., phage K1–5 infects *E. coli* K1 and K5, phage SP6 infects different *Salmonella enterica* serogroups, and phage SFP10 infects *S. enterica* and *E. coli* O157:H7).

Phages from the T7 supergroup infect most *E. coli* strains (Fig. 4). The T7 particle contains six trimeric fibers of gp17 attached to the proximal end of the tail tube (to the dodecameric gp11 ring). Even though the other end of the fiber is free, and that part will eventually interact with a cell surface receptor, in the free state of the phage most of the fibers are bound to the capsid. Similar to the more complex T4 fibers and the side fibers of the siphophage T5, the T7 fibers are not straight but are bent roughly in the middle. In fact, the T7 fibers resembles the letter “L”. The rough LPS forming the bacterial outer membrane is the main receptor for T7 fiber binding. This interaction is accomplished by rigid body rotation of the fibers and leads to ejection of the capsid core proteins. The latter form an extension to the tail and connect it to the cytoplasm. The extended tail then acts as the conduit for DNA transfer across the bacterial membranes while also protecting it from periplasmic nucleases.

N4-like bacteriophages usually encode at least two different tailspike/tail fiber proteins with the larger protein directly attached to the phage tail and the smaller protein bound to the longer one (Fig. 4). The shorter tailspike (gp63.1) of the N4-like phage G7C is a deacetylase, and the host recognition of G7C depends not only on the main chain structure of the O-antigen, but on the presence of the acetyl group at a certain position of that particular O-antigen. N4 was the first phage that was shown to contain a large virion-encapsidated RNA polymerase (vRNAP) inside its capsid. This enzyme exits the capsid through the tail during infection and participates in genome transfer. The accumulating body of data shows that many phages contain vRNAPs in their capsids that are delivered (together with the genome) into the host cell cytoplasm where they initiate RNA synthesis, although such phages are usually much larger than N4 (e.g., AR9 and PBS1, giant phages of *Bacillus subtilis*).

Acknowledgments

The Novo Nordisk Foundation Center for Protein Research is supported financially by the Novo Nordisk Foundation (NNF14CC0001). This work was also supported by an NNF Hallas-Møller Emerging Investigator grant (NNF17OC0031006) to NMIT. NMIT is a member of the Integrative Structural Biology Cluster (ISBUC) at the University of Copenhagen.

Further Reading

- Ackermann, H.W., 1998. Tailed bacteriophages: The order caudovirales. *Advances in Virus Research* 51, 135–201.
- Aksyuk, A.A., Bowman, V.D., Kaufmann, B., *et al.*, 2012. Structural investigations of a Podoviridae streptococcus phage C1, implications for the mechanism of viral entry. *Proceedings of the National Academy of Sciences of the United States of America* 109, 14001–14006.
- Arnaud, C.-A., Effantin, G., Vivès, C., *et al.*, 2017. Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. *Nature Communications* 8, 1953.
- Bebeacua, C., Tremblay, D., Farenc, C., *et al.*, 2013. Structure, adsorption to host, and infection mechanism of virulent lactococcal phage p2. *Journal of Virology* 87, 12302–12312.
- Choi, K.H., McPartland, J., Kaganman, I., *et al.*, 2008. Insight into DNA and protein transport in double-stranded DNA viruses: The structure of bacteriophage N4. *Journal of Molecular Biology* 378, 726–736.
- Fokine, A., Rossmann, M.G., 2014. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* 4, e28281.
- Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* 339, 576–579.
- Lander, G.C., Khayat, R., Li, R., *et al.*, 2009. The P22 tail machine at subnanometer resolution reveals the architecture of an infection conduit. *Structure* 17, 789–799.
- Leiman, P.G., Arisaka, F., van Raaij, M.J., *et al.*, 2010. Morphogenesis of the T4 tail and tail fibers. *Virology Journal* 7, 355.
- Prokhorov, N.S., Riccio, C., Zdoboveno, E.L., *et al.*, 2017. Function of bacteriophage G7C esterase tailspike in host cell adsorption. *Molecular Microbiology* 105, 385–398.
- Sokolova, M., Borukhov, S., Lavysh, D., *et al.*, 2017. A non-canonical multisubunit RNA polymerase encoded by the AR9 phage recognizes the template strand of its uracil-containing promoters. *Nucleic Acids Research* 45, 5958–5967.
- Taylor, N.M.I., Prokhorov, N.S., Guerrero-Ferreira, R.C., *et al.*, 2016. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* 533, 346–352.
- Taylor, N.M.I., van Raaij, M.J., Leiman, P.G., 2018. Contractile injection systems of bacteriophages and related systems. *Molecular Microbiology* 108, 6–15.
- Veesler, D., Cambillau, C., 2011. A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. *Microbiology and Molecular Biology Reviews* 75, 423–433.
- Xu, J., Gui, M., Wang, D., Xiang, Y., 2016. The bacteriophage ϕ 29 tail possesses a pore-forming loop for cell membrane penetration. *Nature* 534, 544–547.

Relevant Website

<https://www.khanacademy.org/science/biology/biology-of-viruses/virus-biology/a/bacteriophages>
Bacteriophages – Viruses – Khan Academy.

Bacteriophage Tail Fibres, Tailspikes, and Bacterial Receptor Interaction

Mateo Seoane-Blanco and Mark J van Raaij, National Center for Biotechnology, Madrid, Spain

Meritxell Granell, National Center for Biotechnology, Madrid, Spain and Institute of Chemical Research of Catalonia (ICIQ), Tarragona, Spain

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Beta-structure Structural element of a protein made up of beta-strands (beta-strands are elongated parts of protein chains connected to each other via multiple main-chain hydrogen bonds, they can be parallel or anti-parallel); beta-structured elements can be in the form of beta-hairpins, beta-sheets or beta-barrels, among others.

Lipopolysaccharide (LPS) Cell wall component of Gram-negative bacteria which consist of a lipid A membrane anchor and a carbohydrate extra-cellular part. The carbohydrate part is composed of an inner core, an outer core and the variable O-antigen. Lipopolysaccharide is also known as endotoxin; its toxicity is due to our immune system reacting to them.

Peptidoglycan Bacterial copolymer consisting of carbohydrates and amino acids that forms a mesh-like layer

outside the plasma membrane. It contains alternating residues of beta-(1,4) linked N-acetylglucosamine and N-acetylmuramic acid. A short peptide chain joins N-acetylmuramic acid residues together, leading to a three-dimensional network. Also called murein.

Tail fiber Long, thin receptor-binding proteins or protein complexes attached to the tail of a bacteriophage. A typical example are the six long tail fibers of bacteriophage T4.

Tailspike Shorter, stubby receptor-binding proteins attached to the tail of a bacteriophage. A typical example is the tailspike of bacteriophage P22.

Teichoic acid Bacterial copolymers of glycerol phosphate (or ribitol phosphate) and carbohydrates that are linked to each other through phosphodiester bonds, fortifying the cell wall of Gram-positive bacteria.

Introduction

Tailed bacteriophages, belonging to the *Caudovirales* order ("cauda" is Latin for tail), are formed by an icosahedral or prolate capsid, a neck complex, and a short or long tail (Fig. 1). For some phages, these tails are short (podoviruses), for others they are long and flexible (siphoviruses) and for yet others they are composed of two concentric tubes, of which the outer one is rigid and contractile (myoviruses). Phage tails are very important for correct host recognition and efficient DNA transfer into the bacterium. For phages infecting Gram-negative bacteria, the genome must avoid the periplasmic space, where many nucleases reside. Just before DNA ejection, members of the *Podoviridae* family (like T7 and P22) expel multiple copies of internal capsid proteins, which then assemble into a tube, spanning the cell wall and contacting the cytoplasmic membrane (Fig. 1(e) and (f)). Members of the *Myoviridae* family contract the outer tail sheath, driving the inner tail tube into the cell wall (Fig. 1(c) and (d)). Less is known about the exact infection mechanism for the *Siphoviridae* family, but they may recognize existing membrane-spanning protein complexes and subvert these to promote DNA transfer into the cytoplasm (Fig. 1(a) and (b)).

Before DNA transfer can take place, however, the correct bacterium has to be identified. Bacteriophages recognize exposed structures of the bacterium they are going to infect (Fig. 1(g) and (h)). These structures may be proteins, like flagella, pilins, porins or transporters, or oligosaccharides, like lipo-polysaccharide (LPS) for Gram-negative bacteria and teichoic acid for Gram-positive bacteria. Accurate host cell recognition is essential, because a phage that ejects its genome into the medium or into the wrong cell can't replicate. Phage receptor-binding proteins perform this selection of encountered bacteria. These receptor-binding proteins are usually present in multiple copies. Furthermore, each of these are in general homo-trimeric, i.e., they consist of three identical subunits. The presence of multiple trimeric copies leads to a strong avidity effect, allowing relatively weak individual interactions to contribute to strong binding of the phage.

Bacteriophages may first encounter a potential host cell with their heads instead of their tails and maintaining contact with this bacterium could be useful. Candidates for these head-host interactions are the head fibers of phi29, the immunoglobulin-like domains found in structural proteins of many phages and the Dec protein on the bacteriophage L head. These interactions may allow additional time for the tail fibers or tailspikes to recognize the primary receptor on the cell surface. Alternatively, these head proteins may also serve to maintain phages in an environment where suitable host bacteria may be encountered, for example by binding to intestinal mucus.

Adsorption of the phage tail to the host often takes place in two steps. First, via a reversible interaction with a specific primary receptor on the surface of the target cell and then, irreversibly binding a secondary receptor. The phage proteins that mediate these interactions are usually tailspikes or tail fibers.

Phage receptor-binding proteins can be divided into several classes (Fig. 2). Fibers are long and thin, and sometimes composed of more than one type of protein chain, like in phage T4 (Fig. 2(a)). Their length provides reach, perhaps necessary to bind distant receptors on the bacterial surface. Tailspikes are shorter and stubby and may contain an enzymatic function, helping to digest host capsule polysaccharides and allowing the phage to reach the host cell membrane. An example are the tailspikes of phage P22 (Fig. 2(e)). Atomic structures of phage receptor-binding proteins have revealed new protein folds consisting of mainly beta-sheets with different topologies. Often, the three monomers are intertwined, stabilizing the proteins and explaining their resistance to heat and denaturants.

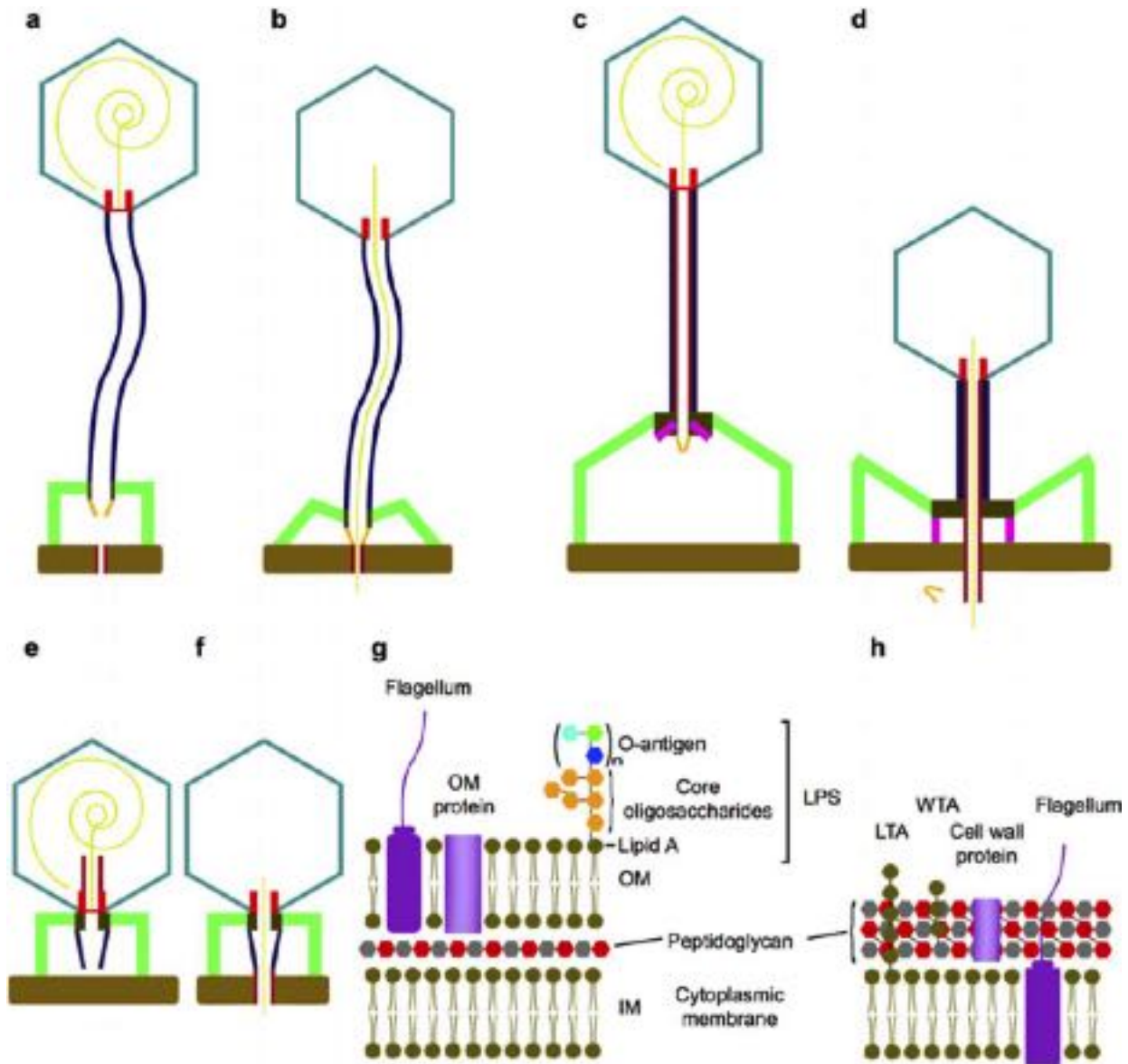


Fig. 1 Tailed bacteriophages and their infection mechanisms. Schematic drawing of a siphovirus attached to the cell wall (in brown) in pre-infection conformation (a) and in post-infection conformation (b). The capsid wall is colored blue, the DNA yellow, the portal complex red, the sheath dark blue, the baseplate dark brown, the tail fibers green and the spike in orange. Schematic drawing of a myovirus in pre-infection conformation (c) and in post-infection conformation (d). Coloring is as for the siphovirus, with the addition of the tail tube in dark red and the short tail fibers in pink. Schematic drawing of a podovirus in pre-infection conformation (e) and in post-infection conformation (f). The tail tube is colored dark blue and the core proteins are colored dark red. Schematic representations of a typical cell wall of a Gram-negative (g) and of a Gram-positive (h) bacterium are also shown, with the different components labeled (LPS, lipopolysaccharide; OM, outer membrane; IM, inner membrane; WTA, wall teichoic acid; LTA, lipoteichoic acid).

In this article, we will review known structures of tail fibers and tailspikes. Detailed knowledge of these receptor-binding proteins will inform us on their receptor-binding properties and allow us to better understand phage infection. It may also allow site-directed mutation to direct phages to different receptors, or the generation of chimeric receptor-binding proteins for the same purpose. Possible applications are detection and elimination of bacteria and incorporation into protein-based biotechnological devices.

Structures of Bacteriophage Tail Fibers

Although most podoviruses appear to have tailspikes, some have relatively long and thin fibrous appendages (Figs. 2 and 3). Coliphage T7, a well-studied example, has six kinked fibers. The T7 tail fibers are each formed by a parallel homotrimer of gp17. The kinked fibers are comprised of a rather large (when compared to other fibers) amino-terminal tail attachment domain, a thin

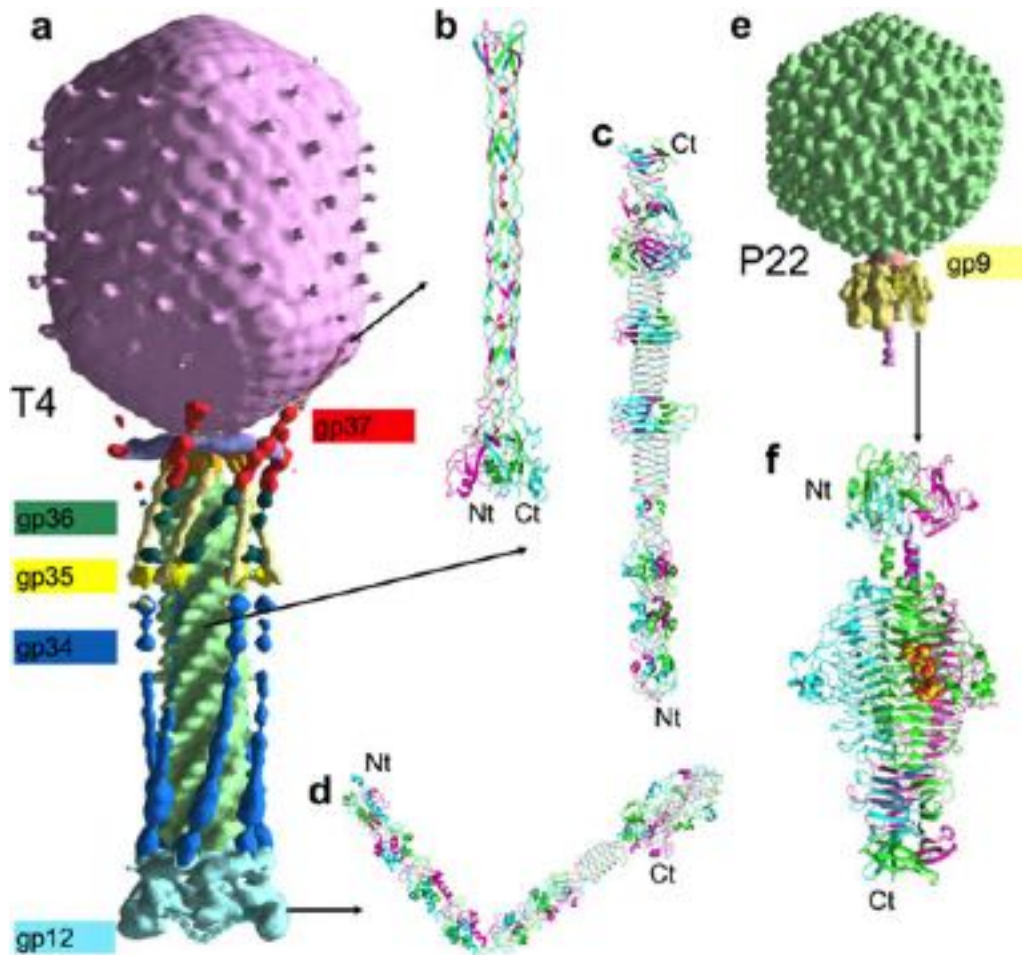


Fig. 2 Tailspikes and tail fibers. (a) Bacteriophage T4 (EMDB entry EMD-2774). Tail fiber proteins are shown in blue (proximal tail fiber protein gp34), yellow ("knee" protein gp35), dark green (distal tail fiber protein gp36) and red (distal tail fiber protein gp37). The structures of the C-terminal regions of gp34 and gp37 (ribbon diagrams) are shown in their approximate positions. The baseplate (light blue) contains the six short tail fibers, which are made up of gp12 (one is shown as a ribbon diagrams). The carboxy-terminal part of gp37 (b; PDB entry 2XGF), the carboxy-terminal part of gp34 (c; PDB entry 5NXH) and full-length gp12 (d; from PDB entry 5IV5) are shown as ribbon diagrams. (e) Bacteriophage P22 (EMDB entry EMD-1220). P22 tailspikes (gp9) are in yellow. A close-up of a gp9 trimer with each chain colored differently is also shown (f; PDB entry 2XC1). The *Salmonella* O-antigen receptor (ball representation in red and yellow; from PDB entry 1TYX) is shown bound to one of the chains. Amino- and carboxy-termini are indicated for each protein (Nt and Ct, respectively).

shaft, a flexible kink, and a carboxy-terminal domain that is apparently composed of four nodules. The structure of the distal three nodules is known (Fig. 3(c)). This structure revealed a pyramid domain and a globular carboxy-terminal domain. The trimeric pyramid domain is composed of three interacting beta-strands that become shorter towards the carboxy-terminus, yielding a pointed shape. Each beta-sheet is formed by one beta-strand from the first monomer, followed by five beta-strands from the next monomer and three strands from the third monomer. A short alpha-helical neck connects the pyramid domain to the carboxy-terminal globular domain, which is composed of three beta-sandwiches, one contributed by each of the three gp17 molecules. The gp17 beta-sandwich has a unique topology, although it looks a lot like the beta-sandwich head domain of adenovirus fibers.

The *Salmonella* podovirus epsilon15 has six fibers, each made up of a trimer of gp20. Like the T7 fibers, they have a slender amino-terminal part and a flexible kink, but they have a much thicker and somewhat longer carboxy-terminal part. The carboxy-terminal part consists of a tailspike-shape domain and ends in a flower-like structure with three petals.

The tail of myovirus bacteriophage T4 is the best-studied contractile system. The tail consists of a sheath, an internal tail tube and a baseplate which is situated at the distal end of the tail (Fig. 2(a)). Attached to the baseplate are the fibers. There are six long or side tail fibers which interact with the primary receptor and six short or central tail fibers, which interact with the secondary receptor. The long tail fibers of bacteriophage T4 are kinked structures composed of four different proteins: gp34, gp35, gp36 and gp37. The proximal half-fiber is formed by a parallel homo-trimer of gp34. The amino-terminal end of gp34 is attached to the baseplate, while the carboxy-terminal end interacts with the distal half-fiber, presumably with gp35 and/or gp36. A monomer of gp35 forms the "knee". The distal half-fiber is composed of trimers of gp36 and gp37. The gp36 protein subunit is located at the proximal end of the distal half-fiber, while gp37 makes up the distal part, including the receptor-recognizing tip. The short tail fibers are trimers of gp12.

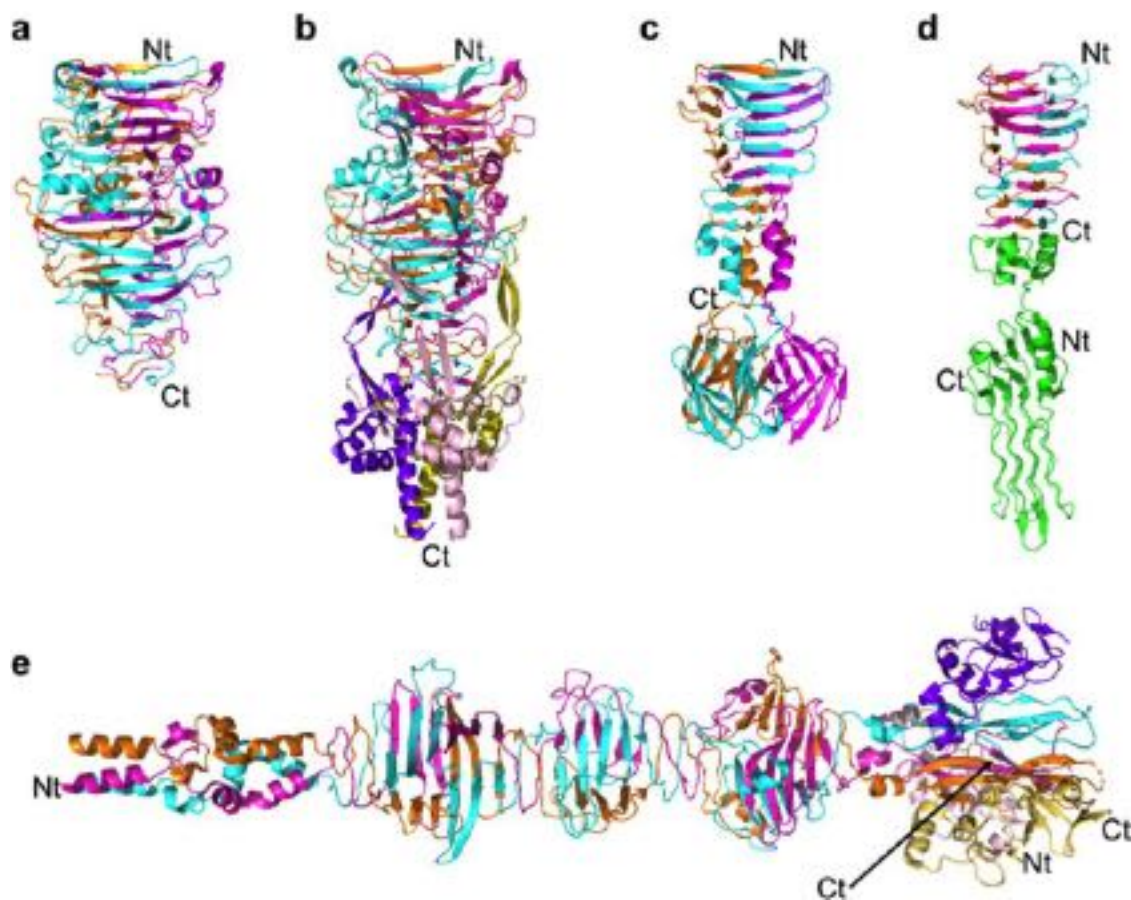


Fig. 3 Structures of tail fibers. Ribbon diagrams of the distal regions of the T5 fiber (a; PDB entry 4UW7), the T5 fiber with its intramolecular chaperone domain (b; PDB entry 4UW8), the T7 fiber (c; PDB entry: 4A0T), the phage S16 fiber consisting of three chains of gp37 and one chain of gp38 (d; PDB entry 6F45) and the Mu fiber with its chaperones (e; PDB entry 5YVQ). Fiber chains are colored magenta, cyan and orange and their chaperones in light pink, purple and dark yellow. Gp38 of phage S16 is colored green. Amino- and carboxy-termini are indicated for each protein (Nt and Ct, respectively).

For the correct folding of gp12, gp34 and gp37 a molecular chaperone, gp57, is necessary. It is a small protein of 79 residues that lacks aromatic amino acids, cysteines and prolines. In vitro, it adopts different oligomeric states. Another chaperone, gp38, must be present for the correct trimeric assembly of gp37. The molecular bases of the gp38 and gp57 chaperone activities are unclear, but it has been proposed that gp57 functions to keep fiber protein monomers from aggregating non-specifically before folding is completed, while gp38 may bring together the carboxy-terminal ends of gp37 monomers to start the folding process. In the related coliphages T2 and T6, in the *Salmonella* phage S16, and presumably many yet other undiscovered T4-like phages, gp38 plays another role. Instead of being just a chaperone, gp38 stays bound to the end of gp37 and serves as the *de facto* receptor-binding protein.

Structures are known for the carboxy-terminal region of gp34 ([Fig. 2\(c\)](#)) and the receptor-binding needle of gp37 ([Fig. 2\(b\)](#)). The carboxy-terminal part of gp34 is important for folding, which is generally true for viral fiber proteins. Crystal structures of carboxy-terminal parts of gp34 revealed three structural repeats containing three central alpha-helices preceded by intertwined loops. Seven more of these repeats are likely present in the unknown amino-terminal part of the structure. Carboxy-terminal to the repeat region, a short triple-helical region followed by two turns is present. The rest of the structure is a long triple beta-helix, interspersed with three tower domains. The tower domains are formed by three anti-parallel beta-sheets, one contributed by each of the three monomers. The first tower domain is decorated with loops, while the loops of the second tower domain are really short and mainly beta-turns. The last tower domain is extensively decorated, including several short alpha-helices and a long beta-hairpin extending towards the carboxy-terminus.

The carboxy-terminal part of gp37 revealed an elaborately interwoven trimeric structure formed by a globular collar domain (about 45 Å wide, at the bottom in [Fig. 2\(b\)](#)), an elongated needle domain (around 15 Å wide) and a small head domain with a width of around 22 Å (at the top of [Fig. 2\(b\)](#)). Each of the three chains runs from the collar domain to the end of the tip and twists around a neighboring chain before turning back, with both the amino- and carboxy-terminus located near the bottom of the collar domain. In the collar domain, each monomer contains a small beta-sandwich of anti-parallel sheets with an alpha-helix inserted into a loop of the outside beta-sheet. The three inner sheets provide the trimer interactions. Next to the collar domain is a small intertwined region. The carboxy-terminus of each gp37 is folded back into the collar domain, making the needle domain an

insertion into the collar domain, projecting outwards. The needle domain is a 150 Å long six-stranded antiparallel right-handed twisted circular sheet, with each of the chains completing one-and-a-half turns (about 540°) around the fiber axis. In the core of the needle domain, hydrophobic and hydrophilic regions alternate, with the latter forming metal-binding sites. Seven iron ions are coordinated octahedrally by two histidine residues from each chain (six histidines per iron ion). Twenty-six residues from each of the three chains form a compact, interwoven head domain of 22 Å in diameter and 18 Å high. The first fourteen amino acids loop around a neighboring chain; the chain then threads through the loop of a neighbor, before turning back into the needle domain. The head domain, being located at the extreme distal end of the long tail fiber, is likely to play a primary role in receptor binding.

The short tail fibers of bacteriophage T4 are parallel homotrimers of gp12. Morphologically, they consist of an amino-terminal virus-binding region, a long thin shaft domain and a more globular carboxy-terminal receptor-binding domain (Fig. 2(d)). The amino-terminal part of the short tail fiber is made up of a base-plate binding region or domain of unknown structure, followed by several of the repeats also present in gp34 (described above). The repeats are followed by a triple beta-helix domain, which is connected to a collar domain by a short alpha-helical neck. From the collar domain, the globular receptor-binding domain protrudes, just like the needle domain protrudes from the gp37 collar domain. The topology of the carboxy-terminal globular part has a complicated "knitted" fold, in which the three protein chains interact extensively with each other. A metal ion was found to be buried deep inside the carboxy-terminal domain. It is positioned on the threefold axis of the protein and is coordinated by the side chains of two histidines from each of the three chains, like the seven iron ions in the long tail fiber.

The distal carboxy-terminal end of the trimeric *Salmonella* phage S16 gp37 protein forms a triangular beta-prism domain and ends in a short triple beta-helix (Fig. 3(d)). The monomeric gp38 adhesin has a modular design. Its amino-terminal attachment domain binds to the tip of gp37 via a hydrophobic platform, whereas the carboxy-terminal specificity domain determines affinity for the host cell receptor. The small amino-terminal gp37 attachment domain forms a helical bundle of three short alpha-helices, from which three tryptophan residues intercalate into three sites in the gp37 platform. Around the hydrophobic platform, a ring of three-fold symmetric hydrogen bonds completes the stable interface.

The specificity domain contains single beta-helix, from which loops project in the direction away from the phage (Fig. 3(d)). Two of these loops are long and contain glycine-rich motifs. There are ten of these glycine-rich motifs, two in one loop and the remaining eight in the other. These motifs form poly-glycine type II helices that are folded into a packed lattice, forming a PGII-sandwich. Sequence variability between related phages is limited to surface-exposed helices and distal loops that form putative receptor-binding sites.

The structure of the bacteriophage mu fiber (gpS) revealed an alpha-helical shaft, a central domain containing beta-sheet domains (called "tower" domains) and triple beta-helix domain, plus a carboxy-terminal beta-strand domain (Fig. 3(e)). The carboxy-terminal domain is relatively thin, elongated and has distal protruding loops which are postulated to form the receptor-binding site. The amino-terminal phage-binding domain was missing from the structure. Interestingly, upon co-expression, the mu fiber chaperone gpU remained bound to the carboxy-terminal domain of gpS and part of its structure could be resolved. GpU contains two domains, an amino-terminal domain that binds to gpS and a carboxy-terminal domain (only partly visible in the structure) that is presumed to be responsible for chaperone oligomerization; this organization is proposed to be general to gpU-like chaperones. Phage particles with and without gpU are both infectious and both the fiber and the chaperone appear to specifically bind LPS.

The structures of the side tail fibers of other "simple" myoviruses like phage P2 are likely similar to the fiber of phage Mu. This is based on the known structures of the fibers from pyocins R1 and R2, which in turn show sequence homology with the fiber of P2. Both the R1 and R2 fibers contain tower domains, thin shaft domains and a carboxy-terminal lectin-like domain. The shaft domains are composed helix-plus-turn motifs, which intertwine with their counterparts in the trimer. The carboxy-terminal lectin-like domain may be involved in receptor-binding.

Siphoviruses lambda and T5, which infect the Gram-negative bacterium *Escherichia coli*, possess fibers at the bottom and sides of their tail tips. The side tail fiber proteins are involved in recognition of cell surface molecules, while the central fiber proteins play an active role in membrane penetration. No high-resolution structure of a siphovirus central tail fiber has been determined, but structural information on some siphovirus side tail fibers is available. Interestingly, the carboxy-terminal part of the lambda side tail fiber has high sequence identity to the carboxy-terminal part of phage T4 gp37, and thus must also form a collar and needle domain. An extra metal-binding motif suggests that the needle domain of the lambda side tail fiber contains eight instead of seven iron ions. The receptor-binding head domain is not homologous and has probably mutated to bind a different receptor. Both the lambda and T5 side tail fibers need a chaperone for folding, but while the lambda chaperone is a separate protein, the T5 chaperone is intramolecular (see below). Interestingly, the lambda side tail fiber chaperone can functionally substitute for the T4 chaperone required for gp37 assembly, gp38.

Bacteriophage T5 has a flexible, three-fold symmetric tail, to which three L-shaped side tail fibers are attached. These fibers recognize oligo-mannose units on the bacterial cell surface prior to infection and are composed of homotrimers of the pb1 protein. Pb1 has 1396 amino acids, of which the carboxy-terminal 133 residues form a trimeric intra-molecular chaperone that is auto-proteolyzed after correct folding. The structure of a trimer of the preceding residues 970–1263 resembles a bullet, with the walls formed by partially intertwined beta-sheets, conferring stability to the structure (Fig. 3(a)). The topology of the structure is unique to pb1 and the protein inhibits T5 infection by competition. A site-directed mutant (Ser1264 to Ala), in which auto-proteolysis is impeded, revealed the structure of the additional chaperone domain (residues 1263–1396; Fig. 3(b)). It consists of a central trimeric alpha-helical coiled-coil flanked by a mixed alpha-beta domain. Three long beta-hairpin tentacles, one from each chaperone monomer, extend into long curved grooves of the bullet-shaped domain.

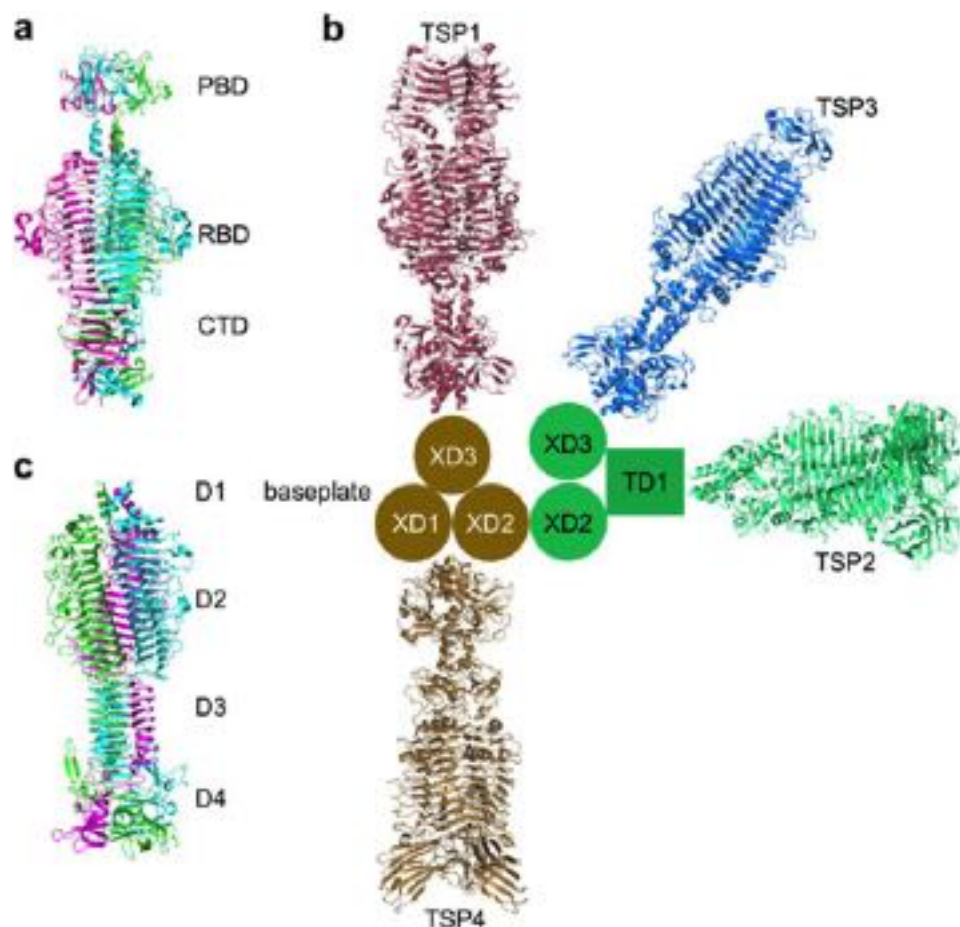


Fig. 4 Structures of beta-helical tailspikes. (a) *Salmonella* phage P22 tailspike (PDB entry 2XC1). The amino-terminal phage-binding domain (PBD), the central beta-helical receptor-binding domain (RBD) and the carboxy-terminal interdigitated domain (CTD) are indicated. The three chains are colored green, magenta and cyan. (b) The host recognition apparatus of coliphage CB120, consisting of tailspikes 1 (purple, PDB entry 4OJ5), 2 (green, PDB entry 5W6P), 3 (blue, PDB entry 5W6F) and 4 (brown, PDB entry 5W6H). Domains for which high resolution structures are unknown, but that link the four tailspikes to each other are shown as circles or squares. (c) *Bacillus* phage Phi29 appendix (PDB entry 3SUC). The locations of the linker (D1) that links to the amino-terminal phage-binding domain, the beta-helical domains D2 and D3, and the intramolecular chaperone domain D4 are indicated. Coloring is as in part (a).

High-resolution structures of fibers from bacteriophages that infect Gram-positive bacteria are not yet known. *Bacillus* bacteriophages PBP1 and AR9 have long helical fibers with which they ensnare bacterial flagella. SPO1-like phages (e.g., A511) have fibers that fold back under the baseplate, similarly to the way T4 gp12 does, but whether there are structural similarities remains to be seen.

Structures of Bacteriophage Tailspikes and Other Receptor-Binding Proteins

The first high-resolution tailspike structure to be solved was that of *Salmonella* phage P22 (Fig. 4(a)). In each tailspike, a small trimeric beta-structured amino-terminal domain connects with the rest of the virus. These virus-binding domains are connected by a short alpha-helical coiled-coil linker to a central domain, allowing for some flexibility. The central domain is a trimer of parallel left-handed beta-helices, with each helix containing thirteen complete turns. This beta-helix domain is responsible for carbohydrate binding and cleavage (endorhamnosidase activity). The carboxy-terminal domain of the P22 tailspike consists of interdigitated beta-strands and has been shown to be important for correct protein trimerisation and folding.

Many of the lessons learnt from the P22 tailspike structure have turned out to be general. The parallel beta-helix fold present in other tailspike proteins is the most obvious example of this. This fold appears to be especially suitable for carbohydrate cleavage reactions. A monomeric beta-helix fold was first discovered in the enzyme pectate lyase; bacteriophage tailspikes have adopted the fold for efficient recognition and cleavage of host cell wall polysaccharides. Another general lesson is that folding of tailspikes (and of tail fibers and most other viral receptor-binding proteins) starts at the carboxy-terminus. The carboxy-termini of three monomers come together, sometimes with the help of a chaperone, and folding starts. It might appear to be counterproductive to wait until the protein chain is synthesized completely before folding starts. However, for proteins containing repetitive structural motifs, folding the carboxy-terminal domains first makes out-of-register interactions, which might lead to unproductive misfolding, less likely.

Tailspike structures have also been determined for other phages. The *Salmonella* myovirus Det7 and siphovirus 9NA both contain a tailspike with high sequence identity to the *Salmonella* podovirus P22 tailspike (except for the phage-binding amino-terminal domain). Not surprisingly, the three structures turned out to be very similar, containing the same beta-helix and carboxy-terminal domains.

The *Shigella flexneri* podovirus Sf6 also has a tailspike with a beta-helix domain comprised of 13 complete turns, although its sequence has no homology to the beta-helix of the P22 tailspike. The Sf6 tailspike beta-helices are wound around each other with a slight left-handed twist. At the distal carboxy-terminal end of the tailspike, each monomer forms a beta-sandwich domain. A similar carboxy-terminal beta-sandwich domain is also observed in the *Escherichia coli* podovirus HK620, although, with the domain rotated by about 180° around the trimer axis.

The tailspike of *Pseudomonas* podovirus LKA1 has a beta-barrel insertion domain in its beta-helix, which contributes to the active site. Here, the carboxy-terminal domain is a longitudinal beta-sandwich. A similar longitudinal beta-sandwich is also present in a second tailspike of the aforementioned phage Det7, which in fact has a total of four different tailspikes. In this second Det7 tailspike the beta-sandwich domain has homology to a lectin domain, so apart from a putative role in trimerisation and folding, this carboxy-terminal domain may also help the tailspike bind to the LPS O-antigen.

The myovirus CBA120 also has four tailspikes, attached to the side of the baseplate (Fig. 4(b)). The four tailspikes allow binding to and infection of different hosts. Tailspike 1 recognizes *Salmonella enterica* serovar Minnesota, tailspike 2 *Escherichia coli* O157, tailspike 3 *Escherichia coli* O77 and tailspike 4 *Escherichia coli* O78 strains. All four tailspikes have a central beta-helix domain, followed by a bend and another beta-helix domain at the carboxy-terminus (although the carboxy-terminal beta-helix of tailspike 3 is very small). The central beta-helix domains are involved in receptor-binding, while the carboxy-terminal beta-helix domains may be responsible for correct folding. Tailspike 4 has five small amino-terminal domains with which it attaches to the baseplate, to tailspike 2 and to tailspike 1 (Fig. 4(b)). The amino terminal domain of tailspike 3, in turn, attaches to an amino-terminal domain of tailspike 2.

Examples of podoviruses with multiple tailspikes are *Salmonella* phage SP6, *Escherichia coli* phage G7C and *Klebsiella* phage KP32. Phage SP6 has two tailspikes, recognizing the O-antigens of *Salmonella enterica* serovars Typhimurium and Newport, respectively. The two tailspikes are attached to the phage via an adapter protein. This adapter protein can rotate, locating the correct tailspike into the correct orientation for host recognition and infection.

In the case of coliphage G7C, one tailspike, gp66, binds to the phage, and the second tailspike, gp63, binds to an amino-terminal region of gp66. Gp63 has six domains, of which the first two resemble those of the phage CBA120 tailspike 1. These domains have thus evolved to organize the relative organization of tailspikes in myovirus baseplates and in podovirus tails. Domain 3 is a connector domain, also seen in tailspikes with endosialidase activity (see below). Domain 4 is an esterase domain, and it has been shown that the esterase activity is necessary for infection. Domains 5 and 6 are homologous to non-catalytic carbohydrate binding domains. This is an example of a tailspike without a beta-helix fold.

Klebsiella phage KP32 has a similar organization to coliphage G7C, with one tailspike, gp37, binding to the phage, and the other, gp38, binding to an amino-terminal domain of gp37. The crystal structure of gp38 is known. A flexible amino-terminal region is followed by a catalytic right-handed parallel beta-helix domain, and two non-catalytic carboxy-terminal carbohydrate binding domains.

Each monomer of the trimeric tailspike of the *Acinetobacter baumannii* phage PhiAB6 has a small amino-terminal beta-structured phage-binding domain. This domain is followed by a short alpha-helix and a beta-helical exopolysaccharide cleavage domain. The carboxy-terminal part is formed by a triangular beta-prism, an interdigitated region, another triangular beta-prism and a carboxy-terminal beta-sandwich. Structures of other *Acinetobacter* phage tailspikes have also been solved.

The appendices of the Gram-positive *Bacillus subtilis* phage Phi29 (Fig. 4(c)) also contain parallel left-handed beta-helices. Each appendix has an amino-terminal domain D1 of unknown structure, which is bound to the phage neck. A short triple coiled-coil domain connects the amino-terminal domain to a barrel-shaped domain D2 composed of three parallel left-handed beta-helices. The beta-helices in this domain contain around 24 residues per turn and, like in the P22 tailspike, have inserts of variable lengths that form loops. This domain looks similar to the central domain of P22 tailspike and may thus be involved in receptor-binding. Supporting this is a bound buffer molecule in one of the crystal structures (PDB entry 3GQ8). Carboxy-terminal to this, another thinner and much more regular beta-helix domain D3, with turns that contain only around fifteen residues (without significant inserts), is found. Between the two beta-helical domains, a cyclical swap of the polypeptide chains takes place.

To infect *Escherichia coli* K1, which is surrounded by a polysialic-acid capsule, bacteriophage K1F is equipped with receptor-binding proteins that have an endosialidase activity. These proteins are also trimers but lack the typical beta-helices of conventional tailspikes. Instead, the receptor-binding proteins of phage K1F have an amino-terminal phage-binding domain of largely unknown structure, central beta-propeller and beta-barrel domains and carboxy-terminal beta-prism and triple beta-helix domains (Fig. 5(a)). Endosialidases depend on alpha2,8-glycosidically linked sialic acid oligomers. The beta-barrel domain is where the enzymatic polysialic-acid cleavage reaction takes place.

Interestingly, the phage K1F tailspikes and phage Phi29 appendices have intramolecular chaperones like that of the phage T5 fiber (and were actually structurally analyzed in these phages first). The central parts of chaperones are structurally homologous to each other, although the chaperone from Phi29 (D4 in Fig. 4(c)) is smaller and lacks the carboxy-terminal triple coiled-coil present in the intramolecular chaperones of the K1F tailspike and the T5 fiber.

The well-studied lactococcal phages p2, TP901-1 and Tuc2009 all have large baseplate assemblies, containing arrays of six or 18 trimeric globular receptor-binding proteins. These receptor complexes can bind various saccharidic moieties and likely interact with teichoic or lipoteichoic acid on the cell surface. Although these receptor-binding proteins are different in structure from tail fibers, they still share a trimeric structure and also possess a short region of beta-helix as is seen in tailspike proteins. Interestingly,

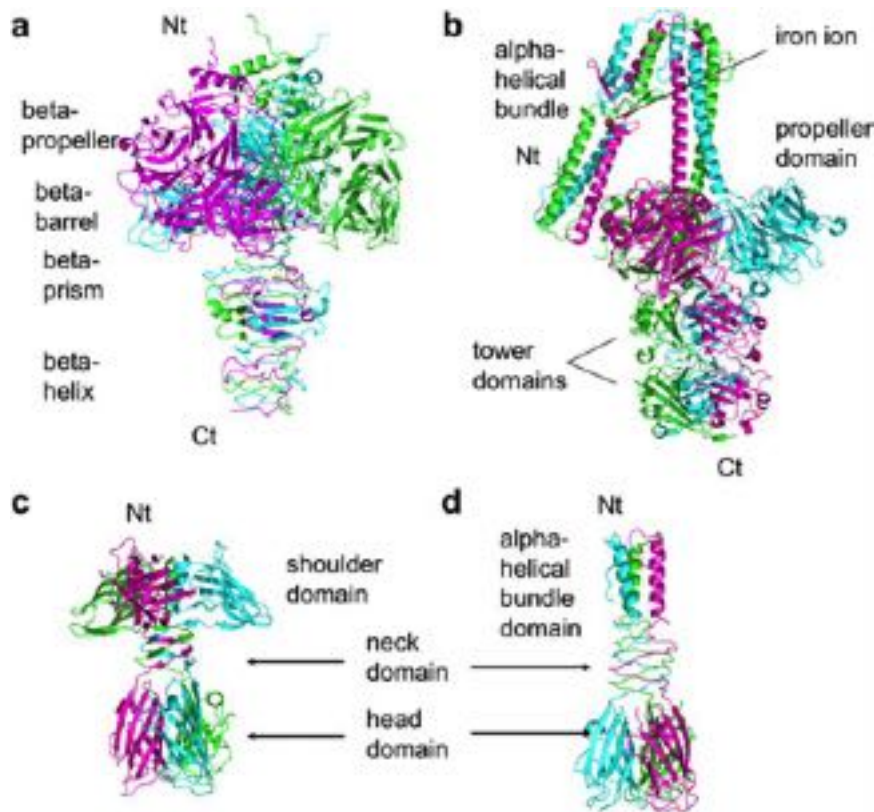


Fig. 5 Atypical tailspikes and other receptor-binding proteins. (a) Coliphage K1F endosialidase (PDB entry 3JU4). (b) Receptor-binding protein from staphylococcal phage phi11 (PDB entry 5EFV). (c) Receptor-binding protein of lactococcal phage p2 (PDB entry 1ZRU). (d) Receptor-binding protein of lactococcal phage TP901-1 (PDB entry 3EJC). For each panel, the three protein chains are colored green, magenta and cyan. Domains, amino-termini (Nt) and carboxy-termini (Ct) are indicated, where appropriate.

phage p2 appears to bind to the cell surface using only its receptor-binding proteins, while phage TP901-1 possesses both receptor-binding proteins and a central fiber protein, which may also be involved in cell surface recognition and/or cell-wall degradation. The receptor-binding proteins achieve high-affinity binding through avidity resulting from the large number of receptor-binding proteins displayed on the baseplate.

The receptor-binding protein of lactococcal phage p2 has an amino-terminal shoulder domain, a central neck domain and a globular head domain (Fig. 5(c)). The shoulder domain has a beta-sandwich fold and serves to anchor the protein to the baseplate. The neck domain is a triple-stranded beta-helix, with the three protein chains of the trimer winding around each other. Finally, each monomer of the head domain forms a seven-stranded beta-barrel with the same topology as that of adenovirus and reovirus fiber head domain (and different from the coliphage T7 fiber head domain). The receptor-binding protein of the lactococcal phage TP901-1 has the same head and neck domains as the receptor-binding protein of phage p2, but the shoulder domains are replaced by a parallel bundle of three helices (Fig. 5(d)).

Two staphylococcal phages, phi11 and P68, have been shown to have receptor-binding proteins with very similar structures. This is likely to be another example of horizontal gene transfer between different phage types. In the siphovirus phi11, the protein is folded away into the baseplate, while in the podovirus P68, the protein extends away from the phage. Both structures consist of a long, kinked, three-membered parallel alpha-helical bundle. This stalk domain is followed by a five-bladed propeller domain, which may harbor the receptor-binding site. At the carboxy-terminal end, two tower domains are present, which probably have a role in trimerization and structural stability. Appreciable differences between the two structures are the hinge, which in the phi11 receptor-binding protein (Fig. 5(b)) is more acutely folded, and the presence of an octahedrally coordinated iron ion in the phi11 receptor-binding protein, but which is absent in the P68 receptor-binding protein.

Host Cell Recognition

Cell-surface recognition by bacteriophages is a two-step process. While the two steps of interaction with the cell surface may involve only one type of surface molecule, many phages interact with more than one receptor. The phage T5 tail first interacts non-specifically with the O-antigen of *E. coli* LPS through its L-shaped tail fibers (pb1) and then makes strong specific interactions with

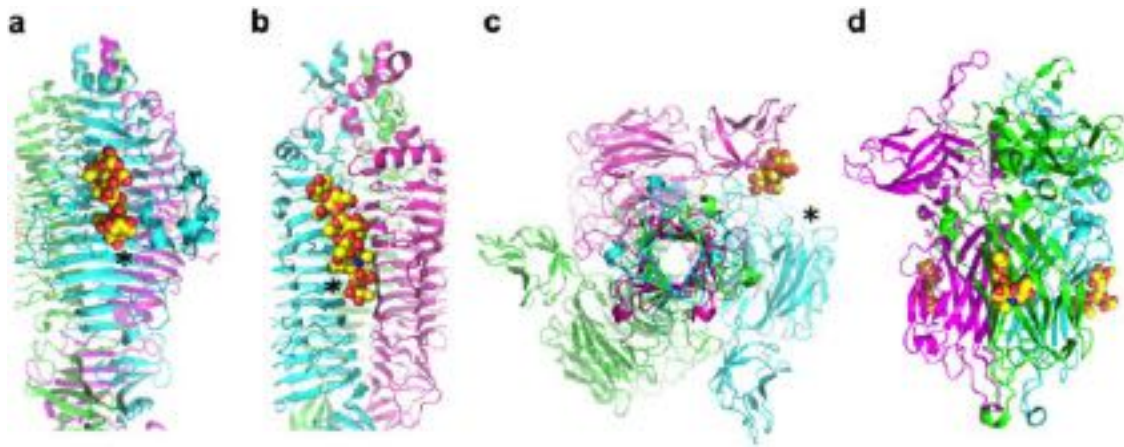


Fig. 6 Structures of bacteriophage receptor-binding proteins in complex with bacterial receptors. (a) *Salmonella* phage P22 tailspike complexed with O-antigen octasaccharide (PDB entry 3TH0). (b) Tailspike protein of Sf6 bacteriophage bound to *Shigella flexneri* O-antigen octasaccharide fragment (PDB entry 4URR). (c) Endosialidase of bacteriophage K1F in complex with oligomeric α -2,8-sialic acid (PDB entry 1V0F). (d) *Lactobacillus* phage 1358 in complex with a bacterial cell wall trisaccharide (PDB entry 4RGA). Approximate locations of active sites for polysaccharide cleavage are shown with asterisks where appropriate. In panels (a), (b) and (d), the receptor-binding proteins are viewed from the side, with the carboxy-terminal ends at the bottom of the figure; in panel (c) the protein is viewed from the bottom, i.e., down from the carboxy-terminal end towards the amino-terminal end.

cell-surface receptor, FhuA, through the central tail fiber, pb5. In an analogous manner, bacteriophage SPP1 binds reversibly to cell-wall teichoic acids and irreversibly to a specific cell surface receptor, YueB.

The tailspikes or fibers mediate initial, reversible, contact of the virion with the target cell, but in general do not trigger DNA release from the virion. The role of primary receptor binding is to speed up infection, facilitating a two-dimensional search on the bacterial surface, rather than a three-dimensional search of a large volume. Once the phage is in the correct orientation and/or a secondary receptor is encountered, irreversible binding takes place. At this point, the phage is committed to DNA transfer into the host. The phage, as a microbiological particle, ceases to exist, but its genetic material will direct the genesis of multiple daughter phages - if the bacterium is a suitable host and does not successfully defend itself by a post-infection mechanism (such as a restriction enzyme or CRISPR-Cas system).

Most of the fibers of podovirus T7 normally point upwards, interacting with the phage head. Occasionally, a detached fiber may bind to the host cell LPS. When this happens, the phage may "walk" over the host cell surface, alternately detaching fibers that recognize the host LPS. When all six fibers are detached and bound to the bacterium, the phage is in the correct orientation for infection, with the short conical tail pointing towards and contacting the bacterial outer membrane. Whether there is a specific secondary receptor interaction between the tail and the membrane or a membrane protein is currently unknown. Core proteins then pass through the tail, forming a protective tube through the periplasm for the phage genome to pass through.

Podovirus P22 tailspikes bind glycans with long grooves along its length. Each of the three monomers has a long groove that binds and cleaves the *Salmonella* O-antigen receptor through endo-rhamnosidase activity (Fig. 6(a)). Other tailspikes, such as that of *Shigella* phage Sf6, have the binding site (and active site) in grooves between monomers, with contributing active site residues from both neighboring monomers for each of the three sites in the trimer (Fig. 6(b)). The O-antigen degradation activity allows the phage to get near the bacterial outer membrane and position itself in the right orientation for infection (with the tail tube pointing towards the cell wall). Apart from these two examples, additional structures for tailspikes in complex with their oligosaccharide receptors have been reported. Furthermore, tailspikes often have additional carbohydrate binding domains, these may help increase the affinity for the bacterial cell wall.

Where bacteria have developed different cell wall structures, phages have adapted to bind and, where necessary, digest these. The endosialidase of bacteriophage K1F is an example of this. Poly- α -2,8-sialic acid binds to the beta-barrel domains of the endosialidase and to the triple beta-helix domain (Fig. 6(c)). Only the former site is thought to be involved in polysaccharide cleavage, with the active site on the surface of the beta-propeller domain of a neighboring monomer (Fig. 6(c)). With regards to the enzymatic reactions, endo-depolymerases appear to be preferred over exo-depolymerases, which prevents bacteria to inhibit their activity by "capping" the polysaccharides.

The siphoviruses lambda and T5 recognize host LPS with their long, L-shaped, tail fibers. The exact receptor for the coliphage lambda long tail fibers has not been determined, but the coliphage T5 long tail fibers binds to α (1,2)-linked oligo-mannose of the O-antigen of O8- and O9-type *Escherichia coli*. Interestingly, it appears that before cleavage of the intramolecular chaperone, the L-shaped fibers of phage T5 do not bind their receptors, suggesting that the long beta-fingers cover up the oligo-mannose binding site. However, it has to be kept in mind that both lambda and T5 can infect perfectly well in the laboratory without their long tail fibers, utilizing only the interaction of the central tail fiber with their secondary receptor proteins LamB and FhuA, respectively, even though the rate of irreversible adsorption of the phage to its host cell is substantially reduced.

Many bacteriophages that infect Gram-positive bacteria are thought to recognize teichoic acid moieties, but, unfortunately, no structures of receptor-binding proteins with teichoic acid analogs have yet been determined. Some Gram-positive bacteria form cell wall structures that cover the teichoic acid layer. An example is *Lactobacillus lactis* strain SMQ388, which forms a polysaccharide cell wall "pellicle" made up of specifically linked N-acetyl-glucosamine and galactose and glucose units. The siphovirus 1358 can recognize this polysaccharide and the structure of the receptor-binding protein in complex with a trisaccharide analog has been determined (Fig. 6(d)). This protein has a structure related to that of the receptor-binding protein of *Lactobacillus* phage p2, with a beta-sandwich shoulder domain and a larger head domain, lacking the triple beta-helical neck.

In the case of the myovirus T4, in adverse conditions for phage multiplication, the long tail fibers are in a retracted conformation, lying against the tail sheath and head of the bacteriophage. In the extended conformation, only the proximal end of the fiber is attached to the baseplate. Primary host cell interaction is mediated by the six long tail fibers. As gp37 is known to interact with the glucosyl- α -1,3-glucose terminus of LPS and protein-saccharide interactions almost always involve stacking of sugar residues onto aromatic amino acid side chains, aromatic surface amino acids are attractive candidates for receptor binding. Basic residues like lysine and arginine may also interact with the LPS phosphate groups. Gp37 may also bind reversibly to the outer membrane protein OmpC. If a certain number of long tail fibers are bound to the host, estimated to be three or four, they initiate a conformational change of the baseplate. In the uncontracted phage, the short tail fibers form a garland around the baseplate, in which the amino-terminal domain, a region forming a kink in the shaft domain and the receptor-binding domain are bound the baseplate. During infection, the conformation of the baseplate changes and only the amino-terminal domain remains bound to the baseplate, releasing the rest of the protein and allowing the receptor-binding domain to strongly bind to the LPS oligosaccharide core. When six of the short tail fibers bind to LPS, the phage becomes effectively bound irreversibly to the host, ready for the extensive conformational changes necessary for DNA transfer. The conformational change of the baseplate triggers contraction of the tail sheath that drives the tail tube through the cell membrane, digesting the peptide glycan layer by the tail lysozyme, and ejection of the phage DNA through the tail tube channel into the bacterial cytoplasm.

Phages are not restricted to having only one type of receptor-binding protein, and several mechanisms to increase the number of hosts that can be infected are known. Phages can carry two or more types of tailspikes that have different target specificities, either of which can function in primary receptor binding. For example, coliphage K1-5 is a T7-like phage that has two different tailspikes in its virion, binding α -polysialic acid (K1) or poly-glucuronic acid-N-acetylglucosamine (K5) capsular polysaccharides. The *Bordetella* phage BPP-1 has a reverse transcriptase mechanism with which it creates sequence variability in the receptor-binding site of its fiber. This variability helps the phage keep up with phase variation of the host, allowing it to evolve fibers that recognize the changed host cell wall. The modular nature of tailspikes and tail fibers favors these gene duplication, domain exchange and mutational events, because changes in the receptor-binding domains do not affect the interaction of the spikes and fibers with the rest of the virion.

Some bacteriophages alter the carboxy-terminal regions of tail fiber proteins by DNA inversion, changing their host range. An example is bacteriophage Mu, where the invertible gene segment is called G, and the corresponding site-specific DNA invertase enzyme is Gin (the gene for Gin maps outside the G-segment). The G-segment starts inside the long tail fiber gene (S) and also contains the tail fiber adapter gene (U) and, in inverted form, the genes U' and S'. U' codes for an alternative tail fiber adapter and S' for an alternative carboxy-terminal end of the long tail fiber. The orientation of the G segment thus determines which carboxy-terminal end of the fiber is expressed and which bacterial hosts can be infected. Mu G(+) phage particles expressing S and U are only infectious for *Escherichia coli*, while Mu G(-) phage particles expressing S' and U' have a wider host range and can also infect *Citrobacter freundii*, *Enterobacter cloacae*, and *Serratia marcescens*.

Conclusion and Perspectives

The increasing occurrence of pathogenic antibiotic-resistant bacteria calls for alternative approaches to combat pathogenic bacteria. Using bacteriophages and their proteins are some of a number of promising strategies. Phages or their endolysin proteins may be used to directly kill, while phage tailspikes and fibers can be used to specifically detect pathogenic bacteria. Enzymatic tailspikes may also be used to weaken their cell walls.

Atomic structures of phage tailspikes and fibers have yielded new protein folds and insights into their stability, their interaction with bacterial receptors and O-antigen degradation mechanisms. These insights have been obtained mainly by X-ray crystallography. The rapid development of cryo-electron microscopy is allowing the structure determination of entire phage tail complexes, although for detailed receptor interactions X-ray crystallography and NMR spectroscopy still have a major role to play.

In general, the genes of receptor-binding proteins are those that show most sequence variability between related bacteriophages. This is presumably the result of adaptation to mutations in the receptor of the phage host, of adaptation to a new receptor, or of adaptation to a new bacterial host strain or species. The fact that receptor-binding proteins protrude from the phage capsid allows for these mutations not to affect phage structure, even if whole domains are replaced. Phage-binding domains of receptor-binding proteins of closely related phages, which do interact with the rest of the virus, usually show more sequence similarity than the receptor-binding domain. In some cases, different phages have very similar receptor-binding proteins. Examples are *Salmonella* phages P22, 9NA and Det7 and the carboxy-terminal domains of the lambda and T4 fibers mentioned before. Where these phages infect the same or a very similar host, horizontal gene transfer between co-infecting phages or an infecting phage and a prophage in the genome is a likely mechanism. For many fibers and tailspikes no sequence similarity is detectable between domains of their receptor-binding

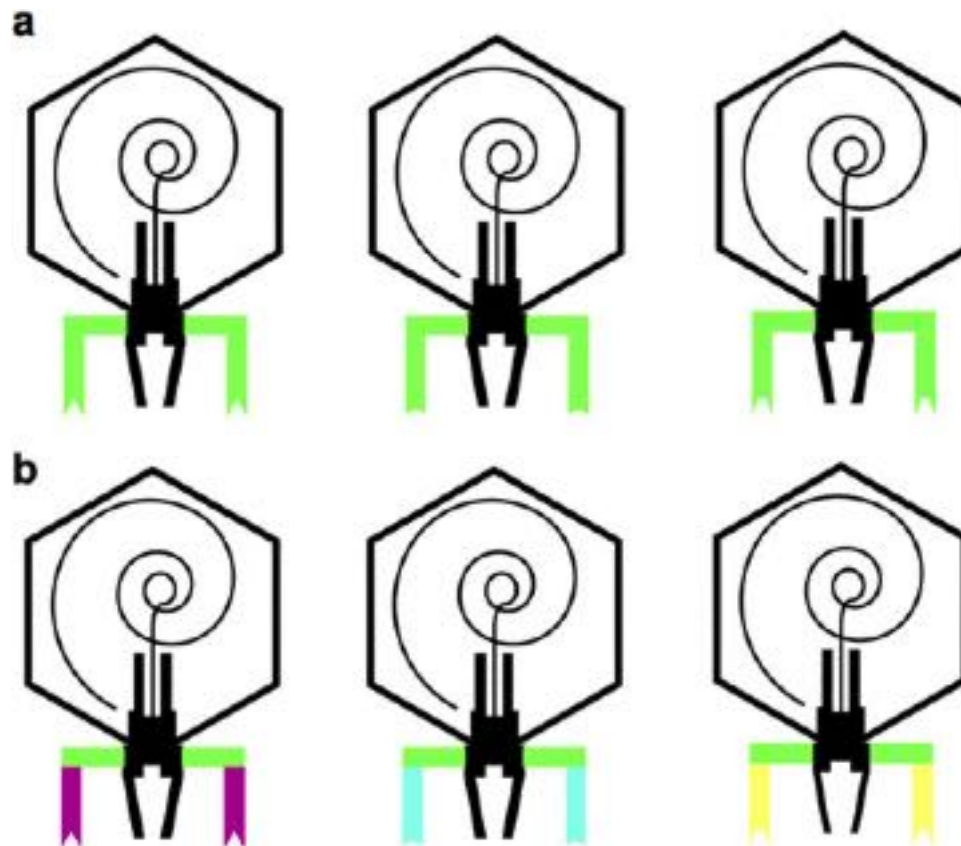


Fig. 7 Schematic drawing of site-directed mutated and chimeric phages. In the top row (a), three phages are shown with site-directed mutations in their tailspikes. These mutations modify the shape of the binding site enabling it to fit different bacterial receptors. In the bottom row, three phages with chimeric tailspikes are shown (b). The carboxy-terminal domains are from other phages with different host ranges.

proteins, but the folds are nevertheless the same, even for phages that infect very different hosts. An example is the tower domain in the fiber protein gp34 of coliphage T4 and in the receptor-binding proteins of phages infecting *Staphylococcus*. These similarities may be the result of convergent evolution of different phages finding the same structural solution independently, or divergent evolution of a phage from which one of the branches adapted to a very different host.

Given the large amounts of phages present in Nature and the rapid evolutionary race with their host bacteria, it is likely more new folds will be discovered for their receptor-binding proteins. Knowledge of these folds will allow the structure, and in favorable cases, receptor-binding properties, of newly discovered genes to be predicted. Experimentally determined and reliably predicted structures will also allow the tailoring of phage tailspikes and fibers to new hosts (Fig. 7). This may be done by site-directed mutation, as has been done for the tail fiber of coliphage T3. Replacing entire domains, i.e., generating chimeric receptor-binding proteins, is another exciting possibility. It is possible that in a relatively near future, a battery of phage receptor-binding proteins, plus a plethora of natural and synthetic bacteriophages, will be available for detection and elimination of most bacterial pathogens.

Acknowledgments

The authors thank the Spanish Ministry of Science, Innovation and Universities for financial support (grant number BFU2017-87022-P; MCIU/AEI/FEDER, EU and Severo Ochoa program SEV 2017-0712).

Further Reading

- Betts, S., King, J., 1999. There's a right way and a wrong way: In vivo and in vitro folding, misfolding and subunit assembly of the P22 tailspike. *Structure* 7 (6), R131–R139.
- Casjens, S.R., Molineux, I.J., 2012. Short noncontractile tail machines: Adsorption and DNA delivery by podoviruses. *Advances in Experimental Medicine and Biology* 726, 143–179.
- Dams, D., Brøndsted, L., Drulis-Kawa, Z., Briers, Y., 2019. Engineering of receptor-binding proteins in bacteriophages and phage tail-like bacteriocins. *Biochemical Society Transactions* 47 (1), 449–460.

- Davidson, A.R., Cardarelli, L., Pell, L.G., Radford, D.R., Maxwell, K.L., 2012. Long noncontractile tail machines of bacteriophages. *Advances in Experimental Medicine and Biology* 726, 115–142.
- Dunne, M., Ruf, B., Tala, M., *et al.*, 2019. Reprogramming bacteriophage host range through structure-guided design of chimeric receptor binding proteins. *Cell Reports* 29 (5), 1336–1350.
- García-Doval, C., van Raaij, M.J., 2013. Bacteriophage receptor recognition and nucleic acid transfer. *Subcellular Biochemistry* 68, 489–518.
- Goulet, A., Spinelli, S., Mahony, J., Cambillau, C., 2020. Conserved and diverse traits of adhesion devices from *Siphoviridae* recognizing proteinaceous or saccharidic receptors. *Viruses* 12 (5), 512.
- Hyman, P., van Raaij, M., 2018. Bacteriophage T4 long tail fiber domains. *Biophysical Reviews* 10 (2), 463–471.
- Leiman, P.G., Arisaka, F., van Raaij, M.J., *et al.*, 2010. Morphogenesis of the T4 tail and tail fibers. *Virology Journal* 7, 355.
- Leiman, P.G., Shneider, M.M., 2012. Contractile tail machines of bacteriophages. *Advances in Experimental Medicine and Biology* 726, 93–114.
- Mitraki, A., Papanikolopoulou, K., van Raaij, M.J., 2006. Natural triple beta-stranded fibrous folds. *Advances in Protein Chemistry* 73, 97–124.
- Nobrega, F.L., Vlot, M., de Jonge, P.A., *et al.*, 2018. Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology* 16 (12), 760–773.
- Sanz-Gaitero, M., Seoane-Blanco, M., van Raaij, M.J., 2019. Structure and function of bacteriophages. In: Harper, D., Abedon, S., Burrowes, B., McConville, M. (Eds.), *Bacteriophages*. Cham: Springer. https://doi.org/10.1007/978-3-319-40598-8_1-1.
- Schultz, E.C., Ficner, R., 2011. Knitting and snipping: chaperones in beta-helix folding. *Current Opinion in Structural Biology* 21, 232–239.
- Yehl, K., Lemire, S., Yang, A.C., *et al.*, 2019. Engineering phage host-range and suppressing bacterial resistance through phage tail fiber mutagenesis. *Cell* 179 (2), 459–469.

Phage Genome and Protein Ejection In Vivo

Ian J Molineux and L Letti Lopez, The University of Texas at Austin, Austin, TX, United States

Aaron P Roznowski, The University of Texas at Austin, Austin, TX, United States and University of Arizona, Tucson, AZ, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Continuum mechanics theory A theory that considers packaged phage DNA as a uniformly charged rod with a specified persistence length (rigidity) organized in a toroidal spool on a hexagonal lattice.

Contractile tail Myophage tails have an internal tail tube encased in an outer sheath. Only the sheath contracts.

Cryo-electron microscopy/tomography (cryoEM/cryoET) Sometimes referred to as electron cryomicroscopy. In both techniques, aqueous samples on a carbon grid are instantaneously frozen in liquid ethane at -196°C . CryoEM, usually specifically referred to as single particle cryoEM, takes a single picture of thousands of particles, which are different orientations in the liquid film. These images are then computationally combined to generate a 3-D image. CryoET is used for larger samples but has much lower resolution. A series of images are obtained from a single sample at different tilt angles to the electron beam. Computation can then generate a 3-D image of the sample.

Higher resolution structures can subsequently be obtained by combining data from different samples.

DNA-condensing agents Among other compounds, low concentrations of the polyamines spermine and spermidine, and the trivalent cobalt hexammine chloride, condense DNA in solution to densities that approach that in phage heads. PEG and high $[\text{NaCl}]$ also induces condensation to form a similar DNA structure.

Extensible tail Some podophages eject internal proteins from their capsid that functionally extend tail length.

Proton motive force The central aspect of Mitchell's chemiosmotic theory. The pmf has two components: a membrane potential (electrochemical gradient) and a pH gradient across the two sides of the cytoplasmic membrane. The protonophore FCCP (carbonyl cyanide-p-trifluoromethoxyphenylhydrazone) is an aromatic, negatively charged compound that can delocalize the charge of bound protons, enabling proton transfer across lipid bilayers. FCCP collapses both the electrochemical and pH gradients of the cytoplasmic membrane.

Protein Ejection and Trans-Envelope Channel Formation

The first observations relating to how phage genomes enter the infected cell were those of Anderson, who obtained electron micrographs showing that phages adsorbed to cells through their tails (although ϕX174 was known at the time it had not been recognized as being tail-less). The following year the classic Hershey-Chase experiment showed that the majority of ^{32}P -labeled T2 DNA but only a minority of ^{35}S -labeled protein became cell-associated, i.e., Waring blender-resistant, after infection. The Hershey-Chase experiment is often described as proving that DNA is the genetic material; however, that had actually been conclusively demonstrated eight years earlier by Avery *et al.* Their demonstration that highly purified DNA constituted the transforming principle was incontrovertible. However, Avery only made the claim privately that DNA was likely to be the genetic material, possibly because the scientific community at the time was not ready to accept his conclusion. The prevailing feelings are perhaps best reflected in a 1946 Nobel Prize being awarded for the structure of Tobacco Mosaic Virus's protein shell. TMV's RNA genome had not been detected, thus protein "must be" the genetic material. Surprisingly, even though their conclusion was in agreement with Avery's data, the Hershey and Chase paper failed to cite Avery *et al.*

Although simple and elegant, the Hershey-Chase experiment has been grossly over-interpreted. That single experiment showed that blending released 75%–82% of T2 ^{35}S -labeled proteins from infected cells but only 15%–35% of T2 ^{32}P -DNA. The paper also reported that less than 1% of the blender-resistant ^{35}S , but a substantial proportion of the ^{32}P , was found in progeny phage. Several additional experiments were also presented in the paper and all the data, not just the usually cited single experiment, are essential for the solid conclusions Hershey and Chase made. However, their conclusion that "protein [of the infecting virion] has no function in the growth of intracellular phage" requires further qualification.

That conclusion is largely correct if one only considers the phage's major structural proteins, which protect the genetic material from environmental nucleases. In addition, long phage tails usually protect the nucleic acid during its translocation into the host's cytoplasm. However, this often involves ejection of virion proteins into the host cell that can protect the infecting virus's genetic material from well-known periplasmic and sometimes cytoplasmic nucleases. These proteins are all likely to be "blender-resistant" in a Hershey-Chase experiment. It remains a conundrum that, at least when using laboratory hosts and laboratory conditions, the DNA genome of the *Bacillus subtilis* phage SPP1 is transiently exposed to the environment and is nuclease-sensitive during infection. As a final comment on phage proteins entering the infected cell, some phages eject their own RNA polymerase, unequivocally essential for phage development, into the host. Thus, some proteins contained within the infecting virion do indeed have a function during intracellular growth. Examples of this phenomenon will be described throughout this article.

We will utilize the phagocentric term "ejection"; phage proteins and DNA are ejected from the virion during infection. The term is much preferable to its counterpart "injection" because it avoids the misleading implication of phages being thought of as a hypodermic syringe; unlike when using the latter, the volume inside a phage capsid does not change significantly during infection.

Proteins are Ejected into the Host Cell at the Initiation of Infection

Hershey himself first showed that “the germinal substance of bacteriophage T2” was ejected into the infected cell; this was most likely the IP (internal proteins) of T-even phages. T4 IP proteins are non-essential for growth in common laboratory hosts but one has been shown to inhibit a restriction enzyme in a hospital-derived strain. Avoiding host restriction by ejecting internal virion proteins is a strategy used by other phages, including the workhorse *Escherichia coli* transducing phage P1. Another T4 IP is gpAlt, which, among other target proteins, ADP-ribosylates one α -subunit of bacterial RNA polymerase in a process that leads to the preferential transcription from T4 promoters after infection. It is not known if these proteins leave the phage capsid before, concomitant with, or after genome ejection, but if the latter there cannot be a significant delay as they need to function rapidly to promote phage infection.

Recently, some giant phages (defined as possessing a genome $> \sim 230$ kb, compared to the ~ 170 kb T4 genome) have been shown to harbor all the subunits of a bacteria-like multi-subunit RNA polymerase inside their capsid. Growth of the *Pseudomonas aeruginosa* phage ϕ KZ has been demonstrated to be independent of the host transcriptional apparatus. The virion-encapsidated RNA polymerase subunits must therefore be ejected into the cell in order to transcribe the phage genome. Whether enzyme ejection precedes, occurs simultaneously with, or follows genome ejection is not known. It should be noted that proteins ejected into the infected cell cytoplasm in an unfolded (perhaps partially) state have the possibility of being refolded by host chaperones. However, unfolded proteins that have to refold in the cell envelope probably do not have this option. Only limited chaperone activity has been observed in the periplasm, and most ejected proteins likely spontaneously fold into their active conformation.

All long-tailed phages must eject at least one protein into the cell before genome translocation into the cytoplasm can occur. Tail length is controlled and made constant by a protein tape measure that occupies the lumen and extends through most of the tail in mature particles. The lumen of the tail is also used as the conduit for genome ejection, and thus the tape measure protein must be removed in order to allow DNA egress. Although there is not yet any direct experimental evidence, there are suggestions that tape measure proteins may facilitate formation of the channel across the cytoplasmic membrane and allow the phage genome to access the cell cytoplasm. Some tape measure proteins have also been shown to harbor a cell wall-degrading activity; these proteins therefore do have at least one defined function inside the infected cell.

Tail-Less Phages

Microviruses, e.g., ϕ X174 and St-1, which lack any external tail on their isometric capsids, have perhaps the most obvious requirement for proteins to be ejected from the virion before any DNA. The capsid of these phages, like those of the more common dsDNA tailed phages, remains on the infected cell surface after genome ejection. ϕ X174 ejects its gpH protein into the cell in order to allow genome transfer into the cell cytoplasm (Fig. 1). gpH was originally called a pilot protein because it appeared to escort the single-stranded (ssDNA) circular genome into the cell cytoplasm. Several molecules of gpH have now been shown to form a tube, functionally equivalent to a tail, through which the ssDNA is transported. Phage protein synthesis is not required but the activity of host proteins, which convert the ssDNA to supercoiled dsDNA (replicative form, RF), may be necessary for complete genome internalization.

The filamentous (ff) ssDNA phages (f1, fd, M13; *Inoviridae*) were once also described as possessing a pilot protein gp3. However, gp3 mainly functions as an adsorption protein that is brought to the surface of the infected cell as the F pilus retracts. Gp3 then interacts with TolA, an inner membrane-anchored component of several interacting proteins spanning the cell envelope of *E. coli*. The TolQRA protein complex mediates disassembly of the f1 coat protein in the inner membrane (where it is

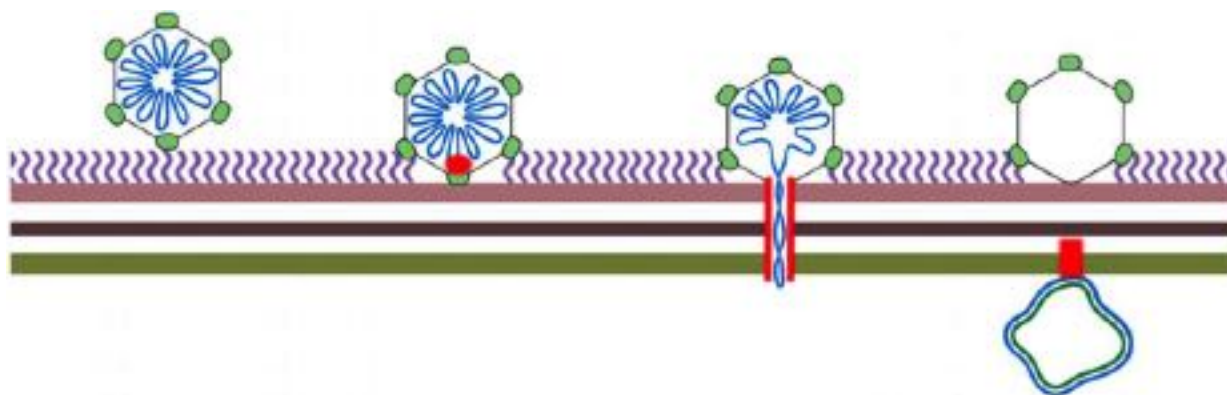


Fig. 1 ϕ X174 infection pathway. Pathway of ϕ X174 genome ejection. After ϕ X174 adsorbs to the bacterial LPS, one of the gpG spikes (green) dissociates. Dissociation causes conformational changes in the newly exposed capsid proteins that result in the opening of a pore at the attached vertex. The pilot proteins gpH (red) emerge and form a tube across the cell envelope through which the ssDNA genome can pass into the cytoplasm. Adapted from Sun, Y., Roznowski, A.P., Tokuda, J.M., *et al.*, 2017. Structural changes of tailless bacteriophage ϕ X174 during penetration of bacterial cell walls. *Proceedings of the National Academy of Sciences of the United States of America* 114, 13708–13713.

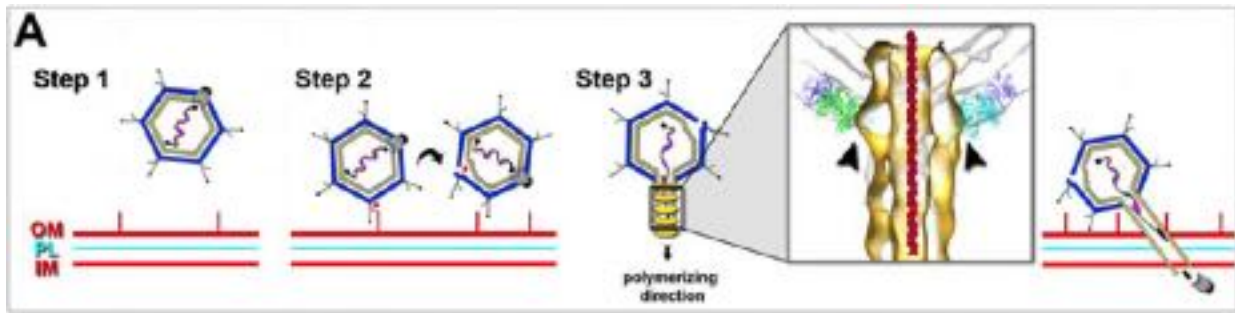


Fig. 2 *PRD1 infection pathway.* After its initial collision with a cellular receptor, PRD1 undergoes random movement over the cell surface, eventually orienting its unique vertex almost perpendicular to the cell surface. De-capping of this vertex triggers reshaping of the internal viral membrane by an osmotic flow from the environment into the cell. Reshaping includes formation of a membrane tube that traverses the cell envelope. Once in the cytoplasm, the tube tip is opened to allow DNA delivery. Reproduced from Peralta, B., Gil-Carton, D., Castaño-Díez, D., *et al*, 2013. Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLOS Biology* 11 (9), e1001667.

subsequently reused for progeny phage particles) and the *ff* genome may therefore simply enter the cytoplasm by default. The single-strand replication origin is then utilized by pre-existing host enzymes to create RF DNA before mRNA synthesis is initiated.

Little is also known about genome internalization of the small single-stranded RNA (ssRNA) phages (*Leviviridae* and *Alloleviviridae*), the best known of which are MS2 and Q β , which infect F plasmid-containing bacteria. Similar phages are known to adsorb to other conjugative plasmid pili. The single copy of the maturation protein per particle is essential for infectivity; it is cleaved into two fragments and leaves the virion together with the RNA genome. Surprisingly, at this stage of infection, at least in the laboratory, the genome is sensitive to extrinsic RNases. The final location of the maturation protein fragments after infection is not known; how the genome avoids degradation by the periplasmic enzyme RNase I, and how the RNA crosses the cytoplasmic membrane are also poorly understood processes. However, recent studies of both MS2 and Q β by cryo-electron microscopy (CryoEM) are beginning to reveal more details.

Internalization of double-stranded RNA (dsRNA) phage genomes (*Cystoviridae*) is more akin to animal viruses. These bacteriophages have segmented genomes packaged as a nucleoprotein complex that is inside a protein shell. The latter is surrounded by an external membrane consisting of phospholipids and proteins. The membrane fuses with the outer membrane of the host bacterium, and the nucleocapsid enters the periplasm. An endopeptidase that is exposed on the nucleocapsid surface likely hydrolyzes the cell wall to allow passage across the periplasm. An invagination of the cytoplasmic membrane and intracellular vesicle formation is thought to allow the nucleocapsid to enter the cytoplasm where the protein shell is removed. The remaining nucleoprotein core is then active for RNA-dependent transcription and replication. There is therefore no genome or protein ejection *per se* by this class of phages.

The marine *Pseudoalteromonas* spp. phage PM2 (*Corticoviridae*) has an outer protein capsid underlain by a lipid membrane. This phage is novel not only in its mode of genome entry but also in that no virion protein is known to enter the infected cell. PM2 adsorbs to an unknown outer membrane component(s) that in some way senses the cytoplasmic ATP concentration - and not the proton motive force. If the [ATP] is high, the capsid protein dissociates, exposing the lipid core that then fuses with the outer membrane. If the [ATP] is low, the infection is aborted. Fusion of the lipid core releases the genome into the periplasm where it enters the cytoplasm through a membrane pore likely composed of cellular proteins. Interestingly, Ca⁺⁺ ions are essential, not only to stabilize the outer membrane of the infected cell but also to facilitate localized cell wall hydrolysis by a host-encoded muralytic enzyme.

Although the structure of PM2 is similar to the Tectivirus PRD1, the mode of PRD1 infection is strikingly different (Fig. 2). Receptor recognition by PRD1 triggers structural rearrangements and decapping of a capsid vertex. This allows the internal membrane to form a tube-like structure that penetrates the infected cell envelope, forming a channel for phage genome translocation. A virion protein that is embedded in the tube facilitates localized degradation of the cell wall. The energy for tube extension is provided by the flow of liquid along the osmotic gradient from the environment (growth medium, low osmolarity), through the filled capsid, into the higher osmolarity composition of the periplasm/cytoplasm of the cell. The empty PRD1 capsid remains attached to the cell surface. Genome internalization of the related phage Bam35, which infects Gram-positive bacteria, has also been studied.

Short-Tailed Phages (*Podoviridae*)

Short-tailed phages infecting Gram-negative hosts also obviously require proteins to be ejected from the virion before DNA. The short, stubby tails of podophages are too short to span the cell envelope from their adsorption site on the cell surface. The coliphage T7 was the first podophage observed to eject proteins (E proteins) into the cell. In mature virions these proteins constitute a distinct morphological structure inside the capsid abutting the portal and they are ejected into the cell before the genome. Inside the cell they form a dramatically different structure that extends from the tip of the phage tail through the outer membrane, and across both the periplasm and the cytoplasmic membrane (Fig. 3(A)). This elongated structure comprises the channel for DNA transport into the cell. T7 has therefore been described as having an extensible tail, in contrast to the contractile tail sheath of myophages. Other podophages, e.g., salmonellaphages SP6 and P22 (Fig. 3(B)), have also been shown to have extensible tails. This feature is likely common to this phage morphotype, at least of phages infecting Gram-negative hosts.

Clearly, these ejected proteins, as do the tape measure protein of long-tailed phages, have a defined and essential function in facilitating genome entry into the cytoplasm and thus the subsequent growth of intracellular phage.

A major question remains in understanding how protein ejection is accomplished. Three proteins constitute the T7 extended tail and the process must be tightly coordinated as 12 copies of gp14 localize exclusively to the outer membrane of the infected cell, whereas the eight gp15 and four gp16 molecules function together in spanning the periplasm and the cytoplasmic membrane. T7 gp16 contains the only predicted helical transmembrane domain of the ejected proteins (E proteins), and disrupting that predicted helix abolishes infectivity. It has also been shown to bind membranes *in vitro*. Gp16 is also responsible for the localized enzymatic degradation of the cell wall that facilitates phage genome internalization. However, the structure of packaged gp16 is far too large to exit through the phage portal and tail intact; the four copies of this 144 kDa protein must at least disaggregate and partially unfold during ejection. Gp16 must therefore also spontaneously refold, without the help of energy-dependent chaperonins, to form an active enzyme in the periplasm. A similar argument can be made for the eight copies of the 84 kDa gp15. Gp15 has been shown to bind DNA, and may thus escort the leading end of the phage genome into the infected cell.

T7 is not unique in having three E proteins; those of P22 are also present in multiple copies in the virion. Their ejection is clearly controlled as distinct intermediate extended tail structures have been observed during infection when one E protein is defective. Presumably all copies of each E protein are ejected together and separately from other E proteins. How this is accomplished is unknown. P22 also discards its tail needle after it has penetrated the outer membrane. The needle, as with the central tail fiber of the Siphophages described below, blocks the exit channel for phage genome translocation.

In Gram-negative bacteria, the cell wall is composed of a single layer of peptidoglycan, an open network that allows passage of globular proteins up to 45–50 kDa. As mentioned above for some specific phages, many virions contain a cell wall-degrading activity. Note that this cell wall-degrading activity is distinct from the endolysin that elicits cell lysis at the end of the phage growth cycle. Lysozyme, lytic transglycosylase and endopeptidase (cleaving the cross-links in peptidoglycan) activities have been found as part of structural virion proteins. These must therefore penetrate into the periplasm of the infected cell. In T4 and T7, enzyme activity is not essential for infection when the host bacteria are growing exponentially at or above 30°C. However, as cells approach stationary phase or are growing at low temperature, plaque formation requires enzyme activity. In its absence the phage latent period during growth in liquid is greatly extended. T-even phages release their tail-associated lysozyme into the periplasm where it can diffuse and cause extensive degradation. High multiplicities of infection by T-even phages lead to lysis-from-without, a phenomenon where the newly infected cell immediately lyses without producing progeny. Lysis-from-without is primarily a property of T-even phages; most phages infecting Gram-negative cells that contain virion-associated muralytic enzymes constrain that activity to the infecting particle and thus the cell wall is only degraded at the site of infection. However, cell-wall degrading activity has not been detected in all phage virions: λ and its relative HK022, P22, and ϕ X174 appear not to degrade peptidoglycan during the initiation of infection. It is unknown whether the genome ejection apparatus of these phages can freely pass through the peptidoglycan layer or whether the infecting phage merely waits for the normal turnover of peptidoglycan during cell growth to generate a sufficiently large gap.

The thick cell walls of Gram-positive cells require that virion enzymes degrade either or both peptidoglycan and teichoic acid to allow phage access to the cytoplasmic membrane. Such enzymes are often present on the baseplate or tailspikes of phages. The podophage ϕ 29, for example, hydrolyzes both polymers as it drills its way through the *B. subtilis* cell envelope so that its tail knob protein gp9 can fuse with the cytoplasmic membrane (Fig. 3(C)).

Non-Contractile Long-Tailed Phages (Siphoviridae)

In several long-tailed phages, including λ , T5, and the contractile tailed P2, it has been shown that mature virions have the leading end of the genome extending into lumen of the tail. Linear phage genomes always enter the cell with a fixed polarity. Most phages that have been studied exhibit a first-in (during packaging), last-out (during ejection) strategy. An exception is T4, which uses a first-in, first-out strategy. In either scenario, the end that exits first must be positionally constrained during packaging. Cross-linking studies with long-tailed phages have shown that DNA penetrates the portal channel and can extend into the lumen of the tail. As these phages assemble their tails independently of the capsid and of DNA packaging, the plug that closes the portal channel prior to tail addition must be remodeled or removed in the final assembly steps. Recent cryoEM reconstructions have extended the observation of a genome end in the portal channel to short-tailed phages.

These are significant observations suggesting that triggering of genome ejection may only require conformational changes at the tail tip and, for long-tailed phages, ejection of the tape measure protein. The latter may serve to transduce a signal to head-proximal components to initiate genome ejection. It should also be noted that positioning one end of the genome inside the tail lumen during virion assembly provides a direct and efficient path for DNA ejection; it could be difficult for a free genome end inside the capsid to find the exit channel.

Some early studies of phage genome ejection employed λ . A repeat of the “Hershey-Chase experiment” allowed the same conclusions originally made with T2: the majority of λ proteins could be removed by blending while the majority of DNA remained cell-associated. It was also shown that λ can eject its genome *in vitro* by the simple addition of Mg^{++} and the purified receptor protein LamB. *Shigella sonnei* LamB is effective by itself although *E. coli* LamB also requires the addition of $CHCl_3$ to cause genome ejection. Pioneering *in vitro* studies using electron microscopy of LamB-containing liposomes or crude membranes prepared from λ -infected cells revealed the presence of two distinct complexes: one showed the central gpJ fiber interacting with the outer membrane

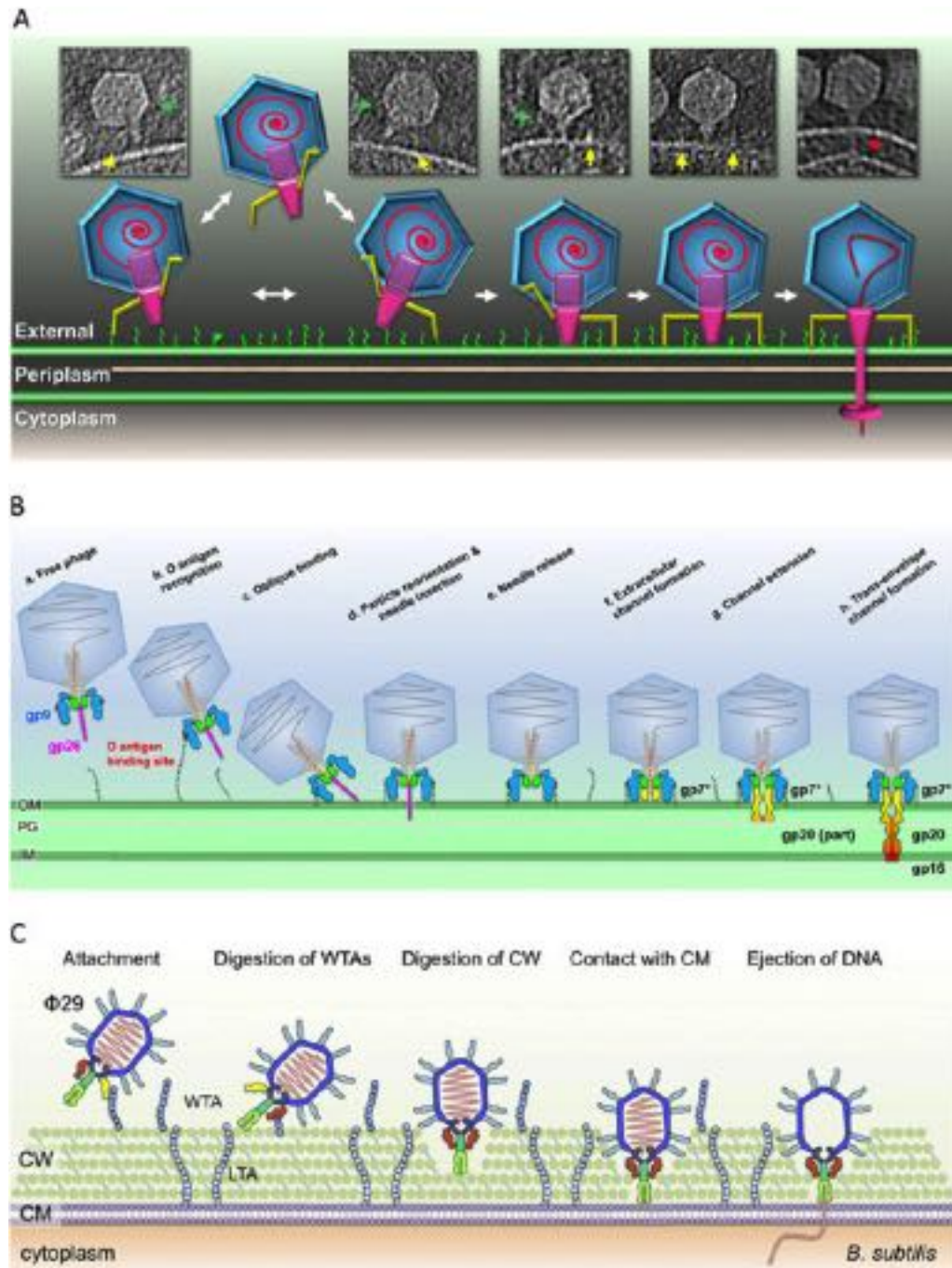


Fig. 3 Podophage infection pathways. 3A. T7: After initial adsorption to outer core LPS, T7 is thought to “walk” over the cell surface using a subset of fibers. At a preferred site, “walking” stops and all tail fibers bind, allowing the tail protein gp12 to interact with its receptor in the lipid A-KDO region of LPS. This interaction initiates ejection of the small gp6.7 and gp7.3 proteins into the outer membrane, where they are normally degraded. The internal core proteins gp14, gp15 and gp16 then disassemble and eject into the cell to form a trans-envelope channel from the tip of the tail extending into the cytoplasm. The T7 genome is then translocated into the cell using molecular motors. For more details, see Hu *et al.* 3B. P22: Tailspikes bind to O antigen and by hydrolysis bring the phage to the cell surface. Needle length causes the virion to initially bind at an oblique angle. The virion then reorients, forcing the needle through the outer membrane where it is released. In turn, needle release triggers ejection of the P22 E proteins that form a complete channel for DNA transport extending from the phage baseplate into the cell cytoplasm. For more details, see Wang *et al.* 3C. ϕ 29: ϕ 29 uses its gp12* appendages (called tailspikes in other phages) to adsorb to and then hydrolyze *B. subtilis* wall teichoic acid to bring the virion close to the cell surface. As with P22, the ϕ 29 tail causes the virion to bind obliquely but the gp13 tail tip, which harbors a muralytic activity, can now interact with the cell wall. Continued enzyme activity, both of gp12* and gp13, causes reorientation of the virion and fusion of the gp9 knob protein with the cytoplasmic membrane to form a channel for DNA ejection. For more details, see Farley *et al.* and Xu *et al.* Reproduced from (A) Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* 339, 576–579. (B) Wang, C., Tu, J., Liu, J., Molineux, I.J., 2019. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nature Microbiology* 4, 1049–1056. (C) Farley, M.M., Tu, J., Kearns, D.B., Molineux, I.J., Liu, J., 2017. Ultrastructural analysis of bacteriophage Phi29 during infection of *Bacillus subtilis*. *Journal of Structural Biology* 197, 163–171. Xu, J., Gui, M., Wang, D., Xiang, Y., 2016. The bacteriophage ϕ 29 tail possesses a pore-forming loop for cell membrane penetration. *Nature* 534, 544–547.

(or LamB in the proteoliposome), whereas the second lacked gpJ and the phage tail contacted the membrane surface. In this type II complex, gpJ is sensitive to proteolysis, suggesting that it does not enter a proteoliposome whereas the tape measure gpH*, which is protease-sensitive in free phage, becomes resistant and thus is likely buried in the membrane. This same complex was subsequently shown to form a transmembrane channel, providing the first suggestion that a phage tape measure protein could be ejected from an infecting virion and form a conduit for genome translocation. Unfortunately, perhaps due to methodological limitations of the time, these experiments were not pursued further. The structures of the counterpart proteins to gpJ from the Gram-positive phages SPP1 and p2 are now known; the structures support the idea that the central tail fiber bends out of the way during infection to allow genome release, a feature that may thus be common to many siphophages,

Contractile Long-Tailed Phages

The prototype contractile tailed phage is obviously T4, and we now have a good understanding of the initial steps of infection, particularly for the process of tail tube penetration of the cell envelope (Fig. 4). Other well-studied myophages likely follow a similar pathway, although their baseplate structures are simpler than that of T4 and not all may possess cell wall-degrading activity.

Atomic structures of many T4 proteins have been obtained by X-ray crystallography, and detailed structures of particles both before and after genome ejection *in vitro* have been solved by cryoEM. Furthermore, cryoET has elucidated trans-envelope channel formation for genome ejection *in vivo*. Reversible adsorption of at least three long tail fibers to the cell surface destabilizes the baseplate. This results in the short fibers rotating downwards where they contact the cell and irreversibly bind. In addition, baseplate destabilization allows the tail sheath to spontaneously contract, which drives the inner tail tube through the outer membrane. The tail tube tip then dissociates, releasing the gp5* lysozyme complex into the periplasm. Degradation of the cell wall facilitates further penetration of the tail tube into the cell.

Following tail sheath contraction, the end of the T4 tail tube (now lacking its tip complex) is in the periplasm; despite innumerable cartoons to the contrary it was clearly shown 50 years ago that the tube only extends to the outer leaflet of the cytoplasmic membrane. CryoET now reveals that the membrane blebs outwards in fusing with the tail tube. For DNA to be transported into the cell cytoplasm, the tail tape measure protein must be ejected and it is currently thought that this protein forms the transmembrane channel. Blebbing may then be the simple consequence of the membrane moving towards the fixed end of the tail tube during the fusion process.

Only in myophages has the source of energy been clearly established for changes in protein conformation that occur during infection initiation. The outer sheath of the contractile tailed phages is assembled in a metastable form. When infection is triggered, the sheath spontaneously contracts to a lower energy state structure, one that may be identical to that obtained with purified protein. Although sheath contraction is an essential step during infection, it does not necessarily lead to genome ejection. Sheath contraction can be induced *in vitro* by treating virions with urea but the resulting non-infective virions retain an encapsidated genome. Comparable detailed studies have not been performed with podophages but the structure of purified and complexed T7 gp15 and gp16 bears more resemblance to the extended tail than their structures in the mature virion. Perhaps the internal E proteins of T7 are also assembled in a metastable form.

Factors Affecting Phage Genome Translocation

In 1956, the first quantitative model of ejection hypothesized that simple Brownian motion was responsible for genome ejection, but it was later calculated that the time that would be spent for complete genome ejection by diffusion greatly exceeded the phage latent period. A theoretical proposal, one that was supported using a 10^5 -scale glass model of the T4 virion with nylon thread representing the DNA (!), invoked osmotic pressure inside the phage head as driving the ejection process. However, the most widespread description derives from the first molecular genetics textbook authored by Gunther Stent in 1963: "One may imagine, therefore, that the DNA is packed into the phage head under constraint and forces its own way out through the sheath [*Myoviridae*] after the contraction and "uncorking" reactions of the tail are triggered".

Although usually depicted in textbooks as an instantaneous event, complete internalization of most phage genomes may occupy a substantial part of the latent period. Some specific values are given below, but cryoET studies of infection initiation by several different phages have all visualized partially emptied capsids, where DNA is presumably in the process of entering the cell cytoplasm. Such structures would not be detected if genome transport is extremely fast but there is a caveat to those experiments: all but one study used minicells rather than intact bacteria. Minicells, which are smaller than whole cells and thus allow acquisition of higher resolution structures, have been shown to support growth of some phages. However, minicells lack a chromosome and cannot grow, and additional experiments are therefore underway in which exponentially growing cells are infected. It should be remembered that many of the other assays that have been employed to estimate genome translocation rates are indirect and require interpretations that can also be questioned. For example, genome ejection, once initiated, may continue even at low temperature during disruption of the phage-bacterium complex, and although a genome may be shown to have exited the virion, it may not have entered the cell cytoplasm. A second source of variability lies in the nature of the assay – are populations being measured, where the fastest fraction determines the earliest time recorded, or are single phage-cell complexes being observed?

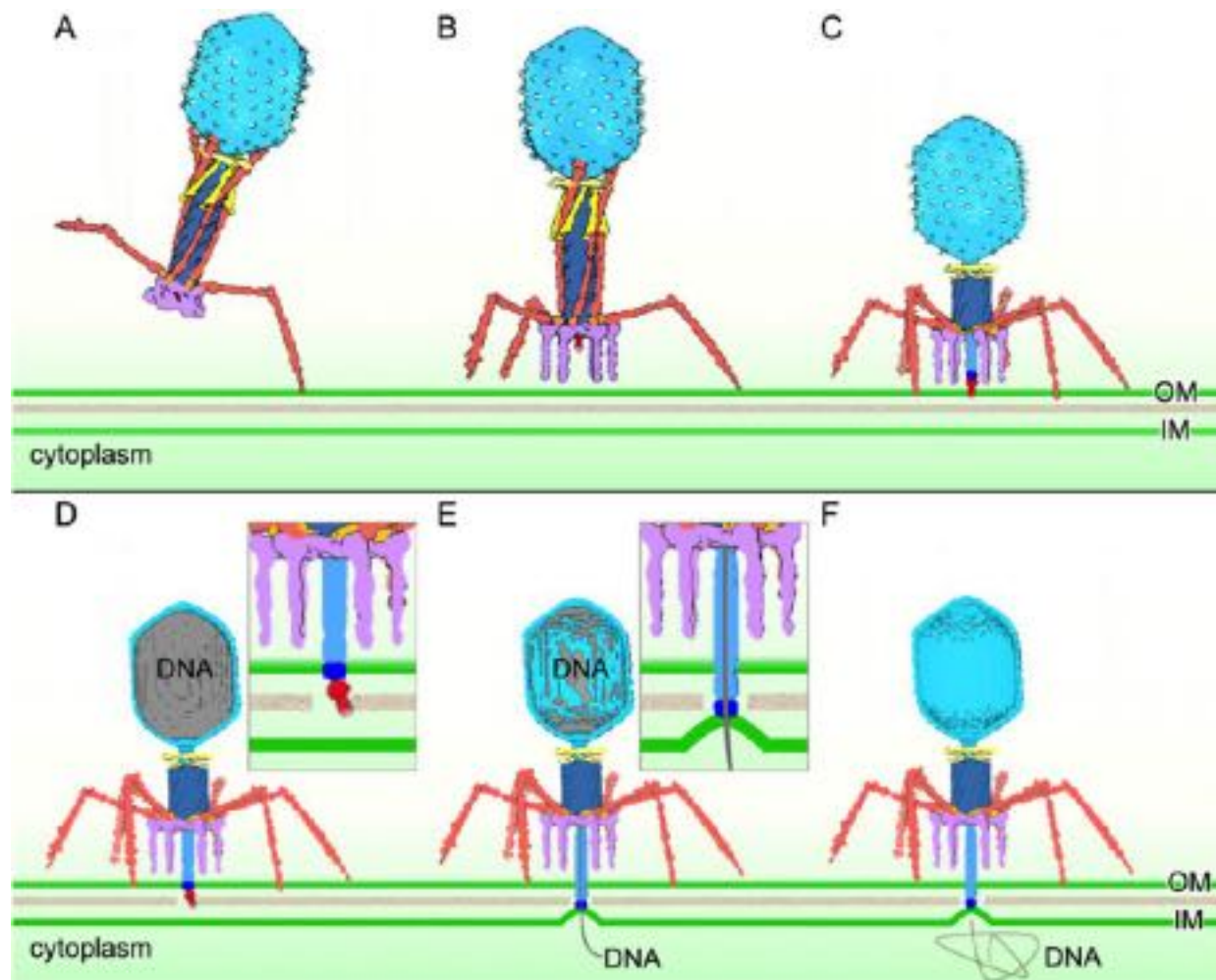


Fig. 4 *T4* infection pathway. Sequential binding of a subset of long tail fibers (LTF) allows T4 to “walk” over the cell surface. When three or more LTF are simultaneously bound, the strain in each LTF–gp9 baseplate junction triggers conformational changes that releases some STF short tail fibers from the baseplate. The latter, in this transient intermediate, rapidly transitions from its hexagonal to a star configuration, which triggers contraction of the tail sheath and the release and binding of the remaining STF and LTF to the cell. Sheath contraction also pushes the tail tube and needle complex through the outer membrane; this complex then dissociates, allowing the gp5* lysozyme to freely digest the cell wall. The gp27 tail tube, now the distal end of the ejection machinery, contacts the cytoplasmic membrane, which bulges out from its normal plane. Bulging may be caused by ejection of the gp29 tape measure that is thought to make the transmembrane channel for DNA translocation into the cytoplasm. For more details, see Hu *et al.* and Taylor *et al.* Reproduced from Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2015. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proceedings of the National Academy of Sciences of the United States of America* 112, E4919–E4928. Taylor, N.M., Prokhorov, N.S., Guerrero-Ferreira, R.C., *et al.*, 2016. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* 533, 346–352.

In addition to the possible presence of DNA phosphate counterions, a function for ion movement has often been proposed during genome ejection. Determining this role has been difficult, in part because ionic strength also affects particle adsorption. Many phages do not adsorb, or do not progress beyond the reversible adsorption step in low ionic strength environments. A very high ionic strength environment is usually also inimical to infection as the initial adsorption steps are governed by electrostatic interactions. In addition, many phages have a specific requirement for Mg^{++} or Ca^{++} , again both for initial adsorption and for the conformational rearrangements of virion proteins necessary to allow genome delivery into the cytoplasm. As examples, Ca^{++} triggers opening of the lactococcal phage p2 baseplate and stabilizes the post-ejection conformation of the distal end of the SPP1 tail *in vitro*. Ca^{++} and/or Mg^{++} are also necessary to stabilize stocks of some, but not all phages.

λ provides an interesting example of the complexities of specific ion requirements for both stability and infection. Free virions and isolated heads are freely permeable to small ions, including the common bacterial polyamines putrescine and spermidine. Isolated tails are slowly (hours) inactivated by Mg^{++} and are usually stored in buffers containing EDTA. Conversely, isolated heads, which have four nucleotides from the single-stranded right end of the genome protruding from the portal, are rapidly inactivated by both EDTA and Mg^{++} due to premature genome release. They are stabilized by putrescine but not by spermidine or spermine, both of which actually markedly increase the rate of genome loss. This is surprising, as spermidine and spermine, unlike

putrescine, are DNA-condensing agents and *a priori* would be expected to stabilize the DNA inside heads. Isolated tails and heads can be mixed *in vitro* to make infective virions that are stable to all polyamines, although Mg^{++} is usually employed for long-term viability. As final complications, λ growth, either by infection or induction of a lysogen, is inhibited in cells grossly deficient in polyamines, but in low (μM) concentrations of Mg^{++} all three common polyamines have been shown to inhibit infection.

Cytoplasmic ions, in particular K^+ , have been shown to leak from cells during the adsorption and genome penetration steps of infection by several phages. It has been suggested that these ion fluxes are responsible for the transient membrane depolarization of the cell that is also associated with the initiation of infection by most phages. Leakage from T5-infected cells occurs only during the two distinct steps of DNA translocation (see below), suggesting that DNA transport and leakage are intimately coupled. However, T5-induced leakage also occurs in cells that have been depolarized, and membrane depolarization can be uncoupled from transport of the SPP1 genome. In addition, K^+ leakage does not require membrane depolarization during PRD1 infection. Ion leakage and/or the transient reduction in membrane potential may simply be a consequence of a phage breaching the cell envelope, an idea first proposed in 1954. Alternatively, membrane depolarization was suggested to be due to protons associating with T1 and T4 DNA as it translocated into the cell. However, in the case of T4, the membrane potential was later shown to be required only for opening or maintaining the channel across the cytoplasmic membrane. In contrast to the above, initiation of infection by T3 or T7 causes neither a reduction in membrane potential nor ion leakage, observations that are currently confined to this phage group. This difference between T7-like and other phages may be the key to understanding their fundamentally different mechanisms of genome ejection.

Transcription-Mediated Genome Internalization In Vivo

The detailed mechanism of DNA translocation from the phage head into the cell cytoplasm has only been established for two phage types; the majority of the T7 and N4 genomes are internalized by transcription. RNA polymerases (RNAP) pull the genome from the phage head into the cell. NTP hydrolysis thus provides the energy both for RNA synthesis and for translocation of the genome out of the phage head into the cytoplasm. Close relatives of these prototype phages likely utilize a similar mechanism of genome internalization, but it has so far only been demonstrated experimentally for the coliphage T3 and the cyanobacteriophage Pf-WMP4.

~ 1 kb of the T7 genome is ejected into the cell at 140 bp/sec ($37^\circ C$) by a process that is consistent with it being enzyme-catalyzed. This reaction then stops and entry of the remainder of the genome requires transcription. It is not known what limits the initial genome internalization process; there is not a specific stopping nucleotide sequence and the ~ 1 kb appears to be a length measurement. Perhaps the leading end of the genome is directly escorted into the cell together with the E proteins gp14, gp15 and gp16. Regardless, promoters for *E. coli* (and T7) RNAP lie on the initial DNA segment that enters the cell, and transcription by the *E. coli* enzyme normally pulls ~ 7 kb of the genome into the cell at 90 bp/sec (Fig. 5). Termination of *E. coli* RNAP-catalyzed transcription at the early terminator T_E is not 100% efficient, and in the absence of phage protein synthesis the *E. coli* enzyme internalizes the entire T7 genome at a constant rate. However, T7 RNAP is encoded on the ~ 7 kb DNA transcribed by *E. coli* RNAP. In normal infections, T7 RNAP is synthesized and the trailing 32 kb of the phage genome is internalized by transcription using the faster phage enzyme. Complete genome internalization takes 5–6 min at $37^\circ C$, about 1/3 of the latent period. Separately measured times of T7 mRNA and late protein syntheses are consistent with the genome internalization rates determined by genome entry experiments.

The transcriptional mode of genome internalization is part of the regulatory process of T7 gene expression. Genes expressed early in infection are the first to enter the cell, whereas late genes, largely encoding structural proteins, enter the cell later and thus can only be expressed after early gene expression. However, this regulation of gene expression is not essential for phage growth. If T7 RNAP is provided in the cell prior to infection, the leftmost T7 promoter (present on the initially ejected ~ 1 kb) is utilized by the exogenously provided T7 RNAP and late mRNAs are synthesized much earlier. Conversely, delaying expression of T7 gene 1 (RNAP) by placing it in the distal 40% of the genome delays genome entry and as a result, synthesis of late RNAs and proteins is delayed.

Mutant virions containing substitutions in two regions of gp16 allow the entire genome to enter the infected cell without the need for transcription. With all substitutions tested, DNA entry occurs at the same constant rate over the entire 40 kb genome without detectable pauses. The substitutions thus appear to inactivate the mechanism that normally stops DNA translocation after the initial ~ 1 kb has been internalized. Importantly, the rate of genome internalization by this mechanism varies with temperature and data can be linearly fitted on an Arrhenius plot, consistent with the process being enzyme-catalyzed. Truncations that remove six or more C-terminal residues of gp16 prevent any DNA from entering the cell cytoplasm; this defect can be suppressed, and viability restored by substitutions affecting gp15. Collapsing the pmf with the protonophore FCCP while the genome is being actively translocated into the cell immediately stops further internalization. Purely by analogy with the F_1F_0 ATP synthase, it was suggested that the ejected gp15 and gp16 proteins formed a rotary pmf-dependent motor that ratchets the genome into the cell base-pair by base-pair. However, it was not determined whether the pmf is directly involved in DNA translocation or in maintenance of the channel across the membrane. RNA polymerase could complete genome internalization (using ATP as its energy source) in the absence of a pmf but only at a much lower rate than in its presence.

In contrast to T7, coliphage N4 virions contain four copies of a phage-specific RNAP (vRNAP). The initial ~ 1 kb of the N4 genome is internalized by transcription but the host RNAP is not involved. The vRNAP must therefore be ejected into the cell cytoplasm before or concomitantly with the leading end of the genome. Extensive genome internalization requires the synthesis of N4 gp1, whose gene is encoded on the initial ~ 1 kb. About 40 kb of the genome is thought to be internalized by this vRNAP-gp1 complex; the remaining ~ 30 kb is catalyzed by transcription using N4 RNAP II and its transcription factor N4 gp2. For many years, N4 was an orphan with no known relatives. However, this phage type now has many examples that infect *Pseudomonas* spp.

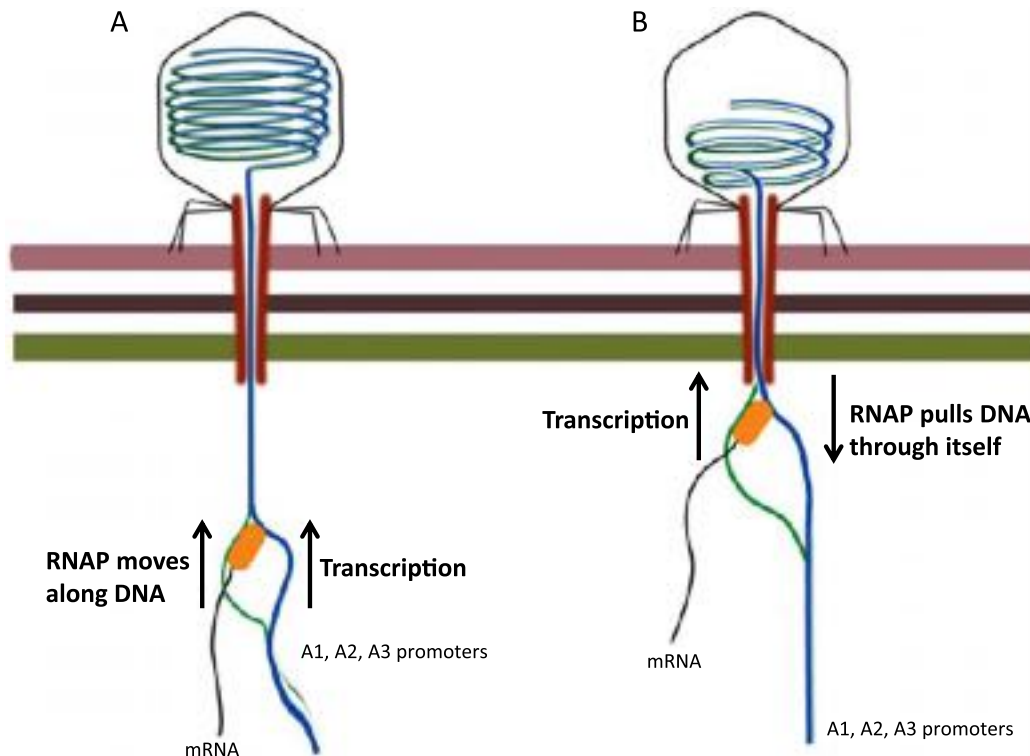


Fig. 5 Transcription-mediated T7 genome internalization. T7 ejects ~ 1 kb of its 40 kb genome into the cell using a molecular motor (see text). On the leading end are three strong promoters for host RNA polymerase (RNAP) and a T7 RNAP promoter (not shown). A. Transcription by RNAP from a promoter proceeds normally. B. Transcription continues unabated once RNAP encounters the cell membrane. Rather than the enzyme moving along its template, duplex DNA must now be pulled through the enzyme. The fate of the superhelical turns that are created ahead of the transcription complex is not known, although topoisomerases do not appear to be involved.

The conservation of the two phage-specific RNAPs and their cofactors gp1 and gp2 make it likely that the description of N4 genome internalization applies to the entire group.

Transcription-mediated genome internalization is currently limited to these two phages and their close relatives, and it seems unlikely that many different phage groups will be shown to employ a similar strategy. It is unsuitable for the ~ 50% of tailed phages that package by a headful mechanism (*pac*-type) and therefore have terminally redundant sequences; appropriate promoters for RNAP would not necessarily be correctly located on the varying genome ends of individual virions. The same argument precludes any sequence-specific DNA-binding protein, be it phage- or host-encoded, from facilitating internalization of *pac*-type genomes after only a short segment has entered the cell. Furthermore, some phage genomes that have fixed genome ends (*cos*-type packaging), e.g., λ and T5, are known not to utilize transcription to catalyze genome entry. However, in addition to T7 RNAP and N4 RNAP II, other phage-encoded proteins are known to function in multi-step mechanisms of internalization of *cos*-type genomes.

Non-Transcriptional Genome Internalization In Vivo

The influence of Stent's 1963 textbook, which informed molecular biologists and geneticists of the time – and then their descendants, coupled with the technological inability to experimentally test the idea that a phage genome forces its own way out of the capsid into the cell cytoplasm led to it becoming widely accepted as “fact” for all phages. However, even in 1963 it was known that T5 genome internalization could not be driven by the release of pressure in the capsid arising from the densely packaged DNA. Even more than five decades later, mechanisms of genome translocation from the capsid into the bacterial cell remain unresolved for most phage groups.

Phage T5 genome entry takes place in two distinct steps, called first-step (FST) and second-step transfer (SST). In FST, the first 10 kb of the 121 kb genome enters the cell in ~ 2 min. The transient leakage of cytoplasmic ions and the accompanying decrease in membrane potential stops after FST completion, without a requirement for phage protein synthesis. FST is normally followed by a ~ 5 min pause when the membrane potential is restored and the A1 and A2 genes essential for complete genome internalization, are transcribed and translated. What causes the pause after FST transfer is unknown. The naturally occurring nicks in the T5 genome, the first of which is close to the end of FST, are not required for pausing. However, the DNA sequence surrounding the first nick is replete with direct and inverted repeats that can be imagined to be involved in FST termination; host proteins that bind to the repeats could also have a role.

After synthesis of A1 and A2, SST begins and is associated with a second transient leakage of cytoplasmic ions and a second drop in membrane potential. A2 is a simple dsDNA-binding protein but A1 has many functions that include binding to A2, host membranes, and RNAP. The actual role(s) of A1 and A2 in facilitating SST has not been determined but genome internalization proceeds at a rate of ~ 500 bp/sec and does not require transcription. The assays used to measure T5 genome internalization would not have been adequate to detect pauses in SST *in vivo*, as demonstrated *in vitro*, but if any do occur they are not as pronounced as the termination of FST. The only mutants known to affect SST lie in the A1 and A2 genes. Internalization of the entire 121 kb genome thus occupies ~ 10 min at 37°C . Use of metabolic poisons both to deplete intracellular ATP and to abolish the pmf had no effect on the DNA transport process of either FST or SST. It was thus concluded that genome translocation occurred by passive, or perhaps facilitated, diffusion. Significantly, removal of the capsid after FST (and after synthesis of A1 and A2) still allowed the now naked ~ 110 kb of SST DNA to translocate into the cell cytoplasm and at about the same rate as in a normal infection. These data are obviously inconsistent with the idea that any physical forces associated with DNA densely packaged in phage heads cause genome ejection into the cell cytoplasm.

The $\phi 29$ genome also enters a *B. subtilis* cell in two phases, but unlike T5, the second requires metabolic activity. The first 65% of the genome is ejected into the cytoplasm in what was described as the “push” phase, where internal virion pressure was considered the driving force. Two proteins synthesized early in infection, gp17 and gp16.7, which are coded on the leading part of the genome, are necessary to complete DNA transfer. This second step was called the “pull” phase as energy from the membrane potential was shown to be required, although no detailed mechanism was proposed.

An early study of phage genome entry *in vivo*, again using the Hershey-Chase approach also employed *B. subtilis*. Using an infective center assay, complete internalization of the 140 kb SP82G (*Spounavirinae*, a relative of SPO1) genome took ~ 6 min at 33°C after an initial adsorption step and was independent of phage protein synthesis. Again using blending or sonication to disrupt phage-cell complexes, a complex marker rescue assay showed that the rate of genome internalization is constant: ~ 2 kb/sec at 33°C , ~ 1 kb/sec at 28°C and ~ 0.6 kb/sec at 25°C ; these data were linearly fitted on an Arrhenius plot. No enzymatic mechanism that could catalyze genome internalization was proposed but the constant rate of DNA transfer at a given temperature and the major effect on rate by varying temperature strongly suggests that something other than, or in addition to, physical processes is critical.

λ genome ejection has been followed *in vivo* using single cell-phage complexes. λ virions containing fluorescently stained DNA were observed by microscopy in real time during infection. The gradual reduction in fluorescence of the phage was used to estimate the amount of DNA ejected. Individual phages showed significant variability in the time for complete ejection into the cell with a mean of ~ 5 min, one to two orders of magnitude longer than for *in vitro* ejection into buffer. Pausing events, where genome ejection temporarily halted, were observed in some but not all phage-bacterium complexes. Comparable experiments were also performed using the λ deletion mutant b221, whose genome lacks 22% of the DNA in wild-type phage. Continuum mechanics theory, popular because it is amenable to analytic procedures, considers packaged phage DNA as a uniformly charged rod with a certain persistence length (rigidity) organized in a toroidal spool on a hexagonal lattice. This theory predicts that after 22% of the wild-type genome had entered the cell it should display the same dynamics as the b221 mutant genome. Although *in vitro* DNA ejection dynamics for both genomes are captured by this theory, the *in vivo* results disagree. The experimental data suggest that the ejection dynamics reflect the amount of DNA that has entered the cell, rather than the amount remaining inside the capsid. As currently formulated, continuum mechanics theory cannot explain phage DNA ejection *in vivo*. The experimental results are also inconsistent with two-step models, in which energy stored in the capsid controls ejection of the leading half of the genome with the distal half being internalized by another mechanism (e.g., $\phi 29$). It was suggested that unspecified bacterium-internal processes dominate ejection *in vivo* but it is difficult to relate the observation that the rate of ejection declines after $\sim 50\%$ of the genome is inside the cell, when the bacterial internal forces proposed to dominate the process would be expected to be increasing. This clever *in vivo* assay comes with few assumptions beyond any effects of the intercalating fluorescent dye, but more experimentation is clearly necessary before any mechanistic conclusions can be made.

The observation that T-even capsids contain polyamines sufficient to neutralize 30%–50% of the DNA phosphates, plus the suggestion that electrostatic repulsion between packaged DNA strands could cause genome ejection was generalized in Stent's model that the phage genome was packaged in the capsid under constraint. An alternative idea, perhaps inspired by Mitchell's Nobel-winning chemiosmotic theory, was proposed by Grinius *et al.*: T4 DNA entered the cell by associating with protons (symport) traveling down their electrochemical gradient. It was estimated that the rate of DNA transfer was 10^3 – 10^4 bp/sec and thus complete internalization of the entire ~ 170 kb genome would take a few seconds to ~ 3 min. These data were obtained using variations of the Hershey-Chase experiment; however, the time taken to form the trans-envelope channel was not considered and the major assumption was made that no DNA transport occurred during the separation of phage-cell complexes. The precise rate of T4 genome translocation may therefore be inaccurate but it does appear to be faster than other phage genomes. Furthermore, although the cellular proton motive force, specifically the membrane potential component, was clearly demonstrated to be important it was subsequently shown that the potential was required for trans-membrane channel opening/maintenance rather than DNA transport *per se*. The mechanism of T4 genome ejection therefore remains undetermined.

Models of Phage Genome Ejection *In Vivo*

The dsDNA genome of all phages examined is packaged at near crystalline density, ~ 500 mg/ml. Beginning with Stent's 1963 textbook and the 1969 model building study of T4 genome ejection by Zárybnický, the idea that phage genomes are packaged

under pressure that is subsequently used to drive genome ejection has become commonly accepted. However, almost all the experiments conducted in order to substantiate the idea have been performed *in vitro*. These experiments have focused on inhibiting λ , T5, and ϕ 29 genome ejection by imposing an opposing osmotic pressure with concentrated solutions of polyethylene glycol, PEG 6000–8000.

However, phage genome ejection into cells *in vivo* is fundamentally different from the *in vitro* experiments and cannot be explained by pressure internal to the phage capsid alone. Bacteria maintain an internal osmotic pressure that is necessarily higher than the surrounding environment so that they can enlarge during growth. This excess osmotic pressure is called turgor. Estimates of turgor in Gram-negative cells range from 1 to 5 atmospheres while that in Gram-positive cells is estimated to be ~ 20 atm. In the same way that suspending phage particles in buffered PEG solutions inhibits genome ejection, turgor achieves the same result *in vivo*. Crucially, turgor opposes the ejection of DNA from the phage head into a bacterial cell, and for example, when $\sim 50\%$ of a λ genome has been internalized, the osmotic pressure of the cell cytoplasm and phage capsid was calculated to have been equalized; ejection using only pressure internal to the capsid must therefore stop. Using current continuum mechanics models, a separate process must always operate *in vivo* to complete genome internalization and allow a productive infection. Various theoretical solutions pertaining to the problem have been put forward but none has yet been supported by an actual *in vivo* experiment.

It is worthwhile considering what biophysical characteristics have been established about mature phage particles. Their genomes occupy about half the available volume inside the capsid, the remainder being filled by internal proteins (in some phages only) and buffer. Capsids of the T-even phages are osmotically sensitive; they are slowly permeable to small ions, but polyamines and larger ions are maintained in the capsid at the concentration found in their last host. Osmotic shock-resistant T4 mutants have been isolated that can exchange polyamines by dialysis. Small ions and polyamines contained within most other phage capsids can be readily exchanged by dialysis and some phages are stable in EDTA solutions, at least at low temperature.

If the DNA phosphates of the genome are ionized then ~ 1.5 M monovalent (or their equivalent) counterions must be present. However, the presence of cations that neutralize ionized DNA phosphates in mature virions has recently been questioned; the packaged genomes of P22 and the eukaryotic virus AAV are reported to be in a largely non-ionic state. If confirmed and generalized by extension to other phages, these observations have the potential to profoundly affect long-standing thinking of genome ejection. If the DNA phosphates are actually neutralized as $-P-O-H$, not only is charge repulsion between helices negated but the lack of free counterions necessarily reduces the osmotic pressure inside the virion.

Several observations, together with the long-term stability of most phages in simple buffers, are also not consistent with the view that phages are “pressurized”. Substantial indentations in a mature λ capsid can be made with the cantilever of an atomic force microscope and the deformation is reversible. A reduction in capsid volume with the concomitant expulsion of internal water is therefore tolerated without loss of viability. The packaged highly supercoiled PM2 genome is reported to exert no detectable pressure on the membrane internal to the phage capsid. Remodeling of the PRD1 internal membrane occurs before genome ejection and is driven by exchange of osmolytes between the environment and the infected cell. Proteins have been shown to be mobile inside the T4 head, and the inner body of giant phages is not found at the same location in different virions and is presumably likewise mobile. Several different phage capsids trapped immediately after rupture on an electron microscope grid reveal that the released genome retains the same dimensions as in the intact head. None of these observations is readily reconciled with a high pressure inside the mature capsid. Furthermore, a single P4 genome, only 30% the length of P2 DNA, has been packaged *in vitro* inside a normal-sized P2 head. The calculated internal pressure of such chimeric particles is less than the osmotic pressure inside the bacterial cytoplasm; in these particles internal capsid pressure cannot therefore drive any genome ejection, yet the chimeras were shown to be infective *in vivo*. By contrast, analysis of electron micrographs of the *Staphylococcus aureus* phage P68 revealed that $\sim 2\%$ of particles had ejected an apparently fixed amount of DNA; it is not known what causes DNA release nor what stops the entire P68 genome from exiting those particles. Quantized partial genome ejection *in vitro* has also been observed from T3 virions.

An alternative process, termed the hydrodynamic model for phage genome ejection *in vivo*, posits that extracellular water flows along the osmotic pressure gradient between the environment and the cell cytoplasm, which, because of turgor, must be at a higher osmotic pressure (Fig. 6). In contrast to phage DNA ejection *in vitro*, which contains only two compartments (phage virion and surrounding buffer) with differing osmotic pressures, and where water may enter the phage head only as DNA leaves, phage-infected cells have three compartments (environment, phage head and cell cytoplasm). Each compartment has its own independent osmotic pressure. However, because water transfer is ultimately between the environment and the cytoplasm of the infected cell, the osmotic pressure of the intermediate compartment, the phage head, is thermodynamically unimportant. Water flowing through the phage head and tail into the cytoplasm transports the genome much as logs are moved downstream by a river current. The DNA in the phage tail exerts a drag force on the water flowing past it; in turn, because action and reaction are equal and opposite, the DNA will be moved by the flowing water into the cytoplasm. It is not immediately obvious how to design experiments that would rigorously test the hydrodynamic model for *in vivo* genome ejection. *In vitro* experiments could demonstrate that water flow from a low osmotic pressure environment can transport DNA out of a phage head into a separate compartment of higher osmotic pressure, but such experiments have not yet been reported.

The fundamental differences between the continuum mechanics model for *in vitro* genome ejection and this idea are that the hydrodynamic mechanism involves three compartments: the bacterial cell, the infecting phage, and the environment. The hydrodynamic model thus posits that the energy required in transporting the phage genome lies not inside the capsid but in the environment surrounding the phage and the bacterial cell. Formal mathematical analyses of all the pressures involved in this three-component situation show that if the osmotic pressure of the cell is artificially set equal to that of the environment, the equations and the overall system reduce to be equivalent to the two-body continuum mechanics model. However, in the *in vivo*

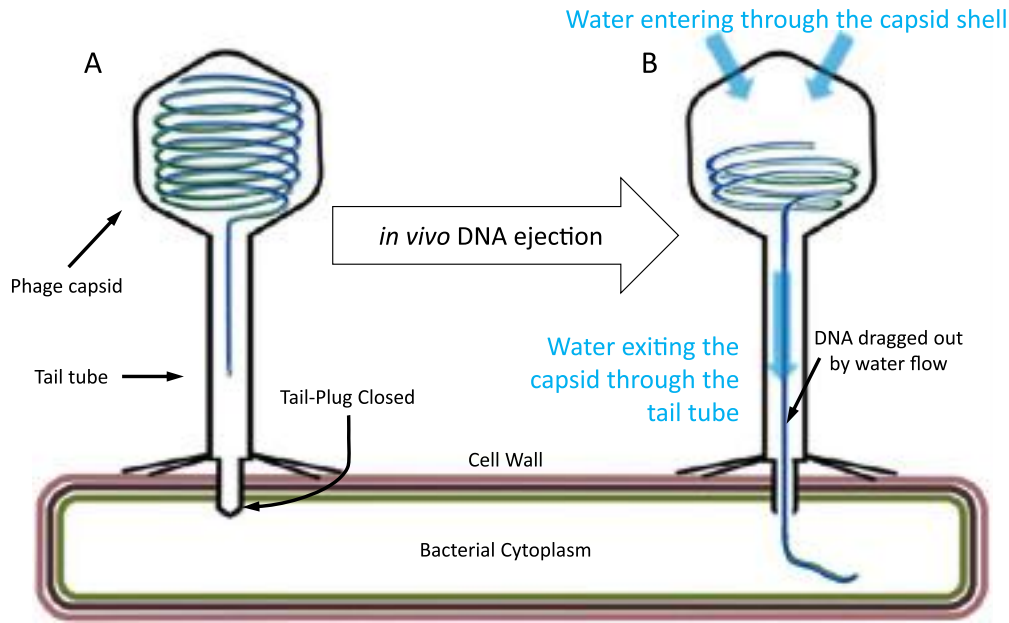


Fig. 6 Hydrodynamic model of DNA ejection. (A). There are two osmotic gradients: low-to-high between the growth medium and the phage, and low-to-medium between the growth medium and the cell cytoplasm (turgor). (B). During infection, the two gradients are connected through the phage and only the overall osmotic gradient between the growth medium and the cytoplasm is of thermodynamic importance. When the tail plug (A) is opened during infection (B), water will osmose across the capsid shell and eventually flow into the cell cytoplasm. Hydrodynamic drag on DNA in the tail tube will cause DNA translocation into the cytoplasm. As long as the channel is open and turgor is maintained, continued water flow will completely internalize the genome and any soluble proteins in the virion. Note that this mechanism cannot apply if phage infection (e.g., T7-likes) does not result in transient K^+ -leakage (and thus water import) from cells. See text for more details.

situation where bacterial cells have a higher osmotic pressure, water will flow from the environment, through the phage head and tail, and into the cytoplasm until the channel across the inner membrane is closed. Under the hydrodynamics model, the osmotic pressure inside the capsid will equilibrate with the environment and not with the cell cytoplasm, as posited by the continuum mechanics model. In theory therefore, the entire phage genome can be transported into the cell by a constant force.

DNA is much more condensed in phage capsids than the bacterial chromosome. In order to package a complete genome *in vivo*, the packaging motor must create room in the capsid by displacing water molecules from the DNA. Some of the work expended by the motor is thus used in expelling water across the capsid shell. DNA inside a mature phage head, although it retains its B-form configuration, is therefore dehydrated relative to the bacterial chromosome and especially relative to free DNA in solution. External water molecules thus have a thermodynamic incentive to enter a mature capsid but they cannot do so until infection is triggered and an open channel is created. Opening the channel allows water to flow into the capsid and out through the tail into the cell cytoplasm. DNA that is already inserted into the long tails of siphon- and myo-phages will then be propelled into the cell.

The hydrodynamic model requires that the channel from the capsid into the cell cytoplasm is wide enough for both DNA and water molecules to pass at the same time. This requirement is reasonable, most phage infections are associated with a transient leakage of intracellular cations; if ions can leak out from the cytoplasm through a DNA-filled channel, water molecules should be readily able to enter the same cytoplasm. Significantly for this model of genome ejection, the only phage group (T7-likes) currently known *not* to leak cations at infection initiation, and thus may not be able to use water flow along an osmotic gradient, is known to require molecular motors to transport the entire phage genome into the cell. T7 has presumably found this mode of genome ejection advantageous for its specific life-style.

The hydrodynamic model also provides a potential explanation for how proteins are ejected into the infected cell, a topic that cannot be addressed by DNA-centric continuum mechanics models. Water flow from the environment, through the phage virion and into the bacterium could provide the necessary force for ejection of the tape measure protein that is common to all long-tailed phages. In this case, the initial open channel may extend only into the periplasm, which, however, is iso-osmotic with the cytoplasm. Some tape-measure proteins have actually been suggested to penetrate the cytoplasmic membrane, completing the channel from the phage capsid into the cytoplasm. Water flow could help transport the intra-capsid proteins that form the extended tail of the short-tailed phage P22, and perhaps facilitate formation of the tail-less ϕ X174 ejection tube. Water flow into the cell cytoplasm is also suggested to cause the extended membrane tube formed by PRD1 during infection initiation. By contrast, as concluded from K^+ leakage experiments T7 does not allow water flow; and as previously suggested, the T7 ejected proteins could be assembled into virions in a metastable state and thus contain the intrinsic energy necessary for their ejection into the infected cell.

It seems unlikely that all the ~ 1000 copies of the T-even IP proteins, and internal proteins of other phages packaged at high copy number are similarly metastable. Unless they are transported while bound to the phage genome, such proteins may require

an external force in order to enter the cell. Importantly, however, there is no temporal constraint on when proteins are ejected under the hydrodynamic model: they can leave the capsid before and/or after genome ejection, or if there is adequate space between the internal wall of the tail and DNA, at the same time. The only requirement is that a channel from the phage capsid into the infected cell is open to allow water flow.

Genome and protein ejection from phage virions remains the least understood aspect of phage development. The continued development of cryoEM and cryoET techniques is yielding much needed structural information of the DNA translocation channel before, during, and after genome ejection. Traditional genetics and physiological studies are therefore now beginning to be fitted into a structural environment. Seeing can be believing, but visualizing structures *in situ* is also instructive – it both informs and constrains interpretations of genome ejection mechanistic data that by their very nature may be indirect and can thus be analyzed in different ways. Finally, the development of better kinetic assays that directly measure how much of a phage genome is internalized in the bacterial cytoplasm would certainly increase our understanding of the mechanisms of DNA translocation used by phages.

Acknowledgments

We thank YeRin Kim for Figs. 1, 5 and 6. Work in the Molineux laboratory is supported by NIH grants GM110243 and GM124378. APR was supported by NSF MCB-1408217 to B. A. Fane at the University of Arizona.

See also: Energetics of the DNA-Filled Head

Further Reading

- Casjens, S.R., Molineux, I.J., 2012. Short noncontractile tail machines: Adsorption and DNA delivery by podoviruses. *Advances in Experimental Medicine and Biology* 726, 143–179.
- Cvirkaitė-Krupovič, V., Krupovič, M., Daugelavičius, R., Bamford, D.H., 2010. Calcium ion-dependent entry of the membrane-containing bacteriophage PM2 into its *Pseudomonas* host. *Virology* 405, 120–128.
- Daugelavičius, R., Cvirkaitė-Krupovič, V., Gaidelytė, A., *et al.*, 2005. Penetration of enveloped double-stranded RNA bacteriophages ϕ 13 and ϕ 6 into *Pseudomonas syringae* cells. *Journal of Virology* 79, 5017–5026.
- Davidson, A.R., Cardarelli, L., Pell, L.G., Radford, D.R., Maxwell, K.L., 2012. Long noncontractile tail machines of bacteriophages. *Advances in Experimental Medicine and Biology* 726, 115–142.
- Farley, M.M., Tu, J., Kearns, D.B., Molineux, I.J., Liu, J., 2017. Ultrastructural analysis of bacteriophage Phi29 during infection of *Bacillus subtilis*. *Journal of Structural Biology* 197, 163–171.
- Fernandes, S., Labarde, A., Baptista, C., *et al.*, 2016. A non-invasive method for studying viral DNA delivery to bacteria reveals key requirements for phage SPP1 DNA entry in *Bacillus subtilis* cells. *Virology* 495, 79–91.
- Gaidelytė, A., Cvirkaitė-Krupovič, V., Daugelavičius, R., Bamford, J.K.H., Bamford, D.H., 2006. The entry mechanism of membrane-containing phage Bam35 infecting *Bacillus thuringiensis*. *Journal of Bacteriology* 188, 5925–5934.
- Hřebík, D., Štveráková, D., Škubník, K., *et al.*, 2019. Structure and genome ejection mechanism of *Staphylococcus aureus* phage P68. *Science Advances* 5, eaaw7414.
- Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* 339, 576–579.
- Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2015. Structural remodeling of bacteriophage T4 and host membranes during infection initiation. *Proceedings of the National Academy of Sciences of the United States of America* 112, E4919–E4928.
- Kemp, P., Gupta, M., Molineux, I.J., 2004. Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. *Molecular Microbiology* 53, 1251–1265.
- Meng, R., Jiang, M., Cui, Z., *et al.*, 2019. Structural basis for the adsorption of a single-stranded RNA bacteriophage. *Nature Communications* 10 (1), 3130.
- Molineux, I.J., 2006. Fifty-three years since Hershey and Chase; much ado about pressure but which pressure is it? *Virology* 344, 221–229.
- Molineux, I.J., Panja, D., 2013. Popping the cork: Mechanisms of phage genome ejection. *Nature Reviews Microbiology* 11 (3), 194–204.
- Panja, D., Molineux, I.J., 2010. Dynamics of bacteriophage genome ejection in vitro and in vivo. *Physical Biology* 7.
- Peralta, B., Gil-Carmona, D., Castaño-Díez, D., *et al.*, 2013. Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLOS Biology* 11 (9), e1001667.
- Roessner, C.A., Ihler, G.M., 1986. Formation of transmembrane channels in liposomes during injection of λ DNA. *Journal of Biological Chemistry* 261, 386–390.
- Samire, P., Serrano, B., Duche, D., *et al.*, 2020. Decoupling filamentous phage uptake and energy of the TolQRA Motor in *Escherichia coli*. *Journal of Bacteriology* 202, e00428.
- Stent, G.S., 1963. *Molecular Biology of Bacterial Viruses*. San Francisco: W. H. Freeman.
- Sun, Y., Roznowski, A.P., Tokuda, J.M., *et al.*, 2017. Structural changes of tailless bacteriophage Φ X174 during penetration of bacterial cell walls. *Proceedings of the National Academy of Sciences of the United States of America* 114, 13708–13713.
- Taylor, N.M., Prokhorov, N.S., Guerrero-Ferreira, R.C., *et al.*, 2016. Structure of the T4 baseplate and its function in triggering sheath contraction. *Nature* 533, 346–352.
- Van Valen, D., Wu, D., Chen, Y., *et al.*, 2012. A single-molecule Hershey-Chase experiment. *Current Biology* 22, 1339–1343.
- Wang, C., Tu, J., Liu, J., Molineux, I.J., 2019. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nature Microbiology* 4, 1049–1056.
- Xu, J., Gui, M., Wang, D., Xiang, Y., 2016. The bacteriophage ϕ 29 tail possesses a pore-forming loop for cell membrane penetration. *Nature* 534, 544–547.
- Xu, J., Wang, D., Gui, M., Xiang, Y., 2019. Structural assembly of the tailed bacteriophage ϕ 29. *Nature Communications* 10, 2366.
- Zárybnický, V., 1969. Mechanism of T-even DNA ejection. *Journal of Theoretical Biology* 22, 33–42.

Dealing With the Whole Head: Diversity and Function of Capsid Ejection Proteins in Tailed Phages

Lindsay W Black, The University of Maryland School of Medicine, Baltimore, MD, United States

Julie A Thomas, Rochester Institute of Technology, Rochester, NY, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Ejection or E protein Protein that is packaged within the DNA of a phage capsid and upon host infection, exits the capsid with the DNA and enters the host cell.

Prohead Spherical, protein-only precursor form of the capsid.

Introduction

Tailed phages of the *Caudovirales* are considered to be among the most genetically diverse viruses, a trait that is reflected in a diversity of virion structures, and infection mechanisms. As their name suggests, the virions of all tailed phages are comprised of a tail and a head of icosahedral, or similar, symmetry. Tails vary dramatically between the main phage taxa: Members of the *Podoviridae* have short, non-contractile tails; members of the *Siphoviridae* have long, non-contractile tails; and members of the *Myoviridae* have long, doubled, contractile tails. The prime function of the tail is the identification and initiation of infection of a suitable bacterium, usually via baseplate and fiber structures, and then delivery of the genome and capsid ejection proteins, or E proteins, into the host cell.

The major role of the head is to package, protect, and deliver both the double-stranded DNA (dsDNA) genome (~12 to ~500 kb), and, more frequently than previously realized, capsid internal proteins. Phage heads are remarkable for so successfully protecting their DNA and protein cargos, as illustrated by the virions of some phages having been shown to be stable and infectious after decades of storage. In contrast to tail structures across phage families, the icosahedral heads of all tailed phages have shared features due to the fact that the basic head structure is formed by the actions and interactions of only three structurally conserved and essential proteins; the major capsid protein (MCP), the portal protein which forms a structure for DNA entry and egress between the head and tail, and the large terminase protein which actively packages DNA into the head. These proteins' ancestries date back to a primordial virus in existence prior to the split of eukaryotes and prokaryotes.

Many copies of the MCP form a thin (~20 Å), but remarkably strong, protein outer shell. The MCP copy number varies as 60T, where the icosahedral triangulation number, or T number, increases from 1 to more than 50. Increasing capsid T numbers correlate with increased capsid dimensions necessary to house a longer genome, and in some instances, extremely large cargos of internal proteins (see below). The stability of the capsid shell can also in some cases be enhanced by MCP crosslinking or surface decoration proteins. Twelve copies of the portal protein form a turbine-like structure located at one unique capsid vertex. The portal has a number of essential functions, including that it is involved in (1) nucleation of prohead assembly and structure, (2) genome packaging into the prohead, (3) attachment of the tail to the head after the genome has been packaged, and (4) acting as a conduit for capsid internal contents as they exit the capsid into the tail upon infection of the host. The third head protein that is conserved in all tailed phages is the terminase which packages the DNA into the head through the inner channel of the portal. The terminase docks to the portal structure as a multimeric motor complex and hydrolyzes ATP to package, as well as often to measure and cut out one genome's worth of DNA from a multi-genomic, replicated and joined head-to-tail DNA concatemer. The mature genome is packed into the prohead to an extraordinarily high, liquid crystalline density (~500 mg/ml), whether there are also cargo proteins packaged before the DNA. In tailed phages, the terminase protein(s) generally dissociate from the portal after packaging and are not part of the mature virion.

Inside the MCP shell, in addition to the highly condensed genome, many phages have internal proteins. The existence of these internal head proteins illustrates beautifully the remarkable diversity that exists between different phages, as internal protein content can vary in sequence, length, copy number, location, and function in the capsid. The existence of internal proteins also highlights the fact that genome packaging occurs successfully in spite of their diversity which results in numerous final internal capsid conditions/states. However, although in every case studied these internal proteins are packaged into the prohead before the DNA, exactly how these proteins impact prohead structure and assembly, DNA packaging, head stability and DNA ejection is incompletely understood for most phages. The presence of internal proteins in different phages also raises the question of: Why are they there? Since, with the exception of a handful of model phages, the knowledge of internal head protein function(s) is limited. Based on knowledge acquired from model phages, such as T7, N4, P22, P1 and T4, there is an expectation that proteins packaged within the capsid outer shell are ejected into the host cell at the time of infection along with the DNA and have roles in host takeover. Consequently, internal head proteins in tailed phages are often referred to as ejection, or E proteins, and in this review we

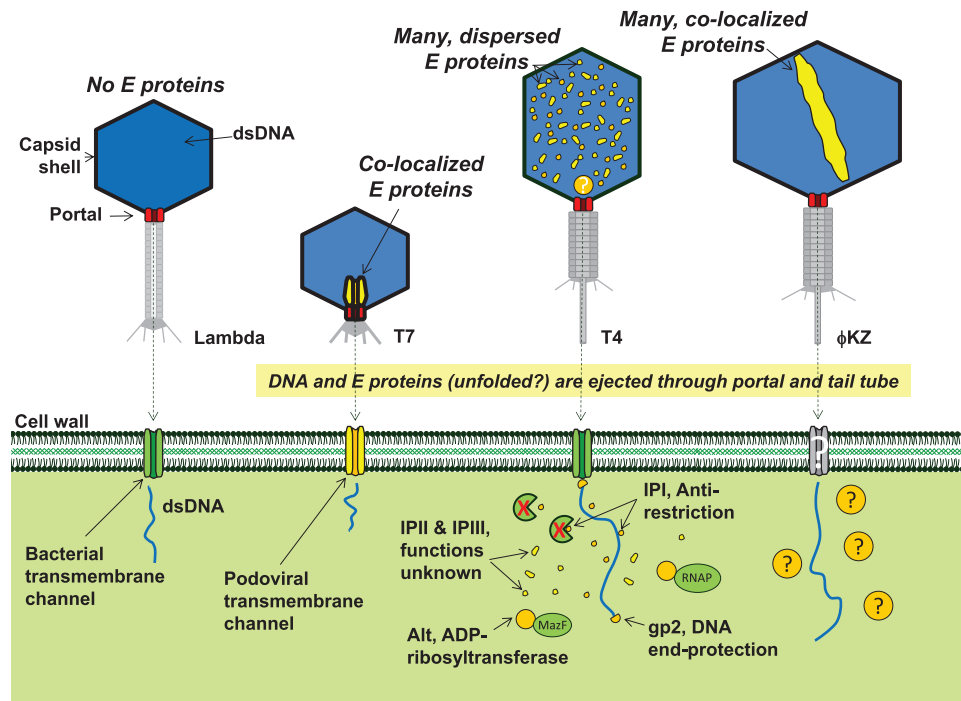


Fig. 1 Tailed phages whose capsids have Ejection, or E, proteins packaged within their DNA that have a diversity of functions (if known). All example phages infect Gram negative bacteria, but there are also phages that infect Gram positive bacteria that have E proteins.

will use both terms interchangeably, while acknowledging that the ejection status for many such proteins requires experimental confirmation.

We previously delineated tailed phage heads into four subcategories based on the amount and location of E proteins within them (see [Black and Thomas, 2012](#)). Continued research has led us to modify slightly these categories into the following: (1) Capsids that contain only packaged DNA and no, or very few, proteins within them; (2) Capsids that contain a relatively small number (<50) of E proteins that have a specific locale within the capsid; (3) Capsids that contain many (over one-thousand) E proteins embedded within the DNA in no identified order; and (4) Capsids that have a large mass of co-localized E proteins that occupy an appreciable fraction of the inner capsid volume, sometimes in a structure with a defined shape and volume (see [Fig. 1](#) for examples). We note that this categorization of E proteins is simply for ease of discussion, and has no taxonomic or functional basis. Examples of each of these categories of E proteins can be found in different phage families. Conversely, some phages, such as T4, have different classes of E proteins, with different functions, locations and/or abundances of each. All these E proteins, regardless of their structure and locales in the capsid, are able to exit the capsid through the narrow (~30 Å diameter) channel in the center of the portal complex and then must pass through the tail structure of their respective virion [sometimes a long (≥1000 Å), narrow (~30 Å diameter) structure!] for delivery to their bacterial target. That target is dependent upon the E protein's function, for some it is the host outer membrane (e.g., the podoviral core structures discussed below), whereas for others it is the host cytoplasm (e.g., the T4 internal proteins). It should also be noted that all icosahedral phage DNAs in capsids have small molecule counterions, such as polyamines and divalent cations, to neutralize the highly negatively charged DNA phosphodiester backbone; as these molecules likely have no specific biological phage function they will not be discussed as a part of this review.

In this review, we will summarize the major features and functions, where known, of examples of phages whose E proteins fall into the different categories listed above ([Fig. 1](#)). While these examples are mostly from a limited number of extremely well-characterized phages, they serve to demonstrate the eclectic nature of E proteins and their functions. In addition, we will include examples from recent research that hint at the there being a virtual smorgasbord of E proteins in the virosphere, and finish with a discussion on some of the most pressing questions raised by the existence of E proteins.

Tailed Phage Capsids With Low Numbers of E Proteins in Specific Locales

Numbers of unrelated phages have been shown to contain small numbers of E proteins (<50) within their capsids that have specific locales within the densely packaged DNA, often close to the portal structure. Some low copy number E proteins are highly organized into structures consisting of multiple proteins, whereas others may be present as just a handful, or one or two copies, of a single localized protein. Despite their generally low copy number relative to the highly abundant E proteins present in some other phage types (see below), characterization of a relatively small number of examples has illustrated a remarkable range of

functions (numbers are apparently multi-functional), from promoting phage replication (often at the transcriptional level) to the manipulation and modification of cellular structures and molecules to allow or enhance infection. Based on this, it is not surprising that many of these low abundance E proteins have been shown to be essential.

Low copy number E proteins with specific locales in the capsid

Numbers of E proteins are not thought to be part of a specific structure but nonetheless have specific locales within the capsid, typically in close proximity to the portal complex. Examples of localized internal proteins are found in *E. coli* phages N4 and T4, and *Bacillus* phages Phi29 and SPP1. Despite having similar locales, these localized inner capsid proteins have a remarkably eclectic set of functions and several of them are apparently multi-functional. One localized protein that stands out for its unusualness is the virion RNA polymerase (vRNAP) gp50 of the podovirus N4 which was among the first phage E proteins discovered (see Lenneman and Rothman-Denes, 2015). This enzyme is essential; transcribing early genes from atypical promoters containing hairpin/stem structures whose formation is assisted by host proteins (*E. coli* DNA gyrase and single-stranded DNA-binding protein). The single-subunit gp50 is surprisingly large (~380 kDa) relative to its homologs in mitochondria and other phages (e.g., the T7 gp1 RNAP, which is not an E protein, is ~99 kDa). In addition to the central domain of gp50 that is responsible for RNAP activity, the enzyme has a C-terminal domain that targets the vRNAP into the head and an N-terminal domain that is required for the ejection of the first 500 bp of the genome. Only about four copies of the vRNAP are present in each N4 virion and cryo-EM comparison of wild-type versus gp50-minus mutants indicated that the vRNAP is located at the base of an inner protein core, close above the internal entrance to the portal channel. Considering its molecular weight, it is remarkable this vRNAP can pass through the portal channel and other structures to enter the cell.

The myovirus T4 also has two low abundance E proteins, Alt and gp2, whose positions within the capsid are localized. Alt is a large protein (~75 kDa) which ADP-ribosylates host RNAP subunits, and a number of other host proteins, e.g., MazF toxin (see Alawneh *et al.*, 2016). Although Alt is nonessential, its absence from the prohead results in capsids of varying size, both smaller and larger than those of the wild-type phage, suggesting it has an additional, assembly-related function. Alt is also located close to, possibly on, the portal complex as supported by the fact that in giant elongated T4 heads formed by certain head mutations the copy number of Alt remains constant, at about 40 copies. This is despite increases in the copy numbers of other internal proteins (IPI, IPII, and IPIII, see below) proportional to increases in head size in these mutant particles.

T4 gp2 is a low copy (~1–2/head) DNA end protection protein that is required to block the *E. coli* exonuclease V (RecBCD); in *E. coli* lacking this nuclease phages lacking gp2 are fully infective. Both ends of the mature T4 DNA are located 80–90 Å apart and close to the portal, indicating gp2 must also be in this region and that it, like Alt, may have a secondary function, which for gp2 is likely a genome plug as mutations in it produce unstable heads that leak DNA. Genome plugs are proteins that function to prevent the packaged genome from escaping through the portal vertex until the appropriate time for DNA ejection or descent into the tail tube.

Podoviral E proteins

The short tails of podoviruses present a unique problem with regard the delivery of their genome to the cytoplasm of the host cell; they are too short to span the cell wall, compared to the tails of phages from other families. Podoviruses overcome this multi-layered barrier (consisting of inner and outer membranes, and periplasm) by having a set of E proteins, each typically present in relatively small copy number (<10), that are ejected ahead of the DNA and form a remarkable channel in the cell wall through which the genome passes, i.e., within their capsids podoviruses package an evertible tail of sorts! Structural evidence for the formation of this E-protein-derived channel has only been observed in two podoviruses to date (*E. coli* phage T7 and *Salmonella* phage P22) using cryo-electron tomography and exploiting minicells for phage adsorption to make visualization feasible.

In T7, the E proteins that form the eversible tail, are derived from a multi-protein structure, or core. That is centralized within the densely packaged DNA in the capsid (Fig. 2). The proteins that form the core are technically E proteins, due to their presence in the core and multi-functionality that this positioning affords them, they are usually referred to as core proteins. The T7 core is roughly cylindrical in shape; 265 Å high and 175 Å wide and positioned immediately above the portal complex. Although the core is not composed of large numbers of proteins, it does occupy a considerable volume in the 510 Å diameter capsid – large enough to be visualized by conventional metal staining electron microscopy.

The T7 core is not a simple cylinder, being composed of three domains, each with a different symmetry twelve-, eight-, and four-fold. The twelve-fold symmetry domain most probably represents the connector complex, whereas the other two domains are composed of three essential proteins: gps 14 (21 kDa), 15 (84 kDa) and 16 (144 kDa), present in ~10–12, 8 and 4 copies, respectively. Along the length of the core is an internal channel/cavity (diameter varies from 35 to 110 Å) that is continuous through to the outer side of the capsid due to the positioning of the core structure immediately above the portal complex. Regarding the genome, the T7 core is multi-functional, acting as a conduit via which the DNA enters the capsid (via the core inner channel). In addition, the core goes through an astonishing deconstruction, with the E proteins exiting the capsid via the portal, and then reassembling, into a channel that transverse the *E. coli* cell wall. This channel is formed by gp14, 15 and 16, with gp14 localizing in the outer membrane of the cell, and gp15 and gp16 localizing in the periplasm and cytoplasmic membrane. The packaged genome is then able to enter the cell via this channel whose development by-passes the need for specific phage or host structures employed by other phages to facilitate genome entry into the host cell (e.g., the long tail and LamB outer membrane porin required by Lambda phage). Even phages without a tail such as PhiX174 transfer their genome into a host cell by forming such an evertible tail.

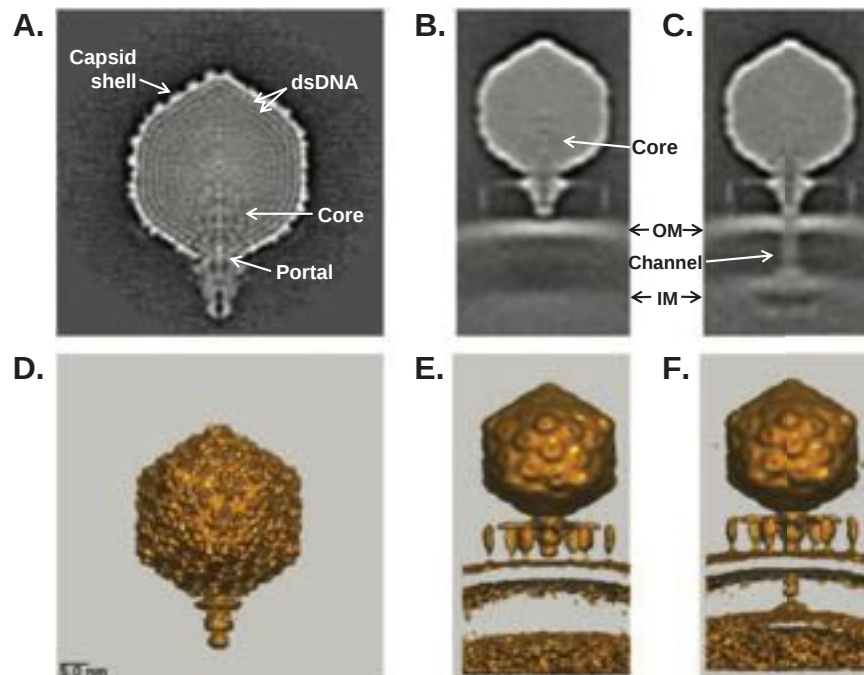


Fig. 2 Cryo-electron microscopy of T7 phage. Central slices of the T7 virion (A, B and C) and orthogonal surface views of the virion (D, E and F). Cryo-Electron Tomography of the T7 virion attached to the *E. coli* cell wall shows the transformation of its inner core (A and B) into a transmembrane channel (C and F) that facilitates genome ejection into the host cytoplasm. IM, inner membrane, OM, outer membrane. Images are reproduced from EMD-1164 (A and D), EMD-5535 (B and E), and EMD-5534 (C and F). Reproduced from Agirrezabala, X., Martin-Benito, J., Caston, J.R., *et al.*, 2005. Maturation of phage T7 involves structural modification of both shell and inner core components. The EMBO Journal 24, 3820–3829. Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. Science 339. doi:10.1126/science.1231887.

Core structures are likely conserved features of podoviruses related to T7, such as *Prochlorococcus* phage P-SSP7, as well as more distantly related phages, such as *Salmonella* phage Epsilon15 and *E. coli* phage N4. There is considerable variation in the size and dimensions of the cores among these and other phages, likely due to variations in core proteins which function to interact with the cell walls of different bacterial hosts. However, overall it is likely most cores serve an analogous function with regard genome delivery to that of the T7 core.

Phage P22 has three essential E proteins, gp7, gp16 and gp20, that also create a channel structure that completely spans from the bottom of the baseplate through *Salmonella* cell wall, a total distance of ~40 nm! The diameter of this channel varies from ~5–10 nm, and has internal cavity of ~4–8 nm through which the DNA passes. The exact functions of the P22 E proteins are unknown, but analyzes of mutants suggest that the mature form of gp7 (gp7* which is cleaved by a *Salmonella* protease) forms the extracellular component of the channel and that the majority of the channel in the periplasm is formed by gp20. In contrast to T7, the locations of the P22 E proteins inside the head are less well defined, but the fact these proteins are ejected ahead of the DNA, and in a specific order, suggests they likely have some sort of consistent locale between particles, possibly residing close to the portal complex, and represent exciting avenues of future research (see Casjens and Molineux, 2012; Hu *et al.*, 2013; Wang *et al.*, 2019; Jin *et al.*, 2015; Agirrezabala *et al.*, 2005).

Capsids With Large Numbers of E Proteins Distributed Throughout the DNA

In contrast to the low copy number E proteins with very specific locales in the capsid, some phages have E proteins that are present in many (hundreds!) of copies with no specific structural locale, of which the capsid internal proteins (IPs) of T4 are the most well studied. There are more than one thousand molecules of three small (~8 to ~20 kDa) proteins, IPI, IPII and IPIII, and several lines of evidence support that they are dispersed throughout the packaged but highly structurally mobile 500 mg/ml DNA. These include: (1) packaged IP-nuclease fusion proteins, whether present in ~200 or ~10 copies per head, are able to cleave the genome into the same capsid-sized fragments, either quickly or slowly, suggesting they are spread throughout the DNA and are mobile within it, (2) packaged IPIII-GFP fusion proteins show gradual formation of fluorescence suggesting mobility of the fusion protein within the capsid, (3) IP proteins sediment together with the genome after both have been released after treatments that disrupt the MCP outer shell, and (4) Solid state NMR has also shown that there are specific close T4 IP-lysine-to-DNA phosphate structural contacts (see Yu and Schaefer, 2008).

T4 can form mature heads without any IP proteins as discovered by genetic analyses. Mutants with knockouts in all three IP genes showed that elimination of these three proteins resulted in only minor effects on capsid volume, and also without significantly affecting the amount of DNA packaged within the normal size capsid. These studies demonstrating the T4 IPs were dispensable were initially surprising, as their high copy numbers had led researchers to infer they had a major function. A major inference from that discovery was that the IP proteins are not essential for the stability of packaged DNA. But if the IP proteins were not required for the stability of packaged DNA, this then raised the question of what was the main function of the IPs? As it was known the T4 IPs are all injected into the host cell, this led to the hypothesis that the IPs interacted with the host in a manner that would enhance infection. It took a number of years to show that IPI binds and protects T4 DNA against a highly specific DNA modification-dependent restriction enzyme (RE), which is only present in some, mostly pathogenic, strains of *E. coli*. About 350 copies of IPI are ejected before or with the DNA, and appear to function immediately to protect the DNA they accompany in the RE containing host, since phenotypic mixing shows a functional IPI gene is not required for the RE protection. A major question still to resolve is what are the targets or functions of the other two, abundant injected proteins, IPII and IPIII.

Another *E. coli* phage with E proteins that also have an anti-restriction endonuclease inhibitor role is the myovirus P1. P1 has two large (60 and 200 kDa) E proteins, Dar A and Dar B, with four other Dar accessory proteins whose functions are also to protect the ejected genome from *E. coli* REs. These proteins enter the host through the long and narrow P1 tail and likely require refolding for the anti-RE functions to emerge. Interestingly absence of the Dar proteins shows they also play a major role in P1 capsid size determination, correlating to formation of small, middle, and normal size capsids see (Piya *et al.*, 2017; Black *et al.*, 1994; Mullaney and Black, 2014).

Capsids With Large Numbers of Co-Localized E Proteins

In recent years, increasing numbers of phages have been identified as having large masses of co-localized E proteins inside their capsids. These masses are expected to be comprised of many different proteins, of which several are present in extremely high copy numbers per capsid. To date these masses have only been identified in myoviruses with capsids with high T numbers and long >200 kb genomes, so-called “giant” or “jumbo” phages [Nb., the terms “giant” and “jumbo” have no taxonomic meaning with regard phages, and are used interchangeably by researchers. These terms were first introduced into the literature in 2002 (for “giant” phage (Mesyanzhinov *et al.*, 2002)) and 2009 (for “jumbo” phage (Hendrix, 2009)). The authors prefer “giant” as the term has historically been used to describe large phage heads, which were the product of head gene mutations in T-even and other phages, e.g., (Doermann *et al.*, 1973; Ackermann *et al.*, 1975)]. These internal protein masses occupy an appreciable fraction of the inner capsid volume, sometimes in a structure with a defined shape and volume. The first of these masses was identified as being present inside the capsid of *Pseudomonas aeruginosa* phage ϕ KZ, called the inner body (IB), in 1984 by Victor Krylov's group. It was another 25 years, and ensuing developments of high resolution CryoEM imaging and proteomics via mass spectrometry, that the structure and major components of this mass were delineated by these modern techniques (Fig. 3). The ϕ KZ IB is ~90 nm in length and 35 nm in diameter – considerably larger than that of the T7 inner core with a total volume that is greater than the total volume of some smaller capsid phages. The ϕ KZ IB has a spring- or spool-like appearance and the long 280 kb genome is spooled around it. Despite a mass of ~15–20 kDa, the exact components of the IB require clarification. Proteomic analyses of the ϕ KZ head supported the IB being comprised of numbers, possibly many, different proteins, which range in abundances. Several high abundance proteins are excellent candidates for the major IB components including gps 93, 95 and 162, all of which are paralogs and members of the Pfam12699 [Pfam is a database for protein families that for each included conserved protein contains information about that protein (e.g., function if known) and an alignment of its homologous proteins, i.e., its family]. The IB proteins are apparently ejected into the host cell based on it disappearing after phage particles adsorb to bacteria, or in particles with contracted tails and DNA-free heads. Whether the IB has an inner channel and/or is anchored to the portal complex, as in podoviral cores, are yet to be determined.

That an internal mass of E proteins is not an unusual feature may be a fairly recent discovery among “giant” or “jumbo” (genomes >200 kb) environmental phages related to ϕ KZ that infect a range of Gammaproteobacteria. The isolation of these new giant phages highlights a need for further studies on the E protein content and structure within the capsids of these phages. Data to date derived from bioinformatics, proteomic, genetic and structural analyses, support that different giant phages related to ϕ KZ, such as *Salmonella* phage SPN3US, all have large masses of co-localized E proteins. Intriguingly, despite these phages all sharing a core set of head genes, including counterparts to the abundant proteins in ϕ KZ, there is likely to be great variability in the structures and E protein content between different phages. For instance, the number of Pfam12699 paralogs in each phage varies, from two in SPN3US to seven in *Pseudomonas* phage ϕ PA3. In addition the amount of E proteins in SPN3US is estimated to be >40 MDa, approximately twice the mass of the ϕ KZ IB. Since the SPN3US genome is ~40 kb shorter than that of ϕ KZ the increase in E protein content in its capsid raises the question of whether some of the E protein mass functions to maintain the density of packaged DNA in their similarly-sized capsids.

Although the function(s) of the E proteins of ϕ KZ, SPN3US and relatives, are mostly a mystery, genetic studies of SPN3US indicate that numbers of them have essential roles. Intriguingly, several minor, essential SPN3US E proteins have predicted transmembrane domains, suggesting they may interact with one of the host membranes, possibly assisting the genome and/or other E proteins to enter the cytoplasm. Other potential functions of the E proteins include assisting with the assembly of such a large capsid, possibly regulating capsid size (i.e., analogous to the shape-determining T4 core or to a tail tape measure protein),

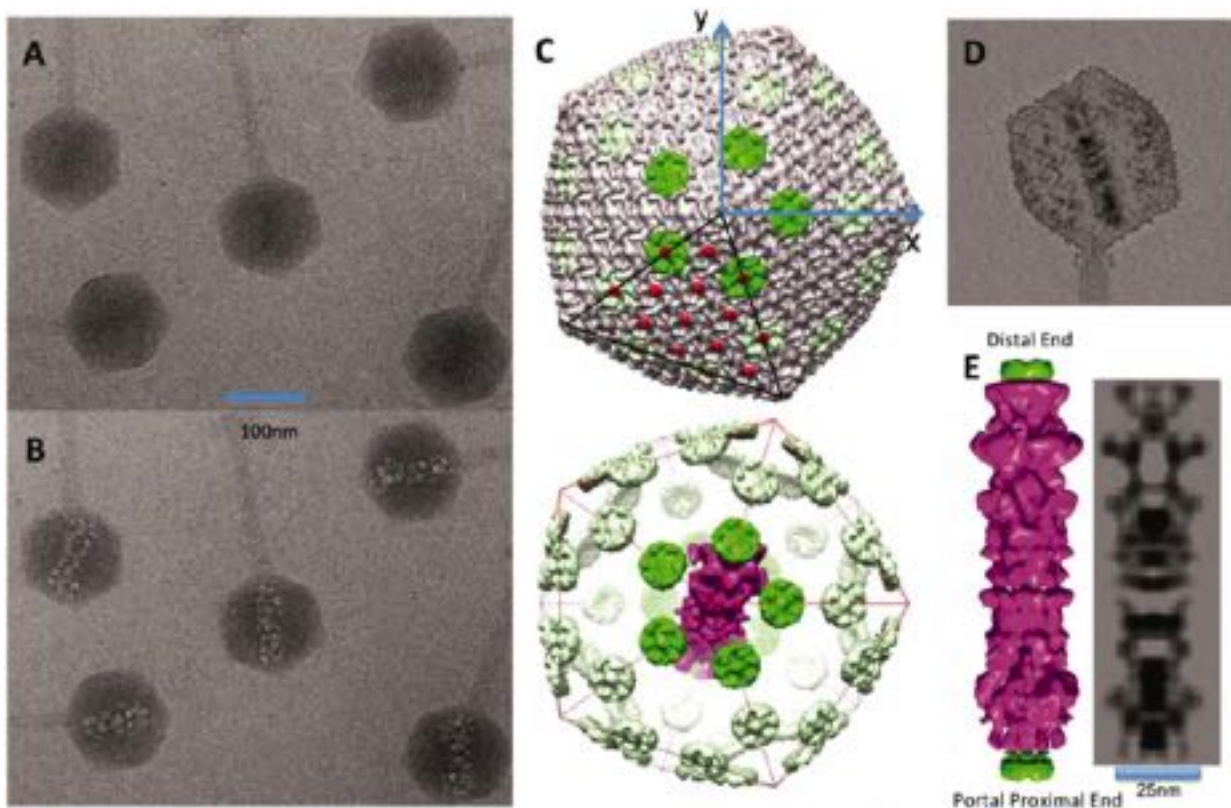


Fig. 3 Cryoelectron micrographs of purified ϕ KZ virions: (A) initial low-dose exposure; (B) subsequent exposure of the same field, with bubbling showing the locale of the E proteins in radiation-damaged virions. (C) Three-dimensional reconstruction of the ϕ KZ capsid viewed along the axis of fivefold symmetry that passes through the portal. The capsid has $T = 27$ icosahedral symmetry. The central axis of the inner body passes through the center of the capsid. (D) Central section of the ϕ KZ head sampled in the plane in which the inner body axis lies. (E) Multitiered structure of the inner body shown in surface rendering (left, magenta) and central gray-scale section (right). Reproduced with permission from Wu, W., Thomas, J.A., Cheng, N., Black, L.W., Steven, A.C., 2012. Bubblegrams reveal the inner body structure of ϕ KZ. *Science* 335, 182. doi:10.1126/science.1214120.

DNA packaging, DNA compaction in the mature head and/or roles in host takeover. The latter is supported by each of the giant phages elated to ϕ KZ and SPN3US having an essential encapsidated multi-subunit vRNAP (with diverged homology to prokaryotic RNAP β and β' , rather than the single subunit podoviral RNAPs discussed earlier) which based on studies on ϕ KZ are required for transcription of early phage genes. Interestingly, the SPN3US vRNAP is encapsidated into the DNA free prohead in an assembled form since knockout of a single one of the 5 vRNAP subunits prevents any one of this ~ 3.2 MDa enzyme from incorporation into the procapsid. Possibly, following DNA packaging by the SPN3US terminase, the vRNAP subunits are unfolded and/or disassembled, which would provide an additional function for the terminase. At very least the vRNAP subunits must be in that state by the time they exit the capsid and transverse the long, narrow SPN3US tail into the host to begin phage development. That the heads of ϕ KZ and SPN3US each have ~ 50 different proteins suggests that their E proteins collectively likely have a multitude of functions, and represent a rich resource for future studies and novel structural precedents (see Wu *et al.*, 2012; Ali *et al.*, 2017; Ceyssens *et al.*, 2014).

The Enigmas of E Proteins

Phages with yet to be identified E proteins

A major question to resolve is how many tailed phages have E proteins? It seems feasible that many phages have E proteins that are yet to be detected, or formally annotated in their genomes. This is because E protein identification is often not straightforward, requiring specialized techniques, possibly mass spectrometry, especially to identify low copy number E proteins, or Cryo-EM which is especially useful for the detection of large masses of co-localized E proteins, such as those in ϕ KZ and its relatives. Homology searches are probably the most straightforward method for inference of the presence of an E protein if there is an existing, characterized homolog in the databases, such as homologs to the core proteins of T7, or the IP proteins of T4. However, such results may be of limited use, especially for small E proteins, even ones with homologs, depending upon the extent of sequence divergence to the characterized protein.

Obviously, homology searches are of no practical use for the identification of E proteins in new phage taxa, whose genomes have a high percentage of novel genes – not an uncommon occurrence based on the overall genetic diversity of tailed phages. In such instances, there needs to be analyzes of at least one representative to assist with the inference of E protein content for related phages. As noted above, such analyzes are not insignificant, and we note that much of the existing knowledge regarding E proteins is derived from numerous and elegant studies on model phages, such as T4, T7, P22, P1, SPP1 and N4, conducted over many years. In addition, much of this knowledge is the by-product of these phages having genetic systems, which are not currently available for many phages. Despite these challenges, and based on the existing precedents, the analyzes of E proteins in novel phage types are likely to reveal many new functional classes of E proteins, possibly such as those which provide defense against prokaryotic CRISPR-Cas adaptive immune systems.

What mechanisms ensure encapsidation of E proteins?

Remarkably E proteins, in all categories discussed above, are not transferred into the procapsid after it forms, as occurs, via the portal, with the DNA. E proteins are internalized into the head at a stage when it is a protein-only prohead and only associate with the DNA after it is packaged. This knowledge is derived from characterization of the proheads in numbers of model phages, as well DNA packaging studies which have shown that proteins are excluded by the terminase from being packaged with the DNA, even proteins with high DNA binding affinities that normally bind to metabolically active cellular DNA. It can then be inferred that potentially all E proteins are incorporated into the capsid via events based on protein-protein interactions. However, it is not at all clear for many phages what the specific determinants of these, often essential, interactions are. Possibly, the mechanisms of E protein encapsidation have evolved to be as diverse as the proteins themselves.

There are several phages for which some of the determinants of E protein encapsidation are better understood, such as T4 phage (Fig. 4). In these phages their E proteins are targeted into the procapsid via attraction to proteins that are critical for prohead formation (e.g., scaffold and core proteins), and in doing so add an additional functional aspect to those proteins. The scaffold protein is typically present in hundreds of copies in the prohead – and as the name suggests – creates a structure around which the MCP shell can assemble, therefore having an important role in determining head dimensions. During head maturation the scaffold protein typically exits the capsid and by doing so creates space within the capsid for the DNA to be packaged. In phages with more complex prohead structures, such as T4, there is a main scaffold protein, gp22, and several other core proteins that are also present in many copies and work together to share the role of a single scaffold protein in other phage types. The T4 IP proteins are incorporated into DNA-free proheads based on the affinity of a short (10–19 residue), highly conserved propeptide consensus sequence found at their N-terminus (the capsid targeting sequence, or CTS) for the major essential gp22 core proteins in the procapsid.

The affinity of the T4 CTS for the core is so strong that this sequence alone can be added to foreign, non-phage proteins which results in their being packaged into the capsid. Once IP, or other targeted proteins are packaged into the core, their interaction with

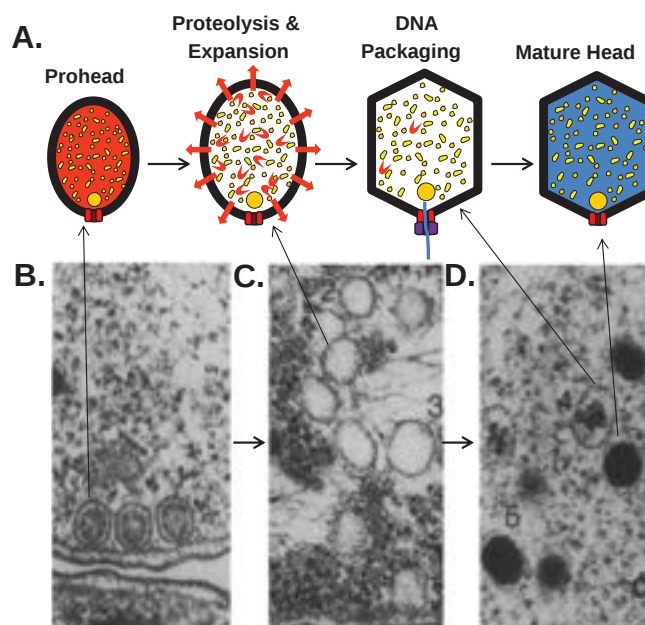


Fig. 4 Maturation of the T4 head containing densely packaged dsDNA (171 kb) and E proteins (IPI, IPII, IPIII, Alt and gp2). (A) Scheme showing T4 head maturation from a protein-only spherical core to the expanded mature DNA- and E protein-containing Iprolate, mature head. (B) Transmission electron micrographs of thin-sections of *E. coli* infected with T4 showing proheads formed on the inner membrane, (C) expanding proheads moving away from the membrane after proteolysis by the prohead protease has removed scaffold protein (gp22) and E protein and major capsid protein propeptides, and (D) proheads undergoing DNA packaging to become mature heads.

the temporary scaffold protein is no longer necessary. At this point, the prohead protease, gp21, which the bulk of the evidence suggests nucleates as a kernel at the center of prohead core, is activated to resolve these CTS-scaffold interactions. The prohead protease also cleaves an *N*-terminal propeptide from the major capsid protein and all transient core proteins into small peptides, most of which exit the capsid, although a few of which are also retained within the mature capsid.

Despite their having major differences in their capsid dimensions, structures and composition to the T4 capsid, giant phages related to SPN3US and ϕ KZ likely target many of their E proteins into the procapsid via a similar mechanism to the T4 CTS system. Support for this includes that these giant phages all have diverged homologs to the T4 major capsid protein, portal, large terminase and prohead protease proteins, suggesting they share major, anciently-derived head assembly steps. In addition proteomic analyses of three phages, SPN3US, ϕ KZ, and the *Pseudomonas chlororaphis* phage 201 ϕ 2-1 identified head proteins to have had *N*-terminal propeptides removed, with cleavage in all cases being C-terminal to a glutamate residue after a short motif comparable to the cleavage substrate for the T4 protease. Proteins identified as being cleaved in all three giant phages included the major capsid protein, portal and E proteins belonging to Pfam12699. The propeptides of many highly abundant E proteins are considerably longer than the T4 CTS, i.e., 125 and 156 residues are removed from the *N*-termini of the highly abundant E proteins, gp53 in SPN3US, and gp93 in ϕ KZ, respectively. These propeptides are longer than T4 IPI (95 residues) and IPII (100 residues) proteins! *In vitro* and *in vivo* data for several of giant phage long propeptides indicate they are cleaved at additional internal cleavage motifs. This apparent redundancy in cleavage sites is likely to ensure the propeptide regions are small enough to leave the capsid, however the need for the longer propeptides in giant phage head proteins compared to those of T4 is unknown; possibly these long propeptides ensure correct localization of the E proteins in the mature head and/or act as supplements to a scaffold protein/core.

Giant phages related to SPN3US and ϕ KZ are likely to have an equivalent to the T4 CTS despite their E protein propeptides being considerably longer than those of T4. Support for this includes that the *N*-terminal 15 residues (MANFVKSKLARESVE) of the SPN3US highly abundant E protein paralogs, gp53 and gp54 are identical and contain sequences (AXE-12, SXE-15) consistent with the known cleavage motifs for SPN3US and ϕ KZ proteases. In addition, the diverged homologous abundant E protein in ϕ KZ, gp93, is processed at SLE-13, in addition to other processing sites. There are many questions yet to be resolved regarding the role(s) of the longer E protein propeptides, however it is likely there are numbers of significant variations in giant phage head assembly mechanisms relative to those of T4, which in conjunction with the functions of their novel E protein cargos represent likely sources for future novel discoveries.

In contrast to E proteins that are targeted to scaffold/core structures, some E proteins are likely targeted into the prohead via interactions with the portal complex. This interaction is based on the locales of numbers of different proteins immediately above the portal complex, such as the N4 vRNAP and T4 Alt. This location of Alt is a yet to be resolved puzzle, as Alt also has a CTS sequence like the IP proteins, albeit slightly shorter. Possibly, Alt's *N*-terminal CTS sequence processing and our recent discovery that its C-terminus is also a CTS processed propeptide removed by gp21, suggest scaffold targeting is followed by portal targeting.

Impacts of E proteins on capsid DNA packaging and ejection

The presence of E proteins raises a number of interesting questions, particularly as to how DNA packaging occurs around or through them, particularly those E proteins which are portal-localized. Do any of these proteins have any specific interactions with packaged DNA? Or conversely, do some of these proteins have limited DNA interactions so as to prevent any hindering of their ejection into the host cell, possibly ahead of the DNA? Some E proteins, such as the N4 vRNAP, Alt, and P1 Dar proteins must impact both the prohead structure and the packaging of DNA and its final structure based on the volume they would occupy in the phage head. Other proteins with specialized locations within the head potentially have less impact on the overall structure of packaged DNA as they are present in low copy number and/or have low molecular weights. However, even their presence is likely to influence the structure of the packaged DNA in specific regions, particularly near or within the portal complex, especially proteins that are bound to genome termini, such as T4 gp2.

Impacts of E proteins on DNA ejection

The successful translocation of the phage genome through the tail and into the bacterial cell is critical for infection; however, with the exception of the T7 and podoviral E protein-to-channel transformations which facilitate DNA ejection (discussed above), the impacts of E proteins on this process are poorly understood. This fundamental gap in knowledge exists in part because of the difficulty of defining these processes experimentally, which is further complicated by the number of variables in play during this process. These include the characteristics of the packaged DNA itself (length, packaged density and structure), and the structures of the tail and cell wall channel employed by the phage (whether it be bacterial or phage encoded). Notably, during the ejection process the DNA must undergo a structural transition from liquid-crystalline-like to fluid-like state, but even the speed of this transition must vary between different capsids.

Different E proteins themselves likely have different impacts on DNA ejection, simply due to dramatic variation in types, numbers, and structures within the head. For instance, the 10 kDa IPI of T4 is apparently ejected through the portal and tail tube with the DNA and without unfolding thereby promoting immediate inhibition of its nuclease target in the infected host. This does not seem feasible for other E proteins, particularly larger ones, such as the N4 vRNAP, and P1 Dar proteins, which based on their dimensions versus the internal diameters of the portal channel and tail, must undergo unfolding in the ejections process and refolding in the bacterial cytoplasm. That such permutations can occur was demonstrated by the packaging of β -galactosidase into the T4 head via an IPIII- β -galactosidase fusion and its subsequent ejection into, and activity within, the host. Physically it is not feasible that packaged β -galactosidase could exit the capsid in its folded state especially considering the structure of the myoviral

tail through which it passes. The double tube contractile tail structure (typically ~100 nm long, ~0.3 nm inner diameter) would seemingly make tail expansion to allow ejection of bulky E proteins unlikely. In addition, the appearance of the 540 kDa tetramer-dependent activity in the infected cell requires host chaperones. Hence, the ejection of the unfolded, or possibly semi-unfolded, protein must be followed by refolding and multimerization within the host. Possibly, unfolding of such non-native and native E proteins, such as β -galactosidase and Alt, that are folded and active upon synthesis prior to encapsidation, is a secondary function of the high force-generating packaging motor and the resulting DNA pressure in the full capsid. A further intriguing feature of E proteins is whether or not they resist CRISPR-Cas nuclease attack. A number of small acr proteins (anti-CRISPR proteins) comparable in size to the T4 IPs have been described that counter CRISPR-Cas attack, but thus far these proteins have been found only to be synthesized by phage origin genes in the host. Thus they are thought to function in an altruistic or cooperative fashion, unlike T4IPI that is an egoistic anti-RE protein. Could this be related to the apparent lysogeny of the acr proteins versus lytic phage proteins such as IPI. Or do E-acr's still await discovery? (see [Nussenzweig and Marraffini, 2018](#)).

Potential of E protein delivery from capsid derived nanocontainers

A related, intriguing feature of E proteins to be resolved is the mechanisms by which large proteins such as Alt, Dar, N4, giant multi-subunit phage vRNAPs, and even the 450 kDa non-phage protein β -galactosidase, are apparently unfolded within the capsid, probably while the DNA is being packaged, permitting them to be in a conformation that allows them to be ejected through the narrow portal and tail tube channels. Could this unfolding be a secondary but essential function of the high energy DNA packaging motor mechanism that, through high DNA pressure within the fully packaged capsid, promotes the ejection of unfolded or partially unfolded large proteins. Evidently these E proteins are then refolded upon ejection to display often essential enzymatic functions within the host cell? Ejection and refolding of β -galactosidase in an *E. coli* host cell argues against the necessity of special E protein structural evolution to permit transfer into a host. This high flexibility of E proteins potentially provides a means of transferring a variety of enzymes, singly or together, (e.g., Cre recombinase, CRISPR targeted Cas9, Lambda beta and exonuclease recombination proteins etc., see ([Liu et al., 2014](#))) into a host cell for gene repair or gene knockout.

Conclusions

In addition to the conserved features of all tailed phage heads – the portal, major capsid shell and DNA packaging enzyme(s) – required to form an icosahedral capsid full of densely packaged DNA, there are clearly many variations on the internal E protein composition between different phages. E proteins vary significantly in number, and locales within the head, some dispersed throughout the DNA, others form distinct structures. The reasons for these variations are in many cases still obscure but are likely linked to their functions which, based on existing precedents, potentially for most E proteins, are linked to the infection of and replication within a new bacterial host. In addition, whether fully packaged ~500 mg/ml DNA in different E protein-containing heads is spooled or folded is uncertain and could vary among these viruses.

As our knowledge of internal head or E proteins is from the phages isolated and studied to date, which represent a relatively narrow pool of total phage diversity, it is clear and exciting to consider that a myriad of internal proteins composition, structures and functions are yet to be discovered. Based on numbers of known E proteins having roles in infection and the potential of phages for biocontrol applications in a multi-drug resistant era, it could be proposed that phage E protein research is a relatively unexplored area that merits concentrated effort.

References

- Ackermann, H.W., Caprioli, T., Kasatiya, S.S., 1975. A large new *Streptococcus* bacteriophage. *Canadian Journal of Microbiology* 21, 571–574.
- Agirrezabala, X., Martin-Benito, J., Caston, J.R., et al., 2005. Maturation of phage T7 involves structural modification of both shell and inner core components. *The EMBO Journal* 24, 3820–3829.
- Alawneh, A.M., Qi, D., Yonesaki, T., Otsuka, Y., 2016. An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the *Escherichia coli* MazF toxin of a toxin–antitoxin module. *Molecular Microbiology* 99, 188–198.
- Ali, B., Desmond, M.I., Mallory, S.A., et al., 2017. To be or not to be T4: Evidence of a complex evolutionary pathway of head structure and assembly in giant *Salmonella* virus SPN3US. *Frontiers in Microbiology* 8, 2251.
- Black, L.W., Showe, M.K., Steven, A.C., 1994. Morphogenesis of the T4 head. In: Karam, J.D. (Ed.), *Molecular Biology of Bacteriophage*. Washington, DC: ASM Press, p. T4.
- Black, L.W., Thomas, J.A., 2012. Condensed genome structure. In: Rossmann, M.G., RAO, V.B. (Eds.), *Viral Molecular Machines*. US: Springer.
- Casjens, S.R., Molineux, I.J., 2012. Short noncontractile tail machines: Adsorption and DNA delivery by podoviruses. *Advances in Experimental Medicine and Biology* 726, 143–179.
- Ceyssens, P.-J., Minakhin, L., Van Den Bossche, A., et al., 2014. Development of giant bacteriophage ϕ KZ is independent of the host transcription apparatus. *Journal of Virology* 88, 10501–10510.
- Doermann, A.H., Eislerling, F.A., Boehner, L., 1973. Genetic control of capsid length in bacteriophage T4. I. Isolation and preliminary description of four new mutants. *Journal of Virology* 12, 374–385.
- Hendrix, R.W., 2009. Jumbo bacteriophages. In: Van Etten, J. (Ed.), *Lesser Known Large dsDNA Viruses*. Berlin Heidelberg: Springer.
- Hu, B., Margolin, W., Molineux, I.J., Liu, J., 2013. The bacteriophage T7 virion undergoes extensive structural remodeling during infection. *Science* 339. doi:10.1126/science.1231887.
- Jin, Y., Sdao, S.M., Dover, J.A., et al., 2015. Bacteriophage P22 ejects all of its internal proteins before its genome. *Virology* 485, 128–134.
- Lennehan, B.R., Rothman-Denes, L.B., 2015. Structural and biochemical investigation of bacteriophage N4-encoded RNA polymerases. *Biomolecules* 5, 647–667.

- Liu, J.L., Dixit, A.B., Robertson, K.L., Qiao, E., Black, L.W., 2014. Viral nanoparticle-encapsidated enzyme and restructured DNA for cell delivery and gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 111, 13319–13324.
- Mesyanzhinov, V.V., Robben, J., Grymonprez, B., *et al.*, 2002. The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. *Journal of Molecular Biology* 317, 1–19.
- Mullaney, J.M., Black, L.W., 2014. Bacteriophage T4 capsid packaging and unpackaging of DNA and proteins. *Methods in Molecular Biology* 1108, 69–85.
- Nussenzweig, P.M., Marraffini, L.A., 2018. Viral teamwork pushes CRISPR to the breaking point. *Cell* 174, 772–774.
- Piya, D., Vara, L., Russell, W.K., Young, R., Gill, J.J., 2017. The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis. *Molecular Microbiology* 105 (3), 399–412.
- Wang, C., Tu, J., Liu, J., Molineux, I.J., 2019. Structural dynamics of bacteriophage P22 infection initiation revealed by cryo-electron tomography. *Nature Microbiology* 4, 1049–1056.
- Wu, W., Thomas, J.A., Cheng, N., Black, L.W., Steven, A.C., 2012. Bubblegrams reveal the inner body structure of ϕ KZ. *Science* 335, 182.
- Yu, T.Y., Schaefer, J., 2008. REDOR NMR characterization of DNA packaging in bacteriophage T4. *Journal of Molecular Biology* 382, 1031–1042.

Jumbo Phages

Isaac T Younker and Carol Duffy, University of Alabama, Tuscaloosa, AL, United States

© 2021 Published by Elsevier Ltd.

Nomenclature

bp Base pairs

kb Kilobase pairs

nm Nanometers

Glossary

CRISPR-Cas system Prokaryotic acquired immune system that confers resistance to foreign genetic elements and phages via degradation of recognized nucleic acid sequences.

ORFan gene Predicted open reading frame without detectable homology to coding sequences in the genomes of other organisms or viruses.

Restriction-modification system Prokaryotic innate immune system that confers resistance to foreign genetic

elements and phages via degradation of invading DNA while precluding degradation of host DNA.

Terminal redundancy Repeated genetic sequence present at both ends of a linear DNA molecule.

Triangulation number (T-number) The number of unique environments occupied by icosahedral capsid subunits; generally representative of the size and complexity of the capsid.

Introduction

Bacteriophages are the most numerous biological entities on the planet. Outnumbering their bacterial hosts by a factor of ten, there are an estimated 10^{31} phages on Earth. Ubiquitous in nature, phages are found in soil, composts, freshwater and marine environments, plants and animals, and even in the free atmosphere. With their varying lifestyles and prominent contributions to horizontal gene transfer, phages are the driving force behind bacterial populations. As such, they fundamentally influence biogeochemical cycles, the composition of microbial communities, bacterial evolution, and plant and animal health. Phages vary greatly in their morphology, virion size, genome composition and length, and have been grouped by various systems. Grouping by genome size, phages with genomes over 500 kb in length have been collectively designated ‘megaphages’ whereas phages with genomes 200–500 kb were originally referred to as ‘giant phages’ and later coined ‘jumbo phages’. Whereas all megaphages discovered to date were identified via analysis of viral metagenomic data and have not been physically isolated, an ever-growing list of jumbo phages have been isolated, characterized, and sequenced. Increased interest in jumbo phage biology has occurred hand-in-hand with a resurgence in both basic and applied phage research. These giants of the phage world offer unique insights into virus evolution, present novel genetic and molecular features, and can be especially useful in phage therapy and biocontrol applications. In this contribution we present a brief overview of jumbo phage history, discuss the challenges and methods used in the isolation and characterization of these oversized viruses and highlight a variety of unique genomic features and advanced capabilities identified in some of the 160+ jumbo phages isolated to date.

History

The story of jumbo phage biology begins, from a research standpoint, in 1968 when Gianfranco Donelli reported the isolation and initial characterization of a *Bacillus megaterium* bacteriophage named phage G. Long before the modern convenience of next generation sequencing technologies, Donelli and his colleagues recognized phage G was incredibly large, estimating the molecular weight of its double-stranded DNA (dsDNA) genome at 4.9×10^8 daltons. Phage G’s icosahedral head is 160 nm in diameter and its tail is 453 nm long. In 2011, the genome of phage G was sequenced and found to be 497,513 bp. To date, phage G holds the record for the largest isolated phage based on both virion size and genome length.

Another early discovery in jumbo phage history, the *Caulobacter crescentus* phage phiCbK was first reported by Nina Agabian-Keshishian and Lucille Shapiro in 1970. Originally characterized with head dimensions of 64 nm × 195 nm and a tail 275 nm long, phiCbK virions were more recently measured at 56 nm × 205 nm for the head and 300 nm for the tail length. The phiCbK dsDNA genome was sequenced in 2012 and found to be 215,710 bp. Due to its large head size, phiCbK helped advance the field of structural biology by serving as one of the first phages for which fine capsid structure was determined by electron microscopic image reconstruction. PhiCbK also aided in studies of the unique developmental program of *Caulobacter*, serving as a cell cycle and morphological indicator through its use of swarmer cells’ flagella and polar pili for adsorption.

The *Pseudomonas aeruginosa* phage phiKZ, originally described by Krylov and Zhazykov in 1978, is one of the most thoroughly studied jumbo phages and claims several firsts in jumbo phage history. In 2001, phiKZ became the first jumbo phage to have its

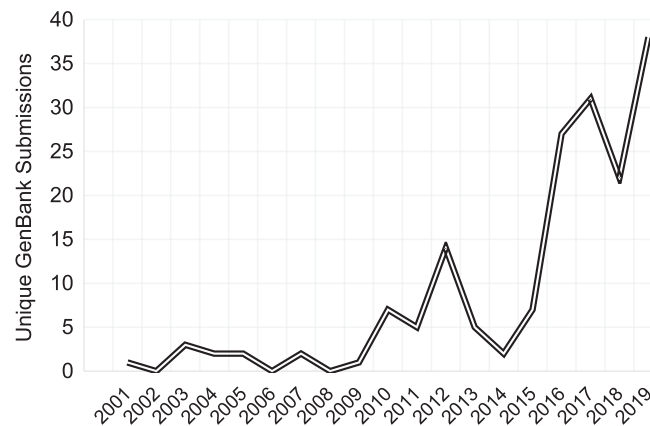


Fig. 1 GenBank jumbo phage genome submissions. Number of unique jumbo phage genome sequences submitted to GenBank each year from 2001 through 2019.

genome completely sequenced and entered in the National Center for Biotechnology Information's GenBank. PhiKZ was the first jumbo phage found to possess a proteinaceous spool-like structure, termed the inner body, inside its head and the first for which the capsid structure was resolved by cryo-electron microscopy (cryo-EM). In addition, although *Pseudomonas chlororaphis* jumbo phage 201phi2-1 was the first phage determined to form a nucleus-like compartment in infected cells, phiKZ has led the way in studies aimed at determining the assembly and function of this fascinating structure.

Changes in isolation methods and advancements in sequencing technologies have greatly eased the discovery of new jumbo phages. Thus, GenBank has seen a dramatic increase in the number of jumbo phage genomes submitted over the past few years. Whereas 0–3 new jumbo phage genomes were submitted to GenBank each year from 2001 to 2009, the past decade shows an unmistakable growth trend in submissions with a record 38 genome sequences from new and unique jumbo phages submitted in 2019 (Fig. 1). All jumbo phages isolated and characterized to date possess linear dsDNA genomes and belong to the *Myoviridae* and *Siphoviridae* families of the order *Caudovirales*. As more of these giant viruses are discovered, only time will tell whether jumbo phages exist that belong to other taxonomic families.

This is an exciting time for phage biology; the accessibility and growing abundance of genomic sequence data allows questions to be asked and comparisons to be made that were previously impossible. The growing number of new phages reported each year has also necessitated a standardization and formalization of the naming process. The history of phage biology includes multiple instances of duplicate names. For example, prior to isolation of the jumbo phage G by Donelli in 1968 a much smaller phage G was reported by James Murphy in 1957. The current naming system, proposed in 2017, aims to avoid this confusion by incorporating the common name into a longer formal name for each newly discovered virus. An example of a formal phage name using this system is vB_DsoM_JA11. This naming scheme tells us this is a virus of *Bacteria* isolated on *Dickeya solani* that belongs to the *Myoviridae* family. The last portion of the name, in this example JA11, is the unique common name given by the researcher(s) who isolated and reported the phage. In this work we have introduced phages with their formal names when available. However, most of the jumbo phages discussed herein were first reported prior to the proposal of this naming scheme and thus are referred to by their published common names. To aid the reader in further exploration of the topic, we have also indicated the host bacterium for all phages discussed.

Isolation and Characterization

Although jumbo phages represent a minority of the total collection of bacteriophages isolated in the past 100 years, they are similarly ubiquitous in nature. Jumbo phages have been successfully isolated and propagated from soil, freshwater and marine environments, sewage, and fecal samples from around the globe. The majority of jumbo phages isolated to date infect gram-negative bacteria including species from *Achromobacter*, *Acinetobacter*, *Aeromonas*, *Burkholderia*, *Caulobacter*, *Cronobacter*, *Dickeya*, *Erwinia*, *Escherichia*, *Klebsiella*, *Pectobacterium*, *Prevotella*, *Prochlorococcus*, *Pseudomonas*, *Ralstonia*, *Salicola*, *Salmonella*, *Serratia*, *Sinorhizobium*, *Synechococcus*, *Tenacibaculum*, *Vibrio*, *Xanthomonas*, and *Yersinia* (Table 1). Only six jumbo phages infecting gram-positive bacteria have been isolated, all of which infect *Bacillus* species. It is unknown whether this phenomenon simply reflects a bias in isolation efforts or if *Bacillus* species possess features unique among the gram-positive *Bacteria* that make them amenable to hosting very large phages.

The most common phage isolation technique is the double-agar overlay method in which a filtered environmental sample is mixed with a liquid bacterial culture and molten agar and then poured onto a solid agar plate. Historically, a 0.7%–0.8% molten agar solution has been used resulting in a final top agar concentration of ~0.5%. As phages infect, lyse, and diffuse through the top agar-bacteria mixture, visible plaques appear as holes in the bacterial lawn and can be picked for further propagation. With this method, diffusivity, adsorption rate and burst size can all contribute to plaque size. Thus, Philip Serwer and colleagues reasoned the great plaque count anomaly (the small number of unique phages successfully isolated relative to the large number observed by electron microscopy) could be due, at least in part, to top agar concentrations that preclude diffusion of very large phages. By using

Table 1 Jumbo phages and bacterial hosts. Jumbo phage genome sequences submitted to GenBank as of December 31st, 2019. Phages are sorted by host name and by genome size within each host. Information regarding hosts and isolation materials was obtained from GenBank and/or published articles

Host	Genome length (bp)	Phage name	Isolation source	Accession number
<i>Achromobacter animicus</i>	221,431	Motura	Soil	MN094788.1
<i>Acinetobacter baumannii</i>	234,900	vB_AbaM_ME3	Wastewater effluent	NC_041884.1
<i>Aeromonas hydrophila</i>	262,021	D3	Sewage water	MN102098.1
	260,310	LAh10	Not provided	MK838116.1
	259,833	D6	Sewage water	MN131137.1
	240,447	PS2	River water	MN453779.2
	238,150	CF8	Aquaculture pond water	MK774614.2
	237,367	PS1	River water	MN032614.2
	233,234	Aeh1	Sewage	NC_005260.1
	231,743	CC2	Sewage	NC_019538.1
	221,116	Ah1	Wastewater	MG250483.1
<i>Aeromonas salmonicida</i>	236,567	65.2	Not provided	KY290955.1
	235,229	65	Not provided	NC_015251.1
	229,957	As-szw	River water	MF498773.1
	229,929	AS-zj	River water	MF448340.1
	225,268	PhiAS5	River water	NC_014636.1
	222,006	PX29	Not provided	NC_023688.1
<i>Agrobacterium tumefaciens</i>	490,380	Atu_ph07	Soil sample	NC_042013.1
<i>Bacillus megaterium</i>	497,513	Phage G	Not provided	NC_023719.1
<i>Bacillus pumilus</i>	255,569	vB_BpuM_BpSp	Not provided	KT895374.1
<i>Bacillus subtilis</i>	252,197	PBS1	Not provided	MF360957.1
	251,042	AR9	Not provided	NC_031039.1
	221,908	SP-15	Soil	KT624200.1
<i>Bacillus thuringiensis</i>	218,948	0305phi8-36	Soil	NC_009760.1
<i>Burkholderia cenocepacia</i>	262,652	BcepSauron	Soil	MK552141.1
<i>Caulobacter crescentus</i>	279,967	CcrColossus	Surface water	NC_019406.1
	229,319	Ccr29	Pond/stream	KY555145.1
	223,720	CcrRogue	Surface water	NC_019408.1
	221,828	CcrKarma	Surface water	NC_019410.1
	220,299	Ccr2	Pond/stream	KY555143.1
	219,348	Ccr10	Pond/stream	KY555142.1
	219,216	CcrSwift	Surface water	NC_019411.1
	218,929	CcrMagneto	Surface water	NC_019407.1
	218,729	Ccr5	Pond/stream	KY555144.1
	216,240	Ccr34	Pond/stream	KY555147.1
	215,779	Ccr32	Pond/stream	KY555146.1
	205,504	phiCbK	Not provided	JX163858.1
<i>Caulobacter vibroides</i>	322,272	CcrBL9	Not provided	MH588546.1
	317,488	CcrSC	Not provided	MH588547.1
	308,141	CcrPW	Not provided	MH588545.1
	220,934	CcrBL10	Not provided	MH588544.1
<i>Cronobacter sakazakii</i>	358,663	vB_CsaM_GAP32	Wastewater	JN882285.1
	223,989	CR5	Farm soil	JX094500.1
<i>Dickeya solani</i>	261,165	vB_DsoM_AD1	River water	MH460463.1
	255,356	vB_DsoM_JA11	River water	MH389777.1
	255,356	vB_DsoM_JA33	River water	MH460462.1
	254,061	vB_DsoM_JA13	River water	MH460460.1
	253,323	vB_DsoM_JA29	River water	MH460461.1
<i>Erwinia amylovora</i>	275,000	vB_EamM_Madmel	Soil	MG655269.1
	273,914	vB_EamM_Mortimer	Soil/foilage near infected tree	MG655270.1
	273,731	Rebecca	Tree	MK514281.1
	273,501	vB_EamM_Deimos-Minion	Branches, blossoms	NC_041972.1
	273,224	vB_EamM_Special G	Branches, blossoms	NC_041975.1
	272,458	vB_EamM_Desertfox	Soil	NC_042098.1
	272,228	vB_EamM_Bosolaphorus	Orchard dirt	MG655267.1
	271,182	vB_EamM_RAY	Leaves, stem	NC_041973.1
	271,088	vB_EamM_Simmy50	Bark	NC_041974.1
	271,084	Ea35-70	Pear tree soil	NC_023557.1
	266,532	vB_EamM_Alexandra	Soil	MH248138.1
	261,365	vB_EamM_Y3	Soil	KY984068.1
	259,700	vB_EamM_Yoloswag	Soil/foilage near infected tree	KY448244.1

(Continued)

Table 1 Continued

Host	Genome length (bp)	Phage name	Isolation source	Accession number
246,390	vB_EamM_Kwan	Soil/foilage near infected tree	NC_031010.1	
246,290	vB_EamM_Asesino	Soil	KX397364.2	
244,950	Wellington	Not provided	MH426724.1	
244,840	vB_EamM_ChrisDB	Soil/foilage near infected tree	KX397366	
243,953	vB_EamM_Stratton	Soil/foilage near infected tree	KX397373	
243,050	phiEaH2	Soil	NC_019929.1	
241,654	vB_EamM_Machina	Soil/foilage near infected tree	NC_042056.1	
241,147	vB_EamM_Caitlin	Soil/foilage near infected tree	KX397365	
241,050	vB_EamM_Parshik	Soil/foilage near infected tree	KX397371	
240,761	vB_EamM_Huxley	Soil/foilage near infected tree	KX397368	
235,374	vB_EamM_Joad	Apple Blossoms	MF459647.1	
235,108	vB_EamM_Rising Sun	Apple Blossoms	NC_042018.1	
229,501	vB_EamM_Phobos	Soil/foilage near infected tree	KX397372	
223,950	Derbicus	Pear tree	MK514282.1	
223,935		vB_EamM_EarlPhillipIV	Soil/foilage near infected tree	NC_031007.1
218,339	PhiEaH1	Soil near infected apple tree	NC_023610.1	
<i>Escherichia coli</i>	386,442	CMSTMSU	Shrimp farm wastewater	MH494197.1
	370,817	vB_EcoM_G17	Pig manure	MK327931.1
	353,081	UB	Orangutan feces	MH383160.1
	352,598	vB_EcoM_phAPEC6	Not provided	MK817115.1
	348,532	121Q	Sewage	KM507819.1
	348,113	PBECO4	Sewage	NC_027364.1
	348,043	vB_Eco_slurp01	Porcine feces	LT603033.1
	347,152	SP27	Not provided	LC494302.1
	237,307	vB_EcoM_Goslar	Duck feces	MK327938.1
<i>Klebsiella</i> ^a	345,809	vB_KleM-Rak2	Sewage polluted pond	JQ513383.1
<i>Klebsiella pneumoniae</i>	346,602	K64-1	Wastewater	NC_027399.1
<i>Pectobacterium carotovorum</i>	378,379	CBB	Wastewater	NC_041878.1
<i>Photobacterium damsela</i>	237,509	PDCC-1	Lollipop catshark	MN562221.1
<i>Prochlorococcus marinus</i>	252,013	P-SSM5	Subtropical open ocean	HQ632825.1
<i>Prochlorococcus</i> sp. NATL1A	252,401	P-SSM2	Sargasso Sea	NC_006883.2
<i>Pseudomonas aeruginosa</i>	309,208	PhiPA3	Sewage	NC_028999.1
	304,671	PA1C	Not provided	MK599315.1
	301,543	vB_PaeM_PS119XW	Not provided	MN103543.1
	288,170	vB_PaeM_MIJ3	Manure runoff	LR588166.1
	286,783	vB_PaeM_PA5oct	Sewage near farm	MK797984.1
	280,334	phiKZ	Sewage	NC_004629.1
	279,696	SL2	Hospital sewage	NC_042081.1
	279,593	KTN4	Sewage from irrigation field	KU521356.1
	279,095	PA02	Sewage	AP019418.1
	258,139	PaBG	Lake water	KF147891.1
	211,215	EL	Pond water	NC_007623.1
<i>Pseudomonas chlororaphis</i>	316,674	201phi2-1	Soil	NC_010821.1
<i>Pseudomonas fluorescens</i>	309,157	Phabio	Compost	MF042360.1
	284,757	OBP	Compost	NC_016571.1
	278,136	Noxifer	Compost	NC_041994.1
<i>Pseudomonas putida</i>	280,538	Lu11	Mango tree soil	NC_017972.1
<i>Pseudomonas syringae</i>	305,260	Psa21	Leaf litter in kiwi orchard	MK552327.1
<i>Ralstonia solanacearum</i>	279,845	RP12	Tomato soil	NC_041911.1
	276,958	RP31	Tomato soil	AP017925.1
	231,255	RSL1	Crop soil	AB366653.2
	223,932	RSL2	Soil	AP014693.1
	222,888	RSF1	Soil	NC_028899.1

Table 1 Continued

Host	Genome length (bp)	Phage name	Isolation source	Accession number
<i>Salicola</i> sp. PV4	440,001	SCTP-2	Solar saltern	MF360958.1
<i>Salmonella</i> ^a	350,103	Munch	Cattle feedlot	MK268344.1
<i>Salmonella enterica</i>	348,718	7t3	Wastewater	MK773491.1
	250,739	SPAsTU	Sewage water	MH221129.1
	240,413	SPN3US	Chicken fecal sample	JN641803.1
	239,461	SEGD1	Not provided	KU726251.1
<i>Salmonella typhimurium</i>	242,624	SPFM1 ^b	Not provided	LR535901.1
	241,405	SPFM13 ^b	Not provided	LR535910.1
	240,198	SPFM6 ^b	Not provided	LR535905.1
<i>Serratia</i> ^a	273,933	Moabite	Mixed swine farm sample	MK994515.1
	212,807	PCH45	Sewage	MN334766.1
<i>Serratia marcescens</i>	357,154	BF	Compost	NC_041917.1
	276,025	2050HW	Not provided	MF285618.1
<i>Sinorhizobium meliloti</i>	206,713	phiN3	Not provided	KR052482.1
<i>Sphingomonas paucimobilis</i>	219,372	PAU	Silkworms	NC_019521.1
<i>Synechococcus</i> ^a	244,930	B3	Sub-arctic lake	MN695334.1
	243,633	B23	Sub-arctic lake	MN695335.1
	232,883	S-SSM6a	Sargasso Sea	HQ317391.1
	232,878	S-SSM7	Sargasso Sea	NC_015287.1
	222,619	ACG-2014f Syn7803C11 ^c	Not provided	KJ019142.1
	216,121	S-CAM7 0910CC49 ^c	Pacific Ocean	NC_031927.1
	213,993	S-B43	Bohai Sea	MN018232.1
	208,857	S-B05	Bohai Sea	MK799832.1
	208,007	S-SKS1	North Atlantic Ocean	HQ633071.1
	204,930	Bellamy	Seawater	MF351863.1
<i>Tenacibaculum discolor</i>	234,670	pT24	Shrimp culture pond water	LC168164.1
<i>Tenacibaculum maritimum</i>	226,876	Ptm5	Fish aquaculture seawater	AP019525.1
	224,680	Ptm1	Fish aquaculture seawater	AP019524.1
<i>Vibrio alginolyticus</i>	250,485	phi-ST2	Coastal seawater	KT919973.1
	248,605	phi-Gm1	Coastal seawater	KT919972.1
	248,088	ValKK3	Marine sediment	NC_028829.1
	247,619	ValB1 HC	Not provided	MK568540.1
	242,446	2 TSL-2019	Not provided	MK368614.1
	238,099	USC-1	Marine water	MK905543.1
	238,053	5	Not provided	MK358448.1
	237,722	Aphrodite1	Not provided	NC_042100.1
	231,998	pVa-21	Not provided	KY499642.1
<i>Vibrio coralliilyticus</i>	288,967	BONAISHI	Sea water near coral reef	MH595538.1
<i>Vibrio mediterranei</i>	278,116	vB_VmeM-Yong MS32 ^b	Not provided	MK308677.1
	290,532	vB_VmeM-Yong XC31 ^b	Not provided	MK308674.1
<i>Vibrio natriegens</i>	247,511	nt-1	Salt marsh mud samples	NC_021529.2
<i>Vibrio parahaemolyticus</i>	246,421	phi-pp2	Aquaculture waterways	JN849462
	244,834	KVP40	Polluted sea water	NC_005083.2
	239,276	pTD1	Not provided	NC_041916.1
<i>Vibrio strain 7D</i>	246,964	VH7D	Abalone farm seawater	NC_023568.1
<i>Xanthomonas citri</i>	384,670	XacN1	Orange grove soil	AP018399.1
<i>Yersinia</i> ^a	352,596	fHe-Yen9-03	Sewage	LT960552.1
<i>Yersinia enterocolitica</i>	262,391	PhiR1-37	Sewage	NC_016163.1
Not provided	257,877	NCTB	Surface water	LT598654.1

^aHost genus but not species was provided.^bPhage listed is one of multiple phage genome sequences submitted to GenBank with >98% sequence identity.^cPhage listed is one of multiple isolates of an individual phage that have been sequenced and submitted to GenBank.

top agar concentrations as low as 0.15% they successfully isolated two new jumbo phages, *P. chlororaphis* phage 201phi2-1 and *Bacillus thuringiensis* phage 0305phi8-36, and opened the door to an era of jumbo phage discovery. The use of dilute top agar solutions, typically 0.1%–0.5%, is now widely practiced by jumbo phage biologists. For example, *Agrobacterium tumefaciens* phage Atu_ph07, which has the second largest genome of all jumbo phages sequenced to date (490 kb), was initially isolated on 0.3% top agar and later propagated on 0.15% top agar to promote the development of larger, clearer plaques. Additional modifications to standard methods also aid the isolation and propagation of jumbo phages. Double-agar plates may be incubated at a temperature lower than the host's optimal growth temperature to prevent bacterial growth from outpacing plaque formation.

Additionally, debris may be removed from environmental samples prior to plating by differential centrifugation rather than filtration to avoid inadvertent removal of very large phages.

While sequencing is the only method to precisely determine the genome length, determining the packaged DNA lengths of jumbo phages is also of interest. Packaged DNAs purified from jumbo phage virions are typically longer than the corresponding genome lengths due to genomic terminal redundancies that vary from phage to phage and can be quite extensive. Knowledge of the virion packaged DNA length, rather than the genome length, is necessary to calculate DNA packaging densities and is of interest in contemplating mechanisms of jumbo phage evolution. Pulsed-field gel electrophoresis (PFGE) is typically used to determine the packaged DNA length. By alternating the direction of the electrical field at regular time intervals, PFGE allows the resolution of DNA molecules greater than 50 kb. To avoid shearing DNA prior to electrophoresis, jumbo phage virions are concentrated, mixed with agarose to form plugs, then treated with proteinase K to digest virion proteins within the plug. The plug containing the liberated packaged DNA then serves as the sample and is loaded directly into a well of the gel. Many jumbo phage genomes are heavily modified which can cause an overestimation of the packaged DNA length. To address this issue, Jianfei Hua and colleagues developed a two-dimensional agarose gel electrophoresis (2D-AGE) protocol. Taking advantage of the fact that length does not detectably affect the mobility of very long (> 50 kb) dsDNA molecules during electrophoresis under constant voltage (nonpulsed electrophoresis), the technique involves a typical PFGE run followed by a perpendicular nonpulsed run. The first run separates DNA by length and mass/charge ratio, and the second run separates by mass/charge ratio only. The distance traveled by a phage DNA band in the second dimension relative to the distance traveled by unmodified marker bands is then used to calculate a corrected migration position predicted to occur in the first dimension if the phage DNA was not modified. In prepping jumbo phages for PFGE, it is important to note that some are sensitive to the concentrated cesium chloride solutions used in step gradients and isopycnic separations. Thus, polyethylene glycol precipitation and sucrose gradient velocity sedimentation are often used to concentrate jumbo phage solutions.

In addition to providing annotated genome sequences, packaged DNA lengths, and genomic terminal redundancy percentages, phage biologists have done an excellent job of characterizing the virion morphologies, plaque morphologies, host ranges, growth cycles, stabilities, virion protein compositions, head packaging densities, and even the transcriptional programs for many of the numerous jumbo phages isolated to date. Each of these characters aids the field in exciting comparative and evolutionary studies. In addition, novel morphological, genetic, and molecular features of several jumbo phages have been examined in greater depth and are highlighted below.

Virion Structure

Morphotypes

Bacteriophages are categorized into three orders: *Caudovirales*, *Ligamenvirales*, and Unassigned. The order *Caudovirales* represents the most numerous and widespread group of bacteriophages characterized to date. Three families belong to the *Caudovirales*: the *Myoviridae*, *Siphoviridae*, and *Podoviridae*. Members of these three families all possess unenveloped icosahedral capsids and are differentiated, morphologically, by their tail structures. Myoviruses have long contractile tails (morphotype A), siphoviruses have long noncontractile tails (morphotype B), and podoviruses have short tails (morphotype C). Each of the three morphotypes also display a range of head lengths, designated 1, 2, and 3 (Fig. 2). The vast majority of jumbo phages isolated to date belong to the *Myoviridae* family and display A1 and A2 morphologies. Exceptions include 16 different siphoviruses of *C. crescentus*, all of which belong to the genus *Phicbkvirus*. This genus of jumbo phages includes the group's namesake, phiCbK, and all members display the B3 morphotype with elongated heads and long, flexible, noncontractile tails.

Icosahedral Geometries

Jumbo phage capsids display a variety of icosahedral geometries with some possessing previously unseen triangulation numbers (T-numbers). *Sphingomonas paucimobilis* phage PAU and *E. coli* phage 121Q were the first examples of true $T = 25$ and $T = 28$ geometries, respectively, and *B. megaterium* phage G was the first example of a virus with $T = 52$ geometry. Accordingly, jumbo phages also display a fairly broad range of capsid sizes, from the 90 nm $\times$ 50 nm prolate head of *Vibrio parahaemolyticus* phage KVP40 to the 160 nm isometric head of phage G. The head structures of several jumbo phages have been determined via cryo-EM at resolutions sufficient for atomic model fitting. Interestingly, all utilize the canonical HK97 capsid fold conserved among smaller phages and simply assemble greater numbers of capsid protein subunits to generate larger heads. In addition, several different arrangements of decoration proteins have been noted among the jumbo phage capsids examined. For example, the heads of phages N3 (*Sinorhizobium meliloti*), PBS1 (*Bacillus subtilis*), G, and Bellamy (*Synechococcus* sp. WH8109) possess extra structures on their penton vertices that may help stabilize these large capsids.

Virion DNA Density

Although viruses with larger capsids are typically capable of packaging larger genomes, jumbo phage head size does not always correlate with genome length. For example, *P. aeruginosa* phage phiPA3 has a 100 nm head and a 309,208 bp genome while phage EL (also of *P. aeruginosa*) possesses a larger 140 nm head but a shorter 211,215 bp genome. The density at which DNA is packaged in phage heads is inherently important to the process of DNA injection during the initial stages of infection. Thus, jumbo phages with small genomes relative to their head sizes package additional DNA in the form of genomic terminal redundancies. Virion

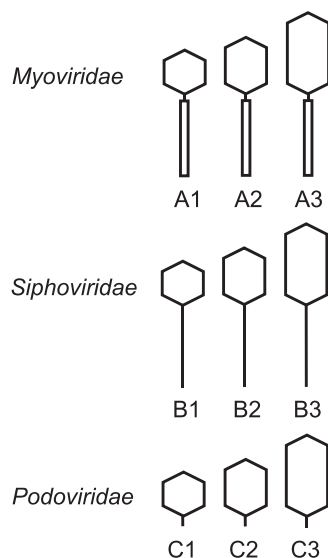


Fig. 2 *Caudovirales* morphotypes. The order *Caudovirales* is comprised of the *Myoviridae*, *Siphoviridae*, and *Podoviridae* families. Jumbo phages characterized to date belong to the *Myoviridae* and *Siphoviridae* and exhibit A1, A2, and B3 morphologies. Modified from Hull, R., Brown, F., Payne, C., 1989. *Virology: A Directory and Dictionary of Animal, Bacterial and Plant Viruses*. London: The Macmillan Press Ltd., p. 165.

packaged DNA lengths, determined via PFGE or 2D-AGE, and head volumes, calculated from cryo-EM measurements, are used in estimating packaged DNA densities. Such estimates have been made for several jumbo phages and found to vary substantially, from 0.39 bp/nm³ for phage G to 0.55 bp/nm³ for phage PAU.

Differences in packaged DNA densities are likely due to variabilities in capsid structures and associated strengths. In addition, some jumbo phages possess an inner capsid structure that decreases the head volume available for packaged DNA. The heads of phiKZ and related *Pseudomonas* jumbo phages 201phi2-1, phiPA3, EL, and OBP contain a highly ordered, proteinaceous, cylindrical structure that spans the inner length of the capsid. Originally visualized by transmission electron microscopy (TEM) in 1978, the protein composition of this 'inner body' has been determined via proteolysis and mass spectrometry. The inner body is encased in DNA in virions. This poses a problem for structural studies as the inner body is indistinguishable from the surrounding DNA in conventional cryo-electron micrographs. The inner body is sensitive to radiation damage, exploding into bubbles of gaseous radiation products at electron doses that leave the surrounding capsid structure intact. Using the information from these 'bubblegrams' the structure of the inner body and its localization within phiKZ capsids was determined and used to generate three-dimensional reconstructions. The phiKZ inner body is ~24 nm wide and ~105 nm long, is tilted ~22° relative to the portal axis and is composed of multiple stacked tiers with some regions possessing six-fold symmetry. It is hypothesized to play a role in organizing the packaged genome thereby reducing internal pressure on the capsid and promoting efficient and rapid genome ejection. Interestingly, the inner body is not detected following infection indicating its constituent proteins may be injected into host cells along with the phage genome.

Head and Tail Fibers

As with small phages, jumbo phage virions frequently contain fibers. Whereas in small phages such fibers are typically found attached to the tail baseplate and/or collar, some jumbo phages display fibers in unique locations and configurations. Phage PAU virions contain a narrow, long baseplate to which several tail fibers are attached. While the baseplate fibers of many jumbo phages are fairly short (20–30 nm), those of phage PAU are longer (~65 nm). These fibers extend in various directions from the bottom of the baseplate and appear somewhat flexible in electron micrographs, giving the tail complex a string mop appearance. The tail of *Pectobacterium carotovorum* jumbo phage vB_PcaM_CBB contains several long (~120 nm) whisker-like structures. Rather than emanating from the baseplate, these whiskers cover the length of the tail sheath. Phage CBB whiskers are highly stable, remaining attached during 1.5 years of storage at 4°C, implying they serve an indispensable role.

A few jumbo phages have also been shown to possess head fibers. Phage 121Q virions contain numerous long (~100 nm), curly, hair-like structures attached to both the capsid surface and all along the sheath section of the tail. The number of these fibers varies, ranging from 0 to >20 per virion. The distal ends of the fibers appear more electron dense in TEM images, suggesting they may terminate in a globular domain. Phage 121Q head and tail fibers likely interact tightly with the host cell as it is difficult to separate virion particles from host debris during phage purification. Jumbo phage Atu_ph07 virions also contain hair-like fibers that extend from both the capsid and along the length of the tail. In TEM images, the head fibers range in length with some appearing close to 200 nm, while the tail fibers are generally shorter (~50 nm). Finally, the *Tenacibaculum maritimum* jumbo phages PTm1 and PTm5 both possess fiber-like appendages emanating from the upper regions of their heads. These fibers are 50–100 nm long and appear to function in host recognition, as follows. Electron micrographs of phage-host cell mixtures show some PTm5 virions attached to cells

via the tops of their heads, while others are attached via their tails. These head fibers are easily lost during phage purification and may serve to enhance infection in ocean currents by providing an addition initial host binding mechanism prior to tail attachment.

Genome Features

The genomes of over 160 unique jumbo phages have now been sequenced and range in length from the 204,930 bp genome of phage Bellamy to the 497,513 bp genome of phage G. Jumbo phages with the 25 largest genomes infect a range of bacterial genera including *Bacillus*, *Agrobacterium*, *Salicola*, *Escherichia*, *Xanthomonas*, *Pectobacterium*, *Cronobacter*, *Serratia*, *Yersinia*, *Salmonella*, *Klebsiella*, *Caulobacter*, and *Pseudomonas* (Table 2). Comparative genomic analyses of jumbo phage genomes have aided immensely in our understanding of these giant viruses by revealing both broader commonalities and features unique to specific groups.

Terminal Redundancies

Genomic terminal redundancies are commonly seen in the packaged DNAs of jumbo phages. Terminal redundancies in smaller phages with linear, circularly permuted genomes, such as phage T4, result from a head-full packaging mechanism in which concatemeric DNA serves as the packaging substrate. Once initiated, concatemer packaging continues until the capsid is full. Because the phage T4 head volume is larger than its genome, the packaged DNA is terminally redundant. A similar head-full packaging mechanism is thought to occur in jumbo phages which also contain linear, circularly permuted, terminally redundant genomes. Terminal redundancies, typically expressed as a percentage of the genome length, are determined by subtracting the genome length (ascertained via sequencing) from the packaged DNA length (estimated by PFGE or 2D-AGE). Whereas the terminal redundancy for phage T4 packaged DNA is 3%, the terminal redundancies for the jumbo phages PBS1, G, and PAU have been estimated at 26%, 26%, and 35%, respectively. Other jumbo phages display smaller terminal redundancies, such as 5% for phage N3, 8% for phage 121Q, and 9% for phage Bellamy. The extent of genomic terminal redundancy is largely a result of the difference in head volume and genome size. Thus, shorter terminal redundancies may be indicative of more evolved phages that have lengthened their genomes through successive acquisitions of additional unique sequences.

Genome Nucleotide Composition

Although not exclusive to jumbo phages, it is worth noting that many of the jumbo phage genomes sequenced to date possess a low GC-content, particularly in comparison to that of their bacterial hosts. The phiKZ-like jumbo phages phiKZ, 201phi2-1, and

Table 2 Largest jumbo phage genomes. The 25 largest jumbo phages based on lengths of phage genome sequences submitted to GenBank as of December 31, 2019. Does not include megaphage sequences obtained via metagenomic analyses

Genome (bp)	Phage name	Host	Accession number
497,513	Phage G	<i>Bacillus megaterium</i>	NC_023719.1
490,380	Atu_ph07	<i>Agrobacterium tumefaciens</i>	NC_042013.1
440,001	SCTP-2	<i>Salicola</i> sp. PV4	MF360958.1
386,442	CMSTMSU	<i>Escherichia coli</i>	MH494197.1
384,670	XacN1	<i>Xanthomonas citri</i>	AP018399.1
378,379	CBB	<i>Pectobacterium carotovorum</i>	NC_041878.1
370,817	vB_EcoM_G17	<i>Escherichia coli</i> DSM 103244	MK327931.1
358,663	vB_CsaM_GAP32	<i>Cronobacter sakazakii</i>	JN882285.1
357,154	BF	<i>Serratia marcescens</i>	NC_041917.1
353,081	UB	<i>Escherichia coli</i> O157:H7	MH383160.1
352,598	vB_EcoM_phAPEC6	<i>Escherichia coli</i>	MK817115.1
352,596	fHe-Yen9-03	<i>Yersinia</i>	LT960552.1
350,103	Munch	<i>Salmonella</i>	MK268344.1
348,718	7t3	<i>Salmonella enterica</i> ssp. <i>enterica</i>	MK773491.1
348,532	121Q	<i>Escherichia coli</i>	KM507819.1
348,113	PBEC04	<i>Escherichia coli</i> OH157:H7	NC_027364.1
348,043	vB_Eco_slurp01	<i>Escherichia coli</i> MG1655	LT603033.1
347,152	SP27	<i>Escherichia</i>	LC494302.1
346,602	K64-1	<i>Klebsiella pneumoniae</i>	NC_027399.1
345,809	vB_KleM-Rak2	<i>Klebsiella isolate</i> KV-3	JQ513383.1
322,272	CcrBL9	<i>Caulobacter vibroides</i>	MH588546.1
317,488	CcrSC	<i>Caulobacter vibroides</i>	MH588547.1
316,674	201phi2-1	<i>Pseudomonas chlororaphis</i>	NC_010821.1
309,208	PhiPA3	<i>Pseudomonas aeruginosa</i>	NC_028999.1
309,157	Phabio	<i>Pseudomonas fluorescens</i>	MF042360.1

OBP have genomic GC-contents of 36%–48%, significantly lower than the 60%–66% GC-contents of their *Pseudomonas* hosts. Likewise, the *Aeromonas salmonicida* jumbo phages 65.2, PX29, and AS5 have genomic GC-contents of 37%, 42%, and 43%, respectively, while their host's genome possesses a GC-content of 58%. Whereas phage genomes are on average 4% richer in A/T than their hosts, this difference is 15%–30% for the above-mentioned jumbo phages. Although common among sequenced jumbo phages, this contrast in phage-host GC-content is not universal and several jumbo phages have genomic GC-contents similar to those of their bacterial hosts. The higher energy cost of synthesizing GTP and CTP combined with the greater abundance of intracellular ATP and TTP may result in increased replication of virulent phages with AT-rich genomes. Thus, higher genomic AT-contents may provide phages and other intracellular pathogens an evolutionary advantage.

Nucleotide Modifications and Substitutions

Many jumbo phage genomes are heavily modified while others incorporate nucleotide substitutions. Restriction modification systems comprise a well-studied and important component of the bacterial arsenal used against infecting phages. To thwart this defense mechanism the genomes of many phages, both large and small, contain nucleotide modifications and/or substitutions that go unrecognized by restriction endonucleases.

The first identified DNA base modification, 5-hydroxymethylcytosine, was discovered in T-even bacteriophages in 1953. Since then, various modifications in which chemical groups are biologically appended to the nucleobase have been shown for all four DNA nucleotides. These chemical additions range from simple methyl groups to amino acids, polyamines, and mono- and disaccharides. Some of the host- and phage-encoded enzymes that perform nucleobase modifications utilize free nucleotides as their substrates, while others perform base modifications post-replication. Genes encoding putative DNA modification enzymes have been found in several jumbo phage genomes. Jumbo phages SPN3US (*Salmonella enterica*) and phiEaH2 (*Erwinia amylovora*) encode adenine methylase homologs, *E. amylovora* jumbo phage Y3 encodes both adenine and cytosine methylase homologs, and *Klebsiella* jumbo phage vB_KleM_RaK2 encodes two putative DNA methyltransferase homologs. Genome modification in some jumbo phages is quite extensive. For example, whereas the length of jumbo phage PAU packaged virion DNA was estimated at 690 kb via PFGE, 2D-AGE provided a calculated corrected length of 296 kb. Thus, the phage PAU genome modifications impart a charge difference great enough to result in a 26% decrease in DNA electrophoresis migration rate.

Several jumbo phages utilize nucleotide substitutions rather than nucleobase modifications. *Yersinia enterocolitica* phage phiR1-37 and the *B. subtilis* jumbo phages PBS1, PBS2, and AR9 all have genomes in which thymidine is > 99% replaced by deoxyuridine. This pyrimidine substitution is accomplished by both altering the composition of intracellular deoxynucleotide pools and protecting the newly replicated uracil-containing phage genomes from host DNA repair pathways. Because the incorporation of deoxyuridine precludes direct sequencing, these phages presented unique challenges during genome sequencing. Nuclear magnetic resonance or liquid chromatography-mass spectrometry was first employed to determine the chemical makeup of the phage genomes and was followed by different creative solutions. For phiR1-37, phage DNA was cloned using *E. coli* CJ236, a strain lacking dUTPase and uracil N-glycosylase which made partial sequencing possible. Phage AR9 was amplified using a Uracil + polymerase for library prep prior to sequencing on the Illumina MiSeq platform. As additional jumbo phages are isolated and chemically analyzed, it will be interesting to discover the novel base modifications and/or nucleotide substitutions utilized by these evolutionarily innovative viruses.

ORFan Genes

One commonality seen across the spectrum of jumbo phage genomes sequenced to date is the high proportion of predicted open reading frames (ORFs) that currently lack detectable homology in public databases. These ORFan genes encode hypothetical proteins and constitute a large percentage of many jumbo phage genomes. For example, 70%–80% of the genomes of *P. aeruginosa* phage PaBG, *Aeromonas hydrophila* phage CC2, *Acinetobacter baumannii* phage ME3, *S. enterica* phage SPN3US, and *E. amylovora* phage phiEaH2 code for hypothetical proteins. This percentage is even higher in the jumbo phages phiCbK, PBECO4 (*E. coli*), and Ea35-70 (*E. amylovora*) with 88%–89% of their genomes encoding hypothetical proteins. As an increasing number of jumbo phage genome sequences are deposited in public databases, homologs are identified and the number of ORFan genes decreases. However, the functions of the predicted proteins remain largely unknown. This is both an exciting aspect and a challenge of jumbo phage biology; it currently imparts difficulty in making functionality predictions while presenting phage biologists with thousands of new and interesting proteins to explore.

Transfer RNA Genes

In addition to the phage-host genome nucleotide composition differences noted above, virulent phages also typically encode more tRNA genes than temperate phages. Phage-encoded tRNAs tend to correspond to codons that are highly prevalent in phage ORFs and simultaneously rare in the coding sequences of their hosts. Although the inclusion of tRNA genes in their genomes is not unique to jumbo phages, the sheer number of tRNAs encoded by some of these giant viruses is nonetheless remarkable. For example, whereas phage T4 encodes eight tRNA genes, the *V. parahaemolyticus* jumbo phages KVP40 and phipp2 encode 30 tRNA genes and the *C. crescentus* jumbo phages CcrKarma, CcrMagnet, and CcrColossus encode 26, 27, and 28 tRNA genes, respectively. As with smaller phages, the tRNAs encoded by jumbo phages are thought to correspond to codons that are abundant in phage coding sequences, especially those encoding structural proteins, and serve to increase translation efficiency and virion production.

Advanced Capabilities

With the advanced functionalities encoded by their large genomes, jumbo phages and other giant viruses are beginning to blur the line between what we consider living and non-living. Many of these advanced capabilities provide a level of host independence not seen in smaller viruses while others aid virus replication by boosting intracellular nucleotide pools, countering host defense systems or precluding superinfection/secondary infections by competing phages.

RNA Polymerases

A dependence on host transcriptional machinery is a hallmark of many dsDNA viruses. Most small dsDNA phages utilize the host's RNA polymerase (RNAP) throughout infection, modulating its function via various mechanisms to ensure ordered viral gene expression. Notable exceptions include the T7-like phages and the N4-like phages. T7-like phages rely on the host RNAP immediately upon infection to transcribe the phage RNAP gene, then utilize the phage-encoded RNAP to transcribe the viral middle and late genes. In a reverse strategy, the N4-like phages inject a virion-encapsulated RNAP (vRNAP) into the host cell upon infection. The vRNAP is used to express phage early genes including a non-virion RNAP (nvRNAP) that is used to transcribe the phage middle genes. Finally, the host RNAP, modified by the phage, is used to carry out late gene transcription. Thus, even though T7-like and N4-like phages encode their own RNAPs they are still dependent on the host RNAP at particular stages of their replication cycles. Jumbo phage phiKZ and its giant relatives, however, are known to replicate completely independently of the host RNAP.

The phiKZ genome encodes multiple ORFs with homology to RNAP subunits. RNA polymerases can be divided into two families on the basis of subunit composition, conserved amino acid motifs, and three-dimensional structures. The family of single-subunit RNAPs (ssRNAPs) are known to transcribe phage, mitochondrial, and chloroplast genomes, while the family of multi-subunit RNAPs (msRNAPs) transcribe genes in cells of all three domains of life. Whereas all other known phage-encoded RNAPs belong to the ssRNAP family, phage phiKZ encodes two sets of proteins homologous to amino-(N-) and carboxy-(C-) terminal fragments of the largest msRNAP subunits. Four proteins from one set are present in phiKZ virions and likely comprise a vRNAP that is injected into host cells upon infection and functions to transcribe phage early genes. The second set of proteins is thought to comprise an nvRNAP that functions in transcription of phage middle and late genes. Similar to the nvRNAP of the N4-like phages, the phiKZ-encoded nvRNAP is found in infected cells but not virions. The phiKZ nvRNAP has been purified from infected cells and analyzed by mass-spectrometry. The enzyme is composed of five polypeptides, all products of early phage genes. Two polypeptides are homologous to N- and C-terminal portions of msRNAP β subunits, two are homologous to N- and C-terminal portions of msRNAP β' subunits, and the fifth polypeptide currently possesses no homologs in public databases other than counterparts in related phiKZ-like viruses. Curiously, the nvRNAP holoenzyme lacks similarity to σ factors utilized by bacterial msRNAPs for promoter specificity, yet specifically transcribes from late phiKZ promoters in vitro. Thus, this jumbo phage may encode a novel mechanism for RNAP promoter recognition.

To examine dependence on the host RNAP, phiKZ transcription levels were quantified at time points throughout infection of *P. aeruginosa* pre-treated with rifampicin, an antibiotic that inhibits bacterial RNAPs. Whereas bacterial transcription was decreased by the addition of rifampicin, phiKZ gene transcription was resistant to rifampicin treatment. The *B. subtilis* jumbo phages PBS2 and AR9 and the *Ralstonia solanacearum* jumbo phage phiRP31 are also resistant to rifampicin. Like phiKZ, phages AR9 and phiRP31 encode two sets of RNAP β and β' subunits. Also like phiKZ, the RNAP purified from phage PBS2-infected cells is comprised of five subunits with molecular weights roughly corresponding to those of phiKZ's nvRNAP. Because the genome of phage PBS2 has not yet been sequenced, its relationship to phiKZ is unknown at this time. Phages AR9 and phiRP31, however, are known to belong to the genus *phiKZlikevirus*. Thus, it appears multiple phiKZ-related phages have evolved a transcriptional apparatus that provides increased host independence.

DNA Repair Enzymes

Viruses are often characterized as being generally ineffective at DNA repair because they typically do not encode the extensive and complex repair mechanisms found in their hosts. Jumbo phages, however, are expanding our knowledge of virus-encoded DNA repair mechanisms. The genomes of *Cronobacter sakazakii* jumbo phage CR5, *S. enterica* jumbo phage SPN3US, and *E. amylovora* jumbo phage vB_EamM_Deimos-Minion all encode homologs of the SbcC and SbcCD nucleases involved in double-strand break repair in *E. coli* and *B. subtilis*. In addition, the genome of *E. amylovora* jumbo phage Y3 encodes homologs of RecA, UvsE, and the MmcB-like family of DNA repair proteins. RecA is well known for its role in DNA maintenance and repair, UvsE is an endonuclease involved in removing nucleotides damaged by ultraviolet radiation, and MmcB-like proteins are endonucleases thought to function in DNA repair by generating the substrates for translesion synthesis. Functional in vivo analyses of these phage-encoded proteins have not yet been performed, but jumbo phages are thought to encode and employ DNA repair mechanisms to increase the number of viable progeny virions generated during infections. The larger genome and virion sizes of jumbo phages result in fewer infectious particles assembled prior to cell lysis. For example, phage Deimos-Minion has a reported burst size of only 5 plaque-forming particles per infected bacterial cell. Therefore, jumbo phages may have evolved to encode DNA repair mechanisms to help ensure sufficient progeny production.

NAD⁺ Salvage Pathway

To boost nucleotide synthesis in their hosts, many large phages encode a nicotinamide adenine dinucleotide (NAD⁺) salvage pathway. First identified in the genome of *V. parahaemolyticus* jumbo phage KVP40, over twenty jumbo phages encode enzymes that catalyze a two-reaction NAD⁺ salvage pathway. While not exclusive to jumbo phages, the smallest phage found to encode these enzymes, *Ruegeria pomeroyi* phage DSS3phi8, still possesses a genome greater than 146 kb. The phage-encoded NAD⁺ salvage pathway, proposed by Lee and colleagues, transforms nicotinamide (NAM) into NAD⁺ via a nicotinamide mononucleoside (NMN) intermediate (Fig. 3). Genes encoding the nicotinamide phosphoribosyltransferase (NAMPTase) and nicotinamide mononucleotide adenylyltransferase (NMNATase) enzymes that catalyze the pathway are found in the genomes of several *Caulobacter* phiCbK-like jumbo phages, jumbo phages that infect *Sinorhizobium*, *Ralstonia*, *Aeromonas*, *Cronobacter*, and *Klebsiella* species, and half of the *Vibrio* jumbo phages sequenced to date. The phage KVP40 NAMPTase and NMNATase homologs were purified and found to possess the requisite activities in vitro. Transcriptional analyses showed these enzymes were expressed early in the phage infection cycle, during the metabolic phase when intracellular NAD⁺ and NADH levels are most pertinent for phage replication. Phage DNA replication rates can be greater than four times those of host DNA replication, necessitating increased deoxynucleotide production during infection. With their exceedingly large genomes, this need is even more pronounced during jumbo phage infections. It appears several of these viruses have taken this matter into their own hands, so to speak, by encoding a mechanism to provide increased levels of DNA precursors.

LPS Biosynthesis

Various temperate phages are known to encode enzymes involved in lipopolysaccharide (LPS) biosynthesis. Such enzymes are thought to function in altering the surface composition of host cells, thereby precluding superinfection/secondary infection by competing phages. Intriguingly, large clusters of genes involved in LPS biosynthesis have now been found in the genomes of two lytic jumbo cyanophages. Their genomic inclusion was originally thought to simply serve as stuffer DNA for headful packaging. However, *Synechococcus* jumbo phage S-SSM7 and *Prochlorococcus* jumbo phage P-SSM2 were isolated from geographically distant locations hundreds of miles apart in the Atlantic Ocean yet encode seven identical homologs to LPS biosynthesis genes. Thus, it is likely these LPS genes are functionally linked. Upon infection of cells undergoing suboptimal growth, some T4-like lytic phages can enter a reversible dormant state in which lytic replication is stalled following early and middle gene expression and resumes once the necessary nutrients become available. It has been suggested a similar condition, referred to as pseudolysogeny, occurs with some marine phages. If this is the case, altering the surface composition of their hosts via phage encoded LPS biosynthesis may serve to prevent secondary infections by other cyanophages, particularly smaller phages capable of faster replication, thereby ensuring production of jumbo phage progeny upon resumption of lytic replication.

Phage Nucleus

Phages and their bacterial hosts are in a continuous co evolutionary arms race in which bacteria evolve to prevent phage infection and phages evolve to counter host defense mechanisms. Bacteria encode various defense mechanisms such as DNA restriction-

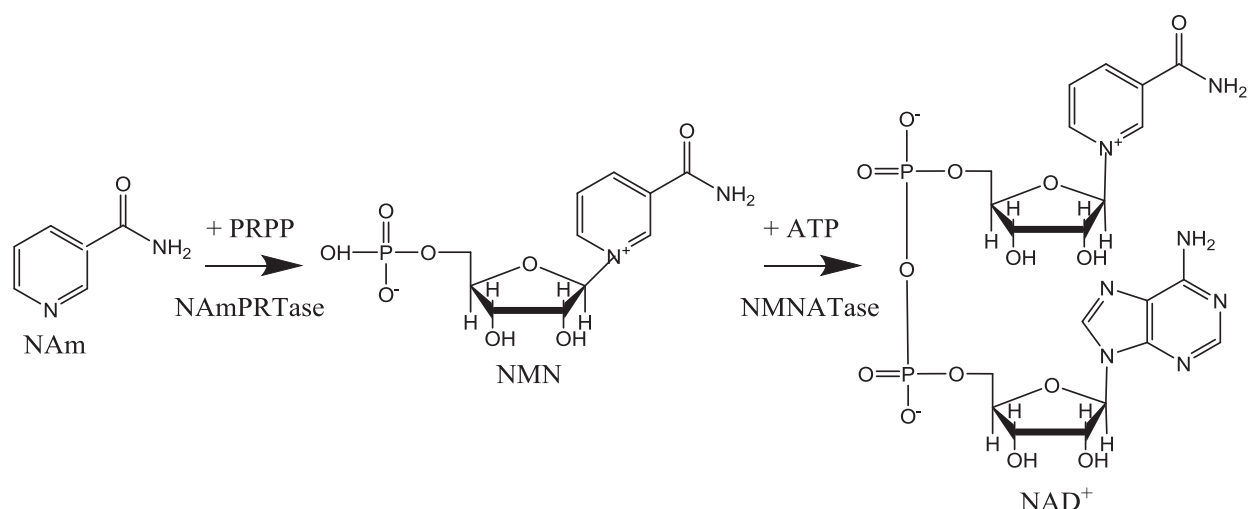


Fig. 3 Jumbo phage encoded NAD⁺ salvage pathway. Nicotinamide adenine dinucleotide (NAD⁺) is synthesized from nicotinamide (NAM), phosphoribosyl pyrophosphate (PRPP), and adenosine triphosphate (ATP) in a two-step pathway catalyzed by phage encoded nicotinamide phosphoribosyltransferase (NAMPTase) and nicotinamide mononucleotide adenylyltransferase (NMNATase) enzymes. Modified from Lee, J.Y., Li, Z., Miller, E.S., 2017. Vibriophage KVP40 Encodes a Functional NAD⁺ Salvage Pathway. *Journal of Bacteriology* 199 (9), e00855.

modification systems and CRISPR-Cas systems. In addition, bacteria often evolve changes in cell surface receptors utilized by phages for attachment. Phages have countered these strategies with genome chemical modifications and/or nucleotide substitutions, anti-CRISPR proteins, and changes to virion structures used in host attachment. Jumbo phages have also been shown to utilize a unique and intriguing counter-defense mechanism coined the 'phage nucleus'. This phage encoded structure forms a proteinaceous barrier in infected cells that separates the phage genome from intracellular defense mechanisms. The compartment is centered in the cell by a bipolar spindle composed of a phage encoded tubulin-like protein. Remarkably, the proteinaceous shell of the phage nucleus allows selective entry and/or retention of specific host and phage proteins. Enzymes involved in transcription and DNA replication are enclosed within the phage nucleus, while ribosomes and metabolic enzymes are relegated to the surrounding cytoplasm. Thus, in many aspects this phage-encoded structure is indeed reminiscent of eukaryotic nuclei. During infection, phage capsids assemble on the bacterial membrane, then migrate to the phage nucleus where they dock for DNA packaging.

Although the phage nucleus was originally described in *P. chlororaphis* cells infected with phage 201phi2-1, the role of this structure as a counter-defense mechanism has been more extensively studied with the jumbo phage phiKZ. While examining anti-CRISPR capacities of *P. aeruginosa* phages, Mendoza and colleagues noticed phiKZ was resistant to a variety of CRISPR systems despite a lack of anti-CRISPR homologs in its genome. Purified phiKZ DNA was cleaved by restriction endonucleases in vitro, thus ruling out genomic chemical modifications and/or nucleotide substitutions as a counter-defense mechanism. Employing confocal microscopy to investigate the mechanism by which phiKZ resists host degradation of its genome, phage DNA was found localized to the center of infected cells, while fluorescently tagged Cas9 protein was clearly excluded from the phage nucleus. The phage nucleus also excludes host restriction enzymes, thereby protecting phage DNA from both defense mechanisms.

It is currently unclear how widespread this counter-defense mechanism is among the various jumbo phage lineages. In addition to phiKZ and 201phi2-1, a phage nucleus is also encoded by phiPA3 and public database alignment tools reveal several other jumbo phages, including vB_PaeM_PS119XW, Phabio, and KTN4, encode homologs to the phiKZ tubulin and phage nucleus shell proteins. All these phages infect *Pseudomonas* species and belong to the *PhiKZlikevirus* genus, indicating the evolutionary distribution of this counter-defense mechanism might be limited. Recently, however, a *Serratia* jumbo phage (PCH45) highly divergent from all jumbo phages sequenced to date has been shown to encode and utilize a phage nucleus structure. Thus, this counter-defense mechanism may be more widespread than previously thought. Although the phage nucleus structures assembled in phiKZ- and PCH45-infected cells protect against Type I CRISPR-Cas systems that target phage DNA, they do not protect against Type III CRISPR-Cas systems which target phage mRNAs exported to the cytoplasm for translation. Many bacteria encode both DNA- and RNA-targeting CRISPR-Cas systems and, thus, the arms race continues.

Evolution

Phages belonging to the order *Caudovirales* vary dramatically in genome lengths and head sizes, ranging from the tiny *Rhodococcus rhodochrous* phage RRH1 to the giant phage G. Across this range, two common rules apply. First, genome lengths are limited by head capacities and second, high packaged DNA densities are necessary for successful DNA injection during the initial stages of infection. As noted above, jumbo phage heads utilize the canonical phage HK97 capsid fold and simply use greater numbers of capsid protein to build larger heads. Noting this conservation at the structural level, Roger Hendrix suggested an elegantly simple 'ratchet model' of phage evolution that accommodates the common rules. In this model, mutations in the major capsid and/or scaffold proteins lead to a larger capsid with an increased T-number while packaged DNA density is maintained through head-full packaging with a concomitant increase in terminal redundancy length. The enlarged head size allows for successive incorporation of additional coding sequences into the phage genome until the genome fills the capsid. Importantly, the newly added sequences need not displace preexisting phage genes and instead simply result in a reduction in terminal redundancy length. At this point, reversion to a smaller capsid size would likely be evolutionarily disfavored as it would require deletion of genes that encode selective advantages. Thus, the larger capsid size would become locked in, as by a ratchet, and evolution could repeat the process over and over generating larger and larger phages in a stepwise manner.

The upper limit of phage size is currently unknown. Several megaphages with genomes of 540,000–735,000 bp have recently been identified during metagenomic studies. Phylogenetic analyses of their terminase genes place these behemoths in the *Myoviridae* and all are predicted to infect members of the Bacteroidetes. As methodologies are developed to isolate these megaphages, capsid analyses will reveal whether structural conservation exists across the continuum of phage head sizes. Interestingly, megaphage genomes analyzed to date possess an apparently low (<70%) coding density when annotated according to the standard genetic code. Fragmentation of predicted coding sequences indicate these phages may use an alternative genetic code in which the canonical TAG stop codon is repurposed to encode glutamine. As larger and larger phages are discovered, additional interesting and exciting evolutionary innovations will surely be revealed.

Concluding Remarks

In 2019 the World Health Organization listed the growing concern of antibiotic resistance as one of the top ten greatest threats to global health. Phage therapy, successfully developed and utilized in Eastern Europe over the past century, is experiencing a resurgence in interest among Western government agencies, health care workers, and the food industry. The typically lytic lifestyles, advanced counter-defense mechanisms, and adequately broad host ranges of many jumbo phages make them good candidates for

phage therapy and/or biocontrol applications. However, detailed knowledge of any phage therapy candidate is a crucial prerequisite to broad clinical use. There is much work to be done in characterizing jumbo phage genomes and further exploring the unique capabilities, host ranges, and infection cycles of these phages. With continuing innovations in technology and the synergistic expertise of phage enthusiasts from diverse fields, the coming decades are sure to be an inspiring and productive time for jumbo phage biology.

Further Reading

- Al-Shayeb, B., Sachdeva, R., Chen, L.X., *et al.*, 2020. Clades of huge phages from across Earth's ecosystems. *Nature* 578, 425–431. doi:10.1038/s41586-020-2007-4.
- Chaikereetisak, V., Nguyen, K., Khanna, K., *et al.*, 2017. Assembly of a nucleus-like structure during viral replication in bacteria. *Science* 355, 194–197.
- Hendrix, R.W., 2009. Jumbo bacteriophages. In: Etten, J. (Ed.), *Lesser Known Large dsDNA Viruses*. Current Topics in Microbiology and Immunology 328. Berlin Heidelberg: Springer-Verlag, pp. 229–240.
- Hua, J., Huet, A., Lopez, C.A., *et al.*, 2017. Capsids and genomes of jumbo-sized bacteriophages reveal the evolutionary reach of the HK97 fold. *mBio* 8 (5), e01579.
- Kawato, Y., Istiqomah, I., Gaafar, A.Y., *et al.*, 2020. A novel jumbo *Tenacibaculum maritimum* lytic phage with head-fiber-like appendages. *Archives of Virology* 165, 303–311.
- Lavysh, D., Sokolova, M., Minakhin, L., *et al.*, 2016. The genome of AR9, a giant transducing *Bacillus* phage encoding two multisubunit RNA polymerases. *Virology* 495, 185–196.
- Lee, J.Y., Li, Z., Miller, E.S., 2017. *Vibrio* phage KVP40 encodes a functional NAD⁽⁺⁾ salvage pathway. *Journal of Bacteriology* 199 (9), e00855.
- Malone, L.M., Warring, S.L., Jackson, S.A., *et al.*, 2020. A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nature Microbiology* 5, 48–55.
- Mendoza, S.D., Niewegloska, E.S., Govindarajan, S., *et al.*, 2019. A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* 577 (7789), 244–248.
- Saad, A.M., Soliman, A.M., Kawasaki, T., *et al.*, 2019. Systemic method to isolate large bacteriophages for use in biocontrol of a wide-range of pathogenic bacteria. *Journal of Bioscience and Bioengineering* 127 (1), 73–78.
- Yakunina, M., Artamonova, T., Borukhov, S., *et al.*, 2015. A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage. *Nucleic Acids Research* 43, 10411–10420.
- Yuan, Y., Gao, M., 2017. Jumbo bacteriophages: An overview. *Frontiers in Microbiology* 8, 403.

CRISPR-Cas Systems and Anti-CRISPR Proteins: Adaptive Defense and Counter-Defense in Prokaryotes and Their Viruses

Asma Hatoum-Aslan and Olivia G Howell, University of Alabama, Tuscaloosa, AL, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Cas CRISPR-associated

CRISPR A genetic locus composed of clustered regularly-interspaced short palindromic repeats

crRNA CRISPR RNA

PAM Protospacer adjacent motif

PFS Protospacer flanking site

Glossary

Conjugation A process by which conjugative plasmids are transferred from one prokaryote to another.

Homologs A group of related proteins that share a common ancestor and likely perform similar functions.

Horizontal gene transfer The exchange of genetic material from one organism to another.

Lysogen A prokaryotic cell that harbors the genome of one or more prophages.

Natural transformation A process by which microorganisms acquire DNA from their environment.

Nuclease An enzyme that degrades nucleic acids (RNA and/or DNA).

Nucleic acid Macromolecules such as DNA or RNA which are composed of mononucleotide building blocks.

Phage A virus that specifically infects prokaryotic organisms.

Prokaryotes Single-celled organisms which lack a nucleus and belong to the domain of *bacteria* or *archaea*.

Prophage A phage genome that has been integrated into a prokaryotic chromosome.

The Phage-Host Arms Race

Phages comprise a diverse range of viruses capable of infecting specific bacterial and archaeal hosts. Independently discovered in 1915 and 1917 by Frederick Twort and Félix d'Hérelle, phages are now recognized as the most abundant entities on the planet. In fact, there are an estimated 10^{31} phage particles in the biosphere, which equates to about ten billion times more phages than there are stars in the observable universe! Phages have been found to persist ubiquitously alongside their prokaryotic hosts and outnumber them by a factor of ten to one. Phages are parasites that rely on their hosts to reproduce. Most phages replicate in one of two manners: some undergo a strictly lytic cycle, while others may choose to enter into a lysogenic cycle (Fig. 1). The former are referred to as virulent or lytic phages, while the latter are called temperate phages. Both types begin their replication cycle by attaching to a specific receptor on their host's surface and injecting their genetic material into the cell. Virulent phages strictly adhere to the lytic replication cycle, during which injection is promptly followed by the synthesis of the phage genome and proteins, and the assembly of progeny phages. These processes rely upon the host's enzymes and reservoir of chemical building blocks. Following a defined latent period, the progeny exit the cell, an event that typically leads to cell lysis and the death of the host. In contrast, following injection, temperate phages decide whether to enter into the lytic cycle or persist in a state of dormancy through the lysogenic cycle. The lysogenic cycle usually entails the integration of phage DNA into the host chromosome immediately after entry into the cell. The integrated phage is termed a prophage, and prokaryotes harboring one or more prophages are referred to as lysogens. After integration, the prophage is replicated along with the host chromosome until a stressor or other signal triggers its excision. Following excision, the prophage completes the lytic replication cycle. While these varying strategies of replication follow distinct time courses and carry differing evolutionary ramifications, it has been estimated that nearly 10^{30} such phage infections occur each day. Therefore, phages impose a substantial selective pressure upon their hosts, who must quickly respond and adapt in order to survive.

This relationship between phages and their hosts is a classic example of the Red Queen Hypothesis. Originally posited by Leigh Van Valen in 1973, the Red Queen hypothesis states that in order to maintain fitness, both the host and parasite must continually co-evolve in a veritable arms race to surmount the defenses of each party involved. Such a process can lead to the near-extinction of one side or the other; however, a series of evolutionary adaptations within host and parasite ensures that both continue to co-exist, albeit in a rather uneasy equilibrium. Accordingly, numerous studies have reported upon the various strategies phages and bacteria have acquired to ensure their continued survival. Prokaryotic defenses against phages can include passive mechanisms such as the modification of cell-surface receptors through random mutagenesis, as well as active mechanisms such as the induction of programmed cell death in the face of a phage invasion (a process known as abortive infection), or the degradation of phage-derived nucleic acids by prokaryotic immune systems. The latter can be further divided into innate immunity, such as that

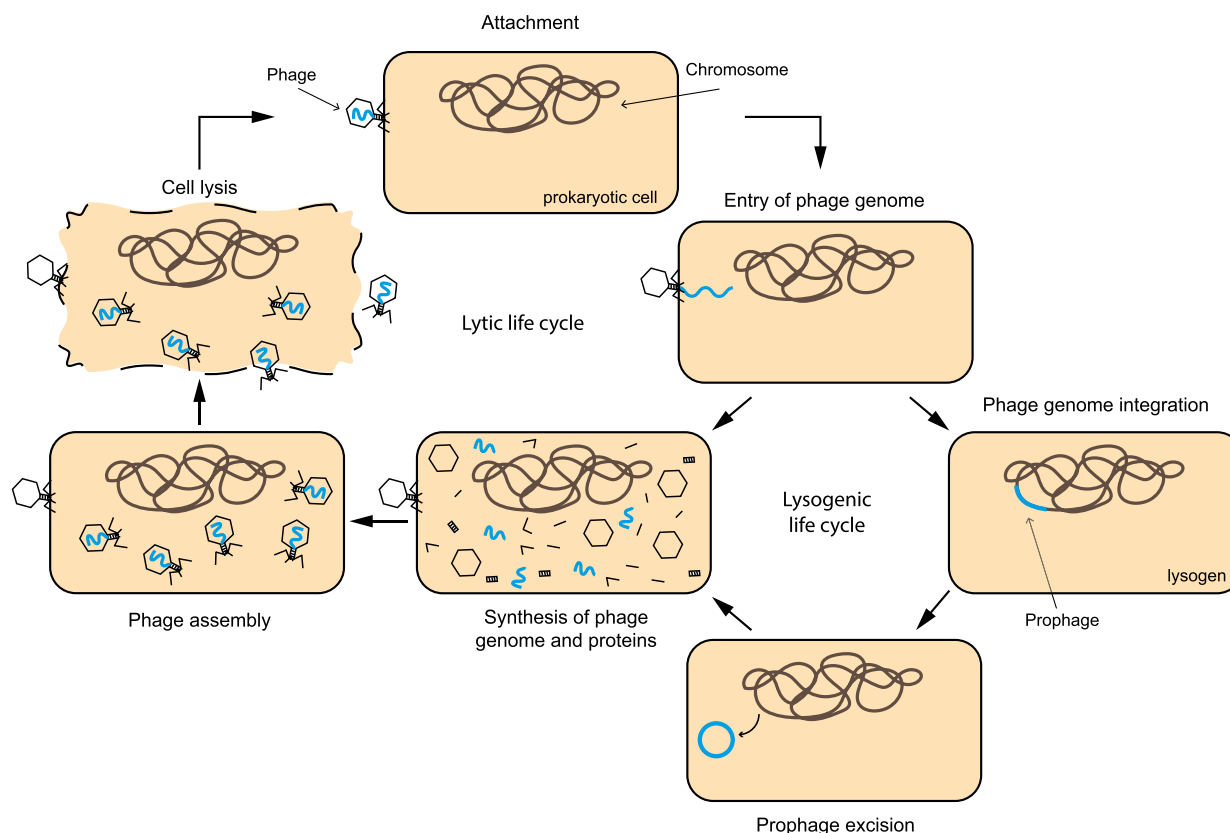


Fig. 1 The lytic (left) and lysogenic (right) cycles of phage replication.

conferred by restriction-modification systems, and adaptive immunity, such as that provided by clusters of regularly-interspaced short palindromic repeat (CRISPR) sequences and CRISPR-associated (*cas*) genes.

Adaptive immune systems were once thought to be found exclusively in animals; however, the discovery of CRISPR-Cas systems proved this notion to be incorrect. These systems utilize short CRISPR RNAs (crRNAs) and Cas proteins to detect and destroy foreign nucleic acids originating from plasmids, phages and other extrachromosomal genetic elements. The general CRISPR-Cas pathway occurs in three steps (**Fig. 2**): adaptation, expression, and interference. During adaptation, short segments of invading nucleic acids (~30–40 nucleotides in length) are captured and inserted into the CRISPR locus as “spacers” in between similarly sized repeat sequences in order to record molecular “memories” of the invader. During expression, the repeats and spacers are transcribed into a long precursor crRNA (pre-crRNA). The pre-crRNA is subsequently processed (*i.e.*, chopped) within each repeat to liberate mature crRNAs that each specifies a single nucleic acid target. Mature crRNAs combine with Cas proteins to form a Cas complex (also known as an effector complex) which patrols the interior of the cell, searching for a match. During interference, foreign nucleic acids that bear sequences complementary to the crRNA (called “protospacers”) are detected and sliced by the Cas enzymes.

CRISPR-Cas systems are relatively widespread in nature – about 40% of bacteria and 90% of archaea have been found to harbor one or more of these systems. They not only protect against phage infection, but also prevent other modes of horizontal gene transfer (HGT) such as conjugation and natural transformation. Since HGT is a major driving force of prokaryotic evolution, these systems profoundly impact the microbial communities within, on, and around us. However, CRISPR-Cas systems by no means represent an impenetrable barrier as phages have been shown to evolve a variety of resistance strategies. As an example, anti-CRISPRs consist of a diverse set of phage proteins acquired in response to the selective pressures imposed by their CRISPR-Cas containing hosts. These proteins are the first-identified viral defense mechanisms that provide sustainable protection against the adaptive immunity conferred by CRISPR-Cas systems. This article will focus on the history and biology of CRISPR-Cas systems and anti-CRISPR proteins, two remarkable molecular innovations that have emerged from the perpetual phage-host arms race.

It is worthwhile mentioning that over the past decade, significant advances in our basic understanding of CRISPR-Cas pathways have enabled the development of a myriad of cutting-edge biotechnologies that continue to revolutionize genetics research and promise to offer transformative treatments for genetic diseases. Notable pioneers of CRISPR-Based technologies include Jennifer A. Doudna of the University of California at Berkeley, Emmanuelle M. Charpentier of the Max Planck Institute, Feng Zhang at the Broad Institute, and George Church at Harvard University. Such technologies include powerful genetic engineering tools that can be used to introduce precise modifications into the genomes of organisms spanning all domains of life, including humans. As a consequence, the birth of the very first CRISPR-engineered human babies was revealed in 2018 through a shocking announcement

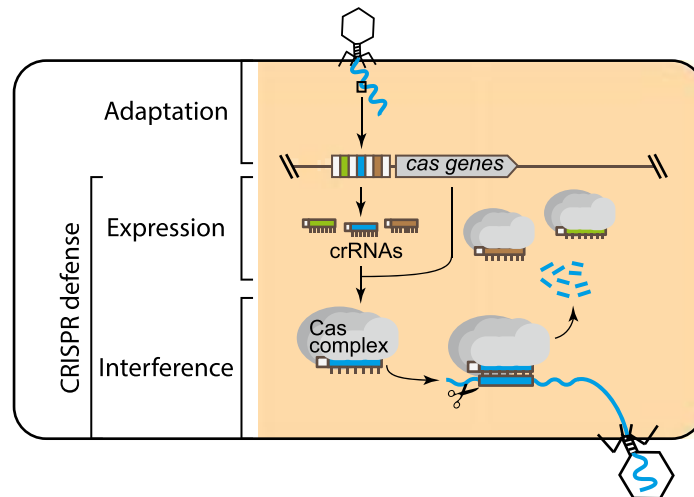


Fig. 2 The CRISPR-Cas pathway.

that has sparked an international controversy and heated ethical debates across the globe, the conclusions of which have yet to be discerned. The topic of CRISPR-based technologies and potential applications in humans lies outside of the scope of this article. However, for more information on these topics, the reader is directed to the relevant articles referenced in the “Further Reading” section.

CRISPR-Cas, A Brief History

At the time of writing (Summer, 2019), a simple search for “CRISPR” in the NCBI (National Center for Biotechnology Information) “PubMed” database (see “Relevant Website section”) returned 13,718 publications – an immense body of literature that reflects the intense interest in the subject, yet stands seemingly at odds with its relatively brief history. The presence of the curious clusters of repeats were first observed in *Escherichia coli* in 1987, in a study lead by Atsuo Nakata at Osaka University, Japan. This study sought to characterize *iap*, a gene thought encode a proteolytic (protein-cleaving) enzyme. As almost an afterthought, it was noted that *iap* is encoded adjacent to an unusual genetic structure consisting of several 29-nucleotide direct repeats interrupted by 32-nucleotide unique sequences. In 1989, a follow-up paper by the same group was published in which these repeats were further characterized and identified in other gram-negative bacteria. Around the same time period, nearly half-way across the world, Francisco J. M. Mojica began his PhD dissertation research at the University of Alicante, Spain. While studying the gas vesicles of *Haloferax mediterranei*, a salt-loving archaeon, Mojica came across regularly-spaced repeats of 30 nucleotides separated by short stretches of unique sequences within the organism’s genome. In the years that ensued, multiple publications independently noted similar structures in diverse *bacteria* and *archaea*, including studies by Groenen *et al.* (1993), Masephol *et al.* (1996), Karlin *et al.* (1998), and Hoe *et al.* (1999). However, despite this increasing awareness of their prevalence, the biological function of these mysterious repeats remained elusive for over a decade.

Undoubtedly, early bioinformatic and computational analyses played pivotal roles in predicting their biological relevance. In 2002, the term “CRISPR” was coined to describe this special class of repeats, an acronym agreed upon by Mojica – by that time a leader of his own research group – and a team from Utrecht University lead by Ruud Jansen. Jansen’s group was the first to conduct extensive *in silico* analyses of CRISPR loci and flanking regions, and identified the conserved *cas* genes in their vicinity. However, the most important clues arose when attention shifted from the repeats and *cas* genes to the short “spacer” sequences in between them. In 2005, Mojica’s group published the first report of a systematic analysis of the spacers – it was discovered that many of them matched sequences found within extrachromosomal genetic elements such as plasmids and phages. These and other observations lead to the hypothesis that CRISPR-Cas systems provide immunity against mobile genetic elements matching the spacers through a small RNA-mediated mechanism akin to eukaryotic RNA interference. Within a few months, similar observations were reported in two additional publications from two independent research groups in France (lead by Gilles Vergnaud and Alexander Bolotin, respectively). Both reports noted that many spacers are derived from phage sequences or other extra-chromosomal elements, and Bolotin’s group also arrived at the hypothesis that CRISPR elements and associated *cas* genes likely protect cells against foreign nucleic acid invasion.

In the years that followed, several seminal studies published in quick succession by a collection of international teams revealed the basics of how CRISPR-Cas systems operate at the molecular level. One such study, published in 2007, was led by Philippe Horvath and Rodolphe Barrangou. They reported the first experimental evidence that spacers and associated *cas* genes indeed provide immunity against phages. At the time, the lead researchers were affiliated with Danisco Inc., a food production company (acquired by DuPont in 2011), and the study focused on *Streptococcus thermophilus*, a key species used in the production of yogurt

Table 1 Cas protein composition of the six CRISPR-Cas types

Class	Type	Subtype designations	Adaptation	Expression	Interference	Large subunit
1	I	A-F, U, Fv ^a	Cas1, Cas2	Cas6	Cas5, Cas7, Cas8, Cas3, SS ^d	Cas8
	III	A-D, Bv	Cas1, Cas2	Cas6 ^c	Cas10, Cas5, Cas7, SS Csm6/CARF ^e	Cas10
	IV	IV, IVv	unknown	Cas6v	Cas5, Cas7, Cas8-like, SS	Cas8-like
2	II	A-C, Cv	Cas1, Cas2, Cas4	RNase III	Cas9	Cas9
	V	A-E ^b	Cas1, Cas2, Cas4	Cas12	Cas12	Cas12
	VI	A, B1, B2, C	Cas1, Cas2	Cas13	Cas13	Cas13

^av, variant.^bAdditional tentative systems have been proposed.^cAdditional host nucleases are also involved.^dSS, subtype-specific small subunit.^eCARF, CRISPR-associated Rossman fold domain-containing protein.

and cheese. Presumably stemming from an interest in developing more robust, phage-resistant bacterial strains, the team investigated the relationship between the bacterium's CRISPR-Cas locus and its phages. They demonstrated that phage infection caused the bacteria to acquire new spacer sequences that matched back to regions of the phage genome. Furthermore, by intentionally adding and removing spacer sequences from the bacterium's genome, the investigators established a positive correlation between the presence of a specific spacer sequence and resistance to the phage that harbors that sequence. One year later, a study on *E. coli*'s CRISPR-Cas system was published by John van der Oost, Stan J.J. Brouns, and co-workers at Wageningen University in the Netherlands. In this study, the investigators unambiguously demonstrated that repeats and spacers are transcribed as a long pre-crRNA, and the spacers are snipped out to produce mature crRNAs that each contain a single spacer sequence. Mature crRNAs were found to associate with a group of Cas proteins, and both components were required to interfere with phage proliferation.

As continuing insights into CRISPR-Cas mechanisms in diverse organisms appeared in the literature, clear distinctions between them began to emerge. As examples, a study published in 2008 by Luciano A. Marraffini and Erik J. Sontheimer at Northwestern University provided evidence which implied that DNA is the most likely target for CRISPR interference in the skin-dwelling bacterium *Staphylococcus epidermidis*. However, one year later, a similar CRISPR-Cas system in the archaeon *Pyrococcus furiosus* was convincingly shown to directly cut protospacer sequences within mRNA in a study lead by Michael and Rebecca Terns at the University of Georgia. Another report published in 2010 by a group at Université Laval in Quebec lead by Sylvain Moineau showed that double-stranded plasmid and phage DNA are directly cleaved within protospacer regions by the CRISPR-Cas system in *S. thermophilus*. Although these initial reports on the targets of CRISPR-Cas immunity seemed to contradict one another, further advances have reconciled these seemingly disparate observations and have led to the current understanding that there exist multiple CRISPR-Cas Types which exhibit remarkable diversity with regard to their genetic architecture, *cas* gene composition, and mechanisms of action.

CRISPR-Cas Diversity and Classification

The variety of CRISPR-Cas systems known today are believed to have evolved from a single common ancestor, which diverged over time due to its dissemination into diverse organisms and the constant selective pressure imposed by the phage-host arms race. In 2006, even before the very first experimental evidence of CRISPR-Cas function became available, Eugene V. Koonin and Kira S. Makarova of the National Institutes of Health, USA, began the process of building a framework for CRISPR-Cas classification. These investigators used computational methods to compare CRISPR loci and *cas* genes from different systems across available prokaryotic genomes in order to infer their evolutionary relationships. They also integrated their findings with structural and functional data to arrive at the current classification scheme. Therefore, it is important to note that the current scheme represents an evolving framework that is continually revisited and refined as new CRISPR-Cas systems are identified and more information about them becomes available. According to this scheme (summarized in Table 1), CRISPR-Cas systems are grouped into two broad classes and six distinct Types (I-VI). Class 1 CRISPR-Cas systems include Types I, III, and IV, while Class 2 systems include Types II, V, and VI. Each CRISPR-Cas Type is further subdivided into subtypes (indicated with a letter or alphanumeric designation). One defining feature that separates the classes is the configuration of their effector complexes: Class 1 systems all possess multi-subunit Cas complexes, while Class 2 systems possess single-subunit effectors. Furthermore, the protein composition of these complexes delineates the different CRISPR-Cas Types. All complexes contain a large Cas protein that is specific to each Type, and Class 1 complexes contain an additional small Cas protein subunit that is characteristic of each subtype. Class 1 systems are the most widespread in nature and have been found to reside in bacterial and archaeal species, while the less common Class 2 systems are found nearly exclusively in bacteria.

It now bears mentioning that the early functional insights into the general three-step CRISPR-Cas pathway emerged from studies on a diverse set of Class 1 and Class 2 systems: Type I-E (found in *E. coli*), Type II-A (found in *S. thermophilus*), and Types III-A and III-B (found in *S. epidermidis* and *P. furiosus*, respectively). Adaptation is considered the most highly-conserved

step in CRISPR-Cas immunity. Accordingly, Cas1 and Cas2, which are essential for adaptation, are found in the majority of CRISPR-Cas Types. On the other hand, the expression and interference steps of immunity, collectively referred to as CRISPR defense, are carried out by diverse Cas and sometimes even non-Cas proteins and/or additional factors. As examples to highlight this remarkable functional diversity, crRNA processing in *E. coli*'s Type I-E system is catalysed by the Cas6 nuclease, which snips the pre-crRNA within repeat sequences. In the *S. epidermidis* Type III-A system, crRNA processing occurs in multiple steps that are carried out by Cas6 and additional host-encoded nucleases. CrRNA processing is even more complicated in Type II systems. As an example, the Type II-A system in *Streptococcus pyogenes* relies upon the host-encoded nuclease RNase III as well as a second small RNA called a trans-activating crRNA (tracrRNA). The tracrRNA bears partial complementarity with the CRISPR repeats and anneals to the pre-crRNA in order to create the double-stranded RNA substrate that is required for cleavage by the host-encoded nuclease RNase III. This remarkable mechanism was first reported in 2011 in a study lead by Emmanuelle Charpentier, then at Umeå University in Sweden. The interference step of immunity is similarly diverse across all CRISPR-Cas Types and relies upon distinct sets of Cas and non-Cas proteins that degrade a variety of nucleic acid targets. For example, Type I, II and V systems recognize and cut DNA targets, Type VI systems degrade strictly RNA, and Type III systems can cut both DNA and RNA. To further illustrate their diversity, some CRISPR-Cas Types have evolved to perform functions beyond adaptive immunity, such as the regulation of endogenous gene expression. Undoubtedly, future explorations into CRISPR-Cas biology will continue to reveal such unexpected mechanisms and contribute a broader understanding of the impacts of these systems on the host organisms in which they reside.

Mechanisms of CRISPR-Cas Immunity

Given the considerable diversity across the six CRISPR-Cas Types, it is difficult to draw generalizations that would encapsulate all of their mechanistic details. Therefore, the following sections will focus on the adaptation and defense mechanisms of the best-characterized examples to serve as a starting point for a more in-depth exploration of the literature. It is worthwhile mentioning that Types I, II, and III systems have historically been the most thoroughly investigated and are considered the main CRISPR-Cas Types, whereas Types IV, V, and VI are newer additions to the CRISPR-Cas collection, and considerably less is known about their biology.

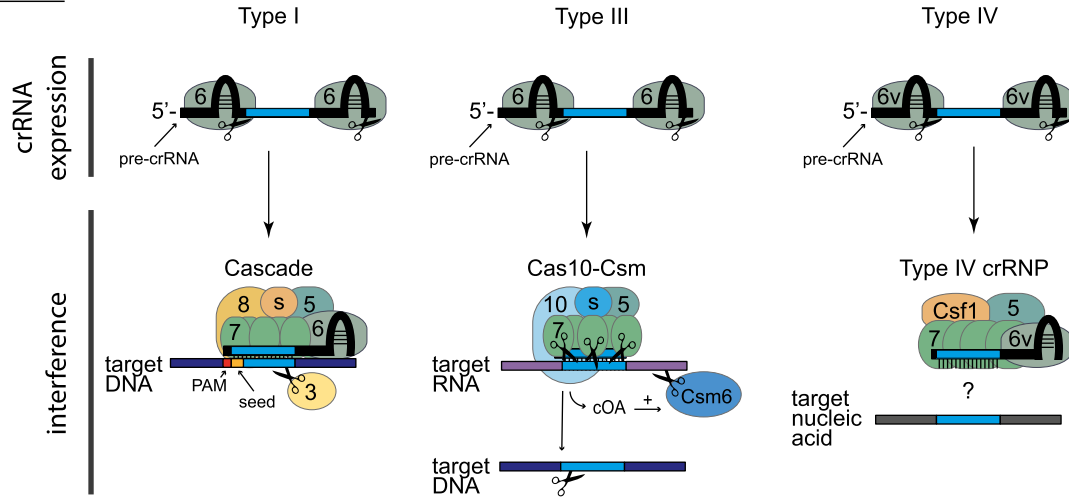
Adaptation

The first step in the CRISPR-Cas pathway, adaptation generally involves four steps: detection of foreign nucleic acids, selection of a protospacer, processing of the protospacer, and integration into the CRISPR locus as a new spacer. Since the mistaken integration of endogenous chromosomal DNA would lead to self-targeting (*i.e.*, autoimmunity) and cell death, the protein machinery catalysing these events must necessarily exhibit a strong preference for foreign nucleic acids and/or remain under tight regulation by the cell. Although adaptation is considered the most highly-conserved step of the CRISPR-Cas pathway, it remains the least well-understood. Much of the information available on adaptation has been derived from studies on Types I-F and II-A CRISPR-Cas systems.

E. coli's Type I-F system is arguably the best-characterized of the six CRISPR-Cas Types, and the first insights into its mechanism of adaptation were reported in 2012 by Udi Qimron's group at Tel Aviv University. These investigators demonstrated that Cas1 and Cas2, even in the absence of other Cas proteins, could mediate the acquisition of new spacer sequences from a plasmid. The investigators observed that new spacers are added on the promoter-proximal end of the repeat-spacer array and integration requires the presence of specific sequences in the "leader" region (*i.e.*, sequences that reside directly upstream of array). Importantly, analysis of the acquired spacers and corresponding targeted regions in the plasmid revealed the presence of conserved sequences adjacent to the protospacer (referred to as a protospacer-adjacent motif, or PAM). It is now known that PAM sequences play an essential role in preventing autoimmunity by allowing the CRISPR-Cas system to distinguish between invader-derived DNA and the spacer sequences of the endogenous CRISPR locus. Subsequent studies on this system by numerous groups, including those lead by Severinov and Semionova (2012) and Brouns (2012), revealed that there are at least two possible modes of adaptation: naïve and primed. Naïve adaptation occurs when the system encounters a brand-new phage bearing little or no resemblance to an existing spacer. This mode of adaptation depends entirely on the activities of Cas1 and Cas2, and occurs at a lower frequency. In contrast, primed adaptation occurs when the system encounters a phage that bears partial complementarity to an existing spacer. Primed adaptation is a more efficient process that relies upon the effector complex to guide Cas1 and Cas2 to the partially-matching target and facilitate further acquisition of new spacers in the vicinity. Following these initial findings, roles for host-encoded proteins in protospacer selection and new spacer integration were later described by Rotem Sorek's group at the Weizmann Institute of Science and Jennifer Doudna's group at the University of California at Berkeley, respectively, in 2015.

In subsequent studies by various groups, mechanisms of adaptation in the Type II-A system came to light. Similarly to Type I-F, adaptation by the Type II-A CRISPR-Cas system was found to rely upon Cas1 and Cas2 in addition to other Cas and non-Cas proteins encoded in the host. Although it is unclear whether this system exhibits primed adaptation, naïve adaptation in Type II-A also displays requirements for sequence motifs in the CRISPR leader region, preferences for adding new spacers to the leading end, and the presence of a PAM.

Class 1



Class 2

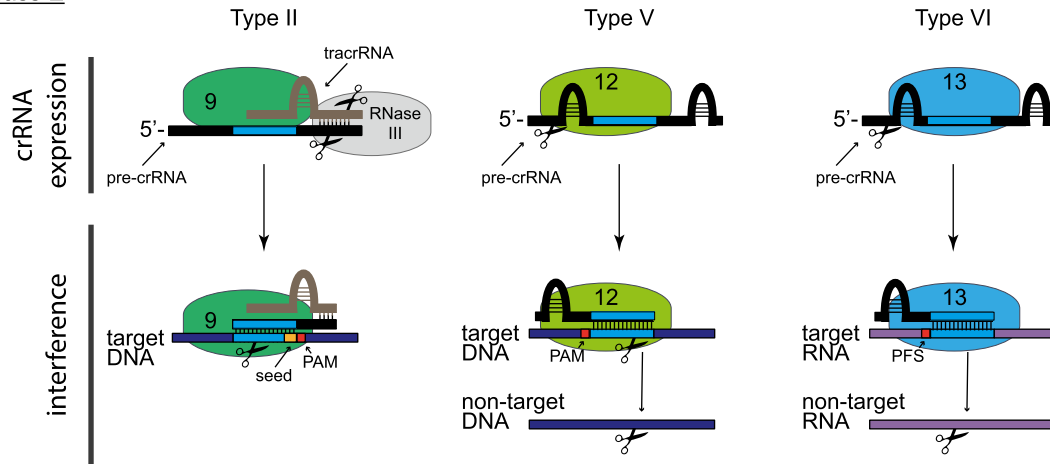


Fig. 3 Examples of the six CRISPR-Cas types.

Defense

Type I CRISPR-Cas systems, sometimes referred to as CRISPR-Cas3, are the most widely distributed in nature. As mentioned above, the best characterized example is the Type I-F system found in *E. coli* (Fig. 3, top panel). CrRNA expression in this system relies upon the endonuclease Cas6 (also known as CasE or Cas6e). This enzyme recognizes and cuts the pre-crRNA within each repeat. Cas6 and the mature crRNA are then joined by four additional Cas protein subunits in varying stoichiometries: one copy of Cas8 (the large Type-specific subunit, also called CasA or Cse1), six copies of Cas7 (also known as CasC), one copy of Cas5 (also called CasD), and two copies of CasB (a small subtype-specific subunit also known as Cse2). The resulting Cas complex is known as Cascade (CRISPR-associated complex for antiviral defense), a term that was coined by John van der Oost, Stan J.J. Brouns, and co-workers in 2008 in their original paper which first described this complex. Cascade can detect nucleic acid sequences that are complementary to the crRNA and positioned adjacent to a two- to six- nucleotide PAM sequence. Once such invading DNA sequences are identified, the helicase-nuclease Cas3, a protein separate from the complex, degrades the targeted DNA. An important feature of Type I systems is the requirement for perfect complementarity in the first 6–8 nucleotides adjacent to the PAM in the crRNA-protospacer match, a region known as the “seed”. Notably, the presence of a PAM and seed in Type I systems render them particularly susceptible to passive mechanisms of phage escape from immunity, whereby if a phage randomly acquires a point mutation in the PAM or seed region, it will evade CRISPR detection altogether.

Type II CRISPR-Cas systems, also known as CRISPR-Cas9 (Fig. 3, bottom panel), have historically been the most commonly used for biotechnological applications due to their relative simplicity compared to Class 1 systems. The Type II-A systems found in *Streptococcus pyogenes* and *S. thermophilus* were among the first to be functionally characterized. As mentioned earlier, crRNA processing in these systems relies upon a small tracrRNA, which bears some complementarity to the repeat-derived sequence of the

pre-crRNA. Pairing between the two RNAs facilitates crRNA processing by the host-encoded nuclease RNase III. During interference, the Cas9 nuclease bound to both RNAs recognizes and cleaves both strands of foreign double-stranded DNA matching the crRNA sequence. Similarly to Type I, Type II systems rely upon the presence of a PAM to distinguish “self” from “non-self” DNA and require perfect complementarity in a seed region to carry out interference. Accordingly, targeted phages can escape Type II immunity by simply acquiring point mutations in these critical regions.

Type III CRISPR-Cas systems are the second-most prevalent in nature and are very likely the most complex of the six Types. One of the best-characterized examples of these is the Type III-A CRISPR-Cas system found in *S. epidermidis*, also known as CRISPR-Cas10 (Fig. 3, top panel). In this system, crRNA processing relies upon the activities of Cas6 as well as host-encoded nucleases including PNPase. The mature crRNAs that result are joined by five protein subunits in varying copy numbers: one copy of Cas10 (the large Type-specific subunit, also called Csm1), one copy of Cas5 (also called Csm4), multiple copies of the Cas7 homologs Csm3 and Csm5, and multiple copies of Csm2 (the small subtype-specific subunit). The resulting effector complex is termed “Cas10-Csm”. Interference by this system has at least two requirements: First, the targeted region must be actively transcribed, and second, the crRNA must bear complementarity to the mRNA. Once the Cas10-Csm complex pairs with a targeted mRNA, it performs interference through the activities of at least three different CRISPR-associated nucleases: Cas10 cuts the targeted DNA, Csm3 cuts the targeted RNA, and Csm6, which is not a member of the complex, cuts RNA outside of the protospacer in a nonspecific (*i.e.*, sequence-independent) manner. During interference, Cas10 also releases cyclic oligoadenylates, which are small second-messenger molecules that bind to the CARF (CRISPR-associated Rossman fold) domain of Csm6 and further stimulate its nuclease activity. This remarkably sophisticated second-messenger signaling mechanism, first described in 2017 by two independent groups (lead by Virginijus Siksnys at Vilnius University and Martin Jinek at the University of Zurich, respectively) explains how Type III systems constrain the activities of Csm6 to the site of the phage infection. Interestingly, additional host-encoded nucleases are also essential for immunity through mechanisms that are still being unraveled. Unlike Types I and II systems, Type III systems do not possess PAM or seed sequences, and therefore exhibit a robust immune mechanism that cannot be easily overcome by point mutations in the phage genome. Notably, this unique feature makes Type III systems particularly suited to facilitate genome editing of lytic phages (see relevant article in “Further Reading”). However, the requirement for active transcription across the targeted region provides a route for lysogenic phages to remain unharmed by the CRISPR-Cas system provided they keep their targeted genes silent while integrating into the host chromosome.

Type IV CRISPR-Cas systems are the least well understood (Fig. 3, top panel). A recent study in 2019 by Lennart Randau and co-workers of the Max Planck Institute for Terrestrial Microbiology showed that in the Type IV CRISPR-Cas system of *Aromatoleum aromaticum*, crRNA processing is carried out by a unique Cas6 variant (Cas6v, also known as Csf5). Cas6v and the mature crRNAs are joined by three additional subunits in yet undetermined stoichiometries: Csf1 (the large Type-specific subunit), multiple copies of the Cas7 homolog Csf2, and the Cas5 homolog Csf3. The resulting ribonucleoprotein complex, termed the Type IV crRNP, has a mechanism of action that remains unknown; however, the investigators speculate that this complex functions similarly to the Type I Cascade complex.

As newer members of the Class 2 group, Types V and VI CRISPR-Cas systems have been the focal point of more recent investigations, particularly Types V-A and VI-A (Fig. 3, bottom panel). Both systems perform crRNA-mediated interference against nucleic acid invaders using single subunit effector complexes. In Type V-A systems, the large subunit is Cas12a (also known as Cpf1), while in Type VI-A systems, the large subunit is Cas13a (also known as C2c2). Both enzymes have been shown to process pre-crRNAs and cleave their targets using separate nuclease active sites. During interference, the Cas12 effector complex recognizes and cuts double-stranded DNA within the protospacer region, as well as non-target single-stranded DNA in the vicinity. In contrast, the Cas13 effector complex recognizes single-stranded RNA protospacers and subsequently degrades non-target RNA. The latter cleavage of non-target DNA or RNA upon binding a matching protospacer is termed collateral cleavage. This indiscriminate degradation of nucleic acids ensures that all invaders in the vicinity of the protospacer are shredded, and may also induce a state of dormancy in the cell due to the unintended degradation of host-derived nucleic acids. Finally, while Type V CRISPR-Cas systems have been shown to require specific PAM sequences to enable interference, Type VI CRISPR-Cas systems have a looser preference for specific nucleotides adjacent to the protospacer in a protospacer-flanking site (PFS). Thus, the latter sequence requirements may promote the facile accumulation of phage escape mutants *via* the acquisition of point mutations. It is worthwhile noting that the collateral cleavage exhibited by Types V and VI CRISPR-Cas systems has enabled the development of powerful CRISPR-based diagnostic tools. Additionally, since Type V CRISPR-Cas systems are even simpler than Type II systems with regard to their mechanisms of action, the former appear to rival the latter in their increasingly more widespread use for genome editing applications.

Anti-CRISPR Proteins

In response to the risk of encountering a CRISPR-containing host, phages have evolved a variety of countermeasures. The most common of these occurs through the passive acquisition of random mutations in the phage genome at or near the protospacer sequence. Most CRISPR-Cas systems are particularly prone to this type of escape due to presence of a seed region which is highly sensitive to mismatches, or the requirement for specific sequence motifs adjacent to it such as the PAM and PFS. Although Type III CRISPR-Cas systems do not have a seed or PAM requirement, the loss of the targeted gene in the phage constitutes a feasible mechanism of passive escape, provided the gene is dispensable for phage survival. Even though such mutations can be effective in

deterring CRISPR-Cas immunity, it is only a temporary solution since CRISPR-Cas systems can simply acquire new spacers from different parts of the phage genome to restore immunity. However, a more permanent solution to this problem lies in the diverse mechanisms of anti-CRISPR proteins.

The majority of known anti-CRISPR proteins are small, ranging in size from 50 to 150 amino acids in length. They can be found encoded in the genomes of prophages, lytic phages, or bacteria (particularly within horizontally-transferred genomic islands). Since their initial identification in 2013, reports of CRISPR inhibitors have rapidly grown to encompass over 20 distinct families targeting three of the six currently known CRISPR-Cas Types (Types I, II and V). Owing to these inhibitors' diversity in size, location, and specificity of targeting, various systematic approaches have proved necessary to elucidate each anti-CRISPR family and characterize their respective functions. The sections below highlight the seminal studies which reported the initial discoveries of anti-CRISPR proteins and the approaches that have since been applied to mine bacterial and phage genomes for these diverse CRISPR-Cas inhibitors.

Prophage-Dependent Functional Approach

Anti-CRISPRs were first reported in 2013 by a group led by Alan R. Davidson and Joseph Bondy-Denomy and co-workers at the University of Toronto, Canada. The proteins were discovered in a group of related temperate phages that infect the bacterial pathogen *Pseudomonas aeruginosa*, which contains a Type I-F CRISPR-Cas system. The authors examined a suite of 44 *P. aeruginosa* lysogens which are genetically identical with the exception of the presence of a distinct prophage within each of their genomes. Although most prophage genes are typically repressed, the authors predicted that some prophages might encode one or more genes that inhibit CRISPR-Cas function. To test for the presence of such anti-CRISPR proteins, the investigators challenged the lysogens with three CRISPR-sensitive phages that should be targeted by the bacterium's CRISPR-Cas system. As expected, the majority of the bacterial strains were able to thwart phage infection; however, a few of the prophage-containing strains failed to protect against the phages despite the presence of a functioning Type I-F CRISPR-Cas system. This suggested the lysogens harbored a prophage capable of inhibiting the CRISPR-Cas system. To identify the genetic elements responsible for this anti-CRISPR activity, the investigators compared the genomes of the CRISPR-inactivating prophages with the CRISPR-sensitive phages, and uncovered a set of five novel phage genes at a single genomic locus that were capable of inhibiting the Type I-F CRISPR-Cas system. These proteins, named AcrF1-AcrF5, constituted the very first anti-CRISPRs identified.

In a second study published in 2014, the same group identified four additional anti-CRISPRs using the same approach; however, these proteins were active against the Type I-E CRISPR-Cas system, and correspondingly were named AcrE1-AcrE4. Notably, the Type I-E inhibitors were located at the same genomic locus as the Type I-F anti-CRISPRs. The nine anti-CRISPRs identified to that date were found to exist in varying combinations across other phage genomes, suggesting that phages must acquire diverse CRISPR inhibitors to maintain their host ranges. Moreover, anti-CRISPRs were discovered outside of the prophages in genomic regions of *P. aeruginosa* that are likely mobile genetic elements, suggesting that while these proteins are subject to horizontal transfer, they are by no means restricted to transfer *via* phage.

The first insights into anti-CRISPR mechanisms were reported shortly thereafter in 2015 by members of the Davidson group. They used biochemical analyses to demonstrate that AcrF1 and AcrF2 bind directly to different protein subunits of the Type I-F effector complex and inhibit its ability to bind to the targeted DNA. In contrast, AcrF3 was unable to associate with the Cas complex and was instead shown to bind and block the function of the Cas3 nuclease-helicase. Follow-up structural analyses published in 2016 by Yongqun Zhu, Yan Zhou, and co-workers at Zhejiang University, China, showed that two copies of AcrF3 bind to each Cas3 subunit and prevent its recruitment to the Cas complex. Now, as more and more anti-CRISPR mechanisms come to light, it is becoming apparent that the majority function in two general manners: by disrupting target binding and/or inhibiting target cleavage.

Guilt-By Association Bioinformatic Approach

The lack of sequence homology between the first reported anti-CRISPRs initially hindered efforts to identify additional inhibitors through genomic comparisons alone. To bypass this challenge, April Pawluk and colleagues from the Davidson lab in 2016 considered the genomic context of the Type I-E and I-F inhibitors. They found that these anti-CRISPRs were consistently located upstream of a highly conserved anti-CRISPR associated gene, *aca1*, which encodes a predicted transcriptional regulator. Accordingly, *aca1* was used as a genomic landmark in bioinformatic screenings across phage genomes to identify diverse anti-CRISPR proteins lying adjacent to it. This method, termed a *guilt-by-association bioinformatic approach*, was used to discover two additional anti-CRISPRs acting against Type I-F systems. Furthermore, one of these was the first example of an anti-CRISPR protein capable of inhibiting two distinct CRISPR systems, Types I-E and I-F. In the same study, the team identified a second transcriptional regulator, *aca2*, adjacent to which three additional Type I-F anti-CRISPRs were identified. Subsequently, anti-CRISPR homologs were discovered across all *Proteobacteria* housing the I-F CRISPR-Cas system. Some of these proteins inhibited I-F CRISPR systems of different bacterial species despite their loosely related Cas proteins, thus underscoring the potential for anti-CRISPRs to act as broad-range inhibitors.

In a subsequent study published in 2016, a group led by Alan Davidson, Karen Maxwell, and Erik J. Sontheimer applied this guilt-by-association approach in conjunction with *aca2* to identify the first Type II CRISPR-Cas inhibitors. Originally discovered in

Neisseria meningitidis, this family of inhibitors was found to block the function of the bacterium's Type II-C CRISPR-Cas system, and thus the three inhibitors were named AcrIIC1-AcrIIC3. These findings proved significant for two reasons: First, they demonstrated that anti-CRISPRs are not limited to Type I CRISPR-Cas systems. Second, since Type II CRISPR-Cas systems are renowned for their broad genomic applications, Type II anti-CRISPR proteins have the potential to be exploited as a powerful mechanism to regulate the activities of Type II CRISPR-based technologies. Indeed, all three AcrIIC proteins discovered successfully inhibited CRISPR-Cas9 mediated genome editing within human cell lines, supporting their potential use as tools to reduce spurious cleavage events in genome editing applications. The mechanisms of action for two of the AcrIIC proteins was reported in a study led by Jennifer Doudna's group: One of these inhibitors was found to disrupt DNA binding by forcing the effector complex into a dimerized state, while the other was shown to bind and prevent Cas9 from cleaving its DNA target.

Self-Targeting Bioinformatic Approach

A distinct bioinformatics-guided method was used to identify additional anti-CRISPR genes acting against the Type II-A CRISPR-Cas system of *Listeria monocytogenes* in a 2017 study lead by Joseph Bondy-Denomy, now at the University of California in San Francisco. This methodology, termed a *self-targeting bioinformatic approach*, relies upon the prediction that the CRISPR-mediated acquisition of spacers from within a bacterium's genome would result in chromosomal cleavage by the endogenous CRISPR system and death of the cell. However, the presence of anti-CRISPR proteins within the same genome are expected to prevent such self-targeting. By identifying bacteria with self-targeting spacers, the investigators were able to discover four prophage-encoded Type II-A CRISPR-Cas inhibitors, two of which were found to block the function of the Type II-A CRISPR-Cas system of *Streptococcus pyogenes*, the very first system exploited for genome editing. The remarkable mechanism of one of these inhibitor proteins, named AcrIIA4, was elucidated in 2017 by three independent groups headed by Dinshaw Patel, Zhiwei Huang, and Jennifer Doudna, respectively. Using structural analyses, these investigators discovered that AcrIIA4 structurally mimics the PAM and binds the PAM-interacting site on Cas9, thereby blocking Cas9's ability to bind a bona fide target.

In 2018, a similar bioinformatic-guided approach was used to identify the very first Type V-A CRISPR-Cas inhibitors – these were reported in two independent studies led by Jennifer Doudna and Joseph Bondy-Denomy, respectively. Altogether, five different inhibitors were identified and experimentally confirmed to function against Type V-A interference. Notably, some of these inhibitors were shown to successfully modulate editing in human cells, again highlighting their utility for use as tools to regulate Type V-A mediated genome editing applications.

Lytic Phage Dependent Functional Approach

In 2017 and 2018, Sylvain Moineau and co-workers used a functional screen to identify additional Type II-A inhibitors within lytic phages. This approach involved engineering a Type II-A CRISPR-Cas system in *Streptococcus thermophilus* to contain spacers targeting an array of lytic phages, and then challenging the strains with each phage – those capable of reproducing within such a bacterial context were presumed to contain one or more anti-CRISPR proteins. Two such phages were identified, and further experiments in which the phages' genes were methodically tested for anti-CRISPR activity identified AcrIIA5, a broadly acting inhibitor of *S. pyogenes* and *S. thermophilus* CRISPR-Cas9 systems. Capitalizing on this successful approach, a second anti-CRISPR, AcrIIA6 was discovered despite its failure to appear concomitantly with any single *aca* gene in available phage genomes. Furthermore, homologs of AcrIIA5 or AcrIIA6 were identified in 38% of virulent *S. thermophilus* phages, indicating the prevalence of anti-CRISPRs within virulent phage genomes.

Given the widespread distribution of CRISPR inhibitors discovered within bacterial phages, it logically follows that archaeal viruses would also house anti-CRISPR genes. In fact, since 90% of sequenced *archaea* maintain CRISPR-Cas systems, it would be expected that these inhibitors might prove even more widespread amongst archaeal viruses. While anti-CRISPRs specific to archaea have yet to be thoroughly studied and characterized, a study lead by Xu Ping published in 2018 identified the first example of an anti-CRISPR protein endemic to the lytic ruidiviruses SIRV2 and SIRV3. This anti-CRISPR, AcrID1, was found to inhibit the activity of Cas10d, which is necessary for interference mediated by the Type I-D CRISPR-Cas system in *Sulfolobus islandicus*. Moreover, it was hypothesized that homologs of the newly identified archaeal anti-CRISPR are likely to be present across a diverse range of archaeal viruses as has been observed for bacterial anti-CRISPRs.

Concluding Thoughts

The microscopic war that has been raging for billions of years between prokaryotes and their viruses has given rise to an immense pool of genetic diversity, with CRISPR-Cas and anti-CRISPRs representing just two manifestations of the countless defense and counter-defense systems employed on the battlefield. Historically, the detailed investigation of such molecular weaponry has provided a fertile ground for discovery and innovation, and CRISPR-Cas research is no exception. Over the past decade, basic insights into CRISPR-Cas mechanisms have inspired the development of powerful genetic technologies, which, in turn, have propelled basic discoveries at an accelerated pace. A similar positive feedback loop is beginning to emerge for anti-CRISPR proteins, in which the continued discovery of new families and elucidation of their mechanisms are expected to not only deepen

our understanding of the evolutionary trajectories of phages and their hosts, but also impact biotechnological and biomedical fields of research. It is anticipated that in the coming years, the collection of distinct CRISPR-Cas systems will continue to expand, and anti-CRISPRs will eventually be discovered for all CRISPR-Cas Types. Beyond this, the presence of CRISPR inhibitors implies the existence of anti-CRISPR inhibitors. Though no such actors have been identified within CRISPR-containing bacterial species to date, it would not come as a surprise to see the Red Queen Hypothesis play out in this manner given such a fertile venue.

Further Reading

- Cyranoski, D., Ledford, H., 2018. International outcry over genome-edited baby claim. *Nature* 563, 607–608.
- Doudna, J.A., Charpentier, E., 2014. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346 (6213), 1258096.
- Hatoum-Aslan, A., 2018. Phage genetic engineering using CRISPR–Cas systems. *Viruses* 10 (335), 1–11.
- Hille, F., Richter, H., Wong, S.P., *et al.*, 2018. The Biology of CRISPR-Cas: Backward and Forward. *Cell* 172, 1239–1259.
- Hsu, P.D., Lander, E.S., Zhang, F., 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157 (6), 1262–1278.
- Keen, E.C., 2015. A century of phage research: Bacteriophages and the shaping of modern biology. *Bioessays* 37 (1), 6–9.
- Klompe, S.E., Sternberg, S.H., 2018. Harnessing a billion years of experimentation: The ongoing exploration and exploitation of CRISPR–Cas immune systems. *The CRISPR Journal* 1 (2), 141–158.
- Koonin, E.V., Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* 37, 67–78.
- Mohanraju, P., Makarova, K.S., Zetsche, B., *et al.*, 2016. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 353 (6299), aad5147.
- Pawluk, A., Davidson, A.R., Maxwell, K.L., 2018. Anti-CRISPR : Discovery, mechanism and function. *Nature Reviews Microbiology* 16 (1), 12–17.
- Reardon, S., 2016. The crispr zoo. *Nature* 531, 160–163.
- Stanley, S.Y., Maxwell, K.L., 2018. Phage-encoded anti-CRISPR defenses. *Annual Review of Genetics* 52, 445–464.
- Stern, A., Sorek, R., 2011. The phage-host arms race: Shaping the evolution of microbes. *BioEssays* 33 (1), 43–51.

Relevant Website

<https://www.ncbi.nlm.nih.gov/pubmed/?term=CRISPR>
Genome engineering using CRISPR-Cas9 system.

Bacteriophage: Therapeutics and Diagnostics Development

Teng-Chieh Yang, Scarsdale, NY, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Antibiotic resistance The ability of microorganisms to grow at high concentrations of an antibiotic.

Antibiotic susceptibility test The test to examine the microbial susceptibility to antibiotics of choice under varying concentrations of antibiotics.

Chemistry, manufacturing, and control The activities for development, manufacturing processes, quality control, and life cycle management of a drug product, to ensure the safety, efficacy, and manufacturing consistency of the product.

CRISPR/Cas CRISPR (clustered regularly interspaced short palindromic repeats) is a family of DNA sequences found in the genomes of prokaryotic organisms, originally derived from DNA fragments of bacteriophages that previously infected the prokaryote. Cas (CRISPR-associated protein) is a family of enzymes that use CRISPR sequences as a guide to recognize and cleave specific strands of DNA that are complementary to the CRISPR sequence. CRISPR and Cas together form the basis of a technology for editing genes within organisms.

Lysogenic cycle A life cycle of bacteriophages where most of the infecting bacteriophages integrate the genome into

the host genome, being passively replicated with bacterial genome as the host replicates. Some bacteriophage genomes exist as plasmids with separated replication and segregation systems, synchronizing the plasmid replication with host replication.

Lytic cycle A life cycle of bacteriophages where the infecting bacteriophage takes control of host cellular machinery to produce its own progeny and ultimately kills the host, releasing new bacteriophages to kill more hosts.

Mass spectrometry An analytical technique that measures the mass-to-charge ratio of ions from the analytes.

Molecular diagnostics An array of diagnostic techniques for analyzing the genomic and proteomic biomarkers using molecular biology tools.

Phage display A technology to study protein–protein, protein–peptide, and protein–nucleic acid interactions using bacteriophages as the vehicle to express proteins and connect proteins with the corresponding genetic information.

Phage typing A method to trace the source of outbreaks of bacterial infections by detecting single strains of bacteria with lytic bacteriophages.

Introduction

Antibiotics are one of the most important life-saving drugs for humankind. The human exposure to antibiotics has been dated back to 350–550 Common Era, where trace amounts of tetracycline have been found in human skeletal remains, likely due to the consumption of tetracycline-containing food. Traditional Chinese medicine practiced at least 1000 years ago was also found to have antibiotic activities. The modern “antibiotic era” started with the discovery of penicillin by Alexander Fleming in 1928 and, 15 years later, the mass production and introduction to the public for treating bacterial infections. In the next decades from 1950s to 1970s, the discovery and development of novel antibiotics flourished and revolutionized medicine by reducing global morbidity and mortality.

The bacterial resistance to antibiotics was reported shortly after their discovery and became prevalent after the massive introduction. The resistance to penicillin became widespread 24 years after the commercialization, whereas the resistance to methicillin was identified in the same year as its introduction (Fig. 1). Most of the resistance is attributed to proteins expressed in bacteria, which inactivate antibiotics via different mechanisms including efflux pumping and enzymatic cleavage. The misuse and overuse of antibiotics further accelerate the resistance by creating a selective pressure to kill the susceptible bacterial strains and allow the resistant strains to thrive. This resistance became a problem after the 1980s as bacteria more rapidly developed resistance to new antibiotics and a dramatic slow-down on the development of new antibiotics with only three new classes of antibiotics marketed after 2000 (Fig. 1).

Over the last decade, Antibiotic Resistant (AR) bacteria are recognized as global threats and projected to cause more annual global deaths than cancer by 2050. Each year in the US, more than 2.8 million people are infected with AR bacteria, causing at least 35,000 people to die directly from the infection. The annual nationwide cost to treat hospitalized patients with AR bacterial infections is estimated between \$30 and \$50 billion (between 2012 and 2015). These infections, if not diagnosed early and treated with effective drugs, can lead to severe diseases like sepsis which has a high mortality.

Current standard diagnostic method for bacterial infections is to amplify bacterial pathogens from culture (i.e., bacterial culture), followed by the species identification with biochemical tests or *mass spectrometry* analysis. Upon identifying the pathogens, a separate *Antibiotic Susceptibility Test* (AST) is used to select effective treatments with the total testing time of two to five days from sampling to obtaining AST results. Patients with acute infections (e.g., bloodstream infection) cannot afford such a prolonged testing time, thus physicians have no choice but to empirically treat these patients with broad-spectrum antibiotics before the test results would be available, which accelerates the resistance to last-resort drugs. Several *molecular diagnostics* emerged for rapid bacterial identification;

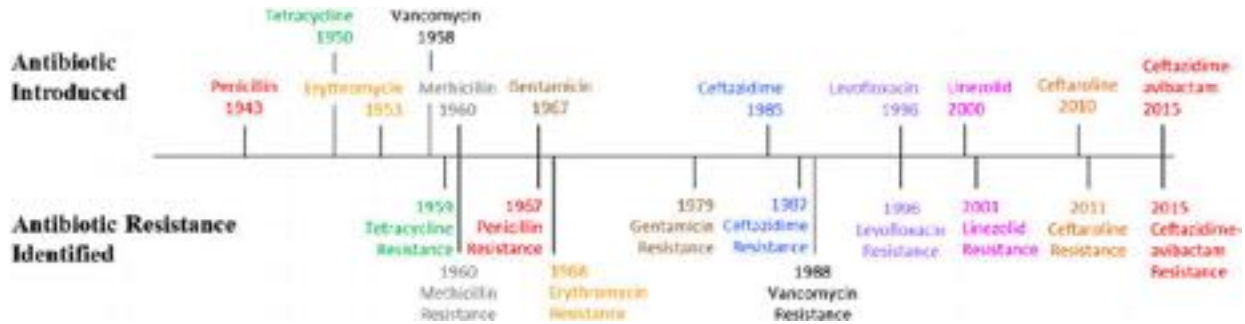


Fig. 1 Timeline of antibiotic introduction and the identification of antibiotic-resistant bacteria from 1940 to 2015. Data modified from Center for Disease and Control and Prevention (<https://www.cdc.gov/drugresistance/about.html>).

however, all of them still require a preliminary bacterial culture, a process that takes at least 8–12 h to amplify culturable bacteria. It is almost impossible to detect bacteria that are difficult to culture or are unculturable. Therefore, we need to develop a rapid and accurate diagnostic tool and combine with evidence-based treatments to combat AR bacteria.

In the past decade, bacteriophages (Phages) have gained attention from scientific community as an alternative therapy for AR bacterial infections. Phages are viruses that infect bacteria with two different life styles, *lytic* and *lysogenic* cycles. In the lytic cycle, the infecting phage takes control of host cellular machinery to produce its own progeny and ultimately lyse (i.e., kill) the host bacteria, releasing new phages to kill more bacteria. In the lysogenic cycle, most of the infecting phages become a *prophage* by integrating the genome into the host bacterial genome, being passively replicated with bacterial genome as the host replicates. If a bacterium with a prophage is under environmental stresses (e.g., UV), the prophage may be excised from the host genome and enter the lytic cycle, a process called induction.

Phages hold great potential to be a new tool for detecting and treating AR bacterial infections because of two major advantages: *specificity* and *self-replication*. However, there are at least two major limitations: *narrow host range* and *resistance of the bacteria to phages*. Many excellent reviews describing the medical application of phage have been published in the last 10 years. The goal of this article is to discuss the history of phage development as therapeutic and diagnostic tools, to dissect the major advantages and limitations of phage technology associated with therapeutic and diagnostic applications.

The Development of Phage as Therapeutics for Bacterial Infections

Phages were independently discovered by two microbiologists: Frederik Twort in 1915 and Félix d'Hérelle in 1917, and named "bacteriophage" after the idea of "bacteria-eater." Shortly after d'Hérelle's discovery of *Shigella* phages, he acknowledged the potential of using phages to treat bacterial infections (i.e., phage therapy) and successfully applied lytic phages to treat *Salmonella gallinarum* infected chickens. This success led to the first human clinical trial in 1921, where five patients with bacillary dysentery were treated with *Shigella* phages and recovered. In 1927, additional clinical trials were conducted in India for treating cholera with phages, which showed an over 50% decrease in mortality compared to the control groups.

After the initial success of clinical trials, more scientists began to test phage therapy for other bacterial infections. The outcome of these tests, however, was mixed success and raised criticisms on the study design and the quality of phage products. The enthusiasm of phage therapy waned after the discovery and mass production of penicillin gained popularity in 1940s, shifting the pharmaceutical focus to finding and developing novel antibiotics as the treatment for bacterial infections. Research on phage instead moved towards the fundamental understanding of phage biology with several phages (e.g., λ , T4, and M13) became powerful model systems for studying gene regulation, protein engineering, restriction modification, and among others, transforming the knowledge learned from phages into modern biology disciplines including molecular biology and genetics.

This shift of focus continues to bring the technologies developed from phages at the center of modern drug development, including the introduction of "phage display" technology in 1985, which created bestselling protein therapeutics and the more recent discovery of "CRISPR/Cas system" which holds a great potential to be the next generation precision therapy by its targeted genome editing capability. Despite the success from phage-derived technologies, phage therapy has received mixed interests around the globe. From the 1940s to 1990s, Western countries largely abandoned the idea of phage therapy, whereas the former USSR still extensively exploited its capability through the 1970s. Today, phage therapy is still available as a treatment option for AR bacterial infections in the Republic of Georgia.

A notable phage therapy center in the Republic of Georgia, Eliava Institute, has developed several phage products for treating septic, topical, and intestinal bacterial infections. Despite these products were subjected to clinical trials and are commercially available in Eastern Europe, they are not approved elsewhere, because the standards of clinical trials and pharmaceutical manufacturing procedures in the former USSR and the Republic of Georgia do not comply with international regulations. With the emergence of multi-drug resistant bacterial strains and the slowed development of new antibiotics, Western countries have a renewed interest in phage therapy.

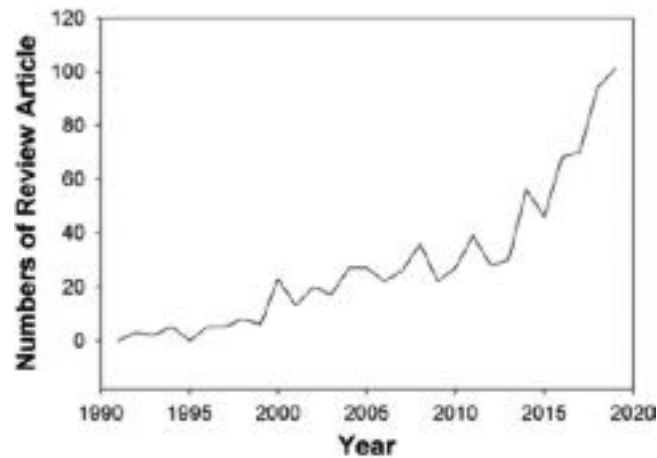


Fig. 2 Interests of phage therapy indicated by the number of review article from 1991 to 2019 (data extracted from PubMed).

Table 1 Overview of randomized and placebo-controlled clinical trials for phage therapy

Industry sponsor	Indication	Target bacterial species	Phase	Study number	Year	Product	Number of phages in the product	Evaluated patient population	References
Intralytix	Venous leg ulcers	<i>Pseudomonas aeruginosa</i> , <i>Staphylococcus aureus</i> , <i>Escherichia coli</i>	I	WIRB, protocol# 20061649	2007–2008	WPP – 201	8	39	Rhoads <i>et al.</i> (2009)
Biocontrol Limited	Chronic otitis	<i>Pseudomonas aeruginosa</i>	I/II	2004–001691–39	2009	T4-like coliphages	6	24	Wright <i>et al.</i> (2009)
Nestle	Pediatric diarrhea	<i>Escherichia coli</i>	I	NCT00937274	2016	Biophage-PA	11	115	Sarker <i>et al.</i> (2016)
Pherecydes Pharma SA	Burn wounds	<i>Pseudomonas aeruginosa</i>	I/II	2014–000714–65	2018	PP1131	12	25	Jault <i>et al.</i> (2019)

This interest gradually grew as indicated by the steady increase of published review articles discussing phage therapy in Western scientific journals (Fig. 2). Additionally, four randomized, placebo-controlled Phase I/II clinical trials were conducted to address one of the biggest criticisms on phage therapy – suboptimal design of clinical trials (Table 1). All four trials pointed towards promising safety; however, three failed to demonstrate the superiority of phage treatment compared to controls or standard of care. In contrast, several case studies of compassionate treatments showed positive outcomes. Phages appear to have a niche in treating some bacterial infections but still have many unsolved questions from the fundamental mechanisms of action to the manufacturing and regulatory pathway. More high quality data is needed to turn phage therapy into the reality.

The Development of Phage as Diagnostics for Bacterial Infections

Current diagnostic tools for bacterial infections primarily rely on bacterial culture combined with (1) secondary biochemical tests or mass spectrometry analysis to identify the pathogen species or (2) molecular diagnostics to detect species-specific and bacterial virulence genes. Both methods require at least 8–12 h to amplify the pathogens and a separate AST for selecting treatments. An ideal diagnostic tool should only detect viable bacteria, identify species, and provide treatment recommendations from a single test. If test results were available in one to three hours, physicians could initiate evidence-based treatments and minimize the empirical use of broad spectrum antibiotics.

Phages have been used to detect bacteria since the development of *phage typing* in 1950s, where bacterial strains can be typed (i.e., classified) by the plaques created from a panel of lytic phages on the bacterial lawn. This technique is still used in some hospitals to monitor the bacterial outbreaks by detecting clusters of virulent strains. A reporter phage was later developed to detect *Mycobacterium tuberculosis*, a slow growing bacterial species, by incorporating a luciferase gene into the phage genome. After phage infection, the luciferase is expressed in minutes with the activity to be detected by the addition of luciferin, dramatically reduced the diagnostic time from weeks to one day. Importantly, this reporter phage is compatible with AST, allowing the selection of effective antibiotic treatments and the screen of new anti-tuberculosis drugs.

With a better understanding of phage genomics and structure, the engineering effort turned to the use of phage capsids to display functional molecules on the surface of phage virions, enabling the detection of phage-infected bacteria either through conjugated fluorophores or direct fluorescent signal from the displayed molecules (Table 2). Despite extensive research on

Table 2 Summary of modern diagnostic phages

Technology	Target bacteria	Diagnostic phage	Signal detection	Limit of detection	Detection time	Compatible with AST	References
Luciferase	<i>Mycobacterium tuberculosis</i>	TM4	Luminescence	500–5000 cells	<24 h	Yes	Jacobs Jr. <i>et al.</i> (1993)
Fluorescent protein on phage	<i>E. coli</i>	T4	Fluorescence	NA	1 h	NA	Tanji <i>et al.</i> (2004)
Biotinylation coupled with quantum dots	<i>E. coli</i>	T7	Fluorescence	10 cells/mL	1 h	NA	Edgar <i>et al.</i> (2006)
Fluorescent protein in bacteria	<i>Mycobacterium tuberculosis</i>	TM4	Fluorescence	<100 cells	<24 h	Yes	Piuri <i>et al.</i> (2009)
Luciferase	<i>E. coli</i> O157:H7	ΦV10	Luminescence	5 CFU	7 h	NA	Zhang <i>et al.</i> (2016)
Biotinylation coupled with fluorescence nanodiamonds	<i>E. coli</i>	λ	Fluorescence	NA	NA	NA	Trinh <i>et al.</i> (2018)

Note: NA, not available.

diagnostic phages, very few products of this type have moved into clinical testing or become commercially available, owing to some practical challenges associated with phages. Below we discuss the major advantages and limitations of phage technology.

Major Advantages of Phage Technology

Specificity

Although broader host range phages do exist, more commonly a phage species is typically limited to infecting one bacterial species or even to several lineages within a species. This specificity is driven by *phage adsorption*, a process generally consisting of three steps, initial contact, reversible binding, and irreversible attachment to the target bacteria. The last step is mediated by a *specific* interaction between the often bacterial species-specific surface receptor and the receptor binding protein on the phage virion, allowing the accurate detection of target bacteria for diagnostic phages and the minimum impact on commensal bacteria for therapeutic phages.

Self-Replication

Lytic phages replicate inside the target bacteria, kill the bacterial host, and release dozens to hundreds of progenies, continuing the bacterial killing. This self-replicating nature allows the sustained release of therapeutic phages from the target bacteria, potentially reducing the dosing amount and frequency. As the phages only replicate in *live* bacteria, this process enables the detection of live pathogens when infected with diagnostic phages.

Major Limitations of Phage Technology

Narrow Host Range

A phage's host range is defined as the spectrum of bacteria that it infects. Some phages have a narrow host range and only infect few strains in a bacterial species, limiting their ability to treat or diagnose infections caused by single or multiple bacterial species. While this limitation has been overcome with the use of multiple phage species to expand the range of host targets, new challenges emerge for therapeutic applications, as discussed later in this article.

Phage Resistance

Bacteria and phages co-exist in an evolutionary arm race, where a bacterial species may already have or will inevitably develop resistance against selected phages. The pre-existing phage resistance will increase the false negative results from diagnostic applications and decrease efficacy in therapeutic applications, while the acquired phage resistance may be difficult to monitor and control in human for therapeutic phages.

Additional Challenges for Phage Therapy

Preclinical Data Translation

Translating results from preclinical models to human is a significant and common challenge for drug development. Early studies showed that the outcome of in vitro phage experiments did not always give the same results as in vivo studies. Additionally, the

in vivo efficacy from one study may not be consistent with another study or translatable to the efficacy in humans. This is largely due to the poor study design and the discrepancies in biology between species. High quality preclinical data from in vitro and representative in vivo models is essential for the approval of phage therapy.

Safety and Efficacy

Phages are generally considered to be safe to human because no major adverse events were reported from previous clinical trials. In animal models, circulating phages are also shown to be rapidly removed by liver and spleen in the absence of target bacteria. Recent studies suggested that, while phages can enter human cells via at least three mechanisms, phagocytosis, transcytosis, and receptor-mediated endocytosis, the entered phage virions were degraded in one day without damaging the viability of cells. However, it is not clear how phages interact with human immune cells. Such interaction may trigger immunogenicity with specific responses against therapeutic phages, leading to the development of anti-phage antibodies. The clinical impact of these antibodies ranges from no apparent effect, to the reduced efficacy of phages by neutralizing and/or eliminating phages from the body, to the safety concern from the induced allergic responses. To ensure the safety and efficacy of phage therapy, more studies are needed to understand the potential immunogenicity triggered by phages.

Dosing Strategies

To maximize the therapeutic activity of phage, several dosing strategies should be considered, *dose determination*, *dose interval*, and *route of administration*, to improve the common pharmacology effects, (1) Pharmacokinetics: the fate of phages after being administered to humans, and (2) Pharmacodynamics: the impact of administered phages on humans. These pharmacology effects are more complex for phage therapy compared to antibiotics or protein therapeutics because of the population (from self-replicating nature of phages) and evolutionary dynamics between phages and bacteria. Several population dynamic models were developed to predict the short-term dynamics (i.e., days) for in vitro phage-bacteria interactions; however, it is still challenging to model these interactions in humans in a clinically relevant time frame (i.e., weeks to months). Additionally, most of the models were developed for systems with a single phage and a single bacterial species, there is limited understanding on systems with multiple phage (i.e., phage cocktail) and bacterial species.

A unique aspect to consider for phage therapy is the *dosage form*: single phage versus phage cocktail. Phage cocktails were developed to expand the host range and overcome phage resistance from bacterial mutants. While the cocktail sometimes outperforms the single phage treatment, it remains difficult to determine the mechanisms of action and the associated synergy (if any) from the cocktail. In-depth understanding of population and evolutionary dynamics among cocktail phages and the target bacteria may facilitate a more rational design of phage cocktails and prevent phage resistance.

The rise of bacterial phage resistance in human clinical trials has been shown to be alleviated by subsequent treatments with different phages selected as able to propagate on the phage-resistant bacteria. This selection process typically continues until the patient showed no symptoms, making the entire dosing process “personal”, as the next dosing regimen depends on the bacteria evolved in a given patient. The major advantage of this process is the use of target phages. However, the selection process is time-consuming and labor intensive with several manufacturing and regulatory challenges, as discussed below.

Chemistry, Manufacturing, and Controls (CMC)

CMC are the activities for the development, the manufacturing processes, the quality control, and the life cycle management of a drug product, to ensure the safety, efficacy, and manufacturing consistency. Several CMC challenges were identified for phage therapy. (1) High levels of bacterial endotoxin in the drug product and loss of infectivity during the shelf-life storage were reported in phages used in recent clinical trials, raising the concerns of product quality. (2) Phage therapy has been evaluated as the personalized treatment process (i.e., personalized phage products over the course of treatment) or as the single off-the-shelf product (i.e., fixed composition of phages in a single drug product). The off-the-shelf products have a clear CMC pathway that is likely to be similar to biologics; however, bacteria will eventually develop phage resistance. On the other hand, the personalized treatments may be more effective in reducing phage resistance; however, there is no clear general strategy to control the product quality and establish shelf-life in a timely manner for each treatment cycle.

Regulatory Pathway

There is no approved phage therapy in Western countries; therefore, the regulatory pathway is not clear. To engage with phage therapy developers, the US Food and Drug Administration and National Institutes of Health hosted workshops to discuss current drug approval processes and other considerations for phage therapy (e.g., phage therapy workshop, Rockville MD, 2017). Experts from both agencies also participated in public phage conferences to share their perspective on phage therapy (e.g., phage congress, DC, 2019). The message from these agencies suggested that, in general, the regulatory pathway will be similar to biologics, and there is a need for more high quality preclinical and clinical data to demonstrate the safety, efficacy, and manufacturing consistency of therapeutic phages.

Environmental Impact

As shown in the antibiotic industry, the mass production (> 100,000 tons per year) causes overwhelming pollution in the environment and accelerates antibiotic resistance. Mass production of phages will inevitably create a similar level (if not worse) of environmental impact from the toxic byproduct during manufacturing to the evolutionary pressure on natural bacterial populations. If phage therapy becomes widely available for treating bacterial infections, it will be essential to have proper regulation of phage production and surveillance programs dedicated to monitor the environmental impact and avoid the unintended consequences from antibiotic industry.

Additional Challenges for Phage Diagnostics

Phage Manipulation

To improve sensitivity, modern diagnostic phages were developed by the genetic manipulation of phage genome, which requires a delicate selection of the target location on the phage genome for editing and the discovery of new functional molecules to improve the sensitivity. Poor engineering strategy will result in damaged phages with little infectivity compared to the wild type phages.

Multiplex Testing

Many bacterial infections are caused by different species; therefore, the ability to test multiple bacterial species (i.e., multiplex testing) becomes critical for improving the diagnostic efficiency and reducing the sample requirement. Current diagnostic phages were designed to target single bacterial species with single channel of detection on the functional molecules. With limited functional molecules (e.g., fluorophores) available for engineering, it is challenging to apply diagnostic phages for multiplex testing.

Intracellular Bacteria

Several bacterial pathogens replicate inside human cells and transport between neighboring cells, which are often difficult to identify unless they outbreak and cause symptoms. Moller-Olsen and co-workers recently showed that K1F phage can enter human epithelial cells and kill the intracellular *E. coli*. It is not clear if this reaction is universal for different human cells, phages, and bacterial species. More studies at clinically relevant conditions are needed to understand the potential of using diagnostic phages to detect intracellular bacteria.

Sample Interference

Sample interference is a common issue of traditional bacterial culture and modern molecular diagnostics, where the interference may come from contaminating bacteria picked up during sample collection or from the sample matrix, impacting the test sensitivity, specificity and predictive values. Some interference may be partially mitigated by repeated tests; however, not all patients (e.g., neonate) can afford the samples for re-testing. Again, future work should address the performance of diagnostic phages in clinically relevant conditions and compare the assay performance to current standard techniques.

Market Acceptance

Bacterial culture has been utilized for several decades as the gold standard for bacterial diagnosis; therefore, customers may be reluctant to switch to other methods. Current diagnostic phages rely on the luminescence detection or the fluorescent imaging analysis; both are low throughput and labor intensive. To compete with bacterial culture, the diagnostic phages should be developed as a high throughput assay with a minimum labor requirement and compatible with current diagnostic workflow in the hospitals and clinical testing laboratories.

Conclusions

Interest in using phages to fight AR bacterial infections remains high. As noted above, the emergence of AR bacteria and the shortage of new antibiotic development prompt the search for alternatives. Phages are approved as prophylactic agents for bacterial contamination in food products in the US and several other countries; therefore, they have the potential to be the alternative diagnostics and therapeutics for bacterial infections. The emergence of successful phage treatments of patients who were otherwise untreatable with antibiotics has gained attention from the media and ignited passion to pave the regulatory approval path for phage therapy.

However, challenges remain at every level of development. For the single phage product, bacterial phage resistance is inevitable. It is not clear how fast bacteria will develop such a resistance and the impact of phage resistance on bacterial virulence. Will

co-evolutionary arm race between phages and bacteria exacerbate bacterial virulence or reduce bacterial fitness? For phage cocktails, the mechanism of bacterial killing is not clear as the individual and synergistic efficacy is difficult to dissect. An important question remains: how to rationally design the treatment with a single phage product, a phage cocktail, or a combination of phage(s) with antibiotics? We also have limited knowledge of human immune responses to phages and similarly the interaction between human cells and phages. Will phages transduce virulent genes from a virulent bacterial host to non-virulent bacteria or even human cells? The limited data available makes it challenging to define the clinical safety and efficacy profile for phage therapy.

While there is tremendous body of work in biochemistry, structural, and genetic aspects of phages, the data is more concentrated on selected model systems. Every phage species is unique with its own properties and should be treated as an individual entity in the product. To produce a single phage product or a phage cocktail, one must monitor the quality attributes of every phage species during the manufacturing processes and the shelf-life. How to define the shelf-life and control the quality of phages are critical CMC issues for phage therapy and depend on the dosage form (cocktail versus single phage product) and the treatment process (off-the-shelf versus personalized). Finally, there is an urgent need for high quality studies to evaluate the clinical efficacy with *sufficient patient population* and the performance of diagnostic phages in *clinically relevant conditions*, which is at present hard to come by. Once these challenges are solved, phages can be considered as the alternative tools to fight AR bacterial infections.

Further Reading

- Abedon, S.T., Kuhl, S.J., Blasdel, B.G., Kutter, E.M., 2011. Phage treatment of human infections. *Bacteriophage* 1, 66–85.
- Bassett, E.J., Keith, M.S., Armelagos, G.J., Martin, D.L., Villanueva, A.R., 1980. Tetracycline-labeled human bone from ancient Sudanese Nubia (A.D. 350). *Science* 209, 1532–1534.
- Brussow, H., 2005. Phage therapy: The *Escherichia coli* experience. *Microbiology* 151, 2133–2140.
- Davies, J., Davies, D., 2010. Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews* 74, 417–433.
- Dedrick, R.M., Guerrero-Bustamante, C.A., Garlena, R.A., *et al.*, 2019. Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant *Mycobacterium abscessus*. *Nature Medicine* 25, 730–733.
- Edgar, R., McKinsty, M., Hwang, J., *et al.*, 2006. High-sensitivity bacterial detection using biotin-tagged phage and quantum-dot nanocomplexes. *Proc. Natl. Acad. Sci. USA* 103, 4841–4845.
- Gandra, S., Barter, D.M., Laxminarayan, R., 2014. Economic burden of antibiotic resistance: How much do we really know? *Clinical Microbiology and Infection* 20, 973–980.
- Jacobs Jr., W.R., Barletta, R.G., Udani, R., *et al.*, 1993. Rapid assessment of drug susceptibilities of *Mycobacterium tuberculosis* by means of luciferase reporter phages. *Science* 260, 819–822.
- Jault, P., Leclerc, T., Jennes, S., *et al.*, 2019. Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by *Pseudomonas aeruginosa* (PhagoBurn): A randomised, controlled, double-blind phase 1/2 trial. *Lancet Infect. Dis.* 19, 35–45.
- Keen, E.C., 2015. A century of phage research: Bacteriophages and the shaping of modern biology. *Bioessays* 37, 6–9.
- Kortright, K.E., Chan, B.K., Koff, J.L., Turner, P.E., 2019. Phage therapy: A renewed approach to combat antibiotic-resistant bacteria. *Cell Host and Microbe* 25, 219–232.
- Krueger, A.P., Scribner, E.J., 1941. The bacteriophage: Its nature and its therapeutic use. *Journal of the American Medical Association* 116, 2269–2277.
- Kutateladze, M., Adamia, R., 2008. Phage therapy experience at the Eliava Institute. *Médecine et Maladies Infectieuses* 38, 426–430.
- Lehti, T.A., Pajunen, M.I., Skog, M.S., Finne, J., 2017. Internalization of a polysialic acid-binding *Escherichia coli* bacteriophage into eukaryotic neuroblastoma cells. *Nature Communications* 8, 1915.
- Levin, B.R., Bull, J.J., 2004. Population and evolutionary dynamics of phage therapy. *Nature Reviews Microbiology* 2, 166–173.
- Lino, C.A., Harper, J.C., Carney, J.P., Timlin, J.A., 2018. Delivering CRISPR: A review of the challenges and approaches. *Drug Delivery* 25, 1234–1257.
- Merril, C.R., Biswas, B., Carlton, R., *et al.*, 1996. Long-circulating bacteriophage as antibacterial agents. *Proceedings of the National Academy of Sciences of the United States of America* 93, 3188–3192.
- Moller-Olsen, C., Ho, S.F.S., Shukla, R.D., Feher, T., Sagana, A.P., 2018. Engineered K1F bacteriophages kill intracellular *Escherichia coli* K1 in human epithelial cells. *Scientific Reports* 8, 17559.
- Nixon, A.E., Sexton, D.J., Ladner, R.C., 2014. Drugs derived from phage display: From candidate identification to clinical practice. *MAbs* 6, 73–85.
- Pires, D.P., Cleto, S., Sillankorva, S., Azeredo, J., Lu, T.K., 2016. Genetically engineered phages: A review of advances over the last decade. *Microbiology and Molecular Biology Reviews* 80, 523–543.
- Piuri, M., Jacobs Jr., W.R., Hatfull, G.F., 2009. Fluoromycobacteriophages for rapid, specific, and sensitive antibiotic susceptibility testing of *Mycobacterium tuberculosis*. *PLoS One* 4, e4870.
- Powers, J.H., 2004. Antimicrobial drug development – The past, the present, and the future. *Clinical Microbiology and Infection* 10 (Suppl 4), 23–31.
- Rhoads, D.D., Wolcott, R.D., Kuskowski, M.A., *et al.*, 2009. Bacteriophage therapy of venous leg ulcers in humans: Results of a phase I safety trial. *J. Wound Care* 18 (237–238), 240–243.
- Salmond, G.P., Fineran, P.C., 2015. A century of the phage: Past, present, and future. *Nature Reviews Microbiology* 13, 777–786.
- Sarker, S.A., Sultana, S., Reuteler, G., *et al.*, 2016. Oral phage therapy of acute bacterial diarrhea with two coliphage preparations: A randomized trial in children from Bangladesh. *EBioMedicine* 4, 124–137.
- Schooley, R.T., Biswas, B., Gill, J.J., *et al.*, 2017. Development and use of personalized bacteriophage-based therapeutic cocktails to treat a patient with a disseminated resistant *acinetobacter baumannii* infection. *Antimicrobial Agents and Chemotherapy* 61.
- Tanji, Y., Furukawa, C., Na, S.H., *et al.*, 2004. *Escherichia coli* detection by GFP-labeled lysozyme-inactivated T4 bacteriophage. *J. Biotechnol.* 114, 11–20.
- Trinh, J.T., Alkhatani, M.H., Rampersaud, I., *et al.*, 2018. Fluorescent nanodiamondbacteriophage conjugates maintain host specificity. *Biotechnol. Bioeng.* 115, 1427–1436.
- Wright, A., Hawkins, C.H., Anggard, E.E., Harper, D.R., 2009. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa*: A preliminary report of efficacy. *Clin. Otolaryngol.* 34, 349–357.
- Zhang, D., Coronel-Aguilera, C.P., Romero, P.L., *et al.*, 2016. The use of a novel NanoLuc -based reporter phage for the detection of *escherichia coli* O157:H7. *Sci. Rep.* 6, 33235.

Bacteriophage Vaccines

Pan Tao and Venigalla B Rao, The Catholic University of America, Washington, DC, United States

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Vaccines are one of the most effective medical interventions in the human history. More than 60 vaccines against 26 pathogens are currently available in the US market. Most of these are whole-pathogen vaccines (CDC data updated on April 13, 2018, for website link see “Relevant Website section”). Historically, vaccines were developed using mainly the whole pathogen, either the attenuated (live-attenuated vaccines) or the inactivated (heat- or formalin-killed vaccines) organism. However, whole-pathogen vaccines generally pose significant safety risks including reversion to a pathogenic form, severe reactions in immunocompromised hosts, and adverse effects such as allergic responses. Therefore, the focus in recent years has shifted to the development of recombinant subunit vaccines, which contain only the well-defined antigenic molecules of the pathogens, which are a much safer alternative to the whole-pathogen vaccines.

Virus-like particles (VLPs) represent one of most promising approaches to design subunit vaccines. VLPs are nanometer scale particles with precise dimensions that are assembled with viral structural proteins. However, they generally lack the viral genome and hence, non-pathogenic. Due to their favorable characteristics as seen in viral pathogens such as size, geometry, highly ordered and repeat structure, they can, by simulating a viral infection, stimulate robust immune responses by the host immune system. Successful VLP vaccines have been developed and commercialized worldwide such as hepatitis B vaccine and human papillomavirus vaccine.

It would be extremely useful if there is a “universal” platform that could assemble any antigen, or antigens, and generate VLP vaccines against one or multiple pathogens. Bacteriophages (phages) are highly promising candidates to develop such a platform because of their size, surface structure, stability, safety, biodegradability, and low cost of manufacture. Here, we describe some unique features of phage T4 that make it an effective VLP vaccine platform, both in terms of design and effectiveness.

Architecture of Phage T4

The phage T4 belongs to *Myoviridae* family and infects the *E. coli* bacterium. Thus, it is innocuous to humans. Phage T4 consists of three major components; head (capsid), tail, and tail fibers. These are assembled by independent pathways and joined together to form the infectious virion. The application of T4 phage as a VLP platform mainly involves the head, which is an elongated icosahedron, 120 nm long and 86 nm wide (Fig. 1). It is built with 930 copies of the major capsid protein gp23* (“*” represents the cleaved form) (48.7 kDa), which form the hexagonal capsid lattice. Eleven of the twelve vertices are occupied by pentamers of the vertex protein gp24* (55 copies per capsid) (46 kDa). The twelfth vertex is a unique vertex where the dodecameric portal protein (61 kDa) resides. A central channel in the dodecamer serves as a portal for entry of DNA during genome encapsidation and for exit of genome during infection. The dodecameric vertex is where the packaging motor docks for DNA packaging and later, neck and tail attach after packaging is terminated and motor dissociated. A unique feature of the T4 head is that it contains two non-essential outer capsid proteins; 870 copies of the small outer capsid protein (Soc, 9.1 kDa) and 155 copies of the highly antigenic outer capsid protein (Hoc, 39 kDa) (Fig. 1). Soc monomers assemble into trimers at the quasi three-fold axes and clamp the adjacent gp23* hexameric capsomers, thus reinforcing an already very stable capsid structure and protect the pressurized capsid against harsh extracellular environment (e.g., pH 11). Hoc is a linear “fiber” containing a string of four domains. It binds at the center of each capsomer as a monomer through its COOH-terminal domain whereas the NH₂-terminus is projected at ~170 Å distance away from the capsid. Hoc might facilitate attachment of phage to bacteria.

Phage T4 as a Vaccine Platform

Although Soc and Hoc provide survival advantages to the virus in the natural environment, they are completely dispensable under laboratory conditions. Deletion of these genes has no significant effect on phage assembly, productivity, or infectivity. Moreover, purified Soc or Hoc proteins bind to Hoc⁺Soc⁺ capsid with high specificity and nanomolar affinity, properties that are not greatly compromised by attachment of an antigen at the NH₂- and/or COOH-termini. Actually, the NH₂- and COOH-termini of Soc and the NH₂-terminus of Hoc are well exposed on the capsid structure, thus allowing efficient display of foreign proteins fused to Soc and Hoc on the capsid surface. Proteins as large as ~83 kDa anthrax protective antigen (PA) or 129 kDa β -galactosidase were efficiently displayed. However, the COOH-terminal domain of Hoc has the capsid binding site and is not desirable for antigen display. The T4 antigen display can be accomplished either *in vivo* or *in vitro*.

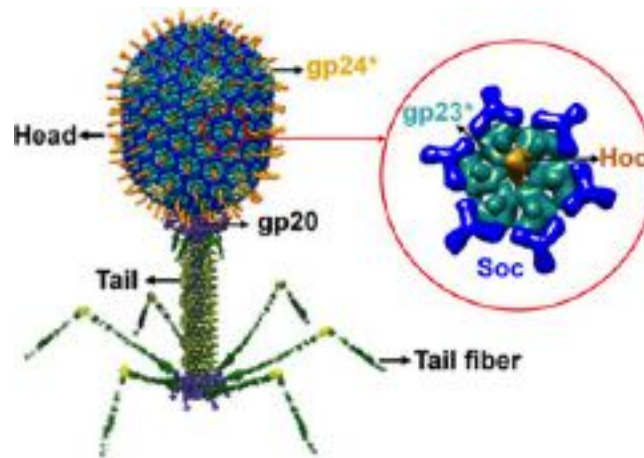


Fig. 1 Structural model of bacteriophage T4. Three major components of phage T4 – Head, tail, and tail fibers – are indicated. The enlarged capsomer (inset) shows the major capsid protein gp23* (930 copies; * represents the cleaved form), Soc (blue, 870 copies), and Hoc (orange, 155 copies). Yellow subunits at the five-fold vertices correspond to gp24*. The unique portal vertex (gp20) connects the head to the tail.

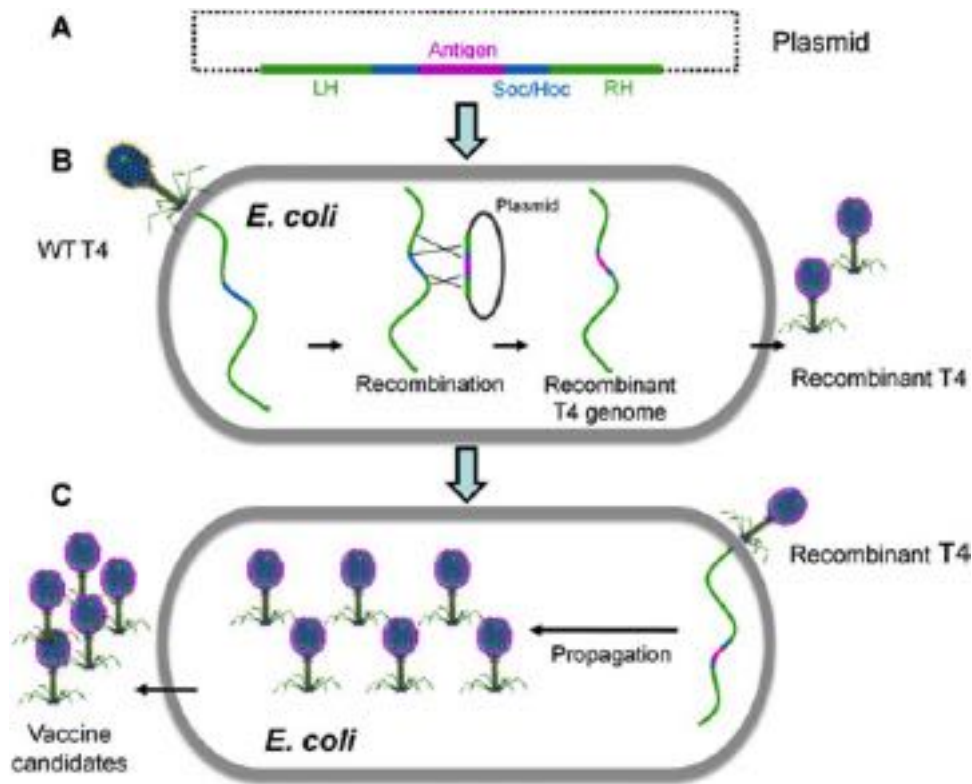


Fig. 2 Schematic of phage T4 *in vivo* VLP display system. A. The Soc- or Hoc-antigen fusion gene is cloned into a donor plasmid flanked by homologous arms. B. The recombinant T4 phage is generated by homologous recombination between wild-type (WT) T4 and donor plasmid upon infection of *E. coli*. C. The fusion proteins are expressed under the control of a T4 promoter and assembled on T4 capsids during the propagation of the recombinant T4 phages.

In Vivo VLP Assembly

For *in vivo* VLP assembly, the antigen gene is inserted into T4 genome at the NH₂- or COOH-termini of Soc or the NH₂-terminus of Hoc by homologous recombination between T4 phage genome and a donor plasmid, which contains an antigen gene flanked by two homologous arms of T4 genome (Fig. 2). The recombinant T4 phage containing the antigen-Soc or -Hoc fusion gene is then selected and amplified. Upon infection of *E. coli*, the fusion proteins are expressed under the control of a T4 promoter and

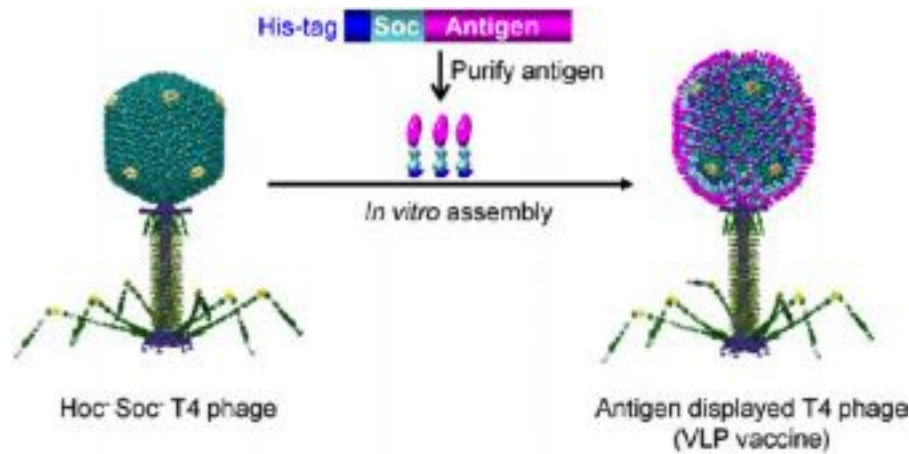


Fig. 3 Schematic of phage T4 *in vitro* VLP display system. The polyhistidine-tagged Soc-fused antigen is purified from *E. coli* and assembled on purified Hoc⁺Soc⁺ T4 phage by mixing the two to generate the VLPs. The same principle is used for the display of Hoc-fused antigens or targeting molecules.

assembled on the capsid *in vivo*. The resulting phage is purified and used as a VLP vaccine. The advantage of *in vivo* assembly is that, once the recombinant phage is constructed, the VLP vaccines can be produced by a simple procedure at the scale needed. However, the copy number of the antigen on the VLP could vary greatly because different Soc or Hoc fusion proteins may be expressed to different extents and the Soc/Hoc binding sites don't have to be completely occupied. Additionally, it would be difficult to assemble multiple antigens to generate multivalent vaccines.

In Vitro VLP Assembly

For *in vitro* VLP assembly (Fig. 3), Soc- or Hoc-antigen fusion proteins are expressed in *E. coli* under the control of a strong promoter such as the phage T7 promoter. The proteins are then purified and displayed on phage by mixing the recombinant protein with the purified Hoc⁺Soc⁺ T4 phage. The Soc- or Hoc-fusion proteins can also be expressed using a mammalian expression system for antigens requiring specific post-translational modifications. This approach has many advantages. First, functional characterization of the purified Soc- or Hoc-fusion proteins can be performed prior to display to insure the production of full-length antigens and their native-like function(s). Second, the composition of the *in vitro* assembly mixture can be adjusted to generate multivalent vaccines against one or more pathogens and to control the copy number (Fig. 3). For example, by adding two different anthrax antigens fused to Soc, LFn-Soc, and Soc-PA4 to the reaction mixture, T4 phage nanoparticles simultaneously displaying both the antigens could be generated. The copy number of each antigen was proportional to the ratio of antigen molecules to binding sites added to the assembly mixture. Third, *in vitro* assembly allows building of oligomeric complexes through interaction with a displayed antigen. The tripartite anthrax toxin complexes containing LFn, PA, and EF could be assembled using the Soc-fused LFn molecules anchored on the capsid. Finally, *in vitro* assembly allows display of antigens that require post-translational modifications such as glycosylation for function by simply producing the antigens using a mammalian expression system.

T4 capsid nanoparticle has several features that make it a desirable platform to design VLP vaccines. The availability of up to 1025 Soc and Hoc binding sites per capsid provide a great deal of flexibility to generate symmetrically arrayed antigens on a nanoparticle at high density. Such ordered and repetitive patterns of display on phage capsid mimic the patterns exhibited by pathogenic viruses and are considered to be important for activation of the innate immune systems. Another useful feature of the T4 VLP is that the interior of the capsid can also be used for DNA vaccine design. This can be accomplished either by filling an empty capsid *in vitro* using the DNA packaging motor with DNAs that can express vaccine antigens upon delivery into human cells or by inserting DNAs into phage genome by CRISPR engineering. Furthermore, the T4 nanoparticles can be targeted to antigen presenting cells such as the dendritic cells by displaying specific ligands such as anti-DEC205 monoclonal antibodies on the capsid. The NH₂-terminus of the Hoc fibers projected 170 Å away from the capsid wall provide a unique opportunity to display molecules with considerable reach to capture the cell surface receptors present on the targeted cells.

Phage T4 VLP Vaccines

T4 nanoparticle platform has been used to develop VLP vaccines against a number of pathogenic bacteria and viruses, such as *Bacillus anthracis*, *Yersinia pestis*, HIV-1, foot-and-mouth disease virus (FMDV), classical swine fever virus (CSFV), and bursal disease virus. Here, we focus on our recent work on the design of biodefense vaccines against *B. anthracis* and *Y. pestis* using the T4 *in vitro*

VLP display system. *B. anthracis* is the causative agent of anthrax, a deadly disease that leads to rapid death upon exposure. It is listed as a tier 1 biothreat agent by the United States Center for Disease Control (CDC). *Y. pestis* is the causative agent of plague also known as Black Death, which is one of the deadliest infectious diseases known to mankind. Currently, there are no Food and Drug Administration (FDA)-approved plague vaccines for mass vaccination in humans. Although there is a licensed anthrax vaccine consisting of a filtered crude culture supernatant of *B. anthracis* strain V770-NP1-R adsorbed to Alum adjuvant, it is limited to military and high-risk health care workers due to its significant reactogenicity in vaccinated individuals. Therefore, a safe and effective subunit vaccine has been a high priority to protect the public against these biothreats.

The anthrax protective antigen (PA) is a primary target for the development of subunit vaccines. To develop an anthrax VLP vaccine, PA fused to the NH₂-terminus of Hoc was displayed on T4 capsid by *in vitro* assembly to the maximum copy number of ~155 molecules per capsid. When intramuscularly injected into mice with no adjuvant, the anthrax VLP vaccine elicited 6.5-fold higher PA-specific IgG when compared to those immunized with soluble PA adjuvanted with Alhydrogel. Similarly, PA could be fused to Soc and assembled on T4 capsid. In a rhesus macaque model, T4 phage displayed with both PA-Hoc and PA-Soc, after three intramuscular injections, again with no adjuvant, induced up to ~500 µg/ml of PA-specific IgG and provided 100% protection against aerosol challenge with lethal *B. anthracis* Ames strain spores (85 LD₅₀). No side effects due to T4 VLP vaccination were observed in either animal model.

Two *Y. pestis* virulence factors, the capsular protein (Caf1 or F1) and the low calcium response V antigen (LcrV or V) are the main targets for plague vaccine design. We engineered a F1 mutant that produces a monomer as opposed to heterogeneous aggregate in the case of the native F1 and fused it to V to generate a bivalent F1mutV antigen. To develop a plague VLP vaccine using T4 platform, the F1mutV was fused to the NH₂-terminus of Soc and assembled on T4 capsid by *in vitro* display up to about 660 copies per capsid. Without any adjuvant, the plague T4 VLP vaccine when immunized in mice or brown Norway rats (natural host for *Y. pestis*) elicited very high titers of F1V-specific antibodies, higher than the soluble antigen adjuvanted with Alhydrogel. Of particular note is that the T4 VLP vaccine also induced IgG2a and IgG1 antibodies whereas the soluble F1mutV vaccine mainly elicited IgG1 antibodies. IgG2a represent T_H1 immune responses whereas IgG1 reflects the T_H2 responses. This and additional data indicated that the T4 VLP vaccines induced more balanced T_H1 and T_H2 immune responses unlike the traditional subunit vaccines. Stimulation of both arms of the immune system, humoral (T_H2) and cellular (T_H1), is considered to be beneficial for protection, in particular for clearance of intracellular pathogens. Indeed, the T4 VLP vaccine provided 100% protection in both the mouse and brown Norway rat models against intranasal challenge with the most lethal *Y. pestis* CO92 strain up to as high ~5000 LD₅₀ dose.

Taking one step further, we combined both the anthrax and plague T4 VLP vaccines into one and tested its efficacy in mouse, brown Norway rat, and New Zealand white rabbit models. With two intramuscular injections at three week intervals, the robust immunogenicity observed with single vaccine was recapitulated in this bivalent anthrax-plague vaccine and no antigen competition was observed. Most importantly, the vaccine conferred 100% protection against dual challenges with 1 LD₁₀₀ of anthrax lethal toxin and 200–400 LD₅₀ of *Y. pestis* CO92. Protection was observed whether the challenges were administered in sequence or simultaneously, in both mouse and brown Norway rat models. The vaccine also provided 100% protection in rabbits against 200 LD₅₀ challenge with lethal *B. anthracis* Ames strain spores. Thus, the phage bivalent vaccine could protect against two deadly biothreat agents, inhalational anthrax and pneumonic plague.

Other Phage Vaccine Platforms

Other phages have also been used as scaffolds for development of VLP vaccines (Fig. 4). In fact, the abundance and diversity of phages provides many choices for designing VLP vaccine according to specific purposes. For example, the size of VLP was reported to be an important factor for eliciting immune responses. The size of T4-VLP described above is about 120 × 86 nm. Most other phage capsids are smaller. The commonly used phages include λ, T7, MS2, and Qβ, which have sizes of 60 nm, 56 nm, 26 nm, and 28 nm respectively. The capsid of λ phage is composed of the major capsid protein gpE and the outer capsid protein gpD. The gpD, which has a copy number of 405–420 molecules per capsid stabilizes the capsid against the internal pressure of the packaged phage genome (48.5 kb). Therefore, it is an essential protein for the WT phage but dispensable for phages with shorter genomes. The COOH-terminus of gpD has been extensively used for display of epitope peptides. However, it is possible to display larger antigens but at reduced copy number. Although both *in vitro* and *in vivo* display have been used to assemble the foreign peptides fused to gpD, *in vivo* display is the preferred method because phages or capsids produced in the absence of gpD are less stable for *in vitro* display.

The capsid of T7 phage is assembled with the major capsid protein gp10. However, a small portion of gp10 undergoes a –1 frame-shift at the COOH-terminus of gp10 (345 amino acids) to produce a slightly longer gp10 (398 amino acids). The former is named gp10A and the latter, gp10B. The gp10B is not essential for capsid assembly and, therefore, has been used for display of peptide epitopes and antigens as fusions to the COOH-terminus of gp10. Although the copy number is generally limited, as high as 415 copies of about 50 amino acid antigen peptides have been displayed on the T7 phage by *in vivo* display.

Two small single stranded RNA phages MS2 and Qβ have also been used for VLP vaccine design. The capsid of these phages is composed of 180 copies of major coat protein (CP) to which a peptide epitope can be fused and self-assembled both *in vivo* and *in vitro*. However, successful capsid assembly requires WT CP, hence the copy number of the epitope-fused CP is limited. Phage Qβ also contains three to five copies of A1 protein, which is a large 196-amino acid extension at the COOH-terminus of CP as a result

Phage	Capsid size (nanometer)	Phage protein(s) used for display (copies/capsid)	Preferred molecule for <i>in vivo</i> display	Main display scheme	Maximum copy number	Display of mammalian expressed antigen
T4	120 X 86	Hec (155) Sec (870)	Full length protein, Peptide	<i>in vivo</i> <i>in vitro</i>	1,025	Yes (<i>in vitro</i>)
A	60	gpD (405-420)	Peptide	<i>in vivo</i> <i>in vitro</i>	420	Possible (<i>in vitro</i>)
T7	56	gp108 (415)	Peptide	<i>in vivo</i>	415	No
MS2	25	CP (180)	Peptide	<i>in vivo</i>	180	No
Q β	28	A1 (3-5) or its mutant (96)	Peptide	<i>in vivo</i>	86	No
M13/fd	900 X 6	pVIII (2,700) pIII (5)	Peptide	<i>in vivo</i>	2,700	No

Fig. 4 Properties of various phage display systems.

of infrequent read-through at the termination codon. This feature has been used as a way to display peptides through fusion to the COOH-terminus of A1.

Filamentous phages such as M13 and fd have also been used for vaccine development, although these phages have been most powerful for display of random peptide libraries. The 900 nm long phage fd filament is composed of about 2700 copies of major coat protein pVIII. At one end of the filament are 5 copies each of minor capsid proteins pIII and pVI, while the other end contains 5 copies each of minor capsid protein pVII and pIX. The pVIII can only display short 6–8 amino acid peptides, while the minor capsid proteins can display proteins although at low copy number, 1–5 molecules per filament.

Many vaccine candidates have been developed using different phage VLP platforms described above. Readers are encouraged to read the review articles list below for details of these vaccine candidates.

Conclusion Statement

Vaccine development thus far has been individualized for each disease. Consequently, it takes enormous time, resources, testing, and manufacturing hurdles, in order to bring a vaccine into human use. The question remains if a “universal” VLP platform can be developed to design vaccines against many infectious diseases. Phages in general, and T4 phage in particular, provide the necessary architectures for such a platform. Phages are highly stable, can be manufactured relatively easily and cost-effectively, is considered to be safe, and has no significant pre-existing immunity in humans. The features of phage T4; large size, high antigen capacity, robust immunogenicity without adjuvant, ability to co-deliver DNAs and targeting molecules, and ease of engineering using CRISPR provide distinct advantages to manipulate this VLP platform and design efficacious vaccines against pathogens, including complex and emerging pathogens such as HIV and Flu. The phage platforms could therefore potentially streamline and accelerate the whole vaccine development process in the future.

Acknowledgments

This research has been supported by grants from the National Institute of Allergy and Infectious Diseases (current: AI111538 and AI081726) and in part from the National Science Foundation, to VBR. We thank many current and past post-doctoral fellows and graduate students who contributed to bacteriophage T4 vaccine research, in particular Jennifer Jiang, Laura Abu-Shilbayeh, Zhihong Zhang, Qin Li, Sathish Shivachandra, Taheri Sathaliyawala, and Guofen Guo, and Wadad Alsalmi; collaborators Drs. Michael G. Rossmann and Andrei Fokine (Purdue University), Ashok Chopra (University of Texas Medical Branch), Stephen H. Leppla (National Institutes of Health), and Carl Alving, Gary Matyas, and Mangala Rao (Walter Reed Army Institute of Research).

Further Reading

- Bao, Q., Li, X., Han, G., *et al.*, 2018. Phage-based vaccines. *Advanced Drug Delivery Reviews*. doi:10.1016/j.addr.2018.12.013, (pii: S0169-409X(18)30317-X).
- Black, L.W., Rao, V.B., 2012. Structure, assembly, and DNA packaging of the bacteriophage T4 head. *Advances in Virus Research* 82, 119–153.
- Chen, Z., Sun, L., Zhang, Z., *et al.*, 2017. Cryo-EM structure of the bacteriophage T4 isometric head at 3.3-Å resolution and its relevance to the assembly of icosahedral viruses. *Proceedings of the National Academy of Sciences of the United States of America* 114 (39), E8184–E8193.

- Mohsen, M.O., Zha, L., Cabral-Miranda, G., Bachmann, M.F., 2017. Major findings and recent advances in virus-like particle (VLP)-based vaccines. *Seminars in Immunology* 34, 123–132.
- Rao, V.B., Feiss, M., 2015. Mechanisms of DNA packaging in large double stranded DNA viruses. *Annual Review of Virology* 2 (1), 351–378.
- Rappuoli, R., Pizza, M., Del Giudice, G., De Gregorio, E., 2014. Vaccines, new opportunities for a new society. *Proceedings of the National Academy of Sciences of the United States of America* 111 (34), 12288–12293.
- Ren, Z., Black, L.W., 1998. Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid. *Gene* 215 (2), 439–444.
- Tao, P., Mahalingam, M., Kirtley, M.L., *et al.*, 2013. Mutated and bacteriophage T4 nanoparticle arrayed F1-V immunogens from *Yersinia pestis* as next generation plague vaccines. *PLOS Pathogens* 9 (7), e1003495.
- Tao, P., Mahalingam, M., Zhu, J., *et al.*, 2018. A bacteriophage T4 nanoparticle-based dual vaccine against anthrax and plague. *mBio* 9 (5). doi:10.1128/mBio.01926-18.
- Tao, P., Wu, X., Tang, W.C., Zhu, J., Rao, V.B., 2017. Engineering of bacteriophage T4 genome using CRISPR-Cas9. *ACS Synthetic Biology* 6 (10), 1952–1961.
- Tao, P., Zhu, J., Mahalingam, M., Batra, H., Rao, V.B., 2018. Bacteriophage T4 nanoparticles for vaccine delivery against infectious diseases. *Advanced Drug Delivery Reviews*. doi:10.1016/j.addr.2018.06.025, (pii: S0169-409X(18)30164-9).

Relevant Website

<https://www.cdc.gov/vaccines/vpd/vaccines-list.html>
List of Vaccines – CDC.

Bacteriophage Diversity

Julianne H Grose, Brigham Young University, Provo, UT, United States

Sherwood R Casjens, University of Utah, Salt Lake City, UT, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Genome mosaicism Phage genomes that have patches of similar sequence and patches of quite different sequence are considered to “mosaically related”. Such patches are said to be “mosaic sections” and are thought to have arisen by horizontal exchange among divergent phages rather than by differential divergence.

Host range The hosts that can be successfully infected by a given phage.

Metagemone A collection of all the nucleic acid sequences present in an environmental sample.

Phage cluster A set of related phages with nucleotide sequence similarity that spans > 50% of their genomes as determined by nucleotide dot plot.

Phageome Collection of sequences identified as bacteriophage-derived in an environmental or metagenomic source.

Plaque assay An assay for phage infection that relies on mixing the bacteria and phage prior to plating, and observing the plaques that arise on plate.

Prophage A phage genome whose lytic genes are not expressed and which replicates in concert with the host genome, either as a plasmid or integrated into the host chromosome.

Protein cluster A cluster of protein sequences derived from adjacent genes in a sequencing contig in metagenomic studies.

Virion The stable virus particle that is released from infected cells and is capable of binding to and infecting other sensitive cells.

Virome Collection of sequences identified as virial due to the extraction method and derived from an environmental or metagemone source.

Introduction

The diversity of the functional, physical and genetic aspects of bacteriophages (or phages), viruses that infect bacteria, seemingly knows no bounds. This is an exciting time of new discovery in this field, and with each new study known phage diversity continues to expand. Phages have been found that infect nearly every bacterial species that has been examined, and since they replicate rapidly and kill host bacteria, they are important players in the ecology of bacteria and thus the overall ecology of Earth. Bacteria, and thus their phage predators, are critically important in such varied processes as global nitrogen and carbon cycling, animal and plant diseases of many types and the human immune response. Although a few phages have been very successfully studied since the 1940s as model systems for the understanding of basic molecular biology of life, their amazingly high abundance was only fully recognized when particles with their unique tailed phage-like morphology were quantified in environmental samples in the 1980s and 90s. These studies showed that there are 10^{31} – 10^{32} tailed phage virions on Earth, which probably corresponds to a global ratio of about ten for each bacterial cell. In addition, the advent of facile nucleotide sequencing of bacterial genomes showed that most harbor one or more quiescent phage genomes called prophages. Other phage types beyond the tailed phages may be less common, but their virions are more difficult to recognize in environmental samples and so their abundance is less well understood.

Phages exist that have several different general behavioral types. (1) Lytic phages always replicate to make progeny virions and kill their host upon infection. (2) Temperate phages can, depending on the conditions at the time of infection, either replicate lytically to make progeny or turn off their lytic, cell killing genes and replicate as a prophage in synchrony with the host chromosome, either integrated into the host chromosome or as a plasmid. (3) Filamentous phages continuously replicate progeny and they escape the cell without overtly killing it. Temperate prophages often express “lysogenic conversion” genes that give the host cell some advantage, and in response to environmental signals many prophages can undergo “induction” to enter the lytic cycle to produce progeny and kill the host cell.

Like all viruses, phages are extremely diverse, and they exist as eight very different general types that include virions that carry ssDNA, dsDNA, ssRNA or dsRNA chromosomes and that have several different morphologies (Table 1). Virion chromosomes can be linear or circular in different phage types. The genome sizes of known phages range from less than 4 to nearly 500 kbp, and the number of encoded proteins varies commensurately. Most phage virions contain one nucleic acid molecule, but the dsRNA phages contain several separate chromosomes. (See the International Committee on Taxonomy of Viruses [ICTV] world-wide-web page: website link provided in “Relevant Websites section” for detailed descriptions of the various types.) In addition, the five different protein folds of the protein that makes up the virion’s protective coat suggests that this critical aspect of phages has been “invented” at least five times during the course of life on Earth, and divergence within these groups indicates that they are all very ancient. The chromosomes of different nucleic acids and the > 100-fold variation in genome size means that there are a multitude of varied molecular strategies for phage replication and survival in nature that cannot all be described in this venue.

Table 1 Eight types of bacteriophages

Group	Well studied members	Virion DNA	Member lifestyles	Virion type
dsDNA				
<i>Caudovirales</i> (tailed phages)				
<i>Myoviridae</i>	T4, P2	Linear, 28–500 kbp	Lytic or temperate	Icosahedral ^a head, long contractile tail
<i>Siphoviridae</i>	λ , SPP1	Linear, 36–125 kbp	Lytic or temperate	Icosahedral ^a head, long non-contractile tail
<i>Podoviridae</i>	T7, P22	Linear ^b , 16–90 kbp	Lytic or temperate	Icosahedral ^a head, short tail
<i>Corticoviridae</i>	PM2 ^c	Circular, 10–11 kbp	lytic (or temperate?) ^d	Icosahedral, internal lipid membrane
<i>Plasmaviridae</i>	AVL2 ^c	Circular, 12 kbp	lytic	Pleomorphic. external membrane
<i>Tectiviridae</i>	PRD1	Linear ^b , 15–18 kbp	Lytic or temperate	Icosahedral, internal lipid membrane
ssDNA				
<i>Inoviridae</i>	M13, f1	Circular, plus strand, 4–9 kb	Lytic or temperate	Filamentous/helical
<i>Microviridae</i>	ϕ X-174	Circular, plus strand, 4–6 kb	Lytic (or temperate?) ^e	Icosahedral
dsRNA				
<i>Cystoviridae</i>	ϕ 6 ^c	Linear, 3 molecules, 13 kbp	Lytic	Icosahedral, external membrane envelope
ssRNA				
<i>Leviviridae</i>	MS2, Q β	Linear, plus strand, 3–5 kb	Lytic ^f	Icosahedral

^aSome species have elongated pseudo-icosahedral heads.

^bA protein is covalently attached to 5'-ends in all *Tectiviridae* and some for *Podoviridae*.

^cOnly a small number of isolates have been studied.

^dThe studied members of this group appear to be lytic, but PM2-like sequences (possible prophages?) are present in the genomes of some bacteria.

^eRecent reports of *Microviridae* genomes embedded in bacterial genomes have led to the suggestion that some may be temperate.

^fA poorly understood carrier state has been reported for phage LeviOr01.

In spite of their staggeringly high abundance and the different kinds of virion structures, the true diversity of phages was not appreciated until genome sequencing became easy to perform. Phages MS2 (ssRNA), ϕ X-174 (ssDNA) and λ (dsDNA) were the first virus genomes to be sequenced with high cost and great effort in the late 1970s and early 1980s, but as sequencing costs and required effort declined over the past several decades the number of phage genome sequences in the public databases has increased dramatically. Indeed, up to 100 phage genomes can now be sequenced in a single Illumina type run for only about ten US dollars per genome, and there are currently (in 2019) over ten thousand available genome sequences for “authentic” phages (defined as phages that have been propagated in the laboratory and not including prophage sequences in bacterial genomes or phage sequences assembled from metagenomic information). As phage genome sequences were compared to one another it was quickly realized that genetic differences are very much greater than easily observable virion morphology differences, and the latter are insufficient to distinguish phages into informative groups of similar individuals. For example, *Siphoviridae* tailed *Escherichia coli* phages λ (Accession No. JQ086376) and *Bacillus subtilis* phage SPP1 (Accession No. NC_004166) virions appear very similar by negative stain electron microscopy, but they have essentially no recognizably similar nucleotide sequence. Thus, the study of the diversity of phages is necessarily the study of the diversity of their genome sequences.

There are eight general phage types (Table 1), and where more than a few individual phages are known, there is substantial sequence diversity among individuals, but only in the *Caudovirales* or tailed phages is this diversity known to extend to very major differences in gene content. The other seven types appear at present to have limited ranges of gene content, for example in the two other best studied types the *Microviridae* have about ten genes that include largely homologous but not identical virion assembly and replication gene complements and variable lysis and other accessory genes, and the *Leviviridae* have four genes that include conserved virion assembly and replication genes and variable lysis genes. Large numbers of authentic phage genome sequences are known for only the tailed phages, so we will focus on the tailed phages in the remainder of this review.

Understanding the Nature of Tailed Phage Diversity and Phage Classification

The tailed phages have been historically separated into three types, the *Podoviridae* that have short tails, *Siphoviridae* that have long non-contractile tails and *Myoviridae* that have long contractile tails. These tail differences, which easily seen by electron microscopy of virions, reflect different strategies for DNA delivery into cells but do not correlate with other features of phage lifestyle details. The tailed phage classification method that is most similar to cellular ribosomal RNA based classification relies on the fact that there are (only) three proteins for which homologs are encoded by all tailed phages, the major capsid protein (MCP, which forms the head shell), portal protein (channel through which DNA is packaged and ejected) and the large terminase subunit (DNA packaging motor protein), but such an approach that uses one or a few proteins can be confused by the extensive horizontal exchange of genetic material that has characterized phage history. In addition, phages and thus these proteins, especially MCP and large terminase, are so diverse that they can have little to no easily recognizable sequence similarity, making comparisons of distantly related phages difficult. Thus, comparison of such a small set of genes, although usually informative, can incorrectly identify overall group membership, but more importantly it does not show important differences within groups of related phages

(discussed in more detail below). More informative methods for overall phage classification utilize whole genome or proteome approaches. Among these are nucleotide dot plot analysis, Average Nucleotide Identity (ANI; nucleotide sequence identity between two genomes), and whole proteome similarity determinations (such a proteomic tree that compares all the encoded proteins or proteome conservation that measured the fraction of genes that are homologs).

Each of these strategies has weaknesses and strengths. A strength of whole genome nucleotide analysis is that it doesn't depend on correct annotation of genes or encoded protein function. An advantage of the proteome-based approaches is that proteins often retain recognizable homology after the encoding nucleotide sequences have diverged beyond the point of having recognizable similarity, so more distant protein relationships among phages can be detected and analyzed. Protein comparison methods are also particularly useful in the analysis of metagenomic data where proteins encoded by genes or gene clusters, rather than whole genomes, are usually compared. A significant disadvantage of whole genome nucleotide or proteome sequence clustering or tree construction, is that such approaches average differences and similarities across the genomes, losing important information on the variable relationships within genomes.

Whole genome nucleotide dot plots on the other hand are capable of showing varied relationships across highly mosaic and rearranged genomes. This operational approach was pioneered by Graham Hatfull and colleagues for easy visualization of relationships within large sets of phages by overall genome similarity in a way that does not disguise differential relationships across the genomes (genome mosaicism, see below). Groups of highly related phages recognized by the method are defined as "clusters" of phages that have dot plot diagonal line similarity that covers $\geq 50\%$ of their genomes, which generally correlates with $> 50\%$ ANI and $> 40\%$ proteome conservation. Such lines indicate regions of genetic similarity and synteny. Thus, phage genomes within a cluster carry homologous genes in the same functional order and so have similar virion structures, transcription pattern and lifestyles. For example, no *Enterobacteriales* cluster contains convincingly temperate and lytic phages; however, some *Actinobacteria* clusters contain both and their transcription patterns are very similar. This suggests that at least some types of phages do not easily move between these two major types; however, a small number of *Enterobacteriales* lytic and temperate clusters have similar virion structure and assembly genes. The ES18-like and P22-like temperate phage clusters have rather divergent virion structural genes that are nonetheless syntenic with the E1-like and IME-EC2-like lytic phage clusters, respectively, indicating that a small number of ancient exchanges of large genome segments has occurred between lytic and temperate phages. We note that analysis of just the MCP has been shown to usually ($\sim 99\%$ of the time) reflect the cluster to which a phage belongs, indicating that horizontal MCP exchange between clusters is in fact quite rare. The exceptions to this rule are phages that have acquired an MCP gene from a more distant relative outside of its cluster. The dot plot approach also reveals hierarchical relationships, and we use the term "subclusters" for more highly related groups (long, strong diagonal similarity lines, generally with $> 80\%$ ANI) within clusters. The higher level "superclusters" include groups of phage clusters that have syntenic genomes that encode homologous but more distantly related proteins without extensive nucleotide sequence similarity (less than 50% of the genomes are similar by dot plot, generally with $< 50\%$ ANI) (Fig. 1). We use this cluster terminology in this review.

A weakness of the dot plot cluster method is that it is difficult to quantitate relationships. Nonetheless, it succeeds in placing nearly all phages into useful and informative groups. However, there are a few known cases where a given phage is about half related to one cluster and half to another within a supercluster. These are apparently rather recently formed hybrid phages between members of two closely related clusters, and the dot plot method shows this clearly. For example, among the *Enterobacteriales* phages (see below) about half of the phage FSL_SP-016 genome (Accession No. KC139516) is similar to the BP-4795-like cluster and about half is similar to the ES18-like cluster. Thus, classification by the dot plot method can in rare instances be somewhat ambiguous, but it provides a platform to easily recognize and understand such "hybrid" phages.

We note that the two most common terminologies for discussing phage relationships, International Committee on Taxonomy of Viruses (ICTV) defined species and genera and dot plot defined clusters, are not unrelated or mutually exclusive in that they both rely on these whole genome nucleotide/proteome methods. The taxa put forward by the ICTV for phage classification (phylum, class, order, family, genus and species) are useful, but since they typically utilize whole genome averaging methods they do not emphasize the mosaic nature of phage genomes in the way dot plots do (see below). Discussion regarding what comprises a phage 'species' and what classification scheme is most useful remains lively.

Strategies for Studying Phage Diversity

Phage diversity can be studied by two opposing tactics, a "top down" strategy where total extant phage sequence diversity is analyzed in environmental samples and a "bottom up" strategy where individual phage genomes are compared. The enormous diversity of phages means that to be meaningful the latter must examine a very large number of individual "authentic" phages or be limited to phages that can infect a given bacterial taxon. Again, each approach has strengths and weaknesses. The top down methods depend on metagenomic analysis of environmental samples which can include mining total metagenomic sequence libraries for sequences that are similar to known phages or isolation and sequencing (not propagation) of purified virion fractions of such samples. The major advantage of this type of approach is that in theory it can see the entire expanse of phage diversity. However, it also has weaknesses; for example it is difficult or impossible to recognize completely novel phages, RNA phages are overlooked when only DNA is sequenced, not all virion types purify in the same fractions, some virions may be unstable and hard to purify, whole genome sequences have been difficult to obtain in high numbers, any complete genomes assembled from such

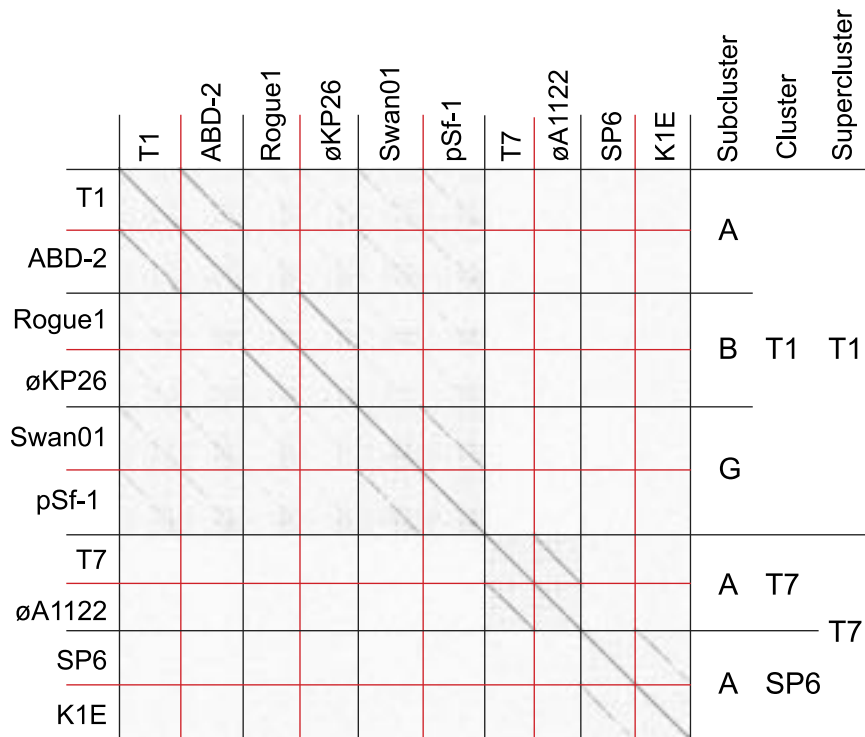


Fig. 1 Hierarchical phage classification by dot plot. Six representative phage genomes in the *Enterobacteriales* T1 supercluster phage genomes and four T7 supercluster genomes are compared with Gephard and a window size of 10. Phage names are shown at the left and top of the figure. The T1 supercluster genomes fall into a single cluster since they all have strong or weak, long diagonal similarity lines. They comprise three (previously named) subclusters whose members form strong lines within the subcluster and weak lines between subclusters. The four members of the T7 supercluster form the T7-like and SP6-like clusters with no diagonal lines between them. Each of these clusters contains two members of a single subcluster. The T7- and SP6-like clusters are members of the same (T7) supercluster because their genomes are highly syntenic when the proteins encoded by their genes are analyzed (*i.e.*, they have very similar gene orders, transcription patterns and molecular lifestyles).

samples are likely composites of similar individual phages, it can be difficult to know the bacterial host species of the phage genomes discovered, phage virions can contain DNA from their bacterial hosts, and, finally, interesting findings cannot be easily be studied in more depth because of the inability to find and propagate the actual phages that correspond to the sequences of interest.

On the other hand, the bottom-up strategy of isolating phages on particular hosts and studying them has the major advantage that the hosts are known and phage features can be studied in as much detail as desired. Its disadvantages are as follows: phage types may vary across the bacterial spectrum, so many phages that infect many different hosts must be studied to determine the complete extent of phage diversity, some phages may be difficult to propagate in the laboratory and so may be missed, and most importantly one is entirely dependent upon random chance for the discovery of new phage types so it is very hard to know rigorously when all extant varieties have been found; even for *E. coli* and *Salmonella enterica*, two of the bacterial species for which phages are best understood, new phage clusters or types are still being discovered.

Examples of Phage Diversity

Top Down Studies

Human phageome

A myriad of top down metagenomic virome and phageome analyses from such diverse sources as human, termite and sea squirt guts, ocean water and lake bottom sediment have shown that phages are spectacularly diverse and that this diversity has not yet been fully described. The majority of top down phage diversity studies have been human-centric or focused on the marine environment. These two kinds of samples provide evidence that phage diversity varies dramatically in different ecological niches. For example, temperate phages appear to be more prominent in the human gut, while lytic phages appear to be more prominent in the marine environment. The temperate phages of the gut, especially in their prophage state, influence human health through the expression of their lysogenic conversion genes within their bacterial hosts, providing novel functions and capabilities, including defense against the human immune system. Initial human phageome studies suggested a great diversity among individuals, but more recent deeper sequencing and better analysis methods have allowed the identification of bacteriophages that are common among many human individuals, including the crAssphage family. The latter is a novel group of viruses that infect at

least in some cases *Bacteroides* hosts and can comprise up to 90% of the human gut virome, making them the most ubiquitous and abundant phages found in human samples to date. These and additional studies support the presence of a stable, diverse and unique personal phageome as well as a population of phages that are highly conserved across most healthy individuals.

Marine phageome

In contrast to the primarily temperate human phageome, the marine phageome is reported to be primarily lytic. With an estimated concentration of 10^8 phage per mL of ocean water, many marine virome metagenomic studies have been performed from geographically diverse sources. These studies report a staggering diversity in that 87%–93% of putative sequences in oceanic “virus preparations” lack similarity to known reference sequences. This lack of similarity also hampers assembly of whole genomes, and therefore, estimates of diversity are primarily derived from analysis of viral protein clusters. In 2015, the analysis of metagenomic marine data suggested a global virome consisting of less than 3.9 million “protein clusters”, and revealed variation according to depth, distance from shore and season of the year. Genetic flow appears to travel from the ocean surface to the deep in that core (ubiquitous) protein clusters are enriched on the ocean surface, while greater diversity is present in areas where the ocean is not illuminated by the sunlight (the aphotic zone). Relatively few whole genome sequences of authentic marine phages have been reported (482 complete sequences as of 2015, of which only 274 have known hosts), and these have been isolated on only a few representative bacterial hosts (22 genera of the 10^6 estimated in the ocean), making phage-host relationships relatively unstudied. This is rapidly changing, however, with advanced methods for bacterial culture, single-cell sequencing of bacteria and the phages that have infected them, and phage-tagging strategies. Although many of the fully sequenced marine phages represent novel types, some have similarities to the *Enterobacteriales* T4-like, T7-like and N4-like non-marine phage clusters.

Bottom Up Studies

Two large diversity studies

The bottom up method has been utilized on a large scale in only two out of the nearly 200 currently defined bacterial orders, the Gram negative order *Enterobacteriales* and the Gram positive order *Actinomycetales* (currently being expanded to include other orders in the *Actinobacteria* phylum). Analyses of these two large phage groups have largely relied on the dot plot and shared gene approaches to define phage clusters. Over 1150 genomes of authentic phages that infect the *Enterobacteriales* and about 2700 that infect the *Actinobacteria* have been sequenced. These *Enterobacteriales* phages were independently isolated and sequenced by a large number of diverse researchers around the world and deposited in the GenBank database. They currently form 82 clusters of which 23 are singleton clusters (clusters with only one phage). The *Actinobacteria* phages were isolated through the Howard Hughes Medical Institute (HHMI) Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA PHAGES) program and are available at NCBI GenBank and website available at: “Relevant Websites section”. They currently form 182 clusters of which 62 are singleton clusters. The significant number of singleton clusters strongly suggests that the diversity of both these large sets of phages remains quite incompletely described. Although the order *Enterobacteriales* is a narrower taxonomic unit than the phylum *Actinobacteria*, the current average number of phages examined for each cluster defined is quite similar, about 14 and 15, respectively. Thus, at this still early stage of the analysis of tailed phages, this aspect of diversity appears to be rather similar in these two very disparate bacterial taxa, and we project that it may well be similar across all the *Bacteria* domain. Nonetheless, currently these bottom up studies still suffer from the fact that a very small fraction of extant phages have been analyzed, so future studies could change current views.

The *Enterobacteriales* tailed phage example

As an example of the rapidly expanding nature of the study of phage diversity, the initial 2014 analysis of the *Enterobacteriales* phages in GenBank included 337 complete bacteriophage genomes sequences. These phages were isolated on 18 bacterial host genera and 31 species and comprised 56 separate clusters (32 lytic and 24 temperate). There are currently about 1150 *Enterobacteriales* phage genomes in GenBank (a 3.4-fold increase in 4.5 years) isolated on 25 host genera and 50 species, and these fall into 82 clusters (50 lytic and 32 temperate). As further indication of the incomplete nature of our current understanding, we note that the bacterial order *Enterobacteriales* currently contains 55 genera (and at least 30 more have been proposed for membership), so phages that infect less than half of this taxon’s genera have been examined. The number of related phage clusters will certainly grow until phages are included that infect all its genera.

As described above, members of a given phage cluster have very similar molecular lifestyles, but the various *Enterobacteriales* tailed phage clusters, nonetheless, have internal diversity, the extent of which can vary dramatically. For example, one of the least diverse clusters is at present the jumbo phage Ea35-70 *Myoviridae* cluster comprised of nine *Erwinia* phages that form a single subcluster level group of highly related phages (>94% ANI with genome sizes that range only from 271 to 275 kbp). These correspond to the ICTV designated Agrican357 virus genus. Other clusters are comprised of more diverse phages, such as the SPN3US-like *Myoviridae* cluster whose 18 known members infect four different host genera including *Escherichia*, *Salmonella*, *Cronobacter* and *Erwinia* and show striking diversity with some members sharing less than 37% ANI. These SPN3US-like phages are diverse even when isolated from a single host in that the 12 *Erwinia* phages form seven different subclusters.

Superclusters designate distant phage relatives with similar protein and content order but little nucleotide similarity (above). Analysis of the *Enterobacteriales* phage clusters reveals at least seven supercluster level groups that contain multiple clusters. The

most commonly sequenced *Enterobacteriales* bacteriophages are members of the T7 supercluster, which is comprised of nine clusters that contain 191 phages that infect 18 host genera. No doubt these phages are very successful in the environment, however, this may reflect ease of propagation in the laboratory rather than ecological domination. The remaining six groups that contain multiple clusters are the lytic superclusters typified by phage N4 (2 clusters, 28 phages, 7 host genera), SETP3 (3 clusters, 88 phages, 8 host genera) and rV5 (2 clusters, 46 phages, 6 host genera) and the temperate superclusters typified by lambda (25 clusters, 169 phages, 15 host genera), P1 (2 clusters, 7 phages, 2 host genera), and P2 (2 clusters, 25 phages, 6 host genera). Thus, superclusters of both lytic and temperate types are internally very diverse, infecting multiple hosts with a wide diversity of phage clusters.

A closer look at diversity within the *Enterobacteriales* T7 supercluster reveals gene synteny with minor variations and conservation across all eight of its clusters for key early genes (host restriction proteins, DNA polymerase and RNA polymerase), middle genes (DNA metabolism functions and lysozyme) and late genes (virion structure and assembly). In general, within each of these clusters member phages share >50% of their proteome, whereas the proteome conservation is often ~30% between these “closely-related” clusters. Phages in the even more diverse lambda supercluster also contain conserved modules of function including the following genes in the same order (with only a few minor exceptions): terminase and head assembly genes, followed by tail assembly genes, lysogenic conversion genes, integrase and homologous recombination genes, the early gene transcriptional control genes (anti-terminator [N], prophage repressor [CI], cro repressor and CII transcriptional activator), DNA replication, “nin region” genes, late gene transcriptional antiterminator and lysis genes. Nearly every lambdoid cluster contains a unique, very different set of head assembly genes; the only exceptions are the P22-/APSE-like and lambda-ø80-/N15-like groups of clusters that have similar head assembly genes but have major differences elsewhere in the genome.

Contribution of Prophages to Phage Diversity

Although it has not been studied quantitatively, isolation of phages from the environment may sometimes underestimate temperate phages. For example, nearly all of the *Enterobacteriales* “model system” temperate phages that have been studied in great depth in the laboratory (e.g., phages P1, P2, P22 and lambda) were isolated after they were released from prophage-carrying bacteria in the laboratory rather than as infectious virions from environmental samples. Perhaps they spend most of their time as prophages and less as virions? This may be less true for the *Actinobacteria* phages, as many environmental temperate phages have been isolated that infect that host taxon. Bacterial genomes are known that harbor >15 prophages, and in one of the few systematic studies in this area it was found in 2016 that analysis of 3298 *Salmonella* genomes identified 9371 tailed phage prophages that belonged to the 20 temperate *Enterobacteriales* phage clusters that had been defined at the time. Prophages are clearly very abundant and diverse. It is very likely that unrecognized prophages that belong to currently undiscovered clusters are also present in bacterial genome sequences, but because of the great diversity of tailed phages no computer search engine currently exists to rigorously discover and classify all of them. Clearly, since bacterial cells are thought to be about 10% as abundant as phage virions in the environment, the number of tailed phage prophage DNAs on Earth is likely within about an order of magnitude of the number of free virions, and they represent a still rather poorly understood source of phage diversity.

Phage Diversity and Horizontal Exchange of Genetic Information

The Nature of Horizontally Exchangeable Mosaic Section Alleles

Many reports have emphasized credible examples of past horizontal exchange of genetic information amongst the tailed phages, and although such exchange does not increase the number of phage gene types, it increases the diversity of gene combinations, increases the diversity within the participating clusters, provides opportunity for further gene divergence and spreads phage genetic information across the *Bacteria* domain. The frequency of such horizontal exchange events between phage types is not known, but it is not high enough to homogenize all such genomes or even disrupt the clear division of phages into clusters. Comparison of phages within any cluster typically shows patchy similarities with regions of high similarity interspersed with regions that are very different. Such genomes are said to be “mosaically related” and the different patches are thought to have arisen through horizontal exchange mechanisms. Genome mosaicism was recognized early in the study of *E. coli* phage lambda and its relatives and has since been shown to be present to varying degrees in most *Enterobacteriales* phage clusters. In general, temperate phage clusters appear to exhibit more extensive mosaicism than virulent ones, and mosaicism has been studied in most detail in the phage lambda.

All members of each supercluster have syntenic genomes where genes with parallel functions have the same chromosomal order and have similar overall transcriptional gene expression patterns. At present genomes of about 170 authentic phage (ignoring thousands of related prophage sequences) members of the *Enterobacteriales*-infecting lambda supercluster have been completely sequenced, and these are currently divided into 25 clusters. Comparison within each cluster shows much more overall similarity than comparisons between clusters; however, even phages within a cluster often show substantial mosaicism, and Fig. 2 shows, as an example, the relationship between two typical members of the temperate phage P22-like cluster within the lambda supercluster. This cluster has been studied in detail and contains at least 25 exchangeable mosaic segments or modules within the cluster of 15 virion assembly genes that make up about half of their genomes. Such segments can encode parts of genes, whole genes or clusters of genes.

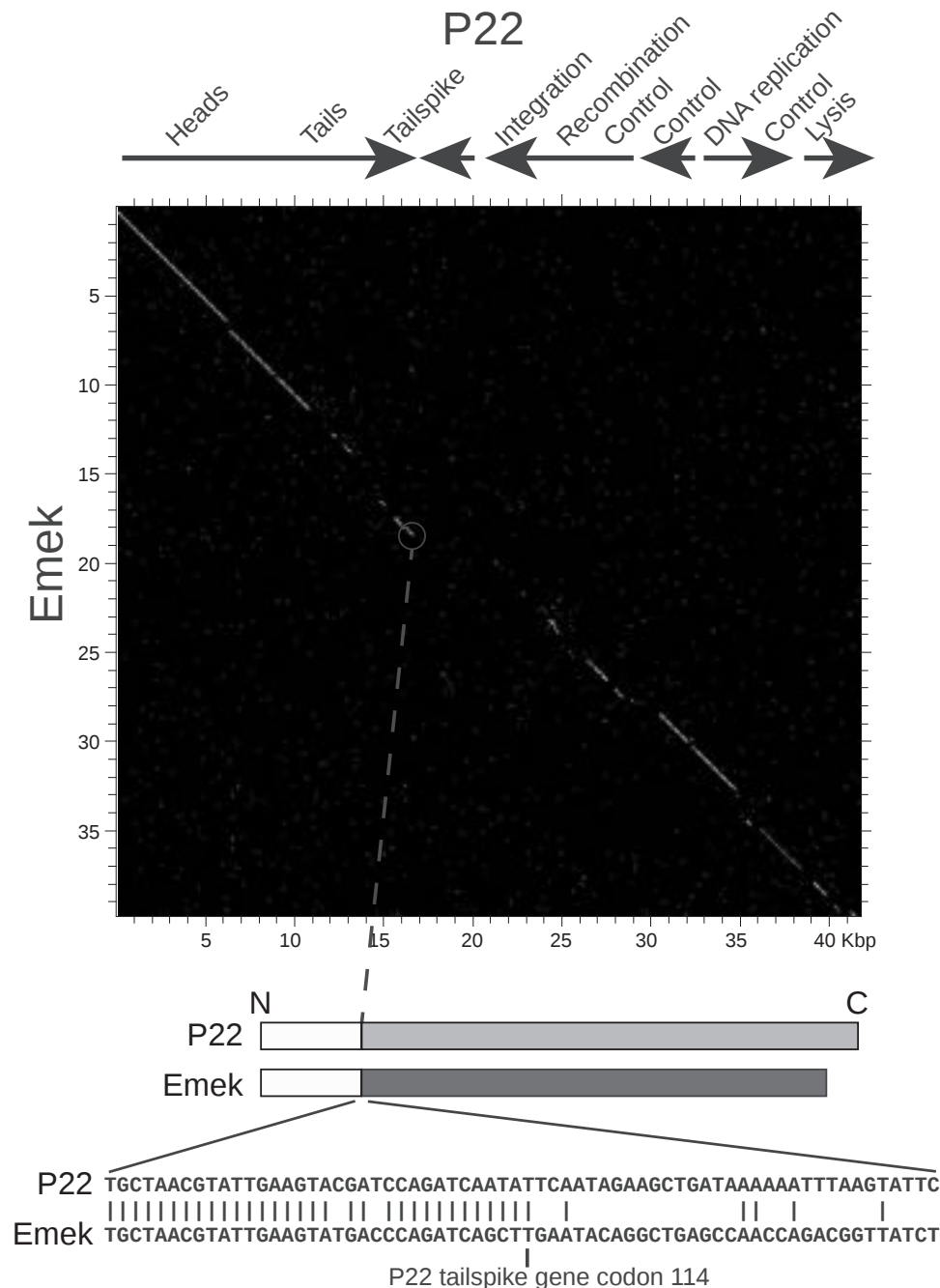


Fig. 2 Phage genome mosaicism. The typical mosaicism present in the genomes of phages P22 and Emek in the P22-like cluster within the phage lambda supercluster is shown in the dot plot (Accession numbers BK000583 and JQ806763, respectively; plot was produced by DNA Strider with a scan window threshold of 17 identities per 23 bp). The transcription pattern and location of functional gene clusters of P22 are indicated above the plot. Below the plot, the two-domain tailspike proteins are diagrammed with white N-terminal domains indicating their strong similarity and two shades of gray indicating that their C-terminal receptor-binding domains have no recognizable homology. At the bottom, the aligned nucleotide sequences at the mosaic boundary between the two phage's tailspike genes' N- and C-terminal domains are shown (translation is from left to right). We note that P22 and Emek infect *Salmonella* serovars Typhimurium and Haardt, respectively. These two *Salmonella* serovars have very different surface polysaccharides, and the C-terminal polysaccharide-binding tailspike domains are essentially unrelated in amino acid sequence.

Three different types of diversity exist within allelic sets of mosaic modules, and the mosaic section that encodes the P22 tailspike C-terminal receptor-binding domain demonstrates this as follows: (1) Some such alleles are divergent but retain recognizable protein sequence similarity (*e.g.*, phage P22 (Accession No. KR296686) and Det7 (Accession No. KP797973) tailspikes); (2) some are ancient homologs that have diverged to the point that recognizable sequence similarity is lost but they retain

similar protein folds (e.g., P22 and Sf6 (Accession No. AF547987) tailspikes); and (3) others are unrelated and have different protein folds that serve the same function (e.g., P22 and CUS-3 (Accession No. CP000711) tailspikes). The number of different alleles that are known for the various mosaic modules along the genome is highly variable. There are, for example, three major sequence types of P22-like major capsid protein, but there are many more types of tailspike C-terminal domain which binds the virion to its polysaccharide receptor and so is the primary determinant of host specificity; each of its >70 known very different sequence types in the P22-like phages is responsible for binding a different polysaccharide. Thus, there can be very extensive diversity even within a phage cluster.

How did Genome Mosaicism Arise?

The patches of similarity and difference (mosaic modules) appear to be the result of recombination after rare horizontal exchange, usually between divergent members of a cluster or supercluster. A few short homologous syntenic sequences have been found between mosaic modules (often transcription terminators) that could allow homologous recombination to create new combinations of adjacent mosaic module alleles; however, no such common sequences have been found between most adjacent modules. Fig. 2 shows that module boundaries are often very abrupt. Thus, generation of new combinations of adjacent alleles (i.e., new mosaic section boundaries) is likely usually due to random nonhomologous recombination. The probability of generation of new combinations by random processes is clearly very low, but it does not seem unreasonable since phages are so abundant that random processes, even if they are infrequent, might be able to generate many different nonhomologous recombinants over evolutionary time. The probability of success is increased within a supercluster of phages by the fact that such phages have the same gene order, so a single rare non-homologous recombination event that happens to occur at the same location between two exchangeable DNA (mosaic) sections of two such phages will result in a genome that contains all the modules that are necessary to be a functional phage (i.e., multiple recombination events are not required). If such a rare recombinant happens to have an evolutionary advantage it will survive and thrive. Once such a nonhomologous recombination event has created a new module boundary, new combinations of nonadjacent module alleles can arise rapidly by homologous recombination within a module allele that two phages have in common.

These exchangeable modules appear to be evolutionarily shuffled more or less randomly within clusters, suggesting that each unit may have evolved to function at least semi-autonomously. This idea is supported by the fact that where information is available, intragenic boundaries coincide with protein domain boundaries, and laboratory construction of new combinations of modules (not previously observed in nature) usually creates a functional phage.

Genetic Exchange Among Superclusters

Rare successful recombination events have also clearly occurred between phages that belong to different superclusters, and most of these involve tail fibers or tailspikes. Two well studied examples are the tailspikes of phages P22, 9NA (Accession No. KJ802832) and Det7 and the long tail fibers of phages T4 (Accession No. AF158101) and lambda (Accession No. J02459). The first three phages are members of different superclusters in the *Podoviridae*, *Siphoviridae*, and *Myoviridae*, respectively. They have essentially no overt nucleotide sequence similarity except between C-terminal domains of the tailspikes that bind to the O-antigen polysaccharide on the surface of *Salmonella enterica* serovar Typhimurium. The second pair of phages above also reside in very different superclusters (a lytic and a temperate cluster) and are members of the *Siphoviridae*, and *Myoviridae*, respectively. They have similar long tail fiber genes, and the lambda fiber can functionally replace the T4 fiber. It seems reasonable that these islands of similarity in otherwise extremely distantly related phages must have been horizontally transferred by very rare nonhomologous recombination and have given the recipient phage access to a new, naïve host that allowed them to flourish.

Phage-host Relationships and Phage Diversity

Phage Cluster-Host Species Relationships With the *Enterobacteriales*

The relationship between phage clusters and their hosts must reflect their evolutionary history. But do phages simply co-evolve with their hosts over long evolutionary periods so that each phage cluster infects a single host taxon (species, genus, etc.), or is jumping to a new taxonomically very different host sufficiently common to disrupt this kind of simple co-evolutionary descent? This aspect of phage diversity is very poorly understood at present. The following paragraphs describe our current knowledge in this arena.

When phage cluster membership is compared with the hosts on which the phages were isolated, there is one strong conclusion. Very few phages from outside the *Enterobacteriales* have been found that clearly fall within *Enterobacteriales* phage clusters, so phages appear to jump between distantly related hosts quite rarely. We suppose that there are at least two major factors that limit such jumps across large host taxonomic chasms. (1) Virions recognize specific bacterial surface receptor molecules and disparate hosts are unlikely to have very similar surface features, so delivery of virion nucleic acid into very distantly related cells should be an uncommon event. (2) Perhaps more importantly, phages must be evolutionarily tuned to optimally infect their primary hosts

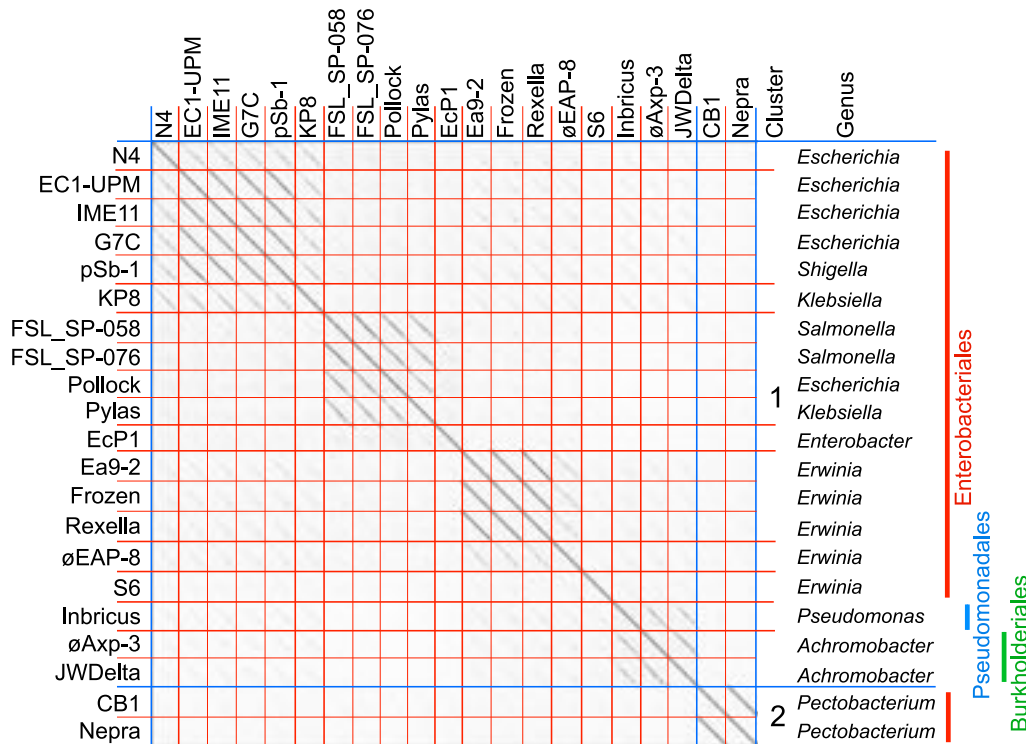


Fig. 3 The N4 supercluster dot plot. Representative N4 supercluster phages are included in the dot plot. Thin red lines separate genomes, thick red lines separate subclusters, and blue lines separate clusters. Clusters and host genera are indicated at the right and thick vertical colored lines there indicate the taxonomic order to which the hosts belong. Note that the orders *Enterobacteriales* and *Pseudomonadales* are in the *Gamma* *Proteobacteria* class, and *Burkholderiales* is in the *Beta* *Proteobacteria*.

in terms of the many ways in which their encoded proteins interact with host components, for example during the commandeering of host macromolecular synthesis machinery. Such tuning would mean that phages would likely infect distantly related bacteria with suboptimal efficiency and so would not compete well with the native phages there.

On the other hand, any given non-singleton *Enterobacteriales* phage cluster often contains phages that have different hosts. This suggests that either phages are more frequently jumping between closely related hosts, or that co-evolution of phages and hosts during host species separation has not driven the phages into different clusters. For example, *Enterobacteriales* clusters with more than four member phages often have hosts from multiple genera that are typically but not always within one family. *Mycobacterium smegmatis* phage Patience (Accession No. JN412589) is a counterexample of an unusual phage that may have switched to a very different host relatively recently. Nonetheless, closely related phages (phages that belong to the same subcluster) often infect the same host. Our 2014 analysis found that 78% of the non-singleton subclusters were populated by phages that infect a single host genus. Thus, there appears to be a strong tendency (but not a requirement) for closely related phages to infect closely related hosts.

We emphasize that phage-host relationships must always be viewed with some caution because of inherent complicating issues. The most prominent issue is that the full host range of most phages is not known. Rigorous determination of the full host range of any phage would require plaque assay or other tests for successful infection on an impractically large number of potential hosts. Therefore, careful and extensive host range studies are rarely performed due to the technically difficult, labor intensive and time consuming nature of testing a wide variety of bacteria, so nearly all host range studies have been limited to a small number of closely related potential hosts, which may not be indicative of true host range (see below).

Narrow and Wide Host Range Phages

Phages have a reputation for having very narrow host ranges, and many phages are indeed very host specific. Some phages like those whose virions adsorb to bacterial surface polysaccharides often successfully infect only a small fraction of cells in their host species. For example, phages such as *S. enterica* phage P22 and *E. coli* phage CUS-3, only adsorb to hosts whose surface polysaccharide has a repeating -mannose-rhamnose-galactose- unit or is polysialic acid, respectively. A few phage virions have up to four or five different tailspikes that allow them to adsorb to, and thus infect, hosts with several different surface polysaccharides. *Salmonella* and *Escherichia* strains are known that have over 200 different surface polysaccharides, so even phages that recognize several such receptors have rather narrow host ranges restricted to a small fraction of these species. At the other end of the scale there are reports of very broad host range phages that can infect quite disparate hosts, but “broad host range” is a poorly defined term that is commonly used to indicate the ability to productively infect many different strains within a species or multiple

relatively closely related species. Nonetheless, a few phages have been reported that are able to infect hosts from two different bacterial orders, *Pseudomonadales* and *Enterobacteriales*, within the class Gamma Proteobacteria (e.g., EMCL-117-like phage øFenriz, Accession No. KT254133), two different orders, *Chroococcales* and *Synechococcales*, within the phylum *Cyanobacteria* (T4-like phage Syn9, Accession No. NC_008296), and even two different bacterial classes, *Alpha* and *Beta* Proteobacteria, within the phylum *Proteobacteria* (phage P14, Accession No. KX660669). One study even suggested some phages are capable of infecting species across several bacterial phyla including *Proteobacteria*, *Actinobacteria* and *Bacteroidetes* (phages øFenriz and close relatives). Such reports of successful infection of extremely distantly related hosts by a single phage species are tantalizing but very rare and should be regarded with caution until the details of infection of the disparate hosts have been described. Nonetheless, if true such phages could be vehicles for spreading diverse phage genes across the bacterial spectrum.

Finally, we re-emphasize that phages from a given cluster are often able to infect different bacterial hosts, especially if the hosts are not very distantly related, indicating that particular lifestyles do not necessarily restrict phages to particular hosts within the *Enterobacteriales*. Fig. 3 shows this for the *Enterobacteriales* phage N4 supercluster, where members are known that infect seven different *Enterobacteriales* genera, *Escherichia*, *Shigella*, *Salmonella*, *Klebsiella*, *Enterobacter*, *Erwinia* and *Pectobacterium*. For example, *Escherichia* phages EC1-UPM (Accession No. KC206278), IME11 (Accession No. NC_019423) and G7C (Accession No. NC_015933) are very close relatives, while *Escherichia* phage Pollock (Accession No. KM236242) is quite different. On the other hand, *Klebsiella* phage Pylas (Accession No. MH899585) is more closely related to phage Pollock than it is to the other *Klebsiella* phage KP8 (Accession No. MG922974) in this group. Although the relationship of most of the phages with hosts outside the *Enterobacteriales* have not yet been compared in detail to all the *Enterobacteriales* phages, a few phages have been identified from other orders in the *Gamma-proteobacteria* class and even other classes in the *Proteobacteria* phylum that fall within the *Enterobacteriales* clusters. For example, Fig. 3 also shows that *Pseudomonas* phage Inbricus (*Pseudomonadales*, *Gamma* Proteobacteria; Accession No. MG018928) and *Achromobacter* phages øAxp-3 and JWDelta (*Burkholderiales* order in *Beta* Proteobacteria; Accession Nos. KT321317 and KF787094, respectively) fall within the N4-like cluster, while other *Enterobacteriales* phages CB1 and Nepra form a separate CB1-like cluster within the N4 supercluster. Similarly, other “outside” phages fall into the *Enterobacteriales* T4-like, T7-like and P2-like clusters or superclusters. Thus, similar molecular lifestyles (superclusters) can extend at least as far as across the *Proteobacteria*, and “T4-like” phages that infect the *Cyanobacteria* phylum suggest it may extend much further. It remains to be seen if such broad reaches are common or exceptions.

Summary

Bacteriophages, like all viruses, are certainly very ancient and appear to have arisen independently multiple times; for example, virions can contain single-stranded RNA or DNA or double stranded RNA or DNA, and the proteins that form the virion’s protective coats are present in different phage types with at least five unrelated polypeptide folds. The dsDNA tailed phages are extremely abundant in the environment and their diversity has been studied in most detail through comparison of whole genome sequences, so this discussion focused on their diversity. Both top down and bottom up methods have shown that the tailed phages are extremely diverse. Multiple types (called clusters) of tailed phages exist that infect essentially all bacterial species that have been examined, and different phages within such a cluster often infect different (but almost always closely related) host species. The number of bacterial species of Earth is not known, but estimates have ranged widely from millions to billions, so the number of tailed phage types or clusters is likely this enormous as well. In addition, very substantial diversity exists within each phage cluster. Thus, phages are almost incomprehensively enormously diverse in general features as well as details, and many of these are yet to be discovered.

Further Reading

- Aiewsakun, P., Adriaenssens, E.M., Lavigne, R., Kropinski, A.M., Simmonds, P., 2018. Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: Steps towards a unified taxonomy. *Journal of General Virology* 99, 1331–1343.
- Brum, J.R., Sullivan, J.B., 2015. Rising to the challenge: Accelerated pace of discovery transforms phage biology. *Nature Reviews Microbiology* 13, 147–159.
- Casjens, S.R., 2005. Comparative genomics and evolution of the tailed-bacteriophages. *Current Opinion in Microbiology* 8, 451–458.
- Casjens, S.R., Thuman-Commike, P.A., 2011. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. *Virology* 411, 393–415.
- Casjens, S.R., Grose, J.H., 2016. Contributions of P2- and P22-like prophages to understanding the enormous diversity and abundance of tailed bacteriophages. *Virology* 496, 255–276.
- Clokier, M.R., Millard, A.D., Letarov, A.V., Heaphy, S., 2011. Phages in nature. *Bacteriophage* 1, 31–45.
- Doore, S.M., Fane, B.A., 2016. The *microviridae*: Diversity, assembly, and experimental evolution. *Virology* 491, 45–55.
- Grose, J.H., Casjens, S., 2014. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family *Enterobacteriaceae*. *Virology* 468–470, 421–443.
- Guerin, E., Shkoporov, A., Stockdale, S.R., et al., 2018. Biology and taxonomy of crAss-like bacteriophages, the most abundant virus in the human gut. *Cell Host & Microbe* 24, 653–664.
- Hatfull, G.F., 2018. Mycobacteriophages. *Microbiology Spectrum* 6, GPP3-0026-2018.
- Hayes, S., Mahony, J., Nauta, A., van Sinderen, D., 2017. Metagenomic approaches to assess bacteriophages in various environmental niches. *Viruses* 9, 127.
- Hendrix, R.W., 2002. Bacteriophages: Evolution of the majority. *Theoretical Population Biology* 61, 471–480.

- Lawrence, J.G., Hatfull, G., Hendrix, R., 2002. The imbrolios of viral taxonomy: Genetic exchange and the failings of phenetic approaches. *Journal of Bacteriology* 184, 4891–4905.
- Mahmoudabadi, G., Phillips, R., 2018. A comprehensive and quantitative exploration of thousands of viral genomes. *eLife* 7, e31955.
- Manrique, P., Bolduc, B., Walk, S.T., *et al.*, 2016. Healthy human gut phageome. *Proceedings of the National Academy of Sciences of the United States of America* 113, 10400–10405.
- Pope, W., Mavrich, T.N., Garland, R.A., *et al.*, 2017. Bacteriophages of *Gordonia* spp. Display a spectrum of diversity and genetic relationships. *mBio* 8 (4), e01069-17.
- Rohwer, F., 2003. Bacteriophage diversity. *Cell* 113, 141.
- Roux, S., Hallam, S.J., Woyke, T., Sullivan, M.B., 2015. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *eLife* 4, e08490.
- Sepulveda, B.P., Redgwell, T., Rihtman, B., *et al.*, 2016. Marine phage genomics: The tip of the iceberg. *FEMS Microbiology Letters* 363, fnw158.
- Yutin, N., Backstrom, D., Ettema, T.J.G., Krupovic, M., Koonin, E.V., 2018. Vast diversity of prokaryotic virus genomes encoding double jelly-roll major capsid proteins uncovered by genomic and metagenomic sequence analysis. *Virology Journal* 15, 67.
- Yutin, N., Makarov, K.S., Gussow, A.B., *et al.*, 2018. Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut. *Nature Microbiology* 3, 38–46.

Relevant Websites

<https://phagesdb.org/>
PhagesDB.org.

<https://talk.ictvonline.org/taxonomy/>
Taxonomy.

Genetic Mosaicism in the Tailed Double-Stranded DNA Phages

Welkin H Pope, University of Pittsburgh, Pittsburgh, PA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

BLAST Basic Local Alignment Search Tool. Algorithm that identifies related nucleotide or protein sequences using local alignments, available at NCBI.

Capsid Protein shell that surrounds the genetic material of a phage, in the Caudovirales, the capsid is twenty-sided – icosahedral, and may be either isometric (all sides are the same length) or prolate (elongated).

Caudovirales In formal phage taxonomy, Caudovirales (“tailed viruses”) is the name of the Order of phages that use double-stranded DNA as genetic material; enclosed in a protein capsid and possessing a protein tail that is used to recognize the host cell and deliver the DNA from the head.

Circularly permuted Describes phage genomes from phages that use the headful packaging strategy; these phages do not have a defined start or end to their genome, but instead package DNA from a long concatemer of the genome sequence. The DNA is cut when the phage head is “full”, and includes more than 100% of the genome sequence resulting in identical sequences at the end of each piece of DNA within a phage head and each virion containing the complete set of genes with in the genome, yet each individual virion contains a different chromosomal sequence.

Direct repeat Property of some phage genome ends, in that the same DNA sequence is found at both the beginning and end of the piece of DNA within the phage head. Allows for circularization of the phage genome via homologous recombination after injection into the host cell, as well as for linear multimer concatemers.

Endolysin Phage encoded enzymes that cleave the bacterial cell wall from within, aiding the release of the virions at the culmination of the lytic cycle.

Exclusion Refers to the mechanism by which a cell prevents infection of phages via modification of cell surface proteins, preventing phage attachment and DNA injection.

Genome The complete DNA sequence of an individual phage.

Host A cell that a phage can infect and propagate within.

Host range The different bacterial strains that a specific phage can infect.

Identity Sequences that are identical.

Immunity Refers to the mechanism by which a cell is resistant to phage infection via expression and binding of an immunity repressor to the invading phage DNA.

Immunity repressor DNA binding protein encoded by the phage genome that prevents expression of phage genes involved in the lytic cycle. The immunity repressor is continually synthesized during lysogeny.

Integrase Enzyme that promotes homologous recombination of the prophage and the bacterial genome at attachment sites with the result of a stable single piece of DNA.

Lysogenic The cycle in which some phages (temperate phages) delay the lytic cycle after DNA injection, and

instead exist within a host cell as a quiescent piece of DNA (a prophage).

Lysogen-y A cell that contains a prophage is a lysogen, and exists in the state of lysogeny.

Lytic Describes the cycle of phage infection from DNA injection, through phage genome replication, synthesis and assembly of virions, and burst of the host cell. Also describes phages that undergo the lytic cycle (“lytic phages”).

Myoviridae The taxonomic Family name for the contractile tailed phages.

Pham (or phamily) Group of related protein sequences as calculated by the program Phamerator.

Plaque A clearing in a lawn of bacteria due to the presence of an initial single virion that subsequently underwent multiple rounds of the lytic cycle, killing the host cells in its proximity.

Plasmid An additional piece of DNA within a host organism, smaller than the larger genome, that is replicated and passed down to daughter cells during division.

Podoviridae The taxonomic family name for the short tailed (extensible tailed) phages.

Prophage The repressed genome of a temperate phage found within a host cell.

Recombination, homologous and non-homologous Process by which two pieces of DNA break and swap ends; such that pieces “A–B” and “C–D” become “A–D” and “C–B”. Homologous recombination occurs between identical sequences with the strands and happens more frequently than non-homologous recombination which requires no sequence identity.

Resistant A cell that is not susceptible to phage infection.

Similarity Sequences that show conservation of related amino acids in corresponding positions, even if the sequence is no longer identical.

Siphoviridae The taxonomic Family name for the non-contractile tailed phages, also called the flexible-tailed phages.

Synteny The conservation of gene order observed within phage genomes.

Tail Protein structure that mediates host recognition and DNA delivery in the tailed phages. Tails are comprised of multiple different proteins and are found in a number of morphologies.

Tail fiber The long, thin, extended proteins attached to the tail tip that are the primary mechanism by which a phage binds to its host cell.

Tape measure The protein that makes up the core of the flexible-tailed phages, the length of the tape measure is directly proportional to the length of the phage tail.

Temperate Describes phages that are capable of delaying the lytic cycle after their DNA is injected within the host cell; instead the phage undergoes the lysogenic cycle in which DNA is replicated within the cell (called lysogens) and passed onto daughter cells after cell division. When host cells are stressed, the lytic cycle may be induced.

Terminase ATP-driven DNA packaging motor encoded by phage genomes. By convention, the terminase gene is frequently assigned position 1 in the genomes of flexible-

tailed phages; in the short-tailed podoviridae, the terminase is found towards the right end of the genome.

Virion An individual virus or phage particle.

Introduction

Bacteriophages – or phages, the viruses of bacteria – are the most numerous biological entities on the planet, with an estimated global population of 10^{31} , and a turn-over time of several days – these numbers are based on measurements taken by the Colwell lab in 2000, and Suttle lab in 2007. Phages are comprised of nucleic acid – dsDNA, ssDNA, dsRNA, or ssRNA – surrounded by a protective proteinaceous shell. The majority (over 90%) of the sequenced phage isolates in the non-redundant “Nucleotide” collection in GenBank as of February 2019 are members of the Caudovirales (also called the tailed phages), these phages use dsDNA as genetic material surrounded by an icosahedral “head” (or capsid), and at a single vertex is attached a complex, multi-protein structure called a “tail” which mediates recognition of the bacterial host. Tailed phage morphology broadly falls into three groups and is the basis for the taxonomic classification of these phages: the Myoviridae, or contractile-tailed phages; the Siphoviridae, the non-contractile or flexible tailed phages, and the Podoviridae, the short-tailed phages. Numerous isolates of each morphological type have been recovered from a variety of hosts and habitats; comparatively fewer of these have had their genomes sequenced and analyzed using tools to predict the location and functions of their genes, and to determine how similar they are to each other. Comparisons of the genome sequences in smaller, host specific-sets, as well across the broad collection as whole, have illuminated a number of properties of phage genes and genome sequences conserved throughout the morphological types, as well as allowed scientists to gain insights into global viral genetic space. This includes the property of “mosaicism” – as phage genomes are collections of individual genes with separate evolutionary histories – much like a mosaic is a piece of art comprised of many different individual tiles. Independently, the tiles – or genes – are less meaningful, but together, the tiles form a single large mosaic picture, just as phage genes collectively encode the instructions for the life cycles of a single phage.

Evidence for Genome Mosaicism Before the Advent of DNA Sequencing

The mosaic nature of tailed bacteriophage genomes was first described in the hybridization studies of DNA from *Escherichia coli* phage lambda and its relatives in 1971. DNA from these phages was melted to single strands, mixed, and allowed to re-anneal and form heteroduplexes; these were examined with transmission electron microscopy, and the lengths and locations of the bound, complementary sections versus unhybridized pieces carefully mapped. It was readily apparent that there were portions within each of the phage genomes that had sufficient identity to form stable duplexes with the others interspersed with regions of non-identity; and that these sequences were not uniformly distributed across the different phage genomes. Some phages shared certain portions of their genomes; other portions of their genomes were shared with other phages that did not contain the first segment.

Subsequent genetic studies of the *Salmonella* phage P22 identified genes and regions of genome organization in common with the *E. coli* lambdaoid phages – surprising investigators who had grouped the P22-like phages with the *E. coli* T7-like phages based on their shared short-tail morphology. These observations of shared genetics between morphologically dissimilar phages, among others, led Susskind and Botstein (1978) to propose a method of phage genome evolution based on swapping of modules of genes; suggesting that this was accomplished by homologous recombination – breaking of the two DNA strands at identical sequences followed by rejoining to the other phages’ genome, creating a new overall sequence. They proposed this was promoted by the existence of short, identical “linker” sequences flanking related modules; such a mechanism could result in the observed mosaic patterns between genomes. In phages in which no linker sequences could be identified, they proposed that periodic loss of linker sequences through mutation could occur, thus preventing their identification in sequence comparisons.

DNA Sequencing Illuminates Properties of Bacteriophage Genomes

With the advent of DNA sequencing, phages were an obvious target due to their short genomes and tractability as a model system; and indeed, the first complete genome sequence of any entity reported belonged to the phage phiX174 by Sanger (1977), with the larger phage lambda following in 1983. Subsequent genome sequencing and comparative analyses through the 1990s revealed sequence similarity and shared genome architecture from phages across of a variety of hosts, including Gram negative *E. coli* and *Salmonella* phages as well as Gram positive *Mycobacterium* and *Streptomyces* phages. Indeed, in 1999, sufficient phage genomes had been fully sequenced such that Hendrix *et al.* proposed that all tailed phages were related through common ancestry, had access to a global pool of genes; and that the variation of identical sequences flanked by unrelated sequences, mosaicism, was a universal property of phage genomes. Interestingly, the mosaicism was not distributed at random through the genomes, but instead mapped most often to gene boundaries, such that intact genes were shared between otherwise unrelated genomes. Nor was

the mosaicism limited to a few mobile elements, such as transposons, found across a large number of genomes, but was identifiable throughout all regions and genes within phage genomes.

With the analysis of additional phage genome sequences, genetic mosaicism continued to be observed throughout phages of different hosts and morphologies, and it became evident that homologous recombination between linker sequences facilitating the exchange of genetic modules could not be the primary driver behind genome diversity and evolution, as multiple versions of genes with significant sequence similarity were found across otherwise unrelated phage genomes. Instead, Hendrix *et al.* (1999) proposed that the driving force behind the mosaicism present in phages genomes must be non-homologous recombination. This argument was based on the sheer magnitude of the numbers involved, with a rationale along these lines, with 10^{31} phage virions on the planet, and a predicted 10^{23} productive infections per second, a certain number of those infections would be co-infections, bringing related or unrelated phage genomes into direct contact with each other inside the same host cell. Within those co-infections, if a fraction of those underwent a nonhomologous recombination event, there could be on the order of 10^{16} events per second. The majority of these non-homologous recombination events are likely to produce non-viable progeny – if you swap the center of your capsid gene for the middle of the other phage's DNA polymerase gene, both progeny will both be unable to produce new phages in the next generation – and so, the recombination events that produce viable progeny will be more likely to occur at gene or protein domain boundaries at a rate of approximately 10^9 per second, – thus leading to the observed pattern of genetic mosaicism in phage genomes.

Comparative Genomics of Complete Bacteriophage Sequences Shows Diversity and Mosaicism

As DNA sequencing costs dropped, more phages were completely sequenced; including groups of *Lactococcus* phages and *E. coli* phages. A detailed analysis of the genomes of lambda-like *E. coli* phages, including HK97 and HK022 by Juhala *et al.* (2000) demonstrated that the mosaicism present in the genomes was not limited to gene or module boundaries, even those sites were overrepresented in the reassortment of the genes between genomes. Indeed, recombination events had clearly taken place between phages within coding sequences, although these were not evenly distributed across all genes, as more events were identified between tail fiber genes, than between others, suggesting that the mechanism behind these patterns was rampant non-homologous recombination between co-infecting phages or prophages that were previously integrated in the host cell. The prevalence of mosaic events in tail fibers could be part of a mechanism to increase the rapid diversification of host range, or could be an artifact of the linear structure of the fibers – long extended proteins like tail fibers could more easily tolerate non-homologous joins than similar events in globular proteins, which could be more likely be lethal to the phage. Closer examination of genomes from phages across hosts and virion morphology demonstrated that even though nucleotide and protein sequences of the genes were unrelated, the order and location of genes within phage genomes, particularly amongst phages with shared morphologies, was conserved. This property is called “synteny”, that analogous genes appear in the same order within phage genomes.

The vast diversity of phages was illuminated by an analysis by Pedulla *et al.* (2003) of a collection of fourteen complete genome sequences from phages infecting the same host, *Mycobacterium smegmatis* mc²155. These phages were all tailed phages, and predominantly flexible tailed (two were contractile-tailed); and many of them were indistinguishable from each other by electron micrograph. However, comparative analysis of the genome sequences revealed that these genomes had *less than 50%* nucleotide sequence identity. This was a tremendous surprise to the scientific community – these phages that appeared identical under the microscope and were in direct genetic contact with each other through their ability to infect the same host did not share significant DNA sequence identity (for comparison, humans share about 60% of their genome sequence with bananas). Not only did these phages not share sequence similarity with each other, the majority of genes within phage genomes were unlike any other sequences in our sequence databases, that only 20% of the genes within a phage genome could be assigned a putative function.

The genes that could be assigned functions, both during these studies in the early 2000s and up through the writing of this article in 2019, were those that played well-characterized roles in the phage lifestyles. During the lytic cycle, after binding to a host cell, the phage nucleic acid is deposited into the cytoplasm, where it undergoes transcription, translation, and replication; the structural proteins are synthesized, the heads and tails self-assemble, DNA is packaged into the head, the head and tail are joined; and the cells are lysed releasing the new virions into the environment. Temperate phages can undergo the lysogenic cycle, during which they integrate their DNA into the host genome or circularize as a large extrachromosomal plasmid and repress their lytic functions by means of a protein called the immunity repressor; the phage DNA is then replicated by the host during normal cell division and passed on to both daughter cells. Lysogens, the cells that contain these phage genomes (called prophages) sometimes gain new abilities via genes expressed from the prophage, included those involved in host defense against subsequent infection from other phages; lysogens can also undergo induction of the lytic cycle during stress, and the prophage can excise from the host genome and begin the lytic cycle. Many of the proteins involved in these processes are phage-encoded, and can be routinely identified in phage genome sequences based on sequence similarity to those genes in other phages that have been characterized at the bench. The different branches of the Caudovirales – Siphoviridae, Myoviridae, and Podoviridae – demonstrate conservation of some genome properties with regard to genetic content as well as variations specific to their morphology; each type is broadly discussed below.

Mosaicism in the Siphoviridae

The flexible, non-contractile tailed phages, the Siphoviridae, have been isolated from all known bacterial phyla and habitats; many of the more recent large scale comparative genomics studies describe phages of the Siphoviridae. Within the genomes of the Siphoviridae (Fig. 1),

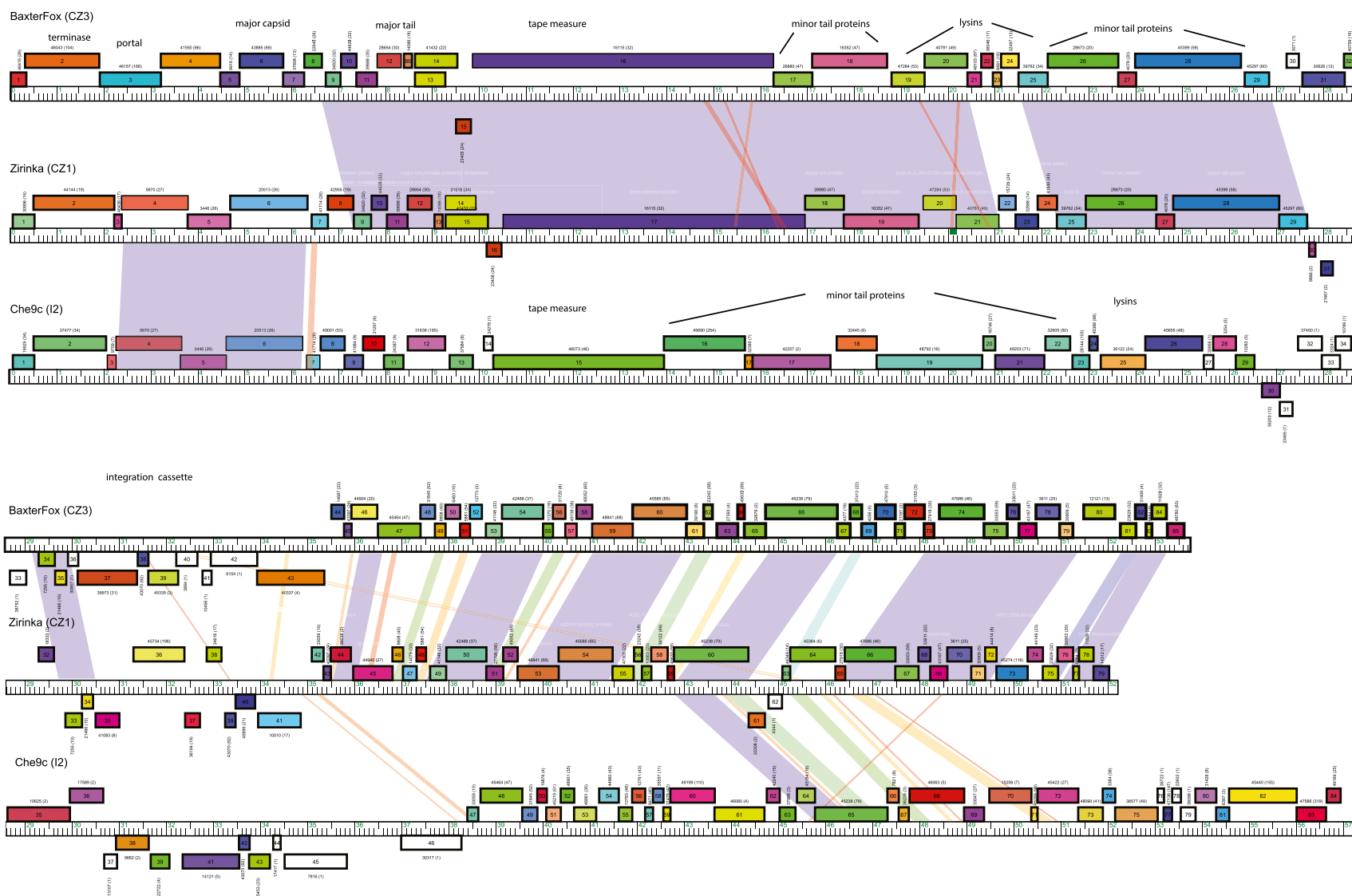


Fig. 1 Three phage genomes – *Gordonia* phages BaxterFox and Zirinka, and *Mycobacterium* phage Che9c – are compared with the program Phamerator. The letters/numbers after each phage name indicates their cluster and subcluster designation. The central ladder indicates the length and position of the nucleotide sequence, in kilobases. Genes are represented by rectangles, and colored according to protein sequence similarity; genes in white share no protein sequence similarity with any other proteins in the Actino-Draft database. Numbers within the rectangles indicate the gene numbers; numbers above the rectangles indicate pham number, the number in parenthesis indicate the number of genes within that pham from the entire database “Actino-Draft”. The direction of transcription is indicated by the placement of the rectangles above or below the central ladder; above, indicates left to right. The colors between the genomes – as opposed to the colors of the gene rectangles – indicates nucleotide sequence similarity as calculated with the program BLAST, purple represents the most similar sequences, shading to red as the least similar, and white indicates no significant sequence similarity.

synteny is evident in the order of the virion structural genes in particular; by convention, newly sequenced phage genomes are oriented such that the structural genes are on the left-side of the genome, and the gene(s) encoding the terminase, the DNA packaging motor, is one of the first, if not the first, gene. Following the terminase are genes encoding the head and tail components, specifically, the portal protein, the capsid maturation protease, the scaffolding protein, the major capsid protein, the head-to-tail connector proteins, the tail terminator, the major tail subunit, the tail assembly chaperones, the tape measure, the minor tail proteins, and the tail fiber. The number and size of the structural genes, and the conservation of their order is pervasive, from the very smallest flexible tailed phages with genomes of approximately 15 kb; to the largest, with genomes closer to 100 kb; and is observed in phages of hosts in multiple phyla, in both Gram-positive and Gram-negative bacteria, and from diverse environments. Interestingly, while the order of these genes is predominantly maintained, the larger phage genomes have more gene insertions between various structural proteins; these genes include parasitic elements like transposons, self-splicing introns, and homing endonucleases, as well as a large number of genes without identifiable functions.

Adjacent to the structural genes are the genes predicted to encode proteins involved in cell lysis, including the endolysin(s) and holin; these are followed by a central region of the genome in found only in temperate phages containing the integration cassette, including the integrase, the attP site (for “att”achment “p”hage, the DNA sequence that is identical to the integration site in the host) and the immunity repressor, which prevents expression of genes involved in lytic functions. The integration cassette region in many temperate phages is highly variable, and otherwise closely related phages can display a subset of a wide range of diverse small genes in this region. These genes are expressed from the prophage during lysogeny, and at least some of them aid in defense of the host from super-infecting related or unrelated phages, via restriction, exclusion, toxin/anti-toxin systems, abortive infection systems, and others; the mechanism(s) that contribute to this extensive localized genetic diversity is unclear.

The right arm of the flexible-tailed phage genomes is replete with small genes (~30–500 aa) of unknown function; as well as larger genes for which functions can be identified; those code for proteins involved in DNA replication, recombination, and nucleotide modification and processing. The smaller genes are characteristic of phage genomes in general, their role in the biology of the phage is unclear, nor are they similar to other sequenced phage genes available in GenBank. These genes of unknown function tend to be smaller than their counterparts to which a function can be assigned, likely both due to the lack of similar sequences of known function being present in the databases, but also due to the limitations of search algorithms in matching smaller sequences with statistically relevant specificity.

Transcriptomic and proteomic studies indicate that these small genes are expressed early during lytic infection, and in at least a few cases, are involved in the defense of the host cell against super-infecting phages – either self or unrelated. A putative role in host defense is further supported through whole genome essentiality studies performed – most of these small genes can be deleted from the genome with no effect on the viability of the progeny under laboratory monoculture conditions. This would make sense if these smaller genes are used to protect the host from further infection of other phages.

Mosaicism in the Myoviridae

The larger comparative genomic studies of the Myoviridae have primarily focused on the on the T4-like phages in Gram negative hosts such as *E. coli* and the marine Cyanobacteria *Synechococcus* and *Prochlorococcus*; indeed, while myoviridae of the Actinobacteria have been isolated, they do not exhibit the large complicated tails of the T4-like phages, and the two types of phage share no genes in common. The genomes of the T4-like phages generally are between 150 kb and 250 kb; they share some gene content in the form of a set of 20 or so “core” genes, present in all of these phages; core genes in this case being defined by functional analogy and protein sequence similarity. The majority of these shared genes are homologs of the structural genes of T4; T4 virions have a complicated structure and elaborate, multi-protein contractile tail designed to deliver the phage DNA by injection through the cell membrane and wall. To date, no temperate phages with a T4-like morphology have been isolated; all are lytic phages. Like the Siphoviridae, the overall gene order of the core genes within the Myoviridae is conserved; and indeed, the order from terminase through major capsid and neck proteins is identical to those of the Siphoviridae. The mosaicism of phage genomes is evident within the insertions of smaller genes of unknown functions between the structural proteins similar to the larger genomes of the Siphoviridae, the overall synteny of the structural proteins is maintained in these larger phages.

Mosaicism in the Podoviridae

The taxonomic classification of Podoviridae, or short-tailed phages, comprises at least two large unrelated groups; the short-tailed phages that are similar to the *Salmonella* phage P22 are more related to the lambdoid phages of *E. coli* such that the melted strands of the genomes can anneal and form heteroduplexes as demonstrated by Campbell (1994), than these phages are to other short-tailed phages. The larger comparative studies performed on Podoviridae have primarily included the phages related to *E. coli* phage T7, including phages of the Enterobacteria and phages of the marine cyanobacteria. These phages share little nucleotide or protein sequence similarity, but their genome size of approximately 40 kb, genome architecture, and functional gene content are highly conserved. The synteny of these genomes is evident in the comparative studies, with an RNA polymerase gene near the left end, followed by genes involved host takeover and in DNA replication, with the structural genes occupying the end of the right arm. Here there is a slight permutation from the gene order in the Siphoviridae; as the T7-like phage structural genes begin with the

portal gene, followed by the head, and tail genes, and end with the terminase. The smaller genes of unknown function show the most variety in terms of gene content across genomes, with the larger structural genes showing more conservation. Interestingly, some of these genomes include genes related to integrases found in temperate phages, but do not include identifiable homologs of the remainder of the machinery required for stable integration.

Clusters and Superclusters: Large Scale Comparative Genome Analysis

The mosaic nature of phage genes and genomes and the lack of a single gene present in all phages has made defining the boundaries of their relatedness a difficult task. A number of classification schemes have been used to group -related phages together; including morphology and host range. As more phage genomes were sequenced through the 2000s and 2010s, the more it became clear that these earliest metrics for classification were not sufficient; *Salmonella* phage P22 and *E. coli* phage lambda being the only two earliest examples.

The larger scale comparative analyses of complete genome sequences of phages infecting specific bacterial genera published in the past decade have upheld the earliest observations of genetic mosaicism across the tailed phages. Phages isolated from a multitude of hosts share genome architectures and functional units while exhibiting little or no nucleotide or protein sequence similarity; and many of the predicted protein encoding genes are small, only 300 bp or so on average, diverse, located in equivalent areas of the genomes, and do not have identifiable functions. The ubiquitous mosaicism and vast diversity of gene sequences has confounded traditional taxonomy based on vertical descent of genes; however, attempts at grouping phages by shared nucleotide sequences or by shared proteomes – also called shared gene content – have proved useful in identifying groups of related phage that share similar biological traits. Phages with similar gene content are assigned to the same “cluster”, however, these boundaries may still be fuzzy, in that phage A may share 35% of its genes with phage B, and phage B may share 35% of its genes with phage C; all three would be assigned to the same cluster even if phages A and C do not share 35% of their genes. The results of this type of cluster building is that given enough members, there may no longer be any single protein that is present in all cluster members; this is seen in the Cluster A of the Actinobacteria. These phages are temperate Siphoviridae with approximately 600 isolates with complete genome sequences listed on website provided in “Relevant Website section”, the Actinobacteriophage database. The most recent comparisons of phage gene content suggests that phages with similar genome architecture and sharing 35% or more protein sequences from across the genome, when shared proteins are measured through pham membership as determined by the program Phamerator, can be used to define cluster membership.

An analysis of Entereobacteriophages suggests the use of the term “supercluster”; these would include phages that share genome architecture, but few or no protein sequences across the group. The phages of the superclusters show diversity of host range across multiple genera; and provide evidence for the model of gene movement throughout the global population of phages proposed by Jacobs-Sera *et al.* (2012), in which different phage types exhibit different but overlapping host preferences (the “stepping stone” model). Recombination events between these co-infecting phages allows all phages access, albeit at different rates, to the continuum of diversity within global phage genetic space.

Different Phage Clusters Acquire New Genes at Different Rates

Comparative genomic analysis between and within phage clusters demonstrated that the rates of gene exchange and mosaicism were not equivalent between different groups of phages. With more than 3000 sequenced phages isolated on hosts from across the domain of bacteria, patterns in gene exchange and gene flow became evident. By examining the gene content and nucleotide sequence similarity of over 2000 phages in direct pairwise comparisons, Mavrich and Hatfull demonstrated that different phage clusters exchange genes at different rates; and moreover, that only temperate phages exhibit “high gene flux” – in which parts of their genomes are relatively identical in terms of gene content and nucleotide sequence (frequently, this corresponds to the left arm of the genome containing the structural proteins) – while the remainder of the genomes shares little to no genome sequence similarity or gene content. High-gene flux phages are more likely to swap entire genes than accumulate nucleotide changes within shared genes. This type of pairwise comparison is in direct contrast to the phages that exhibit “low gene flux”; and in these phages, changes between genomes are spread out evenly across the genome, and nucleotide changes within shared genes occur at about the same rate that entire genes are swapped between genes. To date, all lytic phages and some groups of temperate phages demonstrate low gene flux. These two different modes account for the seemingly irreconcilable observations of discreet phage populations observed within large lytic marine Cyanophages – these are low-gene flux phages with many shared genes, for which individual rates of nucleotide substitutions can be counted; while some groups of the temperate Actinobacteriophages within the high-flux mode share structural genes and little else. It is unclear as to the mechanism through which these phages mediate high gene flux.

Mosaicism, Hybrids, and Single Gene Surveys

Today’s sequencing technologies rely on shearing DNA into small, manageable fragments (100–1000 bp, depending on the sequencing technology used) prior to sequencing, followed by post-sequencing computer assembly of the small pieces into a

complete full-length genome. For purified DNA from phage isolates, assembly is trivial, however, for DNA sequences extracted from complex environmental samples (also called the metagenome of the sample) the mosaicism, vast numbers, and genetic diversity of the phage genomes makes assembly difficult. Nor is it easy to identify phage sequences recovered from metagenomic samples as there are comparatively few full genome phage sequences available for alignments. Some environmental survey studies use amplification of genes such as those encoding phage portal proteins, major capsid proteins, integrases, or DNA polymerase genes, and these are of some utility in determining the presence and diversity of phages containing those particular genes. While the majority of complete phage genome sequences support that single gene survey analyses may represent individual phage types, some exceptions have been discovered and are worth mentioning. First, a number of phage genomes of the Siphoviridae and Podoviridae from diverse hosts such as *Gordonia* and the marine cyanobacteria contain integrases without any other lysogenic components; nor have efforts at isolating lysogens from infections of these phages been successful. In these cases, the “integrase” genes may be merely homologous recombinases that facilitate the gene flow between these phage types and the host or co-infecting phages. Second, structural hybrids have recently been identified between the *Gordonia* and *Mycobacterium* phages (including the three shown in Fig. 1); the bottom two phages exhibit prolate capsids and the top has an isometric head. A similar relationship exists between the phages depicted in the figure, which illustrates a number of properties discussed in this article; including mosaicism, synteny, and conservation of genome architecture; these are not the only examples of structural hybrids, the marine phages of *Pseudoalteromonas* include a Myo/Sipho hybrid – in this case, the contractile-tailed phage is similar to the smaller less complex phage Mu rather than the larger complicated T4.

The rate and formation of hybrids is dependent on the size, type, and locations of functionally equivalent genes. The ratio of genome overall length to that occupied by essential genes directly impacts the amount of flexibility a phage genome has to acquire new genes without losing viability. As capsid sizes increase, phages have room to acquire new genes, but these new genes may be in the form of insertions between essential genes. As essential genes are moved farther apart from each other on the genome, the less likely it is that a non-homologous recombination event could accommodate the formation of a structural hybrid – thus it is unlikely that a lambda-T4 hybrid could even be recovered. It is also unlikely that a phage with a larger capsid size could downsize once enough new genes have been acquired; as truncation of the genome sequence may then include essential genes.

Additional genome sequencing of phages from a wider variety of phages may reveal more insights into the gene flow within the phage population.

Further Reading

- Bondy-Denomy, J., Qian, J., Westra, E.R., *et al.*, 2016. Prophages mediate defense against phage infection through diverse mechanisms. *ISME Journal* 10 (12), 2854–2866.
- Brussow, H., Hendrix, R.W., 2002. Phage genomics: Small is beautiful. *Cell* 108 (1), 13–16.
- Cotinho, F.H., Silveira, C.B., Gregoracci, G.B., *et al.*, 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications* 8.
- Dekel-Bird, N.P., Avrani, S., Sabehi, G., *et al.*, 2013. Diversity and evolutionary relationships of T7-like podoviruses infecting marine cyanobacteria. *Environmental Microbiology* 15 (5), 1476–1491.
- Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., *et al.*, 2014. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* 513 (7517), 242–245.
- Dutilh, B., Cassman, N., McNair, K., *et al.*, 2014. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature Communications* 5.
- Grose, J.H., Casjens, S.R., 2014. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. *Virology* 468–470, 421–443.
- Hendrix, R.W., 2002. Bacteriophages: Evolution of the majority. *Theoretical Population Biology* 61 (4), 471–480.
- Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., Hatfull, G.F., 1999. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America* 96 (5), 2192–2197.
- Mavrich, T.N., Hatfull, G.F., 2017. Bacteriophage evolution differs by host, lifestyle and genome. *Nature Microbiology* 2, 17112.
- Pedulla, M.L., Ford, M.E., Houtz, J.M., *et al.*, 2003. Origins of highly mosaic mycobacteriophage genomes. *Cell* 113 (2), 171–182.
- Pope, W.H., Bowman, C.A., Russell, D.A., *et al.*, 2015. Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *elife* 4.
- Pope, W.H., Mavrich, T.N., Garlena, R.A., *et al.*, 2017. Bacteriophages of *Gordonia* spp. Display a spectrum of diversity and genetic relationships. *mBio* 8 (4).
- Rowher, F., Hisakawa, N., Youle, M., Maughan, H., 2014. *Life in Our Phage World*. Wholon Publication.
- Yuhan, Y., Gao, M., 2017. Jumbo phages: An overview. *Frontiers in Microbiology* 8 (403).

Relevant Websites

- <https://www.ncbi.nlm.nih.gov/>
NCBI – NIH.
- Phamerator.org
Phamerator.
- Phagesdb.org
The Actinobacteriophage Database.

Bacteriophages of the Human Microbiome

Pilar Manrique¹, The Ohio State University, Wexner Medical Center, Columbus, OH, United States

Michael Dills¹ and Mark J Young, Montana State University, Bozeman, MT, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Bacteriophage-phage Virus of bacteria.

Diversity The number of species, their taxonomies, and their abundance in an ecological community. Generally represented by a diversity index, which is a quantitative measure of both, the richness and evenness of a community (e.g., Shannon Diversity Index).

Dysbiosis Microbial system imbalance within body. Dysbiosis is the disruption of a microbiota's commensal and mutualistic capacity. Dysbiosis is not a requisite criterion in health or disease though it is implicated in the etiology, persistence and progression of many diseases.

Eubiosis The healthy state of the human microbiome communities.

Fecal microbial transplantation or transplant (FMT) Transplantation of feces from a healthy individual donor to the gut of a recipient individual suffering a disease in order to restore a healthy microbiota.

Gnotobiotic mice Germ-free mice (devoid of microbes) that are colonized with exogenous microorganisms in order to study the effect of the microorganism on the host.

Human microbiome The human body in total as a conglomeration of human cells and microorganisms (the microbiota). The phage component is in turn referred to as the human phageome.

Lysogenic or temperate phage A phage that has the ability to integrate its DNA into the host chromosome where it is passively replicated with the bacterial chromosome.

Lytic phage A phage that upon injection of its DNA into its susceptible host cell replicates its genome, produces structural proteins and ultimately lyses the cell to release its progeny virions.

Metagenomics Study of DNA isolated directly from complex biological samples. DNA is isolated and deep-sequenced generating sequence reads that are assembled with bioinformatic programs to render an ensemble of DNA sequences from all the organisms present in the microbial community under study. Within this context, viruses can be purified and sequenced separately, rendering a viral metagenome containing sequences from viral entities. In contrast, cellular metagenomes will contain both cellular and viral sequences.

Phage therapy Use of specific bacteriophage or phage cocktails to treat bacterial infections.

Prophage reservoir All the prophages encoded in bacterial members of the gut microbial community.

Prophage A bacteriophage genome integrated in the bacterial host genome.

Introduction

Humans exist in homeostasis with a considerable diversity of microbial life, ranging from bacteria and their viruses-termed bacteriophages (phages), to eukaryotic viruses, archaea, protozoa, and fungi. Together these organisms comprise the human microbiome. Within this community, bacteria are the most abundant cell type and bacteriophages are the most abundant virus type. The composition of the human microbiome varies between individuals and body site, diet, geography and health status. Disease states can correlate with deviations in microbiome structure, termed dysbiosis, with respect to that of the healthy state. The causes of dysbiosis are a central focus of microbiome research.

Bacteriophages are a key ecological determinant affecting the dynamics and function of bacterial communities. Their role in shaping human bacterial communities, and ultimately in human health and disease, remains a key question. The study of phage-bacteria-human relationships spans the paradigms of health and disease, with phages being agents of both. Individual phages have been extensively studied as secondary agents of human disease, often conferring pathogenic traits to commensal bacteria. In other microbial systems bacteriophages are known to strongly influence the dynamics and function of their bacterial host communities. It is assumed therefore, that microbial viruses are partially responsible for the human microbiome's capacity to modulate host phenotype and cause, enhance or ameliorate disease. This assumption is underpinned by evidence that viral fractions of the microbiota alone can generate systemic changes in health. Phages can potentially influence the human host through predation of human pathobionts, and through direct interaction with the human immune system among other modes.

The human phageome has been studied at both a multi-trophic systems level, as well as at the individual virus level. To this end, a number of tools have been developed to further our understanding of the features and dynamics of bacteriophages within the human microbiome. Quantification of purified phages using fluorescent DNA intercalating dyes and direct visualization by electron microscopy have allowed for the estimation of phage densities and provided information on the virion morphologies of those communities. Advances in 'omics' technologies (primarily genomics) and bioinformatic tools have permitted in-depth

¹The authors contributed equally to this work.

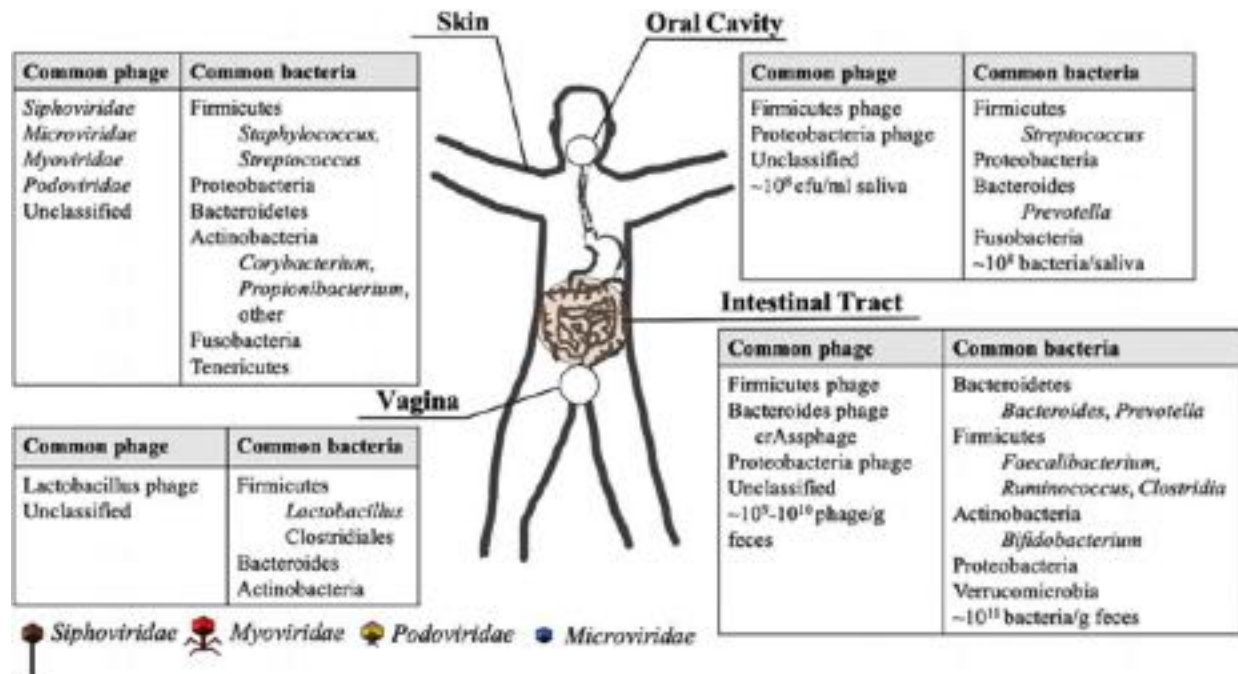


Fig. 1 Human microbiome phages across several body sites.

characterization of human phages at the community level. Specifically, next generation sequencing technologies (NGS) have provided a more complete understanding of phage diversity and community structure associated with different human developmental and health states. Mouse models, both gnotobiotic mice which can replicate specific bacteria-phage dynamics or community dynamics, as well as conventional mouse disease models which replicate disease-associated changes in humans, have been used to investigate the overall phageome composition and its influence on host health. The increase of human fecal microbial transplantation (FMT) to treat several gut associated diseases has provided new opportunities to study human gut phage temporal ecology. *Ex-vivo* and *in-vitro* models have been employed to decipher the same dynamics as the other systems mentioned as well as mechanisms of spatial localization for certain phage. This combination of research tools has provided invaluable insights into the role of phages in health and disease.

This article examines phages in human microbiomes from three different perspectives: the distribution and composition of the human phageome across individuals and body sites, the dynamics of the phage community within specific body sites, and the applications of bacteriophages for human health. Each of these aspects exists at the intersection of multiple dimensions including time, intra- and inter-individual variation, health and disease.

Human Phageome Distribution and Composition

Bacteria and their viruses populate nearly all external features of the body (Fig. 1). These sites range from highly habitable to depauperate. Like the bacteriome (the collection of all bacteria present in a given microbiome), the composition of human-associated phage communities is significantly affected by body site. Micro-habitats of the skin, for instance, are populated by fairly simple resident microbial and viral communities. These communities are generally composed of *Staphylococcus* and *Streptococcus* phage, such as *Staphylococcus capitis* phage STB20, being the dominant prokaryotic phage. Itinerant microbes pass through these habitats frequently from other environments. In contrast, the vaginal environment hosts a more complex and overall more stable microbial community that is usually dominated by *Lactobacillus* species and lactobacillus phages. Gastrointestinal (GI) tract communities are dominated by *Streptococcus* and *Prevotella* species, and presumably their viruses in the oral cavity, and members of the Bacteroidetes, Firmicutes, Proteobacteria and Actinobacteria phyla and their viruses in the gut. The majority of microbes make their home in the GI tract, which is one of the main immune organs in the human body. The GI tract is therefore the largest interface between microbes and human immune cells; and the largest microbe-microbe interface, making it a fruitful area for microbiome research. It encompasses the majority of studies carried out on the human phageome.

Beyond external surface habitats, phages are also found within the body. Active phage particles have been isolated from blood and organs. The process of phage transcytosis, in which phage particles passively move across epithelial layers, likely contributes to their internal presence, although active transport systems may also exist. Using tissue culture models of gastrointestinal epithelium, it has been estimated that approximately 31 billion bacteriophage particles cross the gut epithelium into the body every day. The

presence of internal active phage particles may provide a cooperative immune function for the human host, although its exact implications on both the human host and microbial dynamics have yet to be determined.

Along the length of the GI tract there is significant variation in the abundance of bacteria and viruses. Microbial cell densities are high in the oral cavity ($\sim 10^8$ /ml of saliva) and decline towards the stomach. Beyond the stomach, cell densities increase progressively through the small intestine and number in excess of 10^{11} cells per cm^3 in the colon. Phages reside in the lumen or associated with mucosal membranes (e.g., gums, gut mucus) in similarly high numbers to their bacterial hosts. It has been estimated that between 10^9 and 10^{11} virus particles per cm^3 are present in the lumen compared to 10^8 – 10^{10} virus particles per cm^3 in the mucosa. However, accurate measures of phage densities across both radial and longitudinal axes are difficult to determine. The ratio of phages to bacteria varies between 1:1 and 1:10, which is significantly lower than the phage to bacteria ratios observed in other environments, like the ocean, where phages commonly outnumber bacterial hosts. The differences in phage to bacteria ratios observed between different environments may be linked to cell density, with high cell density favoring lower phage to bacteria ratios. Understanding the spatial and temporal dynamics of phage-bacteria associations at a finer scale will provide critical information on role of phages within the GI ecosystem.

Within an individual, tens to hundreds of phage species can be found at any given time in either the oral cavity, the small intestine or the colon. The highest phage load is found in the colon, followed by the oral cavity and small intestine, however, the oral cavity hosts the most diverse phage community. The majority of phages metagenomic surveys have been carried out in the oral cavity and in stool samples, which serve as a proxy to study the colon microbiome. In both environments, the phage community is relatively novel since >60% of the phage DNA sequences found in these studies are not represented in cultured phage isolates. However, the overall phage diversity is lower than what is observed in other microbial environments. The GI tract phageome is characterized by a lower species richness and a very uneven phage distribution, with a minority of phage types comprising a majority of the total phage particles. Without perturbation, the gut phageome can remain surprisingly consistent within an individual over the course of years. When comparing the phage community among different individuals, the gut phage community is specific to each individual. In contrast, the oral cavity phageome shows more temporal variation within an individual.

Despite being characterized by an individualized phageome, some GI tract phages are globally distributed (Fig. 2). For example, the gut crAss-like phage is distributed broadly across the human population. CrAss phage has been found in most individuals, and can make up nearly 90% of the community in some people. How and when these phages are acquired is an open question. Through environmental exposures such as common interactions and cohabitation, individuals are constantly exposed to phages from other people and the environment. Most of these viruses do not colonize at detectable levels but a small proportion can establish residency. Consequently, gut and oral phage communities tend to be more similar between individuals within a same household and between relatives (e.g., siblings, mother and offspring). A subset of phages is commonly found in healthy people and less so in diseased individuals (e.g., patients with irritable bowel disease-IBD), which contributes to the notion of phages being agents of health in addition to their known roles in acute disease.

The dsDNA phages of the Caudovirales order and ssDNA Microviridae phages are the most commonly observed phage types in the gut. Both viral metagenomics and direct observation by electron microscopy indicate the dominance of these two taxonomic groups. However, it is important to note that over 60% of the genomic sequences obtained from human gut phages do not share sequence homology with known viruses and therefore cannot be classified. Of the classifiable phages, Microviridae generally

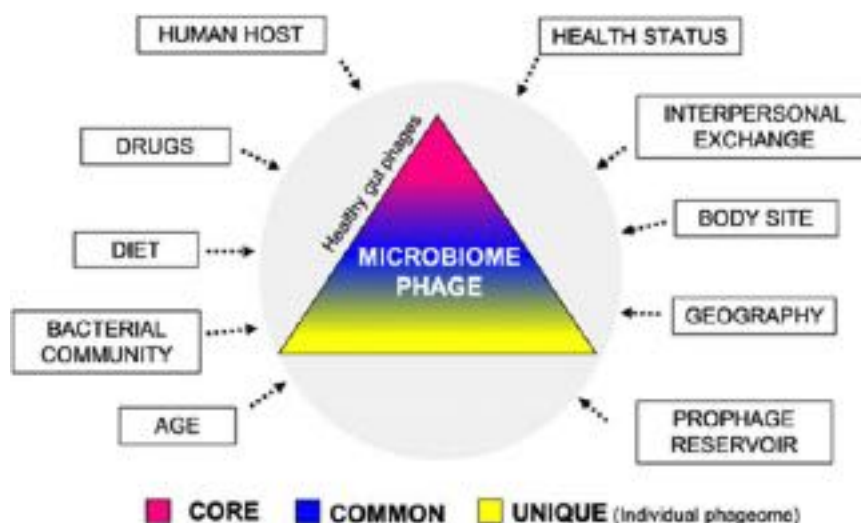


Fig. 2 Factors affecting composition and global distribution of GI tract bacteriophage. The individual composition of the gut phage community is impacted by a complex array of factors. Overall, healthy individuals harbor a mixture of individual-specific phages and a smaller subset of phages globally distributed and commonly found in healthy individuals.



Fig. 3 The adult human gut phage community in health and disease. Human gut phages found in stool samples are a mixture of lytic phages and activated prophages from the prophage reservoir. During certain diseases, the phage community composition is changed. An increase in prophage activation, activation of a different subset of prophage, or an increase of lytic phage potentially contributes to these changes. A balance between lytic and temperate prophage, and an adequate regulation of the prophage reservoir is important to maintain health. Reprinted from Manrique, P., Dills, M., Young, M.J., 2017. The human gut phage community and its implications for health and disease. *Viruses* 9.

dominate in healthy gut communities. An increased abundance of Caudovirales phages has been associated with certain diseases. For instance, in individuals suffering from IBD, the classifiable Microviridae to Caudovirales ratio is significantly decreased. Other prevalent classified gut phages include the 936 group of lactococcal phages, N4, I3, Mu, Hp1, P2, PhiCD119-like viruses. Most RNA human associated viruses are eukaryotic, with only a limited number of RNA phages detected within human fecal matter. It is important to note that the majority of studies characterizing the phage community are carried out using stool samples due to the difficulty and scarcity of access to intestinal tissue for research. In humans, stool phage content significantly resembles that of the colon, providing a good approximation to study human gut phages. In mice, where access to intestinal tissue is easier, differences in both the phage community composition and dynamics between the colon and the small intestine have been identified. However, information about the fine scale spatial distribution of phages within the human gut ecosystem is scarce and should be addressed when possible.

At least 50% of human-associated bacteriophage particles isolated from stool samples can be characterized as prophages derived from lysogenized bacterial strains. The majority of identified prophage sequences in stool samples are found as integrated prophages in Firmicutes, Bacteroidetes and Proteobacteria bacterial species, which comprise the main bacterial phyla found in the gut. Lambda-like and phi80-like coliphages are found in the majority of healthy individuals. Within the Firmicutes phyla, an in-depth characterization of prophages of *Faecalibacterium prausnitzii*, a species highly associated with human health has been carried out, showing that the majority (>70%) of these genomes contain at least one prophage. Bacteroidales-like temperate phages are highly prevalent in cellular metagenomes of healthy adult individuals, but are nonetheless underrepresented in viral metagenomes from stool samples. Some of these prophages are classified as Microvirus, a genus from the Microviridae viral family. Previously, this viral genus was considered to be strictly lytic. This type of prophage is shared among different individuals and can be used to group healthy individuals based on the types of prophages they harbor (i.e., viral enterotypes). However, most gut isolated Bacteroides phages including crAss-like phages, do not have lysogenic capabilities and are thought to be maintained in the gut through persistent infections. The oral cavity and vagina are also rich in temperate phages, with Firmicutes prophages being the most abundant in the oral environment and Lactobacillus phages in the vagina.

The regulation of prophage activation is likely a key determinant of microbiome 'health' with important consequences for human health (Fig. 3). Approximately 70% of gut bacteria contain at least one prophage in their genome. At any given time, a subset of these prophages are induced, producing viral particles which make up a major part of the active phage community. The totality of microbiome prophages is known collectively as the microbiome prophage reservoir. Using murine models, it has been demonstrated that activated temperate phages account for a large fraction of the viral particles found in fecal samples, creating an active viral pool. It is thought that external changes, such as diet or antibiotic treatment, significantly affect the type of prophage that enter the lytic cycle causing a shift in the active temperate bacteriophage community. A model known as the "community shuffling model" proposes that changes in the bacterial community associated with certain diseases arise partly from changes in the prophage activation profile. For instance, individuals suffering from IBD, in which *F. prausnitzii* abundance is significantly lower compared to healthy individuals, present a greater abundance of active *F. prausnitzii* prophages in their stool sample. This, together with changes in the ratio of lytic and temperate phages observed in certain diseases has led to the hypothesis that a balanced prophage activation is necessary to maintain health. A more extensive characterization of human associated temperate phages across different body sites, time, and health state is needed to test this hypothesis and understand their ecological role in the human microbiome.

Both phage and bacterial communities fluctuate through time and display certain trends correlating to the stages of human development. Gut phage diversity at birth is extremely low, but increases significantly within days after birth. The composition of infant gut phageomes is influenced by the type of birth (natural vs cesarean) and the newborn food intake (breast fed vs formula), as most of the prophages are transferred from the mother to the child inside bacterial lysogens. The most abundant phages in newborns are those infecting *Lactobacillus* and *Bifidobacterium* sp. As the infant ages, prophages become active and predate on their host species contributing to the microbial dynamics that take place within the gut of the newborn. Soon after birth, several factors such as time, diet and phage predation alter the microbial community composition and structure. A shift to Bacteroides and

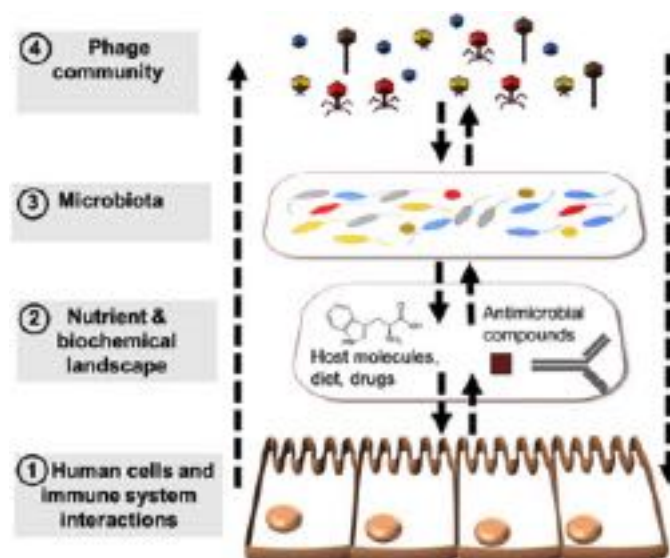


Fig. 4 Phage microbiome trophic interactions.

Firmicutes phages occurs corresponding with a rise of *Bacteroides* and Firmicutes hosts within the bacterial community. The proportion of lytic to temperate phages in the gut is altered as well, with lytic phages predominating in early childhood. As both the bacterial and phage communities gain complexity, this balance switches to favor lysogenic phages and their lysogenized hosts. Much of the phageome temporal stability seen in adulthood is hypothesized to be the manifestation of a predominantly temperate phage community, as integrated or episomal prophages (the prophage reservoir) are able to persist in the gut by sheltering within select bacteria.

Diet and drugs can exert strong selective forces on the gut phageome (Fig. 4). By modulating the nutrient landscape, dietary elements can promote certain bacterial assembly patterns which in turn alter the composition of the phage community. Antagonistic compounds derived from the diet or consumed as medication can also influence gut phage dynamics through depletion of bacterial hosts and activation of prophages. Changes due to dietary intervention are generally reversed upon return to a conventional diet, however, in some cases, the phage community fails to return to its prior state. During and after antibiotic treatment, bacterial diversity is significantly decreased, whereas viral richness is not significantly affected, in all likelihood due to an increase in antibiotic-induced activated prophages. However, antibiotics do cause a shift in the overall phage community membership. It is hypothesized that this apparent greater resilience in the phage community potentially contributes to the ability of the bacterial community to return to equilibrium after disruptions, including disease.

Both acute and systemic diseases often correlate with bacterial and phage dysbiosis in the microbiome (Fig. 3). General changes in phage number, phage diversity and the prevalence of certain phage types can be indicative of unhealthy states. For example, phage taxonomies commonly associated with healthy individuals are underrepresented in patients with IBD. IBD patients with different disease type (colitis vs Crohn's disease) harbor different phage assemblies, generally characterized by an increase in classifiable Caudovirales phages, an increase in strictly lytic phages, a decrease in Clostridiales phages, and an increase in phages associated with pathobiont hosts. The number of detectable phage particles can also change. Mucosa-associated phages are more abundant in individuals with IBD and leukemic disease patients. These features may be attributed to changes in prophage activation profile (community shuffling model) or an increase in lytic bacteriophages. Consistent with this model, a shift from a temperate lactobacillus phage-enriched environment to a more lytic-enriched one in the vaginal microbiome has been associated with bacterial vaginosis. Phage communities are significantly altered in patients suffering from a variety of diseases, ranging from periodontal diseases to bacterial vaginosis, cystic fibrosis, type 2 diabetes, IBD and malnutrition. These changes result in differences in phage diversity; impacting the bacterial community with potential consequences for human health. Phages can also act as vectors of pathogenicity, like cholera toxin CTX ϕ phage and shiga toxin encoding Stx phages. Though many are transient to humans and exist within their host bacteria causing acute disease, a spectrum of these phages can infiltrate the human microbiome. At the cost of increased predation, many pathobionts are highly permissive to temperate phages which can provide advantageous genetic factors. These organisms and their associated phages can exist in a commensal or infectious capacity which dictates their broader distribution across the human population as members of the microbiome.

Dynamics and Implications of the Human Phage Community

Phages engage in a full spectrum of symbiotic interactions with bacteria, from mutualism to parasitism, each contributing to significantly different outcomes in their bacterial host population. Virus predation drives the evolution of both phage and host, likely

through a constant arms-race directional selection, in which selection and propagation of bacterial mutants resistant to phage predation is followed by the emergence of virus mutants that are able to infect the newly evolved strains. Lysogeny offers an alternative host-virus association. Upon integration of a phage genome into the bacterial chromosome, an altered symbiosis is created. The prophage fitness burden can be compensated through prophage-encoded advantages such as protection from superinfection, lysis of competing bacteria, or utilization of advantageous prophage-encoded genes such as antibiotic resistance genes or polymorphic toxins. In complex microbial environments, like the oceans, phages have been shown to have predominantly lysogenic cycles under unfavorable conditions (e.g., low nutrient availability, low host densities). However, increasing host densities (e.g., the gut ecosystem) might also favor the strategy of lysogeny as proposed in the Piggy Back the Winner model (PbW). Some phages may remain in a state between lysogeny and lytic replication known as pseudolysogeny, in which the viral genome remains as an extrachromosomal unit, without being degraded, until it is either actively replicated or integrated into the host genome.

Host-phage dynamics have been primarily explored in the intestinal tract and the oral cavity. Within these microbiomes, the human developmental stage and specific tissue site (e.g., saliva compared to sub- and supragingival areas in the oral cavity or small intestine compared to cecum and colon in the GI tract) are some of the key factors dictating host-phage dynamics. In the gut microbiome of newborns, rapid colonization with the mother's microbiome leads to density dependent (Lotka–Volterra) host-phage dynamics that fade with time as the community reaches an equilibrium and becomes enriched with lysogenic phages. In adulthood, host-phage dynamics change considerably within different sections of the GI tract. For instance, lytic dynamics are more common in the oral cavity than in the gut. Analysis of phage-host abundance and genome evolution in murine models has shown that within the intestinal tract there is reduced directional selection, evident by a decrease in the appearance rate of phage-resistant bacterial mutants. This may be due to epigenetic variability of bacterial host populations in response to different chemical micro-environments within the gut allowing the bacteria to evade phage predation without necessarily modifying their genomes. These observations are our first indications of the mechanisms driving the balance between lysis and lysogeny, and the prevalence and role of pseudolysogeny and their contributions to host-phage dynamics.

The effects of bacteriophages in the gut are beginning to be unraveled primarily using a combination of animal models and NGS technology. Animal models in which germ-free mice are colonized with two bacterial strains— one containing a prophage and a second strain susceptible to the prophage, or with a bacterial consortium representative of the human microbiome, have been used to elucidate the role of GI phages. Within this context, it has been shown that prophage induction rate is significantly higher in the gut than rates observed *in vitro*. The ‘cost’ of prophage induction (lysis) in a subset of the bacterial population encoding the prophage, may be compensated by the ability of the temperate phage particle (activated prophage) to kill competitor strains. It is also possible to establish a more complex microbial community that is more representative of the gut ecosystem by colonizing either germ free mice or *in vitro* chemostat-based bioreactors with human stool samples. This, together with conventional mouse models, has shown that the phage fraction influences the bacterial community structure and that it likely contributes to its resilience after disruptive events such as antibiotic treatment. Recently, more specialized models such as ‘gut on a chip’ are becoming important tools to investigate how the human epithelium and mucosal surfaces can impact bacterial-phage interactions with important consequences to the human host.

The GI tract provides extensive mucosal surfaces onto which some bacteriophages can adhere in high localized densities, altering site-specific phage-bacteria ratios (Fig. 5). In the gut, phage adherence to mucosal surfaces has been described by the bacteriophage adherence to mucus (BAM) model. Briefly outlined, the BAM model proposes that the binding of certain bacteriophages to mucus glycoproteins through immunoglobulin-like domains found on their capsids alters their diffusive properties, increasing the density of phages in outer mucosal layers where they are more likely to encounter susceptible hosts. This mechanism not only allows for the increased success of the adherent phages, but it also confers potential immunity to bacterial invasion of the gut epithelium. The BAM model may operate in other mucosal surfaces of the body as well, such as gums, where bacteriophages can be found bound to the oral mucous membranes in extremely high concentrations compared to their bacterial counterparts (35:1 ratio).

A model that integrates both BAM-based immunity and lysogeny-driven PbW dynamics proposes that bacteriophages can provide a level of protection against bacterial invasion of the epithelium (Fig. 5). In this model, lysogeny, which is prevalent in the lumen and at the surface of the mucosal layer, is thought to confer competitive advantages to commensal bacteria in the highly colonized areas of the gut through symbiotic interaction between the lysogen and the prophage. Simultaneously, lytic phage predation, enhanced in the dense mucus layers, is likely involved in eliminating potential pathogens. Similar dynamics are seen in the oral cavity. Regardless of whether phage infection is of lytic or lysogenic nature, phages can alter the metabolism of their host cell in order to promote their successful replication, influencing the overall functional profile of their microbial communities.

To better understand phage-bacteria dynamics, it is important to determine the host range of the phageome. Bioinformatic analysis of metagenomic datasets suggests that some members of the human phageome may have a broader host range than previously appreciated. Through mechanisms that increase phage genomic diversity, such as single nucleotide mutations and rearrangements of tail fiber genes, gut microbiome phages can alter their host specificity allowing persistence through time. These results have been validated with mouse models and in culture.

At the community level, phages likely stabilize human microbiomes and provide a degree of resistance to external changes and colonization. For instance, the gut microbiota is deeply affected by antibiotic treatment, after which a new, but similar, microbial “steady-state” is restored. Antibiotic pressure leads to an increase of phages that encode antibiotic resistance genes, potentially contributing to the resilience of the bacterial community, which is a hallmark of healthy microbiomes. These observations suggest that phages could contribute to re-establishing a healthy symbiosis altered during changes caused by disease. The bacterial

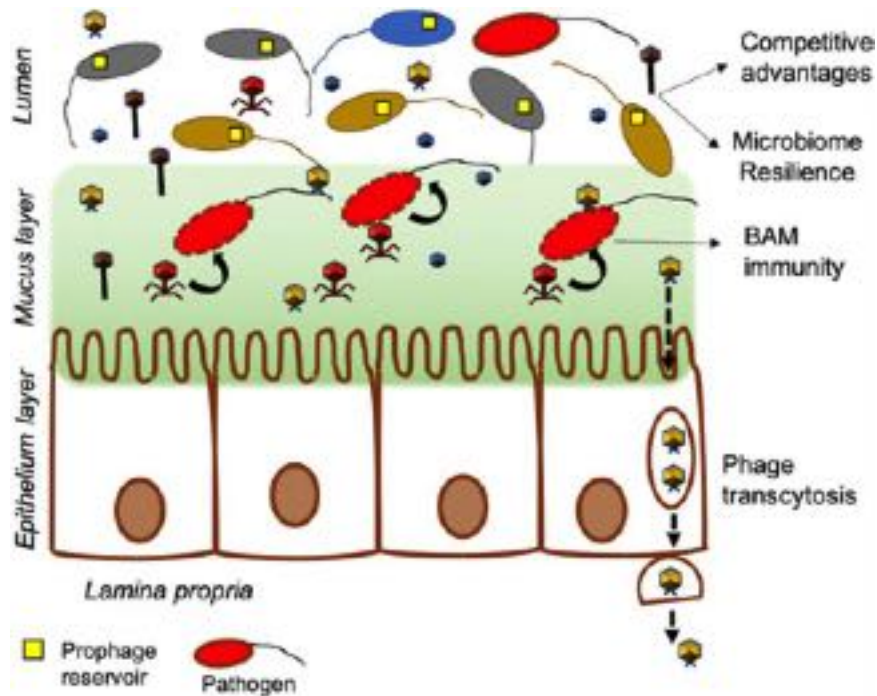


Fig. 5 Phage-bacteria and phage-human interactions within the intestinal tract.

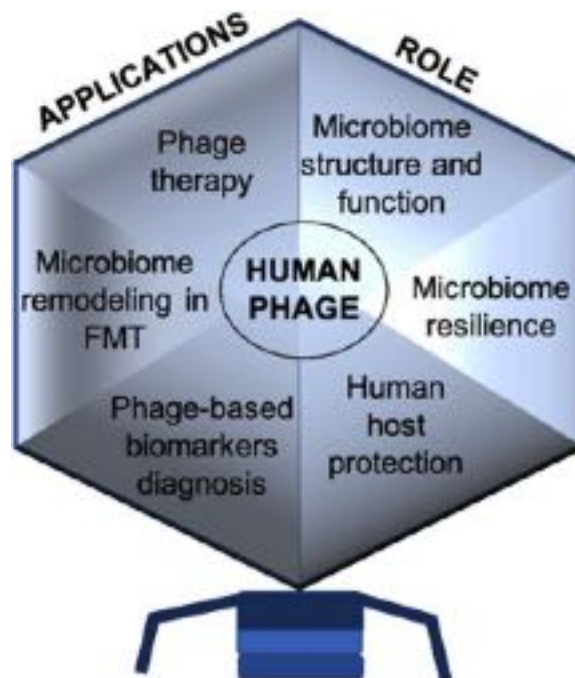


Fig. 6 Role and applications of human associated bacteriophages.

community established after fecal microbial transplant (FMT), is likely influenced by the original phage community. However, it is important to keep in mind that phages may also augment disease development by enhancing dysbiosis in the microbiome.

The use of whole microbial community transplants (e.g., FMT) as an effective treatment of gut-microbiome associated diseases, provides an ideal scenario to study both the influence of bacteriophages on the resilience of the gut microbiome and the capability of bacteriophages to modulate changes in the microbial community structure. Based on the effectiveness of fecal microbial transplantation in treating *C. difficile* infection this practice has been implemented in the treatment of other diseases. The type of

eukaryotic viruses and bacteriophages hosted by a subject have been linked with treatment success, suggesting that both, the viruses that were already in the patient and the viruses which are able to establish after treatment, may affect the outcome of the procedure. Donor-patient compatibility is likely an important factor for phage transfer and treatment outcome.

Clinical Utility of Bacteriophages of the Human Microbiome

Since their discovery in the early 1900s, bacteriophages have been investigated as biotherapeutics to control bacterial infections for diverse applications, ranging from food safety to disease treatment in animals and humans (Fig. 6). It was proposed early on that phage therapy using cocktails of purified phages could eliminate bacterial infections. However, the emergence of commercial antibiotics in the late 1930s and the variability in phage treatment success limited the pursuit of phage therapy in Western medicine. The rise of antibiotic resistant human pathogens has rekindled interest in phage therapy. Phage therapy has been successfully used to eliminate topical (e.g., *Staphylococcus* sp. skin infections), internal or systemic infections (e.g., gut-derived sepsis due to *P. aeruginosa*, multi-drug resistant *E. coli* infection, and *C. difficile* infection), and successfully personalized as intravenous phage therapy for treating *Acinetobacter baumannii* septicemia.

Advancement of phage-based therapeutics will require a better understanding of the ecological forces that shape phage establishment and dynamics within human microbiomes. Elimination of specific pathogens within complex bacterial communities, such as the gut microbiome, is currently a challenge for phage therapy implementation. Important variables included phage host range, host range switching, and the control of the cell lysis process. Ideally, phage therapy could be used to eliminate specific pathogenic bacteria without disruption of the commensal bacterial community, mitigating the potential for secondary infections caused by off target antibiotic effects. If successful, phage therapy may provide a cheap and effective alternative to antibiotics.

From an ecological viewpoint, phages are crucial members of the human microbiota in both health and disease. Therefore, it is reasonable to consider that phages can be used not only to treat specific bacterial infections, but to return a dysbiotic bacterial community to eubiosis. Effective treatment of *C. difficile* infection, an opportunistic pathogen that is associated with a decrease in gut microbiome diversity, can be accomplished by application of fecal filtrates which are greatly enriched in bacteriophages. The use of complex mixtures of bacteriophages isolated from healthy individuals presents a potential therapeutic treatment of more complex human diseases associated with dysbiotic bacterial communities such as IBD, or non-gut related diseases, such as periodontitis or bacterial vaginosis. A long-term goal of phage therapy is to provide safe and effective phage cocktails to control the microbiome bacterial community structure and function by design.

Alongside their role as important biological agents that can modify microbial communities, phages may emerge as useful biomarkers of health and disease. Specific phage sequences may be identified that can separate individuals based on their health status. Phage-based biomarkers have been shown to be useful to determine healthy, at risk individuals and subjects more likely to respond to treatments involving modification of the gut microbial community.

Concluding Remarks

It is well established that human associated microbes are key factors contributing to human health and disease. There is increasing evidence that phages play a direct role in microbiome health and disease. However, our knowledge of bacteriophage communities in human microbiomes remains sparse. Advances in sequencing technologies, phage annotation, animal and in-vitro models, and clinical human trials are significantly contributing to our understanding of phage-bacteria-human interactions. The influence of phage particles inside of the human host also requires further research. Within the next decade, we foresee a significant advance in human phage biology, which will result in applications that improve human health and limit disease.

Further Reading

- Abeles, S.R., Pride, D.T., 2014. Molecular bases and role of viruses in the human microbiome. *Journal of Molecular Biology* 426, 3892–3906.
- De Paepe, M., Leclerc, M., Tinsley, C.R., Petit, M.A., 2014. Bacteriophages: An underestimated role in human and animal health? *Frontiers in Cellular and Infection Microbiology* 4, 39.
- De Sordi, L., Lourenco, M., Debarbieux, L., 2019. The battle within: Interactions of bacteriophages and bacteria in the gastrointestinal tract. *Cell Host & Microbe* 25, 210–218.
- Edlund, A., Santiago-Rodriguez, T.M., Boehm, T.K., Pride, D.T., 2015. Bacteriophage and their potential roles in the human oral cavity. *Journal of Oral Microbiology* 7, 27423.
- Hyman, P., Abedon, S., 2012. Bacteriophages in human health and disease. In: *Advances in Molecular and Cellular Microbiology series 24*. CABI.
- Manrique, P., Dills, M., Young, M.J., 2017. The human gut phage community and its implications for health and disease. *Viruses* 9.
- Maura, D., Debarbieux, L., 2012. On the interactions between virulent bacteriophages and bacteria in the gut. *Bacteriophage* 2, 229–233.
- Mills, S., Shanahan, F., Stanton, C., *et al.*, 2013. Movers and shakers: Influence of bacteriophages in shaping the mammalian gut microbiota. *Gut Microbes* 4, 4–16.
- Nguyen, S., Baker, K., Padman, B.S., *et al.*, 2017. Bacteriophage transcytosis provides a mechanism to cross epithelial cell layers. *MBio* 8.
- Ogilvie, L.A., Jones, B.V., 2015. The human gut virome: A multifaceted majority. *Frontiers in Microbiology* 6, 918.
- Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., Gordon, J.I., 2012. Going viral: Next-generation sequencing applied to phage populations in the human gut. *Nature Reviews: Microbiology* 10, 607–617.
- Shkoporov, A.N., Hill, C., 2019. Bacteriophages of the human gut: The “known unknown” of the microbiome. *Cell Host & Microbe* 25, 195–209.
- Weitz, J.S., 2016. *Quantitative Viral Ecology – Dynamics of Viruses and Their Microbial Hosts*. Princeton University Press.

Bacteriophage: Red Recombination System and the Development of Recombineering Technologies

Lynn C Thomason, Frederick National Laboratory for Cancer Research, Frederick, MD, United States
Kenan C Murphy, University of Massachusetts Medical School, Worcester, MA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

λ Red system The generalized recombination system of bacteriophage λ , which infects *E. coli*. The Red system consists of a 5' $\rightarrow$ 3' dsDNA exonuclease, Exo, and a ssDNA annealing protein, Beta.

Bacteriophage A virus that infects bacteria.

***Escherichia coli* (*E. coli*)** A well-studied genetically tractable Gram-negative bacterium.

Exonuclease An enzyme that digests linear DNA from the ends. Exonucleases are classified according to their degradation properties.

Genetic engineering *In vitro* or *in vivo* manipulation of DNA segments to form new sequences.

Heteroduplex DNA A double-stranded DNA with the two strands derived from different molecules.

Homeologous DNA DNAs with partially but not completely identical nucleotide sequences.

Homologous DNA DNAs with identical or nearly identical nucleotide sequences.

Methyl-directed mismatch repair (MMR) A method that bacteria use for detecting and repairing mistakes created during synthesis of new DNA.

Oligonucleotide For the purposes of this article, a commercially synthesized piece of single-stranded DNA, generally ~ 70 nt long.

Prophage A bacteriophage chromosome integrated into the chromosome of its bacterial host, with most of the viral functions repressed by a phage-encoded repressor protein. An intact prophage can excise from the bacterial chromosome in response to an environmental stimulus and undergo lytic growth to make new viruses.

Rac prophage A defective *E. coli* prophage that contains the RecET system. Defective prophages are unable to excise and undergo a lytic cycle, but they still encode functional genes.

RecET The generalized recombination system of the Rec prophage. The RecET system consists of a 5' $\rightarrow$ 3' dsDNA exonuclease, RecE, and a single-stranded DNA annealing protein, RecT.

Recombinase For the purposes of this article, an enzyme that promotes pairing of two complementary single-strand DNAs.

Recombination The rearrangement of DNA sequence in a new order.

Recombineering *In vivo* genetic engineering with bacteriophage recombination functions.

Selection/countersélection A two-step procedure of genetic manipulation using two adjacent gene cassettes, one of which can be selected for, and the other selected against.

Single-strand annealing Pairing of two complementary single-strand DNAs.

Introduction

Over the past two decades, a phage-based genetic engineering technology in bacteria has emerged that allows any stretch of genomic DNA in a bacterial chromosome, or endogenous plasmid, to be easily altered in a precise manner. Many different genetic changes can be made, including gene knockouts, replacements, insertions, deletions, and point mutations. This technology, termed recombineering (for *in vivo* recombination-mediated genetic engineering) relies on the expression of bacteriophage-encoded homologous recombination functions in the absence of all the other phage functions. This article will first review studies of the phage λ Red recombination system and how it operates during a phage infection, followed by details of how two phage homologous recombination systems, Red from phage λ and RecET from the cryptic Rac prophage, have been utilized to create new genetic engineering tools. What makes the phage systems so practical for *in vivo* engineering is that short DNA homologies (~ 50 bp) are adequate for targeting, such that recombination substrates can be generated using PCRs of drug-resistant markers, with targeting homologies of these substrates incorporated into the PCR primers. Short DNA oligonucleotides can also serve as substrates for recombineering. Thus, allelic exchange substrates are easily generated for gene targeting of bacterial chromosomes, as well as eukaryotic DNA cloned into cosmids and bacterial artificial chromosomes (BACs). This technology has impacted an array of diverse fields such as the study of bacterial pathogenesis, metabolic engineering, synthetic biology, and mouse knockout technologies.

Bacteriophage

Since Twort and d'Herelle first discovered the existence of bacterial viruses in the early 1900s, these organisms have been under intensive study. Bacterial viruses are commonly known as bacteriophages (bacteria eaters) or phage. By the 1930s, they were employed as therapeutic interventions for diseases such as cholera and dysentery, though the details of the reported successes of

these treatments were not properly documented. The advent of antibiotics in the 1940s during WWII, together with the lack of interest by medical professionals in the West, led to only limited use of phage therapy as a modern way to combat bacterial infections. Phage therapy was mostly studied and developed at the Eliava Institute in the Republic of Georgia from the 1930s onward, where to this day, phage formulations are still used to treat bacterial infections of the skin and gut, and other infections that have proved refractory to antibiotic treatments.

By the 1940s, however, bacteriophages moved to center stage in a different venue, as a major tool that led to the development of modern-day molecular biology. The physicist Max Delbrück founded the phage group with a goal of characterizing these invisible creatures to better understand the basis of life. As the new field of molecular biology grew, the phages and their bacterial hosts served as primary model organisms. Early concepts of molecular biology were established by members of this group using bacteriophage and their hosts, which included the discovery of genetic mutations in bacteria, the structure of DNA and how it replicated, the triplet nature of the genetic code, and the capacity of DNA to repair itself, including mechanisms of general DNA homologous recombination.

The phage group studied various bacteriophages, primarily the T-even lytic phages. In 1953, Esther Lederberg reported the isolation of a new phage, λ , as a prophage resident of *Escherichia coli* that could be induced from the host cell to propagate lytically. λ is thus a temperate phage and maintains two life-styles: it is able to integrate its 48.5 kb double-stranded DNA (dsDNA) chromosome into that of its *E. coli* host, allowing passive replication of the phage DNA along with that of the bacterium. In this prophage state, the viral lytic functions are repressed until the virus is triggered to undergo lytic growth and make more phage particles, a process that results in lysis of the host cell by phage encoded enzymes. During the lytic growth cycle phage chromosomes are both replicating and recombining with each other, in a process to generate long concatemeric DNA species, which are the precursors for packaging phage chromosomes into newly formed phage heads. It was this system that led early investigators to use phage λ as a tool to study DNA homologous recombination, as one could easily infect bacteria with two mutant forms of the phage, allow them to recombine with one another during a single infection cycle, then determine the numbers of wild type progeny that were generated following lysis of the bacterial cell. It offered hope to determine, in a very simple system, the mechanisms of initiation and resolution of DNA homologous recombination intermediates.

Homologous Recombination Systems in *E. coli* and λ Phage

DNA homologous recombination is an important universal process in all forms of life. In simple terms, it is genetic exchange between identical or near identical DNA sequences. Its existence is required for lateral transfer of genetic material between bacteria, for the genetic variation generated during sexual reproduction, and for the repair of dsDNA breaks that occur either spontaneously, or because of the presence of DNA damaging agents. In essence, homologous recombination is the guardian of the genome, where mutations in genes encoding recombination functions can lead to the development of cancer (e.g., BRCA1) or to instability of chromosomal structures and numbers. One of the first species to be used in the laboratory to define the genetic requirements of homologous recombination, and subsequently its molecular mechanisms, was *E. coli*, and its infectious particle, phage λ .

E. coli has a homologous recombination system that is important for the incorporation of foreign DNA into the chromosome and for repair of damaged DNA. In 1962, A. John Clark and Ann Margulies isolated mutants of *E. coli* defective for recombination. These mutations were in a gene that became known as *recA*. The RecA protein plays a central role in almost all pathways of recombination in *E. coli*, by binding to single-stranded DNA (ssDNA) and promoting invasion of the RecA-ssDNA filaments into homologous dsDNA, the initiating event in almost all types of both prokaryotic and eukaryotic homologous recombination systems. Soon thereafter, other *E. coli* genes were discovered that also played roles in homologous recombination. These included the genes *recB*, *recC* and *recD*, whose products form a trimeric protein, RecBCD, which binds to the ends of dsDNA and digests both the 5' and 3' strands in a highly processive manner. RecBCD, also known as ExoV, is an *E. coli* dsDNA exonuclease complex capable of degrading linear intracellular dsDNA at a rate of 1000 bp/s. As this powerful exonuclease translocates along, it nicks both strands of dsDNA until encountering an 8 bp sequence called Chi (crossover hotspot initiator – 5' GCTGGTGG 3'). Recognition of the Chi site alters the enzymatic properties of RecBCD such that degradation of the 3'-ended DNA strand is suppressed. Further translocation of RecBCD past the Chi site allows continued production of a long ssDNA with a 3' end, which serves as a substrate for the RecA protein.

As noted above, it was recognized that λ phage was able to recombine its DNA in *E. coli*. The question arose if λ used the host RecA-RecBCD functions to recombine, or whether it possessed its own recombination system. An answer to this question came in 1967 when Brooks and Clark found that if the *E. coli recA* gene was inactivated, λ could still promote recombinant formation during phage infection. Thus, it was clear that the phage possessed a generalized recombination system that could work independently of the host RecA protein. Soon thereafter, mutants in λ were identified that disrupted this new phage recombination system. This system was called λ Red, named after the recombination-defective phenotype displayed by these mutants. It was shown genetically that these λ mutants formed two complementation groups, α and β . Biochemical studies later determined that the α gene encoded a dsDNA 5'-3' exonuclease (called Red α or λ Exo), and the β gene encoded a protein that stimulated annealing between two homologous ssDNA molecules (called Red β or λ Beta protein). A third λ gene, γ , affects recombination indirectly: the encoded protein (λ Gam) was found to inhibit RecBCD by binding to the dsDNA binding site of the enzyme, thus inhibiting RecBCD from binding to dsDNA ends. The Gam protein prevents RecBCD from interfering with λ Exo-Beta complex processing of dsDNA ends, thus indirectly stimulates Red-stimulated recombination. It is noteworthy that the λ phage recombination system is assisted by the Gam protein, which directly inhibits a major component of the *E. coli* recombination system (RecBCD). This turns

out to be a requirement, as the λ chromosome contains no Chi sites, and thus cannot efficiently utilize the host recombination system. Since it lacks Chi sites, the λ chromosome is susceptible to complete digestion by the dsDNA exonuclease activity of RecBCD unless the Gam protein is present.

Over a number of years, work from various laboratories contributed to the genetic and biochemical characterization of the λ Red system. Scientists in F.W. Stahl's laboratory at the University of Oregon, in the C. Radding lab at Yale Medical School, and in the laboratories of T. Poteete and K. Murphy at the University of Massachusetts Medical School were instrumental in defining the *in vivo* activities and biochemical properties of the Red system, as well as the phage P22 recombination system, and distinguishing them from contributions from the hosts. Two key concepts emerged from these studies: (1) In the absence of RecA, the λ Red system required DNA replication of the phage chromosomes to promote high levels of recombination. Many of these studies came from replication-blocked phage crosses, where λ Red recombination was highly dependent on RecA. This observation highlighted the direct link between the mechanisms of DNA replication and recombination for the Red system, a mechanistic coupling unique to the phage that does not occur in its *E. coli* host. (2) A key biochemical finding was that the *E. coli* RecA protein and λ Beta protein promoted different types of reactions with ssDNA. While a RecA-coated ssDNA filament was capable of finding homologous DNA in the context of the DNA duplex (*i.e.*, strand invasion), λ Beta protein bound to and accelerated the annealing of ssDNA to a complementary ssDNA species. The ssDNA substrate, in the case of Beta, was provided *in vivo* by the action of λ Exo when it initiated DNA degradation of a linear phage chromosome.

Biological Roles of the λ Red System

Role in DNA Replication

Enquist and Skalka observed that Red-mediated recombination is required for optimal levels of λ DNA replication. Both the rate of replication and the total amount of λ DNA generated are lower in *red* mutants. This defect is not rescued by RecA function, suggesting that at least partially non-overlapping functions exist for the phage and host generalized recombination systems. The DNA rolling-circle replication forms that predominate during the late stage of the phage replication cycle are likely to be substrates for Red recombination, since they contain dsDNA ends and active replication forks. It was later noted that rolling-circle tails are shorter in a *red* mutant background, suggesting that two replicating linear molecules may be recombined into a longer one by the Red system. The products of rolling-circle replication intermediates are multimers of λ chromosomes, termed concatemers. These concatemers are the DNA substrates utilized by the phage terminase enzyme for packaging individual chromosomes into protein capsids. Red is important in generating larger concatemers, thus making more packageable λ chromosomes. However, long concatemers can also persist in the population in the absence of Red, as long as the Gam protein is present. In this case, Gam, by binding to RecBCD, protects the dsDNA ends of rolling-circles, thus allowing these replication intermediates to form concatemeric DNA. While recombination is not absolutely required for the formation of packageable DNA, it stimulates this process.

Genetic Exchange Among Similar Phages – Generating Genetic Diversity

The lambdoid phages as a group share similar genomic organizations. The recombination, replication, and morphogenesis functions all share similar positions along the length of the chromosome when lambdoid phage genetic maps are compared. Although individual protein factors may differ structurally, and functions from one phage may fail to complement mutations in analogous genes of other lambdoid phages, complete functional modules, which may include several genes, are often readily exchanged between phages. One example of this modularity is the clustering of recombination functions from phage λ and P22 (a *Salmonella* phage). The phage P22 annealase called Erf (essential recombination function) cannot substitute for λ Beta in *E. coli*, and the P22 anti-RecBCD function modifies RecBCD's enzymatic activity and actually increases its affinity for dsDNA ends, while the λ Gam protein inhibits RecBCD from binding to dsDNA ends. Nonetheless, the λ Red system (three genes) and the P22 system (four genes) can efficiently substitute for one another in *E. coli* and *Salmonella* hosts.

Recombination reactions in these functional segments occur in short regions of homology between the modules, leading to sequence mosaicism in the lambdoid family of bacteriophages. λ Red-like systems are capable of recombining with short homologies very efficiently and thus would be expected to easily promote recombination among these modules during mixed phage infections. Furthermore, the Red system is also able to act on homeologous DNAs, recombining DNA that is only partially homologous (as much as 22% divergent). This ability gives lambdoid phages even greater capabilities to promote recombination within small regions of imperfect homology between functional modules. Such a relaxed requirement for DNA shuffling would give lambdoid phages an evolutionary advantage over other phages (and bacteria) that require high sequence identity for exchanging DNA segments.

DNA Repair

When phage DNA enters the cytoplasm of a bacterium in nature, restriction endonucleases often generate double-strand breaks in the viral chromosomal DNA. The dsDNA breaks inflicted by these restriction systems can be repaired by recombination with

another infecting phage. Heterologous prophages already resident in the cell may also participate in genetic recombination with the damaged DNA of the infecting phage, with portions of the broken chromosomes rescued by recombination and incorporated into resident prophages. It was, in fact, the *in vivo* delivery of a dsDNA break by a restriction enzyme during experimental phage λ crosses in the Stahl laboratory that demonstrated the importance of dsDNA ends for the action of λ Red *in vivo*.

DNA damage is the natural inducing agent for excision of lambdoid prophages from the bacterial chromosome, as a result of RecA co-protease-dependent *cI* repressor cleavage and subsequent expression of the phage lytic genes, including the recombination functions. A low level of DNA damage may allow transient expression of phage functions without inducing a full-fledged lytic cycle. Activation of the Red system through this transient prophage induction increases total cellular recombination potential, which may benefit the bacterial host by providing an additional DNA repair pathway. Transient prophage induction would also increase the likelihood that double-strand breaks in the phage DNA will be rescued by other prophages of the same type or by heterologous prophages with limited homology. Under some circumstances a partially induced prophage may undergo DNA replication while still integrated in the *E. coli* chromosome, giving rise to onion-skin replication, a replication bubble with multiple replication forks. Exo and Beta may act to resolve these kinds of aberrant structures.

Classical Models of λ Recombination

After nearly three decades of research involving the Red system, two general mechanisms of λ homologous recombination have been defined. One mechanism is the RecA-dependent pathway of Red recombination, which occurs when phage chromosomes are prevented from replicating. In delineating this pathway, crosses were performed with λ mutants defective for genes responsible for the initiation of chromosomal replication. As noted above, the loss of RecA drastically reduced recombination in such crosses. By this mechanism, it is thought that the Red proteins Exo and Beta act on dsDNA ends to generate ssDNA, where the bound Beta protein is replaced by RecA, with assistance of host *recFOR* proteins (functions normally associated with loading RecA onto ssDNA coated with *E. coli*'s single-stranded binding protein). This intermediate is then shuttled into the host pathway of homologous recombination, involving both branch migration proteins and Holliday junction resolvases of the host (though the phage resolvase Rap could also play a role).

The second major mechanism of Red recombination is the RecA-independent ssDNA annealing pathway. This route is best described in the context of the rolling-circle mode of chromosomal replication (also called the sigma, or σ , mode), as mentioned above. Such rolling-circle intermediates are thought to be initiated by random nicks in a λ chromosome replicating in the circle-to-circle mode (also called the theta, or θ , mode), which consequently generates dsDNA ends that are distributed across all regions of the λ chromosome. Thus, in this scenario, the ends of two rolling-circle intermediates are acted upon by λ Exo and Beta to produce ssDNA coated with Beta protein. Complementary regions of this ssDNA are then annealed together by the action of λ Beta to form a larger heteroduplex recombinant DNA intermediate. After DNA polymerase I fills the remaining ssDNA gaps, the nicks are sealed by DNA ligase. Thus, the recombination event promotes both genetic diversity in a heterogenic community, as well as formation of large concatemeric DNAs refractory to exonuclease degradation for more efficient packaging.

Development of Recombineering

By the late 1990s, it was recognized that the λ Red system might be useful to promote genetic engineering of the *E. coli* chromosome, by virtue of its ability to promote recombination between linear dsDNA substrates and bacterial chromosomes. It was theorized that the Red and Gam functions, if expressed off a plasmid independent of any other phage functions, might create a hyper-recombinogenic environment in *E. coli*, and perhaps other hosts as well, including pathogenic bacteria. Experiments performed by one of us (KM) showed that this was indeed the case. Plasmid-borne expression of Red and Gam in an *E. coli* strain resulted in a strain that could be transformed with linear DNA substrates at very high efficiency. Furthermore, the hyper-rec background of these strains facilitated the insertion of a PCR-amplified linear dsDNA antibiotic resistance gene, flanked on each side by 1 kb homologies to the target gene, at efficiencies up to 100-fold higher when compared to other recombination-proficient strains used at that time. Around the same time, the Francis Stewart laboratory was working with another phage homologous recombination system, the RecET system of the cryptic Rac prophage. This system, which is analogous to the λ Red system, possesses a 5'-3' dsDNA exonuclease (RecE) and a single-strand annealing protein (RecT). Although λ Exo and RecE have similar activities, the full length RecE protein is much larger (nearly 900 amino acids), with the exonuclease activity in the C-terminal domain. The Stewart lab demonstrated that RecET could also promote recombination between bacterial genomic DNA and PCR substrates, but this time with target homologies of about 50 bp. This was a key step in the development of recombineering technologies. This relaxed requirement for the size of flanking homologies allows the preparation of dsDNA substrates to be performed by PCR, where a drug marker serves as a template and flanking homologies are simply added to the 5' ends of the PCR primers. Plasmid constructs containing drug markers with long flanking homologies, typical of the early days of gene replacement in bacteria, were no longer required.

This novel gene replacement technology was quickly followed by similar reports from a number of laboratories. It was found that short flanking homologies (30–50 nt) on the linear dsDNA were adequate for the recombination for both the λ Red and RecET systems, though the former system showed mildly better efficiencies. Both systems required λ Gam for efficient gene replacement.

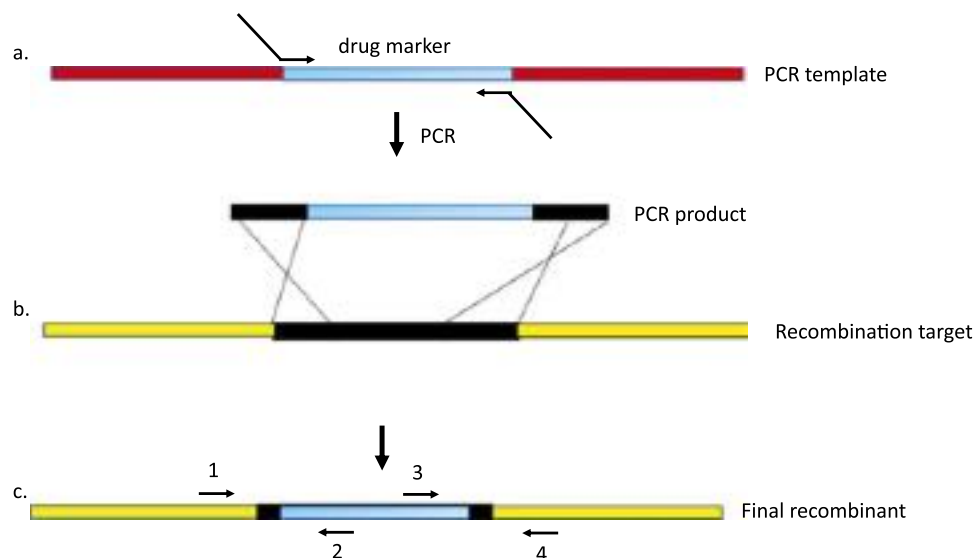


Fig. 1 Linear dsDNA recombineering with PCR-generated substrate. (a) Hybrid primers are designed that are ~70 nt long and contain ~50 nt of 5' homology that flank the intended target sequence. The remaining ~20 nt at the 3' ends prime the DNA template, in this case an antibiotic resistance gene (drug^R marker) indicated in blue. These hybrid primers are used in a PCR reaction to generate a linear dsDNA containing the amplified cassette with homology to the target gene at both ends. (b) The PCR product is used for recombineering (requiring both λ Beta and Exo) and will replace most of the endogenous DNA, indicated in black, between the two target homology sequences. (c) The final recombinant can be verified using two pairs of PCR primers, 1 + 2 and 3 + 4. These primer pairs will amplify the recombinant junctions between the target and the newly inserted DNA and confirm correct insertion. If the inserted DNA is a different size than the original endogenous sequence, primer pair 1 + 4 can be used to confirm that the recombinant strain contains only the modified copy of the gene and that the original sequence is no longer present. Primer pair 1 + 4 can also be used to amplify a PCR product for DNA sequencing.

The phage recombination functions plus Gam, either residing in the bacterial chromosome or encoded by a plasmid, are typically expressed from one of a number of controllable promoters. These promoters are normally inactive but can be up-regulated for gene expression by displacement of a bound repressor protein (e.g., LacI or AraC) from the operator by small molecule inducers (e.g., IPTG or arabinose, respectively), or by inactivation of a heat-labile repressor. The regulated expression of these recombineering functions is important, as constitutive high levels of expression result in poor growth and can be mutagenic. Once induced, the phage recombination functions promote a hyper-recombinogenic environment, where linear DNA species recombine with target chromosomal or episomal sequences at very high frequencies. The DNA substrates can be dsDNA that is either generated by PCR or chemically synthesized, or single-stranded DNA oligonucleotides.

To perform recombineering, the cells are grown to moderate density, the recombination functions are induced, the cells are collected by centrifugation and washed, and linear DNA donor substrates are introduced into the bacteria by electroporation. The cells are then suspended in liquid media and grown for a period of time, to allow recombination between the introduced substrate and its target chromosome (or episome), and to promote expression of any antibiotic resistance cassette. The cells are then plated on petri plates under appropriate conditions to recover recombinants; most often an antibiotic selection is applied. Recombineering is efficient, precise, and accurate. The ease and speed of the process has made it the method of choice for genetic engineers to alter bacterial chromosomes, study bacterial pathogenesis, perform metabolic engineering of industrially-important bacteria, or manipulate eukaryotic DNA cloned into plasmids and bacterial artificial chromosomes (BACs).

Recombineering With dsDNA

The most common type of genetic modification constructed with recombineering is a simple gene replacement, where the target gene is replaced by a DNA cassette conferring drug resistance. For this purpose, a PCR is used to generate linear dsDNA where the targeting homologies are incorporated into the 5' ends of the PCR primers (Fig. 1). The PCR products are made using chimeric (hybrid) primers of about 70 nt in length. The 5' ends provide ~50 nt of homology to the target DNA and the 3' ends provide ~20–25 nt of priming sequence to amplify the drug^R cassette. Although the template for PCR can theoretically be a drug^R marker contained on a plasmid, care must be taken to prevent carryover of the supercoiled template plasmid from the PCR reaction, leading to false positive recombinants following electroporation. Treating the PCR with the restriction enzyme *DpnI* can minimize template carryover, as it is active only on methylated DNA and will not digest PCR products. Other appropriate templates for PCRs used in recombineering include drug-resistant markers on conditionally-replicating plasmids, or bacterial colonies or cultures. For the latter, a small portion of a medium-sized colony, or several microliters of an overnight culture, will serve well as a template in a

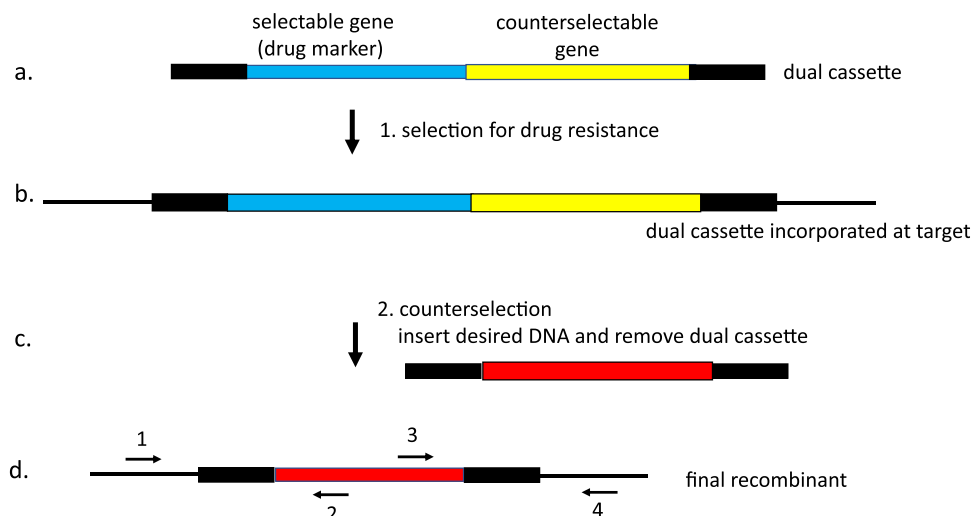


Fig. 2 Two-step selection/counterselection procedure to generate unmarked gene knockouts. (a) A linear dsDNA dual cassette containing two adjacent genes, one encoding drug resistance (indicated in cyan) and the other a counterselectable maker, *i.e.*, *sacB* (indicated in yellow) is generated using PCR with hybrid primers that provide flanking homologies to the desired target, as described in Fig. 1. (b) In the first recombineering step, the linear dsDNA dual cassette is inserted at the desired target, and selection for drug resistance is applied. (c) In the second recombineering step, a linear dsDNA substrate containing the final desired DNA sequence (in red) with flanking homologies is provided. Application of the counterselection allows recovery of recombinants in which the dual cassette has been replaced with the desired DNA. (d) As described in Fig. 3, the final recombinant should be verified with the appropriate PCR reactions and by DNA sequencing.

standard PCR to generate recombineering substrates, provided the PCR program contains an initial (95°C) heating cycle to lyse the cells prior to cycling. To this end, the Court laboratory has engineered a useful bacterial strain called T-SACK, which contains most of the drug-resistant markers typically used for recombineering in *E. coli*. Contained in this strain are the genes *tetA* (Tet^R), *sacB* (sucrose^S), *bla* (Amp^R), *cat* (Cam^R), and *kan* (Kan^R), making this a useful single source for the variety of drug-resistant markers used for recombineering. Use of the T-SACK strain as a PCR template for drug^R dsDNA markers is preferable to using a plasmid template, since the possibility of false positives resulting from contamination by residual supercoiled plasmid template is eliminated. The error-prone Taq polymerase is suitable for this PCR, as any errors generated in amplifying the drug^R cassette will not appear on the antibiotic selection plate.

Markerless Gene Deletions and Counterselection Schemes

Linear dsDNA recombineering can also be used in a two-step selection/counterselection method (Fig. 2), if it is undesirable for an antibiotic resistance gene to remain in the final construct. Several dual cassettes with two genes are commonly used; these cassettes contain a drug marker linked to a second gene whose function can be selected against. The *sacB* gene from *Bacillus subtilis* is frequently used as a counterselection marker, since bacteria containing this gene are unable to grow on solid agar medium containing sucrose. The *sacB* gene has been linked to several different drug markers, to generate selection/counterselection dual cassettes. Other counterselectable genes can also be used, including the wild type *rpsL* gene in a *rpsL31* (strep^R) mutant (where loss of the overexpressed WT *rpsL* gene confers streptomycin resistance) and *tolC* (resistance to colicin E1). In the first recombineering step, the drug resistance encoded by the dual cassette is selected, and the cells are tested to verify that the counterselectable function is expressed. In the second recombineering step, the counterselectable function is selected against, allowing the dual cassette to be replaced with the desired genetic alteration, which may be either a net deletion/small insertion created with a ssDNA oligo, or any desired dsDNA sequence created with a PCR or synthetically designed substrate. This method is useful to insert non-selectable pieces of DNA such as gene tags (*i.e.*, *gfp*). Replacement of the dual cassette is confirmed using PCR analysis and final constructs are verified by DNA sequencing.

Recombineering With ssDNA

The simplest recombineering reaction uses linear ssDNA, usually provided as synthetic DNA oligonucleotides (oligos). For oligo-mediated recombineering, only the expression of λ Beta is required, as the substrate is already single-stranded, *i.e.*, λ Exo is not needed. For oligo-mediated recombineering, ~70 nt ssDNA oligos are optimal in length, with the desired base change(s) in the center of the linear DNA. In wild type *E. coli* cells, oligos yield recombinants at a frequency of about 1 out of 1000 total cells. Thus, a selection scheme is usually required for single base changes, insertions, or deletions. However, the Court lab found that the

frequency of oligo-mediated recombineering can be increased ~100-fold by using strains that are deficient for methyl-directed mismatch repair (MMR), providing further evidence of a mechanistic link between λ Red and host DNA replication machinery. In the absence of MMR, point mutations and changes of a few bases can be made very easily, with 30%–50% of the total viable cells (after outgrowth) containing the oligo-mediated mutation in its chromosome. The model for this mechanism is straightforward. When the ssDNA oligo anneals to the lagging strand template at the replication fork, mismatches present between the oligo and the template strand are recognized by the MMR system, and are corrected in favor of the wild-type DNA sequence. Thus, MMR mutants expressing λ Beta demonstrate higher rates of oligo-mediated recombineering.

However, because cells lacking MMR accumulate mutations, it is preferable to avoid use of such mutants as recombineering hosts. One method to bypass the MMR system is the use of oligos that generate C-C mismatches. Since these mismatches are not recognized by the MMR system in *E. coli*, oligos that generate this particular mismatch generate high frequency mutagenesis. The Court laboratory also found that when modifying genes, incorporation of base changes at several adjacent wobble codons near the desired nucleotide change allows evasion of an active MMR system and can be done without changing the amino acid sequence of a protein of interest. The ssDNA oligo design for this scheme generally results in a high frequency of recombinant formation, even in MMR⁺ bacteria. Other measures to avoid MMR involve the co-expression of dominant negative mutations of Dam methylase or the mismatch recognition function MutS along with λ Beta protein during the induction period. In these circumstances, the inhibition of the MMR system is transient, allowing the oligo to avoid correction for just a short recombinogenic window, leaving MMR system intact for the outgrowth period. When MMR is avoided, high efficiency genetic alterations are possible with oligo recombineering, including single point mutations and small changes of a few base pairs. Large deletions of more than several kb can also be made with ssDNA oligos, but the efficiency of this reaction is much lower and generally requires a selection.

In vivo Cloning

Another useful dsDNA recombineering reaction mediated by the Red system is a recombination reaction that results in a bacterial segment of genomic DNA being incorporated into an electroporated plasmid backbone to form an intact replicating plasmid. It is a method to clone large segments of bacterial chromosomes, or eukaryotic DNA from BAC libraries, without the use of restriction enzymes or PCR, a procedure called *in vivo* cloning. This is especially useful for cloning large sizes of chromosomal segments, where classical methods of restriction enzymes digestion and ligation are not feasible as a result of the lack of unique cutting sites. To perform *in vivo* cloning, a DNA plasmid backbone containing, at a minimum, the origin of DNA replication and a drug^R selection marker, is amplified using chimeric primers. The 5' ends of these primers provide homology to the target DNA to be retrieved, and the 3' ends prime the plasmid DNA sequence (Fig. 3). Cloning by retrieval is a low efficiency reaction but is easily selected for by plating on antibiotic-containing selection plates. The advantage of *in vivo* cloning is that segments of the bacterial genome are cloned without subjecting them to a PCR, which might otherwise introduce polymerase errors. Instead, once retrieved from the chromosome and inserted into the vector backbone, the insert replicates using the high-fidelity system of the *E. coli* replisome (and its proofreading capabilities). Any errors that might persist in the wake of the replisome are caught by the MMR system. The most common false positives occurring with *in vivo* cloning arise when the electroporated plasmid recircularizes without the insert, often as a result of recombination between micro-homologies that exist in the terminal ends of the vector. Such false positives can be minimized by careful design of the homologous sequences to the chromosomal target region, in order to avoid such micro-homologies. Control transformations in the absence of the recombination functions should be done to verify that the plasmid backbone does not circularize at high efficiency without incorporating the desired insertion.

The Stewart laboratory found that the RecET system is superior to λ Red for one type of *in vivo* cloning: the joining of multiple linear dsDNAs containing short terminal homologies to form intact circular plasmids. Recombination occurs at the DNA ends and depends on homology rather than on restriction sites, giving flexibility in plasmid design. For reasons that are not understood, the full length RecE protein is required for this linear DNA assembly to occur at high efficiency.

Mechanism of Oligo-Mediated Recombineering

Oligo-mediated recombineering requires only a single function, a ssDNA annealing protein, Beta or RecT. For any chromosomal target, there are two alternative sequences for a recombinogenic oligo: one targeting either the leading strand or the lagging strand templates. It was observed that for these two complementary substrates, one oligo gave a ~20-fold higher recombination frequency relative to the other. This difference correlates with the direction of DNA replication through the target site, and the higher efficiency oligo is complementary to the lagging strand template. This observation suggested that the recombination is occurring at the DNA replication fork, with oligos annealing to the lagging strand template at single-strand gaps present in the newly synthesized discontinuous lagging strand, as illustrated in Fig. 4. In effect, the oligo becomes a pseudo-Okazaki fragment, with filling-in of the gap behind the 5' end and extension from the 3' end performed by DNA polymerase I, so that the oligo is incorporated into the growing new daughter strand. Subsequent work from several laboratories strengthened the evidence that in

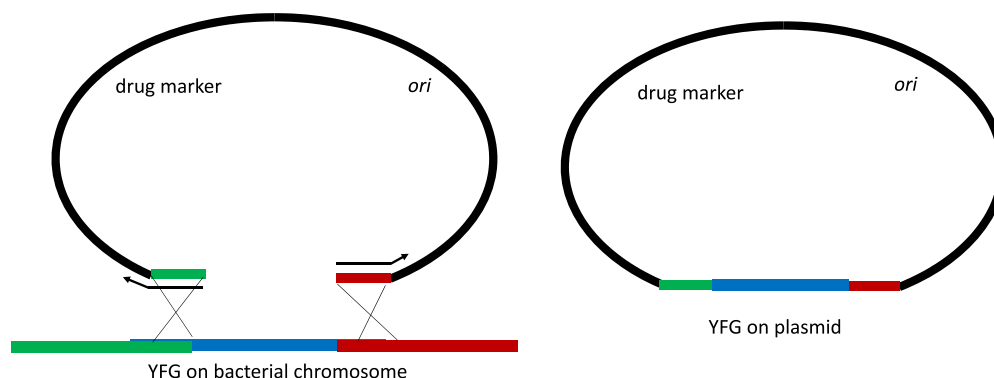


Fig. 3 *In vivo* cloning by retrieval. To retrieve a gene or other DNA sequence from the bacterial chromosome, hybrid primers, shown in black, are used to amplify a plasmid backbone, indicated in gray, containing an origin of DNA replication (*ori*) and usually an antibiotic resistance gene to allow for plasmid selection. The hybrid primers should be designed with the two homology sequences (indicated in green and red) facing inward, so that when the recombination between the plasmid backbone and the bacterial DNA occurs, the desired DNA sequence will be incorporated onto the plasmid backbone. DNA to be retrieved, 'Your Favorite Gene' (YFG, indicated in blue) is located on the bacterial chromosome or BAC. This reaction can also be used to create an intact circular plasmid from two or more PCR products.

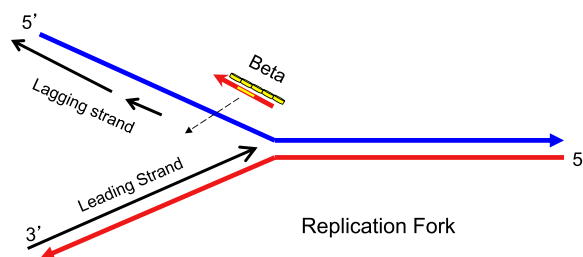


Fig. 4 Model for single-strand DNA annealing at the DNA replication fork promoted by the λ Beta protein. A simplified schematic of a DNA replication fork is presented, with the leading and lagging strand templates in red and blue, respectively, and strand polarities indicated. The newly replicated leading and lagging strands are in black with the direction of DNA synthesis indicated by the arrowheads. Beta (or other single-strand annealing protein) binds a ssDNA oligonucleotide and anneals it to complementary DNA sequence on the lagging strand template in a gap that arises during discontinuous replication of the lagging strand. Beta may also anneal a complementary ssDNA oligo to single-strand DNA on the leading strand template ahead of the newly replicated leading strand, but this recombination is less efficient than a recombination event targeting the lagging strand.

order for recombinants to form, the circular DNA being targeted (either chromosome or plasmid) must be replicating, consistent with the idea that recombination occurs predominately by single-strand annealing, rather than by strand-invasion.

Mechanism of dsDNA Recombineering

Two laboratories have reported that dsDNA recombination likely proceeds through a single-stranded DNA intermediate. In this model, after a dsDNA substrate is introduced into the bacterial cell expressing λ Red, one strand is entirely removed by phage λ Exo, leaving the complementary ssDNA bound by the recombinase, which is then incorporated by single-strand annealing at the cellular target, much like the mechanism described above for oligo-mediated recombineering. The data for this mechanism is suggested by the fact that sequence information in the homologous regions for most recombinants following recombineering is usually derived from one of the two strands in the dsDNA substrate. This mechanism is likely to occur with short dsDNA substrates more often when compared to longer substrates, since λ Exo is a moderately processive (~ 3 kb) enzyme. However, the fact that this is not seen in all recombinants, and less so with long recombinants, suggests that another apparently less-favored mechanism is also at work.

Recombineering in Pathogenic Bacteria

Recombineering technology opened up the toolbox of genetic engineering in *E. coli*. The fact that the λ Red functions could be supplied from an episome suggested early on that the system might be amenable for use in clinically relevant bacteria, simply by transformation of a Red + Gam-producing plasmid into a clinical strain of interest. This proved to be the case, and the plasmid-borne

λ Red system has been used to make gene knockouts in enterohemorrhagic and enteropathogenic *E. coli*, for constructing vaccines in *Shigella* species, and to manipulate the chromosome of *Salmonella enterica*. The successful use of λ Red for genetic modification in these bacteria is not too surprising, given the genetic relatedness between these bacterial species and *E. coli* K12. In more distantly related bacteria such as *Pseudomonas* species, the λ Red system can work, but high efficiency requires longer regions of targeting homologies. Ideally, the use of recombineering technology in non-*E. coli* species is most successful when recombination systems from endogenous phage systems of the bacterium in question are found and characterized. This was the approach taken by the Hatfull lab when investigators there identified a RecET-like system in the mycobacterial phage Che9 and demonstrated its use in both dsDNA and oligo-mediated recombineering in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. Other studies have identified λ Beta or RecT-like annealases in the chromosomes of diverse bacteria and have shown that they are capable of promoting oligo-mediated recombineering in *E. coli* at a wide range of efficiencies. A recent search of λ Beta-like proteins for use in the soil organism *Pseudomonas putida* revealed an annealase activity (the Ssr protein) capable of promoting oligo-mediated recombineering, setting the stage for further development of recombineering in the important strain, used often in metabolic engineering strategies. Similarly, a Beta-like W3 phage recombinase has recently been developed for use in the genus *Shewanella*, a metal-reducing bacterium important in bioremediation.

Combining Recombineering With Other Gene Modification Technologies

The gene-editing tool CRISPR has revolutionized the genetic toolbox for eukaryotic cells. In its simplest manifestation, RNA guides bring the Cas9 nuclease to the target gene to deliver an endonucleolytic double-strand cut at the desired location. Because of the presence of non-homologous end-joining (NHEJ) functions in eukaryotic cells, the cut is often repaired inefficiently by rejoining the dsDNA break after some nucleotide degradation occurs at the cut site, leading to gene inactivation. Bacteria typically lack NHEJ systems, and as such, cuts are either repaired by homologous recombination pathways, or not. In the latter case, the cut is lethal.

It was recognized that CRISPR could be used in bacteria, not for gene inactivation per se, but more importantly to select against non-recombinants following recombineering. The Marraffini lab combined the CRISPR-Cas9 technology of creating targeted *in-vivo* dsDNA breaks with Red-mediated oligo recombineering, using dsDNA cleavage as a counterselection step against unmodified chromosomes in *E. coli* and *Streptococcus pneumoniae*. This methodology drives up the frequency of oligo-mediated recombineering, where half of the total viable cells recovered were recombinants. Since this efficiency is so high, recombinants can be identified by DNA sequencing a few candidates, with no need for selections or screens. These experiments also suggested that Red recombination did not repair the double-strand break created by CRISPR-Cas9, and that Cas9-cleavage was acting solely as a counterselection step against non-recombinants (*i.e.*, parental DNA). Clearly, one requirement in using this combination of technologies is timing; that is, that the phage-mediated linear DNA recombination occurs before Cas9 cutting is activated.

In another manifestation of combining two separate technologies into one genetic modification scheme, oligo-mediated recombineering is used to deliver the P1 phage site-specific attachment site *loxP* to different regions of a bacterial chromosome, separated by 10 kb of DNA or more. Following this step, expression of the site-specific recombinase Cre will promote recombination between the two *loxP* sites, and depending on the relative orientation of these sites, will promote either inversion of the marked segment or its deletion. This has been a common practice in genome reduction strategies, and to identify essential regions of a bacterial genome. In a more recent example of a similar technology in *M. tuberculosis*, Che9 RecT-promoted oligo-mediated recombineering is coupled with the mycobacterial Bxb1 phage site-specific integration system. In this system called ORBIT (Oligo Recombineering followed by Bxb1 Integrase Targeting), the ssDNA oligo contains the Bxb1 phage *attP* site. The *attP*-containing oligo is co-electroporated with a non-replicating plasmid containing Bxb1 *attB*, a drug^R-marker, and a payload (*e.g.*, a *gfp* tag). In this system, the introduction of the *attP* site is followed by integration of the payload plasmid in the same outgrowth period. Thus, gene deletions or fusions can be made by co-electroporating an oligo with a ready-made target-independent plasmid. A major advantage of ORBIT is that no target-specific plasmids or PCR products need to be generated for the construction of gene knockouts or genetic fusions.

Recombinase-Independent Recombination

It has been observed that in some bacteria, electroporation of linear ssDNA substrates gives rise to recombinants that are independent of any known phage recombinase. While not strictly “recombineering”, there are similarities between this type of ssDNA recombination and ssDNA phage-mediated recombineering. B. Swingle and others have found that oligo-mediated recombinants can be found at a frequency of $10^4/10^8$ viable cells when high oligo concentrations are used. The recombination displays a lagging strand bias, and oligos targeting the lagging strand that escape mismatch repair give the highest efficiencies. Recombinase-independent oligo-mediated recombination has so far been demonstrated in *Pseudomonas syringae*, *E. coli*, *S. enterica*, *Shigella flexneri*, *Yersinia pseudotuberculosis*, *Legionella pneumophila* and *Shewanella oneidensis*. Recombinase-independent recombination provides a toe-hold for creating mutations in bacteria that lack recombineering systems, and thus affords a starting point for development of *in vivo* genetic engineering in these organisms. Although the frequencies of this recombinase-independent oligo incorporation are too low to allow direct screening for recombinants, the level of

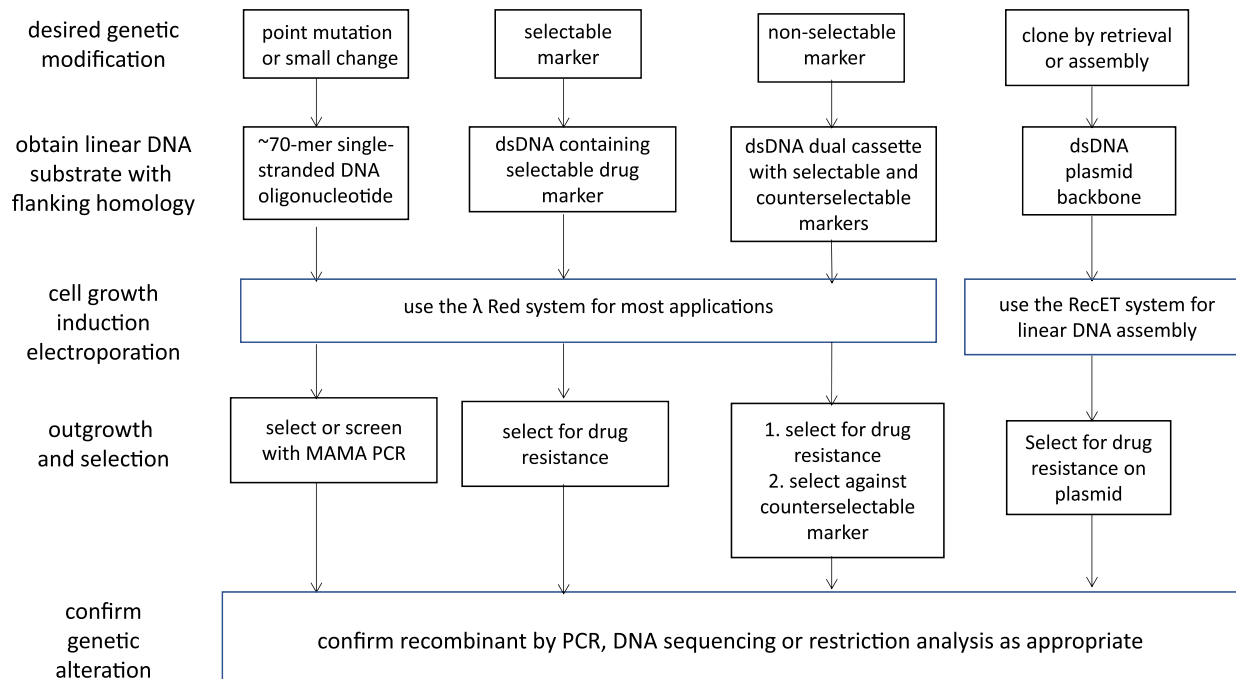


Fig. 5 A flow-chart for recombineering. In all cases the desired genetic constructs should first be designed *in silico*. The appropriate linear DNA substrate and expression system will vary depending on the desired alteration. The outgrowth procedures and methods used to identify the recombinant clone will also differ.

recombination can be dramatically enhanced by using it in combination with CRISPR-Cas9 cutting. The dsDNA break created by Cas9 is used as a counterselection and enriches the recombinant class by eliminating non-recombinants, just as when CRISPR-Cas9 is coupled to canonical recombineering.

Final Thoughts

Over the last ~20 years the development of recombineering technology has revolutionized bacterial genetic engineering. Many papers have been published describing the various possible reactions and helping to elucidate molecular mechanisms involved in this bacteriophage-mediated recombination. A few seminal publications are listed in the following section. For those considering recombineering, a flow chart defining the necessary steps is provided in **Fig. 5**. While this article has described the present state of recombineering, it is a rapidly evolving field, since phages are the most abundant organisms in nature and new isolates are being characterized daily. Many of these newly identified phages will contain generalized recombination functions analogous to the Red and RecET systems; these newly identified functions have the potential to enable gene editing in their specific bacterial hosts.

Further Reading

- Cairns, J., Stent, G.S., Watson, J. (Eds.), 1966. *Phage and the Origins of Molecular Biology*, first ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Costantino, N., Court, D.L., 2003. Enhanced levels of λ Red-mediated recombination in mismatch repair mutants. *Proceedings of the National Academy of Sciences of the United States of America* 100 (26), 15748–15753.
- Court, D.L., Sawitzke, J.A., Thomason, L.C., 2002. Genetic engineering using homologous recombination. *Annual Review of Genetics* 36, 361–388.
- Ellis, H.M., Yu, D., DiTizio, T., Court, D.L., 2001. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-strand oligonucleotides. *Proceedings of the National Academy of Sciences of the United States of America* 98 (12), 6742–6746.
- Fu, J., Bian, X., Hu, S., *et al.*, 2012. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nature Biotechnology* 30 (5), 440–446.
- Jiang, W., Bikard, D., Cox, D., Zhang, F., Marraffini, L.A., 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* 31 (3), 233–239.
- Kutter, E.M., Kuhl, S.J., Abedon, S.T., 2015. Re-establishing a place for phage therapy in western medicine. *Future Microbiology* 10 (5), 685–688.
- Marinelli, L.J., Piuri, M., Hatfull, G.F., 2019. Genetic manipulation of lytic bacteriophages with BRED: Bacteriophage recombineering of electroporated DNA. *Methods in Molecular Biology* 1898, 69–80.
- Murphy, K.C., 1998. Use of bacteriophage λ recombination functions to promote gene replacement in *Escherichia coli*. *Journal of Bacteriology* 180 (8), 2063–2071.
- Murphy, K.C., 2016. λ Recombination and recombineering. *EcoSal Plus* 7 (1), doi:10.1128/ecosalplus.ESP-0011-2015.
- Murphy, K.C., Nelson, S.J., Nambi, S., *et al.*, 2018. ORBIT: A new paradigm for genetic engineering of mycobacterial chromosomes. *mBio* 9 (6), doi:10.1128/mBio.01467-18.

- Sawitzke, J.A., Thomason, L.C., Costantino, N., *et al.*, 2007. Recombineering: In vivo genetic engineering in *E. coli*, *S. enterica*, and beyond. *Methods in Enzymology* 421, 171–199.
- Stahl, M.M., Thomason, L., Poteete, A.R., *et al.*, 1997. Annealing vs. invasion in phage λ recombination. *Genetics* 147 (3), 961–977.
- Thomason, L.C., Sawitzke, J.A., Li, X., Costantino, N., Court, D.L., 2014. Recombineering: Genetic engineering in bacteria using homologous recombination. *Current Protocols in Molecular Biology* 106. (1:16.1–1:16.39).
- Zhang, Y., Buchholz, F., Muirers, J.P., Stewart, A.F., 1998. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genetics* 20 (2), 123–128.

Nanotechnology Application of Bacteriophage DNA Packaging Nanomotors

Tao Weitao, Southwest Baptist University, Bolivar, MO, United States

Lixia Zhou, Zhefeng Li, Long Zhang, and Peixuan Guo, College of Pharmacy, The Ohio State University, Columbus, OH, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ATP Adenosine triphosphate

qRT-PCR Quantitative reverse transcription polymerase chain reaction

Glossary

ATPase gp16 gp16 is an enzyme that can catalyze the hydrolysis of ATP. The gp16 is a DNA packaging ATPase from the virus phi29.

Bacteriophage A virus that can infect and replicate within bacteria.

Biomotor/nanomotor A motor with size in nanometer scale functions to gear molecules moving forward.

Connector/nanopore A hollow channel or pore in nanometer scale for the translocation of molecules, also known as the portal.

Procapsid A preformed protein shell in the virus assembly stage of many bacteriophages.

Introduction

We are all surrounded by traditional macroscale motors including both combustion engines and electric motors. One might ask: can we downscale the way we think of motors to the nanoscale range? The possible applications of nanomotors in nanotechnology will be intriguing but different. However, when we look at the basic mechanisms of both the macroscale and the nanoscale motors, they show many striking parallels and similarities. Analogous to the macroscopic ones, biological nanomotors are manufactured from static modules or parts that form a framework for moving and transporting substrates. Even though the moving parts are not always restricted to rotary or revolving motion, some motors perform the linear motion. Moreover, both macroscopic and nanomotors work by repeating the same cyclical of steps over and over.

Active motion is a typical and prominent feature of life. Motion distinguishes living organisms from inanimate matter. Motion is action derived from motors. All important life activities rely on biomotors – macroscopic respiration, heartbeats, walking, speaking, blinking. Biomotors drive even microscopic DNA replication, DNA repair, RNA transcription, homologous recombination, intracellular transport, molecular transport, and viral assembly. Biomotors are powered by ATP. The amount of ATP produced and consumed by one person is close to body weight, which is 50–80 kg per day.

Biological nanomotors are nanoscale machines that convert a primary energy source to mechanical motion between active and framework components. Motor complexes are vital to biological systems, and they support efficient, directional motility for the transportation of cellular components. Due to the progress of structural biology – single molecule methods and biotechnology approaches – the understanding of the molecular basis of structure in biological movement has progressed to the point where we can start thinking about possible applications of biomotors in nanotechnology. The elegant and delicate structures of viral double-stranded (ds)DNA packaging motors have inspired their application in nanotechnology. For example, studies on the motor packaging RNA (pRNA) of bacterial phage phi29 has led to the emergence of the field of RNA nanotechnology for disease treatment (Fig. 1). With the discovery of the highly stable pRNA three-way junction (3WJ) motif (Fig. 1), the application for constructing multivalent RNA nanoparticles with high chemical and thermodynamic stability leads to the development of RNA nanotechnology. RNA nanotechnology is the bottom-up self-assembly of nanometer-scale RNA architectures that are composed mainly of RNA (Fig. 1). In RNA nanoparticles, the scaffolds, targeting ligands, and therapeutics miRNAs or siRNAs can be composed mainly of RNA (Fig. 1). While classical studies on RNA focus on intra-RNA interactions and 2D/3D structure, more recently, RNA nanotechnology has further extended to inter-RNA interaction and quaternary (4D) structure (Fig. 1). The resulting branched RNA nanoparticles are homogenous, uniform in size and shape, and can harbor different functionalities while retaining their tertiary folding and independent functionalities both in vitro and in vivo.

Another new approach for high potent drug development leads to the study of motor ATPase with high stoichiometry of homologous subunits (Fig. 2). Many ATPase biomotors contain six pRNA copies that can serve as drug targets ($Z = 6$); and inducing a single pRNA subunit inhibits the entire biomotor function ($K = 1$), resulting in high inhibition (Fig. 2(D)). Leading to the creation of a novel drug targeting strategy: increasing the number of subunits that are potential targets (changing Z) further improving drug efficiency. This idea is akin to series circuits, such as those in traditional Christmas lights, where if a single bulb stops working, the entire strand fails (Fig. 2(A)). In contrast, traditional drug development is analogous to parallel circuits'

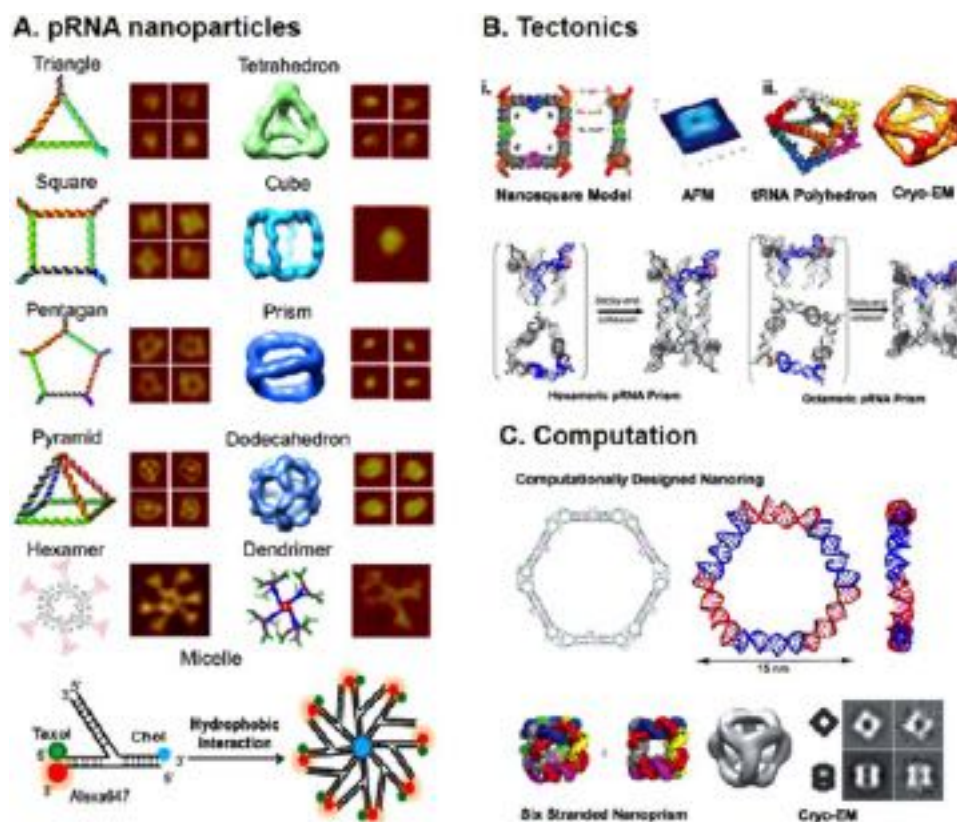


Fig. 1 Emerging field of RNA nanotechnology. (A) Construction and characterization of multivalent pRNA nanoparticles with tunable shape and stoichiometry. (B) Tectonics method to construct RNA nanosquares, polyhedron from tRNA, and nanoprisms from phi29 pRNA. (C) Computational approaches to expedite RNA nanoparticle manufacture and optimization. Adapted with permission from (A) Xu, C., Haque, F., Jasinski, D.L., *et al.*, 2018. Favorable biodistribution, specific targeting and conditional endosomal escape of RNA nanoparticles in cancer therapy. *Cancer Letters* 414, 57–70. Available at: doi:10.1016/j.canlet.2017.09.043 © 2017 Elsevier B.V. All rights reserved. (B) Jasinski, D., Haque, F., Binzel, D.W., Guo, P., 2018. The advancement of the emerging field of RNA nanotechnology. *ACS Nano* 11(2), 1142–1164. Available at: doi:10.1021/acs.nano.6b05737 Copyright © 2017 American Chemical Society. (C) Shu, Y., Yin, H., Rajabi, M., *et al.*, 2018. RNA-based micelles: A novel platform for paclitaxel loading and delivery. *Journal of Controlled Release* 276, 17–29. Available at: doi:10.1016/j.jconrel.2018.02.014 © 2018 Elsevier B.V. All rights reserved.

challenge to circumvent, as a single bulb's failure has no effect on the rest of the strand. By addressing multicomponent systems more like series circuits, it may be possible to create ultra-high inhibitory drugs. The phi29 motor is part of a large ATPase family (AAA +) that was used to elucidate this potent drug inhibition mechanism. Drugs that target viral motors in this manner include rimantadine for influenza, ALLINI for HIV, and bedaquiline for tuberculosis (Fig. 2(E)).

Bacillus phi29 and *Escherichia T7* present themselves as one striking feature in double-stranded (dsDNA) tailed bacteriophages of the *Podoviridae* family. During replication and morphogenesis, they encapsulate and compress their genomic DNA with tremendous velocity into preformed protein shells, called procapsids, driven by hydrolysis of ATP. This energetically unfavorable translocation is accomplished by a viral nanomotor. The viral DNA packaging motor is composed of a few parts. The gearing components are the first; and the second is the motor structure frame, such as the force generation ATPase. The third is the essential components of connector complex that is a dodecameric structure with a central channel (30–60 Å) through which viral DNA is packaged into the capsid and exits during infection. Though the portal proteins from different viruses have little sequence homology and large variation in molecular weight, portal complexes show significant morphological similarities. Yet, the portal connector of phi29 is bound with both ATPase gp16 and prohead RNA (pRNA). Binding of gp16 to the packaging motor relies on the association of pRNA because the central domain of pRNA is there for connector binding, and the 5'/3' paired helical region for recruitment of gp16. In phi29, moreover, the structure of one phage portal protein has been determined at atomic resolution. The connector ring consists of a twelve α -helical subunit, with three long helices of each subunit forming the central channel. The ring is 138 Å across at its wide end and 66 Å at the narrow end. The internal channel is 60 Å at the top and 36 Å at the bottom. The wider end of the connector is located in the prohead and its narrow end partially protrudes out of the capsid. The connector is located at the five-fold vertex of the viral capsid, which leads to a symmetry mismatch between capsid and the portal connector.

Extensive investigation of the group of dsDNA viruses reveals that all DNA packaging motors known so far involve two nonstructural components typical of ATPase. Further scrutiny has found these two proteins can be classified into two categories according to their roles and sizes. One category with larger size is for procapsid binding, such as gp16 of phi29, gpA of λ , gp12 of

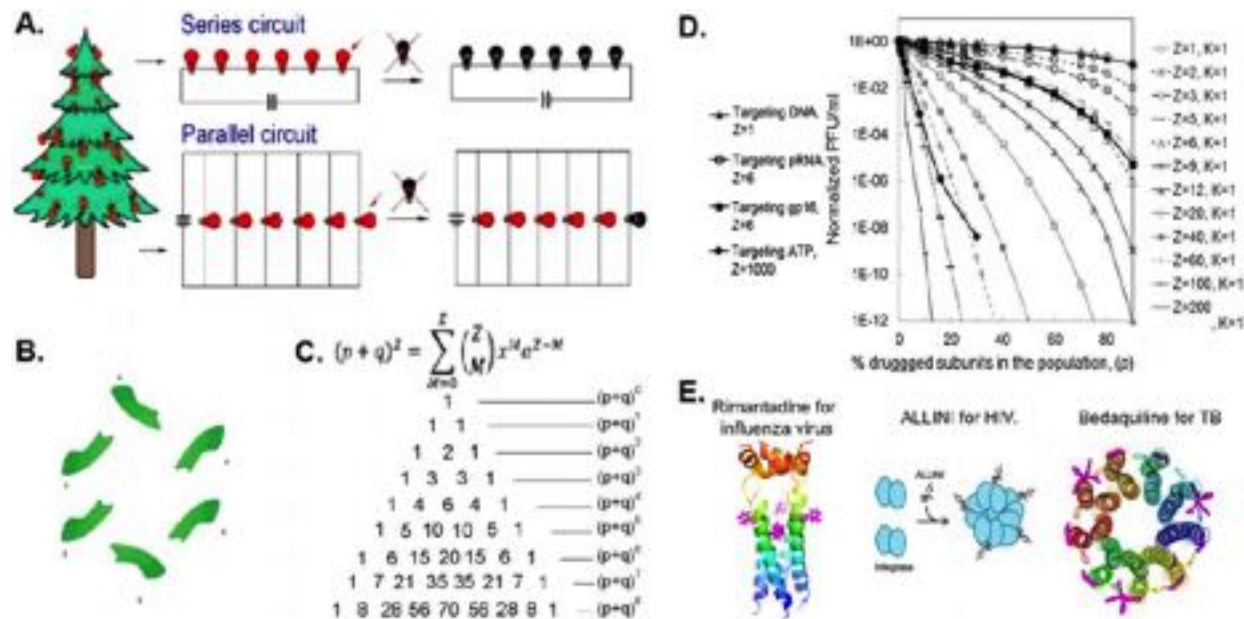


Fig. 2 Biomotors have high stoichiometry of homologous subunits. Drug inhibition efficiency is correlated with the stoichiometry of the targeted biocomplex with a mechanism of series circuit, using the hexameric ATPase as an example. (A) In the series circuit of Christmas lights; one broken bulb will turn off the whole chain. (B) The hexameric ATPase. (C) One key factor regarding drug potency is the stoichiometry of the homo-subunit serving as a target. Interactions between the subunits may follow the binomial distribution and Yang Hui's Triangle. (D) Assay to prove the potent drug development based on the series circuit model. Plotted is virus assembly inhibition effect by drugged components of DNA, pRNA, gp16 ATPase, and ATP with stoichiometry of 1, 6, 6, and 1000, respectively. (E) Drugs that target ATPase have high stoichiometry. (left) Rimantadine for influenza virus, (middle) ALLINI for HIV, and (right) bedaquiline, the recent FDA-approved drug that targets the ATPase of *Mycobacterium Tuberculosis*, one of the most antimicrobial-resistant bacteria. (E) Adapted with permission from Pi, F., Zhao, Z., Chelikani, V., et al., 2016. Development of potent antiviral drugs inspired by viral hexameric DNA packaging motors with revolving mechanism. Journal of Virology 90(18), 8036–8046. Available at: doi:10.1128/JVI.00508-16. Copyright © 2016, American Society for Microbiology.

phi21, gp17 of T4, and gp19 of T3/T7. The second category is the smaller-size protein for DNA binding and processing. This group includes gp16 of T4, gp18 of T3/T7, gpNu1 of λ , and gp1 of phi21. The DNA-packaging reaction is prohead and DNA dependent. One ATP is used for the packaging of two base pairs of DNA. Emerging information reveals that the mechanism of herpes virus DNA packaging is very similar to that of phages.

General structures of phage dsDNA motors include channels and DNA packaging enzymes (Fig. 3). The tailed-bacteriophage assembles the channel into an oligomeric-ring portal (connector), which is composed of 12 protein subunits and embedded in the pentagonal portal vertex of the procapsid. The phi29 connector has a truncated cone-shape. The connector functions as a nucleation point for the assembly of motor components. Both the DNA-packaging enzyme gp16 and pRNA bind the connector. Binding of gp16 to the packaging motor relies on binding of pRNA because the central domain of pRNA is for connector binding, and the 5'/3' paired helical region for recruitment of gp16.

The ingenious structures and functions of the phage biomotors have inspired us to develop nanobiotechnology for medical application. As RNA is a building block in nanotechnology, phi29 pRNA can be used to construct nanoparticle vehicles to deliver various therapeutics to the target cells for treatment of cancers, viral infections, and some genetic diseases. The phi29 DNA-packaging motor can be engineered and embedded into cell membrane to generate a pore for molecular sensing and drug delivery to the target cells for precision medicine. In this work, we intend to delineate the revolving mechanism of the biomotors in bacteriophages, with an emphasis on phi29. With our understanding of the dsDNA translocation mechanism by the phi29 packaging motor, we introduce the application of the packaging motors, especially the motor connectors, in single pore sensing of DNA, RNA, chemicals and proteins.

The nanomotors play significant roles in many biological processes as described above. The phage DNA packaging motors, such as phi29, SPP1, T4, and T7, package their DNAs into the procapsid during genome replication. Herein, due to space limitation, we will use the phi29-gp10 revolution motor as an example to explain how it packages its genome into the procapsid. From the movement mode, the nanomotors can be classified into linear, rotation, and revolution categories. (Please refer to the further reading list to acquire more knowledge on the mechanism of different biomotors). Fig. 4 shows the structure of several revolution nanomotors. Generally speaking, revolution nanomotors have larger size channel with left-handed chirality. Fig. 3 shows a simplified illustration of the mechanisms for the revolution motor of phi29 to drive unidirectional translocation of dsDNA. Briefly, the ring-shaped ATPase present on pRNA binds ATP to trigger the conformational changes of the ATPase subunit to revolve the dsDNA and move it forward. The phi29 gp10 connector has a large diameter and the narrowest part in the channel is ~ 3.6 nm to allow the translocation of dsDNA into the procapsid. The movement of the dsDNA around the motor is similar to how the

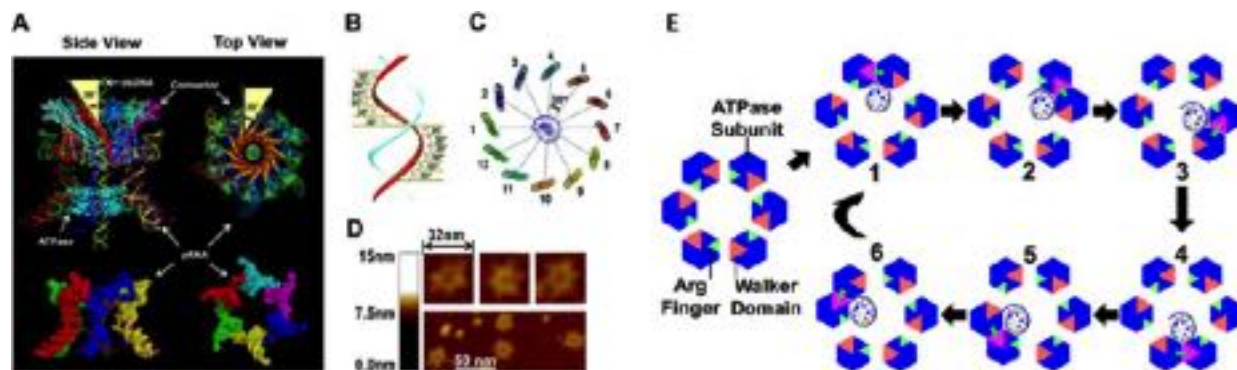


Fig. 3 Structure and function of phi29 DNA packaging motor. (A) Structure of hexameric pRNA and the connector showing a 30° tilt. (B)–(C) dsDNA showing the shift of 30° angle between two adjacent connector subunits. (D) AFM images of hexameric pRNA with 7-nucleotide loops. (E) The mechanisms for the revolution motor. Reprinted with permission from (A)–(C) Schwartz, C., De Donatis, G.M., Zhang, H., Fang, H., Guo, P., 2013. Revolution rather than rotation of AAA + hexameric phi29 nanomotor for viral dsDNA packaging without coiling. *Virology* 443(1), 28–39. doi:10.1016/j.virol.2013.04.019. Copyright 2013 Elsevier. (D) Shu, Y., Haque, F., Shu, D., *et al.*, 2013. Fabrication of 14 different RNA nanoparticles for specific tumor targeting without Accumulation in normal organs. *RNA* 19, 766–777. Copyright 2013 RNA Society. (E) Guo, P., Driver, D., Zhao, Z., *et al.*, 2019. Controlling the revolving and rotating motion direction of asymmetric hexameric nanomotor by arginine finger and channel chirality. *ACS Nano* 13, 6207–6223. Available at: <https://pubs.acs.org/doi/10.1021/acs.nano.8b08849>. Further permissions related to the material excerpted should be directed to the ACS.

Earth revolves around the Sun but without self-rotation of the Earth, which avoids the coiling and tangling that happens in large or lengthy DNAs. From this DNA packaging process, researchers determined that the biomotors drive unidirectional DNA passage through the nanopore into the procapsid.

The single pore sensing is derived from the natural phenomenon of viral DNA packaging. Scientists have discovered that analytes can be driven through the extracted nanopore and each generates a unique electric signal with applied voltage. The proof-of-concept about the nanopore sensing was initially demonstrated by the inspection of DNA translocation through the α -hemolysin nanopore in 1996. In their research, the qRT-PCR verified the successful translocation of the single-stranded DNA (ssDNA) that was able to be driven electronically through the α -hemolysin nanochannel. To date, the nanopore sensing has been developed for different analyte detection as it has many advantages as an analytical tool. First, the detection is label- and amplification-free. Second, it can achieve the single-molecule detection with high specificity and sensitivity and offer real-time identification. Third, it only requires small analyte sample volumes (μ L range) and low concentrations (nM or pM range) for analysis. In the following paragraphs, we will summarize the third type revolving motor mechanism and the application of these motors in single molecule nanopore sensing.

The Revolving Mechanism of Biomotors in Bacteriophages

The Revolution Mechanism Without Rotation Common to the dsDNA Packaging Motors of all the dsDNA Bacteriophages

During replication, the dsDNA viruses translocate their genomic DNA into procapsids using the same mechanism by the nanomotors with ATP as an energy source in an entropically unfavorable process. The force generation mechanism of the bacteriophage packaging motors of phi29, HK97, SPP1, P22, T4, and T7 share the following common revolution mechanism without rotation. The 30° left-handed twist of the channel wall causes an anti-parallel arrangement with the right-handed helix of the dsDNA, leading to one-orientation trend (Fig. 3(B) and (C)). The same twist has been found in motor channels of all dsDNA viruses known so far including phi29, HK97, SPP1, P22, T4, and T7, of which their primary amino acid sequences were non-conserved, but the higher structures of the swivel were impeccably conserved and aligned. One-direction flow loops inside the channel of SPP1 and phi29 help a one-directional processing for the one-way translocation of dsDNA. The electropositive-lysine layers present in all the phage channels interact with one strand of the electronegative-dsDNA phosphate backbone, causing a relaying contact and transitional pausing during dsDNA translocation. In April of 2020, Xiangxi Wang, Hongrong Liu and Zihe Rao reported a high-resolution structural and mechanism of the ATP-driven DNA packaging motor of the double-stranded herpesvirus. The structure clearly shows that the herpesvirus DNA packaging motor is a hexamer exercising the revolution mechanism instead of rotation (Fig. 4(A)). The elucidation of the structural ends the 20-year fervent debate on whether the stoichiometry of the genome packaging motor of dsDNA viruses is pentamer or hexamer, or whether the motion process is rotating or revolving. Like the architecture of the phage packaging motors (Fig. 4(A)), a hexameric ring structure of the herpesvirus DNA packaging complex suggests that the motor translocates dsDNA by a sequential revolution mechanism with the aid of a transacting arginine fingers essential for ATP hydrolysis and DNA translocation. Astonishingly, the human viral motor has the structural properties similar to the bacteriophage motors, implying that the sequential DNA translocation via revolving motion is common during viral motor evolution in both eukaryotic and prokaryotic systems.

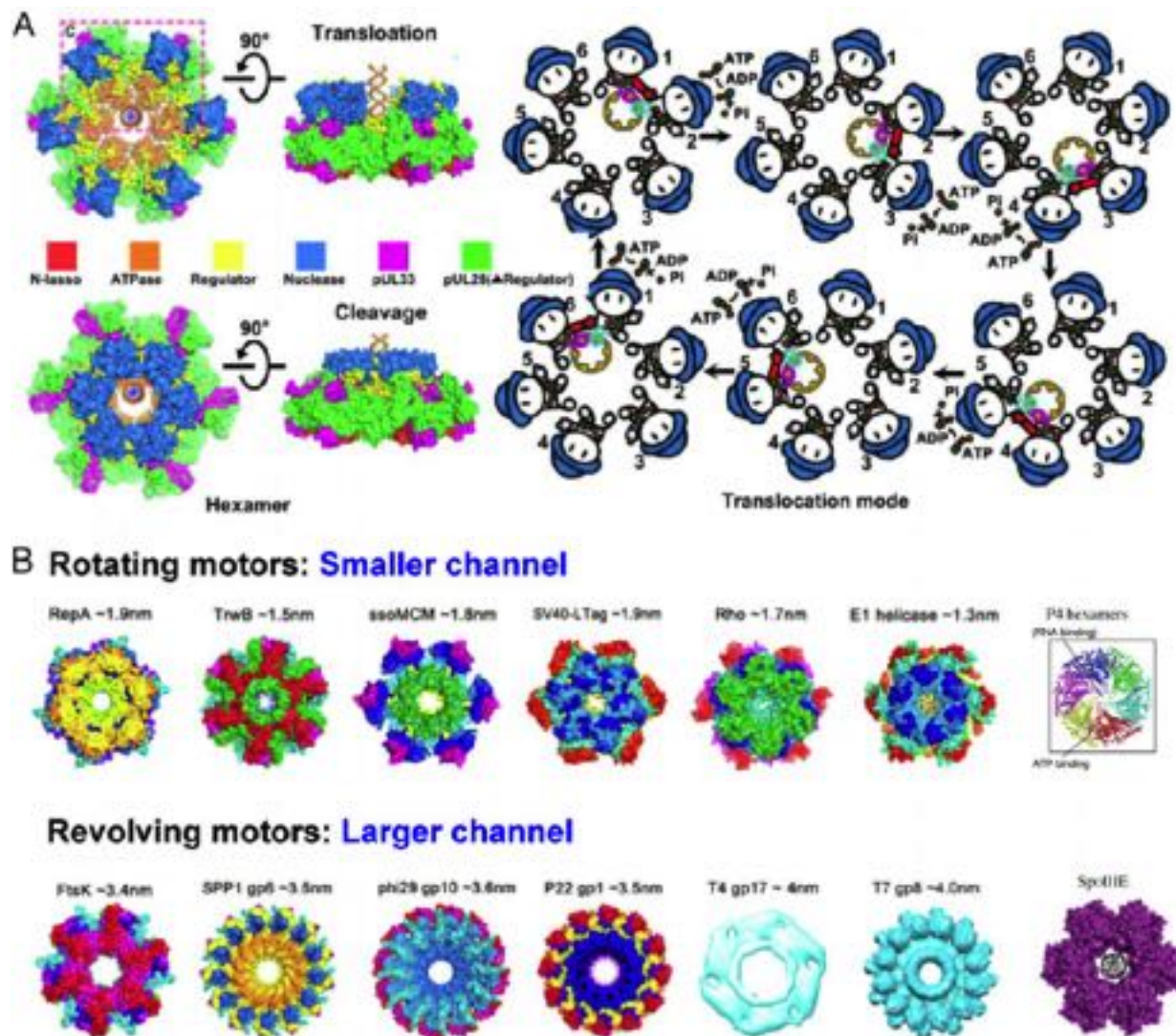


Fig. 4 The structure of hexameric ATPase motor. (A) The DNA packaging motor of the double-stranded DNA herpesvirus (left) and the elucidation of the sequential revolving mechanism in DNA translocation (right). Adapted in part with permission from Creative Commons Attribution 4.0 International in Architecture of the herpesvirus genome-packaging complex and implications for DNA translocation. Yang *et al.* Protein Cell. doi:10.1007/s13238-020-00710-0. Copyright 2020. (B) Channel size to differentiate rotating and revolving mechanism. Rotating motors have channel sizes all ≤ 2.0 nm in diameter to ensure full contact between DNA and channel wall similar to the nut driving the bolt, while revolving motors have channel sizes ≥ 3 nm to have room to accommodate the revolving motion. Reprinted in part with permission from: Guo, P., Grainger, I., Zhao, Z., Vieweger, M., 2014. Two classes of nucleic acid translocation motors: Rotation and revolution without rotation. Cell and Bioscience 4, 54. Copyright 2014 Springer Nature. Mancini E.J., Tuma, R., 2012. Mechanism of RNA packaging motor. Advances in Experimental Medicine and Biology 726, 609–629. Besprozvannaya, M., Burton, B.M., 2014. Do the same traffic rules apply Directional chromosome segregation by SpoIIIE and FtsK. Molecular Microbiology 4, 599–608.

Revolving Mechanisms Determined by Motor Channel Sizes

For dsDNA to rotate through the center of the motor channel, the channel needs to have a size similar to the diameter of dsDNA, 2 nm. Thus, the channel diameter should be approximately 2 nm for dsDNA or 1 nm for ssDNA. For dsDNA to revolve through the motors, dsDNA moves by touching the channel wall rather than proceeding through the center of the channel. Accordingly, the channel diameter should be much larger than the dsDNA's to ensure sufficient room for revolution. The crystal structures and cryoelectron microscopy studies on the connectors of the T7, phi29 and other phages demonstrate that the channel diameters of the revolution motors are >3.5 nm, and those of rotation motors such as helicase and DNA polymerase is < 2 nm. **Fig. 4(B)** shows several rotating motors with smaller channel but revolving motors with larger channel. The revolution of dsDNA along the hexameric ATPase ring has the following parameters: the width of dsDNA helix is 2 nm while the diameters of the narrowest regions of the dodecameric portals of all the connector channels of phi29 SPP1, T4, T7, HK97, and FtsK are 3–5 nm. Even a herpesvirus motor complex forms a central channel with an internal diameter of 3.9 nm, greater than that of dsDNA as revealed by Cryo-electron microscopy (**Fig. 4(A)**).

Revolving Motors Distinguished by Chirality

The anti-parallel arrangement between the phi29 connector subunit and the DNA helices can facilitate dsDNA to revolve in single direction. All 12 subunits of the phi29 connector portal protein, relative to the vertical axis of the channel, tilt at a 30° left-handed angle to form the channel with a configuration anti-parallel to the right-hand dsDNA helix during packaging. This structural arrangement significantly facilitates the controlled motion, providing evidence for dsDNA revolving, instead of rotating, through the connector channel. The revolution neither produces coiling, or torsion force, nor touches each of the 12 connector subunits in 12 discrete steps of 30° transitions for each helical pitch. Fig. 5 demonstrates examples of rotating motors with right-handed channel whereas revolving motors with left-handed channel.

Special Aspects of Revolving Motor Actions

Force generation and energy conversion Several primary chemo-mechanical coupling mechanisms exist in all biomotors studied so far. By these mechanisms, the cycles of nucleotide binding and hydrolysis are coupled to conformational entropy rearrangements of the substrate-binding subunits of the motors. In the sequential mechanism, individual ATP binding and hydrolysis events proceed sequentially. For the concerted mechanism, all active sites on the motors simultaneously hydrolyze ATP. Of the stochastic mechanism, the ATPase sites randomly hydrolyze ATP. For all three mechanisms, ATP binding to the disordered subunits of ATPase stimulates conformational alterations of the ATPase with the entropy change. This fastens the bound ATPase at a less random configuration than the unbound ATPase. Such new conformation entropy facilitates the ATPase subunits to bind dsDNA and to prime ATP hydrolysis. ATP hydrolysis by the ATPase triggers the subsequent entropic and conformational changes. These alterations render the ATPase a low affinity for the dsDNA substrates, thereby leaving dsDNA to the next ATP-bound subunit that has a high affinity with dsDNA. These repetitive actions cause dsDNA to revolve around the subunits through the interior channel of the motor.

Unidirectional dsDNA translocation The translocation direction of dsDNA is controlled by five factors in phi29 motor (Fig. 3). First, ATPase experiences a chain of entropy transitions and conformational changes during ATP and dsDNA binding. ATP hydrolysis causes additional change in entropy and conformation of the ATPase, including the one with a low affinity for dsDNA that pushes dsDNA away and revolves inside the channel. Second, the 30° angle of each subunit of the dodecameric connector channel is anti-parallel to dsDNA, matching with the 12 subunits of the connector channel ($360^\circ \div 12 = 30^\circ$), as shown by crystallography (Figs. 3 and 5). Third, the internal channel loops have the unidirectional flowing property acting as a ratchet valve to avoid dsDNA reversal. Fourth, the $5' - 3'$ single-directional revolution of one strand of dsDNA goes along the connector channel wall. Fifth, the electrostatic force occurs from the relaying interaction of the electropositive-lysine layers with the electronegative-DNA phosphate backbone. Altogether, dsDNA is translocated in one direction by the multiple factors.

Communication/interaction between motor subunits for sequential action In the phi29 motor, a sequential action of both the pRNA and the ATPase gp16 was reported. Particularly, the subunits of ATPase act sequentially and cooperatively. The arginine finger motif plays a role in communication and interaction among the subunits by bridging dimer formation (Fig. 3(D)). Communication between



Fig. 5 Different chiralities of rotating and revolving motors. Rotating biomotors exhibit right-handed chirality to drive the right-handed dsDNA similar to the nut driving the bolt or the screwdriver turning the screw, whereas revolving biomotors exhibit left-handed chirality within the channel. Crystal structure analysis of viral DNA packaging motors reveals that this class of biomotors package DNA using the revolving mechanism. Reprinted with permission from Guo, P., Grainge, I., Zhao, Z., Vieweger, M., 2014. Two classes of nucleic acid translocation motors: Rotation and revolution without rotation. *Cell and Bioscience* 4, 54. Copyright 2014 Springer Nature. and De-Donatis, G.M., Zhao, Z., Wang, S., *et al.*, 2014. Finding of widespread viral and bacterial revolution dsDNA translocation motors distinct from rotation motors by channel chirality and size. *Cell and Bioscience* 4, 30. Copyright 2014 Springer Nature.

each two neighboring subunits is mediated by the finger. Such communication causes an asymmetrical hexameric organization, consistent with the asymmetrical structures in other hexameric ATPase systems. Thus, the arginine finger acts as a bridge between two subunits to form a transient dimeric subunit. Specifically, the arginine finger, located at the interface of two subunits of gp16, extends into the ATP binding pocket of the downstream subunit. The asymmetrical appearance of one dimer and four monomers in high-resolution structural complexes provide evidence for the mechanisms by which the arginine finger promotes inter-subunit interactions and sequential movements of individual subunits. The arginine finger is important for regulating energy transduction and motor function (Fig. 3(D)).

The Application of Bacteriophage DNA Packaging Motors in Single Pore Sensing of DNA, RNA, Chemicals, and Proteins

The single pore sensing technique has built up a powerful platform for numerous applications including the detection of DNAs, peptides, and chemicals. Prior to discussing the use of nanopores for single-molecule sensing, we present background information about the history of single molecule sensing. The well-known biological connectors found in bacteriophages include the phi29, SPP1, T4, P22, and T7. The following lists several well-studied connectors.

Phi29 Connector

The phi29 connector was the first portal protein with the atomic structure solved. The connector consists of 12 protein subunits (gp10) that form a ring-like truncated cone structure. The molecular weight for each subunit is 36 kDa. The end diameters of the cone structure are 6.6 and 13.8 nm, with the wider and narrower ends termed the C- and N-terminals, respectively. The area of the narrowest part of the inner channel is 10 nm², which corresponds to a diameter of ~3.6 nm. In bacteriophage phi29, the C-terminal is located in the procapsid, and translocation of dsDNA during packaging is unidirectional from the N-terminal to C-terminal. The clip region in the phi29 connector can bind with pRNA to help with the DNA packaging. In addition, the negatively charged residues, aspartate and glutamate in the channel, have the potential to keep the DNA in the center of the channel; this central localization is necessary for the phi29 DNA packaging. The phi29 nanochannel also has the positively charged residues, such as lysine and arginine, which can attract the DNA and hence prevent the DNA from slipping out.

SPP1 Connector

There are four domains in the SPP1 connector: clip, stem, wing, and crown. This connector has 13 protein subunits (gp6) that assemble into a cone structure similar to the phi29 connector, with a total molecular weight of 745 kDa. The overall diameter of the SPP1 nanochannel is ~16.5 nm with a height of ~11 nm. The most constricted part in the tunnel is ~2.7 nm. The negatively and positively charged residues in the SPP1 channel are also required for DNA packaging similar to those of the phi 29 motor.

T7 Connector

The T7 connector is formed by 12 protein subunits with a molecular weight of 59 kDa for each subunit. The channel length is 13.1 nm, with external diameters ranging from 5.9 to 17.3 nm. The most restricted region in the T7 interior channel is 3.9 nm, which is relatively wide compared to the other types of bacteriophage connectors. The lysine residue in the stem domain of T7 nanochannel is highly possible to interact with the phosphate backbone to help with DNA translocation.

Apart from phi29, SPP1, and T7 connectors, many other phage connectors such as T3, T4, and P22 were studied. Generally, the phage connectors consist of about 11–13 subunits of proteins, with their height and width on the order of 10's nm and inner diameter within 4 nm.

Over the years, the nanopore-based sensing technique has been developed for the detection of various molecules and has proven its potential for the medical diagnosis of diseases. The following sections discuss the mechanisms and provide examples of single-molecule sensing using nanopores. Please note that we will focus on the discussion of the biological nanopores from bacteriophages in this article. For information on other types of nanopores, please refer to the further reading list.

The Mechanism of Single Pore Sensing

The general mechanism of biological nanopore sensing is based on the resistive pulse technique. Fig. 6(A) shows a schematic diagram of the technique. Briefly, the purified connector is inserted into the lipid bilayer membrane to form the nanopore channel. Voltage is applied across the membrane, which allows ions to pass freely through the channel. When analytes are translocated through the membrane, ion flow is affected. This results in a change to the current and creates fingerprinting signals. Commonly used electronic signatures that help identify different analytes include the current blockage and dwell time. The current blockage represents the percentage of the blocked current relative to the open current of the nanopore channel, while the dwell time measures the length of time the current blockage lasts.

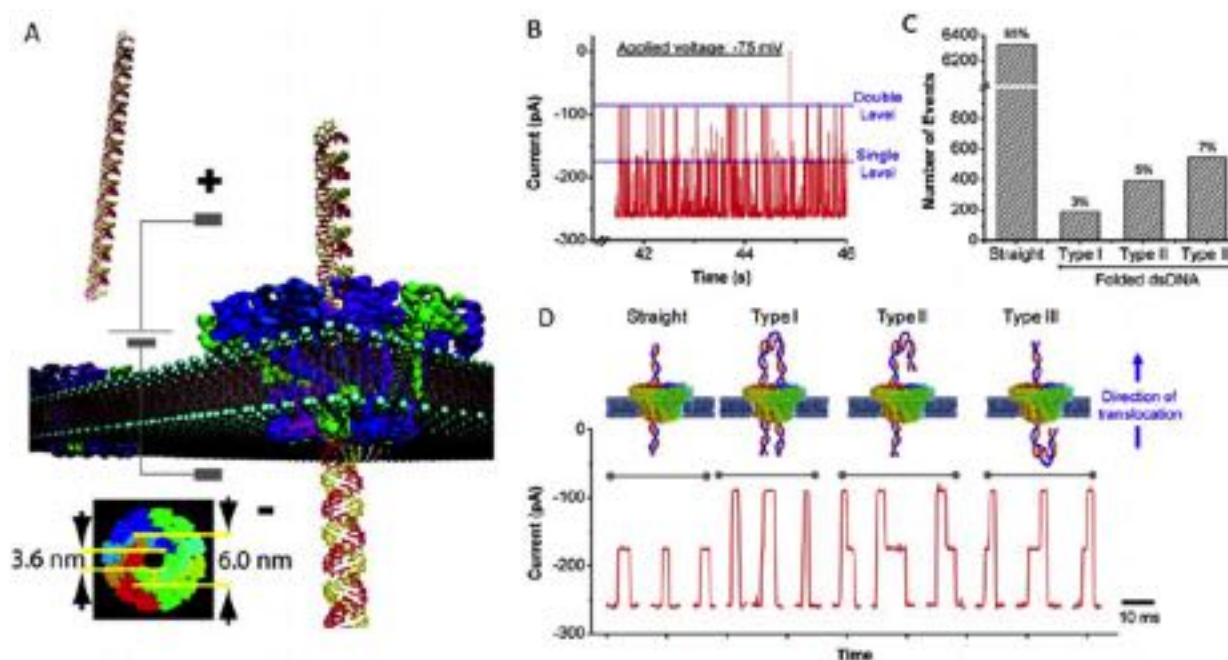


Fig. 6 The sensing of dsDNA with the phi29 connector. (A) The schematic diagram of the single pore sensing of a membrane-embedded connector with applied voltage in vitro. (B) The current trace signals of dsDNA. (C) The distribution of 7500 translocation events for various types of DNA from quantitative analysis. (D) The illustration and corresponding current blockage signals for different types of DNAs. Reprinted with permission from: (A) Jing, P., Haque, F., Shu, D., Montemagno, C., Guo, P., 2010. One-way traffic of a viral motor channel for double-stranded DNA translocation. *Nano Letters* 10 (9), 3620–3627. Further permissions related to the material excerpted should be directed to the ACS. (D) Haque, F., Wang, S., Stites, C., *et al.*, 2015. Single pore translocation of folded, double-stranded, and tetra-stranded DNA through channel of bacteriophage phi29 DNA packaging motor. *Biomaterials* 53, 744–752.

As analytes are driven to translocate or interact with the connector by electric force, many things may affect current blockage and dwell time. These things include intrinsic properties of the targets such as their shape, length, charge, hydrophobicity, or hydrophilicity, as well as the type and modification of the analytes. First, analytes that are smaller than the nanopore may be translocated through the channel to generate electric signatures, and different samples have the potential to generate different signals. Second, analytes that are larger than the diameter of the nanopore cannot pass through the channel due to size limitations. Theoretically, larger analytes or those with longer chains will block more current signals and have a longer dwell time than smaller analytes. In addition, stronger interactions between the analytes and the connector can produce longer dwell times than those have weaker interactions with the nanopore. Moreover, some analytes can interact with the connector on one side and induce conformational changes to the connector, thus generating distinct electric signals without translocation. Since only specific types of analytes can trigger conformational changes to the connector and cause current blockages with different dwell times, this becomes a technique for analyte determination. In both cases concerning small and large analytes, the type and modification of the connector can greatly affect the electric signals. For instance, a larger nanopore can allow larger analytes to be translocated through the pore, while smaller nanopore channels do not. Moreover, modification of the nanopores, such as site-directed mutagenesis or introducing ligands into the nanopore, can significantly influence the interactions between the analytes and the nanopore, resulting in different electric signals. For example, introducing more hydrophobic groups into the nanopore can increase the dwell time for samples with many hydrophobic groups. Please refer to the further reading list for additional information on the modification of the connector for nanopore-based sensing techniques.

In conclusion, biological nanopore sensing is based on distinct electric signatures generated when analytes pass through or interact with the nanopore. The intrinsic properties of the analytes and the nanopore determine the fingerprinting signals, which allow for the differentiation and identification of unique analytes.

Applications of the Biological Nanopore Channel as a Conduit for Single Pore Sensing

As mentioned in the nanopore detection mechanism section above, different analytes can generate distinctive electric signals to be distinguished from the background. The following sub-sections will address several experimental examples to explain the single pore sensing technique further.

The sensing of DNA: DNA carries genetic information and is imperative for all lives. Recognizing the variations in different types of DNA may help people better elucidate their functions, which can benefit the diagnosis of diseases in their early stages, and further assist in the treatment of these diseases.

The single pore sensing of DNA recognizes that different size and conformation can cause various current blockage events. The phi29 gp10 connector has been shown for the differentiation of dsDNA with different conformations by Haque and coworkers. In the experiment, Haque *et al.* examined the translocation and analyzed the current blockage of the folded 5 kbp dsDNA. Fig. 6 shows the current signatures for the dsDNA with different conformations that caused distinct blocking events. The straight dsDNA caused a single level blockage $\sim 32\%$, while the tetra-stranded complex with two dsDNA geared into nanostructure blocked the channel current with $\sim 64\%$. Additionally, the dsDNA and tetra-stranded DNA (tsDNA) also showed difference in their current blockage events. In short, the dsDNA mainly caused the 32% current blockage events whereas the tsDNA produced both the 32% and 64% blockage events. These experiments fortify a foundation for future studies on the detection of biomarkers from the conformational changes.

The sensing of RNA: Different types of RNAs have been found to be able to serve as the biomarkers for diseases diagnosis, demonstrating their importance to lives.

Currently, the sensing of RNAs with bacteriophage nanopores is at the beginning stage. The work in the detection of single-stranded RNA (ssRNA) with a modified phi29 nanopore has promoted the application of bacteriophage connectors in RNA detection. Geng *et al.* determined that the removal of the internal loop segment of phi29 channel could create a modified channel with a cross-sectional area about 40% less than that for the wild-type phi29 connector. The translocation of ssRNA was able to be identified by the modified phi29 connector, which resulted in $\sim 20\%$ current blockage. Their experiments suggested that the modification of nanochannels could achieve different detection purposes. Please refer to the further reading list for more details.

The sensing of chemicals: Chemicals are essential to our everyday life. Biological nanopores are also proven to be capable for the detection of the chemicals after connector modifications.

Presently, the chemical detection is mainly based on the physical blocking of current signals when chemicals interact with the functional groups in the connector. The detection of chemicals with modified bacteriophage connector has been reported. By mutating the lysine-234 to cysteine in the phi29 nanopore, the modified phi29 nanopore can distinguish among ethane, thymine, and benzene with thioester moieties. The binding events between the connector and chemicals can cause current blockage events. The unique electric signals are generated by the physical blockings when the chemicals react with the cysteine residues on the connector during the translocation process. From the analysis of the binding events among ethane, thymine, and benzene, the current blockage for the permanent binding events have been found to allow for the discrimination among the three chemicals (Table 1).

The biological nanopore sensing of peptides and proteins: Polypeptides or proteins play significant roles in biological functioning. Even though the nanopore-based sensing of proteins or peptides is still a nascent technology, it presently can achieve the detection of the proteins or peptides at the single-molecule level.

In the single pore sensing of polypeptides, the short peptides have the possibility to translocate through the nanopore channel to produce current blockage. However, due to the relatively small size of the nanopore, larger sized polypeptides or proteins cannot be translocated. In this situation, a current blockage may occur through nanopore conformational changes caused by specific interactions between the larger-sized proteins and the connector. Therefore, the dwell time indicates either the time required for one complete translocation event or the interaction time between the nanopore and the analytes. As mentioned earlier, as peptides are driven to be translocated or to interact with the connector by electrical force, the intrinsic properties of the polypeptide such as length and charge, type and modification of the connector greatly influence the current blockage and dwell time. Researchers have been able to discriminate peptides with different compositions and lengths from the nanopore technique experimentally. For example, Fig. 7 shows the differences in the current signatures to discriminate peptides with same lengths but different compositions. Ji and coworkers were also able to differentiate among peptides with different numbers of arginine residues using T7 nanopore technology. Fig. 8 compares the current blockage and dwell time to distinguish peptides with the number of arginine residues ranging from 8 to 12. Like other biomolecules, longer peptides generated larger current blockage and longer dwell time. Moreover, the peptides with same length and composition can also be detected using the peptide digestion assay. The main idea for the detection is that peptides with various lengths will be created after specific enzyme digestion. Furthermore, the phi29 connector was inserted into the stable polymer membrane in Oxford MinION flow cell, which

Table 1 Comparison of permanent current blockage events among thioesters

<i>Thioesters: Permanent Binding Events</i>					
<i>Transient binding events</i>		<i>Permanent binding events</i>		<i>Ratio of permanent to transient events</i>	
<i>thioesters</i>	<i>(Cys-X)</i>	<i>(Cys-X + Cys-X)</i>	<i>p value</i>	<i>(Cys-X + Cys-X)/(Cys-X)</i>	<i>p value</i>
ethane	16.4 $\pm$ 2.0%	33.5 $\pm$ 0.5% (N = 63)	<0.001 (thymine); <0.001 (benzene)	2.04 $\pm$ 0.12	<0.001 (thymine); <0.065 (benzene)
thymine	18.9 $\pm$ 2.6%	36.3 $\pm$ 1.2% (N = 44)	<0.001 (ethane); <0.001 (benzene)	1.92 $\pm$ 0.14	<0.001 (ethane); <0.437 (benzene)
benzene	19.5 $\pm$ 6.2%	38.4 $\pm$ 2.0% (N = 66)	<0.001 (ethane); <0.001 (thymine)	1.96 $\pm$ 0.32	<0.065 (ethane); <0.437 (thymine)

Note: Reprinted with permission from Haque, F., Lunn, J., Fang, H., Smithrud, D., Guo, P., 2012. Real-time sensing and discrimination of single chemicals using the channel of phi29 DNA packaging nanomotor. ACS Nano 6 (4), 3251–3261. Copyright (2012) American Chemical Society.

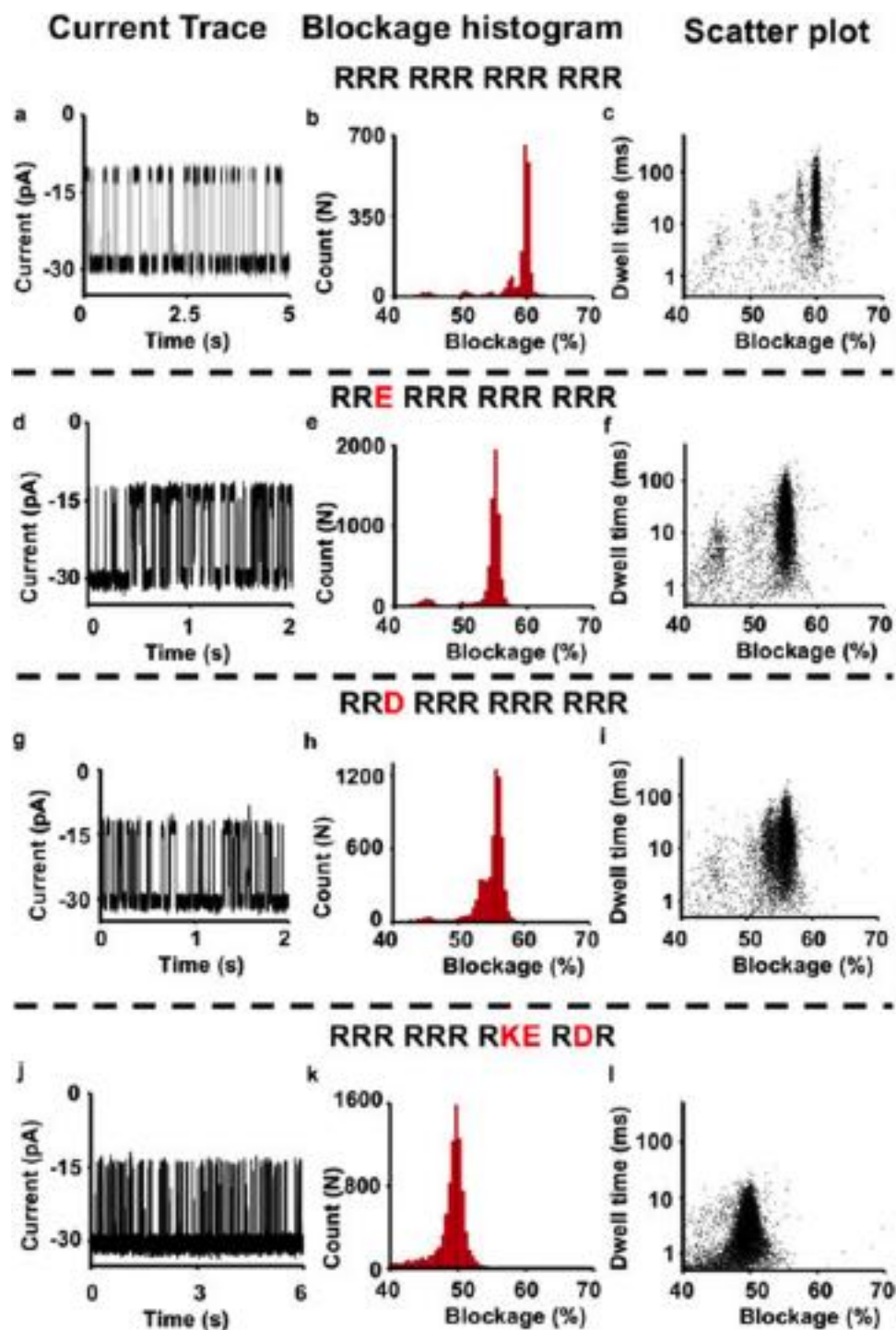


Fig. 7 The identification and differentiation among peptides with same length but distinctive composition. Reprinted with permission from Ji, Z., Guo, P., 2019. Channel from bacterial virus T7 DNA packaging motor for the differentiation of peptides composed of a mixture of acidic and basic amino acids. *Biomaterials* 214, 119222. From Elsevier.

comprises 2048 wells in the portable MinION device for high-throughput peptides sensing. Currently, the TAT, R14, MAR, and P27 peptides were found to be able to generate distinct current blockages signals in the MinION. This innovation has greatly advanced the development in peptides sensing and sequencing. Please refer to the further reading list for more details on the peptides sensing.

Scientists have also developed methods to detect proteins that are too large to pass through a nanopore via the analysis of fingerprinting signals by specific protein-nanopore interactions. For example, Wang and coworkers (Please refer to the further reading list for more details) incorporated an Epithelial Cell Adhesion Molecule (EpCAM) peptide into the C-terminal of phi29 connector channel to

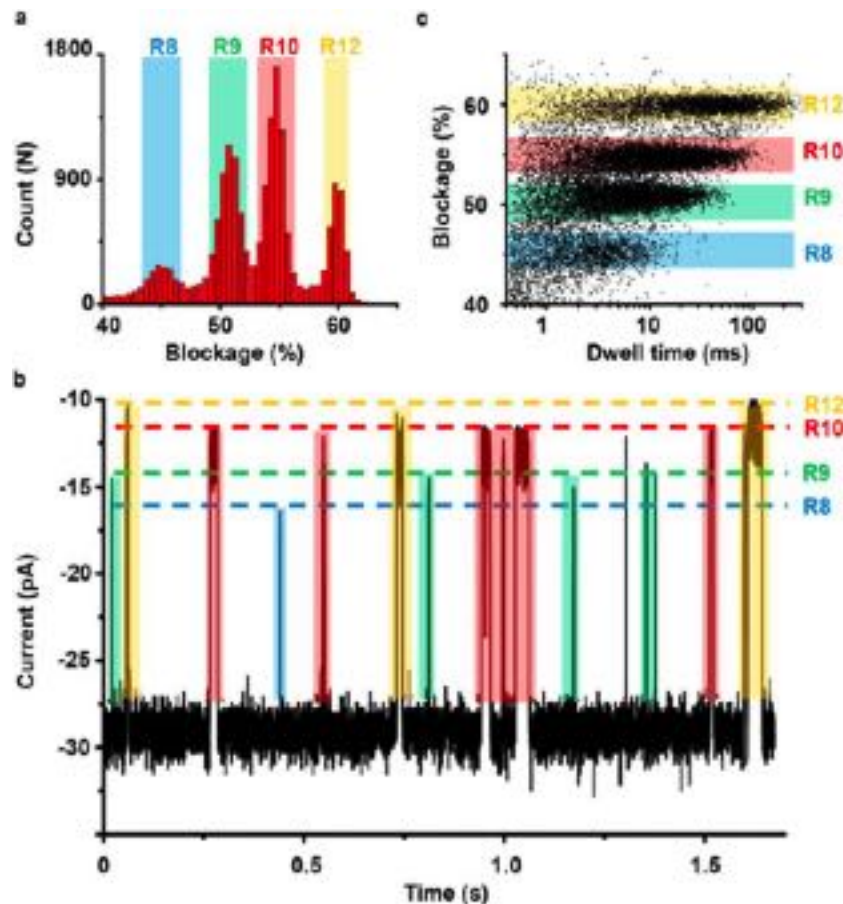


Fig. 8 The differentiation of peptides with different number of arginine residues (R8, R9, R10, and R12). Reprinted with permission from Ji, Z., Kang, X., Wang, S., Guo, P., 2018. Nano-channel of viral DNA packaging motor as single pore to differentiate peptides with single amino acid difference. *Biomaterials* 182, 227–233. From Elsevier.

detect the EpCAM antibody. Upon the specific interaction between the EpCAM peptide and the antibody, it was possible to collect distinct current blockage signals from the background even though many other nonspecific antibody or serum components were present.

Summary

The phage dsDNA-packaging motors have long been thought to be rotation motors. A third type of biomotors recently has been discovered that translocate dsDNA by a revolution mechanism without rotation, and are commonly found in viruses, bacteria, and eukaryotic cells. Analogously, the rotation motion resembles Earth rotating on its own axis every 24 h, but revolution resembles Earth revolving around Sun every 365 days. The revolving motion renders a motor free of friction, coiling, and torque, which has helped clear up many puzzles of viral DNA packaging motor studies over the structure, stoichiometry, and functioning of DNA translocation motors.

In recent years, development and application of biological nanopores have advanced significantly. These biological nanopores can detect different types of molecules, such as DNAs, chemicals, and peptides. The bacteriophage nanopore has proven its ability for discriminating DNAs with a variety of lengths and conformations. These studies indicate likelihood of future applications of the bacteriophage conduits in detection of biomarkers, such as miRNAs, which can be achieved by the conformational differences in the absence and presence of target analytes. Regarding chemical sensing, specific modifications on the nanopore channel can generate unique signals for individual chemicals. Additionally, the biological nanopores can distinguish among short peptides. The analysis of the product peptides with the use of enzymes also shows the evidence of the nanopore system to differentiate among proteins of similar sizes. This differentiation further indicates the potential to detect specific enzymes by the nanopore system. Overall, based on their ability to provide quantitative analyses of molecules of many kinds, nanopores have great potential for disease diagnosis.

The biological nanopores can be used in versatile applications, though several challenges remain. The type and number of analytes that can be discriminated by single pore sensing are inadequate. Different analyte detection will likely require the relevant modification of the nano-channel with distinct properties and specifically designed probes. For example, the detection of certain

chemicals may require the specific types and modifications of the nanopore for specific binding between the analytes and the connector. In addition, the detection or sequencing of proteins remains challenging. Further studies are necessary to build the database for peptide discrimination or sequencing. Moreover, to transfer the nanopore sensing technique to the clinical trials, additional knowledge on how to enhance the sensitivity and specificity in clinical samples or in the presence of many impurities are necessary to be revealed. In sum, the single pore sensing is becoming a powerful analytical tool, but still requires further improvement and exploration of various connectors.

Acknowledgments

We thank Todd Sukany for copy editing and proofreading of this article.

Further Reading

- Bazin, C., King, J., 1985. The DNA translocating vertex of dsDNA bacteriophage. *Annual Review of Microbiology* 39, 109–129.
- Besprozvannaya, M., Burton, B.M., 2014. Do the same traffic rules apply? Directional chromosome segregation by SpoIIIE and FtsK. *Molecular Microbiology* 4, 599–608.
- Cuervo, A., Carrascosa, J.L., 2012. Viral connectors for DNA encapsulation. *Current Opinion in Biotechnology* 23 (4), 529–536.
- De-Donatis, G.M., Zhao, Z., Wang, S., *et al.*, 2014. Finding of widespread viral and bacterial revolution dsDNA translocation motors distinct from rotation motors by channel chirality and size. *Cell and Bioscience* 4, 30.
- Dedeo, C.L., Cingolani, G., Teschke, C.M., 2019. Portal protein: The orchestrator of capsid assembly for the dsDNA tailed bacteriophages and herpesviruses. *Annual Review of Virology* 6, 141–160.
- Erika, J., Tuma, R., 2012. Mechanism of RNA packaging motor. *Advances in Experimental Medicine and Biology* 726, 609–629.
- Geng, J., Wang, S., Fang, H., Guo, P., 2013. Channel size conversion of phi29 DNA-packaging nanomotor for discrimination of single- and double-stranded nucleic acids. *ACS Nano* 7, 3315–3323.
- Guo, P., Driver, D., Zhao, Z., *et al.*, 2019. Controlling the revolving and rotating motion direction of asymmetric hexameric nanomotor by arginine finger and channel chirality. *ACS Nano* 13, 6207–6223.
- Guo, P., Erickson, S., Anderson, D., 1987b. A small viral RNA is required for in vitro packaging of bacteriophage phi29 DNA. *Science* 236, 690–694.
- Guo, P., Grainger, I., Zhao, Z., Vieweger, M., 2014. Two classes of nucleic acid translocation motors: Rotation and revolution without rotation. *Cell and Bioscience* 4, 54.
- Guo, P., Grimes, S., Anderson, D., 1986. A defined system for in vitro packaging of DNA-gp3 of the *Bacillus subtilis* bacteriophage phi29. *Proceedings of the National Academy of Sciences of the United States of America* 83, 3505–3509.
- Guo, P., Peterson, C., Anderson, D., 1987c. Prohead and DNA-gp3-dependent ATPase activity of the DNA packaging protein gp16 of bacteriophage phi29. *Journal of Molecular Biology* 197, 229–236.
- Guo, P., Zhang, C., Chen, C., Garver, K., Trottier, M., 1998. Inter-RNA interaction of phage phi29 pRNA to form a hexameric complex for viral DNA transportation. *Molecular Cell* 2 (1), 149–155.
- Guo, P., 2010. The emerging field of RNA nanotechnology. *Nature Nanotechnology* 5, 833.
- Haque, F., Lunn, J., Fang, H., Smithrud, D., Guo, P., 2012. Real-time sensing and discrimination of single chemicals using the channel of phi29 DNA packaging nanomotor. *ACS Nano* 6 (4), 3251–3261.
- Haque, F., Wang, S., Stites, C., *et al.*, 2015. Single pore translocation of folded, double-stranded, and tetra-stranded DNA through channel of bacteriophage phi29 DNA packaging motor. *Biomaterials* 53, 744–752.
- Ji, Z., Kang, X., Wang, S., Guo, P., 2018. Nano-channel of viral DNA packaging motor as single pore to differentiate peptides with single amino acid difference. *Biomaterials* 182, 227–233.
- Ji, Z., Jing, P., Jordan, M., Jayasinghe, L., 2020. Insertion of channel of phi29 DNA packaging motor into polymer membrane for high-throughput sensing. *Nanomedicine: NBM* 25, 102170.
- Jing, P., Haque, F., Shu, D., Montemagno, C., Guo, P., 2010. One-way traffic of a viral motor channel for double-stranded DNA translocation. *Nano Letters* 10 (9), 3620–3627.
- Maluf, N.K., Feiss, M., 2006. Virus DNA translocation: Progress towards a first ascent of mount pretty difficult. *Molecular Microbiology* 61 (1), 1–4.
- Parent, K.N., Schrad, J.R., Cingolani, G., 2018. Breaking symmetry in viral icosahedral capsids as seen through the lenses of X-ray crystallography and cryo-electron microscopy. *Viruses* 10 (2), 67.
- Pi, F., Vieweger, M., Zhao, Z., Wang, S., Guo, P., 2016. Discovery of a new method for potent drug development using power function of stoichiometry of homomeric biocomplexes or biological nanomotors. *Expert Opinion on Drug Delivery* 13 (1), 23–36.
- Schwartz, C., De Donatis, G.M., Zhang, H., Fang, H., Guo, P., 2013. Revolution rather than rotation of AAA + hexameric phi29 nanomotor for viral dsDNA packaging without coiling. *Virology* 443 (1), 28–39.
- Shu, Y., Haque, F., Shu, D., *et al.*, 2013. Fabrication of 14 different RNA nanoparticles for specific tumor targeting without accumulation in normal organs. *RNA* 19, 766–777.
- Wendell, D., Jing, P., Geng, J., *et al.*, 2009. Translocation of double stranded DNA through membrane adapted phi29 motor protein nanopore. *Nature Nanotechnology* 4 (11), 765–772.
- Yang, Y., Yang, P., Wang, N., *et al.*, 2020. Architecture of the herpesvirus genome-packaging complex and implications for DNA translocation. *Protein Cell*. doi:10.1007/s13238-020-00710-0.

General Ecology of Bacteriophages

Stephen T Abedon, The Ohio State University, Mansfield, OH, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Burst Release of phages from a bacterium, usually upon lysis. Burst size is the number of phage virions released per phage-infected bacterium.

Chronic release The continuous release of phage progeny from infected bacteria in a manner that does not involve bacterial death, infection loss, or host lysis.

Epifluorescent microscopy Microscopic observation of light absorbed and then reemitted from specimen-bound dyes.

Fecundity Measure of the number of offspring produced by an individual.

Latent period The time beginning with phage adsorption to a bacterium and ending with phage release from a bacterium.

Metagenomics The isolation and sequencing of genomic nucleic acid from whole communities including bacteriophage communities such as those found within a sample volume of seawater.

Plaque Localized region of phage population growth occurring in association with a turbid, spatially constrained bacterial population (bacterial lawn), as resulting in localized clearing of that turbidity.

Pseudolysogeny A latent phage infection in which penetration of the phage genome into the bacterial cytoplasm has occurred but subsequent steps towards either productive infection or lysogeny are delayed.

Pseudolysogeny is thought to occur primarily when bacteria that are in a significantly starved state are infected by a phage.

Single-step growth A technique involving adsorption of phages to bacteria that is followed by a determination of the timing of phage-progeny release from the now-infected bacteria, as well as determination of the total number of phages released per bacterium. The word 'single' (or 'one') is used to indicate that the released phages are prevented from infecting or even adsorbing to subsequent bacteria found within the experimental vessel.

Spatial structure Environmental impediments to mixing, diffusion, and organism motility such that the present spatial position of an organism serves as a good predictor of future spatial position. In the microbiology laboratory, spatial structure is often imposed upon cultures via the addition of agar to growth media, e.g., such as may be found within a Petri dish.

Introduction

Bacteriophages are viruses of bacteria. An enormous number of bacteriophages are found the world over, on the order of 10^{30} or more. By way of comparison, 10^{30} is approximately the weight of the Sun, in pounds. This article presents an overview of the interactions between bacterial viruses and their environments: that is, an overview of phage ecology. Phage ecology is an exploration of how bacteriophages exist in the wild, and of their impact there.

Phage Existence Within Environments

Phage populations have been documented in more or less all environments within which bacteria have been documented. Phage ecological characterization begins with efforts to determine what phages are present within specific environments and in what numbers. With the availability of modern molecular, microscopic, flow-cytometric, and sequencing techniques, it is relatively easy to obtain direct counts or to molecularly characterize uncultured phage virions obtained from environmental samples. A variety of techniques may also be employed to study viral infection/production (aka, proliferation) within environments, though none of these methods are ideal. Estimations of lysogeny in environments are possible but similarly are not ideal.

Overview

Determination of phage numbers by viable count, typically meaning plaque count (Fig. 1), once dominated phage environmental endeavors, particularly prior to the 1990s. Direct microscopic counts, however, are generally much higher than viable counts, for example by as much as 10,000-fold. For isolation of viable phages from environmental samples, selective enrichment techniques are often employed. Representative phage numbers, but more generally *virus* numbers determined by direct counts range from 1×10^4 to 2×10^7 per ml of seawater. Assuming virion degradation between sampling and enumeration in the above estimations, these numbers are calculated to be as high as, e.g., $3\text{--}100 \times 10^6$ per ml. Sediment and terrestrial virus counts can be even higher than numbers as found in open water.

Especially using epifluorescence microscopy, phage direct counts are accomplished more easily than either phenotypically characterizing environmental phages or demonstrating phage environmental impact on specific bacteria. Various molecular

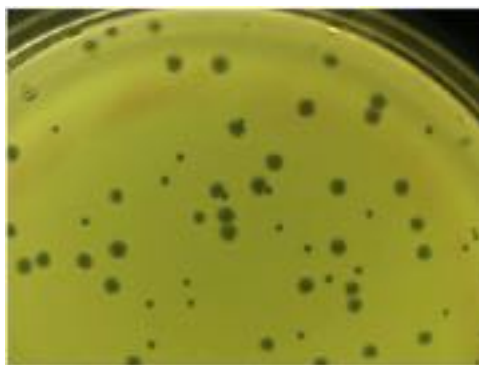


Fig. 1 Bacteriophage plaques. Shown are two strains of phage T4 as grown on the host bacterium, *Escherichia coli*. The larger plaques are made by an *r* mutant, which can display a significantly shorter latent period than the smaller-plaques wild type. Note the abruptness of the ‘collisions’ between *r*-type and wild-type plaques, which is suggestive that phage infections likely extend here beyond the visible portion of the wild-type plaques, an interesting laboratory-ecological phenomenon which was first observed in the late 1940s.

approaches – including pulsed-field gel electrophoresis and metagenomic sequencing of phage environmental populations – allow characterization of phage environmental diversity, especially in terms of genotype. While electron microscopy of environmental samples does allow some phenotypic characterization, particularly of virion morphology, none of these approaches are substitutes for phenotypic analysis of individual phage isolates found in at least ‘phage’ pure cultures. In most cases, however, such characterization is not easily accomplished because the majority of presumptive phage hosts are not easily cultured in the laboratory. Phage culturing and microscopy are considered in greater details as follows.

Phage Isolation and Host Range

Phage presence as well as host range diversity can be assessed by determining phage viable counts. Typically this involves plaquing (Fig. 1). Though rarely employed, it is also possible to dilute phages suspended in liquid such that their isolation and even enumeration, using the most probable number method, can be achieved without plaquing. While directly plating environmental samples can be desirable, phage densities are not always sufficient for either isolation or quantification. Two general approaches exist for increasing phage numbers associated with environmental samples: selective phage enrichment and phage concentration. Selective enrichment involves incubating an environmental sample in liquid culture supplemented with a specific bacterial host. The result is a selective increase in the number of those phages which are able to productively infect that host. Concentration of environmental phages requires more sophisticated techniques and equipment, such as filtration, high-speed centrifugation, or precipitation, but allows quantitative determinations of even low environmental densities of phages. Unfortunately, all handling steps involved in phage isolation are likely to result in enrichment of phages possessing certain properties, such as phages which replicate more effectively under enrichment conditions, which have virions that are better able to resist physical damage, or which produce more-easily noticed plaques.

Even when carefully employed, viable counts, as noted, tend to grossly underestimate total phage numbers. Aside from phage losses due to handling, these underestimations can stem from the relatively narrow host ranges of phages that allow only a small portion of a phage community to infect or form plaques on a given bacterial host. Also, it is uncertain what fraction of total phage virions are viable under even ideal laboratory conditions nor how much laboratory environments deviate from each phage’s ideal growth environment. Furthermore, also as noted, not all bacteria are culturable; in fact, the vast majority of bacterial species currently are either nonculturable or, at least, difficult to culture using standard laboratory procedures. The resulting limitations on testing phage samples against most hosts necessarily limit our understanding of phage host range diversity within environments.

Microscopic Determination

Historically, microscopic determinations have played important roles in the study of phage environmental microbiology, since it was through these observations, especially in the late 1980s and early 1990s, that aquatic environments were found to harbor vast numbers of virus-like particles, especially phages. These observations made it obvious that phage-bacterial interactions could occur within environments with great regularity, suggesting that phages could enormously impact aquatic bacteria.

Early visualization of phages involved almost exclusively electron microscopes, since the small size of phages causes them to appear, at best, as pinpricks of light using dark-field optical microscopy. Furthermore, electron microscopes allow scrutiny of phage morphology, which is helpful in differentiating virus-like particles from non-virus particles of similar size, and in obtaining a first approximation of virion *in situ* diversity. Unfortunately, electron microscopy is fallible because viruses that are not obviously (or exclusively) virus-like in their morphologies are less likely to be enumerated, and thereby may be excluded from viral total

Table 1 Common metrics of phage ecological, and universal, abundance

Measure	Quantity
Approximate number of phage hosts worldwide	10^{30}
Typical ratio of phage to bacteria in environments	10:1
Approximate average phage density in sea water	10^7 ml^{-1}
Typical range of phage densities in aquatic environments	$10^4\text{--}10^8 \text{ ml}^{-1}$
Approximate phage density in soils	10^8 g^{-1}
Approximate phage density in sediments	10^{10} ml^{-1}
Total volume of phage habitats (mostly ocean)	$> 10^{24} \text{ ml}$
Estimated worldwide number of virus particles ^a	10^{31}
Estimated total phage mass assuming 10^8 Da/virion	10^{15} g
Approximate phage mass in 'blue-whale units' (BWUs = 10^8 g)	10^7 BWUs
Cumulative length of phage particles lain end to end	10^8 light years

^aIt is typically assumed that most virus particles observed in the wild are phages.

counts. An additional problem with total counts is that they can determine neither the viability nor the host range of the observed viruses. This includes uncertainty about what fraction of the observed viruses are indeed phages as it is especially difficult to distinguish tailless phages from various groups of eukaryotic viruses. For this reason, researchers who use culture-independent microscopic techniques tend to classify all observations as 'viruses-like particles' and not necessarily as 'phages'. Because of virion morphology and the relative prevalence of bacterial hosts in environments, it is assumed that the majority of observed virions in most environments never the less are phages.

Alternatively for determining virion total counts is epifluorescence microscopy, a technique requiring less-sophisticated equipment and sample preparation than electron microscopy. In epifluorescence, virus-containing samples are stained with nucleic acid-binding fluorescent dyes that allow viruses to be viewed using a fluorescent microscope (again as pinpricks of light, though now pinpricks of illuminated viral genomes). Both electron microscopy and epifluorescence microscopy, like viable counts, often require sample concentrating prior to enumeration. In general, electron microscopy allows determination of virion morphologies while epifluorescence allows differentiation of entities containing nucleic acid from those lacking it.

As these various forms of microscopy and sample preparation have been refined and developed over the past 30 or so years, more accurate determinations of environmental phage numbers have become possible, and estimates of $\sim 10^6$ or more phages per milliliter are typical in pelagic aquatic environments (i.e., within water columns). In marine environments, phage densities appear to be greater in coastal waters (e.g., $\sim 10^7 \text{ ml}^{-1}$ or more) versus more off-shore waters (e.g., $\sim 10^6 \text{ ml}^{-1}$ or more). Densities may be in ranges closer to $\sim 10^8 \text{ g}^{-1}$ for terrestrial environments (i.e., soils) as well as productive lakes and even higher still (up to 10^{10} g^{-1}) in sediments, though note that in general study of the ecology of viruses in soils, or sediments, is more difficult than its study in more fluid environments. Although densities tend to decline in deeper sediments, it is speculated that shallow sediments could serve as reservoirs for pelagic viruses, offering long-term, perhaps many decades, phage storage such that sediment disruption may return phages to overlying waters, serving to a degree as equivalents to plant soil 'seed banks'. Using these methods, it has been estimated that at least 10^{30} phage particles, and perhaps as many as 10^{31} or more, are present on the planet at any given time, where the lower figure is equal to approximately the volume of all the world's oceans, in milliliters, times 10^6 (Table 1).

Phage Ecology

The science of ecology may be differentiated into a number of more-or-less distinct subdivisions. These include organismal, population, community, ecosystem, and landscape ecologies. Also included, especially under the heading of organismal ecology, are physiological and evolutionary ecologies. Though important to an understanding of ecology in general, in most cases little effort is made within the phage ecology literature to explicitly distinguish among these various categories, plus there exist numerous overlaps between categories. Indeed, phage biologists including this author often are trained first as microbiologists, or as molecular geneticists, etc., and only later or much less so as ecologists. As a consequence, phage publications relevant to our understanding of phage ecology typically will not describe their contents from a classically ecological perspective. A major goal of this section, therefore, will be to expose readers to phage ecological thinking as organized within a more general ecological framework, presented in overview as follows.

Overview

What can be described as a phage *organismal ecology* considers phage adaptations within environments, with emphasis on phage survival, reproduction, and dissemination. Included under this heading can be phage physiological ecology, which considers the impact of environments on phenotypes. For phages this would be manifest either through modifications in the physiology of bacterial hosts or in terms of environmental impact on virion properties. Also included under the heading of phage organismal



Fig. 2 Lytic bacteriophage life history. Images © James A. Sullivan, used with permission.

ecology is phage evolutionary ecology, which considers the selective benefits of phage adaptations, i.e., how adaptations impact phage evolutionary fitness (survival and growth rates).

Phage *population ecology* considers phage intraspecific interactions, particularly in terms of how those interactions control phage population size, dispersion, and growth rates. Phage *community ecology* considers phage interactions with other species, especially with bacteria but also with other types of phages. Also are the biotic components of the greater bacterial environment, such as in terms of the impact of phage-encoded exotoxins on the eukaryotic hosts of bacteria. Phage *ecosystem ecology* considers phage impact on nutrient cycling and energy flow within ecosystems, which by necessity represents a much more abiotic emphasis than other approaches. Phage *landscape ecology*, which has not been substantially considered explicitly within the phage literature, would consider the phage impact on interactions between distinct ecosystems, such as between a pond and a surrounding forest, a sewage treatment plant and downstream environments, or an animal (serving as an ecosystem itself) and the extra-animal environment.

Note that these various means of looking at phage ecology build upon one another. That is, ecosystem ecology builds upon community ecology which builds upon population ecology which builds upon organismal ecology which, finally, is understood best in light of a strong molecular appreciation of phage biology.

Phage Organismal Ecology

Phage organismal ecology may be considered from the perspective of environmental impacts on phage 'growth' parameters, particularly as affecting phage reproduction, dissemination, and survival. More specifically are such things as phage adsorption rates, infection timing, productive-infection versus lysogeny decisions, burst sizes, and virion decay rates. Since our understanding of phage growth parameters is derived from the laboratory characterization of phages, this section describes this aspect of phage ecology by employing a very much laboratory-centered emphasis.

The traditional means of studying most phage growth parameters are single-step growth experiments and phage adsorption determinations. We can fit these growth parameters into a generalized scheme of the phage life cycle, and then consider how variation in the assorted steps of this life cycle can impact phage ecological success. These steps, in order, include (**Fig. 2**) (1) a diffusion-driven phage-virion extracellular search for bacteria to infect, (2) the phage adsorption process during which a relatively inert phage-virion and adsorption-susceptible bacterium are together converted into a phage-infected bacterium, (3) the phage infection which can vary in its phenotypic expression from restricted (phage dies, bacterium lives) to abortive (both phage and bacterium die) to productive (phage lives but bacterium dies) to pseudolysogenic (or 'carrier state'; especially, a relatively unstable latent infection) to lysogenic (a somewhat stable latent infection), and (4) some means of phage-virion release into the extracellular environment. The latter, depending on the phage, may or may not involve host-cell and infection death (i.e., via lysis versus chronic/continuous release, respectively).

Consideration of how variation in these assorted steps can impact phage ecological success is presented in the following sections. For example, we can expect rapid virion adsorption – particularly rapid phage attachment given encounter with a phage-susceptible bacterium – to be favored over less-rapid adsorption unless phenotypes conferring rapid adsorption interfere with phage survival or fecundity. For instance, T4-like phages may enhance their survival within extra-colonic environments by taking on a temporarily adsorption-incompetent but inactivation-resistant state. Another counter to the evolution of more-rapid phage adsorption/attachment may be phage growth within relatively spatially structured environments, if there exist tradeoffs between time spent disseminating versus time spent infecting bacteria. Note that these considerations of phage adsorption rates more generally reflect differences between phage properties 'as virions' versus 'as during bacterial infection', differences which are explored more fully in the following section.

Phage infections in the wild also can be less efficient than those studied under laboratory 'ideal' conditions for phage growth (e.g., smaller burst size, longer latent period, slower adsorption). The major bacterial nutrients, carbon and energy, as well as critical minerals, such as phosphorus, can be in short supply within many environments, resulting in such inefficiencies in nature. The ecological study of organismal physiological response to changing environmental conditions is termed physiological ecology, or ecophysiology in short. In addition to developing a better general understanding of phage *in situ* population ecology,

consideration of phage ecophysiology is relevant to estimating levels of phage-induced bacterial mortality within natural environments, an important aspect of *in situ* phage–bacterial community ecology.

Bacteria-like versus virion-emphasizing modes of existence

The selective advantage associated with various phage infection and release strategies, similar to phage adsorption rates, will be dependent on environmental conditions, particularly in terms of numbers of bacteria, bacterial physiology, and rates of phage decay. In general we can assume that conditions that favor bacterial growth and survival over phage population growth and survival will tend to favor more bacteria-like modes of phage replication/survival over more production of phage virions-emphasizing modes; more precisely, we can expect life history biases towards phage existence infecting a bacterium rather than towards greater emphasis on existence as free virions. Thus, low bacterial densities, bacterial physiologies that poorly support significant virion production, or virion-specific environmental antagonists including phage-restricting bacteria will tend to favor extended phage latent periods. These infecting states may include pseudolysogeny, lysogeny, or productive but longer-lasting infections, including infections by phages that release their progeny chronically rather than lytically. Alternatively, we can assume that conditions that favor phage population growth will tend to favor more productive modes of replication, that is, ones that emphasize the virion state over the infected state. This more virion-emphasizing state can include shorter-lengthed productive phage infections – especially lytic infections that favor earlier virion release over higher infection fecundity – but also more-rapid chronic virion release at the expense of longer-term infection survival.

The advantage of displaying a more virion-emphasizing existence may be viewed from two perspectives. The first are conditions that allow for rapid phage-population expansion, i.e., relatively high densities of uninfected host bacteria. Less intuitively, that expansion comes at the expense of bacterial survival, and therefore any phage which persists within a more bacteria-like (extended latent period) mode of infection may be more susceptible to attack by unrelated phages under conditions which are favorable toward phage population expansion (see, e.g., ‘Killing the winner’, below). Also favoring the virion over the infecting state, a temperate phage can increase its target number and reduce its genomic target size via lysogen induction, i.e., by changing from a single bacterial lysogen to multiple copies of phage virions. Such induction can occur following bacterial lysogen exposure to DNA damaging agents. In considering phage population growth and survival, we thus can envisage phage adaptation as involving compromises between conflicting tendencies toward greater emphasis on existence as an infection versus greater emphasis on existence as a virion.

Phage Population Ecology

Considerations of phage population ecology may be differentiated into two categories: (1) the impact of phage adaptations and environmental conditions on phage population growth within well-mixed, fluid environments (such as within broth in the laboratory) and (2) the impact of phage adaptations and environmental conditions on phage population growth within spatially structured environments (such as within soft-agar overlays in the laboratory; [Fig. 1](#)). These categories will be considered in turn. Like phage organismal ecology, much of our understanding of phage population ecology derives from laboratory characterization of phages, but also from exercises in mathematical modeling of phage population growth. We can also distinguish phage population considerations into those operating at low phage densities versus those operating at high phage densities. In both cases density refers to that of a single phage population since inter- as opposed to intra-species interactions are considered under the province of phage community rather than phage population ecology. These latter considerations are much simpler to envisage during phage broth-culture growth.

The growth of phages in broth can be considered to occur over at least four distinct steps. These are as follows:

- (1) Phage entrance into an environment. This can be physical, in terms of phage movement from one place to another, physiological in terms of the induction of an already-present lysogen, or genetical in terms of phage host-range mutation such that a previously phage-insusceptible bacterial population suddenly becomes susceptible to sympatric phages.
- (2) A period of phage-population growth that spans from the point of the initial phage infection of a susceptible bacterium within an environment through to the point of transition where most susceptible bacteria have become phage infected. Most bacteria will become phage infected, however, only if sufficient bacteria are present to support phage population growth to relatively high densities.
- (3) The transition to and then span over which most of the phage-susceptible bacteria within a population have become phage infected.
- (4) A postinfection, particularly post community-wide bacterial lysis period during which phage numbers greatly outnumber those of bacteria.

In many cases, each of these steps is at best a transient phenomenon for a given combination of phage, bacterium, and environment.

Once a phage virion has been released from an infected bacterium, it has entered into a period of what can be described as an extracellular search for new bacteria to infect. The duration of this search will be a function of the product of the density of phage susceptible bacteria within a given environment – which in turn can be a function of the range of bacterial types that a given phage may infect – and the phage adsorption rate constant, which will vary as a function of phage, bacterium, and environmental parameters. The likelihood that a given phage will successfully adsorb to and infect some bacterium is additionally a function of a phage’s decay rate, a.k.a., virion inactivation rate, with higher decay rates reducing the likelihood of successful infection by a virion. Decay can be a function of abiotic insults that result in virion capsid or nucleic acid damage, phage adsorption to abiotic materials,

biotic insults including engulfment by eukaryotes as well as inactivation mediated by phage-adsorbed bacteria, or, from a modeling perspective, even emigration out of an environment. To a phage entering a previously phage-free environment, the likelihood that phage population growth will occur therefore is a function of bacterial numbers, rates of phage adsorption per bacterium, and rates of phage decay.

If successful adsorption occurs, then the likelihood that phage population growth will continue beyond this first adsorption will be a function of phage burst size, with each additional phage produced per infection potentially possessing an additional, nearly identical likelihood of subsequent successful bacterial adsorption. Extending this point, we can assume that with either higher bacterial numbers, more-effective/rapid virion adsorption per bacterium, lower rates of decay, or higher initial virion numbers, then there will be a greater likelihood that phage population growth will be initiated upon phage entrance into a new environment. Overall the potential for a phage population to grow within a given environment can be described as its existence conditions.

Once phage population growth has begun, it seems unlikely that there would exist any intrinsic inhibitions on phage evolution toward faster population growth. Both in terms of intraspecific and inter-specific competition among phages, it is that phage which reaches a bacterium first that should have the greatest potential to 'claim' that bacterium for infection. Thus, a phage type that grows its population fastest will have access to more bacteria sooner, and therefore enjoy among its progeny greater overall numbers of bacterial infections within a given environment. In general, we can expect faster overall phage population growth given greater phage burst sizes, shorter phage latent periods, or faster phage adsorption. Phage decay can also affect rates of phage population growth, and this occurs via reductions in effective phage burst sizes, where the qualifier 'effective' refers to the number of phages released from a given bacterial cell which go on to successfully infect new bacteria.

Things change, within a given environment, as a phage population reaches the limits of its population growth. At this point the selective pressures on phage growth change from ones that are mostly phage-density independent to ones which instead are highly phage-density dependent. These phage-density-dependent limits vary in their relevance, however, depending on bacterial densities along with phage fecundities. At one extreme, bacterial densities are so low, or phage fecundities (effective burst sizes) so small, that phage populations never reach densities of sufficient size to greatly affect bacterial populations. Under these conditions, multiple phage adsorptions to a given bacterium is of low likelihood and competition between phages for individual bacteria small. Selection here should be biased toward greater virion or infection durability, for example, rather than higher rates of phage population growth. We have this expectation because bacterial rarity presumably would put a premium on phage survival prior to bacterium encounter and also because phage potential to migrate to new bacteria-containing environments could be a function of longer-term phage survival, either as virions or as bacterial infections.

At the other extreme, bacterial densities as well as phage fecundities may be sufficiently high that a majority of phage-susceptible bacteria may become phage infected and therefore no longer available for further infection by the same phage type. Prior to this point we may expect selection to favor phages that display rapid population growth, even if such rapidity should come at the expense of effective burst size, such as in terms of actual burst size or in terms of virion resistance to decay. Thus, at higher bacterial densities, while phage densities are still relatively low, a phage could display faster population growth by shortening its latent period. Such shortening, however, is most easily accomplished at the expense of the overall duration of the period during which phage progeny intracellularly mature, resulting in a reduced phage burst size. Alternatively, one could envisage a loss of phage genes that otherwise could contribute to long-term virion survival should the presence of those genes come at the expense of effective phage burst size or further latent period shortening. One possible example of such evolution could be a common loss of abilities to display lysogenic cycles among phages contaminating industrial dairy ferments.

Adaptation that results in declines in phage survival could be shortsighted within environments in which densities of new bacteria to infect come to rapidly decline. Instead, we can envisage an advantage bestowed upon phage infections during 'final rounds' of infection – that is, at the point where most bacteria within a population have become phage infected – that would result from larger effective burst sizes. In this way more or more robust phages are produced to better assure phage survival until access to a new bacterial population occurs. Somewhat equivalently, a phage may accrue advantages by displaying more bacteria-like lifestyle under conditions where bacterial densities are in decline, such as via temperate-phage reduction to lysogeny upon bacterial adsorption rather than display of a lytic infection (above).

Many of these same effects take place during phage population growth within spatially structured environments, such as within bacterial lawns immobilized within soft agar, that is, as phage plaques (Fig. 1) or, as we can speculate, in association with bacterial biofilms. A difference, though, is that in addition to temporal differentiation between phage density-independent and phage density-dependent selection, there also can be spatial differences. Thus, phage multiplicity may be higher toward the center of plaques versus toward their periphery. In other words, at a plaque's leading edge, selection acting upon phage characteristics likely is more phage density independent than toward the center of plaques. Therefore, we can expect selection for faster phage population growth to occur at a plaque's periphery, whereas, by contrast, selection at a plaque's center should be biased more toward a greater effective burst size. As noted, it may be possible to extrapolate, to a degree, the characteristics of phage population growth within phage plaques to that in association with at least some bacterial biofilms.

Phage Community Ecology

Phage community ecology includes issues of phage predation on bacteria as well as the establishment of mutualistic-like relationships such as those between prophages and bacteria, forming lysogens. The phage community ecology literature is more

closely aligned with the ecological literature, which is to say that a good deal of phage community ecology has been done by individuals trained first as ecologists who have then taken to phages as model systems rather than by individuals trained first by studying more organismal or more molecular aspects of phage biology. As such, phage community ecology has a much more theoretical basis than either phage organismal or population biologies, including the employment of simulations of phage-bacterial interactions, especially within continuous cultures (chemostats). In addition to considerations of phage impact on bacterial populations, including as can be associated with other organisms such as ourselves, phage community ecology can also consider interactions between different phage species or grazing on phages by protists.

We can distill three general governing principles of phage community ecology, especially with regard to phage impact on bacteria: (1) The greater a phage population density then, all else held constant, the greater the phage impact on a bacterial population, and the faster that phage populations reach these higher densities, then the sooner their impact. (2) The greater the density of phage-sensitive bacteria, then – potential changes in bacterial physiology aside – the greater the population size that phages can reach and the sooner they will grow to those population sizes. And (3) Bacteria can hide in various ways from phages – numerically, genetically, phenotypically, physiologically, spatially – but doing so can come at a cost in terms of, for example, bacterial growth rates.

Real world applications of these concepts include that of ‘Killing the winner’ where it is expected that populations of same or similar bacterial types, which grow to higher densities, will tend to become more susceptible to phage-mediated population crashes. These crashes result solely from bacteria having reached these higher densities and thereby being able to support phage population growth to higher densities. By culling especially more successful bacterial populations, phages may serve as mediators of a frequency-dependent selection among bacterial communities that results in greater bacterial diversity than environments might otherwise sustain.

A second and related real-world consideration is that bacteria within especially eutrophic, i.e., high-nutrient environments, where phage virions can be especially abundant, may display physiologies that represent tradeoffs between rates of bacterial growth and phage-free survival. That is, bacteria may be subject to physiological and other burdens that come with deploying protective measures against potentially infecting phages. Protective measures are diverse and include such things as restriction-modification systems along with CRISPR-Cas systems. Additionally, bacteria may take on more-protective lifestyles, potentially such as within extracellular polysaccharides as found in association with biofilms.

Bacteria may also display reduced versatility in the course of mutating to phage resistance because of loss (temporarily or permanently; fully or partially) of membrane proteins involved, for example, in nutrient transport, proteins which could otherwise serve as phage receptors. That is, the more proteins or other molecules a bacterium displays on its surface, then the more phage types that bacterium may be susceptible to. Because many of these mechanisms of phage resistance can result in reduced bacterial fitness, we should expect that the strength of selection for phage resistance should be directly proportional to phage density within a given environment. Phages, in turn, can display adaptations which allow them to overcome various bacterium-mediated mechanisms of phage resistance. Bacteria can also evolve to mitigate the costs associated with displaying phage resistance.

A third real-world consequence of phage impact on bacteria can result from phage-mediated transduction of bacterial genetic material. Phages do this by two basic mechanisms: generalized transduction in which random pieces of bacterial DNA are incorporated into consequently defective phage virions, and specialized transduction. The latter traditionally involves the incorporation of that bacterial genomic DNA which flanks bacterial-chromosome inserted prophages, and occurs as a consequence of imprecise prophage excision upon induction. Serving essentially as an extension of the concept of specialization transduction is phage encoding of ‘morons’, which involves the incorporation via illegitimate recombination of bacterial genes into various locations within temperate phage genomes. This ‘more DNA’ is then passed on to bacteria in the course of establishment of lysogeny by moron-carrying phages. These genes, along with prophages generally, can impact bacterial phenotype including the fitness of lysogens, potentially positively. These latter mechanisms of transduction serve as a means of transferring genetic material from bacterium to bacterium while also transferring genetic material from bacterium to phage.

Phage Ecosystem Ecology

The primary emphasis of phage ecosystem ecology is on the impact of phage infection of prokaryotes on movement of nutrients and energy from prokaryotes to higher trophic levels. That is, phage infection influences the acquisition of bacteria including cyanobacteria by bacteria-eating eukaryotes. The real-world significance of this emphasis derives from prokaryotes serving as the base of many, especially aquatic ecosystems, and the role that these organisms play in carbon sequestration. That is, within aquatic environments most dissolved organic material becomes available to eukaryote grazers only once bacteria have assimilated it. With phage-induced bacterial lysis, however, a significant fraction, even 25% of the carbon entering these ecosystems photosynthetically, is converted to a dissolved organic state. This creates a ‘viral loop’ that shunts organic carbon away from what is known as the aquatic ‘microbial loop’, which otherwise moves carbon from prokaryotes to grazers, i.e., especially phagotrophic protists to grazer-consuming animals (Fig. 3). Another way of stating the significance of the phage impact on the survival of aquatic bacteria is that it is approximately equivalent to the impact of protist grazing. The impact of phage infection versus protist grazing, however, varies with habitat, with phages displaying a greater impact especially in more extreme, including low-oxygen environments.

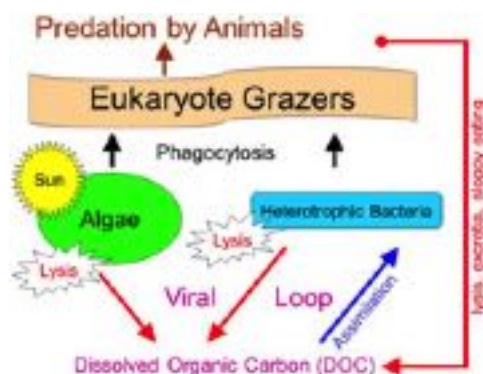


Fig. 3 Microbial loop as short circuited by virus-, particularly phage-induced lysis. The viral loop consists of the generation of dissolved organic carbon via virus-induced lysis of especially heterotrophic bacteria and cyanobacteria (the latter here included under the heading, Algae), which is followed by assimilation of a fraction of that material into heterotrophic bacteria. The viral loop reduces the efficiencies by which photosynthetically fixed organic carbon is transferred to higher trophic levels via the microbial loop, such as to aquatic animals.

Further Reading

- Abedon, S.T., 2006. Phage ecology. In: Calendar, R., Abedon, S.T. (Eds.), *The Bacteriophages*. Oxford: Oxford University Press, pp. 37–64.
- Abedon, S.T., 2007. *Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses*. Cambridge, UK: Cambridge University Press.
- Abedon, S.T., 2009. Phage evolution and ecology. *Advances in Applied Microbiology* 67, 1–45.
- Abedon, S.T., 2011. *Bacteriophages and Biofilms: Ecology, Phage Therapy, Plaques*. Hauppauge, New York: Nova Science Publishers.
- Abedon, S.T., 2017. Commentary: Communication between viruses guides lysis-lysogeny decisions. *Frontiers in Microbiology* 8, 983.
- Abedon, S.T., 2017. Phage “delay” towards enhancing bacterial escape from biofilms: A more comprehensive way of viewing resistance to bacteriophages. *AIMS Microbiology* 3, 186–226.
- Breitbart, M., Rohwer, F., Abedon, S.T., 2005. Phage ecology and bacterial pathogenesis. In: Waldor, M.K., Friedman, D.I., Adhya, S.L. (Eds.), *Phages: Their Role in Bacterial Pathogenesis and Biotechnology*. Washington, DC: ASM Press, pp. 66–91.
- Brüssow, H., Kutter, E., 2005. Phage ecology. In: Kutter, E., Sulakvelidze, A. (Eds.), *Bacteriophages: Biology and Application*. Boca Raton, FL: CRC Press, pp. 129–164.
- Goyal, S.M., Gerba, C.P., Bitton, G., 1987. *Phage Ecology*. Boca Raton, FL: CRC Press.
- Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., Sullivan, M.B., 2017. Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME Journal* 11, 1511–1520.
- Hyman, P., Abedon, S.T., 2010. Bacteriophage host range and bacterial resistance. *Advances in Applied Microbiology* 70, 217–248.
- Hyman, P., Abedon, S.T., 2018. *Viruses of Microorganisms*. Norwich: Caister Academic Press.
- Lenski, R.E., 1988. Dynamics of interactions between bacteria and virulent bacteriophage. *Advances in Microbial Ecology* 10, 1–44.
- Weinbauer, M.G., 2004. Ecology of prokaryotic viruses. *FEMS Microbiological Reviews* 28, 127–181.
- Wommack, K.E., Colwell, R.R., 2000. Virioplankton: Viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews* 64, 69–114.

Relevant Websites

- <http://www.archaealviruses.org>
Archaeal Viruses.
- <http://www.phage.org>
Bacteriophage Ecology Group.
- <http://www.ISVM.org>
International Society for Viruses of Microorganisms.
- https://en.wikipedia.org/wiki/Phage_ecology
Phage Ecology.
- <http://www.phage-therapy.org>
Phage Therapy.

Marine Bacteriophages

Vera Bischoff, Falk Zucker, and Cristina Moraru, Institute for Chemistry and Biology of the Marine Environment, Oldenburg, Germany

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Bathypelagic layer The *bathypelagic layer* extends from the mesopelagic layer to ~ 4000 m below the ocean surface, where the abyssopelagic zone begins. Because no light penetrates in the bathypelagic layer, it is also known as the dark zone.

Epipelagic layer The *epipelagic layer* represents the uppermost layer of the ocean, where sunlight easily penetrates and the photosynthetic autotrophs can be found. It extends from the 0 m to 150/200 m depth.

Fosmid Fosmids are cloning vectors based on the bacterial F-plasmid. They can hold DNA inserts of up to 40 kb in size.

Mesopelagic layer The *mesopelagic layer* represents the part of the pelagic zone in between the photic epipelagic and aphotic bathypelagic zones. It is also known as the twilight zone and it extends from ~ 150/200 m to ~ 1000 m depth. It begins at the depth where only 1% incident light reaches and ends where there is no light.

Nanopore Sequencing Technology Nanopore sequencing technology, as developed by Oxford Nanopore Technologies Inc, produces long reads. The read length usually averages 4 kb, but, with DNA preparations minimizing fragmenting, it can reach even 1 Mb. From here on we will refer to this technology as the *nanopore technology*. It involves several steps. First, the sequencing library is prepared by binding adapters to DNA fragments, while striving to keep the fragment length as high as possible. Then, the library is loaded on a flow cell, which contains more than one thousand protein nanopores set in an electrically resistant polymer membrane. By setting a voltage across the membrane, an ionic current is passed through the nanopore. The protein nanopores can capture single stranded DNA and pass it through. Each time a base passes

through a pore, the electric current is changed and recorded. For each pore, the base sequence of a single DNA fragment passing through it is inferred from the changes in the electrical current.

Read In nucleic acid sequencing, a *read* is defined as the base pair sequence corresponding to a single DNA fragment.

Sequencing by Synthesis Technology Sequencing by synthesis technology, as developed by Illumina Inc., produces short reads of 200–400 bps, in very high throughput. From here we will refer to this technology as the *Illumina technology*. It involves several steps, as follows. First, the sequencing library is prepared by fragmenting the DNA and binding adapters to both ends of the fragments. Then, the library is loaded on a flow cell, where DNA fragments bind randomly to its surface, one fragment at a single location. This is followed by solid-phase amplification, which creates up to 1000 identical copies of each single DNA fragment. All copies from one molecule are found in close proximity, creating a cluster with a diameter of 1 μm or less. In a sequencing flow cell, there are millions of such clusters per square centimeter. Further, in each sequencing cycle deoxynucleotide triphosphates (dNTP) which are fluorescently labeled are added. Because the fluorochrome acts as a terminator, only a single base binds to a DNA molecule per sequencing cycle. After the base incorporation, the fluorescent dyes are imaged and then cleaved away from the DNA molecules, to allow the next cycle of sequencing. Because sequence clusters have a defined spatial localization on the flow cell, the sequence of every single read can be inferred from the order of the fluorescent signals corresponding to the four bases (A, T, C, and G) at the respective location.

A Short Introduction in Marine Phages

In the late 80's it was recognized that, in marine environments, viruses outnumber microbial cells by an order of magnitude. In seawater, virus numbers range in between 10^4 to 10^8 per millimeter. In sediments, virus abundance is even higher, up to 10^9 per cubic centimeter. The majority of these viruses infect bacteria, and they are called bacteriophages, or phages. Phages shape the bacterial community structure through viral lysis and influence both host metabolism through the expression of auxiliary metabolic genes and host evolution through horizontal gene transfer. Furthermore, phages are responsible for the viral shunt, in which organic matter is hijacked by viral lysis from being transferred to higher food chain levels through grazing. Through their various roles, phages are major drivers of marine biogeochemical cycles.

Morphological Diversity of Marine Phages

Generally, phage morphological diversity is low. All phages have a capsid, which is a protein shell surrounding the nucleic acid. Phage capsids are either helical or icosahedral in shape (see [Figs. 1–3](#)). A variation of the icosahedral shape is the prolate capsid, resembling an elongated icosahedron (see [Fig. 1](#)). Helical capsids are known only for ssDNA phages. The icosahedral shape is the

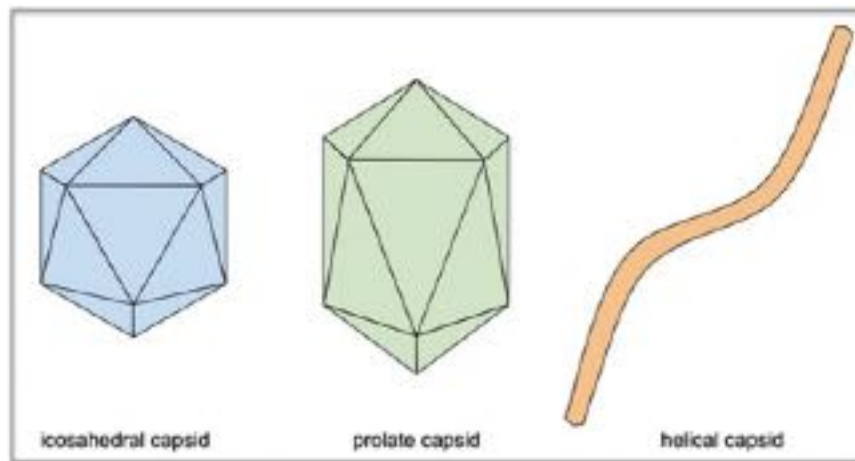


Fig. 1 Phage capsid types.

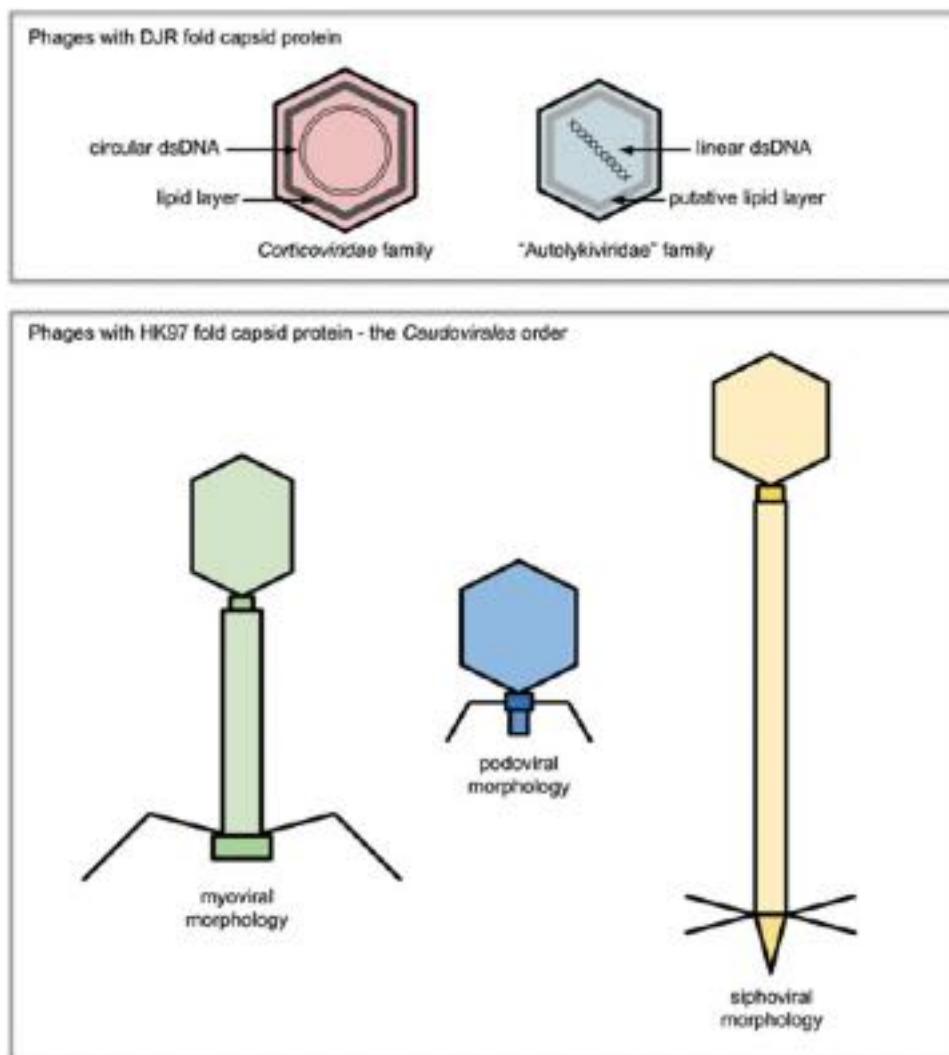


Fig. 2 Morphological and taxonomic diversity of marine dsDNA phages.

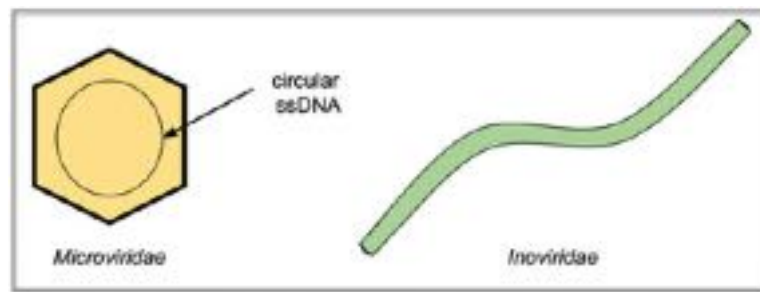


Fig. 3 Morphological and taxonomic diversity of marine ssDNA phages.

most encountered, being present in all dsDNA and RNA phages and in part of the ssDNA phages. In rare instances, icosahedral phages are enveloped by a lipid layer. An additional structure, called tail, is present in the majority of known dsDNA phages. The tail mediates the adsorption to the host and the DNA transfer. There are three basic tail morphologies: podoviral, myoviral and siphoviral (see Fig. 2). In podoviruses, the tail is short and non-contractile. In myoviruses, the tail is long, rigid and contractile. In siphoviruses, the tail is long, flexible and non-contractile. Tailed phages have capsid proteins with an HK97 fold (named after the eponymous bacteriophage), while non-tailed phages have capsid proteins with a double jelly roll fold (DJR).

Because the majority of cultivated marine phages are tailed, and thus belong to *Caudovirales*, until recently it was believed that tailed phages dominate in the environment as well. However, it was shown by quantitative electron microscopy that non-tailed phages make up in between 66%–85% of environmental phages in global surface oceans. The identity and phylogenetic affiliation of these non-tailed phages is currently unknown. Part of them could be dsDNA phage with a DJR-fold capsid, from the *Corticoviridae* or “*Autolykiviridae*” families. Another part could be ssDNA phages from the *Microviridae* family. And another part most likely represents so far unknown phages.

As with their bacterial hosts, morphology alone cannot be used to classify phages or to predict their behavior and ecological role. Behind the few morphological types described, a huge genomic diversity is hidden.

Genomic Diversity of Marine Phages and How do we Study it

The marine phage diversity is large and just being explored. The methods used to study phage diversity fall within two categories: (1) culture-dependent methods, based on phage isolation in pure cultures and (2) culture-independent methods, based on microbial ecology tools, as for example sequencing of marker genes or of phage genomes directly from the environment. The first set of methods allows in-depth characterization of specific phage-host systems. For example, cultured phages can be visualized by electron microscopy, can be (relatively) easily genome sequenced, their interaction with the host can be studied in great detail and their host range can be determined. The second set of methods most often allows access only to phage gene or genome sequences. However, they allow sampling and analysis on a much larger scale than would be possible by using only phage isolates.

By now, phages for several marine bacterial groups have been isolated and characterized. For isolation of phage cultures from an environmental sample, the phage fraction of the respective samples has to be separated from the microbial cells first. Then, the phage fraction is mixed with growing cultures of the hosts of interest. The propagation of host-specific phages is monitored by tracking cell lysis, for example by plaque assays or turbidity measurements. If lysis is observed, the new phages are purified by successive plaque assays when the hosts can grow on solid media, or by dilution to extinction when the hosts grow only in liquid media. The most seminal isolation efforts were focused on phages infecting cyanobacteria, also called cyanophages. More than 7000 cyanophages have been obtained from different studies. However, many isolates have a high degree of genomic relatedness, which reduces their diversity to just a few groups. In comparison, the number of isolates available from other bacterial groups is smaller, but of comparable diversity. For example, to date there have been published 34 phages infecting the *Roseobacter* group, 40 phages infecting *Cellulophaga* and 15 phages infecting *Pelagibacter*. Although having phage cultures is highly advantageous, the inability to cultivate most of the marine bacteria results also in an inability to cultivate their phages and the diversity we see from cultured phages is only the tip of the iceberg. Therefore, researchers need other methods to access environmental phage diversity.

Phage Marker Genes

In microbial ecology, marker genes have long been used to analyze the composition and functions of microbial communities. The most used gene in diversity studies is the one encoding for the 16S rRNA, because it is universally present in cellular organisms and it is conserved, allowing identification and classification of microorganisms. In contrast, phages don't have a universally present gene. Instead, specific marker genes have been used to assess the diversity of certain phage groups in environmental studies. Genes encoding different proteins were used, as for example genes for the major capsid protein or the DNA polymerase. In these approaches, specific primers were used to amplify by Polymerase Chain Reaction (PCR) fragments of the marker genes, followed

by sequencing and sequence analysis. A recent development, the solid-phase single-molecule PCR polony method allows quantification of virus populations in nature.

Viral Metagenomics (Viromics)

Much of our knowledge regarding phage diversity comes from metagenomics studies. A metagenome represents a collection of the sequences of all DNA molecules in a sample. When only the viral fraction is targeted, it is named viral metagenome, or virome. To generate a virome, the phage fraction is separated, the DNA is extracted, fragmented in small pieces and sequenced using one of the high throughput sequencing technologies, most often Illumina. This approach generates many small sequence fragments, also called reads, which then have to be assembled into the original phage genomes (see Fig. 4). Most of the times, due to the huge sequence diversity, only partial phage genomes can be assembled. These partial or complete phage genomes (also called contigs) are then grouped into populations, roughly representing the equivalent of phage species, and then into viral clusters, representing approximately the equivalent of genera or subfamilies (see Fig. 5).

The biggest collection of marine viromes has been generated from three scientific expeditions through the world's oceans: Malaspina, Tara Oceans and Tara Oceans Polar Circle. A total of 145 samples have been collected, mainly from the epipelagic and mesopelagic layers, but also from the bathypelagic layer. The first analysis of these viromes included only the Malaspina and the Tara Ocean expeditions and only the epipelagic and the mesopelagic layers. The resulting dataset was labeled the Global Ocean Virome (GOV) and contained 1,380,834 viral contigs. These contigs were grouped first in 15,222 viral populations and then in 867 viral clusters. From the 867 viral clusters, two-thirds did not affiliate with any cultivated virus, underlining the richness of unexplored marine virus diversity. Five of the 867 viral clusters were relatively ubiquitous, being abundant in more than 20 samples. Only three of the clusters were related to known marine phage groups, that is the T4 superfamily, the Cbaphi381 virus, and the T7virus (see more about these groups in the "Cultivated marine phages" section). The predicted hosts for these viral clusters included *Actinobacteria*, *Alphaproteobacteria*, *Deltaproteobacteria*, *Gammaproteobacteria*, *Bacteroidetes*, *Cyanobacteria*, and *Deferribacteres*.

The second Global Ocean Virome (GOV2) dataset comprises all 145 samples, including the ones in GOV, deeper sequenced. In GOV2 there are of 195,728 viral populations with contig size bigger than 10 kb. From these, 90% could not be assigned to any known viral family, however, the expectation is that a big portion represents phages. The other 10% were assigned to dsDNA viruses and bacteriophages.

Standard viromics has enabled unprecedented access to the diversity of marine phages. However, one of its shortcomings is that microdiversity, defined by the presence in the same sample of similar phage genomes (for example genomes with >95% identity over at least 70% of their genome), hinders assembly of complete genomes from small reads. As a result, at least part of the phage contigs obtained from viromics studies originate probably from the phages with low microdiversity in the respective samples, but not necessarily with high abundances. Several approaches that can circumvent this problem are reviewed below.

Viral Contigs From Fosmid Libraries

A first approach is to clone environmental phage DNA into fosmids, and then to sequence each fosmid using Illumina technology (see Fig. 6). Because one fosmid carries only one phage DNA fragment, this reduces the assembly complexity and, in principle, allows the retrieval of phage contigs from microdiverse phages. On the other hand, large phage genomes cannot be cloned into fosmids, the practical insert size being in the range of 5–48 kb. Despite this limitation, complete phage genomes can be obtained using fosmid libraries, as exemplified by the retrieval of 208 such genomes from the deep chlorophyll maximum (DCM) in the Mediterranean Sea. These phage genomes were grouped in 21 clades of the *Caudovirales* and were predicted to infect SAR11, SAR116, *Cyanobacteria* and the low GC *Actinobacteria*. Fosmid libraries were also used to explore the viral diversity in the 1000 m deep Adriatic Sea and the 3000 m deep Ionian Sea. A total of 28 complete viral genomes, with lengths of 30–41 kb, were obtained. To assess the distribution of these genomes in the water column, raw reads from the Pacific Ocean virome dataset were mapped against each of them. From the 28 genomes, several were found only in the bathypelagic and mesopelagic layers and in high abundances. Five of the phage genomes were predicted to infect the deep ocean SAR11. From these five, uvDeep-CGR2-AD10-C281 was abundant in the mesopelagic and bathypelagic waters and uvDeep-CGR2-KM22-C255 was found both in the deep and surface waters. In addition, 11 contigs represented putative prophages likely infecting *Psychrobacter*, *Flavobacteria*, *Planctomyces*, and *Pelagibacter*.

Single Virus Genomics

A second approach is single-virus genomics. Individual phage particles are separated by flow cytometry, followed by genome extraction, amplification and then sequencing (see Fig. 7). The resulted genomes are called viral single-amplified genomes (vSAGs). Applying this method on seawater samples from the Mediterranean Sea and the Atlantic Ocean resulted in 44 vSAGs. Most of the vSAGs were tentatively assigned to the *Caudovirales*, representing potentially 37 new species and 7 new genera. One of the vSAGs, vSAG-37-F6, was later discovered to infect environmental *Pelagibacteriales* and was ranked as the 13th most abundant marine phage. Furthermore, a high abundance of vSAG-37-F6 transcripts was found in coastal temperate waters from the NE Atlantic, suggesting that this phage was actively infecting its host.

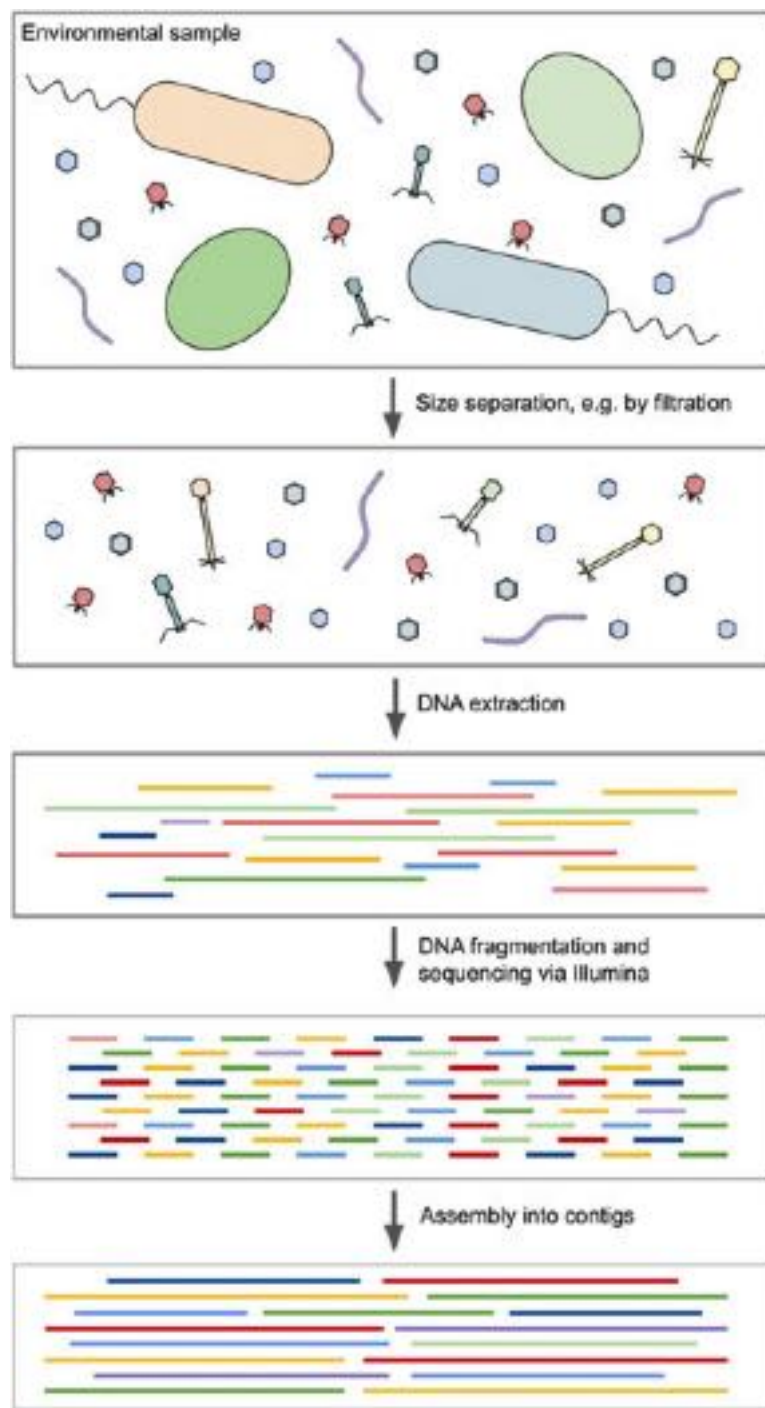


Fig. 4 Workflow for preparation of viral metagenomes.

A limitation of single virus genomics becomes visible when considering the high number of viral particles in environmental samples: this approach becomes impractical at the viral community scale.

Long Read Viromics

A third and very promising approach to capture phage microdiversity is VirION. It combines Illumina sequencing technology, with reads of ~ 300 bases, and nanopore technology, with reads of ~ 4000 bases. The combination is necessary because the long read technologies have a high error rate, and the high-quality Illumina reads are used to correct the errors. When applied to seawater

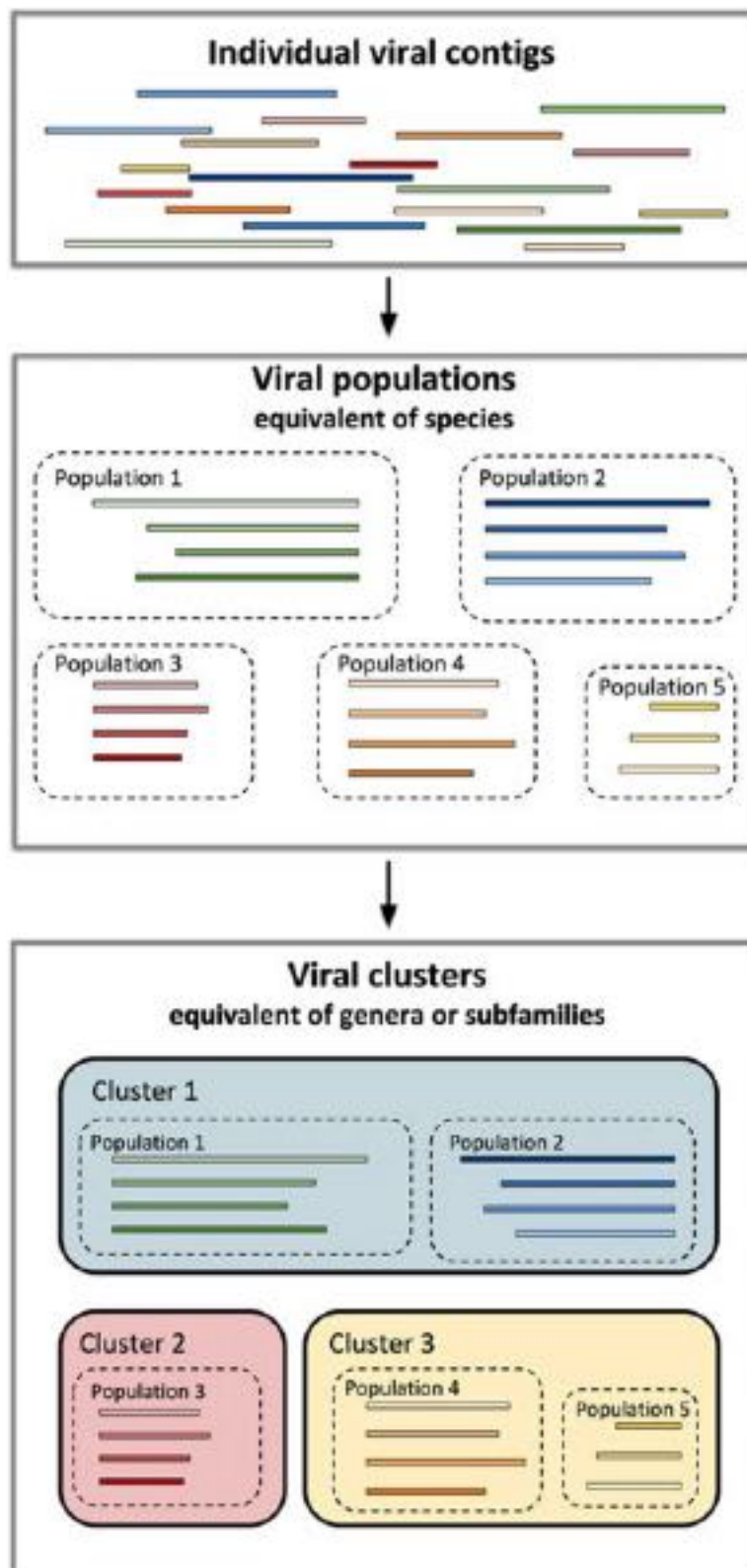


Fig. 5 Grouping of viral contigs in populations (based on sequence similarity) and clusters (based on shared gene content).

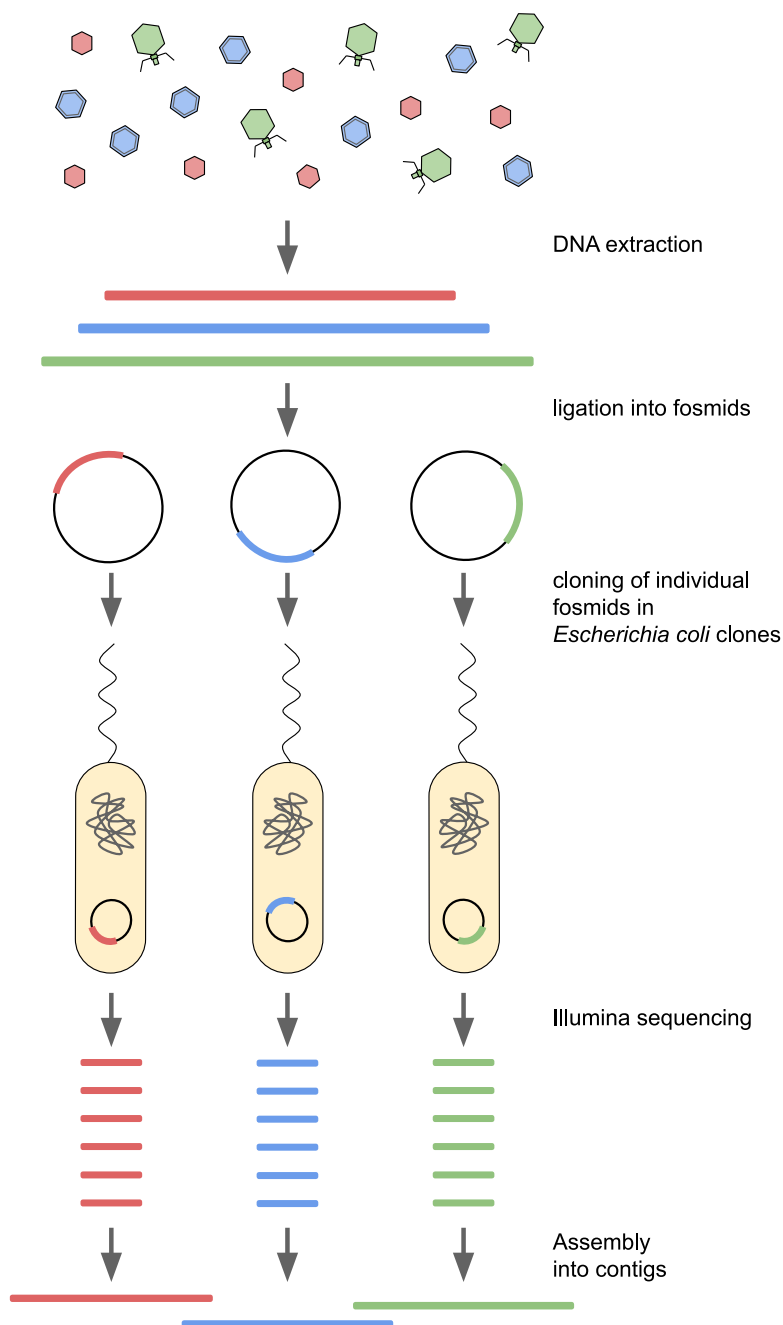


Fig. 6 Preparation of viral contigs from fosmid libraries.

samples from the Western English Channel, VirION was able to capture phage genomes belonging to abundant, microdiverse populations that were previously missed by short-read data. One of the assembled contigs, H_NODE_1248, was identified as the most abundant and ubiquitous viral genome in the marine environment. H_NODE_1248 belongs to the same viral cluster as 57 other contigs, including vSAG-37-F6.

Metagenome Assembled Viral Genomes (MAVGs)

Cellular metagenomes can be a rich source of phage contigs. It is estimated that in between 20%–60% of bacterial cells in the marine environment are infected by phages at any time. Therefore, bacterial metagenomes also contain sequences of phages actively infecting the cells at the time of sampling (see Fig. 8). The deeper the metagenome is sequenced (a higher number of reads

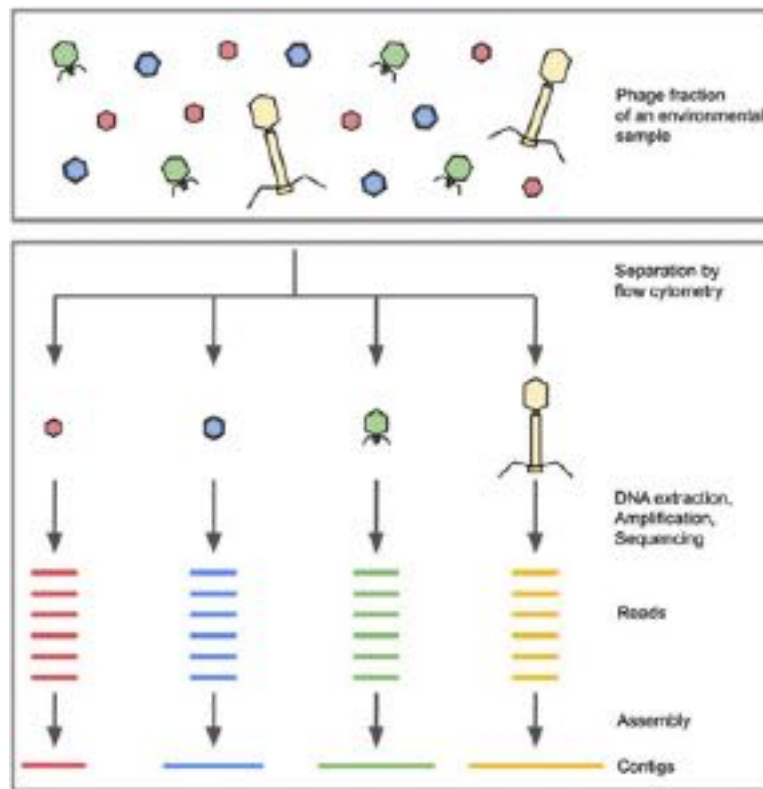


Fig. 7 Workflow for single-virus genomics.

produced), the higher the number of viral sequences retrieved. This approach was recently applied to samples from the Mediterranean Sea, from various depths and sampling sites. More than 1300 viral sequences were obtained, from which 36 contigs represented complete genomes. Host prediction indicated that part of the viral contigs infect *Cyanobacteria*, *Actinomarina*, SAR11, SAR116 or other *Alphaproteobacteria*. Most of the retrieved phage contigs were found exclusively in the Mediterranean Sea, and some were specific to the bathypelagic realm.

Linking Phage Genomes With Host Identity

Obtaining phage genomes using either viromics or single-virus genomics does not preserve the link between the phage and its host. Several computational methods can be used to predict the potential hosts, as for example (1) sequence similarity with known phages, (2) sequence similarity between phage and host genome, (3) similarities between phage genomes and host CRISPR spacers or (4) similarity of phage and host tRNA. However, for the vast majority of environmental phage genomes, the host cannot be predicted using bioinformatics approaches. Outside the realm of bioinformatics, phageFISH and viral tagging allow linking phage and host identity in environmental samples. In phageFISH, microbial cells in a sample are hybridized with fluorescently labeled probes targeting their 16S rRNA, for taxonomic identification. Simultaneously, intracellular phages are hybridized with probes specific for their genes, labeled with different fluorochromes. Hybridized cells and phages are detected by microscopy. Overlapping 16S rRNA and phage fluorescent signals are considered proof that those specific cells are infected by the targeted phage.

Another way to link phages and hosts is to sequence genomes from individual environmental cells. These type of genomes are called single amplified genomes (SAGs), because the DNA from one single cell is extracted and then amplified enzymatically to obtain the amount necessary for sequencing. Single cell genomics recovers all DNA molecules in a cell. Therefore, if infecting phages are found in the cell, they will be sequenced together with the genome of its host, enabling thus the linkage between host cell identity and phage identity (see Fig. 9). Even more, the approximate stage of infection can be estimated by comparing the number of reads per genome base, also called coverage, corresponding to the phage and bacterial genome. A higher coverage of the phage genome indicates a higher number of phages per cell, therefore an advanced infection stage. In SAGs obtained from the Saanich Inlet water column, 69 phage contigs infecting the gammaproteobacterial SUP05 clade were identified. Phylogenetic analysis grouped the contigs into five new genera within the dsDNA *Caudovirales* and the ssDNA *Microviridae*. Several cells were co-infected by dsDNA and ssDNA phages, potentially indicating a cooperative interaction

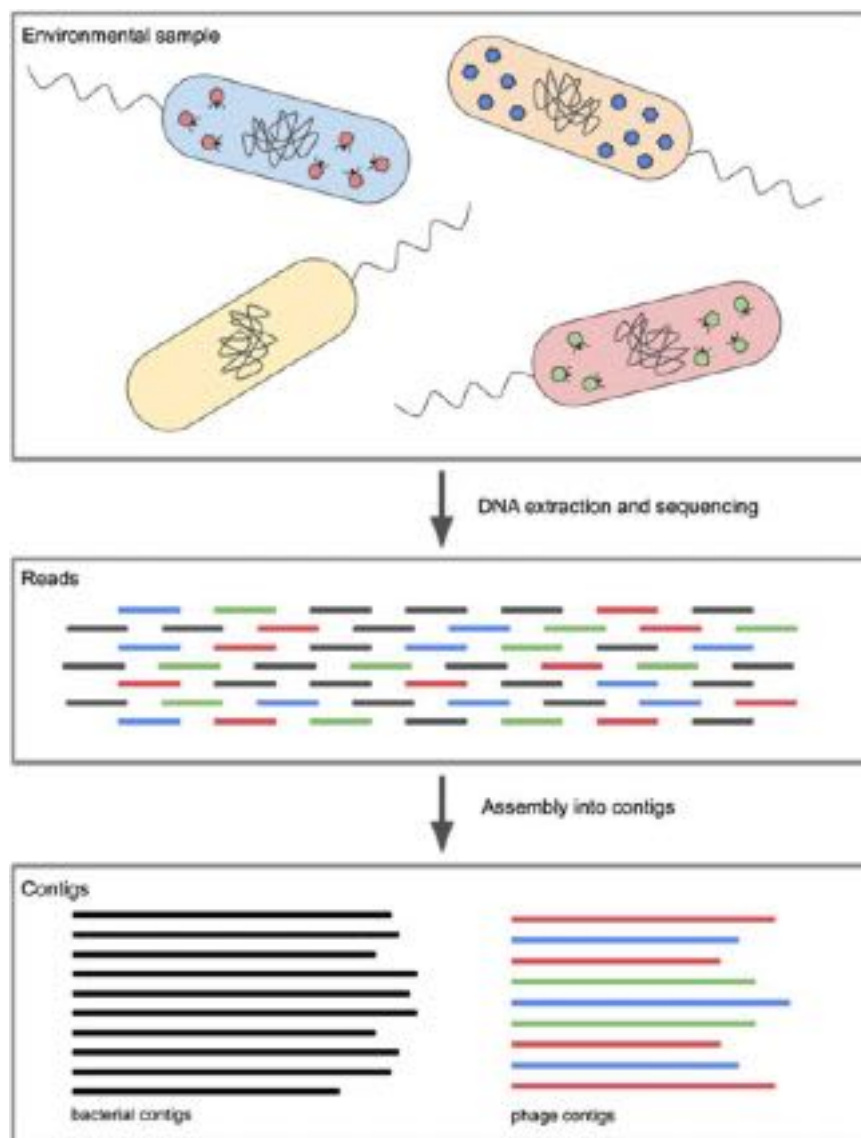


Fig. 8 Preparation of viral contigs from cellular metagenomes (MAVGs).

between these phage types. In another study, twenty complete or near-complete phage genomes were discovered in SAGs from seawater samples from the Gulf of Maine and the North Pacific subtropical gyre. The infected hosts belonged to *Marinimicrobia*, *Verrucomicrobia*, *Gammaproteobacteria* lineages SAR86 and SAR92, *Bacteroidetes* and *Alphaproteobacteria*. Gene content analysis indicated that the assembled genomes belonged to dsDNA, tailed phages in the order *Caudovirales*. Retrieval of an algal virus from a *Verrucomicrobia* cell indicated that unspecific attachment of virus particles to cells can be a problem when obtaining phage genomes from bacterial SAGs.

ssDNA Phages in the Marine Environment

Most of the viral metagenomic studies have focused on dsDNA phages. In comparison, less is known about the diversity of ssDNA and RNA phages in the marine environment, because their nucleic acid cannot be sequenced by the usual methods. To sequence ssDNA phages, their genome has to first be converted to dsDNA. This can be achieved using several methods, from which multiple displacement amplification (MDA) is the most used. However, MDA is not quantitative, because it is amplifying predominantly circular ssDNA phages, to the detriment of dsDNA phages. The method giving the most precise quantitative results from using an adaptase-linker amplification. Applying this method on marine samples, it was concluded that ssDNA phages represent not more than 5% of the marine phage community.

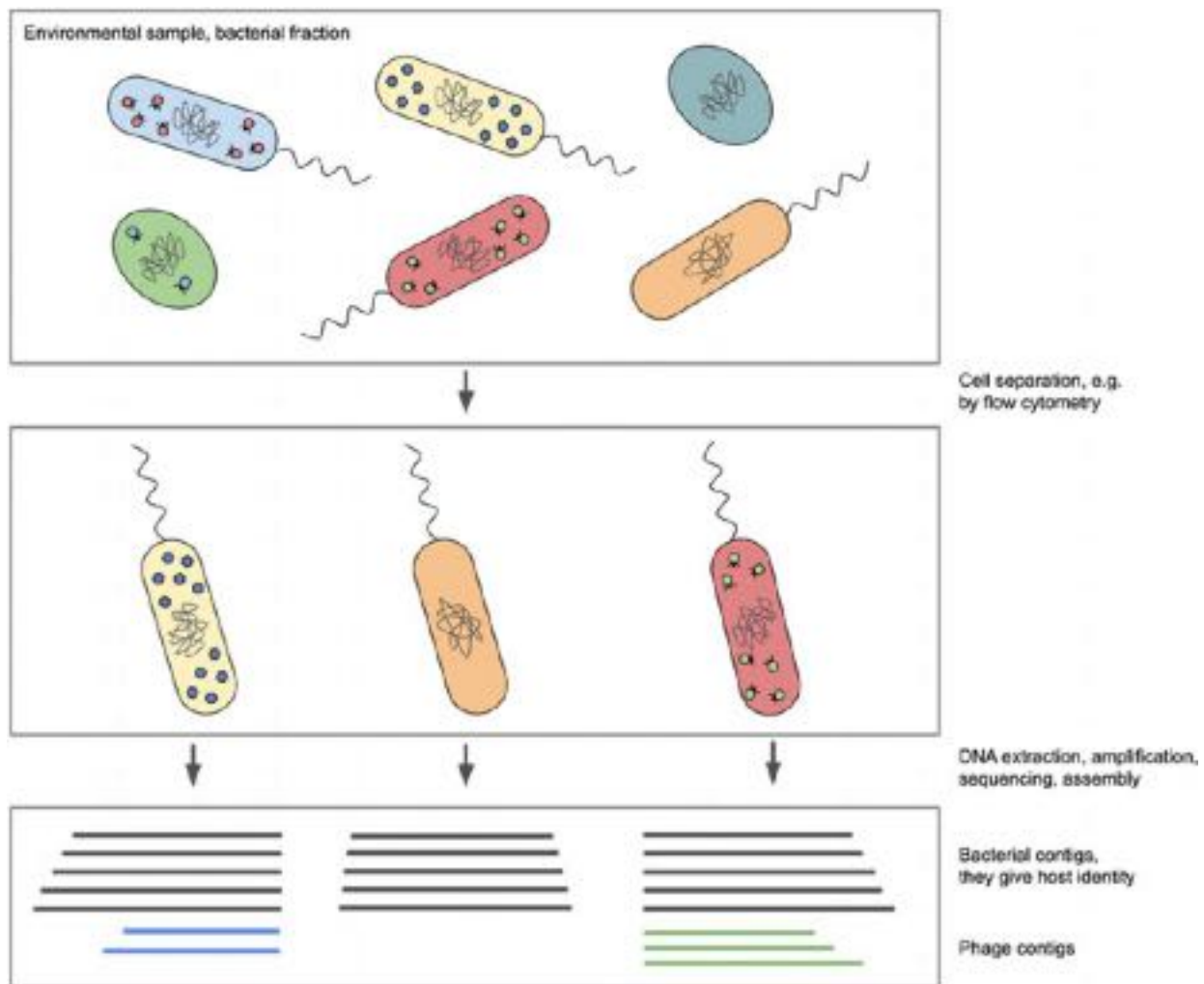


Fig. 9 Linking host identity and phage genomes using single-cell genomics.

Cultivated Phages Infecting Main Groups of Marine Bacteria

Phages of Marine Cyanobacteria – Cyanophages

Cyanophages are viruses infecting cyanobacteria. As oxygenic photoautotrophs, cyanobacteria are the main contributors to the oceanic primary production. Furthermore, many cyanobacteria are diazotrophs and thus they can assimilate N_2 gas and make it available for other organisms. This is especially important in ocean surface waters, where nitrogen is often a limiting nutrient for primary production. Cyanobacteria can be categorized into two major subgroups: unicellular and filamentous. The most abundant cyanobacteria in the oceans are the unicellular *Prochlorococcus* and *Synechococcus*. Due to their small size, they are also called picocyanobacteria. *Prochlorococcus* dominates in oligotrophic oceans between $40^\circ N$ and $40^\circ S$. The co-occurring genus *Synechococcus* is usually less abundant than *Prochlorococcus*, but it occurs also in sub polar regions, brackish and nutrient-rich environments.

Cyanophages have been studied quite intensively in the last decades. However, in comparison to their hosts, they are still poorly understood and the knowledge is severely biased towards certain host species and environments. The majority of the isolates and of the publicly available genomes belongs to phages of *Prochlorococcus* and *Synechococcus*, also called picocyanophages. Phages infecting filamentous and unicellular cyanobacteria form separate phylogenetic clusters.

Phages of unicellular cyanobacteria

More than 7000 picocyanophages have been isolated, from which at least 1000 have been sequenced in different studies. All isolated picocyanophages are tailed (see [Table 1](#)). The T4-like, S-TIM5 and S-CBWM1 picocyanophages have myoviral morphology. The T7-like picocyanophages have podoviral morphology. The other isolates have siphoviral morphology. Most of the picocyanophage in culture are either T4-like ([Fig. 10.1](#)) or T7-like.

Table 1 Cultivated cyanophages infecting unicellular cyanobacteria

Cyanophages	Genome type	Morphology	Life style	Host	Genome size (kb)	G + C content [%]	tRNAs #
T4-like	dsDNA	myoviral	strictly lytic	<i>Prochlorococcus</i> <i>Synechococcus</i>	170–250	~33–45	0–25
S-TIM5				<i>Synechococcus</i>	~161	~41	10
S-CBWM1				<i>Synechococcus</i>	~140	~52	34
T7-like		podoviral	strictly lytic, some potentially temperate	<i>Prochlorococcus</i> <i>Synechococcus</i>	41–48	~38–55	
P-SS2		siphoviral	strictly lytic, potentially temperate	<i>Prochlorococcus</i>	~108	~52	n.a.
S-CBS1					~30	~59	0
S-CBS2				<i>Synechococcus</i>	~72	~55	0
S-CBS3				<i>Synechococcus</i>	~33	~61	0
S-CBS4				<i>Synechococcus</i>	~69	~51	1
KBS2A			Strictly lytic	<i>Synechococcus</i>	~40	n.d.	0
A-HIS1				<i>Acaryochloris</i>	~55	~47	0
A-HIS2				<i>Acaryochloris</i>	~57	~47	0

Most cultivated phages infecting picocyanobacteria are strictly lytic. To date, no temperate picocyanophages have been isolated. However, induction experiments with environmental cyanobacteria suggest the existence of prophages. Furthermore, the *Prochlorococcus* phage P-SSP7 has an integrase gene, potentially being able to integrate into the host genome. Recently, a 100 kb putative prophage was predicted in the genome of *Synechococcus* WH8016. This prophage forms a new picocyanophage clade together with four MAVGs recruited from the Red Sea and Tara Ocean cellular metagenomes. This clade is found all over the oceans, in low abundance.

Several cyanophages, as for example S-CBWM1, S-TIM5 and those infecting *Acaryochloris* (see Table 1), have mitochondria-like DNA polymerase genes. Phylogenetic analysis indicates a likely common ancestor for the phage and mitochondrial DNA polymerase genes. It is likely, but not proven, that the DNA polymerase gene was transferred from a phage to an alpha-proteobacterial mitochondrial progenitor.

Picocyanophage and their host systems are the most studied amongst marine phage-host systems. Because several ecological principles derived from their study can be extended to other phage-host systems, they are further reviewed in this book article separately, in the “Marine Phage Ecology” section.

Phages of filamentous cyanobacteria

Filamentous cyanobacteria are free-living like the aggregate-forming *Trichodesmium* or associated with diatom hosts like the heterocyst-forming *Richelia* and *Calothrix*. They often form blooms in estuarine and marine environments. The production of cyanotoxins and oxygen depletion during and after the blooms can change the functioning of the whole ecosystem and can be harmful to the environment, human health and economy. Global warming is expected to enhance the frequency and intensity of cyanobacterial blooms. Cyanophages are thought to play an essential role in the termination of cyanobacterial blooms, but also in their maintenance due to community shifts between specific ecotypes. For example, phage lysis of *Nodularia* leads to the release of nitrogen and stimulates the growth of *Synechococcus*. A summary of the known phages infecting filamentous cyanobacteria is given in Table 2.

Phages of Marine Alphaproteobacteria

Phages of marine Pelagibacterales

Pelagibacterales, also known as SAR11, are small, vibrio-shaped chemoheterotrophic bacteria that constitute 20%–40% of all planktonic cells in the euphotic ocean and ~20% of cells in the deeper mesopelagic and bathypelagic waters. They are present in the oceans from pole to pole, but their largest numbers are found in stratified, oligotrophic gyres.

For a while, it was believed that the success of SAR11 clade is due to a lack of phage predators. However, this hypothesis was disproved by the isolation of several phages infecting *Pelagibacterales*, also called pelagiphages (Fig. 10.2). To date, there are only 15 known pelagiphage isolates, from which 14 are podoviruses and one is a myovirus related to T4 phages (see Table 3). Most pelagipodovirus isolates are placed within the proposed genus HTVC019Pvirus of the *Autographivirinae* subfamily and have the ability to propagate by both lytic and lysogenic cycles. They encode in their genomes a tyrosine integrase. Most likely this integrase catalyzes the integration of the pelagiphage in the host genome, through site-specific recombination in between the bacterial chromosomal attachment site (*attB*) and phage chromosomal attachment site (*attP*). The *attB* sites are various tRNAs present in the SAR11 genome, while the *attP* site is located in between the integrase and the RNA polymerase gene. Integration of HTVC019Pvirus phages has been shown both in SAR11 cultures and in metagenomic samples, suggesting that temperate pelagiphages could be implicated in ecological processes on broad scales.

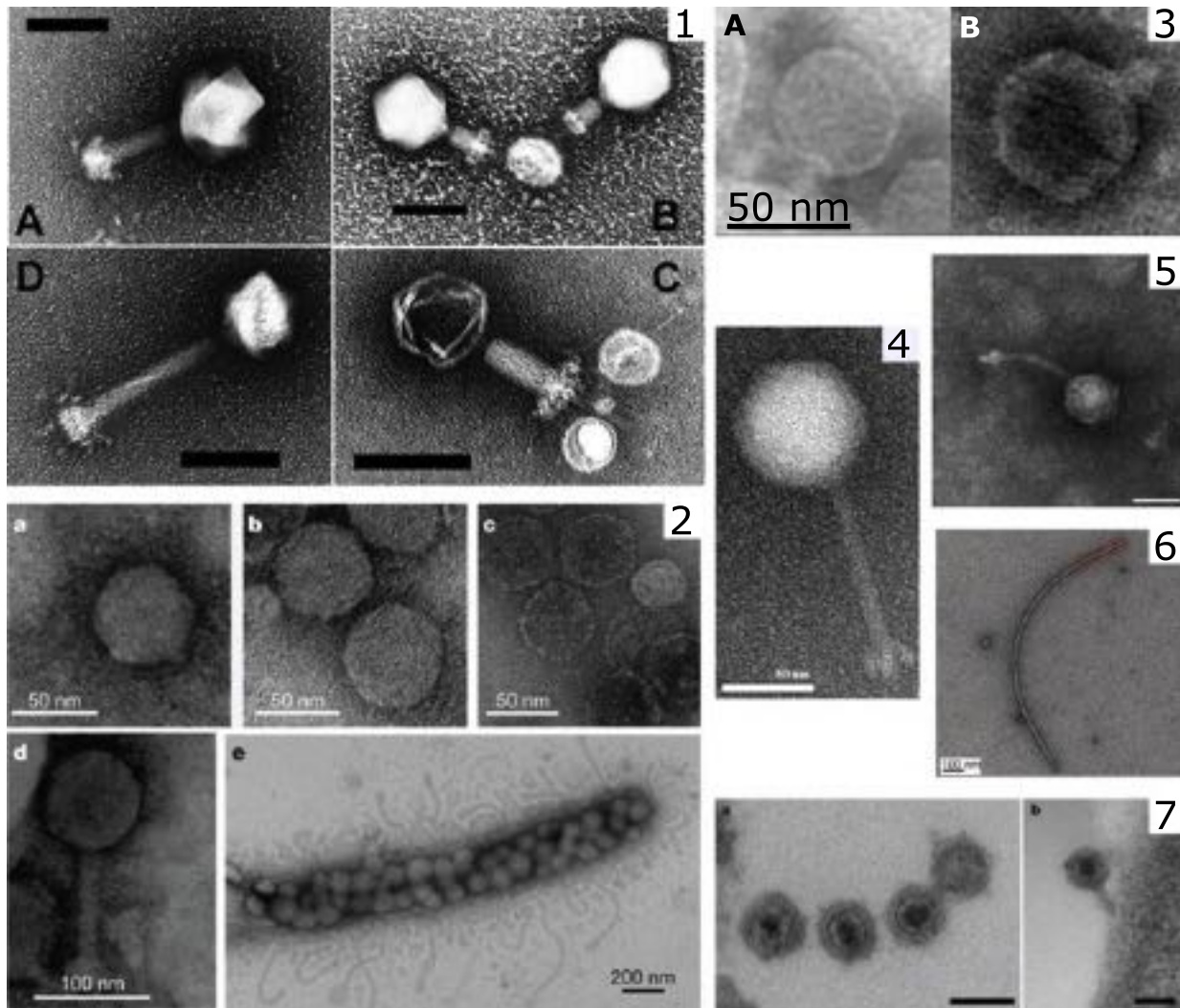


Fig. 10 (1) Electron micrograph of negative-stained *Prochlorococcus* myoviruses P-SSM2 and P-SSM4. Scale bars indicate 100 nm. (A) P-SSM2 with non-contracted tail. (B) P-SSM2 with contracted tail. (C) P-SSM4 with contracted tail. (D) P-SSM4 with non-contracted tail. (2) Electron micrograph of isolated pelagiphages. (A) Pelagipodovirus HTVC011P. (B) Pelagipodovirus HTVC019P. (C) Pelagipodovirus HTVC010P. (D) Pelagimyovirus HTVC008M. (E) Host cell of *Candidatus* P. ubique strain HTCC1062 infected with HTVC011P. (3) Electron micrograph of cobaviruses. (A) Molybdenum stained, cell debris bound Lentibacter virus vB_LenP_ICBM1. (B) Uranyl acetate stained, free Lentibacter virus vB_LenP_ICBM2. Scale bars indicate 50 nm. (4) Electron micrograph of *Pseudoalteromonas phenolica* infecting phage TW1. Scale bar indicates 50 nm. (5) Electron micrograph of *Pseudoalteromonas* phage vB_PspS-H40/1. Tungstophosphoric acid stained. Scale bar indicates 40 nm. (6) Electron micrograph of phage f327 infecting *Pseudoalteromonas* sp. BSI20327. Scale bar indicates 100 nm. (7) Thin-section electron micrograph of autolykiviruses. Scale bar indicates 50 nm. Reproduced from (1) Sullivan, M.B., Coleman, M.L., Weigele, P., Rohwer, F., Chisholm, S.W., 2005. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biology* 3 (5), e144. doi:[10.1371/journal.pbio.0030144](https://doi.org/10.1371/journal.pbio.0030144). Used under CC BY (<http://creativecommons.org/licenses/by/4.0/>). No changes made. (2) Zhao, Y., Temperton, B., Thrash, J.C., et al., 2013. Abundant SAR11 viruses in the ocean. *Nature* 494 (7437), 357–360. doi:[10.1038/nature11921](https://doi.org/10.1038/nature11921). Copyright (2013, Springer Nature). (3) Bischoff, V., Bunk, B., Meier-Kolthoff, J.P., et al., 2019. Cobaviruses – A new globally distributed phage group infecting Rhodobacteraceae in marine ecosystems. *The ISME Journal*. doi:[10.1038/s41396-019-0362-7](https://doi.org/10.1038/s41396-019-0362-7). Used under CC BY (<http://creativecommons.org/licenses/by/4.0/>). (4) Shin, H., Lee, J.-H., Ahn, C.S., Ryu, S., Cho, B.C., 2014. Complete genome sequence of marine bacterium *Pseudoalteromonas phenolica* bacteriophage TW1. *Archives of Virology* 159 (1), 159–162. doi:[10.1007/s00705-013-1776-6](https://doi.org/10.1007/s00705-013-1776-6). Copyright (2013, Springer Nature). (5) Kallies, R., Kiesel, B., Schmidt, M., et al., 2017. Complete genome sequence of *Pseudoalteromonas* phage vB_PspS-H40/1 (formerly H40/1) that infects *Pseudoalteromonas* sp. strain H40 and is used as biological tracer in hydrological transport studies. *Standards in Genomic Sciences* 12, 20. doi:[10.1186/s40793-017-0235-5](https://doi.org/10.1186/s40793-017-0235-5). Used under CC BY (<http://creativecommons.org/licenses/by/4.0/>). No changes made. (6) Yu, Z.-C., Chen, X.-L., Shen, Q.-T., et al., 2015. Filamentous phages prevalent in *Pseudoalteromonas* spp. confer properties advantageous to host survival in Arctic sea ice. *The ISME Journal* 9 (4), 871–881. doi:[10.1038/ismej.2014.185](https://doi.org/10.1038/ismej.2014.185). Copyright (2014, Springer Nature). (7) Kauffman, K.M., Hussain, F.A., Yang, J., et al., 2018. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature*. doi:[10.1038/nature25474](https://doi.org/10.1038/nature25474). Copyright (2018, Springer Nature).

Table 2 Phages of filamentous cyanobacteria

Cyanophages	Genome type	Morphology	Life style	Host	Genome size (kb)	G + C content	tRNAs #
vB_NpeS-2AV2	dsDNA	siphoviral	strictly lytic	<i>Nodularia</i>	~139	~40	1
vB-AphaS-CL131				<i>Aphanizomenon flos-aquae</i>	~113	~40	2
A-1(L)		myoviral		<i>Anabaena</i>	~68	n.d.	n.d.
A-4L		podoviral			~42	~43	0
N1		myoviral		<i>Nostoc</i>	~65	~35	0
A1					~68	~37	0
PaV-LD		podoviral		<i>Planktothrix agardhii</i>	~95	~41	0
PP				<i>Phormidium foveolarum</i>	~43	~46	0
Pf-WMP3					~43	~47	0
Pf-WMP4					~41	~52	0
NCTB ^a		myoviral	n.d.	<i>Trichodesmium</i>	~258	~42	0

^aUncultivated, genome obtained from a decaying bloom of *Trichodesmium* in the New Caledonian lagoon in the South Pacific Ocean.

Note: n.d. = not determined.

Table 3 Cultivated pelagiphages

Pelagiphages	Genome type	Morphology	Life style	Host genus	Genome size (kb)	G + C content	tRNAs #
HTVC019Pvirus	dsDNA	podoviral	lytic and lysogenic	<i>Pelagibacter</i>	37–42	~32–36	0–1
HTVC010P			strictly lytic		~35	~30	0
HTVC008M		myoviral	strictly lytic		~147	~34	0

Amongst cultivated pelagiphages, HTVC010P is the most abundant in the oceans. In the North and South Atlantic, it was found both in surface waters and in the DCM layer. HTVC008M was found only in a few stations in the Southern Atlantic, indicating latitudinal biogeography of the SAR11 phages. A study looking at the ratio between the phage and host transcripts in the coastal waters of the NE Atlantic, found a low ratio between cultivated pelagiphages and SAR11. This is suggestive of a low lytic activity of cultivated pelagiphages in coastal waters, not surprising considering that SAR11 generally has slow growth rates and that cultivated pelagiphages have long latent periods (16–22 h) and lower burst sizes (9–49 phages per infected cell).

SAR116 phages

SAR116 is an important group of heterotrophic bacteria in the surface ocean, with abundances as high as 10%. They have genes for proteorhodopsin-based photoheterotrophy, dimethylsulfoniopropionate, carbon monoxide and C1-metabolism, indicative of their involvement in marine biogeochemical cycles.

To date, only one phage infecting the marine SAR116 has been isolated, named HMO_2011. Its host is "*Candidatus* Punicispirillum marinum" strain IMCC13122. Phage HMO_2011 has a podoviral morphology and a ~55 kb dsDNA genome. As it is typical with phage genomes, most of the encoded genes are hypothetical. Amongst the annotated genes, the DNA polymerase and the putative methanesulfonate monooxygenase (MsmA) are noticeable. The DNA polymerase contains a partial DnaJ domain between the 3'-5' exonuclease and the DNA polymerase domains. Usually, this domain is found in molecular chaperones. The MsmA presence hints toward the involvement of the HMO_2011 phage in the C1 metabolism of SAR116.

The environmental distribution of the HMO_2011 phage was assessed using the Pacific Ocean Virome dataset. It was found that HMO_2011 is distributed mainly in the euphotic zones of both coastal and open ocean, and it is much less abundant in the aphotic zones. In the coastal waters of the NE Atlantic, HMO_2011 was highly active, as judged from the ratio between phage and host transcripts. This is in agreement with the shorter latency time (6 h) and bigger burst size (500 phages/infected cell) observed in culture.

Phages of marine Rhodobacteraceae – Roseophages

Marine *Rhodobacteraceae*, also known as the *Roseobacter* Group, are another important group of heterotrophic bacteria in the marine environment. They have a high metabolic versatility, being able to perform anoxygenic photosynthesis, metabolize a large variety of organic compounds, degrade dimethylsulfoniopropionate and produce various secondary metabolites. Most often, they are associated with micro- and macroalgae, with which they potentially engage in both mutualistic and pathogenic interactions.

Roseophages, that is phages infecting roseobacters, have been isolated for ten out of more than 70 described genera of marine *Rhodobacteraceae*. These ten genera are *Ruegeria*, *Sulfitobacter*, *Roseovarius*, *Dinoroseobacter*, *Roseobacter*, *Lentibacter*, *Celeribacter*, *Loktanella*, *Pelagibaca* and *Thiobacimonas*. Most of the roseophages are tailed, with dsDNA genomes and a podoviral or siphoviral morphology, and only two isolates have a ssDNA genome (see Table 4). There are two main groups of roseopodoviruses, the N4-like and the Cobavirus group, both strictly lytic. The roseosiphophages are more diverse, most of them being classified as

Table 4 Cultivated roseophages

<i>Roseophages</i>	<i>Genome Type</i>	<i>Morphology</i>	<i>Life style</i>	<i>Host genus</i>	<i>Genome size (kb)</i>	<i>G + C content</i>	<i>tRNAs #</i>
N4-like	dsDNA	podoviral	strictly lytic	<i>Ruegeria</i> , <i>Sulfitobacter</i> <i>Roseovarius</i> <i>Dinoroseobacter</i> <i>Roseobacter</i>	73–77	~43–51	0–15
Cobavirus group			strictly lytic	<i>Lentibacter</i> <i>Sulfitobacter</i> <i>Celeribacter</i>	39–41	~46–48	0
phiCB2047-A, phiCB2047-C			temperate	<i>Sulfitobacter</i>	~41	~59	0
Chi-like		siphoviral	strictly lytic, potentially temperate	<i>Ruegeria</i> <i>Loktanella</i>	57–61	~55–64	0
Cbk-like				<i>Ruegeria</i>	~147	~56	24–26
RDJLphi1				<i>Roseobacter</i>	~62	~58	
RDJLphi2				<i>Roseobacter</i>	~63	~57	0
vB_DshS-R5C				<i>Dinoroseobacter</i>	~78	~62	0
vB_ThpS-P1			temperate	<i>Thiobacimonas</i>	~39	~67	0
vB_PeaS-P1				<i>Pelagibaca</i>	~38	~64	0
vB_RpoMi-Mini, B_RpoMi-V15	ssDNA	icosahedral, non-tailed	strictly lytic	<i>Ruegeria</i>	4.2	~58	0

Cbk-like, Chi-like and Mu-like head group. The phages in the Mu-like group are temperate, the induction to lytic cycle being shown in deep-sea roseobacters. The other roseosiphophages are mainly strictly lytic, only some having certain genome characteristics, as for example integrase genes, which indicate a temperate potential.

Cobaviruses (Fig. 10.3) comprise several cultivated roseophages, as well as environmental phage genomes, forming a clade in the proposed *Siovirus* genus. Their genomes are split in two arms, with opposite transcriptional directions and separated by a rho-independent transcriptional terminator. The presence of direct terminal repeats (DTRs) at the end of the genomes indicate that cobaviruses replicate using a mechanism similar to the T7 phages. This involves the formation of long concatemeric DNA molecules as intermediates in replication and packaging, concatemers formed by the annealing of 3' single strands resulted at the DTR level during replication. Cobaviruses most likely lyse their host cells via the canonical holin-endolysin pathway. They are found worldwide in the euphotic ocean, mostly in coastal areas, but also in the open ocean. They are more numerous in bays or estuaries, as for example in the Goseong Bay, Delaware Estuary, Chesapeake Bay and Port of Los Angeles. Some cobaviruses were shown to be cosmopolitan, their genome being found for example both in the North Sea and in the Yellow Sea. In coastal environments, cobaviruses seem to be present throughout the year. Both cultivated and environmental cobaviruses encode in their genomes a cobalamin-dependent ribonucleotide reductase gene, related with similar genes from protist-associated microorganisms. This suggests that the habitat of cobaviruses are protist associated bacterial hosts, most probably vitamin B12 producers.

Recently, two ssDNA roseophages, distantly related to other *Microviridae* have been isolated (see Table 4). At 4.2 kb and 4 predicted ORFs, vB_RpoMi-Mini and B_RpoMi-V15 have the smallest genome size amongst known ssDNA phages. They encode a major capsid protein, a replication initiation protein, a peptidase, and a hypothetical protein. Similar proteins have been found in the genome of the *Novosphingobium tardagens* NBRC 16725, indicating the presence of a prophage or prophage remnant. The similarity of the peptidase gene with other genes from bacterial genomes, including genes located in prophage regions, indicate that ssDNA roseophages participate in horizontal gene transfer.

Phages of Marine Gammaproteobacteria

Vibriophages

Vibriophages are viruses infecting bacteria from the *Vibrionaceae* family. These bacteria represent a genetically and metabolically diverse group of heterotrophic bacteria, ubiquitous in the oceans. They can be free-living, however, they are more abundant in sediments or in associations with other marine organisms or organic particles. Furthermore, the group includes several pathogens, including *Vibrio anguillarum*, *V. parahaemolyticus*, *V. harvey*, and *V. vulnificus*. These pathogens infect more than 50 species of fish, mollusks and crustaceans. They can cause vibriosis, a malignant disease affecting the aquaculture industry globally, and food poisoning in people consuming infected, raw seafood. As a consequence, strictly lytic vibriophages are of special interest due to their potential use in phage therapy, as treatment for vibriosis.

The known marine vibriophages have a high diversity, with representatives in *Caudovirales*, *Autolykiviridae*, and *Inoviridae*. In the environment, vibriophages are found both in the water column and in association with animals. For example, a diverse and abundant vibriophage community has been found in oysters (*Crassostrea virginica*) from Delaware Bay, Ladysmith Harbor in British Columbia and from different estuaries in Gulf of Mexico. Vibriophages were present year-round in oyster tissue, with increased abundances in summer. This suggests that vibriophages are a common occurrence in oysters and most likely they play a

Table 5 Cultivated phages infecting *Pseudoalteromonas*

Pseudoalteromonas phage/genus	Genome type	Morphology	Life style	Host genus	Genome size (kb)	G + C content	tRNAs #
†327	ssDNA	filamentous	chronic	<i>Pseudoalteromonas</i>	~6	n.d. ^a	0
PM2	dsDNA	icosahedral	strictly lytic and temperate		~10	~42	0
Rio1-like virus		podoviral	strictly lytic		43–45	~45	0
<i>pYD6-A</i>					~77	~39	2
PSAHM1-like virus		myoviral	strictly lytic		130	~36	0
PSAHS1-like virus		siphoviral	strictly lytic		36–39	~40	0
PSAHS2-like virus					~38	63–64	0
vB_PspS-H40/1					~45	~41	0
Pq0					~33	~40	0
H105/1			strictly lytic and temperate		~31	~41	0
PSAHS6-like virus					~35	~40	0
B8b			Strictly lytic, possibly temperate		~46	50	0
TW1					~40	~40	0

^an.d. = not determined.

role in regulating *Vibrio* populations. Morphological characterization of vibriophages from oysters showed the presence of both tailed and non-tailed phages.

A few hundred lytic marine vibriophages have been isolated to date, from which 251 make part of the Nahant collection and have been genome sequenced. Most belong to *Caudovirales*, being either podo-, siphoviruses or myoviruses. About 20 isolates belong to the newly proposed *Autolykiviridae* family (see below). Caudoviruses in the Nahant collection have genome sizes ranging in between 21 kb and 348 kb. Notably, 32 phages carry putative CRISPR features. CRISPR/Cas systems are usually used by bacteria to protect themselves against phages. However, phages can encode their own CRISPR/Cas systems to evade host immunity. The first such system was shown in a vibriophage infecting *Vibrio cholera*. This suggests that the CRISPR features found in marine vibriophages could function in a similar way.

Autolykiviruses are non-tailed phages representing a novel family of the double jelly-roll capsid viruses. Their capsid has about 50 nm in diameter and it likely contains an inner lipid layer (see Fig. 10.7). Their linear, dsDNA genomes have 10-kb in length, terminate in inverted repeats and encode for ~20 proteins. Autolykiviruses have a broad host range, commonly killing hosts in multiple *Vibrio* species. This is in contrast to tailed phages, which usually infect few and closely related species. The isolated autolykiviruses are all strictly lytic. Autolykivirus-like prophages have been predicted bioinformatically in bacterial genomes and in contigs from marine metagenomes. Their *in vivo* activity has been recently demonstrated in two marine *Vibrio* species.

Prophages are abundant in marine vibrios and can encode (1) virulence factors, as for example Zonula occludens toxin (Zot); (2) bacterial host fitness factors, as for example RTX-toxins, collagenases, lipases, hemolysins, chondroitin AC lyases; and (3) genes involved in antibiotic and heavy metal resistance. In a recent analysis, 15% of the marine *Vibrio* isolates contained *Inoviridae* prophages encoding Zot. Inoviruses are filamentous ssDNA phages with a chronic life cycle. They integrate into the host genome and, upon induction, they release phage progeny without lysing the cell. Zot is known from *Vibrio cholerae* infected by the chronic CTX phage. Zot localizes in the bacterial membrane and increases intestinal permeability of the mammalian hosts. Zot containing prophages were found in isolates from coastal marine waters (e.g., *V. campbellii*), deep hydrothermal vents (e.g., *V. antiquarius* and *V. diabolicus*) and deep seafloor sediments (e.g., *V. diazotrophicus*). This suggests that harmless, environmental bacteria can acquire virulence traits from pathogenic donors and act as potential biological reservoirs of these genes in the environment.

Pseudoalteromonas phages

Pseudoalteromonas spp. are ubiquitous heterotrophic bacteria in the marine environment. Most often they are associated with particles, where they constitute up to 20%–60% of the microbial communities and they are involved in carbon export in the oceans. There are 12 cultivated phage groups infecting *Pseudoalteromonas*. Most of them are dsDNA phages, belonging either to *Caudovirales* (Fig. 10.4 and 10.5) or *Corticoviridae*, and one is a ssDNA phage (see Table 5).

Phage 327 is the first filamentous phage known to infect a *Pseudoalteromonas* strain from the Arctic Sea ice (Fig. 10.6). In culture, it decreases the host growth rate, cell density, and tolerance to H₂O₂ and NaCl, but it enhances its motility and chemotaxis. The latter might confer a survival advantage during the polar night, when nutrients are scarce.

Pseudoalteromonas virus PM2 is the first isolated and the most studied corticovirus. It infects via the lytic cycle the marine *Pseudoalteromonas espejiana* BAL-31. The capsid has ~56 nm in diameter and it contains an inner lipid bilayer. The genome is composed of a circular, highly supercoiled dsDNA molecule. It has only 10 kb in size and encodes for 21 proteins. Genome replication takes place via a rolling-circle mechanism. Through bioinformatics analysis, PM2-like prophages have been predicted in the genomes of cultivated marine *Proteobacteria*, as well as on contigs from marine metagenomes.

Table 6 Cultivated phages infecting *Bacteroidetes*

Phage/genus	Genome Type	Morphology	Life style	Host genus	Genome size (kb)	G + C content	tRNAs #
phi38:1, Cba401likevirus	dsDNA	podoviral	strictly lytic	<i>Cellulophaga baltica</i>	~73	38	16
Cba183likevirus					~73	33	1
Cba142likevirus					~100	30	0
Cba411likevirus					~146	33	24
CbaSMLikevirus		myoviral			~54	33	0
Cba391likevirus		siphoviral			~29	31	0
Cba461likevirus					~35	38	0
Cba181likevirus					~39	37	0
Cba101likevirus					~57	31	3
Cba131likevirus					~78	30	0
Cba184likevirus	ssDNA	icosahedral, non-tailed			~7	34	0
Cba482likevirus		icosahedral, non-tailed	Strictly lytic, possibly temperate		~12	29	0
11b	dsDNA	siphoviral	strictly lytic	<i>Flavobacterium</i>	~36	~31	0
P12024S				<i>Persicivirga</i>	~36	~36	0
P12024L					~36	~36	0
P2559S				<i>Croceibacter atlanticus</i>	~45	~42	0

Phages of Marine *Bacteroidetes*

One important group of marine heterotrophs belongs to the *Bacteroidetes* phylum. Marine *Bacteroidetes* increase in abundance during algal blooms, being responsible for the degradation of biopolymers, for example polysaccharides, and thus are involved in recycling the organic matter derived from blooms. To date, most phage isolates for *Bacteroidetes* (see Table 6) infect the *Cellulophaga* genus, present in coastal environments and most often associated with macroalgae.

To date, 40 isolates grouped in 12 phage genera are known to lytically infect different strains of *Cellulophaga baltica*. Ten of the genera represent tailed dsDNA phages, with podo-, siphovirus and myovirus morphology. The phi38:1 podoviral isolate belongs to one of the most abundant viral clusters in the sunlit ocean, as discovered by viromics studies. The presence of a high number of tRNAs in its genome, correlated with a broader host range compared with isolates with fewer tRNAs, could explain in part the success of this phage and its relatives in the environment. Likely, the tRNAs allow the phage a better use of the host resources. Although phi38:1 phage has a broad host range, it does not infect all *C. baltica* strains with the same efficiency. In the original isolation host, phi38:1 infects 64% of the host cells, has a latent period of 70 min and a burst size of 8 phages per cell. In the alternative host, the infection efficiency of phi38:1 is lower. Only 20% of the host cells are infected, the latent period is ~5 h and the burst size is 40 phages per cell. Transcriptomics experiments showed that phi38:1 has similar expression patterns in the two hosts. On the other hand, the transcriptional response of the isolation and alternative hosts are significantly different. Upon phage infection, the alternative host overexpresses DNA degradation genes and underexpresses translation genes, resulting in a delayed phage replication and protein translation. These findings have implications for phage-infection ecology in nature. Because the microdiversity of both host and phages is expected to be high in the environment, it is likely that variable infection efficiencies are quite common.

Single-stranded DNA phages infecting *C. baltica* have been classified in two genera. The isolates from the Cba184likevirus genus are strictly lytic, have a genome size of ~6.5 kb, a capsid of ~30 nm and they are thought to represent a new subfamily of the *Microviridae*. The Cba482likevirus genus has strictly lytic phages with 11.5 kb genomes, the biggest length known for ssDNA phages. Furthermore, the capsid has a diameter of ~72 nm, twice bigger than other *Microviridae* representatives. Consequently, the Cba482likevirus is thought to represent a third lineage of ssDNA phages, besides *Microviridae* and *Inoviridae*. Sequence similarity indicates that *Zunongwangia profunda* (*Bacteroidetes*) harbors a Cba482likevirus prophage in its genome. This suggests that phages in the Cba482likevirus genus are capable of integrating into host genomes and becoming temperate.

Marine Phage Ecology

Phage Micro- and Macro-diversity in the Marine Environment

An ongoing discussion in virology regards the existence of the viral species. The biological species concept defines species as interbreeding individuals that remain isolated from other similar groups. Viruses don't reproduce sexually. However, they undergo homologous gene exchange, during simultaneous infection of the same host by multiple viruses. If the gene exchange

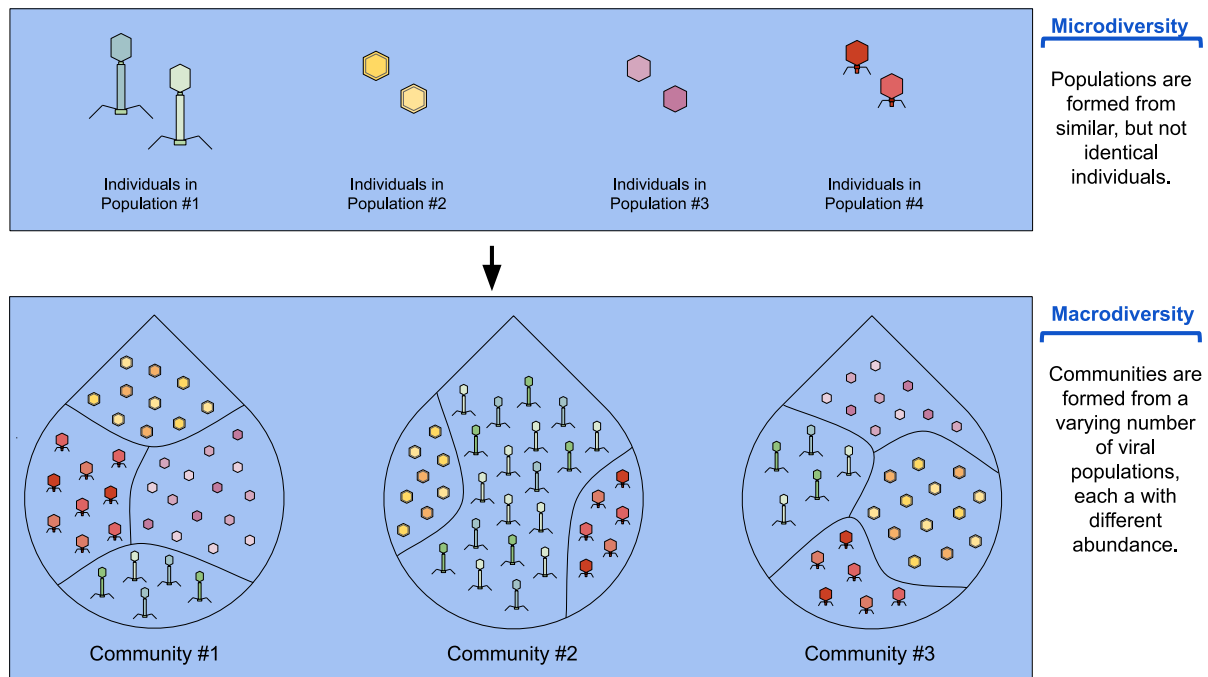


Fig. 11 Schematic representation of viral micro and macrodiversity.

would be confined within groups of similar viruses, then these groups would form discrete viral populations and could be defined as species.

One of the first indications that marine phages form discrete and stable genomic populations came from work on roseophages. Following the isolation of the SIO1 roseophage, a very similar phage was isolated 12 years later, indicating that phages are stable in time. Later on, several studies on cyanophages confirmed the existence of sequence discrete clusters. For example, in coastal waters of Southern New England, different clusters of cyanophages were stably maintained over a time span of 15 years. Furthermore, they exhibited a clear temporal and spatial pattern of abundance, suggesting that they occupy slightly different niches and that they represent viral ecotypes. From the different studies (e.g., [Brum et al., 2015](#); [Roux et al., 2016, 2018](#)), an empiric threshold for defining viral populations was established. This threshold was set at 95% nucleotide identity over at least 70% of the genome. Recently, the existence of discrete viral populations in the marine environment and their delineation by the above threshold was validated using the massive amount of data from the GOV2 dataset.

In an environmental sample, phages can be organized into different ecological levels (see [Fig. 11](#)). The lowest level is the individual phage. All individuals belonging to the same species form a population. The genetic diversity within the population, that is the number and proportion of different individual phages, is defined as microdiversity. The highest level of organization is the community, which is represented by all the phage populations in a particular sample. The number and the abundances of these populations are defined as macrodiversity. Usually, a viral community is formed from multiple viral populations, each with its own degree of microdiversity.

Diversity analysis showed that oceanic viral communities are grouped within meta-communities or ecological zones. In terms of diversity, this means that phage communities within one zone are more similar to each other than to phage communities from other zones. There are only five viral ecological zones within the world's oceans: (1) Arctic, (2) Antarctic, (3) bathypelagic (> 2000 m depth), (4) temperate and tropical epipelagic (0–150 m depth) and (5) mesopelagic (150–1000 m depth). These zones are similar to those observed for the marine bacterial communities, indicating that the main driver of phage diversity is the bacterial community composition. The highest macrodiversity was found in the epipelagic and Arctic zones, and the lowest in the Antarctic zone. The highest microdiversity was found in the mesopelagic and epipelagic zones, and the lowest in the Arctic. Furthermore, the epipelagic and the Arctic zones contained the higher number of zone-specific populations, which indicates high levels of endemism and suggest that these regions are virus biodiversity cradles.

Marine Phages as Factors Driving Bacterial Mortality and Diversity

Until the late 80's, when it was discovered that marine phages are highly abundant, it was believed that the main mortality factor for marine bacteria is grazing by protists. Now, it is recognized that viral lysis contributes significantly to bacterial mortality, shaping microbial community composition and diversity. It is estimated that 10 and 40% of bacterial cells are lysed by viruses.

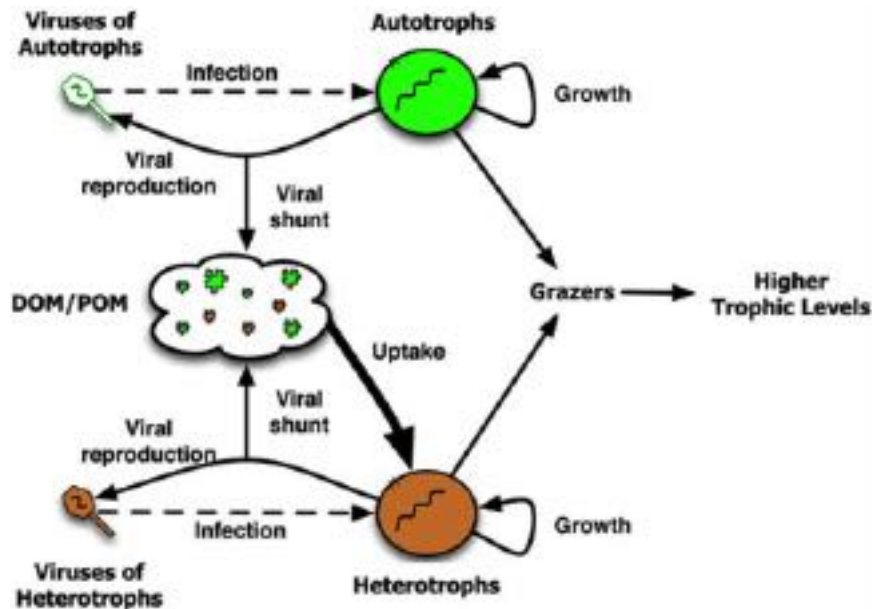


Fig. 12 Viral shunt in the marine environment. Reprinted with permission from Weitz, J.S., Wilhelm, S.W., 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports* 4, 17. Used under CC BY (<http://creativecommons.org/licenses/by/4.0/>). No changes made.

Phage infections in the oceans are believed to follow a kill the winner model. In this model, phage infections are stimulated by increasing host cell densities and in turn, lead to the host population crash. In the same time, the coevolution of bacterial hosts and phages is a veritable arms race, in which the hosts find ways to evade the attack of phages and the phages evolve to counter the new defenses, to ensure that neither the predator nor the prey is driven to extinction.

This arms race results in a high microdiversity of both host and phage. This was shown in chemostat experiments in which a single bacterial strain and its phage were incubated for prolonged periods. Both heterotrophic hosts, e.g., *Cellulophaga baltica* MM#3, and photosynthetic hosts, e.g., *Synechococcus* sp. WH7803, were subjected to such experiments (Middelboe *et al.*, 2009; Marston *et al.*, 2012). In both cases, samples collected at different time points showed the evolution of several host and phage strains, with different degrees of resistance and virulence, respectively.

The mechanisms used by marine bacteria to escape phage predation are multiple, and still largely uncharacterized. In *Prochlorococcus*, resistance to phages is conferred by mutations impairing the attachment of phages to the cell surface. These mutations are localized especially in non-conserved, horizontally transferred genes from a single hypervariable genomic island, genes probably involved in the synthesis of phage receptors. Although these mutations protect *Prochlorococcus* from certain phages, they also decrease its growth rate and even allow more rapid infection by other phages. Phage-resistant strains continue to evolve, until they reach an improved growth rate and a narrower resistance range. This mechanism could explain why in nature many *Prochlorococcus* cells have high growth rates and are resistant to phages.

Bacterial lysis leads to the release of both dissolved organic matter (DOM) and of particulate organic matter (POM) into the environment. The released matter, especially the DOM, is re-utilized by heterotrophic microorganisms, instead of being transferred to upper trophic levels via grazing (see Fig. 12). This phenomenon is termed the “viral shunt”. Part of the released POM will likely aggregate because cell debris resulted from cell lysis are sticky, and will be exported to the deep ocean. Therefore, it has been hypothesized that phages participate in the biological pump through the stimulation of particle formation. Recently, it has been shown that *Synechococcus* and their phages are strongly associated with carbon export at 150 m in the subtropical, nutrient-depleted, oligotrophic ocean.

Diel Rhythms of Phage Infections in the Marine Environment

Most organisms have a circadian rhythm, in which physiological processes are regulated in a ~24 h cycle. These diel oscillations are observed both in pure cultures in the laboratory and in the marine environment as well. In unicellular cyanobacteria, the circadian clock directs oscillations in gene transcription levels, with the majority of transcripts peaking at dawn or dusk. For example, in *Prochlorococcus*, the photosystem I and II genes and the Calvin cycle genes are maximally transcribed at dawn, most likely in preparation for harvesting the light energy through the day. Because unicellular cyanobacteria are main primary producers in the oceans, most likely their diel rhythms influence the diel rhythms of the co-occurring heterotrophic bacteria and of their phages.

In laboratory settings, cyanophage infections are influenced by light at different steps in the infection cycle. Some cyanophages show light-dependent absorption onto their host cells, with a higher fraction of phages absorbing in the light than in the dark. Phage replication and burst size are positively correlated with light intensity, indicating that cyanophages use for their replication resources generated by the photosynthetic metabolism of their host cells.

In the marine environment, the influence of diel cycles on the host and phages was studied in the North Pacific Subtropical Gyre, a habitat representative of oligotrophic oceans. During an eight-day experiment, metagenomics and quantitative meta-transcriptomics were used to monitor the temporal abundances and transcriptional activities of most abundant dsDNA viruses. A small fraction of the assembled viral scaffolds showed a diel periodicity in their transcript abundance, with a peak in the afternoon to early evening. Most likely other viruses, with lower abundances, were subjected to diel rhythms, but the methods used were not sensitive enough to detect their transcripts. Amongst the diel transcripts, those for structural viral genes were highly abundant, indicating that virus production was following a diurnal lytic cycle. The majority of viral scaffolds showing a diel transcriptional activity were attributed to cyanophages, most likely infecting *Prochlorococcus*. The replication and cell division of *Prochlorococcus* also followed a diel periodicity, both processes peaking in the later afternoon to dusk. This indicates connectivity in between the diel cycle of *Prochlorococcus* and of their phages. Three possible scenarios can be envisioned to explain the peak in cyanophage transcripts in the afternoon. In the first, cyanophages infect their hosts through the 24-h cycle, but with different burst sizes. Because the host has more metabolic resources in the afternoon, those phages whose infection cycles overlap with the afternoon period have the highest burst sizes. In the second scenario, cyanophages released in the afternoon burst will infect their hosts during the night, but not advance in their infection cycle until the afternoon increase in host photosynthetic metabolism and DNA replication. In the third scenario, cyanophages absorption is enhanced by light, leading to a higher infection rate during the day and thus to an afternoon burst peak. Not only cyanophages exhibited diel variations in their transcript levels. Amongst the viral scaffolds with diel periodicity, one most likely represented a *Pelagibacter* phage. This indicates that not only phages of photosynthetic bacteria follow a diel cycle, but also those of heterotrophic bacteria.

Auxiliary Metabolic Genes in Marine Phages

With the increased availability of genome sequencing technologies, more and more phages have been genome sequenced. As expected, most of the genes found in phage genomes are essential for the production of new progeny, as for example those involved in nucleic acid replication, capsid, and tail assembly, cell lysis. However, much to our surprise, in the early 2000s it was discovered that cyanophages carry host genes involved in photosynthesis. Since then, more than 200 different genes were found which are not directly involved in phage replication, but rather are part of the host metabolic pathways. Those genes have been named auxiliary metabolic genes (AMGs). Not included in AMGs are genes involved in DNA packaging, nucleotide transport and metabolism, protein metabolism and assembly, DNA synthesis, replication, and repair. AMGs are classified in 2 classes: (1) class I includes AMGs present in the KEGG metabolic pathways and (2) class II includes proteins assigned only a general metabolic function or marginal metabolism (e.g., transport function) and thus, not present in KEGG.

The first photosynthetic AMGs discovered were *psbA* and *psbD*, encoding two major components of the photosystem II (PSII). They were found in the genome of S-PM2 myocyanophage, which infects *Synechococcus*. Since then, several other photosynthetic AMGs were detected. Environmental surveys show that they are common in marine picocyanophages and that they were likely acquired by horizontal gene transfer from infected hosts. Phage PSII genes are actively transcribed in *Prochlorococcus* infected with P-SSP7 podovirus. Likely, they are functional in photosynthesis and increase phage fitness. Recently, genes for both photosystem I and II were discovered in P-TIM68 myovirus infecting *Prochlorococcus*. These genes are expressed during phage infection and the encoded proteins are incorporated into host membranes. Not only is the host photosynthetic capacity maintained during phage infection, but the cycling electron flow in the photosystem I is enhanced. This suggests an increased production of ATP in infected cells, required probably by the phage replication machinery.

With the increase in the number of cultivated and environmental phage genomes, more and more AMGs have been found. Many of them are potentially involved in different biogeochemical cycles in the oceans. For example, *dsrC*, a gene belonging to the dissimilatory sulfur reductase operon, is usually found in sulfate/sulfite-reducing bacteria in anoxic environments and in sulfur-oxidizing bacteria in both oxic and anoxic environments. Recently, a *dsrC* gene was found on several viral contigs from the Global Ocean Virome. Another gene from the sulfur cycle, *soxYZ*, was found as well on phage contigs from the GOV dataset. All contigs with *dsrC* and *soxYZ* genes belonged to the T4-like viral cluster, one of the most abundant and ubiquitous clusters in the oceans. These findings suggest that marine phages manipulate sulfur cycling. Another example of AMGs is the *P-II* gene, which is involved in the regulation of the nitrogen metabolism in bacteria. It was found on several phage contigs in the GOV dataset, suggesting phage involvement in marine nitrogen cycling.

Transfer RNAs (tRNAs) in Phages and Their Role in Phage-Host Interactions

Transfer RNAs are present in the genomes of many phages, including cyanophages. These RNAs recognize specific codons in the mRNA and transfer the corresponding amino acid for incorporation into the growing protein chain. Due to the genetic code degeneracy, part of the 20 amino acids are encoded by multiple codons. In a particular bacterial host, depending on its genome G + C content, certain codons can be more frequent than others. As a result, the cellular tRNA pool can be biased toward the most

frequent codons. Generally, phages infect hosts with similar codon usage, because otherwise, the protein synthesis would be inefficient. To bypass differences in codon usage phages can encode tRNAs in their own genomes. Some cyanophages use this strategy to broaden their host range. For example, cyanomyophages with a higher number of tRNAs in their genomes are able to infect both low G + C content hosts (*Prochlorococcus*) and high G + C content hosts (*Synechococcus*). On the other hand, cyanomyophages with no tRNAs or a low number of tRNAs in their genomes are able to infect only *Prochlorococcus*. The number of tRNAs in the phage genomes is quite variable (see Table 1). The cultivated phage with the highest number of tRNA genes is the S-CBWM1 cyanophages. It has 34 tRNA genes in its genome. And potentially used these tRNAs to increase the translation efficiency for several unique phage genes, which have no homologs in sequences databases.

Conclusions

Phages are most likely present in all marine environments. They regulate the metabolism, structure, and composition of bacterial communities, thus impacting globally the marine biogeochemical cycles. Significant advancements in sequencing technologies have allowed characterization of marine phages on an unprecedented scale. Even so, much more remains to be known. For example, the host is unknown for most of the environmental phage contigs. As opposed to the sunlit ocean, the phage diversity of the deep-ocean and sediments is less known. Similarly, knowledge is sparse regarding marine ssDNA and RNA phages, as compared to dsDNA phages. Phage-host interactions have been studied only for a handful of phage-host systems. Also, little is known about lysogeny, how it shapes microbial communities and which factors are responsible for the lytic/lysogenic switch. To answer all these questions, a combination of cultivation-dependent and independent methods is necessary.

References

- Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., *et al.*, 2015. Patterns and ecological drivers of ocean viral communities. *Science* 348 (6237), 1261498. doi:10.1126/science.1261498.
- Marston, M.F., Pierciey, F.J., Shepard, A., *et al.*, 2012. Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proceedings of the National Academy of Sciences of the United States of America* 109 (12), 4544–4549. doi:10.1073/pnas.1120310109.
- Middelboe, M., Holmfeldt, K., Riemann, L., Nybroe, O., Haaber, J., 2009. Bacteriophages drive strain diversification in a marine *Flavobacterium*: Implications for phage resistance and physiological properties. *Environmental Microbiology* 11 (8), 1971–1982. doi:10.1111/j.1462-2920.2009.01920.x.
- Roux, S., Brum, J.R., Dutilh, B.E., *et al.*, 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537 (7622), 689–693. doi:10.1038/nature19366.

Further Reading

- Allers, E., Moraru, C., Duhaime, M.B., *et al.*, 2013. Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environmental Microbiology* 15 (8), 2306–2318. doi:10.1111/1462-2920.12100.
- Aylward, F.O., Boeuf, D., Mende, D.R., *et al.*, 2017. Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proceedings of the National Academy of Sciences of the United States of America* 114 (43), 11446–11451. doi:10.1073/pnas.1714821114.
- Bischoff, V., Bunk, B., Meier-Kolthoff, J.P., *et al.*, 2019. Cobaviruses – A new globally distributed phage group infecting *Rhodobacteraceae* in marine ecosystems. *The ISME Journal* 13, 1404–1421. doi:10.1038/s41396-019-0362-7.
- Coutinho, F.H., Silveira, C.B., Gregoracci, G.B., *et al.*, 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications* 8, 15955. doi:10.1038/ncomms15955.
- Duhaime, M.B., Solonenko, N., Roux, S., *et al.*, 2017. Comparative omics and trait analyses of marine *Pseudoalteromonas* phages advance the phage OTU concept. *Frontiers in Microbiology* 8, 1241. doi:10.3389/fmicb.2017.01241.
- Enav, H., Kirzner, S., Lindell, D., Mandel-Gutfreund, Y., Béjà, O., 2018. Adaptation to sub-optimal hosts is a driver of viral diversification in the ocean. *Nature Communications* 9 (1), 4698. doi:10.1038/s41467-018-07164-3.
- Gregory, A.C., Solonenko, S.A., Ignacio-Espinoza, J.C., *et al.*, 2016. Genomic differentiation among wild cyanophages despite widespread horizontal gene transfer. *BMC Genomics* 17 (1), 930. doi:10.1186/s12864-016-3286-x.
- Gregory, A.C., Zayed, A.A., Conceição-Neto, N., *et al.*, 2019. Marine DNA viral macro- and microdiversity from pole to pole. *Cell* 177, 1109–1123. doi:10.1016/j.cell.2019.03.040.
- Holmfeldt, K., Solonenko, N., Howard-Varona, C., *et al.*, 2016. Large-scale maps of variable infection efficiencies in aquatic Bacteroidetes phage-host model systems. *Environmental Microbiology* 18 (11), 3949–3961. doi:10.1111/1462-2920.13392.
- Holmfeldt, K., Solonenko, N., Shah, M., *et al.*, 2013. Twelve previously unknown phage genera are ubiquitous in global oceans. *Proceedings of the National Academy of Sciences of the United States of America* 110 (31), 12798–12803. doi:10.1073/pnas.1305956110.
- Hurwitz, B.L., Brum, J.R., Sullivan, M.B., 2015. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *The ISME Journal* 9 (2), 472–484. doi:10.1038/ismej.2014.143.
- Marston, M.F., Martiny, J.B.H., 2016. Genomic diversification of marine cyanophages into stable ecotypes. *Environmental Microbiology* 18 (11), 4240–4253. doi:10.1111/1462-2920.13556.
- Martínez-Hernández, F., Fornas, O., Lluesma Gomez, M., *et al.*, 2017. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nature Communications* 8, 15892. doi:10.1038/ncomms15892.
- Mizuno, C.M., Ghai, R., Saghai, A., López-García, P., Rodríguez-Valera, F., 2016. Genomes of abundant and widespread viruses from the deep ocean. *mBio* 7 (4), doi:10.1128/mBio.00805-16.
- Thompson, L.R., Zeng, Q., Kelly, L., *et al.*, 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences of the United States of America* 108 (39), E757–E764. doi:10.1073/pnas.1102164108.
- Weitz, J.S., Wilhelm, S.W., 2012. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biology Reports* 4, 17. doi:10.3410/B4-17.

Ecology of Phages in Extreme Environments

Tatiana A Demina, Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Nina S Atanasova, Finnish Meteorological Institute, Helsinki, Finland and University of Helsinki, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

CRISPR Clustered regularly interspaced short palindromic repeats

Cryo-EM Cryogenic electron microscopy

ds Double-stranded

MCP Major capsid protein

NaCl Sodium chloride

ss Single-stranded

T number Triangulation number

T4-like Virus morphology resembling bacterial myovirus T4

T7-like Virus morphology resembling bacterial podovirus T7

UV Ultra violet

VLP Virus-like particle

VPR Virus-to-prokaryote ratio

Glossary

Caudovirales Taxonomical order of viruses, which contains viruses with icosahedral heads, different types of tails, and linear dsDNA genomes.

Cryoconite hole Water-filled melt-hole on the surface of a glacier.

Cryogenic electron microscopy Electron microscopy technique where samples are rapidly frozen and then observed at cryogenic temperatures.

Cryopeg Water brine within permafrost.

Cryosphere The areas of the Earth where water is in solid state.

Deep-sea hydrothermal vent Geothermal vents at the seafloor of volcanically active areas.

Extremophile Organism that inhabits extreme environment, where one or more physico-chemical parameters reach their extreme values.

Halophile Organism that requires high salt concentrations for the optimal growth.

Hyperarid Extremely dry, with an aridity index below 0.05.

Metagenomics Metatranscriptomics, and metaproteomics, studies of pools of genomes, mRNA, and proteins, respectively, derived directly from environmental samples.

Myovirus Virus with an icosahedral head and a contractile tail.

Permafrost Ground that has remained frozen for at least two years' time period.

Phage Virus infecting bacteria.

Podovirus Virus with an icosahedral head and a short non-contractile tail.

Polyextreme environment Environment with two or more physico-chemical parameters reaching their extreme values.

Prophage Phage genome integrated into the host cell DNA or existing as a plasmid in the host cell cytoplasm.

Siphovirus Virus with an icosahedral head and a long non-contractile tail.

Triangulation number Square of the distance between two adjacent five-fold vertices of an icosahedral viral capsid.

Virome Specific pool of nucleic acids that constitute a viral community within an ecosystem.

Virophage Small dsDNA virus whose replication is dependent on the co-infecting giant virus.

Virus-like particle In microscopy: particle that morphologically resembles a virus.

Vitrification Rapid cooling process that suppresses the formation of ice crystals.

Introduction

The availability of liquid water within a cell can be considered as one of the prerequisites for life, enabling a proper functioning of essential cellular processes, such as enzymatic reactions and nucleic acid metabolism. These requirements are easily met in an environment with 20–30°C temperature, 1 atm pressure, and available water. However, in the Earth's extreme environments, physico-chemical parameters like temperature, pressure, salinity, ultra violet (UV) radiation and pH reach extreme values. The word *extreme* is derived from the Latin *extrēmus*, which means uttermost, utmost or last. Indeed, the conditions for life in extreme environments can probably best be described as “at the limit”. Majority of organisms cannot tolerate such conditions, indicating that different types of extreme environments may have low species diversity and some large taxonomic groups of organisms can be totally absent.

Nonetheless, several extreme environments on the Earth, such as polar sea ice, hypersaline environments, volcanic springs, and deep-sea hydrothermal vents, are actually flourishing with life. These environments are inhabited by extremophiles, organisms that

not only tolerate, but also require the surrounding extreme conditions for survival. Most extremophiles are microorganisms. Archaea often dominate in various extreme environments, but bacteria and eukaryotes can also be abundant.

Viruses of extremophiles exceed the number of their archaeal, bacterial or eukaryotic host cells by ten- to hundred-fold and are often the only predators in various ecosystems of extreme environments. Due to the low species diversity of their host organisms, viruses of extremophiles are important evolutionary players in shaping the dynamics and structure of microbial communities. Even today, extreme environments remain largely unexplored territories and the overall number of extremophilic phage isolates is low compared to the non-extremophilic ones. Genome sequences of the described phages of extremophiles, as well as metagenomes obtained from extreme environments typically contain low number of hits to common databases. Detailed molecular studies of virus-host interactions in extreme environments are scarce and much more research is needed in order to better understand these unique phage-host systems. Moreover, some extreme environments, such as volcanic hot springs, are suggested to resemble the conditions on the early Earth and could thus provide valuable information about the origin and evolution of life. In addition, studies on extremophiles and their viruses are also important regarding astrobiology, i.e., how microorganisms could survive in extra-terrestrial conditions. The research on viruses and their extremophilic hosts can also add insights into the biodiversity on our planet, providing knowledge needed for predicting the effects of climate change.

This article concentrates on extremophilic bacterial viruses, *phages*, but archaeal viruses are also mentioned when relevant. The available data on extremophilic phages reviewed here have been obtained by different methods, each of which has its own limitations. For example, by electron microscopy, virus-like particles (VLPs), cellular membrane vesicles or staining artefacts may be visualized together with bona fide viruses. For simplicity, the different environments reviewed here have been categorized into e.g., thermal, cold, or hypersaline. In fact, several of the introduced environments are polyextreme, meaning that they are subjected to two or more extreme parameters. As with all the other studied environments, phages with icosahedral heads, different types of tails and linear double-stranded (ds) DNA genomes dominate in most extreme environments, while still very little is known about extremophilic enveloped viruses or viruses with RNA genomes. We expect that future research on extremophilic phages will reveal many novel structures and functions helping to understand the life around us.

Phages in Hypersaline Environments

Hypersaline environments are characterized by high salt concentrations. In such environments, the dominant salt is usually NaCl and the overall proportion of ions corresponds to that in seawater, but some hypersaline waters, e.g., the Dead Sea, may have ions in other proportions. The examples of hypersaline environments are solar salterns, brine springs, hypersaline lakes and ponds, deep-sea brine pools, halite deposits, salt mines, saline soils, and salty food products. *Halophilic*, i.e., “salt-loving”, organisms flourish in hypersaline environments. Although some marine organisms may be slightly halophilic, here, by *halophiles* we refer to organisms that are found in a salinity higher than that of seawater. *Halophiles* may be slight, moderate, or extreme, requiring 1%–3% (0.2–0.5 M), 3%–15% (0.5–2.5 M), or 15%–30% (2.5–5.2 M) NaCl, respectively, for optimal growth. Halophiles are known within all three domains of life, but archaea and bacteria are especially abundant in hypersaline environments.

High numbers of VLPs (up to 10^9 VLPs ml^{-1}) of various morphologies, including spindle-shaped, spherical, tailed, filamentous, as well as some exceptional shapes, such as stars and hooks, have been reported for hypersaline environments. For example, high VLP densities (10^8 – 10^9 VLPs ml^{-1}) and high abundance of tailed icosahedral VLPs, typical for the phages belonging to the order *Caudovirales*, have been found in the alkaline (pH ~10), hypersaline Mono Lake, USA (Fig. 1). In those samples, siphoviruses and myovirus-like particles were the most abundant among tailed VLPs, but podoviruses were also present. However, the VLPs observed in hypersaline environments, especially those of unusual forms, may represent mostly archaeal viruses, given that archaea usually dominate at high salt concentrations. Most of the phages described so far display tailed icosahedral particles, but in hypersaline environments, even tailed icosahedral particles may also refer to archaeal tailed viruses.

Metagenomic, metatranscriptomic, and metaproteomic studies have revealed that viral communities are dynamic, specific, and diverse in hypersaline environments. Metagenomics of the sediment samples from the Great Salt Lake (15% total salinity, pH 7.4), USA, revealed that bacteria dominated in the prokaryotic community, with *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Chloroflexi*, *Cyanobacteria*, and *Planctomycetes* as the most dominant phyla. In the viral fraction of the metagenome, sequences related to the members of the order *Caudovirales* dominated, with lambda-like, T4-like, and *Cellulophaga*-associated phages being particularly abundant. The sequences of tailed phages have been also found in other hypersaline environments, e.g., deep-sea anoxic brines of Red Sea or cold hypersaline Deep Lake in the Antarctica. The presence of prophage sequences, viral functional genes, and virus defense systems, such as clustered regularly interspaced short palindromic repeats (CRISPRs), in bacterial genomes retrieved from high salinity environments suggests active viral-mediated gene transfer and dynamic virus-host interactions.

At least 31 halophilic phages have been isolated to date (Table 1). These phages infect hosts of various genera and have been isolated from distant hypersaline environments all over the world or induced from halophilic bacterial strains. The first eight viruses infecting *Salinibacter ruber*, an extremely halophilic bacterium commonly found in hypersaline environments, have been isolated only recently, while previous studies of metavirome and metatranscriptome from Santa Pola solar saltern, Spain, have already suggested the presence of active phages infecting *S. ruber*. A few phages have been isolated from soda lakes, which are characterized by high sodium carbonate concentrations and high pH values. Both virulent and temperate halophilic phage isolates are known, although for many of them, life cycles have not been studied in sufficient detail. Most of the isolated halophages

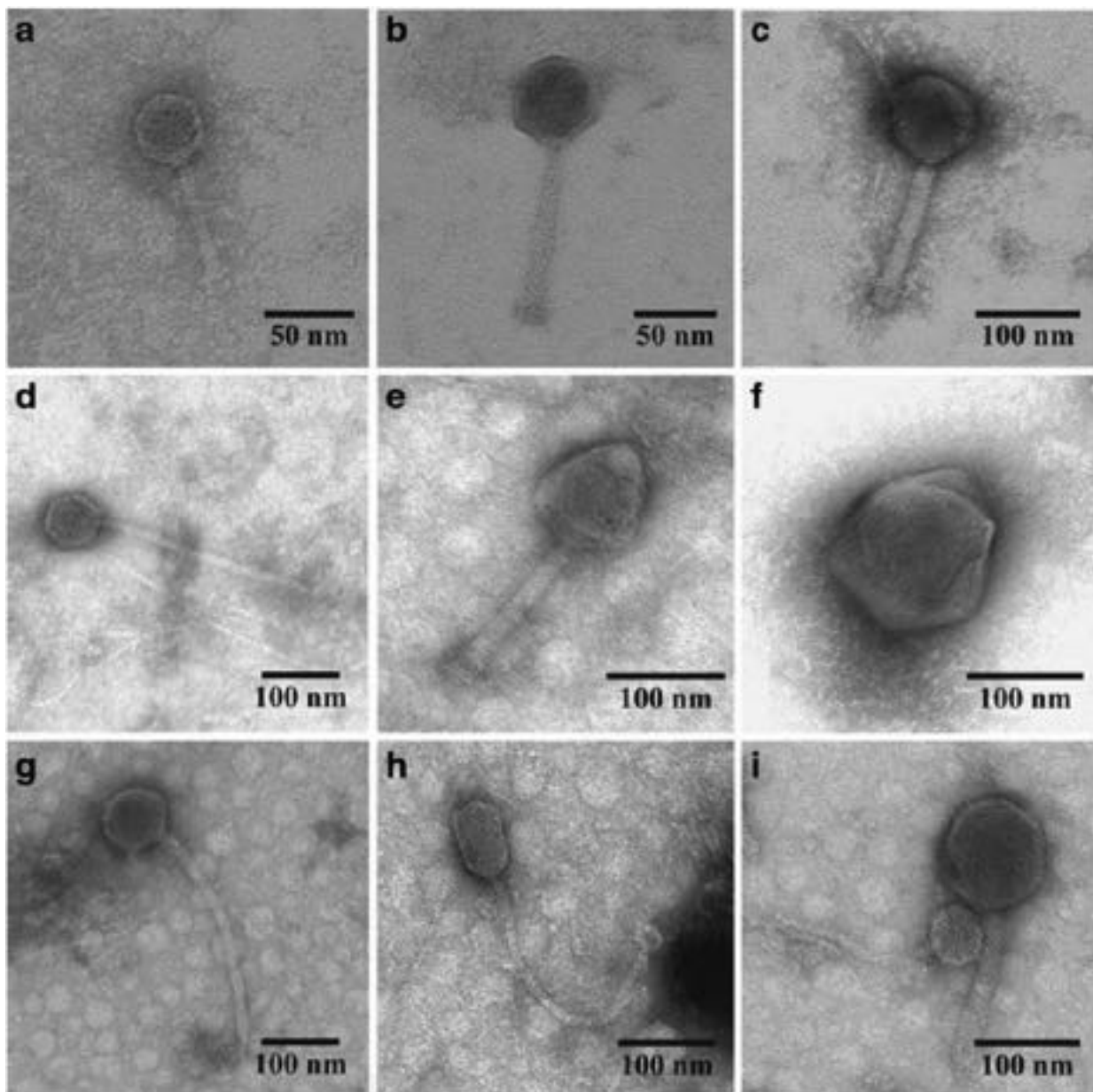


Fig. 1 Transmission electron micrographs of viruses from depths of 2 m (A–C), 16.5 m (D–F), and 35 m (G–I) in Mono Lake. Reproduced from Brum, J.R., Steward, G.F., 2010. Morphological characterization of viruses in the stratified water column of alkaline, hypersaline Mono Lake. *Microbial Ecology* 60 (3), 636–643. doi:10.1007/s00248-010-9688-4.

display tailed icosahedral particles, typical for phages classified into the order *Caudovirales*. As for other morphotypes, one tailless icosahedral virus, SSIP-1, which infects *Salisaeta* sp., was isolated from Sedom Ponds, Israel, and one spherical virus, JMT-1, which infects *Chromohalobacter*, was isolated from Yuncheng Saline Lake, China. For many isolates, there are no available data on their genome organization, others have dsDNA genomes in either linear or circular conformation and of variable length (Table 1).

High-resolution image reconstructions based on cryogenic electron microscopy (cryo-EM) revealed that the icosahedral capsid of CW02, podovirus infecting *Salinivibrio*, is decorated with turret-like structures at icosahedral vertices and its major capsid protein (MCP) has a conserved HK97-fold. A so-called canonical HK97-fold, which was originally observed in phage HK97, is found in tailed phages, eukaryotic herpesviruses and in HSTV-1, an archaeal tailed virus. The conservation of certain protein folds and overall virion architectural principles observed across various viral groups has led to the idea that all viruses may be in principle classified into a limited number of structure-based lineages, including the HK97-like lineage. The other halophilic tailless icosahedral bacteriophage, SSIP-1, (Fig. 2) displays structural features typical for viruses grouped into the PRD1-adenovirus-like structural lineage. This lineage comprises tailless icosahedral inner-membrane containing (except membraneless adenovirus) dsDNA viruses, including both prokaryotic and eukaryotic ones, as well as virophages. Viruses belonging to the PRD1-adenovirus-like structural lineage are either extremophilic or

Table 1 Examples of phages isolated from high salinity environments

<i>Virus name</i>	<i>Host</i>	<i>Isolation site</i>	<i>Particle morphology</i>	<i>Genome (type and size)</i>
F9-11	<i>Halomonas halophila</i>	Induced from <i>H. halophila</i> F9–11, Alicante, Spain	Siphovirus	Nd ^a
G3	<i>Pseudomonas</i> sp. G3	Salt pond, Canada	Myovirus	Nd
HM5	Isolate 131 from La Mala saltern	La Mala saltern, Granada, Spain	Siphovirus	Nd
HM15	Isolate 121 from La Mala saltern	La Mala saltern, Granada, Spain	Siphovirus	Nd
phi7116	<i>Tetragenococcus halophilus</i>	Fermenting soy sauce	Myovirus	Nd
phiD-86	<i>Tetragenococcus halophilus</i>	Fermenting soy sauce	Siphovirus	Nd
F5–4	<i>Halomonas halophila</i>	Induced from <i>H. halophila</i> (Alicante, Spain)	Siphovirus	Nd
F12–9	<i>Halomonas halophila</i>	Induced from <i>H. halophila</i> , Alicante, Spain	Siphovirus	Nd
UTAK	<i>Vibrio</i> sp. B1	Solar saltern, Alicante, Spain	Myovirus	dsDNA, ~80 kbp
phiMono1	<i>Idiomarina</i> sp.	Mono Lake, USA	Nd	Nd
phigspB	<i>Halomonas salina</i>	Saline soil, Great Salt Plains, USA	Myovirus	dsDNA, ~49 kbp
phigspC	<i>Halomonas salina</i>	Saline soil, Great Salt Plains, USA	Myovirus	dsDNA, ~340 kbp
SCTP-1	<i>Salicola</i> sp. PV3	Solar saltern, Margherita di Savoia, Italy	Siphovirus	Nd
SCTP-2	<i>Salicola</i> sp. PV4	Solar saltern, Margherita di Savoia, Italy	Myovirus	Nd
Ø1M3–16	<i>Vibrio metschnikovii</i>	Lake Magadi, Kenya	Siphovirus	dsDNA, <30 kbp
SCTP-3	<i>Salicola</i> sp. s3–1	Solar saltern, Margherita di Savoia, Italy	Myovirus	Nd
CW02	<i>Salinivibrio</i> SA50	Great Salt Lake, USA	Podovirus	Linear dsDNA, 49,390 bp
SSIP-1	<i>Salisaeta</i> sp. SP9–1	Salt water, Sedom Ponds, Israel	Tailless icosahedral with inner lipid membrane	Circular dsDNA ~43,788 bp
QHHSV-1	<i>Halomonas ventosae</i>	Salt mine, Qiaohou, China	Siphovirus	dsDNA, nd
Shbh1	<i>Bacillus</i> sp. MGK1/ERV9	Lake Shala, Ethiopia	Myovirus	Linear dsDNA, 138,081 bp
Shpa	<i>Paracoccus</i> sp. HS3	Lake Shala, Ethiopia	Siphovirus	Linear dsDNA, 38,261 bp
Mgbh1	<i>Bacillus</i> sp. MGK1	Lake Magadi, Kenya	Siphovirus	Linear dsDNA, 58,951 bp
JMT-1	<i>Chromohalobacter</i> sp. LY7–3	Yuncheng Saline Lake, China	Spherical	Linear dsDNA, ~23 kbp
M8CC-19	<i>Salinibacter ruber</i> M8	Bras del Port saltern, Alicante, Spain	Siphovirus	Linear dsDNA, 53,812 bp
M31CC-1	<i>Salinibacter ruber</i> M31	Bras del Port saltern, Alicante, Spain	Siphovirus	Linear dsDNA, 53,808 bp
M31CR41–2	<i>Salinibacter ruber</i> M31	Bras del Port saltern, Alicante, Spain	Siphovirus	Linear dsDNA, 51,326 bp
M31CR41–3	<i>Salinibacter ruber</i> M31	Bras del Port saltern, Alicante, Spain	Siphovirus	Linear dsDNA, 50,128 bp
M8CR30–2	<i>Salinibacter ruber</i> M8	Bras del Port saltern, Alicante, Spain	Siphovirus	Linear dsDNA, 35,009 bp
M8CR30–4	<i>Salinibacter ruber</i> M8	Bras del Port saltern, Alicante, Spain	Siphovirus	Linear dsDNA, 35,434 bp
M8CRM-1	<i>Salinibacter ruber</i> M8	Bras del Port saltern, Alicante, Spain	Siphovirus	Linear dsDNA, 53,197 bp
M1EM-1	<i>Salinibacter ruber</i> M1	Campos saltern, Mallorca, Spain	Siphovirus	Linear dsDNA, 35,883 bp

^aNd, not determined.

moderate, e.g., phages PRD1, PM2, P23-77 (see below), archaeal viruses STIV, SH1, HHIV-2, HCIV-1, algal virus PBCV-1, human adenovirus, and the virophage Sputnik.

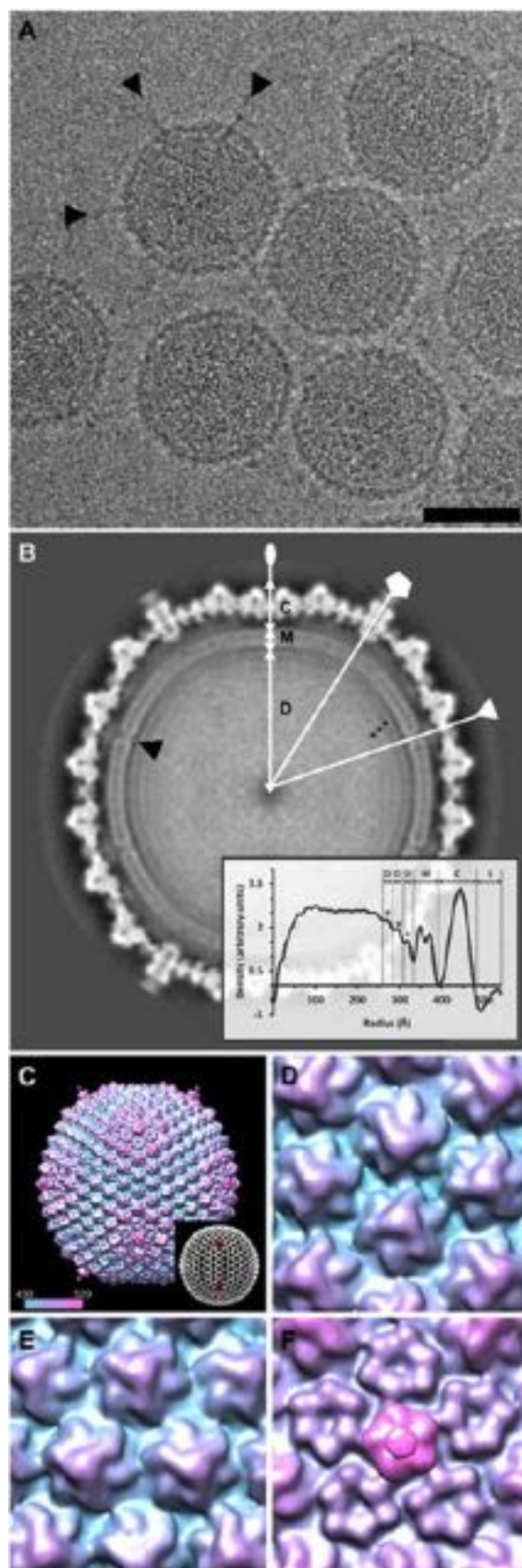
Phages in Thermal Environments

Thermophilic, i.e., “heat-loving”, organisms inhabit hot environments, where temperature ranges from ~40 to ~80°C, while *hyperthermophiles* grow optimally at temperatures above 80°C. Such hot environments include terrestrial volcanic sites, non-polar deserts, hydrothermal springs, submarine hydrothermal systems, and oceanic basement subsurface. In addition, thermophilic microorganisms can be found from various artificially made environments, such as dairy production plants, water heaters, and compost piles.

Hyperthermophilic archaea tend to dominate in the hottest environments (at above 70°C), while thermophilic bacteria are often dominant at milder temperatures (~40–70°C). Viruses are abundant and active players in microbial communities inhabiting thermal terrestrial and marine sites. Moreover, in thermal environments where temperature exceeds the upper limit for eukaryotic life (~60°C), prokaryotic viruses are the only predators, controlling host abundance and affecting organic and inorganic nutrient cycling.

Terrestrial Hot Springs

Terrestrial thermal springs are sites where geothermally heated groundwater is released at the land surface. These sites are characterized by temperatures of ~35°C up to the boiling point, a pH range of 1–9, and various ion composition. Thermal springs are found all over the planet, but some regions, such as Yellowstone National Park in the USA, are particularly rich in hot springs and geysers.



Quite different VLP densities have been reported for hot springs, ranging roughly from 10^4 to 10^7 VLPs ml^{-1} . Virus-to-prokaryote ratios (VPRs) detected in hot springs are typically lower when compared to moderate environments. Electron microscopic examinations of VLPs in Yellowstone hot springs revealed a variety of different morphotypes, including rod-shaped, lemon-shaped, and filamentous particles typical for hyperthermophilic archaea, novel morphotypes (zipper-like, pleomorphic, and other complex particles), and tailed particles. The latter ones were observed only in neutral or mildly alkaline springs and may represent bacteriophages or archaeal viruses. A complete genome sequence of the OS3173 phage, which presumably infects *Aquificales*, was retrieved from a metavirome from the alkaline Octopus hot spring (pH ~ 8.5), Yellowstone. Metagenomic studies of acidic springs in Iceland, Italy, and Yellowstone National park (76–90°C, pH 1.8–5.5) showed the overall predominance of archaeal viruses over bacteriophages. However, the fraction of phages might have been underestimated due to a bias of reference databases, which contain significantly more hyperthermophilic archaeal viral sequences compared to those of thermophilic bacteriophages. Sequences related to the members of the order *Caudovirales* were abundant in samples obtained from a hot spring in Grensdalur, Iceland (85°C, pH 5), and a partial genome of a presumably *Hydrogenobaculum* phage, HP1, was reconstructed from the same sample.

Cyanophages have been reported to dominate in some hot springs. Microscopy analysis of Brandvlei hot spring (60°C, pH 5.7), South Africa, showed the presence of both archaea-specific VLPs and tailed particles. Lambda-like siphoviruses dominated, while podoviruses, as well as regular-sized and jumbo myoviruses, were also present. Metaviromic analysis demonstrated the prevalence of *Caudovirales*-related sequences and revealed a large fraction of genes typical for cyanophages. Partial genomes of two phages, cyanophage BHS3 and putatively *Gemmata* phage BHS4, were also obtained from Brandvlei hot spring metavirome. The predominance of tailed VLPs was also shown in cyanobacteria-dominated phototrophic mats in Porcelana hot spring (46–70°C, pH ~ 7), Chile (Fig. 3). Metagenomic analysis of the same samples confirmed the dominance of phages belonging to the order *Caudovirales* and allowed a full genome reconstruction of a T7-like thermophilic cyanophage (podovirus), TC-CHP58, which presumably infects *Mastigocladus* sp. However, no thermophilic cyanophages have yet been isolated.

The most numerous thermophages isolates have been obtained on *Thermus* strains (Table 2). In a large sampling of alkaline hot springs in Iceland, New Zealand, Russia (Kamchatka), and the USA, arranged by the Promega corporation, 115 phages were isolated on seven *Thermus* strains. About 45% of the obtained isolates were tailed phages (myo- and siphoviruses). In addition, tailless icosahedral and filamentous phages were present. In this sampling, two *Thermus thermophilus* siphoviruses, P23-45 and P74-26, with exceptionally long tails (~ 800 nm) were isolated from Kamchatka hot springs, Russia. Later, TSP4, a similar *Thermus* siphovirus with a 785-nm long tail, was isolated from Tengchong hot spring, China, suggesting a worldwide distribution of this unusual siphovirus type. In addition, another *Thermus thermophilus* siphovirus, G20c, isolated from Kamchatka hot springs shows high sequence similarity to P23-45 and P74-26.

Thermus phages P23-77 and IN93 display similarities to tailless icosahedral inner-membrane containing dsDNA viruses SH1, PH1, HHIV-2, HCIV-1, and SNJ1, which infect halophilic archaea. Consequently, these phages and archaeal viruses have been classified into one viral family, *Sphaerolipoviridae*. In addition, two other *Thermus* phages, P23-72 and P23-65 H, have been proposed to belong to the same family. Cryo-EM-based image reconstructions (Fig. 4) revealed that P23-77 virion consists of an icosahedral protein shell and an internal lipid membrane, which encloses the dsDNA genome. At the five-fold symmetry positions, P23-77 virion is decorated with spike complexes, which most probably serve for host recognition. The capsid lattice arrangement (pseudo-hexameric with $T = 28$ symmetry) seems to be conserved within the family *Sphaerolipoviridae*, while spike complexes display various forms. Overall virion structural organization, as well as MCP folding, places P23-77, together with other sphaerolipoviruses, within the PRD1-adenovirus structural lineage.

Deep-Sea Hydrothermal Vents

Seafloor had been considered as a desert-like environment, where life can be barely found, until the discovery of hydrothermal vent systems in the late 1970s. Since then, most studies on life in the deep ocean have concentrated around underwater thermal systems as the hotspots of life, surrounded by dark, cold and nutrient deficient ocean waters. Deep-sea hydrothermal vents are unique extreme environments with huge gradients of temperature, pH, and dissolved chemical compounds, not to mention high hydrostatic pressure. Vent fluids may reach as high temperatures as 400°C at the top of the chimney, but the temperature of fluids drops sharply, when they are mixed with the surrounding cold seawater (0–4°C). Due to the mixing of extremely hot and cold

Fig. 2 Icosahedral reconstruction of the SSIP-1 virion. (A) Electron micrograph of SSIP-1 virions vitrified at 9% (wt/vol) salt water (SW) buffer (see composition in Table S2 in original publication). Three spikes are indicated with black arrowheads (Scale bar, 50 nm.) (B) Central slice through an icosahedral reconstruction. *Inset* shows a radially averaged density profile. DNA (D), membrane (M), capsid (C), and spikes (S) are indicated. Twofold, threefold, and fivefold axes of icosahedral symmetry are indicated by an ellipse, a triangle, and a pentagon, respectively. Three concentric layers of DNA are indicated with asterisks. Lipid bilayer is interrupted by transmembrane densities at the threefold axes of symmetry (triangle). (C) Radially colored isosurface representation of the reconstruction with an arbitrary handedness is rendered at 2σ above the mean density. Color bar shows radial coloring. *Inset* shows a model lattice exemplifying the $T = 49$ icosahedral triangulation. Geometrical arrangement of the capsomers is given by the relationship $T = h^2 + hk + k^2$, where h and k define the lattice point. Here $h = 7$, $k = 0$. The two frontmost fivefold vertices are in red. (D–F) Six times magnified close-ups of the reconstruction taken along the (D) twofold, (E) threefold, and (F) fivefold axes of symmetry. Reproduced from Aalto, A.P., Bitto, D., Ravantti, J.J., *et al.*, 2012. Snapshot of virus evolution in hypersaline environments from the characterization of a membrane-containing *Salisaeta* icosahedral phage 1. *Proceedings of the National Academy of Sciences of the United States of America* 109 (18), 7079–7084. doi:10.1073/pnas.1120174109.

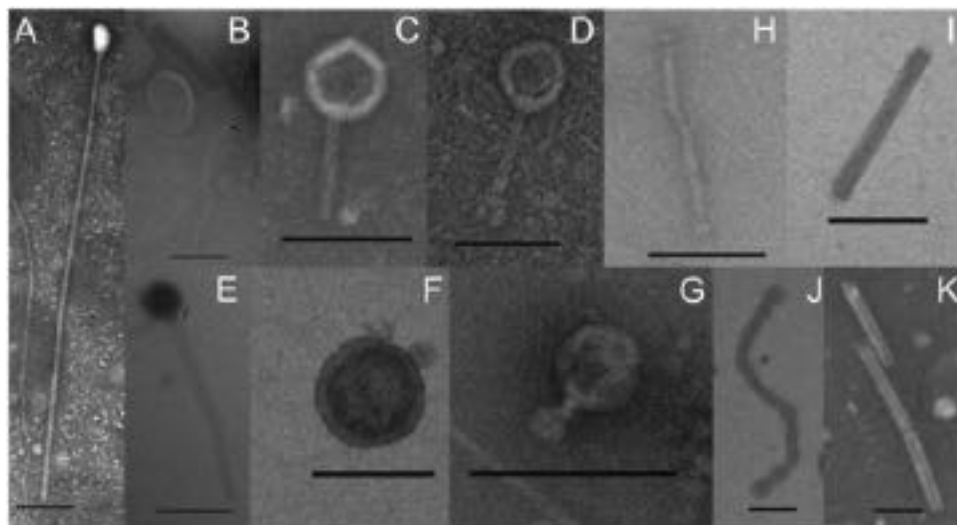


Fig. 3 Transmission electronic micrographs of VLPs obtained from the interstitial fluid of phototrophic microbial mats growing between 62°C and 42°C in Porcelana hot spring. Scale bar: 100 nm. (A–G) Caudovirus-like particles belonging to *Myoviridae*, *Podoviridae*, and *Siphoviridae* families. (H–K) Filamentous and rod shaped VLPs that could be associated with *Lipothrixviridae* and *Clavaviridae* families. Reproduced from Guajardo-Leiva, S., Pedrós-Alió, C., Salgado, O., Pinto, F., Díez, B., 2018. Active crossfire between cyanobacteria and cyanophages in phototrophic mat communities within hot springs. *Frontiers in Microbiology* 9, 2039. doi:10.3389/fmicb.2018.02039.

waters, a steep gradient of pH and fluid chemistry is formed within meters from the fluid emission site. As a result, a variety of microenvironments is formed within hydrothermal vents systems, hosting very diverse forms of life. Chemosynthesizing archaea and bacteria serve as primary producers in these communities.

Viruses are ubiquitous in both deep and shallow underwater vents, except that no VLPs have been found on the interior parts of chimney structures. Virus particles are typically more abundant in vent systems than in the surrounding seawater: 10^4 – 10^5 VLPs ml^{-1} observed in deep-sea waters and 10^5 – 10^7 VLPs ml^{-1} in hydrothermal vent systems. In deep-sea hydrothermal environments, VPR is typically not higher than that in seawater (≤ 10). Sequence and microscopy analyzes have revealed the presence of tailed bacteriophages in various hydrothermal vents and even hundreds of meters below the seabed surface. VLPs of various shapes were observed in the brackish-marine sediments of the Baltic Sea (Fig. 5), including tailless and tailed icosahedral particles and VLPs inside cells. The analysis of g23 gene, which encodes MCP of T4-like myoviruses, revealed diverse groups of T4-like phages, including cyanophages, in the Baltic Sea subsurface sediments.

Sampling of hydrothermal vents is highly difficult not only due to technical and logistical constraints, but also due to the very dynamic nature of these systems. Cells and viruses found in plume fluid samples may in fact originate from the surrounding seawater, surface chimney sediments, or subsurface seabed, as water masses are constantly in movement. Perhaps, this partially explains the fact that the available literature contains controversial reports regarding the specificity of viral community composition, when the vent and background seawater metagenomes are compared. Metagenomes obtained from hydrothermal vents display high enrichment with proviral sequences and mobile genetic elements in cells and auxiliary metabolic genes in viruses, suggesting that lysogeny is a common virus life strategy in these environments. Genomes of phages presumably infecting sulfur-oxidizing bacteria, which are ubiquitous marine chemolithoautotrophs, were retrieved from metagenomic datasets sampled from hydrothermal vent plumes in the Pacific Ocean. These dsDNA viral genomes contained auxiliary metabolic genes involved in elemental sulfur oxidation, suggesting a role in supplementing the metabolism of sulfur-oxidizing bacteria and thus affecting sulfur cycling in deep sea.

To the best of our knowledge, only eight phages infecting bacteria from deep-sea hydrothermal vents have been obtained to date (Table 2). All of them have been either induced or spontaneously isolated from the host cultures, i.e., no phages have been yet isolated directly from deep-sea hydrothermal vent samples. The known isolates are all tailed phages obtained from the strains belonging to *Bacillus*, *Geobacillus*, *Nitratiruptor*, and *Marinitoga* genera isolated from Pacific or Atlantic deep-sea hydrothermal systems. Interestingly, the virions of MPV1, a siphovirus infecting *Marinitoga piezophila*, carry a host plasmid in addition to the viral DNA, highlighting the horizontal gene transfer in the deep-sea thermal systems.

Hot Deserts

Hot hyperarid deserts represent complex environments, where organisms have to adapt to multiple extreme conditions, including broad temperature variations, water limitation, intensive UV radiation and nutrient starvation. In spite of multiple extremities, microbial cells and viruses inhabit the soils of hot deserts, although viral abundances are low. Tailed virus particles were observed in the samples of soils from the Namib, Mojave, and Sahara deserts. Hot desert metaviromes have also shown to

Table 2 Examples of phages isolated from thermal environments

<i>Virus name</i>	<i>Host</i>	<i>Isolation site</i>	<i>Particle morphology</i>	<i>Genome (type and size)^a</i>
Isolates from terrestrial hot springs				
phiYS40	<i>Thermus thermophilus</i>	Hot spring, Japan	Myovirus	Linear dsDNA, 152,372 bp
phiNS11	<i>Bacillus acidocaldarius</i>	Beppu hot springs, Japan	Tailless icosahedral	dsDNA, nd
IN93	<i>Thermus thermophilus</i>	Induced from <i>Thermus aquaticus</i> T22, hot spring, Japan	Tailless icosahedral	Circular dsDNA, 19,604 bp
P23–45	<i>Thermus thermophilus</i>	Alkaline hot spring, Dolina Geyser, Kamchatka, Russia	Siphovirus with a long tail (823 nm)	Linear dsDNA, 84,201 bp
P74–26	<i>Thermus thermophilus</i>	Alkaline hot spring, Uzon, Kamchatka, Russia	Siphovirus with a long tail (823 nm)	Linear dsDNA, 83,319 bp
P23–77	<i>Thermus thermophilus</i>	Alkaline hot spring, New Zealand	Tailless icosahedral	Circular dsDNA, 17,036 bp
P23–72	<i>Thermus thermophilus</i>	Alkaline hot spring, New Zealand	Tailless icosahedral	dsDNA, nd
P23–65 H	<i>Thermus thermophilus</i>	Alkaline hot spring, New Zealand	Tailless icosahedral	dsDNA, nd
PH75	<i>Thermus thermophilus</i>	Thermal spring	Filamentous	ssDNA, ~6.5 kbp
GBSV1	<i>Geobacillus</i> sp. 6K51	Isolated from <i>Geobacillus</i> sp. 6K51 culture, hot spring, Xiamen of China	Myovirus	Linear dsDNA, 34,683 bp
TSP4	<i>Thermus</i> TC4	Tengchong hot spring (65°C, pH 7.0), China	Siphovirus with a long tail (785 nm)	dsDNA, ~80 kbp
BV1	<i>Geobacillus</i> sp. 6k512	Isolated from <i>Geobacillus</i> sp. 6k512 culture, hot spring, Xiamen of China	Myovirus	Circular dsDNA, 35,055 bp
MMP17	<i>Meiothermus</i> TG17	Hot spring (pH 7.3 and 63°C), Eryuan, Yunnan, China	Myovirus	dsDNA, 33.5–39.5 kb
phiTMA	<i>Thermus thermophilus</i>	Atagawa hot spring, Japan	Myovirus	Linear dsDNA, 151,483 bp
RM378	<i>Rhodothermus marinus</i>	Slightly saline geothermal site, Hveravik, Iceland	Myovirus	Linear dsDNA, 129,908 bp
phiOH3	<i>Thermus thermophilus</i>	Obama hot spring (75°C), Nagasaki, Japan	Filamentous	Circular ssDNA, 5,688 bp
G20c	<i>Thermus thermophilus</i>	Hot spring (~65°C pH 7.5), Geyser Valley, Kamchatka, Russia	Nd (probably siphovirus) ^b	Linear dsDNA, 81,291 bp
Isolates from deep-sea hydrothermal systems				
BVW1	<i>Bacillus</i> sp. w13	Isolated from <i>Bacillus</i> sp. w13, Pacific hydrothermal fields	Tailed icosahedral	dsDNA, ~18 kbp
GVE1	<i>Geobacillus</i> sp. E26323	Isolated from <i>Geobacillus</i> sp. E26323, Pacific hydrothermal fields	Siphovirus	dsDNA, ~41 kbp
GVE2	<i>Geobacillus</i> sp. E263	Isolated from <i>Geobacillus</i> sp. E263, deep-sea hydrothermal field, the east Pacific	Nd (probably siphovirus) ^b	Linear dsDNA, 40,863 bp
D6E	<i>Geobacillus</i> sp. E263	Isolated from <i>Geobacillus</i> sp. E263, deep-sea hydrothermal field, the east Pacific	Myovirus	Circular dsDNA, 49,335 bp
NrS-1	<i>Nitratiruptor</i> sp. SB155–2	Induced from <i>Nitratiruptor</i> sp. SB155–2, Iheya North hydrothermal field, Japan	Siphovirus	Linear dsDNA, 37,159 bp
MPV1	<i>Marinitoga piezophila</i>	Induced from <i>M. piezophila</i> KA3, hydrothermal vent chimney (2630 m depth), East-Pacific Rise	Siphovirus	dsDNA, 43,715 bp
MCV-1	<i>Marinitoga camini</i>	Induced from <i>M. camini</i> 97, Lucky strike hydrothermal vent (1700 m depth), Mid Atlantic Ridge	Siphovirus	Circular dsDNA, 53,412 bp
MCV-2	<i>Marinitoga camini</i>	Induced from <i>M. camini</i> 55, black smoker chimney, Menez Gwen site, (840–870 m depth), Mid Atlantic Ridge	Siphovirus	Circular dsDNA, 50,307 bp

^aNd, not determined.^bBased on genome sequence analyzes.

contain phage sequences, e.g., phages associated with bacteria belonging to the following genera: *Bacillus*, *Rhizobium*, *Geobacillus*, *Actinoplanes*, *Mycobacterium*, *Streptomyces*, and *Myxococcus*. The treatment of desert soil samples with mitomycin C has significantly enhanced the recovery of viral particles from the samples, suggesting that lysogeny may be a prevalent lifestyle of viruses in hot desert soils.

Phages in Polar and Other Cold Environments

About 10% of the Earth land surface is permanently covered by ice and snow, extending up to roughly 30% during winter in the Northern hemisphere. Furthermore, ocean deep waters, which account for 90% of the total ocean volume, are cold (0–3°C), and up to ~5% of the area of the global ocean is covered by sea ice. Thus, the biosphere provides a great variety of cold biotopes,

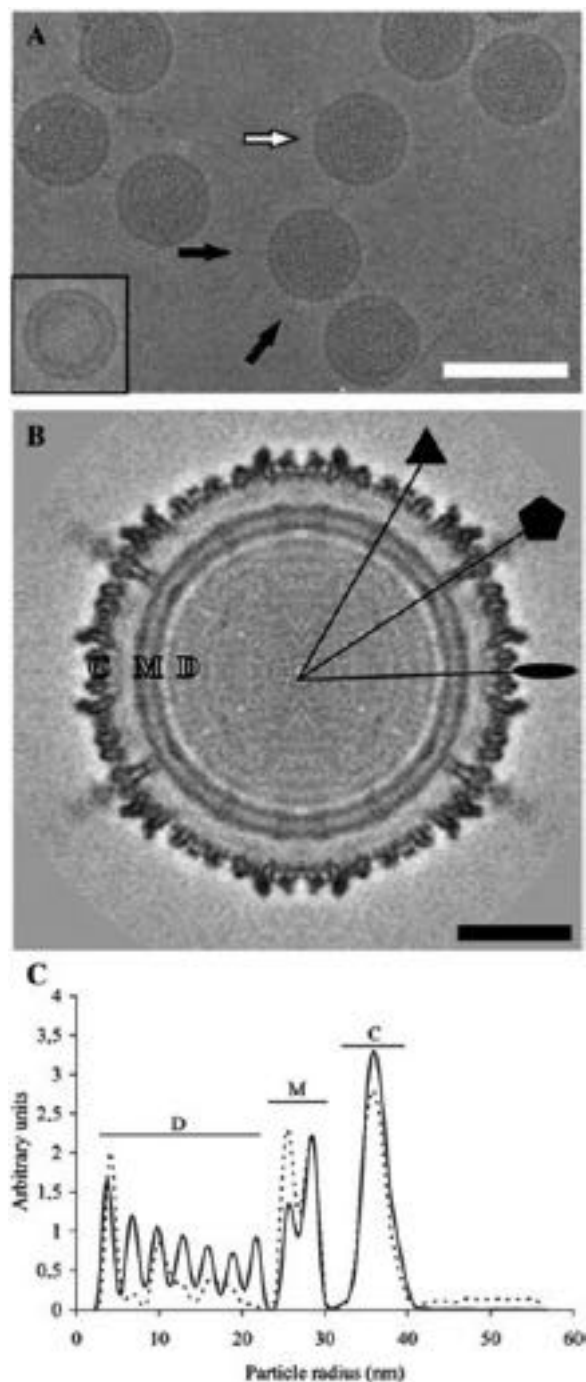


Fig. 4 (A) Organization of P23-77. Cryo-electron micrograph taken at 2.7 μm underfocus showing P23-77 viral particles (white arrow), with a diameter of 78 nm. Thin spikes, which might be used in the infection of the host, are visible on some particles (black arrow). The viral membrane inside the protein capsid is visible in empty particles (inset, underfocus 2.2 μm). Bar, 100 nm. (B) A 0.28 nm thick central section through the virion. Symmetry axes are indicated with a black ellipse (2-fold), triangle (3-fold) and pentagon (5-fold). Proteins connecting the viral capsid and the underlying lipid bilayer at the five-fold vertexes are visible. D (DNA), M (membrane) and C (capsid shell). Bar, 20 nm. Protein is black in A and B. (C) Radial density profiles of the icosahedral reconstruction of the intact virion (solid line) and the empty particle (dotted line). For calculation of the radial profiles both the full and the empty particle reconstructions were calculated to 3.0 nm resolution. Reproduced from Jaatinen, S.T., Happonen, L.J., Laurinmäki, P., Butcher, S.J., Bamford, D.H., 2008. Biochemical and structural characterisation of membrane-containing icosahedral dsDNA bacteriophages infecting thermophilic *Thermus thermophilus*. *Virology* 379 (1), 10–19. doi:[10.1016/j.virol.2008.06.023](https://doi.org/10.1016/j.virol.2008.06.023).

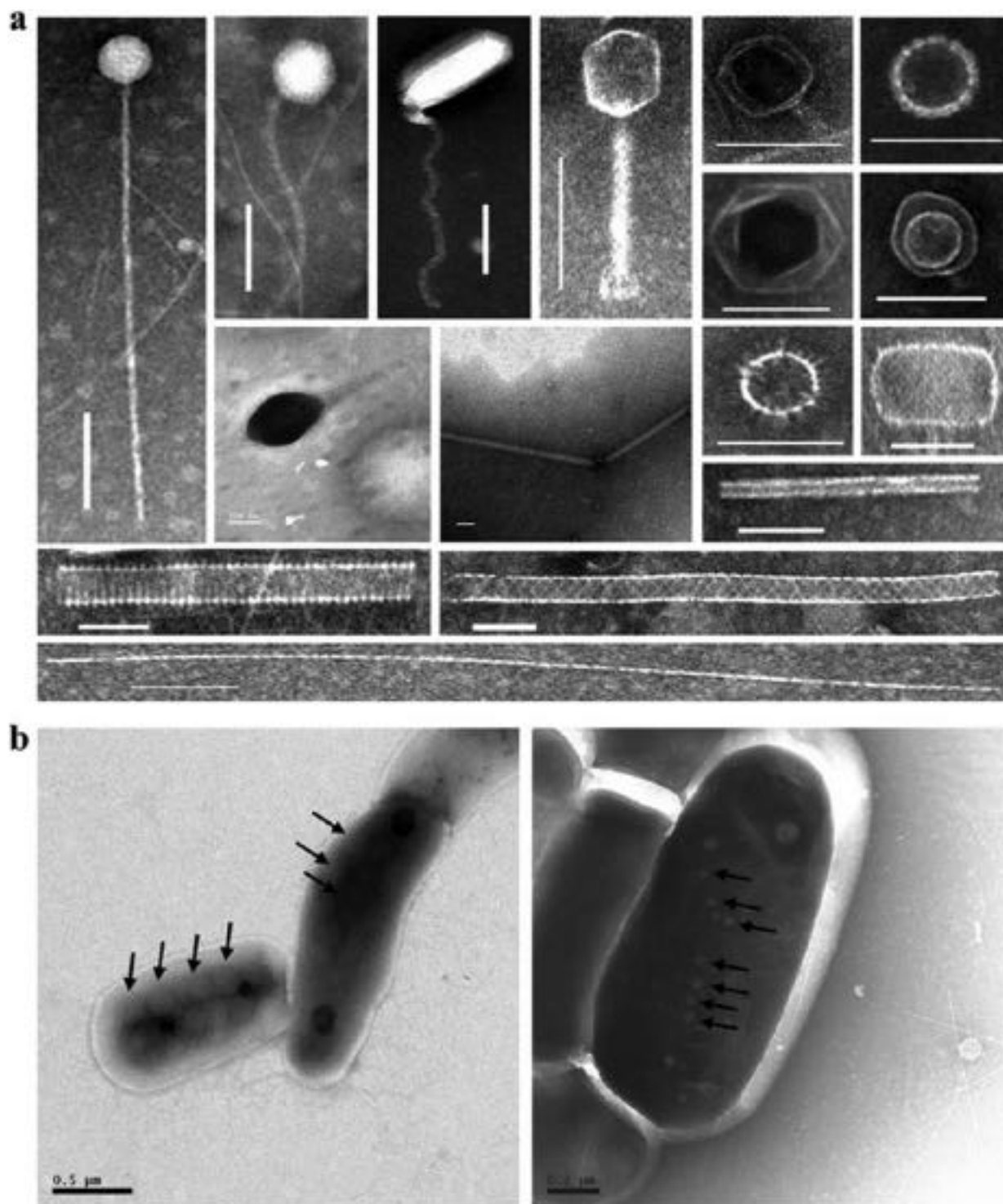


Fig. 5 Transmission electron micrographs showing the morphologies of virus-like particles and infected cells in the deep sediments of the Baltic Sea. (a), Examples of virus-like particles observed. Scale bars: 100 nm. (b), Infective viruses (arrows) in the visibly infected cells. Samples were recovered from deep sediments down to 70 mbsf in Hole M59C. Reproduced from Cai, L., Jørgensen, B.B., Suttle, C.A., *et al.*, 2019. Active and diverse viruses persist in the deep sub-seafloor sediments over thousands of years. *The ISME Journal* 13 (7), 1857–1864. doi:[10.1038/s41396-019-0397-9](https://doi.org/10.1038/s41396-019-0397-9).

where *psychrophilic* or *psychrotrophic* organisms thrive. *Psychrophiles*, i.e., “cold-loving” organisms, grow and reproduce optimally at the temperatures ranging roughly from 0 to 15°C (max. ~20°C) and often are also capable to survive at sub-zero temperatures. Psychrotrophic organisms are cold-tolerant, they may survive at the above mentioned temperatures, but require temperatures

Table 3 Examples of phages isolated from cold environments

<i>Virus name</i>	<i>Host</i>	<i>Isolation site</i>	<i>Particle morphology^a</i>	<i>Genome (type and size)</i>
1a	<i>Shewanella</i> sp. 1A	Arctic sea ice	Myovirus	Nd
21c	<i>Colwellia</i> sp. 21C	Arctic sea ice	Siphovirus	Nd
11b	<i>Flavobacterium</i> sp. 11B	Arctic melt pond samples	Siphovirus	Circular dsDNA, 36,012 bp
MYSP03	<i>Flavobacterium</i> sp. MYB03	Mingyong Glacier, China	Tailed icosahedral	dsDNA, ~66 kbp
9A	<i>Colwellia psychrerythraea</i> 34H	Arctic nepheloid layer seawater	Siphovirus	Linear dsDNA, 104,936 bp
MYSP08	<i>Flavobacterium</i> sp.	Mingyong Glacier, China	Tailed icosahedral	Nd
MYTP08	<i>Flavobacterium</i> sp.	Mingyong Glacier, China	Tailless icosahedral (tequivirus-like)	Nd
1/49	<i>Shewanella</i> sp. 49	Arctic sea ice	Myovirus	Nd
1/32	<i>Flavobacterium</i> sp. 32	Arctic sea ice	Siphovirus	Circular dsDNA, 42,252 bp
1/44	<i>Shewanella</i> sp. 44	Arctic sea ice	Siphovirus	Circular dsDNA, 49,640 bp
3/49	<i>Shewanella</i> sp. 49	Arctic sea ice	Siphovirus	Circular dsDNA, 40,161 bp
1/4	<i>Shewanella</i> sp. 4	Arctic sea ice	Myovirus	Linear dsDNA, 133,824 bp
1/40	<i>Shewanella</i> sp. 40	Arctic sea ice	Myovirus	Linear dsDNA, 139,004 bp
1/41	<i>Shewanella</i> sp. 41	Arctic sea ice	Myovirus	Circular dsDNA, 43,510 bp
VMY22	<i>Bacillus cereus</i> MYB41–22	Mingyong Glacier, China	Podovirus	dsDNA, 18–20 kb
f327	<i>Pseudoalteromonas</i> sp. BSi20327	<i>Pseudoalteromonas</i> sp. culture from Arctic sea ice	Filamentous	ssDNA, 6.1 kbp
VNPH	<i>Aeromonas sobria</i> NPH-1	Napahai plateau wetland, China	Myovirus	dsDNA, 110–120 kb
MuztagBP1	<i>Pseudomonas</i> sp.	Karakul lake, China	Myovirus	Nd
MuztagBP2	<i>Pseudomonas</i> sp.	Karakul lake, China	Siphovirus	Nd
MYBP2A-15	<i>Pseudomonas</i> sp.	Mingyong Glacier, China	Myovirus	Nd
MYSP06	<i>Janthinobacterium</i> sp. MYB06	Mingyong Glacier, China	Siphovirus	dsDNA, 65–70 kb
VSW-3	<i>Pseudomonas fluorescens</i> SW-3	Napahai plateau wetland, China	Podovirus	dsDNA, 40,556 bp
PANV1	<i>Paraglaciecola</i> sp.	Antarctic sea ice	Myovirus	Nd
PANV2	<i>Paraglaciecola</i> sp.	Antarctic sea ice	Siphovirus	Nd
OANV1	<i>Octadecabacter</i> sp.	Antarctic sea ice	Siphovirus	Nd
OANV2	<i>Octadecabacter</i> sp.	Antarctic sea ice	Podovirus	Nd

^aNd, not determined.

above 15°C for the optimal growth (max. ~30°C). In many habitats, psychrophiles have to tolerate also other extreme factors, such as high pressure in cold deep-sea waters or high salinity in some Arctic and Antarctic lakes.

Diverse and active microbial communities inhabit the cryosphere. Because of the absence or low abundance of grazers, these communities are characterized by high prokaryotic mortality due to viral infections. In low temperature ecosystems, viruses have a particularly important role in driving microbial evolution and, indirectly, the cycling of carbon. Although the viral component of cryosphere remains largely unstudied, it is already evident that viruses are abundant in various cold environments, such as polar freshwater and saline lakes, glaciers, sea ice, polar soils, snow, and permafrost.

Polar Oceans

Metagenomic analysis of viromes from four different oceanic regions, including the Arctic Ocean, has revealed that global and regional marine viral diversity is high, varying in different latitudes, and there are both highly widespread and endemic marine viruses in the global ocean. The analyzes of surface water dsDNA viromes sampled in the Antarctic Prydz Bay revealed high abundance of tailed phages, including *Flavobacterium* phage 11b (Table 3). A substantial number of prophage sequences and viral genes, especially phage terminase genes, has been detected within bacterial genomes from both Arctic and Antarctic Ocean waters, indicating active horizontal gene transfer in these cold polar waters.

Glaciers

Glaciers, especially supraglacial (top) biotopes, host a variety of microorganisms, including bacteria and their viruses. Due to the absorption of sun heat by cryoconite, i.e., dark rock particles and organic sediments on the surface of glaciers, the underlying ice melts, forming cryoconite holes filled with water (Fig. 6). These holes are the hotspots of microbial life in glaciers. Positive correlation between bacterial and viral abundances, high numbers of virus-like particles (10^6 – 10^7 VLPs ml⁻¹), high virus infection and production rates, but low virus burst size have been reported for cryoconite holes. Molecular studies targeting T4 MCP gene, g23, on glacier surfaces in Svalbard, Arctic, revealed a diverse community of T4-like phages, including both novel and cosmopolitan phages. The analyzes of metaviromes from cryoconite hole ecosystems of Svalbard glaciers and the Greenland Ice Sheet showed high diversity of bacteriophages, suggesting novel virus groups and unusual virus life strategies. At least eight

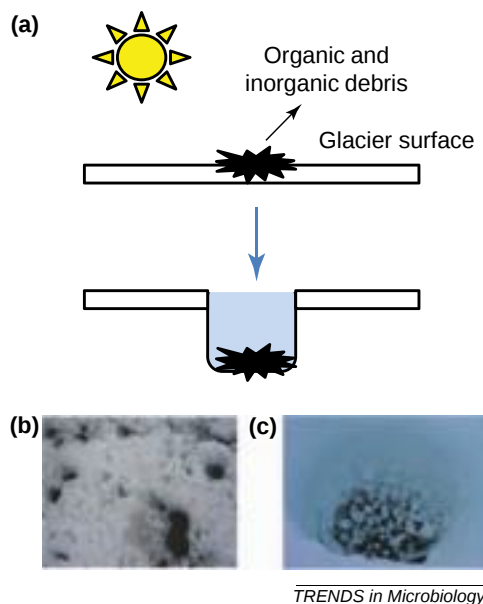


Fig. 6 Cryoconite holes. (a) Schematic diagram of cryoconite hole formation. Cryoconite holes at the surface of glaciers are formed because organic and inorganic particles, which are darker than the surrounding white icy surface, absorb the solar radiation better than the ice. The heated debris therefore melts into the ice forming holes, which provide nutrients and liquid water for microbial activity. (b) Example of open cryoconite holes in the Arctic. (c) Example of entombed cryoconite hole in Antarctica. Photo courtesy Liz Bagshaw. Reproduced from Anesio, A.M., Bellas, C.M., 2011. Are low temperature habitats hot spots of microbial evolution driven by viruses? *Trends in Microbiology* 19 (2), 52–57. doi:10.1016/j.tim.2010.11.002.

cold-active tailed bacteriophages have been isolated from glacial melt water (Table 3): *Flavobacterium* phages MYSP03, MYSP08, and MYTP08, *Janthinobacterium* phage MYSP06, *Bacillus cereus* phage VMY22, and *Pseudomonas* phages MYBP2A-15, MuztagBP1, and MuztagBP2.

Polar Lakes

Culture-independent studies have revealed high virus loads in Antarctic freshwater and saline lakes. VLP concentrations appear to be high in polar saline lakes compared to temperate marine waters, but lower in polar freshwater bodies than in temperate freshwater systems. Microscopic examinations of Antarctic freshwater and saline lakes demonstrated various virus morphologies, including tailed and tailless icosahedral particles. Metagenomic data have also showed high relative abundance of the phages belonging to the order *Caudovirales*. Based on the available viromes, Arctic and Antarctic freshwater viral communities are similar at general taxonomic level, but diverse and specific at a fine level. A large portion of sequences from polar metagenomes has no similarities to the data stored in common repositories, highlighting a need for more comprehensive sampling of cold environments.

Sea Ice

In sea ice, microbial communities thrive in brine channels and pockets, which are formed within the ice in the process of seawater freezing. Algae, bacteria, and their viruses dominate in these unique microenvironments. It has been shown that bacterial growth and viral production in sea ice brines is possible even at -12°C . Bacteriophage densities range roughly from 10^5 to 10^8 VLPs ml^{-1} in sea ice. Higher concentrations of bacteria and viruses have been reported for sea ice compared to seawater, pointing to active virus production within the ice and/or active enrichment with viruses from seawater. The analysis of genome sequences of sea ice bacteria indicated that gene transfer mediated by viruses is abundant in sea ice. To date, the majority of sea ice virus-host systems have been isolated from the Arctic (Fig. 7), but recently, the first four virus-host systems have been obtained also from the Antarctic sea ice (Table 3).

Permafrost

Viruses have been observed in Siberian, Antarctic, and Arctic permafrost. The hotspots of life in frozen soils are cryopegs, highly saline pockets or brines within permafrost layers. The study of cryopegs in permafrost near Barrow, Alaska, showed greater VLP concentrations in cryopegs (up to 10^7 VLPs ml^{-1}) than in the associated ice wedge (up to 10^5 VLPs ml^{-1}). Microscopic images of Barrow cryopeg filtered brine revealed VLPs resembling siphoviruses and virome analysis showed that phages belonging to the order *Caudovirales* dominated in the taxonomic composition among sequences having hits to known viruses. Unclassified

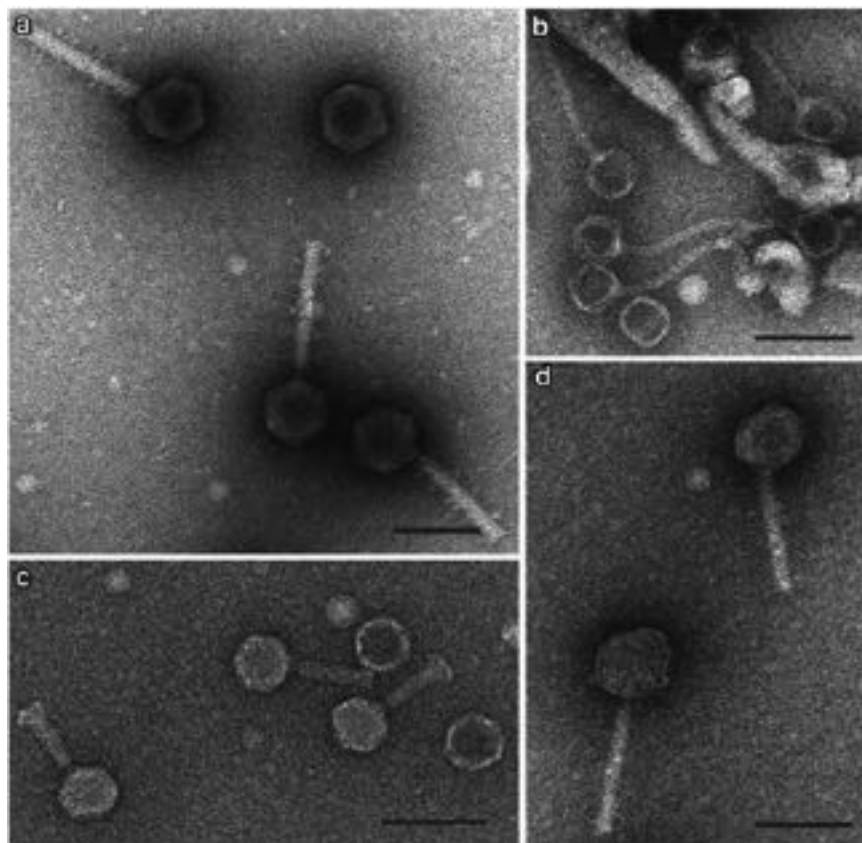


Fig. 7 Transmission electron micrographs of the negative stained samples of '1 × purified' ice phages. a Phage 1/4, b phage 1/32, c phage 1/41, d phage 1/40 isolated from Baltic sea ice. Scale bars 100 nm. Note that only sections A–D from original figure are used here. Reproduced from Luhtanen, A.-M., Eronen-Rasimus, E., Kaartokallio, H., 2014. Isolation and characterization of phage–host systems from the Baltic Sea ice. *Extremophiles* 18 (1), 121–130. doi:10.1007/s00792-013-0604-y.

members of the order *Caudovirales* were also prevalent in the taxonomically identifiable part of the metavirome obtained from peatland soils along a permafrost thaw gradient in Stordalen, Sweden.

Other Cold Environments

Metagenomic studies have revealed the presence of bacteriophages in polar soils and snow. Public snow metagenomes contain a portion of viral sequences, albeit small, and prophages have been induced from *Paenibacillus* isolates obtained from top snow in the Antarctic. Studies of cold hyperarid desert soils in Antarctica showed high VLP densities ($\sim 10^8$ VLPs g^{-1} of dry soil), which is substantially higher than in hot hyperarid desert soils, and very high VPRs (up to 8200). Sequence analysis demonstrated that the members of *Caudovirales* were dominant among identified viruses, with *Mycobacterium* phages being particularly abundant. Interestingly, temperate bacteriophages (siphoviruses SpaA1 and BceA1) with an unusual arrangement of genes were retrieved from the Antarctic soils.

About 20 cold-active lytic tailed bacteriophages have been isolated from the Napahai wetland located in the middle of the Hengduan Mountains, China. Some dozens of phages were isolated on *Flavobacterium psychrophilum*, a psychrophilic pathogen of fish. In addition, many psychrotrophic or psychrophilic phage isolates have been obtained from refrigerated food sources, e.g., meat, fish and milk products. Water droplets in the clouds may also harbor viruses, as these unique microenvironments remain liquid at temperatures close to -40°C and allow microbial cells to survive and metabolize avoiding vitrification at subzero temperatures. However, very little is known about phages residing in the atmosphere (see the article below).

Atmosphere

Atmospheric phages and their airborne host microbes originate from various terrestrial or aquatic sources, such as sea spray, desert dust, plant litter, snow, or sometimes, human or animal sources. Oceans are the major producers of primary biological aerosol particles, which are also called biological aerosols, referring to airborne viruses, bacteria, archaea, fungi, pollen, spores, or

fragments of different microorganisms. Marine biological aerosols are released from sea surface microlayer to the atmosphere by bubble bursting process and transported further by winds. Various studies have shown that bacteria and phages are enriched in the sea surface microlayer indicating that a large fraction of atmospheric viruses are likely phages. At the airborne state, viruses are considered to be associated with small organic aerosol particles ($< 1 \mu\text{m}$) and due to the relatively small size, atmospheric residence times of viruses are longer compared to bacteria and other cellular organisms.

Terrestrial dust is another frequent source of bioaerosols and it has been found to contain various VLPs with icosahedral heads and different types of tails. Saharan dust is known to cross the Pacific Ocean in a few days' time, indicating that phages attached to small dust particles can indeed travel long distances in air. Terrestrial dust contains a size distribution of inorganic and organic particles with the smallest ones being able to penetrate lung alveoli. This suggests that in addition to various environmental sites, phages can be transported also to the lungs and from therein to the blood circulation of humans and animals. Diverse phages have also been observed in different types of soils and plant litter from which they can be aerosolized by winds.

Because environmental phages are major drivers of the carbon cycles by causing cellular lysis, they can also influence climate change. The changing climate, on the other hand, can affect for instance, phage life cycles. Atmospheric dispersal of phages and other viruses has been considered as one of the most plausible explanations for the major question why genetically similar viruses are found from geographically distant environments, in spite of the enormous number of phages in the oceans (up to 10^{31}). Because phages influence horizontal gene transfer, atmospheric phages can transport genes to host bacteria even to distant environments, affecting the global gene pool. Extremophilic archaeal viruses were found to be transported by air between three different geographically distant hot springs of Yellowstone National Park, suggesting that airborne viral dispersion contributes to local population diversity more than mutation rate. It is likely that the same phenomenon is experienced globally for phages and other viruses as well.

The atmospheric significance of phages and their host bacteria is especially crucial in polar environments, where aerosols from terrestrial origin are less prevalent. Atmospheric microorganisms can function as cloud condensation nuclei and ice nuclei, affecting cloud dynamics. Bacteria, such as those belonging to the genus *Pseudomonas*, are known to contain ice nucleation active membrane proteins. In some rare occasions, phages have also been shown to be able to nucleate ice, but to date very little is known about this property.

Environments like the oceans are known to contain numerous phages that are ecologically important, but to our knowledge, specific phage isolates originating from outdoor atmospheric samples, are yet to be discovered. In addition, very little is known about the survival and infectivity of phages in the atmosphere. At the airborne state, phages are subjected to several environmental stressors, such as intense UV radiation, desiccation, rain, ice formation, and low temperature. Phages can be protected from these stressors for instance by remaining in the prophage state within the host cell. Attachment to solid particles, on the other hand, can protect phages from desiccation caused by UV. Flow cytometric studies of air samples have demonstrated that atmospheric samples contain around 10–100 times more viruses than bacteria and approximately 70% of these viruses are attached to organic aggregates. Deposition rates of approximately $10^9\text{--}7 \times 10^9 \text{ m}^{-2}$ in a day were reported for viruses and $10^7\text{--}8 \times 10^7 \text{ m}^{-2}$ in a day for bacteria with the highest virus deposition rates originating from marine bioaerosols. It is presumable that a large percentage of these viruses are phages.

Conclusions

Although characterized by very challenging conditions, extreme environments are in fact inhabited by a large number of extremophiles, which are mostly microorganisms. Viruses and bacteriophages in particular are abundant and active in extreme environments and have globally important ecological roles. Virus-like particle densities differ in various extreme niches, being e.g., $10^4\text{--}10^5 \text{ VLPs ml}^{-1}$ in the deep-sea and up to $10^9 \text{ VLPs ml}^{-1}$ in hypersaline waters. Tailed VLPs typical for bacteriophages, as well as other more unusual particle morphologies, have been observed in extreme environments. Part of these VLPs may represent archaeal viruses. Metagenomic studies have revealed *Caudovirales*-related sequences in numerous extreme sites. Atmospheric movement of phages has been suggested to provide some explanations for the observed genetic similarity between phages originating from geographically distant extreme environments. Sampling of certain environments, e.g., hydrothermal vents, seabed subsurface or the atmosphere, is still very challenging technically, while other sites, e.g., hypersaline lakes and salterns, represent more easily accessible locations. Subsequently, the number of isolates from different types of extreme environments varies. Known extremophilic phage isolates display the following morphologies: tailless or tailed icosahedral (myo-, siphon-, or podoviruses), spherical, and filamentous. Their genomes are linear or circular ds- or ssDNA molecules, many of which are not yet sequenced. Detailed molecular studies of the obtained isolates are still rare, but the available data adds insights into the understanding of evolutionary relations between different viral groups. The research on extremophilic phages widens our view on global viral diversity and helps to determine certain boundaries of the virosphere.

Further Reading

- Atanasova, N.S., Oksanen, H.M., Bamford, D.H., 2015. Haloviruses of archaea, bacteria, and eukaryotes. *Current Opinion in Microbiology* 25, 40–48.
- Deming, J.W., 2010. Sea ice bacteria and viruses. In: Thomas, D.N., Dieckmann, G.S. (Eds.), *Sea Ice*, second ed. Blackwell Publishing Ltd., pp. 247–282.
- Després, V.R., Huffman, J.A., Burrows, S.M., et al., 2012. Primary biological aerosol particles in the atmosphere: A review. *Tellus B: Chemical and Physical Meteorology* 64 (1),

- Hodson, A., Anesio, A.M., Tranter, M., *et al.*, 2008. Glacial ecosystems. *Ecological Monographs* 78 (1), 41–67.
- Lossouarn, J., Dupont, S., Gorlas, A., *et al.*, 2015. An abyssal mobilome: Viruses, plasmids and vesicles from deep-sea hydrothermal vents. *Research in Microbiology* 166 (10), 742–752.
- Pearce, D.A., Wilson, W.H., 2003. Viruses in Antarctic ecosystems. *Antarctic Science* 15 (3), 319–331.
- Reche, I., D'Orta, G., Mladenov, N., Winget, D.M., Suttle, C.A., 2018. Deposition rates of viruses and bacteria above the atmospheric boundary layer. *ISME Journal* 12 (4), 1154–1162.
- Rothschild, L.J., Mancinelli, R.L., 2001. Life in extreme environments. *Nature* 409, 1092–1101.
- Sabet, S., 2012. Halophilic viruses. In: Vreeland, R.H. (Ed.), *Advances in Understanding the Biology of Halophilic Microorganisms*. Springer, pp. 81–116.
- Santos, F., Yarza, P., Parro, V., *et al.*, 2012. Culture-independent approaches for studying viruses from hypersaline environments. *Applied and Environmental Microbiology* 78 (6), 1635–1643.
- Snyder, J.C., Wiedenheft, B., Lavin, M., *et al.*, 2007. Virus movement maintains local virus population diversity. *Proceedings of the National Academy of Sciences of the United States of America* 104 (48), 19102–19107.
- Zablocki, O., Adriaenssens, E.M., Cowan, D., 2015. Diversity and ecology of viruses in hyperarid desert soils. *Applied and Environmental Microbiology* 82 (3), 770–777.
- Zablocki, O., van Zyl, L., Trindade, M., 2018. Biogeography and taxonomic overview of terrestrial hot spring thermophilic phages. *Extremophiles* 22 (6), 827–837.

ARCHAEAL VIRUSES

Diversity of Hyperthermophilic Archaeal Viruses

David Prangishvili, Institut Pasteur, Paris, France and Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

Mart Krupovic, Archaeal Virology Unit, Institut Pasteur, Paris, France

Diana P Baquero, Archaeal Virology Unit, Institut Pasteur, Paris, France and Sorbonne University, Paris, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Å Angstrom

ABV Acidianus bottle-shaped virus

ACV Aeropyrum coil-shaped virus

AFV1 Acidianus filamentous virus 1

APBV1 Aeropyrum pernix bacilliform virus 1

APOV1 Aeropyrum pernix ovoid virus 1

ATV Acidianus two-tailed virus

bp Base pair

cryo-EM Cryo-electron microscopy

dsDNA Double-stranded DNA

GDGT Glycerol dibiphytanyl glycerol tetraether

ITRs Inverted terminal repeats

kb Kilobase

kDa Kilodalton

MCP Major capsid protein

nm Nanometer

NMR Nuclear magnetic resonance

nt Nucleotide

ORF Open reading frame

PCNA Proliferating cell nuclear antigen

PFV1 Pyrobaculum filamentous virus 1

PSV Pyrobaculum spherical virus

SEV1 Sulfolobus ellipsoid virus 1

SIRV1 Sulfolobus islandicus rod-shaped virus 1

SIRV2 Sulfolobus islandicus rod-shaped virus 2

SNDV Sulfolobus neozealandicus droplet-shaped virus

SPV1 Sulfolobus polyhedral virus 1

ssDNA Single-stranded DNA

ssRNA Single-stranded RNA

SSV1 Sulfolobus spindle-shaped virus 1

STIV Sulfolobus turreted icosahedral virus

TSPV1 Thermoproteus spherical piliferous virus 1

Glossary

Acidophilic Thriving under highly acidic conditions.

A-form One of the three major forms of double-stranded DNA, with a 23 Å helical diameter and 11 bp per helix turn.

Capsid Protein shell that encloses the genetic material of the virus.

Convergent evolution Independent evolution of similar features in species of different lineages.

Core genes One or more genes strongly conserved at the nucleotide sequence level among a related group of genomes.

Homologous recombination Recombination between two identical or similar DNA sequences.

Hyperthermophilic Requiring extremely high temperatures for optimal growth.

Inverted terminal repeats Short, related or identical sequences located in reverse orientation at the ends of the viral genome.

Jelly-roll fold A protein fold composed of eight β -strands arranged in two antiparallel four stranded β -sheets.

Pleomorphic viruses Viruses with asymmetric or variable virion morphology.

Protein glycosylation Post-translational modification where a carbohydrate molecule is covalently bound to a predetermined region of a protein to form a glycoprotein.

Proviruses Viral genomes integrated into the host chromosome.

Structural genomics Description of the three-dimensional structure of a protein encoded by a given genome.

Synteny The shared ordering of genomic segments along a chromosome.

Thermophilic Requiring high temperatures for optimal growth.

Viral envelope Lipid layer present in many types of viruses that protects their genetic material.

Virion Infectious mature virus particle.

Introduction

One of the most surprising results of the studies on viral diversity on our planet is the observation of an astounding number of different viral morphologies in the habitats where large-scale biodiversity is least expected – in geothermally heated environments where temperatures exceed 80°C and pH values are often below pH 3. Many virus morphologies observed here have never been detected in environments with less extreme conditions. Besides common filamentous and spherical virions, this diversity includes particles resembling bottles, droplets, coils and spindles, which can be tailless, tailed or two-tailed. Viruses with all these morphologies have been isolated from hot terrestrial springs of Europe, Asia, and North America. The hosts for this collection of viruses are archaea from the phylum Crenarchaeota – members of the genera *Acidianus*, *Aeropyrum*, *Metallosphaera*, *Pyrobaculum*, *Stygiolobus*, *Sulfolobus* and *Thermoproteus*. All these hosts grow optimally at temperatures above 80°C and thus are referred to as hyperthermophiles. Viral infection occurs most efficiently at optimal temperatures of host growth and thus also the viruses are considered to be hyperthermophiles.

Table 1 Viruses of Crenarchaeota

Family name	Virion shape	Example of species	Envelope	Genome type and size (bp)	Accession number
Viruses with particular shapes					
<i>Ampullaviridae</i>	Bottle-shaped, with fibers at the blunt end	Acidianus bottle-shaped virus (ABV)	External	Linear dsDNA, 23,900	EF432053
<i>Bicaudaviridae</i>	Spindle-shaped with two appendages	Acidianus two-tailed virus (ATV)	None	Circular dsDNA, 62,730	AJ888457
<i>Clavaviridae</i>	Bacilliform	Aeropyrum pernix bacilliform (APBV1)	None	Circular dsDNA, 5278	AB537968
<i>Fuselloviridae</i>	Spindle-shaped with fibers at one end	Sulfolobus spindle-shaped virus 1 (SSV1)	External	Circular dsDNA, 15,465	XO7234
<i>Guttaviridae</i>	Droplet/ovoid-shaped	Aeropyrum pernix ovoid virus 1 (APOV1)	External	Circular dsDNA, 13,769	NC_028256
<i>Spiraviridae</i>	Coils with two identical terminal appendages	Aeropyrum coil-shaped virus (ACV)	None	Linear dsDNA, 24,893	HE681887
Spherical viruses					
<i>Globuloviridae</i>	Spherical	Pyrobaculum spherical virus (PSV)	External	Linear dsDNA, 28,337	AJ635162
<i>Ovaliviridae</i>	Ellipsoid	Sulfolobus ellipsoid virus 1 (SEV1)	External	Circular dsDNA, 23,219	MF144115
<i>Portogloboviridae</i>	Icosahedron	Sulfolobus polyhedral virus 1 (SPV1)	Internal	Circular dsDNA, 20,222	KY927925
<i>Turriviridae</i>	Icosahedron with 12 turrets	Sulfolobus turreted icosahedral virus (STIV)	Internal	Circular dsDNA, 17,663	AY569307
Filamentous viruses					
<i>Lipothrixviridae</i>	Flexible filament with terminal structures	Acidianus filamentous virus 1 (AFV1)	External	Linear dsDNA, 21,080	AJ567472
<i>Rudiviridae</i>	Rigid rod, with terminal fibers	Sulfolobus islandicus rod-shaped virus 2 (SIRV2)	None	Linear dsDNA, 35,450	AJ344259
<i>Tristromaviridae</i>	Filamentous, with terminal fibers	Pyrobaculum filamentous virus 1 (PFV1)	External	Linear dsDNA, 17,714	KU307456

Hyperthermophilic viruses are extremely thermostable in aggressive environmental conditions of their natural habitats, as well as in the laboratory conditions, e.g., thermal inactivation of some of them requires autoclaving at 121 °C at least for 40 min.

Based on their diverse morphological and genomic properties, the isolated and characterized viruses of Crenarchaeota are currently classified into 13 families (Table 1), 1 order *Ligamenvirales* and 1 proposed class *Tokiviricetes*. The genomes of members of all families, except *Spiraviridae*, have double-stranded (ds) DNA genomes, circular or linear (Table 1), whereas members of the *Spiraviridae* have circular single-stranded (ss) DNA genome.

Morphology and Structure

The families of crenarchaeal viruses have virions with diverse characteristic morphologies. The bottle-shaped, tailed spindle-shaped, tailless spindle-shaped, bacilliform, coil-shaped, droplet-shaped morphologies of members of the *Ampullaviridae*, *Bicaudaviridae*, *Fuselloviridae*, *Clavaviridae*, *Spiraviridae*, and *Guttaviridae*, correspondingly, are unprecedented among viruses of Bacteria and Eukarya, and represent Archaea-specific virion morphotypes. Specific for Archaea are also dsDNA viruses with filamentous virions, including members of the *Lipothrixviridae*, *Rudiviridae* and *Tristromaviridae*, as well as viruses with spherical virions carrying helical nucleoprotein core – members of *Globuloviridae*, *Ovaliviridae* and *Portogloboviridae*. The only family of crenarchaeal viruses with structural resemblance to viruses from other domains of life is *Turriviridae*. Turriviruses have non-enveloped icosahedral virions, with an internal lipid layer enclosing circular dsDNA genome. The major and minor capsid proteins have the double and single jelly-roll fold, respectively, and are related to those found in a wide range of bacterial and eukaryotic viruses. In other studied cases, capsid proteins of crenarchaeal viruses have unique structural folds.

Structures of virions from the families *Clavaviridae*, *Lipothrixviridae*, *Portogloboviridae*, *Rudiviridae*, and *Tristromaviridae*, were reconstructed at near atomic resolution using cryo-electron microscopy (cryo-EM). The results shed light on the mechanisms of virion morphogenesis and provide information on the molecular basis of high thermostability of the corresponding virions. Remarkably, in the reconstructed virions of members of the four latter families the packed dsDNA could be observed and was found to be in A-form: the phosphate-phosphate distance along DNA backbone is about 5.9 Å, as opposed to 7 Å for common B-form DNA. The common occurrence of A-DNA in hyperthermophilic viruses suggests that it may be the prevalent storage form of DNA in most extreme environments.

Most known viruses of Crenarchaeota are covered with envelopes. In those cases, where this was studied, the viruses form envelopes from lipids selectively acquired from the pool of host lipids.

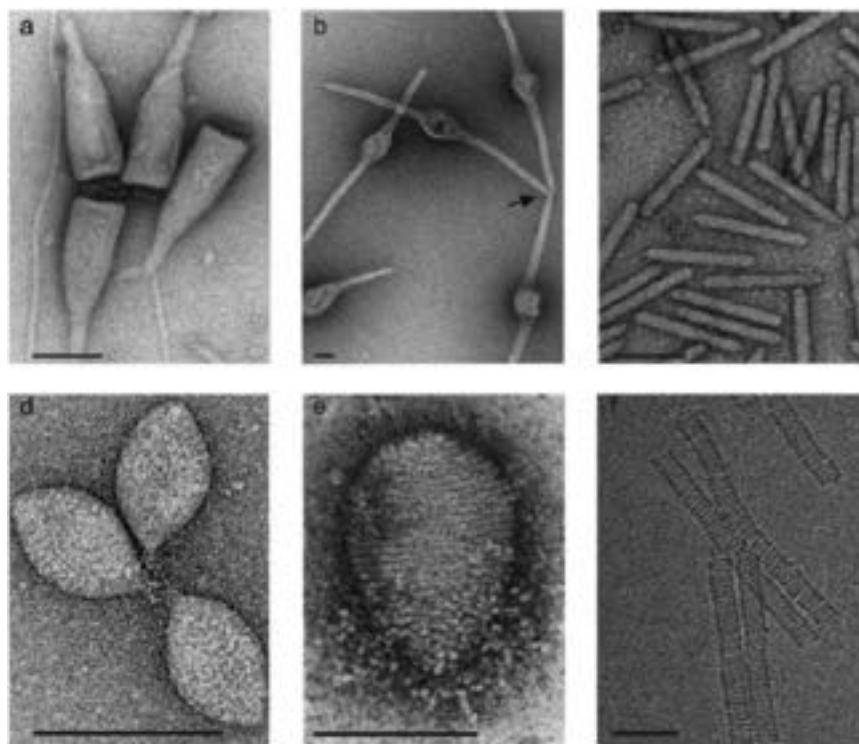


Fig. 1 Electron micrographs of archaeal viruses with particular morphologies. (a) ABV, Acidianus bottle-shaped virus; (b) ATV, Acidianus two-tailed virus, the arrow indicates virion tails which underwent extracellular development; (c) APBV1, Aeropyrum pernix bacilliform virus 1; (d) SSV1, Sulfolobus spindle-shaped virus 1; (e) SNDV, Sulfolobus neozealandicus droplet-shaped virus; and (f) ACV, Aeropyrum coil-shaped virus. Negative stain with uranyl acetate, except for ACV, which is in vitreous ice. Bars, 100 nm. Image modified from Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature reviews Microbiology* 15, 724–739.

Viruses with particular morphologies

Family Ampullaviridae (from Latin ampulla for “bottle”)

The enveloped virion of Acidianus bottle-shaped virus (ABV), the only isolated member of the family *Ampullaviridae*, resembles in its shape a champagne bottle (**Fig. 1(a)**). It has an overall length of 230 nm and a width varying from 75 nm at the broad end to 4 nm at the pointed end. The broad end of the virion is decorated with 20 thin rigid filaments, which appear to be inserted into a disc and interconnected at their bases. The cone-shaped inner core is formed by a toroidally supercoiled nucleoprotein filament 7 nm in width. It is presently unclear whether the pointed end or the filaments on the broader end are involved in adsorption to the cell surface and channelling of viral DNA into host cells. The virion contains six major proteins in the size range between 15 to 80 kDa.

Family Bicaudaviridae (from Latin bi, “two”, and cauda for “tail”)

The virions of Acidianus two-tailed virus (ATV), the only currently classified member of the family *Bicaudaviridae*, are released from host cells as spindle-shaped particles with overall dimensions of approximately 120×300 nm (**Fig. 1(b)**). Upon further incubation at temperatures above 75°C , appendages protrude from both pointed ends of the virion, and the lemon-shaped virion body shrinks to approximately 85×150 nm. The tails are heterogeneous in length, reaching 400 nm; the maximum length of the virion including tails reaches about 1000 nm. They have a tube-like structure and terminate with a narrow channel, which is 2 nm in width, and a terminal anchor-like structure formed by two furled filaments, each with a width of 4 nm. The virion contains at least eleven proteins with molecular masses in the range of 12–90 kDa.

Extracellular morphological development of the ATV virion takes place specifically at temperatures above 75°C , close to that of the natural habitat, and does not require the presence of host cells, an exogenous energy sources or specific co-factors. However, the mechanism of development of the ATV tails remains unknown.

Family Clavaviridae (from Latin clava for “club”, “stick”)

Aeropyrum pernix bacilliform virus 1 (APBV1), the only known member of the family *Clavaviridae*, has non-enveloped bacilli-form, rigid virions with dimensions of about 143×16 nm (**Fig. 1(c)**). The terminal cap structures, one of which is pointed and the other rounded, most likely are involved in DNA packaging and recognition of the host cell.

The virions carry multiple copies of a single major capsid protein (MCP) of about 10 kDa and three minor capsid proteins which have molecular masses in the range of 9.5–21.5 kDa. The MCP consists of two α -helices linked with a β -hairpin, a structural fold not seen in the capsid proteins of other known viruses. The structure of APBV1 virion was determined by cryo-electron microscopy at near-atomic resolution. The structure reveals how the MCPs pack together forming a tubular structure: each MCP molecule makes extensive hydrophobic contacts to six other neighbouring subunits, forming very tight hydrophobic interface which apparently contributes to the virion stability at extremely high temperatures. The inner surface of the tubular structure is positively charged, allowing efficient interactions with the circular dsDNA genome and its packaging as a left-handed superhelix. The structure allowed to propose an assembly model where the dsDNA genome and the capsid assembly in a concerted fashion. According to this model, the virion assembly starts by binding of three specific sites of the circular dsDNA with one of the cap structures; this forms three loops, which gradually intertwine directed by protein assembly into the tubular structure; after all DNA has been covered by protein, the open end of the virion is sealed by another cap structure.

Family Fuselloviridae (from Latin fusello for “little spindle”)

Virions of this family are enveloped and many have the shape of a lemon or spindle, with a bunch of thin filaments attached to one of the two pointed ends (Fig. 1(d)). Other family members are more pleomorphic and elongated, with three relatively thick filaments at one pointed end. The terminal filaments most likely are involved in adsorption to the host cell surface. Virions have dimensions of approximately 60×100 nm. The envelope of *Sulfolobus* spindle-shaped virus 1 (SSV1) contains lipids and proteins VP1 and VP3. The circular dsDNA of SSV1 is positively supercoiled when isolated from virions.

Examination of the SSV1 virion using cryo-electron microscopy (cryo-EM) have provided only limited information on virion architecture, but the study of SSV1 virion assembly and egress using electron tomography was more informative. Both virion assembly and egress were found to be concomitant and occur at the cellular cytoplasmic membrane via a process highly reminiscent of the budding of enveloped viruses that infect eukaryotes, including human immunodeficiency virus, influenza virus, and Ebola virus.

Family Guttaviridae (from Latin gutta for “droplet”)

The guttaviruses have slightly pleomorphic enveloped virions. Virions of *Sulfolobus* neozelandicus droplet-shaped virus (SNDV) resemble elongated droplets with varying dimensions, 110–185 nm in length and 95–70 nm in width (Fig. 1(e)). The pointed end of the virion is covered by a beard of thick filaments. Virions of another member of the family, *Aeropyrum* pennix ovoid virus 1 (APOV1), are ovoid without detectable filamentous attachments.

Family Spiraviridae (from Latin spira for “coil”)

The virions of *Aeropyrum* coil-shaped virus (ACV), the sole member of the family, are hollow, non-enveloped cylindrical particles, measuring about 230×20 nm. A short appendage of about 20 nm protrudes from each end of the cylindrical virion (Fig. 1(f)). Exceptionally, the viral genome is a single-stranded, positive-sense DNA molecule. The virion carries two MCPs with molecular masses of about 23 and 18.5 kDa, and a few minor virion proteins. The observation of partially degraded virions enabled to propose the following model of virion morphogenesis: the circular ssDNA is covered by multiple copies of the MCP and the two halves of the circular nucleoprotein filament intertwine to form a rope-like structure, which is further condensed into a helix of higher order.

Spherical viruses

Family Globuloviridae (from Latin globulus for “small ball”)

The virions of globuloviruses are enveloped, spherical particles, 70–100 nm in diameter (Fig. 2(a)). On the surface of the virion are multiple spherical protrusions about 15 nm in diameter. The viral envelope contains host-derived lipids and encases a tightly-packed superhelical nucleoprotein consisting of linear dsDNA and multiple copies of 33 kDa MCP. The virions carry minor capsid proteins. Virions of *Thermoproteus* spherical piliferous virus 1 (TSPV1) are decorated with numerous highly unusual filaments, which can extend hundreds of nanometers from the virion.

Family Ovaliviridae (from Latin ovalis for “oval”)

The sole known member of the *Ovaliviridae* family, *Sulfolobus* ellipsoid virus 1 (SEV1), has virions of ellipsoidal shape which measure about 115×80 nm. The virions are enveloped by a lipid containing membrane. The membrane encases circular nucleoprotein filament - consisting of circular dsDNA and DNA-binding MCPs, wrapped multiple times around the central axis of the virion (Fig. 2(b)).

Family Portogloboviridae (from Latin porto for “to bear”, and globus for “ball”)

The virions of *Sulfolobus* polyhedral virus 1 (SPV1) are icosahedral, about 90 nm in diameter (Fig. 2(c)). The cryo-EM reconstruction of the virion revealed structural details at near-atomic resolution. The icosahedral capsid is formed by 2 types of MCPs both of which carry variants of the single jelly-roll fold and are arranged in an atypical manner. The protein capsid encloses an internal lipid-containing membrane, which, in turn, encloses a spherical core formed by a circular nucleoprotein filament consisting of dsDNA and DNA-binding MCP. In the circular nucleoprotein, the dsDNA is complexed by a dimeric DNA-binding MCP

and is in A-form. The condensation of the circular nucleoprotein into a spherical core is suggested to occur in two steps: folding of circular nucleoprotein into a raft which is then spooled into concentric shells.

Family Turriviridae (from Latin turris for “tower”)

The overall morphology of non-enveloped icosahedral virions of the *Turriviridae* is highly similar to that of bacterial viruses from the families *Tectiviridae* and *Corticoviridae*: the virions are icosahedral with an inner lipid layer and pack naked dsDNA with the help of virus-encoded ATP-dependent molecular machinery (Fig. 2(d)). Moreover, members of all three families share the structure of MCPs which exhibit the double-jelly roll fold. Cryo-EM reconstruction of *Sulfolobus turreted icosahedral virus* (STIV) revealed unique structural features, including the elaborate turret-like structures at the fivefold vertices and the unusual pseudo- $T = 31$ icosahedral lattice on which the virion is built.

Filamentous viruses

Proposed class Tokiviricetes: order Ligamenvirales: family Rudiviridae (from the Latin rudis for “small rod”) order Ligamenvirales: family Lipothrixviridae (from the Greek lipos for “fat” and thrix for “hair”) family Tristromaviridae (from the Greek tria for “three” and stroma for “layer”)

The virions of members of the three families of the proposed class *Tokiviricetes* are filamentous, about 23 nm in width and the lengths in the range of 400–2000 nm, depending on the size of the linear dsDNA genomes and parameters of the nucleoprotein helix (Fig. 3). At both ends, the virions carry identical terminal structures. In the case of *Acidianus filamentous virus 1* (AFV1), a member of the *Lipothrixviridae*, the terminal structure resembles claws and folds upon interaction with the appendages of the host cells (Fig. 3(a)). The virions of *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2), the type member of the *Rudiviridae* family, are decorated with three thin terminal fibers, whereas virions of *Pyrobaculum filamentous virus 1* (PFV1), a member of the *Tristromaviridae*, possess bundles of filaments that can reach a length of 80 nm (Fig. 3(b) and (c)).

The virion structures of members of the three families have been reconstructed by cryo-EM at near atomic resolution and revealed similar, previously unknown form of virion organization. In all three families the core of the filamentous virion is a tube-like nucleoprotein formed by condensation of linear dsDNA by dimers of the MCPs – homodimers in case of *Rudiviridae* and heterodimers in the case of the other two families (Fig. 4(a)). As a result of binding to DNA, the MCPs form a helix-turn-helix structure and tightly

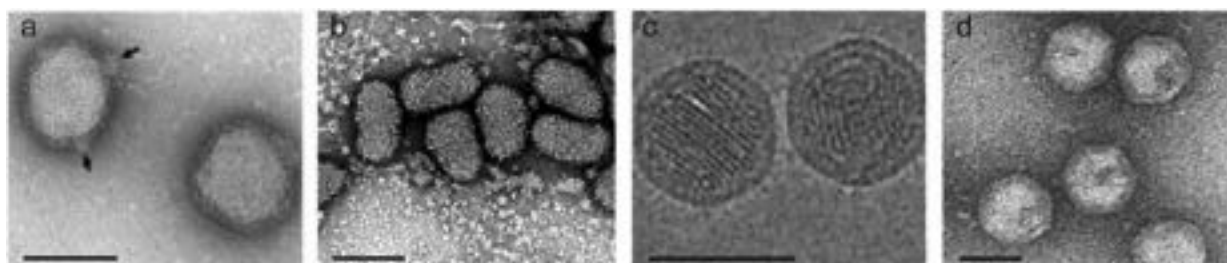


Fig. 2 Electron micrographs of archaeal viruses with spherical morphologies. (a) PSV1, *Pyrobaculum* spherical virus 1, the arrows indicate spherical protrusions; (b) SEV1, *Sulfolobus* ellipsoid virus 1; (c) SPV1, *Sulfolobus* polyhedral virus 1; and (d) STIV, *Sulfolobus* turreted icosahedral virus. Negative stain with uranyl acetate. Bars, 100 nm. Image modified from Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature reviews Microbiology* 15, 724–739. SEV1 image is courtesy of L. Huang.

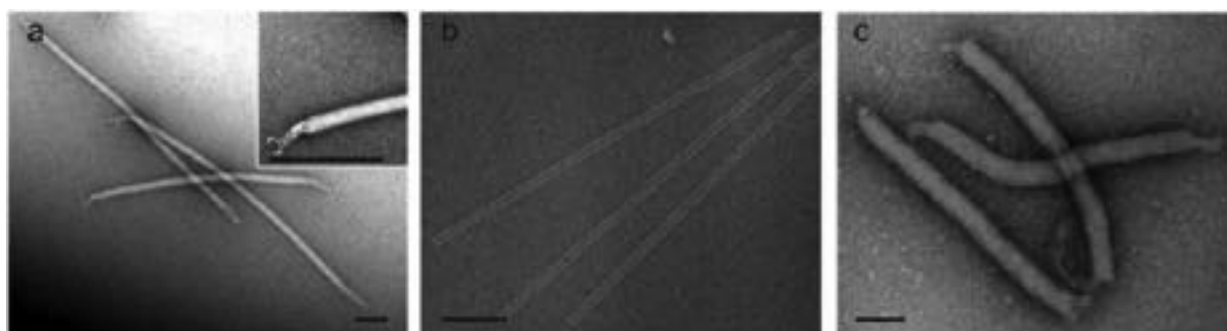


Fig. 3 Electron micrographs of archaeal viruses with filamentous morphologies. (a) AFV1, *Acidianus* filamentous virus 1, the inset displays the terminal structure; (b) SIRV2, *Sulfolobus islandicus* rod-shaped virus 2; and (c) PFV1, *Pyrobaculum* filamentous virus 1. Negative stain with uranyl acetate, except for SIRV2, which is in vitreous ice. Bars, 100 nm. Image modified from Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature reviews Microbiology* 15, 724–739.

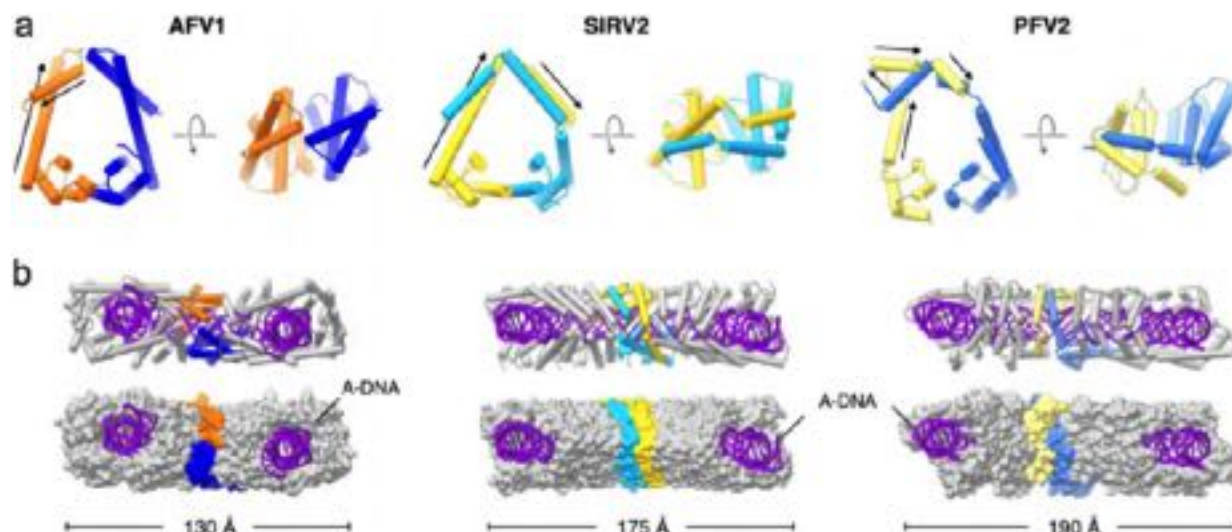


Fig. 4 Virion organization of the filamentous viruses AFV1, SIRV2 and PFV2. (a) Comparison of the MCP dimer (asymmetric unit) of AFV1 (left), SIRV2 (centre) and PFV2 (right). The MCP1 of AFV1, SIRV2 and PFV2 are colored in orange, gold and yellow, respectively. The MCP2 of AFV1, SIRV2 and PFV2 are colored in blue, cyan and light blue, respectively. The N-terminal helices of MCP1 in AFV1, SIRV2 and PFV2 are marked with black arrows; (b) wrapping of A-form DNA in AFV1, SIRV2 and PFV2. One MCP dimer is colored as in (a), other four dimers are colored in gray. Proteins are shown in ribbon representation (top) and as surfaces (bottom). Image reproduced from Wang, F., Baquero, D.P., Su, Z., *et al.*, 2020. Structure of a filamentous virus uncovers familial ties within the archaeal virosphere. *Virus evolution* 6, veaa023.

wrap around dsDNA, in a manner not observed in known viruses, and transforms it into A-form (Fig. 4(b)). Such arrangement keeps DNA inaccessible for solutes and ensures virion stability in highly aggressive environmental conditions. The MCPs have a molecular mass of about 14.5 kDa. *Rudiviridae* encode a single MCP, whereas each of the two other families encodes two paralogous MCPs. Remarkably, although the MCPs from the three virus families do not share significant sequence similarity, their structures are highly similar – they all carry unique four-helix bundle fold, not seen in the capsid proteins of other known viruses. The structural similarity of the MCPs suggests common origin of the MCPs and provides arguments to postulate evolutionary relationship between the three virus families which employ these MCPs. Thus, it was proposed to unify the families *Rudiviridae*, *Lipothrixviridae*, and *Tristromaviridae* into the new virus class *Tokiviricetes*. The former two families reveal certain similarity also at the genomic level, suggesting closer relationship. Thus, *Rudiviridae* and *Lipothrixviridae* were unified also in a taxon of a lower rank, the order *Ligamenvirales*.

The major difference in virion architecture of the three families concerns coating of the nucleoprotein core. In the *Rudiviridae*, it is non-enveloped, in *Lipothrixviridae*, covered with lipid envelope, and in *Tristromaviridae*, covered by a protein matrix which mediates contact with the outer lipid envelope. The matrix layer of the *Tristromaviridae* is formed by an 18 kDa capsid protein, VP3. As in other crenarchaeal viruses, *Lipothrixviridae* and *Tristromaviridae* recruit the lipids for formation of the envelope from the pool of host lipids.

The thickness of the lipid envelope in the reconstructed virions of the lipothrixviruses is about 20–25 Å, only half of the thickness of the cellular membrane. The host membrane, as in all Crenarchaeota, represents a 40 Å monolayer of glycerol dibiphytanyl glycerol tetraether (GDGT) lipids with different numbers of cyclopentane rings. Studies on the composition and structure of the virion envelope showed that the virus selectively acquires more flexible GDGT lipids lacking cyclopentane rings, which can be bent into U-shaped, ‘horseshoe’ conformation, forming the 20 Å thick virion envelope. Apparently, the membrane with lipids in such conformation, which was previously never observed in the living world, has advantages for survival in aggressive conditions of the natural environment of the hyperthermophiles.

General comments

Summarizing the results on studies on morphology and morphogenesis of viruses of hyperthermophilic Crenarchaeota, it could be noted that packaging of viral genomes in the form of a nucleocapsid appears to be the major structural characteristic of crenarchaeal viruses. Based on the available information, the assembly of the virion core apparently occurs through concerted coating of the DNA with MCPs and spontaneous self-assembly of such nucleocapsids into minimum free energy structures of different shapes. The self-assembly pathway of virion morphogenesis is common for RNA viruses but not for dsDNA viruses, where capsid assembly and genome packaging are generally separated. dsDNA viruses of Bacteria, herpesviruses of Eukarya, and crenarchaeal *Turritoviridae* pack dsDNA in the naked state into the pre-assembled empty capsids. This process requires significant energy expenses — mainly because of the mutual repulsion of the negatively charged phosphate backbone of the DNA — and is facilitated by virus-encoded elaborate molecular machinery. Considering the complexity of such process, it is surprising that the pathways of virion morphogenesis which are common for crenarchaeal viruses are so rare in the world of dsDNA viruses.

In most crenarchaeal virus families, the nucleoprotein core is covered by lipid-containing envelopes. Moreover, the core of the *Portogloviridae* is encased by an icosahedral protein shell. The mechanisms of coating of the nucleoprotein cores remain

unknown. For *Fuselloviridae*, the final step of virion assembly was shown to occur in the course of budding from the host cell in a process, resembling the morphogenesis of enveloped viruses of eukaryotes. For other crenarchaeal viruses, the coating of virions with a lipid envelope takes place apparently intracellularly.

Genomes

The genomes of crenarchaeal viruses are generally small in size (Table 1), with the largest ones (60–70 kb) being found in members of the family *Bicaudaviridae*. The 5,278 bp-long genome of the clavavirus APBV1 is among the smallest known dsDNA genomes. The other record holder is the coil-shaped spiravirus ACV – its ssDNA genome with 24.9 kb is the largest among known ssDNA genomes. The genomes can be either circular or linear (Table 1). Viruses with linear genomes often carry inverted terminal repeats (ITRs) of different length and exploit different strategies to protect their genome termini, such as covalently closed hairpins or covalently attached terminal proteins.

Along with the extraordinary diversity of morphotypes, another outstanding feature of archaeal viruses is the uniqueness of their genome content, with a very low proportion of genes with recognizable homologs in public databases. Indeed, family-specific comparison of viral proteomes against the sequence databases revealed that ~85% of the crenarchaeal virus proteins do not have identifiable homologs when an E-value threshold of $< 1e-5$ is used.

Functional annotation of crenarchaeal proteins shows that very few of these proteins are homologous to any sequences in the public databases, be it proteins of other viruses or those of cellular organisms. Consequently, archaeal virus genomes remain a rich source of unknown genes, many of which could be responsible for unique mechanisms of virus-host interactions or possess unexpected properties.

Structural genomics have been performed to promote the functional characterization of *Fuselloviridae*, *Bicaudaviridae*, *Rudiviridae*, *Lipothrixviridae*, *Globuloviridae* and *Turriviridae* proteins. Protein structures were determined by X-ray crystallography, and nuclear magnetic resonance (NMR) spectroscopy. For some viral proteins, the structures displayed unique folds with no homologs in databases, hindering the assignment of putative functions. In few cases, the structural information suggested a function which was verified biochemically. This was the case, e.g., for a novel type of nuclease encoded by the lipothrixvirus AFV1 and replication initiator protein encoded by the rudivirus SIRV2.

Notably, none of the viruses encodes a recognizable DNA-dependent RNA polymerase and only members of two families – *Ampullaviridae* and *Ovaliviridae*, encode recognizable DNA-dependent DNA polymerases. However, crenarchaeal viruses commonly encode multiple transcription factors which may recruit and redirect the host transcription machinery to preferential expression of viral genes.

Comparative genomic analyses have revealed that hyperthermophilic viruses from different families share just a small group of common genes, suggesting independent origins of the distinct groups of archaeal viruses. Ten virus families each have only one or two members with the sequenced genome. Exceptions are *Fuselloviridae* and two families of the order *Ligamenvirales* with higher number of members with sequenced genomes. Comparison of these genomes facilitated functional predictions and contributed to understanding the evolutionary history of the corresponding virus families, and will be briefly detailed below.

Family *Fuselloviridae*

The genomes of all members of *Fuselloviridae* are highly similar at the nucleotide sequence level and display overall gene synteny. A high frequency of homologous recombination has been reported between family members. The majority of predicted ORFs of the fuselloviruses cannot be assigned a function based on homology with sequences in public databases. Comparison of genome sequences has revealed a set of 12 not contiguous “core” genes shared by all family members, suggesting a common evolutionary history despite differences in the geographical context. The “core” proteome includes structural proteins, predicted DnaA-like AAA + ATPase, transcriptional regulators and a tyrosine superfamily integrase, which is involved in the integration of the viral genome into the host tRNA gene. Accordingly, the infected cell cultures contain both the covalently closed circular form of the viral genome and the integrated provirus. Unlike in the case of temperate bacteriophages, the integrase gene of the fuselloviruses is partitioned into two fragments upon the viral genome integration. Notably, studies on the prototypical fusellovirus SSV1 have shown that SSV1 integrase is not essential for the viral cycle. Moreover, nearly half of the predicted SSV1 ORFs were shown to tolerate insertions or deletions without affecting the viral infectivity.

Order *Ligamenvirales*

All members of the order have linear dsDNA genomes. On the example of the genome of the rudivirus SIRV1 it was shown that the two strands of the linear dsDNA genomes are covalently linked at both ends, forming a continuous polynucleotide chain and producing terminal hairpin structures. For other ligamenviruses such detailed analysis of the terminal regions has not been performed. Rudiviruses carry terminal inverted repeats which differ in size and sequence. The longest with 2 kb is found in the rudivirus SIRV1. In both rudiviruses and lipothrixviruses active genome remodelling, involving both deletions and horizontal acquisition of new genes, has been documented. The comparison of the genomes of different members of the same family and different strains of the same species revealed multiple examples of genomic rearrangements caused by insertion/deletion of

sequences that do not disrupt the ORFs. In *Rudiviridae* and *Lipothrixviridae* such sequences were shown to carry 12 bp or multiples thereof. It was suggested that the latter might constitute genetic elements mobilized by archaeal intron splicing.

Members of the *Rudiviridae* share significant similarity in the genome sequence and synteny, and carry a set of conserved core genes, most of which are localized in the middle region of the linear genome. By contrast, the genomes of *Lipothrixviridae* display a considerable variation in gene order and content, suggesting a longer evolutionary history. Based on the sequence similarity, the genomes of *Lipothrixviridae* group into four clusters of related species which form the four genera in the family. The examples of intergenomic recombination have been observed between members of different genera.

Properties of about half of the proteins encoded by the rudiviruses SIRV1 and SIRV2 have been predicted as a result of analysis of their sequences, structures and biochemical characteristics. Such proportion of recognized gene functions is among the highest for crenarchaeal viruses. The proteins with predicted functions include transcription regulators with ribbon-helix-helix and helix-turn-helix motifs, glycosyltransferases, acetyltransferase, Holliday junction resolvase, methyltransferase, dUTPase, ssDNA-binding protein, ssDNA annealing ATPase, Cas4-like exonuclease, as well as a protein specifically interacting with the host proliferating cell nuclear antigen (PCNA) and presumably recruiting it for the replication of the viral genome. Moreover, three proteins of the virus SIRV2 have been identified as anti-CRISPR proteins, which inhibit type I and type III CRISPR-Cas immunity systems of the *Sulfolobus* host. Detailed studies on the life cycle of the rudivirus SIRV2 enabled identification of the viral protein involved in the unique mechanism of virion egress. The function of this protein could not be predicted based on its sequence and was verified only as a results of detailed studies on the virus life cycle, supporting the possibility that many genes of archaeal viruses are implicated in specific aspects of virus-host interactions.

Evolutionary relationships

The unique morphological and genomic features of crenarchaeal DNA viruses raise important questions regarding their origins. The analysis of the evolutionary relationships between all dsDNA viruses using the bipartite network approach, which traces connections between viral genomes through shared gene families, revealed that crenarchaeal viruses are largely disconnected from the global dsDNA virosphere. The families of crenarchaeal viruses are themselves largely disconnected from each other and share just a small group of common genes, suggesting that most families of crenarchaeal viruses have evolved independently of one another.

How and when did archaea-specific viruses originate? Why do they only infect archaea? There are at least two non-mutually exclusive explanations. Some of the archaea-specific virus groups could have emerged during the early stages of cellular evolution and been retained in the Archaea but lost in the domains Bacteria and Eukarya. Other archaeal virus groups could have evolved concomitantly with the Archaea or, even more recently, within specific archaeal lineages. The observed limited gene sharing between different groups of archaea-specific viruses seems to make the latter possibility particularly plausible.

Further Reading

- Arnold, H.P., Ziese, U., Zillig, W., 2000. SNDV, a novel virus of the extremely thermophilic and acidophilic archaeon *Sulfolobus*. *Virology* 272, 409–416.
- Bettstetter, M., Peng, X., Garrett, R.A., Prangishvili, D., 2003. AFV1, a novel virus infecting hyperthermophilic archaea of the genus *acidianus*. *Virology* 315, 68–79.
- Contursi, P., Fusco, S., Cannio, R., She, Q., 2014. Molecular biology of fuselloviruses and their satellites. *Extremophiles: Life Under Extreme Conditions* 18, 473–489.
- DiMaio, F., Yu, X., Rensen, E., *et al.*, 2015. Virology. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* 348, 914–917.
- Häring, M., Peng, X., Brügger, K., *et al.*, 2004. Morphology and genome organization of the virus PSV of the hyperthermophilic archaeal genera *Pyrobaculum* and *Thermoproteus*: A novel virus family, the Globuloviridae. *Virology* 323, 233–242.
- Häring, M., Rachel, R., Peng, X., Garrett, R.A., Prangishvili, D., 2005. Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, *Acidianus* bottle-shaped virus, from a new family, the Ampullaviridae. *Journal of virology* 79, 9904–9911.
- Iranzo, J., Koonin, E.V., Prangishvili, D., Krupovic, M., 2016. Bipartite network analysis of the archaeal virosphere: Evolutionary connections between viruses and capsidless mobile elements. *Journal of virology* 90, 11043–11055.
- Kasson, P., DiMaio, F., Yu, X., *et al.*, 2017. Model for a novel membrane envelope in a filamentous hyperthermophilic virus. *eLife* 6.
- Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., Koonin, E.V., 2018. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Research* 244, 181–193.
- Liu, Y., Ishino, S., Ishino, Y., *et al.*, 2017. A novel type of polyhedral viruses infecting hyperthermophilic archaea. *Journal of Virology* 91 (13).
- Mochizuki, T., Krupovic, M., Pehau-Arnaudet, G., *et al.*, 2012. Archaeal virus with exceptional virion architecture and the largest single-stranded DNA genome. *Proceedings of the National Academy of Sciences of the United States of America* 109, 13386–13391.
- Mochizuki, T., Sako, Y., Prangishvili, D., 2011. Provirus induction in hyperthermophilic archaea: Characterization of *Aeropyrum pernix* spindle-shaped virus 1 and *Aeropyrum pernix* ovoid virus 1. *Journal of Bacteriology* 193, 5412–5419.
- Mochizuki, T., Yoshida, T., Tanaka, R., *et al.*, 2010. Diversity of viruses of the hyperthermophilic archaeal genus *Aeropyrum*, and isolation of the *Aeropyrum pernix* bacilliform virus 1, APBV1, the first representative of the family Clavaviridae. *Virology* 402, 347–354.
- Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature Reviews Microbiology* 15, 724–739.
- Prangishvili, D., Koonin, E.V., Krupovic, M., 2013. Genomics and biology of Rudiviruses, a model for the study of virus-host interactions in Archaea. *Biochemical Society Transactions* 41, 443–450.
- Prangishvili, D., Krupovic, M., 2012. A new proposed taxon for double-stranded DNA viruses, the order “Ligamenvirales”. *Archives of Virology* 157, 791–795.
- Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2006. Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *Journal of Molecular Biology* 359, 1203–1216.
- Quax, T.E., Krupovic, M., Lucas, S., Forterre, P., Prangishvili, D., 2010. The *Sulfolobus* rod-shaped virus 2 encodes a prominent structural component of the unique virion release system in Archaea. *Virology* 404, 1–4.
- Quemin, E.R., Chlanda, P., Sachse, M., *et al.*, 2016. Eukaryotic-like virus budding in archaea. *mBio* 7.

- Rensen, E.I., Mochizuki, T., Quemin, E., *et al.*, 2016. A virus of hyperthermophilic archaea with a unique architecture among DNA viruses. *Proceedings of the National Academy of Sciences of the United States of America* 113, 2478–2483.
- Wang, F., Baquero, D.P., Su, Z., *et al.*, 2020. Structure of a filamentous virus uncovers familial ties within the archaeal virosphere. *Virus Evolution* 6, veaa023.
- Wang, H., Guo, Z., Feng, H., *et al.*, 2018. Novel *sulfolobus* virus with an exceptional capsid architecture. *Journal of Virology* 92.
- Wang, F., Liu, Y., Su, Z., *et al.*, 2019. A packing for A-form DNA in an icosahedral virus. *Proceedings of the National Academy of Sciences of the United States of America* 116, 22591–22597.

Euryarchaeal Viruses

Tatiana A Demina and Hanna M Oksanen, Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Å Ångström, 0.1 nm

bp Base pair

CRISPR Clustered Regularly-Interspaced Short
Palindromic Repeats

ds Double-stranded

EM Electron microscopy

ICTV International Committee on Taxonomy of Viruses

kbp Kilo base pairs

knt Kilo nucleotides

MCP Major capsid protein

nt Nucleotide

ORF Open reading frame

ss Single-stranded

T Triangulation number

Glossary

Alkaliphilic Having a requirement for an environment with high pH.

Circular permutation A change in the sequence of the linear DNA termini that does not alter the relative sequence (e.g., circular permutation of ABCDEFGH could generate BCDEFGHA, CDEFGHAB, etc).

Concatamer Two or more DNA molecules that are linked together to form a long, linear DNA molecule.

Halophilic Having a requirement for an environment with a high salt concentration.

Headful packaging The mechanism of packaging viral DNA based on the size of the virus head, rather than the length of the viral genome.

Integrase An enzyme which can integrate viral DNA into the genome of its host cell.

Lysogen A host cell that has been infected by a virus that remains dormant, despite the presence of viral DNA.

Methanogenic Having the ability to produce methane.

Prophage A virus that is dormant within the host cell either integrated into the host chromosome or replicating free like a plasmid.

Protein-primed DNA replication Duplication of a linear DNA genome being started via the interaction of DNA polymerase with a protein covalently linked to the DNA termini.

Temperate virus A virus that is able to infect a host, but remains dormant within the host cell.

Thermophilic Having a requirement for an environment with high temperature.

Virion Infectious mature virus particle.

Virulent virus A virus that is able to infect a host, replicate, and subsequently leave the host cell by lysing the host cell.

Introduction

Microorganisms belonging to the phylum *Euryarchaeota* inhabit diverse environments: halophilic euryarchaea dominate in hypersaline environments such as solar salterns and salt lakes, methanogenic euryarchaea are found in intestines, anoxic sediments, and sludge digesters, while thermophilic euryarchaea thrive in thermal environments, e.g., hot springs and deep-sea hydrothermal vents. Euryarchaea are also abundant and active in oceanic surface waters.

The first archaeal virus Hs1 of halophilic euryarchaeon *Halobacterium* was found in 1974, being originally described as a bacteriophage before it was recognized that Archaea and Bacteria form two different domains of life. During the last 10 years, the number of described euryarchaeal viruses has tremendously increased due to extensive virus surveys, especially by using samples from hypersaline environments. Today, more than 110 euryarchaeal virus isolates or virus-like particles (VLPs) are described. Euryarchaeal viruses are extremely diverse, having various virion morphologies, genome types, sequences, life cycles, and host ranges. Among the five different morphologies known for euryarchaeal viruses ([Fig. 1](#)), tailed icosahedral viruses form the most numerous group with all three tail types found in tailed dsDNA bacteriophages of *Caudovirales*. Other known morphologies are tailless icosahedral viruses with internal membrane, pleomorphic, spindle-shaped, and spherical viruses. The smallest euryarchaeal virus is the pleomorphic virus HRPV-1 with a diameter of 40 nm, while the largest capsid diameter is found in tailed icosahedral virus HHTV-1 (~95 nm).

So far, all known archaeal viruses have genomes of either linear or circular double-stranded (ds) or single-stranded (ss) DNA molecules. Euryarchaeal pleomorphic virus HRPV-1 was the first described archaeal virus with a ssDNA genome, but most of the euryarchaeal viruses known today have a linear dsDNA genome. The genome sizes range from 7,048 nucleotides (nt) of *Halorubrum* pleomorphic virus HRPV-1 to 143,855 kilo base pairs (kbp) of *Halogrammon* tailed virus HGTV-1. For ~50 euryarchaeal virus isolates, the whole genome sequence is available ([Tables 1–3](#)). Currently, euryarchaeal viruses are classified by the International Committee on Taxonomy of Viruses (ICTV) into the families *Sphaerolipoviridae*, *Pleolipoviridae*, and the newly proposed family *Halspiviridae*, but there is also a significant number of unclassified ones. High number of proviruses found in the chromosomal

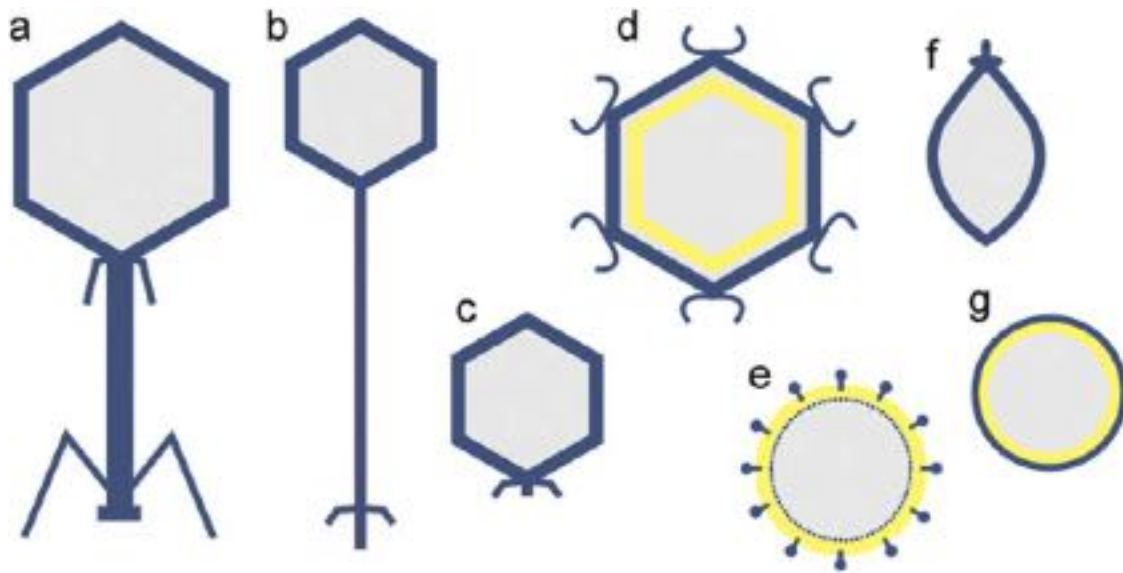


Fig. 1 Euryarchaeal virus morphotypes. (A–C) Icosahedral tailed virus morphotypes. (A) Myovirus morphotype, icosahedral head and long contractile tail. (B) Siphovirus morphotype, icosahedral head and long non-contractile tail. (C) Podovirus morphotype, icosahedral head and short non-contractile tail. (D) Internal membrane-containing icosahedral virus morphotype. (E) Pleomorphic virus morphotype without any protein capsid structure. (F) Spindle-shaped morphotype. (G) Internal membrane-containing spherical virus morphotype. Protein, blue; Lipid, yellow. Drawn not in scale.

Table 1 Icosahedral tailed dsDNA euryarchaeal virus isolates with a complete genome sequence available

Virus	Isolation host	Isolation source	Morphology, size and <i>T</i> number ^a	Genome topology, length (bp) and Acc. No. ^b
phiH	<i>Halobacterium salinarum</i>	<i>Halobacterium salinarum</i> culture	M 64/170	L (C as a provirus) 58,072 MK002701
ChaoS9	<i>Halobacterium salinarum</i>	<i>Halobacterium salinarum</i> culture	M 61/128	L CP 55,145 MK310226
HF2	<i>Halobacterium saccharovorum</i> Ch2	Saltern, Australia	M 58/94	L 77,670 AF222060
HF1	<i>Haloferax</i> sp. Aa2.2	Saltern, Australia	M 68/90	L 75,898 AY190604
HFTV1	<i>Haloferax</i> sp. LR2–5	Lake Retba, Senegal	S 50/60	L CP 38,059 MG550112
HCTV-1	<i>Haloarcula californiae</i>	Saltern, Margherita di Savoia, Italy	S 70/80	L 103,257 KC292029
HCTV-2	<i>Haloarcula californiae</i>	Saltern, Samut Sakhon, Thailand	S	L CP 54,291 KC292028
HCTV-5	<i>Haloarcula californiae</i>	Saltern, Samut Sakhon, Thailand	S	L 102,105 KC292027
HHTV-1	<i>Haloarcula hispanica</i>	Saltern, Margherita di Savoia, Italy	S 55/110	L CP 49,107 KC292025
HHTV-2	<i>Haloarcula hispanica</i>	Saltern, Samut Sakhon, Thailand	S	L CP 52,643 KC292024
HSTV-1	<i>Haloarcula sinaiensis</i>	Saltern, Margherita di Savoia, Italy	P 62 <i>T</i> = 7	L CP 32,189 KC117378
HVTV-1	<i>Haloarcula vallismortis</i>	Saltern, Samut Sakhon, Thailand	S 95.5/73 <i>T</i> = 13	L 102,319 KC117377
HGTV-1	<i>Halogramma</i> sp. SS5–1	Saltern, Samut Sakhon, Thailand	M	L CP 143,855 KC292026
HSTV-2	<i>Halorubrum sodomense</i>	Saltern, Eilat, Israel	M 73.8/101 <i>T</i> = 7	L 68,527 KC117376.1
HRTV-4	<i>Halorubrum</i> sp. s5a-3	Saltern, Margherita di Savoia, Italy	S	L CP 35,722 KC292023
HRTV-5	<i>Halorubrum</i> sp. s5a-3	Saltern, Margherita di Savoia, Italy	M	L 76,134 KC292022
HRTV-7	<i>Halorubrum</i> sp. B2–2	Saltern, Margherita di Savoia, Italy	M	L 69,048 KC292021
HRTV-8	<i>Halorubrum</i> sp. B2–2	Saltern, Samut Sakhon, Thailand	M	L 74,519 KC292020
BJ1	<i>Halorubrum</i> sp. BJ1 B11	Lake Bagaejinnor, Mongolia, China	S 56/71	L 42,271 AM419438
phiCh1	<i>Natrialba magadii</i>	<i>Natrialba magadii</i> culture	M 70/130	L (C as a provirus) 58,498 AF440695 MK450543
Drs3	<i>Methanobacterium formicicum</i> Kh110	Anaerobic sludge digester	S 60/230	L 37,129 MH674343
psiM1	<i>Methanothermobacter marburgensis</i>	Anaerobic sludge digester	S 55/210	L 26,803 AF065411, AF065412
psiM2	<i>Methanothermobacter marburgensis</i>	Anaerobic sludge digester (deletion mutant of psiM1)	S 55/210	L 26,111 AF065411

^aMorphotypes M, S, and P correspond to myovirus (long contractile tail), siphovirus (long non-contractile tail), and podovirus (short non-contractile tail) morphologies. Head diameter and tail length are given for myoviruses and siphoviruses, and head diameter for podoviruses (nm). Head diameters of HSTV-2, HVTV-1, and HSTV-1 are based on cryo-EM (vertex to vertex).

^bL, linear; C, circular; CP, circularly permuted.

Table 2 Icosahedral internal membrane-containing euryarchaeal dsDNA virus isolates with a complete genome sequence available

Virus	Isolation host	Isolation source	Virion size, <i>T</i> number ^a	Genome topology, length (bp) and Acc. No. ^b
SH1	<i>Haloarcula hispanica</i>	Lake Serpentine, Australia	80.0, <i>T</i> = 28 dextro	L 30,889 AY950802
PH1	<i>Haloarcula hispanica</i>	Pink Lake, Australia	50, nd	L 28,072 KC252997
HHIV-2	<i>Haloarcula hispanica</i>	Saltern, Margherita di Savoia, Italy	80.0, <i>T</i> = 28 dextro	L 30,578 JN968479
HCIV-1	<i>Haloarcula californiae</i>	Saltern, Samut Sakhon, Thailand	80.0, <i>T</i> = 28 dextro	L 31,314 KT809302
SNJ1	<i>Natrinema</i> sp. J7-1	Salt mine, Hubei, China (induced from <i>Natrinema</i> sp. J7-1 and grown on <i>Natrinema</i> sp. J7-2)	70-75	C 16,341 AY048850.1

^aDiameter of the capsids are given (nm). SH1, HHIV-2, and HCIV-1 diameters are based on cryo-EM.

^bL, linear; C, circular.

Table 3 Pleomorphic, spindle-shaped, and spherical euryarchaeal virus isolates with a complete genome sequence available

Virus	Isolation host	Isolation source	Morphology and size ^a	Genome topology, length (bp or nt) and Acc. No. ^b
HRPV-1	<i>Halorubrum</i> sp. PV6	Saltern, Trapani, Italy	P 41	ssDNA C 7048 FJ685651
HRPV-2	<i>Halorubrum</i> sp. SS5-4	Saltern, Samut Sakhon, Thailand	P 54	ssDNA C 10,656 JN882264
HRPV-3	<i>Halorubrum</i> sp. SP3-3	Salt water, Sedom Ponds, Israel	P 67	dsDNA C 8770 JN882265
HRPV-6	<i>Halorubrum</i> sp. SS7-4	Saltern, Samut Sakhon, Thailand	P 49	ssDNA C 8549 JN882266
HRPV9	<i>Halorubrum</i> sp. SS5-4	Culture supernatant of <i>Halorubrum</i> sp. B2-2 on <i>Halorubrum</i> sp. SS5-4 culture	P 57	dsDNA C 16,159 KY965934
HRPV10	<i>Halorubrum</i> sp. LR2-17	Lake Retba, Senegal	P 55	dsDNA C 9296 MG550111
HRPV11	<i>Halorubrum</i> sp. LR2-12	Lake Retba, Senegal	P 55	dsDNA C 9368 MG550113
HRPV12	<i>Halorubrum</i> sp. LR1-23	Lake Retba, Senegal	P 55	dsDNA C 9944 MG550110
HPV-1	<i>Haloarcula hispanica</i>	Saltern, Margherita di Savoia, Italy	P 52	dsDNA C 8082 GU321093
HPV-2	<i>Haloarcula hispanica</i>	Saltern, Hulu Island, Liaoning, China	P 50	ssDNA C 8176 KF056323
HPV-3	<i>Haloarcula hispanica</i>	Saltern, Samut Sakhon, Thailand	P 50	dsDNA C 11,648 KX344510
HPV-4	<i>Haloarcula hispanica</i>	Culture supernatant of <i>Haloferax</i> sp. s5a-1 on <i>Har. hispanica</i> culture	P 60	dsDNA C 15,010 KY264020
His2	<i>Haloarcula hispanica</i>	Pink Lakes, Victoria, Australia	P 71	dsDNA L 16,067 AF191797
HGPV-1	<i>Halogeometricum</i> sp. CG-9	Saltern, Cabo de Gata, Spain	P 56	dsDNA C 9694 JN882267
SNJ2	<i>Natrinema</i> sp. J7-1	<i>Natrinema</i> sp. J7-1 culture	P 70-80	dsDNA C 16,992 AJVG01000023 (WGS contig04, 19,792-36,797)
His1	<i>Haloarcula hispanica</i>	Saltern, Avalon, Victoria, Australia,	L 92 × 40	dsDNA L 14,462 AF191796
TPV1	<i>Thermococcus prieurii</i>	Deep-sea hydrothermal vents, East Pacific Rise	L 140 × 80	dsDNA C 21,592 JQ010983
PAV1	<i>Pyrococcus abyssi</i>	<i>Pyrococcus abyssi</i> culture	L 120 × 80	dsDNA C 18,098 EF071488
MetSV	<i>Methanosarcina mazei</i> Gö1	Anaerobic sewage sludge	S 56.6	dsDNA L 10,567 MF186604

^aMorphotypes P, L, and S correspond to pleomorphic, lemon-shaped (spindle-shaped), and spherical morphologies. Diameter (for P and S) or length and width (for L) are given (nm). Nd, not determined.

^bL, linear; C, circular.

sequences of euryarchaea and metagenomic datasets indicate that the viruses known today represent only a minuscule fraction of the true euryarchaeal viral diversity. Here, we review advances in understanding viruses infecting euryarchaea by summarizing the data from culture-independent studies and describing known virus isolates.

Icosahedral Tailed Viruses

To date, about 90 euryarchaeal icosahedral tailed viruses have been isolated, the majority of which are viruses infecting halophilic archaea of the class Halobacteria and the rest infect methanogenic archaea belonging to the family *Methanobacteriaceae* (Table 1). Archaeal tailed viruses morphologically resemble bacteriophages of the order *Caudovirales*, with an icosahedral head and either a long contractile (myovirus) or non-contractile (siphovirus) tail or a short non-contractile tail (podovirus) (Fig. 1(A)-(C)). Of these three morphotypes, myovirus isolates are the most numerous, while only one archaeal podovirus HSTV-1 of *Haloarcula sinaiensis* is known. All currently known euryarchaeal icosahedral tailed viruses have dsDNA genomes (Table 1), which are reminiscent to the genomes of tailed bacteriophages in the order *Caudovirales*. Tailed archaeal viruses resemble tailed bacteriophages by a modular structure and mosaicism of their genomes, although many ORFs have no assigned functions. The gene for the large terminase subunit is one of the most conserved genes in tailed dsDNA phages and it is annotated in all euryarchaeal tailed virus genomes

described to date. Most archaeal viruses have not been taxonomically assigned to any taxa, except some of the oldest isolates, e.g., phiH, Hs1, and HF2, but obvious similarities with tailed dsDNA phages suggest that all archaeal tailed viruses could be members of the order *Caudovirales*, but having their own family level taxa. Notably, no tailed viruses have been yet isolated for crenarchaeal hosts, although microscopic and metagenomics studies of hyperthermic environments, where crenarchaea typically dominate, have revealed the presence of tailed VLPs and sequences related to phages of *Caudovirales*.

Icosahedral Tailed Viruses of Halophilic Euryarchaea

Since the discovery of the first archaeal virus Hs1 in 1974, other tailed halophilic archaeal viruses have been isolated: Ja1 (1975), B10 (1982), phiH (1982), Hh-1 and Hh-3 (1982), S45 (1984), phiN (1988), S5100 (1990), HF1 and HF2 (1993), phiCh1 (1997), S41, S50.2, and S4100 (1998). From these viruses, only a few have been characterized at the molecular level, and some of the early isolates have been lost later. In 2000s, a few more tailed haloarchaeal viruses have been reported: BJ1 (2007), HHTV-1, HRTV-1, and HCTV-1 (2009). Two large virus surveys of hypersaline environmental samples published in 2012 and 2015 yielded close to 60 new tailed archaeal virus isolates infecting halophilic strains of the genera *Haloarcula* or *Halorubrum* and one *Halogramma* tailed virus. The most recently described isolates are ChaoS9 and HFTV1 (2019), of which HFTV1 represents the first virus infecting *Haloferrax* strain. In addition, several putative proviruses related to tailed viruses have been detected in the genomes of halophilic archaea, showing their wide distribution.

One of the earliest euryarchaeal virus isolates is the temperate myovirus phiH, which was obtained from a spontaneously lysed culture of *Halobacterium salinarum* strain R1. It was actively studied until late 1990s, when the work on this virus isolate stopped due to its genomic instability. phiH virion displays a typical myovirus morphotype and it requires high salt concentrations (3 M KCl or NaCl) for stability. The linear dsDNA genome of phiH is packaged into viral particles, but is a circular plasmid-like molecule in infected cells. Viral DNA is terminally redundant and partially circularly permuted, being packed by a headful mechanism. Recently, the genome of phiH1, a predominant variant of phiH, has been fully sequenced and the annotation has been refined (Table 1). During replication rounds, a mixture of genetically rearranged phiH variants is produced due to the activity of insertion sequences. A so-called L-segment, flanked by insertion sequence(s), may undergo inversions or circularize into an autonomous 12-kbp plasmid phiHL, which provides partial immunity from the phiH infection. phiH1 has been largely used in molecular studies of archaeal gene expression and regulation, resulting in e.g., development of transfection methods in archaea.

Another temperate myovirus, phiCh1, has been isolated from a spontaneously lysed culture of haloalkaliphilic *Natrialba magadii* (Table 1). Despite having very different hosts, phiCh1 is closely related to phiH. Peculiarly, in addition to viral DNA, phiCh1 virions contain several host-derived RNA species of 70–800 nt in length. Unlike phiH, phiCh1 integrates into the host's chromosome. The replicative form is circularly permuted and terminally redundant, implying a headful mechanism of packaging. phiCh1 genome of 58,498 bp is methylated and arranged into three transcriptional units. phiCh1 and phiH1 genomes display similar gene synteny and share 63% nucleotide identity. Similarly to phiH1, phiCh1 demonstrates genomic rearrangements. The rearrangements caused by the lambda-like integrase Int1 result in the inversion in the region of two genes encoding putative tail fiber proteins, which affects host recognition specificity. Genomic rearrangements have been also suggested in haloarchaeal myoviruses S41 and S50.2 of *Hbt. salinarum* and siphovirus BJ1 of *Halorubrum*.

A myovirus related to phiH and phiCh1, ChaoS9, has been recently isolated after a spontaneous lysis of a *Hbt. salinarum* culture. The left-end region (first 25 kbp) of ChaoS9 genome (Table 1) shares similarities in sequence and gene synteny with phiCh1 and phiH, while the right end (from 25 to 55 kbp) is significantly more divergent. The viruses share overall nucleotide identity of over 74%. The left-end region of ChaoS9 carries tail fiber genes, which are similar to those in phiCh1 and phiH, and other structural proteins, while the right-end region encodes proteins involved in DNA replication and lysogeny. ChaoS9 major capsid protein (MCP) displays 36% amino acid identity to that of siphovirus HHTV-1 of *Haloarcula hispanica*. Putative head and assembly proteins are unrelated in ChaoS9, phiH, and phiCh1. The revealed similarities in tail morphogenesis module and differences in head morphogenesis module suggest that large recombination events may have shaped the genomes of these viruses. This kind of genomic plasticity is also typical for tailed bacteriophages of the order *Caudovirales*.

One more example of two closely related haloarchaeal tailed viruses is HF1 and HF2, both isolated from Cheetham Saltworks in Victoria, Australia. These myoviruses are virulent, displaying particles of a similar size and genomes of a similar length (76–77 kbp). Both viruses may enter unstable carrier states, when viruses are produced continuously, but eventually switch to a complete lysis. However, HF1 and HF2 have distinct host ranges with no common hosts. HF1 and HF2 genomes are almost identical in their first 48 kbp, while the right-end regions are 87% identical. Such genomic structure may be a result of a recent recombination event with some other HF-related virus. The divergent right-end part of virus genomes probably contains late genes, encoding viral structural proteins, possibly explaining different host ranges of these viruses. Virus genomes are mosaic and encode own DNA polymerases.

In the first half of the 2010s, global sampling of hypersaline environments resulted in the isolation of dozens of new tailed haloarchaeal viruses. It has been shown that tailed haloarchaeal viruses are able to infect host strains originating from geographically very distant sites or from the same location sampled over a few years. Some of these viruses can also infect strains belonging to different genera. Some myoviruses isolated from the Samut Sakhon solar saltern, Thailand, have shown notably wide host ranges, while some of *Halorubrum* strains from the same location were highly susceptible to virus infections.

A few tailed euryarchaeal viruses have been structurally characterized using cryo-electron microscopy (EM) and 3D image reconstruction. The most detailed structure (~9 Å resolution) has been resolved on HSTV-1, which is so far the only known archaeal podovirus (Table 1). *Haloarcula sinaiiensis* virus HSTV-1 is virulent, and its infectivity depends on salinity. At low salinity,

HSTV-1 is reversibly non-infectious, regaining its infectivity when transferred back to high salinity. This kind of characteristic is very advantageous in environments with salinity fluctuations, caused by e.g., changing weather conditions. The majority of HSTV-1 ORFs have no assigned functions, while other ORFs are similar to those involved in DNA replication and nucleotide metabolism or virion assembly in tailed bacteriophages. The order of genes encoding structural proteins is also same as observed in bacteriophage counterparts. HSTV-1 particles display an icosahedral head with a diameter of 560 Å measured facet to facet and 624 Å vertex to vertex (Fig. 2(A) and (B)). The short podovirus-like tail with base plate tail fibers functions in host recognition. The capsid is decorated with 20 large (at least 138 Å long and 74 Å wide), cone-shaped towers, which are located at positions of threefold symmetry (Fig. 2(B)). The capsid shell is 20 Å thick and is arranged into $T = 7$ laevo lattice. HSTV-1 MCP adopts the bacteriophage HK97 fold (Fig. 2(C)), thus suggesting evolutionary links between archaeal and bacterial tailed viruses and eukaryotic herpesviruses, all having the common HK97 fold in their MCPs. The structural similarity between different viral MCPs (without having any sequence similarity) unites these diverse viruses into one of the structure-based viral lineages (HK97-like viruses). In addition, based on homology modeling, phiCh1 MCP could also adopt the HK97 fold. Most probably, all archaeal tailed viruses share the same conserved HK97 fold to build up their capsids.

The other structurally characterized tailed viruses are siphovirus HVIV-1 of *Haloarcula vallismortis* and myovirus HSTV-2 of *Halorubrum sodomense*. Like podovirus HSTV-1, the infectivities of HVIV-1 and HSTV-2 are reversibly lost in low-salt conditions. Both viruses are virulent with quite long intracellular phases: cells are lysed after 10–12 h of infection. HVIV-1 virion consists of an icosahedral capsid enclosing the linear genome of 101,734 bp and a tail typical for siphoviruses – long and flexible (Table 1). Cryo-EM revealed that the icosahedral capsid of HVIV-1 is arranged into $T = 13$ lattice and decorated with five-fold symmetric spikes at vertices and trimeric decorative structures at the center of each MCP hexamer. HSTV-2 genome (Table 1) shares ~70 and ~60% of overall nucleotide identity with HF1 and HF2 genomes, respectively. HSTV-2 capsid (Table 1) is arranged into $T = 7$ lattice being smaller than

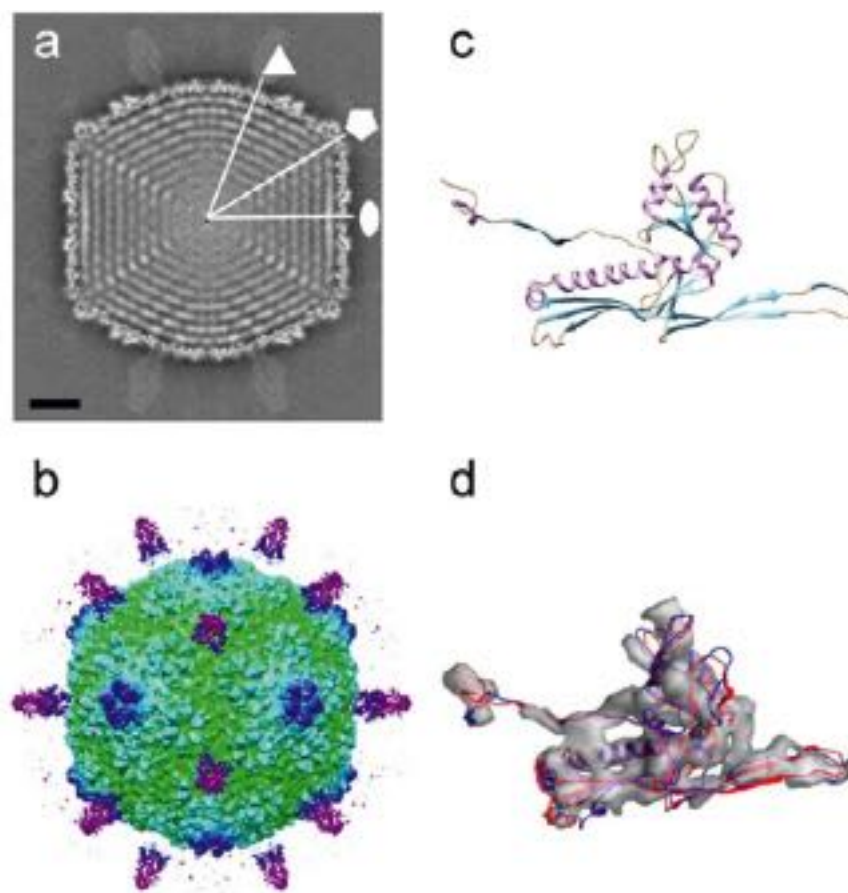


Fig. 2 Structure of the haloarchaeal podovirus HSTV-1 head based on cryo-EM and icosahedral reconstruction, where the tail is invisible. (A) Central section of the icosahedral reconstruction of HSTV-1 is shown viewed along an icosahedral twofold symmetry axis (ellipse). Threefold (triangle), and fivefold (pentagon) symmetry axes are indicated. Bar is 10 nm. (B) 3D isosurface representation of HSTV-1. (C) Bacteriophage HK97 MCP fold (Protein Data Bank code 1OHG). (D) Averaged density showing the MCP fold of HSTV-1 calculated from the six segmented subunits from one HSTV-1 MCP hexamer. HK97 MCP fold (red ribbon) and HSTV-1 MCP fold (blue ribbon) fitted into the EM density in gray (see Panel B). Reproduced from Atanasova, N.S., Bamford, D.H., Oksanen, H.M., 2015. Haloarchaeal virus morphotypes. *Biochimie* 118, 333–343 with a permission from Elsevier.

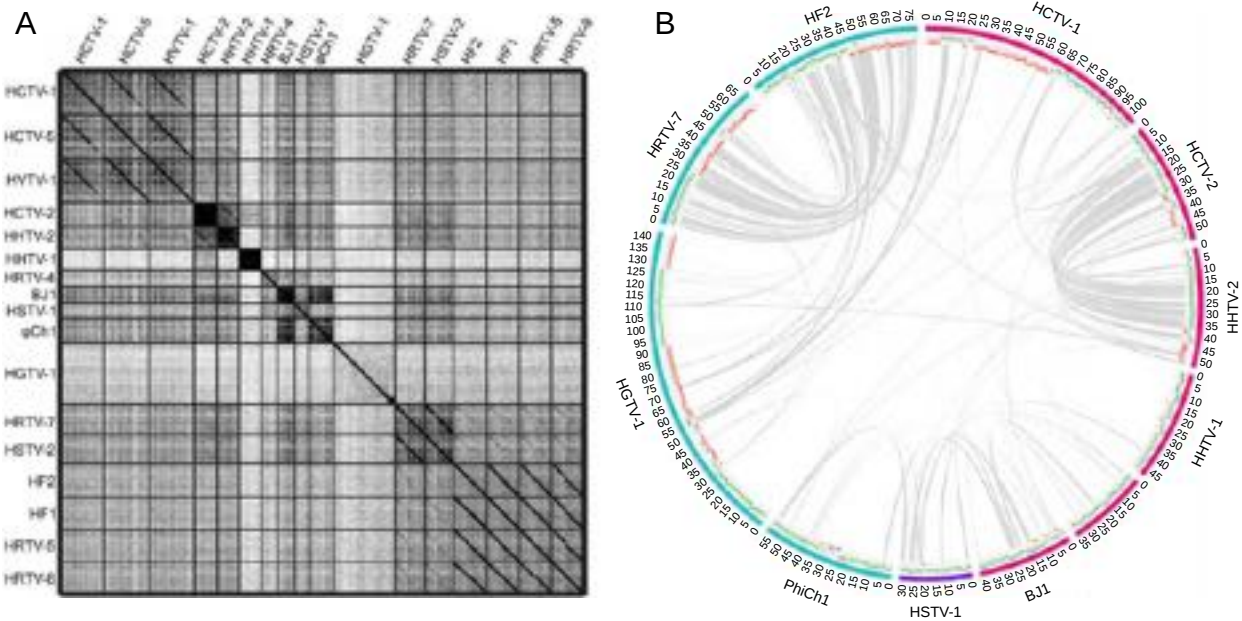


Fig. 3 Genomic comparisons of the 17 haloarchaeal tailed viruses with completely sequenced genomes. All viruses with complete genome sequence can be found in [Table 1](#). (A) Dotplot alignment of the genomes. (B) Circular visualization of the homologous proteins shared between the selected virus representatives from each of the delineated groups and singletons (myoviruses in blue; siphoviruses in pink; podoviruses in purple). The outermost circle represents the genome maps with the coordinates (kbp) followed by the ORFs (on positive strand in green; on negative strand in red). Putative homologs (sharing over 30% identity) are connected by a gray line. Reproduced from Senčilo, A., Roine, E., 2014. A glimpse of the genomic diversity of haloarchaeal tailed viruses. *Frontiers in Microbiology* 5, 84.

that of HVTV-1. Minor trimeric proteins between the MCPs at the local three-fold positions stabilize the capsid, allowing enlargement of HSTV-2 capsid and fitting of a larger genome, which is a novel strategy not previously observed in virus capsids.

Up to date, from the isolated haloarchaeal tailed viruses, 20 have completely sequenced genomes ([Table 1](#)). In addition, the Genbank database contains a record of CGphi46 virus genome sequence. Based on the description available in Genbank, CGphi46 infects *Halorubrum* sp. and was isolated from Cargill Saltern, USA. The genome is a linear dsDNA of 39,784 bp, displaying similarities to tailed haloarchaeal virus sequences, such as *Halorubrum* virus BJ1. Generally, the genomes of haloarchaeal tailed viruses are dsDNA molecules ranging in length roughly from 32 kbp (HSTV-1) to 144 kbp (HGTV-1) with GC content typically similar to that of their hosts. Some conserved bacteriophage genes, such as for the large terminase subunit, are widely found in haloarchaeal tailed virus genomes, while the majority of predicted ORFs have no assigned functions. Based on the whole genome sequences, a few groups of similar viruses can be distinguished: (1) HF1, HF2, HRTV-5, and HRTV-8, (2) HRTV-7 and HSTV-2, (3) HCTV-1, HCTV-5, and HVTV-1, while other viruses are so far singletons ([Fig. 3](#)). Clearly, many more sequences are needed to gain more insights into genetic diversity of tailed archaeal viruses and their taxonomy and to resolve their relatedness to other viral groups.

Tailed Viruses of Methanogenic Euryarchaea

Since 1980s, several viruses infecting methanogenic archaea have been reported: PG (1984), PMS1 (1986), psiM1 and psiM2 (1989), phiF1 and phiF3 (1993), and Drs3 (2018). In addition, tailed VLPs have been found associated with *Methanococcus voltae* PS. Icosahedral VLPs with long contractile or non-contractile tails (myoviruses and siphoviruses) have been found in high abundances in acetate-fed anaerobic digester sludge reactors dominated by *Methanosaeta*. High numbers of VLPs, including tailed ones, have also been reported in other anaerobic digesters inhabited with methanogens. Moreover, several proviruses have been identified in the genomes of methanogenic archaea belonging to the orders *Methanobacteriales*, *Methanococcales*, and *Methanosarcinales*.

Siphovirus psiM1 was isolated from an anaerobic sludge digester (55–60°C) on *Methanothermobacter marburgensis* ([Table 1](#)). Circularly permuted and terminally redundant linear genome of psiM1 is packaged from a concatemeric precursor via a headful mechanism. psiM2 is a deletion derivative of the wild type psiM1 and lacks a 0.7-kbp deletion segment, which makes it more stable than the wild type. Both variants are virulent, encoding pseudomurein endoisopeptidase, an enzyme that cleaves host pseudomurein. However, a putative integrase gene is also found in psiM1 and psiM2. It has been observed that ~15% of psiM1 particles carry multimers of a cryptic 4.5-kbp host plasmid pME2001, suggesting that psiM1 can mediate general transduction. Indeed, the virus has been shown to transduce some host chromosomal markers. A defective prophage related to psiM1 and

psiM2, psiM100, can be found in the chromosome of *Methanothermobacter wolfeii*. Upon hydrogen limitation, *Methanothermobacter wolfeii* cultures spontaneously lyse, however, without producing VLPs, but the autolysate contains pseudomurein endoisopeptidase encoded by psiM100. psiM100 genome is ~28.8 kbp and is 70.8% identical to psiM2, except for the region of 2.8 kbp, which is apparently an insertion from a virus non-related to psiM1 or psiM2.

Two virulent viruses, phiF1 and phiF3, infecting *Methanobacterium thermoformicum* have been isolated from an anaerobic sludge reactor. phiF1 has an icosahedral head of ~70 nm in diameter and a non-flexible tail of ~160 nm in length, while phiF3 has an icosahedral head of ~55 nm in diameter and a flexible 230-nm long tail. The two viruses differ significantly in their host ranges, genome organization, and size. From the tested *Methanobacterium* strains, phiF1 infected *M. thermoautotrophicum* strain ΔH and several strains of *M. thermoformicum*, while phiF3 was able to infect only *M. thermoformicum* strain FF3. phiF1 has a linear dsDNA genome of ~85 kbp, while phiF3 genome is a circular or terminally redundant linear dsDNA molecule of ~36 kbp. No similarity between their sequences were found in hybridization experiments. There is no complete genome sequence available for these viruses. The other siphovirus infecting methanogenic euryarchaea, Drs3, (Table 1) is only distantly related to psiM2 or psiM100, having no predicted functions for the majority of ORFs.

Icosahedral Internal Membrane-Containing Viruses

Currently, five icosahedral internal membrane-containing viruses infecting euryarchaea have been described, SH1 being the first virus isolate of this group (found 2005; Table 2). Their genomes are either linear or circular dsDNA molecules. These viruses are classified into the family *Sphaerolipoviridae* together with the phages P23–77 and IN93 infecting thermophilic *Thermus* bacteria. *Alphasphaerolipovirus* genus includes virus species of *Haloarcula hispanica* viruses SH1, PH1, and HHIV-2, and *Haloarcula californiae* virus HCIV-1. *Natrinema* virus SNJ1 is the sole member of the genus *Betasphaerolipovirus*, and the bacteriophages have their own genus *Gammasphaerolipovirus*.

Archaeal sphaerolipoviruses have been isolated from different hypersaline environments: Australian salt lakes and solar salterns in Italy and Thailand (Table 2). Their host cells are either *Haloarcula* or *Natrinema* species. SH1, PH1, HHIV-2, and HCIV-1 are virulent and their progeny is released by host cell lysis. SNJ1 is temperate and initially found from lysogenic strain *Natrinema* sp. J7–1 originating from a salt mine, Hubei province, China. The strain J7–1 contains a plasmid pHH205, whose sequence is identical to that of SNJ1 and it is the proviral element of SNJ1 replicating freely in the cell. SNJ1 employs a rolling-circle mechanism of genome replication and harbors a gene encoding replication initiation protein. SNJ1 can be induced by mitomycin C and plated on *Natrinema* sp. J7–2 strain to obtain plaques. *Natrinema* sp. J7–1 is also the source of another halophilic virus SNJ2, which is a pleomorphic virus (Table 3).

Archaeal sphaerolipovirus virions are icosahedral, ~80 nm in diameter, and contain more than 12 different protein types, which are either forming the capsid, receptor recognizing spike complexes, or associating with the membrane. Underneath the icosahedral capsid, the protein-rich membrane bilayer encloses the genome. This virion morphology is also found among crenarchaeal (e.g., turritivirus STIV and STIV-2) and bacterial viruses (e.g., tectivirus PRD1 and corticovirus PM2). The recently obtained cryo-EM based whole virion structures of HCIV-1, HHIV-2, and SH1 at 3.7–3.8 Å resolution revealed that virus capsids are arranged in a pseudo $T = 28$ dextro lattice composed of two MCP species (Fig. 4). MCP VP7 is smaller and composed of one single β -barrel, whereas the large MCP VP4 is composed of two β -barrels. One β -barrel is forming the lattice and the other one is located on top of the lattice-forming β -barrel, thus forming the tower structure on the capsomer surfaces (Fig. 4(B)). The MCPs are conserved among the archaeal sphaerolipoviruses and the amino acid similarities of the large MCP VP4 and small MCP VP7 are 90%–97% and 86%–99%, respectively. MCPs VP4 and VP7 form pseudo-hexameric capsomers with either two or three towers, which are built of VP7-VP4 heterodimers and monomeric VP7. The capsid assembly is coordinated by the conserved penton protein complexes at the five-fold vertices and two specific membrane associated GPS-II and GPS-III complexes. The bacterial sphaerolipovirus P23–77 capsid has the same $T = 28$ organization, but the sequences of its MCPs are divergent from those of the archaeal counterparts although having similar single β -barrel folds.

The multi-protein spike complexes are either dimeric horn-shaped structures (SH1, HCIV-2, and PH1) or pentameric propeller-like complexes (HHIV-2) plugged into the five-fold vertices serving as host recognition devices. The viral membranes are composed of virus-encoded proteins and host-derived lipids, which are selectively acquired from the host membranes during the virion assembly. The membrane is closely associated with the capsid, being connected to the capsid at least by the penton and GPS-complexes. The major lipid species in HCIV-1, HHIV-2, and SH1 virions are phosphatidylglycerol, phosphatidylglycerophosphate methyl ester, and phosphatidylglycerosulfate. In a similar way, the fatty acids of SNJ1 are of the host origin, but the composition is a bit different than in viruses of *Haloarcula*.

The ~30 kbp genomes of HCIV-1, HHIV-2, SH1, and PH1 are linear and have inverted terminal repeats and terminal proteins attached to the ends of the molecules, proposing a protein-primed mechanism of genome replication. However, no canonical DNA polymerase-encoding gene such as in the genome of bacteriophage phi29 is found in viral or host genomes, indicating that a different mode of replication might be employed. The genomes of HCIV-1, HHIV-2, SH1, and PH1 share over 56% nucleotide identity. SH1 and PH1 are the most closely related viruses with ~76% identity. The four genomes are co-linear and the organization of their ~50 ORFs is conserved (Fig. 5).

The circular SNJ1 genome sequence shares little similarity to other archaeal sphaerolipoviruses at the nucleotide level. However, the arrangement of the viral core genes encoding putative genome packaging ATPase and the small and large MCPs is

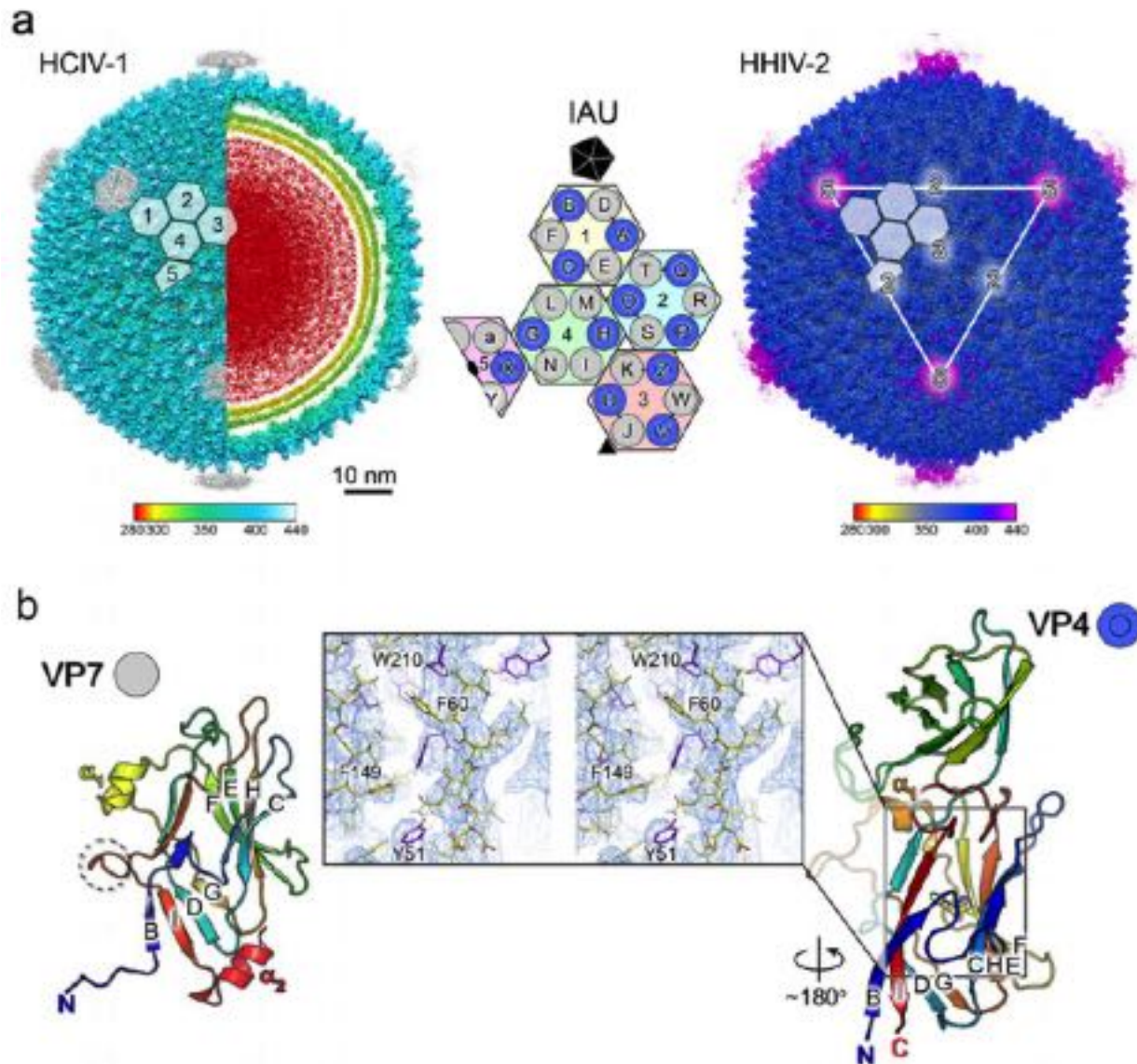


Fig. 4 Cryo-EM structures of HCIV-1 and HHIV-2 virions and their MCPs. (A) Cryo-EM density maps of HCIV-1 (left) and HHIV-2 (right) (vertex complexes have been omitted). HCIV-1 is rendered to display the capsid shell (left-half) and the particle interior (right-half). Color-coding by distance from the center (legend below); genome, red; inner and outer membrane leaflets, yellow and yellow-lime. Icosahedral asymmetric unit (IAU) and capsomer organization are shown on HCIV-1 surface (white-transparent hexagons) and at the center. Capsomers are numbered Nos. 1–5; the three-tower capsomers (No. 1, light-yellow; No. 2, cyan; No. 3, pink) and two-tower capsomers (No. 4, light-green; No. 5, light-magenta). Two MCPs VP4 and VP7 are visualized by blue and light-gray circles. The MCP subunits are labeled (A–X and a). Black short-lines joining the circles identify the VP7–VP4 heterodimers. Black pentagons represents the positions of the penton complex plugging the vertices. Facet of the virion and the icosahedral symmetry axes (two, three, and five-fold) are marked on HHIV-2 by a white triangle and numbers (2, 3, and 5), respectively. (B) Single β -barrel MCPs VP7 (left) and VP4 (right) of HCIV-1. VP4 is composed of two vertical single β -barrels, one standing on top of the other, while VP7 consists of a vertical single β -barrel. The four-stranded sheets BIDG and CHEF are labeled in VP7. A loop in VP7 (residues 149–154) is marked by a dashed circle. Region of the VP4 atomic model is shown as a stereoview (black rectangles) and fitted into the corresponding 3.7 Å resolution density map. Reproduced from Santos-Pérez, I., Charro, D., Gil-Carton, D., *et al.*, 2019. Structural basis for assembly of vertical single β -barrel viruses. *Nature Communications* 10, 1184.

conserved. These three genes are the most conserved among HCIV-1, HHIV-2, SH1, and PH1, and can be also found from bacteriophages belonging to the same family. The least conserved genes are encoding the receptor recognition complexes. The HHIV-2 spike complex genes are different from those found in the same position of the genomes of HCIV-1, SH1, and PH1 (Fig. 5) as are the structures of their vertex complexes. Putative proviruses HalaCibP1, HalaPauP1, and HaloLacP1 identified in the chromosomes of *Haladaptus cibarius*, *Haladaptus paucihalophilus*, and *Halobiforma lacisalsi* resemble HCIV-1, HHIV-2, SH1, and PH1 viruses and have signatures of putative integrase or transposase and tRNA genes flanking the sphaerolipovirus-like provirus sequences, indicating that some of these viruses could be temperate alphasphaerolipoviruses. In addition, two euryarchaeal

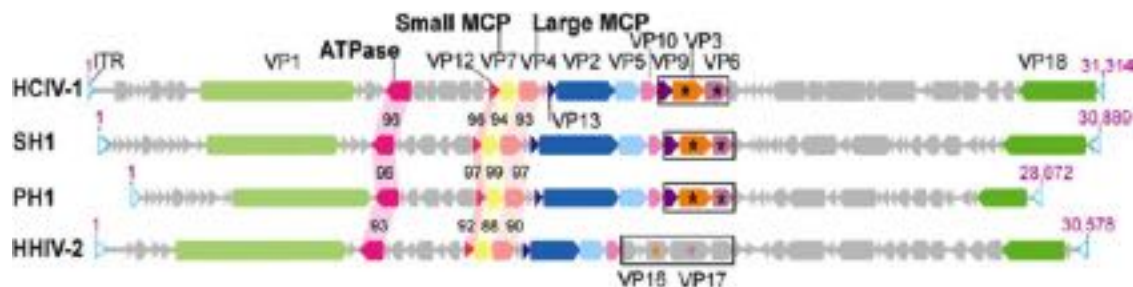


Fig. 5 Genome comparison of icosahedral membrane-containing euryarchaeal dsDNA virus isolates belonging to the family *Sphaerolipoviridae*. Gray arrows, ORFs and genes; other colored arrows, homologous genes coding for structural virion proteins (VPs) indicated above; light blue arrow heads, inverted terminal repeats (ITR). The nomenclature for the virion proteins (VPs) is same in all four viruses except that the vertex complex proteins that are marked with asterisks: HHIV-2 proteins VP16 and VP17 and proteins VP3 and VP6 in other viruses. Genes encoding putative vertex complex proteins are bordered with black lines. Pairwise amino acid similarities (%) between the major membrane proteins (VP12), major capsid proteins (MCPs), or putative packaging ATPases are shown in between the genomes. Reproduced from Demina, T.A., Pietilä, M.K., Svirskaitė, J., *et al.*, 2017. HCIV-1 and other tailless icosahedral internal membrane-containing viruses of the family *Sphaerolipoviridae*. *Viruses* 9, 32.

proviruses TKV4 and MVV (putatively forming icosahedral particles) can be found integrated into the chromosomes of *Thermococcus kodakaraensis* KOD1 and *Methanococcus voltae* A3.

Pleomorphic, Spindle-Shaped, and Spherical Viruses

Pleomorphic, spindle-shaped, and spherical viruses of euryarchaea have non-lytic life cycles. From these groups, halophilic pleomorphic viruses are the most numerous with 18 virus isolates described today (Table 3). A few spindle-shaped (i.e., lemon-shaped) viruses have been isolated on halophilic or thermophilic euryarchaea, whereas the only example of a spherical euryarchaeal virus is MetSV of anaerobic methane-producer *Methanosarcina* (Table 3).

Pleomorphic Viruses

Since the discovery of the first archaeal pleomorphic virus HRPV-1 in 2009, altogether 18 pleomorphic viruses infecting halophilic euryarchaea have been described. Fifteen of those are members of the family *Pleolipoviridae*, which is the first virus family containing viruses with either ssDNA or dsDNA genomes (Table 3). In addition, three virus isolates are tentative species of the family. Here, we briefly summarize the current knowledge on the life cycles, virions, and genomic sequences of pleolipoviruses. For more information, please see the further reading “Vesicle-Like Archaeal Viruses”.

The majority of pleolipoviruses have been isolated on halophilic archaea belonging to the genera *Halorubrum* and *Haloarcula* of the class Halobacteria. Pleolipoviruses establish non-lytic infection, presumably using budding as an exit mechanism. Virus particles are produced continuously, and the host growth is typically retarded. SNJ2 is a temperate virus, and some other viruses harbor also integrase-encoding genes in their genomes. Pleolipoviruses lack a rigid protein capsid, instead, their virions are flexible membrane vesicles (Fig. 6). Virions have only 2–4 structural protein types, of which the spike induces the fusion of viral and host membranes during virus entry.

Pleolipoviruses have either ssDNA or dsDNA circular genomes, except His2, whose genome is a linear dsDNA molecule (Table 3). All pleolipovirus genomes carry a conserved block of collinear genes. Based on whole genome information and gene content, the pleolipoviruses are classified into *Alphapleolipovirus*, *Betapleolipovirus*, and *Gammapleolipovirus* genera. In addition, proviral sequences related to pleolipoviruses are commonly found in genomes of halophilic archaea, indicating that they are quite common in hypersaline environments.

Spindle-Shaped Viruses

Spindle-shaped viruses are so far found only in archaeal viruses. There are several crenarchaeal spindle-shaped viruses belonging to the family *Fuselloviridae*, but His1 is the only halophilic euryarchaeal spindle-shaped virus and it infects *Haloarcula hispanica*. His1 is taxonomically assigned into the genus *Salterprovirus* in the newly proposed family *Halspiviridae*. The linear dsDNA genome of His1 encodes a putative type-B DNA polymerase for protein-primed replication. The spindle-shaped His1 virion with a short tail is elastic and very stable in different conditions (Fig. 6). When inactivated (e.g., by treatment with non-ionic detergent or boiling), the spindle-shaped particle releases its genome and transforms into an empty tube. Similar mechanism might be used when the viral genome is injected upon host infection. His1 particles vary in size and are composed of only one MCP type and a few minor structural protein species. His1 MCP exists in two forms, one of which is modified with lipids. No lipid bilayer has been detected in His1 virion. Virion elasticity might be a consequence of the MCP lipid-modification. His1 life cycle is non-lytic and the infection leads to continuous virus production.

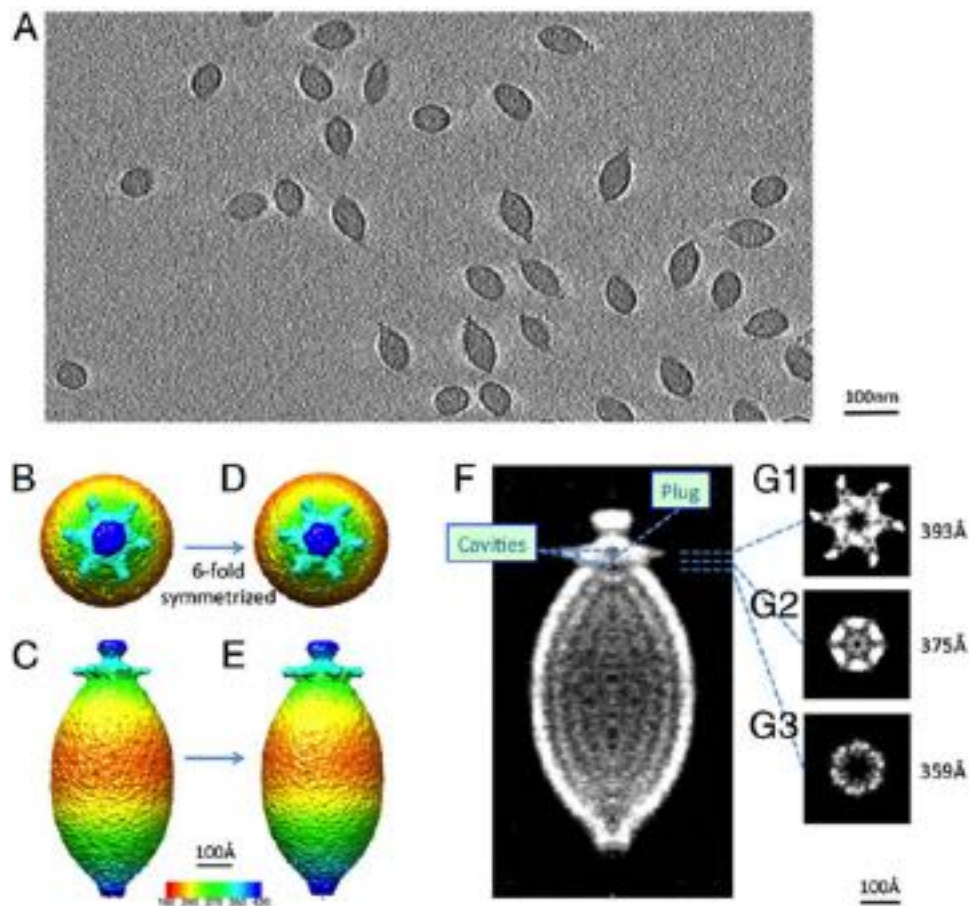


Fig. 6 Spindle-shaped virus His1 of the family *Halspiviridae*. (A) A slice through a tomogram of His1 viruses. (B–E) Structure of His1 virion is shown (B and C top and side views of the symmetry-free map; D and E top and side views of the sixfold symmetrized map). (F) The central slice of the refined sixfold average map showing the two cavities and a plug density. (G1–G3) Slices normal to the long axis of the virion at the indicated positions in F. Reproduced from Hong, C., Pietilä, M.K., Fu, C.J., *et al.*, 2015. Lemon-shaped haloarchaeal virus His1 with uniform tail but variable capsid structure. *Proceedings of the National Academy of Sciences of the United States of America* 112, 2449–2454.

TPV1 is the only described spindle-shaped virus isolate of hyperthermophilic euryarchaeon (**Table 3**). Its circular dsDNA genome can be found freely in high copy numbers in the host cell *Thermococcus prieurii*, from which it is produced continuously in a non-lytic fashion. VLP PAV1 is also spindle-shaped with a short tail but isolated from hyperthermophilic euryarchaeon *Pyrococcus abyssi*, from which it is continuously released (**Table 3**). However, its infectivity has not been demonstrated. PAV1 circular genome exists also as an episome in the *Pyrococcus abyssi* cytoplasm. Half of the PAV1 genome is composed of genes that have homologs in pTN2-like plasmids of *Thermococcus* strains. PAV1 persists most probably in its host in a stable carrier state. In addition, a spindle-shaped A3-VLP has been observed in a methanogenic euryarchaeon *Methanococcus voltae*. The major capsid proteins of the spindle-shaped viruses with a short tail His1 (type virus of *Halspiviridae*), PAV1, TPV1, A3-VLP, and SSV1 (type virus of *Fuselloviridae*) are homologous based on their sequence similarity, suggesting their close evolutionary relationship.

Spherical Virus Metsv

The only euryarchaeal virus with spherical morphology is MetSV, which has been isolated from an anaerobic sewage sludge on *Methanosarcina mazei* Gö1 (**Fig. 1(G)**, **Table 3**). Spherical particles are of ~57 nm in diameter and contain a blackberry-like envelope. Albeit having smaller size and no helical nucleoparticles, MetSV particles generally resemble PSV, the previously reported spherical virus infecting thermophilic crenarchaea of the genera *Pyrobaculum* or *Thermoproteus*. The analysis of viral and host lipids has revealed that virus particles contain an internal membrane composed of lipids obtained from the host. The presence of direct terminal repeats at the ends of the linear genome and a gene encoding a type B DNA polymerase suggest that MetSV is replicated via a protein-primed mechanism. The majority of MetSV ORFs could not be assigned with putative functions. MetSV is virulent, and its host range includes a few *Methanosarcina mazei* strains and *M. barkeri* strain DSM 1311 growing as single cells. However, sarcina-like aggregates are resistant to MetSV infection. MetSV-derived spacers have been identified (with

some mismatches) in Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR) loci in several *Methanosarcina* sp. genomes, but not in the type strain *M. mazei* DSM 3647.

Culture-Independent Studies

Microscopic examinations of hypersaline environments have revealed high VLP concentrations, typically up to 10^9 VLP/mL, and varying VLP/cell ratios (1 – 100). A variety of VLP shapes was observed in hypersaline environments with spindle-shaped VLPs being the most abundant ones, especially in the saltiest waters. Other observed morphotypes include spherical, filamentous, tailed icosahedral and some more unusual shapes. For example, star-shaped VLPs have been detected in the Dead Sea and a lot of unusual particle morphologies in hypersaline Lake Retba, e.g., hairpin-shaped and bacilliform particles, branched filaments, chains of small globules and others. It has to be yet noted that in microscopy, it is impossible to distinguish between bacterial and archaeal viruses, and e.g., pleomorphic VLPs can be confused with membrane vesicles commonly produced by cells.

Square archaeon *Haloquadratum walsbyi* is very abundant in hypersaline environments, but no viruses have been yet isolated for it. However, EM examination of environmental samples has revealed icosahedral tailed and spindle-shaped VLPs associated with square cells of *Haloquadratum* sp. In addition, pulse-field gel electrophoresis of environmental DNA from Santa Pola solar salterns, Alicante, Spain, with subsequent fosmid and plasmid clones construction and sequencing allowed a reconstruction of a near-complete genome sequence of EHP-1 (environmental halophage 1) presumably infecting *Hqt. walsbyi*. Among the other 42 partial viral genome sequences obtained from the same salterns, half could be tentatively assigned to *Hqt. walsbyi*, *Nanohaloarchaea*, or *Hrr. lacusprofundi*, indicating potential virus-host pairs. In addition, single-cell genomics and microarrays have been used for the identification of virus-host pairs in environmental samples of Santa Pola salterns, with NHV-1 (nanohaloarchaeal virus 1) and its nanohaloarchaeal host as the case study.

Metagenomic and metaproteomic analyses of the Deep Lake in Antarctica revealed that the lake is dominated by haloarchaea. Identified viral MCP sequences matched several haloarchaeal virus isolates: HCTV-1, HVTV-1, HCTV-2, HHTV-1, and CGphi46. The matches to HCTV-1-like prohead protease and HRTV-7-like scaffold protein sequences indicate active virus life cycles in the sampled community. The analysis of Antarctic salt lakes metagenomes has also revealed matches between archaeal CRISPR spacers and tailed halovirus isolates ChaoS9, phiH1, and phiCh1.

Virus-host interactions in hypersaline environments still remain largely unresolved. It has been shown that salt concentration changes, which are inevitable in natural environments, may affect adsorption rates, infectivity, and life cycle of haloarchaeal viruses. Virus-host systems in hypersaline environments seem to be diverse and dynamic. Generally, common features are found in meta-viromes from distant hypersaline environments, but finer-level spatial and temporal analyses often suggest controversial trends in viral dynamics. When compared to sequence databases, metagenomic datasets from hypersaline environments usually have a very limited number of hits. Virus defense systems such as CRISPR-Cas, commonly found in archaea, suggest a constant active interplay between viruses and their hosts.

Euryarchaeal viruses have been detected even deep in the ocean basement by metagenomics. A complete genome sequence (JdFR1000234; 55,906 bp) of a putative archaeal virus obtained from geothermally heated fluid samples collected from the ocean basement (117–292 m depth) of the Juan de Fuca Ridge, North America, encodes putative viral MCP, portal protein, and terminase, displaying similarities to archaeal icosahedral viruses with a long contractile tail. In addition, the complete viral genome of ANM-1, which presumably infects an anaerobic methane-oxidizing archaeon, was assembled from subsurface sediments collected from a methane seep (820 m depth) in Santa Monica Basin. ANM-1-related metagenomic sequences found from other methane seeps suggest that these viruses have a wide distribution in such environments. Interestingly, ANM-1 sequence contains diversity-generating retroelements, which together with similar elements found in the genomes of subterranean nanoarchaea represent the first examples of this genetic diversification-driving mechanism in archaeal systems. A variety of VLPs resembling archaeal viruses have been found in deep anoxic sediments of the Lake Pavin, France. Pleomorphic ellipsoid VLPs have been observed associated with filamentous cells, presumably belonging to methanogenic archaea of the genus *Methanosaeta*.

Recently, tens of genomes of viruses putatively infecting marine unculturable euryarchaea of Group II have been assembled from environmental viromes (the Red sea, Tara Oceans, and Osaka Bay viromes). These viruses, called magroviruses, possess dsDNA genomes of up to ~100 kbp, displaying similarities to euryarchaeal tailed viruses. Interestingly, magrovirus genomes appear to encode an extensive set of proteins involved in DNA replication, i.e., a nearly complete replication apparatus of the archaeal type. Marine euryarchaea and magroviruses seem to be globally widespread and abundant, although no such virus-host systems have been yet isolated.

Conclusions

The diversity of euryarchaeal viruses is still undersampled compared to that of viruses infecting bacteria or eukaryotes, limiting our understanding on their evolution and environmental impacts. The existing knowledge suggests that euryarchaeal viruses represent a very diverse group of viruses with various morphologies and life strategies. With detailed molecular studies, unexpected evolutionary links have been revealed between archaeal, bacterial or eukaryotic viruses, while environmental studies have shown that euryarchaea and their viruses are abundant and ecologically relevant in various environments, both moderate and extreme.

Possible evolutionary relations have been also suggested for membrane-enclosed pleomorphic viruses and membrane vesicles, commonly produced by cells. Much more research is needed to understand the relations within and beyond the group of euryarchaeal viruses, their interactions with host cells and the extent of their ecological roles.

Further Reading

- Atanasova, N.S., Bamford, D.H., Oksanen, H.M., 2015. Haloarchaeal virus morphotypes. *Biochimie* 118, 333–343.
- Atanasova, N.S., Bamford, D.H., Oksanen, H.M., 2016. Virus-host interplay in high salt environments. *Environmental Microbiology Reports* 8, 431–444.
- Atanasova, N.S., Demina, T.A., Buivydas, A., Bamford, D.H., Oksanen, H.M., 2015. Archaeal viruses multiply: Temporal screening in a solar system. *Viruses* 7, 1902–1926.
- Atanasova, N.S., Roine, E., Oren, A., Bamford, D.H., Oksanen, H.M., 2012. Global networks of specific virus-host interactions in hypersaline environments. *Environmental Microbiology* 14, 426–440.
- Bamford, D.H., Pietilä, M.K., Roine, E., *et al.*, 2017. ICTV virus taxonomy profile: *Pleolipoviridae*. *Journal of General Virology* 98, 2916–2917. (ICTV Report Consortium).
- De Colibus, L., Roine, E., Walter, T.S., *et al.*, 2019. Assembly of complex viruses exemplified by a halophilic euryarchaeal virus. *Nature Communications* 10, 1456.
- Demina, T.A., Pietilä, M., Svirskaitė, J., *et al.*, 2017. HCIV-1 and other tailless icosahedral internal membrane-containing viruses of the family *Sphaerolipoviridae*. *Viruses* 9, 32.
- El Omari, K., Li, S., Kotecha, A., *et al.*, 2019. The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. *Nature Communications* 10, 846.
- Hong, C., Pietilä, M.K., Fu, C.J., *et al.*, 2015. Lemon-shaped haloarchaeal virus His1 with uniform tail but variable capsid structure. *Proceedings of the National Academy of Sciences of the United States of America* 112, 2449–2454.
- Pietilä, M.K., Atanasova, N.S., Oksanen, H.M., Bamford, D.H., 2013. Modified coat protein forms the flexible spindle-shaped virion of haloarchaeal virus His1. *Environmental Microbiology* 15, 1674–1686.
- Pietilä, M.K., Laurinmäki, P., Russell, D.A., *et al.*, 2013. Structure of the archaeal head-tailed virus HSTV-1 completes the HK97 fold story. *Proceedings of the National Academy of Sciences of the United States of America* 110, 10604–10609.
- Pietilä, M.K., Roine, E., Sencilo, A., Bamford, D.H., Oksanen, H.M., 2016. *Pleolipoviridae*, a newly proposed family comprising archaeal pleomorphic viruses with single-stranded or double-stranded DNA genomes. *Archives of Virology* 161, 249–256.
- Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature Reviews Microbiology* 15, 724–739.
- Senčilo, A., Jacobs-Sera, D., Russell, D.A., *et al.*, 2013. Snapshot of haloarchaeal tailed virus genomes. *RNA Biology* 10, 803–816.
- Senčilo, A., Roine, E., 2014. A glimpse of the genomic diversity of haloarchaeal tailed viruses. *Frontiers in Microbiology* 5, 84.
- Weidenbach, K., Nickel, L., Neve, H., *et al.*, 2017. Methanosarcina spherical virus, a novel archaeal lytic virus targeting *Methanosarcina* strains. *Journal of Virology* 91. (e00955-17) Chapter "Vesicle-Like Archaeal Viruses".

Relevant Websites

- <https://talk.ictvonline.org/>
International Committee on Taxonomy of Viruses (ICTV).
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-dsdna-viruses/w/pleolipoviridae
The Family *Pleolipoviridae*.

Vesicle-Like Archaeal Viruses

Elina Roine, University of Helsinki, Helsinki, Finland

Nina S Atanasova, Finnish Meteorological Institute, Helsinki, Finland and University of Helsinki, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Å Angstrom

CM Cytoplasmic membrane

HAPV-2 *Haloarcula* pleomorphic virus 2

HGPV-1 *Halogeometricum* pleomorphic virus 1

HHPV-1 *Haloarcula hispanica* pleomorphic virus 1

HHPV-2 *Haloarcula hispanica* pleomorphic virus 2

HHPV3 *Haloarcula hispanica* pleomorphic virus 3

HHPV4 *Haloarcula hispanica* pleomorphic virus 4

His2 His2 virus

HRPV *Haloarchaeal* pleomorphic virus

HRPV-1 *Halorubrum* pleomorphic virus 1

HRPV-2 *Halorubrum* pleomorphic virus 2

HRPV-3 *Halorubrum* pleomorphic virus 3

HRPV-6 *Halorubrum* pleomorphic virus 6

HRPV-7 *Halorubrum* pleomorphic virus 7

HRPV-8 *Halorubrum* pleomorphic virus 8

HRPV-9 *Halorubrum* pleomorphic virus 9

ICTV International Committee on Taxonomy of Viruses

kb Kilobase

kDa Kilodalton

nt Nucleotide

NTPase Nucleoside-triphosphatase

ORF Open reading frame

pHK2 *Haloarchaeal* plasmid pHK2

RCR Rolling circle replication

SNJ1 *Natrinema* virus 1

SNJ2 *Natrinema* virus 2

TM Transmembrane domain

tRNA^{Arg} Transfer RNA specific for Arginine

tRNA^{Met} Transfer RNA specific for Methionine

VP Virus protein

wHTH Winged helix-turn-helix

Glossary

Alphapleolipovirus Pleolipoviruses with circular ssDNA genomes encoding putative rolling circle replication initiation proteins.

Archaeal virus Virus infecting organisms in the third domain of life, the *Archaea*.

Betapleolipovirus Pleolipoviruses with circular dsDNA genomes with short single stranded discontinuities.

Gammapleolipovirus Pleolipoviruses with linear dsDNA genomes with inverted terminal repeat sequences and a gene for putative type B DNA polymerase.

Glycosylation Addition of carbohydrate(s), i.e., glycan to another molecule. In this article, glycosylation is limited to proteins as target molecules.

Haloarchaea Extremophilic archaea that require NaCl concentration above 1.5 M for survival.

Integrase An enzyme that catalyzes the integration of viral genomic DNA into the genome of the host cell.

N-Glycosylation A mode of glycosylation where an asparagine (Asn/N) of a protein is modified with an addition of a glycan.

PhiH repressor Repressor protein whose gene was originally identified in the archaeal virus PhiH of

Halobacterium salinarum, with a function to repress the production of an early lytic transcript. PhiH-like repressor genes have been identified in the genomes of those pleolipoviruses that contain a putative integrase gene.

Pleolipovirus Archaeal virus with a vesicle-like virion architecture.

Receptor A molecule, usually exposed on the surface of the host cell, that the virus uses to recognize the host cell.

S-Layer protein (archaeal) Archaeal cell wall glycoproteins, which can self-assemble into a paracrystalline surface layer. S-layer proteins are candidates of pleolipovirus receptors.

Spike Structural virion surface protein, which usually protrudes from the surface of the viral particle and serves as the receptor recognition protein. In pleolipoviruses it also serves as a membrane fusion protein.

VP3-Like protein Pleolipovirus virion membrane associated protein, which resides for the most part, inside the virion. It is a homolog of HRPV-1 VP3 protein.

VP4-Like protein Pleolipovirus virion spike protein, which serves as the receptor recognition protein and fusion protein. It is a homolog of HRPV-1 VP4 protein.

Introduction

Haloarchaeal pleomorphic viruses (HRPVs) belong to a newly established family of *Pleolipoviridae*. The first isolate and now the typical member of the family, *Halorubrum* pleomorphic virus 1 (HRPV-1), was reported in 2009. It was also the first archaeal virus containing a single-stranded DNA (ssDNA) genome. Ever since, new representatives have been reported adding up to almost 20 different isolates that, after tailed viruses, represent the second largest group of viruses infecting haloarchaea (Table 1). Apart from representing the first archaeal viruses containing ssDNA genomes, the family *Pleolipoviridae* is also the first family, officially

Table 1 Reported Haloarchaeal pleomorphic viruses

<i>Virus</i>	<i>Genome type (c/l)^a</i>	<i>Genome size (nt/bp)</i>	<i>Capsid diameter (nm)</i>	<i>Host</i>
HRPV-1	ssDNA, c	7048	41	<i>Halorubrum</i> sp. PV6
HRPV-2	ssDNA, c	10,656	54	<i>Halorubrum</i> sp. SS5–4
HRPV-3	discontinuous dsDNA, c	8770	67	<i>Halorubrum</i> sp. SP3–3
HRPV-6	ssDNA, c	8549	49	<i>Halorubrum</i> sp. SS7–4
HRPV9	discontinuous dsDNA, c	16,159	~57	<i>Halorubrum</i> sp. SS7–4
His2	dsDNA, l	16,067	71	<i>Haloarcula hispanica</i>
SNJ2	dsDNA, c	16,992	~70–80	<i>Natrinema</i> sp. J7–1
HHPV-1	dsDNA, c	8082	52	<i>Haloarcula hispanica</i>
HHPV-2	dsDNA, c	8176	~50	<i>Haloarcula hispanica</i>
HHPV3	discontinuous dsDNA, c	11,648	~50	<i>Haloarcula hispanica</i>
HHPV4	discontinuous dsDNA, c	15,010	~60	<i>Haloarcula hispanica</i>
HGPV-1	discontinuous dsDNA, c	9694	56	<i>Haloquadratum</i> sp. CG–9
HRPV-7	ND ^b	ND	~43	<i>Halorubrum</i> sp. SS5–4
HRPV-8	ND	ND	~51	<i>Halorubrum</i> sp. SP3–3
HAPV-2	ND	ND	~40	<i>Haloarcula</i> sp. SS13–14
HRPV10	discontinuous dsDNA, c	9296	~55	<i>Halorubrum</i> sp. LR2–17
HRPV11	discontinuous dsDNA, c	9368	~55	<i>Halorubrum</i> sp. LR2–12
HRPV12	discontinuous dsDNA, c	9944	~55	<i>Halorubrum</i> sp. LR1–23

^ac, circular; l, linear conformation.^bND, not determined.

classified by the International Committee on Taxonomy of Viruses (ICTV), which contains viral species harboring either single stranded or double stranded DNA (dsDNA) genomes. Despite the increasing number of new isolates and their genomic descriptions, we still have very little information about the detailed mechanisms of different steps in the viral life cycle. An exception to this is the recent report on the atomic structure of the HRPV spike protein, a membrane fusion protein, and its low-resolution structure on the virion (see *Structural proteins*).

Since viruses are obligate parasites, information on their host organisms is also crucial for understanding the interaction. One of the particularly important host factors for viral infections is receptors. According to the current understanding, the cell wall of most of the haloarchaeal strains consists of the cytoplasmic membrane (CM) covered by S-layer proteins that form a para-crystalline layer on the cell surface and a pseudo-periplasmic space between the CM and the surface of the S-layer. Not only S-layer proteins can be highly variable and thus represent the most probable candidate for the receptor, but they are also glycosylated. The haloarchaeal S-layer proteins are N-glycosylated, and as the types of glycosylation in different haloarchaeal hosts can be highly diverse, the glycan structures are also good candidates as receptors or as parts of them.

An Overview of the Virion Structure and Viral Life Cycle

The basic structure of the HRPV virions is a simple membrane vesicle enclosing the naked genome. A typical virion consists of two major structural proteins both of which are associated with the viral membrane. The larger structural protein, VP4 and VP4-like proteins (approx. from 50 to 60 kDa), forms spike structures on the surface of the virion whereas the smaller protein, VP3 or VP3-like proteins (approx. from 9–15 kDa), localizes mostly to the inner surface of the viral membrane (**Fig. 1**) without significant interactions with the genome.

The host range of the known HRPVs seems to be very narrow with only few, if any, additional strains identified in addition to the original isolation host. The life cycle of the pleomorphic viruses starts with the recognition of the host cell receptor by the spike proteins and the release of viral genome into the host cell cytoplasm in a membrane fusion event. Genome replication and transcription take place in the cytoplasm, most probably using at least two different mechanisms (see Family *Pleolipoviridae*: related viruses with different genome types) and the structural proteins together with the replicated genomes are translocated to the cytoplasmic membrane for assembly into the virions. Exit of virions occurs by budding. In that process, virions acquire their membrane from the host CM. Thus, lipid composition of the viral membrane is approximately the same as the one of the host CM. Infected cells produce new progeny viruses for several hours after the first new virions are detected.

The infection cycle for most of the characterized viruses is persistent, but detailed dynamics of different HRPV life cycles can vary substantially. For instance, HRPV-1 infects only approximately 10% of the host population at the start of the infection whereas HRPV-6 is able to infect at least 80% of the cells within the first 30 min. The duration of a life cycle of different HRPVs are relatively fast varying between approximately 2–5 h. Production of viruses by the cells retards the growth of the culture and leads in some instances to an inability to form colonies on rich culture media. The first reported temperate pleolipovirus, SNJ2, is able to integrate into the tRNA^{Met} gene of the host, *Natrinema* sp. J7–1. Interestingly, SNJ2 seems to have some type of a synergistic relationship with an icosahedral virus, SNJ1, as the presence of SNJ1 increases the titer of SNJ2 during co-infection.

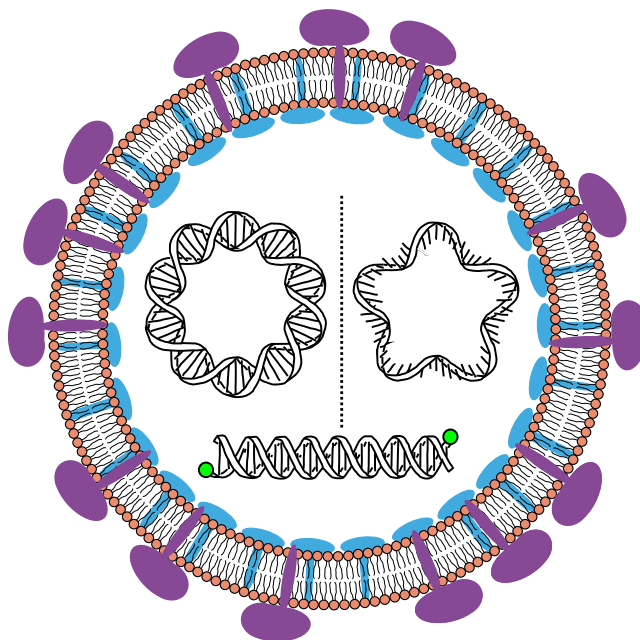


Fig. 1 Schematic representation of pleolipovirus virion structure. Spike proteins are indicated with purple color, internal membrane proteins with turquoise, lipid bilayer with orange and DNA with black and DNA-binding terminal proteins in light green.

Genomic Characteristics

Gene Content and Conserved Genes

The genomes of the first reported HRPVs, *Halorubrum* pleomorphic virus 1 (HRPV-1) and *Haloarcula hispanica* pleomorphic virus 1 (HHPV-1), were circular molecules containing less than 10 predicted open reading frames (ORFs). The identity of the genes encoding the two major structural proteins, VP3 (14 kDa) and VP4 (53 kDa), was verified using N-terminal sequencing. In addition, thorough mass spectrometry analysis of HRPV-1 particles revealed one peptide belonging to gene 8 encoding a putative nucleoside-triphosphatase (NTPase) and possibly being crucial to the particle formation. Comparative genomics, using the two HRPVs mentioned above as well as previously reported haloarchaeal plasmid pHK2, revealed a conserved cluster of the two major structural protein encoding genes as well as ORFs 2, 5 and 6. In addition, all these genetic elements contained an ORF encoding putative rolling circle replication (RCR) protein that was not necessarily conserved in sequence, but in the predicted function.

As new isolates of pleomorphic viruses, *Halorubrum* pleomorphic viruses 2 and 6 (HRPV-2 and HRPV-6) and especially *Halorubrum* pleomorphic virus 3 (HRPV-3) and *Haloarcula hispanica* pleomorphic virus 1 (HHPV-1) were reported, the number of conserved genes and predicted, conserved ORFs was reduced to five. Thus, in the HRPV-1 genome, the conserved gene cluster starts from the most conserved (in sequence) gene 3 and ends after the putative dNTPase encoded by gene 8 (Fig. 2(a)). The conserved genes of HRPV-3 and HGPV-1 show lower sequence similarity with the ones of the other viruses, and they do not contain a homolog of the RCR, but a conserved gene containing predicted winged helix-turn-helix (wHTH) domain. The gene content outside the conserved gene region also varies from genome to genome.

His2 was the seventh virus originally included in the proposal for a new family of viruses, the *Pleolipoviridae*. It was isolated by the laboratory of Mike Dyll-Smith from an Australian salt lake on a lawn of *Haloarcula hispanica*. Unlike all the other genomes characterized to date, the genome of His2 was shown to be a linear dsDNA molecule with yet another suggested replication mechanism, the so-called protein primed replication with proteins covalently attached to the 5' ends of the genome. The genome of His2 is the most divergent from the other pleomorphic virus genomes in that it contains two clearly different parts, one of which contains the gene region conserved in the pleomorphic viruses. The conserved gene region of His2 genome resembles the most those of HRPV-3 and HGPV-1.

Family Pleolipoviridae: Related Viruses With Different Genome Types

The initial proposal of the new family *Pleolipoviridae* included eight viral isolates HRPV-1 (the typical member), HRPV-2, HRPV-3, HRPV-6, HHPV-1, HHPV-2, HGPV-1 and His2 (Fig. 2(a)). As outlined above, they all contain the conserved genomic region encoding the HRPV-1 VP3 and VP4 homologs, the two major structural proteins of the virion. In the other viral isolates, these proteins can be encoded by genes assigned with a different number, which depend on the genomic content. For example, the bigger major structural protein of HRPV-2 is encoded by gene 5. In that case, the protein is VP5 and it is a VP4-like protein.

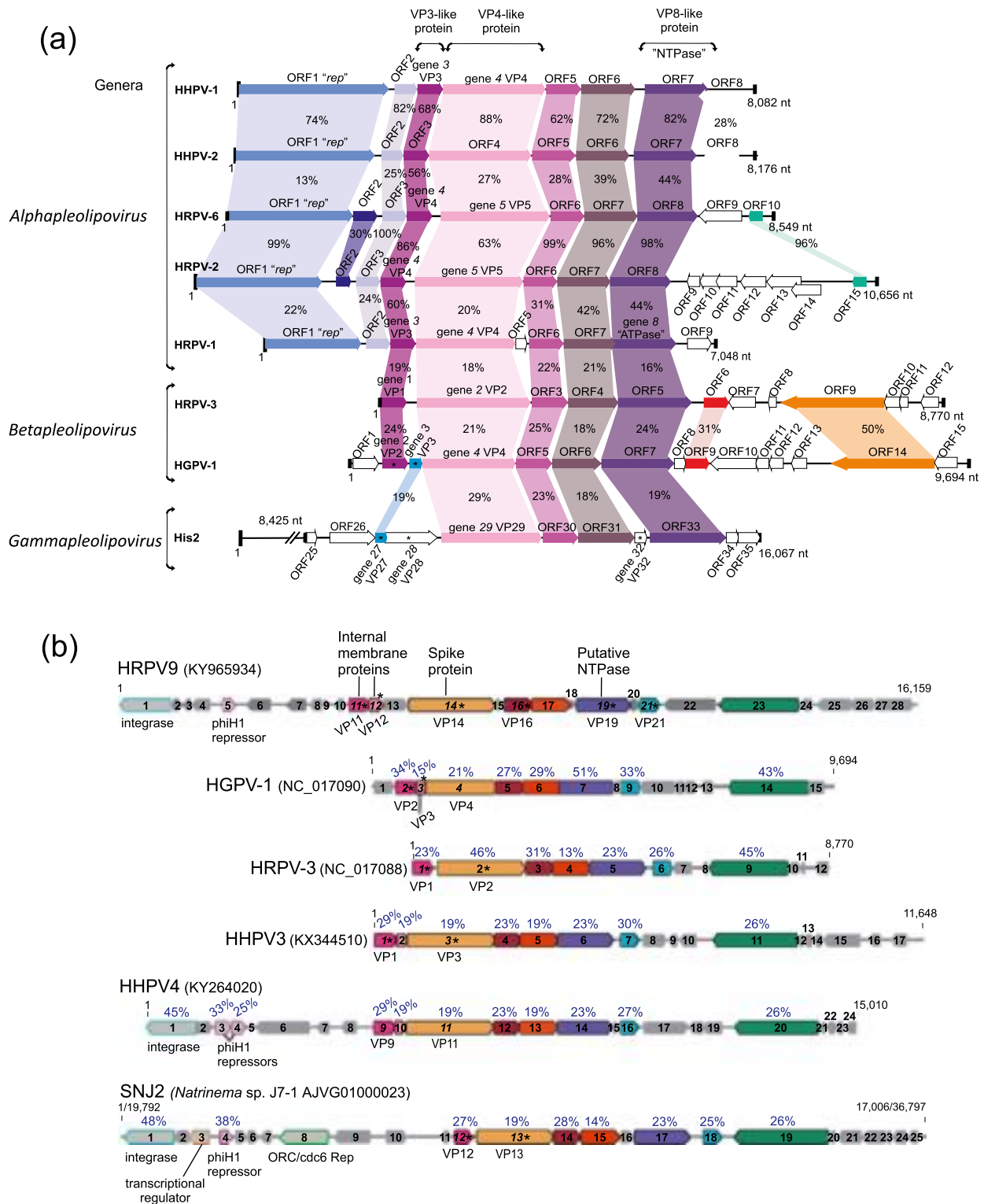


Fig. 2 Genomic comparison of pleolipovirus genomes. (a) Different representative viruses of the Pleolipoviridae family. (b) Betapleolipoviruses. Reproduced from Bamford, D.H., Pietilä, M.K., Roine, E., *et al.*, 2017. ICTV virus taxonomy profile: *Pleolipoviridae*. Journal of General Virology 98, 2916–2917. doi:[10.1099/jgv.0.000972](https://doi.org/10.1099/jgv.0.000972) (under the licence CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/>). Atanasova, N.S., Demina, T.A., Krishnam Rajan Shanthi, S.N.V., Oksanen, H.M., and Bamford, D.H., 2018b. Extremely halophilic pleomorphic archaeal virus HRPV9 extends the diversity of pleolipoviruses with integrases. Research in Microbiology 169, 500–504. doi:[10.1016/j.resmic.2018.04.004](https://doi.org/10.1016/j.resmic.2018.04.004) (under the licence CC BY, <https://creativecommons.org/licenses/by/4.0/>).

Downstream from the genes encoding the two major structural proteins, there are two conserved open reading frames encoding proteins with as yet unassigned functions, as well as the putative NTPase. The assignment of the viral isolates into different genera i.e., *Alpha*-, *Beta*- and *Gammapleolipovirus* is based on the gene content outside the conserved gene region. HHPV-1, HHPV-2, HRPV-1, HRPV-2 and HRPV-6 isolates, which carry the ORF encoding the putative RCR, are included in the *Alphapleolipovirus* genus. The assignment of HRPV-3 and HGPV-1 to the *Betapleolipovirus* genus is due to the presence of conserved hypothetical gene containing the wHTH domain. As mentioned above, the genome of His2 is different in terms of the other gene content and is assigned to the genus *Gammapleolipovirus*.

It is notable that despite of the relatedness of these viral isolates their genome type can be different. In general, the genomes of *Alphapleolipoviridae* members contain ssDNA genomes whereas the genomes of *Betapleolipoviridae* isolates contain dsDNA genomes. An intriguing feature of the genomes belonging to the latter genus is that the otherwise dsDNA genomes contain short genomic regions of ssDNA. In HRPV-3, where these discontinuities were reported for the first time, the ssDNA regions were found to be associated with a specific penta-nucleotide motif GCCCA (5' → 3') which precedes the ssDNA region of the same strand. Later on, similar discontinuous genomes have been reported for pleolipoviruses HHPV3, HHPV4 and HRPV9. Although the biological meaning of these short stretches is unknown, it has been hypothesized that, upon infection, that they would have a role in attracting the host DNA repair system components for active replication. Thus, despite the lack of information on the detailed mechanisms of genome replication, it is evident that these mechanisms are different for isolates belonging to the different genera.

New Isolates and Haloarchaeal Genomic Regions

Analyzes of the HRPV genomes showed that homologs of the *Alphapleolipovirus* genomes were present in previously reported, partially sequenced plasmid pHK2 of *Haloferax* sp. Sequencing of the entire 10.8 kb pHK2-plasmid showed that this virus-related genetic element, in addition to the conserved genomic region similar to the pleolipovirus genome, also contains an open reading frame encoding a putative integrase as well as a putative homolog of the phiH repressor gene. A homologous 12.5 kb genomic region in *Haloferax volcanii* genome spanning from a tRNA^{Arg} gene close to HVO_1434 all the way to HVO_1422 was also identified suggesting that the smaller viral genomes originated from bigger genomes of temperate viruses with a typical propensity to integrate into genomic tRNA genes. In fact, in haloarchaeal genomes there seems to be a wealth of integrated proviruses or proviral remnants related to pleolipoviruses, and especially to the viruses belonging to the *Betapleolipovirus* genus. Indeed, in 2014, a pleolipovirus genome, SNJ2, of the betapleolipovirus type was, for the first time, reported to contain both the integrase and the phiH repressor encoding genes. Also, HHPV4 and HRPV9 genomes have recently been shown to harbor both the integrase and the phiH repressor (Fig. 2(b)), and homologs of the SNJ2-type of integrases have been identified in numerous mobile genetic elements found in haloarchaeal genomes.

Stability of Pleolipovirus Infectivity

Salinity

Haloarchaeal organisms thrive in hypersaline conditions with, for instance, saturated NaCl concentrations (approximately 5.5 M). Thus, the most important environmental condition affecting the infectivity of pleolipoviruses is the high NaCl concentration. Most pleolipoviruses lose infectivity at NaCl concentrations below 1 M. Different pleolipoviruses vary slightly in their NaCl requirements. HRPV-1 is stable in a narrow range of 1–3 M NaCl, while betapleolipoviruses HRPV-3, HRPV-9, and HGPV-1 were found to be stable in NaCl concentrations varying between 0 M to above 5 M (saturation), during the 24 h incubation time, while HHPV3 and HHPV4, require at least 2.5 M NaCl. A peculiar and apparently pH-dependent infectivity at changing NaCl concentration was observed for pleolipoviruses HHPV3 and HHPV4. HHPV3 was able to maintain the infectivity at pH values 4.5–9 when NaCl concentration was 3.5 M. However, when NaCl was lowered to 1.5 M, the virus lost all infectivity at the optimum pH 7.5. At pH 9, on the other hand, the infectivity was maintained high even at 1.5 M NaCl. Similar pH-dependence was observed for HHPV4, which, genetically, is almost identical to HHPV3. The loss of infectivity was irreversible for both viruses. When HHPV-1 adsorption was studied at different salinities, maximum adsorption was detected when NaCl molarity was as high as 3.5 M. In addition to NaCl, pleolipoviruses HHPV3 and HHPV4 require at least 4 mM CaCl₂ concentration for stability and small amounts of CaCl₂ are used standardly for the other viruses as well.

Temperature and Other Significant Factors

Virus stock preparations of all pleolipoviruses are stable for months when stored at 4°C. The natural environments of halophilic archaea are mostly located in tropical or subtropical areas and exposed to intense sunlight. Thus, it is logical that their viruses require temperatures above 30°C for optimal infection. The most commonly used temperature for studying pleolipoviruses is 37°C. Incubation of the virions at temperatures higher than 40°C abolishes infection. Recently, it was shown that in the case of the spike proteins of HRPV-6, heat treatment at 55°C induces a drastic conformational change leading to an extended conformation of the spike proteins (see also: Structural proteins).

In general, pleolipovirus infectivity is lost in detergents and organic solvents as well as negative stains used in transmission electron microscopy. Some viruses, such as HRPV-6, HRPV-8, HRPV9, and HRPV10 stand out as being more resistant to chloroform (in the test conditions used) than other pleolipoviruses. The reasons for this stability of a lipid-membrane containing particle to chloroform are not known. The effect of negative stains on the pleolipovirus infectivity has been studied to answer the question whether the virion morphology observed in the ordinary negatively stained samples represent the one of infective virions. Most of the negative stains abolish the infectivity, but no clear correlation with the virion morphology has been found. Thus, ordinary negative stain transmission microscopy cannot be used for determining the morphology of an infectious pleolipovirus particle.

Structural Proteins

As mentioned above, the basic pleolipovirus virion contains two major structural proteins, both of which are associated with the membrane envelope of the virion. In addition, traces of the conserved ORF encoding the putative NTPase can be found. However, some of the isolates belonging to the *Beta*- and *Gammapleolipovirus* genera, seem to contain some additional structural proteins. For instance, in His2 virions there are two protein species making the spikes as well as two protein species serving the role of the smaller membrane protein. Quantitative biochemical dissociation studies for the smaller membrane protein have shown that it is membrane associated and mostly exposed to the lumen of the virion. However, it does not seem to have strong interaction with the genome.

The Spike Protein

The spike protein has been studied more intensively due to its role in host recognition and membrane fusion. All the spike proteins contain an N-terminal signal peptide for either Sec- or predicted Tat-secretion pathway. They also contain a C-terminal trans-membrane domain (TM) with which the spike protein stays anchored to the viral membrane. This has been shown experimentally for HRPV-1 and HRPV-6 virions. The primary amino acid sequence for the different spike proteins can be highly variable, which is common for the viral proteins involved in host recognition. Thus, as such it cannot be considered to be among the conserved proteins of the virus. However, its position as one of the conserved proteins is warranted as in all viruses it seems to serve the same function as a spike protein, and its position in the genomic context of the gene is always the same.

Different types of modifications bring additional variation to the spike protein properties. The HRPV-1 spike protein is N-glycosylated with a pentasaccharide that comprises glucose, glucuronic acid, mannose, sulphated glucuronic acid and 5-N-formyl-legionaminic acid residue as the terminal sugar. There is also evidence suggesting that this glycan structure is involved in the recognition of the host receptor. This post-translational modification must take place in the pseudo-periplasmic space as described for the S-layer proteins of haloarchaea. None of the spike proteins from other pleolipovirus species have been shown to be glycan modified. However, such a role of glycans in specific receptor recognition may also be true for other pleolipoviruses, but provided by glycan modifications of the host S-layer. As there are no genes predicted to be involved in N-glycosylation pathways in the pleolipoviral genomes, the glycosylation is dependent on the host glycosylation machinery. This was shown by heterologous expression of the HRPV-1 VP4 in *Hfx. volcanii* in which case the N-glycan structure was the same as described for *Hfx. volcanii* S-layer protein N-glycan. In addition to the reported glycan modification, one of the two spike proteins of His2 virus was predicted to contain lipid modifications since the protein was positively stained with Sudan Black. The possible biological function of this modification is not known.

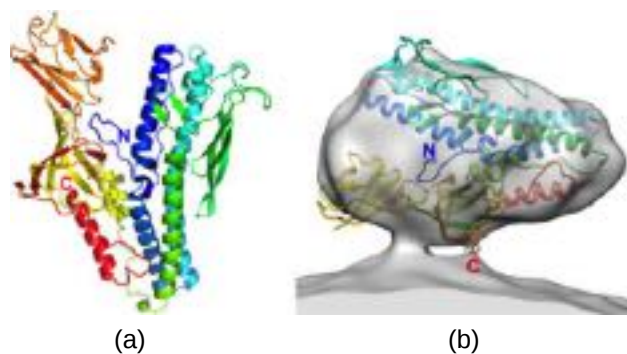


Fig. 3 Structure of the HRPV-6 spike and the spike protein VP5. (a) Cartoon representation of the spike protein colored from the N-terminal (blue) to the C-terminal (red). (b) Density of the spike on virion surface as solved by cryo-electron tomography and subtomogram averaging with HRPV-6 VP5 crystallographic structure fitted into the density. Reproduced from El Omari, K., Li, S., Kotecha, A., *et al.*, 2019. The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. *Nature Communications* 10, 846, doi:10.1038/s41467-019-08728-7, under the licence CC BY 4.0, <http://creativecommons.org/licenses/by/4.0/>.

The atomic structure of the HRPV-2 and HRPV-6 spike proteins (VP4-like protein VP5) has been determined at 2.5 and 2.7 Å resolution, respectively. Both proteins, approximately 65% identical in their primary amino acid sequence, show highly similar structures. The overall protein structure is a previously unreported V-shaped fold dividing the protein in two major domains that are connected with a single residue (Fig. 3). The lowest similarity, both in amino acid sequence and in structure, was shown to be in a domain that, when fitted to a 16 Å resolution cryo-EM structure of the spike on the HRPV-6 virion, was pointing outwards from the spike structure. Thus, this domain was suggested to be involved in host recognition.

In addition to the host recognition, these spike proteins were also shown to be involved in fusion of the viral membrane with the host cell membrane. The highly alpha-helical secondary structure of the N-terminal domain was predicted to be involved in this process. It was hypothesized that upon the interaction of the putative host recognition domain with the host S-layer protein, the N-terminal alpha helical structure will open with concomitant exposure of the fusion peptide at the very N-terminus of the protein to the host membrane.

Further Reading

- Atanasova, N.S., Heiniö, C.H., Demina, T.A., *et al.*, 2018a. The unexplored diversity of pleolipoviruses: The surprising case of two viruses with identical major structural modules. *Genes* 9, 131.
- Atanasova, N.S., Demina, T.A., Krishnam Rajan Shanthi, S.N.V., Oksanen, H.M., Bamford, D.H., 2018b. Extremely halophilic pleomorphic archaeal virus HRPV9 extends the diversity of pleolipoviruses with integrases. *Research in Microbiology* 169, 500–504.
- Bamford, D.H., Pietilä, M.K., Roine, E., *et al.*, 2017. ICTV virus taxonomy profile: *Pleolipoviridae*. *Journal of General Virology* 98, 2916–2917.
- Bath, C., Cukalac, T., Porter, K., Dyall-Smith, M.L., 2006. His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, *Salterprovirus*. *Virology* 350, 228–239.
- Demina, T.A., Atanasova, N.S., Pietilä, M.K., Oksanen, H.M., Bamford, D.H., 2016. Vesicle-like virion of *Haloarcula hispanica* pleomorphic virus 3 preserves high infectivity in saturated salt. *Virology* 499, 40–51.
- El Omari, K., Li, S., Kotecha, A., *et al.*, 2019. The structure of a prokaryotic viral envelope protein expands the landscape of membrane fusion proteins. *Nature Communications* 10, 846.
- Jarrell, K.F., Ding, Y., Meyer, B.H., *et al.*, 2014. N-linked glycosylation in *Archaea*: A structural, functional, and genetic analysis. *Microbiology and Molecular Biology Reviews* 78, 304–341.
- Kandiba, L., Aitio, O., Helin, J., *et al.*, 2012. Diversity in prokaryotic glycosylation: An archaeal-derived N-linked glycan contains legionaminic acid. *Molecular Microbiology* 84, 578–593.
- Liu, Y., Wang, J., Liu, Y., *et al.*, 2015. Identification and characterization of SNJ2, the first temperate pleolipovirus integrating into the genome of the SNJ1-lysogenic archaeal strain. *Molecular Microbiology* 98, 1002–1020.
- Mizuno, C.M., Prajapati, B., Lucas-Staat, S., *et al.*, 2019. Novel haloarchaeal viruses from Lake Retba infecting *Haloferax* and *Halorubrum* species. *Environmental Microbiology* 21, 2129–2147. <https://doi.org/10.1111/1462-2920.14604>.
- Pietilä, M.K., Atanasova, N.S., Manole, V., *et al.*, 2012. Virion architecture unifies globally distributed Pleolipoviruses infecting halophilic archaea. *Journal of Virology* 86, 5067–5079.
- Pietilä, M.K., Roine, E., Paulin, L., Kalkkinen, N., Bamford, D.H., 2009. An ssDNA virus infecting archaea: A new lineage of viruses with a membrane envelope. *Molecular Microbiology* 72, 307–319.
- Senčilo, A., Paulin, L., Kellner, S., Helm, M., Roine, E., 2012. Related haloarchaeal pleomorphic viruses contain different genome types. *Nucleic Acids Research* 40, 5523–5534.
- Wang, J., Liu, Y., Liu, Y., *et al.*, 2018. A novel family of tyrosine integrases encoded by the temperate pleolipovirus SNJ2. *Nucleic Acids Research* 46, 2521–2536.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-dsna-viruses/w/pleolipoviridae
Pleolipoviridae.

Virus–Host Interactions in Archaea

Diana P Baquero, Archaeal Virology Unit, Institut Pasteur, Paris, France and Sorbonne University, Paris, France
David Prangishvili, Institut Pasteur, Paris, France and Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia
Mart Krupovic, Archaeal Virology Unit, Institut Pasteur, Paris, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Acr Anti-CRISPR

cOAs Cyclic oligoadenylates

ESCRT Endosomal sorting complex required for transport

MCP Major capsid protein

ORF Open reading frame

S-layer Surface layer

T4P Type 4 pilus

TA Toxin-antitoxin

Glossary

Adsorption rate Rate at which extracellular virions become attached to the host.

A-form One of the three major forms of double-stranded DNA, with a 23 Å helical diameter and 11 bp per helix turn.

Capsid Protein shell that encloses the genetic material of the virus.

Episomal element Extrachromosomal (non-integrated) form of a mobile genetic element inside the host cell.

Holliday junction resolvase A specialized structure-selective endonuclease that cleaves four-way DNA intermediates formed during DNA replication.

Homologous recombination Recombination between two identical or highly similar DNA sequences.

Hyperhalophiles Organisms thriving in the presence of extremely high salt concentrations.

Hyperthermophiles Organisms requiring extremely high (80°C or above) temperatures for optimal growth.

Integrase Enzyme that catalyzes the integration of a viral genome into the genome of the host cell.

Inverted terminal repeats Short, related or identical sequences located in reverse orientation at the ends of the viral genome.

Jelly-roll fold A structural protein fold composed of eight β-strands arranged in two antiparallel four stranded β-sheets.

Lysogenic life cycle One of the two alternative life cycles of a virus, whereby the virus genome integrates into the host chromosome as a provirus or remains as a dormant episomal element.

Lytic life cycle One of the two alternative life cycles of a virus, whereby the virus replicates inside the host and lyses the infected cell releasing the progeny viruses to infect other cells.

Methanogenic Producing methane as a metabolic by-product in anoxic conditions.

Orthologous genes Homologous genes that are related by vertical descent from a common ancestor and encode proteins with the same functions in different species.

Phase variation system Reversible process that involves the ON/OFF switching of protein expression to adapt to rapidly changing environments generating phenotypical variations within the population.

Protein glycosylation Post-translational modification where a carbohydrate molecule is covalently bound to a predetermined region of a protein to form a glycoprotein.

Provirus Viral genome integrated into the host chromosome.

Pseudomurein endoisopeptidase An enzyme that cleaves pseudomurein cell-wall sacculi of methanogens.

Rolling-circle replication Model of unidirectional DNA replication that can rapidly synthesize multiple copies of circular ssDNA molecules.

S-layer A paracrystalline protein surface layer that is present in nearly all archaea described so far.

Strand-coupled genome replication Model of DNA replication that couples leading-strand and lagging-strand synthesis.

Strand-displacement Displacement of a downstream DNA strand encountered during DNA replication.

Type 4 pili Surface-exposed filaments involved in many functions of the cell, including locomotion, adhesion, microcolony formation and protein secretion.

Virion Infectious mature virus particle.

Introduction

Viruses infecting archaea constitute a distinctive part of the virosphere and exhibit diverse virion morphologies, many of which have never been observed among viruses infecting bacteria or eukaryotes. Diversity of archaeal viruses is also reflected in their genome content, with ~75% of the genes lacking detectable homologs in sequence databases. Based on their diverse virion morphologies and genomic properties, the characterized archaeal viruses are currently classified into 20 families. Although the knowledge of virus-host interactions in archaea remain highly fragmented, the increasing number of genetic tools developed for archaea and their viruses as well as application of advanced structural and functional genomics efforts have yielded valuable information on certain aspects of virus-host interactions. Some of these mechanisms are shared with bacterial and/or eukaryotic viruses, whereas others are unique to archaeal viruses.

The virus infection cycle can be subdivided into several stages. The infection starts with the recognition and binding to specific host receptors on the host cell surface which leads to the delivery of the viral genetic material into the cell interior. Following the entry, many viruses hijack the host replication, transcription, and translation machineries to produce multiple copies of the virus progeny. For some enveloped viruses, the morphogenesis and egress are concomitant, whereas in the case of non-enveloped viruses the virion assembly typically precedes the release. Thus far, two different egress strategies have been elucidated for archaeal viruses: a budding mechanism similar to that of some eukaryotic enveloped viruses and a unique lytic mechanism employed by certain crenarchaeal viruses which involves the formation of pyramidal structures on the host cell surface. Similar to bacteriophages and certain eukaryotic viruses, some archaeal viruses can undergo lysogenic life cycle, whereby the virus genome integrates into the host chromosome as a provirus or remains as an episomal element until the cell host is exposed to certain stimuli or stress conditions that induce viral replication. Archaea and their viruses have evolved diverse defense and counter-defense mechanisms, among which CRISPR-Cas system and anti-CRISPR (Acr) proteins have been the most studied.

In this article, we review the state-of-the-art on virus-host interactions in Archaea. Note that most studies have focused on just a handful of model virus-host systems, including *Sulfolobus islandicus* rod-shaped virus 2 (SIRV2; family *Rudiviridae*), *Sulfolobus* spindle-shaped virus 1 (SSV1; family *Fuselloviridae*), *Sulfolobus* turreted icosahedral virus (STIV; family *Turriviridae*) and *Sulfolobus tengchongensis* spindle-shaped virus 2 (STSV2; family *Bicaudaviridae*), all infecting hyperthermophilic archaea of the genus *Sulfolobus* (phylum Crenarchaeota).

Virus Entry

The infection cycle begins with the recognition of a suitable host through specific interactions between a receptor-binding protein exposed on the virion and a receptor located on the surface of the host cell. Once the viral protein successfully binds to the host receptor, the virus particle typically undergoes a conformational change that results in the delivery of the genetic material into the cytoplasm of the host cell. For the well-studied viruses infecting bacteria, different cell surface structures have been recognized to be targeted by viruses including pili, flagella, peptidoglycan, lipopolysaccharide and integral membrane proteins. Viral attachment and entry have been poorly studied in archaea and only recent efforts have provided the first insights into the entry mechanisms and possible receptors of hyperthermophilic and hyperhalophilic archaeal viruses.

Interaction With Cellular Appendages

Similar to bacteria, archaea possess distinct cell surface appendages, including archaeal flagella (archaella) implicated in cell motility and diverse pili involved in intercellular communication and adhesion to various organic and inorganic substrates. Electron microscopy observations showed that many filamentous viruses bind to pili. For instance, termini of the filamentous virions of members of the *Lipothrixviridae* family are decorated with diverse structures, which resemble claws (AFV1), brushes (AFV2) or mops (SIFV), and have been proposed to play a role in the viral attachment to the host cell, specifically to pilus-like appendages.

The interaction of SIRV2, the prototype virus of the family *Rudiviridae*, with cellular appendages has been studied in more detail. Both ends of the rod-shaped SIRV2 virions contain three terminal fibers which specifically interact with the pili abundantly present on the surface of *Sulfolobus islandicus* LAL14/1 cells. Notably, when the pili were detached from the cells, the virus interacted nearly exclusively with the tips of the pili, whereas in the context of metabolizing cells, the virions were observed both at the tips and on the sides of the pili, suggesting that virions move along the pili toward the cell surface. However, the energy source for this movement remains unclear. Once at the cell surface, the SIRV2 virions appear to disassemble, presumably as a consequence of the delivery of the viral DNA into the host cytoplasm. Similarly, rudivirus SSRV1 and tristromavirus PFV2 have also been shown to interact with the type IV pili (T4P) of *Saccharolobus solfataricus* and *Pyrobaculum arsenaticum*, respectively (Fig. 1(A)–(D)). Deletion of the orthologous pilin genes, *pilA1* and *pilA2*, in *S. islandicus* M.16.4 resulted in loss of T4P and resistance to rudivirus SIRV8, a close relative of SIRV2, which could no longer adsorb to the host cell. These observations further reinforce the critical role of pili during the early stages of archaeal virus infection.

A near-atomic structure of the SIRV2 receptor showed that it is a T4P. Remarkably, the pilus was extremely stable and resisted various harsh treatments, including digestion by trypsin and pepsin as well as boiling in sodium dodecyl sulfate and 5 M guanidinium-HCl. The resilience of the pilus was attributed to extensive surface glycosylation on serine and threonine residues, which are anomalously abundant in the *Sulfolobus* T4P, compared to the bacterial T4P and other *Sulfolobus* proteins. Notably, the cryo-EM structures of the T4P of the acidophilic hyperthermophile *Saccharolobus solfataricus* and neutrophilic hyperthermophile *Pyrobaculum arsenaticum* (Fig. 1(E)–(H)) revealed that the extensive glycosylation previously observed in the *Sulfolobus islandicus* pilus is a response to acidic environments rather than extreme temperatures, as at even higher temperatures but neutral environments, much less glycosylation is observed in *Pyrobaculum* than in *Sulfolobus* and *Saccharolobus* pili.

Attachment to pili-like filaments was also confirmed for *Sulfolobus* turreted icosahedral virus (STIV), the type member of the *Turriviridae* family, which specifically recognizes unidentified pili-like filaments of its host, *Saccharolobus solfataricus*. The icosahedral STIV virions are decorated with turret-like protrusions at each of the fivefold vertexes. A three-dimensional reconstruction of the STIV-pilus interaction using cryo-electron tomography displayed that the turrets physically interact with the *S. solfataricus* pilus. Furthermore, tomographic reconstruction and sub-tomogram averaging unequivocally showed that pilus recognition occurs at the cleft between the second and third jelly-roll domains of the pentameric turret protein C381. Notably, structurally similar jelly-roll

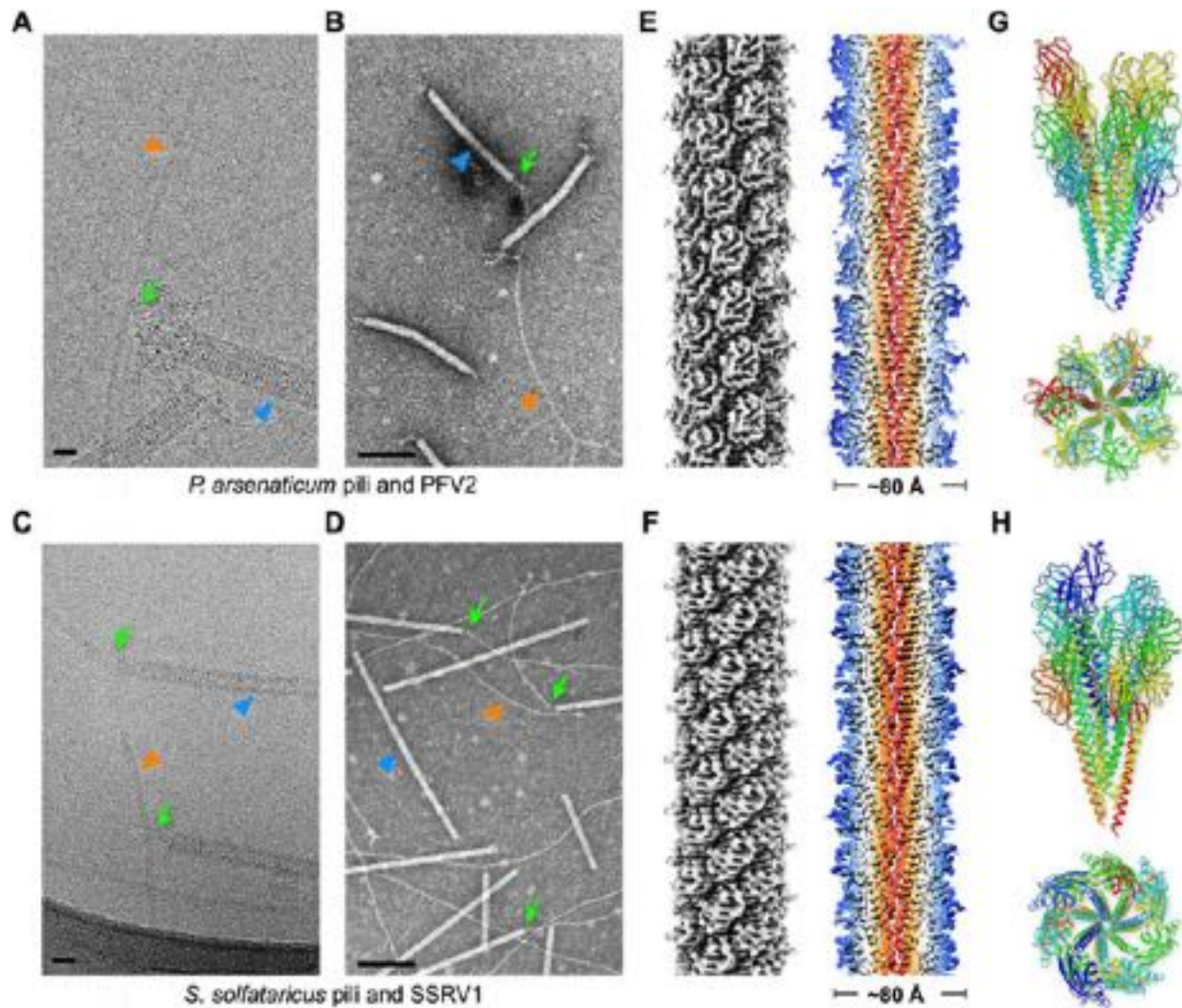


Fig. 1 Virus binding to archaeal type 4 pili. Representative cryo-electron (A) and negative staining (B) micrographs of the *P. arsenaticum* pili interacting with the hyperthermophilic virus Pyrobaculum filamentous virus 2 (PFV2). Scale bar, 20 nm in A and 200 nm in B. Orange arrowhead points to the *P. arsenaticum* pilus; blue arrowhead points to the PFV2 virion; green arrow points to the regions of the pilus-virion interaction. Representative cryo-electron (C) and negative staining (D) micrographs of the *S. solfataricus* pili interacting with the filamentous hyperthermophilic virus Saccharolobus solfataricus rod-shaped virus 1 (SSRV1). Scale bar, 20 nm in C and 200 nm in D. Orange arrowhead points to the *S. solfataricus* pilus; blue arrowhead points to the SSRV1 virion; green arrow points to the regions of the pilus-virion interaction. Cryo-EM reconstruction of the *P. arsenaticum* pilus at 3.8 Å resolution (E) and the *S. solfataricus* pilus at 3.4 Å resolution (F). Thin slices parallel to the helical axis of the pilus are shown, colored by the helical radius (G, H). Side view and top view of the *P. arsenaticum* and *S. solfataricus* pilus atomic models, built into the cryo-EM maps shown in E, F. The model is colored by chain. Image modified from Wang, F., Baquero, D.P., Su, Z., *et al.*, 2020. The structures of two archaeal type IV pili illuminate evolutionary relationships. *Nature Communications* 11, 3424.

domains have been implicated in receptor recognition in highly diverse viruses, including P22-like head-tailed phages (*Caudovirales*), tectiviruses and adenoviruses. Despite the molecular insights into the receptor binding, as in the case of SIRV2, it remains unclear how (and whether) the virus moves along the filaments and how circular virus genome is delivered into the host cytoplasm. Notably, adsorption to the pili of both SIRV2 and STIV was shown to occur exclusively at physiologically-relevant, high temperatures.

Interaction With Cell-Surface

Most archaea are surrounded by a thin proteinaceous surface layer (S-layer), consisting of glycosylated proteins anchored in the cell membrane. The S-layer is proposed to maintain an osmotic balance across the envelope, protect the cell from harsh environmental conditions and contribute to the cell shape. Notably, except for certain methanogens, archaea lack an equivalent of the bacterial peptidoglycan layer. The relative simplicity of the archaeal cell envelope likely underlies the mechanisms of virus entry and egress, and might have played a key role in shaping the virus diversity associated with archaea.

Spindle-shaped viruses of the *Fuselloviridae* family are often attached to membrane-derived vesicles or cellular fragments, suggesting that viral receptors are exposed on the host cell surface, with S-layer itself being a prime suspect. The spindle-shaped virion of the model fusellovirus SSV1 at one of the two pointed ends of the virion contains short terminal fibers, likely composed of protein VP4, which are postulated to mediate host recognition. Despite the observations suggesting possible interactions between SSV1 terminal fibers and the host cell surface, two recent studies on the *Sulfolobus* S-layer produced somewhat contradictory results regarding the role of S-layer for fusellovirus infection. On the one hand, the essential role of the S-layer in SSV1 infection appears to be supported by the finding that *Saccharolobus solfataricus* cells in which *slaB* encoding the membrane-anchoring S-layer protein was downregulated by a CRISPR-based silencing technology are less susceptible to SSV1 infection. On the other hand, cells in which both genes encoding the S-layer proteins *SlaA* and *SlaB* were deleted remained susceptible to infection with SSV9, a close relative of SSV1, suggesting that S-layer is not essential for either adsorption or infection by SSV9. These discrepant results call for additional studies focusing on the entry mechanism of fuselloviruses.

The entry process has also been explored for viruses infecting halophilic archaea. Insights have been obtained into the DNA ejection process in the halophilic spindle-shaped virus His1, the type member of the *Halspiviridae* family. Similar to fuselloviruses, at one of the pointed ends, His1 virion carries a putative receptor-binding module consisting of a central hub and six tail spikes. *In vitro* analysis demonstrated that His1 DNA ejection is unidirectional, occurs at a rate comparable to that of bacteriophage λ and is dependent on external osmotic pressure. Notably, *in vitro*, the His1 DNA ejection was only partial, suggesting that cellular factors are required for completion of the nucleic acid transfer. Interestingly, upon DNA ejection, the lemon-shaped virions transform into empty tubes, indicating that capsid proteins are capable of undergoing substantial quaternary structural changes.

Members of the family *Pleolipoviridae* have pleomorphic virions, which resemble membrane vesicles decorated with protruding spikes that in all likelihood participate in host attachment and membrane fusion processes. It has been suggested that upon binding to the receptor, the spike protein VP5 of Halorubrum pleomorphic virus 6 (HRPV-6) undergoes conformation change and drives fusion of the viral membrane with the host cytoplasmic membrane. Structural analysis of HRPV-6 virion by cryo-electron tomography and crystallography revealed that HRPV-6 VP5 has a unique V-shaped fold that is unrelated to the previously reported class I–III viral fusion protein.

The host recognition by archaeal head-tailed viruses appears to be mediated by tail fiber proteins, resembling the initial interactions of the bacteriophages of the order *Caudovirales*. Interestingly, the head-tailed dsDNA virus ϕ Ch1 infecting the haloalkaliphilic archaeon *Natrialba magadii* encodes a phase variation system. The system consists of an invertible region including a site-specific recombinase of the tyrosine recombinase superfamily interspersed between convergently oriented ORFs 34 and 36, which encode viral tail fiber proteins. Recombination in this region leads to an exchange of the gene fragments encoding the carboxy-termini of the tail fiber proteins, thereby creating a set of heterogeneous gp34 and gp36 proteins with distinct C-termini. Notably, only ORF34 is expressed during the virus infection indicating that fiber proteins of ϕ Ch1 are produced exclusively from this ORF. Binding assays showed that only one type of the heterogeneous tail fibers encoded by ORF34 interacts with *N. magadii*, suggesting that other tail fiber variants enable adsorption to other host strains. Galactose moieties present on the cell surface were implicated in host recognition by ϕ Ch1. Thus, as reported for bacterial viruses, the generation of two types of tail fiber proteins with distinct binding features might expand the ϕ Ch1 host range to quickly respond to habitat changes.

Kinetics of Virus Entry

There is a wide variation in terms of adsorption kinetics among archaeal viruses. Generally, hyperthermophilic archaeal viruses tend to display rapid adsorption rates, whereas halophilic viruses are notoriously slow. For instance, it has been shown that SIRV2 adsorbs very rapidly, with ~80% of the virions bound to the host cell within the first 30 s of infection. Similarly, ~50% of spindle-shaped SMV1 virions were attached to the cells within 1 min post-infection (p.i.). By contrast, only 30% of the salt-erprovirus His1 and siphovirus HHTV-1 virions adsorb to their host in 3 h. Similarly, the adsorption of the tailless icosahedral haloarchaeal virus HCIV-1 is rather efficient but slow, with the 80% of the particles adsorbed to cells at 4–5 h p.i. While the high adsorption rates of hyperthermophilic viruses are thought to minimize the time they are exposed to harsh extracellular conditions, including acidic pH and high temperatures, the low adsorption rates of haloarchaeal viruses are hypothesized to reflect an evolutionary adaptation to the long generation time of the hosts, whereby rapid adsorption might deplete the host population. Alternatively, it has been proposed that the changing salinity conditions of natural hypersaline environments may have favored the slow-adsorbing viruses that bind efficiently only under a specific range of salt concentrations.

Genome Replication

Very few studies have investigated experimentally the mechanisms of archaeal virus genome replication. In most cases, the mode of genome replication has been inferred from recognizable virus-encoded replication-associated genes, including rolling-circle replication initiation endonucleases (RCRE), DNA polymerases, replicative helicases and other components of the host replisome. Members of the families *Ampullaviridae*, *Thaspiviridae*, *Halspiviridae*, and *Ovaliviridae* as well as members of the genus *Gammapleolipovirus* (*Pleolipoviridae*) have linear dsDNA genomes and encode protein-primed family B DNA polymerases. By contrast, certain head-tailed haloarchaeal viruses encode RNA-primed DNA polymerases closely related to the corresponding proteins of their hosts. Generally, there is a correlation between the viral genome size and the completeness of the viral DNA replication machinery. Thus, viruses with small to

medium-sized genomes (5–50 kb) commonly encode only essential components of the replication machinery, with the rest of the components being recruited from the host, whereas viruses with large genomes (> 100 kb) appear to depend minimally on the host replisome, encoding nearly complete DNA replication machineries including DNA polymerases, proliferating cell nuclear antigen (PCNA), primases, replicative helicases, and homologs of the archaeal Orc1/Cdc6 replication initiators. Global analysis of dsDNA viral genomes demonstrated that replicative helicases are the most common replication proteins (75% of the genomes) in the dsDNA virus world. Indeed, some archaeal viruses with medium-sized genomes encode replicative minichromosome maintenance (MCM) helicases. Interestingly, phylogenetic analysis indicates that the *mcm* genes found in haloarchaeal, methanosarcinal, and methanococcal (pro) viruses were acquired from their respective hosts many times independently.

The RCRE are encoded by members of the *Pleolipoviridae* (genus *Alphapleolipovirus*) and *Sphaerolipoviridae* (genus *Betasphaerolipovirus*). However, experimental evidence for rolling circle replication, showing the presence of ssDNA replicative intermediates, has been obtained only in the case of sphaerolipovirus SNJ1. It has been shown that the SNJ1 RepA protein, an RCRE of the HUH superfamily, is indispensable for the genome replication of SNJ1. Closest homologs of SNJ1 RepA are encoded by plasmids of halophilic archaea, while more divergent homologs (13%–19% identity) are also identified in euryarchaea from the orders *Methanosarcinales* and *Thermoplasmatales*, as well as in ammonia-oxidizing archaea of the phylum *Thaumarchaeota*, underpinning a wide distribution of the SNJ1 RepA-like proteins across mobile genetic elements from diverse archaeal lineages. Interestingly, SNJ1 RepA also shares similarity with bacterial transposases of the IS91 family, which transpose via a rolling-circle-like mechanism, highlighting evolutionary connections between viruses, plasmids and transposons.

A divergent RCRE has been also identified in rod-shaped viruses of the *Rudiviridae* family. The protein has been structurally characterized and studied biochemically. *In vitro*, the Rep protein displays the expected nicking activity. However, its role in the viral genome replication remains unclear, because it does not seem to be expressed during the virus life cycle, at least, under laboratory conditions. Although genome replication has been most extensively studied for rudiviruses, the actual mechanism remains enigmatic. Members of the *Rudiviridae*, such as SIRV2, encode several proteins proposed to be involved in DNA replication, recombination and repair. Those proteins include the above mentioned Rep, an ssDNA-binding protein with a unique fold, a ssDNA-annealing ATPase, a Cas4-like ssDNA nuclease, a dUTPase, and a Holliday junction resolvase. Furthermore, yeast two-hybrid analysis showed that five SIRV2 proteins interact with the host DNA sliding clamp PCNA, known as a “molecular toolbelt” which interacts with multiple components of the host replisome. Presumably, SIRV2 recruits the host replication machinery for the assembly of the replisome on the viral DNA template. An immunofluorescence study of SIRV2 infected cells demonstrated that viral DNA synthesis is confined to a focus near the periphery of the cell. The study also provided evidence that the viral ssDNA-binding protein (gp17) as well as host PCNA and DNA polymerase I (Dpo1) are recruited to the site of viral DNA synthesis, confirming their essential role in the viral genome replication. Given that SIRV2 does not encode an identifiable DNA polymerase, the host polymerase Dpo1 found at the viral DNA synthesis site is likely to be involved in the replication of SIRV2 genome. An exceedingly complex model of SIRV2 replication has been proposed, whereby the virus employs a combination of strand-displacement, rolling-circle and strand-coupled replication mechanisms, which yields highly branched intermediates of about > 1200 kb (~34 viral genome units) with unusual ‘brush-like’ structures. However, it remains unclear how these three replication mechanisms are coordinated and whether all of them are essential for SIRV2 genome replication. In addition, the particular role of the host and viral proteins involved in the orchestration of viral DNA replication remains obscure.

It should be noted that most archaea-specific viruses do not encode identifiable DNA replication proteins, implying that these viruses either depend on the host DNA replication machinery or employ novel uncharacterized mechanisms for DNA replication. For instance, the genome replication of the lipothixvirus AFV1 has been suggested to start by the formation of a D-loop and progress by the strand displacement replication mechanism, whereas termination relies on recombination events through the formation of terminal-loop-like structures. However, the genes involved in this unique mechanism of replication are currently unknown. Similarly, the conspicuous absence of recognizable genes encoding type B DNA polymerases in viral genomes with linear dsDNA and terminal proteins, as in the case of the haloarchaeal sphaerolipoviruses PH1, SH1, and HHIV-2, suggests novel mechanisms of genome replication. A model resembling that of the *Streptomyces* linear plasmids has been proposed, whereby the cellular polymerase is employed for the viral genome replication and the terminal proteins are involved in the end patching of the DNA after replication.

Genome Integration

Many archaeal viruses are temperate and can undergo a lysogenic pathway in which they coexist with the host cell as proviruses. In most cases, the temperate viruses integrate their genomes into the host chromosome by the activity of viral site-specific integrases. However, sometimes proviruses can exist as circular extrachromosomal plasmids. For instance, the genomes of the euryarchaeal head-tailed virus ϕ Ch1 (family *Myoviridae*), pleomorphic virus SNJ2 (family *Pleolipoviridae*), and the crenarchaeal spindle-shaped SSV1 (family *Fuselloviridae*) integrate into the chromosome of *Natrialba magadii*, *Natrinema* sp. J7–1 and *Sulfolobus shibatae* cells, respectively, whereas the genomes of the sphaerolipovirus SNJ1 and myovirus ϕ H are stably maintained in a non-integrated circular form in *Natrinema* sp. J7–1 and *Halobacterium halobium* cells, respectively. Notably, the reactivation of the provirus typically occurs as a response to stressful conditions, such as DNA damage (e.g., by UV light or mitomycin C), temperature shock and shift from aerobic to anaerobic conditions. Depending on the virion release mechanism, the provirus induction can lead either to cell lysis, as for the head-tailed archaeal viruses and SNJ1, or to growth retardation, as observed for the fusellovirus SSV1 and pleolipovirus SNJ2.

Site-specific integrases of archaeal viruses, all belonging to the tyrosine recombinase superfamily, have been experimentally studied for two groups of viruses: fuselloviruses SSV1 and SSV2, and pleolipovirus SNJ2. The integrase of SNJ2 catalyzes homologous recombination between the viral attachment site located next to the integrase gene and the homologous site located within one of the tRNA genes on the host chromosome, leading to merger of the two genomes. Interestingly, integrases of fuselloviruses are unique in that the attachment site lies within the integrase gene itself and homologous recombination leads to disruption of the integrase gene into two fragments which flank the integrated provirus. Nevertheless, phylogenetic analysis has shown that SSV-like integrases have evolved from the more typical SNJ2-like integrases.

Transcription

Genomic analysis of archaeal viruses has shown a lack of genes encoding identifiable RNA polymerases as well as the presence of archaeal promoter elements, indicating a strong dependence on the host transcriptional machinery. However, the existence of putative transcription regulators in viral genomes suggests a viral-driven modulation to downregulate the host transcription and redirect the host transcriptional machinery towards the expression of specific viral genes in an efficient and temporal manner. Bioinformatics, biochemical, structural and transcriptomic efforts have provided valuable insights into the role of viral transcription regulators as well as into the modulation of the virus/host gene expression during the course of infection.

Transcription Regulators

A number of archaeal virus ORFs have been predicted to contain DNA-binding motifs typical of cellular transcription factors, with the majority displaying ribbon-helix-helix (RHH), winged helix-turn-helix (wHTH), or zinc (Zn) finger motifs. Notably, RHH and wHTH transcription factor sequences encoded by archaea and their viruses are bacterial-like, whereas archaeal Zn-finger domains share a common origin with those of eukaryotes. During the past years several virus transcription factors have been experimentally characterized and their structures have been solved. For instance, The AFV1p06 protein from the lipothrixvirus AFV1 is the first experimentally characterized archaeal Zn-finger protein. The protein preferentially binds to GC-rich DNA regions and displays a classical Zn-finger motif, albeit with an atypical substitution of one of the residues responsible for chelation of the Zn ion.

Most of the functionally characterized transcription regulators encoded by archaeal viruses are repressors, often also auto-regulating their own expression. For example, the transcription regulator SvtR encoded by the rod-shaped virus SIRV1 contains an RHH fold and forms a homodimer that resembles bacterial repressors CopG, NikR, and MetJ. Functional characterization of SvtR showed that its high-affinity binding-site corresponds to the promoter region of the *gp30* gene, which encodes a structural protein involved in the assembly of terminal fibers. Hence, it has been suggested that the primary function of SvtR is to prevent the premature expression of the structural protein encoded by *gp30*. Furthermore, SvtR binds to the promoter of its own gene (*gp08*), acting as a repressor and downregulating its own expression. The transcription regulator AvtR encoded by the filamentous virus AFV6 is highly conserved in the *Betalipothrixvirus* genus and consists of two RHH motifs connected by a linker. AvtR represses the expression of its own gene (*gp29*), but is able to both activate and repress the expression of the *gp30* gene in a concentration-dependent manner. Although in the case of both genes, AvtR binding sites are distant from the TATA boxes, DNase I footprinting assays showed that AvtR protects a region of approximately 100 nucleotides between the divergently oriented *gp29* and *gp30* promoters. This finding suggests that AvtR regulation depends on protein oligomerization on the DNA template. It has been proposed that AvtR binds to the initial binding site with high affinity and its oligomerization along the viral DNA induces cooperative binding to degenerate secondary sites with weaker affinity.

The transcriptional regulator F55 encoded by the fusellovirus SSV1 is one of the more extensively studied transcription factors encoded by archaeal viruses. Upon exposure to UV light, F55 plays a crucial role in the transition from the carrier to the induced state of SSV1. F55 is a dimer with an RHH DNA-binding motif. F55 recognizes tandem repeat sequences located within the promoters of the immediate early-induced transcripts T5, T6, and Tind as well as in its own promoter. Notably, binding of F55 to the target sequences leads to repression of the gene expression, whereas its dissociation causes activation of the transcription upon exposure of the SSV1 infected cells to UV light. A recent study using a variant of electrophoretic mobility shift assay (EMSA) coupled to mass spectrometry revealed that host RadA recombinase is associated with F55 when bound to the specific promoter sequences, forming a RadA-F55-dsDNA complex. The RadA recombinase belongs to the RecA/RadA/Rad51 protein superfamily and promotes DNA repair and recombination in hyperthermophilic archaea. Therefore, it has been proposed that RadA is a molecular sensor of the SSV1 host DNA damage analogous to the role of bacterial RecA protein in the life cycle of phage lambda. According to the proposed model, the exposure of infected cells to UV light causes massive DNA degradation that leads to accumulation of ssDNA regions. The RadA protein, which is coupled to F55 on the viral dsDNA genome, is recruited to the ssDNA regions and is progressively released from the RadA-F55-dsDNA complex, causing dissociation of F55 from the viral target sequences, thereby leading to activation of transcription of the immediate early transcripts. Notably, however, in vitro addition of ssDNA to the stable RadA-F55-dsDNA complex did not result in the release of the transcriptional block, prompting further studies to understand the specific role of RadA in the activation of the SSV1 transcription.

The regulation of the lysogeny has been also investigated for the haloarchaeal icosahedral virus SNJ1. The product of SNJ1 ORF4 has been recently suggested to controls the lysis-lysogeny switch. The expression level of ORF4 in the host cell appears to play a crucial role in the repression of genes responsible for the lytic pathway. Consistently, in absence of ORF4, SNJ1 produces

clear plaques, whereas in its presence, the plaques are turbid. In addition, ORF4 has been shown to play a key role in conferring immunity to the host cell against subsequent infections by SNJ1 (homotypic superinfection immunity) by repressing the genome replication of the superinfecting virus. ORF4 is conserved in other SNJ1-like proviruses, suggesting that the mechanisms behind the lysis-lysogeny switch and superinfection immunity is conserved in SNJ1-like proviruses.

Bicaudavirus ATV encodes an atypical transcriptional regulator, ORF145, which has apparently evolved from the major capsid protein, with which it shares significant sequence similarity. Similar to the major capsid protein, ORF145 is abundantly present in ATV virions. The protein binds to the host RNA polymerase (RNAP) with nanomolar affinity and inactivates it in a reversible manner via an allosteric mechanism. ORF145, renamed as RNAP inhibitory protein (RIP), binds apically to the DNA-binding channel of the RNAP and locks in a fixed position the RNAP clamp domain that typically switches between open and closed conformations during the transcription cycle. The high-affinity complex formed between RIP and the host RNAP inhibits the formation of transcription pre-initiation complexes and represses abortive and productive initiation as well as transcription elongation. Interestingly, RIP efficiently hinders the transcription directed from both host and viral promoters, suggesting a global inhibitory activity. Although the biological significance of a global transcriptional shutdown is unclear, it has been proposed that global repression prevents the host defense response. Notably, presence of RIP in the ATV virions suggests that host transcription might be repressed during the very early stages of the viral infection, most likely, to prevent the activation of the host type III-B CRISPR system.

Transcriptional Control

Whole-genome transcriptomic analyses using DNA microarray and RNAseq technologies have provided novel insights into virus–host interactions in crenarchaea. For example, SSV1 infected cells exhibit a tight chronological regulation of viral gene expression upon UV irradiation highly reminiscent of the strategy used by many bacterial and eukaryotic viruses. Thus, SSV1 temporal control leads to three clearly distinguishable sets of genes: immediate early, early and late genes. Transcriptomic analysis of the closely related virus SSV2 also displays a temporal regulation of gene expression which occurs in a distributive fashion with expressed genes not being adjacently located. Contrary to SSV1 and SSV2, analysis of gene expression of the lytic viruses SIRV2 and STIV display little temporal regulation of the viral gene transcription, with SIRV2 starting the transcription at multiple sites in the genome. Moreover, an RNAseq analysis on the two-tailed virus STSV2 shows that transcription of the majority of viral genes starts shortly after infection and increases throughout the infection cycle.

Studies on the host gene expression during virus infection showed that host response differs among archaeal viruses. While a small proportion of host genes was differentially expressed during the induction of SSV1 lysogens, the expression of several host genes encoding DNA replication, repair and transcription proteins was increased upon the infection with SSV2. Likewise, the transcription of more than one third of the *Sulfolobus* genes was differentially regulated as consequence of the SIRV2 infection. Notably, transcriptomic analysis of SIRV2 and STSV2 display a strong upregulation of antiviral defense genes including those for CRISPR-Cas and toxin-antitoxin systems. Consistently, infected *Sulfolobus islandicus* cells are actively undergoing CRISPR spacer acquisition from STSV2. These findings indicate that host response is virus-dependent since some viruses trigger a massive host response, whereas others, such as SSV1, are nearly unnoticed by the host. Interestingly, the same set of host genes can be up- and down-regulated by two different viruses, which suggests that distinct groups of host functions are required for the propagation of different archaeal viruses. For instance, the crenarchaeal cell division (*cdv*) operon, encoding homologs of the eukaryotic endosomal sorting complex required for transport (ESCRT) machinery, is downregulated upon SIRV2 and STSV2 infection but upregulated during the STIV infection.

Analysis of the host gene expression during virus infection also revealed significant upregulation of the host genes implicated in replication and DNA repair. The infection with fuselloviruses SSV1 and SSV2 leads to upregulation of genes encoding reverse gyrase, two subunits of the topoisomerase VI and the replication initiation protein Orc1/Cdc6, whereas the expression of the genes encoding the MCM helicase, PCNA and Dpo1 were boosted only in the case of SSV2. Similarly, reverse gyrase and Orc1/Cdc6-like genes were upregulated during the STIV infection, whereas genes implicated in energy production and metabolism were downregulated. The STSV2 infection leads to increased transcription of Holliday junction resolvase, DNA topoisomerase I and several DNA repair proteins, albeit the expression of the reverse gyrase, unlike for other crenarchaeal viruses, was downregulated. Differential regulation of genes encoding proteins involved in transcription suggests that viral transcription also relies on the host cell machinery. For instance, infection with SSV2, SIRV2 and STIV resulted in the upregulation of transcription initiation factor IIB, different subunits of the DNA-directed RNA polymerase and several transcriptional regulators.

Virion Egress

The last stage of the viral infection cycle corresponds to the egress of the new virions from the host cell. So far, release mechanisms have been studied in detail only for a handful of archaeal viruses. Generally, there are two types of archaeal viruses – those that disrupt the host cell upon virion egress (i.e., lytic viruses) and those that do not (i.e., non-lytic viruses). Similar to bacteriophages, some archaeal viruses can undergo a lysogenic life cycle in which the expression of most viral genes is suppressed and the virus is able to persist in the cell either as a provirus integrated into the host chromosome or as an episomal (extrachromosomal) element. In addition, some viruses can establish a carrier state, whereby the virus is stably maintained in a fraction of the cellular population, without causing cell lysis with the remaining cells being transiently resistant to the viral infection.

Cell Membrane Disruption

Lytic archaeal viruses have evolved several unrelated mechanisms for disruption of the cell envelope. The most extensively characterized mechanism involves formation of large pyramidal portals, dubbed virus-associated pyramids (VAP) (Fig. 2). VAPs develop on the surface of the infected cells, protrude through the S-layer (Fig. 2(A)–(D)) and open outward as flower petals (Fig. 2(E)–(H)), generating apertures through which the mature virions exit from the cell. This mechanism is used by viruses from at least three unrelated families, namely, *Rudiviridae*, *Turriviridae* and *Ovaliviridae*. VAPs of the rudivirus SIRV2 and turrivirus STIV exhibit seven-fold symmetry, while those of the ovalivirus SEV1 are six-sided. The VAPs consist of multiple copies of a single 10 kDa viral protein containing a transmembrane domain which promotes its insertion into the cellular membrane. Interestingly, heterologous expression of the SIRV2 pyramid protein P98 in archaeal (*Sulfolobus acidocaldarius*), bacterial (*Escherichia coli*) or eukaryotic (*Saccharomyces cerevisiae*) cells resulted in correct insertion into the membrane and formation of VAPs. Nevertheless, the signal triggering the opening of the pyramid structures appears to be archaea-specific since pyramids expressed in bacteria and eukaryotes were never observed in the open conformation. In addition, pyramidal structures have also been observed on the surface of an unidentified crenarchaeon, probably of the order *Thermoproteales*, infected with an unknown filamentous virus as well as on the surface of *Pyrobaculum oguniense* cells, suggesting that the VAP-based strategy is common among crenarchaeal viruses.

Head-tailed viruses infecting euryarchaea also lyse their host cells, but the underlying mechanism remains unknown. Bacterial relatives of archaeal head-tailed viruses lyse the cells using the holin–endolysin system, in which holin is a small membrane protein which forms lesions in the membrane, whereas endolysin is a peptidoglycan digesting enzyme. Phages infecting gram-negative bacteria often encode additional lysis proteins responsible for disintegration of the outer membrane. Thus far, homologs of the components constituting the holin–endolysin systems have not been identified in archaeal viruses. However, several viruses infecting methanogenic archaea, including siphovirus ψ M1 infecting *Methanothermobacter marburgensis*, encode pseudomurein endoisopeptidases which degrade the pseudomurein layer, which is equivalent, but not evolutionarily related to the bacterial peptidoglycan. Activity assays confirmed the cell wall-degrading activity of the pseudomurein endoisopeptidases which cleave the ϵ -isopeptide bond between alanine and lysine in the peptide chain of the pseudomurein. However, it remains unclear how these enzymes cross the cell membrane to reach the cell wall since no genes encoding potential holins have been identified in the genome of methanogenic viruses. Furthermore, pseudomurein is not universally present in euryarchaea or even methanogens. How head-tailed viruses infecting other archaea, such as halophiles, are released from the host remains unclear.

Lytic life cycle has been also demonstrated for filamentous enveloped viruses of the *Tristromaviridae* family which infect hyperthermophilic *Pyrobaculum* species (phylum *Crenarchaeota*). Electron microscopy analysis has shown that at the late stages of infection, the cells are packed with virions and their envelope is slashed by long, straight cuts not observed for other archaeal virus–host systems. However, the mechanism underlying this lysis mechanism has not been investigated.

Viral Release Without Membrane Disruption

Many archaeal viruses are released from the host without causing cell lysis. This type of virion egress has been most extensively studied on the example of fuselloviruses, in particular, SSV1. The assembly of SSV1 virions is concomitant with the egress and occurs by a mechanism that resembles the budding of enveloped eukaryotic viruses (Fig. 3). Dual-axis electron tomography analysis has shown that viral nucleoprotein complexes are extruded through the host cytoplasmic membrane in the form of tubular intermediate structures which share a continuous envelope with the host membrane (Fig. 3(A)). Subsequently, the SSV1 virions attached to the cell membrane undergo maturation to the characteristic spindle-shaped morphology. Formation of constricted ring-like structures (Fig. 3(B)) at the trailing end of the virion bud precedes the separation of the SSV1 virion from the cell membrane (Fig. 3(C)). The ring-like structures observed during the final step of SSV1 budding resembles the budding necks observed prior to the ESCRT machinery-mediated membrane scission during egress of some eukaryotic viruses, including human immunodeficiency virus and Ebola virus. The ESCRT system drives key membrane-remodeling processes in eukaryotes, such as cytokinesis, multivesicular body biogenesis and viral budding processes. Notably, proteins homologous to eukaryotic ESCRT components are conserved in several members of the *Sulfolobales* and play a central role in cell division, suggesting that, as for eukaryotic viruses, SSV1 budding may rely on a cellular membrane remodeling machinery. Similarly, pleolipoviruses have been suggested to possess a nonlytic life cycle and be released through a budding mechanism. Notably, however, unlike crenarchaea, halophilic archaea do not encode the ESCRT machinery.

Antiviral Defense and Viral Counterdefense Mechanisms

In most environments, viruses and their hosts are engaged in a constant evolutionary arms race. In this context, a broad range of host defense strategies as well as viral counterdefense mechanisms have been described in bacteria, whereas in archaea such mechanisms remain poorly understood, with the exception of the CRISPR–Cas system.

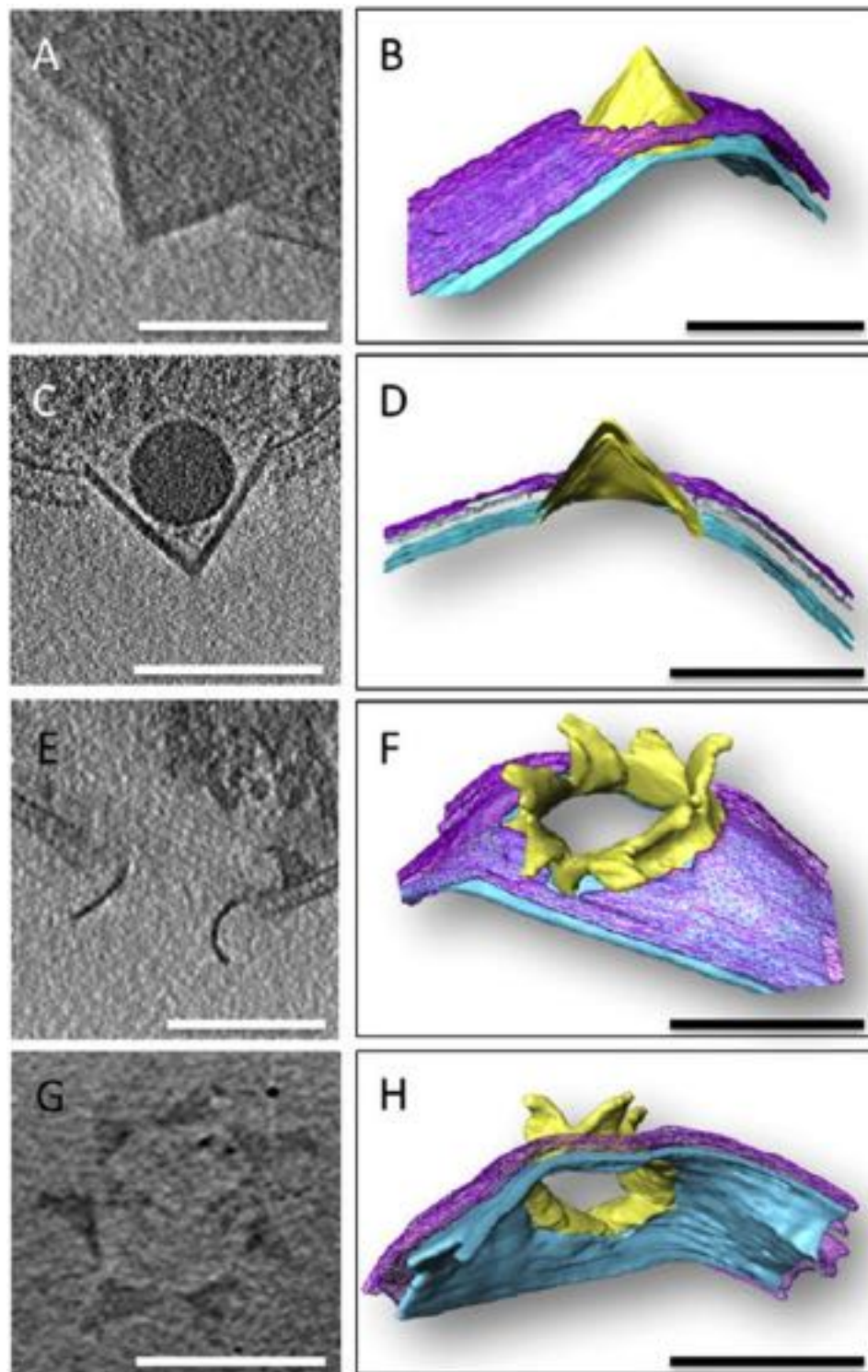


Fig. 2 Virus-associated pyramids (VAPs) in closed and open conformation. Tomographic slice (A, C, E, and G) and segmented, surface-rendered volumes (B, D, F, and H) of VAPs in the membrane of SIRV2-infected *S. islandicus* cells. VAPs are either closed (A–D) or open (E–H). The S-layer is purple, the cell membrane is blue, and the VAP is yellow. Scale bars, 200 nm. Image reproduced from Daum, B., Quax, T.E., Sachse, M., *et al.*, 2014. Self-assembly of the general membrane-remodeling protein PVAP into sevenfold virus-associated pyramids. *Proceedings of the National Academy of Sciences of the United States of America* 111, 3829–3834.

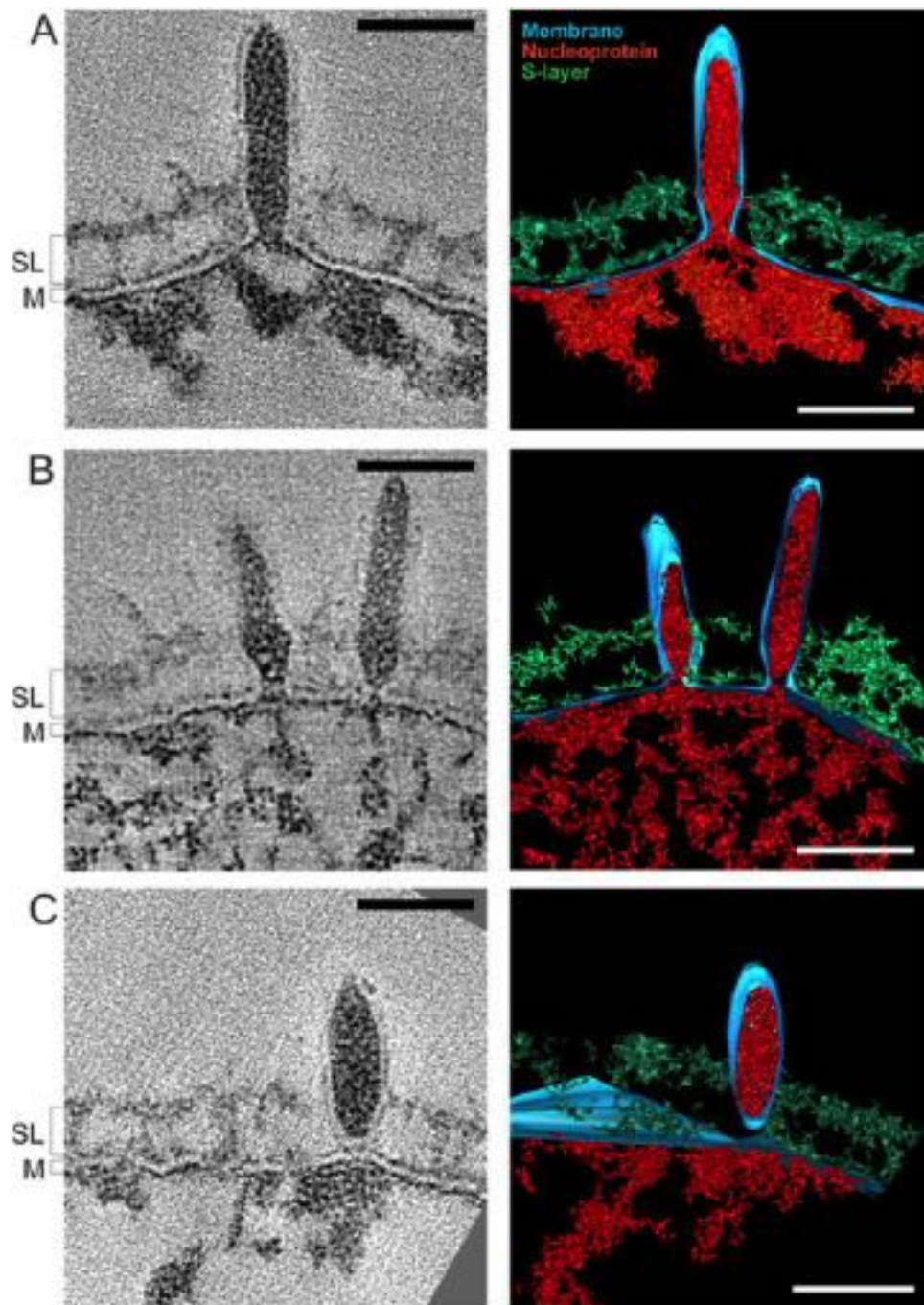


Fig. 3 Different stages of SSV1 budding. Slices through tomograms (left) and volume segmentations (right) showing concomitant assembly and release of SSV1 virions. (A) SSV1 virions are attached to the host cell surface, with their envelope continuous with the cell membrane; (B) Presence of a constricted budding neck at the trailing end of the SSV1 virion bud; (C) SSV1 virion separated from the cell membrane. Red, putative nucleoprotein; blue, lipid membrane (M); green, S-layer (SL). Scale bars, 50 nm. Image reproduced from Queminn, E.R., Chlanda, P., Sachse, M., *et al.*, 2016. Eukaryotic-like virus budding in archaea. *mBio* 7, e01439–16.

Antiviral Defense Mechanisms

Mechanistically, the defense systems of archaea and bacteria can be classified into three main groups: (1) variation of virus receptors, (2) innate and adaptive immunity and (3) dormancy and programmed cell death. The variation of virus receptors includes programmed changes, such as phase variation and physical masking of the receptor in order to hamper the successful binding of the virus to the host cell. Defense mechanisms that rely on immunity involve the recognition and inactivation of invader genetic material either by nonspecific innate immunity, such as restriction modification modules and Argonaute-based

innate immunity, or highly specific adaptive immunity, represented by CRISPR-Cas systems. Finally, strategies based on induction of dormancy or programmed cell death upon viral infection include toxin-antitoxin (TA) systems, in which the infection disrupts the balance of the host toxin-antitoxin complex leading to retardation of the cell growth or death. During the past few years the study of antiviral defense mechanisms in archaea has been focused on CRISPR-Cas systems, whereas hints about the role of toxin-antitoxin system in virus-host interactions has been provided by transcriptomic analysis of infected cells.

CRISPR-Cas systems

The CRISPR-Cas systems represent prokaryotic adaptive immunity mechanism present in about 40% of bacteria and almost all archaea. This system protects cells against invasion of mobile genetic elements, such as viruses and plasmids. CRISPR-Cas immunity involves three main stages: (1) adaptation stage, during which new virus/plasmid-derived sequences (spacers) are integrated into the CRISPR array adjacent to the leader sequence of the CRISPR array, (2) processing stage, during which the CRISPR array is transcribed and processed into separate CRISPR RNAs (crRNAs) containing the spacer sequence with 5' and 3' tags from the flanking repeats of the CRISPR array, and (3) interference stage, whereby the crRNA binds to the assembled CRISPR effector complex (interference module) to recognize and degrade DNA and/or RNA molecules containing the protospacer sequence. Based on gene synteny and the composition of the effector complexes, CRISPR-Cas systems have been divided into two classes, each subdivided into three types and several subtypes. CRISPR-Cas systems are very abundant among archaea and often several different types of CRISPR systems are encoded on the same archaeal genome.

It has been observed that SIRV2 infection leads to massive activation of the CRISPR-Cas systems present in *Sulfolobus islandicus* LAL14/1. The genome of *S. islandicus* LAL14/1 contains six CRISPR-*cas* loci encoding complexes of the three subtypes: I-A, I-D and III-B. The SIRV2 infection led to a sharp increase of the transcription levels of all CRISPR-*cas* arrays, except for one incomplete type III-B CRISPR-Cas module lacking the CRISPR array. Moreover, the expression of CRISPR arrays was activated immediately after SIRV2 infection and steadily increased during the course of the infection, with the highest levels of *cas* expression being reached 1 hpi. Notably, the expression levels of different CRISPR-Cas loci varied upon SIRV2 infection, suggesting their specialized roles during viral infection. The ability of SIRV2 to propagate in *S. islandicus* LAL14/1 despite the presence of active CRISPR-Cas systems carrying spacers matching the viral genome suggested the presence of viral anti-CRISPR (Acr) proteins which could counteract the host immune response. Similarly, a transcriptomic analysis on STSV2 infected cells revealed a differential expression level of the CRISPR-Cas systems upon infection. *S. islandicus* REY15A, the STSV2 host, encodes one type I-A and two type III CRISPR systems. Upon STSV2 infection, the type I-A module was strongly upregulated, whereas the expression level of one type III-B complex was downregulated through the course of STSV2 infection and the second type III-B complex was weakly upregulated. The interaction between tailed spindle-shaped virus SMV1 and the CRISPR-Cas systems of *S. islandicus* REY15A was studied in more detail by constructing strains carrying plasmid-borne mini-CRISPR arrays targeting SMV1 genome. The CRISPR response against SMV1 was type-specific, with the III-B CRISPR complex showing a tight control on the inhibition of the viral replication and proliferation during the infection, whereas the I-A CRISPR complex gradually lost the control of the viral proliferation allowing viral replication and release. The absence of SMV1 escape mutations conferring tolerance to I-A CRISPR system suggests that the virus has evolved a mechanism, probably an Acr protein, specific against the type IA CRISPR system.

Rudivirus SIRV3, a closely related virus to SIRV2, undergoes a host-dependent carrier state infection in *Sulfolobus islandicus* REY15A whereby the virus is maintained in a small fraction of the population over several days without apparent chromosomal DNA degradation, disturbance of the cell growth or induction of detectable CRISPR-Cas response. Notably, coinfection with the bicau-davirus SMV1 did not affect the SIRV3 carrier state and led to the coexistence of both viruses in the culture over 12 days, despite the induction of CRISPR spacer acquisition from SIRV3 DNA and the increased transcription of the subtype I-A CRISPR-Cas module. A plausible explanation for the maintenance of both viruses in the cell cultures seems to be that host CRISPR-Cas systems are inhibited by the virus-encoded Acr proteins encoded by SIRV3 and SMV1, which most likely target different types of CRISPR-Cas complexes encoded by the host. Transcriptomic analysis of *S. solfataricus* P2 cells infected with the fusellovirus SSV2 showed strong upregulation of the six CRISPR loci. By contrast, such effect was not observed when *S. solfataricus* P2 was infected with SSV1, a close relative of SSV2. Surprisingly, the co-infection with both viruses caused the silencing of the host CRISPR response. The major difference between SSV1 and SSV2 lies in the UV-inducible operon present in SSV1 but not in SSV2. Therefore, it has been speculated that SSV1 UV-inducible operon encodes transcription factors that may silence the CRISPR-Cas response in the SSV1-infected strain.

Interestingly, CRISPR-mediated defense systems appear to be employed not only by cells to defend against mobile genetic elements, but also by viruses for interviral conflicts. It was discovered that some archaeal viruses carry mini-CRISPR arrays (with 1–2 repeat-spacer units), which are preceded by promoter-containing leader sequences and genetic determinants required for insertion of new spacers. Remarkably, most of these virus-borne spacers target closely related viruses present in the same population. For instance, SPV1 and SPV2, two closely related viruses of the family *Portogloboviridae* isolated from a Japanese hot spring, were found to possess mini-CRISPR arrays with spacers reciprocally targeting each other. The existence of related viruses carrying mini-CRISPR arrays against each other in the same population suggests that viruses have adapted the host defense system for interviral conflicts and use it as a mechanism of heterotypic superinfection exclusion, in which a cell infected by one virus becomes resistant to another closely-related virus.

Toxin-antitoxin system

The toxin-antitoxin (TA) modules consist of a toxin protein capable of inhibiting the cell growth and an antitoxin that neutralizes the action of the toxin. TA modules can be encoded by both cellular organisms and mobile genetic elements, more commonly, plasmids

but also viruses. The toxin and its cognate antitoxin form a stable complex preventing the toxin from exerting its toxic effect in normally growing cells. Whereas toxin is a stable protein, often a nuclease, the antitoxin is typically labile, with quick turnaround, and has to be constantly synthesized. Under stressful circumstances, when the production of antitoxin is perturbed, the toxin is unleashed, leading to cell dormancy or death. Given that virus infection can lead to disbalance in toxin-antitoxin equilibrium, TA clusters have been proposed to represent a mechanism of abortive infection, in which an infected cell commits “altruistic suicide” to protect the population. Although the role of TA systems in virus-host interactions has been poorly understood in archaea, it has been suggested that *csa5* represents a toxin gene in the type I-A CRISPR system of *Saccharolobus solfataricus* P2, because Csa5 is toxic in the CRISPR-deficient *S. solfataricus* strain. Infection with rudivirus SIRV2 leads to induction of the Csa5 expression and formation of Csa5 oligomers in *Sulfolobus* cells, suggesting that Csa5 may be involved in programmed cell death in response to virus infections. Future work is required to gain insights into the mechanism of Csa5 toxicity and its biological relevance.

Transcriptomic analyses have revealed an upregulation of the type II TA systems upon viral infection. The type II TA systems are widespread among members of the order *Sulfolobales* and consist of an antitoxin protein that directly binds and inhibits the effect of a toxin protein. *S. islandicus* LAL14/1 encodes 16 operons of the family VapBC (virulence associated proteins B and C) and 6 operons of the family HEPN-NT (higher eukaryote and prokaryote nucleotide binding-nucleotidyltransferases), whereby the toxins are proposed to perform RNA cleavage. Notably, the expression of 11 out of 16 VapBC and 3 out of 6 HEPN-NT loci increased after SIRV2 infection at different time points. In most cases, the expression level of both genes coding the toxin and antitoxin was upregulated to similar extents. Likewise, bicaudavirus STSV2 infection caused a strong upregulation of the host TA gene pairs, including *vapBC*. Even though the role of TA operons remains unclear, the increase in TA gene expression after infection strongly suggests a function in host defense response.

Counterdefense Mechanisms

Viruses have evolved strategies to evade the targeting by host CRISPR-Cas adaptive immunity. The simplest mechanism involves mutations in the protospacer adjacent motifs (PAM), a short motif necessary for correct CRISPR targeting. A more sophisticated mechanism relies on the dedicated anti-CRISPR proteins (Acrs) which inhibit CRISPR-Cas by a diversity of mechanisms, including binding to specific subunits of the effector complexes. Although Acrs have been studied in many bacteriophages, only three archaeal Acr proteins have been characterized, despite the high abundance of CRISPR systems in archaeal genomes.

All three archaeal Acrs were discovered in rudiviruses, although some of them display much broader distribution. The AcrID1 protein inhibits subtype I-D CRISPR-Cas system by binding directly to the Cas10d subunit of the I-D CRISPR-Cas effector complex. The AcrID1-Cas10d interaction blocks the interference stage of the I-D CRISPR-Cas response in which the CRISPR RNAs (crRNAs) and Cas proteins form an effector complex that recognizes the viral sequence complementary to that of the crRNA spacer and cleaves it. Notably, AcrID1 is a conserved dimeric $\alpha\beta$ -sandwich protein widely distributed in viruses infecting hyperthermophilic crenarchaeota of the order *Sulfolobales*, including rudiviruses, lipothrixviruses, fuselloviruses, and monocaudaviruses, with about 50 homologs.

The two other Acr proteins disarm the type III CRISPR systems. Type III CRISPR-Cas systems exhibit a high complexity and consists of four subtypes (III-A–III-D), of which subtypes III-A (Csm) and III-B (Cmr) have been the most studied ones. In both cases, the type III effector complex binds to the viral protospacer causing the activation of the Cas10 protein as well as the synthesis of cyclic oligoadenylates (cOAs) from ATP. The presence of cOAs, in turn, triggers the activation of the Csm6 RNase in the III-A system and Csx1 in the III-B system to cleave the viral mRNA. It has been recently demonstrated that SIRV2 gp48, which is conserved in several members of *Rudiviridae* and *Lipothrixviridae*, is an Acr protein that exclusively inhibits the subtype III-B CRISPR-Cas system (AcrIIIB1) of *S. islandicus* LAL14/1. AcrIIIB1 was demonstrated to bind to two distinct effector complexes of the subtype III-B system, Cmr- α and Cmr- γ , suggesting that the mechanism by which AcrIIIB1 inhibits the subtype III-B response is by interfering with the Csx1 RNase activation process. The third archaeal Acr belongs to the DUF1874 protein family and was named AcrIII-1 family, because unlike all other known Acr of bacteria or archaea, it is not specific for a particular subtype, but blocks all subtypes of type III. AcrIII-1 is an enzyme with a ring nuclease activity, rapidly degrading the cyclic tetra-adenylate (cA₄) second messenger into a linear di-adenylate (ApA>P) with a cyclic 2',3'-phosphate, thereby preventing the activation of the type III-associated RNase and, therefore, blocking the host type III CRISPR defense system. Given that the target of AcrIII-1 is a signaling molecule (cA₄) with a constant structure, rather than specific CRISPR effector protein, this Acr is likely to be able to inhibit any type III CRISPR subtype using cA₄ as part of its activation. Consistently, the AcrIII-1 family is widely distributed in viruses infecting both archaea and bacteria as well as plasmids and proviruses.

Conclusions

Since the first archaeal virus was isolated in 1977, a considerable effort has been made to understand their virion architectures, genome contents and modes of interaction with their hosts. However, despite the substantial progress, a comprehensive understanding of molecular mechanisms that underlie these interactions is still lacking. The continued exploration of the archaeal virus diversity and development of new model systems, coupled with advances in molecular biology, genetics and microscopy tools are expected to shed further light on the mechanisms of virus-host interactions in archaea.

Further Reading

- Atukoralage, J.S., McMahon, S.A., Zhang, C., *et al.*, 2020. An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* 577, 572–575.
- Daum, B., Quax, T.E., Sachse, M., *et al.*, 2014. Self-assembly of the general membrane-remodeling protein PVAP into sevenfold virus-associated pyramids. *Proceedings of the National Academy of Sciences of the United States of America* 111, 3829–3834.
- Fusco, S., She, Q., Fiorentino, G., Bartolucci, S., Contursi, P., 2015. Unravelling the role of the F55 regulator in the transition from lysogeny to UV Induction of *Sulfolobus* spindle-shaped virus 1. *Journal of virology* 89, 6453–6461.
- Hanhijärvi, K.J., Žiedaite, G., Pietilä, M.K., Hægström, E., Bamford, D.H., 2013. DNA ejection from an archaeal virus – A single-molecule approach. *Biophysical Journal* 104, 2264–2272.
- Hartman, R., Eilers, B.J., Bollschweiler, D., *et al.*, 2019. The molecular mechanism of cellular attachment for an archaeal virus. *Structure* 27, 1634–1646. (e1633).
- He, F., Bhoobalan-Chitty, Y., Van, L.B., *et al.*, 2018. Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nature Microbiology* 3, 461–469.
- Krupovic, M., Cvirkaite-Krupovic, V., Iranzo, J., Prangishvili, D., Koonin, E.V., 2018. Viruses of archaea: Structural, functional, environmental and evolutionary genomics. *Virus Research* 244, 181–193.
- Luk, A.W., Williams, T.J., Erdmann, S., Papke, R.T., Cavicchioli, R., 2014. Viruses of haloarchaea. *Life* 4, 681–715.
- Makarova, K.S., Wolf, Y.I., Iranzo, J., *et al.*, 2020. Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nature Reviews Microbiology* 18, 67–83.
- Martínez-Alvarez, L., Deng, L., Peng, X., 2017. Formation of a viral replication focus in *Sulfolobus* cells infected by the rudivirus *Sulfolobus islandicus* rod-shaped virus 2. *Journal of Virology* 91, e00486–17.
- Medvedeva, S., Liu, Y., Koonin, E.V., *et al.*, 2019. Virus-borne mini-CRISPR arrays are involved in antiviral conflicts. *Nature Communications* 10, 5204.
- Okutan, E., Deng, L., Mirlashari, S., *et al.*, 2013. Novel insights into gene regulation of the rudivirus SIRV2 infecting *Sulfolobus* cells. *RNA Biology* 10, 875–885.
- Ortmann, A.C., Brumfield, S.K., Walther, J., *et al.*, 2008. Transcriptome analysis of infection of the archaeon *Sulfolobus solfataricus* with *Sulfolobus* turreted icosahedral virus. *Journal of Virology* 82, 4874–4883.
- Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature Reviews Microbiology* 15, 724–739.
- Quax, T.E., Voet, M., Sismeiro, O., *et al.*, 2013. Massive activation of archaeal defense genes during viral infection. *Journal of Virology* 87, 8419–8428.
- Quemin, E.R., Chlanda, P., Sachse, M., *et al.*, 2016. Eukaryotic-like virus budding in archaea. *mBio* 7, e01439-16.
- Quemin, E.R., Lucas, S., Daum, B., *et al.*, 2013. First insights into the entry process of hyperthermophilic archaeal viruses. *Journal of Virology* 87, 13379–13385.
- Sheppard, C., Werner, F., 2017. Structure and mechanisms of viral transcription factors in archaea. *Extremophiles* 21, 829–838.
- Wang, F., Cvirkaite-Krupovic, V., Kreutzberger, M.A.B., *et al.*, 2019. An extensively glycosylated archaeal pilus survives extreme conditions. *Nature Microbiology* 4, 1401–1410.
- Wang, F., Baquero, D.P., Su, Z., *et al.*, 2020. The structures of two archaeal type IV pili illuminate evolutionary relationships. *Nature Communications* 11, 3424.

Antiviral Defense Mechanisms in Archaea

Qunxin She, Shandong University, Qingdao, China

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Anti-CRISPR (Acr) genes Viral genes that code for proteins that interact with effectors of CRISPR-Cas systems to inhibit their antiviral immunity.

cOA signal transduction pathway Cyclic oligoadenylates (cOA) synthesized by type III CRISPR-Cas systems function as a secondary signal that bind to RNase of Csm6/Csx1 family to activate the enzyme for general cellular RNA degradation, leading to cell dormancy or cell death.

CRISPR-Cas systems The prokaryotic adaptive immune system that mediates small RNA-guided target nucleic acids destruction.

PAM-dependent DNA interference One of the CRISPR immunity mechanisms, in which CRISPR-Cas systems specifically recognize protospacer-adjacent motif to distinguish native versus foreign DNA and only target foreign DNA for destruction.

Transcription-dependent CRISPR interference Another unique CRISPR immunity mechanism, in which CRISPR-Cas systems rely on target transcription for mediating DNA interference.

Introduction

Archaea and bacteria coexist with mobile genetic elements (MGEs) such as viruses and plasmids in various environments. To maintain their ecological fitness, these microbes interact with viruses beneficially on one hand and develop antiviral mechanisms to prevent growth inhibition and cell death caused by harmful viruses on the other hand. These include the restriction and modification (R–M) systems, abortive infection, and toxin–antitoxin (TA) systems, prokaryotic Argonaute (pAgo), and most recently identified DNA phosphorothioate modification and virus replication inhibition systems, all of which provide innate antiviral mechanisms. About 12 years ago, antiviral immunity based on the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas (CRISPR-associated) system was discovered, and it is present in about 90% of archaea and 40% of bacteria. To date, CRISPR-Cas represents the only adaptive immune system known for prokaryotes.

Archaea and bacteria often thrive in the same environments, and they exchange genetic materials via horizontal gene transfer. For this reason, it is not surprising that organisms of these two prokaryotic domains share many strategies of antiviral defense. To date, many bacterial antiviral systems have been studied in great detail, and this has facilitated the investigation of corresponding archaeal antiviral systems. In particular, important progress has been made in the study of archaeal CRISPR immunity due largely to the boom in CRISPR biology and CRISPR technology research in the past 12 years. Here, I mainly summarize the current understanding of archaeal CRISPR-Cas systems.

CRISPR-Cas: The Prokaryotic Adaptive Immune System

There are two genetic entities in CRISPR-Cas: CRISPR arrays carry repeats that are interrupted by unique short DNA sequences derived from viruses and other mobile genetic elements (protospacers) whereas *cas* gene cassettes code for structural proteins and enzymes. For antiviral defense, the immune system starts with searching for protospacers on invading viruses and incorporating them into CRISPR arrays as spacers during the first encounter. Upon a subsequent encounter, small RNAs generated from the acquired spacers guide the immune system to specifically recognize the corresponding protospacers and target the viruses for destruction.

Genes coding for Cas proteins are grouped together for each system, forming *cas* gene modules. Although distantly related Cas proteins can share limited or no sequence similarity, those of the same functional category have adopted very similar structures. At present, CRISPR-Cas systems can be categorized into two broad classes and six main types based on their gene synteny, sequence similarity of Cas proteins, and mechanisms of interference. Class 1 CRISPR-Cas systems employ multisubunit effector complexes to mediate CRISPR interference such as the systems belonging to types I, III, and IV, while those of class 2 rely on a single multidomain protein to perform the same function, including types II, V, and VI.

By now, it has been demonstrated that type I, II, and V systems target DNA for destruction whereas those of type VI show RNA interference (RNAi). Furthermore, type III CRISPR systems exhibit multiple CRISPR interference activities, including dual DNA and RNA interference and activation of a signal transduction pathway with cyclic oligoadenylates (cOAs) as the second messenger to activate Cas-accessory RNases, leading to indiscriminate cellular RNA degradation. Our current understanding on molecular mechanisms of archaeal CRISPR-Cas systems is summarized below.

Spacer Acquisition and its Regulation in Archaea

Spacer acquisition is the first step toward the generation of adaptive CRISPR immunity in prokaryotes. The process involves two core proteins, Cas1 and Cas2, that form a hexameric integrase. The enzyme complex contains four subunits of Cas1 and two

subunits of Cas2. It interacts, on one hand, with the first repeat of a CRISPR array and on the other hand, binds to a protospacer, a DNA segment derived from invading nucleic acids. In the end, the integration reaction adds the DNA segment into the CRISPR array at the end of the first repeat, forming the new first spacer. For this reason, the order of spacers in CRISPR arrays provides the chronological record of previous infections on the host chromosome by viruses and other genetic elements.

Archaea and bacteria often contain additional genes in the same locus of *cas1* gene. For example, in the genomes of *Sulfolobus* species, *cas1* and *cas2* are clustered with three other genes, *csa3a*, *csa1*, and *cas4*, forming the adaptation Cas (aCas) module. In this module, *csa3a* is the first gene and codes for a transcriptional factor. Based on the gene organization, Csa3a may function as a regulator to control expression of all aCas genes in the operon. Indeed, Csa3a activates not only the expression of aCas genes but also an array of DNA repair genes. Since bacterial DNA repair proteins are shown to be involved in protospacer selection, these archaeal DNA repair proteins are implicated in the same process.

Csa3-like transcriptional factors are interesting since they carry a ligand-binding domain called CRISPR-associated Rossmann fold (CARF). Other CARF-containing CRISPR-accessory proteins include Csm6 and Csx1 RNases associated with type III CRISPR-Cas systems. Type III systems catalyze the synthesis of cOA second messenger that interacts with the CARF domain of Csm6/Csx1 RNases (see below). However, the ligand that binds to the *Sulfolobus islandicus* Csa3 CARF domain remains to be identified.

Proteins encoded by remaining two genes (*csa1* and *cas4*) belong to the so-called Cas4 superfamily. Cas4 proteins interact with Cas1 and specify the upstream protospacer adjacent motif (PAM). Together they define the spacer length and recognize a downstream motif to ensure proper generation of pre-spacers as well as functional spacer integration. They are essential for efficient spacer acquisition in archaea. Strikingly, *cas4* genes are not only widespread on archaeal and bacterial chromosomes; viruses and plasmids can also code for such proteins albeit they are only distantly related to those present in aCas modules. Possibly, these viral Cas4 proteins can also interact the CRISPR acquisition machinery to downregulate the spacer acquisition or producing nonfunctional spacers to unarm the CRISPR immunity. For this reason, the virus-encoded Cas4 proteins can be classified as anti-CRISPR proteins at the stage of spacer acquisition.

Expression of CRISPR Loci and crRNA Biogenesis

All CRISPR loci show polarity as they provide a fossil record of MGE infection on the host chromosome. At the proximal end, there is a leader sequence that specifies the site for incorporation of new spacer as described above. This leader region also contains a promoter that drives the expression of the entire CRISPR array, producing precursor crRNAs. CRISPR loci can carry hundreds of spacers from which leader-proximal spacers are generally expressed to a high level. Nevertheless, there are also exceptions since crRNAs generated from some leader-distal spacers can also be very abundant. There are two possible reasons for the differential expression. First, spacers can carry promoter-like sequences (internal promoters) that increase the expression level of their downstream spacers; second, some DNA-binding protein may interact with CRISPR loci and modulate their expression such as the *Sulfolobus solfataricus* CRISPR-repeat binding protein. Since CRISPR loci are generally expressed to a high level in archaea, the significance of the modulated CRISPR expression remains elusive.

After transcription, the resultant precursor crRNAs are processed by a specialized enzyme of the Cas6 endonuclease family, producing small RNAs of single repeat-spacer unit. Although some bacterial CRISPR-Cas systems also employ other Cas proteins for crRNA processing, all known archaea only utilize Cas6 for crRNA biogenesis. In fact, many archaea contain several CRISPR-Cas systems and more than one *cas6* gene. For example, *S. solfataricus* has six CRISPR loci with two families of repeats, four *cas6* genes, and three different types of effector complex. The encoded processing enzymes show optimal activity to distinct repeat sequences and this scenario fits with the specific crRNA processing for different types of antiviral immunity. Nevertheless, *S. islandicus* REY15A has only one *cas6* gene although this archaeon contains three active CRISPR systems. Thus, this *cas6* is responsible for production of crRNAs to be used by all three immune systems. Furthermore, since the sizes of crRNAs are different for these antiviral systems, crRNAs generated by Cas6 are further processed and the mechanism of crRNA maturation can be type-specific.

Archaeal Type I CRISPR-Cas Systems

Bioinformatics analysis has revealed eight subtypes of type I CRISPR-Cas systems (I-A through I-G, I-U) among which I-A, I-B, I-C, I-D, I-E, and I-G are found in archaea. CRISPR-Cas systems of 3 type I subtypes are characterized in archaea, including I-A systems of *S. solfataricus*, *S. islandicus*, *Aeropyrum pernix*, and *Pyrococcus furiosus*, and I-B and I-G systems of *Haloferax volcanii* and *P. furiosus*, respectively (Fig. 1). These archaeal systems code for four structural Cas proteins: Cas7 is a backbone subunit that is present in multiple copies and binds to crRNA; Cas5 interacts with the 5'-handle of crRNAs; Cas11 is a small subunit interacting with Cas7 and crRNA; whereas Cas8, the largest subunit, interacts with Cas5 and Cas7 and functions in PAM recognition. They also code for their own Cas6 to generate mature crRNAs and Cas3 to mediate DNA interference. Cas3 is the type-specific Cas protein of type I systems. For most type I systems, the enzyme harbors a helicase and a nuclease domain but in I-A systems, the two domains are split into two proteins, denoted Cas3' and Cas3'' that contain the helicase domain and the nuclease domain, respectively.

Cas proteins of type I systems are known to form CRISPR-associated complexes for antiviral defense (CASCADE) to mediate antiviral immunity. Two basic types of CASCADE complexes are known for type I systems: Those represented by *Escherichia coli* I-E system are not associated with Cas3, and the enzyme is recruited into the CASCADE upon recognition of a PAM before DNA

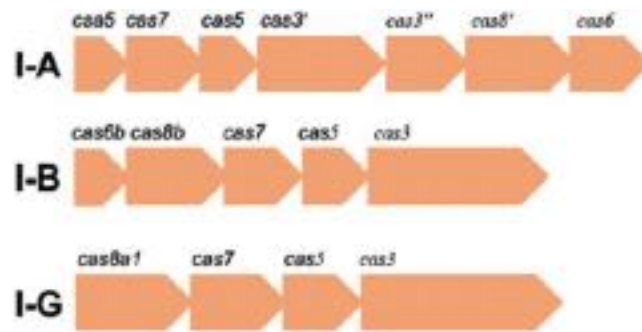


Fig. 1 Gene organization of three type I subtypes of CRISPR-Cas systems. I-A: *S. islandicus* I-A module; I-B: *H. volcanii* I-B system; I-G: *P. furiosus* I-G module. For I-A and I-G systems, their backbone subunits (Cas7) are also called Csa2 and Cst2; large subunits (Cas8) as Csa4 and Cst1, respectively. Small subunit (Cas11) is only present in I-A, i.e., Csa5.

silencing. In I-A subtype systems, Cas3' and Cas3'' are integrated parts of CASCADE as shown for *A. pernix* and *P. furiosus* I-A systems. The other studied archaeal type I systems are *H. volcanii* I-B system and *P. furiosus* I-G complex, and they resemble the I-E effector of *E. coli*. To date, the structure of archaeal type I CASCADE remains to be determined.

Nevertheless, biological function of archaeal type I CRISPR-Cas systems has been demonstrated by using interference plasmid assay (also called invader plasmid assay). In this approach, an interference plasmid was constructed carrying a protospacer that matches one of the spacers in the chromosomal CRISPR arrays. Further, the plasmid also carries a PAM sequence that is absent from CRISPR loci in the host genome. Both interference plasmids and reference plasmids were introduced into archaeal cells by transformation, and the former showed 100–1000 folds of reduced transformation efficiency compared with the latter. Further, all transformants on plates were found to have lost the I-A CRISPR immunity either due to spontaneous deletion of targeting-spacer, or plasmid rearrangement, or mutation of a I-A *cas* gene. They are called escape mutants. Moreover, since PAM is absent from the spacer on the chromosomal CRISPR array, self-targeting by CASCADE is therefore avoided. The mechanism of the *S. islandicus* I-A DNA interference is illustrated in Fig. 2.

Archaeal Type III CRISPR-Cas Systems

Four type III subtypes of CRISPR-Cas systems are known: III-A, III-B, III-C, and III-D among which III-A and III-D are also called Csm systems whereas III-B and III-C are called Cmr systems. While III-A systems dominate in bacteria, III-B systems are primarily found in archaea. To date only a few archaeal type III systems have been characterized, and these include Cmr systems of *P. furiosus*, *S. solfataricus*, and *S. islandicus*, Csm systems in *Thermococcus onnurineus* (III-A) and *S. solfataricus* (III-D) (Fig. 3).

The first characterized type III effector is the *P. furiosus* Cmr ribonucleoprotein complexes, and it specifically cleaves target RNA. By now, a number of III-B effectors have been characterized. These effector complexes contain six different subunits (Cmr1 through Cmr6) among which Cmr1, 4, and 6 are backbone subunits possessing the Cas7 fold; Cmr3 has the Cas5 structure; whereas Cmr2 and Cmr5 are the large and small subunits, respectively. Each system forms effector complexes of two distinct sizes: The larger one has the subunit stoichiometry of Cmr1₁2₂3₂4₄5₃6₁ carrying a crRNA of 46 or 45 nt while the smaller one has three Cmr4 and two Cmr5 subunits with a shorter crRNA (40 or 39 nt). The two complexes share the same general structure: Cmr4 and Cmr5 form the core backbone in which Cmr4 subunits are packed in a head-to-tail fashion. The backbone is capped with Cmr2 and Cmr3 at the head and Cmr6 and Cmr1 in the tail. The crRNA passes through the entire effector complex. Furthermore, Cmr2 and Cmr3 form a heterodimer that binds specifically to the 5'-handle of crRNA, whereas the tail subunits are loosely associated with the core helix. This arrangement in the ribonucleoprotein (binary effector) complex ensures the flexibility for the tail to mediate target RNA capture to yield the ternary effector complex that is active in multiplex interference activities, including target RNA cleavage, RNA-activated DNA cleavage, and synthesis of cOA second messenger.

Target RNA cleavage by archaeal type III-B effector complexes is featured with 6 nt periodicity as for other type III systems. When some of their structures are resolved, it becomes clear that the active sites of RNA cleavage are located on Cmr4. Once target RNAs are cleaved and released from the effector complex, the immune system returns to the binary status that is inactive in DNA interference.

Transcription-dependent DNA interference by the Cmr effector relies on two activities: The RNA-activated DNA cleavage and synthesis of cOAs, both of which are conducted in Cmr2, the large subunit belonging to the Cas10 family. The involved mechanism is summarized as follows: (1) transcription of a target DNA produces cognate target RNA (CTR) that will bind to the effector in the process of target RNA capture; (2) interaction between crRNA and its CTR (that shows mismatches between the 3'-anti-tag of target RNA and the 5'-handle of crRNA) induces conformation changes in the effector to activate the Cmr2 HD motif for DNA cleavage; (3) in the meantime, Cmr2 is also activated for the cyclase activity to produce cOA, which functions as a second messenger to activate Csx1/Csm6 family proteins for cellular RNA destruction to yield cell dormancy or cell death; and (4) target RNA cleavage and release restores the inactive form of Cmr effector complexes (Fig. 4).

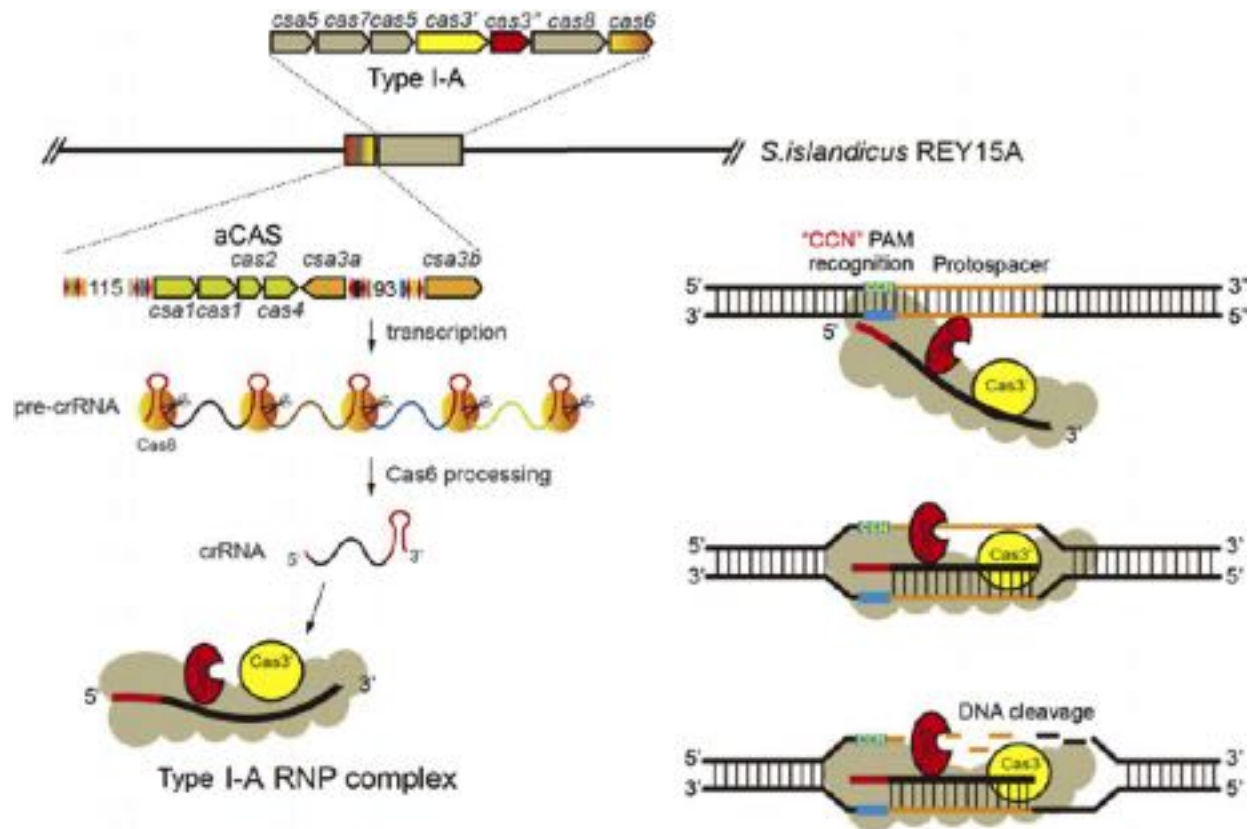


Fig. 2 Schematic of antiviral defense by the *S. islandicus* I-A system. Gene organization of I-A module (type I-A) and adaptation cas module (aCas) as well as two CRISPR loci (with 115 and 93 repeats, respectively) in *S. islandicus* are shown on the top. Transcription of CRISPR loci yields precursor CRISPR RNA (pre-crRNA) that is processed by Cas6 endonuclease to yield short crRNAs with a single spacer (crRNA). Expression of I-A cas genes give I-A Cas proteins including Cas3' nuclease and Cas3'' helicase. I-A Cas proteins and crRNA form ribonucleoprotein (type I-A RNP) complex. In the interference stage, the effector scans for target DNA site on ds-DNA, once a protospacer is found, the large subunit (Cas8) recognizes the CCN protospacer adjacent motif and authenticates the immune response. Cas3' helicase unwinds double-stranded DNA and Cas3'' nuclease cuts the target DNA sequence to yield target DNA silencing.

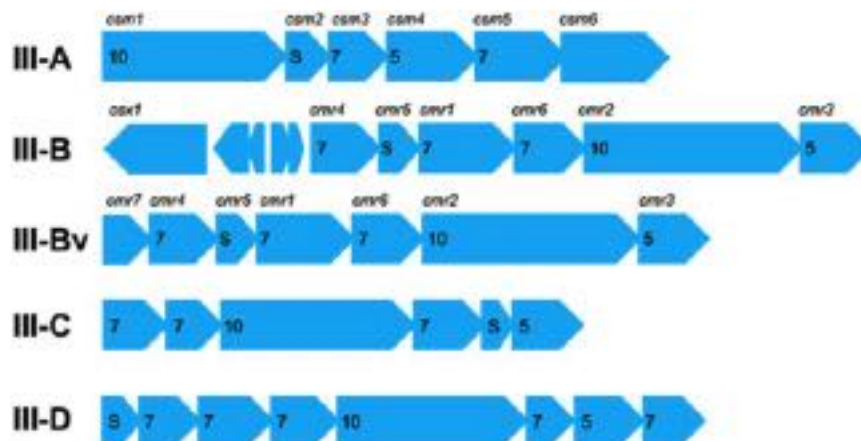


Fig. 3 Gene organization of type III subtypes of CRISPR-Cas systems. Four subtypes (III-A, -B, -C, -D) and one variant, III-Bv, are identified by analysis of gene synteny and sequence similarity of Cas proteins. Representatives of all subtypes except III-C have been characterized. III-B and III-C are also called Cmr systems and III-A and III-D, Csm systems. Annotations of individual subunits are given and their types of Cas proteins are also indicated: 10-Cas10, 7-Cas7, 5-Cas5, and S-small subunit. Cas10 is the type-specific protein and the most conserved Cas protein in this type of CRISPR immunity.

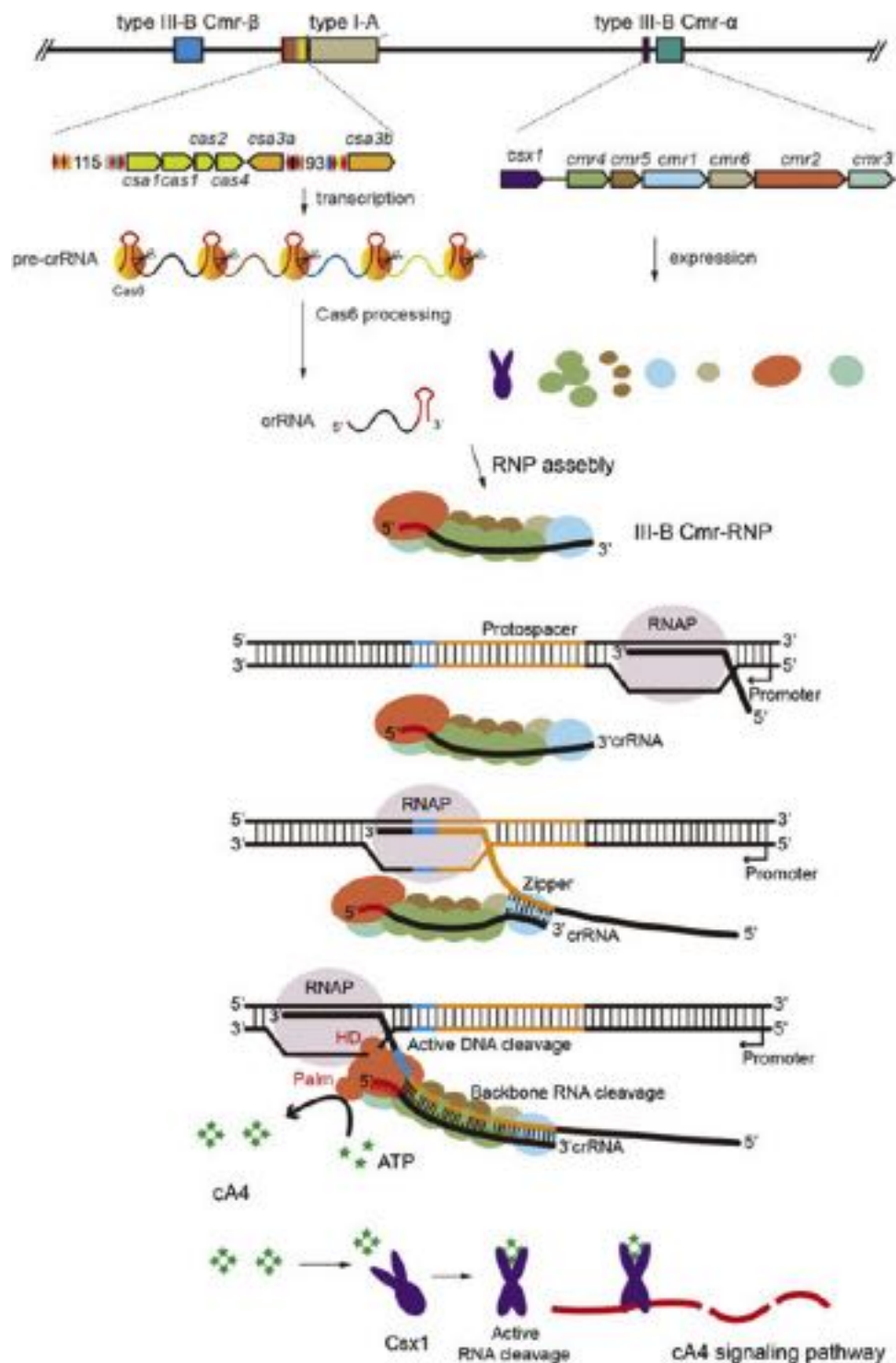


Fig. 4 Schematic of antiviral defense by the *S. islandicus* III-B Cmr-α system. Gene organization of adaptation cas module (aCas) and the III-B Cmr-α module as well as two CRISPR loci (with 115 and 93 repeats, respectively) in *S. islandicus* shown on the top. Transcription of CRISPR loci yields precursor CRISPR RNA (pre-crRNA) that is processed by Cas6 endonuclease to yield short crRNAs with a single spacer (crRNA). Expression of III-B *cmr* genes give 6 Cmr proteins. Cmr proteins and crRNA form ribonucleoprotein (III-B RNP) complex. Transcription of target gene produces mRNAs containing the cognate target RNA. Cmr1 mediates target RNA capture at the seed sequence region and base pairing between the crRNA and the target RNA yield major conformation changes to yield a ternary Cmr-α complex. Mismatches between the 5'-handle of crRNA and the 3'-anti-tag of target RNA authenticate the CRISPR immunity including RNA-activated DNA cleavage and synthesis of cOA second messenger. cOA binds to the CARF domain of Csx1 and activates its RNase from the HEPN domain to indiscriminately cleave viral and cellular RNAs in the cOA signal transduction pathway.

To date, all known type III systems possess both activities even though each activity is theoretically sufficient to mediate type III antiviral immunity. The current hypothesis is the cOA pathway is required for silencing a large amount of invading nucleic acids while the DNase is essential for final clearance of MGEs.

Among the three Cmr activities, the third activity, target RNA cleavage, mainly functions in the spatiotemporal control of the indiscriminate DNase and cOA synthesis. This mechanism ensures that the indiscriminate destruction of cellular DNA and RNA by type III immunity is only allowed to be active during a very brief time period. Prompt target RNA cleavage and cleavage product release readily inactivate Cmr effector. This provides an important mechanism to recover from cell dormancy or even cell death that would otherwise be induced by an uncontrolled type III immune response.

CRISPR arrays are also transcribed on the antisense strand, producing countertranscripts that show the full sequence complementarity with crRNAs including the 5'-handle, called noncognate target RNA. As a result, these target RNAs cannot trigger DNA cleavage and cOA synthesis, providing the mechanism of self-immunity avoidance.

Effective control of type III CRISPR-Cas immunity also requires to inactivate cOA second messenger. *S. solfataricus* code for ring nucleases that specifically degrade cOAs. As for Csx1 and Csm6, ring nucleases are also CARF proteins but they do not possess any conventional nuclease activity. Rather, they have a different CARF pocket that not only binds to cOA but also cleaves the messenger molecule in a metal-independent manner. In a Csx1 RNA cleavage assay, the ring nuclease effectively deactivates the Csx1 nuclease by competitively removing cA4 in the reaction.

Novel Archaeal CRISPR-Cas Systems

More recently, search for CRISPR-Cas systems in metagenomic databases reveals that archaea also code for class 2 CRISPR-Cas systems. These include two type II systems identified in nanoarchaea: One carrying *cas1*, *cas2*, and *cas4* for spacer acquisition with a *cas9*-like gene for interference as well as a hypervariable CRISPR array, and the other including 24 different *cas14* variant genes of three subgroups identified in uncultivated archaea. These are a unique group of archaeal organisms that form symbiosis with other organisms and are characterized by small cell size and genome size. Their type proteins, named Cas14, have 400–700 amino acids and therefore are much smaller than those of bacterial class 2 systems (950–1400 aa). *Trans*-acting CRISPR RNAs (tracrRNA) are also identified from transcriptome metagenomic data, and this rendered it possible to reconstitute Cas14 effector complexes in *E. coli* and study their antiviral immunity. Interestingly, Cas14 is capable of binding to ssDNA and mediate discriminate ssDNA cleavage. This novel type of CRISPR immunity is only present in archaea so far.

Anti-CRISPR Proteins of Archaeal Viruses

Viruses code for proteins that are able to counteract the CRISPR immunity, allowing viruses to proliferate in their hosts. These anti-CRISPR (*acr*) genes are widely present in bacterial and archaeal viruses, mediating resistance to immunity of CRISPR-Cas systems of types I, II, and V. In archaea, anti-CRISPR activity has been investigated for *Sulfolobus* viruses. Although remaining to be identified, the *Sulfolobus* SMV1 codes for Acr-IA proteins since it is resistant to the type I-A immunity in *S. islandicus* Rey15A. The first archaeal *acr* gene was identified in SIRV2 and SIRV3, the lytic rudiviruses. These viruses can infect *Sulfolobus islandicus* LAL14/1 despite the presence of functional CRISPR-Cas subtypes I-A, I-D, and III-B systems and spacers complementary to their genomes. Further studies reveal that the SIRV3 *gp02* gene encodes a protein (Acr-ID1) that interacts with the I-D Cas10 and inhibits the I-D immunity. Interestingly, Acr-ID1 has as many as 50 homologs in archaeal viral genomes. Possibly, archaeal viruses code for even more Acr-IA and Acr-IIIB proteins since these CRISPR-Cas systems are more prevalent in this prokaryotic domain.

Archaeal Innate Antiviral Systems

As in bacteria, innate immune systems in archaea include abortive infection, R–M systems, pAgo, and TA systems identified by comparative genomic studies, but most of these innate antiviral systems remain to be investigated. Nevertheless, a few representatives have been characterized, including the *Sulfolobus acidocaldarius* Sual R–M system, *Pyrococcus furiosus* Argonaute (PfAgo) and TA system, and the *H. jeotgali* DNA phosphorothioate modification and virus replication inhibition system. Here I briefly introduce the last three systems since the Sual R–M system functions as the well-known bacterial ones.

In eukaryotes, Agos function as a key enzyme in the RNAi pathway, and they target RNA transcripts using short 5'-phosphorylated RNA as guides. Their prokaryotic homologs, pAgos, are present in about 32% of archaea and 9% of bacteria. PfAgo shows DNA-guided DNA cleavage activity and the system reduces plasmid formation rate in *P. furiosus* by 30%–50%. Nevertheless, how archaeal Ago proteins function in antiviral defense remains to be investigated.

Archaea code for an arsenal of TA systems, for example, *Sulfolobus islandicus* Rey15A and LAL14-1 code for 21 or 22 TA systems, and many of them are homologs. This high homology has strongly hindered their functional characterization since their activity can only be revealed when all other homologous systems are inactivated in the same archaeal host. Nevertheless, some distantly related *P. furiosus* TA systems function as selection makers in *Pyrococcus yayanosii*. This represents the first example of genetic demonstration of archaeal TA systems that function in inducing programmed cell death. Therefore, archaeal TA systems function under the same principle as for their bacterial homologous systems.

A completely novel archaeal defense system was identified in *Haloterrigena jeotgali*. It consists of the genes homologous to the bacterial *dndCDEA* system that mediates phosphorothioate modification and the PbeABCD system that functions in halting virus propagation via inhibition of DNA replication. This unique combination allows the haloarchaeon to inhibit virus infection, and the defense system is conserved in several archaea.

Future Perspectives

Archaea and bacteria have devoted a large portion of their coding capacity for an arsenal of antiviral defense systems, including R–M systems, abortive infection, TA systems, archaeal Ago, DNA phosphorothioate modification, and virus replication inhibition systems, as well as the prokaryotic adaptive immunity encoded by CRISPR-Cas systems. These antiviral mechanisms constitute multiple layers of virus defense network, allowing archaea to effectively defend invasion of mobile genetic elements. Furthermore, many archaea carry multiple CRISPR-Cas systems that utilize very different immunity mechanisms to protect their hosts against invasion of nucleic acids. This raises important questions such as: How do these antiviral systems cooperate in the immune defense? Why are they all required in the arms race between microbes and their genetic elements? As it is now well established that genetic elements play important roles in horizontal gene transfer and evolution as well as in the maintenance of biodiversity on the Earth, these antiviral systems may contribute to fine-tuning of the arms race between archaea and their viruses, which in turn yields optimal horizontal gene transfer to increase the ecological fitness of archaeal organisms during evolution.

Further Readings

- Athukoralage, J.S., Rouillon, C., Graham, S., Gruschow, S., White, M.F., 2018. Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature* 562, 277–280.
- Deng, L., Garrett, R.A., Shah, S.A., Peng, X., She, Q., 2013. A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Molecular Microbiology* 87, 1088–1099.
- Elmore, J., Deighan, T., Westpheling, J., Terns, R.M., Terns, M.P., 2015. DNA targeting by the type I-G and type I-A CRISPR-Cas systems of *Pyrococcus furiosus*. *Nucleic Acids Research* 43, 10353–10363.
- Garrett, R.A., Shah, S.A., Erdmann, S., *et al.*, 2015. CRISPR-Cas adaptive immune systems of the sulfobolales: Unravelling their complexity and diversity. *Life* 5, 783–817.
- Gudbergsdottir, S., Deng, L., Chen, Z., *et al.*, 2011. Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Molecular Microbiology* 79, 35–49.
- He, F., Bhoobalan-Chitty, Y., Van, L.B., *et al.*, 2018. Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nature Microbiology* 3, 461–469.
- Koonin, E.V., Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology* 37, 67–78.
- Liu, T., Liu, Z., Ye, Q., *et al.*, 2017. Coupling transcriptional activation of CRISPR-Cas system and DNA repair genes by Csa3a in *Sulfolobus islandicus*. *Nucleic Acids Research* 45, 8978–8992.
- Maier, L.K., Stachler, A.E., Brendel, J., *et al.*, 2019. The nuts and bolts of the *Haloflex* CRISPR-Cas system I-B. *RNA Biology* 16, 469–480.
- Majumdar, S., Zhao, P., Pfister, N.T., *et al.*, 2015. Three CRISPR-Cas immune effector complexes coexist in *Pyrococcus furiosus*. *RNA* 21, 1147–1158.
- Terns, R.M., Terns, M.P., 2013. The RNA- and DNA-targeting CRISPR-Cas immune systems of *Pyrococcus furiosus*. *Biochemical Society Transactions* 41, 1416–1421.
- Zhang, J., Graham, S., Tello, A., Liu, H., White, M.F., 2016. Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Research* 44, 1789–1799.
- Zhang, Y., Lin, J., Feng, M., She, Q., 2018. Molecular mechanisms of III-B CRISPR-Cas systems in archaea. *Emerging Topics in Life Sciences* 2, 483–491.

Discovery of Archaeal Viruses in Hot Spring Environments Using Viral Metagenomics

Jennifer Wirth, Montana State University, Bozeman, MT, United States

Jacob H Munson-McGee, Montana State University, Bozeman, MT, United States and Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Mark J Young, Montana State University, Bozeman, MT, United States

© 2021 Elsevier Ltd. All rights reserved.

Introduction

High-temperature environments have intrigued life scientists ever since the first hyperthermophiles were discovered in the hot spring of Yellowstone National Park USA (YNP) in 1966. Since then the microbial communities of hot springs in YNP, as well as other high-temperature environments from around the world, have been extensively studied resulting in the discovery of archaeal and bacterial thermophilic (organisms with a growth optimum of 60–80°C) and hyperthermophilic (organisms with a growth optimum of > 80°C). Surprisingly, no hyperthermophilic eukaryotic species have been identified. While high-temperature environments can support both bacterial and archaeal microbial communities, high-temperature acidic environments (pH < 4) tend to be dominated by archaeal species.

Viruses are ubiquitous with cellular life. In any environment where cells are replicating, one is likely to find viruses associated with those cells. Not surprisingly, in archaeal dominated high-temperature environments, archaeal viruses are expected to be present. However, in part due to the limited number of archaeal species that have been cultured, we know of only a limited number of archaeal viruses from cultured-based studies. In order to overcome this limitation, virologists have increasingly turned to viral metagenomics, the deep sequencing of viral communities directly from environmental samples, to assess viral diversity and to assemble viral genomes. These studies have greatly expanded our knowledge of archaeal viruses in extreme environments, especially with regards to their diversity and unique adaptations for replication in high-temperature environments. There are presently 93 complete archaeal virus genomes deposited at NCBI (June 2019), and many more are expected in the coming years.

Regardless of the geographic location, it is the underlying geochemistry of the hot spring that determines the microbial and virus community structure. That is because these are chemolithotrophic dominated microbial communities which derive their energy from the oxidation of reduced inorganic compounds such as ammonia, ferrous iron, hydrogen sulfide, or hydrogen. Beyond temperature, sulfate levels and pH are major parameters that influence the microbial and virus community composition. As sulfide oxidation decreases pH, the resulting acidity may leach certain trace elements (Fe, Al, Co, Pb, Cd, V, and others) from rock and sediment through which geothermal water passes to reach the surface or from recirculating acid-sulfate water. Likely due to the energy constraints for microbial life, high-temperature low-pH hot springs typically support low cell densities (typically < 10⁶ cells/ml) and low free virus particle concentrations (typically < 10⁵ particles/ml). Cellular and viral metagenomic studies of these environments have revealed relatively simple microbial communities usually consisting of < 10 archaeal species and viral communities made up of ~ 100 different viral types. An additional benefit of these systems is the fact that viruses are the only microbial predator in these hot springs creating microbial systems where the role of viruses can be directly examined. These simplified cellular and viral communities provide a tractable natural system where host-virus interactions can be examined in detail.

Archaeal viruses remain enigmatic biological entities. They are the least characterized group of viruses as compared to viruses infecting Bacteria and Eukarya. Despite having only a limited number of characterized archaeal viruses, they possess a surprising diversity of virion morphologies and viral gene content. This is especially true for archaeal viruses present in hot springs. Archaeal viruses have been isolated from only two of the 18 proposed or recognized archaeal phyla. All known archaeal viruses from hot springs infect members of the Crenarchaeota, while archaeal viruses from high salt environments infect Euryarchaeota. This bias towards these two phyla is likely only a function of the environments in which archaeal viruses have been most searched for. It is likely that all phyla of Archaea support viruses. More recently, metagenomic studies have identified viruses predicted to infect other archaeal phyla but none of these viruses have been isolated to date. In addition to the two phyla with cultured viruses described above metagenomic studies have identified viruses predicted to infect four additional phyla: candidatus Bathyarchaeota, Nanoarchaeota, candidatus Nanohaloarchaeota and Thaumarchaeota.

Below we discuss the advantages and challenges of using metagenomics to identify and characterize archaeal viruses from high temperate environments. First, we review the practical considerations for collecting and processing viral samples for metagenomic analysis from hot spring environments. Next, we discuss the important variables to address in next generation sequencing (NGS) and sequence analysis that are unique to archaeal viral metagenomic studies. Finally, we review the major findings of applying metagenomic and other culture-independent tools in understanding archaeal virus community structure and function in hot springs ([Fig. 1](#)).

Sample Collection and Processing

Acidic terrestrial hot springs offer both unique advantages and challenges for investigators. The harsh chemical and physical environment selects for a small number of specific cell types, and thus their viruses. This selection can be leveraged to address

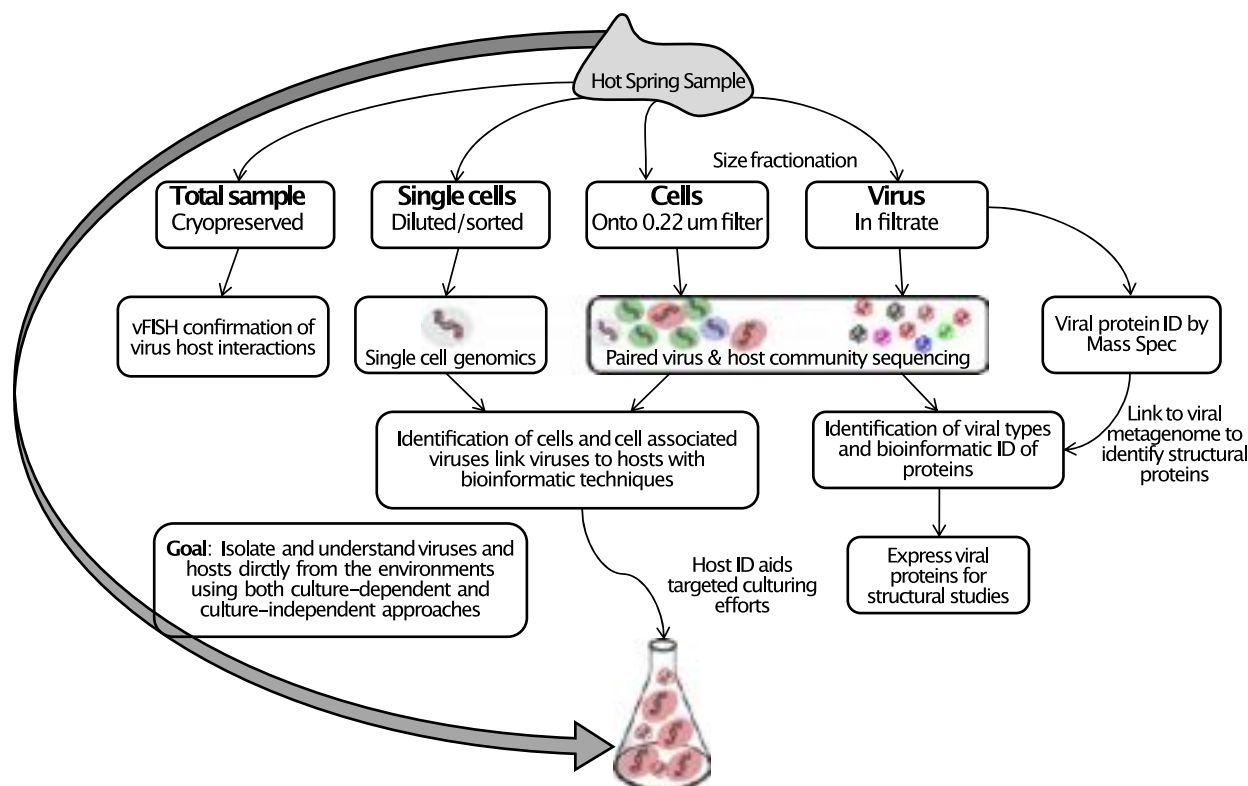


Fig. 1 Work flow of combined culture-dependent and culture-independent approaches to achieve virus characterization, including host identification, from hot spring environments.

complex ecological questions in a naturally simplified community. However, these same conditions present a number of difficulties ranging from personnel safety to technical challenges such as low host and virus densities and preventing acid hydrolysis of biomolecules.

Sampling terrestrial thermal fields requires preplanning and sampling by trained, experienced personnel with appropriate safety gear. It is essential to plan in advance the specific purpose for which the sample will be later utilized. For example, the sampling strategy for culturing new host and viruses is different from the sampling strategy required for the isolation of environmental nucleic acids. Sample site selection is also critical and dependent on the purpose of the sampling. Many thermal features are ephemeral, thus if long term temporal sampling on the time frame of years is required, appropriate site selection is critical. Furthermore, most thermal fields have a variety and fluctuating set of geochemical conditions. From a practical consideration, field measurements of temperature, pH and conductivity can be valuable in choosing an appropriate hot spring to sample. Hot springs, especially acidic hot springs, can be dangerous places to sample due to often turbulent super boiling acid waters and demineralization of the soils present in thermal fields. Walking in thermal areas is often more akin to traveling on glaciers, however in thermal areas the crevasses are boiling acid and steam. Sampling tools, including but not limited to extendable sampling poles, protective clothing, eyewear, and boots are important for safe sampling.

In planning a field sampling trip, one should consider what to sample and how to treat the sample until they are processed in a laboratory. Hot spring sediments often provide a rich source of cells while hot spring waters typically provide a better source of virus particles. Most acidic hot springs create environments where DNA that is not protected within a cell, or within a virus particle is rapidly hydrolyzed, therefore it is important to maintain cell and virus integrity during sampling. Furthermore, the low cellular and viral density present in most hot springs necessitates sampling large volumes, typically 1–100 L. If perfectly preserving a time-critical snapshot is necessary, then a small volume sample can be flash-frozen in liquid nitrogen. For large samples, transporting the samples at ambient temperature is ideal, especially if samples can be further processed within 24 h.

Isolation of cellular and viral fractions from hot spring samples is typically a multistep process. Clay particles are often present in samples and are removed by very low g centrifugation ($< 500 \times g$), which maintains the cells and viruses in solution. Separating the cells and intracellular viruses from free virus particles is accomplished by size filtration through a 0.45 μM and or 0.22 μM filter. The collection of cells on filters provides a convenient method to store and extract cellular DNA and RNA for subsequent NGS applications. Viruses present in filter flow-through can be concentrated by a diversity of methods including tangential flow filtration (TFF), FeCl_3 flocculation, polyethylene glycol precipitation, ultracentrifugation and size selection spin column filtration. TFF, FeCl_3 flocculation, and PEG precipitation methods lend themselves easily to the larger volume samples, while centrifugation-based methods are typically applicable to smaller volumes. After separation from cells, viral fractions can be further purified using

CsCl or CsSO₄ buoyant density gradients, sucrose gradients, or size exclusion chromatography. Of course, each step of any purification protocol has its own limitations, biases, and sample loss that the investigator must be aware of and plan accordingly.

Next Generation Sequencing of Samples

The availability of inexpensive, efficient NGS has greatly advanced archaeal virus discovery. Preparing nucleic acid samples from environmental virus and cellular samples for NGS requires care. It is essential that nucleic acid extraction methods minimize extraction bias while maintaining nucleic acid integrity. The minimum amount of nucleic acid required for NGS is dependent on estimated viral complexity and the DNA library construction protocol that is being utilized. Library preparation for Illumina Nano Miseq platforms requires only nanograms of input DNA while single molecule based long read technologies such as PacBio and Nanopore require high quality long DNA fragments at a higher starting concentration. It is also important to estimate the sequencing read depth required to adequately characterize the viral community, which is dependent on the estimated complexity of the viral community. A general guideline is to aim for 50–100× base coverage for every type of viral genome present. In order to determine the number of reads to generate, a researcher needs to take into account the size of the viral genome (generally 30–50k for high-temperature archaeal viruses), expected number of viruses in the community (typically 50–200 virus types), length of reads (150–300 for most NGS libraries), and desired coverage of the genome (50–100× base coverage). Hot springs with simple archaeal and viral communities can generally be well characterized with 5 million Illumina reads while a more complex community will obviously require more reads to fully sequence the viral community.

Post Sequence Analysis

A diversity of post-sequencing pipelines are available and new ones are constantly being developed for analysis of the viral community structure and the assembly of viral genomes. The overall approach requires the quality assessment of primary sequence data, sequence assembly into contigs, and post-assembly analysis. There are multiple assemblers that have been designed for NGS data. MetaSPAdes and IDBA-UD are two such assemblers. MetaSPAdes is user friendly, contains extensive documentation, and has been shown to perform well on data from a wide range of environments. MetaSPAdes also has the ability to create assemblies using both long and short reads, which generally results in higher quality assemblies and allows for the resolution between closely related viruses. Utilizing previously sequenced genomes as scaffolds can facilitate the assembly of genomes and differentiation of closely related viruses from the same or different environments. Evaluation of the quality of the assemblies is either performed by the assembler or by a stand-alone program such as QUAST. Assembled circular genomes provide confidence in the completeness of viral genomes.

As metagenomic technologies improve and metagenomic assembled viral genomes (MAVs) become more common, assessing the quality and completeness of these sequences has become more important. The Minimum Information for Uncultivated Viral Genomes (MIUViG) has been introduced to standardize the type and quality of data required to publish new MAVs (adapted from similar standards for Metagenome Assembled Microbial Genomes-MAGs). A checklist of mandatory metadata includes statistics about assembly quality, number of contigs, predicted genome type and structure, tools used for assembly and analysis and source dataset type.

One of the most challenging and exciting aspects archaeal viral metagenomic studies is gene annotation since a majority of the viral sequences have little to no similarities to sequences in the reference databases. Traditional BLASTn or BLASTx analysis typically results in only 10%–20% of genes having a significant match to known genes. More sensitive analysis such as Hidden Markov Modeling (HMM) and HHpred have demonstrated the ability to identify more viral genes but can be computationally expensive. A recent analysis of genes in almost 3000 prokaryotic viruses with HMM created ~10,000 profiles of conserved viral orthologous groups (pVOGs), along with the function of the original genes when known. This is an excellent resource for identifying known and unknown viral genes.

Network analysis based on either viral DNA sequence or their encoded proteins, is a relatively new approach to analyze viral metagenomes. These methods identify homologous sequences that are then grouped into viral “clusters” which represent closely related viral genotypes. Relatively small contigs retrieved from conventional assembly pipelines can be grouped into biologically meaningful clusters. This approach overcomes some of the challenges inherent in complex environmental samples where sample heterogeneity can limit assembly platforms that were designed for high quality deep sequencing of relatively homogenous samples. Protein clustering is another approach to reduce the complexity of metagenomic datasets and identify proteins that are shared between viruses and hot springs. However, protein clustering relies on the identification of viral genes, which can be challenging to computationally automate for archaeal viruses due to their genetic diversity from other viruses and microbes. As a result, these steps frequently require labor intensive manual curation of predicted genes greatly complicating and increasing the analysis time.

One final step in characterizing archaeal viruses is the taxonomic classification of newly discovered virus. Traditionally this has relied on the establishment of a virus infected host culture and description of the viral genome and particle morphology. Due to their morphological and genetic diversity the ~100 archaeal viruses comprise 11 recognized viral families (more than all 3000 + bacteriophage represent). However, with the rise of MIUViG culture independent classification is becoming increasingly popular.

A recently described tool vContact2 takes all predicted proteins in a viral genome and creates protein clusters that can be used to classify viral genomes at the genera, family, and order level. Further advancements and streamlining of viral characterization pipelines will greatly assist in the classification of new viral types.

A difficult but important question for any novel uncultivated virus is which cell type is the host? A number of *in silico* techniques can be combined to confidently address this question. In archaeal genomes the prevalence of CRISPR/Cas systems can greatly aid in host identification. Mapping CRISPR spacer sequences to the metagenomic assembled viral genome gives high confidence of a host-virus relationship. K-mer analysis and BLAST comparison of viral and cellular proteins can lend further support to host identification. Candidate host-virus pairs can then be validated with dual-label FISH probing of cells and viruses. This technique can also provide information about numbers of infected cells *in situ* and changes over time. Development of qPCR primers can also be used to follow viral and host population dynamics over time.

As obligate cellular parasites viruses and their nucleic acid are frequently found associated with cells, and it is worth examining cellular metagenomes for viral sequences. Most cellular genomes that have been examined to date have proviral sequences or prophages integrated into their genome. In addition, many cells are infected with actively replicating viruses and the sequencing of cell associated DNA will identify these viruses as well. Programs such as VirSorter and the archaeal virus specific MARVD identify likely viral sequences in cellular metagenomes by identification of viral hallmark genes, genome characteristics and other markers of viruses. Analysis of cellular genomes from YNP hot springs has revealed that 15%–20% of reads are likely of viral origin. While *de novo* identification of reads as viral is very difficult, the comparison to a viral database generated from purified viral particles either from the same hot spring or one with similar geochemical characteristics and microbial members greatly facilitates this process. After the viral database is created and curated, reads can be recruited onto the viral contigs in order to identify which viruses are associated with which cell type, linking viruses to cells, providing information about the host range of particular viruses and their relative abundance. Single cell genomics (SCG) can also now achieve this with much improved accuracy as intracellular viruses will be sequenced along with the host genome.

The difficulty of culturing certain cell types does not preclude further characterization of the virus using culture-independent approaches. The culture-independent methods outlined above that identify the host can aid in isolating the host and virus in culture. This is particularly true if the host is related to other cultured isolates. Virion structural proteins can often be identified from metagenomic sequence data. However, the inability to do so is no longer an insurmountable hurdle to structural studies. A purified virus fraction can be subjected to SDS PAGE and MS proteomic analysis to identify the predominant viral proteins in the fraction, which are typically the major capsid proteins. These structural proteins can then be mapped back to the predicted open reading frames in the viral metagenome to identify coat protein genes and analyzed further to give insights as to viral morphology. The identified viral structural protein genes can be cloned and expressed in a heterologous protein expression system. Expressed proteins can be used for antibody generation for use in ELISA or FISH assays. Purified proteins can also be used to solve the high-resolution structure using x-ray crystallography. Cryo electron microscopy and image reconstruction of purified virus directly from the environment can be used to generate a near atomic resolution structure of the entire virion, providing valuable insights into the virion architecture and function.

Combining viral metagenomics with other “omics” based approaches has proved to be a powerful technique to link viruses to their hosts and to ask important questions about the viral replication cycles. Paired viral and cellular metagenomes are allowing researchers to compare the viral populations within cells with that of the released extra cellular viral populations. Environmental viral metagenomic studies are being combined with environmental RNA transcriptomic studies to address questions of which virus are undergoing active replication. The identification of viral gene transcripts not only confirms that a virus is active in an environment it also can provide insight into which viral genes are the most abundantly transcribed in the system. The more abundant viral genes likely encode viral proteins that are needed in multiple copies in order to form virions (e.g., structural proteins) while less abundant viral transcripts may indicate regulatory genes that help to control the viral replication cycle. This begins to address one of the major problems with viral metagenomes in that the function of most genes was not possible to elucidate. Viral metabolomics aids in the identification of lipids that are part of the envelope in many archaeal viruses, and other components of the viral particle.

The ability to synthesize full length MAV DNA *in vitro* provides a pathway to establishing culture-based host-virus systems. Synthetic biology now allows us to express MAV DNA in suspected host cells. If a candidate host can be identified and is tractable to culturing, then a synthetic clone can potentially be transformed into a potential host and replicated, or launched off a plasmid. This may or may not lead to a productive infection, but can still provide insights into virus life cycle.

What Have we Learned?

The initial viral metagenomic studies focused on the identification of novel viruses and they helped to define the viral community structure. This era of “stamp collecting” revolutionized the study of archaeal viruses by providing insight into archaeal virus diversity and the genetic diversity of the thermophilic archaeal virosphere. For example, viral metagenomic analysis of acidic hot springs from YNP revealed that archaeal viruses dominate, that most of the assembled viruses genomes belong to uncultured viruses and that many are completely new viruses, unrelated to other known archaeal, bacterial or eukaryotic viruses. One acidic YNP hot spring was found to contain 110 virus types, of which 103 were previously unknown. Viral genomes assembled through viral metagenomics provide a roadmap and tools to track, purify and characterize these unknown viruses directly from environmental samples. For example, a

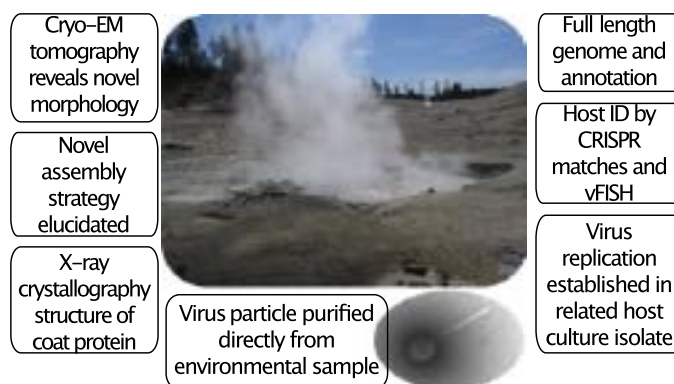


Fig. 2 *Acidianus* tailed spindle virus was purified and characterized primarily through culture independent methods.

recently described virus, *Acidianus* tailed spindle virus (ATSV) was originally isolated, its virion structurally characterized and its host identified all through culture-independent methods. This characterization guided subsequent successful efforts to culture the virus and its host (Fig. 2).

Archaeal MAVs have shown that they are genetically quite diverse. Nearly 90% of the genes present have no significant similarity to other genes in the NCBI nr database, and structural proteins were only identified in ~20% of viral types. Due to this high percentage of uncharacterized proteins, techniques other than database homology are needed to assign function to many MAV proteins. Two of the most common techniques to do this in MAVs are structural determination and genetic analysis. X-ray crystallography was used to solve the structure and determine the function of the archaeal viral protein A197 in *Sulfolobus* turreted icosahedral virus 1 (STIV1). While direct DNA sequence annotation of this protein provided little insight into its function, X-ray based structural analysis was able to identify it as a glycosyltransferase. Combined genetic and biochemical analysis has provided insight into the function of many archaeal viral proteins. This includes the discovery of a novel lysis mechanism encoded by STIV1 gene C92 which was determined to produce seven-sided pyramid shaped structures on the surface of the cell prior to lysis. Genetic studies have also showed that less than half the proteins encoded by *Sulfolobus* spindle shaped virus 1 (SSV-1) are necessary for productive infections in culture.

Despite the continued discovery of novel archaeal viral genomes, the field is rapidly moving past the cataloging of archaeal viral genomes to understanding MAVs in a broader biological context. Longitudinal sampling of YNP hot springs over time by viral metagenomics combined with targeted viral qPCR and vFISH analysis has provided insights into archaeal virus-host associations. Like all predators, viruses have been shown to fluctuate in response to their host abundance and to changes in their geochemical environment.

Temporal and longitudinal sampling allow for the comparison of MAVs across time and in different hot springs with different microbial communities and geochemical conditions. While the membership of the major archaeal virus community members can remain remarkably stable over a time frame of years, the relative abundance of a given viral type may vary considerably. Phylogenetic analysis of closely related viruses from the same hot spring over time can provide insight into viral evolution. Multiple STIV MAV's were recently compared from numerous YNP hot springs, revealing three distinct homologs of the C381 structural protein. The STIV1 and STIV2 C381-like structural proteins formed two distinct phylogenetic clusters while a third C381-like protein formed a third cluster. This suggests that a third type of STIV is present in YNP hot springs. A closer examination of the C381 clades reveals a geographic split of where each clade is predominately found. The first clade was primarily found in hot springs in the YNP Nymph Lake thermal region while the second clade was more abundant in the Crater Hills region. This local adaptation is likely in part due to the different microbial communities present in hot springs in these thermal areas.

Analysis of full length and near full length MAVs from geographically isolated (i.e., YNP, USA and Kamchatka, Russia) and geochemically distinct hot springs has identified different lifestyle strategies in different populations and different environments. For example, the lysogenic SSV strains found in the Multnovsky geyser basin (Kamchatka, Russia) were closely related to each other but distinct and genetically distant from the YNP SSV strains. Comparison across sites and over time has also enabled identification of core genes that are conserved across viral lineages through time and space as well as the flexible genome that appears to provide adaptation to the local conditions. Combining this analysis with host genome CRISPR spacer content has identified different host population immune strategies and viral countermeasures as well. For example, while all SSVs have spacer matches targeting their core genes, SSVs with fewer variable genes have fewer spacer matches to these accessory genes, while longer SSVs have many spacer matches to these accessory genes, suggesting a mechanism by which SSVs may avoid multiple host CRISPR spacer matches, and thus evade host immunity.

Combining environmental single cell genomics with population-based viral metagenomics is a powerful tool to link MAVs to their cellular hosts. A recent study took advantage of a multiyear sequencing effort of the viral community in a single hot spring in YNP and leveraged that knowledge to identify viruses in 250 single cell genomes. Despite low genome coverage of these single cells, viral sequences were identified in over 60% of the cells and sequences from multiple viruses were detected in most of the

cells. This indicates that virus-cell associations are common and leads to the speculation that most cells may be infected by viruses most of the time. Not only did this study identify high levels of virus-cell associations in the hot spring, it also identified viral hosts and host ranges for numerous viral types, and identified which viruses were actively infecting cells. While other single cell genome studies have identified viral sequences within cells they have lacked the thorough knowledge of the viral community that allowed the increased sensitivity of this study.

Looking Forward

Culture-independent approaches to investigating viruses directly in the environment are a valuable complement to traditional culture-based viral studies. This is especially true for studies of archaeal viruses from hot spring environments where there is a lack of suitable culturing systems. Viral metagenomics is rapidly progressing beyond being only a tool to describe archaeal viral diversity to a valued roadmap to address a wider range of questions. Among these questions are what viruses are infecting what cells and what is the host range of these viruses? How many cells are infected with which viruses and by how many viruses at any given moment? Overall, addressing these questions should provide insights into the bigger question of the role of viruses in driving microbial ecology and evolution in natural systems. There is still a role for the application of viral metagenomes to understand archaeal viral diversity, especially in unexplored natural environments, or in environments that contain archaeal phylum for which no viruses have yet to be discovered. However, there is an important need to leverage viral metagenomic datasets to enhance efforts to culture previously uncultured archaea and their viruses. Culture-based techniques will likely remain the gold standard to investigate details of viral processes for the foreseeable future. However, viral metagenomics are rapidly becoming initial experiments that allow researchers to refine their hypothesis and investigate biological processes in more depth.

As new sequencing technologies advance, impacts on archaeal virus discovery and characterization will further expand. The MinION sequencing platform is already allowing researchers to perform sequencing in the field. Concurrent with the improvement of current and development of new sequencing technologies is an ongoing need to generate bioinformatic tool kits and pipelines to process the large amounts of data that will be generated by these technologies. A major challenge is the development of rapid methodologies to annotate archaeal viral genes. This will likely require the combination of computational, genetic, biochemical and structural tools. The use of viral FISH and FACS tools directly on environmental samples will allow for the rapid quantification and isolation of infected cells, while the application of cryo electron microscopy techniques directly to environmental samples will provide new insights in archaeal virus structure, viral attachment and entry, replication and release from into cells.

It is likely that examining host-virus associations in their natural environments will further expand in the years to come. It is clear that many aspects of host-virus associations are evident only in the context of their natural environments and are lost when individual viruses and hosts are studied in the isolation of pure culture systems. This is particularly true when examining the ecological consequences of host-virus associations in terms of virus-virus, host-virus, and host-host competition. The interacting roles of host defense systems along with mechanisms to overcome host resistance are often only evident within the natural context of the environment. The field of environmental virology will continue to expand, embracing a much broader role of viruses in the environment from pathogens to commensal agents.

Single cell genomics and single cell transcriptomics will further our understanding of biochemistry and cell biology of archaeal virus infection. Currently single cell genomics is capable of identifying the presence of viruses in cells, but the details of viral replication are usually limited. Prokaryotic single cell transcriptomics is still in its infancy but the further advancement of the ability to sequence all of the RNA within a cell will prove crucial to identifying replicating viruses in natural environments. At the same time these techniques will provide unprecedented insight into how cells are defending against viral infection.

The field of high-temperature archaeal virology has made tremendous strides over the past two decades characterizing the morphological and genetic diversity of archaeal viruses, identifying virus-host pairs, surveying a wide range of thermal environments, and developing culture-independent techniques for viral characterization. However, the more we learn, the more we realize how much there is still left to discover. Future advances in archaeal virology over the next decade will likely require the isolation and in-depth characterization of many of these viruses and their hosts. A few of the areas that are current focuses of active research are the structural analysis, temporal fluctuations of viruses over time and the role archaeal viruses play in shaping microbial communities in natural environments.

Further Reading

- Bolduc, B., Wirth, J.F., Mazurie, A., Young, M.J., 2015. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. *The ISME Journal* 9 (10), 2162–2177. doi:10.1038/ismej.2015.28.
- Hartman, R., Munson-McGee, J.H., Young, M., Lawrence, C.M., 2019. Survey of high-resolution archaeal virus structures. *Current Opinion in Virology* 36, 74–83.
- Hochstein, R., Amenabar, M.J., Munson-McGee, J.H., Boyd, E.S., Young, M.J., 2016. Acidianus tailed spindle virus: A new archaeal large tailed spindle virus discovered by culture-independent methods. *Journal of Virology* 90 (7), 3458–3468. doi:10.1128/JVI.03098-15.
- Inskeep, W.P., Jay, Z.J., Tringe, S.G., Herrgård, M.J., Rusch, D.B., 2013. The YNP metagenome project: Environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. *Frontiers in Microbiology* 4, 67. doi:10.3389/fmicb.2013.00067.

- Jang, H.B., Bolduc, B., Zablocki, O., *et al.*, 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology* 37, 632–639.
- Liu, Y., Brandt, D., Ishino, S., *et al.*, 2019. New archaeal viruses discovered by metagenomic analysis of viral communities in enrichment cultures. *Environmental Microbiology* 21, 2002–2014.
- Munson-McGee, J.H., Peng, S., Dewerff, S., *et al.*, 2018. A virus or more in (nearly) every cell: Ubiquitous networks of virus–host interactions in extreme environments. *The ISME Journal*. 1–9. doi:10.1038/s41396-018-0071-7.
- Roux, S., Tournayre, J., Mahul, A., Debroas, D., Enault, F., 2014. Metavir 2: New tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics* 15, 76. doi:10.1186/1471-2105-15-76.
- Vik, D.R., Roux, S., Brum, J.R., *et al.*, 2017. Putative archaeal viruses from the mesopelagic ocean. *PeerJ* 5, e3428. doi:10.7717/peerj.3428.

Metagenomes of Archaeal Viruses in Hypersaline Environments

Fernando Santos, María D Ramos-Barbero, and Josefa Antón, University of Alicante, Alicante, Spain

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

GC Guanine + cytosine content in the DNA

TEM Transmission electron microscopy

VLP Virus-like particle

Glossary

Fosmid Cloning vector that allows the cloning of large fragments (between 30 and 40 Kb) and thus can be used for cloning complete viral genomes directly purified from environmental samples.

NGS (next generation sequencing) Massive parallel sequencing technologies that allow the sequencing of a mixture of DNA molecules without the need of previously separating them by cloning.

Introduction

An environment is considered hypersaline when its salt concentration is more than twice that of seawater. This type of systems includes solar salterns, salt lakes, soda lakes, saline soils and deep-sea brines, among others. From an anthropocentric point of view, hypersaline systems are extreme environments, although this view has recently been challenged considering thermodynamic aspects. The organisms thriving in hypersaline systems are known as halophiles or extreme halophiles, depending on their salt requirements. These organisms are adapted to live in the presence of salt and need it for their normal functioning.

At very high salt, the microbiota of hypersaline systems is normally dominated by Archaea, although other organisms are also present. These Archaea include extremely halophilic members of the *Euryarchaea* and the recently described and still uncultured *Nanohaloarchaea*. Within the first group we can find the high GC extremely halophilic archaea, well represented in culture collections (e.g., *Halorubrum*, *Halobacterium*, *Haloferax*, etc.), in addition to the low GC square archaeon *Haloquadratum walsbyi*, very widespread in many environments worldwide. This archaeon, although culturable, is very difficult to grow on plates.

One remarkable characteristic of hypersaline systems is the high number of VLPs (virus-like particles) that they harbor. In fact, brines are the water systems with the highest concentration of viruses reported to date (up to 10^{10} VLPs/ml) and, frequently, with one of the highest virus to cell ratios (up to 300, when the normal values in aquatic systems are around 10 or below). Furthermore, TEM studies show that hypersaline environments harbor a wide diversity of VLP morphologies. In fact, lemon-shaped (or spindle) viruses are considered to infect exclusively archaea, although so far only one has been isolated. Indeed, as discussed elsewhere in this Encyclopedia, archaeal viruses are underrepresented in culture collections, and this is also the case for virus infecting extremely halophilic archaea.

If a microbial community is dominated by Archaea, then very likely the viral assemblage is dominated by archaeal viruses. Thus, the analysis of viral metagenomes from hypersaline environments dominated by Archaea can be very useful to get information about archaeal viruses. This can be especially valuable for learning about viruses infecting the *Nanohaloarchaea* or *Haloquadratum*, for which there are no cultured virus representatives since, to date, the only available isolated haloarchaeal viruses infect members of the high GC *Euryarchaea*. Furthermore, hypersaline systems are good models for studying viral assemblages since at very high salt viruses are likely the main factor controlling the prokaryotic communities.

How can Archaeal Viruses be Studied Using Viral Metagenomes?

In addition to the general characterization of the viral metagenome (discussed elsewhere in this Encyclopedia), a key point in the analysis is to determine which viral genomes from the viral metagenomes correspond to viruses infecting Archaea. This means assigning to their archaeal hosts the virus genomes (or genomic fragments) reconstructed from the viral metagenome.

This assignment can be accomplished by using different approaches. The safest one, albeit the less applicable given the limitation of databases, is to compare viral genomes reconstructed from metagenomes with viral genomes coming from cultured viruses (or *in silico* reconstructed viral genomes from previous works). This is particularly limited for the Archaea since, as mentioned above, all the archaeal haloviruses cultured so far infect the high GC members of the hyperhalophilic Euryarchaea which are only a part of the archaeal community. This, obviously, imposes a bias in the metagenomics-based viral identification. In any case, even for this group of viruses the identification may not be straightforward.

An example of how the tools which are very successful to analyze viruses from many environments fail to recognize viruses from extremely halophilic Archaea has been recently provided by the isolation and characterization of viruses from the hypersaline Lake Retba infecting *Haloferax* and *Halorubrum* (both high GC *Euryarchaea*). Only one out of the four viral genomes described was

recognized as virus with the powerful tool Virsorter, which is frequently used to identify viral genomes retrieved from metagenomic datasets. This, as pointed out by the authors of the study, calls for an improvement of virus hallmark datasets.

Another approach to assign viral genomes from a metagenomic library to their host is based on the fact that with time viruses tend to ameliorate their nucleotide composition towards that of the host. This is reflected in the fact that the GC difference between phages and their microbial hosts is normally around 4%, and have similar oligonucleotide frequency and codon usage, although there are considerable deviations to this trend. In any case, in the absence of other information, GC content and oligonucleotide frequency analysis can be used to assign viruses to their host in hypersaline systems. In fact, these systems are well suited for this approach since frequently the GC profiles of the community are bimodal, with a lower peak at around 45%–50% corresponding to the low GC *Nanohaloarchaea* and *Haloquadratum*, and a higher peak around 60%–65%. Accordingly, viral assemblages also display this bimodal profile.

Finally, in order to assign viruses to hosts, we can also explore the imprint that viral infection leaves in prokaryotic genomes, more specifically in the defense CRISPR/Cas system, CRISPR standing for “clustered regularly interspaced short palindromic repeat”. When exposed to the invaders (like viruses), cells harboring a functional CRISPR/Cas system may become specifically immunized by the acquisition of new spacers (i.e., sequences intervening between the repetitions) that derive from the viral DNA sequences. Thus, knowing the spacers present in a given prokaryotic genome, one can look for the corresponding protospacers in the genome of the infecting viruses. As discussed below, this approach has been used in hypersaline environments to “fish” for virus infecting different extremely halophilic archaea, such as *Haloquadratum walsbyi*, high GC Euryarchaea or members of the *Nanohaloarchaea*.

What Viral Metagenomes are Available From Hypersaline Environments and What are Their General Characteristics?

Fig. 1 shows the hypersaline environments for which viral assemblages have been characterized through metagenomics. There are basically two types of viral metagenomes in the figure that are obtained either by NGS, by directly sequencing environmental DNA, or by cloning the DNA in vectors (plasmids or fosmids) prior to sequencing. In both cases, viral DNA has to be extracted from the extracellular virus fraction in a way that avoids (or at least minimizes) cellular DNA contamination. The use of fosmids, although

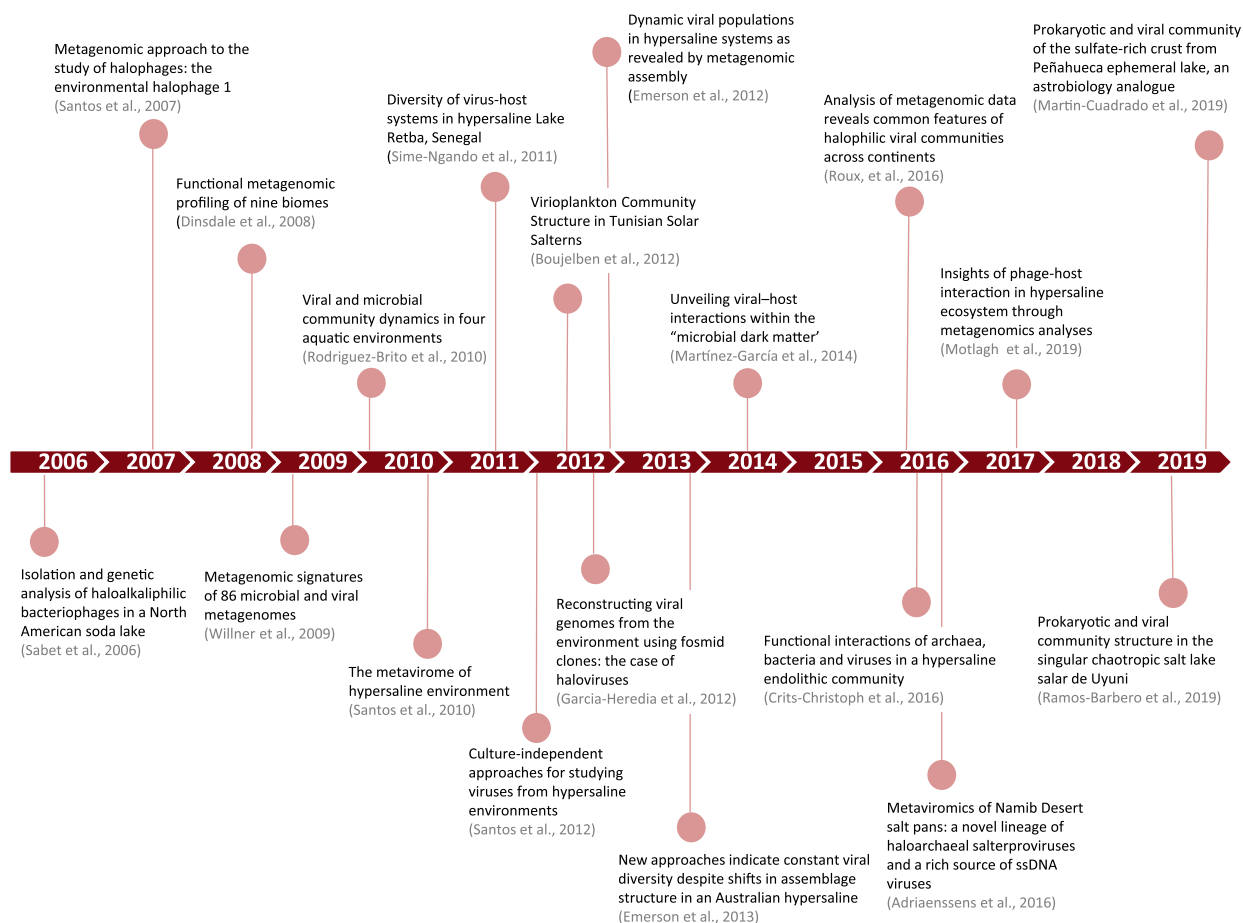


Fig. 1 Publications on metagenome analysis of viral assemblages from hypersaline environments.

not devoid of technical difficulties, allows the cloning of complete virus genomes directly from the environments, providing then “natural” contigs which may simplify considerably the downstream analysis. Furthermore, these cloned viral genomes are, indeed, real genotypes present in the natural sample, and not artifacts derived from the assembly of metagenomic short reads coming from different viral genomes. However, the most frequently used approach currently use is the direct sequencing of viral DNA.

Furthermore, there is still another metagenomic strategy to analyze virus genomes, which consists of sequencing DNA extracted from the complete community, including cells and the viruses attached or replicating inside of them. In this type of metagenomes, viral sequences can only be retrieved when they have some type of similarity with previously described viruses. However, the approach can still yield valuable results like, in the case of the Atacama salt crust (see Fig. 1), the identification of a virus tentatively infecting members of the Nanohaloarchaea.

Although there is not complete agreement among all the works published so far, the comparisons of data from Senegal, Spain, United States and Australia indicate that halophilic viral metagenomes share some common traits worldwide. Thus, in the same way that there is a marine-ness in marine viruses, there is a kind of hypersaline-ness nature in viral metagenomes from hypersaline systems, with genes that are never found in other systems. Furthermore, some related lineages have been found to have a worldwide distribution. However, there is also a considerable degree of specifically local genotypes, which seem to be more important for peculiar environments such as the chaotropic Salar de Uyuni, in Bolivia, which harbors viral communities very different from what was previously known. In any case, salinity is a very strong structuring factor for viral assemblages, as previously found also for their microbial counterparts. In addition, as a rule, the higher the salinity, the higher the contribution of archaeal viruses to the total community. Furthermore, there is a temporal variation of the analyzed hypersaline systems, which can complicate the obtention of general conclusions. Finally, as frequently found in other viral metagenomes, haloarchaeal metaviromes harbor a considerably high level of diversity, as indicated by the low number of genes with matches in databases. However, as in most viruses, haloviral genomes also carry what is known as “accessory metabolic genes” which have been acquired from previous hosts. To this regard, it is specially worth mentioning the detection of rhodopsin coding genes in some uncultured halovirus. The question remains open, however, if this light driven ion pumps play any role in the infection process.

Very frequently, genomes from metagenomes of hypersaline environments correspond to viruses from the Caudovirales (head and tail viruses), that can be identified based on the presence of hallmark genes like the terminases (which are very well represented in the available databases). However, these morphologies are not always dominant in all hypersaline environments, which frequently harbor a high number of lemon-shaped viruses. These are more difficult to identify based only on genomic traits given the low number of cultured representatives to compare with. However, even with these limitation, viral genomes from metagenomes have been assigned to the lemon shaped salterproviruses in samples from Namibia, Senegal, and Australia hypersaline systems.

Although they have to be taken with caution given the limitations of the approach, several studies indicate that very likely most haloarchaeal viruses do not undergo lysogenic cycles in nature, given the low proportion of integrases detected in the analyzed metaviromes. However, to which extent haloarchaeal viruses undergo chronic infection, which is typical of some archaeal viruses, is still an open question that metagenomics has not been able to answer.

Viral metagenomics has also been used to address the dynamics of viral communities, indicating that although stable at the time scale of days, they differ significantly at larger time scales.

In the following sections we will focus on the two groups of haloarchaea from which there are no cultures viruses and thus all our knowledge comes from metagenomic studies.

Viruses of *Haloquadratum Walsbyi*

So far, we have not been able to isolate viruses infecting the square archaeon *Haloquadratum* spp., although the study of plasmid and fosmid libraries has enabled the first insight into this group of viruses as described below. In addition, a metatranscriptomic study carried out with samples from a solar saltern in Santa Pola, Spain, indicated that this part of the viral assemblage was not especially active under natural conditions but increased its activity when the community was submitted to stress.

The “environmental halophage 1”, retrieved in 2007 from a fosmid library constructed with viral DNA from the CR30 crystallizer of “Bras del Port” salterns (Santa Pola, Spain), was the first nearly complete haloviral genome that was tentatively associated to *Haloquadratum* based on its low GC content (51.4%, slightly higher than the GC content of the square archaeon, around 48%). Although this nomenclature was kept for all the viral genomes cloned in fosmid libraries, the name is not correct since the term “phage” may not be appropriate to refer to archaeal viruses.

Soon later, a second fosmid library and a plasmid library were obtained. Assembly of the plasmid library produced “uncultured viral contigs” that were grouped in different clusters together with the genomes of some extremely halophilic prokaryotes, according to their GC contents and their dinucleotide frequencies. Cluster 4, including 25% of the contigs and associated to the genome of *Hqr. walsbyi*, was interpreted to contain viruses infecting low GC haloarchaea (only represented by *Haloquadratum* at that time since nanohaloarchaea had not been described yet, but with a fraction of sequences which could, indeed, correspond to nanohaloarchaeal viruses). A third fosmid library from the same crystallizer was published in 2012, reporting a group of 14 viral genomes (also named “environmental halophages”, eHPs) which were associated to *Haloquadratum* based on genetic features. These eHPs were also reported as “very abundant” in hypersaline systems by recruitment of metagenomic reads; indeed, eHPs can recruit up to 20% of the total nucleotides from a CR30 metavirome.

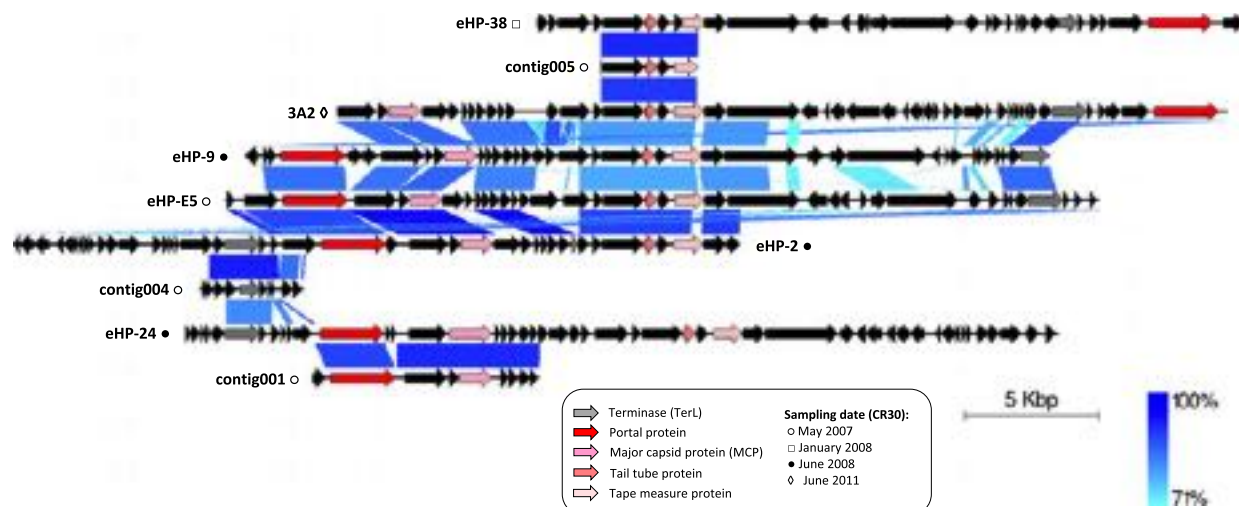


Fig. 2 Some representatives of the EHqrv group I. Six fosmid-derived sequences are represented by five “environmental halophages” (eHPs) and sequence 3A2, from a 2011 fosmid library (unpublished results). Contigs 001, 004 and 005 derive from the assembly of a previous viral metagenomic library cloned in plasmids. The blue color intensity is associated to BLASTN identity percentages. Reproduced from Garcia-Heredia, I., Martin-Cuadrado, A.B., Mojica, F.J., *et al.*, 2012. Reconstructing viral genomes from the environment using fosmid clones: The case of haloviruses. PLoS One 7, e33802.

Many of these sequences, obtained from crytallizer CR30 between May 2007 and November 2014, might be joined in a new group of “*Environmental Haloquadratum Viruses*”, or EHqrv (Fig. 2). EHqrv most likely includes a (micro)diverse population of *Haloquadratum* viruses *sensu stricto*, given the high degree of genetic relatedness among them and their host: an average low GC content, a similar codon usage, very close di/tetra-nucleotide frequencies, and up to 30 proto-spacers that match the three CRISPR systems of *Hqr. walsbyi* C23^T. Also, EHqrv sequences are characterized by a remarkable degree of synteny and share some conserved genes, such as virion structural proteins and the terminase large subunit (TerL), the genetic signature to place EHqrv within the order *Caudovirales*.

Viruses of the Nanohaloarchaea

Until 2012, all extremely halophilic archaea were included within the phylum Euryarchaeota. However, that year a new group of unusually small extremely halophilic Archaea was discovered in Australian solar salterns. This group turned out to be very widespread and rather abundant in hypersaline systems worldwide. It was later named as Nanohalorcheaota and included within the archaeal superphylum DPANN, formed by phyla belonging to the so-called “microbial dark matter”, since no culture representatives are available so far. Soon after, metagenomic analyses allowed the tentative assignment of new haloviruses to this newly discovered host.

The analysis of the fosmid libraries described above also allowed the description of a group of viruses tentatively assigned to the Nanohaloarchaea. This group would also belong to the Caudovirales, given the presence of the hallmark gene terminase, unless a more surprising explanation is waiting to be discovered. In addition to the *in silico* assignment of viruses to their nanohaloarchaea host discussed above for the Atacama samples, tetranucleotide frequency and codon usage, together with CRISPR spacer analysis and the comparison with the above mentioned fosmid libraries, allowed the detection of nanohalorachaeal assembled virus genomes also in salterns of Australia and Senegal.

Although extremely powerful, virus-host assignment based solely on metagenomics remain hypothetical until experimentally proven. A combination of viral metagenomics, single cell genomics and microarrays has allowed the unambiguous characterization of a virus infecting a nanohalorachaeon host. This virus (NHV-1) turned out to be closely related to one of the nanohaloarchaeal viruses retrieved from the fosmid libraries. NHV-1 harbored a terminase (in good agreement with previous results from tentative nanohaloviruses) and as well as a putative arsenical resistance repressor-like gene, which is very widespread in cellular and viral metagenomes from hypersaline environments worldwide.

Concluding Remarks

Metagenomics has been an extremely powerful tool to unveil the diversity of the haloarchaeal virosphere, although some questions remain still open. We anticipate that the use of long read sequencing technologies and new tools of viral metaproteomics will certainly help to solve them.

Further Reading

- Emerson, J.B., Thomas, B.C., Andrade, K., Heidelberg, K.B., Banfield, J.F., 2013. New approaches indicate constant viral diversity despite shifts in assemblage structure in an Australian hypersaline lake. *Applied and Environmental Microbiology* 79, 6755–6764.
- Garcia-Heredia, I., Martin-Cuadrado, A.B., Mojica, F.J., *et al.*, 2012. Reconstructing viral genomes from the environment using fosmid clones: The case of haloviruses. *PLoS One* 7, e33802.
- Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature Reviews Microbiology* 15, 724–739.
- Roux, S., Enault, F., Ravet, V., *et al.*, 2016. Analysis of metagenomic data reveals common features of halophilic viral communities across continents. *Environmental Microbiology* 18, 889–903.
- Santos, F., Yarza, P., Parro, V., *et al.*, 2012. Culture-independent approaches for studying viruses from hypersaline environments. *Applied and Environmental Microbiology* 76, 1635–1643.

Extreme Environments as a Model System to Study How Virus–Host Interactions Evolve Along the Symbiosis Continuum

Samantha J DeWerff and Rachel J Whitaker, University of Illinois at Urbana-Champaign, Urbana, IL, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Antagonism Symbiosis in which one partner benefits.

Chronic virus A virus that buds from the host cell and does not kill its host in horizontal transmission. Chronic viruses may also persist in host cells in integrated or episomal (not integrated) forms for vertical transmission.

CRISPR-Cas immunity Adaptive immune system found in bacteria and archaea that allows for sequence-specific targeting of foreign DNA elements to prevent infection.

Horizontal transmission Movement of a viral particle from an infected host through lysis or budding that leads to adsorption of the viral particle in a new, uninfected host.

Lytic virus A virus that lyses (kills) its host as it is transmitted to the next generation.

Metapopulation Spatially separated populations connected by migration.

Mutualism Symbiosis in which both partners benefit.

Symbiosis A close interaction between two biological entities.

Temperate virus A virus that can either integrate into the host chromosome as a latent lysogen for vertical transmission or kill its host through lysis for horizontal transmission.

Vertical transmission Movement of a virus from a mother cell to a daughter cell through genome replication.

Viral fitness A measure of the number of viral progeny that are able to successfully infect host cells.

Most studies of microbial viruses in ecology and evolution focus on antagonistic lytic viruses whose transmission requires cell lysis and death. However, not all viruses are lytic. Latent proviruses are found in genomes of all cellular forms of life and persistent chronic viruses are common in all three domains. Chronic and temperate lifestyles include parts of their life cycle that are associated with their hosts. These vertically transmitted genetic units rapidly confer new evolutionary traits to their hosts and change their eco-evolutionary interactions with other organisms. Such viral symbioses result in emergent properties arising from the merging of genetic units in a process called infection genomics that links ecological and evolutionary processes and timelines. Here we expand the view of viruses to include properties across the continuum of symbiosis from antagonist to mutualists.

Aspects of Viral Fitness

To expand our understanding of archaeal viruses and their symbiosis, we first consider the virus point of view. Like all evolving genetic units, evolution will select for increased viral fitness, the number of successful progeny transmitted to the next generation (Maynard-Smith, 1989; Hartl, 1988). Viruses are transmitted to the next generation either vertically, by replicating with their host, or horizontally, by successfully establishing a new infection in a new uninfected host. Selection will act on traits (antagonistic or mutualistic) that increase viral fitness (Bull *et al.*, 1991).

Most well-known archaeal viruses are lytic and kill their host when they are horizontally transmitted to the next generation. Well-studied examples in archaea include the viruses *Sulfolobus islandicus* rod-shaped virus (SIRV), *Thermoproteus tenax* virus 1 (TTV1), and most head-tail morphology viruses that have been shown to infect methanogens and halophiles (Prangishvili *et al.*, 2006, 2017; Luk *et al.*, 2014). In conditions where horizontal transmission results in higher viral fitness, selection may act to increase fitness, for example, by increasing the number of particles produced at lysis, decreasing the time to lysis, or increasing the infectivity of the particles. It is likely that increased fitness of the virus *decreases* host fitness. This is defined as antagonism on the symbiosis continuum (Bull *et al.*, 1991; Bull, 1994; Cressler *et al.*, 2016).

Evolutionary theory predicts that horizontal fitness increases when susceptible hosts in which viruses can establish new infections are abundant. Therefore, a naive population of susceptible hosts where resistance, immunity and infection are low, and the abundance of susceptible hosts is high, should favor viral antagonism for lytic viruses. In contrast, when susceptible hosts are rare fitness is not increased by horizontal transmission (there are no new hosts) and instead vertical transmission increases viral fitness (Turner *et al.*, 1998). In this case, horizontal transmission through lysis would be selected against and viruses would evolve to become less virulent (favoring, for example, delayed lysis) (Kerr *et al.*, 2006).

Just as ecological abundance of susceptible hosts impacts the evolution of virus-host interactions, these interactions also impact the ecology of host organisms in mixed communities. For example, lytic viruses impact the abundance of their specific hosts, disrupting the stability of complex microbial communities. Lytic viruses are often modeled as having Lotka-Volterra predator-prey dynamics, causing periodic oscillations of high and low abundance for the host and virus. This impacts competition and interactions between populations of microbes and others in community of viruses and hosts in which they exist in nature (Maslov and Sneppen, 2017). Selection for host resistance to lytic viruses can promote diversification among hosts evolving as resistance or resource specialists in an eco-evolutionary dynamic known as kill the winner (Thingstad, 2000; Weitz *et al.*, 2005).

In contrast to lytic viruses, some viruses of archaea are temperate, which have a lifecycle with two distinct modes of transmission. In the lytic mode, viruses will infect the host, rapidly replicate, and kill the host through cell lysis to be transmitted horizontally; however, there is also a lysogenic or latent phase in which the virus is integrated into the host chromosome and passed vertically from mother to daughter cell through host replication (Stewart and Levin, 1984). *Sulfolobus* turreted icosahedral virus (STITV) exhibits this lifestyle when infecting its host *Sulfolobus acidocaldarius*. While it has been shown to cause a lytic infection, natural *S. acidocaldarius* isolates from Yellowstone National Park have also been found with this virus integrated in the genome (Fu and Johnson, 2012; Brumfield *et al.*, 2009; Anderson *et al.*, 2017). In halophile *Natrinema* sp. J7–1, the sphaerolipoviruses SNJ1 and SNJ2 have also shown a temperate lifestyle; however instead of integrating into the host chromosome the viral genome remains as an extrachromosomal plasmid (Liu *et al.*, 2015; Wang *et al.*, 2016).

For temperate viruses, conditions which favor horizontal transmission, such as a high density of susceptible hosts, would theoretically favor the lytic cycle. Like purely lytic viruses, selection would favor traits that, while increasing the fitness of the infecting temperate virus, would decrease the fitness of the host. However, if there was a low density of susceptible hosts, then the lysogenic cycle may be favored. When vertical transmission is favored the fitness of the virus is aligned with the fitness of the host - if the host is successful then the virus is as well. Therefore, traits that are more benevolent or beneficial to the host are selected for and shift the interaction toward mutualism, where both partners benefit from the interaction (Bull *et al.*, 1991).

While temperate viruses show either horizontal or vertical transmission in the lytic-lysogenic cycle, chronic viruses have mixed-mode transmission by exhibiting both vertical and horizontal transmission at the same time (Lipsitch *et al.*, 1996; Weitz *et al.*, 2019). These viruses produce new viral progeny through non-lytic mechanisms that allow for the continued growth of the infected host. Examples of chronically infected viruses in archaea include *Sulfolobus* spindle-shaped viruses (SSVs) and the pleomorphic viruses infecting haloarchaea (Luk *et al.*, 2014; Quemain *et al.*, 2016; DeWerff *et al.*, 2020). In chronic infection there must be a balance between both modes of transmission. In conditions that favor horizontal transmission, selection for higher fitness traits may decrease the efficacy of the vertical transmission component. The opposite may also occur where vertical transmission is favored; selection for higher fitness may work against efficient horizontal transmission.

While lytic viruses can cause instability within microbial populations, temperate and chronic viruses can shape the ecology of their microbial host by providing fitness advantages in the context of their community. The long-term association between virus and host means that the viral fitness, the ability to succeed to a new generation, is aligned to host fitness, the ability of the host to grow and divide (Bull *et al.*, 1991). When virus and host fitness are aligned, viral traits that improve host fitness can then increase the fitness of the virus as well resulting in a viral mutualism. One example of this can be accessory metabolic genes that can give the host a growth advantage compared to the uninfected population, giving their host a competitive advantage (Mann *et al.*, 2003; Lindell *et al.*, 2004; Anantharaman *et al.*, 2014; Obeng *et al.*, 2016). Another possibility would be utilizing an interference strategy where viral infection gives direct competitive advantage in which an infected cell is antagonistic against an uninfected cell, for example, through use of a toxin (Boynton, 2019; DeWerff, 2020; Gama *et al.*, 2013). Finally, some non-lytic viruses confer protection to their hosts against lytic virus infection through superinfection exclusion (Bondy-Denomy *et al.*, 2016; Labrie *et al.*, 2010; Quemain *et al.*, 2013). Any of these contributions to align virus and host fitness are considered mutualisms and are predicted to have an outsized effect that may stabilize microbial populations against the predations of lytic viruses.

As previously discussed, low abundance of susceptible hosts is predicted to favor vertical transmission however, this is not the only ecological factor that would favor this transmission mode. Spatially structured populations, even when composed of high densities of microbes, may nonetheless have a limited effective population size that is accessible to a virus (Berngruber *et al.*, 2015). Paired with low host migration rates, this can create conditions in which there is little availability of susceptible hosts at a local scale even if the broader population is susceptible. With chronic viruses, vertical transmission would theoretically be favored in the highly structured areas, but production of new viral particles is still important to allow for possible viral migration in the population which may explain why horizontal transmission is still observed in similar natural populations.

Below we explore how environments and specific molecular mechanisms shape the interactions between microbes and viruses found along this symbiosis continuum. To do this we will be reviewing what has been learned from the unique host and viral ecology of extreme environments by focusing on the model system of acidic hot springs and how this might impact their interactions (Fig. 1).

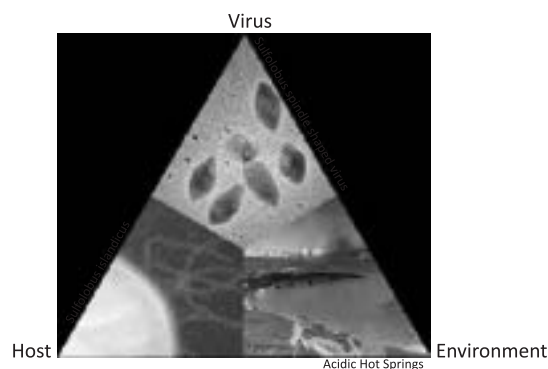


Fig. 1 In this review we explore how the virus, host, and environment together influence the broader ecology of the acidic hot spring system. TEM images courtesy of Elizabeth Rowland.

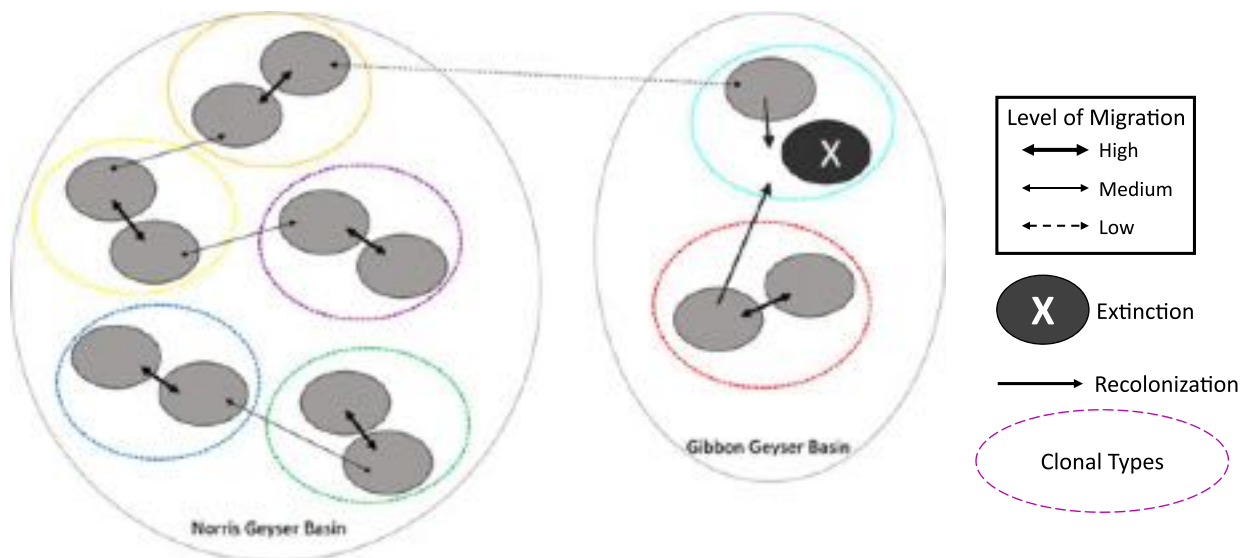


Fig. 2 Metapopulations of *Sulfolobus* hosts found in Yellowstone National Park. Each host spring makes up an island population of locally adapted hosts (and viruses). Migration and gene flow can occur, though at different scales based on distance. When extinction occurs recolonization can occur from within the population, like in the case of a bottleneck event, or competition from a new clonal type.

Environment Impact on Host-Virus Interaction

Extreme environments pose unique and highly structured landscapes for microbial life. One such environment is acidic hot springs, like those found in Yellowstone National Park, USA or Kamchatka, Russia where the model crenarchaea *Sulfolobus islandicus* can be found. *Sulfolobus islandicus* has been studied as a model not only for population structure but for its viruses as well (Anderson *et al.*, 2017; Whitaker *et al.*, 2003; Reno *et al.*, 2009; Bautista *et al.*, 2017; Munson-McGee *et al.*, 2018a). *Sulfolobus* species contain several adaptations that support growth in this environment: the enzyme reverse gyrase to maintain DNA structure at high temperatures, the proteinaceous surface layer that helps to maintain cell structure and osmotic stress, and an active proton pump that transports protons out of the cell to maintain intracellular pH (Zhang *et al.*, 2019; Forterre *et al.*, 1996; Gleißner *et al.*, 1997). The adaptations that allow for growth in the extreme environments also mean that there is low survival of *Sulfolobus* species outside of this environment, where at low temperatures DNA degradation limits the success of migration between hot springs (Hjort and Bernander, 1999). This low migration has led to metapopulations (subpopulations connected by limited migration) not only on a large spatial scale such as between continents, but on a local scale as well such as between hot springs (Fig. 2) (Anderson *et al.*, 2017; Whitaker *et al.*, 2003; Campbell *et al.*, 2017). Limited migration between populations can increase the overall genetic diversity and provide a potential for local adaptation between viruses and their archaeal hosts.

As discussed before this type of spatial structure would be hypothesized to favor either vertical transmission in temperate and chronic viruses or less virulent traits in lytic viruses, leading to local adaption between the virus and the host; however, in experimental studies this has not always been supported. When there is migration of the virus, even at low rates, it can provide enough genetic diversity to maintain the selection of antagonistic traits (Morgan *et al.*, 2005).

Like some other extreme environments, acidic hot springs are also dynamic with sporadic fluctuations in temperature and pH. In Yellowstone National Park, hot springs are subject to seasonal periods of snowmelt and rainfall in addition to geothermal excursions that rapidly change temperature and pH (Campbell *et al.*, 2017). This influx of cooler, pH neutral water in small hot springs can cause a flush or dilution of the host population. This instability can cause a significant selective sweep or bottleneck of the population that removes host diversity. In the hosts *Sulfolobus islandicus* and *Sulfolobus acidocaldarius*, population structure was studied through multi-locus sequence typing or whole genome sequencing respectively in hot springs from around Yellowstone. In populations from smaller hot springs, there was very limited diversity, with hosts belonging to one or two clonal types (Anderson *et al.*, 2017; Campbell *et al.*, 2017). Clonal types were highly associated with a single spring in time and place but were observed to change over time, suggesting an extinction-recolonization dynamic in some but not all springs. It is still not known whether the signatures of extinction are due to viral or environment extinction. Either way, this type of extinction-recolonization dynamic can counteract selection and local adaption and influence virus-host coevolution.

Finally, in acidic hot springs (as in other aquatic environments) the density of hosts is relatively low (Inskeep *et al.*, 2010). Low host density may select for higher vertical transmission or increased stability of the virion particle against prolonged environmental exposure. However, local areas of high density such as in biofilms are often observed in the environment. Biofilms are created when microbes attach to the surface and can be made of one or more microbial members. Microbes in a biofilm form three dimensional communities of relatively high density (Nadell *et al.*, 2016). While this can be a condition that allows for a virus to have access to a susceptible population, there are often extracellular structures that can limit the spread of a virus within the

biofilm itself (Simmons *et al.*, 2018; Chan and Abedon, 2015). While it has been shown that *Sulfolobus* species can form biofilms, it is unknown at this time if in nature they are found within biofilm structures or in the aquatic phase (Henche *et al.*, 2012a,b).

Viruses of Acidic Hot Springs

Viral metagenomic analysis has revealed a wealth of novel viral diversity in geothermal acidic hot springs. Deep sequencing of multiple time points from a single hot spring in Yellowstone resulted in 110 different viral clusters, of which only 7 clusters have representatives in a cultured system (Bolduc *et al.*, 2015). Within this large diversity of viruses, it would be expected that there also contains novel biology that has yet to be discovered. In the few characterized extremophilic viruses there has already been many unexpected discoveries. Some of these are adaptations to the environmental stressors placed on viral particles, such as *Sulfolobus islandicus* rod-shaped virus (SIRV), which contains A-form DNA in the viral particle (DiMaio *et al.*, 2015). A-form DNA is most commonly found within bacterial spores and helps to protect DNA from degradation. However, due to this low stability in the environment it is imperative for survival that viruses continue to infect a host. *Acidianus* two-tailed virus (ATV) leaves an infected cell through lysis as a lemon-shaped particle. Once it is in the extracellular environment it then goes through a maturation process that produces 2 long tails from the particle (Prangishvili, 2013; Prangishvili *et al.*, 2006). It is hypothesized that the extension of these tails helps to increase the chances of contact between the viral particle and the receptor of a new host.

In the host *Sulfolobus islandicus*, we will focus on just two viruses whose virus-host interactions can be found at the two ends of the symbiosis continuum. On the antagonism side of the symbiosis continuum is the *Sulfolobus islandicus* rod-shaped viruses (SIRVs). SIRVs are long stiff rod viruses with short tail fibers at one end (Prangishvili, 2013). These fibers attach to the pili of host cells and then travel down them to infect the host (Quemin *et al.*, 2013). SIRVs are lytic viruses and lyse the cell through the creation of proteinaceous pyramids in the cell membrane that open to release progeny viruses (Prangishvili and Quax, 2011). On the mutualism side we discuss the chronic *Sulfolobus* spindle-shaped viruses (SSV). These viruses are made up of lemon-shaped particle with tail fibers that often interact to form rosette-like structures of multiple viruses. The best characterized example, SSV1, was initially thought to follow a temperate lifecycle in which the virus can integrate into the genome in a lysogenic state and then, under UV induction, create new viral particles (Schleper *et al.*, 1992). However, further study into the egress mechanism showed that SSVs are not lytic but instead exit the cell through budding (Fig. 3) (Quemin *et al.*, 2016). The continued host growth of the during viral production, although with a fitness cost, suggests instead that these viruses are better characterized as chronic viruses (DeWerff *et al.*, 2020).

Like the host population, both SIRV and SSVs are also structured as a metapopulation. In initial studies using a single marker for viral presence it appeared that viruses are highly mobile (Snyder *et al.*, 2007); however, full genome sequences showed that in fact they are both structured like the host as metapopulations in Yellowstone National Park (Bautista *et al.*, 2017; Pauly *et al.*, 2019).

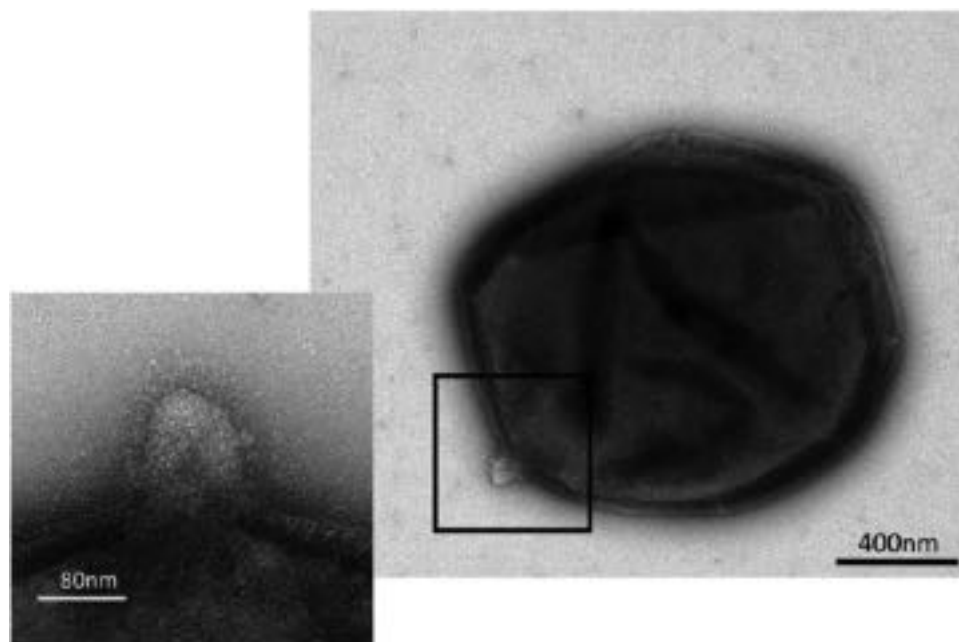


Fig. 3 Budding of the virus *Sulfolobus* spindle-shaped virus 9 from a chronically infected *Sulfolobus islandicus* strain. Chronically infected cells were visualized by transmission electron microscopy with negative staining. The ability of the virus to exit the cell through budding without killing the host allows for mixed-mode transmission with both vertical and horizontal transmission occurring at once. TEM images courtesy of Elizabeth Rowland.

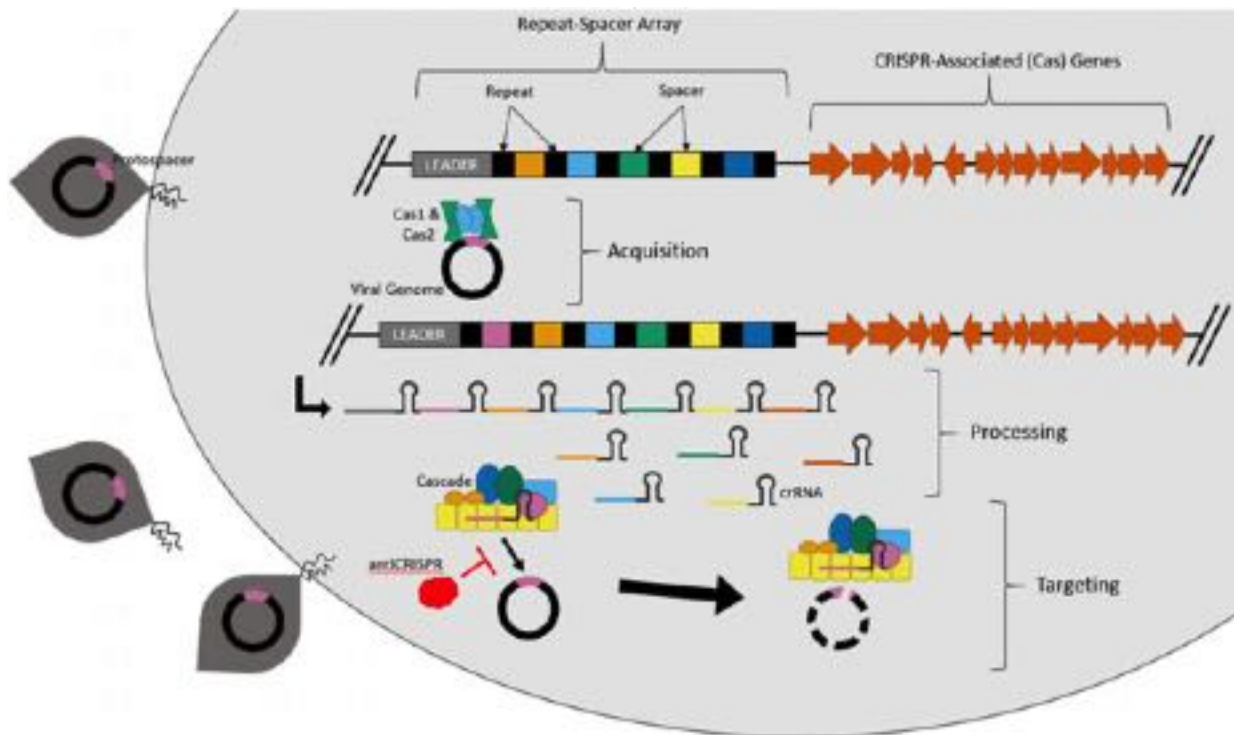


Fig. 4 Model of CRISPR-Cas immunity found in *Sulfolobus islandicus*. On initial infection and spacer can be acquired and integrated into the leader end of the repeat-spacer array. This array is then transcribed and processed into CRISPR RNAs (crRNA) that will interact with the cascade complex. Upon reinfection with a virus carrying the same protospacer, the crRNA-Cascade complex will bind to the viral genome and degrade it to prevent infection. Figure adapted from Maria A. Bautista.

Complexity of Host Virus-Interactions

Archaea have many host defenses against viral infection including surface resistance, restriction modification and CRISPR-Cas immunity (Koonin *et al.*, 2017; Rowland *et al.*, 2019). The molecular mechanism of defense shapes the co-evolutionary dynamics of viruses and hosts. Surface modifications that commonly evolve in the lab, such as deletion of the genes encoding surface appendages, are not seen in natural populations, suggesting this type of resistance is rare in nature because appendages are essential to survival (Rowland *et al.*, 2019). Instead studies of natural populations in thermoacidophilic archaea suggest CRISPR-Cas immunity is the dominant mechanism of defense (Held *et al.*, 2010, 2013).

Clustered regularly interspaced short palindromic repeats (CRISPRs) are repeat-spacer arrays within the host where the spacers can be added that match a viral or other foreign DNA element, called a protospacer (Fig. 4) (Barrangou *et al.*, 2007; Marraffini and Sontheimer, 2008; Amitai and Sorek, 2016). The protospacers that confer immunity are limited only by a short protospacer associated motif (PAM), meaning in any viral genome there are thousands of possible loci that are targeted by CRISPR-Cas immunity. Repeat-spacer elements are transcribed and processed to form a complex with other CRISPR associated (Cas) genes. If the spacer sequence in the protein-RNA complexes matches a viral genome on subsequent infection the Cas machinery will degrade the viral genome to confer immunity. Viruses evade targeting through escape mutations in their protospacer loci or PAM (Amitai and Sorek, 2016).

Theory predicts that the unique biology of the CRISPR-Cas defense system promotes diversification of CRISPR-Cas immunity because every host infection can result in a different spacer addition to the CRISPR repeat spacer array resulting in distributed immunity (Fig. 5) (Childs *et al.*, 2012, 2014; van Houte *et al.*, 2016). High distributed immunity effectively creates a population in which there are few susceptible hosts for a virus, since each single escape mutation gives a virus access only to a very small subset of the population. The nested structure of CRISPR-Cas immunity leads to alternating regimes of stable host populations and epidemic viral outbreaks (Pilosof *et al.*, 2020).

Acr proteins are viral proteins that work to block key components of CRISPR function or regulation. The first identified antiCRISPR in archaea was found in SIRV (He *et al.*, 2018). This protein was found to interact with the CRISPR cascade complex, and this interaction led to the inability of the complex to cleave viral DNA targets (Peng *et al.*, 2020). Recently another Acr protein was found encoded in SIRV that does not directly target components of the CRISPR-Cas system but instead encodes a nuclease that cleaves a signaling molecule responsible for upregulating the CRISPR-Cas system during active infection (Athukoralage *et al.*, 2020). As stated previously, in a population with high distributed immunity theory would suggest that this would limit the susceptible population size and favor vertical transmission or lead to viral extinction (Chabas *et al.*, 2018; Childs *et al.*, 2014). The

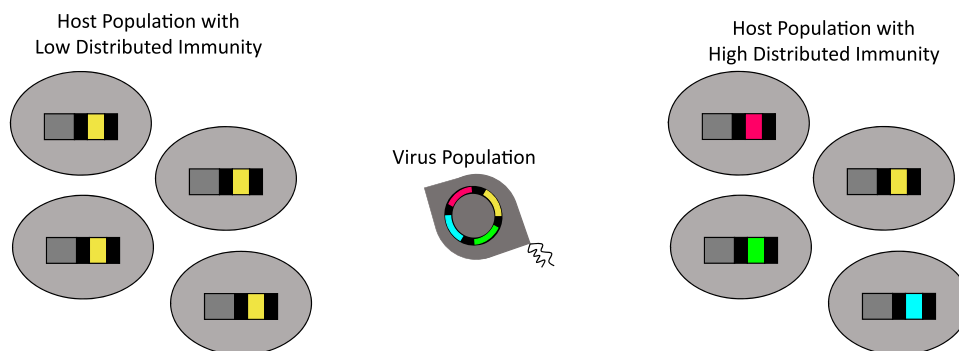


Fig. 5 Model of distributed immunity in host populations against a virus with 4 potential protospacers. If a host population only contains spacers to one of them then it has low distributed immunity. In this scenario the virus would only need an escape mutation in the yellow protospacer to allow infection of the whole population. If there is a diversity of CRISPR spacer targets in the population then it will have high distributed immunity. In the figure above, if the virus evolved an escape mutation in any one of the protospacer it would only be able to infect one member of the host population.

presence of Acr proteins changes the interaction between viruses and their host by removing key elements of CRISPR-Cas immunity and leading to a much larger susceptible population than what spacer sequences would predict. This would ultimately favor horizontal transmission and traits associated with higher virulence.

Since the CRISPR-Cas array retains a small segment of the viral DNA in the order that it was infected, allows for the tracking both the history of infection over space and time (Held *et al.*, 2010, 2013; Held and Whitaker, 2009). Signatures of previous virus infection recorded in the CRISPR-Cas repeat-spacer sequences contain matches to SIRV and SSV but suggest there are many other types undiscovered viruses that can infect *Sulfolobus* species (Anderson *et al.*, 2017; Held *et al.*, 2010; Shah *et al.*, 2009).

CRISPR-Cas arrays in both Yellowstone and Kamchatkan populations of *Sulfolobus* have been shown to be highly diversified (Pauly *et al.*, 2019; Held *et al.*, 2010). In addition, analysis of the content of CRISPR spacer arrays from *Sulfolobus islandicus* suggest that there is a level of recombination between CRISPR arrays as well as shared spacers that were acquired independently (Held *et al.*, 2010). The degree of CRISPR-Cas spacer targeting of different viruses reflect the viruses and their host interactions in natural populations. *Sulfolobus islandicus* isolated from Kamchatka, Russia highly target *Sulfolobus* spindle shaped viruses (SSVs) but lack targeting to *Sulfolobus islandicus* rod-shaped viruses (SIRVs); whereas host populations from Yellowstone National Park, USA highly target SIRVs and only have low immunity to SSVs (Pauly *et al.*, 2019). This matches what has been seen with viral isolation from these locations – while SSVs are quite commonly isolated from Kamchatka hot springs, to date there has yet to be a SIRV isolated. While in Yellowstone National Park hot springs SIRVs are quite common while SSVs, though present, are not the primary viruses isolated from these location.

How do Ecology and Evolution Shape Archaeal Viruses?

Putting the pieces together, can we predict how will viruses evolve along the symbiosis continuum in natural populations of archaea in extreme environments? In the context of a metapopulation the landscape of susceptible hosts may be highly varied. One hot spring may contain a population with highly distributed CRISPR-Cas immunity, while another one nearby that has recently experienced a bottleneck or extinction may lack diversified spacers that can target the population. If viral migration occurs between hot springs or a host bottleneck occurs within a hot spring there could be a viral boom period with a large susceptible population or a bust period where no susceptible hosts are present, effectively causing viral extinction unless escape mutations are present. It is within these two different scenarios that we will explore how selection will move the virus-host interactions along the symbiosis continuum.

During a viral boom period there is a large susceptible population that can be infected by the virus. When finding a new host is no longer the selective pressure, traits that are favored are ones that will increase the chance of horizontal transmission. This can be accomplished by increasing the number or infectivity of viral progeny or decreasing the time needed for a complete infection cycle. Viruses that have a decreased infection time will have a greater chance of infecting and spreading within the susceptible population compared to their longer infection cycle counterparts. However, these traits that are selected for are tied with an increase in virulence for lytic viruses (as described above). This increase in virulence is harmful to the host population and shifts the interaction toward the extremes of antagonism. Selecting for the quick use of susceptible hosts within the population for faster spread can lead to a tragedy of the commons in which the shared resource, in this case the susceptible hosts, is depleted by the viral population, which can ultimately lead to a bust period in which there is a low or absence of susceptible hosts (Kerr *et al.*, 2006). There is highly distributed immunity among populations of *Sulfolobus islandicus* against the virus SIRV in this environment; therefore it is perhaps not surprising that there is not one, but two types of antiCRISPRs present in SIRVs as these genes would be highly selected for to increase the fitness of the virus (Peng *et al.*, 2020).

In a viral bust period in which there are few susceptible hosts, selection will favor viruses that are more prudent with their hosts. For lytic viruses, this would select for lifestyle characteristics that increase the time of infection within the host allowing for production of

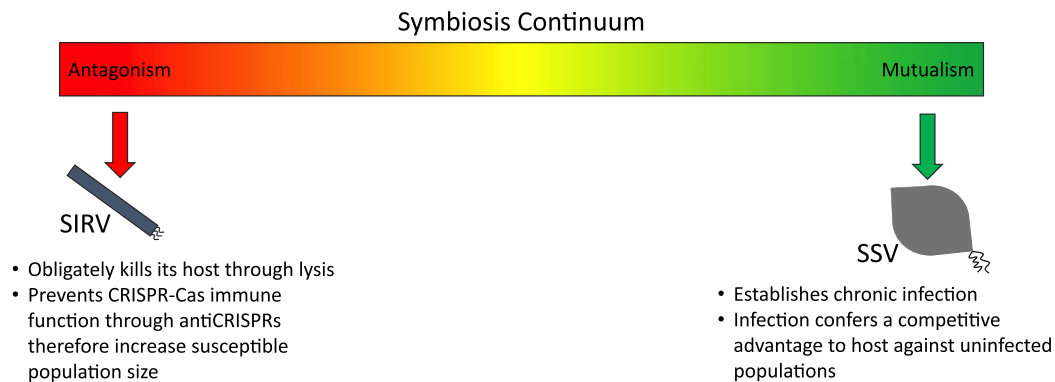


Fig. 6 Virus–host interactions fall along a symbiosis continuum. When a virus decreases host fitness it moves towards antagonism in which one partner benefits at the cost of the other. If both virus and host gain benefits from the interaction, then it is shifted toward mutualism.

either high quantity or quality of new viral progeny. However, this bust period would also strongly favor viruses that are not strictly lytic but instead follow a dual- or mixed-mode transmission such as with temperate or chronic infection. During a bust period, viruses that are already vertically associated with their hosts would be protected from environmental degradation of the genome or virion particle. This would select for mutualistic traits as discussed earlier that would give a benefit to the infected host (Bull *et al.*, 1991). This has been shown recently with SSV9 infection of the archaea *Sulfolobus islandicus*. Kamchatka populations exhibit high distributed CRISPR-Cas immunity to SSVs, which limits the size of the susceptible population (Pauly *et al.*, 2019). Though chronic infection comes at a cost in cell growth, DeWerff *et al.* showed that infected cells produce a toxin that kills uninfected and immune cells. This killing of competitors gives a benefit not only to the infected host cell through toxin immunity but the chronic virus as well (DeWerff *et al.*, 2020). The virus can ensure its continued vertical transmission by the toxin acting in a similar manner to plasmid addiction systems, and selects for successful horizontal transmission by only allowing the infected cells to survive. Given the structure of these metapopulations, a competitive ability would give a substantial benefit to the host. In its established hot spring, the ability to remove uninfected cells would suggest that the chronically infected cell becomes the dominant member. If migration to a new hot spring were to occur, the chronically infected host would be able invade an established uninfected population (even one with distributed CRISPR-Cas immunity) by killing established local hosts.

How do Virus Interactions Impact Ecology and Evolution?

So far, we have considered how both viruses, the lytic SIRV and chronic SSV, interact individually with their host and shape the ecology of the system. With SIRVs the ability to prevent CRISPR-Cas function allows for a larger susceptible population despite the high targeted distributed immunity that can be seen in these populations. This allows for selection of traits that are associated with high virulence in the host and shift the interaction toward antagonism (Fig. 6). On the opposite end of the spectrum, SSVs can establish chronic infection that allows for both vertical and horizontal transmission of the virus. The chronically infected host and virus benefit from this interaction through the competitive advantage that is conferred by production of a virally-encoded toxin that kills uninfected cells. This is in line with the selection of benevolent traits predicted to be associated with vertical transmission and a shift of the symbiotic relationship toward mutualism.

However, in Yellowstone National Park, both SSV and SIRV viruses can be found simultaneously in hot springs. How does the presence of both change the eco-evolutionary dynamic of each individually? It is necessary to go beyond thinking about one virus – one host interactions and consider those when multiple types of viruses are present. Infection of a chronically infected host with a lytic virus would be detrimental not only to the host but the chronic virus infecting this host cell as well.

The metapopulation nature of acidic hot springs and extinction-recolonization dynamic suggest that competition between host strains to invade new habits and quickly establish as the dominant member is a strong driver of selection for viruses that confer competitive advantages and remain with their hosts. Studying natural isolates from these populations as well as single cell genomics has shown that while most cells are infected by at least one virus in nature, chronic infection with SSV does not dominate each subpopulation with Yellowstone National Park (Pauly *et al.*, 2019; Munson-McGee *et al.*, 2018b). Taken together, the interactions discussed here must be more complicated and further research into novel archaeal virus mechanisms is needed to fully understand their eco-evolutionary impact.

Through the lens of extreme environments, the expanding view of viruses to include positive mutualistic impacts on their hosts provides dramatic insight into ecology and evolution in the microbial world especially on contemporary time-scales. In addition, study of viral symbiosis is uncovering the diversity of molecular interactions within cells that contribute to their physiology. In this way, the dual role of viruses as individuals and as genetic elements shows us that integrating molecular mechanisms with their eco-evolutionary impacts will play an increasingly important role in biology going forward.

Acknowledgments

We thank Elizabeth Rowland for help and inspiration on the initial drafts of this manuscript and TEM images. Work included discussed here was funded by NSF DEB #1656869.

References

- Amitai, G., Sorek, R., 2016. CRISPR–Cas adaptation: Insights into the mechanism of action. *Nature Reviews Microbiology* 14, 67–76.
- Anantharaman, K., Duhaime, M.B., Breier, J.A., *et al.*, 2014. Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344, 757–760.
- Anderson, R.E., Kouris, A., Seward, C.H., Campbell, K.M., Whitaker, R.J., 2017. Structured populations of *Sulfolobus acidocaldarius* with susceptibility to mobile genetic elements. *GBE* 9 (6), 1699–1710.
- Athukoralage, J.S., McMahon, S.A., Zhang, C., *et al.*, 2020. An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. *Nature* 577, 572–575.
- Barrangou, R., Fremaux, C., Deveau, H., *et al.*, 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712.
- Bautista, M.A., Black, J.A., Youngblut, N.D., Whitaker, R.J., 2017. Differentiation and structure in *sulfolobus islandicus* rod-shaped virus populations. *Viruses* 9.
- Berngruber, T.W., Lion, S., Gandon, S., 2015. Spatial structure, transmission modes and the evolution of viral exploitation strategies. *PLOS Pathogens* 11, e1004810.
- Bolduc, B., Wirth, J.F., Mazurie, A., Young, M.J., 2015. Viral assemblage composition in Yellowstone acidic hot springs assessed by network analysis. *The ISME Journal* 9, 2162–2177.
- Bondy-Denomy, J., Qian, J., Westra, E.R., *et al.*, 2016. Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal* 10, 2854–2866.
- Boynton, P.J., 2019. The ecology of killer yeasts: Interference competition in natural habitats. *Yeast* 36, 473–485.
- Brumfield, S.K., Ortmann, A.C., Ruigrok, V., *et al.*, 2009. Particle assembly and ultrastructural features associated with replication of the lytic archaeal virus *Sulfolobus turreted icosahedral virus*. *Journal of Virology* 83, 5964–5970.
- Bull, J.J., Molineux, I.J., Rice, W.R., 1991. Selection of benevolence in a host–parasite system. *Evolution* 45, 875–882.
- Bull, J.J., 1994. Virulence. *Evolution* 48, 1423–1437.
- Campbell, K.M., Kouris, A., England, W., *et al.*, 2017. *Sulfolobus islandicus* meta-populations in Yellowstone National Park hot springs. *Environmental Microbiology* 19 (6), 2334–2347.
- Chabas, H., Lion, S., Nicot, A., *et al.*, 2018. Evolutionary emergence of infectious diseases in heterogeneous host populations. *PLOS Biology* 16, e2006738.
- Chan, B.K., Abedon, S.T., 2015. Bacteriophages and their enzymes in biofilm control. *Current Pharmaceutical Design* 21, 85–99.
- Childs, L.M., Held, N.L., Young, M.J., Whitaker, R.J., Weitz, J.S., 2012. Multiscale model of CRISPR-induced coevolutionary dynamics: Diversification at the interface of lamarck and darwin. *Evolution* 66, 2015–2029.
- Childs, L.M., England, W.E., Young, M.J., Weitz, J.S., Whitaker, R.J., 2014. CRISPR-induced distributed immunity in microbial populations. *PLOS One* 9, e101710.
- Cressler, C.E., McLEOD, D.V., Rozins, C., Hoogen, J.V.D., Day, T., 2016. The adaptive evolution of virulence: A review of theoretical predictions and empirical tests. *Parasitology* 143, 915–930.
- DeWerrf, S.J., Bautista, M.A., Pauly, M., Zhang, C., Whitaker, R.J., 2020. Killer archaea: Virus-mediated antagonism to CRISPR-immune populations results in emergent virus–host mutualism. *MBio* 11 (2), (00404-20).
- DiMaio, F., Yu, X., Rensen, E., *et al.*, 2015. A virus that infects a hyperthermophile encapsidates A-form DNA. *Science* 348, 914–917.
- Forster, P., Bergerat, A., Lopex-Garcia, P., 1996. The unique DNA topology and DNA topoisomerases of hyperthermophilic archaea. *FEMS Microbiology Reviews* 18, 237–248.
- Fu, C., Johnson, J.E., 2012. Structure and cell biology of archaeal virus STIV. *Current Opinion in Virology* 2, 122–127.
- Gama, J.A., Reis, A.M., Domingues, I., *et al.*, 2013. Temperate bacterial viruses as double-edged swords in bacterial warfare. *PLOS One* 8.
- Gleißner, M., Kaiser, U., Antonopoulos, E., Schäfer, G., 1997. The Archaeal SoxABCD complex is a proton pump in *Sulfolobus acidocaldarius*. *Journal of Biological Chemistry* 272, 8417–8426.
- Hartl, D.L., 1988. *A Primer of Population Genetics*. Sunderland, MA: Sinauer Associates, Inc.
- He, F., Bhoobalan-Chitty, Y., Van, L.B., *et al.*, 2018. Anti-CRISPR proteins encoded by archaeal lytic viruses inhibit subtype I-D immunity. *Nature Microbiology* 3, 461–469.
- Held, N.L., Herrera, A., Cadillo-Quiroz, H., Whitaker, R.J., 2010. CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLOS One* 5, e12988.
- Held, N.L., Herrera, A., Whitaker, R.J., 2013. Reassortment of CRISPR repeat-spacer loci in *Sulfolobus islandicus*. *Environmental Microbiology* 15, 3065–3076.
- Held, N.L., Whitaker, R.J., 2009. Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environmental Microbiology* 11, 457–466.
- Henche, A.-L., Ghosh, A., Yu, X., *et al.*, 2012b. Structure and function of the adhesive type IV pilus of *Sulfolobus acidocaldarius*. *Environmental Microbiology* 14, 3188–3202.
- Henche, A.-L., Koerdt, A., Ghosh, A., Albers, S.-V., 2012a. Influence of cell surface structures on crenarchaeal biofilm formation using a thermostable green fluorescent protein. *Environmental Microbiology* 14, 779–793.
- Hjort, K., Bernander, R., 1999. Changes in cell size and DNA content in *sulfolobus* cultures during dilution and temperature shift experiments. *Journal of Bacteriology* 181, 5669–5675.
- Inskeep, W.P., Rusch, D.B., Jay, Z.J., *et al.*, 2010. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLOS One* 5, e9773.
- Kerr, B., Neuhauser, C., Bohannan, B.J.M., Dean, A.M., 2006. Local migration promotes competitive restraint in a host–pathogen “tragedy of the commons”. *Nature* 442 (7098), 75–78.
- Koonin, E.V., Makarova, K.S., Wolf, Y.I., 2017. Evolutionary genomics of defense systems in archaea and bacteria. *Annual Review of Microbiology* 71, 233–261.
- Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology* 8, 317–327.
- Lindell, D., Sullivan, M.B., Johnson, Z.I., *et al.*, 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proceedings of the National Academy of Sciences of the United States of America* 101, 11013–11018.
- Lipsitch, M., Siller, S., Nowak, M.A., 1996. The evolution of virulence in pathogens with vertical and horizontal transmission. *Evolution* 50, 1729–1741.
- Liu, Y., Wang, J., Liu, Y., *et al.*, 2015. Identification and characterization of SNJ2, the first temperate pleolipovirus integrating into the genome of the SNJ1-lysogenic archaeal strain. *Molecular Microbiology* 98, 1002–1020.
- Luk, A.W.S., Williams, T.J., Erdmann, S., Papke, R.T., 2014. Viruses of Haloarchaea. *Life* 4, 681–715.
- Mann, N.H., Cook, A., Millard, A., Bailey, S., Clokie, M., 2003. Bacterial photosynthesis genes in a virus. *Nature* 424, 741. (6950).
- Marraffini, L.A., Sonthheimer, E.J., 2008. CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science* 322, 1843–1845.
- Maslov, S., Snieppen, K., 2017. Population cycles and species diversity in dynamic Kill-the-Winner model of microbial ecosystems. *Scientific Reports* 7, 39642.
- Maynard-Smith, J., 1989. *Evolutionary genetics*. Oxford: Oxford University Press.
- Morgan, A.D., Gandon, S., Buckling, A., 2005. The effect of migration on local adaptation in a coevolving host–parasite system. *Nature* 437 (7056), 253–256.
- Munson-McGee, J.H., Peng, S., Dewerrf, S., *et al.*, 2018a. A virus or more in (nearly) every cell: Ubiquitous networks of virus–host interactions in extreme environments. *The ISME Journal* 12, 1706–1714.
- Munson-McGee, J.H., Snyder, J.C., Young, M.J., 2018b. Archaeal viruses from high-temperature environments. *Genes* 9.

- Nadell, C.D., Drescher, K., Foster, K.R., 2016. Spatial structure, cooperation and competition in biofilms. *Nature Reviews Microbiology* 14, 589–600.
- Obeng, N., Pratama, A.A., van Elsas, J.D., 2016. The significance of mutualistic phages for bacterial ecology and evolution. *Trends in Microbiology* 24, 440–449.
- Pauly, M.D., Bautista, M.A., Black, J.A., Whitaker, R.J., 2019. Diversified local CRISPR-Cas immunity to viruses of *Sulfolobus islandicus*. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374. (20180093).
- Peng, X., Mayo-Muñoz, D., Bhoobalan-Chitty, Y., Martínez-Álvarez, L., 2020. Anti-CRISPR proteins in archaea. *Trends in Microbiology* 28, 913–921.
- Pilosof, S., Alcalá-Corona, S.A., Wang, T., *et al.*, 2020. The network structure and eco-evolutionary dynamics of CRISPR-induced immune diversification. *Nature Ecology & Evolution* 4, 1650–1660.
- Prangishvili, D., Bamford, D.H., Forterre, P., *et al.*, 2017. The enigmatic archaeal virosphere. *Nature Reviews Microbiology* 15, 724–739.
- Prangishvili, D., Forterre, P., Garrett, R.A., 2006. Viruses of the archaea: A unifying view. *Nature Reviews Microbiology* 4, 837–848.
- Prangishvili, D., 2013. The wonderful world of archaeal viruses. *Annual Review of Microbiology* 67, 565–585.
- Prangishvili, D., Quax, T.E., 2011. Exceptional virion release mechanism: One more surprise from archaeal viruses. *Current Opinion in Microbiology* 14, 315–320.
- Prangishvili, D., Vestergaard, G., Häring, M., *et al.*, 2006. Structural and genomic properties of the hyperthermophilic archaeal virus ATV with an extracellular stage of the reproductive cycle. *Journal of Molecular Biology* 359, 1203–1216.
- Quemin, E.R.J., Chlanda, P., Sachse, M., *et al.*, 2016. Eukaryotic-like virus budding in archaea. *mBio* 7, e01439-16.
- Quemin, E.R.J., Lucas, S., Daum, B., *et al.*, 2013. First Insights into the entry process of hyperthermophilic archaeal viruses. *Journal of Virology* 87, 13379–13385.
- Reno, M.L., Held, N.L., Fields, C.J., Burke, P.V., Whitaker, R.J., 2009. Biogeography of the *Sulfolobus islandicus* pan-genome. *Proceedings of the National Academy of Sciences of the United States of America* 106, 8605–8610.
- Rowland, E.F., Bautista, M.A., Zhang, C., Whitaker, R.J., 2019. Surface resistance to SSVs and SIRVs in pilin deletions of *Sulfolobus islandicus*. *Molecular Microbiology* 113 (4).
- Schleper, C., Kubo, K., Zillig, W., 1992. The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: demonstration of infectivity and of transfection with viral DNA. *Proceedings of the National Academy of Sciences of the United States of America* 89, 7645.
- Shah, S.A., Hansen, N.R., Garrett, R.A., 2009. Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism. *Biochemical Society Transactions* 37, 23–28.
- Simmons, M., Drescher, K., Nadell, C.D., Bucci, V., 2018. Phage mobility is a core determinant of phage–bacteria coexistence in biofilms. 2. *The ISME Journal* 12, 531–543.
- Snyder, J.C., Wiedenheft, B., Lavin, M., *et al.*, 2007. Virus Movement Maintains Local Virus Population Diversity. *Proceedings of the National Academy of Sciences* 104 (48), 19102–19107. <https://doi.org/10.1073/pnas.0709445104>.
- Stewart, F.M., Levin, B.R., 1984. The population biology of bacterial viruses: Why be temperate. *Theoretical Population Biology* 26, 93–117.
- Thingstad, T.F., 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* 45, 1320–1328.
- Turner, P.E., Cooper, V.S., Lenski, R.E., 1998. Tradeoff between horizontal and vertical modes of transmission in bacterial plasmids. *Evolution* 52, 315–329.
- van Houte, S., Ekroth, A.K.E., Broniewski, J.M., *et al.*, 2016. The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532, 385–388.
- Wang, Y., Sima, L., Lv, J., *et al.*, 2016. Identification, characterization, and application of the replicon region of the halophilic temperate sphaerolipovirus SNJ1. *Journal of Bacteriology* 198, 13.
- Weitz, J.S., Hartman, H., Levin, S.A., 2005. Coevolutionary arms races between bacteria and bacteriophage. *PNAS* 102, 9535–9540.
- Weitz, J.S., Li, G., Gulbudak, H., Cortez, M.H., Whitaker, R.J., 2019. Viral invasion fitness across a continuum from lysis to latency. *Virus Evolution* 5 (1), vez006.
- Whitaker, R.J., Grogan, D.W., Taylor, J.W., 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301, 976.
- Zhang, C., Wipfler, R.L., Li, Y., *et al.*, 2019. Cell structure changes in the hyperthermophilic crenarchaeon *Sulfolobus islandicus* lacking the S-Layer. *mBio* 10.

Further Reading

- Bao, X., Roossinck, M.J., 2013. A life history view of mutualistic viral symbioses: Quantity or quality for cooperation? *Current Opinion in Microbiology* 16, 514–518.
- Ewald, P.W., 1987. Transmission modes and evolution of the parasitism-mutualism continuum. *Annals of the New York Academy of Sciences* 503, 295–306.
- Roossinck, M.J., 2011. The good viruses: Viral mutualistic symbioses. *Nature Reviews Microbiology* 9, 99–108.

FUNGAL VIRUSES

An Introduction to Fungal Viruses[☆]

Nobuhiro Suzuki, Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S.A. Ghabrial, N. Suzuki, Fungal Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd, 2008, doi:10.1016/B978-012374410-4.00563-X.

Glossary

Hyphal anastomosis The union of a hypha with another resulting in cytoplasmic exchange.

Hypovirulence Attenuated fungal virulence mediated by virus infection, mitochondrial defects, or mutations in the fungal genome.

Mycoviruses Viruses that infect and multiply in fungi.

Vegetative incompatibility A genetically controlled self/non-self recognition system in fungi that determines the ability to undergo hyphal anastomosis. This is regarded as antiviral defense that operates at the population level.

Virocontrol One form of biological control using viruses. Control of crops against phytopathogenic fungi utilizing viruses infecting them.

Introduction

Relative to plant virology, animal virology, or bacterial virology, fungal virology (or mycovirology) is new. In 1948, an economically important disease of the cultivated edible mushroom, *Agaricus bisporus* (a basidiomycete), characterized by malformed fruiting bodies and serious yield loss, was first reported in a mushroom house owned by the La France Brothers of Pennsylvania. The disease was called “La France disease”, and similar diseases were reported shortly thereafter from Europe, Japan, and Australia. Different designations, such as “X-disease”, “watery stripe”, “brown disease”, and “die-back” were given to basically the same disease as La France. The significance of the 1948 report lies in the fact that it led to the discovery of fungal viruses (mycoviruses). In 1962, Hollings noted the presence of at least three types of virus particles in diseased mushroom sporophores. This was the first report of virus particles in association with a fungus and is regarded as the dawn of mycovirology. The subsequent discovery that dsRNA of mycoviral origin was responsible for interferon-inducing activities of cultural filtrates of several species of *Penicillium* spp. (ascomycetes) greatly stimulated the search for mycoviruses.

Fungi, like other living organisms, can be infected by a number of viruses, and mycoviruses are found in all the major groups of fungi such as the Ascomycota and Basidiomycota as well as fungal-like protists such as Oomycota (see below). Although mycoviruses are widely prevalent, only those infecting a limited number of fungal host species have been studied, e.g., the yeast *Saccharomyces cerevisiae*, edible mushrooms and phytopathogenic fungi. Given the predicted vast number of fungal species (approximately 120,000 accepted species and much more unknown species), it is expected that a greater number of unrecognized mycoviruses occur in nature. Support for this hypothesis comes from recent extensive searches of field fungal isolates that showed relatively high frequencies of virus infection, e.g., approximately 65%, 20%, and 2%–28% of isolates of *Helicobasidium mompa* (a basidiomycete that causes violet root rot), *Rosellinia necatrix* (an ascomycetes called the white root rot fungus) and *Cryphonectria parasitica* (an ascomycetes called the chest blight fungus) were found to be infected, respectively. Another line of support is from recent next-generation sequencing and metaviromic analyzes of environmental and plant-associated samples discovering numerous viruses most likely of fungal origin.

Biological Properties

Host Range

The natural host range of mycoviruses is likely to be restricted to the same or closely related vegetative compatibility groups that allow lateral transmission. The development of inoculation or introduction methods allows testing of experimental host ranges. Until recently, there were no known experimental host ranges for fungal viruses because of a lack of suitable infectivity assays. Experimental host ranges for some mycoviruses, however, were recently demonstrated and shown to extend to different vegetative compatibility groups, different species, different genera and even to different classes. For example, mycoreovirus 1 (MyRV1, an encapsidated dsRNA virus), the prototype of the genus *Mycoreovirus* in family *Reoviridae*, can replicate and induce phenotypic alterations in different vegetative compatibility groups of *C. parasitica* similar to those exhibited by the original virus-containing strain. Furthermore, *Cryphonectria hypovirus* 1 (CHV1, a capsidless (+)ssRNA virus), the type member of the family *Hypoviridae*, can replicate in and confer hypovirulence onto members of several other ascomycete genera, e.g., *Endothia gyrosa* and *Valsa ceratosperma*, fungi associated with diseases of trees such as eucalyptus or apples and pears, in addition to its natural host, *C. parasitica*. *Cryphonectria parasitica* mitovirus 1 (CpMV1, the genus *Mitovirus*), a mitochondrially replicating capsidless (+)ssRNA virus, can be replicated in *Valsa* spp., but not in fungi more distantly related to its original host *C. parasitica*. Of particular

[☆]In memory of Professor Said A. Ghabrial.

note is the recent discovery of cross-kingdom infections by a few fungal viruses between fungi and plants or fungi and insects (see below).

Symptom Expression

Although the majority of viruses that infect phytopathogenic fungi have been reported to be avirulent, phenotypic consequences of infections with mycoviruses can vary from symptomless to severely debilitating, and from hypovirulence to hypervirulence. Mycoviruses that attenuate the virulence of plant pathogenic fungi provide excellent model systems for basic studies on development of novel biological control measures and for dissecting the mechanisms underlying fungal pathogenesis. In general, infections due to mycoviruses are both symptomless and persistent. Latency benefits the host for survival, and persistence helps the virus in the absence of extracellular modes of transmission. To ensure their retention, some mycoviruses have evolved to bestow selective advantage to their host (e.g., the “killer” phenotypes in yeasts and smuts that are associated with the totivirus infection). Because of their ability to secrete killer toxins, yeast killer strains have been utilized by the brewing industry to provide protection against contamination with adventitious sensitive strains. The genes encoding the smut killer toxins have been used for development of novel transgenic approaches to control the corn smut.

Macroscopic symptoms caused by fungal viruses are a consequence of alterations in complex physiological processes that involve interactions between host and virus factors. Several virally encoded proteins are identified as symptom determinants including the papain-like protease, p29, of some hypoviruses that are phylogenetically related to potyviruses and viruses belonging to the picorna-like superfamily. This protein acts to repress host pigmentation and conidiation regardless of whether it is expressed from host chromosomes or the homologous virus genome. A protein encoded by one of the betachrysovirus (family *Chrysoviridae*) dsRNA genomic segments acts as a symptom determinant of phytopathogenic host fungi. Another interesting example is the proteins encoded by the totivirus *Helminthosporium victoriae* 190S virus (HvV190S, the prototype of the genus *Victorivirus*) infecting the filamentous ascomycete *Helminthosporium victoriae*, the causal agent of Victoria blight of oats. Co-expression of the HvV190S capsid and RdRP proteins results in empty capsid production and phenotypic changes similar to those induced in virus-infected fungal isolates, suggesting that viral replication is not required for symptom development. Host factors involved in symptom expression are not well studied except for host genes involved in the killer phenomenon in yeast infected with the totivirus *Saccharomyces cerevisiae* virus L-A (ScV-L-A) and its associated satellite dsRNAs as exemplified by super killer genes and the symptom expression by CHV1 in *C. parasitica*. A genetic screen for host factors led to the identification of a Mg^{2+} transporter as being necessary for normal symptom induction by CHV1 in *C. parasitica*. The putative mRNA binding protein, termed virus response 1 (*vr1*), and hexagonal peroxisome protein (Hex1) of the filamentous ascomycete *Fusarium graminearum*, the causal agent of fusarium head blight, are associated with symptom expression induced by a betachrysovirus (*Fusarium graminearum* virus China 9, FgV-ch9) and a fusarivirus (*Fusarium graminearum* virus 1, FgV1, a capsidless (+)ssRNA virus related to hypoviruses), respectively. Transcriptome analysis was performed for only limited virus/natural host combinations. In the *C. parasitica*/mycovirus pathosystem, the SAGA (Spt-Ada-Gcn5 acetyltransferase) complex, a general transcriptional co-activator and DCL2 (a Dicer), a key enzyme of antiviral RNA silencing, mediate the upregulation of many host genes. Some of the virus responsive *C. parasitica* genes are associated with CHV1 symptom expression or its mitigation. Of the 2200 *C. parasitica* genes, 13.4% are either up- or down regulated upon infection with the prototype strain (a severe strain) of CHV1, while only 7.5% are altered in their transcription levels by infection with a mild strain of CHV1. Half of the *C. parasitica* genes responsive to the infection with the latter mild strain are commonly altered in transcription by the severe strain, which generally causes greater magnitude of transcriptional changes.

Transmission

Mycoviruses generally lack an extracellular phase to their life cycles. They are transmitted intracellularly during cell division, sporogenesis and cell fusion. Lateral (horizontal) transmission usually occurs only between individuals within the same fungal species, which belong to the same or closely related vegetative compatibility groups. Vegetative compatibility is governed genetically. Mycoviruses may be eliminated during sexual spore formation. Although the totiviruses and narnaviruses (capsidless (+)ssRNA viruses belonging to the family *Narnaviridae*) that infect the budding yeasts are effectively transmitted via ascospores, the mycoviruses infecting the ascomycetous filamentous fungi are in effect eliminated during ascospore formation. Whereas the (+)ssRNA and dsRNA viruses infecting mushrooms are transmitted efficiently via basidiospores, the virus-containing strains of the basidiomycete *H. mompa* are cured during the sexual sporulation processes. Therefore, whether a mycovirus is transmitted through sexual spores depends on the host/virus combination involved. Whereas mycovirus transmission through asexual spores occurs frequently, its rate varies greatly depending on the combination of viral and host fungal strains. It was recently revealed that CHV1 p29 (the papain-like protease p29) might play a role in virus transmission since it enhances the transmission of the homologous virus (CHV1) as well as a heterologous virus (a mycoreovirus, MyRV1) through the conidia in *C. parasitica*.

Though rare in nature, interspecies transmission has been reported between members within the same genus including *Cryphonectria*, *Sclerotinia*, and *Ophiostoma* that share the same habitats. Interspecies transmission is also implicated between distantly related fungi, e.g., between *R. necatrix* and *Entoleuca* spp. It remains unknown whether the interspecies barrier is overcome by physical contacts, by vectors, or by infecting viruses.

Of note is the recent discovery of extracellular routes for mycovirus entry. There is experimental evidence for the transmissibility of an encapsidated ssDNA virus, *Sclerotinia sclerotiorum* hypovirulence associated DNA virus 1 (SsHDV1, the family *Genomoviridae*), by a mycophagous fly between *Sclerotinia sclerotiorum* (an ascomycetous plant pathogen with a broad host range) colonies and the acquisition and replication of a plant (+)ssRNA virus by/in *Rhizoctonia solani* (a basidiomycetous plant pathogen that has a wide host range). See the Section “Cross-Kingdom Infection by Viruses and Viroids” in this article.

Mixed Infections

Mixed infections with two or more related or unrelated viruses and accumulation of defective dsRNA and/or satellite dsRNA molecules are common features of mycovirus infections. Examples include the totiviruses (Sphaeropsis sapinea RNA virus 1 and 2, genus *Victrivirus*) infecting *Sphaeropsis sapinea* (an important opportunistic fungal pathogen on pine trees) and *S. cerevisiae* (ScV-L-A and ScV-LB/C), and *Penicillium stoloniferum* gammapartitiviruses (*Penicillium stoloniferum* virus S and F). There are a number of known examples of mixed infections with plant or animal viruses where one virus either interferes with or enhances the replication of the other one. As a consequence, reduction or increase in symptom severity may arise.

Recent extensive searches for fungal viruses confirmed relatively high mixed infection rates in some phytopathogenic fungi like *C. parasitica*, *R. necatrix*, *Heterobasidion* spp. (a root rot basidiomycete fungus), *S. sclerotiorum* and *H. mompa*. There are at least three types of recognizable virus/virus interactions: synergistic, mutualistic and antagonistic interactions and genome alterations.

Recent studies show synergistic interactions between a hypovirus (CHV1) and a mycoreovirus (MyRV1) through a hypovirally-encoded protein. In this case, one-way synergism is observed in that only the hypovirus infection trans-activates the replication of the mycoreovirus, which has a replication strategy distinct from that of hypoviruses. A similar one-way synergistic interaction is detected between *Cryphonectria hypovirus* 4 (CHV4) and mycoreovirus 2 (MyRV2), in which CHV4 enhances stable infection and replication of MyRV2 in *C. parasitica*. These two examples of synergism are mediated by papain-like cysteine proteases p29 and p24 encoded by CHV1 and CHV4, respectively, both of which also play a role in the suppressor of antiviral RNA silencing. Genome rearrangement occurs in *C. parasitica* co-infected with MyRV1 and CHV1 which is a result of RNA silencing suppression via the action of the CHV1 p29. Synergistic interactions are also found between a megabirnavirus, *Rosellinia necatrix* megabirnavirus 2 (RnMBV2, an encapsidated dsRNA virus, the prototype of the genus *Megabirnavirus* in family *Megabirnaviridae*) and a partitivirus, *Rosellinia necatrix* partitivirus 1 (RnPV1), in which RnMBV2 confers hypovirulence with the aid of RnPV1 and elevates the accumulation of RnPV1. For antagonistic and mutualistic interactions, see the “Future Perspectives” Sections below.

Fungal Virus Taxonomy

A list of the virus families and genera into which mycoviruses are classified or officially proposed to the international committee on taxonomy of viruses (ICTV) is included in [Table 1](#). Mycoviruses have diverse genomes including ones made of linear dsRNA, linear positive-sense (+) ssRNA, linear negative-sense (−) ssRNA, and circular ssDNA.

Mycoviruses with dsDNA genomes are not in the list, but might yet be found since dsDNA viruses of water molds, now classified as protists not fungi, have been reported. Mycoviruses for which three-dimensional (3D) structures of the virions have been reported are further described in “Structural features”.

Members of some mycovirus families, e.g., the families *Narnaviridae*, and *Hypoviridae*, infect only fungi, while members in other families, e.g., the families *Metaviridae*, *Pseudoviridae*, *Reoviridae*, *Totiviridae*, *Chrysoviridae*, *Endornaviridae*, *Alphaflexiviridae*, *Botourmiaviridae* and *Partitiviridae* infect fungi, protozoa, plants or animals.

Many mycoviruses have RNA genomes and have either dsRNA genomes or (+)ssRNA genomes. Viruses with (−)ssRNA (monopartite mymonaviruses and multipartite phenuiviruses) or circular ssDNA genomes have recently been reported. Since the article of the last edition of this series was published, an extremely large number of viruses have been identified.

Visit See “Relevant Website section” for an updated list of all mycoviruses. Note that there are still many unassigned mycoviruses awaiting official taxonomical classification. [Table 1](#) represents these mycoviruses in the “unassigned” sections.

dsRNA Mycoviruses

Mycoviruses with dsRNA genomes represent a major portion of the fungal-infecting viruses so far reported. With the exception of the mycoviruses (family *Reoviridae*), which have spherical double-shelled particles 80 nm in diameter, the typical dsRNA mycoviruses are isometric particles ranging 25–50 nm in diameter. In addition, for polymycoviruses (family *Polymycoviridae*) with multipartite dsRNA genomes, two viral forms, a filamentous encapsidated and a capsid-less RNA-protein complex forms, have been proposed. These dsRNA viruses are classified, based on the number of genome segments and phylogenetic relation, into one genus *Botybirnavirus* and eight families, *Totiviridae*, *Partitiviridae*, *Curvulaviridae*, *Amalgaviridae*, *Chrysoviridae*, *Megabirnaviridae*, *Quadriviridae*, and *Polymycoviridae* ([Table 1](#)).

Viruses in the family *Totiviridae* have non-segmented dsRNA genomes coding for a CP and an RdRP or CP-RdRP fusion protein. At present, five genera have been placed in this family: *Totivirus*, *Victorivirus*, *Giardiavirus*, *Leishmanivirus*, and *Trichomonasvirus*.

Table 1 List of viral families and genera into which fungal viruses are classified^{a,b}

Order/Family/Genus/Species	Virus abbreviation	No. of segments	Accession number
<u>Single stranded DNA viruses</u>			
<u>Geplafuvirales</u>			
<u>Genomoviridae</u>			
<i>Gemicirculavirus</i>	SsHDV1	1	GQ365709
<i>Sclerotinia gemicircularvirus 1</i>			
<u>Gemytripvirus*</u>			
<i>Gemytripvirus fugra1*</i>	FgGMTV1	3	MK430076; MK430077; MK430078
<u>Reverse Transcribing RNA viruses</u>			
<u>Ortervirales</u>			
<u>Pseudoviridae</u>			
<u>Pseudovirus</u>			
<i>Saccharomyces cerevisiae Ty1 virus</i>	SceTy1V	1	M18706
<u>Hemivirus</u>			
<i>Saccharomyces cerevisiae Ty5 virus</i>	SceTy5V	1	U19263
<u>Metaviridae</u>			
<u>Metavirus</u>			
<i>Saccharomyces cerevisiae Ty3 virus</i>	SceTy3V	1	M34549
<u>Double-stranded RNA viruses</u>			
<u>Reovirales</u>			
<u>Reoviridae</u>			
<u>Mycoreovirus</u>			
<i>Mycoreovirus 1</i>	MyRV1	11	AY277888; AY277889; AY277890; AB179636; AB179637; AB179638; AB179639; AB179640; AB179641; AB179642; AB179643
<u>Ghabrivirales</u>			
<u>Totiviridae</u>			
<u>Totivirus</u>			
<i>Saccharomyces cerevisiae virus L-A</i>	ScV-L-A	1	J04692; X13426
<u>Victorivirus</u>			
<i>Helminthosporium victoriae 190SV</i>	Hv190SV	1	U41345
<u>Chrysoviridae</u>			
<u>Alphachrysovirus</u>			
<i>Penicillium chrysogenum virus</i>	PcV	4	AF296439; AF296440; AF296441; AF296442
<u>Betachrysovirus</u>			
<i>Botryosphaeria dothidea chrysovirus</i>	BdCV1	4	KF688736; KF688737; KF688738; KF688739
<u>Megabirnaviridae</u>			
<u>Megabirnavirus</u>			
<i>Rosellinia necatrix megabirnavirus 1</i>	RnMBV1	2	AB512282; AB512283
<u>Quadriviridae</u>			
<u>Quadrivirus</u>			
<i>Rosellinia necatrix quadrivirus 1</i>	RnQV1	4	AB620061; AB620062; AB620063; AB620064
<u>Durnavirales</u>			
<u>Partitiviridae</u>			
<u>Alphapartitivirus</u>			
<i>Rosellinia necatrix partitivirus 2</i>	RnPV2	2	AB569997; AB569998
<u>Betapartitivirus</u>			
<i>Atkinsonella hypoxylon virus</i>	AhV	2	L39125; L39126; L39127 (satellite)
<u>Gammapartitivirus</u>			
<i>Penicillium stoloniferum virus S</i>	PsV-S	2	AY156521; AY156522
<u>Unassigned species</u>			
<i>Agaricus bisporus virus 4</i>	AbV4	–	No entry in Genbank
<i>Gaeumannomyces graminis virus 0196A</i>	GgV-0196A	–	No entry in Genbank
<i>Gaeumannomyces graminis virus T1A</i>	GgV-T1A	–	No entry in Genbank
<u>Amalgaviridae</u>			
<u>Zybavirus</u>			
<i>Zygosaccharomyces bailii virus Z</i>	ZbV-Z	1	KU200450
<u>Curvulaviridae*</u>			
<u>Orthocurvulavirus*</u>			
<i>Curvularia orthocurvulavirus 1*</i>	CThTV	2	EF120984; EF120985

Table 1 Continued

<i>Order/Family/Genus/Species</i>	<i>Virus abbreviation</i>	<i>No. of segments</i>	<i>Accession number</i>
Order unassigned			
Polymycoviridae			
Polymycovirus			
<u>Aspergillus fumigatus polymycovirus 1</u>	AfuTmV1	4	HG975302; HG975303; HG975304; HG975305
Family unassigned			
Botybirnavirus			
<u>Botrytis porri botybirnavirus 1</u>	BpBBV1	2	JF716350; JF716351
Unassigned			
Fusarium virguliforme dsRNA mycovirus 1	FvV1	1	JN671444
yado-nushi virus	YnV1	1	LC061478
Rosellinia necatrix megatotivirus 1	RnMTV1	1	LC333746
Alternaria alternata virus 1	AaV1	4	AB368492; AB438027; AB438028; AB438029
Ustilagoidea virens RNA virus M	UvRV-M	1?	KJ101567
<u>Positive-sense RNA viruses</u>			
Wolframvirales			
Narnaviridae			
Narnavirus			
<u>Saccharomyces 20S narnavirus</u>	ScNV-20S	1	M63893
Cryppavirales			
Mitoviridae			
Mitovirus			
<u>Cryphonectria mitovirus 1</u>	CpMV1	1	L31849
Ourivirales			
Botourmiaviridae			
Botoulivirus			
<u>Botrytis botoulivirus</u>	BOLV	1	LN827955
Magoulivirus			
<u>Magnaporthe magoulivirus 1</u>	MOLV1	1	LT593139
Scleroulivirus			
<u>Sclerotinia scleroulivirus 1</u>	SsOLV1	1	KP900928
<u>Penoulivirus*</u>			
<u>Phaeoacremonium penoulivirus*</u>	PmOLV1	1	MK584843
<u>Rhizoulivirus*</u>			
<u>Rhizoctonia rhizoulivirus*</u>	RsOLV1	1	KP900922
Sobelivirales			
Barnaviridae			
Barnavirus			
Mushroom bacilliform virus	MBV	1	MBU07551
Durnavirales			
Hypoviridae			
Hypovirus			
<u>Cryphonectria hypovirus 1</u>	CHV-1	1	M57938
Martellivirales			
Endornaviridae			
Alpharendornavirus			
<u>Phytophthora alphaendornavirus 1</u>	PEV1	1	AJ877914
Betaendornavirus			
<u>Sclerotinia sclerotiorum betaendornavirus 1</u>	SsEV1		KJ123645
Tymovirales			
Alphaflexiviridae			
Botrexvirus			
<u>Botrytis virus X</u>	BotVX	1	AY055762
Sclerodarnavirus			
<u>Sclerotinia sclerotiorum debilitation-associated RNA virus</u>	SsDRV	1	AY147260

(Continued)

Table 1 Continued

Order/Family/Genus/Species	Virus abbreviation	No. of segments	Accession number
<i>Gammaplexiviridae</i>			
<i>Mycovirus</i>			
<i>Botrytis virus F</i>	BotVF	1	AF238884
<i>Deltaplexiviridae</i>			
<i>Deltaplexivirus</i>			
<i>Sclerotinia sclerotiorum deltaplexivirus 1</i>	SsDFV1	1	KT581451
Unassigned			
Diaporthe RNA virus	DRV	1	AF142094
Fusarium graminearum virus DK21	FgV/DK21	1	AY533037
Oyster mushroom spherical virus	OMSV	1	AY182001
Sclerophthora macrospora virus A	SmV-A	3	AB083060; AB083061; AB083061
Sclerophthora macrospora virus B	SmV-B	1	AB012756
Sclerotinia sclerotiorum RNA virus L	SsRV-L	1	EU779934
Fusarium graminearum virus 1	FgV1	1	AY533037
yado-kari virus 1	YKV1	1	LC006253
Fusarium boothii large flexivirus 1	FbLFV1	1	LC425115
hadaka virus 1	HadV1	11	LC519840; LC519841; LC519842; LC519843; LC519844; LC519845; LC519846; LC519847; LC519848; LC51984; LC519850
<i>Negative-sense RNA viruses</i>			
<i>Mononegavirales</i>			
<i>Myonnaviridae</i>			
<i>Sclerotimonavirus</i>			
<i>Sclerotinia sclerotimonavirus</i>	SsNSRV-1	1	KJ186782
<i>Botrytimonavirus*</i>			
<i>Botrytimonavirus botrytidis*</i>	BcNSRV-7	1	MT157413
<i>Lentimonavirus*</i>			
<i>Lentinula lentimonavirus*</i>	LeNSRV-1	1	LC466007
<i>Phyllomonavirus*</i>			
<i>Phyllomonavirus phyllosphaerae*</i>	SLaNSRV-4	1	KT598229
<i>Auricularimonavirus*</i>			
<i>Auricularimonavirus auriculariae*</i>	AhNSRV-1	1	MT259204
<i>Penicillimonavirus*</i>			
<i>Penicillimonavirus alphapenicillii*</i>	PdNSRV-1	1	MK584858
<i>Plasmopamonavirus*</i>			
<i>Plasmopamonavirus plasmoparae*</i>	PvLAMV8	1	MN557004
<i>Bunyavirales</i>			
<i>Phenuiviridae</i>			
<i>Lentivirus</i>			
<i>Lentinula lentivirus</i>	LeNSRV-2	2	LC466008; LC466009
<i>Entovirus</i>			
<i>Entoleuca entovirus</i>	EnPLV-1	2	MF375882; MK140653
Unassigned			
Fusarium poae negative-stranded RNA virus 1	FpNSRV-1	1?	LC150618
Botrytis cinerea negative-stranded RNA virus 1	BcNSRV-1	1?	LN827956

*The type species of the specified genus is underlined.

^bProposed taxa for new creation or re-classification (awaiting EC consideration and ICTV 2020 ratification) are asterisked.

Viruses in the genera *Totivirus* and *Victorivirus* infect fungi, whereas those belonging to the latter three genera infect parasitic protozoa. Plant-infecting dsRNA viruses related to totiviruses (potential members of genus *Totivirus*) were recently found. At least two distinct RdRP expression strategies have been reported for totiviruses: (1) those that express their RdRP as a fusion protein (CP-RdRP or Gag-Pol) by ribosomal frameshifting, such as the yeast ScV-L-A and the viruses that infect parasitic protozoa and (2) those that synthesize RdRP as a separate non-fused protein by an internal initiation mechanism (e.g., a coupled termination–reinitiation mechanism), as demonstrated for HvV190S (a victorivirus) and others that infect filamentous fungi. Phylogenetic analysis of CP or RdRP sequences of totiviruses reflects these differences, and separate phylogenetic clusters can be generated. HvV190S and other victoriviruses that infect filamentous fungi are closer to each other than to viruses infecting yeast, smut fungi and protozoa. The fact that independent alignments of CP and RdRP sequences give similar phylogenetic relationships supports the conclusion that totiviruses infecting filamentous fungi should reside in a genus of their own. HvV190S and other known totiviruses that infect filamentous fungi have been classified within

the genus *Victorivirus* (Table 1). The genomes of partitiviruses, curvulaviruses, and megabimaviruses consist of two dsRNA segments, while chrysovirus, quadriviruses, and polymycoviruses have three to seven, four, and four to seven dsRNA segments, respectively.

The unclassified dsRNA mycovirus *Agaricus bisporus* virus 1 (AbV1), also designated La France isometric virus, causes a serious disease (named La France disease) of the cultivated mushroom *Agaricus bisporus* (white button mushroom). AbV1 is of special interest because of its historical and economic importance. The AbV1 virions, isolated from diseased fruit bodies and mycelia, are isometric 36 nm in diameter and co-purify with nine dsRNA segments (referred to as disease-associated dsRNAs). The size of dsRNA segments varies from 3.6 kbp to 0.78 kbp, three of which are believed to be satellites. It is not clear at present whether the nine dsRNA segments are encapsidated individually, in various combinations, or all segments are packaged in single particles. Based on the size of the particles, cesium sulfate gradient profile and results of dsRNA and protein analyses of the gradient fractions, it is highly unlikely that all dsRNAs are packaged together in single particles. More realistically, AbV1 represents a multiparticle system in which the various particle classes have similar densities. Interestingly, phylogenetic analysis of the conserved motifs of AbV1 RdRP, encoded by dsRNA segment 1, and other dsRNA mycoviruses showed that AbV1 is closely related to the multipartite chrysovirus. The Blast analyses also supported the above notion.

In addition to AbV1, there are an increasing number of unclassified dsRNA mycoviruses. Table 1 lists only some of those well-characterized.

Single-Stranded RNA Viruses

Single-stranded (ss) RNA mycoviruses are grouped into two with linear positive-sense (+)ssRNA genomes (currently classified into nine families: *Alphaflexiviridae*, *Barnaviridae*, *Deltaflexiviridae*, *Endornaviridae*, *Gammaflexiviridae*, *Hypoviridae*, *Narnaviridae*, *Mitoviridae*, and *Botourmiaviridae*) and with linear (–)ssRNA (family *Mymonaviridae*, and *Phenuiviridae*) (Table 1). Among them are mycoviruses with apparent ssRNA genomes that do not code for capsid proteins and exist more or less predominantly as dsRNA “replicative” forms in their hosts. Because of the lack of true virions, these viruses were easier to isolate and study as their replicative dsRNA forms, and some were grouped with dsRNA viruses (e.g., family *Hypoviridae*). However, there is ample evidence that many of these mycoviruses replicate and express their genomes like (+)ssRNA viruses and that the lineage of their RdRP and helicase genes are within the lineages of (+)ssRNA viruses. The simplest types of these viruses include members of the families *Narnaviridae* and *Mitoviridae*, whose RNA genomes code only for RdRP and the viruses exist as RNA/RdRP ribonucleoprotein (RNP) complexes. The corresponding dsRNA replicative forms can be isolated from infected tissues, usually in smaller molar amounts than the genomic ssRNA. Phylogenetic analysis of RdRPs of members of the families *Narnaviridae* along with those of other mycoviruses and related taxa indicate a distant relationship between members of the family *Narnaviridae* and bacteriophages such as Q β and MS2 belonging to the family *Leviviridae*. Viruses in the family *Botourmiaviridae* are phylogenetically more closely related to narnaviruses than to mitoviruses. The family accommodates four genera: *Ourmiavirus*, *Botoulivirus*, *Magoulivirus*, and *Scleroulivirus* together with two newly proposed genera “*Penoulivirus*” and “*Rhizoulivirus*.” Viruses belonging to the first genus infect plants associated with host diseases and have tri-partite (+)ssRNA genomes, whereas viruses in the other genera infect filamentous fungi, and have mono-partite (+)ssRNA genomes encoding only RdRPs. These fungal viruses are capsidless, and likely are replicated in the cytoplasm.

Lack of true virions (a capsid-less nature) is characteristic of two other groups of classified (+)ssRNA viruses, those belonging to the family *Hypoviridae* and the family *Endornaviridae*. Viruses in the family *Hypoviridae* are phylogenetically related to the (+)ssRNA viruses in the family *Potyviridae* (picorna-like virus supergroup). Comparisons of hypovirus conserved motifs of RdRP, helicase and protease with those of members of the family *Potyviridae* suggest that viruses in the genus *Bymovirus*, which are plant viruses transmitted by soil-borne plasmodiophorid *Polymyxa graminis*, are the closest relatives to hypoviruses. Fungal endornaviruses belong to the genus *Betaendornavirus*. Exceptionally, a few non-plant viruses in the genus *Alphaendornavirus*, such as *Phytophthora endornavirus* 1 and *Rhizoctonia cerealis endornavirus* 1, infect fungi and oomycetes. Most other *alphaendornaviruses* infect plants. Although *Phytophthora* species and other members of the family *Pythiaceae* (Oomycetes) have many biological properties in common with fungi, they are currently classified, based on sequence similarities, in a protist group known as the Stramenopiles. Endornaviruses are believed to have evolved from an alpha-like virus that has lost its capsid gene. This is consistent with the recent finding that RdRPs of PEV1 and other endornaviruses cluster with those of families and genera in the alpha-like virus superfamily of (+)ssRNA viruses.

The mushroom bacilliform virus (MBV; genus *Barnavirus*, family *Barnaviridae*) is the only mycovirus known to have bacilliform virions. MBV has a (+)ssRNA genome that contains seven ORFs, three of which encode a putative chymotrypsin-like serine protease, a putative RdRP and a CP. The polypeptides encoded by the remaining four ORFs have no homology to known proteins. Amino acid sequence comparisons of the putative protease and RdRP suggest that MBV is evolutionarily related to sobemoviruses and poleroviruses, both of which have smaller monopartite (+)ssRNA genomes and spherical particles. Although double infections of cultivated mushrooms with MBV and ABV1 occur commonly, the role of MBV in the ensuing dieback disease of cultivated mushroom remains unknown.

Botrytis virus F (BVF) has flexuous rod-shaped particles comparable in size and morphology to (+)ssRNA plant flexiviruses such as potato virus X (a potexvirus, family *Alphaflexiviridae*). Amino acid sequence identities of the conserved helicase and RdRP regions and the CP genes are most similar to those of potex-like viruses. The main difference between BVF and these plant viruses is the lack of a movement protein (MP) gene(s), which encodes a key protein(s) in cell-to-cell movement of a plant virus. Despite

similar particle morphology along with amino acid sequence similarities of both the replicase and CP genes, it is obvious that the mycovirus BVF is distinct enough to belong to a new family. It has been approved by ICTV as an exclusive genus and family designated as *Mycoflexivirus* and *Gammaflexiviridae* that were created to include BVF (Table 1).

The genome of *Sclerotinia sclerotiorum* debilitation-associated RNA virus (SsDRV) contains a single ORF encoding a protein with significant sequence similarity to the replicases of the “alphavirus-like” supergroup of (+)ssRNA viruses. The putative SsDRV-encoded replicase protein contains the conserved methyl transferase, helicase and RdRP domains characteristic of the replicases of plant alphaflexiviruses (family *Alphaflexiviridae*) and BVF. Although phylogenetic analysis of the conserved RdRP motifs verified that SsDRV is closely related to BVF and to the allelerviruses in the family *Alphaflexiviridae*, SsDRV is distinct enough from these viruses, mainly based on the lack of CP and MP genes, to justify the creation of another genus, *Sclerodarnavirus* in the family *Alphaflexiviridae*. Another mycovirus member of the genus *Sclerodarnavirus* is *Botrytis virus X* (BVX) (Table 1). Recent taxonomical reorganization of related viruses includes the creation of the genus *Deltaflexivirus*, family *Deltaflexiviridae* that accommodates three mycovirus members such as *Sclerotinia sclerotiorum* deltaflexivirus 1 and *Fusarium graminearum* deltaflexivirus 1.

A negative-sense (–)ssRNA virus termed *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNSRV1) was isolated in 2014. SsNSRV1 is now classified in the species *Sclerotinia sclerotimonavirus*, genus *Sclerotimonavirus* (family *Myonnaviridae*, order *Mononegavirales*). SsNSRV1 forms filamentous nucleocapsid structures of 22 nm × 200–2000 nm and appears to employ a gene expression strategy with sequential transcription (“stop-start”) mechanism similar to other mononegaviruses. The family accommodates one genus *Sclerotimonavirus* and other newly proposed or renamed genera “*Botrytimonavirus*,” “*Lentimonavirus*,” “*Phyllomonavirus*,” “*Auricularimonavirus*,” “*Penicillimonavirus*,” and “*Plasmopamonavirus*.” Multipartite (–)ssRNA viruses have also been reported from edible shiitake mushroom (*Lentinula edodes* negative-strand RNA virus 2) and *Entoleuca* sp. infecting avocado (*Entoleuca bunyavirus* 1). Both viruses have a smaller segment with the putative ambisense coding strategy.

Unassigned ssRNA Viruses

It is worth noting that a relatively large number of mycoviruses remain unassigned including some well-characterized ones like *Diaporthe* RNA virus (DRV), and *Sclerophthora macrospora* viruses A and B (SmV-A and SmV-B) (Table 1). DRV is another naked RNA mycovirus that is associated with hypovirulence of its fungal host. It has two large ORFs present in the same reading frame, which are most likely translated by readthrough of an UAG stop codon in the central part of the genome. The longest possible translation product has a predicted molecular mass of about 125 kDa, which shows significant homology to the non-structural proteins of carmoviruses of the (+)ssRNA virus family *Tombusviridae*. Interestingly, transcripts derived from full-length cDNA clones were infectious when inoculated in spheroplasts and the transfected isolates exhibited phenotypic traits similar to the naturally infected isolate.

SmV A found in *Sclerophthora macrospora*, the pathogenic fungus responsible for downy mildew of gramineous plants, is a small icosahedral virus containing three segments of (+)ssRNAs (RNAs 1, 2, and 3). RNA 1 has two ORFs, one of which contains the RdRP motifs, and RNA 2 codes a single ORF for CP, while RNA 3 has any ORFs suggesting it is a satellite-like RNA. Whereas the deduced amino acid sequence of RdRP shows some similarity to RdRPs of members of the family *Nodaviridae*, the amino acid sequence of the viral CP shows similarity to those of members in the family *Tombusviridae*. The capsid of SmV A is composed of two protein species, named CP 1 and CP 2, both encoded in the same ORF in RNA 2. CP 2 is apparently derived from CP 1 via proteolytic cleavage at the N-terminus. The genome organization of SmV A is distinct from those of other known fungal RNA viruses, and suggest that SmV A should be classified into a novel genus of (+)ssRNA viruses.

SmV B, which is also found in *S. macrospora*, has small icosahedral, monopartite virions containing a (+)ssRNA species as a genome. The viral genome has two large ORFs; ORF1 encodes a putative polyprotein containing the motifs of chymotrypsin-related serine protease, and ORF2 encodes a CP. The genome arrangement of SmV B is similar to those belonging to the genera *Sobemovirus*, *Barnavirus*, and *Polerovirus*. The putative domains for the serine protease, VPg (the viral protein linked to the genome), RdRP, and the CP are located in this order from the 5′ terminus to the 3′ terminus. SmV B, however, is distinctive since its genome has only two ORFs, but the others have at least four ORFs. The genome organization of the barnavirus MBV, on the other hand, resembles that of poleroviruses. These results suggest that SmV B, like SmV A, should also be classified into a new genus of (+)ssRNA viruses.

Several unclassified (–)ssRNA viruses distantly related to ophioviruses, which are filamentous plant viruses within the family *Aspiviridae* known to be transmitted by the soil-inhabiting fungus *Olpidium virulentus*, have been reported. Examples include *Fusarium poae* negative-stranded RNA virus 1 and *Rhizoctonia solani* negative-stranded RNA virus 1 and unclassified bunyaviruses (order *Bunyavirales*) such as *Macrophomina phaseolina* negative-stranded RNA virus 1 and *Botrytis cinerea* negative-stranded RNA virus 1.

The mycoviruses belonging to the families *Pseudoviridae*, *Metaviridae*, *Reoviridae* (genus *Mycoreovirus*), *Totiviridae*, *Partitiviridae*, *Chrysoviridae*, *Megabirnaviridae*, *Quadriviridae*, *Hypoviridae*, *Mitoviridae*, *Narnaviridae*, *Barnaviridae* and *Myonnaviridae*, and the members of possible novel families such as *botybimaviruses* (genus *Botybimavirus*), *alternaviruses*, *fusariviruses* and *phlegiviruses* are discussed in more detail in separate articles in this work.

Replication and Gene Expression Strategy

Replication cycles of mycoviruses are not well studied except for a few cases including the prototype totivirus ScV-L-A. For dsRNA mycoviruses including members of the *Totiviridae*, *Partitiviridae*, *Reoviridae*, and *Chrysoviridae*, virus particles or subviral particles,

containing RdRP, are believed to play pivotal roles in RNA transcription and replication. Replication of capsidless (naked) (+)ssRNA mycoviruses represented by members of the family *Hypoviridae* may occur in infection-specific, lipid-membranous vesicles presumed to contain viral RdRP and RNA helicase. These vesicles are able to synthesize in vitro both plus and minus RNA at a ratio of 1:8. The narnaviral genomic RNA, encoding only a single protein (RdRP), is associated with RdRP forming an RNP structure rather than being encapsidated. These RNP complexes seem to play a key role in viral replication.

Mycoviruses, like many RNA viruses of plants and animals, employ non-canonical translational strategies for expressing their genomes. These include -1 frame shifting (totiviruses, trichomonasviruses and giardiaviruses in the *Totiviridae*), -2 frame shifting (trichomonasviruses in the *Totiviridae*), +1 frame shifting (*Pseudoviridae*, *Metaviridae*), presumable +1 frame shifting or ribosomal hopping (leishmaniviruses in the *Totiviridae*), termination-coupled initiation (stop/restart strategies; most of victoriviruses in the *Totiviridae*, *Hypoviridae*), IRES (internal ribosome entry site)-mediated initiation (*Totiviridae*, *Hypoviridae*, *Chrysoviridae*). In addition, proteases such as a cis-acting papain-like cysteine protease (*Hypoviridae*) and 2A-like self-processing peptides (yadokariviruses, probably giardiaviruses in the *Totiviridae* and some others) have also been known. Furthermore, a readthrough of a termination codon strategy is proposed for translation of the 3' proximal ORF of DRV. A non-canonical mechanism may also be required for efficient translation of mRNA of narnaviruses, which lack poly(A) tails. The (CAA)_n repeats found at the 5'-untranslated region (UTR) of chrysoviruses are implicated in translation augmentation, as observed for the 5' UTR sequence of a plant alpha-like virus (tobacco mosaic virus, TMV; family *Virgaviridae*). This possibility needs to be tested, given the recent detection of IRES activities in the UTRs of chrysovirus genomic segments. Translation of the viral genes that are regulated by these mechanisms is considered critical for virus viability.

Recent Technical Advances in Fungal Virology

Fungal virology has been thwarted by many constraints on manipulation of fungal viruses. Many, if not all, plant and animal viruses can be inoculated into individuals/tissue cultures of plant and animal hosts. Some assays with those hosts allow quantitative detection of biologically active viruses. As for mycoviruses, it is rather rare to be able to inoculate into host fungi because of experimental limitations, which often makes the etiology of mycoviruses difficult to establish. Fungal cells have rigid cell walls and are usually difficult to digest for preparation of cell-wall free protoplasts ready for transformation or transfection. Even if protoplasts are made, their maintenance, as with animal cell culturing, is not possible. Furthermore, fungal hosts usually have self/non-self recognition systems operating at inter- and intraspecies levels. Intraspecies barriers are based on vegetative incompatibility/compatibility governed genetically. This is often regarded as one of the host defense barriers that inhibit virus transfer between individuals. To overcome these barriers, a few methods are available. The prototypic hypovirus CHV1 is the first for which a reverse genetics is established. Infection with different CHV1 strains can be launched either from cDNA integrated into host chromosomes or in vitro synthesized viral RNA from the cDNA template. It is noteworthy that via bombardment of mycelia, not protoplasts, the infectious CHV1 cDNA clone can be integrated into chromosomes of fungi other than the natural host *C. parasitica*. cDNA-based transfection systems are now available for other RNA viruses such as DRV (a tombus-like virus), *Saccharomyces cerevisiae* 20S narnavirus (ScNV-20S), *Saccharomyces cerevisiae* 23S narnavirus (ScNV-23S), *Sclerotinia sclerotiorum* uormia-like virus 4, *Sclerotinia sclerotiorum* hypovirus 2, and Yado-kari virus 1 (a calici-like virus).

Mycoviruses in general lack infectivity as purified virions. However, all three members of the genus *Mycoreovirus* including MyRV1, MyRV2 and mycoreovirus 3 (MyRV3, from an *R. necatrix* strain) were found to be infectious as purified particles when applied to fungal protoplasts. It is of interest in this regard that treatment of purified virions with trypsin or chymotrypsin was not required for infectivity. Protoplast fusion provides an alternative approach to introduce mycoviruses into vegetatively incompatible fungal strains that are incapable of hyphal anastomosis. Intra- and interspecies virus transfer via protoplast fusion has been reported in *Aspergillus* spp., *Fusarium* spp. and *Cryphonectria* spp. This method is particularly useful for mycoviruses for which infectious particles or cDNA-derived RNA are unavailable. Another recent revelation is that monokaryotic strains are able to serve as an intermediate mycovirus transmitter between different mycelial incompatibility groups within the same species of *R. necatrix*.

To complete Koch's postulates, virus curing is as important as virus inoculation, because virus must be back inoculated into an isogenic, virus-free strain with the same genetic background as the original virus-infected strain. Virus-free isolates may be obtained from germings of asexual spores if virus transmission through spores is less than 100%. Alternatively, virus-free strains may be isolated by hyphal tip culturing, as in the case for *H. mompa* and *R. necatrix*. This technique is applicable to fungi infected with a virus that is transmitted to 100% of the asexual spores or for fungi that produce little or no spores. Protoplasting is also useful for virus curing. These curing methods can be employed in the presence of antiviral drugs such as ribavirin.

Future Perspectives

Cross-Kingdom Infection by Viruses and Viroids

The ssDNA mycovirus SsHDV1 was the first fungal virus that was shown to be replicated in and transmitted by a mycophagous insect, *Lycoriella ingenua*. Naturally this fly likely serves as a vector of the virus lateral transfer between fungi and also retains the virus across generations through vertical (transovarial) transmission. Of note is that this DNA virus in the purified or crude

preparation can extracellularly enter host mycelia on synthetic media. This feature contributes to field-level virocontrol of rapeseed rot caused by *S. sclerotiorum*.

Other examples of cross-kingdom infections involving mycoviruses include infection of a phytopathogenic basidiomycete, *Rhizoctonia solani*, by one of the best-studied plant viruses, cucumber mosaic virus (CMV, family *Bromoviridae*). The virus is apparently maintained stably in fungal colonies and is transmitted between plants. Namely, not only aphids (Hemiptera; a non-persistent manner) but also fungi act as virus vectors. Furthermore, a two-way facilitative interaction between a plant and fungal alpha-like viruses (TMV and CHV1) has been demonstrated, which could promote cross-kingdom virus infections in both directions. Viroids which are omnipresent plant pathogens having minimal non-protein-coding circular RNA genomes and were not detected from other kingdoms of organisms is also an example. However, hop stunt viroid (HSVd, *Pospiviroidae*), iresine 1 viroid (*Pospiviroidae*) and avocado sunblotch viroid (*Ausunviroidae*), which replicate in the nucleus, nucleus and chloroplast of their plant hosts, respectively, can establish stable infection in a few phytopathogenic fungi including *C. parasitica* and *Valsa mali*, and *F. graminearum*. Interestingly HSVd induces severe growth defects in *V. mali*, but not in the other two fungi.

These examples of cross-kingdom infection provide another layer of research direction. The antiviral RNA silencing is the primary defense mechanism in fungi, invertebrates (such as insects, mites and nematodes) and plants, while key players and mechanisms involved are different in these host organisms. How single viruses counter-defense is an interesting problem to address. Movement proteins are essential for plant viruses to establish a systemic infection in plant hosts, but not for the systemic infection by fungal viruses of fungal hosts. Genome alterations that may occur during long-term maintenance exclusively single kingdom of hosts is worth monitoring. There are a few genera that accommodate fungal viruses and viruses infecting members of other kingdoms such as plants and insects. It will be of great interest to elaborate on their experimental host ranges.

Yeast as a (Model) Host to Study Viral Replication

S. cerevisiae has provided an excellent system to investigate virus assembly and replication of the dsRNA totivirus (ScV-L-A) and the ssRNA narnaviruses (ScNV-20S and ScNV-30S) that infect yeast. With the robust yeast genetics, a number of host factors involved in totivirus replication were identified, many of which are related to translation events (see article on the yeast ScV-L-A for details). Furthermore, the yeast provides an “artificial” viral host model system to explore host genes affecting viral replication on a genome-wide basis. Genetic screens of a collection of 4500–4800 single-gene deletion yeast strains (Yeast Knockout strain collection) have been successfully conducted for identifying host factors involved in replication and recombination with two different tri- and monopartite plant (+)ssRNA viruses, brome mosaic virus (family *Bromoviridae*) and tomato bushy stunt virus (TBSV, family *Tombusviridae*), respectively. Each screen led to the identification of approximately 100 host genes (approximately 1.8% of the entire yeast genes) that affect virus replication. Interestingly, the replication of the two viruses in yeast is affected by a different set of genes. A similar approach was employed to identify genes affecting the recombination of TBSV. This type of use of the yeast system can be expanded to an animal (+) ssRNA virus and a vertebrate (–)ssRNA virus. Recent interesting screening of bacterial effectors, whose targets are known, for antiviral effects in yeast revealed pro- and anti-viral host factors for TBSV replication. It may be surprising that the yeast has yet to be used for viruses infecting filamentous fungi. The yeast should be able to serve as a model host for mycoviruses other than those that naturally infect yeast.

Virus Neo-Lifestyles

Recent mycovirus hunting revealed unique virus lifestyles. The first such lifestyle (called “neo-lifestyle”) is exhibited by a (+) ssRNA mycovirus termed yadokari virus 1 (YkV1) hosted by a dsRNA mycovirus called yado-nushi virus 1 (YnV1). YkV1 encodes RdRP but depends on YnV1 for its encapsidation. The replication and transcription of YkV1 are hypothesized to occur in the YnV1 capsid encasing YkV1 RNA (heterocapsid), thus YkV1 behaving like a dsRNA virus. Intriguingly YkV1 enhances the accumulation of the partner YnV1. YnV1 as a full-fledged dsRNA can complete its replication cycle. There appear to be YkV1-like viruses showing partnership with dsRNA viruses in other filamentous fungi.

Another virus neo-lifestyle was discovered in *Aspergillus fumigatus* (one of the most ubiquitous saprophytic ascomycete) infected by a tetra-segmented RNA virus *Aspergillus fumigatus* polymycovirus 1 (AfuPmV1, family *Polymycoviridae*). AfuPmV1 RdRP has the presumed catalytic amino acid sequence, GDNQ, similar to those of (–)ssRNA viruses (most of mononegaviruses), but is phylogenetically closer to (+)ssRNA viruses such as caliciviruses than to (–)ssRNA viruses. AfuPmV1 is assumed to be associated in a colloidal form with a virally encoded proline-alanine-serine rich protein (PASrp) rather than being encased by capsid protein. More surprisingly, AfuPmV1 is infectious as a purified naked dsRNA as well as a dsRNA/PASrp complex that appears to prevent viral dsRNA from being exposed to RNase. A virus closely related to AfuPmV1 has recently been discovered which is termed Hadaka virus 1 (HadV1). However, HadV1 has an 11-segmented (+)ssRNA genome but lacks a PASrp-encoding segment. Unlike polymycoviruses, HadV1 appears to be accessible in infected mycelial homogenates by nuclease.

Host Defense Against Mycoviruses and Their Counter-Defense

Host defense responses against viruses have not been explored intensively. RNA silencing is regarded as one of the host defense strategies of eukaryotes against molecular parasites including viruses, and has been shown to operate as antiviral defense in a

number of fungi including important models and phytopathogenic fungi. As in the case of other eukaryotes, deficiencies of antiviral-RNA silencing in fungal hosts result in severer symptom induction and enhanced mycovirus replication. There are genetic elements involved in RNA silencing that are conserved widely from fungi to vertebrates. The functional roles of these factors in RNA silencing as antiviral reactions appear to be different between fungi. Two genes, one Dicer gene, dicer-like 2 (*dcl2*), and one Argonaute gene, argonaute-like 2 (*agl2*), are required for antiviral responses in *C. parasitica*, while more players are involved in other fungi such as *F. graminearum* and *S. sclerotiorum*, where at least two Dicer genes (*dcl1* and *dcl2*) and Argonaute genes (*ago1* and/or *ago2*) are involved in the defense redundantly.

The RNA silencing key genes, *agl2* and *dcl2*, of *C. parasitica* are highly induced by over 10-fold on infection by some mycoviruses. This transcription upregulation is mediated by the SAGA complex, a general transcriptional co-activator, that requires DCL2 that, thus, plays a positive feedback role. DCL2 is necessary for the SAGA-mediated upregulation of not only *agl2* and *dcl2* but also many other fungal host genes. Some of the induced host genes contribute to symptom mitigation. The elevated state of antiviral RNA silencing established by a pre-infected virus can eliminate horizontal transmission via hyphal fusion and replication of a second virus susceptible to RNA silencing.

The protein p29 encoded by hypovirus CHV1 has been shown to be a suppressor of RNA silencing targeting a transgene that functions in both plant and fungal cells. In fungal cells, CHV1 p29 suppresses the above mentioned transcriptional upregulation of antiviral silencing key genes (*dcl2* and *agl2*). Unraveling the mechanism by which the CHV1 p29, or other mycovirus-encoded RNA silencing suppressors, may block the RNA silencing pathway will be a very interesting challenge.

Other layers of host defense include vegetative incompatibility which impairs lateral virus transmission between colonies of a single species. This non-self allorecognition is governed genetically, in the case of *C. parasitica*, 6 diallelic loci (*vic*), and therefore, serves as a barrier against virocontrol. It is worth noting that super donor strains with deletions of multiple *vic* genes allow efficient mycovirus transmission at the field level. Non-self-responses include programmed cell death in which reactive oxygen species play a pivotal role. Some mycoviruses are able to suppress vegetative incompatibility reactions in fungi.

Role of Mycoviruses in Plant-Fungal Mutualistic Associations

The question of whether mycoviruses are involved in the mutualistic interactions between endophytic fungi and their host plants is of considerable interest because of the attractive beneficial features of these associations and because of the common occurrences of mycoviruses in all major groups of fungi. This question was recently addressed in an intriguing report that presented evidence for a dsRNA mycovirus being involved in the mutualistic interaction between a fungal endophyte (*Curvularia protuberata*, an ascomycete) and a tropical panic grass (*Dichanthelium lanuginosum*). This association allows both organisms to grow at high soil temperatures. The virus in question, which was designated *Curvularia thermal tolerance virus* (CThTV, the proposed family “*Curvulaviridae*”), has unusual genome organization with an unknown genome expression strategy. CThTV has apparently a bipartite dsRNA genome (RNA1 and RNA2), but no evidence that these RNAs are packaged in the 27-nm isometric particles isolated from the fungal host. Although many questions pertinent to the fungal endophyte, CThTV, and the veracity of the evidence for viral etiology remain unanswered, this report will undoubtedly stimulate the search for mycoviruses in other mutualistic fungal endophytes. In this regard, it is noteworthy that the well-characterized mutualistic ascomycete endophyte, *Epichloë festucae*, was found to harbor a totivirus, but no phenotypes were associated with virus infection.

S. sclerotiorum is a destructive ascomycetous pathogen in many dicotyledonous crops. However, the single-stranded DNA virus SsHADV1, known to induce hypovirulence in the host fungus *S. sclerotiorum*, causes a switch of the lifestyle of the fungal phytopathogen from a pathogenic to an endophytic status. Interestingly, treatments of rapeseed by SsHADV1-infected *S. sclerotiorum* promote plant growth compared to untreated plants, and has protective effects against virulent, virus-free fungal strains of *S. sclerotinia* but also its closely related species, *Botrytis cinerea*. These features of SsHADV1 together with horizontal transmissibility to vegetatively incompatible fungal strains make the virus a potentially robust virocontrol agent (see below). Surprisingly, *S. sclerotiorum* strain DT-8 infected by SsHADV1 grows endophytically in wheat (a monocotyledonous crop), previously believed to be a non-host of the fungus, and increases its grain yield. Furthermore, the fungal strain mitigates the symptoms induced by the *Fusarium* head blight disease (FHB) caused by *Fusarium graminearum* both in the laboratory and field conditions, likely through enhanced defense responses. Similar positive effects can be observed in other cereal monocotyledonous crops, such as rice and barley, against different ascomycetous pathogenic fungi such as *Magnaporthe oryzae*. Specific contribution of SsHADV-1 to this phenomenon is obscure, because these effects appear to be exerted whether *S. sclerotiorum* is infected by SsHADV1 or not.

Mycovirus as Biocontrol Agents and as Tools for Fundamental Studies

The hypovirulence phenotype in the chestnut blight fungus (*C. parasitica*) is an excellent and well-documented example of a mycoviral-induced phenotype that is currently being exploited for biological control. The debilitating disease of Victoria blight (*H. victorae*) and the disease phenotype of the Dutch elm disease fungus (*Ophiostoma novo-ulmi*, an ascomycete) are examples of pathogenic effects of dsRNA mycoviruses. The ssDNA virus, SsHADV has also attracted great attention because of its potential as a virocontrol (biocontrol) agent and its abilities to enter host fungal cells, be replicated in and be transmitted by the mycophagous fly. An understanding of the molecular basis of disease in these fungal-mycovirus systems would provide excellent opportunities for the development of novel biocontrol strategies of plant pathogenic fungi. Mycoviruses also continue to serve as versatile tools to study

the virulence of host fungi, as recognized in studies with the hypovirus/*Cryphonectria* system, in which substantial advances in our understanding of the molecular basis of hypovirulence have been made.

It should be noted that some mycoviruses enhance virulence (hypervirulence), sporulation, and mycelial growth. For example, virus-induced hypervirulence was observed in the *Nectria radicola* mycovirus system. The host fungus causes root rot in ginseng (*Panax quinquefolius*). Another example is the virus (a chrysovirus)-enhanced production of the host-specific toxin (AK-toxin) of *Alternaria alternata* (an ascomycete) that causes black spot in Japanese pear.

Further Reading

- Andika, I.B., Wei, S., Cao, C., *et al.*, 2017. Phytopathogenic fungus hosts a plant virus: A naturally occurring cross-kingdom viral infection. *Proceedings of the National Academy of Sciences of the United States of America* 114 (46), 12267–12272.
- Buck, K.W., 1998. Molecular variability of viruses of fungi. In: Bridge, P.D., Couteaudier, Y., Clarkson, J.M. (Eds.), *Molecular Variability of Fungal Pathogens* CAB International Wallingford 53–72.
- Ghabrial, S.A., 2001. Fungal viruses. In: Maloy, O., Murray, T. (Eds.), *Encyclopedia of Plant Pathology* 1. New York: John Wiley & Sons, pp. 478–483.
- Ghabrial, S.A., 2013. Mycoviruses. Academic Press. (2013/03/19 ed. *Adv Virus Res*, 86).
- Hillman, B.I., Aulia, A., Suzuki, N., 2018. Viruses of plant-interacting fungi. *Advances in Virus Research* 100, 99–116.
- Kondo, H., Kanematsu, S., Suzuki, N., 2013. Viruses of the white root rot fungus, *Rosellinia necatrix*. *Advances in Virus Research* 86, 177–214.
- Koonin, E.V., Dolja, V.V., 2018. Metaviromics: a tectonic shift in understanding virus evolution. *Virus Research* 246, A1–A3.
- Kotta-Loizou, I., Coutts, R.H.A., 2017. Mycoviruses in *Aspergilli*: A comprehensive review. *Frontiers in Microbiology* 8, 1699.
- Luque, D., Mata, C.P., Suzuki, N., Ghabrial, S.A., Caston, J.R., 2018. Capsid structure of dsRNA fungal viruses. *Viruses* 10 (9). doi:10.3390/v10090481.
- Moriyama, H., Urayama, S.I., Higashiura, T., Le, T.M., Komatsu, K., 2018. Chrysovirus in *Magnaporthe oryzae*. *Viruses* 10 (12). doi:10.3390/v10120697.
- Nuss, D.L., 2011. Mycoviruses, RNA silencing, and viral RNA recombination. *Advances in Virus Research* 80, 25–48.
- Rigling, D., Prospero, S., 2018. *Cryphonectria parasitica*, the causal agent of chestnut blight: Invasion history, population biology and disease control. *Molecular Plant Pathology* 19 (1), 7–20.
- Roossinck, M.J., 2019. Evolutionary and ecological links between plant and fungal viruses. *New Phytologist* 221 (1), 86–92.
- Son, M., Lee, K.M., Yu, J., *et al.*, 2013. The HEX1 gene of *Fusarium graminearum* is required for fungal asexual reproduction and pathogenesis and for efficient viral RNA accumulation of *Fusarium graminearum* virus 1. *Journal of Virology* 87 (18), 10356–10367.
- Sutela, S., Poimala, A., Vainio, E.J., 2019. Viruses of fungi and oomycetes in the soil environment. *FEMS Microbiology Ecology* 95 (9). doi:10.1093/femsec/fiz1119.
- Sutela, S., Forgia, M., Vainio, E.J., *et al.*, 2020. The virome from a collection of endomycorrhizal fungi reveals new viral taxa with unprecedented genome organization. *Virus Evolution* 6 (2). doi:10.1093/ve/veaa076.
- Turina, M., Ghignone, S., Astolfi, N., *et al.*, 2018. The virome of the arbuscular mycorrhizal fungus *Gigaspora margarita* reveals the first report of DNA fragments corresponding to replicating non-retroviral RNA viruses in fungi. *Environmental Microbiology* 20, 2012–2025. doi:10.1111/1462-2920.14060.
- Wu, S., Cheng, J., Fu, Y., *et al.*, 2017. Virus-mediated suppression of host non-self recognition facilitates horizontal transmission of heterologous viruses. *PLoS Pathogens* 13 (3), e1006234.
- Xie, J., Jiang, D., 2014. New insights into mycoviruses and exploration for the biological control of crop fungal diseases. *Annual Review of Phytopathology* 52, 45–68.
- Zhang, D.X., Nuss, D.L., 2016. Engineering super mycovirus donor strains of chestnut blight fungus by systematic disruption of multilocus vic genes. *Proceedings of the National Academy of Sciences of the United States of America* 113 (8), 2062–2067.
- Zhang, R., Hisano, S., Tan, A., *et al.*, 2016. A capsidless ssRNA virus hosted by an unrelated dsRNA virus. *Nature Microbiology* 1, 15001. doi:10.1038/NMICROBIOL.2015.1.
- Zhnag, H., Xie, J.T., Fu, Y.P., *et al.*, 2020. A 2-kb mycovirus converts a pathogenic fungus into a beneficial endophyte for Brassica protection and yield enhancement. *Molecular Plant* 13, 1420–1433.

Relevant Websites

<https://talk.ictvonline.org/files/master-species-lists/m/msl>
Master Species Lists.

Cross-Kingdom Virus Infection

Liying Sun, Northwest A&F University, Yangling, China

Hideki Kondo, Okayama University, Kurashiki, Japan

Ida Bagus Andika, Qingdao Agricultural University, Qingdao, China

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Viruses, like other obligate parasites, depend on host organisms for their multiplication and survival. As the hosts do not live forever, the continued existence of viruses in the population is maintained by transmission between individual organisms. The extent to which viruses spread is largely influenced by their host range, transmission pathways, and environmental conditions. The compatibility of virus-host cell interactions, host defenses and the ability of the viruses to counter or evade defense responses are critical in determining viral host range. Viruses can infect new hosts through genetic mutations that enhance their compatibility with host cellular components or enable them to evade host defenses. Furthermore, the mode of virus transmission is also a major factor controlling virus spread. As such, the mobility of biological vectors, human activities, and changes of environmental conditions may increase the possibility of contact between viruses and a potential novel host.

Viruses are known to spread between species; however, virus transfer across kingdoms is previously thought to be less common and naturally occurring cross-kingdom infection by a virus species is rarely documented. There are several classic examples of virus groups that infect hosts belonging to different taxonomical kingdoms, but they are mostly limited to some plant virus groups that replicate in their insect vectors. Nevertheless, certain viruses appear to have an inherent ability to infect different domains of life, as demonstrated by several reports of viruses artificially introduced to new hosts belonging to different biological kingdoms in the laboratory, particularly insect, plant, and fungal viruses. Moreover, several closely related viruses have been found to separately infect organisms of different kingdoms, such as some viruses infecting plants and fungi. This suggests horizontal virus transfer across kingdoms at some time in the past. Thus cross-kingdom viral infections represent an event that has happened before and might happen again.

The distinct biological characteristics of different kingdoms create natural barriers for virus horizontal transfer. Even so, agricultural practices and other human activities have caused drastic changes to the biological and ecological systems of the world. Such a situation might create conditions that conceivably could facilitate cross-kingdom viral infection in the environment. Such a scenario raises concerns for the possible emergence of new animal and plant diseases in the future.

This article summarizes cross-kingdom virus infection of plant, insect, and fungal viruses. The topics cover experimental, ecological and evolutionary aspects of virology.

Viruses That Replicate in Both Plants and Insects

There are some plant virus groups that are known to be transmitted in a persistent propagative manner by their arthropod vectors, and thus they replicate both in plant hosts and insect vectors (**Fig. 1**). They are predominantly negative-sense single-stranded RNA viruses ((–)ssRNA viruses; plant rhabdoviruses, tospoviruses, and tenuiviruses), and segmented double-stranded RNA viruses (dsRNA viruses; plant reoviruses). There are also a small portion of non-segmented positive-sense single-stranded RNA viruses ((+)ssRNA viruses; marafiviruses). Some studies showed that infection by certain plant viruses altered the performance of their insect vectors, such as for longevity, fecundity, and feeding behavior.

Classical plant rhabdoviruses (family *Rhabdoviridae*, order *Mononegavirales*, unsegmented (–)ssRNA viruses) belonging to the genera *Cytorhabdovirus* (type species *Lettuce necrotic yellows cytorhabdovirus*) and *Nucleorhabdovirus* (type species *Potato yellow dwarf nucleorhabdovirus*) infect diverse plants including cereal, maize and several vegetables. They are similarly transmitted by aphids (family *Aphididae*, order *Hemiptera*) such as the cabbage aphid (*Brevicoryne brassicae*) and the sowthistle aphid (*Hyperomyzus lactucae*), leafhoppers (family *Cicadellidae*, *Hemiptera*) such as the green rice leafhopper (*Nephotettix* spp.) and the black-faced leafhopper (*Graminella nigrifrons*) and planthoppers (family *Delphacidae*, *Hemiptera*) such as the small brown planthopper (*Laodelphax striatellus*) and the corn planthopper (*Peregrinus maidis*). Members of the newly established genus *Dichorhavirus* (type species *Orchid fleck dichorhavirus*, two-segmented (–)ssRNA viruses) are transmitted by false spider mites (*Brevipalpus* spp., *Tenuipalpidae*) and are also likely to replicate in their vectors.

Genus *Orthotospovirus* (formerly *Tospovirus*, type species *Tomato spotted wilt orthotospovirus*, segmented (–)ssRNA viruses) in the family *Tospoviridae* (order *Bunyavirales*) contains several members that infect a wide range of plants. They are exclusively transmitted by thrips (family *Thripidae*, order *Thysanoptera*) such as the western flower thrip (*Frankliniella occidentalis*), the onion thrip (*Thrips tabaci*), and the melon thrip (*T. palmi*). The plant hosts of the members of genus *Tenuivirus* (type species *Rice stripe tenuivirus*, segmented (–)ssRNA viruses) in the family *Phenuiviridae* (order *Bunyavirales*) are limited to the family *Poaceae* such as rice and maize, and these viruses are propagatively transmitted by planthoppers such as *L. striatellus*, *P. maidis*, and the brown planthopper, *Nilaparvata lugens*.

Plant-infecting reoviruses (segmented dsRNA viruses) are classified into 3 genera, *Phytoreovirus* (type species *Wound tumor virus*), *Fijivirus* (type species *Fiji disease virus*), and *Oryzavirus* (type species *Rice ragged stunt virus*) in the family *Reoviridae*, and they mainly

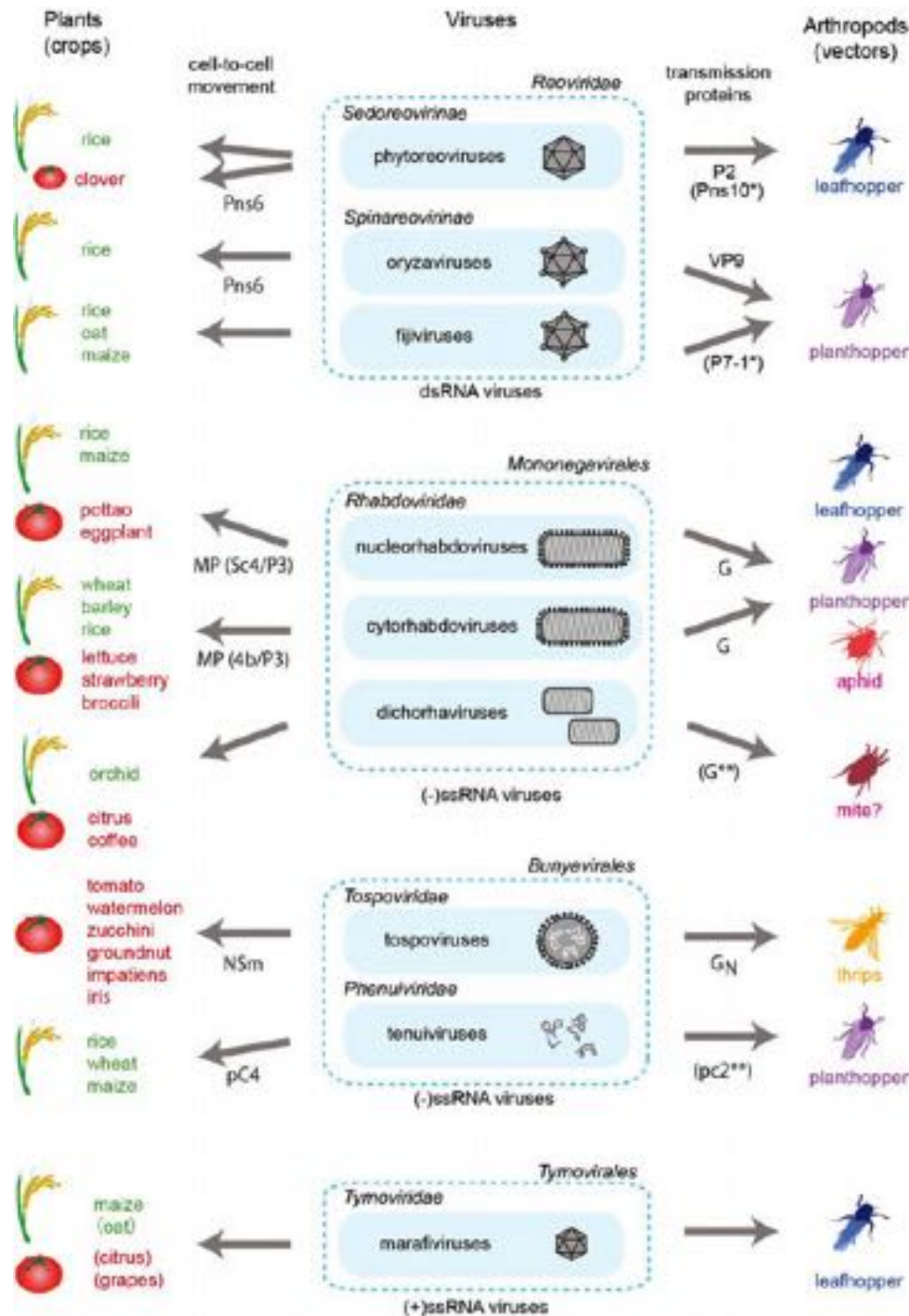


Fig. 1 Virus genera with members that are known to replicate in both plants and insect vectors. The virion structures of each virus genus are illustrated. Viral proteins required for cell-to-cell movement in the plant or vector transmission/infection to the insect vector are presented below the arrow. *Forms tubules facilitating viral spread among insect vector cells. **Possible glycoprotein whose biological function is unknown.

infect the plant family Poaceae. Fijiviruses and oryzaviruses (subfamily *Spinareovirinae*) are propagatively transmitted by planthoppers such as *L. striatellus*, *N. lugens*, and the white-backed planthopper (*Sogatella furcifera*). Phytoreoviruses (subfamily *Sedoreovirinae*, lacking the large surface projection of virus particles) are propagatively transmitted by leafhoppers, such as *Nephotettix* spp., *Agallia* spp., and *Recilia dorsalis*. Members of the genus *Mirafivirus* (type species *Maize rayado fino virus*, plant alpha-like viruses) in the family *Tymoviridae* (order *Tymovirales*) infect diverse host plants including maize, citrus and oat. The corn leafhopper (*Dalbulus maidis*) is known as an efficient vector for maize rayado fino virus. Interestingly, a recent report showed that tobacco ringspot virus (family

Secoviridae, order *Picornavirales*), a plant picorna-like virus with a two segmented (+)ssRNA genome, could replicate in honeybees, *Apis mellifera* (family *Apidae*, Hymenoptera) and spread throughout the entire body of the insect.

Orthospoviruses and classical plant rhabdoviruses have enveloped spherical (~80–110 nm in diameter) and bacilliform (~300–350 nm long and ~75–100 nm in diameter) virions, respectively. Viral membranes are likely required for the entry of viruses into insect cells but not into plant cells. Viral membrane glycoproteins of plant nucleo- and cytorhabdoviruses (G protein) and orthospoviruses (G_N and G_C proteins) are embedded in the outer-membrane envelope and they are believed to be involved in virus entry, interacting with receptor proteins on the insect cell surface. Similarly, P2 and VP9 encoded by phyto-reoviruses and oryzaviruses, respectively, are components of the outer capsid protein that bulge from the surface of the double-shelled virion (~60–100 nm in diameter) and are involved in infection of insect. In addition, these plant- and insect-infecting viruses encode cell-to-cell movement proteins that are specifically required for spread of viruses through host plants (Fig. 1). Notably, an insect small RNA virus, flock house virus (FHV, family *Nodaviridae*, a small non-enveloped spherical virion with a two-segmented (+) ssRNA genome) is able to replicate in mechanically inoculated leaves of several plant species; however, systemic spread of FHV throughout the plant requires expression of a plant virus-encoded cell-to-cell movement protein. Therefore, evidence suggests that these propagatively insect-transmitted viruses are characterized by biological properties required for successful infection in both plant and insect hosts.

Closely Related Viruses Independently Infect Plants and Fungi

There has been a surge of research interest in characterizing and studying fungal viruses, particularly those infecting plant pathogenic fungi. A growing number of newly identified fungal viruses contribute greatly to our understanding of virus diversity and evolution. For example, a number of fungal viruses were found to have close similarities with plant viruses, suggesting horizontal transfer of these viruses between plants and fungi in the not so distant past. One of the most striking observations is that some plant and fungal partitiviruses (family *Partitiviridae*, picorna-like viruses with a segmented dsRNA genome and non-enveloped spherical virion, 25–43 nm in diameter) show close sequence similarities. Phylogenetic analysis shows that plant and fungal partitiviruses were placed together in some clades of alpha- and betapartitiviruses, in addition to clades containing either plant or fungal partitiviruses (Fig. 2(A)). Therefore, some partitiviruses from different kingdoms, plant and fungi, are more similar to each other than to those from within a single kingdom. Additionally, the similarity between plant- and fungus (or oomycetes, including *Phytophthora* spp.) infecting endornaviruses (family *Endornaviridae*, capsid-less alpha-like (+)ssRNA viruses) has been observed. The members of plant partiti- and endornaviruses can establish persistent infections in host cells but lack a typical cell-to-cell movement protein. It is proposed that viruses that cause persistent infections of plants, such as partitiviruses, could have originated from fungal viruses that infect plant associated fungi, including endophytes.

Another finding is that a fungal alpha-like virus, Botrytis virus X (genus *Botrexvirus*, family *Alphaflexiviridae*, *Tymovirales*, a filamentous virion with non-segmented (+)ssRNA genome) isolated from the gray mold fungus *Botrytis cinerea* (family *Sclerotiniaceae*), a necrotrophic plant pathogenic ascomycete, shows close sequence similarity to allxiviruses (genus *Allexivirus*, *Alphaflexiviridae*, plant alpha-like viruses with a filamentous virion) that encode a gene block for a cell-to-cell movement protein and mainly infect the plant genus *Allium* (garlic and onion; Fig. 2(B)). As *B. cinerea* commonly infects horticulture plants including *Allium* spp., it is speculated that *B. cinerea* might have acquired a plant allxivirus or a progenitor during an infection.

Artificially Established Viral Cross Infections Between Plants and Fungi

Pioneering work by Paul Ahlquist established a replication system for a plant alpha-like virus, brome mosaic virus (BMV, family *Bromoviridae*, a tripartite (+)ssRNA genome) in yeast *Saccharomyces cerevisiae*. This has since been achieved for other plant viruses, including tomato bushy stunt (TBSV), Cymbidium ringspot (CymRV), and carnation Italian ringspot viruses (CIRV) (family *Tombusviridae*, with a non-segmented (+)ssRNA genome; Table 1). Replication of these viruses takes place on intracellular membranes derived from organelles in both plant and yeast cells; BMV uses endoplasmic reticulum (ER) membranes, TBSV and CymRV use peroxisomal membranes and CIRV uses mitochondrial membranes. In addition, in the absence of the peroxisomal membrane, TBSV and CymRV can utilize ER membranes for their replication in yeast. Notably, some animal (+)ssRNA viruses (FHV and Nodamura virus, a nodavirus) and DNA viruses (human and bovine papillomaviruses, family *Papillomaviridae*, and adeno-associated virus, family *Parvoviridae*, circular dsDNA and ssDNA genome, respectively) were also shown to replicate in yeast cells.

More recently, some reports show replication of two plant alpha-like viruses, tobacco mosaic virus (TMV, family *Virgaviridae*, non-segmented (+)ssRNA genome) and cucumber mosaic virus (CMV, *Bromoviridae*, tripartite (+)ssRNA genome; see also below), in plant pathogenic filamentous fungi (Table 1). However, it was also shown that CMV was not able to stably infect two other plant pathogenic filamentous fungi, *Cryphonectria parasitica* (family *Cryphonectriaceae*, *Diaporthales*) and *Fusarium graminearum* (family *Nectriaceae*), causal agents of chestnut blight and fusarium head blight diseases, respectively. This demonstrated that there was a fungal-specific host compatibility of CMV infection. TMV virion (non-enveloped rod shape particles, 300 nm long and 18 nm in diameter) enters the fungal hyphae and/or conidia during culture by an unknown mechanism, and virus replication does not alter the pathogenicity and growth of the host fungus. Conversely, two fungal viruses, *Penicillium aurantiogriseum* partiti-like virus 1 and *Penicillium aurantiogriseum* totivirus 1, from a marine plant endophyte (*Penicillium aurantiogriseum*, family

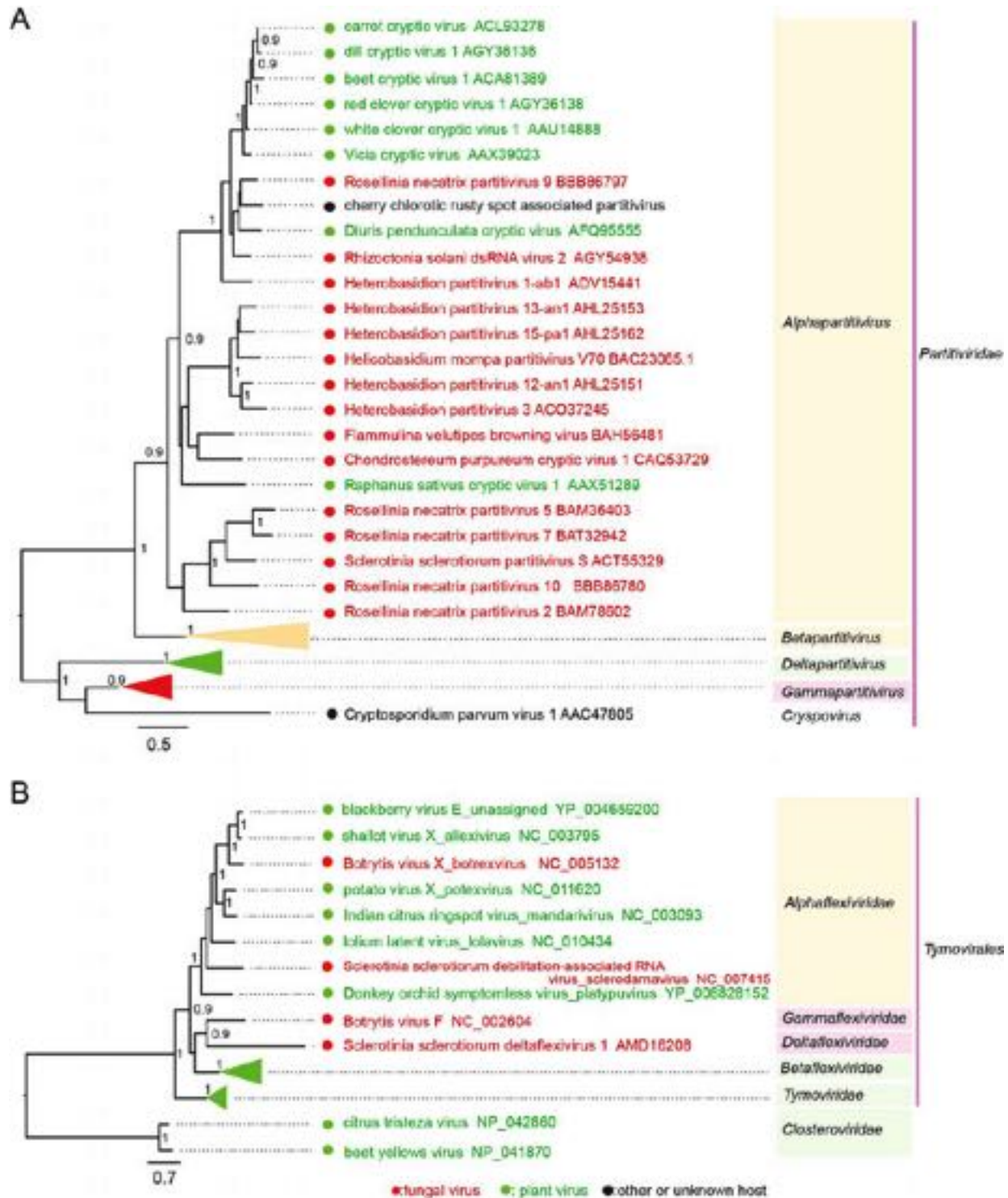


Fig. 2 Phylogenetic relationships of selected plant- or fungus-infecting viruses belonging to families *Partitiviridae* (A) and *Tymoviridae* (B). Maximum likelihood phylogenetic trees were constructed using PhyML 3.0 based on the multiple amino acid sequence alignment of the replicase protein or its candidate sequences. A model LG + I + G + F was selected as the best-fit model for the tree construction. Clades for beta-, gamma-, deltapartitviruses, betaflexi- and tymoviruses are shown by collapsed branches with triangles.

Table 1 The list of viruses used for artificial cross-kingdom infections between plants and fungi

Virus	Original host	New host	Method of virus introduction
Brome mosaic virus (genus <i>Bromovirus</i>)	Plant (<i>Bromus inermis</i> and other grasses)	<i>Saccharomyces cerevisiae</i>	Cell/spheroplast transformation
Tomato bushy stunt virus (genus <i>Tombusvirus</i>)	Plant (mainly vegetables crop and ornamental plants)	<i>S. cerevisiae</i>	Cell/spheroplast transformation
Carnation Italian ringspot virus (genus <i>Tombusvirus</i>)	Plant (some ornamental plants and fruit trees)	<i>S. cerevisiae</i>	Cell/spheroplast transformation
Cymbidium ringspot virus (genus <i>Tombusviruse</i>)	Plant (<i>Cymbidium</i> and clover)	<i>S. cerevisiae</i>	Cell/spheroplast transformation
Mung bean yellow mosaic India virus (genus <i>Begomovirus</i> , family <i>Geminiviridae</i>)	Plant (legume)	<i>S. cerevisiae</i>	Cell/spheroplast transformation
Tobacco mosaic virus (genus <i>Tobamovirus</i>)	Plant (a very wide host range)	<i>Colletotrichum acutatum</i> , <i>C. clavatum</i> , <i>C. theobromicola</i>	Virus particles added to liquid medium
Cucumber mosaic virus (genus <i>Cucumovirus</i>)	Plant (a very wide host range)	<i>Rhizoctonia solani</i> , <i>Valsa mali</i>	Spheroplast transformation
Penicillium aurantiogriseum totivirus 1 (unclassified toti-like virus)	<i>Penicillium aurantiogriseum</i>	<i>Nicotiana benthamiana</i> , <i>N. benthamiana</i> expressing a silencing suppressor (HC-Pro), <i>N. tabacum</i> (BY2 culture cells)	Protoplast transformation
Penicillium aurantiogriseum partiti-like virus (unclassified partiti-like virus)	<i>Penicillium aurantiogriseum</i>	<i>N. benthamiana</i> expressing a silencing suppressor (HC-Pro), <i>N. tabacum</i> (BY2 culture cells)	Protoplast transformation

Trichocomaceae) can replicate in plant cells (Table 1). These reports demonstrate the compatibility between certain plant viruses and fungal cells, as well as between certain fungal viruses and plant cells.

Evidence of the Natural Transmission of a Plant Virus to a Fungus

The first report of the natural infection of a fungus with a plant virus was the discovery of CMV (plant alpha-like virus) infecting a field strain of the plant pathogenic basidiomycete fungus, *Rhizoctonia solani* (family Ceratobasidiaceae). During an attempt to identify fungal viruses infecting several strains of *R. solani* collected from potato fields, one fungal strain was found to be coinfecting with CMV, in addition to other fungal viruses. Further analysis confirmed that CMV infection was stable in this fungal strain and the virus was horizontally transmitted through hyphal anastomosis, but not vertically through basidiospores. In addition, CMV was also shown to replicate in an ascomycete fungus, the apple canker fungus (*Valsa mali*, family Valsaceae, Diaporthales). Therefore, many phytopathogenic fungi may support stable infection of this virus. Under laboratory conditions, it was demonstrated that *R. solani* could acquire CMV from virus-infected potato and *Nicotiana benthamiana*, as well as transmit the virus to uninfected plants. This supported the notion that CMV was transferred from plant to *R. solani* during infection in the natural environment. Interestingly, CMV infection has no effect on *R. solani* growth and morphology in artificial culture media, but infection of plants by the *R. solani* strain carrying CMV induces more severe symptoms than the virus-free strain. Thus, acquiring CMV seems to make *R. solani* more aggressive which has implications for disease control.

It is still not clear how CMV transfers between plant and fungi during infection. As a necrotrophic pathogen, *R. solani* decomposes the host cells for uptake of nutrients. This possibly allows the transfer of relatively big molecules, such as CMV particles (~28 nm in diameter), from the plant cell into fungal cells. Because CMV is infectious in the form of RNA, it is also possible that CMV is transmitted in the form of naked RNA or ribonucleoprotein.

A Fungal DNA Virus That Replicates in an Insect and Uses It as a Vector

Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1), the first DNA virus discovered in fungi, was isolated from a hypovirulent strain of *Sclerotinia sclerotiorum* (family Sclerotiniaceae), an ascomycete fungal plant pathogen known as the white mold fungus. SsHADV-1 belongs to the genus *Gemycircularvirus* (family *Genomoviridae*, a circular ssDNA genome and non-enveloped spherical particles about 20 nm in diameter) whose members are found in various environments, as well as in plant- and animal-associated samples. It was observed that the mushroom fly *Lycoriella ingenua* (family *Sciaridae*, order *Diptera*), an insect that feeds on fungi, could acquire SsHADV-1 when the larvae were fed on virus-infected fungal colonies. SsHADV-1 replicates in *L. ingenua* cells and the virus is transmitted vertically to *L. ingenua* offspring. Under laboratory conditions, *L. ingenua* could transmit SsHADV-1 to virus-free *S. sclerotiorum* and SsHADV-1 was detected in adult *L. ingenua* collected from the field, suggesting that SsHADV-1 could use *L. ingenua* as a vector.

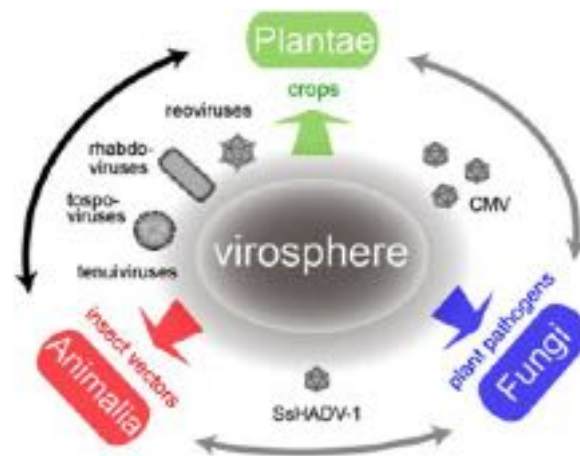


Fig. 3 A diagram showing known examples of natural transmission of viruses among animal, plant and fungal kingdoms. CMV, cucumber mosaic virus; SsHADV-1, *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1.

Given that SsHADV-1 was also detected in two dragonfly species (*Erythemis simplicicollis* and *Pantala hymenaea*, family Libellulidae) and one damselfly species (*Ischnura ramburii*, family Coenagrionidae; order Odonata), and SsHADV-1-related viruses were found in other insects, it is possible that other insects in addition to *L. ingenua* are natural hosts of SsHADV-1. Many insects feed on fungi or are infected with fungi; therefore, horizontal transfer of viruses between insect and fungus may not be specific, but could occur frequently in nature.

The Contribution of Cross-Kingdom Viral Infection to Virus Evolution

The current understanding of RNA virus lineages signifies pervasive horizontal virus transfer between diverse hosts as critical points in RNA virus evolution. Instances where RNA virus groups infect diverse hosts, including organisms of different taxonomic kingdoms includes the ourmiaviruses, plant infecting tripartite (+)ssRNA viruses (genus *Ourmiavirus*, family *Botourmiaviridae*, non-enveloped bacilliform virions with discrete lengths from 30 to 62 nm), that are related to narnaviruses (genus *Narnavirus*, family *Narnaviridae*, capsid-less (+)ssRNA viruses) infecting fungi (budding yeast and filamentous ascomycetes) and oomycete (*Phytophthora infestans*, family Peronosporaceae), and ourmia-like viruses recently discovered from invertebrate and fungi. Thus, ourmiaviruses seem to have evolved from horizontal transfer of narnaviruses or ourmia-like viruses to plants together with the acquisition of genes responsible for viral cell-to-cell movement and capsid formation. *Secoviridae* (non-segmented (+)ssRNA viruses with non-enveloped spherical particles 25–30 nm in diameter), a large virus family limited to plant hosts, belongs to the clade of picorna-like viruses (non-segmented (+)ssRNA genome) infecting a variety of invertebrates, mainly arthropods, suggesting ancestral secoviruses evolved in their host plants after horizontal transfer from arthropods. Furthermore, the genome segmentation of their ancestor may have occurred simultaneously with this event except for the members of genera *Waikavirus* and *Sequivirus*. Likewise, plant rhabdoviruses (family *Rhabdoviridae*, (–)ssRNA viruses) are most likely descended from insect rhabdoviruses because they tend to group according to their insect vectors rather than their host plants. Interestingly, the segmented-genome type of rhabdoviruses (dichorhaviruses), which is similar to the segmented genome feature of many plant viruses, is only known to exist in plants, suggesting evolution of rhabdoviruses in their host plants after horizontal transfer. Recently established viral phylogenies using expansive virus metagenomic data showed unprecedented clustering of plant and invertebrate virus groups, further implying pervasive horizontal virus transfers from invertebrates to plants in the distant past.

Concluding Remarks

Increasing evidence supports the view that some viruses can infect hosts from different kingdoms such as insects, plants and fungi and that may be rather common in nature. Viral phylogenies suggest that virus transmission across kingdom barriers has occurred in the past. In addition, artificial virus inoculation methods have demonstrated that certain viruses are able to replicate in both plants and fungi. Most importantly, examples of natural virus transmissions between insects and plants or fungi, as well as between plants and fungi, have been documented (Fig. 3). Further experimental and ecological studies on cross-kingdom virus infections would deepen our understanding of the origin, host range, dissemination and reservoir of viruses in the environment.

Further Reading

- Andika, I.B., Wei, S., Cao, C., *et al.*, 2017. Phytopathogenic fungus hosts a plant virus: a naturally occurring cross-kingdom viral infection. *Proceedings of the National Academy of Sciences of the United States of America* 114, 12267–12272.
- Balique, F., Lecoq, H., Raoult, D., Colson, P., 2015. Can plant viruses cross the kingdom border and be pathogenic to humans? *Viruses* 7, 2074–2098.
- Dasgupta, R., Garcia, B.H., Goodman, R.M., 2001. Systemic spread of an RNA insect virus in plants expressing plant viral movement protein genes. *Proceedings of the National Academy of Sciences of the United States of America* 98, 4910–4915.
- Dolja, V.V., Koonin, E.V., 2018. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Research* 244, 36–52.
- Hogenhout, S.A., Ammar, E.D., Whitfield, A.E., Redinbaugh, M.G., 2008. Insect vector interactions with persistently transmitted viruses. *Annual Review of Phytopathology* 46, 327–359.
- Liu, S., Xie, J., Cheng, J., *et al.*, 2016. Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. *Proceedings of the National Academy of Sciences of the United States of America* 113, 12803–12808.
- Mascia, T., Nigro, F., Abdallah, A., *et al.*, 2014. Gene silencing and gene expression in phytopathogenic fungi using a plant virus vector. *Proceedings of the National Academy of Sciences of the United States of America* 111, 4291–4296.
- Nagy, P.D., 2008. Yeast as a model host to explore plant virus-host interactions. *Annual Review of Phytopathology* 46, 217–242.
- Nerva, L., Varese, G.C., Falk, B.W., Turina, M., 2017. Mycoviruses of an endophytic fungus can replicate in plant cells: Evolutionary implications. *Scientific Reports* 7, 1908.
- Roossinck, M.J., 2019. Evolutionary and ecological links between plant and fungal viruses. *New Phytologist* 221, 86–92. doi:10.1111/nph.15364.
- Shi, M., Lin, X.D., Tian, J.H., *et al.*, 2016. Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543.
- Zhao, R.Y., 2017. Yeast for virus research. *Microbial Cell* 4, 311–330.

Diversity of Mycoviruses in *Aspergilli*

Ioly Kotta-Loizou, Imperial College London, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Anamorph The asexual reproductive stage of a fungus.

Anastomosis Fusion between branches of hyphae.

Ascocarp Fruiting body of a fungus belonging to the phylum Ascomycota, containing asci filled with ascospores.

Ascospore Sexual spore of a fungus belonging to the phylum Ascomycota, produced inside a sac called an ascus.

Conidiospore Asexual spore of a fungus.

Heterokaryon A multinucleate fungal cell that contains genetically different nuclei.

Heterothallic fungus Fungus not possessing, within a single individual, the resources to reproduce sexually but requiring two distinct individuals of opposite mating types.

Homothallic fungus Fungus possessing, within a single organism, the resources to reproduce sexually.

Teleomorph The sexual reproductive stage of a fungus.

Introduction

Aspergillus is a genus of filamentous fungi belonging to the order Eurotiales, class Eurotiomycetes, phylum Ascomycota. Genus *Aspergillus* includes the anamorphs (asexual stage) of more than 250 species, although some of them are known to also have a teleomorph (sexual stage) belonging to a different genus. For example, *Emericella nidulans* is the teleomorph of *A. nidulans*. Anamorphic fungi reproduce by producing conidiospores, asexual spores, while teleomorphs belonging to the phylum Ascomycota develop ascocarps, fruiting bodies producing ascospores, sexual spores. *Aspergilli* have a worldwide distribution, are ubiquitous and able to occupy numerous niches in terms of temperature, osmotic stress, and nutrient availability, with most of them being saprophytic. Individual *Aspergillus* species are medically, ecologically, and economically important and *Aspergillus* is one of the most well studied fungi in the laboratory; for example, *A. nidulans* was used as a model organism to study fungal genetics and the parasexual cycle in the 1950s and it was also one of the first fungi to have its genome fully sequenced in 2005.

To date, twenty taxa, nineteen families (*Alphaflexiviridae*, *Amalgaviridae*, *Barnaviridae*, *Botourmiaviridae*, *Chrysoviridae*, *Deltaflexiviridae*, *Endornaviridae*, *Gammaflexiviridae*, *Genomoviridae*, *Hypoviridae*, *Megabirnaviridae*, *Metaviridae*, *Mymonaviridae*, *Narnaviridae*, *Partitiviridae*, *Pseudoviridae*, *Reoviridae*, *Totiviridae*, *Quadriviridae*) and one genus that does not belong to a family (*Botybirnavirus*), are officially recognized by the International Committee for the Taxonomy of Viruses (ICTV) to accommodate, exclusively or not, mycoviruses. With the exception of *Genomoviridae*, mycoviruses have linear double-stranded (ds) or single-stranded (ss) RNA genomes. Out of these twenty recognized taxa, only members of the families *Chrysoviridae*, *Partitiviridae*, *Narnaviridae*, and *Totiviridae* have been discovered in *Aspergilli* to date. Other mycoviruses identified that could not be accommodated to the already existing families were often provisionally assigned to proposed families, e.g., *Polymycoviridae*. In general, RNA mycoviruses do not have any known extracellular phase in their replication cycle; they can be transmitted vertically, from parent to offspring, *via* conidiospores and ascospores, and horizontally, from one fungal isolate to another, *via* anastomosis/hyphal fusion. Mycovirus infections are persistent and, once infected, it is very difficult for fungal hosts to eliminate mycoviruses. The majority of mycovirus infections are also asymptomatic, latent, or cryptic.

Prevalence

In *Aspergilli*, the presence of mycoviruses was initially reported in 1970, when spherical virus-like particles (VLPs) containing dsRNA were discovered in *A. foetidus* and *A. niger*. Notably, the mycoviruses infecting *A. foetidus* strain IMI 41871, known as the *A. foetidus* mycovirus complex, are among few fully described at the molecular level, although their characterization was completed only recently, 45 years following their initial discovery.

Since then, investigations have revealed the presence of dsRNA elements, often accompanied by VLPs, in *Aspergillus* sections Nigri, Circumdati, Flavi, and Fumigati. In the cases where large populations of fungi were examined, the percentage of infected isolates ranged from less than 10% to almost 50%, often dependent on the geographic origin of isolates from the same species. The number of dsRNA elements present in each isolate varied from 1 up to 9, suggesting that in at least some of the cases more than one mycovirus was present. The dsRNA elements ranged from less than 0.5 kbp to 10 kbp in size; however, in most instances, the dsRNA elements were not characterized at the molecular level and therefore no conclusions can be drawn regarding the evolutionary relationships among similarly sized segments. Finally, the diameter of the VLPs ranged from 23 to 40 nm.

In *Aspergillus* section Nigri or 'black *Aspergilli*', apart from *A. foetidus* and *A. niger*, the presence of mycoviruses has been reported in *A. awamori*, *A. carbonarius*, *A. heteromorphus*, *A. japonicus*, and *A. tubingensis*. The prevalence of mycoviruses appears to be particularly high in Indonesian isolates. In *Aspergillus* section Flavi, *A. flavus*, *A. leporis*, *A. nomius*, *A. parasiticus* and *A. tamarii*, are known to harbor mycoviruses. In *Aspergillus* section Circumdati, mycoviruses have been found in *A. ochraceus* and *A. petrakii*;

Aspergillus ochraceus virus (AoV) from strain FA0611 has been fully sequenced. Finally, mycoviruses have been reported in both clinical and environmental isolates of *A. fumigatus*. Six of them, *Aspergillus fumigatus* chrysovirus (AfuCV), *Aspergillus fumigatus* partitivirus 1 (AfuPV-1), *Aspergillus fumigatus* partitivirus 2 (AfuPV-2), *Aspergillus fumigatus* narnavirus 1 (AfuNV-1), *Aspergillus fumigatus* narnavirus 2 (AfuNV-2), *Aspergillus fumigatus* mitovirus 1 (AfuMV-1), and *Aspergillus fumigatus* tetramycovirus 1 (AfuTmV-1) have been characterized at the molecular level.

Despite the relative high prevalence of mycoviruses in a number of anamorphic *Aspergilli*, their reported presence in the teleomorphs of the genus *Aspergillus* is much rarer. No dsRNA elements were discovered in a large *E. nidulans* (teleomorph of *A. nidulans*) panel derived from worldwide locations. Nevertheless sporadic mycovirus infections have been reported for *Neosartorya hiratsukae*, *N. quadricincta* (teleomorphs of *Aspergillus* section Fumigati) and *Petromyces alliaceus* (teleomorph of *A. alliaceus*, *Aspergillus* section Flavi).

Additionally, it should be noted that the quantity and even the presence of mycoviruses at detectable levels often depend on the composition of the nutrients in the growth medium and the developmental stage of the fungus. Regarding the latter, there have been a range of reports: in some cases virus replication correlates with the abundance of nutrients in the exponential phase of fungal growth while in others the virus titer is higher at the more advanced stages of the life cycle of the fungus. These results should not necessarily be considered conflicting, since they were derived from different fungal species most likely harboring distinct mycoviruses.

Transmission

Over the years, *Aspergilli* have been used as a model system for investigating all aspects of mycovirus transmission, including vertical transmission *via* conidiospores and ascospores, horizontal transmission *via* anastomosis/hyphal fusion, the role of heterokaryon incompatibility between fungal isolates as a barrier to mycovirus spread and intraspecies transmission.

During the formation of conidiospores from the tip of a differentiated hyphae called the conidiophore, mycoviruses present in the cytoplasm can be trapped within the conidial cell wall. Subsequently, the conidiospores are released from the conidiophore, disperse through the air and eventually give rise to a new fungal colony which is infected with the mycovirus. All relevant studies have shown that vertical transmission of mycoviruses *via* conidiospores is highly efficient in *Aspergilli*. This ensures the maximal dispersion of mycoviruses, which, having no extracellular phase in their replication cycle depend on their host for movement, together with their survival under harsh environmental conditions when the cost-efficient conidiation is the reproductive method of choice.

In contrast, vertical transmission of mycoviruses *via* ascospores is far less efficient. For instance, dsRNA elements introduced in *E. nidulans* (teleomorph of *A. nidulans*) were present exclusively in ascospores produced by self-fertilization and not following crossing with a compatible partner. This observation explains the paucity of mycoviruses in *E. nidulans* and was also confirmed for *N. hiratsukae*. The wider implication is that mycoviruses would not be transmitted *via* ascospores by *Aspergillus* teleomorphs which are exclusively heterothallic (unable to self-fertilize) and not homothallic.

Horizontal transmission of mycoviruses from one fungal strain to another requires their physical proximity and the fusion of their hyphae, a process called anastomosis. Following anastomosis, the cytoplasm, the nuclei and other organelles of two fungal strains, together with any mycoviruses present, co-exist in a joint mycelium called a heterokaryon. When the heterokaryon breaks down to homokaryons, either by forming distinct uninucleate sectors or by conidiation, mycoviruses may be inherited from a previously virus-free strain. In *Aspergilli*, the co-existence of multiple different mycoviruses in the same host is possible; however already infected isolates may be able to accommodate additional dsRNA elements, potentially due to limiting competition with the established mycoviruses.

Notably, not all fungal strains can fuse their hyphae and form heterokarya and this heterokaryon (or vegetative) incompatibility may restrict mycovirus transmission. The (in)compatibility of fungal strains depends on a number of nuclear genes whose specific alleles determine whether anastomosis should be allowed. Nevertheless, no correlation between the presence of dsRNAs and the isolates' mitochondrial RFLP haplotype, specific conserved surface protein types or mating type MAT1-1 and MAT1-2 have been noted. In the case of *A. niger*, mycovirus transfer was possible between isogenic or vegetatively compatible strains following generation and fusion of protoplasts in the laboratory. Mechanical disturbance of mycelia leading to released cytoplasm was necessary to allow mycovirus transfer between vegetatively incompatible strains. In contrast, heterokaryon incompatibility did not prevent mycovirus transfer between *A. nidulans* strains.

The potential of interspecies transmission of dsRNA elements has also been investigated in the laboratory and it has been successfully demonstrated between *A. niger* (as mycovirus donor) and *A. ficuum*, *A. oryzae*, *A. tubingensis*, *A. nidulans*, and in addition to the aforementioned hyphomycetes, the yeast *Saccharomyces cerevisiae* (as mycovirus recipients).

Genomes

To date, only a few mycoviruses from *Aspergilli* have had their genomes fully sequenced and annotated. These include the *A. foetidus* mycovirus complex, comprising of a member of the family *Totiviridae* and two as of yet unclassified viruses, members of the families *Chrysoviridae*, *Partitiviridae* and *Narnaviridae* and a member of the provisionally designated family *Polymycoviridae* (Fig. 1, Table 1). Families *Totiviridae*, *Chrysoviridae*, and *Partitiviridae* accommodate mycoviruses with dsRNA genomes within isometric virions composed of viral proteins. More specifically, *Chrysoviridae*, and *Partitiviridae* have four and two genomic segments, respectively, each one

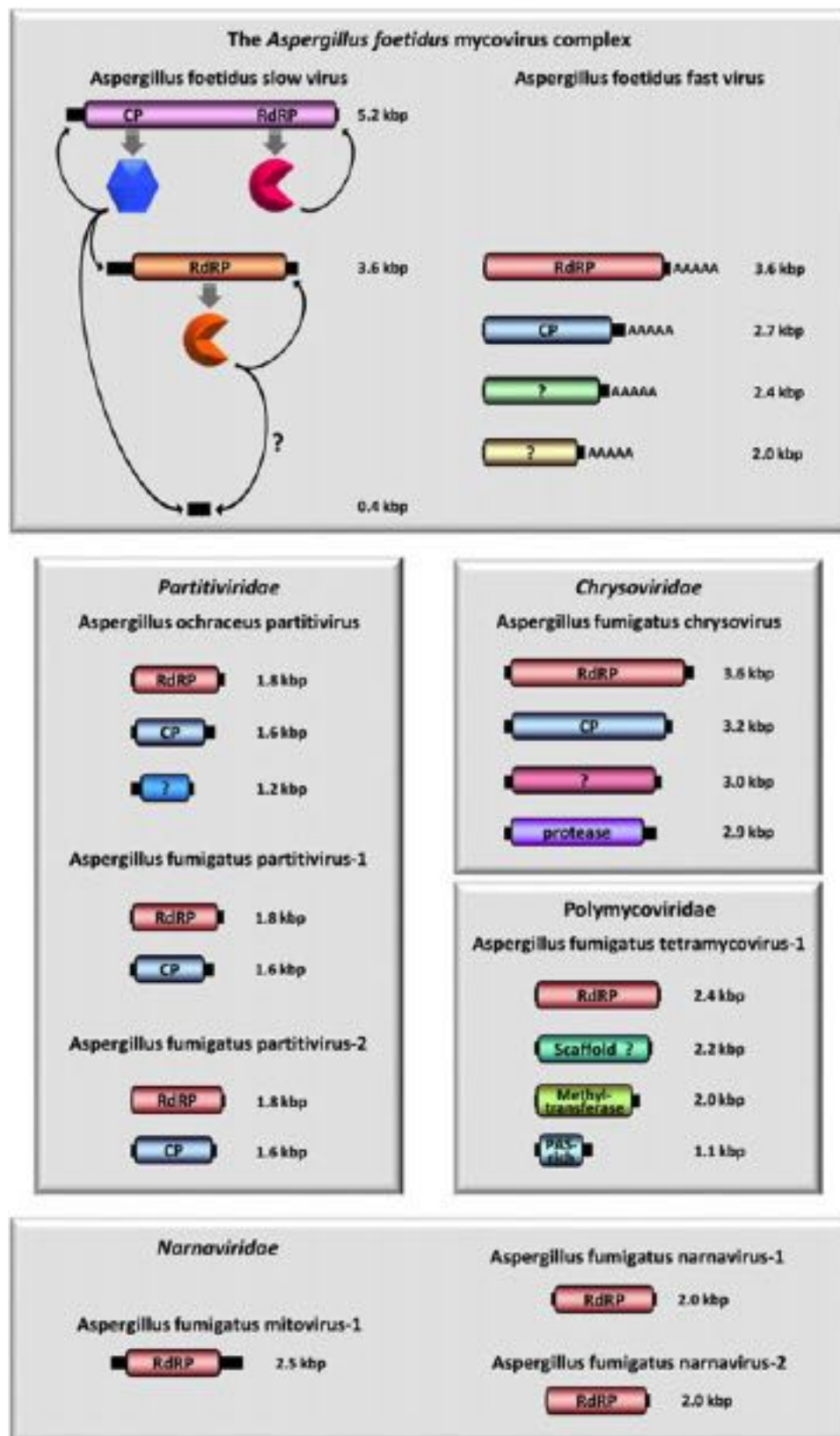


Fig. 1 Schematic representation of the genomic organization of all fully sequenced *Aspergillus* mycoviruses. For each genomic segment, the ORF(s) (colored boxes) are flanked by 5'- and 3'-UTRs (black boxes) and the function of the encoded protein(s) is indicated.

Table 1 Sequence properties of known *Aspergillus* mycoviruses

Virus name	Abbreviation	Original host	Segment (bp)	ORF size (nt; aa; kDa)	UTR length (bp)	
					5'-UTR	3'-UTR
Aspergillus foetidus fast virus	AfV-FRN	<i>A. foetidus</i> IMI 41871	dsRNA 1 (3571)	3375; 1124; 127	51	145
			dsRNA 2 (2734)	2406; 801; 87	48	280
			dsRNA 3 (2418)	2181; 726; 79	50	187
			dsRNA 4 (1961)	1743; 580; 65	50	147
Aspergillus foetidus slow virus 1	AfV-S1	<i>A. foetidus</i> IMI 41871	dsRNA 1 (5194)	4745; 741/839; 78/92	373	76
Aspergillus foetidus slow virus 2	AfV-S2	<i>A. foetidus</i> IMI 41871	dsRNA 1 (3634)	2889; 962; 110	509	236
Aspergillus ochraceus partitivirus	AoPV	<i>A. ochraceus</i> FA0611	dsRNA 1 (1754)	1620; 539; 62	66	68
			dsRNA 2 (1555)	1303; 433; 47	100	153
			dsRNA 3 (1220)	882; 293; 34	180	158
Aspergillus fumigatus partitivirus – 1	AfuPV – 1	<i>A. fumigatus</i> 88	dsRNA 1 (1779)	1629; 542; 63	65	85
			dsRNA 2 (1623)	1329; 442; 48	104	190
Aspergillus fumigatus partitivirus – 2	AfuPV – 2	<i>A. fumigatus</i> V145–13	dsRNA 1 (1822)	1722; 573; 66	34	66
			dsRNA 2 (1638)	1452; 483; 55	95	91
Aspergillus fumigatus chrysovirus	AfuCV	<i>A. fumigatus</i> A – 56	dsRNA 1 (3560)	3345; 1114; 129	128	87
			dsRNA 2 (3159)	2862; 953; 107	167	130
			dsRNA 3 (3006)	2681; 891; 99	169	156
			dsRNA 4 (2863)	2544; 847; 95	154	165
Aspergillus fumigatus narnavirus – 1	AfuNV – 1	<i>A. fumigatus</i> V145–43	dsRNA 1 (2007)	1857; 618; 70	78	72
Aspergillus fumigatus narnavirus – 2	AfuNV – 2	<i>A. fumigatus</i> V145–80	dsRNA 1 (1994)	2031; 676; 71	26	87
Aspergillus fumigatus mitovirus – 1	AfuMV – 1	<i>A. fumigatus</i> V145–13	dsRNA 1 (2500)	1764; 587; 69	300	436
Aspergillus fumigatus tetramycovirus – 1	AfuTmV – 1	<i>A. fumigatus</i> Af293	dsRNA 1 (2403)	2292; 763; 84	35	76
			dsRNA 2 (2233)	2091; 696; 76	70	72
			dsRNA 3 (1970)	1845; 614; 67	51	74
			dsRNA 4 (1131)	840; 279; 29	86	205

encapsidated separately, while *Totiviridae* have non-segmented genomes. In contrast, family *Narnaviridae* accommodate unencapsidated mycoviruses with ssRNA genomes.

The *Aspergillus foetidus* Mycovirus Complex

The *A. foetidus* mycovirus complex, *i.e.*, the seven dsRNA elements found in *A. foetidus* strain IMI 41871, was initially reported in the 1970s and used as a model system to study virus replication and transcription, although the actual sequence of these elements was determined in the 2010s. Its presence is widespread among *Aspergilli*, since it has been noted not only in different isolates of *A. foetidus* but also in *A. niger*. The seven dsRNA elements are accommodated in two different classes of VLPs, designated as ‘fast’ and ‘slow’ based on their electrophoretic mobility. The ‘fast’ and ‘slow’ VLPs are both isometric and similar in diameter, approximately 33–37 nm; however they are serologically unrelated to each other and antibodies specific for one class of VLPs do not cross-react with the other, suggesting that each class of VLPs is constructed with a distinct capsid protein.

The ‘fast’ VLPs accommodate a yet unclassified mycovirus with four dsRNA elements as its genome, each encapsidated separately. Each one of the four dsRNAs has a single open reading frame (ORF) flanked by conserved 5'- and 3'-untranslated regions (UTRs) and a 3' poly(A) tail. The two larger RNAs encode the RNA-dependent RNA polymerase (RdRP) and the capsid protein (CP), respectively, while the other two encode proteins of unknown function. Similar mycoviruses have been described not only in *Aspergilli* but in other ascomycetes as well, such as *Alternaria alternata* and *Fusarium poae*.

The ‘slow’ VLPs accommodate the remaining three dsRNA elements: a member of the genus *Victorivirus*, family *Totiviridae*, which has a linear non-segmented genome and encodes an RdRP and a CP; an unclassified mycovirus with a single open reading frame encoding an RdRP, closely related to viruses found in other ascomycetes, such as *Fusarium poae*, *Penicillium aurantiogriseum*, and *Rosellinia necatrix*, and in basidiomycetes, such as *Rhizoctonia solani*; and a satellite RNA, less than 0.5 kbp in size with no protein coding capacity. All three dsRNA elements depend on the victorivirus, which produces the capsid protein that forms the

'slow' VLPs, for their encapsidation. The victorivirus and the unclassified virus each encode their own RdRP for replicating their own genomes, and the RdRP of the unclassified virus is also responsible for the replication of the satellite RNA.

Partitiviruses

Three viruses belonging to the family *Partitiviridae*, genus *Gammapartitivirus*, have been found in *Aspergilli*, *Aspergillus ochraceus* virus (AoV) was isolated from *A. ochraceus*, while *Aspergillus fumigatus* partitivirus-1 (AfuPV-1) and *Aspergillus fumigatus* partitivirus-2 (AfuPV-2) were discovered in *A. fumigatus*. Notably, the two partitiviruses from *A. fumigatus* are not closely related. Each virus has two dsRNA segments as its genome, containing one ORF flanked by conserved 5'- and 3'-UTRs. As is typical for partitiviruses, the larger dsRNA encodes the RdRP and the smaller the CP. The *Aspergillus ochraceus* virus in particular has an additional component, which has been fully sequenced and is known to contain an ORF encoding a protein of unknown function. Since its UTRs are very similar to those of the other two partitivirus components, it is most likely that it is replicated by the RdRP.

Chrysovirus

The presence of a virus belonging to the genus *Alphachrysovirus*, family *Chrysoviridae* and named *Aspergillus fumigatus* chrysovirus (AfuCV) has been reported in *A. fumigatus*. The AfuCV genome consists of four dsRNAs, each one containing an ORF flanked by conserved 5'- and 3'-UTRs. The two largest dsRNAs encode the viral RdRP and CP, respectively, while the other two dsRNAs encode proteins whose function is not fully understood. More specifically, the smallest dsRNA encodes a putative cysteine protease (alphachryso-P4) while the third dsRNA encodes a protein (alphachryso-P3) that shares significant sequence similarity with the N-terminus of the RdRP. The alphachryso-P3 contains a 'phytoreo S7 domain', initially described in members of the genus *Phytoreovirus*, family *Reoviridae*, which is considered to interact with the RdRP and play a role in viral RNA binding and packaging. Furthermore, the remote protein homology detection HHpred software showed that the AfuCV alphachryso-P3 contains a P-loop NTPase domain near its N-terminus, which has homologous sequences in RdRP. Members of the P-loop NTPase domain superfamily are characterized by conserved nucleotide phosphate-binding motifs, also referred to as the Walker A motif (GxxxxGK [S/T]), and the Walker B motif (hhhh[D/E]), where h is a hydrophobic residue.

Narnaviruses and Mitoviruses

Three members of the family *Narnaviridae* have been sequenced in *A. fumigatus*, two belonging in the genus *Narnavirus* and one in the genus *Mitovirus*. These viruses encode only one protein, the RdRP responsible for replication. The two narnaviruses are not closely related and they are considered the simplest, smallest viruses. Unlike the mitovirus, which is found in mitochondria, narnaviruses are cytoplasmic. These are the only ssRNA mycoviruses reported in *Aspergillus* to date, although it is believed that the unclassified virus in the 'slow' *A. foetidus* mycovirus complex also has an ssRNA genome.

Polymycoviruses

The prototype member of the novel proposed family *Polymycoviridae*, designated as *Aspergillus fumigatus* tetramycovirus-1 (AfuTmV-1), was discovered in *A. fumigatus*. AfuTmV-1 has four dsRNA segments as its genome, similarly to members of the families *Chrysoviridae* and *Quadriviridae* and the unclassified virus in the 'fast' *A. foetidus* mycovirus complex; however, unlike these viruses, AfuTmV-1 segments are smaller in size, ranging from 1 kbp up to 2.5 kbp in length. AfuTmV-1 does not have a capsid and can be isolated from the fungus as long chains of dsRNA associated with a protein. Additionally, it was the first viral entity shown to be infectious as naked dsRNA, following introduction of its purified genome into *A. fumigatus* protoplasts. This was a paradigm-shifting discovery since, unlike the genome of positive-sense ssRNA viruses that can successfully replicate following its introduction in the host cells, dsRNA viruses were previously considered not to be infectious but as whole virions.

The largest dsRNA encodes the viral RdRP, which is related to the RNA picorna-superfamily and has two unusual characteristics that make it unique: firstly, it is evolutionary closer to members of the families *Astroviridae* and *Caliciviridae*, which have positive-sense, single-stranded RNA genomes, and infect birds and mammals; secondly, its GDD motif, responsible for the catalytic activity of the enzyme, has been replaced with GDNQ, characteristic of negative-stranded RNA viruses such as those in the order *Mononegavirales*, which include important human pathogens such as measles, rabies, and Ebola. The second largest dsRNA encodes a protein of unknown function containing a putative endoplasmic reticulum (ER) retaining signal peptide and a short trans-membrane alpha-helix together with a zinc finger-like motif similar to those that bind nucleic acids. It is hypothesized that it acts as a scaffold protein, recruiting the replication complex in the ER and facilitating the interaction among the viral RNA and the viral proteins. Both RNA and DNA viruses, such as hepatitis C and vaccinia, are known to utilize and rearrange the ER membrane for their replication and assembly and it is conceivable that polymycoviruses are also associated with this host organelle. The third

largest RNA encodes a methyl-transferase responsible for capping the positive RNA strands of the virus; a conserved catalytic motif and a FAD/NAD(P)-binding Rossmann fold domain have been identified in the protein sequence. The latter binds a coenzyme, which acts as an acceptor or a donor for the hydrogen anion lost or gained during the redox reaction catalysed by the methyl-transferase. The presence of the capping structure leads to the recognition of the viral RNAs, which act as mRNAs, by the host's ribosomes and to the translation of the viral proteins. Notably this is the first report of a capping enzyme and the RdRP being on distinct dsRNA elements. Finally, the smallest dsRNA encodes a protein hypothesized to coat the viral genome, which is non-conventionally encapsidated. This protein is enriched in the residues proline, alanine, and serine and has long intrinsically disordered regions, which may facilitate its interactions with RNA and proteins.

The aforementioned four dsRNAs and the proteins they encode are common in all known polymycoviruses; however polymycoviruses may have up to four additional genomic segments, encoding small non-homologous proteins of unknown function. It is feasible that these proteins originally participated in the formation of a capsid, since a polymycovirus from plant pathogenic fungus *Colletotrichum camelliae* has eight dsRNA segments and is encapsidated in filamentous virions. Loss of one or more of these genomic segments would not affect the ability of polymycoviruses to replicate, but would coincide with the shredding of the capsid.

Phenotypes

A. fumigatus is the major causative agent of aspergillosis and related respiratory diseases, affecting immunocompromised patients undergoing transplantation or chemotherapy and those vulnerable due to another infection, such as tuberculosis, or a genetic disorder, such as cystic fibrosis. Other *Aspergillus* species are significant food contaminants, may have industrial uses and produce toxins such as aflatoxins. A number of case studies have revealed that mycoviruses may have discernible effects on their fungal host, altering its phenotype. These comparisons typically involve virus-free and virus-infected isogenic lines; *i.e.*, the same fungal isolate with and without the virus, in order to ensure that any observed phenomena are due to the presence of the virus and not to potential differences in the genomic background of the host. The most easily detected phenotypes are alterations in pigmentation of the fungus and may be accompanied by differences in growth, conidiation, pathogenicity, and/or mycotoxin production. For instance, virus infection in an *A. niger* isolate was shown to lead to reduced growth rate and production of conidia.

Virus-mediated aflatoxin production is a very interesting phenotype associated with isolates of *A. flavus*. Aflatoxins are carcinogenic secondary metabolites produced by *A. flavus* and *A. parasiticus*, which are opportunistic plant-pathogens, and can be found on a range of agricultural crops including tree nuts, corn, peanuts, grains, and soybeans, mostly in warm and humid environments. This contamination of food resources with aflatoxins, to which the consumer that can be exposed either directly or indirectly *via* meat and dairy products, is a serious medical and economic concern, particularly in developing countries. The production of aflatoxins is down regulated in the presence of uncharacterised mycoviruses in *A. flavus* strains NRRL 5565 and NRRL 5940. Nevertheless, this correlation is not universal since mycoviruses do not appear to affect aflatoxin levels in other *A. flavus* isolates or the production of the mutagenic patulin by *A. clavatus*. No attempts were made to further characterize the mycoviruses from the NRRL 5565 and NRRL 5940 strains or to take advantage of them in order to control aflatoxin production in the field.

The most extensive phenotypic studies have been performed in *A. fumigatus* isolates: the effect of different mycoviruses on fungal growth has been investigated both in solid and in liquid culture and their effect on virulence has been assessed on mice and on larvae of the greater wax moth *Galleria mellonella*, both widely used model systems for pathogenicity studies. The *A. fumigatus* isolate harboring the chrysovirus produces darker green pigmentation than virus-free isogenic line, while the isolate with the partitivirus has lighter pigmentation. Chrysovirus and partitivirus infections both lead to the formation of aconidial sectors and significantly reduced growth in solid and in liquid culture, but no effects on fungal pathogenicity were discernible. In contrast, infection of *A. fumigatus* with polymycoviruses leads to hypervirulence, or increased pathogenicity, apparently *via* two different mechanisms: in isolate A78 harboring an uncharacterised in terms of sequence polymycovirus, hypervirulence is at least partially a result of increased growth; in Af293 isolate harboring AfuTmV-1, hypervirulence is not concomitant with enhanced growth and depends on other unidentified parameters associated with pathogenicity. For instance, AfuTmV-1 may affect the expression of genes involved in the thermotolerance of the fungus, the constitution and polysaccharide content of the fungal cell wall, together with melanin pigment production, resistance to oxidative stress and other factors that influence the interplay between the fungus and the host's immune system. The association, if any, between mycovirus infection in *A. fumigatus* and clinicopathological parameters, patient survival, and azole resistance is currently unknown.

The next major challenge in the investigations of *Aspergillus* mycoviruses is understanding the molecular mechanisms and signaling pathways mediating the aforementioned phenotypes. Due to the dsRNA nature of their genomes or their replicative forms, mycoviruses act as both triggers and targets of RNA silencing, a process leading to the degradation of viral RNA as part of the anti-viral defense of the fungal host. This phenomenon has been studied in *A. nidulans* and *A. fumigatus*, demonstrating that non-conventionally encapsidated viruses, such as polymycoviruses, and encapsidated viruses, such as chrysoviruses and partitiviruses, can be silenced, while some mycoviruses may in turn suppress RNA silencing in their fungal host. Mycovirus infection may also result in differential expression of small regulatory RNAs, suggesting a putative mechanism underpinning the observed mycovirus effects on fungal hosts.

Further Reading

- Bhatti, M.F., Jamal, A., Petrou, M.A., *et al.*, 2011. The effects of dsRNA mycoviruses on growth and murine virulence of *Aspergillus fumigatus*. *Fungal Genetics and Biology* 48, 1071–1075.
- Hammond, T.M., Andrews, M.D., Roossinck, M.J., Keller, N.P., 2008. *Aspergillus* mycoviruses are targets and suppressors of RNA silencing. *Eukaryotic Cell* 7, 350–357.
- Kanhayuwa, L., Kotta-Loizou, I., Özkan, S., Gunning, A.P., Coutts, R.H.A., 2015. A novel mycovirus from *Aspergillus fumigatus* contains four unique double-stranded RNAs as its genome and is infectious as dsRNA. *Proceedings of the National Academy of Sciences of the United States of America* 112, 9100–9105.
- Kotta-Loizou, I., Coutts, R.H.A., 2017. Mycoviruses in *Aspergilli*: A comprehensive review. *Frontiers in Microbiology* 8, 1699.
- Özkan, S., Coutts, R.H.A., 2015. *Aspergillus fumigatus* mycovirus causes mild hypervirulent effect on pathogenicity when tested on *Galleria mellonella*. *Fungal Genetics and Biology* 76, 20–26.
- van Diepeningen, A.D., Debets, A.J., Hoekstra, R.F., 2006. Dynamics of dsRNA mycoviruses in black *Aspergillus* populations. *Fungal Genetics and Biology* 43, 446–452.
- Zoll, J., Verweij, P.E., Melchers, W.J.G., 2018. Discovery and characterization of novel *Aspergillus fumigatus* mycoviruses. *PLoS One* 13, e0200511.

Evolution of Mycoviruses

Mahtab Peyambari, Vaskar Thapa, and Marilyn J Roossinck, Pennsylvania State University, State College, PA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Alpha helix (α -helix) A common protein secondary structure that is formed when the polypeptide chains twist into a spiral.

Bottleneck An evolutionary event that drastically reduces the size of a population. It can be caused by various factors but here different cell types are the contributing factor. It lowers genetic variability in a population.

Convergent evolution The process whereby organisms not closely related, independently evolve similar traits as a result of having to adapt to similar environments or ecological niches.

Evolutionary lineage A temporal series of organisms, populations, cells, or genes connected by a continuous line of descent from the ancestor to descendant. Lineages are subsets of the evolutionary tree of life. Lineages are often determined by the techniques of molecular systematics.

Jelly roll A protein fold or supersecondary structure composed of eight beta strands arranged in two four-stranded sheets. The jelly roll is the most prevalent fold among viral capsid proteins.

Modular evolution The concept that viruses contain sets of genes or modules that have a common ancestor but may be found in different arrangements in viral genomes.

Polyphyletic A group of organisms that do not share an immediate common ancestor.

Purifying selection The selective removal of mutations or alleles that are not beneficial to an organism.

RNA silencing A sequence-specific degradation of RNA, usually triggered by double-stranded RNA. RNA silencing provides an antiviral defense response at the cellular level, and is an adaptive immune system used by plants, fungi and a few other organisms.

Vegetative Compatibility The ability of hyphae to fuse (anastomosis). Vegetative compatibility is controlled by one to several nuclear genes that limit completion of hyphal anastomosis between fungal colonies to those that belong to the same vegetative compatibility group (usually abbreviated to v-c group).

Introduction

Mycoviruses are commonly found in many taxa of fungi. These viruses lack an extracellular phase but move between hyphal cells during cell division, and naturally transmit horizontally between vegetatively compatible groups through the anastomosis (fusion of hyphae). Some mycoviruses are able to facilitate anastomosis between vegetatively incompatible fungal strains by overcoming non-self recognition, a common phenomenon in fungi that enables distinction of oneself from another fungus. The non-self recognition is activated in fungi when two vegetatively incompatible hyphae come into contact resulting in compartmentalization followed by programmed cell death to interrupt the fusion between hyphae. In many systems vertical transmission of mycoviruses occurs efficiently through fungal asexual spores. The effective horizontal and vertical transmissions in mycoviruses are responsible for their widespread dispersal.

The International Committee for the Taxonomy of Viruses (ICTV) in 2017 reported twelve classified and seven unclassified families of mycoviruses. A majority of mycoviruses have double-stranded (dsRNA) genomes, but this could be partially due to bias in the analysis that often involves isolation of dsRNA as a starting point. Unlike most of the single-stranded (ss) RNA families, the extremely broad range of the dsRNA viruses in fungi appears rather unusual. The three largest families of dsRNA viruses, *Partitiviridae*, *Totiviridae*, and *Reoviridae*, show broad host ranges, infecting two or three kingdoms of eukaryotic hosts, implying ancient origins.

In general, RNA viruses have higher evolutionary rates because the RNA dependent RNA polymerases (RdRps) that replicate their genomes are more error-prone than those for DNA dependent DNA polymerase. The estimated rate of nucleotide substitutions per site per year for animal RNA viruses falls within one order of magnitude of 1×10^{-3} . No parallel studies on mycovirus evolutionary rates are reported. However, unlike most animal RNA viruses, mycoviruses have an exclusively intracellular lifestyle that predominantly limits them to infection of a single fungal host, or to vegetatively compatible populations. Moreover, mycoviruses experience bottlenecks within their hosts. In many fungi such as the genus *Aspergillus*, mycoviruses are transmitted vertically via asexual spores but not through sexual spores. Mutation events in mycoviruses with dsRNA genomes are further constrained by their stamping-machine mode of replication. In stamping-machine replication, one strand of the genome produces multiple copies, but this is then copied only once to produce the genomic RNA. In most ssRNA viruses each strand can be amplified during replication, resulting in a logarithmic increase in progeny strands, and allowing for a higher probability of mutation events to occur. It is likely that such a lifestyle of mycoviruses results in a lower tolerance for mutations, or a higher pressure of purifying selection and hence a lower evolutionary rate over time. Studies in plant persistent viruses that share a similar lifestyle with mycoviruses and belong to the families common to both fungal and plant viruses, show similar slow rates of evolution. For example, an endornavirus that infects the ancestor of domesticated rice and a related virus that infects cultivated rice have diverged by about 24% over the 10,000 years since rice was first cultivated. Another example is a chrysovirus isolated from

ancient maize cobs from about 1000 years ago that can still be found in modern maize. The virus sequences have diverged by only about 3%.

Due to the lack of an extracellular phase for mycoviruses, and dependence on the host for transmission, mycoviruses are generally thought to have co-evolved with their hosts. Their symptomless and persistent infections and, in some cases positive interactions with hosts, support this view. However, phylogenetic studies show mycoviruses are diverse despite sharing the same lifestyle and do not seem to have evolved from a single ancestral virus.

Evolutionary History and Drivers of Evolution

A wide range of evolutionary forces contributes to genetic variability and adaptability of mycoviruses to a changing environment. A comprehensive evolutionary study across the mycoviruses is lacking, however, findings from some mycovirus taxa (e.g., families) and individual species provide us a picture of evolutionary forces shaping the extant population. The family *Totiviridae* contains mycoviruses with undivided dsRNA genomes and is one of the more studied families of mycoviruses. Some members of this family also infect protozoans. Sequence comparisons of predicted RdRp amino acids among members show significant sequence similarity with eight conserved motifs among totiviruses that infect yeast, smut, filamentous fungi, and protozoans. However, the similarity in coat proteins (CPs) among members is not as pronounced as with the RdRps, a common observation among viruses. Two evolutionary lineages are distinct among totiviruses infecting fungi. One of the lineages, now grouped under the genus *Victorivirus* differs from the other (grouped under the genus *Totivirus*) by its RdRp expression strategy. In *Victorivirus* the RdRp open reading frame is separate from that of the CP, whereas these reading frames are fused in members of the *Totivirus* and are expressed by a frame-shifting mechanism. The ancestral relationships of the totiviruses infecting fungi are uncertain. However, the genus *Victorivirus* is phylogenetically closer to the totiviruses infecting protozoans than the totiviruses of yeast and smut fungi indicating that the fungal viruses in the family *Totiviridae* are probably polyphyletic.

Partitiviruses infecting fungi are another widely studied group of viruses. The viruses in the family *Partitiviridae* infect fungi, plants, and oomycetes and possess two dsRNAs genome segments. The members of *Partitiviridae* infecting fungi have three distinct lineages, out of which only one lineage includes exclusively fungi (filamentous, ascomycetes: *Gammapartitivirus*) as hosts, while the two other lineages have both fungal and plant hosts (*Alphapartitivirus* and *Betapartitivirus*). The lineages of partitiviruses with both fungal and plant hosts suggest a horizontal transfer of members between kingdoms.

An analysis of the protein domains of endornaviruses infecting fungi indicates diverse origins suggesting a modular evolutionary history. The endornaviruses infecting fungi have naked ssRNA genomes that encode polyproteins for various function. The helicase domains of *Gremmeniella abietina* type B RNA virus XL1 and *Chalara elegans* endornavirus 1 (CeEV1) are derived from two different protein superfamilies. Similarly, *Helicobasidium mompa* endornavirus 1 and CeEV1 have glycosyltransferase domains originating from the different protein family roots. The occurrence of domains with similar functions but of different origins in different fungal endornaviruses indicates a modular convergent evolutionary strategy. In fact, the evolution of domains of similar function but highly divergent primary sequence in many mycoviruses is sometimes interpreted as ancient divergence, but can equally be understood as convergent evolution.

The highly conserved RdRp domain in endornaviruses is related to the RdRp of closteroviruses, which are plant ssRNA viruses. A phylogenetic analysis of hypovirus RdRps and the gain or loss of other genes support the hypothesis that these naked (+) ssRNA fungal viruses are derived from the potyviruses, another group of plant (+) ssRNA viruses. The mitoviruses, found in the mitochondria of fungi, may have originated from bacteriophages in the *Leviviridae* family, the only (+) ssRNA family in prokaryotes. It seems likely that the mitoviruses evolved from a bacterial virus that lost its coat protein and other genes. The related narnaviruses infecting yeasts are found in the cytoplasm, and could have originated from a mitovirus that escaped to the cytosol. The origin of dsRNA viruses is unclear, but one possibility is that a (+) ssRNA could have been the ancestor of all dsRNA viruses that encapsidated a replicative form (dsRNA) together with the RdRp. Mitoviruses, hypoviruses, and endornaviruses, that have a ssRNA-like RdRp, but are found as unencapsidated dsRNA replicative forms, could be evolutionary intermediates in this process.

The major evolutionary forces affecting more recent evolution of the mycoviruses are mutation and selective sweeps that were discussed above with respect to the different lineages in many groups of mycoviruses. Besides mutation, recombination and horizontal gene transfer (HGT) have been reported often in the literature. Recombination is an evolutionary force that is common in some ssRNA viruses. Recombination occurs when an RdRp switches its template from one molecule to another during virus replication. Since dsRNA viruses replicate exclusively within the virus particle, the opportunity for recombination between templates is probably very uncommon. However, a low frequency of recombination events is reported in *Ustilago maydis* virus H1, a member of the genus *Totivirus*. In members of hypoviruses such as *Cryphonectria hypovirus* 1 (CHV1) that replicate as ssRNA viruses homologous recombination events are noted frequently and some of the recombinant lineages appear to be responsible for the spread of the virus in Europe.

Recent studies suggest that HGT between diverse mycoviruses as an important force contributing to evolution. The homologs of the S7 protein domain, a DNA-binding protein common in phytoreoviruses in the family *Reoviridae*, are reported from many mycoviruses that belong to the families *Endornaviridae*, *Totiviridae*, *Chrysoviridae*, and unclassified monopartite dsRNA fungal viruses. The domain is associated with the RdRp in many chrysoviruses but found with CPs and polyproteins in totiviruses and endornaviruses, respectively. Phylogenetic analysis of the S7 domains from various mycoviruses supports HGT across species. HGT is also reported in *Sclerotinia sclerotiorum* reovirus 1 (SsReV1), a member of the family *Reoviridae*. Virus protein 6 and virus protein 7 of *Sclerotinia sclerotiorum* reovirus 1 (SsReV1) have sequence similarities with dsRNA binding motif (dsRBM) and

reovirus sigma C protein (Reo_ σ C), respectively. Reoviruses can infect mammals, protozoans, insects, plants, and fungi. Homologs of the dsRBM and Reo_ σ C are found in many reoviruses and diverse virus lineages outside the family *Reoviridae*. The phylogeny based on dsRBM and Reo_ σ C sequences supports HGT.

Effects of Host Divergence in the Evolution

The complex relationship between viruses and their hosts can result from any evolutionary event such as adaptation to the hosts. But the question is “How much of the diversification of mycoviruses is linked to the diversification of their fungal hosts?” In one study of mycovirus-host codivergence, results indicated codivergence as a dominant mode of virus diversification in *Partitiviridae* and *Totiviridae*, but not in other mycovirus families. However, the lack of codivergence in other families might be the result of insufficient sampling.

As viruses employ host translation systems, their proteins use the same genetic code as their host organism. There are a few organisms that use modified genetic codes rather than the universal genetic code. If a virus that uses the universal genetic code infects a host that utilizes an alternative genetic code, this would constitute high fitness costs. These differences in genetic codes could inhibit viral jumping between different hosts. Some authors have hypothesized that alternative genetic codes could be used as an antiviral strategy against virus transfers from other hosts. The effects of genetic code shifts on virus-host co-evolution remain poorly understood.

Mitoviruses (family *Narnaviridae*) have the simplest genome among mycoviruses. They are capsid-less ssRNA viruses that encode only an RdRp. While all other known mycoviruses reproduce in the fungal cytoplasm, mitoviruses are unique in reproducing in the mitochondria. Fungal mitochondria often use an alternative genetic code, in which the universal stop codon UGA encodes tryptophan (Trp). Mitoviral RdRps have internal UGA codons, so as these RdRps are active in mitochondria, then UGA also encodes Trp in mitoviruses.

Recently mitovirus-like sequences have been found in plant mitochondria and also integrated into the nuclear genome of many vascular plants. The latter are not transcriptionally active. The plant mitoviruses do not contain the UGA codon for Trp, so they have adapted to this host change. Although, in most fungal mitochondria UGA (Trp) is a common codon, in some fungi UGA (Trp) is a rare mitochondrial codon, and in mitoviruses in these hosts the UGA is mostly absent. Thus, fungal mitoviruses are adapted to their host species by using UGA as Trp codon, but this changes with host species. Recent investigation of plant transcriptomes revealed the presence of many complete genomes of plant mitoviruses, providing strong evidence for genuine plant mitoviruses.

In a study on *Scheffersomyces segobiensis*, a fungus with an alternative genetic code where CUG codes for serine instead of leucine, a dsRNA virus of the genus *Totivirus* has removed all but one of the CUG codons from functional positions of its genome, presumably for adaptation to the host modified genetic code. Without this shift in the genetic code, protein folding would be impacted by replacing a hydrophobic residue with a polar residue.

Based on phylogenetic analysis, *Scheffersomyces segobiensis* virus L evolved from exogenous totiviruses with a standard genetic code. Apparently, genetic code shifts are not serious barriers against virus jumping between different hosts.

Coinfection and Evolution

Coinfection of multiple mycoviruses is common in nature. Coinfecting mycoviruses impose pressure on each other resulting in neutral, synergistic or antagonistic interactions. In many cases, mycoviruses transfer across different fungal strains of vegetative incompatibility by overcoming non-self recognition. However, some mycoviruses can suppress the non-self recognition and facilitate the exchange of mycoviruses from diverse vegetatively incompatible groups. This opens up the opportunity for interactions among different mycoviruses while coinfecting the same host that have evolutionary consequences. One example is the interactions between Mycoreovirus 1 (MyRV1) from the genus *Mycoreovirus* and CHV1 from the genus *Hypovirus*, in the chestnut blight fungus, *Cryphonectria parasitica*. MyRV1 possesses 11 dsRNA segments (S1-S11). The multifunctional protein p29, encoded by CHV1 induces reproducible intragenic rearrangements of the S6 and S10 genomic segments of MyRV1, including duplication in S6 and an internal deletion in S10. Further, p29 copurifies with MyRV1 genomic RNA and binds in vitro to the VP9 protein of MyRV1 that is involved in replication.

Fungi use the adaptive immune system of RNA silencing to counteract RNA viruses. RNA silencing may also contribute to viral RNA recombination. In coinfection, it is likely that RNA silencing may lead to the production of chimeric viral RNAs as a product of intermolecular recombination. Such possibilities have evolutionary consequences and deserve further investigations.

Structural Evolution of Mycoviruses

The origin of the RdRp of RNA viruses is unclear. The protein sequences of RdRps, as hallmark genes in RNA viruses, are divergent, but contain well-conserved domains. The high level of structural similarity based on atomic structure among RdRps of all groups of RNA viruses [dsRNA, (+) ssRNA, and (-) ssRNA] also could imply a common origin for all RdRps. However, it is not possible to be certain whether these critical molecules have diverged or converged for a common function, and if they have diverged, it is difficult to be certain which extant form of RdRp is closest to the ancestral state.

The virion architecture is also a crucial element in understanding virus origins. In spite of tremendous genetic variation, there are a limited number of capsid protein folds, and only a small subset of these have the potential to result in a virus particle.

However, the origin of viral capsids is a poorly understood aspect of virus evolution. Unlike the RdRps, the capsid proteins have no conserved domains, and primary protein sequence analysis, even in viruses in the same family, indicates a polyphyletic origin of these genes. This type of chimeric origin is a common evolutionary scenario for encapsidated mycoviruses.

Recent analyses of virion structure have revealed unexpected similarities in the folds of capsid proteins among these different viruses, suggesting an evolutionary connection among viruses that infect hosts residing in different kingdoms of life. Two different evolutionary trajectories could be concluded from such observations, divergent or convergent evolution. In the divergent scenario, there is a common ancestor for each separate lineage that existed before their host organisms diverged, but these proteins have diverged beyond any recognition at the sequence level. The other scenario would be that the structural similarities are because of convergent evolution to solve the problem of making a stable and viable capsid.

A comparative analysis to investigate the diversity and potential origins of viral capsids resulted in 18 “structure-based viral lineages”, in which 76.3% of viral taxa with defined folds of the major capsid proteins could be categorized. One of these lineages is the dsRNA bluetongue virus (BTV)-like viruses. BTV is a member of the *Reoviridae*.

The icosahedral mycoviruses in the *Totiviridae*, *Partitiviridae*, *Megabirnaviridae*, *Chrysoviridae*, and *Quadriviridae* families have a unique 120-subunit $T = 1$ jelly-roll capsids that also have been described as the internal shell in bacteriophages of the family *Cystoviridae* and members of *Reoviridae*, and as viral capsids in picobirnaviruses, which infect higher eukaryotes. They share a structural signature with the lineage of the dsRNA BTV-like viruses. Reoviruses form a complex capsid consisting of two icosahedral shells similar to cystoviruses, indicating a possible evolutionary relationship between these families.

The icosahedral capsid of dsRNA viruses remains intact during transcription, extruding the ssRNAs that act as mRNAs and as pregenomic RNAs from the intact capsids, probably to avoid triggering host defense mechanisms. To date, four $T = 1$ capsids of mycoviruses have been resolved at the atomic level: *Saccharomyces cerevisiae* 1-A virus (*Totiviridae*), *Penicillium chrysogenum* virus (*Chrysoviridae*), *Penicillium stoloniferum* virus F (*Partitiviridae*), and *Rosellinia necatrix* quadrivirus 1 (*Quadriviridae*). The amino acid sequences of the above four capsid proteins are quite different, however, they all have the same fold, predominantly α -helical, which is also a hallmark of BTV lineage. According to phylogenetic analyses of the RdRps of partitiviruses, they share a common ancestor and RdRp has monophyletic origin in this group of viruses, however, despite the striking similarity of their structure, coat protein genes are often unrelated and are clearly polyphyletic. So, in this case, the evolution of the RdRp alone does not reflect the evolution of partitiviruses, and it is likely that partitiviruses gained coat protein genes at different times, from various sources. All partitiviruses with resolved atomic structures, despite their high coat protein sequence divergence, share distinctive features including surface arches that are unique among dsRNA viruses infecting fungi and not seen in other mycovirus families. The presence of surface arches in partitiviruses, with thin capsid shells, may have a role in stabilization of the capsid, however, their main function is still unknown.

Among dsRNA viruses, partitiviruses and picobirnaviruses, have minimalist genomes consisting of two dsRNA segments, and share capsid structure and distinctive features. However, picobirnaviruses, as vertebrate viruses, are grouped in a different family due to a variety of differences from partitiviruses including host range. Given the structural similarity, two different evolutionary trajectories are suggested for *Partitiviridae* and *Picobirnaviridae* capsids, either they are diverged from a common ancestor that existed before their host organisms diverged or they are converged from different ancestors as a solution to the problem of a thin, unstable capsid protein.

Obviously, not only evolution of the RdRp but also evolution of other genes either gained or lost during evolution are important in the evolution of mycoviruses. There is much to still be learned about mycovirus evolution, largely because this group of viruses is very understudied, and most of the work done has centered on viruses of fungi that are pathogenic in plants or animals.

Further Readings

- Huiquan, L., Yanping, F., Jiatao, X., *et al.*, 2012. Evolutionary genomics of mycovirus-related dsRNA viruses reveals cross-family horizontal gene transfer and evolution of diverse viral lineages. *BMC Evolutionary Biology* 12 (1), 91–105.
- Koonin, E.V., Dolja, V.V., 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiology and Molecular Biology Reviews* 78 (2), 278–303.
- Krupovic, M., Koonin, E.V., 2017. Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences of the United States of America* 114 (12), E2401–E2410.
- Luque, D., Mata, P.C., Suzuki, N., Ghabrial, A.S., Castón, R.J., 2018. Capsid structure of dsRNA fungal viruses. *Viruses* 10 (9).
- Roossinck, M.J., Sabanadzovic, S., Okada, R., Valverde, R.A., 2011. The remarkable evolutionary history of endornaviruses. *Journal of General Virology* 92 (11), 2674–2678.
- Roossinck, M.J., 2019. Evolutionary and ecological links between plant and fungal viruses. *New Phytologist* 221 (1), 86–92.
- Safari, M., Roossinck, M.J., 2014. How does the genome structure and lifestyle of a virus affect its population variation? *Current Opinion in Virology* 9, 39–44.
- Shackelton, L.A., Holmes, E.C., 2008. The role of alternative genetic codes in viral evolution and emergence. *Journal of Theoretical Biology* 254 (1), 128–134.
- Taylor, D.J., Ballinger, M.J., Bowman, S.M., Bruenn, J.A., 2013. Virus-host co-evolution under a modified nuclear genetic code. *Peer Journals* 1 (e50), 2167–8359.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9 (6).

Relevant Website

<https://talk.ictvonline.org>
International Committee on Taxonomy of Viruses (ICTV).

Mixed Infections of Mycoviruses in Phytopathogenic Fungus *Sclerotinia sclerotiorum*

Jiatao Xie and Daohong Jiang, Huazhong Agricultural University, Wuhan, China

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Hypovirulence Reduced virulence of phytopathogenic fungi upon mycovirus infections.

Mycovirus Virus that can replicate and proliferate in fungal cells.

Virocontrol Biological control with hypovirulence-associated mycovirus.

Introduction

All living beings, as well as human beings, can be infected by innumerable viruses. As an important component of virosphere, mycoviruses (or fungal viruses) infect and replicate in fungi, and maybe play important roles in shaping fungal ecosystem under nature. Hypovirulence is a phenomenon that phytopathogenic fungi infected by one or more mycoviruses have reduced or lose virulence on their hosts. Hypovirulence-associated mycoviruses isolated from nature have potential to prevent and control crop fungal diseases, and it is also a useful exploration to open up a new strategy for the management of plant diseases. Hypovirulence-inducing *Cryphonectria hypovirus 1* (CHV1) has been successfully utilized to effectively control chestnut blight disease in Europe, which is a classic virocontrol example of fungal disease and also attracted more phytopathologists to unearthed mycovirus resources. Mycoviruses have been discovered in all major fungal groups during more than 50-year research, and mixed infections with multiple related or unrelated mycoviruses have frequently been confirmed in a single fungal isolate. For example, *Ophiostoma novo-ulmi* Log1/3–8d2 (Ld) with diseased phenotype was reported to be infected with twelve distinct mitoviruses. Two unrelated RNA viruses, yado-nushi virus 1 and yado-kari virus 1, co-infect a single isolate of phytopathogenic fungus *Rosellinia necatrix*, and potentially interplay in vivo. The advanced technique of high-throughput sequencing further supplies new insight into the complexity of the mixed infections in fungi. A hypovirulent strain DC17 of *Rhizoctonia solani* (AG2–2-IV) harbors 17 different mycoviruses that represent different virus lineages. Sixteen mycoviruses occur in a single strain of *Fusarium poae*. Those reports revealed that the mixed infections in fungi are fairly common and more complex than we previously thought.

Sclerotinia sclerotiorum (Lib.) de Bary is a ubiquitous necrotrophic pathogen and has a remarkably broad host range, encompassing over 400 species including important economic crops canola (oilseed rape), soybean, sunflower, and vegetables. In the later stage of the disease, *S. sclerotiorum* usually produces sclerotia, the resting structure for over-winter and summer, to survive for several years in soil. Under congenial conditions (such as moist soil, 10–20°C temperature), sclerotia germinate to produce apothecia (carpogenic germination) or directly produce mycelia (myceliogenic germination), and then results in damage of the plant tissue via the sophisticated pathogenicity mechanism. The disease caused by *S. sclerotiorum* is responsible for yearly several 100 millions of US dollars crop losses worldwide. To effectively control the sclerotinia disease, fungicides have been continuously used widely. However, long term dependence on chemical pesticides for sclerotinia disease leads to the pollution of agricultural environment and ecological damage, which threatens food safety and human health. Therefore, environment-friendly biological control (such as virocontrol) for crop diseases is an alternative strategy for the sustainable development of agriculture.

The first description of hypovirulence and dsRNA elements in *S. sclerotiorum* was published on Canadian Journal of Plant Pathology in 1992. So far, diverse mycoviruses with RNA or DNA genomes have been discovered in *S. sclerotiorum* (Table 1) as well as a DNA mycovirus, *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 (SsHADV1) and the firstly identified negative RNA mycovirus, *Sclerotinia sclerotiorum* negative-sense RNA virus 1 (SsNSRV-1). By application of a high-throughput sequencing-based metatranscriptomic approach, 84 Australian *S. sclerotiorum* isolates were identified to host 57 mycoviruses assigned into 10 distinct viral lineages, and 28 mycoviruses were detected from American *S. sclerotiorum* strains, suggesting that mycovirus infections are quite common in *S. sclerotiorum* population. Herein, we summarized the genomes of mycoviruses identified and molecularly characterized from the reported twenty-three *S. sclerotiorum* strains (Table 1). Some of those identified mycoviruses could confer hypovirulence, and are potential resources for biological control of Sclerotinia disease. For example, SsHADV1, the first example of extracellular transmission, has the potential to develop a virocontrol agent (or viral fungicide), and has been confirmed to be effectively control rapeseed rot disease development under field conditions. In addition, SsHADV1 can convert lifestyle of *S. sclerotiorum* from pathogenic fungus to beneficial endophyte, which enhance rapeseed resistance and yield. Notably, more than 80% (19/23) of those identified strains are mix-infected by multiple related or/and distinct mycoviruses (Table 1). In this article, we focus on characteristics of mixed infections of mycoviruses and their impacting in the individual strains of *S. sclerotiorum*, and also will discuss their potential interactions among viruses or their interactions with *S. sclerotiorum*.

Table 1 Mycoviruses infecting strains of *Sclerotinia sclerotiorum*

Strains	Mycoviruses ^a	Type of genome	Phenotype
Ep-1PN	SsDRV, SsRV-L	+ ssRNA	Hypovirulence
DT-8	SsDAHV1	ssDNA	Hypovirulence
Sunf-M	SsPV-S, SsNsRV-1	dsRNA, + ssRNA	Virulence
SZ-150	SsHV1, SsBV3, SsMTV1	dsRNA, + ssRNA	Hypovirulence
KL-1	SsMV1, SsMV2	+ ssRNA	Hypovirulence
16235	SsMV2, SsMV3, SsMV4, SsMV5, SsMV6, SsMV7	+ SsRNA	Hypovirulence
WF-1	SsPV1	dsRNA	Hypovirulence
11691	RcEV1, SsEV1, SsMV2, SsMV4, SsMV5, SsMV6 SsMV7	+ ssRNA	Virulence
14563	SsMV2, SsMV4, SsMV5, SsMV6, SsMV7, SsNsRV-1	– ssRNA, + ssRNA	Virulence
Lu471	SsEV1, SsMV2, SsMV3, SsMV4, SsMV5, SsMV6, SsMV7, SsNsRV-1	– ssRNA, + ssRNA	Virulence
5472	SsHV2, SsEV1, SsMV2, SsMV3, SsMV4, SsMV5, SsMV6, SsMV7, SsNsRV-1	– ssRNA, + ssRNA	Hypovirulence
AH98	SsNSRV-1, SsHV1	– ssRNA, + ssRNA	Hypovirulence
SX247	SsHV2, SsDRV	+ ssRNA	Hypovirulence
HC025	SsMV1	+ ssRNA	Hypovirulence
SCH941	SsReV1, SsBV1, SsDFV1, SsDFV3, SsYkV1	dsRNA, + ssRNA	Hypovirulence
JMTJ14	SsFV1	+ ssRNA	Hypovirulence
SX466	SsMBV1, SsMV2, SsPV2	dsRNA, + ssRNA	Hypovirulence
AX19	SsDFV1, six unknown dsRNA elements		Hypovirulence
AH16	SsMV2, SsBV2	dsRNA, + ssRNA	Hypovirulence
288	SsDFV2, BpRV1	dsRNA, + ssRNA	Hypovirulence
277	HuSRV1, SsMTV1	+ ssRNA	Hypovirulence
328	SsHV2, SsEV1	+ ssRNA	Hypovirulence
SX10	SsMyRV4	dsRNA	Hypovirulence

^aAbbreviated mycovirus names: SsDRV, Sclerotinia debilitation-associated RNA virus; SsRV-L, Sclerotinia sclerotiorum RNA virus L; SsDAHV1, Sclerotinia sclerotiorum hypovirulence associated DNA virus 1; SsPV, Sclerotinia sclerotiorum partitivirus; SsNsRV-1, Sclerotinia sclerotiorum negative single RNA virus 1; SsHV, Sclerotinia sclerotiorum hypovirus; SsBV, Sclerotinia sclerotiorum botybirnavirus; SsMTV1, Sclerotinia sclerotiorum mycotymovirus 1; SsMV, Sclerotinia sclerotiorum mitovirus; RcEV1, Rhizoctonia cerealis endornavirus 1; SsReV1, Sclerotinia sclerotiorum reovirus 1; SsYkV1, Sclerotinia sclerotiorum yadokarivirus 1; SsFV1, Sclerotinia sclerotiorum fusarivirus 1; SsMBV1, Sclerotinia sclerotiorum megabimavirus 1; SsDFV, Sclerotinia sclerotiorum deltaflexivirus; BpRV1, Botrytis porri RNA virus 1; HuSRV1, Hubei sclerotinia RNA virus 1; SsMyRV4, Sclerotinia sclerotiorum mycoreovirus 4.

Strain Ep-1PN Harbors Two Positive-Single Stranded RNA Mycoviruses

Strain Ep-1PN was originally isolated from a sclerotium collected from a diseased eggplant (*Solanum melongena*) and had lower virulence on rapeseed plants. dsRNA extraction suggested that Ep-1PN contained at least three extra-chromosome dsRNA segments (L, M, and S) with 7.4 kbp (base pairs), 6.4 kb, and 1.0 kbp in size, respectively. Subsequently, the L segment was confirmed to be the genome of Sclerotinia sclerotiorum RNA virus L (SsRV-L), M segment represented the replicative form of the genome of Sclerotinia sclerotiorum debilitation-associated RNA virus (SsDRV), and segment S was a defective RNA derived from SsDRV.

SsDRV is the first molecularly characterized mycovirus based on the full-length genome in *S. sclerotiorum*. SsDRV contains 5419 nucleotides (nt) excluding the poly (A) tail and has a single open reading frame (ORF). This ORF encodes a polyprotein with three conserved domains of methyltransferase, helicase and RNA-dependent RNA polymerase (RdRp). In *S. sclerotiorum* strain SX247 (see strain SX247 section), another strain SsDRV/SX247 was characterized, and share high sequence identity (81%) with SsDRV from strain Ep-1PN. Based on the phylogenetic analysis, SsDRV, the exemplar virus of the species *Sclerotinia sclerotiorum debilitation-associated RNA virus* of the genus *Sclerodarnavirus*, is closely related to the replicases of potex-like plant viruses (potexviruses) and Botrytis virus F (BVF) within the family *Alphaflexiviridae*. Interestingly, SsDRV has a single gene but lacks a coat protein gene, which is significantly distinct from other members of *Alphaflexiviridae* that usually contains five or six genes, and one of them encodes coat protein.

The genome of SsRV-L is 6043 nt in length, with a poly (A) tail. Similar to SsDRV, SsRV-L has only one reading frame and encodes a polyprotein (viral replicase). There is a significant sequence similarity between SsRV-L and human hepatitis E virus (HEV) replicase, which is the first discovery of mycovirus related to the human virus. Phylogenetic analysis showed that SsRV-L was closely related to members from *Rubi-like virus*, *Closterovirus*, *Benyvirus*, *Tobamovirus* and *Omegatetravirus*. Six mycoviruses closely related to SsRV-L have been recently detected in two phytopathogenic fungi *Rhizoctonia solani* (three mycoviruses) and *Sclerotium rolfii* (three mycoviruses), suggesting SsRV-L and its related mycoviruses are common in fungi.

SsDRV was associated with the hypovirulence phenotype of strain Ep-1PN, while SsRV-L has a limited impact on the biological features of *S. sclerotiorum*. Thus, the interaction system of *S. sclerotiorum*-SsDRV helps us to explore the possible mechanisms underlying fungal pathogenicity. 150 genes of *S. sclerotiorum* are down-regulated upon mycovirus infections. One of those genes, *Sclerotinia sclerotiorum integrin-like gene* (SSITL), encodes a secreted protein and is related to the virulence of *S. sclerotiorum*. This protein was confirmed to be secreted into plant cell, and play significant role in the suppression of jasmonic/ethylene (JA/ET) signal pathway mediated resistance via direct interaction with a chloroplast protein at the early stage of infection.

Strain SX247 is Co-Infected With Two + ssRNA Mycoviruses

Strain SX247 has abnormal colony morphology on the PDA medium and is incapable of causing diseased lesions on the rapeseed. Two dsRNA segments, the replicated forms of *Sclerotinia sclerotiorum* hypovirus 2 (SsHV2/SX247) and SsDRV/SX247 (see strain Ep-1PN), were detected and determined in strain SX247. The full-length cDNA of SsHV2/SX247 is 15219 nt excluding poly (A) and contains a large ORF that encoded a putative polyprotein with three conserved domains of papain-like protease, RdRp, and viral RNA helicase. The phylogenetic analysis revealed that SsHV2/SX247 forms an independent phylogenetic branch with hypovirus infecting *Sclerotium rolfsii*. Two other strains (SsHV2/5472 and SsHV2L) of SsHV2, respectively, were molecularly characterized from New Zealand isolate 5472 and American isolate 328 (co-infection with a strain of *Sclerotinia sclerotiorum* endornavirus 1). Compared to SsHV2/SX274 and SsHV2/5472 that are basically all the same in the genomic organization, SsHV2L lacks a fragment of 1.2 kb near the 5' terminus and has an insertion of 524 nt related to *Valsa ceratosperma* hypovirus 1 (NC_017099), suggesting SsHV2L is a recombinant of hypoviruses. In addition, the infectious clone of SsHV2L was successfully established to explore the interactions with *S. sclerotiorum* at the molecular level. All three strains of SsHV2 were directly or indirectly confirmed to be responsible for hypovirulence on *S. sclerotiorum* via methods of transfection with transcripts or horizontal transmission.

Strain AH98 is a Mix-Infection by Two Unrelated Single-Stranded RNA Mycoviruses

Two mycoviruses, *Sclerotinia sclerotiorum* hypovirus 1 (SsHV1) and *Sclerotinia sclerotiorum* negative-sense RNA virus 1 (SsNSRV-1), were identified from a hypovirulent strain AH98. SsHV1 and its biological features will be described in strain SZ-150 (see below). The full-length genome of SsNSRV-1, the first-ssRNA virus infecting fungi, was 10002 nt (KJ186782) with 6 non-overlapped genes (ORF I-VI) which were linearly arranged in the genome. ORF V encodes a large protein (L protein) with 1934 amino acids and is comprised of the conserved mononegaviral RdRp domain that was closely related to mononegaviruses, while other ORFs encode putative proteins with unknown function. The Gene-junction sequence of (A/U)(U/A/C)UAAU(U/A)AA(U/G)AAAACUUAGG(A/U)(G/U) that is widely present and the unique property in mononegaviruses is found at each ORF junction of SsNSRV-1. The sequence "AAAACUUAGG" serves as the transcriptional stop signal of the upstream ORF, while "UAUUUAAUAAAACU" is the transcriptional start signal of the downstream ORF. All six ORFs can be transcribed independently. In addition, ORF V and ORF VI are co-transcribed. The nucleocapsids of SsNSRV-1 are 22 nm in diameter and 200–2000 nm in length. SsNSRV-1 has defective RNAs with different internal deletions of ORF I and ORF II, and with deletions of the 5' termini. Although SsNSRV-1 is a member of *Mononegavirales*, it is different from reported members in genome organization, therefore a new family *Mymonaviridae* and genus *Sclerotimonavirus* being established to accommodate SsNSRV-1 and its related viruses. Moreover, SsNSRV-1 is widely distributed in the natural environment and has been detected in *S. sclerotiorum* strains collected from China, Australia, and USA. Transfection with SsNSRV-1 virions demonstrated that SsNSRV-1 is the key factor for hypovirulence in strain AH98.

The Hypovirulent Strain SZ-150 Contains Three Mycoviruses and a Satellite RNA

Strain SZ-150 was collected in Suizhou County of China, and its virulence on rapeseed plants was severely declined upon mycovirus infection. Application of the conventional method of dsRNA extraction, SZ-150 was originally considered to harbor two mycovirus-related elements, SsHV1 and a co-replicating satellite RNA (SatH). The genome of SsHV1 is 10,398 bp, excluding a poly (A) tail. SsHV1 contains a single ORF encoding a polyprotein with three conserved domains of glycosyltransferase, RdRp, and viral helicase. SsHV1 is phylogenetically related to hypoviruses, and is the first identified hypovirus in fungi other than *Cryphonectria parasitica*. SsHV1 was also detected in two *S. sclerotiorum* strains including strain AH98 coinfecting with SsNSRV-1 and the strain AX19, as well as in strains from Australia. SatH consists of 3643 nt and has an ORF encoding a putative protein (p70) with unknown biological functions. SatH replication is entirely dependent on the replication of SsHV1 which acts as a classic helper virus, and the genome size of SatH is larger than the previously reported viral satellite RNAs. Neither SsHV1 without SatH nor infectious cDNA of SatH without SsHV1 results in hypovirulence on *S. sclerotiorum*, suggesting SsHV1 and SatH are the co-determinant factors in eliciting hypovirulence in strain SZ-150. Further research suggested that the p70 protein encoded by SatH may play an important role as an RNA silencing suppressor to boost expression of SsHV1. It is possible that p70 may suppress the host antiviral response resulting in increased replication of SsHV1 and also guaranteeing the survival of SatH in *S. sclerotiorum*.

In addition to SsHV1 and SatH, SZ-150 was confirmed to harbor the other two mycoviruses, *Sclerotinia sclerotiorum* botybirnavirus 3 (SsBV3/SZ-150) and *Sclerotinia sclerotiorum* mycotymovirus 1 (SsMTV1/SZ-150), by the deep-sequencing of small RNAs. SsBV3/SZ-150 was assumed to be a strain of BpBV1, because of the whole genome of SsBV3/SZ-150 shares more than 95% sequence identity with *Botrytis porri* botybirnavirus 1 (BpBV1) at the nucleotide or amino acid level. SsMTV1/SZ-150 was predicted to contain a large ORF that encodes a putative replication-associated polyprotein. SsMTV1/SZ-150 is related, albeit distantly, to members of the family *Tymoviridae*. Both SsMTV1/SZ-150 and SsBV3/SZ-150 have limited contribution to hypovirulence of strain SZ-150.

Four Mycoviruses Co-Infect a Hypovirulent Strain SCH941

Two novel dsRNA viruses were identified from strain SCH941, a hypovirulent strain of *S. sclerotiorum*. One was a bipartite dsRNA virus named *Sclerotinia sclerotiorum botybirnavirus 1* (SsBV1) and the other was a novel reovirus named as *Sclerotinia sclerotiorum reovirus 1* (SsReV1). SsBV1 comprised two dsRNA segments (dsRNA1 and dsRNA2) and its genome size is 12,422 bp in total. A satellite-like RNA (SatlRNA) of SsBV1 was also found with 1647 bp. SsBV1 had spherical virions with 38 nm in diameter. SsBV1 phylogenetically clustered with BpBV1, a bipartite virus infecting *Botyis porri*, and belongs to the genus *Botybirnavirus*. Compared to the virus-free strain of *S. sclerotiorum*, infection of SsBV1 alone has no significant impacts on biological features of *S. sclerotiorum*, but SsBV1 co-infection with SatlRNA leads to slightly reduced virulence of *S. sclerotiorum*. Besides SsBV1, botybirnaviruses were isolated from *S. sclerotiorum* strain AH16 (co-infection by SsBV2 and a mitovirus) and strain SZ-150 (SsBV3) as well as from Australia strains. SsBV2 is responsible for hypovirulence, whereas SsBV3 shows asymptomatic infection in *S. sclerotiorum* (Ran *et al.*, 2016; Wang *et al.*, 2019).

SsReV1 with spherical virions (approximately 65 nm in diameter) encompasses 11 dsRNA segments (dsRNA1-dsRNA11) and its whole genome is 28,055 bp in length. The conserved 5'-terminal sequences 5'-GAGWUKK-3' (W, U or A; K, U or G) and 3'-terminal sequences 5'-UGCAGUC-3' are found in each dsRNA segment. There is only one ORF in each dsRNA segment excepting dsRNA6 and dsRNA11 in which there were two ORFs. dsRNA1 encoded protein VP1 was predicted to be the RdRp of SsReV1 responsible for virus replication. dsRNA2 encoded protein VP2 is likely to be viral methyltransferase, and dsRNA8 encoded VP8 is likely to be viral kinase and helicase. The biological functions of other proteins from SsReV1 are still unknown. Interestingly, two conserved domains, double-stranded RNA binding motif (dsRBM, Pfam 00035) and reovirus sigma C capsid protein (Reo_σC, pfam04582), were identified in the genome of SsReV1. Polygenetic analysis and multiply alignment indicate that reoviruses indeed have HGT events with other virus lineages, even with cellular organisms on a large scale. SsReV1 has a closer relationship to mammalian coltivirus than to fungal mycoreoviruses. However, dsRNA profile of SsReV1 genome was distinct from those of members within the genera *Coltivirus* and *Mycoreovirus*, suggesting SsReV1 represents a novel genus in the family *Reoviridae*. Reoviruses were reported in *C. parasitica* (MyRV1 and MyRV2), *R. necatrix* (RnMyRV3), and as well in *S. sclerotiorum* (SsMyRV4). The infections by those mycoreoviruses result in virulence debilitation of their hosts, but the *S. sclerotiorum* strain infected by SsReV1 alone shows no symptoms.

Co-infection of strain SCH941 by SsReV1 and SsBV1 failed to cause the typical lesion on rapeseed plant, but infection of a *S. sclerotiorum* strain by SsReV1-alone or SsBV1-alone showed strong virulence on the host. To eliminate the possibility of infection of strain SCH941 by other viruses overlooked by dsRNA extraction, metatranscriptomic analyzes of strain SCH941 was conducted, and finally found three new mycoviruses, *Sclerotinia sclerotiorum yadokarivirus 1* (SsYkV1), *Sclerotinia sclerotiorum deltaflexivirus 1* (SsDFV1), and SsDFV3 coinfecting the strain SCH941 (unpublished data). As a matter of fact, SCH941 was coinfecting with five mycoviruses including two dsRNA mycoviruses and three + ssRNA mycoviruses. Notably, SsBV1 and SsYkV1 always co-exist in a single isolate and SsYkV1 failed to separate from SsBV1, suggesting they could have potential interplay each other. We recently confirmed that neither SsDFV1 nor SsDFV3 is associated with hypovirulence, but hypovirulent phenotypes of *S. sclerotiorum* strains were co-determined by SsBV1, SsYkV1, and SsReV1. Although whether synergistic interaction among three mycoviruses exists in *S. sclerotiorum* still remains unresolved, we found that the accumulation of SsReV1 is significantly increased in *S. sclerotiorum* strains upon co-infections with SsBV1 and SsYkV1 (unpublished data). The SCH941 and its associated mycoviruses supply an excellent system for exploring the quadripartite interaction of fungi-virus-virus-plant.

Megabirnavirus Infects the Hypovirulent Strain SX466 With a Mitovirus and a Partitivirus

Strain SX466 harbors two dsRNA mycoviruses and an ssRNA mycovirus, including *Sclerotinia sclerotiorum megabirnavirus 1* (SsMBV1), SsMV2/SX466, and *Sclerotinia sclerotiorum partitivirus 2* (SsPV2/SX466) (Wang *et al.*, 2015). SsMBV1 forms rigid spherical particles with approximately 45 nm in diameter, and contains two dsRNA segments (dsRNA1 and dsRNA2) each of which possesses two ORFs (ORF1 & 2 on dsRNA1, ORF3 & 4 on dsRNA2). Interestingly, a papain-like protease domain in dsRNA2-encoded protein appears to have been obtained from the single-stranded RNA viruses via horizontal gene transfer events. dsRNA2/SsMBV1 is dispensable for SsMBV1 replication and packaging, but it could enhance transcript accumulation of dsRNA1/SsMBV1 and stability of SsMBV1 particle. This similar result was found in other megabirnaviruses infecting *R. necatrix*. Different from *Rosellinia necatrix megabirnavirus 1* (RnMBV1), SsMBV1 has a slight impact on fundamental biological characteristics of its host regardless of the presence or absence of dsRNA2/SsMBV1. Whether SsMV2/SX466 or SsPV2/SX466, or a combination of three mycoviruses is responsible for hypovirulence in strain SX466 needs to be further explored.

Strain Sunf-M is Coinfected With Two Unrelated dsRNA Mycoviruses

Although strain Sunf-M isolated from the diseased sunflower plant (*Helianthus annuus*) has strong virulence on its hosts, two mycoviruses, *Sclerotinia sclerotiorum nonsegmented virus L* (SsNsV-L) and *Sclerotinia sclerotiorum partitivirus S* (SsPV-S), and a dsRNA element were detected. SsNsV-L is 9124 nucleotides in length without a poly (A) tail. SsNsV-L has two large ORFs (ORF1 and ORF2) and ORF2 is predicted to encode RdRp. Genome comparison and evolutionary analysis suggested SsNsV-L represents a

novel viral evolutionary lineage. Two other strains (above 90% identities with SsNsV-L/Sunf) of SsNsV-L were detected from two strains of *Botrytis* spp.. SsPV-S in Sunf-M is a typical partitivirus based on the RdRp analysis, but its coat protein has the highest amino acid sequence similarity to IAA-leucine-resistant protein 2 (ILR2) of *Arabidopsis*, and its similarity to CPs of other partitiviruses is considerably lower. The systematic analysis further revealed that HGT may have widely occurred between dsRNA viruses and eukaryotic nuclear genomes but previously undetected. The similar result was obtained based on the analysis of Rosellinia necatrix partitivirus 2.

Mitoviruses Widely Exist in and Co-Infect *S. sclerotiorum*

Mitoviruses are usually considered to replicate and locate in mitochondria of fungi. Mitoviruses consist of a single ORF, as revealed by invoking fungal mitochondrial codon usage (UGA coding for tryptophan). Mitovirus encode RdRp, but lacking a protein capsid. Based on the NCBI database, thirty-four *Sclerotinia sclerotiorum* mitoviruses (SsMV) were temporarily named from SsMV1 (accession number JQ013377) to SsMV34 (accession number MF444264), which revealed that mitoviruses are highly prevalent and diverse in *S. sclerotiorum* relative to any other mycoviruses (Fig. 1). Moreover, multiple infection of a single isolate by mitoviruses is very common in fungi, including *S. sclerotiorum*. Two mitoviruses of SsMV1 and SsMV2 were found to co-infect the hypovirulent strain KL-1, and three mitoviruses including SsMV5, SsMV6, and SsMV7 co-infected the strain Lu471. In the following section, we only summarize mitoviruses from hypovirulent strains of *S. sclerotiorum*, and exclude mitoviruses detected via deep-sequencing.

Strain KL-1 was originally isolated from diseased lettuce in USA, and co-infected with two mitoviruses SsMV1 and SsMV2. The full-length cDNA of SsMV1/KL-1 and SsMV2/KL-1 has ~2500 nt, and usually has a lower GC content. Similar to reported mitoviruses, the UTRs of SsMV1/KL-1 and SsMV2/KL-1 have the stem-loop structures. SsMV1/KL-1 and SsMV2/KL-1 are considered to be the causal agents of hypovirulence in strain KL-1. Another strain (SsMV1/HCO25) of SsMV1 was characterized in a Chinese strain HCO25 isolated from the disease soybean. SsMV1/HCO25 shared the highest sequence identity (74%) with SsMV1/KL-1. The

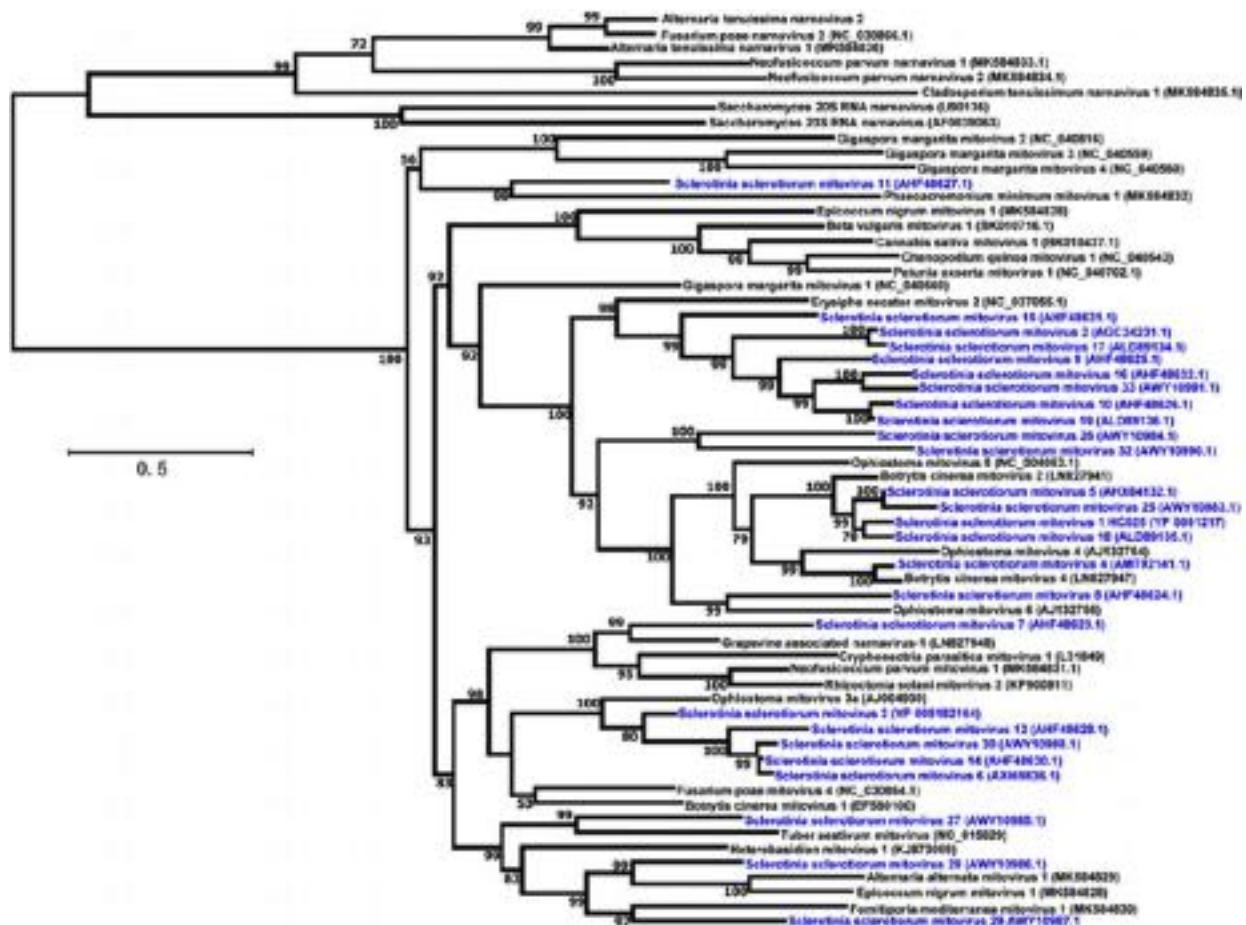


Fig. 1 Maximum likelihood (ML) phylogenetic tree based on multiple alignments of full RdRp region of mitoviruses. Bootstrap values (%) obtained with 1000 replicates are indicated on the branches and only values more than 50 are shown. Branch lengths correspond to genetic distances; the scale bar (0.5) at the left corresponds to the genetic distance. A group of eight narnaviruses was an out-group. Mitovirus infecting *S. sclerotiorum* is indicated by blue typeface. GenBank accession numbers is followed in brackets.

UTRs of SsMV1/HC025 had an inverted complementarity, and could form a potentially stable panhandle structure, which lacks in SsMV1/KL-1. SsMV1/HC025 infection exerted obvious effects on host biological properties, including abnormal colony morphology, lower virulence, and the abnormal morphology of mitochondria.

SsMV2/16235 was detected in a hypovirulent strain 16235 isolated from an infected *Petroselinum crispum* (wild parsley) stem in New Zealand. SsMV2/16235 shares 91.6% amino acid sequence identity with RdRp of SsMV2/KL1, therefore being considered an isolate of SsMV2. A dsRNA profile suggested that at least three dsRNA segments appear in the strain infected by SsMV2/NZ1, and one of the three is the replicate form of SsMV2/16235, the other two segments are those of SsMV3/NZ1 and SsMV4/NZ1. Thus, *S. sclerotiorum* isolate 16235 was co-infected by a minimum of three different mitoviruses. The same research group detected mycoviruses from three *S. sclerotiorum* strains 14563, Lu471, and 11691 via a dsRNA extraction method. They initially dismissed that strain 14563 was co-infected by three mitoviruses (SsMV2/14563, SsMV5/14563, and SsMV6/14563), strain Lu471 harbors three mitoviruses (SsMV5/Lu471, SsMV6/Lu471, and SsMV7/Lu471), and strain 11691 contains SsMV5/11691 and an endornavirus (SsEV1). Shortly afterward, Illumina sequencing revealed more viruses in the same four fungal strains (16235, 14563, Lu471, and 11691): six mitoviruses in strain 16235; five mitoviruses and SsNSRV-1 in strain 14563; six mitoviruses, SsEV1 and SsNSRV-1 in strain Lu471; and five mitoviruses and two endornaviruses in strain 11691. These mycovirus detection patterns are more complex than previous detection patterns obtained with the dsRNA extraction method.

Three Strains of *S. sclerotiorum* Harbor Multiple Unassigned Mycoviruses or dsRNA Elements

Strain AX19 contains at least seven dsRNA segments (dsRNA1 to 7), and the known mycoviruses infecting strain AX19 include hypovirus (dsRNA2), ourmiavirus (dsRNA5), mitovirus (dsRNA6), and an unassigned virus of *Sclerotinia sclerotiorum* deltaflexivirus 1 (SsDFV1, dsRNA3). No information was obtained from the other dsRNA segments. The complete genome of SsDFV1 was predicted to contain a larger ORF and three smaller ORFs. Although a larger ORF encoded protein shares sequence similarities to the members within the order *Tymovirales*, SsDFV1 is markedly distant from all of the members in genomic organization and phylogenetic position. This fact has led the ICTV to create a family *Deltaflexiviridae* that accommodates SsDFV1. The second deltaflexivirus (SsDFV2) was isolated from hypovirulent strain 288 that was coinfecting by a botybirnavirus (BpRV1). SsDFV2 confers hypovirulence on its host and could break the limitation of host vegetative incompatibility to horizontal transmission, but the impact of SsDFV1 on *S. sclerotiorum* still need to be unexplored.

A new +ssRNA mycovirus designated as Hubei sclerotinia RNA virus 1 (HuSRV1) was discovered from the strain 277 of *S. sclerotiorum*. HuSRV1 is 4492 nt long and lacks a poly(A) tail. The genome size of HuSRV1 is smaller than most of mycoviruses except for mitoviruses, but four ORFs were predicted, and encode four putative proteins (protease, RdRp, coat protein, and hypothetical protein with unknown functions). HuSRV1 is placed in a phylogenetic branch distinct from known viruses. In addition, strain 277 was also infected by SsMTV1 that characterized in strain SZ-150, and its hypovirulent phenotypes was shown to be caused by HuSRV1 infection.

The Possible Reasons for the Co-Infections of Mycoviruses in *S. sclerotiorum*

The complex vegetative incompatibility does exist in *S. sclerotiorum* populations. It is usually considered that transmission of mycoviruses is limited by the vegetative incompatibility of host fungi. However, the majority of reported *S. sclerotiorum* strains were infected with multiple mycoviruses (Table 1). More mycoviruses that were not detected by traditional dsRNA extraction could be successfully and accurately detected or discovered by high throughput sequencing. A hypovirulent strain HC051 was confirmed to be co-infected by thirteen mycoviruses including DNA mycovirus, +ssRNA mycoviruses, -ssRNA mycoviruses, and dsRNA mycoviruses (unpublished data), suggesting that co-infections are more complex than we previously thought in *S. sclerotiorum*. Those phenomena also suggested that mycoviruses are likely to overcome the limitation of transmission caused by vegetative incompatibility under nature via some unknown mechanisms. Recently, Thapa and Roossinck (2019) summarized the determinants of co-infection of mycoviruses and how to break the limitation of vegetative incompatibility in filamentous fungi. In *S. sclerotiorum*, we herein list potential reasons for the co-infections of mycoviruses. (1) Some mycoviruses (Such as SsHADV1, SsPV1, and SsDFV2, etc.) are strong infectious. SsHADV1 virions could directly infect hyphae of *S. sclerotiorum* and easily transmit via hyphal contact. SsHADV1 is prevalent among *S. sclerotiorum* strains directly isolated from fields and infects a single strain with other unrelated mycoviruses. (2) Mycovirus (such as SsMYRV4) is capable of promoting horizontal transmission of heterologous mycoviruses between vegetatively incompatible strains. (3) Some mycoviruses act as RNA silence suppressors to resist an antiviral immune system of the host fungus. *S. sclerotiorum* could activate antiviral immune response (such as RNAi) upon mycovirus infections. On the contrary, mycoviruses also have evolved the mechanism to inhibit this host immune response. The satellite RNA of SsHV1 can inhibit the RNAi-related gene expression, therefore increasing the replication of its helper virus. (4) Mycoviruses are capable of escaping from the barrier of vegetative incompatibility system during infection of plants by the fungi hosting the viruses. The horizontal transmission rate of, for example, CHV1 in different vegetative compatible strains is significantly increased when fungal strains infect chestnut than when they grow on artificial medium. A similar phenomenon was found in *S. sclerotiorum* mycoviral transmission (unpublished data), suggesting that the mechanism of horizontal transmission under natural conditions is different from that under laboratory conditions, and remains to be further explored.

References

- Thapa, V., Roossinck, M.J., 2019. Determinants of coinfection in the mycoviruses. *Frontiers in Cell Infection and Microbiology* 9, 169.
- Wang, Q., Cheng, S., Xiao, X., *et al.*, 2019. Discovery of two mycoviruses by high-throughput sequencing and assembly of mycovirus-derived small silencing RNAs from a hypovirulent strain of *Sclerotinia sclerotiorum*. *Frontiers in Microbiology* 10, 1415.
- ## Further Reading
- Bolton, M.D., Thomma, B.P., Nelson, B.D., 2006. *Sclerotinia sclerotiorum* (Lib.) de Bary: Biology and molecular traits of a cosmopolitan pathogen. *Molecular Plant Pathology* 7 (1), 1–16.
- Hillman, B.I., Cai, G., 2013. The family *Narnaviridae*: Simplest of RNA viruses. *Advance Virus Research* 86, 149–176.
- Jiang, D., Fu, Y., Li, G., *et al.*, 2013. Viruses of the plant pathogenic fungus *Sclerotinia sclerotiorum*. *Advances in Virus Research* 86, 215–248.
- Khalifa, M.E., Varsani, A., Ganley, A.R.D., *et al.*, 2016. Comparison of Illumina *de novo* assembled and Sanger sequenced viral genomes: A case study for RNA viruses recovered from the plant pathogenic fungus *Sclerotinia sclerotiorum*. *Virus Research* 219, 51–57.
- Liu, L., Xie, J., Cheng, J., *et al.*, 2014. Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12205–12210.
- Marzano, S.Y.L., Nelson, B.D., Ajayi-Oyetunde, O., *et al.*, 2016. Identification of diverse mycoviruses through metatranscriptomics characterization of the viromes of five major fungal plant pathogens. *Journal of Virology* 90, 6846–6863.
- Mu, F., Xie, J., Cheng, S., *et al.*, 2018. Virome characterization of a collection of *S. sclerotiorum* from Australia. *Frontiers in Microbiology* 8, 2540.
- Nuss, D.L., 2005. Hypovirulence: Mycoviruses at the fungal-plant interface. *Nature Review Microbiology* 3, 632–642.
- Wu, S., Cheng, J., Fu, Y., *et al.*, 2017. Virus-mediated suppression of host non-self recognition facilitates horizontal transmission of heterologous viruses. *PLoS Pathogens* 13 (3), e1006234.
- Xie, J., Jiang, D., 2014. New insights into mycoviruses and exploration for the biological control of crop fungal diseases. *Annual Review of Phytopathology* 52, 45–68.
- Yu, X., Li, B., Fu, Y., *et al.*, 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences of the United States of America* 107, 8387–8392.
- Zhang, H., Xie, J., Fu, Y., *et al.*, 2020. A 2-kb mycovirus converts a pathogenic fungus into a beneficial endophyte for *Brassica* protection and yield enhancement. *Molecular Plant* 13 (10), 1420–1433.
- Zhang, R., Hisano, S., Tani, A., *et al.*, 2016. A capsidless ssRNA virus hosted by an unrelated dsRNA virus. *Nature Microbiology* 1, 15001.

Mycovirus-Mediated Biological Control

Daniel Rigling, Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Cécile Robin, INRAE – French National Research Institute for Agriculture, Food and Environment, UMR BIOGECO, Cestas, France

Simone Prospero, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Biological control A method to control pests and pathogens by using natural enemies.

Horizontal mycovirus transmission Transmission of viruses from infected to non-infected fungal strains via hyphal anastomosis.

Hyphal anastomosis Fusion of fungal hyphae. The formation of hyphal anastomoses allows cytoplasmic exchange and horizontal mycovirus transmission between different fungal strains.

Hypovirulence Reduced ability of a pathogen to cause disease. The term hypovirulence is commonly used in the

context of mycovirus-mediated attenuation of fungal virulence.

Transfection Artificially introducing viral particles or viral transcripts into fungal cells.

Vegetative incompatibility A genetic system that controls the ability of two fungal strains to form hyphal anastomoses and subsequently horizontal mycovirus transmission.

Vertical mycovirus transmission Transmission of viruses to sexual and asexual spores (progeny) produced by virus-infected fungal strains.

Introduction

Fungi represent a major group of pathogens that are able to infect plants and cause diseases in agricultural and forestry production systems. Management of plant diseases mainly depends on resistance breeding and chemical control with fungicides. However, fungal pathogens can overcome plant resistance and can develop tolerance against fungicides. In addition, chemical control can have undesirable side effects on the environment. Biological control offers an alternative disease management tool. Biological disease control utilizes natural enemies or parasites that attack and suppress pathogens. Viruses that infect fungi (so called mycoviruses or fungal viruses) have the potential to act as biological control agents against plant diseases. Mycoviruses have been found in all major groups of plant pathogenic fungi. However, in many cases they live inside the fungal host without causing it harm.

Fungal pathogens can also cause infections in humans and animals, including insects. Mycoviruses have been detected in several of these fungal pathogens. While some are associated with enhanced fungal virulence, others can induce hypovirulence or debilitation in their hosts. This article, however, will focus on mycoviruses that occur in plant pathogenic fungi alone and will not discuss those present in other fungal pathogens.

Disease Cycle of Plant Pathogenic Fungi

Plant pathogenic fungi can have complex life cycles. The disease process typically begins by an initial infection from fungal spores that germinate on, then penetrate into the host plant. This is followed by parasitic growth inside the host tissue. Depending on the type of tissue attacked and the severity of the infection, parts of the plants (e.g., for leaf pathogens) or the entire plant (e.g., for root or stem pathogens) are killed by the pathogens. Finally, the fungal pathogens produce spores on the infected plant tissue. These spores are then dispersed to new hosts by wind, rain, or insect vectors. Many fungal pathogens can reproduce both sexually and asexually. Sexual reproduction takes place after mating of two individuals, which upon meiosis produce genetically recombinant sexual spores. In contrast, asexual spores are produced by single individuals and are all genetically identical. Asexual fungal spores can have two functions. They can cause new infections, but can also act as spermatia during sexual reproduction.

Two main lifestyles are distinguished in plant pathogenic fungi. Biotrophic pathogens derive nutrients from living plant cells, therefore hosts are not killed rapidly. In contrast, necrotrophic pathogens derive nutrients from dead cells that have been killed by the pathogen. Necrotrophic pathogens usually attack weak or damaged plants and can also survive as saprophytes on dead plant tissues.

Basic Concepts for Mycovirus-Mediated Biological Control

For a mycovirus to act as biological control agent, four basic criteria must be fulfilled:

- (1) It must reduce the virulence of the pathogen (i.e., cause hypovirulence) by suppressing at least one phase in the disease cycle of the fungal pathogen. Both the infection or the reproduction process of the pathogen can be targeted. Suppressing either process alone, or in combination will disturb the disease epidemic and reduce the severity of the disease.

- (2) The mycovirus must be transmissible into the fungal pathogen that is causing the disease. Most mycoviruses do not have an extracellular phase. Meaning that they can only be transmitted upon hyphal anastomosis (fusion) and cytoplasmic exchange between fungal individuals. Extracellular transmission of virus particles can also occur, but this seems to be an exception.
- (3) It must be able to overcome the cellular defense mechanisms induced by fungi against virus infections. Fungi use RNA silencing as an antiviral defense pathway. Viral RNA within the fungal host is recognized and broken-down into small RNA fragments. Some mycoviruses can overcome this defense response by suppressing the RNA silencing pathway.
- (4) Finally, the mycovirus must be able to replicate and spread through the fungal mycelia. Most fungi produce networks of filamentous hyphae with actively growing hyphae at the edge of the thallus. Fungal reproductive structures (i.e., fruiting bodies) develop in the older growths. Spread of the virus into actively growing hyphae is important to interfere with mycelial growth and the infection process, as well as to allow for horizontal virus transmission. While spread of the mycovirus into the reproductive structures is required for vertical virus transmission into spores.

Further desired characteristics include the spread and persistence of the biocontrol mycovirus in the pathogen population. Since most mycoviruses do not have an extracellular phase they are fully dependent on their host for their survival and spread. This means the mycovirus can only spread together with its fungal host, which typically disseminates by spores. Consequently, the spread of a mycovirus depends on vertical virus transmission into fungal spores.

General Procedure for Developing a Mycovirus-Mediated Biological Control System

Virus Detection and Characterization

The detection of mycoviruses is the first step towards the development of a biological control system. Isolates of known pathogenic species, which exhibit slow or debilitated growth *in vitro* and/or reduced virulence *in planta* are particularly interesting for virus screening. For necrotrophic pathogens, mycoviruses that cause hypovirulence are more likely detected among fungal isolates recovered from the saprophytic stage of the fungus.

The classical approach for the detection of mycoviruses involves the extraction of double-stranded (ds) RNA from the fungal isolates. This approach exploits the fact that most mycoviruses are RNA viruses that either have a dsRNA genome, or in the case of single-stranded RNA viruses, a dsRNA replicative form. Further detection steps include: synthesis of complementary DNA (cDNA) using reverse transcriptase and sequencing, or cloning of the cDNA fragments into appropriate vectors and sequencing. The alignment of overlapping cDNA clones then allows the assembly of the viral genome. This approach was first used in the 1990s by Donald Nuss and co-workers to determine the complete nucleotide sequence of the *Cryphonectria hypovirus 1* (CHV1) genome, one of the most well-characterized and well-known biocontrol mycoviruses (Table 1).

In recent years, many novel mycoviruses have been detected by using next-generation sequencing, also known as ‘high-throughput’ or ‘deep sequencing’. This detection approach has co-opted RNA-Seq (RNA sequencing) tools that were originally developed for whole transcriptome analysis. The procedure starts with the extraction of total RNA from a fungal isolate, followed by synthesizing and then sequencing the total cDNA. Putative viral sequences are then quality controlled using bioinformatic tools and then identified with BLAST (Basic Local Alignment Search Tool) searches using viral sequence collections. In the most successful searches, a match can be found with conserved motifs in viral genes encoding the RNA-dependent RNA polymerase. These RNA polymerases are responsible for the replication of the viral genomes. The viral RNA polymerase gene is also the first choice for phylogenetic analysis to reveal the taxonomic classification of the detected mycovirus.

Once the nucleotide sequence of a mycovirus has been determined, specific PCR assays can be designed to screen fungal culture collection for the presence of a particular mycovirus. The availability of a large culture collection is important because different fungal isolates can host different virus strains, which often vary in their phenotypic effects on its host. Thus, the potential to explore mycovirus diversity could be crucial for finding suitable biological control agents.

Testing for Phenotypic Effects of the Mycovirus

As outlined earlier, it is essential that a mycovirus has a negative effect on its fungal host to be considered as a biocontrol agent. To confirm a negative mycoviral effect, pairs of isogenic fungal strains that differ only in presence/absence of the particular mycovirus are required. Such pairs can be obtained by: (1) horizontal virus transmission from a virus-infected to a virus-free strain by co-culturing and hyphal fusion (anastomosis) or protoplast fusion; (2) elimination of the mycovirus from an infected strain by using antiviral chemicals (e.g., ribavirin), thermal treatments, or single spore isolations; (3) *in vitro* transfection of protoplasts of a virus-free strain with virus particles or infectious viral transcripts derived from a cDNA clone; or (4) transfection of protoplasts of a virus-free strain with an infectious cDNA clone. Phenotypes that should then be tested must include those related to growth and sporulation on artificial agar medium, as well as *in planta*. Poor or debilitated growth of virus-infected strains on agar medium is, in most cases, a reliable indicator for reduced growth (virulence) *in planta*. The virus effect can be quantified as the difference between the performance (e.g., growth rate, lesion area) of the virus-infected and that of the corresponding isogenic virus-free strain. This should be expressed as a proportion (%) of the performance of the virus-free strain. A negative virus effect therefore results in a negative proportion value.

Table 1 Examples of proven and potential mycovirus-mediated biological control of plant diseases

<i>Fungal pathogen</i>	<i>Mycovirus (Family)</i>	<i>Phenotypic effects on fungal host beside hypovirulence</i>	<i>Transmission properties</i>	<i>Biocontrol experience</i>
<i>Cryphonectria parasitica</i> Necrotrophic bark pathogen causing chestnut blight on <i>Castanea</i> species	CHV1 <i>Cryphonectria hypovirus 1 (Hypoviridae)</i>	Reduced sporulation and pigmentation, female sterility	Horizontal transmission via hyphal contacts, vertical transmission into asexual spores, no transmission into sexual spores	Main method of control of chestnut blight in European orchards and forests: Therapeutic treatments and natural dissemination of hypovirulence, field experiments in North America
	CHV2 <i>Cryphonectria hypovirus 2 (Hypoviridae)</i>	Moderately reduced fungal sporulation and pigmentation	Horizontal transmission via hyphal contacts, scarcely transmitted into asexual spores, no transmission in sexual spores	Laboratory experiments
	CHV3 <i>Cryphonectria hypovirus 3 (Hypoviridae)</i>	Slow growth in culture	Horizontal transmission via hyphal contacts, scarcely transmitted into asexual spores	Therapeutic treatments of chestnut blight cankers in the US
	MyRV1 <i>Mycoreovirus 1 (Reoviridae)</i>	Altered culture morphology	Infectious as virus particles in transfection assays	Laboratory experiments
	CpMV1 <i>Cryphonectria parasitica mitovirus 1 (Narnaviridae)</i>	Mildly reduced mycelial growth	Horizontal transmission via hyphal contacts, vertical transmission into conidia, maternal inheritance in sexual crosses	Laboratory experiments using excised chestnut wood and bark
<i>Ophiostoma novo-ulmi</i> Wilt pathogen causing Dutch elm disease	OnuMV <i>Ophiostoma novo-ulmi mitoviruses (Narnaviridae)</i>	Debilitated growth, reduced conidial viability, mitochondrial malfunction	Horizontal transmission by hyphal contacts, Vertical transmission into conidia	Inoculations studies with elm trees
<i>Sclerotinia sclerotiorum</i> Necrotrophic pathogen causing white mold or stem rot on many plant species	SsHADV1 <i>Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (Genomoviridae)</i>	Abnormal culture morphology, reduced sclerotia size	Horizontal transmission via hyphal contacts, extracellular transmission of virus particles	Preventive application, growth chamber experiments and field trials
	SsPV1 <i>(Sclerotinia sclerotiorum partitivirus 1 (Partitiviridae))</i>	Reduced mycelial growth, reduced sclerotia production	Horizontal transmission via hyphal contacts, transmission into sclerotia, interspecific transmission to other <i>Sclerotinia</i> species.	Laboratory experiments
	SsMV1 <i>Sclerotinia sclerotiorum mitovirus 1 (Narnaviridae)</i>	Reduced mycelial growth, mitochondrial malformation	Horizontal transmission via hyphal contacts	Laboratory experiments with detached leaves
	SsHV2 <i>Sclerotinia sclerotiorum hypovirus 2 (Hypoviridae)</i>	Reduced mycelial growth	Not known	Laboratory experiments with detached leaves
<i>Botrytis cinerea</i> Necrotrophic pathogen causing	BcMV1 <i>Botrytis cinerea mitovirus 1 (Narnaviridae)</i>	Slow mycelial growth, abnormal mycelial sectors, decreased formation of infection cushions	Horizontal transmission via hyphal contacts, vertical transmission to asexual conidia	Laboratory experiments with detached leaves

white mold or stem rot on many plant species	BcRV1 <i>Botrytis cinerea</i> RNA virus 1 (eventually new family)	Reduced mycelial growth, sectorized colony margin	Horizontal transmission via hyphal contacts, vertical transmission to macroconidia	Laboratory experiments with detached leaves
	BcHV1 <i>Botrytis cinerea</i> hypovirus 1 (<i>Hypoviridae</i>)	Decreased formation of infection cushions	Horizontal transmission via hyphal contacts	Laboratory experiments with detached leaves
	BcPV2 <i>Botrytis cinerea</i> partitivirus 2 (<i>Partitiviridae</i>)	Changed culture pigmentation, reduced production of conidia and sclerotia	Horizontal transmission via hyphal contacts	Laboratory experiments with detached leaves
	Bc378V1 <i>Botrytis cinerea</i> CCg378 virus 1 (<i>Partitiviridae</i>)	Reduction of sporulation and laccase activity	Not known	Laboratory experiments with detached leaves
<i>Rosellinia necatrix</i> Necrotrophic pathogen causing root rot in many woody plant species	RnMBV1 <i>Rosellinia necatrix</i> megabirnavirus 1 (<i>Megabirnaviridae</i>)	Reduced mycelial growth	Infectious as virus particles in transfection assays	Greenhouse experiments with apple seedlings
	MyRV3 <i>Mycoreovirus</i> 3 (<i>Reoviridae</i>)	Reduced mycelial growth	Infectious as virus particles in transfection assays	Greenhouse experiments with apple seedlings
<i>Helminthosporium victoriae</i> Necrotrophic pathogen causing Victoria blight of oats	HvV190S <i>Helminthosporium victoriae</i> virus 190S (<i>Totiviridae</i>)	Reduced growth, production of abundant white aerial mycelium, excessive sectoring	Not known, transfection of <i>C. parasitica</i> possible and resulting in a hypovirulent phenotype	Laboratory experiments with apples and branches
<i>Fusarium graminearum</i> Necrotrophic pathogen causing Fusarium head blight of wheat, barley and other small-grain cereals	FgV1 <i>Fusarium graminearum</i> virus 1 (<i>Fusariviridae</i>)	Reduced mycelial growth rate, increased pigmentation, inhibition of mycotoxin production	Horizontal transmission via hyphal contact, vertical transmission to conidia	Laboratory experiments
	FgV-ch9 <i>Fusarium graminearum</i> mycovirus China 9 (<i>Chrysoviridae</i>)	Reduced mycelial growth rate and sporulation, abnormal colony morphology, disorganized cytoplasm	Vertical transmission to conidia	Laboratory and greenhouse experiments
	FgHV2 <i>Fusarium graminearum</i> hypovirus 2 (<i>Hypoviridae</i>)	Reduced mycelial growth rate, conidiation and mycotoxin production	Not known	Laboratory experiments with wheat spikes

Transmission Properties

To be able to exploit a mycovirus's biocontrol ability, a mycovirus needs to be capable of infecting the target pathogen. Since the vast majority of mycoviruses have no extracellular phase, virus infection only occurs upon hyphal fusion between infected and non-infected fungal individuals. Horizontal virus transmission abilities can be tested for on agar medium by co-culturing (pairing) virus-infected and virus-free strains. Mycovirus transmission can be directly observed in the recipient strain in some instances, when the recipient strain adopts the cultural characteristics of the virus donor strain. For example, CHV1 suppresses pigmentation in the chestnut blight fungus, *Cryphonectria parasitica*. Infected fungal isolates exhibit a white culture morphology, compared to orange pigmented virus-free isolates. Horizontal transmission of CHV1 can be followed by observing the change in pigmentation in the virus-free isolate. However, culture morphology is not always a reliable indicator for virus infection. Successful virus transmission must be confirmed by a virus-specific PCR assay or dsRNA extraction. Failure of horizontal mycovirus transmission is often due to vegetative incompatibility (see Glossary) between individuals of the same fungal species. This non-self-recognition system restricts the formation of hyphal anastomoses and consequently the horizontal transmission of mycoviruses.

To be able to spread and persist in a natural environment, the mycovirus needs to be dispersed through the spores of its fungal host. Confirmation of vertical virus transmission into spores requires single-spore cultures derived from a virus-infected isolate. The cultures are then tested for the presence of the mycovirus following the same methodology as horizontal transfer tests. For most mycoviruses, vertical transmission into asexual spores (conidia) is relatively easy to assess because conidia are often formed on agar medium. Transmission into sexual spores is more difficult to assess. In heterothallic fungal species, mating between two sexually compatible fungal isolates is needed for the production of sexual fruiting bodies. For vertical transmission studies, at least one of the two mating partners should be mycovirus-infected. In homothallic species, self-fertilization occurs in a single isolate which produces sexual fruiting bodies without mating type partner. Sexual fruiting bodies may be produced in the laboratory under controlled conditions using defined fungal strains. Alternatively, for highly prevalent mycoviruses, fruiting bodies or sexual spores can be collected in the field and single spore cultures analyzed in the laboratory.

Biocontrol Testing in the Field

Once a mycovirus has been shown to cause transmissible hypovirulence under laboratory and greenhouse conditions, further testing of its biocontrol potential should be conducted in the field. For most applications, a mycovirus-infected fungal strain is used. The mycovirus acts as the biological control agent and the fungal strain is a necessary carrier of the mycovirus. Strain-specific identification of both fungal carrier and mycovirus is important for tracing these organisms after field applications and determining their efficiency and persistence. Traceability markers are based on intra-specific genetic variations and include microsatellite markers for fungal strains and single nucleotide polymorphisms (SNPs) for mycovirus strains. Field trials should also demonstrate the efficacy of the biological control, i.e., a reduction in disease severity and an increase in plant productivity.

Biosafety Issues

Any release of a fungal carrier strain and a mycovirus into the environment requires consideration of biosafety issues. Critical areas of concern include human health (worker and consumer risk) and risk for non-target organisms (e.g., bees, beneficial fungi). In respect to the fungal carrier, the potential production of toxins/secondary metabolites needs to be considered. Such components may not only harm non-target organisms in the field, but could also affect groundwater quality.

Molecular Approaches to Improve the Use of Mycoviruses as Biological Control Agents

Studies in the *Cryphonectria*-hypovirus system have outlined strategies to improve mycovirus-mediated biocontrol using molecular approaches. Spread of the hypoviruses in this system is limited by the fungal vegetative incompatibility system and by exclusion of the virus during sexual reproduction of the fungus. To overcome these limitations, *C. parasitica* strains have been genetically engineered by integrating an infectious cDNA copy of the virus into the fungal genome. In sexual crosses with these strains, half of the progeny (regardless of their vc types) will receive the infectious cDNA copy and become virus-infected. Another promising approach involves gene disruptions at multiple *vic* loci in *C. parasitica* resulting in super donor strains. These strains exhibit enhanced ability to transmit the virus into *C. parasitica* populations with high vc type diversity. The potential to extend the natural host range of mycoviruses was also demonstrated with CHV1, which was introduced as synthetic transcripts into related fungal species.

Selected Examples for Mycovirus-Mediated Biological Control

Mycoviruses that induce hypovirulence may have the potential to be used as biological control agents for several plant pathogenic fungi (Table 1). Hypovirulence in the chestnut blight fungus represents the best-known biocontrol system and will be covered in this section along with other well-known examples.

Hypovirulence in *Cryphonectria parasitica*

The “exclusive transmissible hypovirulence”

Chestnut blight, caused by the ascomycete fungus *Cryphonectria parasitica* (Murr.) Barr., is probably the most well-known example of a tree disease successfully controlled via mycovirus-mediated hypovirulence. Main hosts of this pathogen are species in the genus *Castanea* (family *Fagaceae*), on which the fungus causes perennial bark lesions (so-called ‘cankers’) on stems and branches, eventually leading to wilting of the plant parts distal to the infection point. The fungus originates from Eastern Asia (China, Japan, and Korea) and was accidentally introduced in the 20th century to North America and Europe. In the new continents *C. parasitica* encountered two susceptible chestnut species *Castanea dentata* (Marsh.) Borkh. (American chestnut) and *C. sativa* Mill. (European chestnut).

In North America, the chestnut blight epidemics resulted in the ecological extinction of the formerly forest forming *C. dentata* within its native distribution of 3.6 million hectares. In Europe, after an initial virulent phase of the epidemics that caused high tree mortality, chestnut trees showing atypical superficial, non-lethal bark cankers were observed in Italy and in France. The *C. parasitica* strains isolated from these cankers presented a different phenotype than those recovered from virulent cankers. Specifically, instead of showing an orange pigmentation and a strong sporulation on potato dextrose agar (PDA), they were white and did not sporulate. The atypical strains were reduced in virulence when inoculated on chestnut trees. Further analyses revealed that this phenomenon, which was then called hypovirulence, was associated with the presence of dsRNA in the *C. parasitica* mycelium.

This dsRNA represents the replicative form of a mycovirus in the family *Hypoviridae*, namely *Cryphonectria hypovirus* 1 (CHV1). Hypoviruses are positive-stranded RNA viruses, located in the cytoplasm of the fungal host. Aside from CHV1, the genus *Hypovirus* includes three other well-characterized species: *Cryphonectria hypovirus* 2, *Cryphonectria hypovirus* 3, and *Cryphonectria hypovirus* 4. The exemplar strains of these species, *Cryphonectria hypovirus* 2 (CHV2), CHV3 and CHV4, are phylogenetically related to CHV1, but have different effects on the fungal host. CHV2 and CHV3 also induce a hypovirulent phenotype in *C. parasitica*, whereas the presence of CHV4 does not cause hypovirulence in *C. parasitica*. In North America, CHV3 and CHV4 are the most common hypoviruses, but CHV2 also occurs. In Asia, both CHV1 and CHV2 are present. In Europe, CHV1 is the only hypovirus reported to be present.

CHV1 inhibits *C. parasitica*'s sexual reproduction, strongly reduces asexual sporulation, and attenuates parasitic growth of its fungal host. Hypovirulent *C. parasitica* strains are not able to colonize and kill the cambium of *C. sativa*. As a result, the plant part distal to the infection point survives. Vertical transmission of CHV1 is restricted to asexual spores (conidia) and occurs at variable frequencies. In contrast, sexual spores (ascospores) only spread the virulent form of the pathogen. Horizontal CHV1 transmission to other fungal mycelia via hyphal anastomosis also occurs and is controlled by a vegetative compatibility (vc) system (Fig. 1), involving at least six unlinked, di-allelic vegetative incompatibility (*vic*) loci. Transmission success is highest between vegetatively compatible *C. parasitica* strains, meaning strains that have the same alleles at all *vic* loci. Transmission is severely reduced if strains are heteroallelic (i.e., have different alleles) at five of the six *vic* loci. The sixth locus, *vic4*, does not seem to affect CHV1 transmission. The other five *vic* loci restrict CHV1 transmission in a variable way. In most cases, restriction is asymmetric depending on which *vic* allele is present in the CHV1 donor or recipient strain. Field investigations showed that vc barriers seem to be less restrictive in the field than estimated from in vitro studies. Allelic combinations at the six *vic* loci define a total of 64 *vic* genotypes which correspond to the 64 vc types (EU-1 to EU-64) genetically characterized so far. However, the presence of vc types not compatible with these 64 described types suggest that vegetative compatibility in *C. parasitica* is most likely regulated by more than six *vic* loci or that additional alleles exist at the known *vic* loci.

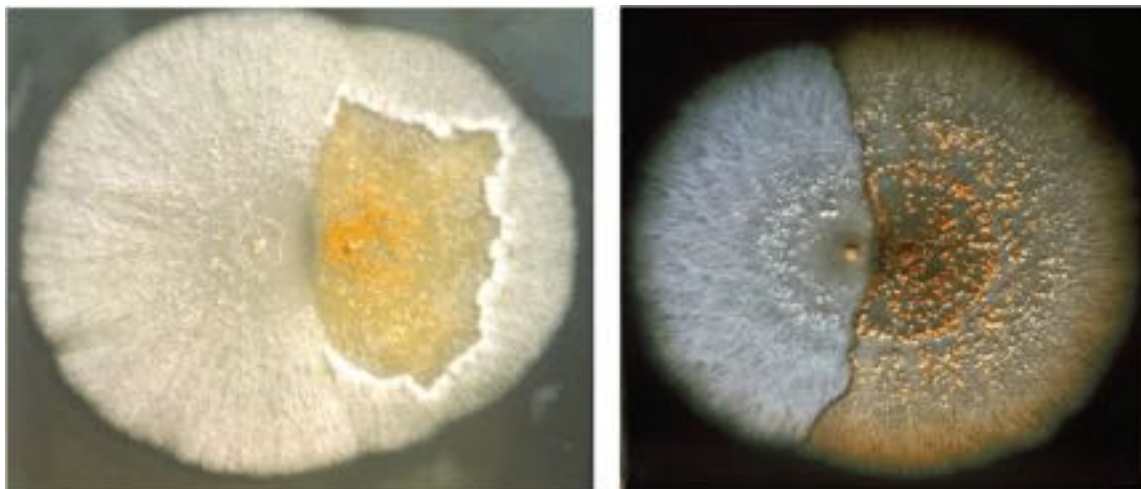


Fig. 1 Horizontal transmission of CHV1. The hypovirus is transmitted between vegetative compatible (left), but not between vegetative incompatible (right) strains of *Cryphonectria parasitica*. CHV1 infected strains exhibit a white culture morphology, which enables the visual assessment of hypovirus transmission. Photos: Phytopathology, WSL.

Several genetically distinct subtypes of CHV1 have been identified in Europe, whose geographic distribution reflects the invasion history of *C. parasitica*. Moreover, several recombination events seem to have contributed to the evolution of CHV1 in Europe. The Italian subtype (subtype I) is the most widespread in central and south-eastern Europe (Italy, Switzerland, south-eastern France, Bosnia-Herzegovina, Croatia, Slovenia, Macedonia, Greece and Turkey). This subtype is represented by the two well characterized strains CHV1/Euro7 and CHV1/EP721. The prototypic hypovirus CHV1 (CHV1/EP713) belongs to the subtype F1. The CHV1 subtype F1 has been found in France and Spain together with subtypes F2, and D/E. In Eurasian Georgia a unique subtype (G) occurs that displays a recombinant pattern between subtypes F2 and D/E. The different subtypes vary in their virulence toward the infected *C. parasitica* strains. Specifically, subtype I has a mild effect on fungal virulence and sporulation and is commonly associated with a high natural prevalence of hypovirulence. In contrast, subtypes F1 and F2 have a severe effect on their fungal host and almost completely inhibit its parasitic growth and sporulation.

Artificial application of hypovirulence

As early as 1965, the field application of hypovirulent strains of *C. parasitica* for biocontrol of chestnut blight was proposed by the French mycologist Jean Grente. In the 1970s in south-eastern France, the hypotheses that the “exclusive transmissible hypovirulence” can be transmitted in the field through inoculations of cankers, as well as that the natural spread of hypovirulence observed in Italy in the 1950s can be accelerated by human intervention, were tested. To this end, virulent cankers were inoculated with hypovirulent *C. parasitica* strains (Fig. 2). Inoculated cankers healed following treatment and hypovirulent strains were isolated from cankers on surrounding trees. Following these first promising results, the French Ministry of Agriculture promoted biological control to help mitigate chestnut blight in France. By 1976, all French orchards had benefited from artificial treatments. Before any field application, the *vc* types present in the designated orchard or stand had to be determined. The necessary canker sampling was organized by INRA (French National Institute for Agricultural Research) in orchards and adjacent coppices. Technical facilities were also provided by INRA to grow and distribute hypovirulent strains. The following biocontrol treatment consisted of stripping the edges of the cankers, making holes with a cork borer in healthy bark tissue around the margins, and filling the holes with a mixture containing one or more hypovirulent strains (in case of different *vc* types). Mixtures were adapted to each production area. From 1998 on, private companies have started to market the hypovirulent strains in France. A license of know-how was established by INRA with these companies in order to control the quality of the product and to be sure that the hypovirulent strains sold are best appropriated for the French *C. parasitica* populations. After 20 years of biological control, in 1995, a new research program on chestnut blight was initiated at INRA in Bordeaux. This program, carried out in close collaboration with several organizations, has allowed to redefine the mixtures of hypovirulent strains to be used, to develop a new method of application of the product, and to characterize the different viral strains and their transmission properties. Chestnut blight is currently well controlled in southern France. Although its incidence in orchards and coppices is still high, chestnut growers no longer consider it a major problem, except for young grafted trees.

In Greece, artificial hypovirulence has also been intensively applied for the biological control of chestnut blight. In 1995, an initial project was started to determine the diversity and distribution of *C. parasitica* *vc* types. Following this (1998–2000), a pilot biocontrol project was carried out in the chestnut coppice stands of Mt. Athos. Roughly 20'000 virulent cankers were treated with a hypovirulent strain of *vc* type EU-12. Promising results from this pilot project led to two nationwide treatment campaigns (2007–2009 and 2014–2016). It is estimated that roughly 170,000 cankers were treated with a CHV1 strain (subtype I), which was naturally present at Mt. Athos. The hypovirulent inoculum was prepared at the Forest Research Institute in Vassilika on a commercial level. In counties where more than one *vc* type were found, a mixture of inoculum of the corresponding *vc* types was applied, as done in France. To help forest managers in organizing the artificial introduction of hypovirulence in areas without naturally occurring hypovirulence, an Integrated Biological Control Plan was developed. Evaluations after the treatment campaigns showed that the treated cankers had healed and the released CHV1 strains had started to naturally infect, and heal, non-inoculated cankers. The type of forest management strongly influenced the spread of the hypovirus, which was faster in coppice forests than in



Fig. 2 Biological control of chestnut blight. Mycelium of a hypovirus-infected *C. parasitica* strain is filled into holes at the margin of the infection. Photo: Phytopathology, WSL.

orchards. Thanks to the successful treatments, Greek forest managers and chestnut growers are currently much less concerned about chestnut blight than 20 years ago and chestnut cultivation has re-gained importance.

Artificial canker treatments with hypovirulent *C. parasitica* strains have also been conducted for more than 20 years in Northern Switzerland. In contrast to Southern Switzerland, no natural hypovirulence has appeared North of the Alps. At roughly 20 sites, more than 2'000 cankers were treated with a local *C. parasitica* isolate of the dominant vc type that was converted with a CHV1 infected isolate from Southern Switzerland. Following the method adopted in France, holes were applied around the virulent cankers and filled with the hypovirulent inoculum. One to two years after treatment, canker morphology was assessed and *C. parasitica* was re-isolated from both treated and untreated cankers. Results revealed the presence of the released CHV1 strain in most treated cankers, which had consequently healed. This confirmed the effectiveness of therapeutic canker treatment in Switzerland. However, there were considerable differences among sites in the spread of the hypovirus to untreated cankers. Although factors determining these differences are still unclear, it seems that the incidence of canker treatment may play an important role. Specifically, a threshold number of cankers must be treated before CHV1 begins to spread into the local *C. parasitica* population.

In North America, natural hypovirulence is associated with the hypovirus CHV3. This has occurred in a few chestnut stands in Michigan and Ontario, which are outside the natural range of the American chestnut. Artificial application of hypovirulence for biological control of chestnut blight in forests and plantations started as early as the mid-1970s in West Virginia, Connecticut, Virginia and Wisconsin. However, results were much less positive than in Europe. Although therapeutic canker treatment has been mostly successful, hypovirulence has largely failed at the population level, meaning that the released hypoviruses did not spread in the local *C. parasitica* populations. Two of the most significant factors responsible for this failure are thought to be the high susceptibility of American chestnut to *C. parasitica* and local vc type diversity which is much higher than in Europe. Molecular approaches have been recently used to enhance the establishment and dissemination of the hypovirus. These include the production of transgenic *C. parasitica* strains carrying an infectious cDNA copy of CHV1 that is integrated into the fungal genome. As well as the production of super hypovirus donor strains, which are able to transmit the hypovirus to recipient strains that are heteroallelic at one-to-five *vic* loci. In a recent field trial in the US, the use of the super donor strains resulted in enhanced hypovirus transmission into treated chestnut blight cankers. This result highlights the potential of these strains to serve as universal hypovirus carriers for biological control of *C. parasitica*, particularly for populations with high vc type diversity.

Therapeutic treatment of individual *C. parasitica* cankers on *C. sativa* and *C. dentata* has been mostly successful and provides an efficient control method for these perennial infections. In contrast, mycovirus-mediated hypovirulence has frequently shown less success at the population level, especially in North America. Natural hypovirulence is still a recent phenomenon in Europe and its success is strongly dependent on characteristics of the three actors involved in this pathosystem, i.e., virus, fungus, and tree. The outcome of the interaction is also affected by the environment. Therefore, predicting the sustainability of this natural biological control system is particularly difficult.

Mycoviruses other than hypoviruses have also been detected in *C. parasitica*. Among these viruses, two mycoreovirus strains and one mitovirus strain can also induce a hypovirulent phenotype. *C. parasitica* isolates infected by either of two mycoreovirus strains (MyRV1 or MyRV2) exhibit a strong hypovirulent phenotype. Both mycoreoviruses are infectious as particles in transfection assays allowing to transfer the virus to fungal isolates in different vc types. The mitovirus strain, CpMV1/NB631, has a mild effect on its host. CpMV1 is transmitted at high frequency into asexual spores and, as expected for a mitochondrial mycovirus is maternally inherited.

Hypovirulence in *Ophiostoma novo-ulmi*

Ophiostoma novo-ulmi Brasier is the causal agent of Dutch elm disease, which is responsible for widespread mortality of elm (*Ulmus* sp.) trees in North America and Europe. The pathogen is spread by elm bark beetles, which transmit fungal spores into feeding wounds on elm twigs. Mycoviruses in *O. novo-ulmi* were originally described as D-factors, inducing severe debilitation of the fungal host. Later on, several D-factors were identified as mitoviruses and assigned to *O. novo-ulmi* mitoviruses 1–7 based on phylogenetic analysis. The mitoviruses are transmitted through asexual conidia but not through sexual ascospores. Horizontal virus transmission can occur via hyphal anastomosis between vegetative compatible strains. The prevalence of mitovirus-infected isolates was found to be high in clonal populations at the epidemic disease front because of lack of restriction by vegetative incompatibility. In post-epidemic populations, the prevalence of viruses decreased as the diversity of vc types increased. Virus-induced hypovirulence in *O. novo-ulmi* is mainly associated with poor mycelial growth and reduced viability of conidia. This means that much more conidia are required for virus-infected isolates to infect elm trees. A proposed approach for biological control is to use elm bark beetles as carrier for virus-infected fungal conidia and thereby introducing the virus into local pathogen populations.

Hypovirulence in *Sclerotinia sclerotiorum*

The ascomycete fungus *Sclerotinia sclerotiorum* (Lib.) de Bary is a necrotrophic plant pathogen that causes white mold or stem rot on a number of different host plants. This includes important agricultural crop species such as bean, rapeseed, soybean, and lettuce. The global pathogen occurs in temperate and subtemperate regions. The pathogen is homothallic and can form sexual fruiting bodies (apothecia) through self-fertilization. The sexual spores (ascospores) spread by wind and serve as primary source for new infections. The pathogen can also reproduce asexually by the production of sclerotia, which serve as survival structure and dissemination propagules.

Several mycoviruses that can cause hypovirulence in *S. sclerotiorum* have been described (Table 1). Most promising as biological control agent is SsHADV-1 (*Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1). This virus is among the few DNA mycoviruses that have been described. To date, it is also the only example where purified virus particles can be used to infect a fungal host. Virus particles that were directly applied to leaves were found to suppress infection by *S. sclerotiorum*. In field trials with rapeseed, the preventive application of hyphal fragments containing SsHADV-1 particles significantly reduced Sclerotinia stem rot and increased crop yield.

S. sclerotiorum also harbors a mycovirus in the family *Hypoviridae*, which induces hypovirulence, namely SsHV2 (*Sclerotinia sclerotiorum* hypovirus 2). Fungal strains infected with this mycovirus exhibit reduced mycelial growth, reduced sporulation, and reduced production of sclerotia. In laboratory experiments with detached leaves, a significant variation in virulence among SsHV2-infected strains was observed. Several of these fungal strains showed a strong reduction in virulence, which could lead to candidate agents for biological control of *S. sclerotiorum*.

Sclerotinia sclerotiorum partitivirus 1 (SsPV1) is another mycovirus that confers hypovirulence in *S. sclerotiorum* and related species. SsPV1 induced phenotypic traits in its fungal host include abnormal colony morphology and severe growth reduction. Horizontal transmission of SsPV1 to other fungal strains has been shown to result in transmission of hypovirulence and associated traits.

A strain of a mitochondrial mycovirus, *Sclerotinia sclerotiorum* mitovirus 1 (SsMV1/HCO25) was also found to induce hypovirulence in *S. sclerotiorum*. An infection by this mitovirus has severe effects on the fungus including reduced mycelial growth, reduced sclerotia production, and mitochondrial malformation. On detached leaves of rapeseed and soybean plants, SsMV1/HCO25-infected fungal strains exhibit a strong hypovirulent phenotype with almost no lesion development. Transmission of SsMV1/HCO25 via hyphal contacts converts recipient virulent strains to the hypovirulent phenotype.

Hypovirulence in *Botrytis cinerea*

Botrytis cinerea Pers. is a cosmopolitan necrotrophic pathogen infecting more than 200 plant species in temperate and subtropical climates, some of which are of high economic importance (e.g., grapes, strawberries, solanaceous vegetable). Symptoms vary across plant organs and tissues but include: leaf blight, blossom blight, and post-harvest fruit rots. In spite of the risk of developing fungicide resistance, the pathogen is still mainly controlled using chemical fungicides. In 2001, *Botrytis*-specific fungicides (so called “botryticides”) represented 10% of the world fungicide market.

Although numerous single- or double-stranded mycoviruses have been reported in *B. cinerea* and some have already been sequenced, not all of them are associated with hypovirulence. The mycoviruses that appear to attenuate the virulence of the infected fungal strains include: *Botrytis cinerea* mitovirus 1 (BcMV1, previously named BcDRV), *Botrytis cinerea* RNA virus 1 (BcRV1), *Botrytis cinerea* hypovirus 1 (BcHV1), *Botrytis cinerea* partitivirus 2 (BcPV2), and *Botrytis cinerea* CCg378 virus 1 (Bc378V1). BcMV1 was found in a hypovirulent *B. cinerea* isolate from oilseed rape in China and appears to debilitate mitochondria in hyphal cells. Laboratory experiments showed that BcMV1 can be transmitted both vertically, to asexual conidia, and horizontally, to other fungal strains. Though the horizontal transmission possibility might be regulated by vegetative incompatibility. BcRV1 was isolated from a Chinese hypovirulent isolate of *B. cinerea* from *Berberis* sp. This mycovirus was found to be capable of vertical transmission via macroconidia and horizontally through hyphal anastomosis. The accumulation level of BcRV1 greatly impacts its attenuation effect on the virulence of *B. cinerea*. In contrast to the two previous mycoviruses, infection by BcHV1 reduces fungal virulence without affecting mycelial growth. The attenuated virulence seems to arise from a significant alteration of the expression of genes associated with the formation of infection cushions, which play a central role during the infection process. BcHV1 can be transmitted horizontally through hyphal anastomosis. BcPV2, a partitivirus in the genus *Alphapartitivirus*, was isolated in an atypically pink strain of *B. cinerea*. Hosts infected with this strain showed a normal vegetative growth, and a reduced production of conidia and sclerotia. Importantly, they also showed an attenuated virulence on several crops, including apples, tomatoes, and potatoes. Following hyphal contact, BcHV1 was successfully transmitted to several virulent *B. cinerea* strains. This resulted in an attenuation of both their pathogenicity and the production of conidia and sclerotia. Bc378V1, a member of the family *Partitiviridae*, was found in a wild-type *B. cinerea* strain that was co-infected by another mycovirus that was not characterized further. Laccase activity and sporulation rates were lower in strains co-infected with both mycoviruses relative to the virus-free strain and strains only infected with Bc378V1. Hypovirulence is therefore most likely conferred by the presence of both mycoviruses.

Despite the presence of several mycoviruses reducing the virulence of the infected *B. cinerea* strains, as of yet none have a real potential as a biocontrol agent. The two main limiting factors are a low competitive ability of mycovirus-infected strains relative to mycovirus-free strains, as well as the limited horizontal transmissibility of the mycoviruses due to vegetative incompatibility.

Hypovirulence in *Rosellinia necatrix*

Rosellinia necatrix is an ascomycete fungus, which causes root rot on many woody plant species including grapevine and many fruit trees. It is a necrotrophic pathogen that also occurs as a saprophyte in the soil. The pathogen is difficult to control due to its soil-borne lifestyle. Different mycoviruses have been detected in *R. necatrix* but often these cause symptomless infections. Two mycoviruses were found to confer hypovirulence: *Rosellinia necatrix* megabirnavirus 1 (RnMBV1) and *Mycoreovirus* 3 (MyRV3). Virus particles of both mycoviruses can be used to transfect virus-free *R. necatrix* strains, which subsequently become hypovirulent.

Hypovirulence in *Helminthosporium victoriae*

The plant pathogenic fungus *Helminthosporium victoriae* (F. Meehan & H.C. Murphy) is the causal agent of Victoria blight of oats (*Avena sativa*). At the end of the 1940s, this fungal pathogen resulted in significant yield losses in most of USA's oat-growing regions. Discovery of a hypovirulence-associated mycovirus occurred in a similar way to that for chestnut blight. In the 1950s, fields of "Victorgrain" oats were observed in Louisiana that, despite an infection by *H. victoriae*, showed only limited yield losses. Isolates recovered from blighted plants presented a stunted and highly sectorized growth. This phenotype was transmitted to 'normally' growing colonies via hyphal anastomosis. The phenotype transmission suggested that the reduced virulence observed in the field was due to a transmissible disease of the fungus. It was later shown that diseased *H. victoriae* isolates harbor two viruses, the victorivirus *Helminthosporium victoriae* virus 190S (HvV190S) and the chrysovirus *Helminthosporium victoriae* 145S (HvV145S). Recent studies have provided strong evidence indicating that HvV190S alone is responsible for the hypovirulence in its fungal host. Coincidentally, laboratory experiments revealed that an infection with HvV190S also induces a hypovirulent phenotype in the chestnut blight fungus *Cryphonectria parasitica*. This might open the door to an artificial introduction of HvV190S-caused hypovirulence in other plant pathogenic fungi.

Hypovirulence in *Fusarium graminearum*

Fusarium graminearum Schwabe is the primary causal agent of Fusarium head blight. This is a severe and worldwide disease of wheat, barley and other small-grain cereals. Quite a few mycoviruses have been reported in this pathogen but three of them (FgV1, FgV-ch9, FgHV2) are associated with hypovirulence. *Fusarium graminearum* virus 1 (FgV1) (proposed family Fusariviridae), causes hypovirulence and reduces rate of mycelial growth, increases pigmentation, and inhibits mycotoxin production. It can be transmitted horizontally via hyphal anastomosis and vertically to asexual conidia. *Fusarium graminearum* mycovirus-China 9 (FgV-ch9) was detected in *F. graminearum* strains recovered from cereals in China. FgV-ch9 can be vertically transmitted to asexual conidia. Based on genomic structure and phylogenetic studies this virus belongs to a novel genus in the family Chrysoviridae. At high and medium FgV-ch9 concentration, infected *F. graminearum* strains show reduced mycelial growth and sporulation (conidia and perithecia), abnormal colony morphology, disorganized cytoplasm, and attenuated virulence on wheat and maize. At low virus concentrations the fungus shows no symptoms of an infection. Finally, the third virus, *Fusarium graminearum* hypovirus 2 (FgHV2) belongs to the genus *Alphahypovirus* and negatively affects mycelial growth rate, conidiation, and mycotoxin production of the infected *F. graminearum* strains.

Future Perspectives

The field of fungal virology has made significant progress in our understanding of the biology of mycovirus-fungus interactions. This research has mainly been driven by the potential of mycoviruses to be used for biological control of fungal pathogens. In recent years a large number of novel mycoviruses have been described in various plant pathogenic fungi. Several of these mycoviruses confer hypovirulence and may have potential as biological control agents for plant pathogens. Thus far, the mycovirus-mediated biological control of chestnut blight is the most successful example of its kind when considering practical applications. Importantly, thanks to the recent promising results in mycovirus research, the perspectives are bright for further successful applications of mycoviruses for biological control of plant diseases.

Further Reading

- Boland, G.J., 2003. Fungal viruses, hypovirulence, and biological control of *Sclerotinia* species. *Canadian Journal of Plant Pathology* 26, 6–18.
- Dawe, A., Nuss, D.L., 2013. Hypovirus molecular biology: From Koch's postulates to host self-recognition genes that restrict virus transmission. *Advances in Virus Research* 86, 109–147.
- Ghabrial, S.A., Castón, J.R., Jiang, D., Nibert, M.L., Suzuki, N., 2015. 50-plus years of fungal viruses. *Virology* 479–480, 356–368.
- Hillman, B.I., Suzuki, N., 2004. Viruses of the chestnut blight fungus, *Cryphonectria parasitica*. *Advances in Virus Research* 63, 423–472.
- Milgroom, M.G., Cortesi, P., 2004. Biological control of chestnut blight with hypovirulence: A critical analysis. *Annual Review of Phytopathology* 42, 311–338.
- Muñoz-Adalia, E.J., Fernández, M.M., Díez, J.J., 2016. The use of mycoviruses in the control of forest diseases. *Biocontrol Science and Technology* 26, 577–604.
- Nuss, D.E., 2005. Hypovirulence: Mycoviruses at the fungal–plant interface. *Nature Reviews Microbiology* 3, 632–642.
- Pearson, M.N., Beever, R.E., Boine, B., Arthur, K., 2009. Mycoviruses of filamentous fungi and their relevance to plant pathology. *Molecular Plant Pathology* 10, 115–128.
- Rigling, D., Prospero, S., 2018. *Cryphonectria parasitica*, the causal agent of chestnut blight: Invasion history, population biology and disease control. *Molecular Plant Pathology* 19, 7–20.
- Suzuki, N., 2017. Frontiers in fungal virology. *Journal of General Plant Pathology* 83, 419–423.
- van de Sande, W.W., Lo-Ten-Foe, J.R., van Belkum, A., et al., 2010. Mycoviruses: Future therapeutic agents of invasive fungal infections in humans? *European Journal of Clinical Microbiology & Infectious Diseases* 29, 755–763.
- Xie, J., Jiang, D., 2014. New insights into mycoviruses and exploration for the biological control of crop fungal diseases. *Annual Review of Phytopathology* 52, 45–68.

Mycoviruses With Filamentous Particles

Michael N Pearson, The University of Auckland, Auckland, New Zealand

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Plasmodesmata Microscopic channels connecting the cytoplasm of adjacent cells in plants.

Triple gene block A set of three genes in overlapping reading frames that function together in facilitating cell-to-cell movement of plant viruses.

Introduction

Mycoviruses have been reported from at least 19 virus families, including viruses with double-stranded (ds) RNA, positive-sense (+), single-stranded (ss) RNA, negative-sense (−) ssRNA(−), and RNA reverse-transcribing genomes, plus at least one ssDNA virus. The majority of these viruses are either unencapsidated or have isometric, encapsidated particles. To date only five mycovirus species with filamentous encapsidated particles have been described (Table 1). Two of these have ssRNA(+) genomes and belong to the order *Tymovirales*: Botrytis virus F (BotV-F), family *Gammaflexiviridae* and Botrytis virus X (BotV-X), family *Alphaflexiviridae*. Two others have ssRNA(−) genomes and belong to the order *Mononegavirales*: Sclerotinia sclerotiorum negative-stranded RNA virus 1 (SsNSRV-1), species *Sclerotinia sclerotimonavirus*, family *Mymonaviridae*, and Fusarium graminearum negative-stranded RNA virus 1 (FgNSRV-1). The fifth is a dsRNA virus Colletotrichum camelliae filamentous virus 1 (CcFV-1) which is phylogenetically related to Aspergillus fumigatus tetramycovirus-1 and Beauveria bassiana polymycovirus-1.

The small number of filamentous mycoviruses reported may simply reflect their rarity compared to other types of mycoviruses but historically it may also reflect the methods most often used to detect mycoviruses. For many years the most commonly used method to screen for the presence of mycoviruses was to extract high molecular weight (MW) dsRNAs, indicative of the presence of either dsRNA viruses or the replicative form of ssRNA viruses. Although this simple technique discovered many RNA mycoviruses it does not detect DNA viruses and also fails to detect some ssRNA viruses, as is well documented for some ssRNA plant viruses, such as luteoviruses and some potyviruses. Amongst the mycoviruses BotV-X and BotV-F are examples of viruses not detected by dsRNA analysis. Although these viruses were discovered during screening of Botrytis cinerea isolates for viral dsRNAs they did not yield sufficient quantities of dsRNA for detection by gel electrophoresis and were only found during virus purification by differential centrifugation, from a B. cinerea isolate that had been chosen as a negative control.

More recently, deep sequencing of total RNA extracts from fungi has detected virus sequences that phylogenetic analysis places in families that include filamentous, encapsidated viruses. For example, the genome of Macrophomina phaseolina tobamo-like virus 1 (MpTLV1) contains four ORFs encoding a methyltransferase and helicase, an RdRp, a putative movement protein, and a coat protein, with a gene order and genome organization typical of members of the genus *Tobamovirus*. Although filamentous particles have not been observed the existence of a coat protein gene suggests that encapsidated particles are produced. In other cases genome analysis may show the absence of a coat protein gene, indicating that encapsidated particles are not produced. For example, Sclerotinia sclerotiorum debilitation associated RNA virus (SsDRV) is classified as a member of the family *Alphaflexiviridae*, a predominantly plant virus family, although it lacks a CP gene and consequently does not form encapsidated particles.

Table 1 Summary of characterized flexuous mycoviruses (September 2018)

Virus name	Genome	Order	Family	Genus	Genome size (Kb)	No. of genome segments	Particle size (nm)	Particle type
<i>Botrytis virus X</i>	ss RNA (+)	<i>Tymovirales</i>	<i>Alphaflexiviridae</i>	<i>Botrexvirus</i>	7	1	13 × 720	filamentous nucleoprotein
<i>Botrytis virus F</i>	ss RNA (+)	<i>Tymovirales</i>	<i>Gammaflexiviridae</i>	<i>Mycoflexivirus</i>	7	1	13 × 720	filamentous nucleoprotein
<i>Sclerotinia sclerotimonavirus</i> ^a	ss RNA (−)	<i>Mononegavirales</i>	<i>Mymonaviridae</i>	<i>Sclerotimonavirus</i>	10.0	6	22 × 200–2000	filamentous enveloped nucleoprotein
<i>Fusarium graminearum</i> negative-stranded RNA virus 1	ss RNA (−)	<i>Mononegavirales</i>	Unassigned	Unassigned	9.1	5	35–50 × ~1200	filamentous enveloped nucleoprotein
<i>Colletotrichum camelliae</i> filamentous virus 1 (CcFV – 1)	dsRNA	Unassigned	Unassigned	Unassigned	12.3	8	12–18 × 1000–4427	filamentous nucleoprotein

^aSyn. = *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNSRV-1).

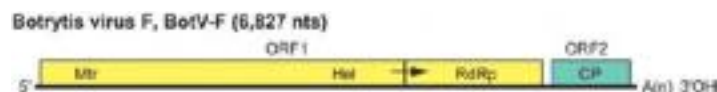


Fig. 1 Genome organization of Botrytis virus F showing the relative positions of the ORFs and their expression products. Mtr, methyltransferase; Hel, helicase; RdRp, RNA-dependent RNA polymerase; CP, capsid protein. The arrow marks the “leaky” opal stop codon in ORF1. Copyright © 2017, International Committee on Taxonomy of Viruses [ICTV].

Botrytis virus F (BotV-F)

Family – *Gammaplexiviridae*

Genus – *Mycovirus*

Genome Structure

BotV-F was the first flexuous mycovirus to be characterized and also the first fully sequenced virus from *B. cinerea*. The genome consists of a single molecule of linear ssRNA(+), 6827 nt in length, excluding the 3'-poly(A) tail. The genomic RNA consists of two major ORFs separated by a 93 nucleotide untranslated region (UTR), a 5'-UTR of 63 nt and a 3'-UTR of 71 nt, followed by a poly(A) tail (Fig. 1). ORF1 contains conserved methyltransferase and helicase motifs, terminating with an opal stop codon (UGA) and yielding a protein of 153 kDa. Readthrough of this codon is expected to extend the protein to 212 kDa that also includes an RNA-dependent RNA polymerase (RdRp) domain. ORF2 encodes the only structural protein, the 32 kDa coat protein. One obvious difference between the genome structure of BotV-F and those of related plant viruses is the apparent lack of a movement protein that many plant viruses encode to mediate the passage of infectious viral material from cell to cell by interaction with the plant host plasmodesmata. However, fungal viruses such as BotV-F presumably do not require movement proteins as the septal pore, which separates individual compartments of the mycelium, is 150–250 nm and therefore unlikely to be a barrier to the movement of mycovirus particles.

Biological Properties

BotV-F was originally detected in *B. cinerea* isolated from strawberry in New Zealand as a co-infection with Botrytis virus X (BotV-X) (genus *Botrexvirus*, family *Alphaflexiviridae*). BotV-F was subsequently detected in *B. cinerea* isolates from grapevine in New Zealand, cucumber in Israel, lettuce in the UK, and tomato and grapevine in France. Surveys of *B. cinerea* isolates from strawberry plants (*Fragaria* sp.) from the Auckland region and isolates from grapevine (*Vitis* sp.) across New Zealand detected BotV-F in 8/59 (13.6%) of isolates. In addition 16% (4/25) of *B. cinerea* isolates from various host plant species around the world also tested positive for BotV-F. Phylogenetic analysis of thirteen BotV-F isolates, based on ~300 bp sequences spanning the intergenic region between ORFs 1 and 2, placed the isolates in three distinct clades. However, the three clades showed no significant correlation with the geographic origin of the isolates suggesting that BotV-F has a long association with *B. cinerea* and has been distributed with its host fungus, most probably in *B. cinerea* infections of plants.

B. cinerea transfected in-vitro with purified BotV-F particles showed changes in hyphal morphology, reduced linear growth in culture, reduced virulence in a *Phaseolus vulgaris* detached leaf assay and significant changes in amino acid, carbohydrate, and fatty acid metabolism.

Virion Morphology

The virions of BotV-F are flexuous nucleocapsid filamentous with a modal length of approximately 720 nm and a diameter of approximately 13 nm (Fig. 2). The virus particles can be partially purified by chloroform treatment of homogenised fungal mycelium and differential centrifugation.

Phylogenetic Relationships

When first characterized in 2001, sequence analysis of BotV-F indicated that it was most similar to plant viruses in the (then) family *Flexiviridae*, although it did not fit well into any of the existing genera. The complete replicase protein showed a highest amino acid identity (22%–23%) to those of the allxivirus, garlic virus X (GVX), the tymovirus, turnip yellow mosaic virus (TYMV), the potexvirus, strawberry mild yellow edge-associated virus (SMYEV), and the vitivirus, grapevine virus B (GVB). In the methyltransferase region, BotV-F showed highest amino acid identity (34%) to the tymovirus, TYMV and the marafivirus, oat blue dwarf virus (OBDV), while in the helicase region it was closer to GVX and SMYEV (28%) and in the RdRp to the allxivirus garlic virus A (GarV-A) (40%); the potexvirus SMYEV (39%), and the capillovirus, cherry virus A (CVA) (39%). The amino acid sequence of the putative coat protein of BotV-F shows highest homology to the conserved central core regions of viruses belonging to the genera *Capillovirus*, *Trichovirus*, and *Vitivirus*, but is unusual in having a long C-terminal region compared with the coat proteins of plant viruses. Another unusual feature of BotV-F, compared to flexuous filamentous ssRNA(+) plant viruses, is the putative

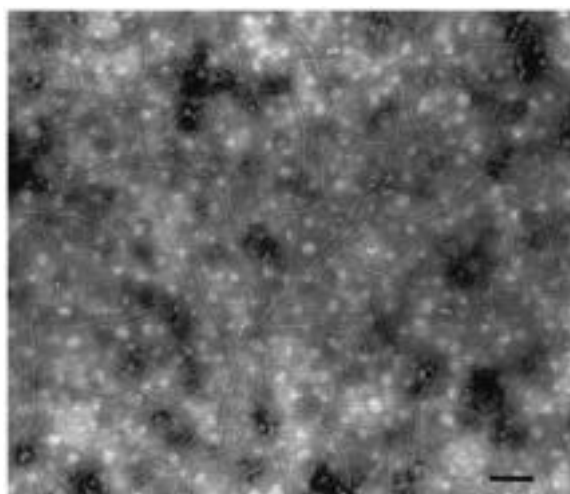


Fig. 2 Electron micrograph of flexuous rod-shaped particles in a partly purified preparation of *Botrytis cinerea* isolate RH106–10, negatively stained with 2% potassium phosphotungstate, pH 4. Bar = 200 nm. From Howitt, R.L.J., Beever, R.E., Pearson, M.N., and Forster, R.L.S., 2001. Genome characterization of Botrytis virus F, a flexuous rod-shaped mycovirus resembling plant ‘potex-like’ viruses. *Journal of General Virology* 82, 67–78.

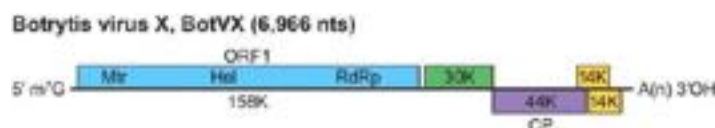


Fig. 3 Genome organization of Botrytis virus X showing the relative positions of the ORFs and their expression products. Mtr, methyltransferase; Hel, helicase; RdRp, RNA-dependent RNA polymerase; CP, capsid protein. Copyright © 2017, International Committee on Taxonomy of Viruses [ICTV].

read-through opal codon in the replicase, which is more typical of read-through codons found in the bipartite genomes of ssRNA straight rod-shaped viruses such as the tobamoviruses and furoviruses.

In the 9th Report of the ICTV (2011) BotV-F was assigned as the type member of a newly created genus *Mycoflexivirus* within a newly created family *Gammaflexiviridae*, order *Tymovirales*.

Botrytis Virus X (BotV-X)

Family – *Alfalflexiviridae*

Genus – *Botrexvirus*

Genome Structure

The BotV-X genome consists of a single molecule of linear ssRNA(+), 6966 nt in length, excluding the 3′-poly(A) tail. The genome comprises five putative ORFs, a 5′-UTR of 95 nt, untranslated regions of 17 nt separating ORFs 1 and 2, and 6 nt separating ORFs 2 and 3, and a 3′-UTR of 149 nt, followed by a poly(A) tail (Fig. 3). The genome has a 5′ cap immediately followed by the nucleotide sequence GAAAAC, typically found in members of the plant virus genus *Potexvirus*. ORF1 encodes the 158 kDa polymerase. ORF2, which is in the +1 frame relative to ORF1, encodes a 30 kDa putative protein (p30) but nucleotide and amino acid sequences of p30 do not reveal any significant homology with other known sequences. ORF3, which is in the same frame as ORF2, encodes the 44 kDa coat protein and terminates at an amber (UAG) codon. ORF4 and ORF5 both encode putative proteins with an expected molecular mass of ~14 kDa. ORF4 overlaps ORF3 and is in the same frame as ORF1, while ORF5 is in the same frame as ORF2 and ORF3. Neither of these proteins show any significant homology with known viral proteins.

Biological Properties

BotV-X was first discovered infecting an isolate of the plant pathogenic fungus *B. cinerea* from a strawberry plant in New Zealand, which was also infected with BotV-F. BotV-X was subsequently detected in other New Zealand *B. cinerea* isolates from grapevine and tomato.

The transmission rate of BotV-X through asexual conidia is >95% and under experimental conditions transmission has also been shown through sexually produced ascospores from both the male parent (34% transmission) and female parent (53% transmission). Since *B. cinerea* is rarely observed to produce ascospores in nature they presumably play a minor role in virus dispersal but sexual reproduction may provide a mechanism of the virus moving into new vegetatively incompatible groups of the fungus.

Immunofluorescent light microscopy of sections through BotV-X infected fungal tissue and fixed mycelia grown on glass slides, using antibodies raised against BotV-X coat protein expressed in *Escherichia coli*, showed (presumed) virus aggregates in hyphal tips, closely associated with fungal cell membranes and walls, and adjacent to septal pores. Although it is generally assumed that mycoviruses move passively through the cytoplasm it is not known exactly how flexuous viruses of filamentous fungi are transported from cell to cell. The occurrence of virus aggregates next to septa suggests that the septal pores of *B. cinerea* may restrict viral movement, although the presence of viral aggregates on both sides of the septum suggests that movement through septa does occur.

In an apple assay, naturally BVX-infected isolates grew slightly slower (i.e., were less virulent) than virus-free isolates. However, *B. cinerea* isolates transfected in-vitro with purified BotV-X particles showed no significant difference in virulence compared to the isogenic virus free line, in either bean leaf or apple fruit assays. In culture on malt extract agar plates some naturally infected *B. cinerea* isolates showed increased growth rates compared to BotV-X-free isolates.

Virion Morphology

The virions of BotV-X are flexuous nucleoprotein filaments with a modal length of ~720 nm and a diameter of ~13 nm (Fig. 2). The virus particles can be partially purified by chloroform treatment of homogenised fungal mycelium and differential centrifugation.

Phylogenetic Relationships

ORF1 encodes the 158 kDa replicase protein (p158) the amino acid sequence of which shows highest identity to replicases from the allxivirus garlic virus A [GarV-A] (49%) and the potexvirus cymbidium mosaic virus [CymMV] (36%). Amino acid alignments of the three internal conserved methyltransferase, helicase, and RdRp domains of GarV-A show 51%, 55%, and 73% identity, respectively, while in CymMV identity to these regions ranges from 45% to 51%. Phylogenetic analysis of the RdRp region clusters BVX amongst the *Allexivirus* and *Potexvirus* genera but closest to the allxiviruses. The first 211 amino acids of the p44 protein (ORF2) do not show significant homology with known proteins, but the C-terminal 189 amino acids (encoding a 20 kDa protein) show significant identity to the coat proteins of CymMV (35%), the allxivirus, garlic virus X [GarV-X] (34%) and the carlavirus, garlic latent virus [GarLV] (34%). Alignment of the conserved residues in the coat protein, associated with the putative salt bridge and corresponding hydrophobic core in flexuous rod-shaped ssRNA plant viruses, also showed highest amino acid identity to CymMV (65%). Phylogenetic analysis of the putative coat protein gene clusters BotV-X with the genus *Allexivirus*.

Although BotV-X has a different genome structure from both allxiviruses and potexviruses it has a number of properties in common with the viruses of these two genera. While it displays high replicase sequence similarity to the allxiviruses, other features, such as the size of the replicase gene, the overall size of the genome, and the conserved 5' terminal nucleotide sequence, are more typical of potexviruses. In the 9th Report of the ICTV (2011) BotV-X was assigned as the type member of a newly created genus *Botrexvirus* within a newly created family *Alfalexiviridae*.

Sclerotinia Sclerotimonavirus

Family – Mymonaviridae

Genus – Sclerotimonavirus

Genome Structure

The genome of the sclerotimonavirus *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNSRV-1) is a 10,002 nt ssRNA (–) that lacks a poly(A) tail at the 3' terminus and has an uncapped 5' terminus (Fig. 4). It consists of six non-overlapping ORFs (I–VI), linearly arranged and separated by conserved gene-junction sequences commonly found in mononegaviruses. ORF V encodes the largest protein, which contains a conserved mononegaviral RdRp domain. ORF II appears to encode the two proteins (p43 and p41) that have been isolated from purified nucleocapsids of the virus. In mononegaviruses ORF IV typically encodes glycoprotein G, however it's unlikely that ORF IV of SsNSRV-1 encodes G protein as it is only 186 bp and lacks a putative signal peptide.

Biological Properties

SsNSRV-1 of *Sclerotinia sclerotimonavirus* confers hypovirulence on *Sclerotinia sclerotiorum*. Symptoms include slow growth on potato dextrose agar (PDA), loss of the ability to produce sclerotia, loss of pathogenicity on rapeseed and zigzag growth of hyphal tips in culture (possibly due to loss of growth polarity). SsNSRV-1 infected *Sclerotinia sclerotiorum* has been isolated from both soybean and rapeseed in China and is widely distributed, having been detected in five provinces (Anhui, Heilongjiang, Hubei, Jiangxi, and Shaanxi) representing a range of climatic regions.

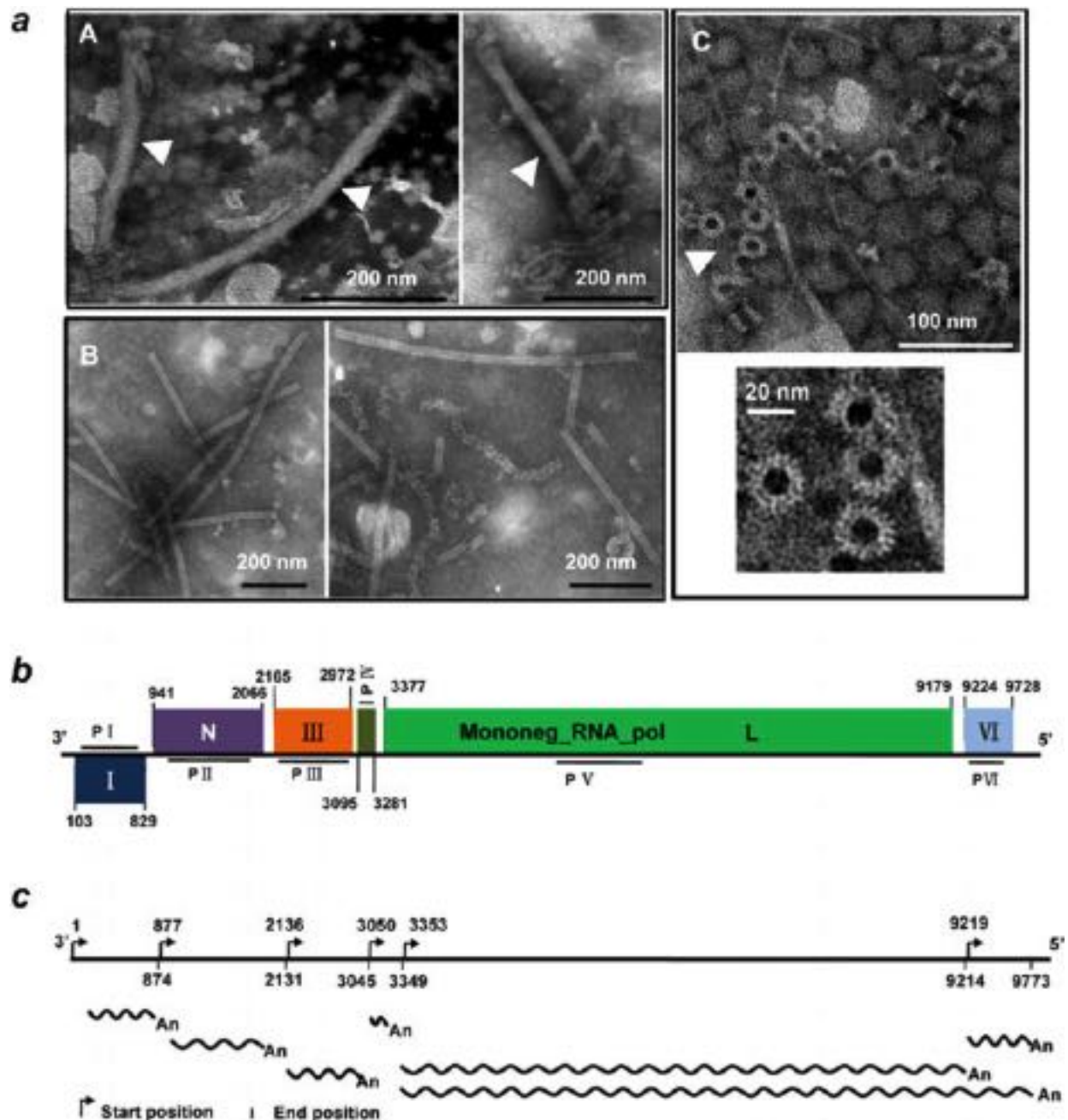


Fig. 4 Virion morphology and genome organization and expression map of *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNSRV-1). (a) Morphology and structure of SsNSRV-1 virions and nucleoprotein–RNA complexes (RNPs). Particles and RNPs were purified from mycelia of strain AH98 and Ep-1PNA367-PT2 and negatively stained with 2% PTA (W/V, pH 7.4) (scale bars are shown at the bottom). (A) Filamentous, possibly enveloped virions and RNPs. (B) Purified tight or loose coils of RNPs. (C) Rings that make up the coils and NP monomers. (b) Genomic organization of SsNSRV-1 showing all six ORFs. (c) Deduced transcription map of SsNSRV-1 based on 5′- and 3′-RACE. Fig. 1 From, 2015.008a-gM.A.v3.Mymonaviridae.pdf (https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/animal-dsrna-and-ssrna-viruses/5831). Modified from Liu, L., Xie, J., Cheng, J., *et al.*, 2014. Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12205–12210.

Virion Morphology

The virions of SsNSRV-1, purified from mycelia and negatively stained with 2% PTA, are filamentous, 25–50 nm in diameter and 200–2000 nm in length, and appear to be enveloped (Fig. 4). The nucleocapsids are long, flexible, helical structures, which, when tightly coiled, have a diameter of 20–22 nm and possibly contain two nucleoproteins of ~43 kDa (p43) and ~41 kDa (p41), both translated from ORF II. The nucleocapsids of SsNSRV-1 (as observed by TEM) consist of tightly stacked turns and have variable

appearance, including both relatively rigid rods and loosely coiled helices, similar to the RNA–nucleoprotein complexes (RNPs) reported from other mononegaviruses. Electron micrographs of ultrathin fungal hyphal sections show tubular structures, similar to the purified virions, consisting of nucleocapsids enclosed within an outer membrane. Mononegaviruses typically have enveloped virions but the apparently stable filamentous shape of SsNSRV-1 is unusual as most mononegaviruses have pleomorphic virions.

Phylogenetic Relationships

SsNSRV-1 shares a number of general characteristics in common with mononegaviruses, although there are some differences. Mononegaviruses typically have nonsegmented ssRNA(–) genomes, 8.9–19 kb in size and virions of variable morphologies that are enveloped with a phospholipid membrane originating from the host. Phylogenetic analysis of RdRp amino acid sequences clusters SsNSRV-1 with viruses of the families *Nyamiviridae* and *Bornaviridae* (order *Mononegavirales*), although SsNSRV-1 does not conform to the typical five-gene pattern (N-P-M-G-L) of mononegaviruses as it has six ORFs, and the NPs are encoded by ORF II instead of ORF I. The 2018 ICTV report classifies SsNSRV-1 in a newly created genus *Sclerotimonavirus*, family *Mymonaviridae*, Order *Mononegavirales*.

Fusarium Graminearum Negative-Stranded RNA virus 1

Family – Mymonaviridae (?)

Genus – Unclassified

Genome Structure

Fusarium graminearum negative-stranded RNA virus 1 (FgNSRV-1) has a ss RNA(–) genome of 9072 nucleotides in length, consisting of five ORFs (ORF I–V) of 774, 1167, 238, 5893 and 552 nt, respectively, and lacks a poly (A) tail at the 3′-terminus. The five ORFs putatively encode five proteins (P I to P V) with predicted molecular masses of 29.1, 42.8, 6.2, 221.6, and 21.1 kDa, respectively. The 3′-untranslated region (UTR) and 5′-UTR are 126 nt and 220 nt long, respectively, and have perfect complementarity for the first six residues (3′-UCCUGC–GCAGGA-5′), a feature common among mononegaviruses. The P II and P IV proteins show significant amino acid sequence similarity to the nucleoprotein N and large protein L of SsNSRV-1 sequences, with the P IV containing mononegaviral-like L polymerase domains. However, similar to SsNSRV-1, the genomic arrangement of FgNSRV-1 does not follow the basic five-gene pattern (N-P-M-G-L) of typical mononegaviruses, the N protein being encoded by ORF II, rather than ORF I, as in most other known mononegaviruses.

Biological Properties

FgNSRV-1 was originally isolated from the plant pathogenic fungus *Fusarium graminearum*. It has no discernable impact on host morphology, mycelial growth, biomass, conidia production, or virulence.

Particle Morphology

The virions of FgNSRV-1 (purified from mycelia) are filamentous, 35–50 nm in diameter and ~1200 nm in length. Short rod-like particles and loose nucleocapsid-like helical structures, similar to the condensed nucleocapsid structures of SsNSRV-1, have also been observed. Given the similarities between FgNSRV-1 and SsNSRV-1, it is possible that the FgNSRV-1 virions are also enveloped by a membrane.

Phylogenetic Relationships

FgNSRV-1 shows high sequence similarity (93.9%–98.8% nucleotide; 97.3%–100% amino acid) to soybean leaf-associated negative-stranded RNA virus 1 (SsNSRV-1), a virus of unknown morphology identified in a soybean leaf metatranscriptomic study. The putative nucleoprotein (P II) and RNA polymerase (P IV) also show significant homology to mononegaviruses with phylogenetic analysis of entire polymerase sequences indicating that FgNSRV-1 is related to the ssRNA(–) mycovirus SsNSRV-1, suggesting that FgNSRV-1 can be classified in the family *Mymonaviridae*.

Colletotrichum Camelliae Filamentous Virus 1

Family – Unclassified

Genus – Unclassified

Genome Structure

The genome of *Colletotrichum camelliae* filamentous virus 1 (CcFV-1) consists of eight separate dsRNAs, of 2444, 2253, 2012, 1299, 1122, 1085, 1053, and 990 bp. DsRNAs 1, 2, 3, 4, 6, and 7 each consist of a single ORF, while dsRNAs 5 and 8 each contain

two ORFs (one >200 aa, the other <50 aa). DsRNA 1 shares 47% amino acid identity with the RNA-dependent RNA polymerase (RdRp) of *Aspergillus fumigatus* tetramicrovirus-1 (AFuTmV-1) and low similarities to the dsRNA1 of *Alternaria tenuissima* virus (ATV), *Cladosporium cladosporioides* virus 1 (CcV-1), and *Botryosphaeria dothidea* virus 1 (BdRV1). The remaining RNAs show no detectable similarity with any known viral RNA sequences. The predicted protein from ORF 1 clusters CcFV-1 with AFuTmV-1, BdRV1, *Beauveria bassiana* polymycovirus-1 (BbPmV-1), and *Beauveria bassiana* polymycovirus-2 (BbPmV-2), separate from other dsRNA viruses and closer to ssRNA(+) viruses belonging to the *Caliciviridae*. It has been suggested that CcFV-1 may be an intermediate link between dsRNA and ssRNA(+) viruses. The protein coded by dsRNA 3 of CcFV-1 is highly similar to the methyltransferase of various bacteria but the functions of the remaining putative proteins are unknown.

Biological Properties

CcFV-1 is a dsRNA virus that was first reported from an isolate of the plant pathogenic fungus *Colletotrichum camelliae*, from tea (*C. sinensis* (L.)) in China. CcFV-1 infected *C. camelliae* exhibited different morphology and pigmentation from virus-free isolates and showed reduced growth rates (5.4 versus 6.3–6.5 mm/day) and reduced virulence with lesion lengths of 1.0 mm compared to 6.1–14.9 mm for virus free isolates. The CcFV-1 infected isolate also showed cytological abnormalities in ultra-thin hyphal sections observed by TEM, including the formation of large membranous vesicles that contained virus-like particles with a width similar to those of the CcFV-1.

Virion Morphology

CcFV-1 infected cultures contain non-enveloped flexuous filamentous particles with widths ranging from 11.9 to 17.7 nm and lengths from 1001 to 4427 nm, with analysis of particles from different fractions of ultracentrifuged sucrose gradients indicating that dsRNAs 1–8 are separately encapsidated. DsRNA 4 codes for the capsid protein, as demonstrated by immunosorbent electron microscopy, Western blotting and indirect enzyme-linked immunosorbent assay, using antisera produced to purified P4 protein.

Phylogenetic Relationships

Colletotrichum camelliae filamentous virus 1 is currently an unclassified virus species. Although the RdRp of CcFV-1 shows some degree of identity to the RdRps of AFuTmV-1, BdRV1, BbPmV-1, and BbPmV-2, the first three of these viruses are capsidless with genomes consisting of 4–5 dsRNAs, while BbPmV-2 has 7 dsRNAs and is possibly capsidless or may form bacilliform virus-like particles. In contrast CcFV-1 forms encapsidated filamentous particles and has eight dsRNAs. In addition the RdRp of CcFV-1 appears to be in an intermediate position between those of ssRNA(+) viruses ssRNA(+) viruses of the extended picorna-like superfamily such as caliciviruses, astroviruses, and picornaviruses, and dsRNA viruses. It has been proposed that CcFV-1 may have originated from a ssRNA(+) viral ancestor belonging to clades three or six of the picorna-like superfamily although the CP has no detectable sequence similarity with those of known filamentous ssRNA(+) plant viruses.

Relationships Between Filamentous Viruses From Fungi and Plants

Mycoviruses have been reported from at least 19 virus families and one genus, eleven of which (*Barnaviridae*, *Chrysoviridae*, *Deltaflexiviridae*, *Gammaplexiviridae*, *Genomoviridae*, *Hypoviridae*, *Megabirnaviridae*, *Myonnaviridae*, *Narnaviridae*, *Quadriviridae*, and *Botybirnavirus*) currently include only fungal viruses. The other nine families include viruses from a range of hosts: *Alphaflexiviridae* (fungi and plants), *Amalgaviridae* (fungi and plants), *Botourmiaviridae* (fungi and plants), *Endornaviridae* (fungi and plants), *Metaviridae* (fungi, invertebrates, plants and vertebrates), *Partitiviridae* (fungi, plants, protozoa), *Pseudoviridae* (fungi, algae, invertebrates, plants, vertebrates), *Reoviridae* (fungi, invertebrates, plants and vertebrates), *Totiviridae* (fungi and protozoa), although in the majority of cases the fungal viruses are classified in distinct genera. Isometric encapsidated mycoviruses are relatively common and are found in at least six of these families and the one genus, but to date (2019) only five encapsidated filamentous mycoviruses have been fully sequenced and described. Although these five filamentous mycoviruses are taxonomically diverse they have one important factor in common, they are all from plant pathogenic fungi.

BotV-F and BotV-X, are ssRNA(+) and belong to the order *Tymovirales* which contains predominantly plant viruses. SsNSRV-1, belonging to the species *Sclerotinia sclerotimonavirus*, and FgNSRV-1, as yet unassigned to a species, have ssRNA(–) genomes and belong to the order *Mononegavirales*, which includes viruses from a wide range of hosts, including vertebrates, invertebrates and plants. However, the filamentous virions of SsNSRV-1 and FgNSRV-1 are somewhat different from other members of the *Mononegavirales* where the virions are typically spherical, bacilliform or plemorphic, although the nucleocapsids may have a filamentous form. The fifth virus is a dsRNA virus CcFV-1, which is phylogenetically related to the unclassified viruses AFuTmV-1 and BbPmV-1. However, the particles of CcFV-1 have a more defined filamentous structure than the other two viruses, which do not encode a conventional CP and are described as being non-conventionally encapsidated, since they encode a proline/alanine/serine (PAS)-rich protein which most likely coats the viral dsRNA.

Interestingly, despite the apparent scarcity of flexuous filamentous mycoviruses, the first two filamentous mycoviruses to be characterized, BotV-X and BotV-F, were found in the same *B. cinerea* isolate. While they differ in genome structure the genomes are of a similar size and virus purification from the co-infected *B. cinerea* isolate yielded only one size of flexuous filamentous particle, indicating that the two viruses have similar-sized particles. BotV-X and BotV-F show only low to moderate sequence identity to each other in the replicase and coat protein genes and there is no significant similarity between other putative proteins encoded by the two genomes. However, the replicase genes of both viruses shows greatest similarity to plant viruses belonging to the order *Tymovirales*, with a highest amino acid identity of 73% between the RdRp of BotV-X and the plant virus GarV-A. While this is remarkably high for interviral homology between plant and fungal viruses there are many examples of homology between plant and fungal viruses from a range of viral families including the *Endornaviridae*, *Totiviridae*, *Chrysovriidae*, *Amalgaviridae*, and *Partitiviridae*. Despite some sequence similarities the genome structures of BotV-F and BotV-X are distinct from those of the plant virus families of the *Tymovirales*, one obvious difference being the absence of a gene encoding a movement protein in BotV-X and BotV-F. In plant viruses of the *Tymovirales*, movement proteins, which interact with plant host plasmodesmata to allow passage of infectious viral material from cell to cell, are present either as a single polypeptide (e.g., *Capillivirus*, *Vitivirus*) or as a triple gene block (e.g., *Allexivirus*, *Potexvirus*). The lack of a movement protein in BotV-X and BotV-F is a feature consistent with the lack of plasmodesmata or similar restrictions to movement in their fungal host.

The remarkably high sequence similarity between the replicase of BotV-X and that of the plant allexivirus, GarV-A, may indicate a relatively recent divergence from a common ancestor. Since *B. cinerea* has an extremely wide host range, including alliums, a possible scenario is that historically *B. cinerea* acquired a plant virus, such as GarV-A, that was infecting the same host and this virus has subsequently lost unnecessary parts of the genome, such as the movement protein gene.

SsNSRV-1 and FgNSRV-1 are also found in plant pathogenic fungi and are phylogenetically related to mononegaviruses that collectively infect a range of hosts, including plants. Similar to BotV-X and BotV-F, there are differences in genome organization between FgNSRV-1 and SsNSRV-1 and related viruses from other hosts. There are also differences in particle morphology, the virions of SsNSRV-1 and FgNSRV-1, which are filamentous, and the virions of other members of the *Mononegavirales* which are typically spherical, bacilliform or pleomorphic, although the nucleocapsids may take on a filamentous form. As with BotV-X and BotV-F the pathogenic association between the fungal host and plants potentially provides an opportunity for the fungus to acquire a plant virus that becomes modified over time. Although most plant viruses do not have lipoprotein envelopes, some, such as members of the ssRNA(-) families *Bunyaviridae* and *Rhabdoviridae* (order *Mononegavirales*) have lipid based envelopes, which are important for specific interaction with their insect vectors. While there is currently no evidence for insect vectors of mycoviruses the envelopes of SsNSRV-1 and FgNSRV-1 virions would be consistent with these viruses being derived from insect transmitted plant viruses.

Shared lineages between plant and fungal viruses are common with both plant and fungal viruses belonging to at least twelve ICTV-recognized families, with phylogenetic evidence of common ancestry. In addition it has been shown that some fungal viruses are able to replicate in plants cells and vice versa. For example the two mycoviruses *Penicillium aurantiogriseum* totivirus-1 and *Penicillium aurantiogriseum* partitivirus-1 are able to replicate in protoplasts of the plant species *Nicotiana benthamiana* and *Nicotiana tabacum*, while the plant virus, cucumber mosaic virus, is able to form a stable infection in cultures of the fungus *Rhizoctonia solani*. It seems less likely that mycoviruses would be acquired by and form an ongoing infection in plants, because of the lack of movement proteins required for systemic movement in plants. However, the possibility should not be totally discounted as it has been proposed that plant ourmiaviruses, which have a tripartite ssRNA(+) genome, have arisen from a narnavirus (monopartite genome) that acquired additional RNAs that encode a movement protein.

Although the dsRNA virus (CcFV-1) was found in the plant pathogen *Colletotrichum camelliae* it shares a highest amino acid identity of 47% with the RNA-dependent RNA polymerase (RdRp) of AfuTmV-1, a virus from the human fungal pathogen *Aspergillus fumigatus*. However it also shares low identity to the dsRNA-1 of ATV, CcV-1, and BdRV1 from the plant pathogenic fungi *Alternaria tenuissima*, *Cladosporium cladosporioides* and *Botryosphaeria dothidea*, respectively. This raises the possibility that CcFV-1 may be derived from a plant virus, although none of the characterized dsRNA plant viruses have long filamentous particles.

The increasing use of deep sequencing is detecting additional, often unencapsidated, mycoviruses with some sequence identity to known filamentous viruses. For example, members of the family *Deltaflexivirus* (order *Tymovirales*), *Fusarium graminearum* deltaflexivirus 1 (FgDFV1), *Sclerotinia sclerotiorum* deltaflexivirus 1 (SsDFV1), and soybean leaf-associated *mycoflexivirus* 1 (SlaMyfV1), are distantly related to viruses of the order *Tymovirales*, although none of these viruses produce coat proteins. However, in other cases it is unknown whether presumed virus sequences belong to viruses that produce encapsidated filamentous particles and further investigation, including virus purification and electron microscopy is required to determine the nature of the virions.

Further Reading

- Arthur, K., Pearson, M.N., 2014. Geographic distribution and sequence diversity of the mycovirus Botrytis virus F. *Mycological Progress* 13, 1249–1253.
- Boine, B., Kingston, R.L., Pearson, M.N., 2012. Recombinant expression of the coat protein of Botrytis virus X and development of an immunofluorescence detection method to study its intracellular distribution in Botrytis cinerea. *Journal of General Virology* 93, 2502–2511.
- Chen, X., He, H., Yang, X., et al., 2016. The complete genome sequence of a novel *Fusarium graminearum* RNA virus in a new proposed family within the order *Tymovirales*. *Archives of Virology* 161, 2899–2903.
- Howitt, R.L.J., Beever, R.E., Pearson, M.N., Forster, R.L.S., 2001. Genome characterization of Botrytis virus F, a flexuous rod-shaped mycovirus resembling plant 'potex-like' viruses. *Journal of General Virology* 82, 67–78.

- Howitt, R.L.J., Beever, R.E., Pearson, M.N., Forster, R.L.S., 2006. Genome characterization of a flexuous rod-shaped mycovirus, Botrytis virus X, reveals high amino acid identity to genes from plant 'potex-like' viruses. *Archives of Virology* 151, 563–579.
- Jia, H., Dong, K., Zhou, L., *et al.*, 2017. A dsRNA virus with filamentous viral particles. *Nature Communications* 8, 168.
- Liu, L., Xie, J., Cheng, J., *et al.*, 2014. Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12205–12210.
- Marzano, S.-Y.L., Nelson, B.D., Ajayi-Oyetunde, O., *et al.*, 2016. Identification of diverse mycoviruses through metatranscriptomics characterization of the viromes of five major fungal plant pathogens. *Journal of Virology* 90, 6846–6863.
- Mu, F., Xie, J., Cheng, S., *et al.*, 2018. Virome characterization of a collection of *Sclerotinia sclerotiorum* from Australia. *Frontiers in Microbiology* 8, 2540.
- Pearson, M.N., Beever, R.E., Boine, B., Arthur, K., 2009. Mycoviruses of filamentous fungi and their relevance to plant pathology. *Molecular Plant Pathology* 10, 115–128.
- Pearson, M.N., Beever, R.E., Boine, B., Tan, C., 2009. Can mycoviruses be used for the biocontrol of the plant pathogenic fungus *Botrytis cinerea*? In: Elad, Y., Freeman, S. (Eds.), *Biological Control of Fungal and Bacterial Plant Pathogens* 43. IOBC/WPRS Bulletin, pp. 7–10. (ISBN978-92-9067-217-3).
- Roossinck, M.J., 2019. Evolutionary and ecological links between plant and fungal viruses. *New Phytologist (Tansley Review)* 221, 86–92.
- Wang, L., He, H., Wang, S., *et al.*, 2018. Evidence for a novel negative-stranded RNA mycovirus isolated from the plant pathogenic fungus *Fusarium graminearum*. *Virology* 518, 232–240.

Prions of Yeast and Fungi

Reed B Wickner, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, United States

Herman K Edskes, National Institutes of Health, Bethesda, MD, United States

Published by Elsevier Ltd.

This is an update of R.B. Wickner, H. Edskes, T. Nakayashiki *et al.*, Prions of Yeast and Fungi, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00407-6.

Glossary

Nonchromosomal gene A gene that segregates 4 + :0 – in meiosis and can be transferred by cytoplasmic mixing, in contrast to chromosomal genes that segregate 2 + :2 – in meiosis, and are not transferred by cytoplasmic mixing.

Nonsense suppressor tRNA A mutant transfer RNA that recognizes a translational stop codon and inserts an amino acid thus allowing the peptide chain to continue.

Introduction and History

The word prion, meaning “infectious protein” without need for a nucleic acid, was coined to explain the properties of the agent producing the mammalian transmissible spongiform encephalopathies (TSEs). The yeast and fungal prions were identified by their unique genetic properties which were unexpected for any nucleic acid replicon, but specifically predicted for an infectious protein.

[PSI] was described by Brian Cox in 1965 as a nonchromosomal genetic element of *Saccharomyces cerevisiae* that increased the efficiency of a weak nonsense suppressor transfer RNA (tRNA). [URE3] was described by Francois Lacroute as a nonchromosomal gene that relieved nitrogen catabolite repression, allowing expression of genes needed for utilizing poor nitrogen sources even when a good nitrogen source was available. The [Het-s] prion was described in 1952 by Rizet as a nonchromosomal gene needed for heterokaryon incompatibility in *Podospora anserina*. In 1994, we showed that [URE3] and [PSI +] were prions based on their genetic properties (see below). The [PIN] prion of yeast was discovered in 1997 by Derkatch and Liebman in their studies of de novo generation of the [PSI] prion. There are now an array of prions known in yeast ([Table 1](#)).

Genetic Signature of a Prion

Viruses of yeast and fungi generally do not exit one cell and enter another, but spread by cell–cell fusion, as in mating or heterokaryon formation. Infectious proteins (prions) should likewise be nonchromosomal genetic elements. To distinguish prions from replicating nucleic acids, three genetic criteria were proposed ([Fig. 1](#)): (1) If a prion can be cured, it can reappear in the cured strain at some low frequency. (2) Overproduction of the protein capable of being a prion should increase the frequency of the prion arising de novo. (3) If the prion produces a phenotype by the simple inactivation of the protein, then this phenotype should resemble the phenotype of mutation of the gene encoding the protein, which gene must be needed for prion propagation.

All three criteria were satisfied by [PSI] and [URE3], strongly indicating, perhaps proving, that they were prions. The [Het-s] prion of *P. anserina* and the [PIN] prion of *Saccharomyces cerevisiae* were likewise proved by application of the same genetic criteria, but because their prion form produces the phenotype, rather than the absence of the normal form, the prion protein gene is needed for propagation of the prion, but the phenotype of the prion is not the same as mutants in the prion protein gene.

Self-Propagating Amyloid as the Basis for Most Yeast Prions

The finding that Sup35p, Ure2p, and HET-s were protease resistant and aggregated in prion-containing cells, and that these proteins (and particularly their prion domains ([Fig. 2](#))) would form amyloid in vitro indicated that a self-propagating amyloid ([Fig. 3](#)) was the basis of these prions. This was confirmed by the finding that the corresponding prions were transmitted by introduction of amyloid formed in vitro from the recombinant proteins. In some cases it was shown that the soluble form or nonspecific aggregates of the protein were ineffective. This argues that the respective amyloids are not by-products or a dead-end stage of these prions, but are themselves the infectious material. All infectious Ure2p amyloids are larger than about 40-mer size. Amyloid formed in vitro is capable, for at least [PSI] and [URE3], of transmitting any of several prion variants. This implies that the amyloids can have any of several structures. Strong variants of [PSI +] (those

Table 1 Prions of *Saccharomyces cerevisiae* and *Podospora anserina*

Prion	Prion protein	Prion phenotype	Normal protein function
[URE3]	Ure2p	Derepressed genes for utilizing poor nitrogen sources in presence of a good nitrogen source; slow growth	Repression of genes for utilizing poor nitrogen sources in presence of a good nitrogen source
[PSI ⁺]	Sup35p	Readthrough of termination codons; slow growth; death (depends on prion variant)	Translation termination
[PIN ⁺]	Rnq1p	Rare generation (by cross-seeding) of [PSI ⁺] or [URE3]	none known
[OCT ⁺]	Cyc8p	Slow growth; impaired mating and sporulation	transcription repressor subunit
[SWI ⁺]	Swi1p	Poor growth on raffinose, galactose or glycerol	chromatin remodeling subunit
[MOT ⁺]	Mot3p	Inappropriate derepression of anaerobic genes; colony polymorphisms	transcription regulator
[MOD ⁺]	Mod5p	Partial azole-resistance; slow growth	tRNA isopentenyltransferase
[BETA]	Prb1p	Active protease B (non-amyloid prion)	Active protease B (this is a functional prion)
[Het-s]	HET-s	Heterokaryon incompatibility	Heterokaryon incompatibility (this is a functional prion)

Note: All but [BETA] are amyloid-based prions. [Het-s] is a prion of the filamentous fungus *Podospora anserina*, and the others are prions of *S. cerevisiae*.

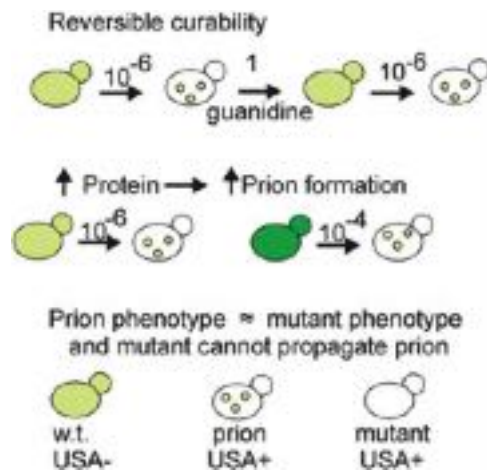


Fig. 1 Genetic signature of a prion. Among nonchromosomal genetic elements, those with the three properties shown here must not be nucleic acid replicons and are almost certainly prions. However, only prions for which the prion form of the protein is inactive (such as [URE3] or [PSI]) will have the third property. The [Het-s] and [PIN⁺] prions are active forms of the HET-s protein and Rnq1 protein, respectively and have the first and second properties but not the third.

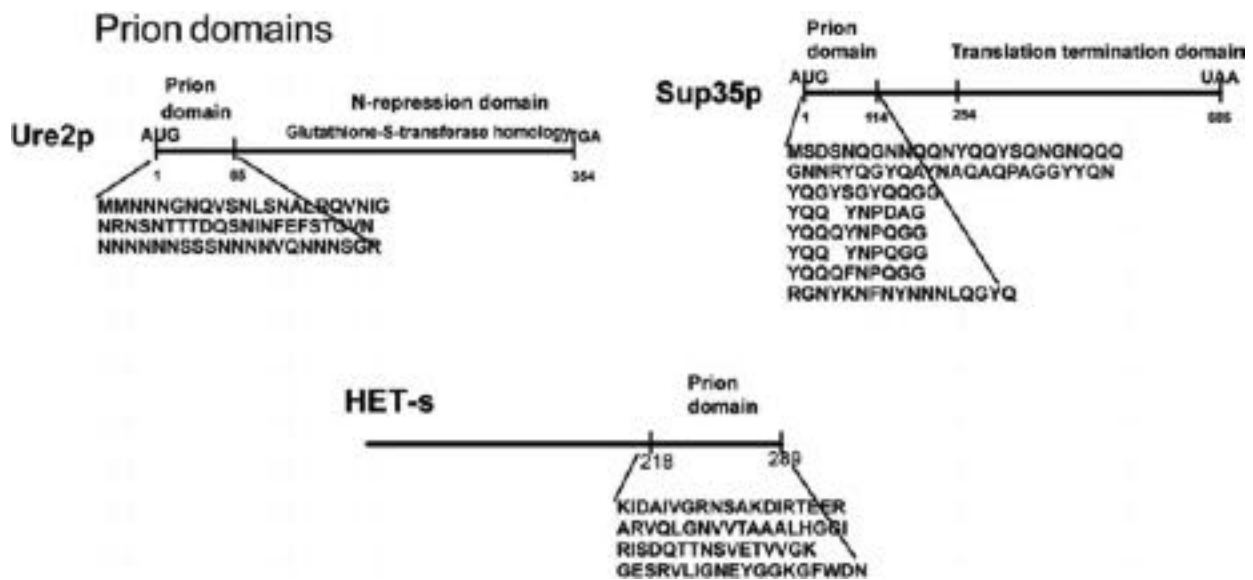


Fig. 2 Prion domains. The prion domains of Ure2p, Sup35p, and HET-s are largely unstructured in the native form and in β -sheet in the prion form.

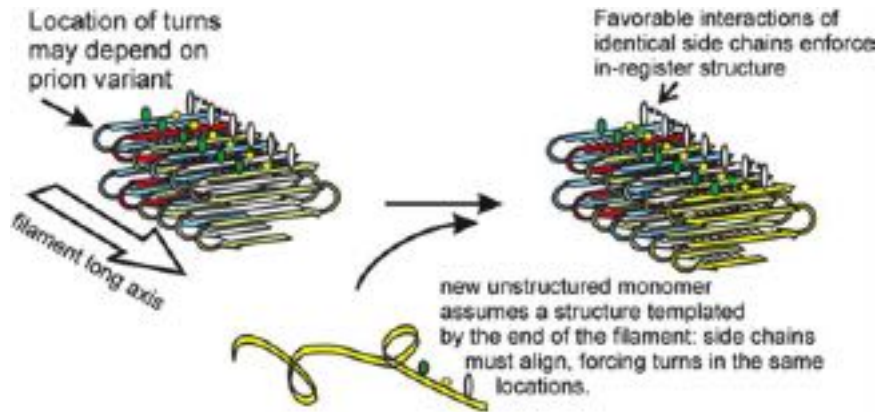


Fig. 3 Infectious amyloids of yeast prion proteins have an in-register parallel folded β -sheet architecture. It is suggested that the amyloids of different prion variants differ in the locations of the folds of the sheets (i.e., the turns of the strands). The in-register architecture is maintained by the favorable interactions among identical side-chains. The same interactions force the new monomer joining the end of the filaments to align in-register and thus have the turns in their strands at the same location as molecules already in the filament. This constitutes conformational templating.

with a more pronounced phenotype) are based on shorter, more fragile amyloid filaments with a shorter β -sheet region than are weak [PSI⁺] variants.

Shuffling Prion Domains and Amyloid Structure

The prion domains (Fig. 2) of Ure2p and Sup35p are quite rich in Asn and Gln residues, and nearly the entire sequence of Rnq1p, the basis of the [PIN] prion, is rich in these amino acids. However, many Q/N-rich proteins are not capable of being prions. Thus, it was assumed that specific sequences in the known prion domains were important for prion formation. The Sup35 prion domain has octapeptide repeats much like those in PrP (the mammalian prion protein), and deletion or duplication of these showed substantial effects on prion generation. In addition, single amino acid changes in the prion domain of Sup35p blocked prion propagation.

To critically test whether the Ure2p prion domain had sequences essential for prion development, the entire Q/N-rich region (residues 1–89) was randomly shuffled (without changing the amino acid content) and each of five shuffled sequences were inserted into the chromosome in place of the normal prion domain. Surprisingly, each of these five shuffled sequences could support prion generation and propagation, although one was rather unstable. Each protein with the shuffled sequence could also form amyloid *in vitro*. This showed that it was the amino acid content of the Ure2p prion domain that determines prion formation, and that sequence plays only a minor role.

Similarly, five shuffled versions of the Sup35p prion domain were each inserted in place of the normal sequence. Again, all five shuffled versions allowed formation and propagation of a [PSI⁺]-like prion. It is likely that the effects of deletion or duplication of the octapeptide repeats observed on prion formation or propagation were due to changes in the length or composition of the prion domain. It appears that repeats *per se* are not important for prion generation or propagation.

Shuffleable Prion Domains Suggests Parallel In-Register β -Sheet Structure

Amyloids are β -sheet structures, but there are at least three kinds of β -sheets. Antiparallel β -sheets have adjacent peptide chains oriented in opposite directions: N->C next to C->N. This results in pairing of largely nonidentical residues. A β -helix also involves pairing of largely nonidentical residues, although they are within the same peptide chain. A parallel β -sheet pairs identical residues if it is in-register, but in principle, one could have an out-of-register parallel β -sheet, in which case nonidentical residues would be bonded to each other.

Prion propagation (and amyloid propagation in general) is a very sequence-specific process. For example, a single amino acid change (at residue 138) can block propagation of scrapie in tissue culture. Humans are polymorphic at PrP residue 129 with roughly equal numbers of alleles encoding M and V. Either M/M or V/V individuals can get Creutzfeldt–Jakob disease (a human prion disease), but M/V heterozygotes rarely do. Similarly, a single amino acid change in the prion domain of Sup35p can block propagation of [PSI⁺] from the normal sequence, but the mutant Sup35p can itself become a prion nonetheless. Thus, if a prion amyloid has an antiparallel, parallel out-of-register, or β -helix structure, there must be some form of complementarity between bonded residues. Shuffling such a sequence would be expected to destroy the complementarity. In contrast, shuffling the sequence of a parallel in-register β -sheet would still leave identical residues paired. This suggested that prion domains that can be shuffled without destroying their prion-forming ability are forming parallel in-register β -sheets.

Infectious Prion Amyloids Have In-Register Parallel Folded β -sheet Architecture

Solid-state NMR experiments, combined with mass per filament length determinations, have shown that the prion domains of Sup35p, Ure2p and Rnq1p are each folded in-register parallel β -sheet structures, like the architecture of such pathologic amyloids of human A β , amylin/IAPP, Tau, and alpha-synuclein. In contrast, the functional prion [Het-s] is composed of amyloid of the HET-s protein having a β -helix structure (Fig. 3).

Prion Variants and the Species Barrier

A single protein or peptide can be the basis of a wide variety of prion “strains” or “variants”, based on differing conformations of the protein in the amyloid, each of which is self-propagating. The manifestation of prion variant differences may be strength of the prion phenotype, stability of prion propagation, sensitivity to the overproduction or deficiency of various chaperones or non-chaperone proteins, or the sensitivity to the various endogenous anti-prion systems (see below). Although prions may be transmissible between species (in mammals or yeast), the efficiency of transmission is generally lower between species than within a species, an effect due to sequence differences between the prion proteins of the donor and recipient species. Barriers are also seen within a single species (“intraspecies barriers”) as a result of sequence polymorphisms, probably selected in the wild to reduce the probability of being infected with a pathologic prion. The ease with which any of these transmission barriers are overcome is sharply dependent on the prion variant, with some completely unable to be transmitted and others transmitted without apparent barrier. It is proposed that these barriers reflect the range of conformers that a given molecule can assume. If the donor conformer is such that the recipient can assume that structure, then there will be little barrier, and the contrary.

The genetics of prion variants makes it clear that some variants are “dominant” to others when introduced together, while other variants are compatible and segregate relative to each other during mitotic growth. While a variant may appear to be stable with respect to one property, it may be unstable with respect to another. Inaccurate templating may cause some prion “mutations”, while severely detrimental effects of a [URE3] or [PSI +] prion on the cells may lead to selection during growth of very rare milder variants or loss of the prion.

Prion Variant Information Templating Mechanism

The folded in-register parallel β -sheet architecture of the infectious yeast prion amyloids determines the structure except for the extent of the β -sheets and the location of the folds in the sheet. The structure is kept in-register by favorable interactions among identical amino acid side chains that can only happen if register is maintained (Fig. 3). This naturally suggests a mechanism by which protein conformation can be templated. In order that their amino acid side chains have these favorable interactions, the unstructured prion domains of the monomers about to join the end of the filament must have their turns at the same locations as those of the molecules already in the filament. This directs the conformation of the new molecule to be essentially that of the rest of the molecules in the filament. Each prion variant is suggested to have the folds of the sheet (=turns of the molecule) in places specific for that variant. This is the only mechanism that has yet been proposed for prion variant propagation/conformational templating.

Chaperones and Prions

Chaperones of the Hsp40, Hsp70, and Hsp104 groups, as well as Hsp90 and its co-chaperones, have been found to be clearly involved in prion propagation (Table 2). Millimolar concentrations of guanidine cure each of the amyloid-based prions by specific inhibition of Hsp104. At least one function of these chaperones is to break large amyloid filaments into smaller ones which can then be distributed at cell division. Overexpressed Hsp104 has a second activity, in curing the [PSI +] prion, and less efficiently, [URE3]. The mechanism of curing is controversial but may involve asymmetric segregation of prion filaments, possibly related to Hsp104's role in collecting damaged proteins and retaining them in the mother cell. This second Hsp104 activity is lost in mutants in (or deletion of) the N-terminal domain (e.g., T160M), but such mutants retain their prion propagation promoting activity and their ability to protect cells from heat shock. Other proteins affecting prion propagation include the nucleotide exchange factors for Hsp70s, namely Fes1p and Sse1p, the cytoskeleton assembly factor Sla1p, and the co-chaperone Sgt2p.

Prion Generation, and [PIN]: A Prion That Gives Rise to Prions

One of the lines of evidence that showed [PSI] was a prion of Sup35p was that overproduction of Sup35p increased the frequency with which [PSI] arises de novo. However, it was found that in some strains, overproduction of Sup35p did not yield detectable emergence of [PSI]-carrying clones. Another nonchromosomal genetic element, named [PIN] for [PSI]-inducibility, was found necessary. [PIN] is a self-propagating amyloid form of the Rnq1 (rich in Asn (N) and Gln (Q)) protein, and it promotes de novo generation of [URE3] as well as [PSI].

Table 2 Interactions of prions with chaperones and other proteins

Mechanism	Components	Yeast prions affected
Filament breakage: removal of a single prion protein molecule from middle of the filament generates new filaments	Hsp104, Hsp70 (Ssa1–4), Hsp40 (Sis1)	All amyloid-based prions
Filament collection ~ Asymmetric segregation	Hsp104 overproduction curing [with Hsp90s and Hsp90 co-chaperones]	[PSI ⁺], ([URE3])
?Blocking filament ends	Btn2 [with Hsp42]	[URE3], (some [PSI ⁺])
?Competition for monomers	Upf1,2,3	[PSI ⁺]
?Occupation of unstructured (prion) domains	Upf1,2,3	[PSI ⁺]
	??	??any amyloid-based prion protein
Solubilizing monomers: a consequence of the “Filament breakage” mechanism above	Hsp104, Hsp70 (Ssa1–4), Hsp40 (Sis1)	All amyloid-based prions
?Blockage of access to filaments for breakage	Excess Hsp104	[PSI ⁺]
?Sequestration of Sis1p	Cur1p overproduction	[URE3]

Biological Roles of Prions: A Help or a Hindrance?

In an attempt to discern whether yeast prions are an advantage or disadvantage to their host organism, cell growth of isogenic [PSI⁺] and [psi⁻] strains have been carried out under a variety of conditions. To what extent the various growth conditions tested represent the normal yeast habitat seems unknowable, although [psi⁻] was an advantage under far more conditions than was [PSI⁺].

An alternative approach was to compare the frequency with which [PSI⁺] or [URE3] was found in wild strains to those of several “selfish” RNA and DNA viruses and replicons known in *S. cerevisiae*. In any organism, an infectious element (such as a virus) may be widely distributed in spite of it causing disease in its host because the infection process overcomes and outraces the loss of infected individuals from the negative effects of the infecting element. Certainly an advantageous infectious element will quickly become widespread as selection and infection operate in the same direction. In fact, while the mildly deleterious RNA and DNA viruses and plasmids of yeast are easily found in wild strains neither [URE3] nor [PSI⁺] was found in any of the 70 wild strains examined. This indicates that the mildest variants of [URE3] and [PSI⁺] produce disease in their hosts, and a rather more severe disease than the mild nucleic acid replicons.

However, more detrimental variants of each prion, either dramatically slowing growth or even killing the host, are more common than the mild variants usually used for laboratory studies. Moreover, the recent discovery of an array of cellular anti-prion systems (see below) indicates that the cells themselves do not regard these prions as beneficial.

The [Het-s] prion of *Podospora* appears to carry out the normal fungal function of heterokaryon incompatibility, thought to be a protection against the sometimes debilitating fungal viruses. Indeed, as one would expect for a prion with a function for the cell, over 90% of wild *Podospora* isolates carry [Het-s], confirming its beneficial effects.

Unlike [PSI⁺] and [URE3], [PIN⁺] is found in wild strains at a frequency similar to that of the parasitic RNA and DNA viruses and plasmids. This suggests that [PIN⁺] is at least not as severe a pathogen as are [URE3] and [PSI⁺].

Anti-Prion Systems

The pathological character of yeast prions suggested that there may be cellular systems that cure prions as they arise, much as DNA repair systems eliminate most mutations as they arise. Indeed, it was found that over 90% of [URE3] variants arising (at increased frequency) in a *btn2 cur1* strain are cured by replacing just the normal amount of these proteins. Btn2p cures [URE3] by collecting Ure2p amyloid filaments at one locus in the cell, increasing the likelihood that one of the daughter cells will lack filaments and thus be cured. Likewise, the frequency of [PSI⁺] arising spontaneously is elevated over 10-fold in a mutant (T160M) that lacks the Hsp104-overproduction-curing activity toward [PSI⁺], and most of those variants are cured by just replacing the normal Hsp104 without overproduction. Mutants in the *UPF* genes, responsible for nonsense-mediated decay, show > 10-fold elevated frequency of [PSI⁺] generation, and, again, most of the variants arising are cured by restoring normal levels of expression of these proteins. Sub3p forms a complex with Upf1p, Upf2p and Upf3p, and it is suggested that this complex either competes with the filaments for free monomers, or blocks the ends of filaments preventing their growth.

The ribosome-associated Hsp70 family chaperones Ssb1p and Ssb2p keep down the frequency of [PSI⁺] arising de novo, but in this case restoring the normal amounts of these proteins does not cure the variants arising. It is inferred that the Ssbs, known to be involved in promoting proper folding of nascent proteins, prevent the generation of [PSI⁺].

The Hsp40 family member Sis1p is essential for propagation of [PSI⁺] and [URE3], but also protects cells from the potential lethality of [PSI⁺] without curing the prion. Similarly, the Lug1/Ylr352w protein prevents the near-lethality of [URE3], apparently via a novel function of Ure2p (not nitrogen regulation) that is essential in the absence of Lug1/Ylr352w.

Inositol Polyphosphates and Prion Propagation

In a screen for anti-prion systems, it was found that nearly all variants of the [PSI⁺] prion require one of several inositol poly/pyrophosphates for their propagation. Inositol hexakisphosphate (IP₆), 5-diphospho-IP₅ and 5-diphospho-IP₄ are each able to support [PSI⁺] propagation. In the absence of the 5-diphospho-IPs, the 1-diphospho-IP₅ has an inhibitory effect on [PSI⁺] propagation. Siw14p, a 5-pyrophosphatase active on 5-diphospho IPs, prevents propagation of about half of [PSI⁺] variants because they need elevated levels of that compound. The mediators of these effects are, as yet, unknown.

Enzyme as Prion

While most of the known prions involve amyloids, the word prion (infectious protein) is more general, requiring only that transmission be by protein alone. If an enzyme is made as an inactive precursor that needs the active form of the same enzyme for its activation, then such a protein can be a prion. The vacuolar protease B (Prb1p) of *S. cerevisiae* can be such a prion in a mutant lacking protease A, which normally activates its precursor. Cells initially carrying only the inactive precursor remain so unless the active enzyme is introduced. Once a cell has some active enzyme, the autoactivation process can continue indefinitely. It is likely that other examples of this type of phenomenon will be found among the many protein kinases, methylases, acetylases, and other protein-modifying enzymes that are known.

Conclusions

The advent of yeast prions has propelled the prion field forward, giving the first proof that an infectious protein can exist, the first structure/architecture of a prion amyloid, and a wide array of information on cellular components favoring and antagonizing prion propagation. With the findings that the common human amyloidosis are actually prions, the continuing insights into the general prion phenomena that are being provided by the yeast and fungal work will continue to push this field forward.

Acknowledgment

This research was supported [in part] by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

See also: Prions of Vertebrates

Further Reading

- Aigle, M., Lacroute, F., 1975. Genetical aspects of [URE3], a non-Mendelian, cytoplasmically inherited mutation in yeast. *Molecular and General Genetics* 136, 327–335.
- Cascarina, S.M., Paul, K.R., Machihara, S., Ross, E.D., 2018. Sequence features governing aggregation or degradation of prion-like proteins. *PLOS Genetics* 14, e1007517. doi:10.1371/journal.pgen.1007517.
- Chernoff, Y.O., Lindquist, S.L., Ono, B.-I., Inge-Vechtomov, S.G., Liebman, S.W., 1995. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [psi⁺]. *Science* 268, 880–884.
- Coustou, V., Deleu, C., Saupe, S., Begueret, J., 1997. The protein product of the het-s heterokaryon incompatibility gene of the fungus *Podospora anserina* behaves as a prion analog. *Proceedings of the National Academy of Sciences of the United States of America* 94, 9773–9778.
- Cox, B.S., 1965. PSI, A cytoplasmic suppressor of super-suppressor in yeast. *Heredity* 20, 505–521.
- Derkatch, I.L., Bradley, M.E., Hong, J.Y., Liebman, S.W., 2001. Prions affect the appearance of other prions: The story of [PIN]. *Cell* 106, 171–182.
- Jung, G., Jones, G., Masison, D.C., 2002. Amino acid residue 184 of yeast Hsp104 chaperone is critical for prion-curing by guanidine, prion propagation, and thermotolerance. *Proceedings of the National Academy of Sciences of the United States of America* 99, 9936–9941.
- King, C.Y., Diaz-Avalos, R., 2004. Protein-only transmission of three yeast prion strains. *Nature* 428, 319–323.
- Liebman, S.W., Chernoff, Y.O., 2012. Prions in yeast. *Genetics* 191, 1041–1072.
- Maddelain, M.L., Dos Reis, S., Duvezin-Caubet, S., Couлары-Salin, B., Saupe, S.J., 2002. Amyloid aggregates of the HET-s prion protein are infectious. *Proceedings of the National Academy of Sciences of the United States of America* 99, 7402–7407.
- Ness, F., Cox, B., Wonwigkam, J., Naeimi, W.R., Tuite, M.F., 2017. Over-expression of the molecular chaperone Hsp104 in *Saccharomyces cerevisiae* results in the malpartition of [PSI⁺] propagons. *Molecular Microbiology* 104, 125–143.
- Paushkin, S.V., Kushnir, V.V., Smirnov, V.N., Ter-Avanesyan, M.D., 1997. In vitro propagation of the prion-like state of yeast Sup35 protein. *Science* 277, 381–383.
- Reidy, M., Sharma, R., Shastry, S., et al., 2014. Hsp40s specify functions of Hsp104 and Hsp90 protein chaperone machines. *PLOS Genetics* 10, e1004720.
- Tanaka, M., Chien, P., Naber, N., Cooke, R., Weissman, J.S., 2004. Conformational variations in an infectious protein determine prion strain differences. *Nature* 428, 323–328.
- Tycko, R., Wickner, R.B., 2013. Molecular structures of amyloid and prion fibrils: Consensus vs. controversy. *Accounts of Chemical Research* 46, 1487–1496.
- Wickner, R.B., 1994. [URE3] as an altered URE2 protein: Evidence for a prion analog in *S. cerevisiae*. *Science* 264, 566–569.
- Wickner, R.B., Bezsonov, E.E., Son, M., et al., 2018. Anti-prion systems in yeast and inositol polyphosphates. *Biochemistry* 57, 1285–1292. doi:10.1021/acs.biochem.1027b01285.
- Wickner, R.B., Shewmaker, F., Bateman, D.A., et al., 2015. Yeast prions: Structure, biology and prion-handling systems. *Microbiology and Molecular Biology Reviews* 79, 1–17.

Single-Stranded DNA Mycoviruses

Daohong Jiang, Huazhong Agricultural University, Wuhan, China

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Biological control A method to prevent pests using beneficial organisms, here specifically means using mycovirus to control fungal diseases.

DNA mycovirus, also called fungal DNA virus A DNA virus that replicates in fungi.

Gemonoviridae A group of small circular single-stranded DNA viruses that are phylogenetically related to plant geminiviruses, but lack move protein genes.

Hypovirulence Reduced virulence of a fungal pathogen caused by the infection of mycovirus; Hypovirulence and hypovirus of chestnut blight/*Cryphonectria parasitica*

system is a classic example for studying mycovirus and biological control of plant fungal diseases.

Sclerotinia sclerotiorum An worldwide spread ascomycetous fungal pathogen that can attack more than 400 species and subspecies of plants. Important disease caused by it includes stem rot on rapeseed (*Brassica napus*).
ssDNA mycovirus single-stranded DNA mycovirus.

Vegetative incompatibility A non-self-recognition system that is determined by *vic* genes in fungi, and is a mechanism for fungi to prevent the infection by molecular parasites, such as mycovirus. It is a key hindrance to successful biological of fungal diseases using mycovirus.

Introduction

DNA viruses are viruses that have DNA genomes and replicate using DNA-dependent DNA polymerase. DNA viruses can be grouped into two classes, double-stranded (ds) DNA viruses and single-stranded (ss) DNA viruses. DNA viruses are very common in both prokaryotic microorganisms and eucaryotic organisms including humans, animals, and plants. The most feared DNA viruses is variola virus which causes smallpox. Three families of DNA viruses have been established in plants, namely *Caulimoviridae*, *Geminiviridae*, and *Nanoviridae*, whereas DNA virus has been very rarely reported in fungi. Previously, a DNA virus named *Rhizidiomyces virus* was discovered in filamentous microorganism *Rhizidiomyces* sp., which used to be grouped in the kingdom Fungi, however, *Rhizidiomyces* is actually a fungi-like water mold that belongs to the kingdom Chromista. In the 1980s, a DNA virus with large virions has been observed in a strain of *Rhizoctonia solani* in China, but this study has not been followed. The presence or absence of DNA virus in fungi was unknown for a considerable period of time. In 2010, a small molecule circular single-stranded DNA virus was discovered and reported in a cosmopolitan necrotrophic pathogenic fungus *Sclerotinia sclerotiorum*. The discovery of DNA mycoviruses has significantly enriched the understanding of mycoviruses, broadened the research field of mycoviruses, and provided new research ideas and methods for the use of mycoviruses to control fungal diseases.

The Host of the DNA Mycovirus

Sclerotinia sclerotiorum is an ascomycetous plant pathogenic fungus that was firstly described in 1837 by Libert, M.A. and named *Peziza sclerotiorum*; the name was changed as *Sclerotinia libertiana* by Fuckel, L. in 1870, finally changed as the current name by de Bary, A. in 1884. *S. sclerotiorum* is a typical necrotrophic plant pathogen; it attacks more than 400 species and subspecies of plants, most of which are dicotyledons, including many economically important crops, such as cruciferous crops, leguminous crops, and compositae vegetable crops. The diseases caused by *S. sclerotiorum* are often called white mold on many vegetable crops or stem rot on rapeseed (*Brassica napus*). *S. sclerotiorum* infects plants via either ascospores or infectious hyphae, produces sclerotia on plant residues at the late stage of infection and uses them to undergo dormancy in the field. When contacting hosts, sclerotia germinate carpogenically to release ascospores into the air and infect aerial parts of plants, or germinate myceliogenically to produce infectious hyphae to infect basal stem and old leaves that contact the ground. The sclerotia, apothecia, ascuses, and ascospores, and the induced symptom on rapeseed are presented in Fig. 1. The stem rot is the major disease on rapeseed worldwide, and no resistant cultivar is available. The control of stem rot is mainly dependent on fungicides which usually is not efficient. Hence, a number of field-collected isolates of *S. sclerotiorum* were screened for hypovirulence-associated mycoviruses to control stem rot, and *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 (SsHADV-1) was discovered.

Discovery of DNA Mycoviruses

Sclerotia collected from rapeseed fields in China were used to screen hypovirulence-associated mycoviruses and one abnormal strain DT-8 was isolated from a sclerotium sampled at Datong Lake District, Yiyang City, Hunan Province, China. Strain DT-8 grows slowly on potato dextrose agar (PDA) medium and forms an abnormal colony morphology rich in aerial hyphae on PDA

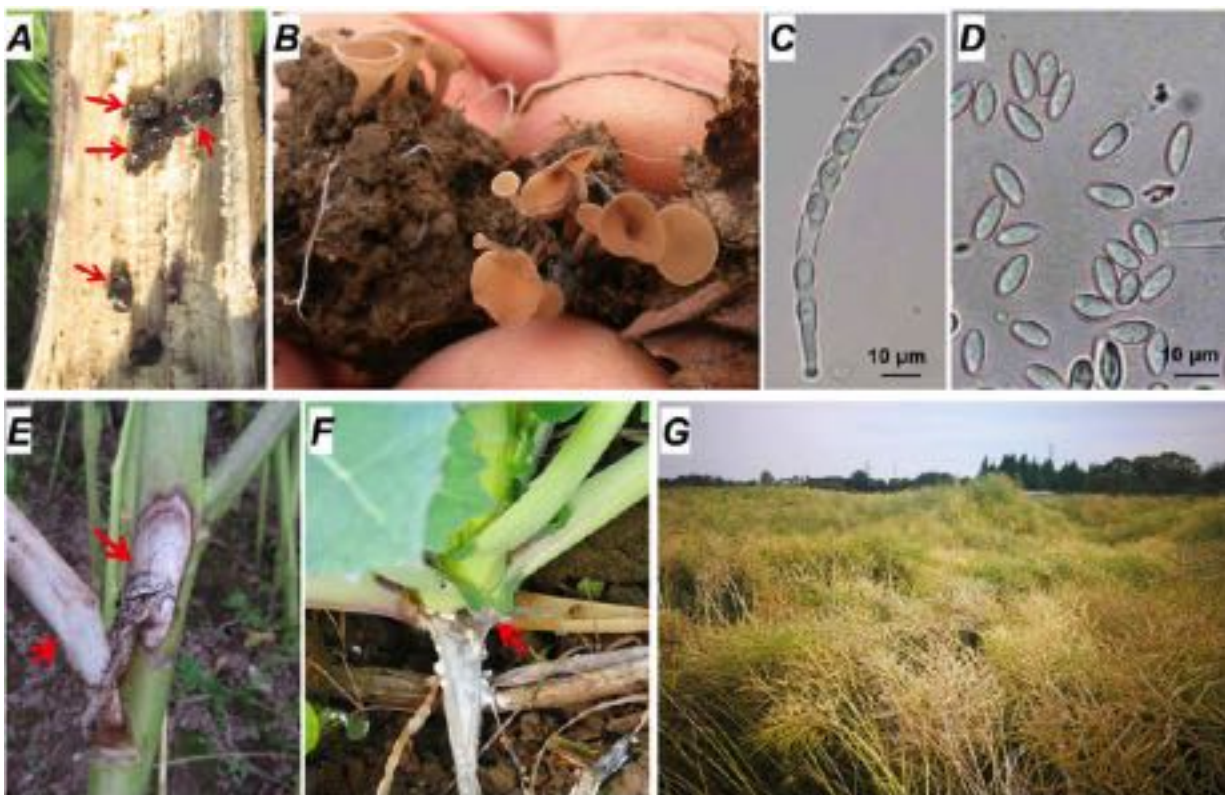


Fig. 1 *Sclerotinia sclerotiorum* and the induced stem rot of rapeseed (*Brassica napus*). A, Sclerotia (arrows) produced on residue of dead rapeseed; B, Apothecia, germinated from sclerotia; C–D, asco and ascospores; E, Symptom on stem (arrows); F, Symptom on stem basal (arrows); G, Premature and lodging, heavy infected rapeseed field, only green plants were not killed.

plate (Fig. 2(A) and (D)). The sclerotia produced by strain DT-8 are usually smaller than these produced by normal strains of *S. sclerotiorum* and are distributed on colony irregularly (Fig. 2(C)). Only very few sclerotia could produce apothecia and ascospores, while ascospore-derived offsprings are fully recovered with no different from the normal strains; occasionally, subcultures derived from hyphal tip of strain DT-8, such as strain DT-8VF, also shows normal phenotype of *S. sclerotiorum*. Strain DT-8 almost loses virulence both on rapeseed and *Arabidopsis thaliana* (Fig. 2(B)). When extracting DNA from strain DT-8 using CTAB method, besides genomic DNA band, two or more additional small DNA bands could be observed under UV light when separated by agarose gel electrophoresis (Fig. 3(A)), while cultures cured did not have these DNA bands, suggesting that these additional DNA were possibly associated with the abnormal characteristics of strain DT-8. The additional DNA was recovered from agarose gel for sequencing analysis, and results showed that the DNA was actually a genome of a small single-stranded circular DNA virus. Isometric particles with a diameter of 22 nm (Fig. 3(C)) were successfully extracted and purified from mycelia of strain DT-8 by ultra-centrifugation, and viral DNA could be successfully extracted from virions (Fig. 3(E)). These results directly confirmed that strain DT-8 was infected by a DNA virus. By dual culture, hypovirulence and the associated traits were co-transmitted with DNA through hyphal contact (hyphal anastomosis). Furthermore purified virions could successfully transfect the protoplast of normal strain of *S. sclerotiorum*, and confer hypovirulence to its host. Hence, the DNA virus is associated with hypovirulence and named as *Sclerotinia sclerotiorum* hypovirulence-associated DNA 1 (SsHADV-1).

The Genome and Proteins of SsHADV-1

The genome of SsHADV-1 is a single-stranded circular DNA with a full length of 2166 nt (Fig. 3(B) and (F)). The genome has only two large open reading frames (ORF) encoding a coat protein (CP) with 312 amino acid residues (aa) and a putative replication initiation protein (Rep) with 324 aa. The CP gene is located on the sense strand of the DNA genome, while the Rep gene is located on the antisense strand of the DNA genome (Fig. 3(F)). The two genes are separated by two non-coding DNA, the large one called the large intergenic region (LIR) which has 133 nt, and the small one called small intergenic region (SIR) with a length of 119 nt. LIR has the nonanucleotide TAATATT ↓ AT at the apex of a stem-loop structure, which is a conserve recognition site (↓) circular ssDNA viruses for Rep to initiate viral DNA replication. Furthermore, there is a bidirectional promoter in the LIR to transcribe CP gene and Rep gene. The coat protein was confirmed by sequencing analysis of protein samples extracted from purified virions,

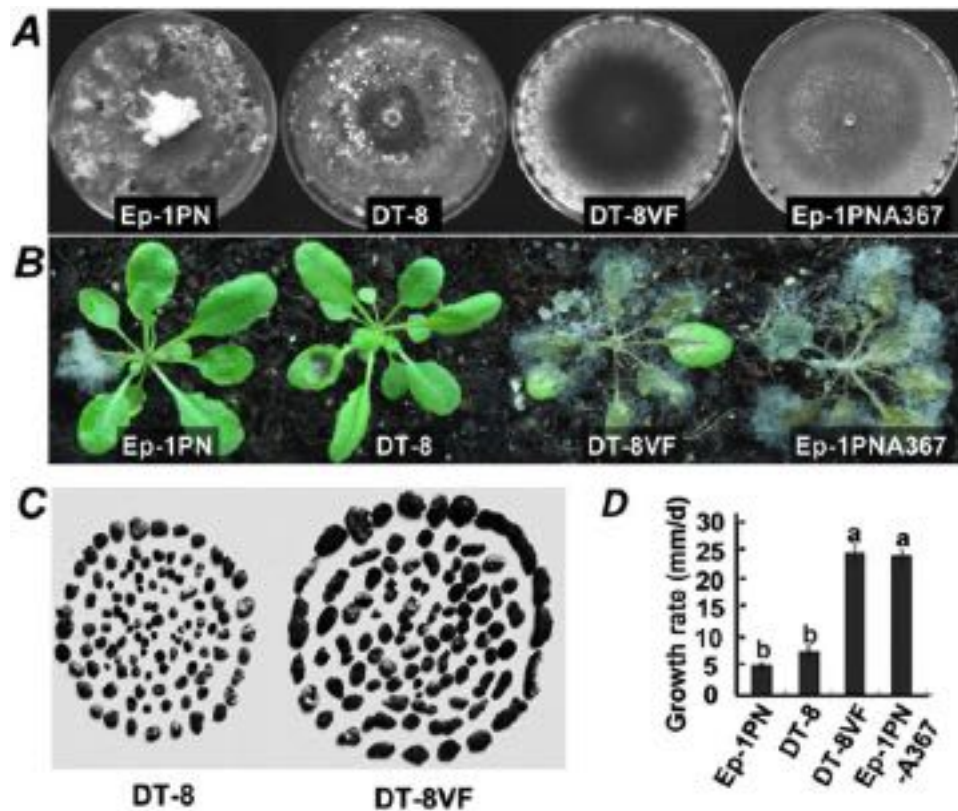


Fig. 2 Abnormal phenotype of strain DT-8 of *Sclerotinia sclerotiorum*. A, Abnormal colony morphology, developed on PDA medium plate at 20°C for 15 days; B, Hypovirulence, could not induce typical lesion on leaf of *Arabidopsis thaliana*, inoculated plants were maintained at 20°C for 4 days post inoculation; C, Small sclerotia produced on a PDA medium (20°C for 30 days); D, Slow growth, bars represent standard deviation from the mean ($n = 8$), the small letters on top of the bars indicate whether the differences are statistically significant ($P < 0.05$). The hypovirulent strain Ep-1PN which is infected by two RNA viruses, its virus-free sexual progeny Ep-1PNA367 and virus-free isogenic strain DT-8VF derived by hyphal tipping culture of strain DT-8 were used as controls. Reproduced from Yu, X., Li, B., Jiang, D., *et al.*, 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences of the United States of America* 107, 8387–8392.

while Rep protein was confirmed by alignment analysis with related viruses. Rep of SsHADV-1 shows high similarity to the Rep proteins of geminiviruses that infect numerous plants. Two conserved domains in the Rep proteins of geminiviruses, namely Rep catalytic domain (Gemini_AL1) (pfam00799) and Rep central domain (Gemini_AL1_M) (pfam08283) are also identified on the Rep of SsHADV-1. Seven conserved motifs which are shared with viruses in the genus *Mastrevirus* in the family *Geminiviridae* are also identified, namely IRD, RCR-I, RCR-II, RCR-III, RBR, NTP Binding-A, and NTP binding-B (Fig. 4(B)).

Host Range of SsHADV-1

SsHADV-1 can infect strains of *S. sclerotiorum* by hyphal anamostosis or just by contacting with colonies of virus-infected strains and non-infected strains. A transfection method has shown SsHADV-1 to be introduced and replicate in species in the genus *Sclerotinia*, such as *S. minor* and *S. nivalis*. However, SsHADV-1 could not replicate in *Botrytis cinerea* which belong to the same family as *Sclerotinia* sp. SsHADV-1 also could neither replicate in *Coniothyrium minitans*, a mycoparasite of *Sclerotinia* spp., *Magnaporthe oryzae* nor *Saccharomyces cerevisiae* which are phylogenetically distant to its natural hosts. As far as we know, there is no any report that SsHADV-1 could infect other fungi in nature. Hence, SsHADV-1 has a narrow fungal host range.

Extracellular Entry of SsHADV-1

Many mycoviruses have no significant effects on their host, and often lack extracellular phase, hence are also considered to be dispensable genetic elements in fungi. Virions of mycoviruses generally are not able to infect their host directly. Unlike most known mycoviruses, SsHADV-1 can infect *S. sclerotiorum* through intact hyphae. When colonies of virus-free strain of *S. sclerotiorum* contacts SsHADV-1 virions which are spread on a PDA plate, the virus enters into fungal cells and successfully infect the colony and convert hypovirulence to the virus-free strain (Fig. 5(A)–(C)). The virions also can infect *S. sclerotiorum* on plants, when sprayed

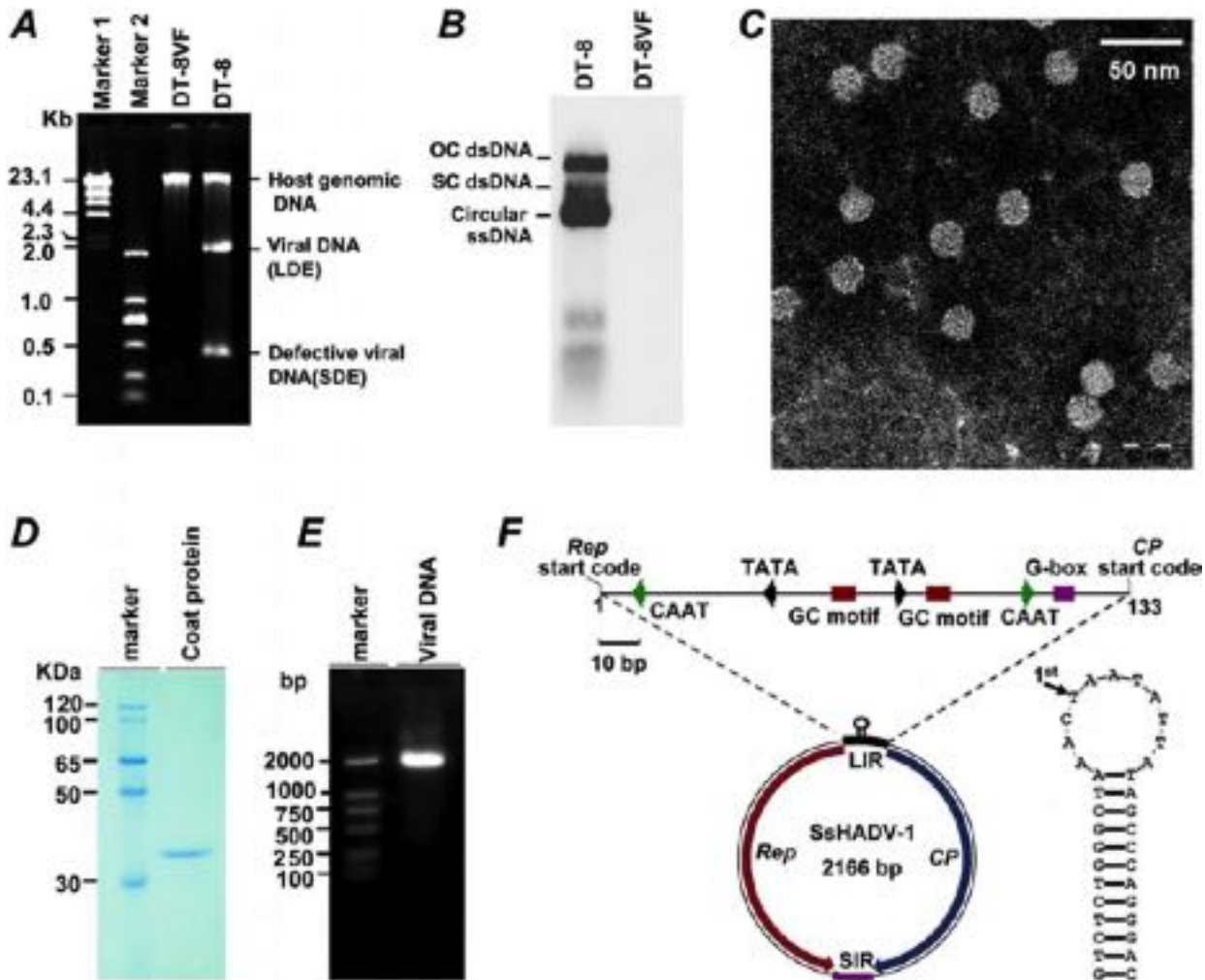


Fig. 3 Genomic structure and virions of *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 (SsHADV-1). A, Extrachromosomal DNA elements in strain DT-8 DNA samples were extracted with CTAB method from mycelia of strain DT-8. DNA samples were separated on 1.0% agarose gel. Lane Marker 1, λ -Hind III-digested DNA Marker; Lane Marker 2, DL2000 DNA ladder marker (TaKaRa). B, Southern blot analysis of total DNA extracted from mycelia. The forms of viral DNA are indicated as “OC dsDNA” for open circular double-stranded (ds)DNA, “SC” dsDNA for supercoiled dsDNA, and “circular” for circular single-stranded (ss) DNA. A 379-bp DNA fragment of the Rep gene was PCR amplified and labeled with α - 32 P dCTP and used as a probe. C, Isometric virions (about 22 nm) purified from mycelia of strain DT-8, observed under transmission electron microscopy. Virions were negatively stained with 1% uranyl acetate before observation. D, The major protein (coat protein) of SsHADV-1 extracted from virions separated by SDS/PAGE. Virions were extracted from mycelial mass of strain DT-8 and purified by ultra-centrifugation. The protein was stained with Coomassie blue before observation. E, Viral genomic DNA extracted from virions separated on 1% agarose gel. F, Genome structure of SsHADV-1. Two large Open Reading Frame (ORFs), one coding for coat protein (CP) and one for putative replication initiation protein (Rep) were presented as thick arrows. The positions of the conserved stem-loop structure for small circular single-stranded DNA viruses, large intergenic region (LIR) and small intergenic region (SIR) were displayed; the detail DNA sequence of stem-loop structure was also shown on the right. LIR is also shown in an expanded form to indicate the elements of the bidirectional promoter. Reproduced from Yu, X., Li, B., Jiang, D., *et al.*, 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences of the United States of America* 107, 8387–8392.

on the aerial parts of plant (Fig. 5(D)–(E)); however, virions on plant are not stable and will be disabled in a short time after sprayed on plants. These properties suggest that SsHADV-1 has a potential to biological control of plant diseases.

Mutualistic Interaction Between SsHADV 1 and Mushroom Sciarid Fly

Mushroom sciarid fly *Lycoriella ingenua* (Diptera: Sciaridae) is a common pest in mushroom houses (Fig. 6(A)). A phenomenon is that SsHADV-1-infected strain DT-8 always specifically attract sciarid fly to lay eggs on its colony, and the larva feeds on the colony and

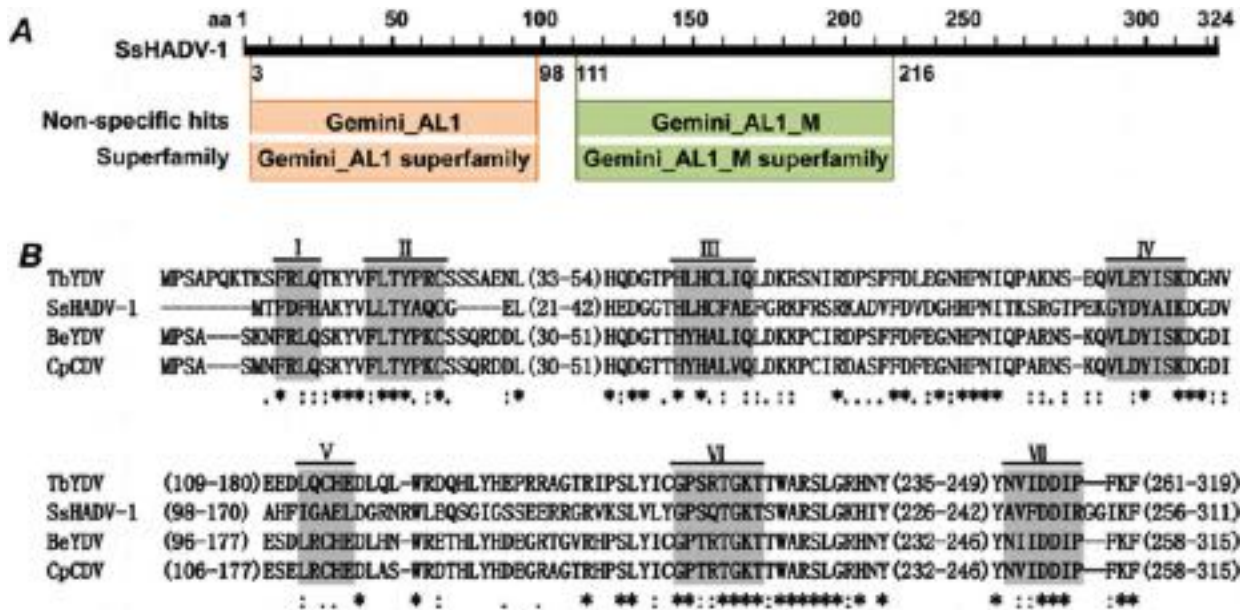


Fig. 4 Conserved domains and motifs on putative Rep of SsHADV-1. A, The locations of two conserved domains, Gemini_AL1 and Gemini_AL1_M are geminivirus Rep catalytic domain (accession no. pfam00799) and geminivirus Pep protein central domain (pfam08283), the E-values are 1.77e-12 and 5.01e-08. This schematic is automatically generated by the Blastp program on NCBI when using Rep of SsHADV-1 as a query sequence. B, Alignment analysis of amino acid sequences among SsHADV-1 and selected viruses from the genus *Mastrevirus* in the family *Geminiviridae* by using CLUSTALX (2.0). Identical and similar amino acid residues are indicated with asterisks and colons. Conserved motifs I to VII are IRD, RCR-I, RCR-II, RCR-III, RBR, NTP Binding-A, and NTP binding-B are shaded with light gray color. "TbYDV" means tobacco yellow dwarf virus (NP_620726), "CpCDV" means chickpea chlorotic dwarf virus (YP_002014712) and "BeYDV" means bean yellow dwarf virus (NP_612221); numbers in brackets are the positions of amino acid residues that are not listed. Reproduced from Yu, X., Li, B., Jiang, D., et al., 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences of the United States of America* 107, 8387–8392.

eats out all hyphae (Fig. 6(B)). After feeding on colony of the virus-infected strain, the insect develops to pupa from larva, and then adult emerges from pupa, after mating, female adult then lays eggs back to the fungal colonies. Viral DNA of SsHADV-1 could be determined by PCR amplification from the larva, pupa, adults (female and male), and eggs of fly that was fed on virus-infected colonies (Fig. 6(C)–(E)), suggesting that this fly carries hyphae attaching on insect surface or in the gastrointestinal system, or that SsHADV-1 enters into insect cells. The insects fed on a virus-infected colony were further examined using immunofluorescence observation with the antibody of coat protein, and results showed that strong fluorescence signal could be observed in the head of larva and pupa, in the digestive tract and oviduct of adults and in eggs (Fig. 6(F)–(I)). These observations suggest that SsHADV-1 enters into tissues and cells of sciarid fly, and replicates in this insect. The virions of SsHADV-1 was used to challenge the cells of *Spodoptera frugiperda* (Sf9), and SsHADV-1 was examined by passage experiment, RT-PCR amplification, Flow cytometry analysis, Northern blot analysis, and immunofluorescence observation under confocal microscope (Fig. 6(K) and (L)). These assays demonstrated that SsHADV-1 could replicate in Sf9 cells and in sciarid fly.

To understand why the virus-infected colony could attract sciarid fly, the volatile compounds released by both virus-infected and virus-free colonies were collected and analyzed with GC–MS. The results showed that virus-infection suppresses the biosynthesis of 1-Octen-3-ol and 3-Octanone, and "Y" shape tube test revealed that high concentrations of these two compounds are likely to be repellents for adults of sciarid fly. The fecundity of female adults can be significantly increased when larvae are fed on the virus-infected colony. This result suggests that SsHADV-1 modifies the secondary metabolites of *S. sclerotiorum*.

Transmission of SsHADV-1

Mycoviruses are often transmitted vertically by host's reproduction, mostly by host asexual reproduction, occasionally by both asexual and sexual reproduction, and also transmitted horizontally by hyphal anastomosis between virus-infected and virus-free strains. Horizontal transmission of mycoviruses is usually suppressed by a fungal vegetative incompatibility system, hence virus cannot efficiently transmit between two vegetative incompatible strains. SsHADV-1 cannot transmit vertically since *S. sclerotiorum* does not form any asexual spores, while sclerotia cannot germinate easily to form apothecia, and ascospore-offsprings are virus-free. Unlike other mycoviruses, SsHADV-1 can transmit between two vegetative incompatible strains (Fig. 7). The possible mechanism is that virions of SsHADV-1 can infect hyphae of *S. sclerotiorum* directly via an extracellular route. While growing, hyphae of virus-infected strain frequently break and tear so that virions are easily released. When hyphae of virus-free strain contact

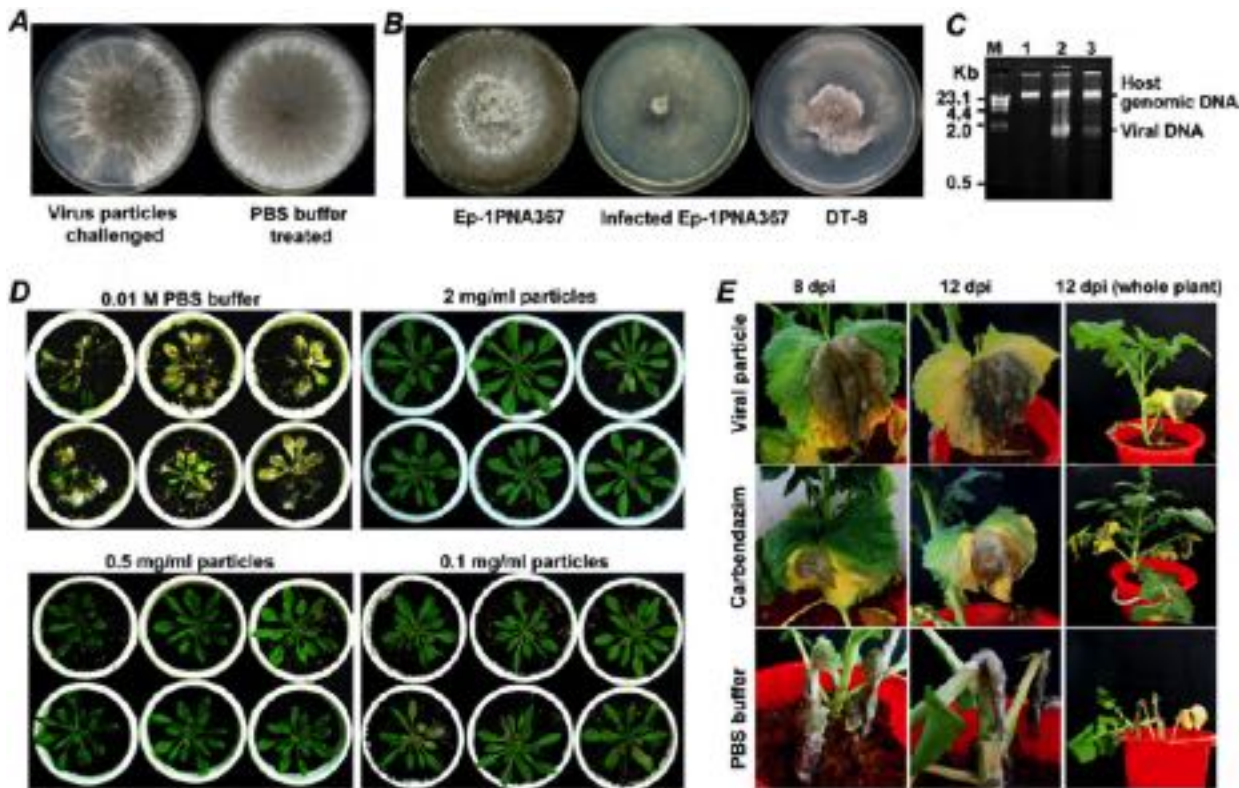


Fig. 5 SsHADV-1 directly infects intact hyphae of *Sclerotinia sclerotiorum* and protects plants against *S. sclerotiorum*. A–C, Purified virions infect hyphae on PDA medium. A, Sectoring observed in a colony of a virus-free strain Ep-1PNA367 at 2 dpi. Strain Ep-1PNA367 grew on a PDA plate for 1 day and then virions were spread on the PDA medium around the margin of the young colony with a distance of about 1–2 cm. PBS buffer was used instead of virion suspension as a control, no sectoring was observed in the colony. The photographs were shot at 4 dpi; B, Abnormal colony morphology of SsHADV-1-infected strain Ep-1PNA367, colony morphology of strain Ep-1PNA367 and strain DT-8 were also presented. All cultures were incubated on PDA medium for 7 d at 20°C; C, Viral DNA could be extracted from infected strain Ep-1PNA367 and observed by agarose gel electrophoresis. Lane M: λ -Hind III digested DNA Marker. Lanes 1–3, DNA samples of strains Ep-1PNA367, virus-infected Ep-PNA367, and DT-8, respectively. D–E, Spraying virions of SsHADV-1 protects plants against the infection of *S. sclerotiorum*. D, Virions suspension prevented *S. sclerotiorum* from killing plants of *A. thaliana*. Virions suspensions at different concentrations were spread on one leaf of each plant, and then mycelial agar disks from the merge of active colony of strain Ep-1PNA367 were placed on the treated leaves. The inoculated plants were maintained in an incubator with 100% relative humidity (RH) at 20°C, and photographs were shot at 6 dpi. E, Therapeutic activity of virions against the attack of *S. sclerotiorum*. Virion crude suspension inhibited the expansion of *S. sclerotiorum* on rapeseed plants and cured the lesions. *S. sclerotiorum* were pre-inoculated on plant leaves for 2 days for lesion development, and then virions suspension, chemical fungicide Carbendazim (100 ppm) or PBS buffer was sprayed on plants. Three leaves were inoculated for each plant. Plants were maintained in an incubator with 100% RH at 20°C, and photographs were shot at 8 and 12 dpi. Reproduced from Yu, X., Li, B., Fu, Y., *et al.*, 2013. Extracellular transmission of a DNA mycovirus and its use as a natural fungicide. *Proceedings of the National Academy of Sciences of the United States of America* 110, 1452–1457.

virions, infection happens. Although transmission vectors of RNA mycoviruses have not been discovered yet, SsHADV-1 can replicate in sciarid fly and enter into eggs, suggesting that sciarid fly can possibly transmit SsHADV-1. Viruliferous larva, pupa, adults, and eggs (acquired virus by feeding on virus-infected colony) and larvae and pupae which are injected with purified virions could transmit virus to virus-free fungal strains. The transmission also occurs on rapeseed plants (Fig. 8), suggesting that SsHADV-1 and sciarid fly undergo a mutualistic lifestyle in the field.

Distribution of SsHADV-1

It is most likely that SsHADV-1 is widely distributed. Firstly, SsHADV-1 has been observed in strains of *S. sclerotiorum* isolated from many places in China; Secondly, it also has been detected in Australia strains of *S. sclerotiorum*; Thirdly, viral DNA of SsHADV-1 has been detected in the river sewage in New Zealand. Additionally, viral DNA of SsHADV-1 is detected in dragonflies (*Pantala hymenaea*). Different viral strains in the same species as SsHADV-1 (qcv23688.1) were also found in Haddock in the United States, with Rep identity of 78% (253/324) and similarity of 83% (272/324). These findings indicate that the distribution of SsHADV-1 is worldwide, at least in Asia, North America, and Oceania.

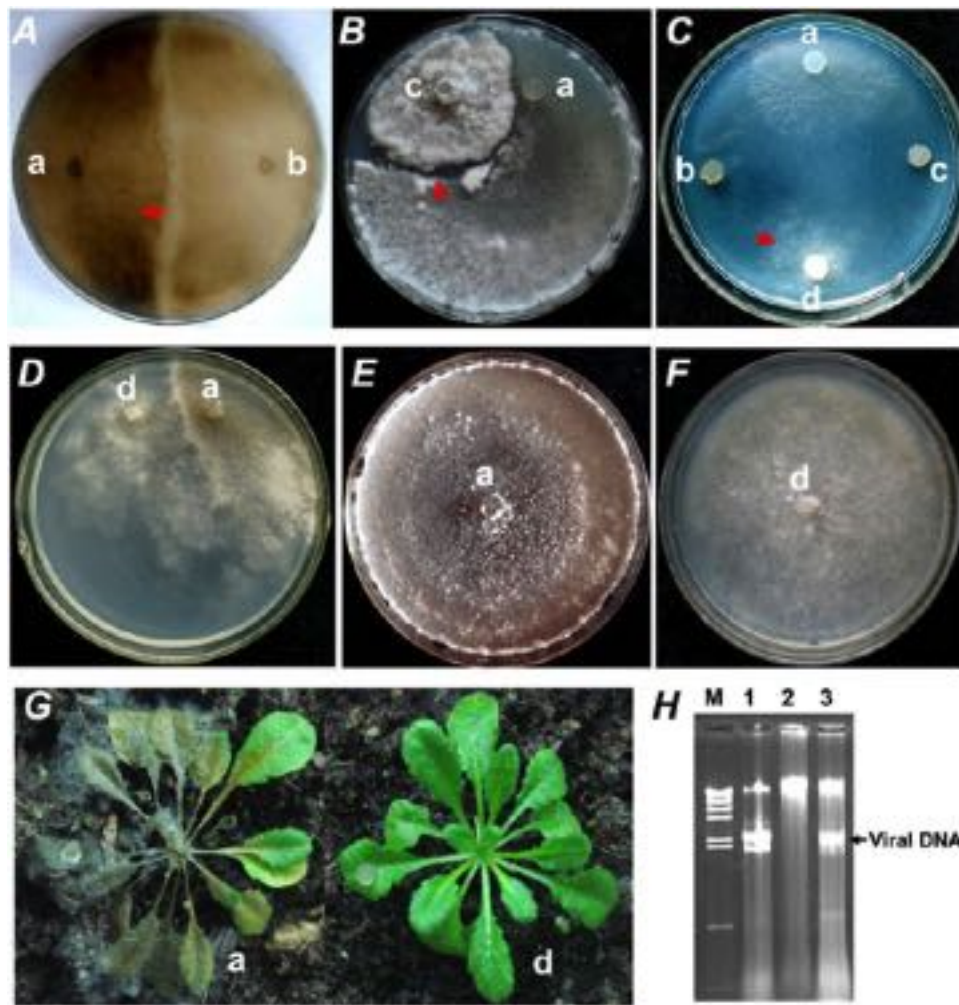


Fig. 6 SsHADV-1 replicates in insect. A, Eggs, larva, pupa, and adults of a mushroom sciarid fly *Lycoriella ingenua* (Diptera: Sciaridae); B, The larva of sciarid fly feeding on colony of *S. sclerotiorum*, red arrowheads show the larva, and white arrowheads show galleries generated by feeding; C, PCR detection of virus in/on sciarid fly after reared on the colony of strain DT-8 with virus-specific primers pairs; D, Determination of SsHADV-1 with Southern blot analysis in sciarid fly. Lane M, DNA markers; Lane 1, DNA from strain DT-8 (positive control); Lanes 2–5, DNA samples extracted from eggs, larvae, pupae and adults which were fed on the colony of strain DT-8, respectively; Lane 6, DNA from larvae reared on strain DT-8VF (negative control). A DIG-labeled probe which was made using a DNA fragment (814 bp) derived from genomic DNA was used to probe the Southern blots. E, Determination of virus retention time in the body of larvae by using PCR amplification. Lane M, DNA marker; Lane 1, DNA sample from strain DT-8; Lanes 2–5, DNA samples from larvae starved for 1, 3, 5, and 7 d, respectively; Lane 6, DNA sample from larvae reared on strain DT-8VF. F–J, Immunofluorescence detection of SsHADV-1 in larva, pupa, midgut, and midgut of female adults, and eggs of *L. ingenua*. Insect developed either on the colony of strain DT-8 (a) or on the colony of strain DT-8VF (b) (negative control) was used for detection. F, Larvae, longitudinal section of larvae (head parts); G, Pupae, longitudinal section of pupae (head parts); H, Midgut of female adults. Photographs shown at the upper right were magnified from a part of midgut (indicated by red boxes); I, Ovarian duct of female adults; J, Eggs dissected from female adults. Bars at the bottom right are 200 μ m for F and G, 100 μ m for H, I, and J, 50 μ m for the upper right in H. Samples were stained with secondary antibody FITC conjugated to virions-specific monoclonal antibody prepared with viral CP protein. The Immunofluorescence reaction was observed under a confocal microscope (fv1000mp, Olympus). K and L, SsHADV-1 replicates in Sf9 cells. K, Immunofluorescence examination of SsHADV-1 in Sf9 cells, cells were inoculated with SsHADV-1 virions and observed at 72 hpi bars = 20 μ m. L, Northern blot analysis for expression of viral Rep and CP gene in Sf9 cells. DIG-labeled DNA probes of 853 bp and 901 bp corresponding to viral DNA for Rep and CP gene, respectively, were used. Reproduced from Yu, X., Li, B., Jiang, D., *et al.*, 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences of the United States of America* 107, 8387–8392.

The Taxonomy of SsHADV-1

SsHADV-1 is phylogenetically closely related to geminiviruses that infect numerous plants, the Rep protein of SsHADV-1 showed high similarity to these of viruses in the family *Geminiviridae*. However, the coat protein gene is different from these of viruses in family *Geminiviridae*. Hence, SsHADV-1 should have some evolutionary relation with geminiviruses, but also significantly different from these viruses. In 2012, a novel genus *Gemycircularvirus*, a name from Gemini-like myco-infecting circular virus was established with SsHADV-1

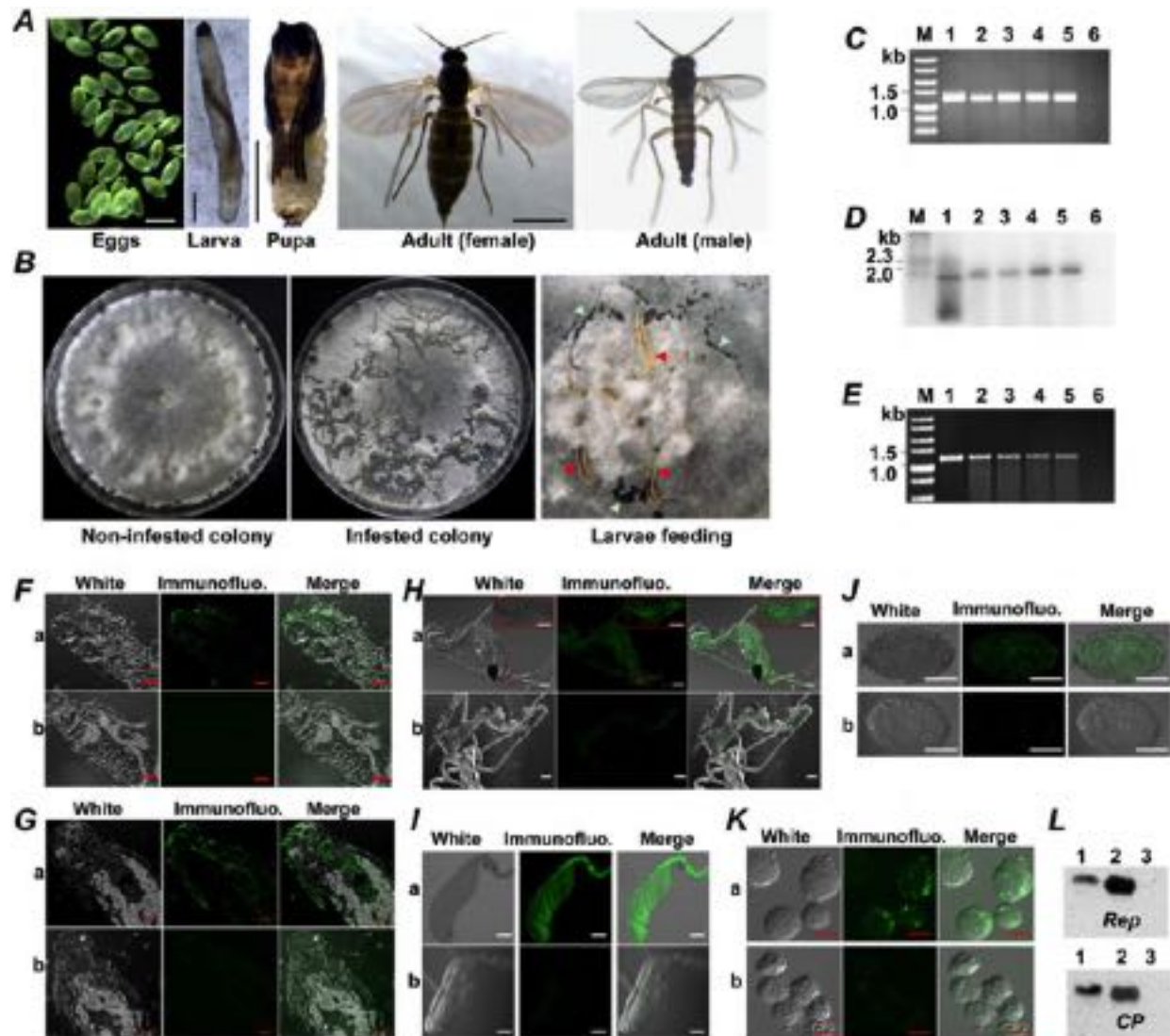


Fig. 7 SsHADV-1 transmits between two vegetative incompatible strains of *S. sclerotiorum*. A and B, Incompatible reaction (cell death zone, red arrow indicated) between a hygromycin-resistant strain Ep-1PNA367R (a) and strain DT-8VF (b) or its isogenic strain DT-8 (c). For A, colonies were dual cultured on a PDA plate for one week. Photograph was shot from the reverse side of the plate, for B, strain DT-8 was inoculated 2 days ahead of inoculation of strain Ep-1PNA367R on a PDA plate, the photograph was taken at 7 days post inoculation of strain Ep-1PNA367R. C, Both strain Ep-1PNA367R and SsHADV-1-infected strain Ep-1PNA367R (d) derived from dual culture in B could grow on hygromycin-amended PDA medium (50 μ g/mL), while strains DT-8 and DT-8VF could not. Both strain DT-8 and SsHADV-1-infected strain Ep-1PNA367R were developed in plate for 4 d, while strain Ep-1PNA367R and strain DT-8VF for one day. D-G, SsHADV-1-infected strain Ep-1PNA367R conferred hypovirulence and its associated trait to strain Ep-1PNA367R by dual cultivation. D, Colony morphology of strain Ep-1PNA367R changed after contacting with the SsHADV-1-infected strain Ep-1PNA367R, which is different from strain Ep-1PNA367R (E), but similar to virus-infected one (F); Colonies in the dual culture were developed for 10 days for virus-infected strain, and for 7 days for strain Ep-1PNA367R, colonies for E and F were developed on PDA medium for 7 d. G, SsHADV-1-infected strain Ep-1PNA367R lost virulence on *Arabidopsis* (d). H, Viral DNA was extracted in newly infected strain Ep-1PNA367R. Lane M, λ -Hind III digested DNA Marker; Lane 1, newly infected strain Ep-1PNA367R; Lane 2, Ep-1PNA367R; Lane 3, strain DT-8. Reproduced from Liu, S., Xie, J., Cheng, J., *et al.*, 2016. DNA mycovirus infects a mycophagous insect and utilizes it as a transmission vector. *Proceedings of the National Academy of Sciences of the United States of America* 113, 12803–12808.

as the exemplar virus. In 2016, a new family, *Genomoviridae* (sigil: Ge- for geminivirus-like, nomo- for no movement protein) was established with SsHADV-1 representative virus strain of the type species, *Sclerotinia gemycircularvirus 1*, in the genus *Gemycircularvirus*. In *Gemycircularvirus*, there are more than 43 sister species reported from the world at different niches. These gemycircularviruses are associated with plants, animal tissues (including humans, mammals, birds, insects, sea fish, and reptile, etc.), and feces of mammals and birds, sewage and sediments and waste water. However, only SsHADV-1 is well-studied, the others are only viral sequences, and their exactly hosts are not known. Based on the complete amino acid sequence of Rep of SsHADV-1 and other selected viruses, a phylogenetic

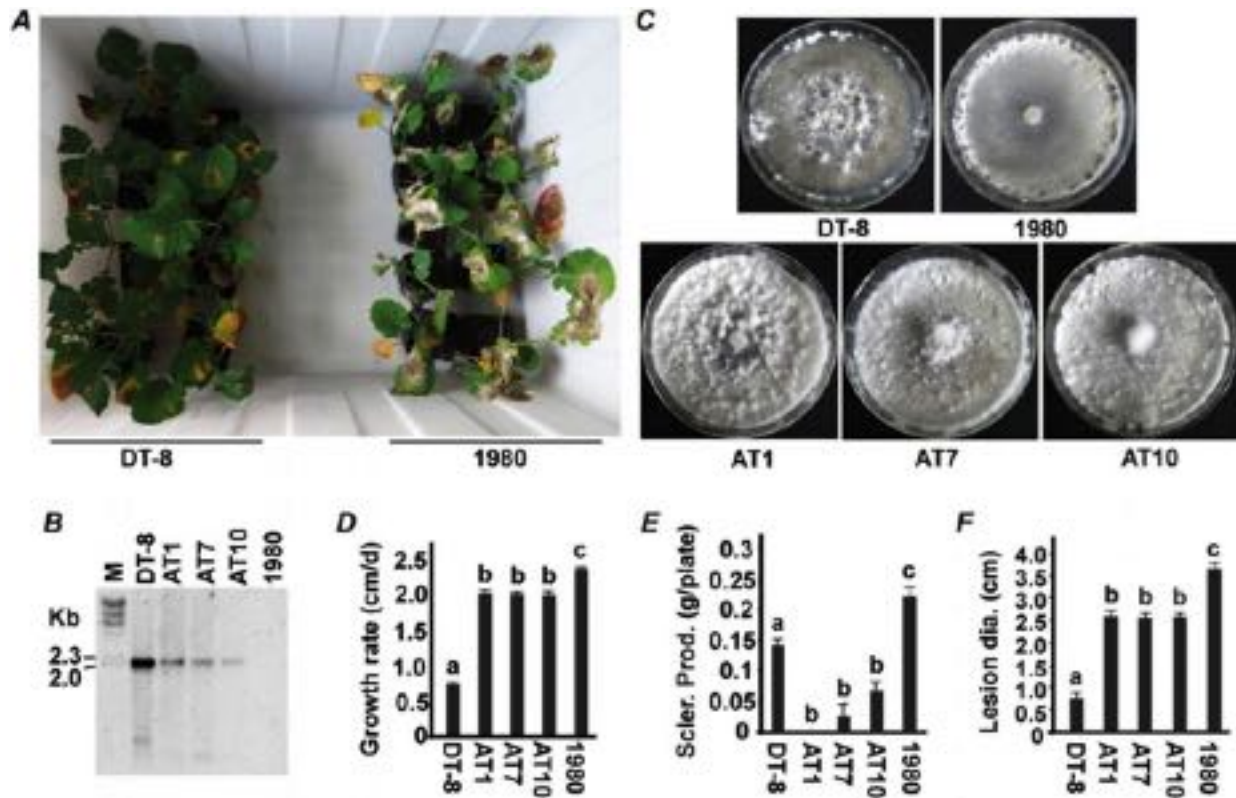


Fig. 8 Transmission of SsHADV-1 by mushroom sciarid fly on rapeseed plants. A, Rapeseed plants were inoculated with strains DT-8 (left) and 1980 (right) with a 30 cm distance between two basins. The basin was sealed and kept 100% RH and then virus-free larvae were placed on small lesions induced by strain DT-8. When right side plants were killed by strain 1980, subcultures recovered from diseased plant residues were subjected to virus detection; B, PCR-positive subcultures (AT1, AT7, and AT10) were confirmed with Southern blot analysis with the same probe used in Fig. 6(D). Viral DNA for Southern blot analysis was amplified with rolling circle amplification (RCA); DNA samples extracted from strains DT-8 and 1980 were used as controls. C-F, Colony morphology, growth rate, sclerotial production and virulence of virus-infected subcultures AT1, AT7, and AT10 were significantly different from that of virus-free strain 1980. Reproduced from Liu, S., Xie, J., Cheng, J., *et al.*, 2016. DNA mycovirus infects a mycophagous insect and utilizes it as a transmission vector. *Proceedings of the National Academy of Sciences of the United States of America* 113, 12803–12808.

tree was constructed using Maximum Likelihood method with a MEG 7.0 program. From the evolutionary tree, viruses in the *Gemycircularvirus* and its-related genera in the family *Genomoviridae* are extremely rich in diversity (Fig. 9).

Explore SsHADV-1 to Control Fungal Disease

The history of using hypovirulence-associated mycovirus to control fungal disease can date back to the early 1970s of the last century. At that time, a transmissible hypovirulent strain was isolated from self-cured cankers caused by the chestnut blight fungus (*Cryphonectria parasitica*), and this hypovirulent strain was successfully used to control chestnut blight in European countries. It became evident later that the transmissible hypovirulent strain was actually caused by the infection of mycovirus (*Cryphonectria hypovirus 1*). This successful example prompted a large number of phytopathologists to screening for similar hypovirulence-associated mycoviruses in different pathogenic fungi and expected to be applied in different fungal disease systems. Unfortunately, these studies have not achieved significant success, even for the hypovirus, the control efficiency of chestnut blight in the United States is not satisfactory. The common consensus is that the efficient transmission of mycoviruses is limited by the host vegetative incompatibility system (also called non-self-recognition system). Mycoviruses can only spread efficiently among host individuals in the same vegetative compatible group, but cannot move in different vegetative compatible groups. Fungi usually have complicated vegetative incompatibility system, making use of mycoviruses to control diseases almost impossible. Of course, a group of scholars have made continuous efforts to overcome this difficulty, which will make it possible to control chestnut blight with hypovirus in the near future.

Unlike most known viruses, SsHADV-1 has very unique characteristics. First, the vegetative incompatibility could not prevent the efficient spread of SsHADV-1 in the population of *S. sclerotiorum*; second, SsHADV-1 can drive insect as transmission vectors, which may help SsHADV-1 spread in the field in a short time; third, virions of SsHADV-1 can directly infect hyphae of *S. sclerotiorum*, that

makes it possible to use virus particles or processed virus particles to control *Sclerotinia* disease. Field experiment demonstrated that spraying hyphal fragment suspension of SsHADV-1-infected strain DT-8 on rapeseed plants at the blossom stage could also reduce disease severity and enhance rapeseed yield up to 16.4% under field conditions. SsHADV-1 and strain DT-8 were introduced to the United States in 2017 by Prof. Weidong Chen in Washington State University for investigating its potential on biological control of white mold of leguminous crops caused by *S. sclerotiorum*.

Further Reading

- Liu, S., Xie, J., Cheng, J., *et al.*, 2016. DNA mycovirus infects a mycophagous insect and utilizes it as a transmission vector. *Proceedings of the National Academy of Sciences of the United States of America* 113, 12803–12808.
- Xie, J., Jiang, D., 2014. New insights into mycoviruses and exploration for the biological control of crop fungal diseases. *Annual Review of Phytopathology* 52, 45–68.
- Xu, L., Li, G., Jiang, D., Chen, W., 2018. *Sclerotinia sclerotiorum*: An evaluation of virulence theories. *Annual Review of Phytopathology* 56, 311–338.
- Yu, X., Li, B., Fu, Y., *et al.*, 2013. Extracellular transmission of a DNA mycovirus and its use as a natural fungicide. *Proceedings of the National Academy of Sciences of the United States of America* 110, 1452–1457.
- Yu, X., Li, B., Jiang, D., *et al.*, 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proceedings of the National Academy of Sciences of the United States of America* 107, 8387–8392.
- Zhang, D., Nuss, D., 2016. Engineering super mycovirus donor strains of chestnut blight fungus by systematic disruption of multilocus *vic* genes. *Proceedings of the National Academy of Sciences of the United States of America* 113, 2062–2067.

Structure of Double-Stranded RNA Mycoviruses[☆]

José R Castón, National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Nobuhiro Suzuki, Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

Said A Ghabrial[†], Department of Plant Pathology, University of Kentucky, Lexington, KY, United States

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Double-stranded RNA (dsRNA) viruses infect practically all organisms, from bacteria to unicellular and/or multicellular simple eukaryotes (fungi, protozoa), through to plants and animals. Although dsRNA viruses are a rather diverse group, they share general architectural principles and numerous functional features. The complexity of the capsid ranges from single shell to multilayered and concentric capsid. Whereas the outer shell has a protective role and is involved in cell entry, the innermost capsid (or inner core) is devoted to the organization of the viral genome and viral polymerase. This inner capsid consists of 120 protein subunits arranged in a T=1 icosahedral shell, i.e., a capsid protein (CP) dimer is the asymmetric unit. The T=1 capsid is also referred to as a “T=2 layer” – an exception to the quasi-equivalence theory introduced by Caspar and Klug.

The T=1 capsids of dsRNA viruses are critical for genome replication (minus-strand synthesis) and transcription (plus-strand synthesis), with the viral RNA-dependent RNA polymerase(s) (RdRp) frequently packaged as an integral component of the capsid. T=1 capsids also function as molecular sieves, allowing the exit of positive single-stranded RNA transcripts for protein synthesis, and the entrance of nucleotides for intra-capsid RNA synthesis. The pores are presumably small enough to exclude potentially degradative enzymes. In addition, T=1 capsid remains structurally undisturbed throughout the viral cycle, isolating dsRNA molecules and any replicative intermediates, thus preventing the triggering of dsRNA sensor-mediated antiviral host defense mechanisms such as RNA silencing, interferon synthesis and apoptosis.

The totiviruses *Saccharomyces cerevisiae* virus L-A (ScV-L-A) and *Ustilago maydis* virus H1 (UmV-H1), which infect the yeast *Saccharomyces cerevisiae* and the smut fungus *Ustilago maydis*, respectively, were the first unambiguously described viruses with a T=1 capsid formed by 12 decamers (rather than 12 pentamers). The conservation of this stoichiometry and architecture is probably related to the stringent requirements of capsid RNA metabolism-associated activity (the capsid organizes the packaged genome and the replicative complex[es]).

T=1 120-subunit capsids have been described for members of the *Reo-* and *Picobirnaviridae*, which mostly infect higher eukaryotic organisms. They have also been described for bacteriophages of the family *Cystoviridae* that infect the prokaryote *Pseudomonas syringae*. These capsids are also present in members of the *Toti-*, *Partiti-*, *Megabirna-*, *Chryso-*, and *Quadriviridae* families, that largely infect unicellular and/or multicellular simple eukaryotes such as fungi, protozoa, but also some plants.

In contrast with the smooth outer surface of T=1 capsid of reoviruses, the 120-subunit capsids of fungal dsRNA viruses have a corrugated outer surface with protuberances rising above the continuous protein shell. Notably, the average thickness of a 120-subunit T=1 CP is 15–30 Å in mammalian dsRNA viruses, but those of mycoviruses are thicker. Unlike their bacteria- and higher eukaryote-infecting counterparts, most mycoviruses are transmitted by cytoplasmic interchange; they never leave the host, and indeed have no strategy for entering host cells. Recent studies of fungal (and protozoan) dsRNA viruses have identified functional and structural features dissimilar to those recorded for reoviruses, as well as evolutionary relationships among T=1 capsid structural proteins. Mycovirus particles accumulate in the fungal cytoplasm; the close relationship between the fungal dsRNA virus and its host probably placed many constraints on the virus that it overcame by increasing CP complexity.

The T=1 capsid proteins of four mycoviruses have been resolved at the atomic level: Gag of the yeast virus ScV-L-A (family *Totiviridae*), CP of *Penicillium chrysogenum* virus (PcV, family *Chrysoviridae*), CP of *Penicillium stoloniferum* virus F (PsV-F, family *Partitiviridae*), and the heterodimer P2-P4 of *Rosellinia necatrix* quadrivirus 1 (RnQV1, family *Quadriviridae*).

Structure of dsRNA Virus Capsids

Totiviruses

Saccharomyces cerevisiae virus L-A is the type species of the genus *Totivirus* (family *Totiviridae*). The ScV-L-A genome is a 4.6 kbp, single-segment dsRNA molecule that encodes a major CP (termed Gag; 680 residues, 76 kDa) and the viral polymerase (Pol; 868 residues, 94 kDa) as a Gag-Pol fusion protein generated by – 1 ribosomal frameshifting (Pol is bound covalently to the inside of the particle wall).

The structure of ScV-L-A was first examined by three-dimensional cryo-electron microscopy (3D cryo-EM) and later by X-ray crystallography at a resolution of 3.4 Å. Scanning transmission electron microscopy (STEM) was used to determine the virus stoichiometry. The rough, icosahedral, ~40 nm diameter T=1 lattice of ScV-L-A has 120 copies of Gag, of which one or two are fused to the Pol moiety (Fig. 1(a) and (b)). The structural unit is an asymmetric Gag dimer. Each Gag monomer can adopt two related conformations, termed A and B, which have notable structural differences and reside in different bonding environments (i.e., make non-equivalent contacts) within

[☆]This work is dedicated to the memory of our friend and colleague, Said A. Ghabrial, who passed away in November 2018.

[†]Deceased.

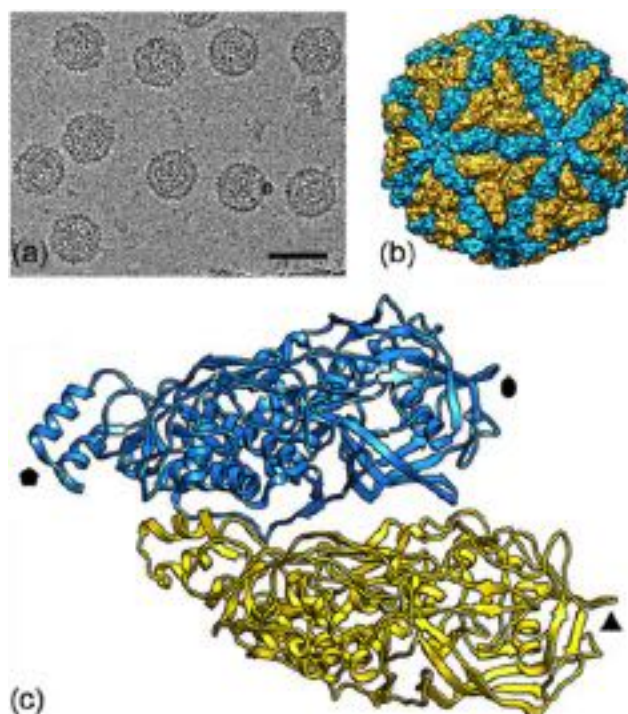


Fig. 1 Structure of the ScV-L-A $T=1$ capsid and X-ray-based structure of the CP. (a) Cryo-EM image of ScV-L-A. Bar, 500 Å. (b) Cryo-EM $T=1$ capsid of ScV-L-A viewed along a twofold axis of icosahedral symmetry, showing the Gag subunits A (blue) and B (yellow). (c) Atomic model of a Gag dimer (top view; PDB accession number 1m1c). Icosahedral symmetry five- (pentagon), three- (triangle) and two-fold (oval) axes are indicated in black.

the capsid (**Fig. 1(c)**). These subunits are arranged in two sets of five: five A subunits directly surrounding the icosahedral five-fold axis, leaving an 18 Å diameter channel for the entry of nucleotide triphosphates and the exit of viral mRNA, and five B subunits intercalated between the A subunits, forming a decamer. Twelve such (A:B)₅ decamers then constitute the complete capsid. This quaternary organization is similar to the 120-subunit $T=1$ inner core of reoviruses, in which adjacent A and B subunits within each decamer are oriented approximately in parallel, suggesting an asymmetric A:B dimer as a possible intermediate for capsid assembly.

Gag functions as an enzyme and has a major role in the sophisticated interaction between ScV-L-A and the host cell. The Gag segment Gln139-Ser182 (in which His154 is the active site) contributes to the rough outer surface of the capsid, mediates decapping of cellular mRNAs and transfers the 7-methyl-GMP (m^7 GMP) cap from the 5' end of the cellular mRNA to the 5' end of the viral RNA. L-A counters a host exoribonuclease that targets uncapped RNAs (such as viral mRNA), allowing the latter to compete with host mRNA for use of the translation machinery.

Helminthosporium victoriae virus 190S (HvV190S), the prototype of the genus *Victorivirus* in the family *Totiviridae*, infects the filamentous fungus *H. victoriae*, and has a similar capsid organization to that of ScV-L-A. The smooth HvV190S capsid is composed of 120 CP monomers, with RdRp incorporated as a separate, non-fused protein. The RdRp is either non-covalently associated with the underside of the capsid (as in reoviruses), or free in the capsid interior, or non-covalently bound to the genome. 3D structures have also been reported for two protozoal members of this family, *Trichomonas vaginalis* virus 1 and *Giardia lamblia* virus, and show strong similarities to the fungal virus structures.

Chrysovirus

Chrysovirus are isometric virions characterized by a multipartite genome. *Penicillium chrysogenum* virus (PcV) is the prototype virus of the *Chrysoviridae*, a family of largely symptomless mycoviruses with a genome typically consisting of four monocistronic genome segments (genome size 2.4–3.6 kbp). Each segment is encapsidated separately in a similar particle, i.e., chrysovirus are multisegmented and multiparticulate virions. dsRNA-1 (3.6 kbp) encodes the RdRp (1117 residues, 128.5 kDa; 1 or 2 copies per virion), dsRNA-2 (3.2 kbp) encodes the CP (982 residues, 109 kDa), and dsRNA-3 and -4 (3 and 2.9 kbp, respectively) code for virion-associated proteins of unknown function (912 residues, 101 kDa and 847 residues, 95 kDa, respectively).

The 3D structures of the capsids of two chrysovirus have been determined by cryo-EM analysis, that of PcV at atomic resolution, and that of *Cryphonectria nitschkei* chrysovirus virus 1 (CnCV1) at subnanometer resolution. Stoichiometric estimates by analytical ultracentrifugation analysis indicated that PcV and CnCV1 virions are exceptions to the most extended tendency among dsRNA viruses – a $T=1$ core with 60 equivalent dimers – since they have an authentic $T=1$ capsid formed by 60 copies of a single monomer. The capsid diameter is 400 Å and the protein shell 48 Å thick (**Fig. 2(a)** and **(b)**). Similar to ScV-L-A, the outer capsid surface of PcV is relatively

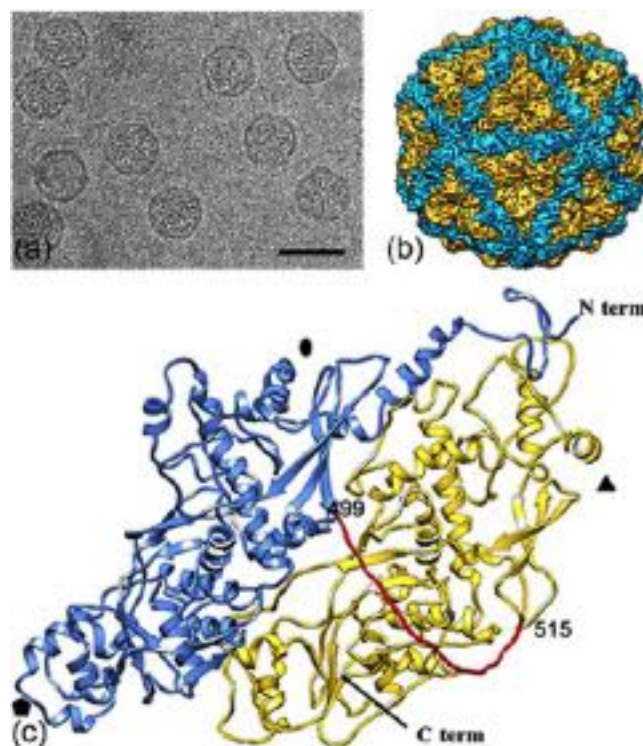


Fig. 2 Cryo-EM- based structure of the PcV T=1 capsid protein. (a) Cryo-EM image of PcV. Bar, 500 Å. (b) T=1 capsid of PcV viewed along a twofold axis of icosahedral symmetry, showing the N-terminal domain A (1–498, blue), the linker segment (499–515, red), and the C-terminal domain B (516–982, yellow). (c) Top view of the atomic model of the PcV CP (3j3i; 982 residues). Symbols indicate icosahedral symmetry axes.

uneven with 12 outwardly protruding pentons (each containing five copies of the CP); this contrasts with the T=1 capsid of reoviruses, in which the CP has a plate-like structure and serve as a template to prime the assembly of the T=13 surrounding capsid. The 982-residue CP of PcV is formed by duplication of an α -helical domain; this is indicative of gene duplication despite negligible sequence similarity between the two roughly parallel α -helical domains (Fig. 2(c)). The N-terminal A domain (residues 1–498; Fig. 2(c), blue) and the C-terminal B domain (residues 516–982; Fig. 2(c); yellow) are connected by a 16-residue linker (Ala499-Ile515; Fig. 2(c), red) accessible from the capsid outer surface. These domains are arranged in two sets of five; five A domains directly surround the icosahedral fivefold axis, and five B domains intercalated between them, forming a pseudodecamer (Fig. 3(a)). This organization is clearly reminiscent of the 120-subunit T=1 lattice of totivirus capsids, in which the two asymmetrical dimer components are arranged approximately in parallel. The structural details of PcV capsid reinforce the idea that a T=1 layer with a dimer as the asymmetric unit provides an optimal framework for managing dsRNA metabolism.

Superimposition of the PcV A and B α -helical domains (Fig. 3(b) and (c)) identifies a single “hotspot” located on the capsid outer surface where variations are introduced by insertion of 50–100 residue polypeptides (Fig. 3(c), orange triangle). A preferential insertion site would allow the acquisition of new functions while preserving basic CP folding. It is plausible that, in addition to its structural role, chrysovirus CP might also have enzymatic activity, like ScV-L-A CP.

Quadriviruses

Rosellinia necatrix quadrivirus 1 (RnQV1) is the exemplar of the type species of the genus *Quadrivirus* in the family *Quadriviridae*. The filamentous ascomycete *Rosellinia necatrix*, a pathogen of many plants, can be infected by dsRNA viruses belonging to at least six families. RnQV1 is associated with latent infections and has a multipartite genome consisting of four monocistronic dsRNA segments (as in chrysoviruses) with genome sizes ranging from 3.7 to 4.9 kbp. DsRNA-1 (4.9 kbp) codes for a protein of unknown function (1602 residues), dsRNA-2 (4.3 kbp) encodes the P2 CP (1356 residues), dsRNA-3 (4 kbp) codes for RdRp (1117 residues), and dsRNA-4 (3.7 kbp) codes for the P4 CP (1061 residues).

RnQV1 strains W1075 and W1118, isolated from different locations in Japan, have been analyzed by 3D cryo-EM and analytical ultracentrifugation. Their P2 and P4 proteins co-assemble into isometric virus particles ~45 nm in diameter, that each package either one or two of the four genome segments (Fig. 4(a)). Whereas most dsRNA virus capsids are based on dimers of a single protein, RnQV1 has a single-shelled T=1 capsid formed by 60 P2 and P4 protein heterodimers (Fig. 4(b)). P2 and P4 of RnQV1 strain W1118 remain nearly intact, but in strain W1075 both proteins are cleaved into discrete polypeptides, apparently without altering capsid structural integrity. The atomic structure of the RnQV1 W1118 capsid at 3.7 Å resolution shows that P2–P4

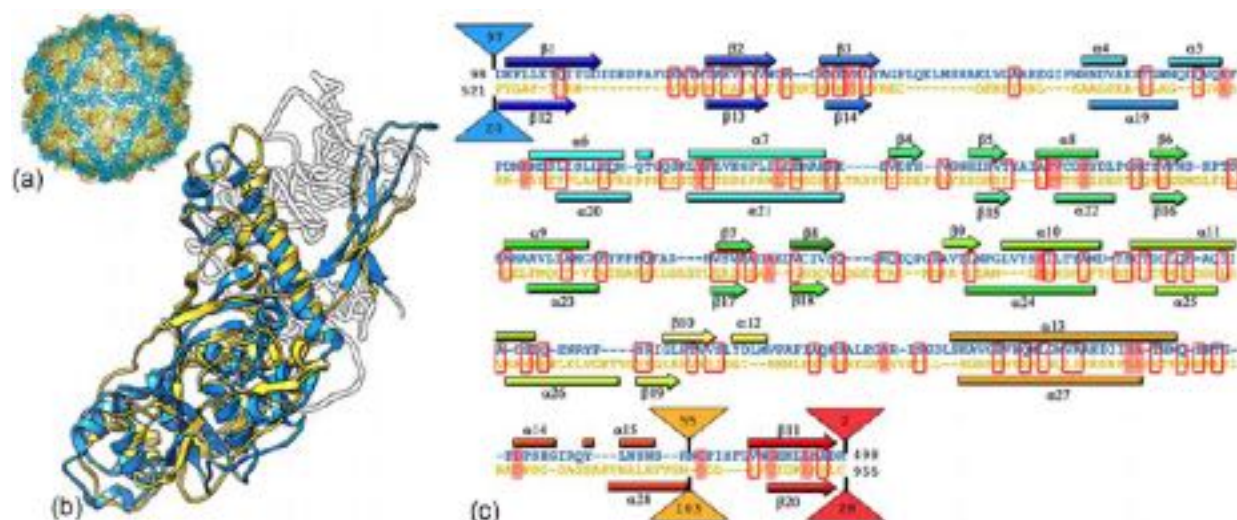


Fig. 3 *PcV* capsid protein is a structural duplication. (a) Atomic model of the *PcV* capsid viewed along a twofold axis (color code as in Fig. 2(c)). (b) Superimposed A and B domains (white segments indicate non-superimposed regions for both domains). (c) Sequence alignment of domains A (blue) and B (yellow) resulting from Dali structural alignment. α -helices (rectangles) and β -strands (arrows) are rainbow-colored from blue (N terminus) to red (C terminus) for each domain. Triangles represent non-aligned segments (sizes indicated): the orange triangle indicates the single “hotspot” on the outer capsid surface. Strictly conserved residues are on a red background and partially conserved residues are in a red rectangle. Adapted from Luque, D., Gómez-Blanco, J., Garriga, D., *et al.*, 2014. Cryo-EM near-atomic structure of a dsrna fungal virus shows ancient structural motifs preserved in the dsRNA viral lineage. *Proceedings of the National Academy of Sciences of the United States of America* 111 (21), 7641–7646.

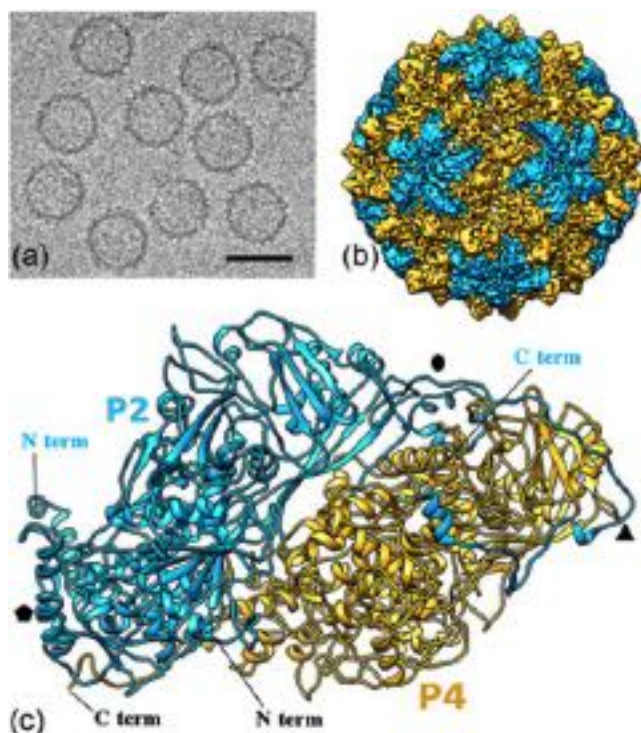


Fig. 4 *RnQV1* $T=1$ capsid cryo-EM-based structure. (a) Cryo-EM image of empty *RnQV1* particles. Bar, 500 Å (b) $T=1$ capsid of *RnQV1* viewed along a twofold axis of icosahedral symmetry, showing P2 (blue) and P4 (yellow). (c) Top view of the atomic models of P2 (blue) and P4 (yellow) (PDB accession number 5nd1).

heterodimers are organized into a quaternary structure similar to that of the homodimers of chrysoviruses and totiviruses (Fig. 4(c)). Although the *RnQV1* capsid, and that of *PcV*, is an exception to the rule that all dsRNA viruses have a $T=1$ capsid with a CP homodimer as the asymmetric unit, it follows the architectural principle that a 120-subunit capsid is a conserved assembly that supports dsRNA replication and organization.

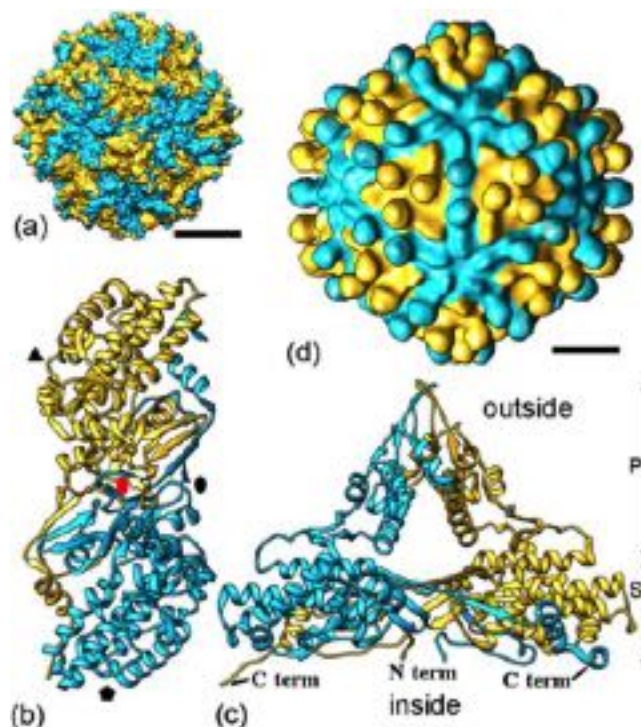


Fig. 5 *PsV-F* and *RnMB1* $T=1$ capsids. (A) X-ray-based structure of the $T=1$ capsid of *PsV-F* viewed along a twofold axis of icosahedral symmetry, showing the CP subunits A (blue) and B (yellow). Bar, 100 Å. (b) Top view of the atomic model of a *PsV-F* CP dimer (3es5). Symbols indicate icosahedral symmetry axes; red oval indicates a local twofold symmetry axis. (c) Side view of a *PsV-F* CP dimer. The protruding arch (P) and shell (S) domains are indicated. (d) Cryo-EM $T=1$ capsid of *RnMB1* viewed along a twofold axis of icosahedral symmetry, showing the rough boundaries of subunits A (blue) and B (yellow). Bar, 100 Å.

Despite their low sequence similarity, the superimposition of P2 and P4 revealed them to have a common α -helical domain. As described for the *PcV* capsid protein, P2 and P4 appear to have also acquired new functions through the insertion of complex domains at preferential insertion sites on the capsid outer surface. These domains are probably related to enzyme activity. The P2 insertion has a fold similar to that of gelsolin and profilin, two actin-binding proteins with a function in cytoskeleton metabolism, whereas the P4 insertion suggests a protease activity involved in cleavage of the P2 383-residue C-terminal region (absent in the mature viral particle). This P2 C-terminal segment might represent an external scaffolding domain.

Partitiviruses

Members of the family *Partitiviridae* have bisegmented, 3.1–4.8 kbp-long genomes. Each segment is encapsidated separately in a similar virus particle. dsRNA1 encodes RdRp (1 copy per particle), while dsRNA2 encodes the CP. The partitiviruses that infect fungi are grouped into three genera: *Alpha-partitivirus*, *Beta-partitivirus*, and *Gamma-partitivirus*. Alpha- and beta-partitiviruses infect plants or filamentous fungi, whereas gamma-partitiviruses infect only the latter. In general, partitivirus infections are largely symptomless.

Four fungal partitivirus structures have been resolved by 3D cryo-EM, including those of the gamma-partitiviruses *Penicillium stoloniferum* virus S (*PsV-S*) and *Penicillium stoloniferum* virus F (*PsV-F*) (by X-ray crystallography at 3.3 Å resolution), and of the beta-partitiviruses *Fusarium poae* virus 1 (*FpV1*) and *Sclerotinia sclerotiorum* partitivirus 1 (*SsPV1*).

The single-layered, 120-subunit capsids of these viruses are 35–42 nm in diameter and distinctive in having “arch-like” surface features that protrude above the continuous capsid shell (Fig. 5(a)–(c)). These $T=1$ capsids have a different quaternary organization, their CP dimer having almost perfect local twofold symmetry (Fig. 5(b)). The quasi-symmetric CP dimer is stabilized by domain swapping within the shell region of the A and B subunits, as well as by intradimeric interactions between equivalent protruding arch domains on the particle surface (Fig. 5(c)). A similar organization has been found in a picobirnavirus, a bisegmented dsRNA virus that infects humans and other vertebrates. Based on their capsid organization, partiti- and picobirnaviruses are proposed to be assembled from dimers of CP dimers (i.e., tetramers) as intermediates. In contrast, the proposed assembly pathway for the 120-subunit capsids of *Toti-* and *Reoviridae* members is based on pentamers of CP dimers (i.e., decamers).

Megabirnaviruses

Rosellinia necatrix megabirnavirus 1 (*RnMBV1*) is the exemplar of the type species of the genus *Megabirnavirus* in the family *Megabirnaviridae*. The *RnMBV1* genome has two dsRNA segments, separately encapsidated in isometric particles of ~50 nm

diameter. dsRNA1 (8.9 kbp) encompasses two partially overlapping ORF's; ORF1 encodes the CP (135 kDa) and ORF2 codes for the RdRp, which is expressed as a fusion protein with CP (250 kDa). dsRNA2 (7.2 kbp) has two non-overlapping ORF's that encode proteins with unknown functions. RnMBV1 causes severe reduction of both mycelial growth of *Rosellinia necatrix* and fungal virulence to plant hosts, and thus has strong potential for virocontrol of white root rot.

RnMBV1 structure has been analyzed by 3D cryo-electron microscopy (Fig. 5(d)). Similar to ScV-L-A, the outer capsid surface of RnMBV1 is rough with 120 spherical large protrusions on the virus surface (~ 50 Å wide). The role of these protruding domains is unknown.

Evolutionary Relationships Based on Structural Comparisons

Structural comparisons of CPs are used to establish relatedness when sequence conservation is limited, and have detected relationships among viruses that infect organisms that, in evolutionary terms, are widely separated. Icosahedral viruses are grouped into four lineages: the dsDNA viruses with an upright double β -barrel CP (the prototypes are phage PRD1 and adenoviruses), the head-tailed phages and herpesviruses that share the Hong Kong 97 (HK97)-like CP fold (also termed the Johnson fold), the picornavirus-like superfamily with a single β -barrel as the CP fold, and the dsRNA or bluetongue virus (BTV)-like viruses. The PRD1- and HK97-like lineages include archaea-, bacteria-, and eukaryote-infecting viruses, suggesting that their last common ancestral hosts were infected by the progenitors of the current viral lineages before the host organisms diverged.

The similarity of the A and B α -helical domains of PcV CP, which have many well-matching secondary structural elements, indicate a common fold in both domains (Fig. 3). It is noteworthy that domain duplication has recently been detected in the capsid protein of other dsRNA fungal virus, botybirnaviruses. Gene duplication (or joined folds) has been a recurrent evolutionary event in other viral lineages. The conserved ~ 350 residue-long PcV fold is also preserved in the Gag of ScV-L-A (Fig. 6). This basic α -helical domain shares many secondary structural elements with ScV-L-A Gag, in particular those regions involved in interactions at the icosahedral symmetry axes. The preserved fold in Gag has three peptide insertion sites facing the outer capsid surface, one of which colocalizes with the single-insertion hotspots of the PcV CP domains. This colocalization suggests that these preferential insertion sites are ancient, and provide a means for the acquisition of new functions without altering the structural and functional motifs of the dsRNA virus CP.

P2 and P4 of RnQV1 also have a common fold some 300 residues long, with two preferential insertion sites on the outer surface. Both coincide with the ScV-L-A Gag insertion sites, and one with the single-insertion site of the PcV A and B α -helical CP domains. Notably, the conserved folds of PcV and ScV-L-A CP are similar to the common fold of P2 and P4, indicating that this fold may have evolved from a common ancestral domain of the dsRNA virus lineage (Fig. 6).

Duplication of an ancestral gene for a CP with the BTV-like fold might have resulted in two separate (as in quadriviruses) or covalently joined folds (as in chrysovirus). This event could direct assembly of a $T=1$ capsid with 120 subunits or domains with a dimer as the asymmetric unit, a necessary arrangement for dsRNA replication/transcription. Separate and joined folds are found in the CP of other virus families such as picornaviruses and comoviruses, respectively. Once the 120-subunit capsid was well established, later divergent evolutionary events would have introduced additional changes in each copy, or even the complete removal of one of them, giving a CP that assembles as a dimer of unfused identical monomers. Alternatively, the ancestral CP could have initially acquired dimer assembly ability, followed by gene duplication.

RdRp and dsRNA Organization Within Mycovirus Capsids

Reovirus $T=1$ cores have 10–12 RdRp complexes per virion, around which the dsRNA is densely coiled. RdRp complexes are non-covalently anchored to the capsid inner surface near the icosahedral five-fold axes, as presumably they are in mycoreoviruses. In addition to RdRp molecules, reovirus replicase complexes include a few minor core proteins with ATPase and/or RNA binding activities. For members of the *Toti*-, *Chryso-Partiti*-, and *Quadriviridae* families, the RdRp molecules are incorporated into 1 or 2 copies per particle, and show more variability than reovirus. For chryso-, partiti- and quadriviruses, the RdRp is expressed as a physically separate protein from a discrete genome segment, and is incorporated into virions via non-covalent interactions with the capsid and/or genome. The same is true for victoriviruses such as HsV190S, except that the RdRp is expressed from the single genome segment of those viruses. For totiviruses such as ScV-L-A, in contrast, the RdRp is expressed as a C-terminal fusion product with the CP. As a result, in ScV-L-A, the 1 or 2 RdRp domains per virion are covalently tethered to the capsid via the fused CP domain, which occupy 1 or 2 subunit positions in the capsid.

The mycovirus $T=1$ capsid wall is perforated by many channels, but none is large enough to pass an A-form 23-Å-diameter genomic dsRNA. Thus the capsid functions as a molecular sieve. Whereas the largest pores (15–20 Å diameter and usually located near the fivefold axis) would allow the passage of nascent mRNA into the host cytoplasm, the smallest holes (5–10 Å in diameter and usually located at the three-fold axis) could be used for nucleotide substrate or pyrophosphate byproduct diffusion. In non-transcribing $T=1$ capsids, the pores are very narrow, but the N- or C-termini or the side chains of residues that face the channel wall might adopt alternative conformations to allow the exit of viral transcripts.

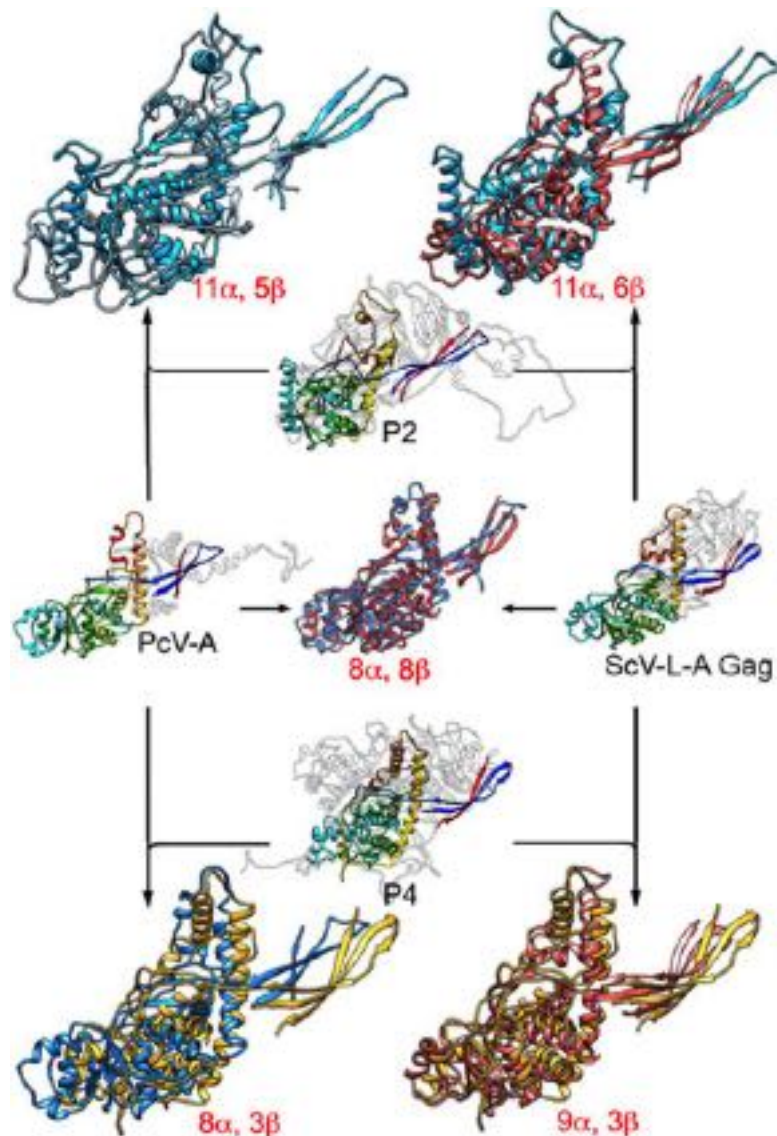


Fig. 6 Structural homology of mycovirus *T=1* CP. The PcV CP A domain (PcV-A, left, center) was structurally aligned with ScV-L-A Gag (right, center), and P2 (top, center) and P4 (bottom, center) with PcV-A and ScV-L-A Gag. Center rainbow-colored structures indicate conserved secondary structure elements within the dsRNA viruses. PcV-A is aligned with ScV-L-A Gag (blue and pink, center). P2 is aligned with PcV-A (blue and light blue, top left) and with ScV-L-A Gag (blue and pink, top right). P4 is aligned with PcV-A (yellow and blue, bottom left) and P4 with ScV-L-A Gag (yellow and pink). Total numbers of secondary structural elements with close relative spatial locations are indicated. Reproduced from Luque, D., Mata, C.P., Nobuhiro, S., Ghabrial, S.A., Castón, J.R., 2018. Capsid structure of dsRNA fungal viruses. *Viruses* 10 (9), 481.

With the exception of totiviruses (which have a single genomic segment), many fungal dsRNA viruses, including chryso-, partiti-, and quadriviruses, have multisegmented dsRNA genomes. In addition, the multisegmented viruses appear to be multiparticulate, i.e., segments are encapsidated separately. Fungal dsRNA viruses have spacious capsids compared to the inner cores of complex eukaryotic dsRNA viruses (Table 1). Whereas reoviruses have 9–12 dsRNA genomic segments packed into liquid crystalline arrays at high density (~ 40 bp/100 nm³ – a spacing between dsRNA strands of 25–30 Å), fungal virus capsids (including ScV-L-A, PcV, PsV-F and RnQV1) contain a single, loosely packed dsRNA molecule (~ 20 bp/100 nm³ – an interstrand spacing of ~ 40 –45 Å). In reoviruses, individual genome segments must be transported through the active sites of the RdRp complexes at the 5-fold axes, and template motion could be a limiting factor. ScV-L-A is a simplified version of these viruses, with a single-segment genome. The looser packing of the dsRNA would probably improve template motion in the more spacious transcriptional and replicative active particles, minimizing electrostatic repulsion between dsRNA strands.

Most mycovirus *T=1* capsids are negatively charged on their inner surface, a feature common to many such capsids of dsRNA viruses. This might facilitate the movement of template and/or product RNA molecules by repulsion, maintaining the RNA layer at ~ 25 Å from the capsid surface (Fig. 7(a), (c), and (d)). The PcV capsid is an exception. It has positively charged regions on the inner

Table 1 Genome packaging densities in fungal dsRNA viruses

Virus family	dsRNA features			Capsid features		
	No segments	Size (kbp)	MW ^a (MDa)	CP (residue)	ϕ^b/ir^c (nm)	dsRNA density (bp/100nm ³) ^d
Reoviridae, MRV	10	~23.5	16	1275	~60/24.5	38
Totiviridae, ScV-L-A	1	~4.6	3.1	680	~43/17	22
Partitiviridae, PsV-S	1 (2) ^e	~1.7 (3.3)	1.2 (2.2)	420	~35/12	23
Chrysoviridae, PcV	1 (4) ^e	~3.2 (12.6)	2.2 (8.6)	982	~40/16	19
Megabirnaviridae, RnMBV1	1 (2) ^e	~8.1 (16.2)	5.5 (11)	1240	~52/19	28
Quadriviridae, RnQV1	1–2 (4)	~4.3 (17.1)	2.9 (11.7)	1356 + 1061	~47/16	25 (50) ^f

^aMW were calculated assuming a mass of 682 Da/bp. HSV dsDNA is assumed to have a B-form.

^bOuter diameter.

^cInner radius.

^dDensities when volume of a perfect sphere is assumed and any other internal components are ignored.

^eFor PsV-S, PcV and RnMBV1 dsRNA, the genome is formed by two, four or two dsRNA molecules, respectively, but a mean value was calculated for one dsRNA molecule/particle in each column.

^f25 if there is one dsRNA molecule/particle; 50 if there are two dsRNA molecules/particle.

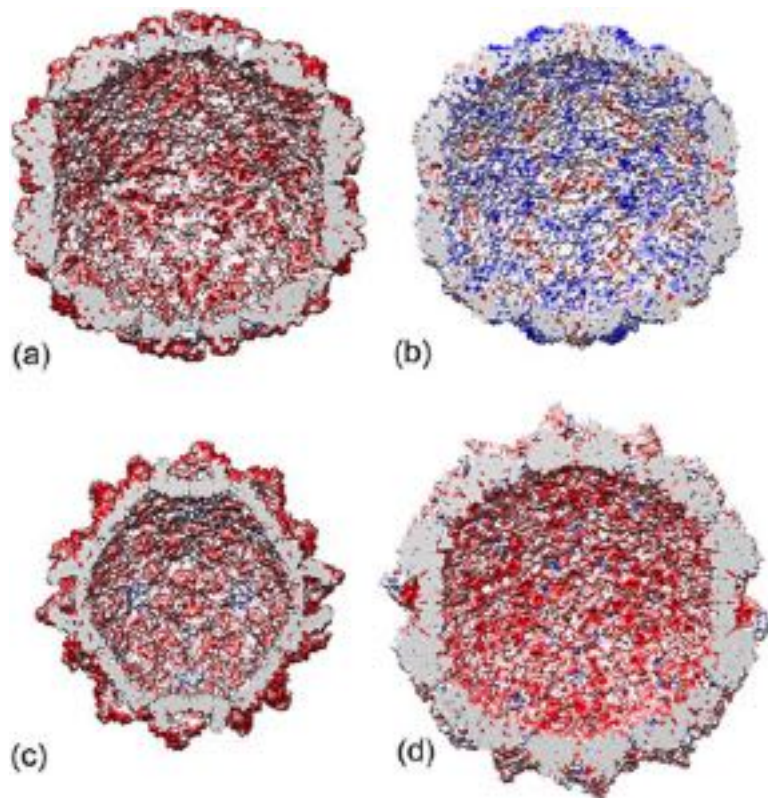


Fig. 7 dsRNA virus $T=1$ capsid inner surfaces with electrostatic potentials. (a)–(d) ScV-L-A (a), PcV (b), PsV-F (c), and RnQV1 (d) capsid inner surfaces viewed along a twofold axis of icosahedral symmetry. The inner surface charge representations of these capsids show the distribution of negative (red) and positive (blue) charges. Note the numerous electropositive areas in PcV. dsRNA and packaged proteins (such as RNA polymerases) were removed computationally.

surface (Fig. 7(b)), and shows numerous interactions with the underlying genome, which is ordered in the outermost RNA layer. As a result, there is almost no space between the latter layer and the inner capsid surface. These contacts have been defined at the atomic level in PcV and PsV-F virions. The lower density of the central region and the associated slight increase in dsRNA mobility might be necessary for maximum RdRp activity in the context of a non-fused RdRp complex.

Comparative analysis of dsRNA packing densities in dsRNA virions have revealed two major tendencies among $T=1$ capsids of dsRNA viruses: (1) those with 9–12 dsRNA segments densely packaged within the same particle and containing 9–12 RdRp complexes, as seen in reoviruses, and (2) those with a single-genomic dsRNA segment with less internal order and one or two copies of the RdRp complex per particle, as seen in mycoviruses.

Concluding Remarks and Future Perspectives

Structural studies of a limited number of fungal viruses have revealed them to hold to the basic concepts of dsRNA viruses, but also to have unexpected features that have contributed to a better understanding of their structure, function, and evolution. dsRNA mycovirus capsids, exemplified by ScV-L-A, PcV, PsV-F, and RnQV1, show structural variations of the same framework optimized for RNA metabolism; they possess 60 asymmetric or symmetric dimers of a single protein (as in ScV-L-A or PsV-F, respectively), dimers of similar domains (as in PcV), or dimers of two different proteins (as in RnQV1). Since mycoviruses are commonly confined to their hosts, their capsids incorporate polypeptides and domains on their outer surfaces for the acquisition of new functions without altering the structure and function of the CP. Such acquisitions would eventually lead to optimal viral-host interactions.

Despite the recent advances made in understanding the structure of dsRNA mycoviruses, many aspects of many fungal viruses remain unknown. Future structural studies should focus on the asymmetric substructures and components of their capsids – such as their RdRp (isolated or packaged inside virions) – and their packaged dsRNA genome.

See also: An Introduction to Fungal Viruses. Chrysovirus (*Chrysoviridae*) - General Features and Chrysovirus-Related Viruses. Fungal Partitiviruses (*Partitiviridae*). Megabirnaviruses (*Megabirnaviridae*). Plant and Protozoal Partitiviruses (*Partitiviridae*). Portal Vertex. Quadriviruses (*Quadriviridae*)

Further Reading

- Abrescia, N.G., Bamford, D.H., Grimes, J.M., Stuart, D.I., 2012. Structure unifies the viral universe. *Annual Review of Biochemistry* 81, 795–822.
- Dunn, S.E., Li, H., Cardone, G., *et al.*, 2013. Three-dimensional structure of victorivirus HvV190S suggests coat proteins in most totiviruses share a conserved core. *PLoS Pathogens* 9, e1003225.
- Ghabrial, S.A., Castón, J.R., Jiang, D., Nibert, M.L., Suzuki, N., 2015. 50-plus years of fungal viruses. *Virology* 479–480, 356–368.
- Luque, D., Gómez-Blanco, J., Garriga, D., *et al.*, 2014. Cryo-EM near-atomic structure of a dsRNA fungal virus shows ancient structural motifs preserved in the dsRNA viral lineage. *Proceedings of the National Academy of Sciences of the United States of America* 111 (21), 7641–7646.
- Luque, D., Mata, C.P., Nobuhiro, S., Ghabrial, S.A., Castón, J.R., 2018. Capsid structure of dsRNA fungal viruses. *Viruses* 10, 481.
- Mata, C.P., Luque, D., Gómez-Blanco, J., *et al.*, 2017. Acquisition of functions on the outer capsid surface during evolution of double-stranded RNA fungal viruses. *PLoS Pathogens* 13, e1006755.
- Mertens, P., 2004. The dsRNA viruses. *Virus Research* 101, 3–13.
- Miyazaki, N., Salaipeth, L., Kanematsu, S., Iwasaki, K., Suzuki, N., 2015. Megabirnavirus structure reveals a putative 120-subunit capsid formed by asymmetrical dimers with distinctive large protrusions. *Journal of General Virology* 96, 2435–2441.
- Naitow, H., Tang, J., Canady, M., Wickner, R.B., Johnson, J.E., 2002. L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nature Structural Biology* 9, 725–728.
- Pan, J., Dong, L., Lin, L., *et al.*, 2009. Atomic structure reveals the unique capsid organization of a dsRNA virus. *Proceedings of the National Academy of Sciences* 106, 4225–4230.
- Patton, J.T., 2008. *Segmented double-stranded RNA viruses. Structure and Molecular Biology*. Norfolk: Caister Academic Press.
- Sato, Y., Castón, J.R., Suzuki, N., 2018. The biological attributes, genome architecture and packaging of diverse multi-component fungal viruses. *Current Opinion in Virology* 33, 55–65.

Ustilago maydis Viruses and Their Killer Toxins

Alexis Williams and Thomas J Smith, The University of Texas Medical Branch, Galveston, TX, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J. Bruenn, *Ustilago Maydis* Viruses, In *Encyclopedia of Virology* (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00519-7.

The Totiviruses

The family *Totiviridae* comprises dsRNA viruses that are classified in five genera; *Totivirus*, *Victorivirus*, *Giardiavirus*, *Leishmanivirus*, and *Trichomonasvirus*. Members of the genera *Giardiavirus*, *Leishmanivirus*, and *Trichomonasvirus* infect protozoa, whereas totiviruses and victoriviruses infect fungi. Only giardiaviruses and trichomonasviruses make capsids that are released from the host to infect other cells. The other members of the family persistently infect their hosts and never leave the cytoplasm.

The word 'toti' is Latin for 'whole' and reflects the fact that the genome is undivided. The positive strand of the genome has two overlapping open reading frames; the coat protein (gag) followed by the polymerase (pol). A fraction of the gag/pol proteins are fused together via a translational frameshift. Even for those genera that do not leave the host, the capsid acts to sequester the dsRNA genome away from the antiviral cellular machinery and form a 'mini nuclei' for genome replication. Early cryo-EM image reconstructions of two representatives of the genus *Totivirus*, *Ustilago maydis* virus H1 (UmV-H1) and *Saccharomyces cerevisiae* virus L-A (ScV-L-A) showed that the capsids have an unusual $T = 1$ icosahedral symmetry where two identical copies of the gag protein are in each icosahedral asymmetric unit. While $T = 2$ is not an acceptable icosahedral triangulation number, these capsids have been called ' $T = 2$ ' to reflect this unusual capsid architecture. These structures were subsequently followed by a crystallographic structure of ScV-L-A, shown in Fig. 1. Structures of several members of the totivirus family have now been determined and are reviewed in the recent paper describing the structure of the giardiavirus, *Giardia lamblia* virus strain Portland I (GLV-I). In spite of low sequence similarity, all members of the family *Totiviridae* share this unusual ' $T = 2$ ' symmetry.

The production of the gag-pol fusion protein plays a critical role in ScV-L-A and UmV-H1 viral replication. In ScV-L-A, there are approximately two copies of the fusion protein per virion where the capsid protein portion forms part of the capsid and the RNA dependent RNA polymerase lies inside the capsid where genome replication occurs. For the toxin encoding species, additional dsRNA segments are separately encapsidated. These are essentially satellite dsRNAs and wholly dependent upon the main portion of the viral genome for their replication and coat protein.

The Killer Phenomena

Several strains of *Ustilago maydis*, a causal agent of corn smut disease, exhibit a 'killer' phenotype that is due to persistent infection by *U. maydis* viruses. These viruses produce potent killer proteins that are secreted by the host. This is a rare example of virus/host symbiosis in that these viruses are dependent upon host survival and, to that end, produce antifungal proteins that kill competing, uninfected strains of *U. maydis*.

Three killer strains of *U. maydis* have been characterized to date: P1, P4 and P6, that secrete KP1, KP4 and KP6 toxins, respectively. Correspondingly, there are three groups of resistant cells where resistance is determined by three independent recessive nuclear genes: p1r, p4r and p6r. The identities of these resistance factors are unknown. These factors are specific for each antifungal protein. Therefore, there are no single resistance alleles that confer simultaneous resistance to all three toxins. All three immunity genes are recessive to their sensitive alleles. Immunity factors are not conferred by the viral genome itself. Resistance to these antifungal proteins was found in ~20% of haploid strains in the natural population. The incidence of strains expressing a toxin is ~1% of the total population and is close to the expected frequency if immunity is a recessive, nuclear encoded factor. This implies that the viruses and their hosts have undergone a co-evolutionary process where the viruses incorporated antifungal proteins, perhaps from other fungi in the distant past, and immunity factors in the host have been selected for that protect it from the effects of the antifungal proteins.

KP4

The KP4 toxin is different from the other *U. maydis* killer proteins. It is a single polypeptide of 105 amino acids and, unlike KP1 and KP6, is not processed by Kex2p protease. There is no significant sequence similarity between KP4, KP6, and other known toxins. While most of the yeast toxins are acidic and the KP6 and KP1 toxins have neutral pI's, KP4 is extremely basic with a pI > 9.0. The structure of KP4 has a unique, compact α/β sandwich structure held together by five disulfide bonds (Fig. 2). α/β proteins are very common and the majority function by interacting with other proteins. KP4 has a single split $\beta\alpha\beta$ motif with a total of seven β -strands

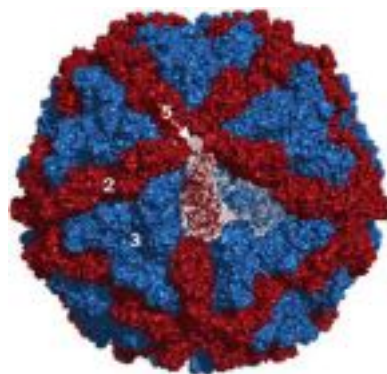


Fig. 1 Structure of the ScVL1 capsid. Subunits A and B are shown here in red and blue, respectively. Ribbon diagrams of one copy of A and B are also shown in this “T = 2” particles. This unusual capsid organization was also found in UmV and other members of the totivirus family. The 2, 3 and 5-fold axes of symmetry are indicated.

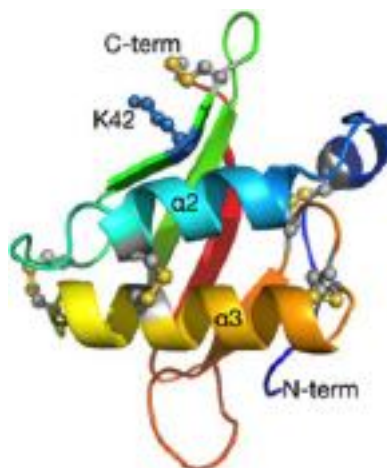


Fig. 2 Atomic structure of KP4. This ribbon diagram is colored red to blue as the chain extends from the N to the C termini. The five disulfide bonds are represented by ball and stick models. Also shown is lysine 42 (K42) which, when mutated to a glutamine or chemically modified abrogates KP4 activity.

(131–137) and three α -helices ($\alpha 1$ – $\alpha 3$). One of the most unusual features of the protein is that the two major helices, $\alpha 2$ and $\alpha 3$, have a left-handed $\beta\alpha\beta$ cross-over conformation. The possible biological function of these unusual arrangements is unclear. It may be to create a more hydrophobic ‘cup-like’ surface on the opposite side of the β -sheet that might be important for protein-protein interactions.

Effects of KP4 on *U. maydis* Cells

The dogma had been that KP4 affects *U. maydis* cells by forming pores in the cell membrane like colicin and some of the small anti-microbial peptides found in animals and yeast. However, from its structure this seemed unlikely. KP4 is only 105 residues in length but contains 5 disulfide bonds and is highly charged. From this, it seemed unlikely that it would form channels on its own or, because of the disulfide bonds, be able to undergo the kinds of conformational changes required to form pores in the membrane. The structure of KP4 is moderately similar to that of scorpion alpha toxins, which are known to target voltage gated sodium channels and inhibit growth in susceptible cells. Therefore, it was proposed KP4 inhibits growth of susceptible fungi through targeting a membrane surface ion channel. To test for this, cell growth was measured in the presence of KP4 with increasing concentrations of K^+ , Na^+ , Mg^{2+} , and Ca^{2+} . Ca^{2+} was very effective at abrogating KP4 inhibition, but, with the exception of a slight effect by K^+ , none of the other metals had any effect even at very high concentrations. This suggested that KP4 might be blocking Ca^{2+} channels. If true, then this further suggests that the block is via reversible protein-protein interactions with the Ca^{2+} channels. To test this hypothesis, sensitive P2 cells were treated for more than 24 h with high concentrations of KP4 and then extensively washed with minimal media. When treated in this way, the cells did recover with about a 6-h delay. Ca^{2+} added to the

wash was very effective at removing all KP4 inhibition. This suggests that Ca^{2+} abrogates KP4 effects by directly blocking toxin binding.

In scorpion toxin AaHII, K58 is a highly reactive lysine and lies near the C-terminus that, when chemically modified, inactivates the toxin. To test whether KP4 and scorpion toxins share this critical feature, KP4 was mutated and chemically modified. There is only one lysine in KP4, K42 (Fig. 2). As with scorpion toxin, when this lysine was chemically modified with acetic anhydride there was a total loss in KP4 activity. This result was confirmed with site directed mutagenesis where K42 was replaced with a glutamine and was also found to block KP4 activity. Finally, as a direct proof that KP4 blocks metal ion import, we demonstrated that KP4 inhibited uptake of $^{45}\text{Ca}^{2+}$ in *U. maydis* cells. Subsequently, experiments were performed with animal cells where electrophysiology is more tractable without the cell wall and the various types of Ca^{2+} channels are well characterized.

KP4 Blocks L-type Voltage Gated Ca^{2+} Channels

The most convincing evidence that KP4 acts on Ca^{2+} channels was that KP4 specifically blocks voltage gated animal channels. KP4 blocked $\text{Ca}_v1.2$, but had no effect on $\text{Ca}_v2.1$ or $\text{Ca}_v2.3$ channels. KP4 specifically blocks L-type Ca^{2+} channels with a weak voltage dependence to the block. In a manner akin to the calciseptine peptide block of $\text{Ca}_v1.2$, KP4 does not affect the voltage-dependence of activation but slightly shifts the voltage-dependence of inactivation of the channel. KP4 activity against both mammalian and fungal cells was blocked by chemical modification of the single lysine residue – akin to what was found with scorpion toxin. Further, as was observed in fungi, the inhibition of the animal $\text{Ca}_v1.2$ channel was abrogated by the addition of Ca^{2+} . Interestingly, we suggested at the time that calcium channel protein (Cch1) was the likely target for KP4 since Cch1 showed homology to the mammalian voltage gated L-type Ca^{2+} channels.

Effect of KP4 on Plants

Since KP4 affects fungal and mammalian calcium influx, it was important to ascertain what effect it might have on plants. Unlike in mammalian cells, calcium ion channels are not as selective in plants when it comes to transporting divalent and monovalent charged ions. A plant homologue of L-type mammalian channels is the *rca*/VDCC2 channel. This channel responds to changes in membrane potential similar to the range of L-type calcium channels, is sensitive to inhibition by dihydropyridine (DHP) derivatives, and has similar calcium ion conductance as their mammalian counterparts.

To measure possible effects of KP4 on plant calcium channels, a non-intrusive and real-time method was chosen to monitor the effects of these antifungal proteins on tip-growing root hair cells. A transgenic line of *A. thaliana* expressing an enhanced yellow fluorescent protein/Rab GTPase fusion protein, EYFP-RabA4b, was used to monitor calcium gradients. RabA4b is involved in a vesicular transport and is localized to the tips of growing root hairs but disappears in mature root hair cells that have stopped expanding. As in fungal hyphae, the root hair tip growth is associated with an apex-high cytosolic free calcium gradient generated by a local influx of calcium at the tip. This calcium gradient was shown to be crucial for RabA4b localization since disruption of the calcium gradient causes a concomitant abrogation of the RabA4b gradient.

The addition of KP4 to growing root hairs causes rapid dissipation of the EYFP-RabA4b gradient at the apical tip and a concomitant cessation of root hair elongation. In all cases, the EYFP-RabA4b gradient was reestablished and root hair extension resumed after KP4 was washed away. Consistent with KP4 blocking a calcium influx channel, the speed of this KP4 effect, and subsequent removal, suggests it targets the outer membrane where it can rapidly bind and cause its effects. Consistent with these results, growth of whole roots was also inhibited by the exogenous application of KP4 and this effect is also blocked by the addition of exogenous calcium. Therefore, KP4 is targeting some conserved portion of calcium channels that are conserved among three kingdoms.

Possible Application of KP4: Fungal Resistance in Plants

Maize smut is a global disease responsible for extensive agricultural losses. In spite of the apparent inhibitory effect of adding KP4 exogenously to *Arabidopsis* root hairs, we created several lines of transgenic maize to see whether constitutive expression of KP4 would confer resistance to *U. maydis* infection. To target KP4 to the extracellular space of transgenic maize, a monocot codon-optimized KP4 gene containing the secretory signal peptide sequence of a plant defensin MsDef1 was placed under the control of a constitutive maize *Ubi1* promoter. KP4 was expressed to high levels *in-planta* as per ELISA and killing assays without any apparent impact on plant development. The fact that we did not see any phenotypic effect of KP4 on root development could be due to less sensitivity of transgenic maize roots to KP4, possible adaptation of the plants to KP4 expression, or that the local concentration of KP4 around the roots is significantly lower than in the experiments described above.

Since transgenic maize lines secreted bioactive KP4 protein, we next determined the ability of this protein to protect transgenic lines from the corn smut disease. Transgenic maize lines expressing KP4 were made and compared to control lines. Seven-day-old seedlings were inoculated with a mixture of the wild-type *U. maydis* KP4-sensitive strains. Both strains generated galls in

nontransgenic maize plants. When the transgenic plants were challenged with *U. maydis* by stem or ear inoculation, they exhibited robust resistance in a dose-dependent manner. These results demonstrate that this family of naturally expressed antifungal proteins holds promise in identifying new targets for antifungal agents and a novel means to protect crops against fungal infections. While KP4 does affect fungi other than *U. maydis* (e.g., *Fusarium graminearum* and *Aspergillus fumigatus* (Smith and Shah, unpublished results)), there has to be strong activity against pathogens of commercial importance. Future studies on the structure/function of KP4 may improve activity against these more commercially important pathogens.

Evolutionary Origin of KP4

Recently, a large number of KP4-like genes/proteins were found in plants and fungi other than *U. maydis*, including the secretome of *Fusarium graminearum*. From sequence analysis, it appears that KP4 likely evolved in a progenitor fungus, possibly a Sordariomycete, and then moved by horizontal gene transfer to other organisms including *U. maydis*. However, it was not clear what function these KP4-like proteins might serve in these pathogenic fungi. Recent studies demonstrated that three KP4-like genes in *F. graminearum* are up-regulated as the fungal infection progresses in wheat. However, it is not clear whether this is causation or just correlation. Interestingly, calcium signaling plays a major role in a wide range of physiological responses in the plant upon fungal infection including cytoskeletal and reactive oxygen species response. It is tempting to speculate that perhaps the original fungus expressing KP4 used it to dampen plant cellular defenses against infection. Consistent with this is the fact that the effects of KP4 on *U. maydis* are abrogated by exogenously adding cAMP, an important secondary messenger for calcium signaling, to the media. It may be that, unlike the channel forming toxins, KP4 evolved to disrupt cytoplasmic signaling pathways regulated by cytoplasmic calcium gradients.

KP6

KP6 is translated as a single polypeptide chain and processed by Kex2p during export through Golgi apparatus to form α and β subunits by the elimination of a 31-residue linker region. There is relatively little known about how KP6 affects *U. maydis* cells. Previous studies have eliminated possible chitinase or protease activity and it is clear that it acts in a manner wholly different to that of KP4. While 100 mM CaCl_2 completely blocks KP4 inhibition and as little as 20 mM CaCl_2 is sufficient to cause significant abrogation, none of the metals had any significant effect on the KP6 killing of *U. maydis* (Smith, unpublished data). The only published data on the mode of action of KP6 suggested that it might be having a strong effect on the membrane integrity or osmotic regulation.

KP6 needs both subunits for activity. Both subunits can be expressed separately in immune *U. maydis* cells and become active when mixed together. Sensitive *U. maydis* cells are able to produce KP6 α but not KP6 β . This suggests that KP6 β added to the extracellular media is not toxic but is when expressed intracellularly. Sensitive *U. maydis* cells can be treated with KP6 α alone without any evidence of antifungal affects. However, if the sensitive cells are killed when treated with KP6 α alone, washed, and then KP6 β is added. However, if sensitive cells are treated with KP6 β , washed, and then KP6 α is added, there is no antifungal activity. Together, this suggests that KP6 α may bring KP6 β into the cell for killing.

The Atomic Structure of KP6

KP6 α forms a single domain structure that has an overall ellipsoid shape and also belongs to the α/β -sandwich family (Fig. 3). The tertiary structure consists of a four-stranded antiparallel β -sheet, a pair of antiparallel α -helices, a short strand along one edge of the sheet, and a short N-terminal helix. Although the fold is reminiscent of toxins of similar size, the topology of KP6 α is distinctly different in that the α/β -sandwich motif has two right-handed $\beta\alpha\beta$ split crossovers. Interestingly, the two KP6 subunits are nearly identical (Fig. 3). The RMSD of the 3D alignment using Chimera was 1.5 Å using 74 of the residues in each of the subunits. Within each subunit, the N and C termini are in close proximity. However, in the $\alpha\beta$ heterodimer, the C terminus of KP6 α is on the opposite side of the N-terminus of KP6 β (Fig. 3). From modeling studies of what the protoxin structure might look like, it seems likely that the $\alpha\beta$ heterodimer undergoes a major structural transition, perhaps activation, during Kex2p processing and export.

This N-terminal region is where the majority of the differences lie between KP6 α and KP6 β (Fig. 3). In KP6 α , there is an amphipathic α -helix at the N-terminus whereas in KP6 β , the N-terminus is shorter by five residues and is entirely a coil structure with the first three residues disordered. In KP6 β , the $\beta 3\text{-}\alpha 3$ loop is much longer than that in KP6 α and lies in the general region of this N-terminal α -helix in KP6 α . The only other major difference is that the $\alpha 2\text{-}\beta 2$ loop is longer and partially disordered in KP6 β .

The α -helix at the N-terminus of the α -subunit is rather unusual in that a disulfide bond holds it in a position that presents a very hydrophobic surface to the aqueous environment rather than burying it within the protein (arrows in Fig. 3). Interestingly, in the structure of KP6 α alone, this hydrophobic patch forms trimeric interactions in the crystal lattice. Clearly there is strong pressure to bury this helix, away from the aqueous environment. It seems likely that this hydrophobic region plays an important role in targeting the fungal cell, perhaps the cell membrane or a membrane associated protein.

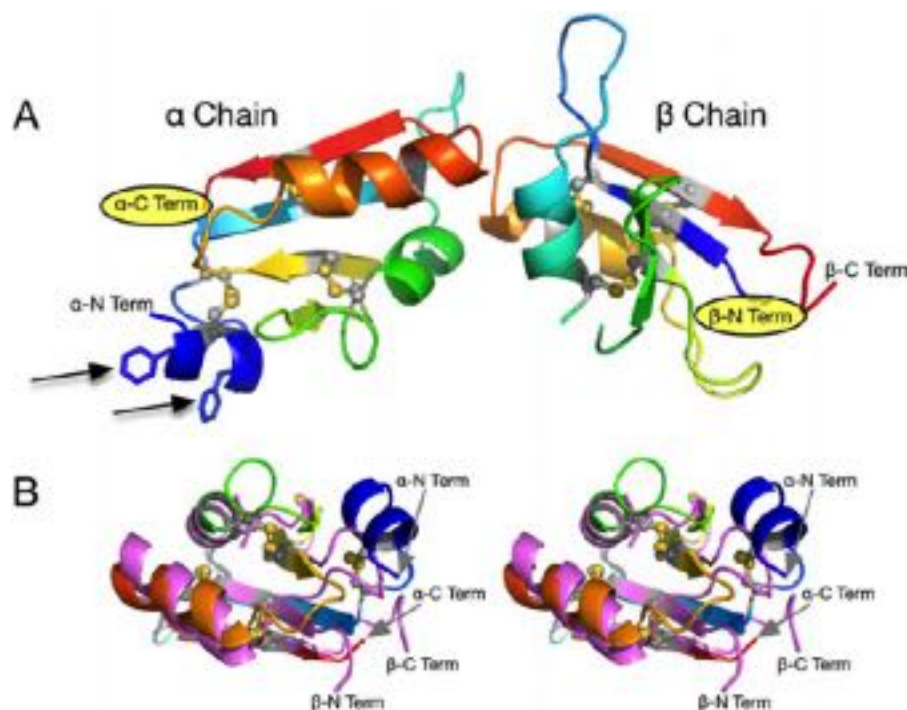


Fig. 3 Atomic structure of KP6. Top: Shown here are the atomic structures of the α and β chains of KP6. The colors of the models change from blue to red as the chains extend from the N to the C termini. The disulfide bonds are represented as ball and stick models. Note the hydrophobic, exposed, phenylalanine residues at the N-terminal helix of the α chain. Bottom: Stereo diagram showing the high degree of structural homology between the α and β chains of KP6. The α chain is colored as a rainbow as in the top figure while the β chain is mauve colored. Note that the core structures of the two subunits are nearly identical and most of the differences are found in the N-termini.

KP1

Relatively little is known about the structure and mode of action of KP1. KP1 is secreted by the P1 *U. maydis* strain, is 12.9 kDa in size, is not glycosylated, and has a basic pI of 8.0. Similar to KP6, KP1 is produced as a pre-protoxin and then processed by Kex2p cleavage to produce two polypeptides, KP1 α and KP1 β . However, unlike KP6, that requires both polypeptides for killing, KP1 only requires the β subunit. Just as KP4 and both KP6 polypeptides, KP1 contains six cysteine residues, all of which are probably connected via disulfide bonds. To date, the structure, the cellular receptor, the mechanism of killing, and the ED₅₀ of KP1 remain unknown.

Comparison of the Killer Proteins

As discussed above, while KP4 and KP6 are both expressed by *U. maydis* viruses, they share no homology with regard to structure or function. For comparison, Fig. 4 shows schematic diagrams for KP1, KP4, KP6, and yeast toxin, K1. From the structures of KP4 and KP6, the disulfide bond pattern is represented by grey lines while the locations of cysteine residues are represented by yellow balls in the KP1 and K1 sequences. At the bottom are the results of transmembrane helical predictions based on the various sequences.

The general organization and sizes of KP6, KP1, and K1 pre-protoxins appear comparable (Fig. 4). However, there is no significant sequence similarity among any of them. Using a transmembrane prediction algorithm, TMHMM2, a very interesting pattern appears among the toxins. As expected, all four toxins are predicted to have transmembrane helices in the N-terminal export signal. For KP1, KP4, and KP6, this is the only part of the protein predicted to have a transmembrane helix. In contrast, there are predicted to be two additional transmembrane helices in the α subunit of K1. This is consistent with the results suggesting that the α subunit of K1 affects the membrane integrity of the target cell. While not shown in the figure, those helices are also predicted to be in the *S. cerevisiae* toxins K2 and K28. This suggests that the *S. cerevisiae* toxins interact with the target in a homologous manner with each other, while the *U. maydis* toxins are quite different among themselves as well as the *S. cerevisiae* toxins.

The other significant difference among the toxins are the disulfide bond patterns and locations of cysteine residues. KP4 is the only single subunit toxin and has five disulfide bonds in the 105 residues. Similarly, both subunits of KP6 are heavily stabilized by disulfide bonds with four and three disulfide bonds in the α and β subunits, respectively. This is important to deduce mode of action since these disulfide bonds make it impossible for the proteins to undergo large conformational changes that could form pores in the membrane like colicin. Therefore, in solution, these are stable hydrophilic proteins that seem more likely to interact

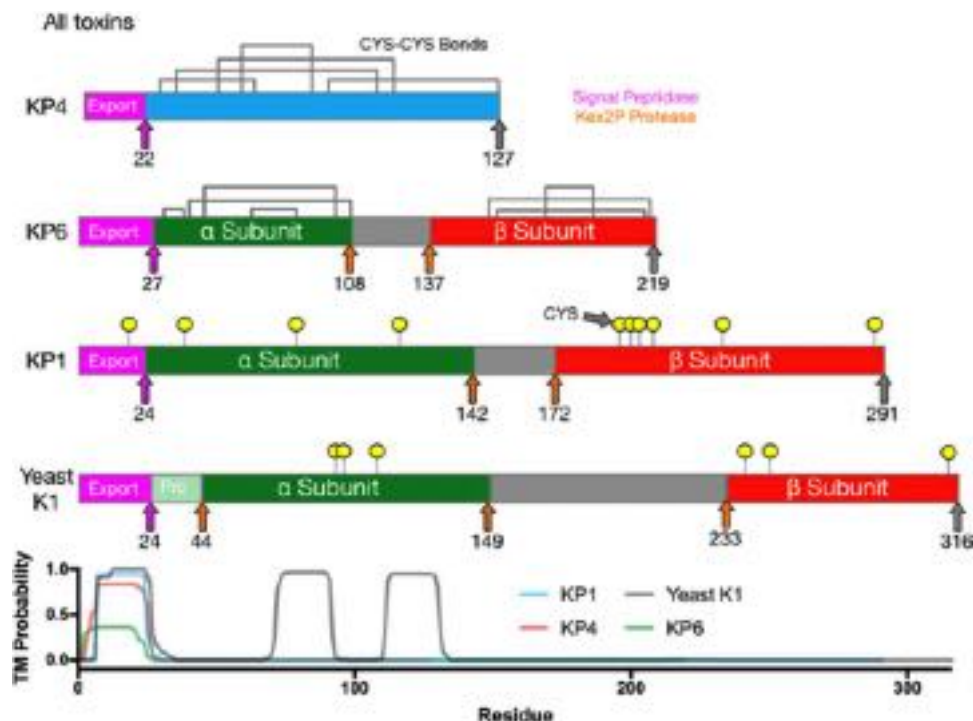


Fig. 4 Similarities and differences among the fungal toxins. In these scaled diagrams of the pre-protoxins, the various regions of the proteins are represented by the various colors. The mauve and orange arrows highlight the signal peptidase and Kex2p protease cleavage sites, respectively. The grey lines in the KP4 and KP6 figures represent the disulfide bonds observed in the crystal structure. For the other diagrams, the locations of cysteine residues are highlighted in yellow. The bottom figure shows the results of the transmembrane helix prediction algorithm, TMHMM2. Note that KP1, KP4, and KP6 only have predicted transmembrane helices in the signal peptide while K1 has two additional predicted helices in the α subunit.

with cell membrane proteins than form channels in the hydrophobic membrane. In the absence of a crystal structure, the disulfide bonding pattern of KP1 is less clear but there are sufficient cysteine residues to make 2–3 disulfide bonds in the two subunits. K1 has three cysteine residues in each subunit with at least one from each forming an inter-subunit disulfide link. Interestingly, the cysteine residues in the α subunit are clustered between the two predicted transmembrane proteins. Therefore, the *S. cerevisiae* toxins appear to be more flexible by virtue of fewer, if any, intra subunit disulfide bonds and more likely to interact with the membrane via the trans membrane helices.

Further Reading

- Allen, A., Chatt, E., Smith, T.J., 2013. The atomic structure of the virally encoded antifungal protein, KP6. *Journal of Molecular Biology* 452, 609–621.
- Allen, A., Islamovic, E., Kaur, J., Gold, S., Smith, T.J., 2011. Transgenic maize plants expressing the Totivirus antifungal protein, KP4, are highly resistant to corn smut. *Plant Biotechnology Journal* 9, 857–864. doi:10.1111/j.1467-7652.2011.00590.x.
- Allen, A., Snyder, A.K., Preuss, M., et al., 2008. Plant defensins and virally encoded fungal toxin KP4 inhibit plant root growth. *Planta* 227, 331–339.
- Brown, D.W., 2011. The KP4 killer protein family. *Current Genomics* 57, 51–62.
- Cheng, R.H., Caston, J.R., Wang, G., et al., 1994. Fungal virus capsids: Cytoplasmic compartments for the replication of double-stranded RNA formed as icosahedral shells of asymmetric Gag dimers. *Journal of Molecular Biology* 244, 255–258.
- Fujimura, T., Ribas, J.C., Makhov, A.M., Wickner, R.B., 1992. Pol of gag-pol fusion protein required for encapsidation of viral RNA of yeast L-A virus. *Nature* 359, 746–749.
- Gage, M.J., Bruenn, J., Fischer, M., Sanders, D., Smith, T.J., 2001. KP4 fungal toxin inhibits growth in *Ustilago maydis* by blocking calcium uptake. *Molecular Microbiology* 41 (4), 775–785.
- Gage, M.J., Rane, S.G., Hockerman, G.H., Smith, T.J., 2002. The virally encoded fungal toxin KP4 specifically blocks L-type voltage-gated calcium channels. *Molecular Pharmacology* 61 (4), 936–944.
- Gu, F., Khimani, A., Rane, S.G., et al., 1995. Structure and function of a virally encoded fungal toxin from *Ustilago maydis*: A fungal and mammalian Ca^{2+} channel inhibitor. *Structure* 3 (8), 805–814.
- Janssen, M.E.W., Takagi, Y., Parent, K.N., et al., 2015. Three-dimensional structure of a protozoal double-stranded RNA virus that infects the enteric pathogen giardia lamblia. *Journal of Virology* 89, 1182–1194.
- Koltin, Y., 1988. The killer system of *Ustilago maydis*: Secreted polypeptides encoded by viruses. In: Koltin, Y., Leibowitz, M. (Eds.), *Viruses of Fungi and Simple Eukaryotes*. New York: Marcel Dekker, pp. 209–242.
- Li, N., Park, C.-M., Erman, M., et al., 1999. The crystal structure of *Ustilago maydis* KP6 killer toxin alpha-subunit. A multimeric assembly with a central pore. *Journal of Biological Chemistry* 274, 20425–20431.

- Lu, S., Edwards, M.C., 2015. Genome-wide analysis of small secreted cysteine-rich proteins identifies candidate effector proteins potentially involved in *Fusarium graminearum*-wheat interactions. *Phytopathology* 106 (2), 166–176. doi:10.1094/PHYTO-09-15-0215-R.
- Martinac, B., Zhu, H., Kubalski, A., *et al.*, 1990. Yeast K1 killer toxin forms ion channels in sensitive yeast spheroplasts and in artificial liposomes. *Proceedings of the National Academy of Sciences of the United States of America* 87, 6228–6232.
- Naitow, H., Tang, J., Canady, M., Wickner, R.B., Johnson, J.E., 2002. L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nature Structural & Molecular Biology* 9, 725–728.
- Park, C.-M., Banerjee, N., Koltin, Y., Bruenn, J.A., 1996. The *Ustilago maydis* virally encoded KP1 killer toxin. *Molecular Microbiology* 20, 957–963.

Vegetative Incompatibility in Filamentous Fungi

Songsong Wu, Daohong Jiang, and Jiatao Xie, Huazhong Agricultural University, Wuhan, China

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Anastomosis Fusion of somatic hyphae followed by exchange of cytoplasmic contents.

Horizontal transmission The transmission of virus, parasite, or other pathogens from one individual to another within the same generation.

Innate immunity Physical and chemical barriers including cells, cytokines, and antiviral proteins that inhibit infection

with little specificity and without generation or adaptation of a protective memory.

NLRs A class of cytosolic nucleotide oligomerization domain (NOD)-like receptors that recognize non-self signals in plants, animals, and fungi.

Vegetative incompatibility A self/non-self recognition system in filamentous fungi that determines the outcome of hyphal fusion between fungal individuals.

Introduction

Self and non-self recognition widely exist among all organisms, which include the immune systems in vertebrates, self-incompatibility during reproduction in plants, mating and vegetative incompatibility (VI) in fungi. Non-self recognition determines the outcome of somatic fusion between different tissues and causes conflicts between individuals that belong to different genetic backgrounds. In filamentous fungi, fungal colony with an interconnected network are formed via hyphal tips extension and branching formation. Different fungal isolates can undergo hyphal fusion to form a heterokaryon, and this process is essential for cytoplasmic communication and sexual reproduction in fungi. If two fungal individuals have one or more different *het/vic* loci, the fusion cells commonly show vacuolization and PCD reaction. Therefore, hyphal contact between two incompatible fungal individuals usually results in a macroscopic phenotype with black-pigmented bands (death cells) at the interface between two contacting fungal isolates, while isolates with the same genetic background can be well integrated and form one colony.

Vegetative compatibility groups (VCGs) are defined as compatible isolates that have the same *het/vic* loci and genetically related within the same species. Isolates from different VCGs differ in colony phenotype and pathogenicity among some fungi, such as *Rhizoctonia solani*. Therefore, VCGs identification is a useful tool to analysis fungal population structure and evaluate the evolution of invasive pathogen fungi, such as *Hymenoscyphus fraxineus* and *Ophiostoma novo-ulmi*. In *Aspergillus flavus*, toxigenic and atoxigenic isolates belong to different VCGs, and atoxigenic VCGs can be used as biocontrol agents to reduce aflatoxin content by increasing the incidence of the atoxigenic isolates among natural population. The application of atoxigenic VCGs supplied an important biocontrol strategy for plant pathogens.

The capable transmission of intracellular parasites among fungal individuals under natural condition is an essential factor for spreading infectious diseases. Mycoviruses as a group of intracellular parasites, are considered to be limited to intracellular transmission, with the exception for *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 (SsHADV-1) which has an extracellular phase. Mycoviruses can spread effectively within a fungal colony or between different strains that belong to the same fungal VCGs. In black *Aspergilli*, *C. parasitica*, and *S. sclerotiorum*, somatic incompatibility can efficiently block virus transfer among vegetatively incompatible individuals. Therefore, VI can effectively reduce the risk of horizontal transmission of mycovirus and its related genotypes among fungal populations.

Heterokaryosis is believed to be controlled by the function of *het* genes in fungi. So far, *het* genes have been cloned in some model species that include eleven *het* loci in *Neurospora crassa*, nine *het* loci in *Podospora anserina*, eight *het* loci in *Aspergillus nidulans* and six *het* loci in *C. parasitica*. Furthermore, the interaction between horizontal transmission of CHV1 (Cryphonectria hypovirus 1) and *vic* systems in *C. parasitica* has been thoroughly studied. Horizontal transmission of virulence attenuating CHV1 is known to be restricted by VI systems. In *C. parasitica*, two super mycovirus donor strains (SD328/SD82) were generated by disrupting four out of six *vic* genes, and a mixture of these two strains can delivery CHV1 to different VCGs both in the laboratory and under natural conditions. In *S. sclerotiorum*, *Sclerotinia sclerotiorum* mycoreovirus 4 (SsMYRV4) could help heterologous viruses to transmission among different VCGs via inhibiting VI-mediated PCD of donor strains. The feature of SsMYRV4 also creates a bridge strain for viruses spreading between different VCGs under natural conditions. These results demonstrated that the biocontrol capability of hypovirulent-associated mycoviruses can be enhanced via modifying the vegetative compatibility type among fungal populations.

In this article, we focus on the process of hyphal fusion between individuals from different VCGs. We analyze the genetic and molecular control of VI reactions. We also examine the ubiquitous immune system of NLRs to analyze the role of HET proteins in processing PCD and mycovirus transmission upon allorecognition in filamentous fungi. These analyses will provide clues for enhancing mycovirus horizontal transmission via block/weaken VI reactions in fungi.

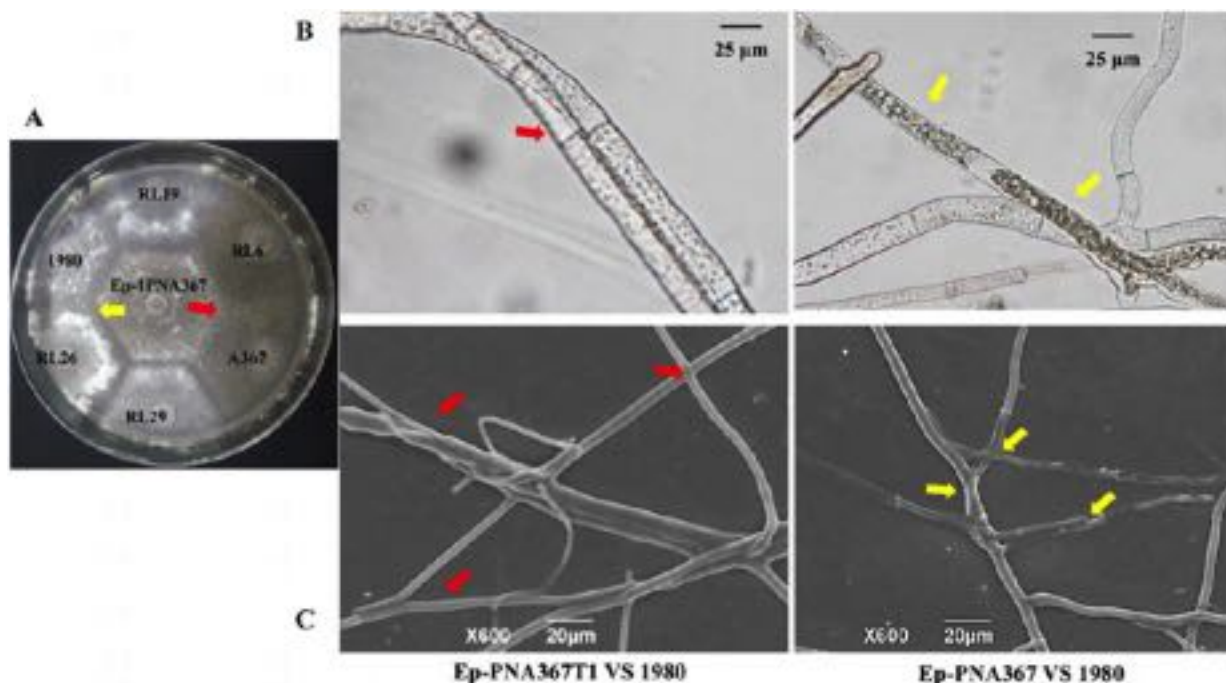


Fig. 1 Cell fusions between *S. sclerotiorum* isolates. (A) Cell fusion between vegetatively compatible and vegetatively incompatible isolates of *S. sclerotiorum*. Strains were cultured on PDA medium supplemented with 75 μ L/L of McCormick's red food coloring for 48 h. Strains Ep-1PNA367 and RL6 belong to the same VCG, while strains RL19, 1980, RL26 and RL29 are from different VCGs. A clear line separates the individual strains that are vegetatively incompatible. (B) Light microscope observation of cell fusion between vegetatively compatible isolates (left panel) and vegetatively incompatible isolates (right panel). (C) SEM observation of the cell fusion (left panel: vegetatively compatible; right panel: incompatible). Strains cultured on PDA plates covered with cellophane membrane. Once hyphal tips of the two strains contact each other, the observation of fusion cells were conducted under microscope.

Characteristics of the Vegetative Incompatibility Reaction

Microscopic and Macroscopic Analyses of Hyphal Fusion

In many organisms, cell fusion plays an essential role in many developmental processes, such as sexual cells fusion and multinucleate cell-based organ formation. In filamentous fungi, during an early stage of hyphal anastomosis, the cell wall synthesizing enzymes and digestive enzymes in secretory vesicles are delivered to the point of contact. After exchange cellular contents of the contacted cells, new septa usually form near to the fusion site. This hyphal anastomosis increases nutrition and water utilization in the fungal colony. While cell fusion from different genetic background often results in the formation of heterokaryon, heterokaryon grows well if fused cells formed from genetically compatible individuals, whereas the fused cells formed from incompatible individuals show vacuolation and PCD, termed VI. The process of VI in some fungal species can be easily observed by the naked eye, when the incompatible pairs of strains grow on agar medium. There usually appears a black-pigmented band at the interface of two contacting fungal isolates (Fig. 1(A), yellow arrow), while there is no clear barrage between the fused vegetatively compatible isolates (Fig. 1(A), red arrow). This vegetative compatibility testing of pair-cultured isolates on agar has been used to analyze VCGs and natural populations in many fungi, including in *S. sclerotiorum*, *P. anserina* and *O. novo-ulmi*. The black-pigmented banding between two contacting fungal isolates is formed by degeneration of the fused cells and neighboring cells. These cells marked as vacuolation with increasing the size of vacuoles and then cytoplasmic granulation. Some contacted cells from different VCGs become shrinks (Fig. 1(B), yellow arrow). Scanning electron microscope (SEM) observation of hyphal fusion between incompatible strains showed that the fusion cells caused cytoplasmic shrinkage (Fig. 1(C), yellow arrow), while the fusion cells from compatible strains grew well (Fig. 1(C), red arrow). Transmission electron microscopy (TEM) observation of the fused cells from compatible pair strains revealed that cell wall fusion with normal vacuoles and new septa formed at the sites of anastomosis. In the incompatible fused cells, the plasma membranes of heterokaryon were disconnected, cytoplasm and vacuoles contents disappeared. Incompatible fuse leads to PCD without hyphal anastomosis. Therefore, the morphological changes of vacuoles and cytoplasm in the fusion cells act as an indicator of VI.

Identification of VCGs

VCGs identification is important to evaluate the genotypic and phenotypic diversity of fungal populations. Multiple techniques have been developed to identify VCGs in filamentous fungi.

Barrage assay

The barrage assay is a convenient method because it does not require additional specific genetic markers, and can be performed directly by pair-culturing of isolates. Potato dextrose agar (PDA) supplemented with McCormick's red food coloring has been applied for VCGs identification in *S. sclerotiorum* (Fig. 1(A)). PDA medium supplemented with bromocresol green pH indicator dye has been used for VCGs tests in *C. parasitica*. The medium supplemented with additional dyes could make the barrage more apparent to the naked eye.

Microscopy

Microscopic analysis of fused cells and neighboring cells to identify VCGs has been applied in some fungal species. Typical features including shrinkage of the plasma membrane and vacuolization of the cytoplasm serve as marker of VI. Mycelia plugs from different VCGs are placed on PDA plates covered with cellophane membrane. Pairs of plugs from the same VCG act as controls. When the hyphae from the two strains contact each other for 3–6 h, we can observe the microstructure of the fused cells under the microscope (Fig. 1(B)).

Visualization of labeled proteins during co-culture

When isolates from the same VCG contact each other, the two isolates can exchange cytoplasmic contents during anastomosis. The contacted cells from different VCGs will undergo PCD and prevent cytoplasmic communication. Therefore, if two tested isolates are labeled with two different fluorescent protein, we can observe the transmission of cytosolic fluorescent proteins during anastomosis; two fluorescent signals in the contacted cells indicate that two strains belong to the same VCG; only one fluorescent signal in the contacted cells at the region of contact indicates that the two strains belong to different VCGs and failed to fuse.

Auxotrophic complementation

Auxotrophic complementation can be used to test for VCGs. Hyphae from two isolates expressing different selective markers are allowed to intertwine at the interface region. If the pair of strains is compatible, the fused cells from the interface region can be able to grow by auxotrophic complementation.

Detecting the PCD of fused cells

In filamentous fungi, the VI reaction usually causes PCD of the fused cells and neighboring cells. We can test for the occurrence of cell death at the contacted area by staining dead cells, such as Evans blue staining. As a feature of early PCD, DNA fragmentation can be observed using TUNEL (terminal deoxynucleotidyl transferase) assays on VCGs identification. These methods for PCD detection are widely used among fungal pathogens.

The Genetics of Vegetative Incompatibility

The outcome of fused cells is dominated by self/nonself-recognition genes. If two contacted isolates have the same recognition genes, the hyphae from the interaction zone will successfully fuse and form a coherent colony. However, when the two isolates have difference at one or more of self-recognition loci, the fused hyphae in the interaction zone usually undergo PCD, manifested as VI. Therefore, the failure of hyphal fusion between VCGs often reflects genetic differences at the *het* or *vic* loci. In *N. crassa* and *P. anserina*, allelic and non-allelic VI systems have been established. In the allelic VI system, co-expression of incompatible alleles at the same locus leads to VI reaction, for instance *het-s*/HET-S in *P. anserina*. In non-allelic VI systems, interactions among unlinked incompatible alleles from different genes results in VI, such as *het-r* and *het-v*-mediated VI in *P. anserina*. Furthermore, the HET proteins that contain a common ~150 amino acid conserved HET domain usually contribute to VI reaction, and HET proteins widely exist in fungal genomes. In *P. anserina*, the proteins with HET domain acted as the mediator of PCD during the VI process. In *N. crassa*, 73 putative HET proteins have been identified via comparative population genomic analysis, but the roles of those HET genes in the process of VI remain to be investigated. Using the same approach, 44 proteins with HET domain were predicted in the genome of *S. sclerotiorum*. To date, the VI systems of three fungal species (*N. crassa*, *P. anserina*, and *C. parasitica*) as model fungi have well been studied.

In *N. crassa*, the locus (*mat-A1* and *mat-a1*) controls mating and sexual development, strains with null mutations in *mat A-1* are both heterokaryon compatible and sterile with *a* mating type strains. However, during vegetative growth, hyphae fusion of opposite mating type leads to growth inhibition and PCD, which is mediated by the tolerant (*tol*) locus. The *tol* gene contains HET domain and leucine-rich repeat regions (LRRs). The mutant *tol* acts as a specific suppressor of heterokaryon incompatibility. Furthermore, interactions between *het* genes (such as *un-24* /*het-6* and *het-c*/*pin-c*) were further confirmed to control VI.

In *P. anserina*, eight VI systems have been characterized; the function of allelic locus *het-s* and the nonallelic *het* loci (*het-c*, *het-e*, and *het-r*) depend on genes encoding HET domain proteins, which have been confirmed to function in triggering PCD. Typical *het* genes contain a HET domain at the N-terminal, a NACHT domain in the central, and a WD40 repeats domain at C-terminal. The NACHT domain contains LRRs and a pyrin domain (PYD). In plants and animals, as recognition of pathogen-associated molecular patterns, NACHT-containing proteins stimulate immune responses and regulate apoptosis. Proteins containing WD40 domain are involved in many life processes, including apoptosis, transcriptional regulation and immune responses. The NACHT-WD40 domain is functionally important in HET-E-triggered PCD reactions in fungi. Furthermore, the *het-s* allele encodes a prion form protein HET-s, it acts as an activation trigger of a cell death execution protein. In *C. parasitica*, six unlinked *vic* loci have been

identified and characterized. Among these VI systems, the function of *vic1*, *vic6*, and *vic7* depends on proteins containing HET domains, and *vic4* encoded protein has a NACHT-WD40 architecture.

The predicted proteins with HET domains are usually not conserved in fungi, thus it is difficult to define *het* genes via its orthologous relationships. Interestingly, heterologous expression of *het-c2* from *N. crassa* can also trigger a VI reaction in *Aspergillus niger*. This result indicated that the *het-c* homolog could function as a *het* gene in a different fungus, although this remains to be explored in different species. The function of un-24 triggering VI-mediated PCD shows the same function in *N. crassa* and its related species, but it does not function as a *het* gene in other species. Therefore, proteins containing a HET domain have diverse functions, and some of these proteins are involved in VI. Exploring the function of these proteins in VI processing will enhance our understanding of the polymorphism of *het* proteins.

The Signaling Pathways of Vegetative Incompatibility

The VI process involves many signaling pathways, including recognition of a non-self-signal, signal transmission, and triggering of PCD. Although the VI systems appear to be complex in different fungi, the cellular reactions to the VI signals might be similar. When two incompatible hyphae contact each other, shrinkage of the plasma membrane and vacuolization are frequently observed in fungi. Interestingly, vegetatively incompatible individuals from *Verticillium albo-atrum* and *Verticillium dahliae* can form heterokaryons via protoplast fusion, not via hyphal fusion. Interspecific and intraspecific heterokaryons formation could be improved by removing the cell wall. Therefore, the outcome of VI is controlled by nuclear and also associated with components affecting the cell wall or cell wall formation. Therefore, cell wall, plasma membranes and vacuoles could play important roles in VI reactions.

Heterotrimeric G proteins play an essential role in the signal-transduction pathways, including pheromone response and mating, pathogenesis, and non-self-recognition in eukaryotes. G proteins consist of three subunits including $G\alpha$, $G\beta$, and $G\gamma$. In *P. anserina*, one α subunit of G proteins (MOD-D) regulates the process of VI. Increasing the cAMP level suppresses the defects in vegetative growth caused by the mod-D1 mutation, while increased cAMP has no influence on the suppressor effect of the mod-D1 mutation during VI reaction, thus, the function of MOD-D in VI reactions is independent of the cAMP pathway. In *S. sclerotiorum*, three homologs of $G\alpha$, one of $G\beta$, and one of $G\gamma$ have been identified as G subunit proteins. With the exception of one $G\alpha$ subunit gene, the expression levels of the other four G protein subunit genes were significantly up-regulated during hyphal contact between incompatible individuals, while there is no expression change during anastomosis. Therefore, G proteins as multifunctional signaling proteins are also involved in transduction of VI signals in fungi.

In all eukaryotes, Calcium (Ca^{2+}), a ubiquitous second messenger, is involved in multiple biological processes. Ca^{2+} levels play an essential role during germling and hyphal fusion. Increasing intracellular Ca^{2+} levels can commonly trigger plasma membrane fusion of secretory vesicles. Furthermore, the process of activation of MAP kinases usually occurs at cell membrane. In *N. crassa*, membrane protein HAM-11 is involved in regulation downstream of the MAK-2 pathway, but the $\Delta ham-11$ mutant fails to undergo self-fusion. HAM-11 protein activates the MAPK signaling complex to achieve cell fusion. Another WD40 repeats-contained protein HAM-5 as a MAP kinase shows physically interaction with MEK-2 and MAK-2. HAM-5 oscillates with MAK-2 during hyphal fusion and controls the disassembly of the MAK-2 MAPK complex, but the $\Delta ham-5$ mutant fails to undergo self-fusion. Therefore, HAM-5 is important for the activation of MAK-2 cascade during VI reaction. MAK-2 pathway is also essential for transmitting the VI-mediated PCD signal to downstream during incompatible fungi fusion (Fig. 2).

In *N. crassa*, HET-C is a transmembrane protein, and the initial self/nonself recognition between two strains occurs in the cell membrane. Another protein HET-E is also related to endomembrane and cytoplasmic. In *P. anserina*, [HET-s] prion and HET-S (a soluble protein) that coexist in one cell will lead to the loss of membrane integrity and cell death. Thus, alteration of the stability of the cell membrane by HET proteins is an initial response after hypha fusion finally lead to VI-associated cell death reactions. In *Rosellinia necatrix*, cytological analysis shows that the VI-mediated PCD is a vacuole-mediated process. The integrity of vacuoles might act as an important role in the switch between compatible and incompatible reactions. In *P. anserina*, among *idi* genes (induced during incompatibility), *idi-1* is associated with the response to nutrient starvation and the VI process. IDI-1 is a cell wall protein localized at the septum, and rapamycin treatment could induce overexpression of *idi-1*. This treatment also shows hallmarks of VI, including increased septation, vacuolization and autophagy. Thus, IDI-1 could be involved in the TOR kinase pathway during VI reaction. *idi-7* is an ortholog of *Atg8* in *Saccharomyces cerevisiae*, which is involved in autophagy. Another *idi* gene, *pspA* (a subtilisin-like serine protease), has an ortholog in *S. cerevisiae*, and encode a vacuolar protease involved in autophagy. These observations suggested that VI reaction may be similar to autophagy, and this process is related to signal protein transportation from cell membrane to vacuole.

Most of the molecularly characterized proteins involved in VI have been identified as possessing a HET domain in fungi. In *S. sclerotiorum*, most HET genes are significantly upregulated during VI; while these candidate *het* genes are downregulated when VI reactions are suppressed. In *N. crassa*, a transcription factor encoded by the *vib-1* gene regulates *het* gene expression. Deletion or mutation of *vib-1* suppressed the VI reaction mediated by genetic differences *het* loci. Thus, proteins containing a HET domain are important for downstream functions during VI reactions. According to the transcriptional analysis in *P. anserina* and *N. crassa*, the top 100 up-regulated genes during the VI-mediated PCD reaction involve in cell signaling pathways or autophagy, which indicated that some common factors have the same function during VI reactions in both *P. anserina* and *N. crassa*. In *N. crassa*, many conserved eukaryotic factors are involved in regulation of germling fusion, such as MAP kinase modules, NADPH oxidases (NOX), and calcium-regulated factors. These signaling pathways are confirmed to play important roles in hyphal fusion in fungi (Fig. 2).

domains (NB-ACC, NACHT and AAA), and 3C-terminal domains (ANX, WD and TPR). Some fungal NLRs have enzymatic activities at N-terminal domains, such as Patatin with phospholipase and peptidase activities. These structures indicate that fungal NLRs have numerous functions in addition to their signaling function. As mentioned above, fungal proteins containing HET domains have function in VI-mediated PCD. In *P. anserina*, 10 of 16 characterized *vic* proteins contain a HET domain, and overexpression of a HET domain could initiate a VI-like PCD. The *vic2* and *vic4* loci in *C. parasitica* encode typical NLRs. In *S. sclerotiorum*, two proteins have an NB-ARC domain, two proteins contain a WD domain, and one protein has a NACHT domain among 44 predicted proteins containing a HET-domain. In *N. crassa*, patatin-like phospholipase-1 (PLP-1) with typical NLRs was identified, and the N-terminal phospholipase activity and central NBD are essential for triggering allorecognition and PCD in *P. anserina* and *N. crassa*. These observations indicated that the functions of fungal NLR-like proteins, including some HET proteins, are similar to the NLRs from animals and plants, and contribute to non-self recognition.

In plants and animals, NLR-mediated innate immunity is triggered by infection by various pathogens. Fungal NLRs-mediated VI also function in bacteria-fungi and virus-fungi interactions as a microbe-associated molecular patterns. Transcriptomic analyses of NLR-controlled VI revealed the activation of processes including cell wall modification, secondary metabolite production and autophagy; this is very similar to the responses induced by bacteria in plant or animal. The *vic2* system in *C. parasitica* has been confirmed to reduce virus horizontal transmission via triggering VI-mediated PCD. In *Arabidopsis*, the patatin domain could enhance the resistance to cucumber mosaic virus via triggering cell death, and the patatin domain in the *vic2* has similar function to initiate cell death in *C. parasitica*. Thus, VI in fungi leads to rapid cell death and increase antimicrobial activity, and NLR-controlled VI could reduce virus horizontal transmission during hyphal fusion of incompatible individuals.

Mycovirus Transmission

Mycovirus was first reported in cultivated mushrooms in 1962. With the application of high-throughput sequencing in mycovirus, more and more mycoviruses with different types were found in all four phyla of fungi. For the most mycoviruses, transmission is limited to intracellular transmission (cell to cell), while there has been reported some exceptions that break the restrictions of VI in fungi. SsHADV-1 has an extracellular phase and can infect its sister species of *Sclerotinia* directly via virus particles. Furthermore, some kind of vectors have been identified that facilitates mycovirus transmission. An insect, *Lycoriella ingénue*, can be a transmission vector to facilitates SsHADV-1 spread. A plant virus, cucumber mosaic virus (CMV), can naturally infect fungus *Rhizoctonia solani*, and *R. solani* isolates can also acquire CMV from plants during infection, which suggests that CMV can be transmitted in *R. solani* isolates via plants vector. Most mycoviruses have neither transmission vector nor extracellular phase, and these mycoviruses are limited to the intracellular life cycles, with lateral transmission among fungal populations via hyphal fusion and vertical transmission via producing sexual/asexual spores. In generally, mycoviruses can be spread through asexual spores with a high frequency rate. In *Basidiomycetes* fungi, mycoviruses can frequently enter into basidiospores. However, in ascomycetous fungi, most of mycoviruses could not enter into ascospores during sexual reproduction, exception for some mitoviruses and ourmia-like viruses, which may relate to mitochondrial inheritance. Mycovirus horizontal transmission between fungal isolates is mainly achieved through anastomosis. Therefore, hyphal anastomosis is essential for maintaining the mycovirus proportion in the natural fungal population.

Vegetatively compatible combination within and between individuals occurs with a high frequency, which promotes hyphal network formation and facilitates mycovirus transmission within a colony. However, most hyphal fusion between different VCGs of the same species triggers cell death. VI as NLR-mediated innate immunity in fungi can effectively block mycovirus transmission among fungal populations. The success of CHV1 application to biocontrol chestnut blight in Europe encouraged similar virocontrol of other fungal pathogens with mycoviruses. The same efforts to virocontrol chestnut blight in North America have failed due to the diversity in VCGs of *C. parasitica* isolates. In black *Aspergilli*, horizontal transmission of mycoviruses only occurs among the same VCGs under laboratory conditions. Therefore, horizontal transmission of mycovirus is usually considered to be limited to the strains belonging to the same VCG. Under the balancing selective forces, diverse organisms maintain polymorphisms of genes involved in innate immunity and VI systems. In *C. parasitica*, VI-related genes are under selective forces due to restriction the transmission of debilitation-associated viruses among fungal populations. This suggests that the polymorphisms of VCGs are the way to prevent transmission of viral disease.

The influence of VI systems on virus transmission has been well-studied in *C. parasitica*. Based on the *C. parasitica* strains isolated from a natural population with different *vic* genes, the ability of CHV1 transmission among different isolates shows negatively correlated with the number of different *vic* alleles between them. When donor and recipient strains have one different *vic* allele, the transmission frequency of CHV1 was 0.5, while transmission frequency decreased to 0.13 or lower when VCGs differed by two or more *vic* loci. In *C. parasitica*, there shows a significant negative correlation between PCD and the frequency of CHV1 transmission. In *S. sclerotiorum*, an RNA virus (SsDRV) cannot be transmitted to the strains that belong to different VCGs. Thus, VI is a main barrier for mycovirus horizontal transmission between different VCGs of *S. sclerotiorum*, and reduction of the frequency of VI-mediated PCD can promote virus horizontal transmission. In the CHV1-*C. parasitica* system, five of the six *vic* loci (except *vic4*) have confirmed to restrict mycovirus horizontal transmission among VCGs.

Mycoviruses can be Horizontally Transmitted Among VCGs

As VI is the main factor to restrict horizontal transmission of mycoviruses, the efficiency of mycoviruses transmission can be promoted by modifying VI systems. Based on the function of these five *vic* genes in CHV1 transmission, a super mycovirus donor

strain with disruption of four *vic* genes (*vic1a-2*, *vic3b-1*, *pix6-2*, and *vic7a-2*) has been engineered, a mixture of the two donor strains with *vic2-1* or *vic2-2* enables CHV1 to be transmitted easily under laboratory and natural conditions. Therefore, it could serve as an enhanced hypovirus vector via breaching the restriction of VI. Similarly modified VI systems in other fungi can be applied to facilitate mycoviruses horizontal transmission.

Current research revealed that some of mycoviruses can escape from the restriction of VI systems under the natural or laboratory conditions. In *S. sclerotiorum*, Sclerotinia sclerotiorum deltaflexivirus 2 (SsDFV2) and Sclerotinia sclerotiorum partiti-virus 1 (SsPV1) can be transmitted between different VCGs without the limitations of the VI system, and SsPV1 can even be transmitted to the other *Sclerotinia* species through hyphal contact. Two partitiviruses (HetRV3-ec1 and HetRV4-pa1) have been confirmed to easily transmit between different VCGs.

Potential Approaches to Enhance Mycovirus Transmission Between VCGs

VCGs shows high diversity in many filamentous fungi, but mycoviruses are still very common in different VCGs, suggesting that viruses can block VI system or have potential transmission bridge under natural conditions. In *R. necatrix*, Rosellinia necatrix fusarivirus 1 (RnFV1) (RNA element N10 carried in fungal strain NW10) can be horizontally transmitted between the two incompatible strains after co-inoculated on an apple tree for three years. It was unclear that any environmental factor changes help RnFV1 to break through the VI system in *R. necatrix* during the process of infecting the host plant. Therefore, unlike simple laboratory conditions, mycovirus transmission can break through the barriers of VI systems under certain natural conditions. It also has been reported that 1.5 mM Zinc chloride treatment was able to facilitate mycovirus transmission via increasing hyphal anastomosis and attenuating PCD in *R. necatrix* and some other fungi. In *N. crassa*, a cell death-activated zinc cluster transcription factor (CZT-1) acts as the regulator for cell death. Using UPLC-HRMS, some secondary metabolite productions have been found significantly shifted in *C. parasitica* during *vic3* incompatibility, such as a farnesyl-S-oxide analog resembling mating pheromones. Adding activated charcoal to the culture medium can also inhibit the VI reactions and promote the transmission of the mycovirus in *R. necatrix*. Some proteins located at cell wall (such as IDI1) may be affected by adding activated charcoal or $\text{Ca}^{2+}/\text{Zn}^{2+}$. Therefore, mycovirus transmission can be modified via eliminating the specific secretions of the strain or suppressing the VI reactions (Fig. 3).

With the development of high-throughput sequencing to gain viral genome information, more and more mycoviruses showed close phylogenetic relationships to plant viruses, suggesting that some mycoviruses could transmit in a cross-kingdom manner.

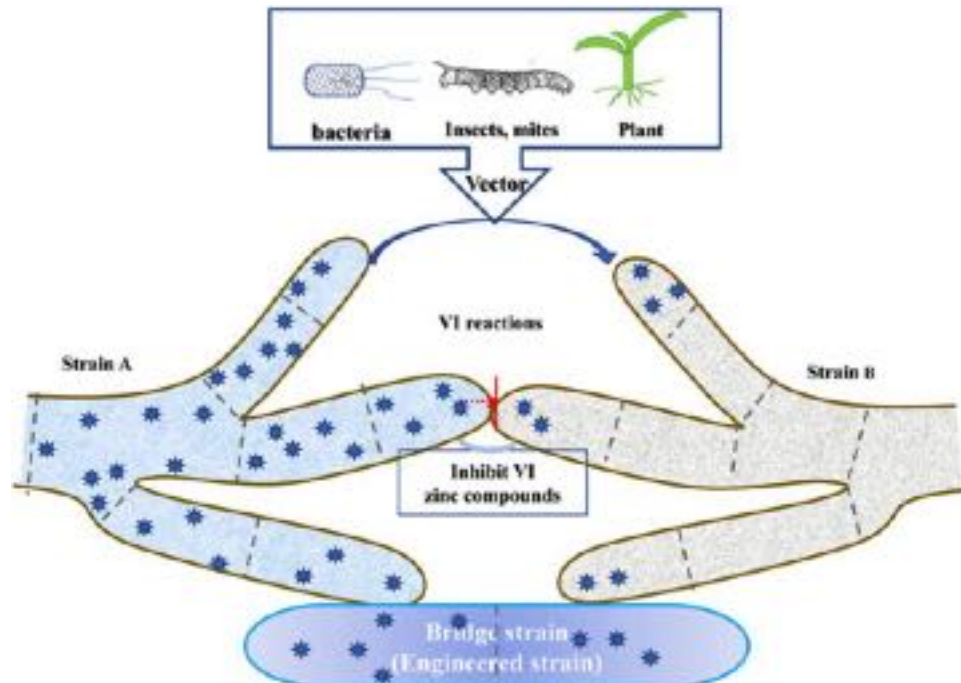


Fig. 3 The potential model of mycovirus horizontal transmission via breakthrough the barrage of VI systems. Strain A, strain B, and bridge strain belong to different VCGs, when strain A and strain B contact each other, the contacting cells will occur strong PCD reactions (red dotted arrow). To attenuate vegetative incompatibility, treatment with zinc compounds can enhance transmission efficiency of mycoviruses to strains with different genetic backgrounds. A bridge strain for mycovirus transmission also can improve mycovirus transmission between different VCGs, this bridge strain can be modulating fungal allelorecognition (disruption of *vic* genes) or the strain that infected by a certain virus (such as SsMYRV4) with suppressing host VI reactions. Mycovirus can be transmitted via different kinds of vectors, such as plants, mites, and insects, etc.

R. solani isolates can obtain and transmit CMV to the host plants, so CMV can be transmitted among different VCGs of *R. solani* in plant. Some mycoviruses are very similar to plant viruses in the order *Tymovirales*, these mycoviruses may replicate in both plant and fungi, and the mycoviruses should shuttle between plants and fungi when pathogenic fungi infect plant tissues, so these groups of mycoviruses may block the barrage of VI systems via vector plants. Conversely, a mycovirus may infect a host plant, bacteria, or protozoa transiently and then virus can be transmitted back to fungi (Fig. 3). In root rot fungi *Heterobasidion* spp., the transmission efficacy of mycovirus HetPV15-pa1 could increase from 0% to 50% among different VCGs if the donor strain was pre-infected with HetPV13-an1. Thus, co-infection with some viruses or bacteria could help other mycovirus transmission among VCGs by blocking VI-mediated PCD.

SsMYRV4 can suppress VI-mediated PCD to block the barriers of VI systems in *S. sclerotiorum*. VI-mediated PCD is controlled via transduction of VI signals, including recognition of strain-specific molecules via G-protein-coupled receptors (GPCRs) and NLRs, transduction of VI signals via MAK2 kinase and induction PCD via ROS signaling (Fig. 2). Thus, suppression of the expression of the key genes in these pathways would effectively block VI-mediated PCD and enable mycovirus horizontal transmission among VCGs (Fig. 3).

Further Reading

- Biella, S., Smith, M.L., Cortesi, P., *et al.*, 2002. Programmed cell death correlates with virus transmission in a filamentous fungus. *Proceedings of the Royal Society B-Biology Science* 269, 2269–2276.
- Daskalov, A., Heller, J., Herzog, S., *et al.*, 2017. Molecular mechanisms regulating cell fusion and heterokaryon formation in filamentous fungi. *Microbiol Spectr* 5 (2).
- Fleißner, A., Herzog, S., 2016. Signal exchange and integration during self-fusion in filamentous fungi. *Seminars in Cell & Developmental Biology* 57, 76–83.
- Gonçalves, A.P., Chow, K.M., Cea-Sánchez, C., *et al.*, 2020. WHI-2 regulates intercellular communication via a MAP kinase signaling complex. *Frontiers in Microbiology* 10, 3162.
- Hutchison, E., Brown, S., Tian, C., *et al.*, 2009. Transcriptional profiling and functional analysis of heterokaryon incompatibility in *Neurospora crassa* reveals that reactive oxygen species, but not metacaspases, are associated with programmed cell death. *Microbiology* 155, 3957–3970.
- Kaneko, I., Dementhon, K., Xiang, Q., *et al.*, 2006. Nonallelic interactions between *het-c* and a polymorphic locus, *pin-c*, are essential for nonself recognition and programmed cell death in *Neurospora crassa*. *Genetics* 172, 1545–1555.
- Li, L., Wright, S.J., Krystofova, S., *et al.*, 2007. Heterotrimeric G protein signaling in filamentous fungi. *Annual Review Microbiology* 61, 423–452.
- Milgroom, M.G., Cortesi, P., 2004. Biological control of chestnut blight with hypovirulence: A critical analysis. *Annual Review Phytopathology* 42, 311–338.
- Milgroom, M.G., Smith, M.L., Drott, M.T., *et al.*, 2018. Balancing selection at nonself recognition loci in the chestnut blight fungus, *Cryphonectria parasitica*, demonstrated by trans-species polymorphisms, positive selection, and even allele frequencies. *Heredity* 121, 511–523.
- Monika, S.F., Wilfried, J., Nuss, D.L., 2019. Integration of self and non-self recognition modulates asexual cell-to-cell communication in *Neurospora crassa*. *Genetics* 211, 1255–1267.
- Paoletti, M., 2016. Vegetative incompatibility in fungi: From recognition to cell death, whatever does the trick. *Fungal Biology Review* 30, 152–162.
- Read, N.D., Goryachev, A.B., Lichius, A., 2012. The mechanistic basis of self-fusion between conidial anastomosis tubes during fungal colony initiation. *Fungal Biology Review* 26, 1–11.
- Uehling, J., Deveau, A., Paoletti, M., 2017. Do fungi have an innate immune response? An NLR-based comparison to plant and animal immune systems. *PLoS Pathogens* 13, e1006578.
- Wu, S., Cheng, J., Fu, Y., *et al.*, 2017. Virus-mediated suppression of host non-self recognition facilitates horizontal transmission of heterologous viruses. *PLoS Pathogens* 13, e1006234.
- Zhang, D.X., Nuss, D.L., 2016. Engineering super mycovirus donor strains of chestnut blight fungus by systematic disruption of multilocus *vic* genes. *Proceedings of the National Academy of Sciences of the United States of America* 113, 2062–2067.

Viral Diseases of *Agaricus bisporus*, the Button Mushroom

Kerry S Burton, Leamington Spa, United Kingdom

Greg Deakin, NIAB-EMR, East Malling, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Casing The upper layer (55 mm) of the bilayer system for *A. bisporus* cultivation, formulated by mixing peat (75%–85%) and lime/chalk (15%–25%) followed by large amounts of water.

Fruitbody Multicellular fleshy structure bearing spores and used as a food and also known as mushroom, sporophore, sporocarp and toadstool.

Hyphal anastomosis The union of two hyphae allowing cytoplasmic exchange.

Mushroom compost The nutritional substrate for cultivated *A. bisporus* made by solid-state fermentation of cereal straw,

animal manure, gypsum and water over 14–20 days and then positioned as the lower layer (200 mm) of the bilayer cultivation system.

ORF Open reading frame.

ORFan ORF with no similarity to other described ORFs and no known function.

Phase 3 compost Mushroom compost colonised by *A. bisporus* mycelium for 17 days at 25°C.

Vegetative incompatibility A genetically controlled system which prevents two fungal strains from undergoing anastomosis.

VLP Virus-like particle.

Introduction

Two disease-causing viruses of the cultivated mushroom, *Agaricus bisporus*, have been characterized. For both diseases the viruses have been identified as part of multiple virus infections i.e., in the presence of other apparently asymptomatic viruses. Four million tonnes of *A. bisporus* mushroom fruitbodies are cultivated annually (value of ~US \$4.7bn) throughout the world and traded internationally enabling viral mixing and movement, and the accumulation of high viral loads. *A. bisporus* is cultivated in a bilayer system with a 55 mm non-nutritious casing layer above a 200 mm layer of composted cereal straw colonised by *A. bisporus* mycelium. This cultivation system is intensive enabling up to nine crops per year to be grown in the same room with cleaning and sterilization between each crop.

In the early 1950s a new disorder of the cultivated mushroom was described and named as La France Disease after the initial observations at the La France Farm in Pennsylvania, USA. The disease has since been reported in other U.S. farms and around the world. The symptoms involve slow mycelial and fruitbody growth, malformed mushroom fruitbodies and major crop loss (up to 100%) often observed as barren zones on the mushroom growing bed. La France disease is strongly associated with a 36 nm isometric virus-like particle (VLP) encapsidating nine double-stranded RNA (dsRNA) elements. Other VLPs have been isolated from diseased tissue, the bacilliform particle (19 nm × 50 nm) of the barnavirus, mushroom bacilliform virus (MBV) and 25 nm, 29 nm and 50 nm isometric particles, however their role in disease has not been established. The nine dsRNA elements of the 36 nm particle were collectively named La France isometric virus, or LIV and later by the more systematic name *Agaricus bisporus* virus 1 (AbV1).

In the late 1990s in the UK a new disease syndrome was described with disease symptoms superficially similar to La France disease, i.e., yield-loss (seen as bare patches with no mushrooms growing) and distorted fruitbodies, but without the characteristic presence of VLPs. Furthermore, this new disease had an additional symptom, a uniform brown discoloration on the mushroom caps. The name given to the new disease syndrome was Mushroom Virus X disease (MVX disease) and it is associated with multiple viral infections of up to 30 viral RNAs. Large scale sequencing of the RNAs associated with MVX-infected fruitbodies identified 18 unique viruses and 8 ORFans. Of these, only a single virus, *Agaricus bisporus* virus 16 (AbV16) and one uncategorized non-host RNA molecule (ORFan8) have been correlated with any symptom; cap browning. Brown Cap Mushroom Disease (BCMD) is now viewed as a viral disease distinct from MVX. Since 2000, the prevalence of the bare patches and distortions caused by MVX has declined while the brown cap symptoms have persisted and reported throughout Europe, North America and Asia. BCMD is economically important as brown discoloration is deleterious to quality (of white mushrooms) and therefore value, and the discoloration can occur after harvest causing friction between mushroom producers and buyers.

AbV1 and AbV16 are often found in co-infections with, MBV. Whether any form of synergistic or antagonistic relationship exists between these viruses, or any of the other many viruses infecting *Agaricus bisporus*, remains unknown. It is of note that the levels of MBV can increase by more than ten-fold when in a multiple infection with AbV1 compared to isolates with undetectable levels of AbV1. However, on its own MBV is not capable of producing any of the symptoms of either disease.

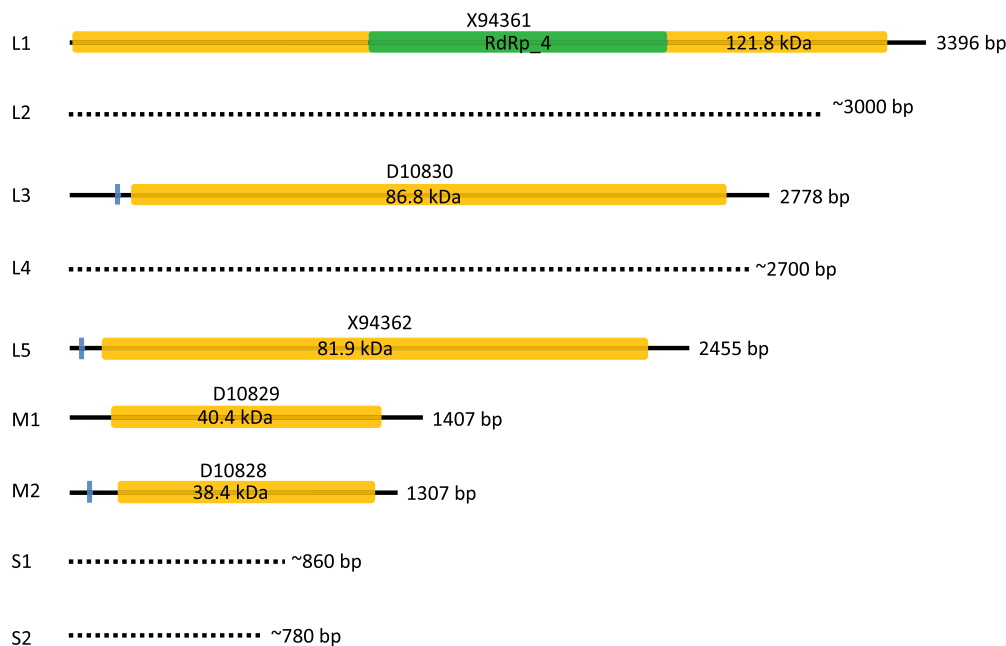


Fig. 1 Genome organization of AbV1. All segments are arranged 5' – 3' and the longest ORFs are shown as yellow annotations. Blue annotations indicate the presence of the shared motif (GGMAACGGCTAGTTKGCC). NCBI accession is shown above each segment. Dashed lines indicate the segment has not been sequenced with the sizes of these segments estimates based on gel electrophoreses.

La France Disease – *Agaricus Bisporus* Virus 1 (AbV1)

Genome Organization

The AbV1 genome comprises of six abundant, always present dsRNAs (L1, L2, L3, L4, L5 and M2) and three less abundant dsRNAs (M1, S1 and S2), with size range of 0.78–3.6 kbp (Fig. 1). If detectable, the less abundant dsRNAs are never found in isolation. Four of the abundant dsRNAs (L1, L3, L5 and M2) and M1 have been sequenced. L1 (3.4 kbp) has a single open reading frame (ORF) which codes for an RNA-dependent RNA polymerase (RdRp) domain containing protein of c. 122 kDa, however the 5' UTR of ~200 bp has not been fully sequenced and so the ORF may be slightly longer. L3 (2.8 kbp) contains a single ORF which codes for an 87 kDa protein and has a 29 nucleotide poly (A) tail on the coding strand. Both L1 and L3 have significant amino-acid similarity to RdRps and structural proteins respectively of members of the *Chrysoviridae*. L5, M1 and M2 code for proteins of size 82 kDa, 42 kDa and 38 kDa respectively, none of which have homology to any known proteins. The polyprotein encoded by L5 is known to be cytoplasmically located and not associated with the AbV1 virions. L3, L5 and M2 share an 18-base motif (GGMAACGGCTAGTTKGCC) in their 5' UTR. The less abundant segment M1 lacks this motif, and it is unknown if the L1 5' UTR contains it due to the incomplete sequencing.

Virion Structure and Composition

AbV1 virions are isometric particles with diameter estimates between 34–36 nm. Three polypeptides are associated with the virions, two abundant molecules of molecular mass 90 kDa and 120 kDa, and a less abundant molecule with estimated mass of either 115 kDa or 129 kDa. The less abundant polypeptide is the RdRp coded by L1. The two abundant polypeptides are presumably structural constituents of the capsid. The 90 kDa protein is coded by the L3 dsRNA, while the 120 kDa molecule is probably coded by L2 as this is the only unsequenced dsRNA of sufficient length to produce a protein of this size. It is not known, or whether it is physically possible, if single virions encapsidated all 9 dsRNA elements, or like other members of the *Chrysoviridae* family the genomic segments are encapsidated separately.

Taxonomy and Classification

Phylogenetic analysis of the RdRp coding region of the L1 segment indicates that *Agaricus bisporus* virus 1 is closely related to members of the family *Chrysoviridae* with dsRNA genomes, and is positioned within the genus *Betachrysovirus* of the *Chrysoviridae* (Fig. 2). It is of note that, unlike most of the classified fungal alphachrysoviruses which are asymptomatic in their hosts, members of the genus *Betachrysovirus* such as *Magnaporthe oryzae* chrysovirus 1 and *Alternaria alternate* chrysovirus 1, typically produce symptoms of retarded growth.

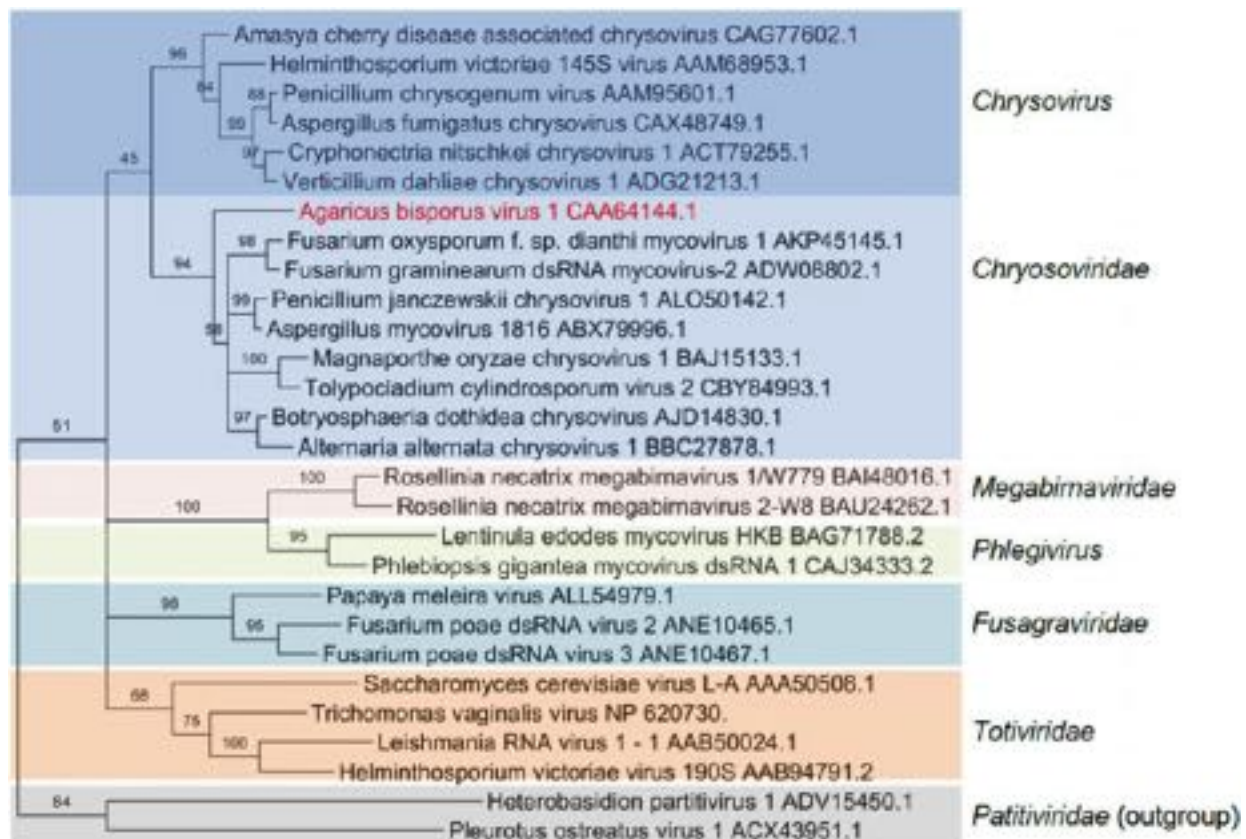


Fig. 2 Phylogenetic relationship of the *Chrysoviridae* and related species. Alignment of the predicated amino acid sequences of the given accessions was performed with MAFFT v7. Uninformative regions of the alignments were removed using Gblocks v0.91. Phylogenetic trees were constructed using Phylml 3 with SMS for best model selection. The numbers refer to the percentage branch support from 1000 bootstrap repetitions. Branches with less than 45% bootstrap support have been collapsed.

Biological Properties (Virus Host Relationships)

As well as La France, the disease has been variously known as X-disease, die-back, watery stipe and brown disease, reflecting the wide range of associated symptoms. The major characteristic symptom is localized bare patches on mushroom growing beds (die-back) caused by failure of the mycelium to colonize the casing and thus failure of mushrooms to emerge. Around the periphery of these die-back zones, the fruitbodies often fail to grow properly and show a characteristic “drumstick” phenotype, where stipes are elongated or swollen and the caps small and misshapen. Other symptoms include brown discoloration of internal tissue (as streaks), bent stipes and a tendency for premature maturation. The severity of the disease varies from a very few abnormal mushrooms to complete loss of the crop. The range of symptoms seems to be affected by the host genotype, the timing of the initial infection with more devastating symptoms occurring when infection occurs at an earlier stage during the mushroom production cycle, and cultural practices. The presence of the AbV1 dsRNAs in a mushroom fruitbody is not in itself sufficient to cause the disease phenotype, though symptomatic fruitbodies from an infected crop always contained the six major dsRNAs.

Epidemiology

AbV1 is spread between mushroom crops via anastomosis; predominantly when infected spores germinate on the mushroom bed, or directly via infected mycelium. It has been noted that infected mushrooms tend to mature early and release a huge number of spores. Like the overwhelming majority of fungal viruses there are no-known external routes to infection or natural vectors.

Diagnostics and Identification

The 36-nm particles of AbV1 can be detected by electron microscope. Gel separation of the dsRNAs has also been successfully used to detect the virus. The more sensitive techniques are PCR-based using the primers to the known RNA sequences of the virus.

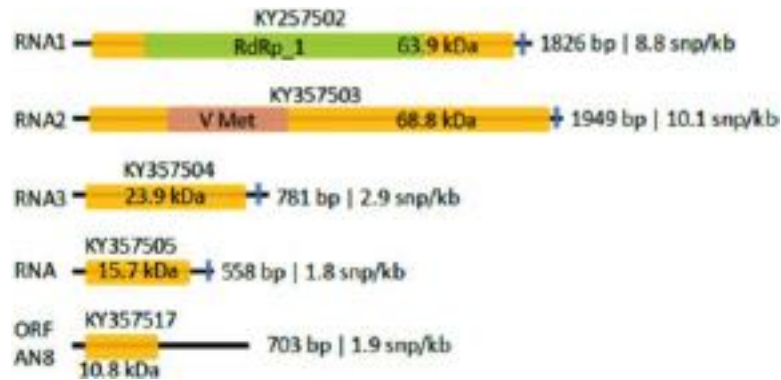


Fig. 3 Genome organization of Abv16 and ORFan8. All segments are arranged 5' – 3' and the longest ORFs are shown as yellow annotations. Blue annotations show the presence and location of the shared motif (STTCAGSGTBBVWSHAGCRGTAAWT). An indication of segment variability is given as the number of single nucleotide polymorphisms per kilobase of sequence (snp/kb). NCBI accession is shown above each segment.

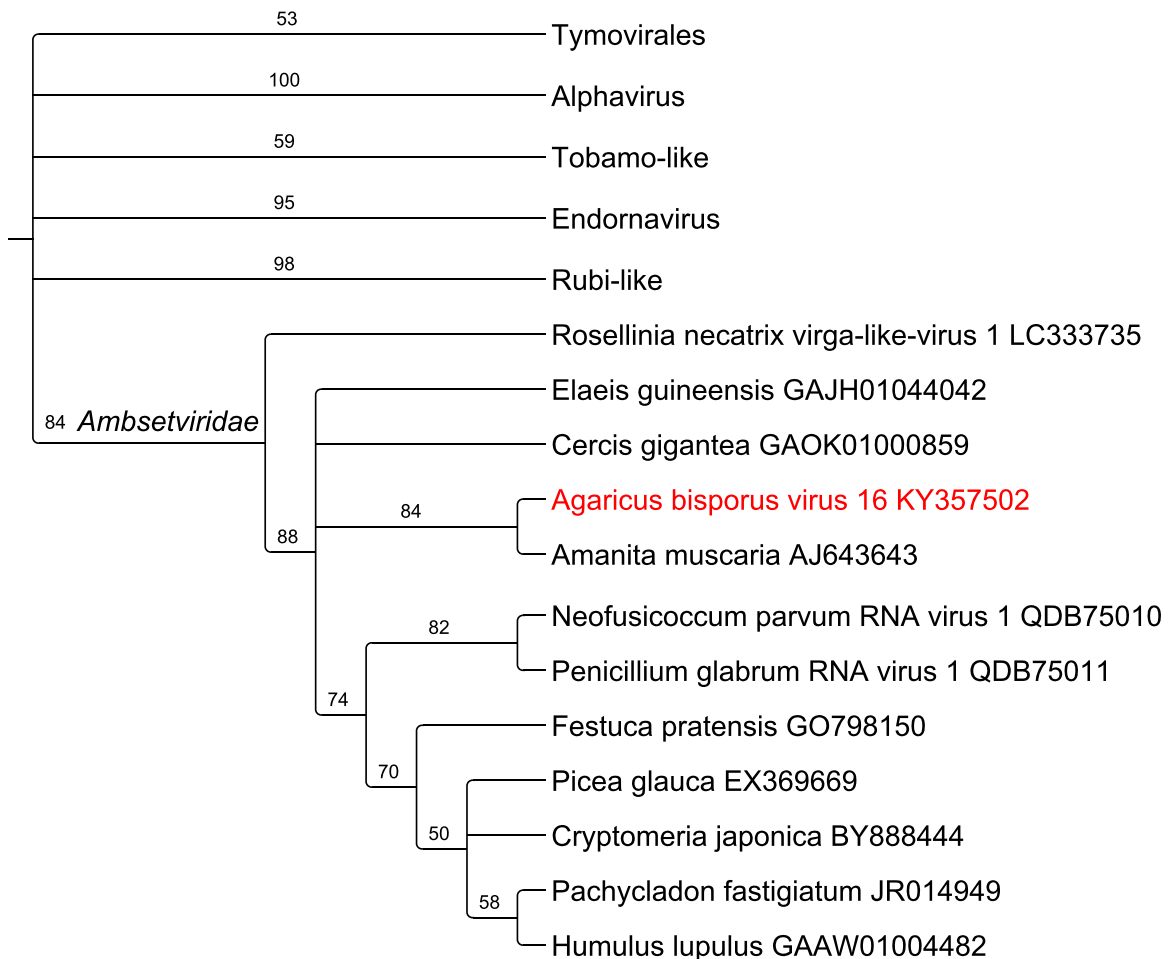


Fig. 4 Phylogenetic relationship of the members of the proposed family *Ambsetviridae* compared to selected members of alphavirus-like supergroup. Alignment of RdRp domains in the predicted amino acid sequences of members of the alphavirus-like supergroup was performed with MAFFT v7. Uninformative regions of the alignments were removed using Gblocks v0.91. Phylogenetic trees were constructed using Neighbor joining. The numbers refer to the percentage branch support from 100 bootstrap repetitions. Branches with less than 50% bootstrap support have been collapsed.

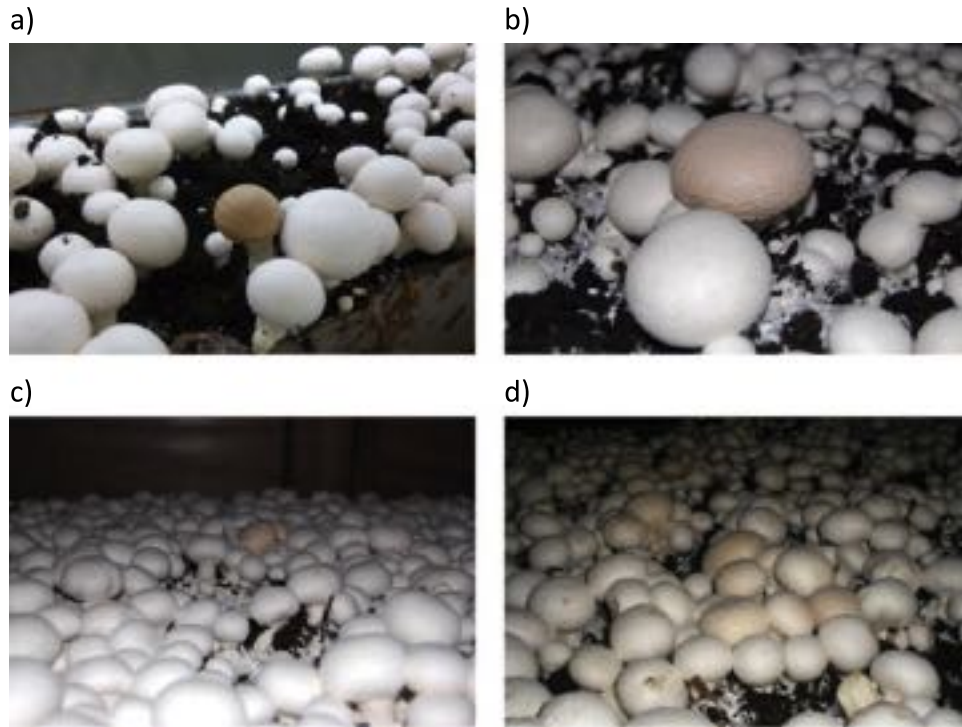


Fig. 5 Typical symptom expression of Brown Cap Mushroom Disease (a) single brown in an experimental bed, (b) single brown in commercial production, (c) group of browns amongst white mushrooms, (d) off-white and brown mushrooms.

Control

There is no natural immunity to AbV1 in any cultivated *A. bisporus* strains. There is a degree of vegetative incompatibility between strains; browns, creams and whites will not readily anastomose therefore rotating strains at the mushroom farm can lead to some protection. However, successful control of La France disease has been achieved by substantial increases in hygiene at mushroom farms; steam treatment of the growing room and equipment and the use of air filtration to exclude infected mushroom spores.

Brown Cap Mushroom Disease – *Agaricus Bisporus* Virus 16 (AbV16)

Genome and Virion Structure

AbV16 is a non-particulate, multipartite virus with four segments ranging in size from 558 to 1949 bases (Fig. 3). The ORFs of AbV16 RNA1 and AbV16 RNA2 encode an RdRp and viral methyltransferase respectively. The function of AbV16 RNAs 3 and 4 are unknown. Each of the AbV16 RNA's contains a 25 base motif in the 3' UTR. ORFan8 does not contain the 3' motif but it is included as part of the causative agent for BCMD, as its levels correlate with the degree of cap browning and with the levels of the components of AbV16. ORFan8 and the AbV16 segments are polyadenylated, and considered to represent the genome of a positive sense, single-stranded RNA virus by close relation to other single-stranded RNA viruses. No virion structure has been identified for this virus.

Classification

AbV16 has been classified as within the alphavirus supergroup by phylogenetic analysis of the Pfam RdRp_2 domain of AbV16 RNA1. It is proposed to be a member of a new viral family, the *Ambsetviridae*, which includes homologs of RNAs previously not classified as viruses but identified from Transcriptome Shotgun Assembly and Expressed Sequence Tag databases (Fig. 4). A recently described virus identified in *Rosellinia necatrix* is a probable member of this family. The organisms identified from the searches suggest the *Ambsetviridae* is a family of fungal and plant viruses. Two of these organisms have homologs to three components of AbV16 and a further seven have homologs to two components. No homologs have been found to the viral segment AbV16 RNA4 or ORFan8.

Viral Expression and Disease Development

AbV16 and ORFan8 are present in infected mushrooms at two distinct levels. Low level infection results in slightly discolored mushroom caps only detectable by precision colorimeters. However, in visibly brown mushrooms, the RNA levels are 10^3 – 10^4 times

higher, intermediate levels have not been detected. The factors determining this persistent to acute transition, from low to high abundance, are not fully understood. There appears to be a stochastic element as brown mushrooms can be observed apparently randomly distributed throughout a bed of white mushrooms (Fig. 5). It is known that the transition occurs early in fruitbody development and so may be associated with host cell division. It is possible the transition may also embrace the “quasi-species” concept, involving the generation of a wide range of sequence variants overcoming host defences, as AbV16 RNAs1 and 2 show high sequence variability.

Epidemiology

Commercial inoculum for mushroom culture involves high standards of hygiene, testing and documentation of axenic cultures of *A. bisporus*. Infection is likely to occur at mushroom growing farms or “phase 3 compost” producers from infected sources (cultivated or wild *A. bisporus*), and associated with inadequate hygiene procedures.

The intra-mycelial movement of AbV16 within compost culture has been shown to be slow at 40 mm/day (compared with 120 mm/day and >200 mm/day for two other *A. bisporus* viruses, AbV12 and AbV7 respectively). Therefore, a point-infection of AbV16 on mushroom growing beds will have limited range for enlargement during the short growing period. Wide-spread outbreaks of BCMD are likely to have resulted from an infection during phase 3 compost production and/or compost mixing, transport and distribution. A further interesting feature of AbV16 is that its levels in mycelium in compost or laboratory culture can decline dramatically over time to become undetectable. The reasons for this decline are unknown, however AbV16 has been found only as part of a multiple viral infection and it is possible that inter-viral competition associated with specific viral combinations may play a role.

Diagnosis

High level infection of AbV16 and ORFan8 can be detected visually as brown discolored mushroom fruitbodies and low-level infection can be detected by sensitive colorimeter ($\Delta E > 15$ or $L/b < 12$).

AbV16 and ORFan8 can be detected qualitatively by agarose gel separation or PCR. The increased sensitivity of quantitative PCR enables detection of early, low level infection in mycelium in compost/casing and fruitbodies.

Further Reading

- Deakin, G., Dobbs, E., Bennett, J., *et al.*, 2017. Multiple viral infections in *Agaricus bisporus* – Characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. *Scientific Reports* 7, 2469. doi:10.1038/s41598-017-01592-9.
- Eastwood, D.C., Green, J., Grogan, H., Burton, K.S., 2015. Characterizing the viral agents causing brown cap mushroom disease of *Agaricus bisporus*. *Applied and Environmental Microbiology* 81, 7125–7134.
- Fleming-Archibald, C., Burton, K.S., Grogan, H.M., 2015. Brown cap mushroom virus (associated with Mushroom Virus X) prevention. *MushTV Factsheet* 02/15. Available at: http://www.hdc.org.uk/sites/default/files/02_15%20Brown%20Cap%20Mushroom%20Virus%20prevention.pdf.
- Fleming-Archibald, C., Ruggiero, A., Grogan, H.M., 2015. Brown mushroom symptom expression following infection of an *Agaricus bisporus* crop with MVX associated dsRNAs. *Fungal Biology* 119, 1237–1245.
- Fletcher, J.T., Gaze, R.H., 2008. *Mushroom Pest and Disease Control: A Colour Handbook*. Mansion Publishing Ltd: London, UK, p. 192.
- Frost, R.R., Passmore, E.L., 1980. Mushroom viruses: A re-appraisal. *Journal of Phytopathology* 98, 272–284.
- Goodin, M.M., Schlagnhauer, B., Romaine, C.P., 1992. Encapsidation of the La France disease-specific double-stranded RNAs in 36 nm isometric virus-like particles. *Phytopathology* 82, 285–290.
- Grogan, H.M., Adie, B.A., Gaze, R.H., Challen, M.P., Mills, P.R., 2003. Double-stranded RNA elements associated with the MVX disease of *Agaricus bisporus*. *Mycological Research* 107, 147–154.
- Harmsen, M.C., Van Griensven, L.J.L.D., Wessels, J.G.H., 1989. Molecular analysis of *Agaricus bisporus* double-stranded RNA. *Journal of General Virology* 70, 1613–1616.
- Harmsen, M.C., Tolner, B., Kram, A., *et al.*, 1991. Sequences of three dsRNAs associated with La France disease of the cultivated mushroom (*Agaricus bisporus*). *Current Genetics* 20, 137–144.
- Marino, R., Saksena, K.N., Schuler, M., Mayfield, J.E., Lemke, P.A., 1976. Double-stranded ribonucleic acid in *Agaricus bisporus*. *Applied and Environmental Microbiology* 31, 433–438.
- Revill, P.A., 2008. Barnaviruses. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Elsevier Ltd, pp. 286–288.
- Romaine, C.P., Goodin, M.M., 2001. Unraveling the viral complex associated with La France disease of the cultivated mushroom, *Agaricus bisporus*. In: Tavantzis, S.M. (Ed.), *dsRNA Genetic Elements- Concepts and Applications in Agriculture, Forestry and Medicine*. CRC Press, pp. 237–257.
- Van der Lende, T.R., Duitman, E.H., Gunnewijk, M.G., Yu, L., Wessels, J.G., 1996. Functional analysis of dsRNAs (L1, L3, L5, and M2) associated with isometric 34-nm virions of *Agaricus bisporus* (white button mushroom). *Virology* 217, 88–96.

Viral Killer Toxins

Manfred J Schmitt and Björn Becker, Saarland University, Saarbrücken, Germany

© 2021 Elsevier Ltd. All rights reserved.

This is an update of M.J. Schmitt, Viral Killer Toxins, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00577-X.

Nomenclature

bp Base pairs

ds Double-stranded

RNAi RNA interference

VBS Viral binding site

IRE Internal replication enhancer

TRE Terminal recognition element

ERAD Endoplasmatic reticulum (ER)-associated protein degradation

NLS Nuclear localization sequence

Glossary

ER-associated degradation (ERAD) A cellular quality control mechanism ensuring efficient removal of misfolded and/or unassembled proteins from the endoplasmic reticulum (ER) lumen and their subsequent elimination by the cytoplasmic ubiquitin–proteasome system.

H/KDEL receptors Seven-transmembrane proteins primarily required for the retention of soluble ER resident proteins in the endoplasmatic reticulum in yeast and mammalian cells.

Heterokaryon Coexistence of two or more genetically different nuclei in a common cytoplasm. Importins: A family of proteins that transport macromolecules into the eukaryotic nucleus.

Killer virus system A killer virus system consists of a helper totivirus and a satellite double-stranded (ds) RNA encoding the unprocessed precursor of a secreted protein toxin that also gives functional immunity.

Preprotoxin Toxin precursor in the cytoplasm of a killer yeast which is post-translationally imported into the secretory pathway for toxin processing, maturation, and secretion.

Satellite dsRNA An encapsidated dsRNA that is dependent on a helper virus for encapsidation and replication.

Spheroplasts A yeast cell whose cell wall has been enzymatically removed.

Introduction

Although double-stranded (ds) RNA viruses had previously been identified in filamentous fungi, dsRNA viruses in yeast were identified as determinants of the killer phenomenon in *Saccharomyces cerevisiae* and are now known to be associated with the presence of cytoplasmically inherited members (totiviruses) of the genus *Totivirus* in the family *Totiviridae*, which are frequently found in various yeast genera. Among these, the killers of *S. cerevisiae*, *Zygosaccharomyces bailii*, *Hanseniaspora uvarum*, *Torulaspora delbrueckii* and *Ustilago maydis* – the latter being the cause of corn smut – are best characterized. Interestingly, the absence of RNA interference (RNAi) seems to be a prerequisite for dsRNA viruses existence as killer systems were so far merely found in RNAi-deficient yeast species whereas RNAi-proficient yeasts did not develop any killer systems during evolution. Characteristic for all killers is the secretion of a certain type of protein toxin that is lethal to sensitive strains of different species and genera without a direct cell-to-cell contact. Cell killing is usually achieved in a receptor-mediated process, requiring initial toxin binding to components of the outer yeast cell surface (such as β -1,6-D-glucans, α -1,3-mannoproteins, or chitin) and subsequent toxin transfer to a secondary plasma membrane receptor. Depending on the toxin, final lethality can be caused by plasma membrane damage, G1- or S-phase cell-cycle arrest, apoptosis, and/or by rapid inhibition of DNA synthesis.

In the yeasts *S. cerevisiae*, *Z. bailii*, *H. uvarum*, *Torulaspora delbrueckii* as well as in the maize smut fungus *U. maydis*, the killer phenotype is cytoplasmically inherited and caused by an infection with dsRNA viruses. Since the majority of fungal mycoviruses are noninfectious and symptomless in the corresponding host, they are often classified as cryptic viruses or virus-like particles (VLPs). All known fungal viruses spread vertically by cell–cell mating and/or heterokaryon formation. In *S. cerevisiae*, diploids formed by mating of a killer with a sensitive strain are likewise killers, as are all haploid progeny of subsequent meiosis. In contrast, virus-free strains are usually sensitive nonkillers, while those containing *Saccharomyces cerevisiae* virus L-A (ScV-L-A) and a toxin-encoding M-dsRNA are killers (see below). Sensitive strains survive mating with killers, and cytoplasmic mixing of the multiple M-dsRNA copies during zygosis accounts for the inheritance pattern during meiosis. Extracellular spread of virions is generally hampered by the rigid yeast and fungal cell wall barrier, and fungal viruses have adopted a strategy of transmission via mating and hyphal fusion (which occurs frequently in nature) making an extracellular route of spread dispensable. While some of these viruses can be associated with adverse phenotypic effects on the fungus (like La France disease in *Agaricus bisporus*, plaque formation in *Penicillium*, and hypovirulence in *Cryphonectria*), dsRNA viruses and their associated satellite dsRNAs are responsible

Table 1 dsRNA encoded killer toxin systems of various yeast species

Totivirus/satellite dsRNA	Virus host	Function of virus	dsRNA (kb)	Encoded protein(s)
ScV-L-A	<i>Saccharomyces cerevisiae</i>	Helper virus	L _A (4.6)	Gag, major capsid protein; Gag-Pol, RDRP ^a
ScV-M1	<i>Saccharomyces cerevisiae</i>	Satellite dsRNA	M ₁ (1.6)	K1 preprotoxin
ScV-M2	<i>Saccharomyces cerevisiae</i>	Satellite dsRNA	M ₂ (1.5)	K2 preprotoxin
ScV-M28	<i>Saccharomyces cerevisiae</i>	Satellite dsRNA	M ₂₈ (1.8)	K28 preprotoxin
ScV-Mlus	<i>Saccharomyces cerevisiae</i>	Satellite dsRNA	M _{lus} (2.1–2.3)	Klus preprotoxin
UmV-H	<i>Ustilago maydis</i>	Helper virus	H (6.1)	Gag, major capsid protein; Gag-Pol, RDRP ^a
UmV-P1	<i>Ustilago maydis</i>	Satellite dsRNA	M _{P1} (1.4)	KP1 preprotoxin
UmV-P4	<i>Ustilago maydis</i>	Satellite dsRNA	M _{P4} (1.0)	KP4 toxin
UmV-P6	<i>Ustilago maydis</i>	Satellite dsRNA	M _{P6} (1.2)	KP6 preprotoxin
HuV-L	<i>Hanseniaspora uvarum</i>	Helper virus	L _A (4.6)	Gag, major capsid protein; Gag-Pol, RDRP ^a
HuV-M	<i>Hanseniaspora uvarum</i>	Satellite dsRNA	M _{Hu} (1.0)	KT470 toxin precursor
ZbV-L	<i>Zygosaccharomyces bailii</i>	Helper virus	L _{Zb} (4.6)	Gag, major capsid protein; Gag-Pol, RDRP ^a
ZbV-M	<i>Zygosaccharomyces bailii</i>	Satellite dsRNA	M _{Zb} (2.1)	prepro-zygocin
TdV-LAbarr	<i>Torulaspora delbrueckii</i>	Helper virus	L _A (4.6)	Gag, major capsid protein; Gag-Pol, RDRP ^a
TdV-Mbarr-1	<i>Torulaspora delbrueckii</i>	Satellite dsRNA	M _{barr-1} (1.7)	Kbarr-1 preprotoxin

^aRDRP, RNA-dependent RNA polymerase.

Adapted from Schmitt, M.J., Breinig, F., 2002. The viral killer system in yeast: From molecular biology to application. FEMS Microbiology Reviews 26, 257–276, with permission from Blackwell Publishing.

for a killer phenotype which is based on the secretion of a polypeptide toxin (killer toxin) that is lethal to a variety of sensitive yeast and fungal strains. With the exception of toxin-secreting strains of *Z. bailii*, killer toxin production is usually associated with specific immunity, protecting the corresponding killer yeast against its own viral toxin.

DsRNA Viruses and Killer Phenotype Expression in *S. Cerevisiae*

On the basis of killing profiles and the lack of cross-immunity, four major killer types (K1, K2, K28, Klus) have so far been identified in *S. cerevisiae*. Each of them produces a specific killer toxin and a self-protecting immunity component. Killer phenotype expression correlates with the presence of two types of dsRNA species stably persisting in the cytoplasm of the infected host: the genomic dsRNA of the helper virus, ScV-L-A, and one of four toxin-coding satellite dsRNAs (ScV-M1, ScV-M2, ScV-M28 or ScV-Mlus). The ScV-L-A and M-dsRNAs are separately encapsidated into capsids encoded by ScV-L-A dsRNA and are present in high copy number in the yeast cell cytoplasm. *In vivo*, ScV-L-A does not confer a phenotype nor does it lead to host cell lysis or slower cell growth. While the killer phenotype can be transmitted to sensitive yeast cell spheroplasts (harboring ScV-L-A) by transfection with purified ScV-M (either during mating or cotransformation with a dsDNA plasmid), extracellular transmission occurs rarely, if at all, in nature. The survival strategy adopted by these dsRNA viruses appears to be a balanced host interaction resulting in stable maintenance, little if any growth disadvantage, and vertical transmission. Mechanisms of exiting and entering the host cell through its tough and rigid cell wall are rendered unnecessary by relatively efficient horizontal transmission during the frequent zygosis events in yeast. Acquisition of a toxin-encoding M satellite dsRNA provides positive selection for both this dsRNA and ScV-L-A, since virus-free segregants are killed.

As summarized in **Table 1**, the linear dsRNA genome of ScV-L-A contains two open reading frames (ORFs) on its plus-strand RNA: ORF1 encodes the major capsid protein Gag necessary for encapsidation and viral particle structure, the second gene (ORF2) represents the RNA-dependent RNA polymerase Pol which is *in vivo* expressed as a Gag–Pol fusion protein by a –1 ribosomal frameshift event. In contrast to ScV-L-A, each M-dsRNA genome contains a single ORF coding for a preprotoxin (pptox) representing the unprocessed precursor of the mature and secreted toxin that also gives functional immunity. Since each toxin-coding ScV-M dsRNA depends on the coexistence of ScV-L-A for stable maintenance and replication, the killer viruses resemble classical satellites of ScV-L-A. Although the presence of all four M-dsRNAs with different killer specificities in a single cell is excluded at the replicative level of the M genomes, this limitation can be by-passed by introducing cDNA copies of other pptox genes into a natural killer strain. Thereby, the artificial construction of a stable triple-killer strain producing three different virus toxins (K1, K2 and K28) and showing multiple toxin immunity at the same time is feasible.

As typical for many dsRNA viruses, the replication cycles of ScV-L-A and M-dsRNAs depend on the presence of specific packaging signals at or near the 3'-end of the viral single-stranded RNA (ssRNA) transcript. These RNA regions function as viral binding sites (VBSs) and consist of a stem–loop structure whose stem is interrupted by an unpaired protruding A residue (**Fig. 1**). For replication of ScV-L-A, two sequence elements are essential: an internal replication enhancer (IRE), which is essentially indistinguishable from the VBS, and a small stem–loop structure 5 bp from the 3'-terminus (3'-TRE). While ScV-L-A contains just a single VBS element, the toxin-coding transcripts of M1, M28, and MZb (encoding the *Z. bailii* viral toxin zygocin) each have two such VBS domains. Interestingly, no VBS elements are present in the dsRNA sequence of Mlus and Mbarr-1, indicating that alternative cis- active elements can be predicted to exist in the 3'-RNA region which act as signal for replication, replication-enhancement and packaging. *In vivo*, these VBS elements are

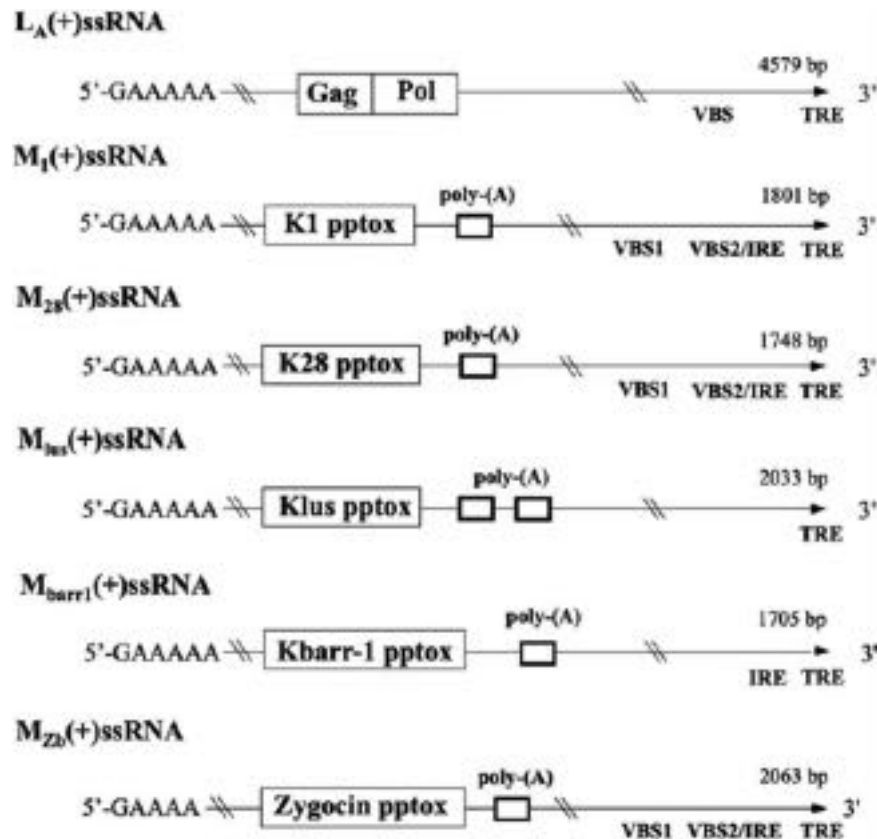


Fig. 1 Comparison of the 5'- and 3'-ssRNA sequences in ScV-L-A and in the toxin-coding transcripts of its satellites ScV-M1, ScV-M28, ScV-Mlus, TdV-Mbarr-1, and ZbV-M. Potential *cis*-active sequences within the 3'-termini as well as poly-A rich regions are indicated (VBS, viral binding site; IRE, internal replication enhancer; TRE, 3'-terminal recognition element). Length of the (+)ssRNA are shown in base pairs (bp).

cis-active sequences that are recognized by Gag–Pol and subsequently packaged into new viral particles. The replicase reaction on the (+)ssRNA template takes place *in vitro* (i.e., within the viral capsid) and requires a correct 3'-end sequence and structure. Within the intact and mature virion, conservative transcription of the plus-strand from the dsRNA template requires recognition of its very 5'-terminal sequence by Gag–Pol. In M28 (as in other dsRNAs), the plus-strand initiates with 5'-GAAAA(A); since there are little additional conserved sequences immediately downstream, this terminal recognition element (5'-TRE) may be all that is necessary for transcription initiation (Fig. 1).

Viral Replication Cycle

Yeast ScV-L-A virions are noninfectious icosahedral particles 39 nm in diameter that show certain similarities to core particles of mammalian orthoreoviruses and rotaviruses. Each ScV-L-A particle consists of a single copy of the 4.6 kb L-A dsRNA genome which is encapsidated by 60 asymmetric dimers of the 76 kDa coat protein Gag and two copies of the 171 kDa Gag–Pol fusion protein. During conservative replication, the single-stranded plus-strand RNA (L-A(+)ssRNA) is transcribed within the viral particle (*in vitro*) and subsequently extruded into the cytoplasm where it serves as (1) messenger RNA for translation into the viral proteins Gag and Gag–Pol and (2) RNA template which is packaged into new viral particles. Once this coat assembly is completed, Gag–Pol functions as replicase, synthesizes a minus-strand, and generates the dsRNA genome of the mature virus. The replication cycle of the toxin-coding M satellite dsRNA resembles that of L-A with the exception that each M virion can accept two copies of the smaller M-dsRNA genome before the ssRNA transcript is extruded into the cytoplasm; in analogy to certain DNA bacteriophages, this phenomenon has been named 'headful packaging'.

Viral Preprotoxin Processing and Toxin Maturation

In totivirus infected killer yeasts, the toxin-encoding M(+)ssRNA transcript is translated on free cytosolic ribosomes into a pptox precursor which is post-translationally imported into the secretory pathway for further processing, maturation, and toxin secretion. Interestingly, intracellular pptox processing in killer strains of either *S. cerevisiae* (K1, K2, K28), *Z. bailii* (zygotin), or *U. maydis*

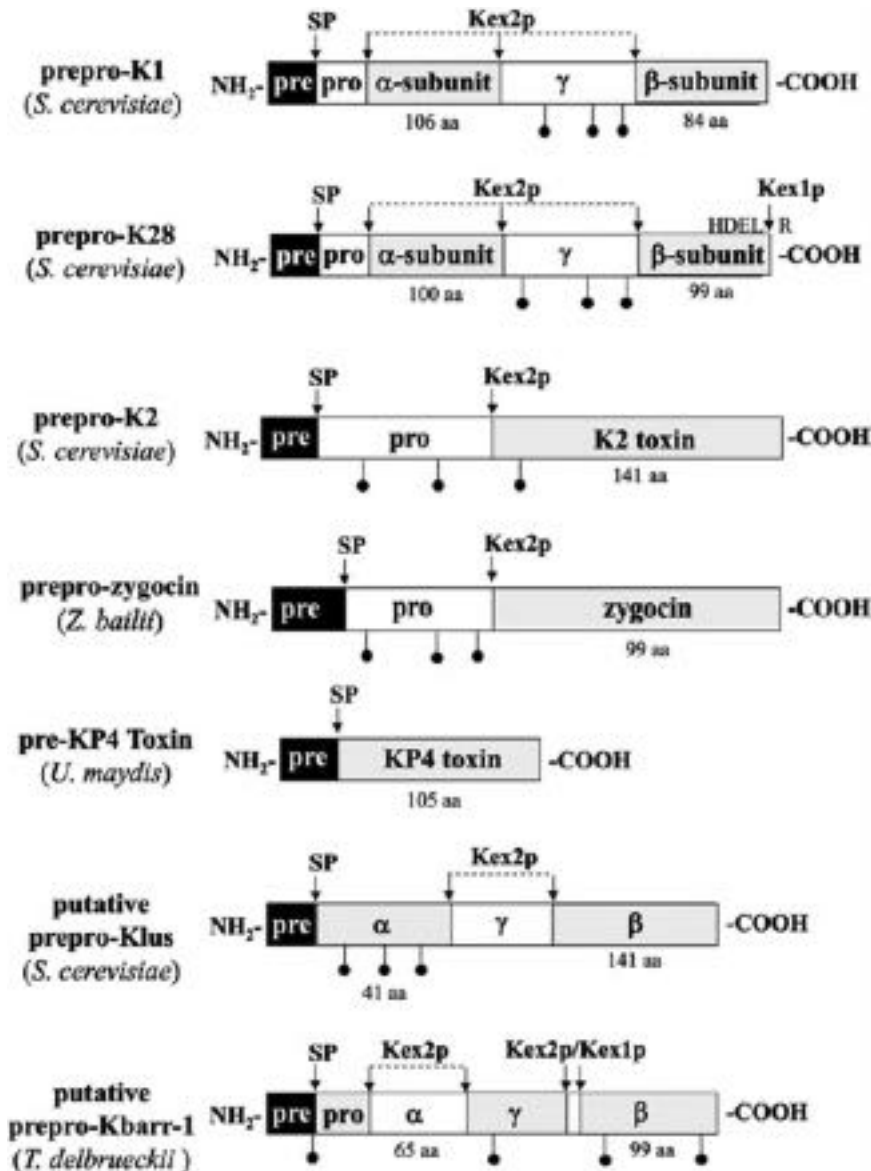


Fig. 2 Analogy of toxin precursor processing in totivirus infected killers of *S. cerevisiae*, *Z. bailii*, *T. delbrueckii* and *U. maydis*. Schematic outline of the known and predicted precursor proteins (preprotoxins) encoding either a heterodimeric α/β toxin (K1, K28, Klus and Kbarr-1) or a monomeric virus toxin (K2, zygocin, KP4). Signal peptidase (SP) cleavage, Kex2p endopeptidase as well as Kex1p carboxypeptidase processing sites are indicated. Potential N-glycosylation sites are indicated with black circles (aa, amino acids).

(KP4) has been demonstrated as being mechanistically conserved, resulting in the secretion of a biologically active monomeric or α/β heterodimeric virus toxin. Based on hydrophobicity plots and *in silico* sequence analysis, a preprotoxin structure is also predicted for the Klus and Kbarr-1 killer toxins, however further experimental data is required to determine the exact nature and function of the postulated subunits (Fig. 2).

For K28 killer strains containing ScV-M28, pptox processing has been intensively studied and is well understood. Mature K28 toxin represents a heterodimer whose α - and β -subunit is covalently linked by a single disulfide bond. Since the unprocessed toxin precursor resembles a secretory protein, it contains a hydrophobic signal sequence for pptox import into the ER lumen followed by the toxin subunits α (10.5 kDa) and β (11.0 kDa) which are separated from each other by a potentially N-glycosylated γ -sequence (Fig. 2). During passage through the yeast secretory pathway the K28 toxin precursor is enzymatically processed to the biologically active heterodimer in a way that is highly homologous to prohormone (e.g., preproinsulin) conversion in mammalian cells. In a late Golgi compartment, the N-glycosylated γ -sequence is removed by the action of the furin-like endopeptidase Kex2p, the C-terminus of β is trimmed by carboxypeptidase Kex1p cleavage, and the biologically active protein is secreted as a 21 kDa α/β heterodimer whose β -C-terminus carries a four-amino-acid epitope (HDEL) that represents a classical ER retention signal. Since

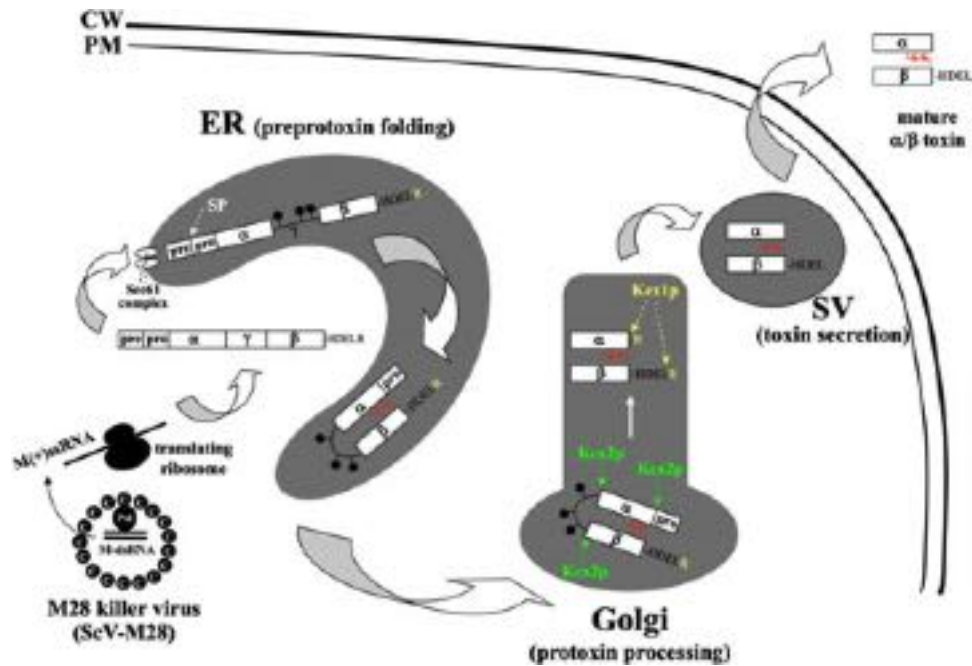


Fig. 3 Preprotoxin processing and toxin secretion in an ScV-M28 infected killer yeast. After *in vivo* translation of the preprotoxin coding killer virus transcript, the toxin precursor is post-translationally imported into the endoplasmic reticulum (ER) through the Sec61 complex. Signal peptidase (SP) cleavage in the ER lumen removes the N-terminal signal sequence (pre-region) and protoxin folding is mediated by luminal ER chaperones. The intervening γ -sequence is N-glycosylated and a single disulfide bond between α and β is generated. In a late Golgi compartment, Kex2p endopeptidase cleaves the pro-region, removes the γ -sequence, and carboxypeptidase Kex1p cleavage trims the C-termini of both subunits, leading to the secretion of mature α/β toxin whose β -C-terminal HDEL motif is uncovered and thus accessible for interaction with the HDEL receptor of the target cell. SV, secretory vesicle; CW, cell wall; PM, plasma membrane. Adapted from Schmitt, M.J., Breinig, F., 2002. The viral killer system in yeast: From molecular biology to application. *FEMS Microbiology Reviews* 26, 257–276, with permission from Blackwell Publishing.

this signal is initially masked by a terminal arginine residue (HDELR), ER retention of the toxin precursor is prevented until the protoxin enters a late Golgi compartment where Kex1p cleavage uncovers the HDEL signal of the toxin (Fig. 3).

Endocytosis and Intracellular Transport of the K28 Virus Toxin

In contrast to most virally encoded yeast killer toxins, which do not enter their host but rather kill a sensitive cell by disrupting plasma membrane function, K28 is taken up by receptor-mediated endocytosis and subsequently targeted to the secretory pathway. After initial binding to α -1,3- and/or α -1,2-mannotriose side-chains of a 185 kDa cell wall mannoprotein, K28 interacts with its secondary toxin receptor, the HDEL receptor Erd2p, at the plasma membrane level (Fig. 4). Responsible and essential for this interaction is a short-amino-acid motif at the carboxyterminus of the K28 β -subunit (HDEL) which normally functions as ER targeting/retention signal for ER-resident chaperones such as Kar2p/BiP and Pdi1p. Once the K28/Erd2p complex is endocytosed, K28 is retrogradely transported from endosomes *via* Golgi to the ER, from where the toxin retrotranslocates into the cytosol to finally reach its target(s) in the nucleus. Due to the high affinity binding between the toxin's HDEL-motif and the cellular HDEL receptor within the low pH-environment of the endosomal and Golgi compartment, the receptor/toxin complexes are likely associated until they reach the neutral pH inside the ER. However, further research is needed to fully understand the exact mechanism(s) of how the toxin/receptor complexes are *in vivo* recognized by the endocytotic machinery and how the internalization and retrograde transport through the different cellular compartments is regulated. In fact, endocytotic uptake and retrograde transport is a common strategy realized in certain prototypes of bacterial toxins such as *Pseudomonas* exotoxin A, *Escherichia coli* heat-labile toxin (HLT), or even Shiga toxin. All these protein toxins are family members of microbial A/B toxins which are usually internalized by receptor-mediated endocytosis, followed by reverse secretion *via* Golgi and ER. Interestingly, many of these toxins contain putative ER retention signals at their C-termini, and H/KDEL-dependent mechanisms have, therefore, been postulated to be of major importance for toxin entry into mammalian cells. In this respect, a major difference between the yeast K28 virus toxin and bacterial A/B toxins is that K28 itself is produced and secreted by a eukaryotic (yeast) cell, and therefore the C-terminal ER-targeting signal in the toxin precursor is initially masked by a terminal arginine residue which ensures successful ptox passage through the early secretory pathway. Once the toxin has reached a late Golgi compartment, Kex1p cleavage removes the β -C-terminal arginine residue and thereby uncovers the ER targeting signal of the virus toxin. To ensure proper access of the ER targeting signal to the H/KDEL receptor of the sensitive target cell, many A/B toxins (including the yeast K28 virus toxin) contain a unique disulfide bond at or near the C-terminus. Consequently,

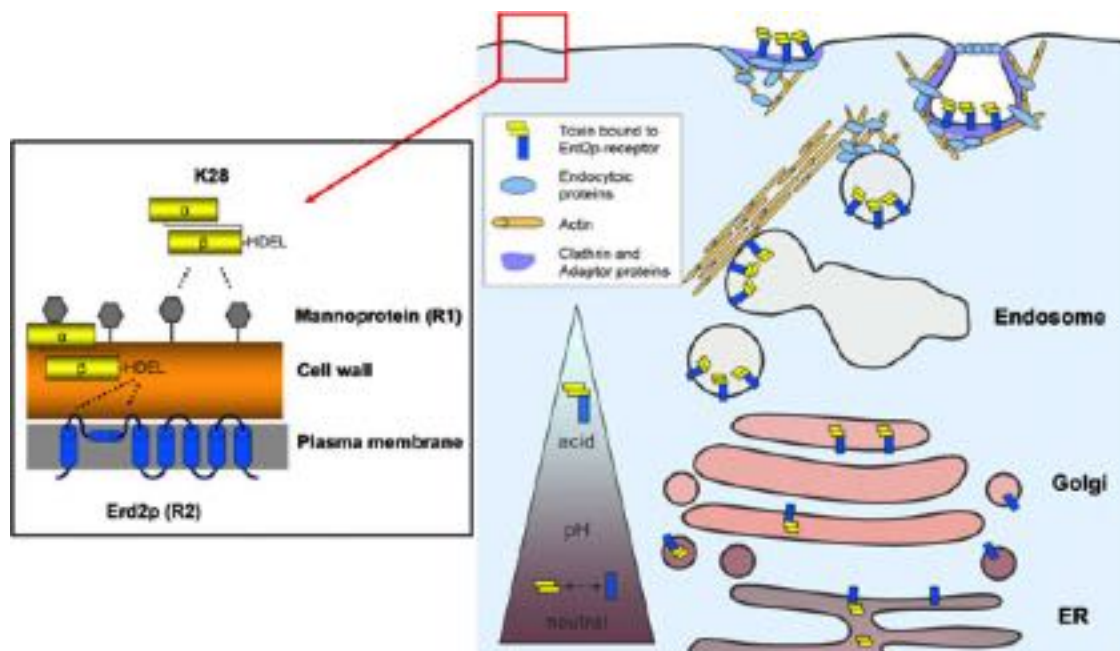


Fig. 4 Schematic model of the endocytotic uptake and retrograde transport of K28 into the ER. In a first step, the β -subunit of the mature K28 α/β heterodimer binds in an energy-independent process to α -1,3- and/or α -1,2-mannotriose side-chains of the primary mannoprotein receptor (R1) in the yeast cell wall (CW). In a second step, K28 crosses the cell wall and subsequently interacts with the yeast HDEL receptor Erd2p (R2) at the plasma membrane level. The C-terminal HDEL motif of the toxin's β -subunit is crucial for the toxin/receptor interaction. Finally, each receptor/toxin complex is internalized by clathrin-mediated endocytosis and retrogradely transported from endosomes through the Golgi apparatus into the ER, where K28 is released from the HDEL receptor and subsequently retrotranslocated into the yeast cytosol. Thereby, K28 binding to and/or release from the HDEL receptor Erd2p critically depends on the pH environment of the respective subcellular compartment; receptor/toxin complexes are stably associated in a more acidic pH environment (pH < 6.8) which is typical for endosomes and Golgi, whereas the complexes dissociate at the neutral pH of the ER lumen (pH > 6.8).

mutant toxin variants with altered inter- and/or intra-subunit disulfide bonding are nontoxic *in vivo* due to the incapability to reach their intracellular target. Thus, disulfide bond formation in microbial and viral A/B toxins is of major importance to ensure interaction competence of the toxins with the H/KDEL-receptor of the particular target cell.

ER Exit and Nuclear Entry of the K28 Virus Toxin

During host cell penetration, K28 retrotranslocates from the ER into the cytosol and dissociates into its subunit components. The β -subunit is subsequently polyubiquitinated and proteasomally degraded while the cytotoxic α -subunit enters the nucleus and causes cell death (Fig. 5). ER exit of the α/β heterodimeric toxin is presumably mediated by the Sec61 complex, termed translocon, which functions as major transport channel in the ER membrane of yeast and higher eukaryotes. In yeast, each translocon resembles a core heterotrimeric complex consisting of the transmembrane protein Sec61p and the two smaller subunits Shh1p and Sss1p. Besides being the major channel for co- and post-translational protein import into the ER, Sec61p is also involved in the export and removal of misfolded and/or mis-assembled proteins from the secretory pathway to initiate their proteasomal degradation in the cytosol. In addition to its central function in protein quality control in the ER, Sec61 is also responsible for ER retrotranslocation of certain plant, microbial and viral A/B toxins such as ricin, cholera toxin, *Pseudomonas* exotoxin A, and the yeast K28 virus toxin. In contrast to microbial and plant A/B toxins, however, retrotranslocation of K28 from the ER lumen is independent of ubiquitination and proteasome activity and classical components normally involved in ER-associated protein degradation (ERAD) are not required for ER exit of this virus toxin. In K28 intoxicated cells, toxin translocation competence in the ER strongly depends on the activity of luminal Hsp70 chaperones (such as Kar2p/ BiP and Pdi1p), Hsp40 cochaperones (such as Scj1p and Jem1p), and additionally requires proper maintenance of calcium homeostasis in the ER (Fig. 5). So far it is not known what cellular component within or near the ER membrane is responsible for toxin exit, but K28 might be a fruitful tool to identify a novel transport pathway for protein transport across the ER membrane.

K28 Affects DNA Synthesis, Cell-Cycle Progression, and Induces Apoptosis

Although the cytotoxic α -component of K28 (10.5 kDa) can enter the nucleus by passive diffusion, extension of α by a classical nuclear localization sequence (NLS) significantly enhances its *in vivo* toxicity due to faster and more efficient nuclear import mediated

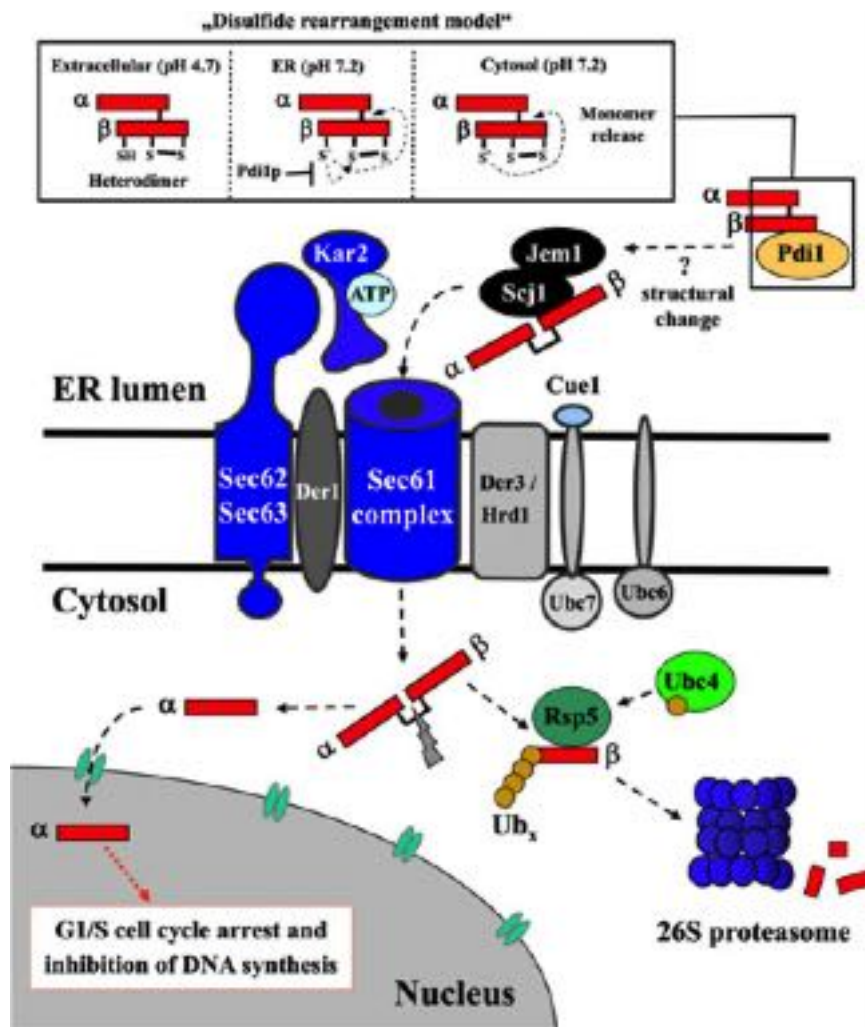


Fig. 5 Retrotranslocation of the K28 virus toxin from the ER and its lethal effect in the nucleus. After endocytotic uptake and retrograde transport via Golgi and ER, the toxin is gated through the Sec61p export channel by the help of luminal ER chaperones such as Pdi1p, Kar2p, Jem1p, and Scj1p. Thereby, Pdi1p plays a major role in the toxin retrotranslocation step. The “disulfide rearrangement model” schematically illustrates how reactivity of the sulfhydryl residues in the β -subunit changes during K28 intoxication. Under acidic pH conditions (pH 4.7), K28 is structurally stable as a disulfide-bonded α/β heterodimer. At neutral pH of the ER lumen, a free sulfhydryl (SH) residue in the β -subunit deprotonates which leads to the formation of an active thiol (S⁻). The nucleophilic attack of the reactive thiol and the heterodimer dissociation is prevented by the activity of Pdi1p, enabling exit of K28 as α/β heterodimer from the ER. In the pH-neutral and Pdi1p-free environment of the cytosol, the nucleophilic attack induces a disulfide bond rearrangement and subsequent release of the monomeric α -subunit from the heterodimer. Thereafter, β is poly-ubiquitinated via Ubc4p (E2) and Rsp5p (E3) and proteasomally degraded while α enters the nucleus and causes cell death. Within the nucleus, the α -toxin interacts with essential host proteins involved in eukaryotic cell-cycle control and causes cell death through G1/S cell-cycle arrest and inhibition of DNA synthesis. Cellular components of the ER quality control system ERAD (such as Cue1p, Ubc7p, Der3p/Hrd1p, Ubc6p, and Der1p) are not involved in ER-to-cytosol export of the K28 virus toxin (ERAD components are shown in gray color). Reproduced and adapted from Leis, S., Spindler, J., Reiter, J., Breinig, F., Schmitt, M.J., 2005. *Saccharomyces cerevisiae* K28 toxin – A secreted virus toxin of the A/B family of protein toxins. In: Schmitt, M.J., Schaffrath, R. (Eds.), *Topics in Current Genetics 11: Microbial Protein Toxins*. Berlin, Heidelberg: Springer, pp. 111–132, with kind permission of Springer Science and business Media.

by α/β importins of the host. Within the nucleus, α interacts with host proteins of essential function in eukaryotic cell cycle control and initiation of DNA synthesis. Thus, as the virus toxin targets evolutionary highly conserved proteins with basic and central function, toxin resistance mechanisms based on mutations in essential chromosomal host genes hardly occur *in vivo*, indicating that the toxin has developed a very efficient strategy to penetrate and kill its target cell. Most interestingly, while high toxin concentrations (10 pmol or higher) cause necrotic cell killing via cell-cycle arrest and inhibition of DNA synthesis, treatment with low doses of viral killer toxins (<1 pmol) results in an apoptotic host-cell response triggered by the accumulation of reactive oxygen species, ROS (Fig. 6). Since toxin concentration is usually low in the natural environment of a killer yeast, toxin-induced apoptosis is probably an important prerequisite for efficient cell killing. Furthermore, since apoptosis is also important in the pathogenesis of virus infections in mammals, it is not surprising that toxin-encoding yeast killer viruses (as has been shown for ScV-M1 and ScV-M2) can also induce

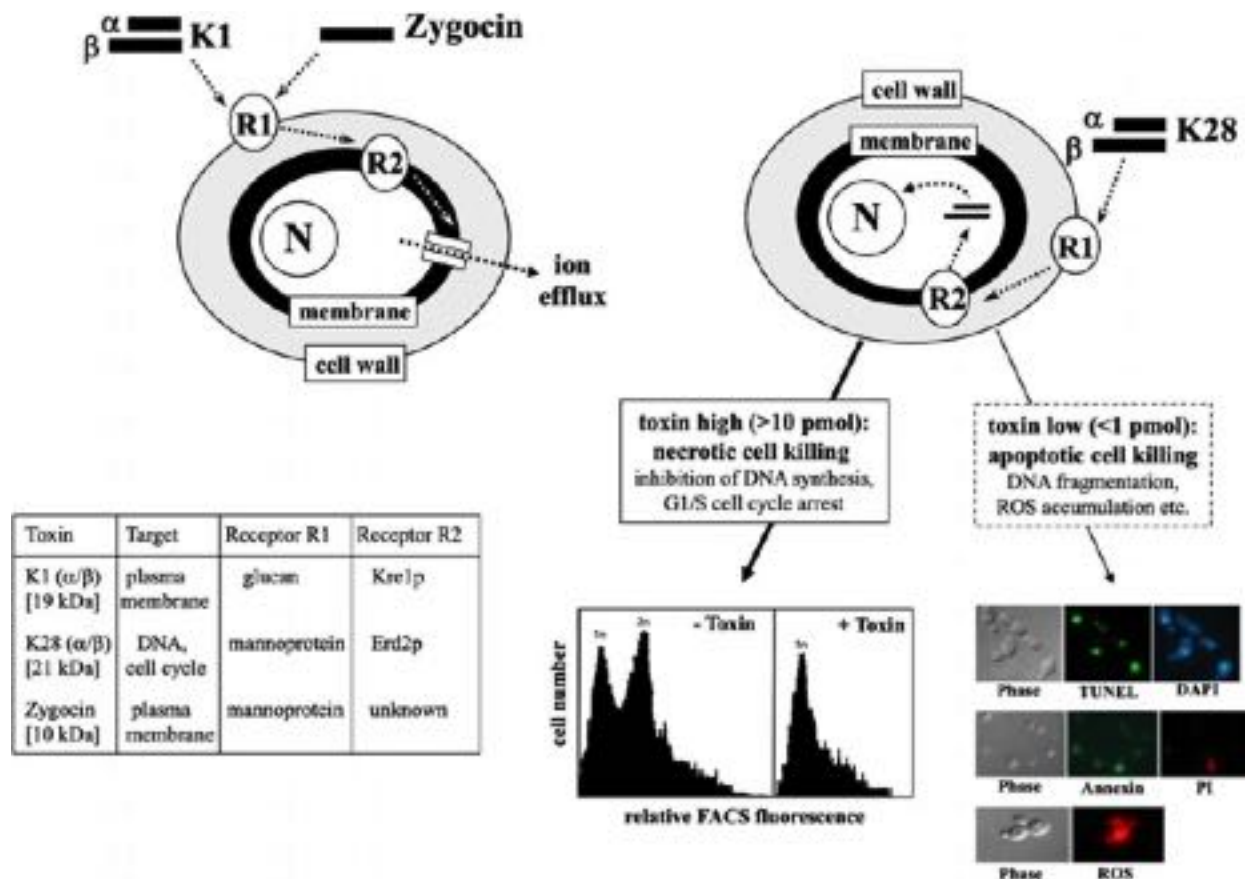


Fig. 6 Receptor-mediated toxicity of the viral killer toxins K1, zygocin, and K28. Killing of a sensitive yeast is envisaged in a two-step process involving initial toxin binding to receptors within the cell wall (R1) and the cytoplasmic membrane (R2). After interaction with the plasma membrane, ionophoric toxins such as K1 and zygocin disrupt cytoplasmic membrane function, while K28 enters the cell by endocytosis and diffuses into the nucleus to cause cell death (note that the cell surface receptors R1 and R2 are different in all three toxins; see also table inset). At high toxin doses (>10 pmol) sensitive cells arrest in the cell cycle with pre-replicated DNA (1n; left panel), while cells treated with K28 in low concentrations (<1 pmol) respond with apoptosis as shown by typical apoptotic markers such as chromosomal DNA fragmentation (TUNEL-positive cells), accumulation of reactive oxygen species (ROS) and phosphatidylserine exposure at the external surface of the plasma membrane detected by annexin-V staining (right panel).

a programmed suicide pathway in noninfected yeast. Although viral killer toxins were shown to be primarily responsible for this phenomenon, yeast killer viruses are not solely responsible for triggering a cell death pathway in yeast.

Lethality of Membrane Damaging Viral Killer Toxins

Yeast viral killer toxins kill sensitive cells in a receptor-mediated fashion by interacting with receptors at the level of the cell wall and the cytoplasmic membrane (Fig. 5). The initial step involves rapid toxin binding to a primary receptor R1 which is localized within the mannoprotein or β -1,6-glucan fraction of the cell wall. In the second step the toxin translocates to the plasma membrane and interacts with a secondary receptor R2 (Fig. 6). To date, only the membrane receptors for K1 and K28 have been identified. Kre1p, an O-glycosylated cell surface protein initially GPI-anchored to the plasma membrane and involved in β -1,6-glucan biosynthesis and K1 cell wall receptor assembly, was identified as plasma membrane receptor of killer toxin K1. In case of K28, the HDEL receptor Erd2p has been demonstrated to be responsible for both efficient K28 binding to and endocytotic uptake from the yeast plasma membrane. Once bound to the plasma membrane, ionophoric virus toxins (such as K1 and zygocin) disrupt cytoplasmic membrane function by forming cation-selective ion channels, while K28 enters the cell and acts in the nucleus: DNA synthesis is rapidly inhibited and cells arrest at the G1/S boundary of the cell cycle (Fig. 6).

Ion channel formation in yeast membranes induced by the K1 virus toxin was initially reported using patch-clamping techniques as a result from direct toxin action. However, this observation is inconsistent with the complete resistance seen in immune yeast cell spheroplasts, and so far receptor-independent channels have been observed, neither in yeast membranes nor in *Xenopus laevis* oocytes. Similar to K1 and K2, zygocin represents a membrane-damaging virus toxin which is produced and secreted

by ZbV-M infected killer strains of the osmotolerant spoilage yeast *Z. bailii*. Zygocin itself is a monomeric nonglycosylated protein toxin with an unusual broad killing spectrum, being equally active against phytopathogenic as well as human pathogenic yeasts including *Candida albicans*, *C. glabrata*, *C. tropicalis*, and *Sporothrix schenckii*. Since even filamentous fungi such as *Fusarium oxysporum* and *Colletotrichum graminicola* are effectively killed by the toxin, zygocin represents a virus toxin with significant antimycotic potential. Similar to K1 and K2 but significantly more efficient, zygocin disrupts plasma membrane integrity and causes rapid cell killing. Its ionophoric mode of action has been reinforced by *in silico* sequence analysis, identifying a stretch of potential α -helical conformation that forms an amphipathic structure characteristic for membrane-disturbing antimicrobial peptides such as alamethicin, melittin, and dermaseptin. In addition, this feature is accompanied by a transmembrane helix at the C-terminus of zygocin which is predicted to favor a membrane permeabilizing potential, not by activating native ion channels but rather by establishing pores by itself after toxin oligomerization. It is therefore assumed that the hydrophobic part in zygocins' amphipathic α -helix is responsible for toxin binding to the target cell. The postulated model of zygocin action resembles that of human α -defensins. In analogy to alamethicin, toxicity of zygocin is probably mediated by incorporation of its transmembrane helix into the plasma membrane, a process solely driven by the natural transmembrane potential of the energized yeast and fungal plasma membrane. Thus, zygocins' mode of action portrays the lethal mechanism of antimicrobial peptides that are produced by virtually all higher eukaryotes. Mechanisms of resistance against antimicrobial peptides are rare and often limited to changes in the composition of the cytoplasmic membrane. In major contrast to mammalian cells, the outer leaflet of yeast and fungal membranes is enriched in negatively charged lipids. Due to the cationic net charge of antimicrobial peptides (including zygocin), an affinity to these lipids facilitates toxin adsorption to the target membrane. Consistent with that, deletion of chromosomal genes whose gene products affect plasma membrane lipid composition (such as *PDR16* and *PDR17*) causes a dramatic decrease in zygocin sensitivity because toxin binding to the plasma membrane is largely prevented in the genetic background of a yeast *pd16/17* mutant. In contrast to K1, a zygocin-specific membrane receptor is not required for its *in vivo* toxicity as the physicochemical properties of zygocin allow efficient plasma membrane interaction independent of any membrane receptor or docking protein. Although the more recently discovered killer toxins Klus and Kbarr-1 likewise show an interesting broad killing spectrum similar to that of zygocin (i.e., killing yeasts such as *S. cerevisiae*, *Hanseniaspora* sp., *Cluyveromyces lactis*, *Schizosaccharomyces pombe*, *Candida albicans*, *C. tropicalis*, *C. dubliniensis*, *C. kefir*, *C. glabrata*, *C. parasitopsis*, *C. krusei*, *Yarrowia lipolytica*, and *Hansenula mrakii*), their precise mode and side of action is completely unknown and needs further investigation.

Self-Protection in Killer Virus-Infected Yeast – Toxin Immunity

For many decades it was unknown how a killer virus-infected yeast protects itself against its own secreted toxin. In killer yeast, functional immunity is essential for survival since the toxins often target and inhibit central eukaryotic cell functions. This is in major contrast to bacterial toxins such as cholera toxin and Shiga toxins which selectively kill eukaryotes, thus making immunity dispensable in a prokaryotic host. So far, only the mechanism of protecting immunity against the K28 virus toxin has been elucidated. It has been shown that K28 killer cells take up external toxin (either produced by itself or by other K28 killers) and translocate it back to the cytosol where the reinternalized α/β toxin rapidly forms a complex with the pptox precursor that has not yet been imported into the ER. Within this complex, the K28 heterodimer is selectively ubiquitinated and proteasomally degraded while the pptox moiety of the complex is in part released to be either imported into the ER (to give active virus toxin) or to complex a newly internalized K28 heterodimer. In this process, the amount of cytosolic ubiquitin is critical for immunity and overexpression of mutant ubiquitin (blocked in polyubiquitin chain formation) results in a significant decrease in toxin secretion and a suicidal phenotype based on nonfunctional immunity. Alternatively, decreasing cytosolic ubiquitin causes an increase in toxin secretion, while immunity is not impaired as sufficient pptox is available for K28 complex formation. This simple and highly efficient mechanism ensures that a toxin-producing killer yeast is fully protected against the lethal action of its own toxin. In contrast to K28 immunity, the precise mechanism(s) of self-protection against the ionophoric virus toxins K1, K2 and zygocin is still obscure and remains largely unknown.

Further Reading

- Becker, B., Blum, A., Giesselmann, E., *et al.*, 2016. H/KDEL receptors mediate host cell intoxication by a viral A/B toxin in yeast. *Scientific Reports* 6, 31105. doi:10.1038/srep31105.
- Becker, B., Schmitt, M.J., 2017. Yeast killer toxin K28: Biology and unique strategy of host cell intoxication and killing. *Toxins* 9. doi:10.3390/toxins9100333.
- Breinig, F., Sendzik, T., Einfeld, K., Schmitt, M.J., 2006. Dissecting toxin immunity in virus-infected killer yeast uncovers an intrinsic strategy of self-protection. *Proceedings of the National Academy of Sciences of the United States of America* 103, 3810–3815.
- Bruenn, J.A., 2002. The double-stranded RNA viruses of *Ustilago maydis* and their killer toxins. In: Tavantzis, S.M. (Ed.), *DsRNA Genetic Elements: Concepts and Applications in Agriculture, Forestry, and Medicine*. Boca Raton, FL: CRC Press, pp. 109–124.
- El-Sherbeini, M., Bostian, K.A., 1987. Viruses in fungi: Infection of yeast with the K1 and K2 killer virus. *Proceedings of the National Academy of Sciences of the United States of America* 84, 4293–4297.
- Ivanovska, J., Hardwick, J.M., 2005. Viruses activate a genetically conserved cell death pathway in a unicellular organism. *Journal of Cell Biology* 170, 391–399.
- Ramirez, M., Velazquez, R., Maqueda, M., Lopez-Pineiro, A., Ribas, J.C., 2015. A new wine *Torulaspora delbrueckii* killer strain with broad antifungal activity and its toxin-encoding double-stranded rna virus. *Frontiers in Microbiology* 6, 983.
- Reiter, J., Herker, E., Madeo, F., Schmitt, M.J., 2005. Viral killer toxins induce caspase-mediated apoptosis in yeast. *Journal of Cell Biology* 168, 353–358.

- Rodriguez-Cousino, N., Maqueda, M., Ambrona, J., *et al.*, 2011. A new wine *Saccharomyces cerevisiae* killer toxin (Klus), encoded by a double-stranded RNA virus, with broad antifungal activity is evolutionarily related to a chromosomal host gene. *Applied and Environmental Microbiology* 77, 1822–1832.
- Ruggiano, A., Foresti, O., Carvalho, P., 2014. Quality control: ER-associated degradation: Protein quality control and beyond. *Journal of Cell Biology* 204, 869–879.
- Schmitt, M.J., Breinig, F., 2006. Yeast viral killer toxins: Lethality and self-protection. *Nature Reviews Microbiology* 4, 212–221.
- Schmitt, M.J., Neuhausen, F., 1994. Killer toxin-secreting double-stranded RNA mycoviruses in the yeasts *Hanseniaspora uvarum* and *Zygosaccharomyces bailii*. *Journal of Virology* 68, 1765–1772.
- Suzuki, Y., Schwartz, S.L., Mueller, N.C., Schmitt, M.J., 2017. Cysteine residues in a yeast viral A/B toxin crucially control host cell killing via pH-triggered disulfide rearrangements. *Molecular Biology of the Cell* 28, 1123–1131.
- Weiler, F., Schmitt, M.J., 2005. Zygocin – A monomeric protein toxin secreted by virus-infected *Zygosaccharomyces bailii*. In: Schmitt, M.J., Schaffrath, R. (Eds.), *Microbial Protein Toxins*. Berlin Heidelberg: Springer, pp. 175–187.
- Wickner, R.B., 1996. Double-stranded RNA viruses of *Saccharomyces cerevisiae*. *Microbiology Reviews* 60, 250–265.

Alternaviruses (Unassigned)

Hiromitsu Moriyama, Tokyo University of Agriculture and Technology, Tokyo, Japan

Nanako Aoki, Kuko Fuke, Kana Takeshita Urayama, Naoki Takeshita, and Chien-Fu Wu, Tokyo University of Agriculture and Technology, Fuchu, Japan

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Mycoviruses Viruses that infect and propagate fungi.

Polyadeniration Addition of a poly (A) tail to 3' UTR of mRNA or viral plus-stranded RNA. In eukaryotes, polyadenylation is part of the process that produces mature mRNA for translation.

Rapid amplified of cDNA ends (RACE) This technique can provide the sequence of an RNA transcript to the 5' end (5' RACE-PCR) or 3' end (3' RACE-PCR) of the RNA.

RNA-dependent RNA polymerase (RdRP) Enzyme that catalyzes the replication of RNA from an RNA template.

Viruses with multipatite genomes The essential genome is divided among several genomic segments (segmented genome) that are either separately encapsidated in identical capsids.

Introduction

Fungi are infected by viruses, named as mycovirus. Recently, numerous mycoviruses are reported to cause epigenetic phenomena. Historically, mycoviruses were found in diseased *Agaricus bisporus* mushroom or *Aspergillus foetidus*, human pathogenic fungi over 50 years ago. Mycoviruses are primarily classified into four groups based on their genomic structure: single-stranded DNA (ssDNA), positive single-stranded RNA (+ ssRNA), negative single-stranded RNA (– ssRNA), and double-stranded RNA (dsRNA). In general, mycoviruses with dsRNA genomes are classified into six families based on the amino acid sequence of the RNA-dependent RNA polymerase (RdRp), the genomic structure, the virion structure. These families are: the *Totiviridae* (nonsegmented genome), the *Partitiviridae* (2 segments), the *Megabirnaviridae* (2 segments), the *Chrysoviridae* (3–5 segments), the *Quadriviridae* (4 segments), and the *Reoviridae* (9–12 segments) (Ghabrial *et al.*, 2015). Reovirus genome consists of 10–12 segments of dsRNA; each of the dsRNA segments has m7GpppNm cap structure at the 5'-end, and lack a 3' poly (A) tail. Partitiviruses genomes generally consist of 2 segments of dsRNA; each of the dsRNA segments has interrupted poly (A) tail at the 3'-end of the coding strands.

Alternavirus has 3–4 dsRNA segments; each of the dsRNA segments has intact stretched poly (A) tail at the 3'-end of the coding strands, which trait is a unique molecular property as segmented double-stranded RNA viruses. Although the number of reports on mycovirus articles has been increasing in recent years, there are only five reports describing alternaviruses. Here, we introduce intriguing characteristic properties of alternavirus, primarily focusing on *Alternaria alternata* virus 1 (AaV1) and *Aspergillus foetidus* dsRNA mycovirus (AfV-F), which are similar in size to dsRNA viruses in the families *Totiviridae*, *Partitiviridae*, and *Chrysoviridae*, and while the genomic structures of the dsRNA viruses in three families are completely different. Based on our findings for AaV1, AfV-F, and other three alternaviruses, we propose a new family, “*Alternaviridae*”, to accommodate them.

Genome Organization

The virions of AaV1 and AfV-F contain four unrelated linear, separately encapsidated, monocistronic dsRNA segments (1.4–3.6 kbp in size; Table 1). The largest segment, dsRNA1 encodes for the RdRp. DsRNA2, and dsRNA3 might encode for the major capsid protein (CP). The dsRNA segment 4 encodes for proteins of unknown function. The genomic structure of AaV1, type species of the genus *Alternaviridae*, and that of AfV-F are schematically represented in Fig. 1. It is noteworthy that all of the coding strands of the four dsRNA genomes have 3'-poly (A) tails ranging from 36 to 50 nt in length and stretches of poly (U) at the 5' terminus of the non-coding strand in all of the four dsRNAs was detected by 5'-RACE.

In the AaV1, both the 5'- and 3'-terminal sequences of all four dsRNAs are significantly conserved. Even in AfV-F, the 5'- and 3'-terminal sequences are highly conserved between the four RNA segments. The nucleotide sequences preserved in each are different between AaV1 and AfV-F (Fig. 2). These terminal sequences may be involved in the replication cycles of these dsRNAs depending on each RdRp of AaV1 and AfV-F. There is no evidence that these four dsRNAs are transcripts from host genomic DNA, which proved by southern hybridization experiments carried out with cDNA probes derived from the four dsRNAs of AaV1 in the EGS 35–193, *Alternaria alternata*.

Virion Properties

The buoyant densities of virions of AaV1 is in the range of 1.35–1.40 g/cm³ in CsCl. After treatment with chloroform–butanol followed by PEG precipitation, virus particles were purified by CsCl density equilibrium centrifugation. AaV1 virions are isometric, about 33 nm in diameter, and are similar in size to those of dsRNA viruses in the families *Totiviridae*, *Partitiviridae*, or *Chrysoviridae*.

Table 1 Member of alternavirus

Host	Virus	acronym	DsRNA segm entno. (length in bp, encoded prote in: size in kDa)	GeneBank accession no.
<i>Alternaria alternata</i>	<i>Alternaria alternata</i> virus 1	AaV1	1 (3613: RdRp, 129) 2 (2795:P2, 90) 3 (2576:P3, 82) 4 (1420:P4, 40)	AB 368492 AB 438027 AB 438028 AB 438029
<i>Aspergillus foetidus</i>	<i>Aspergillus foetidus</i> mycovirus	AfV-F	1 (3588: RdRp, 127) 2 (2770:P2, 87) 3 (2466:P3, 79) 4 (2005:P4, 65)	NC_020103.1 HE588145 HE588146 HE647818
<i>Aspergillus niger</i>	<i>Aspergillus</i> mycovirus 341	AsV341	1 (3588: RdRp, 127)	ABX79997
<i>Fusarium poae</i>	<i>Fusarium poae</i> alternavirus 1	FpAV1	1 (3559: RdRp, 126)	NC_030883.1
<i>Fusarium gaminearum</i>	<i>Fusarium gaminearum</i> alternavirus 1	FgAV1	1 (3524: RdRp, 126)	NC_036596.1

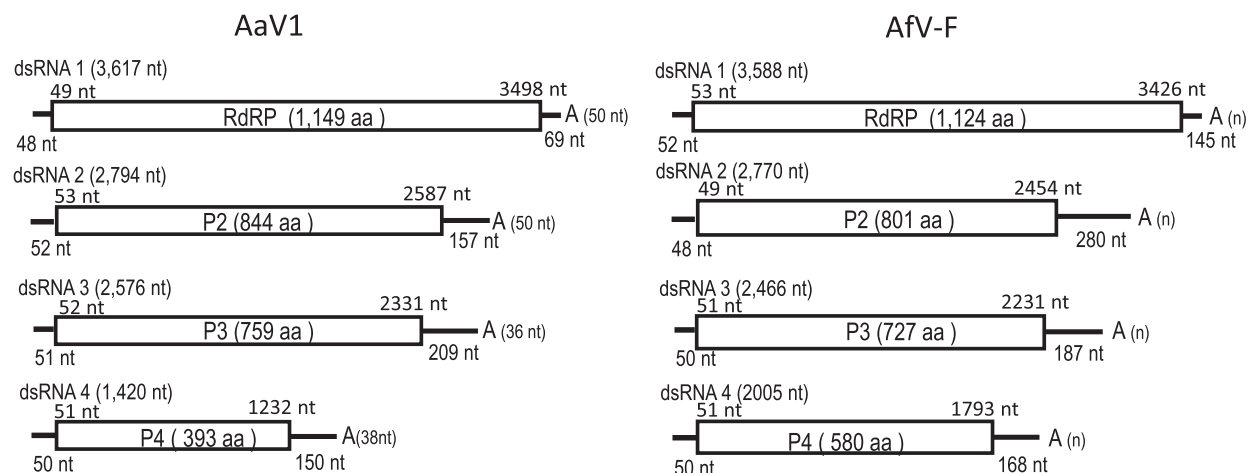


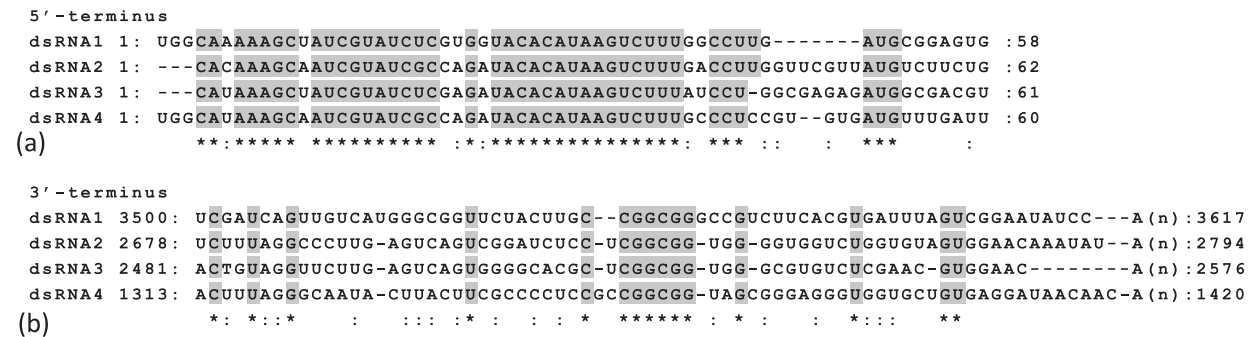
Fig. 1 Genome organization of two alternaviruses. *Alternaria alternata* virus 1 (AaV1), and *Aspergillus foetidus* dsRNA mycovirus (AfV-F). The genome consists of four dsRNA segments; each monocistronic. The RdRp ORFs (nt 49–3498 on AaV1 dsRNA1, nt 53–3426 on AfV-F dsRNA1), the P2 ORFs (nt 53–2587 on AaV1 dsRNA2, nt 49–2454 on AfV-F dsRNA2), the P3 ORFs (nt 52–2331 on AaV1 dsRNA3, nt 51–2231 on AfV-F), the P4 ORFs (nt 51–1232 on AaV1 dsRNA4, nt 51–1793 on AfV-F).

Generally, each virion contains only one of the four genomic dsRNA segments. The four dsRNA segments are packaged separately in individual virus particles, with different buoyant densities; the dsRNA 1 segment is mainly detected in the fraction of buoyant density 1.40 g/cm³, while dsRNA 2 and dsRNA 3 segments are detected in the fractions of buoyant density 1.38–1.39 g/cm³, the dsRNA 4 segment is detectable in the fraction of buoyant density 1.35 g/cm³, although it is barely detectable. Therefore, the dsRNA-containing fractions varied depending on the size of the packaged dsRNA segments in the particles. After purification by CsCl density equilibrium centrifugation, the dsRNA-containing fractions were used for observation of virus particles by electron microscopy. Isometric virus particles with a diameter of about 33 nm were routinely observed (Fig. 3(a)). Approximately 97 kDa protein was detected as the major band, which might be coat proteins encoded by dsRNA 2 and dsRNA 3 (Fig. 3(b)).

Biological Effects of Alternavirus on Their Host

The AaV1 infected strain EGS 35–193 (ATCC), *Alternaria alternata* showed an abnormal phenotype, including reduced mycelial growth, aerial mycelial collapse, and unregulated pigmentation. In order to investigate whether the high concentration of AaV1 dsRNAs are responsible for this impaired growth, we attempted to eliminate or reduce the dsRNAs from the host fungus strain EGS 35–193. By exposing the strain to cycloheximide during hyphal tip isolation, we obtained an isolate E118, in which the amounts of the dsRNAs were reduced. In the isolate E118, the amount of dsRNA was reduced to about one tenth as compared to the parent strain, EGS 35–193. The isolate E118 restored normal mycelial growth and pigmentation, suggesting that the high titer of AaV1 caused attenuated mycelial growth of host fungus, strain EGS 35–193, *Alternaria alternata*. After inoculation on PDA agar medium for 24 h, regular-shaped mycelia were observed in the low titer of the isolate E118, but irregular and poor mycelia were observed in the original high titer strain

AaV1



AfV-F

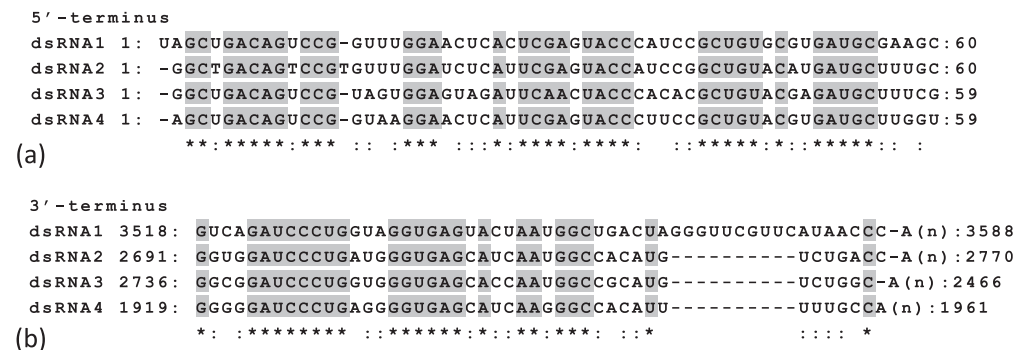


Fig. 2 Comparison of the 5' and 3' UTRs of the four dsRNA segments of AaV1 and AfV-F. Multiple alignments were obtained using CLUSTAL XII (and some manual adjustments) with the nucleotide sequences of the 5' UTR (a) and the 3' UTR (b). Asterisks signify identical bases at the indicated position (shaded) and colons specify that three out of four bases are identical at the indicated positions.

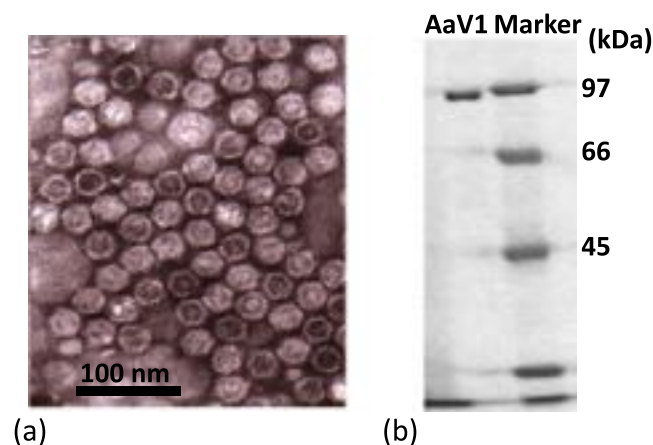


Fig. 3 (a) Negative contrast electron micrograph of particles of *Alternaria alternata* virus 1. (b) Purified AaV1 proteins stained with Coomassie Brilliant Blue.

EGS35-193. In addition, abnormally enlarged vesicles appeared in mycelial cells of EGS 35-193. Bursting of these enlarged vesicles was observed around the hyphal cells. Only a few small vesicles were observed in cells of the low titer isolate E118. These results indicate that the reduced copy number of the dsRNAs in the restored isolates might be responsible for restoring normal morphology.

Proteins Encoded by Alternaviruses

The large dsRNA segment (dsRNA1) of alternaviruses so far sequenced contains a single large ORF coding for RdRp. A multiple alignment of the amino acid sequences of the putative RdRp encoded by the AaV1 dsRNA1 with RdRps of AfV-F, AsV341, FpAV1, and FgAV1 is shown in Fig. 4. Eight conserved motifs characteristic of RdRps of dsRNA viruses of simple eukaryotes were found in

	I	II	III
AaV1	T C R Q H I F S A N F L N A E D L T G W S D (80)	R M A W M S G G A V G R R A K E L L G (36)	K G N E R G S E R T L L A T D L R D Q V S E Y L L (46)
AfV-F	T N R Q H I A S A Y F L N A E D L N G F S D (68)	R S A W A A G G V A G R R S R D V L G (36)	K L D E R G V D R T L S A T D M R D Q T A E S Y V F (45)
AsV341	T N R Q H I A S A Y F L N A E D L N G F S D (68)	R S A W A A G G V A G R R S R D V L G (36)	K L D E R G V D R T L S A T D M R D Q T A E S Y V F (45)
FpAV1	T N L Q H I A S A Y F L N A E D L T G F A D (68)	R T A W A A G G V A G R R A R D V L G (36)	K M D E R G P P R C I Q A T D M R D Q T A E T Y V F (45)
FgAV1	T N L Q H I A S A Y F L N A E D L T G F A D (68)	R T A W A A G G V A G R R A R D V L G (36)	K M D E R G P P R C I Q A T D M R D Q T A E T Y V F (45)
	* : * * : * : * : * * * * * * : *	* * * : * * : * * * : * : * *	* : * * * * * * * * : * * * : * * : *

	IV	V	VI	VII	VIII
AaV1	L A A W D Y S K W D H (92)	G Q Q S G R R S T L E S N T F (23)	L N R A D D V M E I (21)	A N K K K Q V V Q V R T G V Y F R I (10)	F P P R A V Y A C A S A G P
AfV-F	L V A W D W S K W D H (86)	G Q Q S G R R T T L E V N T I (29)	L N R A D D V A E V (21)	A N P K K Q V A Q W R S I V Y F R I (10)	F P P R A V Y A A A S G H P
AsV341	L V A W D W S K W D H (86)	G Q Q S G R R T T L E V N T I (29)	L N R A D D V A E V (21)	A N P K K Q V A Q W R S I V Y F R I (10)	F P P R A V Y A A A S G H P
FpAV1	L T A W D W Q K W D H (90)	G Q Q S G R K T T L E A N T I (29)	L N R A D D V A E V (21)	A N P K K Q V A Q W R S V V Y L R I (10)	F P P R A V Y A A A S G H P
FgAV1	L T A W D W Q K W D H (90)	G Q Q S G R K T T L E A N T I (29)	L N R A D D V A E V (21)	A N P K K Q V A Q W R S V V Y L R I (10)	F P P R A V Y A A A S G H P
	* * * : * * * *	* * * * * : * * * * :	* * * * * : *	* * : * * * : * : * * * *	* * * * * : * * : *

Fig. 4 Multiple alignment of RdRp amino acid sequences from AaV1, AfV-F, AsV341, FpAV1, and FgAV1. The alignment was performed using Clustal X ver. 2.0 and trimmed to include the 8 conserved core regions of RdRp (indicated by solid lines), using Seaview ver. 4.0. Asterisks signify identical residues (shaded) at the indicated positions; colons signify highly conserved amino acid residues within a column; numbers in parentheses correspond to the number of amino acid residues separating the motifs.

the ORFs of the dsRNA1 in the five alternaviruses. The two putative RdRps encoded by AfV-F and AsV341 are very closely related to each other (identity: 99.6%). The two putative RdRps encoded by FpAV1 and FgAV1 are also very closely related to each other (identity: 98%). The RdRp encoded by AaV1 exhibits lower sequence similarities with that of AfV-F (identity: 36%), or that of FpAV1 (identity: 34%), while the identity between the RdRps of AfV-F and FpAV1 is 47%.

Although the putative RdRps encoded by the five alternaviruses appear to be closely related to each other, they are distantly related to members of any dsRNA virus family, including the *Totiviridae*, *Chrysovriidae*, and *Partitiviridae*, where even in the most conserved GDD motif, the glycine (G) residue is replaced by an alanine (A) in cases of the five alternaviruses (Fig. 4).

Phylogenetic analyses of their putative RdRp gene reveal that the five alternaviruses revealed distinct clade apart from the three families, *Totiviridae*, *Chrysovriidae*, and *Partitiviridae* (Fig. 5). Indeed, the genome structures of the dsRNA viruses in the three families are completely different. Alternaviruses have four dsRNA segments of 1.5–3.6 kbp, which are similar in size to dsRNA viruses in the family *Chrysovriidae* that possess four dsRNA segments of 2.4–3.6 kbp. In contrast, viruses in the *Totiviridae* and *Partitiviridae* families have single or two dsRNA segments, respectively. Thus, the alternaviruses appear to be evolutionarily related to but not members of the family *Chrysovriidae*.

3' Poly (A) Structure of Alternaviruses

It is noteworthy the coding strands of genomic dsRNAs in the alternaviruses contain 3' poly (A) tails, which are supposed to range from 33 to 50 nt in length. There was a possibility that a single-stranded (ss) form of dsRNAs was present in larger amounts than double-stranded (ds) form. In order to investigate whether the ratio of the AaV1 ssRNA to dsRNA form, northern blot analyses using three probes types were performed to determine the relative abundance of viral ssRNA and dsRNA; hybridization experiments were performed with probes for positive-strand RNA, negative-strand RNA, or both strands (dsDNA probes). If the positive-strand RNA were present in much greater quantities than the dsRNA, the signals detected by the positive-strand-specific probe should have been stronger than those detected by the negative-strand-specific probe or by a DNA probe that can detect both strands. However, there was no significant difference in the intensities of signals detected by the three probes, indicating that little single-stranded form of dsRNAs exists in the alternavirus, and most of the dsRNAs are present as genomic RNA rather than the replicative form of the ssRNA genome (Table 2).

Evolutionary Relationships Among Chrysoviruses

In recent years, the discovery of many mycoviruses has been reported. Historically, totiviruses and partitiviruses have been found widely not only fungi such as ascomycetes, yeast and mushrooms belonging to basidiomycetes such as *Agaricus* but also other simple eukaryotic organisms such as protozoa or oomycetes. Chrysoviruses discovered from *Penicillium chrysogenum* with four double-stranded RNAs were classified by Gabriel as members of the family *Chrysovriidae* independent from *Partitiviridae*, in EoV

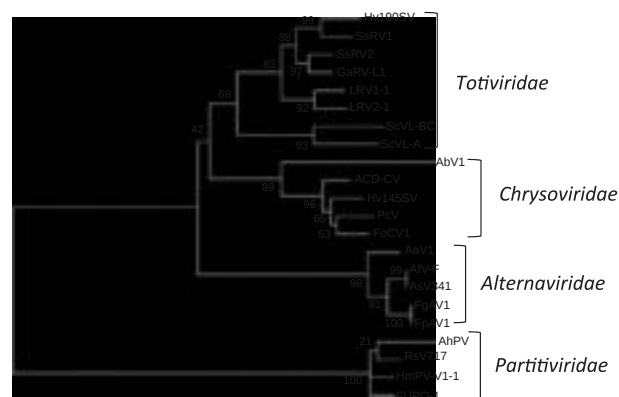


Fig. 5 A phylogenetic analysis of the RdRp sequences of Alternaviruses and selected mycoviruses in the *Totiviridae*, *Chrysoviridae*, *Partitiviridae* families. An unrooted phylogenetic tree based on the neighbor-joining method was performed using the program MEGA 10. The numbers at nodes represent bootstrap values as percentages estimated by 100 replicates. The following viruses in the family *Totiviridae* were included in the phylogenetic analysis (abbreviations in parenthesis): *Helminthosporium victoriae* virus 190S (HmV190S); *Sphaeropsis sapinea* RNA virus 1 (SsRV1); *Sphaeropsis sapinea* RNA virus 2 (SsRV2); *Giardia lamblia* virus L1 (GaRV-L1); *Leishmania* RNA virus 1-1 (LRV1-1); *Leishmania* RNA virus 2-1 (LRV2-1); *Saccharomyces cerevisiae* virus L-A (ScV-L-A); *Saccharomyces cerevisiae* virus L-BC (ScV-L-BC). The following viruses in the family *Chrysoviridae* were included in the phylogenetic analysis (abbreviations in parenthesis): *Agaricus bisporus* virus 1 (AbV1); *Amasya cherry disease associated chrysovirus* (ACD-CV); *Helminthosporium victoriae* virus 145S (HmV145S); *Penicillium chrysogenum* virus (PcV); *Fusarium oxysporum* chrysovirus 1 (FoCV1). The following viruses in the family *Alternaviridae* were included in the phylogenetic analysis (abbreviations in parenthesis): *Alternaria alternata* virus 1 (AaV1); *Aspergillus foetidus* dsRNA mycovirus (AaV-F); *Aspergillus mycovirus* 341 (AsV341); *Fusarium poae* alternavirus 1 (FpAV1); *Fusarium graminearum* alternavirus 1 (FgAV1). The following viruses in the family *Partitiviridae* were included in the phylogenetic analysis (abbreviations in parenthesis): *Atkinsonella hypoxylon* partitivirus (AhPV); *Rhizoctonia solani* virus 717 (RsV717); *Helicobasidium mompa* partitivirus V1-1 (HmPV-V1-1); *Fusarium poae* virus 1 (Fupo-1).

Table 2 Comparison between AaV-1 and dsRNA viruses in three mycovirus families

Virus family (virus name)	Number of dsRNA genome	Segments size (kbp)	Virus particle	3' poly (A)	5' Cap
Alternavirus	4	1.5–3.6	isometric 33 nm	33–50 nt	
Totivirus	1	4.6–6.7	isometric 40 nm	not found	not found
Partitivirus	2	1.4–2.2	isometric 33 nm	20–30 nt	not found
Chrysovirus	4	2.4–3.6	isometric 35 nm	not found	not found
Megabirnavirus	2	6.5–7.0	isometric 50 nm	not found	not found
Reovirus	10 ~ 12	1.1–3.5	isometric 60 nm	not found	found

3rd edition, 2008. The viruses of these families and the proposed family Alternaviridae can be mixed infectiously in one host cell. To date, mycoviruses infecting these hosts have been discovered from more than hundreds of simple eukaryotes in more than 20 families. However, there are only five reports of the alternaviruses as hosts of three saprophytes, namely the genera *Alternaria* spp., *Fusarium* spp., and *Aspergillus* spp.

As one of the characteristics of the alternaviruses, it is to have intact poly (A) structures at the 3' ends of their dsRNA genomes, which is the first discovery as segmented double-stranded RNA viruses. In AaV1, the poly (A) structure at the 3' end has been proved to exist also in the double-stranded RNA genome extracted from within the purified virus particles. Recent studies by the authors have also provided data supporting the presence of a cap structure on the 5' ends of the dsRNA genomes.

Reference

Ghabrial, S.A., Castón, J.R., Jiang, D., Nibert, M.L., Suzuki, N., 2015. 50-plus years of fungal viruses. *Virology* 479–480, 356–368.

Further Reading

Aoki, N., Moriyama, H., Kodama, M., et al., 2009. A novel mycovirus associated with four double-stranded RNAs affects host fungal growth in *Alternaria alternata*. *Virus Research* 140, 179–187.

Hammond, T.M., Andrews, M.D., Roossinck, M.J., Keller, N.P., 2008. *Aspergillus* mycoviruses are targets and suppressors of RNA silencing. *Eukaryotic Cell* 7, 350–357.

He, H., Chen, X., Li, P., Qiu, D., Guo, L., 2018. Complete genome sequence of a *Fusarium graminearum* double-stranded RNA virus in a newly proposed family, Alternaviridae. *Genome Announcements* 6, e00064-18.

Kozlakidis, Z., Herrero, N., Ozkan, S., et al., 2013. Sequence determination of a quadripartite dsRNA virus isolated 135 from *Aspergillus foetidus*. *Archives of Virology* 158, 267–272.

Osaki, H., Sasaki, A., Nomiyama, K., Tomioka, K., 2016. Multiple virus infection in a single strain of *Fusarium poae* shown by deep sequencing. *Virus Genes* 52, 835–847.

Barnaviruses (*Barnaviridae*)

Peter A Revill, The Peter Doherty Institute of Infection and Immunity, Royal Melbourne Hospital, Melbourne, VIC, Australia

© 2021 Published by Elsevier Ltd.

Glossary

Casing A layer of peat moss placed on top of the growing beds to encourage sporophore formation.

Introduction

Mushroom bacilliform barnavirus is the type species and still only member of the *Barnaviridae*. In the U.S.A in 1948, a disease of the cultivated mushroom *Agaricus bisporus* was discovered on a property in Pennsylvania that had a major impact on the mushroom industry and fungal pathology in general. It was characterized by poorly colonised mycelium and misshapen fruiting bodies (sporophores) with long thin stipes and small globular caps producing a drumstick-like appearance, or they were thickened with a barrel-like appearance. The poor colonization of the compost and casing by infected mycelium often produced characteristic bare patches on the growing beds and reduced yields. The disease was named La France disease and a virus was implicated as a possible cause in 1960, after it was shown that the disease could be transmitted to healthy cultures by hyphal anastomosis. In 1962, 3 different virus-like particles were identified, two of which were spherical (25 nm and 29 nm), and the third was a 19 nm × 50 nm elongated or bacilliform particle with rounded ends. Subsequently a 34–36 nm spherical particle with a dsRNA genome (La France infectious virus, LFIV) has been identified as the causal agent of La France disease.

The bacilliform virus particle was of particular interest as almost all mycoviruses identified to that point had a spherical or isometric morphology. 19 nm × 48 nm bacilliform virus-like particles were subsequently identified in the ascomycete *Microsphaera mougeotti*, and 17 nm × 35 nm bacilliform particles were also observed in the deuteromycete *Verticillium fungicola*, itself a pathogen of *A. bisporus*. However no relationship with mushroom bacilliform virus (MBV) and these bacilliform particles has been established.

Originally named mushroom virus 3 (MV3), the virus was subsequently renamed mushroom bacilliform virus (MBV). The viral genome was identified as single-stranded (ss), positive-sense RNA, and the virus was classified by the International Committee on Taxonomy of Viruses (ICTV) as the exemplar virus of the species *Mushroom bacilliform barnavirus* (genus *Barnavirus*, Family *Barnaviridae*). The family name derives its roots from *Bacilliform RNA virus*. MBV remains the only barnavirus identified to date.

MBV remains the only barnavirus characterized to date, however next generation sequencing and metagenomics studies have recently identified another virus with similarity to MBV. Analysis of RNA sequences isolated from the plant pathogenic fungus *Rhizoctonia solani* identified a near-complete viral sequence that grouped closely with MBV on a phylogenetic tree, had a similar genome arrangement, albeit lacking an ORF1 suggesting the sequence was incomplete, and shared amino acid sequence identity, particularly in ORF3, the viral encoded RdRp (47% identity). Termed *Rhizoctonia solani* barnavirus 1 (RsBarV1; KP900904), the incomplete sequence is yet to be ratified as a barnavirus by the ICTV. Subsequent examination of public Transcriptome Shotgun Assembly (TSA) GenBank databases using BLAST search algorithms has since identified another near-complete virus sequence with a high degree of similarity to both BsBarV1 and MBV, named *Colobanthus quitensis* associated barnavirus 1 (CqABV1), detected in the Antarctic pearlwort *Colobanthus quitensis*. If ratified, this would be the first barnavirus identified from a plant species, however it is also possible that the sequence was derived from an associated unidentified fungus, rather than *C. quitensis* itself. Virus characterization studies, including identification of a bacilliform virus particle, are required to determine if these newly identified sequences are true barnavirus genomes.

MBV Virion Properties

The MBV virion M_r is 7.1×10^6 , with a buoyant density in Cs_2SO_4 of 1.32 g/cm³. Virions are stable between pH 6 and 8 and ionic strength of 0.01–0.1 M phosphate.

MBV Virion Structure and Composition

MBV has a bacilliform or bullet shaped morphology, with particles generally 19 nm × 50 nm in size (Fig. 1). Virions contain a single major CP of 21.9 kDa and there are approximately 240 molecules in each capsid. Virions encapsidate a single linear molecule of a positive-sense ssRNA, 4.0 kb in size. The complete 4009-nt sequence is available (Acc No U07551). The RNA has a viral 5' linked protein or viral genome-linked protein (VPg) and lacks a poly(A) tail. RNA constitutes about 20% of virion weight.

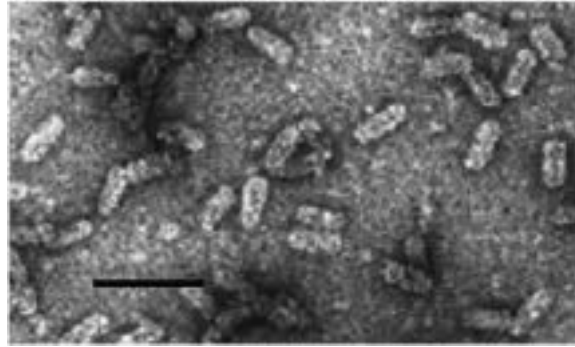


Fig. 1 Electron micrograph of mushroom bacilliform virus particles. The bar represents 100 nm.



Fig. 2 The mushroom bacilliform virus (MBV) genome arrangement.

MBV Genome Organization and Expression

The RNA genome (4009 nt) contains four major and three minor ORFs and has 5'- and 3'-UTRs of 60 nt and 250 nt, respectively (Fig. 2). ORFs 1–4 encode polypeptides of 20, 73, 47, and 22 kDa, respectively. The deduced amino acid sequence of ORF2 contains 3 conserved chymotrypsin-related serine protease sequence motifs. Blast searches of the deduced ORF2 amino acid sequence show similarity to serine proteases encoded by plant sobemoviruses. ORF2 also encodes the VPg. ORF3 contains the $GX_3TX_3NX_nGDD$ amino acid sequence shared by the putative RNA-dependent RNA polymerases (RdRps) of positive-sense ssRNA viruses and has similarity to the RdRps of sobemoviruses, enamoviruses and poleroviruses. ORF4 encodes the capsid protein (CP). ORFs 5–7 encode 8, 6.5, and 6 kDa polypeptides, respectively. The polypeptides potentially encoded by ORFs 1, 5, 6, and 7 show no significant similarity to known polypeptides. The negative strand of MBV contains seven small ORFs of unknown significance. These potentially encode polypeptides ranging from M_r 65K to M_r 105K.

The MBV genome arrangement and transcription/translation strategies are strikingly similar to those of a number of plant viruses, particularly poleroviruses and sobemoviruses. MBV probably also uses similar strategies to express its gene products, including leaky ribosomal scanning for expression of ORF2, ribosomal frameshifting for expression of the RdRP, and subgenomic RNA for expression of the CP. Of these, only subgenomic RNA has been confirmed *in vivo*. In a cell-free system, genomic length RNA directs the synthesis of major 21-kDa and 77-kDa polypeptides and several minor polypeptides of 18–60 kDa. The full-length genomic RNA and a sgrNA (0.9 kb) encoding ORF4 (CP) are found in infected cells. Virions accumulate singly or as aggregates in the cytoplasm. However the MBV life-cycle has yet to be determined.

MBV Evolutionary Relationships

As discussed earlier, the deduced MBV RNA-dependent RNA polymerase (RdRP) sequence shows most similarity to the RdRP of the putative barnavirus RsBarV1, grouping closely together on a phylogenetic tree (Fig. 3). In addition, the deduced MBV RdRP amino acid sequences share striking similarity with those of some plant viruses, particularly sobemoviruses, poleroviruses, and enamoviruses (Fig. 3). This, together with the similarity of the MBV and sobemovirus/polerovirus genome arrangements, suggests that MBV may have shared a common ancestor with these plant virus groups in the distant past.

MBV Transmission and Host Range

MBV is transmitted horizontally through infected mycelium and it is yet to be determined if the virus can be transmitted in spores. There is no known insect vector. Although morphologically similar viruses to MBV have been identified in the field agaric, *A. campestris*, it is unknown whether these particles are related to MBV. Consequently MBV remains the only barnavirus identified to date.

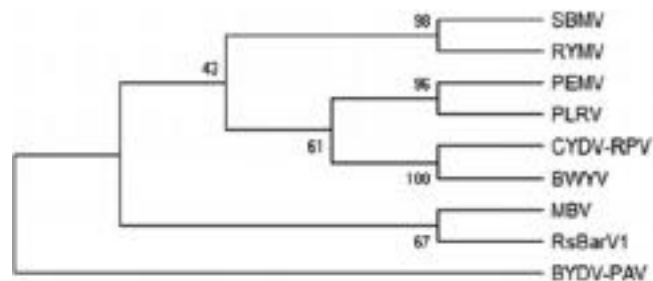


Fig. 3 Neighbor-joining tree of the mushroom bacilliform virus (MBV) RdRp compared to RdRPs of a number of plant viruses and the putative barnavirus RsBarV1, generated using MEGA X v10.0.4. Sequences were aligned using the Clustal W implemented in Genious V11.1.5 (<http://www.geneious.com>), using the cost matrix: BLOSUM default setting (1000 bootstrap replicates), and the tree was constructed with Treeview. Although rooted by outgroup (BYDV-PAV, barley yellow dwarf virus) this is an unrooted tree. CYDV-RPV = cereal yellow dwarf virus-RPV; SBMV = southern bean mosaic virus; RYMV = rice yellow mottle virus; PEMV = pea enation mosaic virus; BWYV = beet western yellows virus; PLRV = potato leafroll virus; RsBarV1 = *Rhizoctonia solani* barnavirus 1. The accession numbers of the sequences used in the analysis were PLRV (D00530), CYDV-RPV (NC004751), BWYV (NC004756), PEMV (NC003629), RYMV (NC001575), SBMV (DQ875594), MBV (U07557), BYDV-PAV (EF043235), RsBarV1 (KP900904).

Further Reading

- Goodin, M.M., Schlagnhauser, B., Romaine, C.P., 1992. Encapsidation of the La France disease specific double-stranded RNAs in 36 nm isometric virus-like particles. *Phytopathology* 82, 285–290.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35 (6), 1547–1549.
- Marzano, S.-Y.L., Nelson, B.D., Ajayi-Oyetunde, O., *et al.*, 2016. Identification of diverse mycoviruses through metatranscriptomics characterization of the viromes of five major fungal plant pathogens. *Journal of Virology* 90, 6846–6863.
- Moyer, J.W., Smith, S.H., 1976. Partial purification and antiserum production to the 19 × 50 nm mushroom virus particle. *Phytopathology* 66, 1260–1261.
- Moyer, J.W., Smith, S.H., 1977. Purification and serological detection of mushroom virus-like particles. *Phytopathology* 67, 1207–1210.
- Nibert, M.L., Manny, A.R., Debat, H.J., *et al.*, 2018. *Archives of Virology* 163, 1921–1926.
- Revill, P.A., Davidson, A.D., Wright, P.J., 1994. The nucleotide sequence and genome organization of mushroom bacilliform virus: A single-stranded RNA virus of *Agaricus bisporus* (Lange) Imbach. *Virology* 202, 904–911.
- Revill, P.A., Davidson, A.D., Wright, P.J., 1998. Mushroom bacilliform virus RNA: The initiation of translation at the 5'-end of the genome and identification of the VPg. *Virology* 249, 231–237.
- Revill, P.A., Davidson, A.D., Wright, P.J., 1999. Identification of a subgenomic mRNA encoding the capsid protein of mushroom bacilliform virus, a single-stranded RNA mycovirus. *Virology* 260, 273–276.
- Romaine, C.P., Schlagnhauser, B., 1991. Hybridization analysis of the single-stranded RNA bacilliform virus associated with La France disease of *Agaricus bisporus*. *Phytopathology* 81, 1336–1340.
- Tavantzis, S.M., Romaine, C.P., Smith, S.H., 1980. Purification and partial characterization of a bacilliform virus from *Agaricus bisporus*: A single-stranded RNA mycovirus. *Virology* 105, 94–102.
- Tavantzis, S.M., Romaine, C.P., Smith, S.H., 1983. Mechanism of genome expression in a single-stranded RNA virus from the cultivated mushroom *Agaricus bisporus*. *Phytopathology* 106, 45–50.

Botybirnaviruses (*Botybirnavirus*)

Mingde Wu, Guoqing Li, Daohong Jiang, and Jiatao Xie, Huazhong Agricultural University, Wuhan, China

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Boty Originates from the Latin name of the host fungus "*Botrytis*"

birna Refers to the bipartite dsRNA genome

Glossary

Anastomosis The fusion between two hyphae leading to their cytoplasmic exchange.

Conidia Asexual, non-motile spores of fungi, a major asexual reproduction structure for most ascomycetes.

Hypovirulence A phenomenon of reduced virulence for fungal pathogens, and sometimes is caused by viral infection.

Sclerotia A hardened granulated structure formed by a compact mass of fungal mycelium with the ability to survive under some extreme environmental conditions.

Vegetative incompatibility A cell-death reaction occurs between two genotypically distinct fungal species or isolates, and mostly resulting in failed transmission of viruses via anastomosis.

Introduction

The first botybirnavirus, *Botrytis porri* botybirnavirus 1 (BpBV1), was identified in the phytopathogenic fungus *Botrytis porri*, as a causal agent of hypovirulence of *B. porri*. BpBV1-infected *B. porri* strains showed dramatically impaired mycelial growth and virulence, and formed numerous mycelial sectors at the colony margin (Fig. 1). Two dsRNA segments (the genome of BpBV1) of approximately 6.2 and 5.8 kbp in size together with the spherical virions were detected in the virus-infected isolate GarlicBc-72. BpBV1 is able to be efficiently transmitted via conidia, and can also be horizontally transferred to other *B. porri* isolates through anastomosis. However, the horizontal transmission seems to be limited by vegetative incompatibility.

Since the firstly discovery of botybirnavirus, botybirnaviruses were soon reported in the close relative of *Botrytis*, *Sclerotinia sclerotiorum*. The infection of botybirnaviruses was determined to be close associated with the hypovirulence of *S. sclerotiorum* as well. Recently, more candidate members in the genus of *Botybirnavirus*, were detected in the fungi of Ascomycota, like fungi of genera *Alternaria* and *Bipolaris*. With the development of sequencing technology, more botybirnavirus-like sequences were also identified during the metatranscriptomic analysis, such as the funding of botybirnavirus-like sequences in the metatranscriptomics survey of soybean phyllosphere phytobiomes. Therefore, botybirnaviruses may have a wide distribution among different fungal groups.

Virion Properties

Electron microscope observation of negative stained botybirnavirus virions (BpBV1) with aqueous uranyl acetate showed that they are spherical in shape, of approximately 35–40 nm in diameter, and probably non-enveloped (Table 1, Fig. 2). At least two structural proteins (SPs) are identified for the BpBV1 virions, encoded by the two dsRNA segments, respectively. Like viruses in *Chrysoviridae* and *Partitiviridae*, the two dsRNAs of BpBV1 are thought to be separately encapsidated in separate virus particles because the virus particles might not be spacious enough for simultaneous packaging of the two dsRNAs. Moreover, variable and unequal molar ratios of two dsRNA segments in purified virus particles of *Sclerotinia sclerotiorum* botybirnavirus 1 (SsBV1) also support the idea that the two segments of botybirnaviruses are separately packaged into particles.

Genome Organization and Replication

The genomes of botybirnaviruses comprise two dsRNAs. The larger dsRNA of BpBV1 is 6215 bp (dsRNA-1, GenBank accession no. JF716350) in length while the size of smaller dsRNA is 5879 bp (dsRNA-2, GenBank accession no. JF716351). The two segments show high sequence identity (~95%) at the 5' termini (500 bp), including the 5'-untranslated regions (UTRs) and partial coding region (Fig. 3). Each dsRNA segment of BpBV1 possesses a large open reading frame (ORF), designated as ORF I (on dsRNA-1) and ORF II (on dsRNA-2) (Fig. 3). The protein encoded by the 5'-proximal coding region of ORF I and ORF II have been determined to be the SPs through peptide mass fingerprinting (PMF) analysis. The protein encoded by the 3'-proximal coding region of ORF I encompasses the RdRp_4 conserved domain sequence (pfam02123 in the conserved domain database of NCBI) and shows



Fig. 1 Colony morphology of the *Botrytis porri* botybirnavirus 1 (BpBV1)-transfected isolate 38T (left) and the BpRV1-free strain GarlicBc-38 (right) on potato dextrose agar (20°C, 15 days).

Table 1 Virion properties of different botybirnaviruses

<i>Virus name</i>	<i>Shape</i>	<i>Diameter (nm)</i>	<i>Genome size (bp)</i>	<i>Genbank Acc. No.</i>
<i>Botrytis porri</i> botybirnavirus 1	spherical	35	6215, 5879	JF716350, JF716350
<i>Sclerotinia sclerotiorum</i> botybirnavirus 1	spherical	38	6457, 5965	KP774592, KP774593
<i>Sclerotinia sclerotiorum</i> botybirnavirus 2	spherical	40	6159, 5872	KT962972, KT962973
<i>Alternaria</i> botybirnavirus 1	spherical	40	6188, 5903	KX784490, KX784491
<i>Bipolaris maydis</i> botybirnavirus 1	ND ^a	ND	6435, 5987	MF034087, MF034086

^aND = Not Done.

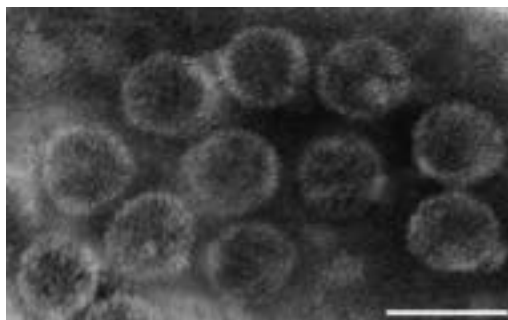


Fig. 2 Transmission electron microscopy image (negative staining) of the *Botrytis porri* botybirnavirus 1 virions from strain GarlicBc-72 of *B. porri* (bar = 50 nm).

sequence similarity with RdRp encoded by viruses in the families *Totiviridae*, *Chrysoviridae*, *Quadriviridae*, and *Megabirnaviridae*. In addition, the proteins encoded by the 5'-proximal region of ORF I and by the entire ORF II lack significant sequence similarities to the proteins of any other known virus groups. SsBV1, a member of a tentative species of the genus *Botyvirus*, has a third dispensable satellite-like dsRNA (SatlRNA) element of ~1.7 kbp (GenBank accession no. KP774594). The expression strategy of botybirnaviruses and functions of remaining regions of the two deduced polyproteins remain unclear. The amino acid sequences of the two SPs also show a significant similarity with each other. Interestingly, whether phylogeny is analyzed based on either of the two SP, their tree topology is similar. This suggests that the ancestral botybirnavirus might have undergone a single duplication event of the SP coding domain.

Taxonomy and Similarity With Other Viruses

Botybirnaviruses are phylogenetically more closely related to viruses of the families *Totiviridae* and *Quadriviridae* than to viruses of the families *Chrysoviridae* and *Megabirnaviridae* based on the RdRp₄ conserved domain sequence similarity and phylogenetic analysis (Fig. 4). The family name, *Botybirnaviridae*, was initially proposed to the International Committee of Taxonomy of Viruses



Fig. 3 Genome organization of *Botrytis porri* botybirnavirus 1 (BpBV1). The genome possesses two dsRNA segments, dsRNA-1 and dsRNA-2. The dsRNA-1 is 6215 bp long and comprises one large ORF (nt positions 404–6112), encoding a polyprotein of 1902 aa. The dsRNA-2 is 5788 bp long and also comprises one large ORF (nt positions 405–5771), which encodes a polypeptide of 1788 aa. p70 and p85/80 are the structural proteins of BpBV1 inferred from peptide fingerprinting analyses. Highly-conserved region (~500 bp) covering the entire 5'-UTR and partial coding region are highlighted in red on the BpBV1 genome.

(ICTV) to accommodate the genus *Botybirnavirus* including four species (Table 1). The viruses belonging to the proposed family can be clearly separated from other viruses with bipartite genomes based on bipartite genome size (~12 kbp, which is smaller than that of *Megabirnaviridae* (~17 kbp) but larger than those of *Partitiviridae* (~3–4.8 kbp), *Picobirnaviridae* (~4 kbp) and *Birnaviridae* (~6 kbp)), genome organization and nucleotide and amino acid sequence similarity of encoded RdRp and CP sequences. The current taxonomic status of botybirnaviruses, approved by ICTV, is that *Botrytis porri* botybirnavirus 1 is only the species in the genus *Botybirnavirus* which is as yet unassigned to a family.

Transmission and Distribution

There are no known natural vectors for the transmission of botybirnaviruses. Under laboratory conditions, the transmission of BpBV1 is similar to most mycoviruses, vertically through sporulation or horizontally by hyphal anastomosis between somatically compatible fungal strains. The vertical transmission via sporulation for BpBV1 is highly efficient in *B. porri* strain GarlicBc-72. BpBV1 could be detected in 34 out of 35 single-conidium (SC) isolates of strain GarlicBc-72 with the vertical transmission rate of 97.1%. In contrast, the horizontal transmission of BpBV1 was not so efficient, it could only be transmitted from strain GarlicBc-72 to one of the recovered SC isolate SC35 but failed to be transmitted to other two strains of *B. porri* during the in-laboratory test. This might be due to the vegetative incompatibility among different *B. porri* strains. As isolate SC35 is isogenic to strain GarlicBc-72, they should be vegetatively compatible to each other resulting in successful viral transmission between the two individuals. Similar unsuccessful horizontal transmission was also observed for *Sclerotinia sclerotiorum* botybirnavirus 2 (SsBV2), and may also be caused by the vegetative incompatibility between the donor and recipient *S. sclerotiorum* strains. Although in-laboratory tests showed the horizontal transmission of BpBV1 was limited, it is surprising that the same botybirnavirus, BpBV1, was detected in both *B. squamosa* and *S. sclerotiorum* in addition to *B. porri*. Moreover, recent results from high through-put sequencing also indicate that BpBV1 may be present in the population of *B. cinerea*. This suggests that interspecific transmission of BpBV1 may occur in field conditions without the limitation by vegetative incompatibility among different fungal species, although the underlining mechanism is still unclear. Botybirnaviruses probably have a wide host range, mainly infecting ascomycetes including *B. porri*, *B. squamosa*, *S. sclerotiorum*, *B. maydis*, and *Alternaria*. Besides being detected in population of *B. porri* and *B. squamosa* of China, high through-put sequencing data also showed that botybirnaviruses were also present in soybean leaf samples from United States and BpBV1 was also found in the *S. sclerotiorum* population of Australia, indicating BpBV1 and other botybirnaviruses may have a geographic distribution all over the world.

Biology

BpBV1 was reported to confer hypovirulence on the phytopathogenic fungus *B. porri*, the causing agent of garlic clove rot, garlic leaf blight, and leek leaf rot. Compared with BpBV1-free isolates of *B. porri*, the BpBV1-infected isolate grew slowly on potato dextrose agar, formed numerous mycelial sectors at the colony margin and caused smaller lesions on leaves of garlic (*Allium sativum*). In addition, abundant vacuole-like membranous structures and membranous vacuoles/vesicles were observed in the cytoplasm of virus-infected isolate. Three lines of experimental evidence support the conclusion that BpBV1 is responsible for the reduced mycelial growth and hypovirulence in *B. porri*. The first clue is based on the vertical transmission experiment. All BpBV1-infected SC isolates of strain GarlicBc-72 were hypovirulent and accompanied with reduced mycelial growth. In contrast, the BpBV1-free SC isolate SC35 became virulent on garlic leaves and grew normally. Secondly, after BpBV1 was transmitted to the virus cured strain SC35 through hyphal contact, three derivative isolates of strain SC35 carrying BpBV1 showed reduced mycelial

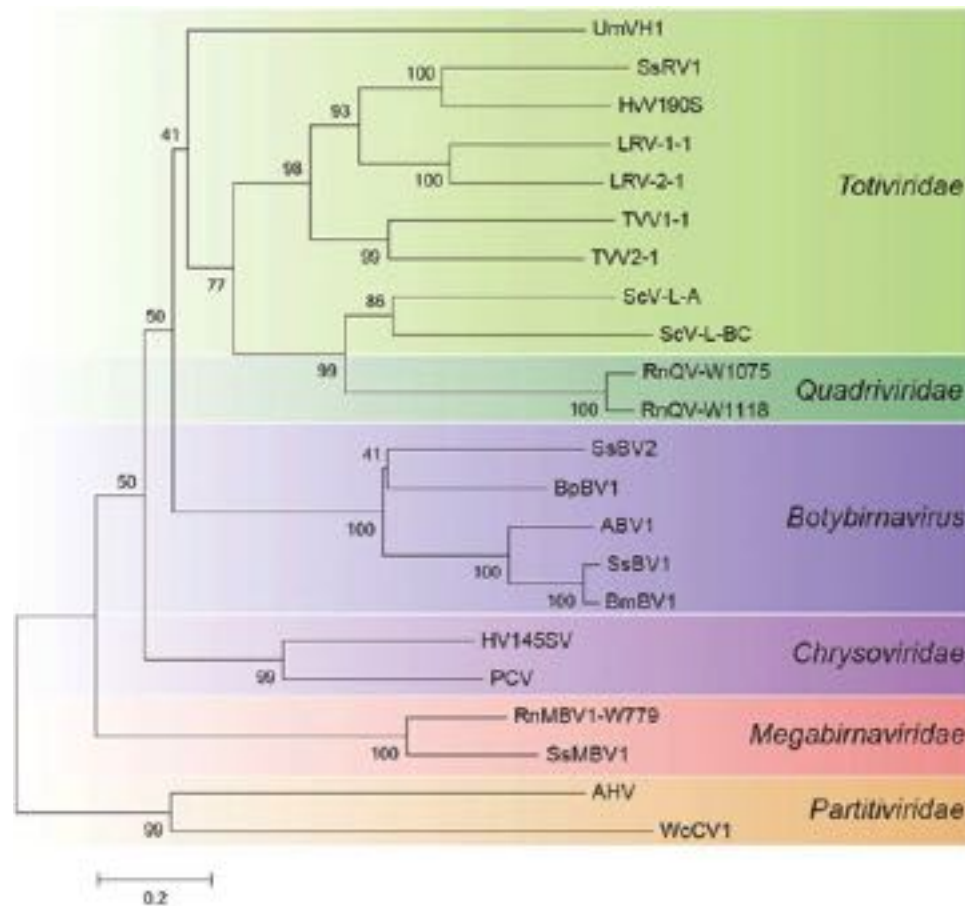


Fig. 4 Phylogenetic analysis of *Botrytis porri* botybirnavirus 1 and other 21 selected RNA viruses presented in an neighbor-joining (NJ) tree inferred from the RdRp sequences. The NJ tree was constructed by using the CLUSTAL_W program in the MEGA 6.0 software. The number labeled at each node indicates the bootstrap percentage ($N = 1000$). Abbreviated virus names and GenBank accession number for viral RdRp: SsRV1, *Sphaeropsis sapinea* RNA virus 1 (NC_001963); HvV190S, *Helminthosporium victoriae* virus 190S (U41345); LRV-1-1, *Leishmania* RNA virus 1-1 (M92355); LRV-2-1, *Leishmania* RNA virus 2-1 (U32108); TVV1-1, *Trichomonas vaginalis* virus 1-1 (U08999); TVV2-1, *Trichomonas vaginalis* virus 2-1 (AF127178); UmVH1, *Ustilago maydis* virus H1 (NC_003823); ScV-L-A, *Saccharomyces cerevisiae* virus L-A (J04692); ScV-L-BC, *Saccharomyces cerevisiae* virus L-BC (U01060); RnQV1-W1075, *Rosellinia necatrix* quadrivirus 1-W1075 (AB620063); RnQV1-W1118, *Rosellinia necatrix* quadrivirus 1-W1118 (AB744679); BpBV1, *Botrytis porri* botybirnavirus 1 (JF716350); SsBV1, *Sclerotinia sclerotiorum* botybirnavirus 1 (KP774592); SsBV2, *Sclerotinia sclerotiorum* botybirnavirus 2 (KT962972); ABV1, *Alternaria botybirnavirus* 1 (KX784491); BmBV1, *Bipolaris maydis* botybirnavirus 1 (MF034087); PCV, *Penicillium chrysogenum* virus (AF296339); HvV145S, *Helminthosporium victoriae* virus 145S (AF297176); RnMBV1-W779, *Rosellinia necatrix* megabirnavirus 1-W779 (AB512282); SsMBV1, *Sclerotinia sclerotiorum* megabirnavirus 1 (KP686398); AHV, *Atkinsonella hypoxylon* virus (L39125); WcCV1, *White clover* cryptic virus 1 (AY705784).

growth and impaired virulence on garlic leaves. The transfection experiment provided the third evidence, in which the purified BpBV1 particles were introduced into the protoplasts of the BpBV1-free virulent strain GarlicBc-38 of *B. porri* artificially. The derivative isolate 38T also displayed debilitation symptoms (Fig. 1), including reduced mycelial growth and hypovirulence on garlic leaves. It is interesting that although BpBV1 caused severe debilitation on *B. porri*, the effects of BpBV1 on *B. squamosa* seems to be mild. Only slightly reduced mycelial growth and virulence on garlic leaves was observed on the BpBV1-infected *B. squamosa* strain. Whether this phenomenon was caused by the minor sequence differences between the two BpBV1 strains from *B. porri* and *B. squamosa*, or the different genetic background of the two hosts remain to be investigated.

Similarly, the infection of botybirnaviruses in *S. sclerotiorum* is also able to reduce its mycelial growth and virulence. The hypovirulent *S. sclerotiorum* strain AH16 was infected by SsBV2 together with a mitovirus (*Sclerotinia sclerotiorum* mitovirus 4, SsMV4). To test whether SsBV2 or/and SsMV4 is responsible for the hypovirulence of strain AH16, both viruses were tested to be transmitted from strain AH16 to Ep-1PNA367R via hyphal contact, or to be eliminated from strain AH16 through protoplast regeneration. Nevertheless, due to the vegetative incompatibility between the donor strain and the recipient strain, the transmission was failed. Moreover, the protoplast regeneration was also unable to cure the virus in strain AH16. Finally, the virions of SsBV2 was successfully introduced into strain Ep-1PNA367R. The results based on dsRNA extraction, or RT-PCR detection revealed that the transfectant Ep-1PNA367RVT carried SsBV2 without the presence of SsMV4. Compared with SsBV2-free strain Ep-1PNA367R, strain AH16 grew slower and showed reduced virulence on detached soybean leaves. In addition, strain Ep-1PNA367R formed sclerotia at

7 days post inoculation (dpi), whereas strain AH16 formed fewer and smaller sclerotia at 15 dpi. It is noteworthy that strain Ep-1PNA367RVT showed more severely debilitated symptoms compared with strain AH16, including slower mycelial growth, more impaired virulence and no sclerotial production. In addition, SsBV2 was able to transmit from strain Ep-1PNA367RVT to SsBV2-free strain Ep-1PNA367, and the newly SsBV2-infected strain exhibited similar biological traits to those of Ep-1PNA367RVT. Therefore, based on these observations, SsBV2 confers hypovirulence to *S. sclerotiorum*. In contrast, infection of SsBV1 alone has no significant impacts on culture morphology and virulence of *S. sclerotiorum*. However, infection of SsBV1 along with its SatlRNA leads to slightly reduced virulence and slower mycelial growth rate on its host. Nevertheless, the culture morphology and sclerotial formation of strains carrying both SsBV1 and SatlRNA was comparable to those of strains without the infection by both SsBV1 and SatlRNA, or by SsBV1 alone. Besides *Botrytis* spp. and *S. sclerotiorum*, botybirnaviruses were also reported in the fungi *B. maydis* and *Alternaria*. Although the genomes of those two botybirnaviruses were fully sequenced, the impacts on their fungal hosts are still unknown.

Further Reading

- Liu, L., Wang, Q., Cheng, J., *et al.*, 2015. Molecular characterization of a bipartite double-stranded RNA virus and its satellite-like RNA co-infecting the phytopathogenic fungus *Sclerotinia sclerotiorum*. *Frontiers in Microbiology* 6, 406.
- Marzano, S.Y.L., Domier, L.L., 2016. Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Research* 213, 332–342.
- Mu, F., Xie, J.T., Cheng, S.F., *et al.*, 2018. Virome Characterization of a collection of *Sclerotinia sclerotiorum* from Australia. *Frontiers in Microbiology* 8, 2540.
- Ran, H., Liu, L., Li, B., *et al.*, 2016. Coinfection of a hypovirulent isolate of *Sclerotinia sclerotiorum* with a new botybirnavirus and a strain of a mitovirus. *Virology Journal* 13, 92.
- Wu, M.D., Jin, F.Y., Zhang, J., *et al.*, 2012. Characterization of a novel bipartite double-stranded RNA mycovirus conferring hypovirulence in the phytopathogenic fungus *Botrytis porri*. *Journal of Virology* 86, 6605–6619.
- Xiang, J., Fu, M., Hong, N., *et al.*, 2017. Characterization of a novel botybirnavirus isolated from a phytopathogenic *Alternaria* fungus. *Archives of Virology* 162, 3907–3911.

Chrysoviruses (*Chrysoviridae*) - General Features and Chrysovirus-Related Viruses[☆]

Ioly Kotta-Loizou, Imperial College London, London, United Kingdom

Robert HA Coutts, University of Hertfordshire, Hatfield, United Kingdom

José R Castón, National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Hiromitsu Moriyama, Tokyo University of Agriculture and Technology, Tokyo, Japan

Said A Ghabrial[†], Department of Plant Pathology, University of Kentucky, Lexington, KY, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Anastomosis The fusion between branches of hyphae.

Conidium Asexual spore of a fungus.

Heterokaryon A multinucleate fungal cell that contains genetically different nuclei.

Introduction

Chrysoviruses were initially discovered in the late 1960s and early 1970s, when the presence of isometric virus particles in many industrial strains of the ascomycete *Penicillium chrysogenum* used for penicillin production was noted. Virus infection was considered to be responsible for the instability of some of these strains, generating significant interest in the study of *Penicillium* viruses. *Penicillium chrysogenum* virus (PcV), the exemplar virus of the genus *Alphachrysovirus*, was one of the first mycoviruses subjected to extensive biochemical, biophysical and structural studies. PcV and the related viruses *Penicillium brevi-compactum* virus (PbV) and *Penicillium cyaneo-fulvum* virus (Pc-fV) have similar isometric particles, 35–40 nm in diameter, and are serologically related. These virus properties were undisputed in these initial studies, however there was confusion as to whether these viruses contain three or four double-stranded (ds) RNA segments. Because none of these three *Penicillium* viruses were characterized at the molecular level, they were originally grouped under the genus *Chrysovirus* and provisionally placed in the family *Partitiviridae* with the assumption that their genomes were bipartite, with dsRNA 1 encoding the RNA-dependent RNA polymerase (RdRp) and dsRNA 2 encoding the major capsid protein (CP). Additional dsRNAs found in these isolates were nominated dsRNA 3 and 4 and were presumed to represent defective or satellite dsRNAs as described for some partitiviruses. Determination of the complete nucleotide sequence and genome organization of each of the four monocistronic dsRNA segments associated with PcV virions together with another chrysovirus *Helminthosporium victoriae* virus 145S (HvV145S) in the early 2000s led to a reconsideration of the classification of the genus *Chrysovirus*. Based on the consistent and simultaneous presence of four dsRNA segments, the existence of extended regions of highly conserved terminal sequences at both ends of the four segments, together with sequence comparisons and phylogenetic analysis, it became clear that PcV and related viruses should not be classified with the family *Partitiviridae*. These observations led to the creation of a new family the *Chrysoviridae* to accommodate isometric dsRNA mycoviruses with multipartite genomes. The name *chryso* means *gold* in Greek and is derived from the specific epithet of *P. chrysogenum*, the fungal host of the prototype strain of the type species, (PcV), which often produces a golden yellow pigment. More recently, the discovery and characterization of a large number of chrysovirus-related viruses resulted in a reorganization of the *Chrysoviridae* family: the original genus *Chrysovirus* was renamed as *Alphachrysovirus* and a new genus *Betachrysovirus* was created with *Botryosphaeria dothidea chrysovirus 1* as the type species. The two genera accommodate viruses that possess three to seven genomic segments and are not only present in ascomycetes and basidiomycetes but also associated with plants and insects.

Virion Properties

Virions of members in the family *Chrysoviridae* are isometric, proteinaceous in nature with no envelope and are 35–40 nm in diameter. The buoyant densities of the virions are in the range of 1.34–1.39 g cm⁻³ and their sedimentation coefficients (in Svedberg units) are in the range of 145S to 150S. Generally, each virion contains only one genomic dsRNA segment, however, purified preparations of PcV and Pc-fV can contain minor distinctly sedimenting components that include empty particles and replication intermediates. Chrysovirus particles possess virion-associated RdRp activity, which catalyzes the synthesis of single-stranded (ss) RNA copies of the (+) strand of each of the genomic dsRNA molecules. *In vitro* transcription occurs by a conservative mechanism, whereby the released ssRNA represents the newly synthesized (+) strand.

[☆]This work is dedicated to the memory of our friend and colleague Said Ghabrial, who sequenced the first chrysovirus in 2000 and passed away in November 2018.

[†]Deceased.

Virion structure and composition

The three-dimensional structure and protein stoichiometry of several chrysoviruses has been analyzed by three-dimensional cryo-electron microscopy (3D cryo-EM) combined with analytical ultracentrifugation analysis. The 3D cryo-EM structures of two chrysoviruses, PcV and *Cryphonectria nitschkei* chrysovirus virus 1 (CnCV1), were determined at sub-nanometer and near-atomic resolutions, respectively.

Virions are isometric, non-enveloped, about 40 nm in diameter with a ~ 50 Å thick protein shell. The capsids of PcV and CnCV1 comprise 60 copies of a 982- and 902-amino acid (aa) polypeptide, respectively, arranged on an authentic T=1 icosahedral lattice. The most prominent features are 12 outwardly protruding pentons, each containing five copies of the CP (Fig. 1(a)). The uneven outer surface is similar to the capsid of *Saccharomyces cerevisiae* virus L-A (ScV-L-A; a totivirus with an undivided genome), and contrasts with the smooth outer surface of reoviruses, where the CP has a plate-like structure.

PcV and CnCV1 capsids constitute partial exceptions to the most extended tendency among dsRNA viruses, the T=1 core with 60 equivalent dimers. The PcV CP is formed by a repeated predominantly α -helical domain, indicative of ancestral gene duplication. These two domains are arranged in two sets of five: five A domains directly surround the icosahedral fivefold axis and five B domains intercalated between them, forming a pseudodecamer. This organization with two similar motifs generates an architecture resembling that of the 120-subunit T=1 lattice of totivirus, reovirus, and cystovirus capsids, in which the two asymmetrical dimer components are arranged in a near-parallel fashion. A 120-subunit (or 120-domain) T=1 layer, which remains structurally undisturbed throughout the viral cycle, is therefore a ubiquitous architecture for management of dsRNA metabolism.

The near-atomic cryo-EM structure of the PcV virion at ~ 4 Å resolution shows that the complete 982-aa CP is built of 33 α -helices and 20 β -strands (Fig. 1(b)). The CP has two domains or halves with a very similar fold, an N-terminal domain (residues 1–498) connected by a 16-residue linker to a C-terminal domain (516–982) (Fig. 1(c)). Both PcV CP domains have a long (> 30 Å) α -helix tangential to the capsid surface (Fig. 1(b); arrows). Despite the lack of sequence similarity between the two halves, the CP is an almost perfect structural duplication of a single domain in which many α -helices and β -chains show good matching (Fig. 2(a)). Superimposition of secondary structure elements shows, in addition to the N- and C-terminal arms, a single “hot spot” on the outer capsid surface into which structural and functional variations can be introduced by insertion of 50–100 residue segments (Fig. 2(b)). A preferential insertion site would allow the acquisition of new functions (e.g., new enzymatic activities) while preserving basic capsid protein folding. Notably, the basic PcV fold is well preserved among dsRNA viruses, and provides information regarding the progenitor fold of the dsRNA virus lineage and its evolutionary mechanisms. Structural comparisons of either of the PcV domains and the ScV-L-A or *Rosellinia necatrix* quadrivirus 1 (RnQV1) CPs highlight the same conserved PcV motif and the hot spot for insertions, in addition to two additional insertion zones that face the outer capsid surface (Fig. 2(b) and (c)). This co-localization suggests that these preferential insertion sites are ancient. Structural comparison of PcV and other dsRNA viruses in the family *Reoviridae* shows ancestral structural motifs or subdomains that have acted as a skeleton.

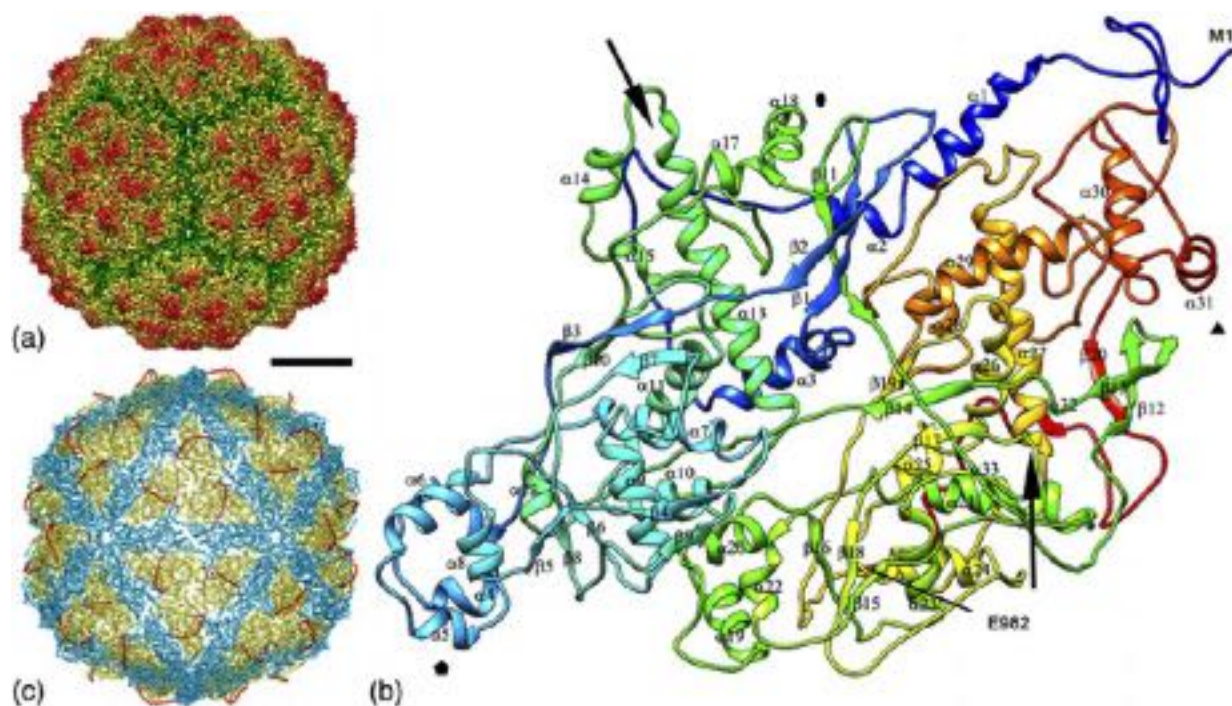


Fig. 1 Three-dimensional cryo-EM reconstruction of *Penicillium chrysogenum* virus (PcV) virions at ~ 4 Å resolution. (a) Radially color-coded surface-shaded virion T=1 capsid showing 12 outwardly protruding pentamers (orange). Bar = 100 Å. (b) Rainbow-colored ribbon diagram of the PcV capsid protein; arrows indicate the longest α -helices. (c) Atomic model of the PcV capsid highlighting the capsid protein with the N-terminal domain (yellow), the linker segment (red), and the C-terminal domain (yellow).

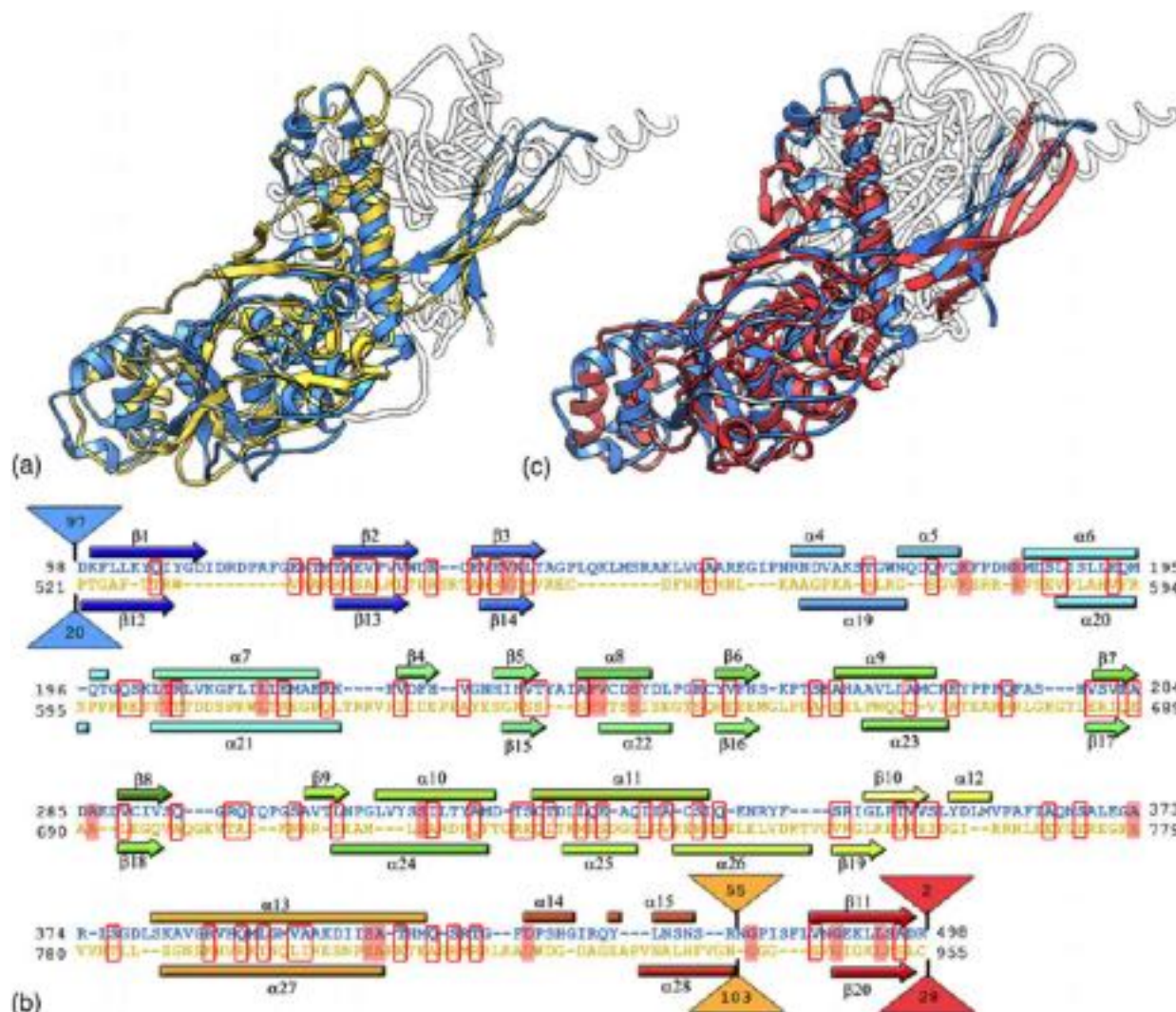


Fig. 2 Structural duplication of the *Penicillium chrysogenum* virus (PcV) capsid protein (CP) and structural homology of the PcV CP fold with the *Saccharomyces cerevisiae* virus L-A CP. (a) Superimposed N-terminal (blue) and C-terminal (yellow) domains of the PcV CP (non-superimposed regions are white). (b) Sequence alignment of N- and C-terminal domains resulting from the structural alignment. α -helices (rectangles) and β -strands (arrows) are rainbow-colored from blue (N terminus) to red (C terminus) for each domain. Triangles represent non-aligned segments (sizes indicated). (c) Structural alignment of the PcV CP N-terminal domain (blue) with L-A capsid protein (red).

Whereas dsRNA in the interior of reovirus cores is very compact (with an average genome density of 40 bp/100 nm³ and a spacing among dsRNA strands of 25–30 Å), PcV and CnCV1 have spacious capsids. Considering the volume available in the capsid interior and average genome size of 3200 bp (each segment is separately encapsidated in a similar particle), the packed dsRNA would have an interstrand spacing of 40 Å, which closely matches the relatively low density in most fungal virus capsids. The looser packing of the dsRNA would probably improve template motion in the transcriptional and replicative active particle that contains 1–2 copies of the RdRp. Although most mycovirus T=1 capsids are negatively charged on their inner surface, the PcV CP inner surface has highly positively charged triskelion-shaped areas that maintain the encapsidated genome in close contact with the inner capsid surface. This arrangement of dsRNA might further facilitate genome mobility within the capsid, and/or have a structural role in capsid stability. The capsid is highly porous; whereas the ~11 Å diameter pores at the fivefold axis would allow the exit of viral transcripts but not of dsRNA, the ~5 Å-diameter pores at the threefold axis would allow nucleotide diffusion.

Genome Organization

The virions of members of family *Chrysoviridae* contain three to seven unrelated linear, separately encapsidated, monocistronic dsRNA segments 0.8–3.7 kbp in size (Table 1); the genomic structures of PcV, the prototype of genus *Alphachrysovirus*, and

Table 1 Members of the *Chrysoviridae* family

Genus	virus name & abbreviation	dsRNA & protein	5'-UTR	3'-UTR	accession number
<i>Alphachrysovirus</i>	Amasya cherry disease associated chrysovirus (ACDACV)	dsRNA 1 (3399 bp); RdRP (124 kDa)	86 nt	49 nt	AJ781166
		dsRNA 2 (3128 bp); CP (112 kDa)	95 nt	48 nt	AJ781165
		dsRNA 3 (2833 bp); alphachryso-P4 (98 kDa)	94 nt	84 nt	AJ781164
		dsRNA 4 (2498 bp); alphachryso-P3 (77 kDa)	105 nt	359 nt	AJ781163
	Anthurium mosaic-associated virus (AMAV)	dsRNA 1 (3550 bp); RdRP (126 kDa)	179 nt	71 nt	FJ899675
		dsRNA 2 (3448 bp); CP (113 kDa)	293 nt	161 nt	FJ899676
		dsRNA 3 (3244 bp); alphachryso-P4 (97 kDa)	393 nt	259 nt	FJ899677
	Aspergillus fumigatus chrysovirus (AfuCV)	dsRNA 1 (3560 bp); RdRP (129 kDa)	128 nt	87 nt	FN178512
		dsRNA 2 (3159 bp); CP (109 kDa)	167 nt	130 nt	FN178513
		dsRNA 3 (3006 bp); alphachryso-P3 (100 kDa)	169 nt	161 nt	FN178514
		dsRNA 4 (2863 bp); alphachryso-P4 (95 kDa)	154 nt	165 nt	FN178515
	Brassica campestris chrysovirus 1 (BcCV1)	dsRNA 1 (3639 bp); RdRP (131 kDa)	101 nt	97 nt	KP782031
		dsRNA 2 (3567 bp); CP (126 kDa)	135 nt	81 nt	KP782030
		dsRNA 3 (3337 bp); alphachryso-P4 (113 kDa)	132 nt	139 nt	KP782029
	Colletotrichum gloeosporioides chrysovirus 1 (CgCV1)	dsRNA 1 (3397 bp); RdRP (126 kDa)	71 nt	59 nt	KT581957
		dsRNA 2 (2869 bp); CP (102 kDa)	71 nt	95 nt	KT581958
		dsRNA 3 (2630 bp); alphachryso-P4 (92 kDa)	66 nt	92 nt	KT581959
		dsRNA 1 (partial; 2978 bp ^a); RdRP (110 kDa)	82 nt	7?	GQ290649
	Cryphonectria nitschkei chrysovirus 1 (CnCV1)	dsRNA 2 (partial; 2980 bp ^a); CP (99 kDa)	242 nt	17?	GQ290645
		dsRNA 3 (partial; 2552 bp ^a); alphachryso-P4 (91 kDa)	?	77	HM013825
		dsRNA 4 (partial; 2960 bp ^a); alphachryso-P3 (81 kDa)	698 nt	30?	HM013826
	Fusarium oxysporum chrysovirus 1 (FoCV1)	dsRNA 1 (partial; 2574 bp ^a); RdRP	?	?	EF152346
		dsRNA 2 (partial; 648 bp ^a); CP	?	?	EF152347
		dsRNA 3 (partial, non-functional; 994 bp ^a); alphachryso-P4	?	?	EF152348
	Helminthosporium victoriae virus 145S (HvV145S)	dsRNA 1 (3612 bp); RdRP (125 kDa)	207 nt	144 nt	AF297176
		dsRNA 2 (3134 bp); CP (100 kDa)	293 nt	153 nt	AF297177
		dsRNA 3 (2972 bp); alphachryso-P4 (93 kDa)	302 nt	150 nt	AF297178
		dsRNA 4 (2763 bp); alphachryso-P3 (81 kDa)	412 nt	209 nt	AF297179
	Isaria javanica chrysovirus 1 (IjCV1)	dsRNA 1 (3593 bp); RdRP (129 kDa)	174 nt	65 nt	KX898416
		dsRNA 2 (3175 bp); CP (109 kDa)	176 nt	74 nt	KX898417
		dsRNA 3 (3165 bp); alphachryso-P3 (108 kDa)	147 nt	84 nt	KX898418
		dsRNA 4 (2874 bp); alphachryso-P4 (92 kDa)	212 nt	130 nt	KX898419
	Macrophomina phaseolina chrysovirus 1 (MpCV1)	dsRNA 1 (3712 bp); RdRP (129 kDa)	256 nt	138 nt	KP900886
		dsRNA 2 (3462 bp); CP (111 kDa)	374 nt	112 nt	KP900887
		dsRNA 3 (2985 bp); alphachryso-P4 (94 kDa)	278 nt	175 nt	KP900889
		dsRNA 4 (2927 bp); alphachryso-P3 (100 kDa)	127 nt	139 nt	KP900888
	Penicillium chrysogenum virus (PcV)	dsRNA 1 (3562 bp); RdRP (129 kDa)	144 nt	64 nt	AF296439
		dsRNA 2 (3200 bp); CP (109 kDa)	157 nt	94 nt	AF296440
		dsRNA 3 (2976 bp); alphachryso-P3 (101 kDa)	161 nt	83 nt	AF296441
		dsRNA 4 (2902 bp); alphachryso-P4 (95 kDa)	162 nt	196 nt	AF296442
	Persea americana chrysovirus (PaCV)	dsRNA 1 (3421 bp); RdRP (126 kDa)	97 nt	42 nt	KJ418374
		dsRNA 2 (3335 bp); CP (122 kDa)	95 nt	99 nt	KJ418375
		dsRNA 3 (2857 bp); alphachryso-P4 (92 kDa)	108 nt	135 nt	KJ418376
	Raphanus sativus chrysovirus 1 (RsCV1)	dsRNA 1 (3638 bp); RdRP (131 kDa)	106 nt	115 nt	JQ045335
		dsRNA 2 (3517 bp); CP (124 kDa)	128 nt	83 nt	JQ045336
		dsRNA 3 (3299 bp); alphachryso-P4 (110 kDa)	134 nt	177 nt	JQ045337
	Shuangao chryso-like virus (SCLV)	dsRNA 1 (3461 bp); RdRP (127 kDa)	45 nt	59 nt	MF176340
		dsRNA 2 (3140 bp); (108 kDa)	200 nt	57 nt	MF176342
		dsRNA 3 (3080 bp); (98 kDa)	153 nt	218 nt	MF176341
		dsRNA 4 (3059 bp); (106 kDa)	150 nt	74 nt	MF176343
	Verticillium dahliae chrysovirus 1 (VdCV1)	dsRNA 1 (3594 bp); RdRP (127 kDa)	91 nt	176 nt	HM004067
		dsRNA 2 (3313 bp); CP (113 kDa)	204 nt	58 nt	HM004068
		dsRNA 3 (2983 bp); alphachryso-P3 (84 kDa)	635 nt	53 nt	HM004069
		dsRNA 4 (2932 bp); alphachryso-P4 (90 kDa)	264 nt	196 nt	HM004070

Table 1 Continued

Genus	virus name & abbreviation	dsRNA & protein	5'-UTR	3'-UTR	accession number
<i>Betachrysovirus</i>	<i>Alternaria alternata</i> chrysovirus 1 (AaCV1)	dsRNA 1 (3647 bp); RdRP (124 kDa)	199 nt	94 nt	LC350277
		dsRNA 2 (2857 bp); CP (82 kDa)	389 nt	152 nt	LC350278
		dsRNA 3 (2785 bp); betachryso-P4 (83 kDa)	366 nt	94 nt	LC350279
		dsRNA 4 (2772 bp); betachryso-P3 (84 kDa)	334 nt	107 nt	LC350280
		dsRNA 5 (836 bp); 13 kDa	102 nt	386 nt	LC350281
	<i>Botryosphaeria dothidea</i> chrysovirus 1 (BdCV1)	dsRNA 1 (3654 bp); RdRP (126 kDa)	230 nt	73 nt	KF688736
		dsRNA 2 (2773 bp); CP (80 kDa)	272 nt	254 nt	KF688737
		dsRNA 3 (2597 bp); betachryso-P3 (82 kDa)	293 nt	63 nt	KF688738
		dsRNA 4 (2574 bp); betachryso-P4 (77 kDa)	292 nt	128 nt	KF688739
	<i>Colletotrichum fructicola</i> chrysovirus 1 (CfCV1)	dsRNA 1 (3620 bp); RdRP (126 kDa)	154 nt	67 nt	MG425969
		dsRNA 2 (2801 bp); CP (89 kDa)	233 nt	90 nt	MG425970
		dsRNA 3 (2687 bp); betachryso-P3 (76 kDa)	273 nt	290 nt	MG425971
		dsRNA 4 (2437 bp); betachryso-P4 (72 kDa)	340 nt	109 nt	MG425972
		dsRNA 5 (1750 bp); 54 kDa	61 nt	168 nt	MG425973
		dsRNA 6 (1536 bp); 33 kDa	67 nt	560 nt	MG425974
		dsRNA 7 (1211 bp); 10 kDa	252 nt	689 nt	MG425975
	<i>Fusarium graminearum</i> dsRNA mycovirus 2 (FgV2)	dsRNA 1 (3580 bp); RdRP (128 kDa)	82 nt	84 nt	HQ343295
		dsRNA 2 (3000 bp); betachryso-P4 (95 kDa)	93 nt	279 nt	HQ343296
		dsRNA 3 (2982 bp); CP (94 kDa)	105 nt	306 nt	HQ343297
		dsRNA 4 (2748 bp); betachryso-P3 (91 kDa)	78 nt	162 nt	HQ343298
		dsRNA 5 (2414 bp); 80 kDa	97 nt	184 nt	HQ343299
	<i>Fusarium oxysporum</i> f. sp. <i>dianthi</i> mycovirus (FodV)	dsRNA 1 (3555 bp); RdRP (129 kDa)	82 nt	53 nt	KP876629
		dsRNA 2 (2809 bp); betachryso-P4 (95 kDa)	84 nt	88 nt	KP876630
		dsRNA 3 (2794 bp); CP (92 kDa)	97 nt	138 nt	KP876631
		dsRNA 4 (2646 bp); betachryso-P3 (92 kDa)	97 nt	56 nt	KP876632
		dsRNA 1 (3554 bp); RdRP (125 kDa)	118 nt	52 nt	AB560761
	<i>Magnaporthe oryzae</i> chrysovirus 1-A (MoCV1-A)	dsRNA 2 (3250 bp); betachryso-P3 (99 kDa)	343 nt	102 nt	AB560762
		dsRNA 3 (3074 bp); betachryso-P4 (84 kDa)	144 nt	530 nt	AB560763
		dsRNA 4 (3043 bp); CP (85 kDa)	161 nt	443 nt	AB560764
		dsRNA 5 (2879 bp); 63 kDa	178 nt	865 nt	AB700631
		dsRNA 1 (3698 bp); RdRP (127 kDa)	168 nt	149 nt	KT601115
	<i>Penicillium janczewskii</i> chrysovirus 1 (PjCV1)	dsRNA 2 (2899 bp); CP (84 kDa)	186 nt	400 nt	KT601116
		dsRNA 3 (2942 bp); betachryso-P3 (90 kDa)	223 nt	253 nt	KT601117
		dsRNA 4 (2506 bp); betachryso-P4 (74 kDa)	179 nt	263 nt	KT601118
		dsRNA 1 (3540 bp); RdRP (124 kDa)	94 nt	86 nt	KT950836
	<i>Penicillium janczewskii</i> chrysovirus 2 (PjCV2)	dsRNA 2 (2699 bp); CP (84 kDa)	194 nt	153 nt	KT950837
		dsRNA 3 (2535 bp); betachryso-P3 (80 kDa)	164 nt	100 nt	KT950838
		dsRNA 4 (partial; 2155 bp ^a); betachryso-P4 (70 kDa)	190 nt	?	KT950839

^alength of dsRNA is underestimated because only partial sequence is available.

Botryosphaeria dothidea chrysovirus 1 (BdCV1), the prototype of genus *Betachrysovirus*, each comprising four dsRNA segments, are schematically represented in Fig. 3.

For the genus *Alphachrysovirus*, the largest segment, dsRNA 1, encodes the RdRp and the second largest segment, dsRNA 2, encodes the major CP. The earlier conflicting reports on whether PcV contains three or four segments were explained when studies on cDNA cloning and sequencing of the viral dsRNAs were completed, revealing that dsRNAs 3 and 4 only differ in size by 74 bp (Table 1) and co-migrate when separated by agarose gel electrophoresis. Previous studies on sequencing analysis and *in vitro* coupled transcription-translation assays showed that each of the four dsRNAs is monocistronic, as each dsRNA contains a single major open reading frame (ORF) and each is translated into a single major product of the size predicted from its deduced aa sequence. Therefore, the fact that PcV virions contain four distinct dsRNA segments was clearly established. Unlike PcV, HvV145S dsRNAs 3 and 4 are clearly resolved from each other when purified dsRNA preparations are subjected to agarose gel electrophoresis and, as shown in Table 1, are significantly different in size. Assignment of numbers 1–4 to PcV dsRNAs was made according to their decreasing size. Following the same criterion as used for PcV, the dsRNAs associated with all members of the genus *Alphachrysovirus* were assigned the numbers 1–4. Sequence comparisons, however, indicated that dsRNA 3 of HvV145S, Amasya cherry disease associated chrysovirus (ACDACV) and *Macrophomina phaseolina* chrysovirus 1 (MpCV1) are in fact the counterparts of PcV dsRNA 4 rather than dsRNA 3. Likewise, dsRNA 4 of these three chrysoviruses are the counterparts of PcV dsRNA3. Since PcV was

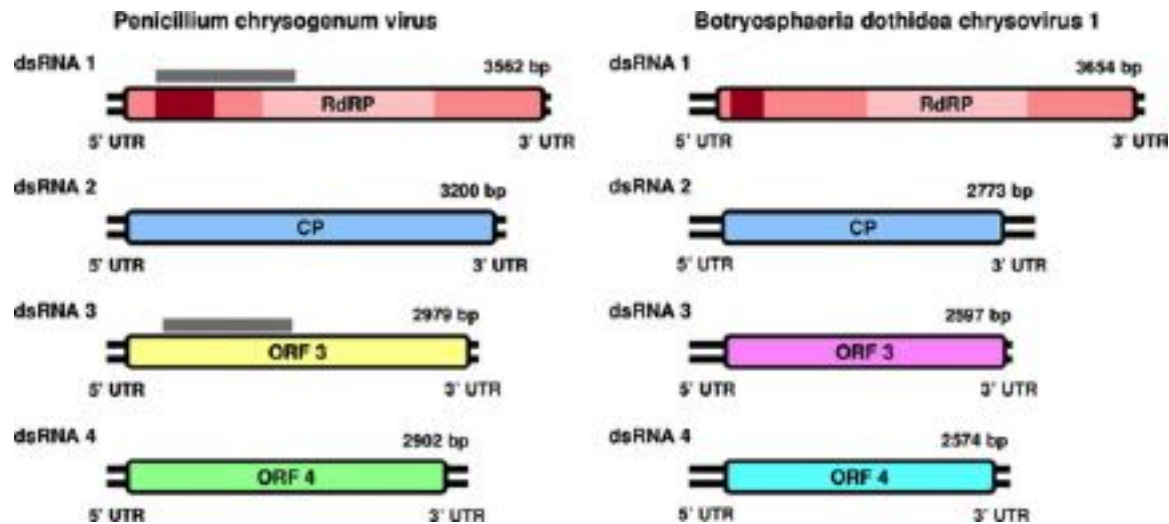


Fig. 3 Schematic representation of the genomic organization of *Penicillium chrysogenum* virus (PcV), exemplar virus for the type species of genus *Alphachrysovirus*, and *Botryosphaeria dothidea* chrysovirus 1 (BdCV1), exemplar virus for the type species of genus *Betachrysovirus*. Each genome consists of four dsRNAs, each containing one open reading frame (ORF; colored boxes) flanked by 5'- and 3'-untranslated regions (UTRs; black double lines). The light colored box in dsRNA 1 represents the RdRP₄ motif; the dark colored box in dsRNA 1 represents an independent P-loop NTPase domain predicted by HHpred; the gray colored thick lines represent a region of homology between the proteins encoded by dsRNAs 1 and 3.

the first chrysovirus to be characterized at the molecular level and to avoid confusion, the protein designations P3 and P4 as used for PcV were adopted and referred to as alphachryso-P3 and alphachryso-P4. Therefore, whereas the alphachryso-P3 protein represents the gene product of PcV dsRNA 3, it comprises the corresponding gene product of, for instance, HvV145S dsRNA4. Additionally, some members of the genus *Alphachrysovirus*, including *Brassica campestris* chrysovirus 1 (BcCV1), *Colletotrichum gloeosporioides* chrysovirus 1 (CgCV1), *Fusarium oxysporum* chrysovirus 1 (FoCV1), *Persea americana* chrysovirus (PaCV), and *Raphanus sativus* chrysovirus 1 (RsCV1) appear to have only 3 dsRNA segments, lacking the dsRNA encoding the alphachryso-P3 protein.

Regarding members of the genus *Betachrysovirus*, the RdRp is encoded by the largest segment, dsRNA 1, while the major CP is usually encoded by dsRNA 2, but it may also be encoded by dsRNA 3, in the case of *Fusarium graminearum* dsRNA mycovirus 2 (FgV2) and *Fusarium oxysporum* f. sp. *dianthi* mycovirus 1 (FodV), or dsRNA 4, in the case of *Magnaporthe oryzae* chrysovirus 1-A (MoCV1-A). It should be noted that betachrysovirus CPs share no detectable homology with alphachrysovirus CPs. Similarly to the approach used for alphachrysoviruses, the betachrysovirus proteins encoded by dsRNA 3 and 4 of the exemplar virus BdCV1 were designated as betachryso-P3 and betachryso-P4, respectively, and are not necessarily produced by dsRNA 3 and 4 of all known betachrysoviruses. For instance, MoCV1-A dsRNA 2 encodes betachryso-P3, while dsRNA 3 encodes betachryso-P4. Unlike alphachrysoviruses, no betachrysoviruses with 3 genomic segments have been reported; however, betachrysoviruses with five genomic segments, such as MoCV1-A, FgV2 and *Alternaria alternata* chrysovirus 1 (AaCV1), and one Betachrysovirus with 7 genomic segments, *Colletotrichum fruticola* chrysovirus 1 (CfCV1), has been described.

Although the proteins alphachryso-P3 and alphachryso-P4, encoded by PcV dsRNA 3 and dsRNA 4, respectively, are of unknown function, protein database searches reveal that PcV alphachryso-P3 sequence shares a 'phytoreo S7 domain' with a family consisting of several phytoreovirus P7 proteins known to be viral core proteins with nucleic acid binding activities (see below). The PcV alphachryso-P4 (and comparable proteins of other alphachrysoviruses) contains the motifs that form the conserved core of the ovarian tumor gene-like superfamily of predicted cysteine proteases. In contrast, little is known about the possible functions of the betachryso-P3 and betachryso-P4 proteins which share no homology with the alphachryso-P3 and alphachryso-P4 proteins or the betachrysoviruses proteins encoded by the dsRNAs 5–7.

The majority of the 5'- and 3'-untranslated regions (UTRs) of chrysovirus dsRNAs are relatively long, up to 698 nt and 865 nt, respectively (Table 1). The longest 5'-UTR belongs to CnCV1 dsRNA 4, while the longest 3'-UTR belongs to MoCV1-A dsRNA 5. The only 5'-UTR shorter than 50 nt belongs to dsRNA 1 of the insect-associated Shuangao chryo-like virus (SCLV). Both PcV (alphachrysovirus) and BdCV1 (betachrysovirus) UTRs have strictly conserved 5'- and 3'-termini (Fig. 4) and there appear to be some sequence similarity between them; for instance the poly A sequence at the 5'-terminus or the 'GUGU' sequence at the 3'-terminus. In addition, a 40–75 nt region with high sequence identity is present in the 5'-UTR of all four PcV dsRNAs. A second region of strong sequence similarity is present immediately downstream, which consists of a stretch of 30–50 nt containing a reiteration of the sequence 'CAA'. The (CAA)_n repeats are found to a lesser extent in the 5'-UTRs of BdCV1 and are similar to the enhancer elements present at the 5' UTRs of tobamoviruses.

		5'-UTRs	3'-UTRs
PcV	dsRNA 1	gataaaaaaagaataattcctttgg	ttaggctttaaaataagtgt
	dsRNA 2	gataaaaaaacaataattccgcagc	ttaggctttaaaataagtgt
	dsRNA 3	gataaaaaaacgataagcggtagcc	cgagggtttaaaataagtgt
	dsRNA 4	gataaaaaaacaataattcctagca	ttcggctttaaaataagtgt
		***** .****	. ** .*****
BdCV1	dsRNA 1	cgcaaaaaagaagaaaaggggaact	cgtgctgctagaaattgtgt
	dsRNA 2	cgcaaaaaagaagaaaaggggttaag	atgcgctccagaaattgtgt
	dsRNA 3	cgcaaaaaagaagaaaaggggagat	agtggcgaggatattgtgt
	dsRNA 4	cgcaaaaaagaagaaaaggggaaag	gtgctgcttagaaattgtgt
		*****	. ** *****

Fig. 4 Comparison of the 5'- and 3'-untranslated regions (UTR; termini of all dsRNAs of the alphachrysovirus *Penicillium chrysogenum* virus (PcV) and the betachrysovirus *Botryosphaeria dothidea* chrysovirus 1 (BdCV1). Asterisks and colons signify identical and highly conserved nucleotides at the indicated positions, respectively. Identical nucleotides in the same position for both viruses are shaded.

	1	2	3	4	5	6	7	8
A AfuCV	LVGRS [73]	WLTGSLVY [64]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDIVL [33]	EFFRN [12]	RALAS
PcV	LVGRG [73]	WLTGGLVY [64]	NKDRILL [52]	OWADPNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDIDL [33]	EFFRN [12]	RALAS
IjCV1	LVGRG [73]	WLTGGLVY [64]	NKDRILL [52]	OWADPNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDIDL [33]	EFFRN [12]	RALAS
MpCV1	LIGRA [73]	WLTGSLVY [64]	NKDRILL [52]	OWADPNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDIDL [33]	EFFRN [12]	RALAS
CnCV1	LVGR [74]	WLTGSLVY [64]	NKDRILL [52]	OWADPNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
VdCV1	LIGRT [74]	WLTGSLVY [65]	NKDRILL [52]	OWADPNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDIDS [33]	EFFRN [12]	RALAS
HvV145S	LLGR [73]	WLTGSLVY [64]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
PcCV1	LLGKE [73]	WLTGSTVF [64]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
CgCV1	MLGKE [72]	WLTGSTVF [64]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
ACDACY	LVGR [73]	WLTGSTVF [64]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
BoCV1	YTGR [72]	WVAGSTVF [60]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
ReCV1	YTGR [72]	WVAGSTVF [60]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
AMAV	LYGR [70]	WVAGSTVF [59]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
PcCV	LYGR [70]	WVAGSTVF [59]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
SCLV	LYGR [69]	WVAGSTVF [59]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
B FyV2	LAGRS [77]	WVAGSATG [79]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
PdV	LAGRS [77]	WVAGSATG [79]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
MoCV1-A	LAGRY [76]	WVAGSATG [76]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
PjCV1	LVGRH [76]	WVAGSATG [76]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
CfCV1	LAGRL [78]	WVAGSATG [76]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
PjCV2	LAGRV [75]	WVAGSATG [75]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
BdCV1	LAGRL [75]	WVAGSATG [76]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS
AaCV1	LAGRI [75]	WVAGSATG [76]	NKDRILL [52]	OWANFNEQHS [49]	GLTSGNRGTTWINTVLNFC [22]	HGGDDVDA [33]	EFFRN [12]	RALAS

Fig. 5 Comparison of the eight conserved motifs of all chrysovirus RdRps presented in Table 1. Numbers in parentheses correspond to amino acid residues present between the motifs. Asterisks and colons respectively signify identical and highly conserved amino acid residues at the positions indicated.

Genome Expression and Replication

Chrysovirus RdRps

The largest dsRNA segment (dsRNA 1) of all chrysoviruses so far sequenced contains a single large ORF encoding the RdRp. The molecular mass of chrysovirus RdRps ranges from 110 to 131 kDa (Table 1). These values are consistent with those estimated by SDS-PAGE of the *in vitro* translation products of full-length transcripts derived from both PcV and Hv145SV dsRNA 1 cDNAs. Examination of the deduced amino acid sequence of the RdRp ORF reveals the presence of the eight conserved motifs characteristic of RdRps of dsRNA viruses present in simple eukaryotes (Fig. 5). Additionally, there is an independent P-loop NTPase domain at the N-terminus of alphachrysovirus and betachrysovirus RdRps as shown by HHpred, a remote protein homology detection software. In the case of alphachrysoviruses this domain overlaps with a region homologous to alphachryso-P3. Members of the P-loop NTPase domain superfamily are characterized by conserved nucleotide phosphate-binding motifs, also referred to as the Walker A motif (GxxxxGK[S/T]) and the Walker B motif (hhhh[D/E], where h is a hydrophobic residue).

Chrysovirus CPs

The second largest dsRNA segment (dsRNA 2) of alphachrysoviruses currently sequenced contains a single large ORF encoding the CP, while the CP of betachrysoviruses currently sequenced may also be produced by dsRNA 3 or 4. The molecular mass of chrysovirus CPs ranges from 80 to 126 kDa (Table 1). The predicted size of PcV CP (109 kDa) is similar to that estimated by

SDS-PAGE of purified PcV virions as well as that determined for the *in vitro* translation product of a full-length transcript of dsRNA 2 cDNA. Direct evidence that dsRNA 2 of both the alphachrysovirus PcV and the betachrysovirus BdCV1 encodes CP was provided by amino acid sequencing of a tryptic peptide derived from a gradient-purified PcV capsid.

Alphachryso-P3 Shares a 'Phytoreo S7 Domain' With Core Proteins of Phytoreoviruses

DsRNA 3 of PcV, *Aspergillus fumigatus* chrysovirus (AfuCV), *Isaria javanica* chrysovirus 1 (IjCV1), and *Verticillium dahliae* chrysovirus 1 (VdCV1) encodes the alphachryso-P3 protein, whereas dsRNA 4 of Hv145SV, ACDACV, and MpCV1 encodes the corresponding alphachryso-P3. Although the function of alphachryso-P3 is not known, sequence analysis and database searches offer some clues. ProDom database searches reveal that chryso-P3 sequences share a 'phytoreo S7 domain' with a family consisting of several phytoreovirus P7 proteins known to be viral core proteins with nucleic acid binding activities. The consensus for the three chrysoviruses is [X(V/I)V(M/L)P(A/M)G(C/H)GK(T/S)T-(L/I)]. Phytoreovirus P7 proteins bind to their corresponding P1 (transcriptase/replicase) proteins, which bind to the genomic dsRNAs. It is of interest, in this regard, that the N-terminal regions of all alphachryso-P3s (encompassing the amino acids within positions 1–500) share significant sequence similarity with comparable N-terminal regions of the putative RdRps encoded by chrysovirus dsRNA1s. The regions in the dsRNA 1-encoded proteins with high similarity to alphachryso-P3 occur upstream of the eight highly conserved motifs characteristic of RdRps of dsRNA viruses of simple eukaryotes. The significance of this sequence similarity to the function of alphachryso-P3 is not known for certain, but one may speculate that the N-terminal region of these proteins might play a role in viral RNA binding and packaging. In contrast, betachryso-P3 shares no detectable homology with alphachryso-P3, contains no phytoreo S7 domain and has no similarity with the betachrysovirus RdRps.

Chryso-P4 is a Putative Protease and Virion Associated as a Minor Protein

Present evidence, based on amino acid sequencing of a tryptic peptide derived from gradient-purified PcV virions, strongly supports the conclusion that PcV alphachryso-P4 is a virion-associated minor protein. Additionally, alphachryso-P4 encoded by some chrysoviruses contains the motif PGDGXCXXHX. This motif (I), along with motifs II (with a conserved K), III, and IV (with a conserved H), form the conserved core of the ovarian tumor gene-like superfamily of predicted cysteine proteases. Multiple alignments showed that motifs I–IV are also present in other viruses including *Agaricus bisporus* virus 1, a tentative member of the family *Chrysoviridae*. Whether the RNAs of these viruses indeed code for the predicted proteases remains to be investigated.

Replication of Chrysoviruses

There is very limited information on how chrysoviruses replicate their dsRNAs. The virion-associated RdRp catalyzes *in vitro* end-to-end transcription of each dsRNA to produce mRNA by a conservative mechanism. Purified virions containing both ssRNA and dsRNA have been isolated from *Penicillium* spp. infected with PcV or Pc-fV and may represent replication intermediates.

Taxonomy and Phylogenetic Analysis

The *Penicillium* chrysoviruses PcV, Pc-fV, and PbV are serologically related and have similar biochemical and biophysical properties. Although molecular data is only available for PcV, the three viruses could be considered as strains of the same virus for all practical purposes. The fact that these closely related viruses occur in different fungal species suggested that transmission by means other than anastomosis may occur naturally, since heterokaryon formation between different fungal species is unlikely. Horizontal transmission of fungal viruses in nature however, has yet to be demonstrated and in the case of viruses of *Penicillium* spp. may not need to occur since the viruses replicate in parallel with their hosts and are carried intracellularly during vegetative growth of the host. Furthermore, the viruses are efficiently disseminated by vertical transmission *via* the conidia of *Penicillium* spp. It seems feasible, however, that virus infection arose early in the phylogeny of *P. brevi-compactum*, *P. chrysogenum*, and *P. cyaneo-fulvum* before they diverged and that the resident virus remained associated with them during their subsequent evolution.

BLAST searches of chrysovirus RdRp amino acid sequence showed significantly high sequence similarity with the RdRps of several members of the family *Totiviridae*. Interestingly, no significant hits were evident with any of the viruses in family *Partitiviridae*, another validation for the removal of chrysoviruses from the family *Partitiviridae* and their placement in the newly created family *Chrysoviridae*. The conclusion that chrysovirus RdRps are more closely related to those of totiviruses than to those of partitiviruses is supported by published results of phylogenetic analysis of RdRp conserved motifs and flanking sequences of chrysoviruses and viruses in families *Totiviridae* and *Partitiviridae* and comparisons of the conserved motifs of chrysovirus RdRps with those of totiviruses and partitiviruses. For instance, the database Pfam, which includes annotations and multiple sequence alignments generated using hidden Markov models, classifies chrysovirus and totivirus RdRps in of the protein family RdRp_4 (PF02123), while partitivirus RdRps belong to RdRp_1 (PF00680).

The recent discoveries of new viruses related to chrysoviruses have led to an expansion and reorganization of the family *Chrysoviridae*. The family has now two genera, *Alphachrysovirus* and *Betachrysovirus*, accommodating seventeen and eight species

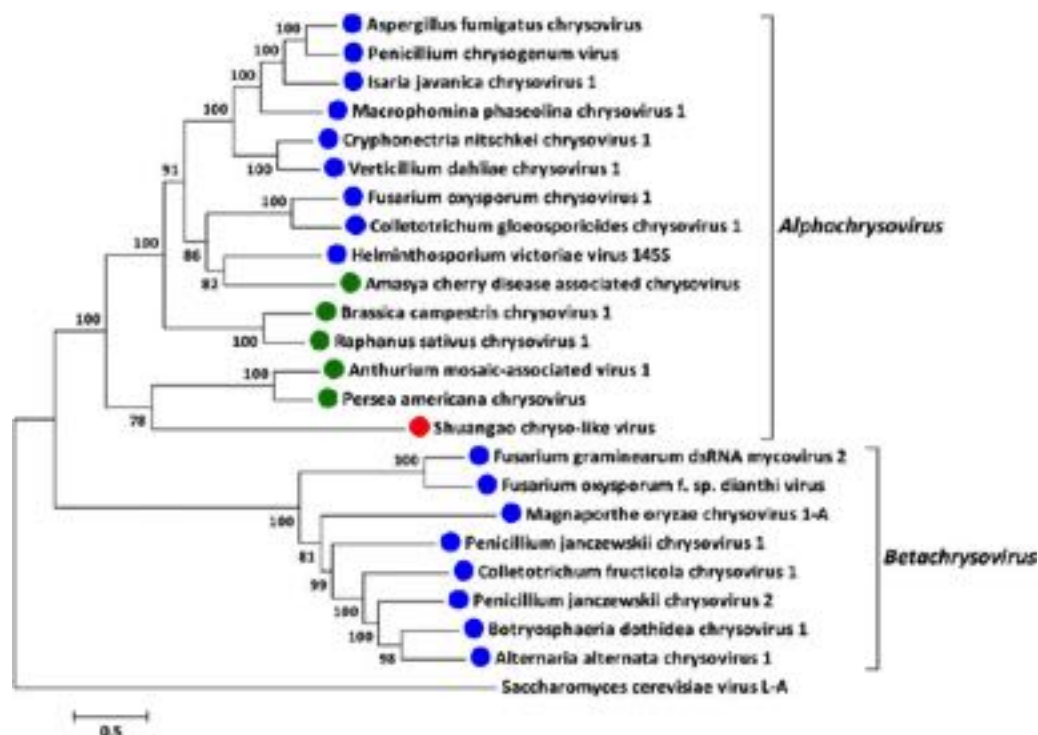


Fig. 6 Maximum likelihood phylogenetic tree created based on the RdRp sequences of chrysoviruses. The sequences were aligned with MUSCLE as implemented by MEGA 6, all positions with less than 30% site coverage were eliminated and the LG + G + I + F substitution model was used. At the end of the branches, blue circles indicate that the virus infects fungi; green circles indicate that the virus infects or is associated with plants; red circles indicate that the virus infects or is associated with insects. *Saccharomyces cerevisiae* virus L-A (ScV-L-A), a member of the genus *Totivirus*, family *Totiviridae*, was used as the outgroup.

respectively, based on the formation of two distinct clades as evidenced by phylogenetic analysis (Fig. 6). The analysis was based on the sequence of chrysovirus RdRp, which is the only protein with detectable homology among all viruses belonging to the two genera. The rest of the viral proteins, including the CP, are conserved within but not among genera. The one exception is SCLV, an insect-associated virus belonging to the genus *Alphachrysovirus*, whose proteins, the RdRp notwithstanding, are non-homologous to those of the other members of genus *Alphachrysovirus* or *Betachrysovirus*.

An interesting feature of chrysoviruses is a variability in the number of their genomic segments, which range from three to seven. More specifically, the known members of the genus *Alphachrysovirus* can have three or four segments as their genome. Six members of the genus *Alphachrysovirus* have three genomic segments and most of them infect or are associated with plants. Therefore alphachryso-P3 is not necessary for virus replication, at least under certain conditions that have not been characterized yet. Additionally, since FoCV1 alphachryso-P4 appears to be non-functional due to the presence of nonsense mutations in the open reading frame, it is feasible that the replication cycle of chrysoviruses can be completed in the presence of only two proteins, the RdRp and the CP.

The members of the genus *Betachrysovirus*, a very recent addition to the family *Chrysoviridae*, are not as well characterized. They typically have four or five genomic segments and, in the case of CfCV1, up to seven segments. The RdRp, the CPs, and betachryso-P3 and betachryso-P4 are conserved among all betachrysoviruses, while the rest, if present, are non-homologous among members of the genus. The RdRp of the betachrysoviruses is most closely related to that of their sister taxon *Alphachrysovirus*, but the proteins encoded from the other segments, including the CP, demonstrate no clear homology with those of the alphachrysoviruses or other proteins available in the databases.

Biology and Effects of Chrysoviruses on Fungal Hosts

Mycoviruses may reduce or increase the virulence of their fungal hosts and as such are considered to have potential as biological control agents. *Cryphonectria hypovirus 1* (CHV1), a positive-sense single-stranded RNA mycovirus belonging to the family *Hypoviridae*, is the best example for attenuation of the host fungus of *Cryphonectria parasitica*. Viruses belonging to the family *Chrysoviridae* have also been reported to alter the pathogenicity of their host fungi. Within the genus *Alphachrysovirus*, AfuCV reduces growth and pathogenicity of its human pathogenic host. Additionally, five mycoviruses which belong to genus *Betachrysovirus*, have been reported

to alter the pathogenicity of their host fungi against plants including MoCV1-A, AaCV1, BdCV1, *Fusarium graminearum* mycovirus-China 9 (FgV-Ch-9) and *Aspergillus thermomutatus* chrysovirus 1 (AthCV1). For instance, MoCV1-A infection generally confers hypovirulence to the fungus but is also a driving force to generate different pathogenic races of the fungus. In another example, AaCV1 exhibits two contrasting effects: it impairs the growth of the host fungus while rendering the fungus hypervirulent to the plant, with increased production of a host-specific toxin.

MoCV1-A and MoCV1-B are the first viruses reported to cause hypovirulence traits in the rice blast pathogen *Magnaporthe oryzae*, including impaired growth of host cells, altered colony morphology, and reduced pigmentation. The hyphal morphology of the host fungus infected with MoCV1-B revealed a remarkable albino phenotype on potato dextrose agar (PDA) medium, indicating that melanin biosynthesis appears to be suppressed by MoCV1-B infection. The cell wall of a MoCV1-B infected strain was loose and enlarged, and staining with calcofluor-white showed that the cell wall was also damaged by MoCV1-B infection. No conidial formation was observed *M. oryzae* infected with MoCV1-B on PDA. Pathogenic races of *M. oryzae* are determined by a gene-for-gene system, where an avirulence gene in the pathogen induces disease resistance in a rice variety with a corresponding resistance (R) gene. To examine whether MoCV1-A infection affects pathogenic races of *M. oryzae*, a virus-free and a MoCV1-A-infected *M. oryzae* strain were inoculated onto different rice varieties. Inoculation of the R gene-free rice variety showed that MoCV1-A infection resulted in reduced fungal virulence. However, when spray or leaf-sheath inoculation methods were used to inoculate monogenic rice lines carrying different R genes MoCV1-A-infected and MoCV1-A-free *M. oryzae* strains caused different lesion types (resistance to susceptible or *vice versa*) on individual rice varieties. These data suggest that MoCV1-A infection can alter the pathogenicity of the host *M. oryzae* from avirulence to virulence, or from virulence to avirulence, depending on the rice variety.

Recently, it has been found that infection of the Japanese pear pathotype fungus, *Alternaria alternata* with AaCV1 simultaneously impaired growth of the host fungus and increased levels of the host-specific AK-toxin. This is another example of a mycovirus infection causing changes in pathogenic races of the host fungus, because *A. alternata* can infect some specific varieties of Japanese pear (*Pyrus pyrifolia* Nakai) cultivar Nijisseiki, but not others. It is likely that the enhancement of host fungal pathogenicity in some varieties, without any mutation in the avirulence genes, is one of the strategies used by mycoviruses to survive in the agricultural ecosystem, where humans cultivate many plant varieties with different R genes to reduce damage caused by plant disease. Mycoviruses may increase survival rates by retaining the diversity of avirulence genes in their host fungi, which are important for fungal adaptation to plants. Since only limited sequence information is currently available for betachrysovirus dsRNA genomes it is difficult to assign potential roles for individual dsRNA encoded products in plant pathogenicity. Recently attempts at dissecting the roles of the five MoCV1-A dsRNA encoded proteins in pathogenicity have been reported using over expressed proteins. The putative protein encoded by MoCV1-A dsRNA1 contains eight conserved motifs characteristic of RdRps and is assumed to function in this role. However, based on BLAST searches, it was not possible to predict the functions of the other four putative proteins. Therefore, shuttle vectors were constructed by ligating each of the remaining four MoCV1-A ORF sequences predicted from the genomic sequences of dsRNAs 2–5 downstream of a low expression promoter (*ADH1*) or a high expression promoter (*TDH3*) for expression in yeast cells. The influence of each protein on yeast cell growth was investigated in liquid cultures. Optical density, viable cell number, glucose concentration, and pH were measured together with observations of cell morphology and immunological analyses. Abnormalities in cell morphology including the appearance of enlarged vacuoles and vesicles were observed when the CP (encoded by dsRNA 4; [Table 1](#)) was over expressed in yeast cells. A series of cultivation tests revealed that CP expression also caused a decrease in the rate of cell proliferation and a decrease in cell life span. Over expression of the CP in the human pathogenic budding yeast *Cryptococcus neoformans* also caused a decrease in the growth rate, increase in emergence and enlargement of vacuoles. Additionally, the formation of capsules, which are involved in the pathogenicity of *C. neoformans*, was also reduced in the CP-expressing cells, suggesting a reduction in pathogenic potential. Expression of a CP-GFP fusion protein in *S. cerevisiae* resulted in the formation of abnormal cell aggregates. In MoCV1-A, CP expression also resulted in significant inhibition of growth at high temperatures (35°C and 37°C) as compared to cells grown at the optimal temperature (30°C) plus reduced expression of stress-response genes and increased expression of translation-related genes. The MoCV1-A CP showed significant similarity to related proteins from other viruses in genus *Betachrysovirus*. Multiple alignments of the CP-related protein sequences showed that their central regions (aa 210–591 in MoCV1-A CP) are relatively conserved. Indeed, yeast transformants expressing the conserved central region of the MoCV1-A CP protein (325–575 aa) showed similar impaired growth phenotypes to those observed in yeast expressing the full-length MoCV1-A CP protein. The yeast heterologous expression system revealed that the CP proteins of MoCV1-A and AaCV1 (in this case encoded by dsRNA 2) were responsible for growth inhibition of the yeasts *S. cerevisiae* and *C. neoformans*, suggesting that the CP in addition to its structural role may have been preserved in betachrysoviruses as a protein implicated in hypovirulence. This hypothesis is supported by a recent report showing that the structural protein of another betachrysovirus, FgV-Ch-9 (closely related to FgV2), acts as symptom determinant.

Acknowledgment

This work is dedicated to the memory of our friend and colleague Said Ghabrial, who sequenced the first chrysovirus in 2000 and passed away in November 2018.

Further Reading

- Bhatti, M.F., Jamal, A., Petrou, M.A., *et al.*, 2011. The effects of dsRNA mycoviruses on growth and murine virulence of *Aspergillus fumigatus*. *Fungal Genetics and Biology* 48, 1071–1075.
- Ejmal, M.A., Holland, D.J., MacDiarmid, R.M., Pearson, M.N., 2018. The effect of *Aspergillus thermomutatus* chrysovirus 1 on the biology of three *Aspergillus* species. *Viruses* 10, pii: E539.
- Gómez-Blanco, J., Luque, D., Gonzalez, J.M., *et al.*, 2012. Cryphonectria nitschkei virus 1 structure shows that the capsid protein of chrysoviruses is a duplicated helix-rich fold conserved in fungal double-stranded RNA viruses. *Journal of Virology* 86, 8314–8318.
- Luque, D., Gómez-Blanco, J., Garriga, D., *et al.*, 2014. Cryo-EM near-atomic structure of a dsRNA fungal virus shows ancient structural motifs preserved in the dsRNA viral lineage. *Proceedings of the National Academy of Sciences of the United States of America* 111, 7641–7646.
- Luque, D., González, J.M., Garriga, D., *et al.*, 2010. The T=1 capsid protein of *Penicillium chrysogenum* virus is formed by a repeated helix-rich core indicative of gene duplication. *Journal of Virology* 84, 7256–7266.
- Moriyama, H., Urayama, S.I., Higashiura, T., Le, T.M., Komatsu, K., 2018. Chrysovirus in *Magnaporthe oryzae*. *Viruses* 10, E697.
- Okada, R., Ichinose, S., Takeshita, K., *et al.*, 2018. Molecular characterization of a novel mycovirus in *Alternaria alternata* manifesting two-sided effects: Down-regulation of host growth and up-regulation of host plant pathogenicity. *Virology* 519, 23–32.
- Urayama, S., Kimura, Y., Katoh, Y., *et al.*, 2016. Suppressing effects of mycoviral proteins encoded by *Magnaporthe oryzae* chrysovirus 1 strain A on conidial germination of the rice blast fungus. *Virus Research* 223, 10–19.
- Wang, L., Jiang, J., Wang, Y., *et al.*, 2014. Hypovirulence of the phytopathogenic fungus *Botryosphaeria dothidea*: Association with a coinfecting chrysovirus and a partitivirus. *Journal of Virology* 88, 7517–7527.
- Zhai, L., Zhang, M., Hong, N., *et al.*, 2018. Identification and characterization of a novel hepta-segmented dsRNA virus from the phytopathogenic fungus *Colletotrichum fructicola*. *Frontiers in Microbiology* 9, 754.

Fungal Partitiviruses (*Partitiviridae*)

Eeva J Vainio, Natural Resources Institute Finland (Luke), Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S. Tavantzis, Partitiviruses of Fungi, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00405-2.

Glossary

Ascomycete/basidiomycete Ascomycetes (phylum Ascomycota) are fungi that produce sexual spores inside sac-like structures called asci, whereas basidiomycetes (phylum Basidiomycota) produce sexual spores externally on specialized cells called basidia.

Capsid Protective shell composed of multiple copies of protein subunits that encapsidate the viral genome.

Heterokaryotic/homokaryotic Nuclear condition of hyphal cells among many basidiomycetes. The germination of a single basidiospore results in the formation of a homokaryotic (i.e., multinucleate but effectively haploid) primary mycelium. A compatible mating between two homokaryotic mycelia produces a heterokaryotic (i.e., multinucleate but effectively diploid) secondary mycelium with two different nuclear haplotypes. Two different heterokaryons are somatically incompatible unless

they are very closely related. In ascomycetes, the vegetative mycelium is usually haploid and uninucleate.

Hyphal anastomosis Fusion of hyphal cells allowing cytoplasmic exchange between contacting mycelia.

Hypovirulence Reduced ability to cause disease. Often associated with the presence of mycoviruses in plant pathogenic fungi.

Icosahedral symmetry Arrangement of the capsid protein subunits of isometric (spherical) viruses in the symmetry of an icosahedron consisting of 60 identical objects or asymmetric units. In the so called “ $T = 2$ ” symmetry, 120 identical protein monomers form 60 asymmetric dimers.

Somatic incompatibility Process of allorecognition in fungi that keeps different fungal genotypes separated. Somatic (vegetative) incompatibility leads to programmed cell death of heterokaryotic cells formed after anastomosis between hyphae of genetically incompatible strains.

Introduction

Viruses infecting fungi (mycoviruses) were first discovered by M. Hollings in diseased strains of the commercially cultivated mushroom, *Agaricus bisporus* in the 1960s. Later during the same decade, the first members of current virus family *Partitiviridae* were found in isolates of the ascomycetous mold fungus *Penicillium* spp. Mycoviruses are now known to be very common among fungi. They are found in diverse fungal phyla including ascomycetes, basidiomycetes, chytridiomycetes and zygomycetes with various lifestyles ranging from symbionts to pathogens. The following classified virus families include viruses infecting fungi: *Alphaflexiviridae*, *Barnaviridae*, *Botourmiaviridae*, *Deltaflexiviridae*, *Endornaviridae*, *Gammapartitiviridae*, *Hypoviridae* and *Narnaviridae* with positive-sense single-stranded (ss) RNA genomes; *Myonnaviridae* with negative-sense ssRNA genomes; *Amalgaviridae*, *Chrysovriidae*, *Megabirnaviridae*, *Partitiviridae*, *Reoviridae*, *Totiviridae* and *Quadriviridae* with double-stranded (ds) RNA genomes; and *Genomoviridae* with ssDNA genomes. In addition, the classified genus *Botybirnavirus* includes dsRNA viruses found so far only in fungi. Virus families *Metaviridae* and *Pseudoviridae* contain retrotransposable elements found in ascomycetous yeast fungi, plants, insects and nematodes.

Viruses belonging to the family *Partitiviridae* are classified into five different genera with characteristic hosts for members of each genus: either plants or fungi for genera *Alphapartitivirus* and *Betapartitivirus*; fungi for genus *Gammapartitivirus*, plants for genus *Deltapartitivirus* and protozoa for genus *Cryspovirus*. The name “partitivirus” originates from the Latin word “*partitius*”, which means “divided” and refers to the genome structure of these viruses, which is bipartite and consists of two linear monocistronic dsRNAs. The term “cryptovirus” (from Greek “*crypto*,” meaning “hidden, covered, or secret”) was formerly used in genus names of plant-infecting *Partitiviridae* members, but since many of the fungal partitiviruses can also be considered cryptic (symptomless), this term is no longer used to differentiate between plant and fungal partitiviruses. Plant and protozoal partitiviruses are discussed in detail elsewhere in this encyclopedia.

Classification

This article focuses on the genera *Alphapartitivirus*, *Betapartitivirus*, and *Gammapartitivirus*, which include viruses that infect filamentous fungi. In addition to characteristic hosts ranges within each genus, partitivirus genera are demarcated based on genome segment and protein lengths within a typical size range for each genus, separate phylogenetic grouping of RdRP sequences from each genus, and level of amino acid sequence identity (less than 24% in pairwise comparisons of viruses from different genera).

There are currently 28 fungal partitiviruses that have been assigned to a genus: ten alphapartitiviruses, ten betapartitiviruses, and eight gammapartitiviruses. In addition, the genomes of several more partitiviruses have been completely sequenced and

Table 1 List of fungal viruses in the family *Partitiviridae* classified by the ICTV. The genera *Alpha-* *Beta-* and *Gammapartivirus* are indicated using the corresponding Greek alphabets. The two essential genome segments are named as dsRNA1 and dsRNA2

Virus species	Genus	GenBank accession numbers			
		dsRNA1/RdRP	dsRNA2/CP	dsRNA3	dsRNA4
<i>Cherry chlorotic rusty spot associated partitivirus</i>	α	AJ781168	AJ781167	AM749118	AM749120
<i>Chondrostereum purpureum cryptic virus 1</i>	α	AM999771	AM999772		
<i>Flammulina velutipes browning virus</i>	α	AB465308	AB465309		
<i>Helicobasidium mompa partitivirus V70</i>	α	AB025903			
<i>Heterobasidium partitivirus 1</i>	α	HQ541323	HQ541324		
<i>Heterobasidium partitivirus 3</i>	α	FJ816271	FJ816272		
<i>Heterobasidium partitivirus 12</i>	α	KF963175	KF963176		
<i>Heterobasidium partitivirus 13</i>	α	KF963177	KF963178		
<i>Heterobasidium partitivirus 15</i>	α	KF963186	KF963187		
<i>Rosellinia necatrix partitivirus 2</i>	α	AB569997	AB569998		
<i>Atkinsonella hypoxylon virus</i>	β	L39125	L39126	L39127	
<i>Ceratocystis resinifera virus 1</i>	β	AY603052	AY603051		
<i>Fusarium poae virus 1</i>	β	AF047013	AF015924		
<i>Heterobasidium partitivirus 2</i>	β	HM565953	HM565954		
<i>Heterobasidium partitivirus 7</i>	β	JN606091	JN606090		
<i>Heterobasidium partitivirus 8</i>	β	JX625227	JX625228		
<i>Heterobasidium partitivirus P</i>	β	AF473549			
<i>Pleurotus ostreatus virus 1</i>	β	AY533038	AY533036		
<i>Rhizoctonia solani virus 717</i>	β	AF133290	AF133291		
<i>Rosellinia necatrix virus 1</i>	β	AB113347	AB113348		
<i>Aspergillus ochraceus virus</i>	γ	EU118277	EU118278	EU118279	
<i>Discula destructiva virus 1</i>	γ	AF316992	AF316993	AF316994	AF316995
<i>Discula destructiva virus 2</i>	γ	AY033436	AY033437		
<i>Fusarium solani virus 1</i>	γ	D55668	D55669		
<i>Gremmeniella abietina RNA virus MS1</i>	γ	AY089993	AY089994	AY089995	
<i>Ophiostoma partitivirus 1</i>	γ	AM087202	AM087203		
<i>Penicillium stoloniferum virus F</i>	γ	AY738336	AY738337	AY738338	
<i>Penicillium stoloniferum virus S</i>	γ	AY156521	AY156522		

described in scientific publications. The classified species of the family *Partitiviridae* found in fungi are listed in [Table 1](#), and tentative new *Partitiviridae* members with complete published genome sequences are shown in [Table 2](#). Sequence information is lacking from the classified partitiviruses *Agaricus bisporus virus 4*, and *Gaeumannomyces graminis* viruses 019/6-A and T1-A, and they remain unassigned to a genus.

Virion Structure

Partitivirus particles are isometric, non-enveloped, and 25–43 nm in diameter ([Fig. 1](#)). The capsid consists of 120 copies of a single capsid protein (CP) arranged as 60 dimers with T=1 icosahedral symmetry. Dimeric “arch-like” surface protrusions are frequently observed on partitivirus capsids. Virion buoyant densities in CsCl gradients range from 1.34 to 1.44 g cm⁻³ for particles with nucleic acid.

The structures of four fungal partitiviruses have been determined by electron cryomicroscopy and 3D reconstruction ([Fig. 2](#)). These include the gammapartitiviruses *Penicillium stoloniferum virus F* and *S* (PsV-F and PsV-S), and the betapartitiviruses *Fusarium poae virus 1* (FpV1) and *Sclerotinia sclerotiorum partitivirus 1* (SsPV1). The structure of PsV-F has also been resolved with X-ray crystallography.

Partitiviruses possess two essential genome segments that are individually encapsidated in separate particles. Virion-associated RNA polymerase activity is present, and one or two molecules of RNA dependent RNA polymerase (RdRP) are packaged inside each particle. Purified virion preparations contain mature virions and a heterogenous population of particles thought to be replicative intermediates that may contain only one ssRNA molecule (the genomic positive strand), or both ssRNA and dsRNA, or two molecules of dsRNA.

Genome

The two essential genome segments of partitiviruses are separately encapsidated, and each linear segment contains one large ORF on the positive strand RNA molecule. One segment/ORF (dsRNA1) encodes the RdRP, whereas the other segment/ORF (dsRNA2) encodes the capsid protein (CP) ([Fig. 3](#)). The CP encoding genome segment is usually smaller than the polymerase encoding

Table 2 List of tentative fungal partitiviruses with complete genome sequences. Putative affiliation with the classified genera *Alpha-* *Beta-* and *Gammapartitivirus* is indicated using the corresponding Greek alphabets. The two essential genome segments are named as dsRNA1 and dsRNA2

Virus isolate	Genus	GenBank accession numbers			
		dsRNA1/RdRP	dsRNA2/CP	dsRNA3	dsRNA4
Rhizoctonia solani partitivirus 2	α	KF372436	KF372437		
Rosellinia necatrix partitivirus 7	α	LC076694	LC076695		
Heterobasidion partitivirus 20	α	MG566085	MG566086		
Rhizoctonia solani partitivirus 3	α	KX914900	KX914901		
Rhizoctonia solani partitivirus 4	α	KX914902	KX914903		
Botrytis cinerea partitivirus 2	α	MG011707	MG011708		
Rhizoctonia oryzae-sativae partitivirus 1	α	MK015642	MK015643		
Hypomyces chrysospermus partitivirus 1	α	MK652147	MK652146		
Sclerotinia sclerotiorum partitivirus 1 strain SsPV1-WF – 1	β	JX297511	JX297510		
Ustilaginoidea virens partitivirus 3 strain HP – 30	β	KF680478	KF680479		
Rosellinia necatrix partitivirus 6 strain W113	β	LC010952 [†]	LC010953		
Fusarium poae partitivirus 2	β	LC150608	LC150609		
Lentinula edodes partitivirus 1	β	KX354971	KX354972		
Rosellinia necatrix partitivirus 8	β	LC314788	LC314789		
Grifola frondosa partitivirus 1	β	LC425601	LC425602		
Ustilaginoidea virens partitivirus 1	γ	KC503898	KC503899	KC503900	KC503901
Aspergillus fumigatus partitivirus 1 isolate 88	γ	FN376847	FN398100		
Colletotrichum acutatum partitivirus 1 isolate CaRV1	γ	KC572132	KC572133		
Ustilaginoidea virens partitivirus 2 isolate Uv0901	γ	KF361014	KF361015		
Verticillium dahliae partitivirus 1 strain Vd08284	γ	KC422244	KC422243		
Penicillium aurantiogriseum partitivirus 1	γ	KT601103	KT601104		
Pythium nunn partitivirus 1	γ	LC371062	LC371063		
Nigospora oryzae partitivirus 1	γ	MF742398	MF742399		
Talaromyces marneffe partitivirus 1	γ	KM235311	KM235304		
Pseudogymnoascus destructans partitivirus pa	γ	KY207543	KY207544		
Beauveria bassiana partitivirus 1	γ	LN896303	LN896304		
Beauveria bassiana partitivirus 2	γ	LN896305	LN896306		
Magnaporthe oryzae partitivirus 1	γ	KX119172	KX119173		
Botryosphaeria dothidea partitivirus 1	?	KF688740	KF688741	KF688742	
Alternaria alternata partitivirus 1	?	KY352402	KY352403		
Aspergillus fumigatus partitivirus 2	?	MH192991	MH192992		

segment. The total size of partitivirus genomes is typically 3.6–3.9 kbp for members of genus *Alphapartitivirus*, 4.3–4.8 kbp for members of genus *Betapartitivirus*, and 3.1–3.4 kbp for members of genus *Gammapartitivirus*. The genome segments are individually 1.4–2.4 kbp in size. **Table 3** shows the size ranges of individual genome segments in classified alpha- beta- and gammapartitiruses, as well as the predicted sizes of RdRP and capsid proteins.

The first 7–20 nucleotides in the 5′–non translated region of the coding RNA are usually highly conserved between the two genome segments of single viruses and may form stem-loop structures. Such conserved terminal sequences are believed to be involved in polymerase recognition for replication and/or RNA packaging. Members of genera *Alphapartitivirus* and *Betapartitivirus* typically possess poly(A) tracts which may be interrupted by other nucleotides near the coding-strand 3′ terminus of one or both genome segments, whereas gammapartitiruses usually lack terminal poly(A) tracts.

Besides the two essential genome segments, some partitiviruses have additional (satellite or defective) dsRNA elements. Apparent RNA satellites encoding small (237–303 aa) proteins, homologous to each other but not to either CP or RdRP, are present in several classified and tentative gammapartitiruses, including *Aspergillus ochraceus* virus, *Discula destructiva* virus 1, *Gremmeniella abietina* RNA virus MS1, and *Ustilaginoidea virens* partitivirus 1 (**Tables 1** and **2**). RNA satellites appearing not to encode any proteins are found associated with studied isolates of the alphapartitivirus Cherry chlorotic rusty spot associated partitivirus, the betapartitivirus *Atkinsonella hypoxylon* virus, and gammapartitiruses *Discula destructiva* virus 1 and *Penicillium stoloniferum* virus F (**Tables 1** and **2**). The third genome segment of alphapartitivirus *Rosellinia necatrix* partitivirus 2 (**Table 1**) is a truncated version of the RdRP that may function as defective-interfering RNA.

Life Cycle

The model for partitivirus replication has been inferred based on in vitro studies of virion-associated RdRP of the gammapartitivirus *Penicillium stoloniferum* virus S, and isolation of particles that represent various stages in the replication cycle from naturally infected mycelium (**Fig. 4**). The virion-associated RdRP mediates positive-strand RNA synthesis within the virus particle

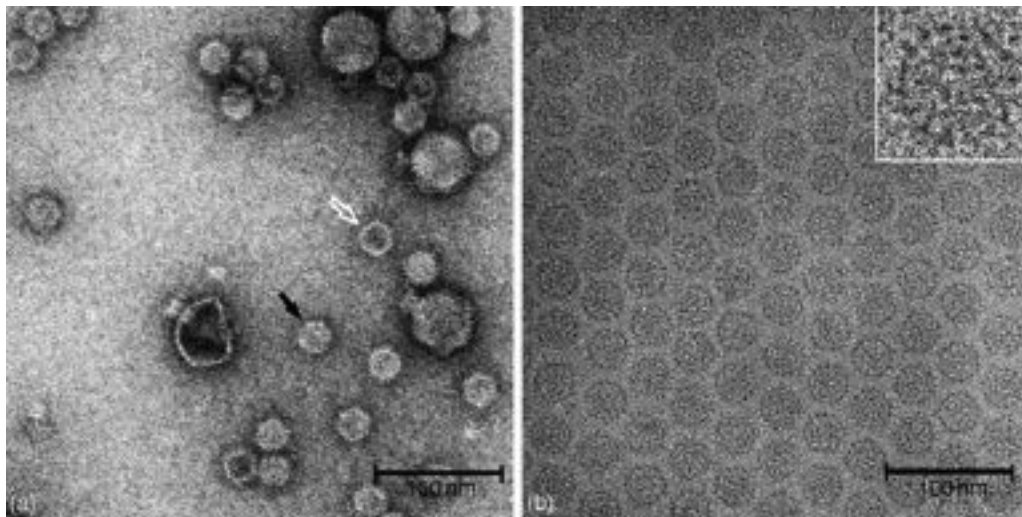


Fig. 1 Electron micrographs of *Penicillium stoloniferum* virus S. Samples were negatively stained in 2% uranyl acetate (a) or prepared unstained and vitrified (b). Micrographs were recorded on a CCD detector in an FEI Polara transmission electron microscope operated at 200 keV with samples at (a) room or (b) liquid nitrogen temperatures. In (a) bacteriophage P22 (five largest particles) was mixed with PsV-S to serve as a calibration reference. Heavy metal stain surrounds virions (e.g., black arrow) and contrasts their surfaces against the background carbon support film. Stain penetrates into the interior of 'empty' capsids (e.g., white arrow), resulting in particle images in which only a thin, annular shell of stain-excluding material (capsid) is seen. In (b) the unstained PsV-S sample was vitrified in liquid ethane. Here particles appear dark (higher density) against a lighter background of surrounding water (lower density). The inset shows a three-times enlarged view of an individual particle, in which several knobby surface features are clearly visible. Reproduced from Ghabrial, S.A, Ochoa, W.F, Baker, T.S, Nibert, M.L., 2008. Partitiviruses: General features. In: Mahy, B.W.J., van Regenmortel, M.H.V (Eds.), *Encyclopedia of Virology*, third ed. Oxford: Elsevier, pp. 68–75. with permission from Elsevier.

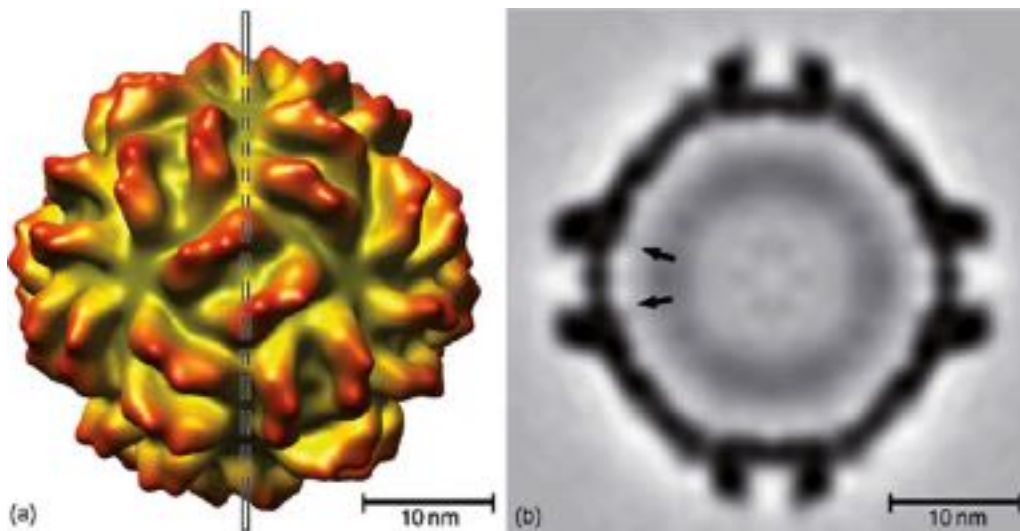


Fig. 2 Three-dimensional (3D) structure of PsV-S. (a) Shaded, surface representation of PsV-S 3D reconstruction viewed along an icosahedral twofold axes. The 3D map is color-coded to emphasize the radial extent of different features (yellows and greens highlight features closest to the particle center, and oranges and reds those farthest from the center). A total of 60 prominent protrusions extend radially outward from the capsid surface. Each protrusion exhibits an approximate dyad symmetry, which is consistent with the expectation that the partitivirus capsid consists of 120 capsid protein monomers, organized as 60 asymmetric dimers in a so-called 'T=2' lattice. (b) Density projection image of a central, planar section through the PsV-S 3D reconstruction (from region marked by dashed box in (a)). Darker shades of gray correspond to higher electron densities in the map section and lighter shades represent low-density features such as water outside as well as inside the particles. The capsid shell appears darkest because it contains a closely packed, highly ordered (icosahedral) arrangement of capsid subunits. The genomic dsRNA on the inside appears at lower density, in part because the RNA is not as densely packed and in part because the RNA adopts a less ordered arrangement. The protrusions seen in (a) appear as large 'bumps' in the central section view that decorate the outside of a contiguous, ~2 nm thick, shell. Arrows point to faint density features that appear to form contacts between the inner surface of the protein capsid and the underlying RNA. These contacts occur close to the fivefold axes of the icosahedral shell. Reproduced from Ghabrial, S.A, Ochoa, W.F., Baker, T.S., Nibert, M.L., 2008. Partitiviruses: General features. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Oxford: Elsevier, pp. 68–75. with permission from Elsevier.

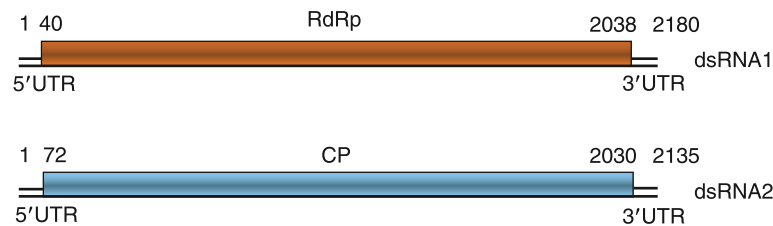


Fig. 3 Atkinsonella hypoxylon virus (AhV), an isolate of the type species of the genus *Betapartitivirus*, has a bipartite genome consisting of dsRNA1 and dsRNA2. dsRNA1 contains the RdRp ORF (nt positions 40–2038) and dsRNA2 codes for the CP ORF (nt positions of 72–2030). The RdRp and CP ORFs are represented by rectangular boxes. Reproduced from Ghabrial, S.A., Buck, K.W., Hillman, B.I., Milne, R.G., 2005. *Partitiviridae*. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, San Diego, CA: Elsevier Academic Press, pp. 581–590. with permission from Elsevier.

Table 3 Size ranges of individual genome segments in classified members of genera *Alpha-* *Beta-* and *Gammapartitivirus*, and the predicted sizes of their RdRP and capsid proteins

Genus	dsRNA1 (bp)	dsRNA2 (bp)	RdRP		CP	
			aa	Mr (kDA)	aa	Mr (kDA)
<i>Alphapartitivirus</i>	1873–2027	1708–1866	581–621	68–73	463–521	51–57
<i>Betapartitivirus</i>	2180–2444	2135–2354	663–746	77–87	636–686	71–77
<i>Gammapartitivirus</i>	1645–1787	1445–1611	519–539	60–62	413–443	44–47

using the negative strand of the genome dsRNA as a template by a semi-conservative mechanism. The newly synthesized positive-strand RNA is retained inside the particle as part of the dsRNA while the parental genomic positive strand (encoding RdRP or CP) is released from the particle and used in translation or in packaging into a new partitivirus particle. The monocistronic positive-sense transcripts of both genomic dsRNAs are translated by cellular ribosomes and virions accumulate in the cytoplasm. The replicase activity of the RdRP is believed to catalyze the synthesis of negative-strand RNA on the positive-strand template within the assembling virion, thereby reconstituting a genomic dsRNA segment.

Epidemiology

Fungal partitiviruses are transmitted intracellularly during cell division, hyphal anastomosis (i.e., cell fusion between fungal strains), and sporogenesis. Vertical virus transmission may occur by sexual spores (basidiospores or rarely ascospores) derived from virus-infected fruiting bodies as well as vegetative spores (conidia). In ascomycetous fungi, such as *Penicillium stoloniferum*, *Gremmeniella abietina*, and *Ustilaginoidea virens*, transmission of partitiviruses through asexual spores is highly efficient, and 90%–100% of single conidia receive these viruses. Only part of basidiomycetous fungi produce asexual spores, and the proportion of virus-infected conidia seems to vary even among isolates of the same host species. In different isolates of the basidiomycete *Heterobasidion parviporum*, two different unclassified alphapartitiviruses were found to be transmitted to 3%–55% of germinated conidia. The basidiospores of this fungus also carry dsRNA elements consistent with partitiviruses, which have been detected in 10%–84% of basidiospores derived from virus-infected fruiting bodies. In many ascomycetes (e.g., *Gaeumannomyces graminis*, *Rosellinia necatrix*, and *Gremmeniella abietina*), partitiviruses are usually eliminated during ascospore formation.

There are no known natural vectors for partitiviruses. Therefore, the horizontal transmission of fungal partitiviruses is regulated by host systems governing hyphal fusion, i.e., somatic (vegetative) incompatibility, mating type incompatibility and intersterility. Inoculation experiments in field conditions have revealed that the unclassified alphapartitivirus *Heterobasidion partitivirus* 4 (HetPV4) (GenBank HQ541325; JX271788-9) is transmitted between conspecific host isolates by two different means: (1) mating between an introduced heterokaryotic virus donor strain and a resident homokaryotic *Heterobasidion* strain, and (2) mycelial contact between two somatically incompatible heterokaryons. In *Rosellinia necatrix*, the transmission of partitiviruses between somatically incompatible isolates of the host fungus can be enhanced by the addition of zinc ions, which promotes hyphal fusion and retards the cell death reaction during an incompatible anastomosis. On the other hand, horizontal transmission of *Rosellinia necatrix* partitivirus 1 (RnPV1) is hindered by the presence of a mycoreovirus, MyRV3, of *Rosellinia necatrix* in the recipient fungal colony. Therefore, pre-existing viral infections may affect the transmissibility of partitiviruses between host isolates.

Partitiviruses are occasionally transmitted across species borders in both laboratory and field conditions. Dual fungal cultures have been used for the transmission of *Heterobasidion* partitiviruses between congeneric host species. Microscopic examination has revealed occasional anastomosing cells between two intersterile *Heterobasidion* species, *H. ecrustosum* and *H. abietinum*, which resulted in subsequent death of the cells involved, but allowed for between-species transmission of the alphapartitivirus HetPV3

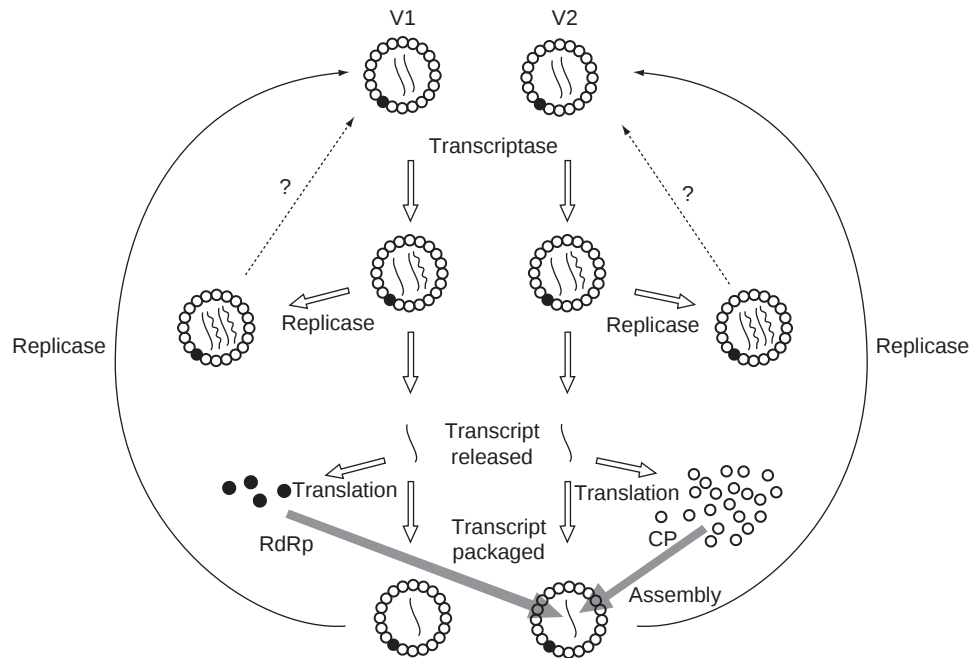


Fig. 4 Model for the replication strategy of *Penicillium stoloniferum* virus S. The open circles represent capsid protein (CP) subunits and the closed circles represent RNA-dependent RNA polymerase (RdRP) subunits. Solid lines represent parental RNA strands whereas wavy lines represent newly synthesized progeny RNA strands. Virion preparations include mature virions with either dsRNA1 or dsRNA2, but also a small proportion of particles that contain only one ssRNA molecule corresponding to the genomic positive strand of the respective dsRNAs, and a heterogeneous population of particles more dense than the mature virions. These heavy particles are believed to represent different stages in the replication cycle, including particles containing the individual genomic dsRNAs with ssRNA tails of varying lengths, particles with one molecule of dsRNA and one molecule of its ssRNA transcript, and particles containing two molecules of dsRNA. Reproduced from Ghabrial, S.A., Hillman, B.I., 1999. Partitiviruses-fungal (*Partitiviridae*). In: Granoff, A., Webster, R.G. (Eds.), Encyclopedia of Virology, second ed. San Diego: Academic Press, pp. 1477–1151. with permission from Elsevier.

(Table 1). Partitiviruses have also been introduced to more distantly related host species by transfecting fungal protoplasts with purified virions. This has been accomplished in *C. parasitica* using partitiviruses of *Rosellinia necatrix* (RnPV1 and RnPV2; Tables 1 and 2), and in *Botrytis cinerea* with SsPV1. In addition, protoplast transfection has been used to transmit viruses between con-specific strains in *B. cinerea* and *Aspergillus fumigatus*.

Sequence analysis has revealed nearly identical alphapartitivirus strains (HetPV11-au1 and HetPV11-pa1; GenBank HQ541328–9 and MG948857–8) in two intersterile species of *Heterobasidion* (*H. australe* and *H. parviporum*) within the same region in Bhutan, suggesting natural transmission of HetPV11 between these fungal species. Conspecific betapartitiviruses have also been described in the ascomycetous blue-stain fungi *Ceratocystis resinifera* (Table 1) and *C. polonica* (GenBank AY247204–5). Similar partitiviruses sharing over 90% RdRp aa sequence identity (the threshold for species demarcation in the family *Partitiviridae*) are sometimes found in divergent host species. For example, gammapartitiviruses found in the ascomycetes *Colletotrichum truncatum* (a plant pathogen) and *Sodiomyces alkalinus* (an alkalophile) seem to be conspecific (GenBank KR074421 and KY484539–40), and alphapartitiviruses infecting the basidiomycete fungi *Heterobasidion parviporum* (order Russulales) and *Megacollybia platyphylla* (order Agaricales) are likewise closely related (JN606085; KI733076).

The genera *Alphapartitivirus* and *Betapartitivirus* include viruses isolated from both fungi and plants. Therefore, a capacity for occasional successful transmission of these viruses between fungal and plant hosts appears likely. Some of the viruses in these genera are found in plant pathogenic fungi, which suggests a potential route for horizontal transmission between fungi and plants. Notably, highly similar partitiviruses have been detected in sugar beet (*Beta vulgaris*) and a basidiomycete fungus infecting its leaves, *Helicobasidium purpureum*. Thus, the partial polymerase sequence of *Helicobasidium purpureum* partitivirus (GenBank AY949837) is 99% identical to that reported for Beet cryptic virus 1, suggesting that the virus may have been transmitted between the associated organisms. This view is further supported by an in vitro transfection study showing that *Penicillium aurantiogriseum* partitivirus 1 (GenBank KT601103–4) is capable of replicating in the *Nicotiana tabacum* BY2 cell line.

Pathogenesis

Partitiviruses are found in all parts of an infected mycelium and can accumulate to very high concentrations (at least 1 mg of virions per one gram of mycelial tissue for *Penicillium stoloniferum* virus F). The transmission of betapartitivirus RnPV1 within

and between *R. necatrix* colonies has been visualized by a colony-print immunoassay, which revealed that the virus was distributed throughout host colonies. However, growing hyphal tips may remain virus-free for a period of time, and then single hyphal tip isolation can be used for curing fungal isolates of partitiviruses. In nature, clonally spreading mycelia of long-living root rot pathogens (such as *Heterobasidion* spp. and *Helicobasidium mompa*) typically consist of virus-infected and virus-free hyphal sections.

Mixed infections of fungal hosts with two or more different partitiviruses have been observed in several species. The gammamapartitiviruses PsV-S and PsV-F were originally found in a single isolate of *Penicillium stoloniferum*. Stable partitivirus co-infections are also found in strains of *Heterobasidion* spp. that may host two different alphapartitiviruses (HetPV16 and HetPV20; GenBank KY859977 and MG566085–6) or alpha- and betapartitiviruses (HetPV9 and HetPV7; GenBank JN606085 and JN606090–1). Co-infections by alpha- and betapartitiviruses occur also in *Helicobasidium mompa* (GenBank AB110979 and AB110980), and mixed infections by two different alphapartitiviruses are found in *Rhizoctonia solani* (KX914900–3). The most striking example of mixed partitivirus infections has been described in single isolates of mycorrhizal *Ceratobasidium* fungi that may host at least five and up to ten different alpha- and betapartitiviruses (GenBank KU291902–22). These fungi occur as root symbionts of *Pterostylis sanguinea* orchids in Australia.

Mixed infections of fungal isolates with partitiviruses and unrelated viruses are also relatively common. The *Rosellinia necatrix* strain hosting RnPV1 is naturally co-infected with *Rosellinia necatrix* megabirnavirus 2 (RnMBV2), and a single strain of the conifer pathogen *Gremmeniella abietina* has been shown to harbor three unrelated viruses (the gammapartitivirus *Gremmeniella abietina* RNA virus MS2, a mitovirus and a totivirus). Similarly, partitiviruses infecting strains of *Heterobasidion* spp. may co-occur with mitoviruses or isolates of the unclassified species *Heterobasidion* RNA virus 6, and *Botryosphaeria dothidea* partitivirus 1 (GenBank KF688740–1) occurs in a mixed infection with a chrysovirus. A mixed infection by a betapartitivirus (GenBank KY296404–5), gammapartitivirus (KY484539–40), and a fusarivirus was described in an isolate of *Sodiomyces alkalinus*.

In general, fungal partitiviruses seem to be associated with symptomless infections of their hosts. However, some of them have deleterious effects on host cells. The unclassified alphapartitivirus *Rhizoctonia solani* partitivirus 2 (Table 2) causes hypovirulence in *Rhizoctonia solani*, a soil-borne basidiomycete fungus. Similarly, the unclassified betapartitivirus SsPV1 (Table 2) has been shown to reduce the virulence of its natural host as well as *Botrytis cinerea* after protoplast transfection. Moreover, *Aspergillus fumigatus* partitivirus 1 (Table 2) seems to mediate reduced growth rate, conidiation and pigmentation in *Aspergillus fumigatus*, and the alphapartitivirus HetPV13-an1 (Table 1) causes drastic growth reduction in several strains of *Heterobasidion annosum* and *H. parviporum*. In contrast, the gammapartitivirus *Talaromyces marneffe* partitivirus 1 (Table 2) seems to enhance the virulence of its native host, which causes opportunistic infections in mammals, including immunosuppressed humans. In some cases, host symptoms may only develop during viral co-infection: a mixed infection by RnPV1 and RnMBV2 leads to host hypovirulence, while the individual viruses exhibit asymptomatic infections.

Taxonomic and Phylogenetic Considerations

Phylogenetic analyses of the predicted amino acid sequences of partitivirus RdRPs show that the classified fungal partitiviruses are included in three distinct clusters corresponding to the genera *Alphapartitivirus*, *Betapartitivirus*, and *Gammapartitivirus* (Fig. 5). However, certain unclassified partitiviruses seem to form a lineage distinct from established partitivirus genera. Based on sequence identity scores and phylogenetic analysis, this virus clade includes *Botryosphaeria dothidea* partitivirus 1, *Alternaria alternata* partitivirus 1 and *Aspergillus fumigatus* partitivirus 2 (Table 2), which are only distantly related to members of genus *Gammapartitivirus* (Fig. 5).

The discovery of a gammapartitivirus in the oomycete *Pythium nunn* in 2018 (GenBank accession LC371062 and LC371063; Fig. 5) expands the host range of partitiviruses to the phylum Stramenopila, which is evolutionarily distinct from fungi and includes organisms with flagellated zoospores ranging from plant pathogenic oomycetes to macroscopic brown algae. It should also be noted that virus sequences resembling known partitiviruses have been discovered in various insect species with high throughput sequencing. Whether these viruses are hosted by the insects or associated microbes remains to be investigated. The putative RdRP of the unclassified *Penicillium aurantiogriseum* partiti-like virus (GenBank accession KT601105) seems to be related to these insect-derived partiti-like sequences, but its taxonomical status remains to be determined, and it seems to form a lineage separate from the classified partitivirus genera.

The evolutionary rate of partitivirus CPs is considerably higher than that of the RdRPs, and pairwise amino acid sequence identities between members of different partitivirus genera are typically less than 10%. Therefore, phylogenetic analyses based on the CP sequences show weaker statistical support than analysis of RdRP sequences. Dendrograms separately constructed from either RdRP or CP sequences seem to reveal essentially the same grouping, but are not entirely congruent. In some cases, the RdRP and CP sequences of a single virus strain show slightly different phylogenetic grouping, suggesting that segment reassortment may occasionally occur among congeneric virus strains in the family *Partitiviridae*. The *Heterobasidion* betapartitivirus HetPV8 and alphapartitivirus HetPV20 (Tables 1 and 2) are examples of such cases.

Based on genome organization and sequence similarity, members of the family *Partitiviridae* are related to a separate and currently unclassified group of fungal dsRNA viruses with two small genome segments (each 1.7–2.4 kbp). These include the mutualistic *Curvularia* thermal tolerance virus, at least five other viruses with published genome sequences (Table 4), and several isolates with unpublished and/or incomplete genome sequences from the fungal genera *Cryphonectria*, *Gremmeniella*, *Lactarius*, *Myriodontium*, *Rhizoctonia*, *Sclerotium*, and *Trichoderma*. The larger genome segment of these viruses has a single ORF encoding an

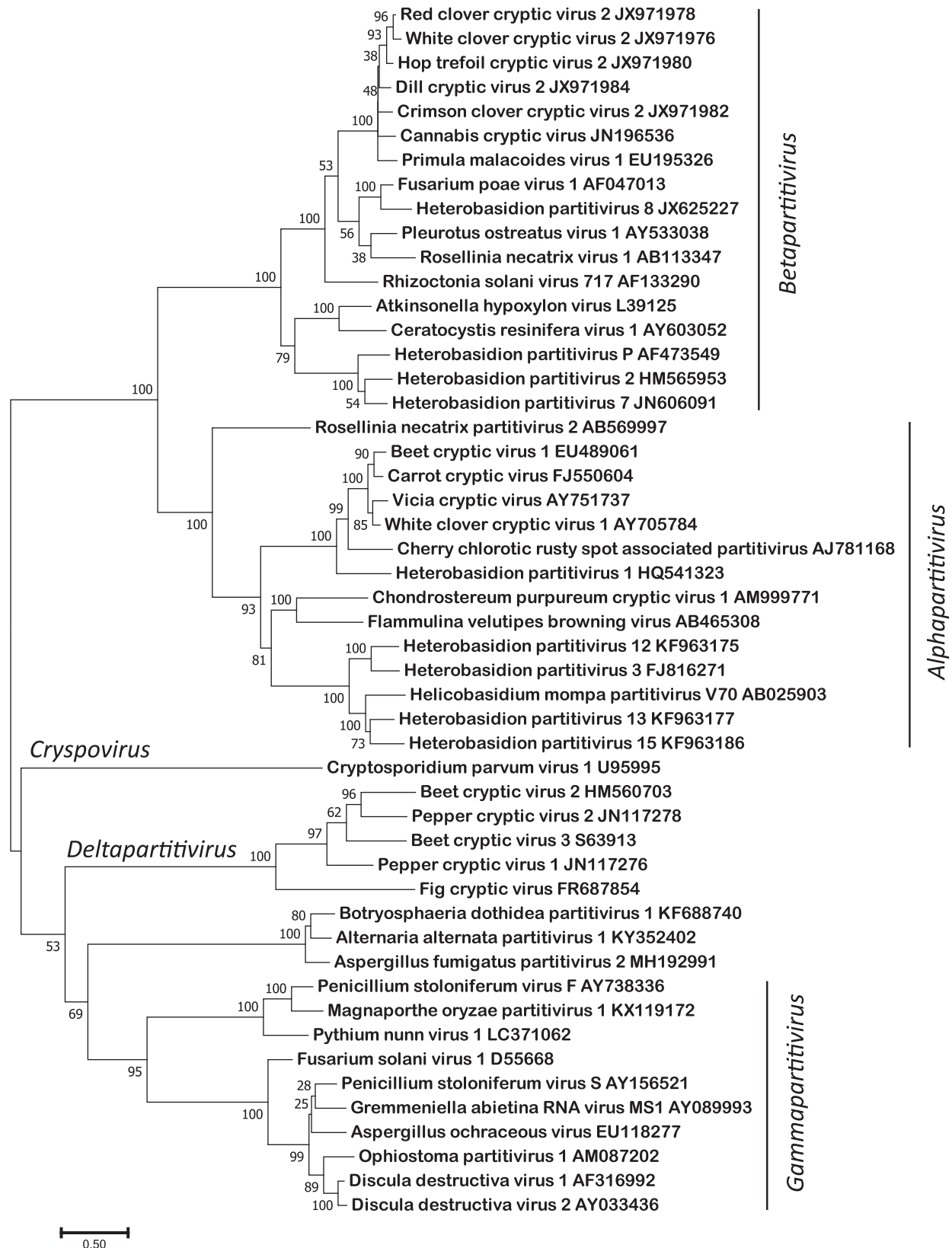


Fig. 5 Unrooted Maximum Likelihood dendrogram constructed based on the complete deduced amino acid sequences of RdRPs of approved members of the family *Partitiviridae* and selected unclassified partitiviruses. The sequences were aligned using MAFFT v7.388 in Geneious R10 (Biomatters Ltd.) and the evolutionary history was inferred using MEGA 7. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (100 replicates) are shown next to the branches. The evolutionary distances were computed using the Le Gascuel 2008 model (LG + G + I) and are in the units of the number of amino acid substitutions per site. Virus names and GenBank accession numbers corresponding to RdRP nucleotide sequences are given.

Table 4 Exemplar isolates of unclassified viruses forming a separate phylogenetic clade related to members of virus family *Partitiviridae*, including the mutualistic *Curvularia* thermal tolerance virus

<i>Virus isolate</i>	<i>GenBank accession</i>	
	<i>dsRNA1</i>	<i>dsRNA2</i>
<i>Curvularia</i> Thermal Tolerance Virus	EF120984	EF120985
<i>Penicillium aurantiogriseum</i> bipartite virus 1	KT601101	KT601102
<i>Fusarium graminearum</i> dsRNA mycovirus 4	GQ140627	GQ140628
<i>Rhizoctonia solani</i> dsRNA virus 1	JX976612	JX976613
<i>Sclerotium hydrophilum</i> virus 1	KU886558	KU886559
<i>Heterobasidion</i> RNA virus 6	HQ189459	MK468678

RdRP of 592–692 aa, whereas the smaller segment has 1–2 ORFs of unknown function. Members of a recently described group of viruses provisionally designated as “Unirnaviruses” also share moderate levels of RdRP homology with partitiviruses, but their genomes are nonsegmented. Viruses with quadripartite genomes currently classified in the family *Chrysoviridae* were formerly included in the family *Partitiviridae*, but a separate family was formed in 2002 to accommodate these viruses as described elsewhere in this encyclopedia.

Members of the family *Megabirnaviridae* and genus *Botybirnavirus* have bisegmented genomes and infect fungi, but their genomes are substantially larger than those of partitiviruses (8.9 kbp and 7.2 kbp for the exemplar strain RnMBV1; 6.2 kbp and 5.9 kbp for *Botrytis porri* RNA virus 1, the only classified member of the genus *Botybirnavirus*). The larger genome segment of megabirnaviruses and botybirnaviruses encodes a CP and RdRP. These virus families are discussed in detail elsewhere in this encyclopedia. Finally, members of the family *Partitiviridae* have properties similar to members of the family *Picobirnaviridae*, e.g., the genomes are bisegmented, and the capsids are small (<45 nm in diameter) with 120 subunits organized as 60 dimers in “T=2” symmetry. The picobirnaviruses, however, are phylogenetically distinct and probably have other basic differences including co-packaging of both genome segments and capacity for extracellular transmission. Picobirnaviruses are found in vertebrates rather than plants, fungi and protozoa.

Further Reading

- Bao, X., Roossinck, M.J., 2013. Multiplexed interactions: Viruses of endophytic fungi. *Advances in Virus Research* 86, 37–58.
- Buck, K.W., Kempson-Jones, G.F., 1973. Biophysical properties of *Penicillium stoloniferum* virus S. *Journal of General Virology* 18, 223–235.
- Ghabrial, S.A., Ochoa, W., Baker, T.B., Nibert, M., 2008. Partitiviruses: General features. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Oxford: Elsevier, pp. 68–75.
- Kondo, H., Kanematsu, S., Suzuki, N., 2013. Viruses of the white root rot fungus, *Rosellinia necatrix*. *Advances in Virus Research* 86, 177–214.
- Marquez, L.M., Redman, R.S., Rodriguez, R.J., Roossinck, M.J., 2007. A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. *Science* 315, 513–515.
- Nerva, L., Silvestri, A., Ciuffo, M., et al., 2017. Transmission of *Penicillium aurantiogriseum* partiti-like virus 1 to a new fungal host (*Cryphonectria parasitica*) confers higher resistance to salinity and reveals adaptive genomic changes. *Environmental Microbiology* 19, 4480–4492.
- Nibert, M.L., Ghabrial, S.A., Maiss, E., et al., 2014. Taxonomic reorganization of family *Partitiviridae* and other recent progress in partitivirus research. *Virus Research* 188, 128–141.
- Nibert, M.L., Tang, J., Xie, J., et al., 2013. 3D structures of fungal partitiviruses. *Advances in Virus Research* 86, 59–85.
- Oh, C.S., Hillman, B.I., 1995. Genome organization of a partitivirus from the filamentous ascomycete *Atkinsonella hypoxylon*. *Journal of General Virology* 76, 1461–1470.
- Sasaki, A., Kanematsu, S., Onoue, M., Oyama, Y., Yoshida, K., 2006. Infection of *Rosellinia necatrix* with purified viral particles of a member of *Partitiviridae* (RnPV1-W8). *Archives of Virology* 151, 697–707.
- Shi, M., Lin, X.D., Tian, J.H., et al., 2016. Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543.
- Tavantzis, S., 2008. Partitiviruses of fungi. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Oxford: Elsevier, pp. 63–68.
- Vainio, E.J., Hantula, J., 2016. Taxonomy, biogeography and importance of *Heterobasidion* viruses. *Virus Research* 219, 2–10.
- Vainio, E.J., Chiba, S., Ghabrial, S.A., et al., 2018. ICTV virus taxonomy profile: Partitiviridae. *Journal of General Virology* 99, 17–18.
- Xiao, X., Cheng, J., Tang, J., et al., 2014. A novel partitivirus that confers hypovirulence on plant pathogenic fungi. *Journal of Virology* 88, 10120–10133.

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsrna-viruses/w/partitiviridae
Partitiviridae - dsRNA Viruses
ICTV.
- https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/fungal-official/4815/
Proposal - ICTV.

Fusariviruses (Unassigned)

Sotaro Chiba, Nagoya University, Nagoya, Japan

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Hypovirulence The reduced pathogenicity of fungal pathogens, in terms of abilities to infect, grow, colonize, and destruct the host organisms, and is generally associated with mycovirus infections.

Sub-genomic RNA An RNA segment shorter than the full-length genome of an RNA virus that expresses protein encoded by downstream ORFs.

Introduction

Mycoviruses (fungal viruses) are increasingly reported with an aid of improved sequencing technology and this has uncovered the highly diverse nature of mycoviruses. Consequently, in the past decade, several new fungal viral taxa, such as the families *Megabirnaviridae*, *Quadriviridae*, and *Mymonaviridae*, have been established. Many reported viruses have still not been assigned to existing viral taxa, even though it has been a long time since the initial discovery and characterization of some of these viruses. One good example is the *Fusarium graminearum* virus 1 strain DK21 (FgV1-DK21), which was identified in Korea in 2002 and thoroughly characterized in 2007. This virus is still taxonomically floating and had been recorded as an unassigned or unclassified virus in the 9th and 10th editions of the International Committee on Taxonomy of Viruses (ICTV) report, published in 2011 and 2018–2019. The related article in the current edition of ICTV report can also be accessed online (see “Relevant Website section”).

Since 2014, viruses that are evolutionarily associated with FgV1-DK21 have been continuously reported, although the genomic structures of those viruses are not identical to that of FgV1-DK21. Therefore, the need for grouping these viruses becomes obvious today. With this in mind, Japanese researchers previously proposed the provisional new genus, *Fusarivirus*, in the new family, *Fusariviridae*, in 2014. The proposed genus/family name was created after the host name of FgV1-DK21, *Fusarium graminearum* (reclassified as *F. boothii*), and thereafter, the terms became widely used to classify the potential members of the group. However, as of 2019, the creation of this new taxon is still incomplete. In this article, the known information on fusariviruses reported so far is briefly summarized, with a view to establish the *Fusarivirus* genus under the family *Fusariviridae*. The future direction of research on fusariviruses is also discussed below.

Taxonomy and Classification

Potential members of the provisional genus *Fusarivirus* are listed in [Table 1](#). From this point onwards, viruses that are members of this genus are described as fusariviruses. The representative species in the provisional genus is *Fusarium graminearum* virus 1, which was first discovered in 2002. A phylogenetic analysis based on the amino acid sequences of the RNA dependent RNA polymerase (RdRp) domain in replication-associated proteins (motifs 1–8) resulted in two major clades: group I, which accommodates FgV1-DK21 and typical fusariviruses with two clearly separated subgroups in the clade, and groups II and III, which are in the same clade that is formed independently of group I (groups II and III are differentiated by their genomic structures) ([Fig. 1](#)). These three groups have the potential to establish independent genera. *Rhizoctonia solani* fusarivirus 3 (RsFV3) with a unique monocistronic genome structure is phylogenetically distant from groups I, II, and III and is thus separately categorized into group IV. Likewise, *Agaricus bisporus* virus 10 and 11 (AbV10 and AbV11) are more distantly related to these four groups and are further categorized as group I/II-like viruses in this article owing to the underlying similarity in their genomic structures and sizes.

Currently, there are no criteria for fusarivirus species and genus demarcations. Amino acid identities of fusariviral replication-associated proteins are not greater than 60%, except for *Alternaria alternata* fusarivirus 1 (AaFV1) and *Alternaria brassicicola* fusarivirus 1 (AbFV1), which share identities as high as 86%. Therefore, species demarcation might possess an amino acid percent identity of ≤ 60 for the replication-associated protein that carry the helicase (Hel) and RdRp domains.

Virion Property

None of the currently reported fusariviruses have been shown to form virions. Fusariviruses are considered not to form rigid particles, but to be the so-called nude viruses or capsid-less viruses. This is reasonable because hypoviruses, which are closely related to fusariviruses from an evolutionary standpoint, do not form virions. *Cryphonectria hypovirus* 1 (CHV1), the representative virus in the family *Hypoviridae*, exists in the fungal host cells as a complex of viral replicase and dsRNA (considered a replicative intermediate), which is capsuled in double-layered small lipid vesicles. Such a subcellular structure associated with

Table 1 List of members in the provisional genus *Fusarivirus*

Group	Virus name	Abbreviation	Genome size (nts)	ORF	5'-UTR (nts)	3'-UTR (nts)	Accession
1-a	<i>Fusarium graminearum</i> virus 1	FgV1	6621 ^{a,b}	4	53	46 ^c	AY533037
1-a	<i>Penicillium roqueforti</i> ssRNA mycovirus 1	PrRV1	6002 ^d	2	31	55	KJ817266
1-a	<i>Pleospora typhicola</i> fusarivirus 1	PtFV1	6733 ^a	2	29	408	KT601107
1-a	<i>Rosellinia necatrix</i> fusarivirus 1	RnFV1	6286 ^a	2	17	92 ^f	AB915829
1-a	<i>Sclerotinia alkalini</i> fusarivirus 1	SaFV1	6239 ^{a,b}	2	5	57 ^c	KT983618
1-a	<i>Neurospora discreta</i> fusarivirus 1	NdFV1	6625 ^a	2	4	455	MK279503
1-a	<i>Fusarium poae</i> fusarivirus 1	FpFV1	6379 ^d	2	74	296	LC150611
1-b	<i>Rutstroemia firma</i> fusarivirus 1	RfFV1	6628 ^a	2	88	134	MK279504
1-b	<i>Macrophoma phaseolina</i> single-stranded RNA virus 1	MpRV1	6356 ^a	2	200	101	KP900890
1-b	<i>Neofusicoccum luteum</i> fusarivirus 1	NlFV1	6244 ^{a,b}	2	46	35	KY906213
1-b	<i>Penicillium aurantiogriseum</i> fusarivirus 1	PaFV1	6193 ^d	2	46	70	KT601099
1-b	<i>Zymoseptoria tritici</i> fusarivirus 1	ZtFV1	5969 ^d	2	36	42	MK279506
1-b	<i>Aspergillus ellipticus</i> fusarivirus 1	AeFV1	6253 ^d	2	34	49	MK279500
1-b	<i>Gaeumannomyces tritici</i> fusarivirus 1	GtFV1	6332 ^a	2	48	115	MK279501
1/2-like	<i>Agaricus bisporus</i> virus 10	AbV10	7033 ^{b,d,g}	2	98	108	KY357495
1/2-like	<i>Agaricus bisporus</i> virus 11	AbV11	6981 ^{b,d,g}	2	44	163	KY357496
2	<i>Nigrospora oryzae</i> fusarivirus 1	NoFV1	7004 ^a	3	109	244	KU980909
2	<i>Sclerotinia homoeocarpa</i> fusarivirus 1	ShFV1	7171 ^d	2	114	41	MK279505
2	<i>Sclerotinia sclerotiorum</i> fusarivirus 1	SsFV1	7754 ^{a,b}	4	83	416	KP842791
2	<i>Morchella importuna</i> fusarivirus 1	MiFV1	7823 ^a	3	20	164	MK279502
2	<i>Alternaria alternata</i> fusarivirus 1	AaFV1	6647 ^d	3	83	219	LC209862
2	<i>Alternaria brassicicola</i> fusarivirus 1	AbFV1	6639 ^{a,b}	3	84	219	KT581960
2	<i>Botrytis cinerea</i> fusarivirus 1	BcFV1	8411 ^{a,b}	2	510	364	MG554633
3	<i>Rhizoctonia solani</i> fusarivirus 1	RsFV1	10,776 ^d	4	161	235	MK558257
3	<i>Rhizoctonia solani</i> fusarivirus 2	RsFV2	10,710 ^a	4	111	152	MK558256
4	<i>Rhizoctonia solani</i> fusarivirus 3	RsFV3	5959 ^a	1	474	97	MK558258
—	<i>Gremmeniella</i> fusarivirus 1	GFV1	2228 ^{d,h}	—	—	—	LR031261
—	<i>Rosellinia necatrix</i> fusarivirus 2	RnFV2	3685; 407; 1287 ^{d,h}	—	—	—	LC333742–4

^a3'-poly(A) sequence is present and excluded from nucleotide counting.^bRACE analysis is performed.^cSubgenomic RNA production is expected.^dNo poly(A)-sequence is detected and actual form of 3'-end is unknown.^eIntegrated poly(A)-like sequence at 3'-terminal is included.^fSubgenomic RNA production is denied.^g5'-end is determined but 3'-end is not.^hOnly partial sequences are available.

fusariviral infection has only been reported as virus-like particle (VLP) for FgV1-DK21. Purified VLPs were aggregates of irregular shaped and sized spherical vesicular structures that were unlike rigid particles. This VLP was shown to strongly associate with viral dsRNAs, thus the structures may stand for replication site of FgV1-DK21. Similar characteristics for other fusariviruses have not been demonstrated.

Genome Organization

Although the type of nucleic acid for fusarivirus genomes is not yet experimentally determined, it is expected to be a positive-sense single-stranded RNA [(+)ssRNA], as is the case for hypoviruses and narnaviruses. The form of the 5' end is unknown, but that of the 3' end is polyadenylated, at least for those complete genomes determined by RACE analyses (Table 1). The length of the RNA genome is 5959–10,776 nucleotides (nts), excluding 3'-poly(A), and the 5' and 3' untranslated regions (UTRs) are in ranges from 5 to 510 nts for the 5' UTR and from 35 to 455 nts for the 3' UTR, excluding poly(A) but including integrated poly(A)-like sequences (Table 1). The replication-associated protein is generally encoded by the largest ORF (ORF1) at the 5' proximal portion, and this is followed by one to three small ORFs in the downstream (Fig. 2). Hel and RdRp domains are present in the replication-associated proteins. This typical genome structure was found in fusarivirus groups I and II. However, viruses in group III have a different genome structure that has a larger sized genome with four coding genes, and the first two and last ORFs encode proteins with no similarities to any proteins in current public databases, whereas the replication-associated protein is encoded by the third ORF (Fig. 2). Additional Hel domain is detectable in the first ORF of viral genomes in group III. Finally, a virus in group IV has a monocistronic nature that encodes only the replication-associated protein. Viruses from groups III and IV have been found only from a basidiomycetous fungus *Rhizoctonia solani*, and these expanded the diversity of fusariviruses.

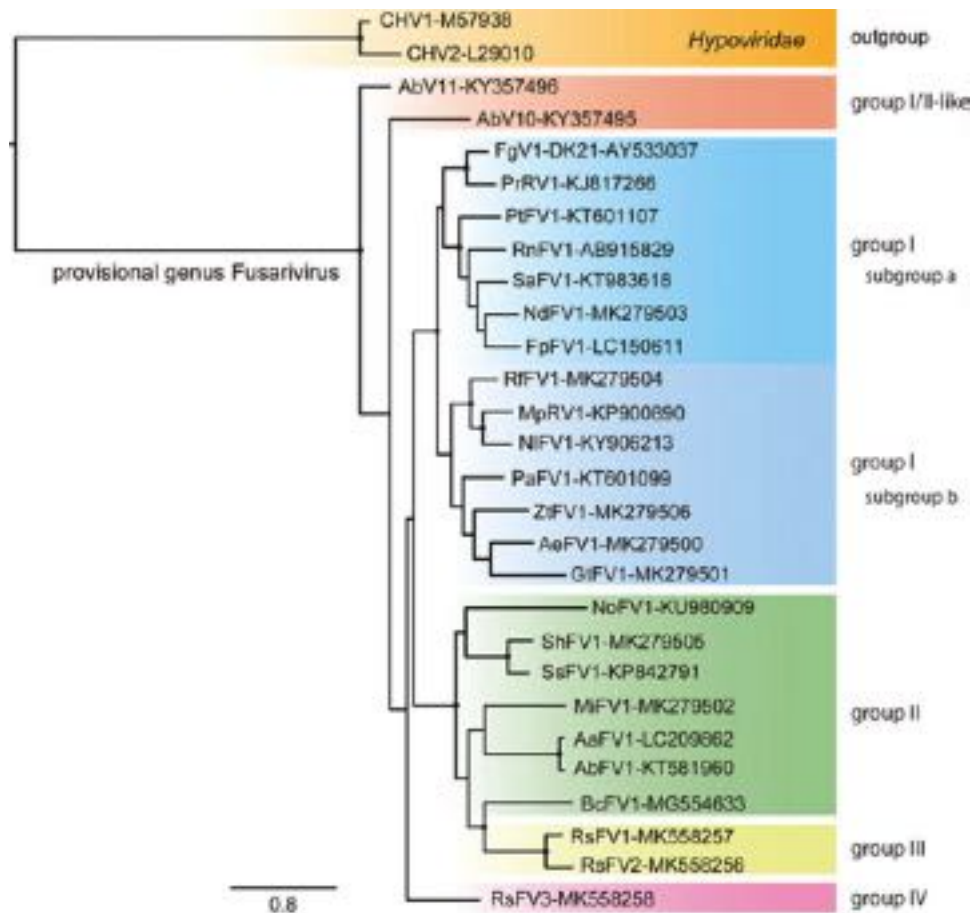


Fig. 1 Maximum-likelihood (ML) phylogenetic tree of fusariviruses. The ML tree was drawn based on multiple amino acid sequence alignment of RdRp domain (motif 1–8) of fusariviral and hypoviral replication-associated proteins. The branches supported by aLRT (over 0.9) are denoted by dots at nodes. Phylogenetic relationship and genomic structure are taken into account for grouping of fusariviruses.

ORF1 and the second largest ORF (ORF 2, 3, or 4) are commonly encoded by members of classic fusariviruses in groups I and II. These secondly largest proteins exhibit low sequence similarities with each other in general, even between proteins in the same groups. It is notable that the structural maintenance of chromosomes (SMC)-like motif or Spc7 kinetochore protein-like motif was detected in the second largest proteins of fusariviruses in group I but were generally not detected for those of the II and I/II-like groups. This suggests that the viruses in group I have common ancestors.

Gene Expression

The gene expression strategy for the 5' proximal ORF of fusariviruses is largely unknown, but there is some information for downstream ORFs. The ORFs 2, 3, and 4 of FgV1-DK21 are expected to be expressed by two sub-genomic RNAs (Fig. 2). A sub-genomic RNA for ORF2 of *Sodiomyces alkalinus* fusarivirus 1 (SaFV1) is suspected, however, the detected fragment in this instance was vague, so it needs verification. Contrary to this, the possibility to synthesize sub-genomic RNA for ORF2 of *Rosellinia necatrix* fusarivirus 1 (RnFV1) was experimentally denied, suggesting poor commonality of gene expression strategies among the different fusariviruses.

Virus Transmissions

FgV1-DK21, BcFV1, RnFV1, and *Sclerotinia sclerotiorum* fusarivirus 1 (SsFV1) are capable of transmission *via* hyphal anastomosis between virus-donor and virus-recipient strains (horizontal transmission), but others are not confirmed. Similarly, transmission through asexual and sexual spores (vertical transmission) has been observed for FgV1-DK21 and SaFV1. A natural transmission vector for fusariviruses has not yet been discovered.

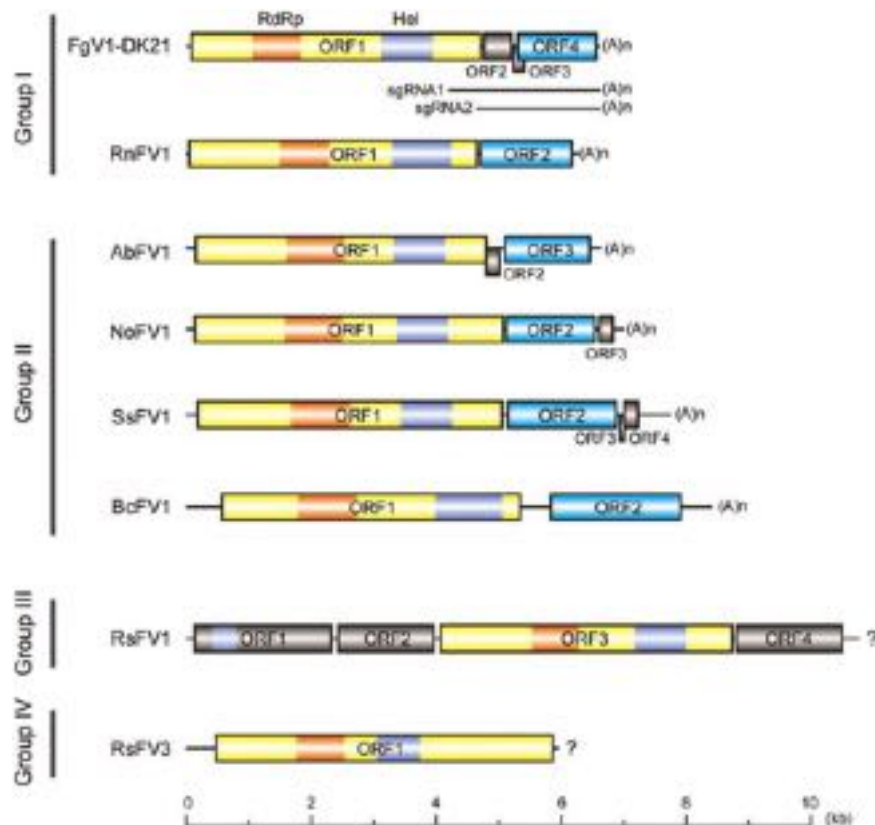


Fig. 2 Genome composition of representative fusariviruses. Genomic structures of representative fusariviruses are schematically shown. Non-coding regions are indicated by lines and coding regions are shown by rectangles. RdRp and Hel domains are illustrated in the protein coding regions.

Virus–Host Interactions

Among all reported fusariviruses, only FgV1-DK21 exhibited an association with host morphological and physiological changes such as impaired mycelial growth, increased pigmentation, reduced mycotoxin production, and attenuated virulence of its natural host, *F. boothii*. FgV1-DK21 also showed hypovirulence-conferring ability to its heterologous hosts *F. asiaticum*, *F. graminearum*, *F. oxysporum*, and *C. parasitica*. RnFV1, AbFV1, SsFV1, and probably BcFV1 have no impact on the biological properties of their host. The effects of infection by other fusariviruses on host fungal properties are unknown.

FgV1-DK21 is targeted by the host anti-viral RNA silencing or RNA interference. It suppresses this defensive machinery of the host through the repression of key genes in RNA silencing such as *FgDICER2* and *FgAGO1* by the function of ORF2 protein (pORF2). This pORF2 function is similar to the mode of an RNA silencing suppressor encoded by CHV1, p29. Moreover, FgV1-DK21 is known to interact with host cellular proteins. For example, hexagonal peroxisome protein (Hex1) facilitates the replication of FgV1-DK21 *via* a direct interaction with viral RNA. Hex1 is a component of Woronin bodies and plays a crucial role in the asexual reproduction and virulence of host fungus. No other virus–host interactions are characterized for other fusariviruses but FgV1-DK21.

Summary and Future Prospective

The behavior of the representative fusarivirus, FgV1-DK21, has been thoroughly investigated, and some questions particularly related to viral propagation and pathogenicity have been uncovered. However, the basic characteristics of fusariviruses as a genus or a family are still unknown. These include the function of genes encoded by fusariviruses, the gene expression strategies commonly taken by fusariviruses, the 5' end form of fusariviral genomes, and the stable form of fusariviruses in host cells. The use of infectious cDNA clone of fusariviruses will be a powerful tool for understanding of above-mentioned characteristics of fusarivirus. Furthermore, since unique fusariviruses such as RnFV1 (group III), RnFV3 (group IV), and AbV10 (group I/II-like) are reported, more information on the viruses that are most closely related to these are required for a clearer understanding of the evolutionary history and proper taxonomical arrangement of fusariviruses.

Further Reading

- Chu, Y.M., Jeon, J.J., Yea, S.J., *et al.*, 2002. Double-stranded RNA mycovirus from *Fusarium graminearum*. *Applied and Environmental Microbiology* 68 (5), 2529–2534.
- Deakin, G., Dobbs, E., Bennett, J.M., *et al.*, 2017. Multiple viral infections in *Agaricus bisporus* – Characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. *Scientific Reports* 7 (1), 2469.
- Gilbert, K.B., Holcomb, E.E., Allscheid, R.L., Carrington, J.C., 2019. Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes. *PLoS One* 14 (7).
- Hrabáková, L., Grum-Grzhimaylo, A.A., Koloniuk, I., *et al.*, 2017. The alkalophilic fungus *Sodiomyces alkalinus* hosts beta-and gammapartitiviruses together with a new fusarivirus. *PLoS One* 12 (11), e0187799.
- Kwon, S.J., Lim, W.S., Park, S.H., Park, M.R., Kim, K.H., 2007. Molecular characterization of a dsRNA mycovirus, *Fusarium graminearum* virus-DK21, which is phylogenetically related to hypoviruses but has a genome organization and gene expression strategy resembling those of plant potex-like viruses. *Molecules and Cells* 23 (3), 304.
- Lee, K.M., Yu, J., Son, M., Lee, Y.W., Kim, K.H., 2011. Transmission of *Fusarium boothii* mycovirus via protoplast fusion causes hypovirulence in other phytopathogenic fungi. *PLoS One* 6 (6), e21629.
- Picarelli, M.A.S., Forgia, M., Rivas, E.B., *et al.*, 2019. Extreme diversity of mycoviruses present in isolates of *Rhizoctonia solani* AG2-2 LP from *Zoysia japonica* from Brazil. *Frontiers in Cellular and Infection Microbiology* 9, 244.
- Son, M., Lee, K.M., Yu, J., *et al.*, 2013. The HEX1 gene of *Fusarium graminearum* is required for fungal asexual reproduction and pathogenesis and for efficient viral RNA accumulation of *Fusarium graminearum* virus 1. *Journal of Virology* 87 (18), 10356–10367.
- Son, M., Choi, H., Kim, K.H., 2016. Specific binding of *Fusarium graminearum* Hex1 protein to untranslated regions of the genomic RNA of *Fusarium graminearum* virus 1 correlates with increased accumulation of both strands of viral RNA. *Virology* 489, 202–211.
- Son, M., Lee, Y., Kim, K.H., 2016. The transcription cofactor Swi6 of the *Fusarium graminearum* is involved in *Fusarium graminearum* virus 1 infection-induced phenotypic alterations. *Plant Pathology Journal* 32 (4), 281.
- Yu, J., Park, J.Y., Heo, J.I., Kim, K.H., 2019. The ORF2 protein of *Fusarium graminearum* virus 1 suppresses the transcription of *FgDICER2* and *FgAGO1* to limit host antiviral defences. *Molecular Plant Pathology*, 230–243.
- Zhang, R., Liu, S., Chiba, S., *et al.*, 2014. A novel single-stranded RNA virus isolated from a phytopathogenic filamentous fungus, *Rosellinia necatrix*, with similarity to hypo-like viruses. *Frontiers in Microbiology* 5, 360.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/

Virus Taxonomy: The Classification and Nomenclature of Viruses.

Giardiavirus (*Totiviridae*)

Juliana Gabriela Silva de Lima, João Paulo Matos Santos Lima, and Daniel Carlos Ferreira Lanza, Federal University of Rio Grande do Norte, Natal, Brazil

© 2021 Elsevier Ltd. All rights reserved.

This is a reproduction of Juliana Gabriela Silva de Lima, João Paulo Matos Santos Lima, Daniel Carlos Ferreira Lanza, 2017, Giardiaviruses, In Reference Module in Life Sciences, Elsevier Inc, doi:10.1016/B978-0-12-809633-8.11003-9.

History

In 1986, a ~6.3 kbp linear double-stranded (ds) RNA molecule was observed in nucleic acid extracts of *Giardia lamblia* Portland I trophozoites obtained by Dr. D. G. Lindmark of Cleveland State University. Further analysis of this dsRNA showed it to be a genome from a small isometric virus that infects this protozoan. The virus found was named *Giardia lamblia virus* (GLV), referring to the host it infects (Wang and Wang, 1986). *Giardia lamblia*, also known as *Giardia intestinalis* or *duodenalis*, is an anaerobic parasitic flagellate that inhabits the gastrointestinal tract of humans as well as many other mammals. When infecting humans, the parasitic protozoan causes giardiasis, an acute diarrhea that often progresses to chronic carrier-stage in adults and severe malnutrition in children. In chronic form, the parasites adhere to the human gut wall and cause losses in host's digestion and absorption of nutrients. In addition to GLV, there is the *Giardia canis virus* (GCV), a ~6.3 kbp dsRNA virus that infects *Giardia canis*, a protozoan that affects dogs and other canids. In 2001, the GCV was isolated from different strains of *G. canis* and identified by Jianjun *et al.* Unlike *G. lamblia* infection in humans, most *Giardia* infections in dogs are asymptomatic. However, some infected dogs may suffer from acute or chronic diarrhea, weight loss or poor weight gain, in spite of having a normal appetite and, less commonly, vomiting and lethargy (Inger *et al.*, 2007). GLV and GCV probably correspond to the same species, as they share more than 94% of identity between nucleotide sequences, in addition to infecting the same giardia species.

Taxonomy, Classification and Evolution

Currently, according to the International Committee on Taxonomy of Viruses (ICTV), the GLV belongs to the *Giardiavirus* genus of the *Totiviridae* family. *Totiviridae* are characterized by having a dsRNA non-segmented genome and virions with simple structures. Two other genera of dsRNA viruses which also infect protozoa, *Leishmanivirus* and *Trichomonasvirus*, are included in this family, in addition to *Totivirus* and *Victorivirus* comprising viruses that infect fungi (King *et al.*, 2012; ICTV, 2015). Some putative *Totiviridae* viruses are known to infect plants such as maize, the herb *Panax notoginseng* and the red alga *Delisea pulchra* (Chen *et al.*, 2016; Guo *et al.*, 2016; Lachnit *et al.*, 2016). Some others were found in some fish species such as salmon and golden shiner baitfish (Wiik-Nielsen *et al.*, 2012; Mor and Phelps, 2012).

When compared with *Saccharomyces cerevisiae virus L* (ScV), one of the best-known viruses of the *Totiviridae* family, GLV and ScV are similar in many ways, namely: (1) single molecules of genomic dsRNAs of similar sizes; (2) single major capsid polypeptides of similar sizes; (3) virion-associated RdRp with similar amino acid sequence motifs deduced from the genomic sequence, (4) ability to synthesize viral messages from intact virions in vitro; and (5) use of ribosomal frameshifting for the synthesis of viral polymerase. However, the two viruses do not cross-infect and the two dsRNAs do not cross-hybridize in Northern blots (Wang and Wang, 1991). Sequence alignments from GLV cDNAs and those from ScV-L, *Leishmania* RNA virus (LRV) and *Trichomonas vaginalis* virus (TVV) indicate that, despite the similar organization of the viral genomes, GLV share very little overall sequence identity with any of these viruses whose genomic sequences have been completely determined. GLV is therefore not closely evolutionarily related to any of these viruses.

The GCV has not yet been given a completely defined classification. Phylogenetic analysis of RdRp gene sequences demonstrate that GCV is clustered with GLV and piscine myocarditis virus (PMCV) a virus that affects fish myocardial cells, forming the GLV-like clade. Within this same family, the closest related group to GLV-like is the IMNV-like clade, composed of some viruses characterized by affecting arthropods (insects and crustaceans) (Oliveira *et al.*, 2014). The IMNV-like clade was recently proposed as a new genus called Artivirus, based on their distinct genomic organization in relation to other members of the *Totiviridae* family (Dantas *et al.*, 2015). The striking similarity in genome organization and sequence identity as well as the more close phylogenetic proximity of their hosts may indicate that GCV and GLV share a common evolutionary origin; this can be further elucidated with more detailed genetic studies. The phylogenetic relationships among representative members of the *Totiviridae* family regarding RdRp amino acid sequences are shown in Fig. 1.

Host Range and Geographic Distribution

Purified GLV readily infects many virus-free isolate of *G. lamblia* trophozoites, but not any other parasitic protozoa tested, including *Tritrichomonas foetus* and *Trichomonas vaginalis* (Sepp *et al.*, 1994). The virus has also been shown to not infect two transformed human

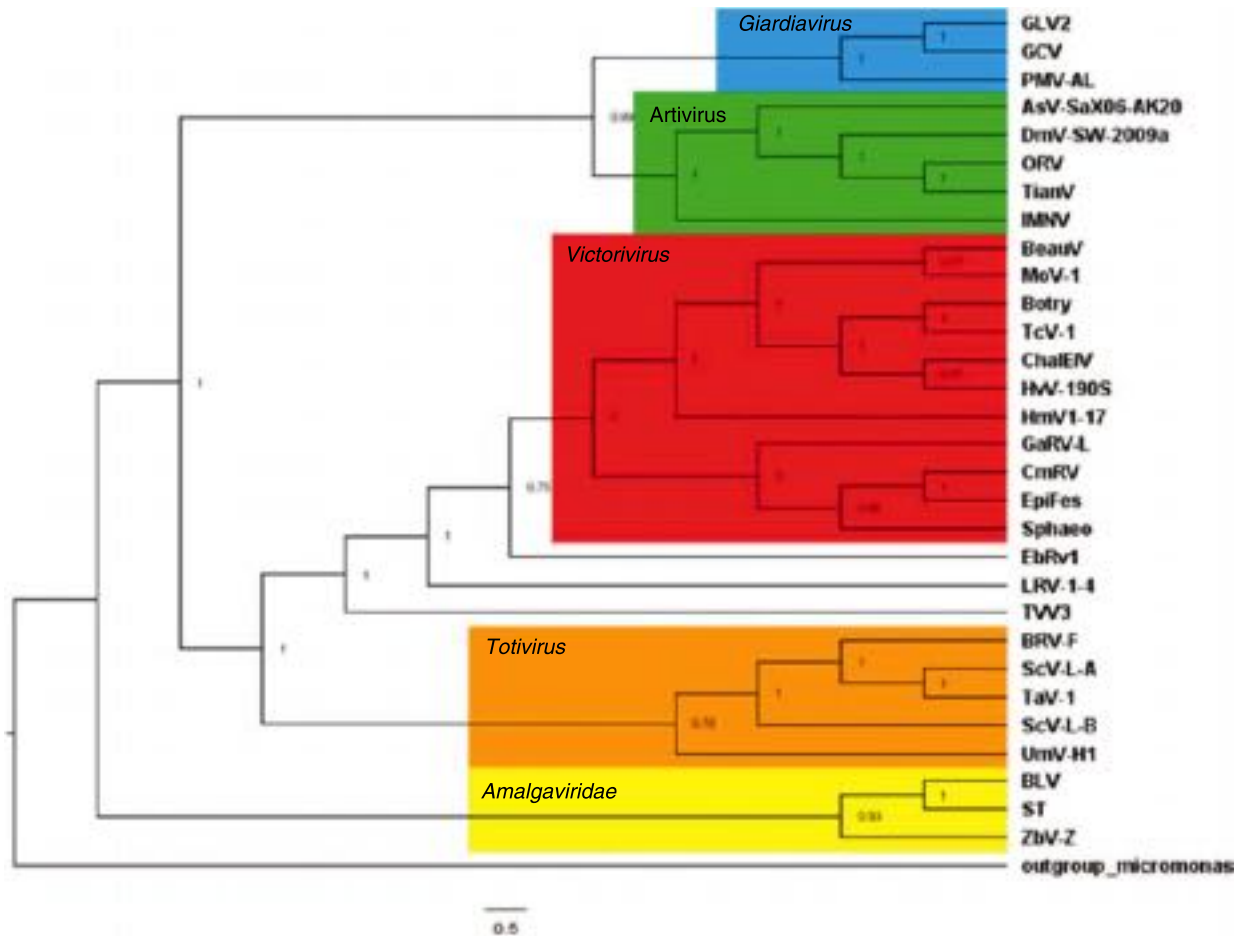


Fig. 1 Phylogeny of the *Totiviridae* family members calculated from an alignment of RdRp aminoacid sequences using Bayesian inference (Oliveira *et al.*, 2014). Numbers in branch nodes indicate posterior probabilities and colors represent the genera in according to ICTV. The artivirus group was not formally considered as a genus of the *Totiviridae* family to date.

intestinal cell lines (Wang *et al.*, 1988). It is therefore believed that GLV has a rather narrow host range; it probably only infects *G. lamblia* in nature. On the other hand, the cellular host for this virus (*G. lamblia*) parasitizes many mammals other than humans. Viruses that are identical in shape and size to genomic dsRNA and share dsRNA sequence homology with one another have been detected from many *G. lamblia* strains and isolates obtained from humans, guinea pigs, cats, beavers, llamas, and sheep. The human isolates were collected from Belgium, Poland, England, Israel, Ecuador, Puerto Rico, and various states in the USA (Miller *et al.*, 1988).

GCV is known to affect *G. canis* trophozoites. Some studies have reported *G. canis* isolates affecting dogs in Italy, Spain, USA, Australia, China, Canada, Germany and Sardinia, and in wildlife animals such as coyotes from California and African painted dogs reviewed in Thompson and Ash (2016). *Giardia* is common in domestic dogs throughout the world, and is often the most common enteric parasite. Since *Giardia* species are found in almost all parts of the world, affecting organisms in developing as well as developed countries, it is expected that GCV and GLV follows their respective hosts and are also distributed worldwide.

Physical and Biochemical Characteristics

From the 3D structures obtained by transmission electron cryomicroscopy (Cryo-TEM), it can be observed that GLV has a "T=2" type considered icosahedral shaped capsid composed of 120 polypeptide subunits called capsid proteins (CP), and arranged in a "T=1" lattice (Fig. 2(A)). Most of the exterior capsid surface forms a mostly flat raised plateau located at the center of each of the icosahedral 5-fold axes. This plateau is composed by 10 CP subunits. According to analysis of the equatorial sections of the virus (Fig. 2(B)), it is suggested that the CP subunits have prevalence of α -helical secondary structures, with each subunit having a maximum length of approximately 100 Å and very similar tertiary structures (Janssen *et al.*, 2015). The capsid has a single layer with a maximum thickness of approximately 60 Å (median ~40 Å) and outer diameter of approximately 485 Å, which makes GLV the virus with the largest capsid among all totiviruses studied to date, including the IMNV capsid (Janssen *et al.*, 2015).

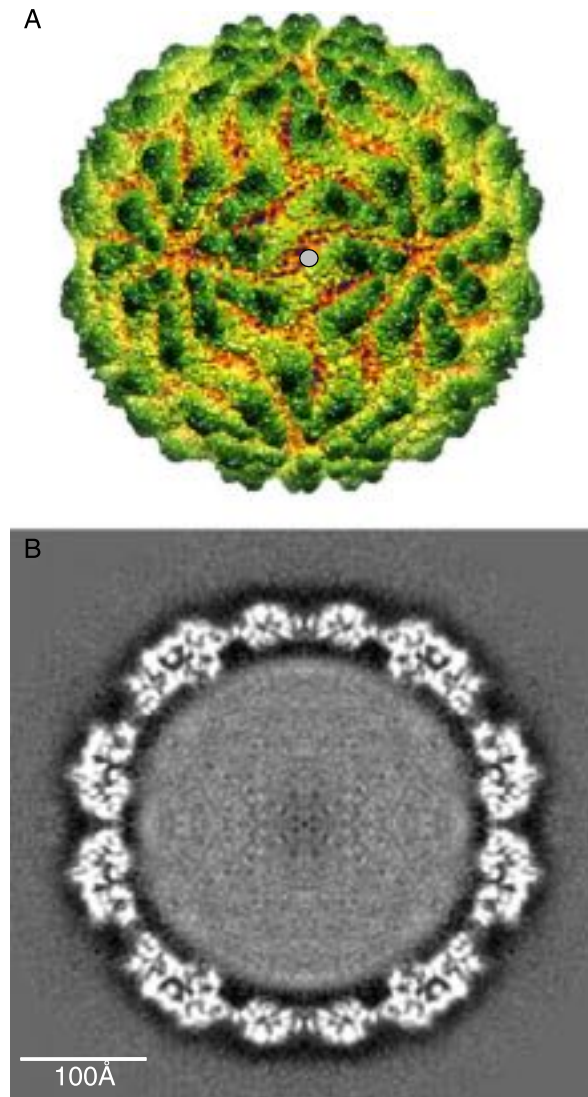


Fig. 2 Three dimensional structure of GLV virions. (A) Space-filling stereo view of the viral particle based on an icosahedral 2-fold axis (gray point). The colors are correspondent to the radius being dark green outermost and red innermost regions. (B) Central thick section through the virion. Black and white corresponding to highest and lowest densities in the density map, respectively. (Janssen *et al.*, 2015; EMDB - EMD-5948).

There are no studies on GCV capsid physical characteristics yet, but according to its high genomic sequence similarity to GLV, it can be assumed that GCV has a capsid with similar structure.

Regarding the GLV genome array within the virus, it is believed that it is homogeneously distributed within the capsid, except for some points of greatest concentration near the shell. Thus, GLV has a different genome distribution when compared with other totiviruses (Janssen *et al.*, 2015). As the GLV dsRNA can be readily radiolabeled by [32 P]pCp and RNA ligase, this molecule must have a free 3'-OH group at one or both of its 3' end. The exact structure at the 5' end of the dsRNA has not yet been elucidated. It is known that the dsRNA molecule cannot be phosphorylated by T4 polynucleotide kinase, and that the denatured dsRNA molecule can be circularized with T4 RNA ligase. Therefore, the 5' terminus of the GLV genomic dsRNA is probably phosphorylated. Additionally, no covalently linked protein is found at this extremity such as in the case of the poliovirus (Garlapati and Wang, 2005; Furfine *et al.*, 1989; Wang and Wang, 1986).

Both GLV and GCV viruses encode two different proteins, with one being a polypeptide of approximately 100 kDa (P100) that gives rise to the CP subunits, and the viral RNA-dependent RNA polymerase (RdRp), consisting of a 190 kDa polypeptide (Wang *et al.*, 1993). GLV purified virions are good antigens in mice. Polyclonal antibodies raised against whole virions react with P100 in Western blots and can effectively block viral infection in the *in vitro* culture of *G. lamblia* trophozoites. The same antiserum also reacts positively with a minor component of the virion, which is the 190 kDa viral RdRp. Indeed, GLV-encoded RdRp activity has been detected in the infected cell extract as well as the purified GLV fractions. Antisera raised against synthetic dsRNA only cross-reacts with the viral dsRNA and are not protective against viral infection of *G. lamblia* trophozoites *in vitro*. Giardiaviruses do not contain any lipids, nor is the P100 polypeptide glycosylated (Wang *et al.*, 1988).

Table 1 Structural properties of some small dsRNA viruses

Virus	Shape	Diameter (nm)	Density (gm/ml ⁻¹)	dsRNA (μm)	dsRNA (kbp)	Capsid protein (kDa)
GLV	Isometric	48.5	1.368	1.50	6.3	98
GCV	Isometric	48.5 ^a			6.3	98
TVV	Isometric	35	1.468	1.50	4.6–5.9	74–82
LRV	Isometric	40			5.3	82
ESV	Isometric	35		1.63	6.2	86.5
ScV-L	Isometric	44	1.368	1.31	4.6	76
UmV-P1	Isometric	41–43	1.418		6.1	73
IMNV	Isometric	45			8.2	99
HvV190S	Isometric	36			5.2	81

^aExpected.

Abbreviations: GLV, *Giardia lamblia* virus; GCV, *Giardia canis* virus; TVV, *Trichomonas vaginalis* virus; LRV, *Leishmania RNA* virus; ESV, *Eimeria stiedae* virus; ScV-L, *Saccharomyces cerevisiae* virus L; UmV-P1, *Ustilago maydis* virus; IMNV, Infectious myonecrosis virus; HvV190S, *Helminthosporium victoriae* virus 190S.

Experiments in which *Trichomonas vaginalis* virus 1 (TVV1) and GLV virions were exposed to different temperatures and pHs showed that the GLV is more thermoresistant than the TVV1 virions. In addition to GLV being able to mediate transcription through a wide pH range, it has maximum transcription efficiency at pHs close to 8.0, while TVV1 virions mediate transcription in a narrower pH range with maximum efficiency at pHs near 6.5. These characteristics correspond to different intracellular conditions encountered in their respective hosts, since *T. vaginalis* trophozoites ideal growth occurs under more acidic conditions (pH < 6.5) than *G. lamblia* trophozoites (pH > 7) (Janssen *et al.*, 2015). The structural properties of the GLV and GCV comparing with other totiviruses are shown in Table 1.

Organization and Molecular Biology of GLV and GCV dsRNA Genomes

The 6277 bp GLV and 6276 bp GCV dsRNA genomes contain only two large open reading frames (ORFs), both on the same RNA strand. The two genomes share nucleotide identity higher than 94%, presenting the most number of single nucleotide variations at ORF1 region. GLV and GCV first ORFs (nts 367–3027 and 367–3030, respectively) encode the precursor polypeptide for the capsid protein. Thirty-two amino acid residues from the N-terminus of this precursor protein are apparently removed by a cellular cysteine protease before the processed capsid protein is assembled into the GLV virion (Yu *et al.*, 1995; Chen *et al.*, 2007; Liu *et al.*, 2008).

The second ORF (nts 2805–5978 and 2808–5981 for GLV and GCV, respectively) is encoded in –1 frame (in relation to ORF 1), and the two ORFs overlap by 220 bp. RdRp motifs have been found in many ORF2 from different totiviruses (Oliveira *et al.*, 2014; Dantas *et al.*, 2015). It is now known that the GLV RNA polymerase (190 kDa) is synthesized as a fusion protein of ORFs 1 and 2 at a level that is 2–5% of P100. Apparently, ribosomes carrying the growing nascent polypeptide chain are stalled when they encounter the pseudoknot structure residing within the 220 nt overlap of these two ORFs. At a 2–5% frequency, the ribosomes slip back one reading frame and proceed to translate ORF 2 as the C-terminus of a fusion protein. The ability of this GLV overlap fragment to induce a –1 ribosomal frameshift has been demonstrated in a reporter system in yeast (Wang *et al.*, 1993). The schemes of GLV and GCV genomes are shown in Fig. 3.

Flanking the 5' and 3' ends of the two ORFs are two untranslated regions. For the GLV, the 5' and 3' UTRs have 367 and 294 nts respectively, and in GCV the corresponding regions have 367 and 298 nts (Wang *et al.*, 1993; Dantas *et al.*, 2015). It was identified that these two regions contain sequence elements that are critical for initiation of transcription and replication of GLV RNA. For example, deletion of a single nucleotide from the 5' terminus of the (+) strand GLV RNA totally abolishes transcription of GLV mRNA. Similarly, deletion or alteration of sequences in these two regions drastically reduces the level of progeny viral RNA (Garlapati *et al.*, 2001; Garlapati and Wang, 2004, 2005).

GLV has been successfully used as a vector for the introduction of foreign genes into *G. lamblia*. When a portion of the GLV genome is replaced with the firefly luciferase gene in a cDNA construct downstream from a T7 promoter, chimeric RNA can be synthesized in vitro using T7 polymerase. *Giardia* cells that are infected with wild-type GLV and electroporated with the chimeric RNA show luciferase activity of a millionfold above background. The chimeric RNA is replicated as dsRNA and packaged into recombinant virions that are shed into the culture supernatant. These recombinant viruses together with wild type GLV can in turn infect naive *Giardia* trophozoites to produce luciferase activity (Garlapati *et al.*, 2001). The list of foreign genes tested in this system include gene encoding neomycin phosphotransferase, hygromycin phosphotransferase, and green fluorescent protein. Moreover, they can be delivered and expressed at high levels (Yu *et al.*, 1996). Inclusion of a nts 368–631 fragment in the coding region of P100, such as the N-terminus of a fusion protein with the foreign gene product has been shown to enhance translation level by 5000-fold (Yee and Nash, 1995). This region therefore may contain elements that can promote interaction between GLV mRNA and ribosomes of *G. lamblia*.

Giardiavirus mRNA is neither capped nor adenylated. Translation is initiated at an internal ribosomal entry site (IRES) in the mRNA consisting of a 253 nt segment in the 5' untranslated region (5' UTR) and a 264 nt stretch in the ORF immediately downstream from the 5' UTR. The initiation codon is localized at the center of an unstructured 31 nt segment flanked by complex

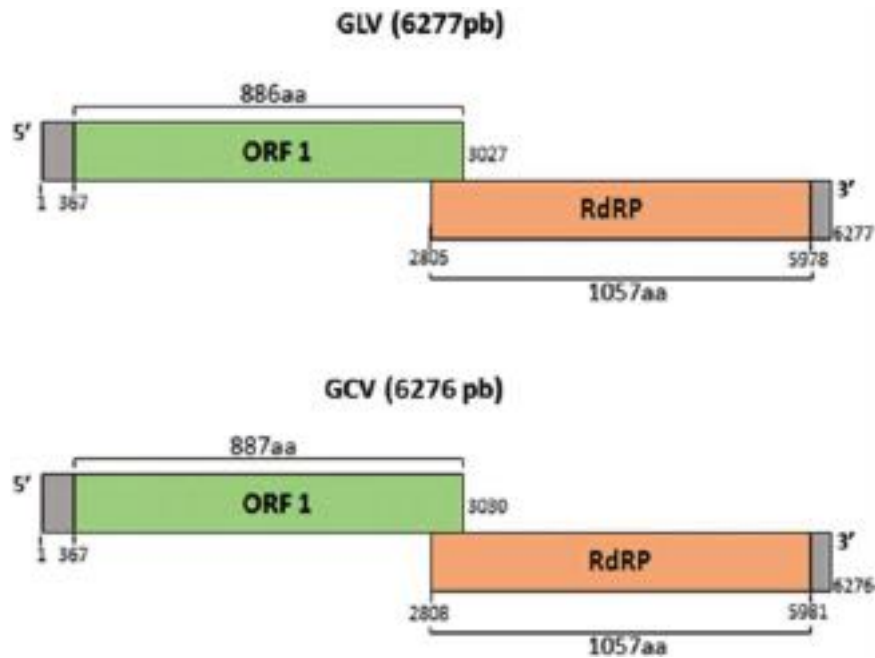


Fig. 3 Organization of the *Giardia lamblia* virus and *Giardia canis* virus dsRNA genomes.

secondary structures in the 5' UTR and the ORF of the IRES. The precise position of the initiation codon is critical for translation initiation. It also suggests a direct recruitment of the 40 S small ribosomal subunit to the initiation codon without ribosomal scanning, a process that is known to be absent from the translation machinery in *Giardia* (Garlapati and Wang, 2005).

Infection and Replication

GLV infects susceptible isolates of *G. lamblia* very efficiently. It has been estimated that demonstrable infection can be achieved at an infection multiplicity as low as 10 GLV particles per cell (Miller *et al.*, 1988). The virus is not observed in *Giardia* cysts, and is not known to occur during the transition between cyst and trophozoite. The exact mode of viral entry into the cell has not yet been elucidated. With infection course, it is observed that the viruses enter into susceptible parasite cells via endocytosis, where the peripheral lysosomal vacuoles serve as virus translocation ports to cytoplasm (Tai *et al.*, 1993). Sepp *et al.* (1994) suggests that this internalization is mediated by endocytosis receptors. However, a number of virus-free *G. lamblia* strains have been found to be resistant to GLV infection. When trophozoites of these strains are electroporated with the single-stranded GLV RNA (ssRNA), they become infected by GLV and can support and complete the replication cycle, producing progeny virions that are fully infective to sensitive strains (Furfine and Wang, 1990). It is expected that these receptors are present on the surface of cells susceptible to the virus, are responsible for binding and internalizing the virus, and are absent in resistant protozoa. GLV is the only known protozoan viruses which can be efficiently transmitted by extracellular pathways, indicating that virion have the necessary protein machinery to efficiently mediate the entry.

Using purified GLV to infect a virus-free *G. lamblia* WB strain, in situ hybridization was used to show that dsRNA replication was first detected in the cytoplasm. Toward the late stage of infection, viral dsRNA was found to spread into the twin nuclei of the infected flagellate. Transmission electron microscopic examination of thin sections of infected cells at stationary phase also reveals paracrystals of virus-like particles in the nuclei (Fig. 4). It has been estimated that an infected cell may harbor as many as 10^5 GLV particles without lysis. Meanwhile, mature and infectious GLV particles begin to appear in the supernatant of *Giardia* culture medium 42 h after infection (Wang and Wang, 1986; Tai *et al.*, 1991). It is not known whether a specific cellular process is involved in the GLV release into the culture supernatant, although cell lysis has not been observed as a consequence of viral infection. The viruses are excreted into the culture medium and can infect these same isolates of many other free *Giardia*. The *G. lamblia* growth rate of newly infected cells decreases with increasing infection multiplicity. As the ratio of the infecting virus to cell increases, the percentage of the non-adhering (non-dividing) trophozoites also increases. However, in established infected cell lines or at moderate multiplicity of infection (<1000), the infected cell assumes normal appearance and maintains the same growth rate as its uninfected counterpart. Furthermore, GLV infection persists indefinitely throughout repeated subcultures (Wang and Wang, 1986; Miller *et al.*, 1988).

It is suggested by Janssen *et al.* (2015) that GLV can transcribe its genome inside the icosahedral capsid right after entry into the *G. lamblia* cytoplasm. The genome is transcribed by the RdRp domain of the CP/RdRp subunit(s) still inside the "T=2" icosahedral capsid, and the plus-strand transcripts are then released from the particle, allowing translation of the GLV proteins and packed into nascent progeny virions. This strategy is thought to be common to all dsRNA viruses, possibly except for Bimaviruses (Luque *et al.*, 2009).

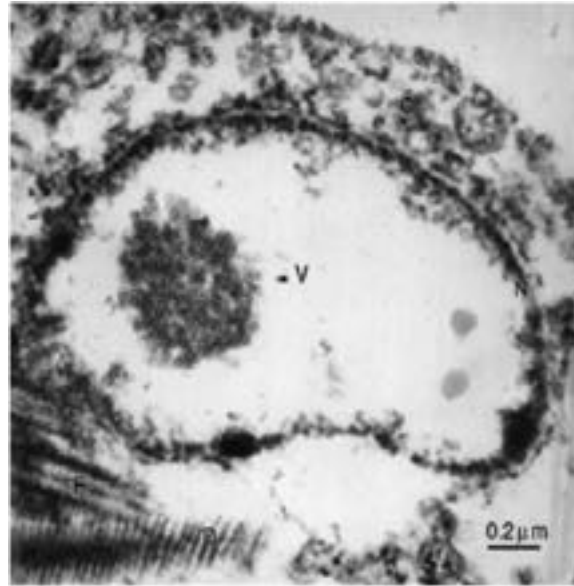


Fig. 4 Thin section of a giardiavirus-infected cell. C, cytoplasm; N, nuclear envelope; F, flagella; D, ventral disk; V, aggregates of virus particles.

In addition to the GLV 6.3 kbp genomic dsRNA, a single stranded RNA (ssRNA) of identical length that is homologous to only one of the two strands of GLV dsRNA is also found in the GLV-infected cellular extract (Furfine *et al.*, 1989). Studies of the infection time course showed that, in contrast to GLV dsRNA which increases steadily from 23 to 141 h after infection, ssRNA becomes detectable at about the same time, peaks at 42–50 h after infection, then gradually declines. Electroporation of gel-purified ssRNA into the uninfected WB cells resulted in the recovery of dsRNA from the cell extract and infectious GLV particles in culture supernatant, demonstrating that SS is the viral messenger RNA as well as the full-length replicative form of the viral genome (Furfine and Wang, 1990). Nucleotide sequence analysis of GLV cDNA clones also verifies that it is the ssRNA strand that encodes the large ORFs. There is no subgenomic viral RNA detected inside the infected cell. The RNA products synthesized in vitro by the virion-associated RdRp are homologous to ssRNA and complementary to the negative strand. Transcription of viral message therefore must proceed conservatively by utilizing the dsRNA as its template.

Studies performed by Cao *et al.* (2009) indicate that GCV may be released by budding. TEM micrographs of *G. canis* infected with GCV show a gradual process of budding formation during the virus discharge. Alternatively, they suggest that the virus may also be released by lysis of infected trophozoites because of the apparent loss of the cell membranes when the trophozoites were crowded.

Since the successful cloning of the full length cDNA from dsRNA template of GLV (Wang *et al.*, 1993), the modified giardiavirus RNA has been used as a genetic tool for the study of possible intervention and pathogenesis (Liu *et al.*, 2005). Similarly, GCV has been used to construct a stable transfection system which was used to stably transfect trophozoites of *G. canis* (Chen *et al.*, 2007; Liu *et al.*, 2008). Liu *et al.* (2008) successfully established a stable transfection system based on the GCV with stable expression of green fluorescent protein in *G. canis*, while Chen *et al.* (2007) cloned a hammerhead ribozyme flanked with various lengths of antisense Krr1 RNA into a viral vector derived from GCV genome. In the same study, RNA plasmid transcripts showed high cleavage activities on *G. canis* Krr1 mRNA in vitro causing a decrease in its levels and leading to trophozoite deformation. The extent to which a person is affected by giardiasis varies widely. The outcome of the severity of parasitic infection depends on many variables derived not only from the parasite, but also from the human host. Attempts at correlating the presence or absence of GLV with the severity of giardiasis have been inconclusive, and we still do not know the role GLV might play in the delicate balance of the host–parasite interaction, if any.

References

- Cao, L., Gong, P., Li, J., *et al.*, 2009. *Giardia canis*: Ultrastructural analysis of *G. canis* trophozoites transfected with full length *G. canis* virus cDNA transcripts. *Exp. Parasitol.* 123, 212–217.
- Chen, S., Cao, L., Huang, Q., Qian, Y., Zhou, X., 2016. The complete genome sequence of a novel maize-associated totivirus. *Arch. Virol.* 161, 487–490.
- Chen, L.F., Li, J.H., Zhang, X.C., *et al.*, 2007. Inhibition of Krr1 gene expression in *Giardia canis* by a virus mediated hammerhead ribozyme. *Veterinary Parasitol.* 143, 14–20.
- Dantas, M.D.A., Chavante, S.F., Teixeira, D.I.A., Lima, J.P.M., Lanza, D.C., 2015. Analysis of new isolates reveals new genome organization and a hypervariable region in infectious myonecrosis virus (IMNV). *Virus Res.* 203, 66–71.
- Furfine, E.S., Wang, C.C., 1990. Transfection of the *Giardia lamblia* double-stranded RNA virus into *Giardia lamblia* by electroporation of a single-stranded RNA copy of the viral genome. *Mol. Cell. Biol.* 10 (7), 3659–3662.
- Furfine, E.S., White, T.C., Wang, A.L., Wang, C.C., 1989. A single-stranded RNA copy of the *Giardia lamblia* virus double-stranded RNA genome is present in the infected *Giardia lamblia*. *Nucleic Acids Res.* 17 (18), 7453–7467.

- Garlapati, S., Chou, J., Wang, C.C., 2001. Specific secondary structures in the capsid-coding region of giardiavirus transcript are required for its translation in *Giardia lamblia*. J. Mol. Biol. 308, 623–638.
- Garlapati, S., Wang, C.C., 2004. Identification of a novel internal ribosome entry site in *Giardiavirus* that extends to both sides of the initiation codon. J. Biol. Chem. 279, 3389–3397.
- Garlapati, S., Wang, C.C., 2005. Structural elements in the 5'-untranslated region of giardiavirus transcript essential for internal ribosome entry site-mediated translation initiation. Eukaryotic Cell 4, 742–754.
- Guo, L., Yang, X., Wu, W., et al., 2016. Identification and molecular characterization of Panax notoginseng virus A, which may represent an undescribed novel species of the genus *totivirus*, family *Totiviridae*. Arch. Virol. 161, 731–734.
- ICTV – International Committee on Taxonomy of Viruses, 2015. Virus Taxonomy: Release. Available at: <http://www.ictvonline.org/virusTaxonomy.asp> (accessed 30.10.16)
- Inger, S.H., Bjørn, K.G., Lucy, J.R., 2007. A longitudinal study on the occurrence of *Cryptosporidium* and *Giardia* in dogs during their first year of life. Acta Vet. Scand. 49, 22.
- Janssen, M.E.W., Takagi, Y., Parent, K.N., et al., 2015. Three-dimensional structure of a protozoal double-stranded RNA virus that infects the enteric pathogen *Giardia lamblia*. J. Virol. 89, 1182–1194.
- King, A.M., Adams, M.J., Lefkowitz, E.J., Carstens, E.B., 2012. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Amsterdam: Elsevier.
- Lachnit, T., Thomas, T., Steinberg, P., 2016. Expanding our understanding of the seaweed holobiont: RNA viruses of the red alga *Delisea pulchra*. Front. Microbiol. 6, 1489.
- Liu, C., Li, J., Zhang, X., Liu, Q., et al., 2008. Stable expression of green fluorescent protein mediated by GCV in *Giardia canis*. Parasitol. Int. 57, 320–324.
- Liu, Q., Zhang, X., Li, J., et al., 2005. *Giardia lamblia*: Stable expression of green fluorescent protein mediated by giardiavirus. Experimental Parasitology 109 (3), 181–187. (Epub 2005 Jan 20).
- Luque, D., Saugar, I., Rejas, M.T., et al., 2009. Infectious bursal disease virus: Ribonucleoprotein complexes of a double-stranded RNA virus. J. Mol. Biol. 386, 891–901.
- Miller, R.L., Wang, A.L., Wang, C.C., 1988. Identification of *Giardia lamblia* isolates susceptible and resistant to infection by the double-stranded RNA virus. Exp. Parasitol. 66 (1), 118–123.
- Mor, S.K., Phelps, N.B., 2012. Molecular detection of a novel totivirus from golden shiner (*Notemigonus crysoleucas*) baitfish in the USA. Arch. Virol. 161, 2227–2234.
- Oliveira, R.A., Almeida, R.V., Dantas, M.D., et al., 2014. In silico single strand melting curve: A new approach to identify nucleic acid polymorphisms in Totiviridae. BMC Bioinformatics 15, 243.
- Sepp, T., Wang, A.L., Wang, C.C., 1994. Giardiavirus-resistant *Giardia lamblia* lacks a virus receptor on the cell membrane surface. J. Virol. 68, 1426–1431.
- Tai, J.H., Ong, S.J., Chang, S.C., Su, H.M., 1993. Giardiavirus enters *Giardia lamblia* WB trophozoite via endocytosis. Exp. Parasitol. 76, 65–74.
- Tai, J.H., Wang, A.L., Ong, S.J., et al., 1991. The course of giardiavirus infection in the *Giardia lamblia* trophozoites. Exp. Parasitol. 73 (4), 413–423.
- Thompson, R.C.A., Ash, A., 2016. Molecular epidemiology of *Giardia* and *Cryptosporidium* infections. Infection Genetics. Evolution 40, 315–323.
- Wang, A.L., Miller, R.L., Wang, C.C., 1988. Antibodies to the *Giardia lamblia* double-stranded RNA virus major protein can block the viral infection. Mol. Biochem. Parasitol. 30 (3), 225–232.
- Wang, A.L., Wang, C.C., 1986. Discovery of a specific double-stranded RNA virus in *Giardia lamblia*. Mol. Biochem. Parasitol. 21 (3), 269–276.
- Wang, A.L., Wang, C.C., 1991. Viruses of the protozoa. Annu. Rev. Microbiol. 45, 251–263.
- Wang, A.L., Yang, H.M., Shen, K.A., Wang, C.C., 1993. Giardiavirus double-stranded RNA genome encodes a capsid polypeptide and a gag-pol- like fusion protein by a translation frameshift. Proc. Natl. Acad. Sci. 90, 8595–8599.
- Wiik-Nielsen, J., Alarcón, M., Fineid, B., Rode, M., Haugland, Ø., 2012. Genetic variation in Norwegian piscine myocarditis virus in Atlantic salmon. *Salmo salar* L.J. Fish Diseases 36, 129–139.
- Yee, J., Nash, T.E., 1995. Transient transfection and expression of firefly luciferase in *Giardia lamblia*. Proc. Natl. Acad. Sci. 92, 5615–5619.
- Yu, D.C., Wang, A.L., Wang, C.C., 1996. Stable coexpression of a drug-resistance gene and a heterologous gene in an ancient parasitic protozoan *Giardia lamblia*. Mol. Biochem. Parasitol. 83, 81–91.
- Yu, D.C., Wang, A.L., Wu, C.H., Wang, C.C., 1995. Virus-mediated expression of firefly luciferase in the parasitic protozoan *Giardia lamblia*. Mol. Cell. Biol. 15, 4867–4872.

Hypoviruses (*Hypoviridae*)

Dong-Xiu Zhang and Donald L. Nuss, University of Maryland, Rockville, MD, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of D.L. Nuss, Hypoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00406-4.

Glossary

Anastomosis The fusion of fungal hyphae resulting in exchange of cytoplasmic material and hypovirus transmission.

Hypovirulence Virus-mediated attenuation of fungal virulence.

Vegetative incompatibility A system controlled by at least six genetic loci that determines the ability of two fungal strains to undergo anastomosis.

Introduction

The discovery of hypoviruses, a group of RNA viruses that reduce the virulence (hypovirulence, see article on Hypovirulence; volume 3, this encyclopedia) of the chestnut blight fungus *Cryphonectria parasitica*, stimulated intensive research into the potential of using fungal viruses for biological control of fungal diseases. Documented examples of virus-mediated hypovirulence have been reported for fungal diseases of plants that range from trees to turfgrass and involve mycoviruses which include representatives from the *Totiviridae*, *Chrysoviridae*, *Reoviridae*, *Narnaviridae* and *Hypoviridae* (the hypoviruses). However, the hypoviruses remain the most thoroughly studied of the hypovirulence-associated viruses, primarily due to several significant advances in hypovirus molecular biology.

In 1977, Peter Day and co-workers at the Connecticut Agricultural Experiment Station reported that hypovirulent *C. parasitica* strains harbor double stranded (ds) RNAs, providing the first indication of the nature of cytoplasmic elements responsible for the hypovirulence phenotype. Subsequent surveys of dsRNAs associated with different North American and European hypovirulent strains revealed considerable variations in concentration, number and size of dsRNA components. By the late 1980s, it was clear that a detailed molecular analysis of the dsRNAs associated with a single hypovirulent strain was required to bring some measure of order to the mounting confusion generated by such surveys. This resulted in the cloning and complete sequence determination of the prototypic hypovirus, now designated CHV1-EP713, in 1991. This milestone was followed in 1992 by the construction of an infectious full-length (12,712 bp) cDNA clone of CHV1-EP713 RNA by Choi and Nuss. This development furnished direct evidence that hypoviruses are indeed the causative agents responsible for transmissible hypovirulence and provided the means for facile manipulation of the hypovirus genome. Current interest in hypoviruses extends past biological control potential to their utility as unique experimental tools for probing fundamental processes underlying fungal pathogenesis and mycovirus-fungal host interactions.

Taxonomy and Genetic Organization

Hypoviruses are classified within the family *Hypoviridae*, consisting of the genus *Hypovirus* and four species with the exemplar strains designated as *Cryphonectria parasitica* hypovirus 1–4 (CHV1–CHV4). Hypovirus taxonomy is not based on virus structure, since this group of viruses does not encode a coat protein, but on genome organization, sequence similarity and symptom expression. Hypovirus genetic information is found predominantly as dsRNA associated with membrane vesicles ranging in diameter from 50 to 80 nm. As observed for fungal viruses generally, hypoviruses exhibit no extracellular phase in their life cycle. Infections cannot be initiated by inoculation with an infected cell extract or enriched fractions. Instead, these viruses are transmitted by cytoplasmic mixing as a result of fusion (anastomosis) between vegetatively compatible strains or to a variable degree in asexual spores (conidia).

The hypoviruses were assigned the designations CHV1–CHV4 in the order in which their genome sequence was completed. The primary nucleotide sequence for the coding strand of the prototypic member of the genus *Hypovirus*, CHV1-EP713, specifies two large open reading frames (ORFs) designated ORF A and ORF B (Fig. 1). A second closely related member, CHV1-Euro7, has the same organization and shares approximately 90% identity at the nucleotide level with CHV1-EP713. The type member of the *Cryphonectria parasitica* hypovirus 2, CHV2-NB58, shares only about 60% nucleotide sequence identity with CHV1-EP713 and lacks a portion of ORF A that, for the CHV1, encodes a functional cis-acting cysteine protease. Sequenced members of *Cryphonectria parasitica* hypovirus 3 and 4, CHV3-GH2 and CHV4-SR2, respectively, are several kb shorter than CHV1-EP713 and CHV2-NB58, ~9.2–9.8 kb versus 12.5–12.7 kb, contain a single large ORF rather than two ORFs and are both more distantly related phylogenically to CHV1-EP713 than CHV1-EP713 is to CHV2-NB58.

Unassigned hypovirus-related viruses have been reported to infect filamentous fungi other than *C. parasitica*, including *Sclerotinia sclerotiorum*, *Valsa ceratoperma*, *Phomopsis longicolla*, *Fusarium* ssp., *Macrophomina phaseolina* and *Agaricus bisporus*.

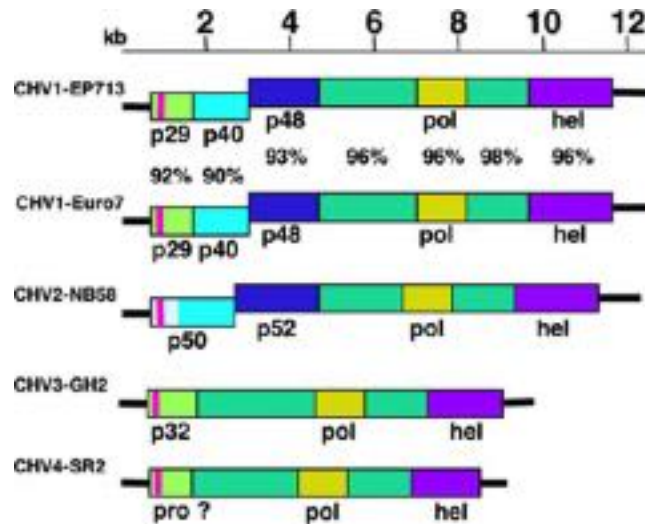


Fig. 1 Genetic organization of sequenced hypovirus genomes representing the four species that comprise the genus *Hypovirus*, family *Hypoviridae*. Amino acid identity levels for coding regions of the two sequenced members of *Cryphonectria hypovirus 1* species, CHV1-EP713 and CHV1-Euro7, are indicated between representations of the two viral genomes. Protein coding regions homologous to CHV1-EP713 encoded p29, p40, p48, polymerase and helicase are color coded. The magenta regions represent a short conserved cysteine rich domain. Note that the genomes of CHV3-GH2 and CHV4-SR2 contain a single ORF. Modified from Dawe, A.L., Nuss, D.L., 2001. Hypoviruses and chestnut blight: Exploiting viruses to understand and modulate fungal pathogenesis. Annual Review of Genetics 35, pp. 1–29, with permission from Annual Reviews.

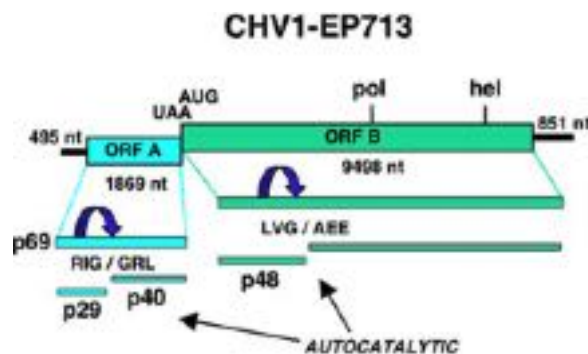


Fig. 2 Expression strategy for prototypic hypovirus CHV1-EP713. The CHV1-EP713 coding strand consists of 12,712 nucleotides excluding a poly(A) tail, and contains two major coding domains designated ORF A and ORF B. Details are discussed in the text. Modified from Dawe, A.L., Nuss, D.L., 2001. Hypoviruses and chestnut blight: Exploiting viruses to understand and modulate fungal pathogenesis. Annual Review of Genetics 35, pp. 1–29, with permission from Annual Reviews.

Hypovirus Gene Expression Strategy

Although hypovirus genetic information is readily recovered from infected cultures as linear dsRNA, the absence of a discrete virus particle and an extracellular infection phase presents some difficulties in precisely defining the hypovirus genome. Synthetic copies of the coding strand are infectious by electroporation into fungal spheroplasts and phylogenetic analyses suggests a common ancestry with the positive strand RNA plant potyviruses. Thus, one could consider hypoviruses as having a positive strand RNA genome and the dsRNA as representing accumulated replicative form RNA. Irrespective of this complication, direct analysis of hypovirus dsRNAs, cDNA cloning studies and *in vitro* translational analyses has provided the following view of genetic organization and expression strategies for the prototypic hypovirus CHV1-EP713 shown in Fig. 2. One strand contains a 3'-poly A tail while the complementary strand contains a 5' poly U tract. All of the CHV1-EP713 coding information appears to reside within two contiguous ORFs on the 12,712 nt long polyadenylated strand. Translation of ORF A is facilitated by an internal ribosome entry site (IRES) located at the end of the 5'-non-coding region and extending into the beginning of the coding domain. ORF A encodes two polypeptides, p29 and p40, that are released from a polypeptide precursor, p69, by an autoproteolytic event between Gly-248 and Gly-249, mediated by a cysteine-like protease catalytic domain located within p29. ORF B has the capacity to encode a polypeptide of 3165 amino acids and contains unmistakable RNA-dependent RNA polymerase and helicase motifs. Proteolytic processing of only a portion of the ORF B polypeptide has been elucidated in the form of the autoproteolytic release of a 48 kDa

protein, p48, from the N-terminus. This cleavage event occurs between Gly-418 and Ala-419 and is catalyzed by essential residues Cys-341 and His-388 within p48. The junction between ORF A and ORF B is defined by the sequence 5'-UAAUG-3', in which the UAA portion clearly serves as the termination codon for ORF A and the AUG portion is thought to serve as the initiation codon for ORF B. The mechanism by which ribosomes transition through the junction involves coupled termination/reinitiation. This unusual pentanucleotide sequence is found at the ORF A/ORF B junction for all confirmed strains of *Cryphonectria hypovirus 1*. There is clearly a need for additional fine detailed mapping of the processing cascades for hypovirus-encoded polyproteins.

The genome of CHV2-NB58, the type strain of *Cryphonectria hypovirus* 2, also consists of a two ORF configuration with a UAAUG junction (Fig. 2). However, ORF A lacks the p29 papain-like catalytic or cleavage sites and directs the translation of a 50-kDa protein product. ORF B of CHV2-NB58 does contain a p48 homologue, p52. The N-terminal portion of the single ORF of type strain CHV3-GH2 (*Cryphonectria hypovirus* 3), contains a protease, p32, with similarity to p29. A putative protease domain has been identified at the N-terminal portion of the type strain CHV4-SR2 of *Cryphonectria hypovirus* 4, but protease cleavage has not been demonstrated.

In most hypovirus-infected *C. parasitica* isolates, the full-length viral dsRNA is accompanied by a constellation of shorter dsRNA species. These ancillary dsRNAs appear to be generated by internal deletion events, are replicated only in the presence of the full-length viral RNA and are not associated with any function or phenotypic effect.

Hypovirus Functional Domains

Hypovirus-encoded symptom determinants and important expression and replication elements have been mapped (**Fig. 3**) by a combination of approaches that include (a) the construction of recombinant chimeras from hypoviruses that differ in their influence on host phenotype, (b) mutagenesis of a hypovirus infectious cDNA clone and (c) cellular expression of viral coding domains independent of virus infection.

In analogy with plant viruses, CHV1-EP713 and CHV1-Euro7 can be viewed as severe and mild hypovirus isolates, respectively. Although these two CHV1 isolates cause quite different phenotypic changes in their fungal host, they share a high level of sequence similarity that has allowed the construction of viable chimeric viruses to begin mapping the determinants responsible for the differences in phenotypic changes.

Differences in host colony morphologies were found to map to a region extending from a position just downstream of the p48 coding domain (map position 3575) to map position 9879, with clear indications of multiple discrete determinants. More specifically, the region extending from position 3575 to 5310 was able to confer a CHV1-EP713-like colony morphology when inserted into a CHV1-Euro7 genetic background. The CHV1-EP713 p48 coding region was found to be a dominant determinant contributing to suppression of asexual spore formation on the canker face. In a separate study, p48 was shown to have the unusual property of being required for the initiation but not the maintenance of viral RNA replication.

The chimeric hypoviruses also proved to be very useful reagents when coupled with a pathway specific promoter/reporter system to map viral determinants responsible for altering host G-protein/cAMP-mediated signaling. A common undesired side effect of hypovirus-mediated virulence attenuation is a significantly reduced ability of the fungal host to colonize and produce spores on the corresponding plant host. This reduces the ability of hypovirulent fungal strains to persist and spread through the ecosystem. Thus, from a practical perspective, a better understanding of the nature of viral symptom determinants and their relative effects on specific regulatory pathways and expression of gene clusters provides the means for a more rational approach for engineering hypoviruses that exhibit a desired balance between virulence attenuation and ecological fitness.

Additional insights into the functional role of viral coding regions were indirectly provided during efforts to develop hypoviruses as gene expression vectors. The nucleotide sequence corresponding to the first 24 codons of p29 was found to be required for viral replication, while the remaining 598 codons of ORF A, including all of the p40 coding region, was found to be dispensable. The 24 N-terminal codons were subsequently shown to function as an essential element of an internal ribosome entry

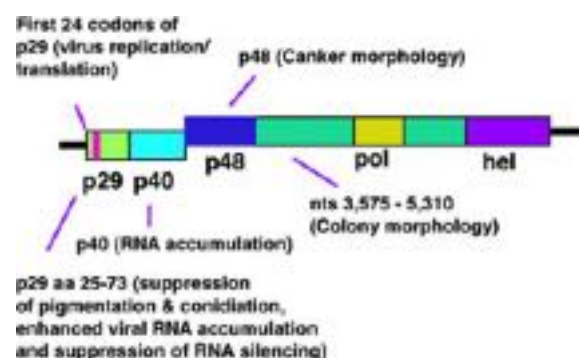


Fig. 3 Map of CHV1-EP713 symptom determinants and essential/dispensable replication elements as described in the text. Modified from Dawe, A.L., Nuss, D.L., 2001. Hypoviruses and chestnut blight: Exploiting viruses to understand and modulate fungal pathogenesis. *Annual Review of Genetics* 35, pp. 1–29, with permission from Annual Reviews.

site (IRES) involved in ORF A translational initiation. Substantial alterations were also tolerated in the pentanucleotide UAAUG that contains the ORF A termination codon and the overlapping ORF B initiation codon. For example, replication competence was maintained following either a frame-shift mutation that caused a two-codon extension of ORF A or a modification that produced a single-ORF genomic organization. Further characterization of p40 revealed a role as an accessory function in viral RNA amplification with a functional domain extending from Thr(288) to Asn(313) and a role of a p40 coding domain just upstream of the UAAUG pentamer in coupled translational termination/reinitiation.

Expression of the CHV1-EP713 encoded papain-like protease, p29, in the absence of virus infection was shown to cause a subset of phenotypic changes exhibited by CHV1-EP713-infected strains, e.g., a white phenotype (reduction in orange pigmentation), reduced asexual sporulation and a slight reduction in the production of fungal laccase activity. By deleting all but the first 24 N-terminal codons of p29 in the context of the CHV1-EP713 infectious cDNA clone (mutant virus Δ p29), it was also possible to show that the p29 protein is dispensable for viral replication and to demonstrate a near restoration of orange pigment production and a moderate increase in conidiation levels relative to wild-type CHV1-EP713-infected fungal colonies. Deletion of p29 had no effect on virus-mediated virulence attenuation. A gain-of-function analysis involving progressive repair of the Δ p29 mutant was also devised to map the p29 symptom determinant domain to a region extending from Phe-25 to Gln-73. When expressed from a chromosomally integrated cDNA copy, p29 elevated RNA accumulation and vertical transmission (through conidia) of the Δ p29 mutant virus to levels observed for wild-type CHV1-EP713. Additional mutational studies indicated a linkage between p29-mediated changes in host phenotype in the absence of virus infection and p29-mediated *in trans* enhancement of viral RNA accumulation and transmission.

The multifunctional nature of p29 was extended to include suppressor of RNA silencing both in the natural fungal host and in a heterologous plant system. This activity was predicted based on similarities between p29 and the well-characterized potyvirus-encoded suppressor of RNA silencing HC-Pro and provided the first circumstantial evidence that RNA silencing in fungi may serve as an antiviral defense mechanism, a well-established function in plants.

Anti-Hypovirus Defense Mechanisms

Direct evidence for RNA silencing as a defense against hypovirus infection was obtained by systematic disruption of the *C. parasitica*-encoded Dicer and Argonaute genes, core RNA silencing pathway components. Dicers nucleases recognize and process viral double-stranded and structured RNAs into small 21–24 nt long RNAs, termed virus-derived small (vs) RNAs. With the aid of an Argonaute family protein, the vsRNAs are incorporated into an effector complex termed the RNA-induced silencing complex (RISC). One strand of the vsRNA is then degraded and the remaining strand guides the effector complex to the cognate viral RNA, which is then cleaved by the Argonaute-associated RNase H-like activity. Disruption of the two Dicer-like genes and the four Argonaute-like genes found in *C. parasitica* resulted in no observable phenotypic changes. However, infection of the mutant strains with hypoviruses or an unrelated mycoreovirus MYRV1-Cp9B21 resulted in increased viral RNA levels and severe virus-induced symptoms only in the Δ dcl2 and Δ agl2 mutants. The dependence on a single Dicer and a single Argonaute for antiviral defense in *C. parasitica* contrasts with dependence on multiple Dicer and Argonaute genes for antiviral defense in plants.

Dicer *dcl2* transcript levels were shown to increase 10–12-fold in an *agl2*-dependent manner, following infection by Cryphonectria hypovirus 1 (CHV1) and to exhibit super-induction (over 30-fold) after infection by CHV1- Δ p69, lacking the p29 viral suppressor of RNA silencing. A genetic screen subsequently identified the conserved transcriptional activator SAGA (Spt-Ada-Gen5 acetyltransferase) as a regulator of this up-regulation. Mutational analysis demonstrated that the SAGA-associated histone acetyltransferase (HAT) activity, but not the histone deubiquitinase (DUB) activity, was the major influence on the SAGA-mediated induction of *dcl2* transcription following virus infection.

Early efforts to characterize hypovirus RNA at the molecular level revealed the accumulation of internally deleted defective interfering (DI) RNAs generated from full-length viral RNA by a combination of deletion and recombination events. Recombination-mediated deletion of non-viral nucleotide sequences also limited efforts to use recombinant hypoviruses to express foreign genes. Interestingly, DI RNAs failed to accumulate and recombinant hypovirus expression vectors were shown to be stable in the Δ dcl2 and Δ agl2 mutants. The combined results were the first to link an RNA silencing pathway to viral RNA recombination and are of potentially broad significance with possible applications for stable production of biologicals by mycovirus-based vectors.

While the RNA silencing pathway serves as an anti-viral defense at the cellular level, the *C. parasitica* vegetative incompatibility (*vic*) non-self recognition system provides antiviral defense at the population level. In the absence of an extracellular phase to their replication cycle, mycoviruses rely on anastomosis (fusion of hyphae) as a major avenue for horizontal transmission. The fungal *vic* system restricts mycovirus transmission due to incompatible reactions triggered when genetically distinct, *vic*-incompatible individuals of the same species interact. This reaction results in localized programmed cell death along the zone of contact, limiting transmission of viruses and other cytoplasmic elements.

The *C. parasitica* *vic* system is controlled by at least six di-allelic loci, designated *vic1*, *vic2*, *vic3*, *vic4*, *vic6* and *vic7*, that were recently identified and characterized at the molecular level. Gene disruption analysis formally confirmed that five of the six loci contribute to restriction of virus transmission. An allelic difference at *vic4* does not restrict virus transmission. Systematic disruption of four of the five virus-restricting *vic* gene alleles was reported to engineer a super hypovirus donor (SD) strain formulation that was able to transmit hypovirus under laboratory conditions to uninfected strains that were heteroallelic at one to five of the virus-restricting *vic* loci. These results predict that the SD formulation could circumvent *vic*-imposed restrictions to virus

transmission by serving as an effective vector to introduce hypovirus into natural field strains with widely diverse *vic* genotypic combinations of the six defined diallelic *vic* loci. This property should find application in efforts to integrate biological control with increased disease resistance derived from backcross breeding programs or transgenes for effective woodland restoration of the American chestnut.

Further Reading

- Dawe, A.L., Nuss, D.L., 2013. Hypovirus molecular biology: From Koch's postulates to host self-recognition genes that restrict virus transmission. *Advances in Virus Research* 86, 109.
- Hillman, B.I., Suzuki, N., 2004. Viruses of the chestnut blight fungus, *Cryphonectria parasitica*. *Advances in Virus Research* 63, 423.
- Nuss, D.L., 2005. Hypovirulence: Mycoviruses at the fungal-Plant interface. *Nature Reviews Microbiology* 3, 632.
- Suzuki, N., Ghabrial, S.A., Kim, K., *et al.*, 2018. ICTV virus taxonomy profile: *Hypoviridae*. *Journal of General Virology* 99, 615.
- Zhang, D.X., Nuss, D.L., 2016. Engineering super mycovirus donor strains of chestnut blight fungus by systematic disruption of multilocus *vic* genes. *Proceedings of the National Academy of Sciences of the United States of America* 113, 2062.
- Zhang, D.X., Spiering, M.J., Dawe, A.L., Nuss, D.L., 2014. Vegetative incompatibility loci with dedicated roles in allorecognition restrict mycovirus transmission in chestnut blight fungus. *Genetics* 197, 701.

Megabirnaviruses (*Megabirnaviridae*)

Yukiyo Sato, Okayama University, Kurashiki, Japan

Nobuhiro Suzuki, Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Ascomycetous fungi Fungi phylogenetically classified into the division Ascomycota, which is characterized by the formation of spores in asci in the fungal life cycle.

Hypovirulence Mycovirus-mediated attenuation in virulence of phytopathogenic fungi to plants.

Mycoviruses Viruses that infect fungi.

Ribosomal-1 frameshifting A translation mechanism of the polycistronic gene, in which ribosome slip through one

nucleotide on a slippery or shifty heptamer (typically “X XXY YYZ”, where the same letter represents the same set of nucleotides), followed by an RNA element forming a pseudo-knot or stem-loop structure.

Spheroplasts Isolated cells enclosed in partially digested cell walls.

T = 1 icosahedral symmetry Symmetry of capsid composed of 60 subunits, which forms 12 pentameric capsomeres.

Introduction

Rosellinia necatrix megabirnavirus 1 (RnMBV1, species *Rosellinia necatrix megabirnavirus 1*, genus *Megabirnavirus*, family *Megabirnaviridae*) was first isolated in 2009 from a phytopathogenic ascomycetous fungus, *Rosellinia necatrix*, which is the causal agent of white root rot disease and causes severe damage to many perennial crops. The virus was discovered during routine screens for mycoviruses conferring hypovirulence to the host fungus as part of a project led by Dr. Naoyuki Matsumoto that was aimed at virocontrol, a form of biocontrol using viruses that was initiated in the 1990s in Japan. The research group was inspired by the success of the virocontrol of chestnut blight disease using hypoviruses in Europe. RnMBV1 inhibits growth of and reduces virulence of *R. necatrix*. It is composed of a bipartite double-stranded RNA (dsRNA) genome that is encapsidated in non-enveloped spherical particles that are 52 nm in diameter. Among the established taxa to which multipartite dsRNA viruses are assigned, namely the family *Reoviridae*, *Partitiviridae*, *Chrysoviridae*, *Quadriviridae* and *Megabirnaviridae*, and the genus *Botybirnavirus*, the polycistronic gene encoding strategy is observed only in megabirnaviruses. Thus, in 2011 RnMBV1 was assigned to the newly established taxa, the genus *Megabirnavirus* in the family *Megabirnaviridae*. “Birna-” is derived from bipartite segments and “Mega-” means much greater genome size (approximately 16 kbp) than those of *birnaviridae* and *pico-birnaviridae* (approximately 6 kbp and 4 kbp, respectively). The family *Megabirnaviridae* is currently comprised of a single genus, *Rosellinia necatrix megabirnavirus 1*, to which only the type member is assigned. Besides, several mycoviruses that are phylogenetically related to and share common properties with RnMBV1 have been reported in the last five years, although they have not been officially assigned to the taxa. In this article the properties of RnMBV1 are summarized and compared with those of related viruses.

Virion Properties

The megabirnavirus genome is encapsidated in non-enveloped spherical particles that are 52 nm in diameter (Fig. 1). Virus particles contain a bisegmented dsRNA genome (Fig. 2). Each segment seems to be packaged separately, based on the two following observations: first, purified virus particles include an unequal and variable molar ratio of two dsRNA segments; second, artificial virion transfection into hosts can trigger the emergence of mutants that completely lack dsRNA2. Virions are constructed from the major capsid protein (CP), which is 136 kDa. Purified virions also contain minor CP-RdRP fusion proteins that are greater than 250 kDa (Fig. 1). Purified virions can be introduced to spheroplasts from the natural and experimental fungal hosts, *R. necatrix* and *Cryphonectria parasitica*, respectively.

Genome Organization

The megabirnavirus genome consists of two segments, termed dsRNA1 and dsRNA2 (Fig. 2, Table 1). Each segment has two non-overlapping putative open reading frames (ORFs). The larger segment (dsRNA1, 8.9 kbp) contains ORF1 and ORF2, which encode the major structural protein (CP, P1) and RNA-dependent RNA polymerase (RdRP, P2), respectively. The smaller segment (dsRNA2, 7.2 kbp) contains ORF3 and ORF4, encoding hypothetical proteins P3 and P4, respectively, with unknown function. Both genome segments contain an extremely long 5'-untranslated region (UTR) of over 1.6 kbp. 3'-UTRs of both segments are relatively short (< 0.4 kbp). The terminal sequence of 5'-UTR and 3'-UTR is conserved between both segments. The 5'-terminal 24-bp and 3'-terminal 8 bp-sequence are strictly conserved between the RnMBV1 segments (Fig. 3).

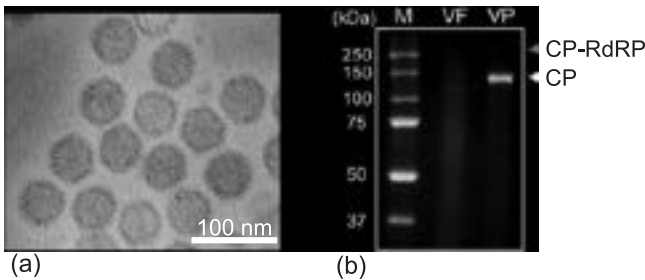


Fig. 1 Transmission electron micrograph (TEM) image and structural proteins of purified RnMBV1 particles. (a) TEM image of negatively stained RnMBV1 virions, the species in the family *Megabirnaviridae*. The scale bar represents 100 nm. (b) Viral proteins composing purified RnMBV1 viral particles (VP). The viral proteins were run on SDS-PAGE gel and stained with Coomassie brilliant blue R-250. The black and gray arrows indicate the bands for CPs or CP-RdRP fusion proteins, respectively. The virus-free (VF) control is included. M indicates molecular size marker for proteins. These figures are reproduced from Chiba, S., Salaipeth, L., Lin, Y., *et al.*, 2009. A novel bipartite double-stranded RNA mycovirus from the white root rot fungus *Rosellinia necatrix*: Molecular and biological characterization, taxonomic considerations, and potential for biological control. *Journal of Virology* 83, 12801–12812, with permission from the American Society for Microbiology.

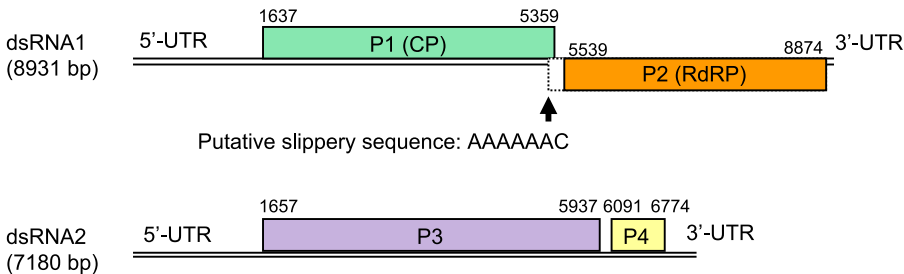


Fig. 2 Genome organization of RnMBV1. Doubled lines and boxes indicates dsRNAs and ORFs, respectively. Numbers near the boxes indicate position of each ORF on the positive strands. The position of putative slippery sequence for ribosomal frameshifting is indicated with an arrow. Reproduced from Sato, Y., Miyazaki, N., Kanematsu, S., *et al.*, 2019. ICTV virus taxonomy profile: *Megabirnaviridae*. *Journal of General Virology* 100, 1269-1270, with permission from the Microbiology Society.

Table 1 Single member in the genus *Megabirnavirus*, the family *Megabirnaviridae*

Virus species	Abbreviation	DsRNA segment no. (length in bp, encoded proteins; size in kDa)	GenBank accession no.
<i>Rosellinia necatrix megabirnavirus 1</i>	RnMBV1	1 (8931; CP, 136; CP + RdRP, 266) 2 (7180; P3, 152; P4, 24)	AB512282 AB512283

Genome Expression and Replication

Replication of megabirnavirus is predicted to occur in virus particles, as is the case for other dsRNA viruses. The fact that purified virions include RdRP-CP fusion proteins supports this hypothesis. RdRP-CP fusion products are likely translated from bicistronic transcripts of dsRNA1 via ribosomal-1 frameshifting. ORF2 encoding RdRP lies downstream -1 reading frame of CP-encoding ORF1 on the same positive strand (Fig. 2). There is a putative slippery sequence (5'-A AAA AAC-3') immediately upstream of the stop codon for ORF1. There is also a sequence element that potentially forms pseudo-knot/stem-loop structures upstream and downstream of the ORF1 stop codon.

The long 5'UTR of megabirnavirus contains multiple mini-ORFs and hypothetical internal ribosomal entry sites (IRESs). It remains to be experimentally determined whether translation initiation of ORF1 and ORF3 is mediated by these IRESs. It is also unknown whether or how ORF4 is expressed.

Virion Structure

The three-dimensional structure of RnMBV1 was determined by cryo-electron microscopy at 15.7 Å resolution (Fig. 4). Each capsid is composed of 60 homodimers of CP with *T* = 1 icosahedral symmetry like many other fungal dsRNA viruses. A distinct features of virions of megabirnavirus is putative 120 large protrusions on the surface of capsid shell. The protrusions are observed, with a width of ~45 Å and a height of ~50 Å, around icosahedral two- and threefold axes, while trenches are observed around icosahedral fivefold axes.

5'-UTR of RnMBV1

dsRNA1	1	GCATAAAAGAGAAGGAAGGTTTAACTTCTGGAACCTTTTATGC---GAACTCGATGAAAA	57
dsRNA2	1	GCATAAAAGAGAAGGAAGGTTTATCTTCTGGA-CTTTTATGCAAAGAACTCGATGAAAA	59

		24 nt	

dsRNA1	58	ATTCTCAAAAGTTTTTCATTTTCAGT---TCTAATCCTTGGATATGAA-AGTCATACCAC-	112
dsRNA2	60	TGAGTGGTAAATTTTCATCCCGGTGTTTCTGGCGCTTTTGC GCGTTCATTCTTCTCT	119
		* * * * *	

(a)

3'-UTR of RnMBV1

dsRNA1	8849	CTTTGTCCAAA---ATGATGAAGATGTGATATTACCTTATTTAAGTTTATATTTTGCAA	8905
dsRNA2	7109	CATTGTAATTGTGATGGTGGGCTGGAGCGCGAGC-----TTCAGA-----GCAT	7154
		* * * * *	

dsRNA1	8906	AAAAGCGGCTGGACCCCTTTTTCGC	8931
dsRNA2	7155	AAAAACGATTGACATCGTTTTCGC	7180
		* * * * *	
		8 nt	

(b)

Fig. 3 Comparison of the terminal sequence of 5' and 3' UTRs of two dsRNA segments of RnMBV1. Multiple sequence alignments of the 5' UTRs (a) and 3' UTRs (b) of two dsRNA segments were performed with ClustalW2. Asterisks indicate the position where an identical base is present. Accession no. of sequence is listed in Table 1.

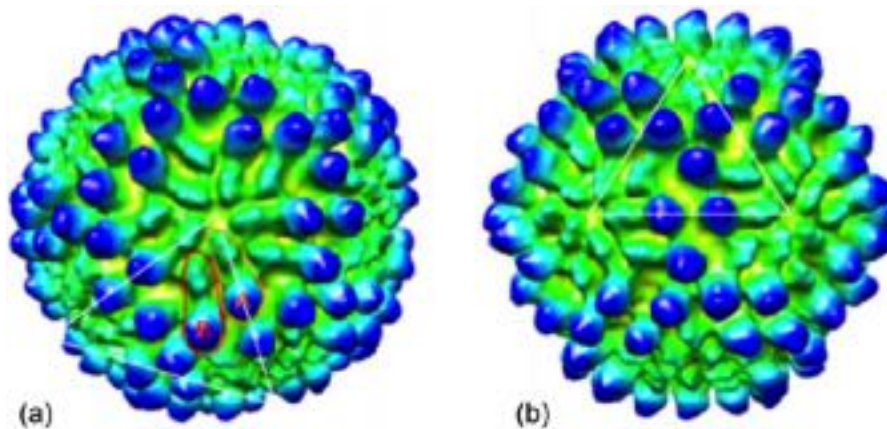


Fig. 4 Three-dimensional cryo-electron microscopy reconstruction of virions of RnMBV1 at a resolution of 15.7 Å. The surface representation of RnMBV1 particles viewed along an axis with fivefold symmetry (a) or twofold symmetry (b). Molecular boundaries of the A and B subunits are drawn with continuous orange and red, respectively. Courtesy of Naoyuki Miyazaki.

The inner and outer diameter of RnMBV1 capsid shell, excluding the protrusions, are 190 and 210 Å, respectively. For *Saccharomyces cerevisiae* virus L-A in the family *Totiviridae*, grooves on capsid shells could play an enzymatic role in cap snatching from cellular mRNA. An investigation into the functions of the protrusions exposed on the surface of RnMBV1 virions could be interesting future work.

Biological Properties

The biological traits of megabirnavirus have been extensively studied in the RnMBV1-W779 strain. RnMBV1 persistently infects the natural host *R. necatrix* via inter-cellular cytoplasmic exchange and cell division. Furthermore, RnMBV1 can be horizontally transmitted to compatible fungal hosts and inhibits the growth of *R. necatrix*, which is a devastating phytopathogen that causes white root rot in fruit trees. Due to the hypovirulence it confers to *R. necatrix* (Fig. 5), RnMBV1 is regarded as a potential biological control agent for plant disease management.

RnMBV1 can be inoculated into the ascomycetous fungus *Cryphonectria parasitica*. *C. parasitica* is not only the causal agent of chestnut blight disease, but it is also a model fungus for studying of virus-virus and virus-host interactions. RnMBV1 is also responsible for growth inhibition and hypovirulence of *C. parasitica*, as well as *R. necatrix*. In a *C. parasitica* mutant ($\Delta dcl2$) that lacks a component of antiviral RNA interference (RNAi) machinery (dicer-like protein 2), accumulation of RnMBV1 is ~20 fold higher than in the wild type. The symptoms induced by RnMBV1 are promoted in $\Delta dcl2$. These observations suggest that RNAi in

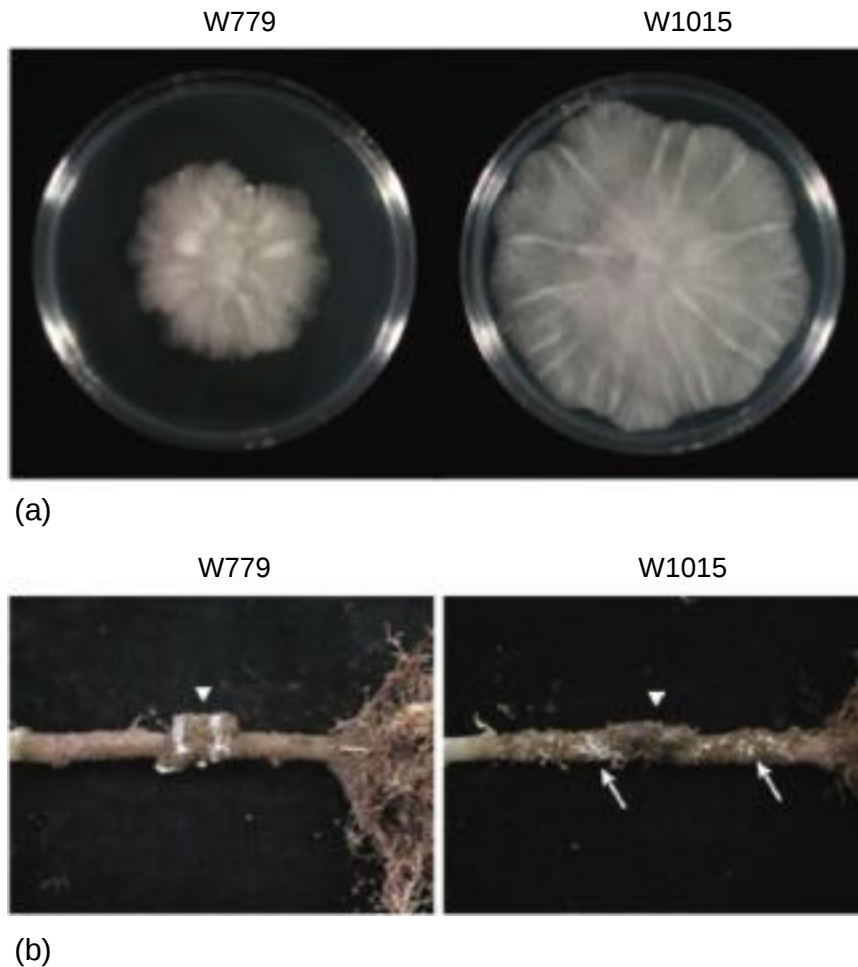


Fig. 5 Mycelial growth of virus-carrying field strain W779 and virus-cured isogenic strain W1015 of *R. necatrix*. (a) Colony morphology of *R. necatrix* strains W779 and W1015. The fungal strains, W779 and W1015, were grown on PDA for 5 days in the dark and photographed. (b) Mycelial growth on apple rootstocks. Twigs of Japanese pear (1.5 cm long) were placed for 3 weeks on 5-day-old PDA cultures of the fungal strains W779 and W1015. Japanese pear twigs covered with mycelia of each fungal strain were fixed on stems of apple rootstocks (*M. prunifolia* var. *ringo*) that had been pre-cultured in a soil medium for horticulture. Inoculated apple plants were continued to be cultured in the soil for additional 2 weeks. The mycelial expansion is shown by white arrows while inoculation sites are denoted by arrowheads. Levels of the mycelial growth on apple rootstocks are equivalent to virulence levels. These figures are reproduced from Chiba, S., Salaipeth, L., Lin, Y., *et al.*, 2009. A novel bipartite double-stranded RNA mycovirus from the white root rot fungus *Rosellinia necatrix*. Molecular and biological characterization, taxonomic considerations, and potential for biological control. *Journal of Virology* 83, 12801–12812, with permission from the American Society for Microbiology.

C. parasitica contributes to its tolerance to RnMBV1, which is stably maintained in both the wild type and $\Delta dcl2$ genotypes during repeated subculture on media.

There is little information on the vertical transmission of the megabirnavirus in nature. The vertical transmission rate of RnMBV1 was investigated in *C. parasitica*, in which asexual spores can be easily formed, and it was found to be substantially low (<1%) in both the wild type and $\Delta dcl2$ genotypes of *C. parasitica*.

All of the three well-characterized megabirnaviruses, namely RnMBV1 and two unassigned viruses termed *Sclerotinia sclerotiorum* megabirnavirus 1 (SsMBV1) and *Rosellinia necatrix* megabirnavirus 2 (RnMBV2), have bi-segmented dsRNA genomes and are associated with host phenotypic alterations. However, virion transfection of host fungal strains often leads to genome alterations, including genome rearrangements and segment loss. Examination of such RnMBV1 transfectants suggests the dispensability of ORF3- and ORF4-encoded proteins. Even the complete loss of dsRNA2 is tolerated for virus viability, as discussed below.

Taxonomic and Phylogenetic Considerations

The Megabirnaviridae family currently consists of a single genus, *Megabirnavirus*, to which only one species, *Rosellinia necatrix* megabirnavirus 1, is assigned. Based on identity of the deduced amino acid sequence of RdRP, megabirnaviruses are placed in a

Table 2 List of the unclassified viruses closely related to megabirnavirus

<i>Virus</i>	<i>Abbreviation</i>	<i>DsRNA segment no. (length in bp, encoded proteins; size in kDa)</i>	<i>GenBank accession no.</i>
Sclerotinia sclerotiorum megabirnavirus 1	SsMBV1	dsRNA1 (8806; P1, 137; P1 + RdRP, 267)	KP686398
Rosellinia necatrix megabirnavirus 2	RnMBV2	dsRNA2 (7909; P3, 179; P4, 40)	KP686399
		dsRNA1 (8985; P1, 138; CP + RdRP, 266)	LC062704
		dsRNA2 (7959; P3, 165; P4, 35)	LC062705
Fusarium pseudograminearum megabirnavirus 1	FpgMBV1	dsRNA1 (8951)	MH057692
		dsRNA2 (5337)	MH057693
Pleosporeales megabirnavirus 1	PMbV1	dsRNA1 (8845; P1, 131; P1 + RdRP, 258)	KT601119
		dsRNA2 (5136; P3, 117)	KT601120
Rosellinia necatrix megabirnavirus 3	RnMBV3 ^a	dsRNA1 (8967; P1, 131; P1 + RdRP, 265)	LC333756
Entoleuca megabirnavirus	EnMBV1 ^a	dsRNA1 (8927; P1, 131; P1 + RdRP, 265)	MF375886
Rhizoctonia solani megabirnavirus 1	RsMBV1 ^b	Not determined ^b	KX349071 ^b
Rhizoctonia solani RNA virus-HN008	RsRV-HN008	dsRNA1 (7596; P1, 128; RdRP, 140)	KP861921

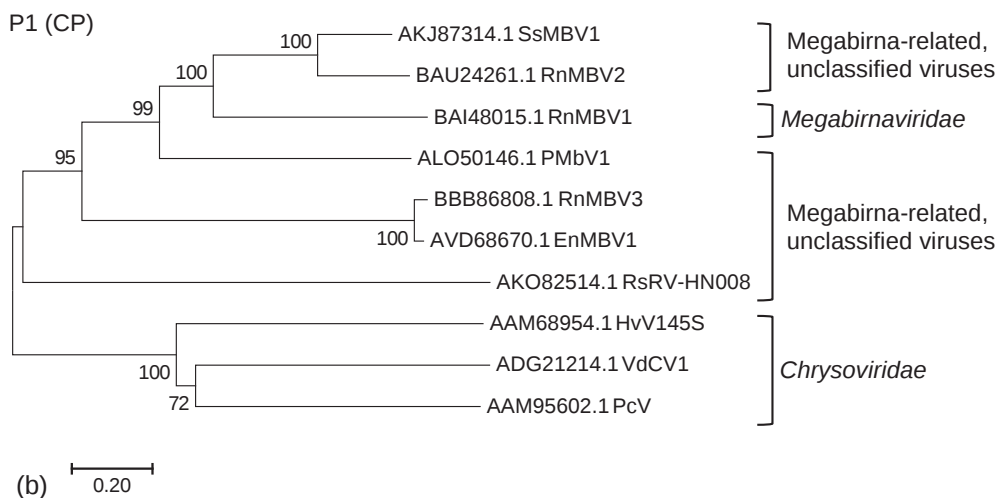
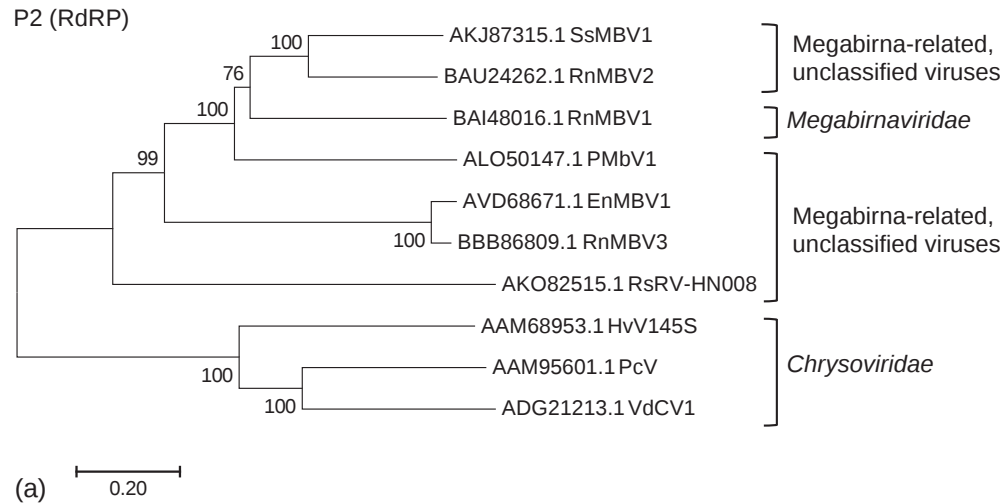
^aPresence/absence and sequence of dsRNA2 remain to be determined.^bOnly partial sequence of RdRP is available.

Fig. 6 Phylogenetic analysis of megabirnavirus and related, unclassified viruses. Phylogenetic trees were constructed based on the complete deduced amino acid sequences of RdRP (a) and P1 (b). Full name of megabirnavirus and related viruses can be found to [Tables 1 and 2](#). The chrysoviruses (HvV145S: *Helminthosporium victoriae* 145S virus; PcV: *Penicillium chrysogenum* virus; and VdCV: *Vorticillium dahliae* chrysovirus 1) are also included as an outgroup. The sequences obtained from GenBank were aligned using MUSCLE. The evolutionary history was inferred using the Neighbor-Joining method and the evolutionary distances were computed using the Poisson correction method. The percentage of reproducibility of trees in the bootstrap test (1000 replicates) are indicated next to the branches. The whole step was conducted using MEGA7 software.

phylogenetic group distinct from the known dsRNA mycoviruses, such as members of the *Chrysoviridae* and *Totiviridae* families. Although megabirnaviruses are phylogenetically related to these dsRNA virus families, megabirnaviruses possess genome organizations and virion structures that are distinct from these viruses.

Recently, several unclassified mycoviruses that are closely related to a member of the *Megabirnaviridae* family have been discovered (Table 2, Fig. 6). Among them, SsMBV1 and RnMBV2 are most closely related to RnMBV1 and have been characterized in detail. The genome organizations and sizes of these two viruses are very similar to RnMBV1 (Tables 1 and 2, Fig. 7). SsMBV1 and RnMBV2 also have bi-segmented linear dsRNA genomes, which consists of 8.8–8.9 kbp of dsRNA1 and 7.2–7.9 kbp of dsRNA2. Both segments of each virus contain two ORFs, long 5'-UTRs that are more than 1.6 kbp and relatively short 3'-UTR. The signature sequence for translation of RdRP via ribosomal -1 frameshifting is also conserved between RnMBV1, SsMBV1 and RnMBV2. The conservation of 5'-UTR sequence between dsRNA1 and dsRNA2 is stricter in SsMBV1 and RnMBV2 than in RnMBV1.

SsMBV1 was found in an ascomycete fungus, *Sclerotinia sclerotiorum*. Unlike the P3 of RnMBV1 or RnMBV2, the P3 of SsMBV1 contains a conserved papain-like protease coding domain, which is commonly found in positive-strand, single-stranded RNA viruses, hypoviruses. Phylogenetic analysis suggests that this papain-like protease domain in the SsMBV1 genome might have been horizontally transferred from *Cryphonectria hypovirus* 1. SsMBV1 moderately attenuates the growth of *S. sclerotiorum*.

RnMBV2 was isolated from the same host species, *R. necatrix*, as RnMBV1, although the two infected host strains were collected from different prefectures in Japan. RnMBV2 would be classified as a distinct species from RnMBV1, based on phylogenetic distances, identity of amino acid sequence of hypothetical proteins and virulence to the host. Phylogenetic analysis based on the

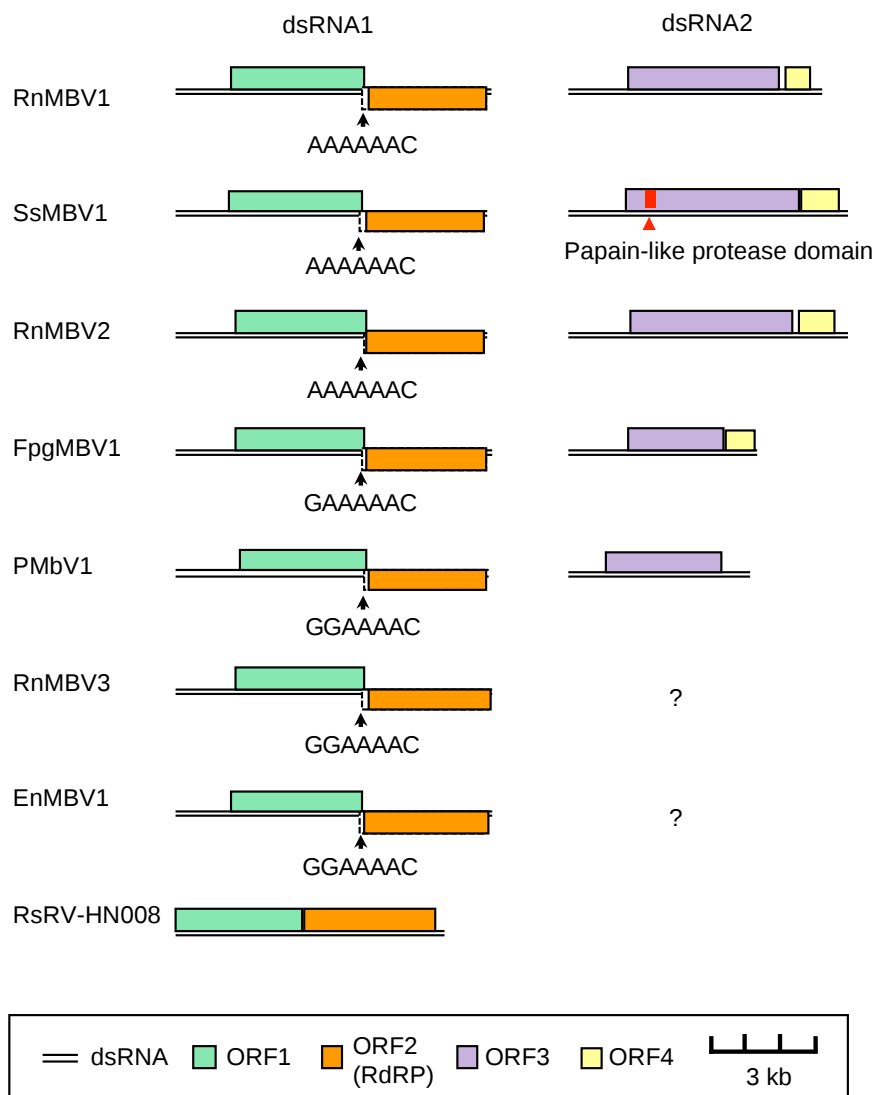


Fig. 7 Genome organization of megabirnavirus and related, unclassified viruses. Abbreviations of virus names corresponds to viruses listed in Tables 1 and 2. Doubled lines indicate dsRNA. ORF1, ORF2, ORF3 and ORF4 are represented by green, orange, purple and yellow boxes, respectively. The position and sequence of putative slippery heptamer adjacent to the stop codon of ORF1 is indicated with an arrow. The position of papain-like protease domain in ORF3 of SsMBV1 is indicated with a red box.

deduced amino acid sequence of RdRP and CP indicates that RnMBV2 is more closely related to SsMBV1 than to RnMBV1 (Fig. 6). A single infection by RnMBV2 causes no obvious effects on *R. necatrix*, which is in contrast to infection by RnMBV1. However, co-infection by RnMBV2 and a partitivirus, *Rosellinia necatrix* partitivirus 1 (RnPV1), induces apparent growth inhibition of *R. necatrix*, although single-infections involving either virus does not. The enhanced RnPV1 replication likely underlies this synergistic interaction between RnMBV2 and RnPV1.

In addition to SsMBV1 and RnMBV2, several unclassified viruses closely related to megabirnaviruses were found in diverse ascomycetous and basidiomycetous fungi. These include *Fusarium pseudograminearum* megabirnavirus 1 (FpgMBV1), *Pleosporales* megabirnavirus 1 (PmbV1), *Rosellinia necatrix* megabirnavirus 3 (RnMBV3), *Entoleuca* megabirnavirus 1 (EnMBV1), *Rhizoctonia solani* megabirnavirus 1 (RsMBV1) and *Rhizoctonia solani* RNA virus HN008 (RsRV-HN008) (Table 2, Fig. 6). FpgMBV1, PmbV1 and RsRV-HN008 have been detected in single infections of the respective original host fungi, while RnMBV3, EnMBV1 and RsMBV1 have been detected in mixed infections with other viruses by next-generation sequencing-based virome analyses. Thus, only the sequence of dsRNA1 (RnMBV3 and EnMBV1) or the partial sequence of RdRP (RsMBV1) have been reported, and the complete genome organizations of these three viruses remain unknown. The dsRNA1 sequences of FpgMBV1, PmbV1, RnMBV3 and EnMBV1 are similar to that of RnMBV1 in several aspects: (1) dsRNA size (8.8–9.0 kbp), (2) 5'-UTR length (> 1.6 kbp), (3) location of ORF2 (RdRP encoding gene) in -1 reading frame, relative to ORF1 and (4) presence of atypical shifty heptamer (5'-G RAA AAC-3', here R represents A or G) resembling a typical one (5'-A AAA AAC-3') immediately upstream of the stop codon of ORF1 (Fig. 7). The dsRNA2 sequence of FpgMBV1 and PmbV1 (5.1–5.3 kbp) is much shorter than dsRNA2s of RnMBV1, SsMBV1 and RnMBV2 (7.2–8.0 kbp). RsRV-HN008 is distinct from these other megabirna-related viruses; RsRV-HN008 has a monosegmented genome with a short 5'-UTR (39 bp) and two ORFs in the same reading frame (Fig. 7). All of these viruses, including RsRV-HN008, are placed on the same phylogenetic clade based on the identity of the deduced amino acid sequence of CP or P1, as well as RdRP (Fig. 6). Further consideration is required for the cogent classification of these viruses and the setting of criteria for the family *Megabirnaviridae*.

Future Perspectives

Functions of Megabirna-P3 and -P4

It has been confirmed that P3s, but not P4s, of RnMBV1 are expressed in infected fungal tissues; however, elucidation of the functions of P3 and P4 is a future research challenge. Artificial virion transfection of *R. necatrix* by RnMBV1 occasionally leads to the loss of dsRNA2, which simultaneously triggers the emergence of rearranged segments derived from dsRNA1. RnMBV1 mutants lacking dsRNA2 are capable of replicating and packaging, but showed impaired replication and attenuated hypovirulence to fungal hosts in comparison with wild type RnMBV1. Furthermore, dsRNA2-lacking RnMBV1 mutants appear to require rearranged dsRNA1 derivatives, in addition to intact dsRNA1, to stably infect the host fungus. Similarly, SsMBV1 mutants lacking dsRNA2 emerge by transfection of wild type SsMBV1 into *S. sclerotiorum*. SsMBV1 mutants without dsRNA2 also accumulate lower amounts of dsRNA1 compared to wild type. Thus, dsRNA2 of megabirnaviruses seems to be involved in efficient replication, translation and/or stability of dsRNA1. DsRNA2 is also likely to contribute to increased virulence to hosts. It remains to be elucidated whether megabirna-P3 and -P4 or the dsRNA itself plays a role in these contexts.

Further Reading

- Chiba, S., Salaipeth, L., Lin, Y., *et al.*, 2009. A novel bipartite double-stranded RNA mycovirus from the white root rot fungus *Rosellinia necatrix*: Molecular and biological characterization, taxonomic considerations, and potential for biological control. *Journal of Virology* 83, 12801–12812.
- Kanematsu, S., Shimizu, T., Salaipeth, L., *et al.*, 2014. Genome rearrangement of a mycovirus *Rosellinia necatrix* megabirnavirus 1 affecting its ability to attenuate virulence of the host fungus. *Virology* 450–451, 308–315.
- Miyazaki, N., Salaipeth, L., Kanematsu, S., *et al.*, 2015. Megabirnavirus structure reveals a putative 120-subunit capsid formed by asymmetrical dimers with distinctive large protrusions. *Journal of General Virology* 96, 2435–2441.
- Salaipeth, L., Chiba, S., Eusebio-Cope, A., *et al.*, 2014. Biological properties and expression strategy of *rosellinia necatrix* megabirnavirus 1 analysed in an experimental host, *Cryphonectria parasitica*. *Journal of General Virology* 95, 740–750.
- Sato, Y., Miyazaki, N., Kanematsu, S., *et al.*, 2019. ICTV virus taxonomy profile: *Megabirnaviridae*. *Journal of General Virology* 100, 1269–1270.
- Wang, M., Wang, Y., Sun, X., *et al.*, 2015. Characterization of a novel megabirnavirus from *Sclerotinia sclerotiorum* reveals horizontal gene transfer from single-stranded RNA virus to double-stranded RNA virus. *Journal of Virology* 89, 8567–8579.

Mitoviruses (*Mitoviridae*)

Bradley I Hillman and Alanna B Cohen, Rutgers University, New Brunswick, NJ, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Anastomosis Fusion between hyphal branches allowing for exchange of cytoplasm, ions, and organelles, and the most common means of horizontal transmission of mycoviruses.

Conidia Asexual, non-motile fungal spores.

Haplotype Group of alleles inherited together. In the case of haploid fungi, it constitutes the genetic makeup of the individual.

Non-retroviral endogenous RNA (NERVE) Genetic elements comprised of fragments or complete genomes

derived from non-retroviral RNA viruses found embedded in eukaryotic genomes.

Protoplast A cell that has its cell wall removed, allowing for membrane fusion and permeabilization.

RNA silencing The negative regulation or inhibition of gene expression by non-coding RNA elements. Also called RNA interference (RNAi).

Vegetative incompatibility Genetic differences between fusing vegetative hyphae of fungi that result in unviable heterokaryons and prevent horizontal mycovirus transmission.

The genus *Mitovirus* contains arguably the simplest of all viruses known: an RNA molecule often of less than 3 kb encodes only RNA-dependent RNA polymerase (RdRp) that serves to direct replication of the simple genome. They are called mitoviruses because they reside in mitochondria. In common with several other fungal viruses, mitoviruses have no capsid protein, but rather consist of RNA, likely encapsulated in lipid vesicles together with virally-encoded RNA-dependent RNA polymerase. Mitoviruses are now known to be abundant in fungi – possibly the most common of all fungal viruses – and they are now suspected to be components of the plant virome as well.

Most mitoviruses have been identified through double-stranded (ds) RNA screening of fungal cultures, a commonly used method to screen fungal culture collections for virus presence. Mitovirus dsRNA is usually relatively abundant in infected cultures, consistent with either a dsRNA or positive sense (+) RNA genome. Free + RNA is more abundant in cultures than free negative-sense (−) RNA, consistent with a + RNA mitovirus genome. Phylogenetic relatedness of mitovirus RdRp coding sequences to + RNA viruses rather than − RNA viruses or dsRNA viruses supports their classification as + RNA viruses.

Although they were first characterized in the mid 1990s, relatively little progress has been made on details of their replication and life cycles. This is in large part because of the lack of good infectivity systems. As mitoviruses have no protein-containing particles to deliver infectious nucleic acid and delivery into mitochondria is required to initiate infection, transmission has been accomplished to date only vertically via asexual or sexual spores, or horizontally by hyphal anastomosis. No one yet has developed a robust reverse genetics system – that is, infectious cDNA clones – for any mitovirus. Recent demonstration of mitovirus transmission to marked heterologous species, genera, and families opens the door for deeper examination of host range and phenotypic effects of a single mitovirus in different host backgrounds.

Genome Structure

There is relatively little variation in mitovirus genome size and structure (Fig. 1). The largest mitovirus genome sequenced to date is ~5.0 kb from the basidiomycete *Heterobasidion annosum*; the smallest is ~2.2 kb from the ascomycete *Sclerotinia sclerotiorum*. In

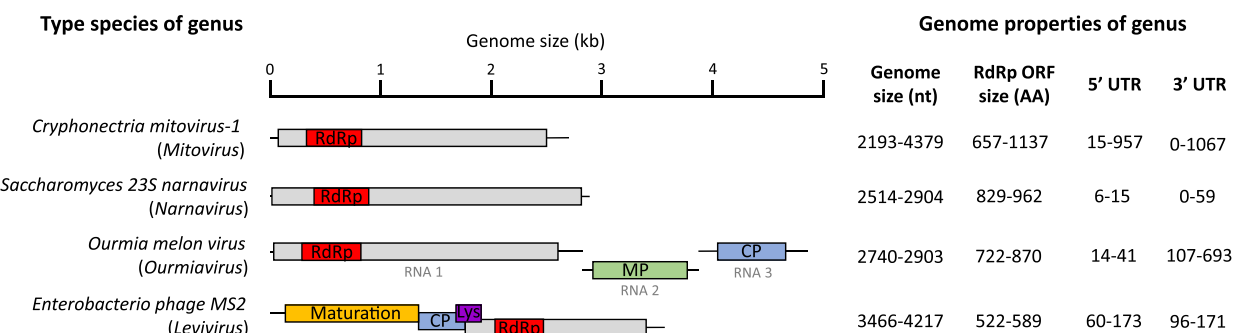


Fig. 1 Genome maps drawn to scale of representative viruses belonging to the genus *Mitovirus* and close relatives. Color-coded regions represent conserved domains including: RdRp; RNA-dependent RNA polymerase (red), MP; movement protein (green), CP; capsid protein (blue), maturation protein (orange), and Lys; lysis protein (purple). Within each genus, there is variability with respect to genome size, RdRp protein size, and number of nucleotides in the non-translated sequences at 5'- and 3' termini. Size ranges of each of these elements within members of each genus are provided.

all cases, a single ORF encodes the RdRp and no other known proteins. A hallmark of fungal mitovirus RNAs is the presence of UGA codons distributed throughout the ORF in varying numbers encoding tryptophan (Try) rather than serving as a terminator of translation (see below). There is no evidence for post-translational processing of mitovirus-encoded proteins and no evidence for the use of additional shorter-than-full-length subgenomic messenger (m) RNAs for mitovirus gene expression. Non-translated sequences at the 5'-end of the RNA vary from 15 to 957 nucleotides; non-translated sequences at the 3'-ends vary from 0 to 1788 nucleotides. Specific host and viral factors required for mitochondrial genome replication and expression have not been investigated in any depth primarily because of the lack of infectious viral cDNA clones that can be manipulated and because of the extreme difficulty associated with manipulating host mitochondrial gene expression.

Accessory RNAs Associated With Mitovirus Infections

In addition to the single, replication-competent RNA of mitoviruses, additional accessory RNA molecules may be found as well. Accessory RNAs that may be associated with RNA virus infections include satellite RNAs, defective or defective-interfering RNAs, or satellite/defective RNA hybrids. Satellite RNAs (S-RNAs) are unrelated by nucleotide sequence to the parent virus, except possibly for 5'- or 3'-terminal nucleotides; defective RNAs (D-RNAs), also called defective genomes, are incomplete deletion mutants derived from the parent RNA; defective-interfering RNAs (DI-RNAs) are defective RNAs that have been demonstrated to interfere with parent virus replication or function. S-RNAs, D-RNAs, and DI-RNAs are all common in fungal RNA virus infections, and all have been identified in mitovirus infections.

A heterogeneous S-RNA population was associated with infection of the mitovirus *Ophiostoma novo-ulmi* mitovirus 3A (OnuMV-3A) in the fungus *Sclerotinia homoeocarpa*, but there was no evidence that the satellite affected accumulation of the parent virus replication. Similarly, D-RNAs were identified in *O. novo-ulmi* strain Ld infected with multiple mitoviruses, but there was no evidence for interference of parent virus replication. In the *Botrytis cinerea* isolate CanBc-1c-78, the mitovirus BcMV-1 was found to be associated with a deletion mutant of the parent viral genome that fits the criteria of a DI-RNA by reducing accumulation of that parent viral RNA. As with other mitovirus investigations, studies on S-RNAs, D-RNAs, and DI-RNAs are hampered by the lack of robust infectivity systems.

Phenotypic Effects of Mitovirus Infection

Many fungal viruses have little measurable effect on fungal colony morphology, growth rate in culture, or virulence, while others may have profound effects. Mitoviruses reflect this spectrum of biological consequences as well, some resulting in symptomless infections but others causing measurable and perhaps severe pathology. The first mitoviruses were identified based on their effects on virulence of the Dutch elm disease fungus in natural settings and colony phenotypes of *O. novo-ulmi* in culture, but that system was particularly difficult to tease apart from the standpoint of satisfying Koch's postulates for individual mitovirus elements because they were numerous, coinfecting, and resulted in different phenotypes in the fungus. The effects of the individual mitoviruses were difficult to quantify in the absence of traditional virology methods such as particle isolation and inoculation to uninfected individuals. In contrast, the first mitovirus identified based on genetic and molecular properties in *C. parasitica* was a single element that had only a slight measurable effect on phenotype, so the challenge was to develop isogenic, virus-free fungal isolates and determine whether the minor changes to growth rate in culture and fungal virulence were caused by the virus itself.

Only recently did Suzuki and colleagues develop a system to separate unambiguously mitovirus effects from host nuclear and mitochondrial backgrounds. Hygromycin-resistant, virus-free isolates were used as recipients in protoplast fusions with virus-containing, hygromycin-sensitive isolates. Following repeated anastomosis and single conidial isolation, resulting mitovirus-containing isolates bearing the nuclear and mitochondrial genotypes of the original recipient isolate were obtained and maintained in continuous culture.

Several mitoviruses appear to have profound effects on fungal phenotype, including virulence on plant hosts. In the ascomycete pathogen of many field crops, *Chalara elegans* (= *Thielaviopsis basicola*), reduced fungal growth, reduced melanin production, and reduced virulence have been associated with mitovirus presence. Furthermore, smaller and fewer mitochondria were associated with mitovirus-bearing isolates compared to their mitovirus-free counterparts. Similarly, a mitovirus from the broad host range plant pathogen *Botrytis cinerea* was shown to have deleterious effects on its fungal host and on the infected mitochondria themselves. In each of these two cases, the mitovirus was not completely eliminated from the fungal host, so conclusions of direct causality of the virus and quantifying disease effects were tempered.

Mixed infection of a single fungal thallus by multiple mitoviruses has been demonstrated in several instances. In *O. novo-ulmi*, single conidial isolates of debilitated, multiply-infected isolates were used to help identify mitoviruses that affected fungal phenotype and those that did not. Multiple mitoviruses have also been shown to infect single isolates of *Sclerotinia* and *Botrytis*. Still unknown is whether multiple mitoviruses infect and replicate together in single mitochondria, or whether instead each mitochondrion within the population of mitochondria in a fungal thallus is infected with a single mitovirus. As noted for other mitovirus infectivity studies, experiments to address this question are challenging.

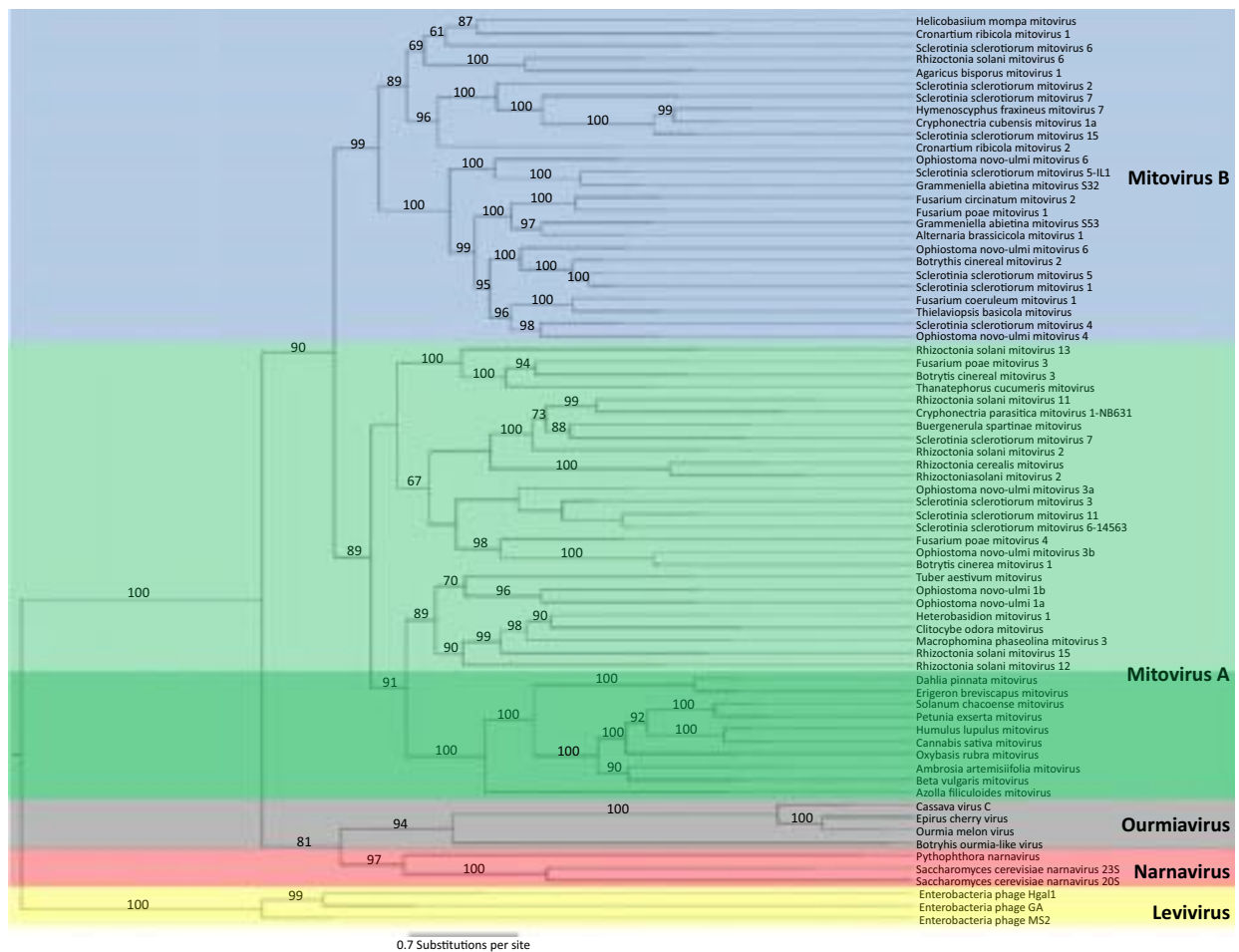


Fig. 2 Maximum likelihood (ML) phylogenetic tree based on multiple sequence alignment of full-length RNA-dependent RNA polymerase (RdRp) protein sequences from 130 mitoviruses. Sequences were aligned with MUSCLE and implemented in IQ-TREE with model selection parameters. A VT + F + I + G4 model of evolution was selected based on the Akaike information criterion (AIC) and Bayesian information criterion (BIC). Bootstrap support above 60% is displayed above branches, calculated from 2000 ultrafast bootstrap replicates. Major clades are colored to denote the following groups: genus *Levivirus* (yellow), genus *Narnavirus* (red), genus *Ourmiavirus* (gray), genus *Mitovirus* clade B (blue), and genus *Mitovirus* clade A (green). Assembled mitovirus RNA sequences from plants form a monophyletic clade within mitovirus clade A and are denoted by a darker shade of green.

Phylogenetic Relationships

Phylogenetic reconstruction of 130 mitovirus sequences reveal a well-supported monophyletic group containing two major clades; a smaller subset of those isolates is provided in Fig. 2. Clade A contains mitovirus sequences determined from a range of Ascomycete and Basidiomycete fungi as well as all identified plant mitoviruses, which form an independent monophyletic group. Mitovirus clade B is comprised of virus isolates from similar fungi, but no plant mitoviruses. High bootstrap values support the separation of the mitovirus clade from its sister clade containing the narnaviruses and ourmiaviruses.

It is notable that several fungal taxa harbor mitoviruses from both clades A and B, and that individual mitovirus species from a single fungal species are distributed broadly within those clades. These observations suggest that mitoviruses have entered fungal taxa such as *Ophiostoma*, *Sclerotinia*, and *Botrytis* independently multiple times.

Taxonomy and Nomenclature

Under current ICTV taxonomy, the family *Narnaviridae* contains two genera: the genus *Narnavirus* and the genus *Mitovirus*. Continued reassessment of the overall taxonomy of these and related viruses is needed, as members of the genus *Narnavirus* are more closely related to the plant-infecting viruses in the genus *Ourmiavirus*, in the recently approved family *Botourmiaviridae*, than they are to the mitoviruses. Thus, it is likely that the mitoviruses and narnaviruses will eventually be separated into different families.

The genus *Mitovirus* currently contains all members discussed in this article. At this time, other than the association with different host taxa, there are no solid biological criteria on which to subdivide the taxon. Based on the strong evidence for at least

two distinct mitovirus clades, one of which includes all known presumed plant-associated mitoviruses and related sequences, it might be useful to subdivide this broad taxon into smaller taxa. As the numbers of characterized mitoviruses and putative mitovirus sequences accumulates, it is expected that such a taxonomic proposal would soon be entertained by the ICTV.

Transmission

Mitovirus transmission is not fundamentally unlike transmission of other fungal viruses, but their mitochondrial localization leads to some important differences compared to cytoplasmic viruses. Like other fungal viruses, mitoviruses depend on hyphal anastomosis for horizontal transmission. As such, vegetative incompatibility (VIC) genes and the associated VIC phenotype play a fundamental role in transmission from one fungal thallus to another in natural settings. Mitoviruses are efficiently transmitted upon anastomosis of vegetatively compatible isolates in culture, but details of transmission are only recently coming to light. Considering the mechanism of transmission as it pertains to the mitochondria themselves, two basic possibilities present themselves: mitovirus-containing mitochondria from an infected isolate could enter an uninfected isolate after hyphal anastomosis and then fuse with virus-free mitochondria in the previously uninfected isolate, resulting in delivery of virus by mitochondrial fusion; or the virus-containing mitochondria could enter the uninfected strain and simply take over, eventually replacing the population of uninfected mitochondria with infected ones. In the former model, the resulting mitochondrial haplotype of the newly infected strain would either be the same as it was previously, or would be a hybrid of the two mitochondrial haplotypes; in the latter model, the mitochondrial haplotype of the newly infected strain would be that of the infecting strain. Evidence to date strongly supports the first model: incoming mitochondria deliver virus to resident mitochondria and do not themselves replace the resident population. This has recently allowed for mitovirus transmission among different species of the same genus, among genera within a family, and even across family lines. These transmission experiments have also shown that a mitovirus from one fungal host species may or may not be stable in a new host. More such experiments are needed to better understand factors that regulate mitovirus/host stability.

Vertical transmission of mitoviruses occurs both via sexual and asexual spores. Most mitoviruses have been characterized in ascomycetes, and their transmission has been more thoroughly characterized in those taxa. Transmission through asexual spores such as conidia in ascomycetes is often very efficient, at or near 100% in single conidial isolates derived from an infected isolate. Experimental transmission through sexual ascospores in ascomycetes has been shown to follow a somewhat predictable pattern. In *C. parasitica*, there are two mating types, and mitochondria have been shown to be inherited maternally: all mitochondria in ascospore progeny are derived from the female parent in sexual crosses. This pattern is common, but not universal in fungi. The *C. parasitica* mitovirus CpMV1-NB631 was shown to be transmitted to ~50% of ascospores derived from matings in which the female isolate was infected with CpMV1-NB631, but no transmission occurred when the male isolate in the matings was the infected partner. While that general pattern of virus inheritance follows the pattern predicted from mitochondrial inheritance, it is unclear why only approximately 50% rather than 100% of progeny in these experiments were infected.

Engineering Mitoviruses for Infectivity

Reverse genetics can be effected for most positive strand RNA viruses through the use of infectious cDNA clones. Usually this is a fairly straightforward process: a full-length cDNA clone of the viral RNA is produced in a manner allowing for in vitro transcription of positive-sense RNA representing a replica of the viral RNA itself. The transcript is infectious when introduced into cells. The cDNA clones can then be manipulated through traditional molecular methods to examine functions of specific sequences. To date, no robust reverse genetics system has been reported for mitoviruses. The fundamental problem in developing a true reverse genetics system is delivery of viral-length RNA to mitochondria to initiate the infection process. It is difficult to know how many attempts at such infections have been made – we know of several – but none have yet been reported as successful.

Another approach to examining mitovirus replication involves attempting replication and expression not in mitochondria but in the cytosol, as with most other positive-sense RNA viruses. For mitoviruses, this involves mutating any of the UGA codons to UGG, thereby allowing them to direct cytoplasmic incorporation of the amino acid tryptophan (Try) rather than translation termination. This approach would effectively turn a mitovirus into a narnavirus, which has the same fundamentally simple genome structure, encoding only RdRp, but is naturally localized to cytoplasm rather than mitochondria. Limited success in this approach has been reported, but we are unaware of peer-reviewed journal articles summarizing such studies. Such experiments are of interest from the standpoint of basic virology and examination of factors differentiating narnaviruses and mitoviruses, but they are not replacements for studies examining mitovirus RNA replication in mitochondria.

Host Defense

Within fungi, a primary host defense mechanism against virus attack is thought to be vegetative incompatibility systems, which limit intraspecific hyphal fusion and accordingly limit horizontal virus transmission. This barrier is well-understood in some fungi, including *C. parasitica*, and has been demonstrated to restrict mitovirus transmission experimentally in that fungus and several other fungi. Presumably it functions in nature in fungi with VIC systems as well.

Active, nucleic acid-based antiviral defense systems in plants and fungi are similar to each other, but differ in details and complexity. Both plants and fungi have RNA silencing-based defense systems. Viral suppressors of silencing encoded by related plant and fungal viruses, including the fungal-infecting members of the family *Hypoviridae* and the plant-infecting *Potyviridae*, have been characterized in detail, and the fungal virus hypovirus silencing suppressor, p29, has been shown to function when expressed in plants. There is currently no evidence that mitoviruses encode suppressors of RNA silencing, and comparison of CpMV1 accumulation among various RNA silencing-deficient and competent strains indicated that the virus was unaffected by RNA silencing. CpMV1 did not induce expression of either the dicer-like protein DCL2 or the Argonaut-like protein AGL2, the two RNAi-associated proteins that are known to be components of the RNA virus defense response pathway in fungi. Furthermore, there was no difference in susceptibility of these strains compared to the wild-type strain, as measured by replication level of CpMV1 in fungal strains deleted in the *dcl2* or *agl2* genes. Studying the interactions between fungal – or plant – RNA defense systems and mitoviruses will help answer important fundamental questions about mitovirus biology.

Codon Usage and Implications for Mitovirus Biology and Evolution

Mitoviruses were initially described as mitochondrial elements based on their biology in the Dutch elm disease fungal system, *Ophiostoma novo-ulmi*, and their molecular, genetic, and subcellular properties in the chestnut blight system, *Cryphonectria parasitica*. Although, few comprehensive studies have been done to date in the many mitovirus systems described at the primary sequence level, deep sequencing and subsequent mining of genomic and transcriptomic data have resulted in important advances in our understanding of the biology and evolutionary history of mitoviruses. One of the hallmarks of mitovirus genome structure is their use of the mitochondrial genetic code for translation of their only gene product, the viral RdRp. The observation in the *C. parasitica* mitovirus CpMV1-NB631 that the genome contains numerous UGA codons in an otherwise open reading frame (ORF) encoding a putative RdRp led to the hypothesis that the element is obligately mitochondrial. UGA is a termination codon in the cytosol but encodes tryptophan in most fungal mitochondria, and therefore this CpMV1-NB631 element was predicted to encode a functional RdRp in mitochondria, but have no capacity for translation in the cytosol. As mitovirus sequences accumulated, the universal presence of UGA codons in predicted RdRp ORFs became evident as a general property of these elements.

The accumulation in databases of many mitovirus sequences allowed for systematic analysis of Try-encoding UGA vs. UGG codons in the context of fungal host UGA vs. UGG codon use. These analyses showed that mitovirus UGA codon use mirrored that of the host fungus: that is, fungal taxa that rarely used UGA to encode Try tended to host mitoviruses with few UGA codons.

Origin and Evolution

The phylogenetic tree of plant and fungal mitovirus sequences does not exclude the possibility that fungal mitoviruses derive from their plant counterparts. However, the likely ancient origin of eukaryotic mitoviruses, coupled with the earlier appearance on earth of fungi rather than plants argues for fungi as the earlier mitovirus host. Possible scenarios to envision include entry of a plant mitovirus into a closely associated sympatric fungus, such as an endophyte or mycorrhizal fungus, and subsequent expansion of fungal mitovirus host range and adaptation to fungal mitochondrial codon preferences – i.e., progressive gain of UGA codons within coding sequences – following heterologous fungal anastomosis events and associated mitochondrial fusion events. It is also easy to envision this scenario moving in the other direction: that a fungal mitovirus devoid of UGA codons within the coding sequence similarly entered a plant mitochondrion, took up residence, and expanded from there into other plants.

Perhaps more than other viruses, the extreme simplicity of mitovirus genomes raises questions about how close they are to progenitors of RNA viruses. Recent deep and comprehensive phylogenetic analysis supports the idea that mitoviruses are ancient in origin, with the intriguing possibility that progenitors of contemporary mitoviruses entered the eukaryotic lineage along with the bacterial progenitors of the mitochondrial endosymbiont, possibly by the loss of a capsid protein gene from a progenitor bacterial levivirus. Evolution of mitoviruses within fungi would then parallel evolution of fungal (and plant) mitochondria. Further downstream evolution of viruses in the lineage represented by mitoviruses would presumably involve escape from mitochondria and then either (1) adaptation to a similarly cell-bound but non-mitochondrial, cytosolic lifestyle in the case of narnaviruses and possibly fungal botourmiaviruses, or (2) acquisition of simple capsid protein and movement protein genes and escape from the cell in the case of plant ourmiaviruses.

Plant Mitoviruses

Signature sequences of mitoviruses were among the first non-retroviral endogenous RNA (NERVE) sequences identified in plants, described initially in the late 1990s. Mitovirus sequences have been identified in both mitochondrial and nuclear DNA genomes of plants. Their initial entry and presumably subsequent integration into plant genomes appears to have been an ancient event, possibly dating from the early evolution of clubmosses nearly a half billion years ago. The extent to which mitoviruses have been active members of the plant virome remains a matter of active investigation that will be explored further as broader and deeper genomic coverage of plant phyla becomes available. Some studies suggest that mitoviruses may have entered into plant lineages

multiple times, possibly being lost from a given lineage and then regained through subsequent mitochondrial exchange. Recent evidence based on mining of transcriptome data strongly suggests that mitoviruses are active, contemporary residents of plants and that they form a monophyletic cluster embedded within one of the two major clades of fungal mitovirus elements (see Fig. 2). Furthermore, analysis of plant mitovirus-related NERVE sequences also supports the finding that these elements form a monophyletic group that clusters with the putative contemporary plant mitoviruses discovered from transcriptome data. Unlike fungal mitochondria, UGA codons in plant mitochondria encode a translational stop codon rather than tryptophan; thus, putative plant mitoviruses would not be expected to contain any UGA codons in ORFs, and this is indeed the case.

As mitoviruses are now understood to be a component of the plant virome, we anticipate that other studies will investigate their relative importance in plants. A key question is whether they are cryptic in plants or whether they have a measurable and significant effect on plant development and biology. An interesting finding from associative plant transcriptome studies suggested higher mitovirus expression levels in flowers than other tissues, but these studies were not directed toward asking questions about mitovirus biology. Experimental studies with isogenic plants, similar to those that have been performed in fungi, will help address the possibility. Many of the same types of studies that have been performed in fungal systems are possible in plant systems, and many of the same specific questions will be addressed, for example: is there a measurable effect on plant phenotype; can horizontal transmission be demonstrated and what are mechanistic details; how do RNA defense systems impact plant mitoviruses?

While genome and transcriptome data mining are very valuable tools for mitovirus discovery and analysis, examination of archival dsRNA studies that were performed for discovery of unknown viruses in plants may also be a useful approach. As it is with fungi, most plant viruses contain RNA genomes; in the case of plants, a preponderance of viruses contain positive-sense RNA genomes, and thus replicative form dsRNA accumulates to visually detectable levels in most. Initial discovery of mitoviruses in fungi, and subsequent discovery of most other fungal mitoviruses, has come through screening of samples from culture collections for dsRNA presence. Mitovirus-infected plants might be expected usually to contain a single signature genome-size dsRNA of ~2.5–3 kb, identifiable in archival and published gel photographs, but evidence to date suggests that plant mitovirus dsRNA accumulation is minimal, so such data mining might not lead to further plant mitovirus discovery.

Further Reading

- Bruenn, J.A., Warner, B.E., Yerramsetty, P., 2015. Widespread mitovirus sequences in plant genomes. *PeerJ* 3, e876.
- Ghabrial, S.A., Suzuki, N., 2009. Viruses of plant pathogenic fungi. *Annual Review of Phytopathology* 47, 353–384.
- Hillman, B.I., Cai, G., 2013. The family *Narnaviridae*: Simplest of RNA viruses. *Advances in Virus Research* 86, 149–176.
- Hong, Y., Dover, S.L., Cole, T.E., Brasier, C.M., Buck, K.W., 1999. Multiple mitochondrial viruses in an isolate of the Dutch elm disease fungus, *Ophiostoma novo-ulmi*. *Virology* 258, 118–127.
- Nerva, L., Forgia, M., Ciuffo, M., et al., 2019. The mycovirome of a fungal collection from the sea cucumber *Holothuria polii*. *Virus Research* 273, 197737.
- Nibert, M.L., 2017. Mitovirus UGA(Trp) codon usage parallels that of host mitochondria. *Virology* 507, 96–100.
- Nibert, M.L., Vong, M., Fugate, K.K., Debat, H.J., 2018. Evidence for contemporary plant mitoviruses. *Virology* 518, 14–24.
- Polashock, J.J., Bedker, P.J., Hillman, B.I., 1997. Movement of a small mitochondrial double-stranded RNA element of *Cryphonectria parasitica*: Ascospore inheritance and implications for mitochondrial recombination. *Molecular Genetics and Genomics* 256, 566–571.
- Polashock, J.J., Hillman, B.I., 1994. A small mitochondrial double-stranded (ds) RNA element associated with a hypovirulent strain of the chestnut blight fungus and ancestrally related to yeast cytoplasmic T and W dsRNAs. *Proceedings of the National Academy of Sciences of the United States of America* 91, 8680–8684.
- Shahi, S., Eusebio-Cope, A., Kondo, H., Hillman, B.I., Suzuki, N., 2019. Investigation of host range and host defense against a mitochondrially replicating mitovirus. *Journal of Virology* 93, e01503–e01518.
- Vainio, E.J., 2019. Mitoviruses in the conifer root rot pathogens *Heterobasidion annosum* and *H. parviporum*. *Virus Research* 271, 19768.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., et al., 2018. Origins and evolution of the global RNA virome. *mBio* 9, e02329-18. doi:10.1128/mBio.02329-18.
- Wu, M., Zhang, L., Li, G., Jiang, D., Ghabrial, S.A., 2010. Genome characterization of a debilitation-associated mitovirus infecting the phytopathogenic fungus *Botrytis cinerea*. *Virology* 406, 117–126.

Mycoreoviruses (*Reoviridae*)

Bradley I Hillman and Alanna B Cohen, Rutgers University, New Brunswick, NJ, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of B.I. Hillman, Mycoreoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00569-0.

Glossary

Anastomosis Fusion between hyphal branches allowing for exchange of cytoplasm, ions, and organelles, and the most common means of horizontal transmission of mycoviruses.

Ascospores Sexual spores of an ascomycete.

Conidia Asexual, non-motile fungal spores.

Haplotype Group of alleles inherited together. In the case of haploid fungi, it constitutes the genetic makeup of the individual.

Protoplast A cell that has its cell wall removed, allowing for membrane fusion.

Reassortment Exchange of genome segments resulting from co-infection of two or more strains of a multi-segmented virus such as a reovirus.

RNA silencing The negative regulation or inhibition of gene expression by non-coding RNA elements. Also called RNA interference (RNAi).

Vegetative incompatibility Genetic differences between fusing vegetative hyphae of fungi that result in unviable heterokaryons and prevent horizontal mycovirus transmission.

The 9th Report of the International Committee for the Taxonomy of Viruses (ICTV) lists 15 genera of viruses in the family *Reoviridae* divided into two subfamilies, *Sedoreovirinae* and *Spinareovirinae*, that infect mammals, invertebrates, plants, protists, and fungi. Members of many of the reovirus genera in the 9th Report replicate in organisms representing more than one kingdom or phylum: for example, all plant reoviruses replicate in and persistently infect their insect vectors, and similarly many of the mammalian reoviruses replicate in their invertebrate vectors.

The Genus *Mycoreovirus* comprises three fungus-infecting reovirus species, *Mycoreovirus 1*, *Mycoreovirus 2*, and *Mycoreovirus 3*. The exemplar strains of the first two species, mycoreovirus 1 (MyRV1) and MyRV2 were isolated from the filamentous ascomycete fungus *Cryphonectria parasitica*, and that of the last species is from the soilborne fungus *Rosellinia necatrix*, also an ascomycete but representing a different fungal order. A fourth proposed species of the genus infecting the fungal plant pathogen *Sclerotinia sclerotiorum* has been described (here designated MyRV4). Another recently described fungal reovirus, also infecting *S. sclerotiorum*, falls phylogenetically outside of the established genus *Mycoreovirus*. That virus, *Sclerotinia sclerotiorum* reovirus 1 (SsReV1), will nevertheless be discussed in this article.

Mycoreoviruses often cause disease in their infected hosts, resulting in greatly reduced virulence, reduced growth in culture, reduced laccase accumulation, and reduced sporulation relative to wild-type, virus-free cultures of the same genetic background (Fig. 1). The mycoreoviruses are most closely related to members of the genus *Coltivirus* of tick-borne reoviruses (Fig. 2). Both Colorado tick fever virus (CTFV) and the closely related Eyach virus (EyaV) cause disease in mammals, and both viruses replicate in their arthropod vectors as well as in their mammalian hosts. A recently characterized coltivirus from bats was also found to replicate in mammalian cells. The coltiviruses are not well understood at the molecular level and so have shed only a little light on the study of mycoreoviruses. Coltiviruses are more closely related to the genus *Orthoreovirus*, which includes the common human pathogen mammalian reovirus (MRV), than to members of the other two genera of the family *Reoviridae* (*Orbivirus* and *Rotavirus*) that have been well studied at the structural and molecular levels.

Filamentous fungi are valuable subjects for investigation of eukaryotic viruses. This has been especially true for ascomycetes, which usually are haploid throughout their vegetative phases and thus readily amenable to tools of classical genetics and to relatively simple gene knockout and knockdown experiments. The plant pathogenic ascomycete fungi *Cryphonectria parasitica*, *Rosellinia necatrix*, and *S. sclerotiorum* have all been exceptional hosts for examination of fungal viruses: they cause important plant disease, are very stable in culture, showing a consistent morphology upon continued maintenance and subculture, are easily transformed and transfected, and can be examined through classical genetics. Because of their historical importance as plant pathogens and the potential for their control using natural or genetically engineered viruses, a large array of diverse viruses that cause stable morphological changes and/or changes in virulence of their fungal hosts have been identified and characterized in these fungi.

Structure-Function Relationships

Relatively little is known of the details of mycoreovirus structure, but all indications from electron micrographs of negatively stained virus particles and genome sequence analysis are consistent with the *Spinareovirinae* subfamily, best studied in the orthoreoviruses. The orthoreoviruses, typified by mammalian orthoreovirus (MRV), are distinguished from the members of the subfamily *Sedoreovirinae* such as orbiviruses and rotaviruses in that they have identifiable pentameric turrets on top of each

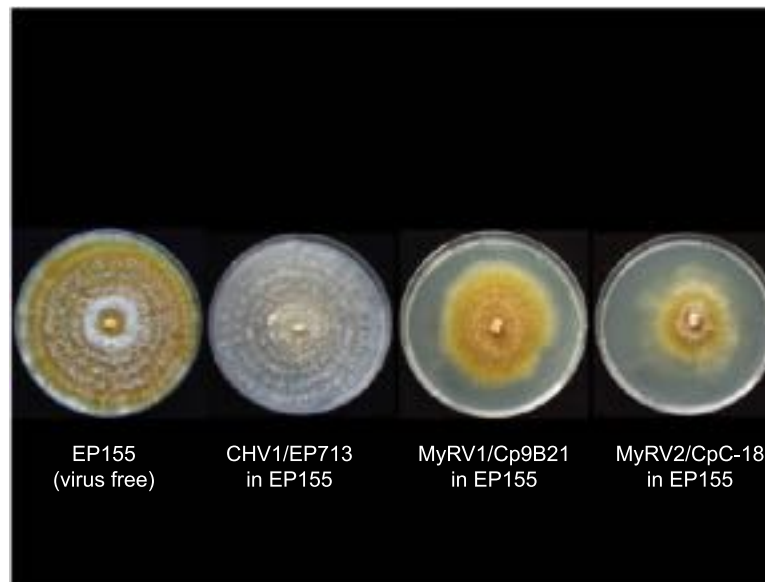


Fig. 1 Morphologies of isogenic *Cryphonectria parasitica* colonies infected with three mycoviruses. (a) Uninfected colony, strain EP155; (b) Strain EP155 infected with hypovirus CHV1/EP713; (c) Strain EP155 infected with mycoreovirus MyRV1/Cp9B21; (d) Strain EP155 infected with mycoreovirus MyRV2/CpC18.

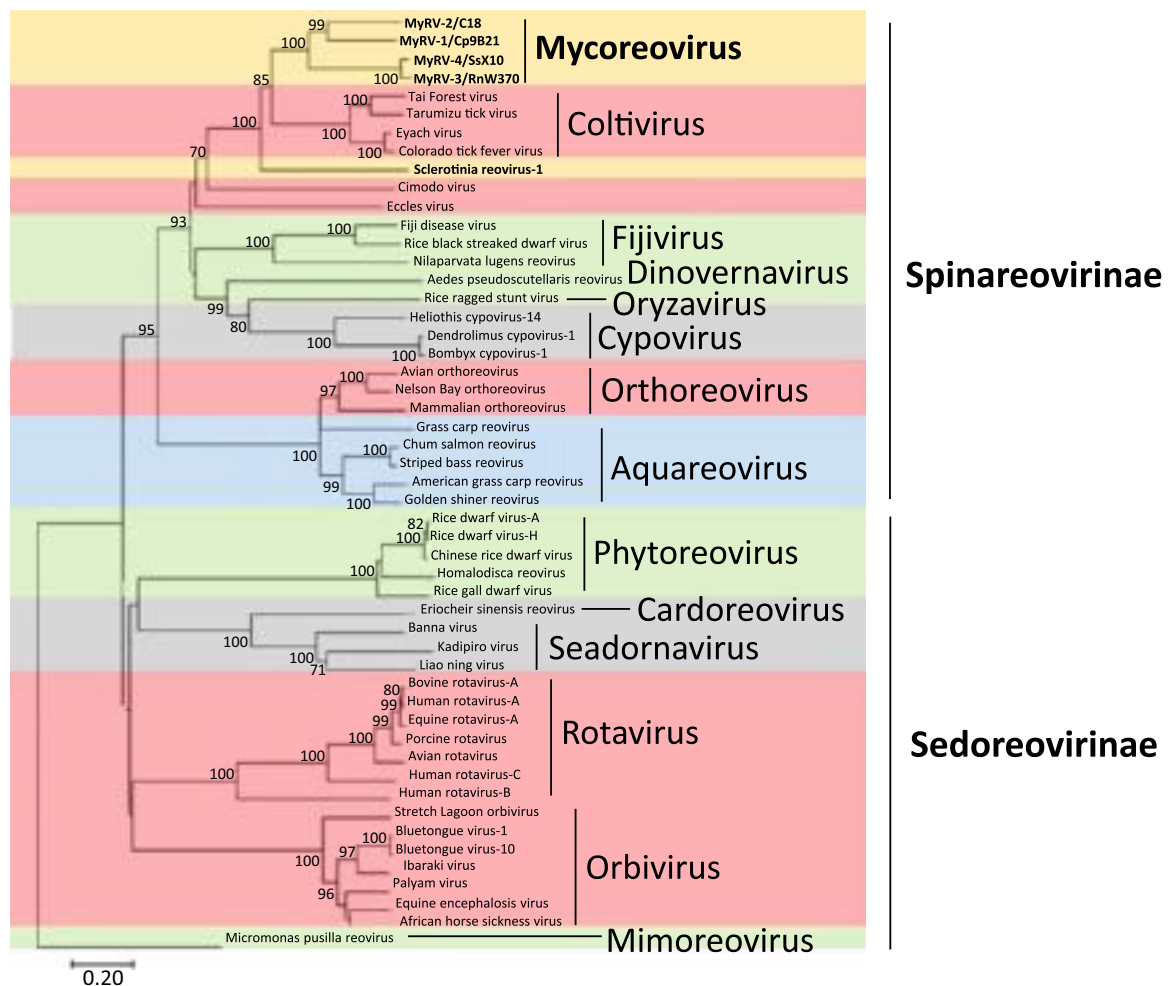


Fig. 2 Neighbor-joining (NJ) phylogeny of members of the family *Reoviridae* based on a multiple sequence alignment of full-length RdRp protein sequences. Amino acid sequences were aligned using MUSCLE. Phylogenetic tree was reconstructed using MEGA version 7 under the Poisson model with 1000 bootstrap replicates. Numbers above branches represent bootstrap support above 70%. Taxa that do not belong to a named clade represent unclassified viruses within Reoviridae. Taxa are colored according to host type: fungus (orange), mammal (red), invertebrate (gray), plant (green), and fish (blue).

fivefold axis of the core particle, and these have been demonstrated to be involved in capping of nascent mRNA. The protein that was experimentally demonstrated to be the mycoreovirus guanylyltransferase is homologous to turret proteins of orthoreoviruses that have the same function, supporting the grouping of mycoreoviruses within the *Spinareovirinae* subfamily of the family *Reoviridae*.

Movement of fungal viruses within the mycelium and during horizontal or vertical virus transmission is also relatively poorly understood. Viruses that have been investigated are generally found in hyphal tip cells and move with the growing mycelium; however, location of mycoreoviruses within the mycelium has not been investigated. The steps in the infection process of the mammalian orthoreovirus have been investigated in considerable detail. Virus entry into cells via vesicles occurs by receptor-mediated endocytosis, whereupon the m1 protein, which is myristoylated at its N-terminus and inserted into the inner membrane of the vesicle, cleaves autocatalytically, resulting in pore formation in the vesicle membrane and subsequent virus release into the cytoplasm. Both of the *C. parasitica* mycoreoviruses have homologs of the orthoreovirus m1 protein, with strong myristoylation signals at the amino termini of the corresponding proteins and putative sites for autocatalytic cleavage. This suggests that a similar mechanism of egress from vesicles is a component of the infection cycle of these reoviruses. Interestingly, although the *R. necatrix* mycoreovirus, MyRV3/RnW370, contains a homolog of the m1 protein, it does not contain a glycine residue at the penultimate position of the N-terminus of this deduced protein, and based on its sequence it has a very low predicted probability of myristoylation. It does, however, contain the N/P putative cleavage motif at a similar position and in a similar environment as those in the other viruses.

It is relatively easy to cure *C. parasitica*, *R. necatrix*, and *S. sclerotiorum* from their associated mycoreoviruses. In *C. parasitica*, single asexual spores are usually virus free, and in *R. necatrix*, cultures initiated from excised hyphal tips (the terminal 2–8 cells) are also virus free. In *S. sclerotiorum*, colonies regenerated from individual protoplasts derived from virus-infected cultures may be virus-free. These are in contrast, for example to cultures infected with the well-studied hypovirus of *C. parasitica*, *Cryphonectria hypovirus 1* (CHV1), in which most or sometimes all conidia contain virus and virus infected cultures cannot easily be cured by hyphal tip isolation. It is likely the case that details of mycoreovirus movement within the mycelium as well as horizontal and vertical transmission are quite different from those properties of hypoviruses, which have no capsid protein, are more closely related to positive-sense single-stranded RNA viruses such as plant potyviruses and animal picornaviruses, and whose replication is associated with the fungal trans-Golgi network.

Taxonomy and Nomenclature

Naming of fungal viruses is often problematic because infectivity studies are very difficult with fungal viruses and have not been done with most. In nature, fungal viruses do not exit completely from an infected isolate and enter an uninfected one exogenously. Instead, they move from one isolate to another only after the hyphae of the two isolates have fused (anastomosed), in which case the contents of one or usually more cells from the infected and uninfected isolate are mixed. Mixed infections are common and symptomless infections are the norm for fungal viruses. For these reasons, fungal virus names usually contain reference not only to the host genus and species, but to the host strain of origin as well. The genus name *Mycoreovirus* was a natural one to use because it is descriptive. For the sake of simplicity and consistency with other viral nomenclature, species are numbered progressively as they are described. The fungal species and isolate in which a particular virus was identified is provided after the species number.

In this system, the first mycoreovirus species described was designated *Mycoreovirus 1*. The only isolate of this virus species identified to date was from *C. parasitica* strain 9B21, so the virus is designated mycoreovirus 1/Cp9B21 or MyRV1/Cp9B21. The reason that MyRV2 represents a separate virus species even though the two were isolated from the same fungal species and only a few miles away is because it shares much less sequence similarity at both the nucleotide and amino acid levels (<50%) than would be expected for two viruses in a single species. Surprisingly, both of these species are monotypic: these two virus isolates, each representing a different species, are the only two mycoreoviruses isolated to date from *C. parasitica*, even though thousands of isolates infected with members of the family *Hypoviridae* have been identified worldwide. In contrast, mycoreoviruses from dozens of isolates of *R. necatrix* from different parts of Japan have been isolated, but all are closely related to each other, indicating that they represent strains of a single virus species, *Mycoreovirus 3*.

The two distinct reoviruses characterized from the phytopathogenic ascomycete *S. sclerotiorum* are not closely related to each other and most likely represent independent introductions to the fungus by viruses evolved elsewhere. One of the viruses falls within the same clade as the other mycoreoviruses and for consistency here and in keeping with the ICTV species names is designated *Mycoreovirus 4*. In fact, the predicted amino acid sequences of the 12 MyRV4 segments share an average of ~80% identity with homologous segments of MyRV3, and terminal nucleotide sequences of MyRV3 and 4 are identical, indicating that they are recently diverged species.

The other *Sclerotinia* reovirus, which has been designated *Sclerotinia sclerotiorum* reovirus 1 (SsReV1), falls well outside of the clade containing the four viruses within the genus *Mycoreovirus*; in fact, SsReV1 is more distant from members of the genus *Mycoreovirus* than it is to members of the genus *Coltivirus* (Fig. 2). SsReV1 shows some properties distinct from those other viruses, discussed below, and has been proposed to constitute the basis for a new genus within the family *Reoviridae* based on these phylogenetic relationships and predicted coding for proteins not found in the four species of the genus *Mycoreovirus*.

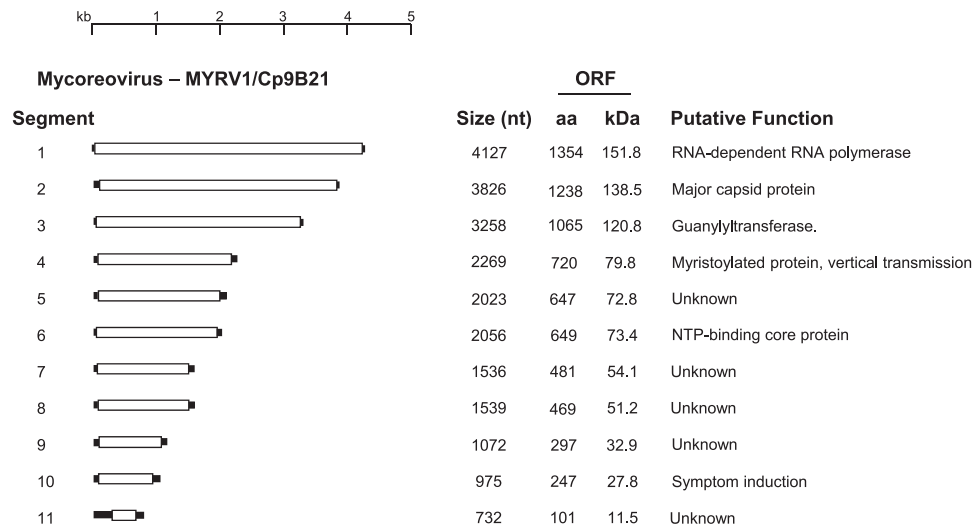


Fig. 3 Diagram illustrating the genome segments of mycoreovirus1/Cp9B21 drawn to scale with respect to size, open reading frame (ORF), and 5' and 3' untranslated regions (UTR). Segment size, ORF in number of amino acids and predicted protein product size, and putative protein function assigned to each genome segment are shown.

Genome Structures, Organizations, and Relationships

Three members of the genus *Mycoreovirus* examined to date appear to have 11 segments of dsRNA that are required for infection. Isolates of one of the viruses, MyRV3, have been found to have either 12 or 11 segments (see below). Whether any of the 12 dsRNA segments of the closely related virus MyRV4 is dispensable for stable infection of *S. sclerotiorum* is unknown. There is significant sequence similarity among the larger segments of the four mycoreovirus species and between homologous segments of the mycoreoviruses and their closest relatives, the coltiviruses, but this similarity becomes less apparent in the middle segments and is not apparent at all in the smallest segments. This feature is common with other members of the family *Reoviridae*, in which the more distant relationships are generally revealed in only the large segments. Each of the 11 required segments of mycoreoviruses appears to contain a single open reading frame (Fig. 3).

Segments 1–5 of the mycoreoviruses are homologous to segments 1–5 of the coltiviruses (segments 4 and 5 of MyRV2 are transposed relative to the others). Based largely on similarity with other reoviruses, their predicted functions are: S1: RNA-dependent RNA polymerase (core protein); S2: dsRNA-binding; putative methyltransferase (core protein); S3: Guanylyltransferase (turret protein); S4: Myristoylated membrane penetration protein (outer capsid); S5: Cytoskeleton-interacting (core). Segment 6 of the mycoreoviruses is predicted to encode a nucleic acid binding core protein and is homologous to segment 10 of the two coltiviruses. Segment 7 of MyRV1 encodes a proline rich protein with similarity to several viral and nonviral proline-rich domains. Homologies among the smaller mycoreovirus proteins (from S8–11 or 12) and other reovirus deduced proteins, including those of coltiviruses, are unknown.

One of the indications of close relationship among reoviruses is conserved terminal sequences. This also is a good indication of whether or not pseudorecombinants can be generated by co-infection with two related viruses. In the case of the mycoreoviruses, the MyRV1, MyRV2, and MyRV3 have different conserved terminal sequences, consistent with their taxonomic separation. Terminal sequences of MyRV3 and MyRV4 are identical, supporting their close relationship.

Mycoreovirus 1

Although MyRV1 was not the first of the two *C. parasitica* reoviruses identified, fungal cultures infected with this virus proved to be more stable and easily studied than those infected with MyRV2 (see below), so early molecular investigations have focused on this virus. When grown on solid media in Petri dishes (e.g., a defined complete medium or potato dextrose agar (PDA)), MyRV1-infected colonies are deep orange in color and have little aerial hyphae compared to their uninfected counterparts. As with other *C. parasitica* viruses, the phenotype of the infected culture has very little to do with the host isolate but is determined almost entirely by the virus.

Fungal isolates infected with MyRV1 are much less virulent than uninfected isolates, and are among the most debilitated of any virus-infected *C. parasitica* cultures studied to date. Although the virus accumulates to reasonably high concentrations in infected colonies, it is transmitted very poorly through conidia, at rates of only 2%–5%. This may account in part for the rarity of the virus in nature.

Expression of MyRV1 Gene Products

Functional analysis of the MyRV1 genome was initiated by cloning the 11 individual segments into a baculovirus expression vector and expressing them in insect cells. All 11 segments were expressed, resulting in 11 identifiable protein products on polyacrylamide gels. Only one of the proteins, the segment 3 product, has been studied functionally. This protein was found to be active in autoguanylation assays, confirming it as the viral guanylyltransferase. Deletion and site-directed mutational analysis determined that the amino acid sequence EPAGYHPRPSIVVPHYFVFR constituted the catalytically active site of the MyRV-1/Cp9B21 guanylyltransferase. The Hx8H motif was identified as absolutely conserved in all three members of the genus *Mycoreovirus*, as well as in the structurally related genera *Coltivirus*, *Orthoreovirus*, *Aquareovirus*, *Cypovirus*, *Dinovernavirus*, *Oryzavirus*, and *Fijivirus*. In all of the above genera in which the guanylyltransferase has been identified functionally, the Hx8H motif has been found within the sequence. The core consensus sequence for the guanylyltransferase within this group of genera was a/vxxHxxxxxxxxHhyf/lvf, with only the H residues being absolutely conserved.

Mycoreovirus 2

The first virus that was tentatively identified as a mycoreovirus was MyRV2. Like many of the virus infected *C. parasitica* cultures, the one that was found to contain this virus was isolated from a canker on an American chestnut tree. The circumstances of the discovery were somewhat unusual in that only one of 36 fungal isolates from that particular canker was virus infected; the rest were virus free. Although it is not unusual to find mixed fungal infections within a canker or to isolate virus-containing and virus-free cultures from one canker, the ratio of 1/36 is extraordinary. The phenotype of the infected culture, designated C-18 (the 18th isolate from canker C on that particular tree) was distinct from the phenotype of MyRV-1/Cp9B21-infected cultures described above in that it was light brown in color and had more aerial mycelium (Fig. 1). MyRV1 and MyRV2 have very similar, dramatic negative effects on fungal virulence. The recent finding that the original C-18 culture is co-infected with the unrelated virus CHV4 and that CHV4 helps stabilize MyRV2 in *C. parasitica*, likely as a result of suppression of RNA silencing, adds interesting complexity to understanding of mycoreoviruses.

A major difference between the two *C. parasitica* mycoreoviruses is their stability and transmissibility. MyRV2 is easily lost upon subculture of the fungus, while this has never been observed with MyRV1. Furthermore, MyRV2/CpC18 is extremely difficult to transmit from an infected isolate to an isogenic uninfected isolate by hyphal anastomosis, a property that is not seen with other *C. parasitica* viruses. It is presumed that these properties are related, but their reasons have not been fully elucidated. Both MyRV1 and MyRV2 have been transmitted to uninfected fungal isolates by inoculating *C. parasitica* protoplasts with purified virus particle preparations and allowing the protoplast to regenerate cell walls, form a hyphal network, and grow into a single colony. The ability to infect protoplasts efficiently using purified virus particle preparations is an interesting and useful feature of *C. parasitica* reoviruses and has been used for a number of other viruses recently. This has allowed for making isogenic virus-infected and virus-free isolates, starting with different genotypes of *C. parasitica* regardless of vegetative incompatibility group and for experimental transmission of mycoreoviruses to different species, genera, and families of fungi.

Mycoreovirus 3

White rot is a root disease of fruit trees that can be limiting to production. Control of the fungus that causes the disease, *Rosellinia necatrix*, by chemical means is difficult and not economically feasible. Plant pathologists in Japan, where the disease is particularly severe, have sought to use virus-infected strains to control the disease, leading to the identification of several viruses. This plant/fungus interaction represents an interesting contrast to the chestnut/*C. parasitica* interaction. As a root disease, there are challenges and opportunities for biological control of a fungal pathogen with viruses that do not apply to aerial diseases. One of the viruses under investigation for biocontrol of *R. necatrix* is the mycoreovirus MyRV3, which causes reduced virulence of the fungus. Unlike the monotypic *C. parasitica* reoviruses, different strains of MyRV3 have been isolated from a variety of *R. necatrix* strains from around Japan.

Of the *R. necatrix* reoviruses, the virus isolate that has been most thoroughly characterized is MyRV3/RnW370. In surprising contrast to the *C. parasitica* viruses, MyRV3/RnW370 was found to contain 12 rather than 11 segments. However, the presence of 12 segments is not a necessary feature of MyRV3: *R. necatrix* isolates derived from subculture of original virus-infected isolates may have either 12 or 11 segments. Experiments to investigate virus composition and transmission have been performed on hyphal tip cultures from infected *R. necatrix* isolates, resulting in demonstration that these viruses behave like the *C. parasitica* mycoreoviruses.

When only 11 segments are present in MyRV3 subculture isolates, segment 8 is the one that is absent from the full complement. With most reoviruses, sequence conservation among species and genera is evident in the larger segments, but much less so in the smaller segments, and this is true in the mycoreoviruses and related genera. Consistent with this general trend, sequence comparison between the 12 segments of MyRV3 and the 12 segments of the two coltiviruses has suggested nothing about possible function of the apparently dispensable segment 8.

It is intriguing to think that perhaps MyRV3 segment 8 is vestigial for reoviruses in fungi and is required only in another host, past, or present. This would be similar to the leafhopper-transmitted phytoreovirus, wound tumor virus (WTV), in which deletion

mutations in any of three segments may be found upon successive serial, insect-free virus passage, or long-term virus maintenance in plants, whereupon resulting mutant viruses become defective in their transmission properties and incapable of replicating in their leafhopper vectors. An apparent major difference between MyRV3 and WTV is that in MyRV3 there is no evidence for terminal remnants of segment 8 in deletion mutant strains; it appears to be either present or entirely absent. In the well-characterized mutants of WTV, the deleted segments were not completely absent; rather, shorter mutant segments containing the two termini and varying amounts of adjacent sequence remain, ensuring that a total of 12 segments remain. Whether this represents a difference between the dsRNA segment sorting and packaging mechanisms of phytoreoviruses and mycoreoviruses is not known. In this line of inquiry, the close association of fungi with mites in natural settings may be significant to the evolutionary biology of mycoreoviruses: it may be no coincidence that their closest relatives are the tick-borne coltivirus, with ticks and mites both in the arachnid subclass Acari. Unfortunately, mites are very difficult experimental subjects and cell cultures are not currently available. Furthermore, much less is known about coltivirus gene function than is known about many of the other members of the family *Reoviridae* that are pathogenic to humans, making it more difficult to pursue this line of research from a strictly bioinformatics standpoint.

Sclerotinia Sclerotiorum Reoviruses

Mycoreovirus 4

MyRV4 from *S. sclerotiorum* is most closely related to MyRV3 from *R. necatrix*, and even though the two were identified in fungi representing distinct classes of ascomycetes, they are more similar to each other than are MyRV1 and MyRV2 to each other, even though the latter two were both isolated from *C. parasitica* only a few miles apart from each other. MyRV4 infection of *S. sclerotiorum* reduces virulence of the fungus on its plant hosts and also suppresses host non-self recognition, or vegetative incompatibility, in the fungus. In doing so, MyRV4 appears to enhance its own spread in nature by lowering the existing barriers to spread within heterogeneous field populations. Like its closest relative MyRV3, MyRV4 contains 12 segments of dsRNA. Each of the 12 segments of MyRV4 has a homologous MyRV3 segment, so these viruses will be good candidates for comparative genomics and for potential coinfection studies. Unlike MyRV3, in which stable virus isolates containing only 11 dsRNA segments (i.e., lacking segment 8) have been identified, no MyRV4 isolates containing only 11 dsRNA segments have yet been reported. Segment 8 of MyRV4 shares 77% nucleotide identity with MyRV3, so it would not be surprising to find similar MyRV4 isolates lacking this segment.

MyRV4 is also associated with distinct alterations to fungal colony morphology in culture, and to alterations in fungal gene expression. Suppression of *het* gene expression, the genes associated with heterokaryon, or vegetative compatibility explains the observed suppression of vegetative incompatibility and increased virus transmission among diverse fungal strains. Reduction in expression of reactive oxygen species (ROS) genes by MyRV4 infection also likely contributes to the same phenotype.

Sclerotinia Sclerotiorum Reovirus 1

In addition to its greater phylogenetic distance from the four members of the genus *Mycoreovirus*, the other *Sclerotinia* reovirus, SsReV1, differs from those in that it causes no identified morphological or phenotypic change to the host fungus. SsReV1 was originally isolated from a hypovirulent strain of *S. sclerotiorum*, but infection with purified virus demonstrated that it was not the cause of the hypovirulent phenotype. A notable feature of the SsReV1 genome is the presence of two domains that were identified in several other viruses and eukaryotes, but not in mycoreoviruses. One of the domains, a dsRNA binding motif, was found especially in DNA viruses and also in cellular host genomes. The other domain, the Reo σ C capsid domain, was more commonly found in other reoviruses but also in DNA-containing viruses. A possible new reovirus genus encompassing SsReV1 has not yet been described.

Effects of Mycoreoviruses on Fungal Gene Expression

Considerable information has been amassed on the impact of CHV1, a positive-sense RNA virus, on its fungal host (see *Hypoviridae*). In contrast, studying the mechanisms of mycoreovirus infection on fungi is in its infancy. The first study addressing these questions was done by microarray analysis using mRNA isolated from isogenic fungal isolates of *C. parasitica* infected with either MyRV1 or MyRV2, and comparing results to the same fungal strain that was uninfected or infected with either of two different CHV1 strains, and with fungal mutants defective in virulence characteristics. To date, these experiments have been performed only on EST-based arrays representing ~20% of the total *C. parasitica* gene complement. Overall, there was consistency in the effects of the two *C. parasitica* mycoreoviruses on host gene expression: MyRV1 infection resulted in differential expression of 6.5% of the genes on the array, whereas MyRV2 infection affected expression of 5.8% of those genes. As might be expected based on their phenotypes, similar but distinct suites of genes were up- or downregulated in isogenic fungal isolates infected with the two reoviruses. Approximately 60% of the genes whose expression was affected were the same whether infection was by MyRV1 or MyRV2, and all but one of those genes were altered in the same direction. Some of these groups of genes are in common with

those that are differentially regulated in cultures infected with the unrelated hypoviruses, but there are predictable differences. For example, hypovirus infection of *C. parasitica* results in female infertility, whereas mycoreovirus infection does not, and this is reflected in the expression of two genes predicted to be involved in the *C. parasitica* mating response. Both *mf2-1*, which encodes the fungal pheromone precursor, and *Csp12*, which encodes a homolog of the yeast *Ste12*-like transcription factor, were substantially downregulated in hypovirus-infected fungal isolates, which are female sterile, whereas there was much less effect on expression of these genes in either of the mycoreovirus-infected *C. parasitica* isolates. Virus is transmitted to ascospores at a rate of ~50% or less when the female parent is infected, but there is no virus transmission to ascospore progeny if the male parent in a mating is infected. Sequencing the complete genome of several strains of *C. parasitica* has been accomplished, and current investigations based on high-throughput RNAseq methodologies are underway.

Coinfections of Mycoreovirus and Other Viruses

Double infections of the well-characterized hypovirus CHV1/EP713 and MyRV1/Cp9B21 revealed that presence of the hypovirus caused an increase in both the concentration and the vertical transmission rate through conidia of the mycoreovirus, but that the mycoreovirus did not affect the concentration or transmission rate of the hypovirus. Furthermore, transgenic expression of only the hypovirus protein p29 also resulted in increased accumulation and transmission of MyRV1/Cp9B21. This is consistent with evidence that p29 serves as a suppressor of RNA silencing (= RNA interference) during hypovirus infection, and that this effect acts in trans to support enhanced mycoreovirus replication.

Coinfection of *C. parasitica* with CHV4 was recently shown to facilitate infection with MyRV2. Single infections of either the original host strain, C-18, or the commonly used virus-free strain EP155 with MyRV2 were less stable upon serial subculture than were the same strains when co-infected with CHV4, which itself is symptomless but stable in *C. parasitica*. Using *C. parasitica* mutants defective in the dicer-like gene *dcl2*, which is an integral component of the fungal RNA silencing-based defense response against virus infection, it was shown that the putative silencing suppressor of CHV4 was responsible for trans-stabilization of MyRV2.

SsReV1 was also first identified as a coinfection with another virus. In this case, a bipartite dsRNA virus was identified along with the 11-segment reovirus in *S. sclerotiorum* isolate SCH94. The doubly-infected isolates had a distinct colony morphology compared to virus-free isolates, but isolates infected with only SsReV1 were indistinguishable from uninfected isolates, indicating that SsReV1 alone did not have a discernable effect on the fungus. Unlike the above case with MyRV2, there was no evidence that SsReV1 was unstable without its original coinfecting virus.

Mycoreovirus Genome Rearrangements

Numerous alterations and rearrangements of mycoreovirus genome segments have been identified and characterized. The first of these, discussed above, involves complete loss of segment 8 of MyRV3 in some fungal isolates. As far as we are aware, this is the only example of contemporary observed loss of an entire reovirus segment. Whether this segment is dispensable because it is not required for virus infection of the fungus in culture or because it is required for infection of an as-yet unknown arthropod host is unknown.

Other segment alterations and rearrangements have been characterized during experimental infection of MyRV1. These include deletion of part or all of specific coding regions and duplication of all or part of coding regions of specific segments. Some of these segment alterations occurred in isolates singly-infected with MyRV1, some resulted in isolates coinfecting with MyRV1 and the hypovirus CHV1, and some from MyRV1 infection of transgenic fungal isolates expressing the p29 RNA silencing suppressor protein. Many of the MyRV1 segment mutants are associated with growth phenotypes and differences in viral RNA accumulation.

Coding region duplication among MyRV1 segments is an interesting phenomenon that has not been observed in other systems. When MyRV1 is influenced by the CHV1 silencing suppressor p29, in-frame extensions of the coding sequences of segments 1, 2, 3, and 6 have been observed, resulting in predicted larger protein products from those segments. Each of these products is predicted to be a structural protein, and while the predicted longer than wildtype proteins have been observed, in infected cells, it is not yet known whether and how these might be incorporated into particles. This type of segment coding region extension is unusual: in observed coding region duplications within rotaviruses, the most thoroughly studied example of such rearrangements, the size of the expressed coding region is not extended in-frame, but rather a second coding region is inserted head-to-tail downstream of the segment's unaltered cognate coding sequence. Such rearrangements would not be predicted to alter the natural coding sequence of the rearranged segment.

Deletions in MyRV1 segments 4 and 10, which are non-structural proteins, are tolerated. These deletions may be in-frame, resulting in a predicted fusion gene product, or out-of-frame, resulting in predicted truncated gene products. Surprisingly, virus that contains complete replacements of both segments 4 and 10 with their deletion mutant counterparts is viable, probably representing the only known example of a reovirus mutant that can tolerate loss of coding capacity of two segments. Also interesting is that this double deletion mutant replicates to similar levels as wildtype MyRV1.

The next step in mycoreovirus research is to develop a robust reverse genetics system to complement studies such as those above. Furthermore, discovery and characterization of new members of the genus *Mycoreovirus* – and similar fungal reoviruses – will continue as high throughput sequence-based methods dominate the realm of virus discovery.

Further Reading

- Attoui, H., Mertens, P.P.C., Becnel, J., *et al.*, 2012. Family Reoviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Ninth Report of the International Committee for the Taxonomy of Viruses*. New York: Elsevier, Academic Press, pp. 541–637.
- Aulia, A., Andika, I.B., Kondo, H., Hillman, B.I., Suzuki, N., 2019. A symptomless hypovirus, CHV4, facilitates stable infection of the chestnut blight fungus by a coinfecting reovirus likely through suppression of antiviral RNA silencing. *Virology* 533, 99–107.
- Enebak, S.A., Hillman, B.I., MacDonald, W.L., 1994. A hypovirulent *Cryphonectria parasitica* isolate with multiple, genetically unique dsRNA segments. *Molecular Plant-Microbe Interactions* 7, 590–595.
- Eusebio-Cope, A., Sun, L., Tanaka, T., *et al.*, 2015. The chestnut blight fungus for studies on virus-host and virus/virus interactions: From a natural to a model host. *Virology* 477, 164–175.
- Eusebio-Cope, A., Suzuki, N., 2015. Mycoreovirus genome rearrangements associated with RNA silencing deficiency. *Nucleic Acids Research* 43, 3802–3813.
- Hillman, B.I., Supyani, S., Kondo, H., Suzuki, N., 2004. A reovirus of the fungus *Cryphonectria parasitica* that is infectious as particles and related to the Coltivirus genus of animal pathogen. *Journal of Virology* 78, 892–898.
- Hillman, B.I., Suzuki, N., 2004. Viruses of *Cryphonectria parasitica*. *Advances in Virus Research* 63, 423–472.
- Kanematsu, S., Arakawa, M., Oikawa, Y., *et al.*, 2004. A reovirus causes hypovirulence of *Rosellinia necatrix*. *Phytopathology* 94, 561–568.
- Liu, L., Cheng, J., Fu, Y., *et al.*, 2017. New insights into reovirus evolution: Implications from a newly characterized mycoreovirus. *Journal of General Virology* 98, 1132–1141.
- Osaki, H., Wei, C.Z., Arakawa, M., *et al.*, 2002. Nucleotide sequences of double-stranded segments from hypovirulent strain of the white root rot fungus *Rosellinia necatrix*: Possibly of the first member of the *Reoviridae* from fungus. *Virus Genes* 25, 101–107.
- Supyani, S., Hillman, B.I., Suzuki, N., 2006. Baculovirus expression of the 11 Mycoreovirus-1 genome segments and identification of the guanylyltransferase-encoding segment. *Journal of General Virology* 88, 342–350.
- Suzuki, N., Supyani, S., Maruyama, K., Hillman, B.I., 2004. Complete genome sequence of Mycoreovirus 1/Cp9B21, a member of a new genus in the family Reoviridae isolated from the chestnut blight fungus, *Cryphonectria parasitica*. *Journal of General Virology* 85, 3437–3448.
- Tanaka, T., Eusebio-Cope, A., Sun, L., Suzuki, N., 2012. Mycoreovirus genome alterations: Similarities to and differences from rearrangements reported for other reoviruses. *Frontiers in Microbiology* 3, 186.
- Wu, S., Cheng, J., Fu, Y., *et al.*, 2017. Virus-mediated suppression of host non-self recognition facilitates horizontal transmission of heterologous viruses. *PLoS Pathogens* 13 (3), e1006234. doi:10.1371/journal.ppat.1006234.

Mymonaviruses (*Mymonaviridae*)

Daohong Jiang, Huazhong Agricultural University, Wuhan, China

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Biological control A method to prevent pests and parasites using beneficial organisms, here specifically means using mycovirus to control fungal diseases.

Hypovirulence Reduced virulence of a fungal pathogen caused by the infection of mycovirus; Hypovirulence and hypovirus of chestnut blight/*Cryphonectria parasitica* system

is a classic example for studying mycovirus and biological control of plant fungal diseases.

Mymonaviruses A group of mononegaviruses that was originally identified from a fungus.

Sclerotinia sclerotiorum An ascomycetous pathogen that can attack more than 400 species and subspecies of plants.

Introduction

The mycovirus was first discovered by Hollings in 1962 from a diseased mushroom, *Agaricus bisporus*. In the same year, an active substance capable of stimulating the production of interferon in mammals was found in *Penicillium* spp., and this substance was identified as double-stranded (ds)RNA which was extracted from mycovirus particles. The dsRNA elements also were found in a hypovirulent strain of chestnut blight pathogen (*Cryphonectria parasitica*) at 1970s. The dsRNA elements represented the first identified virus (hypovirus) which are devoid of coat protein. These studies led to a notion that the genome of the fungal virus is dsRNA. Since then, dsRNA became a marker for mycoviruses and was used to judge whether a fungal strain was infected by mycoviruses or not still today. DsRNA extraction screening with cellulose has detected a group of mycoviruses from various fungi, which produce dsRNA, as their genomes or replicative forms, in infected cells. Some of these viruses are phylogenetically related to plant single-stranded, positive-sense RNA (+ ssRNA) and are capsidless. Such fungal RNA viruses exemplified by hypoviruses and endornaviruses are now classified as + ssRNA viruses, rather than as dsRNA viruses, by the International Committee on Taxonomy of Viruses (ICTV). By recent high-throughput RNA sequencing or RNA Seq of hypovirulent strain AH98 of *Sclerotinia sclerotiorum*, an important necrotrophic ascomycetous pathogen that attacks numerous dicotyledonous plants, we found that strain AH98 was infected by a negative-stranded (–) RNA virus, designated as *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNSRV-1). This is the first discovery of a -ssRNA virus as an infectious entity. Meanwhile, Dr. H. Kondo's research group in Japan discovered the incomplete sequence of L-protein-like gene in the genome of the powdery mildew fungus by performing an exhaustive search using sequences of known negative-strand RNA viruses. They also found two entire L-protein genes and one partial gene of unknown mononegaviruses present in two independent transcriptomic data of *Clariireedia* (*Sclerotinia*) *homoeocarp*. In 2016, these findings led ICTV to create the new family *Mymonaviridae* within the order *Monogavirales* containing one genus *Sclerotimonavirus* with *Sclerotinia sclerotimonavirus* as the type species, the exemplar virus of which is SsNSRV-1. The name *Mymonaviridae* is originated from *Myco* and *Mononegavirales*.

Phylogenetic Status of *Mymonaviridae* and its Related Families

The family *Mymonaviridae* belongs to the order *Monogavirales* (class *Monjiviricetes*, subphylum *Haploviricoyina*, phylum *Negarnaviricota*, realm *Riboviria*). Based on the phylogenetic analysis using the full-length L protein (RNA replicase), the members of *Mymonaviridae* are closely related to those of *Lispiviridae*, *Paramyxoviridae*, *Filoviridae*, *Pneumoviridae* and *Rhabdoviridae*, and are also phylogenetically related to members of *Nayamiviridae*, *Bornaviridae*, *Xinmoviridae*, *Artoviridae* and *Chuviridae*. Currently, there is only one genus (*Scleromonavirus*) in this family, however, the members in *Mymonaviridae* can be grouped into three clusters, and more genera should be established possibly to adopt viruses which are significantly different from those in *Scleromonavirus* (Fig. 1).

Species and Tentative Species in the Family *Mymonaviridae*

At present, the genus *Scleromonavirus* has seven species, namely *Sclerotinia sclerotimonavirus*, *Hubei sclerotimonavirus*, *Drop sclerotimonavirus*, *Dadou sclerotimonavirus*, *Illinois sclerotimonavirus*, *Glycine sclerotimonavirus* and *Phyllosphere sclerotimonavirus*. However, it is clear that there are additional eight viruses at least; *Penicillium cairnsense* negative-stranded RNA virus 1, *Narnavirus* sp.-QDH90729.1 (tentative name), *Lentinula edodes* negative-strand RNA virus 1, *Penicillium adametzoides* negative-stranded RNA virus 1 or *Alternaria tenuissima* negative-stranded RNA virus 1 represents novel species; *Botrytis cinerea* mymonavirus 1 and *Sclerotinia sclerotiorum* negative-stranded RNA virus 7 together represent a novel species, and TSA1, -2, -3 from *Clariireedia homoeocarpaa* may also represent a novel species (Table 1).

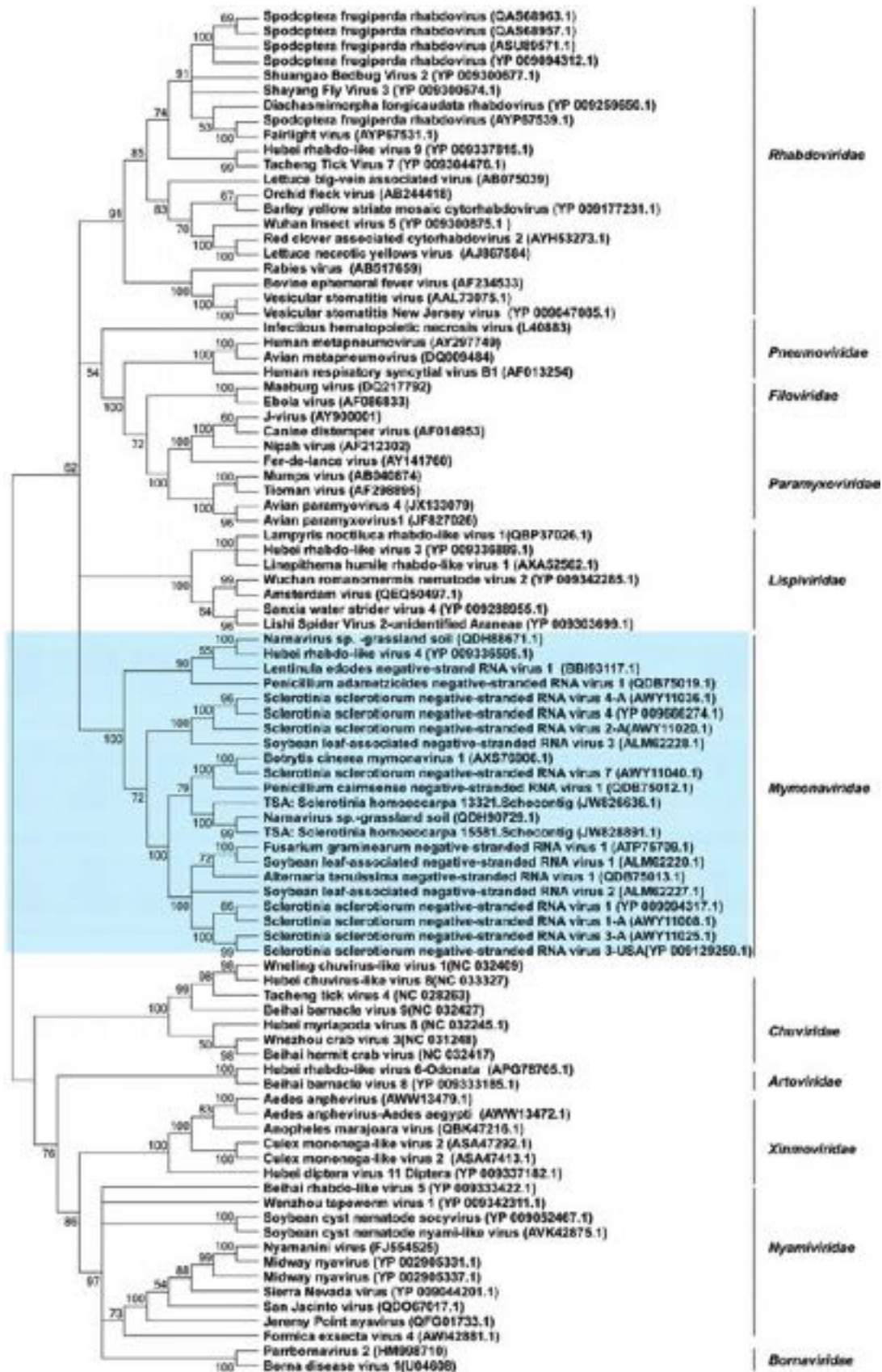


Fig. 1 Phylogenetic analysis of members and tentative members in the family *Mymonaviridae* and other related families by Maximum Likelihood method. The viruses whose complete sequence of RNA replicase is known were selected to construct this phylogenetic tree. In total 83 RNA replicase sequences from viruses in the order *Mononegavirales* were analyzed. The phylogenetic analyses were conducted in MEGA7 (Kumar, S., Stecher, G., Tamura, K., 2016. Molecular Biology and Evolution 33,1870–18742). Bootstrap numbers out of 1000 replicates are indicated at the nodes. Virus name and accession number were on the right at the correspondence position.

Table 1 List of members and tentative members in the family *Mymonaviridae*

Species name	Virus name	Available sequence (nt)	Accession number	Host
<i>Sclerotinia sclerotimonavirus</i>	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 1	10002	NC_025383	Fungi, <i>Sclerotinia sclerotiorum</i>
	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 1-A	7735	MF444277	Fungi, <i>S. sclerotiorum</i>
	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 3-A	9919	MF444280	Fungi, <i>S. sclerotiorum</i>
	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 3-USA	10009	NC_026732.1	Fungi, <i>S. sclerotiorum</i>
	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 4	10003	NC_032783.1	Animal, Arthropod
<i>Hubei sclerotimonavirus</i>	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 4-A	9564	MF444282.1	Fungi, <i>S. sclerotiorum</i>
<i>Drop sclerotimonavirus</i>	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 4	9707	NC_043483.1	Fungi, <i>S. sclerotiorum</i>
	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 2-A	9450	MF444279.1	Fungi, <i>S. sclerotiorum</i>
<i>Dadou sclerotimonavirus</i>	Soybean leaf-associated negative-stranded RNA virus 3	6218	KT598228.1	Plant or associated fungi, soybean or Soybean leaf microbe
<i>Illinois sclerotimonavirus</i>	Soybean leaf-associated negative-stranded RNA virus 2	7321	KT598227	Plant or associated fungi, soybean or Soybean leaf microbe
<i>Glycine sclerotimonavirus</i>	Soybean leaf-associated negative-stranded RNA virus 1	9041	KT598225	Plant or associated fungi, soybean or Soybean leaf microbe
	<i>Fusarium graminearum</i> negative-stranded RNA virus 1	9072	MF276904.1	Fungi, <i>Fusarium graminearum</i>
<i>Phyllosphere sclerotimonavirus</i>	soybean leaf-associated negative-stranded RNA virus 4	5317	KT598229	Plant or associated fungi, soybean or soybean leaf microbe
Tentative members	<i>Penicillium adametzioides</i> negative-stranded RNA virus 1	6713	MK584858	Fungi, <i>Penicillium adametzioides</i>
	<i>Alternaria tenuissima</i> negative-stranded RNA virus 1	8904	MK584852	Fungi, <i>Alternaria tenuissima</i>
	<i>Narnavirus</i> sp. ^a	9843	QDH88671.1	Grassland soil
	<i>Lentinula edodes</i> negative-strand RNA virus 1	11563	BBI93117.1	Fungi, <i>Lentinula edodes</i>
	<i>Penicillium cairnsense</i> negative-stranded RNA virus 1	8759	MK584851	Fungi, <i>Penicillium cairnsense</i>
	<i>Narnavirus</i> sp. ^a	9885	MN035745	Grassland soil
	<i>Botrytis cinerea</i> mymonavirus 1	7863	MH648611.1	Fungi, <i>Botrytis cinerea</i>
	ShTAS3 ^b	8062	JW828891.1	Fungi, <i>Clariireedia homoeocarpa</i>
	ShTAS2 ^b	8645	JW826636.1	Fungi, <i>Clariireedia homoeocarpa</i>
	ShTAS1 ^b	9775	JU091017	Fungi, <i>Clariireedia homoeocarpa</i>
	<i>Sclerotinia sclerotiorum</i> negative-stranded RNA virus 7	7819	MF444285	Fungi, <i>Sclerotinia sclerotiorum</i>

^aThis virus is most likely to incorrectly be named, narnaviruses that belong to the family *Narnaviridae*, with + ssRNA genomes of <3.0 kb.

^bShTAS1, -2, -3 are not really virus names, but are names for transcriptome shotgun assemblies from *Sclerotinia homoeocarpa* (renamed as *Clariireedia homoeocarpa*).

Virion

So far, virion morphology of SsNSRV1 virus and *Fusarium graminearum* negative-stranded RNA virus 1 (FgNSRV-1) has been defined, whereas the virions of other viruses are still unknown. The morphology of virions of SsNSRV1 and FgNSRV-1 is similar. SsNSRV-1 virions are filamentous, 25–50 nm in diameter and about 1000 nm in length, while FgNSRV-1 particles are 35–50 nm in diameter and about 1200 nm in length. The virions have enveloped-like structures (EVLS), which are similar to the virion of viruses in the family *Filoviridae*, but lacking spikes on the surface. When the virion is broken, Nucleocapsid complexes (RNPs) are released. RNPs are filamentous, branched, left-handed, and helical structures with tight or loose coils. RNPs of SsNSRV-1 have a diameter of 20–22 nm and a length of 400–2000 nm, while the size of RNPs is unknown for FgNSRV-1. Nucleocapsid consists of polymerized Nucleoprotein (N) monomers (Fig. 2).

The Genomic Structure

Since the full-length genome sequences of some viruses are not available and novel viruses with larger or smaller genomes than the fully-sequenced mymonaviruses are frequently discovered, the size range of viral genome cannot be accurately determined.

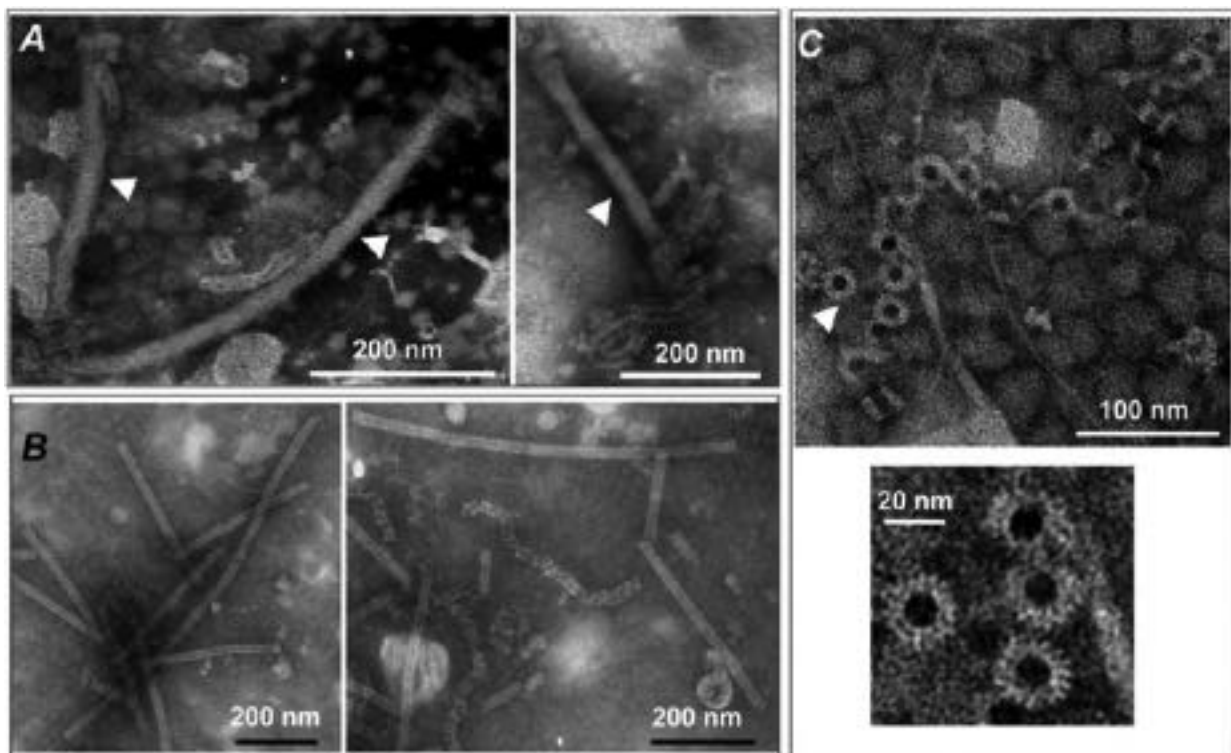


Fig. 2 Morphology and structure of virions of *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNSRV-1). (A) Filamentous, possibly enveloped virions (marked by arrowheads), and nucleoprotein-RNA complexes (RNPs). (B) Purified tight (left) or loose (right) coils of RNPs. (C) Rings that make up the coils and nucleoprotein (NP) monomers (marked by arrowhead). Virions and RNPs were purified from mycelia of strain AH98 and virus-transfected strain Ep-1PNA367-PT2 of *Sclerotinia sclerotiorum* and negatively stained with 2% PTA [(wt/vol), pH 7.4]. Reproduced from Liu, L, Xie, J, Cheng, J, *et al.*, 2014. Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12205–12210.

From the current known viruses in the family *Mymonaviridae*, the genome size may be between 9000 and 11600 nt, and the largest virus in the genome is *Lentinula edodes* negative-strand RNA virus 1. There are 5–7 open reading frames (ORFs) on the genome; SsNSRV-1 has six ORFs on its genome (Fig. 3(A)). There are conserved non-coding regions between genes. For example, in the SsNSRV-1 genome, the sequences of non-coding regions are (3′ – (A/U) (U/A/C) UAUU (U/A) AA (U/G) AAAACUUAGG (A/U) (G/U) – 5′) (Fig. 3(B)). The N protein is encoded by ORF II, and the L protein (RNA replicase) is encoded by ORF V. There is usually an ORF (ORF VI) following ORF V. The small ORF IV and ORF VI are undetectable in some mymonaviruses. The gene arrangement of mymonaviruses is different from that of other mononegaviruses. The most conserved L protein shares sequence similarity among those of different members of the same genus, whereas other proteins have no significant similarity among them. The functions of the N and L protein are determined, while the functions of other proteins are unknown. In fact, only the transcriptional pattern of the SsNSRV-1 genome is known: six ORFs can be transcribed independently, while ORF V and ORF VI are transcribed together (Fig. 3(C)). The genome of this family of viruses has occasionally defective RNA molecules. For example, in *S. sclerotiorum*, there are defective RNAs in which both 3′-terminus (including ORF I and ORF II, even ORF III) and 5′-terminus are deleted (Fig. 3(D)). It is likely that similar phenomenon occurs in other mymonaviruses and the "defective" RNAs replicate in fungi. *Botrytis cinerea* mymonavirus 1 has only three ORFs supports this notion. This observation may suggest that certain shorter version of the mymonavirus genomes, resulting from deletions of dispensable genes, are replication competent.

The Impact on the Host

The impact of mymonaviruses on their host is very difficult to determine. SsNSRV-1 was successfully transfected into a virus-free *S. sclerotiorum* strain by protoplast transfection. The pathogenicity of the strain infected with the virus showed a significant decline compared with the strain without the virus. *Botrytis cinerea* mymonavirus 1 was also found to have a significant inhibitory effect on the pathogenicity of its host *B. cinerea*; in addition, *Lentinula edodes* negative-strand RNA virus 1 may have an effect on the formation and development of fruitbody of the host shiitake mushroom (*Lentinula edodes*). FgNSRV-1 had no significant effects on the growth, pathogenicity, sporulation and toxin synthesis of the host *F. graminearum*. The impact of other members in the family *Mymonaviridae* has not been reported.

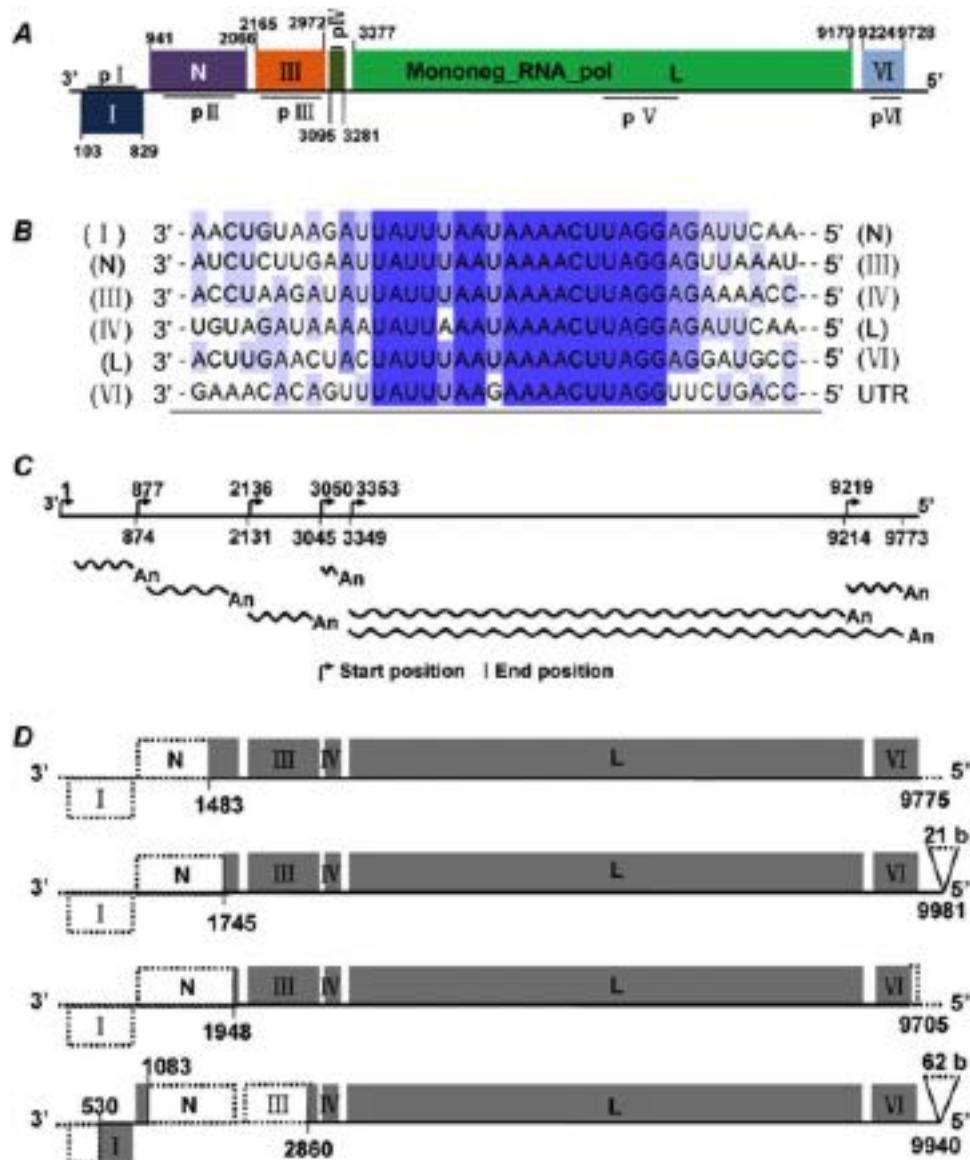


Fig. 3 Genome organization and characteristics of *Sclerotinia sclerotiorum* negative-stranded RNA virus 1 (SsNSRV-1). (A) Genome length and organization. Boxes indicate position and length of each ORF, which are labeled with Roman numerals, except for two ORFs, N (II) and L (V), which encode the nucleoprotein (N) and RNA-dependent RNA polymerase (L). A conserved domain in L is shown. (B) Alignment of the putative gene-junction sequences between ORFs in a 3'-to-5' orientation. Conserved sequences [3' - (A/U)(U/A/C)UAAU(U/A)AA(U/G) AAAACUUAGG(A/U) (G/U) - 5'] are highlighted in blue. Different shades indicate levels of conservations with the darkest color indicating the highest conservation. (C) Deduced transcription map based on 5'- and 3'-RACE. (D) Schematic diagrams for four defective genome RNA of SsNSRV-1, dotted lines represents deleted regions on the genome. Reproduced from Liu, L, Xie, J, Cheng, J, *et al.*, 2014. Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12205–12210.

Host and Distribution

The hosts of viruses in family *Mymonaviridae* are mostly fungi, but two of them (tentatively called *Narnavirus* sp.-QDH90729.1 and -*Narnavirus* sp.-QDH88671.1) were identified from grassland soil, and Hubei rhabdo-like virus 4 was discovered from an arthropods or arthropods-associated organisms. It is interesting that *Narnavirus* sp.-QDH88671.1 belongs to the same species as Hubei rhabdo-like virus 4, suggesting that they may share the same hosts. Three viruses were identified from phyllosphere. It is considered that some viruses from phyllosphere are closely related to viruses from pathogenic fungi that grow on the same plants; hence, those viruses likely have fungal hosts.

Currently, it is known that viral strains of the species *Sclerotinia sclerotimonavirus* such as SsNSRV-1 are distributed widely in China; it is also detected in both the United States and Australia. SsNSRV-4 (species *Drop sclerotimonavirus*) is detected in

S. sclerotiorum in the United States and Australia. *Botrytis cinerea* mymonavirus 1 is distributed in 11 provinces of China and is also found in *S. sclerotiorum*. Hence, viruses in the family *Mymonaviridae* are widely spread around the world.

Further Reading

- Amarasinghe, G.K., Ayllón, M.A., Bào, Y., *et al.*, 2019. Taxonomy of the order Mononegavirales: Update 2019. *Archives of Virology* 164, 1967–1980.
- Hao, F., Wu, M., Li, G., 2018. Molecular characterization and geographic distribution of a mymonavirus in the population of *Botrytis cinerea*. *Viruses* 10, 432.
- Kondo, H., Chiba, S., Toyoda, K., Suzuki, N., 2013. Evidence for negative-strand RNA virus infection in fungi. *Virology* 435, 201–209.
- Lin, Y.H., Fujita, M., Chiba, S., *et al.*, 2019. Two novel fungal negative-strand RNA viruses related to mymonaviruses and phenuiviruses in the shiitake mushroom (*Lentinula edodes*). *Virology* 533, 125–136.
- Liu, L., Xie, J., Cheng, J., *et al.*, 2014. Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12205–12210.
- Marzano, S.L., Domier, L.L., 2016. Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Research* 213, 332–342.
- Marzano, S.L., Nelson, B.D., Ajayi-Oyetunde, O., *et al.*, 2016. Identification of diverse mycoviruses through metatranscriptomics characterization of the viromes of five major fungal plant pathogens. *Journal of Virology* 90, 6846–6863.
- Mu, F., Xie, J., Cheng, S., *et al.*, 2017. Virome characterization of a collection of *Sclerotinia sclerotiorum* from Australia. *Frontiers in Microbiology* 8, 2540.
- Nerva, L., Turina, M., Zanzotto, A., *et al.*, 2019. Isolation, molecular characterization and virome analysis of culturable wood fungal endophytes in esca symptomatic and asymptomatic grapevine plants. *Environmental Microbiology* 21, 2886–2904.
- Shi, M., Lin, X., Tian, J., *et al.*, 2017. Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543.
- Wang, L., He, H., Wang, S., *et al.*, 2018. Evidence for a novel negative-stranded RNA mycovirus isolated from the plant pathogenic fungus *Fusarium graminearum*. *Virology* 518, 232–240.

Narnaviruses (*Narnaviridae*)

Rosa Esteban and Tsutomu Fujimura, Institute of Biology and Functional Genomics, CSIC/University of Salamanca, Salamanca, Spain

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Ribozyme RNA with a catalytic activity.

Introduction

The narnaviruses 20S RNA and 23S RNA (ScV20S and ScV23S, respectively) are positive strand RNA viruses found in the yeast *Saccharomyces cerevisiae*. These two viruses are ascribed to the genus *Narnavirus* of the family *Narnaviridae*. Massive RNA sequencing has started detecting many narnavirus-like sequences in plant and insect sources. Like other fungal viruses, 20S and 23S RNAs have no extracellular transmission pathway. They are transmitted horizontally by mating, or vertically from mother to daughter cells. It is believed that the high frequency of mating or hyphal fusion that occurs in the host life cycle makes an extracellular route of transmission dispensable for the viruses. The thick cell wall of fungi may also form a formidable barrier. The lack of extracellular transmission may, in turn, explain two prominent features found in narnaviruses. First, they are persistent viruses and do not kill the host cells. If their infection caused damages or disadvantages to the host, then the viruses might have perished during the course of evolution because of the lack of an escape route. Secondly, because there is no extracellular phase, the viruses do not need to form virions to protect their RNA genomes in the extracellular environment. In addition, they do not need machinery to ensure exit or reentry to a new host. The lack of a virion structure may sound peculiar to those who are familiar with infectious viruses. Considering that viruses are selfish parasites, however, it will be natural for them to shed genes or functions unnecessary for their existence. Consequently, the genomes of narnaviruses are simple and small: they only encode a single protein, the RNA-dependent RNA polymerase (RdRp). This may contribute to their persistence by reducing a number of viral proteins that might interfere with metabolism vital for the host. The simplicity of their RNA genomes encoding a single protein, together with the development of 20S and 23S RNA virus launching systems from yeast expression vectors, makes narnaviruses a good model system to investigate replication and the molecular basis for intracellular persistence of RNA viruses.

Historical Background

20S RNA was first described in 1971 as a single-stranded RNA (ssRNA) species that accumulated in yeast cells transferred to 1% potassium acetate, a standard procedure to induce sporulation in yeast under nitrogen starvation conditions. Because of its mobility relative to 25S and 18S rRNAs, the species was named 20S RNA. Later it was found, however, that the accumulation of 20S RNA was not related to the sporulation process because haploid cells that do not sporulate also accumulate 20S RNA under nitrogen starvation conditions. It was also found that 20S RNA is a cytoplasmic genetic element. The realization of 20S RNA as a viral entity, however, had to wait several years, until the characterization of 20S RNA by cloning and sequencing in 1991. 23S RNA was reported first time in 1992. Both viruses were recognized as members of a new family, *Narnaviridae* (naked RNA virus), a taxonomic group that appeared first time in the 2000 edition of the ICTV report. The other genus in the family is *Mitovirus*, whose members are found in mitochondria of fungi, many of them are pathogenic to plants. All members of the family have small RNA genomes (2–3 kilobases) that encode single proteins, their RNA-dependent RNA polymerases, and reside either in the cytoplasm (narnaviruses) or in the mitochondria (mitoviruses) of the host.

Viral Genomes

Many laboratory strains of *S. cerevisiae* harbor 20S RNA virus and fewer strains contain 23S RNA virus. Both viruses are compatible in the same host. The presence of 20S and 23S RNA viruses does not confer phenotypic changes to the host. Under nitrogen starvation conditions, the amounts of the viral genomes become almost equivalent to those of rRNAs (>100,000 copies/cell; Fig. 1).

In contrast, vegetative growing cells contain much lower amounts of the viral RNAs (5–20 copies/cell). Fig. 2(A) shows the genome organization of narnaviruses. Both 20S and 23S RNAs are small (2514 and 2891 nt, respectively) and each genome encodes a single protein; a 91 kDa protein (p91) by 20S RNA and a 104 kDa protein (p104) by 23S RNA.

The 5' untranslated regions in both RNAs are extremely short: 12 nt in the case of 20S RNA and only 6 nt in 23S RNA. These RNAs lack poly(A) tails at the 3' ends. The same RNA can serve as template for translation and also for negative strands

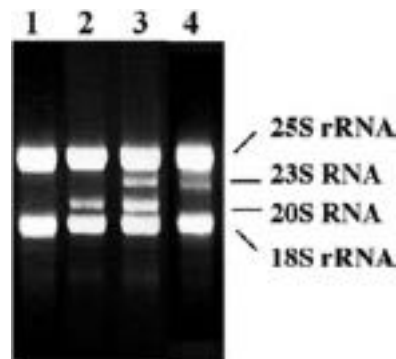


Fig. 1 RNA extracted from nitrogen-starved yeast cells carrying none (1), 20S RNA virus alone (2), 23S RNA virus alone (4) or both viruses together (3), was separated in an agarose gel and visualized by Ethidium bromide-staining. 20S and 23S RNAs, together with the rRNAs are indicated.

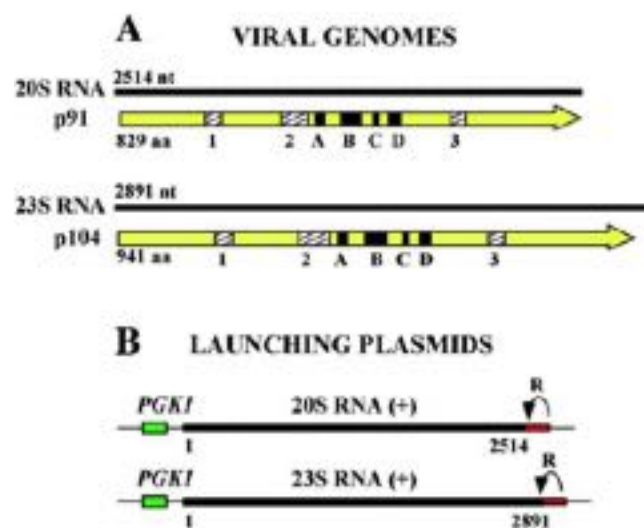


Fig. 2 Genomic organization of 20S and 23S RNA viruses (A) and their launching plasmids (B). (A). Diagrams of 20S and 23S RNAs and the proteins encoded by them, p91 and p104, respectively. (A–D) represents motifs conserved among RdRps from positive strand and dsRNA viruses and 1–3 indicates amino acid stretches conserved between p91 and p104. (B). The complete cDNA of 20S and 23S RNA genome is inserted downstream of the constitutive *PGKI* promoter in a yeast expression vector in such a way that positive strands are transcribed from the promoter. The HDV ribozyme is adjoined directly to the 3' end of the viral genome.

synthesis. The antigenomic (or negative strand) RNAs have no coding capacity for protein and are present at much lower copy numbers (a few per cent) compared to the genomic (or positive strand) RNAs under the induction conditions. The double stranded forms of 20S and 23S RNAs are known and called W and T, respectively. These double-stranded RNAs (dsRNAs) accumulate when the cells are grown at 37°C, a rather high temperature for yeast (the optimal temperature for growth is about 28°C). These dsRNAs are not intermediates of replication but byproducts. Replication proceeds from a positive strand to a negative strand and then to a positive strand.

The proteins encoded in the viral genomes are not processed to produce smaller fragments with distinct functions. Both proteins contain amino acid motifs well conserved among RdRp from positive strand and double-stranded RNA viruses (**Fig. 2(A)**). Their RdRp consensus motifs are most closely related to those of RNA bacteriophages such as *Qβ*. In addition, p91 and p104 share stretches of amino acid sequences (denoted by 1–3 in **Fig. 2(A)**) in the same order throughout the molecules, indicating a close evolutionary relationship between these two viruses.

Ribonucleoprotein Complexes as a Viral Entity

Yeast is also a natural host for dsRNA totiviruses, called ScV-L-A and ScV-L-BC. Like narnaviruses, totiviruses have no extracellular transmission pathway. However, these viruses have *gag* and *pol* genes, and their dsRNA genomes are encapsidated into intracellular

viral particles. In contrast, 20S and 23S RNA viruses have no capsid genes to form virion structures. Then, how do these viruses exist inside the cell and establish a persistent infection without a protective coat? Earlier studies demonstrated that 20S RNA migrated as “naked RNA” in sucrose gradients. Furthermore, deproteination with phenol had no apparent effect on its mobility. Because protein provides a large part of the molecular mass in virions, these data clearly indicate that narnaviruses are capsidless. When specific antibodies against their RdRps became available, however, it was realized that each RNA genome is associated with its RdRp and this interaction is specific; that is, p91 is associated only with 20S RNA and p104 only with 23S RNA. These ribonucleoprotein complexes reside in the cytoplasm and are not associated with the nucleus, mitochondria, or intracellular membranous structures. Further studies indicated that most of the positive strands of 20S and 23S RNA viruses under induction conditions are associated with their own RdRps in a 1:1 stoichiometry. The complexes lack host proteins and can be formed even in *E. coli* if the RNAs and their RdRps are expressed from vectors. These complexes are called “resting complexes” to distinguish them from the “replication complexes” described in the following section. Negative strands are present at much lower amounts compared to positive strands. Available data indicate that they also form complexes with their own RdRps. These findings suggest that the formation of ribonucleoprotein complexes between the viral RNA and its RdRp is important for the life of 20S and 23S RNA viruses.

Replication Intermediates

Lysates prepared from virus-induced cells have an RNA-dependent RNA polymerase activity. The activity is insensitive to actinomycin D or α -amanitin, thus independent of a DNA template. The majority of *in vitro* products are positive strands of 20S RNA. Synthesis of negative strands accounts for a few percent compared to that of positive strands, thus reflecting the positive/negative strand ratio in the lysates. There is no, or very little, *de novo* synthesis *in vitro*. Therefore, radioactive nucleotides are unevenly distributed into 20S RNA positive strand products with more incorporation into the 3′ end region. Replication complexes that synthesize 20S RNA positive strands have a single-stranded RNA backbone and migrate in native agarose gels as a broad band corresponding to ssRNA in the size ranging from 2.5 to 5 kb long. These complexes consist of a full-length negative strand template (2.5 kb) and a nascent positive strand of less than unit-length, probably held together by the polymerase machinery. Deproteination with phenol converts them to double-stranded RNA. Therefore, W dsRNA is not a replication intermediate but a byproduct. It is likely that the high temperature (37°C) for growth may destabilize replication complexes, thus resulting in the accumulation of double-stranded RNA. Upon completion of RNA synthesis *in vitro*, the positive strand products as well as the negative strand templates are released from replication complexes. It is likely that, in the cell, the released negative strands are immediately recruited to another round of positive strand synthesis, because the majority of negative strands in lysates are present in replication complexes engaging in the synthesis of positive strands. Interestingly, both positive and negative strands released from replication complexes are associated with protein. Because replication complexes contain at least one p91 molecule per complex, p91 is a good candidate for the protein. Although resting complexes can be formed in *E. coli*, lysates prepared from these cells showed no viral RNA polymerase activity. It is likely that formation of replication intermediates requires a host factor(s).

Generation of Narnaviruses *in Vivo*

As mentioned earlier, the presence of narnaviruses does not render phenotypic changes to the host. This has hindered studies on replication or virus/host interactions using yeast genetics. This obstacle has been overcome by developments in generating 20S and 23S RNA viruses *in vivo* from a yeast expression vector (Fig. 2(B)). In either case, the complete viral cDNA was inserted in the vector downstream of a constitutive promoter in such a way that positive strands can be transcribed from the promoter. The 3′ end of the viral sequence was directly fused to the hepatitis delta virus (HDV) antigenomic ribozyme. Therefore, intramolecular cleavage by the ribozyme will create transcripts *in vivo* having the 3′ termini identical to the viral 3′ end. The efficiency of virus launching is high. Twenty to 70% of the cells transformed with the vector generated the virus. The primary transcripts expressed from the vectors have non-viral sequences (about 40 nt) at the 5′ ends. The generated viruses, however, possessed the authentic viral 5′ ends without the extra sequences. It is likely that the 5′ non-viral extension was eliminated by a 5′ exonuclease. Using these launching systems, it has been demonstrated that each RdRp is essential and specific for replication of its own viral RNA. Because negative strands cannot be decoded to the RdRps, vectors in which the viral cDNAs were reversed failed to generate the virus. These negative strand-expressing vectors, however, successfully generated narnaviruses, if active polymerases were provided *in trans* from a second vector. Therefore, both 20S and 23S RNA viruses can be generated from either positive or negative strands expressed from a vector.

Cis-Acting Signals for Replication

20S and 23S RNA genomes share the same 5 nt inverted repeats at the 5′ and 3′ termini (5′-GGGGC...GCCCC-OH). Extensive analysis was done modifying each nucleotide at the 3′ ends. It was found that the 3rd and 4th Cs from the 3′ termini are essential for replication in both viruses. While the 3′ terminal and penultimate Cs can be eliminated or changed to other nucleotides

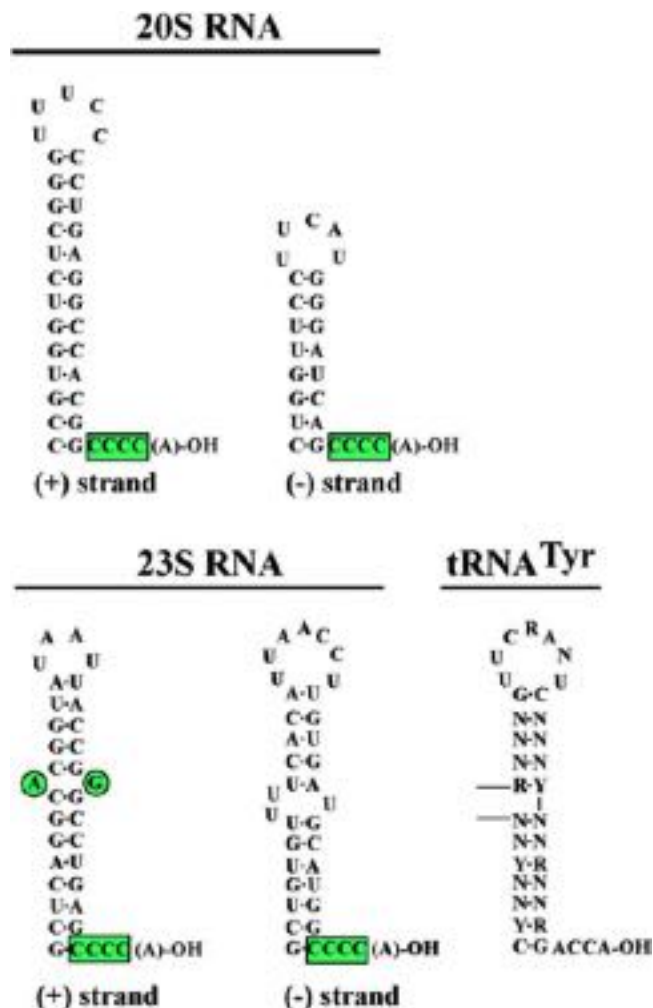


Fig. 3 Comparison of the 3' terminal secondary structures in the positive (+) and negative (–) strands of 20S and 23S RNA viruses, with the top half domain of tRNA^{Tyr}. The non-templated A residues at the viral 3' termini are placed in parenthesis. The consecutive four Cs essential for replication are boxed (green). A second *cis*-signal (the mismatched pair of purines) present in the positive strand of 23S RNA virus is circled (green). Y, R, and N stand for pyrimidine, purine, and any base, respectively.

without affecting virus generation, the generated viruses recovered the wild type Cs at the termini. Therefore, the consecutive four Cs at the 3' terminus are essential for these viruses (Fig. 3).

In contrast, the G immediately upstream the 4 Cs from the 3' end is dispensable for replication in both viruses. 23S RNA virus requires an additional 3' *cis* signal for replication. The stem-loop structure proximal to the 3' end contains a mismatched pair of purines in the stem (Fig. 3). This mismatched pair is essential for replication but the virus tolerates any combination of purines at the pair. On the other hand, changing the purines to pyrimidines or eliminating one of the purines at the mismatched pair blocked virus generation. The distance between the mismatched pair and the 3' terminal four Cs and/or their spatial configuration appears to be critical, because shortening or increasing the length of the stem between the two sites by more than one base pair abolished virus launching. It is not known whether 20S RNA virus has a similar *cis* signal in the stem-loop structure proximal to the 3' end. The first G upstream the 4 Cs from the 3' end is located at the bottom of the stem structure. This G, as mentioned earlier, can be changed to another nucleotide without impairing replication, as far as the modified nucleotide is hydrogen-bonded at the bottom of the stem.

The negative strands of 20S and 23S RNA viruses also possess four consecutive Cs at the 3' ends. Using the two-vector system mentioned above, it has been found that the 3rd and 4th Cs from the 3' end are essential for replication. Similar to the positive strands, the 3' terminal and penultimate Cs can be eliminated or changed to other nucleotides without affecting virus generation and the generated viruses recovered the wild type four Cs at the 3' ends. Therefore, the consecutive four Cs at the 3' end of the negative strand are again a *cis* signal for replication. The 5' ends of viral positive strands have not been analyzed extensively. Elimination of the 5' terminal G or changing it to other nucleotide had no effect on virus generation and the generated viruses recovered this G at the 5' ends.

Cis-Signals for Formation of Ribonucleoprotein Complexes

Narnaviruses, as mentioned earlier, exist as ribonucleoprotein complexes in the host cytoplasm. In the absence of the HDV ribozyme, RNA transcribed from the launching vectors failed to generate viruses because of the presence of non-viral extensions at the 3' end. The transcripts, however, can be decoded to viral polymerases and the polymerases can form complexes *in vivo* with the transcripts, thus providing an assay system to analyze *cis*-signals for formation of ribonucleoprotein complexes. By immunoprecipitation with antiserum specific to p104, it has been found that the bipartite 3' *cis*-signal for replication (more specifically, the mismatched pair of purines and the 3rd and 4th Cs from the 3' end) is essential for 23S RNA positive strand to form complexes with the polymerase. In the case of 20S RNA virus, a similar *in vivo* assay indicates that the 3' *cis* signal for replication (in particular the 3rd and 4th Cs from the 3' end) is also important for formation of ribonucleoprotein complexes with p91. When isolated resting complexes of 20S RNA virus were analyzed *in vitro*, however, it was found that p91 physically interacts with 20S RNA at three different sites, not only at the 3' end but also at the 5' terminal and central regions of the molecule. The 5' binding site is located at the second stem-loop structure from the 5' terminus and is essential for replication. The central site interacts less strongly to p91 than the other two sites, and remains to be defined further. Computer-predicted analysis indicates that the 5' and 3' termini of 20S RNA (and 23S and 23S RNA and 23S RNANA) are brought together into close proximity by a long distance RNA/RNA interaction (Fig. 4). This may allow a single molecule of p91 to interact simultaneously with both ends of 20S RNA genome in a resting complex.

Narnavirus Persistence in the Host

mRNA degradation in yeast, like in other eukaryotes, is initiated by shortening the 3' poly(A) tail followed by decapping at the 5' end. Then the decapped mRNA is degraded by the potent Xrn1p/Ski1p 5' exonuclease as well as by a 3' exonuclease complex called exosome. The RNA genomes of narnaviruses, as mentioned earlier, have no 3' poly(A) tails, thus resembling intermediates of mRNA degradation. This suggests that these RNA genomes are vulnerable to the exonucleases involved in mRNA degradation. In fact, the copy numbers of 20S and 23S RNAs increase several-fold in strains having mutations in *SKI* genes such as *SKI2*, *SKI6*, and *SKI8*. These mutations were originally identified by their failure in lowering the copy numbers of L-A dsRNA totivirus and its satellite RNA M. It is known that the *SKI2*, *SKI6*, and *SKI8* gene products are components or modulators of the exosome. These

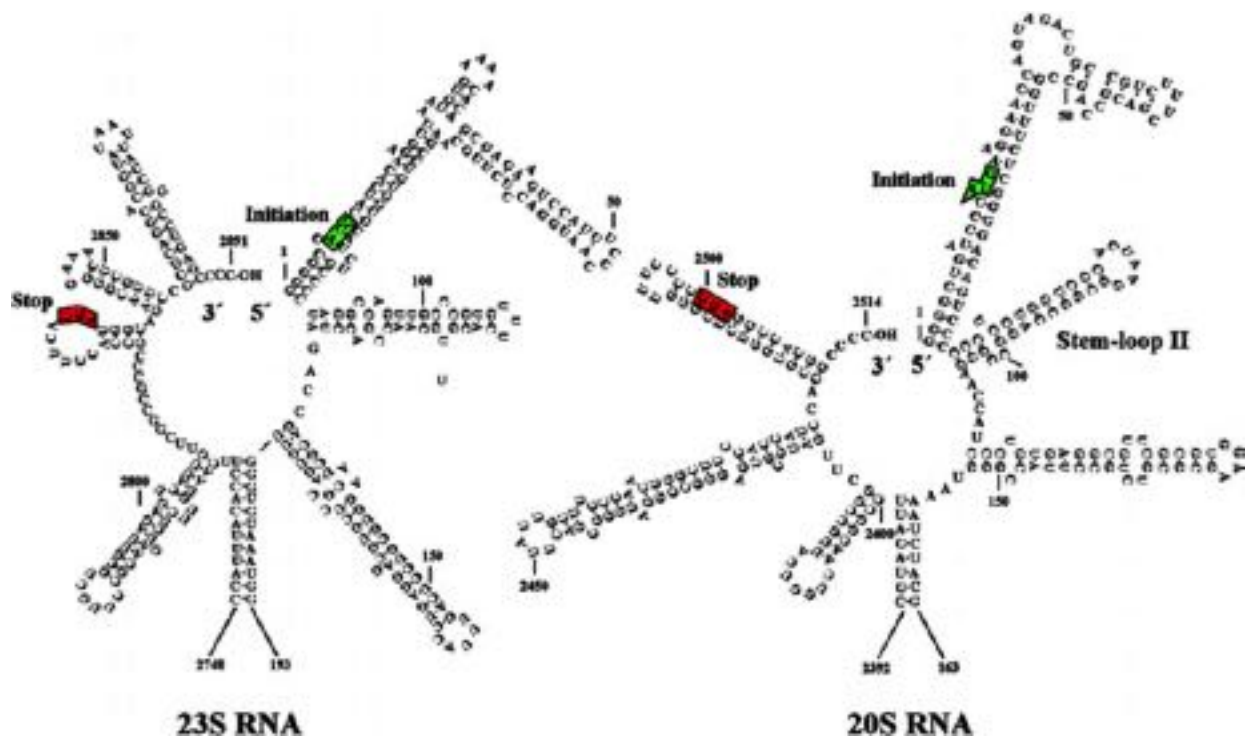


Fig. 4 Secondary structures at the 5' and 3' end regions of 23S and 20S RNA positive strands, as predicted by the MFOLD program. The AUG initiation codons (green) and the stop codons (red) for p104 and p91 are boxed. Stem-loop-II where the 5' binding site for p91 is located is shown. Note that about 150 nt from the ends in each viral genome there are inverted repeats of 8–12 nt long that bring both 5' and 3' ends to a close proximity.

observations suggest that the 3' end of the viral genome is constantly nibbled by 3' exonucleases. Therefore, one of the reasons for narnaviruses to form ribonucleoprotein complexes may be to protect their 3' ends from exonuclease cleavage. The fact that the 3rd and 4th Cs from the 3' end are important to form complexes in both viruses fits this hypothesis because binding of the RdRp to these nucleotides would block progression of the exonuclease and protect the internal region. As described earlier, mutations introduced at the terminal and penultimate positions at the 3' end had no deleterious effects on virus launching and the generated viruses recovered the wild type sequences. This suggests that the terminal and penultimate positions at the 3' ends are not only vulnerable to cleavages but also accessible to the repair machinery. The 3' ends of these viruses may undergo constant turnover at these positions.

It is not known whether the 3' end repair is carried out by the replicase machinery during the replication process, or by host enzymes. The following evidence may favor the latter case. The 3' terminal structures of 20S and 23S RNAs resemble a half of tRNA, the so-called "top-half" domain, consisting of the acceptor stem and T stem (Fig. 3). The domain provides the determinants necessary for specific interactions with tRNA-related enzymes such as the tRNA nucleotidyltransferase (CCA-adding enzyme). This raises the possibility that 20S and 23S RNAs nibbled at the 3' ends by 3' exonucleases are repaired to the wild type sequences by the CCA-adding enzyme. The fact that 15%–30% of both positive and negative strands of 20S and 23S RNAs possess an unpaired A at the 3' ends supports this possibility. Furthermore, that the 3' repair is confined to the terminal and penultimate positions is consistent with the catalytic activity expected for the CCA-adding enzyme.

Concerning the 5' end, the first nucleotides in both 20S and 23S RNAs are four consecutive Gs (Fig. 4). It is known that oligo G tracts inhibit progression of the Xrn1/Ski1 5' exonuclease. Furthermore, these consecutive Gs are buried at the bottom of a long stem structure in both viruses. These features suggest that 20S and 23S RNAs by themselves are resistant to the 5' exonuclease. This was confirmed by modifying the 5' terminal stem structure of 20S RNA. Destabilizing the stem structure severely affected 20S RNA virus generation from a vector in wild type cells, but the same vector successfully generated the virus in *SKI1*-deleted cells. Therefore, the 5' terminal stem structure is critical for the virus to evade *SKI1* surveillance. The initiation codon of p91 is located in the middle of the 5' terminal stem structure. If p91 bound to this stem in the ribonucleoprotein complex, then such a stable binding would interfere with translation of new p91 molecules from the RNA. In this context, it may make sense that the 5' binding site of p91 in the complexes is located at the second stem-loop structure from the 5' end. By binding simultaneously to this site and also to 3' end of the same RNA molecule, p91 may stabilize the long distance RNA-RNA interactions that bring the 5' and 3' ends of the RNA into proximity, and help the RNA to form an organized structure by further interacting with the central binding site. Furthermore, because the host cytoplasm is filled with a great variety of RNA molecules, formation of complexes between the viral polymerase and its template RNA may also facilitate its replication by discriminating non-viral RNAs as templates.

Further Reading

- Buck, K.W., Esteban, R., Hillman, B.I., 2005. Narnaviridae. In virus taxonomy, VIIIth report of the ICTV. Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), London: Elsevier/Academic Press, pp. 751–756.
- Esteban, L.M., Rodríguez-Cousiño, N., Esteban, R., 1992. T double-stranded (dsRNA) sequence reveals that T and W dsRNAs form a new RNA family in *Saccharomyces cerevisiae*: Identification of 23S RNA as the single-stranded form of T dsRNA. *Journal of Biological Chemistry* 267, 10874–10881.
- Esteban, L.M., Fujimura, T., García-Cuellar, M.P., Esteban, R., 1994. Association of yeast viral 23 S. RNA with its putative RNA-dependent RNA polymerase. *Journal of Biological Chemistry* 269, 29771–29777.
- Esteban, R., Fujimura, T., 2003. Launching the yeast 23S RNA narnavirus shows 5' and 3' *cis*-acting signals for replication. *Proceedings of the National Academy of Sciences of the United States of America* 100, 2568–2573.
- Esteban, R., Vega, L., Fujimura, T., 2005. Launching of the yeast 20S RNA Narnavirus by expressing the genomic or anti-genomic viral RNA *in vivo*. *Journal of Biological Chemistry* 280, 33725–33734.
- Esteban, R., Vega, L., Fujimura, T., 2008. 20S RNA narnavirus defies the antiviral activity of *SKI1/XRN1* in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* 283, 25812–25820.
- Fujimura, T., Esteban, R., 2004. Bipartite 3' *cis*-acting signal for replication in Yeast 23 S RNA virus and its repair. *Journal of Biological Chemistry* 279, 13215–13223.
- Fujimura, T., Solórzano, A., Esteban, R., 2005. Native replication intermediates of the yeast 20S RNA virus have a single-stranded RNA backbone. *Journal of Biological Chemistry* 280, 7398–7406.
- Kadowaki, K., Halvorson, H.O., 1971. Appearance of a new species of ribonucleic acid synthesized in sporulation cells of *Saccharomyces cerevisiae*. *Journal of Bacteriology* 105, 826–830.
- Matsumoto, Y., Wickner, R.B., 1991. Yeast circular RNA replicon: Replication intermediates and encoded putative RNA polymerase. *Journal of Biological Chemistry* 266, 12779–12783.
- Rodríguez-Cousiño, N., Esteban, L.M., Esteban, R., 1991. Molecular cloning and characterization of W double-stranded RNA, a linear molecule present in *Saccharomyces cerevisiae*: Identification of its single-stranded RNA form as 20S RNA. *Journal of Biological Chemistry* 266, 12772–12778.
- Solórzano, A., Rodríguez-Cousiño, N., Esteban, R., Fujimura, T., 2000. Persistent yeast single-stranded RNA viruses exist *in vivo* as genomic RNA:RNA polymerase complexes in 1:1 stoichiometry. *Journal of Biological Chemistry* 275, 26428–26435.
- Vega, L., Sevillano, L., Esteban, R., Fujimura, T., 2014. Resting complexes of the persistent yeast 20S RNA narnavirus consist solely of the 20S RNA viral genome and its RNA polymerase p91. *Molecular Microbiology* 93, 1119–1129.
- Wejksnora, P.J., Haber, J.E., 1978. Ribonucleoprotein particle appearing during sporulation in yeast. *Journal of Bacteriology* 134, 246–260.
- Wesolowski, M., Wickner, R.B., 1984. Two new double-stranded RNA molecules showing non-Mendelian inheritance and heat inducibility in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 4, 181–187.

Phlegiviruses (Unassigned)

Karel Petrzik, Biology Center CAS, Institute of Plant Molecular Biology, České Budějovice, Czech Republic

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Although the first reported virus-associated fungal disease (La France disease) was described in the basidiomycete of the cultivated button mushroom, *Agaricus bisporus*, the vast majority of currently known mycoviruses are derived from ascomycetous fungi. Despite their rarity, the basidiomycete-infecting viruses are classified as belonging to several virus families including *Endornaviridae*, *Narnaviridae*, *Partitiviridae* and *Megabirnaviridae*. The exception is phlegiviruses, wherein all currently known viruses infect basidiomycetous hosts. Like many other mycoviruses, phlegiviruses are not associated with any adverse impact on fruiting bodies or mycelia. *Lentinula edodes* virus (LeV) dsRNA has been found in malformed cultures of shiitake mushroom and also in healthy-looking fruiting bodies and actively growing mycelia.

Although phlegiviruses have yet to be officially classified by the International Committee on Taxonomy of Viruses (ICTV), the genus “Phlegivirus” was proposed to accommodate previously described “phlegiviruses” belonging to 5 species (Table 1). Justification for creating the genus “Phlegivirus” is based on biological properties, percent sequence identity and similarity in genome size and organization. Furthermore, members of the proposed genus are distinctly different from other ICTV-recognized virus genera.

Virion Properties

Although repeated attempts were made to isolate virus particles from purified preparations of LeV, PgLV1, and RfV1, no viral particles have been observed by electron microscopy. A purification protocol using polyethylene glycol precipitation, ultracentrifugation, and sucrose gradient centrifugation has been successful in purifying dsRNA but not in isolating virions. At about 3539 nm, the length of the observed linear dsRNA corresponds to the genome size of about 11.8 kb for this virus. Most probably,

Table 1 Members of the proposed genus “Phlegivirus”

Virus	Abbreviation	GenBank accession no.	Genome size nt	Reference
Lentinula edodes virus	LeV	AB429556	11,282	Magae (2012)
Phlebiopsis gigantea large virus 1	PgLV1	NC_013999	11,563	Kozlakidis <i>et al.</i> (2009)
Pterostylis sanguinea leaf-associated virus A	PsVA	KU291925	^a 10,716	Ong <i>et al.</i> (2018)
Rhizoctonia fumigata virus 1	RfV1	KM657432	^a 9907	Li <i>et al.</i> (2015)
Thelephora terrestris virus 1	TtV1	NC_028921	^a 10,316	Petrzik <i>et al.</i> (2016)

^aPartial sequence.

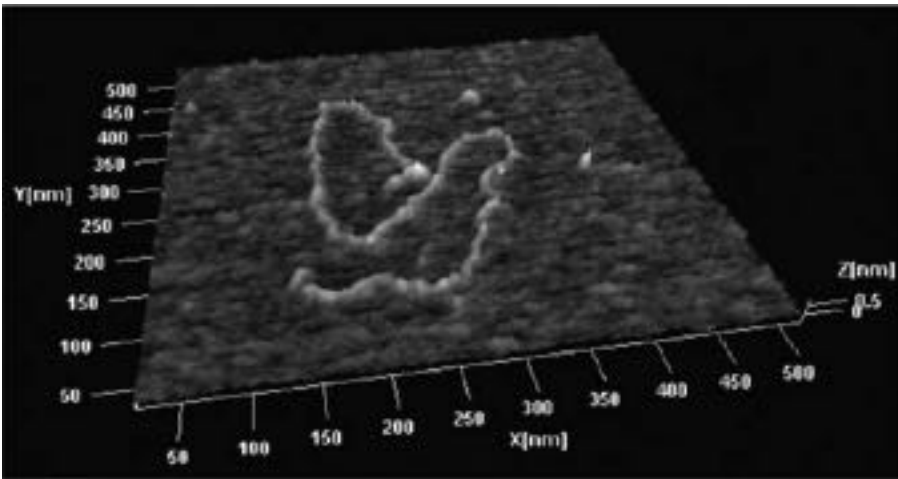


Fig. 1 Atomic force microscopy image of a single linear *Lentinula edodes* virus dsRNA. Picture from: Magae, Y., 2012. Molecular characterization of a novel mycovirus in the cultivated mushroom, *Lentinula edodes*. *Virology Journal*. 9, 60.

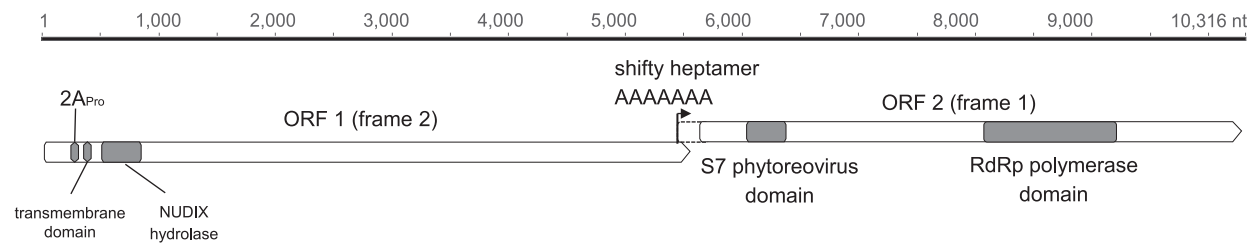


Fig. 2 Genome arrangement and motifs of *Thelephora terrestris* virus 1. Schematic representation of the genomic organization of TtV1 shows the presence of two ORFs (ORF1 and ORF2). The dotted-line box indicates a possible extension of ORF1 via a translational frameshift mechanism. Positions of 2A-like motif, transmembrane motif, NUDIX hydrolase, S7 phytoreovirus domain, and RdRp conserved domain are indicated by boxes.

Virus	AC		↓
PgLV1	YP_003541122.1	A L A D N I G L V	G I E T N P G P
TtV1	YP_009209481.1	V K A D C V G L V	G V E T N P G P
Rosellinia necatrix mycovirus 2-W1032/S6	BAM36407	P K R D L T V D G	D V E K N P G P
Fusarium graminearum hypovirus 1 gp2	YP_009011065	V I A R N A D D A	D V E L N P G P
Fusarium graminearum hypovirus 1 gp2	YP_009011065	V D P Y G A D D H	D V E K N P G P
Fusarium graminearum hypovirus 2 JS16	YP_009130646	K R A R D S S M A	D I E P N P G P
Consensus mycoviruses			⁶ / ₀ E N P G P

Consensus positions are highlighted in gray. Arrow indicates position of putative cleavage between the 2A^{Pro} and the N-terminal proline of 2B protein.

Fig. 3 Comparison of the amino acid sequences flanking 2A^{Pro}-like motifs in mycoviruses.

Virus	AC	
LeV	BAG71789.2	G I G G K I D S R E D L I A G L N R R V M E E V S I D I
PgLV1	YP_003541122.1	G V G G S L E L G E D P S A G I A R E V R E E V G I D I
PsVA	AOX47547.1	L I G G K C E E G E S L E A G L V R E V R E E V G M D I
RfV1	AJE29744	F P G G K V D A G E S I Q D A L K R E V M E E T G L T V
TtV1	YP_009209481.1	G V G G S A E K D E S P S D T L T R E A K E E V G I D I
Consensus phlegiviruses		* * * * *

Conserved positions are marked with asterisks, consensus amino acid positions of NUDIX motif are highlighted in gray.

Fig. 4 NUDIX hydrolase motif (pfam00293) in phlegiviruses.

	I	II	III	IV	V	VI	VII	VIII
LeV	LSGRA 73	LAPTGA 52	GLRSRQIV 59	DYADFNFLH 65	LWSGWRRTTSVINNVFNEVYG 21	NGDD 37	EYTR 13	RGIAS
PgV1	LVGRA 79	VAPRGS 52	GLRLRQII 57	DYADFNFLH 66	LWSGWRRTTMINNMTNLVYN 21	NGDD 37	EYLR 13	RSCAS
PsVA	LAGRI 73	TAPMGS 52	QAKLRQIL 57	DYADFNFLH 65	LWSGWRRTTSAINNTLNFVYG 21	NGDD 37	EYLR 13	RSISS
RfV1	LAGRY 72	VGASGS 50	G-KLRQLL 61	DYADFNITH 65	LWSGWRSTMFNFILFNLYLA 21	MGDD 37	EFLR 12	RSICG
TtV1	LVGRA 79	IAPRGS 52	GLRLRQII 57	DYADFNFLH 66	LWSGWRRTTMINNMTNLVYN 21	NGDD 37	EYLR 13	RTCAS
	* **	:: ::	:...*:.	***** :*	*****: * :*	:***	*::*	*::..

Identical residues are indicated by asterisks. Conserved and semi-conserved aa residues are indicated by colons and dots, respectively.

Fig. 5 Alignment of the conserved motifs (pfam02123) of RdRp for phlegiviruses.

phlegiviruses are present in hosts as naked dsRNA and are somehow stabilized, perhaps by host proteins. Furthermore, the phlegivirus genomes do not code for capsid protein (CP) (Figs. 1 and 2).

Genome Organization

The monopartite dsRNA genome of phlegiviruses is about 11 kbp long and contains two large ORFs in different frames on the genomic plus strand. ORF1 encodes a protein with calculated molecular mass of about 200 kDa, and ORF2 encodes a protein with molecular mass in the range of 140–170 kDa. There are no in-frame stop codons upstream from the ORF2 start codon, but there is a ribosomal -1 frameshift sequence (AAAAAAA in TtV1, GGGAAAC in LeV, GGGUUUU in PgLV1, and AAAUUUC in RfV1) with a pseudoknot structure. This raises the possibility that ORF2 could be translated as a fusion protein with ORF1. The ribosomal

Table 2 Summary of genome features of phlegiviruses

Virus	2A cleavage	NUDIX	S7 phytovereovirus	Slippery motif	ORF1 (kDa)	ORF2 (kDa)
LeV	No	Yes	No	GGGAAAC	218	162
PgLV1	Yes	Yes	Yes	GGGUUUU	219	158
PsVA	No	Yes	Yes	No	208	140 ^a
RfV1	No	Yes	No	AAAUUUC	198	144
TtV1	Yes	Yes	Yes	AAAAAAA	202	174

^aPartial sequence.
*See Table 1 for full names of viruses.

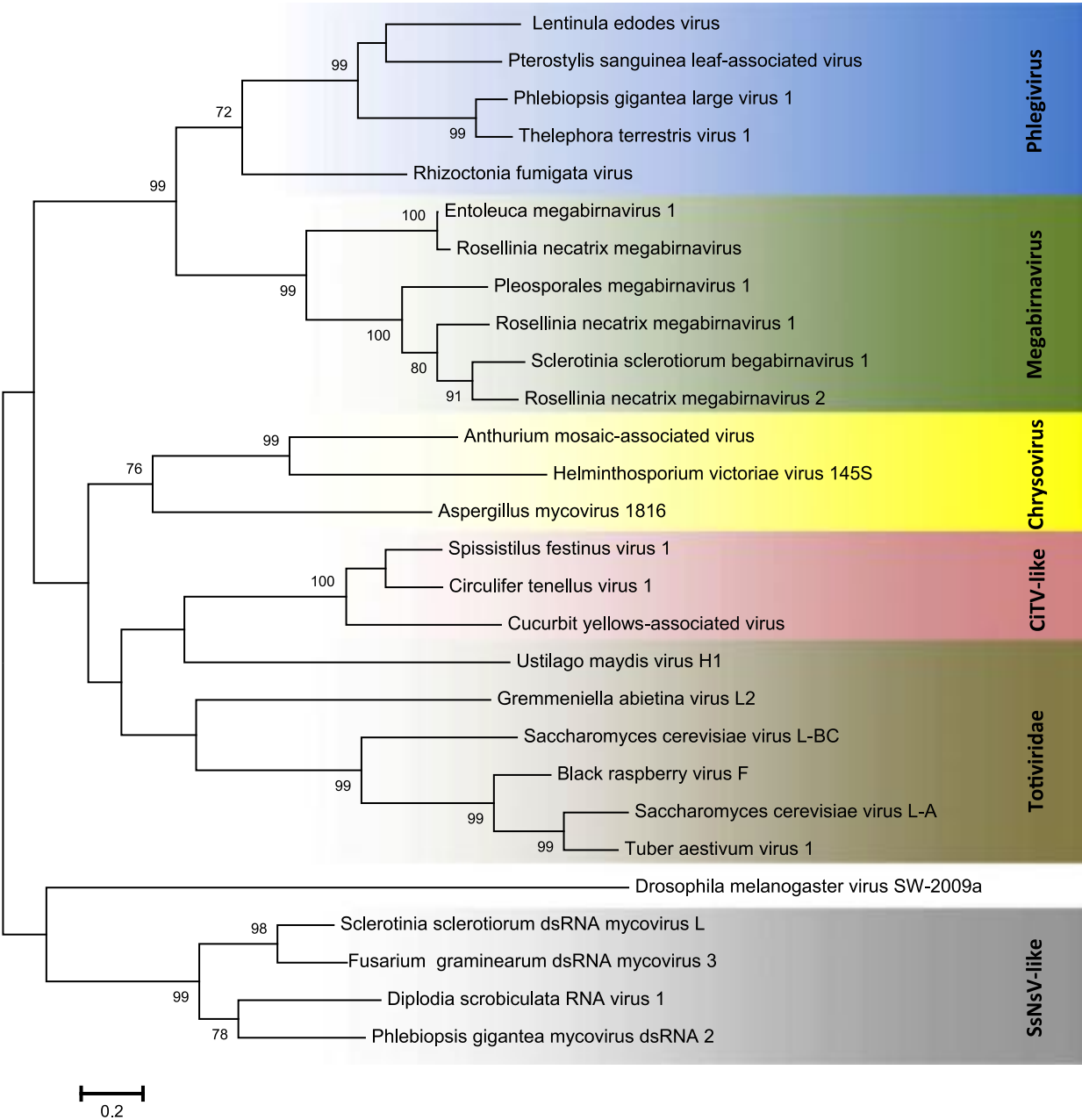


Fig. 6 An unrooted maximum-likelihood tree of phlegiviruses and related mycoviruses. Scale bar corresponds to 0.2 amino acid substitution per site. The tree is based on the amino acid sequences of the region between motifs I and VII of the RdRp. Numbers on nodes show bootstrap values above 70% (1000 replicates).

Table 3 Pairwise amino acid identity (%) of complete RdRp and ORF1 of phlegiviruses

	LeV	PgLV1	PsVA	RfV1	TtV1
LeV*		20.9	20.1	11.6	20.2
PgLV1	27.8		19.4	13.4	43.7
PsVA	27.6	29.9		14.2	19.6
RfV1	20.9	20.9	12.3		12.2
TtV1	25.4	55.2	27.0	19.0	

RdRp below diagonal, ORF1 above diagonal.
See [Table 1](#) for full names of viruses.

frameshift sequence has not been detected in PsVA, but stable stem-loop structures that could assist in pausing translating ribosomes were predicted in this virus close to the slippery site.

Polyprotein processing is intrinsic for many viruses. In phlegiviruses PgLV1 and TtV1, the picornavirus proteinase 2A^{Pro}-like motif was found on the N-terminal part of ORF1. This supports a hypothesis that the protein could be further processed by cleavage. This motif comprises the seven aa residues G/DxExNPG and the N-terminal proline of 2B protein ([Fig. 3](#)). The motif has been found also in *Rosellinia necatrix* mycovirus 2-W and in *Fusarium graminearum* hypoviruses 1 and 2. Based on the expected activity of 2A^{Pro}, the N-terminal 301 and 90 aa fragments of PgLV1 and TtV1 polyprotein, respectively, could be released.

The ORF1-encoded protein is unrelated to any protein in GenBank. In all phlegiviruses, however, there is a region having similarity to the NUDIX hydrolase domain (pfam00293) that is close to the N-terminal part of ORF1. NUDIX hydrolases constitute a diverse superfamily of pyrophosphatases that are widespread among eukaryotes, bacteria, archaea, and some viruses. They catalyze hydrolysis of nucleoside diphosphates with various substrate specificities and function to clear the cell of potentially deleterious endogenous metabolites. Previously, NUDIX domain has been found only in dsDNA poxviruses, but the NUDIX domain in phlegiviruses differs from that of the dsDNA viruses and forms a distinct cluster in phylogenetic analysis. With the exception of RfV1, one or more transmembrane domains are predicted on the N-terminal part of the ORF1 protein upstream from the NUDIX domain in all phlegiviruses.

ORF2 encodes a protein with calculated molecular mass of 140–170 kDa and having the RNA-dependent RNA polymerase (RdRp) catalytic domain. An S7-phytoreovirus-like domain is located in PgLV1, PsVA, and TtV1 close to the N-terminus of the ORF2-encoded protein. The phytoreovirus S7 domain is widely distributed in diverse dsRNA viruses within proteins known to be viral core proteins having nucleic acid binding activities. One can speculate that ORF2 encodes a structural protein that binds and protects the genomic RNA.

The 5'UTR region is 25 to 1154 nt long and the 3'UTR is 29–113 nt long. Disparity in the UTR sequences could be caused by deletions in the 5'UTR sequences. There is no polyadenylation or secondary structure on the 3' ends ([Figs. 4](#) and [5](#), [Table 2](#)).

Biological Properties

All phlegiviruses are associated with infections of basidiomycetes. In the case of PsVA, the mycorrhizal fungi have not been identified to species level, but they do sequence closely to a *Ceratobasidium* sp. and a *Rhizoctonia* sp. Like many other mycoviruses, phlegiviruses are not associated with an adverse impact on fruiting bodies or mycelia. LeV dsRNA has been found in malformed cultures of *Lentinula edodes* and also in healthy-looking fruiting bodies and actively growing mycelia. No correlation has been determined between LeV incidence and growth rate of the mycelium. Furthermore, the fact that more than 40% of monokaryotic progeny originating from the basidiospore of the fruiting body have been shown to contain LeV dsRNA indicates the persistence of dsRNA during sexual reproduction. The absence of capsid protein in phlegiviruses is speculated to be related to the feeding habits of the bark- and fungus-feeding beetles and oribatid mites (Acari), which can efficiently transfer the virus-infected fungus to a new environment. Thereafter, the CP would no longer be necessary for the virus to exit from the host cell, for its protection from the environment, and for it to create a new infection.

Evolutionary Relationships among Phlegiviruses

All five phlegiviruses have almost identical genome arrangements and are monophyletic in a maximum-likelihood phylogenetic tree inferred from the RdRp region between motifs I and VII of the polymerase ([Fig. 6](#)). ClustalW comparison has shown 41%–54% nt identity across the genomes. Amino acid identity among the homologous proteins was 12%–44% for ORF1 proteins and 12%–55% for RdRps ([Table 3](#)).

Megabirnaviruses (family Megabirnaviridae) are closely related to phlegiviruses, while chrysoviruses (family *Chrysoviridae*) and totiviruses (family *Totiviridae*) represent more distant groups as revealed by phylogenetic analysis based on RdRp. Megabirnaviruses differ, however, in genome components and their arrangement. They have two genome segments separately encapsidated in

isometric particles. dsRNA1 encompasses two overlapping ORFs. ORF1 encodes the CP and ORF2 encodes the RdRp, which is expressed as a fusion product with CP, probably via ribosomal frameshifting.

Future Perspectives

It seems highly probable that phlegiviruses translate their genome as polyproteins having molecular mass exceeding 350 kDa. The genomes of potyviruses, which possess genomic nucleic acids of similar length (about 10 kb) but positive ssRNA, is translated as a polyprotein that is cleaved by virus-coded proteinases up to 10 mature proteins. Neither protease motif nor putative cleavage sites have been identified, however, on either ORF1 or ORF2 of phlegiviruses, and thus the presence of a presumed cleavage enzyme (if one exists) remains obscure. The 7aa long motif resembling that of picornavirus proteinase 2A^{Pro} cleavage site has been found in two out of six viruses only, which makes it improbable that cleavage via this motif is important and common for this group.

Other than the RdRp, the presence of the NUDIX domain across all members of the proposed genus is the only constant feature of phlegiviruses indicating their origin from a common ancestor and supporting the phylogenetic relationships as inferred from the RdRp sequence. The role of this domain in the life-cycle of the viruses still needs to be clarified.

Further Reading

- Donnelly, M.L.L., Hughes, L.E., Luke, G., *et al.*, 2001. The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. *Journal of General Virology* 82, 1027–1041.
- Ghabrial, S.A., Castón, J.R., Jiang, D., Nibert, M.L., Suzuki, N., 2015. 50-plus years of fungal viruses. *Virology* 479–480, 356–368.
- Kozlakidis, Z., Hacker, C.V., Bradley, D., *et al.*, 2009. Molecular characterisation of two novel double-stranded RNA elements from *Phlebiopsis gigantea*. *Virus Genes* 39, 132–136.
- Li, Y., Xu, A., Zhang, L., *et al.*, 2015. Molecular characterization of a novel mycovirus from *Rhizoctonia fumigata* AG-Ba isolate C-314 Baishi. *Archives of Virology* 160, 2371–2374.
- Liu, H., Fu, Y., Xie, J., *et al.*, 2012. Evolutionary genomics of mycovirus-related dsRNA viruses reveals cross-family horizontal gene transfer and evolution of diverse viral lineages. *BMC Evolutionary Biology* 12, 91.
- Magae, Y., 2012. Molecular characterization of a novel mycovirus in the cultivated mushroom, *Lentinula edodes*. *Virology Journal* 9, 60.
- McLennan, A.G., 2006. The nudix hydrolase superfamily. *Cellular and Molecular Life Sciences* 63, 123–143.
- Nibert, M.L., 2007. '2A-like' and 'shifty heptamer' motifs in penaeid shrimp infectious myonecrosis virus, a nonsegmented double-stranded RNA virus. *Journal of General Virology* 88, 1315–1318.
- Ong, J.W.L., Li, H., Sivasithamparan, K., *et al.*, 2018. Novel and divergent viruses associated with Australian orchid-fungus symbioses. *Virus Research* 244, 276–283.
- Petrzik, K., Sarkisova, T., Stary, J., *et al.*, 2016. Molecular characterization of a new monopartite dsRNA mycovirus from mycorrhizal *Thelephora terrestris* (Ehrh.) and its detection in soil oribatid mites (Acari: Oribatida). *Virology* 489, 12–19.
- Won, H.K., Park, S.J., Kim, D.K., *et al.*, 2013. Isolation and characterization of a mycovirus in *Lentinula edodes*. *Journal of Microbiology* 51, 118–122.

Plant and Protozoal Partitiviruses (*Partitiviridae*)

Hanna Rose and Edgar Maiss, Leibniz University Hannover, Hannover, Germany

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

A Adenine
aa Amino acids
b Bases
BCV Beet cryptic virus
bp Base pairs
C Cytosine
C. Cryptosporidium
CCCV2 Crimson clover cryptic virus 2
CCV Carrot cryptic virus, Cannabis cryptic virus
CCV1 Carnation cryptic virus 1
CP Coat protein
CsCL Caesium chloride
CSpV1 Cryptosporidium parvum virus 1
DCV2 Dill cryptic virus 2
dsRNA Double-stranded RNA
FCCV *Fragaria chiloensis* cryptic virus
FCV Fig cryptic virus
G Guanine
HTCV2 Hop trefoil cryptic virus 2
ICTV International Committee on Taxonomy of Viruses
Isl. Isolate

NCBI National Center for Biotechnology Information
ORF Open reading frame
PCV Pepper cryptic virus
PeCV Persimmon cryptic virus
PmV1 *Primula malacoides* virus 1
PTGS Post-transcriptional gene silencing
RCCV2 Red clover cryptic virus 2
RCV-1 Rose cryptic virus 1
RdRp RNA-dependent RNA polymerase
RsCV2 *Raphanus sativus* cryptic virus 2
T Thymine
T. *Trifolium*
U Uracil
UTR Untranslated region
VCV *Vicia* cryptic virus
VLP Virus-like-particle
VP Viral protein
WCCV White clover cryptic virus
α Alpha
β Beta
δ Delta

Glossary

Buoyant densities in CsCL Density, equal to that of CsCL. Determined by density gradient centrifugation (also called isopycnic centrifugation), a method to separate molecules in CsCL according to their density.

Mono-, bi- and tripartite Class of genomes: One, two or three individually packaged genome segments of a virus.

Monocistronic (m)RNA that contains one open reading frame with one initiation and one stop codon translated into one protein.

Negative-contrast electron microscopy Method, in which the background is more electron dense as the one of the object of interest, which thus is stained lighter than the background.

Post-transcriptional gene silencing (PTGS) Defense mechanism of plants in which dsRNA, e.g., originated from a virus, is degraded into small-interfering RNA which targets homologous ssRNA.

Root-nodules Structures located at roots of plants forming a symbiotic relationship with nitrogen-fixing bacteria.

Silencing suppressor (VSR) Viral suppressors of gene silencing: Proteins, that are able to inhibit the plant defense mechanism “Post-transcriptional gene silencing” (PTGS) at different steps.

Type species A Type Species is a species whose name is linked to the use of a particular genus name. The genus so typified will always contain the Type Species (Mayo *et al.*, 2002: “The Type Species in virus taxonomy”).

Introduction

Members of the family *Partitiviridae* (partitivirids) appear as small isometric particles with a size of 25–43 nm owning a segmented double-stranded (ds) RNA genome. They cause no symptoms in their hosts, are not mechanically or vectorially transmissible and can infect plants, fungi or protozoa. Partitivirids firstly discovered in 1968/69 by Pullen, were described as spherical, virus-like-particles (VLPs) with a diameter of about 28 nm. These VLPs occurred in apparently healthy sugar beet varieties and beet seeds. It was neither possible to transmit them to several herbaceous test plants via aphids or mechanical methods nor to remove them with heat therapy. During this time, the term “cryptic” was established. Kassanis *et al.* confirmed the wide distribution of this virus in beet species in 1977 and named it beet cryptic virus 1 (BCV1). They also discovered a high seed transmission rate in commercial beet varieties (90%), identifying non-infected beet species and establishing a purification method for BCV1. In 1980 two further VLPs with similar properties were discovered, *Vicia* cryptic virus (VCV) and carnation cryptic virus 1 (CCV1), both showing similar characteristics like BCV1. In 1981, Lisa *et al.* showed that the genome of CCV1 consists of segmented dsRNA. A few years later (1988) it was shown that isolated CCV1 particles own RNA-dependent RNA polymerase

(RdRp) activity catalyzing the synthesis of dsRNA in vitro. During this time period, Natsuaki's research group in Japan also discovered VLPs in plants designated as "temperate viruses", such as spinach temperate virus and beet temperate virus. It turned out that cryptic and temperate viruses are synonymous. In 1984 Kassanis proposed to retain the name module "cryptic" in order to distinguish these viruses from latent viruses, which in contrast to cryptic viruses are usually mechanically or vectorially transmissible. With the isolation of white clover virus 1, 2, and 3 (WCCV1, 2, and 3), Boccardo and team members were able to show that VLPs are indeed cryptic viruses and originate from the plant in which they replicate. The genome of these three viruses consists for each one of two dsRNA segments and confirms previous assumptions of a segmented genome. In the fifth ICTV report in 1991, the family *Partitiviridae* with the genus *Partitivirus* was included for the first time, in which cryptic and temperate viruses were classified. The criteria for this genus were: isometric particles and a genome consisting of two dsRNA species. Two years later, Xie *et al.* found a third cryptic virus in beet and named it beet cryptic virus 3 (BCV3). It was also possible to determine the nucleotide and derived amino acid sequence of one dsRNA segment and to identify it as a putative RdRp with conserved motifs that are also present in RdRps of other viruses. The taxonomy of the *Partitiviridae* changed in 1995–2000 with the sixth and seventh ICTV reports, in that the genus *Chrysovirus*, which contained two fungal infecting viruses, was assigned to the family. With the eighth ICTV Report 2005, this genus formed its own family, the *Chrysoviridae*, since the respective species had clearly characteristics distinguishable from the partitivirids, such as a genome consisting of four dsRNA segments. In 1995–2000, the family *Cryptoviridae*, containing plant-infecting cryptoviruses was removed and its genera *Alphacryptovirus* and *Betacryptovirus* inserted to the *Partitiviridae*. Members of these two genera are serologically and morphologically distinguishable. From then on more and more partitivirids were discovered and sequenced, for example in 2005 WCCV1 by Boccardo *et al.* and in 2007 VCV by Blawid *et al.* A milestone in the decoding of the 3D capsid structure was achieved in 2008 by the research team surrounding Ochoa through transmission electron cryomicroscopy and subsequent 3D reconstruction of the particles of the fungus-infecting gammapartitiviruses *Penicillium stoloniferum* virus S and F (formerly genus *Partitivirus*). In the following two years further structures were analyzed and in general the capsid is composed of 120 subunits consisting of 60 coat protein (CP) dimers in a T1-icosahedral symmetry. In the ninth ICTV report 2012, the new genus *Cryspovirus* was created in the family *Partitiviridae* named after the only protozoan infecting virus *Cryptosporidium parvum* virus 1, which was discovered in 1997 by Khramtsov *et al.*. By the availability of more sequence data from potential members of this family, better phylogenetic comparisons could be made and taxonomic restructuring was needed. Nibert *et al.* explained in 2014 that species from the genus *Partiti-*, *Alpha-*, and *Betacryptovirus* separated into four distinct clades, with two clades (now *Alpha-* and *Betapartitivirus*) simultaneously containing plant and fungal infecting species. As a result, in 2017 the changes were approved by the tenth ICTV report, listing the *Partitiviridae* with the genera *Alpha-*, *Beta-*, *Gamma-*, and *Deltapartitivirus* as well as *Cryspovirus*. To date, partitivirids have been discovered worldwide in fungi and plants (as well as in one protozoan) and due to modern high-throughput sequencing techniques their number is continuously increasing.

Current Taxonomy of the *Partitiviridae*

The family *Partitiviridae* currently comprises the five genera *Alpha-*, *Beta-*, *Gamma-*, and *Deltapartitivirus*, as well as *Cryspovirus*. The alpha- and betapartitiviruses include fungal as well as plant infecting species, whereas deltapartitiviruses infect plants and gammapartitiviruses fungi. The only species in the genus *Cryspovirus* infects protozoa. The phylogenetic tree, based on the amino acid sequences of the partitviral RdRps gives an overview about the taxonomic organization in this encyclopedia.

Overall, the family comprises 45 assigned members whereof 16 belong to the plant infecting species, 15 unassigned members comprising twelve plant infecting partitivirids and two unclassified, related fungal viruses. The unclassified but related viruses within the individual genera are listed separately. [Table 1](#) gives an overview of the composition of the individual genera.

To assign a species into the *Partitiviridae*, genus and species demarcation criteria have been established by the ICTV and need to be fulfilled. One of the criteria is the host, since the members of different genera may have different hosts (plants and fungi) or only one host (plants or fungi or protozoans). With regard to the genome segments and the proteins expressed, each genus has characteristic lengths and sizes. Additionally, there are guidelines for the percentage identity values between RdRp and CP amino acid sequences: Viruses assigned to different genera have a maximum RdRp amino acid sequence identity of less than 24%, different species show a amino acid sequence identity of less than 90% for the RdRp and less than 80% for the CP, respectively.

Until February 2020 there are 28 completely sequenced genomes of plant infecting species available in the NCBI GenBank. [Table 2](#) lists all assigned and unclassified but related plant infecting species with the respective sizes of genome segments and proteins as well as accession numbers.

Species, being suggested a member of the *Partitiviridae* which are not yet listed by the ICTV are shown in [Table 3](#).

Table 1 Overview of the composition of the individual genera of the *Partitiviridae*

Genus	Assigned species	Host: plant/fungal	Unclassified, related species	Host: plant/fungal
<i>Alphapartitivirus</i>	14	4/10	8	6/2
<i>Betapartitivirus</i>	17	7/10	4	–/4
<i>Gammapartitivirus</i>	8	–/8	6	–/8
<i>Deltapartitivirus</i>	5	5/–	4	5/–
<i>Cryspovirus</i>	1	protozoan	3	protozoan

Table 2 Overview of the plant infecting partitivirids with sizes of genome segments, proteins and accession numbers given

<i>Genus/virus name</i>	<i>Size of dsRNA segments [b] and protein [aa (kDa)]</i>	<i>GenBank accession</i>
<i>Alphapartitivirus</i>		
Beet cryptic virus 1 (BCV1)	1: RdRp 2008; 616 (72) 2: CP 1783; 489 (53)	NC_011556.1 NC_011557.1
Carrot cryptic virus (CCV)	1: RdRp 1971; 616 (72) 2: CP 1776; 490 (54)	NC_038824.1 NC_038823.1
Vicia cryptic virus (VCV)	1: RdRp 2,012; 616 (72) 2: CP 1,779; 487 (53)	NC_007241.1 NC_007242.1
White clover cryptic virus 1 (WCCV1)	1: RdRp 1,955; 616 (72) 2: CP 1,708; 487 (54)	NC_006275.1 NC_006276.1
<i>Unclassified – Alphapartitivirus related</i>		
Arabidopsis halleri partitivirus 1	1: RdRp 1,959; 585 (68) 2: CP 1,763; 487 (55)	NC_030889.1 NC_030890.1
Dill cryptic virus 1 isl. IPP_hortorum	1: RdRp 2,013; 616 (73) 2: CP 1,837; 490 (54)	NC_022614.1 NC_022615.1
Diuris pendunculata cryptic virus isl. SW3.3	1: RdRp 2,010; 621 (73) 2: CP 1,806; 496 (55)	JX156424.1 JX891460.1
Raphanus sativus cryptic virus 1	1: RdRp 1,866; 573 (67) 2: CP 1,791; 505 (56) 3: CP 1,778; 505 (55)	NC_008191.1 NC_008190.1 DQ181927
Red clover cryptic virus 1 isl. IPP_Nemaro	1: RdRp 1,936; 616 (73) 2: CP 1,710; 487 (54)	NC_022616.1 NC_022617.1
Rose partitivirus isl. PB	1: RdRp 1,937; 586 (68) 2: CP 1,811; 487 (54)	KU896858.1 KU896859.1
<i>Betapartitivirus</i>		
Cannabis cryptic virus (CCV)	1: RdRp 2,397; 746 (87) 2: CP 2,266; 672 (74)	NC_031134.1 NC_031130.1
Crimson clover cryptic virus 2 (CCCV2)	1: RdRp 2,444; 746 (87) 2: CP 2,354; 674 (75)	NC_038837.1 NC_038838.1
Dill cryptic virus 2 (DCV2)	1: RdRp 2,430; 745 (87) 2: CP 2,354; 673 (75)	NC_021147.1 NC_021148.1
Hop trefoil cryptic virus 2 (HTCV2)	1: RdRp 2,431; 746 (87) 2: CP 2,349; 673 (75)	NC_021098.1 NC_021099.1
Primula malacoides virus 1 (PmV1)	1: RdRp 2,390; 723 (84) 2: CP 2,344; 673 (75)	NC_013109.1 NC_013110.1
Red clover cryptic virus 2 (RCCV2)	1: RdRp 2,430; 745 (87) 2: CP 2,353; 673 (76)	NC_021096.1 NC_021097.1
White clover cryptic virus 2 (WCCV2)	1: RdRp 2,435; 746 (87) 2: CP 2,348; 673 (76)	NC_021094.1 NC_021095.1
<i>Deltapartitivirus</i>		
Beet cryptic virus 2 (BCV2)	1: RdRp 1,575; 475 (54) 2a: VP1 1,598; 426 (49) 2b: VP2 1,522; 393 (45)	NC_038846.1 NC_038845.1 NC_038847.1
Beet cryptic virus 3 (BCV3)	<i>partial genome only</i>	
Fig cryptic virus (FCV)	1: RdRp, 1,696; 472 (54) 2: CP 1,415; 337 (38)	NC_015494.1 NC_015495.1

(Continued)

Table 2 Continued

<i>Genus/virus name</i>	<i>Size of dsRNA segments [b] and protein [aa (kDa)]</i>	<i>GenBank accession</i>
Pepper cryptic virus 1 (PCV1)	1: RdRp 1,563, 479 (54) 2: CP 1,512, 412 (48)	NC_037095.1 NC_037096.1
Pepper cryptic virus 2 (PCV2)	1: RdRp 1,609, 478 (54) 2: CP 1,525, 430 (49)	NC_034159.1 NC_034167.1
<i>Unclassified – Deltapartitivirus related</i>		
Fragaria chiloensis cryptic virus isl. NCGR CFRA 9089	1: RdRp 1,734, 479 (56) 2: CP 1,479, 348 (39) 3: CP? 1,465, 346 (39)	NC_009519.1 NC_009521.1 NC_009520.1
Persimmon cryptic virus isl. SSPI	1: RdRp 1,577, 477 (54) 2: CP 1,491, 415 (47)	NC_017989.1 NC_017988.1
Raphanus sativus cryptic virus 2	1: RdRp 1,717, 477 (55) 2: CP 1,521, 346 (38) 3: VP 1,485, 347 (39)	NC_010343.1 NC_010344.1 NC_010345.1
Rose cryptic virus 1 isl. ShB – 1	1: RdRp 1,749, 479 (56) 2: CP 1,485, 348 (39) 3: VP 1,446, 346 (39)	NC_010346.1 NC_010347.1 NC_010348.1
<i>Unassigned</i>		
Alfalfa cryptic virus 1 (ACV1)	No entry in GenBank	
Carnation cryptic virus 1 (CCV1)	No entry in GenBank	
Carrot temperate virus 1 (CTV1)	No entry in GenBank	
Carrot temperate virus 2 (CTV2)	No entry in GenBank	
Carrot temperate virus 3 (CTV3)	No entry in GenBank	
Carrot temperate virus 4 (CTV4)	No entry in GenBank	
Hop trefoil cryptic virus 1 (HTCV1)	No entry in GenBank	
Hop trefoil cryptic virus 3 (HTCV3)	No entry in GenBank	
Radish yellow edge virus (RYEV)	No entry in GenBank	
Ryegrass cryptic virus (RCV)	No entry in GenBank	
Spinach temperate virus (STV)	No entry in GenBank	
White clover cryptic virus 3 (WCCV3)	No entry in GenBank	

isl.: isolate; VP: viral protein.

Virion Properties

Partitivirids appear as small isometric particles, 25–43 nm in diameter in this encyclopedia. The capsid is composed of 120 subunits consisting of 60 dimers in a T=1-icosahedral symmetry. The viruses have at least two separately encapsidated genome segments encoding a RdRp and a CP. Partitivirids could be detected in numerous plant organs, for example stems, roots, cotyledons and petals. The virions are mainly located in the cytoplasm and less in the nucleus or nucleoli of parenchyma cells. Replication takes place in a semi-conservative mode.

Alphapartitivirus

Alphapartitiviruses are between 25 nm and 40–50 nm in diameter, the type species WCCV1 has an average size of 34 nm. In negative-contrast electron micrographs, the particles appear rounded and ring-shaped because the coloration penetrates in the middle (**Fig. 1**). The buoyant densities in CsCl of WCCV1 virions are 1392 g cm⁻³ and of VCV virions 1.37 g cm⁻³. Encapsidation is performed by a single major CP with a molecular weight of 51–57 kDa. The RdRp has sizes between 68 and 73 kDa. In case of *Raphanus sativus* cryptic virus 1, a third genome segment was found with so far unknown function.

Betapartitivirus

Members of the genus *Betapartitivirus* are between 25 nm and 38 nm in diameter. By negative-contrast electron microscopy WCCV2 appears as rounded particles, which are not penetrated by stain (**Fig. 2**). The buoyant density in CsCl of WCCV2 virions is 1.375 g cm⁻³ and analogous to the alphapartitiviruses there is also a single-major CP with a molecular weight of 71–77 kDa and an RdRp of 77–87 kDa. Betapartitiviruses have the largest proteins among all genera in this family.

Table 3 Overview of the plant infecting partitivirids which are not yet listed by the ICTV. Sizes of available genome segments, proteins and accession numbers are given

<i>putative Genus/Virus name</i>	<i>Size of dsRNA segments [b] and protein [aa (kDa)]</i>	<i>GenBank accession</i>
<i>Alphapartitivirus</i>		
Alopecurus myosuroides partitivirus 1 (Black grass cryptic virus 1)	1: RdRp 1984; 616 (72) 2: CP 1835; 489 (54)	HG005156.1 LN713935.1
Alopecurus myosuroides partitivirus 2 (Black grass cryptic virus 2)	1: RdRp 2037; 515 (60) 2: CP 1956; 520 (54)	LN713936.1 LN713937.1
Medicago sativa alphapartitivirus 1	1: RdRp 1922; 586 (69) 2: CP 1707; 491 (54)	MF443256.1 MF443257.1
Medicago sativa alphapartitivirus 2	1: RdRp 1939; 586 (68) 2: CP 1764; 491 (54)	MK292288.1 MK292289.1
Pyrus pyrifolia cryptic virus 2	1: RdRp 1945; 586 (68) 2: CP 1788; 491 (54)	LC221826.1 LC221827.1
Raphanus sativus cryptic virus 4	1: RdRp 1976; 616 (72) 2: CP 1751; 490 (55)	MF686921.1 MF686922.1
Spinach cryptic virus 1 isl. SRR1766311	1: RdRp 1966; 616 (73) 2: CP 1762; 488 (54)	KX784754.1 KX784755.1
Spinach cryptic virus 1 isl. SRR1766329	1: RdRp 1961; 616 (73) 2: CP 1753; 488 (54)	KX784756.1 KX784757.1
<i>Deltapartitivirus</i>		
Carnation cryptic virus 3	1: RdRp 1573; 475 (54) 2: CP 1562; 411 (47)	KX826917.1 KX826918.1
Citrullus lanatus cryptic virus	1: RdRp 1603; 477 (54) 2: CP 1466; 407 (46)	KY081285.1 KY081284.1
Cucumis melo cryptic virus	1: RdRp 1502; 477 (55) 2: CP 1715; 480 (55)	MH479772 MH479773
Medicago sativa deltapartitivirus 1	1: RdRp 1584; 477 (54) 2: CP 1353; 394 (46)	MF443258.1 MF443259.1
Pittosporum cryptic virus 1	1: RdRp 1967; 482 (55) 2: CP 1525; 405 (47)	LN680393.2 LN680394.2
Pyrus pyrifolia cryptic virus 1	1: RdRp 2: CP 1523; 419 (48) 3: CP 1481; 411 (48)	no entry in GenBank LC221824.1 LC221825.1
Tea-oil camellia deltapartitivirus 1	1: RdRp 1712; 477 (56) 2: CP 1504; 343 (39) 3: CP 1353; 344 (38)	MH814756.1 MH814757.1 MH814758.1
<i>Genus unclassified</i>		
Lolium rigidum partitivirus 1	1: RdRp 2: CP 1939; 516 (57)	no entry in GenBank HG005148.1
Lolium rigidum partitivirus 2	1: RdRp 2: CP 1841; 520 (58)	no entry in GenBank HG005149.1
Lolium rigidum partitivirus 3	1: RdRp 2: CP 1236; 350 (39)	no entry in GenBank HG005150.1
Raphanus sativus cryptic virus 3	1: RdRp 1609; 481 (55) 2: CP 1581; 374 (43)	NC_011705.1 NC_011706.1

isl.: isolate; VP: viral protein.

Cryspovirus

Cryptosporidium parvum virus 1 appears as isometric and non-enveloped virions of about 31 nm in diameter. Negative-contrast electron microscopy revealed single-layered particles, which are not penetrated by stain showing short protrusions on their surfaces. The buoyant density in CsCl is 1.39–1.44 g cm⁻³ and there is a single-major CP with a molecular weight of 37 kDa and an RdRp of 62 kDa.

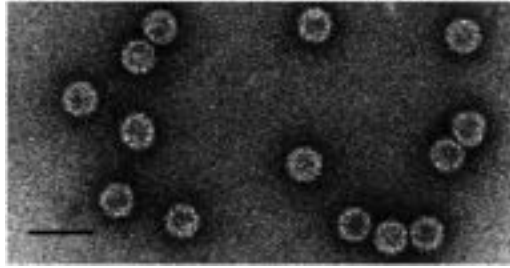


Fig. 1 *Alphapartitivirus*. Negative-contrast electron micrograph of particles of an isolate of *White clover cryptic virus 1*, the type species of the genus *Alphapartitivirus*. The bar represents 50 nm. Reproduced from Ghabrial, S.A., Bozarth, R.F., Buck, K.W., Martelli, G.P., Milne, R.G., 2000. *Partitiviridae*. In *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses*, New York: Academic Press, pp. 503–513.

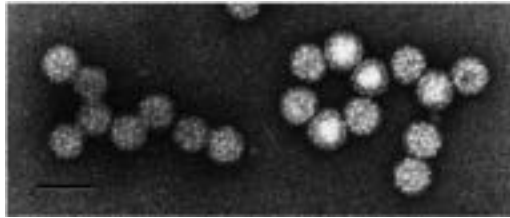


Fig. 2 *Betapartitivirus*. Negative-contrast electron micrograph of particles of an isolate of *White clover cryptic virus 2*. The bar represents 50 nm. Reproduced from Ghabrial, S.A., Bozarth, R.F., Buck, K.W., Martelli, G.P., Milne, R.G., 2000. *Partitiviridae*. In *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses*, New York: Academic Press, pp. 503–513.

Deltapartitivirus

Deltapartitivirus particles of PCV1 and PCV2 are approximately 30 nm in diameter. Like the betapartitiviruses, these particles are not penetrated by stain in negative-contrast electron micrographs and the buoyant density in CsCl of BCV2 was defined as 1.36 g cm^{-3} . The single major CP ranges in molecular weight from 38 to 49 kDa and the RdRp is approximately 54 kDa. A putative member of this genus, *Cucumis melo cryptic virus*, which was recently described, shows a longer dsRNA2 encoding a CP of 55 kDa. Beet cryptic virus 2 and *Fragaria chiloensis cryptic virus* (isl. NCGR CFRA 9089), *Raphanus sativus cryptic virus 2* and rose cryptic virus 1 (isl. ShB-1) contain a third genome segment, the function of which has not yet been conclusively clarified.

Genome Organization and Replication Strategy

Genome Organization

Members of the family *Partitiviridae* have at least two monocistronic genome segments, in few cases a third segment could be detected (see above). There are several theories about the function of this additional genome segment. Tzanetakis et al suggested in 2008 that due to the similarities of segments 2 and 3 of *Fragaria chiloensis cryptic virus*, Rose cryptic virus 1 and *Raphanus sativus cryptic virus 2* to each other, it could possibly be another CP, which was created by a duplication event that took place before species segregation. Thus, the CPs might have retained their function. Another hypothesis of Chen in 2006 was that they could be satellite elements. For the viruses BCV2 and rose cryptic virus 2 it was possible to detect strains in plants which do not contain the third segment. In a review by Nibert et al in 2014 it was suspected that the strains with the additional CP-like segment were present in a mixed infection and the associated RdRp could not yet be detected or was lost. If the RdRp had been lost, the other RdRp would have to be able to replicate all dsRNA segments.

The sizes of the two main genome segments of partitivirids are different between the genera and specific for the respective one. Table 4 gives an overview of the ranges of genome and protein sizes, respectively, from segments 1 and 2 (taken from Table 2 in Section “Current Taxonomy of the *Partitiviridae*”).

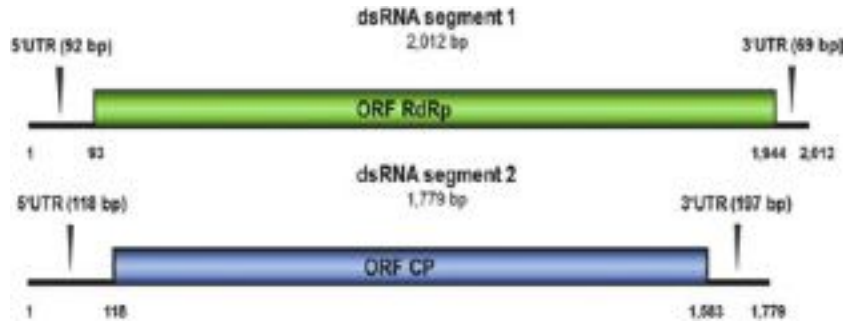
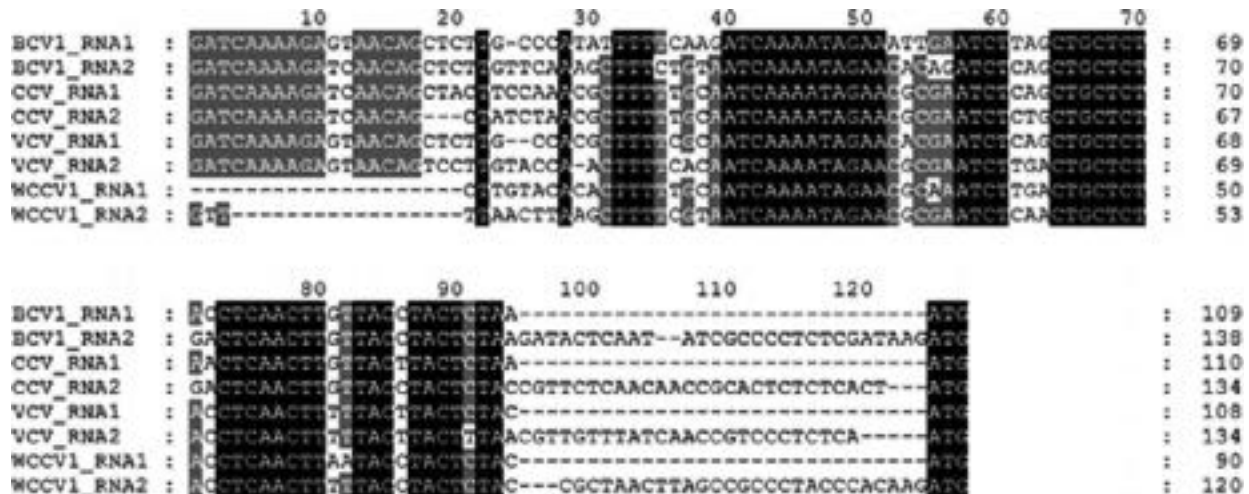
A schematic representation of the alphapartitviral genome of *Vicia cryptic virus* is shown in Fig. 3.

5' and 3'UTRs

The 5'UTRs of partitivirids show some genus-specific conserved sites, which is shown exemplarily in Fig. 4 for the alphapartitiviruses. The 5'UTRs of both dsRNA segments of the assigned species BCV1, CCV, VCV, and WCCV1 were aligned and 100% conserved regions are shown black-shaded. Gray-shaded sequences are not conserved in all viruses. In the case of WCCV1 the

Table 4 Overview of the ranges of genome and protein sizes from segments 1 and 2. (taken from Table 2 in Section “Current Taxonomy of the Partitiviridae”)

Genus	dsRNA1 [bp]	dsRNA2 [bp]	RdRp [aa]	CP [aa]
<i>Alphapartitivirus</i>	1866–2013	1708–1837	573–621	487–505
<i>Betapartitivirus</i>	2390–2444	2266–2354	723–746	672–674
<i>Cryspovirus</i>	1786	1374	524	319
<i>Deltapartitivirus</i>	1563–1749	1479–1521	477–479	346–415

**Fig. 3** Schematic genome organization of the alphapartitivirus *Vicia cryptic virus* (NC_007241.1, NC_007242.1). The RNA is represented by a black bar on which the open reading frames are illustrated as green (RdRp) and blue (CP) boxes. The numbers outline the first nucleotide position of the respective feature.**Fig. 4** Alignment of alphapartitviral 5'UTRs of both dsRNA segments. Areas with 100% identity are black-shaded and gray-shaded sequences are not conserved in all viruses. *BCV1*: beet cryptic virus 1; *CCV*: carrot cryptic virus; *VCV*: *Vicia cryptic virus*.

5'ends differ from the two dsRNA segments, which could possibly be due to the fact that not all nucleotides could be determined correctly. It is noteworthy, that all 5'UTRs of the second genome segments are longer.

Similar findings were described in the 5'UTRs of betapartitiviruses, which almost all begin with the pentanucleotide AGAUIU and show a highly conserved region at positions 39–56 (Fig. 5).

Furthermore, two conserved stem-loop structures are present, potentially being involved in dsRNA replication and/or virion assembly (Fig. 6).

Deltapartitiviruses seem to have three different types of extreme 5'ends (assuming all ends have been correctly determined). For example, nearly all tripartite species despite of BCV2 (FCCV, RSCV2, and RCV-1) begin with the septanucleotide GAUAAUG (Fig. 7, type 1). There is another variation in which the genome segments of most bipartite species (PCV 1 & 2, BCV2, and RNA2 of PeCV) begin with AGAAUU (type 2). Only FCV and RNA1 of persimmon cryptic virus show the nucleotides GGA(A/U)U at the 5'end (type 3). The 5'UTR of *Cryptosporidium parvum* virus 1 does not fit to one of these types and both segments start with GGAAA(A/G).

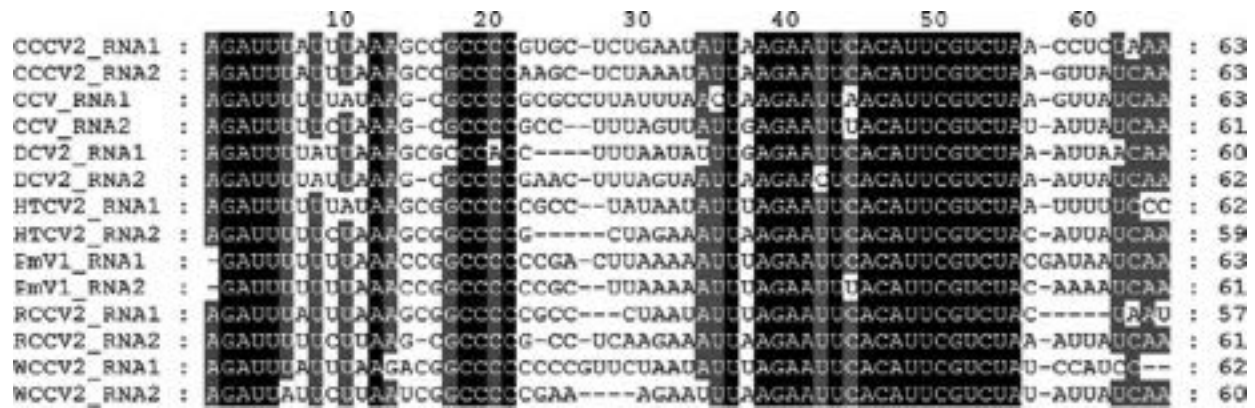


Fig. 5 Alignment of betapartitviral 5'UTRs of both dsRNA segments. Areas with 100% identity are black-shaded and gray-shaded sequences are not conserved in all viruses. *CCCCV2*: crimson clover cryptic virus 2; *CCV*: Cannabis cryptic virus; *DCV2*: dill cryptic virus 2; *HTCV2*: hop trefoil cryptic virus 2; *PmV1*: Primula malacoides virus 1; *RCCV2*: red clover cryptic virus 2; *WCCV2*: White clover cryptic virus 2.

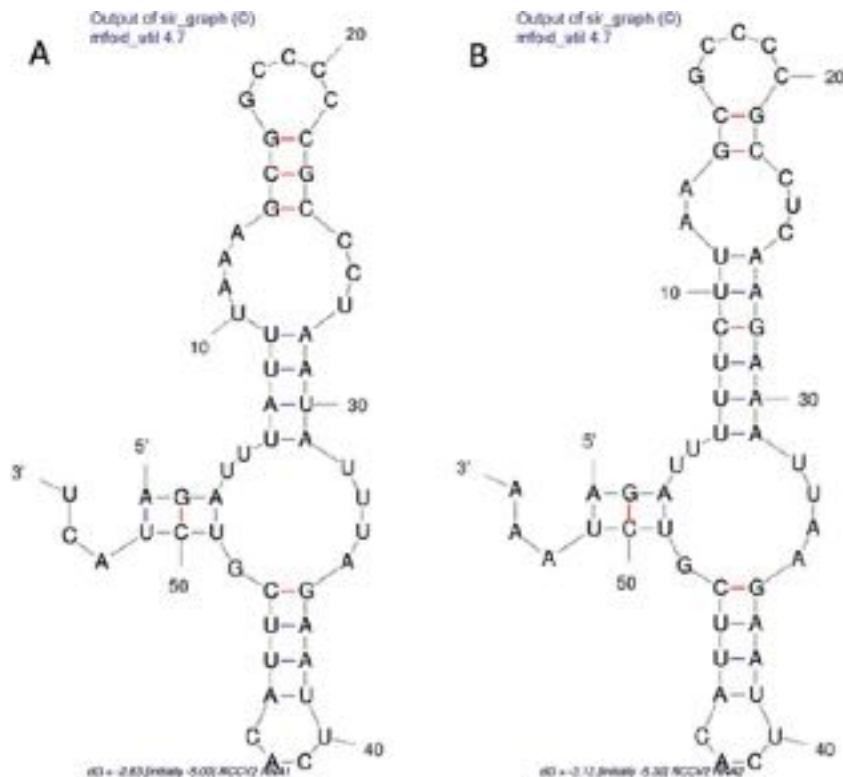


Fig. 6 RNA foldings of partial 5'UTRs (nucleotides 1–54) of RCCV2 dsRNA1 (A) and dsRNA2 (B).

Within the 3'UTRs, there are less conserved sequences than it was shown for the 5'UTRs before. The 3'UTRs of alpha- and betapartitiviruses show poly(A) stretches, which can be interrupted by other nucleotides, deltapartitiviruses and *Cryptosporidium parvum* virus 1 lack this feature (Fig. 8). In general, these A-rich areas seem to be longer in RNA2 than in RNA1. From these findings the question arises, if this could be an imitation of common poly(A) tails. Additionally, the adenosine content of the last 3' located 50 nucleotides is significantly higher in alpha- and betapartitivirids than in the other genera. Betapartitiviruses show another relatively conserved feature - they end with three to four Cs at the extreme 3' end (Fig. 8).

RNA1& 2

In general, RNA1 is larger than RNA2 and encodes an RdRp. The size differs between the genera, so the RdRps of alphapartitiviruses are about 72 kDa in size, of betapartitiviruses about 87 kDa, of CSpV1 62 kDa and of deltapartitiviruses about 54 kDa. There are several

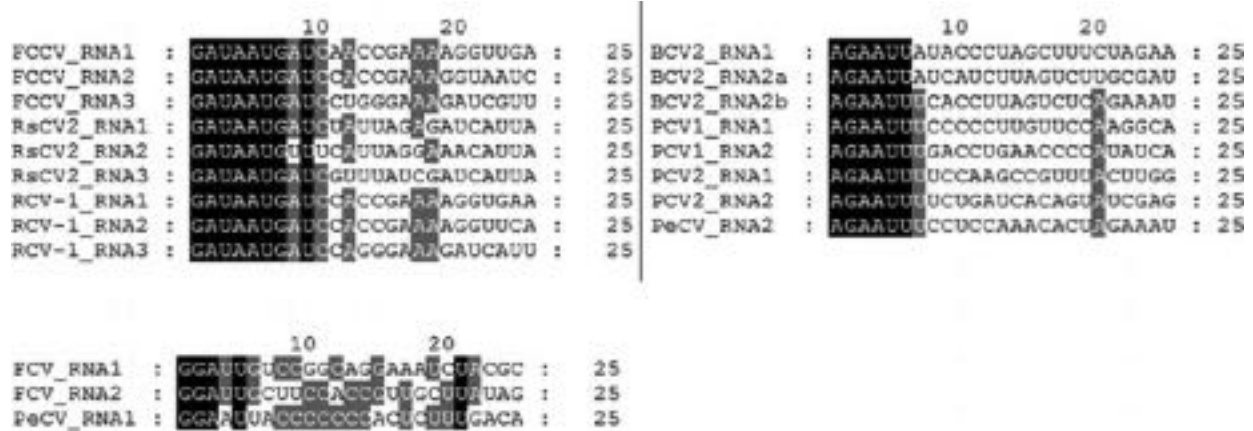


Fig. 7 Comparison of three different types of 5' ends of deltapartitiviruses (not aligned). Areas with 100% identity are black-shaded. Gray-shaded sequences are not conserved among all viruses. *BCV2*: beet cryptic virus; *FCV*: fig cryptic virus; *FCCV*: *Fragaria chiloensis* cryptic virus; *PCV1/2*: pepper cryptic virus 1/2; *PeCV*: persimmon cryptic virus; *RCV1*: rose cryptic virus 1; *RscV2*: *Raphanus sativus* cryptic virus 2.



Fig. 8 3' poly(A) stretches (underlined) and terminal located Cs (edged) of both segments of one alpha- and betapartitivirus. *VCV*: *Vicia* cryptic virus, α ; *WCCV2*: *White clover* cryptic virus 2, β .

known motifs that occur in RdRps of partitivirids. Motifs I and II are not exactly localizable, but motifs III to VIII can be determined with slight variations and examples are given for WCCV1 (α), WCCV2 (β), CSpV1 (*Cryspovirus*) and PCV1 (δ). Motif III can be detected as K-X-R-3X-(A,G,P), motif IV as D-W-2X-(F, Y)-D, motif V as G-(I, V)-X-S-G, motif VI as G-D-D, motif VII as I-(G, S, V)-Y, and motif VIII as R.

The smaller RNA2 encodes a CP with sizes of about 54 kDa for alphapartitiviruses, 75 kDa for betapartitiviruses and 37 kDa for CSpV1. In case of the deltapartitiviruses there seem to be two types of CP which can be distinguishable by size, one is about 49 kDa and the second one about 38 kDa. Compared to the RdRps, there are not so many clearly conserved sequence stretches to be found in the CPs. Some motifs are described, the first of which is called "SQLY" and varies a lot between different species and genera. A second motif "PGPL3XF" can be found in alphapartitiviruses and in a deviant form in CSpV1.

Replication Strategy

In the plant cell, dsRNA usually triggers RNAi as plant defense mechanism leading to post-transcriptional gene silencing (PTGS), which is responsible for its degradation. To circumvent this defense strategy, viruses have evolved silencing suppressors (VSRs). No VSR could be detected for partitivirids yet, so they are believed to perform their transcription/replication within the particle. Every virus particle contains one or two RdRp molecules anchored to the CP as well as one or two dsRNA segments either coding for the CP or the RdRp. The RdRp acts in a semi-conservative manner, producing plus-strands from the minus-strand in this encyclopedia. In a 3D reconstruction of the capsid of the PsV-S, possible pores could be discovered that are probably associated with the transport of the plus-strand RNA from the particle into the cytoplasm during transcription. The newly synthesized plus-strands remains in the particle while the parental plus-strand is released into the cytoplasm. There, translation of RdRps and CPs by cellular ribosomes takes place and packaging of new virions is performed. Therefore, RdRps and RNAs associate and interact with the CPs whereupon the particle is fully assembled. Protein-protein interaction studies of alpha- and betapartitiviruses revealed varying results. For example, the expected interaction between RdRp and CP could only be demonstrated for alphapartitiviruses and not for betapartitiviruses. Furthermore, in the case of alphapartitiviruses, the CP dimers seem to be located in the nuclear membrane, those of betapartitiviruses in inclusions within the cytoplasm of epidermis cells. These findings indicate different replication strategies and still have to be elucidated.

Transmission

Partitivirids are not able to be transmitted mechanically, by grafting or natural vectors. Within the plant or protozoon, no cell-to-cell movement is observed except during cell division or gamete fusion in case of CSpV1. From one plant to another partitivirids can be transmitted by seeds, ovules and pollen in high rates. It was found in crosses of BCV infected and non-infected sugar beet plants

(either female or pollen) that a female non-infected plant crossed with pollen from an infected plant led to an infection rate in progeny plants of 43% and vice versa of 82%. When both plants were infected, the infection rate in progeny plants is 100%. Partitivirids can not be eliminated by meristem culture or thermotherapy and stay life-long in their hosts. In experiments on *Dianthus* Szego and colleagues showed that after 16 years of in vitro cultivation carnation cryptic virus 1 (CCV1) was still present in the plants.

Virus-Host-Interaction

An infection with partitiviruses is asymptomatic, which is also referred to as a latent infection. To date, there is hardly any evidence of the effects that partitiviruses have on their host. One study, however, showed that the expression of the gene TrEnodDR1 from *Trifolium repens* in *Lotus japonicus* resulted in reduced growth of the transformants and a smaller number of root nodules formed by infection with *Mesorhizobium loti* due to an increase of endogenous abscisic acid. Within this gene, sequences corresponding to the CP gene of WCCV1 were detected, leading to the conclusion that an infection of *T. repens* with WCCV1 could be mutualistic. Another study screened perennial ryegrass concerning salt toxicity and one result was a clone containing a partial sequence of a RdRp from the putative deltapartitivirus RsCV2. These findings may suggest evidence for mutualistic relationships between plants and partitivirids. This effect does not seem to be limited to plants only, because there are also studies in the case of CsPV1 whose results indicate that strain-dependent increased virus concentration in its protozoan host is associated with increased *C. parvum* replication in cell cultures and higher oocyte count in dairy calves.

Another impact on the host plant is the horizontal gene transfer. It has been shown that CP-like sequences are present in *Arabidopsis* and *Nicotiana* species. However, most inserts are pseudogenes with internal stop codons. It has also been reported that in infected plants no inserts could be detected in the genome.

Further Reading

- Blawid, R., Stephan, D., Maiss, E., 2008. *Alphacryptovirus* and *betacryptovirus*. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Amsterdam: Elsevier, pp. 98–104.
- Boccardo, G., Milne, R.G., Luisoni, E., Lisa, V., Accotto, G.P., 1985. Three seedborne cryptic viruses containing double-stranded RNA isolated from white clover. *Virology* 147 (1), 29–40.
- Chiba, S., Kondo, H., Tani, A., *et al.*, 2011. Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLoS Pathogens* 7 (7), e1002146.
- Ghabrial, S.A., Ochoa, W.F., Baker, T.S., Nibert, M.L., 2008. Partitiviruses. General features. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. Amsterdam: Elsevier, pp. 68–75.
- Khrantsov, N.V., Upton, S.J., 2000. Association of RNA polymerase complexes of the parasitic protozoan *Cryptosporidium parvum* with virus-like particles. *Heterogeneous system*. *Journal of Virology* 74 (13), 5788–5795.
- Lesker, T., Maiss, E., 2013. In planta protein interactions of three alphacryptoviruses and three betacryptoviruses from white clover, red clover and dill by bimolecular fluorescence complementation analysis. *Viruses* 5 (10), 2512–2530.
- Lesker, T., Rabenstein, F., Maiss, E., 2013. Molecular characterization of five betacryptoviruses infecting four clover species and dill. *Archives of Virology* 158 (9), 1943–1952.
- Nakatsukasa-Akune, M., Yamashita, K., Shimoda, Y., *et al.*, 2005. Suppression of root nodule formation by artificial expression of the TrEnodDR1 (coat protein of White clover cryptic virus 1) gene in *Lotus japonicus*. *Molecular Plant-Microbe Interactions (MPMI)* 18 (10), 1069–1080.
- Nibert, M.L., Ghabrial, S.A., Maiss, E., *et al.*, 2014. Taxonomic reorganization of family *Partitiviridae* and other recent progress in partitivirus research. *Virus Research* 188, 128–141.
- Nibert, M.L., Woods, K.M., Upton, S.J., Ghabrial, S.A., 2009. *Cryspovirus*: A new genus of protozoan viruses in the family *Partitiviridae*. *Archives of Virology* 154 (12), 1959–1965.
- Ochoa, W.F., Havens, W.M., Sinkovits, R.S., *et al.*, 2008. Partitivirus structure reveals a 120-subunit, helix-rich capsid with distinctive surface arches formed by quasisymmetric coat-protein dimers. *Structure* 16 (5), 776–786.
- Roossinck, M.J., 2010. Lifestyle of plant viruses. *Philosophical Transactions of The Royal Society B Biological Sciences* 365 (1548), 1899–1905.
- Sabanadzovic, S., Valverde, R.A., Brown, J.K., Martin, R.R., Tzanetakis, I.E., 2009. Southern tomato virus. The link between the families *Totiviridae* and *Partitiviridae*. *Virus Research* 140 (1–2), 130–137.
- Tzanetakis, I.E., Price, R., Martin, R.R., 2008. Nucleotide sequence of the tripartite *Fragaria chiloensis* cryptic virus and presence of the virus in the Americas. *Virus Genes* 36 (1), 267–272.
- Vainio, E.J., Chiba, S., Ghabrial, S.A., *et al.*, 2018. ICTV virus taxonomy profile: *Partitiviridae*. *Journal of General Virology* 99 (1), 17–18.

Relevant Websites

- <https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=11012>
Complete genomes: Partitiviridae.
NCBI.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsrna-viruses/w/partitiviridae
Partitiviridae.
dsRNA Viruses.
ICTV.
- https://viralzone.expasy.org/168?outline=all_by_species
Partitiviridae ~ ViralZone page.

Quadriviruses (*Quadriviridae*)

Hideki Kondo, Okayama University, Kurashiki, Japan

José R Castón, National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Nobuhiro Suzuki, Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Quadrupartite virus Virus with the four genomic segments separately packaged into four different particles.

Quadrupartite genome The essential genome is separated into four genomic segments.

$T = 1$ icosahedral symmetry Symmetry of capsid composed of 60 subunits, which forms 12 pentameric capsomeres.

Introduction

Screens involving filamentous fungi, particularly phytopathogenic fungi, have detected an enormous number of novel mycoviruses (or fungal viruses) that have unique biological and molecular features, and that enhanced our knowledge of virus diversity and evolution. Many of them have double-stranded RNA (dsRNA) genomes and isometric particles. DsRNA mycoviruses are currently classified into six families, *Chrysoviridae* (includes both traditional chrysovirus with four segments as well as proposed others with three to seven segments), *Megabirnaviridae* (two segments), *Partitiviridae* (mostly two segments), *Quadriviridae* (four segments), *Reoviridae* (11 or 12 segments), and *Totiviridae* (nonsegmented), and one genus *Botybirnavirus* (two segments, family unassigned). Some dsRNA mycoviruses have also been proposed as the members of tentative families or genera, such as *Alternaviridae* (four segments), *Fusagraviridae* (nonsegmented), *Megatotiviridae* (nonsegmented) and *Phlegivirus* (nonsegmented), the virion morphology for which is largely unknown, except for alternaviruses. The *Quadriviridae* is a recently established virus family, and its members are non-enveloped spherical viruses with quadrupartite dsRNA genomes. The genus *Quadrivirus* within the family currently contains a single species *Rosellinia necatrix quadrivirus* 1. *Rosellinia necatrix quadrivirus* 1 (RnQV1) was discovered during a screen of over 1000 Japanese isolates of the white root rot fungus, *Rosellinia necatrix* (class Sordariomycetes), which is an ascomycete able to infect over 400 plants. The name quadrivirus originates from the 'quadrupartite' dsRNA genome. At present, no species demarcation criteria exist for the genus *Quadrivirus* since only one species has been assigned to this genus. The virion structure of quadriviruses has unique features as compared to those of other known dsRNA viruses. Quadriviruses appear to be phylogenetically closely related to totiviruses (members of the genus *Totivirus* with an unsegmented dsRNA genomes). The goals of this article are to give an overview of the features of the well-studied quadriviruses and compare them with those of tentative members of the family *Quadriviridae* and other related dsRNA viruses.

Virion Structure and Composition

The virion of RnQV1 is non-enveloped, isometric, about ~45 nm in diameter (Fig. 1) and likely separately packages each of the four dsRNA genome segments, which are 3.5–5.0 kbp in length. The buoyant densities of quadrivirus virion are not known. The size of quadriviruses are similar to that of chryso-, partiti-, totiviruses (25–43 nm) and megabirnaviruses (~52 nm), which also separately encapsidate their segments, but are much smaller than that of mycoreoviruses (~80 nm), which package all of their dsRNA segments into a single multi-layered virion. Most of spherical dsRNA mycoviruses have a single-shelled $T = 1$ capsid formed by 60 asymmetric homodimers of two capsid protein (a 120-subunit $T = 1$ capsid). An exception is the chrysovirus particle, which is composed of 60 copies of a single capsid protein with a duplicated helix-rich domain (a genuine $T = 1$ capsid). The RnQV1 capsids are comprised of 60 copies of heterodimers of two structural proteins, P2 (1356 or 1357 residues) and P4 (1061 or 1059 residues), which are encoded by dsRNA2 and dsRNA4, respectively, but P2 383-residue C-terminal region is absent in the mature viral particle (Fig. 2). The P2-P4 heterodimers are organized in a quaternary structure similar to that of reovirus, chrysovirus and totivirus. The two structural proteins of RnQV1 isolate W1075 are cleaved into several polypeptides without altering capsid structural integrity, likely during virion purification and/or in the infected fungal host, whereas in isolate W1118 P2 and P4 remain nearly intact (Fig. 3(a)). Similar capsid assembly involving two structural proteins may be suggested for botybirnaviruses, *Botrytis porri* RNA virus 1 and *Alternaria alternata* botybirnavirus 1 and for a betachrysovirus, *Magnaporthe oryzae* chrysovirus, which possibly belongs to the newly proposed genus *Betachrysovirus* (currently designated Cluster II) of the family *Chrysoviridae* (see Fig. 5 for their phylogenetic relationships). Posttranslational cleavage/degradation similar to RnQV1-W1075 virions was also observed for the two putative structural proteins of these dsRNA mycoviruses, even though their structural characteristics are still unclear.

RnQV1 P2 and P4 lack sequence similarity, but they have a similar α -helical domain that is the structural signature shared with most dsRNA viruses. In addition to organizing the viral genome and RNA-dependent RNA-polymerase (RdRp) molecules, P2 and

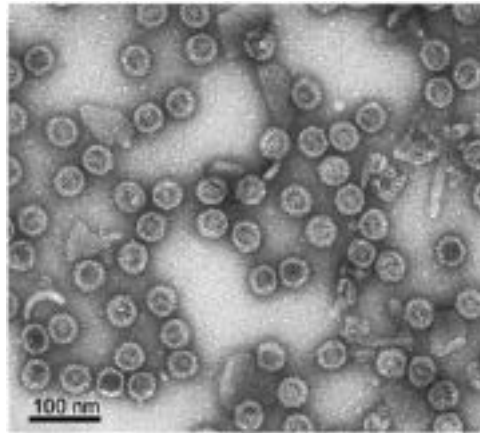


Fig. 1 Quadrivirus particle structure. Electron micrograph of negatively stained particles of RnQV1-W1118 (scale bar, 100 nm).

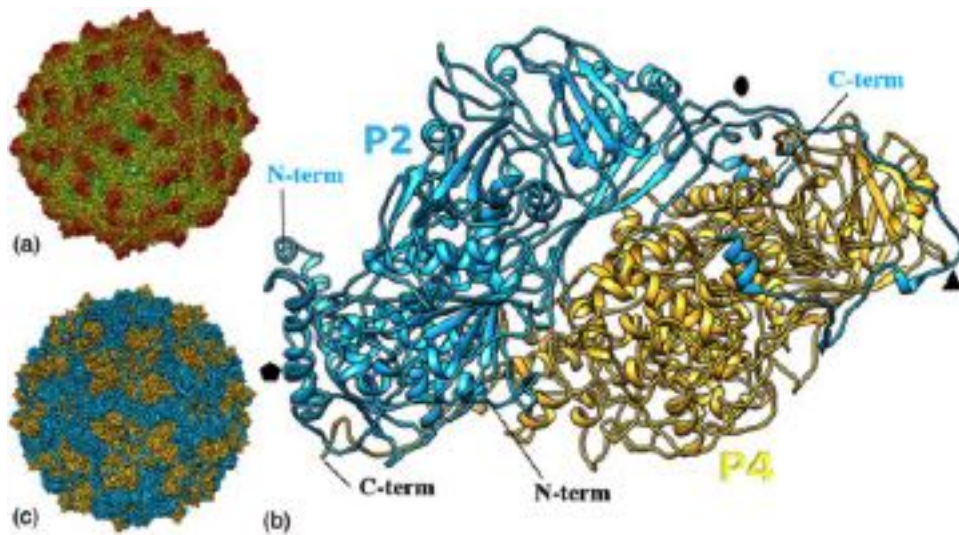


Fig. 2 Three-dimensional cryo-EM reconstruction of RnQV1 virions at 3.7 Å resolution. (a) Radially color-coded surface-shaded virion $T = 1$ capsid viewed along an icosahedral two-fold axis showing 12 outwardly protruding decamers (orange). Bar = 100 Å. (b) Ribbon diagrams of P2 (blue) and P4 (yellow) proteins (top view). N and C termini are indicated. Symbols indicate icosahedral symmetry axes. (c) Surface-shaded virion $T = 1$ capsid viewed along an icosahedral two-fold axis highlighting the A (blue) and B (yellow) structural subunits. P2 C-terminal regions are clearly seen as hooks embracing P4 outer surface. Reproduced from Fig. 1(f) of Mata, C.P., Luque, D., Gómez-Blanco, J., *et al.*, 2017. Acquisition of functions on the outer capsid surface during evolution of double-stranded RNA fungal viruses. *PLoS Pathogens* 13, e1006755.

P4 may have acquired new functions by inserting complex domains in preferential insertion sites at the capsid outer surface, likely related to enzyme activity. The P2 insertion has a fold similar to that of gelsolin and profilin, two actin-binding proteins with a function in cytoskeleton metabolism, whereas the P4 insertion suggests protease activity involved in cleavage of the P2 C-terminal region.

Except for reoviruses, whose 10–12 dsRNA genome segments densely packaged within a single particle (~ 40 bp/100 nm³ that corresponds to a spacing between dsRNA strands of 25–30 Å), most of the known multi-segmented dsRNA mycoviruses have a multi-particulate nature, in which dsRNA segments are encapsidated separately and they might have a lower density at the central region of the capsid, with a single loosely-packed dsRNA molecule (~ 20 bp/100 nm³, with an interstrand spacing of ~ 40 –45 Å in the case of chrysovirus and partitivirus). A similar value for packed dsRNA, 25 bp/100 nm³, if there is one dsRNA molecule in the capsid, was proposed for quadriviruses. Such low density at the central region of the capsid may increase dsRNA mobility, which might be necessary for maximum RdRp activity with the RdRp complex. Interestingly, the amount of dsRNA1 and dsRNA4 in the quadrivirus virions were consistently less than those of dsRNA2 and dsRNA3 (Fig. 3(b)), and a similar proportion of the virion dsRNA profile was observed in a chrysovirus, *Helminthosporium victoriae* virus 145S (HvV145S, a proposed genus *Alphachrysovirus*, formerly the genus *Chrysovirus*). Thus, it is also tempting to speculate that two heterogeneous dsRNA sets, dsRNA1 + dsRNA4 (8.6 kbp total) and dsRNA2 + dsRNA3 (8.5 kbp total), where each set always shows a similar accumulation ratio, are separately packed in the two different capsids. If this is the case, an alternative density for the quadrivirus capsid with two

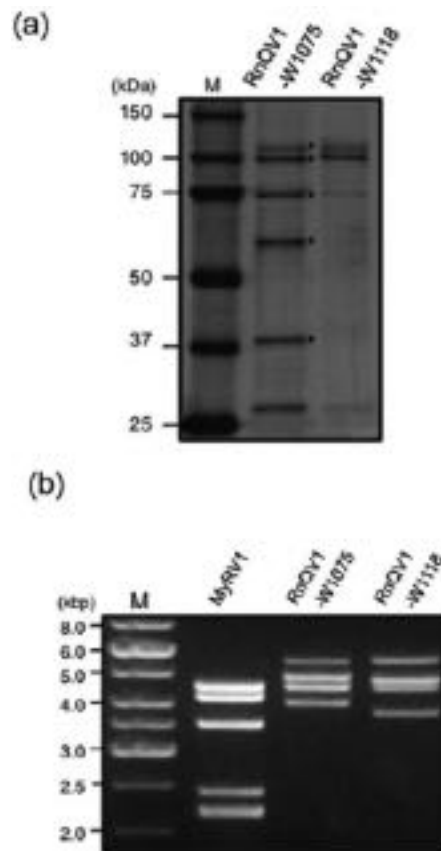


Fig. 3 Quadrivirus dsRNA and protein components. (a) SDS-PAGE pattern of structural proteins of RnQV1-W1118 and -W1075. (b) Agarose gel electrophoresis of RnQV1-W1118 and -W1075 genomic dsRNA segments. The genomic dsRNA segments of a mycoreovirus (MyRV1 S1–S6, and others are not shown) and a DNA size marker (M) were used as size standards. Reproduced from Fig. 1(b) and (c) of Lin, Y.H., Hisano, S., Yaegashi, H., Kanematsu, S., Suzuki, N., 2013. A second quadrivirus strain from the phytopathogenic filamentous fungus *Rosellinia necatrix*. *Archives of Virology* 158, 1093–1098, with permission from Springer.

heterologous dsRNA molecules could conceivably be calculated as 50 bp/100 nm³. See the other article of this section “Structure of Double-Stranded RNA Mycoviruses” by Castón *et al.* for further discussion of this possibility.

Genome Organization and Expression

Quadriviruses have a multi-segmented genome consisting of four monocistronic linear dsRNA segments. The complete genome sequence of the RnQV1-W1075 isolate is a total of 17,078 bp. Each dsRNA segment ranges from 3.7 to 4.9 kbp (GC contents are 51.9%–54.3%) and has a single large open reading frame (ORF) covering 86%–97% of its segment size (Fig. 4). The largest segment, dsRNA1 (4942 bp), codes for a protein of unknown function (P1, 1602 amino acid residues, 178.2 kDa). The third segment, dsRNA3 (4099 bp) codes for RdRp (P3, 1310 amino acids, 146.8 kDa). The second and fourth segments, dsRNA2 (4352 bp) and dsRNA4 (3685 bp) code for the two structural proteins, P2 (1356 amino acids, 147.4 kDa) and P4 (1061 amino acids, 113.2 kDa), respectively. The quadripartite nature of quadriviruses resembles that of the traditional chrysoviruses including several members of the proposed genus Alphachrysovirus and alternaviruses members of the proposed family Alternaviridae. However, the total size of quadrivirus genome (3.7–4.9 kbp, 17.1 kbp total) is approximately 1.5–1.7 fold larger than those of traditional chrysoviruses (2.5–3.6 kbp, ~12.6 kbp total) and alternaviruses (1.4–3.6 kbp, 10.4 kbp total).

The 5'- and 3'-terminal nucleotide sequences of the four dsRNA genome segments of RnQV1 are conserved (the positive-sense strand sequence of the RnQV1-W1075 is 5'-C/UACGAU-CAUGAGAAUUAUCG/A-3', Fig. 4). Several conserved sequence stretches are also present at both untranslated regions (UTRs), and the 3'-ends are better conserved than the 5'-end regions. The heterogeneity at the extreme termini, C or U for the 5'-end and G or A for 3'-end, is notable. As is found in typical chrysoviruses and some partitiviruses, “CAA” repeats were also found in the 5'-UTR, or within the adjacent coding region of each RnQV1 segment, except for dsRNA3. The (CAA) repeat sequences are known to serve as enhancer elements present at the 5'-UTRs of tobamoviruses, plant alpha-like viruses. Recently, the 5'-UTRs of two typical chrysoviruses, *Cryphonectria nitschkei* chrysovirus 1 and HvV145S (215 and 293 nt, respectively), were found to have internal ribosomal entry site (IRES) activities, while those of

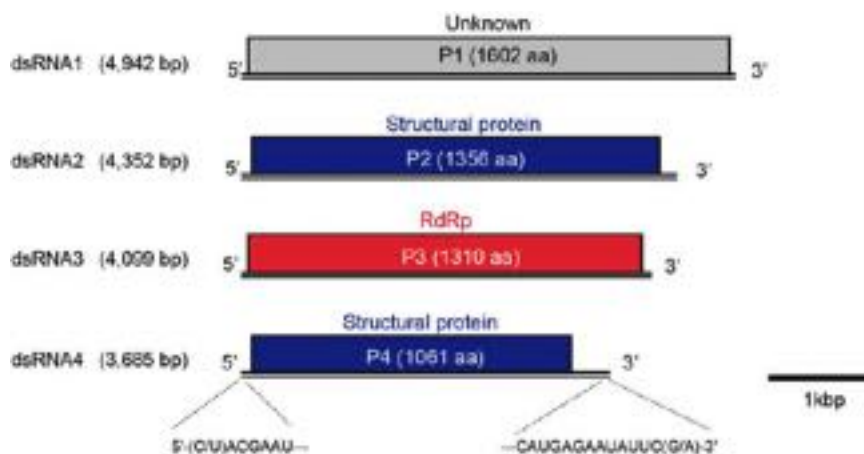


Fig. 4 Genomic organization of RnQV1 isolate W1075. The genome consists of four monocistronic dsRNA segments. dsRNA3 encodes the putative replicase (RdRp), while dsRNA2 and dsRNA4 encode structural proteins. dsRNA1 encodes a putative non-structural protein with unknown function. The sequence stretches conserved among below the dsRNA4. Solid lines and colored open boxes represent genomic dsRNAs and open reading frames on the positive-sense strands, respectively.

quadriviruses are relatively short (23–45 nt) and contain no IRES activity. The translation significance of (CAA)_n in translation may differ between quadriviruses and chrysovirus. Alternaviruses and members of genera *Alphapartitivirus* and *Betapartitivirus* have poly(A) tails or poly(A) tracts either at the 3'-terminus or near the 3'-terminus of the coding-strand, whereas quadriviruses and most other dsRNA viruses have no such poly (A) rich sequences.

Although little is known about the replication of multi-segmented dsRNA mycoviruses, their replication and transcription appear to occur inside particles involving virion-associated RdRps, similar to other well studied dsRNA viruses such as rice dwarf virus (a plant reovirus, genus *Phytoreovirus*, family *Reoviridae*) and yeast L-A virus (*Saccharomyces cerevisiae* virus L-A, genus *Totivirus*, family *Totiviridae*). Therefore, RnQV1 P3 (RdRp) may be involved in the genomic dsRNA replication and transcription, although the presence of RdRp in the quadrivirus virions remains to be confirmed. The putative small pores are located at the five-fold (~11 Å-diameter hole) and three-fold axes (~7 Å-diameter hole) of the RnQV1 capsid. This feature resembles that of other single-shelled *T* = 1 capsids of dsRNA mycoviruses, suggesting that the conformational changes in P2 and/or P4 that face the channel wall might allow the exit of viral transcripts. In addition, the presence of two putative insertion sites in the P2 and P4 proteins, which allocate the virus capsid outer surface, is hypothesized to possess enzyme activities. Interestingly, in the case of the yeast L-A virus capsid, the inserted domain is located at a trench on the outer surface and can be responsible for catalyzing the unique “cap-snatching” reaction, which is known as an alternative method for capping mRNA.

Molecular and Biological Properties

The three RnQV1 isolates were found in *R. necatrix* strains W1075, W1118 and W726 were obtained from the Japanese pear orchards in the Saga, Ibaraki, and Saga Prefectures of Japan. RnQV1-W1075 and W1118 have been completely sequenced, while only the complete sequence of dsRNA3 is available for RnQV1-W726. Nucleotide sequence identities between RnQV1-W1075 and W1118 vary from 67% (dsRNA1 and dsRNA3) to 70% (dsRNA2). Amino acid sequences identities between two virus isolates showed 72% (P1), 81% (P2), 74% (P3) and 82% (P4). The W726 dsRNA3 (RdRp) shows 86% and 66% nucleotide sequence (98% and 75% amino acid sequence) identities to that of W1075 and W1118, respectively. Rabbit polyclonal antibodies against RnQV1-W1075 P2 recognize RnQV1-W1118 P2 by western blot analysis. Since a number of field isolates displayed similar dsRNA profiles to that of quadriviruses, it seems that quadriviruses are widespread in the *R. necatrix* population.

The RnQV1 persistently infects its natural fungal host *R. necatrix* and is transmitted intracellularly during cell division and/or mycelial cell fusion (hyphal anastomosis). Transfection of protoplasts from *R. necatrix* and *Cryphonectria parasitica* (an experimental model fungal host) by purified RnQV1 particles was unsuccessful. Little is known about the virus vertical transmission during sexual or asexual sporulation, or about extracellular transmission in nature. The RnQV1 is stably maintained in the host fungus during successive subcultures for a long period of time. However, viral infection has no discernible phenotypic effects on their host colony morphology, suggesting it is a symptomless infection. The spread of RnQV1 throughout the fungal colonies is relatively slower than other *R. necatrix*-infecting viruses, such as Rosellinia necatrix megabirnavirus 1 and Rosellinia necatrix partitivirus 1, because its dsRNA distribution within a colony is uneven, forming sectors with low levels of virus accumulation. RnQV1 does not block transgene silencing, and thus, it appears to have no RNA silencing suppressor activity. RnQV1 does not upregulate antiviral RNA silencing-related genes, while some other viruses do in *R. necatrix*. Next-generation sequencing analyses suggested that sense-strand-specific viral small RNA peaks were found at the 3' termini of RnQV1 segments. Nevertheless, the

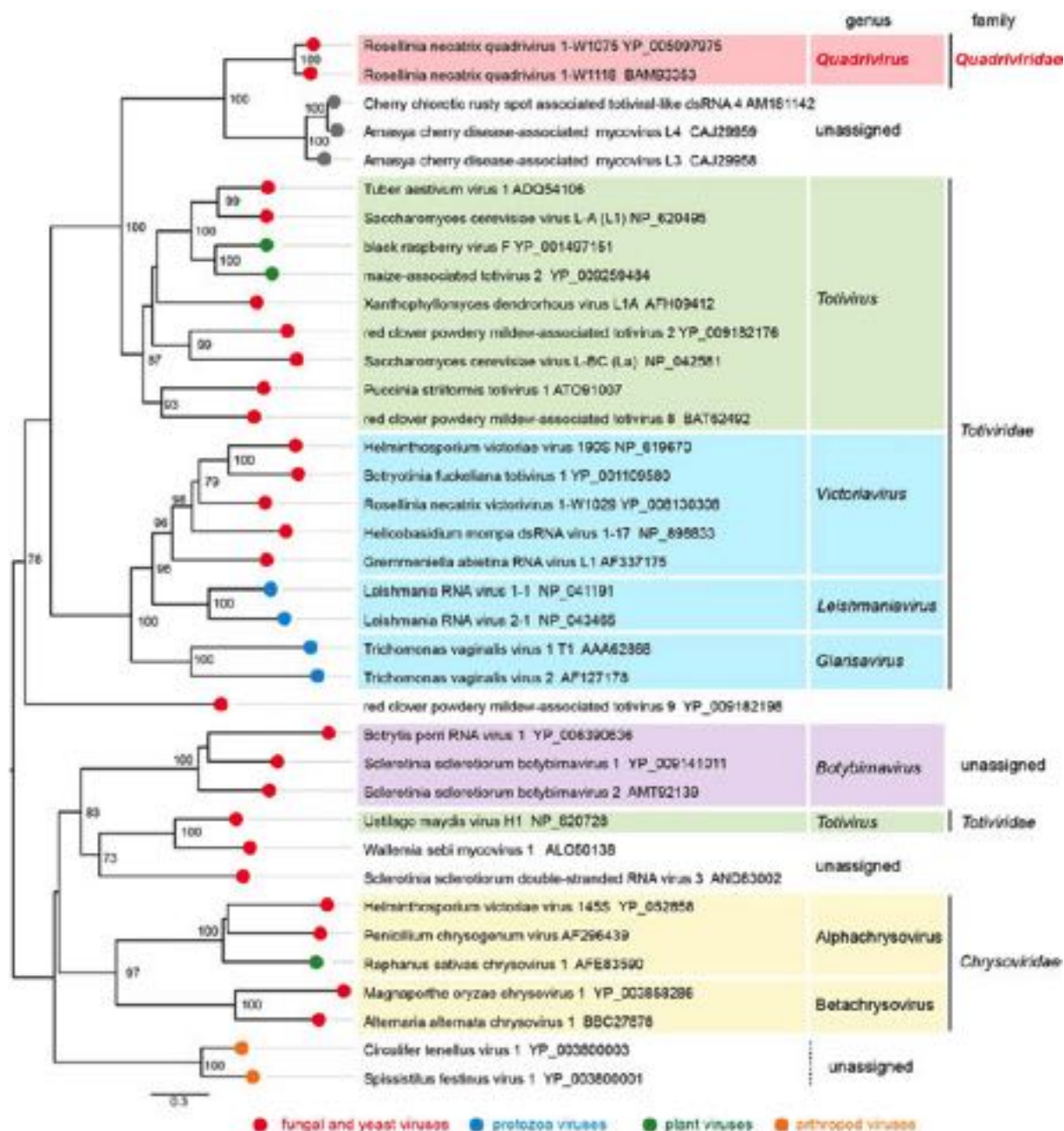


Fig. 5 Phylogenetic relationships of RnQV1 and other related dsRNA viruses. A neighbor-joining tree was constructed from the amino acid sequences alignment based on the RdRp proteins of quadriviruses and selected dsRNA viruses including totiviruses, chrysovirus and botybinaviruses. Numbers at the nodes represent bootstrap percentages (1000 replicates). The scale bar represents the amino acid distances.

abundance of RnQV1 small RNAs was significantly lower than those of other studied *R. necatrix* dsRNA viruses, suggesting that RnQV1 is sequestered from the RNA silencing machinery, or that it has unknown counter-defense mechanism(s).

Phylogenetic and Evolutionary Relationships of RnQV1 to Other dsRNA Mycoviruses

Virus-like dsRNA elements, namely Amasya cherry disease (ACD)-associated dsRNAs with sequence similarity to RnQV1 dsRNA segments together with two novel mycoviruses (chrysovirus and partitivirus), were identified from diseased cherry trees. RnQV1 P3 shows moderate levels of amino acid sequence identities to polypeptides encoded by ACD large (L) dsRNA3 (35%) and L dsRNA4 (36%), which putatively encode RdRps that are related to each other, and are clustered together with quadriviruses in the

phylogenetic tree (Fig. 5). RnQV1 P1 shows weak amino acid sequence identities to polypeptides encoded by ACD L dsRNA1 (19%) and L dsRNA2 (27%). Although no ACD-associated dsRNA segments encoding structural proteins similar to RnQV1 P2 (dsRNA2) and P4 (dsRNA4) have been detected, it might be that these quadrivirus-like dsRNAs are derived from two related mycoviruses. Similarly, quadrivirus-like dsRNA elements, together with two other mycoviruses, were also isolated from the cherry leaves that had Cherry chlorotic rusty spot disease (CCRS). The etiology of ACD and CCRS occurring in Turkey and Italy, respectively, is still unknown. However, a Spanish fungal disease called Cherry leaf scorch (CLS), that has very similar symptoms to CCRS and mycovirus-like dsRNA-elements, is suggested to be caused by *Apiognomonia erythrostoma* (class Sordariomycetes). Thus, ACD- and CCRS-associated dsRNAs are likely of a fungal virus origin.

RnQV1 P3 shows 22%–33% amino acid sequence identities to RdRps of other dsRNA mycoviruses, including the members of the families *Totiviridae* (unsegmented dsRNA genome) and *Chrysoviridae* (multi-segmented dsRNA genome). Phylogenetic analyses show that quadriviruses are more closely related to members of the family *Totiviridae* than to other known quadripartite dsRNA viruses, such as chrysovirus and alternaviruses (Fig. 5, alternaviruses are not included). Quadriviruses and members of the genus *Totivirus* that have unsegmented genomes with two large ORFs for a capsid protein and a RdRp, form a well-supported monophyletic cluster within the *Totiviridae* family clade, suggesting that they descended from a common ancestor except for a unique totivirus, *Ustilago maydis* virus H1 (UmV-H1), that has a single ORF encoding a polypeptide with a capsid protein and RdRp domains. UmV-H1 and yeast L-A totivirus commonly have additional components (i.e., satellite dsRNAs), which encode secreted toxin precursor proteins responsible for the killer phenomena. UmV-H1 is phylogenetically related to recently discovered unsegmented dsRNA mycoviruses (relatives of UmV-H1) and botybirnaviruses (two-segmented dsRNA genome) (Fig. 5). Therefore, it is necessary to revise the taxonomy of the current members of the *Totivirus* genus based on phylogenetic relationships. Similar to quadriviruses, the betachrysovirus and botybirnavirus capsid may be composed of two different structural proteins. In addition, some plant-infecting deltapartitiviruses may have three dsRNA genome segments and encode two different structural proteins. The acquisition of two different structural protein genes (quadriviruses and probably betachrysoviruses), the duplication of a structural protein gene (botybirnaviruses and presumably some tripartite partitiviruses) and the duplication of a helix-rich domain within one single gene (alphachrysoviruses) possibly occurred independently during the course of evolution of these dsRNA mycoviruses.

At the time of writing this article, a new quadrivirus, designated as *Leptosphaeria biglobosa* quadrivirus 1, with a quadripartite dsRNA genome, was reported from the blackleg disease fungus *Leptosphaeria maculans* (anamorph *Phoma lingam*, class *Dothideomycetes*) infecting oilseed rape (*Brassica napus*). This virus shares many molecular characteristics with RnQV1, but its RdRp is more closely related to the putative RdRps encoded by ACD L dsRNA3 and dsRNA4, and CCRS L dsRNA3 and dsRNA4.

See also: Structure of Double-Stranded RNA Mycoviruses

Further Reading

- Chiba, S., Castón, J.R., Ghabrial, S.A., Suzuki, N., ICTV Report Consortium, 2018. ICTV virus taxonomy profile: *Quadriviridae*. *Journal of General Virology* 99, 1480–1481.
- Kondo, H., Kanematsu, S., Suzuki, N., 2013. Viruses of the white root rot fungus, *Rosellinia necatrix*. *Advances in Virus Research* 86, 177–214.
- Kozlakidis, Z., Covelli, L., Di Serio, F., *et al.*, 2006. Molecular characterization of the largest mycoviral-like double-stranded RNAs associated with Amasya cherry disease, a disease of presumed fungal aetiology. *Journal of General Virology* 87, 3113–3117.
- Lin, Y.-H., Chiba, S., Tani, A., *et al.*, 2012. A novel quadripartite dsRNA virus isolated from a phytopathogenic filamentous fungus, *Rosellinia necatrix*. *Virology* 426, 42–50.
- Lin, Y.-H., Hisano, S., Yaegashi, H., Kanematsu, S., Suzuki, N., 2013. A second quadrivirus strain from the phytopathogenic filamentous fungus *Rosellinia necatrix*. *Archives of Virology* 158, 1093–1098.
- Luque, D., Mata, C.P., Gonzalez-Camacho, F., *et al.*, 2016. Heterodimers as the structural unit of the $T = 1$ capsid of the fungal double-stranded RNA *Rosellinia necatrix* quadrivirus 1. *Journal of Virology* 90, 11220–11230.
- Luque, D., Mata, C., Suzuki, N., Ghabrial, S., Castón, J., 2018. Capsid structure of dsRNA fungal viruses. *Viruses* 10, 481.
- Mata, C.P., Luque, D., Gómez-Blanco, J., *et al.*, 2017. Acquisition of functions on the outer capsid surface during evolution of double-stranded RNA fungal viruses. *PLoS Pathogens* 13, e1006755.
- Sato, Y., Castón, J.R., Suzuki, N., 2018. The biological attributes, genome architecture and packaging of diverse multi-component fungal viruses. *Current Opinion in Virology* 33, 55–65.
- Shah, U.A., Kotta-Loizou, I., Fitt, B.D., Coutts, R.H., 2019. Identification, molecular characterization, and biology of a novel quadrivirus infecting the phytopathogenic fungus *Leptosphaeria biglobosa*. *Viruses* 11, 9.
- Yaegashi, H., Sawahata, T., Ito, T., Kanematsu, S., 2011. A novel colony-print immunoassay reveals differential patterns of distribution and horizontal transmission of four unrelated mycoviruses in *Rosellinia necatrix*. *Virology* 409, 280–289.
- Yaegashi, H., Shimizu, T., Ito, T., Kanematsu, S., 2016. Differential inductions of RNA silencing among encapsidated double-stranded RNA mycoviruses in the white root rot fungus, *Rosellinia necatrix*. *Journal of Virology* 90, 5677–5692.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsna-viruses/w/quadriviridae/Quadriviridae.

Totiviruses (*Totiviridae*)¹

Bradley I Hillman and Alanna B Cohen, Rutgers University, New Brunswick, NJ, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of S.A. Ghabrial, Totiviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00518-5.

Glossary

Anastomosis Fusion between hyphal branches allowing for exchange of cytoplasm, ions, and organelles, and the most common means of horizontal transmission of mycoviruses.

Conidia Asexual, non-motile fungal spores.

Mycoviruses Viruses that infect and multiply in fungi.

Protoplast A cell that has its cell wall removed, allowing for membrane fusion.

Pseudoknot A secondary structure in viral mRNA that slows movement of the ribosome and may cause a frameshift that allows entry to an alternative reading frame during translation.

Ribosomal frameshifting Ribosomes switching reading frame on an mRNA, in response to the presence of a slippery site and/or a pseudoknot, to synthesize a fusion protein or a polyprotein from two overlapping reading frames.

RNA silencing The negative regulation or inhibition of gene expression by non-coding RNA elements, also called RNA interference (RNAi).

Totivirids Viruses within the family *Totiviridae*.

Totiviruses Viruses within the genus *Totivirus*.

Vegetative incompatibility Genetic differences between fusing vegetative hyphae of fungi that result in unviable heterokaryons and prevent horizontal mycovirus transmission.

Introduction

The discovery of the killer phenomenon in the 1960s in yeast (*Saccharomyces cerevisiae*) and in the smut fungus (*Ustilago maydis*) eventually led to the discovery of the isometric double-stranded (ds) RNA mycoviruses with nonsegmented genomes, currently classified in the genus *Totivirus* (family *Totiviridae*), distinguishing them by that name from the partitiviruses, which had segmented genomes. An interesting feature of these viruses became apparent: virus-infected killer strains of yeast or smut secrete a protein toxin to which they are immune, but which is lethal to sensitive cells. The precursor to the killer toxin is now known to be encoded by a satellite dsRNA, which is dependent on a helper totivirus for encapsidation and replication.

Unlike the totiviruses associated with the yeast and smut killer systems, other members of the family *Totiviridae* are not known to be associated with killer phenotypes. The other accepted genera in the family *Totiviridae* include the viruses that infect filamentous fungi (members of the genus *Victorivirus*) and those that infect parasitic protozoa (members of the genera *Giardiavirus*, *Leishmaniavirus*, and *Trichomonasvirus*). Genera proposed to be included in the family *Totiviridae* include *Artivirus* and *Eimeriavirus*, neither of which contain toxin-producing member viruses. The viruses constituting the family *Totiviridae*, including proposed members, will be referred to here as totivirids, distinguishing them from members of the genus totivirus.

Totivirids that infect fungi and protozoa were the first dsRNA viruses with undivided genomes. The new dsRNA virus family *Amalgaviridae* also has undivided genomes with an architecture similar to members of the *Totiviridae*, but they are phylogenetically more closely related to the *Partitiviridae* family of viruses. Purified preparations of some of several totivirids and similar viruses contain dsRNA species suspected of being satellite or defective dsRNAs. An extreme example of this phenomenon is the relationship between yado-nushi virus 1 (YnV1), an unclassified totivirus most closely related to members of the genus *Giardiavirus*, which hosts and encapsidates a viral RNA, yado-kari virus 1 (YkV1), that is most closely related to members of the positive-sense (+)RNA virus family *Caliciviridae*.

The realization by the scientific community of the breadth of the family *Totiviridae* is leading to broader interest in these viruses among virologists. Further spurring this interest is the realization that several of these viruses have measurable effects on hosts of economic importance, including reduction of virulence or fitness in filamentous fungi, disease in crustaceans and fish, and enhanced virulence in protozoans that cause human disease. As a group, the totivirids constitute a rapidly expanding family, with metagenomic information revealing new members on a continual basis. In many instances, host-virus relationships are unknown, but the growing set of tools available for study is demonstrating the biological importance of these viruses.

In addition to examining the similarities and differences among members of the family *Totiviridae* (totivirids), this article will focus on the totivirids that infect filamentous fungi, a group of viruses that are phylogenetically more closely related to each other

¹Authors' note: This article is an update of the same chapter in the Third Edition written by Dr. Said A. Ghabrial. Dr. Ghabrial died before the rewrite of this article began, and thus was never able to have input into the content of the updated chapter. So even though roughly half of the content here is originally Dr. Ghabrial's, he is not listed as a co-author.

than to other totiviruses and utilize a different strategy to express their genomes. The yeast and smut totiviruses and associated killer systems, as well as the totivirids infecting parasitic protozoa, are discussed elsewhere in this encyclopedia.

Taxonomy of and Evolutionary Relationships Among Totivirids

The family *Totiviridae* currently accommodates five genera, *Giardiavirus*, *Leishmanivirus*, *Totivirus*, *Trichomonasvirus*, and *Victorivirus*. There are two additional proposed genera within the family, *Artivirus* and *Eimeriavirus*, which are not yet recognized by the International Committee on Taxonomy of Viruses (ICTV). The genus *Victorivirus*, derived from the specific epithet of *Helminthosporium victoriae*, was erected to accommodate HvV190S, and currently the largest in the family *Totiviridae*, with 14 ICTV-recognized members and many tentative victoriviruses that infect filamentous fungi. The members of one genus can be distinguished from those of another genus in genome expression strategy, phylogenetic relation, and/or natural host range (see below). The family encompasses a broad range of viruses characterized by isometric particles, ~40 nm in diameter, that contain a monosegmented dsRNA genome coding for a capsid protein (CP) and an RNA-dependent RNA polymerase (RdRp). Viruses within the *Totiviridae* are found in the major clade that contains most of the dsRNA viruses, including the largest dsRNA family, the *Reoviridae*. This branch that includes dsRNA viruses is thought to have arisen from the branch of (+)ssRNA viruses containing the alpha and flaviviruses, and the dsRNA branch in turn is hypothesized to have given rise to the negative-sense RNA viruses.

Sequence comparisons of the predicted amino acid sequences of totivirid RdRps indicate that they share significant sequence similarity and characteristically contain eight conserved motifs. This sequence similarity was common to all totivirids so far characterized including the totivirids that infect the yeast, smut, filamentous fungi, and plants as well as those infecting parasitic protozoa. It is notable that an expanded phylogenetic tree based on RdRp sequences of all dsRNA viruses makes it apparent that viruses currently contained within – or proposed to be contained within – the family *Totiviridae* does not form a monophyletic group. In particular, the genera *Victorivirus*, *Leishmanivirus*, *Trichomonasvirus*, and proposed genus *Eimeriavirus* consistently group together, but the genera *Totivirus* and *Giardiavirus* and the proposed genus *Artivirus* do not group with the others. Instead, each of those last three represents a separate small clade within the greater clade of dsRNA viruses related to the totivirids, with viruses representing families including the *Chrysoviridae* and *Quadriviridae*, as well as several clades not yet classified intercalated into that larger “Toti-Chryso” clade. These facts suggest the necessity of taxonomical reorganization of these members of the family *Totiviridae* and related viruses with a similar genome organization but different phylogenetic relationships.

Phylogenetic analysis based on multiple alignments of amino acid sequences of totivirids RdRp conserved motifs (Fig. 1) shows that the viruses infecting filamentous fungi (victoriviruses) are most closely related to each other and form a distinct large well-supported cluster (bootstrap value of 97%). Likewise, phylogenetic trees based on the CP sequences showed similar topology to those based on RdRp sequences, although in some cases with less support (Fig. 2). The leishmaniviruses LRV1, LRV2, and LRV4 (percent sequence identities of 46%) are most closely related to each other and form a discrete cluster with 100% bootstrap value. This is also true for the two yeast viruses ScV-L-A and ScV-LBC (RdRp identity of 32%), members of the genus *Totivirus*. UmV-H1, the third member in the genus *Totivirus* forms a phylogenetic clade distinct from the cluster of the two yeast totiviruses. This may reflect the difference in RdRp expression strategy between these viruses (see below). It is of interest that HvV190S and related victoriviruses are phylogenetically more closely to the protozoan-infecting members of the *Totiviridae*, including trichomonasviruses, leishmaniviruses, and the proposed genus *Eimeriavirus*, typified by the unclassified *Eimeria brunetti* RNA virus 1 (EbrV-1), which infects an apicomplexan parasitic protozoan, than to the yeast viruses (Figs. 1 and 2). It was previously hypothesized that leishmaniviruses and the fungal totivirids are of old origin having existed in a single-cell-type progenitor prior to the divergence of fungi and protozoa. The closer relationship of victoriviruses to protozoal viruses than to yeast viruses in the genus *Totivirus* is consistent with this idea.

Virion Properties

The buoyant densities in CsCl of virions of members of the family *Totiviridae* range from 1.36 to 1.43 g cm⁻³, and the sedimentation coefficients of these virions range from 160S to 190S (S20w, in Svedberg units). Particles lacking nucleic acid sediment with apparent sedimentation coefficients of 90–105S. Isolates of ScV-L-A and UmV-H1 may have additional components, containing satellite or defective dsRNAs, with different sedimentation coefficients and buoyant densities. Virion-associated RdRp activity can be detected in all totivirids examined to date. Protein kinase activity is associated with HvV190S virions; capsids contain phosphorylated forms of CP.

Virion Structure and Composition

The totivirids have isometric particles, approximately 40–50 nm in diameter, with icosahedral symmetry (Fig. 3). The capsids are single-shelled and are composed of a single major polypeptide. The capsids consist of 120 CP subunits of molecular mass in the range of 76–98 kDa. The capsid structures of six members or tentative members of the family *Totiviridae* have been determined

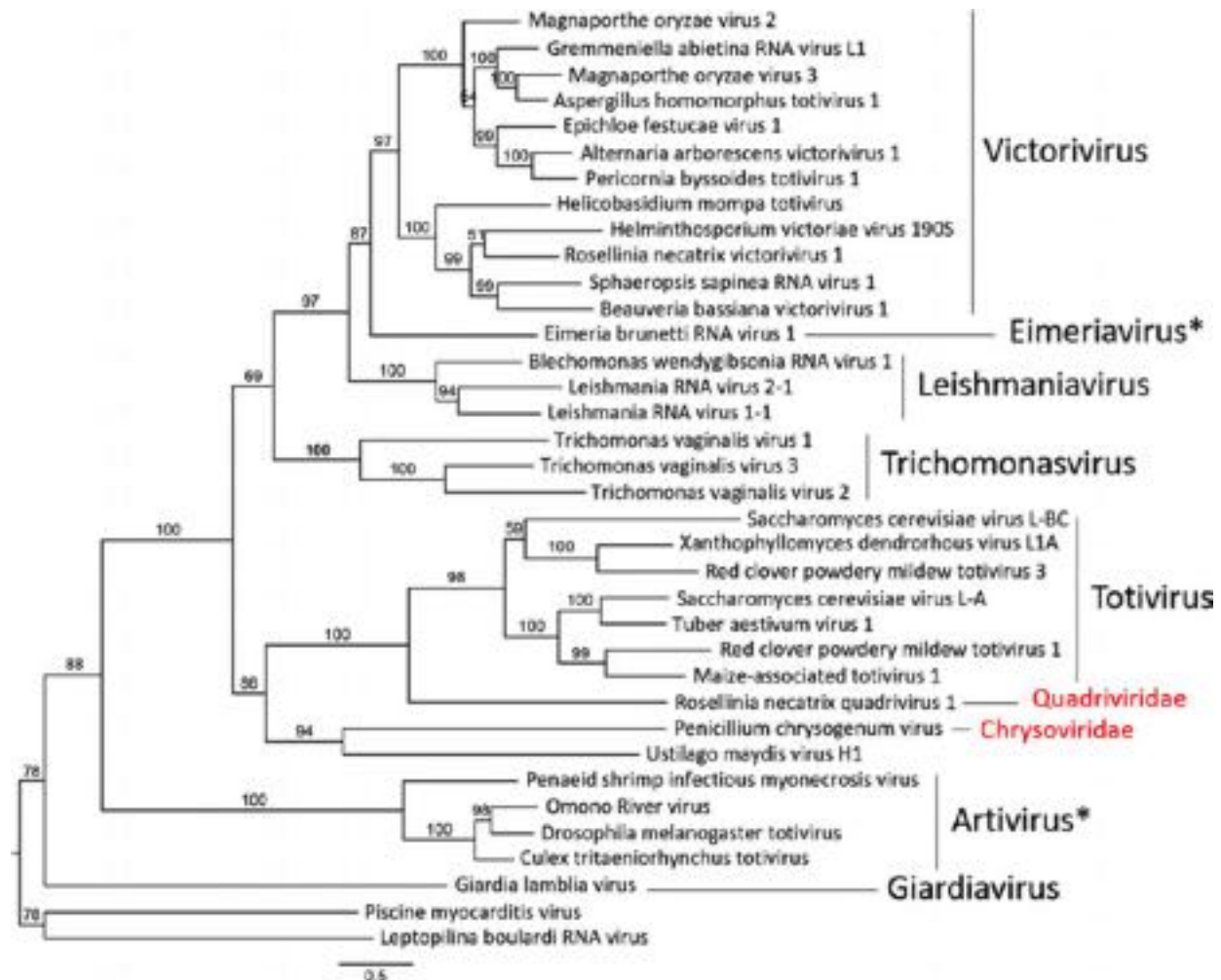


Fig. 1 Maximum Likelihood (ML) phylogenetic tree constructed based on the RNA-dependent RNA polymerase (RdRp) sequences of selected members and proposed member of the family *Totiviridae*. Not all family members were included for space considerations, but the overall topology of the tree generally reflects the topology of the tree when all available sequences were included. A single member of the family *Quadriviridae* and the family *Chrysoviriidae* was included to indicate positions within the larger clade that currently constitutes the family *Totiviridae*. The RdRp sequences were derived from aligned deduced amino acid sequences of members of the family *Totiviridae* using the program MUSCLE. The phylogenetic tree was generated using the program IQ-TREE with 1000 ultrafast bootstrap replicates. A VT + F + I + G4 model of evolution was selected based on Akaike and Bayesian information criteria. Bootstrap support above 70% are indicated on branches.

using cryo-transmission electron microscopy combined with three-dimensional image reconstruction (Fig. 4). In all cases, the capsids of the fungal totivirids are made up of 60 asymmetric CP dimers arranged in a so-called 'T=2' layer, equivalent to a T=1 structure in which the structural units are homodimers rather than monomers. There are structural differences among the viruses examined to date. Surface differences of particles are apparent across the family. Compared to the yeast L-A capsid, the HvV190S capsid shows relatively smoother outer surfaces, but both are more apparently icosahedral, while the trichomonasvirus TVV1, giardiavirus GLV, and proposed artiviruses OmRV and IMNV are more apparently spherical in cryoEM reconstructions. IMNV has an overall shape like OmRV, but it uniquely has fiber protrusions at the pentamers. The quaternary organization of the HvV190S particle is notably similar to that of yeast L-A and the cores of the larger dsRNA viruses of plants and animals: the A-subunits cluster around the fivefold axis and B-subunits around the threefold axis.

The ScV-L-A CP removes the 5' cap structure of host mRNA and covalently attaches it to the histidine residue at position 154. The decapping activity is required for efficient translation of viral RNA. The published yeast L-A capsid structure reveals a trench at the active site of decapping.

Like other totivirids, a single gene encodes the capsid of HvV190S, but the HvV190S capsid comprises two closely related major CPs, either p88 and p83 or p88 and p78. The capsids of all other totivirids so far characterized appear to contain only a single major CP. Interestingly, HmTV-1-17, a totivirid infecting a filamentous fungus, has similar capsid heterogeneity to that of HvV190S. Purified HvV190S virion preparations contain two types of particles, 190S-1 and 190S-2, which differ slightly in sedimentation rates (190S-1 is resolved as a shoulder on the slightly faster sedimenting component 190S-2) and capsid

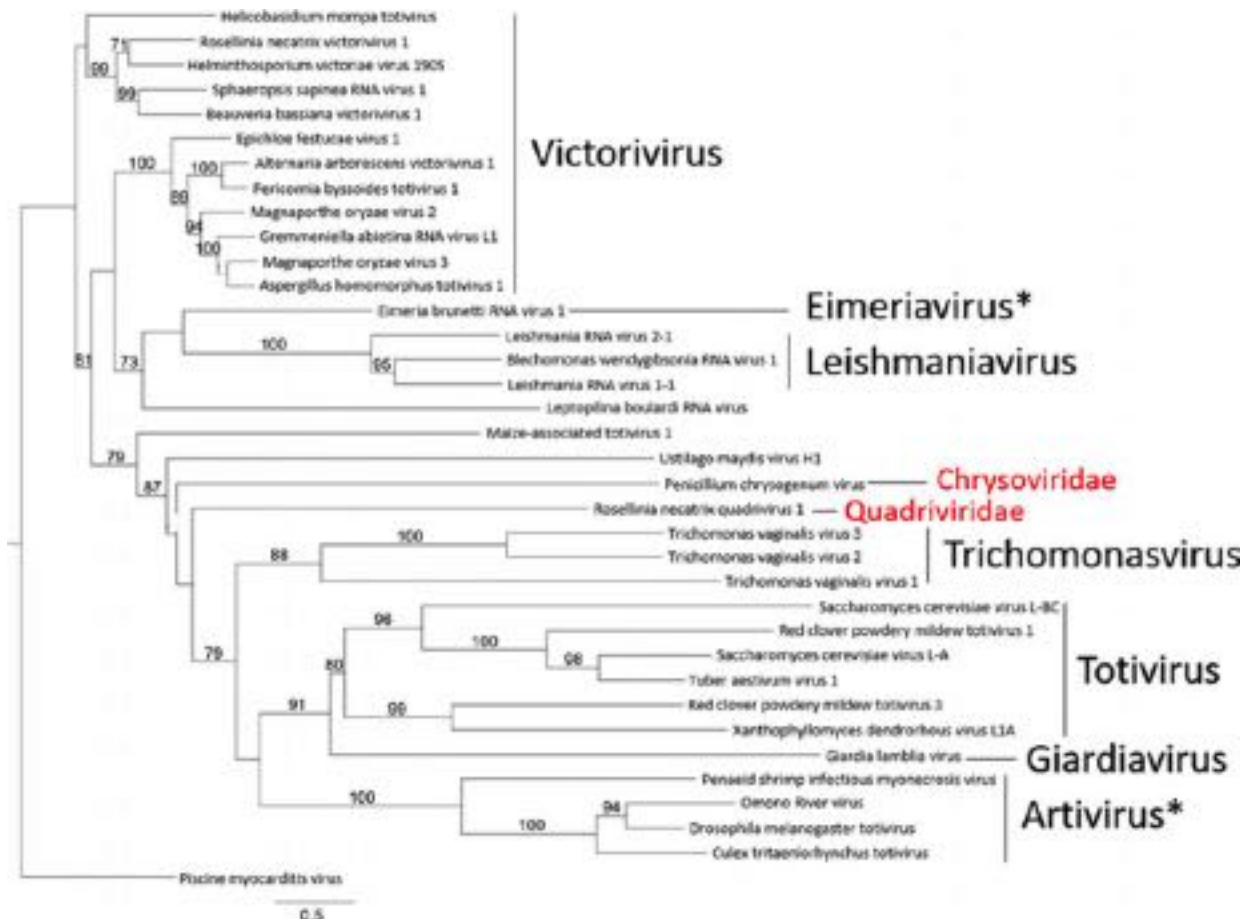


Fig. 2 Maximum Likelihood phylogenetic tree constructed based on the capsid protein (CP) sequences of selected members and proposed member of the family Totiviridae. The viruses and strains selected and the methods used were the same as those used in Fig. 7. A VT + F + G4 model of evolution was selected based on Akaike and Bayesian information criteria.

composition. The 190S-1 and 190S-2 virions are believed to represent different stages in the virus life cycle. The 190S-1 capsids contain p88 and p83, occurring in approximately equimolar amounts, and the 190S-2 capsids comprise similar amounts of p88 and p78. p88 and p83 are phosphoproteins, whereas p78 is nonphosphorylated (Fig. 5). Totivirid virions encapsidate a single molecule of dsRNA, 4.6–7.0 kbp, with tentative members of the family ranging up to 7.8 kbp. Some totivirids may additionally contain satellite dsRNAs or defective dsRNAs, which are encapsidated separately in capsids encoded by the totivirid genome.

Genome Organization and Expression

In general, the genome organization of the totivirids infecting fungi and other eukaryotes are similar: each virus genome contains two large open reading frames (ORFs); the 5' proximal ORF encodes a CP and the 3' ORF encodes an RdRp (Fig. 6).

Two distinct RdRp expression strategies have been reported for viruses in the family Totiviridae: those that express RdRp as a fusion protein (CP–RdRp) by ribosomal frameshifting, such as *Saccharomyces cerevisiae* virus L-A and the viruses that infect parasitic protozoa; and those that synthesize RdRp as a separate nonfused protein by an internal initiation mechanism, as shown for HvV190S and proposed for all of the other totivirids that infect filamentous fungi. An exception to this rule is *Ustilago maydis* virus H1 (Umv-H1), currently being classified into the genus Totivirus. The Umv-H1 genome contains only a single ORF that encodes a polyprotein, predicted to be autocatalytically processed by a viral papain-like protease to generate the CP and RdRp proteins.

In most *Giardiavirus*, *Totivirus*, and proposed *Artivirus* members, a –1 frameshift regulates translation between the two main ORFs; in members of *Trichomonasvirus* a –2 frameshift regulates translation, and in *Leishmanivirus* a +1 frameshift regulates translation between the two main ORFs. In each of these cases, a fusion protein consisting of the CP and RdRp occasionally results, and that fusion product is incorporated in virions. In most members of the genus *Victorivirus*, a translational stop-start separates the two main ORFs, with the termination codon of the CP ORF overlapping the initiation codon of the RdRp ORF in the tetranucleotide sequence AUGA or UAAUG. Other translational details in members or proposed members of the family Totiviridae

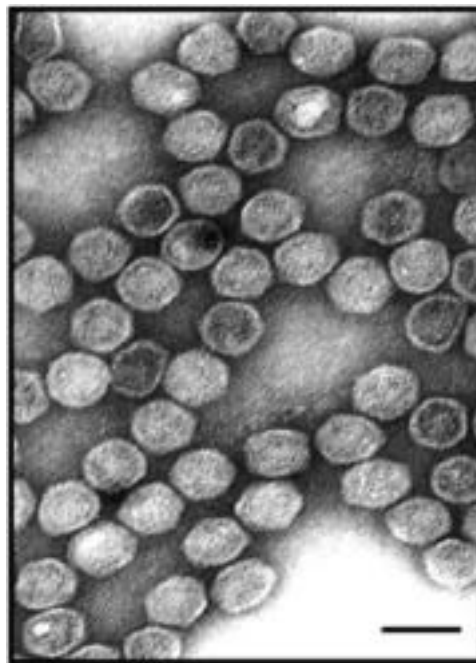


Fig. 3 Negative contrast electron micrograph of particles of an isolate of *Helminthosporium victoriae* 190SV, the type species of the genus Victorivirus. Scale = 50 nm. Reprinted from Ghabrial, S.A., 2008. Totiviruses. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), Encyclopedia of Virology (Third Edition), New York: Academic Press, pp. 163–174.

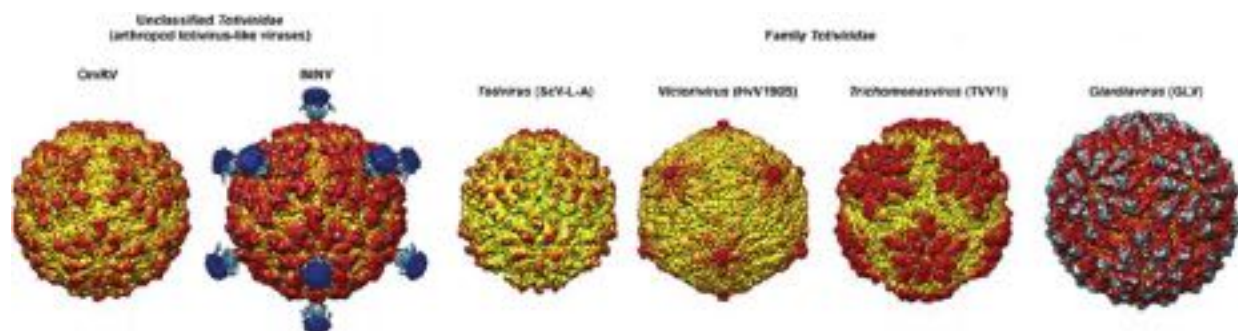


Fig. 4 Structural comparison of totivirid, OmRV, and other totivirus-related viruses. Electron density volume map of Omono River virus (OmRV, 8.3 Å resolution), infectious myonecrosis virus (IMNV, 8.0 Å resolution), *Saccharomyces cerevisiae* virus L-A (ScV-L-A, 8.5 Å resolution), *Helminthosporium victoriae* virus 190S (HvV190S, EMD Accession code: 5726, 7.5 Å resolution), *Trichomonas vaginalis* virus 1 (TVV1, EMD Accession code: 2184, 6.7 Å resolution), and *Giardia lamblia* virus (GLV, EMD Accession code: 5948, 6.0 Å resolution). The 8.5 Å electron density volume map of ScV-L-A was generated from the crystallographic capsid model (PDB ID: 1M1C) using Chimera60. The surface of the structures was colored according to the distance from the center of the virion (green < 131 Å, yellow < 193 Å, red < 220 Å, light blue < 250 Å, blue < 280 Å). Reprinted from Okamoto, K., Miyazaki, N., Larsson, D.S.D., *et al.*, 2016. The infectious particle of insect borne totivirus-like Omono River virus has raised ridges and lacks fiber complexes. Springer. Scientific Reports 6, 1–14.

may include additional small ORFs, additional protease cleavage sites, 2A-like and pseudo 2A-like sequences, and pseudoknot sequences that facilitate frameshifting.

The totivirids express their RdRps either as CP–RdRp (gag-pol-like) fusion proteins or as separate nonfused proteins. Expression of RdRp as a CP–RdRp fusion protein via -1 ribosomal frameshifting has been well documented for ScV-L-A. Virion-associated CP–RdRp has been detected as a minor protein in the capsids of ScV-L-A, ScV-L-BC, TVV, and GLV. Expression of RdRp as a fusion protein in LRV1-1, LRV1-4, or LRV2-1 by $+1$ ribosomal frameshifting or ribosomal hopping has been proposed.

The overlap region between ORF1 and ORF2 of ScV-L-A (130 nt), LRV1-1 and LRV1-4 (71 nt), and GLV (122 nt) contains the structures necessary for ribosomal frameshifting including a slippery site and a pseudoknot structure to promote fusion of ORF1 and ORF2 *in vivo*. Although the overlap region in TVV is short (14 nt), it contains a potential ribosomal slippage heptamer.

There is currently no evidence for CP–RdRp fusion proteins in HvV190S or other victoriviruses. The overlapping region in the dsRNA genomes of HvV190S is AUGA, where the initiation codon of the RdRp ORF overlaps the termination codon of the CP ORF, suggesting

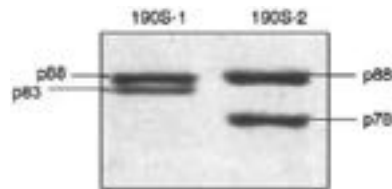


Fig. 5 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of HvV190S sedimenting components 190S-1 and 190S-2. Purified preparations of HvV190S contain two types of particles, 190S-1 and 190S-2, which differ slightly in sedimentation rates. The 190S-1 capsids contain p88 and p83, occurring in approximately equimolar amounts, and the 190S-2 capsids comprise similar amounts of p88 and p78. The capsid proteins p88 and p83 are phosphorylated, whereas p78 is nonphosphorylated. Reprinted from Ghabrial, S.A., 2008. Totiviruses. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), Encyclopedia of Virology (Third Edition), New York: Academic Press, pp. 163–174.

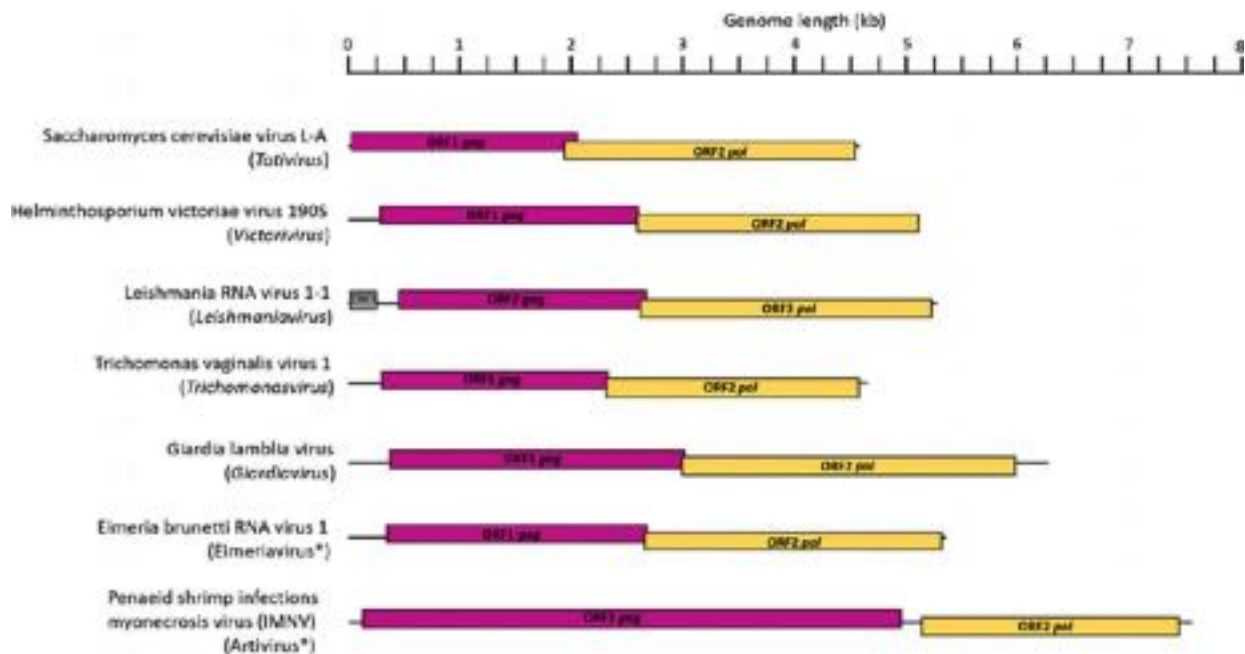


Fig. 6 Genome organizations of selected members and proposed members of the family Totiviridae. The dsRNA genome encompasses two large open reading frames (ORFs) with the 5' ORF encoding a capsid protein (CP) and the 3' ORF encoding an RNA-dependent RNA polymerase (RdRp). Details are described in the text. Except for the members of the genus *Victorivirus*, translation of the RdRp ORF proceeds as a consequence of a frameshift from the CP ORF, resulting in a CP-RdRp fusion protein that is a component of the capsid. Translation of the RdRp of members of the genus *Victorivirus* proceeds by internal ribosome initiation following termination of translation of the CP sequence at a translational stop-start site.

that expression of RdRp occurs by a mechanism different from translational frameshifting. For HvV190S, the stop codon of the CP ORF (nucleotide position 2606-UGA-2608) overlaps with the start codon (nucleotide position 2605-AUG-2607) for the RdRp ORF in the sequence AUGA. Other victoriviruses have either the pentanucleotide UAAUG or, in the case of *Alternaria alternata* victorivirus 1 (AaVV1), a putative –1 frameshift. But in none of these instances has a CP-RdRp fusion protein been identified in translation experiments or in virus particles; rather, each appears to represent a slightly different mechanism for termination of the CP and initiation of the RdRp.

The 5' end of the plus strand of all totivirids dsRNAs examined to date is uncapped and the 3' end is not polyadenylated. As noted, ScV-L-A uses a cap-snatching mechanism to provide capped mRNAs for efficient translation. The 5' end of HvV190S dsRNA is uncapped and highly structured and contains a relatively long (289 nucleotides) 5' leader with two minicistrons. These structural features of the 5' untranslated region (UTR) of HvV190S dsRNA suggest that the CP-encoding ORF1 (with its AUG present in suboptimal context according to the Kozak criteria) is translated via a cap-independent mechanism. The 5'-UTR of the leishmanivirus LRV-1-1 functions as an internal ribosome entry site (IRES). Translation of the uncapped giardiavirus GLV mRNA in *Giardia lamblia* is initiated on a unique IRES element that contains sequences from a part of the 5'-UTR and a portion of the capsid coding region. IRES activities are also observed in the 5'-UTR of the victoriviruses HvV190S and *Rosellinia necatrix* victorivirus 1.

The UGA codon at position 2606–2608 of the HvV190S genomic plus strand was verified by site-directed mutagenesis to be the authentic stop codon for ORF1. The RdRp-encoding downstream ORF2 of HvV190S is in a –1 frame with respect to ORF1 and is expressed via an internal initiation mechanism (a coupled termination–initiation mechanism is proposed). The HvV190S RdRp is detectable as a separate, virion-associated component, consistent with its independent translation from ORF2.

The viral genome of IMNV, a monosegmented dsRNA virus infecting penaeid shrimp and tentatively assigned to the family *Totiviridae* in the proposed genus *Artivirus*, encompasses two nonoverlapping ORFs with ORF1 encoding a polyprotein comprised of a putative RNA-binding protein and a CP. The coding region of the RNA-binding protein is located in the first half of ORF1 and contains a dsRNA-binding motif in the first 60 amino acids. The second half of ORF1 encodes a CP, as determined by amino acid sequencing, with a molecular mass of 99 kDa. ORF2 encoded a putative RdRp with the eight conserved motifs characteristic of totivirids. Phylogenetic analysis based on the RdRp clustered IMNV with other invertebrate-infecting members in the family *Totiviridae* and separate from the five established genera. Important novel features of the genome organization of IMNV have significant bearing on how the viral proteins are expressed, and these features appear to be shared among closely related viruses and are continuing to evolve. These features include two encoded '2A-like' motifs, which are likely involved in ORF1 polyprotein 'cleavage', a 199 nt overlap between ORF1 and ORF2, and the presence a 'slippery heptamer' motif and predicted RNA pseudoknot in the region of ORF1–ORF2 overlap. The latter features probably allow ORF2 to be translated as a fusion with ORF1 by –1' ribosomal frameshifting. The generation of CP as a polyprotein and the potential involvement of the IMNV-encoded '2A-like' peptides (GDVESNPGP and GDVEENPGP) in processing of the polyprotein to release the major CP represent novel features that are not shared by closely related viruses in the proposed *Artivirus* genus of totivirids. Interestingly, more divergent viruses are predicted to have "pseudo 2A" sequences that could give rise to functional "2A-like" peptides by random mutation. In contrast, the potential expression of RdRp as a CP–RdRp fusion protein via ribosomal frameshifting is a common strategy utilized by other totivirids for expressing their RdRps.

Virus Replication Cycle

The replication cycle of totivirids has mainly been derived from in vitro studies of virion-associated RNA polymerase activity and the isolation of particles representing various stages in the replication cycle. In in vitro reactions, the RNA polymerase activity associated with virions of the fungal totivirids ScV-L-A, UmV-H1, and HvV190S, isolated from lag-phase cultures, catalyzes end-to-end transcription of dsRNA, by a conservative mechanism, to produce mRNA for CP, which is released from the particles. Purified ScV-L-A virions, isolated from log-phase cells, contain a less-dense class of particles, which package only plus-strand RNA. In in vitro reactions, these particles exhibit a replicase activity that catalyzes the synthesis of minus-strand RNA to form dsRNA. The resultant mature particles, which attain the same density as that of the dsRNA-containing virions isolated from the cells, are capable of synthesizing and releasing plus-strand RNA.

A proposed life cycle of HvV190S is depicted in [Fig. 7](#). Host-encoded protein kinase and protease have been shown to be involved in post-translational modification of CP. Phosphorylation and proteolytic processing are proposed to play a role in the virus life cycle; phosphorylation of CP may be necessary for its interaction with viral nucleic acid and/or phosphorylation may regulate dsRNA transcription/replication. Proteolytic processing and cleavage of a C-terminal peptide, which leads to dephosphorylation and the conversion of p88 to p78, may play a role in the release of the plus-strand RNA transcripts from virions ([Fig. 7](#)).

Biological Properties

The overall host range of the *Totiviridae*, including proposed members, includes an increasing breadth of eukaryotes. In addition to fungal and protozoal hosts, totivirids have been isolated from oomycetes, plants, insects, crustaceans, fish, and bats. The first such virus identified was a nonsegmented dsRNA virus with isometric particles, designated penaeid shrimp infectious myonecrosis virus (IMNV), isolated from diseased penaeid shrimp and tentatively assigned to the family *Totiviridae*. IMNV is the causal agent of the shrimp myonecrosis disease characterized by necrosis of skeletal muscle, particularly in the distal abdominal segments and tail fan. Phylogenetic analysis based on the RdRp region of viruses in the family *Totiviridae* suggests that *Giardia lamblia* virus (GLV; genus *Giardiavirus*) is the closest relative to IMNV. Related viruses infecting fish, insects, and bats cluster together phylogenetically and are proposed to constitute a new genus, *Artivirus*, which is not yet recognized by the International Committee on Taxonomy of Viruses (ICTV). The complete nucleotide sequences of many members and tentative members of the family *Totiviridae* have now been described.

There are no known natural vectors for the transmission of the fungal totivirids. They are transmitted intracellularly during cell division, sporogenesis, and cell fusion. Although the yeast totiviruses are effectively transmitted via ascospores, the victoriviruses infecting the ascomycetous filamentous fungi are essentially eliminated during ascospore formation. The leishmaniviruses and trichomonasviruses are propagated during cell division and are not infectious as purified virions. Members of the genera *Giardiavirus*, *Totivirus*, *Victorivirus*, and the proposed genus *Artivirus* such as IMNV, on the other hand, are infectious as purified virions. Successful transfection of the protozoan *Giardia lamblia* has also been accomplished via electroporation with plus-strand RNA transcribed in vitro from GLV dsRNA. Whereas IMNV causes a myonecrosis disease in its crustacean host, GLV is associated with latent infection of its flagellated protozoan human parasite *G. lamblia*. GLV is released into the medium without lysing the host cells and the extruded virus can infect many virus-free isolates of the protozoan host.

There is limited information on experimental host ranges for the fungal totivirids because infectivity as particles has been demonstrated for only a few. As a consequence of their intracellular modes of transmission, the natural host ranges of fungal totivirids are limited to individuals within the same or closely related vegetative compatibility groups, but particle-based and RNA-based infectivity systems are allowing for broader host range studies. Furthermore, mixed infections with two or more unrelated viruses are

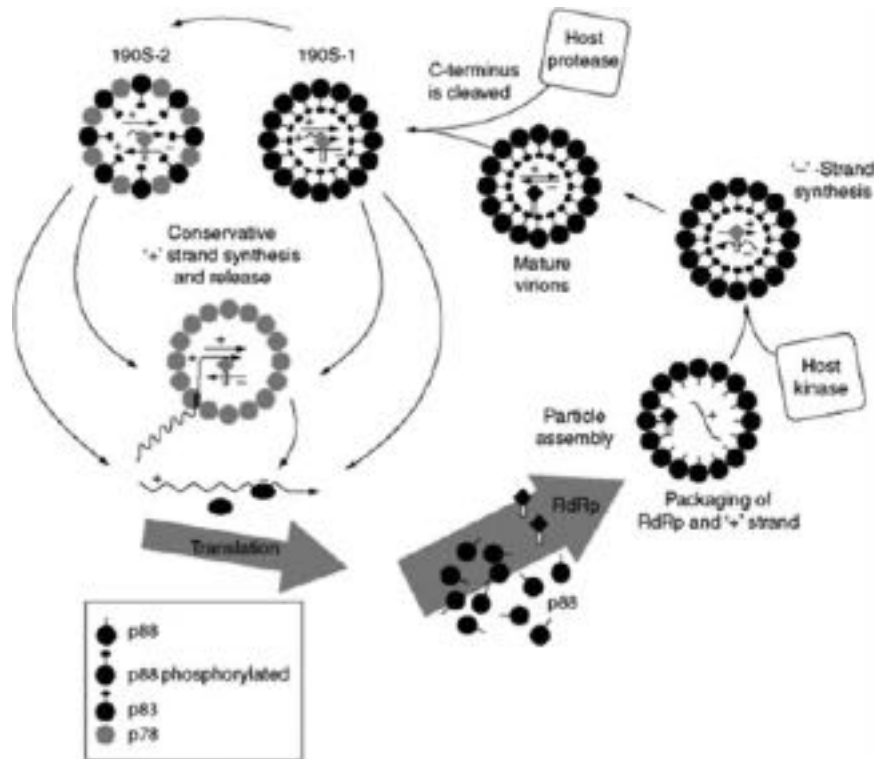


Fig. 7 Life cycle of HvV190S. Mature virions contain a single dsRNA molecule and their capsids are composed entirely or primarily of the capsid protein (CP) p88. Virions representing different stages of the virus life cycle can be purified from the infected fungal host *Helminthosporium victoriae* including the well-characterized 190S-1 and 190S-2 virions. These two types of virions differ in sedimentation rate, phosphorylation state, and CP composition; 190S-1 capsids contain p88 and p83, whereas the 190S-2 capsids contain p88 and p78 (p88 is the primary translation product of the CP gene; p83 and p78 represent post-translational proteolytic processing products of p88 at its C-terminus). p88 and p83 are phosphorylated, whereas p78 is nonphosphorylated. The virions with phosphorylated CPs (p88 + p83) have significantly higher transcriptase activity *in vitro* than those containing the nonphosphorylated p78. Transcription occurs conservatively and the newly synthesized plus-strand RNA is released from the virions. Phosphorylation of CP is catalyzed by a host kinase, and is proposed to play a regulatory role in transcription/replication. A host-encoded protease catalyzes the proteolytic processing of phosphorylated p88; this occurs in two steps, leading first to p83 (the generation of the 190S-1 virions) and then to p78 (190S-2 virions). The conversion of p88-p83-p78 is proposed to play a role in the release of the plus-strand RNA transcripts from virions. The released plus-strand RNA is the RNA that is translated into CP and RNA-dependent RNA polymerase (RdRp) and packaged in capsids assembled from the primary translation product p88. It is not known whether p88 is phosphorylated before or after assembly. Synthesis of minus-strand RNA occurs on the plus-strand RNA template inside the virion; phosphorylation may be involved in turning on the replicase activity. Reprinted from Ghabrial, S.A., 2008. Totiviruses. In: Mahy, B.W.J., Van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology* (Third Edition), New York: Academic Press, pp. 163–174.

common, probably as a consequence of the ways by which fungal viruses are transmitted in nature. Dual infection of yeast with ScV-L-A and ScV-L-BC and the filamentous fungus *Sphaeropsis sapinea* with SsRV-1 and SsRV-2 are examples of mixed infections involving totivirids. Dual infection of *H. victoriae* with the victorivirus HvV190S and the chrysovirus *Helminthosporium victoriae* virus 145 S (HvV145S) is an example of a mixed infection involving two unrelated viruses belonging to two different families.

The victorivirus from the phytopathogenic ascomycete fungus *Rosellinia necatrix*, *Rosellinia necatrix* victorivirus 1 (RnVV1), was eliminated from its host fungus by hyphal tipping and was reintroduced to its natural host and introduced to the fungus *Cryphonectria parasitica* by transfection. Negligible pathogenic effect was observed in either host. However, transfection of RnVV1 to an RNA silencing-deficient strain of *C. parasitica* revealed phenotypic effects of the virus in the RNA silencing-deficient strain, demonstrating that the virus is targeted by anti-viral RNA silencing in the fungus.

Many totivirids appear to be associated with symptomless infections of their hosts, but there is increasing evidence of biological impact of others. IMNV, an unclassified virus tentatively assigned to the proposed new genus Artivirus in the family Totiviridae, causes a myonecrosis disease in shrimp. The closely related Omono River virus isolated from *Culex* mosquitoes causes cytopathic effect in infected cells. Several recent studies suggest that members of the genera *Leishmanivirus* and *Trichomonasvirus* enhance their respective parasite-associated disease in humans by inducing proinflammatory responses at infection sites. These relationships are becoming recognized as increasingly important impacts of members of the family Totiviridae, and are covered in other chapters of this volume.

Multiple infections of totivirids with other totivirids or with unrelated viruses are common. In most instances, such mixed infections have not been dissected biologically. For example, a single isolate of *Trichomonas vaginalis* may be infected with all of the

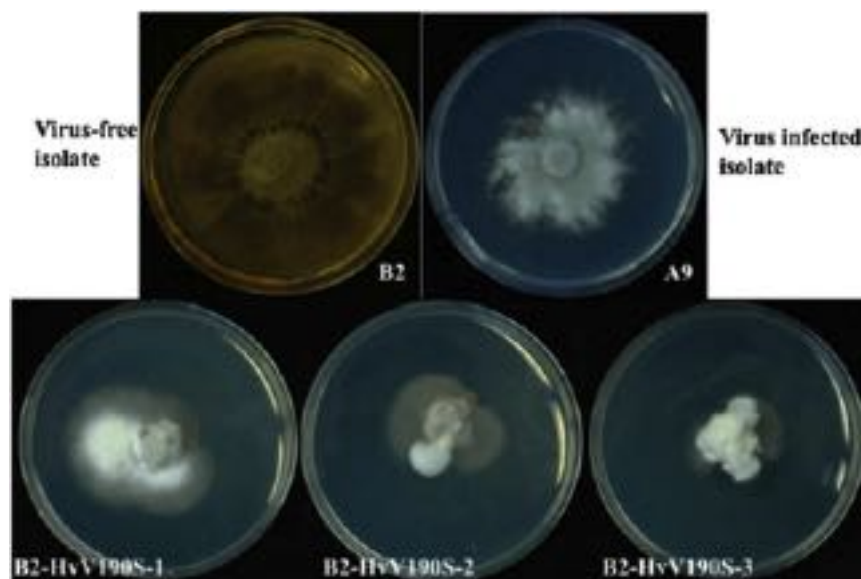


Fig. 8 Colony morphology of virus-free and virus infected isolates B2 and A9, respectively, and three representative transfectants (B2-HvV190S-1, B2-HvV190S-2 and B2-HvV190S-3) of *Helminthosporium victoriae*. All fungal isolates were cultured on potato dextrose agar containing 0.5% (wt/vol) yeast extract medium. Reprinted from Xie, J., Havens, W.M., Lin, Y.-H., Suzuki, N., Ghabrial, S.A., 2016. The victorivirus *Helminthosporium victoriae* virus 190S is the primary cause of disease/hypovirulence in its natural host and a heterologous host. *Virus Research* 213, 238–245, Copyright (2015), with permission from Elsevier.

four known species of TVV. Mixed infections with HvV190S and HvV145S were initially associated with a debilitating disease of the fungal host. The two viruses are transmitted together via hyphal anastomosis and diseased isolates are characterized by reduced growth, excessive sectoring, aerial mycelial collapse, and generalized lysis. The disease phenotype of *H. victoriae*, the causal agent of Victoria blight of oats, is of special interest not only because diseased isolates are hypovirulent, but also because it was among the first examples of the viral etiology of hypovirulence of an important plant pathogen. It was only relatively recently that the role of the victorivirus HvV190S in the fungal disease was elucidated. Colonies regenerated from fungal protoplasts infected singly with HvV190S, but not HvV145S showed the disease phenotype (Fig. 8).

Virus–Host Relationships

The yeast killer system, comprised of a helper totivirus (ScV-L-A) and associated satellite dsRNA (*m*-dsRNA), is one of the few known examples where virus infection is beneficial to the host. The ability to produce killer toxins by yeast strains that in turn are immune to those toxins confers an ecological advantage over sensitive strains. The use of killer strains in the brewing industry provides protection against contamination with adventitious sensitive strains. Totiviruses maintain only the genes that are essential for their own survival (RdRp and CP), but killer viruses host accessory satellite segments that may be beneficial to their survival. The host cells have evolved to support only a defined level of virus replication, beyond which virus infection may become pathogenic. Because of amenability to genetic studies, the yeast–virus system has provided significant information on the host genes required to prevent viral cytopathology. Classical genetic studies with *S. cerevisiae* defined the first systems of chromosomal genes that regulate the copy number of a eukaryotic virus: the copy number of the totivirus ScV-L-A and its satellite M-dsRNAs were found to be downregulated by genes designated SKI (superkiller) or upregulated by genes designated MAK (maintenance of killer). More recently, the ecological flexibility of host/virus interactions in this system has been revealed: two host genes required for RNA silencing (Argonaute, AGO1, and Dicer, DCL1), and thus for interference with ScV-L-A multiplication, were found to be repeatedly gained or lost in different yeast strains, leading to the hypothesis that retaining these general virus resistance genes is more beneficial to the yeast in some circumstances, whereas eliminating these resistance genes and thus facilitating the presence of the killer viruses is more beneficial to the host in other circumstances.

While much less well-defined than the yeast system, the HvV190S victorivirus that infects the plant pathogenic fungus *H. victoriae* also utilizes host-encoded proteins, in this case a protein kinase and a protease for post-translational modification of its CP. Phosphorylation and proteolytic processing of CP may play a role in regulating transcription and the release of plus-strand transcripts from virions (Fig. 7).

The *H. victoriae*–virus system provides a well-studied and useful model system for examining virus–host interactions in a plant pathogenic fungus. A major attribute of this system is that the virus-infected *H. victoriae* isolates exhibit a disease phenotype

(Fig. 8). Furthermore, HvV190S can infect the heterologous fungal host *C. parasitica*, and also results in a disease phenotype in strains deficient in the RNA silencing pathway ($\Delta dcl-2$). It was previously demonstrated that the fungal gene Hv-p68, a novel alcohol oxidase/RNA-binding protein, is upregulated as a result of virus infection suggesting that upregulation of this gene might play a role in virus pathogenesis. Overexpression of Hv-p68 in virus-free fungal isolates, however, resulted in a significant increase in colony growth and did not induce a disease phenotype. Thus, overexpression of Hv-p68 per se is not sufficient to induce the disease phenotype in the absence of virus infection. Because dsRNA-binding proteins are known to sequester dsRNA and suppress antiviral host defense mechanisms, it is feasible that overexpression of the dsRNA-binding protein Hv-p68 may lead to the induction of the disease phenotype by suppressing host defense, but additional studies are needed to determine whether or not Hv-p68 upregulation has a role in viral pathogenesis.

Conclusions

Viruses in the family *Totiviridae* have common properties such as a two-ORF genome architecture and a T=1 capsid composed of 60 asymmetrical capsid protein dimers. Taxonomic considerations of totivirids have proceeded in large part by expanding the family *Totiviridae* to accommodate new genera as they are formally described. The family now contains five ICTV-approved genera and two tentative genera, among virus families with the most diverse host organisms. The genera are differentiated from one another in host range and gene expression. Elucidating the capsid structures of representatives of different genera and proposed genera of the family *Totiviridae* help to inform taxonomy of these viruses. More definitive information on variation in life cycles and genome expression strategies will also aid in putting this large and expanding family of viruses in taxonomic perspective. It is easy to envision a taxonomy in which the viruses considered here represent multiple families, but the utility for the community of virologists and biologists of such a restructuring would be needed, and perhaps should proceed only after the breadth of the family is better understood.

Further Reading

- Castón, J.R., Luque, D., Trus, B.I., *et al.*, 2006. Three-dimensional structure and stoichiometry of *Helminthosporium victoriae* 190S totivirus. *Virology* 347, 323–332.
- Chiba, S., Lin, Y.-H., Kondo, H., Kanematsu, S., Suzuki, N., 2013. A novel victorivirus from a phytopathogenic fungus, *Rosellinia necatrix*, is infectious as particles and targeted by RNA silencing. *Journal of Virology* 87 (12), 6727–6738.
- De Lima, J.G.S., Teixeira, D.G., Freitas, T.T., Lima, J.P.M.S., Lanza, D.C.F., 2019. Evolutionary origin of 2A-like sequences in Totiviridae genomes. *Virus Research* 259, 1–9.
- Fichorova, R., Fraga, J., Rappelli, P., Fiori, P.L., 2017. *Trichomonas vaginalis* infection in symbiosis with *Trichomonasvirus* and *Mycoplasma*. *Research in Microbiology* 168, 882–891.
- Ghabrial, S.A., 2008. Totiviruses. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology* (Third Edition), New York: Academic Press, pp. 163–174.
- Ghabrial, S.A., Castón, J.R., Jiang, D., Nibert, M.L., Suzuki, N., 2015. 50-plus years of fungal viruses. *Virology* 479–480, 356–368. doi:10.1016/j.virol.2015.02.034.
- Ghabrial, S.A., Dunn, S.E., Li, H., Xie, J., Baker, T.S., 2013. Viruses of *Helminthosporium* (*Cochliobolus*) *victoriae*. *Advances in Virus Research* 86, 289–325.
- Goodman, R.P., Ghabrial, S.A., Fichorova, R.N., Nibert, M.L., 2011. *Trichomonasvirus*: A new genus of protozoan viruses in the family Totiviridae. *Archives of Virology* 156 (1), 171–179.
- Jamal, A., Sato, Y., Shahi, S., *et al.*, 2019. Novel victorivirus from a Pakistani isolate of *Alternaria alternata* lacking a typical translational stop/restart sequence signature. *Viruses* 11, 577.
- Li, F., Du, J., Wu, Z., *et al.*, 2019. Identification and genetic analysis of a totivirus isolated from the *Culex tritaeniorhynchus* in Northern China. *Archives of Microbiology* 202 (4), 807–813. doi:10.1007/s00230-019-01788-9.
- Nibert, M.L., 2007. '2A-like' and 'shifty heptamer' motifs in penaeid shrimp infectious myonecrosis virus, a nonsegmented double-stranded RNA virus. *Journal of General Virology* 88 (2007), 1315–1318.
- Okamoto, K., Miyazaki, N., Larsson, D.S.D., *et al.*, 2016. The infectious particle of insect-borne totivirus-like Omono River virus has raised ridges and lacks fibre complexes. *Scientific Reports* 6, 33170.
- Sasai, S., Tamura, K., Tojo, M., *et al.*, 2018. A novel non-segmented double-stranded RNA virus from an Arctic isolate of *Pythium polare*. *Virology* 522, 234–243.
- Xie, J., Havens, W.M., Lin, Y.-H., Suzuki, N., Ghabrial, S.A., 2016. The victorivirus *Helminthosporium victoriae* virus 190S is the primary cause of disease/hypovirulence in its natural host and a heterologous host. *Virus Research* 213, 238–245.
- Soldevila, A.I., Ghabrial, S.A., 2001. A novel alcohol oxidase/RNA-binding protein with affinity for mycovirus double-stranded RNA from the filamentous fungus *Helminthosporium* (*Cochliobolus*) *victoriae*. *Journal of Biological Chemistry* 276, 4652–4661.
- Wickner, R.B., Ghabrial, S.A., Nibert, M.L., Patterson, J.L., Wang, C.C., 2012. Family Totiviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier Academic Press, pp. 639–650. doi:10.1016/B978-0-12-384684-6.00052-5.
- Xin, C., Wu, B., Li, J., *et al.*, 2016. Complete genome sequence and evolution analysis of *Eimeria stiedae* RNA virus 1, a novel member of the family Totiviridae. *Archives of Virology* 161, 3571–3576.

Yado-kari Virus 1 and Yado-nushi Virus 1 (Unassigned)

Subha Das, Okayama University, Kurashiki, Japan

Nobuhiro Suzuki, Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

© 2021 Elsevier Ltd. All rights reserved.

Glossary

2A-like motif A conserved self-processing peptide motif that cleaves a growing polypeptide chain co-translationally at a particular cleavage site. These motifs were first identified in members of *Picornaviridae*, separating structural proteins from replication-associated proteins. The motif is now known to be present in diverse RNA viruses.

Capsid protein The protective shell of viruses usually composed of multiple copies of same or different protein subunits for encasing viral genomes.

Mutualism Interspecific cooperation where two organisms belonging to two different species live together and benefit each other.

Trans-encapsidation or hetero-encapsidation Packaging of nucleic acid entities such as satellites by a capsid protein encoded by unrelated independent viruses.

Yado-kari In Japanese, it means borrowing a room in which to live or hermit crab biologically.

Yado-nushi In Japanese, it means someone who is the owner of the room.

Introduction

Typically, mycoviruses have either double-stranded (ds) or positive-sense single-stranded ((+)ss) RNA genomes. However, a few studies have also indicated the presence of mycoviruses with negative-sense (−) ssRNA genomes and geminivirus-like circular ssDNA genomes. Mixed mycovirus infections are common among fungi with occasional documented interplays between two unrelated viruses. Such virus-virus interactions can be categorized into three types: apparently neutral, synergistic and antagonistic. Synergistic interactions enhance the titer of one virus by another co-infecting virus, subsequently resulting in a different symptom expression than that induced in the host singly infected by either of the viruses. One such example is an enhancement of accumulation and vertical transmission of mycoreovirus 1 (MyRV1) in chestnut blight-causing fungus *Cryphonectria parasitica* by a co-infecting *Cryphonectria hypovirus* 1 (CHV1). In addition, in *Rosellinia necatrix*, *Rosellinia necatrix* megabarnavirus 2 (RnMBV2) reduces host virulence and growth only in the presence of *Rosellinia necatrix* partitivirus 1 (RnPV1). In return, RnMBV2 enhances accumulation of RnPV1 by two-fold. Individually, neither of the viruses causes any alterations to host virulence. In case of antagonistic interaction, one virus impairs the replication or accumulation of another co-infecting virus. For example, in *C. parasitica*, impairment of replication and lateral transmission of *Rosellinia necatrix* victorivirus 1 by either a co-infecting MyRV1 or a mutated CHV1 (lacking the RNA silencing suppressor). Replication impairment by MyRV1 is mediated by enhanced antiviral RNA silencing via transcriptional induction of its key genes namely dicer-like 2 (*DCL2*) and argonaute-like 2 (*AGL2*).

Recently, a unique mutualistic interplay between two RNA mycoviruses, Yado-nushi virus 1 (YnV1) and Yado-kari virus 1 (YkV1) has been described from a debilitated field isolate of *R. necatrix* strain W1032 from Nagano Prefecture, Japan. YnV1, an independent genuine double-stranded RNA (dsRNA) virus, trans-encapsidates the replicative-form dsRNA of the capsidless YkV1 (+)ssRNA (genomic RNA) along with its RNA-dependent RNA polymerase (RdRp), and is thereby responsible for YkV1 replication inside the host. In this way, despite being a (+)ssRNA virus, YkV1 likely replicates inside the hetero-capsid (particles) comprised of YnV1 capsid protein like a dsRNA virus. In return, YkV1 trans-enhances the accumulation of YnV1 by two- to four-fold, making the relationship rather mutualistic than commensal. This unique mutualistic interaction is different than the aforementioned virus-virus synergistic interactions, where two bona fide viruses (with the ability to infect and replicate inside host independently) interact and complement each other. This YkV1/YnV1 interaction is also distinct from those interactions between helper viruses and subviral elements (like satellite or defective viruses). Unlike satellite or defective viruses, YkV1's genome is replicated by its own RdRp.

In this article, we will be discussing genome characteristics of the two viruses and their unique mutualistic interplay in *R. necatrix*. In addition, we give an overview of other viruses or virus-like entities that may have similar YkV1/YnV1-like partnerships.

Virion Morphology

Particles formed by YnV1 are spherical in shape and approx. 40 nm in diameter (Fig. 1). The YkV1 virus genome and RdRp are also encased by the same capsid.

Genome Characteristics of YnV1

At least three different strains of YnV1 (designated as YnV1-A, B and C) have been reported so far from *R. necatrix*, while a strain from *Sclerotium rolfsii* with sequence similarity to YnV1 has been partially characterized. All three YnV1 genomes are composed of a single

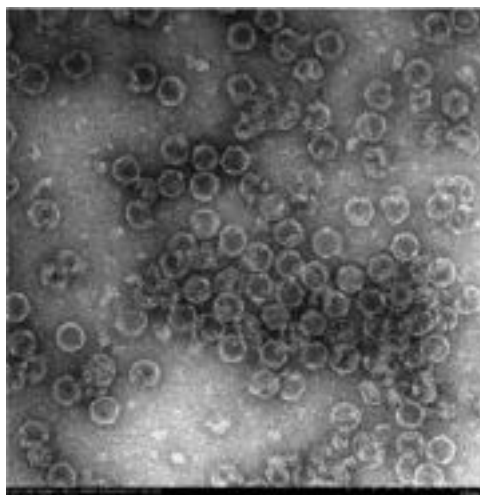


Fig. 1 Electron micrograph of negatively stained virus particles. The scale bar represents 20 nm. Courtesy of Dr. Hideki Kondo at Institute of Plant Science and Resources, Okayama University.

dsRNA element. The lengths of YnV1-A (GenBank Accession No. LC061478), B (LC006254) and C (LC006256) are 8971 bp, 8958 bp, and 8952 bp, respectively. The genome of YnV1-A contains two open reading frames (ORFs), ORF1 (4965 bp) and ORF2 (3339 bp). YnV1-A harbors the 5' and 3' untranslated regions (UTRs) of lengths 150 bp and 392 bp, respectively. YnV1-A ORF1 encodes a capsid protein with the presence of a zinc finger-like motif at the N-terminal side (Fig. 2). The ORF2 encodes a putative RdRp that contains a Phytoreo_S7 (pfam07236) domain upstream of the RdRp motifs (Fig. 2). All the domains identified on YnV1-A are also conserved in YnV1-B and C.

The Phytoreo_S7 domain has been conserved across different virus families including *Reoviridae*, *Chrysoviridae* and *Endornaviridae*. It is assumed that this domain has been transferred horizontally between fungal and plant viruses over evolutionary time. In several non-phytoreoviruses, the domain is located downstream of RdRp, while like members of *Endornaviridae* and *Chrysoviridae*, the YnV1 Phytoreo_S7 domain is located upstream of the RdRp domain. Regarding the function of this domain, a previous study demonstrated that the minor inner core protein of rice dwarf virus (a phytoreovirus) with Phytoreo_S7 domain possessed dsRNA binding capability. Therefore, this domain could be associated with viral RNA synthesis occurring at the inner core of particles. The exact role of the Phytoreo_S7 domain in YnV1 is yet to be determined.

Pairwise comparison showed 8%–10% sequence divergence at the amino acid level in the CP regions among all three strains of YnV1, while the sequence divergence was only 6%–7% between their corresponding RdRp regions. It is predicted that YnV1 CP is processed either by self-cleavage activity of YnV1 itself, or by host-derived proteases.

Genome Characteristics of YkV1

The ssRNA genome of YkV1 is 6310 bp in length, presumably existing in the host as dsRNA replicative form. The genome contains a single large ORF (4293 bp), encoding a putative polyprotein with the RdRp domain, situated close to the center of N-terminal portion (Fig. 2). The lengths of its 5' and 3' UTR regions are 796 bp and 1221 bp, respectively. In addition to the RdRp domain, YkV1 also contains a picornavirus 2A-like domain (Fig. 2). This domain was first reported in the foot-and-mouth disease virus (FMDV) and later in several other members of the family *Picornaviridae* and diverse viruses like animal dsRNA toti-like viruses and reoviruses.

The picornavirus 2A is a self-processing or ribosome skipping peptide that separates structural proteins from replication-associated proteins by cleaving a growing polyprotein at the C-terminal side of a conserved motif, DxE_xNPGP (where x is any amino acid). The cleavage or ribosomal skip occurs co-translationally, which is totally distinct from that of bona fide protease-associated cleavages of viral polyproteins. The YkV1 polyprotein is also predicted to undergo self-processing or ribosome skipping mediated by the 2A-like motif.

Recently, a next-generation sequencing approach combined with Sanger sequencing identified a single Spanish *R. necatrix* isolate (Rn454) to harbor three additional yado-kari viruses designated as YkV2, YkV3 and YkV4. While YkV2 contains a single large ORF like its YkV1 counterpart, the genome of YkV4 consists of two ORFs (Fig. 2). Interestingly, YkV3 has two genome variants, one with a single-ORF organization and the other one with a double-ORF arrangement just like YkV4. Notably, another Spanish isolate, Rn95-16, of *R. necatrix* harbored YkV4. YkV2 to YkV4 share greater sequence similarity (approximately 60% identity in the RdRp region) with one another than with YkV1 (approximately 30% identity). A feature distinguishing YkV1 from YkV2 to YkV4 is the presence of a poly(A) tail at the 3' end of the genome. Notably, the 2A-like motifs of YkV1 and YkV2 are situated almost at the same position where ORF1 of YkV3 (two-ORF variant, not shown in Fig. 2) and YkV4 ends (Fig. 2). This

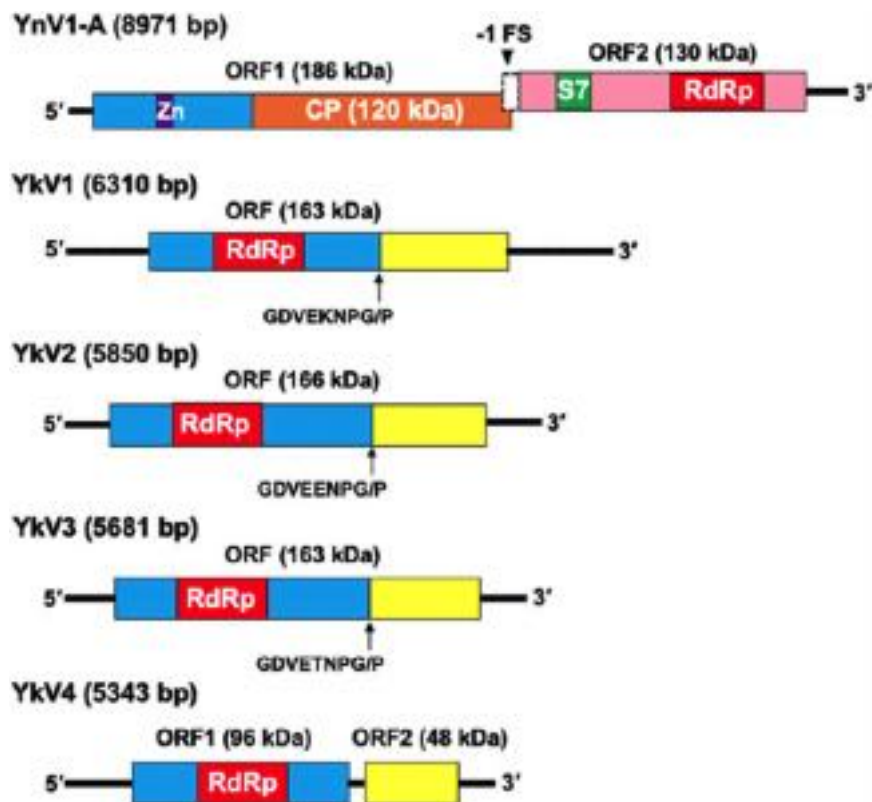


Fig. 2 Schematic representation of the genomic organization of yado-nushi virus 1 (YnV1) and all four yado-kari viruses (YkV1, YkV2, YkV3 and YkV4). The RdRp domains are highlighted in red for all the viruses. The zinc-finger and Phytoreo_S7 motifs on YnV1 are highlighted in purple and green, respectively. In the YnV1 genome, -1 FS indicates a possible -1 frameshift. The arrows indicate the cleavage positions within the conserved 2A-like domains (between amino acids G and P) of YkV1, YkV2 and YkV3. Molecular weight (in kDa) of each putative polypeptide has been mentioned on top of their corresponding open reading frames (ORF).

suggests the possibility that yado-kari viruses may encode at least two distinct polypeptides, *i.e.*, the RdRp for genome replication and transcription and another putative protein with unknown function.

Phylogenetic Placements of YnV1 and YkV1

The ORF1-encoded CP of YnV1 has no significant sequence similarity with any of the viral proteins deposited so far to GenBank. However, the ORF2-encoded putative RdRp showed modest levels of sequence identity to corresponding regions of dsRNA viruses such as *Giardia lamblia* virus (GLV) and *Glomus* sp. RF1 medium virus (GRF1V-M). Phylogenetically, YnV1 is also distantly related to relatively common dsRNA mycoviruses, *i.e.*, members of the families *Toti*-, *Quadri*-, *Megabirna*-, *Chrysoviriidae*, and the proposed family *Fusagraviridae* such as *Sclerotinia sclerotiorum* dsRNA mycovirus-L (also known as *Sclerotinia sclerotiorum* non-segmented virus L or SsNsV-L) (Fig. 3(A)).

Phylogenetic analysis with putative RdRp region of YkV1 showed that it is most closely related to the viruses like *Penicillium aurantiogriseum* foetidus-like virus 1 (PaFIV1), *Aspergillus foetidus* slow virus S2 (AfV-S2) and *Fusarium poae* mycovirus 2 (FpMyV2) (Fig. 3(B)). Interestingly, phylogenetically, YkV2, YkV3, and YkV4 are more closely related to FpMyV2 than to their Japanese counterpart YkV1. Indeed, based on RdRp-based phylogenetic analysis, it can be assumed that yado-kari viruses are distantly related to animal-infecting (+)ssRNA viruses belonging to the *Caliciviridae* family. It is noteworthy that both the families, *Totiviridae* and *Caliciviridae* are members of the extended picorna-like superfamily and are assumed to share the same ancestor.

Proposed Replication Model for YnV1 and YkV1

It is anticipated that, like other typical dsRNA viruses, YnV1 replicates and transcribes its genome using its own RdRp. YkV1 hijacks capsids from YnV1 and utilizes them as its replication site (Fig. 4). As the RdRp of YkV1 is also encased inside the YnV1-encoded capsid, both replication (synthesis of negative-strand RNA) and transcription (synthesis of positive-strand RNA) are presumed to be carried out by its own RdRp. This way, despite being a ssRNA virus, YkV1 likely exploits a replication strategy similar to a dsRNA

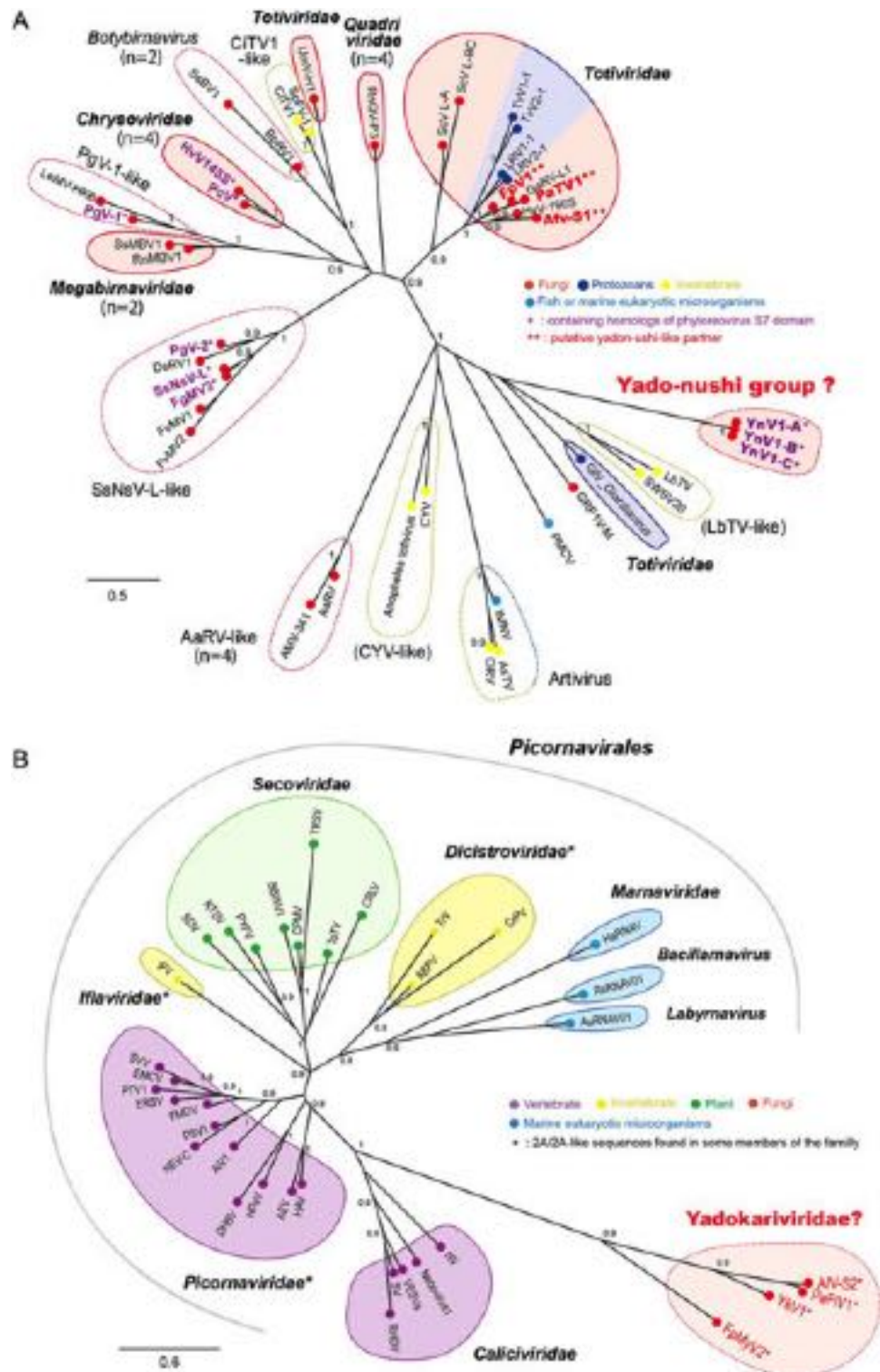


Fig. 3 Phylogenetic trees constructed with amino acid sequences of RNA-dependent RNA polymerase (RdRp) regions from (A) yado-nushi virus 1 (YnV1) and (B) yado-kari virus 1 (YkV1), and their related viruses. The figure has been reproduced with permission from Hisano, S., Zhang, R., Faruk, M.I., Kondo, H., Suzuki, N., 2018. A neo-virus-lifestyle exhibited by a (+)ssRNA virus hosted in an unrelated dsRNA virus: taxonomic and evolutionary considerations. *Virus Research* 244, 75–83.

virus inside its fungal host. This model clearly distinguishes YkV1 from other satellite RNAs or satellite RNA viruses that may or may not encode their own CP, but surely do not encode their own RdRp, therefore depending upon RdRp of their corresponding helper viruses.

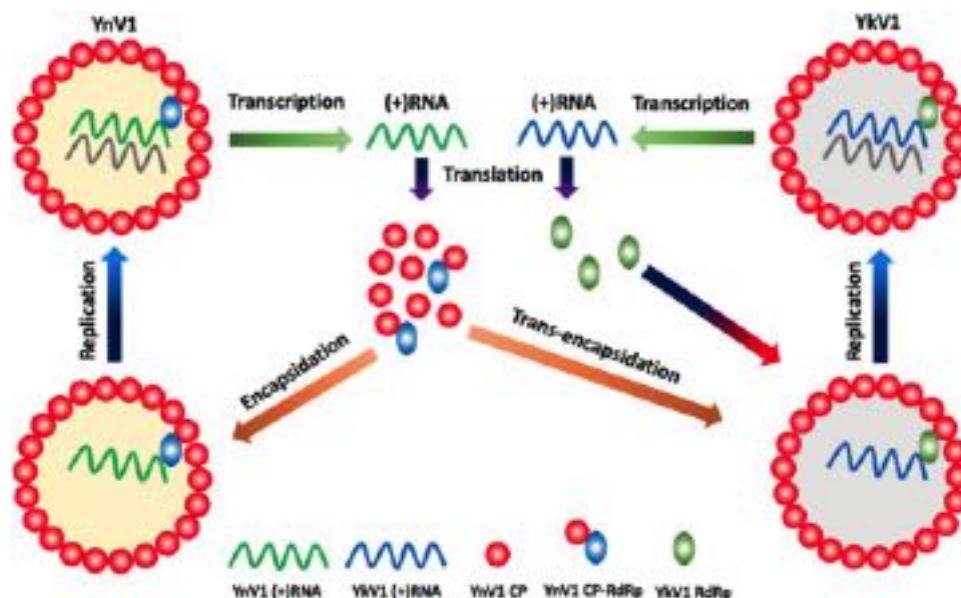


Fig. 4 Proposed model for mutualistic interactions between yado-nushi virus 1 (YnV1) and yado-kari virus 1 (YkV1).

Molecular Entities Sharing Similar YkV1/YnV1-Like Interactions

The novel interplay between YkV1 and YnV1 can be misinterpreted with similar satellite and their helper/dependent virus interactions. Satellite or defective RNAs and viruses are associated with helper viruses. For genome replication, they must depend upon their corresponding helper viruses. By contrast, YkV1 encodes its own RdRp and therefore, for genome replication, it does not depend upon YnV1, although YkV1 diverts the capsid of YnV1 as its replication site.

It seems that YkV1/YnV1-like mutualistic virus-virus interactions between unrelated viruses may not be so uncommon in nature. Umbraviruses with a (+)ssRNA genome do not encode CP, therefore forming no particles. Unlike YkV1, umbraviruses are speculated to replicate in rearranged host membranes like other (+)ssRNA viruses, infecting their hosts systematically and independently of their helper viruses. For encapsidation and aphid transmission, they solely depend on (+)ssRNA assistant viruses, *i.e.*, luteoviruses, which are distinct from umbraviruses. In the absence of luteoviruses, the umbravirus genomes form filamentous ribonucleoprotein structures and their ORF3 and ORF4 products are predicted to help them in long-distance (to cause systemic infection) and cell to cell movement, respectively. Apparently, YkV1 does not have this type of survival mechanism in the absence of YnV1.

Papaya sticky disease or “meleira” is a common disease of papaya in Brazil and Mexico. The causal agent of the disease is a complex mixture of a dsRNA virus, a tentative member of the proposed family Fusagraviridae, named papaya meleira virus (PMeV) and a (+)ssRNA virus termed papaya meleira virus 2 (PMeV2). The genome of PMeV is approximately 10 kbp in size, contains two ORFs, encoding a putative CP and an RdRp. This virus can only produce typical disease symptoms when being accompanied by another 4.5 kb ssRNA virus, PMeV2. The genome of PMeV2 contains two ORFs and is closely related to umbraviruses, while no homolog of the umbravirus movement protein is detected in PMeV2. Like YkV1, the genome of PMeV2 is trans-encapsidated by the CP of PMeV. Unlike the aforementioned umbravirus-luteovirus interplay, PMeV2, a ssRNA virus, behaves like YkV1 and acquires CP from the dsRNA fusagravirus, PMeV, which acts like YnV1 in this case. PMeV2 helps PMeV to establish severe disease symptoms to papaya plants. Despite the fact that PMeV does not encode any movement proteins, it can cause systemic infection in papaya. It is assumed that the virus uses laticiferous vessels for systemic movement in the plant. It would be interesting to examine the role of the PMeV2 ORF1-encoded hypothetical protein in cell to cell or long-distance virus movement.

Another example of comparable virus-virus interplay is the interaction between rice tungro bacilliform virus (RTBV) and rice tungro spherical virus (RTSV). Rice tungro is a devastating disease of rice especially in the south and south east Asia causing significant impairment in rice production each year. The disease is caused by two morphologically different viruses: RTBV, a reverse-transcribing dsDNA virus (pararetrovirus) that forms bacilliform capsids, and RTSV, a (+)ssRNA virus that forms isometric capsids. In the case of mixed infection, both the viruses are transmitted in a semi-persistent manner by leafhoppers. However, on its own, RTBV is not transmitted by leafhoppers; for this, it depends on RTSV. While RTSV can infect plants individually, causing either asymptomatic or mild disease symptoms, severe disease manifestation occurs only when it is accompanied by RTBV.

A similar mutualistic relationship also exists between human hepatitis B virus (HBV, a dsDNA hepadnavirus) and a viroid-like circular (–)ssRNA virus satellite, hepatitis D virus (HDV, deltavirus). HDV is well-known for encoding a ribozyme and hepatitis delta antigen (HDAg). However, for the completion of its life cycle (virion assembly, transmission and cell to cell movement), HDV requires its helper virus HBV. Again, while YkV1 encodes its own RdRp for independent genome replication, HDV depends on host RNA polymerase II and its own ribozyme to replicate in a rolling-circle manner and transcribe its genome in a way similar to the

replication strategy of plant-infecting viroids belonging to the family *Avsunviroidae*. This feature is different from general satellite RNAs or satellite RNA viruses. Thus, HDV is regarded as satellite-like RNA by the International Committee of Taxonomy of Viruses.

YkV1/YnV1-Like Virus Combinations in Other Fungi

Phylogenetically, YkV1 is related to other (+)ssRNA viruses infecting different filamentous fungi than *R. necatrix* that include PaFIV1 (infecting *Penicillium aurantiogriseum*), AfV-S2 (infecting *Aspergillus foetidus*) and FpMV2 (infecting *Fusarium poae*). It is noteworthy that like YkV1, genomes of these viruses also have single-ORF architecture, encoding an RdRp with 2A-like motif at the C-terminal side of the polyprotein. Interestingly, like YkV1, PaFIV1, AfV-S2, and FpMV2 are also accompanied by either totivirus or toti-like viruses with the two-ORF genome organization. This suggests that there could be an intimate relationship between (+) ssRNA and dsRNA viruses similar to YkV1/YnV1 in other fungi as well. These combinations of viruses also share some common features. For example, the genome size of these YkV1-like (+) ssRNA viruses is always smaller than their presumed partner dsRNA virus partners. It is noteworthy that *R. necatrix* strains carrying YkV2 to YkV4 are also co-infected by multiple dsRNA viruses with the two-ORF, undivided genome architecture. However, none of these viruses, detected in the fungi harboring yado-kari viruses shared moderate levels of sequence similarity to YnV1.

Predicted Past and Future of YkV1

Phylogenetic analyses showed that YkV1 is distantly related to calici-like viruses. This suggests that it might have evolved from a calici-like virus after losing the CP. This possibility has been proposed for other capsidless ssRNA viruses as well. Most probably, after losing its CP gene, YkV1's progenitor survived inside the host by capturing CP from a co-infecting full-fledged toti-like dsRNA virus and used it as the site of replication. Another prediction is YkV1 might have evolved from a satellite RNA that somehow acquired the RdRp gene from a calici-like virus and then started hijacking CP from a co-infecting toti-like dsRNA virus. The theory of encountering a toti-like host virus is not implausible as mixed viral infection is very common in fungi, including *R. necatrix*.

In future, to become a full-fledged dsRNA virus, YkV1 needs to acquire either a CP gene from a compatible dsRNA virus (most probably a toti-like virus) or it needs to find a way to utilize host-derived membranous vesicles as its replication site, like hypoviruses.

Concluding Remarks and Future Directions

The neo-viral lifestyle of two unrelated viruses (YkV1 and YnV1) was first described in 2016. Since then, potential YkV1/YnV1-like virus combinations have been detected in other fungal hosts. It would be worth examining whether these viruses also share similar neo-viral lifestyle to the one exhibited by YkV1 and YnV1. Indeed, many aspects of this unique lifestyle are still not fully understood. Some open questions remain regarding this mutualistic interplay. (1) Does the 2A-like motif cleave the yado-kari polyprotein cotranslationally by ribosome slippage mechanism? (2) How does YnV1 efficiently trans-encapsidate the genome and RdRp of YkV1? (3) Are the YkV1 and YnV1 genomes encapsidated together or separately? (4) Where does the assembly origin of YkV1 reside in its genome? (5) What is the molecular mechanism responsible for the enhancement of YnV1 replication by YkV1? The influence of YkV1 and YnV1 on their host biology is also not fully understood. Hopefully, future research will find answers to these questions.

Further Reading

- Arjona-Lopez, J.M., Telengech, P., Jamal, A., *et al.*, 2018. Novel, diverse RNA viruses from Mediterranean isolates of the phytopathogenic fungus, *Rosellinia necatrix*: Insights into evolutionary biology of fungal viruses. *Environmental Microbiology* 20, 1464–1483.
- Hibino, H., 1996. Biology and epidemiology of rice viruses. *Annual Review of Phytopathology* 34, 249–274.
- Kozlakidis, Z., Herrero, N., Ozkan, S., Bhatti, M.F., Coutts, R.H., 2013. A novel dsRNA element isolated from the *Aspergillus foetidus* mycovirus complex. *Archives of Virology* 158, 2625–2628.
- Littlejohn, M., Locarnini, S., Yuen, L., 2016. Origins and evolution of hepatitis B virus and hepatitis D virus. *Cold Spring Harbor Perspectives in Medicine* 6 (1), a021360.
- Osaki, H., Sasaki, A., Nomiyama, K., Tomioka, K., 2016. Multiple virus infection in a single strain of *Fusarium poae* shown by deep sequencing. *Virus Genes* 52, 835–847.
- Ryabov, E.V., Taliany, M.E., Robinson, D.J., *et al.*, 2012. Umbravirus. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier Academic Press, pp. 1191–1195.
- Sá Antunes, T.F., Amaral, R.J., Ventura, J.A., *et al.*, 2016. The dsRNA virus papaya meleira virus and an ssRNA virus are associated with papaya sticky disease. *PLoS One* 11 (5), e0155240. doi:10.1371/journal.pone.0155240.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9. doi:10.1128/mBio.02329-18.
- Yaegashi, H., Nakamura, H., Sawahata, T., *et al.*, 2013. Appearance of mycovirus-like double-stranded RNAs in the white root rot fungus, *Rosellinia necatrix*, in an apple orchard. *FEMS Microbiology Ecology* 83, 49–62.
- Yang, X., Cheng, A., Wang, M., *et al.*, 2017. Structures and corresponding functions of five types of picornaviral 2A proteins. *Frontiers in Microbiology* 8, 1373. doi:10.3389/fmicb.2017.01373.
- Zhang, R., Hisano, S., Tani, A., *et al.*, 2016. A capsidless ssRNA virus hosted by an unrelated dsRNA virus. *Nature Microbiology* 1, 15001.

Yeast L-A Virus (*Totiviridae*)

Reed B Wickner, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, United States

Tsutomu Fujimura and Rosa Esteban, Institute of Biology and Functional Genomics, CSIC/University of Salamanca, Salamanca, Spain

Published by Elsevier Ltd.

Glossary

Decapping Removal of the methylated GMP in 5'-5' linkage at the 5' end of most eukaryotic mRNAs.

Ribosomal frameshifting Ribosomes changing reading

frame on an mRNA, in response to a special mRNA structure, to synthesize a protein from two overlapping reading frames.

Introduction

The L-A virus of bakers/brewers yeast *Saccharomyces cerevisiae* is the exemplar strain (ScV-L-A) of the species *Saccharomyces cerevisiae virus L-A* (genus *Totivirus*, family *Totiviridae*). L-A is one of several RNA viruses infecting this organism, each of which spreads via the cell–cell fusion of mating, rather than by the extracellular route. The totivirus L-A, like members of the very similar L-BC family of viruses (ScV-L-BC), is a single-segment 4.6 kbp double-stranded RNA (dsRNA) virus encapsidated in icosahedral particles with a single major coat protein called Gag. The 20S (*Saccharomyces* 20S RNA narnavirus, ScNV20S) and 23S RNA replicons (*Saccharomyces* 23S RNA narnavirus, ScNV23S) are naked cytoplasmic single-stranded RNA (ssRNA) replicons except for their bound RNA-dependent RNA polymerases. The L-A virus serves as the helper virus for any of several smaller satellite dsRNAs, called M dsRNAs, each encoding a secreted protein toxin and immunity to that toxin, producing the 'killer' phenomenon. Killer strains can eliminate some of the competition by this means, although less than 5% of wild strains harbor a killer dsRNA, suggesting there are costs to carrying this replicon. The killer phenotype was used to study the genetics of M dsRNAs and the helper L-A. Several functional variants of L-A were defined based on their interactions with different M dsRNAs and with the host, and host mutants affected in virus expression or propagation were also examined.

History

In 1963, Makower and Bevan reported that some yeast strains secrete a toxin that kills other yeasts. This led to the identification by Bevan and by Fink of cellular dsRNAs and later viral particles correlated with the killer phenomenon. Studies of the chromosomal genes affecting the killer system revealed that the Kex2 protease, identified by its requirement for toxin secretion, was the long-sought proinsulin-processing enzyme. The Mak3 N-acetyltransferase, whose acetylation of Gag is needed for viral assembly, established consensus sequences for such enzymes. The loss of the L-A virus in mak3 mutants revealed a second dsRNA species of the same size, called L-BC, and unrelated to the killer system.

Virion Structure

L-A virions have icosahedral symmetry, but contain 120 Gag monomers per particle, in apparent violation of the rules of quasi-equivalence. In fact, each virion is composed of 60 asymmetric dimers of Gag (Fig. 1), a feature that is common to the cores of all dsRNA viruses that have been characterized. One type of Gag molecule makes contact with the icosahedral fivefold and twofold axes, but not with the threefold axes. The second type of Gag finds itself in contact with the threefold axis, but not with either the five- or twofold axes (Fig. 1). The two environments of Gag lead to two distinct conformations, suggesting that Gag may be more flexible than some other coat proteins. The L-A virion has more volume per nucleotide than do ssRNA or dsDNA viruses. These facts may reflect the requirement that the dsRNA moves inside the particle as it is transcribed by the RNA-dependent RNA polymerase that is fixed to the inner virion wall (see below). Pores at the fivefold axes are assumed to allow entry of nucleotides and exit of (+) strand transcripts, but retention of the dsRNA genome. A trench on the outside contains His154, the central active site residue of the mRNA-decapping activity to which the 7-methyl-GMP structure becomes covalently attached (Fig. 2), and from which it can then be transferred to new viral (+) strands. A layered density observed for the viral dsRNA may reflect the dsRNA's rigid structure and the fact that it is forced to press against the inner capsid wall.

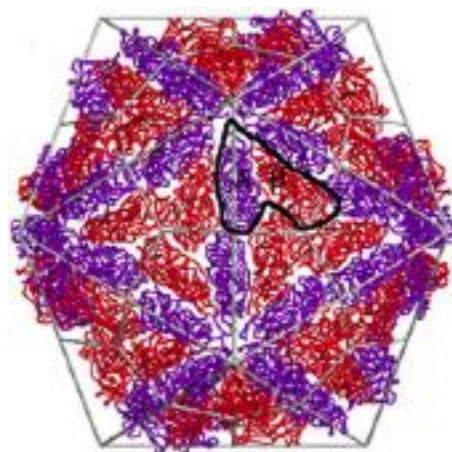


Fig. 1 Wire diagram of the L-A virus capsid. The major coat protein, Gag, is found in two nonequivalent positions: 'A' molecules contact the fivefold and twofold axes, while 'B' molecules contact the threefold axes. Reproduced from Naitow, I., Tang, J., Canady, M., Wickner, R.B., Johnson, J.E., 2002. L-A virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nature Structural Biology* 9, 725–728, with permission from Nature Publishing Group.

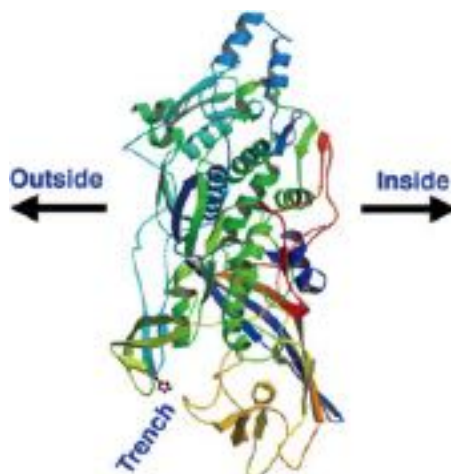


Fig. 2 Ribbon diagram of a single Gag molecule. The trench on the outer surface includes His154, the Gag residue to which 7meGMP is covalently attached by the decapping activity.

Genome Organization

The single segment of the L-A genome is a 4.6 kbp dsRNA with two long open reading frames (ORFs) (**Fig. 3**). The 5' ORF encodes the major coat protein, called Gag in analogy with retroviruses, while the longer 3' ORF, called Pol, encodes the RNA-dependent RNA polymerase, and has homology with similar enzymes of other ssRNA and dsRNA viruses. Pol is expressed only as a fusion protein whose amino end is a nearly complete Gag molecule. The fusion is carried out by a -1 ribosomal frameshift event in the region of overlap of the two ORFs. An RNA pseudoknot, in the region of overlap of the ORFs, slows ribosome progression at a point on the mRNA (of the form X XXY YYZ (0 frame indicated)) where bonding of the A-site and P-site tRNAs to the mRNA is nearly as good in the -1 frame as in the 0 frame. About 1% of ribosomes slip into the -1 frame, and, continuing translation, make the Gag–Pol fusion protein.

Replication Cycle

Transcription

Viral particles have an RNA-dependent RNA polymerase that is due to the Pol part of the Gag–Pol fusion protein. Transcription is a conservative reaction, meaning that the parental strands remain together after the synthesis of a new (+) strand. The L-A (+) strand transcripts are extruded from the particles into the cytoplasm where they serve as both mRNA and as the species that is

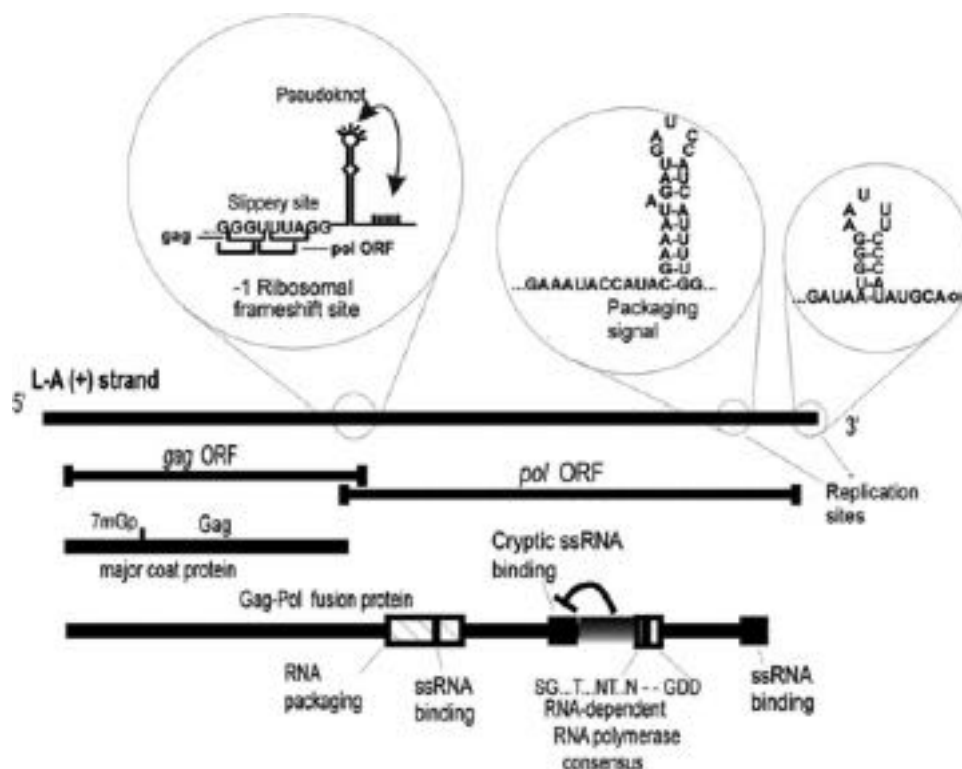


Fig. 3 Genome organization. The L-A (+) strand encodes Gag and Pol in overlapping reading frames. The mRNA lacks 3' polyA structures, but gets a 5' cap stolen from cellular mRNAs. The location of the packaging site on the RNA and the region of the Pol protein segment that recognizes the packaging site are indicated. The RNA sites for replication, and the parts of Pol with homology to other RNA-dependent RNA polymerases (RdRp) are indicated. There is a cryptic in vitro RNA binding site that is inhibited by the protein region to its immediately C-terminal.

packaged by coat proteins to make new viral particles. L-A (+) strand transcripts that are not degraded immediately have two possible fates: they may be translated or they may be encapsulated. The cap-stealing activity of Gag transfers the cap 7meGMP from cellular mRNAs to some L-A (+) strands directing them to the translation apparatus, while only uncapped L-A (+) strands are encapsidated. The (+) strand transcripts of the smaller M dsRNA or of deletion mutants of L-A are often retained within the particle where they may be replicated. This is called 'headful replication' because the volume of the particle seems to be the determinant of how many dsRNA molecules may accumulate in each particle.

RNA Packaging

The RNA packaging site (Fig. 3) is a stem-loop structure about 400 nucleotides from the 3' end of the L-A (+) strand. This stem-loop has an essential A residue protruding on the 5' side of the stem. The loop sequence is also important, but the stem sequence is not critical, as long as the stem structure can form. The packaging site on the RNA is recognized by the proximal part of the Pol domain of the Gag-Pol fusion protein. The Gag part of Gag-Pol is part of the capsid structure, so the Gag-Pol fusion protein structure assures packaging of viral (+) strands in new particles.

RNA Replication

The (-) strand synthesis reaction is called replication (Fig. 4). This reaction involves recognition of the internal binding site on the L-A (+) strands by Pol, followed by interaction with the now nearby 3' end and initiation of new RNA (-) chains. Multiple rounds of RNA synthesis can proceed within the viral particle in the 'headful replication' process discussed above.

Viral Translation

The translation apparatus is the battleground of RNA viruses with their hosts. Poliovirus protease cleaves eIF4G so that host capped mRNAs cannot be translated but its own internal ribosomal entry site-containing mRNAs are used. Influenza virus steals caps from host mRNAs, as does L-A. The L-A coat protein removes caps from cellular mRNAs and transfers them to newly synthesized L-A (+) strands.

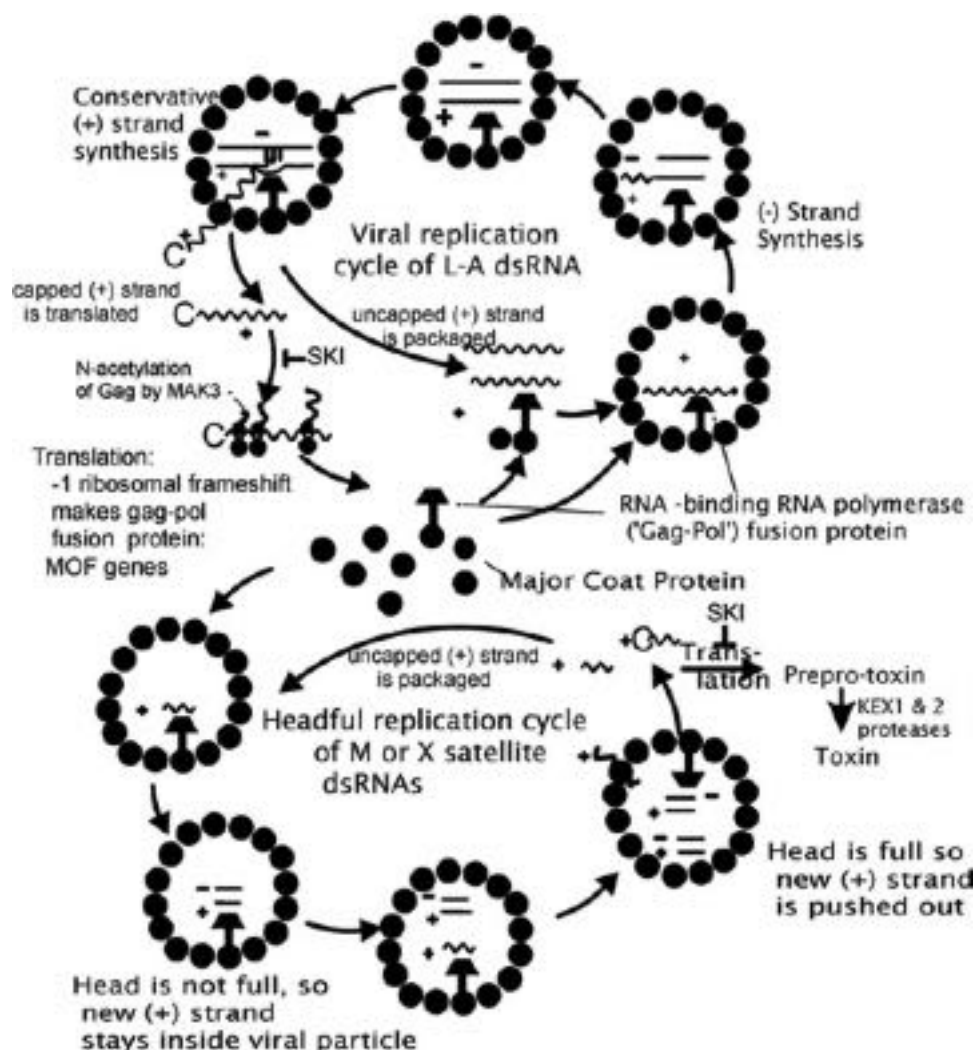


Fig. 4 Replication cycle of the L-A virus. Both (+) and (−) RNA strand synthesis occur within the viral particles, but at different stages of the cycle. RNA (+) strands (transcripts) are extruded from the particles and serve as both mRNA and the species packaged to make new particles. The (+) strands with a cap are translated while those without a cap are encapsulated or degraded. Translation of viral mRNA is blocked by Ski proteins. N-terminal acetylation of Gag by Mak3p is necessary for viral assembly.

L-A (+) strand transcripts lack the 3′ polyA structures typical of eukaryotic mRNAs, and so are at a distinct disadvantage for translation. The requirement for 3′ polyA structure for translation, is imposed by the ribosome-associated Ski (superkiller, after the phenotype of the mutants) proteins (except Ski1p/Xrn1p). These Ski proteins both degrade non-polyA mRNAs and block their translation. In the absence of these Ski proteins, non-polyA mRNAs are expressed with the same kinetics as polyA + messages. The SKI1/XRN1 gene encodes a 5′ → 3′ exoribonuclease specific for uncapped mRNA. As a defense against its degradation of viral mRNA, L-A's Gag protein has a decapping activity that removes caps from cellular mRNAs, transfers them to viral (+) strands protecting them from Ski1p, and leaves some cellular mRNAs as 'decapitated decoys', uncapped RNAs that may serve as alternative targets for the exonuclease action. In the absence of this decapping activity, viral mRNAs are not expressed, but deletion of the SKI1/XRN1 gene restores viral mRNA expression.

Together with the Ski proteins, a mitochondrial nuclease, Nuc1p, is released during meiosis/sporulation and protects cells from toxin overexpression during this process. Nuc1p is homologous to EndoG, an apoptosis nuclease of mammals, suggesting an evolutionary origin of apoptosis as an antiviral process.

L-A Genetics

Several natural variants of L-A have been described. The ability of an L-A variant to support M dsRNA replication was first called [HOK] (helper of killer) and then H when it was found to be a property of L-As such as L-A-H, L-A-HN, or L-A-HNB. Making

several chromosomal MAK genes dispensable for M propagation was named [B] (bypass) and is found on L-A-HNB. Ability to exclude L-A-H was named [EXL] (exclusion) and then just E on L-A-E. Insensitivity to the action of [EXL] was called [NEX] (nonexcludable), and then shortened to N as in L-A-HN or L-A-HNB. The molecular basis of these interactions has resisted study because of the current inability to obtain L-A dsRNA from a cDNA clone. M dsRNA can be supported from a cDNA clone of L-A, but the L-A virus has not been shown to be regenerated from these transcripts.

Other RNA Replicons in Yeast: L-BC, 20S RNA, 23S RNA

20S RNA was discovered as an RNA species which appeared when cells were placed under conditions that induce meiosis and spore formation, namely near-starvation for nitrogen and provision of acetate as a carbon source. It was shown that 20S RNA is an independent replicon, and can be made independent of the sporulation or meiosis processes. 20S RNA encodes its own RNA-dependent RNA polymerase which is bound to the otherwise naked cytoplasmic RNA. The mechanism of 20S RNA replication control by culture conditions remains to be elucidated. 23S RNA is a related, but independent yeast replicon which was found as a dsRNA form (called T). 23S RNA also encodes its RNA-dependent RNA polymerase. Both 20S RNA and 23S RNA can be launched from cDNA clones to form replicating virus. 20S RNA (ScNV20S) and 23S RNA (ScNV23S) are classified into the genus *Narnavirus* within the family *Narnaviridae*. L-BC is a totivirus (genus *Totivirus*, family *Totiviridae*), related to L-A but independent of it. Its dsRNA is essentially the same size as that of L-A although its copy number is usually about tenfold lower. Thus it was not detected until chromosomal mutants that lose L-A were examined. L-BC does not interact with the killer system, and confers no obvious phenotype, so its genetics has not been extensively explored.

See also: Totiviruses (*Totiviridae*). Viral Killer Toxins

Further Reading

- Bevan, E.A., Herring, A.J., Mitchell, D.J., 1973. Preliminary characterization of two species of dsRNA in yeast and their relationship to the 'killer' character. *Nature* 245, 81–86.
- Bostian, K.A., Elliott, Q., Bussey, H., *et al.*, 1984. Sequence of the preprotoxin dsRNA gene of type I killer yeast: Multiple processing events produce a two-component toxin. *Cell* 36, 741–751.
- Dinman, J.D., Icho, T., Wickner, R.B., 1991. A –1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proceedings of the National Academy of Sciences of the United States of America* 88, 174–178.
- Esteban, R., Fujimura, T., Wickner, R.B., 1989. Internal and terminal cis-acting sites are necessary for in vitro replication of the L-A double-stranded RNA virus of yeast. *The EMBO Journal* 8, 947–954.
- Fujimura, T., Esteban, R., 2016. Diphosphates at the 5' end of the positive strand of yeast L-A double-stranded RNA virus as a molecular self-identity tag. *Molecular Microbiology* 102, 71–80.
- Fujimura, T., Esteban, R., 2011. Cap-snatching mechanism in yeast L-A double-stranded RNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 108, 17667–17671.
- Fujimura, T., Esteban, R., Esteban, L.M., Wickner, R.B., 1990. Portable encapsidation signal of the L-A double-stranded RNA virus of *S. cerevisiae*. *Cell* 62, 819–828.
- Fujimura, T., Ribasm, J.C., Makhov, A.M., Wickner, R.B., 1992. Pol of gag-pol fusion protein required for encapsidation of viral RNA of yeast L-A virus. *Nature* 359, 746–749.
- Fuller, R.S., Brake, A., Thorner, J., 1989. Intracellular targeting and structural conservation of a prohormone-processing endoprotease. *Science* 246, 482–486.
- Gao, J., Chau, S., Chwdhury, F., *et al.*, 2019. Meiotic viral attenuation through an ancestral pathway. *Proceedings of the National Academy of Sciences of the United States of America* 116, 16454–16462.
- Icho, T., Wickner, R.B., 1989. The double-stranded RNA genome of yeast virus L-A encodes its own putative RNA polymerase by fusing two open reading frames. *Journal of Biological Chemistry* 264, 6716–6723.
- Leibowitz, M.J., Wickner, R.B., 1976. A chromosomal gene required for killer plasmid expression, mating, and spore maturation in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 73, 2061–2065.
- Masison, D.C., Blanc, A., Ribas, J.C., *et al.*, 1995. Decoying the cap-mRNA degradation system by a dsRNA virus and poly(A)-mRNA surveillance by a yeast antiviral system. *Molecular and Cellular Biology* 15, 2763–2771.
- Naitow, H., Canady, M.A., Wickner, R.B., Johnson, J.E., 2002. L-A dsRNA virus at 3.4 Å resolution reveals particle architecture and mRNA decapping mechanism. *Nature Structural & Molecular Biology* 9, 725–728.
- Sommer, S.S., Wickner, R.B., 1982. Yeast L dsRNA consists of at least three distinct RNAs; Evidence that the non-Mendelian genes [HOK], [NEX] and [EXL] are on one of these dsRNAs. *Cell* 31, 429–441.
- Tercero, J.C., Wickner, R.B., 1992. MAK3 encodes an N-acetyltransferase whose modification of the L-A gag N-terminus is necessary for virus particle assembly. *Journal of Biological Chemistry* 267, 20277–20281.
- Wickner, R.B., 2013. Viruses and prions of yeasts, fungi and unicellular eukaryotes. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology*, sixth ed. Philadelphia, PA: Wolters Kluwer/Lippincott Williams & Wilkins, pp. 2355–2383.

ALGAL VIRUSES

Algal Marnaviruses (*Marnaviridae*)

Marli Vlok and Curtis A Suttle, University of British Columbia, Vancouver, BC, Canada

Andrew S Lang, Memorial University of Newfoundland, St. John's, NL, Canada

© 2021 Elsevier Ltd. All rights reserved.

This is an update of A.S. Lang, C.A. Suttle, Marnaviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00567-7.

RNA Viruses Infecting Marine Protists

Most studies on marine RNA viruses have been on those infecting either charismatic megafauna or animals of economic importance. Over the last three decades this research has extended to include RNA virus isolates infecting protists as well as RNA metagenomic studies. Through this, the known marine RNA viruses now include representatives of all four Baltimore RNA virus groups, with hosts spread across the eukaryotic tree of life. Of these viruses, those infecting single-celled eukaryotes are thought to be the dominant portion of the free-floating RNA virus community.

The first RNA virus isolated and characterized that infects a marine protist was Heterosigma akashiwo RNA virus (HaRNAV). This virus was isolated from the Strait of Georgia in coastal British Columbia and it infects and causes the lysis of the marine, toxic bloom-forming, unicellular, photosynthetic alga *H. akashiwo*. HaRNAV particles are non-enveloped icosahedrons, with an average diameter of 25 nm (Table 1). The positive-sense monocistronic genome is 8.6 kb in size and has distinct domain organization characteristics that contributed to its classification as the founding member of the family *Marnaviridae* within the order *Picornavirales*. The polyprotein translation is under the control of a putative 5' internal ribosome entry site (IRES), with the non-structural proteins (helicase [S3H], 3C protease [3CPro], and RNA-dependent RNA polymerase [RdRp]) encoded at the N terminus and the structural proteins (VP2, -4, -3, and -1) at the C terminus (Fig. 1). HaRNAV particles become visible within infected cells 48 h post infection. Like other RNA viruses it has a relatively large burst size, at 460–520 particles per cell. One notable sign of viral-induced intracellular cytopathology is the appearance of membranous vesicles within the cytoplasm; such structures are the site of RNA replication for other positive-sense single-stranded (ss) RNA viruses and are well characterized for picornaviruses such as poliovirus. This is followed by the appearance of assembled viral particles throughout the cytoplasm, and these can be at sufficient density for the particles to form crystalline arrays.

Additional icosahedral (25–32 nm in diameter) ssRNA viruses (genomes ranging from 4.4 to 13.2 kb) infecting other marine protists have since been isolated and characterized. These include the dinoflagellate-infecting *Heterocapsa circularisquama* RNA virus (HcRNAV) of the family *Alverniviridae*, *Chaetoceros tenuissimus* RNA virus types I and II (CtenRNAV01 and CtenRNAVII), *Chaetoceros socialis* f. *radians* RNA virus (CsfrRNAV), *Chaetoceros* sp. (CspRNAV) and *Rhizosolenia setigera* (RsRNAV) RNA viruses that infect centric diatoms, the pennate diatom-infecting *Asterionellopsis glacialis* RNA virus (AglRNAV), and *Aurantiochytrium* RNA virus (AuRNAV), which infects a heterotrophic marine flagellate. There is also one known double-stranded (ds) RNA virus infecting a marine protist, MpRNAV-01B, which infects the photosynthetic *Micromonas pusilla* and represents the only species in the genus *Mimoreovirus*, subfamily *Sedoreovirinae*, family *Reoviridae*. Some studies have suggested other protist RNA viruses might exist in cultures. Two notable examples are the observation of RNA virus-like particles in *Ostreococcus* sp. and the determination of RdRp sequences for two viruses that infect the pennate diatom *Cylindrotheca closterium*. Recently, the potential host range of protist-infecting RNA viruses has also been extended to the holobiont of the multicellular protist *Delisea pulchra*. The detected viruses included + ssRNA viruses affiliated to the *Picornavirales* as well as some dsRNA viruses with sequence similarity to members of the *Totiviridae*. Metagenomic studies focused on planktonic marine RNA viruses have also revealed many novel viruses, including some complete viral genomic sequences. The majority of these show affiliation with the order *Picornavirales*.

Development of a Taxonomic Framework and Resulting Changes in Taxonomy

In addition to HaRNAV's classification as the founding member of the family *Marnaviridae*, four of the above-mentioned + ssRNA viruses had previously been formally taxonomically classified. These are AuRNAV, classified within the unassigned genus *Labymavirus*, and CtenRNAV01, CsfrRNAV and RsRNAV, which were classified within the unassigned genus *Bacillamavirus*. Until recently, the International Committee on the Taxonomy of Viruses (ICTV) taxonomic classification system required inclusion of biological information such as host range, replication cycle, virus particle structure, and serology. However, recognizing the need to update the taxonomic system, the ICTV recently announced they would accept the formal taxonomic classification of viruses known only from metagenomic studies. We have recently applied a sequence-based framework for analysis of twenty marine RNA viruses that provides the justification for taxonomic classification of these viruses within the family *Marnaviridae*. The 20 viruses include the original and sole representative of the family, HaRNAV, 7 additional viruses represented by isolates, and 12 viruses discovered using metagenomics. Phylogenetic analysis of the RdRp domain sequences placed the 20 marine RNA virus sequences in a strongly supported monophyletic group relative to other *Picornavirales* sequences (Fig. 2). The original genus within the *Marnaviridae*, *Marnavirus*, is basal within the clade and the analysis places the 20 viruses into seven clades that we defined as genera within the family *Marnaviridae*. These include the

Table 1 Isolates in the family *Marnaviridae* that infect marine protists

<i>Virus</i>		<i>Host</i>		<i>Genus</i>	<i>Particle diameter (nm)</i>	<i>Genome size (kb)</i>	<i>Latent period (hours)</i>	<i>Burst size (particles per cell)</i>
HaRNAV	Heterosigma akashiwo RNA virus	<i>Heterosigma akashiwo</i>	Raphidophyte	<i>Marnavirus</i>	25	8.6	48	460–520
RsRNAV	Rhizosolenia setigera RNA virus	<i>Rhizosolenia setigera</i>	Diatom	<i>Bacillarnavirus</i>	32	11.2	48	1010–3100
CtenRNAV01	Chaetoceros tenuissimus RNA virus 01	<i>Chaetoceros tenuissimus</i>	Diatom	<i>Bacillarnavirus</i>	31	9.4	<24	1.0×10^5
CtenRNAVII	Chaetoceros tenuissimus RNA virus type II	<i>Chaetoceros tenuissimus</i>	Diatom	<i>Sogarnavirus</i>	35	9.5	24–48	287
CspRNAV	Chaetoceros species RNA virus 02	<i>Chaetoceros</i> sp.	Diatom	<i>Sogarnavirus</i>	32	9.0–10.0	<48	ND ^a
CsfrRNAV	Chaetoceros socialis f. radians RNA virus 01	<i>Chaetoceros socialis f. radians</i>	Diatom	<i>Bacillarnavirus</i>	22	9.5	<48	66
AuRNAV	Aurantiochytrium single stranded RNA virus	<i>Aurantiochytrium</i>	Thraustochytrid	<i>Labyrnavirus</i>	25	9.0	8	5.8×10^3 – 6.4×10^4
AglaRNAV	Asterionellopsis glacialis RNA virus	<i>Asterionellopsis glacialis</i>	Diatom	<i>Kusarnavirus</i>	31	9.5	48	ND

^aND, not determined.

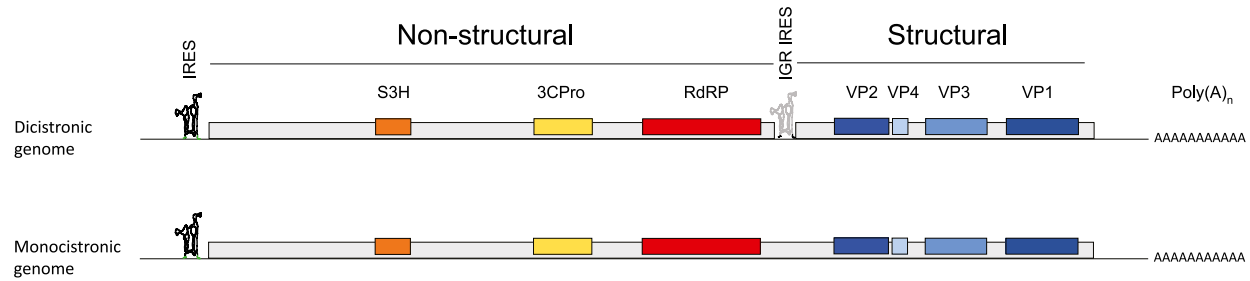


Fig. 1 Viral genomic organizations of members within the family *Marnaviridae*. The isolates and metagenomic-assembled viruses have either a mono- or dicistronic genome organization, which include a putative 5' internal ribosome entry site (IRES), S3 helicase, 3C protease/peptidase, RNA-dependent RNA polymerase (RdRp), putative intergenic region (IGR) IRES (in dicistronic genomes), four capsid domains (VP2, VP4, VP3, and VP1), and a poly(A) tail.

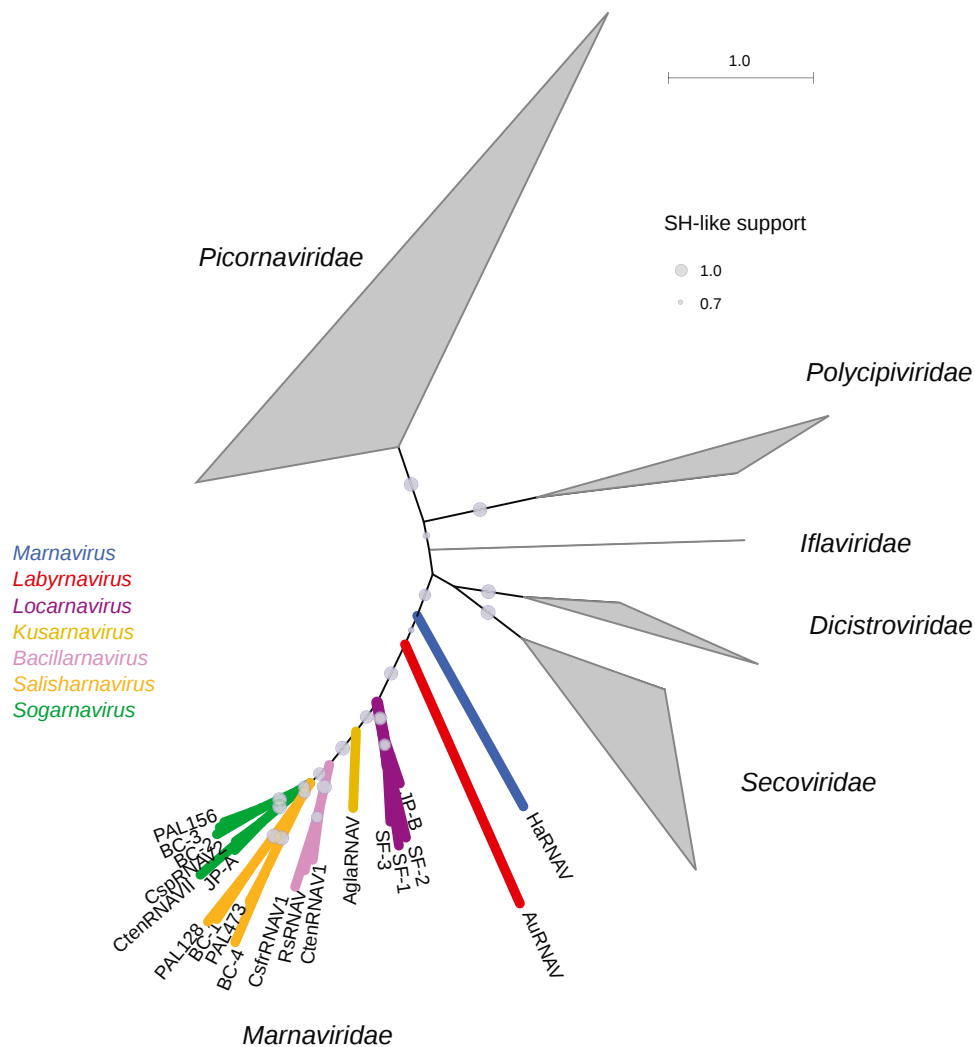


Fig. 2 Unrooted maximum-likelihood phylogeny of the RNA-dependent RNA polymerase amino acid sequences of members in the order *Picornvirales*. Sides of grey triangles represent longest and shortest branches with the base width indicating of the number of genera present in the collapsed clade. Members of the family *Marnaviridae* are coloured by genus: *Marnavirus* (blue), *Labynavirus* (red), *Locarnavirus* (purple), *Kusarnavirus* (yellow), *Bacillarnavirus* (pink), *Salisharnavirus* (orange), and *Sogarnavirus* (green). SH-like branch support values are indicated at nodes when >0.70 and the scale bar indicates substitutions per site.

Table 2 Species in the family Marnaviridae

Genus	Species	Virus name	Abbreviation	Accession number	Size (nt)	Source
Marnavirus	<i>Heterosigma akashiwo</i> RNA virus ^a	<i>Heterosigma akashiwo</i> RNA virus	HaRNAV	NC_005281	8587	<i>Heterosigma akashiwo</i>
Labyrnavirus	<i>Aurantiochytrium</i> single stranded RNA virus ^a	<i>Aurantiochytrium</i> single stranded RNA virus	AuRNAV	NC_007522	9035	<i>Aurantiochytrium</i> sp.
Locarnavirus	<i>Jericarnavirus B</i> ^a	Marine RNA virus JP – B	JP – B	NC_009758	8926	Coastal marine
	<i>Sanfarnavirus 2</i>	Marine RNA virus SF – 2	SF – 2	KF412901	9321	Coastal wastewater
	<i>Sanfarnavirus 1</i>	Marine RNA virus SF – 1	SF – 1	JN661160	8970	Coastal wastewater
	<i>Sanfarnavirus 3</i>	Marine RNA virus SF – 3	SF – 3	KF478836	8648	Coastal wastewater
Kusarnavirus	<i>Astarnavirus</i> ^a	<i>Asterionellopsis glacialis</i> RNA virus	AglaRNAV	NC_024489	8842	<i>Asterionellopsis glacialis</i>
Bacillarnavirus	<i>Chaetoceros tenuissimus</i> RNA virus 01	<i>Chaetoceros tenuissimus</i> RNA virus 01	CtenRNAV1	AB375474	9431	<i>Chaetoceros tenuissimus</i>
	<i>Rhizosolenia setigera</i> RNA virus ^a	<i>Rhizosolenia setigera</i> RNA virus	RsRNAV	AB243297	8877	<i>Rhizosolenia setigera</i>
	<i>Chaetoceros socialis</i> f. <i>radians</i> RNA virus 01	<i>Chaetoceros socialis</i> f. <i>radians</i> RNA virus 01	CsfrRNAV1	NC_012212	9467	<i>Chaetoceros socialis</i> f. <i>radians</i>
Salisharnavirus	<i>Britarnavirus 4</i> ^a	Marine RNA virus BC – 4	BC – 4	MH171300	8593	Coastal/oceanic marine
	<i>Palmarnavirus 473</i>	Marine RNA virus PAL473	PAL473	NC_029309	6360	Coastal marine
	<i>Britarnavirus 1</i>	Marine RNA virus BC – 1	BC – 1	MG584187	8638	Coastal marine
	<i>Palmarnavirus 128</i>	Marine RNA virus PAL128	PAL128	NC_029306	8660	Coastal marine
Sogarnavirus	<i>Britarnavirus 2</i>	Marine RNA virus BC – 2	BC – 2	MG584188	8843	Coastal marine
	<i>Palmarnavirus 156</i>	Marine RNA virus Pal156	PAL156	NC_029307	7897	Coastal marine
	<i>Britarnavirus 3</i>	Marine RNA virus BC – 3	BC – 3	MG584189	8496	Coastal marine
	<i>Jericarnavirus A</i>	Marine RNA virus JP – A	JP – A	NC_009757	9236	Coastal marine
	<i>Chaetenuissarnavirus II</i> ^a	<i>Chaetoceros tenuissimus</i> RNA virus type II	CtenRNAVII	NC_025889	9562	<i>Chaetoceros tenuissimus</i>
	<i>Chaetarnavirus 2</i>	<i>Chaetoceros</i> species RNA virus 02	CspRNAV2	AB639040	9417	<i>Chaetoceros</i> sp.

^aType species.

previously unassigned genera *Bacillarnavirus* and *Labyrnavirus*, and we chose the names *Kusarnavirus*, *Locarnavirus*, *Salisharnavirus* and *Sogarnavirus* for the four new genera. Capsid polyprotein sequence comparisons were used to delineate the species within genera (Table 2), with a cut-off of 75% pairwise amino acid identity.

Among the 20 viruses and 7 genera, there is a mixture of mono- and dicistronic genome organizations (Table 1; Fig. 1). While the majority of the 20 viruses have a dicistronic genome organization, HaRNAV and SF-3 have a single predicted polyprotein encoded in their genomes. HaRNAV is the only representative of the genus *Marnavirus*, but SF-3 is one of four viruses falling within the genus *Locarnavirus*. There is precedent for such different genomic organizations within the *Picornavirales*, with both mono- and bipartite genomes found within the *Secoviridae*.

Marine RNA Virus Quasispecies

RNA viruses exist in a mutation-selection balance where a genetically diverse population of variant genomes are ordered around the fittest genome from a single progenitor, known as the quasispecies concept, and this is important in the evolution of RNA viruses. Single-nucleotide variance analysis of six of the *Marnaviridae* members' sequences identified in RNA virus metagenomic data sets from around the globe suggest that these viruses also exist as quasispecies, with many low frequency mutations detectable across conserved domains. The distribution of viral genotypes across marine waterbodies is poorly understood. Patterns in the variant analysis showed that some of the viruses exhibited quasispecies-like populations with synonymous low-frequency mutations in locations distinct from the original site of discovery. Some mutations provide a competitive advantage to the virus, and they can become established in the viral population. Mutations occurring at >99% frequency were detected in these datasets across multiple viral domains and geographic locations, which indicates turnover of the dominant genotypes had occurred.

Importance of RNA Viruses in Marine Microbial Ecology

It has traditionally been believed that most of the estimated 10^7 marine virus particles per ml of seawater have DNA genomes and infect bacteria. Viral abundance is most frequently determined using epifluorescence or electron microscopy, or flow cytometry. These methods are good for detecting larger DNA viruses, but they lack the sensitivity and specificity to detect small RNA genomes

or, in the case of electron microscopy, differentiate RNA viruses from small DNA viruses. Use of fluorometry to determine the abundance of RNA- and DNA-containing virus particles suggested that at least 50% of the viroplankton in coastal waters have RNA genomes. Using quantitative RT-PCR, picorna-like RNA virus genome abundances were estimated to range from 0.2 to 5×10^7 per ml, thereby representing 14%–90% (mean approximately 40%) of the total coastal marine virus populations sampled. If this holds true for other marine environments, it could have huge implications for our understanding of viral production and the contribution of RNA viruses to biogeochemical processes. Considering that the majority of viruses identified in the study were picorna-like viruses, known to infect eukaryotes, we can assume that protists actually contribute more to marine viral dynamics than previously thought.

The hosts for the isolated *Marnaviridae* members (Table 1) include ecologically important protists and these viruses are thought to have significant effects on both the frequency and duration of plankton blooms. For example, HaRNAV infects the bloom-forming Raphidophyte *Heterosigma akashiwo*. Only some strains of *H. akashiwo* are susceptible to infection by HaRNAV, and the pattern of susceptibility does not appear to relate to geographic origin of the strain in question. When the virus is added to a growing culture of susceptible *H. akashiwo*, complete lysis of the culture generally occurs within 1 week or less. Diatoms are integral members of marine phytoplankton communities and represent the largest proportion of host organisms for viruses in the family *Marnaviridae*. Diatoms are distributed worldwide and are responsible for 35%–75% of the primary production in the oceans. Diatom blooms occur seasonally in nutrient-rich ecosystems, where a succession of species is observed. Like HaRNAV, the studied diatom viruses are highly host specific, frequently exhibiting strain specificity. Although burst sizes range widely, from 66 (CsfrRNAV) to 1.0×10^5 (CtenRNAV01) particles per cell, these viruses can decimate host populations and alter population dynamics, species succession and the functioning of the biological carbon pump.

The survival of host cells within a pool of viral pathogens suggests that there are regulatory factors affecting successful infection. For example, *C. tenuissimus* was found to be consistently detected at high abundances in both seawater and sediments off the coast of Japan where viruses infecting it are also known to occur in high abundance. While virus resistance could account for some of this, environmental niche partitioning was also shown to be an important contributor to viral infection, with temperature and salinity affecting viral infections of the same host organism.

Summary

Work in the ocean, with both culturing and metagenomics approaches, has shown that marine systems harbor a whole new world of ssRNA viruses. The majority of these have shown affiliation with the order *Picornavirales*, and many of these have now recently been formally classified within the family *Marnaviridae*. However, despite their shared characteristics and phylogenetic affiliations based on conserved domains, most of the genomes have no recognizable similarity to any other known sequences, viral or otherwise. More work is clearly needed with the individual marine RNA viruses, and even more viruses need to be brought into culture, sequenced and characterized. The ssRNA viruses isolated from marine systems infect basal eukaryotes, protists, and so these viruses may be ancestral to the related viruses infecting higher organisms such as mammals. There is no doubt that study of marine RNA viruses will further our understanding of RNA virus evolution.

Further Reading

- Allen, L.Z., Mccrow, J.P., Ininbergs, K., *et al.*, 2017. The Baltic Sea virome: Diversity and transcriptional activity of DNA and RNA viruses. *mSystems* 2, e00125–16.
- Armbrust, E.V., 2009. The life of diatoms in the world's oceans. *Nature* 459, 185–192.
- Barrett, T., Visser, I.K.G., Mamaev, L., *et al.*, 1993. Dolphin and porpoise morbilliviruses are genetically distinct from phocine distemper virus. *Virology* 193, 1010–1012.
- Brussaard, C.P.D., Noordeloos, A.A.M., Sandaa, R.-A.A., Heldal, M., Bratbak, G., 2004. Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. *Virology* 319, 280–291.
- Burki, F., 2014. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harbor Perspectives in Biology* 6, a016147.
- Comeau, A.M., Arbiol, C., Krusch, H.M., 2010. Gene network visualization and quantitative synteny analysis of more than 300 marine T4-like phage scaffolds from the GOS metagenome. *Molecular Biology and Evolution* 27, 1935–1944.
- Correa, A.M.S., Welsh, R.M., Vega Thurber, R.L., 2012. Unique nucleocytoplasmic dsDNA and +ssRNA viruses are associated with the dinoflagellate endosymbionts of corals. *The ISME Journal* 7, 13–27.
- Culley, A.I., 2018. New insights into the RNA aquatic virosphere via viromics. *Virus Research* 244, 84–89.
- Culley, A.I., Lang, A.S., Suttle, C.A., 2006. Metagenomic analysis of coastal RNA virus communities. *Science* 312, 1795–1798.
- Culley, A.I., Lang, A.S., Suttle, C.A., 2007. The complete genomes of three viruses assembled from shotgun libraries of marine RNA virus communities. *Virology Journal* 4, 69.
- Culley, A.I., Mueller, J.A., Belcaid, M., *et al.*, 2014. The characterization of RNA viruses in tropical seawater using targeted PCR and metagenomics. *mBio* 5, 1–11.
- Dolja, V.V., Koonin, E.V., 2018. Metagenomics reshape the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Research* 244, 36–52.
- Gleason, F.H., Jephcott, T.G., Küpper, F.C., *et al.*, 2015. Potential roles for recently discovered chytrid parasites in the dynamics of harmful algal blooms. *Fungal Biology Reviews* 29, 20–33.
- Greninger, A.L., DeRisi, J.L., 2015. Draft genome sequences of ciliovirus and brinovirus from San Francisco wastewater. *Genome Announcements* 3, e00651–15.
- Khrantsov, N.V., Upton, S.J., 2000. Association of RNA polymerase complexes of the parasitic protozoan *Cryptosporidium parvum* with virus-like particles: heterogeneous system. *Journal of Virology* 74, 5788–5795.
- Kimura, K., Tomaru, Y., 2015. Discovery of two novel viruses expands the diversity of single-stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine diatom. *Applied and Environmental Microbiology* 81, 1120–1131.
- Kimura, K., Tomaru, Y., 2017. Effects of temperature and salinity on diatom cell lysis by DNA and RNA viruses. *Aquatic Microbial Ecology* 79, 79–83.

- Koonin, E.V., Wolf, Y.I., Nagasaki, K., Dolja, V.V., 2008. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nature Reviews Microbiology* 6, 925–939.
- Kraeva, N., Butenko, A., Hlaváčová, J., *et al.*, 2015. *Leptomonas seymouri*: adaptations to the dixenous life cycle analyzed by genome sequencing, transcriptome profiling and co-infection with *Leishmania donovani*. *PLoS Pathogens* 11, 1–23.
- Lachnit, T., Thomas, T., Steinberg, P., 2016. Expanding our understanding of the seaweed holobiont: RNA viruses of the red alga *Delisea pulchra*. *Frontiers in Microbiology* 6, 1–12.
- Lang, A.S., Culley, A.I., Suttle, C.A., 2004. Genome sequence and characterization of a virus (HaRNAV) related to picorna-like viruses that infects the marine toxic bloom-forming alga *Heterosigma akashiwo*. *Virology* 320, 206–217.
- Lang, A.S., Rise, M.L., Culley, A.I., Steward, G.F., 2009. RNA viruses in the sea. *FEMS Microbiology Reviews* 33, 295–323.
- Lawrence, J.E., Suttle, C.A., 2004. Effect of viral infection on sinking rates of *Heterosigma akashiwo* and its implications for bloom termination. *Aquatic Microbial Ecology* 37, 1–7.
- Levin, R.A., Voolstra, C.R., Weynberg, K.D., van Oppen, M.J.H., 2016. Evidence for a role of viruses in the thermal sensitivity of coral photosymbionts. *The ISME Journal* 11, 808–812.
- Løvoll, M., Wiik-Nielsen, J., Grove, S., *et al.*, 2010. A novel totivirus and piscine reovirus (PRV) in Atlantic salmon (*Salmo salar*) with cardiomyopathy syndrome (CMS). *Virology Journal* 7, 309–315.
- Malviya, S., Scalco, E., Audic, S., *et al.*, 2016. Insights into global diatom distribution and diversity in the world's ocean. *Proceedings of the National Academy of Sciences of the United States of America* 113, 1516–1525.
- Miranda, J.A., Culley, A.I., Schvarcz, C.R., Steward, G.F., 2016. RNA viruses as major contributors to Antarctic viroplankton. *Environmental Microbiology* 18, 3714–3727.
- Mizumoto, H., Tomaru, Y., Takao, Y., Shirai, Y., Nagasaki, K., 2007. Intraspecies host specificity of a single-stranded RNA virus infecting a marine photosynthetic protist is determined at the early steps of infection. *Journal of Virology* 81, 1372–1378.
- Moniruzzaman, M., Wurch, L.L., Alexander, H., *et al.*, 2017. Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. *Nature Communications* 8, 16054–16064.
- Munn, C.B., 2006. Viruses as pathogens of marine organisms – From bacteria to whales. *Journal of the Marine Biological Association of the United Kingdom* 86, 453–467.
- Nagasaki, K., Tomaru, Y., Katanozaka, N., *et al.*, 2004. Isolation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Applied and Environmental Microbiology* 70, 704–711.
- Nagy, P.D., Pogany, J., 2006. Yeast as a model host to dissect functions of viral and host factors in tombusvirus replication. *Virology* 344, 211–220.
- Nelson, D.M., Tréguer, P., Brzezinski, M.A., Leynaert, A., Quéguiner, B., 1995. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochemical Cycles* 9, 359–372.
- Scheffter, S.M., Ro, Y.T., Chung, I.K., Patterson, J.L., 1995. The complete sequence of *Leishmania* RNA virus LRV2-1, a virus of an Old World parasite strain. *Virology* 212, 84–90.
- Shirai, Y., Tomaru, Y., Takao, Y., *et al.*, 2008. Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. *Applied and Environmental Microbiology* 74, 4022–4027.
- Shi, M., Lin, X.-D., Tian, J.-H., *et al.*, 2016. Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543.
- Smetacek, V., 1999. Diatoms and the ocean carbon cycle. *Protist* 150, 25–32.
- Sommer, U., Adrian, R., De Senerpont Domis, L., *et al.*, 2012. Beyond the Plankton Ecology Group (PEG) model: Mechanisms driving plankton succession. *Annual Review of Ecology, Evolution, and Systematics* 43, 429–448.
- Steward, G.F., Culley, A.I., Mueller, J.A., *et al.*, 2013. Are we missing half of the viruses in the ocean? *The ISME Journal* 7, 672–679.
- Steward, G.F., Rappe, M.S., 2007. What's the 'meta' with metagenomics? *The ISME Journal* 1, 100–102.
- Steward, G.F., Wikner, J., Cochlan, W.P., Smith, D.C., Azam, F., 1992. Estimation of virus production in the sea: ii. Field results. *Marine Microbial Food Webs* 6, 79–90.
- Suttle, C.A., 2005. Viruses in the sea. *Nature* 437, 356–361.
- Tai, V., Lawrence, J.E., Lang, A.S., *et al.*, 2003. Characterization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (Raphidophyceae). *Journal of Phycology* 39, 343–352.
- Takao, Y., Mise, K., Nagasaki, K., Okuno, T., Honda, D., 2006. Complete nucleotide sequence and genome organization of a single-stranded RNA virus infecting the marine fungoid protist. *Journal of General Virology* 87, 723–733.
- Tarr, P.I., Aline, R.F., Smiley, B.L., Keithly, J., Stuart, K., 1988. LR1: A candidate RNA virus of *Leishmania*. *Proceedings of the National Academy of Sciences of the United States of America* 85, 9572–9575.
- Tomaru, Y., Katanozaka, N., Nishida, K., *et al.*, 2004. Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga *Heterocapsa circularisquama*. *Aquatic Microbial Ecology* 34, 207–218.
- Tomaru, Y., Mizumoto, H., Nagasaki, K., 2009a. Virus Research in the toxic bloom-forming dinoflagellate *Heterocapsa circularisquama* to single-stranded RNA virus infection. *Environmental Microbiology* 11, 2915–2923.
- Tomaru, Y., Nagasaki, K., 2007. Flow cytometric detection and enumeration of DNA and RNA viruses infecting marine eukaryotic microalgae. *Journal of Oceanography* 63, 215–221.
- Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T., Nagasaki, K., 2009b. Isolation and characterization of a single-stranded RNA virus infecting the bloom-forming diatom *Chaetoceros socialis*. *Applied and Environmental Microbiology* 75, 2375–2381.
- Tomaru, Y., Toyoda, K., Kimura, K., 2018. Occurrence of the planktonic bloom-forming marine diatom *Chaetoceros tenuissimus* Meunier and its infectious viruses in Western Japan. *Hydrobiologia* 805, 221–230.
- Tomaru, Y., Toyoda, K., Kimura, K., *et al.*, 2012. First evidence for the existence of pennate diatom viruses. *The ISME Journal* 6, 1445–1448.
- Tomaru, Y., Toyoda, K., Kimura, K., *et al.*, 2013. Isolation and characterization of a single-stranded RNA virus that infects the marine planktonic diatom *Chaetoceros* sp. (SS08-C03). *Phycological Research* 61, 27–36.
- Vlok, M., Lang, A.S., Suttle, C.A., 2019a. Application of a sequence-based taxonomic classification method to uncultured and unclassified marine single-stranded RNA viruses in the order *Picornavirales*. *Virus Evolution* 5, vez056.
- Vlok, M., Lang, A.S., Suttle, C.A., 2019b. Marine RNA virus quasispecies are distributed throughout the oceans. *mSphere* 4, e00157–19.
- Wang, A.L., Wang, C.C., 1986. Discovery of a specific double-stranded RNA virus in *Giardia lamblia*. *Molecular and Biochemical Parasitology* 21, 269–276.
- Wang, A.L., Wang, C.C., 1986. The double-stranded RNA in *Trichomonas vaginalis* may originate from virus-like particles. *Proceedings of the National Academy of Sciences of the United States of America* 83, 7956–7960.
- Weinbauer, M.G., 2004. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* 28, 127–181.
- Wolf, Y.I., Kazlauskas, D., Iranzo, J., *et al.*, 2018. Origins and evolution of the global RNA virome. *mBio* 9, e02329-18.
- Wu, B., Zhang, X., Gong, P., *et al.*, 2016. *Eimeria tenella*: A novel dsRNA virus in *E. tenella* and its complete genome sequence analysis. *Virus Genes* 52, 244–252.

Algal Mimiviruses (*Mimiviridae*)

Ruth-Anne Sandaa and Håkon Dahle, Department of Biological Sciences, University of Bergen, Bergen, Norway

Corina PD Brussaard, NIOZ Royal Netherlands Institute for Sea Research, Den Burg, Texel, The Netherlands and Utrecht University, Utrecht, The Netherlands

Hiroyuki Ogata and Romain Blanc-Mathieu, Institute for Chemical Research, Kyoto University, Kyoto, Japan

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Algal viruses Viruses infecting eukaryotic algae.

Intein An intein is a “protein intron” that catalyzes self-splicing at the protein level. The self-splicing is carried out by the autocatalytic excision of the intein from a precursor host protein where it is located, and the concomitant ligation of the flanking protein segments (“exteins”) of the precursor. Inteins often possess a homing endonuclease domain, which confers the mobility on the intein-encoding DNA.

MIGE (Major Interspersed Genomic Element) Repeated sequences throughout genomes.

NCLDV Nucleocytoplasmic large DNA viruses is a proposed order of viruses containing several families that all share certain genomic and structural characteristics.

NCVOG Nucleo-Cytoplasmic Virus Orthologous Groups is a database for homologous genes vertically inherited within at least one viral family in the NCLDV group.

ORF Open reading frame, the part of a reading frame that has the ability to be translated.

Phylogenetic diversity A measurement of biodiversity which incorporates phylogenetic difference between species.

Phylogenetics The study of evolutionary history and relationship among individuals or groups of species or populations.

Taxonomic richness The number of different taxa represented in an ecological community.

Virophage Small double stranded DNA viruses that require the co-infection of another virus for the replication.

History

During the last decades it has become clear that viruses are the most abundant biological entities in aquatic environments. By their lytic activity, viruses are linked to several vital ecological processes, such as the structuring of host communities, and biochemical cycling. They are also key players in the evolution of their hosts. Viral abundances in aquatic systems are typically in the order of 10^6 – 10^7 viruses per mL, and it is estimated that viral lysis is responsible for releasing as much as 10 billion tons of organic carbon into the ocean per day. Viral activity contributes not only to nutrient recycling (that fuels the productivity in the upper euphotic ocean) but also impacts the transfer efficiency of organic matter to higher trophic levels and the transport of organic matter to the depths of the oceans. The viral shunt is also proposed to increase the efficiency of the carbon pump. This is because the viral lysis of host cells releases labile NP-rich organic matter (e.g., peptides) that can be re-used to grow new cells, while the less labile, carbon enriched (e.g. cell wall), part of the host cell tend to aggregate and sink.

Most of the marine viruses infect the numerically dominant microorganisms, including the unicellular algae (phytoplankton). As early as 1981 large icosahedral viruses infecting *Chlorella* were reported and in 1982 a virus infecting the photosynthetic protist *Micromonas pusilla* was described. Currently viruses have been isolated for all taxonomic groups of phytoplankton. Originally, all double-stranded (ds) DNA icosahedral viruses infecting photosynthetic protists were first classified into the family *Phycodnaviridae*. Recent phylogeny based on Nucleocytoplasmic Large DNA Viruses (NCLDV) orthologous genes, however, shows that many of the large dsDNA algal viruses belong to the family *Mimiviridae* (Fig. 1). The enormous diversity found within this family suggests that *Mimiviridae* has a deep co-evolutionary history with its marine protists hosts, dating back to the early history of life. The families *Mimiviridae* and *Phycodnaviridae* have been shown to dominate the NCLDV populations in marine metagenomes with concentrations up to $>10^4$ genomes per mL in the photic zone. In this article we describe the phytoplankton viruses belonging to the *Mimiviridae* family.

One of the first isolated large dsDNA viruses, that was later assigned to the *Mimiviridae* family, was *Phaeocystis pouchetii* virus (PpV) isolated in 1995. A few years later (1998) two more viruses, belonging to this subgroup, were characterized, namely *Chrysochromulina ericina* virus (CeV 01B) and *Pyramimonas orientalis* virus (PoV 01B). The genome size of these viruses were the largest ever reported at that time with sizes being 474 and approx. 560 kbp, respectively. All these viruses were isolated from coastal waters in a Norwegian fjord (Raunefjorden). Some years later (2005), characterization of different *Phaeocystis globosa* viruses (PgVs), isolated from the southern North Sea, was published. Some of these PgV viruses belong to the family *Phycodnaviridae* and are denoted PgV Group II, whilst those belonging to the family *Mimiviridae* are called PgV Group I. PgV-16T (Group I, genome size: 460 kbp) was the first of the algal mimiviruses (members of the family *Mimiviridae*) to be fully sequenced (in 2013). In the following decade several new viruses have been linked to this family, namely *Aureococcus anophagefferens* virus (AaV), *Haptolina ericina* virus (HeV RF02), two *Prymnesium kappa* viruses (Pkv RF01, Pkv RF02), *Prymnesium parvum* virus (PpDNAV), and *Tetraselmis* virus (TetV-1). In addition, metagenome assembled genomes have been recovered for three non-isolated *Mimiviridae*, i.e., two organic lake phycodnaviruses (OLPV 1 and 2) and Yellowstone lake mimivirus (YSLGV). The

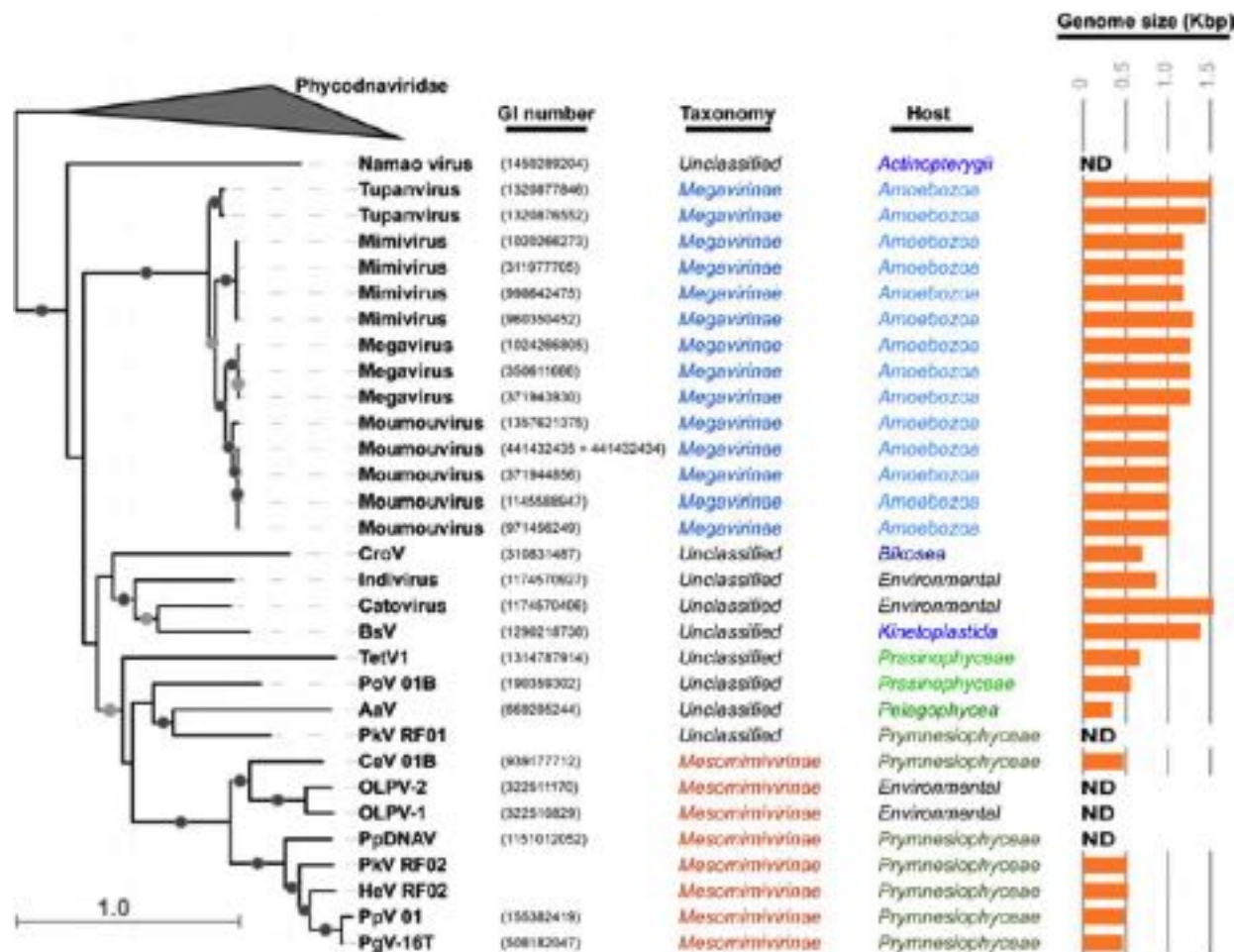


Fig. 1 Maximum likelihood phylogenetic reconstruction on DNA polymerase family B (PolB) using RAXML. Members of *Phycodnaviridae* were used as an outgroup. Circles between internal nodes indicate bootstrap values above 70% (light gray), and above 90% (dark gray). Viral host are color coded by heterotrophs (blue) and phototrophic algae (green). The scale bar indicates the average number of substitutions per site. The proposed subfamilies *Megavirinae* and *Mesomimivirinae* are not yet approved by the ICTV.

two firsts are nearly-complete genomes while the third one is partial. These three viruses are phylogenetically most related to algae-infecting mimiviruses suggesting that phytoplankton are their hosts. As early as 2005 the search of mimivirus proteomes against Sargasso Sea metagenomes, both concomitantly released a year before, indicated the presence of mimivirus-relatives in marine environment. Following analyses of several marker gene sequences (DNA polymerase type B, DNA-directed RNA polymerase (RNAPol)), asparagine synthetase (AsnS) and DNA mismatch repair enzyme (MutS) from various marine genomic datasets hinted to the vast phylogenetic diversity of *Mimiviridae* and their large taxonomic richness. Recently, diversity analysis of DNA-directed RNA polymerase subunits 1 and 2 fetched from bacteria-sized metagenomes collected during the *Tara* Oceans expedition indicated that the taxonomic richness of the family *Mimiviridae* exceeds that of a domain of cellular organisms (i.e., Bacteria). Importantly, the vast majority of this untapped diversity falls in the vicinity of the algae-infecting rather than heterotroph-infecting clade, although their host may include non-photosynthetic organisms.

General Properties

All so far isolated algae-infecting viruses in the *Mimiviridae* family have icosahedral structures between 130 up to 310 nm in diameter and dsDNA genomes. They all replicate in the cytoplasmic area (defined as virus factories) of the host cells, thanks to their own well-conserved transcription machinery, and produce lytic viruses that are transmitted horizontally. The latent time ranges from 10 to 72 h (Table 1). Phylogenetic analysis of DNA polB (Fig. 1) shows that they group in a strongly supported monophyletic group within the *Mimiviridae* family, most similar to viruses infecting heterotrophic protists and less to viruses infecting other autotrophic protists in the *Phycodnaviridae* family. Nevertheless, several genes of *Phycodnaviridae* origin are detected in the genomes of *Mimiviridae* viruses, suggesting a common origin. A comparative genomics study between NCLDV reported

Table 1 Viruses infecting unicellular eukaryotic algae belonging to the family *Mimiviridae*

Viruses		Host		Characteristics							
Strains	Group/subfamily	Strain	Class	Size (nm)	Genome size (Kbp)	GC %	Linear/circular genome	Latent Period (h)	Replication in host cytoplasm	Virophage	Origin of isolation
AaV	Unclassified algae-infecting <i>Mimiviridae</i>	<i>Aureococcus anophagefferens</i>	Pelagophyceae	140	340	29	Linear	60–72	+	No	East-coast, USA
CeV 01B	<i>Mesomimivirinae</i>	<i>Haptolina ericina</i> ^a	Prymnesiophyceae	160	474	25	Linear	14–19	+	No	West-coast, Norway
HeV RF02	<i>Mesomimivirinae</i>	<i>Haptolina ericina</i> and <i>Prymnesium kappa</i>	Prymnesiophyceae	190 × 160	530	ND	ND	14–18	+	ND	West-coast, Norway
PgV-16T	<i>Mesomimivirinae</i>	<i>Phaeocystis globosa</i>	Prymnesiophyceae	150	470	32	Linear	10	+	Yes	Southern North Sea
PkV RF01	unclassified algae-infecting <i>Mimiviridae</i>	<i>Prymnesium kappa</i> and <i>Haptolina ericina</i>	Prymnesiophyceae	310	ND	ND	ND	24–32	+	ND	West-coast, Norway
PkV RF02	<i>Mesomimivirinae</i>	<i>Prymnesium kappa</i>	Prymnesiophyceae	160	507	ND	ND	12–16	+	ND	West-coast, Norway
PoV 01B	Unclassified algae-infecting <i>Mimiviridae</i>	<i>Pyramimonas orientalis</i> and <i>Haptolina ericina</i> ^a	Prasinophyceae	220 × 180	560	ND	ND	14–19	+	ND	West-coast, Norway
PpV 01	<i>Mesomimivirinae</i>	<i>Phaeocystis pouchetii</i>	Prymnesiophyceae	130–160	485	ND	ND	12–18	+	ND	West-coast, Norway
PpDNAV	<i>Mesomimivirinae</i>	<i>Prymnesium parvum</i>	Prymnesiophyceae	221	ND	ND	ND	24–48	+	ND	East-coast, England
TetV1	Unclassified algae-infecting <i>Mimiviridae</i>	<i>Tetraselmis</i> sp.	Prasinophyceae	226–257	668	41	Circular	nd	+	No	Coast of Oahu, Hawaii

^ahost taxonomy reassigned from *Chrysochromulina ericina* to *Haptolina ericina*.

Source: ND: Not determined.

Table 2 Characteristics proteins for algae infecting *Mimiviridae*. (DNA polB) DNA polymerase type B, (MCP 1) Major capsid protein 1, (RNA pol II) DNA-directed RNA polymerase, (AsnS) asparagine synthetase, (MutS7) DNA mismatch repair enzyme, (eIF-4E) eukaryotic translation initiation factor 4E

Strains	DNA polB	MCP1	A32-like virion packing ATPase	RNA pol II	MutS7	AsnS	eIF-4E
AaV	+	+	+	+	+	No	No
CeV-01B	+	+	+	+	+	+	+
PgV-16T	+	+	+	+	+	+	+
PkV RF01	+	+	+	+	+	+	+
TetV-1	+	+	+	+	+	+	+
HeV-RF02	+	+	+	+	+	+	+
PkV-RF02	+	+	+	+	+	+	+
PoV 01B	+	+	ND	ND	+	ND	ND
PpV 01	+	+	ND	ND	+	ND	ND
PpDNAV	+	+	ND	ND	ND	ND	ND

Source: ND: not determined.

22 proteins unique (not seen in other NCLDV) to the *Phycodnaviridae-Mimiviridae* clade and a conservative estimates of 74 proteins present in their last common ancestor. As a comparison these numbers were respectively 27 and 66 for the Chordo/Entomo-poxvirus clade, whose shared ancestry is well acknowledged. It should be noted that the genomes of algae-infecting mimiviruses as well as the distant phycodnavirus *Heterosigma akashiwo* virus HaV53 were not available at that time for comparison. Among the 22 previously reported proteins unique to the *Phycodnaviridae-Mimiviridae* clade, those with notable functions and also encoded in some algae-infecting mimiviruses are transcription factor TFIIB, putative glycosyltransferases, prolyl hydroxylase, DNA J like chaperone and cytidine/deoxycytine deaminase.

Four of the isolated viruses are fully sequenced with genome sizes ranging from 340 to 668 kbp (Table 1). Three of the viruses have linear genomes (CeV, PgV-16T, and AaV), while the genome of TetV-1 is circular. Most of these genomes exhibit high A + T contents, with G + C contents between 23%–32%. One exception is TetV, that has a G + C content of 41% (Table 1). A general feature for these viruses is a relatively low share of common “core” genes and a high proportion of unique genes that is not shared with other members within the *Mimiviridae* family. Of the NCLDV orthologous genes, DNA polB (NCVOG0038), Major capsid protein 1 (MCP1, NCVOG0022), A32-like virion packing ATPase (NCVOG0249), RNAPol (RNA pol subunit I: NCVOG0274 and subunit II: NCVOG0271), asparagin synthase (AsnS, NCVOG0061), eukaryotic translation initiation factor 4E (eIF-4E, NCVOG4661) and MutS7 (NCVOG2626) are all shared in the sequenced mimivirus genomes (Table 2), except AaV that is missing AsnS and eIF-4E homologs. DNA repair enzymes are essential for cellular life but have been detected in some viruses with large genomes. Viral homologs of MutS7 shows homology to MutS7 in *Epsilonproteobacteria* and in the mitochondrial genome of octocorals. The presence of MutS7 homologs in viral genomes therefore provide strong evidence for the placement in the *Mimiviridae* family. Among these core genes, the presence of an additional copy of the RNA pol subunit 2 in the genomes of algae-infecting mimiviruses, with the exception of AaV, is a notable feature that can be used to discriminate between algae- and heterotroph-infecting members of the family *Mimiviridae*. The seven core genes mentioned above are present in the unpublished/partially sequenced genomes of PkV RF01, PkV RF02, and HeV RF02 (Table 2). For PpV, PoV, and PpDNAV, presence of only some of these genes have been tested by PCR followed by sequencing. All three viruses have DNAPolB and MCP1, while PoV and PpV contain MutS7 in addition (Table 2). None of the viruses have been tested for presence of A32-like virion packing ATPase or RNA polymerase subunit II.

Another shared feature between algae and heterotroph-infecting *Mimiviridae* viruses is the presence of several tRNAs and other proteins involved in translation. In line with their smaller genome size (Fig. 1), algae-infecting *Mimiviridae* possess less genes coding for translation related proteins compared to heterotroph-infecting mimiviruses. While mimiviruses infecting heterotrophs encode several aminoacyl-tRNA synthetases, up to 20 for *Tupanvirus soda lake* (genome size: 1.4 Mb), 7 for *Megavirus chiliensis* (genome size: 1.2 Mb) to as few as one (Ile-aaRS) for *CroV* (genome size: 730 kbp), none have been reported in TetV1, CeV 01B, PgV-16T, nor AaV. The notable exception is a catalytic domain for Asn/Asp-tRNA synthetase encoded in the genome of CeV 01B. The closest homologs to this ORF are found in *Prochlorococcus* phages P-SSM2, 5, and 7. However, algae-infecting mimiviruses possess no less tRNA genes (AaV and PgV: 8, CeV: 12, TetV:10) compared to other mimiviruses. tRNA genes tend to occur in clusters in both algae and heterotroph-infecting mimiviruses. CeV, PgV and TetV encode the translation initiation factor 4F (eIF-4E) shared and monophyletic with other mimivirus eIF-4E proteins. The stramenopile-infecting mimivirus AaV encodes the translation elongation factor eEF-1 alpha. eEF-1 alpha is shared with heterotroph-infecting mimiviruses, but not with other algae-infecting mimiviruses. AaV eEF-1 alpha has its closest homologs in genomes of diatoms (62% amino acid (aa) identity) and is very distantly related to the eEF-1 alpha of heterotroph-infecting mimiviruses (aa identity <30%). AaV additionally codes for the translation initiation factor eIF-5A, which presents 56% aa identity to a *Aureococcus anophagefferens* homolog, and the only, yet distant, viral homolog is seen in *Orpheovirus*. TetV-1 encodes the only viral homolog for the eukaryotic translation initiation factor eIF-1A. PgV-16T possesses the translation elongation factor eEF3, not seen in other *Mimiviridae*, and harboring 63% aa identity with *Emiliania huxleyi*'s eEF3. Interestingly, chloroviruses (genus *Chlorovirus*, family *Phycodnaviridae*) also encode an eEF3 sequence most similar (60% aa identity) to a homolog in *Chlorella variabilis*. The sparsity of translation factors in algae-infecting mimiviruses, and the similarity with homologs in host (or relatives), indicate recurrent acquisition of these genes from their host.

The biological underpinnings for the acquisition of these translation factors is unclear. But, it may be needed to counteract the shutdown of host translation machinery upon infection.

Seven of the ten cultured algal mimiviruses infect haptophytes belonging to the class Prymnesiophyceae, two infect hosts belonging to the Prasinophyceae and one infects a host belonging to the Pelagophyceae. These hosts are deeply separated in the eukaryotic tree demonstrating the wide span in host range for viruses belonging to this group. Most viruses infecting eukaryotic hosts are thought to be specialists, infecting only one species or strain within a species. Most (7) of the cultured algal mimiviruses are species- or strain-specific, while the other three can infect different species (HeV RF02, PkV RF01, PoV 01B). HeV RF02 and PkV RF01 both infect different strains of *Haptolina ericina* and one strain of *Prymnesium kappa*, while PoV 01B infects both one strain belonging to the *Pyramimonas orientalis* and *Haptolina ericina*, respectively (Table 1). The host groups of these algal mimiviruses are ubiquitous and diverse, found in the epipelagic layer of tropical, temperate and Arctic oceans. They play ecologically important roles in the epipelagic ocean, moreover some are forming blooms or are mixotrophs, grazing on bacteria and protists. Viral control of these host population dynamics can thus have imperative consequences for carbon and nutrient fluxes and trophic transfer efficiency.

Currently, there is no official taxonomic subdivision within the *Mimiviridae* specific for the algal viruses. In recognition of the phylogenetic affinity of the distinct clade of algae-infecting mimiviruses, and to acknowledge the clear separation of these viruses from the *Phycodnaviridae* family, the subfamily *Mesomimivirinae* was recently proposed. It encompasses CeV01B, HeV RF02, PgV-16T, PkV RF02, PpDNAV, and PpV01, together with OLPV 1 and 2 (Fig. 1). However, not all algal *Mimiviridae* viruses cluster within the *Mesomimivirinae* subfamily. Some separate into distinct lineages (Fig. 1). These are the viruses PkV RF01, AaV, and PoV 01B together with TetV-1. This group of unclassified algae-infecting mimiviruses indicates a huge diversity to be discovered and novel distinct taxonomic groups will certainly be described as more related viruses are isolated and sequenced.

Proposed Subfamily *Mesomimivirinae*

This subfamily consists of six virus isolates, all infecting haptophytes belonging to the Prymnesiophyceae. Virus characteristics vary strongly within this subfamily, as reflected in capsid size (130–221 nm) and gene contents of those two representatives that have been sequenced, namely PgV-16T (GenBank accession no. KC662249) and CeV 01B (GenBank accession no. KT820662). Both PgV-16T and CeV 01B contain genomic repeats, MIGE (Major Interspersed Genome Element), with twelve copies in the PgV genome and six copies in CeV. The MIGEs of PgV and CeV are grouped in separate clusters based on their phylogeny, indicating a separate evolution from a single copy to multicopy duplication. Another common feature between these two viral genomes is the detection of inteins. PgV-16T contains only a single intein, within the coding region of a helicase, while the genome of CeV 01B contains an unusually large number (8) of inteins that are inserted in seven different genes, strongly biased against DNA processing enzymes. Two similar inteins are detected in two unrelated genes, a Lon protease and a DEAD-box helicase. This is quite unusual as an intein normally spreads over homologous (but intein-free) genes with the use of the intein encoded homing endonuclease.

PgV-16T also contains two transposases with strong similarity with viral homologs in OLPV1 (also in two copies). Furthermore, PgV-16T also contained several DNA methylases and type II restriction endonucleases, suggesting that the viral genome contains modified nucleotides as also found for *Paramecium bursaria* Chlorella Virus 1 (*Phycodnaviridae*). An interesting feature is the finding of a virophage genome (PgVV) in the PgV-16T genome data with only three genes homologous to virophages detected in the OLPVs metagenomes and Mavirus infecting the CroV. Surprisingly, no small virions are detected in the cytoplasm together with PgV during infections. Nor is a capsid like gene detected in the PgVV genome. It is therefore suggested that PgVV exists either as a linear plasmid-like molecule, or as a randomly integrated provirus. If so, PgVV might be the first example of a virophage/transposonvirovion intermediate mobile element.

The partially assembled genome sequences of two other PgV strains (PgV-12T and PgV-14T, GenBank accession no. HQ634147.1 and HQ634144.1, respectively) showed a similarity in their gene content and identity (as high as 98% to the fully sequenced PgV-16T genome). All these PgVs are part of the Group I PgV, with similar large dsDNA genomes and clustering based on phylogeny (DNA pol B) and phenological characteristics. PgV-07T, also part of Group I, have shown to contain a lipid membrane, detected by staining with lipophilic dye and chloroform and diethyl ether treatment. Intact polar lipids (IPL) analysis displayed that the IPL composition of PgV-07T was different than its host *P. globosa* G(A). In contrast to EhV infecting the prymnesiophyte *Emiliania huxleyi*, no viral glycosphingolipids (vGSLs) were detected. PgV seems to obtain its lipid membrane in the host cytoplasm and not from the cell membrane as observed for EhV. The lipid membrane of PgV is argued to be an intracapsid membrane, similar to what is proposed for PpV (infecting *P. pouchetii*) using electron cryomicroscopy.

PgV-07T is highly sensitive to a combination of phosphate (P) limited and light stressed growth of its host, resulting in a 95% loss in infectivity of the progeny viruses. Host metabolism strongly regulates the proliferation of PgV, with prolonged latent period and reduced burst size under low and high light intensity (host energy limitation and inhibition) as well as nutrient limitation. Not only phosphorus was shown to strongly limit PgV production but also iron and in particular nitrogen limited host growth resulted in lowered burst sizes as compared to resource replete conditions (i.e., 70%, 72% and 92% reduction, respectively). *P. globosa* occurs in temperate coastal waters with strong gradients in light and nutrient availability. Overall, all hosts of the *Mesomimivirinae* viruses are found in waters with strong seasonal variability in growth-related environmental variables. With their important ecological function, the magnitude of viral control of these host species will vary temporarily, geographically and with depth.

Unclassified Algae-Infecting Members in the Family *Mimiviridae*

AaV (GenBank accession no. KJ645900) infects *Aureococcus anophagefferens* that causes “brown tides” in coastal and estuarine waters of the United States, China and South Africa. The ecological and economical consequences of these blooms is enormous as blooms inhibit penetration of light and kill seagrass and juvenile shellfish. The virus was isolated from the east coast of the United States outside New York. AaV is the smallest member of viruses infecting photosynthetic protists within the *Mimiviridae* family with a sequenced genome of 370 kbp and a capsid size of 140 nm. The genome architecture of AaV demonstrates that the ancestral virus had even a smaller genome, which has been expanded through gene duplication and assimilation of genes from several sources including its host.

AaV has been shown to be an outlier within the *Mimiviridae* family. In sequence-based analyses, AaV encodes many genes that are more similar to genes of mimiviruses than of other NCLDV and its core NCLDV genes showed higher sequence similarity, and in a number of cases, clear orthologous relationship with mimivirus homologs. However, AaV is missing the asparagine synthase, the polyA polymerase and eIF-4E present in other mimiviruses. An additional unique feature of AaV compared to other mimiviruses is a duplication of RNA polymerase II largest subunit (rpb1). Although it possesses a second, yet incomplete (222 aa against >1000 aa for canonical proteins) version of the RNA polymerase second largest subunit (rpb2), these two seem to be only distantly related as it is for other algal mimiviruses. It is also notable that twenty-two AaV genes had highest sequence similarity with phycodnavirus genes. In addition several photolyase genes and genes involved in DNA repair are found, as also found in several large DNA viruses. Phylogenetically it makes a long branch within the group unclassified algae-infecting mimiviruses, with the closest (but still distant) relative being PkV RF01 (Fig. 1).

PkV RF01, infecting the haptophytes *P. kappa* and *H. ericina*, has the largest capsid size (310) of all the so far cultured algae-infecting mimiviruses. Interestingly, PkV RF01 infects the same hosts as HeV RF02 and PkV RF02 that belong to the subfamily *Mesomimivirinae* indicating the complexity of virus-host interactions and further the difficulties in naming viruses. Similar to HeV RF02, PkV RF01 is able to proliferate in more than one genus indicating that broad host range may perhaps be common in marine ecosystems as the ecological advantage seems obvious. The persistent hosts of PkV RF01, *H. ericina* and *P. kappa*, do normally not form massive blooms, as many of the hosts of other isolated algal viruses (e.g., PgVs, AaV, EhV), but are present at low densities at different seasons. As evolution between host and its parasite always will aim towards host survival rather than high mortality, these viral-host systems may represent the most typical systems in the marine environments; low-density hosts (non-blooming) and viruses that appear to have co-evolved for a long time. Although its closest yet very distant relative is AaV, PkV RF01 encodes the AsnS and eIF4-4E mimivirus core genes (Table 2).

Two mimiviruses infecting Chlorophyta have been described, namely PoV 01B and TetV-1. Both viruses have large capsid size (220–257 nm) and genomes (560–668 kbp). Based on DNA polB and MutS7 phylogeny they are more closely related to each other than to other viruses within this algal *Mimiviridae* group. PoV 01B was isolated from the western coast of Norway while TetV-1 was isolated from the coast of Hawaii. PoV 01B has unfortunately been lost from our culture collection. The genome of TetV-1 (Genbank accession no. KY322437) contains several key genes never detected in viruses before. These genes include a mannitol metabolism enzyme, a saccharide degradation enzyme and the key fermentation genes pyruvate formate-lyase and pyruvate-lyase activating enzyme. The presence of the two latter enzymes suggest that TetV-1 has the potential to manipulate its host fermentative pathway. It is speculated that green algae use fermentation to survive periods with low oxygen concentration. Thus, TetV-1 might use the viral homolog of these genes to ensure energy requirement for virus production especially after increased bacterial respiration that follows bloom termination by viruses. The saccharide degradation enzyme in TetV-1, alpha-galactosidase, is suggested to enhance cell lysis as the cell wall of *Tetraselmis* contain up to 7% galactose and 21% galacturonic acid. The same function has been suggested for the mannitol 1-phosphate dehydrogenase gene as the host, *Tetraselmis*, uses mannitol as a osmolyte under hypersaline conditions. Manipulation of such system might therefore induce host lysis.

Further Reading

- Baudoux, A., Brussaard, C., 2005. Characterization of different viruses infecting the marine harmful algal bloom species *Phaeocystis globosa*. *Virology* 341, 80–90.
- Claverie, J.-M., Abergel, C., 2018. *Mimiviridae*: An expanding family of highly diverse large dsDNA viruses infecting a wide phylogenetic range of aquatic eukaryotes. *Viruses* 10, 506.
- Gallot-Lavallee, L., Blanc, G., Claverie, J.-M., 2017. Comparative genomics of *Chrysochromulina ericina* virus (CeV) and other microalgae-infecting large DNA viruses highlight their intricate evolutionary relationship with the established *Mimiviridae* family. *Journal of Virology* 91 (14), e00230-17.
- Hingamp, P., Grimsley, N., Acinas, S.G., et al., 2013. Exploring nucleocytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *The ISME Journal* 7, 1678–1695.
- Iyer, L.M., Balaji, S., Koonin, E.V., Aravind, L., 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Research* 117, 156–184.
- Jacobsen, A., Bratbak, G., Heldal, M., 1996. Isolation and characterization of a virus infecting *Phaeocystis pouchetii* (Prymnesiophyceae). *Journal of Phycology* 32, 923–927.
- Johannessen, T.V., Bratbak, G., Larsen, A., et al., 2015. Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). *Virology* 476, 180–188.
- Moniruzzaman, M., LeClerc, G.R., Brown, C.M., et al., 2014. Genome of brown tide virus (AaV), the little giant of the megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. *Virology* 466–467, 60–70.

- Sandaa, R.A., Heldal, M., Castberg, T., Thyraug, R., Bratbak, G., 2001. Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* 290, 272–280.
- Santini, S., Jeudy, S., Bartoli, J., *et al.*, 2013. Genome of *Phaeocystis globosa* virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proceedings of the National Academy of Sciences* 110, 10800–10805.
- Schvarcz, C.R., Steward, G.F., 2018. A giant virus infecting green algae encodes key fermentation genes. *Virology*. 518, 423–433.
- Wagstaff, B.A., Vladu, I.C., Barclay, J.E., *et al.*, 2017. Isolation and characterization of a double stranded DNA Megavirus infecting the toxin-producing haptophyte *Prymnesium parvum*. *Viruses*. 9, 40.

Miscellaneous Algal Viruses (*Alvernnaviridae*, *Bacilladnaviridae*, *Dinodnavirus*, *Reoviridae*)

Keizo Nagasaki, Kochi University, Nankoku, Japan

Yuji Tomaru, Japan Fisheries Research and Education Agency, Kanagawa, Japan

Corina PD Brussaard, NIOZ Royal Netherlands Institute for Sea Research, Den Burg, Texel, The Netherlands and Utrecht University, Utrecht, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

This is an update of K. Nagasaki, C.P.D. Brussaard, Algal Viruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00359-9.

Introduction

With the realization during the last three decades that viruses are highly abundant in various aquatic environments, interest in aquatic viruses significantly increased. Viruses are now recognized as important biological agents not only regulating population dynamics, and succession and diversity of the host organisms in marine systems, but also influencing the functioning of aquatic food webs and biogeochemical cycles (energy and matter fluxes). Members within the microbial food web can become infected by viruses, including eukaryotic algae. In the present article, we summarize the characteristics of some RNA and DNA algal viruses that do not belong to the virus families thus far established officially, e.g., *Phycodnaviridae*, *Mimiviridae*, and *Marnaviridae* (separately discussed in other chapters).

Algal single-stranded RNA Viruses

Genus *Dinornavirus*

Heterocapsa circularisquama RNA virus (HcRNAV) is a positive-sense ssRNA virus infecting the bivalve-killing, bloom-forming dinoflagellate, *H. circularisquama*, which is a type species of the genus *Dinornavirus* in the family *Alvernnaviridae*. The virion is icosahedral, approximately 30 nm in diameter, harbouring an ssRNA genome of around 4.4 kb in length, and is propagated in the host cytoplasm, often forming a crystalline array. Its genome is a linear positive-sense ssRNA, lacking 5'-cap structure and 3'-polyA tail, and has a stem-loop structure at the 3'-end. HcRNAV clones can be roughly divided into two types (types UA and CY) based on their intraspecies host specificity patterns; each type shows its own strain-specific infectivity that is complementary to each other. These two HcRNAV types can coexist in natural water. Typical HcRNAV clones of type UA and CY (HcRNAV34 and 109, respectively) were fully sequenced (DDBJ accession numbers: AB218608 and AB218609, respectively); they are approximately 97% identical at the nucleotide sequence level. Each genome has two open reading frames (ORFs). ORF-1 encodes a putative polyprotein having at least the serine protease domain and the RdRP domain. In about a half of the tested virus clones, a specific 15-nt deletion was found; however, it is not related in the determination of intraspecies host specificity. ORF-2 coding for the single major capsid protein (MCP) is unlikely to be a polyprotein gene because the molecular mass directly estimated by SDS polyacrylamide gel electrophoresis agrees well with the value predicted by the deduced amino acid sequence of ORF-2. Between the two virus clones tested, the stem-loop structures at the 3'-end are different in stem length and loop-size; this may affect their replication efficiency.

Similarity analysis for the deduced amino acid sequences of RdRP revealed that HcRNAV is evolutionarily quite distant from any of the land and aquatic viruses that have been genetically studied. This is also supported by a phylogenetic analysis of the RdRP amino acid sequence.

Genomic comparison revealed that complementary host ranges of HcRNAV may be related to the amino acid substitution patterns in ORF-2, within the MCP-encoding gene. In addition, the tertiary structure of the MCPs predicted by using computer modelling indicated that many of the amino acid substitutions were located in regions on the outer portion of the viral capsid proteins exposed to the ambient water environments. This suggests that the intraspecies host specificity of HcRNAV is determined by nano-structural differences of the viral surface, which may affect its binding affinity to the host cell.

By a cross-reactivity test, *H. circularisquama* clones are also divided into two types according to the sensitivity spectra of the two HcRNAV types; however, the two host types are indistinguishable when their morphology or the sequences of the internal spacer regions of the ribosomal RNA genes are compared. There are at least two distinct and independent host/virus systems between *H. circularisquama* and HcRNAV; i.e., multiple types of host and virus systems coexist within natural blooms of *H. circularisquama* and their combinations are regulated by exquisite molecular mechanisms.

Algal double-stranded RNA Viruses

Genus *Mimoreovirus*

The *Micromonas pusilla* reovirus (MpRV, originally termed as MpRNAV) is a double-stranded (ds) RNA virus infecting the cosmopolitan prasinophyte, *M. pusilla*. The polysegmented dsRNA genome of MpRV identified it as a member of a large family, *Reoviridae*, arranging

nonenveloped viruses with segmented (9–12 linear segments) dsRNA genomes. The virus was characterized by sequence, morphological and physiochemical properties and is thus far the first and only member of the genus *Mimoreovirus*. Its genome is composed of 11 segments ranging between 741 and 5792 bp, with a total size of 25,563 bp. When isolated, the virus coexisted with a large genome-sized dsDNA virus infecting the same *M. pusilla* LAC38 strain (recently renamed as *M. commoda*; van Baren *et al.*, 2016). It has a narrow host range, a latent period of 36 h (based on a one-step lytic virus growth cycle and TEM enumeration of the virus particles), and infectivity sensitivity to temperatures $>35^{\circ}\text{C}$. Non-ionic detergents and chloroform or diethyl ether did not affect infectivity, while alcohol and acetone did.

The size of intact virus particles is 90–95 nm in diameter (based on TEM images), wherein CsCl purified capsids show a compact inner structure of 75 nm surrounded by a 15 nm outer layer of protein. The smooth surface of the 50 nm subcore particles indicate that MpRV belongs to non-turreted *Reoviridae* (subfamily *Sedoreovirinae*). Within the phylogenetic tree built with *Reoviridae* polymerase (RdRP) sequences, the branch of MpRV dissects the tree, separating the group of turreted and non-turreted viruses. As *M. pusilla* is evolutionarily older than the hosts of other members of the *Reoviridae*, the topology of the tree suggests that the branch of MpRV could be ancestral.

An interesting feature is the unusual length of segment 1 (5792 bp), encoding an outer layer protein of 1877 aa long (VP1) similar to non-turreted intact particles of other viruses. If this is indeed a transient envelope structure, MpRV would be the first of the *Reoviridae* to have acquired a constitutive additional outer coat or pseudo-envelope structure, the consequence of budding of virus particles from the cell membrane or budding into the endoplasmic reticulum during morphogenesis.

The many repeats within the V1 sequence suggest that it may have arisen from amino acid fragment duplication, followed by diversification of the sequence. The mechanism and the constraints which have driven such an evolution are not clear, although very recently a model of stem-loop formation (based on the crystal structure of the polymerase of a mammalian orthoreovirus) was proposed for explaining such duplications.

Algal single-stranded DNA Viruses

Genus *Bacilladnavirus*

Diatoms are important for marine primary production and for support of coastal fisheries. To date, approximately 20 viruses capable of infecting diatoms, including *Centric* and *Pennate* genera have been isolated and characterized. Diatom viruses are divided into two groups based on their genome conformation: single-stranded (ss) RNA or ssDNA viruses. Those harbouring ssRNA are now classified in the genus *Bacillarnavirus* within the family *Marnaviridae*. Diatom-infecting viruses which harbour ssDNA belong to “genus *Bacilladnavirus*”.

The virions of bacilladnaviruses are icosahedral, 32–38 nm in diameter and accumulate in the host nucleus (Fig. 1). In addition to icosahedral virions, fibrous or rod-shaped structures, 17–27 nm in width and $>0.5\ \mu\text{m}$ in length, are often observed in diatom nuclei infected with bacilladnaviruses. The infectivity and role of the rod-shaped particles are still unknown. The ssDNA diatom virus genome is primarily a 5–6 kb, covalently closed, circular ssDNA with a $<1\ \text{kb}$ linear, complementary strand. An exception is CsetDNAV (*Chaetoceros setoensis* DNA virus), which infects the bloom-forming diatom *C. setoensis*. The CsetDNAV genome is composed of a covalently closed circular ssDNA and eight short complementary fragments (67–145 nt in length). The genome of bacilladnaviruses possesses at least two major open reading frames (ORFs) encoding a putative replication-related protein and structural proteins. The replication proteins of bacilladnaviruses display a low similarity to those of circoviruses. Their infection is highly species-specific and strain-specific. As mentioned above, they are also considered to have a significant impact on the ecology of their host populations.

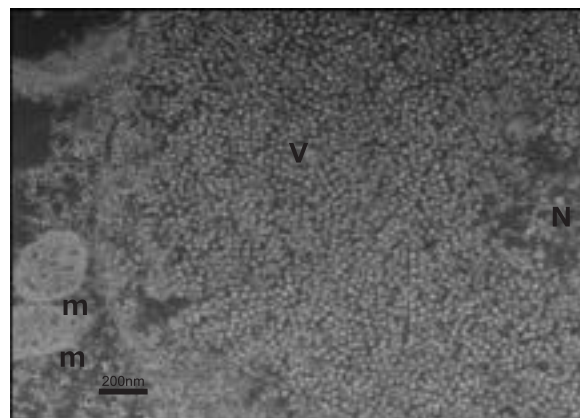


Fig. 1 Transmission electron micrograph of an ultra-thin section of *Chaetoceros setoensis* host strain, NIES-3712, 48 h after CsetDNAV inoculation. Virus-like particles (V, approximately 33 nm in diameter) are accumulated in the host nucleus (N). ‘m’ indicates mitochondrion. Photographed by Y. Tomaru.

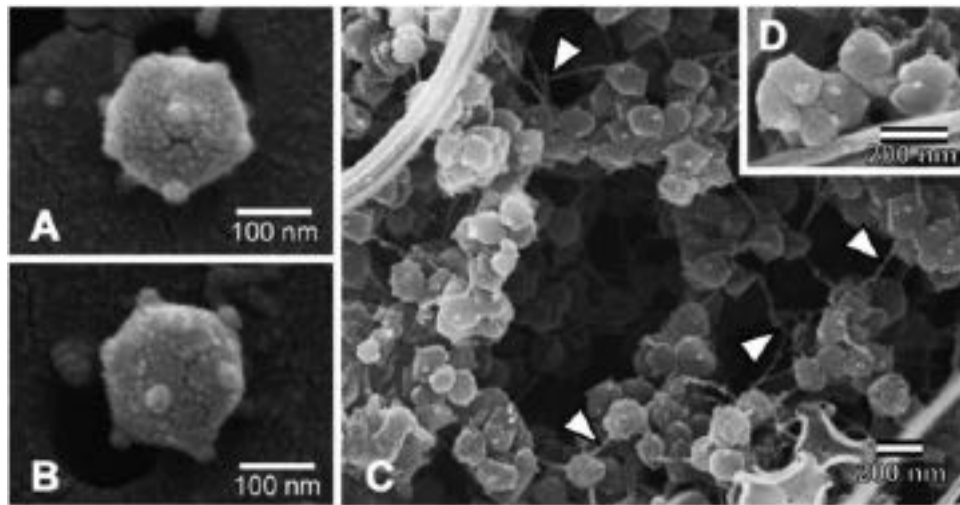


Fig. 2 Field emission SEM images of HcDNAV particles; pentagonal (A) and hexagonal projection (B), and the viroplasm of HcDNAV-infected host *Heterocapsa circularisquama* (C, D). Note that many virus particles do not completely form icosahedral shapes in the viroplasm (i.e., they have deformed shapes), but are equipped with protrusions at the vertex position, and that fibrous materials (arrowheads) are also present. Reproduced from Takano, Y., Tomaru, Y., Nagasaki, K., 2018. Visualization of a dinoflagellate-infecting virus HcDNAV and its infection process. *Viruses* 10, 554.

Algal double-stranded DNA Viruses (not Belonging to Either of the Families: *Phycodnaviridae* or *Mimiviridae*)

Genus *Dinodnavirus*

Heterocapsa circularisquama DNA virus (HcDNAV) is a dsDNA virus infecting the bloom-forming dinoflagellate, *H. circularisquama*, which is a type species of the genus *Dinodnavirus*. The virion is icosahedral and around 200 nm in diameter; each five-fold vertex is decorated with a protrusion (Fig. 2). As well as HcRNAV mentioned above, HcDNAV is reported to play a significant role in the disintegration process of *H. circularisquama* blooms. HcDNAV has a large icosahedral capsid (180–210 nm in diameter) and its genome size is estimated to be approximately 356 kbp in length. During its multiplication process, virions emerge from a specific cytoplasm compartment, called “viroplasm,” or “virus factory,” which is induced during infection by the virus (Fig. 2).

HcDNAV PolB sequence is closely related to the PolB sequence of asfaviruses with dsDNA genomes, exemplified by African swine fever virus. In addition, the amino acid sequence of HcDNAV PolB showed a rare amino acid substitution within a sequence containing a highly conserved motif.

Reference

van Baren, M.J., Bachy, C., Reistetter, E.N., *et al.*, 2016. Evidence-based green algal genomics reveals marine diversity and ancestral characteristics of land plants. *BMC Genomics* 17, 267.

Further Reading

- Attoui, H., Jaafar, F.M., Belhouchet, M., *et al.*, 2006. *Micromonas pusilla* reovirus: A new member of the family Reoviridae assigned to a novel proposed genus (Mimoreovirus). *Journal of General Virology* 87, 1375–1383.
- Brussaard, C.P., Noordeloos, A.A., Sandaa, R.A., *et al.*, 2004. Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. *Virology* 319, 280–291.
- Coy, S.R., Gann, E.R., Pound, H.L., *et al.*, 2018. Viruses of eukaryotic algae: Diversity, methods for detection, and future directions. *Viruses* 10, 487.
- Nagasaki, K., 2008. Dinoflagellates, diatoms and their viruses. *Journal of Microbiology* 46, 235–243.
- Nagasaki, K., Shirai, Y., Takao, Y., *et al.*, 2005. Comparison of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate *Heterocapsa circularisquama*. *Applied and Environmental Microbiology* 71, 8888–8894.
- Nagasaki, K., Tomaru, Y., Takao, Y., *et al.*, 2005. Previously unknown virus infects marine diatom. *Applied and Environmental Microbiology* 71, 3528–3535.
- Takano, Y., Tomaru, Y., Nagasaki, K., 2018. Visualization of a dinoflagellate-infecting virus HcDNAV and its infection process. *Viruses* 10, 554.
- Tomaru, Y., Hata, N., Masuda, T., *et al.*, 2007. Ecological dynamics of the bivalve-killing dinoflagellate *Heterocapsa circularisquama* and its infectious viruses in different locations of Western Japan. *Environmental Microbiology* 9, 1376–1383.
- Tomaru, Y., Katanozaka, N., Nishida, K., *et al.*, 2004. Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga *Heterocapsa circularisquama*. *Aquatic Microbial Ecology* 34, 207–218.
- Tomaru, Y., Toyoda, K., Kimura, K., 2015. Marine diatom viruses and their hosts: Resistance mechanisms and population dynamics. *Perspectives in Phycology* 2, 69–81.

Phycodnaviruses (*Phycodnaviridae*)

James L Van Etten and David D Dunigan, University of Nebraska–Lincoln, Lincoln, NE, United States

Keizo Nagasaki, Kochi University, Nankoku, Japan

Declan C Schroeder, University of Reading, Reading, United Kingdom and University of Minnesota, St. Paul, MN, United States

Nigel Grimsley, Integrative Biology of Marine Organisms Laboratory, Banuyls-sur-Mer, France and Sorbonne University, Banuyls-sur-Mer, France

Corina PD Brussaard, NIOZ Royal Netherlands Institute for Sea Research, Den Burg, Texel, The Netherlands and Utrecht University, Utrecht, The Netherlands

Jozef I Nissimov, University of Waterloo, Waterloo, ON, Canada

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J.L. Van Etten, M.V. Graves, Phycodnaviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy and Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00571-9.

History

The concept that viruses might have a major impact on the marine environment began almost 30 years ago with the discovery that seawater contains around 10^{10} viruses per liter. At any one time, 20%–40% of the photosynthetic microorganisms in the ocean are infected with a virus. Consequently, these viruses contribute to microbial composition and diversity, nutrient cycling, carbon flow, and other biogeochemically-important processes in aqueous environments. This huge viral population consists of both bacterial and algal viruses and they are important because phytoplankton, consisting of cyanobacteria and eukaryotic algae, fix about half of the global carbon dioxide.

Beginning in the early 1970s, viruses or virus-like particles (VLPs) have been reported in many taxa of eukaryotic algae. However, most of the early reports described single accounts of microscopic observations. This situation changed in the 1980s with the discovery of a family of large double-stranded (ds) DNA-containing viruses (referred to as chloroviruses) that infect and replicate in certain strains of unicellular, eukaryotic, symbiotic, chlorella-like green algae; these symbiotic algae are often referred to as zoochlorellae. Known chlorovirus hosts are either associated with the protozoan *Paramecium bursaria* (Fig. 1(A)), the coelenterate *Hydra viridis* or the heliozoae *Acanthocystis turfacea*. Zoochlorellae are resistant to viruses in their symbiotic state. Fortunately, the zoochlorellae from *P. bursaria*, as well as those from *A. turfacea*, can be grown independently in culture, and these cultured, naturally endosymbiotic algae serve as hosts for many large dsDNA viruses. The lytic chloroviruses can be produced in large quantities and assayed by plaque formation (Fig. 1(B)) using standard bacteriophage techniques. The prototype chlorovirus is *Paramecium bursaria chlorella virus 1* (PBCV-1).

Large polyhedral, dsDNA-containing viruses that infect marine algae are also ubiquitous throughout the world. These include viruses that infect the unicellular coccolithophore *Emiliania huxleyi* (EhV viruses, genus *Coccolithovirus*) and the other bloom-forming haptophyte *Phaeocystis globosa* (PgV viruses, genus *Prymnesiovirus*), the noxious bloom-forming raphidopyte *Heterosigma akashiwo* (HaV viruses, genus *Raphidovirus*), viruses that infect filamentous brown algae, e.g., *Ectocarpus* sp. (EsV viruses, genus *Phaeovirus*) and prasinoviruses (genus *Prasinovirus*) that infect the smallest eukaryotic cell, *Ostreococcus* and related species in the class *Mamiellophyceae*. Prasinoviruses infecting *Micromonas* were first reported in 1979 in samples of coastal seawater from British Columbia. However, more *Micromonas* viruses were not discovered until the 1990s, which led to the appreciation of their large diversity. The genomic sequence of the first prasinovirus, *Ostreococcus tauri* virus 5 (OtV5), was published in 2008. Subsequently, other viruses infecting *Ostreococcus* sp. have been sequenced, and the genomes of viruses infecting the genera *Bathycoccus* and *Micromonas* were analyzed. Now many genomes from the genus *Prasinovirus* have been sequenced, and their worldwide diversity and abundance is well documented.

Although all the phycodnaviruses probably arose from a common ancestor, they can have different lifestyles. For example, phaeoviruses have a lysogenic phase in their life-cycle (which is rare among algal viruses); the virus particles are only seen in sporangial cells of their host. Moreover, phaeoviruses also infect the giant kelp seaweed that separated from *Ectocarpus* brown algae (genus *Ectocarpus*) around 90 million years ago, further confirming the ancient evolutionary history of algal viruses.

Because the first algal viruses discovered were large dsDNA viruses, it was initially assumed that algae were only infected by large dsDNA viruses. However, this perception has changed and algal viruses with single-stranded (ss)RNA-, dsRNA- and ssDNA genomes have been discovered in the past two decades. These more recently discovered algal viruses are described in other articles in this 4th Edition of Encyclopedia of Virology entitled “Other algal viruses” and “Marnaviruses (*Marnaviridae*).”

Taxonomy and Classification

Members of the family *Phycodnaviridae* constitute a large genetically and morphologically diverse group of viruses with eukaryotic algal hosts from both inland and marine waters. Accumulating genetic evidence indicates that the phycodnaviruses together with the poxviruses, iridoviruses, asfarviruses, ascoviruses, mimiviruses, marseilleviruses, pithoviruses, molliviruses, faustoviruses and possibly pandoraviruses have a common evolutionary ancestor, perhaps, arising at the point of eukaryogenesis 3 billion or more

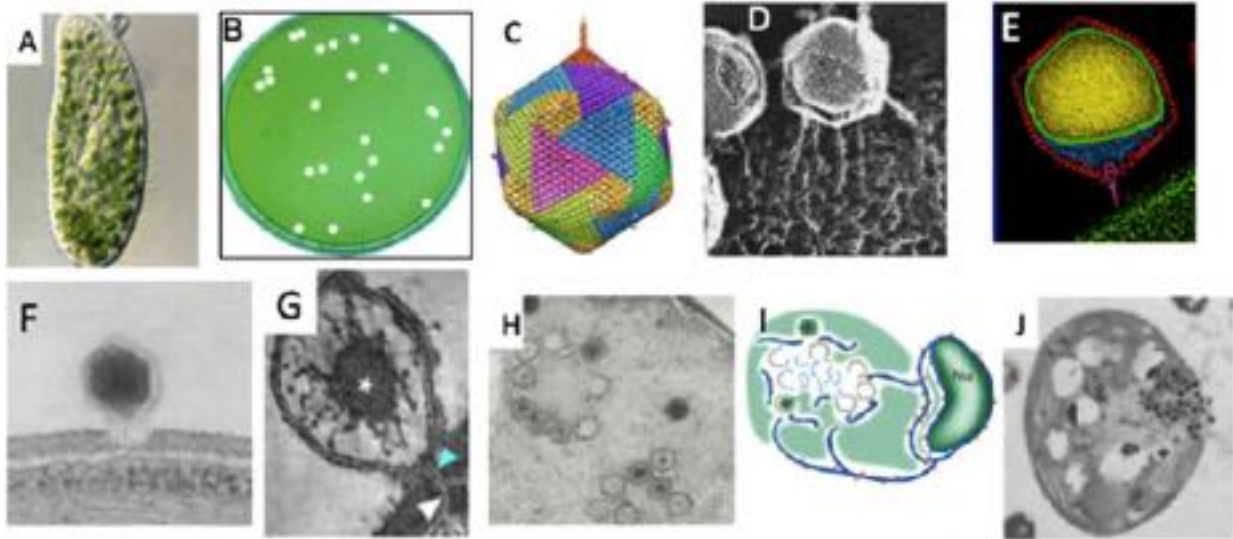


Fig. 1 Chlorella cells and chlorovirus PBCV-1. (A) *Parmecium bursaria* and its symbiotic chlorella cells. (B) Plaques formed by PBCV-1 on a lawn of *Chlorella variabilis* NC64A. (C) Five-fold averaged cryo-electron micrograph of PBCV-1 reveals a long narrow cylindrical spike structure at one vertex and fibers extending from one unique capsomer per trisymmetron. (D) PBCV-1 attached to the cell wall as viewed by the quick-freeze, deep etch procedure. Note the virions attached to the wall by fibers. (E) Initial attachment of PBCV-1 to a *C. variabilis* NC64A cell wall. (F) Attachment of PBCV-1 to the algal wall and digestion of the wall at the point of attachment. This occurs within 1–3 min PI. (G) The generation of a continuous membrane lined tunnel from the virus to the host (arrows). (H) Virion particles assembly in defined areas in the cytoplasm named virus assembly centers at ~4 h PI. Note both DNA containing (dark centers) and empty capsids. (I) A model depicting PBCV-1 assembly into infectious particles including generation of the internal lipid membrane. (J) Localized lysis of cell plasma membrane and cell wall and release of progeny viruses at 8 h PI. This figure is modified with permission. (A)–(F) and (H)–(J) are modified from Van Etten, J.L., Dunigan, D.D., 2016. Giant chloroviruses – Five easy questions. *PLoS Pathogens* 12, e1005751, and (G) is modified from Milrot, E., *et al.*, 2017. *PLoS Path.* 13, e1006562, with permission.

years ago. Collectively, these large dsDNA viruses are referred to as nucleocytoplasmic large DNA viruses (NCLDV) and it has been proposed that these viruses should be included in a new order named *Megavirales*.

Phylogenetic analyses of B-type DNA polymerases encoded by all members of the *Phycodnaviridae* indicate that they are more closely related to each other than to other large dsDNA viruses and that they form a monophyletic group, consistent with a common ancestor. The phycodnaviruses fall into six clades based on phenotypic characteristics and host species and were originally given genus status. Often, the genera can be distinguished by additional properties, e.g., lytic versus lysogenic life styles or linear versus circular genomes. Among viruses with known host species, members of the genus *Chlorovirus* infect fresh water algae; however, chlorovirus DNA sequences are often reported in metagenomic studies from the ocean, which suggests that at least some chloroviruses infect marine algae or some other marine organism(s). In contrast, members of the other five genera currently listed in the International Committee on the Taxonomy of Viruses (ICTV) database (*Coccolithovirus*, *Phaeovirus*, *Prasinovirus*, *Prymnesiovirus*, and *Raphidovirus*) infect marine algae, although DNA sequences from inland water sources indicate that prasinoviruses also exist in inland water.

Complicating the taxonomy of the *Phycodnaviridae* is the fact that genomic sequencing of several recently isolated large dsDNA algal viruses indicates that they are more closely related to viruses in the family *Mimiviridae* than they are to the *Phycodnaviridae*. The largest viruses listed in the genus *Prymnesiovirus* (PgV group I viruses) belong to the *Mimiviridae* family and so these viruses have been removed from this article. However, the PgV group II viruses will probably remain in the *Phycodnaviridae* family and so they are included in this article. In addition, a couple of viruses that infect prasinophytes also belong in the *Mimiviridae* family. These mimivirus-like viruses are described in another article in this 4th Edition of the Encyclopedia of Viruses entitled “Algal viruses belonging to a subgroup within the *Mimiviridae* family.” However, the majority of the viruses that infect prasinophytes belong in the *Phycodnaviridae* and so these members remain in this article. [Table 1](#) lists representative viruses from the six genera in the *Phycodnaviridae*. Another complication is that the predicted amino acid sequences between homologous genes coded by many viruses, now classified as belonging to the same genus, can differ significantly and we expect that these genera will eventually be given family status and that many of the individual viruses will be given genus status.

Virion Structure and Composition

Morphology

Originally, all phycodnaviruses were assumed to be large icosahedral structures (100–220 nm in diameter) with a multi-laminate shell surrounding an internal membrane enclosing an electron dense core. However, recent experiments indicate that not all

Table 1 Representative members of the *Phycodnaviridae*

Genera	Type	Representative	Genome sizes	Number of CDSs
Chloroviruses	NC64A viruses	PBCV-1	330 kb	416
	SAG viruses	ATCV-1	288 kb	329
	Pbi viruses	MT325	325 kb	331
	Osy viruses	Osy-NE5	327 kb	357
	Hydra viruses	HVCV-1	300 kb	unknown
Coccolithoviruses	Emiliana huxleyi viruses	EhV-86	407 kb	472
Phaeoviruses	Ectocarpus viruses	EsV-1	336 kb	231
	Feldmannia viruses	FsV	155 kb	150
	Hicksia viruses	HincV-1	240 kb	unknown
	Myriotrichia viruses	Mclav-1	320 kb	unknown
	Pilayella viruses	PlitV-1	280 kb	unknown
	Laminaria digitata viruses	Ldif	unknown	unknown
	Micromonas viruses	MpV1	184 kb	244
Prasinoviruses	Ostrococcus viruses	OtV1	192 kb	232
	Bathycoccus viruses	BpV1	199 kb	203
Rhaphidoviruses	Heterosigma akashiwo viruses	HaV	275 kb	246
Pymnesioviruses	Phaeocystis globosa group II viruses	PgV II	unknown	unknown

phycodnavirus structures are identical. For example, fivefold-symmetry averaging 3D reconstructions revealed that one of the vertices in the chlorovirus PBCV-1 has a spike-structure (Fig. 1(C)). External fibers also extend from one capsomer in each of the 66 trisymmetron capsomes of PBCV-1 that facilitate attachment to the host (Fig. 1(D)).

Coccolithoviruses and phaeoviruses are reported to have an external membrane in addition to an internal membrane. It is thought that these viruses utilize an animal-like virus infection strategy by entering their hosts via either an endocytotic or an envelope fusion mechanism. The capsid is then disrupted and the viral DNA is transferred to the nucleus of their hosts. During the formation of these viruses, they acquire their external membrane by budding through the host plasma membrane.

Physicochemical and Physical Properties

The Mr of PBCV-1 is $\sim 1 \times 10^9$ daltons and the $S_{20,w}$ is > 2000 S. Some chlorovirus particles are disrupted in CsCl and by freeze-thawing. Chloroviruses are generally stable in sucrose but if not, they are stable in iodixanol. Non-ionic detergents do not affect infectivity, whereas organic solvents do.

The MpoV prasinoviruses, infecting *Micromonas polaris*, are isolated on temperate and Arctic host strains. Temperate MpVs (infecting *M. pusilla* and *M. commode*) remain infectious after freezing, whereas the Arctic MpoVs are sensitive to low temperature. Some prasinoviruses do not survive freezing-thawing in seawater as such, but do survive if a cryoprotectant such as 10% glycerol or dimethyl sulfoxide is added. The Arctic MpoVs are highly sensitive to temperature shifts (0–7°C) comparable to that found during a season or with depth. MpoV-45T has highest infectivity at 3°C with over an 80% drop at lower and higher temperatures, whereas MpoV-46T has the same dynamics but with optimal infectivity at 0°C. The infectivity of MpoV-44T is highest at 7°C while that of MoV-47T is largely independent of temperature.

Nucleic Acids

The viruses currently listed in the ICTV as members of the *Phycodnaviridae* contain dsDNA genomes, ranging from 170 to ~ 560 kbp. However, as noted above some of the viruses with the largest genomes will eventually be transferred to the *Mimiviridae* family. Therefore a more realistic size for the genomes in the *Phycodnaviridae* family range from ~ 100 to 440 kbp. The G + C content of the phycodnavirus genomes ranges from 30% to 55%. Chlorovirus genomes contain methylated bases, both 5-methylcytosine (5mC) and N⁶-methyladenine (6mA). The percentage of methylated bases in chloroviruses ranges from 0.1% 5mC and no 6mA to 47% 5mC and 37% 6mA. The methylation status of the genomes of other algal viruses is unknown.

Proteins

The chlorovirus PBCV-1 virion has 148 different virus-encoded proteins and at least one host-encoded protein ranging in size from < 10 to > 200 kDa. PBCV-1 has many predicted membrane proteins and multiple capsid-like proteins (CP). In addition, the PBCV-1 virion has three glycoproteins, at least two myristylated proteins and several phosphoproteins. The PBCV-1 major CP, Vp54, consists of two eight-stranded, antiparallel- β barrel, “jelly-roll” domains connected by an α helix. A recent 3.5 Å resolution structure of the PBCV-1 virion identified 14 minor proteins associated with the virion structure.

Proteomic analysis of coccolithovirus EhV-86 indicated that it is composed of at least 28 virus-encoded proteins, 23 of which are predicted to be membrane proteins, one major CP, two lectin-binding proteins, a thioredoxin and a Ser/Thr protein kinase.

Lipids

The PBCV-1 virion contains 5%–10% lipid. The lipid is located in a single bilayered membrane located inside the glycoprotein shell and is required for virus infectivity. The coccolithovirus EhV-86 has an external lipid membrane and it probably also has an internal membrane. Glycosphingolipids (GSLs) comprise more than 80% of the EhV-86 lipidome.

Many of the prasinoviruses, including the Arctic MpoVs (infecting *Micromonas polaris*), are sensitive to chloroform, which suggests that they probably have a lipid membrane. However, other MpV isolates do not lose their infectivity after treatment with chloroform. Prasinovirus MpV-08T, infecting the temperate *Micromonas*, also contains a lipid membrane, which is enriched in intact polar lipids (IPLs) in comparison with its host (e.g., the proportion of phosphatidylglycerols was greatly reduced). Other IPLs were only found in MpV (i.e., digalactosyldiacylglycerol, sulfoquinovosyldiacylglycerol and diacylglycerylhydroxymethyltrimethyl-alanine), implying that these are produced by altering IPL-bound fatty acids during viral replication.

Carbohydrates

At least three of the chlorovirus PBCV-1 proteins are glycosylated including the major CP Vp54. Vp54 has four Asn-linked glycans, which are unusual in many aspects: (1) the glycans are attached to Asn by a β -glucose linkage, which is rare in nature; (2) the asparagines are not located in the typical Asn-X-Thr/Ser consensus sequon; (3) the glycans are highly branched and consist of 9–10 neutral monosaccharides; (4) all four glycoforms contain a dimethylated rhamnose as the capping residue of the main chain, a hyper-branched fucose residue and two rhamnose residues with opposite absolute configurations; and (5) these glycans do not resemble any structures previously reported in the three Domains of Life. Unlike other glycoprotein-containing viruses, PBCV-1 encodes most, if not all, of the machinery required to glycosylate its major CP.

Genomes

Many complete or nearly complete draft genome sequences are currently available for members of the *Phycodnaviridae*, including about 45 chloroviruses, 14 coccolithoviruses, 4 phaeoviruses and 19 prasinoviruses. There is considerable variation in genome structure among members of the *Phycodnaviridae*.

The chloroviruses have linear and nonpermuted genomes that are 290–370 kb in length, which code for 330–416 predicted proteins (coding sequences, CDSs). However, in total ~640 unique CDSs are predicted to be coded by 41 chloroviruses and because any one virus encodes no more than 416 CDSs, the chloroviruses exhibit extensive genetic diversity. Furthermore, some of the chlorovirus genes encode proteins with as many as 3 distinct domains that serve different functions, which contributes to their genetic diversity. Therefore, the genetic information coded by the chloroviruses is larger than the number of genes.

The termini of the 331 kb chlorovirus PBCV-1 genome consist of 35 nucleotide-long covalently closed hairpin loops that exist in one of two forms (flip and flop); the two forms are complementary when the 35-nucleotide sequences are inverted. Each hairpin loop is followed by an identical 2.2 kb inverted repeat sequence; the remainder of the genome consists primarily of single-copy DNA. The PBCV-1 genome has 416 CDSs that have 40 or more codons. Many of the core protein genes are in conserved clusters and are referred to as gene gangs. The coding regions of some of the genes slightly overlap and early and late genes are dispersed throughout the PBCV-1 genome, although there is some clustering of the early virus genes. The CDSs are evenly distributed on both strands and intergenic space is minimal. One exception is a 1788 bp sequence near the middle of the genome, which is a polycistronic gene containing 11 tRNAs.

Marine prasinoviruses with known host unicellular green algal species in the class *Mamiellophyceae* have G + C contents that range from 35% to 50% with 203–269 compactly arranged CDSs (with an average intergenic distance of about 40 bp), 4–11 tRNAs, and terminal inverted repeats of 250 to 2150 bp. Among these prasinoviruses, viruses infecting *Bathycoccus* sp. have the fewest CDSs, but they each encode 2 very large predicted proteins (ranging from 3126 to 5689 amino acids) of unknown function that make up to 14% of the genome. About half of the sequenced *Ostreococcus lucimarinus* viruses carry a central inversion of 32 kbp. Remarkably, the overall diversity of predicted proteins in marine prasinoviruses is less than that observed between the genomes of their host genera, suggesting that these viruses may have evolved by host-switching, a hypothesis supported by the incongruence of host and viral phylogenies. Four genomes of freshwater viruses resembling prasinoviruses both in their genome structures and their phylogenetic positioning (but with unknown host species), have been assembled from metagenomic samples taken from lakes in North America and a coastal lake in China. They have genomes of 171–181 kb encoding 225–248 predicted CDSs.

Members of the genus *Coccolithovirus* have large dsDNA genomes (376–421 kbp), 444–548 predicted CDSs, and a G + C content of 40%. These viruses also encode up to 6 tRNAs and the middle of their genomes is characterized by a large section of repeats, inversions and insertions. Complete genome and DNA polymerase-based phylogenetic analyses cluster EhV isolates into distinct sub-clades, based on geographical region and time of isolation. Importantly, EhVs harbor genes for a near-complete sphingolipid biosynthesis pathway, which are homologous to those found in the host genome (see below). Although most of the predicted EhV genes have no assigned function, and most resemble genes in the Domain Eukarya (and by extension, their unicellular algal host), more than 15% are most similar to bacterial genes.

The definitive structure of the 335 kbp phaeovirus EsV genome is unknown. Several experiments suggested that the genome is circular; however, DNA sequencing indicates that it has defined ends with inverted repeats. The virus is predicted to encode 231 CDSs, 48% of which resemble proteins in the public databases. About 12% of the EsV genome consists of tandem repeats and

portions of the genome have ssDNA regions. The genome of the HaV53 strain (274,793-bp in length) contains 246 CDSs and 3 tRNA-coding sequences.

Virus Replication

Chlorovirus PBCV-1 attaches rapidly and specifically to the external surface of cell walls, but not to protoplasts, of its host *Chlorella variabilis* NC64A. The PBCV-1 virion spike structure makes the first contact with the host cell wall (Fig. 1(F)). The virus spike is too narrow to deliver DNA and likely serves to puncture the wall before it is jettisoned. Following host cell wall degradation by a virus-associated enzyme(s) (Fig. 1(G)), the virus internal membrane fuses with the host membrane, forming a membrane-lined tunnel between the virus and the host (Fig. 1(H)), and leaving an empty capsid attached to the surface. Interestingly, the membrane-lined tunnel between the virus and host is so narrow that the compressed DNA (ca. 0.2 bp nm⁻³) in the virion must pass into the host in a linear manner and then once inside the host it appears to condense again. The virus-host membrane fusion process triggers rapid depolarization of the host plasma membrane, probably initiated by a virus encoded K⁺ channel located in the virus internal membrane. This hypothesis is supported by the fact that infection by the chloroviruses is inhibited by the K⁺ channel blockers barium and/or cesium. The depolarization results in the release of K⁺ from the cell. This rapid loss of K⁺ from the host and associated water fluxes reduce the host turgor pressure, which aids ejection of viral DNA and virion-associated proteins into the host. Host membrane depolarization also inhibits many host secondary transporters and prevents infection by a second virus.

None of the chloroviruses encode a recognizable DNA-dependent RNA polymerase (DdRp) gene, so it is assumed that PBCV-1 DNA and viral-associated proteins quickly move to the nucleus and commandeer the host transcription machinery. Early transcription begins 5–10 min post-infection (PI). Viral DNA synthesis begins at 60–90 min PI and presumably starts in the nucleus before moving to the cytoplasm. At 3–4 h PI, assembly of PBCV-1 capsomers begins in localized regions of the cytoplasm, which become prominent 4–5 h PI (Fig. 1(H)). These localized regions, called virus assembly centers or virus factories, consist of host cisternae that are derived from the endoplasmic reticulum next to the nuclear membrane (Fig. 1(I)). The cisternae are localized at the periphery of the viral assembly centers and are cleaved into single bilayered membranes, which then move to the central region of the assembly centers. Capsomers form around these membranes leading to the formation of empty virions at the periphery of the virus assembly centers where DNA packaging occurs. By 5–6 h PI the cytoplasm fills with infectious progeny viruses and localized lysis of the host cell membrane and cell wall releases progeny at 6–8 h PI (Fig. 1(J)). Each infected algal cell releases ~1000 newly formed virus particles, of which ~25% form plaques. Mechanical disruption of cells releases infectious virus particles 30–50 min prior to cell lysis, indicating the virus is mature and does not acquire its glycoprotein capsid by budding through the host plasma membrane as it is released from the cell.

The coccolithovirus EhV virion has an external lipid membrane and the virus encodes 6 RNA polymerase subunits, none of which are packaged in the virion. EhV enters the host cell via either endocytosis or an envelope fusion mechanism by interacting with lipid rafts on the host membrane, which are used as focal points for attachment and entry. Upon entry, EhV internalization and breakdown of the virion capsid takes place instantaneously, quickly followed by the release of the virion genome into the nucleus. There, early virus transcription occurs using host RNA polymerases, followed by transcription of the rest of the viral genes in the cytoplasm, where capsid assembly and viral DNA synthesis occurs. The newly assembled viruses are transported to the plasma membrane where they acquire the inner lipid layer by association with intracellular lipid bodies, and they are released from the cell by budding through lipid rafts where they acquire a GSL rich outer membrane. A large proportion of newly-formed viruses are also released by eventually lysing the host cell.

The EhV latent period can be as short as 4–8 h PI and EhV burst sizes range from 50 to 1000 new particles per lysed cell. The number of infectious EhV progeny can be as high as 30% or as low as 1%, depending on the type of EhV and host (e.g., calcifying or not). Prymnesioviruses and prasinoviruses have somewhat longer latent periods than the coccolithoviruses. Burst size for the prymnesioviruses is between 275 and 410 and up to 330 per lysed host cell for prasinovirus MpV.

Host responses to EhV infection are diverse and can often be used as early markers of infection. For example, EhV infection triggers a series of metabolic events, including host lipid and fatty acid remodeling that includes the production of viral encoded GSLs that are essential for infection, differential induction of nitric oxide and reactive oxygen species, induction of caspase activity, enhanced metacaspase expression, activation of program cell death and autophagy pathways, and impairment of the host photosystem apparatus. Furthermore, EhV infection results in the excess production of “sticky” extracellular polysaccharides, which are exuded from infected cells and facilitate aggregation of cellular content and debris into larger organic carbon-containing particles.

Phaeovirus EsV initiates its life-cycle by infecting free-swimming, wall-less gametes of its host. Virus particles enter the cell by fusion with the host plasma membrane and release a nucleoprotein core particle into the cytoplasm, leaving remnants of the capsid on the surface. The viral core moves to the nucleus within 5 min PI. One important feature that distinguishes the EsV life-cycle from the other phycodnaviruses is that the viral DNA becomes integrated into the host genome and is transmitted mitotically to all cells of the developing alga. The viral genome remains latent in vegetative cells until it is expressed in the algal reproductive cells, the sporangia or gametangia. Massive viral DNA replication occurs in the nuclei of these reproductive cells, followed by nuclear breakdown and viral assembly that continues until the cell becomes densely packed with virus particles. Virus release is stimulated by the same factors that induce discharge of gametes from the host, i.e., changes in temperature, light, and water composition. This synchronization facilitates interaction of viruses with their susceptible host cells.

Less is known about replication of the prasinoviruses and raphidoviruses, although formation of the raphidovirus HaV occurs in the cytoplasm and the nucleus remains intact and separate from the viroplasm that consists of a fibrillar matrix. Ultimately, viral production results in the disruption of organelles, lysis of the cell and release of the virus particles. Unlike the chloroviruses, which are very host specific, some of the raphidoviruses can infect a range of host isolates within specific algal species; however, there is no evidence that they cross the species barrier.

Virus Transcription

After entry of the chlorovirus PBCV-1 DNA into the cell its DNA and viral-associated proteins quickly move to the nucleus and commandeer the host transcription machinery. Early transcription of PBCV-1 begins 5–10 min PI. In this immediate-early phase of infection, host transcription rates decrease and the host transcription machinery is reprogrammed to transcribe viral DNA. Some early viral transcripts are synthesized in the absence of protein synthesis. Details of reprogramming are unknown but host chromatin remodeling is probably involved because PBCV-1 encodes and packages an enzyme that methylates Lys-27 in histone 3. Circumstantial evidence indicates that this methylation represses host transcription following PBCV-1 infection. In addition, host chromosomal DNA degradation begins within 5 min PI, presumably by PBCV-1 encoded and packaged DNA restriction endonucleases, which also aids in inhibition of host transcription and release of the host transcription machinery for the benefit of the virus.

PBCV-1 dominates the infected cells in the early phase prior to initiation of viral DNA synthesis, which begins at 60–90 min PI. Microarray experiments indicated that 63% of the PBCV-1 genes are expressed early, whereas late genes are expressed (after 60 min PI) until the end of virus replication; 43% of the genes are expressed both early and late. Remarkably, an RNA-seq based transcription profile revealed that ~50 PBCV-1 genes are expressed within the first 7 min PI. By 60 min PI, essentially all of the PBCV-1 genes are expressed at some level and ~40% of the poly (A +) containing RNAs in the infected cell are PBCV-1 transcripts. As expected, the synthesis of late transcripts requires translation of early virus genes.

Consensus promoter regions for early and late PBCV-1 genes have not been identified definitively; however, the sequence AATGACA is common in the 100 nucleotides preceding the ATG start codon of most early PBCV-1 genes. Furthermore, the 50 nucleotides preceding the ATG start codons are usually >70% A + T. Transcription of some PBCV-1 genes appears to be complex. For example, some gene transcripts exist as multiple bands and these patterns can change between early and late times in the virus life-cycle.

Transcription of prasinovirus genomes has been analyzed by RNA-Seq in *Micromonas pusilla* infected with MpV-SP1 in continuous light or in *Ostreococcus tauri* infected with OtV5 in a 12 h day 12 h night growth regime. In both cases, the availability of complete host and viral genomes allowed the expression of all host and viral genes to be monitored over the viral life-cycles. In *M. pusilla* the continuous light experiments focused on comparing the effects of different nutrient phosphate levels, whereas in *O. tauri* a day/night cycle was used to mimic natural growth conditions because host gene expression is influenced strongly by light and the cell cycle. When *M. pusilla* growth was limited by 10-fold less phosphate, only 30% of host cells were lysed by MpV-SP1 virus at 24 h PI, compared to 70% lysis in controls; about 5.8% of the differentially regulated genes were due to the viral infection. In *O. tauri* the host cell growth was mainly synchronous, with S-phase (DNA synthesis) occurring in late afternoon-early night and cell division in early night. Under these conditions, prasinovirus transcripts were much less abundant in the day and host cell lysis occurred mainly during the following day. During the night, about 4.4% of host genes showed increased expression relative to uninfected controls, likely representing both genes de-repressed for viral virulence and induced for host defense. Practically all of MpV-SP1 and OtV5 genes were expressed during infections, except for a few CDSs residing in the terminal inverted repeat regions.

Coccolithovirus EhV transcription also begins in the host nucleus using the host DdRp even though EhV encodes 6 RNA polymerase subunits; however, none of them are packaged in the virus. Transcriptional analysis suggests that viral RNA polymerase gene expression occurs in a secondary transcriptional phase (occurring at least one hour PI and demarcated by a transition from a unique-promoter driven, locus-specific transcriptional activity to a genome wide transcriptional activity), potentially indicative of a biphasic replication cycle with both nuclear and cytoplasmic phases.

Ecology

Eukaryotic algae are important components of both inland water and marine environments; however, the significance of viruses in these systems is only beginning to be appreciated. The diversity of marine prasinoviruses was first noted by using the DNA polymerase gene found in all NCLDV as a genetic marker. Using degenerate oligonucleotide primers and subsequent sequencing revealed viruses that were likely to infect the microalga *Micromonas*. Subsequently, diverse viruses infecting the species *Ostreococcus tauri* and *Bathycoccus* sp. (also members of the algal class *Mamiellophyceae*) were identified in coastal samples from the North Sea and Mediterranean Sea. The availability of several prasinovirus genomes permitted screening of very large metagenomic datasets, such as those produced by the *Tara-Oceans* expedition, for the presence of prasinoviruses and showed them to be one of the most abundant types of NCLDV in marine waters. A new degenerate primer pair, in combination with next generation sequencing of PCR products, revealed enormous viral diversity in different marine samples. The viral diversity and distribution are dependent on environmental conditions affecting host growth, e.g., nutrient and light availability and temperature.

Controlled one-step infection experiments with prasinoviruses demonstrated that nutrient limitation typically results in longer latent periods, reduced burst sizes and delayed lysis of the host as compared to control treatments. For example, the burst size of MpV-08T was reduced by 70% under phosphate limitation as well as under nitrogen and iron limitation. *Micromonas* growth is not strongly affected by light intensity and this may be the reason that MpV proliferation is also not impaired by low or high light, independent of whether the host is phosphorus limited or not. Likely MpV proliferation is not dependent on photophosphorylation during infection, similar to chlorovirus production in *Chlorella*. Besides nutrients and light, temperature is also an important host growth-regulating variable and indeed in the Arctic it impacts MpoV viral burst size and progeny virus infectivity.

Phycodnaviruses can play major roles in the termination of marine algal blooms and research efforts are actively focusing on elucidating the natural history of these algal/virus systems. For example, EhV and PgV viruses are one of the primary mechanisms for terminating *Emiliania huxleyi* and *Phaeocystis globosa* blooms, respectively. EhV infection and replication is affected by light, and newly formed virus particles can be transmitted to surrounding regions by 'hitch-hiking' on zooplankton and by marine aerosols. Likewise, infection by the rhabdovirus HaV is regarded as one of the factors influencing the dynamics and termination of *Heterosigma akashiwo* blooms. HaV has a significant impact on *H. akashiwo* populations and changes clonal composition of the alga. Based on host-virus cross reactivity tests, HaV isolates were divided into three groups (I-III) according to their intraspecies host range.

Chloroviruses are ubiquitous in inland water collected throughout the world and titers as high as thousands of infectious particles per ml have occasionally been reported. However, typically the titer is 1–100 infectious particles per ml. The titers are seasonal with the highest titers in the spring and late fall as monitored in a lake in the temperate zone. It is not known if chloroviruses replicate exclusively in zoochlorellae symbiotic with paramecia and heliozoae or if the viruses have another host (s). In fact, it is not known if these zoochlorellae exist free of their hosts in natural environments. However, zoochlorellae contained within their symbiotic hosts are refractory to infection by chloroviruses because the viruses cannot contact the algae.

In the case of the ciliate *Paramecium bursaria*, chloroviruses can attach to the outer membrane of the paramecia without infecting them, which puts the viruses in good position to encounter zoochlorellae if the paramecium is disrupted by a feeding predator. For example, copepods feed on paramecia by engulfing them and pass viable zoochlorellae through their digestive system intact, which leads to virus infection and a rapid increase in viral titer. This gives rise to the concept of predation being an ecological catalyst. Other predators such as *Didinium* can forage on zoochlorellae-bearing paramecia and rupture the cell, which results in the release of zoochlorellae where they can be infected.

The lysogenic phaeoviruses EsV and FsV infect the small filamentous brown algae *Ectocarpus* sp and *Feldmania* sp, respectively. Lysogeny is consistent with the observation by early investigators that VLPs appeared infrequently in eukaryotic algae and only at certain stages of algal development. The apparent lack of infectivity by many of these previously observed VLPs in eukaryotic algae is also consistent with a lysogenic lifestyle. The VLPs might either infect the host and resume a lysogenic relationship or be excluded by pre-existing lysogenic viruses. The ultimate role of the viruses in the ecology of their host is still unknown.

Lysogenic phaeoviruses have also recently been reported to infect the large perennial *Laminaria* macroalgae (kelp). Kelp form underwater forests that dominate coastlines on all continents and form some of the most productive and structurally complex ecosystems in the world, which support diverse marine communities. The role that the phaeoviruses play in kelp ecology is completely unknown.

To summarize, phycodnaviruses contribute to microbial composition and diversity, nutrient cycling, carbon flow, and other bio-geochemically-important processes in aqueous environments.

Resistance to Phycodnavirus Infections

Resistance to virus infection has been studied to some degree for members of three genera within the family Phycodnaviridae. In the case of the chloroviruses, resistance to virus infection occurs quite frequently and most of the time this is due to a change in the host receptor such that the chloroviruses are unable to attach to the host.

Coccolithoviruses infect the marine microalga, *Emiliania huxleyi*. The alga has different life forms including: (1) diploid coccolith-bearing or naked, non-motile cells that are susceptible to the virus, (2) diploid coccolith-bearing or naked, non-motile cells that are resistant to the virus; and (3) haploid, non-calcifying, organic scale-bearing swarming cells that were originally reported to be resistant to virus infection. However, recent evidence indicates that virus-derived lipids are present in *E. huxleyi* haploid cells, indicating that the cells are actually infected. In addition, virus transcripts are detected in these cells in the absence of cell lysis due to virus infection. This superimposed state of being both infected and resistant is reminiscent of Schrödinger's cat; i.e., of being simultaneously both dead and alive.

In the green microalga *Ostreococcus tauri*, resistance to prasinovirus infection arises frequently in culture and in *Micromonas* sp. the frequency of spontaneous resistance is temperature-dependent. RNA-Seq transcription analysis of many independent OtV5-resistant clonal lines obtained from OtV5-infected cultures revealed overexpression of all genes spanning half of the physical length of chromosome 19, and physical rearrangements of this chromosome are also observed in karyotypic analyses of these lines. This chromosome, which carries many predicted glycosyltransferase genes, is variable in its genetic structure in different wild-type algal strains, and appears to be involved in viral defense.

Phycodnavirus Genes Encode Some Interesting and Unexpected Proteins

Many chlorovirus-encoded proteins are either the smallest or among the smallest proteins of their class. In addition, homologous genes in the chloroviruses can differ in nucleotide sequence by as much as 50%, which translates into amino acid differences of 30%–40%. Therefore, comparative protein sequence analyses can identify conserved amino acids in proteins as well as regions that tolerate amino acid changes. The small sizes and the finding that many chlorovirus-encoded proteins are 'user friendly' have resulted in the biochemical and structural characterization of many PBCV-1 encoded proteins. Examples include: (1) 5mC and 6mA DNA methyltransferases and their companion DNA site-specific endonucleases (some of which are sold commercially); (2) the smallest eukaryotic ATP-dependent DNA ligase, which is also sold commercially; (3) the smallest type II DNA topoisomerase (the virus enzyme cleaves dsDNAs ~50 times faster than the human type II DNA topoisomerase); (4) a small prolyl-4-hydroxylase that converts Pro-containing peptides into hydroxyl-Pro-containing peptides in a sequence-specific fashion; and (5) one of the smallest proteins (82 amino acids) to form a functional K^+ channel. In fact, the prasinovirus MpV-12T encodes a functional 78 amino acid K^+ channel protein. These virus-encoded K^+ channels are under intensive investigation and there are over 60 publications on them.

The chloroviruses are also unusual because they encode enzymes involved in sugar metabolism. Three PBCV-1-encoded enzymes, glutamine:fructose-6-phosphate aminotransferase, UDP-glucose dehydrogenase, and hyaluronan synthase are involved in the synthesis of hyaluronan, a linear polysaccharide composed of alternating β -1,4-glucuronic acid and β -1,3-N-acetylglucosamine residues. All three genes are transcribed early in PBCV-1 infection and hyaluronan accumulates on the external surface of the infected cells. Other chloroviruses encode a chitin synthase and chitin accumulates on the external surface of these infected cells. Two PBCV-1 encoded enzymes, GDP-D-mannose dehydratase and fucose synthase, comprise a three-step pathway that converts GDP-D-mannose to GDP-L-fucose and GDP-L-rhamnose. Both of these monosaccharides are in the glycans attached to the virus major CP. As noted earlier, PBCV-1 encodes most if not all of the machinery to glycosylate its major CP including several glycosyltransferases.

PBCV-1 is the first virus known to encode 5 functional enzymes involved in polyamine biosynthesis and catabolism including: ornithine decarboxylase (ODC), homospermidine synthase, agmatine iminohydrolase, N-carbamoyl-putrescine amidohydrolase and a polyamine acetyltransferase. ODC catalyzes the decarboxylation of ornithine to putrescine, which is the first and the rate limiting enzymatic step in polyamine biosynthesis. Not only is the PBCV-1-encoded ODC the smallest known ODC, the PBCV-1 enzyme is also interesting because it decarboxylates arginine more efficiently than ornithine. Homospermidine synthase is a virion-associated protein.

The chloroviruses also encode many other interesting and unexpected enzymes that have been characterized including an aquaglyceroporin, a Ca^{2+} transporting ATPase, a K^+ transporter, a pyrimidine dimer specific glycosylase, an aspartate transcarbamylase, and a thymidylate synthase X.

Another set of unexpected proteins is encoded by the EhV coccolithoviruses. These include enzymes that comprise a near-complete *de novo* sphingolipid biosynthesis pathway believed to have been acquired from the host via horizontal gene transfer (HGT). Importantly in eukaryotes, sphingolipids are common building blocks of membrane lipids and lipid rafts. The first, and likely rate-limiting enzyme in this pathway is serine palmitoyltransferase (SPT). Its inhibition by myriocin (a sphingosine equivalent which is a specific and effective inhibitor of SPT in yeast and mammalian cells) halts successful EhV replication in a dose dependent manner. Other putative proteins coded by the different EhVs include: a phosphate permease, polyubiquitin, and a sialidase.

In contrast to the chloroviruses, the gene content of many prasinoviruses has only come to light in the last decade and very few of the predicted gene functions have been experimentally verified. Exceptions to this are certain cation transporters, whose function has been tested by heterologous expression in yeast or mammalian cells. Some prasinoviruses encode phosphate or ammonium transporters that are thought to confer nutritional advantages to their host cells during infection. These, and numerous other genes are likely to have been acquired by HGT, either from the host or from other microbes. About 30% of the prasinovirus genes show similarity to genes with known functions, and include a diversity of interesting genes whose number increases with every new virus sequenced. All prasinoviruses encode methylases and glycosyltransferases that are likely involved in modifications of host or viral DNA or proteins important for the viral life-cycle. Many *Ostreococcus* or *Micromonas* viruses encode putative enzymes for amino acid synthesis (for example 3-dehydroquinate synthase, acetolactate synthase, oxovalerate aldolase) or sugar synthesis (for example 6-phosphofructokinase). Near-complete genomes of some prasinoviruses with unknown host species have been assembled from brackish or freshwater lakes, and encode genes for a multidrug resistance protein and for histone H3, which is likely to be involved either in host or viral gene expression and/or virus DNA packaging.

Perspectives

Eukaryotic algae are important components of both inland water and marine environments. However, the significance of viruses on host population dynamics in these systems is only beginning to be appreciated. It is obvious that identifying and characterizing large dsDNA viruses that infect eukaryotic algae is in its infancy. For example, metagenomic studies, such as DNA sequencing in the Sargasso Sea, indicate that many algal viruses exist in a variety of aqueous environments with unknown hosts. This lack of knowledge certainly carries over to soil samples. Furthermore, with a few possible exceptions, little is known about the replication

cycles of most algal viruses. The few algal-virus systems that are under investigation indicate that the various systems differ extensively, e.g., in their infection processes, in their replication cycles, consequences of infection (lytic versus lysogenic), and probably in many other aspects. The development of efficient systems that allow scientists to conduct molecular manipulations of the algal virus genomes would provide a huge boost to the study of algal viruses. Such techniques are currently lacking.

Finally, it should be noted that algal viruses encode many interesting proteins, and even some that have potential for commercial exploitation. Unfortunately, it is also true that most of the putative proteins in these viruses have unknown functions and they often have no homology to proteins in the public domain databases. Who knows how many more unexpected algal virus encoded proteins await discovery?

Acknowledgments

We want to thank all of the many researchers who have contributed to the growing field of algal virology, especially close colleagues for useful discussions, and those whose work has not been mentioned because of space limitations.

Further Reading

- Bidle, K.D., 2016. Programmed cell death in unicellular phytoplankton. *Current Biology* 26, R594–R607.
- Claverie, J.M., Abergel, C., 2018. *Mimiviridae*: An expanding family of highly diverse large dsDNA viruses infecting a wide phylogenetic range of aquatic eukaryotes. *Viruses* 10, 506.
- Derelle, E., Yau, S., Moreau, H., Grimsley, N.H., 2018. Prasinovirus attack of *Ostreococcus* is furtive by day but savage by night. *Journal of Virology* 92, e01703–e01717.
- Maat, D.S., Biggs, T., Evans, C., *et al.*, 2017. Characterization and temperature dependence of Arctic *Micromonas polaris* viruses. *Viruses* 9, 134–154.
- Martinez Martinez, J., Boere, A., Gilg, I., *et al.*, 2015. New lipid envelope containing dsDNA virus isolates infecting *Micromonas pusilla* reveal a separate phylogenetic group. *Aquatic Microbial Ecology* 74, 17–28.
- Nissimov, J.I., Pagarete, A., Ma, F., *et al.*, 2017. Coccolithoviruses: A review of cross-kingdom genomic thievery and metabolic thuggery. *Viruses* 9, 52.
- Schroeder, D.C., 2015. More to Phaeovirus infections than first meets the eye. *Perspectives in Phycology* 2, 105–109.
- Van Etten, J.L., Agarkova, I.V., Dunigan, D.D., 2019. Chloroviruses. *Viruses* 12, E20.
- Van Etten, J.L., Dunigan, D.D., 2015. Voyages with chloroviruses. In: Rohwer, F. (Ed.), *The Phage World*. USDA, pp. 326–337.
- Van Etten, J.L., Dunigan, D.D., 2016. Giant chloroviruses – Five easy questions. *PLoS Pathogens* 12, e1005751.
- Weynberg, K.D., Allen, M.J., Wilson, W.H., 2017. Marine Prasinoviruses and their tiny plankton hosts: A review. *Viruses* 9, E43.
- Yau, S., Seth-Pasricha, M., 2019. Viruses of polar aquatic environments. *Viruses* 11, 189.

INVERTEBRATE VIRUSES

An Introduction to Viruses of Invertebrates

Peter J Krell, Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

BV Budded virus (baculoviruses)

G Virion glycoprotein, structural protein

gRNA Genomic (or virion) RNA

GV Granulovirus

ICTV International Committee on Taxonomy of Viruses

IRES Internal ribosome entry site on ssRNAs

L Long segment of a tripartite RNA genome

M Medium segment of a tripartite RNA genome

N Virion nucleoprotein, structural protein of core

NPV Nucleopolyhedrovirus

NS Nonstructural protein

ODV Occlusion derived virus (baculoviruses)

P Virion phosphoprotein, structural protein

RdRp RNA dependent RNA polymerase (of RNA viruses)

S Short segment of trisegmented RNA genome

sgRNA Subgenomic RNA especially for *Nidovirales* viruses

TRS Transcription regulating sequence (*Nidovirales* viruses)

+ve ssRNA positive sense (mRNA sense) single stranded RNA

–ve ssRNA antisense (anti mRNA sense) single stranded RNA

VP Virion protein

Glossary

Granulovirus It is a term used for insect baculoviruses which produce granules composed of granulin each containing a single virion and which protects the virus from the environment.

Internal ribosome entry site (IRES) It is a sequence in a mRNA which allows for cap-independent translation. This is common to the *Picornavirales* order but occurs in the RNAs of others as well.

International Committee on Taxonomy of Viruses It is an international committee which has a defined organizational structure which meets on an annual basis to review proposals on new taxonomies, or changes in taxonomies. Once proposals are formally accepted the committee updates the taxonomy and posts it to the ICTV Website at “See Relevant Websites section”.

Nucleopolyhedrovirus It is a term used for the insect baculoviruses which produce large polyhedral particles

composed of polyhedrin in which virions are embedded to protect them from the environment.

RNA dependent RNA polymerase (RdRp) It is an RNA polymerase which synthesizes RNA copies from an RNA template strand. It is an enzyme which allows an RNA virus to transcribe its genome into antigenome RNA, viral mRNA or subgenomic mRNAs.

Subgenomic RNAs (sgRNAs) They are smaller segments of RNA produced from a long RNA template with each segment encoding a different ORF, or in the case of the *Nidovirales* different combinations of ORFs.

Transcription regulating sequence (TRS) It is a short sequence in viral RNA at the beginning of the first ORF or between subsequent ORFs, especially for viruses in the *Nidovirales* order, that allow transcription of subgenomic mRNAs with each sgRNA being translated into a different ORF.

Introduction

Considering that invertebrate species outnumber the total of all other global species by about 9 to 1 and, other than bacteria, have a longer evolutionary history, it is not surprising that they harbor the majority of virus species on earth. Although most people think of invertebrates as arthropods, including insects and spiders, they also include protozoans, annelids, echinoderms and molluscs. For the sake of simplicity and acknowledging the paucity of information on viruses in other groups of invertebrates, this review will focus mostly on viruses of arthropods, specifically insects, arachnids and crustaceans and, due to their economic importance, molluscs.

While many viruses, notably the arboviruses of mammals and plants, infect both their invertebrate vectors and vertebrate or plant hosts, the rest are restricted to individual species of animals. [Table 1](#) summarizes the different species of invertebrate viruses based on the Family taxon. For the sake of simplicity names of only taxa from the Family level (and Order) and below are given. Higher orders of Class, Subphylum, Phylum, Kingdom and Realm, favored by the International Committee on the Taxonomy of Viruses (ICTV) may be of relevance to those interested in virus evolution and bioinformatics, but from a practical level have little value to most practicing virologists and are not included in these summaries. As one example for the complete taxonomy of a virus of invertebrates, the taxonomy of nonviruses would have to include all of Realm *Ribovairia*, Kingdom *Orthornavirae*, Phylum *Pisuviricota*, Subphylum none, Class *Pisonvircetes*, Order *Nidovirales*, Suborder *Ronidovirineae*, Family *Roniviridae*, Subfamily *Okanovirinae*, Genus *Okavirus*, Subgenus *Tipravirus*, Species *Gill-associated virus*. It would be too cumbersome to provide such a complete taxonomy for all viruses in this article and would probably be of little interest. Nevertheless those interested in the higher levels of classification ordained by the

Table 1 Families and Genera of Viruses Replicating in or Associated with Invertebrate Hosts

Family/Subfamily Order	Genus, Subgenus, Type Species (total # in genus)	Genome type, segments and size	Morphology, T value for icosahedral capsids	Major invertebrate host
<i>Alphatetraviridae</i> (<i>Hepelivirales</i>) ^a	<i>Betatetravirus</i> , <i>Nudaurelia capensis beta virus</i> (7) <i>Omegatetravirus</i> , <i>Nudaurelia capensis omega virus</i> (3)	Beta-monopartite, + v ssRNA, 6.6 kb Omega-bipartite 2.5/5.3	Icosahedral T=4, 42 nm diameter	lepidoptera
<i>Artoviridae</i> (<i>Mononegavirales</i>)	<i>Hexartovirus</i> , <i>Caligid hexartovirus</i> (2) <i>Peropuvirus</i> , <i>Pteromalus puparum peropuvirus</i> (6)	– ve ssRNA, monopartite, 12.5 kb	Spherical, flexuous nucleoprotein core, 100 to 130 nm enveloped	Wasps, barnacles, copepods, pill worms, woodlice, ticks
<i>Ascoviridae</i> (<i>Pimascovirales</i>)	<i>Ascovirus</i> , <i>Spodoptera frugiperda ascovirus 1a</i> (3) <i>Toursvirus</i> , <i>Diadromus pulchellus toursvirus</i> (1)	Circular dsDNA, 100 to 200 kb	Core 80 × 300 nm, enveloped 130 × 200 to 400 nm allantoid or reniform virion	Lepidoptera, parasitoid wasps
<i>Baculoviridae</i>	<i>Alphabaculovirus</i> , <i>Autographa californica multiple nucleopolyhedrovirus</i> (55) <i>Betabaculovirus</i> , <i>Cydia pomonella granulovirus</i> (26) <i>Deltabaculovirus</i> , <i>Culex nigripalpus nucleopolyhedrovirus</i> (1) <i>Gammabaculovirus</i> , <i>Neodiprion lecontei nucleopolyhedrovirus</i> (2)	Circular ds DNA, 82 to 180 kb	Budded virus helical nucleocapsids 30 to 60 nm × 250 to 300, enveloped Occlusion derived virus form (ODV) embedded in a polyhedral occlusion body NPV polyhedral 0.15 to 5.0 μM diameter GV polyhedral 0.12 to 0.5 μM diameter	Lepidoptera, hymenoptera diptera
<i>Bidnaviridae</i> (<i>Poliovirales</i>)	<i>Bidensovirus</i> , <i>Bombyx mori bidensovirus</i> (1)	Linear ssDNA, ambisense, bipartite, 6 and 6.5 kb	Icosahedral T=1, 20 to 24 nm, separate capsid for each of 4 ssRNAs	silkworm
<i>Birnaviridae</i>	<i>Entomobirnavirus</i> , <i>Drosophila X virus</i> (2)	dsRNA, bipartite, 3.2, 3.4 kb	Icosahedral T=3, 60 to 70 nm diameter	Drosophila, mosquitoes
<i>Carmotetraviridae</i> (<i>Tollivirales</i>)	<i>Alphacarmotetravirus</i> , <i>Providencia virus</i> (1)	+ ve ssRNA, monopartite, 6.2 kb	Icosahedral T=4, 40 nm diameter	lepidoptera
<i>Circoviridae</i> (<i>Cirivirales</i>)	<i>Circovirus</i> , <i>Porcine circovirus 1</i> (43) <i>Cyclovirus</i> , <i>Human associated cyclovirus 8</i> (51)	ssDNA, circular, 1.7 to 1.9 kb	Icosahedral T=1, 15 to 25 nm diameter	Cockroach, dragonfly, mosquito, ants, spiders, ticks
<i>Dicistroviridae</i> (<i>Picornavirales</i>)	<i>Aparavirus</i> , <i>Acute bee paralysis virus</i> (6) <i>Cripavirus</i> , <i>Cricket paralysis virus</i> (4) <i>Triatovirus</i> , <i>Triatoma virus</i> (5)	+ ve ssRNA, monopartite 8.0 to 10.2 kb,	Icosahedral, pseudo T=3, 30 nm diameter	Bees, ants, aphids, drosophila, triatomine bugs
<i>Genomoviridae</i> (<i>Geplafuvirales</i>)	<i>Gemycircularvirus</i> , <i>Sclerotinia gemycircularvirus 1</i> (43) <i>Gemyduguivirus</i> , <i>Dragonfly associated gemyduguivirus 1</i> (1) <i>Gemykibivirus</i> , <i>Dragonfly associated gemykibivirus 1</i> (16)	ssDNA, circular, 2.2 kb	Icosahedral T=1, 29 nm diameter	Dragonfly, mosquito
<i>Hytrosaviridae</i>	<i>Glossinavirus</i> , <i>Glossina hytrosavirus</i> (1) <i>Muscavirus</i> , <i>Musca hytrosavirus</i> (1)	dsDNA, circular, 124 to 190 kb	Rod shaped, 50 to 65 nm × 500 to 1,000 nm, helical core, enveloped	Diptera and perhaps hymenoptera
<i>Iflaviridae</i> (<i>Picornavirales</i>)	<i>Iflavirus</i> , <i>Infectious flacherie virus</i> (15)	+ ve ssRNA, linear, monopartite 8-10.5 kb, VPg at 5' end, poly A at 3' end	Icosahedral, pseudo T=3, 22 to 30 nm diameter	Honey bees, parasitic mites, lepidoptera, aphids, planthoppers
<i>Iridoviridae</i> <i>Betairidovirinae</i> (<i>Pimascovirales</i>)	<i>Chloriridovirus</i> , <i>Invertebrate iridescent virus 3</i> (5) <i>Decapodiridovirus 1</i> , <i>Decapod iridescent virus 1</i> (1) <i>Iridovirus</i> , <i>Invertebrate iridescent virus 6</i> (2)	dsDNA, linear 163 to 220 kb, circularly permuted terminally redundant, methylated	Icosahedral T=147 to 217, 150 to 200 nm diameter, inner membrane, naked or enveloped	Mosquitoes, lepidoptera, isopods, crabs, shrimp
<i>Lispiviridae</i> (<i>Mononegavirales</i>)	<i>Arlivirus</i> , <i>Lishi arlivirus</i> (6)	– ve ssRNA, linear, bisegmented, 7.1, 4.4 kb	Unknown, identified solely from sequence data	Spiders
<i>Malacoherpesviridae</i> (<i>Herpesvirales</i>)	<i>Aurivirus</i> , <i>Haliotid herpesvirus 1</i> (1) <i>Ostreavirus</i> , <i>Ostreid herpesvirus 1</i> (1)	dsDNA, linear, 134 to 212 kb	Icosahedral T=16 nucleocapsid, spherical 150 to 200 nm diameter virion, enveloped	Molluscs
<i>Mesoniviridae</i> <i>Hexponivirinae</i> (<i>Nidovirales</i>)	<i>Alphamesonivirus</i> (10) <i>Subgenus Namcalivirus</i> , <i>Alphamesonivirus 1</i> (2 species)	+ ve ssRNA, linear, monopartite, 20 kb, 5' cap 3' polyA tail	Spherical, 50 to 120 nm enveloped with spike protein	Mosquito

<i>Nimaviridae</i>	<i>Whispovirus</i> , <i>White spot syndrome virus</i> (1)	dsDNA, circular, 280 to 307 kb	Ellipsoid, nucleocapsid 70 to 170 nm × 210 to 420 nm, with tegument and envelope	Crustaceans, shrimp
<i>Nodaviridae</i> (<i>Nodamuvirales</i>)	<i>Alphanodavirus</i> , <i>Nodamura virus</i> (5)	+ ve RNA, linear, bipartite, 1.4 and 3.1 kb	Icosahedral T=3, 25 to 33 nm diameter	coleoptera
<i>Nudiviridae</i>	<i>Alphnudivirus</i> , <i>Oryctes rhinoceros nudivirus</i> (2) <i>Betanudivirus</i> , <i>Heliothis zea nudivirus</i> (1)	dsDNA, circular, 96 to 232 kb	Nucleocapsid rod-shaped helical core, enveloped virus 81 to 100 × 200 to 415 nm	Beetles, lepidoptera, crustaceans
<i>Nyamiviridae</i> (<i>Mononegavirales</i>)	<i>Berhavirus</i> , <i>Sipunculid berhavirus</i> (3) <i>Crustavirus</i> , <i>Wenzhou crustavirus</i> (3) <i>Nyavirus</i> , <i>Nyamanini nyavirus</i> (3) <i>Orinovirus</i> , <i>Orinoco orinovirus</i> (1) <i>Socyvirus</i> , <i>Soybean cyst nematode socyvirus</i> (1) <i>Tapwovirus</i> , <i>Tapeworm tapwovirus</i> (1)	– ve ssRNA, linear, monopartite, 12.2 kb (bipartite for <i>Tapwovirus</i>)	Enveloped spherical particles, 100 to 130 nm diameter	Ticks, insects, echinoderms, spiniculids, crustaceans, nematodes, tapeworms (land and sea birds)
<i>Parvoviridae</i> <i>Densovirinae</i> (<i>Piccovirales</i>)	<i>Aquambidensovirus</i> , <i>Decapod aquambidensovirus</i> 1 (2) <i>Blattambidensovirus</i> , <i>Blattella blattambidensovirus</i> 1 (1) <i>Hemiambidensovirus</i> , <i>Hemipteran hemiambidensovirus</i> 1 (2) <i>Iteradensovirus</i> , <i>Lepidopteran iteradensovirus</i> 1 (5) <i>Miniambidensovirus</i> , <i>Orthopteran miniambidensovirus</i> 1 (1) <i>Pefuambidensovirus</i> , <i>Blattodean pefuambidensovirus</i> 1 (1) <i>Protoambidensovirus</i> , <i>Lepidopteran protoambidensovirus</i> 1 (2) <i>Scindoambidensovirus</i> , <i>Orthopteran scindoambidensovirus</i> 1 (3)	ssDNA, linear 4 to 6 kb	Icosahedral T=1, 21 to 28 nm diameter	Crickets, crayfish, Hymenoptera, Orthoptera, Hemiptera, Lepitoptera, aphids, crustaceans, mosquitoes, echinoderms, prawns
<i>Parvoviridae</i> <i>Hamaparovirinae</i> (<i>Piccovirales</i>)	<i>Brevihamaparvovirus</i> , <i>Dipteran brevihamaparvovirus</i> 1 (2) <i>Hepanhamaparvovirus</i> , <i>Decapod hepanhamaparvovirus</i> 1 (1) <i>Penstylhamaparvovirus</i> , <i>Decapod penstylhamaparvovirus</i> 1 (1)	ssDNA, linear, 4 to 6 kb	Icosahedral T=1, 21 to 28 nm diameter	Decapods, diptera
<i>Permutotetraviridae</i> <i>Phasmaviridae</i> (<i>Bunyavirales</i>)	<i>Alphapermutotravivirus</i> , <i>Thosea asigna virus</i> (2) <i>Feravirus</i> , <i>Ferak feravirus</i> (1) <i>Jonvirus</i> , <i>Jonchet jonvirus</i> (1) <i>Orthophasmavirus</i> , <i>Kigluaik phantom orthophasmavirus</i> (9) <i>Sawastrivirus</i> , <i>Sancia sawastrivirus</i> (1) <i>Wuhivirus</i> , <i>Insect wehuvirus</i> (1)	+ ve ssRNA, monopartite, 5.7 kb – ve ssRNA, tripartite Fera-1.5, 4.2, 6.9 kb Jonvirus-1.7, 5.4, 6.9 kb Ortho-2.2, 2.8, 6.7 kb Sawastri -1.6, 4.2, 7.2 kb	Icosahedral T=4, 40 nm diameter Nucleocapsid pseudocircular ribonucleoprotein core, spherical, 80 nm diameter virus, enveloped. Jonchet virions can also appear rod shaped with a diameter of 60 nm by up to 600 nm in length	lepidoptera Mosquitoes, cockroach, midges, water striders, dragon flies, bees
<i>Phenuiviridae</i> (<i>Bunyavirales</i>)	<i>Goukovirus</i> , <i>Gouléako virus</i> (3) <i>Phasivirus</i> , <i>Badu phasivirus</i> (5) <i>Phlebovirus</i> , <i>Rift Valley fever phlebovirus</i> (60) <i>Wenrivirus</i> , <i>Shrimp wenrivirus</i> (1)	– ve ssRNA, tripartite Gouko-1.1, 3.2, 6.4 kb Phasi-1.3, 4.1, 6.8 kb Phlebo-1.7, 3.2, 6.4 kb	Virions typically spherical with diameters 80 to 130 nm with surface proteins arranged in a T=12 icosahedral lattice with three RNP coils in the core	Gouko, mosquitoes Phasi, mosquitoes and flies Phlebo, mosquito vector Wenri, shrimp Ants, midges
<i>Polycipiviridae</i> (<i>Picornavirales</i>)	<i>Chipolycivirus</i> , <i>Chironomus riparius virus</i> 1 (2) <i>Hupolycivirus</i> , <i>Hubei hupolycivirus</i> (1) <i>Sopolycivirus</i> , <i>Solenopsis invicta virus</i> 2 (11)	+ ve ssRNA, monopartite, 10 to 12 kb	Icosahedral, pseudo T=3, 33 nm diameter	Ants, midges
<i>Polydnaviridae</i>	<i>Bracovirus</i> , <i>Cotesia melanoscela bracovirus</i> (32) <i>Ichnovirus</i> , <i>Campoletis sonorensis ichnovirus</i> (21)	dsDNA, circular, multipartite Bracovirus, 15 to 30 segs, 2 to 31 kb each, 25 to 600 kb aggregate size Ichnovirus, 20 to 50 segs, 2 to 20 kb each, 150 to 250 kb aggregate size	Bracovirus, multiple rod shaped nucleocapsids, enveloped Ichnovirus, prolate ellipsoid, enveloped	Parasitoid hymenoptera

(Continued)

Table 1 Continued

Family/Subfamily Order	Genus, Subgenus, Type Species (total # in genus)	Genome type, segments and size	Morphology, T value for icosahedral capsids	Major invertebrate host
<i>Poxviridae</i> <i>Chordopoxvirinae</i> (<i>Chitovirales</i>)	<i>Centapoxvirus</i> , <i>Yokapox virus</i> (2)	dsDNA genome phylogenetically more <i>Orthopoxvirus</i> -like than any entomopoxvirus	Poxvirus like morphology	Mosquito but kills suckling mice and replicates in Vero cells
<i>Poxviridae</i> <i>Entomopoxvirinae</i> (<i>Chitovirales</i>)	<i>Alphaentomopoxvirus</i> , <i>Melolontha melolontha entomopoxvirus</i> (7) <i>Betaentomopoxvirus</i> , <i>Amsacta moorei entomopoxvirus</i> (16) <i>Deltaentomopoxvirus</i> , <i>Melanoplus sanguinipes entomopoxvirus</i> (1) <i>Gammaentomopoxvirus</i> , <i>Chironomus luridus entomopoxvirus</i> (2)	dsDNA, linear, 230 to 375 kb	Brick shaped 140 to 260 × 220 to 450 nm enveloped complex morphology	Mosquitoes, coleoptera, lepidoptera, orthoptera,
<i>Reoviridae</i> <i>Sedoreovirinae</i> (<i>Reovirales</i>)	<i>Cardoreovirus</i> , <i>Eriocheir sinensis reovirus</i> (1) <i>Phytoreovirus</i> , <i>Wound tumor virus</i> (3) <i>Seadornavirus</i> , <i>Banna virus</i> (3)	dsRNA all with 12 segments <i>Cardoreovirus</i> 0.7 to 3.7 kb <i>Seadornavirus</i> 0.9 to 3.7 kb <i>Phytoreovirus</i> 1.4 to 4 kb	Icosahedral, 2 shells, T=2 (inner shell) T=13 (intermediate shell), 60 to 70 nm diameter ¹	Cardio, crabs Phyto, leafhopper (and plants) Seadorna, mosquitoes (and mammals)
<i>Reoviridae</i> <i>Spinavirinae</i> (<i>Reovirales</i>)	<i>Aquareovirus</i> , <i>Aquareovirus A</i> (7) <i>Coltivirus</i> , <i>Colorado tick fever coltivirus</i> (5) <i>Cypovirus</i> , <i>Cypovirus 1</i> (16) <i>Dinovernavirus</i> , <i>Aedes pseudoscutellaris reovirus</i> (1) <i>Idnoreovirus</i> , <i>Idnoreovirus 1</i> (5)	dsRNA 9 to 12 segments Aqua-11 segs, 0.8 to 3.9 kb Colti-12 segs, 0.7 to 4.3 kb Cypo-10 segs, 1 to 4.2 kb Dinoverna-9 segs, 1.1 to 3.8 kb Idno, 10 segs, 1 to 4.8	Icosahedral single (T=2) or dual (T=2/T=13) shell Aqua-dual shell, 75 nm diameter Colti-dual shell, 60 to 80 nm diameter Cypo-single shell, 65 nm diameter Dinoverna-single shell T=2 only, 50 nm diameter Idno, dual shell T=2/T=13, 70 nm diameter	Aqua, crab and oyster Colti, ticks and mosquitoes Cypo, lepidoptera Dinoverna, <i>Aedes</i> mosquitoes Idno, hymenoptera
<i>Rhabdoviridae</i> (<i>Mononegavirales</i>)	<i>Almendravirus</i> , <i>Puerto Almendras alمندravirus</i> (6) <i>Alphanemrhavirus</i> , <i>Xingshan alphanemrhavirus</i> (2) <i>Caligrhavirus</i> , <i>Lepeophtheirus caligrhavirus</i> (3) <i>Sigmavirus</i> , <i>Drosophila melanogaster sigmavirus</i> (7)	– ve ssRNA, monopartite, 10.8 to 16.1 nm	Bullet shaped, helical rod shaped nucleocapsid 45 to 100 nm × 100 to 430 nm with matrix and envelope	mosquito, butterflies bed bugs, nematodes, sea lice, flies
<i>Roniviridae</i> (<i>Nidovirales</i>)	<i>Okavirus</i> , (Subgenus <i>Tipravirus</i>) <i>Gill-associated virus</i> (3)	+ ve ssRNA, monopartite, 26 kb	Bacilliform 40 to 60 nm × 150 to 200 nm, enveloped with peplomers	Crustaceans (prawns)
<i>Sarothroviridae</i>	<i>Macronovirus</i> , <i>Macrobrachium satellite virus 1</i> (1)	+ ve ssRNA, monopartite, 0.8 kb	Icosahedral T=? 15 nm diameter naked	Prawns, crayfish, butterflies, water bugs
<i>Solinviviridae</i> (<i>Picornavirales</i>)	<i>Invictavirus</i> , <i>Solenopsis invicta virus 3</i> (1) <i>Nyfulvavirus</i> , <i>Nylanderia fulva virus 1</i> (1)	+ ve ssRNA, monopartite, 8 to 11 kb 5' Vpg, 3' polyA	Icosahedral T=3 27 to 29 nm diameter	Ants
<i>Tospoviridae</i> (<i>Bunyavirales</i>)	<i>Orthotospovirus</i> , <i>Tomato spotted wilt orthotospovirus</i> (26)	– ve ssRNA, tripartite, 3, 4.8, 8.8 kb	Pseudocircular ribonucleoprotein core, 80 to 120 nm diameter, enveloped	Thrips
<i>Xinmoviridae</i> (<i>Mononegavirales</i>)	<i>Anphevirus</i> , <i>Xincheng anphevirus</i> (7)	– ve ssRNA, monopartite, 12 kb	Not seen, based solely on sequence	Mosquitoes, drosophila, orthoptera

^aThe taxon Order, if any is in brackets below the family or subfamily name. The taxon Subfamily, if any, is below Family name and is indented. The number of species in a genus is in brackets after the Genus name.

Note: Some families in this list also include viruses that infect more than invertebrates.

ICTV of any of these viruses can consult the ICTV Taxonomy Database for any virus species mentioned, at “See Relevant Websites section”. This site is updated on a regular basis following approval of new taxa by the ICTV. Any reference to taxa in the articles in this Volume are based on that posted as of Aug 31, 2020.

Below are provided brief synopses of the virus families, in alphabetical order, which include viruses which infect invertebrates. In addition to family names, names of additional taxa including Order, Subfamily, Genus and Species are also provided. The total number of species for each genus are included in brackets after the genus name and is based on the ICTV Taxonomy list as of August 31, 2020. Also provided are the years that the corresponding names for the taxa were first approved by the ICTV. Note that by the rules of orthography of the ICTV, the names of taxa (e.g., Order, ending in “*virales*”, Family, ending in “*viridae*”, Subfamily ending in “*virinae*”, genus, one word ending in “*virus*”, and Species, two or more words ending in “*virus*”) are in italics and names of viruses are in normal Roman letters. Also given are the names of exemplar viruses of each type species. It will be noticed that in most cases the species names are derived directly from the names of a virus classified into that species, the only difference being that the species name is in italics and the virus name is in normal font.

A fuller description of most of these are provided in the rest of the articles in Viruses of Invertebrates.

Artoviridae

Family: *Artoviridae* [ICTV 2018] (Order *Mononegavirales*, ICTV 1990)

Genus: *Hexartovirus* [ICTV 2019] (2 species)

Type Species: *Caligid hexartovirus* [ICTV 2019]

Species Exemplar: *Lepeophtheirus salmonis* negative-strand RNA virus-1

Genus: *Peropuvirus* [ICTV 2016] (6 species)

Type Species: *Pteromalus puparum peropuvirus* [ICTV 2016]

Species Exemplar: *Pteromalus puparum* negative strand RNA virus-1

Artoviruses, originally discovered by high throughput sequencing, belong to the order *Mononegavirales* and hence have a single negative sense (–ve) ssRNA genome which is up to 12.5 kb in length. The two genera in the *Artoviridae* family (*Arto* from arthropod), *Peropuvirus* (from *Pteromalus puparum* negative-strand RNA virus, PpNSRV-1) with six species and *Hexartovirus* (from the crustacean host crab, *Hexanauplia* and *artovirus*) with two species. PpNSRV-1 was initially reported in 2017 in a pteromalid wasp, *Pteromalus puparum*, hence the name of the virus. Other members of the *Peropuvirus* genus have been found in other parasitic wasps, pillworms, woodlice, odonates and copepods, while members of the *Hexartovirus* genus are associated with copepod sea lice or barnacles. PpNSRV-1 virions, the only artovirus virions to be described, are enveloped with a diameter of 100–130 nm. PpNSRV-1 is transmitted vertically through both male and female wasps and is present in various tissues throughout the life stages. While PpNSRV-1 increases adult longevity, infection does have a fitness cost. PpNSRV-1 particles are present in the follicular cells and ovaries of the parasitoid wasp host and in intracellular vesicles of the digestive tract. Virions for the other artoviruses have not been described as they have been identified solely from metagenomic studies. Collectively artovirus genomes encode nucleocapsid protein (N), glycoprotein (G) RNA directed RNA polymerase domain in the large (L) protein along with some proteins of unknown function, U1 to U3. Transcription initiation and termination motifs, typical of other mononegavirales viruses flank the independently transcribed ORFs.

Ascoviridae

Family: *Ascoviridae* [ICTV 1998] (Order *Pimascovirales*, ICTV 2019)

Genus: *Ascovirus* [ICTV 1998] (3 species)

Type Species: *Spodoptera frugiperda ascovirus 1a* [ICTV 1998]

Species Exemplar: *Spodoptera frugiperda* ascovirus 1a

Genus: *Toursvirus* [ICTV 2015] (1 species)

Type Species: *Diadromus pulchellus toursvirus* [ICTV 1998]

Species Exemplar: *Diadromus pulchellus* toursvirus 4a (DpTV)

Ascoviruses are unusual viruses of lepidopteran insects that appear to be horizontally transmitted by endoparasitoid wasps during oviposition and have a wide tissue tropism. They are bacilliform or allantoid shaped virions about 130–150 nm in diameter and 300–400 nm in length with an inner, enveloped, 80 × 300 nm core surrounded by a second, outer layer of protein subunits. Their large circular dsDNA genome is 100–200 kb in size encoding 117–180 genes. In addition to their highly complex structures, ascoviruses initiate an unusual cytopathology starting in the nucleus followed by enlargement of the infected cell and its eventual cleavage into 10–30 vesicles similar to apoptotic vesicles. Virion morphogenesis, including development of the inner core surrounded by de novo synthesized membrane and envelopment by an outer membrane is completed within the vesicles as the cell ruptures. As the virions accumulate near the periphery of the vesicles, they often form occlusion bodies in a “foamy” vesicular matrix. When these abundant, highly refractile vesicles accumulate in the hemolymph they form a milky white appearance. Ascoviruses share 9 core genes with iridoviruses and the giant marseilliviruses infecting amoeba. Phylogenetic analysis suggests that iridoviruses and ascoviruses share a common ancestor which is shared with marseilliviruses.

Baculoviridae

Family: *Baculoviridae* [ICTV 1975]

Genus: *Alphabaculovirus* [ICTV 2008] (55 species)

Type Species: *Autographa californica multiple nucleopolyhedrovirus* [ICTV 1979]

Species Exemplar: *Autographa californica multiple nucleopolyhedrovirus* (C6)

Genus: *Betabaculovirus* [ICTV 2008] (26 species)

Type Species: *Cydia pomonella granulovirus* [ICTV 1991]

Species Exemplar: *Cydia pomonella granulovirus* (Mexican 1)

Genus: *Deltabaculovirus* [ICTV 2008] (1 species)

Type Species: *Culex nigripalpus nucleopolyhedrovirus* [ICTV 2004]

Species Exemplar: *Culex nigripalpus nucleopolyhedrovirus* (Florida 1997)

Genus: *Gammabaculovirus* [ICTV 2008] (2 species)

Type Species: *Neodiprion lecontei nucleopolyhedrovirus* [ICTV 2002]

Species Exemplar: *Neodiprion lecontei nucleopolyhedrovirus*

Baculoviruses are among the most studied of any invertebrate virus. These viruses infect only insects, specifically lepidoptera (alpha and betabaculoviruses), diptera (gammabaculoviruses) and hymenoptera (*deltabaculoviruses*). Although baculovirus-like viruses in occlusion bodies have been described in shrimp, their taxonomic status has not been verified.

From a historical perspective, baculoviruses were initially recognized as agents which caused jaundice (or flacherie) disease of silkworm. The “polyhedrosis” which was associated with this disease was due to polyhedral bodies observed in liquified silkworm which died of the disease. In the late 1860s it was Louis Pasteur who raised the hypothesis that the “flecherie” disease of silkworm was due to a bacterium, later in the 1920s identified as a virus by Glaser and Paillot. The involvement of polyhedrosis diseases on reducing pest insect populations led to a lot of research interest in identifying and characterizing the causative agent and using them for biological control. Perhaps the first evidence of their potential for insect control was the unforeseen introduction of gHNPV, a nuclear polyhedrosis virus (NPV) through parasitoids from Finland, into a population of the European spruce sawfly, *Gilpinia hercyniae*, a serious pest of spruce in Canada in 1938. As reported by a mentor for the author of this article, Fred Bird of the Insect Pathology Research Institute in Sault Ste Marie, Ontario, Canada, by 1945 the sawfly infestation essentially collapsed and is still being kept in check by this virus. This was followed shortly by control of another sawfly pest, *Neodiprion sertifer* by introduction from Sweden of NeseNPV, now commercialized as Neocheck-s, by Gerry Wyatt of Watson and Crick fame. Gerry observed that in NPV DNA the ratio of A:T and of G:C were 1 irrespective of the NPV source, leading to the theory of the double stranded nature of the DNA double helix. As is often the case with fundamental research, the study of the molecular biology of baculoviruses, largely by the laboratories of Lois Miller (University of Georgia) and Max Summers (Texas A&M University) led to the development of the baculovirus expression system (BEVS) now in universal use for production of bioreagents and vaccines, including against COVID-19, and as gene delivery vectors.

Baculoviruses are characterized by two morphological forms, budded virus (BV) and occlusion-derived virus (ODV). Budded virus, as the name implies, are virions which have budded from infected cells to form rod-shaped virions with helical nucleocapsids measuring 30–60 nm in diameter by 250–300 nm in length surrounded by an envelope derived from the plasma membrane and containing the major envelope protein GP64 or F. The ODV has an almost identical nucleocapsid and is surrounded by an inner nuclear membrane-derived envelope and, as the name implies, is occluded in an occlusion body, made of either polyhedrin (nucleopolyhedroviruses) or granulin (granuloviruses, GVs). The circular dsDNA genome of a given baculovirus is identical between the budded and occlusion derived form and ranges in size from 82 (NeleGV) to 179 kb (XecnGV), encoding 81–183 ORFs. Unsurprisingly the baculovirus phylogeny, based on 38 core baculovirus genes, matches that of the taxonomy, with the *deltabaculovirus* and *gammabaculoviruses* being more ancient and the *alphabaculoviruses* being divided into two clades representing Groups I and II, respectively.

During infection, the polyhedra or granules disassemble in the alkaline insect gut, releasing the ODVs which pass through the peritrophic membrane of the gut lumen, aided by metalloprotease enhancers. Virion nucleocapsids enter the microvilli of midgut epithelial cells via membrane fusion, assisted by a complex of *per os* infectivity factors. In contrast to the ODVs, BVs infecting other cells of the host or tissue culture cells enter via receptor-mediated endocytosis. In both cases the nucleocapsids are released into the cytoplasm of the cells to initiate infection. Nucleocapsids are transported to the nuclei where they become uncoated and initiate a temporally regulated cascade of immediate early, delayed early, late and very late virus gene transcription. Following nuclear hypertrophy, viral DNA replication and viral protein synthesis, nucleocapsids begin to be formed within the nucleus.

Early in infection, nucleocapsids escape the ring zone in the nucleus and through actin-mediated transport, traverse the cytoplasm to the plasma membrane, via microtubules, collecting at so called budding sites where the major envelope protein GP64 (or F for some baculoviruses) has formed foci, in association of another viral protein ME53. GV envelopes do not contain a fusion protein suggesting that budded GVs might not play a role in vivo. As the nucleocapsids bud through the plasma membrane they are enveloped by the membrane and are released as BVs which spread through the hemocoel or trachea to other tissues, or cell to cell transmission in tissue cultures.

Late in infection, and in the same cells producing BVs, the nucleocapsids remain in the nucleus. They become enveloped by a membrane derived from the inner nuclear envelope which contains, among other viral proteins, the *per os* infectivity factors forming the occlusion-derived viruses. The ODVs are then assembled within polyhedra or granules based on the major occlusion body protein, polyhedrin (alpha, delta and gamma-baculoviruses) or granulin (betabaculovirus).

At the end of the infection the concerted activity of viral chitinase and cathepsin enzymes results in the liquefaction of the larvae and dissemination of the polyhedra or granules in the environment to infect other insects. Assisted in this dissemination is a curious baculovirus modification of behavior in which the infected larvae migrate to the tops of the plants they are on prior to death, a character called "Wipfelkrankheit".

As noted earlier, baculoviruses have also been developed as high-level protein expression systems, now widely commercialized and used for the development of bioactive reagents. This was a consequence of basic fundamental research completely unrelated to their use as biological control agents. At regular seminars held on Friday afternoons in the early 1980s at the Chicken Oil Company in College Station Texas, home of Texas A&M University, Gale Smith and others from the Max Summers lab noted that polyhedrin was the major protein produced in baculovirus-infected cells amounting to as much as 50% of the total protein. This led to a more in-depth characterization by Gale Smith, Max Summers and Malcolm Fraser, in 1983, of the polyhedrin promoter (in the absence of sequencing) showing that sequences upstream of the polyhedrin ORF did contain a powerful polyhedrin promoter and was capable of directing expression of human β interferon. This promoter was the basis of developing the baculovirus expression vector system (BEVS). This system has been expanded to utilize the p10 promoter, another highly expressed late promoter and others such as the orf46 promoter of SeMNPV.

This BEVS system led to the production of literally thousands of different proteins, some of pharmacological and diagnostic importance, such as specific antibodies, immunotherapeutic agents, virus-like particles and vaccines. Baculoviruses are also being considered as gene therapy vectors showing some early encouraging results. For example, baculoviruses expressing a Shigatoxin A subunit known to kill cancer cells, successfully reduced tumors in mice with MCF7-induced breast cancer after injection of the tumor. This author was not surprised to learn that BEVS was being used to generate diagnostic reagents and vaccines to address the 2019/2021 COVID-19 pandemic.

Bidnaviridae

Family: *Bidnaviridae* [ICTV 2011] (Order *Poliovirales*, ICTV 2008)

Genus: *Bidensovirus* [ICTV 2011] (1 species)

Type Species: *Bombyx mori bidensovirus* [ICTV 2011]

Species Exemplar: *Bombyx mori bidensovirus* 3 (VD1)

Bidensoviruses, which caused a denonucleosis in infected silkworm larvae were originally classified as parvoviruses because of their single stranded linear DNAs encapsidated separately into small icosahedral capsids. However bidensoviruses BmBDV-2 and BmBDV-3 are quite different from other densoviruses in that they have a bisegmented dsDNA genome that in aggregate (13 kb) is larger than the 4–6 kb of densoviruses, encodes a DNA polymerase, no densovirus ORFs, has a different transcription strategy and has no identity with densovirus genomes. Consequently, in 2012 they were re classified in the new family *Bidnaviridae*, genus *Bidensovirus*, type species *Bombyx mori bidensovirus*.

Bidensoviruses are naked, 20–24 nm diameter icosahedral viruses with a bisegmented linear ssDNA genome (6.5 kb VD1 and 6 kb VD2) with ITRs in which complementary strands are separately encapsidated into 4 virion types. Both strands of both segments generate transcripts expressing different proteins and hence the genome has an ambisense organization. To date, bidensoviruses have been isolated only from silkworm and hence the three isolates (Yamanashi, Zhenjiang and Indian) are referred to as BmBDVs. In terms of evolution, bidensoviruses are quite unusual since different parts of the genome appear to originate from different sources. Based on bioinformatic analysis the BmBDV NS1 and VP may have been derived from parvoviruses, the DNAPolB might have evolved from the hypothetical ancestral polintoviruses, P133 may have come from an insect reovirus and NS3 from granuloviruses.

One strand of ambisense VD1 encodes transcripts for nonstructural proteins NS2 and NS1 and structural protein VP while the other strand encodes transcripts for protein-primed PolB. Though ORFs for NS2 and NS1 overlap, two different transcripts, with alternative initiation sites, are made. Leaky scanning of the VP ORF results in two VPs. One strand of ambisense VD2 encodes transcripts for the nonstructural NS3 and the other strand encodes the structural protein P133.

NS1 has DNA binding, ATPase and helicase activities suggesting it is a multifunctional protein important in virus replication. The function of NS2 is still unknown, though it shares slight sequence similarity with dnaA, a chromosomal replication initiator protein, and a DNA-binding response regulator. A homolog of NS3 in a densovirus is important for viral DNA replication. The VPs are the major structural proteins which can self-assemble if expressed in a baculovirus expression system, while P133 is a minor structural protein. Interestingly P133 has a homolog with the outer capsid protein of cypoviruses, members of the *Reoviridae* family. PolB is a protein-primed DNA polymerase that appears to be processed by post translational cleavage into three peptides with sizes of 53 (detected in purified virions), 70 and 120 kDa.

Infection of silkworm intestinal epithelial cells appears to be by receptor-mediated endocytosis through attachment of the virion VPs with a NSD-2, an amino transporter protein cell receptor protein. During infection there is nuclear hypertrophy with the appearance, late in infection, of a dense homogenous structure in the nuclei and concomitantly vacuolization of mitochondria and endoplasmic reticulum and formation of a large autophagosome. The dead columnar epithelial cells are sloughed out, along with the released viruses into bead-like frass. The mode of replication of BmBDV DNA is unknown though absence of replicative intermediates typical of parvovirus DNA replication seem to discount rolling circle replication. Rather, the protein-primed nature of the BmBDV DNAPolB and the presence of ITRs at the genome ends which could form a panhandle, suggest replication of BmBDV DNA might occur by protein priming, as is well described for replication of adenovirus DNA.

Infection of early instars results in the lengthening of times between instars while infection of late instars can result in pupation and adult emergence, but the weight and survival of pupae is compromised.

Birnaviruses of Invertebrates (*Entomobirnavirus*)

Family: *Birnaviridae* [ICTV 1984]

Genus: *Entomobirnavirus* [ICTV 1993] (2 species)

Type Species: *Drosophila X virus* [ICTV 1991]

Species Exemplar: *Drosophila X virus*

Viruses belonging to three genera of the family *Birnaviridae* infect non mammalian vertebrates. The fourth genus represents entomobirnaviruses infecting exclusively insects and are classified in the genus *Entomobirnavirus*. Originally discovered as a contaminant in a fruit fly cell culture, at least four other birnaviruses have also been found in free-living mosquitoes and drosophila. One birnavirus recovered in a mosquito cell line, Espirito Santo virus, was found in a sample from a Dengue-3 virus-infected patient from Brazil.

Like other birnaviruses, they have an icosahedral $T = 3$ symmetry and a dsRNA genome which, as the name implies, is bisegmented. One segment is about 3.2 kb and the other about 3.4 kb in length, with VPg, a virion protein derived from VP1 and linked to the RNA genome, covalently attached at the 5' end of the positive sense strand of each ds genomic RNA. The 3.4 kb segment A encodes a 114 kDa polyprotein precursor which is proteolytically cleaved by the VP4 protease into structural capsid proteins in the order on the polyprotein, preVP2, VP4 and VP3. A fourth VP, VP5 is encoded in frame with VP4. In addition to being a capsid protein, VP3 has anti RNAi activity suppressing antiviral RNAi. The 3.2 kb segment B encodes the RdRp. Little is known of the entomobirnavirus replication cycle, but it is likely that genome replication and transcription are similar to that of other birnaviruses, covered elsewhere in the encyclopedia. Entomobirnaviruses cause a cytopathic effect in infected insect cells including cell aggregation and ending in death. *Drosophila X virus* can result in a persistent infection often resulting in oxygen starvation causing death.

Dicistroviridae

Family: *Dicistroviridae* [ICTV 2002] (Order *Picornavirales*, ICTV 2008)

Genus: *Aparavirus* [ICTV 2009] (6 species)

Type Species: *Acute bee paralysis virus* [ICTV 2004]

Species Exemplar: *Acute bee paralysis virus*

Genus: *Cripavirus* [ICTV 2002] (4 species)

Type Species: [ICTV 1982] *Cricket paralysis virus*

Species Exemplar: *Cricket paralysis virus*

Genus: *Triatovirus* [ICTV 2015] (5 species)

Type Species: [ICTV 2002] *Triatomavirus*

Species Exemplar: *Triatoma virus*

Dicistroviruses are a group of highly stable +ve sense ssRNA viruses which, unlike other picornavirales viruses, has a bicistronic genome encoding two ORFs separated by an intergenic ribosome entry site (IRES), a characteristic which led to their name as dicistroviruses. They are 30 nm diameter icosahedral viruses with a pseudo $T = 3$ symmetry and composed of 4 viral proteins, VP1, 2, 3 and 4. As a group dicistroviruses have a broad host range including many insects like bees, ants, crickets, aphids, drosophilids, lepidoptera and triatomes, the vector of Chagas disease, and other arthropods such as shrimp and crabs. Thus, they are of great economic and medical importance and some dicistroviruses have been touted as biological control agents for them.

The 8.5–20 kb +ve sense ssRNA genome has a VPg covalently linked to the 5' end containing two ORFs, ORF1 and ORF2. ORF1 is preceded by a 5' IRES which, along with the intergenic IRES between the two ORFs and a 3' polyA tail, allows for translation of both ORFs in a cap-independent manner. ORF1 encodes nonstructural proteins including a helicase, the VPg, protease and RNA dependent RNA polymerase (RdRp). ORF2 encodes the four structural proteins VP1, 2, 3 and 4. While the receptor for attachment to cells is still unknown, dicistroviruses enter cells via clathrin-mediated endocytosis. The rest of the replication cycle is similar to the picornaviruses, including the synthesis of the ORF1 and ORF2 polyproteins which are auto proteolytically cleaved into the smaller functional proteins. The VPg is uridylylated at a tyrosine residue by the viral RdRp allowing it to serve as a primer for replication of both the -ve and +ve sense strands.

Most of the dicistroviruses infect mostly the gut and shed virus into the gut lumen, eliminating it through the feces which enables efficient transmission of the virus. However, dicistroviruses infecting hymenoptera infect all developmental stages and replicate in most tissues. Infection is associated with nutritional stress and in bees can be manifest in neurological signs such as paralysis, trembling and flightlessness and ultimately death. Infections due to dicistroviruses have been associated, along with other factors, with the worldwide colony collapse disorder of beehives in 2006–2007 and continues to plague the honey industry. Dicistroviruses also cause losses in the crab and shrimp industries, being the causative agents of sleeping disease by Mud crab virus and Taura syndrome in penaeid shrimp due to Taura syndrome virus. Many dicistroviruses are transmitted horizontally through the fecal-oral route, while in hymenoptera, virus transmission can be food-borne, venereal, vector borne and vertical.

Hytrosaviridae

Family: *Hytrosaviridae* [ICTV 2011]

Genus: *Glossinavirus* [ICTV 2011] (1 species)

Type Species: *Glossina hytrosavirus* [ICTV 2011]

Species Exemplar: *Glossina pallidipes* salivary gland hypertrophy virus (Uganda)

Genus: *Muscavirus* [ICTV 2011] (1 species)

Type Species: *Musca hytrosavirus* [ICTV 2011]

Species Exemplar: *Musca domestica* salivary gland hypertrophy virus

Hytrosaviruses are large dsDNA viruses of diptera, mostly tsetse fly and houseflies and syrphid fly. Though the symptoms of salivary gland hypertrophy in tsetse flies were first described in the 1930s, the etiological viral agent exhibiting as virus-like rods, was not described until 1978. In hytrosavirus-infected insects the salivary glands swell to up to four times normal size, leading to the origin of the name salivary gland hypertrophy virus. They are now classified in the family *Hytrosaviridae* which has two genera, *Glossinavirus* and *Muscavirus*, named after the host of the virus in the genus. Each genus has a single species which is the type species, *Glossina hytrosavirus* in genus *Glossinavirus* and *Musca hytrosavirus* in genus *Muscavirus*. Though not always lethal, infection is associated with reproductive dysfunctions including infertility and distorted mating behaviors. The *Musca hytrosavirus* *Glossina pallidipes* salivary gland hypertrophy virus is particularly problematic in raising tsetse flies, the causative agent of African sleeping sickness, for the very effective sterile male release program using the sterile insect technique (SIT). Once in a colony, the hytrosavirus results in collapse of the colonies as happened in the Seibersdorf, Austria tsetse fly culture.

Virions are rod-shaped with a diameter of 50–65 nm and length of 500–1000 nm with a bilipid membrane envelope and circular dsDNA genomes 124–190 kb in length. They are not only structurally similar to other large dsDNA invertebrate viruses like baculoviruses, nudiviruses and nimaviruses but they share about 12 of the 38 core genes common to them. Little is known of their replication cycle. They appear to be infected through the tracheal system. Viral capsids are released into the cytoplasm through membrane fusion. Through the cell's microtubule system, the capsids move to the nucleus to initiate transcription of immediate early, early and late genes. More detailed studies of these early stages are hampered by lack of a more tractable cell culture system. Nucleocapsids can be observed leaving the nucleus through nuclear pores and exiting the cell by budding through the membrane of cells bordering the salivary gland lumen.

Transmission of *Glossina hytrosavirus* in tsetse flies is both horizontal, due to feeding of contaminated blood and vertical through trans-ovum transmission or through infected milk. Transmission of *Musca hytrosavirus* is largely horizontal through gregarious feeding of the houseflies.

Iflaviridae

Family: *Iflaviridae* [ICTV 2011] (Order: *Picornavirales*, ICTV 2008)

Genus: *Iflavirus* [ICTV 2008] (15 species)

Type Species: *Infectious flacherie virus* [ICTV 2002]

Species Exemplar: *Infectious flacherie virus*

Iflaviruses are small icosahedral viruses 22–30 nm in diameter with a monopartite 8–10.5 kb + ve ssRNA genome with a VPg at the 5' end and a templated polyA tail at the 3' end. They infect only arthropods, namely insects including honeybees, lepidoptera, aphids and planthoppers and arcnids, such as mites parasitic on honeybees. Genus, *Iflavirus* numbering 15 species to date includes the type species *Infectious flacherie virus* represented by a silkworm virus, *Infectious flacherie virus* from which the genus and species names are derived. Their cellular replication cycle is thought to be similar to that of other picornaviruses in the *Picornavirales* including cap-independent translation due to an IRES 5' to the polyprotein ORF and involving auto catalytic cleavage into structural proteins and the RdRp.

Iflaviruses are transmitted largely through contaminated food and often result in a permissive infection sometimes resulting in vertical transmission. Some symptomatic infections do occur resulting in diarrhea, developmental abnormalities and death of the host. Iflavirus infections of honeybees have been linked, along with other bee viruses, with colony collapse disorders and as such they are of economic importance.

Iridoviruses of Invertebrates (*Betairidovirinae*)

Family: *Iridoviridae* [ICTV 1975] (Order: *Pimascovirales*, ICTV 2019)

Subfamily: *Betairidovirinae* [ICTV 2016]

Genus: *Chloriridovirus* [ICTV 1981] (5 species)

Type Species: *Invertebrate iridescent virus 3* [ICTV 1999]

Species Exemplar: Mosquito iridescent virus

Genus: *Decapodiridovirus* [ICTV 2018] (1 species)

Type Species: [ICTV 2018] *Decapod iridescent virus 1*

Species Exemplar: Shrimp hemocyte iridescent virus

Genus: *Iridovirus* [ICTV 1971] (2 species)

Type Species: *Invertebrate Iridovirus 6* [ICTV 1999]

Species Exemplar: Chilo iridescent virus

The *Irodoviridae* family represents large icosahedral viruses classified into two subfamilies. Viruses in the subfamily *Alphairidovirinae* infect ectothermic vertebrates including bony fishes, amphibians and reptiles. The second subfamily, *Betairidovirinae*, includes all iridescent viruses of invertebrates. It has three genera, *Chloriridovirus* with five species, *Decapodiridovirus* with one species and the original genus *Iridovirus*, with two species. The invertebrate iridescent viruses are among the largest icosahedral viruses with diameters up to 200 nm and symmetries of $T = 147$ to 217. Though they have an icosahedral structure, they can be more complex including having an internal membrane enveloping the nucleoprotein core containing a linear dsDNA genome of 163–220 kb that is circularly permuted with terminal redundancies. While some are naked, iridoviruses that bud from infected cells will have an additional membrane surrounding it. They have a fairly broad host range infecting 6 orders of insects including mosquitoes and lepidoptera and crustaceans, like isopods, molluscs and oysters (*Decapodiridoviruses*) and even some vertebrates. For example, iridolike viruses cause gill necrosis and oyster velar virus disease in oysters. The viruses are highly stable in water but thermolabile and sensitive to dry conditions. Viruses enter susceptible cells via clathrin-mediated endocytosis, in the case of IIV-6 using ORF 096 L as the virus attachment molecule. Transcription is temporally regulated into immediate early and delayed early phases and a viral RNA polymerase II-mediated late phase. Genome size or slightly longer DNA molecules are first generated in the nucleus and after transport to the cytoplasm form concatemeric DNA facilitated by their terminal redundancy and are methylated. The concatemeric viral DNA is packaged by a “headful” process in which slightly longer than genome length DNA is packaged into virions as circularly permuted terminally redundant molecules. Virions are released from infected cells by cell lysis or by budding from the host cell membrane. Like many DNA viruses, iridoviruses code for anti-apoptotic proteins, like IIV-6 ORF193R, that are expressed early to protect the virus from the cells’ apoptotic machinery. However, late in infection they also induce apoptosis, mediated by viral proteins like the Chilo iridescent virus 389 L protein (iridoptin) which aids in viral dissemination to adjacent cells. Transmission between insects can be perorally or by endoparasitic hymenoptera. Infection can be patent, with an accumulation of high numbers of virions, giving rise to a visible iridescence, from which the viruses originally got their name. Infection is normally lethal or covert resulting in only a low level of virus particles, but still resulting in reduced fitness.

Malacoherpesviridae

Family: *Malacoherpesviridae* [ICTV 2008] (Order: *Herpesvirales*, ICTV 2008)

Genus *Aurivirus* [ICTV 2012] (1 species)

Type Species: *Haliotid herpesvirus 1* [ICTV 2012]

Species Exemplar: *Haliotid herpesvirus 1*

Genus *Ostreavirus* [ICTV 2008] (1 species)

Type Species: *Ostreid herpesvirus 1* [ICTV 2008]

Species Exemplar: *Ostreid herpesvirus 1*

The malacoherpesviruses are herpesviruses that infect a variety of marine invertebrates notably molluscs, like oysters (superfamily *Ostreioidea*) and abalones (sea snails in the family *Halitodidae*) and have a global economic impact, particularly in hatcheries. The oyster herpesvirus, OsHV-1, was the first malacoherpesvirus discovered and that was in adult Eastern oysters, *Crassostrea virginica* in 1972. It was also found to be responsible for mass mortality in Pacific oyster (*C. gigas*), European flat oyster (*Ostrea edulis*) and other oyster species. Disease outbreaks due to a second malacoherpesvirus, *Haliotis herpesvirus* HaHV-1, were also discovered in a variety of abalone species such as *Haliotis diversicolor supertexta*, *H. laevigata* and *H. rubra* and others beginning in 2005. The malacoherpesviruses have a typical herpesvirus morphology consisting of a 70–80 nm diameter icosahedral capsid with $T = 16$, surrounded by a trilaminar unit membrane giving an overall diameter of about 100–180 nm. However, the malacoherpesviruses lack the tegument usually associated with herpesviruses but instead have a 5 nm translucent gap between the capsid and outer viral membrane filled with fine fibrils. The linear dsDNA genomes of only two malacoherpesviruses have been sequenced including the ostreavirus OsHV-1 genome with 207,339 bp (NC_005881) and the aurivirus, abalone herpesvirus Victoria/AUS/2009 genome with a size of 211,518 bp (NC_018874). Both genomes are structured as in other herpesviruses, in the order $TR_L-U_L-IR_L-X-IR_S-U_S-TR_S$. For example, for OsHV-1, the genome is organized into a 3,370 bp unique short U_S and a 167,843 bp unique long U_L regions having two orientations each flanked by inverted repeats. U_L is flanked by 7,584 bp TR_L and IR_L and U_S is flanked by 9,774 bp TR_S and IR_S . However, the malacoherpesviruses differ from amniote herpesviruses in the family *Herpesviridae* and herpesviruses of fishes and amphibians in the family *Alloherpesviridae* by infecting only invertebrate hosts and their genomes are phylogenetically distinct from them. Consequently, during the re-examination of the order *Herpesvirales* they were classified in 2008 in a new family *Malacoherpesviridae* with two genera *Ostreavirus* for the oyster herpesviruses and *Aurivirus* for the abalone herpesviruses. The prefix “malaco” for this family name is derived from the word malaco, meaning soft, a term used to denote molluscs.

Mesoniviridae

Family: *Mesoniviridae* [ICTV 2012] (Order: *Nidovirales*, ICTV 1996)

Subfamily: *Hexponivirinae* [ICTV 2018]

Genus: *Alphamesonivirus* [ICTV 2002] (9 species)

Subgenus: *Namcalivirus* [ICTV 2018]

Type Species: [ICTV 2012] *Alphamesonivirus 1*

Species Exemplar: Cavally virus

Mesoniviruses are 50–120 nm diameter enveloped mosquito viruses with a monopartite 20 kb +ve sense ssRNA genome with a 5' cap and 3' polyA tail. For such a small group of very similar viruses in terms of morphology and a relatively tight nucleic acid phylogeny, mesoniviruses have a complicated taxonomy with a suborder *Mesonidovirineae* of the order *Nidovirales*, subfamily *Hexponivirinae*, family *Mesoniviridae*, one genus, *Alphamesonivirus* with a type species *Alphamesonivirus 1*. The genus is further subdivided into eight subgenera. The subgenera *Karsalivirus* and *Namcalivirus* each have two species, while each of the other six subgenera have only one species in the subgenus.

The genome consists of seven ORFs. The replicative enzymes ORF1a and 1b overlap and are translated by a ribosomal frame shift, facilitated by overlapping GGAUUUU sequences. ORF3a and 3b might similarly be translated by a ribosomal frame shift. ORF1a encodes a viral protease for autocatalytic processing of pp1a and pp1ab into smaller proteins. ORF1a and ORF1b encodes a polyprotein processed by autocatalytic cleavage to RdRp, helicase, exoribonuclease and two methyl transferases. The ORFs downstream of ORF1b, are for structural proteins, two Spike protein subunits, S1 and S2, the nucleocapsid protein N and 4 membrane spanning proteins (M). These are translated from two subgenomic RNAs, the synthesis of which are regulated by transcription regulating sequences (TRSs). The envelope consists of protruding spike-like projections and an integral membrane protein anchoring it to the nucleoprotein core.

To date mesoniviruses have been identified in only mosquitoes and one report of a mesonivirus from aphids based on next generation sequencing. While they replicate in some insect cells, they could not replicate in several mammalian and avian cell lines studied. Infection results in a variety of cytopathic effects, associated with a high and rapid yield of progeny viruses. Some mesoniviruses are maintained in the population by vertical transmission with transmission from female mosquitoes to her progeny, including male offspring. Horizontal transmission may occur via contaminated legs, wings and even saliva from naturally infected mosquitoes.

Nimaviridae

Family: *Nimaviridae* [ICTV 2002]

Genus: *Whispovirus* [ICTV 2002] (1 species)

Type Species: *White spot syndrome virus* [2008]

Species Exemplar: White spot syndrome virus-CN

White spot syndrome virus is the only species in the *Nimaviridae* family. Variants of White spot syndrome virus (WSSV), the only virus in the species, causes a highly contagious disease, White spot disease (WSD), after which the virus was named. The first WSD outbreak was in China and Taiwan around 1992 and spread to other shrimp farming countries like Japan, Thailand, the Americas, France and the Middle East. The virus has a broad host range among crustaceans, including shrimp, crayfish, crab and even lobster resulting in large economic losses in the related industries. Originally, because of the baculovirus-like rod shape of the nucleocapsid and its large ds circular DNA genome, their earlier names included Chinese baculo-like virus, *Penaeus monodon* baculovirus and white spot baculovirus. The names were changed to WSSV followed by letters designating their geographic origins and were first classified as species *White spot syndrome virus 1* in the family *Baculoviridae*. However due to significant differences in morphology and phylogeny between these viruses and baculoviruses, they were reassigned to a new family *Nimaviridae* (named for the threadlike tail, “nima” in Latin, of the virions) with a single genus, *Whispovirus*. The genus *Whispovirus* derives its name from the sole *White spot syndrome virus* species with the White spot syndrome virus-CN (WSSV-CN) as the exemplar isolate. This species is represented by 15 other isolates whose genomes have been completely sequenced including WSSVs from China, Japan, Taiwan, Thailand, Mexico, Australia, India, Brazil and Ecuador.

WSSVs are large, 70–170 in diameter by 210–320 nm in length, enveloped virions with a thread-like extension with an intermediate tegument layer containing a 65 × 300 nm rod-shaped nucleocapsid, reminiscent of baculovirus nucleocapsids. In addition to about 50 structural proteins, the virus also has lipid components. The large circular dsDNA genome is 281–309 kb in size and has nine internal homologous repeats similar to the *hrs* of baculovirus genomes. Only about 150 of the predicted 530 ORFs are thought to be functional. The WSSV genome also encodes some microRNAs targeting both viral and host genes. While there is some genotypic variability among the isolates, all are considered to be from the same species.

WSSV enters cells by caveola or clathrin-mediated endocytosis and alters host metabolism. Replication is nuclear with temporally regulated transcription divided into immediate early, early, late and very late phases. While most mRNAs have 3' polyA tails and 5'-caps there is no evidence of splicing, but some have IRESs which could be used to regulate translation. Like baculoviruses, the early genes have host-like RNAPolIII promoters, while late genes have a late gene promoter. Replication takes place entirely in the nucleus, including *de novo* formation of intranuclear viral envelope and morphogenesis of nucleocapsids.

The virus appears to target stomach, gills, lymphoid organs and connective tissue but not hepatopancreas and midgut, both of endodermal origin. Signs of infection include lethargy, a reddish coloration of the body and white spots, from which it derives its name.

Nodaviridae

Family: *Nodaviridae* [ICTV 1981] (Order *Nodamuvirales*, ICTV 2019)

Genus: *Alphanodavirus* [ICTV 1997] (5 species)

Type Species: *Nodamura virus* [ICTV 1981]

Species Exemplar: Black beetle virus

Alphanodaviruses replicate exclusively in insects, notably wax moth larvae, grass grubs, tsetse fly, reduviid bugs and mosquitoes. Viruses from genus *Betanodavirus*, the second of two genera in the family, infect fish. In 1975, the black beetle virus was the first

nodavirus discovered which was from black beetles (*Heteromychus arator*) in New Zealand. The virions are icosahedral with $T = 3$ symmetry and diameters of 32–33 nm and consisting of a single capsid protein (CP) or its two cleavage products. The Flock house virus CP is 44 kDa and the two cleavage products β and γ are 39 and 4 kDa in size, respectively. The +ve ssRNA genome is bisegmented with sizes of 1.3 and 1.4 kb. Following entry to the cytoplasm, three mRNAs, RNA1, which codes for the catalytic subunit of RdRp, RNA2 which codes for CP, and RNA3, which codes for B1 and B2 which function as suppressors of RNAi, are produced. Infection of hosts usually results in stunting, paralysis and death. Complete genome sequences have been deposited for black beetle virus, Boolarra virus, five Flock House viruses, two Nodamura viruses and one Pariacoto virus.

Nudiviridae

Family: *Nudiviridae* [ICTV 2013]

Genus: *Alphanudivirus* [ICTV 2013] (2 species)

Type Species: *Oryctes rhinoceros nudivirus* [ICTV 2013]

Species Exemplar: *Oryctes rhinoceros nudivirus* (Ma07)

Genus: *Betanudivirus* [ICTV 2013] (1 species)

Type Species: *Heliothis zea nudivirus* [ICTV 2013]

Species Exemplar: *Heliothis zea nudivirus* (1)

The first nudivirus recognized as a virus is probably the “non occluded baculovirus” that infected the Rhinoceros beetle in Samoa and which was used in 1967 as an effective biological control agent against this beetle. Since then it has had a convoluted taxonomic history, originally classified as a baculovirus, except that it did not form occlusion bodies like other baculoviruses. Through phylogenetic analysis, the nudiviruses are now classified in their own family, *Nudiviridae* organized into two genera, *Alphanudivirus* with two species including *Oryctes rhinoceros nudivirus* as the type species and *Betanudivirus* with *Heliothis zea nudivirus* as the sole member and type species. Nudivirus virions have a rod-shaped nucleocapsid and a circular dsDNA 96–232 kb genome, much like that of baculovirus genomes and are enveloped. The alphanudiviruses are more rounded in shape measuring 100×200 nm, while the betanudiviruses are more rod-shaped measuring 80×300 – 400 nm. They differed initially from baculoviruses in that they were non occluded, or naked, resulting in their current taxonomic designation as “nudi” viruses. However, some nudiviruses have enveloped or non enveloped occlusion bodies, which in some cases are facultative. They infect many orders of insects and are a major problem for the shrimp industry. Not much is known about their replication cycle.

OrNV attaches to cells and is taken in by pinocytosis. It is released into the cytoplasm upon fusion with lysosomal membranes and degeneration of the lysosomes. Presumably the nucleocapsids are transported to the nucleus where the DNA is released, initiating replication around 7 h post infection. Transcription is temporally regulated into early, intermediate and late stages with at least 100 transcripts being expressed. Viral envelopes and nucleocapsids and episomally replicated viral DNA assemble in the nucleus to be released into the cytoplasm followed by budding through the plasma membrane.

Heliothis zea nudivirus 1 (HzNV-1) was originally found in a persistently infected cell line IMC Hz-1. HzNV-1 causes gonadal atrophy in adult *Galleria mellonella*. In cells, HzNV-1 has both a productive (lytic) infection resulting in a high titer of virus and cell death and in some cells a persistent (latent) infection. Persistently infected cells can be passaged without cell lysis and viral DNA can be detected in episomal or integrated form. Two viral genes appear to be correlated with the switch between productive and latent infections. When a high concentration of HzNV-1 viral *hh1* is expressed in latently infected cells a lytic infection ensues. However if *pag1* encoded miRNAs targeting *hh1* are produced then a latent infection occurs.

As mentioned, nudiviruses have a broad host range in insects and shrimp. Transmission is through a *peroral* feeding of feces from virus-infected midgut epithelia and *per parenteral* mating routes. Consequences of infection can include reduced size, crippling, reduced fecundity, host sterility and death by fat body disintegration of infected individuals.

Nyamiviridae

Family: *Nyamiviridae* [2013] (Order *Mononegavirales*, ICTV 1990)

Genus: *Berhavirus* [ICTV 2018] (3 species)

Type Species: *Sipunculid berhavirus* [ICTV 2018]

Species Exemplar: Běihǎi rhabdo-like virus 4

Genus: *Crustavirus* [ICTV 2015] (3 species)

Type Species: *Wenzhou crustavirus* [ICTV 2015]

Species Exemplar: Běihǎi rhabdo-like virus 6

Genus: *Nyavirus* [ICTV 2013] (3 species)

Type Species: *Nyamanini nyavirus* [ICTV 2013]

Species Exemplar: Sierra Nevada virus

Genus: *Orinovirus* [ICTV 2018] (1 species)

Type Species: *Orinoco orinovirus* [ICTV 2018]

Species Exemplar: Orinico virus

Genus: *Socyvirus* [ICTV 2015] (1 species)

Type Species: *Soybean cyst nematode socyvirus* [ICTV 2013]

Species Exemplar: Soybean cyst nematode virus 1
 Genus: *Tapwovirus* [ICTV 2018] (1 species)
 Type Species: *Tapeworm tapwovirus* [ICTV 2018]
 Species Exemplar: Wēnzhōu tapeworm virus 1

Nyamiviruses, being in the Order *Mononegavirales*, have a non-segmented -ve ssRNA genome ranging from 8 to 12 kb in size, with the exception of *Tapwovirus* which has a bi-segmented – ve ssRNA genome with an aggregate size of 10.5 kb. The family *Nymaviridae* consists of six genera and a total of 12 species. The family is named for the *Nyavirus* genus (derived from *Nyamanini* Pan in South Africa from where the first *Nyavirus*, *Nyamanini virus*, was isolated). Viruses in that genus are tick-borne and can also infect birds. Many viruses in this family have been implied to exist only through next generation sequencing, so little virological and biological information, except their association with a broad range of invertebrates is available. Viruses in the genus *Berhavirus* are associated with marine echinoderms and sipunculids, those of genus *Crustavirus*, as the name implies, are associated with crustaceans, those of *Orinovirus* with moths, those of *Socyrivirus* with plant parasitic nematodes and those from *Tapwovirus* are from tapeworms. The *Nyamanini virus* virions are enveloped and spherical in shape with diameters of 100–130 nm and consisting of 5–6 structural proteins including the nucleocapsid protein (N), glycoprotein (G) and RdRP (L) based on sequence similarity to other mononegavirales sequences. The genomes of the different nyamiviruses encode 5–6 ORFs encoding structural proteins including N, G, P, L and some of unknown function. *Nyamanini virus* replicates in the nucleus but other aspects of virus replication is unknown, a major limitation of identifying viruses solely through metagenomics.

Parvoviruses of Insects (*Densovirinae* and *Hamaparvovirinae*)

Note: The family taxonomy of the *Parvoviridae* underwent a major reorganization as this article was being written. Below is the updated taxonomy as of August 2020.

Family: *Parvoviridae* [1975] (Order *Picovirales*, ICTV 2019)

Subfamily: *Densovirinae* [ICTV 1993]

Genus: *Aquambidensovirus* [ICTV 2019] (2 species)
 Type Species: *Decapod aquambidensovirus 1* (2) [ICTV 2019]
 Species Exemplar: *Cherax quadricarinatus densovirus*
 Genus: *Blattambidensovirus*, (1) [ICTV 2019] (1 species)
 Type Species: *Blattella blattambidensovirus 1* [ICTV 2019]
 Species Exemplar: *Blattella germanica densovirus*
 Genus: *Hemiambidensovirus* [ICTV 2019] (2 species)
 Type Species: *Hemipteran hemiambidensovirus 1* (2) [ICTV 2019]
 Species Exemplar: *Dysaphis plantaginea densovirus*
 Genus: *Iteradensovirus* [ICTV 2013] (5 species)
 Type Species: *Lepidopteran iteradensovirus 1* (5) [ICTV 2013]
 Species Exemplar: *Bombyx mori densovirus 1*
 Genus: *Miniambidensovirus* [ICTV 2019] (1 species)
 Type Species: *Orthopteran miniambidensovirus 1* (1) [ICTV 2019]
 Species Exemplar: *Acheta domestica miniambidensovirus*
 Genus: *Pefuambidensovirus* [ICTV 2019] (1 species)
 Type Species: *Blattodean pefuambidensovirus 1* (1) [ICTV 2019]
 Species Exemplar: *Periplanata fuliginosa densovirus*
 Genus: *Protoambidensovirus* [ICTV 2019] (2 species)
 Type Species: *Lepidopteran protoambidensovirus 1* (2) [ICTV 2019]
 Species Exemplar: *Galleria mellonella densovirus*
 Genus: *Scindambidensovirus* [ICTV 2019] (3 species)
 Type Species: *Orthopteran scinambidensovirus 1* (3) [ICTV 2019]
 Species Exemplar: *Acheta demostica densovirus*

Subfamily: *Hamaparvovirinae*

Genus: *Brevihamadensovirus* [ICTV 2019] (2 species)
 Type Species: *Dipteran brevihamadensovirus 1* (2) [ICTV 2019]
 Species Exemplar: *Aedes albopictus C6/36 cell densovirus*
 Genus: *Hepanhamaparvosovirus* [ICTV 2019] (1 species)
 Type Species: *Decapod hepanhamaparvosovirus 1* (1) [ICTV 2019]
 Species Exemplar: *Panemus merguensis hepandensovirus*
 Genus: *Penstylhamaparvovirus* [ICTV 2019] (1 species)
 Type Species: *Decapod penstylhamaparvovirus 1* (1) [ICTV 2019]
 Species Exemplar: *Decapod penstyldensovirus 1*

Viruses in the family *Parvoviridae* are small, naked icosahedral viruses with a small, linear, single stranded DNA genome about 4–6 kb in length. Parvoviruses are classified into three subfamilies based largely on the nature of their host. Those in the *Parvovirinae* subfamily infect vertebrates while those in the two subfamilies *Densovirinae* and *Hamaparvovirinae* infect invertebrate hosts. *Densovirinae* have eight genera including the original *Itaradensovirus*, and what used to be genus *Ambidensovirus*, which due to recent phylogeny studies (2019) has been reorganized into seven genera, as indicated above. A second new subfamily, *Hamaparvovirinae*, was created and accommodates densoviruses in three of the five genera which infect invertebrates namely genera, *Brevihamadensovirus* (originally *Brevidensovirus*), *Hepanhamadensovirus* (originally *Hepandensovirus*) and *Penstylhamadensovirus* (originally *Penstyldensovirus*).

Densoviruses were first discovered in the early 1960s as an epizootic of denosonucleosis (the name reflecting the dense appearance of nuclei due to the accumulation of viroplasm in nuclei of infected cells) in a commercial mass rearing of the greater wax moth larvae used for fishing bait in France. The virions have a naked icosahedral $T = 1$ symmetry measuring 21–26 nm in diameter with a monopartite ssDNA genome, 4–6 kb in length encoding two or three NS proteins and two to four VPs depending on the species. The ambisense densoviruses (in the seven ambidensovirus genera) package both the –ve and +ve ssDNA strands, but in separate virions while the other, monosense capsids, package primarily the -ve ssDNA strand. The imperfect palindromic sequences at each end of the genome fold into hairpins providing a free 3' OH for initiating DNA replication using host DNA polymerase. There is a wide variety of expression strategies to generate the different proteins, including splicing, leaky scanning and overlapping promoters and ORFs. As a group, the densoviruses have a very broad host range, including six orders of insects, decapod crustaceans, like shrimp and crayfish and echinoderms, like starfish. Parvovirus infections are polytropic in the tissues they infect and are often lethal and spread easily. Disease can be devastating to commercial enterprises, especially in industrial shrimp and crayfish farming, silk production and in the insects for food industry such as cricket farming. Some insect parvoviruses might be potentially developed as biocontrol agents against insect pests like ants and mosquitoes and perhaps some agricultural insect pests. For example, densovirus was used to successfully control *Limacodidae* pests of oil palm plantations and *Galleria mellonella* in beehives.

Polydnaviridae

Family: *Polydnaviridae* [ICTV 1984]

Genus: *Bracovirus* [ICTV 1999] (32 species)

Type Species: *Cotesia melanoscela bracovirus* [ICTV 1990]

Species Exemplar: *Cotesia melanoscela bracovirus*

Genus: *Ichnovirus* [ICTV 1990] (21 species)

Type Species: *Campoletis sonorensis ichnovirus* [ICTV 1999]

Species Exemplar: *Campoletis sonorensis ichnovirus*

Polydnaviruses were first recognized in the late 1970s in the laboratories of Don Stoltz (Dalhousie University, Canada) and Brad Vinson (Texas A&M University, USA) who noticed virus-like particles from the calyx region of the ovaries of first *Apanteles* (later renamed *Cotesia*) *melanoscelus* and then in other parasitic hymenoptera. Polydnaviruses are associated with the reproductive tracts of braconid (bracoviruses) and ichneumonid (ichnoviruses) parasitoids. Injected, along with parasitoid eggs, into host caterpillars, polydnaviruses subvert the host immune and physiology systems benefiting developing parasitoid eggs and larvae. Polydnaviruses are also characterized as having a multipartite circular dsDNA genome, referred to originally as a polydisperse genome, from which the family *Polydnaviridae* took its name. Because of the distinct differences, particularly in morphology and phylogeny between polydnaviruses from braconid wasps and those from ichneumonid ones, the viruses were classified in two genera *Bracovirus* and *Ichnovirus*, respectively. Another unique feature of polydnaviruses is that their genomes exist in an integrated and extrachromosomal form. The integrated provirus form allows for vertical transmission, and some of the integrated genes are responsible for generation of virions in the calyx cells. The extrachromosomal form is packaged into virions and encodes genes expressed to take over host immunology and physiology to benefit the developing parasitoid subsequent to their delivery coincident with oviposition.

The polydnavirus virion DNA is a collection of discrete numbers of circular dsDNA molecules in non equimolar abundance. The number and size of individual circles varies depending on the genus and species.

Bracovirus virions measure 50 nm in diameter x 30–80 nm long, consist of rod-shaped nucleocapsids with one or more of them surrounded by a nucleus-derived membrane and resemble baculovirus budded viruses. It was because of this morphology that bracoviruses were originally classified along with the baculoviruses in 1982 with the bracovirus *Apanteles melanoscelus virus* (now *Cotesia melanoscelus polydnavirus*) being recognized as a species in the then genus *Baculovirus* in the family *Baculoviridae*. The bracovirus dsDNA genome consists of 15–30 circles ranging in size from 2 to 50 kb with aggregate sizes of 125–600 kb. BV nucleocapsids assemble and acquire an envelope from the nucleus forming intranuclear virions in calyx cells near the lumen. These “mature” calyx cells fill with virions, eventually rupturing the nuclear envelope and escaping into the calyx lumen when the cells disrupt.

Ichnoviruses have quite a different morphology. The lenticular nucleocapsid measures 80–90 nm in diameter x up to 300 nm in length. Single (from *Campopleginae* wasps) or multiple (from *Banchinae* wasps) nucleocapsids are surrounded by two, two-unit membranes, the inner one derived from within the nucleus, often surrounding a hook at one end and the outer one derived from the plasma membrane during budding into the calyx lumen. Ichnovirus circular dsDNA genomes consist of 20–50 circles ranging in size

from 2 to 38 kb with aggregate sizes of around 250 kb. Ichnovirus replication is similar to that of bracoviruses in that nucleocapsids assemble in a virogenic stroma in the nucleus where they also acquire an envelope. From that point ichnovirus development differs from the bracoviruses as the enveloped ichnovirus subvirions exit the nucleus and then bud into the calyx lumen through the apical plasma membrane.

There has been a flurry of research activity surrounding the nature of the integrated and extrachromosomal (virion) DNA, the origin of the genomes and the expression of viral genes within the parasitoid and following oviposition in the parasitized host. During virus production the viral DNA segments of the integrated genomes are amplified, excised and circularized prior to encapsidation into polydnavirus nucleocapsids destined for envelopment and secretion into the calyx lumen. The “replication” genes remain integrated but are amplified in calyx cells. They are then expressed to produce mostly proteins associated with virus production. Many polydnavirus genes are derived from the parasitoids which harbor them and consequently the phylogeny of the virus often reflects the phylogeny of the parasitoids which harbor them.

Poxviruses of Invertebrates (*Entomopoxvirinae*)

Family: *Poxviridae* [ICTV 1978] (Order *Chitovirales*, ICTV 2019)

Subfamily: *Entomopoxvirinae* [ICTV 1978]

Genus: *Alphaentomopoxvirus* [ICTV 2004] (7 species)

Type Species: *Melonontha melonontha entomopoxvirus* [ICTV 1976]

Species Exemplar: *Melonontha melonontha entomopoxvirus*

Genus: *Betaentomopoxvirus* [ICTV 2004] (16 species)

Type Species: *Amsacta moorei entomopoxvirus* [ICTV 1976]

Species Exemplar: *Amsacta moorei entomopoxvirus*

Genus: *Deltaentomopoxvirus* [ICTV 2019] (1 species)

Type Species: *Melanoplus sanguinipes entomopoxvirus* [ICTV 1976]

Species Exemplar: *Melanoplus sanguinipes entomopoxvirus*

Genus: *Gammaentomopoxvirus* [ICTV 2004] (6 species)

Type Species: *Chironomus luridis entomopoxvirus* [ICTV 1976]

Species Exemplar: *Chironomus luridis entomopoxvirus*

Subfamily: *Chordopoxvirinae* [ICTV 1978]

Genus: *Centapoxvirus* [ICTV 2016] (2 species)

Type Species: *Yokapox virus* [ICTV 2016]

Species Exemplar: *Yoka poxvirus*

The *Poxviridae* family is subdivided into two subfamilies. Poxviruses of vertebrates, like the familiar smallpox virus, are classified in the subfamily *Chordopoxvirinae*, while those in the *Entomopoxvirinae* subfamily infect only insects. The *Entomopoxvirinae* subfamily is currently divided into three genera, *Alphaentomopoxvirus* with 7 species, *Betaentomopoxvirus* with 16 species and *Gammaentomopoxvirus* with 6 species. The entomopoxviruses, first discovered by Constantine Vago in 1963, have been isolated from four orders of insects, namely Coleoptera, Orthoptera, Lepidoptera and Diptera. One poxvirus, Yoka poxvirus, was isolated from pools of mosquitoes in 1972, but phylogenetically it is more similar to orthopoxviruses than to any entomopoxvirus. Yoka poxvirus is classified in the subfamily *Chordopoxvirinae*, *Centapoxvirus* genus, species *Yokapoxvirus*. Moreover, it replicates in Vero and other mammalian cells and causes death in intracranially-inoculated suckling mice suggesting it may not be an insect virus. Given that it was isolated from a pool of mosquitoes it is more likely to be a contaminant of a chordopoxvirus, perhaps from a blood meal. Entomopoxvirus virions are ovoid shaped measuring 140–260 nm × 220–450 nm with a complex morphology. The entomopoxvirus genome is a linear dsDNA ranging in size from 229 to 308 kb with an A+T content of 79%–82% and shares 49 core genes with the chordopoxviruses. Unlike the chordopoxviruses of vertebrates, entomopoxviruses are embedded in a spherical shaped spheroid, made of spheroidin, much in the same manner that baculoviruses are embedded in a polyhedrin or granulin based occlusion body.

Similar to other poxviruses, as described elsewhere in this encyclopedia, entomopoxviruses replicate in the cytoplasm in viroplasms, first as immature particles that mature into enveloped brick-shaped intracellular mature virions which can be released by cell lysis or by budding through the plasma membranes as external enveloped virions allowing systemic spread of the infection. Late in infection, and unlike chordopoxviruses, mature entomopoxvirions become severally occluded in spheroids, remaining intracellular till cell lysis with the spheroids allowing for insect to insect spread. Another unique feature of entomopoxviruses is that they produce a viral protein called fusolin which can be occluded with virions into spheroids which can also result in the formation of “spindles” which are not occluded. The spindles are thought to disrupt the peritrophic membrane allowing access of the virus to the epithelial cells of the larval midgut. Fusion of the outer EPV membrane with the plasma membrane allows access to the cell cytoplasm where replication occurs. Virus spread in the insects starts with the hemocytes and spreads to the fat body, the main site for virus replication, and other tissues. The signs of the disease vary depending on the virus and the host. EPV-infected lesser cornstalk borer become unresponsive and sluggish with their cutical changing from brown to orange, while in black-soil scarabs they develop a white spotted or mottled epidermis. EPV-infected lepidoptera can become whitish while some increase in size due to delayed pupation. Although they are insect viruses that can kill the insect, the time to kill is too long to be feasible for use in the control of insect pests.

Reoviruses of Invertebrates (*Sedoreovirinae* and *Spinavirinae*)

Family: *Reoviridae* [ICTV 1974] (Order *Reovirales*, ICTV 2019)

Subfamily *Sedoreovirinae* [ICTV 2009]

Genus: *Cardoreovirus* [ICTV 2008] (1 species)

Type Species: *Eriocheir sinensis reovirus* [ICTV 2008]

Species Exemplar: *Callinectes sapidus reovirus* 1

Genus: *Phytoreovirus* [ICTV 1978] (3 species)

Type Species: *Wound tumor virus* [ICTV 1976]

Species Exemplar: *Wound tumor virus* (34)

Genus: *Seadornavirus* [ICTV 2004] (3 species)

Type Species: *Banna virus* [ICTV 1999]

Species Exemplar: *Banna virus-China*

Subfamily: *Spinavirinae* [ICTV 2009]

Genus: *Aquareovirus* [ICTV 1990] (7 species)

Type Species: *Aquareovirus A* [ICTV 1999]

Species Exemplar: *American oyster reovirus*13p2

Genus: *Coltivirus* [ICTV 2009] (5 species)

Type Species: *Colorado tick fever coltivirus* [ICTV 1974]

Species Exemplar: *Colorado tick fever-Florio*

Genus: *Cypovirus* [ICTV 1990] (16 species)

Type Species: *Cypovirus 1* [ICTV 1999]

Species Exemplar: *Bombyx mori cypovirus* 1

Genus: *Dinovernavirus* [ICTV 2008] (1 species)

Type Species: *Aedes pseudoscutellaris reovirus* [ICTV 2008]

Species Exemplar: *Aedes pseudoscutellaris reovirus*

Genus: *Idnoreovirus* [ICTV 2004] (5 species)

Type Species: *Idnoreovirus 1* [ICTV 2004]

Species Exemplar: *Diadromus pulchellus idnoreovirus* 1

Genus: *Orthoreovirus* [ICTV 1991] (10 species)

Type Species: *Mammalian orthoreovirus* [ICTV 1999]

Species Exemplar: *Mahlapitsi orthoreovirus*

The reoviruses have a broad host range, including insects, algae, fungi, fish, crustaceans, crabs as well as humans and other vertebrates and their invertebrate vectors and of plants and their invertebrate vectors. The *Reoviridae* family is subdivided into two subfamilies, *Sedoreovirinae* with six genera and *Spinavirinae* with nine genera. The two subfamilies differ in the nature of the virion surface with the *Spinavirinae* having a turreted or “spiked” surface (hence the prefix “Spina”, Latin for spike) while the *Sedoreovirinae* have a smooth surface (“Sedo” is Latin for smooth). Of the more recognized reoviruses are those infecting humans, other mammals and other vertebrates. Those from *Rotavirus* and *Orthoreovirus* genera infect only vertebrates, while viruses in genera *Coltivirus*, *Orbivirus* and *Seadornavirus* are considered arboviruses which infect both vertebrates and their insect vectors (mostly ticks, mosquitoes, midges, gnats and sandflies). Lesser known reoviruses infect algae (*Mimoreovirus*) and fungi (*Mycoreovirus*). In addition to infecting plants, reoviruses in genera *Fijivirus*, *Oryzavirus* and *Phytovirus* also infect their insect vectors (e.g., glassy winged sharpshooter). Reoviruses infecting only invertebrates are in genera *Cardoreovirus* (crabs), *Cypovirus* (insects), *Dinovernavirus* (mosquitoes) and *Idnoreovirus* (hymenoptera and diptera). The aquareoviruses, as the name implies, infect aquatic vertebrates like fish and but also aquatic invertebrates including shellfish and crustaceans such as oysters.

This summary includes six genera of invertebrate-infecting reoviruses, *Cardoreovirus*, *Phytoreovirus* and *Seadornavirus* in the *Sedoreovirinae* and *Cypovirus*, *Dinovernavirus* and *Idnoreovirus* in the *Spinavirinae*. Of the *Spinavirinae* insect reoviruses, most cypoviruses have 10 dsRNA segments and infect lepidopteran insects like *Lymantria dispar* (Gypsy larvae), idnoreovirus have 10 dsRNA segments and infect *Diadroma pulchella* (a hymenopteran parasitoid) and dinovernavirus has 9 dsRNA segments and infects *Aedes* mosquitoes. Most of the *Sedoreovirinae* reoviruses have 12 dsRNA segments (though *Macrobrachium nipponense reovirus* has 10 dsRNA segments). Seadornavirus is an arbovirus of humans, pigs and cattle but also infects mosquito vectors and two seadornaviruses appear to infect only mosquitoes.

As noted some reoviruses, like the arboviruses in the genera *Coltivirus* (Colorado tick fever virus), and *Orbivirus* (Bluetongue virus) are vectored by invertebrates which themselves are infected and include midges, mosquitoes, ticks and sandflies. Also, some reoviruses infect plants and are vectored by invertebrates carrying the viruses. These are described in more detail in other articles in this Encyclopedia.

Like other reoviruses, most invertebrate reoviruses have a double shelled icosahedral capsid with diameters ranging from 60 to 80 nm (though idnoreovirus capsids are smaller at about a 50 nm diameter). While cypoviruses and dinovernaviruses have a single shell with $T = 2$, aquareoviruses, coltiviruses, idnoreoviruses, cardcoreoviruses, phytoviruses and seadornaviruses have two shells, the inner with $T = 2$ and the other with $T = 13$ icosahedral symmetry. As noted above they have 9–12 dsRNA segments. These range in size from 0.76 to 4.8 kb with each segment, except for cypoviruses, encoding one protein. For cypoviruses, segment 5 encodes two proteins by ribosomal skipping.

Cypovirus was the first invertebrate reovirus discovered and that was in 1934 as polyhedra in the cytoplasm of infected cells of silkworm larvae and hence were initially named cytoplasmic polyhedrosis viruses from which the current genus name was derived.

Aquareoviruses (*Aqua* for their common aqueous environment of their hosts) have been isolated from oysters and crabs. Coltiviruses (from *Colorado tick fever virus*) replicate in ticks as well as mammalian hosts. Idnoviruses (from “idno” meaning water) were first isolated from houseflies in 1978, while the lone dinovernavirus (from *double stranded RNA icosahedral nove*, meaning nine, dsRNA segments) was isolated from a cell line from *Aedes pseudoscutellaris* in 2005. The first cardoreovirus (from *carcinus* for crab and *dodeca* for 12 dsRNA segments) was isolated in 2004 from diseased mitten crabs (*Eriocheir sinensis*). The Banna virus, in the genus *Seadornavirus* (from Southeast Asian *dodeca* for 12 dsRNA segments), was first discovered in 1987 in a patient with encephalitis and fever in China. Phytoreoviruses (from “phyto” Greek for plant) were initially discovered in 1941 as wound tumor viruses of plants transmitted by leaf hoppers in which the virus also replicates.

The invertebrate reoviruses range from having a limited host range to a broad one. Aquareoviruses have been isolated from American oyster and Geoduck, shore, swimmer and mitten crabs. Coltiviruses replicate in ticks, the main vector for Colorado tick fever and mosquitoes. Cypoviruses have been isolated from lepidoptera and hymenoptera. Idnoreoviruses infect hymenoptera, and dipterans like the housefly, olive fly and fruit flies. So far dinovernavirus has been isolated solely from persistently infected *Aedes pseudoscutellaris* cells. Seadornaviruses have been found in mosquitoes, cardoreovirus in crabs and prawn and phytoreovirus in various plants and leafhopper vectors.

Infection and replication of the invertebrate reoviruses closely mimic that of other reoviruses, so the reader is referred to other reovirus chapters. One unique feature of the cypoviruses is that they form polyhedra which can retain infectivity for years. The major polyhedrin protein is translated in the cytoplasm late in infection in large excess and virions are occluded in the polyhedrin at the periphery of the virogenic stroma in which morphogenesis of the virions occurs.

Rhabdoviruses of Invertebrates (*Rhabdoviridae*, *Almendravirus*, *Alphanemrhavirus*, *Caligrhavirus*, *Sigmavirus*)

Family: *Rhabdoviridae* (Order *Mononegavirales*, ICTV 1990)

Genus: *Almendravirus* [ICTV 2016] (6 species)

Type Species: *Puerto Almendras almdravirus* [ICTV 2016]

Species Exemplar: Puerto Almendras virus (Lo-39)

Genus: *Alphanemrhavirus* [ICTV 2018] (2 species)

Type Species: *Xingshan alphanemrhavirus* [ICTV 2018]

Species Exemplar: Xingshan nematode virus 4 (XSNXC3292)

Genus: *Caligrhavirus* [ICTV 2018] (3 species)

Type Species: *Lepeophtheirus caligrhavirus* [ICTV 2018]

Species Exemplar: *Lepeophtheirus salmonis* rhabdovirus 127 (LSRV127)

Genus: *Sigmavirus* [ICTV 2012] (7 species)

Type Species: *Drosophila melanogaster sigmavirus* [ICTV 2012]

Species Exemplar: *Drosophila melanogaster* sigmavirus (HAP23)

Like reoviruses, rhabdoviruses as a group, composing 30 genera, have a very broad host range, including humans and other mammals (e.g., genera *Lyssavirus*, *Tibrovirus*, *Vesiculovirus*), fish (genera *Novirhabdovirus*, *Perhabdovirus* and *Sprivivirus*), birds (genus *Tupavirus*) and plants (*Varicosavirus* nucleorhabdoviruses). Many of the vertebrate rhabdoviruses, (like those in genera *Curiovirus*, *Ephemerovirus*, *Tibrovirus* and *Vesiculovirus*) are also arboviruses that have been isolated from arthropod vectors like mosquitoes, midges and ticks. The arthropod specific rhabdoviruses belong to four genera in the *Rhabdoviridae* family, *Almendravirus* with five species, *Alphanemrhavirus* with two species, *Caligrhavirus* with three species and the most studied *Sigmavirus*, with seven species. Not covered in this introductory section are some rhabdoviruses in other genera that are arboviruses which infect vertebrates and are found in arthropod vectors such as midges, sandflies and mosquitoes and those found in insects such as aphids, planthoppers, leafhoppers and mites that transmit virus to plants, but do not replicate in the vector. The higher-level taxonomic designation of these viruses is highly convoluted but for the sake of completeness, these viruses belong to the Order *Mononegavirales*, Class *Monjiviricetes*, Subphylum *Haploverticota*, Phylum *Negarnaviricota* and Realm *Riboviria*.

In terms of morphology and virus replication the insect rhabdoviruses are similar to other rhabdoviruses covered elsewhere in this Encyclopedia. They have a helical nucleocapsid, surrounded by an envelope with spikes and a matrix between the nucleocapsid and envelope. The bullet shaped virions measure 80 nm diameter x 100 nm in length (sigmaviruses) or 40–55 nm diameter x 190–460 nm in length (almendraviruses). Alphanemrhavirus was identified only through metagenomic sequencing of pooled parasitic nematode populations, so it is not sure if an infectious virion form exists for this species.

The 10.5–14.5 kb ssRNA genome encodes in a 3′ to 5′ direction the canonical four rhabdovirus structural genes for, nucleoprotein (N) which is the major protein of the nucleocapsid, phosphoprotein (P) which ensures proper placement of L on the genome, matrix protein (M) which is the major structural protein between the nucleocapsid and virion and helps regulate genome RNA synthesis, and, glycoprotein G which is the major envelope glycoprotein and acts as the virus attachment protein and the non structural large protein (L) which is responsible for RdRp activity, 5′ capping and 3′ polyadenylation. In addition, the genomes of sigmavirus has an additional ORF X between P and M, involved in host-virus interaction and in almdravirus encodes an additional class 1a viroporin, between G and L. In some cases, such as sigmaviruses, invertebrate rhabdovirus sequences are found integrated into the genomes of drosophila and some mosquitoes.

The insect rhabdovirus life cycle is similar to that of other rhabdoviruses. Virions attach to host cell receptors through the G glycoprotein inducing low pH-dependent receptor-mediated endocytosis. Transcription occurs in the cytoplasm in a “stop start” L and P dependent process to generate separate monocistronic mRNAs for each of the proteins in decreasing molar abundance from the 3' to 5' end of the genome. Each mRNA has a common leader 5' end which is 5' capped and ending in a reiterative copying-derived polyA at the 3' end. Late in infection, following translation of viral proteins and when the concentration of N protein has increased to sufficient levels, transcription switches to genome replication. During genome replication the full length antigenome +ve ssRNA template is produced and serves as the replicative intermediate for full length -ve sense genome ssRNA. The genome ssRNA is then encapsidated in N protein in a P protein dependent fashion. The formed nucleocapsids and M protein attach to the plasma membrane of infected cells and bud through it to release mature virions which can then infect adjacent cells.

Invertebrate rhabdoviruses have been found in several insects such as fruit flies, butterflies, mosquitoes and bees, as well as in nematodes and sea lice copepods. For the most part, transmission is vertical through either the female or male parents, though for sigmaviruses they can be experimentally transmitted by injection.

Roniviridae

Family: *Roniviridae* [ICTV 2002] (Order *Nidovirales*, ICTV 1996)

Subfamily: *Okavirinae* [ICTV 2018]

Genus: *Okavirus* [ICTV 2002] (3 species)

Subgenus: *Tipravirus* [ICTV 2018]

Species: *Gill-associated virus* [ICTV 2002]

Species Exemplar: Yellowhead virus

To date all roniviruses (from “rod shaped nidovirus”) are from crustaceans, the prawn *Penaeus monodon*. Two viral diseases are associated with roniviruses, one caused by Yellowhead virus in black tiger shrimp which first appeared in Thailand in 1990 and the second, Eriocheir sinensis ronivirus which causes black gill disease in Chinese mitten crab. Yellowhead virus appears to be transmitted horizontally, while gill-associated virus can be transmitted vertically. Roniviruses differ from other nidoviruses in having a bacilliform 40–60 nm × 150–200 nm enveloped virion. Like other nidoviruses it has a long, 26 kb, +ve ssRNA monopartite genome. The ronivirus genome encodes a 5' terminal replicase complex (ORF1) followed by ORF2 for p20 nucleoprotein, ORF3 for pp3 (producing gp116 and gp64 envelope glycoproteins) and ORF4 of unknown function. The replicase ORFs ORF1a and ORF1b have a slippery sequence and RNA pseudoknot allowing for a – 1 ribosomal frameshift and read through to a pp1ab replicase. Unlike other nidoviruses, roniviruses (and toroviruses) lack the common 5' leader. Preceding ORFs 2 and 3 are intergenic regions (IGRs) acting as a terminator and promoter. Replication probably follows that of other nidoviruses, like coronavirus described elsewhere, to generate a nested set of subgenomic mRNAs. Unlike other nidoviruses that use a discontinuous strategy to generate the different sgRNAs, roniviruses appear to use a “continuous” transcription strategy.

Sarothroviridae

Family: *Sarothroviridae* [ICTV 2015]

Genus: *Macronovirus* [ICTV 2015] (1 species)

Type Species: *Macrobrachium satellite virus 1* [ICTV 2015]

Species Exemplar: Extra small virus

Extra small virus (XSV) derives its name based on the extremely small size of its naked icosahedral particle, of about 15 nm in diameter, with the name sarthrovirus derived from small *arthropod* virus. These viruses are +ve ssRNA satellite viruses, depending on their replication on the RdRp of *Macrobrachium rosenbergii* nodavirus (MrNV). XSV belongs to the sole and type species *Macrobrachium satellite virus 1*, in the single genus *Macronovirus* of the family *Sarothroviridae*. They affect mostly different species of prawns, but have also been isolated from crayfish, beetles and water bugs and replicate in mosquito cell lines suggesting mosquitoes might be a vector. The genomic RNA is +ve sense, 796 ntd in size with a 3' polyA tail and an upstream polyadenylation signal. It encodes two ORFs for coat proteins CP16 and CP17 made in equimolar amounts, with the ORF for CP16 initiating at an internal methionine codon. Infected prawns present with lethargy and opaqueness of the abdominal muscle and degeneration of the telson and uropods, followed by death around five days after observation of gross signs of infection. Immersion in a mixture of MrNV and XSV results in 100% mortality by 12 days post infection. They are a major problem for the shrimp farms in China, India, Australia, French West Indies, Taipei and Thailand.

Solinviviridae

Family: *Solinviviridae* [ICTV 2016] (Order *Picornavirales*, ICTV 2008)

Genus: *Invictavirus* [ICTV 2016] (1 species)

Type Species: *Solenopsis invicta virus 3* [ICTV 2016]

Species Exemplar: *Solenopsis invicta virus* (DM)

Genus: *Nyfulvavirus* [ICTV 2016] (1 species)

Type Species: *Nylanderia fulva virus 1* [ICTV 2016]

Species Exemplar: *Nylanderia fulva virus 1* (Florida initial)

Soliniviruses were first discovered through a bioprospecting approach using metagenomics of ant populations, showing the value of this approach to identify novel and perhaps useful biological control agents. Soliniviruses are icosahedral in shape measuring 27–29 nm in diameter with $T = 3$ symmetry and showing surface projections and cup-like depressions. The virions are composed of two capsid proteins, VP1 and a VP1 fused to the capsid projection domain, CPD (VP1-CPD) and an about 10 kb +ve ssRNA genome with a 5' VPg and a 3' polyA tail. The genome encodes a single polyprotein ORF (*Nylanderia fulva* virus) or two polyprotein ORFs via frame shifting (*Solenopsis invicta* virus). The polyprotein for the nonstructural proteins, helicase, protease and RdRp, are located at the amino end of the polyprotein while the structural proteins VP1 (or VP1-CPD) and VP2 are at the carboxy ends. It is thought that the polyprotein is cleaved to the individual proteins by the viral protease, much like poliovirus polyproteins are. There is some evidence of subgenomic mRNAs for the structural proteins.

Viruses enter the host by feeding to access the alimentary canal of larval stage ants (*NflV-1*) or adults (*SINPV-3*). Based on the genome organization, replication is thought to be similar to that of polioviruses and caliciviruses. Replication is in the cytoplasm with the genomic +ve ssRNA acting as a transcript and later as a template for the production of genome length –ve sense ssRNA replicative strand. Virus assembly presumably occurs by encapsidation of the genomic +ve sense ssRNA by the viral capsid proteins. Virus spread is intercolonial through trophallaxis. Soliniviruses are specific to *Solenopsis* genera fire ants which exhibit signs of infection including Queen weight decline, reduced feeding and fecundity and larval mortality. Virus can be transmitted mechanically by consumption of crickets that consume dead, SIV-3-infected worker ants. Soliniviruses are being developed as biological control agents using protein baits with lower doses resulting in chronic infection and higher doses of 10^7 to 10^9 virus particles per ml leading to colony collapse. The major problem for biopesticide development is the production of sufficient virus.

Tetraviruses (Families, *Alphatetraviridae*, *Carmotetraviridae*, *Permutotetraviridae*)

Family: *Alphatetraviridae* [ICTV 2016] (Order *Hepelivirales*, ICTV 2019)

Genus: *Betatetravirus* [ICTV 1997] (7 species)

Type Species: *Nudaurelia capensis beta virus* [ICTV 1995]

Species Exemplar: *Nudaurelia cytheria capensis* β virus

Genus: *Omegatetravirus* [ICTV 1997] (3 species)

Type Species: *Nudaurelia capensis omega virus* [ICTV 1995]

Species Exemplar: *Nudaurelia capensis* ω virus

Family: *Carmotetraviridae* [ICTV 2011] (Order *Tolivirales*, ICTV 2019)

Genus: *Alphacarmotetravirus* [ICTV 2011] (1 species)

Type Species: *Providence virus* [ICTV 2004]

Species Exemplar: *Providence virus*

Family: *Permutotetraviridae* [ICTV 2011]

Genus: *Alphapermutotetravirus* [ICTV 2011] (2 species)

Type Species: *Thosea asigna virus* [ICTV 2004]

Species Exemplar: *Thosea asigna virus*

The tetraviruses were all at one time members of the *Tetraviridae* family with the common features of a unique $T = 4$ icosahedral symmetry, a common jelly roll architecture of the capsid protein which form a monophyletic group all infecting insects and all having a +ve sense ssRNA genome with a VPg at the 5' end and a polyA tail at the 3' end. However, there are some features which differed among the original species in the family, particularly the nature of the 3' end, existence of either a mono or bipartite genome, phylogeny of the RdRp of the tetraviruses and nature of the ORFs and their translation. In particular the alphatetraviruses (family *Alphatetraviridae*) have a monopartite (*Betatetravirus* genus) genome encoding both the replicase and through a subgenomic RNA the capsid protein and ending in a structured 3'tRNA-like end or a bipartite (*Omegatetravirus* genus) genome one encoding a replicase and the other the capsid protein. The replicase of both genera are alpha-like Supergroup replicases. Viruses in family *Carmotetraviridae* have a monopartite genome with one ORF encoding a Carmo-like supergroup replicase generated by a read through stop signal and the second ORF a capsid ORF generated from a subgenomic RNA. The permutotetraviruses, also have a monopartite genome with a highly structured 3' terminal sequence. The replicase and VPg are like those in the dsRNA bimaviruses including a permuted active site, hence the family name *Permutotetraviridae*.

Except for the, perhaps, subtle differences particularly in RdRp phylogeny and genome organization, most other characteristics are shared so these viruses are here treated collectively as tetraviruses. The monopartite genomes of the betatetraviruses of the *Alphatetraviridae*, carmotetraviruses, and permutotetraviruses are 6.6, 6.2, and 5.7 kb in size. Those of the bipartite genomes of the omegaviruses of the *Alphatetraviridae* are about 2.4 and 5.3 kb in size with the larger one encoding at least replicase and the smaller one encoding capsid protein ORFs. The capsid proteins are derived from subgenomic RNAs from the gRNA of either the monopartite genomes or the smaller omegavirus gRNA. The iconic $T = 4$ icosahedral capsid is about 40 nm in diameter and made from capsid proteins from all tetraviruses being monophyletic. While they usually encapsidate viral RNA, the capsids can also capsidate non-viral RNAs, allowing for generation of reassortant viruses and gene exchange. When just capsid proteins are expressed, in a baculovirus expression system, they self-assemble and under acidic conditions condense to tetravirus-like VLPs. It has long been thought that tetraviruses replicate only in insects (particularly midgut cells of larvae) and cell culture but it was recently shown that they also infect and replicate in lab-infected HeLa and MCF7 cancer cells, as well as in cowpea plants. They are found mostly as persistent infections.

Taxa of Other Viruses of Invertebrates

Not included in these summary descriptions above of viruses of invertebrates are taxa for which little is known and those for whose designations have relied largely on only sequence data from next generation high throughput sequencing and metagenomic studies of environmental samples. For many of these, if they are actually viruses of insects have not been otherwise confirmed. Since environmental samples were the source of the sequences the nucleic acid might simply have been a contaminant, or even present due to a blood meal from infected animals. While sequence data can predict phylogeny and protein composition, except for the source of the material which was analyzed, it cannot accurately provide information on basic virological data such as morphology, replication cycle, biological characteristics and true hosts. Below is a listing of some of these taxa along with some minimal associated information.

Circoviridae [ICTV 1993] (Order *Cirivirales*, ICTV 2019) hosts two genera, *Circovirus* [ICTV 1993] and *Cyclovirus* [ICTV 2015]. Circoviruses are characterized by having a circular, 1.7–1.9 kb ssDNA genome within an icosahedral capsid with $T = 1$ symmetry. The monopartite covalently closed ssDNA genome is ambisense encoding two proteins, capsid protein and replication associated protein for rolling circle DNA replication. Two of the most known circoviruses are the porcine circovirus and beak and feather disease virus of parrots. Commercially available vaccines like CircoFLEX, Circogard and Circumvent PCV2 G2, derived from virus-like particles synthesized by the insect baculovirus expression vector system, are effective against some porcine circovirus infections. While viruses from most species in both genera of the *Circoviridae* family infect humans, other mammals and birds, some virus sequences are associated with spiders and insects including ants, mosquitoes cockroaches and dragonflies. All of the insect-specific species, like *ant associated cyclovirus 1* are referred to as “associated cyclovirus” or “associated circovirus”. This is because the viruses have been identified solely through virus discovery approaches using next generation sequencing or metagenomics approaches. Consequently, we have no information about the true nature of these viruses.

Genomoviridae [ICTV 2015] (Order *Gepfluvirales*, ICTV 2019) genus *Gemyduguivirus* [ICTV 2016], with the single and type species *Dragonfly associated gemyduguivirus 1*, genus *Gemykibivirus* [ICTV 2016], with 16 species including the type species *Dragonfly associated gemykibivirus* [ICTV 2016], and genus *Gemycircularvirus* [ICTV 2015] with 43 species including species *Dragonfly associated gemycircularvirus* and *mosquito associated gemycircularvirus* are all associated with dragonflies and mosquitoes. Genomoviruses, at least for those based on more than sequence analyses, and despite being in a different order are similar to circoviruses in having a circular ambisense ssDNA genome encoding Rep and Capsid proteins in an icosahedral capsid with $T = 1$ symmetry. The genome is 2.2 kb in size and the capsids measure 29 nm in diameter. Following entry to the cytoplasm the genomovirus DNA becomes double stranded and replication is thought to be by a rolling circle mechanism. They are associated with a broad range of sources including invertebrates, humans, other mammals, birds and fungi.

Lispiviridae [ICTV 2018] (Order *Mononegavirales*, ICTV 1990) family has a single genus, *Arlivirus* [ICTV 2015] with six species, *Lishi arlivirus* [ICTV 2015] being the type species. Viruses in this family have a –ve ssRNA genome and are associated with spiders. Based on the sequence from Lishi spider virus 1 from China the lispivirus genome is bisegmented –ve ssRNA. The larger segment is 7,051 ntd in size and encodes the L protein, while the smaller one, 4,426 ntd in size encodes the G, N and VP4 proteins. As the only evidence provided for this virus is through genome sequencing, there is no further information on, for example virion morphology, replication cycle, transmission or host range.

Phasmaviridae [ICTV 2016] (Order *Bunyavirales*, ICTV 2016), genus *Sawatrivirus* [ICTV 2018] with a single and type species, *Sanxia sawatrivirus* [ICTV 2018] from the Sanxia water strider-2 virus is associated with water strider insects. Genus *Orthophasmavirus* [ICTV 2016] viruses infect mosquitoes, cockroaches and midges. The genome is tripartite -ve ssRNA with S (for N protein), M (for G protein) and L (for L protein, RdRp) segments with sizes of 1.7, 5.2 and 6.8 kb, respectively. Virions in the family are spherical about 80 nm diameter and are enveloped. Genus *Jonvirus* [ICTV 2018] and *Feravirus* (from Ferak Virus) [ICTV 2018] viruses were initially isolated from *Aedes* cell line C6/36 from pools of mosquitoes collected in 2004 in the Tai National forest from Côte d'Ivoire. Both show typical bunyavirus morphology, except that Jonvirus also exhibits as enveloped rods measuring 60 nm by up to 600 nm (the genus name *Jonvirus* refers to Jonchet, a French pick-up sticks game where the shape of the sticks mimicked this unusual morphology).

Polycipiviridae [ICTV 2017] (Order *Picornavirales*, ICTV 2008) has three genera, *Chipocyclovirus*, *Hupolycivivirus* and *Sopolycivivirus*, [ICTV 2017]. The viruses are thought to be about 33 nm in diameter, icosahedral in shape with pseudo $T = 3$ symmetry and a 10–12 kb monopartite +ve ssRNA genome. Genomes of chipocycloviruses were isolated from midges like *Chironomus riparias*, while those of sopolyciviruses were isolated from a variety of ant species. Most of the information on this family is based on large scale sequencing of environmental samples and metagenomic approaches so little is known of the biology of its members, like morphology and replication strategy.

Phenuiviridae [ICTV 2016] (Order *Bunyavirales*, ICTV 2016) viruses contain a tripartite –ve ssRNA genome in an RNP structure within an enveloped virion, 80–120 nm in diameter and surface proteins organized in a $T = 12$ icosahedral structure. Some members of the family infect insects. Viruses from genus *Wenivirus* [ICTV 2018] with a single and type species *Shrimp wenivirus* [ICTV 2018] is associated with shrimp. Genus *Goukovirus* [ICTV 2016] (from Gouléako Village where it was first found) viruses and Genus *Phasivirus* [ICTV 2016] viruses are found exclusively in mosquitoes. Genus *Phlebovirus* [ICTV 1981] viruses infect mosquito vectors of some viruses in this genus.

Tospoviridae [ICTV 2018] (Order *Bunyavirales*) is a family of plant viruses. One genus in the family, *Orthotospovirus* [ICTV 2016] which includes the Tomato spotted wilt tospoviruses consists of viruses which replicate in the midgut of thrip vectors of the virus. After replication, the virus moves to the salivary glands from where it can be transmitted to plants. Like other bunyaviruses, these viruses are pseudocircular in shape with a tripartite -ve ssRNA genome of 3, 4.8 and 8.8 kb in size.

Xinmoviridae [ICTV 2018] (Order *Mononegavirales*, ICTV 1990) family has a single genus, *Anphevirus* [ICTV 2015] with seven species, with *Xincheng anphevirus* being the type species. The sequence of the viral genome of Xincheng mosquito virus is a monopartite 12,774 ntd

Table 2 Families of viruses with an invertebrate arthropod host and mammalian host, often causing human diseases

Family	Genus, type species (number of species in genus)	Genome type and size	Morphology	Invertebrate vector
<i>Asfarviridae</i> (<i>Asfuvirales</i>) ^a	<i>Asfivirus</i> , African swine fever virus (1)	dsDNA, linear, 170 to 194 kb	Spherical, 175 to 215 nm diameter, with icosahedral capsid core (170 to 190 nm diameter), internal and external envelope	Soft ticks (<i>Ornithodoros</i>), psyllids, odonates
<i>Flaviviridae</i> (<i>Amarillovirales</i>)	<i>Flavivirus</i> , Yellow fever virus (53)	+ ve ssRNA, 9 to 13 kb	Spherical, 40 to 60 nm diameter, enveloped	Mosquitoes, ticks
<i>Nairoviridae</i> (<i>Bunyavirales</i>)	<i>Orthonairovirus</i> , Dugbe orthonairovirus (15)	– ve ssRNA, tripartite, 1 to 3 kb; 3.2 to 4.9 kb; 6.8 to 12 kb	Spherical, 80 to 120 nm, enveloped	Ticks, crustaceans, nematodes
<i>Nyamiviridae</i> (<i>Mononegavirales</i>)	<i>Nyavirus</i> , Nyamanini nyavirus (3)	– ve ssRNA, 1 or 2 segments, total 12.2 kb	Spherical, 100 to 130 nm, enveloped	Ticks
<i>Orthomyxoviridae</i> (<i>Articulavirales</i>)	<i>Socycivirus</i> , Soybean cyst nematode socycivirus (1)	– ve ssRNA, tripartite, 6 to 7 segs 0.9 to 2.3 kb	Spherical, 80 to 120 nm diameter	Mosquitoes, sand flies
<i>Peribunyaviridae</i> (<i>Bunyavirales</i>)	<i>Quaranjavirus</i> , Quarafil quaranjavirus (2)	– ve ssRNA, tripartite, 1, 4 and 6 to 8 kb	Pleomorphic to spherical, 80 to 120 nm diameter	Mosquitoes, cockroaches water striders, drosophila, midges
<i>Phasmaviridae</i> (<i>Bunyavirales</i>)	<i>Thogotovirus</i> , Thogoto thogotovirus (2)	– ve ssRNA, tripartite, 1.5, 4.2 and 6.8 kb	Spherical, 80 to 120 nm diameter	Mosquitoes
<i>Phenuiviridae</i> (<i>Bunyavirales</i>)	<i>Orthobunyavirus</i> , Bunyamwera orthobunyavirus (88)	– ve ssRNA' tripartite 1,7, 3.2, 6.4 kb	Spherical, 80 to 120 nm diameter, surface envelope with T=12 symmetry	Mosquitoes, midges, gnats, sand flies, ticks
<i>Reoviridae</i> <i>Sedoreovirinae</i> (<i>Reovirales</i>)	<i>Orbivirus</i> , Bluetongue virus (22)	dsRNA, segmented	Orbi-icosahedral T=2/T=13, 80 nm diameter	Ticks, mosquitoes
	<i>Seadornaviru,s</i> Banna virus (3)	Orbi-10 segs, 0.8 to 4.0 kb Seado-12 segs 0.9 to 4.0 kb	Seado-icosahedral T=2/T=13, 60 to 70 nm diameter	
<i>Reoviridae</i> <i>Spinavirinae</i> (<i>Reovirales</i>)	<i>Coltivirus</i> , Colorado tick fever coltivirus (5)	12 segments, 0.7 to 4.4 kb	Icosahedral T=2/T=12 60 to 80 nm diameter	
<i>Rhabdoviridae</i> (<i>Mononegavirales</i>)	<i>Curiovirus</i> , Curionopolis curiovirus (41)	– ve ssRNA, 10.7 to 16.1 kb	Bullet shaped, rod shaped nucleocapsid 45 to 100 nm × 100 to 430 nm with matrix and envelope	Mosquitoes, midges, sandflies, ticks, fleas, blackflies
	<i>Ephemerovirus</i> , Bovine fever ephemerovirus (8)			Vertebrate hosts, mammalian, avian, lizards, reptiles
	<i>Hapavirus</i> , Flanders hapavirus (16)			
	<i>Ledantevirus</i> , Le Dantec ledantevirus (16)			
	<i>Sripuvirus</i> , Niakha sripuvirus (8)			
	<i>Tibrovirus</i> , Tibrogargan tibrovirus (7)			
	<i>Vesiculovirus</i> , Indiana vesiculovirus (16)			
<i>Togaviridae</i> (<i>Martellivirales</i>)	<i>Alphavirus</i> , Sindbis virus (31)	+ ve ssRNA, 10 to 12 kb	Spherical, 65 to 70 nm enveloped	Mosquitoes, ticks, lice

^aThe taxon Order is in brackets below the family or subfamily name. The number of species in each genus is given in brackets after the genus name.

Table 3 Some common arboviruses, taxonomy, listed by Order, vertebrate host and their invertebrate vectors

Arbovirus	Order	Family	Genus	Vertebrate host	Insect and arachnid vectors
Bagaza virus	Amarillovirales	Flaviviridae	Flavivirus	Game birds	Culex mosquitoes, midges
Entebbe bat virus	Amarillovirales	Flaviviridae	Flavivirus	bats	Mosquitoes
Dengue hemorrhagic fever virus	Amarillovirales	Flaviviridae	Flavivirus	Humans, primates	Aedes aegypti and Aedes albopictus, mosquitoes,
Japanese encephalitis virus	Amarillovirales	Flaviviridae	Flavivirus	Humans, wading birds, pigs, horse	Culex mosquitoes
Kyasanur forest disease virus	Amarillovirales	Flaviviridae	Flavivirus	Camels, rodents, monkeys	Hyalomma hard ticks
Louping ill virus	Amarillovirales	Flaviviridae	Flavivirus	Grouse, sheep	Ixodes ricinus ticks
Murray Valley encephalitis virus	Amarillovirales	Flaviviridae	Flavivirus	Humans, Ardeid water birds	Culex annulirostris mosquitoes
Powassan virus	Amarillovirales	Flaviviridae	Flavivirus	Humans	Ixodes and Dermacentor ticks
Spondweni virus	Amarillovirales	Flaviviridae	Flavivirus	Humans	Aedes circumluteolus mosquitoes
Saint Louis encephalitis virus	Amarillovirales	Flaviviridae	Flavivirus	Humans, birds	Culex pipiens mosquitoes
Tick-borne encephalitis virus	Amarillovirales	Flaviviridae	Flavivirus	Humans, rodents	Ixodidae ticks
West Nile virus	Amarillovirales	Flaviviridae	Flavivirus	Humans, birds, horse	Culex mosquitoes
Yellow fever virus	Amarillovirales	Flaviviridae	Flavivirus	Humans, primates	Aedes, Hemagogus and Sabethes mosquitoes
Zika virus	Amarillovirales	Flaviviridae	Flavivirus	Humans, primates	Aedes mosquitoes
Thogoto virus	Articulavirales	Orthomyxoviridae	Thogotovirus	Humans, wild and domestic vertebrates	Hyalomma hard ticks
African swine fever virus	Asfarvirales	Asfarviridae	Asfivirus	Pig, warthog, bushpig	Ornithodoros soft ticks,
Bunyamwera virus	Bunyavirales	Peribunyaviridae	Orthobunyavirus	Humans	Aedes aegypti mosquitoes
California encephalitis virus	Bunyavirales	Peribunyaviridae	Orthobunyavirus	Humans	Mosquitoes
Crimean-Congo fever virus	Bunyavirales	Nairoviridae	Orthonairovirus	Humans, wild and domestic animals	Hyalomma hard ticks
Jamestown Canyon virus	Bunyavirales	Peribunyaviridae	Orthobunyavirus	Humans, white tailed deer	Mosquitoes
La Crosse virus	Bunyavirales	Peribunyaviridae	Orthobunyavirus	Humans, chipmunks, squirrels	Ochlerotatus triseriatus mosquitoes
Oropouche fever virus	Bunyavirales	Peribunyaviridae	Orthobunyavirus	Humans, primates, sloth, marsupials, birds	Aedes serratus and Culex quinquefasciatus Mosquitoes, midges
Heartland virus	Bunyavirales	Phenuiviridae	Bandavirus	Humans, white tailed deer raccoons	Amblyomma americanum ticks
Rift Valley fever virus	Bunyavirales	Phenuiviridae	Phlebovirus	Humans, ovine, goat, sheep ruminants	Culex tritaeniorhynchus and Aedes vexans mosquitoes
Sandfly fever virus	Bunyavirales	Phenuiviridae	Phlebovirus	Humans	Psychodidae sandflies
Toscana virus	Bunyavirales	Phenuiviridae	Phlebovirus	Humans, dogs, goats	Phlebotomus sandflies
Barmah forest fever virus	Martellivirales	Togaviridae	Alphavirus	Humans, marsupials	Aedes vigilax and Culex annulirostris mosquitoes
Chikungunya	Martellivirales	Togaviridae	Alphavirus	Humans, primates	Aedes albopictus, Aedes aegypti mosquitoes
Eastern equine encephalitis virus	Martellivirales	Togaviridae	Alphavirus	Humans, birds	Culiseta melanura and Cs. morsitans mosquitoes
Mayaro fever virus	Martellivirales	Togaviridae	Alphavirus	Humans, primates	Haemagogus mosquitoes
O'nyong -nyong	Martellivirales	Togaviridae	Alphavirus	Humans	Anopheles mosquitoes
Ross River virus	Martellivirales	Togaviridae	Alphavirus	Humans, macropods	Ochlerotatus vigilax and Culex mosquitoes
Sindbis disease virus	Martellivirales	Togaviridae	Alphavirus	Humans, passerine birds	Aedes, Culex and Culiseta mosquitoes
Venezuelan equine encephalitis virus	Martellivirales	Togaviridae	Alphavirus	Humans, horses, donkeys, zebras	Aedes albopictus and Culex mosquitoes
Western equine encephalitis virus	Martellivirales	Togaviridae	Alphavirus	Humans, birds	Culex tarsalis mosquitoes
Bovine ephemeral fever virus	Mononegavirales	Rhabdoviridae	Ephemerovirus	Bovine	Culicoides oxystoma and nipponensis biting midges
Vesicular stomatitis virus	Mononegavirales	Rhabdoviridae	Vesiculovirus	Humans, bovine, swine, equine	Culicoides midges, sandflies, blackflies
African horse sickness virus	Reovirales	Reoviridae Sedoreovirinae	Orbivirus	Horses, donkeys, mules, zebras	Culicoides imicola midges
Banna virus	Reovirales	Reoviridae Sedoreovirinae	Seadornavirus	Humans, pigs, cattle	Culex, Aedes and Anopheles mosquitoes
Bluetongue virus	Reovirales	Reoviridae Sedoreovirinae	Orbivirus	Goats, cattle, wild ruminants	Culicoides imicola and Culicoides variipennis midges
Colorado tick fever virus	Reovirales	Reoviridae Spinareovirinae	Coltivirus	Rodents, ground squirrels	Dermacentor ticks
Equine encephalosis virus	Reovirales	Reoviridae Sedoreovirinae	Orbivirus	Horses, donkeys, zebras	Culicoides midges

Table 4 Viruses of invertebrate vectors of plant viruses (The Order is given in brackets below the family name. The number of species in the genus is given by the number in brackets after the genus name)

Family	Genus (type species)	Genome type and size	Morphology	Invertebrate vector (transmission type)
<i>Betaflexiviridae</i> (Quinvirinae) (Tymovirales)	<i>Carlavirus</i> , Carnation latent virus (53)	+ v ssRNA, 5.8 to 9 kb	Helical, 12 to 13 × 470 to 1,000 nm	Aphid and whitefly (non-circulative, semi-persistent)
<i>Bromoviridae</i> (Martellivirales)	<i>Alfavirus</i> , Alfalfa mosaic virus (1) <i>Cucumovirus</i> , Cucumber mosaic virus (4)	+ ve ssRNA, tripartite Alfa- 2.1 to 2.2/2.6 Cucumo- 3.0/3.2 to 3.4 kb	Alfa-bacilliform 18 × 30 to 57 nm Cucumo- icosahedral T=3, 26 to 35 nm diameter	Aphid (non-circulative, nonpersistent)
<i>Caulimoviridae</i> (Ortervirales)	<i>Caulimovirus</i> , Cauliflower mosaic virus (13) <i>Cavenovirus</i> , Cassava vein mosaic virus (2) <i>Roseadnavirus</i> , Rose yellow vein virus (1) <i>Soymovirus</i> , Soybean chlorotic mottle virus (4) <i>Tungovirus</i> , Rice tungro bacilliform virus (1)	dsDNA, open circular, 7.0 to 8.2 kb	Icosahedra T=7, 45 to 50 nm diameter or bacilliform (Tungovirus) 30 nm × 60 to 900	Aphid, mealybug, leafhopper (non-circulative, semi-persistent)
<i>Closteroviridae</i> (Martellivirales)	<i>Ampelovirus</i> , Grapevine leafroll-associated virus 3 (12) <i>Closterovirus</i> , Beet yellows virus (16) <i>Crinivirus</i> , Lettuce infectious yellows virus (14)	+ ve ssRNA Ampelo- and Closterovirus- monopartite, 15.5 to 19.3 kb Crini-bisegmented 7.8 and 9.1 kb	Ampelo- and Closterovirus- single flexuous helix, 3.4 to 3.8 × 650 to 2,200 nm Criniviruses, bipartite helices, 10 to 13 × 650 to 85/10 to 13 × 700 to 900	Aphid, whitefly, softscale mealybug (non-circulative, semi-persistent)
<i>Fimoviridae</i> (Bunyavirales)	<i>Emaravirus</i> , European mountain ash ringspot-associated emaravirus (11)	– ve ssRNA, quatrpartite, 1.3/1.6/2.0/6.9 kb	Spherical 80 to 90 nm, enveloped	Eriophyd Mite (?)
<i>Geminiviridae</i> (Gepfuvirales)	<i>Begomovirus</i> , Bean golden yellow mosaic virus (424)	+ ve ssDNA, bipartite, 2.5/2.6 kb	Twined icosahedra T=1, 22 × 38 nm	Whitefly, leaf hopper, aphid, tree hopper (circulative, non-propagative, persistent)
<i>Luteoviridae</i> (Tolivirales)	<i>Enamovirus</i> , Pea enation mosaic virus 1 (5) <i>Luteovirus</i> , Barley yellow dwarf virus (13) <i>Polerovirus</i> , Potato Leafroll virus (26)	+ ve ssRNA, monopartite, 5.3 to 5.7 kb	Icosahedral T=3, 23 to 30 nm diameter	Aphid (circulative, nonpropagative, persistent)
<i>Nanoviridae</i> (Mulpavirales)	<i>Babuvirus</i> , Banana bunchy top virus (3) <i>Nanovirus</i> , Subterranean clover stunt virus (8)	+ ve ssDNA, multipartite, 6 to 8 segments, 1 kb each	Icosahedra T=1, 18 to 19 nm diameter	Aphid, planthopper (circulative, nonpropagative, persistent)
<i>Phenuiviridae</i> (Bunyavirales)	<i>Tenuivirus</i> , Rice stripe tenuivirus (8)	– ve and ambisense ssRNA, 4 to 5 segments	4 to 5 helical circular helices 3 to 10 nm diameter	Planthopper (circulative, propagative, persistent)
<i>Potyviridae</i> (Patatavirales)	<i>Macluravirus</i> , Maclura mosaic virus (10) <i>Potyvirus</i> , Brugmansia mosaic virus (183) <i>Tritimovirus</i> , Wheat streak mosaic virus (6) <i>Rymovirus</i> , Ryegrass mosaic virus (3) <i>Phytoreovirus</i> , Wound tumor virus (3)	+ v ssRNA, 8.0 to 12 kb	Flexuous, helix, 12 to 15 nm × 650 to 850 nm	Aphid, mite, whitefly (non-circulative, nonpersistent)
<i>Reoviridae</i> (Reovirales)		dsRNA, 12 segments, 1.1 to 4.4 kb	Icosahedral, double shell T=2/T=13, 70 nm diameter	Leafhopper (circulative, propagative, persistent)
<i>Rhabdoviridae</i> (Mononegavirales)	<i>Cytorhabdovirus</i> , Lettuce necrotic yellows cytorhabdovirus (23) <i>Alphanucleorhabdovirus</i> , Potato yellow dwarf alphanucleorhabdovirus (9)	– ve ssRNA 12.0 to 14.5 kb	Bullet to bacilliform shape, 45 to 100 nm × 130 to 350 nm, enveloped	Aphid, leafhopper, planthopper (circulative, propagative, persistent)
<i>Secoviridae</i> (Picornavirales)	<i>Nepovirus</i> , Tobacco ringspot virus (40) <i>Fabavirus</i> , Broad bean wilt virus 1 (7) <i>Sequivirus</i> , Parsnip yellow fleck virus (3)	+ ve ssRNA Nepo- bipartite, 3.9/7.5 kb Faba-bipartite, 1.58/3.4 kb Sequi-monopartite, 9 kb	Icosahedral, naked, pseudo T=3, Nepo- 28 to 30 nm diameter, Faba- 28 to 38 nm diameter Sequi- 25 to 30 nm diameter	Nepo- nematode, mite, thrip Faba- aphid, beetle, leafhopper (non-circulative) Sequi- aphid, leafhopper (circulative, propagative)
<i>Tospoviridae</i> (Bunyavirales)	<i>Orthospovirus</i> , Tomato spotted wilt orthospovirus (26)	– ve ssRNA, tripartite, 3, 4, 8, 8.k kb	Spherical enveloped, 80 to 120 nm diameter, 3 RNPs	Thrip (circulative, propagative)
<i>Virgaviridae</i> (Martellivirales)	<i>Tobravirus</i> , Tobacco rattle virus (3)	+ ve ssRNA, bipartite, 4.5/6.8 kb	Helical 22 × 180 to 215; 22 × 46 to 115	Nematodes

Note: Adapted from Whitfield, A.E., Falk, B.W., Rotenberg, D., 2015. Insect vector-mediated transmission of plant viruses. *Virology* 479–480, 278–289. Dietzgen, R.G., Mann, K.S., Johnson, K.N., 2016. Plant virus-insect vector interactions: Current and potential future research directions. *Viruses* 8, 303. doi:10.3390/v8110303.

-ve ssRNA encoding N, G, L and an unknown ORF protein. These viral genomes have been associated with *Anopheles* mosquitoes, *Drosophila* and orthoptera. However as only sequence data is available, no other characteristics, like morphology or replication are known.

Arboviruses in Their Invertebrate Vectors

Only superficially, if at all, covered in this article are viruses that are arboviruses involved in transmitting arboviral disease, like yellow fever, Dengue, Zika even though part of the life cycle involves replication in the insect vectors. Nevertheless summary Table 2 provides a summary of the families of arboviruses and their insect vectors and summary Table 3 is for some better known arboviruses, their taxonomy and vector species. The reader is referred to other chapters in this Encyclopedia which are devoted to more detailed coverage of these viruses.

Plant Viruses in Their Invertebrate Vectors

While many plant viruses are vectored by invertebrates (Table 4), they are not further discussed here but the reader is referred to the corresponding chapters on plant viruses. In terms of the nature of transmission of plant viruses, these can be defined as circulative (virus enters the insect body and circulates, usually through the hemolymph, prior to transmission), non circulative (virus does not enter the insect body but is transmitted physically through contaminated stylet or foregut of the insect) propagative (virus enters the insect and replicates in the insect prior to transmission, usually through the salivary glands) and non propagative (virus enters the insect, usually transported to salivary glands, but does not replicate in it prior to transmission). Transmission of plant viruses can also be defined as persistent, in which the virus is taken up by the insect following circulative means and invade the salivary gland, semi-persistent in which viruses are internalized and retained in the insect foregut, but do not enter the tissues and non persistent in which viruses are retained on the distal tip of the stylet only until it is released during feeding. A summary of some of these invertebrate-vector transmitted plant viruses is provided in Table 4.

Retrotransposons Associated With Invertebrates

Three virus families associated with invertebrates, all in order *Ortervirales* (for reverse transcribing viruses), are actually transposons. While isolated mostly from plants, retrotransposons from these three families were also found to be associated with invertebrates. Retrotransposons in family *Pseudoviridae* [ICTV 1998] genus *Hemivirus* [ICTV 1998], type species *Drosophila melanogaster copia virus* are associated with *Drosophila* and mosquitoes. Retrotransposons in family *Balpaoviridae* [ICTV 2017] with genus *Semotivirus* [ICTV 2017], type species *Ascaris lubricoides virus* are associated with roundworm, lepidoptera and *Drosophila*. Retrotransposons in family *Metaviridae*, [ICTV 1998] genus *Errantivirus*, [ICTV 1998], type species *Drosophila melanogaster Gypsy virus* and genus *Metavirus* [ICTV 1998], type species *Saccharomyces cerevisiae Ty3 virus* have retrotransposons associated with *Drosophila*, *Trichoplusia ni* and *Bombyx mori*. The genome of the erantivirus *Trichoplusia ni* TED virus is 7,510 bp.

Acknowledgments

Much of the information for this summary was gleaned from preprints of the articles on Invertebrate Viruses prepared for this Encyclopedia of Virology 4th edition. I acknowledge the contributions of these authors for being able to use their information. I leaned heavily on the ICTV web sites for the wealth of knowledge provided on up to date taxonomy (some of which changed as I was writing this article), summary articles on different taxa in the 10th report of "The ICTV Report on Virus Classification and Taxon Nomenclature" and the numerous "proposals" that accompany the "history" provided through the taxonomy site.

The articles in the main body of this volume do not cover all the families of viruses infecting invertebrates covered in this summary. In part that is because often not much is known, or in some instances only sequence data is available.

I hope the reader will find the summary Tables (Tables 1–4) of use, as the tables attempt to collate the wealth of the diversity of invertebrate viruses and their taxa. I have made every effort for accuracy, but considering the breadth of information, it is easy to overlook some important information or err on transcription (the writing kind). For these I apologize.

See also: Ascoviruses (*Ascoviridae*). Baculoviruses: General Features (*Baculoviridae*). Bidsenoviruses (*Bidnaviridae*). Bunyaviruses of Arthropods (*Mypoviridae*, *Nairoviridae*, *Peribunyaviridae*, *Phasmaviridae*, *Phuniviridae*, *Wupedeviridae*). Dicistroviruses (*Dicistroviridae*). Entomobirnaviruses (*Birnaviridae*). Hytrosaviruses (*Hytrosaviridae*). Iflaviruses (*Iflaviridae*). Iridoviruses of Invertebrates (*Iridoviridae*). Mesoniviruses (*Mesoniviridae*). Nimaviruses (*Nimaviridae*). Nodaviruses of Invertebrates and Fish (*Nodaviridae*). Nudiviruses (*Nudiviridae*). Parvoviruses of Invertebrates (*Parvoviridae*). Poxviruses of Insects (*Poxviridae*). Reoviruses of Invertebrates (*Reoviridae*). Reoviruses (*Reoviridae*) and Their Structural Relatives. Rhabdoviruses of Insects (*Rhabdoviridae*). Sarthroviruses (*Sarthroviridae*). Solinviviruses (*Solinviviridae*). Tetraviruses (*Alphatetraviridae*, *Carmotetraviridae*, *Permutotetraviridae*)

Further Reading

- Bateman, K.S., Stentiford, G.D., 2017. A taxonomic review of viruses infecting crustaceans with an emphasis on wild hosts. *Journal of Invertebrate Pathology* 14, 86–110.
- Bonning, B.C., 2019. *Insect Molecular Virology: Advances and Emerging trends*. Caister Academic Press.
- Dietzgen, R.G., Mann, K.S., Johnson, K.N., 2016. Plant virus-insect vector interactions: Current and potential future research directions. *Viruses* 8, 303. doi:10.3390/v8110303.
- Kibenge, S.B., Godoy, M.G., 2016. *Aquatic Virology*. New York Academic Press.
- Whitfield, A.E., Falk, B.W., Rotenberg, D., 2015. Insect vector-mediated transmission of plant viruses. *Virology* 479–480, 278–289.
- Williams, T., Bergoin, M., van Oers, M., 2017. Diversity of large DNA viruses of invertebrates. *Journal of Invertebrate Pathology* 147, 4–22.
- Young, P.R., 2018. Arboviruses: A family on the move. In: Hilgenfeld, R., Vasudevan, S. (Eds.), *Dengue and Zika: Control and Antiviral Treatment Strategies*. *Advances in Experimental Medicine and Biology* 1062. Springer. doi:10.1007/978-981-10-8727-1_1.

Relevant Websites

<https://talk.ictvonline.org/information/w/faq/386/how-to-write-virus-and-species-names>

How to write virus, species, and other taxa names.

<https://talk.ictvonline.org>

International Committee on Taxonomy of Viruses (ICTV).

https://talk.ictvonline.org/ictv-reports/ictv_online_report

The ICTV Report on Virus Classification and Taxon Nomenclature.

<https://viralzone.expasy.org>

ViralZone root.

<https://talk.ictvonline.org/taxonomy>

Virus Taxonomy: 2019 Release.

International Committee on Taxonomy of Viruses (ICTV).

Ascoviruses (*Ascoviridae*)

Sassan Asgari, The University of Queensland, Brisbane, QLD, Australia

Dennis K Bideshi, California Baptist University, Riverside, CA, United States and University of California, Riverside, CA, United States

Yves Bigot, INRAE – French National Research Institute for Agriculture, Food and Environment, Nouzilly, France

Brian A Federici, University of California, Riverside, CA, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of B.A. Federici, Y. Bigot, Ascoviruses, In Encyclopedia of Virology (Third Edition), Edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00347-2.

Glossary

Apoptosis Genetically programmed cell death.

Apoptotic bodies Cell vesicles resulting from apoptosis.

Caspase Protease that activates a major portion of programmed cell death.

Endoparasitic wasps Species of insect parasites belonging to the order Hymenoptera, which lay their eggs in insects where the wasp larvae develop.

Per os infection Infection by feeding.

Programmed cell death Genetically programmed cascade proteases and nucleases that cleave DNA and proteins within a cell leading to its death.

Reniform Shaped like a kidney.

Transovarial transmission Transmission of virus inside the egg.

Virion-containing vesicles Vesicles containing virions formed by ascoviruses by rescue of apoptotic bodies induced by ascovirus infection.

Introduction

The family *Ascoviridae* is one of the newest families of viruses, established in 2000 to accommodate several species of a newly recognized type of DNA virus that attacks larvae of insects of the order Lepidoptera. It consists of two genera, *Ascovirus* and *Toursvirus*; for practicality, here we refer to the collective members of the family *Ascoviridae* as ascoviruses. Viruses of this family produce large, enveloped virions, measuring 130 nm in diameter by 300–400 nm in length, and when viewed by electron microscopy have a reticulated surface appearance. They are typically bacilliform or reniform in shape, and contain a circular double-stranded DNA genome that, depending on the species, ranges from 119 to 200 kbp. Whereas the virions of ascoviruses are structurally complex like those of other large DNA viruses that attack insects, such as those of iridoviruses (family *Iridoviridae*) and entomopoxviruses (family *Poxviridae*), they differ from these in two significant aspects. First, ascoviruses are transmitted from diseased to healthy lepidopteran larvae or pupae by female endoparasitic wasps when these lay eggs in their hosts. Second, ascoviruses have a unique cell biology and cytopathology in which shortly after infecting a cell, they induce apoptosis and then rescue the developing apoptotic bodies and convert these into virion-containing vesicles. This aspect of viral reproduction apparently evolved to disseminate virions to the larval hemolymph (blood) where they could contaminate the ovipositors of female wasps so that the virus could be transmitted to new hosts. Ascoviruses appear to occur worldwide, wherever there are endoparasitic wasps and larvae of species belonging to the lepidopteran family Noctuidae. However, as these viruses have been discovered relatively recently and their signs of disease are not commonly known in the scientific community, relatively few ascovirus species have been described.

History

The first ascoviruses were discovered during the late 1970s in southern California where they were found causing disease in larvae of moths belonging to the lepidopteran family Noctuidae. Diseased larvae were recognized by the presence of hemolymph that was very white and opaque, in marked contrast to the hemolymph of healthy larvae which is translucent and slightly green (Fig. 1). The color and opacity of the hemolymph in diseased larvae was shown to be due to the presence of high concentrations of vesicles that contained virions (Fig. 2). The white hemolymph and virion-containing vesicles are diagnostic for the disease, and the name for this group, ascoviruses (derived from the Greek *asco* meaning “sac”), was chosen on the basis of the latter characteristic. Since the discovery of the first ascovirus, ascoviruses have been isolated as the cause of disease in many species of noctuid larvae. In addition, an ascovirus that attacks the pupal stage of a species belonging to the family Yponomeutidae was discovered in the 1990s in France.

Distribution and Taxonomy

With respect to distribution, ascoviruses have been reported from the United States, Europe, Australia, and Indonesia, and it is highly probable that they occur worldwide. This is because their most common hosts, larvae of lepidopteran species belonging to

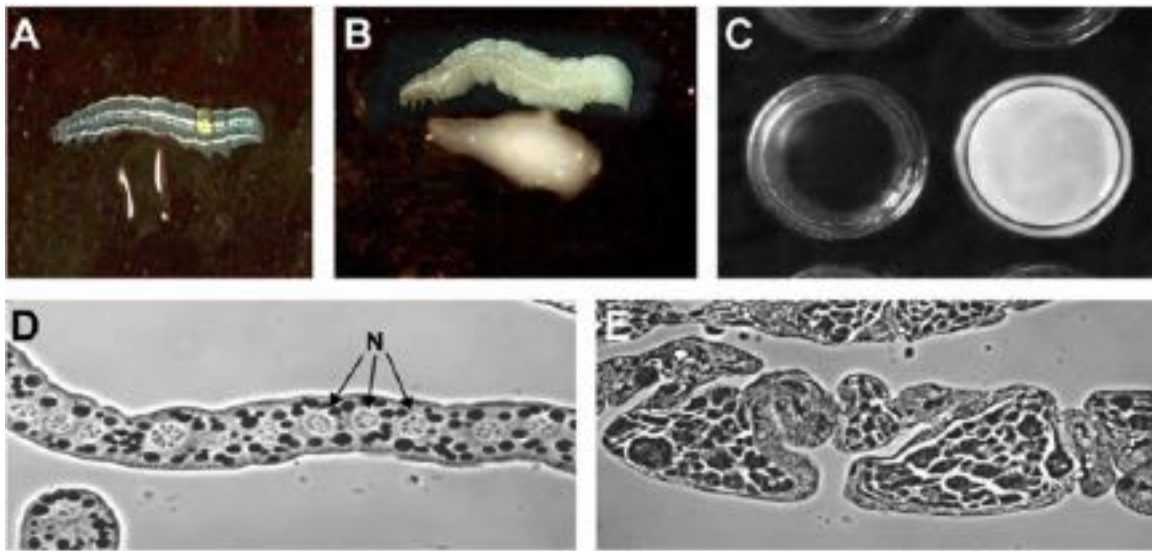


Fig. 1 Major characteristics of the disease typically caused by ascoviruses in lepidopteran larvae. (A) and (B) Healthy and ascovirus-infected larvae, respectively, of the cabbage looper, *Trichoplusia ni*, infected with *Trichoplusia ni* ascovirus. Note the opaque white hemolymph in the infected larva. (C) Spot plate containing hemolymph from healthy (left) and infected larvae (right). (D, E) Sections through lobes of fat body from a healthy and infected larva, respectively, of the fall armyworm, *Spodoptera frugiperda*, infected with *Spodoptera frugiperda* ascovirus. Note the greatly hypertrophied cells in the fat body of the infected larva. The cells in most of this tissue have already cleaved into viral vesicles. N, nuclei.

the family Noctuidae, the largest family of the order Lepidoptera, as well as their most common vectors, endoparasitic wasps of the families Braconidae and Ichneumonidae, are distributed throughout the world. Although only a few ascovirus species have been described to date, there are probably many, including variants, that occur worldwide. Thus, given the common occurrence of their hosts and vectors, it is possible that ascoviruses are very common insect viruses. That they have not been discovered in large numbers, for example like baculoviruses (family *Baculoviridae*), is probably because they cause a chronic disease with few easily detectable signs, making it difficult for individuals not familiar with the disease to recognize diseased larvae in field populations.

At present, five species of ascoviruses are officially recognized based on a combination of properties including the relatedness of 28 key genes coding among others for the DNA polymerase and major capsid protein, the percent identity between their sequenced genomes, their lepidopteran host range, and their tissue tropism (Table 1). They are split into two genera, *Ascovirus* and *Toursvirus*. The type species of the *Ascovirus* genus is *Spodoptera frugiperda* ascovirus 1a (SfAV-1a), with the other species being *Trichoplusia ni* ascovirus 2a (TnAV-2a) and *Heliothis virescens* ascovirus 3a (HvAV-3a). To date, the *Toursvirus* genus contains only one virus species originally called *Diadromus pulchellus* ascovirus 4a (DpAV-4a), and now named *Diadromus pulchellus* *toursvirus* (DpTV-4a). The Arabic numeral reflects the order in which each species was formally recognized, whereas the lower case letter indicates variants of the type species. Variants from the type species are recognized by different consecutive lower case numbers; for example, TnAV-2b and 2c would represent two different variants of TnAV-2a recognized subsequently. Herein, these viruses are referred to by their acronyms without the numerical and lower case suffix.

Members of the *Ascovirus* genus have been isolated from many more insect species than those listed above, but these isolates have turned out to be variants of known ascoviruses, and therefore they have not been named after the host from which they have been isolated. For example, ascoviruses related to TnAV and HvAV have been isolated from noctuid species such as *Autographa precationis*, *Helicoverpa zea*, *Helicoverpa armigera*, and *Helicoverpa punctigera*; however, they do not bear the name of their host of isolation. What this implies is that ascoviruses belonging to the TnAV and HvAV species have a broad and overlapping host range among different noctuid species, although this has only been tested experimentally to a limited extent.

Virion Structure and Composition

Depending on the species, the virions of ascoviruses are either bacilliform or reniform in shape, with complex symmetry, and very large, measuring about 130 nm in diameter by 300–400 nm in length. Although variations occur in the virion structure among various ascovirus species, their structural components appear to be essentially the same. Ascovirus virions consist of an inner particle surrounded by an outer envelope (Fig. 2). The inner particle is complex containing a DNA/protein core as well as an apparent internal lipid bilayer surrounded by a distinctive layer of protein subunits. Thus, the virion appears to contain two lipid membranes: one associated with the inner particle and the other forming the lipid component of the envelope. In negatively stained preparations, virions have a distinctive reticulate appearance, which is thought to be due to superimposition of subunits on the surface of the internal particle with those in the envelope.

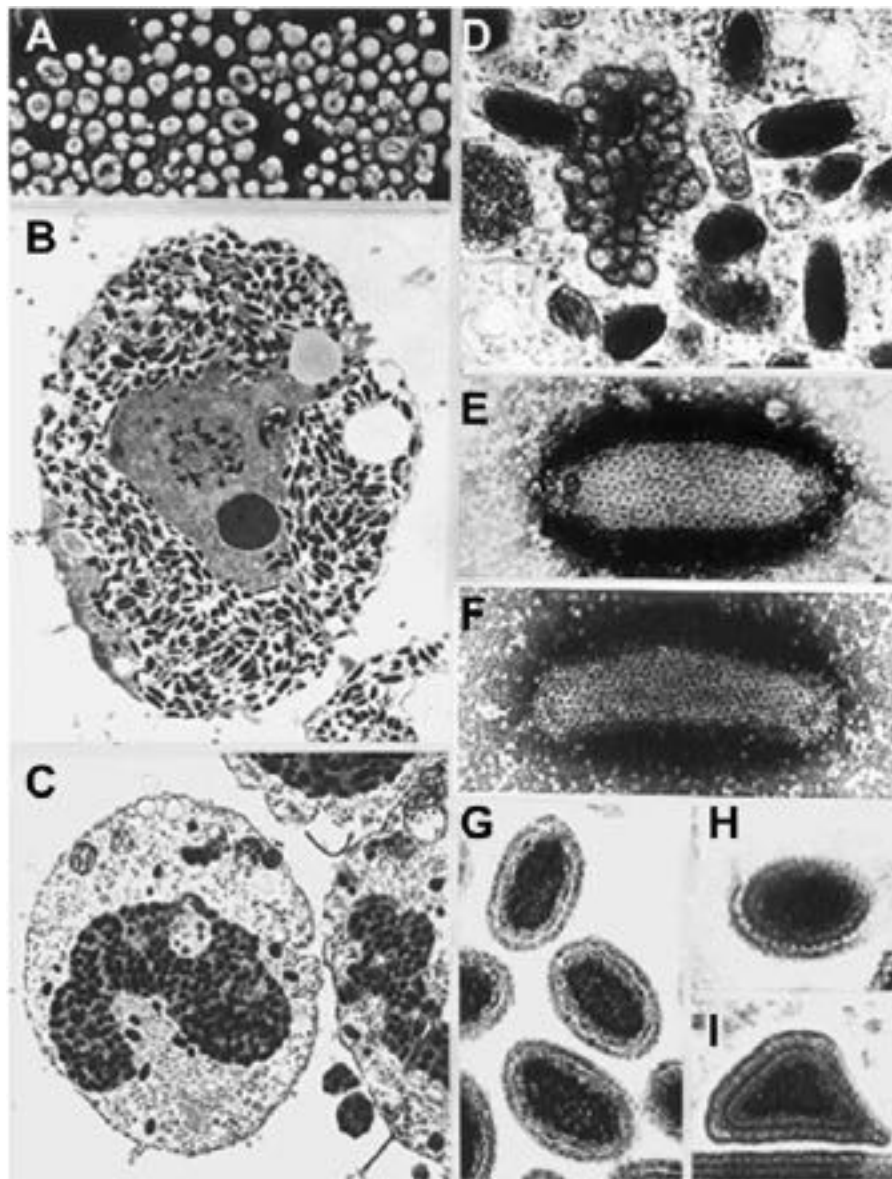


Fig. 2 Structural and morphological characteristics of ascovirus virions and virion-containing vesicles. (A) Wet mount preparation viewed with phase microscopy of hemolymph from a *Spodoptera frugiperda* larva infected with *Spodoptera frugiperda* ascovirus (SfAV). The spherical refractile bodies are virion-containing vesicles. (B, C) Electron micrographs of ultrathin sections through viral vesicles produced by *Trichoplusia ni* ascovirus (TnAV) and SfAV, respectively. (D) Matrix of the occlusion body produced by SfAV. The occlusion consists of virions, protein, and small spherical vesicles. (E, F) Negatively stained virions of SfAV and TnAV, respectively. Note the reticulate appearance of the virions. (G, H) Electron micrographs of ultrathin cross sections through inner particles of SfAV after formation (G) and during envelopment (H). (I) Ultrathin cross section through a fully developed virion of TnAV.

As indicated by the size and complexity of the virions, the genome of ascoviruses is large, and consists of a single molecule of double-stranded circular DNA. The genomes of SfAV, TnAV, HvAV, or variants of these, in the genus *Ascovirus*, and DpTV-4a the sole virus in the genus *Toursvirus* have been sequenced (Table 1). The SfAV-1a genome is 157 kbp and codes for at least 120 proteins, whereas the TnAV genomes are slightly larger, 174–186 kbp, and code for at least 165–178 proteins. The genomes of HvAV variants range from 186 to 200 kbp and code for approximately 180–190 potential proteins. The DpTV-4a genome is 119 kb in size and encodes about 119 proteins. Based on gel analyzes, ascovirus virions contain more than 12 structural polypeptides ranging in size from 12 to 200 kDa. Furthermore, proteomic analysis revealed that SfAV-1a virion is composed of at least 21 proteins, whereas similar analyzes identified at least 67 proteins in the virion of HvAV-3i. The marked difference in the virion protein profiles of SfAV-1a and HvAV-3i could be due to inadvertent contamination with non-structural proteins or a higher level of structural complexity in the HvAV-3i virion. Regardless, the two major virion proteins in ascovirus virions are, invariably, the major capsid protein and members of an unusual large cationic protein family (ascovirus P64 homologs) that functions in

Table 1 Members in the genus *Ascovirus* and *Toursvirus* belonging to the family *Ascoviridae*

Species name ^a	Virus abbreviation	Genome (bp) ^b /ORFs ^c	Accession number
<i>Ascovirus</i>			
<i>Spodoptera frugiperda</i> ascovirus 1a	SfAV-1a	156,922/123	[AM3988432]
<i>Trichoplusia ni</i> ascovirus 2a	TnAV-2a	N/A ^d	[AJ312707]
	TnAV-6a	174,059/165	[DQ517337.1]
	TnAV-6b	185,664/178	[KY434117]
<i>Heliothis virescens</i> ascovirus 3a	HvAV-3a	N/A ^d	[AJ279817.1]
	HvAV-3e	186,262/180	[EF133465.1]
	HvAV-3f	198,157/190	[KJ755191.1]
	HvAV-3g	199,721/194	[JX491653.1]
	HvAV-3h	190,519/181	[KU170628.1]
	HvAV-3i	185,650/188	[MF781070.1]
	HvAV-3j	191,718/189	[LC332918.1]
<i>Toursvirus</i>			
<i>Diadromus pulchellus</i> toursvirus 4a	DpTV-4a	119,343/119	[CU469068.1]
Tentative species			
<i>Spodoptera exigua</i> ascovirus 5a	SeAV-5a		
<i>Spodoptera exigua</i> ascovirus 6a	SeAV-6a		
<i>Helicoverpa armigera</i> ascovirus 7a	HaAV-7a		
<i>Helicoverpa punctigera</i> ascovirus 8a	HpAV-8a		

^aRecognized type species (bold text) and variants of type species (plain text).^bComplete genome sequence; bp, base pairs.^cPutative number of open reading frames (ORFs).^dComplete genome sequence not available.

condensing and encapsidating ascovirus genomic DNA. In addition to proteins and the DNA genome, the presence of an envelope as detected by electron microscopy, as well as experiments with detergents and organic solvents, indicate that virions contain a substantial lipid component. And, as in other enveloped viruses of eukaryotes, it is likely that the virion also contains carbohydrate in the form of glycoproteins, though at present none have been identified.

Transmission and Ecology

One of the most interesting features of ascoviruses is that their transmission from host to host appears to be dependent on their being vectored by female endoparasitic wasps (order Hymenoptera). Ascoviruses are extremely difficult to transmit *per os*, with typical infection rates averaging less than 15% even when larvae are fed as many as 10^5 virion-containing vesicles in a single dose. In contrast to this, infection rates for caterpillars injected with as few as 10 virion-containing vesicles are typically greater than 90%. Moreover, experiments with parasitic wasps show that they can effectively transmit ascoviruses to their noctuid hosts. For example, when females are allowed to lay eggs in ascovirus-infected noctuid caterpillars, thereby contaminating their ovipositor, and then allowed to lay eggs in healthy larvae, the majority of the latter contract ascovirus disease. Interestingly, though the parasitoid eggs hatch in their infected noctuid hosts, the parasitoid larvae die as the ascovirus disease develops in the caterpillar. Under field conditions, the prevalence of ascovirus disease in caterpillars is correlated with rates of parasitization by endoparasitic wasps. When wasps from these populations are collected in the field and allowed to oviposit in healthy caterpillars reared in the laboratory, the latter often exhibit ascovirus disease within a few days. Thus, laboratory and field studies provide sound evidence that the primary mechanism for the transmission of ascoviruses attacking noctuid larvae is through being vectored mechanically by parasitic wasps. No evidence has been found in the lepidopteran hosts for transovum or transovarial transmission.

In the case of DpTV (genus *Toursvirus*), the association of the virus with its wasp and caterpillar hosts is much more intimate. DpTV DNA is carried in wasp nuclei as a circular molecule, and small numbers of virions are produced in the oviducts of females. However, the virus does not cause noticeable pathology in the wasp host. The females lay eggs in the pupal stage of the lepidopteran host, *Acrolepiopsis assectella*, introducing small numbers of toursvirus virions along with the wasp eggs. These virions invade lepidopteran host cells, replicate, and initiate destruction of major host tissues. The wasp larva then emerges from the egg and feeds on the host tissues and toursvirus virions. The DpTV genome is carried extrachromosomally by both male and female wasps, where it is apparently transmitted from generation to generation transovarially. These observations make ascoviruses the only known group of viruses pathogenic to insects that are primarily dependent on vectors for their transmission.

Now that the characteristics of the disease are known, field studies in the southeastern United States and California are beginning to show that ascoviruses are probably the most common type of virus to occur during most of the year in populations of several important noctuid pests, including the cabbage looper, *Trichoplusia ni*, fall armyworm, *Spodoptera frugiperda*, and the corn

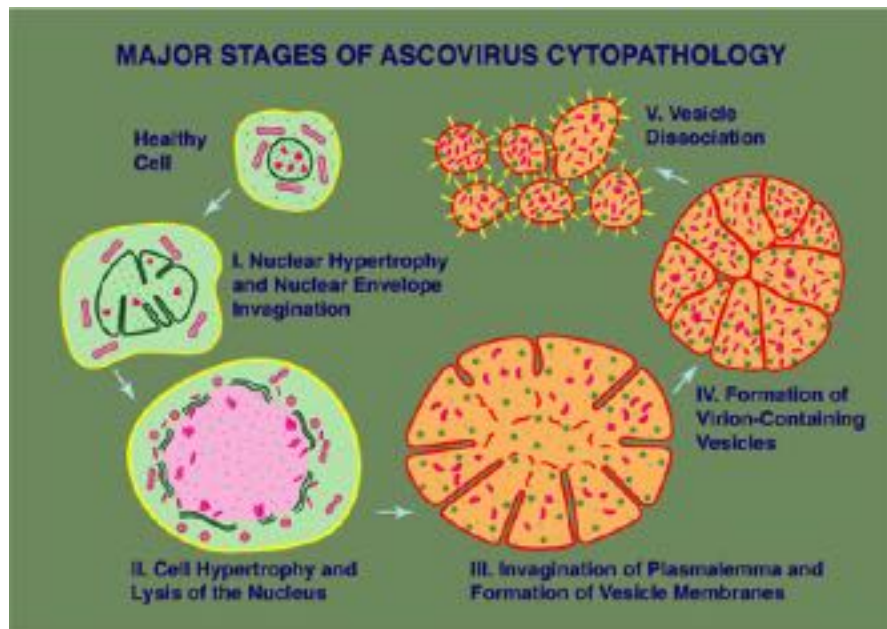


Fig. 3 Major stages of cellular pathogenesis caused by a typical ascovirus, a process that resembles apoptosis. After infection, the nucleus enlarges and the nuclear membrane invaginates, and then lyses. Subsequently, the plasmalemma of the cell invaginates and coalesces with cytoplasmic membranes, apparently formed *de novo*, thereby dividing the cell into a cluster of virion-containing vesicles. These vesicles dissociate and are liberated into the hemolymph as the basement membrane of infected tissues degenerates. Virion assembly becomes apparent as the nuclear membrane lyses, and continues throughout all subsequent stages of vesicle formation.

earworm, *Helicoverpa zea*. Prevalence rates range from 10% to 25%, depending on the species and time of the year, with the highest rates of infection, as noted above, being correlated with high levels of parasitization. In South Carolina, ascovirus infection rates as high as 60% have been reported in populations of noctuid larvae at the end of summer.

Host Range

The experimental host range of ascoviruses varies with the viral species. TnAV, HvAV, and SeAV have a broad host range and are capable of replication in a variety of noctuid species, as well as in selected species belonging to other families of the order Lepidoptera. Alternatively, the experimental host range of SfAV is limited to other species of the genus *Spodoptera*. DpTV can replicate in hymenopteran and lepidopteran hosts closely related to its natural host species, *A. assectella*. To propagate virus in the laboratory, all ascoviruses can be grown in their larval or pupal hosts. To infect caterpillars, they are injected with virus in the third, fourth, or early fifth instar, and virion-containing vesicles are harvested from the hemolymph 5–7 days later. While ascoviruses can be propagated in lepidopteran host cell lines, virus pathology usually is not as distinct as that of *in vivo*. In addition, virus replication *in vitro* does not appear to be as high as that *in vivo*.

Pathology and Pathogenesis

Signs of Disease

In both natural and laboratory environments, the initial signs of ascovirus disease are very subtle, and this probably accounts for why ascoviruses were discovered only recently. The most obvious sign of disease within 24 h of infection is a decrease in the normal rate of feeding. The feeding rate continues to slow as the disease progresses, and as a result larvae fail to gain weight or advance in development. Healthy larvae, particularly in the early stages of development, will easily quadruple their weight and size in a period of 3 or 4 days, whereas ascovirus-infected larvae cease to grow and may actually lose weight. This feature of ascovirus disease is almost impossible to detect in infected larvae in the field. However, it is easily noticed under laboratory conditions when infected and healthy larvae are reared side by side over a period of a few days. A second feature easily noted in the laboratory is that ascovirus diseases are chronic, though usually fatal. When infected during early stages of development, ascovirus-diseased larvae often survive for 2 or 3 weeks beyond the time at which most healthy larvae have completed their development and pupated. Signs of disease other than these are minor, but include the inability to completely cast the molted cuticle, a bloated thoracic region, and a white or creamy discoloration and hypertrophied appearance of the larval body at advanced stages of disease development.

Cytopathology and Cell Biology

In comparison to all other known viruses, the most unique property of ascoviruses is the unusual cytopathology that leads to the formation of the virion-containing vesicles. This process resembles apoptosis, and previous studies of the SfAV-1a genome have shown that it encodes a functional executioner caspase, synthesized 9 h after infection, which by itself is capable of inducing apoptosis (Fig. 3); at present, no other virus is known to encode a functional caspase. Although *in vitro* studies have demonstrated that knock-down of SfAV-1a caspase by RNA interference significantly reduced the formation of apoptotic vesicles, it is unclear whether caspase homologs in other ascoviruses participate in this process. For example, in a similar study a caspase-like gene in HvAV-3e was shown not to be implicated in apoptosis but was essential for virus replication.

At the cellular level *in vivo*, the disease begins with extraordinary hypertrophy of the nucleus accompanied by invagination of sections of the nuclear envelope, followed by a corresponding enlargement of the cell. Cells typically grow from 5 to 10 times the diameter of uninfected cells. As the nucleus enlarges, the nuclear envelope ruptures and disintegrates into fragments. At about this stage, the cell plasmalemma begins to invaginate along “planes” toward the now anucleate cell center. Concomitantly, sheets of membrane form closely adjacent to mitochondria that accumulate along the planes. As this process continues, the membrane sheets coalesce and join the invaginating plasmalemma, thereby cleaving the cell into a cluster of 20 to more than 30 vesicles, ranging in size from 5 to 10 μm in diameter. This aspect of ascovirus cellular pathology resembles the formation of apoptotic bodies during apoptosis. However, rather than dissipate as the cell dies, the developing apoptotic bodies are rescued by the virus and progress to form vesicles in which virions continue to assemble. These virion-containing vesicles, also referred to as viral vesicles, typically remain in the tissue until the basement membrane ruptures, though on occasion cell hypertrophy can be so great that the enlarging cell erupts out through the basement membrane of the infected tissue, releasing large fragments of the infected cell directly into the hemolymph. Analysis of ascovirus genomes shows that, unlike many other large DNA viruses, ascoviruses encode several lipid-metabolizing enzymes that are likely involved in the process of converting developing apoptotic bodies into virion-containing vesicles. Recent transcriptome analysis have shown that genes coding for a patatin-like phospholipase, a PlsC phosphate acyltransferase, a fatty acid elongase, and an esterase/lipase in SfAV-1a are highly expressed during later stages of virus infection *in vivo*. These enzymes are known to modify fatty acid chain length and the degree of lipid saturation, presumably generating novel lipid components in the virion and/or the membranes of viral vesicles. Regarding the latter, it is possible that these lipids, in part, function in providing stability to viral vesicles as they circulate in hemolymph for an extended period encompassing several days or weeks, and when these vesicles are disseminated through oviposition by female parasitic wasps. Another interesting aspect of this cytopathology is that the mitochondria, apparently under ascovirus control, are reprogrammed to assist synthesis of membranes that cleave the cell into the viral vesicles in which most replication occurs.

Although the process by which viral vesicles are cleaved from cells varies among different ascoviruses, the histopathology is similar among virtually all ascoviruses. Vesicles accumulate in the tissues where they are formed, but as these tissues degenerate during disease progression, the basement membrane of infected tissues deteriorates and ruptures, allowing the vesicles to spill out into the hemolymph. There they accumulate reaching concentrations as high as 10^7 – 10^8 vesicles per ml within 3–4 days of infection. There is some evidence that viral replication proceeds within the vesicles as they circulate in the hemolymph, and thus this tissue must also be considered one of the tissues attacked by ascoviruses. In fact, because such high concentrations of viral vesicles are found in the hemolymph, this tissue could be considered a major site of infection, particularly if it is eventually shown that these viruses continue to replicate in the vesicles as they circulate in the hemolymph.

Despite the chronic nature of the disease caused by ascoviruses, virion-containing vesicles are present in the hemolymph within 2 or 3 days of infection. When the virus replicates in cells *in vitro*, the vesicles are formed within 12–16 h of infection. The rapid development and circulation of the viral vesicles in the hemolymph probably evolved to enhance transmission of the virus by parasitic wasps.

Tissue Tropism

The cytopathology of ascoviruses is consistent among different viral species; however, considerable variation occurs with respect to the tissues attacked, that is, in which replication occurs. TnAV, HvAV, and SeAV exhibit a relatively broad tissue tropism infecting the tracheal matrix, epidermis, fat body, and connective tissue. Differences exist between these species in that some HvAV variants infect the epidermis much more extensively than TnAV variants, whereas some of the latter can also replicate more extensively in fat body cells, but appear only to do this when larvae are infected early in their development. Alternatively, the type species, SfAV, and its variants have a very narrow tissue tropism, with the fat body being the primary site of infection. DpTV occurs in the nuclei of all tissues of its wasp host, but appears to only produce progeny in ovarian tissues. In its lepidopteran pupal host, it attacks and replicates in a wide variety of tissues.

Replication and Virion Assembly

Although there have been few biochemical studies of viral DNA replication or protein synthesis, studies carried out with ascoviruses *in vivo* and *in vitro* show that progeny virions first appear about 12 h after infection. Virion assembly is initiated after the nucleus ruptures, and occurs prior to and during the cleavage of the cell into viral vesicles. The first recognizable structural

component of the virion to form is the multilaminar layer of the inner particle. Based on its ultrastructure, this layer consists of a unit membrane and an exterior layer of protein subunits. As the multilaminar layer assembles, a dense nucleoprotein core aggregates on the interior surface. This process continues until the inner particle is complete. After formation, the inner particle is enveloped by membranes within the cell or vesicle. These membranes are apparently synthesized *de novo*. Thus, the assembly of the virions is reminiscent of that in other viruses with complex virions, such as the iridoviruses, herpesviruses, and poxviruses, where the virions differentiate after association of the precursors of virion structural components.

After formation, the virions of the TnAV ascovirus accumulate toward the periphery of the vesicle where they often form inclusion bodies, that is, aggregations of virions (Fig. 2). In SfAV, occlusion bodies are formed in which the virions are actually occluded in a “foamy” vesicular matrix that consists of a mixture of protein and minute spherical vesicles. When viewed with phase microscopy, these viral inclusion and occlusion bodies are phase bright, and are largely responsible for the highly refractile appearance of the vesicles. Ascoviruses do not typically form the types of occlusion bodies characteristic of other types of DNA insect viruses, such as baculoviruses and entomopoxviruses.

Origin and Evolution

The subject of viral evolution over millions of years has received relatively little study due to the lack of a fossil record. Moreover, viruses are considered polyphyletic, and thus most of the more than 70 families of viruses are thought to have originated independently. In this regard, ascoviruses may provide a unique opportunity to obtain insights into virus evolution over long periods. Phylogenetic comparisons of ascovirus genes sequenced to date including those coding DNA polymerase and major capsid protein as well as 20 other proteins indicate that these viruses evolved from a lepidopteran iridovirus (family Iridoviridae). Iridoviruses, in turn, appear to share a common ancestor with two families of *Megavirales*, *Marseilleviridae* and *Pithoviridae* which attack certain aquatic unicellular organisms.

Future Perspectives

At present, too little is known about ascoviruses to assess whether they are or will turn out to be of economic importance. Their poor infectivity *per os* makes it highly unlikely they will ever be developed as viral insecticides, especially given the successful advent of insect-resistant transgenic crops. However, as more entomologists become familiar with the disease caused by ascoviruses, it may be shown that in habitats rarely treated with chemical insecticides, such as transgenic crops, these viruses are responsible for significant levels of natural pest suppression, particularly where parasitic wasps are abundant. Such findings would encourage even greater emphasis on the development of biological control and other more environmentally sound methods of pest control. With respect to the cell biology of viral vesicle formation, ascoviruses provide an interesting model for how apoptosis can be manipulated at the molecular level. Additionally, study of the unusual process by which ascoviruses rescue the developing apoptotic bodies to form viral vesicles could lead to insights into how cells manipulate the cytoskeleton and mitochondria. Finally, it is possible that viral vesicles will provide a unique anucleate cellular system for studying the replication of a complex type of enveloped DNA virus *in vitro*.

Further Reading

- Asgari, S., 2006. Replication of *Heliothis virescens* ascovirus in insect cell lines. *Archives of Virology* 151, 1689–1699.
- Asgari, S., 2007. A caspase-like gene from *Heliothis virescens* ascovirus (HvAV-3e) is not involved in apoptosis but is essential for virus replication. *Virus Research* 128, 99–105.
- Asgari, S., Bideshi, D.K., Bigot, Y., Federici, B.A., Cheng, X.-X., 2017. ICTV virus taxonomy profile: *Ascoviridae*. *Journal of General Virology* 98, 4–5.
- Asgari, S., Davis, J., Wood, D., Wilson, P., McGrath, A., 2007. Sequence and organization of the *Heliothis virescens* ascovirus genome. *Journal of General Virology* 88, 1120–1132.
- Bideshi, D.K., Bigot, Y., Federici, B.A., Spears, T., 2010. Ascoviruses. In: Asgari, S., Johnson, K.N. (Eds.), *Insect Virology*. Great Britain: Caister Academic Press, pp. 2–34. (ISBN 978-1-904455-71-4).
- Bideshi, D.K., Demattei, M.V., Rouleux-Bonnin, F., *et al.*, 2006. Genomic sequence of *Spodoptera frugiperda* ascovirus 1a, an enveloped, double-stranded DNA insect virus that manipulates apoptosis for viral reproduction. *Journal of Virology* 80, 11791–11805.
- Bideshi, D.K., Spears, T., Zaghloul, H.A.H., *et al.*, 2018. Ascovirus P64 homologs: A novel family of large cationic proteins that condense viral genomeic DNA for encapsidation. *Biology* 7, 44.
- Bideshi, D.K., Tan, Y., Bigot, Y., Federici, B.A., 2005. A viral caspase contributes to modified apoptosis for virus transmission. *Genes and Development* 19, 1416–1421.
- Bigot, Y., Rabouille, A., Doury, G., *et al.*, 1997. Biological and molecular features of the relationships between *Diadromus pulchellus* ascovirus, a parasitoid hymenoptera wasp (*Diadromus pulchellus*) and its lepidopteran host, *Acrolepiopsis assectella*. *Journal of General Virology* 78, 1149–1163.
- Chen, Z.S., Cheng, X.-W., Wang, X., *et al.*, 2018. Proteomic analysis of the *Heliothis virescens* ascovirus 3i (HvAV-3i) virion. *Journal of General Virology* 100, 301–307.
- Cheng, X.W., Wang, L., Carner, G.R., Arif, B.M., 2005. Characterization of three ascovirus isolates from cotton insects. *Journal of Invertebrate Pathology* 89, 193–202.
- Federici, B.A., 1983. Enveloped double-stranded DNA insect virus with novel structure and cytopathology. *Proceedings of the National Academy of Sciences of the United States of America* 80, 7664–7668.
- Federici, B.A., Govindarajan, R., 1990. Comparative histopathology of three ascovirus isolates in larval noctuids. *Journal of Invertebrate Pathology* 56, 300–311.
- Federici, B.A., Vlak, J.M., Hamm, J.J., 1990. Comparison of virion structure, protein composition, and genomic DNA of three ascovirus isolates. *Journal of General Virology* 71, 1661–1668.

- Govindarajan, R., Federici, B.A., 1990. Ascovirus infectivity and the effects of infection on the growth and development of noctuid larvae. *Journal of Invertebrate Pathology* 56, 291–299.
- Pellock, B.J., Lu, A., Meagher, R.B., Weise, M.J., Miller, L.K., 1996. Sequence, function, and phylogenetic analysis of an ascovirus DNA polymerase gene. *Virology* 216, 146–157.
- Piégu, B., Asgari, S., Bideshi, D., Federici, B.A., Bigot, Y., 2015. Evolutionary relationships of iridoviruses and divergence of ascoviruses from invertebrate iridoviruses in the superfamily megavirales. *Molecular Phylogenetics and Evolution* 84, 44–52.
- Wang, L., Xue, J., Seaborn, C.P., Arif, B.M., Cheng, X.W., 2006. Sequence and organization of the *Trichoplusia ni* ascovirus 2c (*Ascoviridae*) genome. *Virology* 354, 167–177.
- Zaghloul, H.A.H., Hice, R., Arensburger, P., Federici, B.A., 2018. Transcriptome analysis of the *Spodoptera frugiperda* ascovirus *in vivo* provides insights into how its apoptosis inhibitors and caspase promote increased synthesis of viral vesicles and virion progeny. *Journal of Virology* 91 (23), e00874.

Baculovirus–Host Interactions: Repurposing Host-Acquired Genes (*Baculoviridae*)

A Lorena Passarelli, Kansas State University, Manhattan, KS, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Apoptosis A type of programmed cell death important for the normal development of organisms and used as a defense mechanism by cells to prevent the replication of a pathogen and their further propagation. The process is characterized by DNA fragmentation and cell blebbing. Cysteine-aspartases are caspases which cleave proteins after aspartic acids and are the active enzymes that drive the process.

Co-option Existing genes acquiring a different function and/or regulation, resulting from natural selection.

Horizontal gene transfer Lateral transfer of genes between genomes, having a role in evolution and adaptation and differing from vertical gene transfer or the transfer of genes from a parent to an offspring.

Host range Group of host species that are permissive and susceptible to infection by a virus.

Introduction

Viruses are obligate intracellular parasites that depend on cellular metabolism and factors to support their multiplication and propagation between hosts. Viral replication cycles, involving the disassembly and reassembly of viral components to generate new infectious viral particles, provide ample opportunities for viral and host proteins to interact, genome segment exchange and rearrangement, and competition for mutually needed abundant or limited resources used by viruses and hosts. These interactions lead to an arms race between viruses trying to multiply and host cells mounting immune defenses to halt virus multiplication and ameliorate any cytopathic effects. The process leads to co-evolution of viruses and their hosts. Viruses need to carefully choreograph their host defense weaponry, as killing a host too quickly can result in a dead-end host, curtailing desired levels of virus multiplication and spread. There are many known examples of viruses having acquired host genes that benefit their replication and, in some cases, these genes have been co-opted to defeat the host. The horizontal transfer of genes is bidirectional—viruses acquiring genes from a host or another co-infecting pathogen and hosts acquiring viral genes.

This review will provide background on gene acquisition by viruses from a host or other organism within that host. It will also summarize gene acquisition from a virus to a host. The general role and examples of host-captured genes by baculoviruses will be discussed in more detail, focusing on two broad groups, viral host range genes and genes that affect host defense pathways.

Viruses Acquiring Genes From Their Hosts

Viruses may acquire genes from other organisms and these may be maintained in viral genomes if they are beneficial for their replication. Acquired genes may serve to manipulate the immune system of the host, by evading, avoiding or hindering immune factors that curtail their replication. In addition, virus-acquired host genes can enhance virus replication in their native host or expand their host range, further ensuring their successful spread. Captured genes typically evolve faster than the cellular counterpart, often making the lineage of acquired genes hard to determine and study. In some cases, specific genes have been acquired independently multiple times, for example, IL-10 in herpesviruses and poxviruses. Viruses may acquire genes from their host, a co-infecting pathogen (*e.g.*, virus or bacteria) or the microbiome of the host ([Fig. 1](#)). The transfer of DNA may be complex involving three or more pathogen-host interactions. For example, bacterial viruses that along with their obligate intracellular bacterial host co-infect a eukaryotic organism may facilitate DNA exchange between these organisms. Bacteriophage WO from *Wolbachia*, Gram-negative bacteria that infect several arthropods, carries a eukaryotic association module, a segment of DNA in the bacterial virus with potentially functional eukaryotic gene domains (*e.g.*, spider toxic, protease cleavage sites), indicating the lateral transfer of genes and the potential role of these domains in factor mimicry.

Hosts Acquiring Genes From Infecting Viruses

Although it is more widely discussed how viruses capture genes from their hosts to supplement their replication advantages, the reciprocal also occurs, namely, host organisms acquiring genes from the viruses that infect them. As with virus co-opted genes, hosts can use viral genes for specific purposes and the function of the gene product may be modified. Viruses can integrate into host genomes and may remain as part of the host genome indefinitely. Acquired genes may be co-opted and function as antiviral defense genes, lending a twist to the arms race between viruses and their hosts, where the virus or its host acquired a gene from its

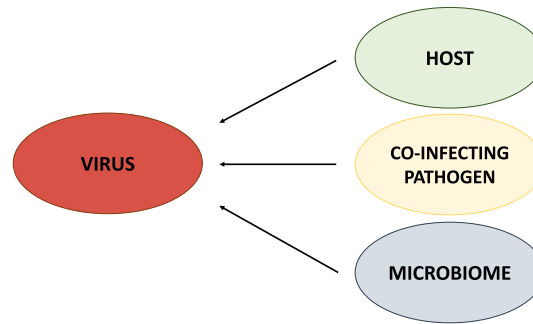


Fig. 1 *Gene flow to viruses.* Movement of genes to viruses from the host organism, co-infecting pathogens or organisms in the microbiome of the host. Reciprocal movements may also take place.

foe to defeat it. Analyses have indicated that viral RNA-dependent RNA polymerase and capsid genes, two viral-specific genes, have been found nested in cellular genomes and, in some cases, expressed, but their specific function in the host genome remains to be determined. Horizontal gene transfer of genes may also occur in non-viral invaders; for example, the pillbug *Armadillidium vulgare* was recently shown to have acquired a gene from the bacteria *Wolbachia*. Gene transfer can also occur by the integration of virus genes into a host genome, where viral genes usually evolve at the pace of host genes. Some viruses use integration of their genome into host genomes as part of their replication strategy, while others may integrate depending on the environment or to hide from host defense pressures. There is ample evidence of viral genomes “fossilized” in host genomes, underscoring the long relationship and co-evolution between viruses and hosts.

A startling example of virus evolution is exemplified by polydnaviruses. Polydnaviruses have symbiotic relationships with two large groups of parasitic wasps, ichneumonid and braconid wasps, that are evolutionarily different. The wasps parasitize lepidoptera larvae as they oviposition. During oviposition, the eggs are injected along with polydnavirus particles. The encapsidated genomes of polydnaviruses encode genes that interfere with lepidopeteran larvae immunity and physiology, allowing egg development. In turn, polydnaviruses require the wasp host to spread the virus, since the viral genes that encode structural and replication proteins are encoded in the wasp genome and are not in the encapsidated polydnavirus genome. The ovaries of braconid wasps were reported to express 29 genes with homologs in nudiviruses and derived from nudiviruses (viruses evolutionarily closely related to baculoviruses). A significant number of these 29 genes (62% or 18 genes) are also baculovirus core gene homologs, as well as being in nudiviruses. Baculovirus core genes are carried by all baculoviruses sequenced to date and are essential for baculovirus replication. Thus, it appears that polydnaviruses evolved from the integration of nudiviruses into the genomes of wasps, aiding wasp reproduction success and expanding the gene repertoire of wasps to more effectively parasitize hosts.

Baculovirus Replication

Baculoviruses belong to the family *Baculoviridae* and the prototype baculovirus is *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV). Baculoviruses naturally infect insects, yet baculovirus virions can enter—but cannot replicate in—many other types of animal cells. The majority of baculoviruses discovered to date infect the larval stage of insects in the order Lepidoptera, which comprises moths and butterflies. A few other baculoviruses, considered more ancient, infect insects in the orders Diptera (mosquitoes) and Hymenoptera (sawflies). Baculoviruses carry a double-stranded circular DNA genome, ranging in length from 80- to 180-kilobasepairs that is contained in an enveloped rod-shaped nucleocapsid. Baculovirus genes are transcribed from both DNA strands with short intergenic regions, and their genomes can be engineered to incorporate large segments of foreign DNA.

During prototypical baculovirus replication, two types of infectious virions are produced, budded and occlusion-derived virions. Both virion types are produced in the nucleus of infected cells and have identical genomes and capsids, but different envelopes, envelope proteins, and virion morphology. The composition of their envelopes and an environmentally-stable protein matrix surrounding the occlusion-derived virions are essential to carry out distinct functions in the replication cycle of the virus. The budded virus exits the nucleus and buds through the cell plasma membrane at late times post infection, obtaining the viral-encoded fusion protein necessary for entry into new cells. Thus, the budded virus is responsible for spreading infection between cells and is the form of the virus used in cell culture. The occlusion-derived virus is produced at very late times post infection. In some baculoviruses, multiple nucleocapsids (MNPVs) are co-enveloped and then occluded in a protein matrix composed of the protein polyhedrin. In single NPVs (SNPVs), single nucleocapsids are enveloped and embedded in polyhedra. Polyhedra are protein crystals containing virions that form an occlusion body. Polyhedrin protects virions in the environment once virions are released from an infected host, allowing transmission of the virus to other hosts. Additional information on the replication of baculoviruses is described in other articles in this volume.

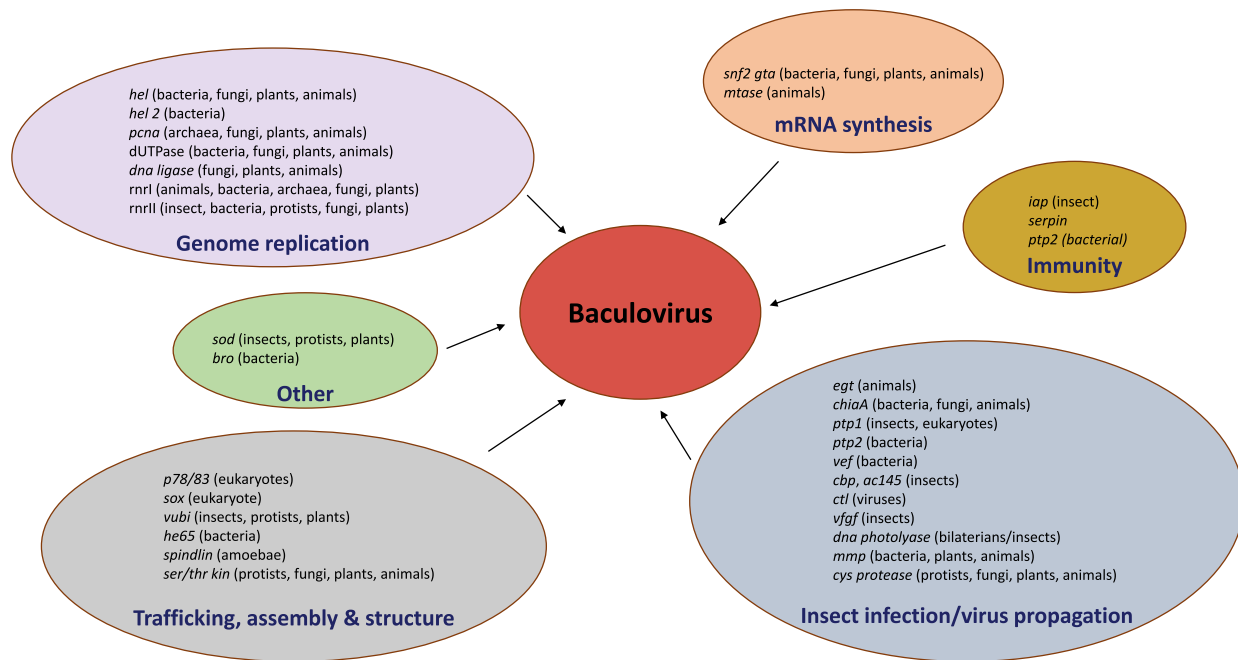


Fig. 2 *Baculovirus* genes with homology to genes in other organisms. Baculoviruses have acquired genes that fulfill several purposes. Proposed origin of genes or homologies to cellular genes is indicated in parentheses. Additional genes from those presented here have cellular homologs. *Bro*, baculovirus repeated open reading frame; *chiaA*, chitinase A; *ctl*, conotoxin-like; *cys protease*, cysteine protease; *dUTPase*, deoxyuridine 5'-triphosphate nucleotidohydrolase; *egt*, ecdysteroid UDP-glucosyltransferase; *he65*, hr4 left/*EcoRI* 65-kilodaltons; *hel*, helicase; *iap*, inhibitor of apoptosis; *mmp*, matrix metalloproteinase; *mtase*, methyltransferase; *pcna*, proliferating cell nuclear antigen; *ptp*, protein phosphatase; *rrn*, ribonucleotide reductase; *ser/thr kin*, serine/threonine kinase; *snf gta*, sucrose non-fermenting global transactivator; *sod*, superoxide dismutase; *sox*, sulfhydryl oxidase; *vef*, viral enhancing factor; *vfgf*, viral fibroblast growth factor; *vubi*, viral ubiquitin. Information was obtained mainly from Hughes, A.L., Friedman, R., 2003. Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. *Molecular Biology and Evolution* 20, 979–987. Thèzè, J., Takatsuka, J., Nakai, M., Arif, B., Herniou, E.A., 2015. Gene acquisition convergence between entomopoxviruses and baculoviruses. *Viruses* 7, 1960–1974.

Baculovirus Acquisition of Genes

At least 25% of baculovirus proteins have been shown to directly interact with host factors. The lineages of many of these interacting viral factors can be traced to cellular counterparts, indicating horizontal gene transfer taking place in different baculoviruses at different times in evolution and interactions on an evolutionary scale. A systematic study of virus-host interacting proteins using published work revealed that 29 viral proteins had been shown to interact with cellular host factors or pathways.

The study of an overlapping host repertoire between two insect viruses, entomopoxviruses and baculoviruses, suggested analogous selection pressures from their respective host, and this may have led to similar genetic acquisitions and repurposing of genes to ensure virus replication. Phylogenetic analyses using these acquired genes indicated that genes had been transferred to insect viruses following coinfections of insects with viruses and other organisms. The gene origin of several virus-acquired genes was traced to genes obtained from bacteria, and multicellular or unicellular eukaryotes (Fig. 2). In some cases, acquired genes could be traced to the host order, Lepidoptera, and there was clear indication that horizontal gene transfer may have taken place. The authors also cited examples of viral genes that seem to have no cellular corresponding homolog and are only carried by a virus (es). Since some of these genes are conserved in both virus families, it implied that acquired genes conferred a fitness to the virus, retaining these genes.

In addition to gene acquisition from the host and co-option of genes, many baculovirus genes are under positive selection, and this serves to avoid the immune system of the host or other pressures. Genes under positive selection include viral genes involved in DNA replication (e.g., *helicase*), structural genes (e.g., *ODV-e66*), transcription genes (e.g., *lef-4*), apoptosis inhibitor genes (e.g., *p35*) and others, painting a complete picture of adaptation that ensure fitness. Some of these genes have homologies to host genes (e.g., *helicase*) and others do not (e.g., *p35*).

A total of 15 proteins or 11% of proteins from the *Bombyx mori* NPV had recognizable counterparts in the host *Bombyx mori* and a number of them affect the physiology of the host but are not essential for virus replication. These included *protein kinases 1 and 2*, viral *superoxide dismutase*, *DNA polymerase*, *desmoplakin*, *inhibitor of apoptosis 1 and 2*, *UDP-glycosyltransferase*, *viral fibroblast growth factor*, *viral ubiquitin*, *snf2 global transactivator*, *methyltransferase*, *viral chitinase A*, *viral cathepsin*, and *protein tyrosine phosphatase*. In addition, other baculovirus genes with other functions or essential for virus replication also had homology to *Bombyx mori* genes, indicating broad acquisition of genes from lepidopteran hosts.

A number of viral genes appear to be devoted to relieving translational blocks imposed by the host to prevent the translation of viral products, many in the eIF2 α – PKR pathway. Some baculoviruses encode protein kinase 2 (PK2), which inhibits eIF2 α by binding to the N-lobe of PKR. This forms a pseudokinase that targets the insect heme-regulated inhibitor-like eIF2 α kinase thereby enhancing virus production. It was concluded that the viral PK2 may have been acquired from a lepidopteran host to compete for shared translation-specific factors.

One pathway used by several viruses for replication or spread within or away from the infected cell is actin polymerization. Baculoviruses encode P78/83, a capsid-associated protein with Wiskott-Aldrich syndrome protein-like domains similar to actin nucleating promoting factors. P78/83 is necessary for virus replication. It is a functional factor that activates the host Arp2/3 (proteins that regulate the actin cytoskeleton, serving as nucleation sites that seed the origin of new actin filaments) after it is trafficked to the nucleus of infected cells, promoting actin polymerization. Baculoviruses can mobilize by assembling G-actin monomers from the host into F-actin filaments. Baculoviruses use actin to move to the nucleus of cells and to egress from the nucleus of cells and transit through the cytoplasm before budding. Furthermore, nuclear F-actin filaments can embed into the nuclear envelope, disrupt it and facilitate virus egress. Although nuclear actin in eukaryotes and its roles in chromatin remodeling, transcription and processing of RNAs have been more widely studied in recent years, the role of nuclear actin in virus replication is still in the discovery phase. Thus, baculoviruses have co-opted a common cytoplasmic function to promote actin assembly in the nucleus of cells and as an instrument that disrupts nuclear envelope integrity for viruses to transit into the cytoplasm rather than exiting through the nuclear envelope pores.

Baculovirus genes acquired from the host may supply functions that optimize virus replication or facilitate host expansion; however, in some cases, the acquired gene has been co-opted to carry out a new essential function. The envelope fusion protein F of some baculoviruses may have been derived from an insect gene, as indicated by the similarity with the F predicted proteins of three *Drosophila* species and the mosquito *Anopheles gambiae*. Alternatively, baculoviruses may have acquired the *f* gene from *f*-like genes carried in endogenous retroviruses embedded in insect hosts.

Baculovirus Host Range Genes

The host range of baculoviruses has been explored in insects and in cells. Studying the replication of baculoviruses in diverse insects defined alternative hosts in the laboratory that were permissive for virus replication in addition to the host from which the virus was isolated. This defined a relatively broader host range for some members of the *Baculoviridae*. Surveying infected individuals in nature is expanding our knowledge of the host range of each virus, while genomic sequencing of baculoviruses is further defining viral species that can be found in more than one host. In cells, host determining factors that expanded the host range of a virus have been described, and it appears that baculoviruses have acquired genes that allow more efficient replication or transmission in some hosts. Baculoviruses can be considered generalists, infecting several insects, or specialists, infecting a small number of insects, depending on the species. In general, they have a relatively narrow host range, a characteristic that is not optimal to agriculturists needing to combat more than one insect pest during a growing season but advantageous in that they do not target beneficial insects (*e.g.*, pollinators). AcMNPV was originally isolated from the host species *Autographa californica* and can infect over 35 insect species, while genetically similar viruses isolated from other species (*e.g.*, BmNPV isolated from *Bombyx mori*) infect slightly over a dozen insect species in a laboratory study but has only been successfully isolated from *Bombyx mori* in the wild. This poses the question of whether additional or specific genes allow some viruses to expand their host range.

Having the appropriate host receptor binding protein for viral attachment may allow virus entry into cells, but successful and complete virus replication may not be supported unless viral genes that properly interact with host genes and defeat, reduce or circumvent immune responses are effective. Several baculovirus genes have been characterized that are necessary for a productive infection in another insect host or cell line derived from a different host, suggesting acquisition of genes occurred to expand the host range of a virus. These genes include six factors with seemingly little in common, either structurally or functionally, which suggests that they are affecting different mechanisms to alter host range, and viruses may have acquired these genes at different times. Briefly, the six baculovirus genes defined as host range genes are described as follows. (1) The viral *helicase p143* (*ac95*) is a conserved gene in all baculovirus genomes. Its product, P143, is a helicase required for viral DNA replication and late gene expression. Changing two amino acids in P143 allowed AcMNPV to be lethal in non-permissive *Bombyx mori* larvae. (2) The AcMNPV *p35* gene (*ac135*), an apoptosis inhibitor gene, carried by some baculoviruses, is necessary to prevent apoptosis in cell lines derived from *Spodoptera frugiperda* but not necessary for successful replication in cell lines derived from *Trichoplusia ni*. Mutant viruses lacking *p35* are significantly less lethal than wild-type in *S. frugiperda* larvae but not in *T. ni* larvae. Baculoviruses have other apoptosis inhibitors (P49, IAPs) and in some cases more than one gene copy is present. This indicates that virus replication is dependent on blocking the process of apoptosis in some species, but not others, and several factors may be encoded to ensure the proper host-virus interactions. (3) The *late expression factor-7* gene (*lef-7*, *ac125*) enhances viral late gene expression and DNA replication. Its deletion affects virus replication in some hosts and cell lines (SF-21) but not in others (TN-368). The BmNPV *lef-7* showed little or no replication defects in BmN cells or *B. mori* larvae. (4) The *host cell factor-1* (*hcf-1*, *ac70*) is found only in some baculoviruses. It stimulates a late viral promoter in some cell lines (*e.g.*, TN-368), where it is necessary for optimal virus DNA synthesis and virion production. Deletion of *hcf-1* affects mortality of *T. ni* larvae. It is not required in SF-21 cells and a virus with a deletion of *hcf-1* can replicate in SF-21 cells and *S. frugiperda* larvae, but protein synthesis shut-off was impaired. Expression of

HCF-1 allowed *Hyphantria cunea* MNPV to replicate in non-permissive TN-368 cells. (5) The *host range factor-1* gene, *hrf-1*, is a gene encoded by *Lymantria dispar* MNPV. When expressed in AcMNPV, it allows AcMNPV to replicate in semi-permissive cells derived from *L. dispar* and in *L. dispar* larvae by allowing translation to proceed. A homolog of *hrf-1* is also carried by *Orgyia pseudotsugata* MNPV, a virus that replicates in *L. dispar*, suggesting acquisition of the gene was necessary for these viruses to productively infect *L. dispar*. However, the OpMNPV *hrf-1* did not allow AcMNPV to replicate in *L. dispar*-derived cells, suggesting differences in functionality or in the interactions with other host or viral factors. (6) The AcMNPV *ie-2* (*ac151*) gene is a transcriptional transactivator in the presence of the powerful transactivator IE-1. IE-2 also enhances viral DNA replication and arrests the cell cycle. It is not necessary in *T. ni*-derived cells compared to *S. frugiperda*-derived cells, but AcMNPV mutants lacking *ie-2* still have diminished replication in *T. ni*. It is possible that specific genes encoded by some baculoviruses optimize virus replication in some hosts and depending on host innate immunity, the presence of such genes may determine if virus replication is successful or not. The inability to test alternative hosts in some cases precludes obtaining a complete picture of specific gene advantages in those hosts. Altogether, various auxiliary genes functioning at the levels of transcription, translation and DNA biosynthesis are often non-essential for the ability of a virus to infect some hosts but essential in others. Their conservation in some virus species may be indicative of efficacious infectivity in specific hosts. Acquisition of host range genes also suggests that viral genes interact with specific host genes to augment viral replication.

Virus Defense and Resistance to Baculoviruses

Several genes that block apoptosis are encoded by baculoviruses to prevent host cell death and thereby allow virus replication. These genes work at different steps in the apoptotic pathway, suggesting different acquisitions during their evolution and the importance of carrying genes that effectively block apoptosis. The gene *p35* was discovered in a spontaneously generated mutant of AcMNPV that had a mutation within *p35*. A virus with this mutation prevented specific cells (*i.e.*, Sf-21) from undergoing apoptosis during infection. The *p35* gene product, P35, inhibits effector or executioner caspases, cysteine-aspartic proteases, by binding to them in a stoichiometric manner and preventing proteolytic cleavage of their aspartic acid-containing substrate proteins. A similar gene product, P49, is encoded by some baculoviruses but, works earlier in the apoptotic cascade than P35, inhibiting initiator caspases. Thus, P49 can rescue *p35* mutant viruses by working upstream of P35. Inhibitors of apoptosis genes, *iaps*, have homologs widespread in metazoans. The viral IAP from *Orgyia pseudotsugata* multiple NPV, Op-IAP3, binds the insect IAP, stabilizing it and preventing its transitory ability to block cell death. This ensures effective cell death and virus replication.

The baculovirus *Hemileuca* sp. NPV encodes a functional serpin. Serpins are serine proteinase inhibitors and mediate the regulation of proteinases involved in the phenoloxidase process, a process that melanizes tissues or pathogens in invertebrates and has an important role in immune response and tissue healing. The baculovirus serpin, Hesp018, inhibited trypsin, chymotrypsin and plasmin. In addition, it blocked bacteria-activated phenoloxidase responses in the hemolymph of the tobacco hornworm *Manduca sexta*. Expression of *hesp018* in AcMNPV resulted in increased budded virus production in Sf9 cells and increased virulence in *T. ni* larvae. Other baculoviruses encoding serpin-like genes have not been discovered, indicating that *Hemileuca* sp. NPV has a singular advantage to encode *hesp018* and replicate in its host. The viral serpin gene differs from noctuid and bombycoid serpin-4 genes. Thus, it was concluded that *hesp018* was either acquired long ago and slowly diverged or it was a recent acquisition and the gene changed at a fast rate compared to the cellular homolog. Other baculoviruses may have other genes that help control insect humoral immunity.

For many years it was thought that an advantage of using baculoviruses as pest control agents of important agricultural and forestry plants was that the pests would have difficulty in developing resistance to the viruses they had co-evolved with for many years. However, in recent years, it has been found that a specific viral gene in the *Cydia pomonella* granulovirus involved in transcription, *pe38*, is necessary for infectivity in codling moths. Resistant insects were identified that were infected by a virus carrying a mutation within *pe38*. Acquired host resistance in codling moths has also been documented and has been mapped to dominant gene mutation(s) in sex-linked chromosomes, but the specific resistance genes have not yet been identified. Although insects have acquired resistance to virus isolates, indicating adaptation, this is an area that deserves more attention and has been understudied. In the past, the limitation of lepidopteran genetic tools and genomic sequences has hindered developments in the field, although this is now beginning to change.

In addition to baculovirus genes that allow a productive infection in additional hosts and that interfere with the host immunity, other baculovirus genes that aid in other processes, including virus assembly, viral DNA replication, host physiology manipulation, and virus dispersal have homologs in other organisms (Fig. 2).

Concluding Remarks

A significant percentage of baculovirus genes have counterparts in cellular organisms, conveying the close interaction between the two and showing how changing environmental situations leads to gene exchanges to benefit one and counteract the other. Studies on baculovirus gene lineages indicate that baculoviruses have usurped many of their genes from their insect hosts (some specifically traced to lepidopterans) or bacteria. Homologs of many baculovirus genes are found in insects, prokaryotes, other

eukaryotes and archaea, heightening the richness of tools and strategies acquired through evolution. Thus, baculoviruses have evolved to contain a toolbox suitable to replicate in diverse insect hosts. Host-acquired genes include genes important for immunity, spread, genome replication, transcription and other processes. Detailed analyses of the phylogenetics of each gene will provide a more comprehensive picture as to when gene acquisition took place, the cellular organism from which genes were acquired, additional host interactions, and insights into gene function. The fluidity of gene transfer suggests the need for baculoviruses to optimize their replication strategies in different hosts through time and following diverse burdens.

Genes from a host or from another organism present in a host can be acquired by a virus to meet selective advantages depending on the environmental cues. These can include fighting host defenses, mounting offenses to debilitate host processes and host range expansion. Viruses can use different strategies to further their replication capabilities: mimicry, where gene products mirror host proteins and serve as decoys; co-option, the adaptation of different functions of an original host protein to aid in virus replication; and gene acquisition, to have an additional function not previously encoded. Altogether, gene acquisition adds tools to a viral toolbox that can be effectively used to either defeat the host or enhance virus replication.

Viral genes acquired from other organisms generally have low similarity to and simpler domains than the cellular gene, making it difficult to define their origin. In addition, viral genes incur additional and more rapid changes than host proteins, making it harder to discern the directionality of horizontal gene transfer events. Co-opted genes can acquire a different function, further muddying identification of gene origins. Genes transferred between viruses and their hosts have co-evolved in an impending and continuous arms race for survival of each. Virus-host networks can be studied using molecular methods that provide a footprint of gene acquisition. More specifically, the transferal of genes and relationships (genetic or behavioral) between viruses and their hosts can be studied using diverse systems, including high-throughput methods and genome-wide screens that map host genes that affect various viral replication steps; next generation sequencing of host and viral genomes combined with detailed and structure-based phylogenetics to define evolutionary lineages; molecular interactomes using the yeast two-hybrid system, mass spectrometry complex identification or computational algorithms to identify the interactions between factors; identification of microbiome components to further define the interactions and gene transfer between organisms and viruses; and host experiments supporting *in vitro* ones, when possible. The amalgamation of various methods will help define the cross-talk and mobility of genes between viruses, their hosts, host microbiomes and additional parasitic organisms replicating in a host. Co-option of genes aids in our understanding of the different potential functions of gene products in different settings throughout evolution.

Classification (Compact)

Family: *Baculoviridae*; Genera: *Alphabaculovirus*, *Betabaculovirus*, *Deltabaculovirus*, *Gammabaculovirus*; Species: *Autographa californica multiple nucleopolyhedrovirus* (for example).

Virion Structure

Virion structure is complex. Capsids have helical symmetry (200–450 nm in length, 30–100 nm in diameter). Two enveloped forms of virus are produced: budded and occluded viruses.

Genome

One double-stranded circular DNA genomes, 80- to 180-kilobasepairs in length.

Replication Cycle

Insects ingest occlusion bodies containing occlusion-derived viruses and released virions infect midgut epithelial cells. Following budding from midgut cells, the budded virions infect insect cells in the hemocoel, including hemocytes and cells in various tissues. Infected cells produce more budded virus for systemic spread within an insect and occluded virus for insect-to-insect spread. Upon host death, occluded virions are dispersed for new hosts to ingest. In a few baculoviruses, particularly *deltabaculoviruses*, the infection is midgut restricted.

Epidemiology

Insects acquire infection through consumption of the occluded form of the virus as a contaminant of their food. In some cases, they can be transmitted transovum and transovarially. Baculoviruses have a relatively narrow host range and can cause epizootics, altering insect population dynamics.

Pathogenesis

The tissue and cuticle of infected insects liquefies through the concerted activity of baculoviral chitinase and cathepsin enzymes, dispersing occluded virus. Cuticle discoloration may be apparent. At the cellular level, cells round up and the nucleus is enlarged. At very late times post infection, cells exhibit occlusion or granular bodies, containing sets of one enveloped or more co-enveloped nucleocapsids.

Diagnosis

Infected insect larvae have a lighter color and their cuticle is fragile, easily breaking to the touch and dispersing virus. Infected larvae climb to the top of plant foliage, a virus-induced behavior called *Wipfelkrankheit* or tree-top disease, and this allows more efficient virus spread on the foliage below, upon larva liquefaction or melting.

Further Reading

- Clem, R.J., 2015. Viral IAPs, then and now. *Seminars in Cell & Developmental Biology* 39, 72–79.
- Hill, T., Unckless, R.L., 2017. Baculovirus molecular evolution *via* gene turnover and recurrent positive selection of key genes. *Journal of Virology* 91, e01319.
- Hughes, A.L., Friedman, R., 2003. Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. *Molecular Biology and Evolution* 20, 979–987.
- Katsuma, S., Kawaoka, S., Mita, K., Shimada, T., 2008. Genome-wide survey for baculoviral host homologs using the *Bombyx* genome sequence. *Insect Biochemistry and Molecular Biology* 38, 1080–1086.
- Kong, M., Zuo, H., Zhu, F., *et al.*, 2018. The interaction between baculoviruses and their insect hosts. *Developmental and Comparative Immunology* 83, 114–123.
- Thézé, J., Takatsuka, J., Nakai, M., Arif, B., Herniou, E.A., 2015. Gene acquisition convergence between entomopoxviruses and baculoviruses. *Viruses* 7, 1960–1974.

Baculoviruses: General Features (*Baculoviridae*)

Vera ID Ros, Wageningen University and Research, Wageningen, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AcMNPV	<i>Autographa californica</i> multiple (multiple refers to the multiple nature of nucleocapsids within individual ODV virions of nucleopolyhedroviruses as opposed to those where only a single nucleocapsid is present within each virion from an ODV) nucleopolyhedrovirus	CpGV	<i>Cydia pomonella</i> granulovirus
AgMNPV	<i>Anticarsia gemmatilis</i> multiple nucleopolyhedrovirus	CrleGV	<i>Cryptophlebia leucotreta</i> granulovirus
BmNPV	<i>Bombyx mori</i> nucleopolyhedrovirus	CuniNPV	<i>Culex nigripalpus</i> nucleopolyhedrovirus
CfMNPV	<i>Choristoneura fumiferana</i> multiple nucleopolyhedrovirus	HearNPV	<i>Helicoverpa armigera</i> nucleopolyhedrovirus
		LdMNPV	<i>Lymantria dispar</i> multiple nucleopolyhedrovirus
		NeleNPV	<i>Neodiprion lecontei</i> nucleopolyhedrovirus
		SeMNPV	<i>Spodoptera exigua</i> multiple nucleopolyhedrovirus
		SpliNPV	<i>Spodoptera littoralis</i> nucleopolyhedrovirus
		SpltNPV	<i>Spodoptera litura</i> nucleopolyhedrovirus
		XecnGV	<i>Xestia c-nigrum</i> granulovirus

Glossary

Budded virus (BV) One of the two virion types of baculoviruses that is produced when nucleocapsids bud from the plasma membrane of an infected cell and mediates the systemic cell-to-cell spread within the infected insect.

Granule The OB of granuloviruses (plural granules).

Granulin The major protein in granulovirus OBs.

Granulovirus (GV) Name used for all baculoviruses that form granular-shaped OBs of which the major matrix protein is granulin, comprising all viruses from the genus *Betabaculovirus*.

Nucleopolyhedrovirus (NPV) Name used for all baculoviruses that form polyhedral-shaped OBs of which the major matrix protein is polyhedrin, comprising all

viruses from the genera *Alphabaculovirus*, *Deltabaculovirus* and *Gammabaculovirus*.

Occlusion body (OB) Proteinaceous crystalline structure protecting the infectious ODVs, responsible for horizontal insect-to-insect spread.

Occlusion derived virus (ODV) One of the two virion types of baculoviruses that is assembled in the nucleus where nucleocapsids obtain an inner nuclear membrane-derived envelope, and that becomes occluded within the proteinaceous crystalline matrix of the occlusion body.

Polyhedrin The major protein in nucleopolyhedrovirus OBs.

Polyhedron The OB of nucleopolyhedroviruses (plural polyhedra).

Historical Perspective

The history of the discovery of baculoviruses is closely linked to the development of the silk industry, which originated in China and arrived in Europe during medieval times. The dramatic effect of baculoviruses on insect hosts was already described in 1527 by Marcus Hieronymus Vida, bishop of Alba in Italy, long before the identity of baculoviruses was revealed. His poem “*Bombycum*” (The silkworm) includes a description of liquefying silkworms (caterpillars of the silk moth *Bombyx mori*), later known as jaundice disease and characteristic for baculovirus infections. After the invention of light microscopy, Maestri and Cornalia independently described refractile occlusion bodies which were commonly polyhedron shaped, leading to the naming of the disease observed in the silkworms as ‘polyhedrosis’ by the mid-1800s. Glaser, Paillot, and colleagues, between 1913 and 1928, demonstrated that the polyhedrosis disease of caterpillars was due to a filterable agent and therefore must have a viral etiology. With the perfection of insect tissue culture methodology, William Trager in 1935 reported on the ability of the Grasserie virus, now known as *Bombyx mori* nucleopolyhedrovirus (BmNPV), to replicate in silkworm tissue culture cells. This led to the opportunity of studying the replication of these viruses under more controlled conditions than possible when working with whole insect larvae and free of other microbial agents.

Nomenclature, Taxonomy, and Classification

The family *Baculoviridae* comprises insect viruses with 80–180 kbp circular double-stranded DNA (dsDNA) genomes encoding 100–200 proteins. The family name is derived from the rod-shaped (baculum, Latin for stick) nucleocapsids. In the 10th report of the ICTV (2019, as per see “Relevant Website section”) a total of 82 baculovirus species are recognized and a further 19 tentative species are listed (Table 1). Four genera are distinguished within the *Baculoviridae*, based on phylogeny, genome characteristics,

Table 1 Number of species per baculovirus genus

Genus	Member species (ICTV) ^a	Related, unclassified species
<i>Alphabaculovirus</i>	53	15
<i>Betabaculovirus</i>	26	0
<i>Deltabaculovirus</i>	1	2
<i>Gammabaculovirus</i>	2	2
Total	82	19

^aBased on the 10th report of the International Committee on Taxonomy of Viruses (ICTV) (2019, as per https://talk.ictvonline.org/ictv-reports/ictv_online_report/ds dna-viruses/w/baculoviridae).

host range and occlusion body (OB) morphology; *Alphabaculovirus* (lepidopteran-specific nucleopolyhedroviruses (NPVs)) with 53 species, *Betabaculovirus* (lepidopteran-specific granuloviruses (GVs)) with 16 species, *Deltabaculovirus* (dipteran-specific NPVs) with one species, and *Gammabaculovirus* (hymenopteran-specific NPVs) with two species. Based on phylogenetic inferences but not yet recognized taxonomically, lepidopteran-specific alphabaculoviruses are further divided into Group I and Group II alphabaculoviruses (Fig. 1). These two groups differ significantly in gene content, most notably Group I alphabaculoviruses encode GP64 as the BV fusion protein, whereas Group II alphabaculoviruses encode Fusion protein (F) but lack *gp64*.

Baculoviruses are unique among viruses in having a biphasic replication cycle with two morphologically distinct virion types, budded viruses (BVs) and occlusion derived viruses (ODVs) (Fig. 2). Early in infection, BVs are produced when newly formed nucleocapsids bud out of the insect cell, obtaining an envelope from the cell membrane (Fig. 3). BVs are responsible for cell-to-cell transmission within the infected insects (as well as in cell tissue culture). Later in infection, ODVs are formed in the host nucleus where nucleocapsids obtain a membrane from microvesicles derived from the inner-nuclear membrane. The ODVs are subsequently embedded in a polyhedral (NPV) or granular (GV)-shaped proteinaceous crystalline matrix, forming viral OBs. The occlusion body form is responsible for horizontal insect-to-insect spread. ODVs may have one or several nucleocapsids within a single envelope, depending on the species of baculovirus, forming single (S) or multiple (M) NPVs, respectively.

Baculoviruses infect larval stages of species belonging to the insect orders Lepidoptera, Diptera (family Culicidae) and Hymenoptera (suborder Symphyta). NPVs are found in these three insect orders, while GV is restricted to the Lepidoptera. Baculoviruses are normally named according to the initial host from which they were isolated and the type of OB/ODV associated with it. While this might cause some confusion, since the same virus might occur in a range of insect species, this approach is still applied today. Originally baculovirus abbreviations used the first letter of the genus and species name of the host insect followed by NPV or GV depending on the type of occlusion body. Hence the MNPV from the alfalfa looper *Autographa californica* is abbreviated as AcMNPV and the MNPV from *Choristoneura fumiferana* is abbreviated as CfMNPV. Those baculoviruses which have historically had a two-letter species code continue to maintain those abbreviations. However, as more baculoviruses were being described, a two-letter species designation system was untenable and now a four-letter species identification based on the first two letters of each of the genus and species names of the host insect from which the virus was first isolated is used. Hence the NPV from *Helicoverpa armigera* is abbreviated as HearNPV and that from *Neodiprion lecontei* is abbreviated as NeleNPV. If identical letterings might result, the third or later letters are used instead of the second, as seen for SpltNPV for an NPV from *Spodoptera litura* that was chosen since SpliNPV was already used for an NPV from *Spodoptera littoralis*. Although the S and M are useful morphological descriptors in the names indicating if nucleocapsids exist either singly or in multiples within ODV virions, they do not appear to have any taxonomic relevance and so in more recently described baculoviruses these designations have been dropped but have been maintained for those baculoviruses in which there is a strong historical precedence. To determine whether an isolate belongs to an existing virus species or should be the basis of a new species requires biological and molecular information. A Kimura two-parameter method including the information from all 38 baculovirus core genes can be used for species demarcation, along with additional information such as host range and genome organization.

Morphology

Viruses in all four baculovirus genera have two structurally and biochemically distinct virion phenotypes (BVs and ODVs; Fig. 2) and a biphasic replication cycle (Fig. 3). These virions consist of cylindrical, rod-shaped, nucleocapsids contained within a lipid bilayer envelope. The nucleocapsids contain the dsDNA genome and range in size from 30 to 60 nm in diameter and, depending on the size of the genome, 250–300 nm in length. They consist of a single circular supercoiled dsDNA packed into a protein capsid. The nucleocapsids have a flat base structure on one end and a nipple-like cap on the other end. BVs contain a single nucleocapsid surrounded by a loose-fitting envelope. The BV envelope, obtained from the cellular plasma membrane during budding, is modified at the nipple end by the occurrence of spikes or peplomers consisting of viral glycoproteins such as GP64 or F proteins. ODVs contain one or several nucleocapsids present in a parallel array, enveloped in a membrane derived from the inner nuclear membrane in the nucleus (for alphabaculoviruses) or in the nuclear-cytoplasmic milieu after loss of the nuclear membrane (for betabaculoviruses). ODVs are embedded in paracrystalline occlusions forming either polyhedra (NPVs) or granules (GVs). Polyhedra are roughly cuboidal (although other shapes have been described) containing multiple ODVs. Alphabaculovirus OBs are 0.15–5.0 µm in size and contain many ODVs and,

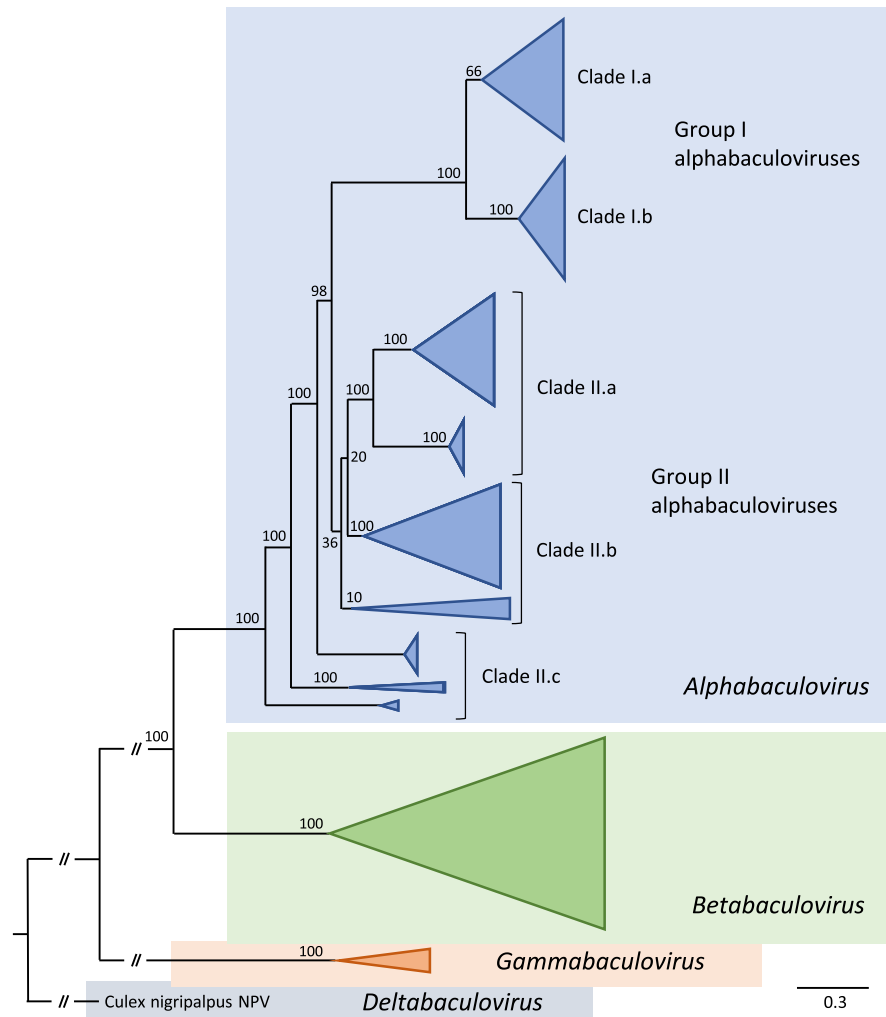


Fig. 1 Schematic representation of baculovirus phylogeny, showing the different genera and other (sub) groups, based on the baculovirus core-genome phylogeny of Thézé, J., Lopez-Vaamonde, C., Cory, J.S., Herniou, E.A., 2018. Biodiversity, evolution and ecological specialization of baculoviruses: A treasure trove for future applied research. *Viruses* 10, 366. The original phylogeny was obtained from maximum likelihood inference analysis of the concatenated amino acid alignment of 37 baculovirus core genes. Values at nodes represent bootstrap values (100 replicates).

depending on the virus species, containing either single (S) or multiple (M) nucleocapsids. Betabaculovirus OBs (granules) are ovoid cylindrical in shape and much smaller in size, measuring 0.12–0.50 μm . Each granule characteristically contains a single virion each containing a single nucleocapsid. Within the genus *Deltabaculovirus* only a single species has been classified, *Culex nigripalpus* nucleopolyhedrovirus. The ODVs of CuniNPV are embedded in an OB composed of a crystalline matrix of a viral protein with no homology to polyhedrin or granulins. The OBs range in size from 0.5 to 5.0 μm and contain few (1–4) or many (> 50) ODVs depending on the strain of the virus. Gammabaculovirus OBs contain multiple ODVs, each containing a single nucleocapsid. Genome sequences of viruses from the three classified species in this genus revealed that they do not encode an envelope fusion protein as found in other baculoviruses, questioning the role of the BV phenotype in the biology of gammabaculoviruses.

Genomes, Gene Content, Organization

Baculovirus genome size varies considerably for the currently sequenced genomes, ranging from 81,755 bp for the gammabaculovirus *Neodiprion lecontei* NPV (NeleNPV), encoding 89 open reading frames (ORFs) to 178,733 bp for *Xestia c-nigrum* GV (XecnGV), encoding 181 ORFs. Baculovirus ORFs are defined as those encoding proteins of at least 50 amino acids with minimal (< 75 nucleotides) overlap with each other. The guanine-cytosine (GC) content of baculovirus genomes ranges from 32.4% for *Cryptophlebia leucotreta* granulovirus (CrleGV) to 57.5% for *Lymantria dispar* multiple nucleopolyhedrovirus (LdMNPV). Baculovirus genes are distributed on both strands, in both orientations. The gene order does not follow a general pattern but is conserved among related baculoviruses. Intergenic regions are short and promoters and/or 3' untranslated regions of flanking genes

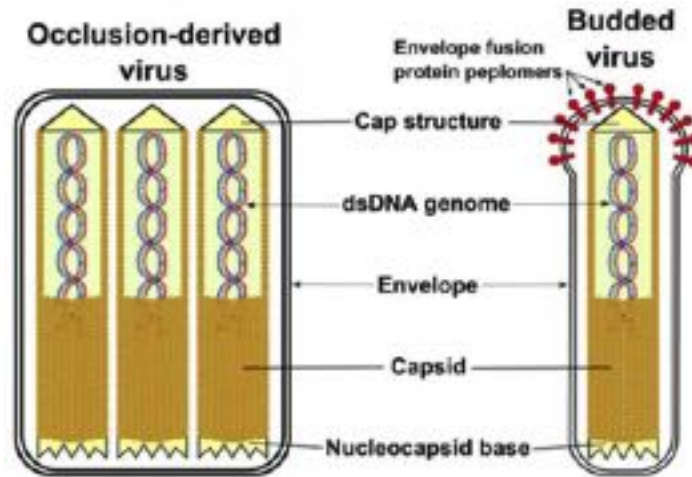


Fig. 2 Baculovirus virions and nucleocapsids. The two baculovirus virion phenotypes, occlusion derived virus (ODV) and budded virus (BV), are illustrated as diagrams with shared and phenotype-specific components. Taken from the International Committee on Taxonomy of Viruses (ICTV) website (<https://talk.ictvonline.org/>).

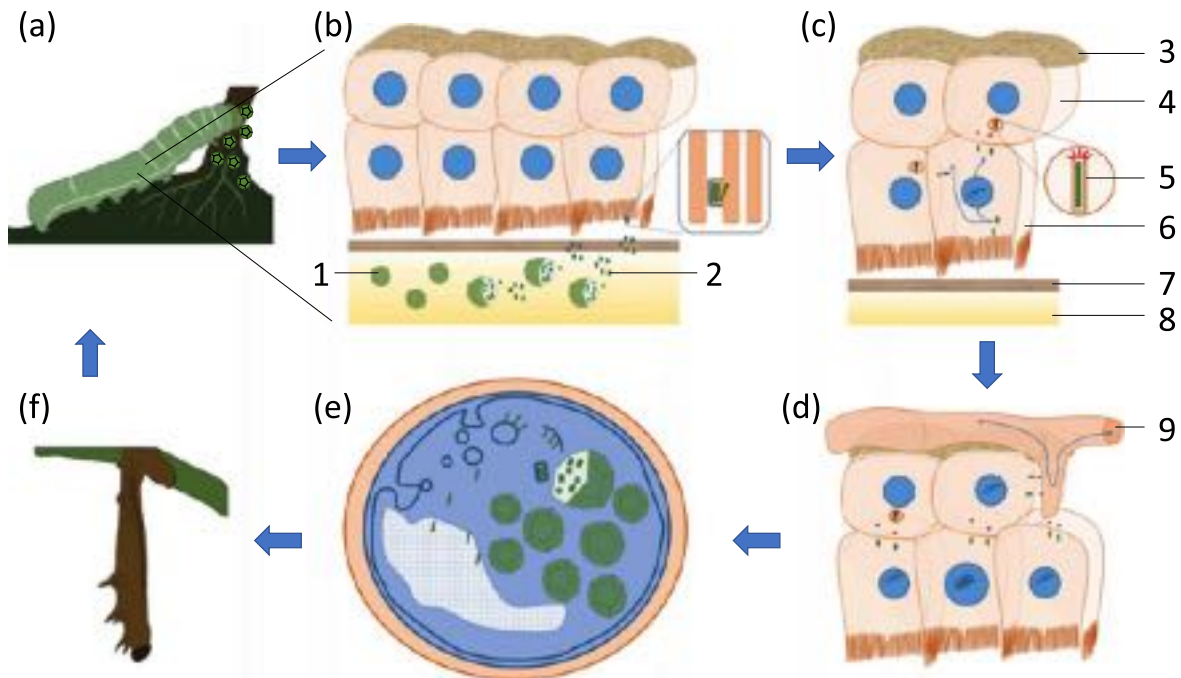


Fig. 3 Schematic representation of the alphabaculovirus infection cycle. A lepidopteran larva ingests plant material contaminated with occlusion bodies (OBs) (a). The OBs dissolve in the alkaline midgut, resulting in release of the occlusion derived viruses (ODVs). The ODVs pass the peritrophic membrane and bind to and fuse with the midgut epithelial cell membrane, after which the nucleocapsids enter the cell (primary infection) (b). The parental nucleocapsids migrate towards the nucleus for the first round of replication forming new nucleocapsids that migrate to the plasma membrane, or directly migrate to the plasma membrane. The nucleocapsids bud out of the cell, acquiring an envelope from the plasma membrane which contains the viral fusion protein (GP64 or F-protein; depicted as red spikes), forming budded viruses (BVs) (c). BVs are responsible for cell to cell spread (secondary or systemic infection). The BVs infect surrounding columnar or regenerative midgut cells via endocytosis. From the midgut, the infection is spread to other larval tissues via the hemolymph and the trachea (d). At the very late stage of infection, the nucleocapsids remain inside the host cell nucleus to be enveloped by a microvesicle membrane that is derived from the inner-nuclear membrane, forming new ODVs. These are subsequently occluded in a protein matrix, forming new OBs. Finally, the cells lyse and the exoskeleton is degraded, resulting in liquefaction of the larvae and release of the OBs into the environment. 1. Occlusion body; 2. Occlusion derived virus; 3. Basal lamina; 4. Regenerative midgut cell; 5. Budded virus; 6. Epithelial midgut cell; 7. Peritrophic membrane; 8. Midgut lumen; 9. Trachea. Drawings produced by Bob Boogaard.

Table 2 Overview of the baculovirus core genes and the function of the encoded proteins

<i>Baculovirus</i> core genes ^a	Function of the encoded protein	
<i>lef-1</i> (<i>ac14</i>)	DNA primase	Replication
<i>lef-2</i> (<i>ac6</i>)	DNA replication/associated with LEF-1	
<i>dnapol</i> (<i>ac65</i>)	DNA polymerase	
<i>p143</i> (<i>ac95</i>)	Helicase	Gene expression
<i>alkexo</i> (<i>ac133</i>)	Alkaline exonuclease	
<i>lef-4</i> (<i>ac90</i>)	mRNA capping enzyme	
<i>lef-8</i> (<i>ac50</i>)	RNA polymerase subunit	
<i>lef-9</i> (<i>ac62</i>)	RNA polymerase subunit	
<i>p47</i> (<i>ac40</i>)	RNA polymerase subunit	
<i>lef-5</i> (<i>ac99</i>)	Transcription initiation factor	
<i>vlf1</i> (<i>ac77</i>)	Very late factor 1 (<i>p10</i> and <i>polh</i> expression)	Nucleocapsid (NC) assembly
<i>ac53</i>	U-box/RING-like domains, NC formation	
<i>vp1054</i> (<i>ac54</i>)	NC assembly	
<i>vp39</i> (<i>ac89</i>)	Major capsid protein	
<i>38k</i> (<i>ac98</i>)	NC assembly	
<i>p49</i> (<i>ac142</i>)	NC assembly	
<i>P6.9</i> (<i>ac100</i>)	DNA condensation, NC protein	
<i>desmop</i> (<i>ac66</i>)	Desmoplakin, NC protein	
<i>odv/bv c42</i> (<i>ac101</i>)	NC associated protein	
<i>p18</i> (<i>ac93</i>)	Egress of NCs	
<i>p48</i> (<i>ac103</i>)	BV production and ODV envelopment	ODV morphogenesis/oral infectivity
<i>p74</i> (<i>pif-0</i>) (<i>ac138</i>)	Per os infectivity factor	
<i>pif-1</i> (<i>ac119</i>)	Per os infectivity factor	
<i>pif-2</i> (<i>ac22</i>)	Per os infectivity factor	
<i>pif-3</i> (<i>ac115</i>)	Per os infectivity factor	
<i>pif-4</i> (<i>ac96</i>)	Per os infectivity factor	
<i>odv-e56</i> (<i>pif-5</i>) (<i>ac148</i>)	Per os infectivity factor	
<i>pif-6</i> (<i>ac68</i>)	Per os infectivity factor	
<i>pif-7</i> (<i>ac110</i>)	Per os infectivity factor	
<i>P95</i> (<i>pif-8</i>) (<i>ac83</i>)	Per os infectivity factor and NC protein	
<i>odv-e18</i> (<i>ac143</i>)	ODV envelope-protein	
<i>odv-e25</i> (<i>ac94</i>)	ODV envelope-protein	
<i>odv-e27</i> (<i>ac144</i>)	ODV envelope-protein	
<i>gp41</i> (<i>ac92</i>)	Tegument protein	Other
<i>ac109</i>	Structural protein with unknown function	
<i>p33</i> (<i>ac92</i>)	Sulfhydryl oxidase/preventing oxidative stress	
<i>ac78</i>	Interacts with sulfhydryl oxidase	
<i>ac81</i>	Non-structural protein of unknown function	

^aName in brackets refers to the gene number on the AcMNPV genome, the first baculovirus complete genome to be sequenced.

may overlap. A characteristic of most baculovirus genomes is the presence of several homologous repeat regions (*hr*'s), consisting of repeats of short, often palindromic sequences, typically 30–40 nt in length. Cell culture assays revealed that *hr*'s function as origins of replication and enhancers of transcription.

Comparison of all fully sequenced baculovirus genomes of all four genera has revealed a set of 38 core genes (Table 2), which are shared among all baculovirus species. Most of these genes encode proteins involved in basic functions, including DNA replication, gene transcription, progeny virion assembly and structure and oral infectivity. Additional sets of genes are shared within genera, or among groups of related baculoviruses. Other genes are unique to one or a few baculovirus species.

Baculovirus transcription is temporally regulated into four phases: immediate early, delayed early, late and very late. Expression of genes in one phase is dependent on the expression of genes in the preceding phase. The expression of some genes extends through several phases. Early genes are expressed prior to DNA replication and are transcribed by host RNA polymerase II. Late and very late genes are transcribed by a viral RNA polymerase and the very late genes comprise two highly expressed genes: the polyhedrin or granulins gene encoding the OB matrix protein and the p10 gene encoding a 10 kDa protein that forms fibrillar structures important for OB formation and release.

Evolution

Baculoviruses are related to other insect viruses with large dsDNA genomes, including nudiviruses (*Nudiviridae*), bracoviruses (endogenized viruses that are derived from the nudiviruses) (genus *Bracovirus* from the family *Polydnnaviridae*) and hytrosaviruses

(*Hytrosaviridae*), all of which are exclusively pathogenic to arthropods, harbor rod-shaped enveloped nucleocapsids and replicate in the nucleus. Nudiviruses and baculoviruses are sister groups, bracoviruses are nested within the nudiviruses and hytrosaviruses are an outgroup. Nudi- and baculoviruses first evolved around 310 Mya in the Paleozoic Era during the Carboniferous Period with the first insects. Further virus diversification is linked to the diversification of insect orders. More distantly related to these insect viruses are whispoviruses, infecting shrimp, of the family *Nimaviridae*.

The baculovirus lineage appeared shortly after the appearance of holometabolous insects – insects with larval stages undergoing complete metamorphosis. Baculoviruses have been characterized in only holometabolous insects from the orders Diptera, Hymenoptera and Lepidoptera. Phylogenetic analyses revealed ancient co-evolution between baculoviruses and these insects, and also showed that the baculovirus ancestor was already specialized to infect larval stages of holometabolous insects.

Within the baculoviruses, the four genera form distinct and major lineages, reflecting co-evolution of baculoviruses and their insect hosts. The deltabaculoviruses are the most ancient baculoviruses and are ancestral to the gammabaculoviruses. Both are ancestral to the lepidopteran-infecting alpha- and betabaculoviruses, which are sister groups. Furthermore, alphabaculoviruses are further subdivided into two groups, the monophyletic Group I alphabaculoviruses that is nested within the Group II alphabaculoviruses (Fig. 1). Two additional subgroups, clade Ia and Ib can be distinguished within Group I alphabaculoviruses, while clade IIa, IIb and IIc are found within Group II alphabaculoviruses (Fig. 1). Phylogenetic analyses of whole genome sequences have greatly expanded our understanding of baculovirus diversity and evolution and showed that host shifts played a major role in the diversification of baculoviruses. Colonization of new ecological niches has sometimes led to viral radiation. Furthermore, extensive horizontal gene transfer between baculoviruses and insect hosts and between baculoviruses and other viruses has occurred.

Infection Cycle

Insects become infected by baculoviruses when an insect larva consumes viral OBs during feeding (Fig. 3). The OBs dissolve rapidly in the alkaline environment of the larval midgut, releasing the ODVs. In order to access the target columnar midgut epithelial cells, the released virus must first penetrate the peritrophic membrane (a mesh of fibers consisting of chitin fibrils linked to glycoproteins and proteoglycans) lining the gut. This process is facilitated by enhancins (metalloproteases), which are encoded by most GV and some NPV. Enhancins are also thought to increase the fusion efficiency between the viral and cellular membranes during infection. The ODV envelope fuses with the host cell plasma membrane, releasing the nucleocapsids into the cytoplasm. Fusion is preceded by binding of the ODV to the microvilli by a specific, protease-sensitive receptor. Attachment and fusion requires a set of proteins, called *per os* infectivity factors (PIFs). These proteins are encoded by core genes, and currently eight *pif* genes are known (Table 2). The released nucleocapsids migrate along cytoplasmic actin filaments to nuclear pores. The nucleocapsid is uncoated either at the nuclear pore or as it enters the nuclear pore to release the DNA core into the nucleus. In the nucleus, transcription of viral genes is initiated, followed by DNA replication, late gene transcription, viral protein translation and nucleocapsid assembly. Nucleocapsids are transported to the basal site of the infected midgut cell where they form BVs by budding out of the cell, thereby acquiring an envelope from the host cell plasma membrane which has been modified to contain viral proteins such as GP64 or F. BVs are responsible for cell to cell spread within the insect host (secondary or systemic infection). Cell entry of BVs occurs via clathrin-mediated endocytosis. Most baculoviruses disseminate throughout the insect infecting all tissues, via the hemolymph (after passing the basal lamina) and/or via the tracheal system. For some baculovirus species, including some GVs, gammabaculovirus and deltabaculovirus, infections are limited to host midgut epithelial cells. The role of BVs in these infections, if any, remains unclear.

At the very late stage of infection, cells switch to ODV production and OB formation (Fig. 3). Within the nucleus, nucleocapsids acquire an envelope from the inner-nuclear membrane. The virions then become occluded within occlusion bodies (polyhedra or granules). This systemic infection is so effective that following death, up to 25% of the larval mass is due to polyhedra. The late expressed virally encoded enzymes chitinase and cathepsin aid cell lysis and exoskeleton degradation, liquefying the larvae and releasing the OBs into the environment. These enzymes are not encoded by baculoviruses whose infection is limited to the midgut. For these, the virus particles are released by sloughing of infected midgut cells and by excretion of viruses.

Baculovirus Transmission and Host Behavioral Manipulation

OBs are highly persistent in the environment. OBs can even survive passage through the gastrointestinal tract of birds, facilitating their long distance dispersal. Baculoviruses also ensure enhanced dispersal by modifying host behavior. Baculovirus-induced behavioral changes were first reported in 1891 by Hofmann, representing the oldest written record of parasitic manipulation of host behavior. Infected *Lymantria monacha* larvae lost their appetite and migrated to the top of trees where they liquefied, termed by Hofmann as “Wipfelkrankheit” (tree-top disease). Such climbing prior to death has been observed for a range of baculovirus-caterpillar combinations, and infected caterpillars were found to move towards the light. In addition, infected caterpillars may show increased locomotion behavior (hyperactivity or hypermobility). These behavioral changes are thought to enhance viral spread, by increasing the area over which the virus is distributed and by making deceased caterpillars more visible to predators including birds. Recent studies have revealed part of the molecular mechanisms underlying these virus-induced behavioral changes. A virally encoded protein tyrosine phosphatase (PTP) is involved in the induction of hyperactivity in BmNPV-infected

B. mori larvae as well as in AcMNPV-infected *S. exigua* larvae. Since *ptp* is restricted to Group I alphabaculoviruses, most likely other viral genes are involved as well for hyperactivity in non Group I alphabaculoviruses. Furthermore it was found that the *egt* gene, present in nearly all baculoviruses and encoding an ecdysteroid UDP-glucosyl transferase (EGT), plays a role in baculovirus-induced tree-top disease in certain virus-host combinations. In LdMNPV-infected *L. dispar* larvae, *egt* is involved in the induction of tree-top disease, however, the *egt* gene of AcMNPV did not induce tree-top disease in *S. exigua* or *Trichoplusia ni* larvae. Baculoviruses apparently possess multiple mechanisms to modify host behavior, dependent on the specific virus-host combination studied.

In addition to horizontal transmission of OBs, by either consumption of OB contaminated food or by cannibalism, baculoviruses are also vertically transmitted. Natural and laboratory populations often harbor covert baculovirus infections. The molecular basis for covert baculovirus infections is currently unclear, however, these covert viruses can be vertically transmitted from parents to offspring. Therefore baculoviruses have a mixed-mode transmission, with both horizontal and vertical transmission, possibly allowing virus persistence when opportunities for horizontal transmission is limited. Covert infections can be re-activated to produce lethal disease, followed by horizontal transmission.

Baculovirus Expression Technology

The discovery in the 1980s that the polyhedrin promotor of AcMNPV could be exploited to produce high levels of recombinant proteins in insect cells marks the beginning of the baculovirus expression vector (BEV) system. Since then, baculovirus-based expression technology has developed extensively and is now widely used to produce proteins of interest and commercial vaccines for both human and veterinarian use, like the Circumvent®PCV-M G2 vaccine developed by Merck against both circovirus and *Mycoplasma hyopneumoniae* in pigs, and the Cervarix vaccine against human papillomavirus. As a BEV is a eukaryotic expression vector, many of the common post-translational modifications that occur to produce mammalian proteins also occur in the BEV system, although glycoproteins produced in insects generally display more uniform, but less complex-glycans than mammals. Attempts are being taken to further “mammalianize” glycoprotein processing in insect cells to further improve the authenticity of baculovirus-expressed mammalian proteins.

Furthermore, recombinant baculoviruses are used as gene delivery vectors for mammalian cells and as expression vectors for adeno-associated virus (AAV)-based gene therapy products.

Baculoviral Insecticides

Since their discovery, baculoviruses have been studied mainly for their value as potent, biologically based insect pesticides. Baculoviruses are naturally occurring pathogens, with therefore a much smaller environmental impact compared to chemical insecticides in the management of agricultural and forest pest insects. Furthermore, baculoviruses are highly specific with a limited host range, have no negative effects on non-target organisms and can be mass-produced, formulated, packaged and stored. More than 60 baculovirus-based pesticides have been utilized worldwide to control diverse insect pests. One of the most effective ones include *Anticarsia gemmatilis* MNPV (AgMNPV) which was successfully used to control the velvet bean caterpillar *A. gemmatilis*, an abundant soybean pest in Brazil and other South American countries. The use of AgMNPV reduced larval populations by 80%, the same level as for insecticide treatment. Other highly successful examples of baculovirus insecticides include *Cydia pomonella* GV (CpGV) to control the codling moth *C. pomonella* in apples, pears and walnuts, *Helicoverpa armigera* NPV (HearNPV) to control the cotton bollworm *H. armigera*, a highly polyphagous pest worldwide and *Spodoptera exigua* MNPV (SeMNPV) to control the beet armyworm *S. exigua*, a highly polyphagous and widely distributed pest both in the field and in greenhouses.

Although baculoviruses have been successfully applied, they have some limitations compared to other (chemical) insecticides, including a slower speed of kill, a very narrow host specificity (which may hinder commercialization), a susceptibility to UV light and high production costs (large scale production still relies on in vivo systems). In addition, resistance against baculoviruses may develop, as occurred in some *C. pomonella* populations in Europe. Since the speed of kill is much faster for smaller instars (and since older instars may bore into stems or fruits being unreachable for any pesticide), the effectiveness of baculovirus application can be optimized by proper monitoring in the field enabling a timely application. To further lower the time to kill and to increase the efficacy of baculoviruses, recombinant baculoviruses have been developed in the laboratory, by deleting baculovirus genes, by adding genes encoding insect-specific toxins from scorpions or spiders, by exchanging genes among different baculoviruses or by reorganizing the genome. The stability of baculoviruses in pesticide formulations can be further improved by adding new and more effective additives. Nonetheless, baculoviruses have a promising future as biopesticides, and can be successfully implemented in integrated pest management strategies.

Further Reading

- Arif, B.M., 2005. A brief journey with insect viruses with emphasis on baculoviruses. *Journal of Invertebrate Pathology* 89, 39–45.
 Arif, B., Escasa, S., Pavlik, L., 2011. Biology and genomics of viruses within the genus *Gammabaculovirus*. *Viruses* 3, 2214–2222.
 Beas-Catena, A., Sánchez-Mirón, A., García-Camacho, F., Contreras-Gómez, A., Molina-Grima, E., 2014. Baculovirus biopesticides: An overview. *Journal of Animal & Plant Sciences* 24, 362–373.

- Bezier, A., Thézé, J., Gavory, F., *et al.*, 2015. The genome of the nucleopolyhedrosis-causing virus from *Tipula oleracea* sheds new light on the *Nudiviridae* family. *Journal of Virology* 89, 3008–3025.
- Gasque, S.N., van Oers, M.M., Ros, V.I.D., 2019. Where the baculoviruses lead, the caterpillars follow: Baculovirus-induced alterations in caterpillar behaviour. *Current Opinion in Insect Science* 33, 30–36.
- Han, Y., van Oers, M.M., van Houte, S., Ros, V.I.D., 2015. Virus-induced behavioural changes in insects. In: Mehlhorn, H. (Ed.), *Hostmanipulations by Parasites and Viruses*. Parasitology Research Monograph. Springer International Publishing, pp. 149–174.
- Harrison, R.L., Herniou, E.A., Jehle, J.A., *et al.*, 2018. ICTV virus taxonomy profile: *Baculoviridae*. *Journal of General Virology* 99, 1185–1186.
- Harrison, R., Hoover, K., 2012. Baculoviruses and other occluded insect viruses. In: Vega, F., Kaya, H. (Eds.), *Insect Pathology*, second ed. Amsterdam: Elsevier, pp. 73–131.
- Rohrmann, G.F., 2019. *Baculovirus Molecular Biology*, fourth ed. Bethesda, MD: National Center for Biotechnology Information.
- Thézé, J., Bézier, A., Periquet, G., Drezen, J.-M., Herniou, E.A., 2011. Paleozoic origin of insect large dsDNA viruses. *Proceedings of the National Academy of Sciences of the United States of America* 108, 15931–15935.
- Thézé, J., Lopez-Vaamonde, C., Cory, J.S., Herniou, E.A., 2018. Biodiversity, evolution and ecological specialization of baculoviruses: A treasure trove for future applied research. *Viruses* 10, 366.
- van Oers, M.M., Pijlman, G.P., Vlak, J.M., 2015. Thirty years of baculovirus-insect cell protein expression: From dark horse to mainstream technology. *Journal of General Virology* 96, 6–23.
- Wang, M., Hu, Z., 2019. Cross-talking between baculoviruses and host insects towards a successful infection. *Philosophical Transactions of the Royal Society B* 374, (20180324).
- Wennmann, J.T., Keilwagen, J., Jehle, J.A., 2018. Baculovirus Kimura two-parameter species demarcation criterion is confirmed by the distances of 38 core gene nucleotide sequences. *Journal of General Virology* 99, 1307–1320.
- Williams, T., Bergoin, M., van Oers, M.M., 2017. Diversity of large DNA viruses of invertebrates. *Journal of Invertebrate Pathology* 149, 4–22.
- Williams, T., Virto, C., Murillo, R., Caballero, P., 2017. Covert infection of insects by baculoviruses. *Frontiers in Microbiology* 8, 1337.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsDNA-viruses/w/baculoviridae
ICTV online.

Baculoviruses: Molecular Biology and Replication (*Baculoviridae*)

Monique M van Oers, Wageningen University and Research, Wageningen, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AcMNPV	<i>Autographa californica</i> multiple nucleopolyhedrovirus	LeseNPV	<i>Leucania separata</i> nucleopolyhedrovirus
BmNPV	<i>Bombyx mori</i> nucleopolyhedrovirus	MaviNPV	<i>Maruca vitrata</i> nucleopolyhedrovirus
ChchNPV	<i>Chrysodeixis chalcites</i> nucleopolyhedrovirus	OpMNPV	<i>Orgyia pseudotsugata</i> multiple nucleopolyhedrovirus
CuniNPV	<i>Culex nigripalpus</i> nucleopolyhedrovirus	SeMNPV	<i>Spodoptea exigua</i> multiple nucleopolyhedrovirus
HearNPV	<i>Helicoverpa armigera</i> nucleopolyhedrovirus		

Glossary

Budded virus (BV) The type of baculovirus virion that is formed by budding from the plasma membrane of infected cells and mediates the systemic spread of infection in the infected insect.

Granulovirus (GV) A betabaculovirus, forming granular-shaped occlusion bodies containing only a single virion each.

Homologous repeat (hr) A DNA sequence that is typically comprised of a series of imperfect palindromes and repeated at various locations in a baculovirus genome. Baculovirus homologous repeats function as origins of DNA replication and as transcriptional enhancers.

Nucleocapsid (NC) Nucleoprotein helical structure forming the core of the enveloped baculovirus.

Nucleopolyhedrovirus (NPV) Indication used for all baculoviruses that form polyhedral-shaped occlusion

bodies (OBs) in the nucleus (compare granulovirus, GV, with granular OBs).

Occlusion body (OB) A typical polyhedral or granular-shaped structure made of a paracrystalline protein matrix in which the baculovirus ODVs are embedded and that is found in the nucleus of infected cells.

Occlusion derived virus (ODV) The type of baculovirus virion that is assembled in the nucleus and becomes embedded within the paracrystalline matrix of the occlusion body.

PIF proteins A conserved series of crucial *per os* infectivity factors encoded by baculovirus genomes and being specifically required for oral infection.

Polyhedrin The major occlusion body protein of nucleopolyhedroviruses.

Introduction

Baculoviruses are insect-infecting viruses characterized by rod-shaped, enveloped virions that are packaged in viral occlusion bodies (OBs). These OBs allow the virus to survive outside a host insect for long times. Baculoviruses replicate in the nucleus of infected cells and are further characterized by their double-stranded (ds), circular DNA genomes, ranging in size from 80 to 180 kbp. Historically, baculoviruses were divided based on their OB appearance into granuloviruses (GVs), with a single virion per granular-shaped OB, and nucleopolyhedroviruses (NPVs) with many virions per polyhedral-shaped OB. The family *Baculoviridae* is subdivided into four genera, Lepidopteran-infecting baculoviruses are classified in the genus *Alphabaculovirus* or *Betabaculovirus*, while the *Gammabaculovirus* and *Deltabaculovirus* genera encompass the hymenopteran and dipteran insect infecting baculoviruses, respectively. So far, only one deltabaculovirus is known, *Culex nigripalpus* (Cuni) NPV that infects mosquito larvae. The alphabaculoviruses, on the other hand, form the most widely studied group of baculoviruses, which is primarily due to the availability of permissive insect cell cultures for many of these viruses.

Autographa californica multiple nucleopolyhedrovirus (AcMNPV) represents the most in-depth analysed alphabaculovirus. Furthermore, AcMNPV and *Bombyx mori* (Bm) NPV are used as expression vectors to produce (therapeutic) proteins, sub-unit vaccines and virus-like particles in cultured insect cells or caterpillars. AcMNPV is also an important tool to produce viral vectors for gene therapy purposes, especially for recombinant Adeno-associated virus (AAV). In addition, AcMNPV is being used for mammalian cell transduction and research is ongoing to use AcMNPV in human medicine as a gene delivery vector. A number of alpha- and betabaculoviruses have been developed into bio-insecticides, while biocontrol programs based on gammabaculoviruses to control sawflies also exist. The remainder of this article will focus on viruses belonging to the genus *Alphabaculovirus*.

Infection Cycle

Two Virion Phenotypes

Two distinct types of virions are produced in the infection cycle of alphabaculoviruses: the budded virus (BV) and the occlusion-derived virus (ODV). The latter type is embedded in OBs. BVs represent the first phenotype to be produced by an infected cell. BVs consist of a single nucleocapsid surrounded by an envelope obtained by budding from the plasma membrane of the infected cell. As the infection progresses, ODVs are assembled in the nucleus of infected cells and acquire an envelope that is derived from the inner nuclear membrane. As a consequence of the different sources of the BV and ODV envelopes, they have a significantly different lipid and protein composition (see Fig. 1). Unlike the ODVs of the baculoviruses classified in the other taxa, the ODVs of alphabaculoviruses can contain single (S) or multiple (M) nucleocapsids (NCs) per virion, depending on virus species. The S and M designation (SNPV and MNPV; *e.g.*, AcMNPV) does not hold taxonomic significance above the species level but is in practice an easy indicator to discriminate virus isolates obtained from the same host, but that belong to different viral species.

Towards the end of the infection, the ODVs become occluded in a para-crystalline protein matrix to form OBs that accumulate in the nucleus of the infected cell. OBs represent a feature that is common to all the currently classified baculoviruses, but the shape and size of the OBs depends on the virus isolate. The AcMNPV OBs are approximately 10 µm in diameter and are clearly visible using a light microscope. In NPVs these OBs have polyhedral shapes, hence the historical name “nuclear polyhedrosis” for the observed pathology that gave the NPVs their name. The major component of the OBs of most baculoviruses is the OB matrix protein, polyhedrin, which forms the bulk of the para-crystalline array in which the ODVs are embedded. Betabaculoviruses have an evolutionarily related, though slightly different protein, called granulins. Mature OBs are surrounded by a carbohydrate layer known as the polyhedral envelope or calyx. The alphabaculovirus-specific phosphoprotein called PP34 or the polyhedral envelope protein, is the major protein associated with the calyx in this group of viruses. The OBs of an AcMNPV *pp34* deletion mutant were found to have an increased sensitivity to alkaline disruption and an enhanced virulence in fourth instar *Spodoptera exigua* larvae, suggesting that the polyhedral calyx may stabilize the OBs in the environment. EM images of such a mutant showed OBs with a pitted appearance, apparently due to the loss of ODV particles.

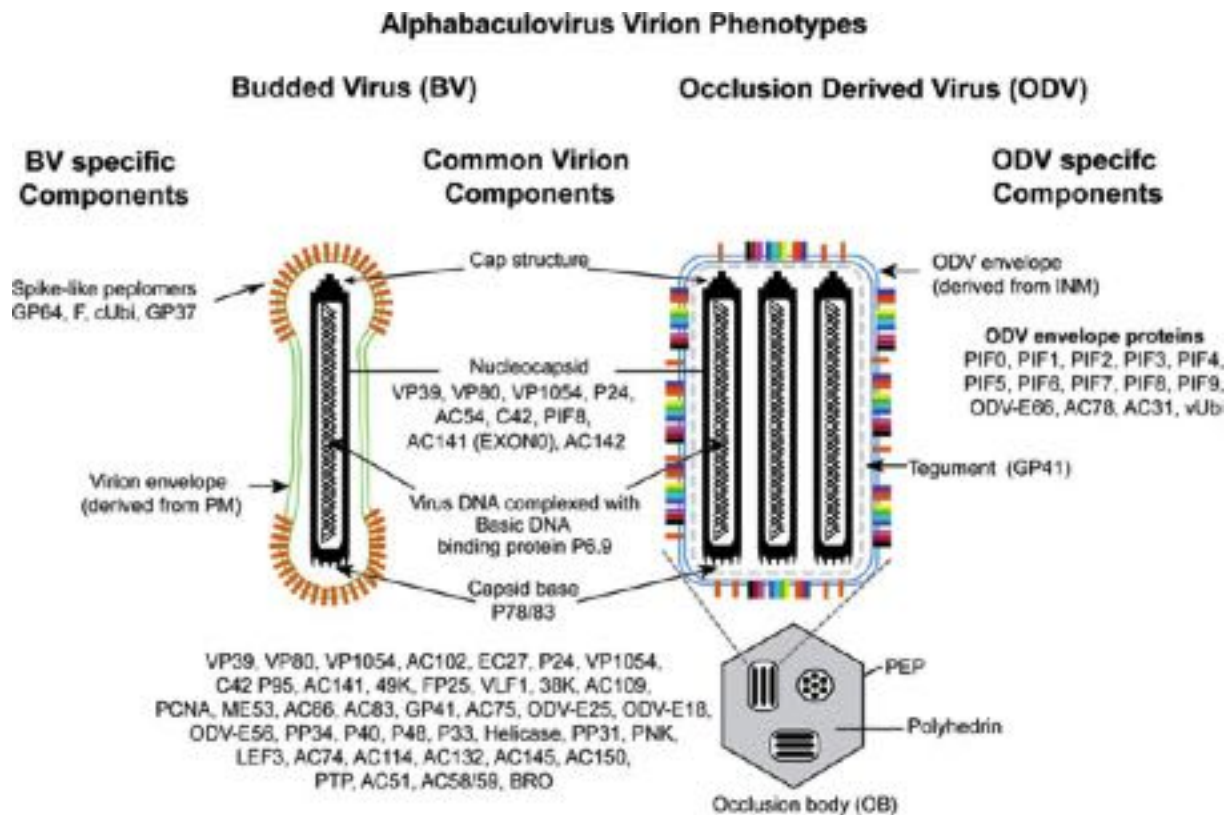


Fig. 1 Schematic diagram of alphabaculovirus virion phenotypes showing BV and ODV structures. Components common to both virion phenotypes are shown in the center and components unique to each phenotype are indicated on the left and right. Note the BV glycoprotein spikes at both ends of the BV, as well as the ODV envelope proteins, which include a complex of *per os* infectivity factors (PIFs). PM = plasma membrane; INM = inner nuclear membrane. Adapted from Encyclopedia of Virology 3rd edition based on new insights (courtesy of D.A. Theilmann) and originally derived from Blissard, G.W., 1996. Cytotechnology 20, 73–93 with kind permission from Springer Science and Business Media.

Oral and Systemic Infection

The infection cycle is initiated when insects orally ingest OBs by consuming contaminated plant material. The OBs pass through the foregut, enter the midgut, and disassemble in response to the alkaline environment in there. This process releases the ODVs, which then traverse the peritrophic membrane (a protein–chitin structure that lines the gut) in order to infect the columnar epithelial cells of the lepidopteran midgut. This infection process may be aided by metalloproteases called enhancins, which degrade the peritrophic membrane and are encoded by certain alpha- and most betabaculoviruses. The ODVs bind to microvilli on the surface of midgut epithelial cells and are believed to enter the cells by direct fusion of the ODV envelope with the plasma membrane. A set of ten baculovirus *per os* infectivity factors (PIF proteins), embedded in the ODV envelope, is essential for oral infectivity and is involved in binding and fusion. Upon entry into the midgut epithelial cell, the NCs are released and transported along actin filaments to the nucleus, where transcription of viral genes is initiated.

Viral DNA replication and NC assembly occur within the nucleus in a viral structure known as the virogenic stroma (see Fig. 2). After an initial round of replication in the gut, BVs are produced at the basal side of the columnar cells and these BVs spread the infection systemically to the other tissues. BVs carry a viral fusion protein in their envelope allowing them to enter host cells by endocytosis. A recent study showed that internalization of AcMNPV BVs is facilitated by actin polymerization and dynamin. Microtubules, on the other hand, are involved in intracellular transport of BV-loaded endosomes. While still in an early endosomal stage, the viral envelope fuses with the endosomal envelope, releasing the NCs into the cytoplasm. These NCs then move to the nucleus along actin filaments. Most cells and tissues of the host insect are permissive for infection by BVs. These secondary infected cells first start to produce more BVs, and subsequently also start to produce ODVs and OBs.

Dissemination of OBs

Finally, the nuclei of the infected host cells will be filled with OBs. At least two viral enzymes, chitinase and cathepsin, assist in the release of OBs into the environment in a process called liquefaction: the disruption of infected cells and tissues by the breakdown of chitin and proteins, respectively. A third protein, called P10, helps to disintegrate the infected cell nuclei, thereby further dispersing individual OBs. The OBs released into the environment can remain stable for years and remain infectious if protected from ultraviolet light. The spread of OBs into the environment is probably aided by viral modification of host behaviour. Many alphabaculovirus-infected caterpillars show pre-death climbing behaviour (Wipfelkrankheit) and finally can be found hanging from the upper segments of the plants on which they fed. Release of the OBs at such elevated locations is believed to enhance transmission by contaminating the lower regions of the plant. Up in the plants, the infected caterpillars are also an easier prey for predating birds and these have been shown to aid in dispersal of OBs over large distances. In some baculovirus species, the *egt* gene, which encodes the enzyme ecdysteroid glucosyl transferase (EGT) that prevents larval moulting, is needed for pre-death climbing behaviour. In other species EGT is not required for climbing behaviour or rather plays an indirect role by prolonging the life span of infected caterpillars, thereby providing the time in which the caterpillars start to climb. The exact induction mechanism therefore remains unclear and most likely varies depending on the combination of virus and host species. In addition, it was shown that both AcMNPV and BmNPV induce hyperactivity in infected caterpillars for which the viral protein tyrosine phosphatase (PTP) is responsible, a protein specific for the clade of alphabaculoviruses to which these two viruses belong.

Genome Organization and Content

All baculoviruses have circular, double-stranded DNA genomes that are negatively supercoiled. The complete genomic sequences of over 200 alphabaculoviruses have been deposited to GenBank (NCBI). The smallest alphabaculovirus genome is that of *Maruca vitrata* (Mavi) NPV at 111,953 bp, while *Leucania separata* (Lese) NPV has with 168,041 bp the largest alphabaculovirus genome. The numbers of predicted genes with open reading frames (ORFs) of 50 amino acids or greater range from 126 (MaviNPV) to 169 (LeseNPV). The best studied alphabaculovirus, AcMNPV, has a genome size of 133,894 bp with 154 predicted genes and was the first baculovirus genome to be sequenced in its entirety. A distinct feature of nearly all baculoviruses is the presence of regions with homologous repeats (*hrs*) distributed over the genome. The *hr* regions contain a series of repeated palindromic sequences and can vary significantly in length. In the AcMNPV genome seven *hrs* are present, which contain repeats of an imperfect palindrome with a central EcoRI restriction site. Transient transcription and plasmid replication assays have shown for alphabaculoviruses that the *hr* sequences function as transcription enhancers and as replication origins. In addition, single copy *non-hr* replication origins have been identified in the genomes of AcMNPV, *Orgyia pseudotsugata* (Op) MNPV, and *Spodoptera exigua* (Se) MNPV.

Baculovirus genomes contain tightly packed ORFs with very short intergenic regions. Early and late genes are dispersed over the genome and both strands of the dsDNA genome carry approximately equal numbers of ORFs. Baculovirus genomes are relatively stable, but do display some plasticity, resulting in small variations in genome sequence when comparing virus isolates from the field. During evolution, recombination events and transposon insertions have led to the integration of new genes from co-infecting viruses or from the host insect. This resulted in sometimes extensive differences in gene content between baculoviruses belonging to different viral species. However, a set of 38 core genes is conserved in all sequenced baculoviruses (marked blue in Fig. 3) and as such, form the baculovirus genomic hall mark. The core genes are often involved in crucial processes in the viral replication cycle,

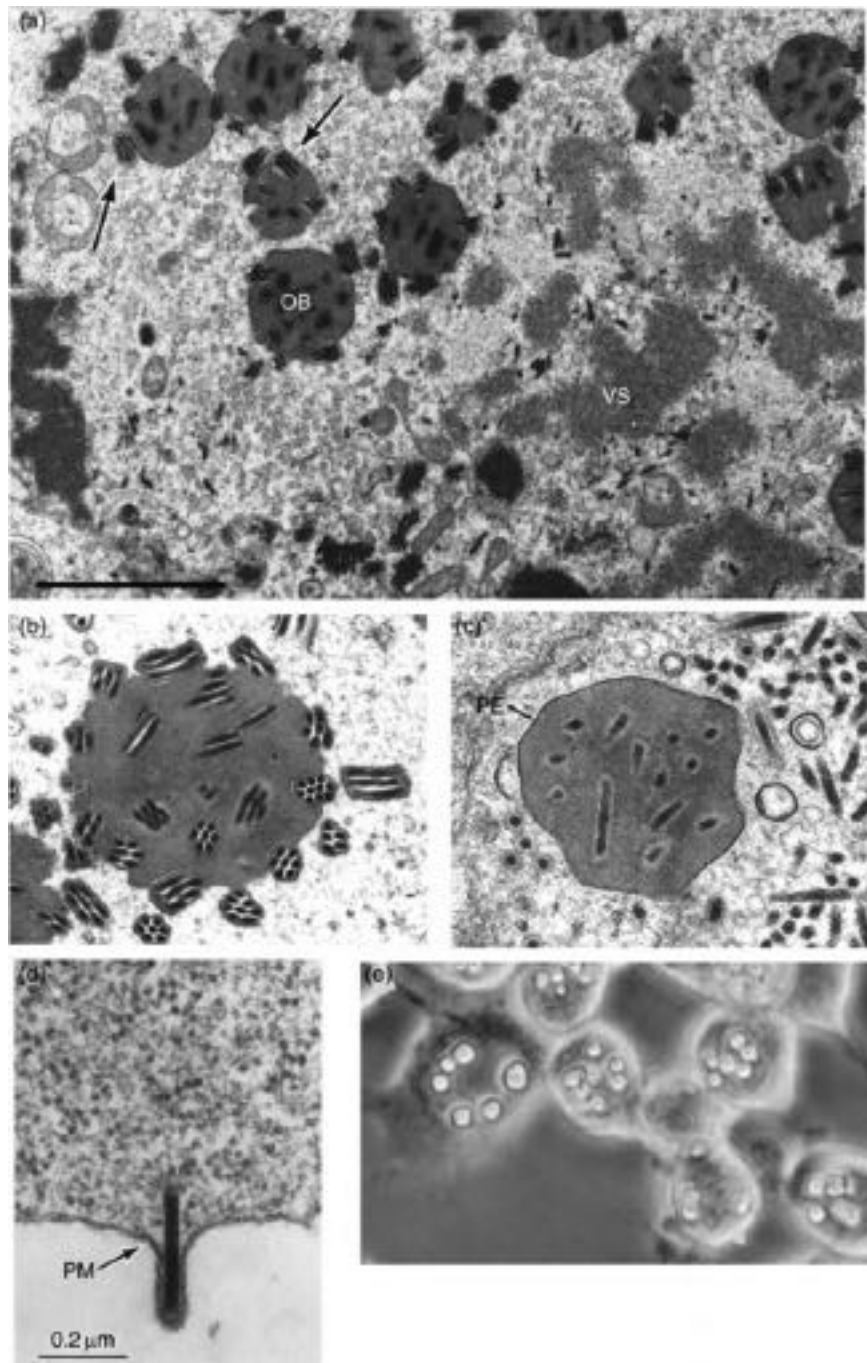
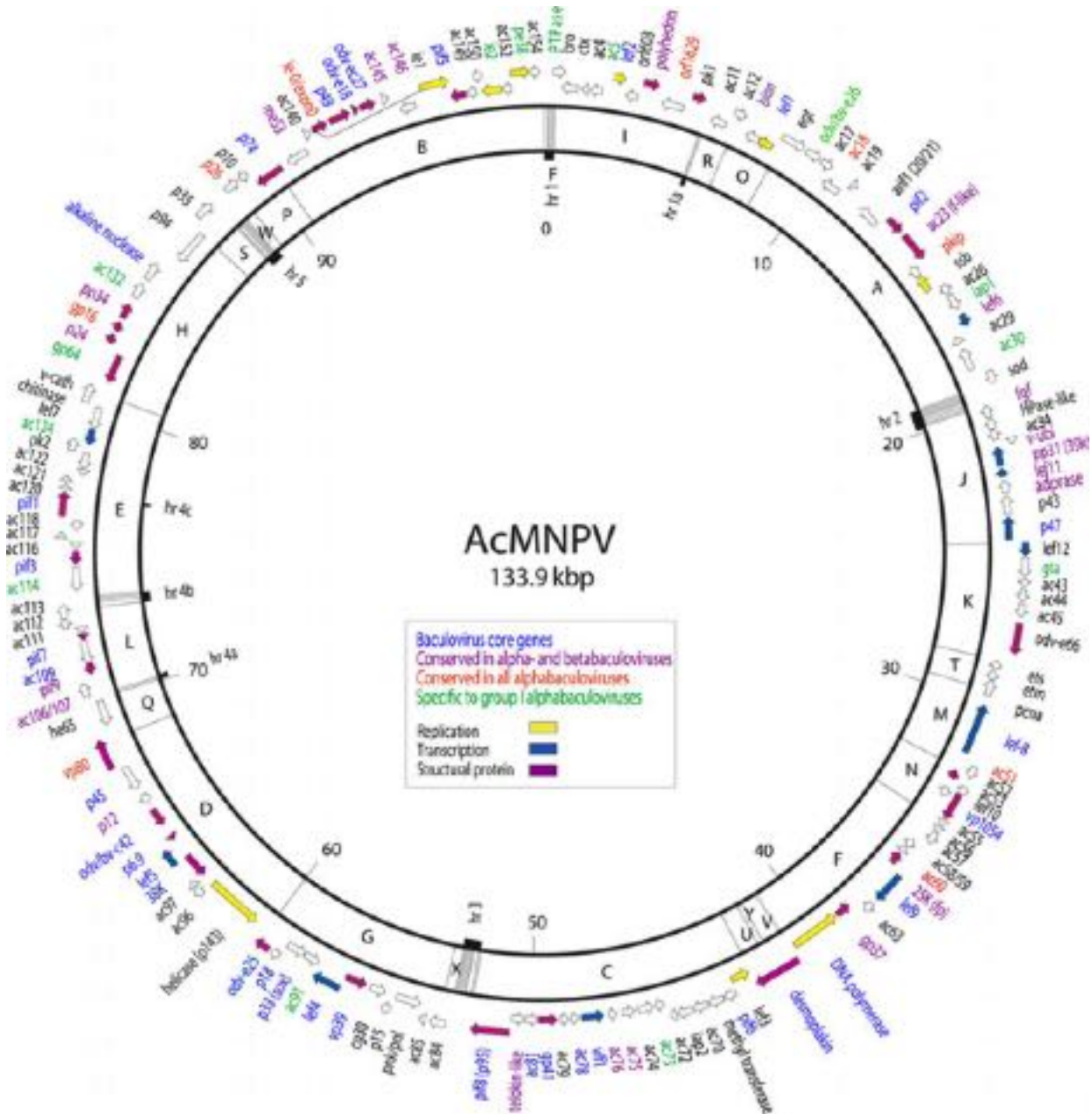


Fig. 2 Electron and light micrographs of baculovirus-infected insect cells. (a) An insect cell infected with AcMNPV showing the nuclear virogenic stroma (VS) that contains developing nucleocapsids and the OBs showing ODVs in the occlusion process (arrows). Scale = 1 μm . (b) An OB of the MNPV type, showing the process of occlusion of ODVs, containing multiple nucleocapsids in each virion. Many ODVs can be found in an OB. (c) A mature OB of the SNPV type, showing a single nucleocapsid in each virion (ODV) and many ODVs per occlusion body. Also note the formation of the polyhedral envelope (PE). (d) An NPV nucleocapsid budding from the plasma membrane (PM) of an infected cell. (e) Light microscopy of Tn5b1-4 cells infected with AcMNPV producing large relatively uniform OBs. (a-d) Courtesy of R.R. Granados and (e) kindly provided by G.W. Blissard.

such as midgut entry, DNA replication, transcription of viral genes and virion assembly. (For an overview table of the core genes please see the supplementary material). However, not all core genes are essential to allow the virus to infect a host; they may also play a role in virus fitness. On the other hand, genes that are not categorized as core genes may be crucial to the infection process for the viruses that carry these genes.

Phylogenetic analysis using the combined sequences of the core genes divides the baculoviruses in four major evolutionary clades and these correspond with the four genera in the family *Baculoviridae*, which were mentioned above (Fig. 4). Such a



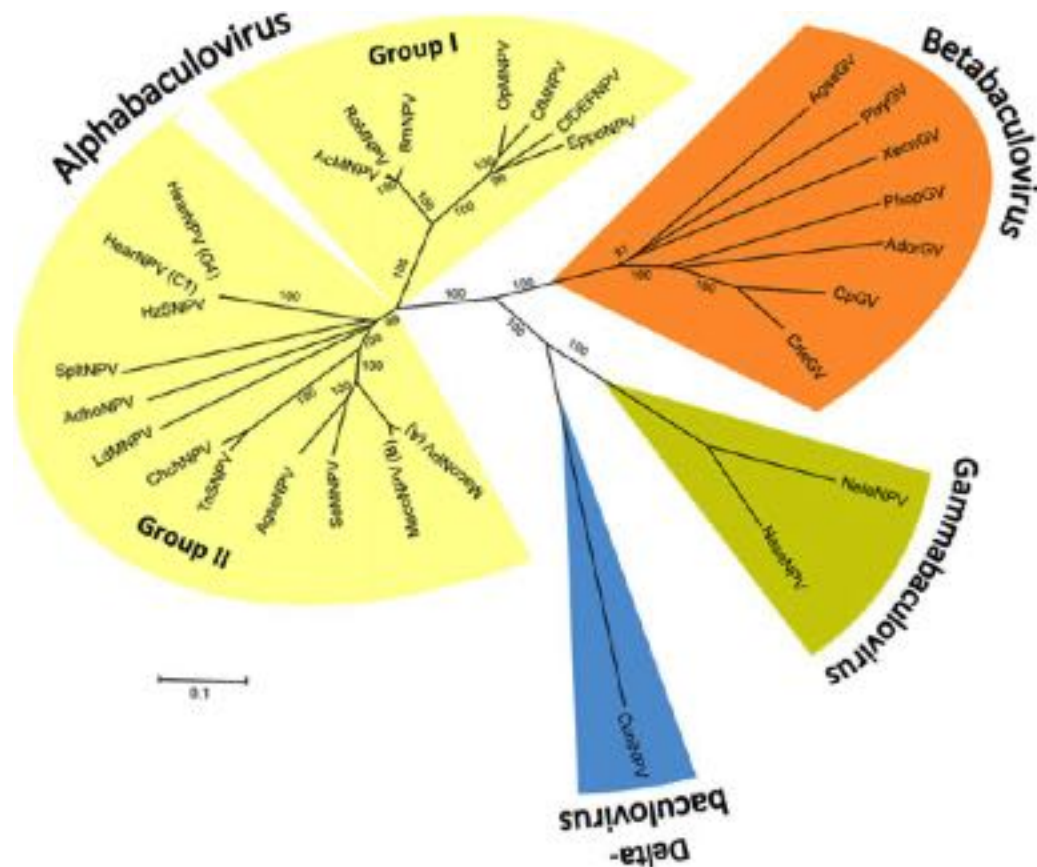


Fig. 4 Neighbour-joining tree based on the amino acid alignment of 29 (of 38) core genes as identified in the first 29 sequenced baculovirus genomes. The alignment comprised 16349 positions. All branches have bootstrap values exceeding 50%. Bootstrap values > 95% are given along the branches. Adapted from Jehle, J.A., Blissard, G.W., Bonning, B.C., *et al.*, 2006. Archives of Virology 151, 1257–1266 with permission from Springer Nature Switzerland AG.

clades. Baculoviruses that lack particular orthologues of non-core genes may encode evolutionarily unrelated, but functionally similar proteins to be able to execute these crucial tasks. Functions of a number of genes important for alphabaculoviruses are described in the sections below.

Baculovirus Gene Expression

Temporal Regulation of Transcription

Baculovirus genes are expressed in a temporal cascade, beginning with immediate early and delayed early gene expression, which is followed by late and very late gene expression. After the initial uncoating of the viral genome in the nucleus, transcription of viral early genes is initiated by host RNA polymerase II, resulting in early transcription and the subsequent production of proteins associated with viral DNA replication and late gene expression. A variety of regulatory and even structural protein genes are also transcribed in the early phase. Production of early gene products results in the initiation of viral DNA replication, and the assembly and activity of a virus-encoded RNA polymerase. This viral RNA polymerase recognizes unique viral late promoters resulting in the transcription of genes encoding viral structural proteins as well as other late genes. Late gene transcription starts just prior to or concomitantly with the onset of DNA replication. Very late gene expression occurs at the terminal part of the replication cycle and includes the hyper-expressed viral genes *polh* and *p10*, encoding the major occlusion body protein polyhedrin and the 10K protein (P10) mentioned above that forms fibrillar structures in the nucleus and cytoplasm.

Early gene expression

Transfected purified baculovirus DNA is infectious, indicating that no viral proteins are required to initiate or mediate early transcription from the viral genome. Early viral gene expression is dependent though upon the host cell's RNA polymerase II complex and is sensitive to the drug α -amanitin. The early genes have been divided into two categories, immediate early (IE) and delayed early (DE) genes. IE genes require only cellular factors for expression whereas DE genes are either dependent on, or substantially upregulated by prior viral

Table 1 Conservation of non-core genes in lepidopteran baculoviruses^{a,b}

A. Genes conserved in all alpha- and betabaculoviruses	
<i>polh</i> (ac8) ^c	Occlusion body matrix, granulin in GVs
<i>bion</i> (ac13)	Baculovirus protein associated with inner and outer nuclear membrane (BION), involved in nuclear lipid accumulation
<i>f</i> protein gene (ac23) ^d	BV fusion protein (group II NPVs) or truncated homolog (group I NPVs)
<i>dbp</i> (ac25) ^e	DNA binding protein
<i>lef6</i> (ac28)	Unknown role in late gene transcription
<i>fgf</i> (ac32)	Fibroblast growth factor
<i>v-ubi</i> (ac35)	Ubiquitin-like protein
<i>pp31</i> (ac36)	PP31; phosphoprotein with ss DNA binding capacity
<i>lef11</i> (ac37)	Essential factor for DNA replication
<i>adprase</i> (ac38)	ADP-ribose pyrophosphatase (ADPRase).
<i>fp25</i> (ac61) ^e	Nuclear transport of ODV envelope proteins
<i>gp37</i> (ac64)	Fusolin-like protein, unclear function
<i>ac82</i>	Telokin-like protein
<i>p12</i> (ac102)	ODV-associated, nuclear localisation of G actin
<i>pif9</i> (ac108) ^e	Per os infectivity factor
<i>p24</i> (ac129)	Capsid protein
<i>pp34</i> (ac131) ^e	Polyhedral envelope-associated phosphoprotein (PEP)
<i>me53</i> (ac139)	NC associated protein that interacts with GP64 in virus budding
<i>ac145</i> ^e	Possible role in oral infectivity
<i>ac146</i>	Structural protein of BV and ODV
<i>ac75</i> ^e , <i>ac76</i> ^e , <i>ac106/107</i> ^e	Unknown function
B. Genes conserved in all alphabaculoviruses	
<i>orf1629</i> (ac9)	P78/83; Essential structural protein of the nucleocapsid
<i>egt</i> (ac15) ^e	Ecdysteroid UDP-glucosyltransferase, inhibits moulting
<i>pkip</i> (ac24)	Stimulates the activity of the viral protein kinase-1 in vitro
<i>ac51</i>	DNAJ domain protein, unknown function
<i>chaB</i> (ac60) ^e	Unknown function
<i>vp80</i> (ac104)	Essential structural protein of the nucleocapsid
<i>p10</i> (ac137) ^e	Major fibrillar structural protein, OB formation and dissemination
<i>exon0</i> (ac141) ^e	NC associated, BV nuclear egress
<i>ie0</i> (ac141 and ac147)	Transcriptional trans-activator, DNA replication
<i>ie1</i> (ac147) ^e	Transcriptional trans-activator, DNA replication
<i>ac18</i> , <i>ac102</i> ^e	Unknown functions
C. Genes conserved in group I and in almost all group II alphabaculoviruses	
<i>arif1</i> (ac20/21)	Rearrangement of cellular actin
<i>odv-e66</i> (ac46) ^f	ODV-E66
<i>lef10</i> (ac53a) ^f	Late essential factor 10
<i>iap2</i> (ac71)/ <i>iap3</i>	RING domain protein-possible apoptosis inhibition
<i>gp16</i> (ac130)	Unknown function
<i>p26</i> (ac136)	Unknown function
<i>ac17</i> , <i>ac19</i> , <i>ac26</i> , <i>ac29</i> ^d , <i>ac43</i> , <i>ac55</i> , <i>ac56</i> , <i>ac58/59</i> , <i>ac120</i>	Unknown functions
D. Genes conserved in all group I alphabaculoviruses	
<i>ptp</i> (ac1)	Protein tyrosine phosphatase, OB formation, behavioural manipulation
<i>bv-odve26</i> (ac16)	BV and ODV envelope associated protein
<i>iap1</i> (ac27)	RING domain protein-possible apoptosis inhibition
<i>gta</i> (ac42)	Global trans-activator-like protein
<i>ac44</i> ^g	Homolog of ascovirus iap/RING finger protein
<i>gp64</i> (ac128)	BV specific glycoprotein required for virion entry and budding
<i>lef7</i> (ac125) ^g	Late gene expression, role in DNA replication
<i>ie2</i> (ac151)	RING finger domain protein, transcriptional activation
<i>pe38</i> (ac153) ^g	RING domain protein, transcriptional activation
<i>ac4</i> ^h , <i>ac14</i> ^h , <i>ac30</i> , <i>ac73</i> , <i>ac74</i> ^h , <i>ac79</i> , <i>ac91</i> ^h , <i>ac111</i> ^h , <i>ac114</i> , <i>ac117</i> ^h , <i>ac124</i> , <i>ac132</i>	Unknown functions

^aConservation levels were retrieved from Rohrmann, G., 2010. Chapter 12 – The AcMNPV genome: Gene content, conservation, and function. *Baculovirus Molecular Biology*, third ed. Bethesda, MD: National Center for Biotechnology Information. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK138304/>. Information on the function of BION (ac13) was retrieved from Nagamine, T., Inabe, T., Saiko, Y., 2019. *Virology* 532, 108–117. Available at: <https://doi.org/10.1016/j.virol.2019.04.006>.

^bA similar Table, but then for the baculovirus core genes, can be found in the chapter *Baculoviruses: General Features* by V.I.D. Ros.

^cAlso present in gammabaculoviruses.

^dAlso present in the deltabaculovirus CuniNPV.

^eAlso present in individual members of one or more other lineages.

^fAlso present in individual members of one or more other lineages.

^gAlso present in individual members of one or more other lineages.

^hAlso present in individual members of one or more other lineages.

gene expression. Early gene products are involved primarily in gene regulation, host modification, DNA replication, or are required for late gene expression. Many IE genes have a common motif, TATA-N_(24–26)CAGT, at their transcription start site. The promoters of IE and DE genes resemble typical eukaryotic RNA pol II promoters and contain host cell transcription factor binding sites.

The key viral regulatory proteins for early (and late) gene transcription are the immediate early proteins IE0 and IE1, encoded by the essential *ie0* gene complex. This *ie0* gene has two promoters and produces spliced and non-spliced transcripts. It is the only known baculovirus gene that produces two viral proteins via spliced transcripts, although several more baculovirus mRNAs are known to be generated through splicing. The IE0 protein contains the entire IE1 amino acid sequence, but in addition has an N-terminal extension that is variable in length depending on the viral species (54 amino acids in AcMNPV). Both the IE0 and IE1 proteins are present during the entire infection process due to the fact that the *ie0* transcript can also produce IE1, in addition to IE0, by internal translation initiation at the start codon of the *ie1* sequence. IE0 has peak expression prior to viral DNA replication, while IE1 becomes more abundant than IE0 when replication starts. IE1 levels continue to increase until the final stages of the infection. IE0 appears to be specific to alphabaculoviruses. IE0/1 is a typical acidic domain transcriptional trans-activator, similar to the herpesvirus VP16 protein, and activates viral gene transcription by both enhancer-dependent and independent mechanisms. Both proteins form dimers that bind to the genomic *hr* sequences, which serve as the transcriptional enhancers. IE0/1 is also essential for the viral replication complex, where it is believed to play the role of an origin binding protein. IE0 and IE1 can support viral DNA replication and BV production equally, but for some promoters (*lef4*, *lef6*, *ac79*, *p35*, *ac18*, *lef3*, *ac111*, and *39K*) IE0 is a more potent activator than IE1. This difference in potency is seen only when IE1/IE0 are expressed at relatively low levels, in line with the situation at early times post-infection. It has been suggested that IE0 expression in midgut cells may stimulate rapid DNA replication and BV production to allow the progeny virus to quickly escape the midgut and allow systemic dissemination of the infection.

Additional IE genes include the alphabaculovirus-specific *ie2* and *pe38* genes, which have been shown to augment viral transcription and, as indicated previously, viral DNA replication. Unlike *ie0/1*, both *ie2* and *pe38* can be deleted without interfering with the capacity to replicate. For viruses with *ie2* or *pe38* gene deletions, however, BV production and viral DNA replication are less efficient. Some of the most highly expressed viral IE genes are *he65* and *me53*. The presence of G-actin within the nucleus upon infection has been shown to be in part due to the function of HE65, whereas ME53 is needed for optimal virus production and in infected cells is associated with GP64 at foci thought to be budding sites. Overall, IE genes play crucial regulatory roles that coordinate the infection cycle. The *gp64* gene, encoding the AcMNPV BV fusion protein, is transcribed early in infection by host factors, but also has a late promoter element.

DE gene expression requires other viral factors, such as IE0/IE1 to achieve above basal levels of expression. An example of a DE promoter is the AcMNPV *pp31* promoter, which contains a TATA-CAGT sequence at the transcription start site and is dependent on IE1 for activated transcription. Also, the above-mentioned genes *lef4*, *lef6*, *ac79*, *p35*, *ac18*, *lef3*, *ac111*, and *39K* belong to the DE genes. In general, DE gene products are required for DNA replication and/or late gene expression.

Late and very late gene expression

Baculovirus late genes are transcribed by a viral-encoded DNA-dependent α -amanitin-insensitive RNA polymerase that recognizes baculovirus late promoters. Studies of the purified NPV RNA polymerase have shown that it is comprised of four major proteins, LEF4, LEF8, LEF9, and P47. Homologs of the corresponding genes have been identified in all baculovirus genomes sequenced to date. LEF4 has been shown to have enzymatic activity for RNA capping, and LEF8 and LEF9 contain conserved motifs found in other RNA polymerases. With only very rare exceptions, the baculovirus late promoters recognised by this viral RNA polymerase contain the sequence TAAG and transcriptome analysis has shown that the 5' end of late transcripts corresponds with the second position of this motif. In model late promoters that have been examined, only very short sequences immediately up- and downstream of the TAAG sequence were identified as components of the late promoter. Regions farther downstream have been shown to affect levels of very late gene expression.

Using transient assays, twenty AcMNPV genes were identified as important or necessary for late gene expression: the late expression factor genes (*lef1* to *lef12*), *pp31*, *p47*, *DNApol*, *vlf1*, *helicase*, *p35*, *ie1/0*, and *ie2*. Many of the *lef* genes are required for DNA replication, a prerequisite for late gene expression (see below). Typical examples of late genes are those encoding the structural proteins of the nucleocapsids and the envelope proteins of BVs and ODVs. Both *gp64* and *me53* have an early and a late gene promoter, allowing expression in both phases.

At very late times post infection, transcription of the two hyper-expressed late genes (*polh* and *p10*) is upregulated, resulting in accumulation of the polyhedrin and P10 proteins to extremely high levels. The burst of very late gene transcription is mediated by the late viral factor 1 (VLF1), which binds to an A/T-rich region, called a burst sequence, located downstream of the DTAAG motif in the *polh* and *p10* promoter regions. The hyper-expression of these two genes during the very late phase forms the basis of the baculovirus expression system. This recombinant protein production platform makes use of the promoters of either or both of these genes to efficiently express heterologous genes of interest in cultured insect cells.

Viral DNA Replication

Baculovirus DNA replication occurs in the nuclei of infected cells and is first detected 6–8 h post-infection of cultured *Spodoptera frugiperda* cells with AcMNPV and reaches top levels at 16–18 h post-infection. Several viral proteins have been shown to be crucial for AcMNPV DNA replication. These include DNA polymerase, the helicase P143 and several LEF proteins (LEF1; LEF2; LEF3,

LEF11). LEF1 and LEF2 mutually interact and possess primase activity. LEF3 is a single-stranded DNA binding protein that forms a complex with P143. Chemical cross-linking experiments showed that the viral trans-activator IE1 interacts with this complex. As shown in transient assays and by using mutant viruses, IE0 or IE1 can each support viral DNA replication, but both proteins are required to achieve full levels of viral replication. IE1 and presumably IE0 bind *hr* sequences in the baculovirus genome and therefore may function as origin-activating proteins. ME53 is a protein with two zinc-finger motifs, but its exact role in virus production is not clear. The exact role of LEF11 also remains enigmatic.

LEF7 is an F-box protein that is required for optimal levels of DNA replication in AcMNPV. Despite its important role in DNA replication, the *lef7* gene is conserved in the genomes of group I and a limited number of group II alpha- and betabaculoviruses. A number of other viral proteins have been shown to augment viral DNA replication in AcMNPV without being absolutely essential. These include the anti-apoptotic protein P35, the Inhibitor of apoptosis protein 1 (IAP1) and the group I alphabaculovirus-specific proteins IE2 and PE38. All three of IAP1, IE2 and PE38 are RING domain proteins. Both IE2 and PE38 are transcriptional activators (see above) and have ubiquitin ligase activities. However, it is not known whether this enzymatic function plays a role in DNA replication.

The mechanism by which the baculoviral genome replicates is not exactly known, but high molecular mass viral DNA molecules that are suggestive of genomic concatemers have been reported in infected cells. It has therefore been suggested that baculovirus DNA may replicate following a rolling circle model. More recently, it has been proposed that baculoviruses may use a recombination-dependent mechanism of viral DNA replication. In reality, it may even be a combination of the two, where recombination is used to resolve the concatemeric products of the rolling circle process. A protein that plays a role in producing genome length DNA molecules is Alkaline nuclease (AN). Alkaline nuclease family members repair dsDNA breaks and are involved in homologous recombination. The baculovirus AN may, therefore, be required in DNA processing, before the progeny genomes are packaged into the nucleocapsids. The ss-DNA binding protein LEF3 associates with AN, and this complex may play a crucial role in resolving replicative intermediates. Another protein that might be involved in this process is DBP (DNA binding protein), a protein that protects DNA against nuclease activity.

Virion Morphogenesis and Protein Composition

Nucleocapsids

The baculovirus life cycle is biphasic and produces two virion phenotypes, BV and ODV. NCs of both virion phenotypes are structurally identical. The NCs are rod-shaped, and measure approximately 30–60 nm by 250–300 nm. The NCs have characteristic cap and base structures (see Fig. 1) and several viral proteins are particularly associated with either of these two structures. NCs assemble in the ring zone of the nuclear virogenic stroma. In this process the viral dsDNA genome is densely packed within the NC in a spiral fashion. The small highly basic protein P6.9 is directly associated with the viral DNA and contains ca. 40% of arginine and 30% of serine/threonine residues. The highly basic nature of this protein functions in neutralizing the positive charges of the nucleic acid and aids in the condensation and packaging of the viral DNA. The condensed viral DNA is coated with the major capsid protein VP39. Recently, it was found that the 38K protein is crucial for the dephosphorylation of p6.9 during NC assembly. In contrast, uncoating of the viral DNA upon entry into the nucleus involves phosphorylation of p6.9.

Other AcMNPV proteins associated with the NCs include: P24, VP80, VP1054, BV/ODV-C42, P95, ORF1629, and EXON0 (AC141), AC142 and VLF1, and for several of these proteins more information about their function is available. AC142 is an essential gene required for NC formation. ORF1629 encodes the essential phosphoprotein P78/P83 that is associated with the basal end of the NC and needed for NC transport towards the nucleus. Deletion of BV/ODV-C42, VP80, VP1054, p6.9 and P95 showed that these genes are also essential for NC/BV assembly. In addition, P95 also has a domain required for oral infectivity of ODVs. VP80 and EXON0 are required for efficient egress of NCs (see also next section). Immunogold labelling has shown that VP80 binds to one end of the NCs (although it is not yet clear which end). VP80 also has DNA binding affinity, which may explain why VP80 deletion mutants do not assemble properly. Interestingly, the very late gene transcription factor VLF1 is also required for proper NC assembly and like VP80, typically associated with one end of the NCs. Q-PCR analysis showed that deletion of either the *vp80* or *vlf1* gene did not lower the levels of viral DNA produced. VLF1 is also not required for the production of genome-size DNA molecules, as previously suggested. Another NC-associated protein, VP1054, shares characteristics with the cellular DNA binding protein PUR α , and like this cellular protein, VP1054 has affinity for GGN-repeat motifs in DNA and RNA. GGN-repeat islands appear to be conserved in baculovirus genomes. A long (GGN)_n stretch is for instance present in AcMNPV on the complementary strand of *orf1629*, in which a CCN-repeat corresponds to a proline-rich domain in the essential protein P73/P87. As such, VP1054 might play a role in DNA packaging. Testing this hypothesis is rather difficult though, as artificial mutation of the GGN repeats will directly interfere with the function of P78/83 encoded on the complementary strand.

BV Morphogenesis

After their assembly in the virogenic stroma, NCs move along virus-induced, nuclear actin filaments to the nuclear periphery. The interaction with these actin filaments requires the NC protein VP80 that contains paramyosin-like domains. The NCs are believed to acquire an initial envelope when they egress from the nucleus to the cytoplasm to become BVs. By an unknown mechanism, the

initial envelope is lost and NCs are further transported to the cell surface, most likely utilizing microtubules. The AcMNPV protein EXON0 (AC141) seems to play a crucial role here, as it is found to co-purify with beta-tubulin and to bind to kinesin-1. Deletion of the *exon0* gene strongly reduced BV formation. At the cell surface, the NCs bud from the modified plasma membrane and as a consequence acquire an envelope carrying viral-encoded glycoproteins. Typical for the BV envelope of group I alphabaculoviruses is the presence of the major envelope glycoprotein GP64, which is a class III envelope fusion protein. GP64 is not present in ODVs and is therefore a BV-specific protein. Group II NPVs, GVs and deltabaculoviruses carry a functional equivalent, the F protein, which is a class I envelope fusion protein. Gammabaculoviruses do not have a BV fusion protein and as a consequence do not produce BVs, and, hence, are gut restricted.

GP64 glycoprotein characteristics

GP64 proteins are highly conserved, with approximately 80% amino acid sequence identity among all GP64 protein ectodomains examined. As indicated above, GP64 is structurally a class III glycoprotein that is phosphorylated, palmitoylated, and heavily glycosylated. It is found on the cell surface of infected cells and on the virion as a disulphide-linked trimer of GP64 monomers. GP64 functions in both viral entry and exit. In viral entry, GP64 is important for host cell binding and membrane fusion. Little is known of the details of GP64 interactions with host cell receptor(s) and the molecules that can serve as the host cell receptor for BV binding also remain unknown. However, GP64 binding appears to be highly promiscuous and this feature has been exploited for instance in the use of AcMNPV as a transduction vector for mammalian cells and in the use of GP64 for pseudo-typing retrovirus-based gene therapy vectors. The GP64 protein has also been used for peptide display on BV particles *e.g.*, for immunization purposes or drug screening.

After host cell binding mediated by GP64, the BVs are internalized via endocytosis and the low pH of the endosome triggers a conformational change in the GP64 trimers, resulting in activation of membrane fusion. Studies of GP64-mediated membrane fusion have indicated that large short-lived complexes of approximately 10 or more GP64 trimers form immediately prior to membrane fusion and are likely the unit structure of the membrane fusion mechanism. In addition, the opening of the fusion pore occurs rapidly after low pH triggering. Unlike F and many other membrane fusion proteins, the GP64 protein does not require a prior internal cleavage for maturation or activation of its fusogenic capacity.

Studies of a *gp64* gene knockout virus showed that GP64 is also necessary for efficient budding of progeny BVs. Because GP64 is found on the surface of infected cells, concentrated in discrete areas, it is thought that these concentrations of GP64 represent the sites of virion budding. ME53 co-localizes with the envelope glycoprotein GP64 in the same plasma membrane foci and does so in a GP64-dependent manner. ME-53 is encoded by all alphabaculovirus genomes and BV production is strongly attenuated in mutants lacking *me53*. Interestingly, ME53 is also found in association with isolated nucleocapsids and as such it may bridge the NCs with the GP64-containing envelope (possibly via the cytoplasmic tails of GP64) and by doing so assist in budding.

F protein properties

F proteins have been studied most extensively in SeMNPV (SE8), HearNPV (HA133), and LdMNPV (LD133). F proteins presumably serve a similar role in host receptor binding as GP64 although they appear to recognize a different host cell receptor. Both F and GP64 are glycosylated low-pH-activated membrane glycoproteins present as trimeric spikes in the BV envelope. F is a class I envelope fusion protein that becomes fusogenic after an internal cleavage by the cellular protease furin. This cleavage leads to the F1 and F2 components, which are interconnected by disulphide bonds. Unlike GP64, F proteins are not specific to BVs. Proteomic analyses of AcMNPV, HearNPV, ChchNPV and the deltabaculovirus CuniNPV have also identified F proteins in the ODVs. However, more detailed studies will be necessary to understand the potential role of the F protein in ODVs.

All viruses that encode a GP64 protein (group I alphabaculoviruses) also encode a homolog of the F protein (AC23 in AcMNPV, OP21 in OpMNPV). This F-homolog can be deleted with no substantial effect in cell culture infections as long as the *gp64* gene is intact. However, the conservation of *f* gene homologs in the genomes of the group I alphabaculoviruses suggests an important function and it was recently shown that deletion of the *ac23* from the AcMNPV genome results in delayed mortality of infected larvae. Thus, while F proteins found in group II alphabaculoviruses are essential entry proteins, the F homologs in group I are not essential, but appear to serve an accessory role that is important in the pathogenicity and/or virulence of the virus. Evolutionarily, the F protein is most likely the original baculovirus BV fusion protein as homologs of this protein are also encoded by insect genomes. An ancestor of the group I alphabaculoviruses has probably obtained a *gp64* gene by horizontal gene transfer, maybe from an insect retrovirus or via a retrotransposon, and the encoded GP64 protein has taken over the envelope fusion function in this clade of viruses.

Recent cryo-EM analysis of BV particles has shown that BVs are rather ovoid than rod-shaped (as obtained when using chemical fixation and as shown in previous drawings). Cryo-EM has also shown that both GP64 and F are found concentrated on both ends of the virion, although spikes are sometimes also found on the lateral sites of the ovoid (**Fig. 5**).

ODV Assembly and ODV Envelopes

Nucleocapsids destined to be incorporated into ODVs are retained in the nucleus and are enveloped by a lipid bilayer that is derived from the inner nuclear membrane. AcMNPV ORF142 and ODV-EC27 are essential for successful ODV assembly and for the production of infectious virus. ODV envelopes have been shown to contain a number of proteins that are not present in BVs (see **Fig. 1**). The most abundant AcMNPV ODV-envelope specific proteins include ODV-E18, ODV-E25, PIF5 (ODV-E56/ ODVP-6E), ODV-E66, and PIF0 (P74). In more recent years, additional ODV envelope proteins have been identified that are less

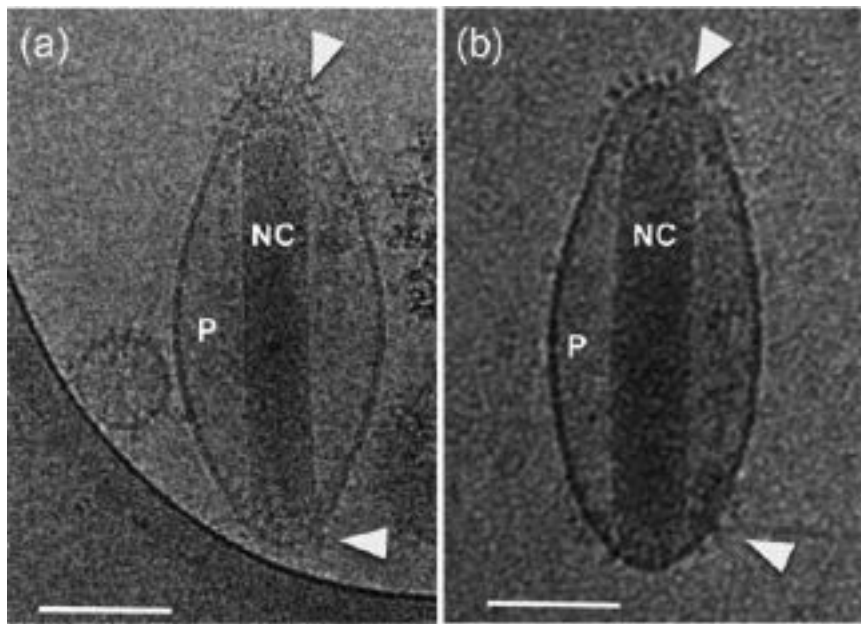


Fig. 5 Cryo-EM images of budded virus particles of group I and group II alphabaculoviruses. The ovoid BVs of AcMNPV (a) and SeMNPV (b) have an internal nucleocapsid (NC) surrounded by an envelope derived from the plasma membrane from which spikes (arrows) protrude at both ends. The spikes in AcMNPV are trimers of the envelope protein GP64, while SeMNPV has trimeric F protein spikes. Spikes might also be found along the lateral sides of the BVs. A pocket (P) of unknown composition is present between the lateral sites of the nucleocapsid and the envelope in these cryo-EM images. Scale bars represent 100 nm. The images are a courtesy of J.W.M. van Lent and have been published before as part of a larger figure in Wang, Q., Bosch, B.J., Vlak, J.M., *et al.*, 2016. Budded baculovirus particle structure revisited. *Journal of Invertebrate Pathology* 134, 15–22 and used with permission of Elsevier, Amsterdam, the Netherlands.

abundantly present but are absolutely required for oral infectivity. Therefore, they have been termed *per os* infectivity factors (PIFs). The previously discovered P74 and ODVE-56, and VP95 were subsequently renamed to PIF0, PIF5 and PIF8, respectively. With the recent discovery of PIF9 (AC108), ten PIFs have been allocated in total. Viruses with *pif* gene mutations are infectious by injection of their BVs into caterpillars, but do not show mortality upon oral infection with ODVs. The only exception is PIF8, which contains a DNA domain (NAE domain) which is required for BV formation. PIF8 is also associated with NCs, but the other PIFs are ODV-envelope specific. Most of the studies on PIF genes were done with AcMNPV or HearNPV.

PIF proteins, with the exception of PIF5, associate into a macromolecular protein complex, called the ODV entry complex. The macromolecular complex has a stable core consisting of PIF1–4. The other PIFs in the entry complex are associated with a relatively lower affinity to the core. One of the biological functions of the entry complex, apart from mediating midgut infection, is the protection of the individual PIF proteins against proteolytic degradation under alkaline circumstances. PIF proteins are highly specific and cannot easily be exchanged between viruses, possibly due to these complex protein-protein interactions in the ODV entry complex.

Tegument Proteins

GP41 is an ODV-specific glycoprotein encoded by a core gene. GP41 does not co-purify with either the NC or the envelope fraction and is therefore considered as a tegument protein, located in between the NC and envelope. A temperature-sensitive mutant of AcMNPV with a codon replacement in the *gp41* gene, showed that GP41 plays a critical role in virus assembly. When grown at the non-permissive temperature, the *gp41* ts mutants failed to produce ODVs and OBs, and, in addition, NCs did not egress from the nucleus to form BVs. GP41 therefore appears to play a key role in the assembly of both virion phenotypes, even though it has been detected only in ODVs. Tegument proteins appear to be acquired within the nucleus, when the ODV NCs become enveloped. Since BVs do not contain GP41, it was suggested that proteins that loosely associate with the NCs in the nucleus, might be lost when these migrate from the nucleus to the cytoplasm to become BVs.

Appendix A Supplementary Material

Supplementary data associated with this article can be found in the online version at [doi:10.1016/B978-0-12-809633-8.21550-1](https://doi.org/10.1016/B978-0-12-809633-8.21550-1).

See also: Baculoviruses: General Features (*Baculoviridae*)

Further Reading

- Blissard, G.W., Theilmann, D.A., 2018. Baculovirus entry and egress from insect cells. *Annual Review of Virology* 5, 113–139.
- Boogaard, B., van Oers, M.M., van Lent, J.W.M., 2018. An advanced view on baculovirus *per os* infectivity factors. *Insects* 9, 84.
- Chen, Y.R., Zhongm, S., Feim, Z., *et al.*, 2013. The transcriptome of the baculovirus *Autographa californica* multiple nucleopolyhedrovirus in *Trichoplusia ni* cells. *Journal of Virology* 87, 6391–6405.
- Han, Y., van Oers, M.M., van Houte, S., Ros, V.I.D., 2015. Virus-induced behavioural changes in insects. In: Mehlhorn, H. (Ed.), *Host Manipulations by Parasites and Viruses*. Parasitology Research Monographs 7. Cham: Springer, pp. 149–174.
- Harrison, R.L., Herniou, E.A., Jehle, J.A., *et al.*, 2018. ICTV virus taxonomy profile: *Baculoviridae*. *Journal of General Virology* 99, 1185–1186.
- Huang, Z., Pan, M., Zhu, S., *et al.*, 2017. The *Autographa californica* multiple nucleopolyhedrovirus *ac83* gene contains a *cis*-acting element that is essential for nucleocapsid assembly. *Journal of Virology* 91. (e02110-16).
- Lai, Q., Wu, W., Li, A., *et al.*, 2018. The 38k-mediated specific dephosphorylation of the viral core protein P6.9 plays an important role in the nucleocapsid assembly of *Autographa californica* multiple nucleopolyhedrovirus. *Journal of Virology* 92. (e01989-17).
- Marek, M., Merten, O.W., Galibert, L., Vlak, J.M., van Oers, M.M., 2011. Baculovirus VP80 protein and the F-actin cytoskeleton interact and connect the viral replication factory with the nuclear periphery. *Journal of Virology* 85, 5350–5362.
- Ohkawa, T., Volkman, L.E., Welch, M.D., 2010. Actin-based motility drives baculovirus transit to the nucleus and cell surface. *Journal of Cell Biology* 190, 187–195.
- Rohrmann, G.F., 2010. Chapter 12 – The AcMNPV genome: Gene content, conservation, and function. In: *Baculovirus Molecular Biology*, third ed. Bethesda, MD: National Center for Biotechnology Information. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK138304/>.
- Stewart, T.M., Huijskens, I., Willis, L.G., Theilmann, D.A., 2005. The *Autographa californica* multiple nucleopolyhedrovirus *ie0–ie1* gene complex is essential for wild-type virus replication, but either *IE0* or *IE1* can support virus growth. *Journal of Virology* 79, 4619–4629.
- van Oers, M.M., Pijlman, G.P., Vlak, J.M., 2015. Thirty years of baculovirus-insect cell protein expression: From dark horse to mainstream technology. *Journal of General Virology* 96, 6–23.
- Vanarsdall, A.L., Okano, K., Rohrmann, G.F., 2006. Characterization of the role of very late expression factor 1 in baculovirus capsid structure and DNA processing. *Journal of Virology* 80, 1724–1733.
- Wang, Q., Bosch, B.J., Vlak, J.M., *et al.*, 2016. Budded baculovirus particle structure revisited. *Journal of Invertebrate Pathology* 134, 15–22.
- Wang, X., Shang, Y., Chen, C., *et al.*, 2019. Baculovirus *per os* infectivity factor complex: Components and assembly. *Journal of Virology* 93. (e02053-18).
- Williams, T., Bergoin, M., van Oers, M.M., 2017. Diversity of large DNA viruses of invertebrates. *Journal of Invertebrate Pathology* 149, 4–22.

Bidensoviruses (*Bidnaviridae*)

Qin Yao, Zhaoyang Hu, and Keping Chen, Jiangsu University, Zhenjiang, China

© 2021 Elsevier Ltd. All rights reserved.

Classification

Bombyx mori bidensovirus (BmBDV) is one of the causative agents that causes fatal densovirus in the silkworm. The single-stranded DNA genome is similar to that in parvovirus densoviruses. BmBDV was previously known as *Bombyx mori* densovirus type 2 (BmDENV-2) and was previously assigned to the *Densovirinae* in the *Parvoviridae* family. However, unlike parvoviruses the BmBDV genome consists of two ssDNA segments in separate capsids and encodes DNA polymerase B. Considering these unusual properties, in the Eighth Report of the International Committee on Taxonomy of Viruses (ICTV), this virus was excluded from the family *Parvoviridae*. In 2012, BmBDV was assigned to the new, sole and type species *Bombyx mori bidensovirus* in the new genus *Bidensovirus* within the new family *Bidnaviridae*. BmBDV is currently the only virus that possesses a single stranded DNA that encodes for a DNA polymerase. Characterizing its basic biology is important for understanding this new virus family and we can expect that other bidnaviruses will be isolated and characterized in the future.

Genome Organization and Expression Strategy

BmBDV has two linear single-stranded DNA segments about 6.5 kb (viral DNA1, VD1) and 6 kb (viral DNA2, VD2) in size. The complementary strands of the genomic DNA, but probably also VD1 and VD2, are encapsidated separately. The genome compartmentalization of such a bipartite virus gives rise to an obvious cost, because the different viral particle types containing at least one copy of each segment have to co-infect the same host cell during primary and systemic infection. If the different particle types are present at the same frequency, this cost would be minimized, while deviations from a one to one ratio would increase the cost of genome multipartition. However, genome segments accumulate at different frequencies in BmBDV virions and in BmBDV-infected midguts. The copy number of short-segment VD2 is higher than that of long-segment VD1. Faba bean necrotic stunt virus (FBNSV, family *Nanoviridae*) has an even more complex genome composition of eight circular ssDNA molecules encapsidated in individual icosahedral particles. The various segments of FBNSV present very different frequencies within individual host plants and virus particles as well as aphid vectors.

Several BmBDV isolates were obtained from silkworm and genomes of three of them, Yamanashi isolate (Japan), Zhenjiang isolate (China), and Indian isolate, have been sequenced. Sequence analysis of cloned VD1 and VD2 demonstrates that they are different DNA molecules that share a common 3' terminal sequence of 53 nucleotides. Both DNAs contain complementary terminal repeats in which the characteristic terminal palindrome observed in parvoviruses are not found (Fig. 1). The genomic structure of VD1 is similar to the self-synthesizing DNA transposons (hypothetical viruses termed polintovirus) characterized by terminal-inverted repeats and a unique set of proteins such as protein-primed DNA polymerase B (PolB) and ATPase. VD1 encodes four genes, *ns2*, *ns1*, and *vp* located on the 5'-half of one strand and *polB* on the 5'-half of the complementary strand. *ns2* and *ns1* consist of 381 and 951 nts respectively, and there is a 218 nts overlap between them. *vp* and *polB* consist of 1500 nts and 3318 nts respectively. There is an 8 nts overlap between *ns1* and *vp*. VD2 contains two genes which are located on the 5'-half of the complementary strands respectively, the *ns3* gene consists of 669 nts and the *p133* gene consists of 3483 nts. The conserved initiator elements TATA box and CAGT box, and the polyadenylation signal were found in the upstream and downstream regions of these genes respectively. Between the Zhenjiang isolate and Yamanashi isolate of BmBDV, VD1 shares 98.4% identity and VD2 shares 97.7% identity.

Transcript mapping shows that three mRNA molecules with sizes of 1.1, 1.5, and 3.3 kb respectively are expressed during transcription of VD1 of BmBDV. The 1.1 kb transcript corresponds to *ns2* and *ns1*. Further studies suggest that *ns2* and *ns1* are controlled by overlapping promoters, P5 and P5.5, respectively. Transcriptional start sites of *ns2* and *ns1* are located 21 nts upstream and 3 nts downstream of the *ns2* initiation codon, respectively. Transcription of *ns2* and *ns1* terminate at the same site. The 1.5 kb transcript derived from *vp*, starts at 57 nts upstream of the initiation codon and terminates 22 nts downstream of the terminal codon. The major structural protein VPs are synthesized by a leaky scanning mechanism from this mRNA. The 3.3 kb transcript corresponding to *polB*, transcribes from 22 nts upstream of the start codon and stops 13 nts after the terminal codon. Sequence analysis showed conserved initiator elements, a TATA box and a CAGT box, which were observed in the upstream regions of the viral genes. The activity of promoter P5 was investigated using luciferase reporter assay and the core elements of P5 were determined by a series of truncation mutants. The results suggest that P5 can drive reporter expression *in vitro*, and the TATA box, CAGT box and the upstream nt-205 to nt-236 region are important for the promoter activity. In VD2, the transcription of *ns3* initiates 29 nts upstream of the first ATG and terminates at the 59 nts downstream of the stop codon. The transcriptional initiation site and terminal site of *p133* are located at nt 5367 and nt 1771 of VD2, respectively. Alternative splicing of mRNA, which is common in parvoviruses, is not observed in BmBDV.

Pathology of Silkworm Associated With BmBDV

BmBDV replicates predominantly in the nuclei of columnar cells of the larval midgut epithelium, and leads to fatal densovirus in the silkworm. Symptoms of BmBDV-infected silkworm larvae include anorexia, diarrhea, and flaccidity. In diseased larvae, the

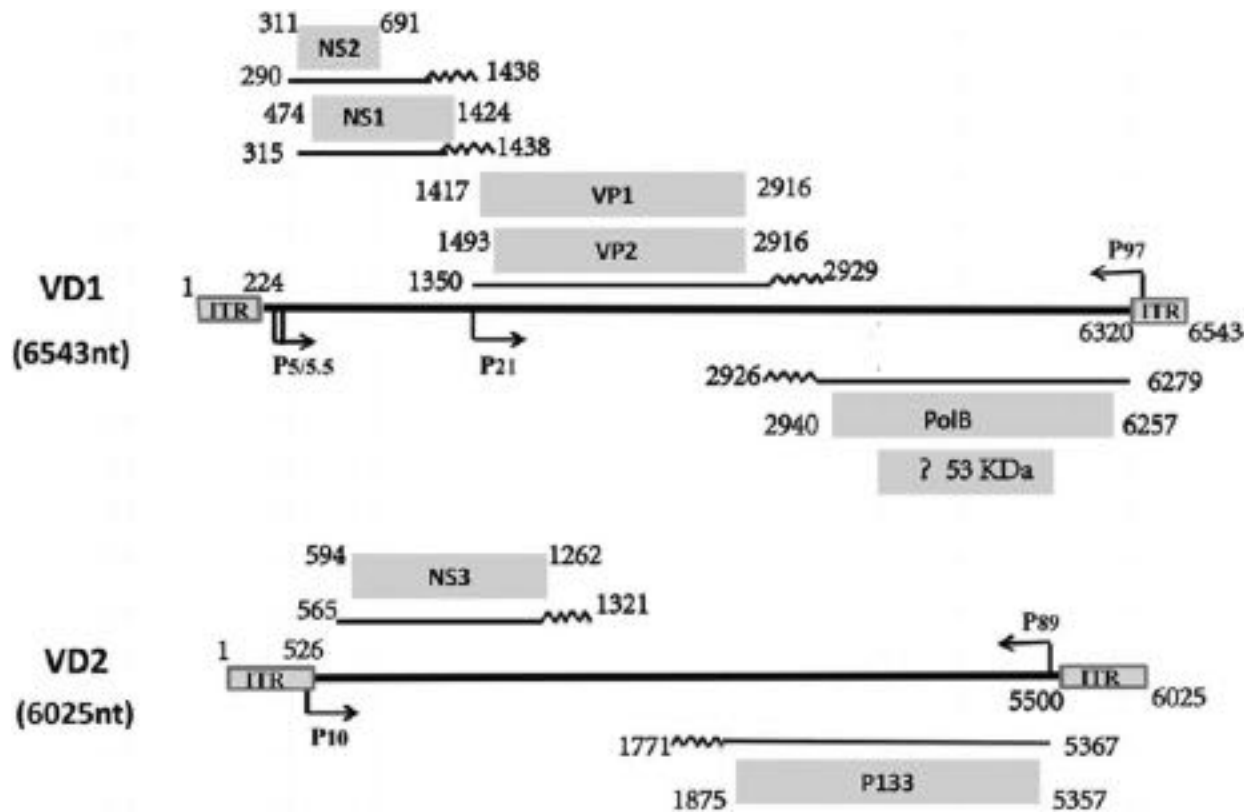


Fig. 1 Transcription map of BmBDV Zhenjiang isolate. The viral genome segments (VD1 and VD2) are depicted, flanked by terminal repeats (ITR), including the viral promoters (P5/5.5, P21, P97, P10 and P89). The viral mRNAs, transcription initiation and termination sites, and the proteins that they encode are shown. Polyadenylation is shown by a squiggly line. The open reading frames (ORFs) used to encode the viral proteins are shown.

color of the postero-midgut changed to pale yellow. Silkworms infected by BmBDV before the first day of the 4th instar discharged bead-like frass. Virus infection caused slow growth and extension of lengths of instar times. Different development stages of the silkworm larvae have different sensitivities to infection by BmBDV. The LT_{50} of first stage larvae is about 21 days, indicating that BmBDV has a long pathogenic process. BmBDV-infected 5th instar larvae could develop to the adult stage and produce the next generation, but the weight and survival rate of the pupa and the cocoon production decreases slightly. The main lesions occur in the nuclei of infected cells. Because of the accumulation of viral DNA, the nuclei of virus-infected cells become hypertrophied and stain intensely with methyl green or Feulgen reagent or exhibited an orange yellow or red fluorescence after staining with acridine orange at acidic pH. The hypertrophied nuclei of infected cells is 2.5 times larger than that of the normal cell and a voluminous dense homogeneous structure appeared in each of the infected nuclei at the late stage of infection. Infected columnar cells eventually degenerate and are discharged into the gut lumen, after which the mature virions in the degenerated infected cells were excreted along with the frass. However, cells of the midgut epithelium infected by BmBDV are not as readily discharged as those from larvae infected by *Bombyx mori* densovirus (BmDENV). This was consistent with the fact that BmBDV infection was chronic, while BmDENV caused acute disease. Two non-viral 45.5 kDa and 44 kDa proteins accumulate in BmBDV-infected degenerated cells in the midgut epithelium of the silkworm, and this accumulation is inversely correlated with that of viral structural proteins. TUNEL staining showed the accumulation of these proteins in degenerated cells infected by BmBDV was not due to apoptosis. Pathological changes are also observed in the goblet cells of infected midgut epithelium at the late stages of infection. Although the nuclei of goblet cells were infected and degenerated, the extent of damage was far less than that of columnar cells.

The ultrastructural changes in cells infected by BmBDV were observed both in the cytoplasm and the nucleus. In the cytoplasm, vesicles or cisternae near the granular endoplasmic reticulum, in which small spherical particles accumulated, could be observed, indicating the accumulation and transport of viral protein to the nucleus. During the replication of the virus, cellular organelles such as mitochondria and endoplasmic reticulum become vacuolized and degenerated. Meanwhile, a great many lysosomes and a large autophagosome containing degenerated organelles emerged. In the nucleus, the heterochromatin becomes very condensed and discrete. Hypertrophy of the virus-infected nucleus follows infection and a virogenic stroma appears. The virogenic stroma is less electron-dense than the surrounding nuclear matrix and occupies most of the nucleus. Simultaneously, the fibrillar and granular components of the nucleolus are segregated, and the hypertrophied nucleolus is pushed to the nuclear periphery. Electron-dense discrete sites of replication within the virogenic stroma appear early in infection. Following the multiplication of

the virus, these discrete sites in which virions were assembled increase in size. At the end of infection, the greatly enlarged nuclei are filled with thousands of small spherical virions, 20–24 nm in diameter. The virions enter the cytoplasm through the enlarged nucleopores or by destruction of the nuclear envelope. The complete virions are then liberated into the gut lumen along with the degenerated cells.

Viral Non-Structural Proteins

The BmBDV virus non-structural (NS) proteins are likely essential for viral replication as they are in parvoviruses. Bidensoviruses encode the DNA polymerase PolB which clusters with the DNA pol of polintoviruses. NS1 of BmBDV shows homology to that of densovirus and could be involved in the replication of the virus. A superfamily 3 helicase motif, which is conserved in NS1 of parvoviruses, was identified in the BmBDV NS1. However, another conserved initiator (replicator) motif in parvoviruses was lacking in BmBDV NS1. In parvoviruses, NS1 possesses multiple functions such as sequence specific DNA-binding, as well as helicase, ATP-dependent site-specific endonuclease and ATPase activities, involved in viral DNA replication. The NS1 of BmBDV also exhibits DNA binding, ATPase as well as helicase activities and NS1 interacted with the PolB. These facts implied that BmBDV NS1 is also a multifunctional protein, which possibly plays an important role in virus replication. BmBDV NS1 is a phosphorylated protein and the phosphorylation regulates its ATPase activity.

The *ns3* gene encodes a 27 kDa protein named NS3. Bioinformatics analysis suggests that NS3 possesses two zinc-finger motifs, 6 putative N glycosylation sites, and 4 putative phosphorylation sites. Homologs of BmBDV NS3 have also been identified in some densoviruses and baculoviruses. The homolog of NS3 in *Junonia coenia* densovirus is absolutely required for viral DNA replication, implying that BmBDV NS3 might play a similar and important role in the BmBDV viral life cycle. The NS3-like proteins from diverse viruses share a conserved Zn-finger motif. Inhibitors of apoptosis (IAP proteins) in baculoviruses also have Zn-finger motifs. It was hypothesized that NS3-like proteins might represent a novel family of apoptosis inhibitors in arthropod-infecting viruses.

NS2 was identified from BmBDV-infected silkworm midguts via comparative proteomics but this protein was not detected in virions, indicating this protein is a non-structural protein. NS2 of BmBDV shares no homology with other NS2 from parvoviruses and little is known about it. NS2 shares slight sequence similarity with some chromosomal replication initiator protein *dnaA* and DNA-binding response regulator. In virus-infected larvae, the transcription of the *ns2* gene was first detected 28 h post inoculation in low amounts, but in higher amounts at late stages of infection. Immunofluorescence showed NS2 ultimately concentrated on the nuclear membrane in BmN cells at late stages. Like the NS2, adenovirus death protein (ADP) is an Asn-glycosylated integral membrane protein which ultimately localizes to the nuclear membrane, and it is expressed early but is greatly amplified at late stages of infection. As ADP is required for efficient cell lysis and virus release, BmBDV NS2 might have the same function in facilitating the cell's lysis and virus release. As yet, the function of NS2 is still obscure.

Viral Structural Protein

Similar to parvoviruses, bidensoviruses are non-enveloped small spherical viruses with a diameter of 20 – 24 nm with equiaxial symmetry. However, the detailed ultrastructure of this virus remains to be elucidated. Viral structural proteins perform a wide variety of structural and biological functions during the virus life cycle, including host cell surface receptor recognition, pathogenicity determination, viral genomic encapsidation and host immune response detection and evasion.

BmBDV contains two genes *vp* and *p133* encoding the major capsid proteins (VPs) and minor capsid protein, respectively. Two VPs are produced by leaky scanning from the first and second ATG of *vp*. The VPs contain an N-terminal glycine-rich region as well as jelly-roll domains similar to parvovirus VPs. The conserved phospho-lipase A2 domain, which is essential for virus entry in parvoviruses, is not found in BmBDV. The BmBDV VPs expressed in a baculovirus expression system can auto-assemble into virus-like particles. *p133* encodes a 133 kDa protein named P133. Western blot analysis showed that the anti-P133 N-terminus and anti-P133 C-terminus serums recognized the same specific 133 kDa protein in virions, but no other proteins could be detected, suggesting that the *p133* gene encodes a single structural protein. Amino acid sequence alignment illustrated that the C-terminus is conserved between P133 and some viral structural proteins in family *Reoviridae* such as VP3 of cypovirus. In reoviruses, P133 homologs are located at the outer capsid shell and are involved in host recognition and virus entry. It appears most likely that P133 has a similar function. A 53 kDa protein was detected in purified virions using a specific antiserum against PolB, and three peptides with molecular masses of around 120, 70, and 53 kDa were also detected in virus-infected silkworm midgut, demonstrating that *polB* encodes a structural protein processed by post-translational cleavage.

Viral Replication

Unlike other ssDNA viruses, including parvoviruses, bidensoviruses encode the protein-primed DNA polymerase B rather than an endonuclease, indicating that bidensoviruses employ a different replication mechanism from that of parvoviruses. In fact, the replicative intermediates of viral DNAs similar to those of parvoviruses were not detected during virus infection, strongly suggesting that bidensovirus genomic DNAs does not replicate by a rolling circle mechanism, which is thought to be used by all

parvoviruses. However, details of the replication mechanisms of the bidensoviruses have not yet been defined. It is important to further characterize the viral specific DNA polymerase to understand the replication mechanism of bidensoviruses. Tijssen and Bergoin suggested that BmBDV DNA could replicate like the DNA of adenoviruses. In this model, the terminal protein (TP) which is covalently linked at the 5'-terminus of genome DNAs, serves as primers for viral DNA replication initiation. The terminal complementary repeats of bidensoviruses would form the panhandle structures, resulting in a double-stranded molecule that would then serve as a template for further replication of both minus and plus progeny strands. The 53 kDa peptide encoded by the *PolB* gene might be the TP of BmBDV, but the details of this peptide remain to be elucidated. Moreover, the terminus of BmBDV and polintovirus genomes start with the tetranucleotide GTGT/GTGT repeat, which is similar to the terminal GTAGTA sequence of adenovirus DNAs. The GTA at positions 4–6 of adenovirus DNA, rather than 1–3, are used as a template for pTP-CAT intermediate formation, and then, this intermediate jumps back to position 1 of the template to start elongation. This jumping-back mechanism ensures the integrity of the terminal sequences during replication of the linear genome.

Receptor-mediated attachment and entry are essential first steps in the virus life cycle. The interaction between the BmBDV attachment protein and its receptor is very specific. The silkworm *nsd-2* gene determines the resistance of silkworm to BmBDV. Its allele encodes a 12-pass transmembrane protein, +NSD2, which is homologous with amino acid transporter family and is expressed in predominantly the posterior part of the midgut. A deletion of 9 transmembrane domains of +NSD2 at the C-terminus results in resistance to BmBDV. +NSD-2 can interact with the VP of BmBDV *in vitro*. These facts suggest that +NSD-2 serves as a functional receptor for BmBDV and the target site recognized by the virus is present in the deleted portion of the membrane protein. However, the BmBDV genome can replicate at a low level and the virus VP can be detected before 48 h post-infection in resistant silkworm, indicating that +NSD-2 regulates virus replication in addition to its function as a cell receptor protein for BmBDV.

Virus Evolution

The linear single-stranded DNA genomic structure of bidensoviruses as well as the dimensions and morphology of the virions are like those of parvoviruses. Similar to the insect densovirus, BmBDV infection leads to a fatal densovirus. BmBDV NS1 is homologous to densovirus NS1, possessing a conservative superfamily 3 helicase domain. Although the major structural protein VP of BmBDV didn't share significant sequence similarity to that of parvoviruses, the BmBDV VP encompasses counterparts to all 8 β -strands that can form the core of the jelly-roll fold as parvoviruses, suggesting that VP was a homolog of the VP of parvoviruses. These features strongly suggest that bidnaviruses originated from a parvovirus ancestor. The main feature of bidensoviruses that differs from ssDNA viruses is that the VD1 of bidensoviruses encode the protein-primed DNA polymerase B. A phylogenetic tree indicates that the PolB of BmBDV was derived from polintovirus. In BmBDV VD2, *p133* encodes the 133 kDa minor structural protein (P133) that shares sequence similarity to structural proteins of some reoviruses. Phylogenetic analysis demonstrates that bidensoviruses cluster with members of the genus *Cypovirus*. The non-structural protein NS3 shows some homology with NS3 of some densovirus, and the homologs of NS3 were also found in some granuloviruses and nucleopolyhedroviruses. Phylogenetic analysis shows that BmBDV NS3 clusters with granuloviruses, suggesting that BmBDV most likely acquired the NS3 from a granulovirus rather than a densovirus. All together, it appears that bidnaviruses evolved from a parvovirus ancestor from which they inherited a jelly-roll capsid protein and a superfamily 3 helicase and that later, the genome of the ancestral parvovirus integrated into a polintovirus genome acquiring the polintovirus PolB gene along with terminal inverted repeats to generate the VD1 genome segment. It is possible that the VD2 segment of BmBDV evolved directly from the VD1 segment by replacement of the parvoviral and polintoviral genes with genes from reoviruses and baculoviruses.

Further Reading

- Bao, Y., Chen, L., Wu, W., *et al.*, 2013. Direct interactions between bidensovirus BmDNV-Z proteins and midgut proteins from the virus target *Bombyx mori*. *FEBS Journal* 28 (3), 939–949.
- Chen, H., Qin, Yao, Bao, F., Chen, K., Liu, X., 2012. Comparative Proteome analysis of silkworm in its susceptibility and resistance responses to *Bombyx mori* densovirus. *Intervirology* 55, 21–28.
- Guo, H., Sun, Q., Wang, B., *et al.*, 2019. Sp1 is downregulated by multiple viruses to elevate ERK signaling and ensure viral reproduction in silkworm. *Developmental and Comparative Immunology* 98, 1–5.
- Hu, Z., Deng, Y., Zhang, X., *et al.*, 2018. Selection and validation of reference genes for reverse transcription quantitative real-time PCR (RT-qPCR) in silkworm infected with *Bombyx mori* bidensovirus. *Biologia* 73 (9), 897–906.
- Ito, K., Fujii, T., Yokoyama, T., Kadono-Okuda, K., 2018. Decrease in the expression level of the gene encoding the putative *Bombyx mori* bidensovirus receptor during virus infection. *Archives of Virology* 163 (12), 3327–3338.
- Kapitonov, V.V., Jurka, J., 2006. Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 103, 4540–4545.
- Krupovic, M., 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Current Opinion in Virology* 3, 578–586.
- Kumar, D., Sun, Z., Cao, G., *et al.*, 2019. *Bombyx mori* bidensovirus infection alters the intestinal microflora of fifth instar silkworm (*Bombyx mori*) larvae. *Journal of Invertebrate Pathology* 163, 48–63.
- Li, G., Zhou, Q., Hu, Z., *et al.*, 2015. Determination of the proteins encoded by BmBDV VD1-ORF4 and their interacting proteins in BmBDV-infected midguts. *Current Microbiology* 70 (4), 623–629.

- Li, G., Zhou, Q., Qiu, L., *et al.*, 2017. Serine protease Bm-SP142 was differentially expressed in resistant and susceptible *Bombyx mori* strains, involving in the defence response to viral infection. *PLoS One* 12 (4), e0175518.
- Lü, P., Pan, Y., Yang, Y., *et al.*, 2018. Discovery of anti-viral molecules and their vital functions in *Bombyx mori*. *Journal of Invertebrate Pathology* 154, 12–18.
- Lü, P., He, Y., Lin, F., *et al.*, 2019. Rapid detection of *Bombyx mori* bidensovirus by loop-mediated isothermal amplification based lateral flow dipstick assay for field applications. *Journal of Invertebrate Pathology* 163, 75–81.
- Sun, Q., Jiang, L., Guo, H., *et al.*, 2018. Increased antiviral capacity of transgenic silkworm via knockdown of multiple genes on *Bombyx mori* bidensovirus. *Developmental and Comparative Immunology* 87, 188–192.
- Zhang, P., Miao, D., Zhang, Y., *et al.*, 2016. Cloning and rescue of the genome of *Bombyx mori* bidensovirus, and characterization of a recombinant virus. *Virology journal* 8 (13), 126.

Bunyaviruses of Arthropods (*Mypoviridae*, *Nairoviridae*, *Peribunyaviridae*, *Phasmaviridae*, *Phunuiviridae*, *Wupedeviridae*)

Sandra Junglen, Charité - University Medicine Berlin, Berlin, Germany

© 2021 Published by Elsevier Ltd.

Glossary

Ambisense Coding strategy used by some bunyaviruses in which a single stranded genome segment encodes proteins in both positive and negative sense orientations.

Arbovirus A virus transmitted to vertebrates by hematophagous (blood-feeding) insects.

Arthropod-specific virus A virus that infects only arthropods and cannot be transmitted to vertebrates.

Bunyavirus Any member of the order *Bunyavirales*.

Cap snatching Cleavage of the capped 5' end of a cellular mRNA to produce short oligonucleotides that are used to prime viral mRNA transcription. The cellular sequences are incorporated into the viral mRNAs.

Insect-specific virus A virus that infects only insects and cannot be transmitted to vertebrates.

Unclassified A virus not yet classified as a member of a taxon.

Introduction

At the time when Gouléako virus and Herbert virus as the first insect-specific bunyaviruses were discovered, bunyaviruses were organized as a family comprising the five genera *Phlebovirus*, *Orthobunyavirus*, *Hantavirus*, *Tospovirus* and *Nairovirus*. All bunyaviruses detected previously in blood-feeding arthropods were able to replicate in vertebrate cells and were considered to be arboviruses. The detection of bunyaviruses in mosquitoes that replicated efficiently in insect cells but were unable to infect vertebrate cells was a peculiar novelty. Moreover, these viruses did not group with any of the classical arthropod-associated bunyavirus genera (*Phlebovirus*, *Orthobunyavirus* and *Nairovirus*) and defined new deep rooting lineages in the bunyavirus phylogeny. Many other insect-specific bunyaviruses were discovered in subsequent years, giving rise to the reorganization of the family into an order comprising twelve families at present. [Table 1](#) gives an overview of selected classified arthropod-specific bunyaviruses. As most of these viruses were identified by metagenomic sequencing of pooled arthropod specimens and no virus isolates are available, the following paragraphs are limited to mainly those viruses for which the phenotypic and molecular characteristics have been determined.

Virion Structure

The morphology of virions has been analyzed for viruses of the genera *Herbevirus*, *Feravirus*, *Goukovirus*, *Phasivirus* and *Jonvirus*. Herbert virus (genus *Herbevirus*) and Badu virus (genus *Phasivirus*) particles are spherical, enveloped and 90–110 nm or 130 nm in diameter, respectively ([Fig. 1](#)). The enveloped virions of Gouléako virus (genus *Goukovirus*) and Ferak virus (genus *Feravirus*) are pleomorphic with a diameter of 70 – 120 nm. Jonchet virus (genus *Jonvirus*) has a unique morphology in the Order *Bunyavirales*. Jonchet virus has two types of enveloped virions with either a tubular morphology of about 60 nm up to 600 nm or a spherical morphology with a diameter of about 80 nm. The virions consist of four structural proteins, termed Gn, Gc, N, and L, in analogy to the bunyavirus terminology.

Genome and Coding Strategies of the Viral Genomes

The genomes consist of three single-stranded RNA segments of negative-sense or ambisense polarity. The segments are designated L (large), M (medium) and S (small) and have complementary terminal sequences that are conserved among viruses of the same genus and may also be conserved within the same family but differ between families. The terminal nucleotides have been identified for only the insect-specific bunyaviruses shown in [Table 2](#). The conserved terminal nucleotides of the 3'- and 5'-prime ends are complementary allowing the RNA segments to base-pair and form panhandle structures.

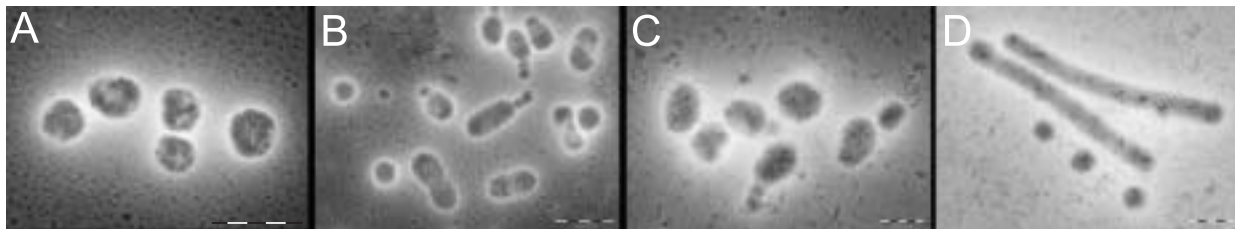
The complete genomes of all classified bunyaviruses have been sequenced. However, genome annotations are available for only those viruses that were isolated in cell culture and analyzed by molecular methods such as protein sequencing by mass spectrometry. The genome organization for those species is shown in [Fig. 2](#).

The L segment encodes the RNA-dependent RNA polymerase (RdRp) protein, also named L protein.

The M segment encodes the two glycoproteins which are translated as a precursor polypeptide (glycoprotein precursor, GPC) and posttranslationally cleaved into Gc and Gn. Viruses of the genera *Feravirus* and *Jonvirus* in the family *Phasmaviridae* encode an

Table 1 Selected classified arthropod-specific bunyaviruses in the order *Bunyvirales*

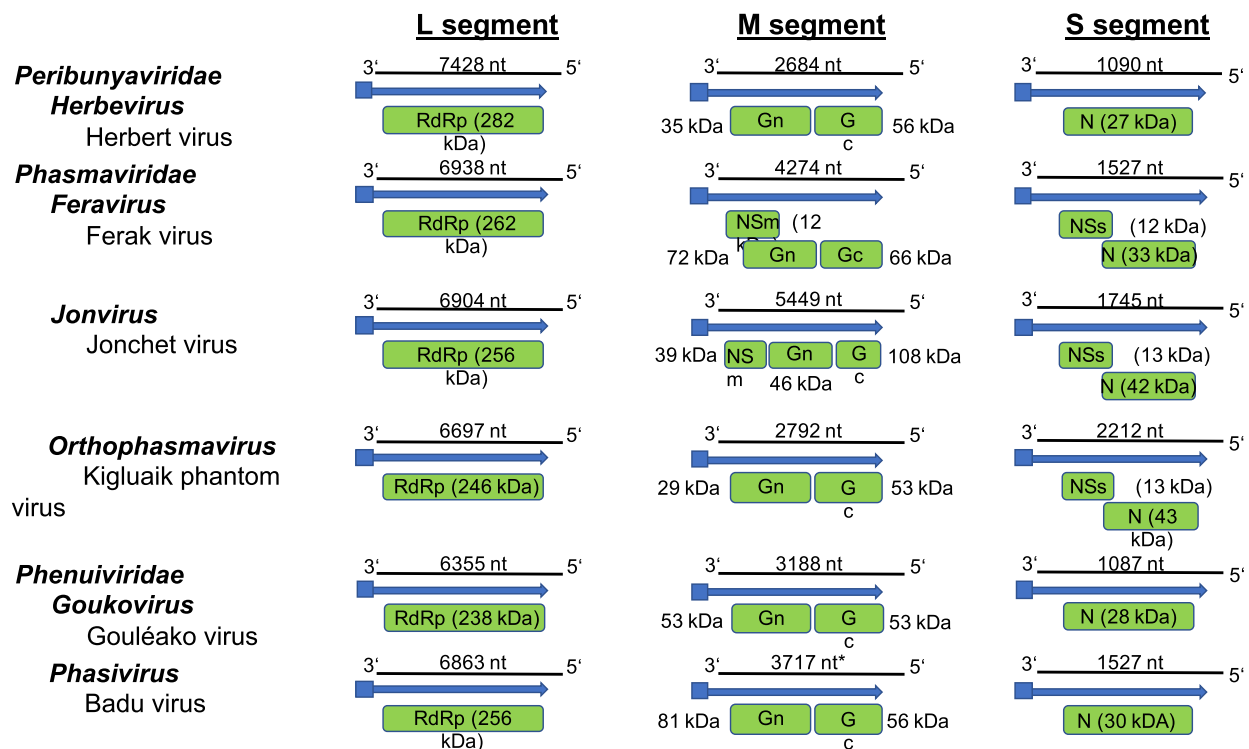
Family	Genus	Species	Host
<i>Mypoviridae</i>	<i>Hubavirus</i>	<i>Myriapod hubavirus</i> ^a	Millipede
<i>Nairoviridae</i>	<i>Shaspivirus</i>	<i>Spider shaspivirus</i> ^a	Spider
	<i>Striavirus</i>	<i>Strider striavirus</i> ^a	Water strider
<i>Peribunyaviridae</i>	<i>Herbevirus</i>	<i>Herbert herbevirus</i> ^a	Mosquito
		<i>Kibale herbevirus</i>	Mosquito
		<i>Tai herbevirus</i>	Mosquito
		<i>Insect shangavirus</i> ^a	n.d.
	<i>Shangavirus</i>		
<i>Phasmaviridae</i>	<i>Feravirus</i>	<i>Ferak feravirus</i> ^a	Mosquito
	<i>Jonvirus</i>	<i>Jonchet jonvirus</i> ^a	Mosquito
	<i>Orthophasmavirus</i>	<i>Anopheles orthophasmavirus</i>	Mosquito
		<i>Culex orthophasmavirus</i>	Mosquito
		<i>Ganda orthophasmavirus</i>	Bee
		<i>Kigluaik phantom orthophasmavirus</i> ^a	Phantom midge
		<i>Odonate orthophasmavirus</i>	Dragonfly
		<i>Qingling orthophasmavirus</i>	Dragonfly
		<i>Wuchang cockroach orthophasmavirus 1</i>	Cockroach
		<i>Wuhan mosquito orthophasmavirus 1</i>	Mosquito
		<i>Wuhan mosquito orthophasmavirus 2</i>	Mosquito
		<i>Sanxia sawastrivirus</i> ^a	Water strider
	<i>Sawastrivirus</i>		
	<i>Wuhivirus</i>	<i>Insect wuhivirus</i> ^a	n.d.
<i>Phenuiviridae</i>	<i>Beidivirus</i>	<i>Dipteran beidivirus</i> ^a	n.d.
	<i>Goukovirus</i>	<i>Cumuto goukovirus</i>	Mosquito
		<i>Gouleako goukovirus</i> ^a	Mosquito
		<i>Yichang insect goukovirus</i>	n.d.
	<i>Horwuvirus</i>	<i>Horsefly horwuvirus</i> ^a	Fly
	<i>Hudivirus</i>	<i>Dipteran hudivirus</i> ^a	n.d.
	<i>Hudovirus</i>	<i>Lepidopteran hudovirus</i> ^a	Moth
	<i>Ixovirus</i>	<i>Blackleg ixovirus</i> ^a	Tick
		<i>Norway ixovirus</i>	Tick
		<i>Scapularis ixovirus</i>	Tick
		<i>Laurel Lake virus</i> ^a	Tick
	<i>Laulavirus</i>		
	<i>Mobuvirus</i>	<i>Mothra mobuvirus</i> ^a	Moth
	<i>Phasivirus</i>	<i>Badu phasivirus</i> ^a	Mosquito
		<i>Dipteran phasivirus</i>	n.d.
		<i>Fly phasivirus</i>	n.d.
		<i>Phasi Charoen-like phasivirus</i>	Mosquito
	<i>Pidchovirus</i>	<i>Wutai mosquito phasivirus</i>	Mosquito
		<i>Pidgey pidchovirus</i> ^a	Moth
<i>Wupedeviridae</i>	<i>Wumivirus</i>	<i>Millipede wumivirus</i> ^a	Millipede

^atype species for the corresponding Genus, n.d. not determined.**Fig. 1** Morphology of selected insect-specific bunyaviruses. Negative-stained virions of Herbert virus (A), Gouléako virus (B), Ferak virus (C), and Jonchet virus (D) sedimented by ultracentrifugation. Scale bars represent 100 nm.

additional non-structural protein named NSm. The NSm protein of Jonchet virus is encoded at the N-terminal part of the GPC protein and posttranslationally cleaved as observed in phleboviruses. In contrast the NSm protein of Ferak virus shows a unique coding strategy as the NSm is encoded in a second small ORF upstream of the GPC ORF that overlaps the N terminus of the GPC

Table 2 Consensus 3' and 5' terminal nucleotides of selected arthropod-specific bunyaviruses

Family	Genus	Virus	Terminal nucleotides
<i>Phasmaviridae</i>	<i>Feravirus</i>	Ferak virus	3' UCAUCAUUUGU ... 5' AGUAGUAAACA ...
	<i>Jonvirus</i>	Jonchet virus	3' UCAUCAU ... 5' AGUAGUA ...
	<i>Orthophasmavirus</i>	Kigluaik phantom virus	3' UCGUCGUGCG 5' AGCAGCAUGC
<i>Peribunyaviridae</i>	<i>Herbevirus</i>	Herbert virus	3' UCAUCAC ... 5' AGUAGUG ...
<i>Phenuiviridae</i>	<i>Goukovirus</i>	Gouléako virus	3' UGUGU ... 5' ACACA ...
	<i>Phasivirus</i>	Badu virus	3' UGUGUUUCUG 5' ACACAAAGAC

**Fig. 2** Schematic coding strategy of selected arthropod-specific bunyaviruses. The viral RNA genome is shown as a black line with the length in nucleotides indicated above each segment. Viral mRNAs are represented as blue arrows with boxes indicating the host-derived primer sequences. Proteins are shown as green boxes with predicted protein sizes given in or beside the boxes.

ORF. The S segment encodes the N protein and in the case of fera- and jonviruses also a nonstructural protein, named NSs, in different overlapping reading frames in the complementary-sense RNA. The NSs ORF is located upstream of the N ORF. The two proteins are translated from the same mRNA species, the result of alternative initiation at different AUG codons. While Shuangou insect virus 1 (genus *Shangavirus*) lacks a NSs, it has been predicted to encode a NSm protein. All other arthropod-specific bunyaviruses seem not to encode NSm or NSs proteins which is in contrast to the vertebrate-infecting bunyaviruses that encode mostly those two proteins.

Goukoviruses contain the shortest bunyavirus genomes identified to date consisting of S, M and L segments of about 1.1, 3.2 and 6.4 kb, respectively.

The L protein of herbeviruses contains an additional region of about 150 amino acids and is thus longer than the related orthobunyaviruses. However, the glycoprotein precursor of herbeviruses is ca. 600 amino acids shorter compared to orthobunyaviruses as the variable region of the Gc protein as well as the NSm protein are lacking.

Viral Replication Cycle

Infection of the target cell is believed to be mediated by interaction of the glycoproteins with cellular receptors. However, no cellular receptors nor the role of each glycoprotein attachment and entry has been identified for any arthropod-specific bunyavirus.

Viral replication takes place in the cytoplasm. The negative-strand RNA is replicated via a complementary full-length positive-sense RNA. The mRNA contains a 5'-methylated capped nonviral (primer) sequence and is truncated at its 3'-end compared to the genomic full-length and is used for transcription and translation. Further analyses on Herbert virus, Gouléako virus, Jonchet virus and Ferak virus have shown that the 5'-nonvirally templated elements are obtained from host cell mRNAs by a cap-snatching mechanism that is mediated by an endonuclease activity of the viral L protein. The 3'-end of the mRNAs terminate ca. 100–200 nucleotides before the end of the genomic RNA. There is no apparent termination signal present. Transcription in arthropod-specific bunyaviruses is similar to that of the plant- and vertebrate-infecting bunyaviruses.

Assembly and release of bunyaviruses occurs at the Golgi complex. Similar observations have been made for Herbert virus. Two types of premature spherical viral particles of high and low electron densities, termed intracellular annular viruses (IAV) and intracellular dense viruses (IDV), were detected in Golgi vesicles. Maturation and budding of Herbert virions occurs at the Golgi membrane of Golgi vesicles filled with IAV and IDV.

Host Associations, Virus Maintenance Cycles and Pathogenicity

The progress in the area of metagenomic sequencing enabled the detection of bunyavirus-like sequences in a huge diversity of arthropods. For example, bunyavirus sequences were found in beetles, bees, dragonflies, moths, water striders, millipedes, spiders and springtails to name just a few, indicating that bunyaviruses have coevolved with their arthropod hosts. Despite the accumulation of sequence data from diverse arthropod hosts, there is little knowledge on the routes of transmission of these viruses and their pathogenicity in arthropod hosts.

Further Reading

- Auguste, A.J., Carrington, C.V., Forrester, N.L., *et al.*, 2014. Characterization of a novel Negevirus and a novel Bunyavirus isolated from *Culex* (Culex) declarator mosquitoes in Trinidad. *Journal of General Virology* 95, 481–485.
- Ballinger, M.J., Bruenn, J.A., Hay, J., *et al.*, 2014. Discovery and evolution of bunyavirids in arctic phantom midges and ancient bunyavirid-like sequences in insect genomes. *Journal of Virology* 88, 8783–8794.
- Hobson-Peters, J., Warrilow, D., McLean, B.J., *et al.*, 2016. Discovery and characterisation of a new insect-specific bunyavirus from *Culex* mosquitoes captured in northern Australia. *Virology* 489, 269–281.
- Käfer, S., Paraskevopoulou, S., Zirkel, F., *et al.*, 2019. Re-assessing the diversity of negative strand RNA viruses in insects. *PLoS Pathogens* 15 (12), e1008224.
- Li, C.X., Shi, M., Tian, J.H., *et al.*, 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* 4, e05378.
- Makhsous, N., Shean, R.C., Droppers, D., *et al.*, 2017. Genome sequences of three novel bunyaviruses, two novel rhabdoviruses, and one novel nyamiviruses from Washington state moths. *Genome Announcements* 5, e01668.
- Marklewitz, M., Handrick, S., Grasse, W., *et al.*, 2011. Gouléako virus isolated from West African mosquitoes constitutes a proposed novel genus in the family *Bunyaviridae*. *Journal of Virology* 85, 9227–9234.
- Marklewitz, M., Zirkel, F., Kurth, A., Drosten, C., Junglen, S., 2015. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. *Proceedings of the National Academy of Sciences of the United States of America* 112, 7536–7541.
- Marklewitz, M., Zirkel, F., Rwego, I.B., *et al.*, 2013. Discovery of a unique novel clade of mosquito-associated bunyaviruses. *Journal of Virology* 87, 12850–12865.
- Shi, M., Lin, X.D., Tian, J.H., *et al.*, 2016. Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543.
- Tokarz, R., Sameroff, S., Tagliaferro, T., *et al.*, 2018. Identification of novel viruses in *Amblyomma americanum*, *Dermacentor variabilis*, and *Ixodes scapularis* ticks. *mSphere* 3, e00614–e00617.

Dicistroviruses (*Dicistroviridae*)

Yanping Chen, Bee Research Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States
Steven M Valles, Center for Medical, Agricultural and Veterinary Entomology, Agricultural Research Service, US Department of Agriculture, Gainesville, FL, United States

© 2021 Elsevier Ltd. All rights reserved.

This is an update of P.D. Christian, P.D. Scotti, Dicistroviruses, In Encyclopedia of Virology (Third Edition), Edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00608-7.

Glossary

Dipteran A member of the insect order Diptera: true flies.

Hemipteran A member of the insect order Hemiptera: true bugs (including aphids).

Horizontal transmission Transmission from one individual to another in the same generation.

Hymenopteran A member of the insect order Hymenoptera: wasps and bees.

Intergenic region Region between the two open reading frames in the dicistrovirus genome.

Lepidopteran A member of the insect order Lepidoptera: moths and butterflies.

Orthopteran A member of the insect order Orthoptera: crickets and grasshoppers.

Pathogenicity The ability of a virus to cause damage and disease in a host.

Penaeid A shrimp from the family Penaeidae.

Polyprotein A protein that is cleaved after synthesis to produce a number of smaller functional proteins.

Vertical transmission Transmission of virus directly from an infected mother to her offspring.

Virion Infective form of a virus outside a host cell, consisting of a nucleic acid core and a protein coat.

VPg A virally encoded protein covalently linked to the 5' end of the viral genome.

Introduction

For years, viruses in the *Dicistroviridae* family had no definitive taxonomic placement and were referred to as “picorna-like” based simply on their similarities to mammalian picornaviruses possessing an RNA genome, the size of virions, the composition of capsids, and other biophysical characteristics. Genome sequencing revolutionized biological science across many fields including virus taxonomy. In 1998, the first complete genome sequence of an insect picorna-like virus, *Drosophila C virus* (DCV) was published and predicted to possess a genome organization different from that of mammalian picornaviruses. Unlike other picornaviruses with monopartite, monocistronic genomes, DCV has a monopartite bicistronic genome with replicase proteins encoded by a 5'-proximal ORF and capsid proteins encoded by a 3'-proximal ORF, which are separated by an intergenic region (IGR), suggesting this virus belongs to a previously undescribed virus family. Subsequently, the increasing availability of complete genome sequences of insect picorna-like viruses and phylogenetic data have confirmed the distinct genome architecture of the dicistroviruses, and have conclusively defined *Dicistroviridae*, a taxonomic designation which was officially adopted by the ICTV in 2002. The family currently includes fifteen members officially recognized by the ICTV and many additional unclassified species.

Taxonomy and Classification

The *Dicistroviridae* family is a member of the Order *Picornavirales* and currently comprises 15 species classified into three genera, *Aparavirus* (derived from acute bee paralysis virus), *Cripavirus* (derived from cricket paralysis virus), and *Triatovirus* (derived from *Triatoma* virus). There are several other potential candidates for the family but these have yet to be accepted as species by the ICTV. For the purposes of this article, we will limit our discussion to only those officially classified species shown in Table 1.

Biophysical Properties

Dicistrovirus virions are roughly spherical as observed by electron microscopy with a particle diameter of approximately 30 nm and no envelope (Fig. 1(a)). The virions are composed of 60 protomers, each comprised of a single molecule of each of the major capsid proteins VP1, VP2, and VP3 (Fig. 1(b)). The major capsid proteins VP1 to VP3 exhibit a jelly roll β -sandwich fold and form the capsid shell with pseudo- $T = 3$ icosahedral symmetry. A smaller protein, VP4, is also present inside the virion and in contact with the RNA genome. In Cricket paralysis virus (CrPV) and Israeli acute paralysis virus (IAPV), VP4 is located on the internal surface of the five-fold axis below VP1, but in contrast, is disordered in *Triatoma* virus (TrV). In most species, a protein precursor (VP0) is present which is cleaved to yield capsid proteins VP3 and VP4 during virion maturation. An Asp-Asp-Phe (DDF) motif, which is part of the VP1 subunit and conserved among dicistroviruses may be involved in VP0 cleavage. There are spike-like

Table 1 Members of the family *Dicistroviridae*. Isolate and vernacular names are shown in brackets

Genus	Species (isolate name)	Accession #	Abbreviation
<i>Aparavirus</i>	<i>Acute bee paralysis virus</i> ^a (Acute bee paralysis virus)	[AF150629]	ABPV
	<i>Israeli acute paralysis virus</i> (Israeli acute paralysis virus)	[EF219380]	IAPV
	<i>Kashmir bee virus</i> (Kashmir bee virus)	[AY275710]	KBV
	<i>Mud crab virus</i> (Mud crab virus)	[HM777507]	MCV
	<i>Solenopsis invicta virus 1</i> (<i>Solenopsis invicta</i> virus 1)	[AY634314]	SINV – 1
	<i>Taura syndrome virus</i> (Taura syndrome virus)	[AF277675]	TSV
<i>Cripavirus</i>	<i>Cricket paralysis virus</i> ^a (Cricket paralysis virus)	[AF218039]	CrPV
	<i>Aphid lethal paralysis virus</i> (Aphid lethal paralysis virus)	[AF536531]	ALPV
	<i>Drosophila C virus</i> (<i>Drosophila</i> C virus)	[AF014388]	DCV
	<i>Rhopalosiphum padi virus</i> (<i>Rhopalosiphum padi</i> virus)	[AF022937]	RhPV
<i>Triatovirus</i>	<i>Triatoma virus</i> ^a (<i>Triatoma</i> virus)	[AF178440]	TrV
	<i>Black queen-cell virus</i> (Black queen cell virus)	[AF183905]	BQCV
	<i>Himetobi P virus</i> (<i>Himetobi</i> P virus)	[AB017037]	HiPV
	<i>Homalodisca coagulata virus 1</i> (<i>Homalodisca coagulata</i> virus 1)	[DQ288865]	HoCV – 1
	<i>Plautia stali intestine virus</i> (<i>Plautia stali</i> intestine virus)	[AB006531]	PSIV

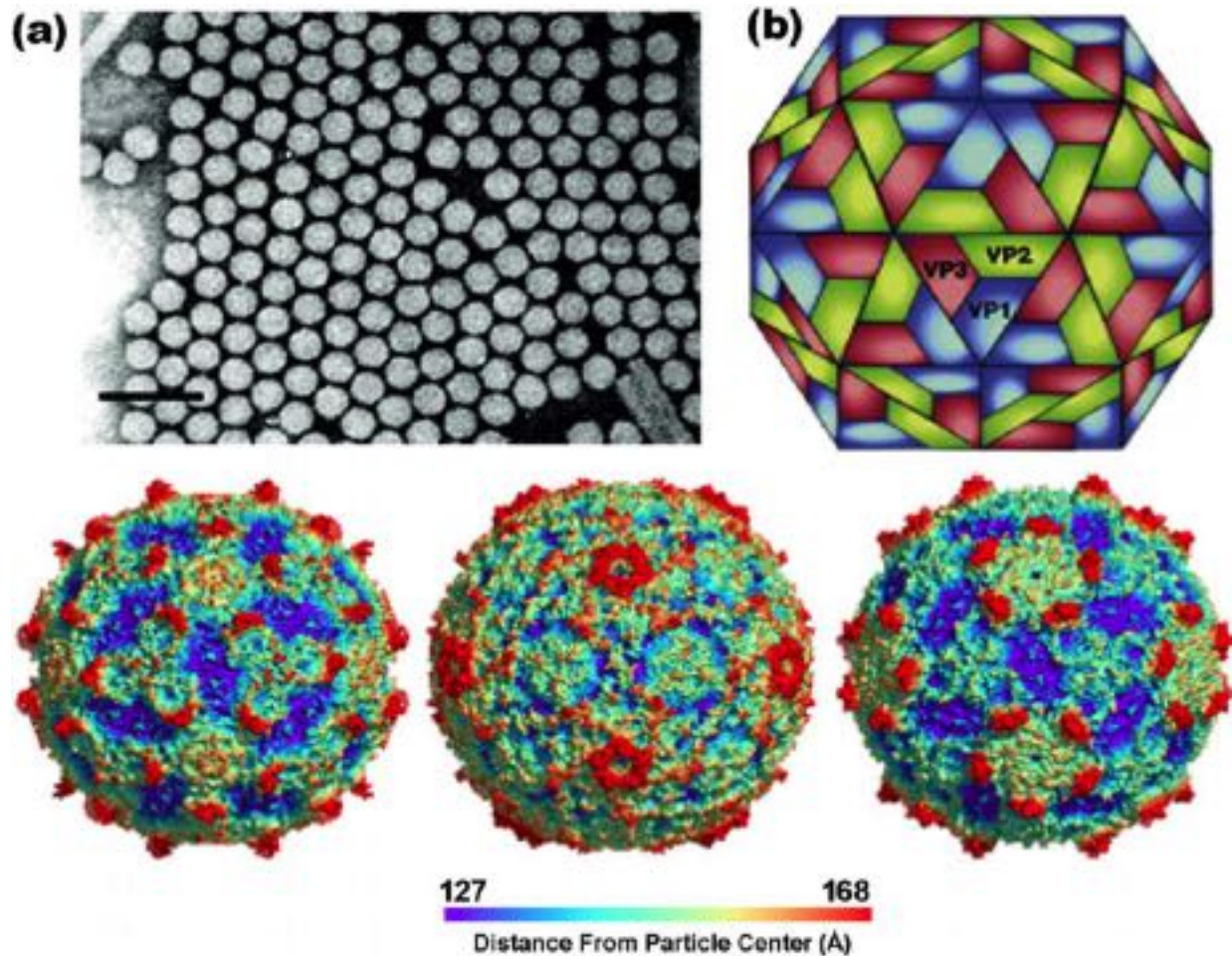
^aType species for the Genus.

Fig. 1 Virion structure. (a) Negative stained electron micrograph of CrPV virions that appear spherical in shape. (b) Schematic diagram of a dicistrovirus virion showing the surface packing of the coat proteins (VP1, VP2, and VP3) (courtesy of A. E. Mechaly, with permission). The comparison of surface structures of, from left to right, *Triatoma* virus, Israeli acute bee paralysis virus, and Cricket paralysis virus. Color scale indicates the distance from the particle center (Å). Protrusions are shown in red and depressions are shown in blue (courtesy of A. E. Mechaly, with permission). (a) Courtesy of Carl Reingannum, with permission.

protrusions that are formed by two antiparallel β -strands from the CD loop of VP3 and a C-terminal β -strand of VP1 on the surface of virions; however, the size of the protrusions varies among different dicistroviruses (Fig. 1(c)–(e)). Unlike many picornaviruses, some dicistroviruses such as Cricket paralysis virus (CrPV), Black queen cell virus (BQCV), and IAPV do not contain a hydrophobic pocket in capsid protein VP1, which is naturally occupied by a lipid and can be targeted by capsid-binding antiviral compounds.

Proteins account for 70% of virion weight. The approximately 200 kDa nonstructural polyprotein and 100 kDa structural polyprotein are encoded by ORF 1 and ORF 2, respectively. Virions contain three major structural (capsid) viral proteins, VP1, VP2, and VP3. The size of these capsid proteins ranges from 24 to 40 kDa; an exception is Taura syndrome virus (TSV), in which VP1 is 55 kDa. A fourth smaller capsid protein (VP4) of roughly 4.5–9 kDa has also been reported in some species.

Virions are stable in acidic conditions (to pH = 3.0) and resistant to detergents and organic solvents such as ether and chloroform. Mature virions have a buoyant density of 1.34–1.39 g ml⁻¹ in CsCl in the pH range of 7–9 and sedimentation coefficients of between 153 and 167 S. However, physicochemical properties have not been fully established for all members of the family. A summary of the biophysical properties of dicistrovirus virions is provided in Table 2.

Organization of the Dicistrovirus Genome

The positive-sense single-stranded RNA genomes of dicistroviruses are monopartite and dicistronic and approximately 8500–10,000 nt in size. The RNA genomes comprise two non-overlapping ORFs of approximately 5500 and 2600 nt that are separated by an untranslated intergenic region (IGR) of ~190 nt and flanked by 5' and 3' untranslated regions (UTRs) (Fig. 2). The length of the UTRs at each end of the genome varies. The 5'-proximal and 3'-proximal ORFs encode non-structural and structural protein precursors, respectively. Each ORF is preceded by a specific RNA structure identified as an internal ribosome entry site (IRES), which allows for the initiation of translation in a cap-independent manner. Components of the non-structural

Table 2 Summary of some biophysical properties of dicistroviruses.

Virus		Molecular weight of major capsid proteins (kDa) ^a	Buoyant density in CsCl (g ml ⁻¹)	Particle diameter (nm)
Genus	Species			
<i>Aparavirus</i>	<i>Acute bee paralysis virus</i>	35, 33, 24, 9	1.34	30
	<i>Israeli acute paralysis virus</i>	no data available	1.33	27
	<i>Kashmir bee virus</i>	41, 37, 25, 6	1.37	30
	<i>Mud crab virus</i>	53, 35, 22	no data available	30
	<i>Solenopsis invicta virus 1</i>	41, 36, 25 (22)	1.35	31
	<i>Taura syndrome virus</i>	55, 40, 24 (58)	1.34	31
<i>Cripavirus</i>	<i>Cricket paralysis virus</i>	35, 34, 30 (43)	1.37	27
	<i>Aphid lethal paralysis virus</i>	34, 32, 31 (41)	1.34	27
	<i>Drosophila C virus</i>	31, 30, 28, 9 (37)	1.34	27
	<i>Rhopalosiphum padi virus</i>	31, 30, 28 (41)	1.37	27
<i>Triatovirus</i>	<i>Triatoma virus</i>	39, 37, 33	1.39	30
	<i>Black queen-cell virus</i>	34, 32, 29, 6	1.34	30
	<i>Himetobi P virus</i>	37, 33, 28	1.35	29
	<i>Homalodisca coagulata virus 1</i>	no data available	no data available	no data available
	<i>Plautia stali intestine virus</i>	33, 30, 26, 5	no data available	30

^aMinor virion components are shown in brackets. These are presumed to be precursors of VP4–VP3.

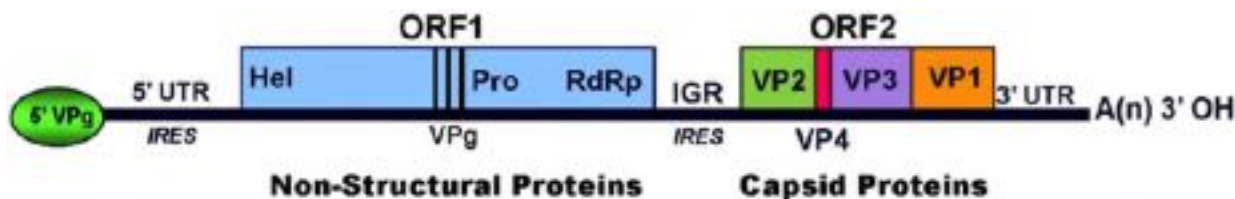


Fig. 2 Organization of the dicistrovirus genome. The RNA genome contains two non-overlapping open reading frames (ORFs) separated by an intergenic region (IGR) and flanked by the 5' and 3' untranslated regions (UTRs). The 5' proximal ORF encodes the nonstructural proteins: RNA helicase (Hel), 2 x VPg, cysteine protease (Pro), and RNA-dependent RNA polymerase (RdRp). The 3'-proximal ORF encodes structural proteins: VP2, VP4, VP3, and VP1. The internal ribosome entry sites (IRESs) are located within the 5'-UTR and the IGR.

polyprotein include an RNA helicase (Hel), 3C-like cysteine protease (Pro), and RNA-dependent RNA polymerase (RdRp) lying in the order (5′–3′) Hel-Pro-RdRp. The structural proteins that build the capsid are arranged in the order of VP2-VP4-VP3-VP1. VP4 is cleaved from VP3 during the maturation of the virion. A small genome-linked virus protein (VPg), derived from ORF1, is covalently attached to the 5′ end of the genome and plays an important role in RNA replication. The VPg sequence is repeated in most dicistrovirus genomes with the number of repeats being species-dependent. The presence of multiple VPg copies is predicted to allow for multiple progeny RNA molecules to be produced per template. The poly(A) tail at the 3′ end of the genomes is presumed to function synergistically with the IRES to enhance IRES-mediated translation efficiency,

Virus Replication and Genome Expression

The ways in which the dicistrovirus viral capsid proteins interact with host cellular receptors to facilitate entry is not fully understood. Likely, however, dicistroviruses enter host cells via clathrin-mediated endocytosis. The recent development of infectious full-length clones of dicistroviruses will make the study of virus replication and genome expression possible. Dicistroviruses are considered to use the same general replication mechanisms common to picornaviruses. Following viral attachment to the host cell, conformational changes occur in the capsid facilitating genome release into the cytoplasm. The RNA genome is immediately translated into polyproteins. The multiplication of the viral genomic RNA involves the synthesis of a negative-strand RNA intermediate from the positive strand, which, in turn, serves as a template for the production of progeny positive-strand RNAs. The viral RNA-dependent RNA polymerase (RdRP) named 3Dpol is responsible for primarily RNA synthesis. The 3Dpol uses VPg (also called 3B) at the 5′ terminus of the genomic RNA as a primer to initiate the replication process. The first step in protein-primed initiation is the uridylation of a tyrosine residue at position 3 of VPg by the 3D viral polymerase. The resulting VPg-pUpU then serves as a primer for the synthesis of both negative as well as positive-strand RNA at the 3′- or 5′-end of the template genomes.

The viral genomic RNA also functions as mRNA which is translated into proteins for the production of new virion materials. The initiation of protein synthesis occurs immediately upon entry into the host cell and coincides with the shutdown or downregulation of host cell protein synthesis. For example, CrPV infection inhibits the host translational machinery, a process called host shutoff which is concomitant with an increase in CrPV viral protein synthesis. Translation of the dicistronic RNA genome of dicistroviruses proceeds directly from internal ribosome entry sites (IRESs) and RNA stem-loop structures located within the 5′-UTR and the IGR. Although the mechanism of IRES-directed translation initiation has not been fully defined and varies among different IRES elements, the IGR-IRES of dicistroviruses is one of the best-studied IRES elements. The IGR IRES can initiate translation without codon-anticodon base pairing between the initiation AUG codon and the initiator Met-tRNA. Moreover, the IGR IRES can directly recruit and bind the ribosome in the absence of canonical initiation factors. The activities of the IGR IRES depend on two conserved and independently folded domains that work in concert to mediate ribosome recruitment and direct translation initiation. In most cases, the initiation codon is CCU (Pro) while the first codon of VP2 encodes an alanine residue. It has been shown empirically that the glutamine at the 5′ end of PSIV VP2 can be replaced with any other amino acid to produce a mature protein in an in vitro translation system. Conversely, the 5′-UTR IRES is not well conserved within the group and there are no clear structural homologies between the 5′-UTR IRES and IGR IRES. Translation activity of the IGR IRES is relatively stronger than that of the 5′-UTR IRES. However, a study with CrPV showed that IGR IRES translational activity is coupled with 5′UTR IRES translational activity and the increased IGR IRES translation of capsid proteins is stimulated when the expression of non-structural proteins is active.

Host Range

To date, all members of the *Dicistroviridae* have been isolated from arthropods. Most dicistroviruses have a range of multiple hosts. The hymenopteran viruses, Acute bee paralysis virus (ABPV), Kashmir bee virus (KBV), IAPV, KBV, and BQCV which were originally identified in the European honey bee (*Apis mellifera*) have recently been detected in other bee species, which is likely due to the sharing of foraging sites as well as the high mutation rates of the RNA viruses in driving viral adaptation and host expansion. In addition, it is evident that these bee viruses are capable of replicating in the ectoparasitic mite *Varroa destructor*, which acts as a vector of viruses in honey bee colonies. Studies have shown that the host of predilection for *Solenopsis invicta* virus 1 is the red imported fire ant, *Solenopsis invicta*. However, in areas where *S. invicta* and the congener, *Solenopsis geminata* are sympatric, the virus will infect both species. SINV-1 has also been detected in the black imported fire ant, *S. richteri*.

Dicistroviruses found in Hemiptera (the combined orders of Heteroptera and Homoptera), also have multiple host ranges. *Rhopalosiphum padi* virus (RhPV) has been isolated from laboratory and field populations of the aphids *Rhopalosiphum padi*, *R. maidis*, *R. rufiabdominalis*, *Schizaphis graminum*, *Diuraphis noxia*, and *Metopolophium dirhodum*. Aphid lethal paralysis virus (ALPV) has been isolated from several species of aphids, including *R. padi*, *M. dirhodum*, and *Sitobion avenae*; more recently, it has been identified in the honey bee *A. mellifera*. The Himetobi P virus (HiPV) and *Plautia stali* intestine virus (PSIV) are found in true bugs rather than aphids with HiPV having been isolated from the leafhoppers *Laodelphax striatellus*, *Sogatella furcifera*, and *Nilaparvata lugens*, and PSIV from the brown-winged green bug, *Plautia stali*. TrV is also a virus of hemipterans and has been found in the hematophagous triatomine bug, *Triatoma infestans*, which is also a vector of the protozoan agent responsible for Chagas disease in South America.

DCV has a host range restricted to dipterans and has been isolated from *Drosophila melanogaster* and the sibling species *D. simulans*. Taura syndrome virus (TSV) is a virus of penaeid shrimps and has been isolated from a number of shrimp species including *Litopenaeus vannamei*, *L. stylirostris*, *Metapenaeus ensis*, and *Penaeus monodon*. *P. chinensis* is highly susceptible to infection in experimental bioassays. CrPV was originally isolated from the field crickets, *Teleogryllus oceanicus* and *T. commodus*. This virus has an exceptionally broad host range as it has subsequently been isolated from a further 20 species belonging to five taxonomic orders: Orthoptera, Hymenoptera, Lepidoptera, Hemiptera, and Diptera, as well as a range of cultured insect cells. CrPV is the only dicistrovirus isolated from lepidopterans – in fact from ten lepidopteran species.

All of the above records refer to the natural host range of the viruses. To a certain extent, studies on the experimental host range of dicistroviruses are limited and those that have been carried out have not substantially extended the known host ranges. Again the exceptions are CrPV and DCV. CrPV replicates in a number of established insect cell lines including those from *Drosophila*, the hemipteran *Agallia constricta*, and the lepidopterans *Pieris rapae*, *Plutella xylostella*, *Spodoptera ornithogalli*, and *Trichoplusia ni*. In addition to insect cell lines, CrPV has also been found to replicate readily in larvae of the greater waxmoth, *Galleria mellonella*. This is an easy insect to rear and maintain and virus yields can be very high. Apart from CrPV, the only other dicistrovirus shown to replicate in cultured cells is DCV, which multiplies in several *Drosophila* cell lines (some DCV isolates also replicate in the greater waxmoth). Reports that TSV can replicate in some mammalian cell lines have never been substantiated and may simply be attributable to the production of a cytopathic effect in the absence of virus replication.

Pathology and Transmission

Dicistroviruses are significant pathogens of agricultural and health importance. Dicistrovirus infections vary considerably in virulence and pathogenicity. While in some instances, dicistroviruses may not produce noticeable pathological symptoms, most dicistrovirus infections result in impaired functions, severe diseases, reduced life expectancy, and increased mortality in infected hosts. RNA interference (RNAi) is a major immune defense mechanism against viruses in arthropods. Viruses, in turn, have evolved to encode RNA silencing suppressor proteins to counterattack antiviral defense of hosts. The activity of the virus-encoded suppressors of host RNAi (VSRs) has been identified in dicistroviruses including DCV, CrPV, and IAPV. This evolutionary arms race between dicistroviruses and their hosts defines the pathogenicity of viruses in hosts.

Most of the dicistroviruses exhibit a tissue tropism toward some part of the alimentary canal, often replicating in epithelial cells of the gut and subsequently shedding virus particles into the gut lumen where the virus accumulates in the feces and serves as inoculum. DCV is a viral pathogen affecting the model organism, *D. melanogaster*, which has often been used for deciphering virus-host interactions. The DCV infection triggers a nutritional stress in infected flies, which is the result of intestinal obstruction; the virus displays an affinity for the food-storage organ of the digestive tract, the crop, in infected hosts. The related virus, CrPV, does not trigger the same pathology.

The hymenopteran viruses, ABPV, IAPV, KBV, and BQCV are highly pathogenic to the honey bee, *A. mellifera*. The viruses attack all developmental stages and castes of the honey bees and replicate in most tissues of infected bees, including the tissues of the gut, fat body, mandibular glands, hypopharyngeal glands, hemolymph, and nerve. ABPV and IAPV infections are associated with paralysis of adult bees, abnormal trembling of the wings and body, flightlessness and mortality. ABPV, IAPV, and KBV infections can be lethal at times, especially in the presence of the parasitic *Varroa* mite. BQCV affects mainly developing queen larvae and pupae, causing infected pupae to turn dark and quickly die, with the wall of the queen cell becoming a dark color, a characteristic symptom of infection. The hymenopteran virus infections in honey bees have often been linked to worldwide colony losses, including the losses caused by the colony collapse disorder (CCD), a devastating malady that wiped out bees by the millions in the U.S. during the winter of 2006–2007.

The fecal-oral route features prominently in the transmission of SINV-1. Uninfected fire ants acquire SINV-1 infections when exposed to baits containing crude homogenates of SINV-1-infected worker ants and purified preparations of the virus. Virus transmission in this manner is easy and facilitates its use as a biopesticide (bait formulation) to control this pest ant. Fire ant colonies infected with SINV-1 can exhibit a chronic asymptomatic condition, or, under certain stressors, the colony exhibits significant mortality. Interestingly, SINV-1 prevalence in field colonies of fire ants is correlated strongly with increasing temperature; the virus prevalence is very high during warmer seasons and nearly undetectable during colder parts of the year.

TSV and Mud crab virus (MCV) are two dicistroviruses of crustaceans. MCV is associated with sleeping disease (SD) in farmed Portunidae crabs. TSV is responsible for one of the more devastating diseases of penaeid shrimp, Taura syndrome, which affects the aquaculture industry worldwide. TSV infects post larvae, juveniles and adult stages of shrimps; it has a broad range of tissue tropism and the infection can spread to different types of tissues, including the antennal gland, stomach, gill, lymphoid organ, sub-cuticle and connective tissues. TSV infection involves acute and chronic phases at the farm level. During the acute phase of the infection, infected shrimp display a variety of disease symptoms including anorexia, discoloration, lethargy and erratic swimming behavior. The mortality of infected hosts can be up to 95%. Shrimp that survive the acute stage pass through a transition phase and then become long-term (chronic) carriers of the virus.

The *Dicistroviridae* family consists of members whose hosts are agricultural and urban insect pests. SINV-1 was the first virus identified in the red imported fire ant, *Solenopsis invicta*, a pest whose sting is known to cause serious medical problems in people. SINV-1 infects all developmental stages and caste members of the fire ant, exhibiting a tissue tropism for gut tissue especially midgut. *S. invicta* queens infected with SINV-1 have lower body weights, reducing the probability of colony

founding. The potential use of SINV-1 as a biopesticide for pest control is currently under investigation. RhPV, ALPV, HiPV, and PSIV infect hemipteran pests that are important vectors of plant diseases. The viral infections primarily affect gut tissues of the hemipteran hosts. The host of TrV is a blood-sucking reduviid bug, which is an important vector of the protozoan parasite *Trypanosoma cruzi* responsible for Chagas disease in humans. TrV is the sole viral pathogen of triatomine bugs and replicates within the intestinal epithelial cells of the hosts, causing a delayed molting cycle and high mortality rate; its use as a biological control agent for vectors of Chagas disease has been proposed. HoCV-1 infects and causes increased mortality in the polyphagous glassy-winged sharpshooter, an important insect vector of the xylem-limited bacterial plant pathogen *Xylella fastidiosa*. CrPV. HoCV-1 grows rather efficiently in *Drosophila*, is highly pathogenic to the European olive fruit fly – a possible target pest for this virus.

Transmission of dicistroviruses involves both horizontal and vertical pathways. Transmission via ingestion and the alimentary canal feature prominently in dicistrovirus infection acquisition and transmission. Studies with hymenopteran dicistroviruses, ABPV, IAPV, KBV, and BQCV demonstrate that transmission in honey bees involves multiple pathways including food-borne, venereal, vector-borne, and vertical transmission. The detection of honey bee dicistroviruses in colony food stores, the digestive tract of infected bees, and feces suggests that horizontal transmission via the fecal-oral route is a prominent mechanism of dissemination. The detection of viruses in adult drones, semen, and the spermatheca of queens (the organ in which sperm is stored) implies transmission via the venereal route, where viruses are transmitted from infected males to females during mating. The route of horizontal transmission with the most consequence is by the parasitic mite *Varroa*. The varroa mite acquires and transmits the virus from infected bees to health bees. In addition to facilitating virus transmission, *Varroa* parasitism also suppresses honey bee immunity, thereby activating virus replication and increasing the susceptibility of infected hosts to further pathogenic infection. The detection of viruses in eggs as well as in the larval stages of bee hosts that are not normally associated with virus vector, suggests that viruses might be transmitted vertically from infected queens to her offspring.

Evidence of vertical transmission has also been identified in other dicistroviruses. ALPV RNA can be detected in the developing embryos inside infected females of the aphid host, *R. padi*, suggesting that ALPV can be vertically transmitted. RhPV can be transmitted not only vertically between aphid hosts but horizontally via the plant by using the plant as a passive reservoir. However, RhPV does not replicate in the plant. DCV is vertically transmitted and is associated with transovum transmission on the egg surface. A similar transovum transmission is also found with CrPV where the surface sterilization of eggs with dilute hypochlorite blocks the transmission of the virus to emerging nymphs.

Evidence of horizontal transmission is also documented for other dicistroviruses. The transmission of TSV is predominantly by cannibalism of infected or dead shrimp by healthy shrimp. The virus can also be spread from one farm to another farm by seagulls and aquatic insects. In the case of TrV and its host, *T. infestans*, the virus is transmitted through the fecal-oral pathway, where infected insects excrete the virus in their feces and healthy insects become infected by coprophagy. The leafhopper-infecting virus Homalodisca coagulata virus 1 (HoCV-1) spreads most readily in high population densities through contact among infected individuals, contact with virus-contaminated surfaces, and/or as an aerosol in leafhopper excreta. With DCV, the natural route of transmission in *Drosophila* is poorly characterized, as the majority of the virus infection assays in *Drosophila* utilize injection. However, when uninfected males are placed with uninfected females, the males become infected (and vice versa). In fact, even if uninfected female flies are placed on the same media that infected males have been allowed to feed on it for several hours, the females become infected, clearly indicating that the virus is transmitted horizontally among individuals.

Geographic and Strain Variation

Some dicistroviruses infect specific hosts with a wide geographic distribution or have been found to infect a number of species spread over a large geographic range. Several studies have looked at strain variations between geographic isolates of dicistroviruses using a variety of techniques. CrPV and DCV possess a range of biological and serological characteristics that show differentiation between geographical isolates. With CrPV, two major serogroups have been identified that separate Australian and New Zealand isolates from North American isolates. With DCV, isolates from different localities varied both in their pathogenicity and virus yield after being injected into virus-free flies. Additional genetic studies on CrPV and DCV have utilized ribonuclease T1 fingerprinting and subsequently polymerase chain reaction and restriction endonuclease analysis. Estimates of the maximum nucleotide divergence between isolates of CrPV and DCV were about 10%. In the case of CrPV, the North American isolates were quite distinct from the antipodean isolates, a finding that reflects the serological data.

More recently, a number of molecular studies with the honeybee dicistroviruses have been undertaken to determine the levels of genetic variation between isolates of these viruses. In most instances, slightly different regions of the virion protein coding regions (usual regions of VP3) have been used, which makes a direct comparison between studies quite difficult. Nevertheless, IAPV and ABPV isolates show levels of nucleotide identity between 90% and 100%, but isolates from the same geographic region are more closely related than isolates from more distant locations. One such study has revealed that viruses isolated from different central European regions were more similar to each other than to isolates from North America or from the UK. Similar patterns are also evident for BQCV and KBV with nucleotide identity within species at 90%–100%. To put these values into perspective, the region used for the KBV studies was 75% identical to the same region from ABPV.

For TSV, the situation is slightly different with nucleotide identities between isolates from North and Central America and Asia ranging from 95% to 100%. These lower levels of diversity may indicate that TSV has rapidly spread into many of the regions and hosts where it is now found, a hypothesis that is supported to some extent by the rapid emergence of the disease over the last two decades. There is also some evidence that suggests there are serological differences between isolates.

The RdRp region of the SINV-1 genome was sequenced from infected ant colonies collected from across the USA, northern Argentina, and northern Taiwan. Nucleotide sequences were calculated to exhibit an overall identity of >90% between geographically-separated samples. A total of 171 nucleotide variable sites (representing 22.4% of the region amplified) were mapped across the SINV-1 RdRp alignment and no insertions or deletions were detected. Phylogenetic analysis at the nucleotide level revealed the clustering of Argentinean sequences, distinct from the USA sequences. SINV-1 RdRp sequences derived from populations of *S. invicta* from northern Taiwan resided within the multiple USA groupings.

Relationships Within the Family

Phylogenetic relationships with the *Dicistroviridae* family are depicted in Fig. 3. The three genera, *Aparavirus*, *Cripavirus* and *Triatovirus*, form distinct clades. Triatoviruses are distinct from aparaviruses and cripaviruses by the presence of prominent projections on the virion surface and a DDF motif in the homologous location in VP3 and the absence of icosahedrally ordered VP4. Aparaviruses and cripaviruses are different from each other in their distinctive features exhibited in the IRES. The typical IGR-IRES of cripaviruses has a conserved bulge sequence (UGAUCU and UGC), while those of aparaviruses have a different bulge sequence (UGGUUACCCAU and UAAGGCCUU) and an additional stem loop in the 3' region of the IGR-IRES. Phylogenetic analysis of deduced amino-acid sequences of the three major coat proteins (ORF 2) shows three distinct clades (Fig. 3), supporting the classification of species into three genera in the *Dicistroviridae* family.

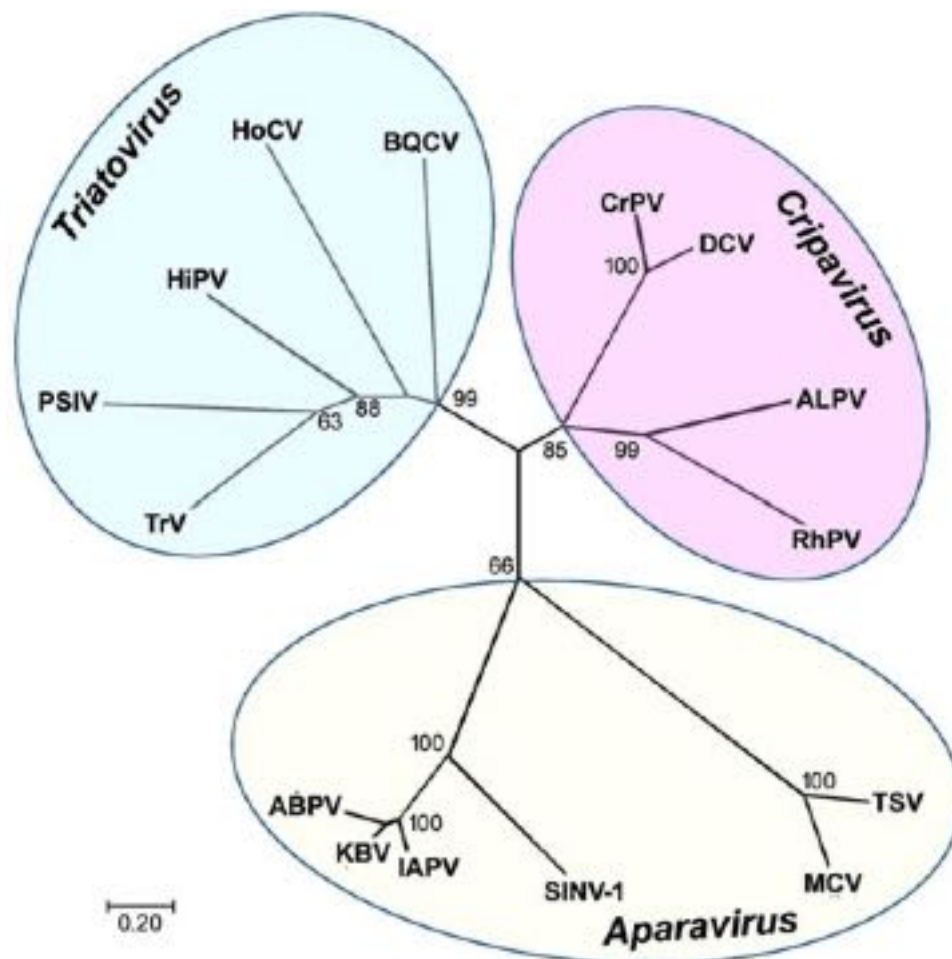


Fig. 3 Phylogenetic tree constructed from the amino acid identity of the structural proteins encoded by ORF2 shows the relationships among members of the family *Dicistroviridae*. Numbers on the branches indicate percentage of bootstrap support out of 1000 replications. The scale indicates amino acid distance. The tree was generated using neighbor-joining algorithm implemented in MEGA7.

Similarity With Other Taxa

Members in the family *Dicistroviridae* have similarities with viruses in the families *Iflaviridae*, *Picornaviridae*, *Marnaviridae*, and *Secoviridae*. The genomes of viruses of these taxa are positive sense ssRNAs with a VPg and a poly (A) tail, and are translated into autoproteolytically processed polypeptide(s). Nonstructural proteins contain sequence motifs for helicase (Hel), 3C-like cysteine proteinase (Pro) and RNA-dependent RNA polymerase (RdRp) with the characteristic gene order: (Hel)-(Pro)-(RdRp). Virions contain capsid proteins organized in a module containing three related jelly-roll domains that form non-enveloped, isometric particles of pseudo $T = 3$ symmetry and 30 nm diameter.

Acknowledgments

We would like to thank Dr. Peter D. Christian for all his help and support.

Further Reading

- Bonning, B.C., Miller, W.A., 2010. Dicistroviruses. *Annual Review of Entomology* 55, 129–150.
- Bonning, B.C., 2009. The *Dicistroviridae*: An emerging family of invertebrate viruses. *Virologica Sinica* 24, 415–427.
- Chen, Y.P., Nakashima, N., Christian, P., *et al.*, 2012. Dicistroviridae. In: King, A.M., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, pp. 840–845.
- Gall, O.L., Christian, P., Fauquet, C.M., *et al.*, 2008. Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo- $T = 3$ virion architecture. *Archives of Virology* 153, 715–727.
- Valles, S.M., Chen, Y.P., Firth, A.E., *et al.*, 2017. ICTV virus taxonomy profile: Dicistroviridae. *Journal of General Virology* 98, 355–356.
- Valles, S.M., Chen, Y., Firth, A.E., *et al.*, 2019. Dicistroviridae in the ICTV report on virus taxonomy, 10th report. ICTV. Available at: https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/dicistroviridae.

Entomobirnaviruses (*Birnaviridae*)

Marco Marklewitz, Institute of Virology, Charité – University Medicine Berlin, Berlin, Germany

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Antagonist An agent that binds a target consequently blocking or dampening a biological process.

Cytopathic effect Structural changes of infected cells resulting from a viral infection.

Metagenome Summary of all genomes and genes obtained from an environmental sample. The data are generated through non-targeted shotgun sequencing of nucleic acids extracted from the sample.

Open reading frame Continuous stretch of codons that begins with a start codon and ends at a stop codon which has the ability to be translated into a protein.

Phylogenetic clade A group of organisms derived from a common ancestor species.

Positive-sense Also known as 'plus-stand' which refers to the translation sense of the viral RNA (5' to 3') that may be directly translated into the desired viral proteins.

Protease An enzyme that cuts or cleaves proteins.

Prototype virus The first representative virus of a species or genus.

Ribonucleoprotein complex A nucleoprotein that contains RNA.

RNA interference A biological mechanism of inhibiting gene expression or translation. Interfering RNA molecules bind certain mRNA molecules, leading to their neutralization.

Classification (Compact)

Entomobirnaviruses are small RNA viruses that infect exclusively insects and they are the only ones in the *Birnaviridae* family that do. When discovered in 1958 during a series of experiments with Sigma virus (genus *Sigmavirus*, family *Rhabdoviridae*), the prototype entomobirnavirus *Drosophila* X virus was identified as a contaminant in laboratory-reared fruit flies of the species *Drosophila melanogaster* (Table 1). The virus has also been detected in various *Drosophila melanogaster* cell lines. However, *Drosophila* X virus has not yet been detected in wild populations and thus its origin is still unclear. It was speculated that it could have pre-existed in *Drosophila* broods in a non-pathogenic form or that it might have originated as a contaminant in fetal bovine serum used in cell culture. Over 50 years later, published in 2012, the first tentative entomobirnavirus was identified in free-living insects.

Table 1 Classified and unclassified entomobirnaviruses, their host, country, and year of initial detection. Accession numbers provided refer to the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>).

Virus	Host	Country	Year	Accession numbers
<i>Drosophila</i> X virus ^a	<i>Drosophila melanogaster</i> ^b	France	1958	U60650 AF196645
mosquito X virus	<i>Anopheles sinensis</i>	Yunnan province, China	2009	JX403941 JX403942
unclassified:				
<i>Culex</i> Y virus ^a	<i>Culex pipiens</i> complex	Bad Segeberg, Germany	2010	JQ659254 JQ659255
culicine-associated Z virus	<i>Ochlerotatus caspius</i> <i>Ochlerotatus detritus</i>	Camargue region, France	2011	KF298271 KF298272
Espirito Santo virus ^a	unknown	Brazil	2010	JN589003 NJ589002
Eridge virus	<i>Drosophila immigans</i>	United Kingdom	2011	KU754527 KU754528
Thirlmere virus ^a	unknown	United Kingdom	1980	not available
Port Bolivar virus ^a	<i>Aedes sollicitans</i>	Galveston County, Texas, United States	2013	MT263973 MT263974

^avirus isolate available.

^blaboratory contamination.

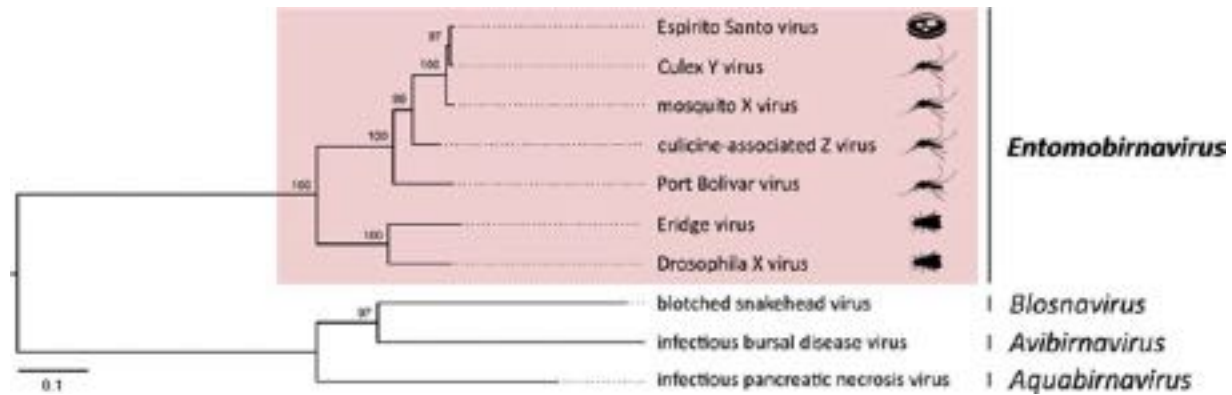


Fig. 1 Phylogenetic relationship of entomobirnaviruses. Phylogenetic analysis of entomobirnaviruses and prototype members of the birnavirus genera *Avibirnavirus*, *Blosovirus*, and *Aquabirnavirus*. Pictograms illustrate the source of the respective virus: laboratory contamination, mosquito, fruit fly.

Culex Y virus was isolated from hibernating *Culex pipiens* complex mosquitoes collected in a cave in Bad Segeberg, Germany. Today, natural infections with entomobirnaviruses have been detected in *Drosophila immigrans* and in various Culicidae sp. mosquitoes of France, China, Germany, and the United Kingdom.

The genus *Entomobirnavirus* belongs to the family *Birnaviridae* and is currently one of four established genera (Fig. 1). Entomobirnaviruses form a monophyletic clade within the family *Birnaviridae* and are distantly related to the other three birnavirus genera that comprise vertebrate infecting viruses. Apart from the species *Drosophila X virus*, which is represented by the prototype entomobirnavirus *Drosophila X virus*, a second species, *mosquito X virus*, has formally been established in the genus. In addition, there are six unclassified but suggested entomobirnaviruses that have not yet been officially assigned to species in the genus (Table 1). *Drosophila X virus* is a widely used model pathogen for the study of different RNAi mechanisms in insect cells. As an insect virus, *Drosophila X virus* can be propagated in cell culture under standard laboratory conditions.

Together with Eridge virus, which has been detected in *Drosophila immigrans* during a metagenomic study, *Drosophila X virus* forms a sister clade to the other five mosquito-associated entomobirnaviruses: Espirito Santo virus, *Culex* Y virus, *mosquito X virus*, culicine-associated Z virus, and Port Bolivar virus (Fig. 1). However, the host of Espirito Santo virus remains unknown as it has been detected in an *Aedes albopictus* cell line inoculated with serum from a Dengue virus (genus *Flavivirus*, family *Flaviviridae*) infected patient. In addition, Thirlmere virus has been isolated from a water sample in cells derived from *Drosophila melanogaster* (Table 1). The virus failed to replicate in other insect, plant and various vertebrate cells. A relationship between Thirlmere virus and *Drosophila X virus* has been determined based on antigenic, biochemical, and physical properties. The genome of Thirlmere virus has not yet been sequenced.

Virion Structure

Like other members of the family *Birnaviridae*, entomobirnavirus virions are non-enveloped with a capsid made by the viral protein 2 (VP2) and a diameter of about 60–70 nm (Fig. 2). *Drosophila X virus* particles have a density of 1.345 g/ml. The virions contain ribonucleoprotein complexes formed by the genome and multiple copies of the capsid protein.

Genome

Like all birnaviruses, entomobirnaviruses possess a bisegmented, linear, double-stranded RNA genome encoding several viral proteins (VP). Segment A of *Drosophila X virus* is about 3.4 kbp in length and encodes a 114 kDa polypeptide precursor preVP2-VP4-VP3. The VP4 (24 kDa) is a viral serine-alanine protease that releases preVP2 (55 kDa) and VP3 (35 kDa) by cleavage of its own N- and C-termini in the polypeptide (Fig. 3). In *Drosophila X virus* the cleavage sites are located at Ser₅₀₀/Ala₅₀₁ (preVP2 and VP4) and Ser₇₂₃/Ala₇₂₄ (VP4 and VP3), respectively (Fig. 3). The mature VP2 (47 kDa), which represents the viral capsid protein and type-specific antigen, is generated by subsequent serial cleavages of small peptides at the C-terminus of preVP2 most likely via cellular proteases. In mature virus particles VP2 dominates, however progenitor proteins, including preVP2, can also be found. Multiple copies of the nucleoprotein VP3 are associated with the two genome strands forming ribonucleoprotein complexes inside the virion. It has also been shown that VP3 has an RNAi antagonistic function to suppress antiviral RNAi. An additional protein of unknown function (VP5, 27 kDa) is encoded by a small 714 nucleotide, internal open reading frame. In contrast to *Drosophila X virus*, Eridge virus, Espirito Santo virus, and Port Bolivar virus the initiation codon AUG of this open reading frame is mutated to GUG in *Culex* Y virus, *mosquito X virus*, and culicine-associated Z virus (Fig. 3). This open reading frame is likely to be expressed

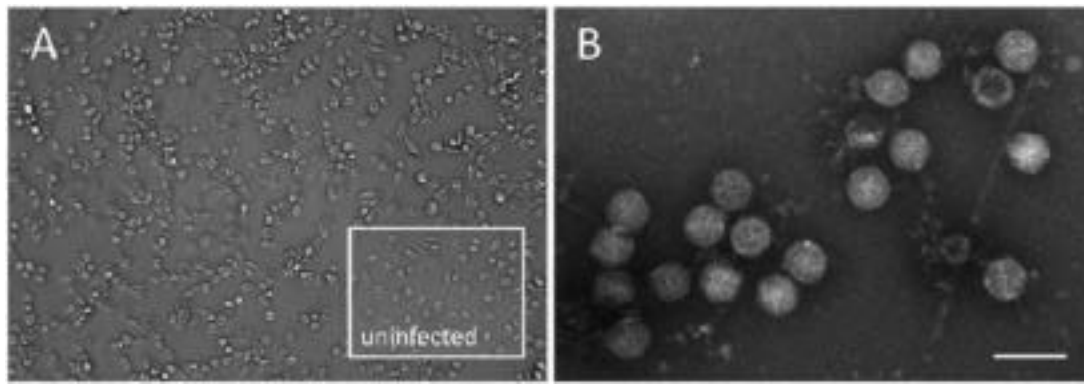


Fig. 2 Infected cells and purified virus particles of Culex Y virus. *Aedes albopictus* cells infected with Culex Y virus 24 h post infection. Cells show cell aggregation as a consequence of the infection with the virus (A). Electron microscopy of purified Culex Y virus particles. Particles were negative stained using uranyl acetate and suggest icosahedral symmetry. Scale bar represents 100 nm (B).

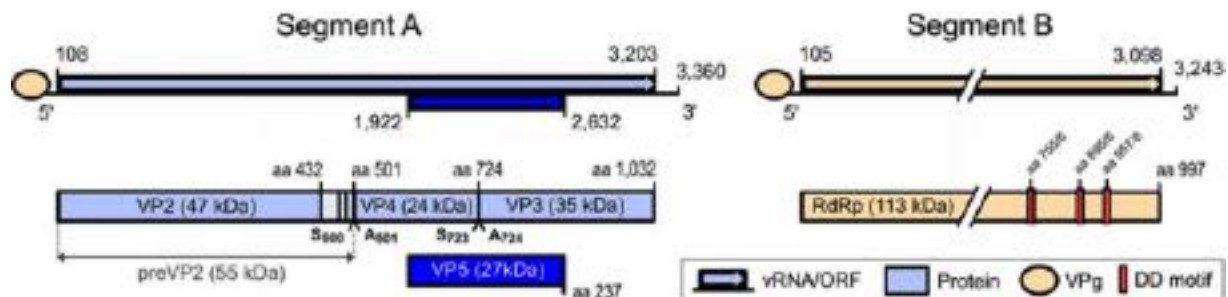


Fig. 3 Schematic representation of entomobirnavirus genome organization. Numbers refer to nucleotide and amino acid sequence position in the genome and proteins of Drosophila X virus. Protease cleavage sites are shown in bold. VPg = genome-linked RNA-dependent RNA polymerase (VP1).

by a -1 ribosomal frame shift that gets facilitated by a particularly shift-prone (“slippery”) sequence (5′-...UUUUUUA...) located 22 nucleotides downstream of the VP5 initiation codon and is conserved among entomobirnaviruses.

Segment B of Drosophila X virus is about 3.2 kbp in length. It encodes the VP1 protein which represents the viral RNA-dependent RNA polymerase. This protein is covalently linked to the 5′-end of the positive-sense genome strand of both segments (VPg), though free forms can be found inside the virion (Fig. 3). Like all RNA viruses, entomobirnaviruses carry highly conserved motifs in their polymerases that are believed to correspond to RNA polymerase functions. However, entomobirnaviruses are lacking the highly conserved Gly-Asp-Asp (GDD) motif that is postulated to form the catalytic site of RNA virus polymerases. Instead, the entomobirnavirus Drosophila X virus RdRp contains three downstream Asp-Asp (DD) residues. They may function as homologs but are located in a more spatially distinct region than the conserved GDD motif in RNA virus polymerases. Due to this the polymerase of entomobirnaviruses may form a distinct subgroup of RNA virus polymerases.

Life Cycle

Knowledge of the replication cycle of entomobirnaviruses is scarce. Transcription and replication are facilitated by the viral RNA-dependent RNA polymerase (VP1). The VP1 protein of Drosophila X virus contains a consensus GTP-binding site and likely has a self-guanlylation activity. Genome replication takes place in the cytoplasm. Translation is performed by the cellular machinery using capped mRNAs that are lacking 3′-poly (A) sequences. Mature virus particles can be found scattered or arranged as crystalline arrays in the cytoplasm.

Epidemiology

Due to their insect-specificity, entomobirnaviruses do not have epidemiological significance. However, entomobirnaviruses have been shown to be present in free-living insects in China, the United States, and different countries in Europe.

Clinical Features

Infections with the entomobirnavirus *Drosophila* X virus have been determined to cause oxygen starvation in *Drosophila melanogaster*. The symptoms occur three to four days before death in dose-dependent *in vivo* experiments. However, these flies and *Drosophila*-derived cell lines can also be persistently infected.

Pathogenesis

Knowledge on the pathogenesis of entomobirnaviruses is scarce. *Drosophila melanogaster* flies infected with *Drosophila* X virus showed a wide distribution of virus particles in the brain, digestive tract, muscles, Malpighian tubules, ovaries, testis, and thorax. Flies that have been examined directly after the onset of anoxia showed that gut and trachea cells, as well as the muscle sheath of different organs were mainly affected. It is speculated that an early invasion of trachea cells by the virus leads to the sensitivity of flies to anoxia.

Experimental infections of various vertebrate and plant cell lines showed no replication of these viruses in non-insect cells. *Culex* Y virus, Espirito Santo virus, and Port Bolivar virus, cause a cytopathic effect (CPE) in different cell lines derived from *Aedes albopictus* (Fig. 2). The CPE is lethal to the cells and they die shortly after being infected with the virus. In contrast, *Drosophila* X virus and Thirlmere virus are not able to infect cells of *Aedes albopictus*.

Further Reading

- Chung, H.K., Kordyban, S., Cameron, L., Dobos, P., 1996. Sequence analysis of the bicistronic *Drosophila* X virus genome segment A and its encoded polypeptide. *Virology* 225 (2), 359–368.
- Cook, S., Chung, B.Y.W., Bass, D., *et al.*, 2013. Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *PLoS One* 8 (11), e80720.
- Franzke, K., Leggewie, M., Sreenu, V.B., *et al.*, 2018. Detection, infection dynamics and small RNA response against *Culex* Y virus in mosquito-derived cells. *Journal of General Virology* 99 (12), 1739–1745.
- Huang, Y., Mi, Z., Zhuang, L., *et al.*, 2013. Presence of entomobirnaviruses in Chinese mosquitoes in the absence of Dengue virus coinfection. *Journal of General Virology* 94 (Pt. 3), 663–667.
- Kelly, D.C., Ayres, M.D., Howard, S.C., *et al.*, 1982. Isolation of a Bisegmented Double-stranded RNA Virus from Thirlmere reservoir. *Journal of General Virology* 62, 313–322.
- L'Heritier, P.H., 1958. The hereditary virus of *Drosophila*. *Advances in Virus Research* 5, 195–245.
- Marklewitz, M., Gloza-Rausch, F., Kurth, A., *et al.*, 2012. First isolation of an entomobirnavirus from free-living insects. *Journal of General Virology* 93 (Pt. 11), 2431–2435.
- Shwed, P.S., Dobos, P., Cameron, L.A., Vakharia, V.N., Duncan, R., 2002. Birnavirus VP1 proteins form a distinct subgroup of RNA-dependent RNA polymerases lacking a GDD motif. *Virology* 296 (2), 241–250.
- Teninges, D., 1979. Protein and RNA composition of the structural components of *Drosophila* X virus. *Journal of General Virology* 45, 641–649.
- Teninges, D., Ohanessian, A., Richard-Molard, C., Contamine, D., 1979. Isolation and biological properties of *Drosophila* X virus. *Journal of General Virology* 42, 241–254.
- Tesh, R.B., Bolling, B.G., Guzman, *et al.*, 2020. Characterization of Port Bolivar virus, a novel entomobirnavirus (*Birnaviridae*) isolated from mosquitoes collected in East Texas, USA. *Viruses* 12 (4), (Epub ahead of print).
- van Cleef, K.W.R., Van Mierlo, J.T., Miesen, P., *et al.*, 2014. Mosquito and *Drosophila* entomobirnaviruses suppress dsRNA- and siRNA-induced RNAi. *Nucleic Acids Research* 42 (13), 8732–8744.
- Vancini, R., Paredes, A., Ribeiro, M., *et al.*, 2012. Espirito Santo virus: A new birnavirus that replicates in insect cells. *Journal of Virology* 86 (5), 2390–2399.
- Webster, C.L., Longdon, B., Lewis, S.H., Obbard, D.J., 2016. Twenty-five new viruses associated with the Drosophilidae (Diptera). *Evolutionary Bioinformatics* 12 (Suppl. 2), 13–25.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsrna-viruses/w/birnaviridae/1017/genus-entomobirnavirus
 Genus: *Entomobirnavirus*.
Birnaviridae.
 dsRNA Viruses.
 ICTV.

Hytrosaviruses (*Hytrosaviridae*)

Henry M Kariithi, Kenya Agricultural and Livestock Research Organization, Nairobi, Kenya

Irene K Meki, French National Center for Scientific Research, Montpellier, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Hytrosa Is derived from two Greek words: “*hypertrophia*” (“excess nourishment”) and “*sialoadenitis*” (“salivary gland inflammation”)

Hytrosaviridae Is a family of entomopathogenic dsDNA viruses that infect dipteran insects

Glossary

Aposymbiotic Is an organism that is functionally devoid of its natural symbiont (mutualist, commensalist, or parasitic).

Corpora allata/cardiac Are a complex of endocrine glands that secrete and store juvenile hormones, which are involved in the regulation of metamorphosis and tissue development in insects.

Horizontal transmission Is the direct (e.g., air-borne, food-borne or venereal) or indirect (e.g., intermediate host vector) transmission of an organism among related or unrelated individuals of the same generation of an ecosystem.

Hyperplasia Is enlargement of an organ or tissue caused by an increased proliferation of cells, which become multi-layered and are capable of replication.

Hypertrophy Is enlargement of an organ or tissue from increased size of cells, which are incapable of dividing.

Hytrosa Is derived from “*hypertrophia sialoadenitis*”, the Greek words meaning “excess nourishment” and “salivary gland inflammation”, respectively.

Peritrophic matrix Is a semipermeable membranous layer that lines the insect gut and is composed of proteins, glycoproteins and chitin microfibrils.

Vertical transmission Is the transovum (egg surface) or transovarian (within the egg) transmission of an organism among individuals of the same generation (e.g., from a mother to its offspring).

Vitellogenesis Is the process of egg yolk formation through deposition of nutrients into oocytes or female germ cells involved in reproduction of lecithotrophic (non-feeding larval development) organisms.

Introduction

Hytrosaviruses (also known as salivary gland hypertrophy viruses or SGHVs; *Hytrosaviridae* family) are rod-shaped, enveloped viruses with large, circular double-stranded DNA (dsDNA) genomes of 120–190 kbp in length. These viruses infect the hematophagous tsetse fly (Diptera: Glossinidae), the filth-feeding common housefly, *Musca domestica* (Diptera: Muscidae) and the phytophagous syrphid fly, *Merodon equestris* (Diptera: Syrphidae). The distinguishing diagnostic feature of SGHV infections is the induction of salivary gland hypertrophy (SGH) syndrome in the salivary glands (SGs) of their adult insect hosts. Observations of a morphologically (rod-shaped viral particles) and symptomologically (occurrence of SGH symptoms) similar virus in the male accessory gland filaments of the solitary braconid wasp, *Diachasmimorpha longicaudata* Ashmead, (Hymenoptera: Braconidae) suggests existence of SGHVs in other insect hosts. The intrinsic properties of SGHVs (i.e., chronic covert infection of adult stages of their hosts), combined with the need to dissect SGs for diagnosis probably hinder the discovery of more members of the *Hytrosaviridae* family.

The SGH symptoms were first reported in the tsetse species *Glossina pallidipes* in the early 1930s in Zululand, South Africa, during a survey of the prevalence of African trypanosomes, the causative agents of Africa's ancient scourge, African trypanosomiasis (sleeping sickness and nagana in humans and animals, respectively) (see [Table 1](#)).

Three decades later (1970s and 1980s), SGH symptoms were noted to occur twice as often in males than in females, to favor development of *Trypanosoma* parasites, and to cause reproductive dysfunctions in the infected flies. The second description of SGH was in 1970s in southern France in 31% and 54% of sampled populations of *M. equestris* var. *nobilis* and *M. equestris* var. *transverlis*, respectively. The third report of SGH symptoms was in the 1990s in adult populations of *M. domestica* at a dairy farm in Florida, USA, during a survey of parasitic nematode infections.

Of particular interest is the negative impacts of SGHV infections in tsetse mass production factories. These factories are required for production of sexually sterile males that are used for tsetse vector control via the environmentally friendly sterile insect technique (SIT), an insects' birth control with a solid track record. For instance, in the 1980s, a *G. pallidipes* colony maintained at the now defunct Kenya Trypanosomiasis Research Institute in Kenya collapsed unexpectedly a few years after its establishment due to poor productivity. Subsequent SGH outbreaks caused collapse of two more *G. pallidipes* colonies at a tsetse production facility in Seibersdorf, Austria (1987 and 2001), and another colony at Kaliti in Ethiopia (in 2012). The poor productivity and eventual collapse of the Seibersdorf colonies jeopardized the implementation of an ambitious SIT-mediated control of *G. pallidipes* from a

Table 1 Chronological history of the discovery and characterization of SGHVs

Year	Milestones in SGHV research
1930–1940	First observation of salivary gland hypertrophy (SGH) syndrome in <i>Glossina pallidipes</i> in Zululand, South Africa. SGH symptoms noted to preferential occur more in males than females
1970–1979	- SGHV reported in other tsetse species (<i>G. morsitans</i> and <i>G. fuscipes</i>). - SGHV reported in two varieties of the narcissus bulb fly (<i>Merodon equestris</i> vr. <i>nobilis</i> and <i>M. equestris</i> vr. <i>transversalis</i>) in southern France.
1980–1989	- SGH reported to be a common diagnostic feature in wild populations of <i>G. pallidipes</i>. - Collapse of a <i>G. pallidipes</i> colony in Kenya attributed to SGHV infections. - Demonstration of <i>per os</i> transmission of infectious GpSGHV particles in tsetse flies. - SGHV shown to reduce insemination rates, fecundity and lifespan of colonized <i>G. pallidipes</i>. - Reports of SGH symptoms in tsetse flies from Zimbabwe and Côte d'Ivoire. - SGHV linked to poor productivity of two <i>G. pallidipes</i> colonies maintained at the Insect Pest Control Laboratory (IPCL) in Seibersdorf, Austria.
1990–1999	- SGHV demonstrated to be transmitted from artificially infected mothers to their offspring. - First description of SGHV cytopathology in tsetse salivary glands. - SGHV infections reported in <i>G. morsitans Swyenatoni</i> and <i>G. brevipalpis</i>. - SGHV reported for the first time in the housefly, <i>Musca domestica</i> L in Florida, USA. - SGHV shown to infect milk glands, mid-gut and male accessory glands of <i>G. pallidipes</i>.
2000–2009	- Collapse of an Ethiopian-derived <i>G. pallidipes</i> colony at IPCL, Seibersdorf, Austria. - SGHV shown to infect male accessory glands of <i>G. m. morsitans</i> Westwood. - Full sequencing of genomes of GpSGHV and MdSGHV. - GpSGHV and MdSGHV proposed and established in a new virus family, <i>Hytrosaviridae</i>. - Transcription analysis of MdSGHV.
2010–2018	- Description of the full proteome, ultra-structural features and morphogenesis of GpSGHV. - Demonstration of global distribution of MdSGHV in <i>M. domestica</i> populations. - SGHV-like virus described in accessory gland filaments of the solitary braconid wasp <i>Diachasmimorpha longicuada</i>. - Antiviral drugs (valacyclovir/acyclovir) demonstrated to interrupt GpSGHV replication in <i>G. pallidipes</i>. - Demonstration that microbiota influence vertical transmission of GpSGHV in <i>G. pallidipes</i>. - Successful management of GpSGHV infections and complete eradication of overt SGH symptoms in <i>G. pallidipes</i> colonies using combination of antiviral drugs and modified feeding regimes. - GpSGHV shown to modulate different molecular pathways in <i>G. pallidipes</i> and <i>G. m. morsitans</i>. - Full sequencing of genome a GpSGHV strain from Ethiopian tsetse flies. - The first evidence that microRNAs (miRNA) and RNA interference (RNAi) are involved in GpSGHV infections in <i>G. pallidipes</i>.

25,000 square kilometers of under-utilized fertile land in the Southern Rift Valley of Ethiopia. The Ethiopian SIT project was to be a model demonstrating that tsetse flies can be eradicated from the mainland Africa, after the hailed successful eradication of *G. austeni* populations on Zanzibar Island. These turns of events precipitated into concerted research efforts to characterize the tsetse virus and develop integrated mitigation strategies to salvage tsetse mass production. The research also included the housefly-infecting virus, which was deemed as a potential risk in the magmeal production industry and as a biocontrol agent due to its ability to suppress ovarian development in the insect.

To date, genomes of two strains of SGHVs from different tsetse fly populations, and one strain from the housefly have been fully sequenced.

Taxonomy and Classification

When first discovered, the tsetse SGHVs were described as “virus-like particles” due to their morphological resemblance to viruses that had already been described in other organisms such as the *Drosophila*, mosquito, and nematodes. The SGHVs were erroneously presumed to be arboviruses (based on their morphological similarities to other viruses that replicate in the SGs and are transmitted by hematophagous insects such as mosquitoes, ticks, sand flies and gnats), or nudiviruses (based on their sizes and enveloped rod-shaped morphology). However, full genome sequencing data of SGHVs in 2008 did not support the placement of these viruses within any of the established families of invertebrate viruses. The tsetse fly and housefly SGHVs showed limited but significant gene homologies, and were subsequently classified into separate genera within a new *Hytrosaviridae* family. *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) is classified in the type species *Glossina hytrosavirus* of the genus *Glossinavirus* and *Musca domestica* salivary gland hypertrophy virus (MdSGHV) is classified in the species *Musca hytrosavirus* of the genus *Muscavirus*. The name of the virus family (Hytrosa) was derived from the Greek words *Hypertrophia* and *sialoadenitis*, meaning “excess nourishment” and “salivary gland inflammation”, respectively. Compared to GpSGHV, MdSGHV is more structurally and symptomologically similar to the virus that induces SGH symptoms in *M. equestris* (MeSGHV).

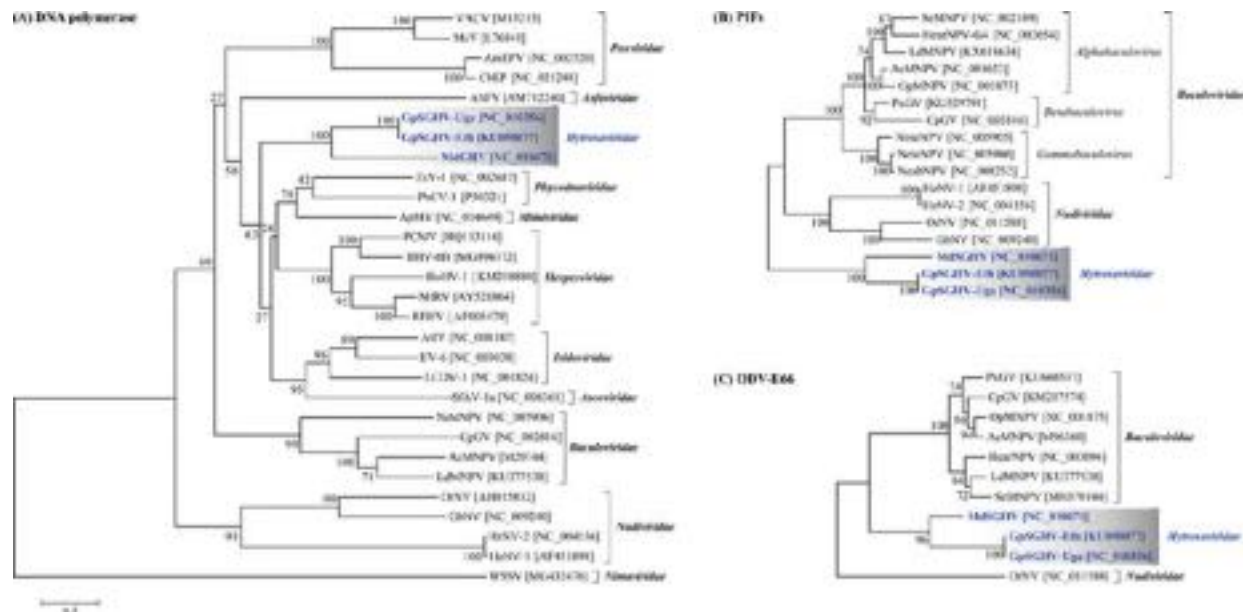


Fig. 1 Maximum likelihood (ML) phylogenetic trees of conserved hytrosaviruses (SGHVs) genes. Based on DNA polymerase, SGHV relate more closely with linear dsDNA viruses (e.g., *Herpesviridae*, *Iridoviridae* and *Phycodnaviridae*) than circular dsDNA viruses (A). Panels B and C show ML trees of concatenated sequences of *per os* infectivity factors (P74 or PIF-0, PIF-1, PIF-2 and PIF-3) and ODV-E66, respectively. SGHV PIFs branched more distantly from nudiviruses than from baculoviruses. GenBank accession numbers of the viruses are indicated in tree branches (square brackets). *Virus abbreviations*: ASFV, African swine fever virus; GpSGHV-Uga/Eth, *Glossina pallidipes* SGHV Uganda/Ethiopia strain; MdSGHV, *Musca domestica* SGHV; HHV-6B, Human betaherpesvirus 6B; PCMV, Porcine cytomegalovirus; BoHV-1, Bovine herpesvirus 1; RFHV, Retroperitoneal fibromatosis-associated herpesvirus; MfRV, *Macaca fuscata* rhadinovirus; ApMV, *Acanthamoeba polyphaga* mimivirus; PbCV-1, *Paramecium bursaria* Chlorella virus 1; EsV-1, *Ectocarpus siliculosus* virus 1; SfAV-1a, *Spodoptera* ascovirus 1a; LDV-1, Lymphocystis disease virus 1; IIV-6, Invertebrate iridescent virus 6; AtIV, *Aedes taeniorhynchus* iridescent virus; AmEPV, *Amsacta moorei* entomopoxvirus L; ChEP, *Choristoneura biennis* entomopoxvirus L; VACV, *Vaccinia* virus; McV, *Molluscum contagiosum* virus; WSSV, Shrimp white spot syndrome virus; NeleNPV, *Neodiprion lecontei* nucleopolyhedrovirus; NeabNPV, *Neodiprion abietis* nucleopolyhedrovirus; HearNPV, *Helicoverpa armigera* nucleopolyhedrovirus; CpGV, *Cydia pomonella* granulovirus; PxGV, *Plutella xylostella* granulovirus; AcMNPV, *Autographa californica* multiple nucleopolyhedrovirus; LdMNPV, *Lymantria dispar* multiple nucleopolyhedrovirus; SeMNPV, *Spodoptera exigua* multiple nucleopolyhedrovirus; OpMNPV, *Orygia pseudotsugata* multiple nucleopolyhedrovirus; OrNV, *Oryctes rhinoceros* nudivirus; HznV-2, *Helicoverpa zea* nudivirus 2; HznV-1, *Heliothis zea* nudivirus 1; GbNV, *Gryllus bimaculatus* nudivirus; PmNV, *Penaeus monodon* nudivirus; HgNV, *Homarus gammarus* nudivirus.

Similarities With Other Virus Taxa

GpSGHV and MdSGHV are structurally similar to members of other arthropod-infecting virus families such as *Baculoviridae*, *Nudiviridae*, and *Nimaviridae*. The SGHVs share 12 of the 38 core genes that have been described in baculoviruses, nudiviruses, nimaviruses, and some bracoviruses. Among the structural and genomic features shared by SGHVs and other large dsDNA viruses include possession of enveloped, rod-shaped virions, circular dsDNA genomes and their replication in the nucleus of infected insect cells. However, the SGHVs functionally differ from viruses such as the baculoviruses in the lack of occlusion bodies and lower lethality (i.e., SGHVs rarely kill their insect host).

Based on the DNA polymerase II (B-family; polB) gene, which is present and conserved in all large dsDNA viruses, SGHVs relate more closely with invertebrate large linear dsDNA viruses compared to circular dsDNA viruses. The linear dsDNA viruses that cluster together with SGHVs include members of families *Herpesviridae* (120–240 kb), *Iridoviridae* (140–303 kb), *Poxviridae* (130–375 kb), *Phycodnaviridae* (100–560 kb), and *Mimiviridae* (1200 kb) (Fig. 1(A)). At the amino acid level, the polB of GpSGHV and MdSGHV are 31% identical, and their best matches are to the alcelaphine herpesvirus polB (24.9% and 22.6% identities, respectively).

The most significant similarity between the SGHVs and other large invertebrate dsDNA viruses is the possession of genes encoding four of the core and highly conserved *per os* infectivity factor proteins (PIFs 0 or P74, 1, 2, and 3) (Fig. 1(B)). The PIFs are essential for entry of baculoviruses and nudiviruses in their host's midgut cells. At the amino acid level, the GpSGHV and MdSGHV PIF homologs are closely related (31%–39% identical) and appear to have evolutionarily diverged more distantly from nudiviruses than from baculoviruses. The genomes of SGHVs also encode homologs to a late expression protein, and a major structural component of occlusion-derived virus (ODV) envelopes of lepidopteran baculoviruses (ODV-E66 protein). ODV-E66 is a chondroitinase that plays roles during baculovirus infections in the midguts of lepidopteran larvae. The ODV-E66 of SGHVs cluster together with their homologs in baculoviruses (Fig. 1(C)), but its role during the SGHV infection has not been established.

In addition to polB and PIFs, the two SGHVs encode homologs to four of the six subunits of the DNA-dependent RNA polymerase (DdRp) complex reported in baculoviruses and nudiviruses. DNA viruses use the DdRp to transcribe their mRNAs. The DdRp complex subunits present in SGHVs include the late expression factors 4, 5, 8, and 9 (LEF-4, LEF-5, LEF-8, and LEF-9). The multifunctional LEF-4 is an mRNA capping enzyme, LEF-5 is a transcription initiation factor, and LEF-8 plays critical roles in the activation of the late gene promoter, while LEF-9 forms a part of the catalytic center of the DdRp complex. It is notable that whereas LEF-9 is encoded by a single open reading frame (ORF) in MdSGHV, baculoviruses and nudiviruses, GpSGHV has two adjacent ORFs coding for the N- and C-terminals of this enzyme. Both GpSGHV and MdSGHV lack homologs to the remaining two of the six DdRp subunits, i.e., P47 and very late expression factor 1 (VLF-1) proteins, which play roles in systemic viral infections, and capsid assembly/viral DNA packaging, respectively.

Taken together, the presence of the above-mentioned gene homologs in the SGHVs suggests their common ancestry and similar modes/mechanisms of entry into the host cells and gene expression.

Virion Structure

The MdSGHV and GpSGHV virions are 50–65 nm in diameter, non-occluded, non-icosahedral rod-shaped particles measuring 500–1000 nm in length, and contain a bilayer lipid envelope. The SGHV particles have a density of 1.153 g cm^{-3} on a 10%–60% Nycodenz gradient. There are however distinct morphological differences between the SGHV virions. The rigid and rod-shaped GpSGHV virions are longer (50–65 nm $\times$ 1000 nm) compared to those of MdSGHV (65 nm $\times$ 550 nm). The enveloped virions of the unclassified MeSGHV measure 50–60 nm in diameter and 500–600 nm in length. The GpSGHV virions are asymmetrical with a rounded and a conical end, and consist of a thin, central, electron-dense helical nucleocapsid core (40 nm in diameter). The nucleocapsid core is surrounded by a 10 nm-thick proteinaceous tegument matrix, which is encapsulated by an outer lipid bilayer envelope. The outer surface of the virion is studded with left-handed helical polymeric spikes (13 nm long with 15 nm periodicity) composed of viral and host-derived protein dimers (23 spikes $\times$ 24 helical turns = 1104 envelop dimers). GpSGHV particles are more fragile (the viral envelopes are readily depleted by treatment with common buffers), which partially accounts for their low infectivity compared to the MdSGHV particles. The MdSGHV virions have regularly spaced braided, bead-like surface topology and are rounded on both ends.

Genome Organization

Detection of relaxed and supercoiled bands after gel electrophoresis of purified viral DNA, and the absence of end-labeling of undigested DNA suggested that SGHV genomes were circular dsDNA molecules. Currently, two strains of GpSGHV isolated from Ugandan (GpSGHV-Uga; GenBank Acc. no. EF568108) and Ethiopian *G. pallidipes* (GpSGHV-Eth; GenBank Acc. no. KU050077.1) have been fully sequenced. The organization and the main features of the genome of GpSGHV-Eth compared to that of GpSGHV-Uga are schematically represented in Fig. 2. The conserved regions of GpSGHV-Uga and GpSGHV-Eth genomes are 98.1% similar at the nucleotide level. However, compared to the GpSGHV-Uga, the GpSGHV-Eth genome has insertions and deletions in 17 and 20 ORFs, respectively, while 11 ORFs are deleted and 24 ORFs are novel. Only one MdSGHV strain has so far been fully sequenced (MdSGHV GenBank Acc. no. NC_010671.1), but other uncharacterized isolates of the virus have been detected in housefly populations sampled from various geographical locations. The MdSGHV genome presumably has similar genome organization as that of GpSGHV.

The GpSGHV-Uga, GpSGHV-Eth and MdSGHV differ in their genome sizes (190,032 bp, 190,291 bp and 124,279 bp, respectively), and G + C contents (28.0%, 27.9% and 43.5%, respectively). Approximately 2.8% and 1.7% of the GpSGHV and MdSGHV genomes, respectively, harbor tandem direct repeat sequences (*drs*). The numbers and sizes of the *drs* differ among the SGHV strains: 14 *drs* (52–147 bp) in GpSGHV-Eth, 15 *drs* (52–246 bp) in GpSGHV-Uga, and 18 *drs* (30–380 bp) in MdSGHV. The *drs* are clustered in some genomic regions, but they are less clustered in MdSGHV than in GpSGHV. Three GpSGHV-Eth *drs* (dr9, 11, and 13) and seven MdSGHV *drs* (dr5, 6, 7, 10, 16, 17, and 18) are localized within coding regions of specific ORFs. All the GpSGHV-Uga *drs* are localized in noncoding regions. In baculoviruses, similar repeat regions are usually located between genes, and are hypothesized to influence viral replication (as transcription enhancers and/or origins of DNA replication), evolution, and pathobiology. The SGHVs lack the palindromic homologous repeat (*hrs*) regions found in other invertebrate circular dsDNA viruses such as baculoviruses, nudiviruses and nimaviruses.

Most of the ORFs in the genomes of GpSGHV harbor the TATA-like box promoter elements (89.7%), poly(A) signals (83.9%), and late transcriptional initiation motifs (T/G/A)TAAG (35.5%). The transcriptional elements in the MdSGHV genome are largely unknown, but most of its ORFs are enriched with the TAAG motifs, poly(A) signals and TATA-like box promoter elements. Notably, 13 GpSGHV-Eth genes are non-canonical, five of which are homologs to known regulatory genes, including the *MAL7P1.132* gene, *lef-9*, thymidylate synthase (*ts*), cluster of differentiation antigen protein-48 (*cd-48*) and metalloendopeptidase (*mp-nase*).

Modes of Infection and Gene Expression

The mechanism(s) of SGHVs life cycle within their insect hosts is unclear. Proteomic and transcriptomic studies have however provided evidence for the expression of *pif* genes, but it is unclear whether PIFs are involved in the virus initial attachment and

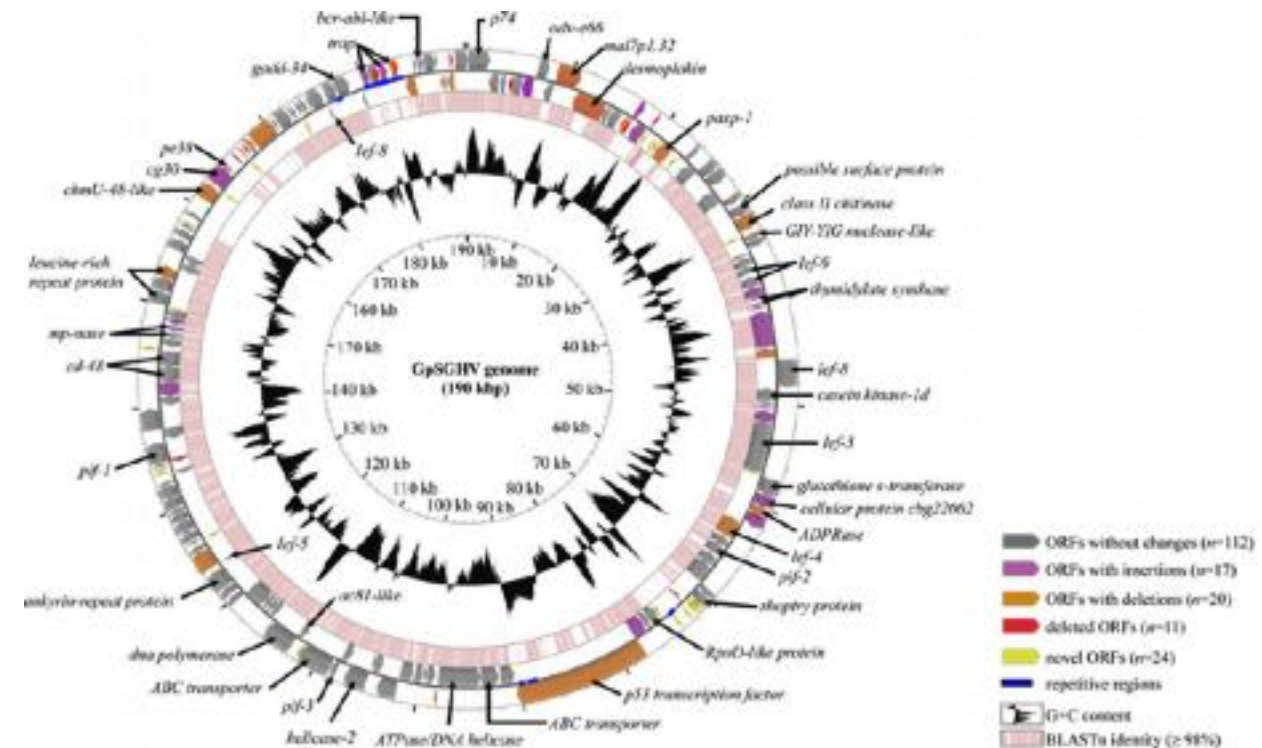


Fig. 2 Circular representation of the GpSGHV-Eth genome and its comparison (percentage identity) to the GpSGHV-Uga genome. The open reading frames (ORFs) are presented starting with the ATG initiating codon of the *p74* gene (coded by SGHV-Eth001). The outermost two concentric rings and the arrows represent the positions and orientations of the transcription potential of each ORF. BLASTn comparison (98.1% identity at the nucleotide level) between the two SGHV genomes is indicated by the third ring (from outside). The black peaks represent the deviation from the average % G + C content of the entire genome. ORFs are color-coded based on their homologies between the two genomes as shown in the figure key. The names of the annotated ORFs are indicated. Unlabeled ORFs (light gray color) represent hypothetical/uncharacterized ORFs. The figure is drawn to scale. *Gene abbreviations:* *odv-e66*, occlusion-derived virus envelope protein 66; *paxp-1*, paired box protein Pax-1; *lef*, late expression factor; *ADPRase*, ADP-ribose pyrophosphatase; *pif*, per os infectivity factor; *RpoD*, RNA polymerase sigma (σ) factor; *cd-48*, Cell division cycle protein 48; *chmU*, chlacomycin gene U; *cg-30*, Zinc finger protein CG30; *pe38*, Major immediate early protein 38; *gadd-34*, growth arrest and DNA damage protein 34; *trap*, trp RNA-binding attenuation protein; *bcr-abl*, an oncogene arising from the fusion of the breakpoint cluster (*bcr*) gene with chromosomal Abelson murine leukemia (*c-abl*) proto-oncogene; *ABC transporter*, ATP-binding cassette transporter.

infection of the host's midgut epithelial cells. Available data suggest low frequency of *per os* infection, most likely due to the peritrophic matrix (PM) barrier that may reduce the penetration of viral particles into the hemocoels. Following initial oral infection, SGHVs exploit the host's tracheal system as a conduit to breach the basal laminal barrier in the midgut to access the SGs, the primary tissue of viral replication. Viral capsids are released into the cell cytoplasm via fusion of viral and host's vesicle membranes, and then trafficked (via the host's microtubule motor complexes) to the cell nucleus for viral gene transcription, viral DNA genome replication, packaging and assembly of progeny nucleocapsids in the virogenic stroma.

Based on the complexity and significant homologies of SGHV-expressed proteins to well-annotated proteins in other viruses, the current school of thought is that the viral replication involves sequential expression of the immediate early (transcription factors), early (DNA replication genes), and late (viral replication proteins) genes. However, a lack of a suitable cell culture to support SGHV replication has so far hindered the elucidation of the precise transcriptional timings of the SGHV genes. Transmission electron microscopy (TEM) has provided evidence that the SGHV nucleocapsids exit the nucleus of infected SG cells via the nuclear pore complex, and subsequently acquire their glycoprotein-containing envelopes in the cell cytoplasm, presumably via the ER-Golgi system. After maturation, the enveloped viral particles egress by migrating to and budding out of the cell membranes bordering the SG lumens (for MdSGHV) or lysis of the luminal cell membranes (for GpSGHV).

Major Structural Proteins

Proteomic studies have revealed that the SGHV structural complex comprise of a blend of virally-encoded and host-derived proteins, with the nucleocapsid and envelope sections of the GpSGHV and MdSGHV virions composed of at least 45 and 29 structural proteins, respectively. Of the GpSGHV structural proteins, 10, 15, and 20 proteins are localized in the envelope, nucleocapsid, and tegument matrix, respectively. The structural proteins in GpSGHV and MdSGHV are almost identical, most of

Table 2 Twenty-nine homologous structural proteins of SGHVs and other dsDNA viruses

<i>GpSGHV proteins^a</i>		<i>MdSGHV homolog</i>	<i>Homologs in other dsDNA viruses</i>	<i>Functional annotation</i>
<i>GpSGHV</i>	<i>GpSGHV-Eth</i>			
<i>Homologs to nucleocapsid structural proteins</i>				
SGHV103	Viral capsid associated-like protein (GpSGHVEth113)	–	Viral capsid-associated protein – 1054 <i>NeabNPV</i> (VP1054)	Viral DNA encapsidation.
SGHV062	P53 transcription factor-like protein (GpSGHVEth069)	–	Smc protein <i>TnIAV</i> – 2c (ORF147)	Chromosomal structural positioning
SGHV083	GpSGHVEth091	MdSGHV013	<i>AmEPV</i> (AMV214)	–
SGHV052	Nucleocapsid protein (GpSGHVEth053)	–	Major core protein P4a (A10L gene) <i>MsEPV</i> (MSV152)	Viral DNA assembly.
Cell division-cycle protein 48 like protein (cd – 48) (SGHV107/108)	cd – 48 (GpSGHVEth117/118)	vacuolar sorting ATPase (MdSGHV033)	LCDV CDC48 protein (ORF 209R)	Proteins export to ubiquitin-proteasome complex during late viral infection.
SGHV154	Nucleocapsid protein (GpSGHVEth169)	–	<i>Wiseana</i> iridovirus ORF026	–
<i>Homologs to tegument structural proteins</i>				
SGHV010	Desmoplakin (GpSGHVEth009)	–	<i>MsEPV</i> desmoplakin (Ac66)	Nucleocapsid egress from virogenic stroma.
ABC transporter (SGHV064)	ABC transporter (GpSGHVEth071)	–	<i>AmEPV</i> ABC transporter (ORF AMV130)	Multifunctional protein.
SGHV071	putative capsid protein 3 (GpSGHVEth078)	MdSGHV090	<i>ApMV</i> capsid protein 3	–
SGHV086	Tegument protein (GpSGHVEth094)	–	Megavirus chileensis ankyrin repeat protein (ORF398)	Hijacking of the host's cell ubiquitin machinery.
SGHV110/111	Metalloproteinase (<i>mp</i> -nase) – 38 (GpSGHVEth121/122)	matrix metalloproteinase (MdSGHV036)	<i>Spl</i> /GV <i>mp</i> -nase – 38	Disrupts plasma membrane to initiate infection.
<i>Homologs to envelop proteins</i>				
P74 (SGHV001)	P74 (GpSGHVEth001)	P74 (MdSGHV039)	<i>Spl</i> /NPV P74	Essential for <i>per os</i> infection.
PIF – 1 (SGHV102)	PIF – 1 (GpSGHVEth112)	PIF – 1 (MdSGHV029)	<i>NeabNPV</i> PIF – 1	
PIF – 2 (SGHV053)	PIF – 2 (GpSGHVEth054)	PIF – 2 (MdSGHV089)	<i>AmFV</i> PIF – 2	
PIF – 3 (SGHV076)	PIF – 3 (GpSGHVEth083)	PIF – 3 (MdSGHV106)	<i>EupSNPV</i> PIF – 3	Virus replication and late gene expression. Required for correct virus assembly.
SGHV045	LEF – 3 (GpSGHVEth046)	MdSGHV083	<i>Mav</i> /MNPV LEF – 3	
SGHV072	Thiol oxidase (GpSGHVEth079)	MdSGHV102	ASFV FAD-dependent thiol oxidase (ORF pB119L)	
SGHV088	Metalloprotease (GpSGHVEth096)	MdSGHV017	<i>AmEPV</i> vaccinia G1L metalloprotease (AMV258)	Morphogenesis of infectious virions.
<i>Homologs to infected cell-specific viral proteins (ICSVs)</i>				
Class-II chitinase (SGHV027)	Class-II chitinase (GpSGHVEth027)	–	<i>PrGV</i> type II chitinase (ORF010)	Promote terminal host liquefaction.
Core-like protein (SGHV051)	LEF – 4 (GpSGHVEth052)	MdSGHV087	<i>GbNV</i> LEF – 4 (Ac90)	Promoter recognition and stabilization of late and very late viral transcripts.
SGHV040	LEF – 8 (GpSGHVEth041)	MdSGHV070	<i>Spl</i> /NPV LEF – 8 (Ac50)	Viral DNA replication.
SGHV032/33	LEF – 9 (GpSGHVEth033/34)	MdSGHV074	<i>GbNV</i> LEF – 9 (ORF24)	
DNA polymerase (SGHV079)	DNA polymerase (GpSGHVEth087)	DNA polymerase (MdSGHV001)	<i>AmEPV</i> RNA polymerase TF (AMV054)	

(Continued)

Table 2 Continued

<i>GpSGHV proteins^a</i>		<i>MdSGHV homolog</i>	<i>Homologs in other dsDNA viruses</i>	<i>Functional annotation</i>
<i>GpSGHV</i>	<i>GpSGHV-Eth</i>			
SGHV078	Ac81-like (GpSGHVEth086)	Ac81-like (MdSGHV108)	OrNV Ac81-like protein (ORF004)	Virus interaction with host cell cytoskeleton.
Thymidylate synthase (TS) (SGHV035/36)	TS (GpSGHVEth036/37)	TS (MdSGHV012)	MsEPV TS (ORF MSV238)	<i>de novo</i> nucleotide biosynthesis.

^aHypothetical (non-annotated) proteins are indicated by their respective ORF numbers in the viral genomes.

Abbreviations: - PIF, *per os* infectivity factor; LEF, late expression factor; *EupSNPV*, *Euproctis pseudoconspersa* single nucleopolyhedrovirus; *AmFV*, *Apis mellifera* filamentous virus; *MaviMNPV*, *Maruca vitrata* multi-nucleopolyhedrovirus; *SpliNPV*, *Spodoptera litura* nucleopolyhedrovirus; ASFV, African swine fever virus; *MsEPV*, *Melanoplus sanguinipes* entomopoxvirus; *GbNV*, *Gryllus bimaculatus* nudivirus; *NeabNPV*, *Neodiprion abietis* nucleopolyhedrovirus; *OrNV*, *Oryctes rhinoceros* nudivirus; *AmEPV*, *Amsacta moorei* entomopoxvirus; *PrGV*, *Pieris rapae* granulosis virus; *TnIAV-2c*, *Trichoplusia ni* ascovirus 2c; SMC, structural maintenance of chromosomes protein; BP-NLS-Bipartite nuclear localization signal; LCDV, Lymphocystis disease virus; *ApMV*, *Acanthamoeba polyphaga* mimivirus.

which are homologs to structural proteins described in several insect viruses. The major structural proteins of GpSGHV, their homologs in other invertebrate dsDNA viruses and their functional annotations are summarized in **Table 2**.

As shown in **Table 2**, the most notable GpSGHV nucleocapsid structural proteins include homologs involved in viral DNA assembly and encapsidation, chromosomal structural positioning, and protein turnover during late viral infection. Amongst the homologous structural tegument proteins include those involved in the disruption of the plasma membrane to initiate infection, nucleocapsid egress from virogenic stroma, and hijacking of the host cell ubiquitin machinery. Notable envelope structural proteins include homologs involved in oral infection, viral replication and late gene expression, correct virus assembly, and morphogenesis of infectious virions. Additionally, there are proteins that could not be associated with the structural (nucleocapsid, tegument, envelope) components of GpSGHV, and were assigned as “infected cell-specific viral proteins” (ICSVs). Notable ICSVs include homologs to proteins essential for viral DNA replication, late and very late viral transcription, nucleocapsid envelopment, and terminal host liquefaction (**Table 2**). Other than the virally-encoded proteins, a mature GpSGHV particle contain at least 50 host-derived cellular proteins, almost 50% of which are tegument proteins. At least 13 of the host-derived cellular proteins are potentially incorporated into the virions, which may serve specific and auxiliary roles in viral morphogenesis. Although there is no evidence for the presence of cellular-derived host proteins in MdSGHV, the derivation of the viral envelope during the cytoplasmic envelopment process implies incorporation of a complex of host proteins and lipids.

Host Range

Tsetse fly is the only known natural host for GpSGHV, in which the virus predominantly causes chronic asymptomatic (covert) infections. GpSGHV is highly specific to tsetse species, and there is no evidence for its potential infection or replication of heterologous hosts such as the housefly. The susceptibility of tsetse flies to GpSGHV infections differ widely amongst different tsetse species, of which *G. pallidipes* is the most susceptible. There is evidence for the existence of at least 15 GpSGHV haplotypes whose prevalence differs spatially and genetically among different wild populations of tsetse flies in East, Central, and West Africa.

The common housefly (*M. domestica* L.) is the natural host for MdSGHV in which the virus causes only acute symptomatic (overt SGH symptoms) infections. Under laboratory conditions, MdSGHV can infect other insects such as the obligate hematophagous stable fly (*Stomoxys calcitrans*), the autumn housefly (*Musca autumnalis*), and a larval predator of the housefly, the black dump fly (*Hydrotaea aeneascens*). Although MdSGHV is incapable of inducing the diagnostic overt SGH symptoms in other insects, the virus does significantly affect ovarian development and fly mortality in the stable fly and black dump fly. These observations suggest that MdSGHV may reside asymptotically in other muscids as alternative or reservoir hosts. It is however unknown whether these alternative hosts can transmit infectious viral particles to healthy conspecifics under field/natural conditions. MdSGHV also fails to infect or replicate in the tsetse flies when virus suspensions are injected into the flies.

Pathogenesis and Tissue Tropism

All SGHVs induce the same gross pathology (i.e., SGH symptoms) in their respective host insects (adult stages), but the cytopathologies are distinct for each genus (**Fig. 3**). The SG pair are equally affected, swollen up to four times or higher than their normal size, and the enlargement typically covers the entire length of the distal regions of the glands. The precise mechanism(s) underlying the development of SGH symptoms are unknown. In tsetse flies, GpSGHV significantly alters the synthesis of host proteins, notable of which are proteins associated with blood feeding, immunity, cellular proliferation, homeostasis, cytoskeletal traffic and regulation of protein turnover.

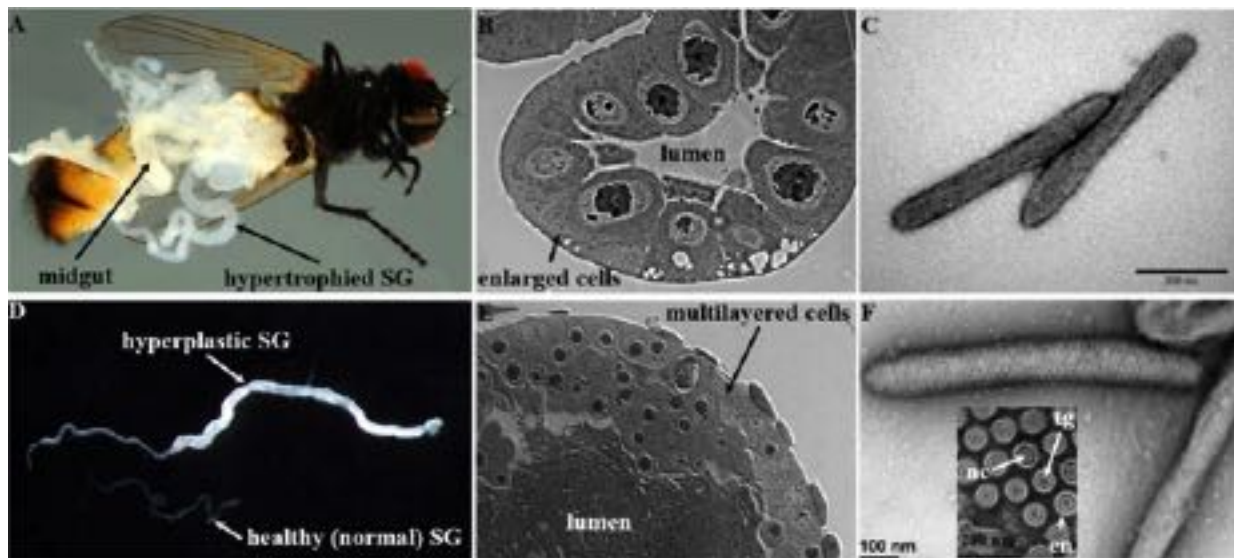


Fig. 3 Pathogenesis and the main structural/morphological features of SGHVs. The infected salivary glands (SGs) are swollen up to four times or higher than their normal sizes (A and D). MdSGHV induces hypertrophy (i.e., enlargement of both cytoplasm and nucleus of SG cells) (B), while GpSGHV induces hyperplasia (i.e., enlarged, multi-layered cytoplasm of SG cells; the nucleus is not enlarged) (E). The MdSGHV and GpSGHV particles are shown in panels C and F, respectively. The nucleocapsid core (nc), tegument (tg) and envelope (en) structural components of GpSGHV particle are shown in the inset of panel F. Scale bar in C is 300 nm. Adapted from. Kariithi, H.M., Meki, I.K., Boucias, D.G., Abd-Alla, A.M.M., 2017. Hytrosaviruses: Current status and perspective. *Current Opinion in Insect Science* 22, 71–78. doi:10.1016/j.cois.2017.05.009. (with permission).

Infections of non-SG tissues by the SGHVs is associated with various pathological effects including reproductive dysfunctions, infertility in the females and distorted mating behaviors. These effects are more apparent in flies that exhibit overt SGH symptoms, and the characteristics of the pathologies are distinct for each genus (see below).

Pathogenesis in the Salivary Glands

In the tsetse fly, GpSGHV causes SGs hyperplasia, whereby the cytoplasmic (but not nuclear) compartments of the infected cells are enlarged. The SG cells become multi-layered (Fig. 3) and are capable of replicating. The cellular proliferation is thought to be part of the host's responses to GpSGHV infections, which result from virus-induced reprogramming of differentiated SG cells. Notably, only *G. pallidipes* exhibits overt SGH symptoms, and even in this species, overt SGH symptoms is an exception rather than the rule. Other tsetse species harbor only chronic asymptomatic GpSGHV infections, without any notable host's fitness cost (development and survival). The mechanism(s) underlying the switch from asymptomatic to symptomatic infection states are currently unknown, but it is suspected that certain host's and/or viral factors, and host's microbiota communities influence development of overt SGH. When adult flies are artificially (hemocoelic) injected with GpSGHV suspensions, overt SGH symptoms do not develop in the same parental generations, rather the symptoms develop in the F₁ progenies produced by the virus-injected mothers.

In the housefly, MdSGHV causes SG hypertrophy whereby both the cytoplasmic and nuclear compartments of the infected gland cells proliferate, but the SG cells are incapable of dividing. Unlike in the case of GpSGHV, within 3 days post infection, MdSGHV induces overt SGH symptoms in 100% of the infected individuals (i.e., the virus does not infect asymptotically). Additionally, older adult houseflies show increased resistance to MdSGHV infection, which is partially attributed to the maturation of the PM barrier in older flies.

Pathogenesis in Other Host Tissues

In tsetse flies, overt SGH symptoms are associated with testicular degeneration, ovarian abnormalities, severe necrosis, and degeneration of germaria, reduction of the development, survival, fertility and fecundity of the flies. The testicular degeneration is characterized by vacuolation of the follicles, thus causing complete shutdown of spermatogenesis. GpSGHV infection also causes disintegration of the accessory reproductive glands of males. Virus infection of tsetse fly tracheal cells and milk glands may cause extensive hypertrophy, but without hyperplasia. Infected milk glands become necrotic, and the reservoir organelles that store the milk secretions are depleted. Virus-induced midgut cellular necrosis impair nutrient assimilation, which, combined with the SG

pathologies, results in starvation of infected flies. A combination of these pathological effects of the virus largely accounts for reduce productivity and collapse of tsetse fly colonies.

In the housefly, MdSGHV replicates rapidly, resulting in complete shutdown of vitellogenesis via blocking the production of sesquiterpenoids. The replication of MdSGHV in the cells of corpora-allata and corpora-cardiaca partially accounts for a shutdown of vitellogenesis. In addition, the ovaries of viremic housefly females are arrested at a pre-vitellogenic stage. MdSGHV infection also alters housefly-mating behaviors, whereby viremic females refuse to copulate when paired with healthy males and viremic males show reduced avidity to initiate courtship with healthy females. Examination of various tissues obtained from flies with overt SGH symptoms under TEM provided evidence for presence of viral particles in multiple other tissues including the crop, midgut lumens, and ovarioles.

Viral Latency

One of the outstanding questions that has emerged in the course of SGHVs research is how GpSGHV is maintained at an asymptomatic state in the tsetse host. Some of the proposed hypotheses include: (1) that the virus may be integrated in the host chromosomes; or (2) may occur as an episome in the nuclei; or (3) persist at a low-level, covert infection in the tsetse fly organs. In all these cases, it appears that GpSGHV is 'under the control' of the host and the question is what type of control is exerted. Southern blot analysis using selected GpSGHV genes as probes did not provide any evidence for integration of SGHVs into host genomes as proviruses.

The predominant asymptomatic GpSGHV infection state is thought to represent either a sub-lethal persistent infection state, or viral latency. During persistent infections, a virus remains in specific cells, whereby progeny virions are perpetually produced at low-levels without excessive damage to the host's cells. During latency, viral genome copies and proteins are present in infected cells for a period of time, but without detectable formation of infectious viral particles. Additionally, depending on the tissue tropism, a virus can cause both persistent and latent infections in the same host concurrently, but in different tissues. Based on its pathobiology during asymptomatic infections, GpSGHV is thought to exist in both persistent and latent infection states at the same time. A persistent infection state in the tsetse SG tissue is supported by the release of low amounts of virus particles ($\sim 10^2$ viral genome copies) via saliva during feeding by an asymptomatic fly. In this state, low virus replication levels does not result in the development of overt SGH symptoms. At the same time, the virus may latently infect non-SG tissues such as the tracheal cells, in which viral DNA is detectable but no transcripts. During viral latency or persistence, only minimal numbers of viral genes are expressed to evade the host's immune system, and the virus does not induce overt SGH symptoms or reproductive dysfunctions as observed during the symptomatic infection state.

A specific RNA interference (RNAi) is one of the possibilities of how GpSGHV latency is maintained in tsetse flies. This possibility was informed by findings from other insect viral systems that have been well studied, in which viruses are under the control of the host's RNAi machinery, particularly the small interfering RNA (siRNA) and micro RNA (miRNA) pathways. Research into this aspect of GpSGHV infection has revealed that the virus provokes an RNAi defense response in *G. pallidipes* as evidenced by significant downregulation of some key genes in the siRNA arm of the RNAi machinery (e.g., *Argonaute*) in symptomatic *G. pallidipes* flies as compared to artificially or naturally-infected, asymptomatic flies. A knockdown of *Argonaute-2* gene led to an increased susceptibility to GpSGHV infection as evidenced by increased viral replication. Additionally, evidence suggests the virus alters the host's miRNA expression profiles in *G. pallidipes* (by targeting the host's immune genes and participate in viral immune evasion).

Transmission and Epidemiology

The transmission dynamics of GpSGHV and MdSGHV are distinct in both wild and laboratory-bred tsetse fly and housefly species, which could be attributed to differences in the hosts' ecologies and life histories.

GpSGHV Transmission Dynamics in the Tsetse Fly

GpSGHV is transmitted both vertically (mother-to-offspring, either trans-ovum or through infected milk gland secretions) and horizontally (fly-to-fly). Horizontal transmission is more prevalent and well documented in tsetse mass-rearing facilities where the flies are fed using an *in vitro* membrane-feeding regime. Under these conditions, up to 10 cages with flies (at average densities of approximately 75 flies per cage) are fed on the same membrane in succession. It is estimated that, during a single 10–15 min feeding event, $\sim 10^{6-9}$ viral genome copies are released via saliva secretions by an infected fly. The viral particles released by infected flies are infectious *per os* to susceptible uninfected flies that feed on the same membrane. Infection of the tracheal system, milk glands, germline cells, nurse cells, and the oocytes of the ovaries, ensures vertical viral transmission to the larvae that develop within the mother's womb (adenotrophic viviparity) and on milk gland secretions. Additionally, GpSGHV is efficiently transmitted to the F_1 progenies produced by intra-hemocoelically injected mothers. These F_1 progenies exhibit overt SGH symptoms, the proportion of which gradually increase from the first to fourth gonotrophic (G) cycles (0%–3% in G_1 , 10%–30% in G_2 ,

40%–60% in G_3 , and 100% in G_4 onwards). Except in *G. pallidipes*, injection of GpSGHV into third-instar larvae of tsetse flies does not lead to the development of overt SGH symptoms in the adults that emerge from the virus-injected larvae.

Some of the maternally inherited tsetse microbiota have been implicated in GpSGHV vertical transmission as evidenced by absence of overt SGH symptoms in the F_1 progenies produced by aposymbiotic mothers. Interestingly, the absence and/or presence of certain microbiota such as *Wolbachia* may influence vertical GpSGHV transmission of the virus. Evidence suggests a causal link between the expression of overt SGH symptoms and microbiota as attested to by the absence of the symbiont in *G. pallidipes* (exhibits overt SGH) and its presence in most other tsetse species (without overt SGH). There is no evidence for the transmission of GpSGHV from father-to-progeny, which could be due to reduced virus transmission rates, rather than to a total failure of transmission.

The mechanisms responsible for the circulation and maintenance of GpSGHV in the wild tsetse populations are unclear. Further, there is no evidence for mechanical transmission of the virus via contacts between flies, mating, or fecal contamination. This limits horizontal transmission as the main mode GpSGHV dispersal amongst wild tsetse populations in the field, especially because tsetse flies are known to aggregate on specific mammalian hosts for blood feeding. When feeding on an animal, tsetse flies create blood pools at the bite sites on their hosts. GpSGHV-infected flies could deposit infectious viral particles via saliva into the blood pools, which would be picked up and infect other susceptible flies. This may facilitate transmission of the virus from fly-to-fly without necessarily replicating in the mammalian hosts. The dynamics of such horizontal GpSGHV transmission probably depends on feeding behaviors of specific tsetse species, their host preferences, feeding regimes and susceptibility to GpSGHV infections. Nevertheless, these factors partially explain the significantly low prevalence of GpSGHV in the wild tsetse populations (0.4%–15%) compared to colonized laboratory populations (almost 100% based on a diagnostic polymerase chain reaction [PCR] method).

MdSGHV Transmission Dynamics in the Housefly

MdSGHV is transmitted mainly horizontally during the gregarious feeding of the houseflies. There is no evidence of vertical or sexual transmission of MdSGHV from mother to the progeny in infected houseflies. During feeding, viremic houseflies release an estimated 10^6 MdSGHV genome copies onto the food substrates, which are infectious and serve as inoculum to infect most ($\geq 65\%$) of the co-feeding healthy conspecifics. When newly eclosed (~ 2 h-old) adult houseflies are force-fed with MdSGHV preparations under laboratory conditions, $\geq 50\%$ of the treated flies display SGH symptoms. Unlike in the GpSGHV-tsetse infection system, MdSGHV-infected houseflies remain infectious until their death. However, dynamics of MdSGHV *per os* infection under a laboratory setting may be different in nature because newly eclosed flies do not imbibe food until after several hours. Within 12–24 h post-eclosion, synthesis of the protective gut PM increases the housefly's resistance to orally-ingested virus. Under laboratory conditions, oral infections of newly eclosed adults with low viral quantities is successful only during the initial six hours post eclosion. Further, oral challenge of housefly larvae with MdSGHV suspensions does not result in detectable SGH symptoms in the adults, implying that MdSGHV is acquired during only the adult stages. This raises the question of how MdSGHV is maintained within wild populations of the housefly.

Recently, mechanical transmission of MdSGHV amongst wild populations of houseflies has been suggested as an alternative route for *per os* transmission. This is thought to be through cuticular-wounding events, which circumvents the PM barrier in the midgut by introducing the virus directly into the fly's hemocoel. The likelihood of the occurrence of the mechanical transmission of MdSGHV is partially supported by the increased activity of male houseflies, which increase cuticular wounding, especially during courtship. Notably, males harbor higher viral titers and prevalence than the females. The mechanical transmission of MdSGHV is also supported by the fact that topical applications of virus preparations to the labellum and spiracular regions of the older housefly adults, which are more likely to be wounded, induces SGH symptoms in 10% and 70% of the flies, respectively.

Diagnosis and Management of SGHV Infections

GpSGHV is presumably introduced into tsetse mass-rearing facilities from asymptomatic field-collected materials (e.g., pupae), or those derived from already existing colonies, that are used to establish new colonies or replenish existing ones. The asymptomatic infections and vertical transmission of GpSGHV allow the spread and long-term maintenance of the virus in the colonies. Undefined stress- and/or genetic-related factors may trigger outbreaks of SGH, which result in fly mortality, reduced fecundity and colony collapse.

Diagnosis

There are no obvious external clinical signs of SGHV infections in both tsetse flies and houseflies. However, in tsetse flies, hyperplastic SGs appear as pale outlines with irregular ridges in the abdomens of male flies. The discoloration is probably due to the extension of gland cells towards the cell lumina, resulting to constricted gland lumens. The enlarged and chalky-white glands are often observed in MdSGHV-infected houseflies. In the case of GpSGHV-tsetse model, a simple, sensitive and reliable non-destructive PCR-based diagnostic assay has been developed, which allows the screening of virus infections in individual live

tsetse flies. For this method, a single intermediate fly leg is excised, followed by DNA extraction, PCR amplification of a conserved viral gene (*odv-e66*), and analysis of the PCR products on agarose gels.

Management

The best option to manage GpSGHV infections in tsetse mass rearing is an integrated approach, which involves two components. The first component, which is referred to as the clean feeding system (CFS), is based on strict sanitation, regular and routine monitoring of virus infections and occurrence of overt SGH symptoms, and quarantine of infected fly colonies. The CFS is a modification of the *in vitro* feeding membrane regime aimed to interrupt horizontal virus transmission within the tsetse colonies, and consists of three feeding rounds. On the same membrane, teneral flies are always fed first, after which the rest of the fly cages are fed in a second feeding round, and finally oldest colony flies are fed last in the third round. This prevents uninfected young colony flies from picking up the virus that may be deposited onto feeding membranes by older infected flies. The second component is the supplementation of blood meals with antiviral drugs (e.g., valacyclovir), which are administered at low doses (non-detrimental to fly's DNA synthesis). Virally encoded thymidylate synthase enzymes convert the drugs into active metabolites, which subsequently block viral replication and thus reduce viral titers and shedding. In practice, CFS breaks the cycles of SGH outbreaks and thus rapidly restores colony productivity (within six months) without additional investments in specialized equipment or reagents. The SGH outbreaks are eliminated by the administration of antiviral drugs, which can be withdrawn in case of undesirable side effects, or development of drug resistance. After withdrawal of the antiviral drugs, CFS protocols alone can reduce viral infections to low levels (i.e., <10% virus prevalence) that do not compromise colony productivity. Overall, CFS is recommended in colonies with low SGH or non-observable SGH outbreaks to prevent any development and outbreaks of SGH symptoms.

Conclusions

The SGHVs are a small group of large dsDNA viruses currently restricted to dipteran insects. However, the detection of virus particles that induce the SGH symptoms in Hymenoptera, and are morphologically similar to GpSGHV and MdSGHV, suggest possibility of more members of the *Hytrosaviridae* family. The majority of the SGHV genes are hypothetical (unknown functions) and have limited or no homologies to known genes in other invertebrate viruses. The pathological differences observed from the SGHVs (MdSGHV vs. GpSGHV, and between different GpSGHV strains) appear to reside in the genetics of both the SGHVs and their respective hosts. The ecologies and the life histories of SGHV hosts may have significantly influenced the SGHVs-host coevolution. The roles of other co-infecting microorganisms, particularly the microbiota, cannot be overlooked. The GpSGHV-induced differential expression of overt SGH symptoms in certain *Glossina* species and not others could be attributed to the host's inherent abilities to mount a robust immune response to counteract the viral infections. It is now evident that, different wild and laboratory-bred populations of tsetse fly species are infected by different lineages of the same GpSGHV virus, even in *G. pallidipes* (the only *Glossina* species known to exhibit overt SGH symptoms). Further, the RNAi is a major contributor to the maintenance of asymptomatic/latent infection state of GpSGHV in tsetse as evidenced by the virus-induced provocation of an RNAi defense response via the siRNA arm of the RNAi machinery, and modulation of the host's miRNA profiles during symptomatic infection state.

Further Reading

- Abd-Alla, A.M.M., Cousserans, F., Parker, A.G., *et al.*, 2008. Genome analysis of a *Glossina pallidipes* salivary gland hypertrophy virus reveals a novel, large, double-stranded circular DNA virus. *Journal of Virology* 82 (9), 4595–4611.
- Abd-Alla, A.M., Kariithi, H.M., Cousserans, *et al.*, 2016. Comprehensive annotation of *Glossina pallidipes* salivary gland hypertrophy virus from Ethiopian tsetse flies: A proteogenomics approach. *Journal of General Virology* 97 (4), 1010–1031.
- Abd-Alla, A.M.M., Vlak, J.M., Bergoin, M., *et al.*, 2009. *Hytrosaviridae*: A proposal for classification and nomenclature of a new insect virus family. *Archives of Virology* 154 (6), 909–918.
- Amargier, A., Lyon, J., Vago, C., Meynadier, G., Veyrunes, J., 1979. Discovery and purification of a virus in the gland hyperplasia of insects. a study on *Merodon equestris* F. (Diptera, Syrphidae). *Comptes rendus hebdomadaires des séances de l'Académie des sciences. Série D: Sciences naturelles* 289 (5), 481–484.
- Coler, R., Boucias, D., Frank, J., *et al.*, 1993. Characterization and description of a virus causing salivary gland hyperplasia in the housefly, *Musca domestica*. *Medical and Veterinary Entomology* 7 (3), 275–282.
- Garcia-Maruniak, A., Maruniak, J.E., Farmerie, W., Boucias, D.G., 2008. Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of *Musca domestica*, MdSGHV. *Virology* 377 (1), 184–196.
- Jehle, J.A., Abd-Alla, A.M., Wang, Y., 2013. Phylogeny and evolution of *Hytrosaviridae*. *Journal of Invertebrate Pathology* 112 (Suppl. 1), S62–S67.
- Kariithi, H.M., 2013. *Glossina hytrosavirus* control strategies in tsetse fly factories: Application of infectomics in virus management. Wageningen, The Netherlands: Wageningen University and Research.
- Kariithi, H.M., Boucias, D.G., Murungi, E.K., *et al.*, 2018. Coevolution of hytrosaviruses and host immune responses. *BMC Microbiology* 18 (Suppl. 1), 183.
- Kariithi, H.M., van Lent, J.W., Boeren, S., *et al.*, 2013. Correlation between structure, protein composition, morphogenesis and cytopathology of *Glossina pallidipes* salivary gland hypertrophy virus. *Journal of General Virology* 94 (1), 193–208.
- Kariithi, H.M., Yao, X., Yu, F., *et al.*, 2017. Responses of the housefly, *Musca domestica*, to the hytrosavirus replication: Impacts on host's vitellogenesis and immunity. *Frontiers in Microbiology* 8, 583.

- Lietze, V.-U., Abd-Alla, A., Vreysen, M., Geden, C.C., Boucias, D.G., 2011. Salivary gland hypertrophy viruses: A novel group of insect pathogenic viruses. *Annual Review of Entomology* 56, 63–80.
- Meki, I.K., 2018. Hytrosavirus in tsetse flies: Phylogeography and molecular mode of action. Wageningen, The Netherlands: Wageningen University and Research.
- Orlov, I., Drillien, R., Spehner, D., *et al.*, 2018. Structural features of the salivary gland hypertrophy virus of the tsetse fly revealed by cryo-electron microscopy and tomography. *Virology* 514, 165–169.
- Vreysen, M.J.B., Saleh, K.M., Lancelot, R., Bouyer, J., 2011. Factory tsetse flies must behave like wild flies: A prerequisite for the sterile insect technique. *PLoS Neglected Tropical Diseases* 5 (2), e907.

Relevant Website

www.ictv.global/report/hytrosaviridae

Hytrosaviridae.

Hytrosaviridae.

dsDNA Viruses.

ICTV.

Iflaviruses (*Iflaviridae*)

Bryony C Bonning, University of Florida, Gainesville, FL, United States

Sijun Liu, Iowa State University, Ames, IA, United States

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Endogenous viral element (EVE) Virus-derived DNA sequence in the genome of an organism.

Internal ribosome entry site (IRES) Complex secondary RNA structure that allows translation initiation independent of a 5' cap.

UTR Untranslated region in an mRNA at the 5' or 3' end of an open reading frame.

VPg Viral protein genome-linked attaches to the 5' end of positive strand viral RNA and functions to prime RNA synthesis.

Classification

The family *Iflaviridae* is classified under the order *Picornavirales* (Realm *Ribovira*) and shares a number of important characteristics with other families in that order. Iflaviruses have non-enveloped, icosahedral virions ranging from 22 to 30 nm in diameter (Fig. 1). The virions encase a single copy of a positive sense, single-stranded RNA. The genome is 8–10.5 k nucleotides (excluding the polyA tail) and non-segmented with a polyA tail. Translation occurs directly from the genomic RNA via an internal ribosome entry site (IRES) to produce a single polyprotein. This polyprotein is post-translationally cleaved to generate the structural (capsid) and nonstructural proteins.

The type species of the family *Iflaviridae* is *Infectious flacherie virus* (IFV), and all members of the family as recognized by the International Committee on Taxonomy of Viruses (ICTV; Table 1) belong to the genus *Iflavirus*. The name “ifla” is derived from *Infectious flacherie virus*. Iflaviruses are demarcated by host range and <90% amino acid identity in the sequence of the capsid protein precursor. Based on this, *Deformed wing virus* (DWV) and *Varroa destructor virus 1* (VDV1) with 95% amino acid identity appear to be different isolates of the same species.

An additional 17 unclassified RNA viruses appear to belong to the genus *Iflavirus* on the basis of their genome organization (Table 2).

The diversity of iflaviruses is evident from phylogenetic analysis which shows three major clades (Fig. 2) with no relationship between virus relatedness and host range. The dicistroviruses *Cricket paralysis virus* (CrPV) and *Drosophila C virus* (DCV) group separately from *Perina nuda virus* (PnV) and *Ectopis obliqua virus* (EoV) which in turn distinctly differ from other members of the family.

Virion Structure

The smooth surfaced, icosahedral virions are comprised of four structural proteins; three major capsid proteins (VP1-VP3, ranging from 28 to 35 kDa in size) and a minor structural protein (VP4, 4–12 kDa in size). The virion structure is similar to that of the dicistroviruses, for which VP4 is located on the interior of the capsid and may function to bind genomic RNA within the virion. The icosahedral particles have a $T = 3$ symmetry and are made up of 60 protomers with a basic structure similar to those of mammalian picornaviruses. Low pH promotes genome release from the virion implicating host cell entry via endosomes.

Genome

The primary characteristics of the iflaviruses genome are a non-segmented genome of 8–10.5 kb excluding the 3' polyA tail. The genome is translated into a single polyprotein (Fig. 3). Cap-independent translation of the iflaviruses polyprotein is mediated by an internal ribosome entry site (IRES) within the 5' UTR. The polyprotein is cleaved by the virus-encoded protease. Structural proteins are encoded at the 5'-proximal region while non-structural proteins are encoded at the 3' proximal region of the genome. The structural proteins are preceded by a leader protein (L) of unknown function that is removed from VP2 prior to capsid assembly. The nonstructural proteins are encoded in the order RNA helicase (Hel), a chymotrypsin-like 3C protease (pro), and RNA-dependent RNA polymerase (RdRP) similar to other small RNA viruses. The iflaviruses RdRP has a canonical RGD motif, which is typical for RNA polymerases of positive sense, single-stranded RNA viruses. The genome organization of iflaviruses strongly resembles that of picornaviruses.

The 11.5 kDa VPg protein covalently associates with the 5' of the positive and negative strand viral RNA and may be required to prime RNA synthesis similar to the situation in *Picornaviridae*. The location of sequence encoding the VPg protein has yet to be resolved. The construction of an infectious clone of a *Deformed wing virus* genome provides a useful tool for increasing understanding of iflaviruses biology.

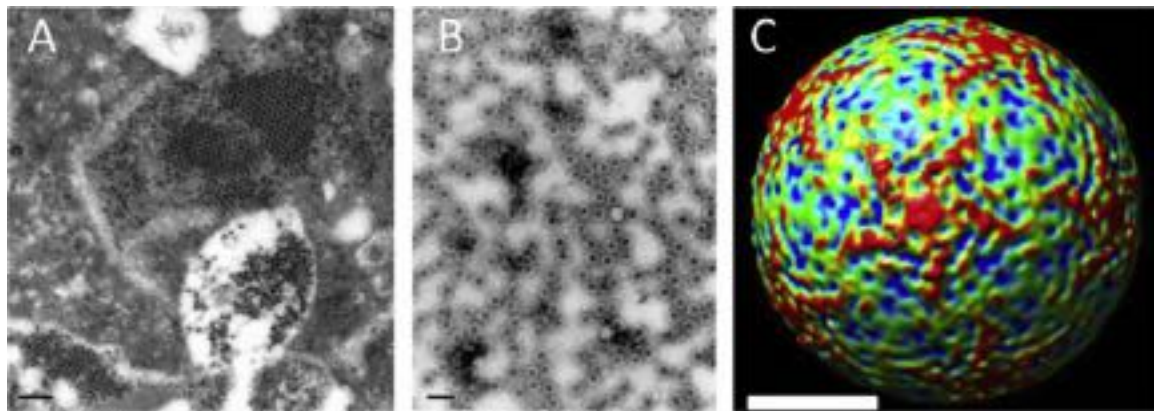


Fig. 1 Iflavirus particles. (A) Transmission electron micrograph of *Varroa destructor virus 1* (VDV1) particles within a *Varroa* mite. Bar, 150 nm. (B) Purified VDV1 particles. Bar, 50 nm. Photographs courtesy of Dick Peters. (C) Surface view of a virion of *Infectious flacherie virus* (IFV) along a 5-fold axis reconstructed by cryo-electron microscopy. Bar, 10 nm. Courtesy of J. Hong.

Table 1 Species within the Genus *Iflavirus*

Species	Virus abbreviation	Accession number	Host
<i>Antheraea pernyi iflavirus</i>	LnApIV	KF751885	Chinese (oak) tussar moth
<i>Brevicoryne brassicae virus</i>	BBV	EF517277	Cabbage aphid
<i>Deformed wing virus</i>	DWV	AJ489744	Honey bee
<i>Dinocampus coccinellae paralysis virus</i>	DcPV	KF843822	Braconid wasp
<i>Ectopis obliqua virus</i>	EoV	AY365064	Tea looper
<i>Infectious flacherie virus</i>	IFV	AB000906	Silkworm
<i>Lygus lineolaris virus 1</i>	LyLV1	JF720348	Tarnished plant bug
<i>Lymantria dispar iflavirus 1</i>	LdIV1	KJ629170	Gypsy moth
<i>Nilaparvata lugens honeydew virus 1</i>	NLHV1	AB766259	Brown planthopper
<i>Perina nuda virus</i>	PnV	AF323747	Ficus transparent wing moth
<i>Sacbrood virus</i>	SBV	AF092924	Honey bee
<i>Slow bee paralysis virus</i>	SBPV	EU035616	Honey bee
<i>Spodoptera exigua iflavirus 1</i>	SEIV1	JN091707	Beet armyworm
<i>Spodoptera exigua iflavirus 2</i>	SEIV2	JN870848	Beet armyworm
<i>Varroa destructor virus 1</i>	VDV1	AY251269	Honey bee

Table 2 Unclassified, putative members of *Iflaviridae*

Proposed name	Virus abbreviation	Accession number
<i>Bombyx mori iflavirus</i>	BMIV	LC068762
<i>Ceratitis capitata iflavirus 1</i>	CcIV1	GAMC01001920
<i>Ceratitis capitata iflavirus 2</i>	CcIV2	GAMC01020602
<i>Formica exsecta virus 2</i>	FeV2	KF500002
<i>Graminella nigrifrons virus 1</i>	GnV1	KP866792
<i>Halyomorpha halys virus</i>	HhV	KF699344
<i>Heliconius erato iflavirus</i>	HeIV	KJ679438
<i>King virus</i>	KgV	KX779454
<i>Kinkell virus</i>	KkV	KU754510
<i>La Jolla virus</i>	LJV	KP714074
<i>Laodelphax striatella honeydew virus 1</i>	LsHV1	KF934491
<i>Laodelphax striatellus iflavirus 2</i>	LsIV2	KM272628
<i>Moku virus</i>	MV	NC031338
<i>Nilaparvata lugens honeydew virus 2</i>	NIHV2	AB826459
<i>Nilaparvata lugens honeydew virus 3</i>	NIHV3	AB826460
<i>Osiphanes invirae iflavirus 1</i>	OiIV1	KR534892
<i>Thaumetopoea pityocampa iflavirus 1</i>	TpIV1	KP217032

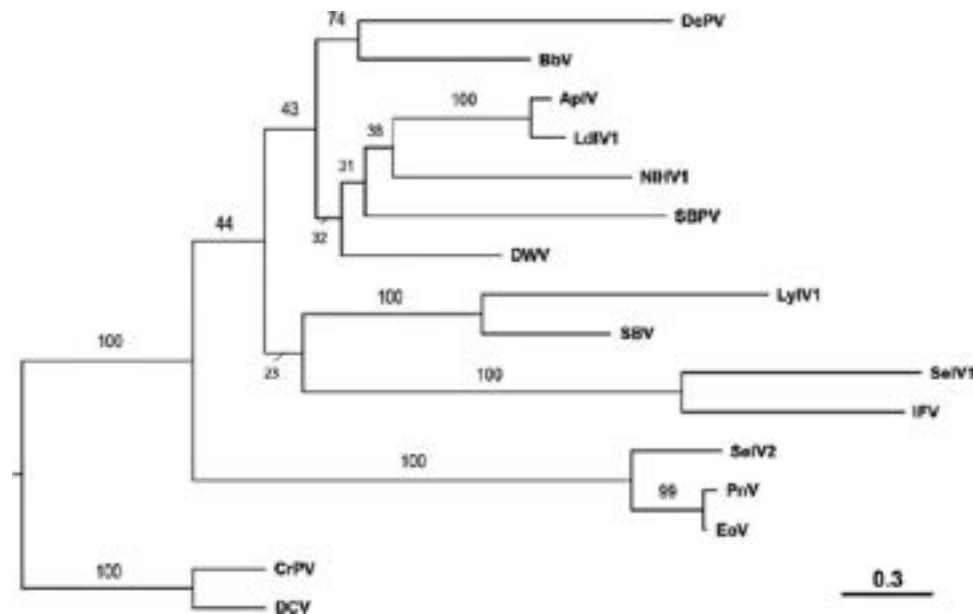


Fig. 2 Phylogenetic relationship between viruses in the family *Ifloviridae* based on RdRP sequences. Dicistroviruses CrPV (*Cricket paralysis virus*), NP_647481.1; DCV, (*Drosophila C virus*), NP_044945.1 were used as outgroups. RdRP sequences were aligned using MAFFT version 7 (Multiple Alignment using Fast Fourier Transform) and phylogenetic analysis was conducted using the maximum likelihood (ML) method with 10,000 bootstraps. Evolutionary distance marker indicates 0.3 substitutions per amino acid position. RdRP sequences used were BbV (*Brevicoryne brassicae virus*), YP_001285409.1; DWV (*Deformed wing virus*), AST47884.1; DcPV, (*Dinocampus coccinellae paralysis virus*), YP_009111311.1; EoV (*Ectopis obliqua virus*), NP_919029.1; IFV (*Infectious flacherie virus*), NP_620559.1; LdIV1 (*Lymantria dispar iflavirus 1*), YP_009047245.1; NLHIV1 (*Nilaparvata lugens honeydew virus 1*), YP_009505599.1; PnV (*Perina nuda virus*), NP_277061.1; SBV (*Sacbrood virus*), ATA66298.1; SBPV (*Slow bee paralysis virus*), ADI46683.1; SelV1 (*Spodoptera exigua iflavirus 1*), YP_004935363.1; SelV2 (*Spodoptera exigua iflavirus 2*), YP_009010984.1.

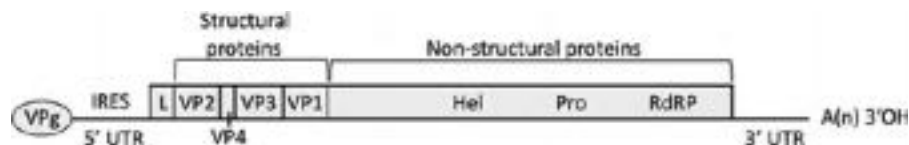


Fig. 3 Genome organization of viruses within the family *Ifloviridae*. The genome encodes a single polyprotein with structural proteins (VP1 to VP4) toward the N-terminus, and nonstructural proteins (helicase, protease, RNA-dependent RNA polymerase) toward the C-terminus. An internal ribosome entry site (IRES) in the 5' untranslated region (UTR) allows for translation directly from the genomic RNA. A covalently linked protein, VPg is present at the 5' end of the genome. The polyprotein produced on translation of the single open reading frame, is auto-catalytically cleaved to generate the individual structural and nonstructural proteins. There are two systems for naming of iflavirus structural proteins, based on genome position and sequence identity, or on size or molecular mass. The system used by the ICTV based on molecular mass is adopted here.

Life Cycle

Ifloviruses have been isolated from only arthropods (**Table 1**) with host ranges limited to one or a few closely related species. Ifloviruses may be vertically and / or horizontally transmitted. For example, DWV is vertically transmitted from queen honey bees to workers and drone progeny, but may also be transmitted from worker to worker either directly through food exchange, or indirectly via the *Varroa* mite. In the latter case, virus is released directly into the body cavity of the honey bee.

Epidemiology

Analysis of the incidence and distribution of honey bee viruses highlighted the dynamic nature of virus infection across hives, locations and seasons. The *Varroa* mite plays a key role in changing the viral landscape with increased DWV prevalence and titer, and selection of a subset of viral variants. Co-infection of a host by ifloviruses and viruses from other families is typical.

Clinical Signs

Iflaviruses are transmitted primarily on ingestion of contaminated diet. Many iflaviruses infect the host persistently without inducing disease symptoms. These viruses are transmitted vertically, which is characteristic of covert, persistent or chronic infection. There are several exceptions to this however with viruses resulting in symptomatic infections including discoloration, diarrhea, developmental abnormalities and death of the host.

Infectious flacherie virus infects mainly the goblet cells in the midgut epithelium of the silkworm and results in accumulation of fluids known as the flacherie phenotype. This flacherie disease causes diarrhea and results in significant losses to the sericulture industry in China and Japan. The flacherie phenotype is also observed on infection of other lepidopteran species by other iflaviruses. In addition fluid accumulation under the skin affects honey bee larvae infected with *Sacbrood virus* resulting in fatal sacbrood disease.

Iflaviruses of the honey bee have received particular attention as a result of research into global honey bee declines. Viral infection (by DWV in particular) is exacerbated by the presence of the Varroa mite, which vectors viruses and weakens the honey bee thereby reducing anti-viral defenses. At high titers, DWV impairs honey bee development, resulting in insects with deformed wings that are unable to fly. DWV can also affect learning behavior and aggression, in addition to lifespan.

Pathogenesis

Iflaviruses replicate in the cytoplasm of host cells and can be observed as crystalline arrays of virions ([Fig. 1](#)). Virus-induced vesicular structures appear to play an important role in virus replication, similar to other picornaviruses. The impact of replication on host cell macromolecular synthesis has not been established for members of this family. Disease manifests when viral titers reach high levels. Information on the manner of disease development is currently limited.

Diagnosis/Detection Methods

Serological methods such as western blot and enzyme-linked immunosorbent assay (ELISA) can be used for detection of viral proteins. Iflaviruses (with the exception of DWV and VDV1 which are viruses of the same species) are serologically distinct, with no serological relationships between family members.

Reverse transcriptase-polymerase chain reaction (RT-PCR) with primers based on unique genome sequences provides an accurate and specific approach for detection of viral sequence. The use of multiple primer pairs is recommended to reduce the likelihood of viral sequence resulting from endogenous viral elements, rather than complete viral genomes. RT-PCR can also be used for detection of negative strand RNA (indicative of virus replication) provided that the appropriate controls (including positive sense RNA) are included. Quantitative PCR (RT-qPCR) can be used for determination of virus titer (genome equivalents) with reference to an appropriate viral sequence standard. A high viral titer would indicate virus replication in a given species, rather than virus that may be associated with an insect, on the surface or in the gut for example.

Cell lines that support replication are available for some iflaviruses, including DWV, PnV, and *Lymantria dispar virus 1* (LdIV1). The availability of such cell lines will facilitate further study of the biology of this virus family.

Prevention

Maintenance of virus free insect colonies and cell lines in isolated, clean facilities is recommended. Management of the Varroa mite in honey bee colonies is essential for preventing escalation of viral titers to potentially damaging levels. Colonies with enhanced hygienic behavior are better able to survive in the presence of Varroa mites.

Further Reading

- Carrillo-Tripp, J., Bonning, B.C., Miller, W.A., 2015. Challenges associated with research on RNA viruses of insects. *Current Opinion in Insect Science* 8, 62–68.
- Carrillo-Tripp, J., Krueger, E.N., Harrison, R.L., *et al.*, 2014. *Lymantria dispar* iflavirus 1 (LdIV1), a new model to study iflaviral persistence in lepidopterans. *Journal of General Virology* 95, 2285–2296.
- de Miranda, J.R., Genersch, E., 2010. Deformed wing virus. *Journal of Invertebrate Pathology* 103 (Suppl. 1), S48–S61.
- Fannon, J.M., Ryabov, E.V., 2016. Iflavirus (deformed wing virus). In: Liu, D. (Ed.), *Molecular Detection of Animal Viral Pathogens*. Boca Raton: CRC Press, pp. 37–46.
- Lamp, B., Url, A., Seitz, K., *et al.*, 2016. Construction and rescue of a molecular clone of Deformed Wing Virus (DWV). *PLoS One* 11 (11), e0164639.
- Martin, S.J., Highfield, A.C., Brettell, L., *et al.*, 2012. Global honey bee viral landscape altered by a parasitic mite. *Science* 336 (6086), 1304–1306.
- McMenamin, A.J., Genersch, E., 2015. Honey bee colony losses and associated viruses. *Current Opinion in Insect Science* 8, 121–129.
- Oers, M.M.V., 2010. Genomics and biology of iflaviruses. In: Johnson, K.N., Asgari, S. (Eds.), *Insect Virology*. Norfolk, UK: Caister Academic Press, pp. 231–250.
- Runckel, C., Flenniken, M.L., Engel, J.C., *et al.*, 2011. Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, nosema, and crithidia. *PLoS One* 6 (6), e20656.
- Wilfert, L., Long, G., Leggett, H.C., *et al.*, 2016. Deformed wing virus is a recent global epidemic in honeybees driven by Varroa mites. *Science* 351 (6273), 594–597.

Relevant Websites

<https://www.ncbi.nlm.nih.gov/genome/?term=Iflavirus>

GEO DataSets Result.

NCBI.

NIH.

https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/picornavirales/w/iflaviridae

ICTV Report on Iflaviridae.

<https://viralzone.expasy.org/278>

ViralZone.

Iridoviruses of Invertebrates (*Iridoviridae*)

İkbal Agah İnce, Department of Medical Microbiology, Acibadem University School of Medicine, Istanbul, Turkey

© 2021 Elsevier Ltd. All rights reserved.

Glossary

DE Delayed early gene transcription.

IE Immediate early gene transcription.

IFN Interferon.

IIV Insect Iridovirus.

Iridovirid Generic designation of members of the family

Iridoviridae.

L Late gene transcription.

NCLDV Nucleocytoplasmic large DNA viruses, a non taxonomic grouping of large DNA viruses that replicate exclusively in the cytoplasm, or like iridoviruses, in both the cytoplasmic and nucleus.

Introduction

Iridovirids, a generic designation for members of the family *Iridoviridae*, are nucleocytoplasmic large dsDNA viruses (NCLDVs) that infect both invertebrates (IIVs) and ectothermic vertebrates (Vertebrate Iridoviruses, VIVs). Infections may be asymptomatic or range in severity from minor reductions in host fitness to systemic disease and large-scale mortality.

Currently, Iridovirids are classified according to virion particle size, host preference, presence of a DNA methyltransferase gene, GC content and phylogeny based on the amino acid sequence of the major capsid protein (Table 1).

Although new iridovirids are continuously being discovered, the genomes of only a limited number of IIVs have been sequenced, and thus, it is hard to precisely quantify or estimate the genetic heterogeneity and variability within each genus. Additionally, it is still not clear what type of information or criteria would be sufficient to determine whether newly discovered iridovirids represent new species or are just variants or strains.

The advent of phylogenetic analysis based on complete genome sequences now provides a superior method of differentiation and classification, (Fig. 1), and using pan-genomic data, which needs the raw sequencing data of the available complete genomes, might very well advance the analysis of viral phylogenies, which may ultimately lead to the development of new criteria for virus classification in general.

Classification of Iridovirids

The invertebrate iridoviruses belong exclusively to the *Betairidovirinae* subfamily of the Family *Iridoviridae*. This subfamily consists of three genera, *Iridovirus* with two species, including the type species *Invertebrate Iridovirus 6*, *Chloriridovirus* with 5 species including the type species *Invertebrate Iridovirus 3* and recently *Decapodiridovirus* with a single species, the type species *Decapod Iridovirus 1*.

There are 26 core genes present in all 45 iridovirid genome sequences, forming four clades. Three clades for each of the vertebrate-infecting genera (*Ranavirus*, *Lymphocystivirus* and *Megalocytivirus*) and one clade for the invertebrate iridoviruses, recently split into three genera (*Iridovirus*, *Chloriridovirus* and *Decapodiridovirus*). Due to the independent monophyletic origins of these vertebrate and invertebrate iridoviruses the subfamilies, *Alphairidovirinae* and *Betairidovirinae* are now recognized.

The currently accepted approach for iridovirid taxonomy is to use the conserved iridovirid genes (26 genes) in phylogenetic reconstructions. The criteria distinguishing taxa for Subfamilies are based on primary host species (i.e., vertebrate vs. invertebrate) and methylation content.

Genera are distinguished by principal host species, sequence-based phylogenetic analysis and sequence identity, with members of different genera displaying <50% sequence identity among a concatenated set of 26 core genes. Species share >95% sequence identity within the set of 26 core genes, phylogenetic relatedness, a co-linear gene order, and similar genome size ($\pm 10\%$) and

Table 1 Comparison of *Betairidovirinae* genera based on their different qualities

Genera	Virion size	Hosts	GC content	DNA methylation
<i>Iridovirus</i>	120–130 nm	arthropods, particularly insects	29%–32%	Absent
<i>Chloriridovirus</i>	180 nm	Diptera with aquatic larval stages, mainly mosquitoes	48%	Absent
<i>Decapodiridovirus</i>	158 nm	Crustaceans	34.58%	Absent

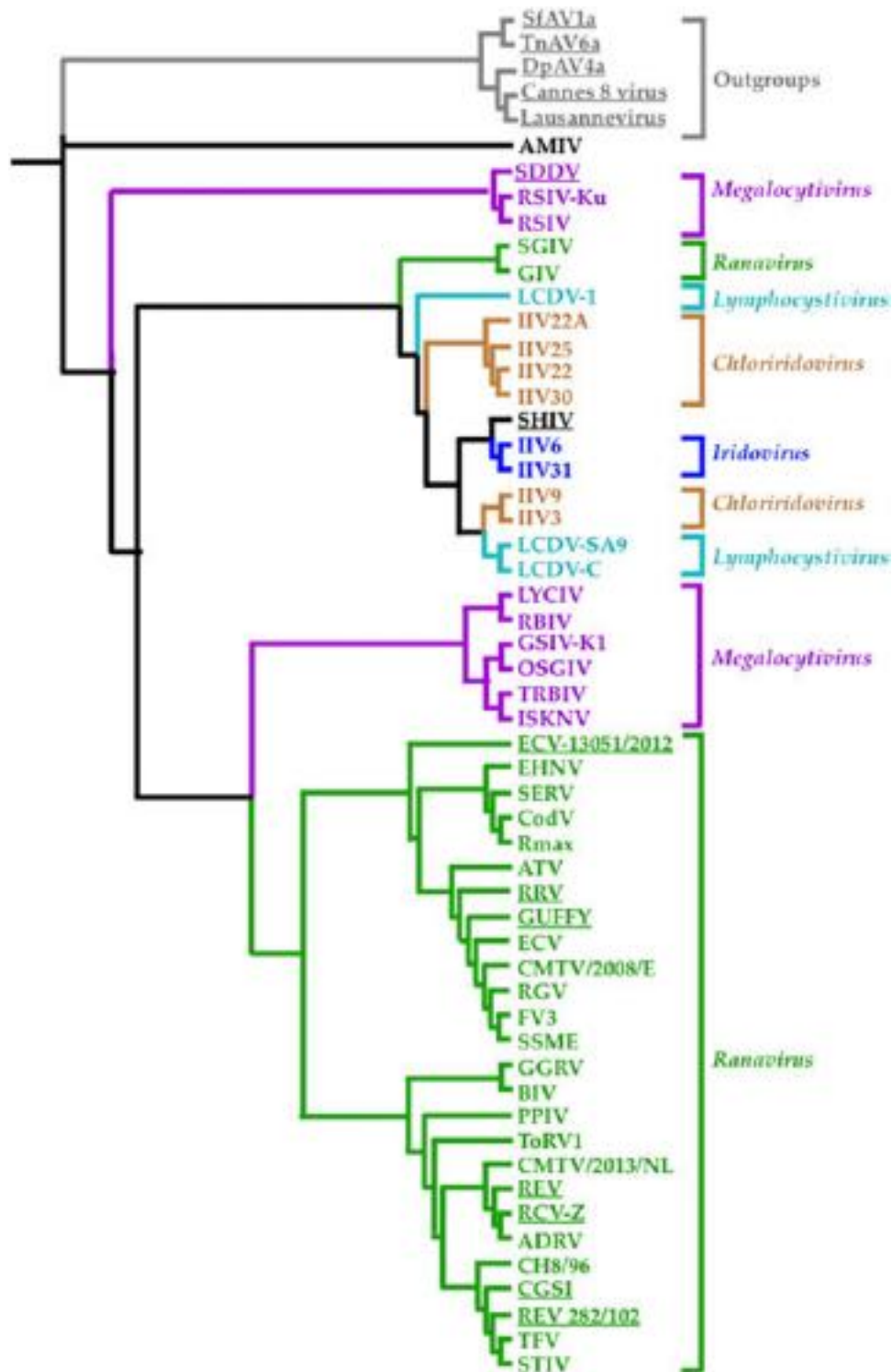


Fig. 1 Complete iridovirid genome sequences based on RAxML Maximum likelihood phylogeny of family *Iridoviridae* as described in Ince, I.A., Özcan, O., Ilter-Akulke, A.Z., Scully, E.D., Özgen, A., 2018. Invertebrate iridoviruses: A glance over the last decade. *Viruses* 10, E161.

G + C content. Although these criteria provide a rationale for the current taxonomy, they are rather poorly defined in terms of assigning new members to iridovirids.

The genomes of many iridovirids have now been sequenced completely, and phylogenetic reconstructions based on complete genome sequences or genome methylation patterns may provide additional resolution when compared to approaches that use only a subset of genes.

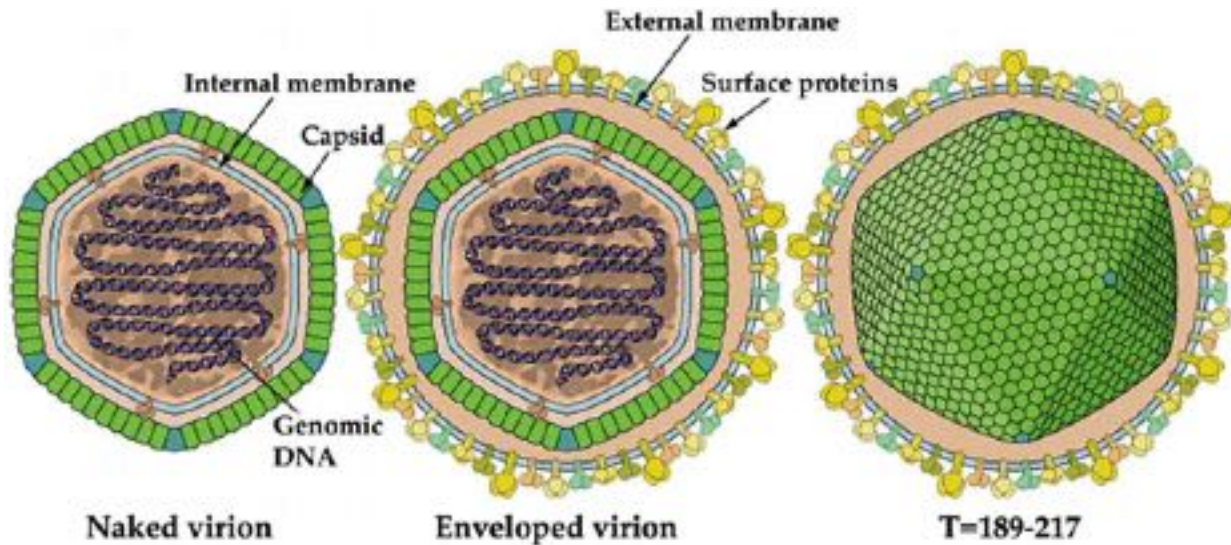


Fig. 2 Graphical representation of the iridovirid structure. Source: ViralZone: www.expasy.org/viralzone. SIB Swiss Institute of Bioinformatics.

Morphology and Composition

IIV-6, formerly called Chilo Iridescent virus, is the most studied invertebrate iridescent virus (IIV). The 212 kbp IIV-6 genome is packed into an icosahedral capsid with a $T = 147$ lattice. The virion is comprised of four concentric domains: an outer membrane derived from the plasma membrane of host in the virions that are released by budding, a proteinaceous capsid layer, an intermediate lipid membrane with associated polypeptides and a DNA-protein core (Fig. 2). Initially measured at 120–130 nm diameter in ultrathin sections, CryoEM and 3D image reconstruction showed a maximum diameter of 185 nm including fibrils emanating from the surface of the virion.

IIV-6 Persistence and Sensitivity to External Factors

IIVs are highly stable in water, but dry soil rapidly inactivates the infectivity of IIV-6. IIV-6 is thermolabile and can be rapidly inactivated once temperatures rise above 55°C. Furthermore, solar UV light and ultraviolet radiation reduces IIV-6 infectivity, especially in aqueous habitats. Exposure to various solvents and enzymes can also impact IIV-6 infectivity, and sensitivity exists both in cell culture and whole insects. In addition, differences in sensitivity to suspensions, based on whether the assay was performed in cell culture or in whole insects, were observed. Notably, when the IIV-6 lipid bilayer was disrupted, the virus was still capable of infecting cultured cells, but not insect larvae when the inoculum was injected into the haemocoel.

These observations highlight the need for studies on the role of the lipid component of the virus (1) in the process of entry (2) host specificity in different hosts and cell lines. This may also explain the observed differences in sensitivity to organic solvents and enzymes documented between vertebrate and invertebrate viruses.

Host Range and Pathology

The host range of IIV-6 includes over 100 insect species belonging to six orders; Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera and Orthoptera. Additionally, IIV-31 can infect several different isopod species, as well as Drosophilidae. Iridovirid-like particles have also been detected in marine invertebrates. Patent infection with IIV turns the host an iridescent blue and is mostly considered lethal. However, (inapparent) infections are more common, reducing the reproductive capacity and longevity of adult hosts. In some cases, the link between viral presence and mortality is obscure.

For instance, a link between co-infection of *Nosema* species and IIV-6 with insect mortality was suggested but subsequent studies failed to find a significant correlation between IIV-6 and colony collapse disorder in the USA.

IIV-6 can be propagated in nineteen different insect cell lines, with varying levels of susceptibility. In some insects, extracellular host factors may influence viral infectivity and not all cell lines derived from susceptible hosts can be infected with IIV-6. Interestingly, reptiles and amphibians fed IIV-infected insects appear to become infected and invertebrate iridovirus (IIV-6) propagation has also been achieved in some poikilothermic vertebrate cell lines. Intraperitoneal injections of large doses of native or ultraviolet-irradiated IIV-6 were lethal to frogs and mice, while heat- or antiserum-inactivated IIV-6 had no lethal toxicity.

Vertebrates can become infected with IIVs and lethality was observed under laboratory conditions and at high doses, but natural infections are rarely symptomatic or lethal, suggesting that IIV replication is more limited in vertebrate hosts and that the host immune system is able to inhibit viral replication.

Mass produced feeder insects might be a source of viral diseases for reptiles and amphibians kept as pets. However, large-scale IIV outbreaks have not yet been observed.

In vertebrate and invertebrate hosts, IIV-6 infection causes different immune responses. In invertebrates, autophagy and/or apoptosis may not be a typical host response to IIV-6 e.g., *Drosophila*. IIV-6 encoded proteins interfere with insect RNAi immune responses while the JAK-STAT pathway does not provide immunity against an IIV-6 infection.

In vertebrates IIV-6, like many other DNA viruses, can induce a type I IFN-dependent antiviral response in mammalian cells, which is likely triggered by a cytosolic sensor upon viral dsRNA recognition, which is known to be produced in the course of IIV-6 transcription.

Overall, studies suggest that IIVs can occasionally cross-infect certain vertebrates, but that their immune systems may be able to inhibit replication under natural conditions, reducing or preventing symptoms. Another possibility is that IIVs do not replicate productively in ectotherms and thus do not result in disease.

Genome Organization and Codon Usage

The IIV-6 genome consists of a single linear 212,482 bp dsDNA molecule with 28.63% G + C content. Compared to other animal viruses, the IIV-6 genome has a unique arrangement, with circular permutations and terminal redundancies that cause variations in the number of direct terminal repeats during DNA replication and genome packaging. While the coding capacity of the genome was estimated initially at 215 non-overlapping open reading frames (ORFs), the total number of ORFs rises to 468 when both non-overlapping and overlapping ORFs are considered. However, also taking overlapping ORFs into account is still a point of debate.

Importantly, gene transcripts of iridovirids show no evidence of introns, and viral mRNAs lack poly(A) tails. However, biochemical and in silico evidence suggests the existence of viral microRNAs (miRNA) that might modulate viral gene expression. In addition, unlike VIVs, apart from IIV-3, IIVs do not encode a DNA methyltransferase except for SGIV. In silico prediction of the methylation sites of 50 iridovirid genomes has demonstrated that there is a clear distinction in methylation patterns between IIVs and VIVs, except for IIV-3 from IIVs and SDDV and LCDV-1 from VIVs, which have an as yet unclear phylogenetic status.

To date, 26 core genes have been identified in both IIVs and VIVs. They have been linked to virus replication, gene transcription, protein structure and nucleocapsid assembly. The rest of the encoded proteins may have roles in virus-host interactions, hijacking of host cell machinery and establishment of viral infection.

Virion Proteins

Virion proteins have been identified for IIV-1, IIV-2, IIV-6, IIV-9, IIV-22 and IIV-25. However, we have observed that the different experimental settings in these studies do affect the resolution of the gel analyses and thus influence previous results and comparisons.

More recently, a comprehensive proteomic analysis (LC-MS/MS) detected 54 proteins in IIV-6. Possible functions of fifteen of the putative ORFs were inferred to be: serine-threonine kinase, dual specificity protein phosphatase, DNA polymerase (viral) N terminal domain, carboxy terminal domain phosphatase, ribonuclease III, tyrosine protein kinase, cathepsin, DNA-binding protein, protein disulphide isomerase, lysosome associated membrane glycoprotein, and a ranavirus enveloped protein homolog. The remaining 39 proteins lacked similarity to any other annotated viral proteins and should be investigated further.

Genomic analysis of IIV-9 identified 191 predicted genes, with 20% of its repeated sequences located mainly within the coding regions. 97 of 211 IIV-6 genes have detectable orthologs in IIV-9, whereas 108 out of 191 IIV-9 genes have orthologs in IIV-3.

Phosphorylation reactions represent common host responses to iridovirid infection. During IIV-6 infection, increased phosphorylation of the ribosome associated proteins in the early phase of the infection cycle requires the expression of the viral genome. However, phosphorylation events were observed in ribosomal proteins when permissive cells were infected with UV-irradiated IIV-6, suggesting that increasing phosphorylation reactions were likely a host cell response rather than being related to an expression of the viral genome.

Iridovirid genomes contain multiple genes coding for kinases and phosphatases, regulators of protein activities over the course of the infection. Although kinases and phosphatases are among the most abundant proteins, it is challenging to assign functions to these enzymes without expressing them individually in cell culture, making knock-out viruses, unravelling their subcellular localization, or identifying the protein targets that are specifically phosphorylated. One of the viral kinases of IIV-6, using a purified serine/threonine protein kinase 389L (also called iridoptin) was found to induce apoptosis in insect cells but it remains uncertain whether this is a virion associated protein kinase or not.

The functions of many of these viral kinase enzymes remain to be determined however some significant successes have been achieved with nucleoside analogues which are “activated” by phosphorylation by viral and/or host-cell nucleoside kinases, the final target being principally the viral polymerase. Hence, they represent possible new targets for antiviral therapy. Targeting viral polymerases would serve as a safe model system for example in areas like fish farming to develop viral disease mitigation strategies, as kinases play crucial roles in guaranteeing successful viral infections.

Furthermore, virally coded kinases present apoptotic features which could be candidates for developing pest insect controlling agents via generation of transgenic crops with using viral kinases showing pest insect specific toxicity.

Viral Entry, Replication, and Release Strategy

Iridovirids undergo nucleocytoplasmic replication. Initially, the virus attaches to cell surface receptors and the virus is engulfed by the host cell via clathrin-mediated endocytosis or macropinocytosis in a pH-dependent manner. IIV-6 ORF 096L is proposed to act as an insect cell adhesion molecule (CAM) which may facilitate “fusion” between the virus shell and the cellular membranes, or possibly between the internal lipid membrane and cellular membranes. Following the entry into the host cell, immediate early and delayed early transcripts are synthesized. These transcripts encode proteins that are crucial for viral DNA replication and expression of late genes. Then, newly synthesized viral DNA is translocated from the nucleus to the cytoplasm where a second stage of viral DNA replication results in the formation of DNA concatemers (Fig. 3). Late viral transcripts are synthesized by a viral encoded RNA polymerase that has not yet been identified in IIVs. However, there are ORFs with homology to RNA polymerase II indicating that IIVs likely encode a unique viral transcriptase.

Evidence from IIVs and VIVs suggests that a virus-encoded transcriptase synthesizes late viral messages. For example, knock down of the large subunit of the viral transcriptase of SGIV eliminates late protein synthesis.

When replicated DNA reaches the cytoplasm, it forms large concatemers which eventually serve as “feedstock” for packaging of virions. The underlying mechanism is not yet understood. Progeny virions accumulate in large paracrystalline arrays, or egress from the cell by either budding or cell lysis. Viral particles are released from host cells either by budding from the host cell membrane or together with mature virions from infected cells via lysis. (Fig. 3).

The specific classes of lipids in the virion structure can influence virus infectivity. Besides, changes in host cellular lipid composition associated with cell proliferation, apoptosis, cellular stress and viral infection can also occur during viral replication, impacting the virus’ ability to infect its hosts.

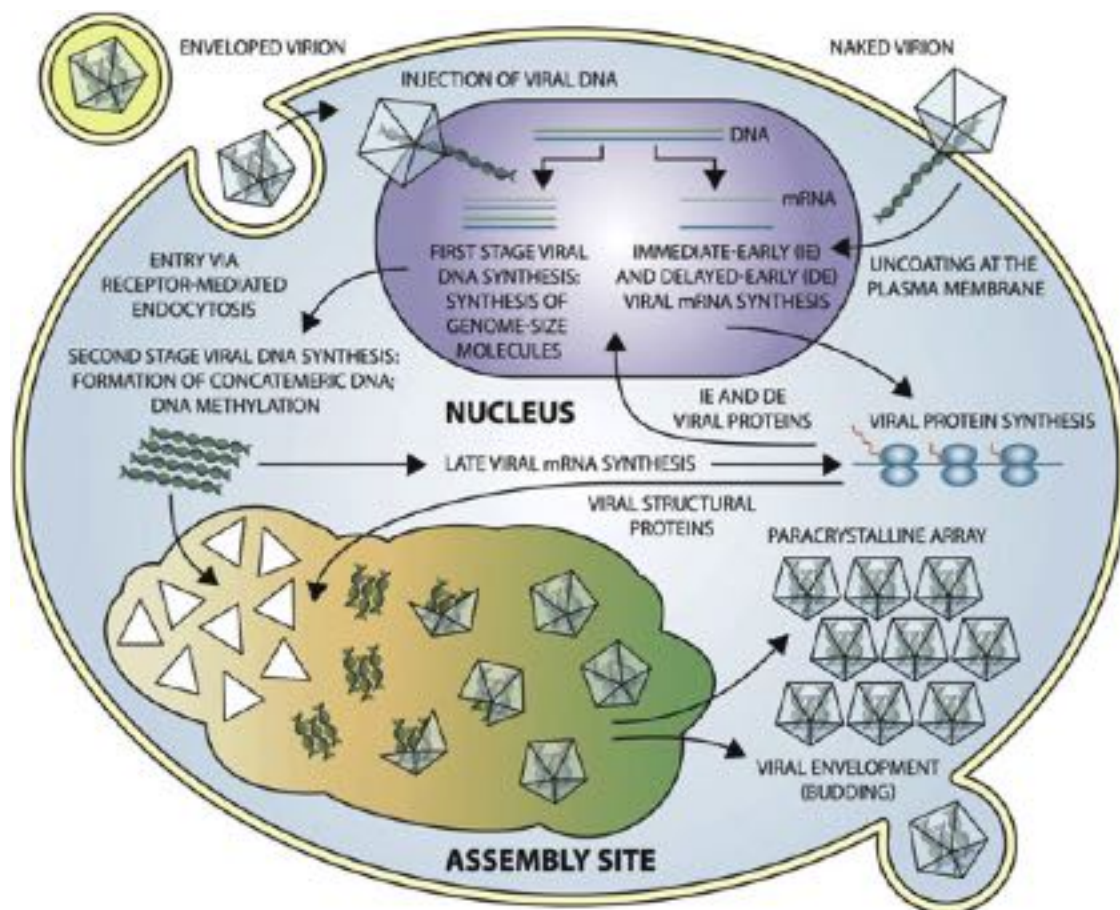


Fig. 3 Schematic representation of iridoviral replication. This figure was obtained from the 10th report of ICTV with permission Ince, I.A., Özcan, K., Vlak, J.M., van Oers, M.M., 2013. Temporal classification and mapping of non-polyadenylated transcripts of an invertebrate iridovirus. *Journal of General Virology* 94 (1), 187–192. Chinchar, V.G., Duffus, A.L.J., 2019. Molecular and ecological studies of a virus family (Iridoviridae) infecting invertebrates and ectothermic vertebrates. *Viruses* 11, 538.

Transcriptional Regulation

In IIV-6 infected cells, viral transcription occurs in a regulated temporal gene expression cascade: as immediate-early (IE), delayed-early (DE) and late (L). However, some IE and DE transcripts are still present during the L-stage, and no meaningful correlation has yet been identified between transcription and the prevalence of particular proteins during the course of infection.

A quantitative proteomic study provided detailed information about the kinetics of IIV-6 viral protein levels, showing that the transcripts belonging to the IE class are involved in mainly nucleoside metabolism, blocking host cell apoptosis, post-translational protein modifications and transcriptional activation of DE genes.

The DE genes function as viral DNA polymerases, protein kinases, and transcriptional activators of late genes. The major capsid protein MCP (274L), one of the most abundant proteins in IIV-6, was expressed with other late class transcripts.

The way viral transcriptional regulation is regulated in different target hosts is yet to be investigated.

Promoter Elements and Transcriptional Regulation

It is difficult to conclusively identify promoter elements that regulate the temporal expression patterns of iridovirid genes. A very limited number of studies identified motifs in putative promoter regions that may be important for transcriptional regulation.

IIV-6 virion proteins are likely able to either directly or indirectly activate the promoters to trigger transcription of some/ a few IE genes. Assays confirm that without virion proteins the promoters remained inactive even though the reporter plasmid constructs carrying these promoters were transfected.

Interestingly, DNAPol, helicase, and mcp transcripts of IIV-6 lack polyA tails, and clear polyadenylation signals downstream of these three IIV-6 ORFs were not found.

To identify the transcription termination signals of IIV-6 genes, ligation-based amplification of cDNA ends was used, showing that about half of all IIV-6 genes contained complementary TAATG and CATTA motifs in the 3' UTR regions of their mRNAs. These CATTA motifs may be conserved features that enable formation of hairpins which, in the absence of polyadenylation signals, may serve as transcriptional terminators.

Induction/Inhibition of Apoptosis in Infections

The principles and the dynamics of apoptosis in response to iridovirid infection are similar in vertebrates and invertebrates.

During the early stages of infection, host cells trigger apoptotic machinery to impede the cell-to-cell movement of progeny viruses. However, many viruses have evolved mechanisms to protect their progeny viruses from the adaptive immune system, producing anti-apoptotic proteins to enhance the likelihood of spreading to neighbouring cells. At the same time, viruses also contain genes that may stimulate apoptosis, using it to promote viral dissemination to neighbouring cells.

Studies found that initially, expression of early viral gene/s (IE) inhibited apoptosis, which suggests that one or more IE genes likely function as anti-apoptotic factors (e.g., IIV-6 ORF193R, *iap* gene). But IIVs also code for pro-apoptotic factors, which trigger apoptosis at later stages during the infection with the goal of facilitating viral dissemination.

Concluding Remarks

Notwithstanding the above information, knowledge gaps about iridovirus still exist. It is not known which host and/or viral factors trigger viral entry into and replication in host cells, and the regulatory mechanisms of genes expressed at various stages during infection remain uncharacterized and will likely only be answered through using functional genomics. Identifying viral proteins responsible for viral infectivity is essential for developing a genetic recombination system for iridovirids, as iridovirid DNA by itself is not infectious.

A genetic recombination system will expand the utility of IIVs as a model for identifying and characterizing iridovirid genetic factors driving virus-host interactions, impacting host fitness, and influencing immune system response, ultimately helping to limit or prevent disease in aquatic ecosystems, as well as leading to the development of IIVs as biocontrol agents. IIV-6 can be a potential biopesticide, as studies have identified virally coded genes (e.g., viral kinases) possibly linked to host toxicity, which could be exploited as biocontrol compounds.

Additional functional genomics studies and the expansion of genomic resources may lead to the identification of new virulence factors and toxins that could be exploited and engineered for targeted control against certain groups of insects.

Furthermore, virally coded kinases could represent potential antiviral drug targets in areas like fish farming, as they play crucial roles in guaranteeing successful viral infections.

Further Reading

- Chinchar, V.G., Duffus, A.L.J., 2019. Molecular and ecological studies of a virus family (Iridoviridae) infecting invertebrates and ectothermic vertebrates. *Viruses* 11, 538.
- Chinchar, V.G., Hick, P., Ince, I.A., *et al.*, 2017a. ICTV virus taxonomy profile: Iridoviridae. *Journal of General Virology* 98, 890–891.
- Chinchar, V.G., Waltzek, T.B., Subramaniam, K., 2017b. Ranaviruses and other members of the family Iridoviridae: Their place in the virosphere. *Virology* 511, 259–271.
- Colson, P., De Lamballerie, X., Yutin, N., *et al.*, 2013. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Archives of Virology* 158, 2517–2521.
- Foster, L.J., 2011. Interpretation of data underlying the link between colony collapse disorder (CCD) and an invertebrate iridescent virus. *Molecular & Cellular Proteomics* 10. (M110-006387).
- Ince, I.A., Özcan, K., Vlak, J.M., van Oers, M.M., 2013. Temporal classification and mapping of non-polyadenylated transcripts of an invertebrate iridovirus. *Journal of General Virology* 94 (1), 187–192.
- Ince, I.A., Özcan, O., Ilter-Akulke, A.Z., Scully, E.D., Özgen, A., 2018. Invertebrate iridoviruses: A glance over the last decade. *Viruses* 10, E161.
- King, A.M., Lefkowitz, E., Adams, M.J., Carstens, E.B., 2011. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Amsterdam: Elsevier.
- Liu, Y., Tran, B.N., Wang, F., Ounjai, P., Wu, J., Hew, C.L., 2016. Visualization of assembly intermediates and budding vacuoles of Singapore grouper iridovirus in grouper embryonic cells. *Scientific Reports* 6, 18696.
- Tokarz, R., Firth, C., Street, C., Cox-Foster, D.L., Lipkin, W.I., 2011. Lack of evidence for an association between Iridovirus and colony collapse disorder. *PLOS ONE* 6, e21844.
- Williams, T., 2008. Natural invertebrate hosts of iridoviruses (Iridoviridae). *Neotropical Entomology* 37, 615–632.
- Wu, J., Chan, R., Wenk, M.R., Hew, C.L., 2010. Lipidomic study of intracellular Singapore grouper iridovirus. *Virology* 399, 248–256.
- Yan, X., Olson, N.H., Van Etten, J.L., *et al.*, 2000. Structure and assembly of large lipid-containing dsDNA viruses. *Nature Structural & Molecular Biology* 7, 101–103.
- Yan, X., Yu, Z., Zhang, P., *et al.*, 2009. The capsid proteins of a large, icosahedral dsDNA virus. *Journal of Molecular Biology* 385, 1287–1299.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsDNA-viruses/w/iridoviridae
 Iridoviridae.
 Iridoviridae.
 dsDNA Viruses.

Mesoniviruses (*Mesoniviridae*)

Jody Hobson-Peters and Daniel Watterson, Australian Infectious Diseases Research Centre, School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD, Australia

© 2021 Elsevier Ltd. All rights reserved.

Classification

The *Mesoniviridae* is a new family of positive sense single-stranded RNA viruses belonging to the order *Nidovirales*. The *mesonivirus* genome is organized similarly to other nidoviruses and its size of approximately 20–21 kb falls intermediately between those of the large nidovirus families (*Coronaviridae*, *Roniviridae*: 26–32 kb) and the small nidoviruses (*Arteriviridae*: 13–16 kb), prompting the name – *mesos* meaning “in the middle” in greek and “ni” referring to nidovirus. Indeed, the *mesoniviruses* represent an evolutionary bridge across a large gap in RNA genome size. Similar to other members of the order, the enveloped *mesonivirus* virions are often observed with projections, at times with large globular heads attached to low density stalks. Mesoniviruses have no known disease association and host range analysis indicates that their replication is restricted to mosquitoes and that they are unable to replicate in vertebrate cells in culture. Mesoniviruses have been isolated from a variety of mosquito species collected from diverse geographical locations, suggesting that these viruses are likely to be very common in mosquito populations. However, most of the *mesonivirus* species have been isolated from *Culex* spp. mosquitoes (Table 1).

Current taxonomy for the *Mesoniviridae* family categorizes the viruses into a single genus, *Alphamesonivirus*, within the subfamily *Namcalivirus* (ICTV, 2018). Within eight subgenera, there are nine species designated *Alphamesoniviruses* 1–9 (Table 1). These assignments were based on genomic sequence analysis using the computational comparative genomics framework DEmARC (DivErsity pArtitioning by hierarchical Clustering) and five replicative protein domains (3CLpro, NiRAN, RdRp, ZBD, and HEL1) that are characteristic of nidoviruses.

The type species, *Alphamesonivirus* 1, accommodates the two closely related viruses Nam Dinh virus and Cavally virus that were both described in 2011 and isolated from various mosquito species collected in Vietnam and Cote d’Ivoire, respectively. New isolates of *Alphamesonivirus* 1 were subsequently cultured from mosquitoes collected in Australia, Austria, China, Indonesia, Mexico, Thailand, and U.S.A (Table 1). Literature suggests that *Alphamesonivirus* 1 is the most frequently detected *mesonivirus* across both geographic location and mosquito genera (Table 1).

The remaining 8 classified species are represented by mostly single reports of the species from one country and often isolated from a single genus of mosquito (Table 1). Apart from the isolates of *Alphamesonivirus* 1, only Casuarina virus (*Alphamesonivirus* 4) and Nsé virus (*Alphamesonivirus* 8) have been isolated from multiple genera of mosquitoes.

Three viruses, Yichang, Dianke, and Moumo viruses, remain to be assigned (Table 1). Yichang virus was detected at a frequency of 16.5% from pools of mosquitoes collected in Hubei, in China. Pairwise evolutionary distances (PED) based on the conserved domains of ORF1b indicate that this virus is divergent from those that have already been classified. Several attempts to isolate Moumo virus from pools of mosquitoes collected in Côte d’Ivoire were not successful. However, a sequence fragment of 7985 nt was elucidated and enabled phylogenetic comparison with other *mesoniviruses*.

Using antiserum produced to a mix of *Alphamesoniviruses* 1 and 8, conservation of antigenic sites in the S and M proteins between *Alphamesoniviruses* 1, 5, 8, and 9 was confirmed. The phylogenetic relationship between the four *mesonivirus* species was further supported with cross-reactivity analysis in Western blot using numerous peptide-specific antisera, whereby cross-reactivity was more likely to be observed between those viruses that were more closely related genetically.

Virion Structure

Electron microscopy studies are limited but indicate that *mesonivirus* virions are spherical, enveloped viruses, ranging in size from 50 to 120 nm (Table 1). Only one study has assessed *mesonivirus* particles by cryo-electron microscopy, determining Casuarina virus (*Alphamesonivirus* 4) virions to have a diameter of 65 nm. On the surface of the particles, most *mesoniviruses* carry large club-shaped projections. Cryo-tomographic analysis revealed that these spike-like projections are attached to the virion via a low density stalk and display a globular head, with distinct similarities to the well characterized spikes of the coronaviruses, severe acute respiratory syndrome coronavirus (SARS-CoV) and murine hepatitis virus (MHV) (Fig. 1). The reported size of the projections ranges from 3 to 4 nm for *Alphamesonivirus* 1 through to 15 nm for *Alphamesonivirus* 4, but may be attributable to differences in imaging (Table 1). Studies using *Alphamesonivirus* 4 indicate that these viruses lack a dense nucleoprotein core complex.

Genome

Mesoniviruses have a positive sense single stranded RNA genome with a 5′ cap and a 3′ poly A tail. A feature consistent with other nidoviruses is the organization of the *mesonivirus* genome into discrete ORFs, each encoding separate functional domains. The *mesonivirus* genome encodes seven ORFs, ordered ORF1a-ORF1b-ORF2a-ORF2b-ORF3a-ORF3b-ORF4, which are bounded by

Table 1 Species in the Family Mesoniviridae, Genus Alphamesonivirus

Subgenus	Species	Name identification	Isolation region	Date of mosquito collection	Mosquito species	Titer in C6/36 cells/ml	Particle morphology
Namcalivirus	Alphamesonivirus-1	Cavally virus	Côte d'Ivoire	2004	<i>Culex</i> spp. <i>Aedes</i> spp. <i>Anopheles</i> spp. <i>Uranotaenia</i> spp. <i>Culex vishnui</i> <i>Culex tritaeniorhynchus</i>	10 ^{9.7} TCID ₅₀	120 nm with 12 nm spikes
		Nam Dinh virus	Vietnam	2002	<i>Cx. annulirostris</i> , <i>Cq. xanthogaster</i> , <i>Ae. vigilax</i>		60-80 nm
			Australia	2007 – 2014	<i>Uranotaenia unguiculata</i> <i>Culex quinquefasciatus</i> <i>Culex quinquefasciatus</i>		50-80 nm, 3 nm spikes
			Austria China U.S.A	2016–2017 2004, 2010	<i>Ae. albopictus</i> <i>Culex quinquefasciatus</i> <i>Culex quinquefasciatus</i> <i>Ae. taeniorhynchus</i> <i>Culex pipiens</i>		
		Houston virus	Mexico Mexico	2017 2007/08	<i>Cx. vishnui</i> <i>Cx. tritaeniorhynchus</i> <i>Cx. vishnui</i> <i>Cx. australicus</i> <i>Mosquito pool</i>		55–65 nm
		Bontang Baru virus	South Korea Indonesia	2012 1981			~50 nm
		Ngewotan virus	Indonesia	1981			~50 nm
		Kamphaeng Phet virus	Australia Thailand	2015 1984–85			80 nm
							65 nm with 15 nm spikes
Karsalivirus	Alphamesonivirus-2	Karang Sari virus	Indonesia	1981	<i>Cx. vishnui</i>	10 ¹¹ PFU	80 nm
Karsalivirus	Alphamesonivirus-3	Dak Nong virus	Vitenam	2007	<i>Cx. tritaeniorhynchus</i>	10 ^{8.5} TCID ₅₀	65 nm with 15 nm spikes
Casualivirus	Alphamesonivirus-4	Casuarina virus	Australia	2010, 2006	<i>Cq. xanthogaster</i> <i>Culex annulirostris</i>		
Hanalivirus	Alphamesonivirus-5	Hana virus	Côte d'Ivoire	2004	<i>Culex</i> spp.	10 ^{9.3} TCID ₅₀	120 nm with 12 nm spikes
Ofalivirus	Alphamesonivirus-6	Ofaie virus	Brazil	2010	<i>Mansonia</i> spp.		
Kadilivirus	Alphamesonivirus-7	Kadiweu virus	Brazil	2010	<i>Culex</i> spp.		
Enselivirus	Alphamesonivirus-8	Nsé virus	Côte d'Ivoire	2004	<i>Culex</i> spp. <i>Aedes</i> spp. <i>Anopheles</i> spp. <i>Anopheles</i> spp.	10 ^{8.8} TCID ₅₀	
Menolivirus	Alphamesonivirus-9	Méno virus	Côte d'Ivoire	2004	<i>Uranotaenia chorleyi</i>	10 ^{8.3} TCID ₅₀	120 nm with 12 nm spikes
Unassigned		Yichang	China	2014	<i>Culex</i> spp. <i>Armigeres</i> spp. <i>Anopheles</i> spp.	10 ⁷ PFU	80 nm with 10 nm spikes
Unassigned		Moumo virus	Côte d'Ivoire	2004	<i>Cx. nebulosus</i>		65 nm with 5 nm spikes
Unassigned		Dianke virus	Senegal	2012/2013	<i>Anopheles</i> spp. <i>Culex</i> spp. <i>Aedes</i> spp. <i>Mansonia</i> spp.		5 nm spikes

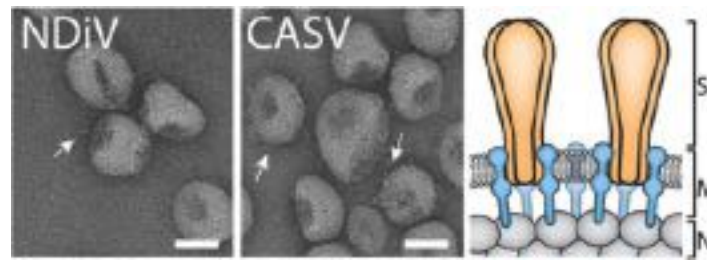


Fig. 1 *Mesonivirus particle morphology and spike protein architecture.* Left and middle panels: Negative stained transmission-electron micrographs of Nam Dinh virus (Alphamesonivirus 1) and Casuarina virus (Alphamesonivirus 4) particles. Spikes on the virion surface are highlighted with white arrows. Right panel: Proposed *mesonivirus* virion architecture, with trimeric spike proteins (S) projecting from an ordered M protein/lipid bilayer support. Scale bars indicate 50 nm. Images produced by Dr Daniel Watterson, The University of Queensland, Australia.

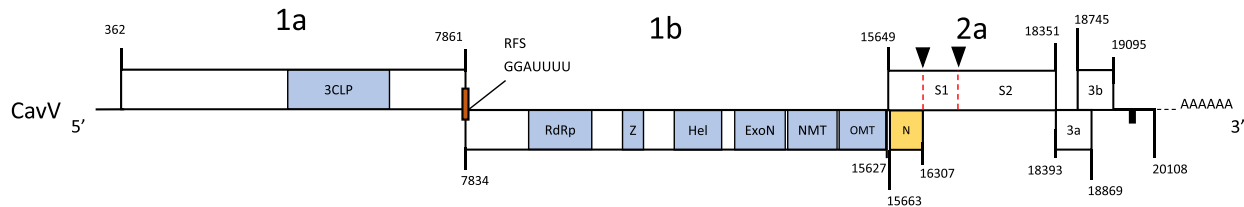


Fig. 2 *Mesonivirus genome organization.* Schematic representation of the typical genome of *mesoniviruses* using Cavally virus (Alphamesonivirus 1) as the representative. The open reading frames are shown by boxes. In ORFs 1a and 1b, the functional domains (3C-like serine protease (3CLP), RNA-dependent RNA polymerase (RdRp), zinc-binding (Z), helicase (Hel), exoribonuclease (ExoN), N7-methyltransferase (NMT) and 2'-O-methyltransferase (OMT)) are shaded gray and the ribosomal frame shift (RFS) element is shown. ORF2a encodes the S glycoprotein, which is cleaved (indicated by inverted triangles) to generate S1, S2 and an uncharacterized N-terminal fragment. An alternative ORF (2b) is encoded at the N-terminal end of ORF2a and expresses the N protein (colored in orange). The putative membrane protein is encoded by ORF3a. The function of the predicted protein encoded by ORF3b is not known and the putative ORF4 is indicated by a solid black rectangle.

5' and 3' untranslated regions (Fig. 2). The two largest ORFs, ORF1a and ORF1b, are located at the 5' end of the genome and overlap by a few nucleotides, with translation regulated by a ribosomal frameshift motif (RFS), with the sequence GGAUUUU, allowing the two respective polypeptides, pp1a and pp1ab, to be expressed. A second possible RFS also exists between ORFs 3a and 3b.

Downstream of ORF1b are multiple smaller ORFs that code for the virion proteins. These ORFs at the 3' end of the genome are expressed from two subgenomic RNAs, as confirmed by Northern blot. Transcription regulating sequences (TRSs) have been identified which are predicted to control the synthesis of the subgenomic RNAs.

ORF1a/ORF1b

ORFs 1a and 1b code for the key replicative enzymes. The replicase genes include conserved functional domains that are present in *mesoniviruses* and larger *nidoviruses*. The viral 3C-like chymotrypsin-like protease (3CL^{pro}) domain is flanked by transmembrane domains (TM) and is translated from ORF1a (Fig. 2). Consistent with the role of 3CL^{pro} for other *nidoviruses*, studies with Alphamesonivirus 1 3CL^{pro} have confirmed the autocatalytic activity of 3CL^{pro}, as well as its role in the processing of the enzymes encoded by ORF1b (Fig. 2). However, the *mesonivirus* 3CL^{pro} does feature distinct substrate specificities in comparison to other *nidoviruses*. Another attribute of some *mesoniviruses* is that the N terminus of ORF1a can be variable in length, due to the presence of large block insertions of up to almost 200 amino acids for some viruses such as Yichang, Karang Sari, and Bontang Baru viruses. Apart from variations on the sequence SKRKKG commonly seen at the termini of these insertions, function of these insertions is unclear.

Encoded by ORF1b are the remaining replicative enzymes – a RNA-dependent RNA polymerase (RdRp), a complex zinc-binding motif (Z) linked to a superfamily 1 helicase (Hel), 3' to 5' exoribonuclease (ExoN), N7-methyl-transferase (NMT) and 2'-O-methyltransferase (OMT) from ORF1b (Fig. 2). As detailed above, these enzymes are processed from the polyprotein by 3CL^{pro} and putative host cell protease(s). NendoU is the only replicase domain not encoded by *mesoniviruses*, but for which is part of the typical constellation of seven conserved domains of other *nidoviruses*.

Of note is the acquisition of ExoN by an ancestral group from which the *mesoniviruses* descended. This facilitated the required increase in replication fidelity, necessary for the evolution of the *coronaviruses*, which boast the largest RNA genome size. It is the presence of conserved functional ExoN and OMT domains within the genomes of *mesoniviruses* and other larger *nidoviruses* which distinguish them from the smaller *nidoviruses*.

Structural Proteins

The structural and accessory proteins are encoded by five ORFs at the 3' end of the *mesonivirus* genome and are translated from subgenomic mRNAs (Fig. 2). Analysis of the *mesonivirus* genome suggests the presence of eight structural proteins, including glycoprotein spike (S protein, S1 and S2 subunits) (ORF2a), nucleocapsid protein (N) (ORF2b), and four membrane-spanning proteins (M protein – p17, p18, p19, and p20) (ORF3a). Protein expression analysis could not detect proteins from ORF3b or ORF4.

The spike protein encoded by ORF2a is translated as a precursor polypeptide, which undergoes post-translational proteolytic processing to form the spike proteins. N-terminal sequence analysis of Cavally virus and Dak Nong virus spike proteins confirmed the processing of the ORF2a peptide by a signal peptidase to create S (p77–80) and a furin-like protease to generate the two spike proteins, S1 (p23) and S2 (p55–57), which share a common N terminus and C terminus with S, respectively. While the signalase cleavage site appears to be relatively variable between the *mesoniviruses*, the furin-like motif is conserved. The S1 and S2 proteins are post-translationally modified by N-linked glycosylation.

The extreme N-terminus of ORF2a translation product has not been detected, either as part of a full-length ORF2a polypeptide, or the signalase-cleaved product and exhibits high sequence variability between the *mesoniviruses*. An alternative ORF (ORF2b) is, however, encoded within this region (Fig. 2). Translation of this ORF results in the production of the putative 25 kDa hydrophilic nucleoprotein.

Immediately downstream of ORF2a are two overlapping ORFs – ORF3a and ORF3b, the former encoding the putative glycosylated membrane (M) protein. In silico analysis of ORF3b proteins suggests features consistent with a class II transmembrane protein, although there is poor sequence conservation and a function of the protein remains to be established. In the 3'-terminal end of the genome, most *mesoniviruses* encode an additional putative ORF (ORF4). While it is not known if ORF4 is expressed in cells and analysis of the putative protein does not yield any indication of possible function, there is a high level of sequence identity at the N-terminal region. Experiments using anti-serum produced to the first 223 aa of ORF2a, and the predicted proteins of ORF3b and ORF4 of Cavally virus did not provide any evidence for the presence of these proteins in virus-infected cells.

Life Cycle

The tropism of *mesoniviruses* is restricted to insect cells, and perhaps only to mosquitoes, although robust analysis for replication in a variety of insect cells is warranted. Numerous in vitro infection studies have been performed on a range of mammalian and avian cell lines to provide evidence that *mesoniviruses* do not replicate in vertebrate cell lines. In mosquito cell culture, the viruses cause varying degrees of cytopathic effect depending on the virus. In the *Aedes albopictus* mosquito cell line, C6/36, *mesoniviruses* replicate to very high titers, frequently exceeding 10^8 infectious units per ml (Table 1). The replication in C6/36 cells is also rapid. Within 24 h, 10^9 to 10^{10} RNA copies/ml can be detected in culture supernatant, peaking at 10^{11} RNA copies/ml by 48 h post-infection. The viruses will also replicate readily in cell lines of *Culex* mosquito origin, but to lower titers.

There is evidence for maintenance of insect-specific viruses of other families within the mosquito population via vertical transmission. The detection of Alphamesonivirus 1 RNA in both male and female mosquitoes provides some evidence that *mesoniviruses* are similarly transmitted from the female mosquito to her progeny. The detection of Dianke virus in pools of male mosquitoes further support the likelihood of vertical transmission. However, more studies in this area are warranted. It has been shown that *mesoniviruses* disseminate into the legs and wings of naturally infected mosquitoes and intriguingly the viral nucleic acid in mosquito saliva expectorated onto honey-baited cards was detected. These data suggest that there may be numerous routes of transmission between mosquitoes.

Reference

ICTV, 2018. ICTV Master Species List 2018b.v2. Available at: <https://talk.ictvonline.org/files/master-species-lists/m/msl/8266>.

Further Reading

Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Nidovirales: Evolving the largest RNA virus genome. *Virus Research* 117 (1), 17–37.

Hall, R.A., Bielefeldt-Ohmann, H., McLean, B.J., *et al.*, 2016. Commensal viruses of mosquitoes: Host restriction, transmission, and interaction with arboviral pathogens. *Evolutionary Bioinformatics Online* 12 (Suppl. 2), 35–44.

Laufer, C., Goeman, J.J., del Carmen Parquet, M., *et al.*, 2013. The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathogens* 9 (7), e1003500.

Nimaviruses (*Nimaviridae*)

Peter J Krell, Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Emine Ozsahin, University of Guelph, Guelph, ON, Canada

© 2021 Elsevier Ltd. All rights reserved.

This is an update of J.-H. Leu, J.-M. Tsai, C.-F. Lo, White Spot Syndrome Virus, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00776-7.

Glossary

DNA dependent RNA polymerase (DdRp) An RNA polymerase that copies mRNA from a DNA only template.

Homologous repeat (hr) a sequence of DNA with several copies of imperfect palindromes, that are repeated two or more times in the genomes of baculoviruses or whispoviruses.

In situ hybridization (ISH) A technique that uses a labeled complementary DNA or RNA probe to hybridize and detect a specific DNA or RNA sequence in a section of tissue (in situ).

Internal ribosome entry site (IRES) A sequence in mRNA, either at the 5' end or internal between ORFs allowing for translation of a downstream internal ORF in a cap-independent manner.

Microarray A DNA microarray is a collection of microscopic dsDNA or oligonucleotide spots that are deposited on a solid glass slide at high density. Each spot commonly represents a single gene. DNA microarrays use DNA–DNA or DNA–RNA hybridization to perform simultaneous large-scale analyses of the expression levels of the corresponding genes.

RNAi (RNA interference) RNAi refers to the introduction of homologous double-stranded RNA to specifically interfere with the target gene's expression in the cells/organisms. It was originally discovered in *C. elegans*, and is now widely used in many organisms for null mutations. Recent studies suggest that RNAi can be applied in shrimp as well through intramuscular injection of dsRNA.

Introduction

White spot syndrome viruses (WSSVs) cause White spot disease resulting in a major economic problem in worldwide industrial shrimp and crayfish fisheries. They have a wide host range causing lethal diseases in shrimp, prawns, crabs, lobster, crayfish and copepods. The first whispovirus to be documented was called white spot bacilliform virus, now named White spot syndrome virus-CN. As evidenced by earlier names of isolates such as Chinese baculo-like-virus, *Penaeus monodon* non-occluded baculoviruses II and III and white spot baculovirus, these viruses were originally thought to be baculoviruses and were classified with the *Baculoviridae*. However due to unique features in their morphology and phylogeny they were placed in a new family, *Nimaviridae* in 2002.

Taxonomy

The *Nimaviridae* family has a single genus, *Whispovirus*, with a single and type species *White spot syndrome virus*, all taxa names approved by the International Committee on Taxonomy of Viruses (ICTV) in 2002. Mercifully there are no higher orders of classification (Order, Class, Phylum, Kingdom, Realm) for *Nimaviridae*. There is one exemplar virus in the species, white spot syndrome virus CN (WSSV-CN). Thirteen additional viruses are recognized by ICTV, though many more isolates have been characterized. While the etymology of the species name is more self-evident (from white spot syndrome virus), that of the family name relates to the viral morphology not the disease. A distinguishing and unique feature of the whispoviruses is that they have a polar thread-like tail, “nima” in Greek, which gave rise to the family name *Nimaviridae*. Prior to 2002, due to the circular nature of the large dsDNA genome and similarity in morphology to baculoviruses, the white spot syndrome viruses were initially misclassified in the *Baculoviridae* (subfamily *Nudibaculovirinae*, genus *Nonoccluded baculovirus*). However, their unique morphology, genome structure and phylogeny showed they were clearly distinct from the baculoviruses and were deserving of their own *Nimaviridae* family.

Virion Structure and Composition

WSSV virions are enveloped, ellipsoid to bacilliform in shape measuring 70–170 nm diameter and 210–420 nm in length with a rod-shaped nucleocapsid 50–80 nm in diameter and 300–420 nm long (Figs. 1 and 2(b)). One end of the nucleocapsid appears open and flattened and the other end is more rounded. By electron microscopy the nucleocapsids have a distinct pattern of a

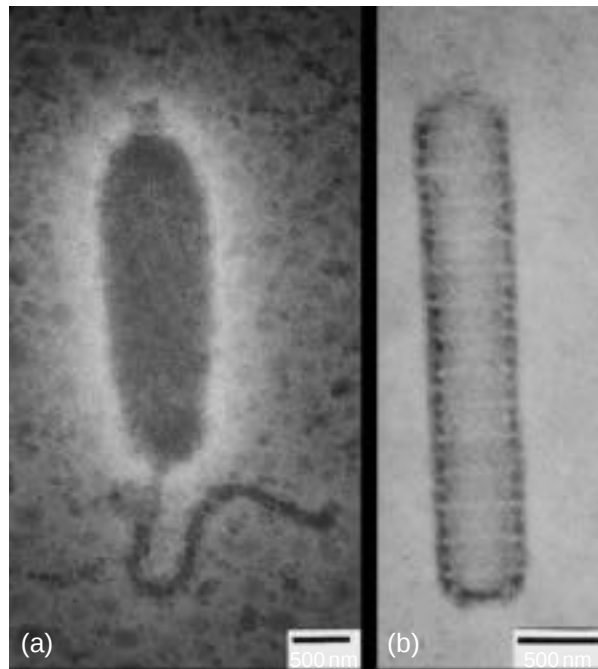


Fig. 1 Morphology of WSSV virions (a) intact WSSV virion with tail like extension at the smaller end. (b) WSSV nucleocapsid showing the stacked ring structures made up of two rows of VP664-based globular subunits. From Leu, J.H., Tsai, J.M., Lo, C.F., 2008. White spot syndrome virus. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. pp. 450–459.

stacked series of about 16 rings, each about 20 nm in thickness and composed of two rows of 12–14 globular units each about 10 nm in diameter. Between the envelope and nucleocapsid is a tegument layer. A thread-like tail, or “nima” responsible for the family name, extends from the narrower end of the virus. WSSV virions have a buoyant density of 1.2 g cm^{-3} and nucleocapsids have a buoyant density of 1.31 g cm^{-3} and are quite stable in pond water, being viable for 3–4 days.

SDS PAGE (Fig. 2(a)) and proteomic analyses have identified about 40 structural proteins ranging from 11 to 664 kDa in size. WSSV virions have a complex SDS-PAGE protein profile (Fig. 2(a)). Early structural protein studies used SDS-PAGE coupled with Western blot analyses and/or N-terminal sequencing of proteins to identify at least six structural proteins. More recent studies have used mass spectrometry to analyze the protein profiles of purified WSSV virions separated by SDS-PAGE and 2-D electrophoresis, and more than 40 structural proteins have now been identified. The major structural proteins are VP664, VP28, VP26, VP24, VP19, and VP15 (Fig. 2(a, b)). VP15 is a basic capsid protein with *in vitro* DNA-binding activity, and it has high homology to the lysine- and arginine-rich DNA-binding proteins of the insect baculoviruses. This protein is therefore thought to be responsible for packaging of the WSSV DNA into the nucleocapsid. VP26, one of five tegument proteins, can interact with nine host proteins including actin, but the importance of this interaction remains unknown. VP26 interacts with beta-integrin to potentially facilitate virus entry, and antiviral proteins such as C-type lectins (MjLecA and LvCTL1) and prohibitins (PHBs). VP19 and VP28 are envelope proteins, and an *in vitro* physical interaction between VP28 and tegument proteins VP24 and VP26 has been reported. VP664, the major capsid protein, takes its name from its calculated molecular mass of 664 kDa. Not only is it the largest protein encoded by the WSSV genome, but it is also the largest viral structural protein ever found. This protein is encoded by the intronless giant ORF (*wssv419*) of 18,234 nt. Immunoelectron microscopy of purified virions has shown that VP664 forms the globular subunits visible in the nucleocapsid ring structures. VP664 antibodies bind to the nucleocapsid but not envelope of purified virions (Fig. 3). In addition to VP15, a very basic DNA binding nucleoprotein, and VP664, the nucleocapsid contains at least four other minor proteins: VP160A (WSSV344), VP160B (WSSV094), VP60B (WSSV474), and VP51C (WSSV364) (Fig. 2(b)).

Neutral lipids, phosphatidylcholine and phosphatidylethanolamine form the major lipids of the envelope, along with the fatty acids stearic acid, eicosapentaenoic acid and docosahexaenoic acid.

Genome and Phylogeny

The nimavirus genome is a large, circular dsDNA (Fig. 2(a)), one of the largest viral DNAs published rivaling that of some phycodnaviruses which range from 154,641 bp for a phaeovirus to 459,984 bp for a prymnesiovirus. There are currently 15 fully sequenced nimavirus genomes accessible from Genbank (Table 1), showing a range in size from 280,591 bp for WSSV-IN-AP4RU from India to 309,286 bp for WSSV-CN01 from China. Note, not included in Table 1 is the genome sequence for WSSV-EG3 (KR083866.1) since it is a duplicate of the WSSV-CN (AF332093) genome and the supporting publication for the WSSV-EG3

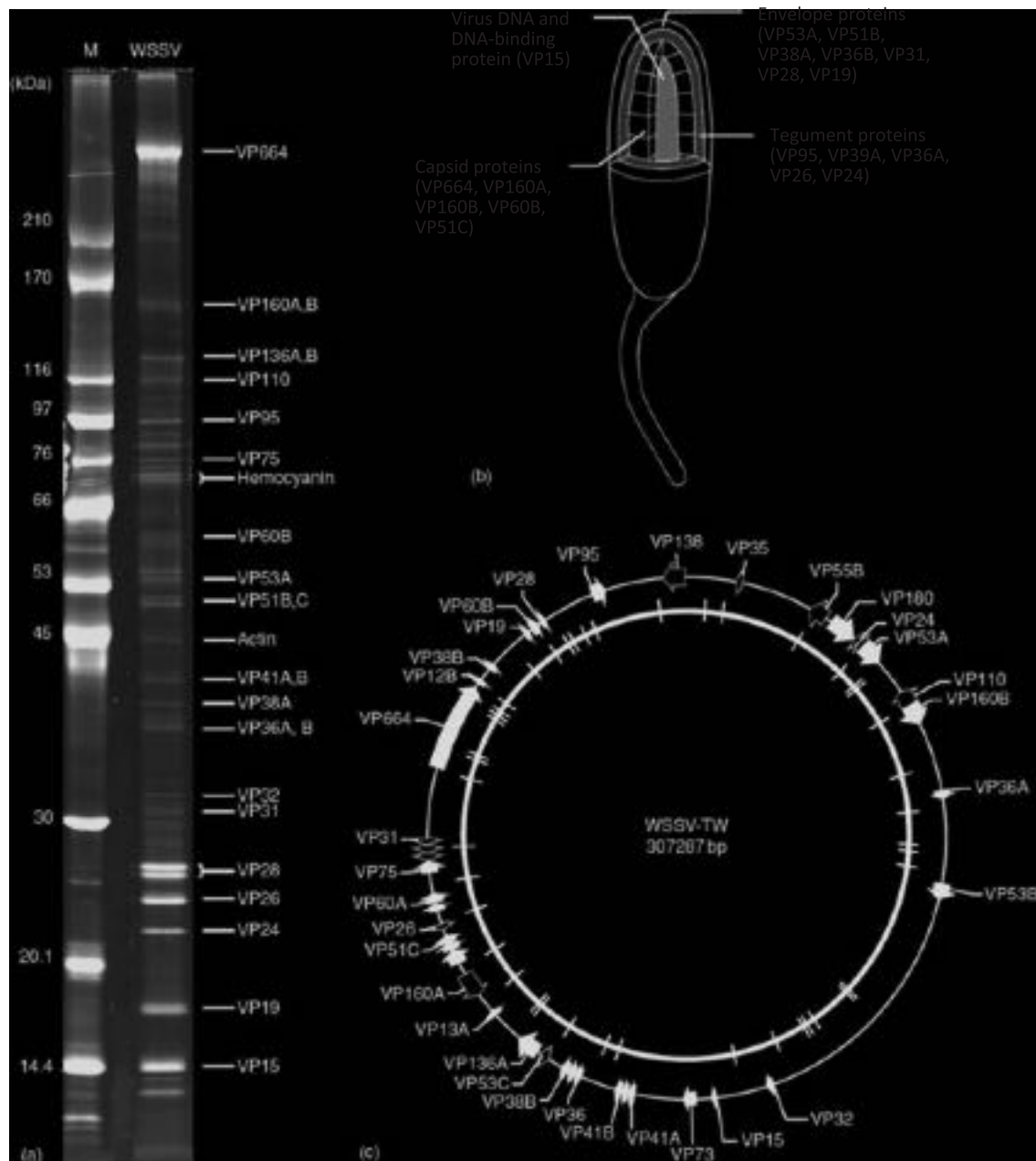


Fig. 2 WSSV structural proteins and genome. (a) SDS-PAGE profile of purified WSSV virions. The six major and some minor structural protein bands are indicated. The bands corresponding to two host proteins, hemocyanin and actin co-purified with the virion are also indicated. (b) A proposed schematic diagram to show the WSSV virion structure and the location of some WSSV structural proteins and its DNA. (c) Distribution of 33 WSSV structural protein genes in the WSSV-TW genome. The inner circle shows predicted HindIII restriction enzyme cutting sites. Modified from Leu, J.H., Tsai, J.M., Lo, C.F., 2008. White spot syndrome virus. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. pp. 450–459.

genome was withdrawn by the authors. Also not included in this table and article are nimavirus-like endogenous genomes found in host genomes.

All genomes have a G + C content of about 41%. Although most WSSV genome sequences are unique, 3% of the genome consists of highly repetitive sequences, which are organized into about nine homologous regions (*hrs*) with the number of repeats of about 250–300 bp within an *hr* varying among isolates and localized to intergenic regions. The *hrs* include a variety of repeats including direct repeats, atypical inverted repeats and imperfect palindromic sequences. Their function in the nimaviruses is unknown but in baculoviruses, *hrs* have been implicated in DNA replication and as enhancers of early gene transcription. Differences in genome lengths have been attributed to various indels, like the insertion of a 1.3 kb transposase sequence in the genome of WSSV-TW and deletions in the WSSV-TH genome and in variations in the number and size of *hrs*.

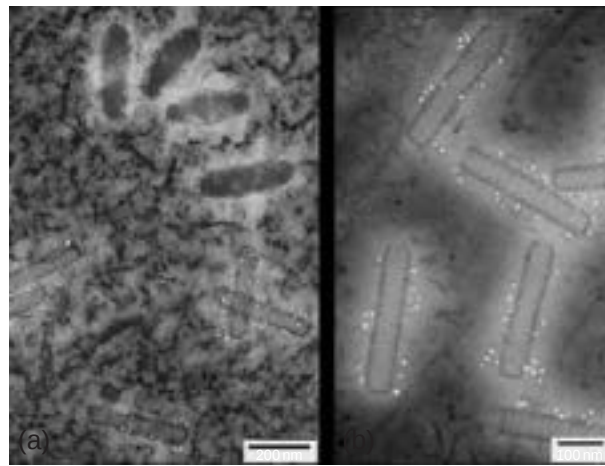


Fig. 3 Immunoelectron microscopy analysis of purified WSSV virions detected with VP664 antibody followed by gold-labeled secondary antibody. (a) The anti-VP664 antibody specifically binds to the nucleocapsid and not to the viral envelope. (b) Most of the gold particles are localized at the perimeter of the nucleocapsid. From Leu, J.H., Tsai, J.M., Lo, C.F., 2008. White spot syndrome virus. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), Encyclopedia of Virology, third ed. pp. 450–459.

Table 1 Characteristics of fifteen WSSV isolates and their complete genomes (derived from Genbank as of Aug 31, 2020)

Virus Abbreviation	Country of Origin	Species source	Year collected	Date seq submitted	Sequence Length	Protein number	% GC	Accession Number
WSSV-AU	Australia	<i>Penaeus monodon</i>	2016	2017 Aug	285,973	904 ^a	41	MF768985.1
WSSV-CN	China	<i>Marsupenaeus japonicus</i>	1996	2014 Nov	305,119	524 ^a	41	AF332093.3
WSSV-CN01	China	<i>Marsupenaeus japonicus</i>	1994	2015 Nov	309,286	177	40.9	KT995472.1
WSSV-CN02	China	<i>Procambarus clarkii</i>	2010	2015 Nov	294,261	164	41	KT995470.1
WSSV-CN03	China	<i>Litopenaeus vannamei</i>	2010	2015 Nov	281,148	154	41	KT995471.1
WSSV-CN04	China	<i>Marsupenaeus japonicus</i>	2012	2017 Mar	281,054	157	41	KY827813.1
WSSV-CN – 95-DFPE ^b	China	<i>Litopenaeus vannamei</i>	1995	2019 Dec	305,094	173	41	MN840357.1
WSSV-CN-Pc	China	<i>Procambarus clarkii</i>	2015	2016 Aug	300,223	191	41	KX686117.1
WSSV-CN-PC02 ^c	China	<i>Procambarus clarkii</i>	2018	2018 Jul	287,179	180	41	MH663976.1
WSSV-EC – 15098	Ecuador	<i>Litopenaeus vannamei</i>	Pre 2017	2018 Mar	288,997	150	40.9	MH090824.1
WSSV-IN-AP4RU	India	<i>Litopenaeus vannamei</i>	2014	2017 Dec	280,591	442 ^a	40.8	MG702567.1
WSSV-K-LV1	South Korea	<i>Litopenaeus vannamei</i>	2011	2012 Nov	295,884	515 ^a	40.9	JX515788.1
WSSV-MEX2008	Mexico	<i>Litopenaeus vannamei</i>	2008	2015 Nov	300,087	184	41	KU216744.2
WSSV-TH	Thailand	Not reported	1996	2001 Mar	292,967	184	41.1	AF369029.2
WSSV-TW	Taiwan	<i>Penaeus monodon</i>	1994	2005 Mar	307,287	532 ^a	41	AF440570.1

^aprotein numbers in excess of 200, which may be due to over-annotation, including overlapping ORFs in the count.

^bFrom archival paraffin embedded tissues of shrimp infected with a China CN95 strain of WSSV.

^cWSSV-CN-PC02 was originally named *Procambarus clarkii* virus. In this review the abbreviation WSSV-CN-PC02 is used to maintain the consensus in the naming of WSSV isolates beginning with WSSV followed by letters indicating its origin (CN for China), Pc for *Procambarus clarkii* and isolate number, 02 since another WSSV-CN-Pc was previously described. Common names: *Penaeus monodon*, Asian tiger shrimp or black tiger shrimp; *Marsupenaeus (Penaeus) japonicus*, Japanese tiger prawn or Kuruma shrimp; *Litopenaeus vannamei*, Whiteleg shrimp; *Procambarus clarkii*, Red swamp crayfish.

By analysis of the 15 current whispovirus genomes using MAFFT alignment, the overall genome sequence identity between genomes is quite high, between 92.44% and 98.91% (Table 2). Despite the overall identities, the genomes separated into two phylogenetic clades (Fig. 4). The most populous clade is clade B including two subclades B1 and B2. For the four isolates in subclade B1, the % identities range from 95.6% to 97.3%, all above 95% for all and their genomes range from 281 to 287 kb in size. The ten isolates within subclade B2 have % identities ranging from 95.28 to 98.91, all above 95%. Genome sizes in this subclade are among the largest and range from 289 to 309 kb. Nevertheless, the % identities between WSSV genomes from subclade B2 compared to genomes from subclade B1 range from 93.97% to 97.33% suggesting they should be part of the same clade B. The single isolate in clade A, WSSV-IN-AP4RU from India, has the lowest level of % identity with any of the other 14 isolates, ranging from 92.44% to 93.62% and has the smallest genome at 280,591 bps. Moreover, WSSV-IN-AP4RU appears to be an earlier genome compared to all the other WSSV genomes. The clades/subclades described here mimic the clusters observed in

Table 2 Sequence identities among 15 WSSV complete genomes

Clade/Subclade	Nucleotide Sequence Identity (%)														
	A	B1					B2								
WSSV Isolate	IN-AP4RU	CN04	AU	CN03	CN-Pc02	MEX2008	TH	TW	CN02	CN01	EC-15098	CN-Pc	CN-95-DFPE	CN	K-LV1
Genome Size	280,591	281,054	285,973	284,148	287,179	300,087	292,967	307,287	294,261	309,286	288,997	300,223	305,094	305,119	295,884
WSSV-IN-AP4RU		93.62	92.96	92.58	92.44	93.35	93.56	93.29	93.22	93.36	93.31	93.10	93.35	93.38	93.50
WSSV-CN04	170,774		95.89	95.96	95.81	96.60	96.73	96.51	96.55	96.90	96.70	96.69	97.33	97.32	96.70
WSSV-AU	215,159	232,907		96.94	95.60	94.34	95.01	94.57	95.35	95.14	95.15	93.97	94.79	94.80	95.20
WSSV-CN03	170,764	277,931	234,036		97.30	95.34	95.80	95.73	95.62	95.92	95.78	94.88	95.83	95.82	95.94
WSSV-CN-Pc02	170,821	277,562	232,216	280,245		95.28	95.83	95.73	95.62	95.93	95.79	95.06	96.12	96.08	95.93
WSSV-MEX2008	220,481	234,685	283,889	235,543	238,270		96.46	96.63	96.49	97.03	96.82	95.87	96.53	96.50	96.54
WSSV-TH	218,455	234,854	279,251	235,773	238,532	291,649		97.70	97.16	97.96	97.66	96.37	97.36	97.32	97.58
WSSV-TW	193,491	262,460	256,409	263,406	266,138	270,440	263,434		97.28	98.18	97.80	96.75	96.86	96.81	97.62
WSSV-CN02	172,073	278,144	232,611	278,740	283,134	241,395	241,657	270,214		97.81	97.71	97.33	96.78	96.75	97.32
WSSV-CN01	174,648	279,703	234,639	280,370	282,785	245,735	245,826	274,566	292,768		98.39	97.45	97.35	97.29	97.85
WSSV-EC-15098	173,647	278,716	233,479	279,167	280,564	241,500	241,609	270,253	285,119	288,808		97.47	97.03	96.98	97.66
WSSV-CN-Pc	253,636	152,149	199,728	151,986	153,397	211,370	204,184	183,988	155,711	159,945	155,771		97.51	97.46	96.97
WSSV-CN-95-DFPE	175,508	280,738	235,388	281,159	283,644	246,838	247,113	275,773	293,674	302,806	288,556	160,970		98.91	97.32
WSSV-CN	175,388	280,660	235,256	281,090	283,559	246,705	246,978	275,629	293,686	302,659	288,419	160,899	304,896		97.31
WSSV-K-LV1	170,249	275,525	231,296	275,719	278,261	240,337	240,623	268,505	285,944	289,577	280,382	154,889	290,859	290,797	

Sequence identities in percentage above the diagonal, number of nucleotide match below the diagonal. Similarities (calculated from the distances identified by using ape package in R) and the number of the nucleotides that match in the alignments were calculated in R. White background for clade A (in row WSSV-IN-AP4RU only), light grey background for Subclade B1, and grey background for Subclade B2.

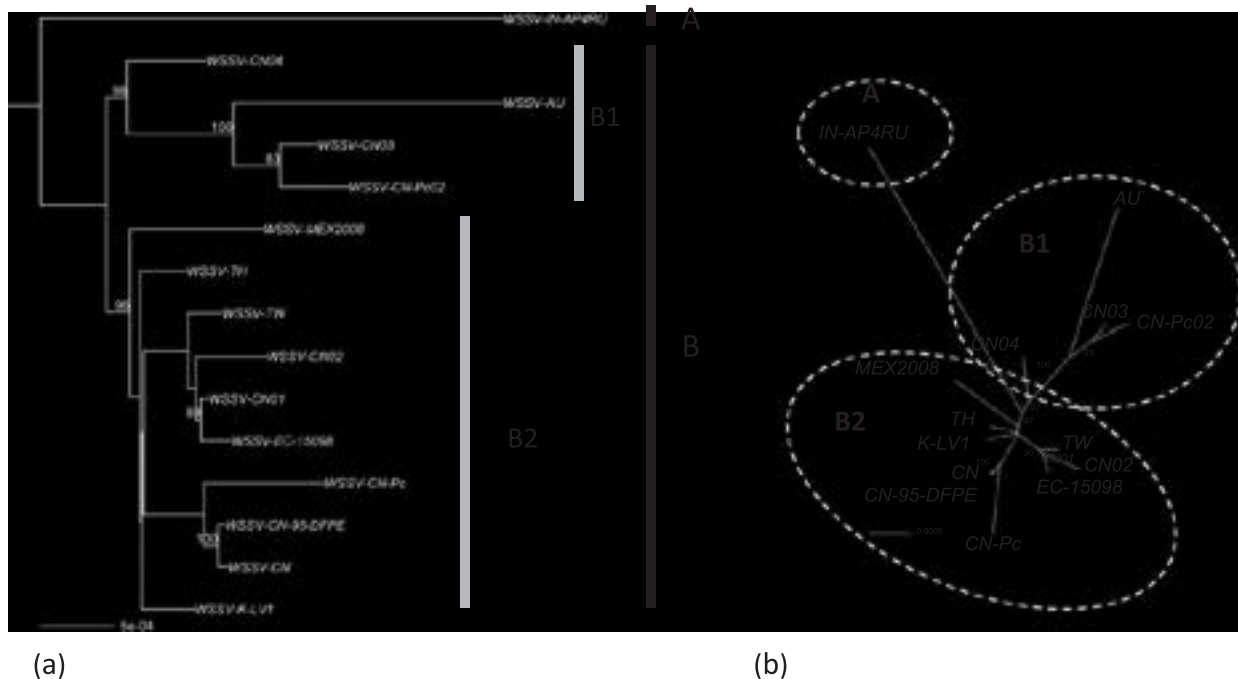


Fig. 4 Alignment of 15 WSSV genome sequences using MAFFT (version 7.397). Neighbor joining clustered trees were used to calculate the Maximum Likelihoods and bootstrap support values based on 100 bootstraps in R using the Phangorn package. (a) Ancestry relationships of strains (rooted tree). (b) Relatedness of the strains (unrooted tree). Virus clades are shown as A or B and subclades of B are shown as subclade B1 and B2. from: Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30 (4), 772–780. doi:10.1093/molbev/mst010. Schliep, K.P., 2011. Phangorn: Phylogenetic analysis in R. *Bioinformatics* 27 (4), 592–593. doi:10.1093/bioinformatics/btq706.

the unrooted phylogenetic tree (Fig. 3(b)) and for subclades B1 and B2 mimic the two clusters derived for nine genomes using ClustalW alignment reported by Oakey and Smith (2018).

Both strands of nimavirus DNAs have protein coding genes. For those genomes that appear to be well annotated, whipovirus genomes encode from 150 proteins for WSSV-EC-15098 to 191 proteins, for WSSV-CN-Pc (based on NCBI “Protein” tables) (Table 1). However, for some isolates, identified by “a” in Table 1, the number of proteins is much higher, between 442 for

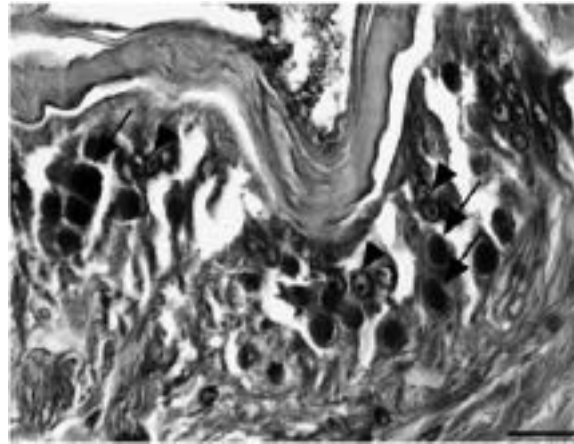


Fig. 5 Hematoxylin and eosin (HE) staining of tissue section from *Penaeus monodon* cuticular epithelium of eyestalk infected with WSSV. Degenerated cells characterized by hypertrophied nuclei (black arrows) are readily seen. The cells with normal nuclei are also indicated with arrowheads. Scale = 20 μ m. From Leu, J.H., Tsai, J.M., Lo, C.F., 2008. White spot syndrome virus. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), Encyclopedia of Virology, third ed. pp. 450–459.

WSSV-IN-AP4RU to 904 for WSSV-AU. These very high numbers, considering the genome size, suggest their genomes were over annotated, e.g., including smaller ORFs that clearly overlap larger ones and counting all ORFs even if fully overlapping.

Proteomic and genomic studies have identified genes for at least 48 structural proteins including 31 envelope proteins such as the major envelope proteins VP28 and VP19, five tegument proteins like the major tegument proteins VP26 and VP24 and nine nucleocapsid proteins like the major VP664 component of nucleocapsids and VP15, the major DNA binding protein. A number of nonstructural proteins were also identified. Of the 18 immediate early genes, *ie1*, *ie2*, *ie3*, and *ubiquitin E3 ligase* were identified. Genes for four antiapoptotic proteins (AAP), i.e., AAP1, 2, 3 and 4 were identified. Genes for at least nine enzymes were detected including for thymidylate synthase, protein kinases 1 and 2, a nuclease, DNA polymerase, two ribonucleotide reductases (RR1 and RR2), thymidine kinase/thymidylate kinase and a dUTPase. Given the large size of its genome, it is not surprising to find some additional interesting nimavirus genes including for two anti-melanization proteins, two latency-related genes, and genes for a histone-binding DNA mimic, mitochondria targeting protein and a TATA-box binding protein.

Life Cycle

WSSV replication occurs in the tissues of mainly ectodermal (cuticular epidermis, fore- and hindgut, gills and nervous tissue) and mesodermal (lymphoid organ, antennal gland, connective tissue and hematopoietic tissue) origin but not in tissues of endodermal origin (hepatopancreas and midgut). In the early stage of infection, affected cells display nuclear hypertrophy (Fig. 5), nucleoli dissolution, and chromatin margination, and the central area changes into a homogeneous eosinophilic region. The infected cells then develop an intranuclear eosinophilic Cowdry A-type inclusion, which becomes a light basophilic, denser inclusion separated by a transparent zone from the margined chromatin. The cytoplasm becomes less dense and more lucent. Later in infection, the nuclear membrane is disrupted, causing the intranuclear transparent zone to fuse with the lucent cytoplasm. At the end of cellular degeneration, the nucleus or whole cell disintegrates, leading to loss of cellular architecture. In moribund shrimp, most tissues and organs are heavily infected with the virus, and they exhibit severe multifocal necrosis.

Following infection, the first WSSV-positive signals detected are in the stomach, gills, cuticular epidermis, and connective tissue of the hepatopancreas (Fig. 6). At later stages of infection, the lymphoid organ, antennal gland, muscle tissue, hematopoietic tissue, heart, midgut, and hindgut also become positive. As infection proceeds, the stomach, gill, hematopoietic tissue, lymphoid organ, antennal gland, and cuticular epidermis become heavily infected with WSSV, which leads to serious damage and necrosis and/or apoptosis.

The entire WSSV replication cycle is nuclear and, in an acutely infected shrimp, can be completed within 24 h post infection. WSSV enters susceptible cells via caveola- or clathrin-mediated endocytosis facilitated, in part by the binding of the major envelope glycoprotein VP28 and cell receptors such as integrins α C1qR and Rab7 and a complex of about 11 virion proteins interacting with cell surface chitin-binding protein (CBP) which may also facilitate endocytosis. Several viruses use the Arg-Gly-Asp (RGD) motif to bind to cellular integrins during infection and this cell attachment site signature can be identified in at least six WSSV structural proteins. One of these six structural proteins, VP110 (WSV035), can attach to shrimp host cells, and this adhesion can be blocked by synthetic RGDT peptides, suggesting that the RGD motif in VP110 may play a role in WSSV infection.

Following endocytosis, WSSV travels through endosomes and as a consequence of acidification of the endosomes and the interaction between VP28 and Rab7, the nucleocapsids are released to the cytoplasm. By an unknown mechanism WSSV nucleocapsids enter the nucleus (structurally similar baculovirus rod-shaped nucleocapsids enter nuclei through the nuclear pore

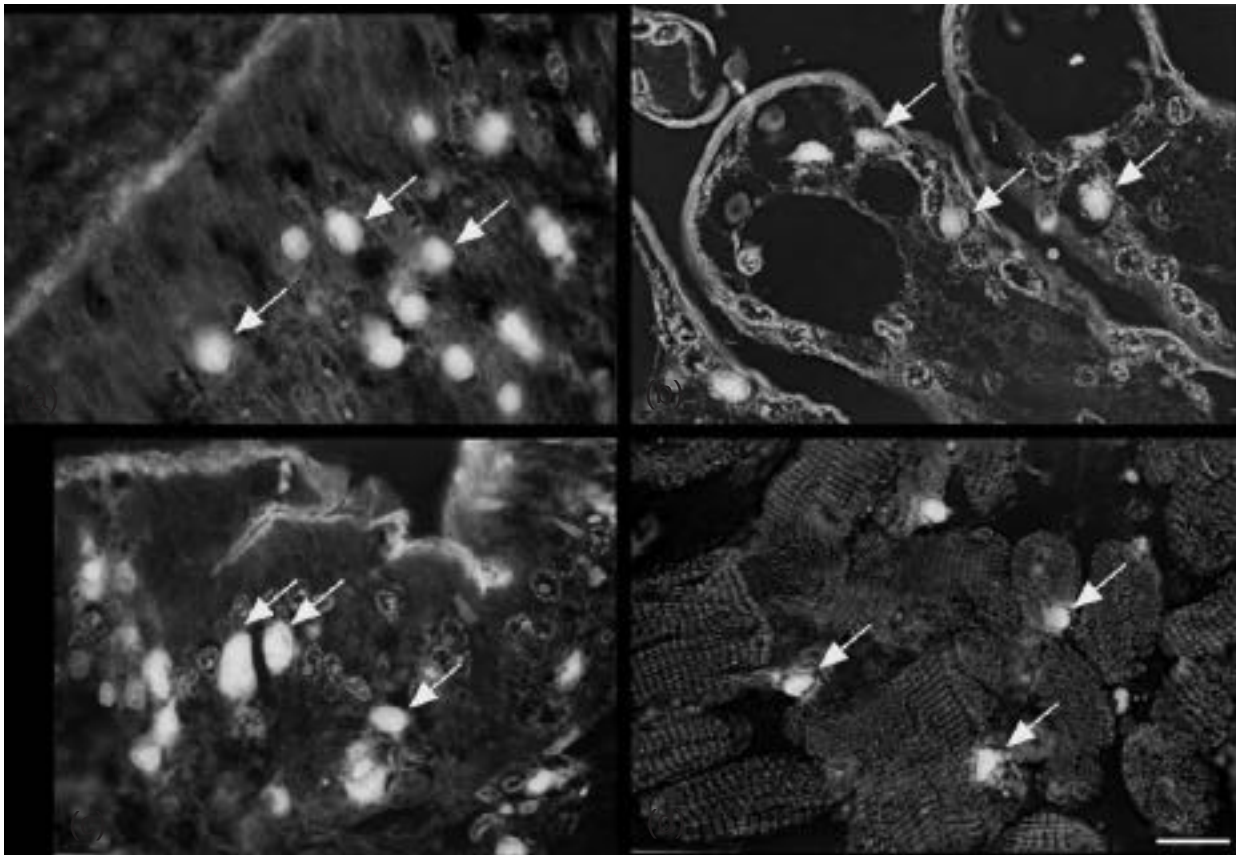


Fig. 6 *In situ* hybridization analysis of WSSV positive cells (arrows) in (a) the integument, (b) the gill, (c) the stomach, and (d) the heart from experimentally infected *Penaeus monodon* at 60 hpi. The infected cells shown here in the heart are connective tissue cells; muscle cells are not usually targeted by WSSV. Scale = 20 μ m. From Leu, J.H., Tsai, J.M., Lo, C.F., 2008. White spot syndrome virus. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. pp. 450–459.

or uncoat at the nuclear pore, so WSSV might use a similar nuclear entry mechanism). Once in the nucleus, transcription of viral genes can begin. Most virus RNAs are 5' capped and 3' polyA tailed, with no evidence of splicing. Some virion protein genes such as *vp31*, *vp39b*, and *vp11* are arranged in a cluster, producing polycistronic mRNAs which use internal ribosome entry sites (IRES) to regulate their translation. Several IRES elements have been identified, for example upstream of *vp28*, *wsv493* (*icp35*) and across the coding region of *vp31* and *vp39b*. Immediate early gene expression, like that for *immediate early gene 1* (*ie1*), is mediated by host transcription factors and RNAPolIII, initiating a temporal transcriptional cascade consisting of immediate early (IE), early (E), late (L) and very late (VL) phases. *ie1* genes, can be expressed in the presence of cycloheximide indicating their expression relies solely on host factors. The *ie1* promoter is characterized by an upstream TATA-box and binding site for the host transcription factor Sp1. About 16 early ORFs have transactivation activity, presumably assisting in later gene transcription. Structural gene promoters incorporate a degenerate ATNAC motif that could be involved in late gene transcription. Unlike baculoviruses, whispoviruses show no evidence of a viral DNA dependent RNA polymerase (DdRp) and must rely on the host RNAPolIII and host factors. Thus, early viral transcription factors like IE1, IE2 and IE3 might recognize late gene promoters recruiting the cellular RNAPol and transcription factors allowing for late gene transcription.

Viral and host miRNAs may be involved in WSSV infection and latency. For example, *M. japonicus* miR-7 targeted viral *wsv477*, an early gene involved in virus replication and when injected, reduced WSSV copies by 1000 fold late in infection. Studies of WSSV expression identified as many as 89 WSSV miRNAs, the majority in the early phases. Some miRNAs showed tissue specific expression. For example, WSSV-miR-N24 inhibits apoptosis by downregulating host *caspase 8*. WSSV-miR-66 targets viral *wsv094* and *wsv177* and WSSV-miR-68 targets *wsv248* and *wsv309* genes related to latency to virus replication.

Not much is known about whispovirus DNA replication. Nevertheless, the WSSV genome contains several genes identified through homology searches that encode proteins involved in DNA metabolism and replication. These genes include *thymidylate synthase*, *dUTPase*, *ribonucleotide reductases* (*rr1* and *rr2*; two separate ORFs encoding the two subunits), chimeric *thymidylate/thymidine kinase* (TK-TMK), and *DNA polymerase*. The chimeric *thymidylate/thymidine kinase* gene is a unique feature of WSSV, as these genes in other large DNA viruses are encoded in separate ORFs. The enzymatic activities of dUTPase, ribonucleotide reductase, and TK/TMK have been demonstrated by using purified recombinant proteins. Since it does not encode all genes needed for DNA replication, WSSV depends on the host for other replication proteins. It has therefore been proposed that WSSV, like some

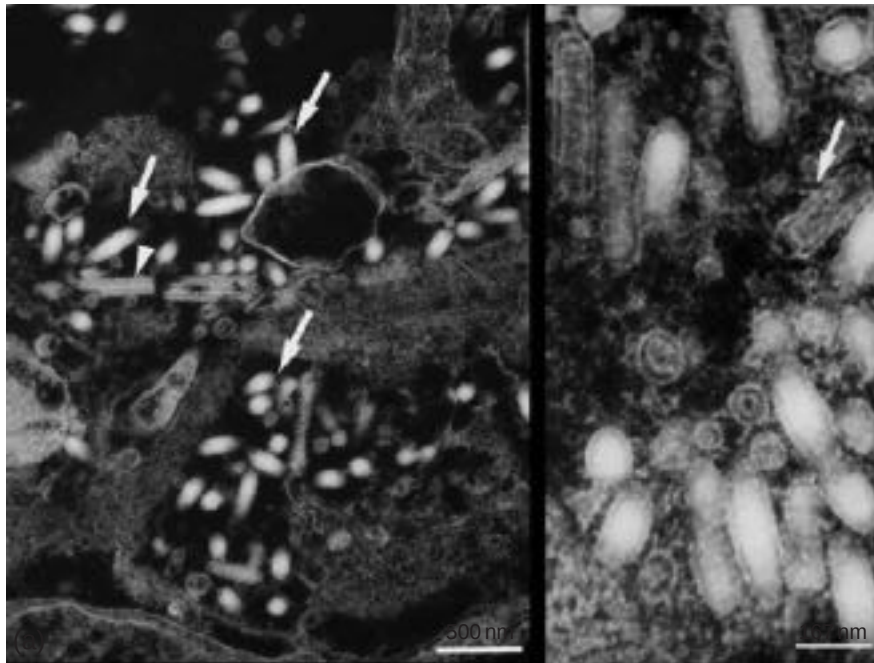


Fig. 7 Transmission electron micrograph of WSSV-infected tissues from beneath the cephalothoracic exoskeletal cuticle. (a) The viral particles spread among the necrotic area. The complete viral particles are indicated with arrows and the long empty tubules with an arrowhead. (b) High magnification of virus particles with rod-shaped morphology. A viral particle with an empty nucleocapsid is indicated with an arrow. From Leu, J.H., Tsai, J.M., Lo, C.F., 2008. White spot syndrome virus. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. pp. 450–459.

other DNA viruses like herpesviruses and SV40 promotes S-phase arrest to maximize the DNA replication machinery. Part of the mechanism for S-phase arrest could involve WSSV IE1 and WSSV056 which bind to an Rb homolog in shrimp. This binding could activate host E2F1 leading to S1 arrest, as demonstrated in *Drosophila* cells overexpressing these two proteins.

Viral morphogenesis begins with *de novo* formation of the viral envelope, seen at first as fibrillar fragments in the nucleoplasm. These fragments form into the membranous and/or vesicular precursors of the viral envelope. Virus assembly begins with the packaging of the 300 kb viral genome into nucleocapsids. This packaging is mediated by binding of VP15 a basic viral DNA binding protein forming nucleoprotein complex fibers. The development of nucleocapsids begins with the formation of long empty tubules (Fig. 7) with a diameter like that of empty nucleocapsids and a segmentation similar to that of nucleocapsids. They can exist individually or may be laterally aligned in groups of two or three to form a larger structure. The assembled tubules break into fragments of 12–14 rings to form empty naked capsids which are then surrounded by envelopes, leaving an opening at one end. The filamentous nucleoprotein complex enters the capsid through this open end, while simultaneously increasing the diameter of the virion. Finally, the open end of the nucleocapsid is closed, and the envelope narrows at the open end to form the polar tail of the mature virion, which has now obtained its characteristic olive-like shape.

Apoptosis

Apoptosis is a cell suicide program that enables multicellular organisms to direct and control cell numbers in tissues and to eliminate cells, including virus-infected cells that may be harmful to the survival of the organism. If apoptosis occurs during the early stage of infection, virus production is severely limited. It is because it inhibits the spread of progeny virus in the host that apoptosis is recognized as an antiviral defense. The classical signs of apoptosis can be identified in WSSV-infected shrimp, including nuclear disassembly, fragmentation of DNA into a ladder, and increased caspase3 activity. Furthermore, there is a positive correlation between the severity of WSSV infection and the number of apoptotic cells. Among the WSSV target tissues, the subcuticular and abdominal epithelia are the most seriously damaged, and these epithelial tissues exhibit the highest incidence of apoptotic cells. However, it is significant that cells displaying apoptotic characteristics do not contain virions, whereas those containing WSSV virions are not apoptotic. It is therefore reasonable to suppose that apoptosis is employed by shrimp as a protective response to prevent the spread of WSSV. Viruses have evolved to inhibit apoptosis early in infection, allowing virus to replicate and to induce it later in infection to enhance virus release from infected cells, and WSSV is no exception. For example, WSSV anti-apoptosis protein 1 (AAP1 or WSSV449) binds to PmCaspase (in *P. monodon* animals). The resultant cleavage (at one site) of AAP1 by the caspase results in its inhibition. In another system pmCaspase is bound by WSSV134 and WSSV322

preventing caspase cleavage activity. Viral proteins, such as VP38 and VP41B can also influence apoptosis by binding the promoter of *PjCaspase* (in *P. japonicum*) to either repress expression (VP38) or activate it later in infection (VP41B). Another way WSSV induces apoptosis is through WSSV ICP11 which disrupts the binding of histones to DNA and disrupting nucleosome assembly, leading to apoptosis. Another viral protein, WSSV208, binds a mitochondrial protein COX5a, another protein which induces apoptosis by upregulating the transcription of apoptosis-related genes.

Hemocyte Responses to WSSV Infection

In crustaceans, the hemocytes (blood cells) mediate many defense-related activities that are essential for nonspecific immunity, including phagocytosis, melanization, encapsulation, cytotoxicity, and clotting. They also produce a vast array of antimicrobial proteins, such as agglutinins, antimicrobial peptides, and lysozymes. WSSV infection decreases the total hemocyte count (THC) in shrimp but not in crayfish. The reduced THC in shrimp is probably due to lysis by the virus of infected hemocytes as well as due to virus-induced apoptosis in both circulating hemocytes and hematopoietic tissue. In addition, circulating hemocytes migrate to tissues that have a high number of virus-infected cells, although this is probably a general defensive response rather than a specific antiviral response. WSSV infection also induces apoptosis in crayfish hemocytes but the percentage of apoptotic hemocytes is much lower in crayfish (1.5%) than in shrimp (20%).

WSSV infection differentially affects the three morphologically and functionally distinct crustacean hemocyte types, the semigranular cells (SGCs), granular cells (GCs), and hyaline cells (HCs). In shrimp and crayfish, both SGCs and GCs can be infected with WSSV but SGCs are the preferential target. This suggests that SGCs are more susceptible to WSSV and that the virus replicates more rapidly in SGCs than in GCs. However, the melanization activity of GCs in WSSV-infected crayfish is reduced. The HC type of hemocyte is refractory to WSSV infection in shrimp but their role in antiviral defense is still not clear. On the other hand, GCs and SGCs each play a major defensive role and WSSV infection of these cell types is likely to weaken shrimp defenses and reduce shrimp health. Since hemocytes are also necessary for clotting, the low THC also accounts for the phenomenon that hemolymph withdrawn from WSSV-infected shrimp and crayfish always has a delayed (or sometimes completely absent) clotting reaction.

Epidemiology

White spot disease is a highly contagious viral disease of penaeid shrimp. Its onset is rapid, and high levels of mortality can result within just a few days. Following the first outbreaks of WSD in Fujian Province, China in 1992 and Taiwan in 1993, this viral disease spread quickly to other shrimp-farming areas, including Japan, Thailand and Korea (1993), the United States (1995), other regions including Vietnam, Malaysia and Indonesia of Southeast Asia (1996), India (1998), Central and South America (1999), and France and Iran (2002). In 2008 it was detected in Mexico, then between 2010 and 2012 in Madagascar, Mozambique and Saudi Arabia and, more recently since 2016 it affected Australia. Essentially, WSD is considered endemic in almost all shrimp-producing countries particularly in Asia and the Americas, causing serious economic damage to the shrimp culture industry.

WSSV is highly virulent and can cause a cumulative mortality of up to 100% within 3–10 days of the first signs of disease at an annual loss of about one billion dollars. WSSV can be transmitted horizontally, either *per os* when the shrimp feed on diseased individuals, contaminated food or infected carcasses, or else through exposure to virus particles in the water, in which case the route of infection is primarily through the gills or other body surfaces. The virus is also transmitted vertically from brooder to offspring. However, transmission is not transovarial because the virus appears to attack only young developing oocytes, which die before reaching maturation. It is therefore more likely that vertical transmission is caused by contamination of the egg mass. Penaeid shrimp are highly susceptible to WSSV. Although there is evidence of resistance during the larval and early (younger than PL6) post-larval stages, WSSV can cause disease in shrimp at any growth stage.

WSSV has a remarkably broad host range. Almost every species of penaeid shrimp is susceptible to WSSV infection. Moreover, the virus can infect other marine, brackish, and freshwater crustaceans, including crayfish, crabs, and spiny lobsters. However, in contrast to penaeid shrimp, infection is often not lethal in these species, and consequently they may serve as reservoirs and carriers of the virus. Furthermore, at least one insect, the shore fly (a member of the family *Ephyridae*), as well as copepods collected from WSSV-affected farms, have been diagnosed as WSSV-positive by PCR, suggesting that they are also possible reservoir hosts.

In epidemiological studies of WSSV, variations in the number of the 54 bp tandem repeats located between the *rr1* and *rr2* genes have been used as a strain-specific, genetic marker. The number of repeats varies greatly when infected shrimps are collected from different ponds or from the same ponds at different times, but in almost every outbreak of WSD, shrimps from the same pond usually have the same number of repeats. This suggests that in each pond, a single WSSV isolate is the causative agent.

Clinical Features and Pathology

The most commonly observed clinical sign of WSD in shrimp is white spots in the exoskeleton and epidermis. These may range in size from minute spots to disks several millimeters in diameter which may coalesce into larger plates. The spots may result from

abnormal deposits of calcium salts by the cuticular epidermis or result from disruption to the transfer of exudates from the epithelial cells to the cuticle. Infected animals are lethargic, reduce their food consumption, and display a reddish to pink body discoloration due to expansion of cuticular chromatophores. Moribund shrimp exhibit systemic destruction of target tissues with many infected cells showing homogeneous hypertrophied nuclei. At advanced stages of infection, numerous virus particles are released into the hemolymph from the lesions, causing a general viremia.

It should be noted that although the white spots are a typical and characteristic clinical sign of WSD, white spots on the carapace of shrimp can also be caused by other environmental stress factors, such as high alkalinity or a bacterial shell disease. Conversely, moribund shrimp with WSD may have few, if any, white spots.

To date, no species of penaeid shrimp is known to show significant resistance to WSD.

Protection of Shrimp Against WSSV Infection Using Vaccination and RNAi Strategies

Considering the economic impact of WSSV on the shrimp and crab industries, it is not surprising that a lot of research has gone into trying to mitigate the disease. These include efforts to enhance the immune system, use of vaccines and RNAi and discovery or development of antivirals. When *P. japonicus* shrimps survive either natural or experimental WSSV infections, they sometimes show resistance to subsequent challenge with WSSV. This 'quasi-immune response' suggests that shrimps may have an innate immune system that includes specific memory. Several vaccines including sub-unit, whole virus-inactivated, DNA and RNA based vaccines using different modes and times of vaccination have been developed with varying degrees of success. In earlier studies inactivated bacteria expressing VP28 (but not VP19) provided survival up to about 71%, but the protection is relatively short lived to less than 21 days postvaccination. For example, oral administration by continuous feeding of *Bacillus subtilis* spores expressing VP28 antigen even at low concentrations, provided about 70% protection. Through a metanalysis, Shu-Ying Feng and colleagues demonstrated that RNA-based vaccine showed the best protection and that of the subunit vaccines, VP26 had the best protective effects. Among different routes of administration, oral, immersion and injection, oral administration was the best. RNAi against different WSSV genes has also been tested as a treatment method. For example, injection of *M. japonicus* with vp28-siRNAs encapsulated with β -1,3-D-glucan inhibited WSSV replication. In a later study, injection of dsRNA targeting the VP28 mRNA provided nearly complete protection for WSSV infection. While RNAi seems to be effective, injection would be too labor intensive as a treatment regimen.

Many natural products have been shown to be particularly effective in limiting virus infection. Alginate from the seaweed *Sargassum siliquosum*, provided about 56% survival rates. Epigallocatechin-3-gallate (EGCG) extracted from green tea reduced WSSV production in the mud crab *Scylla paramamosain* and Kuruma shrimp (*Marsupeneaus japonicus*). An additional effect of EGCG treatment was that it correlated with increased expression of immune-related genes such as those for proPO, Rho, Rab7, p53, TNF- α , MAPK and NOS, improving innate immunity. Hesperetin, a flavanone compound from lemons and oranges, has multiple anti-inflammatory, anti-cancer and anti-viral (e.g., against herpes simplex virus type 1, poliovirus, respiratory syncytial virus) characteristics. Treatment of crayfish with hesperetin reduced mortality in WSSV infections and increased expression of innate immunity-related genes such as NF- κ B and C-type lectin. Other extracts that have shown promise are ones derived from *Cynodon dactylon* (Bermuda grass), *Ceriops tagal* (Indian mangrove), *Momordica charantia* (bitter melon) and *Agathi grandiflora* (vegetable hummingbird). Some host proteins including LvGP (*Litopenaeus vannamei*, Glycogen phosphorylase), PmGILT enzyme (*Penaeus monodon*, Gamma-interferon-inducible lysosomal thiol reductase) and LvSrc64B (*L. vannamei*, Src family kinases) showed antiviral activity against WSSV.

Further Reading

- Bao, W., Tang, K.F.J., Acivar-Warren, A., 2020. The complete genome of an endogenous nimavirus (Nimav-1_LVa) from the Pacific whiteleg shrimp *Panaeus (Litopenaeus) vannamei*. *Genes* 11, 94. doi:10.3390/genes11010094.
- Feng, S.Y., Liang, G.F., Zu, Z.S., 2018. Meta-analysis of antiviral protection of white spot syndrome virus vaccine to the shrimp. *Fish and Shellfish Immunology* 81, 260–265.
- Kawato, S., Shitara, A., Wang, Y., et al., 2019. Crustacean genome exploration reveals the evolutionary origin of White spot syndrome virus. *Journal of Virology* 93 (3), e01144. [18].
- Leu, J.H., Tsai, J.M., Lo, C.F., 2008. White spot syndrome virus. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. pp. 450–459.
- Leu, J.H., Tsai, J.M., Wang, H.C., et al., 2005. The unique stacked rings in the nucleocapsid of the white spot syndrome virus virion are formed by the major structural protein VP664, the largest viral structural protein ever found. *Journal of Virology* 79, 140–149.
- Oakey, H.J., Smith, G.S., 2018. Complete genome sequence of a white spot syndrome virus associated with a disease incursion in Australia. *Aquaculture* 484, 152–159.
- Robalino, J., Bartlett, T., Shepard, E., et al., 2005. Double-stranded RNA induces sequence-specific antiviral silencing in addition to nonspecific immunity in a marine shrimp: convergence of RNA interference and innate immunity in the invertebrate antiviral response? *Journal of Virology* 79, 13561–13571.
- Tsai, J.M., Wang, H.C., Leu, J.H., et al., 2006. Identification of the nucleocapsid, tegument, and envelope proteins of the shrimp white spot syndrome virus virion. *Journal of Virology* 80, 3021–3029.
- van Hulten, M.C.W., Witteveldt, J., Peters, S., et al., 2001. The white spot syndrome virus DNA genome sequence. *Virology* 286, 7–22.
- Verbruggen, B., Bickley, L.K., van Aerle, et al., 2016. Molecular mechanisms of White spot syndrome virus infection and perspectives on treatments. *Viruses* 8. doi:10.3390/v8010023.
- Wang, H.C., Hirono, I., Maningas, M.B.B., et al., 2019. ICTV virus taxonomy profile: Nimaviridae. *Journal of General Virology* 100, 1053–1054.
- Witteveldt, J., Cifuentes, C.C., Vlak, J.M., van Hulten, M.C.W., 2004. Protection of *Penaeus monodon* against white spot syndrome virus by oral vaccination. *Journal of Virology* 78, 2057–2061.
- Yang, F., He, J., Lin, X., et al., 2001. Complete genome sequence of the shrimp white spot bacilliform virus. *Journal of Virology* 75, 11811–11820.

- Zhang, X., Huang, C., Hew, C.L., 2004. Use of genomics and proteomics to study white spot syndrome virus. In: Leung, K.Y. (Ed.), *Molecular Aspects of Fish and Marine Biology*, Vol. 3: Current Trends in the Study of Bacterial and Viral Fish and Shrimp Diseases. Singapore: World Scientific, pp. 204–236.
- Zheng, S.C., Xu, J.Y., Liu, H.P., 2019. Cellular entry of white spot syndrome virus and antiviral immunity mediated by cellular receptors in crustaceans. *Fish and Shellfish Immunology* 93, 580–588.
- Zuidema, D., van Hulte, M.C.W., Marks, H., *et al.*, 2004. Virus–host interaction of white spot syndrome virus. In: Leung, K.Y. (Ed.), *Molecular Aspects of Fish and Marine Biology*, Vol. 3: Current Trends in the Study of Bacterial and Viral Fish and Shrimp Diseases. Singapore: World Scientific, pp. 237–255.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsdna-viruses/w/nimaviridae
Nimaviridae.
dsDNA Viruses.
International Committee on Taxonomy of Viruses (ICTV).

Nodaviruses of Invertebrates and Fish (*Nodaviridae*)

Kyle L Johnson, The University of Texas at El Paso, El Paso, TX, United States and CQuentia, Memphis, TN, United States
Jacen S Moore, University of Tennessee Health Science Center, Memphis, TN, United States

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AHPND Acute hepatopancreas necrosis disease
BBV Black beetle virus
BFNNV Barfin flounder nervous necrosis virus
BoV Boolarra virus
CMD Covert mortality disease
CMNV Covert mortality nodavirus
CPE Cytopathic effects
cryoEM Electron cryomicroscopy
EMS Early mortality syndrome
FHV Flock House virus
MrNV *Macrobrachium rosenbergii* nodavirus
NNV Nervous necrosis virus
NoV Nodamura virus
PaV Pariacoto virus

PL Post-larvae
qPCR Quantitative PCR
qRT-PCR Quantitative reverse-transcription PCR
RdRp RNA-dependent RNA polymerase
RGNNV Redspotted grouper NNV
RT-LAMP Reverse-transcription loop-mediated rapid isothermal amplification
RT-PCR Reverse-transcription PCR
SJNNV Striped Jack NNV
TPNNV Tiger puffer NNV
VCMD Viral covert mortality disease
VER Viral encephalopathy and retinopathy
VLP Virus-like particle
VNN Viral nervous necrosis
WTD White tail disease

Classification (Compact): Family *Nodaviridae*, genera *Alphanodavirus* and *Betanodavirus*

The family *Nodaviridae* in the order *Nodamuvirales* is divided into two canonical genera: *Alphanodavirus* with five species including *Nodamura virus* as the type species and *Betanodavirus* with four species including *Striped jack nervous necrosis virus* as the type species.

Alphanodaviruses are primarily virus pathogens of insects. They include Nodamura virus (NoV, a member of the type species), Flock House virus (FHV), black beetle virus (BBV), Boolarra virus (BoV), and Pariacoto virus (PaV).

Viruses of the *Betanodavirus* genus are the causative agents of viral nervous necrosis (VNN) in infected fish. Consequently, the nomenclature used for most betanodaviruses is the name of the host species followed by “nervous necrosis virus” (NNV). The viruses are classified into four genotypes based on genome homology: Redspotted Grouper NNV (RGNNV), Barfin Flounder NNV (BFNNV), Striped Jack NNV (SJNNV), and Tiger Puffer NNV (TPNNV). RGNNV and SJNNV are serologically distinct from one another and from the cold-water betanodaviruses BFNNV and TPNNV. Corresponding serogroups A (SJNNV), B (RGNNV), and C (BFNNV/TPNNV) have been proposed to reflect these serological relationships. The placement of reassortant viruses that contain RNA1 from RGNNV and RNA2 from SJNNV (RGNNV/SJNNV) in serogroup A and those that RNA1 from SJNNV and RNA2 from RGNNV (SJNNV/RGNNV) in serogroup C indicates that the capsid protein encoded by RNA2 is a major antigenic determinant.

The discovery of nodaviruses that infect shrimp and their low sequence homology with either Alphanodaviruses or Betanodaviruses has led to the proposal of a third genus, *Gammanodavirus*, to include *Macrobrachium rosenbergii* nodavirus (MrNV) and *Penaeus vannamei* nodavirus (PvNV). These viruses have been isolated from the giant river prawn (*M. rosenbergii*) and the white leg or Pacific white shrimp (*P. vannamei*), respectively, and share approximately 80% similarity at the amino acid level. Two other shrimp nodaviruses, covert mortality nodavirus (CMNV) and *Farfantepenaeus duorarum* nodavirus (FdNV), have also been proposed for inclusion in this group. While *Macrobrachium rosenbergii* nodavirus (MrNV) remains in the family *Nodaviridae*, a satellite of MrNV called extra small virus (XSV) that was originally classified as a nodavirus has been reclassified as the sole member of a newly established virus family, the *Sarothroviridae*. The *Sarothroviridae* were named for the small arthropod hosts (shrimp, some insects) they infect.

Provisional Nodaviruses

Additional viruses have been proposed as potential members of the family *Nodaviridae*, as summarized in [Table 1](#), although none have yet been officially classified as nodaviruses by the ICTV.

Proposed Six Clade Taxonomic Structure

While previous nodavirus phylogenies have relied on the relationship between the capsid proteins, yielding the current 2 genera (*Alphanodavirus* and *Betanodavirus*) and the proposed *Gammanodavirus* genus, more recent alignments using the RdRp protein sequence have suggested a broader taxonomic structure, particularly when considered together with a number of newly discovered viruses that share significant sequence homology with one or more recognized nodaviruses. Clade 1 includes all of the recognized

Table 1 Provisional nodaviruses

<i>Virus name</i>	<i>Abbreviation</i>	<i>Source of isolate</i>
Bat guano-associated virus	BGNV	Bat species, including the Brazilian free-tailed bat (<i>Tadarida brasiliensis</i>), cave myotis (<i>Myotis velifer</i>), evening bat (<i>Nycticeus humeralis</i>), and/or the tricolored bat (<i>Perimyotis subflavus</i>)
Culannivirus	CuNV	<i>Culex annulirostris</i> mosquitoes
Hypsignathus nodavirus	HSNV	Fruit bats (<i>Hypsignathus monstrosus</i>)
Le Blanc nodavirus	LBNV	Nematodes (<i>Caenorhabditis briggsae</i>)
Lutzomyia Plaiui nodavirus	LPNV	Phlebotomine sand flies (<i>Lutzomyia longipalpis</i>)
Mosinivirus	MoNV	<i>Culex nebulosus</i> and <i>Culex</i> sp. mosquitoes
Orsay virus	ONV	Nematodes (<i>Caenorhabditis elegans</i>)
Santeuil nodavirus	SNV	Nematodes (<i>Caenorhabditis briggsae</i>)
Wuhan nodavirus	WhNV	Small cabbage white butterfly larvae (<i>Pieris rapae</i>)

alphanodaviruses except PaV; Clade 2 includes the betanodaviruses and several newly discovered noda-like viruses isolated from different arthropod hosts. Clade 3 includes PaV, MrNV, and CulV and is considerably more diverse than the other proposed clades, particularly MrNV. Clade 4 includes several newly sequenced isolates; Clade 5 includes one isolate from *Helicoverpa zea* and several newly sequenced isolates from diverse insect hosts, and Clade 6 includes SNV, ONV, and LBNV, all isolated from nematodes.

Virion Structure

Nodavirus particles are comprised of a protein shell co-encapsidating one copy each of the two genomic RNA molecules. The three dimensional structures of several members of the *Nodaviridae* have been solved by cryo electron microscopy (cryoEM) and/or x-ray crystallography (BBV, FHV, NoV, PaV, MrNV). The unique features of each are characterized below.

The structures of alphanodaviruses BBV, FHV, and NoV are all closely similar. The viral capsid consists of 180 coat protein subunits arranged with icosahedral symmetry on a T=3 quasi-equivalent lattice. Although the primary structures of the coat protein subunits are identical, they occupy three quasi-identical locations (A, B, and C) in the icosahedral asymmetric unit. The A subunits are related by 5-fold symmetry, while the B and C subunits are related by three-fold symmetry. However, only ~15% of the viral RNA is visible in these structures. Strikingly, cryoEM image reconstruction and the crystal structure of PaV revealed the presence of a dodecahedral cage comprised of duplex RNA, representing approximately 35% of the viral single-stranded RNA genome. The highly basic N-terminal regions of the capsid protein that form pentamers (the A subunits) make contacts with the RNA cage structure.

The capsid structure of recombinant virus-like particles (VLPs) of a malabaricus grouper nervous necrosis virus (MGNNV) (*Epinephelus malabaricus*) nodavirus VLP differs from that of the insect nodaviruses. Its protein fold more closely resembled those of Norwalk virus and tomato bushy stunt virus than that of the alphanodaviruses.

The structural arrangement of the *Machrobrachium rosenbergii* nodavirus (MrNV) capsid also differs from those reported for the alphanodaviruses. While MrNV shares an 8-stranded antiparallel beta barrel capsid protein structure with the other nodaviruses, it lacks the canonical nodavirus trimeric “spike” at the isocahedral vertex at the 5-fold axis and the MrNV capsid protein forms a dimeric spike in the equivalent position. Many similarities were seen between the MrNV capsid and that of tombusviruses, including the coordination of calcium ions (3 pairs in each asymmetric unit) and an overall similar fold to that of the tombusvirus spike protein.

Genome

As shown schematically in [Fig. 1](#), the nodavirus genome is divided into two segments of positive strand ssRNA with non-methylated 5' cap structures and lacking poly(A) tails. The larger segment, RNA1, encodes the viral RNA-dependent RNA polymerase (RdRp), while RNA2 encodes the capsid protein precursor (see below). During replication of RNA1, a small subgenomic RNA (RNA3) is synthesized and subsequently replicated. RNA3 encodes proteins B1 and B2, whose functions are described below. For FHV, replication of RNA2 suppresses RNA3 synthesis, yet viral mutants that do not make RNA3 fail to replicate RNA2. If RNA3 is provided *in trans* from an exogenous source, RNA2 replication is restored. This interdependence may assist in coordinating replication of the two segments, which must be co-encapsidated during virus assembly.

Proteins

RNA-Dependent RNA Polymerase (RdRp)

Protein A (110 kDa) is the sole catalytic subunit of the RNA-dependent RNA polymerase that replicates the genomic RNA. It also possesses an RNA capping (guanylyl transferase) activity. This protein is membrane-associated (via either a transmembrane region or a membrane-association region that may involve only a single leaflet of the membrane, depending on the virus) and establishes

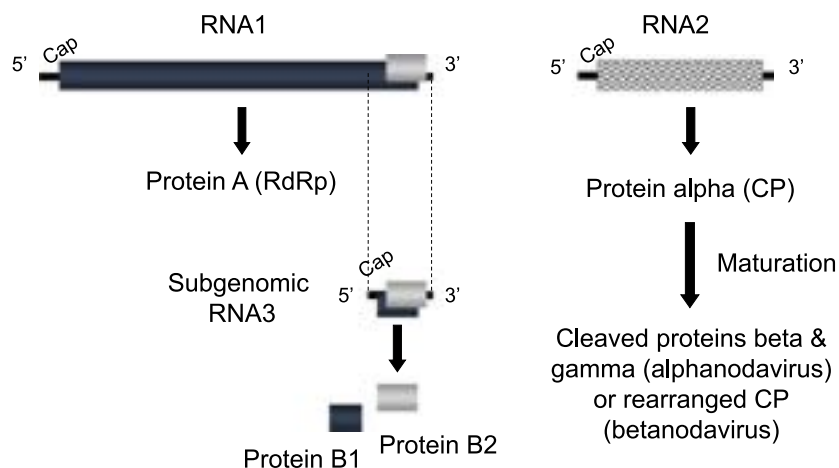


Fig. 1 Schematic of nodavirus genome organization. Genomic RNA1 and RNA2 as well as subgenomic RNA 3 are shown schematically. All three are capped (unmethylated cap 0 structure) but not polyadenylated. RNA1 encodes the viral RdRp, RNA3 encodes small proteins B1 and B2 in overlapping reading frames, and RNA2 encodes the capsid protein (CP; betanodaviruses) or the capsid protein precursor alpha, which is autocatalytically cleaved into mature capsid proteins beta and gamma (alphanodaviruses). RNAs 1 and 2 are copackaged into the same virion, while subgenomic RNA3 is not packaged.

viral replication complexes in association with the cellular outer mitochondrial membrane. For betanodaviruses, the N-terminus of the RdRp is associated with their ability to replicate at colder temperatures (thermotolerance).

Capsid Proteins

The nodavirus capsid consists of 180 copies of the viral capsid protein (CP). For alphanodaviruses, the capsid protein (CP) is initially translated as a precursor, protein alpha, which undergoes an auto-catalytic maturation cleavage after assembly to yield the mature capsid proteins beta and gamma found in the virion. For betanodaviruses, this cleavage is not observed but the capsid protein undergoes an essential conformational change to produce mature virions. The betanodavirus CP contains an N-terminal signal for localization to the nucleolus, where it plays a role in cell cycle arrest. Its accumulation later in infection induces apoptosis via a caspase-dependent mechanism. The C-terminus of the CP may play a role in host specificity.

Proteins B1 and B2

Two small proteins, B1 and B2, are translated from RNA3. B1 is in the same reading frame as the RdRp and is identical to its C-terminus, although not all betanodaviruses are predicted to encode B1. B2 is translated in an overlapping reading frame. In alphanodaviruses, the function of B1 is unknown. For betanodaviruses, B1 serves as an anti-necrotic “death factor” that results in loss of mitochondrial membrane potential. For RGNNV, the subcellular localization of B1 transitions from the cytoplasm to the nucleus late in infection, resulting in cell cycle arrest. B2 suppresses RNA interference (RNAi) by binding to and sequestering double-stranded RNA replication intermediates. Betanodavirus B2 serves as a necrotic “death factor” that upregulates proapoptotic protein Bax and induces loss of mitochondrial membrane potential but not release of mitochondrial cytochrome C. However, in alphanodaviruses, B2 is not associated with apoptosis and while FHV infection does induce caspase-dependent apoptosis in infected cells, NoV does not.

Physical Properties

Nodaviruses are non-enveloped, small (29–32 nm diameter) icosahedral viruses that contain bipartite positive-strand ssRNA genomes. Alphanodavirus particles are resistant to chloroform, diethyl ether, 1 M urea, 20 mM EDTA and 1% sodium deoxycholate and all except Boolarra virus (BoV) are stable in 1% SDS. Immature provirions are sensitive to these denaturants and to freezing. FHV is stable at 40°C for 10 min, but loses infectivity when heated at 58°C. FHV, BoV, and NoV are stable over a wide range of pH values (pH 3–10). NoV is unstable in the presence of chloride ions and BBV and LNV_C are unstable when stored in cesium chloride (CsCl), while FHV is stable in CsCl.

Betanodaviruses are resistant to aquatic environments and can survive at low temperatures in seawater. Infectivity of viruses contained in seawater decreases at temperatures above 25°C while that in drying conditions lose infectivity at 21°C. Betanodavirus particles are inactivated by hydrogen peroxide, benzalkonium chloride, iodine, and sodium hypochlorite but resistant to formalin. While MrNV particles are stable to storage in a cell lysate or tissue sample at –20°C and –70°C, their infectivity is lost upon experimental heat treatment (65°C for two hours). The chemical stability of MrNV particles is unknown.

Life Cycle

The ecological life cycle of Alphavodaviruses is unknown. Nodamura virus was isolated from mosquitoes and found to cause lethal infections in suckling mice when introduced experimentally via the intracranial (IC) route. NoV can also be transmitted from an infected mouse to a naïve one by mosquitoes and cause lethal infections in both vertebrates (suckling mice and suckling hamsters) and invertebrates (mosquitoes, honeybees, larvae of the greater wax moth, and fruit flies). The remaining alphavodaviruses have been isolated from only insects, including FHV, BBV, BoV, but little is known about their transmission in the wild. However, the infectious cycle of the viruses is well understood due to their ready replication in a variety of cultured cells from both vertebrates and invertebrates, and in transformed yeast cells. Betanodaviruses infect primarily larvae and juvenile fish. They can be transmitted to juvenile farmed fish populations from wild reservoir species or through importation of infected juveniles from other farms. The literature suggests that betanodaviruses most likely infect the intestinal epithelium and peripheral nervous system, are rapidly transmitted to the central nervous system, and result in either death or persistence in survivors. The virus is thought to spread to the environment by decomposition of dead fish.

Little is known about the natural transmission cycle of *Macrobrachium rosenbergii* nodavirus (MrNV). Both horizontal and vertical (*trans*-ovum) transmission by the water-borne route have been observed. Outbreaks of WTD disease may be induced or enhanced by rapid environmental changes such as those in pH, salinity, and temperature. Similarly, survival of shrimp with symptoms of covert mortality disease or viral covert mortality disease decreases sharply with water temperatures above 28°C.

Epidemiology

Over 50 species of fish are susceptible to betanodavirus infections. In particular the marine fish species striped jack (*Pseudocaranx dentex*), European sea bass (*Dicentrarchus labrax*), groupers (*Epinephelus* species), and flatfishes. These diseases primarily affect farmed fish, which are particularly vulnerable to infectious diseases due to the high densities at which the fish are raised. In general, the earlier the age at which betanodavirus disease signs and symptoms occur, the greater the mortality. In larval striped jack, the disease can occur within two days of hatching with the peak mortality observed at 10 days post-hatching, resulting in almost complete loss of the larvae. In European sea bass, mortality is most frequently seen 30-days post-hatching but may occur in adults as well as larvae and juveniles. The disease has been reported from countries in the Mediterranean region (France, Greece, Israel, Italy, Malta, Portugal, Spain, and Tunisia), in the Atlantic region (Norway and the United Kingdom), in North America (Canada and the United States), and in Asia (Chinese Taipei, India, Indonesia, Iran, Japan, Korea, Malaysia, the Philippines, the Peoples Republic of China, Thailand, and Vietnam). In Japan, a representative sample of 30 species in two geographically distinct regions confirmed that most farmed and wild fish tested positive for VNN.

White tail disease (WTD) in the giant river prawn *Macrobrachium rosenbergii* is caused by infection with MrNV. The disease has been reported in Australia, Chinese Taipei, the French West Indies (the islands of Guadeloupe and Martinique), India, Malaysia, the Peoples Republic of China, and Thailand. PCR-positive results but no active infection have been obtained with other crustacean species, including brine shrimp (*Artemia* sp.), red claw crayfish (*Cherax quadricarinatus*), Monsoon River prawn (*Macrobrachium malcolmsonii*), hairy river prawn (*M. rude*), Indian white prawn (*Penaeus indicus*), kuruma prawn (*Penaeus japonicus*), and the giant tiger prawn (*Penaeus monodon*). In the latter three marine shrimp species, infection could not be initiated by the oral route, but intramuscular inoculation resulted in accumulation of particles in abdominal muscle, gill, hemolymph, intestine, and stomach that were infectious when inoculated into *M. rosenbergii* postlarvae (PL). Similar results have also been described in several insect species, including dragonfly (*Aeshna* sp.), giant water bug (*Belostoma* sp.), beetle (*Cybister* sp.), and backswimmer (*Notonecta* sp.). For MrNV, only larvae, PL, and early juveniles are susceptible, while adults are resistant. No mortality has been reported in subadult or adult prawns naturally or experimentally infected with MrNV.

Covert mortality disease (CMD) or viral covert mortality disease (VCMD) is an emerging shrimp disease caused by covert mortality nodavirus (CMNV). The name CMD derives from moribund shrimp hiding in deep water rather than swimming to the surface or in shallow water. Initially identified in *Litopenaeus vannamei* (formerly *Penaeus vannamei*; common name white leg shrimp or Pacific white shrimp), specimens of *Marsupenaeus japonicus* (Japanese tiger prawn or Kuruma prawn) and *Fenneropenaeus chinensis* (Chinese white shrimp or fleshy prawn) isolated from diseased ponds exhibit similar symptoms. An outbreak of VCMD in 2012–2014 resulted in heavy loss of farmed *M. japonicus* and *F. chinensis* in the municipalities of Tianjin and Hebei and Shandong Provinces in China. A recent three-year epidemiological survey determined the prevalence of CMNV in samples of *L. vannamei*, *F. chinensis*, *Marsupenaeus japonicus*, *Macrobrachium rosenbergii*, and *Penaeus monodon* collected from 145 sampling site shrimp farms in ten coastal provinces in China. The prevalence was highest in Guangdong and Hainan Provinces (approximately 50%), 25%–30% in Hebei, Shandong, Jiangsu, Zhejiang, and Fujian Provinces, 17% in Guangxi Province, and approximately 10% in Liaoning and Tianjin Provinces. Other studies have reported on isolation of CMNV-positive shrimp from Ecuador, Thailand, and Vietnam, suggesting a wide distribution.

Clinical Features

VNN

Betanodaviruses are the causative agents of viral nervous necrosis (VNN), also known as viral encephalopathy and retinopathy (VER). Most common clinical signs included aberrant swimming (horizontal swimming, looping, darting, spiral swimming, or whirling), loss of

appetite, hyperinflation of the swim bladder, and lightening or darkening of the coloration. Nerve cells in the cerebellum and olfactory lobe exhibit vacuolization and infected cells have been detected in the spinal cord, medulla oblongata, thalamus, and preoptic area. However, the clinical signs depend on the fish species, developmental stage, water temperature, and the phase of the disease. For example, while juvenile fish may be severely affected, adults are often less so. Further information on VNN may be obtained from the OIE Reference Laboratory for Viral encephalopathy and retinopathy and the excellent review by Bandín and Souto (see Further Reading).

WTD

Clinical signs include lethargy and whitening or opaqueness of the abdominal muscles. PL become opaque and develop a milky, white appearance. This whitish discoloration appears first in the second or third abdominal segment and diffuses gradually in both anterior and posterior directions. In severe cases, the last pair of abdominal legs (uropods) and the tail fan (comprised of uropods and telson) may degenerate. Mortality generally peaks 5–6 days after onset of symptoms and very few PL survive beyond 15 days. PL that do survive can reach normal market size. Due to the possibility of persistent infection, adults can serve as carriers.

VCMD

Clinical signs include hepatopancreatic atrophy with color fading and necrosis, empty stomach and guts, soft shell, slow growth, and often abdominal muscle whitening and necrosis. Daily reports of mortality within a population begin within one month and increase during a 60–80 day period post-stocking, with a total mortality of approximately 80%. Although these symptoms are somewhat similar to those caused by early mortality syndrome (EMS)/acute hepatopancreas necrosis disease (AHPND), caused by a bacterial pathogen, CMD and EMD/AHPND are distinct from one another. As a result, covert mortality disease (CMD) was renamed viral covert mortality disease (VCMD) to reduce confusion between the diseases.

Pathogenesis

Betanodaviruses replicate in brain, spinal cord, and retina of a number of fish species, resulting in vacuolization of the brain and retina together with cellular necrosis and neural degeneration. Often large vacuolar lesions are seen in larvae and juveniles while adults exhibit mild or absent symptoms. Infected striped jack larvae also exhibit hyperplasia and degeneration of epithelial cells of the skin, oral cavity, and gills. The tissue distribution appears to differ between larvae/juvenile fish and adults, with greater vacuolization in the spinal cord and medulla oblongata seen in the earlier developmental stages.

In MrNV-infected PL and early juveniles, striated muscles of the tail, abdomen, and cephalothorax are affected. Histological examination reveals acute Zenker's necrosis of striated muscles, with severe hyaline degeneration, muscular lysis, and necrosis. Abnormal open spaces, moderate edema, and basophilic cytoplasmic inclusions are observed among infected muscle cells. Transmission electron microscopy reveals a disorganized cytoplasm and the presence of not only MrNV but also a satellite virus (extra small virus, XSV). XSV was initially classified as a member of the *Nodaviridae* but was moved recently into its own family, the *Sarothroviridae*.

Covert mortality nodavirus (CMNV) causes multifocal myonecrosis in striated muscle, accumulation of vacuoles in the cytoplasm of hepatopancreocytes, eosinophilic inclusions in the tubular epithelium of the hepatopancreas, and spheroids in the lymphoid organ with inclusions and nuclear condensation of the chromatin.

Immune Responses to Nodavirus Infection

Innate Immune Responses

Host immune systems are equipped with a variety of tools in their antiviral arsenal. Innate immune responses to viral infection can typically be classified within seven primary categories: (1) activation of Pattern Recognition Receptors (PRR) capable of binding Pathogen-Associated Molecular Patterns (PAMPs); (2) production of siRNAs, miRNAs, and piwiRNAs that target viral RNA for degradation and inhibit replication; (3) activation of signaling pathways such as Toll, NF- κ B, and Imd; (4) activation of Jak-STAT pathways that act in a fashion similar to interferon pathways in mammals which control transcription and antiviral replication; (5) triggering of autophagy which can recycle damaged organelles and engage in signaling through the PI3K-Akt and TOR pathways; (6) production of interferon-like molecules and inflammatory cytokines; and (7) cytotoxic T cell responses that activate granzymes and perforins.

Activation of the PRR through the binding of viral-associated nucleic acids to Toll-like receptors lead to the activation of a number of signaling molecules including TRIF, MyD88, NF- κ B, Interferon Regulatory Factors (IRF), and MAPK that stimulate the production of cytokines, chemokines, and interferon-like proteins to suppress viral replication. Anti-microbial Toll-responses from PRRs led to activation of serine proteases and signal transduction, resulting in nuclear translocation and regulation of effector genes. Both Toll and Imd pathways are linked to NF- κ B signaling in innate immune responses by activation of cellular transcription. Upregulation and expression of genes in both pathways has been observed in nodavirus-infected species.

Silencing of viral RNA occurs most commonly through the production of complementary silencing RNAs (siRNAs) and microRNAs (miRNAs) that prevent transcription and regulate gene expression. *P*-element Induced WImpy (PIWI) or piRNAs are a type of RNA distinct from miRNAs whose role is to exert epigenetic control of germline genomic elements in vertebrates. MiRNAs

and piRNAs have been identified in invertebrates, however, their role in transcriptional regulation particularly in insects is not well-understood. Evaluation of their role has suggested that production of piRNA precedes that of siRNA then wanes. Blockage of piRNA leads to increased viral replication and the accumulation of viral particles, suggesting a regulatory role. In response to these regulatory pathways, nodaviruses have developed mechanisms to suppress the function of both host siRNA and miRNA activity by producing proteins and subgenomic RNAs including B2 and dicer that bind and promote cleavage of dsRNA. Jak-STAT pathways can respond to viral activation through multiple mechanisms including dephosphorylation, nuclear export, or negative feedback. The activation of this pathway results in changes in the transcription profile of infected organisms. This upregulation is not due to dsRNA alone, but requires active viral replication and production. Jak-STAT pathway mRNA expression is increased in nodavirus-infected hosts, supporting an anti-viral role for this pathway. Importantly, infection of hosts with different nodaviruses led to differential expression of Jak-STAT pathway members, indicating that other pathways may be involved. Furthermore, STAT can also repress gene activity by forming complexes with other transcription factors and chromatin-modifying proteins that regulate NF- κ B immune responses. Toll ligands, cecropins, and defensins can also be downregulated, supporting the notion that JAK-STAT activation can lead to inhibition of Toll-regulated genes, thus regulating anti-viral responses.

Autophagy is another mechanism that infected hosts use to regulate viral transcription and translation. In this process, *de novo* synthesized membranes encompass large cytoplasmic components including damaged organelles or protein aggregates containing viral components that join with lysosomes to form autolysosomes which degrade and recycle digested materials. Positive and negative regulation of autophagy occurs primarily through the phosphoinositol 3-kinase-related-kinase (PI3K-Akt) and Target of Rapamycin (TOR) pathways, respectively. Knockdown of autophagy-related genes results in higher viral production and prevents the formation of autolysosomes, supporting its role in viral suppression. Phosphorylation of PI3K-Akt was shown to downregulate TOR activation and increase autophagy which subsequently leads to decreased viral replication.

Interferons and other inflammatory cytokines are a crucial aspect of the host antiviral response and help by inhibiting viral replication and promoting a pro-inflammatory phenotype. Viral pathogens such as NoV can promote activation of Interferon Specific Response Elements (ISRE) through two common pathways. Pattern Recognition Pathways (PRRs) that include RIG-I (retinoic acid-inducible gene I) and MDA5 (melanoma differentiation-associated protein 5) can sense double-stranded RNA in the cytoplasm. Once this occurs, these proteins can trigger the activation of MAVS (mitochondrial antiviral signaling protein) which, in turn, subsequently activates a complex composed of Tank Binding Kinase 1 (TBK1) and I κ B kinase- ϵ (IKK ϵ). Stimulation of the TBK1/IKK ϵ complex drives the phosphorylation of interferon regulatory factors 3 and 7 (IRF3 and IRF7), which translocate to the nucleus and bind to the ISRE promoters that upregulate transcription and translation of interferon- β and other Interferon-stimulated genes (ISGs). Viral PAMPs are recognized through Toll-like receptors 3, 7 and 8 (TLR3, 7, and 8), which specifically recognize viral dsRNA and ssRNA in the endosome and signal through TRIF (TIR-domain-containing adapter-inducing interferon- β) and MyD88 (Myeloid differentiation primary response 88). Activation of these factors triggers phosphorylation of IRF5 and IRF7 and downstream upregulation of ISGs and Type I Interferon production. TRIF can also activate the TBK1/IKK ϵ complex, thus amplifying the antiviral response.

Multiple mechanisms used by viruses to inactivate this pathway have been identified including the prevention of dephosphorylation keeping signaling proteins like MDA5 in an inactive state, blocking the ATPase activity or preventing ubiquitination of RIG-I, modulation of miRNAs that block MDA or RIG-I upregulation, or direct cleavage or degradation of MDA5, MAVS and RIG-I. Nodaviral RNA replication complexes promote progressive clustering and aggregation of mitochondria within the cell culminating in malformation of muscle fibers, tissue damage and paralysis. Because nodavirus RdRp's localize to the outer mitochondrial membrane, it is critical that these viruses prevent or modulate host antiviral activity. Because the NoV RdRp associates with MAVS (Johnson, personal communication), mechanisms using such strategies as direct cleavage and degradation of mitochondrial membrane-associated proteins like MAVS and likely STING (Stimulator of interferon genes) or their sequestration in virally-induced autophagosomes (Johnson, unpublished) are likely to be required to control the innate host response.

Adaptive Immune Responses in Nodavirus Infection

T cells are a crucial component of antiviral immunity and can determine whether a viral infection will be controlled and whether lasting immunity will be elicited. T cell responses have been characterized in only a limited number of studies in response to nodavirus infection. Data suggest that there is a specific role for cytotoxic T lymphocytes (CTL) in the modulation of nodavirus infection that is dependent upon MHC Class I and antigen-presenting cells. CTL responses have been characterized by the upregulation of cell markers, inflammatory cytokines and mediators of genes associated with cell lysis including granzyme, perforin and others. Importantly, this activity was limited in some species and allowed virally infected cells to evade killing. Other studies have confirmed observation of increased levels of T cell markers and Type II interferon following immunization with nodavirus, further suggesting that CD4⁺ T cells may also play a critical role in the reduction of viral loads.

Diagnosis

VNN

Diagnosis of a suspect case of VNN (VER) depends on at least one of the following criteria: abnormal swimming behavior in susceptible species; typical histopathological lesions in a susceptible species; typical cytopathic effects (CPE) observed in cell cultures; detection of specific antibody reactivity; any positive result obtained by RT-PCR and subsequent genomic sequencing,

quantitative PCR, or histopathology; transfer of live fish from an infected farm to another site; or the existence of epidemiological links between an infected farm and another farm. A confirmed case is one in which positive results are obtained from at least two of the following methods: a suspect case that has produced typical CPE in cell culture plus identification of the causative agent by a molecular test or an antibody-based test and a second diagnostic assay that might include isolation from cultured infected cells followed by immunostaining, reverse transcription PCR (RT-PCR), or quantitative RT-PCR (qRT-PCR). While serological diagnostic methods are not yet used as a screening method for assessing the presence of *Betanodavirus* species in fish populations, a recent study generated novel antibodies suitable for development as a serological test for diagnosis of VNN.

WTD

Diagnosis of a suspect case of WTD requires that at least one of these criteria are met: clinical signs or histopathology consistent with MrNV infection or a positive result by RT-PCR or qRT-PCR. A confirmed case fulfills two or more of the following: MrNV-positive histopathology, positive result in target tissues, RT-PCR followed by sequencing, or qRT-PCR. MrNV can be detected by *in situ* hybridization of histological specimens using a digoxigenin-labeled probe. Although no serological methods have been developed to date, both genome (qRT-PCR) and antibody-based (ELISA) tests are available for diagnostic use. The virus can be propagated in nodavirus-susceptible cultured cell lines including those of the mosquito *Aedes albopictus* line C6/36 and the fish *Channa* (Ophicephalus) *striatus* (striped snakehead) line SSN-1. No specific CPE is observed but viral replication can be detected by acridine orange staining, quantitative RT-PCR, and EM.

VCMD

Diagnosis of shrimp exhibiting suspected clinical signs of CMD has been accomplished using various nucleic acid amplification tests including RT-LAMP, nested RT-PCR, and qRT-PCR. The RT-LAMP and RT-PCR assays test RNA isolated from the cephalothorax of suspected CMNV-infected shrimp. Because the RT-LAMP assay uses a fluorescent dye for detection rather than gel electrophoresis or qRT-PCR, it is suitable for use as a rapid detection method in the field.

Treatment

No treatment is available for nodavirus infection.

Prevention

VNN

Virus contaminated water can be sterilized by UV treatment. Fertilized fish eggs may be washed in ozone-treated seawater to remove viral particles from their outer surfaces. Treatment of rearing water with chlorine or ozone appears effective in controlling the disease in larvae of striped jack, sevenband grouper, barfin flounder, and Atlantic halibut. Immunization with recombinant viral capsid protein or virus-like particles or with formalin-inactivated virus has proven effective in preventing VNN. An inactivated RGNNV preparation that protects sevenband grouper from RGNNV infection is commercially available in Japan. More recently, the production of antisera that are specific for the capsid proteins of SJNNV, RGNNV, TPNNV, and BFNNV suggested the possibility of developing antibody-based vaccines that might prove cross-protective against all four betanodavirus genotypes.

WTD

Prevention may be accomplished using standard procedures for control of crustacean viral diseases, including use of iodophors or formalin to disinfect eggs and larvae. Preventive measures include disinfection of tanks and water supplies and screening of brood stock and PL to maintain pathogen-free populations. The OIE Manual of Diagnostic Tests for Aquatic Animals recommends against mixed cultures of *M. rosenbergii* and the giant tiger prawn *Penaeus monodon* or rotation between the two species in the same tank to decrease the likelihood of spread. No vaccine is available for WTD.

VCMD

Preventative measures include deployment of a reverse transcription loop-mediated rapid isothermal amplification (RT-LAMP) detection kit for screening brood stock for CMNV prior to stocking of ponds and adoption of indoor culture methods. No vaccine is available for VCMD.

Further Reading

- Ball, L.A., Johnson, K.L., 1998. Nodaviruses of insects. In: Miller, L.K., Ball, L.A. (Eds.), *The Insect Viruses*. New York: Plenum Publishing Corporation.
- Bandin, I., Souto, S., 2020. Betanodavirus and VER disease: A 30-year research review. *Pathogens* 9.
- Chaves-Pozo, E., Valero, Y., Esteve-Codina, A., *et al.*, 2017. Innate cell-mediated cytotoxic activity of European Sea bass leucocytes against nodavirus-infected cells: A functional and RNA-seq study. *Scientific Reports* 7, 15396.
- Chen, N.C., Yoshimura, M., Guan, H.H., *et al.*, 2015. Crystal structures of a piscine betanodavirus: Mechanisms of capsid assembly and viral infection. *PLoS Pathog* 11, e1005203.
- Fan, X., Dong, S., Li, Y., Ding, S.W., Wang, M., 2015. RIG-I-dependent antiviral immunity is effective against an RNA virus encoding a potent suppressor of RNAi. *Biochemical and Biophysical Research Communications* 460, 1035–1040.
- Gant Jr, V.U., Moreno, S., Varela-Ramírez, A., Johnson, K.L., 2014. Two membrane-associated regions within the Nodamura virus RNA-dependent RNA polymerase are critical for both mitochondrial localization and RNA replication. *Journal of Virology* 88, 5912–5926.
- Ho, K.L., Gabrielsen, M., Beh, P.L., *et al.*, 2018. Structure of the *Macrobrachium rosenbergii* nodavirus: A new genus within the Nodaviridae? *PLOS Biology* 16, e3000038.
- Hua, Y., Huang, Y., Liua, J., *et al.*, 2018. TBK1 from orange-spotted grouper exerts antiviral activity against fish viruses and regulates interferon response. *Fish and Shellfish Immunology* 73, 92–99.
- Huang, Y., Huang, X., Yang, Y., *et al.*, 2015. Involvement of fish signal transducer and activator of transcription 3 (STAT3) in nodavirus infection induced cell death. *Fish and Shellfish Immunology* 43, 241–248.
- Langevin, C., Alekseeva, E., Passoni, G., *et al.*, 2013. The antiviral innate immune response in fish: Evolution and conservation of the IFN system. *Journal of Molecular Biology* 425, 4904–4920.
- OIE, 2018a. Infection with *Macrobrachium rosenbergii* nodavirus (white tail disease). *Manual of Diagnostic Tests for Aquatic Organisms*. Paris, France: World Organisation for Animal Health.
- OIE, 2018b. Viral encephalopathy and retinopathy. *Manual of Diagnostic Tests for Aquatic Organisms*. Paris, France: World Organisation for Animal Health.
- Overgard, A.C., Patel, S., Nostbakken, O.J., Nerland, A.H., 2013. Atlantic halibut (*Hippoglossus hippoglossus* L.) T-cell and cytokine response after vaccination and challenge with nodavirus. *Vaccine* 31, 2395–2402.
- Panzarin, V., Toffan, A., Abbadi, M., *et al.*, 2016. Molecular basis for antigenic diversity of genus Betanodavirus. *PLoS One* 11, e0158814.
- Sahul Hameed, A.S., Ninawe, A.S., Nakai, T., *et al.*, 2018. ICTV virus taxonomy profile: Sarthroviridae. *Journal of General Virology* 99, 1563–1564.
- Sahul Hameed, A.S., Ninawe, A.S., Nakai, T., *et al.*, 2019. ICTV virus taxonomy profile: Nodaviridae. *Journal of General Virology* 100, 3–4.
- Sullivan, C.S., Ganem, D., 2005. A virus-encoded inhibitor that blocks RNA interference in mammalian cells. *Journal of Virology* 79, 7371–7379.
- Warrilow, D., Huang, B., Newton, N.D., *et al.*, 2018. The taxonomy of an Australian nodavirus isolated from mosquitoes. *PLoS One* 13, e0210029.

Nudiviruses (*Nudiviridae*)

Yu-Chan Chao, Chih-Hsuan Tsai, and Sung-Chan Wei, Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Baculovirus Member of the *Baculoviridae* family of occluded viruses that are pathogenic to insects and other invertebrates. The salient features of the viruses are their rod-shaped nucleocapsids and large circular, double-stranded, supercoiled DNA genome.

Episome A non-integrated extrachromosomal closed circular DNA molecule that may be replicated in the nucleus with chromosomal DNA.

Latent infection Viral infection without cell lysis, the viral genome but not virus particle is detectable.

Nucleocapsid Protein shell that encloses the viral genetic material.

Nudivirus Viruses that are generally not packed in an occlusion body.

Occlusion body A viral-encoded protein lattice that packs and protects the viruses in the environment.

Persistent infection Viral infection without cell lysis, but with the appearance of virus particles.

Productive/lytic infection A form of viral infection in which large numbers of viruses are produced and that usually results in cell lysis.

Classification

History

Nudiviruses are a highly diverse group of rod-shaped, enveloped, and circular double-stranded DNA (dsDNA) viruses. These viruses infect a wide variety of arthropod hosts including insects and crustaceans (shrimp). The word “Nudi-” comes from the Latin “*nudus*”, meaning naked, as several of the initially described members of this viral group do not produce occlusion bodies in which virion particles are embedded, as is the case for other large dsDNA insect viruses such as baculoviruses. Due to their structural features and viral replication processes being similar to those of baculoviruses, nudiviruses were previously classified as “non-occluded baculoviruses” (NOBs) in the *Baculoviridae*. However, due to differences in morphological and genomic characters, they were later removed from the family *Baculoviridae* and orphaned for years without having a clear classification. Several nudiviruses were also characterized as “intranuclear bacilliform viruses” (IBVs) during this period. Ultimately, in 2013, the new family *Nudiviridae* was assigned by the International Committee on Taxonomy of Viruses (ICTV), consisting of two new genera, *Alphanudivirus* and *Betanudivirus*. To date, a variety of nudiviruses and nudivirus-like viruses have been reported from various host species belonging to the Lepidoptera, Trichoptera, Diptera, Siphonaptera, Hymenoptera Neuroptera, Coleoptera, Homoptera, Thysanura, Orthoptera, Acarina, Araneina, and Crustacea (Table 1).

Criteria for Classification

Although nudiviruses were first characterized according to the absence of an occlusion body, increasing numbers of the *Nudiviridae* have been found to generate occlusion bodies, so this feature is no longer suitable for classification. Several criteria have been proposed by nudivirus researchers for classifying a virus as a member of the *Nudiviridae*: (1) virion morphology is typically rod-shaped and enveloped; (2) the virus has a large circular dsDNA genome and possesses the conserved nudivirus core genes (described further below); (3) the virus propagates in the nuclei of infected host cells and induces nucleus hypertrophy; (4) the virus can be transmitted orally or parenterally; and (5) the virus infects hosts at larval and/or adult stages and exhibits diverse tissue and cell tropisms.

As most of these morphological criteria may not clearly distinguish nudiviruses from other dsDNA arthropod viruses, molecular phylogenetic analyses using genomic data have become the most reliable means of classifying nudiviruses. These viruses exhibit phylogenetic distinctiveness from their closest viral taxa, e.g., baculoviruses (see Fig. 1). Most new members of the *Nudiviridae*, either re-assigned from other virus families or new isolates, are classified based on genetic analyses in spite of their various morphologies and physiologies.

Members of the *Nudiviridae*

Table 1 lists the current well-studied or confirmed nudiviruses, though several potential nudiviruses remain to be definitively described. Current classification puts both OrNV and GbNV in the genus *Alphanudivirus*, with HzNV-1 and HzNV-2 belonging to the genus *Betanudivirus*, and the rest remaining unassigned (Fig. 1). Apart from newly isolated nudiviruses, members of the *Nudiviridae* have mostly been re-assigned from other virus families and reclassified in the *Nudiviridae*. These re-classified viruses were renamed using the species name of the host from which they were first isolated. For instance, *Oryctes rhinoceros nudivirus* is now the type species of the genus *Alphanudivirus*. OrNV, from the beetle *Oryctes rhinoceros*, is a member of that species and is first

Table 1 Nudiviruses and hosts from which they were isolated

Host		Nudivirus (abbreviation)
Order	Species (common name)	
Lepidoptera	<i>Heliothis zea</i> (corn earworm)	Heliothis zea nudivirus – 1 (HzNV – 1)
	<i>Heliothis zea</i> (corn earworm)	Heliothis zea nudivirus – 2 (HzNV – 2)
Diptera	<i>Tipula oleracea</i> (cabbage crane fly)	Tipula oleracea nudivirus (ToNV)
	<i>Tipula paludosa</i> (European crane fly)	Tipula paludosa nudivirus (TpNV)
	<i>Drosophila innubila</i> (mycophagous fly)	Drosophila innubia nudivirus (DiNV)
	<i>Drosophila melanogaster</i> (fruit fly)	Kallithea virus (KV)
Orthoptera	<i>Gryllus bimaculatus</i> (two-spotted cricket)	Gryllus bimaculatus nudivirus (GbNV)
Coleoptera	<i>Oryctes rhinoceros</i> (rhinoceros beetle)	Oryctes rhinoceros nudivirus (OrNV)
Decapoda	<i>Penaeus monodon</i> (tiger shrimp)	Penaeus monodon nudivirus (PmNV)
	<i>Litopenaeus vannamei</i> (whiteleg shrimp)	Penaeus vannamei single nucleopolyhedrovirus (PvSNPV)
	<i>Homarus gammarus</i> (European lobster)	Homarus gammarus nudivirus (HgNV)

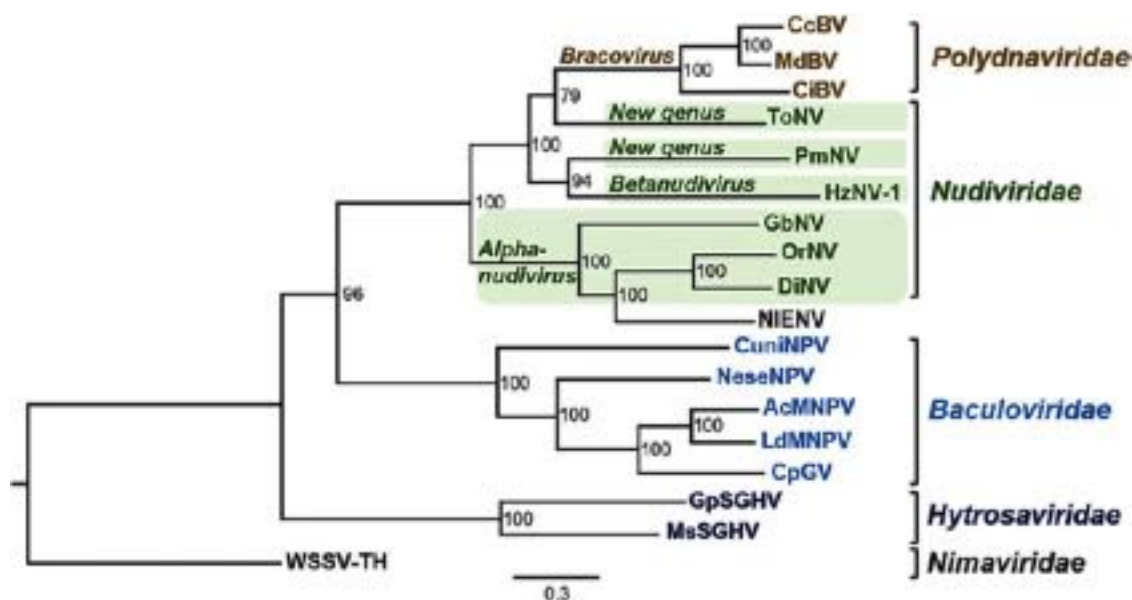


Fig. 1 Phylogeny of large dsDNA arthropod viruses. The tree was generated by Maximum Likelihood (ML) inference analysis using a concatenated amino acid alignment of 37 nudivirus-related genes. Numbers on the nodes indicate ML non-parametric bootstrap support (100 replicates). The white spot syndrome virus (WSSV-TH) was used as the outgroup. This tree has been reproduced with permission from: Bézier, A., Thézé, J., Gavory, F., *et al.*, 2015. The genome of the nucleopolyhedrosis-causing virus from *Tipula oleracea* sheds new light on the *Nudiviridae* family. *Journal of Virology* 89, 3008–3025.

nudivirus discovered had been variously named “Rhabdionvirus oryctes”, “baculovirus of Oryctes”, or “Oryctes baculovirus” and was assigned as a subtype in different virus families, e.g., the family *Baculoviridae* Subgroup C (non-occluded rod-shaped nuclear viruses) and the Oryctes virus family. GbNV that was isolated from the two-spotted cricket, *Gryllus bimaculatus*, was previously called the “baculovirus of cricket” or “cricket baculovirus”. *Heliothis zea nudivirus* is the type species of the *Betanudivirus* genus with HzNV-1 as a member of that species. HzNV-1 was isolated from *Heliothis zea* ovarian tissue cell lines and its genome was the first nudivirus genome to be sequenced. It has variously been known as “IMC-HZ-I-NOV”, “baculovirus X”, “Hz-1 baculovirus”, and “Hz-1 virus (Hz-1V)”, and it was considered to belong to the type species of the subfamily *Nudibaculovirinae* in the *Baculoviridae*.

Reassignment of PmNV into the *Nudiviridae* was more complicated as this shrimp-infecting virus forms occlusion bodies (as do baculoviruses) and it had been designated “monodon baculovirus” (MBV), “*Penaeus monodon* baculovirus”, “*Penaeus monodon* singly enveloped nuclear polyhedrosis virus” (PmSNPV) in the subgenus of single nucleocapsid SNPVs, and “*Penaeus monodon*

nucleopolyhedrovirus" (PemoNPV) in the former *Nucleopolyhedrovirus* genus. This virus is now considered as an unclassified nudivirus belonging to the family *Nudiviridae* based on molecular phylogenetic analysis and consequently was renamed "Penaeus monodon nudivirus (PmNV)". Genomic data on PmNV supports this new classification and may even suggest that a potentially distinct genus under *Nudiviridae* should be assigned (described below). However, the name MBV is still commonly used in the literature to describe this virus. Another occluded nudivirus, TpNV, which infects crane flies, was previously named "T. paludosa baculovirus (TpBV)" and "T. paludosa nucleopolyhedrovirus (TpNPV)".

Apart from the above-mentioned nudiviruses characterized according to their structural features of the virion particle, HzNV-2 representing the second nudivirus to be isolated from the corn earworm *Heliothis zea*, which was previously named as "gonad-specific virus (GSV)", "*H. zea* reproductive virus", and "Hz-2V", primarily due to pathogenies in its hosts. Newly isolated nudiviruses are now more practically designated according to the hosts from which they have been isolated, such as DiNV that infects the mushroom-feeding fly *Drosophila innubila*, and ToNV that infects the cabbage crane fly *Tipula oleracea*. Kallithea virus (KV) is the first DNA virus to be isolated from wild *D. melanogaster* populations and has temporarily been named according to the collection locality.

Phylogeny (Evolution of *Nudiviridae* and Other dsDNA Arthropod Viruses)

Phylogenetic analyses of viral core genes (described in "Genome" section) reveals that nudiviruses can be separated from baculoviruses (Fig. 1, shaded in green). The *Nudiviridae* exhibits several clades, including from the *Alphanudivirus* and *Betanudivirus* genera. The newly sequenced PmNV and ToNV genomes suggest the existence of two additional genera. Notably, the phylogeny demonstrates that the bracoviruses (Family *Polydnviridae*) may derive from nudiviruses. Bracoviruses form a symbiotic relationship with braconid wasps and are obligatorily parasitic on their Lepidopteran hosts. The presence of a significant number of genes homologous to those of nudiviruses in bracoviruses suggests that these polydnviruses may have originated from the integration of an ancient nudivirus into the genome of a braconid wasp. Nudivirus-like genes have also been identified in *Nilaparvata lugens* endogenous nudivirus (NIENV) of the plant sap-sucking insect, *Nilaparvata lugens* (brown planthopper), but whether these genes are expressed or functional is uncertain.

The phylogeny displayed in Fig. 1 may reflect the evolution of these viruses. The common ancestor of baculoviruses and nudiviruses seems to have diverged upon the appearance of holometabolous insects (insects exhibiting completely different appearances in immature and mature stages). The baculovirus lineage became specialized in infecting larvae of the newly evolved dipteran, hymenopteran, and lepidopteran insects. In contrast, the nudivirus lineage evolved to infect larval and adult hosts from a variety of diverse arthropods and by means of complex infection strategies. Such strategies, including the establishment of latent infections, might have been instrumental for the nudiviral ancestor of the bracoviruses to integrate its genome into a parasitoid wasp chromosome.

Virion Structure

The enveloped virion of nudiviruses typically harbors a single rod-shaped nucleocapsid. The virion structures of the two recognized *Nudiviridae* genera exhibit different morphologies, with alphanudiviruses (e.g., OrNV and GbNV) having more rounded virion particles of $\sim 200 \times 100$ nm, whereas those of betanudiviruses (e.g., HzNV-1 and HzNV-2) are longer and thinner at $\sim 300\text{--}400 \times 80$ nm. Several nudiviruses exhibit protruding structures within their virions, such as that of OrNV with thickened or "capped" nucleocapsid ends and a tail-like appendage at one end of the nucleocapsid. A filamentous structure has frequently been observed between the nucleocapsids and dilated envelopes of GbNV and HzNV-2, whereas HzNV-1 does not exhibit this structure.

Although nudiviruses such as HzNV-1 and GbNV are non-occluded, many nudiviruses present occlusion body formation during viral transmission. The morphology of occlusion bodies also differs among nudiviruses. ToNV and TpNV form enveloped occlusion bodies, PmNV forms non-enveloped occlusion bodies, and both OrNV and HzNV-2 form facultative occluded virions. Although the occlusion body proteins or potential genes encoding nudivirus occlusion bodies usually bear low sequence similarities to those of baculoviruses (described further below), the similar morphologies of baculovirus and nudivirus occlusion bodies would seem to preclude using this feature as a criterion to distinguish these two families of viruses.

Transmission and Pathogenesis

Nudiviruses infect a broad diversity of host species and they exhibit tissue tropism and transmission modes distinct from those of their sister group, the baculoviruses. Unlike baculoviruses that infect primarily the midgut epithelium of lepidopteran larvae and are transmitted by feeding (*per oral* route), nudiviruses infect a variety of tissues of both larvae and adult hosts in a range of arthropod orders (Table 1), and they can be transmitted through feeding or mating (*per oral* and *per parenteral* routes, respectively). Pathogenies are therefore diverse and depend on the stage of the infected host. Specifically, OrNV infects both larvae and adult *O. rhinoceros* beetles via *per oral* transmission and targets primarily midgut epithelial cells. In larvae, the virus spreads to other tissues and causes the infected larvae to die by fat body disintegration. Infected adult beetles exhibit few obvious symptoms,

carrying abundant newly-synthesized viruses and releasing them in their feces. GbNV infects its host crickets through fecal-contaminated foods. It targets the fat body cells of cricket nymphs and adults. GbNV infection is lethal in early nymphs, whereas infected adult crickets may exhibit reduced size or become crippled. PmNV is transmitted via oral ingestion of foods contaminated with feces containing viral occlusion bodies. PmNV targets the anterior midgut epithelium of very young post-larvae of tiger shrimp. When infecting the post-larvae, juveniles and adults, PmNV targets the hepatopancreatic tubule epithelium and duct epithelium. PmNV infectivity has been shown to reduce shrimp productivity. The fly nudivirus, DiNV, diminishes the viability of infected flies and reduces the number of eggs they lay. TpNPV targets the blood cells and the infection delays the crane fly larval development before the larvae die.

Differences in the pathogenies of HzNV-1 and HzNV-2 infection are interesting. HzNV-1 can replicate in only cultured cells of lepidopteran tissue but the infection is not evident in its insect host, the corn earworm. In cultured cell lines, HzNV-1 results in productive infection which kills most of the cells, leaving very few survivals to become viral latently infected cells. During viral latency, viral DNA was found to present as an episome or as an insertion into the host's genome. In contrast, HzNV-2 can replicate and persistently infect corn earworm. The predominant means of transmission of HzNV-2 is through mating between infected moths. HzNV-2 infects the reproductive tissues of infected moths, causing malformation of those tissues and host sterility. The mating pheromone generated by infected females attracts mate-seeking males, further transmitting the HzNV-2 nudivirus.

Virus Life Cycle

Viral life cycles in infected cells have been studied for only a few nudiviruses. OrNV attaches to and enters cultured cells by pinocytosis, whereby the virus is brought into cells by means of a small vesicle formed by cell membrane invagination. OrNV may enter the nucleus in the form of an unenveloped nucleocapsid, although the uncoating process in the cytoplasm remains unknown. Initial virus replication occurs at ~7 h post infection (hpi) with a sign of cleared chromatin-free areas showing up in the nucleus. In these areas, envelopes and nucleocapsid shells form first, and the viral nucleoprotein core and DNA condense into these enveloped shells. As early as 12 hpi, the resulting virions enter the cytoplasm and bud out through the plasma membrane, where they acquire a second membrane. Viral budding is greatest at 16 hpi and lasts until at least 36 hpi. The assembly processes of OrNV and some nudiviruses are very different from baculoviruses; baculoviruses produce naked nucleocapsids in the nucleus and acquire their envelopes during budding from the plasma membrane or before occlusion into polyhedra or granule cores. In OrNV-infected cells, enveloped nucleocapsids have always been observed although the virus purified from the cells with only a single envelope exhibits similar infectivity as the virus acquiring a second envelope by budding.

In the cell line DSIR-HA-1179, OrNV produces at least 27 proteins of which 14 are thought to be envelope proteins and 13 are associated with the nucleocapsid. The most abundant protein, P13, can be synthesized up to 36 hpi but most of the protein syntheses have ceased before it. Although the homologs of late or very-late genes appear in the genome, synthesis of these late proteins does not occur in OrNV.

HzNV-1 exhibits two infection stages in cultured cells, i.e., productive (lytic) and persistent (latent) infections (**Fig. 2**). DNA replication and virus assembly processes in productive infection of HzNV-1 are similar to those of OrNV. That is, enveloped, rod-shaped virus particles are first formed in the nucleus of infected cells and then filled with capsids containing the viral DNA. Viruses are released from infected cells by cell lysis. The early sign of viral DNA replication occurs at 4 hpi, and the replication cycle can be divided into three stages based on the timing that viral proteins appear: early (0–4 hpi), intermediate (4–8 hpi), and late (beyond 8 hpi). As high titers of virus progeny are produced, most infected cells ultimately die. However, a small percentage (usually less than 5%) of infected cells become persistently infected. During this stage, virus progenies are detected with low titers for a prolonged period of time and for many culture passages without obvious cell lysis. The persistently infected cells can be established by the viruses in both productive and latent infection cycles. Upon the establishment of persistent viral infection, virions are not detectable in most of the infected cells and viruses exist either as episomes or are inserted into the host genome (**Fig. 2**). At the same time, in some very few cells (less than 0.2%), viruses undergo the productive viral infection cycle, resulting in viral titers of approximately 10^3 plaque-forming unit (PFU)/ML in the culture medium. The phenomenon of HzNV-1 persistent infection is more appropriate to be described as a latent infection, in which virus is usually undetectable, and intermittent acute symptoms may occur in infected cells or tissues. During latency, insect cells are resistant to infection with homologous viruses (the same viruses), but they can be reactivated to result in a productive infection by infecting them with heterologous viruses (different viruses).

After several serial, high-multiplicity passages of HzNV-1 virus in cell culture, defective interfering particles (DIPs) show up in the culture. These particles contain smaller genomes, which might be acquired from deletions in the HzNV-1 genome. Interestingly, 37 HzNV-1 viral proteins identified in virus with a complete genome are all synthesized by DIPs, but the relative rates of synthesis of many of these proteins are altered in DIPs (thirteen decelerate and five accelerate). The contribution of DIPs to the establishment of latency remains unknown. Temporal gene expression profiles have also been analyzed for HzNV-1. During productive infection, gene expression can be categorized into early stage (0–2 hpi), intermediate stage (2–6 hpi), and late stage (after 6 hpi), with at least 101 transcripts expressed; during latent infection, there is only one detectable transcript, i.e., the persistency-associated transcript 1 (PAT1) that is expressed by the *persistence-associated gene 1* (*pag1*).

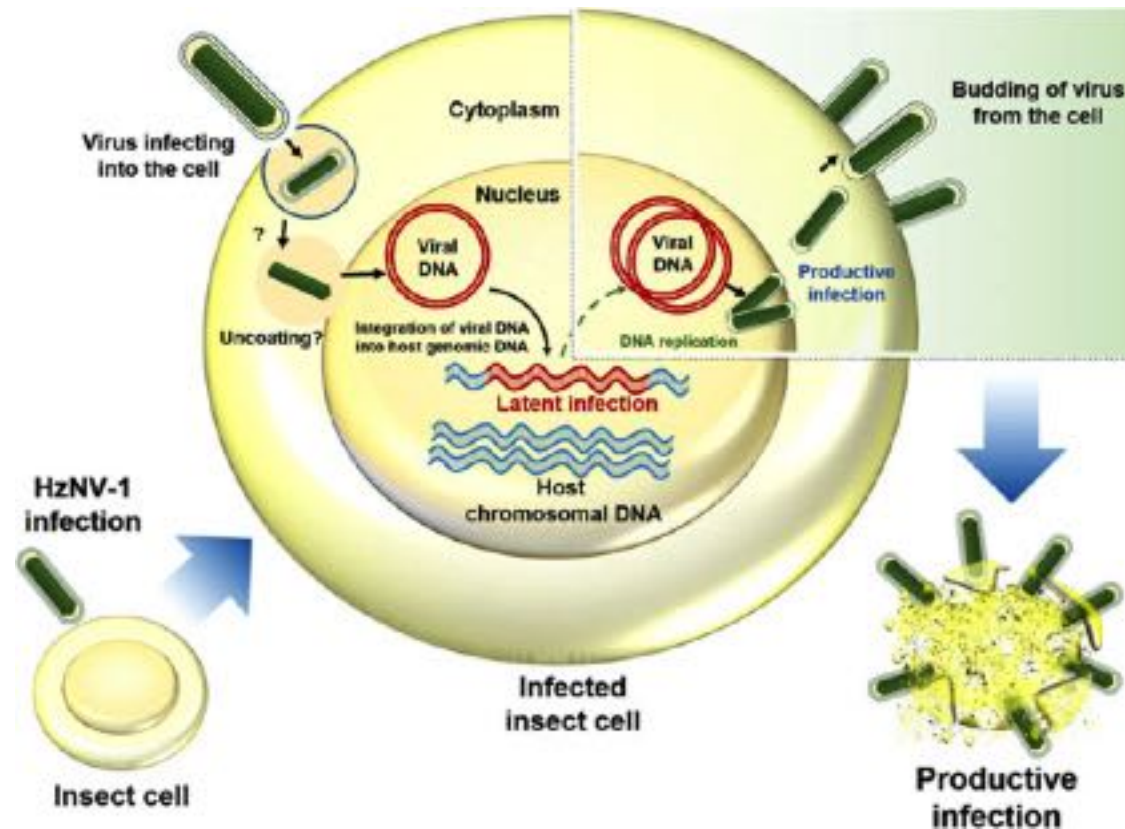


Fig. 2 Bi-phasic infection of HzNV-1 in insect cells. After HzNV-1 enters the infected cell, the virus may be uncoated and enter the nucleus in the form of an unenveloped nucleocapsid. In the nucleus, the viral DNA may insert into the host chromosomal DNA or exist as episomes for latent infection. In this infection stage, viral particles are not detectable and only the *pag1* gene expresses its transcript. Occasionally, HzNV-1 switches to productive infection in which viral DNA and proteins are abundantly produced. The enveloped, rod-shaped viral nucleocapsids are formed and filled with viral DNA in the nucleus. Viruses obtain their second envelope from the plasma membrane when they bud out from the cell. The budding eventually leads to the lysis of the infected cell.

Table 2 Comparison of nudivirus genomes

Virus	Genome size (base pairs)	No. of ORFs ^a	G + C content (%)	Gene density (per kilobase pairs)
HzNV – 1	228,089	155	41.8	1.47
HzNV – 2	231,621	113	42.0	2.05
OrNV	127,615	139	42.0	0.92
GbNV	96,944	98	28.0	0.99
PmNV	119,638	115	34.5	1.04
ToNV	145,704	131	25.5	1.11
DiNV	155,555	107	30.0	1.45
KV	152,388	95	38.9	1.60
HgNV	107,063	97	35.3	1.10

^aORFs: open-reading frames.

Genome

General Features

Nudiviruses whose genomes have been fully sequenced are listed in [Table 2](#). These viral genomes vary greatly in size, G + C content, and number of coding sequences, except for the genomes of HzNV-1 and HzNV-2 that share 94% sequence identity. G + C contents of some nudivirus genomes, e.g., GbNV and ToNV, are particularly low, with a lowest for ToNV at 25.5%. Similar to other large dsDNA viruses, nudivirus genomes feature direct repeat sequences that are extremely AT-rich (up to 98%) and with 2–3 copies of short reiterated sequences. However, these direct repeats are not homologous to each other within or between nudivirus genomes, or with those of other large dsDNA viruses. Numbers of direct repeats also differ among nudivirus genomes.

Table 3 The conserved baculovirus core genes in nudivirus genomes

Gene name	Encoding product
<i>dnapol</i>	DNA polymerase
<i>helicase</i>	Helicase
<i>p47</i> , <i>lef-4</i> , <i>lef-8</i> , <i>lef-9</i>	RNA polymerase subunit
<i>lef-5</i>	Late expression factor
<i>vlf-1</i>	Very late expression factor
<i>p74</i> (<i>pif-0</i>), <i>pif-1</i> , <i>pif-2</i> , <i>pif-3</i> , 19 kDa (<i>pif-4</i>), <i>odv-e56</i> (<i>pif-5</i>), <i>ac68</i> (<i>pif-6</i>), <i>vp91</i> (<i>pif-7</i>)	<i>per os</i> infectivity factor (PIF)
<i>vp39</i>	Major capsid protein
<i>38 K</i>	Viral phosphatase
<i>p33</i>	Sulfhydryl oxidase
<i>ac81</i>	Unknown function

For instance, the HzNV-2 genome harbors six direct repeats, the ToNV genome has five, the GbNV genome contains 14, and the DiNV genome has 156 simple direct repeats that account for 5.1% of the genome. Homologous repeat regions composed of direct repeats but with an imperfect palindromic core sequence commonly exist in the genomes of baculoviruses. In contrast, no homologous repeats have been found in most sequenced nudivirus genomes, apart from in the ToNV genome that hosts several direct repeats containing imperfect palindromic cores that are reminiscent of homologous repeats.

Gene Content

Nudivirus genomes typically share a set of 30–32 common genes, so-called “nudivirus core genes”. Although the members of these nudivirus core genes are likely to be debated as more nudivirus genomes become available, currently 20 nudivirus core genes homologous to baculovirus core genes (Table 3) are present in the nudivirus genomes sequenced to date. These genes are responsible for different functions in both baculovirus and nudivirus life cycles. Their presence evidences the close relationship between nudiviruses and baculoviruses, but phylogenetic analyses of these genes also support their divergence (described in “Phylogeny (Evolution of Nudiviridae and Other dsDNA Arthropod Viruses)” section).

Among the 20 nudivirus core genes shared with baculoviruses, eight encode homologs of baculovirus *per os* infectivity factor (PIF). PIFs are conserved among all baculoviruses and are essential for oral infection of insect hosts. Since many nudiviruses also exhibit *per oral* transmission, the occurrence of baculovirus PIF homologs suggests that a conserved infection mode exists in both nudiviruses and baculoviruses.

Genes encoding the four RNA polymerase units homologous to baculovirus variants – P47, LEF-4, LEF-8, and LEF-9 – are also present among the nudivirus core genes, as is a late expression factor, LEF-5, and a very late expression factor, VLEF-1. Thus, nudiviruses seem to have a baculovirus-like RNA transcription apparatus and transcribe late gene expression in a similar manner to baculoviruses.

Aside from these 20 shared core genes, some nudiviruses present a gene homolog to the baculovirus *polyhedrin/granulin* (*polh/gran*) gene. In baculoviruses, *polh/gran* encodes the polyhedrin protein that forms the occlusion body. The ToNV nudivirus obligatorily forms occlusion bodies, and its occlusion body protein is encoded by a baculovirus *polh* homolog. However, the gene encoding the major occlusion body protein of another occluded nudivirus, PmNV, does not share sequence similarities to baculovirus *polh* or *gran*. Whether the occlusion bodies of other occluded nudiviruses are encoded by *polh/gran* homologs awaits investigation.

Apart from genes encoding protein products, nudiviruses such as HzNV-1 and KV encode miRNAs. miRNAs are involved in the establishment of HzNV-1 latency and regulation of host histone modification. KV transcribes many small RNAs (21 nucleotides) during infection of *D. melanogaster*. These small RNAs may be the targets of *Drosophila* antiviral RNA interference (RNAi) responses.

Gene Regulation for Switching Productive and Latent Infections

The molecular mechanism responsible for switching between productive and latent infections has been analyzed for HzNV-1. A 6.2-kilobase early transcript encoded by the *hhi1* gene reactivates viruses from latently infected cells. The *hhi1* transcript has been detected very early during productive HzNV-1 virus infection, but it is not detectable during latent infection. When *hhi1* is expressed in latently infected cells, most of the cells lyse and release high titers of viral progeny. In contrast, when *hhi1* expression is suppressed (e.g., by siRNA), the number of latently infected cell clones is dramatically increased. The expression of *hhi1* transactivates the expression of viral early genes, which may contribute to switching the virus from latent infection to productive infection. Latency of HzNV-1 is also well regulated at the molecular level. During viral latency, the non-coding RNA PAT1 (transcribed by *pag1*) is the only detectable viral transcript. PAT1 constantly exists in both latent and productive infections. Overexpression of PAT1 in productively infected cells enhances HzNV-1 latent viral infection. Interestingly, *pag1* encodes at least two miRNAs that target the *hhi1* coding region. Both of these miRNAs can reduce the amount of *hhi1* transcript within cells to establish latent viral infection. The interactions between *hhi1* and *pag1* are illustrated in Fig. 3. During HzNV-1 infection, cells

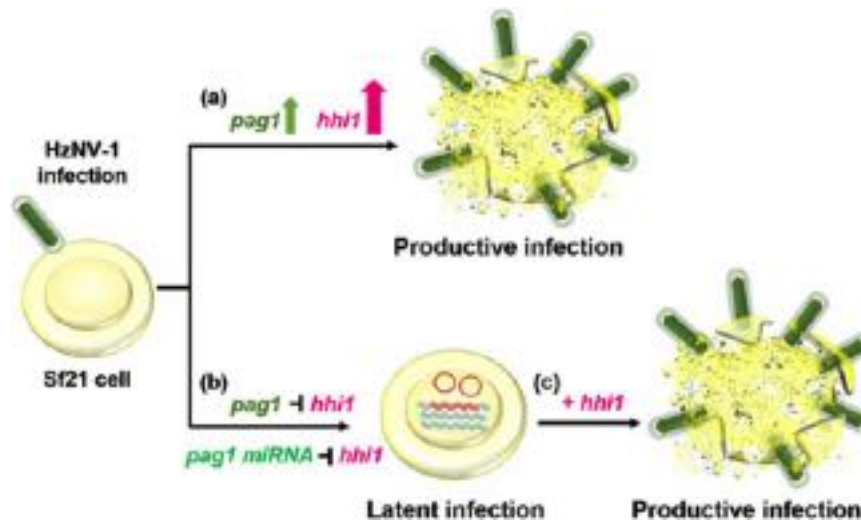


Fig. 3 Model for the establishment of productive or latent HzNV-1 viral infection via *hhi1* and *pag1* interactions. (a) Infection of HzNV-1 results in high levels of *hhi1* expression, and a moderate level of *pag1* expression. The abundant *hhi1* transcript can tolerate the RNA degradation caused by the *pag1* miRNAs, and transactivate viral early genes for activating the productive infection in most of the infected cells. (b) In a low percentage of cells, the level of *hhi1* transcript is low or undetectable. Suppression of *hhi1* transcript by *pag1* miRNA allows the virus to enter latency. (c) Viruses in the latently infected cells can be reactivated by introducing *hhi1* transcript, which results in virus production and cell lysis.

exhibiting high levels of *hhi1* expression can tolerate the RNA degradation caused by the *pag1* miRNAs and enter productive viral infection (Fig. 3(a)). However, if the expression level of *hhi1* in a cell becomes lower, or if *pag1* miRNAs are introduced, *hhi1* transcript expression will be suppressed by *pag1* via miRNA degradation and the virus enters latency (Fig. 3(b)). Re-activation of latent virus can be induced by introducing *hhi1* transcripts, resulting in virus production (Fig. 3(c)).

Negative Impacts and Potential Applications

PmNV infection was first identified in Taiwanese shrimp in 1977 and has since been reported worldwide. The infection of PmNV was initially thought to have resulted in the collapse of the penaeid shrimp culture industry in Taiwan in 1988. Although subsequent analyses suggested the collapse may have been due to environmental management or infection by other pathogens, PmNV is still closely linked to losses in many shrimp culture facilities in Taiwan and the Indo-Pacific region. To control PmNV infection, infected stocks need to be eradicated and the hatcheries must be disinfected using chemicals such as formalin, chlorine, and iodophor at sufficient concentrations.

Use of OrNV as a biocontrol agent to suppress its beetle host *O. rhinoceros* is a classical but rare example of (initially) successful biological control. *O. rhinoceros* damaged both coconut and oil palms in many Pacific islands in the early 1900s. OrNV was first used in Samoan fields in 1967 to control this beetle. Symptomless adult beetles infected with the nudivirus become natural vectors for OrNV spread, and beetle populations were effectively under control for over 30 years. However, a new wave of invasive *O. rhinoceros* became established in 2007 that were resistant to OrNV infection. Since resistant beetles had already been detected in 1989, it is unclear if the OrNV biocontrol process catalyzed the emerging resistance in *O. rhinoceros*.

Further Reading

- Bézier, A., Thézé, J., Gavory, F., et al., 2015. The genome of the nucleopolyhedrosis-causing virus from *Tipula oleracea* sheds new light on the Nudiviridae family. *Journal of Virology* 89, 3008–3025.
- Burand, J.P., 1998. Nudiviruses. In: Miller, L.K., Ball, L.A. (Eds.), *The Insect Viruses*. New York: Plenum, pp. 69–90.
- Federici, B.A., Bigot, Y., 2003. Origin and evolution of polydnaviruses by symbiogenesis of insect DNA viruses in endoparasitic wasps. *Journal of Insect Physiology* 49, 419–432.
- Harrison, R.L., Herniou, E.A., Bézier, A., et al., 2019. *Nudiviridae* in the ICTV report on virus taxonomy 10th report. ICTV. Available at: https://talk.ictvonline.org/ictv-reports/ictv_online_report/dsna-viruses/w/nudiviridae.
- Harrison, R., Hoover, K., 2012. Baculoviruses and other occluded insect viruses. *Insect Pathology*. Elsevier, pp. 73–131.
- Huger, A.M., 2005. The *Oryctes* virus: Its detection, identification, and implementation in biological control of the coconut palm rhinoceros beetle *Oryctes rhinoceros* (Coleoptera: Scarabaeidae). *Journal of Invertebrate Pathology* 89, 78–84.
- Jehle, J.A., Burand, J., Herniou, E., et al., 2013. Creation of a new family *Nudiviridae* including two new genera and three species. *Taxonomy Proposal*. ICTV. Available at: https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/invertebrate-official/4819.
- Lin, C.L., Lee, J.C., Chen, S.S., et al., 1999. Persistent Hz-1 virus infection in insect cells: Evidence for insertion of viral DNA into host chromosomes and viral infection in a latent status. *Journal of Virology* 73, 128–139.

- Reil, J.B., Doorewaard, C., San Jose, M., *et al.*, 2018. Transpacific coalescent pathways of coconut rhinoceros beetle biotypes: Resistance to biological control catalyses resurgence of an old pest. *Molecular Ecology* 27, 4459–4474.
- Wang, Y., Burand, J.P., Jehle, J.A., 2007. Nudivirus genomics: Diversity and classification. *Virologica Sinica* 22, 128–136.
- Wang, Y., Jehle, J.A., 2009. Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: New insights on an old topic. *Journal of Invertebrate Pathology* 101, 187–193.
- Williams, T., Bergoin, M., Van Oers, M.M., 2017. Diversity of large DNA viruses of invertebrates. *Journal of Invertebrate Pathology* 147, 4–22.
- Wu, Y.L., Wu, C.P., Liu, C.Y., *et al.*, 2011. A non-coding RNA of insect HzNV-1 virus establishes latent viral infection through microRNA. *Scientific Report* 1, 60.

Parvoviruses of Invertebrates (*Parvoviridae*)¹

Judit J Péntzes and Hanh T Pham, National Institute of Scientific Research – Armand-Frappier Health Research Centre, Laval, QC, Canada

Qian Yu, School of Life Sciences, Jiangsu University, Zhenjiang, China

Max Bergoin, National Institute of Scientific Research – Armand-Frappier Health Research Centre, Laval, QC, Canada

Peter Tijssen, National Institute of Scientific Research – Armand-Frappier Health Research Centre, Microbiology and Immunology, Laval, QC, Canada

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Ambisense and monosense An ambisense single-stranded RNA virus (as in segmented negative stranded viruses) refers to a virus in which both the –ve and +ve sense RNA segments encode proteins. For the ssDNA parvoviruses both the +ve, and its –ve copy, ssDNA genomes are packaged and, after conversion to a dsDNA in the recipient host, encode mRNAs and proteins. In contrast, for monosense ssDNA parvoviruses, genes are located on only the -ve strands that are predominantly packaged.

APOBEC ("apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like") is a family of evolutionarily conserved cytidine deaminases. A mechanism of generating protein diversity is mRNA editing as well hypermutation.

EVE-derived immunity (EDI) widespread recruitment of endogenous viruses as regulatory elements for immune genes that points to a systematic evolutionary process in their co-option for host immunity.

Monophyly Members of a monophyletic taxon are all descendants of one common ancestor. Modern taxonomy aims to establish exclusively monophyletic taxa. A good example is the vertebrate taxon of amniotes, unifying mammals, birds and reptiles.

Paraphyly A taxon is considered paraphyletic if it excludes at least one lineage from its definition, even though given excluded lineage(s) still clearly derive(s) from the same common ancestor as the lineages included in the defined taxon. A well-known example is the definition of reptiles as taxon without the inclusion of birds.

Polyphyly A taxon is polyphyletic if it includes lineages derived from a common ancestor, which itself is excluded from the taxon definition. Warm-blooded animals make a good example, as it unifies birds and mammals but excludes reptiles, though all three derive from the same reptile-like common amniote ancestor.

Introduction

Parvoviruses of invertebrates are named densoviruses (DVs, formerly denso-nucleosis viruses or DNVs) although their demarcation characteristics from vertebrate parvoviruses (PVs) are increasingly blurred through the discovery of a third lineage (now classified as subfamily *Hamaparvovirinae*), which encompasses both vertebrate and invertebrate parvoviruses and is related to several endogenous viral elements (EVEs) identified as integration into the genomes of members from all major arthropod lineages. Linear single-stranded DNA (ssDNA) virus families of animals are relatively scarce since: (1) ssDNA cannot be transcribed to yield translation products such as a viral DNA polymerase; (2) linear ssDNA cannot, in contrast to linear single-stranded RNA, be replicated without loss of genome termini (corresponding to Okazaki fragments) of genomic integrity unless complex hairpin structures, or a protein-primed mechanism, are introduced to create double-stranded DNA first; (3) the absence of a complementary DNA strand, as a repair template, increases the mutation rate, and therefore limits the maximum length (2–12 kb with very few exceptions; to that of RNA viruses. PVs and DVs have small icosahedral capsids with a diameter of 18–27 nm with a limited packaging capacity). Although most DVs are highly pathogenic, some are essential for the life cycle of their hosts. Infection with *Dysaphis plantaginea* densovirus (DpIDV) is required to produce the winged morph in asexual clones of the rosy apple aphid and its colony dispersal to neighboring plants. This mutualistic relationship between the rosy apple aphid and its viruses results both in a negative impact of DpIDV on rosy apple aphid reproduction, but also contributes to the survival of aphids by inducing wing development and promoting dispersal.

The ssDNA viruses are, apart from those of the bidnaviruses, exceptionally widespread in all domains of life despite the limited number of virus families. Metagenomic studies dramatically boosted our knowledge about the distribution of ssDNA viruses in the biosphere, from the human gut to hot springs. The diversity of ssDNA viruses appears to be determined by two principal factors: extremely high nucleotide substitution rates that approach those of RNA viruses, facilitating adaptation to different environments and pervasive recombination, both by DNA and RNA donors.

¹This article is dedicated to Prof. Michael Rossmann for his essential contributions to the structure of arthropod densoviruses and to Prof. Su De Ming, a pioneer of parvovirology in China. Both forerunners passed away while this article was in preparation.

Discovery, Taxonomy and Evolution of Densoviruses; of Polyphyly, Paraphyly and its Tangled Relationship With Vertebrate-Infecting Parvoviruses

DVs are defined as small icosahedral viruses containing a linear, monopartite, ssDNA genome with terminal hairpins which facilitate rolling hairpin replication (RHR). DV genomes are monopartite, though for the ambidensoviruses both + and – ssDNA strands are coding in an ambisense strategy.

The first discovery of a DV was in the early 1960s when an epizootic decimated a mass rearing of *Galleria mellonella* L. larvae, reared as fishing bait in France. Electron microscopic examination of thin sections from diseased larvae revealed hypertrophied nuclei containing an electron-dense viroplasm with thousands of tiny viral particles. After purification, these particles were observed as small isometric, nonenveloped virions, 22–25 nm in diameter, and their biochemical analysis showed that they contained single-stranded DNA. The virus causing the disease was designated as densovirus, which later was simplified to the name densovirus (DV). Diseases presenting the same type of histopathological features, and caused by the same type of viruses, were later described not only in Lepidoptera but also in other invertebrates. The cloning and sequencing of many DV genomes during the last two decades, while confirming their relatedness to the vertebrate parvoviruses, revealed an extraordinary diversity, reflecting very likely the evolutionary diversity of their hosts. To date, and despite the current *Parvoviridae* taxonomy which reflects the actual evolutionary relationships among the family members, regardless of host affiliation, parvoviruses (family *Parvoviridae*) infecting vertebrates are still referred to as parvoviruses and parvoviruses of invertebrates are traditionally united under the densovirus designation even though this kind of classification of the family would introduce polyphyly. In contrast to most DNA viruses, DVs lack the DNA polymerase gene and use a pleiotropic non-structural protein 1 (named NS1 or Rep) for Rolling-Hairpin Replication (RHR).

It became apparent throughout the *Parvoviridae* family that the SF3 helicase superfamily domain of the NS1 is the only conserved motif at the protein sequence level, hence NS1-based phylogeny was introduced as a basis of classification since 2014. Although this method initially conformed with the segregation of the family into subfamilies *Parvovirinae* and *Densovirinae* (sensu lato; referring to all invertebrate parvoviruses at the time), following the vertebrate-invertebrate host affiliation-based classification, the discovery of the first members from the current invertebrate-infecting genera *Brevihamaparvovirus*, *Penstylhamaparvovirus*, and *Hepanhamaparvovirus* revealed a diversity among densoviruses, which was unmatched with the rather well-conserved nature of the vertebrate-infecting *Parvovirinae*. Members of the latter share both NS1 and VP proteins of a clearly homologous nature proven by detectable sequence similarity, throughout basically the entire length of these proteins. In contrast, densoviruses could be linked together only by the above-mentioned short (approx. 200 aa long) SF3 helicase domain of detectable sequence similarity. If only the helicase domain is considered, however, certain densoviruses, such as hepanhamaparvoviruses, harbor the same amount of amino acid (aa) sequence similarity to certain densoviruses as they do to members of the vertebrate *Parvovirinae*.

Since 2012 and on a frequent basis, parvoviruses of a novel, divergent lineage, unified under the name “chapparvovirus”, have been detected in association with kidney and liver tissue, as well as with various secretions (such as blood) and excretions (outbreaks of enteritis) of vertebrates. As unexpected as it may be, nevertheless, phylogeny evidence revealed a close relationship of these vertebrate-infecting parvoviruses to invertebrate-infecting densoviruses, such as members of genera *Hepanhamaparvovirus*, *Penstylhamaparvovirus*, and *Brevihamaparvovirus*. To further complicate the situation, EVEs have been detected in arachnid, chilopod, entognath and coleopteran arthropod genomes, evidently derived from ancient members of the “chapparvovirus” lineage, previously believed to be yet another new genus of *Parvovirinae*. These findings led to the establishment of the first *Parvoviridae* family of mixed host affiliation, designated *Hamaparvovirinae*. Although the subfamily itself is well-supported, *Hamaparvovirinae* still appears to be the most heterogeneous one of the three subfamilies, as relationships between its three major clades, namely *Penstylhamaparvovirus*-*Brevihamaparvovirus* vs. *Hepanhamaparvovirus* and affiliated unclassified relatives vs. *Ichthamaparvovirus* and *Chaphamaparvovirus* have low support and they lack homologous capsid protein sequences (Fig. 1). This could suggest that *Hamaparvovirinae* may not be long-lived once novel parvoviruses related to these three clades will be discovered and provide more insights into evolutionary relationships within this diverse clade.

Back in 2014, the unification of several ambisense densovirus genera to establish the large and diverse genus, called Ambidensovirus, created another major, unforeseen problem. Even then Ambidensovirus did not conform with the then-established demarcation criteria, albeit the ambisense genome organization proved to be a strong enough argument to support its existence. Since then, however, multiple densoviruses have been described, eventually rendering Ambidensovirus paraphyletic as opposed to monophyletic, as some members were more closely related to monosense iteradensoviruses than they are to other ambidensoviruses. This suggests that the orientation of the genome coding order is subjected to plasticity and either the monosense or the ambisense strategy could probably be “re-invented” on multiple occasions (Fig. 1). The selective constraints driving this conversion, however, are not well-understood. Today, to reflect the diversity of the clade and to avoid the above-mentioned paraphyly problem, ambisense DVs segregate into seven genera, namely *Miniambidensovirus*, *Protoambidensovirus*, *Aquambidensovirus*, *Scindoambidensovirus*, *Hemiambidensovirus*, *Blattambidensovirus*, *Pefuambidensovirus*.

Despite the divergence in genome organization, members of the above-mentioned seven genera and *Iteradensovirus* share some features in common with each other but not with members of any other genus in the *Parvoviridae*, including (1) a shared NS1-based and helicase domain-based phylogeny; (2) their VP proteins appear to share a common ancestral VP protein gene and similar capsid structures; (3) at least 32% identity can be detected between the NS1 proteins of any itera- or ambidensovirus, but only much lower identities with other densoviruses; (4) possession of a phospholipase A2 (PLA2) domain in their largest minor capsid protein sequence. Based on these criteria, ambisense DVs and monosense iteradensoviruses are now united in one

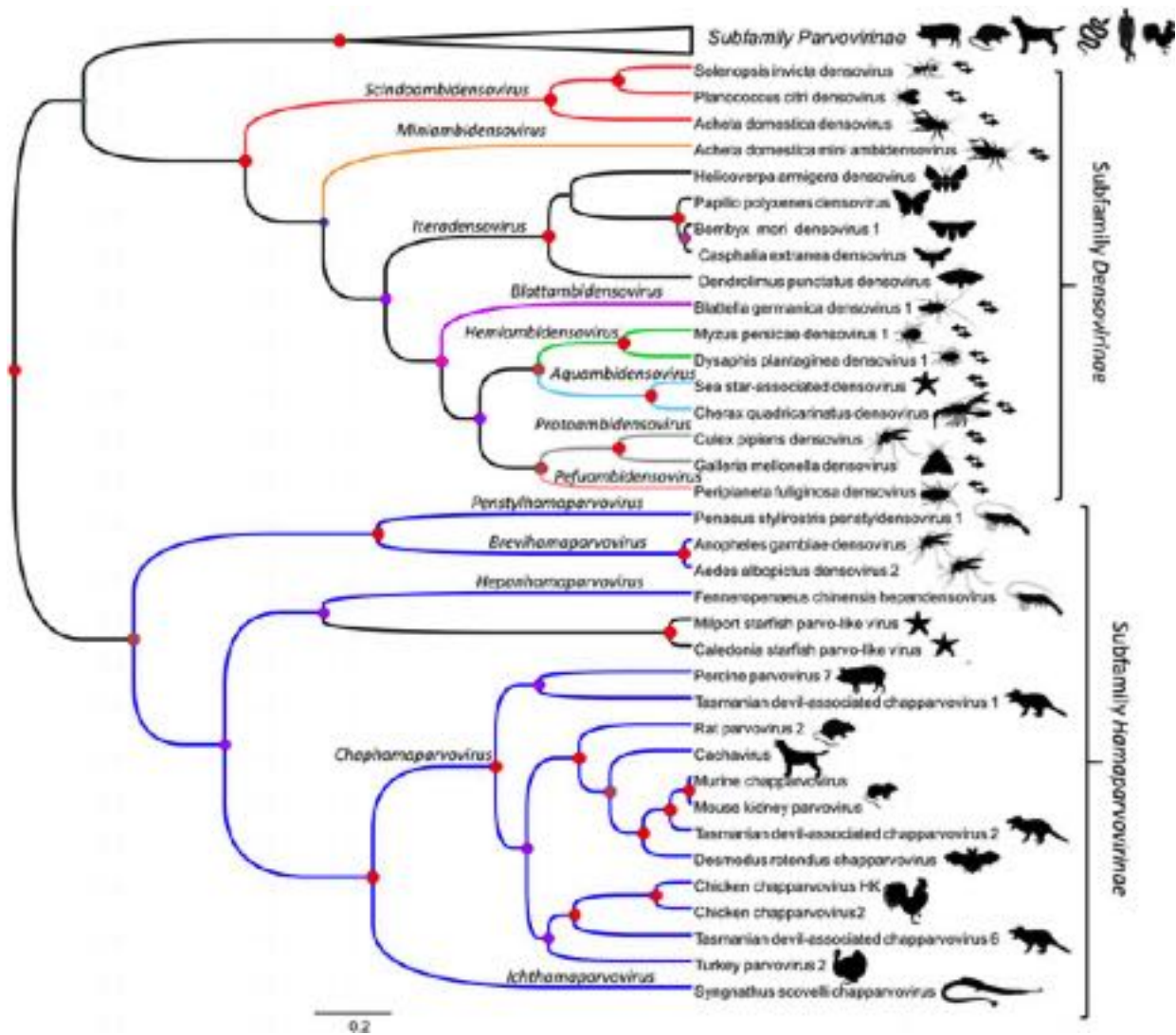


Fig. 1 Bayesian phylogenetic inference of family *Parvoviridae* based on the SF3 helicase domain. For the easier visibility subfamily *Parvovirinae* has been collapsed. Each tip label represents a separate species of its assigned genus (though murine chapparravirus is the same species as mouse kidney parvovirus). The current (2020) taxonomy is presented as branch labels, subfamily level taxonomy is presented to the right. Blue branches indicate lineages, which lack the phospholipase A2 domain from their large minor structural proteins. Branches presented in multiple colors correspond with genera introduced by the 2020 new taxonomy to resolve the former paraphyletic situation of ambisense densoviruses, previously classified as one genus. Ambisense genome organization is indicated by two small arrows facing toward each other. Host spectrum of each species or subfamily is indicated by the silhouette of the host animal. The reliability of the topology is indicated by posterior probability values, represented by the circles at the nodes; the larger and redder a circle is the higher the support is of given node as bright, large red circles indicate maximal support. This figure was created by BEAST v1.10.4.

subfamily on their own, which is subfamily *Densovirinae*. Given the large number of EVEs, discovered in various invertebrate genomes, which appear to be closely-related to this clade, it is to be expected that the current eight genera will soon be accompanied by several new ones, when the current dynamics of virus discovery are considered. The current taxonomy of DVs is summarized by [Table 1](#).

Biology, Pathology and Host Range

So far, DVs have been found in almost all major clades of the phylum Arthropoda, including arachnids, diplopods, decapod crustaceans and hexapods. As of the latter two, six different insect orders, as well as penaeid shrimps (family Penaeidae), crabs, and a crayfish (*Cherax quadricarinatus*) are known to be infected by DVs. Recent studies have derived DVs from members of the vast and diverse phylum of Mollusca, infecting oysters. The DV host spectrum also spans deuterostome invertebrates, derived from sea stars

Table 1 Master list of densovirus species. Former classification (Cotmore *et al.*, 2019): rows 1: Ambidensovirus, rows 2 Brevidensovirus, rows 3 Hepandensovirus, row 4 Iteradensovirus and row 5 Penstyldensovirus with acronyms and GenBank accession numbers. References are to be found in the GenBank listings. Revision of the family, accepted in 2020 (Penzes *et al.*, 2020): subfamily a: *Densovirinae* with ambidensoviruses, in the former genus *Ambidensovirus* (i) and genus *Iteradensovirus* (ii) and subfamily b: *Hamaparvovirinae* with genera *Brevihamaparvovirus* (x), *Hepanhamaparvovirus* (y), *Penstylghamaparvovirus* (z). The 4th column contains the names of the proposed genera of ambisense densoviruses

Former and current taxonomy	Virus Name	Name, abbreviated	GenBank Accession #	Genera of ambisense DVs
1.a.i	Periplaneta fuliginosa densovirus	PfDV	AF192260	<i>Pefuambidensovirus</i>
	Blattella germanica densovirus1	BgDV1	AY189948	<i>Blattambidensovirus</i>
	Culex pipens densovirus	CpDV	FJ810126	<i>Protoambidensovirus</i>
	Planococcus citri densovirus	PcDV	AY032882	<i>Scindoambidensovirus</i>
	Diatraea saccharalis densovirus	DsDV	AF036333	<i>Protoambidensovirus</i>
	Galleria mellonella densovirus	GmDV	L32896	<i>Protoambidensovirus</i>
	Helicoverpa armigera densovirus	HaDV1	JQ894784	<i>Protoambidensovirus</i>
	Junonia coenia densovirus	JcDV	S47266	<i>Protoambidensovirus</i>
	Mythimna loreyi densovirus	MIDV	AY461507	<i>Protoambidensovirus</i>
	Pseudoplusia includens densovirus	PiDV	JX645046	<i>Protoambidensovirus</i>
	Acheta domesticus densovirus	AdDV	HQ827781	<i>Scindoambidensovirus</i>
	Acheta domesticus minidensovirus	AdMDV	KF275669	<i>Miniambidensovirus</i>
	Acheta domesticus segmented densovirus	AdSDV		
	Asteroid densovirus	SsaDV	PRJNA253121	<i>Aquambidensovirus</i>
	Cherax quadricarinatus densovirus	CqDV	KP410261	<i>Aquambidensovirus</i>
	Solenopsis invicta densovirus	SiDV	SRX023381	<i>Scindoambidensovirus</i>
	Myzus persicae densovirus 1	MpDV1	AY148187	<i>Hemiambidensovirus</i>
	Dysaphis plantaginea densovirus	DplDV	EU851411	<i>Hemiambidensovirus</i>
2.b.x	Aedes aegypti densovirus1	AaeDV	M37899	
	Aedes albopictus densovirus1	AalDV1	AY095351	
	Culex pipiens pallens densovirus	CppDV	EF579756	
	Anopheles gambiae densovirus	AgDV	EU233812	
	Aedes aegypti densovirus2	AaeDV2	FJ360744	
	Aedes albopictus densovirus 2	AalDV2	X74945	
	Aedes albopictus densovirus 3	AalDV3	AY310877	
3.b.y	Haemagogus equinus densovirus	HeDV	AY605055	
	Penaeus monodon hepadensovirus 1	PmoHDV1	NC_007218	
	Penaeus chinensis hepadensovirus	PchDV	AY008257	
	Penaeus monodon hepadensovirus 2	PmoHDV2	EU247528	
	Penaeus monodon hepadensovirus 3	PmoHDV3	EU588991	
	Penaeus merguensis hepadensovirus	PmeDV	DQ458781	
	Penaeus monodon hepadensovirus	PmoHDV4	J410797	
4.a.ii	Fenneropenaeus chinensis hepadensovirus	FchDV	FJN082231	
	Bombyx mori densovirus	BmDV1	AY033435	
	Casphalia extranea densovirus	CeDV	AF375296	
	Sibine fusca densovirus	SfDV	JX020762	
	Dendrolimus punctatus densovirus	DpDV	AY665654	
	Papilio polyxenes densovirus	PpDV	JX110122	
	Helicoverpa armigera densovirus	HaDV2	HQ613271	
5.b.z	Penaeus stylirostris penstyldensovirus 1	PstDV1	AF273215	Q411199
	Penaeus monodon penstyldensovirus 1	PmoPDV1		
	Penaeus monodon penstyldensovirus 2	PmoPDV2	AY124937	
	Penaeus stylirostris penstyldensovirus 2	PstDV2	GQ475529	

and sea urchins (phylum Echinodermata). Moreover, using a metagenomic approach, DVs have been derived from eumetazoan animals even a primitive as cnidarians while the presence of endogenous DV-like elements in nematodes and platyhelminths implies that the DV host spectrum, in fact, might span most eumetazoan invertebrate phyla. DVs, which have been characterized thoroughly, were derived mainly from insects (orders Lepidoptera, Diptera, Hymenoptera (e.g., red imported fire) Blattodea, Hemiptera and Orthoptera but also decapod crustaceans (giant tiger prawn, crayfish), and echinoderms (starfish and sea urchins)).

However, our knowledge on the true extent of the DV host spectrum is still limited. Invertebrates comprise multiple phyla and constitute over 1 million species. Moreover, some species, e.g., the common house cricket, are infected by at least four different ssDNA viruses. Phylogenetic analysis revealed that these densoviruses may be as diverse as their hosts, therefore a definite classification is not yet possible (Fig. 1). The hypertrophied nuclei intensively stained by Feulgen reagent is a salient feature shared by all densovirus-infected tissues to date, where characterized. Thin sections of heavily infected cells observed by EM show hypertrophied nuclei filled by a virogenic stroma containing thousands of virions and cytoplasmic paracrystalline virion arrays.

General Features

Most of the DVs so far isolated are lethal for their natural hosts, the first symptoms being anorexia and lethargy followed by flaccidity. After a progressive paralysis, a slow melanization and death follow. The smokybrown cockroach (*Periplaneta fuliginosa*), infected by PfDV, displays characteristic hind leg paralysis, uncoordinated movements and a hypertrophied abdomen. The oil palm pests *Casphalia extranea* and *Sibine fusca* (in Ivory Coast and Colombia, respectively), when infected with CeDV and SfDV display tumor lesions in the gut. Infections are highly contagious and rapidly spread leading to severe epizootics both in noxious insects such as lepidopteran defoliators, mosquito larvae, aphids or cockroaches but also unfortunately in mass rearing facilities of commercially significant invertebrates such as *Bombyx mori* in sericulture farms, crickets in cricket farms and industrial shrimp farming. A densovirus is the causative agent of the recent extensive outbreak of sea-star wasting disease (SSWD). Due to their high pathogenicity, DVs have been successfully used to control Limacodidae larvae in oil palm plantations.

Most DVs are polytropic and infect the fat body, hypodermis, central nervous system, silk gland, muscular membrane, tracheal cells, Malpighian tubules, foregut, hindgut, hemocytes, ovaries, and molting glands but not the midgut. However, some iteradenosoviruses such as BmDV (*B. mori*), CeDV (*Casphalia extranea*) and SfDV (*Sibine fusca*), multiply predominantly in the midgut.

BmBDV, though not a densovirus, is a bidnavirus which causes fatal flacherie disease in the silkworm, replicates in contrast to most parvo-like densoviruses, but similarly to BmDV, only in midgut columnar cells. *B. mandarina* and *B. mori* are considered to have a common ancestor and both (at least native strains) are susceptible to BmBDV. Many Japanese strains acquired BmBDV-resistance through improved breeding. These resistant strains became susceptible to the parvovirus-like BmDV. The bidnavirus BmBDV is, therefore, likely the native disease of *Bombyx*, whereas the origin of BmDV is likely to be a pathogen from other insects. The pyralid *Glyphodes pyloalis*, a pest of mulberry plantations of sericultural farms is also susceptible to BmDV and probably the real/original host of BmDV.

The host range of the different DVs varies considerably. Some, like GmDV of *Galleria mellonella*, and CeDV are restricted to their original hosts, some, like AdDV of *Acheta domestica* and PfDV of *P. fuliginosa* to closely related hosts while others, e.g., Junonia coenia densovirus (JcDV), Mythimna loreyi densovirus (MlDV) and the mosquito DVs infect many different host species. Although those infecting the midgut have a restricted host range, the inverse is not true for the polytropic viruses. Mosquito larvae infected with mosquito DVs exhibit symptoms of paralysis. Mosquito cells cultured in vitro often contain DVs, despite the lack of cytopathic effects (CPE), that are pathogenic for mosquito larvae by per os infection. The Ld652 cell line derived from the gypsy moth (*Lymantria dispar*) also seems to be contaminated with a DV despite the lack of a CPE. Iteradenosoviruses replicate essentially in the columnar cells of midgut epithelium. Hepadenosoviruses (genus *Hepanhamaparvovirus*) of penaeid shrimps infect hepatopancreatic tubule epithelial cells characterized by basophilic intranuclear inclusions. The penstylidenosoviruses (genus *Penstylhamaparvovirus*) are cosmopolitan and infect different species of several penaeid shrimp genera.

Densoviruses do not infect vertebrates or cells derived from vertebrates, nor was any pathogenic effect observed after inoculation of rabbits or mice with these viruses. However, their ability to integrate into the host chromosome has been exploited for stable expression of foreign proteins in insect cells and somatic transformation of insects.

Shrimp Densoviruses

Two groups (genera) of shrimp densoviruses were identified as pathogens responsible for economically significant and virulent disease in farmed shrimp (infectious hypodermal and hematopoietic necrosis virus, PstDV1) and *Fenneropenaeus chinensis* hepadenosovirus (FchDV), now both members of the *Hamaparvovirinae* subfamily. They are widespread in nature but are no longer a major economic problem because tolerant shrimp populations have been developed.

Penaeus stylirostris penstylidensovirus 1 (PstDV1)

Penaeus stylirostris penstylidensovirus 1 (PstDV1) (genus *Penstylhamaparvovirus*) was originally discovered in cultured blue shrimp (*Penaeus stylirostris*) in Hawaii in 1981. Formerly known as infectious hypodermal and hematopoietic necrosis virus (IHNV), PstDV1 causes acute infection with up to 100% mortality in *P. stylirostris*. In Pacific white shrimp *P. vannamei* and giant tiger prawn (*P. monodon*), infection is mostly chronic and results in stunted growth and runt deformity syndrome (RDS) with wrinkled antennae, cuticular deformities of the rostrums and deformed sixth abdominal segments.

PstDV1 was reported to infect various decapod species of the Penaeidae family, from wild to cultured shrimps worldwide. Only *P. merguensis* seems to be refractory to PstDV infection. PstDV1 infection occurred in all stages of the shrimp life cycle, including eggs, larvae, postlarvae, juveniles, and adults. Histologic and in-situ hybridization analysis showed Cowdrey type A inclusion

bodies in ectodermal and mesodermal tissues of penaeids such as gills, cuticular epidermis, lymphoid organs, hematopoietic tissues, antennal glands and connective tissues.

PstDV1-related EVE sequences, lacking telomers, have been detected in the germline of *P. monodon* in Australia, Madagascar and Thailand. These would likely have been eliminated unless they provide beneficial effects, such as EVE-derived immunity (EDI, e.g., mediated by TRIM5a and APOBEC).

***Fenneropenaeus chinensis* hepadensovirus (FcHVD)**

FcHVD or previously named hepatopancreatic parvovirus (HPV) (genus *Hepanhamaparvovirus*) was first reported in banana shrimp, *P. merguensis*, and in Indian prawns, *P. indicus*, in Singapore. FcHVD later was found to infect *P. chinensis*, *P. monodon* and many other penaeid species worldwide in Asia, Australia, America, and Africa. The virus was also reported in giant freshwater prawn.

Infected shrimps do not always show diagnostic gross signs of FcHVD disease, but infection is associated with reduced growth rate in juveniles and can cause chronic mortalities at early- and post-larva stages. FcHVD infected shrimps appear to be more susceptible to other viral disease agents such as White Spot Syndrome Virus (WSSV) or, *Penaeus monodon* nudivirus (previously called, erroneously, *Monodon Baculovirus* (MBV)). Target tissues for viral replication include the hepatopancreas system and the tubule epithelial cells of the digestive gland.

Transmission routes of PstDV1 and FcHVD can be either horizontal or vertical. Horizontal transmission is by cannibalism or by water contamination and vertical transmission is via infected eggs. Multiple infection with different viruses is commonly found in cultured shrimps.

New invertebrate DVs

Sea stars, infected with sea star-associated densovirus (SsaDV), inhabiting the Northeast Pacific Coast have recently experienced an extensive outbreak of wasting disease, leading to their degradation and disappearance from many coastal areas. The signs of the disease are behavioral changes, lesions, loss of turgor, limb autotomy, and death characterized by rapid degradation. SsaDV could be detected in plankton, sediments and in nonasteroid echinoderms, providing a possible mechanism for viral spread. SsaDV was detected in museum specimens of asteroids from 1942, suggesting that it has been present on the North American Pacific Coast. Since June 2013, millions of sea stars (asteroids) off the west coast of North America have wasted away into slime and ossicle piles, due to the disease known as sea-star wasting disease (SSWD). SsaDV is related to densoviruses in the Hawaiian sea urchins (Echinoidea) *Colobocentrotus atratus*, *Echinometra mathaei*, and *Tripneustes gratilla*, placing it to the same species with other DVs of echinoderms.

Cherax quadricarinatus densovirus (CqDV), the causative agent of an epizootic in freshwater crayfish *Cherax quadricarinatus*, in Australia, has a large, ambisense genome of 6334 nt. It is closely related to the marine sea star densovirus (75% identity) and is the first ambidensovirus to be found in decapod crustaceans. Mortalities after infection with CqDV occurred over 4 weeks, with up to 96% cumulative mortalities in two earthen ponds stocked with juveniles. The crayfish were weak, anorexic and lethargic. A transmission trial was conducted, using filtered, cell-free extract prepared from infected crayfish as inoculum. The disease was reproduced, with ongoing mortalities occurring in inoculated crayfish over 55 days. Experimentally inoculated crayfish showed gross signs of malaise, anorexia, and disorientation before dying. Two types of intranuclear inclusion bodies (INIBs) were seen in tissues of endodermal, ectodermal and mesodermal origin by light microscopy with hematoxylin and eosin (H&E) stained sections.

Using metagenomics, densoviral nucleic acids in oysters (*Crassostrea ariakensis*) was identified, designated CaaDV1 and CaaDV2, assembling the nearly complete genome of both (5860 nucleotides (nt) and 4034 nt, respectively). Both genomes displayed an ambisense organization but were divergent from the existing densoviral species. The NS1 protein of the two CaaDVs shared 43.3% ~ 61.5% amino acid identities with SsaDV and CqDV as well as clustering with these in phylogeny calculations. The CaaDVs are potential new members of genus *Aquambidensovirus*, accompanying the echinoderm DVs and CqDV, all of aquatic host origin.

Solenopsis invicta densovirus (SiDV) was the first DNA virus discovered in fire ants from Argentina and is related to AddV and PcDV. The red fire ant, *Solenopsis invicta* (Buren) was introduced into the United States and due to the paucity of natural enemies is estimated to cause damage of about 6 billion USD annually. SiDV causes a high mortality rate and is a prime candidate for biological control.

Recently, a metagenomic study has derived partial DV genomes from several novel invertebrate hosts, of various, distant arthropod orders. All sequences encompassed at least the complete NS1-encoding open reading frame (ORF). Myriapoda DV 1 and 2 are DVs detected in diplopods for the first time and appear to be closely-related to mosquito-infecting brevidensoviruses (genus *Brevihamaparvovirus*). *Perisesarma bidens* hepadensovirus could be a potential member of the *Hepanhamaparvovirus* genus, the first one, which is not associated with penaeid shrimps, but with the red-clawed crab (*Perisesarma bidens*), another decapod species, instead. The partially sequenced genome of *Tetragnatha maxillosa* DV which is derived from the long-jawed orb weaver (*Tetragnatha maxillosa*), suggests it is a novel arachnid DV, probably a divergent member of the *Densovirinae* subfamily.

Biochemical Properties and Purification of Densoviruses

Infectious particles of DVs are stable in a broad pH range from 3 to 9 in various buffers, and at 56°C for 60 min, except for GmDV (rapidly loses viral infectivity when raising temperature over 56°C). Therefore, DVs are still infectious and survive for a long

period in the environment. Purified virus stocks can be stored at -20°C in Tris-Cl buffer pH 7.6 without significant loss of viral infectivity for years.

The virions lack lipids and the buoyant density of DVs in CsCl gradients are in the range $1.4\text{--}1.44\text{ g cm}^{-3}$, reported values can be slightly different depending on techniques or buffers used. The major band at $1.39\text{--}1.42\text{ g cm}^{-3}$ contains mostly infectious particles, although a slightly higher density, minor band at $1.45\text{--}1.47\text{ g cm}^{-3}$ is also infectious. Those banding from 1.31 to 1.37 g cm^{-3} are empty or contain incomplete genomes (defective particles). The sedimentation coefficient of a DV is about 110 S. The virion contains 25% DNA and 75% protein and a low percentage of polyamines, such as spermidine.

Purification of DVs includes 3 steps: extraction of virus from infected tissues, followed by virus concentration of viruses and final purification. Virus-containing tissues are lysed by several freeze-thaw cycles and by homogenization in a low ionic strength pH 6–7 buffer. Reducing agents like phenylthiourea, 20 mM sodium ascorbate and EDTA are added to reduce melanisation. Final steps include differential and isopycnic centrifugation.

Structural Features of Virions

The metastable capsid is an essential part of the infectious virion and has multiple functions. These include the protection and isolation of the viral nucleic acid from its environment, the delivery of the viral genome to its replication site and interaction with the host environment while executing the previous two functions. Like their vertebrate-infecting counterparts, densoviruses, possess non-enveloped capsids of 21–25 nm in diameter, adopting a $T = 1$ icosahedral symmetry. The capsid is comprised of 60 monomers, of which an asymmetric unit interacts with the others at a twofold, threefold and fivefold symmetry axes when incorporated into the viral particle.

The structural protein encoding ORF, also known as *cap* can express up to four different VPs of varying lengths, which all share a common C-terminal region. Some densoviruses may possess a split VP gene, which requires alternative splicing in order to be expressed (detailed later). In the case of penstyldensoviruses only one VP of 37 kDa gets expressed, deeming each of the 60 subunits to be identical. This implies that each monomer can execute the same function throughout the viral life cycle of PstDV. Nevertheless, generally, the smallest VP comprising the common C-terminal region, is expressed at a higher level and is therefore considered the major VP. The larger, less abundant VPs are N-terminal extended variants of the major VP, which contain multiple domains and motifs important in the viral life cycle such as the PLA2 domain, a calcium-binding domain, and nuclear localization and export signals. Due to the common C-terminus, the larger VPs are also incorporated into the capsid, albeit at low copy number. In different DVs the size of the major VP varies between 37 and 70 kDa in size.

In contrast to members of *Parvovirinae* for which numerous capsid structures have been determined, only four crystal structures are available for *Densovirinae* viruses along with two low resolution structures that have been determined using cryo-EM (Table 2). Two of the high-resolution structures, for GmDV and AdDV, were determined for DNA packaged (full) infectious virions where only AdDV displayed three ordered pyrimidine bases at the 3-fold symmetry axis luminal surface. This ordering is unexpected given the lack of icosahedral symmetry of the packaged genome. Similarly, to the *Parvovirinae* VP structures, a significant portion of the N-terminal region of the major capsid VP is disordered in DVs. Interestingly, the BmDV1 structure currently represents the only parvovirus VP structure where the last 40C-terminal residues are also disordered.

The parvovirus capsid is comprised of monomers of a single jelly roll core. This highly conserved structural fold is comprised of eight beta strands arranged in two, four-stranded antiparallel beta sheets packed across a hydrophobic interface, with the strands in the sheets traditionally labeled from B to H. When folded, each strand faces opposite its neighboring letter in the alphabetical order, hence a BIDG and CHEF sheet is formed, linked by various numbers of loops, responsible for creating the often surface-exposed variable regions (VRs). For the *Parvoviridae* the BIDG sheet is complemented by an N-terminal fifth strand, designated βA .

In the GmDV VP structure, which is the prototype for the *Densovirinae*, the EF and GH loops are further divided into five and four sub-loops, respectively (Fig. 2). While the GH loop is the longest and forms most of the surface features, its length is significantly shorter compared to the corresponding loop in the *Parvovirinae* members at 97 aa compared to 226 aa in canine parvovirus. As for the *Parvovirinae*, the GH loop is the most variable among ambidensoviruses. At the twofold symmetry axis, all parvoviruses harbor an alpha helix (αA), although densoviruses also contain a second α -helix within the EF loop. PstDV also possesses a third helix in the CD loop and displays significantly shorter EF and GH loops than any of the known densoviral capsid structures (Fig. 2).

Table 2 Densoviral capsid structures resolved to date

Virus	Empty/Full	Structure determination method	Resolution in Å	PDB-ID	Ordered residues included in the structure of total residue number (major capsid protein)
AdDV	Full	X-Ray Crystallography	3.5	4MGU	23–418 of 418 (VP4)
GmDV	Full	X-Ray Crystallography	3.6	1DNV	22–437 of 437 (VP4)
JcDV	Empty	Cryo-EM	8.7	N/A	N/A
AalDV2	Full	Cryo-EM	15.6	N/A	N/A
BmDV1	Empty	X-Ray Crystallography	3.1	3POS	43–454 of 494 (VP3)
PstDV	Empty	X-Ray Crystallography	2.5	3N7X	31–329 of 329 (VP1)

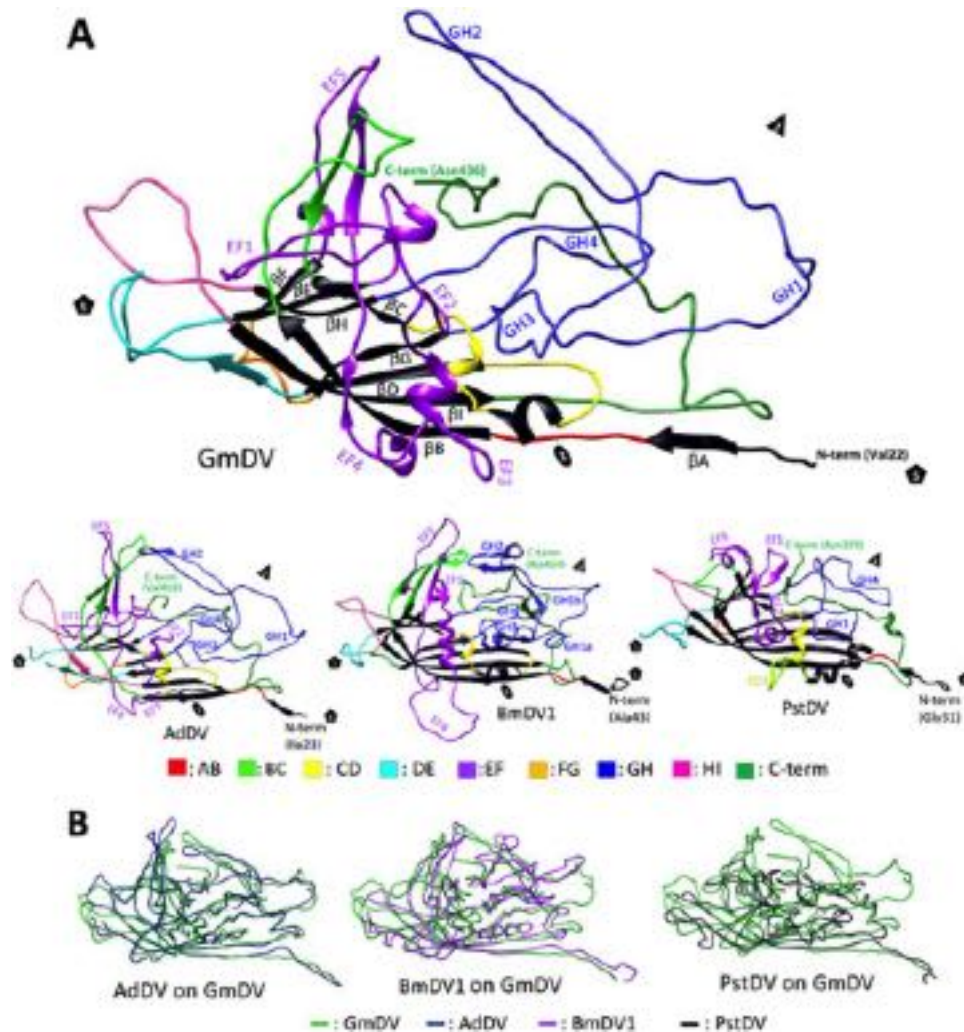


Fig. 2 Densovirus VP structure. (A) Cartoon ribbon diagrams of the ordered common VP structures of *Galleria mellonella* densovirus (GmDV) (top), and *Acheta domestica* densovirus (AdDV), *Bombyx mori* densovirus (BmDV1), and *Penaeus stylirostris* densovirus (PstDV) (bottom). The first ordered N-terminal residue and last C-terminal residue are labeled. The conserved β -core and α A helix are colored in black and labeled in GmDV. Loops and subloops within the major loops are labeled by the major loop name and a number, given according to their order in GmDV. The approximate fivefold symmetry axis is marked by a pentagon, the threefold by a triangle, and the twofold by an ellipsoid. (B) The GmDV VP structure (*Ambidensovirus*, proposed *Protoambidensovirus*) superimposed on the VPs of AdDV (left), BmDV1 (middle), PstDV (right). Diversity on the surface loops are evident.

An important and differentiating feature of the densoviral VP is the domain swapping confirmation observed at their N-terminus (Fig. 3). The densoviral β A strand is a direct extension of the β B strand and it interacts via either hydrogen bonds, metal ions, or hydrophobic interactions with the twofold-related neighboring monomer instead of the β B strand of the very same subunit, as happens in the *Parvovirinae* capsid structures (Fig. 3). Thus, the luminal β BIDG sheet of the jelly roll core is still extended into a β ABIDG sheet as in the *Parvovirinae* and the first observed N-terminal residue is also in a position underneath the 5-fold axis, but in this case, that of the neighboring VP subunit. This swapped confirmation is hypothesized to provide extra stability. The PstDV VP monomer, however, despite displaying the domain-swapped confirmation, is incapable of providing this assessed structural stability.

The overall capsid morphology of DVs can be divided into two types: The larger capsids with diameters of ~ 235 to ~ 260 Å in the depressions and protrusions, respectively, while the smaller capsids measure 215–250 Å, being the smallest capsids so far described for the *Parvoviridae* (Fig. 4). For the former, including GmDV, AdDV, and BmDV1, the capsid surface is smooth with small spike-like protrusions surrounding the 5-fold axes. In GmDV the spikes, formed by the EF4 sub-loop, appear to be smaller compared to BmDV1 and AdDV, due to the protruding GH2 sub-loop filling up the depression surrounding them. In GmDV, a smaller second protrusion is formed by the BC loop. The twofold axes are covered by a depression. In the second group, containing PstDV and AalDV2, there are prominent protrusions surrounding the 5-fold axis, forming two rim-like concentric circles. The 2- and 3-fold symmetry axes have depressions (Fig. 4).

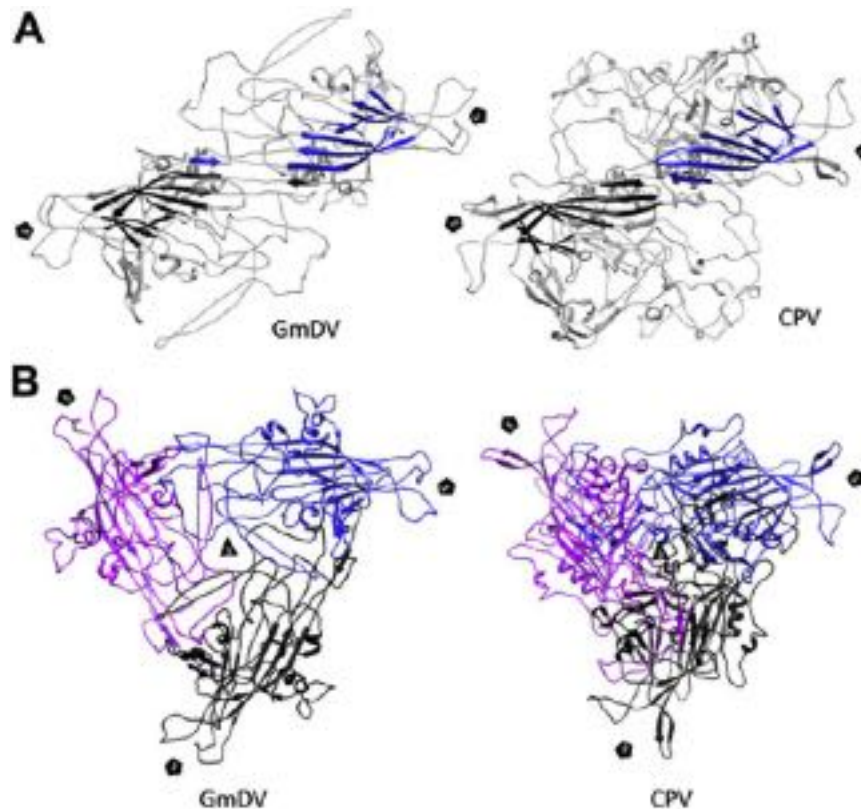


Fig. 3 Multimeric interactions of densoviral and parvoviral VPs. (A) Ribbon diagrams of the interactions between βA and βB at the twofold symmetry axis of *Galleria mellonella* densovirus (GmDV) and canine parvovirus (CPV). The eight-stranded core, with the additional βA , which performs the domain swapping, are colored. (B) Interaction of three threefold-symmetrical VPs for GmDV and CPV are shown. The triangle indicates the threefold axis and the pentagon the fivefold axis. Note the open annulus-like structure at the threefold axis of the densovirus, compared to the more closed arrangement in the vertebrate parvoviruses.

The 5-fold symmetry axis of the DV capsids, like the *Parvovirinae* ones, contains a channel, with a direct opening to the surface. The inner wall of the channel is lined by large hydrophobic residues in all four structures, proposed to provide an interacting surface to a glycine-rich stretch of residues when the N-terminus is externalized during endosomal escape. The densoviral threefold axis, however, is radically different from that of the *Parvovirinae*, occupied by an additional opening besides the fivefold pore instead of the protrusions observed in the *Parvovirinae* (Fig. 4). The opening can be small, formed by the GH loops, which do not even interdigitate between neighboring monomers (PstDV1 of genus *Penstylhamaparvovirus*) or can be occupied by a β -annulus-like structure (subfamily *Densovirinae*). The latter is comparable to the threefold axis of (+) ssRNA viruses, such as Tomato Stunt Mosaic Virus of *Tomoviridae* and Southern Bean Mosaic Virus of *Solemoviridae*. The annulus is formed by charged and flexible residues, with a ~ 10 -Å wide opening in GmDV. This opening is the least pronounced in BmDV1.

Biophysical Features and Functions Associated With the Densovirus Capsid

Compared to members of the *Parvovirinae*, little is known about the functions of densovirus VPs and the available information is based on studies of mostly lepidopteran ambidensovirus. By comparing the VP4 major capsid proteins of two closely related lepidopteran ambidensovirus, GmDV and JcDV, eight variable, exposed regions were identified, located in the vicinity of the fivefold and threefold symmetry axes. Mutating the GH-loop-associated residues in JcDV to their counterpart in GmDV resulted in a decrease of the ability to cross the host midgut epithelium and a reduction of JcDV virulence if introduced through the natural, gastrointestinal pathway. When infecting *Spodoptera frugiperda* hosts *ex vivo*, the mutated virus became mis-targeted and accumulated in subcellular compartments of midgut epithelial cells instead of reaching their target receptors in the basal tight junctions.

Recently, the *Helicoverpa armigera* densovirus 2 (HaDV2) VPs were shown to enhance the structural promoter activity by 35-fold compared to the activation by NS. This suggests that the densoviral capsid proteins may have a non-structural function besides the expected structural ones.

The biophysical properties to date have been investigated only in case of AdDV. Heating of infectious AdDV particles to 70°C resulted in increased PLA2 activity accompanied by genome ejection while capsids remained intact when checked by negative

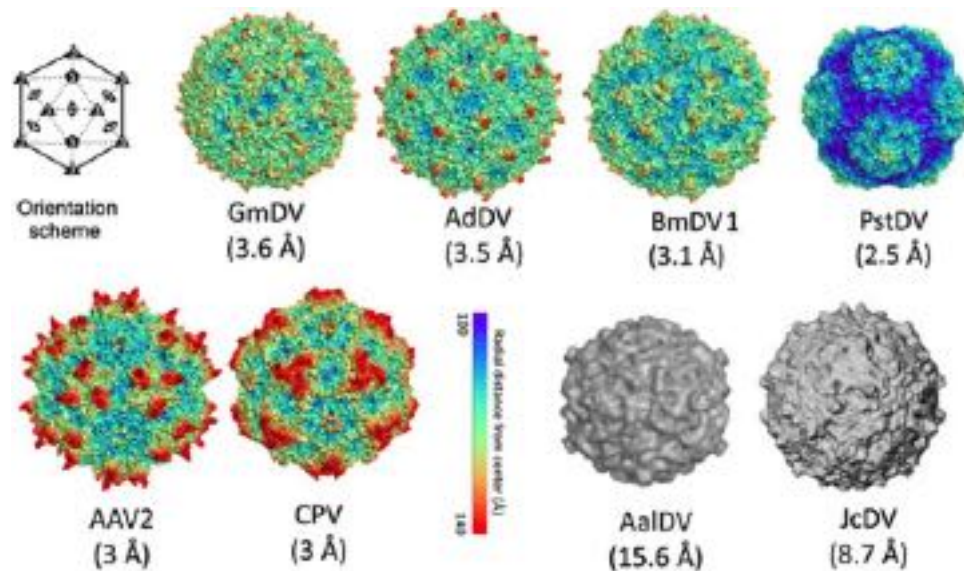


Fig. 4 Densoviral capsid structures. The capsid surface images show the high resolution structures of *Galleria mellonella* densovirus (GmDV), *Acheta domestica* densovirus (AdDV), *Bombyx mori* densovirus BmDV1 and *Penaeus stylirostris* densovirus PstDV, in comparison with the structure of two vertebrate-infecting parvoviruses, namely adeno-associated virus 2 (AAV2) (Dependoparvovirus) and canine parvovirus (CPV) (Protoparvovirus). The resolution of the structures is provided in parenthesis. The scale bar shows the color code of the radial distance from the capsid center. An icosahedral symmetry diagram indicating the positions of the visible symmetry axes in the capsid images is also shown. The cryo electron microscopy-obtained structures (in gray) of lower resolution are shown for *Aedes albopictus* densovirus 2 (AalDV) and *Junonia coenia* densovirus (JcDV) at the bottom right hand side.

stained EM. Structural studies of these emptied particles, however, did not display increased electron density in the fivefold channels, associated with the VP N-terminal externalization required to expose the PLA2.

Our discovery of PLA2 in capsids of most vertebrate and invertebrate parvoviruses shed light on how the densovirus enters the cell and later escapes the endosome or lysosome. During clathrin-mediated endocytosis the PLA2-containing region of VP1u is externalized through the channel at the 5-fold axis which then enables the virion to breach the endosomal membrane. Almost all parvoviruses contain a calcium-binding loop (GPGN) and the active site (DxxAxxHDxxY) required for PLA2 activity in the N-terminus of the largest minor capsid protein. Brevidensoviruses, penstylidenosoviruses and hepandenosoviruses (subfamily *Hamaparvovirinae*) lack this enzymatic activity and enter probably by direct membrane fusion/penetration reactions at the plasma membrane and might have evolved other, yet unexplored, mechanisms to breach the endosomal membrane.

Densovirus Genome Structure and Replication

As a general rule, capsids of monosense DVs package primarily the “–” DNA strand (complementary to mRNA) while the ambisense DVs package both the “+” and the “–” strands in separate particles. There is a great variety in the organization of genome structures, such as size of DV genomes and genome termini, imperfect palindromic sequences at each extremity able to fold into duplex telomeric structures designated hairpins.

The structural similarities of DV and vertebrate parvovirus genomes suggest common strategies in the replication of their DNA by a single-strand displacement by a DNA polymerase and encapsidation by helicase hexamers into the channel at the 5-fold axis of the preformed capsids. The self-priming hairpin at the 3′ end of the genome serves as primers for complementary strand synthesis according to the rolling-hairpin replication model to generate concatemeric intermediate chain resolved in monomers by the nicking activity of Rep. Based on their viral replication model, the 5′ and 3′ terminal hairpins of all DVs with inverted terminal repeats, including terminal hairpins (ITRs), can exist in either of two orientations termed “flip” and its reverse-complement “flop” orientation. The significant differences in size and structure of palindromic 5′- and 3′-terminal sequences within the different genera of DVs might reflect their dependence on specific cellular factors necessary for the replication of their genome or promote their encapsidation. Ambisense DVs have ITRs whereas monosense genomes can have unique terminal hairpins (e.g., brevidensoviruses) or ITRs (e.g., iteradenosoviruses). An N-terminal HuHuuu amino acid motif (u-bulky hydrophobic residue), with one or two other motifs, was identified in the Rep protein (HUH endonuclease superfamily) that is conserved in two vast classes of proteins, one of which is involved in initiation and termination of rolling-circle DNA replication, or RCR (Rep proteins), and the other in mobilization (conjugal transfer) of plasmid DNA (Mob proteins). One of the two additional conserved motifs is located upstream and the other downstream with a conserved Tyr residue. The major role of these HUH endonucleases is processing a range of mobile genetic elements by catalyzing cleavage and rejoining of single-stranded DNA using an active-site Tyr residue to

make a transient 5'-phosphotyrosine bond with the DNA substrate, such as in rolling-circle replication, in various types of transposition and in intron homing. HUH enzyme activities require a divalent metal ion, coordination which is provided by the HUH motif, to facilitate cleavage by locating and polarizing the scissile phosphodiester bond.

During DV replication the 3' hairpin' provides a free 3'-OH group from which leading-strand synthesis can be initiated using host cell enzymes. The replication of this 3' hairpin is thus incomplete and requires a site-specific nick to generate a second 3'-OH at a specific site (the terminal resolution site, *trs*) in the hairpin so that the complementary strand can be extended to completion. Both terminal hairpins on the two strands can then thus be refolded while the hairpins at the two ends are separated by a duplex.

Rep proteins also contain a C-terminal superfamily 3 (SF3) hexameric 3'-5' helicase domain. RCR uses this 3'-5' helicase activity acting on the template strand to facilitate DNA unwinding at the replication fork. The crystal structure of an SF3 DNA helicase, Rep40, from adeno-associated virus 2 (AAV2) delineates the expected Walker A and B motifs, but also reveals an unexpected "arginine finger" that directly implies the requirement of Rep40 oligomerization for ATP hydrolysis and helicase activity. The peptide linker between the HUH endonuclease and the SF3 helicase domains has a critical role in helicase oligomerization.

Expression Strategies

The expression strategies among the DVs differ considerably (Figs. 5 and 6). The promoters of the *rep* and *cap* cassettes of the ambisense DV genomes tend, at least partially, to be in the ITRs. In addition to unspliced transcripts, splicing and leaky scanning, and even overlapping promoters/ORFs (brevidensovirus), are used to produce a large array of proteins from a compact genome. For ambisense DVs, it is striking that an overlapping 3' untranslated region (UTR) exists, which is shared by both the NS and the VP transcripts.

In GmDV two types of transcription strategies have been described. The transcription and translation map of GmDV (Fig. 5(a)) is a model for closely related densovirus (JcDV, MLDV, PiDV, DsDV, HaDV1, etc) from butterflies. Analogous to vertebrate parvoviruses, the NS gene cassette is to the left and reading from left to right. The VP-encoding cassette, however, is positioned on the complementary strand and is read from right to left. Both expression cassettes are driven by promoters positioned in the ITRs. The relative amounts of the various NS products depend on (1) the strength of the splicing signal and (2) the relative strength of the NS1 and NS2 initiation codons. The relative amounts of VP products depend on the leakiness of the scanning ribosomes and does not involve splicing.

CpDV on the other hand, and though sharing the leaky scanning VP cassette with the lepidopteran ambidensovirus, has evolved a remarkably complex NS expression cassette. The NS-1, NS-2, and NS-3 in the 5' half of one strand are organized into five open reading frames (ORFs) due to the split of both NS-1 and NS-2 into two ORFs. The ORF encoding capsid proteins are positioned in the 5' half of the complementary strand. The expression of NS proteins is controlled by two promoters, P7 and P17, driving the transcription of a 2.4-kb mRNA encoding NS-3 and of a 1.8-kb mRNA encoding NS-1 and NS-2, respectively. The two NS mRNA species have a small 53-nt intron in the middle of the sequence.

The NS expression strategy of AdDV (Fig. 5(b)) is rather like that of GmDV. However, different combinations of splicing donor and acceptor sites result in different N-terminal sequences unlike the truncated sequences as for other parvoviruses. The intron in the NS mRNA occurs in about half of the NS transcripts. Ia, Ib, and II are introns that occur in alternative VP transcripts. NS transcripts start at nt 192 and yield NS3. However, a fraction of these transcripts is spliced just upstream of this start codon (intron), leading to translation of NS1 from 856-AUG, with a poor initiation environment, and, through leaky scanning, of NS2 from 875-AUG. Like the case for all densovirus, the 3' ends of AdDV NS and VP transcripts overlap in the middle of the genome at a 3' UTR sequence, a shared trait by all ambisense DVs to date. VP transcripts are initiated at nt 5235, and VP1 initiation is at nt 5230. The short 5'-UTR predicts an inefficient initiation (leaky scanning) and could be responsible for production of a nested set of N-terminally extended viral proteins. However, removal of either of the two alternative introns in ORF-B (Ia or Ib) did not connect the exons in ORF-B and ORF-A in frame, so only nonstructural proteins could be produced from nt 5230 and VP2 could be produced directly from the first AUG in ORF-A when this splicing occurred. An alternative intron II, which is mutually exclusive with introns Ia and Ib since it removes the ORF-B splice acceptor, connects ORF-B and ORF-A in frame so that VP1 can be produced from nt 5230 (Fig. 5(b)).

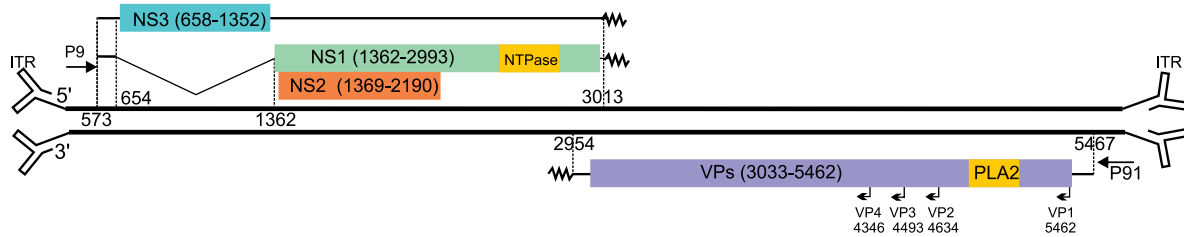
The NS expression strategy of the ambisense pefuambidensovirus, PfDV, is the same as for GmDV. Structural analysis of cDNAs complementary to mRNAs from the region coding for structural proteins suggested alternative splicing and polyadenylation as a means for generation of the structural proteins of PfDV.

The NS cassette of the ambisense blattambidensovirus, BgDV1, has an extra NS product (through an extra splice variant, identical to last 273 codons of NS1 (Fig. 5(c))). Alternative splicing in the VP transcript yields a VP2 product with a unique N terminus as for AdDV (Fig. 5(b)).

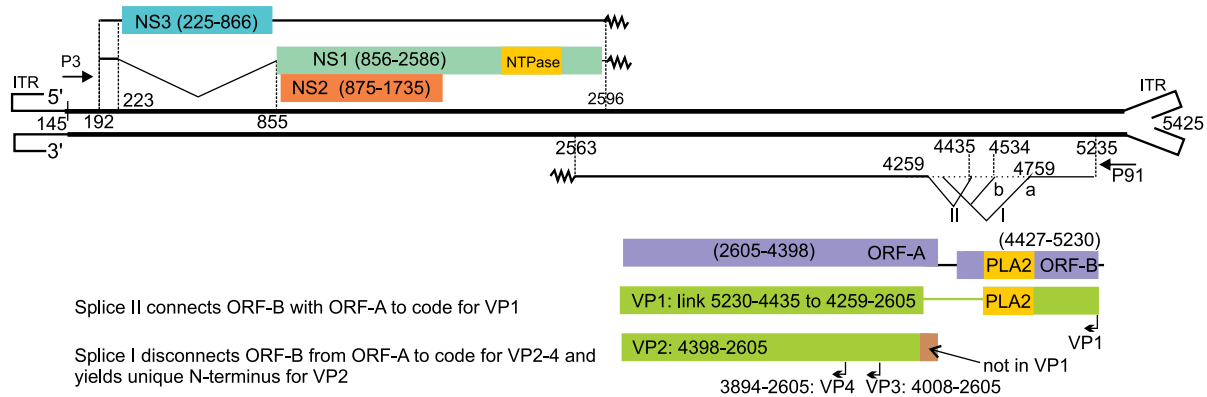
Although their transcription strategy has not been resolved yet, in some cases, uninterrupted potential ORFs corresponding to non-structural viral proteins or capsid proteins of *Myzus persicae* densovirus (MpDV) and *Dysaphis plantaginea* densovirus (DpIDV), both ambisense hemiambidensovirus, were found within EVEs identified in the aphid genome. Little is known about the expression strategy of the *Aquambidensovirus* genus members. The predicted expression of the SSaDV contig in sea stars is represented in Fig. 5(d).

The phenomenon of possessing a VP expression cassette comprised of two ORFs, united by alternative splicing, appears at four separate DV lineages (e.g., Figs. 5(b) and (c)). The variation in donor- and acceptor site sequence and position as well as in the number of spliced transcripts suggests that this strategy has evolved independently at multiple times during ambisense DV evolution.

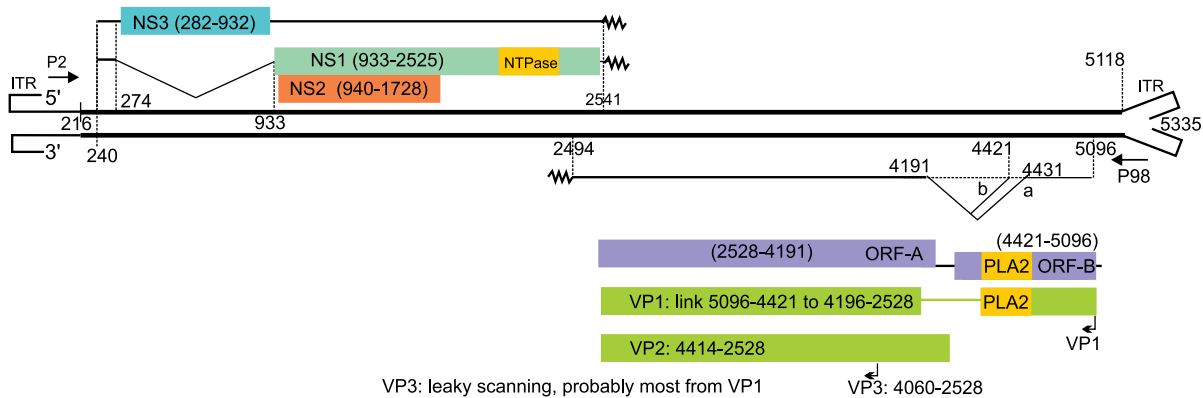
a. Protoambidensovirus (GmDV; GenBank accession no. L32896)



b. Scindoambidensovirus (AdDV; GenBank accession no. HQ827781)



c. Blattambidensovirus (BgDV; GenBank accession no. AY189948)



d. Aquambidensovirus (SSaDV; GenBank accession no. PRJNA253121)

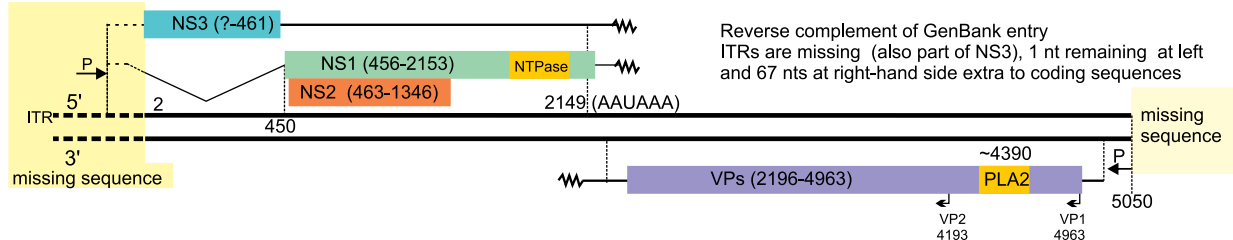


Fig. 5 Transcription strategies of ambisense densoviruses. Panels (a, b, c, d) display representatives of four ambisense densoviral genera, according to the new taxonomy classification of subfamily *Densovirinae*. The thick line in the middle represents the coding genome, while genome termini are drawn in thinner lines. The fine lines indicate the starts, ends and intron positions on the mRNAs and promoters at the 5' ends are indicated by small arrows according to the direction of the transcription they drive. Genes and proteins are shown as colored boxes. Thick, bent arrows indicate the leaky scanning translation sites.

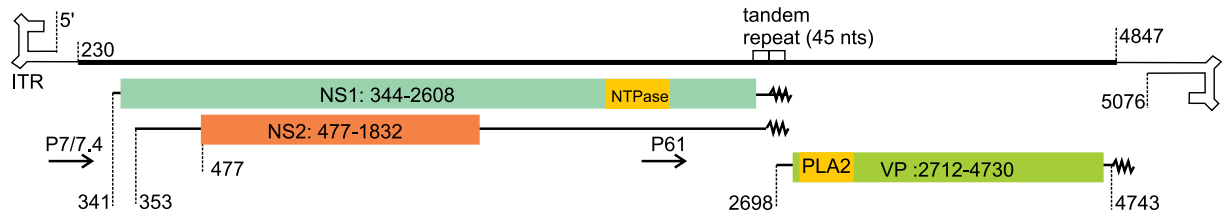
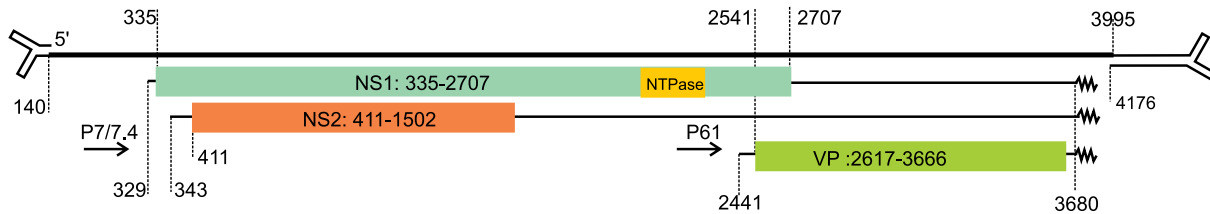
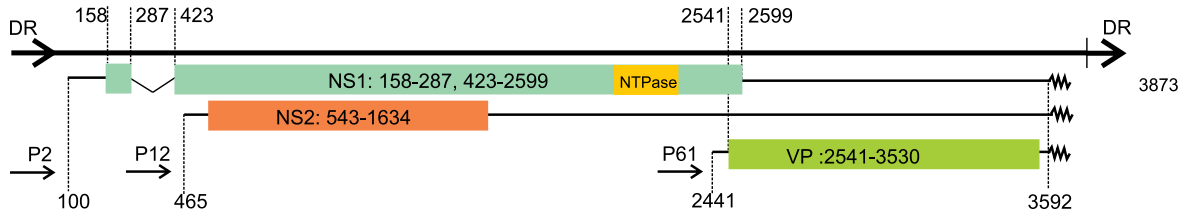
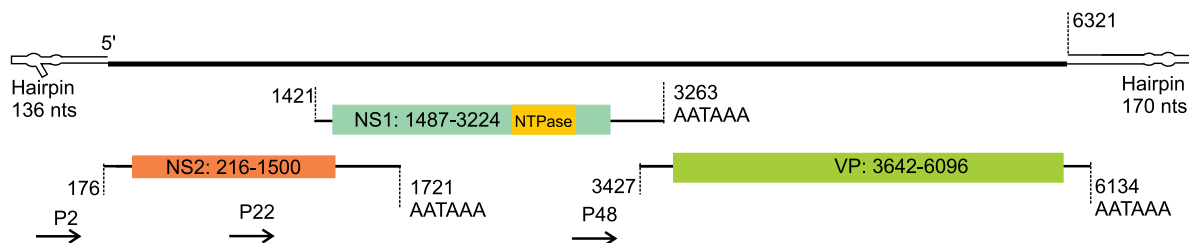
a. *Bombyx mori* densovirus 1 (BmDV1; GenBank accession AY033435.1)b. *Aedes albopictus* densovirus 1 (AalDV; GenBank accession no. AY095351)c. *Penaeus stylirostris* penstyldenovirus 1 (PstDV; GenBank accession no. AF273215)d. *Penaeus monodon* hepadenovirus 1 (PmoHDV1; GenBank accession no. NC_007218.1)

Fig. 6 Transcription strategies of monosense densoviruses. Panels (a, b, c, d) display representatives of four monosense genera, according to the proposed new taxonomy of family *Parvoviridae*, classifying *Iteradensovirus* to the original subfamily *Densovirinae*, whereas the remaining three genera are all members of *Hamaparvovirinae*. The thick line in the middle represents the coding genome, while genome termini are drawn in thinner lines. The fine lines indicate the starts, ends and intron positions on the mRNAs and promoters are indicated by small arrows according to the direction of the transcription they drive. Genes and proteins are shown as colored boxes. Thick, bent arrows indicate the leaky scanning translation sites.

As for monosense DVs, the transcription maps and expression of the iteradensoviruses do not show great variety (Fig. 6(a)). The two nonstructural (NS) genes are expressed by overlapping promoters with alternate transcription starts at either side of the NS1 start codon. Leaky scanning yields 4 VP products of which VP1 has PLA2 activity. So far, BmDV1 is the only iteradensovirus with a tandem repeat of 45 nts in the intergenic region.

As for the iteraviruses, brevidensoviruses do not need mRNA splicing since they possess two overlapping P7/7.4 NS promoters (monocytic) located closely to the alternate transcription initiation sites, positioned at either side of the NS1 initiation codon (Fig. 6(b)). Their genomes harbor only one polyadenylation signal, at which both the NS and VP transcript populations co-terminate. VPs are expressed by leaky scanning from only one transcript.

In the penstylidenoviruses (**Fig. 6(c)**) NS1 and NS2 have separate promoters (P2 and P12; Pham, 2014) for NS1 and NS2 where the location of P12 overlaps the intron of NS1, possibly through sequence and motif constraints. The absence of hairpins in the telomeres (Direct Repeats, DR) questions whether PstDV is an authentic parvovirus.

Little is known about the transcription strategy of *Penaeus monodon* hepadensovirus. It has been reported to have a unique expression strategy (**Fig. 6(d)**). Whereas in other densoviruses the initiation codon of NS2 is downstream of that of NS1, the *ns2* gene is located upstream of NS1. The ORF of VP is exceptionally large compared to the protein observed in SDS-PAGE (92 kD vs 57 kD).

Conclusions

Advances in viral metagenomic approaches in recent years have revolutionized the discovery of novel viruses (EVEs and exogenous viruses) from a high diversity of invertebrate hosts. This resulted in suggestions for a significant overhaul of the classification and taxonomy of ambisense densoviruses and the addition of an ancient lineage of highly divergent parvoviruses (*Hamaparvovirinae* subfamily) that infect an exceptionally broad range of vertebrate and invertebrate hosts. For instance, chapparvoviruses found in fish are more closely related to those from invertebrates than they are to those of amniote vertebrates, suggesting parvoviruses to have evolved to infect vertebrates on at least two independent occasions. Currently, host species susceptible to DVs and/or possessing EVEs of suspected DV origin encompass the basal animal phylum Cnidaria as well as both proto- and deuterostome animals. The future directions of DV research must assess and cope with this enormous diversity; a product of parvovirus evolution may flank a 500-million-year-long interval.

Acknowledgments

The authors wish to acknowledge the support received from the Natural Sciences and Engineering Research Council of Canada.

Reference

Cotmore, S.F., Agbandje-McKenna, M., Canuti, M., *et al.*, 2019. ICTV virus taxonomy profile: Parvoviridae. *Journal of General Virology* 100, 367–368.

Further Reading

- Cotmore, S.F., Tattersall, P., 2006. Structure and organization of the viral genome. In: Kerr, J.R., Cotmore, S.F., Bloom, M.E., Linden, R.M., Parrish, C.R. (Eds.), *Parvoviruses*. London, UK: Hodder Arnold, pp. 73–94.
- Feschotte, C., Gilbert, C., 2012. Endogenous viruses: Insights into viral evolution and impact on host biology. *Nature Reviews Genetics* 13, 283–296.
- Flegel, T.W., 2006. Shrimp parvoviruses. In: Kerr, J.R., Cotmore, S.F., Bloom, M.E., Linden, R.M., Parrish, C.R. (Eds.), *Parvoviruses*. London, UK: Hodder Arnold, pp. 487–493.
- Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Research* 20, 3279–3285.
- Krupovic, M., 2013. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Current Opinion in Virology* 3, 578–586.
- Mietzsch, M., Péntzes, J.J., Agbandje-McKenna, M., 2019. Twenty-five years of structural parvirology. *Viruses* 11. doi:10.3390/v11040362.
- Péntzes, J.J., de Souza, W.M., Agbandje-McKenna, M., Gifford, R.J., 2019. An ancient lineage of highly givergent parvoviruses infects both vertebrate and invertebrate hosts. *Viruses* 11. doi:10.3390/v11060525.
- Rosario, K., Duffy, S., Breitbart, M., 2012. A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics. *Archives of Virology* 157, 1851–1871.
- Tijssen, P., Bando, Y., Li, Y., *et al.*, 2006. Evolution of densoviruses. In: Kerr, J.R., Cotmore, S.F., Bloom, M.E., Linden, R.M., Parrish, C.R. (Eds.), *Parvoviruses*. London, UK: Hodder Arnold, pp. 55–68.
- Tijssen, P., Péntzes, J.J., Yu, Q., Pham, H.T., Bergoin, M., 2016. Diversity of small, single-stranded DNA viruses of invertebrates and their chaotic evolutionary past. *Journal of Invertebrate Pathology* 140, 83–96.
- Zádori, Z., Szelei, J., Lacoste, M.C., *et al.*, 2001. A viral phospholipase A2 is required for parvovirus infectivity. *Developmental Cell* 1, 291–302.

Relevant Websites

<https://en.wikipedia.org/wiki/Ambidensovirus>

Ambidensovirus.
Wikipedia.

https://talk.ictvonline.org/ictv-reports/ictv_online_report/ssdna-viruses/w/parvoviridae/1047/subfamily-densovirinae

Subfamily: Densovirinae.
Parvoviridae.
ssDNA Viruses.
ICTV.

Polydnaviruses (*Polydnaviridae*)

Anne-Nathalie Volkoff, Diversity, Genomes and Insects-Microorganisms Interactions, National Institute of Agricultural Research, University of Montpellier, Montpellier, France

Elisabeth Huguet, Research Institute on Insect Biology, French National Center for Scientific Research, University of Tours, Tours, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AsIV *Apophua simplicipes* Ichnovirus

BV Bracovirus

BVSPER Bracovirus Structural Proteins Encoding Regions

CcBV *Cotesia congregata* Bracovirus

CiBV *Chelonus inanitus* Bracovirus

CpBV *Cotesia plutellae* Bracovirus

CrV *Cotesia rubecula* Virus (bracovirus)

CsIV *Campoletis sonorensis* Ichnovirus

CvBV *Cotesia vestalis* Bracovirus

DRJ Direct Repeat Junction

Egf Epidermal growth factor

GfBV *Glyptapanteles flavicoxis* Bracovirus

GfIV *Glypta fumiferanae* Ichnovirus

HdIV *Hyposoter didymator* Ichnovirus

HfIV *Hyposoter fugitivus* Ichnovirus

HIM Host Integration Motif

ICTV International Committee on Taxonomy of Viruses

IV Ichnovirus

IVSPER Ichnovirus Structural Proteins Encoding Regions

MdBV *Microplitis demolitor* Bracovirus

PDV Polydnavirus

PTP Protein Tyrosine Phosphatase

RU Replication Unit

TnBV *Toxoneuron nigriceps* Bracovirus

TrIV *Tranosema rostrale* Ichnovirus

Glossary

Encapsulation Encapsulation is a cellular immune response used against pathogens that are too large to be phagocytosed. This response is employed by insect larvae in response to parasitism by wasp parasitoid eggs for example. Encapsulation involves the aggregation of insect immune cells (hemocytes) around the invading body.

Hemocyte The insect blood cells.

Koinobiont parasitoid A parasitoid whose host continues to develop after parasitization.

Melanization In insect immunity, melanization is an immune effector mechanism involved in the killing of pathogens and the eggs of parasitoid wasps. Melanization involves a series of reactions such as the conversion of tyrosine to melanin precursors and the cross-linking of proteins to form a layer of melanin that surrounds and

sequesters an invading body. The death of the parasitoid egg presumably occurs either by oxidative damage or by starvation.

Oviposition Act of laying an egg or eggs.

Parasitoid wasp A wasp the adult stage of which is free-living and whose immature stages feed and develop within or on the bodies of other organisms (usually an insect) that will eventually be killed.

Rep/Helicase A single-stranded DNA-dependent ATPase involved in DNA replication.

Virulence genes Genes responsible for the biological alterations observed in the organism where they are expressed. This term is borrowed from bacterial pathogens that use “virulence factors” to establish on or within a host and enhance their potential to cause disease.

Classification (Compact)

Polydnaviruses (PDVs) are enveloped large DNA viruses with a segmented double-stranded DNA (dsDNA) genome. All PDVs are mutualistically associated with parasitic wasps (also named parasitoids) of lepidopteran larvae.

The Polydnaviridae family was proposed by Don Stoltz in 1984 and recognized by the ICTV in 1991. PDV are divided into two genera, *Bracovirus*, for PDVs associated with wasp subfamilies of the family *Braconidae*, and the *Ichnovirus* for those associated with wasp subfamilies of the family *Ichneumonidae*. BVs are found in wasps from the “microgastroid complex”, a monophyletic assemblage of ~ 50,000 species belonging to six braconid subfamilies. IVs are presently described in two ichneumonid subfamilies, the Campopleginae and the Banchinae, and they are estimated to be associated with at least 14,000 species of wasps.

Life Cycle of PDVs

Among dsDNA viruses, PDVs are the only ones known to be mutualistically associated with multicellular organisms, parasitoid wasps. These insects are characterized by a free-living adult stage and parasitic immature stages. Their host, usually an insect, dies at the end of parasitoid development. Parasitoid wasps associated with PDVs are all koinobiont endoparasites of lepidopteran larvae, i.e. female wasps lay eggs in the body of a host caterpillar, which continues to develop as the parasitoid's offspring matures.

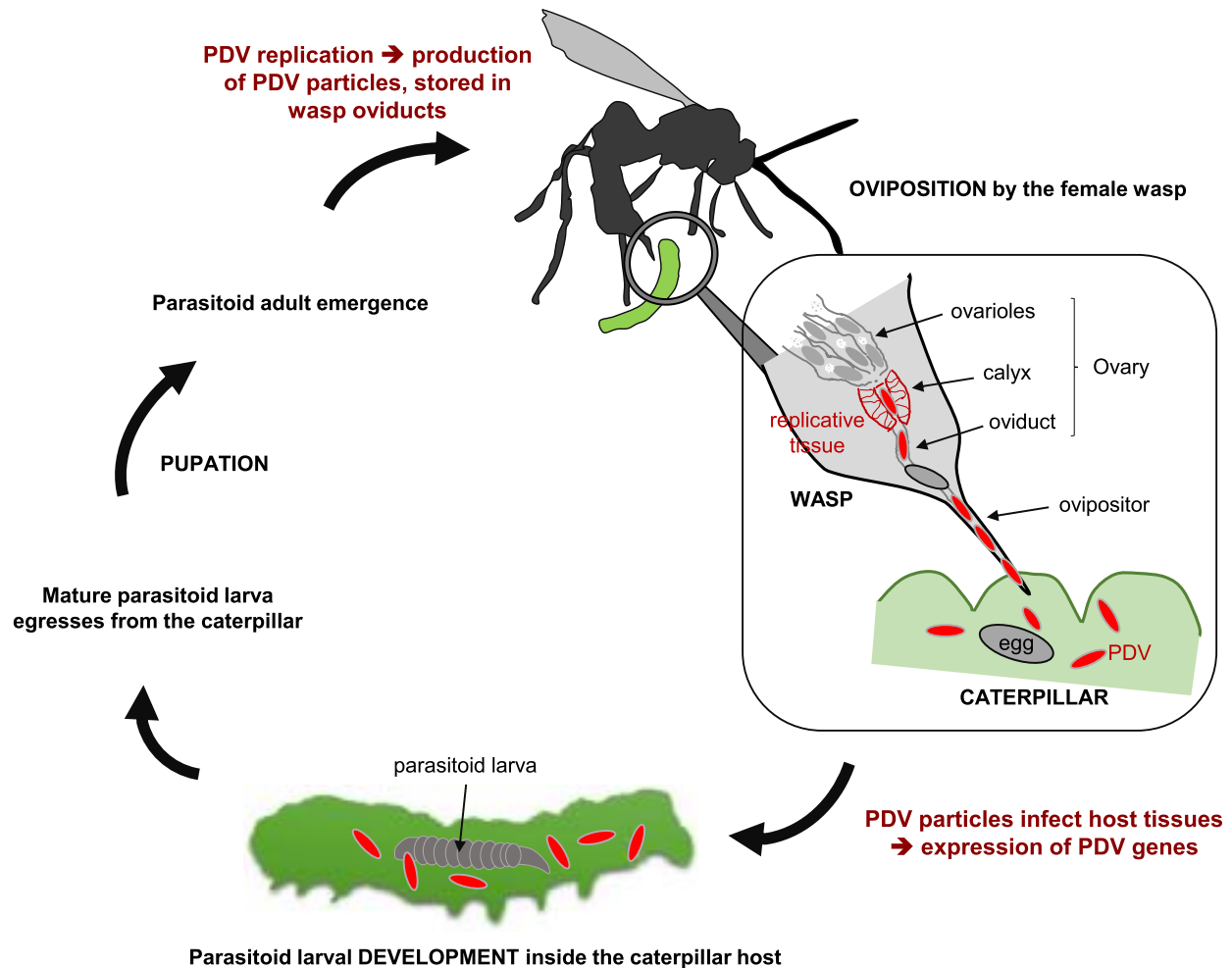


Fig. 1 Life cycle of a parasitoid wasp associated with a polydnavirus. Virus particles are produced in the wasp female ovaries during the pupal and adult stages. Production occurs exclusively in a specialized tissue, the calyx. The produced virus particles are released in the oviducts, and then injected in the parasitoid insect host, a caterpillar, when the female wasp lay its egg(s). Once in the caterpillar, polydnavirus particles infect all the tissues of the insect, viral DNAs reach the nuclei and the genes encoded by the packaged polydnavirus genome are expressed. Successful development of the wasp progeny depends on the expression of these viral genes, which lead to physiological alterations in the caterpillar that are beneficial for the wasp (inhibition of the caterpillar host immune response, modulation of its development, etc.). Once the parasitoid larva (or larvae) ends its development, the mature parasitoid larva egresses from the caterpillar, which dies. The parasitoid then spins a cocoon within which pupation takes place, leading to emergence of a new adult wasp.

BVs and IVs share a common life cycle. They both persist in all cells of the wasp as integrated proviruses. Consequently, PDVs are transmitted vertically through the germline of wasps. During the pupal and adult stage, PDVs replicate in the nuclei of specialized ovarian cells, in a region named calyx (**Fig. 1**). Many virions are produced by calyx cells. BVs are released by lysis of the replicative cells whereas IVs are released by budding. Virions are then stored in the lumen of the oviducts to be injected in the parasitoid host upon oviposition by the female wasp.

PDVs rapidly infect host cells, but no viral replication occurs in the parasitized insect. Infection of the parasitoid host tissues is followed by expression of viral genes on which the survival of wasp offspring depends. Indeed expression of PDV genes disables host immune responses and alters host development and metabolism.

Virion Structure

The virions of BVs and IVs are morphologically and structurally distinct (**Fig. 2**).

BV virions resemble some non-occluded baculoviruses. They have cylindrical, often tailed nucleocapsids surrounded by a single envelope, formed in the nucleus of calyx cells. BV virions contain one or several nucleocapsids. BV nucleocapsids each contain a single DNA molecule. The capsids can vary in length from 30 to 80 nm (excluding tails), with a width around 50 nm.

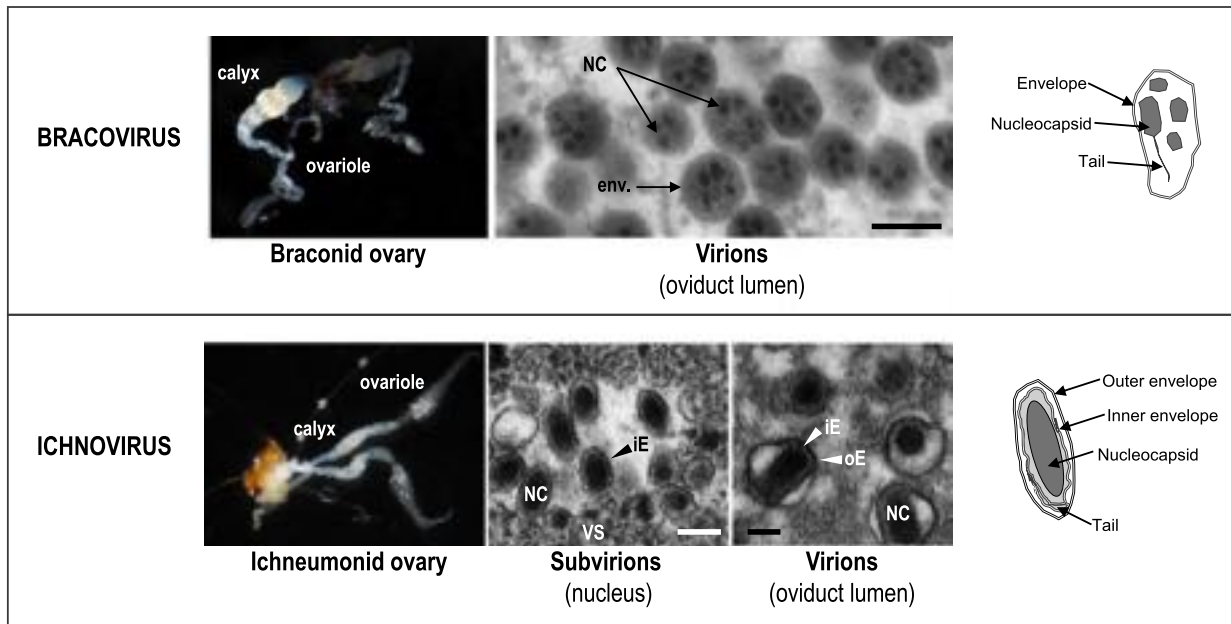


Fig. 2 Morphology of polydnavirus particles: differences between *Bracoviruses* (upper panel) and *Ichnoviruses* (lower panel). Both bracoviruses and ichnoviruses are produced in a specialized region of the female wasp ovaries, named the calyx; on the left, views of dissected ovaries of parasitoid wasps indicating localization of the calyx. In the center, the upper panel shows bracovirus virions in the oviduct lumen of a braconid female wasp observed by transmission electron microscopy (TEM); the lower panel shows on the left ichnovirus subvirions assembled in the nucleus of a calyx cell, and on the right mature virions present in the oviduct lumen of an ichneumonid female wasp. NC, nucleocapsid; env., envelope; VS, virogenic stroma; iE, inner envelope; oE, outer envelope. Scale bar: 200 nm. Pictures by M. Frayssinet (ovaries) and M. Ravallec (TEM), DGIMI, INRA. On the right of the figure, schematic representations of bracovirus and ichnovirus virions. For bracoviruses, a virion containing multiple nucleocapsids is represented.

IVs have lenticular nucleocapsids surrounded by two-unit membranes. The inner envelope is formed in the nucleus of calyx cells and often presents a “hook-like” structure (or tail) of variable length. The outer envelope originates from the plasma membrane and is acquired during virion budding. Nucleocapsids are uniform in size, they measure more than 300 nm in length and approximately 80–90 nm in diameter. Virions from campoplegine wasps contain a single nucleocapsid, whereas virions of banchine wasps may have multiple nucleocapsids. Note that the virions described in the campoplegine *Bathyplectes* spp. enclose multiple punctiform nucleocapsids. There is no data presently on the number of circular DNA molecules contained in IV nucleocapsids.

Morphogenesis of PDV Particles

PDVs replicate in the nuclei of calyx cells. Both BV and IV nucleocapsids are assembled in virogenic stroma and acquire an envelope in the nucleus (Fig. 3). The origin of the nuclear envelope, and whether it has the same origin in BVs and IVs, remain unknown. BV morphogenesis has been extensively followed for *Chelonus inanitus* BV. During the first steps of replication, virions are assembled and progressively fill the nucleus; “mature” calyx cells, located basally near the oviducts, undergo successive lysis of the nuclear envelope, then of the plasma membrane, leading to the release of the mature virions in the oviduct lumen. On the other hand, the production of IV particles relies on another process. Enveloped subvirions formed in the nucleus exit the later by budding through the nuclear envelope. They finally access the oviduct lumen by budding through the apical plasma membrane.

The PDV Packaged Genome

The genome packaged into PDV virions consist in several multiple circular, double-stranded DNAs that are non-equimolar in abundance. The number and size of DNA segments vary depending on the species (Table 1). BV genomes usually have from 15 to 30 DNA segments ranging in size from 2 to 50 kb, with estimated genome sizes from 125 kb to almost 600 kb. IV genomes consist in 20 to 50 DNA segments, ranging in size from 2 to 28 kb, and estimated genome sizes around 250 kb. The genomes of IV associated with banchine wasps have slightly larger genomes and display a higher degree of genome segmentation compared to campoplegine IVs.

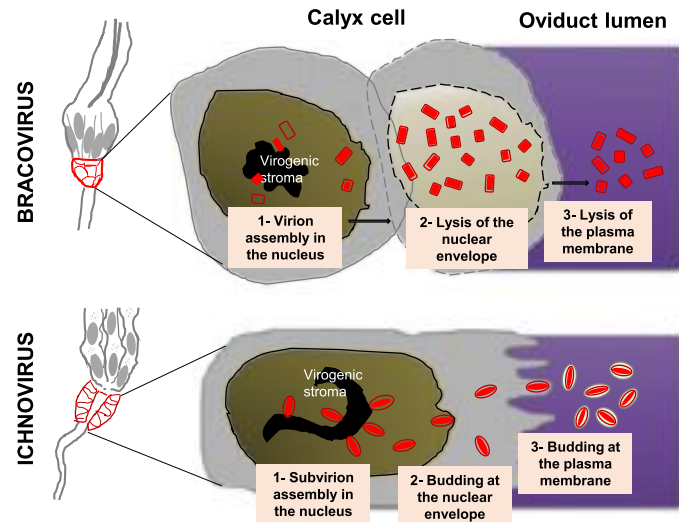


Fig. 3 Schematic representation of the different steps of Bracovirus and Ichnovirus morphogenesis in the wasp calyx cells. On the left, a schematic representation of the wasp ovary indicates the position of the calyx (in red) in braconids and ichneumonids. Bracovirus (BV) and ichnovirus (IV) virions are assembled in the nucleus of the calyx cells, in specific regions named virogenic stroma. For BVs, along with nucleocapsid production, two-unit membrane envelopes appear in the nucleus. Nucleocapsids are enveloped to form mature virions that accumulate in the nucleus. Release of virions occurs by disintegration first of the nuclear and then of the plasma membrane. For IV, assembly of sub-virions (i.e. nucleocapsids surrounded by a single two-unit membrane envelope) takes place in the vicinity of the virogenic stroma. Sub-virions then migrate towards the nuclear envelope where they bud to the calyx cell cytoplasm. Sub-virions then reach the plasma membrane and bud through it to finally reach the calyx lumen where they accumulate as mature virions surrounded by two envelopes. In contrast to what occurs in BV-producing cells, IV-producing cells do not undergo cell lysis; each calyx cell can a priori produce virus particles during the entire female adult life.

Table 1 Features of the packaged genome of selected polydnaviruses. Are indicated the name of the parasitoid species carrying the polydnavirus, the number and length of the segments present in the packaged genome; the total genome size, and the number of predicted genes encoded by the genome

Parasitoid species	Polydnavirus	Segment number and size	Genome size	Number of predicted genes	Reference
<i>Braconid wasps</i>					
<i>Microplitis demolitor</i>	MdBV	15 (3.6–34.3 kb)	190 kb	61	Webb <i>et al.</i> (2006)
<i>Cotesia congregata</i>	CcBV	35 (5.0–41.6 kb)	580 kb	222	Bézier <i>et al.</i> (2013); Chevignon <i>et al.</i> (2014)
<i>Cotesia vestalis</i>	CvBV	35 (2.6–39.2 kb)	540 kb	157	Chen <i>et al.</i> (2011)
<i>Toxoneuron nigriceps</i>	TnBV	> 27 (3.9–13.9 kb)	> 203 kb	> 42	Ibrahim (2010); PhD thesis
<i>Glyptapanteles flavicoxis</i>	GfBV	> 29 (3.8–50.7 kb)	> 581 kb	> 193	Desjardins <i>et al.</i> (2008)
<i>Ichneumonid campoplegine wasps</i>					
<i>Campoletis sonorensis</i>	CsIV	24 (6.1–19.6 kb)	247 kb	101	Webb <i>et al.</i> (2006)
<i>Hyposoter didymator</i>	HdIV	50 (2.5–3.6 kb)	250 kb	135	Dorémus <i>et al.</i> (2014)
<i>Hyposoter fugitivus</i>	HfIV	56 (2.6–8.9 kb)	246 kb	150	Tanaka <i>et al.</i> (2007)
<i>Tranosema rostrale</i>	TrIV	> 40 (4.1–10.1 kb)	~ 250 kb	> 89	Tanaka <i>et al.</i> (2007)
<i>Ichneumonid banchine wasps</i>					
<i>Glypta fumiferanae</i>	GfIV	105 (1.5–5.2 kb)	~ 291 kb	101	Lapointe <i>et al.</i> (2007)
<i>Apophua simplicipes</i>	AsIV	> 132 (~ 1.0–4.0 kb)	~ 304 kb	186	Djoumad <i>et al.</i> (2013)

MdBV, *Microplitis demolitor* Bracovirus; CcBV, *Cotesia congregata* Bracovirus; CvBV, *Cotesia vestalis* Bracovirus; TnBV, *Toxoneuron nigriceps* Bracovirus; GfBV, *Glyptapanteles flavicoxis* Bracovirus; CsIV, *Campoletis sonorensis* Ichnovirus; HdIV, *Hyposoter didymator* Ichnovirus; HfIV, *Hyposoter fugitivus* Ichnovirus; TrIV, *Tranosema rostrale* Ichnovirus; GfIV, *Glypta fumiferanae* Ichnovirus; AsIV, *Apophua simplicipes* Ichnovirus.

Sequencing of PDV packaged genomes revealed they generally encode more than 100 predicted genes. The genes encoded by PDV genomes differ in nature in BVs and campoplegine IVs (Table 2); interestingly banchine IVs share gene families with BV (PTPs and viral ankyrins). The most represented gene family in BVs, in terms of number of genes, are the PTPs, with generally more than 30 genes identified in the sequenced genomes. For IVs, the most represented gene family is the repeat-element genes, with also often more than 20 members identified in IV genomes; they have no similarities with known genes, but all contain a 540 nt motif that make them related. Both BV and IV genomes possess genes encoding proteins harboring ankyrin domains (viral ankyrins);

Table 2 Gene families identified in polydnavirus genomes. Are listed the main viral gene families encoded by bracoviruses (BV, on the left) and by ichnoviruses (IV) associated with campoplegine wasps (on the right). Conserved families are those encountered in all sequenced genomes for a polydnavirus genera; for BVs are also indicated gene families found in several species but not all; Specific gene families are examples of genes found in a single polydnavirus species, knowing that only a few polydnavirus packaged genomes have been sequenced so far

Gene families conserved in all sequenced BVs (and in IVs associated with banchine wasps)	Gene families conserved in all sequenced IVs associated with campoplegine wasps
Protein tyrosine phosphatase (PTP)	Repeat-element gene
Viral ankyrin	Cys-motif protein
Gene families conserved in some BVs	Viral ankyrin
Cystatin	Viral innexin
CrV – 1 like protein	Polar-residue rich protein (PRRP)
Lectin-like	N-gene
Specific gene families	Specific gene families
Egf-motif (MdBV)	Glycine-proline rich (HdIV)

MdBV, *Microplitis demolitor*; Bracovirus; HdIV, *Hyposoter didymator* Ichnovirus.

however whether BV and IV viral ankyrins are related or derive from independent acquisitions in the two wasp lineages has not yet been demonstrated.

As a rule, based on analysis of all the PDVs sequenced, no typical virus replication genes have been identified in PDV packaged genomes. Nonetheless a helicase gene was identified in *Cotesia vestalis* BV; however, further work revealed that it was actually a Rep/Helicase from an intact Helitron transposable element inserted into a virus segment.

Function of the Genes Encoded by PDV Packaged Genome

The packaged genome is transferred to the host of the parasitoid upon parasitism, where the genes are for the most part expressed. These genes are thus considered as “virulence” genes, responsible for the physiological alterations observed in the parasitized insect and on which rely successful development of the wasp progeny. A number of these genes, such as PTPs, viral ankyrins, cystatins, and viral innexins have similarities with eukaryotic genes. The hypothesis is that these genes have an insect, probably a wasp, origin rather than a viral origin.

PDVs have early been shown to be necessary for successful development of the wasp. They are indeed involved in immune suppression (e.g., changes in hemocyte composition, cellular spreading, encapsulation and melanisation) and alterations of development (e.g., modification of growth hormone titers) or metabolism (e.g., changes in levels of total carbohydrates, lipids or proteins) in the parasitized insect. PDV physiological effects have been shown by phenotypic approaches but also indirectly by transcriptomic approaches. However, amongst the hundred genes encoded by PDVs, the physiological roles in parasitism of only a few has been investigated so far (Table 3). This is probably due to the absence of similarity to known functional proteins for the majority of PDV encoded proteins, but also, and most probably the main reason, due to the technical difficulty to study these viruses which are not replicative in the insect where genes are expressed and functional.

Apart from physiological effects, recent studies have amazingly also highlighted ecological roles for PDVs. PDV infection affects protein expression in salivary glands of the parasitized insect; and particularly proteins that act as elicitors of the plant response are affected in terms of abundance. This leads to induction of the plant immune response and emission of volatiles that differ from those induced by a non-infected phytophagous insect, and consequently impact trophic interactions between the plant, the phytophagous insect, its parasitoids and even the hyperparasitoids.

Because PDVs are essential for parasitoid survival, it is expected that these mutualistic viruses impact parasitoid host range. Indeed, PDVs were shown in at least one case to be involved in the specialization of parasitoids to their hosts.

Fate of the PDV Packaged Genome in the Parasitized Insect

Larval development of parasitoid larvae within their lepidoperan hosts last several days, a time period during which PDV continue to be active. Expression of PDV genes is detected early after parasitism and during all the parasitoid development in the parasitized insect. Since PDV particles are non replicative (absence of replication genes), this raises the question of how PDV DNA is maintained in the lepidopteran host of the parasitoid. Early studies in cultured cells showed that portions of PDV segments were capable of integration into the genome of lepidopteran cells. More recently, integration of BV segments in genomic DNA of the host was shown in vivo. Integration into host cells occurs rapidly after infection and in association with a domain termed the “host integration motif” (HIM) (Fig. 4), which was identified in most MdBV genomic DNAs and several CcBV circles. This finding suggests that BV genomic segments integrate into host cells through a shared mechanism. However, no specific host genomic target could be identified, despite the fact that integration was not totally random within the lepidopteran host genome. Nonetheless, this process allows PDV to persist and throughout parasitism.

Table 3 Some examples of polydnavirus (PDV) or PDV genes effect on the physiology of the parasitized caterpillar host. PDV infection affects either the host immune response (Immunity) or its development (Development). For each gene studied, the observed effect is indicated

Name of the gene	PDV	General function affected in the parasitized caterpillar		Effect of the PDV gene	Reference
<i>Lectin-like</i>	CpBV	Immunity	Cellular immunity	Inhibition of encapsulation	Lee <i>et al.</i> (2008)
<i>Glc1.8</i>	MdBV	Immunity	Cellular immunity	Inhibition of encapsulation	Beck and Strand (2005)
<i>CpBV-CrV1</i>	CpBV	Immunity	Cellular immunity	Inhibition of encapsulation	Kumar and Kim (2015)
<i>CcV1</i>	CcBV	Immunity	Cellular immunity	Inhibition of phagocytosis	Labropoulou <i>et al.</i> (2008)
<i>Glc1.8, ptp-H2, ptp-H3</i>	MdBV	Immunity	Cellular immunity	Inhibition of phagocytosis	Pruijssers and Strand (2007)
<i>ptp-H2</i>	MdBV	Immunity	Cellular immunity	Induction of hemocyte apoptosis	Suderman <i>et al.</i> (2008)
<i>TnBVank1</i>	TnBV	Immunity	Cellular immunity	Induction of hemocyte apoptosis	Salvia <i>et al.</i> (2017)
<i>p-vank – 1</i>	CsIV	Immunity	Cellular immunity	Inhibition of hemocyte apoptosis	Fath-Goodin <i>et al.</i> (2009)
<i>ank-h4, ank-n5</i>	MdBV	Immunity	Humoral immunity	Inhibition of antimicrobial peptides production	Bitra <i>et al.</i> (2012), Thoetkiattikul <i>et al.</i> (2005)
<i>Not determined</i>	CvBV	Immunity	Humoral immunity	Inhibition of PO activity	Wang <i>et al.</i> (2018)
<i>Not determined</i>	HdIV	Immunity	Antiviral immunity	Activation of antiviral gene transcription	Darboux <i>et al.</i> (2019)
<i>TnBVank3, TnBVank1</i>	TnBV	Development	Hormonal regulation	Arrest of ecdysteroidogenesis	Ignesti <i>et al.</i> (2018), Salvia <i>et al.</i> (2018)
<i>CpBV-H4</i>	CpBV	Development	Molting	Down-regulation of insulin signaling	Kumar <i>et al.</i> (2016)
<i>CpBV_E94k</i>	CpBV	Development	Growth	No data	Kim and Hepat (2016)
<i>vinnexinG</i>	CsIV	Development	Growth	No data	Hasegawa <i>et al.</i> (2017)
<i>TnBVank1</i>	TnBV	Development	Growth	Decrease of amino acid transport	Di Lelio <i>et al.</i> (2014)

CpBV, Cotesia plutellae Bracovirus; CsIV, Campoletis sonorensis Ichnovirus; CvBV, Cotesia vestalis Bracovirus; HdIV, Hyposoter didymator Ichnovirus; MdBV, Microplitis demolitor Bracovirus; TnBV, Toxoneuron nigricaps Bracovirus; TnBV, Toxoneuron nigricaps Bracovirus.

A consequence of this property of PDVs is occurrence of lateral transfers on lepidoptera genomes. Indeed, insertions of bracovirus sequences are present in the genomes of certain moth and butterfly lineages; the viral genes present in these sequences have been co-opted by lepidopteran species to confer some protection against pathogens.

The Proviral Genome Maintained in the Wasp Genome

PDV sequences persist in all cells of the wasp as integrated proviruses. PDV chromosomal genomes include two functional components, the viral segments encoding the virulence genes and the viral replication genes (Fig. 4). The segmented proviral sequences are amplified, excised and circularized as dsDNA molecules that are packaged into the particles to be delivered to the caterpillar host. On the other side, DNA for the replication genes, which are also amplified, are not encapsidated and remain in the wasp. The latter are expressed specifically in calyx cells during virus replication and encode proteins found, for the majority of them, associated with the virus particles (by proteomic analyses of purified particles); they constitute the viral machinery that produces the PDV particles. These viral machineries derive from the PDV viral ancestor, which differs between BVs and IVs.

Organization of PDV Proviral Segments in the Wasp Genome

Proviral segment organization in the wasp genome differs between BVs and IVs. The first data came from the analysis of genomic clones from two *Glyptapanteles* braconid wasps. It was the first demonstration that BV proviral segments were organized in “macro-loci” containing several segments arranged in tandem array. The organization in macro-loci of BV proviral segments was confirmed by later studies on several *Cotesia* species and *M. demolitor*. The number of macro-loci and the number of segments differ between species, but globally, braconid wasps harbor in their genome a dozen viral loci containing from 18 to a single segment. For *M. demolitor*, availability of the whole genome allowed the demonstration that the eight loci containing MdBV proviral segments were flanked by large distances of wasp genomic DNA, suggesting a wide distribution of the viral loci within the wasp genome. Organization in clusters may facilitate concerted segment DNA amplification during BV replication. Indeed, it was shown that BV segments are amplified within replication units (RU) composed of several segments and including intersegmental sequences that are not encapsidated. The hypothesis suggested by the authors is that the concatemer is first excised and then sub-divided into individual segments. The organization of proviral segments is quite different in ichneumonids. The genome of two campoplegine species *Hyposter didymator* and *Compoletis sonerensis* reveals that IV segments are highly dispersed across the wasp genomes. Therefore the question on how the coordinated amplification of all the viral loci is regulated remains unanswered.

In its linear form, each PDV segment is flanked in its extremities by a direct repeat, named “direct repeat junction” (DRJ). DRJs are the regions where homologous recombination takes place allowing excision of the segment; consequently, a single copy of the DRJ sequence is present in the circular molecule (Fig. 3). In IVs, no consensus sequence has been identified. Conversely, in BVs,

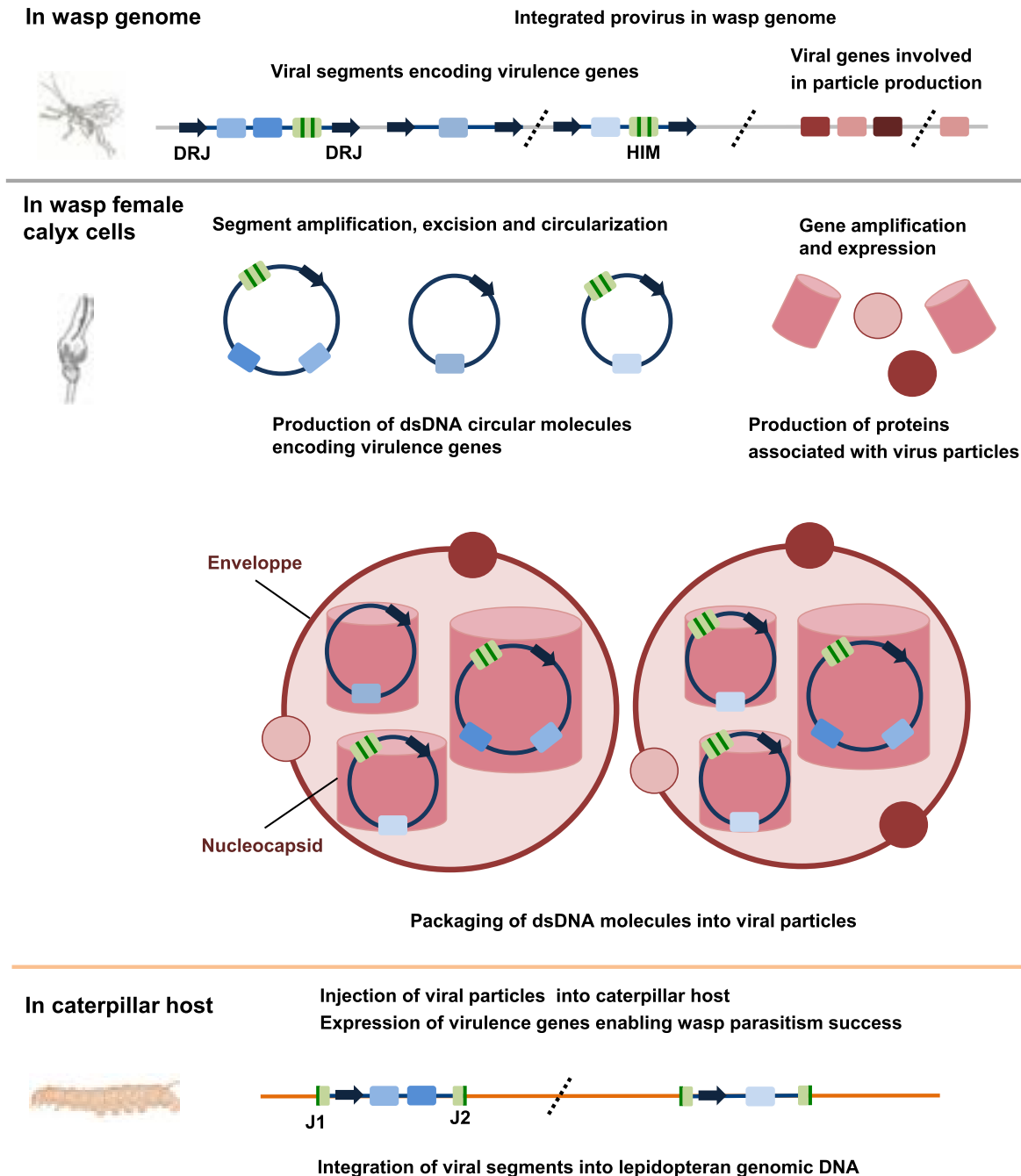


Fig. 4 The different forms of PDV genomes. In the wasp, PDV genome sequences are present as integrated proviruses composed of two components. One component is composed of proviral segments, each of which are flanked by direct repeat junctions (DRJ) (dark blue arrows) and encode virulence genes (blue squares). Homologous recombination at the level of the DRJs allows segment excision and leads to the formation in the calyx cells of double stranded circular DNA molecules. For BVs, amplification of segments within replication units was shown to precede segment excision. The other proviral component corresponds to genes belonging to “BVSPERs” or “IVSPERs” “BV or IV structural proteins encoding regions” (red squares). In the case of BVs, these regions correspond to genes of nudiviral origin, whereas in the case of IVs the genes belong to a yet to be described virus. Expression of these genes in the calyx cells allows production of viral particles and the packaging of the double stranded circular molecules. These non-replicative PDV particles are then injected inside the caterpillar along with wasp eggs upon parasitism. Expression of PDV virulence genes occurs early after parasitism and is in part necessary for successful wasp development. PDV DNA was shown to persist in the genome of cultured lepidopteran cells. In the case of BVs, integration of BV DNA segments into genomic DNA of the host during parasitism was described for molecules containing a 100 bp “Host Integration Motif” (HIM). During integration BV circles are opened within the HIM, a short 40–50 bp sequence of the circle is lost and the remaining HIM sequences flanking this lost sequence become the extremities of the BV sequences integrated in the host DNA, these extremities are named Junction 1 and 2 (J1 and J2) and are represented by dark green lines. The fate of molecules that do not contain HIM sequences awaits investigation. The wasp genome is represented by a gray line. DRJs are represented by dark blue arrows. PDV virulence genes are represented by blue squares, virus-derived genes encoding virion structural proteins are represented by red squares. The lepidopteran genome is represented by an orange line. The HIM is represented by a green square, and J1 and J2 sequences by dark green lines.

DRJs contain a conserved tetramer AGCT, shown to be the site where the segment circularizes. Note that this motif differs from the host integration motif (HIM) mentioned above.

The PDV Viral Machineries

One of the major findings of the last 10 years in the field of PDVs has been the discovery by coupling calyx transcriptome and virus particle proteome analyses, of the virus ancestors and of the virus-derived machineries allowing for the production of PDV particles. Importantly, BVs and IVs do not share a common viral ancestor, they have different evolutionary origins. BVs descend from the integration in the genome of an ancestral wasp of a nudivirus genome, whereas IVs originate from the integration of a still undetermined virus.

Nudiviruses represent a sister-group of baculoviruses, with which they share a set of core-genes. A large proportion of nudiviral structural core genes was found to be maintained in the braconid wasp genome. These nudivirus-like genes potentially (1) code for the different subunits of a DNA-dependent RNA polymerase and the *lef-5* initiation factor involved in transcription; (2) are involved in producing and assembling viral nucleocapsids; (3) code for a set of *per os* infectivity factors (PIFs) that could be involved in infection of host cells; or (4) are constituents of the viral particles (see Table 4). Several lines of evidence suggest that nudivirus-like genes in braconid wasp genomes have retained their ancestral function. First, nudivirus-like gene expression dynamics are consistent with the different steps in virion formation. Secondly, proteomic analysis showed that products corresponding to capsid or envelope components can indeed be present in the BV virions. Finally, RNA-interference experiments in *M. demolitor* showed that the *Lef-4* and *Lef-9* RNA polymerase subunits are necessary for transcription of viral structural genes involved later in the virion cycle. RNA-interference in this model also revealed that *vp39*, *p74*, and *pif-1* are required for virion formation; while *vlf-1* and the integrase display functions required for DNA excision and circularization. Notably, however, none of the braconid genomes sequenced so far possess a DNA polymerase of viral origin, suggesting viral amplification is under the control of wasp DNA polymerases. Half of the nudiviral genes identified within the genome of the braconid wasp *Cotesia congregata* are localized in a 17 kb region referred to as the “nudiviral cluster” whereas the other nudiviral genes are dispersed in the wasp genome. A nudiviral cluster was similarly found in the *M. demolitor* genome displaying a conserved synteny of structural genes with the *Cotesia congregata* nudiviral cluster. This observation shows that these nudiviral clusters have remained stable since divergence 53 MYA between the *C. congregata* and *M. demolitor* species.

Regarding IVs, virus-derived genes encoding virion structural proteins were all found in distinct regions of the genomes of the campoplegine *Hyposoter didymator* and *Camponotus sonorensis*, and of the banchine *Glypta fumiferanae*. These regions were termed “IV

Table 4 Nudivirus genes, with known or predicted functions, identified in braconid wasp genomes

Function	Gene name	MdBV	CcBV	CiBV
DNA replication/processing	<i>helicase</i>	+	nd	nd
	<i>integrase</i>	+	nd	+
Transcription by DNA-dependent RNA polymerase	<i>p47 RNA polymerase subunit</i>	+	+	nd
	<i>lef – 4 RNA polymerase subunit</i>	+	nd	+
	<i>lef – 5 initiation factor</i>	+	+	nd
	<i>lef – 8 RNA polymerase subunit</i>	+	+	+
	<i>lef – 9 RNA polymerase subunit</i>	+	nd	nd
Packaging, assembly, morphogenesis	<i>38k phosphatase</i>	+	+	+
	<i>vp39 major capsid protein</i>	+	+	+
	<i>vlf1 multifunctional</i>	+	nd	+
	<i>p33 sulphydryloxidase</i>	+	nd	nd
Envelope component/infectivity	<i>pif – 0 (p74)</i>	+	+	+
	<i>pif – 1</i>	+	nd	+
	<i>pif – 2</i>	+	nd	+
	<i>pif – 3</i>	+	+	nd
	<i>pif – 4</i>	+	+	+
	<i>pif – 5</i>	+	+	+
	<i>pif – 6</i>	+	+	nd
	<i>pif – 8 (vp91)</i>	+	nd	+
	<i>11k</i>	–	+	+
Particle component	<i>HZNVOrf9-like</i>	+	+	+
	<i>HZNVOrf106-like</i>	+	+	+
	<i>HZNVOrf118-like</i>	+	+	+

MdBV, *Microplitis demolitor* Bracovirus; CcBV, *Cotesia congregata* Bracovirus; CiBV, *Chelonus inanitus* Bracovirus. +, gene present; –, gene absent; nd, gene not detected to date. Gene names in bold correspond to core genes that are also conserved in baculoviruses.

Structural Proteins Encoding Regions" (IVSPERs). Five IVSPERs ranging in size from 2 to 30 kbp were identified in the *H. didymator* genome, and *C. sonorensis* genomes, and three IVSPERs were also identified in the *G. fumiferanae* genome. So the large majority of the replication genes are clustered in the wasp genome, although both campoplegine wasps genomes also contain one or two isolated genes. IVSPERs of these closely or distantly related species showed clear inter-taxonomic homology detected for several IVSPER genes. The function in particle assembly and trafficking was demonstrated for six *H. didymator* replication genes through RNA-interference experiments, clearly showing that IVSPER genes have classical viral functions.

Concluding Remarks

Polydnaviruses are unusual viruses due to their atypical life cycle and to their genome composed of two functional components. The virus particles are non replicative and merely consist in gene transfer systems allowing delivery of virulence wasp genes to the host of the parasitoid. As a consequence, the virus status of PDVs could be questioned. Nonetheless, via the replication genes that derive from the viral ancestors, PDVs still possess typical virus characteristics, as the endogenous viral elements present in the wasp genomes have clearly retained their ancient viral functions.

Polydnaviruses have a complex life style and they are still wrapped in many mysteries, especially in regard to the field of classical, molecular and evolutionary virology. For example, how these genomes are regulated to produce the viral particles, what are the mechanisms underlying particle assembly, where do the envelopes that form in the nuclei come from, how do these integrated genomes evolve, etc ... so many research questions that future generations will have to answer before the mystery of these unique domesticated viruses becomes somewhat clearer.

Further Reading

- Beckage, N.E., Drezen, J.M., 2012. Parasitoid Viruses: Symbionts and Pathogens. Academic Press. p. 312. doi:10.1016/C2009-0-64055-1.
- Burke, G.R., 2019. Common themes in three independently derived endogenous nudivirus elements in parasitoid wasps. *Current Opinion in Insect Science* 32, 28–35. doi:10.1016/j.cois.2018.10.005.
- Burke, G.R., Simmonds, T.J., Sharanowski, B.J., Geib, S.M., 2018. Rapid viral symbiogenesis via changes in parasitoid wasp genome architecture. *Molecular Biology and Evolution* 35, 2463–2474. doi:10.1093/molbev/msy148.
- Drezen, J.M., Leobold, M., Bézier, A., *et al.*, 2017. Endogenous viruses of parasitic wasps: Variations on a common theme. *Current Opinion in Virology* 25, 41–48. doi:10.1016/j.coviro.2017.07.002.
- Feschotte, C., Gilbert, C., 2012. Endogenous viruses: Insights into viral evolution and impact on host biology. *Nature Reviews Genetics* 13, 283–296. doi:10.1038/nrg3199.
- Lorenzi, A., Ravallec, M., Eychenne, M., *et al.*, 2019. RNA interference identifies domesticated viral genes involved in assembly and trafficking of virus-derived particles in ichneumonid wasps. *PLOS Pathogens* 15 (12), e1008210. doi:10.1371/journal.ppat.1008210.
- Pichon, A., Bézier, A., Urbach, S., *et al.*, 2015. Recurrent DNA virus domestication leading to different parasite virulence strategies. *Science Advances* 1, e1501150. doi:10.1126/sciadv.1501150.
- Strand, M.R., Burke, G.R., 2014. Polydnaviruses: Nature's genetic engineers. *Annual Review of Virology* 1, 333–354. doi:10.1146/annurev-virology-031413-085451.
- Strand, M.R., Burke, G.R., 2019. Polydnaviruses: Evolution and function. *Current Issues in Molecular Biology* 6, 163–182. doi:10.21775/cimb.034.163.

Poxviruses of Insects (*Poxviridae*)

Basil Arif and Lillian Pavlik, Laboratory for Molecular Virology, Great Lakes Forestry Centre, Sault Ste Marie, ON, Canada

Remziye Nalçacıoğlu, Hacer Muratoğlu, and Cihan İnan, Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Mehtap Yakupoğlu, Trabzon University, Trabzon, Turkey

Emine Özşahin, University of Guelph, Guelph, ON, Canada

Ismail Demir, Kazım Sezen, and Zihni Demirbağ, Department of Biology, Karadeniz Technical University, Trabzon, Turkey

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Alkaline midgut The midgut of the lepidopteran larvae has a pH about 10.5 and contains proteases active at this high pH.

Apoptosis Programmed cell death. Viruses, toxins, etc. induce apoptosis in certain cells.

Ecdysteroids Arthropod hormones responsible for molting and development.

Epithelial cells A layer of cells the form part of the midgut and is permissive to virus infection.

Fusolin It is the main protein of the spindle body. The N-terminus is associated with degradation of the peritrophic membrane and enhances virus access to epithelial cells.

gfp A gene encoding the green fluorescent protein, GFP.

Juvenile hormone A group of hormones that regulate larval growth and prevent metamorphosis.

Lateral bodies Proteinic bodies present in the virion as a single unit on one side of a unilateral concave core in alphaentomopoxviruses or as two units on either side of a cylindrical viral core (beta- and gammaentomopoxviruses).

Malpighian tubules A system for the elimination of metabolic waste as well as regulators of osmolarity.

Peritrophic membrane It is a non-cellular, semi-permeable membrane composed mainly of chitin, which lines the larval midgut. It improves digestion, protects against chemical and mechanical damage and acts as a barrier to pathogenic infections.

Phylogeny It is the evolution of genetically related organisms (here viruses) by studying their gene or protein evolution through comparison of homologous sequences.

Spheroid A proteinic body into which mature virions are embedded. The spheroid also contains an endogenous alkaline protease. It serves as a protective agent as it delivers the virus to the larval midgut.

Spindles A para crystalline body that in some entomopoxviruses become occluded into the spheroid late in the infection cycle.

Tracheoblasts Cells derived from epidermal cells that line the trachea to give rise to the tracheoles.

Introduction

Over the decades, poxviruses have been associated with a most serious smallpox infection caused by two poxviruses, Variola major and Variola minor (Genus *Orthopoxvirus*), which resulted in millions of human deaths before it was eradicated with the last case reported in October 1977. Entomopoxviruses (EPVs) were discovered by C. Vago in 1963 where he noted very close structural similarities to poxviruses of vertebrates. EPVs, however, are distinguished by the synthesis of a spherical-like structure, termed spheroid or occlusion body into which virions are embedded at the end of the viral replication cycle. The spheroid consists mainly of a protein, spheroidin of 110–115 kDa and functions much like the polyhedra of nucleopolyhedroviruses. Spheroids and polyhedra afford the occluded virions a certain amount of protection against inactivating environmental conditions such as desiccation, UV light, heat, etc. but are also the means to deliver the virus to permissive midgut cells in the host insect. To date, EPVs have been isolated from Coleoptera, Orthoptera, Lepidoptera and Diptera and the German cockroach, *Blattella germanica* (Order: Blattodea). As in other poxviruses, EPVs replicate in the cytoplasm, primarily because the virus particle carries a DNA-dependent RNA polymerase to initiate the early stages of gene expression. The virus forms an electron-dense virogenic stroma or viral factories where virions are assembled as spherical immature particles. These virus particles develop into brick-shaped intracellular mature virions (IMVs) that are either released upon cell lysis or acquire a second double membrane from trans-Golgi and bud as external enveloped virions (EEVs). Mature virions become occluded in spheroids late in the infection cycle and remain intracellular until cell lysis. Intracellular particles are highly infectious and, indeed, cells infected with *Amsacta moorei* EPV (AMEV) can cause a systemic infection in the gypsy moth, *Lymantria dispar* by *per os* inoculation. Virus purified from spheroids does not cause a systemic infection in *L. dispar* by *per oral* inoculation as this insect is not its natural host.

Classification

The family *Poxviridae* is divided into two subfamilies, *Chordopoxvirinae*, which encompasses several genera of viruses that infect chordates and *Entomopoxvirinae*, which encompass poxviruses of insects and has been subdivided to contain four genera:

- *Alphaentomopoxvirus* of which the species *Melolontha melolontha entomopoxvirus* is represented by *Melolontha melolontha* EPV (MMEV). Members of this genus infect coleopteran insects (beetles). Virions are 450×250 nm in size and ovoid in shape. They contain a unilateral concave core and a single lateral body. The genome size of these viruses is in the order of 260–370 kb.
- *Betaentomopoxvirus* of which the *Amsacta moorei entomopoxvirus* is the type species. Viruses of this genus infect lepidopteran insects (butterflies and moths). The virions contain a cylindrical core, a “sleeve-shaped” lateral body and are similar in shape to alphaentomopoxviruses (ovoid) but smaller in size (250×350 nm).
- *Deltaentomopoxvirus* of which *Melanoplus sanguinipes entomopoxvirus* is the type species. Members include Orthopteran EPVs such as *Melanoplus sanguinipes entomopoxvirus* ‘O’ (MSEV).
- *Gammaentomopoxvirus* of which the *Chrinomus luridus entomopoxvirus* is the type species. Members of this genus that infect dipteran insects (mosquitoes and true flies) are brick shaped ($320 \times 230 \times 110$ nm) and the virions contain two lateral bodies and a biconcave core.
- Unassigned species in *Entomopoxvirinae*: *Diachasmimorpha entomopoxvirus*

Host range, morphology and virus size are the criteria for the different genera.

Poxviridae and *Iridoviridae* have been grouped as eukaryotic nucleocytoplasmic large DNA viruses (NCLDV) with double-stranded genomes and complete part or all their replication and capsid assembly in the cytoplasm. *Ascoviridae* may also be included in this group. There is a proposed order *Megavirales* to include the above virus families along with *Mimiviridae* and *Phycodnaviridae*.

Virus/Midgut Interactions

Infection starts when larvae ingest plant material contaminated with spheroids that contain occluded virus particles (Fig. 1). Spheroids dissolve in the alkaline environment of the midgut with the aid of an endogenous alkaline protease. Released virions negotiate the peritrophic membrane and attach to permissive epithelial cells. Infection is enhanced in part by a virus-encoded protein termed fusolin which appears to cause significant disruption of the peritrophic membrane. Fusolin is the major protein of spindle-like structure and is occluded into the spheroids of most EPVs (Fig. 2). Some EPVs do not occlude the spindles such as the *Anomala cuprea* EPV (ACEV) and *Pseudaletia separata* EPV (PSEV). Fusolins are conserved at the amino acid level and exhibit significant identity with an apparently unrelated baculovirus (BV) protein, GP37. Both fusolin and GP37 have a 20 amino acid leader sequence that is cleaved post translationally. Interestingly, GP37 forms spindle-like structures in the cytoplasm of certain nucleopolyhedrovirus infected cells (Fig. 3). Fusolin not only enhances the infectivity of EPVs, it also appears to interact functionally with BVs and boosts the infectivity of some nucleopolyhedroviruses (NPVs). The spindle protein, not the spheroid, causes the disintegration of the peritrophic membrane in the silk moth to aid the penetration of greater numbers of NPV virions into microvilli.

A series of deletion ACEV fusolin mutants were constructed, expressed in a BV expression system and assayed for enhancement of BmNPV infectivity. It was observed that the N-terminus of fusolin (aa 1–253) is essential to enhance the infectivity of BmNPV in

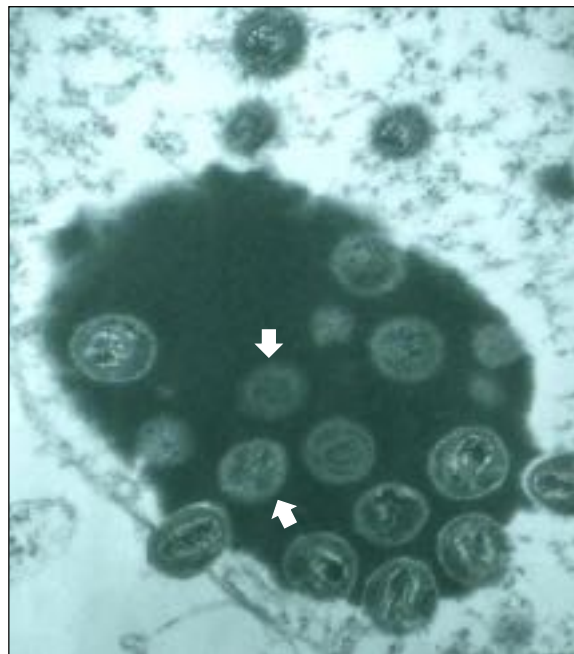


Fig. 1 *Choristoneura biennis entomopoxvirus* is occluded within a spheroid. Arrow points to virions maturing within the spheroid.

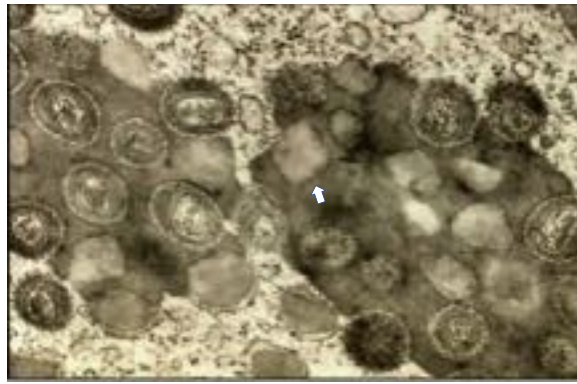


Fig. 2 Virions and spindles (arrow) occluded within a spheroid.

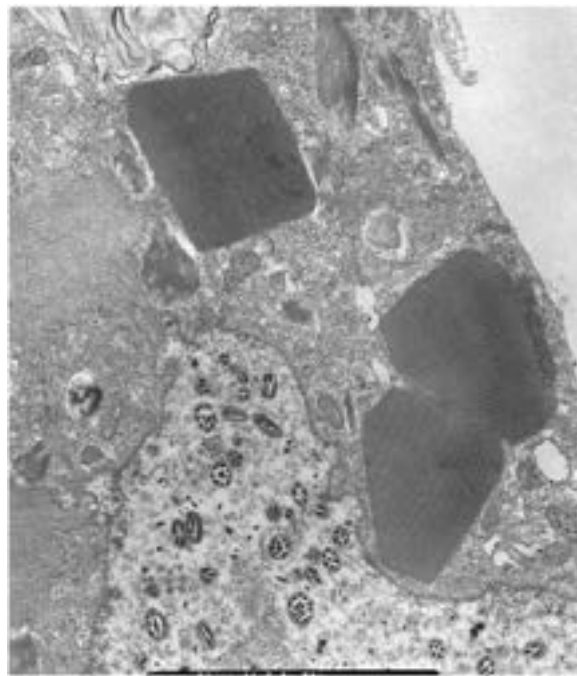


Fig. 3 GP37 spindles in the cytoplasm of CF70 cells infected with *Choristoneura fumiferana* defective nucleopolyhedrovirus (CfDEFNPV).

the silk moth. Interestingly, that part of the N-terminus is a chitin-binding domain that causes degradation of the peritrophic membrane. Also, essential to fusolin activity and stability in midgut juices is the glycosylation of N191. The carboxy-terminus does not appear to play a role in this function.

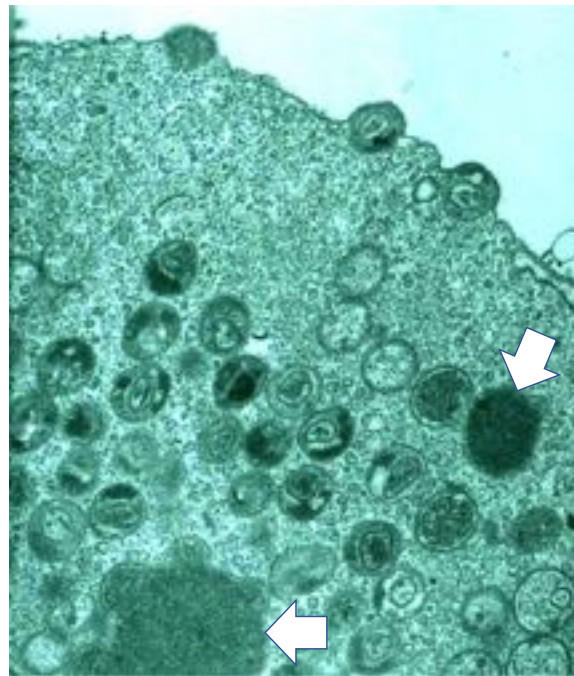
As with many other viruses, EPVs induce apoptosis upon infection of certain cells. Most but not all EPVs contain inhibitor of apoptosis genes (*iap*) but their anti-apoptotic properties have not been established. BVs express a virus-encoded protein, P35, that targets cellular caspases to inhibit apoptosis. AMEV encodes P33, which is a functional homolog of the baculovirus P35 protein, which inhibits programmed cell death in certain insect and human cell lines. It is specific for effector but not initiator caspases and mutagenesis of the caspase cleavage site annuls cleavage and subsequent anti-apoptotic activity.

Replication in Larvae

Much like the occlusion bodies of baculoviruses, the function of the spheroid is not only to give the virions a certain amount of protection in the environment but, more importantly to deliver the virus to permissive midgut cells. The spheroid contains an endogenous alkaline protease that aids in dissolution to efficiently release the virions. Initiation of infection starts when the outer envelope protein of the virion attaches and fuses with the plasma membrane allowing entry of the viral core and lateral bodies into the cell cytoplasm. The entry–fusion complex in chordopoxviruses (ChPVs) has been substantially elucidated, primarily in vaccinia

Table 1 Potential EPV genes related to fusion entry complex and orthologs in Fowlpox virus

Predicted Function Structural	Homologs to AMEV	Homologs to MSEV	Homologs of Chordopoxvirus
Myristylated protein (Cop-G9R)	AMEV – 035	MSEV-Tuc – 121	FWPV-Iowa – 127
Myristylated Entry/Cell-cell fusion protein (Cop-A16L)	AMEV – 118	MSEV-Tuc – 090	FWPV-Iowa – 181
Unknown MP (Cop-J5L)	AMEV – 232	MSEV-Tuc – 142	FWPV-Iowa – 136
S-S bond formation pathway protein (Cop-F9L)	AMEV – 243	MSEV-Tuc – 094	FWPV-Iowa – 112
Entry and Cell-Cell Fusion (Cop-A21L)	AMEV – 249	MSEV-Tuc – 209	FWPV-Iowa – 186
IMV MP/Virus entry (Cop-A28L)	AMEV – 186	MSEV-Tuc – 132	FWPV-Iowa – 192
Entry and Cell-Cell Fusion (Cop-H2R)	AMEV – 127	MSEV-Tuc – 060	FWPV-Iowa – 139

**Fig. 4** Viroplasm of CBEV infected cells and maturing budded virions. Intracytoplasmic virus particles are at various stages of maturity. Arrows point to electron-dense virogenic stroma.

virus and the H3 component of the complex is a homolog of the AMEV, *amu248*. **Table 1** delineates AMEV ORFs potentially encoding components of the entry–fusion complex and orthologues in Fowlpoxvirus. Virus replication is exclusively in the cytoplasm of infected cells in electron dense viroplasm or virus factories where immature virion cores are formed. Maturation of the virions in some hosts occurs after occlusion into the spheroids (**Figs. 1 and 2**). Also, extracellular virus may mature following budding from infected cells (**Fig. 4**).

EPVs infect four orders of insects, Lepidoptera, Orthoptera, Diptera and Coleoptera and, while the virus is infective to larvae, an EPV of the European spruce bark beetle, *Ips typographus*, appears to target adult insects only.

The primary larval tissue that supports virus replication is the fat body although other tissues also become infected. *Melanoplus sanguinipes* EPV (MSEV) is more restricted and infects fat tissue and haemocytes. A more comprehensive pathway is seen when AMEV tagged with a *gfp* marker under the control of the very late *spheroidin* promoter is used to infect the gypsy moth, *Lymantria dispar* by the per oral route. Since *gfp* is under the control of the very late *spheroidin* promoter, GFP expression indicates the virus has completed productive replication. The first signs of fluorescence are in the haemocytes that appear to carry the virus to the rest of the larval tissues. Unlike BVs, where the trachea is the conduit to disseminate the virus systemically, AMEV does not appear to infect tracheoblasts (**Fig. 5**) until the very late stages of infection of larva and by that time the virus has spread to all susceptible tissues. The fat tissue however is still the main site that supports virus replication where the virus progressively invades all cells and the tissues become the main site for productive replication (**Fig. 6**). Virus replication also occurs in the silk glands, the hind- and midgut. The malpighian tubules are refractory to infection. At the latter stages of infection, the gypsy moth becomes totally fluorescent with GFP (**Fig. 7**).

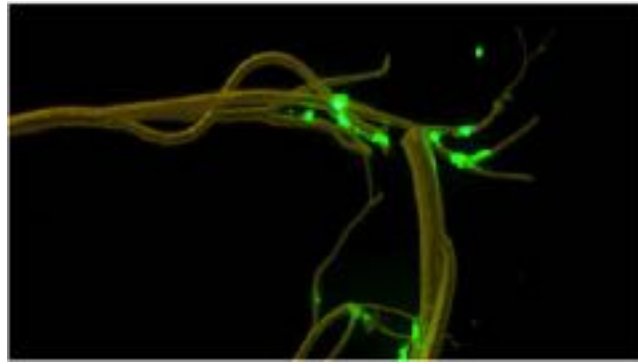


Fig. 5 AMEV expressing GFP in haemocytes proximal to uninfected trachea. Reproduced from Perera, S., Zhen, L., Pavlik, L. and Arif B. (2010). Entomopoxviruses. In *Insect Virology* (Edited by: Sassan Asgari and Karyn N. Johnson), pp 83–102. Caister Academic Press, UK.

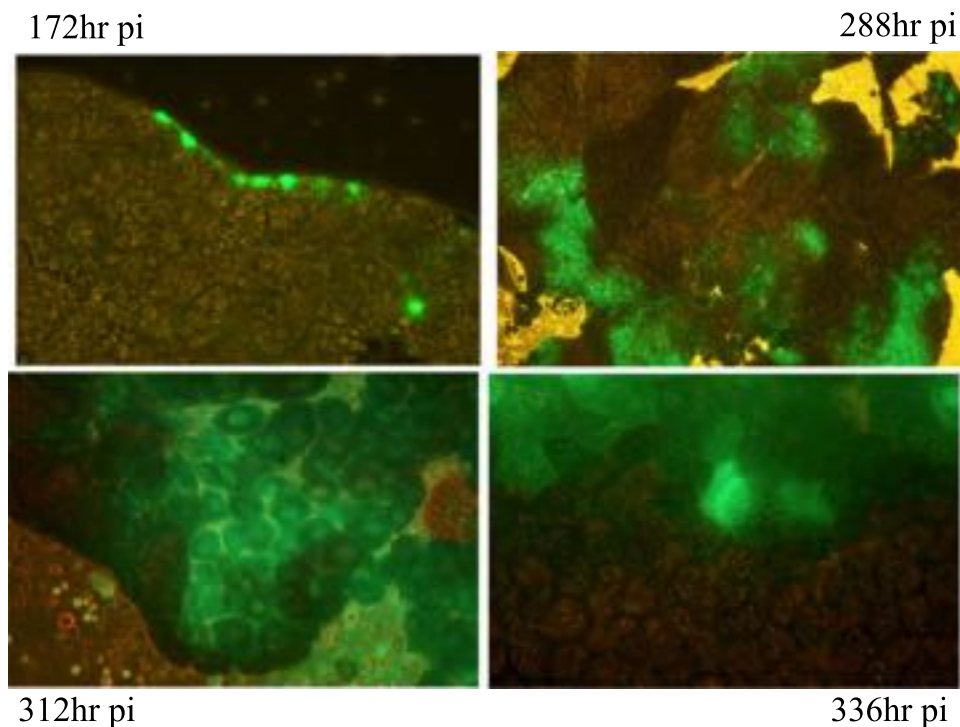


Fig. 6 Gypsy moth fat tissue infected by AMEV *gfp*⁺ *sph* at different hours post infection (pi). Reproduced from Perera, S., Zhen, L., Pavlik, L. and Arif B. (2010). Entomopoxviruses. In *Insect Virology* (Edited by: Sassan Asgari and Karyn N. Johnson), pp 83–102. Caister Academic Press, UK.

To date, EPVs have not been used to any extent in the control of agricultural or forest insect pests primarily because the virus takes a long time to kill the insect, usually 4–5 weeks. There has been some correlation observed between the host and the amount of time it takes to kill the larvae. For example, coleopteran EPVs could take up to 37 weeks, and may be temperature dependent. On the other hand, lepidopteran EPVs take about 4 weeks. Apart from general morbidity, symptoms of infected larvae differ from host to host. Larvae of the lepidopteran lesser cornstalk borer, *Elasmopalpus lignosellus* become unresponsive and sluggish just before they die. Their cuticle changes from brown to reddish in color and the hemolymph turns white or light blue. EPV-infected black-soil scarab, *Othnonius batesi* develop a characteristic white-spotted or mottled exterior triggered by the accumulation of infected fat body in the last segment that enclose the rectal sac. Lepidopteran larvae such as the Winter moth, *Operophtera brumata* and the Army cutworm, *Euxoa auxiliaris* become whitish but the Salt marsh caterpillar, *Estigmene acrea* remain dark after infection. Infected larvae of the Eastern spruce budworm, *Choristoneura fumiferana* (CFEV) or the Two-year-cycle budworm, *C. biennis* EV (CBEV) become much larger in size because pupation is either delayed or inhibited until they die. EPVs interfere with metamorphosis by modulating the levels of juvenile hormone (JH) and ecdysteroids in the hemolymph. Larvae treated with JH, appear to undergo similar symptoms. When infected with EPV, many die as larval/pupal or pupal/adult intermediates. It appears that upon infection with an EPV, the level of JH is elevated with a concomitant reduction of the ecdysteroid titer.

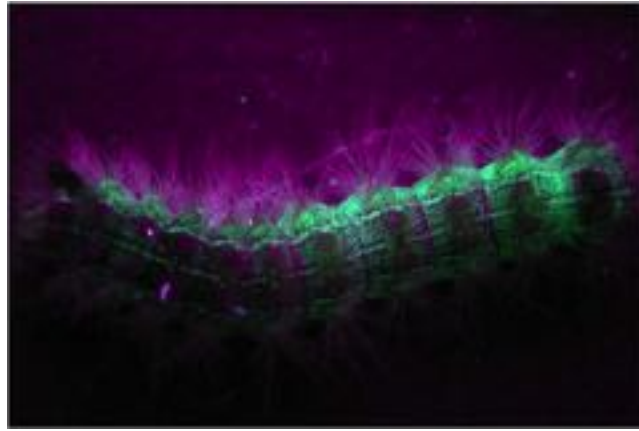


Fig. 7 Latter stages of a whole gypsy moth larva infected with AMEV *gfp*⁺ *sph*[−] exhibiting GFP expression due to virus infection.

Table 2 Sequenced entomopoxvirus genomes

Viral Genome	Size in bp	Size of ITR in bp
Anomala Cuprea EPV	245,717	22,978
Amsacta moorei EPV	232,392	9458
Choristoneura biennis EPV	307,691	23,817
Choristoneura rosaceana EPV	282,895	13,406
Adoxophyes honmai EPV	228,750	5617
Mythimna separate EPV	281,182	7347
Melanoplus sanguinipes EPV	236,120	7201

Genomics

Conserved and Shared Genes

Poxvirus genomes share 49 core genes completely conserved among members of *Chordopoxvirinae* and *Entomopoxvirinae*. The conserved genes play roles in key functions such as transcription, DNA replication and virion structural assembly and are considered essential for virus survival and may even reflect a minimum poxvirus genome. At least 90 genes are common to all ChPVs genera including the 49 core genes. While the central region of the genome contains genes required for essential functions, the termini contain genes specific to a particular genus and some are involved in virus/host interactions. Some genes in the termini encode virulence proteins to counteract the host's immune responses.

During the co-evolution of ChPVs with their hosts, the virus acquired genes encoding proteins needed to evade the host's immune responses. A similar process has most likely developed in EPVs and will probably be elucidated as more studies are conducted on the interactions of EPVs with their natural hosts.

Several EPV genes and their homologs in ChPVs have been characterized with respect to their temporal transcription and viral functions. The AMEV protein kinase gene (*amv197*) is expressed early and encodes a protein of 299 amino acids with conserved serine/threonine protein kinases (PK) domains. In a yeast two-hybrid system, this enzyme and another PK encoded by *amv153* interacted specifically with a number of viral proteins including two involved in the fusion entry complex. Interestingly, *amv133* encodes a protein with two types of enzyme activities, an esterase and a protease.

In comparison to ChPV, to date, only seven EPV genomes have been completely sequenced (Table 2). The EPV genomes are distinguished by having a very high A + T content measuring 79%–82%. A characteristic poxvirus genome, including EPVs, comprises a highly conserved central region where genes and ORFs show collinearity flanked by variable terminal regions. In relation to ChPVs, the central region in the EPV genomes have undergone significant gene rearrangements during the evolutionary separation of the two poxvirus subfamilies. Indeed, comparison of gene order and arrangements in the central part of the genomes of AMEV, CREV, CBEV and MSEV (Fig. 8) reveals almost complete collinearity in the first three betaentomopoxvirus genomes. Gene order has been practically totally rearranged in the genome of MSEV thus ascertaining its divergence from the betaentomopoxviruses, as recognized by the ICTV by being classified in the new (2019) *Deltaentomopoxvirus* genus (Fig. 9).

The first EPV genome to be completely sequenced was the 236-kb MSEV genome containing 267 ORFs. In 2000, the sequence of AMEV (232-kb) became available and was assumed to have 294 protein-coding ORFs. The sequences of five other genomes were reported later (Table 2) and, as in other poxviruses, EPV genomes contain inverted terminal repeat regions (ITR) which range from 5.6 kb to about 24 kb in length (Table 2).

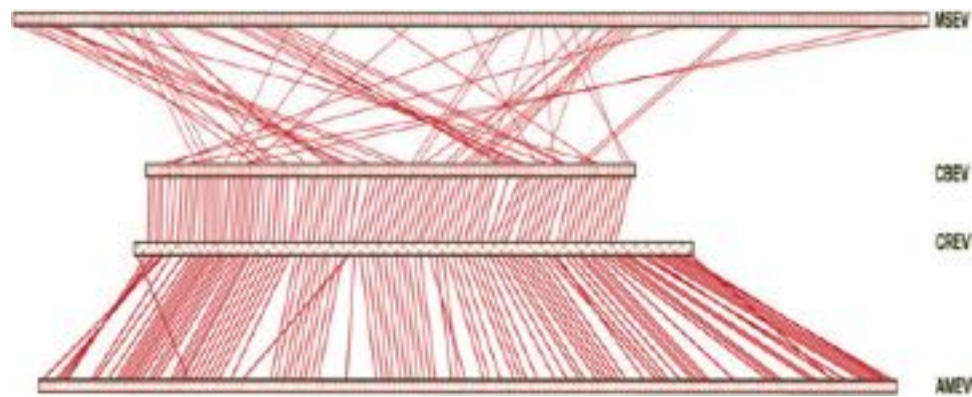


Fig. 8 Gene collinearity in the central part of the genomes of AMEV, CREV, CBEV and MSEV.

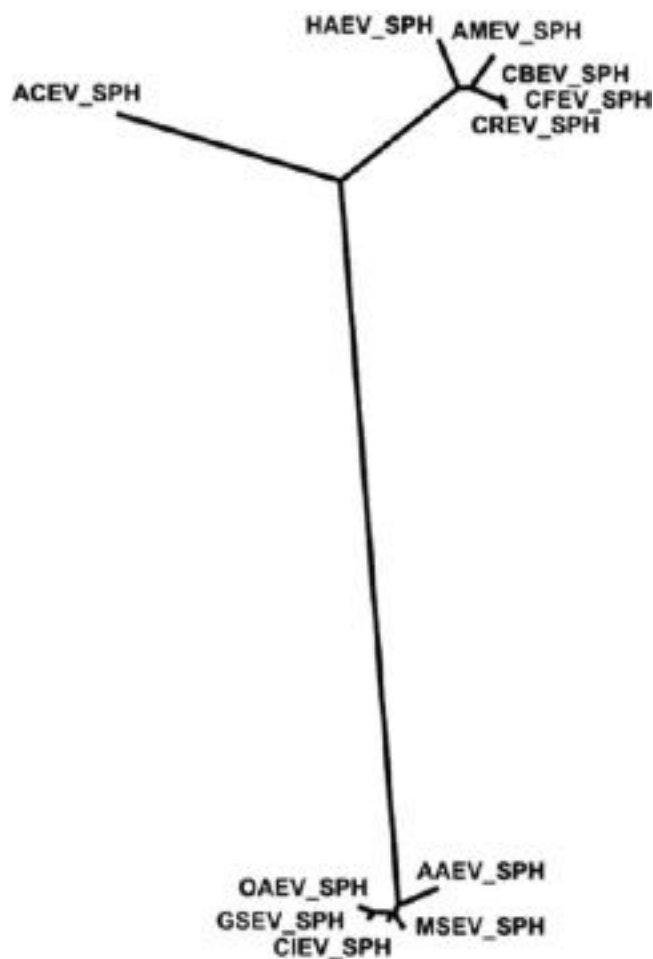


Fig. 9 Unrooted phylogeny of EPVs based on the protein, spheroidin. Reproduced from Perera, S., Zhen, L., Pavlik, L. and Arif B. (2010). Entomopoxviruses. In *Insect Virology* (Edited by: Sassan Asgari and Karyn N. Johnson), pp 83–102. Caister Academic Press, UK.

The initial prediction of 294 protein-coding ORFs in the genome of AMEV is probably an over estimate and the actual number may be significantly less. Based on purine content and the average aspartate-glutamate-serine composition, a method was described to predict functional ORFs in A + T rich genomes. By this method, 45 of the previously considered ORFs in the AMEV genome are likely to be non-coding and cannot be considered as genes. By using the Z-curve method, two other studies predicted that 38–48 previously annotated AMEV ORFs do not encode proteins. Based on the latter two studies, the AMEV genome encodes 246–256 functional proteins, which is significantly less than the 294 ORFs reported earlier.

Table 3 Poxvirus gene families with orthologs in MMEV and MSEV

Family	AMEV ORF	MSEV ORF
ALI (alanine-leucine-isoleucine motif)		
Subgroup 1	AMV257	MSV024, 026, 196, 204
Subgroup 2	AMV 055, 057, 175, 177	MSV023, 194, 195
Leucine-rich repeats (LRR)	AMVITR01, AMV005, 014, 076,134	MSV008, 009, 010, 011, 013, 014, 015, 016, 018, 186, 227, 228, 239, 240, 241, 253, 254, 255, 257, 260, 261
Methionine-threonine-glycine (MTG)	AMV194, 207, 209	MSV021, 191, 198, 199
Tryptophan (W) repeat	AMV029, 254	MSV027, 029, 034, 197, 205, 252
AMV176	AMV56, 176, 178	
Serine-cystein-glycine (SCG)		MSV062, 214, 215, 216, 217
17K ORF (KilK-N domain-containing proteins)		MSV062, 214, 215, 216, 217

The genome sequence of the EPV infecting the Oriental beetle, *Anomala cuprea* (genus *Alphaentomopoxvirus*) revealed that it is 245,717 bp long, smaller than previously reported and contains the 49 core genes conserved in poxviruses. Because only one *Alphaentomopoxvirus* genome has been sequenced and five from the genus *Betaentomopoxvirus*, it is a too early to make generalized conclusion as to the major differences between viruses from the two genera. However, some differences are worth mentioning. For example, ACEV does not have a homolog to the apoptosis inhibitor, P33 but contains serpins known to inhibit apoptosis in ChPVs. The ITRs of ACEV and CBEV are in the order of 23–24 kb, which are significantly longer than those in other EPVs (Table 2). Also, in the terminus of each ITR of ACEV, there is a concatemer resolution motif, which has been reported to be essential for resolving concatemeric DNA during replication of ChPV genomes.

Gene Families

EPVs contain genes that can be assembled into seven gene families (Table 3). Of these families, the AMEV genome contains 23 genes in six families and MSEV has 43 genes in five families. To date, the function of these genes has not been ascertained. However, it is quite reasonable to assume that the ALI (alanine-leucine-isoleucine) family members and 17K ORF family play functional roles in transcription and replication of DNA. The ALI gene family is characterized by an amino-terminal motif that contains invariant residues of alanine, leucine, and isoleucine. With the exception of AMV257, all members of the ALI motif family contain the Bro-N domain (N-terminus of baculovirus BRO proteins). The function of the baculovirus BRO proteins is still not clear but has been suggested that they are DNA binding proteins and may also be involved in regulating host DNA replication and/or transcription. The 17K ORF gene family is comprised of five AMEV proteins and homologs of the 17K ORF of *Heliothis armigera* entomopoxvirus (HAEV). This family has a putative DNA binding domain (KilA-N domain) present in many proteins in eukaryotic and bacterial DNA viruses, which implies a role in transcription and/or replication.

Entomopoxvirus Phylogeny

A variety of methods have been employed in phylogeny analysis including morphological, gene contents, order and arrangements or even individual genes. Relatively recently, single nucleotide polymorphism (SNP) and alignment-free methods have been used to illustrate phylogeny of poxviruses. Collinearity between any two genomes reflects conservation of genes and their order within the genomes. Based on phylogeny of concatenated amino acids of the 49 core genes, MSEV was observed to be distinctly divergent from betaentomopoxviruses. Similarly, phylogeny based on spheroidins revealed that the betaentomopoxviruses cluster together and are divergent from the alphaentomopoxviruses and deltaentomopoxviruses (Fig. 9). Indeed, a new genus was proposed, now called *Deltaentomopoxvirus* and comprising orthopteran EPVs: MSEV, *Calliptamus italicus* EPV (CIEV), *Gomphocerus sibiricus* EPV (GSEV), *Oedaleus asiaticus* EPV (OAEV), and *Anacridium aegyptium* EV (AAEV). These five EPVs are evolutionarily divergent from the alpha- and betaentomopoxviruses. Among the same genes, betaentomopoxviruses have a low degree of identity with MSEV. For example, spheroidin is about 20% identical between MSEV and other betaentomopoxviruses such as AMEV, CBEV, HAEV as well as the spheroidins of the alphaentomopoxviruses, MMEV and ACEV. In contrast, 76%–92% amino acid identity was detected among betaentomopoxviruses spheroidins. Moreover, the organization and gene order among the lepidopteran betaentomopoxviruses are quite conserved particularly the central region in the genomes. Such conservation is not observed between betaentomopoxviruses and those unassigned in the classification scheme. The molecular data that revealed the divergence between betaentomopoxviruses and MSEV validate the decision by the ICTV to remove the latter from the genus *Betaentomopoxvirus* and reclassify it as member of the *Deltaentomopoxvirus* genus.

While some virus groups such as EPVs and BVs have descended from different ancestral origins, they evolved to have remarkably similar processes of infecting the natural host. Analysis by secondary structure alignments revealed 33 clusters of homologous genes to be common to EPVs and BVs and not surprisingly, *DNA polymerase* is among the homologous clusters. The genes were independently acquired from the insect host, other viruses, eukaryotes and bacteria. This gene acquisition convergence may indicate adaptations to conserved interactions of the virus with the natural host. In fact, 15 genes appear to have been acquired by EPV and BV core genomes from the natural host by horizontal gene transfers.

Acknowledgment

We are indebted to Ms. Erika Provenzano for expending time and effort in literature searches and procuring relevant published manuscripts to aid in the preparation of this review.

Genomics research on EPVs was supported by a major grant from Genome Canada through the Ontario Genomics Institute and three grants from the Scientific and Technological Research Council of Turkey, TUBITAC.

It is acknowledged that Figs. 5, 6 and 9 have been published previously in the book "Insect Virology", Caister Academic Press, Norfolk, UK. The publisher is hereby acknowledged.

Further Reading

- Arif, B.M., 1995. Recent advances in the molecular biology of entomopoxviruses. *Journal of General Virology* 76, 1–13.
- Becker, M., Moyer, R., 2007. In: Poxviruses, A.A., Mercer, A., Schmidt Weber, O. (Eds.), Subfamily *Entomopoxvirinae*. Basel, Switzerland: Birkhauser Verlag, pp. 253–271.
- Kurstak, E., 1991. The entomopoxviruses. In: Arif, B.M., Kurstak, E. (Eds.), *Virus of Invertebrates*. Marcel Dekker Inc, pp. 179–195.
- McLysaght, A., Baldi, P.F., Gaut, B.S., 2003. Extensive gene gain associated with adaptive evolution of poxviruses. *Proceedings of the National Academy of Sciences* 100, 15655–15660.
- Perera, S., Li, Z., Pavlik, L., Arif, B., 2010. Entomopoxviruses. In: Asgari, S., Johnson, K. (Eds.), *Insect Virology*, Caister Academic Press, pp. 83–102.
- Roberts, D.W., Yendol, W.G., 1973a. Insect Poxviruses: Pathology, morphology and development. In: Granados, R.R. (Ed.), *Some Recent Advances in Insect Pathology*. Miscellaneous Publications of Entomological Society of America, pp. 73–94.
- Senkevich, T.G., Bugert, J.J., Sisler, J.R., *et al.*, 1996. Genome sequence of a human tumorigenic poxvirus: Prediction of specific host response-evasion genes. *Science* 273, 813–816.
- Théze, J., Takatsuka, J., Nakai, M., Arif, B., Herniou, E.A., 2015. Gene acquisition convergence between entomopoxviruses and baculoviruses. *Viruses* 7, 1960–1974.

Reoviruses of Invertebrates (*Reoviridae*)

Peter J Krell, Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

© 2021 Elsevier Ltd. All rights reserved.

This is an update of P.P.C. Mertens, H. Attoui, Insect Reoviruses, In Encyclopedia of Virology (Third Edition), edited by Brian W.J. Mahy, Marc H.V. Van Regenmortel, Elsevier Ltd., 2008, doi:10.1016/B978-012374410-4.00610-5.

Glossary

Diptera (also known as true flies) An order of insects that comprises mosquitoes, gnats, midges, sand flies, and other flies. They possess a single pair of wings on the mesothorax (the middle of the three segments of the thorax of an insect) and a pair of halteres which are derived from the hind wings present on the metathorax (the posterior of the three segments of the thorax of an insect). The name is derived from the number of wings (*di* means two in Greek and *pteron* refers to the wings).

Hymenoptera This order of insects comprises the sawflies, wasps, bees, and ants. The name refers to the membranous wings of the insects and is derived from ancient Greek (*hymen*: membrane and *pteron*: wings). The hind wings of hymenopterans are connected to their forewings by hooks.

Lepidoptera This order of insects comprises butterflies, skippers, and moths. The name is derived from ancient Greek and refers to the minute scales (*lepidon*) that cover the membranous lanceolate wings (*pteron*) of adults.

Reovirus A virus in the family *Reoviridae* characterized by an icosahedral capsid of about 80 nm with one or two shells encompassing a segmented (usually 10–12) dsRNA genome with an aggregate size of about 25 kb. The name reovirus comes from respiratory enteric orphan virus, since these viruses caused respiratory and enteric infections. Since they were originally thought not to be associated with disease they were erroneously considered orphan viruses.

RNA dependent RNA polymerase (RdRp) An RNA polymerase which produces an RNA product from an ssRNA template.

Taxonomy

[in brackets is first year in which ICTV accepted the taxon name].

Family: *Reoviridae* [ICTV 1974].

Order: *Reovirales* ICTV (2019).

Higher Orders: Realm *Riboviria*, Kingdom *Orthornavirae*, Phylum *Duplornaviricota*, Class *Resentoviricetes*.

Subfamily *Sedoreovirinae* [ICTV 2009] Subfamily *Spinareovirinae* [ICTV 2009].

<i>Sedoreovirinae</i>	<i>Spinareovirinae</i>
Genus: <i>Cardoreovirus</i> [ICTV 2008] (1 species)	Genus: <i>Aquareovirus</i> [ICTV 1990] (7 species)
Type Species: <i>Eriocheir sinensis reovirus</i> [ICTV 2008]	Type Species: <i>Aquareovirus A</i> [ICTV 1999]
Species Exemplar: <i>Callinectes sapidus reovirus 1</i>	Species Exemplar: American oyster reovirus13p2
Genus: <i>Phytoreovirus</i> [ICTV 1978] (3 species)	Genus: <i>Cypovirus</i> [ICTV 1990] (16 species)
Type Species: <i>Wound tumor virus</i> [ICTV 1976]	Type Species: <i>Cypovirus 1</i> [ICTV 1999]
Species Exemplar: <i>Wound tumor virus (34)</i>	Species Exemplar: <i>Bombyx mori cypovirus 1</i>
Genus: <i>Seadornavirus</i> [ICTV 2004] (3 species)	Genus: <i>Dinovernavirus</i> [ICTV 2008] (1 species)
Type Species: <i>Banna virus</i> [ICTV 1999]	Type Species: <i>Aedes pseudoscutellaris reovirus</i> [ICTV 2008]
Species Exemplar: <i>Banna virus-China</i>	Species Exemplar: <i>Aedes pseudoscutellaris reovirus</i>
	Genus: <i>Fijivirus</i> [ICTV 1978] (9 species)
	Type Species: <i>Fiji disease virus</i> [ICTV 1979]
	Species Exemplar: <i>Fiji disease virus</i>
	Genus: <i>Idnoreovirus</i> [ICTV 2004] (5 species)
	Type Species: <i>Idnoreovirus 1</i> [ICTV 2004]
	Species Exemplar: <i>Diadromus pulchellus idnoreovirus 1</i>
	Genus: <i>Orthoreovirus</i> [ICTV 1991] (10 species)
	Type species: <i>Mammalian orthoreovirus</i> [ICTV 1999]
	Species Exemplar: <i>Mahlapitsi orthoreovirus</i>

Introduction

As a group, the reoviruses have a broad host range, including insects, algae, fungi, fish, crustaceans, crabs, humans and other vertebrates and their invertebrate vectors and plants and their invertebrate vectors. Since the 2008 Encyclopedia of Virology 3rd Edition, the

Reoviridae family in the Ninth Report of the International Taxonomy of Viruses in 2009, was subdivided into two subfamilies, *Sedoreovirinae* with six genera and *Spinareovirinae* with nine genera. The two subfamilies differ in the nature of the virion surface with the *Spinareovirinae* having a turreted or “spiked” surface (hence the prefix “Spina”, Latin for spike) while the *Sedoreovirinae* have a smooth surface (“Sedo” is Latin for smooth). Of the more recognized reoviruses are those infecting humans, other mammals and other vertebrates. Those from *Rotavirus* and *Orthoreovirus* genera infect only vertebrates, while viruses in genera *Coltivirus* (e.g., Colorado tick fever virus), *Orbivirus* (e.g., bluetongue virus) and one virus in genus *Seadornavirus* are considered arboviruses which infect both vertebrates and their invertebrate vectors (mostly ticks, mosquitoes, midges, gnats and sandflies). Lesser known reoviruses infect algae (*Mimovirus*) and fungi (*Mycovirus*). In addition to infecting plants, reoviruses in genera *Fijivirus*, *Oryzavirus* and *Phytovirus* also infect their insect vectors (grasshoppers). These are described in more detail in the article on *Reoviridae* in this Encyclopedia.

There are also reoviruses that infect exclusively invertebrates. There are many reoviruses infecting only insects in the genera *Cypovirus* and *Idnoreovirus*. Other reoviruses infecting only insects are found in genera *Dinovernavirus*, *Fijivirus*, *Orthoreovirus*, *Pytoreovirus* and *Seadornavirus*. Reoviruses in two genera *Aquareovirus* and *Cardoreovirus*, infect only crustaceans including oysters and crabs, some of economic importance.

Like other reoviruses, the reoviruses of invertebrates have a single or double shelled icosahedral capsid with diameters ranging from 60 to 80 nm (though idnoreovirus capsids are smaller at about a 50 nm diameter). While cypoviruses and dinovernaviruses have a single shell with $T=2$, aquareoviruses, coltivoruses, idnoreoviruses, cardeoviruses, phytoviruses and seadornaviruses have two shells, the inner with $T=2$ and the outer layer with $T=13$ icosahedral symmetry. They have 8–12 dsRNA segments. These range in size from 0.76 to 4.47 kb with each segment, in general, encoding one protein. For cypoviruses, segment 5 encodes two proteins by ribosomal skipping. Aggregate genome sizes range from 18.1 kb to 25.72 kb (although a size of 32.7 kb was reported for an unclassified *Dendrolimus punctatus* cypovirus). With a G + C content (based on those reoviruses whose whole genomes have been sequenced) ranging from 33.25% to 43.81%.

Morphology, infection and replication of the invertebrate reoviruses closely mimic that of other reoviruses and will be covered only briefly here. Instead, the reader is referred to the article in this Encyclopedia on Reoviruses (*Reoviridae*) and Their Structural Relatives.

Spinareovirinae

Much of the material on the *Spinareovirinae* viruses in this article is based on Insect Reoviruses by Peter Mertens and Houssam Attoui in the previous, 3rd, edition of the Encyclopedia of Virology and provided a good foundation for this update.

Some spinareoviruses infecting only invertebrates are found in genera *Cypovirus* (insects), *Dinovernavirus* (mosquitoes) and *Idnoreovirus* (hymenoptera and diptera). One reovirus in genus *Aquareovirus* (which as the name implies, infect aquatic animals) infects only oysters, while others in the same genus infect vertebrates like fish. In the genus *Fijivirus* of plant-infecting reoviruses, one virus, *Nilaparavata lugens* reovirus, has been found only in plant hoppers. One virus in genus *Orthoreovirus*, *Mahlapitsi orthoreovirus* infects only bat flies, while Chiqui virus, another, unclassified, orthoreovirus was isolated from mosquitoes. [Table 1](#) summarizes the recognized species in the *Spinareovirinae* with viruses infecting invertebrates within the six genera listed above.

Of the *Spinareovirinae* insect reoviruses, most cypoviruses have 10 dsRNA segments but can have 8, 11 or even 16 segments (*Dendrolimus punctatus* cypovirus) and infect lepidoptera like *Lymantria dispar* (Gypsy larvae) and one infects the hymenoptera *Polistes hebraeus*. Idnoreoviruses have 10 dsRNA segments and infect hymenoptera like *Diadroma pulchellas* (a hymenopteran parasitoid) and diptera like *Musca domestica*. Dinovernavirus has 9 dsRNA segments and was found in an *Aedes* mosquito AP61 cell line.

Cypovirus was the first invertebrate reovirus discovered and that was in 1934 as polyhedra in the cytoplasm of infected cells of silkworm larvae and hence were initially named cytoplasmic polyhedrosis viruses from which the current genus name was derived. Aquareoviruses (*Aqua* for their common aqueous environment of their hosts) have been isolated from oysters and crabs. Idnoreoviruses (from “idno” meaning water) were first isolated from houseflies in 1978, while the lone dinovernavirus (from double stranded RNA icosahedral *nove*, meaning nine, dsRNA segments *RNA*) was isolated from a cell line from *Aedes pseudoscutellaris* in 2005.

The cypoviruses can produce severe to chronic diseases in their infected hosts and represent one of the most significant threats to the silkworm industry. The idnoreoviruses have a core structure that is directly comparable to the cypovirus virion but they also have an additional ‘outer layer’ of capsid proteins. They do not usually cause a cytopathic effect or disease in their insect hosts but may be involved in modifying the phenotype or sex of the host insect. The dinovernaviruses have been isolated only from cultured insect cells and like cypoviruses have a single-shelled capsid, but do not produce polyhedra.

The *Cytoplasmic polyhedrosis virus* group was initially recognized as a genus name with viruses infecting insects in 1971 by the International Committee on Taxonomy of Viruses (ICTV). Unlike baculoviruses which replicate in the nucleus, these viruses replicate and produce polyhedra in the cytoplasm, giving rise to their name as cytoplasmic polyhedrosis viruses. By 1990 the genus name was changed to *Cypovirus* (sigla: *cytoplasmic polyhedrosis virus*) and recognized by ICTV. Cypovirus particles are ~70 nm in diameter, contain a segmented dsRNA genome of 10 segments, and belong to the “turreted” reoviruses, with surface projections (turrets) situated at each of the 12 “fivefold” vertices. Unlike most other reoviruses, cypoviruses are single-shelled, with no outer capsid layer. Cypoviruses have one unique feature among reoviruses in that they form polyhedra which can retain infectivity for years. The major polyhedrin protein is translated in large excess in the cytoplasm late in infection and virions are occluded in the polyhedrin at the periphery of the virogenic stroma in which morphogenesis of the virions occurs and in which virus particles become singly or multiply embedded. More than 230 cypovirus isolates have been described, with evidence for at least 22 distinct species (based on differences in the migration pattern of their genome segments during electrophoresis–electropherotype and

Table 1 *Spinareovirinae* reoviruses infecting invertebrates, their hosts and some of their genome characteristics

Species	Virus	Host	# Segs	Aggregate genome size (bp)	Range of segment sizes (kb)	% C + C
Genus <i>Aquareovirus</i>						
<i>Aquareovirus A</i>	American oyster reovirus 13p2	Oyster	11			
Genus <i>Cypovirus</i>						
<i>Cypovirus 1</i>	Lymantria dispar cypovirus 1	Lepidoptera	10	24,732	0.94–4.16	43.22
<i>Cypovirus 2</i>	Inachis io cypovirus 2	Lepidoptera	10	24,041	1.19–3.74	38.59
<i>Cypovirus 3</i>	Anaitis plagiata cypovirus 3	Lepidoptera				
<i>Cypovirus 4</i>	Antheria mylitta cypovirus 4	Lepidoptera	11		S9, 1473	
<i>Cypovirus 5</i>	Heliothis armigera cypovirus 5 China	Lepidoptera	10	24,818	0.88–4.12	36.42
<i>Cypovirus 6</i>	Aglais urticae cypovirus 6	Lepidoptera				
<i>Cypovirus 7</i>	Mamestra brassicae cypovirus 7	Lepidoptera				
<i>Cypovirus 8</i>	Abaraxas grossulariata cypovirus 8	Lepidoptera				
<i>Cypovirus 9</i>	Agrotis segetum cypovirus 9	Lepidoptera				
<i>Cypovirus 10</i>	Aporophyla lutulenta cypovirus 10	Lepidoptera				
<i>Cypovirus 11</i>	Heliothis armigera cypovirus 11	Lepidoptera				
<i>Cypovirus 12</i>	Autographa gamma cypovirus 12	Lepidoptera				
<i>Cypovirus 13</i>	Polistes hebraeus cypovirus 13	Hymenoptera				
<i>Cypovirus 14</i>	Lymantria dispar cypovirus 14	Lepidoptera	10	25,325	0.96–4.44	40.29
<i>Cypovirus 15</i>	Trichoplusia ni cypovirus 15	Lepidoptera	11	25,120	0.20–4.36	36.40
<i>Cypovirus 16</i>	Choristoneura occidentalis cypovirus 16	Lepidoptera	8	19,016	1.17–3.77	40.65
Unclassified	Daphnis nerii cypovirus Nanchang	Lepidoptera	8	18,088	1.14–3.97	38.79
Unclassified	Dendrolimus punctatus cypovirus 22 Macheng	Lepidoptera	16	32,753	1.14–4.05	39.94
Unclassified	Biston robustus cypovirus	Lepidoptera	10	23,954	1.12–4.03	
Genus <i>Dinovernavirus</i>						
<i>Aedes pseudoscutellaris reovirus</i>	<i>Aedes pseudoscutellaris</i> reovirus	Mosquito AP61 cells	9	23,147	1.15–3.82	33.95
Unclassified	Fako virus CSW77	Mosquito	9	23,170	1.14–3.82	33.25
Genus <i>Fijivirus</i>						
<i>Nilaparvata lugens reovirus</i>	<i>Nilaparvata lugens</i> reovirus Izumo ^a	Plant hopper	10	1.43–4.39	28,699	34.9
Genus <i>Idnoreovirus</i>						
<i>Idnoreovirus 1</i>	Diadromus pulchellus idnoreovirus 1	Hymenoptera	10/11	6 segs: 0.99–4.23	25.15 kb	
<i>Idnoreovirus 2</i>	Hyposoter exiguae idnoreovirus 2	Hymenoptera	10	1.35–3.9 kb	23.4 kb	
<i>Idnoreovirus 3</i>	Musca domestica idnoreovirus 3	Diptera	10			
<i>Idnoreovirus 4</i>	Dacus oleae idnoreovirus 4	Diptera	10			
<i>Idnoreovirus 5</i>	Ceratitis capitata idnoreovirus 5	Diptera	10			
Unclassified	Drosophila S virus	Diptera	10			
Genus <i>Orthoreovirus</i>						
<i>Mahlapitsi orthoreovirus</i>	Mahlapitsi orthoreovirus 2511 ^b	Bat flies	11	23,195	1.07–3.93	43.81
Unclassified	Chiqui virus COBD38d	mosquitoes	9	25,028	1.15–4.01	36.20

^a*Nilaparvata lugens* reovirus is the only fijivirus which infects exclusively insects (plant hoppers). All other fijiviruses infect plants and their invertebrate vectors.

^b*Mahlapitsi orthoreovirus* is the only orthoreovirus which infects exclusively insects (bat flies), all other orthoreoviruses infect only vertebrates.

nucleotide sequence divergence). However, very few cypoviruses have so far been isolated from Africa, and undoubtedly many more species are sure to be recognized.

The genus *Idnoreovirus* (sigla: insect-derived nonoccluded reovirus) was formally recognized by ICTV in 2005. *Idnoreoviruses* have ten segmented dsRNA genomes and are icosahedral particles, ~70 nm in diameter with a turreted “core”. However, unlike the cypoviruses, they also have an outer capsid protein layer and do not produce polyhedra. *Idnoreoviruses* have been isolated from various flies including the fruit fly, the small fruit fly (*drosophila*), the olive fly, and housefly. They have also been isolated from parasitic wasps.

The genus *Dinovernavirus* (sigla: double-stranded insect, novem (nine in Latin, the number of genome segments) *RNA virus*) includes the first nine-segmented dsRNA virus. Like the cypoviruses, *dinovernaviruses* are single-shelled, turreted, and lack an outer capsid layer, but do not produce polyhedra or a “polyhedrin” protein.

While the genus *Aquareovirus* contains mostly reoviruses that infect fish like salmon, shiner and catfish, one virus in species *Aquareovirus A*, American oyster reovirus, infects oysters and have been isolated from Geoduck, shore, swimmer and mitten crabs.

This virus has 11 dsRNA segments. Virions are about 80 nm in diameter and have multiple capsid layers. The viruses replicate efficiently in fish and mammalian cell lines within a temperature range of 15–25°C.

While most viruses in the genus *Fijivirus* infect plants one *fijivirus*, *Nilaparvata lugens* reovirus (NLRV) has been found exclusively in planthopper but does not seem to replicate in rice. Nevertheless the virus can be transmitted from hopper to hopper through contaminated rice plants. Virions are double shelled icosahedra about 65–70 nm in diameter with 10 dsRNA segments ranging from 1.43 to 4.39 kp in size. The NLRV RNA genome has a G + C content of 34.9% and an aggregate size of 28.7 kb.

Virus members of the genus *Orthoreovirus* are vertebrate viruses usually spread through the fecal oral route. However, one orthoreovirus, *Mahlapitsi orthoreovirus*, infects bat flies and Chiqui virus was found only in mosquitoes. While most orthoreoviruses have a genome of 10 dsRNA segments, the *Mahlapitsi orthoreovirus* has 11 segments ranging in size from 1.07 to 3.93 kb with an aggregate size of 23,195 bp and a G + C content of 43.81%. Chiqui virus has 9 dsRNA segments ranging in size from 1.15 to 4.01 kb, an aggregate size of 25,028 bp and a G + C content of 36.20%.

Historical Overview

Cypoviruses were first recognized by Ishimori in 1934, who observed polyhedra in the cytoplasm of midgut cells of diseased silkworm larvae. Cytoplasmic polyhedrosis has subsequently been recognized as an important cause of economic losses in the Japanese sericulture industry. Cypoviruses have a wide host range, infecting many lepidopteran insect species (more than 200 host species have been identified). Recent reports of cypovirus infections in mosquitoes confirm that they can also infect Diptera. The high levels of sequence divergence between equivalent genome segments of different cypovirus species, and the large numbers of distinct species, suggest that this is an ancient group. Indeed, cypoviruses have been detected in biting midges trapped in amber, indicating that they existed almost 100 million years ago.

The American oyster reovirus in genus *Aquareovirus* was first identified in 1979 in *Crassostrea virginica* by J.R. Winton and colleagues and was later characterized as a reovirus with 11 dsRNA segments in 1987.

The genus *Dinovernavirus* currently contains a single species (*Aedes pseudoscutellaris dinovernavirus*; the virus *Aedes pseudoscutellaris dinovernavirus* 1 (ApDNV-1)) was isolated in 2005 from the AP61 cell line established by Varma *et al.* from the mosquito *Aedes pseudoscutellaris*.

The first idnoreovirus was isolated from the housefly *Musca domestica* in 1978. Several other idnoreoviruses have been isolated. *Diadromus pulchellus idnoreovirus* 1 (DpIRV-1), the most recently isolated idnoreovirus, has a genome composed of 10 dsRNA segments in functional male insects, with an additional segment (derived from genome segment 3) in particles isolated from females or sterile male wasps.

The first report of an insect reovirus in genus *Fijivirus* was in 1991 when Noda and colleagues found *Nilaparvata lugens* reovirus in apparently healthy brown planthoppers *Nilaparvata lugens* and then characterized it as a reovirus with 10 dsRNA segments. Other *fijiviruses* infect plants and their invertebrate vectors.

An insect orthoreovirus in the genus *Orthoreovirus* was first described by PJ Van Vuren and colleagues in 2016 when they isolated *Mahlapitsi orthoreovirus* from bat flies, *Eucampisipoda Africana* associated with African bats *Rousettus aegyptiacus* in South Africa. As part of their studies they determined the sequence of the 10 dsRNA segments determining an aggregate genome size of 23,195 bp and a G + C content of 43.81%. A second invertebrate orthoreovirus was Chiqui virus which was isolated from mosquitoes collected in Columbia by M.A. Contreras-Gutierrez in 2018. Based on genome sequencing of the 9 dsRNA segments with an aggregate size of 25,028 bp and G + C content of 36.20%, they concluded that Chiqui virus was in the subfamily *Spinareovirinae*.

Host Range, Diseases, Transmission, and Distribution

Cypoviruses

Cypoviruses have been isolated from only arthropods, including members of over 45 genera of Lepidoptera. However, they have also been isolated from Hymenoptera, including the wasps *Polistes hebraeus* and *Diadromus pulchellus*, and from Diptera, including the mosquitoes *Culex restuans* and *Uranotaenia sapphirina*. One isolate was reported from the freshwater daphnid *Simocephalus expinosus*. Experimental studies indicate that cypoviruses cannot infect mammals or mammalian cell cultures.

The majority of cypovirus infections produce chronic disease, often without extensive larval mortality. Consequently, many individuals become adults even though they are heavily diseased. Infected larvae can stop feeding as early as 2 days post infection, producing symptoms of starvation. Their body size and weight are often reduced and diarrhea is common. The duration of the larval stage can also increase by ~1.5 times. Infected pupae are frequently smaller and many diseased adults do not emerge correctly, can be malformed, and flightless. Infected females may exhibit a reduced egg-laying capacity.

Cypoviruses can be transmitted on the surface of eggs, producing high levels of infection in the subsequent generation. However, disinfection of the egg surface destroys the virus, indicating that cypoviruses are not transovarially transmitted. The minimum infectious dose increases dramatically in later larval instars, although virulence varies significantly with virus strain and host species. Larvae can sometimes recover from cypovirus infections, possibly because the gut epithelium has considerable regenerative capacity and infected cells are shed at each larval molt.

Cypoviruses are normally transmitted by an oral–fecal route. Ingested polyhedra on contaminated food materials dissolve in the high pH environment of the insect gut, releasing occluded virus particles, which can then infect the cells lining the gut wall.

Infection is generally restricted to the columnar epithelial cells of the larval midgut, although goblet cells may also become infected. Cypovirus infection spreads throughout the midgut region (and sometimes infect the fat body), although infection of the entire gut has occasionally been observed in some species. The production of very large numbers of polyhedra gives the gut a characteristically creamy-white appearance. The endoplasmic reticulum of infected cells is progressively degraded, mitochondria enlarged, and the cytoplasm becomes highly vacuolated. In most cases, the nucleus shows few pathological changes. However a cypovirus strain has been detected that produces polyhedra within the cell nucleus. In the later stages of infection, cellular hypertrophy is common and microvillae are reduced or absent. Polyhedra are released into the gut lumen by cell lysis and are excreted. The gut pH is lowered during infection preventing dissolution of progeny polyhedra in the gut fluid.

Idnoreoviruses

All known members of the genus *Idnoreovirus* infect insects, with host species that include *Diadromus pulchellus* and *Hyposoter exiguae* (wasps: Hymenoptera); *Musca domestica* (housefly); *Dacus oleae* (olive fly); *Ceratitis capitata* (Mediterranean fruit fly); and *Drosophila melanogaster* (small fruit fly) (flies: Diptera).

Unlike the cypoviruses, idnoreoviruses appear to cause few pathological effects in their hosts, although they may significantly influence the biological properties of individual insects. *Drosophila S* virus appears to be associated with the “S” phenotype in *D. simulans* (a maternally inherited morphological trait associated with abnormal bristles). The presence of an additional 3.33 kb dsRNA segment in DpIRV-1 is related to the sex and ploidy of the host, and may play a role in the biology of the host wasp species, possibly by providing information necessary for larval development.

Dinovernavirus

ApDNV-1, the only representative of this genus, was isolated from persistently infected *A. pseudscutellaris* cells.

Aquareovirus

American oyster reovirus 13p2 is the only aquareovirus not isolated from vertebrates but rather from oysters. Although it was originally isolated from oysters, the virus can replicate in Golden shiner fish at temperatures of 23 and 28°C.

Fijivirus

Unlike other fijiviruses which replicate in plants and their invertebrate vectors, *Nilaparvata lugens* reovirus replicates only in invertebrates, specifically the brown planthopper *Nilaparvata lugens* and transmission appears to involve contaminated rice plants. The virus does not appear to replicate in rice.

Orthoreovirus

While most orthoreoviruses infect vertebrates including mammals, birds and reptiles, only two orthoreoviruses, Mahlapitsi orthoreovirus and Chiqui virus infect invertebrates, specifically bat flies and mosquitoes, respectively.

Virion Properties, Genome, and Replication

Cypoviruses

Large proteinaceous inclusion bodies are characteristically produced during cypovirus infections, usually within the cytoplasm of infected insect cells. The crystalline matrix of these “polyhedra” is primarily composed of a single viral “polyhedrin” protein that is often encoded by the smallest genome segment (Seg-10). The polyhedrin structure appears to protect the occluded virus particles, possibly filling the role of outer capsid proteins found in other reoviruses. Polyhedra are frequently symmetrical (cubic, icosahedral, or irregular) depending on both the virus strain (polyhedrin sequence) and the host species. The polyhedrin molecules appear to be arranged as a face-centered cubic lattice with center-to-center spacing varying between 4.1 and 7.4 nm. In some virus species, the virus particles are occluded singly within small occlusion bodies, which can also aggregate. However, some cypoviruses produce large polyhedra containing large numbers of particles, often regularly spaced within the polyhedrin matrix. “Empty” polyhedra containing no virus particles have also been observed.

Cypovirus particles have a single-layered capsid, composed of a central shell, ~57 nm in diameter (by cryoelectron microscopy), which extends to 71.5 nm when the 12 “turrets” or “spikes” on the icosahedral fivefold vertices are included. These turrets are hollow, up to 20 nm in length and 15–23 nm wide (by conventional electron microscopy and negative staining). The cypovirus capsid has a central compartment, ~35 nm in diameter, which normally contains the genomic dsRNA segments.

Cypovirus virions are structurally comparable to the cores of other reoviruses, particularly those from the genera with “turreted” cores (e.g., *Orthoreovirus*, *Aquareovirus*, and *Oryzavirus*) (Figs. 1 and 2). Cypovirus particles generally contain five to six distinct proteins, three of which have been identified as the T2 protein (120 copies, equivalent to the VP3(T2) subcore shell protein of bluetongue virus, or orthoreovirus lambda1); “large protrusion protein” (LPP, 120 copies, comparable to orthoreovirus sigma2); and “turret protein” (TP, 60 copies, comparable to orthoreovirus lambda2). Cypoviruses also contain transcriptase enzyme complexes attached to the inner surface of the capsid shell at the icosahedral fivefold vertices. Two or three of the cypovirus structural proteins are usually > 100 kDa.

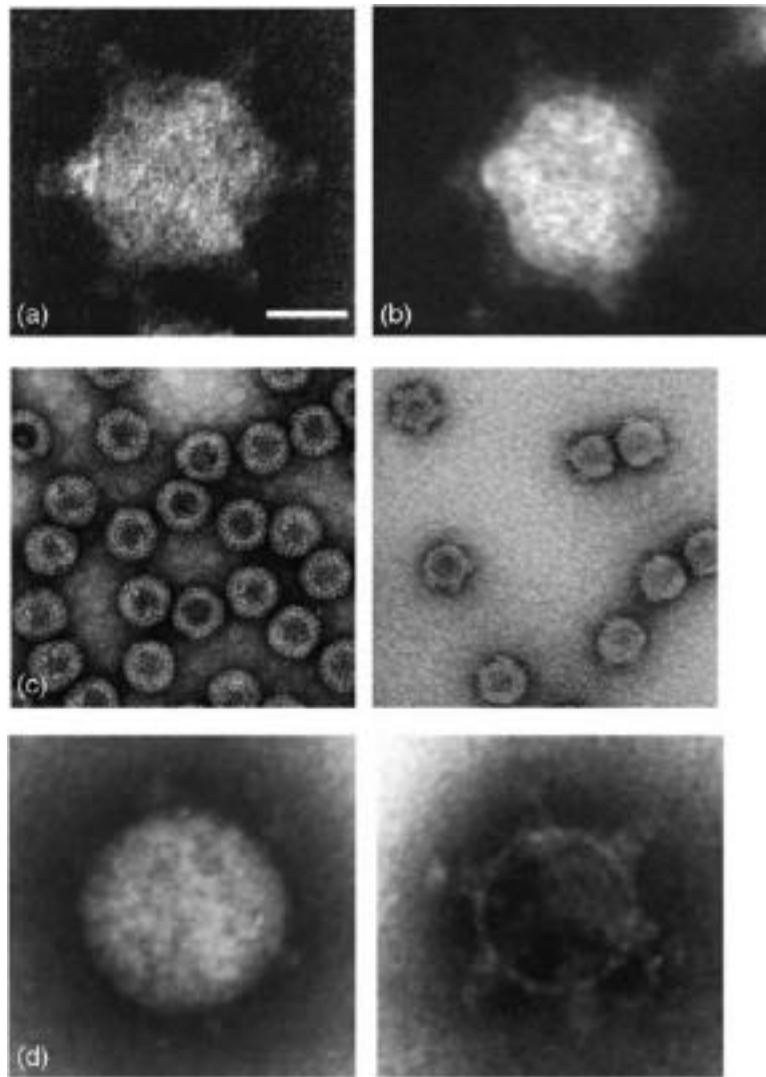


Fig. 1 Electron micrographs of cypovirus, dinovirnavirus, and idnoreovirus particles. (a) Contrast electron micrograph of nonoccluded *Orgyia pseudosugata* cypovirus 5 (OpCPV-5) virion. (b) Contrast electron micrograph of *Aedes Pseudoscutellaris* reovirus particles. The bars represent 20 nm. (c) Electron micrographs of purified virus particles (left) and core particles (right) of *Hyposoter exiguae* idnoreovirus-2 (HeIRV-1) stained with uranyl acetate. (d) Electron micrographs of a virus particle (left) and core particle (right) (stained with sodium phosphotungstate) of purified *Dacus oleae* idnoreovirus-4 (DoIRV-4). DoIRV-4 virions have small icosahedrally arranged surface projections (probably up to 12 in number), while the cores show much larger 'spikes' or 'turrets', which may lose a portion near to the tip (like those of the cypoviruses). (a) Courtesy of C. L. Hill. (c) Courtesy of Andrea Makkay and Don Stoltz. (d) Courtesy of Max Bergoin.

The structural proteins of *Bombyx mori* cypovirus 1 (BmCPV-1) are 148 (VP1), 136 (VP2), 140 (VP3), 120 (VP4), 64 (VP6), and 31 kDa (VP7) (Table 2). Polyhedra also contain the 25–37 kDa "polyhedrin protein" (28.5 kDa for BmCPV-1) which represents approximately 95% of the polyhedra protein (by dry weight). The buoyant density of virions in CsCl is 1.44 g cm^{-3} , c. 1.30 g cm^{-3} for empty particles, and 1.28 g cm^{-3} for polyhedra. Due to the high level of variation among different cypoviruses, it is unlikely that their homologous proteins can be identified simply by their migration order during polyacrylamide gel electrophoresis (PAGE).

Cypoviruses retain infectivity for several weeks at -15 , 5 , or 25°C . The virus retains full enzymatic activity (dsRNA-dependent single-stranded RNA polymerase and capping activity) even after repeated freeze–thawing (up to 60 cycles), even though this disrupts the virus particle into ten active but distinct enzyme/template complexes. Each transcription complex contains one genome segment and a complete transcriptase complex that includes one of the 'spike' structures derived from the fivefold axes of the virion capsid.

Cations have relatively little effect on cypovirus structure. Heat treatment of virions at 60°C for 1 h leads to degradation and release of genomic RNA. Virus particles are relatively resistant to treatment with trypsin, chymotrypsin, ribonuclease A, deoxyribonuclease, or phospholipase. Virion enzyme functions also show some resistance to treatment with proteinase K. However, this

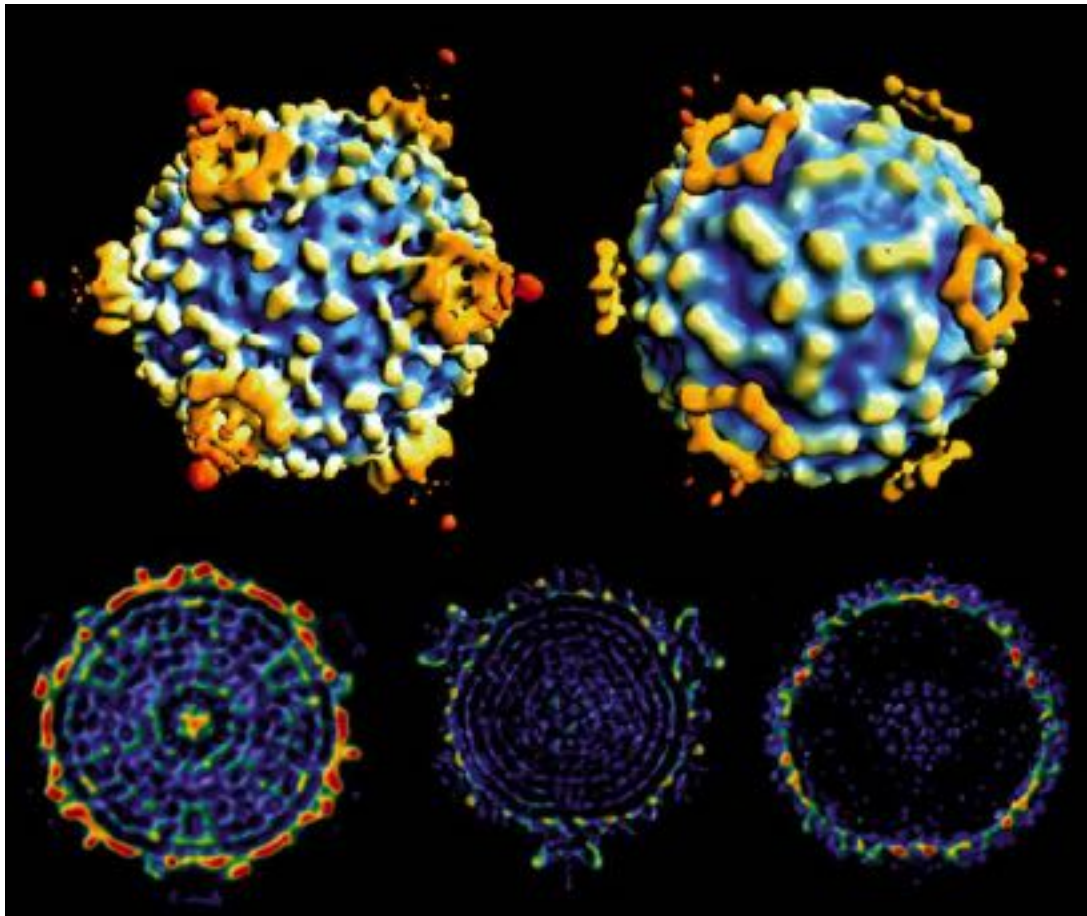


Fig. 2 Cryoelectron microscopy reconstruction of a virion of *Orgyia pseudosugata* cypovirus-5 (OpCPV-5) to 26 Å resolution. Top left: Reconstruction of a nonoccluded virion; top right: reconstruction of an occluded virion. Bottom left: Cross section of a cryoelectron microscopy reconstruction of a fully occluded virion; bottom center: cross section of a full nonoccluded virion; and bottom right: cross section of an empty virion. The cross sections show evidence of dsRNA packaged as distinct layers and suggest localization of the transcriptase complexes at the fivefold axes of symmetry. Courtesy of C. L. Hill.

Table 2 Coding assignments of the CPV type 1 genome segments

Segment	Size (bp)	Protein	Function/location
Seg-1	4190	VP1 (VP1)	Major core (virion)
Seg-2	3854	VP2 (VP2)	RdRp (virion)
Seg-3	3846	(VP3)	(Virion)
Seg-4	3262	VP3 (VP4)	Possible methyltransferase (virion)
Seg-5	2852	NS1 ^a (NS5)	Nonstructural, similar to FMDV 2Apro
Seg-6	1796	VP4 (VP6)	Leucine zipper ATP/GTP binding protein (virion)
Seg-7	1501	NS3, NS4 (VP7)	Nonstructural, with 'structural' cleavage products
Seg-8	1328	VP5 (NSP8)	Unknown (anomalous PAGE migration at 55 kDa)
Seg-9	1186	NS5 (NSP9)	Nonstructural, dsRNA binding
Seg-10	994	Polyhedrin	Polyhedrin protein (Pod)

^aNS1 (NS5): cleaved into NS2 (NS5a) and NS6 (NS5b). RdRp, RNA-dependent RNA polymerase.

Note: From information available at: http://www.iah.bbsrc.ac.uk/dsRNA_virus_proteins/BmCPV-1.htm.

Under the 'Protein' heading, an alternative nomenclature for the viral proteins is given in brackets. The nomenclature that is used in the Eighth Report of the ICTV and presented on the website: http://www.iah.bbsrc.ac.uk/dsRNA_virus_proteins/protein-comparison.htm.

may reflect retention of enzyme activities despite particle disruption, particularly during the early stages of digestion. Cypovirus particles are resistant to detergents including sodium deoxycholate (0.5%–1%) but are disrupted by 0.5%–1% sodium dodecyl sulfate (SDS), which releases the genomic dsRNA. Treatment with triton X-100, NP40, or urea also causes disruption of the virus

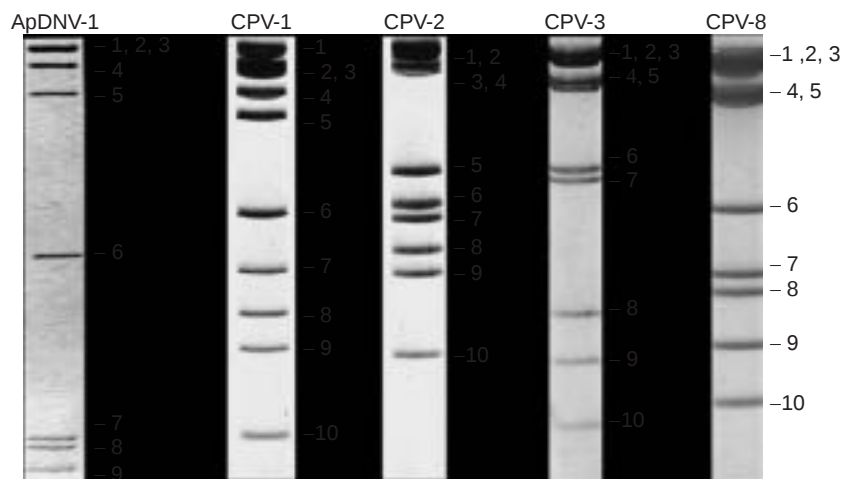


Fig. 3 The genomes of dinovernaviruses and cypoviruses. Genome electropherotypes of members of genera *Dinovernavirus* (ApDNV-1) and *Cypovirus* (CPV – 1, – 2, – 3, and – 8). The segment position is indicated to the right of each profile.

particle structure. One or two fluorocarbon treatments have little effect on virus infectivity, though treatment with ethanol leads to release of RNA from virions. Viruses and polyhedra are readily inactivated by ultraviolet (UV) irradiation which dissociates the dsRNA template from the transcriptase complexes. Polyhedra remain infectious for years at temperatures below 20°C. Virions can be released from polyhedra by treatment with carbonate buffer at pH greater than 10.5 but are disrupted below pH 5. As in permissive insects' midguts, high pH treatment completely dissolves the polyhedral protein matrix. This process is partly due to increased solubility of polyhedrin at high pH but is also aided by alkaline-activated proteases associated within the polyhedra.

Polyhedra (but not virions) contain significant amounts of adenylate-rich oligonucleotides. In the majority of cases, cypovirus particles contain 10 dsRNA genome segments. However, virus particles containing an 11th small segment (e.g., *Trichoplusia ni* cypovirus 15, TnCPV-15 and *Antheria mylitta* cypovirus 4), 8 segments (in *Choristoneura fumiferana* cypovirus 16 and unclassified *Daphnia merii* cypovirus) and up to 16 segments (in unclassified *Dendrolimus punctatus* cypovirus 22) have been detected. The genome segments of BmCPV-1 vary from 4,190–994 bp with a total genome size of 24,809 bp. The genome segments of other cypoviruses have been estimated by electrophoretic comparisons, and have calculated sizes between 5.6 and 0.6 kb, with a total genome size of 29.2–33.3 kb.

The size distribution of the genome segments varies widely among different cypoviruses (e.g., the smallest dsRNA has an estimated size that varies between 0.53 and 1.44 kb). These size differences formed the initial basis for recognition and classification of distinct species (electropherotypes) of cypoviruses (which differ significantly in the migration of at least three genome segments during electrophoresis using 1% agarose or 3% SDS-PAGE). The genome segment migration patterns of types 1, 12, and 14 have some overall similarity, suggesting that they are more closely related than some other cypoviruses (Fig. 3). More accurate sizes are derived from whole genome sequencing (Table 1). By sequence analysis, the aggregate sizes of cypoviruses ranges from 18,088 bp to 25,325 bp, although the largest size of 32,753 bp was reported for *Dendrolimus punctatus* cypovirus 22. Aggregate G + C content ranges from 36.42% to 43.22%. The smallest segment was 0.88 kb (*Heliothis armigera* cypovirus 5) and the largest was 4.44 kb (*Lymantria dispar* cypovirus 14).

The termini of different genome segments within a single cypovirus species are often highly conserved but differ from those reported for other species (Table 3). *Choristoneura fumiferana* cypovirus 16 (CfCPV-16) shows high levels of sequence variation when compared to CPV-1, -2, -5, -14, or -15 and is therefore considered to be a distinct species, even though it has a similar 5' end but different 3' ends to representatives of CPV-5. These data demonstrate that different cypovirus electropherotypes are likely to have different conserved RNA terminal sequences.

Large size variations in the genome segments of most cypovirus species (apart from CPV-1, -12, and -14) indicate that the gene assignments for one "type" or species cannot simply be applied to the other cypoviruses. Genome segment coding assignment by *in vitro* translation of individual genome segments have been published for isolates of CPV-1 and -2. These data and subsequent sequencing studies indicate that the polyhedrin protein is often encoded by the smallest segment.

Cypovirus uptake appears to be a relatively inefficient process in insect cell cultures but can be significantly improved by lipofection. Liposomes deliver cypovirus particles into the cytoplasm where replication occurs. Cypoviruses do not require particle modification to activate transcriptase and capping enzymes. The outer coat proteins of other reoviruses need to be removed to activate transcription. Cypovirus polymerase activity can show very pronounced dependence on the presence of S-adenosyl-l-methionine or related compounds, although this varies between different cypovirus species.

Virus replication and assembly occur within the host cell cytoplasm, although viral RNA synthesis can sometimes occur within the nucleus. Replication is accompanied by the formation of viroplasm (virogenic stroma or "VIB") within the cytoplasm, containing

Table 3 The conserved terminal sequences of member viruses of genera *Cypovirus*, *Idnoreovirus*, *Dinovernavirus*, and their comparison to those of genera *Fijivirus* and *Oryzavirus* (both are plant viruses)

Virus	(Isolate)	Conserved RNA terminal sequences
Cypovirus 1	(BmCPV-1)	5'-AGUAA.....GUUAGCC-3'
	(DpCPV-1)	5'-AGTAA.....GUUAGCC-3'
	(LdCPV-1)	5'-AGU ^A / _G ^A / _GG ^U / _C UAGCC-3'
Cypovirus 2	(LiCPV-2)	5'-AGUUUUA..... UAGGUC-3'
Cypovirus 4	(ApCPV-4)	5'-AAUCGACG..... GUCGUAG-3'
	(AaCPV-4)	5'-AAUCGACG..... GUCGUAG-3'
	(AmCPV-4)	5'-AAUCGACG..... GUCGUAG-3'
Cypovirus 5	(OpCPV-5)	5'-AGUU.....UUGC-3'
Cypovirus 14	(LdCPV-14)	5'-AGAA.....CAGCU-3'
Cypovirus 15	(TnCPV-15)	5'-AUUAAAAA..... GC-3'
Cypovirus 16	(CfCPV-16)	5'-AGUUUUU.....UUUGUC-3'
Idnoreovirus 1	DpIRV-1	5'- ^A / _U / _G CAAUUU.....AGUAAAAAAUn ^A / _G -3'
Dinovernavirus	APDNV	5'-AGUU ^A / _U ^A / _U AGU-3'
Fijivirus	NLRV	5'-AGU.....GUUGUC-3'
	MRCV	5'-AAGUUUUUU.....GUC-3'
	FDV	5'-AAGUUUUUU.....GUC-3'
	RBSDV	5'-AAGUUUUUU.....GUC-3'
Oryzavirus	RRSV	5'-GAUAAA..... GUGC-3'

Note: Data concerning the terminal regions of members of the *Reoviridae* are listed at: http://www.iah.bbsrc.ac.uk/dsRNA_virus_proteins/CPV-RNA-Termin.htm.

large amounts of viral proteins and virus particles. The mechanism used to select the individual genome segments for packaging and assembly into progeny particles (exactly one copy of each segment) is unknown. The importance of the conserved terminal regions in this process is indicated by packaging and transcription of a mutant Seg-10 of CPV-1 that contained only 121 bp from the 5' end and 200 bp from the 3' end. Particles are occluded within polyhedra apparently at the periphery of the VIB, from about 15 h post-infection onward. The polyhedrin protein is produced late in infection and in large excess compared to the other viral proteins. It is unknown how polyhedrin synthesis is regulated.

Idnoreoviruses

Electron microscopy and negative staining of the idnoreovirus particles show that they have no prominent features, are spherical in appearance (icosahedral symmetry), and ~70 nm in diameter. However, they do have a clearly defined outer capsid layer that can be removed to reveal the virus core. Unlike the cypoviruses, idnoreoviruses do not encode a polyhedrin protein, and the virus particles are "nonoccluded". Idnoreovirus core particles have an estimated diameter of ~60 nm, with 12 icosahedrally arranged surface "turrets" or "spikes".

Limited studies of some viruses within the genus indicate that they are resistant to freon (trichlorotrifluoroethane) and CsCl. They may also be resistant to chymotrypsin. Intact virus particles and cores of the prototype isolate *Diadromus pulchellus* idnoreovirus-1 (DpIRV-1) have densities of 1.370 and 1.385 g ml⁻¹, respectively, while intact virions and empty particles of *Dacus oleae* idnoreovirus-4 (DoIRV-4) have a density of ~1.38 and ~1.28 g ml⁻¹, respectively (determined by CsCl gradient centrifugation). The outer capsid layer of *Hyposoter exiguae* idnoreovirus-2 (HeIRV-2) can be disrupted by brief exposure to 0.4% sodium sarcosinate, releasing the virus core.

The genome of most idnoreoviruses consists of 10 linear segments of dsRNA that are conventionally identified as "genome segment 1" to "genome segment 10" (Seg-1 to Seg-10) in order of reducing molecular weight and increasing electrophoretic mobility during agarose gel electrophoresis (AGE). The total genome of DpIRV-1 contains an estimated 25.15 kb of dsRNA, with individual segments ranging between ~4.8 and ~0.98 kb, and an electrophoretic migration pattern (by 1% AGE) showing five larger and five smaller segments (a "5/5" pattern). However, the virions of DpIRV-1 may be atypical since they can sometimes also contain an 11th, 3.33-kb genome segment, the presence of which is related to the sex and ploidy of the individual wasp host. This additional dsRNA (migrating between genome segments 3 and 4) contains sequences similar to and therefore possibly derived from Seg-3 (3.8 kb).

The genome segments of HeIRV-2 range in size from ~3.9 to ~1.35 kb, with a "4/6" electrophoretic migration pattern (by PAGE). DoIRV-4 contains an estimated 23.4 kb, with segments estimated from ~3.8 to ~0.7 kb and a "5/3/2" electrophoretic migration pattern (by PAGE). *Ceratitis capitata* idnoreovirus-5 (CcIRV-5) has a "3/3/4" genome segment migration pattern when analyzed by PAGE and has clear similarities to *Drosophila melanogaster* idnoreovirus-5 (DmIRV-5), suggesting that despite some

Table 4 Coding assignments for the DpIRV-1 genome segments

Segment	Size (bp)	Protein	Function/location
Seg-1	4800	VP1	
Seg-2	4230	VP2	
Seg-3	3812	VP3	Contains RdRp motifs
Seg-x	3333		Sequence closely related to Seg-3. Presence related to sex and ploidy of the host
Seg-4	3000	VP4	
Seg-5	2700	VP5	
Seg-6	1750	VP6	
Seg-7	1652	VP7	
Seg-8	1318	VP8	
Seg-9	1240	VP9	
Seg-10	985	VP10	

Note: Seg-x: the supernumerary segment that is linked to the sex and ploidy of the insect host.

RdRp, RNA-dependent RNA polymerase.

More information is available at: http://www.iah.bbsrc.ac.uk/dsRNA_virus_proteins/Idnoreovirus.htm.

serological differences, they belong to the same virus species, specifically *Idnoreovirus* 5. It is unclear how closely drosophila S virus is related to the other drosophila viruses. It is therefore, currently, classified as a “tentative species” within the genus.

By analogy with other reoviruses (e.g., the orbiviruses), genome segment migration patterns (by AGE) are likely to be characteristic of each *Idnoreovirus* species. Initial sequencing studies suggest that the 3' termini of DpIRV-1 genome segments are more variable than most other reoviruses with little sign of conservation. However, conserved sequences were detected at the 5' termini (Table 3), which are different from those of other reovirus species. Except for the sequences of 6 segments from DpIRV-1, no sequence data are currently available for other members of this genus.

Although the proteins of DpIRV-1 have not been extensively characterized, purified virions contain 11 structural proteins with M_r 21–140 kDa (as analyzed by SDS-PAGE). Three of these appeared to be glycosylated (M_r approximately of 21, 15, and 35 kDa). Some of the viral genome segments have been sequenced (as indicated in Table 4). The viral proteins are named as VP1–VP10 based on the molecular weight of the genome segment (segment number) from which they are translated.

On the basis of their structural and biochemical similarity, it seems likely that many aspects of the genome organization and replication of the idnoreoviruses will show similarities to those of other reoviruses (particularly the other turreted spinareoviruses). On this basis it is likely that the virus core will contain transcriptase complexes that synthesize mRNA copies of the individual genome segments. These will be exported and translated to produce the viral proteins within the host cytoplasm. These positive-sense RNAs are also likely to form templates for negative-strand synthesis during progeny virus assembly and maturation. The genome segments that have been characterized represent single genes, with a large ORF, and relatively short noncoding terminal regions.

Dinovernavirus

Transmission electron microscopy of ApDNV (the prototype and only dinovernavirus) showed that purified virions have a structure similar to core particles of turreted reoviruses (Fig. 1(b)), with particular similarities to the single-shelled cypoviruses, with no outer capsid layer. The mean diameter of the virus particle is ~49–50 nm, with a central space (for the viral genome) that is 36–37 nm in diameter. Turrets were clearly visible projecting from the particle surface.

The infectivity of ApDNV-1 particles is destroyed by treatment with 0.1% SDS but unaffected by 1% deoxycholate or repeated treatments with Freon 113 or Vertrel-XF. Freezing at –20 or –80°C, or heating to 50–60°C for 30 min abolished ApDNV infectivity, even in presence of 50% fetal bovine serum (FBS). However, infectivity was stable for up to 3 weeks at room temperature and for at least 5 months at 4°C. ApDNV is stable at pH 6–8, but infectivity was reduced tenfold at pH 4–5 or pH 9–10. The virion morphology (observed by electron microscopy) becomes distorted below pH 5 and is completely disrupted below pH 3.5.

ApDNV persistently infects AP61 cells without any visible cytopathic effects and has been detected in these cells from several different sources. The number of copies of the ApDNV genome in each persistently infected AP61 cell was estimated by quantitative polymerase chain reaction (PCR) as between 1 and 5, although treatment with 2-aminopurine (2-AM) can increase the number to 60–80 genome copies per cell. The virus can also infect and replicate in mosquito cells, including AA23 and C6/36 (from *Aedes albopictus*) cells and A20 cells (from *Aedes aegypti*), but not in AE cells (*A. aegypti*) or Aw-albus (*Aedes w-albus*). ApDNV failed to replicate in mammalian cells or in mice.

Unlike other reoviruses, the genome of ApDNV contains only 9 segments of dsRNA, with a 4/1/3 AGE migration profile (Fig. 3). The aggregate genome size is 23,147 bps and the 9 dsRNA segments range from 1.15 to 3.82 kb in size. The G + C content is 33.95% compared to 34.8% for fijiviruses, 44.7% for oryzaviruses, 43% for the cypoviruses, and 63% for idnoreoviruses. The termini of ApDNV contain conserved terminal sequences (Table 3). The first three nucleotides “AGU” are conserved between ApDNV, some fijiviruses, and some cypoviruses; though, unlike the cypoviruses and fijiviruses, the first and last nucleotides of the ApDNV genome segments are complementary (A and U). The mean G + C content of the ApDNV genome is 34%.

Table 5 Coding assignments for the ApDNV-1 genome segments

Segment	Size (bp)	Protein	Function by similarity
Seg-1	3817	VP1	Possible cell attachment
Seg-2	3752	VP2	RdRp
Seg-3	3732	VP3	T2
Seg-4	3375	VP4	Nonstructural
Seg-5	3227	VP5	Possible capping enzyme
Seg-6	1775	VP6	Possible NTPase
Seg-7	1171	VP7	Nonstructural
Seg-8	1151	VP8	Possible translational regulation
Seg-9	1147	VP9	Viral inclusion bodies

Note: More information is available at: http://www.iah.bbsrc.ac.uk/dsRNA_virus_proteins/Dinovernavirus.htm.

The putative functions of ApDNV proteins VP1–VP9 (based on sequence similarities to other reovirus proteins) are shown in **Table 5**. These comparisons suggest that there is no equivalent to the cypovirus polyhedrin gene and appear to confirm the status of ApDNV as a distinct and authentic nine-segmented dsRNA reovirus.

Antigenic and Genetic Relationships

Cypoviruses

Cypoviruses are currently classified within 16 species that were initially identified by their distinctive dsRNA migration patterns. Cross-hybridization analysis of the dsRNA, serological comparisons of the viral proteins, and comparison of RNA sequences confirmed the validity of this classification and have identified further tentative species within the genus. However, only relatively few cypovirus isolates have been characterized, suggesting that there may be many more species yet to be identified.

Virus isolates within a single cypovirus electropherotype exhibit high levels of antigenic cross-reaction, as well as efficient cross-hybridization of their denatured genomic dsRNA, even under high-stringency conditions. In contrast, there is little to no evidence of serological cross-reaction between viruses from different electropherotypes (representing different cypovirus species). However, CPV-1 and CPV-12 are exceptions, showing a low but significant level of serological cross-reaction. These viruses also show some overall similarity in their electropherotype pattern and detectable levels of cross-hybridization of their genomic RNA. CPV-14 also shows some similarity in its RNA electropherotype pattern to both CPV-1 and CPV-12, suggesting that it may also show some antigenic relationship and RNA sequence homology with these viruses.

The nomenclature currently used to identify different cypovirus isolates takes account of both the virus species and the host species from which the virus was originally isolated (e.g., an isolate of CPV-1 from *Bombyx mori* would be identified as BmCPV-1). The sequence data that are available for isolates of CPV-1, -2, -5, -14, -15, and -16 allow a comparison of some genes from these viruses. There is significantly higher conservation in the largest genome segments (possibly as a result of functional constraints) although the level of variation is relatively uniform across the whole genome. Earlier cross-hybridization studies suggested that the level of nucleotide sequence variation is also relatively uniform across the whole cypovirus genome, possibly reflecting the absence of neutralizing antibodies (and therefore antibody selective pressure) in their insect hosts. Different isolates within a single cypovirus species usually show very high levels of nucleotide sequence identity. For example, different isolates of CPV-5 show >98% identity in genome Seg-10 (the polyhedrin gene), while different isolates of CPV-1 show 89%–98% nucleotide sequence identity in this gene. In contrast, comparisons of unrelated species showed only relatively low levels (20%–23%) of sequence identity (**Fig. 4**).

The level of amino acid (aa) identity in the putative viral polymerase varied from 92.9% to 99.5% within a single cypovirus type, and 42.3%–43.3% between different types (**Fig. 5**); (see **Table 6** for abbreviations and accession numbers).

Idnoreoviruses

Antigenic relations between different idnoreoviruses have not been analyzed. The phylogenetic relationship between different idnoreoviruses is also unknown (the genome of DpRV is the only idnoreovirus one sequenced to date). However, the sequence of DpRV shows homologies to other reoviruses (aa identities reaching up to 20% with fijiviruses, oryzaviruses, rotaviruses, sea-dornaviruses, and orbiviruses).

Dinovernaviruses

Sequence comparison of the *Aedes pseudoscutellaris* reovirus genome to other reoviruses has shown significant aa identities in each of the viral proteins. Members of the genera *Cypovirus*, *Oryzavirus*, and *Fijivirus*, which are also turreted, showed the highest aa identities (19%–31%). Amino acid identity values <30% between RdRps of the various reoviruses can be used to distinguish the

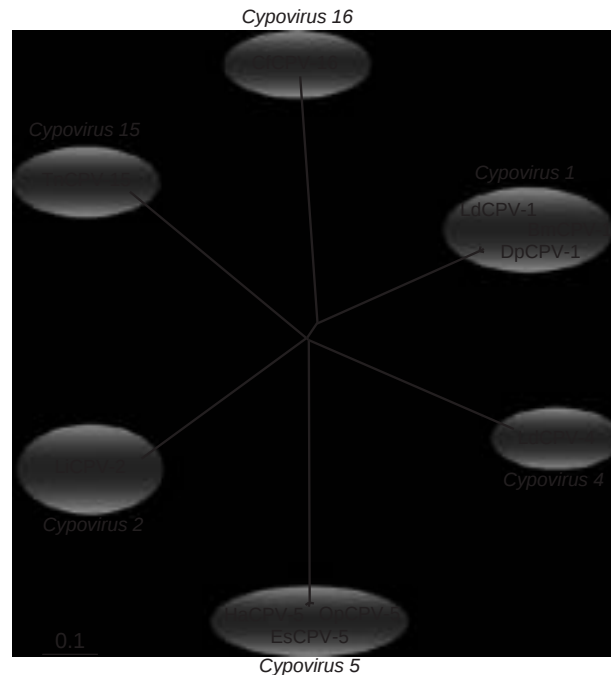


Fig. 4 Phylogenetic tree for polyhedrin proteins from isolates belonging to six *Cypovirus* species. The scale bar indicates the number of substitutions per site.

members of different genera. The relationship of ApDNV to other reoviruses is illustrated by a phylogenetic tree for the RdRp gene (Fig. 5). Although these sequence comparisons confirm the validity of ApDNV classification within a separate genus, they also indicate significant relationships with the turreted cypoviruses, fijiviruses, and oryzaviruses (aa identities of 26%, 23%, and 22%, respectively). Lower levels of aa identity (20%) were also detected with the mycoreoviruses (which are also turreted).

Sedoreovirinae

The *Sedoreovirinae* subfamily encompasses six genera, *Cordoreovirus*, *Mimoreovirus*, *Orbivirus*, *Phytoreovirus*, *Rotavirus* and *Seadornavirus*. Viruses in genus *Cardoreovirus* infect only crustaceans. Viruses in the single and type species of genus *Mimoreovirus* have 11 segments of dsRNA and infect the unicellular and ubiquitous alga *Micromonas pusilla*. Reoviruses in the genus *Orbivirus* have 10 dsRNA segments and are considered as arboviruses infecting a wide range of vertebrate (mammalian and avian) hosts and their arthropod vectors, like gnats, mosquitoes, phlebotomines and ticks. While orbiviruses infect invertebrate vectors they have little to no effect on the infected arthropod and will not be described in this article. Genus *Rotavirus* viruses have 11 dsRNA segments and infect only vertebrates and are transmitted by the fecal-oral route. *Homalodisca vitrepenis* reovirus, is the only reovirus in genus *Phytoreovirus* that infects only an insect, the glassy-winged sharpshooter and is not found in plants frequented by this pest. While most reoviruses in the genus *Seadornavirus* are considered arboviruses infecting vertebrates and their invertebrate vectors, two viruses Kadipiro virus and Liao Ning virus have so far been found only in mosquitoes. A summary of the recognized species or reoviruses in the *Sedoreovirinae* infecting only invertebrates is provided in Table 7.

Cardoreovirus

The first cardoreovirus (from *carcinus* for crab and *dodeca* for 12 dsRNA segments) was isolated in 2004 from diseased mitten crabs (*Eriocheir sinensis*). The genus has a single species, the type species *Eriocheir sinensis reovirus* with *Eriocheir sinensis reovirus* (ESRV905) as an exemplar virus member. There are at least five other possible members of the genus, *Callinectes sapidus reovirus* from blue crab, *Carcinus mediterraneus reovirus* W2 (CMRV) from shore crab, *Macrobrachium nipponense reovirus* from prawn, *Macropipus depurator reovirus* P from swimming crab and *Scylla serrata reovirus* from mud crab. The cardoreovirus icosahedral virions are about 70 nm in diameter and have a smooth non turreted surface, with an RNA containing core of about 55 nm diameter. All cardoreoviruses, except *Macrobrachium nipponense reovirus* which has 10 dsRNA segments, have a genome of 12 dsRNA segments. Replication occurs in the cytoplasm of crab cells and virions appear as rosettes around viral inclusion bodies.

Vago in 1966 described the first crustacean virus ever to be identified, Paralysis virus of *Macropipus depurator* virus, now known as the cardoreovirus *Macropipus depurator reovirus* P (the "P" stands for paralysis) which causes a slowly developing paralysis and darkening exoskeleton.

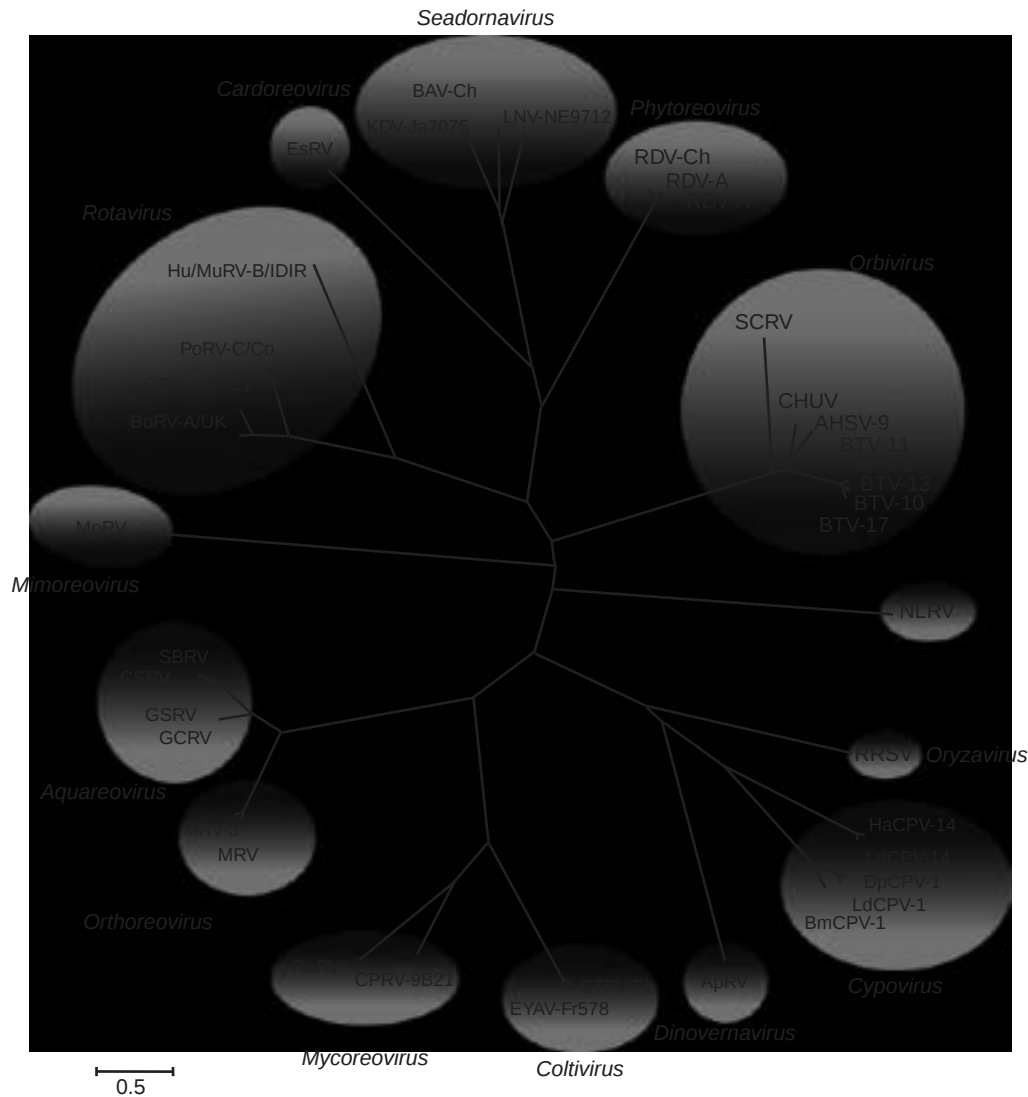


Fig. 5 Neighbor-joining phylogenetic tree, built with available polymerase sequences for representative members of 15 recognized genera of family *Reoviridae*. The abbreviations and accession numbers are those provided in [Table 6](#). The scale bar indicates the number of substitutions per site.

Carcinus mediterraneus reovirus W2 (CMRV) isolated from the hepatopancreas of the shore crab infects interstitial cells of the hepatopancreas, midgut, gills and haemocytes.

Infection of blue crabs in crab fisheries on the Atlantic coast of North and South America with *Callinectes sapidus* reovirus leads to mortalities of up to 100% in both wild and farmed populations. Diagnosis of CSRV is by in situ hybridization and RT-PCR.

Eriocheir sinensis reovirus (EsRV905) was first reported in 2004 from freshwater crab *Eriocheir sinensis*. ESRV is associated with trembling disease, including trembling of the legs, sluggishness and loss of appetite, in farmed mitten crabs in China. This is an economically important disease of freshwater cultured Chinese mitten crabs in Jiangsu Province. The virus infects connective tissue of the gills, midgut, hepatopancreas and hemocytes with mortality reaching 30% in experimental infections. The ESRV WX-2012 genome has an aggregate size of 23,913 bp with a G + C content of 43.80% and the 12 dsRNA segments range from 1.18 to 4.24 kb in size. Genome segment 1 encodes the RdRp of 1217 aa.

Scylla serrata reovirus, originally named mud crab reovirus, was first isolated from mud crab in southern China in 2007. It causes sleeping disease, a severe disease in marine-cultured mud crab in southern China. The highly pathogenic virus infects connective tissues of the hepatopancreas, gills and intestines with mortality reaching 70% in affected farms. Clinical signs include lethargy, anorexia and a gray discoloration. The virus causes vacuolar degeneration of connective tissues in the heart, stomach and intestines. During infection the presence of high numbers of virions results in crystal-like inclusion bodies in the cytoplasm. The *Scylla serrata* reovirus dsRNA genome has an aggregate size of 24,464 bp with 12 segments ranging in size from 1.16 to 4.33 kb.

Macrobrachium nipponense reovirus, was first described in 2016 in farmed freshwater prawn (oriental river prawn) *Macrobrachium nipponense*. Though there are no clinical signs of infection the virus can be associated with smaller shrimp and high levels of mortality.

Table 6 Sequences of the RNA-dependent RNA polymerases (RdRps) used in phylogenetic analysis of different reoviruses (Fig. 4)

<i>Species</i>	<i>Isolate</i>	<i>Abbreviation</i>	<i>Accession number</i>
Genus <i>Seadornavirus</i> (12 segments)			
<i>Banna virus</i>	Ch	BAV-Ch	AF168005
<i>Kadipiro virus</i>	Java-7075	KDV-Ja7075	AF133429
<i>Liao ning virus</i>	LNV-NE9712	LNV-NE9712	AY701339
Genus <i>Coltivirus</i> (12 segments)			
<i>Colorado tick fever virus</i>	Florio	CTFV-FI	AF134529
<i>Eyach virus</i>	Fr578	EYAV-Fr578	AF282467
Genus <i>Orthoreovirus</i> (10 segments)			
<i>Mammalian orthoreovirus</i>	Lang strain	MRV-1	M24734
	Jones strain	MRV-2	M31057
	Dearing strain	MRV-3	M31058
<i>Mahlapitsi orthoreovirus</i>	2511		KU198603
Genus <i>Orbivirus</i> (10 segments)			
<i>African horse sickness virus</i>	Serotype 9	AHSV-9	U94887
<i>Bluetongue virus</i>	Serotype 2	BTv-2	L20508
	Serotype 10	BTv-10	X12819
	Serotype 11	BTv-11	L20445
	Serotype 13	BTv-13	L20446
	Serotype 17	BTv-17	L20447
<i>Palyam virus</i>	Chuzan	CHUV	Baa76549
<i>St. Croix river virus</i>	SCRv	SCRv	AF133431
Genus <i>Rotavirus</i> (11 segments)			
<i>Rotavirus A</i>	Bovine strain UK	BoRV-A/UK	X55444
	Simian strain SA11	SiRV-A/SA11	AF015955
<i>Rotavirus B</i>	Human/murine strain IDIR	Hu/MuRV-B/IDIR	M97203
<i>Rotavirus C</i>	Porcine Cowden strain	PoRV-C/Co	M74216
Genus <i>Aquareovirus</i> (11 segments)			
<i>Golden shiner reovirus</i>	GSRV	GSRV	AF403399
<i>Grass carp reovirus</i>	GCRV-873	GCRV	AF260511
<i>Chum salmon reovirus</i>	CSRv	CSRv	AF418295
<i>Striped bass reovirus</i>	SBRV	SBRV	AF450318
Genus <i>Fijivirus</i> (10 segments)			
<i>Nilaparvata lugens reovirus</i>	Izumo strain	NLRV-Iz	D49693
Genus <i>Phytoreovirus</i> (10 segments)			
<i>Rice dwarf virus</i>	Isolate China	RDV-Ch	U73201
	Isolate H	RDV-H	D10222
	Isolate A	RDV-A	D90198
<i>Homalodisca vitripennis reovirus</i> (unclassified)	Fillmore		FJ497789
Genus <i>Oryzavirus</i> (10 segments)			
<i>Rice ragged stunt virus</i>	Thai strain	RRSV-Th	U66714
Genus <i>Cypovirus</i> (10 segments)			
<i>Bombyx mori cytoplasmic polyhedrosis virus 1</i>	BmCPV-1	BmCPV-1	AF323782
<i>Dendrolimus punctatus cytoplasmic polyhedrosis 1</i>	DpCPV-1	DpCPV-1	AAN46860
<i>Lymantria dispar cytoplasmic polyhedrosis 1</i>	LdCPV-1	LdCPV-1	NC_003017
<i>Lymantria dispar cytoplasmic polyhedrosis 14</i>	LdCPV-14	LdCPV-14	AAK73087
<i>Heliothis armigera cypovirus 14</i>	HaCPV-14	HaCPV-14	DQ242048

Table 6 Continued

Species	Isolate	Abbreviation	Accession number
Genus <i>Mycoreovirus</i> (11 or 12 segments)			
<i>Rosellinia anti-rot virus</i>	W370	RaRV	AB102674
<i>Cryphonectria parasitica reovirus</i>	9B21	CPRV	AY277888
Genus <i>Mimoreovirus</i> (11 segments)			
<i>Micromonas pusilla reovirus</i>	MpRV	MPRV	DQ126102
Genus <i>Dinovernavirus</i> (9 segments)			
<i>Aedes pseudoscutellaris dinovernavirus</i>	ApDNV	ApDNV	DQ087277
Genus <i>Cardoreovirus</i> (12 segments)			
<i>Eriocheir sinensis reovirus</i>	Isolate 905	EsRV	AY542965

Table 7 Sampling of *Sedoreovirinae* reoviruses infecting invertebrates, their hosts and some of their genome characteristics

Species	Virus	Host	# Segs	Aggregate genome size (bp)	Range of segment sizes (kb)	% G + C
Genus <i>Cardoreovirus</i>						
<i>Eriocheir sinensis reovirus</i>	<i>Eriocheir sinensis reovirus</i> WX-2012	Mitten crab	12	23,913	1.18–4.24	43.80
Unclassified	<i>Carcinus mediterraneus</i> W2 virus	Shore crab	12			
Unclassified	<i>Callinectes sapidus reovirus</i> SZ-2007	Blue crab	12			
Unclassified	<i>Macrobrachium nipponense reovirus</i>	Prawn	10			
Unclassified	<i>Macropipus depurator reovirus</i> P	Swimming crab	12			
Unclassified	<i>Scylla serrata reovirus</i>	Mud crab	12	24,464	1.16–4.33	
Genus <i>Phytoreovirus</i>						
Unclassified	<i>Homalodisca vitripennis reovirus</i> -Fillmore	Glassy winged sharpshooter ^a	12	25,724	1.04–4.47	38.70
Genus <i>Seadornavirus</i>						
<i>Banavirus</i>	Banna virus strain JKT-6423	mosquitoes	12	20,682	0.86–3.75	
<i>Kadipirovirus</i> ^b	Kadipiro virus JKT-7075	mosquitoes	12	20,985	0.77–3.77	37.68
<i>Liao ning virus</i>	Liao ning virus LNV-NE9712	mosquitoes	12	20,739	0.76–3.74	41.36

^aAlthough a phytoreovirus, *Homalodisca vitripennis reovirus* has been detected solely in the glassy winged sharpshooter, it has not been detected in any plants (e.g. citrus trees) where the insect is commonly found. Other phytoreoviruses are known to be plant viruses vectored by invertebrates.

^bWhile considered arboviruses replicating in vertebrate hosts and their invertebrate vectors, the *Seadornavirus* Kadipiro virus has been isolated solely from mosquitoes.

Phytoreovirus

Phytoreoviruses (from “phyto” Greek for plant) were initially discovered in 1941 as wound tumor viruses of plants transmitted by leaf hoppers in which the virus also replicates. The phytoreovirus icosahedral virions appear to be double shelled with a diameter of about 70 nm diameter. While phytoreoviruses are generally considered as plant viruses vectored by cicadellid leafhoppers, they induce no marked disease in the leafhopper vector. However, one unclassified reovirus, *Homalodisca vitripennis reovirus* (HoVRV) in the genus *Phytoreovirus* appears to have only an insect host, the hemipteran glassy winged sharpshooter. This virus was first described in 2009 and found to be closely related to Rice dwarf virus. It appears to be maintained in the absence of a plant host. The genome of *Homalodisca vitripennis reovirus* has an aggregate size of 25,724 bp with a G + C content of 38.70% and the 12 ds RNA segments range in size from 1.04 to 4.47 kb.

Seadornavirus

The Banna virus, in the genus *Seadornavirus* (from Southeast Asian dodeca for 12 dsRNA segments), was first discovered in 1987 in a patient with encephalitis and fever in China and is an arbovirus transmitted by *Anopheles* and *Culex* mosquitoes. However, two seadornaviruses have been found only in mosquitoes, Kadipiro virus in *Culex* and Liao ning virus in *Aedes dorsalis*. Seadornavirus icosahedral virions have a well-defined dual shell surface capsomeric structure with a diameter of 60–70 nm and a 40–50 nm core. Kadipiro virus was first discovered following growth in mosquito C6/36 cells of material collected from *Culex fusccephalus* from Java, Indonesia in 1981. Since then it has been isolated from other mosquitoes, particularly from China. The Kadipiro virus

genome has an aggregate size of 20,985 bp with a G + C content of 37.68% and the 12 dsRNA segments range in size from 0.77 to 3.77 kb. Liao ning virus, another seadornavirus, was initially isolated from *Aedes dorsalis* mosquitoes in North East China in 1997. The Liao ning virus genome has an aggregate size of 20,739 bp with a G + C content of 41.36% and the 12 dsRNA segments range in size from 0.76 to 3.74 kb.

See also: Reoviruses (*Reoviridae*) and Their Structural Relatives

Further Reading

- Attoui, H., Jaafar, M.F., Belhouchet, M., *et al.*, 2005. Expansion of family *Reoviridae* to include nine-segmented dsRNA viruses: Isolation and characterization of a new virus designated *Aedes pseudoscutellaris* reovirus assigned to a proposed genus (Dinovernavirus). *Virology* 343, 212–223.
- Attoui, H., Mertens, P.P.C., Becnel, J., *et al.*, 2012. Reoviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press, pp. 541–637.
- Bateman, K.S., Stentiford, G.D., 2017. A taxonomic review of viruses infecting crustaceans with an emphasis on wild hosts. *Journal of Invertebrate Pathology* 14, 86–110.
- Kibenge, F.S.B., Godoy, M.G., 2016. Reoviruses of aquatic organisms. In: Kibenge, S.B., Godoy, M.G. (Eds.), *Aquatic Virology*. New York: Academic Press, pp. 205–236.
- Mertens, P.P.C., Attoui, H., 2008. Insect reoviruses. In: Mahy, B.W.J., van Regenmortel, M.H.V. (Eds.), *Encyclopedia of Virology*, third ed. , pp. 450–459.
- Payne, C.C., Mertens, P.P.C., 1983. Cytoplasmic polyhedrosis viruses. In: Joklik, W.K. (Ed.), *The Reoviridae*. New York: Plenum, pp. 425–504.
- Plus, N., Gissman, L., Veyrunes, J.C., Pfister, H., Gateff, E., 1981a. Reoviruses of drosophila and ceratitis populations and of drosophila cell lines: A new genus of the *Reoviridae* family. *Annales de Virologie* 132E, 261–270.
- Plus, N., Veyrunes, J.C., Cavalloro, R., 1981b. Endogenous viruses of *Ceratitis capitata* Wied. 'J.R.C. Ispra strain' and *C. capitata* permanent cell lines. *Annales de Virologie* 132E, 91–100.
- Varma, M.G., Pudney, M., Leake, C.J., 1974. Cell lines from larvae of *Aedes (Stegomyia) malayensis* Colless and *Aedes (S) pseudoscutellaris* (Theobald) and their infection with some arboviruses. *Transactions of the Royal Society for Tropical Medicine and Hygiene* 68, 374–382.
- Zhou, Y., Qin, T., Xiao, Y., *et al.*, 2014. Genomic and biological characterization of a new cypovirus isolated from *Dendrolimus punctatus*. *PLoS One* 9, e113201. doi:10.1371/journal.pone.0113201.

Rhabdoviruses of Insects (*Rhabdoviridae*)

Andrea González-González, Nicole T de Stefano, David A Rosenbaum, and Marta L Wayne, University of Florida, Gainesville, FL, United States

© 2021 Elsevier Ltd. All rights reserved.

Classification

Rhabdoviruses are negative sense, single stranded RNA viruses that infect a wide range of hosts including mammals, birds, fish, reptiles, plants, and arthropods, and have a considerable impact on crop production and public health. The family *Rhabdoviridae* belongs to the order *Mononegavirales* and is classified as a member of group V from the Baltimore system based on its method of mRNA synthesis. The *Rhabdoviridae* are composed of 20 genera and 144 defined species as well as many unclassified isolates. Each genus corresponds to a monophyletic group in a phylogenetic reconstruction inferred using a maximum likelihood algorithm and the full-length sequence of the large L gene. The L gene is a reliable molecular marker for taxonomic classification due both to its low rate of recombination and because its high level of conservation, with some regions involved in functions such as transcription and replication. Besides sharing a common evolutionary history, members of each genus have similar genomic structure, transmission mode and host range.

Insect rhabdoviruses can be divided into two groups: one whose members infect vertebrate hosts through arthropod vectors (i.e., arboviruses), and one whose members infect solely insects. Members of the genera *Curiovirus*, *Ephemerovirus*, *Hapavirus*, *Ledantavirus*, *Sripuvirus*, *Tibrovirus*, *Vesiculovirus* and *Caligrhavirus* have been isolated from arthropod vectors and can also infect vertebrate hosts. In contrast, the genera *Almendravirus* and *Sigmavirus* comprise viruses capable of infecting only arthropods. A phylogenetic analysis using the conserved regions of the L protein led to the suggestion of new genera tentatively called *Bahivavirus* and *Sawgravivirus* (not yet approved by ICTV) from arthropod hosts; it is not yet known whether they can infect non-arthropod hosts.

Additional novel rhabdoviruses have been identified in insect hosts using next-generation sequencing technologies. Some of these are still in the process of being classified into particular genera, such as the Withyham and Lye Green viruses isolated from *Drosophila obscura*, Merida virus isolated from mosquitoes (MERDV), Apis rhabdovirus-1 (ARV-1) and Apis rhabdovirus-2 (ARV-2) isolated from bees, bumble bees and their mites. Again, the breadth of host range for these viruses remains to be determined.

Almendravirus and *Sigmavirus* are the largest genera whose members infect only insects. So far, five different species are officially classified as *Almendravirus* and seven as *Sigmavirus* (Table 1). The *Almendravirus* species are isolated from mosquitoes in the Americas and the *Sigmavirus* species from different species of flies collected around the world (Spain, USA, Kenya, France, Ghana and the UK). Recent metatranscriptomic studies have suggested the existence of more *sigmavirus* species (Table 1) that remain unclassified.

Table 1 Member species of invertebrate Rhabdoviruses^a

Genus	Species	Virus name abbreviation	Host
<i>Sigmavirus</i>	<i>Drosophila affinis sigmavirus</i>	DAffSV	Fruit fly
	<i>Drosophila ananassae sigmavirus</i>	DAnaSV	Fruit fly
	<i>Drosophila immigrans sigmavirus</i>	DImmSV	Fruit fly
	<i>Drosophila melanogaster sigmavirus</i> (Type species)	DMelSV	Fruit fly
	<i>Drosophila obscura sigmavirus</i>	DObssV	Fruit fly
	<i>Drosophila tristis sigmavirus</i>	DTrisV	Fruit fly
	<i>Drosophila stabulans sigmavirus</i>	MStaSV	Fruit fly
	^b <i>Drosophila sturtevanti sigmavirus</i>	DStuSV	Fruit fly
	^b <i>Drosophila busckii sigmavirus</i>	DBusSV	Fruit fly
	^b <i>Drosophila algonquin sigmavirus</i>	DAlgSV	Fruit fly
	^b <i>Drosophila montana sigmavirus</i>	DMonSV	Fruit fly
	^b <i>Scaptodrosophila deflexa sigmavirus</i>	SDefSV	Fruit fly
	^b <i>Ceratitis capitata sigmavirus</i>	CCapSV	Mediterranean fruit fly
	^b <i>Pararge aegeria rhabdovirus</i>	PAerRV	Speckled wood butterfly
<i>Almendravirus</i>	<i>Arboretum almdravirus</i>	ABTV	<i>Psorophora albigena</i> mosquitoes
	<i>Balsa almdravirus</i>	BALV	<i>Culex erraticus</i> mosquitoes
	<i>Coot Bay almdravirus</i>	CBV	<i>Anopheles quadrimaculatus</i> mosquitoes
	<i>Puerto Almendras almdravirus</i> (Type species)	PTAMV	<i>Ochlerotatus fulvus</i> mosquitoes
	<i>Rio Chico almdravirus</i>	RCHV	<i>Anopheles triannulatus</i> mosquitoes

^aTable elaborated using information from Longdon *et al.* (2015), Walker *et al.* (2018) and from the ICTV Taxonomy Report at: https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/rhabdoviridae.

^bNewly discovered, unclassified *Sigmavirus* species. Membership in the *Sigmavirus* genus is implied by isolation from CO₂ sensitive flies using RNA-Seq.

Virion Structure

As other members of the family *Rhabdoviridae*, virions from the *Sigmavirus* and *Almendraviruses* genera are bullet shaped and enveloped. Sigmavirus particles have an approximate diameter of 80 nm and a length of 100 nm, while Almendraviruses particles have a 40–55 nm diameter and 190–460 nm length. The outer surface of the virions have spikes made by glycoprotein that cross the lipidic envelope. The inside of the virion is composed of the nucleocapsid consisting of one single molecule of negative sense single-stranded RNA attached to a nucleoprotein, RNA-dependent RNA polymerase and a phosphoprotein. In addition there is also a matrix protein that links the nucleocapsid to the membrane glycoproteins.

Genome

Rhabdoviruses have negative sense, single stranded, unipartite RNA genomes 11–16 kb length consisting of five canonical structural genes that encode (3′–5′) the nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and large protein (L) (Fig. 1). The nucleoprotein constitutes the major component of the viral nucleocapsid and binds the template bases to the polymerase. The phosphoprotein interacts physically with L ensuring it is properly placed on the N-RNA template. It also secures the correct synthesis of N. The matrix protein confers the inner structural component of the virion and is thought to regulate genomic RNA transcription. The glycoprotein binds to host cell receptors and induces endocytosis of the virus. The large protein is a component of the nucleocapsid. It is involved in functions related to transcription and replication, like RNA-dependent RNA polymerase (RdRp) 5′ capping and 3′ polyadenylating. Each gene is separated by short untranscribed intergenic sequences and has conserved transcription stop and start sequences of about 10 nt in length. The genome of some rhabdoviruses

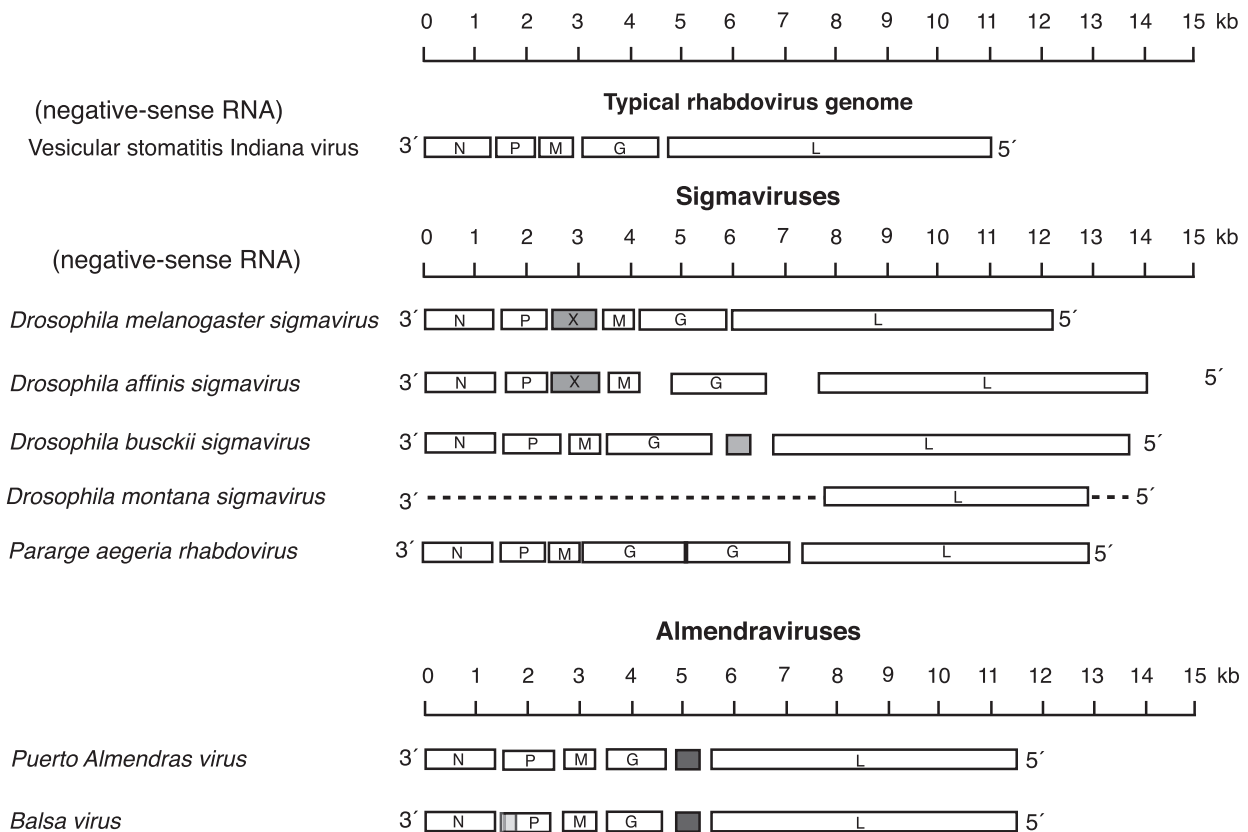


Fig. 1 Genome organization of different insect rhabdoviruses. In contrast to the typical rhabdovirus genome structure consisting on N, P, M, G and L ORFs, the typical *Sigmavirus* genomes encode the X protein whose function is unknown. *Drosophila busckii* sigmavirus encodes an extra protein between G and L with unknown function as well. Almendraviruses genomes do not code for the X protein but encode a class 1a viroporin-like protein ORF which is located between G and L (filled rectangle). Empty spaces between ORFs corresponds to untranscribed intergenic sequence regions. The broken line indicates genome regions not sequenced. Modified from Dietzgen, R.G., Kondo, H., Goodin, M.M., Kurath, G., Vasilakis, N., 2017. The family Rhabdoviridae: Mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Research* 227, 158–170. Longdon, B., Murray, G.G.R., Palmer, W.J., *et al.*, 2015. The evolution, diversity, and host associations of rhabdoviruses. *Virus evolution* 1, 1–12. Walker, P.J., Blasdel, K.R., Calisher, C.H., *et al.*, 2018. ICTV Virus Taxonomy Profile: Rhabdoviridae. *Journal of General Virology* 99, 447–448.

includes additional genes located between the structural proteins, therefore genomes of viruses from different genera may vary in length and organization.

Sigmavirus Genome

Members of the *Sigmavirus* genus have genomes that range from 12.4 to 14.5 kb and typically encode six proteins. The N, P, M, G and L proteins have the same structural characteristics and are homologous to proteins of other rhabdoviruses. The X gene is located between P and M, and codes for the protein PP3 (also called X protein) which is presumably involved in host-virus interactions. Different genome sizes are explained by the variation in length of the untranscribed intergenic sequences. For example, there is a 664 nucleotide untranscribed region between G and L in the G *Drosophila affinis sigmavirus* (DAffSV) (Fig. 1).

Some of the newly discovered related and unclassified sigmaviruses isolated from different species of flies share the canonical genome structure of six genes present in *Drosophila melanogaster sigmavirus* (DMelSV) while others deviate from the prototypic genome organization. For example, the *Drosophila busckii* sigmavirus (DBusSV) genome does not code for the X gene however, it has an accessory ORF of unknown function between the G and L genes. Pararge aegeria rhabdovirus (PAerRV) does not code for the X protein, and the G protein is split into two contiguous ORFs. Other sigmaviruses have been identified from RNAseq data including *Drosophila montana* sigmavirus (DMonSV), *Drosophila sturtevantii* sigmavirus (DStuSV) and *Scaptodrosophila deflexa* sigmavirus (SDefSV); however, their genomes are not yet fully sequenced (Fig. 1).

Almendravirus Genome

Members of the *Almendravirus* genus have genomes that range from 10.5 to 11.5 kb. They also encode the canonical five proteins present in genomes of members of the *Rhabdoviridae* family, with the same structural characteristics and homology of the proteins of other rhabdoviruses. However, *Almendravirus* encode an additional class 1a viroporin-like protein as an independent transcriptional unit between the G and L genes. The Balsa virus (BALV) genome, has an extra ORF (Px) within the P gene which function is unknown (Fig. 1).

Other Insect Rhabdovirus Genomes

The recently discovered *Apis mellifera* Rhabdoviruses 1 (ARV-1) and 2 (ARV-2) have genomes ~14.5 kb long. They both have the prototypic rhabdovirus genome structure consisting of the five open reading frames N, P, M, G and L. In contrast, the genome of other unclassified rhabdoviruses, isolated from bedbugs (Shuangao Bedbug Virus 2), mosquitoes (Wuhan Mosquito Virus 9) and flies (Shayang Fly virus 3), codes only for the G and L proteins plus three different capsid proteins (VP1, VP2 and VP3).

Life Cycle

Sigma virus has the same basic steps as other members of other rhabdoviruses for cell entry, transcription, translation, replication, and assembly, all of which occur predominately in the cytoplasm. Rhabdoviruses have five essential proteins that aid in these processes: nucleoprotein (N), glycoprotein (G), RNA-dependent RNA polymerase large subunit (L), phosphoprotein (P), and matrix protein (M). Initially, the G protein aids in the attachment of the virus to the host cell membrane. Entry of the virus into the host cell occurs by receptor mediated endocytosis, utilizing clathrin coated pits. The low pH levels in the environment facilitate fusion of the viral envelope and the host cell membrane by causing a conformational change of the G protein which subsequently releases the nucleocapsid into the cytoplasm where the virus can undergo transcription.

Transcription of the viral RNA is initiated by the detachment of the M protein from the nucleocapsid. Primary transcription of the negative-sense RNA is catalyzed by the viral transcriptase enzyme and follows a stop-start pattern starting at the 3' end of the RNA. Each mRNA transcript has an initiation sequence and is terminated by a polyadenylation signal, a string of A's at the end of the sequence. The essential genes are transcribed as capped poly-adenylated mRNAs in the order N → P → M → G → L. Small RNAs are also produced that are uncapped and positive sense, typically considered leader RNAs. Once transcribed, the mRNA sequences are translated into proteins and these proteins are transported to varying areas of the cell. Protein translation of N, M, P, and L occur in cytoplasmic polysomes. However, in the case of the G protein, it is translated on membrane-bound polysomes, and then transported from the endoplasmic reticulum to the Golgi apparatus for processing and finally to the plasma membrane.

After a sufficient amount of protein is produced, genome replication is initiated. This focuses mainly on viral encapsidation, where proteins N and P are vital components in this process. The synthesized positive sense antigenome provides a template for formation of the negative sense RNA to be encapsidated. Once initial RNA replication occurs, secondary transcription, translation, and replication can take place. After formation of the nucleocapsid, it attaches to the cell membrane and the synthesized M protein aids in the budding process where virions are released and are able to proliferate the infection to adjacent cells.

Integration Into the Host Genome

SIGMAV-like and other rhabdovirus-like elements homologous to genes coding for protein L (RNA-dependent RNA polymerase), N (Nucleoprotein), P (polymerase-associated protein) and G (glycoprotein) are integrated into the genomes of a range of different *Drosophila* species as well as the genomes of the mosquito *Aedes aegypti* and tick *Ixodes scapularis*. There is evidence that suggests that the sequence encoding the viral protein P was integrated into the *D. yakuba* genome by retrotransposon template-switching, and that it is expressed by the host.

Epidemiology

Distribution, Spatial and Temporal

Rhabdoviruses have been found in worldwide distributed natural populations of insects from the Drosophilidae, Tephritidae, Nymphalidae, Culicidae and, Apidae families.

Transmission

The most studied rhabdovirus in insects is sigma virus, which has been isolated from different *Drosophila* species ([Table 1](#)). DmelSV (sigma virus isolated from *D. melanogaster*), DAffSV (sigma virus isolated from *D. affinis*) and DObsSV (sigma virus isolated from *D. obscura*) are vertically transmitted by both parents to their offspring, a condition thought to ensure the persistence of vertically transmitted virus in natural populations despite the fitness cost such viruses can impose on infected flies.

In the wild, transmission efficiency of DmelSV and DAffSV by females is higher compared to males. In contrast, DObsSV is transmitted at similar rates by both parental sexes. Additionally, if a fly is infected by its female parent, the probability of transmitting DmelSV to its offspring is higher than when it is infected by its male parent meaning that while there can be patrilineal transmission for multiple generations, the transmission efficiency decreases with each generation. In contrast, *D. affinis* and *D. obscura* males infected by their male parents do not transmit the virus to their offspring. Conveniently, although sigma virus is vertically transmitted in nature, sigma virus can be experimentally injected into naïve female flies where it can replicate and be transmitted to the next generation by these flies.

Prevalence

D. melanogaster females tend to have slightly higher prevalence of sigma virus than males in the field and this prevalence varies from season to season in natural populations ranging from 28% to 6.2% depending on the year. This prevalence may be influenced by a latitudinal or climatic gradient (lower latitudes having higher prevalence compared to higher altitudes). In *D. obscura*, the presence of DObsSV ranges from 22% to 73% depending on the collecting site in the UK.

Laboratory experiments have demonstrated that DmelSV, DAffSV and DObsSV can persist and replicate in closely related novel hosts.

Population Dynamics

Sigma virus isolated from different natural populations of *D. melanogaster* across Europe shows almost null genetic diversity and zero levels of recombination. At the same time, this diversity is structured since viral isolates from a particular population share a unique pool of genetic variants compared to other populations. Like DmelSV, DObsSV has very low genetic variation. However, in DObsSV these genetic variants are distributed homogeneously among different geographical populations. Using mathematical modeling, it has been demonstrated that DObsSV invades natural populations of *D. obscura* flies at a fast pace, dramatically increasing its population size. At the single fly level, infections by multiple strains of DmelSV have been reported by measuring their plaque size and incubation times to CO₂ sensitivity. Sequence variation consistent with multiple infections has also been detected.

Clinical Features

Infection with sigma virus decreases female fecundity in natural populations of *D. melanogaster*. Also, eggs laid by uninfected females had greater viability and faster developing times than those laid by infected females in laboratory population.

Pathogenesis

The paralysis experienced by *D. melanogaster* flies after being exposed to CO₂ has been attributed to the damage caused to the nervous tissue by the replication of sigma virus in the fly thoracic ganglia. It also infects other tissues like the eyes, brain and other

nerves. However, the exact mechanism remains unknown. Although not directly studied in the *D. melanogaster*-sigma virus system, it can be hypothesized that the protein G plays an important role in this CO₂ sensitivity. In other members of the Rhabdovirus family such as rabies, the viral protein G attaches to different host cell receptors, allowing the entrance of the virus into nervous cells and mediates the spread of the virus among neurons.

Diagnosis

Sigma virus-infected flies die or become permanently paralyzed after being exposed to high levels of CO₂ and this CO₂ sensitivity is correlated with the viral titers. This symptom is commonly used to screen flies in natural and laboratory populations. Adult flies are exposed to CO₂ for 5–15 min at 12°C and after 30 min, infected flies rarely recover. While death after exposure with CO₂ is very strongly associated with sigma virus infection in *D. melanogaster*, in other *Drosophila* species like *D. obscura*, some infected flies are not sensitive to CO₂. Another way to test for sigma virus presence is by quantitative real-time PCR (qRT-PCR) or by first-strand synthesis followed by quantitative PCR.

Further Reading

- Ammar, E.-D., Tsai, C.-W., Whitfield, A.E., Redinbaugh, M.G., Hogenhout, S.A., 2008. Cellular and molecular aspects of Rhabdovirus interactions with insect and plant hosts. *Annual Review of Entomology* 54, 447–468.
- Brun, G., 1963. Étude d'une association du virus σ et de son hôte la Drosophile: L'état stabilisé. These, Paris-Orsay, p. 254.
- Carpenter, J.A., Obbard, D.J., Maside, X., Jiggins, F.M., 2007. The recent spread of a vertically transmitted virus through populations of *Drosophila melanogaster*. *Molecular Ecology* 16, 3947–3954.
- Contreras, M.A., Eastwood, G., Guzman, H., *et al.*, 2017. Almendraviruses: A proposed new genus of rhabdoviruses isolated from mosquitoes in tropical regions of the Americas. *The American Journal of Tropical Medicine and Hygiene* 96, 100–109.
- Dietzgen, R.G., Kondo, H., Goodin, M.M., Kurath, G., Vasilakis, N., 2017. The family Rhabdoviridae: Mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Research* 227, 158–170.
- Fleuriet, A., 1976. Presence of the hereditary rhabdovirus sigma and polymorphism for a gene for resistance to this virus in natural populations of *Drosophila melanogaster*. *Evolution* 30, 735–739.
- Fleuriet, A., 1988. Maintenance of a hereditary virus, the Sigma virus in populations of its host, *Drosophila melanogaster*. *Evolutionary Biology* 23, 1–30.
- Levin, S., Galbraith, D., Sela, N., *et al.*, 2017. Presence of Apis Rhabdovirus-1 in populations of pollinators and their parasites from two continents. *Frontiers in Microbiology* 8, 2482.
- L'Héritier, P.L., Teissier, G., 1945. Transmission héréditaire de la sensibilité au gaz carbonique chez la Drosophile. *Pubis Lab. Ecole Normale Supérieure Biologie* vol. 1, pp. 35–76.
- Longdon, B., Murray, G.G.R., Palmer, W.J., *et al.*, 2015. The evolution, diversity, and host associations of rhabdoviruses. *Virus evolution* 1, 1–12.
- Rittschof, C.C., Pattanaik, S., Johnson, L., *et al.*, 2013. Sigma virus and male reproductive success in *Drosophila melanogaster*. *Behavioral Ecology and Sociobiology* 67, 529–540.
- Tsai, C.W., McGraw, E.A., Ammar, E.D., Dietzgen, R.G., Hogenhout, S.A., 2008. *Drosophila melanogaster* mounts a unique immune response to the rhabdovirus sigma virus. *Applied and Environmental Microbiology* 74, 3251–3256.
- Walker, P.J., Blasdel, K.R., Calisher, C.H., *et al.*, 2018. ICTV Virus Taxonomy Profile: Rhabdoviridae. *Journal of General Virology* 99, 447–448.
- Walker, P.J., Firth, C., Widen, S.G., *et al.*, 2015. Evolution of genome size and complexity in the Rhabdoviridae. *PLOS Pathogens* 11. (e1004664).
- Wayne, M.L., Blohm, G.M., Brooks, M.E., *et al.*, 2011. The prevalence and persistence of sigma virus, a biparentally transmitted parasite of *Drosophila melanogaster*. *Evolutionary Ecology Research* 13, 323–345.

Relevant Websites

- <https://viralzone.expasy.org/7276>
Almendraviruses
ViralZone page.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/rhabdoviridae/788/genus-almendraviruses
Genus: Almendraviruses
Rhabdoviridae
Mononegavirales.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/rhabdoviridae/799/genus-sigmavirus
Genus: Sigmavirus
Rhabdoviridae
Mononegavirales.
- https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/rhabdoviridae
Rhabdoviridae
Rhabdoviridae
Mononegavirales.
- <https://viralzone.expasy.org/4656>
Sigmavirus
ViralZone page.
- <https://www.genome.jp/virushostdb>
Virus-Host Database.

Sarothroviruses (*Sarothroviridae*)

Azeez Sait Sahul Hameed, C. Abdul Hakeem College, Melvisharam, India

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Cell line It is a permanently established cell culture that will proliferate indefinitely in appropriate cell culture medium.

Pathogenicity Ability of infectious agent to cause disease.

Post-larvae One of early developmental stages in freshwater prawn after larval stages.

Satellite virus Virus dependent on another virus for replication and transcription.

Vertical transmission Transmission of virus from an infected brooder to its progeny.

Introduction

The giant freshwater prawn, *Macrobrachium rosenbergii* is the most favored crustacean species by the prawn farmers and is being cultured on a large-scale in many countries. Disease is one of the important factors which limit the aquaculture production worldwide. Like other animals, prawns are affected by viruses, bacteria, fungi, and metazoan parasites. Very few diseases have been reported in freshwater prawn when compared to marine shrimp because *M. rosenbergii* is a moderately disease-resistant species. Among the pathogens, prawn viruses are very important and responsible for huge economic losses, particularly in the hatchery and nursery phases. Reported prawn viruses include *Macrobrachium* hepatopancreatic parvo-like virus (MHPV), *Macrobrachium* muscle virus (MMV), infectious hypodermal and hematopoietic necrosis virus (IHHNV) and white spot syndrome virus (WSSV). MHPV, MMV, and IHHNV have been reported only once in freshwater prawn (Anderson *et al.*, 1990; Tung *et al.*, 1999; Hsieh *et al.*, 2006). *M. rosenbergii* is not a natural host of WSSV, but the syndrome was experimentally induced in this species via different routes of infection. A new viral disease namely white tail disease caused by *Macrobrachium rosenbergii* nodavirus and *Macrobrachium* satellite virus 1 was reported to cause severe mortality in freshwater prawn hatcheries and nursery ponds in different parts of world.

Geographical Distribution

This virus was first reported in the French West Indies later in China, Chinese Taipei, Thailand, Australia, and Malaysia. The occurrence of WTD is being reported in other prawn growing countries.

Clinical Signs of WTD and Histopathology

The clinical signs of WTD in infected post-larvae include lethargy and opaqueness of the abdominal muscle that gradually extends from the center to the anterior and the posterior sections of the muscle. Degeneration of the telson and uropods is observed in severe cases. The prevalence of opaque and milky post-larvae increases dramatically 1–2 days later and changes are particularly obvious in the abdomen (tail). The mortality rate reaches a maximum at five days after the first observation of gross signs. The most affected tissues in WTD-infected post-larvae are striated muscles in the abdomen and cephalothorax and the connective tissues of all organs. There are basophilic cytoplasmic inclusions with a diameter of 1–40 µm in striated muscles of the abdomen, cephalothorax, and intratubular connective tissue of the hepatopancreas. No viral inclusions were observed in epithelial cells of the hepatopancreatic tubules or in midgut mucosal epithelial cells. Muscles exhibit multifocal areas of hyaline necrosis of the fibers, with moderate edema.

Morphology of Virus

Extra small virus particles are non-enveloped, icosahedral and about 15 nm in diameter with a density in CsCl of 1.325 g/ml and serologically unrelated to those of *Macrobrachium rosenbergii* nodavirus (Fig. 1). Two overlapping coat proteins of about 17 kDa (CP-17) and 16 kDa (CP-16) are found in extra small virus particles. Since the particles are very small, this viral agent has been called extra small virus (XSV).

Genome Organization

The genome of XSV is a linear positive-sense RNA molecule and contains 796 nucleotides with a single ORF located between nucleotides 63 and 587, and a short poly (A) tail of 15–20 nucleotides at the 3'-end. A potential polyadenylation signal (AAUAAA)

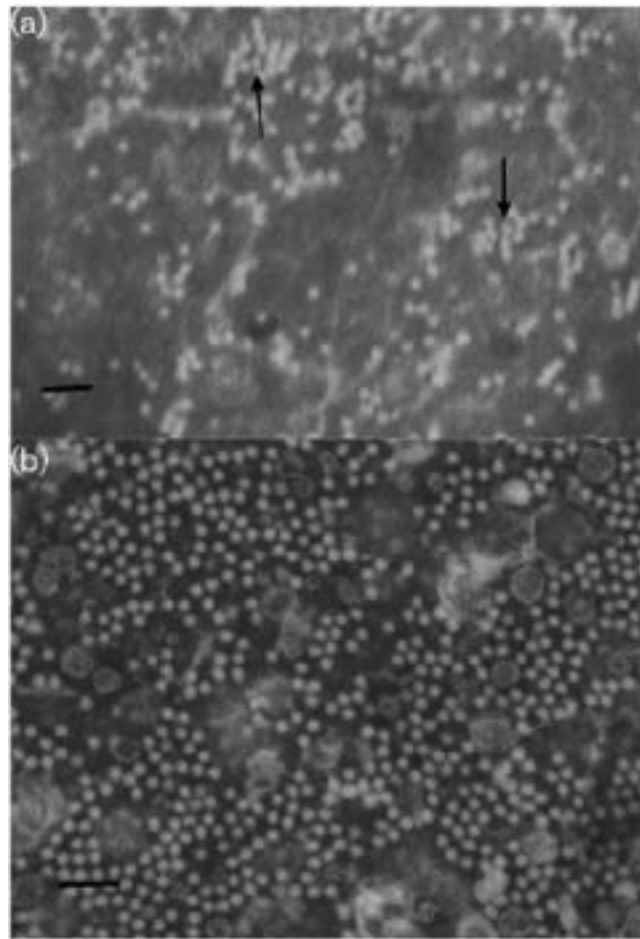


Fig. 1 Negatively stained purified (a) extra small virus (bar = 50 nm) and (b) *Macrobrachium rosenbergii* nodavirus (bar = 100 nm).

is located 6 nt upstream of the poly(A). The ORF encodes two capsid proteins namely CP16 and CP17 (**Fig. 2**). Amino acid sequence analysis indicates that CP16 is synthesized by internal initiation at a second methionine codon 11 amino acid codons downstream of the first initiation codon. The two forms of the capsid protein are synthesized in an equimolar ratio.

Taxonomic Position

The genome sequence of XSV does not have any significant homology with known virus genomes. The genome of XSV encodes only the capsid proteins and this virus depends on the RdRp of MrNV for replication and transcription. This virus meets the primary conditions of satellite viruses and appears more similar to the known satellite viruses infecting plants (e.g., tobacco necrosis satellite virus-like) than those infecting insects (e.g., chronic bee paralysis-associated satellite virus). XSV is a virus member of the single species, *Macrobrachium satellite virus 1* in the single genus, *Macronovirus* belonging to the family *Sarothroviridae*. Isolates from different geographical locations, including French West Indies, Thailand, Taiwan, China and India, have been reported. All isolates belong to different geographical locations showed 96%–99% nucleotide sequence identity in their capsid protein genes but the encoded proteins are not recognizably similar to proteins in the public sequence databases.

Pathogenicity and Transmission of Disease

Mixtures of both MrNV nodavirus and XSV sarothroviruses caused 100% mortality in post-larvae 12 days post infection (d p.i.) by immersion and the results indicate the possibility of horizontal transmission of disease. In the virus-infected group, post-larvae started showing whitish muscles at 7 d p.i. and reached the highest proportion at 12 d p.i. MrNV and XSV were purified and post-larvae were challenged with different combinations of these two viruses by immersion ([Zhang et al., 2006](#)). Clinical signs of WTD and high mortality were observed in post-larvae challenged with combinations containing a relatively high proportion of MrNV and low proportion of XSV. In contrast, there was little sign of WTD and low mortality in post-larvae challenged with a higher proportion of

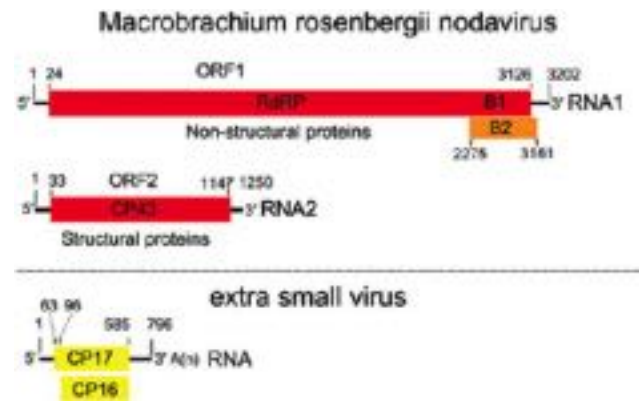


Fig. 2 Genome organization of MrNV and XSV (*Sarothroviridae*). RdRP: RNA-dependent RNA polymerase, CP: capsid protein. Numbers indicate nucleotide positions.

XSV than MrNV, indicating that MrNV plays a key role in causing disease in *M. rosenbergii*. RT-PCR analysis confirmed the presence of both viruses in experimentally infected post-larvae and in gill tissue, head muscle, stomach, intestine, heart, hemolymph, pleopods, ovaries, and tail muscles of experimentally injected adult prawns. Previous studies revealed that MrNV and XSV viruses were always found to be associated with WTD (Zhang *et al.*, 2006). Despite these studies, the role played by XSV in the disease and particularly in its pathogenicity remains unclear and needs further research. Experimental studies confirmed vertical transmission from infected broodstock to post-larvae.

XSV affects mainly post-larvae of *M. rosenbergii* and is responsible for severe mortalities in the early life stages of freshwater prawn. The vertical mode is the main route of WTD transmission. Experimental infection of brooders with both MrNV and XSV indicated that the vertical route is the main route of disease transmission. In the infected brooders, ovarian tissue and fertilized eggs were found to be positive for MrNV/XSV as evidenced by RT-PCR and mortality of up to 100% was observed in post-larvae hatched from eggs of infected brooders. Some virus-infected post-larvae can survive to adulthood and can act as carriers of the disease. *Artemia* which plays an important role in nutrition of shellfish at the larval stages, may also serve as a vector for MrNV/XSV. The results of experimental infection and RT-PCR assay revealed that both MrNV and XSV responsible for WTD, are possibly transmitted to post-larvae of freshwater prawn via the oral route.

Host Range

The early life stages such as larvae, post-larvae and early juvenile of *M. rosenbergii* were found to be susceptible to XSV. No mortality was observed either in naturally or experimentally (MrNV/XSV) infected subadult and adult prawns. In addition, the PCR results showed XSV-positive in kuruma prawn (*Penaeus japonicus*), Indian white prawn (*Penaeus indicus*), giant tiger prawn (*Penaeus monodon*), dragonfly (*Aeshna* sp.), giant water bug (*Belostoma* sp.), beetle (*Cybister* sp.), backswimmer (*Notonecta* sp.), hairy river prawn (*Macrobrachium rude*), monsoon river prawn (*Macrobrachium malcolmsonii*), brine shrimps (*Artemia* sp.), and red claw crayfish (*Cherax quadricarinatus*).

Susceptibility of Cell Line to XSV

XSV can be easily propagated in the C6/36 mosquito *Aedes albopictus* cell line. This cell line can be cultured in Leibovitz L-15 medium containing 100 International Units ml⁻¹ penicillin, 100 µg ml⁻¹ streptomycin and 2.5 µg ml⁻¹ fungizone supplemented with 10% fetal bovine serum at 28°C. The C6/36 cell line was found to be useful for propagation of XSV and viral replication was confirmed by RT-PCR, acridine orange staining, infectivity studies and electron microscopy. A specific cytopathic effect was not observed in XSV-infected cell lines, but multiple vacuolations were observed.

Diagnostic Tools

XSV can be detected by genome-based and protein-based methods. Various diagnostic methods including histopathology have been developed in different laboratories throughout the world to detect XSV. Other methods include immunological methods, RT-PCR, loop-mediated isothermal amplification (LAMP), in-situ hybridization, and dot blot hybridization. All the above works were aimed at detecting MrNV and XSV in the samples.

Control Measures

RT-PCR and other user-friendly kits for detection of XSV are available for commercial use. Therefore, broodstock and seed screening are strongly recommended to prevent WTD in the farms. Broodstock or seed that test positive for WTD are recommended to be discarded by proper zoo-sanitary methods. Usual sanitation and control protocols for viral infections are recommended to avoid WTD in the prawn farm.

Acknowledgment

Author thanks the Department of Biotechnology, Government of India, New Delhi, India for financial support to carry out the research on viruses of shrimp and prawn under Viral Repository Project (BT/PR12660/AAQ/3/710/2014).

References

- Anderson, I.G., Law, T., Shariff, M., Nash, G., 1990. A parvo-like virus in the giant freshwater prawn, *Macrobrachium rosenbergii*. *Journal of Invertebrate Pathology* 55, 447–449.
- Hsieh, C.Y., Chuang, P.C., Chen, L.C., *et al.*, 2006. Infectious hypodermal and haematopoietic necrosis virus (IHHNV) infections in giant freshwater prawn, *Macrobrachium rosenbergii*. *Aquaculture* 258, 73–79.
- Tung, C.W., Wang, C.S., Chen, S.N., 1999. Histological and electron microscopic study on *Macrobrachium* muscle virus (MMV) infection in the giant freshwater prawn, *Macrobrachium rosenbergii* (de Man), cultured in Taiwan. *Journal of Fish Diseases* 22, 319–323.
- Zhang, H., Wang, J., Yuan, J., *et al.*, 2006. Quantitative relationship of two viruses (MV and XSV) in white tail disease of *Macrobrachium rosenbergii* de Man. *Diseases of Aquatic Organisms* 71, 11–17.

Relevant Website

https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/sarothviridae
 Sarothviridae.
 Positive-sense RNA Viruses.
 ICTV.

Soliniviruses (*Soliniviridae*)

Steven M Valles, Center for Medical, Agricultural and Veterinary Entomology, Agricultural Research Service, US Department of Agriculture, Gainesville, FL, United States

Andrew E Firth, University of Cambridge, Cambridge, United Kingdom

Published by Elsevier Ltd.

Glossary

Biological Control (= biocontrol) A method to control insect pests using organisms that are natural predators, parasites, or pathogens.

Biopesticide (= biological pesticide) A pesticide whose active ingredient is biologically based, including predators, parasites, or pathogens.

Entomopathogen A pathogen of insects.

Ribosomal Frameshifting Method of translation exhibited by some viruses to fuse proteins encoded by two overlapping open reading frames.

Trophallaxis In social insects, the exchange of food or other nutrients between colony members by mouth-to-mouth or mouth-to-anus feeding.

Introduction

Soliniviridae is a new family of viruses that possess non-segmented, single-stranded, positive-sense RNA genomes of 10–11 kilobases. This is a highly divergent group exhibiting phylogenetic, genome, and replication characteristics with aspects that are both picornavirus- and calicivirus-like. Among the two classified species of the family, gene order appears conserved. However, whereas the genome of *Nylanderia fulva* virus 1 (NfV-1) contains a single large open reading frame, in the genome of *Solenopsis invicta* virus 3 (SINV-3) this is split into two large open reading frames with the second open reading frame being accessed by a ribosomal frameshifting mechanism, resulting in the translation of a trans-frame fusion polyprotein. In both species, the non-structural proteins involved in replication (picorna-like helicase, protease and RNA-dependent RNA polymerase) are encoded towards the 5' end of the genome, and the capsid proteins (a jelly-roll fold domain, an extension domain, and an additional capsid-associated protein) are encoded towards the 3' end. The capsid proteins may be expressed from genomic or subgenomic RNAs. SINV-3 is the best characterized of the two species and infects the invasive red imported fire ant (*Solenopsis invicta* Buren); NfV-1 infects the tawny crazy ant (*Nylanderia fulva* (Mayr)). Although SINV-3 and NfV-1 infect ant species (Insecta: Formicidae), additional unclassified related sequences have been identified in various other arthropods.

Classification (Compact)

The *Soliniviridae* name is derived from the type species, *Solenopsis invicta* virus 3. The family currently comprises two genera, each currently containing a single species, *Invictavirus* (derived from *Solenopsis invicta* virus 3) and *Nyfulvavirus* (derived from *Nylanderia fulva* virus 1). Comparative analyses of protein complement, gene expression, and genome organization of the two members suggest that the family may form a sister group to the *Caliciviridae*. However, this conclusion will require additional sequence discoveries and analyses of related viruses to resolve the taxonomic position of the group.

The *Soliniviridae* exhibit characteristics that are picornavirus-like and calicivirus-like. Similar to both picornaviruses and caliciviruses, the *Soliniviridae* encode a superfamily III helicase, a 3C-like chymotrypsin-related cysteine protease and a superfamily I RNA-dependent RNA polymerase. Additional calicivirus-like characteristics exhibited by soliniviruses include a single jelly-roll fold capsid domain, capsid protein encoding in the 3' end of the genome, production of a capsid-encoding subgenomic RNA (empirically established in SINV-3 only), similar nucleotide sequences at or near the genomic and subgenomic RNA 5'-termini, and the presence of a capsid projection domain. Another calicivirus characteristic exhibited by SINV-3 is its environmental stability and resistance to heat inactivation. SINV-3 remains viable for years within ant corpses.

Currently, the genomes of three SINV-3 isolates have been sequenced in entirety, including the DM isolate from North American *Solenopsis invicta* (GenBank Accession FJ528584), the SF isolate from South American *Solenopsis invicta* (GU017972), and the Hybrid isolate from *Solenopsis invicta* x *richteri* hybrid ants (MF797911). The isolates exhibit highly similar amino acid sequences (>98% identical) with the majority of nucleotide differences being synonymous. The genome of NfV-1 (isolate Florida initial) has also been sequenced in entirety (KX024775).

Virion Structure

Electron microscopy of SINV-3 revealed spherical particles with irregularly spaced surface projections, cup-shaped depressions, and a diameter of 27.3 ± 1.3 nm (Fig. 1). NfV-1 particles were also spherical with irregularly spaced projections, and a diameter of 28.7 ± 1.1 nm. This morphology was also reported in related unclassified viruses, including Kelp fly virus and *Riptortus pedestris* virus 1. The irregularly spaced surface projections appear to originate at the five-fold axes of symmetry in Kelp fly virus.

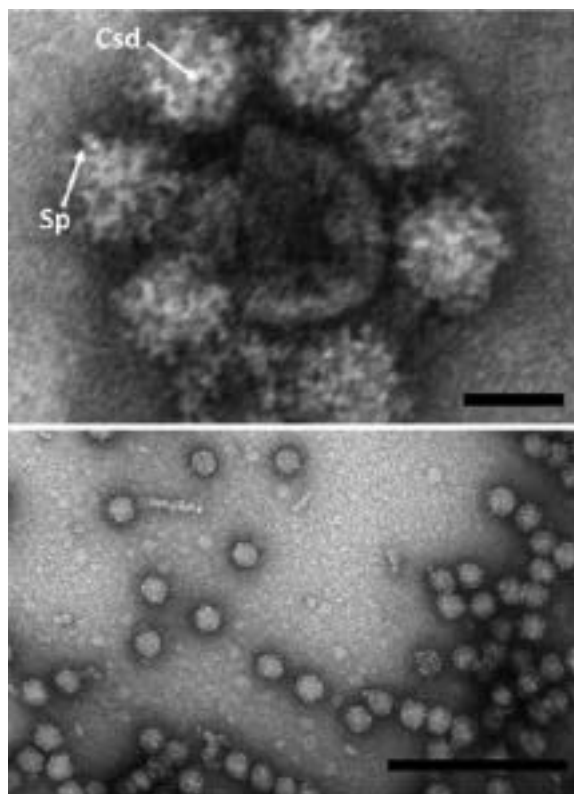


Fig. 1 Electron micrographs of a negative stain of a *Solenopsis invicta* virus 3 preparation purified by cesium chloride isopycnic centrifugation. Surface projections (Sp) and cup-shaped depressions (Csd) are evident. Bar size represents 25 nm and 200 nm in the upper and lower panels, respectively.

SINV-3 virions contain two major proteins, viral protein 1 (VP1) and VP1 fused to the capsid projection domain (VP1-CPD) (**Fig. 2**). VP1 is a single jelly-roll capsid protein and is the major component of the virion; particles are presumed to contain 180 copies of VP1 in a $T = 3$ icosahedral shell arrangement. The capsid projection domains of VP1-CPD proteins are thought to form the irregularly spaced projections. The VP1: VP1-CPD ratio is estimated at approximately 5:1 in SINV-3. A third protein, viral protein 2 (VP2) encoded downstream of CPD is detected at only sub-stoichiometric levels in SINV-3 virions. It is presumed that VP2 is proteolytically cleaved from the CPD region but this has not been determined empirically. VP2 is detected at very low levels, and its function or contribution to the virion structure or its formation are not yet known.

For NfV-1, the virion proteins may be slightly different because the genome contains only one long open reading frame. In contrast to SINV-3, no frameshift separates the VP1 and CPD coding sequences (**Fig. 2**) and it is presumed that all VP1 and CPD polypeptides are produced as the single linked form (VP1-CPD). It is possible that this VP1-CPD polypeptide might be subsequently cleaved to produce free VP1. This latter notion is supported by evidence from the related, unclassified Kelp fly virus whose genome, like that of NfV-1, also contains only a single open reading frame. Two proteins were identified in virions of Kelp fly virus in a molar ratio of 5:1. These proteins may correspond to VP1 and VP1-CPD, respectively.

Genome

Members of the family *Soliniviridae* have single-stranded, positive-sense RNA genomes of 10–11 kb with polyadenylated 3' tails. The genome of NfV-1 contains a single large open reading frame. Conversely, the genome of SINV-3 contains a large 5'-proximal open reading frame (spanning approximately $\frac{3}{4}$ of the genome), and a second smaller 3'-proximal open reading frame accessed via a -1 ribosomal frameshift at the 3' terminus of open reading frame 1. Thus, open reading frame 1 can be translated alone, or in conjunction with open reading frame 2 as a trans-frame fusion polyprotein.

Regardless of species, the protein order is similar. ORFs for structural proteins follow the non-structural proteins. Superfamily III helicase, 3C-like chymotrypsin-related protease and superfamily I RNA-dependent RNA polymerase domains are encoded towards the 5' end and followed by a jelly-roll fold domain (VP1), the capsid projection domain (CPD), and the additional capsid-associated protein (VP2) towards the 3' end. Although not confirmed empirically, a viral protein of the genome (VPg) is thought to be encoded upstream of the 3C-like protease and covalently attached to the 5' end of the genome. A predicted dsRNA binding domain with homology to the 1A suppressor of silencing protein of the dicistrovirus, *Drosophila C virus*, and perhaps acting as a viral suppressor of RNA interference, is encoded between the RNA-dependent RNA polymerase and

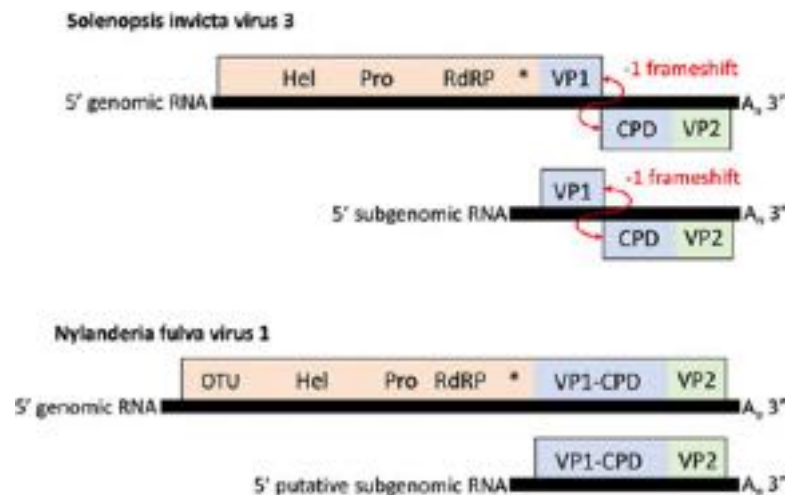


Fig. 2 Genome organization of *Solenopsis invicta* virus 3 (*Invictavirus*) and *Nylanderia fulva* virus 1 (*Nyfulvavirus*). Predicted protein domains and their relative positions for the genomic and subgenomic RNA include Helicase (Hel), Protease (Pro), RNA dependent RNA polymerase (RdRP), Jelly-roll capsid viral protein 1 (VP1), Capsid projection domain (CPD), Viral protein 2 (VP2), dsRNA binding protein (*), and Ovarian tumor domain (OTU).

jelly-roll fold domains in both SINV-3 and NfV-1. NfV-1 also encodes a predicted Ovarian Tumor (OTU) domain upstream of the helicase.

In SINV-3, and presumably NfV-1, the capsid proteins may be expressed from both genomic and subgenomic RNAs, which is a calicivirus-like characteristic. A 24 nucleotide perfect repeat (AAAGUCCAGUAAGGUUACUGGCAU) occurs in the untranslated region at the 5' terminus (genome position 18–41) and just upstream of the sequence encoding the dsRNA binding domain (genome position 6664–6687) in SINV-3. This sequence repeat may be a promoter sequence for genomic and subgenomic RNA production. Indeed, the 36 nucleotide sequence at genome position 6652 through 6687 exhibits strong similarity with the sequence of the 5'-terminal 41 nucleotides of the genome. Thus, the start position for the subgenomic RNA likely occurs at genome position 6652, with translation starting at the AUG at genome position 6800–6802. Consistently, northern analysis revealed a subgenomic RNA species of approximately 3800 nucleotides. Based on analogy with the SINV-3 sequence, the potential start site for an NfV-1 subgenomic RNA occurs at genome position 7312 as similar sequence elements (75% identity) are detected at the 5' end of the genome and at the sequence beginning with nucleotide 7312.

Life Cycle

Solinviviruses are thought to follow a replication cycle similar to picorna/caliciviruses. Once the virus gains cell entry it replicates in the cytoplasm and no DNA stage is formed. The viral genome serves a dual purpose as transcript for polyprotein synthesis and template for production of the complementary minus strand (i.e., replicative strand) of the genome, which, in turn, serves as template for production of the infectious plus strand genome. Viral assembly ultimately occurs by encapsidation of a plus strand genome.

SINV-3 and NfV-1 gain entry into the host through the alimentary canal by ingestion. Because the alimentary canal is the primary means of acquiring SINV-3 and NfV-1 infections, virus transmission into their ant hosts is easily accomplished by feeding. After the virus infects an individual ant worker, it is spread intra-colonially by trophallaxis. SINV-3 and NfV-1 exhibit stage-dependent tropism. SINV-3 replication is detected in only adult worker ants. Conversely, the replicative form of NfV-1 appears limited to the larval ant stage.

Epidemiology (Host Specificity/Prevalence/Transmission)

Establishment of host specificity is of paramount importance to establish the suitability for development and ultimate release of an entomopathogen for control purposes. If the entomopathogen (e.g., virus) exhibits broad host specificity, it is possible that non-target insect hosts could be negatively and irreparably impacted.

Host specificity tests have been conducted extensively for SINV-3. Passive and high-dose active exposure of 23 ant species in 14 genera and 4 subfamilies revealed that SINV-3 was capable of replication in only *Solenopsis invicta*, *Solenopsis richteri*, and *Solenopsis invicta* x *richteri* hybrids. Thus, SINV-3 is host specific for ant species within the *saevissima* complex of South American fire ants. Studies also revealed that SINV-3 could not replicate in the honey bee (*Apis mellifera*). The narrow host specificity exhibited by SINV-3 supported intentional release into the *Solenopsis invicta* United States population as a classical biological control agent. Furthermore, SINV-3 was detected naturally occurring in *Solenopsis invicta* ants from several U.S.

states. However, the incidence of SINV-3 was rather low and irregular. Population surveillance studies detected SINV-3 in *Solenopsis invicta* from the native range (Formosa, Argentina), the United States, and parts of the Caribbean. Because the incidence in the United States was low and irregular, intentional releases of SINV-3 have been completed in California and Florida with successful establishment of the virus. Interestingly, SINV-3 exhibits a seasonal phenology. The prevalence of the virus is higher during fall and winter months in Florida. Analysis of the relationship between temperature and prevalence showed a strong negative correlation between SINV-3 presence and warmer temperature. Temperature can have a significant impact on viral replication, assembly, or host resistance.

Based on laboratory studies, SINV-3 transmission occurs through ingestion of virus. Midgut epithelial cells are thought to be the specific target for the virus. After the infection is established in midgut cells the virus spreads to additional tissues by an unknown path. The mode of natural transmission of SINV-3 in *Solenopsis invicta* is not known. However, the virus was successfully transmitted by crickets that had consumed dead, SINV-3-infected ant workers, which had subsequently fallen prey to uninfected *Solenopsis invicta* worker colonies. The transmission was simply mechanical because crickets did not support replication of SINV-3. This mode of transmission facilitates use of SINV-3 as a biopesticide. Indeed, baits have been developed that successfully transmit SINV-3 to fire ant colonies in the laboratory and field.

Perfunctory host specificity tests with NfV-1 suggest that the virus is specific for *Nylanderia fulva* ants. However, more extensive testing will be required to firmly establish the host specificity of the virus.

Clinical Features

Not applicable.

Pathogenesis

The sequelae exhibited by SINV-3 infection of *Solenopsis invicta* begins with a cessation of food retrieval by the worker caste, followed by larval, pupal, and worker mortality and decreased queen fecundity. Very large midden piles composed of brood and worker corpses develop (Fig. 3(A)). Interestingly, in the laboratory the midden piles are created directly on the food source (Fig. 3(B)). SINV-3 titer in moribund worker ants typically exceeds 10^9 SINV-3 genome copies per ant.

Queen weight declines dramatically in the latter stages of SINV-3 infections in fire ant colonies. Decreases in developing ova and reduced fecundity are also associated with infection. It is not known whether the decrease in feeding causes the fecundity/ovary changes observed, or if SINV-3 has a direct impact on the queen. Despite significant larval mortality in colonies infected with SINV-3, the virus does not appear to replicate or assemble in the larval ant stage. Plus-strand SINV-3 is regularly detected in larval stages by RT-PCR, but virus replication or assembly do not appear to occur in this stage. It has been hypothesized that cessation of food retrieval/feeding and/or behavioral changes whereby the workers neglect the developing larvae is the cause of larval mortality. Reduced foraging in response to virus infection has been reported to occur in other arthropod species. NfV-1 is also associated with a decrease in fecundity in *Nylanderia fulva* colonies, although this response occurs inconsistently.

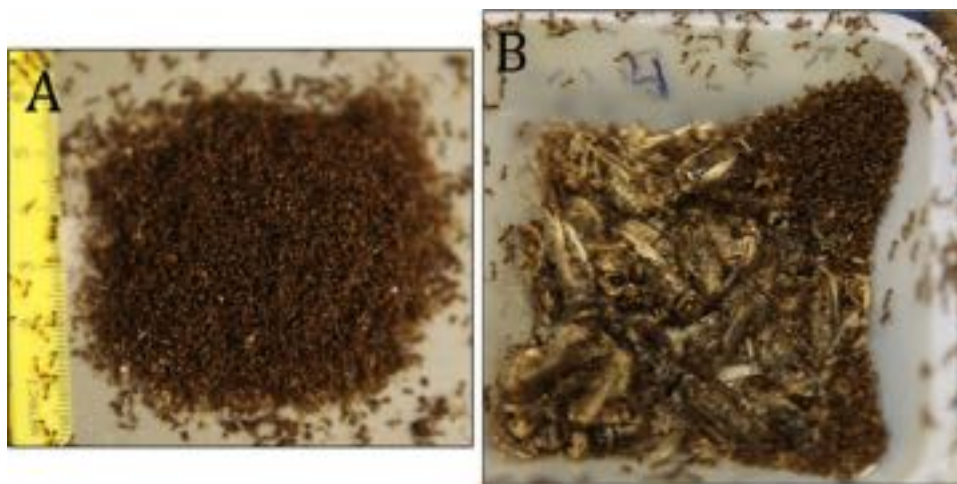


Fig. 3 Large midden pile composed of accumulating corpses from a *Solenopsis invicta* virus 3-infected fire ant colony (A). Often, the midden piles are created on the food source (B). In this case, the food source is the house cricket, *Acheta domesticus*.

Diagnosis (Detection)

Detection of colonies infected with solinviviruses is typically made by molecular analysis of a portion of the genome using RT-PCR or quantitative PCR. In the laboratory, SINV-3 infections may be identified by declining colony health, changes in food consumption and distribution, and increased mortality (midden piles). However, in the field, RT-PCR is required. NfV-1 is best detected by RT-PCR, in the laboratory or field.

Prevention (Biocontrol)

SINV-3 and NfV-1 were discovered using a metagenomics approach coupled with next generation sequencing to detect pathogens within the transcriptomes of pest ant species. The sole purpose of the work was pathogen bioprospecting and discovery. SINV-3 and NfV-1 are being investigated for their use as natural control agents (biological control) against the invasive ant species, *Solenopsis invicta* and *Nylanderia fulva*, respectively. Therefore, from an entomological perspective, introduction and establishment of these viruses into target ant host populations is the objective, rather than disease prevention.

SINV-3 has been released as a classical biological control agent into areas where the virus is not found in *Solenopsis invicta*. In addition, SINV-3 is being developed as a biopesticide. SINV-3-containing baits have been developed and successfully transmit the virus in the laboratory and field. Protein-based baits were superior to sugar- or oil-based baits. The minimum SINV-3 dose required for causing colony collapse was in the 10^7 to 10^9 viral particles per microliter range. Lower doses resulted in a chronic infection with limited impact on colony health. A limitation to commercial development of SINV-3 as a biopesticide is in vitro production of virus. Virus genome production was possible using a baculovirus-driven expression system, but virus assembly failed to occur.

Further Reading

- Adams, M.J., Lefkowitz, E.J., King, A.M.Q., *et al.*, 2017. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses. *Archives of Virology* 162, 2505–2538.
- Brown, K., Olendraitte, I., Valles, S.M., *et al.*, 2019. ICTV virus taxonomy profile: Solinviridae. *Journal of General Virology* 100, 736–737. doi:10.1099/jgv.0.001242.
- Hartley, C.J., Greenwood, D.R., Gilbert, R.J.C., *et al.*, 2005. Kelp fly virus: A novel group of insect picorna-like viruses as defined by genome sequence analysis and a distinctive virion structure. *Journal of Virology* 79, 13385–13398.
- Oi, D.H., Valles, S.M., Porter, S.D., *et al.*, 2018. Introduction of fire ant biological control agents into the Coachella Valley of California. *Florida Entomologist* 102, 284–286.
- Porter, S.D., Valles, S.M., Oi, D.H., 2013. Host specificity and colony impacts of *Solenopsis invicta* virus 3. *Journal of Invertebrate Pathology* 114, 1–6.
- Valles, S.M., 2012. Positive-strand RNA viruses infecting the red imported fire ant, *Solenopsis invicta*. *Psyche* 2012, 1–14.
- Valles, S.M., Bell, S., Firth, A.E., 2014. *Solenopsis invicta* virus 3: Mapping of structural proteins, ribosomal frameshifting, and similarities to *Acyrtosiphon pisum* virus and *Kelp fly virus*. *PLoS One* 9, e93497.
- Valles, S.M., Hashimoto, Y., 2009. Isolation and characterization of *Solenopsis invicta* virus 3, a new positive-strand RNA virus infecting the red imported fire ant, *Solenopsis invicta*. *Virology* 388, 354–361.
- Valles, S.M., Oi, D.H., Becnel, J.J., *et al.*, 2016. Isolation and characterization of *Nylanderia fulva* virus 1, a positive-sense, single-stranded RNA virus infecting the tawny crazy ant, *Nylanderia fulva*. *Virology* 496, 244–254.
- Valles, S.M., Oi, D.H., Porter, S.D., 2010. Seasonal variation and the co-occurrence of four pathogens and a group of parasites among monogyne and polygyne fire ant colonies. *Biological Control* 54, 342–348.
- Valles, S.M., Oi, D.H., Yu, F., Tan, X.X., Buss, E.A., 2012. Metatranscriptomics and pyrosequencing facilitate discovery of potential viral natural enemies of the invasive Caribbean crazy ant, *Nylanderia pubens*. *PLoS One* 7, e31828.
- Valles, S.M., Porter, S.D., 2015. Dose response of red imported fire ant colonies to *Solenopsis invicta* virus 3. *Archives of Virology* 160, 2407–2413.
- Valles, S.M., Porter, S.D., Choi, M.Y., Oi, D.H., 2013. Successful transmission of *Solenopsis invicta* virus 3 to *Solenopsis invicta* fire ant colonies in oil, sugar, and cricket bait formulations. *Journal of Invertebrate Pathology* 113, 198–204.
- Valles, S.M., Porter, S.D., Firth, A.E., 2014. *Solenopsis invicta* virus 3: Pathogenesis and stage specificity in red imported fire ants. *Virology* 461, 66–71.
- Valles, S.M., Varone, L., Ramirez, L., Briano, J., 2009. Multiplex detection of *Solenopsis invicta* viruses — 1, — 2, and — 3. *Journal of Virological Methods* 162, 276–279.

Relevant Websites

- https://talk.ictvonline.org/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/solinviridae
 Solinviridae.
 Positive-sense RNA Viruses.
 ICTV.
- https://viralzone.expasy.org/7936?outline=all_by_species
 Solinviridae.
 ViralZone page.

Tetraviruses (*Alphatetraviridae*, *Carmotetraviridae*, *Permutotetraviridae*)

Rosemary A Dorrington, Rhodes University, Grahamstown, South Africa

Tatiana Domitrovic, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Meesbah Jiwaji, Rhodes University, Grahamstown, South Africa

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Tetraviruses are a diverse group of small, non-enveloped RNA viruses that infect lepidopteran (butterflies and moths) insects. In general, virus particles vary in size from 38 to 41 nm in diameter and share a unique $T = 4$ icosahedral capsid architecture. It is this feature that led to the name “tetravirus”, derived from the Greek *tettares* meaning “four”. Tetravirus particles assemble as procapsids comprising 240 copies of a single capsid precursor protein (CP). During maturation, the CP undergoes autoproteolytic cleavage to form the β (58–67 kDa) and γ (6–8 kDa) subunits of the mature capsid. Tetravirus particles encapsidate one or two single-stranded (ss), positive sense (+ve) genomic RNAs. While some tetraviruses have bipartite genomes, the majority have monopartite genomes ranging from 5.3 to 6.6 kb in size and encoding up to four genes.

The first tetravirus, *Nudaurelia capensis* β virus (N β V), was isolated from diseased South African pine emperor moth (*Nudaurelia cytherea capensis*) larvae in 1968. Structural studies on N β V particles revealed a new type of virus capsid with unusual $T = 4$ icosahedral symmetry that led to N β V becoming the type-member of a new family of viruses then named the *Tetraviridae*. The *Tetraviridae* were classified into two genera according to their genome organization, where viruses with monopartite genomes were classified into the *Betatetraviridae* genus, while viruses with bipartite genomes resided in the *Omegatetraviridae* genus. Comparative analysis of available genome sequences revealed the *Tetraviridae* to be a highly diverse group of viruses, despite the structural conservation of their capsid. Phylogenetic analysis showed that the RNA-dependent RNA polymerase (RdRp) domains of tetravirus replicases can be clustered into three groups, each related to a different supergroup of RNA viral replicases. Consequently, the International Committee for the Taxonomy of Viruses reclassified the *Tetraviridae* into three new families based upon the characteristics of their replicases, namely the *Alphatetraviridae*, the *Permutotetraviridae* and the *Carmotetraviridae* (Table 1). The increasing availability of metatranscriptomic sequence data and the elucidation of the invertebrate virosphere will serve to clarify the viral diversity present in the invertebrates.

Genome Organization

Alphatetraviridae

The *Alphatetraviridae* are classified into two genera according to their genome organization with the *Betatetraviridae* containing virus species with monopartite genomes and viruses with bipartite genomes classified within the *Omegatetraviridae* genus. Currently, there are three omegatetraviruses and seven betatetraviruses species (Table 1). The genome sequence of only one betatetravirus (N β V) is available, so it is possible that other species classified in this genus could in fact be reclassified into the *Permutotetraviridae* or *Carmotetraviridae* families based upon the characteristics of their replicases.

In the omegatetravirus, *Helicoverpa armigera* stunt virus (HaSV), REP is encoded by RNA1, along with three small ORFs, p11, p15 and p8 that overlap with the 3' end of REP (Fig. 1). The HaSV REP (as do those of all sequenced alphatetraviruses) includes the three conserved functional domains, namely methyltransferase (MET), helicase (HEL) and RdRp domain in an N- to C-terminal modular arrangement, characteristic of the alpha-like virus superfamily. The three small ORFs are arranged tandemly and in an in-frame arrangement with p11 and p15 separated by a UGA stop, and UAA GGG separating p15 and p8. These ORFs are translated in HaSV-infected cells via a mechanism that is as yet unknown, but their translation products are thought to be involved in initiating the viral replication complex. The HaSV CP is encoded by the smaller RNA2. At its 5' end, RNA2 encodes a second protein, p17 that is expressed at low levels in HaSV-infected *Helicoverpa armigera* larval midgut tissue and is involved in packaging of viral RNA (Fig. 1). The viral RNAs (vRNAs) of all sequenced genomes of alphatetraviruses (HaSV), *Dendrolimus punctata* tetravirus and, N β V, have a 5' cap and terminates in a tRNA-like structures at its 3' ends that is thought to function in virus replication. The N β V genome encodes the viral REP that overlaps the CP ORF located at their 3' end and the CP is translated from a subgenomic (sg) RNA originating just upstream of the CP coding sequence (Fig. 1).

Permutotetraviridae

There are currently only two permutotetravirus species, namely *Euprosteria elaeasa* virus (EeV) and the type species, *Thosea asigna* virus (TaV). Both viruses have monopartite genomes that are highly conserved (82% and 84% identity, respectively) encoding two overlapping ORFs, with the viral REP at the 5' end and the CP at the 3' end (Fig. 1). In contrast to the alphatetraviruses, the 5' terminus of the TaV/EeV vRNA is predicted to be blocked by a viral protein genome linked, VpG, while the 3' terminus is predicted to form a conserved pseudoknot. The permutotetravirus CP is translated from a sgRNA originating just upstream of the CP ORF.

Table 1 Taxonomic classification of tetraviruses

<i>Virus spécifs</i>		Acronym	Lepidopteran host family	Geographical location	Particle diameter (nm)	GenBank accession no
Family: <i>Alphatetraviridae</i>						
Genus: <i>Omegatetravir</i>	<i>Nudaurelia capensis</i> ω virus ^a	N ω V	<i>Saturniidae</i>	South Africa	41	[S43937.1]
	<i>Helicoverpa armigera</i> stunt virus	HaSV	<i>Noctuidae</i>	Australia	38	[KX423453.1]; [L37299]
	<i>Dendrolimus punctatus</i> Tetravirus	DpTV	<i>Lasiocampidae</i>	China	40	[AY594352]; [AY594353]
Genus: <i>Betatetravir</i>						
	<i>Nudaurelia capensis</i> β virus ^a	N β V	<i>Saturniidae</i>	South Africa	38	[AF102884]
	<i>Antheraea eucalypti</i> virus ^b	AeV	<i>Saturniidae</i>	Australia	32	NA
	<i>Darna trima</i> virus	DtV	<i>Lymacodidae</i>	Malaysia	38	NA
	<i>Dasychira pudibunda</i> virus	DpV	<i>Lymantriidae</i>	UK	38	NA
	<i>Calliteara pudibunda</i> virus	CpV				
	<i>Philosamia cynthia</i> x <i>ricini</i> virus	PxV	<i>Saturniidae</i>	UK	35	NA
	<i>Pseudoplusia includens</i> virus	PiV	<i>Noctuidae</i>	USA	40	NA
	<i>Trichoplusia ni</i> virus	TnV	<i>Noctuidae</i>	USA	38	NA
Family: <i>Permutotetraviridae</i>						
Genus: <i>Alphapermutotetravir</i>	<i>Thosea asigna</i> virus ^a	TaV	<i>Lymacodidae</i>	Malaysia, USA	38	[AF282930], [AF062037]
	<i>Euprosterna elaeasa</i> virus	EeV	<i>Zygaenidae</i>	Peru	38	[AF461742]
<i>Carmotetraviridae</i>						
Genus: <i>Alphacarmotetravir</i>	<i>Providence</i> virus ^a	PrV	<i>Noctuidae</i>	USA	40	[AF062037]

^aType strain for the genus.^bSerologically indistinguishable from N β V (Grace and Mercer, 1976).

The activity of 2A-like processing site located near the N-terminus of the CP results in the production of a 17 kDa amino-terminal peptide (p17) and the 65 kDa α protein (Fig. 1). Permutotetraviruses REPs are structurally related to those of birnaviruses and the Drosophila A virus and distantly related to members of the picorna-like virus superfamily. In both viruses, the MET and HEL domains of REP are absent and the REP displays an internally-permuted C-A-B arrangement of the canonical A, B and C motifs found in palm subdomain of the RdRP, upstream of which is a putative nucleotidylation (VPg) signal.

Carmotetraviridae

Providence virus is the sole species of the *Carmotetraviridae* family. The virus has a monopartite genome with the CP ORF, located at the 3' end of the vRNA and translated from a sgRNA originating just upstream of the CP AUG. Two 2A-like processing sites are present at the amino terminus of the ORF that likely result in the translation of two small peptides of 7 and 8 kDa, along with the α protein of 68 kDa that is subsequently cleaved by maturation-dependent autoproteolytic cleavage to form the β and γ subunits of the mature virus particle (Fig. 1). The PrV REP ORF is upstream of the CP CDS and encodes a read-through stop codon that results in the translation of a smaller replication accessory protein (p40). The RdRp domain of the PrV REP is related to replicases of tombus- and umbraviruses, which are (+ve) ssRNA plant viruses in the carmovirus-like supergroup. Unlike any other tetravirus, the PrV genome encodes a third ORF (p130) at its 5' end, upstream of REP. The p130 ORF encodes a 2A-like processing site at its N-terminus would result in the translation of two products of 17 kDa and 113 kDa, respectively. The function of p130 in the PrV lifecycle is unknown.

Viral Replication

Our understanding of tetravirus replication is based on studies of HaSV (*Alphatetraviridae*) and PrV (*Carmotetraviridae*) because these two viruses infect and replicate in tissue culture cell lines. In studies on HaSV-infected *H. armigera* larvae, RNA1 was detected within three days post infection and RNA2 appeared after a further two days. This implies that RNA1 is replicated in the early stages of infection; delayed RNA2 replication results in the later expression of CP and p17 in the virus lifecycle. Immunofluorescence microscopy studies of HaSV-infected *Spodoptera frugiperda* tissue culture cells shows that virus replication occurs in punctate structures in the host cell cytoplasm. The REP is trafficked to detergent-resistant membranes derived from the secretory pathway and membrane targeting is dependent on the presence of two domains – one located at the N-terminus and the other within the C-terminal within the RdRp domain of the protein.

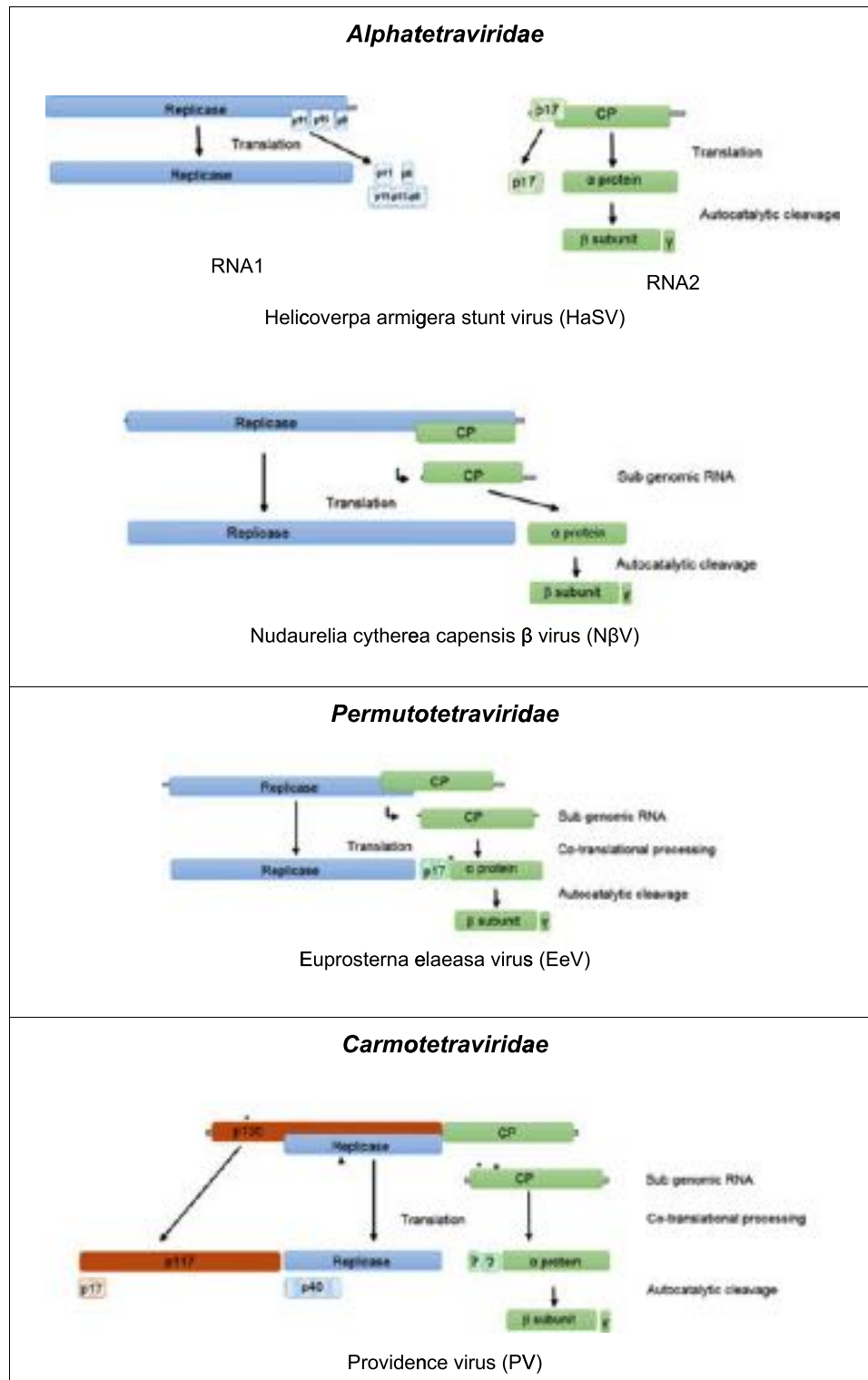


Fig. 1 Genome organization and gene expression of representative tetraviruses of the *Alphatetraviridae*, *Permutotetraviridae* and *Carmotetraviridae* families. Viral genes encoding the replicase, Capsid Protein precursor (CP) and replication accessory protein of PrV (p40) are indicated on the maps as well as gene products produced by co-translational (2A) processing in EeV and PrV as well as maturation-dependent autoproteolytic cleavage of the capsid α protein to produce the β and γ capsid subunits. Translation products labeled with (?) are predicted to be produced but have yet to be identified in virus infected cells. The positions of 2A-like processing sequences (*) and read-through stop codon ($\blacktriangle$) are indicated. The bent arrow shows the start of the sgRNAs of the monopartite viruses (N β V, EeV and PrV) from which the viral CP is translated. GenBank Accession numbers: KX423453.1 (HaSV RNA1), L37299 (HaSV RNA2), AF102884 (N β V), AF461742 (EeV) and AF062037 (PrV).

The PrV REP also localizes to detergent resistant, punctate structures in the cytoplasm of persistently infected *Helicoverpa zea* MG8 tissue culture cells. In *S. frugiperda* cells expressing EGFP-REP fusion proteins, the REP co-localizes with replicating vRNA so these punctate structures are the site of viral replication in PrV-infected cells. As with HaSV, the PrV REP appears to be associated with vesicles originating from the Golgi and secretory vesicles, thus both viruses likely share the same site of replication. The PrV REP is translated into two viral gene products; a truncated accessory replication protein (p40) resulting from the activity of a readthrough stop codon and the full length REP, which encodes the RdRp domain (Fig. 1). p40 also co-localizes with vRNA and is thought to be involved in the establishment of the viral replication complex.

The Tetravirus Capsid

Capsid Structure

X-ray crystallography and cryo-electron microscopy image reconstruction (cryoEM) have been used to examine the different structural states of tetravirus capsids. The crystal structure of authentic *Nudaurelia capensis* omega virus (N ω V) was determined in 1996–2.8 Å resolution. The virion is structurally indistinguishable from the N ω V virus-like particles (VLPs) made by expressing the 70 kDa CP in a baculovirus system. The VLP packages random cellular RNA and can be purified in a metastable procapsid state, allowing for maturation steps to be exquisitely controlled by changes in pH in a time scale that is amenable to biochemical and structural analysis. These characteristics made maturation of this alphatetravirus one of the most accessible systems for the study of eukaryotic icosahedral virus maturation and dynamics.

The N ω V crystal structure showed that the capsid is icosahedral, 420 Å in diameter, and composed of 240 capsid proteins arranged with $T = 4$ quasi-symmetry. Thus, the 60 icosahedral asymmetric units each have four subunits, named A–D, in unique chemical environments (Fig. 2(a)). The 644 amino acid capsid subunit has three domains: an exterior immunoglobulin (Ig)-like fold, a central β -barrel in tangential orientation, and an interior helix bundle (Fig. 2(a)). The helix bundle is created by the N- and C-termini of the polypeptide before and after the β -barrel fold, and the Ig-like domain is an insert between the E and F β -strands of the barrel. The exterior Ig domain is unique in nonenveloped viruses, has the least conserved sequence among tetravirus subunits and is presumed to be involved in cell binding and virus tropism.

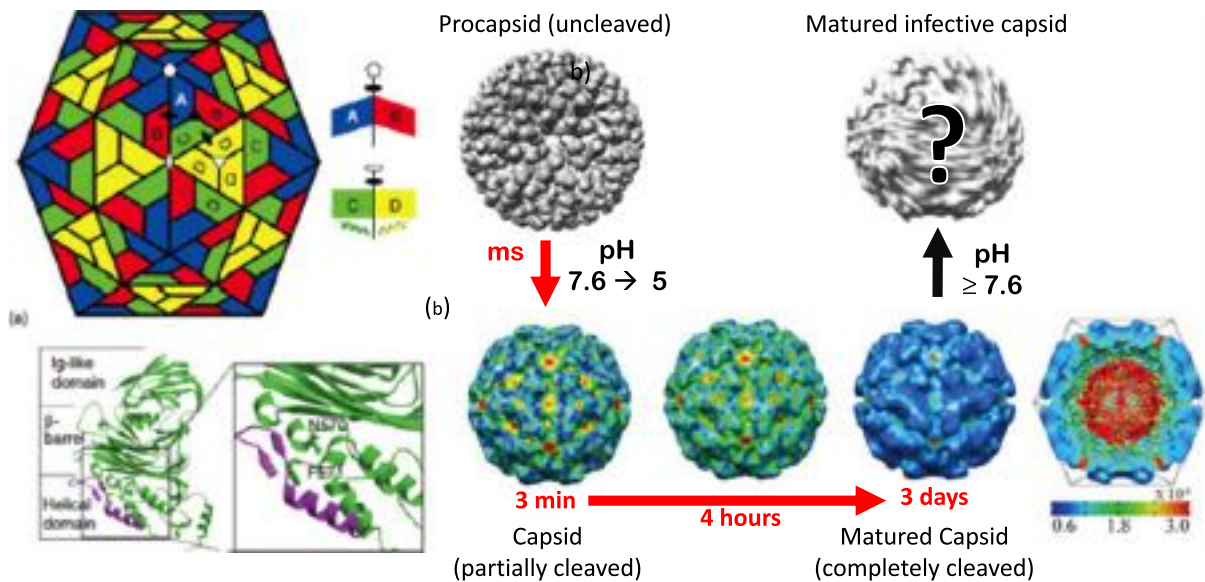


Fig. 2 Structure and dynamics of the alphatetravirus, N ω V, capsid. (a) The subunit organization in the N ω V $T=4$ capsid crystal structure. There are four subunits in the asymmetric unit (A in blue, B in red, C in green, and D in yellow). Icosahedral symmetry axes are marked with white symbols, quasi-symmetry axes are marked with black symbols. Quasi-twofold dimer interfaces occur with the bent A–B or flat C–D conformations (right). The well-ordered C-terminal helices in the C–D subunits comprise a molecular switch that stabilizes the flat interface. (b) Ribbon diagram of the C subunit illustrates the three domain tertiary structure present in all 240 subunits (exterior of capsid on top, interior on bottom). Maturation cleavage occurs within the helical domain between residues Asn₅₇₀ and Phe₅₇₁. The molecular switch, present only in the C or D subunits, is highlighted in magenta. (c) Effect of pH on N ω V maturation and activity. Red arrows indicate the maturation pathway. In the wild-type virus, the procapsid converts to uncleaved capsid with the drop of pH, followed by a maturation cleavage, which locks the particle in capsid conformation after $\sim 10\%$ of the subunits are cleaved. While the large conformational change takes milliseconds, cleavage is a slow process taking, minutes to several hours. The surface of the partially matured N ω V reconstructions are colored according to its dynamic properties as determined by standard deviation maps. Note the progressive particle stabilization promoted by cleavage. In the cross section of matured particles, is possible to visualize a dynamic region at the five-fold axis where the lytic γ peptide resides. Infectivity depends on another conformational transition promoted by high pH (black arrow), when the cleaved particle gains lytic activity. No structural model is available for this condition.

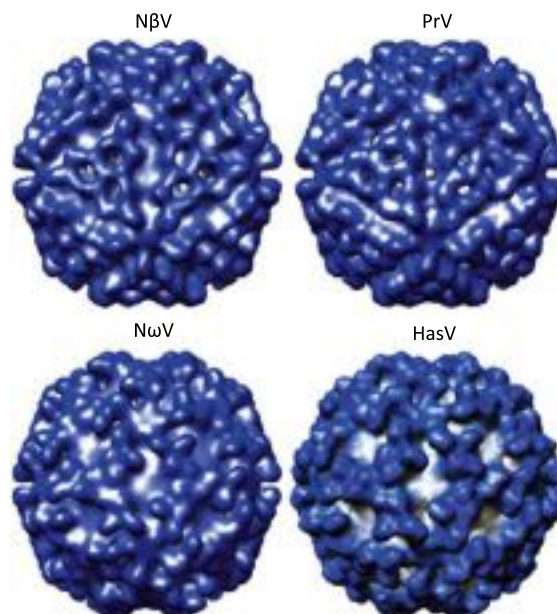


Fig. 3 Cryoelectron microscopy reconstructions of the icosahedral tetravirus virions. All four structures are at $\sim 25\text{--}30$ Å resolution and are viewed down their icosahedral twofold axes. The capsids are ~ 400 Å in diameter and have $T = 4$ quasi-icosahedral symmetry. The capsid morphologies of monopartite and bipartite viruses are closely similar (monopartite at top, bipartite at bottom), but differ between the groups. Most notably, within the triangular facets, $N\beta V$ and PrV have a pitted surface where the alphatetraviruses, $N\omega V$ and $HasV$, surfaces are filled in. The permutotetravirus, TaV , also displays the same pitted surface in negative-stain EM photographs as observed for $N\beta V$ and PrV . The recent crystal structure determination of PrV showed this is mainly due to different rotations of the Ig-like domains relative to the β -barrels when compared to the subunits in the $N\omega V$ crystal structure (see Fig. 2). This supports the suggested role of the Ig-like domain in determining the restrictive host cell specificity of these viruses by displaying distinctly different surfaces to cellular receptors.

The interior helix domain is believed to interact with the packaged vRNA genome and have a role in capsid polymorphism and maturation (see below). In mature capsids, all 240 of the $N\omega V$ capsid subunits have undergone autoproteolysis between residues N_{570} and F_{571} after assembly, leaving the cleaved portion of the C-terminus, called the γ peptide, non-covalently associated with the capsid. This peptide has membrane lytic activity and is essential for virus infectivity. The γ peptide also functions as a quasi-equivalent molecular switch (Fig. 2(a) and (b)). In $N\omega V$, a segment of the γ peptide (residues 608–641) is only ordered in the C and D subunits and functions as a wedge between the ABC and DDD morphological units to prevent curvature and create a flat contact. The interface between two ABC units is bent partly due to the lack of ordered γ peptide. The crystal structure of $HasV$ VLPs is very similar to the $N\omega V$ capsid, presenting a root-mean-square deviation of 1.08. Importantly, tetraviruses subunit structure and the mechanism of autoproteolysis displays a strong relationship with the $T = 3$ insect nodaviruses. Indeed, the β -barrel folds and autocatalytic sites can be superimposed with little variation, with the γ peptide in the nodaviruses also involved in quasi-equivalent switches and membrane lytic activity.

The differences between tetravirus capsids are most pronounced in the Ig-like domain. The crystal structure of PrV at 3.8 Å is the only high-resolution structure of a tetravirus with a monopartite genome. The PrV subunit retains the three domains and a superimposable autoproteolysis site (all four subunits show a clear break at the cleavage site), but the Ig-like domain and the first and last ordered residues of the termini in the helical domain have different positions and structure compared to $N\omega V$ and $HasV$. The different Ig domain position was first observed in the cryo-EM reconstruction of $N\beta V$ as a distinctly different surface structure compared to the omega-like particle surfaces (Fig. 3). Interestingly, the use of subunit termini in PrV is more like that seen in the nodaviruses. The quasi-symmetry switch in PrV has swapped elements compared to the omegatetraviruses and represents a new and unique mixture of structural features among the insect viruses.

Capsid Maturation, Auto-Proteolysis and Dynamics

$N\omega V$ and $HasV$ coat proteins expressed in the baculovirus system assemble into a round and porous metastable VLPs that are 480–490 Å in diameter, called procapsids, when purified at neutral pH 7.6. The procapsid coat proteins remain in the uncleaved precursor form until a reduction in pH. When exposed to acidic conditions (pH 5.0), procapsids undergo a large-scale structural rearrangement (maturation) in which the spherical and porous procapsid condenses into the smaller, angular mature capsid. In $N\omega V$ the structural transition triggers the autocatalytic cleavage, where the 70 kDa α protein begins to cleave into the β (62 kDa) and γ (8 kDa) polypeptides, in a reaction that proceeds even at neutral pH (Fig. 2(b)). Small angle X-ray scattering demonstrated that the pH-induced large conformational change occurs in less than 100 ms in $N\omega V$, while the autocatalytic cleavage of all subunits takes several hours to complete. Structure Analysis

Mutagenesis studies have shown the asparagine at the scissile bond (N₅₇₀) in N ω V is required for cleavage. Substituting threonine for asparagine (N₅₇₀T) results in assembly of procapsids that do not cleave but continue to undergo the conformational change reversibly as a function of pH (Fig. 2(b)). Cryo-EM structures at 8 Å resolution of cleaved and N₅₇₀T VLPs demonstrated the role of cleavage in stabilizing the capsid form. There are more extensive inter-subunit contacts in cleaved particles that result in more ordered C-terminal helices, which account for the additional stability of the particle.

Cryo-EM was used to compare particles at intermediate stages of cleavage. Since these intermediates have the same size as a mature capsid, they are superimposable, allowing a detailed analysis of the structural differences in the cleavage site. N ω V maturation has also served as a model for the development of an innovative method for Cryo-EM image reconstruction that captures the dynamic character of viruses displayed in the CryoEM ensemble of particles at the moment of freezing. The result is a quantitative portrait of the structural variance derived from protein dynamics. With this method it was possible to quantify the progressive capsid stabilization during maturation (Fig. 2(b)). Another relevant finding reproduced by both analysis approaches was that the four quasi-equivalent subunit positions of N ω V cleave at different rates and have specialized functions in the particle. γ peptide derived from the pentamer subunits are produced first and are organized in a vertical helical bundle that is projected towards the particle surface. The identical polypeptides in other quasi-equivalent subunits are produced later and act as molecular switches. They make extensive monovalent contacts and are trapped in positions inappropriate for release. Functional data derived from liposome assays provided experimental support for the hypothesis that the pentameric helical bundle is a specialized lytic domain of the capsid structure. The lytic activity of mature capsids against cellular membranes is triggered by alkaline conditions (Fig. 2(b)). This condition is found in the mid-gut of the larvae and seems to be required for infectivity of HaSV virions (see next).

Pathology

Symptoms and Transmission

Tetraviral infection of lepidopteran larvae occurs primarily by ingestion and the symptoms of infection vary widely between the tetraviruses. *Trichoplusia ni* virus (TnV), when it infects *Trichoplusia ni* larvae, causes only a slight retardation in the growth of the larvae. This is in contrast to N β V, which are lethal to infected *N. capensis* larvae and can act rapidly. N β V-infected *Antheraea eucalypti* larvae could appear healthy and be feeding normally but within 12 h, they would be dead. Larvae dying of an N β V infection exhibit certain characteristics: they either hang from the branches by the prolegs or fall to the ground and the contents of their bodies liquefy. All infection studies have shown that the stage at which the tetravirus infection occurs is critical to the outcome. Early instar (L1-L3) *H. armigera* larvae infected with HaSV cease to feed within 24 h, stunt dramatically, liquefy and die within 4 days. Infection of larvae as later instars does not result in any detectable symptoms or lethality, although the larvae do test positive for the presence of HaSV. There have been no reports of infection of pupae or adults by tetraviruses, but there is evidence that vertical transmission may occur.

Host Range

It was a long held belief that tetraviruses replicate only in lepidopteran midgut cells. Tetraviral infection, based on data collected from HaSV infection of *H. armigera* larvae is focused in the midgut of the animal. Regardless of whether HaSV was fed to the larvae or injected into the hemocoel of the animal, vRNA could be detected only in the midgut. Infected cells undergo apoptosis and are sloughed from the midgut of the animal. It is proposed that the virus exploits this host cell response because apoptosis induces cellular acidification, which would induce maturation of the tetravirus particles. The carmotetravirus, PrV, is also targeted to the larval midgut, being first identified as a persistent infection in a *H. zea* midgut (MG8) cell line.

Recent findings, however, indicate that HaSV and PrV can infect other lepidopteran cell types as well. Both PrV and HaSV readily infect and replicate in *S. frugiperda* Sf9 cells, which were derived from *S. frugiperda* wing bud tissue and PrV has also been shown to infect *Spodoptera exigua* 1 cells. An explanation for the apparent tropism of tetraviruses to midgut tissue is provided by the observation that HaSV virus particles require structural rearrangements brought about by exposure to alkaline pH to facilitate binding host cell receptors followed by entry and infection of *S. frugiperda* Sf9 cells. These alkaline pH conditions are provided by only the lepidopteran midgut and hence infection is confined to midgut tissues.

The host range of carmotetraviruses has recently broadened beyond lepidopteran insects. PrV has been shown to infect and replicate in human cervical (HeLa) and breast (MCF7) cancer cell lines. Remarkably, PrV particles purified from *H. zea* MG8 cells are able to productively infect *Vigna unguiculata* (cowpea) plants. vRNA can be detected at sites distant from the site of infection and PrV particles purified from cowpea plants are capable of infecting and establishing replication in mammalian cancer cell lines. These data reveal that PrV is a virus that has crossed Kingdom boundaries retaining the ability to infect and replicate in animal and plant cells. While species of the *Nodaviridae* family have a similar host range, the initiation of replication in yeast, plant and mammalian cells needs to be launched from transfected cDNA or RNA.

Persistent Infections

PrV was the first tetravirus known to replicate in tissue culture. The virus was first isolated from a persistently infected *H. zea* MG8 cell line and subsequently, PrV has been observed to establish persistent infections in other lepidopteran cell lines including

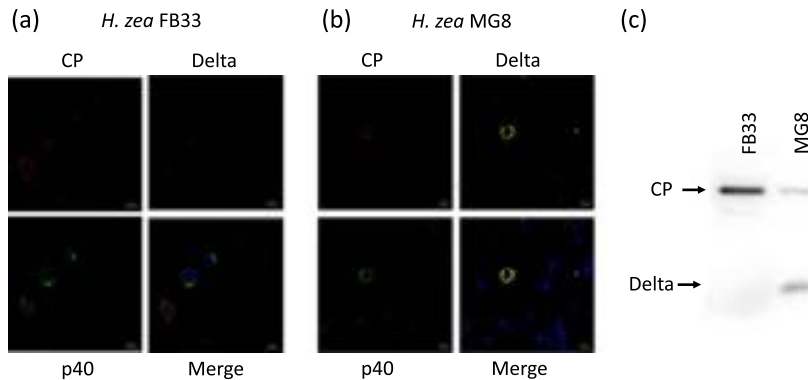


Fig. 4 PrV infected *H. zea* midgut cells express the intestinal stem cell marker Delta. Detection of CP, Delta and p40/REP in persistently infected *H. zea* FB33 (a) and MG8 (b) tissue culture cell lines. Cells were probed with rabbit polyclonal anti-PrV CP and anti-rabbit AF633 secondary antibodies to detect the PrV CP. Mouse anti-Delta and anti-mouse AF546 were used to detect the Delta protein. Viral replicase was stained with biotin-conjugated anti-PrV p40 polyclonal antibodies and streptavidin AF488 to detect the PrV REP. All images represent 1 μ m optical slices taken using a Zeiss LSM780 laser scanning confocal microscope using a X63, 0.75 NA objective. Scale bar represents 10 μ m. (c) Cell-free lysates from PrV-infected *H. zea* FB33 and MG8 cells were separated on an SDS-PAGE gel and western analysis to detect the PrV CP using anti-PrV CP (rabbit) polyclonal and anti-rabbit horse-radish-peroxidase (HRP)-conjugated (goat) secondary antibodies and Delta was detected using anti-Delta (mouse) primary and anti-mouse HRP (goat) secondary antibodies.

an *H. zea* fat body (FB33) and *S. frugiperda* (Sf9) wing bud cell lines. PrV is also able to replicate persistently in human HeLa cells. Lepidopteran larval midgut tissues contain multipotent intestinal stem cells that could be involved in the establishment of persistent PrV infections. Delta, a marker of insect stem cells, is involved in Notch signaling and in the midgut of insects, cellular Delta protein levels are thought to determine the cell fate. High levels of Delta protein activate the Notch pathway, which in turn leads to the down-regulation of Delta in the daughter cells resulting in the cells becoming enterocytes. On the other hand, cells with low levels of Delta are destined to become enteroendocrine cells. Therefore it is the cellular levels of Delta that decides the fate of the cell. PrV infection appears to either target Delta expressing cells or to induce Delta expression in *H. zea* MG8 cells (Fig. 4(b,c)). However, Delta is not expressed in persistently infected *H. zea* FB33 cell line (Fig. 4(a,c)) suggesting that PrV persistence does not involve the infection of insect midgut stem cells.

Concluding Remarks

Since the discovery of N β V, tetraviruses have been defined as small (+ve) ssRNA viruses that share a distinctive $T = 4$ capsid architectures. Their non-structural genes, however, are representative of three diverse virus supergroups, the alpha-like (type IV), the carmo-like (type II) and the picorna-like (type I) supergroups. This implies that modern tetraviruses are the result of converging evolutionary paths that resulted from re-assortment between diverse animal and plant (+ve) ssRNA ancestral viruses. The availability of metatranscriptomic sequence data has significantly extended the extent of the invertebrate virosphere, and there are reports of potentially new tetravirus species. However, these data are based on homology with tetravirus REP sequences. The conserved tetravirus CP is not present on the reconstructed viral genomes and therefore these viruses cannot be considered to be tetraviruses based upon the current definition. The ability of tetravirus capsids to package non-viral RNAs certainly facilitates the evolution of new, re-assorted viruses in the lepidopteran midgut that must be ongoing and undoubtedly, will result in the discovery of new tetravirus species in the future.

Reference

Grace, T.D.C., Mercer, E.H., 1976. Serological relations between twelve small RNA viruses of insects. *J. Gen. Virol.* 31, 131–134.

Further Reading

- Azad, K., Banerjee, M., Johnson, J.E., 2017. Enzymes and enzyme activity encoded by nonenveloped viruses. *Annual Review of Virology* 4, 221–240.
- Doerschuk, P.C., Gong, Y., Xu, N., Domitrovic, T., Johnson, J.E., 2016. Virus particle dynamics derived from CryoEM studies. *Current Opinion in Virology* 18, 57–63.
- Domitrovic, T., Matsui, T., Johnson, J.E., 2012. Dissecting quasi-equivalence in non-enveloped viruses: Membrane disruption is promoted by lytic peptides released from subunit pentamers, not hexamers. *Journal of Virology* 86, 9976–9982.
- Dorrington, R.A., Jiwaji, M., Awando, J.A., de Bruyn, M.-M., 2019. Advances in tetravirus research: New Insight into the infectious virus lifecycle and an expanding host range. In: Bonning, B.C. (Ed.), *Insect Molecular Virology: Advances and Emerging Trends*. Norfolk: Caister Academic Press, pp. 145–162.

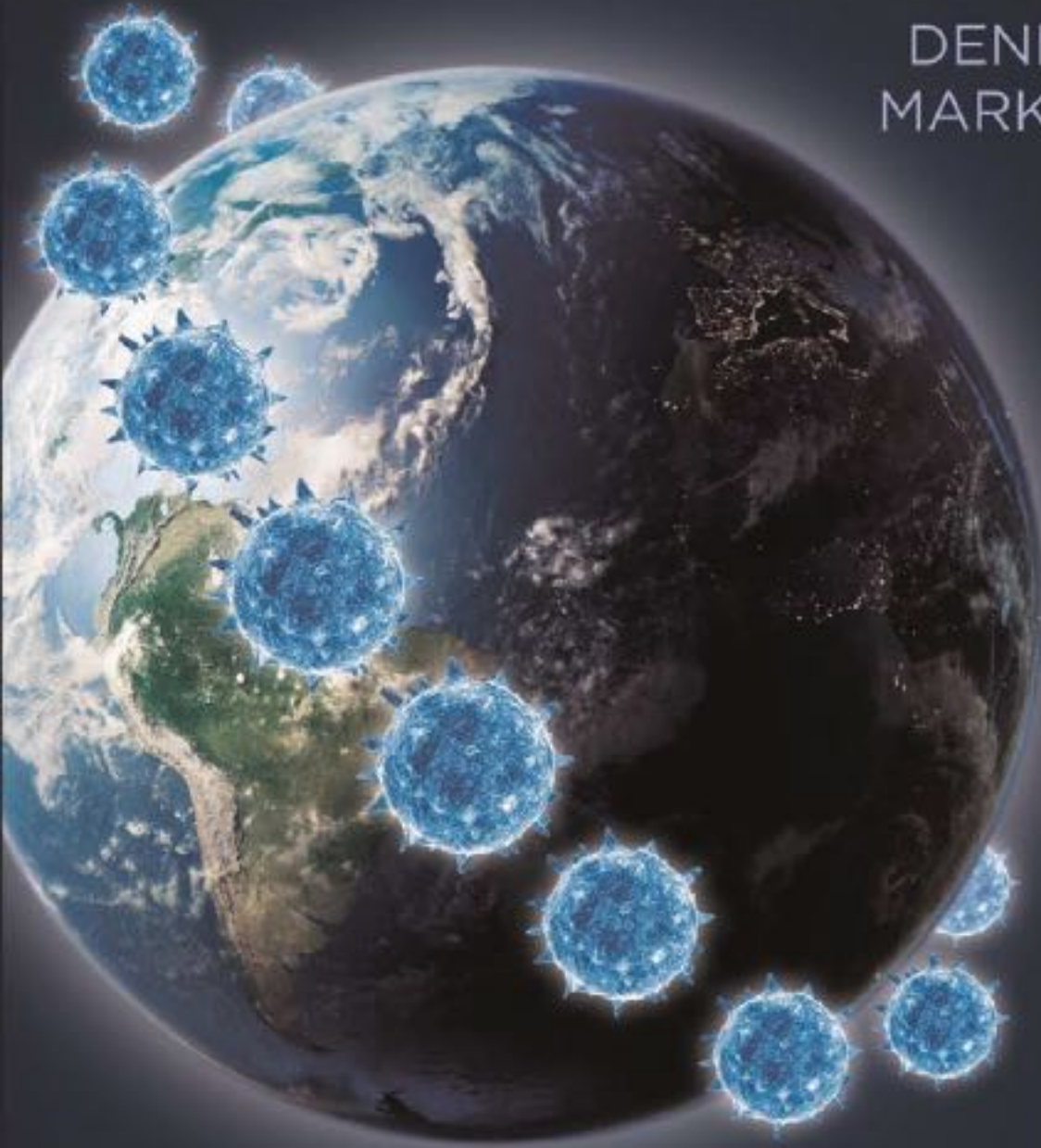
- Jiwaji, M., Matcher, G.F., de Bruyn, M.-M., *et al.*, 2019. Providence virus: An animal virus that replicates in plants or a plant virus that infects and replicates in animals? *PLoS One* 14, e0217494.
- Jiwaji, M., Short, J.R., Dorrington, R.A., 2016. Expanding the host range of small insect RNA viruses: Providence virus (*Carmotetraviridae*) infects and replicates in a human tissue culture cell line. *Journal of General Virology* 97, 2763–2768.
- Penkler, D.L., Jiwaji, M., Domitrovic, T., *et al.*, 2016. Binding and entry of a non-enveloped T= 4 insect RNA virus is triggered by alkaline pH. *Virology* 498, 277–287.
- Routh, A., Domitrovic, T., Johnson, J.E., 2012. Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 109, 1907–1912.
- Speir, J.A., Taylor, D.J., Natarajan, P., *et al.*, 2010. Evolution in action: N and C termini of subunits in related T= 4 viruses exchange roles as molecular switches. *Structure* 18, 700–709.
- Tang, J., Kearney, B.M., Wang, Q., *et al.*, 2014. Dynamic and geometric analyses of *Nudaurelia capensis* ω virus maturation reveal the energy landscape of particle transitions. *Journal of Molecular Recognition* 27, 230–237.
- Zeddam, J.-L., Gordon, K.H., Lauber, C., *et al.*, 2010. *Euprosterna elaeasa* virus genome sequence and evolution of the Tetraviridae family: Emergence of bipartite genomes and conservation of the VPg signal with the dsRNA Birnaviridae family. *Virology* 397, 145–154.

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

Editors in Chief

DENNIS BAMFORD
MARK ZUCKERMAN



ENCYCLOPEDIA OF VIROLOGY FOURTH EDITION

Volume 5

ENCYCLOPEDIA OF VIROLOGY

FOURTH EDITION

EDITORS IN CHIEF

Dennis H. Bamford

*Molecular and Integrative Biosciences Research Programme
Faculty of Biological and Environmental Sciences
University of Helsinki, Helsinki, Finland*

Mark Zuckerman

*South London Specialist Virology Centre
King's College Hospital NHS Foundation Trust
London, United Kingdom*

and

*Department of Infectious Diseases
School of Immunology and Microbial Sciences, King's College London Medical School
London, United Kingdom*

Volume 5



AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge MA 02139, United States

Copyright © 2021 Elsevier Ltd. unless otherwise stated. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers may always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

ISBN 978-0-12-814515-9

For information on all publications visit our
website at <http://store.elsevier.com>

		Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org		

Publisher: Oliver Walter

Acquisitions Editor: Priscilla Braglia

Content Project Manager: Katarzyna Miklaszewska

Associate Content Project Manager: Gayathri S.

Designer: Matthew Limbert

EDITORS IN CHIEF



Dennis H. Bamford, PhD, is Professor Emeritus of Virology at the Faculty of Biological and Environmental Sciences, University of Helsinki, Finland. He obtained his PhD in 1980 from the Department of Genetics, University of Helsinki. During 1981–1982 he was an EMBO postdoctoral fellow at the Public Health Research Institute of the City of New York, United States, and during 1983–1992 he worked as a Senior Scientist at the Academy of Finland. In 1993 he was appointed Professor of General Microbiology at the University of Helsinki. He was awarded the esteemed Academy Professorship twice, in 2002–2007 and 2012–2016, and he also served twice as the Director of the Finnish Center of Excellence (in Structural Virology, 2000–2005, and in Virus Research, 2006–2011). Prof. Bamford has had continuous external research funding (e.g., from several European Union, Academy of Finland, TEKES and Jusélius Foundation funds, as well as the Human Frontier Science Program). He is an EMBO member and has held several positions of trust in scientific and administrative organizations.

Prof. Bamford has published approx. 400 articles in international peer-reviewed journals in virology, microbiology, biochemistry, and molecular biology (36 of them in high impact journals). About half of the primary articles have been published with international collaborators showing high international integration. He has also been invited to give 56 keynote and plenary presentations in major international meetings. Prof. Bamford has supervised over 35 Master's and

over 40 PhD theses. Seven of his graduate students or post docs have obtained a professorship and a similar number have a principal investigator status. Prof. Bamford has studied virus evolution from a structure-centered perspective, showing that seemingly unrelated viruses, such as bacteriophage PRD1 and human adenovirus have similar virion architecture. When the corona virion architecture was gradually revealed, it was observed that its structural elements were close to those seen in RNA bacteriophage phi6 so that phi6 has been actively used as surrogate for pathogenic viruses - quite a surprise!



Dr. Mark Zuckerman is Head of Virology, Consultant Medical Virologist, and Honorary Senior Lecturer at South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust and King's College London Medical School, Department of Infectious Diseases, School of Immunology and Microbial Sciences in London, United Kingdom. His interests include the clinical interface between developing molecular diagnostic tests relevant to the local population of patients, respiratory virus infections, herpesvirus infections in immunocompromised patients and blood-borne virus transmission incidents in the healthcare setting. He has chaired the UK Clinical Virology Network, Royal College of Pathologists Virology Specialty Advisory Committee and Virology Examiners Panel and is a member of the Specialty Advisory Committee on Transfusion Transmitted Viruses. He is a co-author on four editions of the "Mims' Medical Microbiology" textbook, has written chapters in a number of other textbooks and has over 100 publications in international peer-reviewed journals and is an associate editor for two journals.

EDITORIAL BOARD

Editors in Chief

Dennis H. Bamford

Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Mark Zuckerman

South London Specialist Virology Centre, King's College Hospital NHS Foundation Trust, London, United Kingdom and Department of Infectious Diseases, School of Immunology and Microbial Sciences, King's College London Medical School, London, United Kingdom

Section Editors

Claude M. Fauquet

St Louis, MO, United States

Michael Feiss

Department of Microbiology and Immunology, Carver College of Medicine, University of Iowa, Iowa City, IA, United States

Elizabeth E. Fry

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom

Said A. Ghabrial[†]

Department of Plant Pathology, University of Kentucky, Lexington, KY, United States

Eric Hunter

Department of Pathology and Laboratory Medicine, Emory University School of Medicine and Emory Vaccine Center, Emory University, Atlanta, GA, United States

Ilkka Julkunen

Institute of Biomedicine, University of Turku, Turku, Finland

Peter J. Krell

Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada

Mart Krupovic

Archaeal Virology Unit, Institut Pasteur, Paris, France

Maija Lappalainen

HUS Diagnostic Center, HUSLAB, Clinical Microbiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Hubert G.M. Niesters

Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Massimo Palmarini

MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

David Prangishvili

Institut Pasteur, Paris, France and Ivane Javakhishvili Tbilisi State University, Tbilisi, Georgia

David I. Stuart

Department of Structural Biology, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom and Diamond Light Source, Didcot, United Kingdom

Nobuhiro Suzuki

Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

[†]Deceased.

SECTION EDITORS



Claude Fauquet received his PhD in biochemistry from University Louis Pasteur in Strasbourg, France in 1974. Dr. Fauquet joined the Institut de Recherche pour le Développement (IRD) and worked there as a plant virologist for 28 years, and served in Ivory Coast, West Africa for 14 years. In 1991, he founded the International Laboratory for Tropical Agricultural Biotechnology (ILTAB) at The Scripps Research Institute, CA, United States. ILTAB was then hosted by the Donald Danforth Plant Science Center, St. Louis, MO, from 1999 to 2012. In 2003, he co-founded the Global Cassava Partnership for the 21st Century (GCP21), which he directed until 2019 and which goal is to improve the cassava crop worldwide.

Dr. Fauquet is an international leader in plant virology including taxonomy, epidemiology, molecular virology, and in gene-silencing as an antiviral strategy. He was Secretary of the International Committee on Taxonomy of Viruses (ICTV) for 18 years and the editor of several ICTV Reports including the VIIIth ICTV Report in 2005.

He has published more than 300 research papers in reviewed journals and books. He is a fellow of the American Association for the Advancement of Science, of the American Phytopathological Society and a member of the St. Louis Academy of Sciences. In 2007, Dr. Fauquet was knighted “Chevalier de l’Ordre des Palmes Académiques” by the French Minister of High Education and Research.



Dr. Michael Feiss is Professor Emeritus in the Department of Microbiology and Immunology of the Carver College of Medicine at the University of Iowa, IA, United States. Dr. Feiss received his PhD in Genetics at the University of Washington followed by a postdoctoral traineeship in the laboratory of Dr. Allan Campbell at Stanford. Dr. Feiss is a microbial geneticist who studies virus assembly with an emphasis on how a DNA virus, bacteriophage lambda, packages viral DNA into the empty prohead shell. The lab investigates how sites in the viral DNA orchestrate the initiation and termination of the DNA packaging process. This work includes comprehensive examination of the DNA recognition sites. A related interest is study of terminase, the viral DNA packaging enzyme, including the functional domains for protein–DNA and protein–protein interactions. A second focus has been the roles of the bacterial host’s IHF and DnaJ proteins in the lytic life cycle of the virus. More recent work has involved a genetic dissection of the role of terminase’s ATPase center that powers translocation of viral DNA into the prohead. This interest in the ATP hydrolysis-driven packaging motor involves a multidisciplinary collaboration examining the kinetics of DNA packaging during individual packaging events. Finally, recent studies have also looked at how the packaging process has diverged among several lambda-like phages, including phages 21, N15, and Gifsy-1.



Elizabeth E. Fry is a senior postdoctoral scientist in structural biology at the University of Oxford, Oxford, United Kingdom, where she received her DPhil. for studies relating to the structure determination of Foot-and-Mouth Disease Virus. Specializing in structural virology, Dr. Fry has studied many virus/viral protein structures but her primary focus is on picornavirus structure and function, in particular receptor interactions and virus uncoating. She is particularly interested in rationally designing virus-like-particles as next generation vaccines to reduce the inherent risks in handling live viruses.



Said A. Ghabrial[†] received his BSc in 1959 from Cairo University, Cairo, Egypt, and his PhD from Louisiana State University, Baton Rouge, LA, United States, in 1965. Dr. Ghabrial did postdoctoral research at the University of California, Davis, CA, United States, before returning to Cairo, where he served as a plant virologist in the Ministry of Agriculture. He returned to the United States in 1970 to do postdoctoral research at Purdue University, West Lafayette, IN. In 1972, he joined the Plant Pathology Department at the University of Kentucky, Lexington, KY, United States, where he rose to the rank of professor in 1986 and worked until 2013.

Dr. Ghabrial has served as an associate and senior editor of *Phytopathology*. He served on the editorial boards of the *Encyclopedia of Virology*, 3rd edition and *Encyclopedia of Plant Pathology*, and edited a thematic issue of *Advances in Virus Research* on "Mycoviruses". He was a member of the American Phytopathological Society (APS) and the American Society for Virology (ASV); in July 2002 he was elected as a Fellow of the American Phytopathological Society. He also acted as Chair of the ICTV Subcommittee on Fungal Viruses in 1987–1993 and 2011–2014.

His long professional career allowed him to make many scientific achievements in phytopathology and virology. Among them are molecular dissection of a legume-infecting RNA virus, bean pod mottle virus (BPMV), development of BPMV-based vectors, discovery of a transmissible debilitation disease of the phytopathogenic ascomycete, *Helminthosporium victoriae* (*Cochliobolus victoriae*), establishment of a viral etiology of the *H. victoriae* disease, and advancement of structural biology of diverse fungal viruses.



Eric Hunter, PhD, is Professor of Pathology and Laboratory Medicine at Emory University, Atlanta, GA, United States. He serves as Co-Director of the Emory Center for AIDS Research and is a Georgia Research Alliance Eminent Scholar.

Dr. Hunter's research focus has been the molecular virology and pathogenesis of retroviruses, including human immunodeficiency virus. He has made significant contributions to the understanding of the role of retroviral glycoprotein structural features during viral entry and providing unique insights into the assembly and replication of this virus family. In recent years the emphasis of his research has been on HIV transmission and pathogenesis, defining the extreme genetic bottleneck and selection of viruses with unique traits during HIV heterosexual transmission. He has described the selection of fitter viruses at the target mucosa, a gender difference in the extent of selection bias, and a role for genital inflammation in reducing selection. His research has defined the impact of HIV adaptation to the cellular immune response on immune recognition and control of HIV after

transmission, as well as on virus replicative fitness in vitro and in vivo. Recent work highlights the roles that virus replicative fitness and sex of the host play in defining disease progression in a newly infected individual. His bibliography includes over 300 peer-reviewed articles, reviews, and book chapters. He has also been the recipient of four NIH merit awards for his work on retrovirus and HIV molecular biology.

Dr. Hunter served as the Editor in Chief of the journal *AIDS Research and Human Retroviruses* for 10 years. He was Chair of the AIDS Vaccine Research Subcommittee which is charged with providing advice and consultation on AIDS vaccine research to the National Institute of Allergy and Infectious Diseases and continues to serve on editorial boards for several academic journals and on external advisory committees for several government, academic, and commercial institutions.



Ilkka Julkunen graduated as an MD/PhD in 1984 from the Department of Virology, University of Helsinki, Helsinki, Finland. He worked as a postdoctoral research fellow at Memorial Sloan-Kettering Cancer Center in New York, United States, in 1986–1989, followed by positions as a senior scientist, group leader and research professor at Finnish Institute for Health and Welfare in 1989–2013. In 2013 he became a Professor of Virology at the University of Turku, Turku, Finland. The research interests of Dr. Julkunen have concentrated on innate and adaptive humoral immunity in viral and microbial infections. He has studied intracellular signaling and RIG-I and TLR-mediated activation of interferon system in human macrophages and dendritic cells and stable cell lines in response to human and avian influenza, Sendai, Zika and coronavirus infections. In addition, he has analyzed the downregulation of innate immunity by viral regulatory proteins from influenza, HCV, flavi-, filo- and coronaviruses. He has expertise in vaccinology, biotechnology and development of methods to analyze antiviral immunity, he has also been actively involved in research training and collaborations with biotechnological industry.

[†]Deceased.



Peter Krell started his career in virology early as a summer high school student working for the Canadian Forestry Service studying the resistance of nuclear polyhedrosis viruses (now called baculoviruses) to environmental exposure with Dr. Fred T. Bird at the Insect Pathology Research Institute in Sault Ste. Marie, ON, Canada. He received his BSc and MSc in biology from Carleton University studying the iridovirus *Tipula Iridescent Virus* with Dr Peter Lee, in Ottawa, the Canadian capital. For his PhD he headed east to Dalhousie University in Halifax, Nova Scotia on the Atlantic coast. In addition to enjoying the salt sea air, fresh cod, lobster and mussels, he studied the molecular biology of polydnaviruses under the guidance of Dr Don Stoltz. Heading south to Texas A&M University in College Station, TX, United States, as a Postdoctoral Fellow he worked with Dr. Max Summer, of baculovirus fame, and Dr. Brad Vinson continuing to study polydnaviruses, but also became steeped in the early days of molecular baculovirology. He then accepted a faculty position in the Department of Microbiology and Immunology at the University of Guelph in Guelph, ON, Canada. There he switched to baculovirus research, which was more tractable, due in part to available cell cultures and focused on viral DNA replication and functional genomics,

particularly on chitinase, cathepsin and ME53. In collaboration with Dr. Eva Nagy he studied molecular biology of different animal viruses, notably Fowl Avian adenoviruses and their development as vaccine vectors, but also on the birnavirus infectious pancreatic necrosis virus, the coronavirus porcine endemic diarrhea virus, fowlpox virus and the paramyxovirus Newcastle disease virus. He has been involved extensively with virus taxonomy, being active in the International Committee on Taxonomy of Viruses (ICTV) as member of the *Polydnaviridae* and *Baculoviridae* study groups, national representative of Canada on the ICTV, member of the Executive Committee for the ICTV and Chair of the ICTV Invertebrate Virus Subcommittee. In terms of governance, Peter Krell was President of the Canadian Society of Microbiology, Secretary and later President of the Society for Invertebrate Pathology, as well as being on the Editorial Boards of the *Canadian Journal of Microbiology* and the *ASM Journal of Virology*. While at the University of Guelph, he rose through the ranks to Professor and is currently University Professor Emeritus.



Mart Krupovic is the Head of the Archaeal Virology Unit in the Department of Microbiology at the Institut Pasteur of Paris, France. He received his MSc in Biochemistry in 2005 from the Vilnius University, Vilnius, Lithuania and PhD in 2010 in general microbiology from the University of Helsinki, Helsinki, Finland. His current research focuses on the diversity, origin, and evolution of viruses, as well as molecular mechanisms of virus–host interactions in archaea. He has published over 170 journal articles and serves as an editor or on the editorial boards of *Biology Direct*, *Research in Microbiology*, *Scientific Reports*, *Virology*, and *Virus Evolution*. He is also a member of the Executive Committee of the International Committee on Taxonomy of Viruses (ICTV) and chairs the Archaeal Viruses Subcommittee of the ICTV.



Maija Lappalainen, MD, PhD, Associate Professor of Clinical Microbiology, is the Head of Clinical Microbiology in the HUS Diagnostic Center, HUSLAB, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. In her thesis during the years 1987–1992 she studied the incidence and diagnostics of congenital toxoplasmosis. After PhD, her research interest has been in diagnostic clinical virology, viral hepatitis, respiratory infections, viral infections in the immunocompromised patients and viral infections during pregnancy.



Hubert G.M. Niesters (1958) studied biology and chemistry in Nijmegen, the Netherlands. After obtaining his PhD in Utrecht (Prof. dr. M. Horzinek and Prof. dr. B. van der Zeijst, 1987) on the molecular epidemiology of infectious bronchitis virus, he worked as a post-doctoral fellow with Prof. dr. Jim Strauss at the California Institute of Technology (Pasadena, United States) on the replication of Alphaviruses. He received a Niels Stensen fellowship (The Netherlands) and an E.S. Gosney fellowship (Caltech) during this period.

After returning to the Netherlands (1989), he became a research associate in medical microbiology at the Diagnostic Medical Center (Delft) but moved back to clinical virology as a senior research associate in 1991 at the Erasmus University Medical Center Rotterdam (Head Prof. dr. Ab Osterhaus). From 1993 to 2007, he was responsible for the molecular diagnostics unit. During this period, he was involved in the discovery and characterization of several new viruses and variants. In 2007, he became full professor and director of the Laboratory of Clinical Virology within the Department of Medical Microbiology at the University Medical Center Groningen and University of Groningen.

He has been actively involved in the implementation and development of new technologies like real-time amplification and automation within clinical virology. He has been focusing on molecular diagnostics and its use and the clinical value in a transplant setting, as well as in monitoring treatment of hepatitis viruses. Recently, his interest focuses on rapid regional epidemiology, automation including MiddleWare solutions for molecular diagnostics, as well as the cost-benefit of rapid point-of-impact molecular testing. Special interest is focused on raising awareness for the detection of enteroviruses (enterovirus D68) and its relationship with acute flaccid myelitis (AFM).

Since 2017, he is the Chair of the executive board of QCMD (Quality Control of Molecular Diagnostics, Glasgow). He is an auditor and team leader for the Dutch Council of Accreditation and Co-Editor in Chief of the *Journal of Clinical Virology*. He is an (co)-author of more than 250 peer-reviewed papers, chapters and reviews including emerging viruses, such as enterovirus D68 and hepatitis E virus (H-index 80).

For his entire work, he received in 2016 the “Ed Nowakowski Senior Memorial Clinical Virology Award” from the Pan American Society for Clinical Virology.



Massimo Palmarini is the Director of the MRC-University of Glasgow Centre for Virus Research and Chair of Virology at the University of Glasgow, Glasgow, United Kingdom. A veterinarian by training, his research programs focus on the biology, evolution and pathogenesis of arboviruses and the mechanisms of virus cross-species transmission. His work is funded by the MRC and the Wellcome Trust. Massimo Palmarini has been elected Fellow of the Academy of Medical Sciences, of the Royal Society of Edinburgh and of the Royal Society of Biology and he was a Wolfson-Royal Society Research Merit Awardee. He is a Wellcome Trust Investigator.



David Prangishvili, PhD, Honorary Professor at the Institut Pasteur, Paris, France, and Professor at Tbilisi State University, Tbilisi, Georgia, is one of the pioneers in studies on the biology of Archaea and their viruses. His scientific career spans ex-USSR (Institute of Molecular Biology, Moscow; 1970–1976), Georgia (Georgian National Academy of Sciences, Tbilisi; 1976–1991), Germany (Max-Planck Institute for Biochemistry, Munich; University of Regensburg; 1991–2004) and France (Institut Pasteur, Paris, 2004–2020). In the research groups headed by him, several dozens of new species and eight new families of archaeal viruses have been discovered and characterized, which display remarkable diversity of unique morphotypes and exceptional genome contents. The results of his research contribute to the knowledge on viral diversity on our planet and change the field of prokaryotic virology, leading to the notion that viruses of

hyperthermophilic Archaea form a particular group in the viral world, distinctive from viruses of Bacteria and Eukarya, and to the recognition of the virosphere of Archaea as one of the distinct features of this Domain of Life. David Prangishvili is a member of the Academia Europaea, the European Academy of Microbiology, and the Georgian National Academy of Sciences.



David I. Stuart is MRC Professor of Structural Biology in the Nuffield Department of Medicine, Oxford University, Oxford, United Kingdom, Life Science Director at Diamond Light Source and Director of Instruct-ERIC (pan-European organisation providing shared access to infrastructure and methods for structural biology). He has diverse interests in structural virology from picornaviruses, double-stranded RNA viruses and enveloped RNA viruses. His drive to develop structural techniques led to the determination of the structure of Bluetongue virus (1995) and then the first membrane containing virus, PRD1. More recently, he has been at the fore-front of bringing Cryo-EM technology to bear on virus structure determination and its future role in visualizing virus function in cellulose. In addition to basic science he has a strong commitment to structural vaccinology and the development of antiviral drugs.



Dr. Nobuhiro Suzuki, PhD, received his MSc (1985) in phytopathology and PhD (1989) in virology from Tohoku University in Sendai, Japan. Dr. Suzuki currently serves as a full Professor of the Institute of Plant Stress and Resources, formerly Institute of Plant Sciences and Bioresources at Okayama University and as an Editor of *Virus Research*, *Frontiers in Virology*, *Journal of General Plant Pathology*, *Virology Journal*, and *Biology*. He has also been Guest Editor to *PLoS Pathogens*, *PNAS*, and *mBio*, and an Editorial Board member of *Virology* and *Journal of Virology*.

Suzuki Laboratory focuses on characterization of diverse viruses infecting phytopathogenic fungi and exploration of their interplays. Recent achievements include the discovery of a neo-virus lifestyle exhibited by a (+)ssRNA virus and an unrelated dsRNA virus in a plant pathogenic fungus and of multilayer antiviral defense in fungi involving Dicer. Prior to coming to Kurashiki, Okayama Prefecture, he was a visiting fellow of the Center for Agricultural Biotechnology at the

University of Maryland Biotechnology Institute (UMBI), College Park, MA, United States, for 4 years (1997–2001) to study molecular biology of hypoviruses in the laboratory of Professor Donald L. Nuss. Before visiting UMBI, he served as an assistant professor and a lecturer of the Biotechnology Institute at the Akita Prefectural College of Agriculture, Japan, for 11 years (1988–1998) where he was engaged in a project on molecular characterization of rice dwarf phytoecovirus, a member of the family *Reoviridae*. He received awards from the Japanese Phytopathological Society of Japan and Japanese Society for Virology for his outstanding achievements in plant and fungal virology.

FOREWORD

I am delighted to write the foreword to this wonderful Fourth Edition of the *Encyclopedia of Virology*. The Third Edition was published in 2008, how the world has changed in the intervening years. The release of the updated fourth edition could not be more timely or more prescient. It is superb and a huge tribute to the authors, Elsevier the publisher, and to the brilliant editors, Dennis Bamford and Mark Zuckerman.

SARS-CoV-2 has dominated the world since it emerged in 2019 and affected every continent and every aspect of life. A reminder, if it were needed, of the impact of infectious diseases, the importance of virology and the vulnerability and interconnectivity of our world.

There is no doubt that with rapidly changing ecology, urbanization, climate change, increased travel, and fragile public health systems, epidemics and pandemics will become more frequent, more complex and harder to prevent and contain. Most of these epidemics will be caused by viruses, those we know about and maybe able to predict and some we do not know of that will emerge from animals, plants or the environment. Our changing climate will change the epidemiology of viruses, their vectors and the infections they cause, hence the critical importance of this totally revised Fourth Edition of the *Encyclopedia of Virology* which brings together research and an understanding of viruses in animals, plants, bacteria and fungi, the environment, and among humans. Never has a holistic, one-health understanding been more important.

That starts with an understanding of the fundamentals of virology, a field of science that has been transformed in the years since the Third Edition. An understanding transformed by embracing traditional fields of molecular and structural biology, genomics, and influenced by immunology, genetics, pharmacology and increasingly by epidemiology and mathematics. Events of 2020 and 2021 also show why it is so important to integrate within traditional virology an understanding of the animal and human health and behavior, of climate change and its impact on the ecology of viruses, plant sciences and vectors. And why we must understand the viruses we think we know well, and those viruses less extensively studied. Research is critical to this, research that pushes the boundaries of what we know, has the humility to seek answers to things we do not understand and shares that knowledge with the widest possible community. That research will be most exciting at the interface between disciplines, most impactful when dynamic, open, inclusive, global, and collaborative.

This is what the Fourth Edition of the *Encyclopedia of Virology*, the largest reference source of research in virology sets out to achieve. It is a wonderful contribution to a critical field of knowledge. It contains new chapters, every chapter revised and updated by a dedicated global community who have come together to provide what is a brilliant and inspiring reference. It is an honor to contribute in a very small way to the timely release of the Fourth Edition of the *Encyclopedia of Virology*.

Jeremy Farrar

PREFACE

The fourth edition of the Encyclopedia of Virology is encyclopedic, but we wanted to move away from an alphabetical list, apart from where it was more logical, to a vision that encompassed a different structure. Articles describing novel trends as well as original discoveries in specific subfields of virology have been distributed into a set of five volumes, namely Fundamentals of Virology, Human and Animal Viruses, Plant Viruses, Bacterial, Archaeal, Fungal, Algal and Invertebrate Viruses, and Diagnosis, Treatment and Prevention of Virus Infections.

We had hoped that the new edition would 'go viral' but it was ironic that the time to publication 12 years after the previous edition had been made a bit longer due to a virus infection. The world encountered a devastating global pandemic, COVID-19, caused by a new type of a coronavirus, SARS-CoV-2. Scientists in many disciplines all over the world started immediate efforts to discover solutions as to how to mitigate and stop the spread of the pandemic. Virology moved from being a highly specialized subject to one in which everyone became a virologist, proving just how significant the different aspects of virology are in terms of understanding the nature of viral infection. Since the previous edition, the growth in the field of general virology has been enormous, including huge advances in basic science, identification of novel viruses, diagnostic methods, treatment and prevention. Taking this into account, the introduction of the articles within the Encyclopedia are very timely and crucial for providing a wealth of knowledge of the latest findings in the field of virology to a vast range of people, whether school students, undergraduates, postgraduates, teachers, scientists, researchers, journalists and others interested in infections and the conflict between the host and the pathogen.

Pandemic viruses have become a serious public concern in the changing world. We can ask ourselves whether we have reached the point in which nature can no longer cope with the consequences of increased population density and human activities that are harmful to the environment. Although several pandemics have threatened mankind before, this COVID-19 pandemic has highlighted the massive adverse economic consequences towards the wellbeing of society and the importance of research in virology.

We aimed to produce a Major Reference Work that differs in approach to others and binds all the virology disciplines together. Chapters have been included on origin, evolution and emergence of viruses, environmental virology and ecology, epidemiology, techniques for studying viruses, viral life cycles, structure, entry, genome and replication, assembly and packaging and taxonomy and viral–host interactions. Information has been included on all known species of viruses infecting bacteria, fungi, plants, vertebrates and invertebrates. Additional topics include antiviral classification and examples of their use in management of infection, diagnostic assays and vaccines, as well as the economic importance of viral diseases of crops and their control.

This edition used viral classification according to the 9th Report of the International Committee on Taxonomy of Viruses published in 2012. Updating it to the 10th Report in 2020 was affected by the pandemic and can be found online at <http://ictv.global/report/>.

We wish to acknowledge the hard work, interest, flexibility and patience, during such difficult times both socially and professionally, of everybody involved in the process of writing this edition of the Encyclopedia of Virology, especially Katarzyna Miklaszewska, Priscilla Braglia, Sam Crowe and colleagues at Elsevier. We sincerely thank all the authors and section editors for their excellent contributions to this edition.

Book Cover Image: Viruses are obligate parasites and all cells have their own viruses increasing the total number of viruses to the estimated astronomical number of 10^{31} that extends the number of stars in the universe. The viral string illustrates how pandemic viruses surround the globe. The original picture was created by Dr. Nina Atanasova (Finnish Meteorological Institute and University of Helsinki) and amended by Matthew Limbert at Elsevier.

*Dennis H. Bamford
Mark Zuckerman*

HOW TO USE THE ENCYCLOPEDIA

Structure of the Encyclopedia

All articles in the encyclopedia are arranged thematically as a series of entries within subjects/sections, apart from volume 2 where there it was more logical to have articles arranged alphabetically.

There are three features to help you easily find the topic you are interested in: a thematic contents list, a full subject index, and contributors.

1. Thematic contents list: The alphabetical contents list, which appears at the front of each volume, lists the entries in the order that they appear in the encyclopedia.
2. Index: The index appears at the end of volume 5 and includes page numbers for quick reference to the information you are looking for. The index entries differentiate between references to a whole entry, a part of an entry, and a table or figure.
3. Contributors: At the start of each volume there is a list of the authors who contributed to all volumes.

LIST OF CONTRIBUTORS

Stephen T. Abedon

*The Ohio State University, Mansfield, OH,
United States*

Peter Abrahamian

*Agricultural Research Service, US Department of
Agriculture, Beltsville, MD, United States*

Jônatas S. Abrahão

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Florence Abravanel

*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Nicola G.A. Abrescia

*Center for Cooperative Research in Biosciences, Basque
Research and Technology Alliance, Derio, Spain;
Ikerbasque, Basque Foundation for Science, Bilbao,
Spain; and Center for Biomedical Research in the Liver
and Digestive Diseases Network, Carlos III Health
Institute, Madrid, Spain*

Gian Paolo Accotto

*Institute for Sustainable Plant Protection, National
Research Council of Italy, Torino, Italy*

Elisabeth Adderson

*St. Jude Children's Research Hospital, Memphis, TN,
United States and University of Tennessee Health
Sciences Center, Memphis, TN, United States*

Mustafa Adhab

University of Baghdad, Baghdad, Iraq

Alexey A. Agranovsky

Lomonosov Moscow State University, Moscow, Russia

Nasim Ahmed

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Maher Al Rwahnih

University of California, Davis, CA, United States

Olufemi J. Alabi

*Texas A&M AgriLife Research and Extension Center,
Weslaco, TX, United States*

Aurélien A. Albertini

*Institute for Integrative Biology of the Cell (I2BC),
French Alternative Energies and Atomic Energy
Commission, French National Center for Scientific
Research, Paris-Sud University, University of
Paris-Saclay, Gif-sur-Yvette, France*

Aleksandra Alimova

*The City University of New York (CUNY), School of
Medicine, The City College of New York, New York,
NY, United States*

Juan C. Alonso

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Imran Amin

*National Institute for Biotechnology and Genetic
Engineering, Faisalabad, Pakistan*

Stephanie E. Ander

*University of Colorado School of Medicine, Aurora, CO,
United States*

Danielle E. Anderson

Duke-NUS Medical School, Singapore, Singapore

Ida Bagus Andika

Qingdao Agricultural University, Qingdao, China

Ana C.d.S.P. Andrade

*Federal University of Minas Gerais, Belo Horizonte,
Brazil*

Juana Angel

Pontifical Javeriana University, Bogota, Colombia

Vanesa Anton-Vazquez

King's College Hospital, London, United Kingdom

Guido Antonelli

Sapienza University of Rome, Rome, Italy

Josefa Antón

University of Alicante, Alicante, Spain

Nanako Aoki

*Tokyo University of Agriculture and Technology, Fuchu,
Japan*

Timothy D. Appleby

King's College Hospital, London, United Kingdom

Miguel Arenas

*Department of Biochemistry, Genetics and Immunology,
University of Vigo, Vigo, Spain and CINBIO
(Biomedical Research Center), University of Vigo, Vigo,
Spain*

Basil Arif

*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Vicente Arnau

*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain*

Gaurav Arya

Duke University, Durham, NC, United States

Leyla Asadi

University of Alberta, Edmonton, AB, Canada

Sassan Asgari

The University of Queensland, Brisbane, QLD, Australia

Nina S. Atanasova

*Finnish Meteorological Institute, Helsinki, Finland and
University of Helsinki, Helsinki, Finland*

Houssam Attoui

*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Silvia Ayora

*National Biotechnology Center–Spanish National
Research Council, Madrid, Spain*

Walid Azab

Free University of Berlin, Berlin, Germany

Sasha R. Azar

*The University of Texas Medical Branch, Galveston, TX,
United States*

Fengwei Bai

*The University of Southern Mississippi, Hattiesburg, MS,
United States*

Dalan Bailey

The Pirbright Institute, Pirbright, United Kingdom

S.C. Baker

*Loyola University of Chicago, Maywood, IL,
United States*

Fausto Baldanti

*University of Pavia, Pavia, Italy and Scientific Institute
for Research, Hospitalization and Healthcare, San
Matteo Polyclinic Foundation, Pavia, Italy*

Logan Banadyga

*Public Health Agency of Canada, Winnipeg, MB,
Canada*

Ashley C. Banyard

*Animal and Plant Health Agency, Addlestone,
United Kingdom; University of West Sussex, Falmer,
United Kingdom; and St. George's Medical School,
University of London, London, United Kingdom*

Xiaoyong Bao

*The University of Texas Medical Branch, Galveston, TX,
United States*

Yiming Bao

*Beijing Institute of Genomics, Chinese Academy of
Sciences, University of Chinese Academy of Sciences,
Beijing, China*

Alan D.T. Barrett

*The University of Texas Medical Branch, Galveston, TX,
United States*

Diana P. Baquero

*Archaeal Virology Unit, Institut Pasteur, Paris, France
and Sorbonne University, Paris, France*

Moshe Bar-Joseph

*Agricultural Research Organization, Volcani Center, Bet
Dagan, Israel*

Rachael S. Barr

*Bristol Royal Hospital for Children, Bristol,
United Kingdom*

Ralf Bartenschlager

Heidelberg University, Heidelberg, Germany

David L.V. Bauer

Francis Crick Institute, London, United Kingdom

Oliver W. Bayfield

University of York, York, United Kingdom

Sally A. Baylis

Paul-Ehrlich-Institute, Langen, Germany

Philippa M. Beard

*The Pirbright Institute, Pirbright, United Kingdom and The
Roslin Institute, University of Edinburgh, United Kingdom*

Paul Becher

University of Veterinary Medicine, Hannover, Germany

Björn Becker

Saarland University, Saarbrücken, Germany

Karen L. Beemon

Johns Hopkins University, Baltimore, MD, United States

Martin Beer

Friedrich-Loeffler-Institute, Insel Riems, Germany

Jose Miguel Benito

*Health Research Institute of the Jiménez Díaz
Foundation, Autonomous University of Madrid and Rey
Juan Carlos University Hospital, Móstoles, Spain*

Mária Benkő

*Institute for Veterinary Medical Research, Center for
Agricultural Research, Budapest, Hungary*

Max Bergoin
National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada

Sabrina Bertin
Council for Agricultural Research and Economics,
Research Center for Plant Protection and Certification,
Rome, Italy

Shweta Bhatt
University of Copenhagen, Copenhagen, Denmark

Dennis K. Bideshi
California Baptist University, Riverside, CA,
United States and University of California, Riverside,
CA, United States

Yves Bigot
INRAE – French National Research Institute for
Agriculture, Food and Environment, Nouzilly, France

Richard J. Bingham
University of York, York, United Kingdom

Vera Bischoff
Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany

Kate N. Bishop
Francis Crick Institute, London, United Kingdom

Lindsay W. Black
The University of Maryland School of Medicine,
Baltimore, MD, United States

Romain Blanc-Mathieu
Institute for Chemical Research, Kyoto University, Kyoto,
Japan

Soile Blomqvist
National Institute for Health and Welfare, Helsinki,
Finland

Bryony C. Bonning
University of Florida, Gainesville, FL, United States

Lisa M. Bono
Rutgers, The State University of New Jersey,
New Brunswick, NJ, United States

Alexia Bordigoni
Aix-Marseille University, CNRS, IRD, Mediterranean
Institute of Oceanography, Marseille, France and
Aix-Marseille University, IRD257, Assistance-Publique
des Hôpitaux de Marseille, UMR Microbes, Evolution,
Phylogeny and Infections (MEPHI), IHU Méditerranée
Infection, Marseille, France

Mihnea Bostina
University of Otago, Dunedin, New Zealand

Maxime Boutier
University of Liège, Liège, Belgium

P.R. Bowser
Cornell University, Ithaca, NY, United States

Daniel Bradshaw
Public Health England, London, United Kingdom

Claude Bragard
University of Louvain, Louvain-la-Neuve, Belgium

Aaron C. Brault
Centers for Disease Control and Prevention, Fort Collins,
CO, United States

Nicolas Bravo-Vasquez
St. Jude Children's Research Hospital, Memphis, TN,
United States

Rob W. Briddon
University of Agriculture, Faisalabad, Pakistan

Thomas Briebe
Columbia University, New York, NY, United States

Paul Britton
The Pirbright Institute, Pirbright, United Kingdom

Thomas J. Brouwers
Athena Institute, VU Amsterdam, Amsterdam, The
Netherlands

Kevin E. Brown
Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom

Corina P.D. Brussaard
NIOZ Royal Netherlands Institute for Sea Research,
Den Burg, Texel, The Netherlands and Utrecht
University, Utrecht, The Netherlands

Harald Brüssow
Laboratory of Gene Technology, Department of
Biosystems, KU Leuven, Leuven, Belgium

Joachim J. Bugert
Bundeswehr Institute of Microbiology, Munich, Germany

Jozef J. Bujarski
Northern Illinois University, DeKalb, IL, United States
and Polish Academy of Sciences, Poznan, Poland

Laura Burga
University of Otago, Dunedin, New Zealand

Sara H. Burkhard
University Hospital of Zurich, Zurich, Switzerland

Cara C. Burns
Centers for Disease Control and Prevention, Atlanta,
GA, United States

Felicity Burt
University of the Free State, Bloemfontein, South Africa

Kerry S. Burton
Leamington Spa, United Kingdom

Sarah J. Butcher
University of Helsinki, Helsinki, Finland

Mathias Büttner
Leipzig University, Leipzig, Germany

Jesse Cahill
Sandia National Labs, Albuquerque, NM, United States

Marianna Calabretto
Sapienza University of Rome, Rome, Italy

Thierry Candresse
The National Research Institute for Agriculture, Food and the Environment, University of Bordeaux, Villenave d'Ornon, France

Alan J. Cann
University of Leicester, Leicester, United Kingdom

Lorenzo Capucci
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Irene Carlon-Andres
University of Oxford, Oxford, United Kingdom

José M. Casasnovas
National Center for Biotechnology, Spanish National Research Council (CSIC), Madrid, Spain

J.W. Casey
Cornell University, Ithaca, NY, United States

R.N. Casey
Cornell University, Ithaca, NY, United States

Sherwood R. Casjens
University of Utah, Salt Lake City, UT, United States

Antonella Casola
The University of Texas Medical Branch, Galveston, TX, United States

José R. Castón
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Carlos E. Catalano
University of Colorado Anschutz Medical Campus, Skaggs School of Pharmacy and Pharmaceutical Sciences, Aurora, CO, United States

Roberto Cattaneo
Mayo Clinic, Rochester, MN, United States

Patrizia Cavadini
The Lombardy and Emilia Romagna Experimental Zootechnic Institute, Brescia, Italy

Supranee Chaiwatpongsakorn
Nationwide Children's Hospital, Columbus, OH, United States

Supriya Chakraborty
Jawaharlal Nehru University, New Delhi, India

Yu-Chan Chao
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan

Tyler P. Chavers
Centers for Disease Control and Prevention, Atlanta, GA, United States

Keping Chen
Jiangsu University, Zhenjiang, China

Xiaorui Chen
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Yanping Chen
Bee Research Laboratory, Agricultural Research Service, US Department of Agriculture, Beltsville, MD, United States

Dayna Cheng
National Cheng Kung University, Tainan, Taiwan

Quentin Chesnais
University of Strasbourg, Colmar, France

Sotaro Chiba
Nagoya University, Nagoya, Japan

Wah Chiu
Stanford University, Stanford, CA, United States

David Chmielewski
Stanford University, Stanford, CA, United States

Irma E. Cisneros
The University of Texas Medical Branch, Galveston, TX, United States

Lark L. Coffey
University of California, Davis, CA, United States

Alanna B. Cohen
Rutgers University, New Brunswick, NJ, United States

Jeffrey I. Cohen
Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Seth Coleman
Rice University, Houston, TX, United States

Miquel Coll
*Institute for Research in Biomedicine, Barcelona, Spain
 and Institute for Molecular Biology of Barcelona,
 Barcelona, Spain*

John Collinge
*UCL Institute of Prion Diseases, London,
 United Kingdom*

Carina Conceicao
The Pirbright Institute, Pirbright, United Kingdom

Gabriela N. Condezo
*National Center for Biotechnology, Spanish National
 Research Council, Madrid, Spain*

Michaela J. Conley
*MRC-University of Glasgow Centre for Virus Research,
 Glasgow, United Kingdom*

Charles A Coomer
University of Oxford, Oxford, United Kingdom

Anne K. Cordes
*Hannover Medical School, Institute of Virology, Hannover,
 Germany*

Mauricio Cortes Jr.
*Department of Chemistry, College of Arts and Sciences,
 Fort Wayne, IN, United States*

Robert H.A. Coutts
University of Hertfordshire, Hatfield, United Kingdom

Jeff A. Cowley
CSIRO Livestock Industries, Brisbane, QLD, Australia

Robert W. Cross
*The University of Texas Medical Branch, Galveston, TX,
 United States*

Henryk Czosnek
The Hebrew University of Jerusalem, Rehovot, Israel

Håkon Dahle
*Department of Biological Sciences, University of Bergen,
 Bergen, Norway*

Janet M. Daly
*University of Nottingham, Sutton Bonington,
 United Kingdom*

Subha Das
Okayama University, Kurashiki, Japan

Indranil Dasgupta
University of Delhi, New Delhi, India

Sibnarayan Datta
*Defence Research Laboratory, Defence Research and
 Development Organisation (DRDO), Tezpur, Assam,
 India*

Amy Davis
*St Jude Children's Research Hospital, Memphis, TN,
 United States*

William O. Dawson
*Citrus Research and Education Center, Lake Alfred, FL,
 United States and University of Florida, Lake Alfred, FL,
 United States*

Erik De Clercq
*Rega Institute for Medical Research, KU Leuven,
 Leuven, Belgium*

Raoul J. de Groot
Utrecht University, Utrecht, The Netherlands

Juan C. de la Torre
*The Scripps Research Institute, La Jolla, CA,
 United States*

Marcelo De las Heras
University of Zaragoza, Zaragoza, Spain

Juliana Gabriela Silva de Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Athos S. de Oliveira
University of Brasília, Brasília, Brazil

Nicole T. de Stefano
University of Florida, Gainesville, FL, United States

Greg Deakin
NIAB-EMR, East Malling, United Kingdom

Philippe Delfosse
University of Luxembourg, Esch-sur-Alzette, Luxembourg

Natacha Delrez
University of Liège, Liège, Belgium

Tatiana A. Demina
*Molecular and Integrative Biosciences Research Program,
 Faculty of Biological and Environmental Sciences,
 University of Helsinki, Helsinki, Finland*

Ismail Demir
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

Zihni Demirbağ
*Department of Biology, Karadeniz Technical University,
 Trabzon, Turkey*

X. Deng
*Loyola University of Chicago, Maywood, IL,
 United States*

Cécile Desbiez
*Plant Pathology Unit, INRAE – French National
 Research Institute for Agriculture, Food and
 Environment, Montfavet, France*

Christelle Desnues

Aix-Marseille University, CNRS, IRD, Mediterranean Institute of Oceanography, Marseille, France and Aix-Marseille University, IRD 257, Assistance-Publique des Hôpitaux de Marseille, UMR Microbes, Evolution, Phylogeny and Infections (MEPHI), IHU Méditerranée Infection, Marseille, France

Samantha J. DeWerff

University of Illinois at Urbana-Champaign, Urbana, IL, United States

Daniele Di Carlo

Sapienza University of Rome, Rome, Italy

Arturo Diaz

La Sierra University, Riverside, CA, United States

Alfredo Diaz-Lara

University of California, Davis, CA, United States

Ralf G. Dietzgen

The University of Queensland, St. Lucia, QLD, Australia

Michele Digiario

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Michael Dills

Montana State University, Bozeman, MT, United States

Wayne Dimech

National Serology Reference Laboratory, Fitzroy, VIC, Australia

Savithramma P. Dinesh-Kumar

University of California, Davis, CA, United States

Linda K. Dixon

The Pirbright Institute, Pirbright, United Kingdom

Valerian V. Dolja

Oregon State University, Corvallis, OR, United States

Aušra Domanska

University of Helsinki, Helsinki, Finland

Leslie L. Domier

Agricultural Research Service, US Department of Agriculture, Urbana, IL, United States

Pilar Domingo-Calap

Institute for Integrative Systems Biology (I2SysBio), University of Valencia-CSIC, Valencia, Spain

Tatiana Domitrovic

Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Sarah M. Doore

Michigan State University, East Lansing, MI, United States

Lucy Dorrell

University of Oxford, Oxford, United Kingdom

Rosemary A. Dorrington

Rhodes University, Grahamstown, South Africa

Andor Doszpoly

Hungarian Academy of Sciences, Budapest, Hungary

Andreas Dotzauer

University of Bremen, Bremen, Germany

Simon B. Drysdale

St George's University Hospitals NHS Foundation Trust, London, United Kingdom and St George's, University of London, London, United Kingdom

Robert L. Duda

University of Pittsburgh, Pittsburgh, PA, United States

Carol Duffy

University of Alabama, Tuscaloosa, AL, United States

Siobain Duffy

Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

David D. Dunigan

University of Nebraska–Lincoln, Lincoln, NE, United States

Stéphane Duquerroy

University of Paris-Saclay, Orsay, France and Institut Pasteur, Paris, France

Bas E. Dutilh

Utrecht University, Utrecht, The Netherlands and Radboud University Medical Center, Nijmegen, The Netherlands

Michael Edelstein

Faculty of Medicine, Bar Ilan University, Ramat Gan, Israel

Herman K. Edskes

National Institutes of Health, Bethesda, MD, United States

Rosina Ehmann

Bundeswehr Institute of Microbiology, Munich, Germany

Toufic Elbeaino

International Center for Advanced Mediterranean Agronomic Studies (CIHEAM), Mediterranean Agronomic Institute of Bari, Valenzano, Italy

Joanne B. Emerson

University of California, Davis, CA, United States

Ann Emery

University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Christine E. Engeland
*University Hospital Heidelberg and German Cancer
 Research Center, Heidelberg, Germany and
 Witten/Herdecke University, Witten, Germany*

Luis Enjuanes
*National Center for Biotechnology – Spanish National
 Research Council (CNB-CSIC), Madrid, Spain*

Katri Eskelin
University of Helsinki, Helsinki, Finland

Rosa Esteban
*Institute of Biology and Functional Genomics,
 CSIC/University of Salamanca, Salamanca, Spain*

Mary K. Estes
Baylor College of Medicine, Houston, TX, United States

Cassia F. Estofolete
*São José do Rio Preto School of Medicine, São José do
 Rio Preto, Brazil*

Alyssa B. Evans
*National Institutes of Health, Hamilton, MT,
 United States*

Øystein Evensen
Norwegian University of Life Sciences, Oslo, Norway

Alex Evilevitch
*Department of Experimental Medical Science, Lund
 University, Lund, Sweden*

Montserrat Fàbrega-Ferrer
*Institute for Research in Biomedicine, Barcelona, Spain and
 Institute for Molecular Biology of Barcelona, Barcelona,
 Spain*

Francesco Faggioli
*Council for Agricultural Research and Economics,
 Research Center for Plant Protection and Certification,
 Rome, Italy*

Bentley A. Fane
University of Arizona, Tucson, AZ, United States

Brian A. Federici
University of California, Riverside, CA, United States

F. Fenner
Australian National University, Canberra, ACT, Australia

Isabel Fernández de Castro
*Cell Structure Laboratory, National Center for
 Biotechnology – Spanish National Research Council
 (CNB-CSIC), Madrid, Spain*

Giovanni Ferrara
University of Alberta, Edmonton, AB, Canada

Elvira Fiallo-Olivé
*Institute for Mediterranean and Subtropical Horticulture
 “La Mayora”–Spanish National Research Council–University
 of Malaga, Algarrobo-Costa, Málaga, Spain*

Andrew E. Firth
University of Cambridge, Cambridge, United Kingdom

Roland A. Fleck
King’s College London, London, United Kingdom

Ricardo Flores
*Polytechnic University of Valencia, Higher Council of
 Scientific Research, Valencia, Spain*

Ervin Fodor
University of Oxford, Oxford, United Kingdom

Anthony R. Fooks
*Animal and Plant Health Agency, Addlestone,
 United Kingdom; University of Liverpool, Liverpool,
 United Kingdom; and St. George’s Medical School,
 University of London, London, United Kingdom*

Patrick Forterre
*Archeal Virology Unit, Institut Pasteur, Paris, France
 and French National Center for Scientific Research,
 Institute of Integrative Biology of the Cell, University
 of Paris-Saclay, Gif sur Yvette, France*

Rennos Fragkoudis
*University of Nottingham, Sutton Bonington,
 United Kingdom and University of Edinburgh,
 Edinburgh, United Kingdom*

Manuel A. Franco
Pontifical Javeriana University, Bogota, Colombia

Giovanni Franzo
*Department of Animal Medicine, Production and Health
 (MAPS), Padua University, Padua, Italy*

Graham L. Freimanis
The Pirbright Institute, Pirbright, United Kingdom

Juliana Freitas-Astúa
*Brazilian Agricultural Research Corporation (Embrapa)
 Cassava and Fruits, Cruz das Almas, Brazil*

Elizabeth E. Fry
*Department of Structural Biology, Nuffield Department
 of Medicine, University of Oxford, Oxford,
 United Kingdom*

Marc Fuchs
Cornell University, Geneva, NY, United States

Tsutomu Fujimura
*Institute of Biology and Functional Genomics,
 CSIC/University of Salamanca, Salamanca, Spain*

- Kuko Fuke
Tokyo University of Agriculture and Technology, Fuchu, Japan
- Toshiyuki Fukuhara
Tokyo University of Agriculture and Technology, Fuchu, Japan
- To S. Fung
South China Agricultural University, Guangzhou, China
- Yahya Z.A. Gaafar
Julius Kuehn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany
- Toni Gabaldon
Barcelona Supercomputing Center-National Center for Supercomputing, Institute of Research in Biomedicine, and Catalan Institution for Research and Advanced Studies, Barcelona, Spain
- Morgan Gaïa
University of Paris-Saclay, Evry, France
- José Gallardo
National Center for Biotechnology, Spanish National Research Council, Madrid, Spain
- Hernan Garcia-Ruiz
University of Nebraska–Lincoln, Lincoln, NE, United States
- Juan A. García
National Center for Biotechnology-Spanish National Research Council, Madrid, Spain
- Matteo P. Garofalo
The University of Texas Medical Branch, Galveston, TX, United States
- Yves Gaudin
Institute for Integrative Biology of the Cell (I2BC), French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, Paris-Sud University, University of Paris-Saclay, Gif-sur-Yvette, France
- Andrew D.W. Geering
The University of Queensland, St. Lucia, QLD, Australia
- Thomas W. Geisbert
The University of Texas Medical Branch, Galveston, TX, United States
- Andrea Gentili
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy
- Volker Gerdts
University of Saskatchewan, Saskatoon, SK, Canada
- Said A. Ghabrial[†]
Department of Plant Pathology, University of Kentucky, Lexington, KY, United States
- Clément Gilbert
Evolution, Genomes, Behavior and Ecology Laboratory, CNRS University of Paris-Sud UMR 9191, IRD UMR 247, Gif-sur-Yvette, France
- Robert L. Gilbertson
University of California, Davis, CA, United States
- Efstathios S. Giotis
Imperial College London, London, United Kingdom and University of Essex, Colchester, United Kingdom
- Laurent Glais
French Federation of Seed Potato Growers/Research, Development, Promotion of Seed Potato, Paris, France and Institute for Genetics, Environment and Plant Protection, Agrocampus West, French National Institute for Agriculture, Food and Environment, University of Rennes 1, Le Rheu, France
- Miroslav Glasa
Biomedical Research Center, Slovak Academy of Sciences, Bratislava, Slovakia
- Ido Golding
University of Illinois at Urbana-Champaign, Urbana, IL, United States
- Esperanza Gomez-Lucia
Complutense University of Madrid, Madrid, Spain
- Zheng Gong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
- Andrea González-González
University of Florida, Gainesville, FL, United States
- Michael M. Goodin
University of Kentucky, Lexington, KY, United States
- Alexander E. Gorbalenya
Leiden University Medical Center, Leiden, The Netherlands
- Paul Gottlieb
The City University of New York (CUNY), School of Medicine, The City College of New York, New York, NY, United States
- M.-A. Grandbastien
INRAE – French National Research Institute for Agriculture, Food and Environment, Versailles, France

[†]Deceased.

Meritxell Granell
*National Center for Biotechnology, Madrid, Spain and
 Institute of Chemical Research of Catalonia (ICIQ),
 Tarragona, Spain*

Patrick L. Green
*The Ohio State University, Columbus, OH,
 United States*

Sandra J. Greive
University of York, York, United Kingdom

Diane E. Griffin
*Johns Hopkins Bloomberg School of Public Health,
 Baltimore, MD, United States*

Jonathan M. Grimes
University of Oxford, Oxford, United Kingdom

Nigel Grimsley
*Integrative Biology of Marine Organisms Laboratory,
 Banuyls-sur-Mer, France and Sorbonne University,
 Banuyls-sur-Mer, France*

Bruno Gronenborn
*Institute for Integrative Biology of the Cell, CNRS,
 University of Paris-Sud, CEA, Gif sur Yvette, France*

Julianne H. Grose
Brigham Young University, Provo, UT, United States

Scott Grytdal
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

Duane J. Gubler
Duke-NUS Medical School, Singapore, Singapore

Peixuan Guo
*College of Pharmacy, The Ohio State University,
 Columbus, OH, United States*

Tongkun Guo
*Beijing Institute of Genomics, Chinese Academy of
 Sciences, Beijing, China*

Anne-Lise Haenni
*Institut Jacques Monod, French National Center for
 Scientific Research, Paris Diderot University, Paris,
 France*

Susan L. Hafenstein
*Pennsylvania State University, Hershey, PA,
 United States*

Ahmed Hafez
*Biotechvana, Valencia, Spain; Pompeu Fabra University,
 Barcelona, Spain; and Minia University, Minya, Egypt*

Marie Hagbom
Linköping University, Linköping, Sweden

Sébastien Halary
*National Museum of Natural History, UMR 7245
 CNRS/MNHN Molécule de Communication et
 Adaptation des Micro-organismes, Paris, France*

Aron J. Hall
*Centers for Disease Control and Prevention, Atlanta,
 GA, United States*

John Hammond
*Floral and Nursery Plants Research, Agricultural
 Research Service, US Department of Agriculture,
 Beltsville, MD, United States*

Rosemarie W. Hammond
*Agricultural Research Service, US Department of
 Agriculture, Beltsville, MD, United States*

Virginia Hargest
*St Jude Children's Research Hospital, Memphis, TN,
 United States and University of Tennessee Health
 Science Center, Memphis, TN, United States*

Scott J. Harper
Washington State University, Prosser, WA, United States

Balázs Harrach
*Institute for Veterinary Medical Research, Center for
 Agricultural Research, Budapest, Hungary*

Masayoshi Hashimoto
The University of Tokyo, Tokyo, Japan

Muhammad Hassan
University of Agriculture, Faisalabad, Pakistan

Asma Hatoum-Aslan
University of Alabama, Tuscaloosa, AL, United States

Philippa C. Hawes
The Pirbright Institute, Pirbright, United Kingdom

Janelle A. Hayes
*University of Massachusetts Medical School, Worcester,
 MA, United States*

Guijuan He
Virginia Tech, Blacksburg, VA, United States

Klaus Hedman
*University of Helsinki, Helsinki, Finland and Helsinki
 University Hospital, Helsinki, Finland*

Albert Heim
Hannover Medical School, Hanover, Germany

Gary L. Hein
*University of Nebraska–Lincoln, Lincoln, NE,
 United States*

Manfred Heinlein
*IBMP-CNRS, University of Strasbourg, Strasbourg,
 France*

Mercedes Hernando-Pérez
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Carmen Hernández
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Etienne Herrbach
University of Strasbourg, Colmar, France

Stephen Higgs
*Biosecurity Research Institute, Kansas State University,
Manhattan, KS, United States*

Bradley I. Hillman
Rutgers University, New Brunswick, NJ, United States

Deborah M. Hinton
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Judith Hirsch
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Jody Hobson-Peters
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Natalie M. Holste
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Jin S. Hong
Kangwon National University, Chunchon, South Korea

Margaret J. Hosie
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Olivia G. Howell
University of Alabama, Tuscaloosa, AL, United States

Liya Hu
Baylor College of Medicine, Houston, TX, United States

Zhaoyang Hu
Jiangsu University, Zhenjiang, China

Kuan-Ying A. Huang
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Yu Huang
Peking University, Beijing, China

Natalia B. Hubbs
Hanover College, Hanover, IN, United States

Elisabeth Huguet
*Research Institute on Insect Biology, French National
Center for Scientific Research, University of Tours,
Tours, France*

Roger Hull
John Innes Centre, Norwich, United Kingdom

Kiwamu Hyodo
Okayama University, Kurashiki, Japan

Eugénie Hébrard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

Martin Hölzer
University of Jena, Jena, Germany

Tetsuro Ikegami
*The University of Texas Medical Branch at Galveston,
Galveston, TX, United States*

Niina Ikonen
*Finnish Institute for Health and Welfare, Helsinki,
Finland*

Cihan İnan
*Department of Molecular Biology and Genetics,
Karadeniz Technical University, Trabzon, Turkey*

İkbal Agah İnce
*Department of Medical Microbiology, Acibadem
University School of Medicine, Istanbul, Turkey*

Katsuaki Inoue
Diamond Light Source, Didcot, United Kingdom

Toru Iwanami
Tokyo University of Agriculture, Tokyo, Japan

Jacques Izopet
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Fauziah Mohd Jaafar
*UMR1161 Virologie, INRAE – French National
Research Institute for Agriculture, Food and
Environment, ANSES, Ecole Nationale Vétérinaire
d'Alfort, University of Paris-Est, Maisons-Alfort, France*

Andrew O. Jackson
China Agricultural University, Beijing, China

Daral J. Jackwood
*The Ohio State University/OARDC, Wooster, OH,
United States*

Jean-Rock Jacques
*Cellular and Molecular Epigenetics (GIGA), Liège,
 Belgium and Molecular Biology (TERRA), Gembloux,
 Belgium*

Tiffany Jenkins
*Nationwide Children's Hospital, Columbus, OH,
 United States and The Ohio State University, Columbus,
 OH, United States*

Jeffrey D. Jensen
Arizona State University, Tempe, AZ, United States

Daohong Jiang
Huazhong Agricultural University, Wuhan, China

Zhihao Jiang
China Agricultural University, Beijing, China

Allison R. Jilbert
The University of Adelaide, Adelaide, SA, Australia

Peng Jing
*Department of Chemistry, College of Arts and Sciences,
 Fort Wayne, IN, United States*

Xixi Jing
Central South University, Changsha, China

Meesbah Jiwaji
Rhodes University, Grahamstown, South Africa

Kyle L. Johnson
*The University of Texas at El Paso, El Paso, TX,
 United States and CQuentia, Memphis, TN, United
 States*

Welkin E. Johnson
Boston College, Chestnut Hill, MA, United States

Ian M. Jones
*University of Reading, Reading, United Kingdom and
 London School of Hygiene and Tropical Medicine,
 London, United Kingdom*

Ramon Jordan
*Agricultural Research Service, US Department of
 Agriculture, Beltsville, MD, United States*

Thomas Joris
*Cellular and Molecular Epigenetics (GIGA), Liège,
 Belgium and Molecular Biology (TERRA), Gembloux,
 Belgium*

Ilkka Julkunen
*Institute of Biomedicine, University of Turku, Turku,
 Finland*

Sandra Junglen
Charité - University Medicine Berlin, Berlin, Germany

Masanori Kaido
Kyoto University, Kyoto, Japan

Laura Kakkola
University of Turku, Turku, Finland

Hannimari Kallio-Kokko
*University of Helsinki and Helsinki University Hospital,
 Helsinki, Finland*

Nassim Kamar
*Toulouse University Hospital, Toulouse, France and
 Toulouse University Paul Sabatier, Toulouse, France*

Phyllis J. Kanki
*Harvard T.H. Chan School of Public Health, Boston,
 MA, United States*

Peter Karayiannis
University of Nicosia, Nicosia, Cyprus

Henry M. Kariithi
*Kenya Agricultural and Livestock Research Organization,
 Nairobi, Kenya*

Brian A. Kelch
*University of Massachusetts Medical School, Worcester,
 MA, United States*

Karen E. Keller
*Horticultural Crops Research Unit, Agricultural Research
 Service, US Department of Agriculture, Corvallis, OR,
 United States*

Japhette E. Kembou-Ringert
University of Tel Aviv, Tel Aviv, Israel

Peter J. Kerr
*University of Sydney, Sydney, NSW, Australia and
 CSIRO Health and Biosecurity, Black Mountain
 Laboratories, Canberra, ACT, Australia*

Tiffany King
*Nationwide Children's Hospital, Columbus, OH,
 United States and The Ohio State University College
 of Medicine, Columbus, OH, United States*

Andrea Kirmaier
Boston College, Chestnut Hill, MA, United States

Thomas Klose
Purdue University, West Lafayette, IN, United States

Barbara G. Klupp
*Friedrich-Loeffler-Institute, Greifswald-Insel Riems,
 Germany*

David M. Knipe
Harvard Medical School, Boston, MA, United States

Nick J. Knowles
The Pirbright Institute, Pirbright, United Kingdom

Guus Koch
*Wageningen Bioveterinary Research, Lelystad,
 The Netherlands*

Renate Koenig
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

Susanne E. Kohalmi
*The University of Western Ontario, London, ON,
Canada*

Hideki Kondo
Okayama University, Kurashiki, Japan

Jennifer L. Konopka-Anstadt
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Eugene V. Koonin
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States and
National Institutes of Health, Bethesda, MD, United States*

Marion P.G. Koopmans
Erasmus Medical Center, Rotterdam, The Netherlands

Richard Kormelink
*Wageningen University and Research, Wageningen,
The Netherlands*

Ioly Kotta-Loizou
Imperial College London, London, United Kingdom

Peter J. Krell
*Department of Molecular and Cellular Biology,
University of Guelph, Guelph, ON, Canada*

Mart Krupovic
Archaeal Virology Unit, Institut Pasteur, Paris, France

Andreas Kuhn
University of Hohenheim, Stuttgart, Germany

Jens H. Kuhn
*National Institutes of Health, Frederick, MD,
United States*

Richard J. Kuhn
Purdue University, West Lafayette, IN, United States

Suvi Kuivanen
University of Helsinki, Helsinki, Finland

Ranjababu Kulasegaram
*Guy's and St Thomas' NHS Foundation Trust, London,
United Kingdom*

Raghavendran Kulasegaran-Shylini
*Department of Pathogen Infection, Faculty of Infectious
and Tropical Diseases, London School of Hygiene and
Tropical Medicine, London, United Kingdom*

Gaurav Kumar
University of Delhi, New Delhi, India

Manish Kumar
Jawaharlal Nehru University, New Delhi, India

Gael Kurath
*US Geological Survey, Western Fisheries Research
Center, Seattle, WA, United States*

Satu Kurkela
*University of Helsinki and Helsinki University Hospital,
Helsinki, Finland*

Wan-Chun Lai
Chang Gung Memorial Hospital, Taoyuan, Taiwan

Kevin Lamkiewicz
University of Jena, Jena, Germany

Rebecca K. Lane
*University of Texas Health Science Center at San
Antonio, San Antonio, TX, United States*

Andrew S. Lang
*Memorial University of Newfoundland, St. John's, NL,
Canada*

Daniel Carlos Ferreira Lanza
Federal University of Rio Grande do Norte, Natal, Brazil

Maija Lappalainen
*HUS Diagnostic Center, HUSLAB, Clinical
Microbiology, University of Helsinki and Helsinki
University Hospital, Helsinki, Finland*

Katherine LaTourrette
*University of Nebraska–Lincoln, Lincoln, NE,
United States*

Chris Lauber
*TWINCORE – Center for Experimental and Clinical
Infection Research, Hannover, Germany*

Antonio Lavazza
*The Lombardy and Emilia Romagna Experimental
Zootechnic Institute, Brescia, Italy*

C. Martin Lawrence
Montana State University, Bozeman, MT, United States

Hervé Lecoq
*Plant Pathology Unit, INRAE – French National
Research Institute for Agriculture, Food and
Environment, Montfavet, France*

Young-Min Lee
Utah State University, Logan, UT, United States

Kristen N. LeGault
University of California, Berkeley, CA, United States

James Legg
*International Institute of Tropical Agriculture, Dar es
Salaam, Tanzania*

Anne Legreve
University of Louvain, Louvain-la-Neuve, Belgium

Petr G. Leiman
The University of Texas Medical Branch, Galveston, TX, United States

Stanley M. Lemon
Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Sebastian Leptihn
Zhejiang University-Edinburgh University Institute, Zhejiang University, Haining, China

Dennis J. Lewandowski
University of Florida, Lake Alfred, FL, United States

Sébastien Lhomme
Toulouse University Hospital, Toulouse, France and Toulouse University Paul Sabatier, Toulouse, France

Dawei Li
China Agricultural University, Beijing, China

Guangdi Li
Central South University, Changsha, China

Guoqing Li
Huazhong Agricultural University, Wuhan, China

Yi Li
Peking University, Beijing, China

Zhefeng Li
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Zhenghe Li
Zhejiang University, Hangzhou, China

Jia Q. Liang
South China Agricultural University, Guangzhou, China

Sebastian Liebe
Institute of Sugar Beet Research, Göttingen, Germany

João Paulo Matos Santos Lima
Federal University of Rio Grande do Norte, Natal, Brazil

Bruno Lina
HCL Department of Virology, National Reference Center for Respiratory Viruses, Institute of Infectious Agents, Croix-Rousse Hospital, Lyon, France and Virpath Laboratory, International Center of Research in Infectiology (CIRI), INSERM U1111, CNRS—UMR 5308, École Normale Supérieure de Lyon, University Claude Bernard Lyon, Lyon University, Lyon, France

Walter Ian Lipkin
Columbia University, New York, NY, United States

Jan G. Lisby
Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

Ding X. Liu
South China Agricultural University, Guangzhou, China

Qiang Liu
University of Saskatchewan, Saskatoon, SK, Canada

Sijun Liu
Iowa State University, Ames, IA, United States

Carlos Llorens
Biotechvana, Scientific Park University of Valencia, Valencia, Spain

L. Sue Loesch-Fries
Purdue University, West Lafayette, IN, United States

George P. Lomonosoff
John Innes Centre, Norwich, United Kingdom

L. Letti Lopez
The University of Texas at Austin, Austin, TX, United States

Alan T. Loynachan
University of Kentucky, Lexington, KY, United States

Garry A. Luke
University of St. Andrews, St. Andrews, United Kingdom

M. Luo
University of Alabama at Birmingham, Birmingham, AL, United States

Juan J. López-Moya
Center for Research in Agricultural Genomics and Spanish National Research Council, Barcelona, Spain

Che Ma
Genomics Research Center, Academia Sinica, Taipei, Taiwan

Stuart A. MacFarlane
The James Hutton Institute, Invergowrie, United Kingdom

Saichetana Macherla
J. Craig Venter Institute, La Jolla, CA, United States

Kensaku Maejima
The University of Tokyo, Tokyo, Japan

Fabrizio Maggi
University of Pisa, Pisa, Italy and University of Insubria, Varese, Italy

Melissa S. Maginnis
The University of Maine, Orono, ME, United States

Edgar Maiss
Leibniz University Hannover, Hannover, Germany

Kira S. Makarova
National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, United States

Ariana Manglli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Annette Mankertz
Robert Koch-Institute, Berlin, Germany

Pilar Manrique
The Ohio State University, Wexner Medical Center, Columbus, OH, United States

Shahid Mansoor
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Marco Marklewitz
Institute of Virology, Charité – University Medicine Berlin, Berlin, Germany

Giovanni P. Martelli[†]
University of Bari Aldo Moro, Bari, Italy

Darren P. Martin
University of Cape Town, Cape Town, South Africa

Robert R. Martin
Horticultural Crops Research Unit, Agricultural Research Service, US Department of Agriculture, Corvallis, OR, United States

Manuel Martinez-Garcia
University of Alicante, Alicante, Spain

Francisco Martinez-Hernandez
University of Alicante, Alicante, Spain

Natalia Martín-González
Autonomous University of Madrid, Madrid, Spain

Joaquín Martínez Martínez
Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Manja Marz
University of Jena, Jena, Germany

Andrea Marzi
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, United States

Hema Masarapu
Sri Venkateswara University, Tirupati, India

Chikara Masuta
Hokkaido University, Sapporo, Japan

Carlos P. Mata
University of Leeds, Leeds, United Kingdom

Jelle Matthijnsens
Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

Claire P. Mattison
Centers for Disease Control and Prevention, Atlanta, GA, United States and Cherokee Nation Assurance, Arlington, VA, United States

William McAllister
Rowan University School of Osteopathic Medicine, Stratford, NJ, United States

Alison A. McBride
National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, United States

Michael McChesney
University of California, Davis, CA, United States

Elaine McCulloch
Quality Control for Molecular Diagnostics (QCMD), Glasgow, United Kingdom

Andrew J. McMichael
University of Oxford, Oxford, United Kingdom

Alexander McPherson
University of California, Irvine, CA, United States

Irene K. Meki
French National Center for Scientific Research, Montpellier, France

Ulrich Melcher
Oklahoma State University, Stillwater, OK, United States

Tomas A. Melgarejo
University of California, Davis, CA, United States

Michael J. Melzer
Department of Plant and Environmental Protection Sciences, University of Hawaii, Honolulu, HI, United States

Luiza Mendonça
University of Oxford, Oxford, United Kingdom

Xiang-Jin Meng
Virginia Polytechnic Institute and State University, Blacksburg, VA, United States

Peter P.C. Mertens
University of Nottingham, Sutton Bonington, United Kingdom

[†]Deceased.

Thomas C. Mettenleiter
Friedrich-Loeffler-Institute, Greifswald-Insel Riems, Germany

Philip D. Minor
St Albans, United Kingdom

Ali Mirazimi
National Veterinary Institute, Uppsala, Sweden and Karolinska Hospital University, Huddinge, Sweden

Nischay Mishra
Columbia University, New York, NY, United States

Edward S. Mocarski
Emory University School of Medicine, Atlanta, GA, United States

Florian Mock
University of Jena, Jena, Germany

Volker Moennig
University of Veterinary Medicine, Hannover, Germany

Ian J. Molineux
The University of Texas at Austin, Austin, TX, United States

Aderito L. Monjane
Norwegian Veterinary Institute, Oslo, Finland

Jacen S. Moore
University of Tennessee Health Science Center, Memphis, TN, United States

Marc C. Morais
The University of Texas Medical Branch, Galveston, TX, United States

Cristina Moraru
Institute for Chemistry and Biology of the Marine Environment, Oldenburg, Germany

Hiromitsu Moriyama
Tokyo University of Agriculture and Technology, Tokyo, Japan

Sergey Y. Morozov
Lomonosov Moscow State University, Moscow, Russia

Thomas E. Morrison
University of Colorado School of Medicine, Aurora, CO, United States

Léa Morvan
University of Liège, Liège, Belgium

Benoît Moury
Plant Pathology Unit, INRAE – French National Institute for Agriculture, Food and Environment, Montfavet, France

Muhammad Mubin
University of Agriculture, Faisalabad, Pakistan

Nicolas J. Mueller
University Hospital of Zurich, Zurich, Switzerland

Emmanuelle Muller
The French Agricultural Research Center for International Development, Joint Research Units–Biology and Genetics of Plant-Pathogen Interactions, Montpellier, France and Biology and Genetics of Plant-Pathogen Interactions, University of Montpellier, The French Agricultural Research Center for International Development, French National Institute for Agricultural Research, Montpellier SupAgro, Montpellier, France

John S. Munday
Massey University, Palmerston North, New Zealand

Jacob H. Munson-McGee
Montana State University, Bozeman, MT, United States and Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, United States

Hacer Muratoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Kenan C. Murphy
University of Massachusetts Medical School, Worcester, MA, United States

Ugrappa Nagalakshmi
University of California, Davis, CA, United States

Keizo Nagasaki
Kochi University, Nankoku, Japan

Nazia Nahid
GC University, Faisalabad, Pakistan and University of Agriculture, Faisalabad, Pakistan

Venugopal Nair
The Pirbright Institute, Pirbright, United Kingdom

Remziye Nalçacıoğlu
Department of Molecular Biology and Genetics, Karadeniz Technical University, Trabzon, Turkey

Shigetou Namba
The University of Tokyo, Tokyo, Japan

Rubab Z. Naqvi
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Rachel Nash
The Pirbright Institute, Surrey, United Kingdom

C.K. Navaratnarajah
Purdue University, West Lafayette, IN, United States

Maria A. Navarrete-Muñoz
Biotechvana, Madrid, Spain; Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid; and Rey Juan Carlos University Hospital, Móstoles, Spain

Jesús Navas-Castillo
Institute for Mediterranean and Subtropical Horticulture "La Mayora"—Spanish National Research Council—University of Malaga, Algarrobo-Costa, Málaga, Spain

Muhammad S. Nawaz-ul-Rehman
University of Agriculture, Faisalabad, Pakistan

Christopher L. Netherton
The Pirbright Institute, Pirbright, United Kingdom

Thu V.P. Nguyen
Baylor College of Medicine, Houston, TX, United States

Annette Niehl
Julius Kühn Institute – Federal Research Center for Cultivated Plants, Braunschweig, Germany

Hubert G.M. Niesters
Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

Jozef I. Nissimov
University of Waterloo, Waterloo, ON, Canada

Norman Noah
London School of Hygiene and Tropical Medicine, London, United Kingdom

Mauricio L. Nogueira
São José do Rio Preto School of Medicine, São José do Rio Preto, São Paulo, Brazil

Johan Nordgren
Linköping University, Linköping, Sweden

C. Micha Nübling
Paul-Ehrlich-Institute, Langen, Germany

Visa Nurmi
University of Helsinki, Helsinki, Finland

Donald L. Nuss
University of Maryland, Rockville, MD, United States

M. Steven Oberste
Centers for Disease Control and Prevention, Atlanta, GA, United States

Hiroyuki Ogata
Institute for Chemical Research, Kyoto University, Kyoto, Japan

Ane Ogbe
University of Oxford, Oxford, United Kingdom

Hanna M. Oksanen
Molecular and Integrative Biosciences Research Program, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

Graziele P. Oliveira
Federal University of Minas Gerais, Belo Horizonte, Brazil

Francesco Origi
University of Bern, Bern, Switzerland

Nikolaus Osterrieder
Free University of Berlin, Berlin, Germany

Robert A. Owens
Beltsville Agricultural Research Center, Beltsville, MD 20705, United States

Emine Özsahin
University of Guelph, Guelph, ON, Canada

Sergi Padilla-Parra
University of Oxford, Oxford, United Kingdom; Department of Infectious Diseases, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom; and Randall Division of Cell and Molecular Biophysics, King's College London, London, United Kingdom

Joshua Pajak
Duke University, Durham, NC, United States

Massimo Palmarini
MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom

Amanda R. Panfil
The Ohio State University, Columbus, OH, United States

Marcus Panning
Institute of Virology, Freiburg University Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

Vitantonio Pantaleo
National Research Council, Research Unit of Bari, Bari, Italy

Anna Papa
Aristotle University of Thessaloniki, Thessaloniki, Greece

Nikolaos Pappas
Utrecht University, Utrecht, The Netherlands

Hanu R. Pappu
Washington State University, Pullman, WA, United States

Kristin N. Parent
*Michigan State University, East Lansing, MI,
United States*

Colin R. Parrish
Cornell University, Ithaca, NY, United States

A. Lorena Passarelli
Kansas State University, Manhattan, KS, United States

Basavaprabhu L. Patil
*ICAR–Indian Institute of Horticultural Research,
Bengaluru, India*

Jade Pattyn
University of Antwerp, Antwerp, Belgium

T.A. Paul
Cornell University, Ithaca, NY, United States

Lillian Pavlik
*Laboratory for Molecular Virology, Great Lakes Forestry
Centre, Sault Ste Marie, ON, Canada*

Susan L. Payne
Texas A&M University, College Station, TX, United States

Michael N. Pearson
The University of Auckland, Auckland, New Zealand

Mark E. Peeples
*Nationwide Children's Hospital, Columbus, OH,
United States and The Ohio State University College of
Medicine, Columbus, OH, United States*

Ben Peeters
*Wageningen Bioveterinary Research, Lelystad,
The Netherlands*

Joseph S.M. Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Malik Peiris
The University of Hong Kong, Pok Fu Lam, Hong Kong

Judit J. Péntzes
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Miryam Pérez-Cañamás
*Institute for Plant Molecular and Cell Biology (Spanish
National Research Council–Polytechnic University of
Valencia), Valencia, Spain*

Marta Pérez-Illana
*National Center for Biotechnology, Spanish National
Research Council, Madrid, Spain*

Jaume Pérez-Sánchez
*Institute of Aquaculture Torre de la Sal, Spanish
National Research Council, Castellon, Spain*

Jean-Marie Peron
*Toulouse University Hospital, Toulouse, France and
Toulouse University Paul Sabatier, Toulouse, France*

Karin E. Peterson
National Institutes of Health, Hamilton, MT, United States

Karel Petrzik
*Biology Center CAS, Institute of Plant Molecular
Biology, České Budějovice, Czech Republic*

Mahtab Peyambari
*Pennsylvania State University, State College, PA,
United States*

Sujal Phadke
J. Craig Venter Institute, La Jolla, CA, United States

Hanh T. Pham
*National Institute of Scientific Research – Armand-
Frappier Health Research Centre, Laval, QC, Canada*

Mauro Pistello
University of Pisa, Pisa, Italy

Daniel Ponndorf
John Innes Centre, Norwich, United Kingdom

Leo L.M. Poon
The University of Hong Kong, Pok Fu Lam, Hong Kong

Welkin H. Pope
University of Pittsburgh, Pittsburgh, PA, United States

Minna M. Poranen
University of Helsinki, Helsinki, Finland

Claudine Porta
*The Pirbright Institute, Pirbright, United Kingdom and
University of Oxford, Oxford, United Kingdom*

Samuel S. Porter
*National Institute of Allergy and Infectious Diseases,
National Institutes of Health, Bethesda, MD,
United States and University of Maryland, College Park,
MD, United States*

Frank A. Post
*King's College Hospital NHS Foundation Trust, London,
United Kingdom*

Nils Poulicard
*Interactions Plantes Microorganismes Environnement,
Institut de Recherche pour le Développement, Centre de
coopération internationale en recherche agronomique
pour le développement, University of Montpellier,
Montpellier, France*

David Prangishvili
*Institut Pasteur, Paris, France and Ivane Javakhishvili
Tbilisi State University, Tbilisi, Georgia*

B. V. Venkataram Prasad
Baylor College of Medicine, Houston, TX, United States

Lalita Priyamvada
Centers for Disease Control and Prevention, Atlanta, GA, United States

Simone Prospero
Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

Elisabeth Puchhammer-Stöckl
Medical University of Vienna, Vienna, Austria

Jianming Qiu
University of Kansas Medical Center, Kansas City, KS, United States

S.L. Quackenbush
Colorado State University, Fort Collins, CO, United States

Killian J. Quinn
King's College Hospital, London, United Kingdom

Diego F. Quito-Avila
Department of Life Sciences, ESPOL Polytechnic University, Guayaquil, Ecuador

Frank Rabenstein
Julius Kühn Institute, Quedlinburg, Germany

Sheli R. Radoshitzky
United States Army Medical Research Institute of Infectious Diseases, Frederick, MD, United States

Saleem U. Rahman
National Institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan

Mbolarinosa Rakotomalala
FOFIFA, Antananarivo, Madagascar

Norma Rallon
Institute of Health Research-Jiménez Díaz Foundation, Autonomous University of Madrid and Rey Juan Carlos University Hospital, Móstoles, Spain

Robert P. Rambo
Diamond Light Source, Didcot, United Kingdom

Bertha Cecilia Ramirez
The Institute for Integrative Biology of the Cell, The French Alternative Energies and Atomic Energy Commission, French National Center for Scientific Research, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

María D. Ramos-Barbero
University of Alicante, Alicante, Spain

Chris M. Rands
University of Geneva Medical School and Swiss Institute of Bioinformatics, Geneva, Switzerland

Venigalla B. Rao
The Catholic University of America, Washington, DC, United States

Janne J. Ravantti
University of Helsinki, Helsinki, Finland

Mandy Ravensbergen
Wageningen University and Research, Wageningen, The Netherlands

Georget Y. Reaiche-Miller
The University of Adelaide, Adelaide, SA, Australia

D.V.R. Reddy
International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India

Vishwanatha R.A.P. Reddy
The Pirbright Institute, Pirbright, United Kingdom

Juan Reguera
Aix-Marseille University, French National Center for Scientific Research, Marseille, France and French National Institute of Health and Medical Research, Marseille, France

William K. Reisen
University of California, Davis, CA, United States

Jingshan Ren
University of Oxford, Oxford, United Kingdom

Renato O. Resende
University of Brasilia, Brasilia, Brazil

Peter A. Revill
The Peter Doherty Institute of Infection and Immunity, Royal Melbourne Hospital, Melbourne, VIC, Australia

Félix A. Rey
Institut Pasteur, Paris, France

Lara Rheinemann
University of Utah, Salt Lake City, UT, United States

Simone G. Ribeiro
Embrapa Genetic Resources and Biotechnology, Brasília, Brazil

Daniel Rigling
Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Cristina Risco
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Efraín E. Rivera-Serrano

Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, NC, United States and Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, NC, United States

Cécile Robin

INRAE – French National Research Institute for Agriculture, Food and Environment, UMR BIOGECO, Cestas, France

Rodrigo A.L. Rodrigues

Federal University of Minas Gerais, Belo Horizonte, Brazil

Elina Roine

University of Helsinki, Helsinki, Finland

Maria R. Rojas

University of California, Davis, CA, United States

Marilyn J. Roossinck

Pennsylvania State University, State College, PA, United States

Vera I.D. Ros

Wageningen University and Research, Wageningen, The Netherlands

Cristina Rosa

Pennsylvania State University, University Park, PA, United States

Hanna Rose

Leibniz University Hannover, Hannover, Germany

David A. Rosenbaum

University of Florida, Gainesville, FL, United States

Shannan L. Rossi

The University of Texas Medical Branch, Galveston, TX, United States

Michael G. Rossmann[†]

Purdue University, West Lafayette, IN, United States

L. Roux

University of Geneva Medical School, Geneva, Switzerland

Simon Roux

US Department of Energy Joint Genome Institute, Walnut Creek, CA, United States

J. Rovnak

Colorado State University, Fort Collins, CO, United States

David J. Rowlands

University of Leeds, Leeds, United Kingdom

Polly Roy

Department of Pathogen Infection, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom and University of Reading, Reading, United Kingdom

Aaron P. Roznowski

The University of Texas at Austin, Austin, TX, United States and University of Arizona, Tucson, AZ, United States

Luisa Rubino

Institute for Sustainable Plant Protection, National Research Council, Bari, Italy

Olli Ruuskanen

Turku University Hospital, Turku, Finland

Eugene V. Ryabov

USDA, Agricultural Research Service, Beltsville, MD, United States

Martin D. Ryan

University of St. Andrews, St. Andrews, United Kingdom

Ki H. Ryu

Seoul Women's University, Seoul, South Korea

Hanns-Joachim Rziha

Eberhard Karls University of Tübingen, Tübingen, Germany

Sead Sabanadzovic

Mississippi State University, Starkville, MS, United States

Roghayeh Safari

Cellular and Molecular Epigenetics (GIGA), Liège, Belgium and Molecular Biology (TERRA), Gembloux, Belgium

Azeez Sait Sahul Hameed

C. Abdul Hakeem College, Melvisharam, India

Nicole Samies

University of Alabama at Birmingham, Birmingham, AL, United States

Carmen San Martín

National Center for Biotechnology, Spanish National Research Council, Madrid, Spain

Ruth-Anne Sandaa

Department of Biological Sciences, University of Bergen, Bergen, Norway

Hélène Sanfaçon

Agriculture and Agri-Food Canada, Summerland, BC, Canada

[†]Deceased.

Rafael Sanjuán
*Institute for Integrative Systems Biology (I2SysBio),
University of Valencia-CSIC, Valencia, Spain*

Neeraja Sankaran
Utrecht University, Utrecht, The Netherlands

Fernando Santos
University of Alicante, Alicante, Spain

Cecilia Sarmiento
Tallinn University of Technology, Tallinn, Estonia

Takahide Sasaya
*National Agriculture and Food Research Organization,
Fukuyama, Japan*

Preethi Sathanantham
Virginia Tech, Blacksburg, VA, United States

Panayampalli S. Satheshkumar
*Centers for Disease Control and Prevention, Atlanta,
GA, United States*

Yukiyo Sato
Okayama University, Kurashiki, Japan

Andreas Sauerbrei
Jena University Hospital, Jena, Germany

Eugene I. Savenkov
*Swedish University of Agricultural Sciences, Uppsala,
Sweden and Linnean Center for Plant Biology, Uppsala,
Sweden*

Carita Savolainen-Kopra
*National Institute for Health and Welfare, Helsinki,
Finland*

Kay Scheets
*Oklahoma State University, Stillwater, OK,
United States*

Uffe V. Schenider
*Copenhagen University Hospital Hvidovre, Hvidovre,
Denmark*

Richard H. Scheuermann
*J. Craig Venter Institute, La Jolla, CA, United States;
University of California, San Diego, CA, United States;
La Jolla Institute for Immunology, La Jolla, CA,
United States; and Global Virus Network, Baltimore,
MD, United States*

Manfred J. Schmitt
Saarland University, Saarbrücken, Germany

James E. Schoelz
University of Missouri, Columbia, MO, United States

Jason R. Schrad
*Michigan State University, East Lansing, MI,
United States*

Declan C. Schroeder
*University of Reading, Reading, United Kingdom and
University of Minnesota, St. Paul, MN, United States*

Stacey Schultz-Cherry
*St. Jude Children's Research Hospital, Memphis, TN,
United States*

Thomas F. Schulz
*Hannover Medical School, Institute of Virology,
Hannover, Germany and German Center for Infection
Research, Hannover-Braunschweig Site, Braunschweig,
Germany*

Catherine A. Scougall
The University of Adelaide, Adelaide, SA, Australia

Kimberley D. Seed
University of California, Berkeley, CA, United States

Joaquim Segalés
*Departament of Animal Health and Anatomy, Faculty of
Veterinary Medicine, Autonomous University of
Barcelona, Barcelona, Spain; Animal Health Research
Center (CReSA) – Institute of Agrifood Research and
Technology (IRTA), Campus UAB, Barcelona, Spain;
and OIE Collaborating Center for the Research and
Control of Emerging and Re-emerging Swine Diseases in
Europe (IRTA-CReSA), Barcelona, Spain*

Mateo Seoane-Blanco
National Center for Biotechnology, Madrid, Spain

Madhumati Sevvana
Purdue University, West Lafayette, IN, United States

Kazım Sezen
*Department of Biology, Karadeniz Technical University,
Trabzon, Turkey*

Arvind Sharma
Institut Pasteur, Paris, France

Sumit Sharma
Linköping University, Linköping, Sweden

James M. Sharp
*University of Zaragoza, Zaragoza, Spain and Edinburgh,
United Kingdom*

Qunxin She
Shandong University, Qingdao, China

Keith E. Shearwin
The University of Adelaide, Adelaide, SA, Australia

Hanako Shimura
Hokkaido University, Sapporo, Japan

Reina S. Sikkema
Erasmus Medical Center, Rotterdam, The Netherlands

Aaron Simkovich
Agriculture and Agri-Food Canada, London, ON,
Canada and The University of Western Ontario,
London, ON, Canada

Peter Simmonds
University of Oxford, Oxford, United Kingdom

Tarja Sironen
University of Helsinki, Helsinki, Finland

Susanna Sissonen
Finnish Institute for Health and Welfare, Helsinki,
Finland

Michael A. Skinner
Imperial College London, London, United Kingdom

Douglas E. Smith
University of California, San Diego, La Jolla, CA,
United States

Melvyn Smith
Viapath Analytics, Specialist Virology Centre, King's
College NHS Foundation Trust, London, United
Kingdom

Thomas J. Smith
The University of Texas Medical Branch, Galveston, TX,
United States

Teemu Smura
Helsinki University Hospital and University of Helsinki,
Helsinki, Finland

Eric J. Snijder
Leiden University Medical Center, Leiden,
The Netherlands

Gisela Soboll Hussey
Michigan State University, East Lansing, MI,
United States

Maria Söderlund-Venemo
University of Helsinki, Helsinki, Finland

Merike Sõmera
Tallinn University of Technology, Tallinn, Estonia

Eun G. Song
Seoul Women's University, Seoul, South Korea

Milan J. Sonneveld
Erasmus University Medical Center, Rotterdam,
The Netherlands

Beatriz Soriano
Biotechvana, Scientific Park University of Valencia and
Institute for Integrative Systems Biology (I2SysBio),
University of Valencia–Spanish National Research
Council, Valencia, Spain

Thomas E. Spencer
University of Missouri, Columbia, MO, United States

Pothur Sreenivasulu
Sri Venkateswara University, Tirupati, India

Ashley L. St. John
Duke-NUS Medical School, Singapore, Singapore

David K. Stammers
University of Oxford, Oxford, United Kingdom

John Stanley
John Innes Centre, Colney, United Kingdom

Glyn Stanway
University of Essex, Colchester, United Kingdom

Thilo Stehle
University of Tuebingen, Tuebingen, Germany and
Vanderbilt University School of Medicine, Nashville, TN,
United States

Gregory W. Stevenson
Iowa State University, Ames, IA, United States

Lucy Rae Stewart
Agricultural Research Service, US Department of
Agriculture, Wooster, OH, United States

C.C.M.M. Stijger
Wageningen University and Research Center, Bleiswijk,
The Netherlands

Peter G. Stockley
University of Leeds, Leeds, United Kingdom

David Stone
Weymouth Laboratory, Centre for Environment, Fisheries
and Aquaculture Science, Weymouth, United Kingdom

Ashley E Strother
The University of Texas Medical Branch, Galveston, TX,
United States

Sundharraman Subramanian
Michigan State University, East Lansing, MI,
United States

William C. Summers
Yale University, New Haven, CT, United States

Liyang Sun
Northwest A&F University, Yangling, China

Wesley I. Sundquist
University of Utah, Salt Lake City, UT, United States

Petri Susi
University of Turku, Turku, Finland

Curtis A. Suttle
University of British Columbia, Vancouver, BC, Canada

Nobuhiro Suzuki
Institute of Plant Stress and Resources (IPSR), Okayama University, Kurashiki, Japan

Lennart Svensson
Linköping University, Linköping, Sweden and Karolinska Institute, Stockholm, Sweden

Ronald Swanstrom
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Daniele M. Swetnam
University of California, Davis, CA, United states

Moriah L. Szpara
Pennsylvania State University, University Park, PA, United States

Keisuke Tabata
Heidelberg University, Heidelberg, Germany

Anna Taglienti
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Naoki Takeshita
Tokyo University of Agriculture and Technology, Fuchu, Japan

Kana Takeshita Urayama
Tokyo University of Agriculture and Technology, Fuchu, Japan

Michael E. Taliansky
The James Hutton Institute, Dundee, United Kingdom

Pan Tao
The Catholic University of America, Washington, DC, United States

Jacqueline E. Tate
Centers for Disease Control and Prevention, Atlanta, GA, United States

Satyanarayana Tatineni
Agricultural Research Service, US Department of Agriculture, Lincoln, NE, United States and University of Nebraska–Lincoln, Lincoln, NE, United States

Sisko Tauriainen
University of Turku, Turku, Finland

Norbert Tautz
University of Luebeck, Luebeck, Germany

Paulo Tavares
Institute for Integrative Biology of the Cell, CEA, CNRS, University of Paris-Sud, University of Paris-Saclay, Gif-sur-Yvette, France

Nicholas M.I. Taylor
University of Copenhagen, Copenhagen, Denmark

Xu Tengzhi
University of California, Davis, CA, United States

Raquel Tenorio
Cell Structure Laboratory, National Center for Biotechnology – Spanish National Research Council (CNB-CSIC), Madrid, Spain

Robert B. Tesh
The University of Texas Medical Branch, Galveston, TX, United States

Vaskar Thapa
Pennsylvania State University, State College, PA, United States

John E. Thomas
The University of Queensland, Brisbane, QLD, Australia

Julie A. Thomas
Rochester Institute of Technology, Rochester, NY, United States

Lynn C. Thomason
Frederick National Laboratory for Cancer Research, Frederick, MD, United States

Elizabeth Ashley Thompson
The University of Southern Mississippi, Hattiesburg, MS, United States

Jeremy R. Thompson
Cornell University, Ithaca, NY, United States

Antonio Tiberini
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Peter Tijssen
National Institute of Scientific Research – Armand-Frappier Health Research Centre, Microbiology and Immunology, Laval, QC, Canada

Yuji Tomaru
Japan Fisheries Research and Education Agency, Kanagawa, Japan

Laura Tomassoli
Council for Agricultural Research and Economics, Research Center for Plant Protection and Certification, Rome, Italy

Ruben Torres
National Biotechnology Center–Spanish National Research Council, Madrid, Spain

Jia Q. Truong
The University of Adelaide, Adelaide, SA, Australia

- Erkki Truve[†]
Tallinn University of Technology, Tallinn, Estonia
- Chih-Hsuan Tsai
Institute of Molecular Biology, Academia Sinica, Taipei, Taiwan
- Roman Tuma
University of Leeds, Leeds, United Kingdom and University of South Bohemia, České Budějovice, Czech Republic
- Topi Turunen
Infectious Disease Unit, Espoo, Finland and Finnish Institute for Health and Welfare, Helsinki, Finland
- Reidun Twarock
University of York, York, United Kingdom
- Ioannis E. Tzanetakis
University of Arkansas, Fayetteville, United States
- Antti Vaheri
University of Helsinki, Helsinki, Finland
- Eeva J. Vainio
Natural Resources Institute Finland (Luke), Helsinki, Finland
- Anna M. Vaira
Institute for Sustainable Plant Protection, National Research Council of Italy, Torino, Italy
- Steven M. Valles
Center for Medical, Agricultural and Veterinary Entomology, Agricultural Research Service, US Department of Agriculture, Gainesville, FL, United States
- Adrián Valli
National Center for Biotechnology-Spanish National Research Council, Madrid, Spain
- Rodrigo A. Valverde
Louisiana State University Agricultural Center, Baton Rouge, United States
- Pierre Van Damme
University of Antwerp, Antwerp, Belgium
- Rene A.A. van der Vlugt
Wageningen University and Research Center, Wageningen, The Netherlands
- Bernard A.M. Van der Zeijst
Leiden University Medical Center, Leiden, The Netherlands
- Koenraad Van Doorslaer
University of Arizona, Tucson, AZ, United States
- James L. Van Etten
University of Nebraska–Lincoln, Lincoln, NE, United States
- Suzanne van Meer
University Medical Center Utrecht, Utrecht, The Netherlands
- Monique M. van Oers
Wageningen University and Research, Wageningen, The Netherlands
- Mark J. van Raaij
National Center for Biotechnology, Madrid, Spain
- Marc H.V. Van Regenmortel
University of Strasbourg, Strasbourg, France
- Piet A. van Rijn
Wageningen Bioveterinary Research, Lelystad, The Netherlands and North-West University, Potchefstroom, South Africa
- Alain Vanderplasschen
University of Liège, Liège, Belgium
- Dana L. Vanlandingham
College of Veterinary Medicine, Kansas State University, Manhattan, KS, United States
- Olli Vapalahti
Helsinki University Hospital and University of Helsinki, Helsinki, Finland
- Mark Varrelmann
Institute of Sugar Beet Research, Göttingen, Germany
- Nikos Vasilakis
The University of Texas Medical Branch, Galveston, TX, United States
- Michael Veit
Free University of Berlin, Berlin, Germany
- Česlovas Venclovas
Vilnius University, Vilnius, Lithuania
- H. Josef Vetten
Julius Kühn Institute, Braunschweig, Germany
- Marli Vlok
University of British Columbia, Vancouver, BC, Canada
- Anne-Nathalie Volkoff
Diversity, Genomes and Insects-Microorganisms Interactions, National Institute of Agricultural Research, University of Montpellier, Montpellier, France
- Ian E.H. Voorhees
Cornell University, Ithaca, NY, United States
- Alex Vorsters
University of Antwerp, Antwerp, Belgium

[†]Deceased.

Jonathan D.F. Wadsworth
UCL Institute of Prion Diseases, London, United Kingdom

Peter J. Walker
The University of Queensland, St. Lucia, QLD, Australia

Paul Wallace
*Quality Control for Molecular Diagnostics (QCMD),
Glasgow, United Kingdom*

Aiming Wang
*Agriculture and Agri-Food Canada, London, ON,
Canada*

Jen-Ren Wang
National Cheng Kung University, Tainan, Taiwan

Lin-Fa Wang
Duke-NUS Medical School, Singapore, Singapore

Nan Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiangxi Wang
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Xiaofeng Wang
Virginia Tech, Blacksburg, VA, United States

Katherine N. Ward
University College London, London, United Kingdom

Matti Waris
University of Turku, Turku, Finland

Ranjit Warrier
Purdue University, West Lafayette, IN, United States

Daniel Watterson
*Australian Infectious Diseases Research Centre, School of
Chemistry and Molecular Biosciences, The University of
Queensland, St Lucia, QLD, Australia*

Marta L. Wayne
University of Florida, Gainesville, FL, United States

Friedemann Weber
*FB 10 – Institute for Virology, Justus Liebig University
Giessen, Giessen, Germany*

Sung-Chan Wei
*Institute of Molecular Biology, Academia Sinica, Taipei,
Taiwan*

Robin A. Weiss
University College London, London, United Kingdom

Tao Weitao
*Southwest Baptist University, Bolivar, MO,
United States*

Kerstin Wernike
Friedrich-Loeffler-Institute, Insel Riems, Germany

Rachel J. Whitaker
*University of Illinois at Urbana-Champaign, Urbana, IL,
United States*

K. Andrew White
York University, Toronto, ON, Canada

Anna E. Whitfield
*North Carolina State University, Raleigh, NC,
United States*

Richard Whitley
*University of Alabama at Birmingham, Birmingham,
AL, United States*

Reed B. Wickner
*National Institute of Diabetes and Digestive and Kidney
Diseases, National Institutes of Health, Bethesda, MD,
United States*

Luc Willems
*Cellular and Molecular Epigenetics (GIGA), Liège,
Belgium and Molecular Biology (TERRA), Gembloux,
Belgium*

Brian J. Willett
*MRC-University of Glasgow Centre for Virus Research,
Glasgow, United Kingdom*

Alexis Williams
*The University of Texas Medical Branch, Galveston, TX,
United States*

Stephen A. Winchester
*Frimley Park Hospital, Frimley, United Kingdom and
Immunisation and Countermeasures Division, Public
Health England, London, United Kingdom*

Clayton W. Winkler
*National Institutes of Health, Hamilton, MT,
United States*

Stephan Winter
*Leibniz Institute – DSMZ – German Collection of
Microorganisms and Cell Cultures GmbH, Braunschweig,
Germany*

William M. Wintermantel
*Agricultural Research Service, US Department of
Agriculture, Salinas, CA, United States*

Jennifer Wirth
Montana State University, Bozeman, MT, United States

Yuri I. Wolf
*National Center for Biotechnology Information, National
Library of Medicine, Bethesda, MD, United States*

Thorsten Wolff
Robert Koch Institute, Berlin, Germany

Blaide Woodburn
University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Michael E. Woodson
The University of Texas Medical Branch, Galveston, TX, United States

Courtney Woolsey
The University of Texas Medical Branch, Galveston, TX, United States

Chien-Fu Wu
Tokyo University of Agriculture and Technology, Fuchu, Japan

Mingde Wu
Huazhong Agricultural University, Wuhan, China

Songsong Wu
Huazhong Agricultural University, Wuhan, China

Yan Xiang
University of Texas Health Science Center at San Antonio, San Antonio, TX, United States

Jiatao Xie
Huazhong Agricultural University, Wuhan, China

Zhuang Xiong
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Hajime. Yaegashi
Institute of Fruit Tree and Tea Science, NARO, Morioka, Japan

Mehtap Yakupoğlu
Trabzon University, Trabzon, Turkey

Yasuyuki Yamaji
The University of Tokyo, Tokyo, Japan

Meng Yang
China Agricultural University, Beijing, China

Teng-Chieh Yang
Scarsdale, NY, United States

Qin Yao
Jiangsu University, Zhenjiang, China

Tianyou Yao
Baylor College of Medicine, Houston, TX, United States

Nobuyuki Yoshikawa
Iwate University, Morioka, Japan

George R. Young
Francis Crick Institute, London, United Kingdom

Lawrence S. Young
University of Warwick, Coventry, United Kingdom

Mark J. Young
Montana State University, Bozeman, MT, United States

Ry Young
Texas A&M University, College Station, TX, United States

Isaac T. Younker
University of Alabama, Tuscaloosa, AL, United States

Qian Yu
School of Life Sciences, Jiangsu University, Zhenjiang, China

Sang-Im Yun
Utah State University, Logan, UT, United States

Fauzia Zarreen
University of Delhi, New Delhi, India

Francisco M. Zerbini
Federal University of Viçosa, Viçosa, Brazil

Dong-Xiu Zhang
University of Maryland, Rockville, MD, United States

Jianqiang Zhang
Iowa State University, Ames, IA, United States

Junjie Zhang
Texas A&M University, College Station, TX, United States

Long Zhang
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Pan Zhang
Central South University, Changsha, China

Peijun Zhang
University of Oxford, Oxford, United Kingdom and Electron Bio-Imaging Centre, Diamond Light Source, Didcot, United Kingdom

Tao Zhang
Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

Yongliang Zhang
China Agricultural University, Beijing, China

Zhenlu Zhang
Shandong Agricultural University, Tai'an, China

Lixia Zhou
College of Pharmacy, The Ohio State University, Columbus, OH, United States

Ling Zhu
*Institute of Biophysics, Chinese Academy of Sciences,
Beijing, China*

Heiko Ziebell
*Julius Kühn Institute – Federal Research Center for
Cultivated Plants, Braunschweig, Germany*

John Ziebuhr
*The Queen's University of Belfast, Belfast,
United Kingdom*

Jeffrey J. Zimmerman
Iowa State University, Ames, IA, United States

Falk Zucker
*Institute for Chemistry and Biology of the Marine
Environment, Oldenburg, Germany*

CONTENT OF ALL VOLUMES

Editors in Chief	v
Editorial Board	vii
Section Editors	ix
Foreword	xv
Preface	xvii
Guide to Use	xix
List of Contributors	xxi

VOLUME 1

The Virus as a Concept – Fundamentals of Virology

A Brief History of Virology <i>David J Rowlands</i>	3
The Origin of Viruses <i>Patrick Forterre and Morgan Gaïa</i>	14
The Virocell Concept <i>Patrick Forterre</i>	23
Virus Taxonomy <i>Jens H Kuhn</i>	28
The Greater Virus World and Its Evolution <i>Eugene V Koonin and Valerian V Dolja</i>	38
The Virus Species Concept <i>Peter Simmonds</i>	47
Genetic Diversity and Evolution of Viral Populations <i>Rafael Sanjuán and Pilar Domingo-Calap</i>	53
Mechanisms of RNA Virus Evolution <i>Lisa M Bono and Siobain Duffy</i>	62
Mechanisms of DNA Virus Evolution <i>Moriah L Szpara and Koenraad Van Doorslaer</i>	71
Paleovirology <i>Clément Gilbert</i>	79
Evolution Steered by Structure <i>Nicola GA Abrescia</i>	87

Pairwise Sequence Comparison in Virology <i>Tao Zhang, Zheng Gong, Tongkun Guo, Zhuang Xiong, and Yiming Bao</i>	100
Computational Analysis of Recombination in Viral Nucleotide Sequences <i>Miguel Arenas</i>	108
Phylogeny of Viruses <i>Alexander E Gorbalenya and Chris Lauber</i>	116
Virus Bioinformatics <i>Nikolaos Pappas, Simon Roux, Martin Hölzer, Kevin Lamkiewicz, Florian Mock, Manja Marz, and Bas E Dutilh</i>	124
Metagenomics in Virology <i>Simon Roux, Jelle Matthijnsens, and Bas E Dutilh</i>	133
Database and Analytical Resources for Viral Research Community <i>Sujal Phadke, Saichetana Macherla, and Richard H Scheuermann</i>	141
Classification of the Viral World Based on Atomic Level Structures <i>Janne J Ravanntti and Nicola GA Abrescia</i>	153
Isolating, Culturing, and Purifying Viruses With a Focus on Bacterial and Archaeal Viruses <i>Katri Eskelin and Hanna M Oksanen</i>	162
High Throughput Sequencing and Virology <i>Graham L Freimanis and Nick J Knowles</i>	175
Single-Virus Genomics: Studying Uncultured Viruses, One at a Time <i>Manuel Martinez-Garcia, Francisco Martinez-Hernandez, and Joaquín Martínez Martínez</i>	184
Biophysical Characterizations in the Solution State <i>Robert P Rambo and Katsuaki Inoue</i>	191
Virus Crystallography <i>Jonathan M Grimes</i>	199
Advanced Light and Correlative Microscopy in Virology <i>Sergi Padilla-Parra, Charles A Coomer, and Irene Carlon-Andres</i>	208
Atomic Force Microscopy (AFM) Investigation of Viruses <i>Alexander McPherson</i>	218
Cryo-Electron Microscopy (CEM) Structures of Viruses <i>David Chmielewski and Wah Chiu</i>	233
Analysis of Viruses in the Cellular Context by Electron Tomography <i>Peijun Zhang and Luiza Mendonça</i>	242
Mathematical Modeling of Virus Architecture <i>Reidun Twarock</i>	248
Principles of Virus Structure <i>Madhumati Sevvana, Thomas Klose, and Michael G Rossmann[†]</i>	257
Structures of Small Icosahedral Viruses <i>Elizabeth E Fry, Jingshan Ren, and Claudine Porta</i>	278
Structural Principles of the Flavivirus Particle Organization and of Its Conformational Changes <i>Stéphane Duquerroy, Arvind Sharma, and Félix A Rey</i>	290
Reoviruses (Reoviridae) and Their Structural Relatives <i>Liya Hu, Mary K Estes, and B V Venkataram Prasad</i>	303

[†]Deceased.

Structures of Tailed Phages and Herpesviruses (<i>Herpesviridae</i>) <i>Montserrat Fàbrega-Ferrer and Miquel Coll</i>	318
Adenoviruses (<i>Adenoviridae</i>) and Their Structural Relatives <i>Gabriela N Condezo, Natalia Martín-González, Marta Pérez-Illana, Mercedes Hernando-Pérez, José Gallardo, and Carmen San Martín</i>	329
Negative Single-Stranded RNA Viruses (<i>Mononegavirales</i>): A Structural View <i>Juan Reguera</i>	345
Structure of Retrovirus Particles (<i>Retroviridae</i>) <i>David K Stammers and Jingshan Ren</i>	352
Structure of Helical Viruses <i>C Martin Lawrence</i>	362
Giant Viruses and Their Virophage Parasites <i>Rodrigo AL Rodrigues, Ana CdSP Andrade, Grazielle P Oliveira, and Jônatas S Abrahão</i>	372
Viral Replication Cycle <i>AJ Cann</i>	382
Viral Receptors <i>José M Casasnovas and Thilo Stehle</i>	388
Bacterial and Archeal Virus Entry <i>Minna M Poranen and Aušra Domanska</i>	402
Nonenveloped Eukaryotic Virus Entry <i>Ian M Jones and Polly Roy</i>	409
Enveloped Virus Membrane Fusion <i>Aurélien A Albertini and Yves Gaudin</i>	417
Genome Replication of Bacterial and Archaeal Viruses <i>Česlovas Venclovas</i>	429
Viral Transcription <i>David LV Bauer and Ervin Fodor</i>	439
Translation of Viral Proteins <i>Martin D Ryan and Garry A Luke</i>	444
Recombination <i>Jozef J Bujarski</i>	460
Assembly of Viruses: Enveloped Particles <i>CK Navaratnarajah, R Warriar, and RJ Kuhn</i>	468
Assembly of Viruses: Nonenveloped Particles <i>M Luo</i>	475
Virion Assembly: From Small Picornaviruses (<i>Picornaviridae</i>) to Large Herpesviruses (<i>Herpesviridae</i>) <i>Ling Zhu, Nan Wang, and Xiangxi Wang</i>	480
Genome Packaging <i>Richard J Bingham, Reidun Twarock, Carlos P Mata, and Peter G Stockley</i>	488
Virus Factories <i>Isabel Fernández de Castro, Raquel Tenorio, and Cristina Risco</i>	495
Release of Phages From Prokaryotic Cells <i>Jesse Cahill and Ry Young</i>	501
Virus Budding <i>Lara Rheinemann and Wesley I Sundquist</i>	519

Vesicle-Mediated Transcytosis and Export of Viruses <i>Efraín E Rivera-Serrano and Stanley M Lemon</i>	529
Vector Transmission of Animal Viruses <i>Houssam Attoui, Fauziah Mohd Jaafar, Rennos Fragkoudis, and Peter PC Mertens</i>	542
The Human Virome <i>Alexia Bordigoni, Sébastien Halary, and Christelle Desnues</i>	552
Epidemiology of Human and Animal Viral Diseases <i>Michael Edelstein</i>	559
Zoonosis, Emerging and Re-Emerging Viral Diseases <i>Janet M Daly</i>	569
Antiviral Innate Immunity: Introduction <i>Friedemann Weber</i>	577
Humoral and T Cell-Mediated Immunity to Viruses <i>Ane Ogbe and Lucy Dorrell</i>	584
Antigenicity and Antigenic Variation <i>Kuan-Ying A Huang, Xiaorui Chen, Che Ma, Dayna Cheng, Jen-Ren Wang, and Wan-Chun Lai</i>	597
Antigen Presentation <i>Andrew J McMichael</i>	601
Defense Against Viruses and Other Genetic Parasites in Prokaryotes <i>Kira S Makarova, Yuri I Wolf, and Eugene V Koonin</i>	606
Defective-Interfering Viruses <i>L Roux</i>	617
Ecology and Global Impacts of Viruses <i>Joanne B Emerson</i>	621
The Role of Retroviruses in Cellular Evolution <i>Andrea Kirmaier and Welkin E Johnson</i>	627
The Role of Bacteriophages in Bacterial Evolution <i>Chris M Rands and Harald Brüssow</i>	633
Viruses and Their Potential for Bioterrorism <i>Dana L Vanlandingham and Stephen Higgs</i>	644
The Use of Viral Promoters in Expression Vectors <i>Ian M Jones</i>	652
Oncolytic Viruses <i>Laura Burga and Mihnea Bostina</i>	658
Biotechnology Approaches to Modern Vaccine Design <i>George P Lomonosoff and Daniel Ponndorf</i>	662
Viruses: Impact on Science and Society <i>Neeraja Sankaran and Robin A Weiss</i>	671

VOLUME 2

Viruses as Infectious Agents: Human and Animal Viruses

Adenoviruses (<i>Adenoviridae</i>) <i>Balázs Harrach and Mária Benkő</i>	3
---	---

African Horse Sickness Virus (<i>Reoviridae</i>) <i>Piet A van Rijn</i>	17
African Swine Fever Virus (<i>Asfarviridae</i>) <i>Linda K Dixon, Rachel Nash, Philippa C Hawes, and Christopher L Netherton</i>	22
Akabane Virus and Schmallenberg Virus (<i>Peribunyaviridae</i>) <i>Martin Beer and Kerstin Wernike</i>	34
Alphaviruses Causing Encephalitis (<i>Togaviridae</i>) <i>Diane E Griffin</i>	40
Anelloviruses (<i>Anelloviridae</i>) <i>Fabrizio Maggi and Mauro Pistello</i>	48
Animal Lentiviruses (<i>Retroviridae</i>) <i>Esperanza Gomez-Lucia</i>	56
Animal Morbilliviruses (<i>Paramyxoviridae</i>) <i>Carina Conceicao and Dalan Bailey</i>	68
Animal Papillomaviruses (<i>Papillomaviridae</i>) <i>John S Munday</i>	79
Astroviruses (<i>Astroviridae</i>) <i>Virginia Hargest, Amy Davis, and Stacey Schultz-Cherry</i>	92
Avian Hepadnaviruses (<i>Hepadnaviridae</i>) <i>Allison R Jilbert, Georget Y Reaiche-Miller, and Catherine A Scougall</i>	100
Avian Herpesviruses (<i>Herpesviridae</i>) <i>Vishwanatha RAP Reddy and Venugopal Nair</i>	112
Avian Influenza Viruses (<i>Orthomyxoviridae</i>) <i>Nicolas Bravo-Vasquez and Stacey Schultz-Cherry</i>	117
Avian Leukosis and Sarcoma Viruses (<i>Retroviridae</i>) <i>Karen L Beemon</i>	122
Bluetongue Virus (<i>Reoviridae</i>) <i>Raghavendran Kulasegaran-Shylini and Polly Roy</i>	127
Borna Disease Virus and Related Bornaviruses (<i>Bornaviridae</i>) <i>Susan L Payne</i>	137
Bovine Leukemia Virus (<i>Retroviridae</i>) <i>Thomas Joris, Roghaiyeh Safari, Jean-Rock Jacques, and Luc Willems</i>	144
Bovine Viral Diarrhea, Border Disease, and Classical Swine Fever Viruses (<i>Flaviviridae</i>) <i>Paul Becher, Volker Moennig, and Norbert Tautz</i>	153
Capripoxviruses, Parapoxviruses, and Other Poxviruses of Ruminants (<i>Poxviridae</i>) <i>Philippa M Beard</i>	165
Chikungunya Virus (<i>Togaviridae</i>) <i>Thomas E Morrison and Stephanie E Ander</i>	173
Circoviruses (<i>Circoviridae</i>) <i>Giovanni Franzo and Joaquim Segalés</i>	182
Coronaviruses: General Features (<i>Coronaviridae</i>) <i>Paul Britton</i>	193
Coronaviruses: Molecular Biology (<i>Coronaviridae</i>) <i>X Deng and SC Baker</i>	198

Crimean-Congo Hemorrhagic Fever Virus and Nairoviruses of Medical Importance (<i>Nairoviridae</i>) <i>Ali Mirazimi, Felicity Burt, and Anna Papa</i>	208
Dengue Viruses (<i>Flaviviridae</i>) <i>Ashley L. St. John and Duane J. Gubler</i>	218
Ebola Virus (<i>Filoviridae</i>) <i>Andrea Marzi and Logan Banadyga</i>	232
Enteroviruses (<i>Picornaviridae</i>) <i>Carita Savolainen-Kopra, Soile Blomqvist, and Petri Susi</i>	245
Enveloped, Positive-Strand RNA Viruses (<i>Nidovirales</i>) <i>L. Enjuanes, AE Gorbalenya, RJ de Groot, JA Cowley, J Ziebuhr, and EJ Snijder</i>	256
Epstein-Barr Virus (<i>Herpesviridae</i>) <i>Lawrence S. Young</i>	267
Equine Herpesviruses (<i>Herpesviridae</i>) <i>Gisela Soboll-Hussey, Nikolaus Osterrieder, and Walid Azab</i>	278
Equine, Canine, and Swine Influenza (<i>Orthomyxoviridae</i>) <i>Janet M. Daly and Japhette E. Kembou-Ringert</i>	287
Feline Calicivirus (<i>Caliciviridae</i>) <i>Margaret J. Hosie and Michaela J. Conley</i>	294
Feline Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>Brian J. Willett and Margaret J. Hosie</i>	300
Fish and Amphibian Alloherpesviruses (<i>Herpesviridae</i>) <i>Maxime Boutier, Léa Morvan, Natacha Delrez, Francesco Origi, Andor Doszpoly, and Alain Vanderplasschen</i>	306
Fish Retroviruses (<i>Retroviridae</i>) <i>TA Paul, RN Casey, PR Bowser, JW Casey, J Rovnak, and SL Quackenbush</i>	316
Fish Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Gael Kurath and David Stone</i>	324
Foot-and-Mouth Disease Viruses (<i>Picornaviridae</i>) <i>David J. Rowlands</i>	332
Fowlpox Virus and Other Avipoxviruses (<i>Poxviridae</i>) <i>Efstathios S. Giotis and Michael A. Skinner</i>	343
Hantaviruses (<i>Hantaviridae</i>) <i>Tarja Sironen and Antti Vaheri</i>	349
Henipaviruses (<i>Paramyxoviridae</i>) <i>Lin-Fa Wang and Danielle E. Anderson</i>	355
Hepatitis A Virus (<i>Picornaviridae</i>) <i>Andreas Dotzauer</i>	362
Hepatitis B Virus (<i>Hepadnaviridae</i>) <i>Peter Karayiannis</i>	373
Hepatitis C Virus (<i>Flaviviridae</i>) <i>Ralf Bartenschlager and Keisuke Tabata</i>	386
Hepeviruses (<i>Hepeviridae</i>) <i>Xiang-Jin Meng</i>	397
Herpes Simplex Virus 1 and 2 (<i>Herpesviridae</i>) <i>David M. Knipe and Richard Whitley</i>	404

History of Virology: Vertebrate Viruses <i>F Fenner</i>	414
Human Boca- and Protoparvoviruses (<i>Parvoviridae</i>) <i>Maria Söderlund-Venermo and Jianming Qiu</i>	419
Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (<i>Coronaviridae</i>) <i>Ding X Liu, Jia Q Liang, and To S Fung</i>	428
Human Cytomegalovirus (<i>Herpesviridae</i>) <i>Edward S Mocarski</i>	441
Human Immunodeficiency Virus (<i>Retroviridae</i>) <i>Blaide Woodburn, Ann Emery, and Ronald Swanstrom</i>	460
Human Metapneumovirus (<i>Pneumoviridae</i>) <i>Antonella Casola, Matteo P Garofalo, and Xiaoyong Bao</i>	475
Human Norovirus and Sapovirus (<i>Caliciviridae</i>) <i>Sumit Sharma, Marie Hagbom, Lennart Svensson, and Johan Nordgren</i>	483
Human Papillomaviruses (<i>Papillomaviridae</i>) <i>Alison A McBride and Samuel S Porter</i>	493
Human Parainfluenza Viruses (<i>Paramyxoviridae</i>) <i>Elisabeth Adderson</i>	502
Human Pathogenic Arenaviruses (<i>Arenaviridae</i>) <i>Sheli R Radoshitzky and Juan C de la Torre</i>	507
Human Polyomaviruses (<i>Papillomaviridae</i>) <i>Melissa S Maginnis</i>	518
Human T-Cell Leukemia Virus-1 and -2 (<i>Retroviridae</i>) <i>Amanda R Panfil and Patrick L Green</i>	528
Infectious Bursal Disease Virus (<i>Birnaviridae</i>) <i>Daral J Jackwood</i>	540
Infectious Pancreatic Necrosis Virus (<i>Birnaviridae</i>) <i>Øystein Evensen</i>	544
Influenza A Viruses (<i>Orthomyxoviridae</i>) <i>Laura Kakkola, Niina Ikonen, and Ilkka Julkunen</i>	551
Influenza B, C and D Viruses (<i>Orthomyxoviridae</i>) <i>Thorsten Wolff and Michael Veit</i>	561
Jaagsiekte Sheep Retrovirus (<i>Retroviridae</i>) <i>James M Sharp, Marcelo De las Heras, Massimo Palmarini, and Thomas E Spencer</i>	575
Japanese Encephalitis Virus (<i>Flaviviridae</i>) <i>Sang-Im Yun and Young-Min Lee</i>	583
Kaposi's Sarcoma-Associated Herpesvirus (<i>Herpesviridae</i>) <i>Anne K Cordes and Thomas F Schulz</i>	598
Marburg and Ravn Viruses (<i>Filoviridae</i>) <i>Courtney Woolsey, Thomas W Geisbert, and Robert W Cross</i>	608
Measles Virus (<i>Paramyxoviridae</i>) <i>Roberto Cattaneo and Michael McChesney</i>	619
Molluscum Contagiosum Virus (<i>Poxviridae</i>) <i>Joachim J Bugert and Rosina Ehmann</i>	629

Mumps Virus (<i>Paramyxoviridae</i>) <i>Stephen A Winchester and Kevin E Brown</i>	634
Murine Leukemia and Sarcoma Viruses (<i>Retroviridae</i>) <i>George R Young and Kate N Bishop</i>	643
Newcastle Disease Virus (<i>Paramyxoviridae</i>) <i>Ben Peeters and Guus Koch</i>	648
Orthobunyaviruses (<i>Peribunyaviridae</i>) <i>Alyssa B Evans, Clayton W Winkler, and Karin E Peterson</i>	654
Parapoxviruses (<i>Poxviridae</i>) <i>Hanns-Joachim Rziha and Mathias Büttner</i>	666
Parechoviruses (<i>Picornaviridae</i>) <i>Sisko Tauriainen and Glyn Stanway</i>	675
Parvoviruses of Carnivores, and the Emergence of Canine Parvovirus (<i>Parvoviridae</i>) <i>Colin R Parrish, Ian EH Voorhees, and Susan L Hafenstein</i>	683
Polioviruses (<i>Picornaviridae</i>) <i>Philip D Minor</i>	688
Porcine Reproductive and Respiratory Syndrome Virus and Equine Arteritis Virus (<i>Arteriviridae</i>) <i>Jianqiang Zhang, Alan T Loynachan, Gregory W Stevenson, and Jeffrey J Zimmerman</i>	697
Prions of Vertebrates <i>Jonathan DF Wadsworth and John Collinge</i>	707
Pseudorabies Virus (<i>Herpesviridae</i>) <i>Thomas C Mettenleiter and Barbara G Klupp</i>	714
Rabbit Hemorrhagic Disease Virus and European Brown Hare Syndrome Virus (<i>Caliciviridae</i>) <i>Lorenzo Capucci, Patrizia Cavadini, and Antonio Lavazza</i>	724
Rabbit Myxoma Virus and the Fibroma Viruses (<i>Poxviridae</i>) <i>Peter J Kerr</i>	730
Rabies and Other Lyssaviruses (<i>Rhabdoviridae</i>) <i>Ashley C Banyard and Anthony R Fooks</i>	738
Respiratory Syncytial Virus (<i>Pneumoviridae</i>) <i>Tiffany King, Tiffany Jenkins, Supranee Chaiwatpongsakorn, and Mark E Peebles</i>	747
Rhinoviruses (<i>Picornaviridae</i>) <i>Matti Waris and Olli Ruuskanen</i>	757
Rift Valley Fever Virus and Other Phleboviruses (<i>Phenuiviridae</i>) <i>Tetsuro Ikegami</i>	765
Roseoloviruses: Human Herpesviruses 6A, 6B and 7 (<i>Herpesviridae</i>) <i>Katherine N Ward</i>	778
Rotaviruses (<i>Reoviridae</i>) <i>Juana Angel and Manuel A Franco</i>	789
Rubella Virus (<i>Picornaviridae</i>) <i>Annette Mankertz</i>	797
Saint Louis Encephalitis Virus (<i>Flaviviridae</i>) <i>William K Reisen, Lark L Coffey, Daniele M Swetnam, and Aaron C Brault</i>	805
Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS) (<i>Coronaviridae</i>) <i>Joseph SM Peiris and Leo LM Poon</i>	814

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (<i>Coronaviridae</i>) <i>Malik Peiris</i>	825
Simian Immunodeficiency Virus (SIV) and HIV-2 (<i>Retroviridae</i>) <i>Phyllis J Kanki</i>	827
Sindbis Virus (<i>Togaviridae</i>) <i>Satu Kurkela</i>	837
Tick-Borne Encephalitis Virus (<i>Flaviviridae</i>) <i>Teemu Smura, Suvi Kuivanen, and Olli Vapalahti</i>	843
Transmissible Gastroenteritis Virus of Pigs and Porcine Epidemic Diarrhea Virus (<i>Coronaviridae</i>) <i>Qiang Liu and Volker Gerds</i>	850
Vaccinia Virus (<i>Poxviridae</i>) <i>Yan Xiang and Rebecca K Lane</i>	854
Varicella-Zoster Virus (<i>Herpesviridae</i>) <i>Jeffrey I Cohen</i>	860
Variola and Monkeypox Viruses (<i>Poxviridae</i>) <i>Lalita Priyamvada and Panayampalli S Satheshkumar</i>	868
Vesicular Stomatitis Virus and Bovine Ephemeral Fever Virus (<i>Rhabdoviridae</i>) <i>Peter J Walker and Robert B Tesh</i>	875
West Nile Virus (<i>Flaviviridae</i>) <i>Fengwei Bai and Elizabeth Ashley Thompson</i>	884
Yellow Fever Virus (<i>Flaviviridae</i>) <i>Ashley E Strother and Alan DT Barrett</i>	891
Zika Virus (<i>Flaviviridae</i>) <i>Nikos Vasilakis, Shannan L Rossi, Sasha R Azar, Irma E Cisneros, Cassia F Estofolete, and Mauricio L Nogueira</i>	899

VOLUME 3

Viruses as Infectious Agents: Plant Viruses

An Introduction to Plant Viruses <i>Roger Hull</i>	3
Emerging and Re-Emerging Plant Viruses <i>Sabrina Bertin, Francesco Faggioli, Andrea Gentili, Ariana Manglli, Anna Taglienti, Antonio Tiberini, and Laura Tomassoli</i>	8
Emerging Geminiviruses (<i>Geminiviridae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, and Muhammad Mubin</i>	21
Movement of Viruses in Plants <i>Manfred Heinlein</i>	32
Plant Antiviral Defense: Gene-Silencing Pathways <i>Vitantonio Pantaleo, Chikara Masuta, and Hanako Shimura</i>	43
Plant Resistance to Viruses: Engineered Resistance <i>Marc Fuchs</i>	52
Plant Resistance to Viruses: Natural Resistance Associated With Dominant Genes <i>Mandy Ravensbergen and Richard Kormelink</i>	60

Plant Resistance to Viruses: Natural Resistance Associated With Recessive Genes <i>Masayoshi Hashimoto, Kensaku Maejima, Yasuyuki Yamaji, and Shigetou Namba</i>	69
Plant Viral Diseases: Economic Implications <i>Basavaprabhu L Patil</i>	81
Retrotransposons of Plants <i>M-A Grandbastien</i>	98
Vector Transmission of Plant Viruses <i>Etienne Herrbach and Quentin Chesnais</i>	106
Viral Suppressors of Gene Silencing <i>Hernan Garcia-Ruiz</i>	116
Virus-Induced Gene Silencing (VIGS) <i>Xu Tengzhi, Ugrappa Nagalakshmi, and Savithramma P Dinesh-Kumar</i>	123
Alfalfa Mosaic Virus (<i>Bromoviridae</i>) <i>L Sue Loesch-Fries</i>	132
Alphaflexiviruses (<i>Alphaflexiviridae</i>) <i>Sergey Y Morozov and Alexey A Agranovsky</i>	140
Alphasatellites (<i>Alphasatellitidae</i>) <i>Rob W Briddon and Muhammad S Nawaz-ul-Rehman</i>	149
Amalgaviruses (<i>Amalgaviridae</i>) <i>Ioannis E Tzanetakis, Sead Sabanadzovic, and Rodrigo A Valverde</i>	154
Badnaviruses (<i>Caulimoviridae</i>) <i>Andrew DW Geering</i>	158
Banana Bunchy Top Virus (<i>Nanoviridae</i>) <i>John E Thomas</i>	169
Barley Yellow Dwarf Viruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	176
Bean Common Mosaic Virus and Bean Common Mosaic Necrosis Virus (<i>Potyviridae</i>) <i>Ramon Jordan and John Hammond</i>	184
Bean Golden Mosaic Virus and Bean Golden Yellow Mosaic Virus (<i>Geminiviridae</i>) <i>Francisco M Zerbini and Simone G Ribeiro</i>	192
Beet Curly Top Virus (<i>Geminiviridae</i>) <i>Robert L Gilbertson, Tomas A Melgarejo, Maria R Rojas, William M Wintermantel, and John Stanley</i>	200
Beet Necrotic Yellow Vein Virus (<i>Benyviridae</i>) <i>Sebastian Liebe, Annette Niehl, Renate Koenig, and Mark Varrelmann</i>	213
Benyviruses (<i>Benyviridae</i>) <i>Annette Niehl, Sebastian Liebe, Mark Varrelmann, and Renate Koenig</i>	219
Betaflexiviruses (<i>Betaflexiviridae</i>) <i>Nobuyuki Yoshikawa and Hajime Yaegashi</i>	229
Betasatellites and Deltasatellites (<i>Tolecasatellitidae</i>) <i>Muhammad S Nawaz-ul-Rehman, Nazia Nahid, Muhammad Hassan, and Muhammad Mubin</i>	239
Bluner-, Cile-, and Higreviruses (<i>Kitaviridae</i>) <i>Diego F Quito-Avila, Juliana Freitas-Astúa, and Michael J Melzer</i>	247
Brome Mosaic Virus (<i>Bromoviridae</i>) <i>Guijuan He, Zhenlu Zhang, Preethi Sathanantham, Arturo Diaz, and Xiaofeng Wang</i>	252

Bromoviruses (<i>Bromoviridae</i>) <i>Jozef J Bujarski</i>	260
Bymoviruses (<i>Potyviridae</i>) <i>Annette Niehl and Frank Rabenstein</i>	268
Cacao Swollen Shoot Virus (<i>Caulimoviridae</i>) <i>Emmanuelle Muller</i>	274
Carmo-Like Viruses (<i>Tombusviridae</i>) <i>Miryam Pérez-Cañamás and Carmen Hernández</i>	285
Cassava Brown Streak Viruses (<i>Potyviridae</i>) <i>Basavaprabhu L Patil</i>	293
Cassava Mosaic Viruses (<i>Geminiviridae</i>) <i>James Legg and Stephan Winter</i>	301
Caulimoviruses (<i>Caulimoviridae</i>) <i>James E Schoelz and Mustafa Adhab</i>	313
Cheraviruses, Sadwaviruses and Torradoviruses (<i>Secoviridae</i>) <i>Toru Iwanami and René AA van der Vlugt</i>	322
Citrus Tristeza Virus (<i>Closteroviridae</i>) <i>Moshe Bar-Joseph, Scott J Harper, and William O Dawson</i>	327
Closteroviruses (<i>Closteroviridae</i>) <i>Marc Fuchs</i>	336
Comoviruses and Fabaviruses (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	348
Cotton Leaf Curl Disease (<i>Geminiviridae</i>) <i>Nasim Ahmed, Imran Amin, and Shahid Mansoor</i>	355
Cowpea Mosaic Virus (<i>Secoviridae</i>) <i>George P Lomonossoff</i>	364
Cucumber Mosaic Virus (<i>Bromoviridae</i>) <i>Judith Hirsch and Benoît Moury</i>	371
Dianthovirus (<i>Tombusviridae</i>) <i>Kiwamu Hyodo and Masanori Kaido</i>	383
Endornaviruses (<i>Endornaviridae</i>) <i>Toshiyuki Fukuhara</i>	388
Fimoviruses (<i>Fimoviridae</i>) <i>Toufic Elbeaino and Michele Digiaro</i>	396
Furoviruses (<i>Virgaviridae</i>) <i>Annette Niehl and Renate Koenig</i>	405
Geminiviruses (<i>Geminiviridae</i>) <i>Jesús Navas-Castillo and Elvira Fiallo-Olivé</i>	411
Hordeiviruses (<i>Virgaviridae</i>) <i>Zhihao Jiang, Meng Yang, Yongliang Zhang, Andrew O Jackson, and Dawei Li</i>	420
Idaeoviruses (<i>Mayoviridae</i>) <i>Robert R Martin and Karen E Keller</i>	430
Ilarviruses (<i>Bromoviridae</i>) <i>Aaron Simkovich, Susanne E Kohalmi, and Aiming Wang</i>	439

Luteoviruses (<i>Luteoviridae</i>) <i>Leslie L Domier</i>	447
Machlomovirus and Panicoviruses (<i>Tombusviridae</i>) <i>Kay Scheets</i>	456
Maize Streak Virus (<i>Geminiviridae</i>) <i>Darren P Martin and Aderito L Monjane</i>	461
Nanoviruses (<i>Nanoviridae</i>) <i>Bruno Gronenborn and H Josef Vetten</i>	470
Necro-Like Viruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Giovanni P Martelli[†]</i>	481
Nepoviruses (<i>Secoviridae</i>) <i>Hélène Sanfaçon</i>	486
Ophioviruses (<i>Aspiviridae</i>) <i>Anna M Vaira and John Hammond</i>	495
Orthospoviruses (<i>Tospoviridae</i>) <i>Renato O Resende and Hanu R Pappu</i>	507
Ourmiaviruses (<i>Botourmiaviridae</i>) <i>Gian Paolo Accotto and Cristina Rosa</i>	516
Papaya Ringspot Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	520
Pecluviruses (<i>Virgaviridae</i>) <i>Hema Masarapu, Pothur Sreenivasulu, Philippe Delfosse, Claude Bragard, Anne Legreve, and DVR Reddy</i>	528
Pepino Mosaic Virus (<i>Alphaflexiviridae</i>) <i>Rene AA van der Vlugt and CCMM Stijger</i>	539
Plant Reoviruses (<i>Reoviridae</i>) <i>Yu Huang and Yi Li</i>	545
Plant Resistance to Geminiviruses <i>Basavaprabhu L Patil, Supriya Chakraborty, Henryk Czosnek, Elvira Fiallo-Olivé, Robert L Gilbertson, James Legg, Shahid Mansoor, Jesús Navas-Castillo, Rubab Z Naqvi, Saleem U Rahman, and Francisco M Zerbini</i>	554
Plant Rhabdoviruses (<i>Rhabdoviridae</i>) <i>Ralf G Dietzgen, Michael M Goodin, and Zhenghe Li</i>	567
Plant Satellite Viruses (<i>Albetovirus, Aumaivirus, Papanivirus, Virtovirus</i>) <i>Mart Krupovic</i>	581
Plum Pox Virus (<i>Potyviridae</i>) <i>Miroslav Glasa and Thierry Candresse</i>	586
Poleroviruses (<i>Luteoviridae</i>) <i>Hernan Garcia-Ruiz, Natalie M Holste, and Katherine LaTourrette</i>	594
Pomoviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	603
Potato Virus Y (<i>Potyviridae</i>) <i>Laurent Glais and Benoît Moury</i>	612
Potexviruses (<i>Alphaflexiviridae</i>) <i>Ki H Ryu, Eun G Song, and Jin S Hong</i>	623

[†]Deceased.

Potyviruses (<i>Potyviridae</i>) <i>Adrián Valli, Juan A García, and Juan J López-Moya</i>	631
Quinviruses (<i>Betaflexiviridae</i>) <i>Ki H Ryu and Eun G Song</i>	642
Reverse-Transcribing Viruses (<i>Belpaoviridae, Metaviridae, and Pseudoviridae</i>) <i>Carlos Llorens, Beatriz Soriano, Maria A Navarrete-Muñoz, Ahmed Hafez, Vicente Arnau, Jose Miguel Benito, Toni Gabaldon, Norma Rallon, Jaume Pérez-Sánchez, and Mart Krupovic</i>	653
Rice Tungro Disease (<i>Secoviridae, Caulimoviridae</i>) <i>Gaurav Kumar, Fauzia Zarreen, and Indranil Dasgupta</i>	667
Rice Yellow Mottle Virus (<i>Solemoviridae</i>) <i>Eugénie Hébrard, Nils Poulicard, and Mbolarinosy Rakotomalala</i>	675
Satellite Nucleic Acids and Viruses <i>Olufemi J Alabi, Alfredo Diaz-Lara, and Maher Al Rwahnih</i>	681
Secoviruses (<i>Secoviridae</i>) <i>Jeremy R Thompson</i>	692
Sequiviruses and Waikaviruses (<i>Secoviridae</i>) <i>Lucy Rae Stewart</i>	703
Solemoviruses (<i>Solemoviridae</i>) <i>Cecilia Sarmiento, Merike Sõmera, and Erkki Truve[†]</i>	712
Tenuiviruses (<i>Phenuiviridae</i>) <i>Bertha Cecilia Ramirez and Anne-Lise Haenni</i>	719
Tobacco Mosaic Virus (<i>Virgaviridae</i>) <i>Marc HV Van Regenmortel</i>	727
Tobamoviruses (<i>Virgaviridae</i>) <i>Ulrich Melcher, Dennis J Lewandowski, and William O Dawson</i>	734
Tobraviruses (<i>Virgaviridae</i>) <i>Stuart A MacFarlane</i>	743
Tomato Leaf Curl New Delhi Virus (<i>Geminiviridae</i>) <i>Supriya Chakraborty and Manish Kumar</i>	749
Tomato Spotted Wilt Virus (<i>Tospoviridae</i>) <i>Hanu R Pappu, Anna E Whitfield, and Athos S de Oliveira</i>	761
Tomato Yellow Leaf Curl Viruses (<i>Geminiviridae</i>) <i>Henryk Czosnek</i>	768
Tombusvirus-Like Viruses (<i>Tombusviridae</i>) <i>K Andrew White</i>	778
Tombusviruses (<i>Tombusviridae</i>) <i>Luisa Rubino and Kay Scheets</i>	788
Tritimoviruses and Rymoviruses (<i>Potyviridae</i>) <i>Satyanarayana Tatineni and Gary L Hein</i>	797
Triviruses (<i>Betaflexiviridae</i>) <i>Yahya ZA Gaafar and Heiko Ziebell</i>	805
Tymoviruses (<i>Tymoviridae</i>) <i>Rosemarie W Hammond and Peter Abrahamian</i>	818

[†]Deceased.

Umbraviruses (<i>Tombusviridae</i>) <i>Eugene V Ryabov and Michael E Taliansky</i>	827
Varicosaviruses (<i>Rhabdoviridae</i>) <i>Takahide Sasaya</i>	833
Virgaviruses (<i>Virgaviridae</i>) <i>Eugene I Savenkov</i>	839
Viroids (<i>Pospiviroidae</i> and <i>Avsunviroidae</i>) <i>Ricardo Flores and Robert A Owens</i>	852
Watermelon Mosaic Virus and Zucchini Yellow Mosaic Virus (<i>Potyviridae</i>) <i>Cécile Desbiez and Hervé Lecoq</i>	862

VOLUME 4

Viruses as Infectious Agents: Bacterial, Archaeal, Fungal, Algal, and Invertebrate Viruses

Bacterial Viruses

History of Virology: Bacteriophages <i>William C Summers</i>	3
Icosahedral Phages – Single-Stranded DNA (ϕ X174) <i>Bentley A Fane and Aaron P Roznowski</i>	10
Single-Stranded RNA Bacterial Viruses <i>Peter G Stockley and Junjie Zhang</i>	21
Enveloped Icosahedral Phages – Double-Stranded RNA (ϕ 6) <i>Paul Gottlieb and Aleksandra Alimova</i>	26
Membrane-Containing Icosahedral DNA Bacteriophages <i>Roman Tuma, Sarah J Butcher, and Hanna M Oksanen</i>	36
Tailed Double-Stranded DNA Phages <i>Robert L Duda</i>	45
Helical and Filamentous Phages <i>Andreas Kuhn and Sebastian Leptihn</i>	53
Replication of <i>Bacillus</i> Double-Stranded DNA Bacteriophages <i>Silvia Ayora, Paulo Tavares, Ruben Torres, and Juan C Alonso</i>	61
Lytic Transcription <i>William McAllister and Deborah M Hinton</i>	69
Lysogeny <i>Keith E Shearwin and Jia Q Truong</i>	77
Decision Making by Temperate Phages <i>Ido Golding, Seth Coleman, Thu VP Nguyen, and Tianyou Yao</i>	88
Mobilization of Phage Satellites <i>Kristen N LeGault and Kimberley D Seed</i>	98
Portal Vertex <i>Peng Jing and Mauricio Cortes Jr.</i>	105

Prohead, the Head Shell Pre-Cursor <i>Marc C Morais and Michael E Woodson</i>	115
Enzymology of Viral DNA Packaging Machines <i>Carlos E Catalano</i>	124
DNA Packaging: DNA Recognition <i>Sandra J Greive and Oliver W Bayfield</i>	136
DNA Packaging: The Translocation Motor <i>Janelle A Hayes and Brian A Kelch</i>	148
Biophysics of DNA Packaging <i>Joshua Pajak, Gaurav Arya, and Douglas E Smith</i>	160
Energetics of the DNA-Filled Head <i>Alex Evilevitch</i>	167
Bacteriophage Receptor Proteins of Gram-Negative Bacteria <i>Sarah M Doore, Kristin N Parent, Sundharraman Subramanian, Jason R Schrad, and Natalia B Hubbs</i>	175
Tail Structure and Dynamics <i>Shweta Bhatt, Petr G Leiman, and Nicholas MI Taylor</i>	186
Bacteriophage Tail Fibres, Tailspikes, and Bacterial Receptor Interaction <i>Mateo Seoane-Blanco, Mark J van Raaij, and Meritxell Granell</i>	194
Phage Genome and Protein Ejection In Vivo <i>Ian J Molineux, L Letti Lopez, and Aaron P Roznowski</i>	206
Dealing With the Whole Head: Diversity and Function of Capsid Ejection Proteins in Tailed Phages <i>Lindsay W Black and Julie A Thomas</i>	219
Jumbo Phages <i>Isaac T Younker and Carol Duffy</i>	229
CRISPR-Cas Systems and Anti-CRISPR Proteins: Adaptive Defense and Counter-Defense in Prokaryotes and Their Viruses <i>Asma Hatoum-Aslan and Olivia G Howell</i>	242
Bacteriophage: Therapeutics and Diagnostics Development <i>Teng-Chieh Yang</i>	252
Bacteriophage Vaccines <i>Pan Tao and Venigalla B Rao</i>	259
Bacteriophage Diversity <i>Julianne H Grose and Sherwood R Casjens</i>	265
Genetic Mosaicism in the Tailed Double-Stranded DNA Phages <i>Welkin H Pope</i>	276
Bacteriophages of the Human Microbiome <i>Pilar Manrique, Michael Dills, and Mark J Young</i>	283
Bacteriophage: Red Recombination System and the Development of Recombineering Technologies <i>Lynn C Thomason and Kenan C Murphy</i>	291
Nanotechnology Application of Bacteriophage DNA Packaging Nanomotors <i>Tao Weitao, Lixia Zhou, Zhefeng Li, Long Zhang, and Peixuan Guo</i>	302
General Ecology of Bacteriophages <i>Stephen T Abedon</i>	314
Marine Bacteriophages <i>Vera Bischoff, Falk Zucker, and Cristina Moraru</i>	322

Ecology of Phages in Extreme Environments <i>Tatiana A Demina and Nina S Atanasova</i>	342
---	-----

Archaeal Viruses

Diversity of Hyperthermophilic Archaeal Viruses <i>David Prangishvili, Mart Krupovic, and Diana P Baquero</i>	359
Euryarchaeal Viruses <i>Tatiana A Demina and Hanna M Oksanen</i>	368
Vesicle-Like Archaeal Viruses <i>Elina Roine and Nina S Atanasova</i>	380
Virus–Host Interactions in Archaea <i>Diana P Baquero, David Prangishvili, and Mart Krupovic</i>	387
Antiviral Defense Mechanisms in Archaea <i>Qunxin She</i>	400
Discovery of Archaeal Viruses in Hot Spring Environments Using Viral Metagenomics <i>Jennifer Wirth, Jacob H Munson-McGee, and Mark J Young</i>	407
Metagenomes of Archaeal Viruses in Hypersaline Environments <i>Fernando Santos, María D Ramos-Barbero, and Josefa Antón</i>	414
Extreme Environments as a Model System to Study How Virus–Host Interactions Evolve Along the Symbiosis Continuum <i>Samantha J DeWerff and Rachel J Whitaker</i>	419

Fungal Viruses

An Introduction to Fungal Viruses <i>Nobuhiro Suzuki</i>	431
Cross-Kingdom Virus Infection <i>Liyang Sun, Hideki Kondo, and Ida Bagus Andika</i>	443
Diversity of Mycoviruses in Aspergilli <i>Ioly Kotta-Loizou</i>	450
Evolution of Mycoviruses <i>Mahtab Peyambari, Vaskar Thapa, and Marilyn J Roossinck</i>	457
Mixed Infections of Mycoviruses in Phytopathogenic Fungus <i>Sclerotinia sclerotiorum</i> <i>Jiatao Xie and Daohong Jiang</i>	461
Mycovirus-Mediated Biological Control <i>Daniel Rigling, Cécile Robin, and Simone Prospero</i>	468
Mycoviruses With Filamentous Particles <i>Michael N Pearson</i>	478
Prions of Yeast and Fungi <i>Reed B Wickner and Herman K Edskes</i>	487
Single-Stranded DNA Mycoviruses <i>Daohong Jiang</i>	493
Structure of Double-Stranded RNA Mycoviruses <i>José R Castón, Nobuhiro Suzuki, and Said A Ghabrial[†]</i>	504

[†]Deceased.

<i>Ustilago maydis</i> Viruses and Their Killer Toxins <i>Alexis Williams and Thomas J Smith</i>	513
Vegetative Incompatibility in Filamentous Fungi <i>Songsong Wu, Daohong Jiang, and Jiatao Xie</i>	520
Viral Diseases of <i>Agaricus bisporus</i> , the Button Mushroom <i>Kerry S Burton and Greg Deakin</i>	528
Viral Killer Toxins <i>Manfred J Schmitt and Björn Becker</i>	534
Alternaviruses (Unassigned) <i>Hiromitsu Moriyama, Nanako Aoki, Kuko Fuke, Kana Takeshita Urayama, Naoki Takeshita, and Chien-Fu Wu</i>	544
Barnaviruses (Barnaviridae) <i>Peter A Revill</i>	549
Botybirnaviruses (Botybirnavirus) <i>Mingde Wu, Guoqing Li, Daohong Jiang, and Jiatao Xie</i>	552
Chrysoviruses (Chrysoviridae) - General Features and Chrysovirus-Related Viruses <i>Ioly Kotta-Loizou, Robert HA Coutts, José R Castón, Hiromitsu Moriyama, and Said A Ghabrial[†]</i>	557
Fungal Partitiviruses (Partitiviridae) <i>Eeva J Vainio</i>	568
Fusariviruses (Unassigned) <i>Sotaro Chiba</i>	577
Giardiavirus (Totiviridae) <i>Juliana Gabriela Silva de Lima, João Paulo Matos Santos Lima, and Daniel Carlos Ferreira Lanza</i>	582
Hypoviruses (Hypoviridae) <i>Dong-Xiu Zhang and Donald L Nuss</i>	589
Megabirnaviruses (Megabirnaviridae) <i>Yukiyo Sato and Nobuhiro Suzuki</i>	594
Mitoviruses (Mitoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	601
Mycoreoviruses (Reoviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	607
Mymonaviruses (Mymonaviridae) <i>Daohong Jiang</i>	615
Narnaviruses (Narnaviridae) <i>Rosa Esteban and Tsutomu Fujimura</i>	621
Phlegiviruses (Unassigned) <i>Karel Petrzik</i>	627
Plant and Protozoal Partitiviruses (Partitiviridae) <i>Hanna Rose and Edgar Maiss</i>	632
Quadriviruses (Quadriviridae) <i>Hideki Kondo, José R Castón, and Nobuhiro Suzuki</i>	642
Totiviruses (Totiviridae) <i>Bradley I Hillman and Alanna B Cohen</i>	648

[†]Deceased.

Yado-kari Virus 1 and Yado-nushi Virus 1 (Unassigned) <i>Subha Das and Nobuhiro Suzuki</i>	658
Yeast L-A Virus (<i>Totiviridae</i>) <i>Reed B Wickner, Tsutomu Fujimura, and Rosa Esteban</i>	664
Algal Viruses	
Algal Marnaviruses (<i>Marnaviridae</i>) <i>Marli Vlok, Curtis A Suttle, and Andrew S Lang</i>	671
Algal Mimiviruses (<i>Mimiviridae</i>) <i>Ruth-Anne Sandaa, Håkon Dahle, Corina PD Brussaard, Hiroyuki Ogata, and Romain Blanc-Mathieu</i>	677
Miscellaneous Algal Viruses (<i>Alvernnaviridae, Bacilladnaviridae, Dinodnavirus, Reoviridae</i>) <i>Keizo Nagasaki, Yuji Tomaru, and Corina PD Brussaard</i>	684
Phycodnaviruses (<i>Phycodnaviridae</i>) <i>James L Van Etten, David D Dunigan, Keizo Nagasaki, Declan C Schroeder, Nigel Grimsley, Corina PD Brussaard, and Jozef I Nissimov</i>	687
Invertebrate Viruses	
An Introduction to Viruses of Invertebrates <i>Peter Krell</i>	699
Ascoviruses (<i>Ascoviridae</i>) <i>Sassan Asgari, Dennis K Bideshi, Yves Bigot, and Brian A Federici</i>	724
Baculovirus–Host Interactions: Repurposing Host-Acquired Genes (<i>Baculoviridae</i>) <i>A Lorena Passarelli</i>	732
Baculoviruses: General Features (<i>Baculoviridae</i>) <i>Vera ID Ros</i>	739
Baculoviruses: Molecular Biology and Replication (<i>Baculoviridae</i>) <i>Monique M van Oers</i>	747
Bidensoviruses (<i>Bidnaviridae</i>) <i>Qin Yao, Zhaoyang Hu, and Keping Chen</i>	759
Bunyaviruses of Arthropods (<i>Myppoviridae, Nairoviridae, Peribunyaviridae, Phasmaviridae, Phunuiviridae, Wupedeviridae</i>) <i>Sandra Junglen</i>	764
Dicistroviruses (<i>Dicistroviridae</i>) <i>Yanping Chen and Steven M Valles</i>	768
Entomobirnaviruses (<i>Birnaviridae</i>) <i>Marco Marklewitz</i>	776
Hytrosaviruses (<i>Hytrosaviridae</i>) <i>Henry M Kariithi and Irene K Meki</i>	780
Iflaviruses (<i>Iflaviridae</i>) <i>Bryony C Bonning and Sijun Liu</i>	792
Iridoviruses of Invertebrates (<i>Iridoviridae</i>) <i>İkbal Agah İnce</i>	797
Mesoniviruses (<i>Mesoniviridae</i>) <i>Jody Hobson-Peters and Daniel Watterson</i>	804

Nimaviruses (<i>Nimaviridae</i>) <i>Peter Krell and Emine Ozsahin</i>	808
Nodaviruses of Invertebrates and Fish (<i>Nodaviridae</i>) <i>Kyle L Johnson and Jacen S Moore</i>	819
Nudiviruses (<i>Nudiviridae</i>) <i>Yu-Chan Chao, Chih-Hsuan Tsai, and Sung-Chan Wei</i>	827
Parvoviruses of Invertebrates (<i>Parvoviridae</i>) <i>Judit J Péntzes, Hanh T Pham, Qian Yu, Max Bergoin, and Peter Tijssen</i>	835
Polydnaviruses (<i>Polydnaviridae</i>) <i>Anne-Nathalie Volkoff and Elisabeth Huguet</i>	849
Poxviruses of Insects (<i>Poxviridae</i>) <i>Basil Arif, Lillian Pavlik, Remziye Nalçacıoğlu, Hacer Muratoğlu, Cihan İnan, Mehtap Yakupoğlu, Emine Özşahin, İsmail Demir, Kazım Sezen, and Zihni Demirbağ</i>	858
Reoviruses of Invertebrates (<i>Reoviridae</i>) <i>Peter Krell</i>	867
Rhabdoviruses of Insects (<i>Rhabdoviridae</i>) <i>Andrea González-González, Nicole T de Stefano, David A Rosenbaum, and Marta L Wayne</i>	883
Sarothroviruses (<i>Sarothroviridae</i>) <i>Azeez Sait Sahul Hameed</i>	888
Solinviviruses (<i>Solinviviridae</i>) <i>Steven M Valles and Andrew E Firth</i>	892
Tetraviruses (<i>Alphatetraviridae, Carmotetraviridae, Permutotetraviridae</i>) <i>Rosemary A Dorrington, Tatiana Domitrovic, and Meesbah Jiwaji</i>	897

VOLUME 5

Diagnosis, Treatment and Prevention of Virus Infections

Diagnosis

Introduction to Virus Diagnosis and Treatment <i>Maija Lappalainen and Hubert GM Niesters</i>	3
Electron Microscopy for Viral Diagnosis <i>Roland A Fleck</i>	5
Serological Approaches for Viral Diagnosis <i>Klaus Hedman and Visa Nurmi</i>	15
A Brief History of the Development of Diagnostic Molecular-Based Assays <i>Hubert GM Niesters</i>	22
Sequencing Strategies <i>Sibnarayan Datta</i>	27
Validating Real-Time Polymerase Chain Reaction (PCR) Assays <i>Melvyn Smith</i>	35
Rapid Point-of-Care Assays <i>Jan G Lisby and Uffe V Schenider</i>	45

Standardization of Diagnostic Assays <i>Sally A Baylis, C Micha Nübling, and Wayne Dimech</i>	52
Quality Assurance in the Clinical Virology Laboratory <i>Paul Wallace and Elaine McCulloch</i>	64
Biosafety and Biosecurity in Diagnostic Laboratories <i>Hannimari Kallio-Kokko and Susanna Sissonen</i>	82
Screening for Viral Infections <i>Walter Ian Lipkin, Nischay Mishra, and Thomas Briebe</i>	91
Clinical Diagnostic Virology <i>Marcus Panning</i>	98
Virus Diagnosis in Immunosuppressed Individuals <i>Elisabeth Puchhammer-Stöckl and Fausto Baldanti</i>	105
Diagnosis; Future Prospects on Direct Diagnosis <i>Marianna Calabretto, Daniele Di Carlo, Fabrizio Maggi, and Guido Antonelli</i>	112

Treatment

Antiviral Classification <i>Guangdi Li, Xixi Jing, Pan Zhang, and Erik De Clercq</i>	121
Antiretroviral Therapy – Nucleoside/Nucleotide and Non-Nucleoside Reverse Transcriptase Inhibitors <i>Timothy D Appleby and Killian J Quinn</i>	131
Protease Inhibitors <i>Vanesa Anton-Vazquez and Frank A Post</i>	139
HIV Integrase Inhibitors and Entry Inhibitors <i>Daniel Bradshaw and Ranjababu Kulasegaram</i>	145
Management of Respiratory Syncytial Virus Infections (<i>Pneumoviridae</i>) <i>Rachael S Barr and Simon B Drysdale</i>	155
Management of Influenza Virus Infections (<i>Orthomyxoviridae</i>) <i>Bruno Lina</i>	160
Management of Herpes Simplex Virus Infections (<i>Herpesviridae</i>) <i>Nicole Samies and Richard Whitley</i>	175
Management of Varicella-Zoster Virus Infections (<i>Herpesviridae</i>) <i>Andreas Sauerbrei</i>	181
Treatment and Prevention of Herpesvirus Infections in the Immunocompromised Host <i>Sara H Burkhard and Nicolas J Mueller</i>	190
Management of Adenovirus Infections (<i>Adenoviridae</i>) <i>Albert Heim</i>	197
Management of Hepatitis A and E Virus Infection <i>Sébastien Lhomme, Florence Abravanel, Jean-Marie Peron, Nassim Kamar, and Jacques Izopet</i>	206
Management of Patients With Chronic Hepatitis B (<i>Hepadnaviridae</i>) and Chronic Hepatitis D Infection (<i>Deltavirus</i>) <i>Milan J Sonneveld and Suzanne van Meer</i>	217
Studying Population Genetic Processes in Viruses: From Drug-Resistance Evolution to Patient Infection Dynamics <i>Jeffrey D Jensen</i>	227

Virus-Based Cancer Therapeutics <i>Roberto Cattaneo and Christine E Engeland</i>	233
Prevention	
Surveillance of Infectious Diseases <i>Norman Noah</i>	247
Preparing for Emerging Zoonotic Viruses <i>Reina S Sikkema and Marion PG Koopmans</i>	256
Use of Immunoglobulins in the Prevention of Viral Infections <i>Leyla Asadi and Giovanni Ferrara</i>	267
Vaccine Production, Safety, and Efficacy <i>Thomas J Brouwers and Bernard AM Van der Zeijst</i>	281
Vaccines Against Viral Gastroenteritis <i>Scott Grytdal, Tyler P Chavers, Claire P Mattison, Jacqueline E Tate, and Aron J Hall</i>	289
Human Papillomavirus (HPV) Vaccines and Their Impact <i>Jade Pattyn, Pierre Van Damme, and Alex Vorsters</i>	295
Influenza Vaccination <i>Topi Turunen</i>	300
Polio Eradication <i>M Steven Oberste, Cara C Burns, and Jennifer L Konopka-Anstadt</i>	310
Subject Index	315

DIAGNOSIS

Introduction to Virus Diagnosis and Treatment

Maija Lappalainen, HUS Diagnostic Center, HUSLAB, Clinical Microbiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Hubert GM Niesters, Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

There are more than 320,000 mammalian viruses, of which a little over 200 are known to infect humans (Anthony *et al.*, 2013; Woolhouse *et al.*, 2012). The number, however, is steadily increasing. During recent years, we have witnessed the emergence of new viral diseases, such as MERS (Memish *et al.*, 2020), infection caused by Zika virus (Rather *et al.*, 2017), and the COVID-19 pandemic caused by SARS-CoV-2 in 2020 (Hu *et al.*, 2020).

Viral diagnostics is increasingly important in clinical medicine. The growing number of new diseases, the expansion of specific antiviral drugs and the need for effective vaccines poses a challenge for the diagnostic tools available. This volume focuses on clinical virology. It leads us through the presentation of traditional, modern and future diagnostic tools and provides views on how to implement these tools in various clinical settings.

Electron microscopy is a highly valued method allowing visualization of virus particles and combined with antibody-based labeling approaches. It is still the method of choice for identifying ultrastructural changes associated with virus-host cell invasion and replication. Molecular methods, in particular the PCR, have replaced viral culture in many laboratories, limiting the use of this traditional method of virus detection or replacing it altogether. We have to realize that most viruses cannot be grown easily or cannot be grown *in vitro* at all. In special circumstances, however, viral culture may still be needed. Traditional serological assays have largely been replaced by enzyme immunoassays and serodiagnostics are at their best in identifying and pinpointing the time of primary infection, in assessing immune status vs. susceptibility, and in pre-transplant donor-recipient matching.

Challenges with conventional diagnostic methods have been stated in several studies. Only 62% of viral respiratory infections among children can be confidently attributed to known pathogens (Arnold *et al.*, 2008) and in another study, it was demonstrated that in 42.7% of cases of viral gastroenteritis, the pathogen could not be identified with conventional methods (Vu *et al.*, 2019). It has been estimated that up to 40% of viral encephalitis infections remain undiagnosed with modern clinical tests (Kennedy *et al.*, 2017). Therefore, new approaches, like microarrays, sensitive multianalyte methods on microfluidic platforms, and high throughput sequencing (HTS) are needed. In addition, we have to also realize that pathogens can follow a hit-and-run procedure before a clinical syndrome is detected or diagnosed. The data provided with new technology could not only augment diagnostics and help to choose the correct therapy, but also facilitates virological research, since more precise, low-cost and accessible solutions are necessary.

The ongoing COVID-19 outbreak emphasizes the importance of biosafety and biosecurity procedures as well as quality assurance and standardization of diagnostic assays to ensure the appropriate performance of the assays used. Many laboratories already have accreditation according to ISO15189:2015, but as more and more hospitals undergo comprehensive on-site surveys conducted by Joint Commission International (JCI) in order to achieve accreditation, this also highlights the accreditation status of the laboratories they use.

Viruses have a quite simple structure with an encapsidated nucleic acid. They borrow molecular equipment from host cells to complete their replication cycle, thus having only a few targets for antiviral agents. In addition to high costs, the existing antiviral treatments still have safety and efficacy challenges. Fortunately, success stories have also been seen with, for instance, treatment of HIV-1 and hepatitis C, encouraging the development of therapies targeting different steps in the viral replication cycle. Updated reviews about management and treatment of clinically relevant viral diseases will be presented.

Vaccine development normally consists of a series of steps that can take many years. However, urgent need, as noted for the development of a SARS-CoV-2 vaccine, may change the development process. Some of the steps in the research and development process are happening in parallel, while still maintaining strict clinical and safety standards. As an example of triumph of vaccine efficacy, will be the permanent cessation of poliovirus.

The number of potentially pathogenic viruses is very large, while the resources for disease research and development (R&D) are limited. To ensure efforts under WHO's R&D Blueprint are focused and productive, a list of diseases and pathogens are prioritized for R&D in public health emergency contexts. The priority 10 viral diseases are discussed.

In the future, with the developing techniques, the clinical virology laboratories are able to produce an enormous amount of detailed data about viruses, pathogenesis and recovery from the infection. The leading challenge is to fully utilize this data in the patient's clinical management and treatment. The prerequisite for this is good communication and collaboration between laboratory personnel, researchers and clinicians. The improved exchange of information and growing information and knowledge, in turn, significantly aids clinicians in battling viral infections.

References

- Anthony, S.J., Epstein, J.H., Murray, K.A., *et al.*, 2013. A strategy to estimate unknown viral diversity in mammals. *mBio* 4. doi:10.1128/mBio.00598-13.
- Arnold, J.C., Singh, K.K., Spector, S.A., Sawyer, M.H., 2008. Undiagnosed respiratory viruses in children. *Pediatrics* 121, e631–e637. doi:10.1542/peds.2006–3073.

- Hu, B., Guo, H., Zhou, P., Shi, Z.-L., 2020. Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology*. doi:10.1038/s41579-020-00459-7.
- Kennedy, P., Quan, P.-L., Lipkin, W., 2017. Viral encephalitis of unknown cause: Current perspective and recent advances. *Viruses* 9, 138. doi:10.3390/v9060138.
- Memish, Z.A., Perlman, S., van Kerkhove, M.D., Zumla, A., 2020. Middle East respiratory syndrome. *Lancet* 395, 1063–1077.
- Rather, I.A., Lone, J.B., Bajpai, V.K., Paek, W.K., Lim, J., 2017. Zika virus: An emerging worldwide threat. *Frontiers in Microbiology* 8, 1417. doi:10.3389/fmicb.2017.01417.
- Vu, D.-L., Sabrià, A., Aregall, N., *et al.*, 2019. Novel human astroviruses: Prevalence and association with common enteric viruses in undiagnosed gastroenteritis cases in Spain. *Viruses* 11, 585. doi:10.3390/v11070585.
- Woolhouse, M., Scott, F., Hudson, Z., Howey, R., Chase-Topping, M., 2012. Human viruses: Discovery and emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367, 2864–2871. doi:10.1098/rstb.2011.0354.

Electron Microscopy for Viral Diagnosis

Roland A Fleck, King's College London, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Å Ångström a unit of length used to define very small distances. One ångström is equal to 10^{-10} m (one ten-billionth of a meter or 0.1 nanometers).

Bacteriophage (Phage) a virus which infects only bacteria.

DNA See Nucleic acid.

Cryo From the Ancient Greek *κρύος* (*kríos*, “icy cold, chill, frost”) and used as a prefix cryo- for the study of biological materials at very low temperatures by electron microscopy.

Diffraction Limit The resolution of a microscope is principally limited by the physics of diffraction. An optical system with performance at the instrument's theoretical resolution limit is said to be diffraction-limited. Ernst Karl Abbe, described the diffraction limit of a microscope as $d = \lambda / 2n \sin \alpha$, where d is the resolvable feature size (resolution), λ is the wavelength of the illumination source, n is the index of refraction of the medium the sample is imaged in, and α is the half-angle subtended by the objective lens. The electron wavelength λ is $\sim 100,000$ times shorter than that of visible light and this difference is key to the greater resolving power of the electron microscope.

ELISA (Enzyme-linked immunosorbent assay) or enzyme immunoassay (EIA) is a technique designed for detecting and quantifying substances such as peptides, proteins, antibodies and hormones. In an ELISA, detection is by a highly specific antibody-antigen interaction.

Freeze Etching Is the sublimation of surface ice under vacuum to reveal details of the freeze fractured face that were originally hidden. The fracture face of the sample can be imaged directly in a SEM or as a metal replica of the fracture face in a TEM.

Freeze Fracture Is a technique where a frozen specimen is broken to reveal internal structures.

Leidenfrost Effect A physical phenomenon in which a liquid, close to a surface that is significantly hotter than the liquid's boiling point, produces an insulating vapor layer that prevents the liquid making physical contact with the hot surface. When a warm sample is placed in a liquid cryogen the film of insulating vapor slows down cooling. Due to the small temperature difference between solid and boiling point for nitrogen even when a very small sample is placed in liquid nitrogen a film of vaporized liquid nitrogen forms around the sample and impedes heat transfer. This limits the cooling rate achieved and prevents vitrification.

Lyophilization Also referred to as freeze-drying, is a technique which utilizes low pressure and low temperature to dry a material by the sublimation of water. Sublimation converts water directly from a solid to a gas without the effect of a change in surface tension (the tendency of fluid

surfaces to shrink into the minimum surface area possible) which increases as liquid water evaporates causing changes to virus morphology.

Negative Stain Contrasting a thin specimen with an optically opaque electron dense fluid. Once dry, the background is stained, leaving the actual specimen untouched, and thus visible in the transmission electron microscope.

Nucleic acid The genetic information of a virus is carried on nucleic acid. Viruses may be divided into those of ribonucleic (RNA) or deoxyribonucleic (DNA) forms.

RNA See Nucleic acid.

SEM Scanning Electron Microscope, a microscopy technique in which a focused beam of electrons is scanned across a surface in a raster scan pattern, and the position of the beam is combined with the intensity of the detected signal to produce an image.

Single Particle Analysis An image processing technique which is applied to TEM images. The approach digitally combines many images of similar small particles (e.g., a virus) to create an image with stronger and more easily interpretable features. Commonly many individual virus particles will be captured in a single photomicrograph and in many photomicrographs thousands of individual randomly orientated viruses exist. Each individually reveal the different unique faces of the virus. Each individual virus may be “boxed out” or digitally cut out and then processed to build up a three-dimensional reconstruction of the particle. Using cryo-electron microscopy it has become possible to generate reconstructions with near-atomic resolution.

TEM Transmission Electron Microscope, a microscopy technique in which a beam of electrons is transmitted through a specimen to form an image.

Tomography A technique for obtaining detailed 3D structures of sub-cellular macro-molecular objects. Electron tomography is an extension of traditional TEM where single 2D images are collected. Instead, the sample is tilted in the TEM at incremental degrees of rotation around the center of the target sample. At each tilt angle the TEM electron beam is passed through the sample and a 2D image collected. Each image is collected and digitally processed using a weighted back projection into a 3D volume of the sample.

Vitrification The transformation of a water by rapid cooling into a non-crystalline amorphous solid, “glass” thus avoiding the crystalline ice states of water and change in volume associated with freezing.

μl A unit of volume used to define a very small volume. 1 μl is equal to one millionth of a litre.

Introduction to Virus Diagnosis by Electron Microscopy

Organisms, causative of disease, commonly occupy a length scale below that visible by the naked eye. Light microscopes readily aid the observation of amoeba and bacteria but are unable to directly observe viruses, which are also a major cause of disease. A simple change to host phenotype may imply the presence of a pathogen further analysis, using specialist techniques and equipment is needed to confirm and identify a virus infection.

The electron microscope (EM) developed by Ernst Ruska and Max Knoll in 1931, to exploit the shorter wavelength of electrons, created an instrument with the potential to overcome the diffraction limits of the light microscope (restricted by the longer wavelengths of visible light). Eight years after the initial EM development, Ruska with Kausche and Pfankuch were able to visualize a virus for the first time. This first view of a tobacco mosaic virus (TMV) heralded the start of modern virology. The initial proof of concept was rapidly followed by images of bacteriophage. The first EM studies on pathogenic viruses were published shortly after, in 1938 by von Borries *et al.*, detailing the size and fine structure of smallpox, vaccinia virus and Ectromelia mouse pox. Soon after, the EM became the *de facto* tool of choice for the virologist. Providing a simple and almost direct way to observe nanometer (nm) scale biological structures. Today, the latest cryo transmission EMs (cryoEM) reveal near atomic level structural information of the virus and its assembly.

Initially, positive staining by the application of phosphotungstic acid and osmium tetroxide to bacteriophage and proteins was tried, however, disruption to structures was observed due to the low pH of the staining solutions. Early studies of fine virus ultrastructure and morphology used the shadow casting techniques devised by Williams and Wyckoff in 1945, whereby a fine metal (e.g., palladium or platinum) was evaporated under vacuum and applied to a dried preparation of virus from a fixed low angle. Shadowed with $\sim 8 \text{ \AA}$ thickness of metal, at an angle of about 12° from the horizontal plane of the mount, the low angle and unidirectional coating created a shadow of the virus as the metal collected on the side of the virus facing the metal evaporation source. When viewed in the transmission electron microscope (TEM) the enhanced contrast between metal shadowed and shaded areas revealed the morphology and fine ultrastructural details of the virus. Although shadow-cast specimens provided useful details from both air-dried and freeze-dried material, the protocol itself is challenging and requires specialist vacuum metal evaporation technology and high quality “near pure” samples.

The method of choice for direct virus detection is negative staining (NS) (Fig. 1). Established by Brenner and Horne as a fast, robust and universal EM technique NS has been used by virologists for over 60 years. The EM NS preparation technique is effective for fresh, aged, partly degenerated, or dried specimens (e.g., vesicle fluid, fecal samples) and can be applied readily to biological tissues; homogenized or mortar ground organs, biopsy (including wart and molluscum contagiosum skin lesions), autopsy specimens and cell cultures (both supernatant and lysed cell pellets). The versatility of NS also allows samples collected from soil and aquatic sources, e.g., from streams, urban irrigation systems, sewage outlets, standing water, lakes etc to be studied and can be used to detect agricultural or human sewage contamination of a eutrophicated water body.

The potential of EM in the laboratory diagnosis of infectious diseases was first demonstrated in the control of smallpox outbreaks and developed into a routine procedure for the detection and diagnosis of common viral diseases by skilled operators. These virus diagnostic EM (DEM) laboratories became ubiquitous in hospitals and public health laboratories globally, providing a routine service for circulating gastrointestinal and respiratory infections. Routinely applied to human pathogens, e.g., viruses of medical importance, with around 21 viral families and genera being deemed medically significant (Table 1), DEM also found wide acceptance in veterinary virology. Due to the potentially serious consequences of viral outbreaks, rigorous, and universally recognized training of personnel dealing with virus identification was required. Today training in Diagnostic Electron Microscopy (DEM) of human and veterinary infectious diseases can be obtained via an annual External Quality Assurance (EQA) scheme prepared and distributed by the Robert Koch Institute (RKI), Berlin.

Diagnostic Transmission Electron Microscopy

The transmission electron microscopy (TEM) is a powerful initial step in virus diagnosis for several reasons. The TEM provides an unbiased direct and rapid overview of the sample. The viral burden (amount), size and shape (critical for initial identification) of virus or viruses present is immediately apparent. The observed morphology, even for unexpected or emerging pathogens normally allows preliminary classification to family level based on particle morphology (Table 1). This “catch all” diagnostic facility is due to TEM sample preparation protocols targeting proteins, the viral capsid or ribonucleoprotein (RNP) complexes rather than the virus RNA or DNA genomes and makes TEM uniquely unbiased as a detection method. Initial identification of a virus requires only very small volumes ($\sim 1 \mu\text{l}$) of samples carrying high virus ($> 10^6$ particles ml^{-1}) loads and can then direct the virologist to further methods (e.g., bioassays, serological, or molecular biology approaches) to identify virus genus and species.

As a technique where identification is by morphology, DEM developed rapidly. During the early 1950s until the mid-1960s many new viruses were discovered and names were commonly assigned on Latinized abbreviations attributed to their morphology or niche and -viridae for virus families and -virus for viral genera (Table 1). Viral morphology and niche provided the basis for grouping viruses into families, be they a virus family able to replicate only in vertebrates, only in invertebrates, only in plants, only in bacteria or less commonly viruses able to replicate in more than one of these hosts. Viruses were then often further grouped or classified on the basis of size and shape, i.e., helical morphology is seen in nucleocapsids of many filamentous and pleomorphic viruses; helical nucleocapsids consist of a helical array of capsid proteins wrapped around a helical filament of nucleic acid.

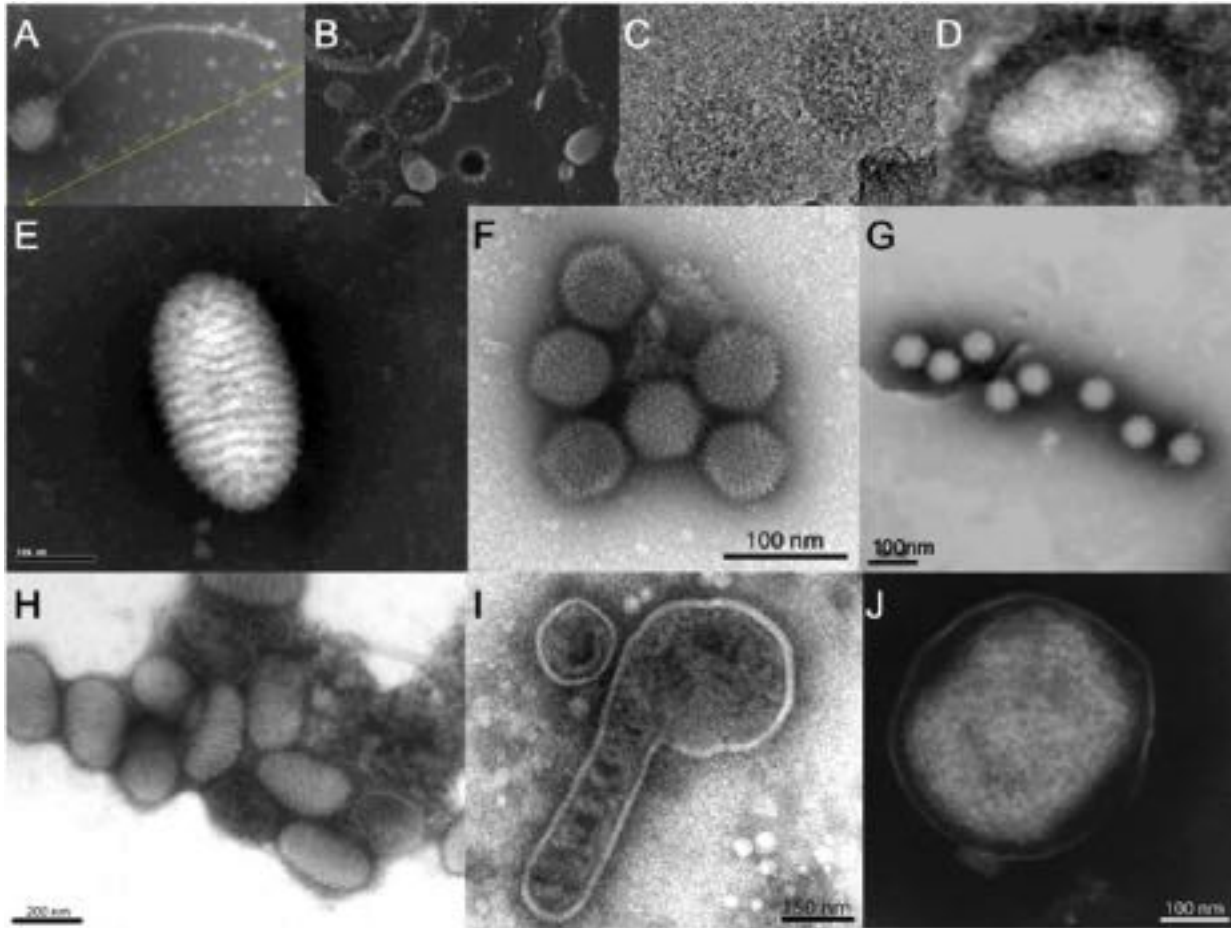


Fig. 1 Virus particles as revealed by different electron microscopy methods. (A) bacteriophage collected using a transmitted electron detector in a field emission scanning electron microscope (FEGSEM). B-C Influenza virus particles by different electron microscopy techniques (B) prepared by high pressure freezing, freeze fracture and direct imaging in the FEGSEM, (C) prepared by plunge freezing into liquid ethane and imaging at 200 kV by low dose cryo transmission electron microscopy, (D) conventional negative staining of air dried particles with 2% sodium silicotungstate (SST). (E-J) virus particles by negative staining, imaged at 120 kV on a TEM E, (H) parapoxvirus, (F) adenovirus, (G) reoviruses virus particles with monomorph morphology (I) paramyxovirus (J) orthopoxvirus.

Icosahedral morphology is characteristic of the nucleocapsids of many “spherical or cubic” viruses. Some viruses (e.g., herpesvirus) also have an outer envelope which can be readily observed in the EM and can help in identifying the virus.

Despite the role of EM as the first line diagnostic tool for gastrointestinal and influenza infections being superseded by genetic screening approaches (e.g., PCR). EM remains an important diagnostic tool in several clinical settings. Indeed, guidance issued by the UK's Royal College of Pathologists cites the Association of Clinical Pathologists Best Practice No. 160, which states that ‘Many respondents [to a review of laboratory practice in renal pathology] expressed the opinion that to carry out evaluation of renal biopsy specimens without at least having the availability of electron microscopy is negligent’. Other specialized areas include the diagnosis of primary ciliary dyskinesia; certain skin/connective tissue disorders, e.g., inherited bullous lesions and Ehlers–Danlos syndrome; and specialized areas of ophthalmic pathology. The versatility of DEM and the power to directly observe the morphology of virus particles, particularly for unknown or emerging diseases and complex co-infections remains its unique value. This is a specialist niche role and with the reduction in the number of DEM facilities globally one whose essential skilled operator base is increasingly difficult to maintain. Specialist DEM units are now largely maintained at a national level in specialist disease surveillance and control centers and continue to play a critical role as demonstrated by their contribution to the investigations into recent unexplained diseases, potential bioterrorism incidents and zoonoses. The identification of the Severe Acute Respiratory Syndrome (SARS) agent, the unexpected diagnosis of monkeypox virus infections in the USA, Buffalopox infections of humans, arboviruses and viral hemorrhagic fevers in India and Hendra paramyxovirus infections in Australia were by these specialist centers (e.g., the Center for Disease Control and Prevention, Georgia, USA and the National Institute of Virology, Pune, India). Based on its ability to recognize contaminations, rapid application of TEM has an important role for quality control of reference material or of virus preparations used for antibody production or directly for vaccination.

Table 1 Viruses families known to cause disease in humans, the origin of their name, nucleic acid relationship, symmetry of the capsomere as described by classical negative stain diagnostic electron microscopy and commonly observed disease

<i>Virus family</i>	<i>Nucleic acid</i>	<i>Symmetry</i>	<i>Explanation of the origin of family name</i>	<i>Symptoms and associated diseases attributed to the virus family</i>
Picornaviridae	RNA	Icosahedral	pico (small) and RNA	e.g., Coxsackieviruses associated with neurological disease, hand foot and mouth disease and respiratory tract illness in man
Reoviridae	RNA	Icosahedral	respiratory, enteric, and orphan viruses because the agents were found in both respiratory and enteric specimens	Widespread in domestic animals and man, first isolated from respiratory and stool samples
Papovaviridae	DNA	Icosahedral	papilloma, polyoma, and vacuolating agent, now split into Papillomaviridae and Polyomaviridae families	e.g., Human Papilloma Virus (HPV) causing cervical cancers and genital warts
Retrovirus	RNA	Complex	reverse transcriptase	e.g., causing Acquired Immune Deficient Syndrome (AIDS) following infection with the Human Immunodeficiency Virus (HIV)
Hepadnaviridae	DNA	Icosahedral	replication of the virus in hepatocytes and their DNA genomes, as seen in hepatitis B virus.	Including the Hepatitis B virus
Adenoviridae	DNA	Icosahedral	adeno, "gland"; refers to the adenoid tissue from which the viruses were first isolated	Infections are commonly associated with diarrhea in children and adults
Astroviridae	RNA	Icosahedral	astron "star"	small viruses associated with childhood diarrhea
Arenaviridae	RNA	Complex	arena "sand" describes the sandy appearance of the virion.	The causative agent of Lassa fever in man
Bunyaviridae	RNA	Helical	Bunyamwera, the place in Africa where the type strain was isolated	Highly pathogenic in man causing Crimean-Congo hemorrhagic fever, Rift Valley Fever and Nephropathia epidemica
Ebolavirus	RNA	Helical	Ebola River, close to the two simultaneous 1976 outbreaks: one in Nzara (a town in South Sudan) and the other in Yambuku village (Democratic Republic of the Congo) where the virus was first identified.	Morphologically similar to Marburg virus causing hemorrhagic fever in man with high levels of mortality
Calicivirus	RNA	Icosahedral	calix, "cup" or "goblet" from the cup-shaped depressions on the viral surfaces	Commonly infecting cats causing respiratory tract infections in adult cats and fatal pneumonia in kittens, pigs and sealions, also infections in man
Coronaviridae	RNA	Helical	corona, "crown" describes the appearance of the peplomers protruding from the viral surface	often reported as causing "common cold" symptoms in man, human respiratory infections and "gassing disease" in chickens
Filoviridae	RNA	Helical	from the Latin filum, "thread" or "filament" describes the morphology	including the Marburg virus which is highly infectious and pathogenic in man
Herpesviridae	DNA	Icosahedral	herpes, "creeping" describes the nature of the lesions	Including the Epstein Barr virus responsible for Glandular fever (mononucleosis) and cold sores in man
Orthomyxoviridae	RNA	Helical	ortho, "true," plus myxo "mucus," a substance for which the viruses have an affinity.	the viruses responsible for Influenza
Orthopoxviruses	RNA	Complex	combining ortho with pox (pustule) descriptive of the infection	the vaccinia, cowpox, and variola (smallpox) viruses
Paramyxoviridae	RNA	Helical	para, "closely resembling" myxo	including the viruses causing measles, mumps, respiratory syncytial infections and Newcastle disease
Parvoviridae	DNA	Icosahedral	parvus means, "small"	one of smallest DNA viruses and including human parvovirus B19 which was initially discovered in human sera being screens for hepatitis B antigen. A canine variant causes bloody diarrhea in puppies and can lead to death
Poxviridae	DNA	Complex	pock means, "pustule"	Including the pseudopoxvirus which causes udder infections in cows and is commonly named cow pox
Rhabdoviridae	RNA	Helical	rhabdo, "rod" describes the shape of the viruses	causing Rabies, an infection almost always fatal in man
Togaviridae	RNA	Icosahedral	toga, "cloak" refers to the tight viral envelope.	Including the virus responsible for Rubella "German Measles"

To partially address the combined challenges of fewer specialist viral DEM units and the essential requirement for experienced virus DEM specialists, Laue and Möller at the RKI have generated a publicly available database of EM images. Expansion of such archives will facilitate recognition of newly emerging as well as unexpected viruses. For instance, an increase in the local prevalence

of plant viruses can serve as indicators of environmental stress and predict underlying contamination of irrigation water with pathogenic human viruses. Thus, TEM remains an integral part of the surveillance role of national reference laboratories, where prompt, reliable and comparable results are essential for efficient diagnostics during unexpected regional virus outbreaks or epidemics (e.g., rare zoonoses, such as Buffalopox). Despite this longstanding and continued value and capacity to inform, DEM as a tool in the clinical setting has been largely stagnant with little change in the technical procedures or their application for decades. In marked contrast, TEM and particularly cryoEM in the research setting has experienced an explosion in capacity and capability, heralding the “resolution revolution”.

Biological EM Challenges and Sample Preparation

Biological EM is not without its challenges. The electron beam is generated under vacuum at pressures and temperatures that are nominally incompatible with liquid water, yet water is the most abundant cellular constituent. Biological tissues (carbon-based life forms) have poor contrast in the electron microscope because they are composed mainly of light elements. In fact, carbon is so “transparent” in the electron microscope that it is often employed as a film to support biological samples. To overcome these challenges, hydrated “live” tissue is converted to an EM stable state; conventional protocols employ a series of steps, including chemical fixation, alcohol dehydration and resin infiltration. Heavy metal salts are added for positive staining of fixed and embedded specimens or as negative stains of whole structures that have been deposited on a support film. More recently, cryo-fixation has been adopted and allows the “solidification of a biological specimen by cooling with minimal displacement of its components”. By using low temperature as a physical fixation strategy, the morphology and dimensions of the living material are retained, and soluble cellular components are not displaced, which means that processing artefacts commonly encountered in more conventional room-temperature EM techniques are either reduced or removed. CryoEM often allows direct observation of specimens that have not been stained or chemically fixed. However, due to the requirement for the cryo fixed sample to be maintained at cryogenic temperatures and the challenge of avoiding warming and/ or contamination during transfer into the TEM the approach is not regarded as suitable for DEM. Instead, cryoEM is reserved for high resolution research studies.

Small particles like bacteria, viruses and proteins readily lend themselves to a simple protocol able to preserve their ultra-structure and introduce sufficient contrast to make their observation easy. Referred to as negative staining (NS) it is an easy, rapid, qualitative method for examining the structure of isolated organelles, individual macromolecules and viruses at the EM level (Fig. 2). This is the opposite of the stain binding to the sample and excess being removed, in which case a particle would appear dark against bright background (positive staining). Despite providing stunningly beautiful images of viruses, NS does not allow the highest resolution examination of samples.

Negative staining involves the deposition of heavy atom stains, e.g., uranyl acetate, around a small particle to make it visible in the electron beam. Structural artefacts such as flattening of spherical or cylindrical structures are common. Nevertheless, negative staining is a very useful technique because of its ease and speed (taking only a few minutes for a skilled operator), and also because it requires no

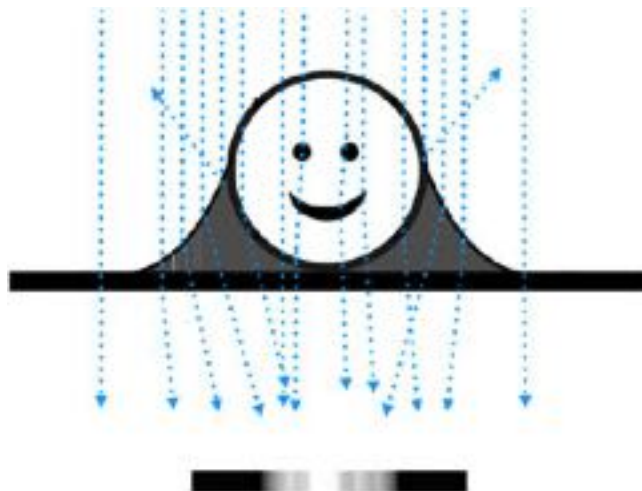


Fig. 2 Representation of a negatively stained virus particle on a TEM grid. Negative staining requires very little specialized equipment and virus suspensions can be mounted directly on TEM support grids (e.g., carbon formvar) and examined after air-drying. However, due to their small size and the low atomic numbers of the elements forming the virus particle they provide little electron scattering on interaction with the electron beam (blue dotted lines). This means that they have very low contrast and do not stand out from the background. To visualize the virus, heavy metal (high atomic number, high scattering) stain such as uranyl acetate or phosphotungstic acid is used. The virus particles (smiley face) are surrounded by the stain, which scatters electrons strongly creating negative contrast, whereby the particle appears bright against the stain. As the stain has a small grain size, as it covers the virus particle it is also trapped in surface features. Differences in *z* “thickness” of stain which the electron beam passes through en route to the detector is responsible for variations in the level of signal detected and thus contrast, to reveal gross morphological features and details of subunits forming the protein structures.

Table 2 Common negative stains used for diagnostic electron microscopy. Other less common stains include; aluminum formate, uranyl oxalate, and uranyl sulfate and gold thioglucose

<i>Common negative stains:</i>	<i>Normal range for use (pH)</i>	<i>Normal concentration (w/v)</i>
Phosphotungstic acid (PTA)	5–8	1%–3% aqueous solution
Uranyl acetate (UA)	4.2–4.5	1%–3% aqueous solution
Uranyl formate	4–4.5	0.5%–1% aqueous solution
Sodium silicotungstate (SST)	5–7	1%–5% aqueous solution
Ammonium molybdate	6–7	1%–2% aqueous solution
Methylamine tungstate	5–8	2% aqueous solution

specialized equipment other than that found in a regular EM laboratory. The NS should not react with the specimen in a 'positive staining' manner (i.e., it should not bind to the specimen). However, uranyl based stains will bind to proteins and sialic acid carboxyl groups and to lipid and nucleic acid phosphate groups. One effect of this is to induce aggregation of the material.

General Protocol for the Study of Viruses by Negative Staining

The method of surrounding or embedding specimens in opaque dyes is commonly used in light microscopy and dates from about 1884 (e.g., microbiologists use India ink to stain micro-organisms heat fixed to a slide. The background is stained with the ink while the organisms remain clear, creating a negative stain of the organism). A common clinical application of this procedure is to confirm the morphology of the encapsulated yeast *Cryptococcus* spp. which cause cryptococcal meningitis.

The equivalent preparative technique applied to electron microscopy was derived from experiments on a bacteriophage by high resolution EM in the Cavendish Laboratory, Cambridge, during 1954. When mixed with 1% potassium phosphotungstate and air-dried, gross morphological features of a human adenovirus and details of subunits forming the protein structures were revealed. The approach was first described by Brenner & Horne in 1959 and confirmed the earlier observation of Hall in 1955 working at the Massachusetts Institute of Technology that "bushy stunt virus stained with 5% phosphotungstic acid at pH 4.6 and insufficiently washed is the opposite to what I usually sought by the use of electron stains the viability of particles of low scattering power can be enhanced as well, if not better, by surrounding them with dense material rather than impregnating them with dense material" (*sic*).

Negative staining is as diverse in how it is performed as it is ubiquitous as a rapid technique able to reveal the structure of small biological particles (e.g., viruses and proteins) in the EM. Many heavy metal contrasting stains are available (Table 2). Each has a "fine grain structure" suitable to reveal the fine structure of the virus particle without obscuring virus morphology, is electron dense to create strong contrast between the unstained specimen (the virus) and the surrounding (negative) stain and in almost all cases has the capacity to stabilize or otherwise fix the sample to preserve its structure in the EM.

Fixed or unfixed samples may be used. With unfixed specimens there is the potential problem of changes occurring due to osmotic shock (or to changes in ionic composition) since most negative stains are made up in distilled water. Also, there may be risk of operator infection when examining unfixed bacterial or viral samples.

Samples should be concentrated and purified to allow the operator the best chance of identifying and imaging relevant particles. Cell debris can be removed by centrifugation before loading samples onto grids. Due to the small volumes used in the preparation methods detailed below, the number of biological particles in the sample must exceed 10^6 particles/ml.

There are a number of methods for preparing samples for examination by NS; the method used depends on either the sample to be stained or the method of choice in the laboratory. Below are outlined some of the simpler methods.

Method A Negative Staining Single-droplet method.

- (1) Glow discharge carbon-formvar coated grids just before use to increase their hydrophilicity. This is essential to allow even wetting of the grid with the sample and stain and ensures even NS of the sample.
- (2) Place 1–3 μ l of the sample on the grid (sufficient to cover the grid surface).
- (3) After ~ 10 s slowly pipette 20 μ l of stain on to sample, while gently absorbing the stain from the opposite side using a wedge of filter paper. The staining procedure should take ~ 30 –60 s.
- (4) After absorbing as much stain as possible allow the grid to dry and examine the grid as soon as possible, preferably the same day in the TEM. Some methods include a wash in distilled water after drying which can be helpful if the buffer becomes crystalline when dried or when the sample is too thick on the grid.

Method B.

Sequential two-droplet method.

- (1) A suspension of particles/organelles is made in a suitable buffer or in distilled water.
- (2) Place a drop (~ 5 –10 μ l) of the suspension on to a carbon-formvar glow discharged grid.
- (3) When the suspension has partly dried the grid is washed by touching it three times to the surface of a drop of distilled water.
- (4) Remove excess water by touching the grid to a filter paper.
- (5) A small drop of stain is then applied to the grid.

- (6) After 10 s the excess stain is removed by touching the edge to a filter paper. The sample is wicked to dryness, dried at room temperature and viewed in the TEM.

Alternatively, sequential droplets of; sample, stain, wash can be laid out on a hydrophobic surface (dental wax or parafilm). The carbon formvar TEM grid can be moved between each droplet with short timed exposures to complete the staining procedure.

Negative Staining for Diagnosis of Fecal Samples

Make a 10%–20% suspension of feces in PBS. Centrifuge at 3000 rpm for 30 min to clarify. Collect the supernatant (i.e., the fecal extract) and store at 4°C until EM examination has been completed. Original fecal specimens and extracts may be stored long term at –40°C.

Fecal extracts are examined by one of the following methods.

- (1) **Ultracentrifugation**
Centrifuge 2 ml fecal extract at 20,000 rpm for one hour. Tip off the supernatant and drain the tube well. Reconstitute the deposit in a few drops of distilled water. Mix a drop of this suspension with an equal volume of 3% phosphotungstic acid pH 6.3. Place a drop of mixture on a formvar-carbon coated grid and then touch a piece of filter paper to the edge of the grid to blot dry. Examine the grid in the electron microscope at an instrument magnification around $\times 30,000$ to $\times 60,000$.
If fecal specimens are very fatty and do not adhere to the grid well, mix one-part resuspended pellet with one part 0.05% bovine plasma albumin. Add one-part stain, place a drop on the grid and blot dry.
- (2) **Lyphogel method**
Weigh out 0.1 g lyphogel into a small screw capped bottle. Add 0.55 ml fecal extract. Leave at room temperature for four hours or at 4°C overnight. Place a drop of the residual fluid on a formvar-carbon-coated grid. Blot off the excess with a piece of filter paper. Wash the grid with 1–2 drops of distilled water and blot. Apply a drop of negative stain to the grid and again remove excess stain with filter paper. Examine grids in the microscope.
- (3) **Sucrose method**
Place 4 ml sucrose solution (30% w/v in distilled water) in a centrifuge tube. Layer 1 ml fecal extract onto the sucrose. Centrifuge at 40,000 rpm for two hours. Drain the tubes well – use a swab or tissue if necessary, to ensure that all excess sucrose solution has been removed. Reconstitute the pellet in a few drops of distilled water and stain as in method 1.
These methods are suitable for the detection of most of the usual gastroenteritis viruses eg rotavirus, adenovirus, astrovirus, etc., and are usually all that are required for examining fecal specimens from babies. Further concentration on cesium chloride gradients may be required to detect small round viruses and in particular the small, round, featureless, parvovirus like particles which are more usually associated with gastroenteritis in older children and adults. This method is used in addition to one of the three methods already described.
- (4) **Cesium chloride gradients**
Place 2 ml fecal extract in a centrifuge tube and top up with phosphate buffered saline (PBS). Centrifuge at 40,000 rpm for one hour. Resuspend the deposit in 0.3 ml distilled water. Place 4.5 ml cesium chloride solution in a centrifuge tube (Cesium chloride is made up in distilled water to give a starting density of 1.38g/cc. Cesium chloride solution can be obtained from BDH as a 60% w/w solution for ultracentrifuge work. This solution diluted to 28% in distilled water gives a suitable starting concentration).
Lift the grid up with a pair of forceps and stain by placing a drop of phosphotungstic acid on the grid and blotting dry as before. Examine the grids in the microscope.

Preparation of Agarose Slides

Place 2 ml agarose (0.9% in distilled water) on a microscope slide. Allow the agarose to set and leave in a damp petri dish at 4°C for at least one hour before use. One slide will accommodate three grids satisfactorily. Cesium chloride diffuses into the agarose leaving virus particles attached to the grid (Fig. 3).

A general working assumption is that the threshold of visibility for virus particles is approximately 10^6 particles per ml. Further direct concentration of samples directly onto the TEM grid is possible. The Airfuge by Beckman Colter equipped with an EM-90 Electron Microscopy Particle Counting Rotor will concentrate up to 6 separate samples (100 μ l) directly onto the EM grid and significantly improves the detection sensitivity of DEM.

Limitations of Traditional Processing Techniques and Advanced Protocols

Traditional EM processing strategies introduce artefacts (perturbations to structure) that affect tissue structure, e.g., by shrinkage or swelling of the tissue under examination, the shrinkage of cellular organelles, and/or the extraction or redistribution of cellular constituents such as lipids, proteins and DNA. In addition, most chemical fixatives react with proteins and cross-link peptide chains as part of their action. These reactions can be highly deleterious to epitopes, significantly compromising immunohistochemical studies.



Fig. 3 Preparation of fecal samples for negative staining using an agarose slide method. 2 ml of agarose is added to a glass microscope slide and allowed to spread to form a uniform film. Once the agarose has set the sample collected from a cesium chloride gradient can be pipetted onto the agarose film and a TEM grid with previously glow discharged carbon-formvar film placed on top of the sample droplet. The cesium chloride will diffuse into the agarose leaving virus particles attached to the grid.

Specialist EM sample processing strategies have emerged to minimize or avoid these limitations. The capacity of EM can be enhanced by correlative light and electron microscopy (CLEM) workflows, where light microscopy (LM) events are co-localized with underlying ultrastructure. In its simplest form, LM generates a “Google Earth”-like map to identify areas of interest, which are then studied at the ultrastructural level by EM, so overcoming both the physical size limitations imposed on samples by EM and the diffraction limits of resolution imposed by LM. Pioneering groups now employ correlation in workflows which allow single particle analysis cryoEM techniques to be applied directly to structures in cells and tissues. Electron tomography is required to first determine the 3D structures of irregular samples such as cells and organelles. Once good tomograms are available, it is possible to extract multiple copies of a structure of interest and then apply single particle methods to obtain averages with greatly improved signal to noise ratios.

By minimizing the use of chemical fixatives, immuno-electron microscopy (IEM) is also possible. IEM is based on the serological principles of the ELISA and can be used for specific virus identification during routine TEM diagnosis. IEM has the advantage that it works directly with raw serum, so no purification of immunoglobulins or conjugation steps are necessary and only small reaction volumes are required. If cell or tissue samples are prepared following the protocol as described by Tokuyasu in 1973, sections of tissue can be labeled with gold conjugated antibodies and ultrastructural details of virus replication processes revealed.

A TEM DEM laboratory can maintain, refrigerated, active stable collections of antisera specific to a broad spectrum of virus species and isolates. Depending on the composition of antigens, as well as the epitopes present in the original virus purification, polyclonal antisera can be suitable for capturing multiple viruses within a genus. This versatility can be valuable in DEM, where serological relationships are revealed in the EM through strength of antibody attachment in the labeling step. A “strong” homologous reaction will be seen as a virion tightly packed with antibodies. For emerging isolates or a “less specific” heterogenous virus only weak antibody labeling will be seen. It is also possible to use monoclonal antibodies which targeting single epitopes to provide high specificity and reproducibility for distinguishing different virus isolates.

Immuno-electron microscopy may be performed on fecal extract by; centrifugation of 1 ml fecal extract at 20,000 rpm for one hour. Remove the supernatant and resuspend the deposit in 0.2 ml PBS. Add 0.1 ml serum at the required dilution (dilute in PBS). Cover the tube with parafilm and leave at room temperature for one hour and at 4°C overnight. Add cold PBS to fill the tube. Centrifuge at 18,000 rpm for one hour. Drain the tube well, resuspend the deposit in a few drops of distilled water and stain as in the ultracentrifugation method described above (method 1).

Cryo Electron Microscopy

Recently, cryo-fixation has been adopted as a physical fixation strategy. The morphology and dimensions of the living material are retained, and soluble cellular components are not displaced, which means that processing artefacts commonly encountered in more conventional room-temperature EM techniques are either reduced or removed. Cryo-EM often allows direct observation of specimens that have not been stained or chemically fixed.

Cryo-fixation has two distinct advantages over chemical fixation. It is rapid (measured in milliseconds), which means that the sample is preserved, hydrated and in a “close-to-life” state at the point of initiating fixation, and ensures simultaneous immobilization of all macromolecular components. Many protein networks are labile and prone to disruption with even a slight osmotic or temperature change; cryo-fixation minimizes these effects. Cryo-techniques also allow the study of biological samples with improved ultrastructural preservation and may facilitate the study of dynamic processes. Rapid freezing prevents artefactual aggregations of proteins because tissue fixation and processing can be performed with no or minimal use of cross-linking fixatives, which means that samples retain higher antigenicity.

This level of preservation relies on vitrification (the transformation of water from a liquid to an amorphous state without inducing the nucleation of ice crystals). The nucleation of ice crystals is temperature- and pressure-dependent. Crystallization depends on the cooling rate, itself dependent on the thermal properties of water, the sample thickness, and the heat extraction flow at the surface of the specimen. Freezing is a time-dependent process. Vitrification sits at the beginning of a workflow where the sample is either subsequently processed to a stable room-temperature state for imaging or maintained in its vitreous state throughout. Due to the poor thermal conductivity of water, cooling of virus samples by simply plunging into liquid nitrogen (LN) would be too slow to allow for vitrification and samples would thus be badly damaged by ice crystals.

Water was vitrified by vapor deposition on a cold surface as early as 1935 and by fast cooling of micro-droplets in 1980, however, these were either not useful or not amenable for direct observation at the TEM. It was Jacques Dubochet’s group which first successfully vitrified thin films of water containing small particles supported by the surface tension of water itself between the bars of

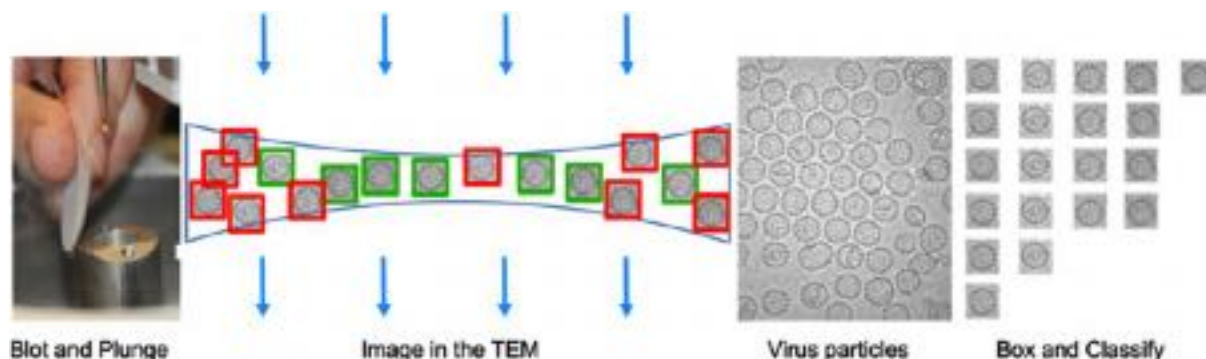


Fig. 4 Cryo electron microscopy for near atomic resolution structural studies. The workflow is presented from left to right. The first step is the blotting of a small volume of pure virus particles ($\sim 4 \mu\text{l}$) on a TEM grid with a suitable support film containing small regular holes. The blotting step produces a thin film of virus particles with many supported by surface tension across the holes in the film. The sample is then rapidly frozen by plunging into a cryogenic liquid to create a vitrified film of virus in water. The vitrified film will contain many individual particles which are imaged in the transmission electron microscope. These particles are randomly arranged and can be categorized as “ideal” (green boxed) when single isolated particles are held in the vitrified solution or “non-ideal” (red boxed) when particles are in contact with the air-water interface which can cause denaturing of structure. In “thick” areas of the film, resolution can be limited by the volume or overlapping of particles. The resulting TEM images are relatively low contrast as no heavy metal stains are added to the sample. Individual particles are subsequently “boxed out” of many individual micrographs and classified on a basis of similarity before finally being processed to generate a high resolution 3D structure.

EM grids. The water films could be vitrified because they were plunged into liquid ethane rather than into liquid nitrogen or helium, which are not good cryogens due to the influence of the Leidenfrost effect (a film of insulating vapor which slows down cooling). The frozen samples could then be observed directly at their frozen hydrated state in the cryo-electron microscope and vitrification could be verified by electron diffraction. The plunging technique developed by this group is the basis of sample preparation used for modern single particle analysis and reconstruction, reaching near atomic resolution (Fig. 4). The success of the method lies not only in choosing the right cryogen but also in the high cooling rates that can be achieved during the freezing of extremely thin aqueous films ($\sim 100 \text{ nm}$), which are in the range of 10^7 C s^{-1} . As Jacques Dubochet wrote: “It is the good fortune of the electron microscopist that, reducing sample size in order to increase the cooling speed, vitrification becomes easy just when the dimensions of the specimen are those suitable for electron microscopy.” This technique as the preparation step for single particle analysis was, together with transformative advances in understanding of EM image formation and image processing, rewarded by the award of the Noble Prize for Chemistry in 2017 for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution to Jacques Dubochet, Richard Henderson and Joachim Frank.

There are a variety of methods for rapid freeze fixation of larger samples e.g., plant or animal tissue or even cell monolayers that are few μm in thickness. The depth of vitrification is often limited (e.g., a few microns) for samples plunged into liquid cryogen. The simplest add cryoprotectants to prevent ice crystal growth and achieve vitrification by plunging the sample into liquid nitrogen. An alternative technology is high-pressure freezing (HPF). First introduced by Moor and Riehle in 1968, HPF exploits the physical benefit of high pressure (210 MPa) to reduce the cooling rate required for the vitrification of water from several $100,000 \text{ K s}^{-1}$ to a few 1000 K s^{-1} , making vitrification of relatively thick samples practicable. HPF machines synchronize pressurization and cooling of the sample (to below the glass transition temperature (T_g)) within 20 ms, in a highly reproducible manner, extending the depth of vitrification to as much as 200 μm . These HPF machines are now incorporating added functionality to achieve a high degree of temporal resolution. For example, light-sensitive cation channels can be expressed in cells and tissues, stimulated in the HPF pressure chamber at different time intervals (ranging from ms to s) and immediately cryo immobilized. Electron micrographs acquired from many time points reveal in minute detail sequences of morphological changes at ms precision. Using this approach, the process of ultrafast endocytosis has been characterized during neurotransmission. Recently, electrical stimulation has also been adapted to the HPF allowing electron microscopy to reveal complex events previously identified by electrophysiology.

Tissue samples prepared this way require further modification to make them compatible with the transmission electron microscope. They must either be converted to a room-temperature stable state (e.g., by freeze substitution into a resin) or be sectioned to make them thin enough for the TEM electron beam to pass through the sample. This sectioning must be carried out whilst remaining below the glass transition temperature (T_g); when this is done, images close to the native structure of biological specimens can be achieved. The significant technical challenges of these techniques preclude them from use for DEM.

Advances in Scanning Electron Microscopy

Cryo-FEGSEM has become well established as a research tool. The stability and emitter life of a FEG source, coupled with the stability and low keV sensitivity of a modern scanning electron microscope, is a powerful tool. Unlike TEM, in which high accelerating voltages are used to transmit the electrons through the specimen, SEM benefits from low accelerating voltages that

limit beam interaction with the specimen. This is key to successful cryo-SEM but creates a challenge around generating sufficient signal for high-resolution imaging. Freeze fracture followed by cryo-FEGSEM can reveal ultrastructural details of viruses and permits high resolution direct visualization of cell membranes and the budding of virus particles (**Fig. 1**). Modern FEGSEM systems (with improved detector sensitivity and signal-to-noise ratios) are now achieving sub-nm resolution at 1 kV, making them highly capable tools for life science research, particularly when combined with high-resolution coating and fracture techniques. When equipped with a STEM detector it is possible to collect images of NS samples using the SEM similar to those generated by the TEM (**Fig. 1**). However, the use of a rasterizing scan to generate the image makes rapid screening of NS samples for DEM challenging. The operator must wait until the scan is complete before seeing the full scanned area. This prevents rapid DEM screening of samples. RKI have developed a field-based bioterrorism SEM protocol for the rapid detection of bacterial spores using portable “table-top” SEM’s. This approach provides a rapid direct test to determine if an unidentified powder is biological (i.e., anthrax spore) or chemical (i.e., talcum powder).

Conclusion

Electron microscopy of viruses is largely determined by the quality of the sample preparation technique used. By optimizing the sample preparation steps, remarkable structural details of viruses, their organization and replication strategies can be revealed. Cell culture can be used to study the mode of replication and when combined with EM and protocols which allow thin sections of cells and tissues to be studied, direct observation of virus replication achieved.

Selection of the most suitable technique is important. The diagnostic role of EM is principally by negative staining. Rapid sample preparation readily and flexibly applied to a wide range of samples facilitates rapid detection and identification of a viral pathogen. Higher resolution cryoEM approaches are better adapted for research applications where sufficient time is available for optimization of all conditions necessary to achieve near atomic resolution structures. The versatility of EM for virus detection maintains its critical role as a tool for diagnosis of viral diseases.

Further Reading

- Danev, R., Yanagisawa, H., Kikkawa, M., 2019. Cryo-electron microscopy methodology: Current aspects and future directions. *Trends in Biochemical Sciences* 44 (10), 837–848.
- Dubochet, J., McDowell, A.W., 1981. Vitrification of pure water for electron microscopy. *Journal of Microscopy* 124, RP3–RP4.
- Fleck, R.A., Humbel, B.M., 2019. *Biological Field Emission Scanning Electron Microscopy*. London: Wiley, p. 752.
- Gelderblom, H.R., 1996. Chapter – 41 Structure and classification of viruses. In: Baron, S. (Ed.), *Medical Microbiology*, fourth ed. Boston, USA: Addison-Wesley Publishing Company, Incorporated, Health Sciences Division.
- Gentile, M., Gelderblom, H.R., 2005. Rapid viral diagnosis: Role of electron microscopy. *New Microbiologica* 28 (1), 1–12.
- Gentile, M., Gelderblom, H.R., 2014. Electron microscopy in rapid viral diagnosis: An update. *New Microbiologica* 37 (4), 403–422.
- Glauert, A.M., Lewis, P.R., 1998. *Biological Specimen Preparation for Transmission Electron Microscopy*. London: Portland Press, p. 326.
- Griffiths, G., Lucocq, 2014. Antibodies for immunolabeling by light and electron microscopy: Not for the faint hearted. *Histochemistry and Cell Biology* 142, 347–360.
- Horne, R.W., Wildy, P., 1979. An historical account of the development and applications of the negative staining technique to the electron microscopy of viruses. *Journal of Microscopy* 117 (1), 103–122.
- Laue, M., Möller, L., 2016. The VirusExplorer DEM - A Database for Diagnostic Electron Microscopy of Viruses, Zenodo. doi:10.5281/zenodo.221574.
- Madeley, C.R., Field, A.M., 1988. *Virus Morphology*, second ed. Edinburgh, London: Churchill Livingstone.
- Nermut, M.V., Frank, H., 1971. Fine structures of influenza A2 (Singapore) as revealed by negative staining, freeze-drying and freeze-etching. *Journal of General Virology* 10, 37–51.
- Robards, A.W., Sleytr, U.B., 1995. Low temperature methods in biological electron microscopy. In: Glauert, A.M. (Ed.), *Practical Methods in Electron Microscopy* 10. Amsterdam: Elsevier, pp. 5–146.
- Stirling, J., Eyden, B., Curry, A., 2012. *Diagnostic Electron Microscopy: A Practical Guide to Tissue Preparation and Interpretation*, first ed. John Wiley & Sons.
- Tokuyasu, K.T., 1973. A technique for ultracyotomy of cell suspensions and tissues. *Journal of Cell Biology* 57 (2), 551–565.

Relevant Websites

- <http://www.era.rothamsted.ac.uk/eradoc/OCR/ResReport1964-128-144>
Report for 1964.
e-RA.
- <https://www.pirbright.ac.uk/cross-cutting-capabilities/bioimaging>
Specialists in imaging viruses and virus-host interactions.
- <https://www.gu.se/core-facilities/centre-for-cellular-imaging-cci?>
Standard protocols for negative staining and chemical fixation of biological tissues with downloadable and printable pdf protocols.
- <https://zenodo.org/record/221574>
The VirusExplorer DEM - A database for diagnostic electron microscopy of viruses.
- <https://zenodo.org/record/221524>
The VirusExplorer DEM - A database for diagnostic electron microscopy of viruses.
- https://www.rki.de/EN/Content/infections/Diagnostics/SpecialLab/EM_content.html
The virus diagnostic unit at the Robert Koch Institute.

Serological Approaches for Viral Diagnosis

Klaus Hedman, University of Helsinki, Helsinki, Finland and Helsinki University Hospital, Helsinki, Finland

Visa Nurmi, University of Helsinki, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

This is an update of R. Vainionpää, M. Waris, P. Leinikki, 2015. Diagnostic Techniques: Serological and Molecular Approaches, In Reference Module in Biomedical Sciences, Elsevier Inc., doi:10.1016/B978-0-12-801238-3.02558-7.

Introduction

Virus serodiagnostics (serology) can be employed for a number of clinical indications, including determination of the etiology and respective time(s) of primary infections or of secondary infections or recurrences, the latter comprising endogenous reactivations and exogenous reinfections. Based on several characteristics of virus-specific B-cell immune response, serology firstly aims at identifying and dating a recent active (i.e., replicative) infection (when infected?), and secondly at determining an individual's lifelong infection history (ever infected?). The former question serves the clinician in diagnosis of diseases associated temporally with the corresponding infection types (primary; other). The latter question can serve two wide (and diverse) purposes: (1) Assessment of an individual's status of immunity versus susceptibility; and (2) Uncovering the possibilities of infection chronicity or reactivation (via latency), or of other longer-term manifestations linked with lifelong history of contagion by some of the many human viruses that exist.

Serology, in a nutshell, suits best for diagnostic assessment of such viruses that tend to infect the human body systemically (widely; deeply) rather than superficially (mucosa; skin). Understandably, for the immune response to ensue, the individual ought to be immunocompetent.

This article briefly describes the principles of serological approaches and their utilities in essential (exemplary) clinical settings. Excluded is virus antigen detection (using antibodies as a tool), despite its ongoing or growing importance in management of e.g., some "heavyweight" pathogens such as HIV or hepatitis B.

Following transmission, the virus begins to multiply, and after an ostensibly silent incubation period, the production of infectious virus begins, and the clinical symptoms appear simultaneously or subsequently, pending virus species, patient and clinical context. I.e., the appearance of virus-specific antibodies follows the primary contagion by a "window" of initial seronegativity lasting days to weeks, and a subsequent phase of antibody maturation lasting months. Having reached the detection level, the first antiviral immunoglobulin (Ig) class is IgM, followed by IgG. Around this time the infectious (or antigenic) virus in blood, termed viraemia, as a rule (pending methods sensitivity) disappears. In the first encounter with a given virus – termed a primary infection, or immune response – the IgG level tends to rise relatively slowly; as opposed to a subsequent encounter with the same virus/antigen – a secondary infection/response – in which the IgG level rises and peaks more rapidly; whereas the corresponding IgM may remain undetectable. Typical course of virus infection and antiviral antibody response are summarized in **Fig. 1**. The antiviral antibodies may be searched in two (or even more) successive serum samples taken at the acute and convalescent phases. In selected cases, other body fluids such as cerebrospinal fluid, saliva or potentially even urine can be analyzed for antiviral antibodies.

Principles of Serological Assays

Serological diagnosis is usually based on either the demonstration of the presence of specific IgM or a significant increase in the levels of specific IgG between two consecutive samples taken 1–4 weeks apart. The antigen for the test can be either viable or inactivated virus or some of its components (proteins; epitopes) prepared by virological or molecular methods. Specific markers or physical separation are used to demonstrate the isotype of the antiviral antibody. In some cases, even IgG subclass specificities are determined although they have limited value in diagnostic work.

In primary infection the antiviral IgM usually peaks at 7–10 days after onset of symptoms, and typically disappears during the subsequent weeks or months. However, it is not exceptional for the circulating IgMs to remain detectable even for years (IgM persistence). By contrast, in secondary infections – endogenous reactivation or exogenous reinfection – an IgM response often remains absent altogether.

After primary infection the antiviral IgG usually becomes detectable a week or two later than the IgM, and often persists through life. This characteristic renders IgG an important hallmark of an individual's lifelong infection history (with regard to a given immunogen/pathogen) – and thereby (pending assay calibration) an indicator of antiviral immunity vs. susceptibility – and (pending taxonomy and crossreactivity versus specificity) of infection latency, linked with the potentials for secondary infections, chronicity or other sequelae.

In the early phase of primary infection, the avidity (functional affinity; average binding force) of antiviral IgG tends to be low, followed by increase during several months along with maturation of the immunoglobulin response. Diagnostic applications of IgG-avidity assays have been developed to help distinguish serological responses due to acute infections from those of chronic or past infections, and for a number of indications addressed below.

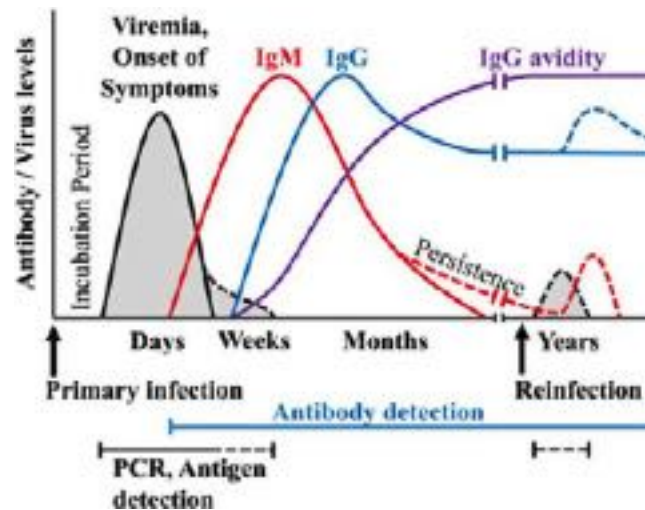


Fig. 1 The course of virus infection, including shedding of infectious virus after incubation period and typical antiviral antibody response. Some typical diagnostic approaches have been marked.

Serological assays in general provide information of many types and perspectives, and for a number of clinical contexts/conditions. By dating primary infection they can tell the etiology of an acute illness even when infectious virus or its components no longer are detectable. The assays are used widely for determination of the immune status, in screening of blood products for the risk of infectious-agent transmission, and for the need of prophylactic procedures. They are also widely used for epidemiological studies, determination of vaccine-induced immunity, and other similar public health purposes.

IgG Avidity

IgG avidity, or functional affinity, is the average binding strength of an antibody population towards an antigen. Low-avidity IgG antibodies usually appear within 2–3 weeks after primary infection or immunization. Reflecting B-cell affinity maturation, IgG avidity increases during the following months. Subsequently, including re-infections and re-activations, the avidity remains elevated. On top of IgM and IgG, IgG avidity measurement provides an additional tool for sorting out the different stages of infection (low avidity during primary infection – increasing avidity during convalescence – high avidity as a marker of past infection, including chronic or recurrent infections). Indeed, valuable information is achieved when traditional IgM and IgG assays are combined with those of IgG avidity. During the past 3 decades, IgG avidity measurement has in the literature been adopted into use with numerous viral pathogens. During pregnancy, adverse clinical outcomes with certain pathogens (e.g., cytomegalovirus, parvovirus B19 and rubella virus) are mainly due to primary infection, moreover with the clinical picture and severity depending on the stage of pregnancy of the maternal infection. This highlights the importance of IgG avidity measurement in timing (dating) of obstetric infections. In addition to viral diagnostics, IgG avidity is increasingly utilized in other branches of clinical microbiology. More recently, the clinical relevance of antibody avidity is being investigated in many autoimmune diseases as well. In epidemiology, IgG avidity (along with IgM) measurement in a single sample has been proposed to be a cost-effective alternative to consecutive sampling, for assessment of new cases (primary infections).

Most IgG avidity assays employ a chaotropic protein denaturant (e.g., urea) to disrupt the antigen-antibody bond. As high-affinity bonds tend to be more resistant to denaturation than those of low affinity, the result can be expressed by comparison of residual (continuously antigen-bound) antibody levels obtained with and without the denaturant. An approach more recently adopted from basic immunology into clinical use employs a soluble antigen to capture the high-avidity antibodies, while those of low avidity remain to bind selectively to the solid phase antigen (displaying higher epitope density required for low-affinity interaction). In analogy with the chaotrope approach, the avidity result can be expressed by comparing the results obtained with versus without the antigen in solution. Indeed, this approach lacks the chaotrope that can crystallize and hamper assay automation.

Unlike the unequivocal affinity of a monoclonal antibody, the avidity of a polyclonal antibody population represents an approximation of the average of affinities, and tends to be expressed in arbitrary/relative values. The simplest is a ratio of signals obtained from two parallel measurements, often referred to as an avidity index ($AI = (\text{signal with denaturation})/(\text{signal without denaturation}) = \text{relative proportion of denaturant resistant high avidity antibodies}$). As the AI is, to some extent, affected by the concentration of the corresponding IgG, the working dilution of the measurement may be chosen according to the IgG level. Due to its simplicity and acceptable sensitivity and specificity, the AI has gained notable popularity. Another approach employs serial dilutions of sample, whereupon avidity is calculated by the ratio of IgG end-point titers generated from titration curves. This latter approach has a broader dynamic range than the former and permits avidity measurement from samples the IgG concentration of which is too low for the AI approach. The sample titration though, understandably is more laborious and costly.

Clinical specimens other than serum/plasma are useful in some situations. Whole blood e.g., dried on filter paper (Guthrie card) facilitates transport over or storage in challenging conditions (climatic; financial). Cerebrospinal fluid can identify immune response (reflecting infection) within the central nervous system (CNS), even if nucleic acid testing (NAT) is overtaking large parts of CNS diagnostics. Increasing attention in serodiagnostics is being paid to the use of noninvasive sample materials such as saliva – and recently, urine. The former has found use e.g., in epidemiology, while the latter's value in clinical virology calls for in-depth assessment!

General Considerations and Sources of Error

The antigens used in serological assays play a central role in their clinical performance. Some epitopes are more immunogenic than others; and the immunoreactivity towards certain antigens/epitopes even correlates with the stage of infection (or duration/time from immunization). e.g., the antibodies against hepatitis B virus s-antigen indicate past infection or vaccination and are generally undetectable during acute or chronic infection. Low purity of antigen conceivably leads to increased background via nonspecific binding and some of the serological methods are more prone to this than others. Antibody responses elicited against certain viruses can also react against other, usually closely related, viral species through shared or similar epitopes. These cross-reacting antibodies can thus cause false positives unless: only unique epitopes are assayed; or cross-reacting antibodies are blocked and depleted with other viral antigens (e.g., anti-dengue virus antibodies with anti-zika virus cross-reactivity and anti-zika antibodies with anti-dengue cross-reactivity can be blocked with dengue antigen, leaving only zika specific antibodies unaffected). Not only serodiagnostics, but the entire antibody response against a virus can be influenced by cross-reacting antibodies of a previous infection. In the phenomenon called original antigenic sin, immune response against a virus can profoundly affect subsequent responses elicited against other viruses. e.g., infection by human bocavirus 1 produces a classical primary immune response. Subsequent infection by human bocavirus 2 can lead to activation of cross-reacting anti-bocavirus 1 memory B cells instead of primary immune response against bocavirus 2. Original antigenic sin not only hampers serodiagnostics but can also lead to weaker immune protection and more severe disease. Rheumatoid factor is an autoantibody and a diagnostic marker for autoimmune diseases such as rheumatoid arthritis and Sjögren's syndrome. It binds the Fc portion of IgG and is usually of IgM class. Rheumatoid factor can thus cause false positives in indirect IgM assays by binding to antigen-specific IgG, if the sample contains said antiviral IgG. In current IgM assays this is rarely an issue since they either are based on methodology not affected by rheumatoid factor, include removal of the bridging IgG or the sample is assayed for rheumatoid factor. A rare but extremely elusive source of false positives are serum antibodies against assay components that are generally considered immunogenically inert. Human antibodies against both streptavidin and its ligand biotin, a popular tool in molecular biology, have been reported in the literature and shown to cause false positives in several assays utilizing the streptavidin-biotin bond.

Restrictions and Limitations

Serological assays have their restrictions. In some infections, e.g. superficially confined to the skin or mucosal membranes, the antibody response is not strong enough. Immunocompromised patients are often unable to mount a B-cell response. In newborns or infants the presence of (residual) maternal IgG can hamper identification of the baby's own antibodies (hallmarking intrauterine infection). The converse also holds; functional comparison with the mother's antibodies especially combined with follow-up can reveal significant qualitative differences, e.g., in antibody isotype, antigen-recognition or avidity, to indicate the emergence of the infant's endogenous immunoreactivity.

Principles of Widely Used Serological Tests

Immunoassays

Immunoassays by definition take advantage of the highly specific antigen-antibody bond to recognize analytes (which in the context of this review are antibodies) in the sample. The majority of immunoassays used in serodiagnostics of viruses are based on sequential binding reactions that, step by step, build a molecular complex onto a solid surface, often beginning with antibodies in the sample targeting the antigen. Each reaction is separated by a washing step to remove unbound reagents. Provided that antiviral antibodies are present in the sample, the process ends at binding of a label generating a measurable signal that is proportional to the amount of those antiviral antibodies. Although many of the assays are rather time consuming and labor intensive, they also are robust, in general have high sensitivity and specificity, and are available in many optional formats. Commercial kits are available for numerous viruses and the assays can be run either manually or in automated systems, up to fully automated high-throughput pipelines used with many clinically relevant viruses. The different immunoassay approaches can also be adapted into multiplexing (i.e., simultaneous detection of antibodies against different viruses or epitopes).

Solid-phase microwell immunoassays

Indirect immunoassays: The by far most popular format in antibody detection is indirect solid-phase enzyme immunoassay (EIA, Fig. 2(A)) alternatively termed enzyme-linked immunosorbent assay (ELISA). Antigens are first immobilized onto the surface of a

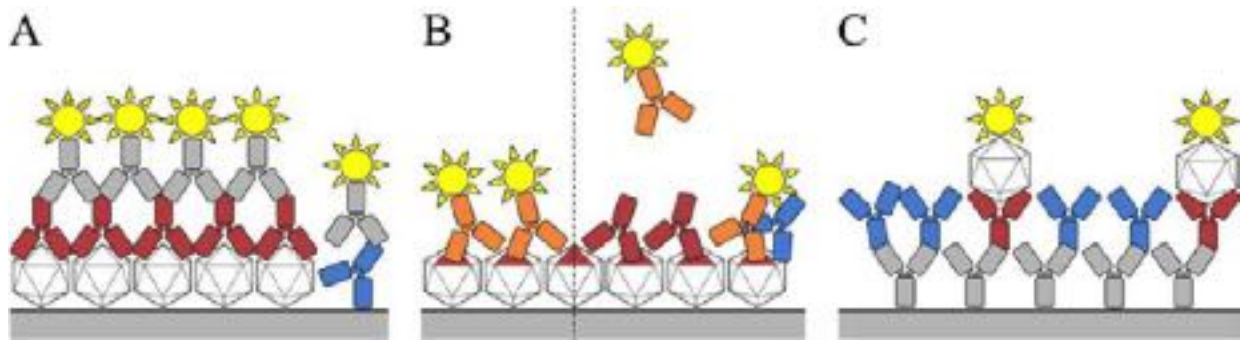


Fig. 2 The most popular immunoassay formats: indirect (A), competitive (B) and capture (C) immunoassay.

microwell and then incubated with the sample. Alternatively, immobilization onto soluble magnetic microspheres can provide increased total surface area leading to higher sensitivity as well as to faster reaction times, as diffusion from the solution phase onto the surface of the microwell is no longer the limiting factor. The specific antibodies bind to the solid-phase antigen and are then detected with anti-immunoglobulin secondary antibodies. Commonly either anti-IgG or anti-IgM are used to assay IgG and IgM antibodies, respectively. With appropriate secondary antibodies, other immunoglobulin classes or subclasses can be assayed. The secondary antibody is coupled with a label that allows its detection. The signal furthermore is amplified, as several secondary antibodies can bind a primary antibody. Depending on the label, the assay nomenclature can vary, although the principle remains the same. In EIAs the label is an enzyme (e.g., horseradish peroxidase or alkaline phosphatase) that can activate multiple substrate molecules thus amplifying the signal (e.g., color reaction measured via absorbance). If fluorescent label is used the method may be called fluoro- (or fluorescence) immunoassay (FIA). A method utilizing chemiluminescence can be called chemiluminescence immunoassay (CLIA). The label can even be a strand of DNA that is amplified and quantified in a subsequent PCR-step. Nonspecific binding of antibodies can cause problems since they are equally recognized by the secondary antibody. This can be largely prevented by blocking the solid-phase with an inert protein or detergent. The indirect immunoassays remain, however, susceptible to cross reacting antibodies if additional measures are not taken to block them.

Competitive immunoassays: In competitive immunoassays (Fig. 2(B)) labeled mono- or polyclonal antibody generates the signal by binding to its antigen. If a sample contains antibodies against the same epitope(s) they compete with the labeled antibody preventing it from binding. This results in a lower signal and the level of this inhibition is proportional to the amount of specific antibodies in the sample. Competitive immunoassays are highly specific since nonspecific or cross-reacting antibodies binding to other epitopes are not detected, unless they are bound close enough to cause steric hindrance. Provided that suitable monoclonal antibodies exist, competitive immunoassays can measure antibodies against single epitopes (e.g., a neutralizing epitope) without the need to separate said epitope from the rest of the antigen (which might not even be feasible). Competitive assays as such cannot distinguish between immunoglobulin classes, but the IgG can be removed or precipitated from the sample to reveal possible underlying IgM reactivity.

Capture immunoassays: Instead of antigens, the solid-phase of capture immunoassays (Fig. 2(C)) consists of anti-IgM or anti-IgG capture-antibodies (or IgG-binding proteins such as protein A or G) that bind and "capture" the total IgM or IgG populations in the sample. Labeled antigen is then added and bound only by antibodies specific to it if present in the captured immunoglobulin population. This makes the method highly specific, although cross-reacting antibodies can still cause false positives. With IgG only a small fraction of the captured total antibody population is specific to a given antigen which can cause problems with sensitivity if signal is not sufficiently amplified. With IgM this is less of an issue since it is primarily produced during acute infections and generally does not persist afterwards. Capture immunoassays are also not affected by rheumatoid factor which makes them especially well suited for IgM diagnostics.

Immunoblotting

Immunoblotting is a classical method for measuring antibody responses against multiple different antigens (i.e., a multiplex assay, actually). The separated antigen molecules are adsorbed as discrete bands on a solid strip and then incubated with the clinical sample. Antibodies present in the sample bind each to their respective antigens and are detected with a labeled secondary antibody visualizing or otherwise identifying the antibody-tagged bands. The latter are then compared to controls on another strip. Immunoblotting is useful (1) for control of antibody specificity, and (2) when the labeling pattern towards several molecules is required for diagnosis; although improved multiplex immunoassays of other types have decreased the popularity. Immunoblotting is still commonly used to ascertain the results of sensitive first-line serodiagnostics for e.g., HIV and hepatitis C virus (HCV).

Immunofluorescence microscopy

Immunofluorescence assay (IFA) follows the same basic principle as the other indirect solid-phase assays. Infected cells immobilized on a microscope slide act as the antigen, and bound antibodies of a sample are detected by a fluorophore-labeled secondary (anti-immunoglobulin) antibody. The method is specific and sensitive, but quite labor intensive, and reading of the

result with a microscope calls for experience. IFA has been a popular approach in the past but has now, to a large extent, been replaced by immunoassays of EIA-type.

Lateral-flow immunoassays

In lateral-flow immunoassays the sample is applied on one end of a porous capillary strip and then diffuses along it, passing through several sections containing different reagents (analogous to distinct reaction steps of other immunoassays). Labeled secondary antibodies can bind those of the sample and move along with them. A band of immobilized antigen captures the specific antiviral antibodies whereby the labeled secondary antibody makes the band detectable. A control band further downstream captures the labeled secondary antibody indicating that excess reagents and nonspecific antibodies have successfully diffused through the system. One strip can hold several antigenic bands, and many of the assays are compatible with whole blood as sample. Performance of the test is often simple and the result is available in a few minutes to hours, making the approach, together with latex agglutination, the most popular form of POC serodiagnostics. While the sensitivity and specificity of these tests can be good, they tend to fall behind those of the more laborious immunoassays such as EIA.

Homogenous wash-free immunoassays

Unlike the other, homogenous assays lack the laborious washing steps, whereby the results are obtained simply by combining the reagents with the sample. The assay can rely on proximity sensing techniques such as Förster resonance energy transfer, two-photon excitation energy transfer or their fluoro-chemical analogs. Briefly, included are two indicator molecules (or complexes) that generate a distinct signal only when in close proximity of each other. When coupled to antigens and/or antibodies, their binding brings the two indicators sufficiently close to generate the signal. In the absence of binding the indicator remains quiescent. Although such techniques have been successfully employed with PCR and microscopic imaging, they still are in their infancy in antibody detection. Commercial kits are available for antigen detection, but presently are more research than diagnostics oriented. Since the homogenous assay format is faster and simpler than the conventional immunoassays, it has great potential in POC diagnostics. Compared with lateral flow and latex agglutination it does require particular instrumentation, but with ongoing development of inexpensive portable technology this is becoming less of an issue.

Neutralizing Antibody Assay

All antiviral antibodies by definition bind viral antigens. Neutralizing antibodies are such that, through this binding and without other components of the immune system, reduce or “neutralize” the infectious capacity of a virus (e.g., blocking the interaction between the virus and its receptor). They can be of IgG, IgM or IgA class and generally persist during lifetime.

In neutralizing antibody assay (Fig. 3(A)), known amounts of infectious virus are mixed with the sample and incubated for a short period after which the residual infectivity is measured by infecting cultured cells and observing the outcome. This infectivity is then compared with the infectivity obtained without the sample and the neutralizing capacity is calculated by comparing the two results. Neutralizing antibody assay is specific and sensitive, but time-consuming and laborious, and therefore not widely used in routine diagnostic services. It measures, however, something the other serological methods do not: the function and efficacy of antiviral immunity, not its mere presence. For this reason neutralizing antibody assays are useful in more in-depth studies of the humoral immune response and especially with vaccines, as successful vaccination (i.e., immunity against the pathogen) generally both requires and can be demonstrated by the presence of neutralizing antibodies.

Neutralizing antibodies can also be measured, to some extent, with conventional immunoassays with regard to neutralizing epitopes, e.g., using neutralizing antibodies in a competitive assay. This requires that, for a given virus: such epitopes have been discovered; a combination of all the relevant ones can be assayed; and they can be assayed independently from the non-neutralizing epitopes.

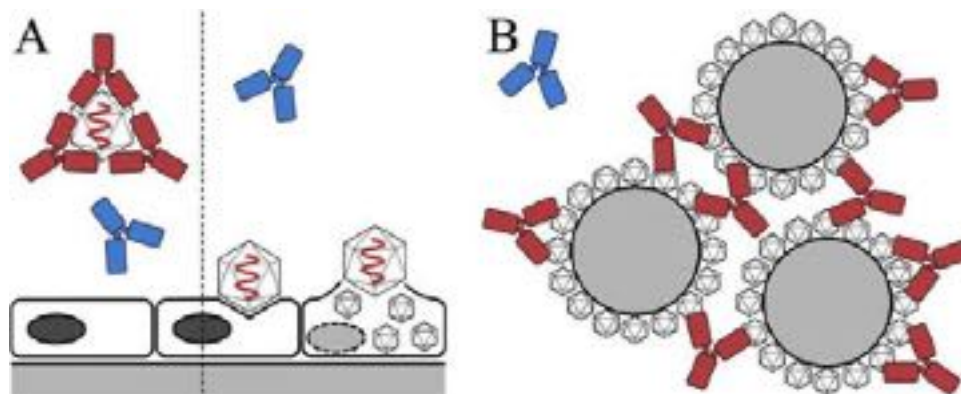


Fig. 3 Neutralizing antibody (A) and latex agglutination (B) assays.

Latex Agglutination

Latex agglutination (or fixation) assays, **Fig. 3(B)** are based on aggregation of antigen-coated beads (e.g., of latex) mixed with a sample containing antibodies against the antigen. When the bi- or multivalent antigen-binding sites of an antibody (IgG or IgM, respectively) attach to different latex beads, the latter are cross-linked and aggregate. The aggregation (or agglutination) may be observed visually or by increased light-scattering spectrophotometrically. The agglutination tests tend to be rapid (within minutes) and easy to perform. Their popularity as POC assays also is due to independence of complex/expensive equipment. The reaction, however, is sensitive to the prozone effect. Paradoxically, excessive concentrations of specific antibodies can bring about a negative result, when the cross-linking fails due to bead coverage. This in turn can be overcome by assaying at serial dilutions of the sample.

Hemagglutination Inhibition

Many viruses bind to hemagglutinins, molecules at the surface of red blood cells (RBC) of various animal species, and causing in suitable conditions RBC aggregation leading to (hem)agglutination. Inhibition of the phenomenon by antiviral antibodies in patients' sera has historically been used widely for diagnostic purposes. In the test, a virus preparation of predetermined hemagglutinating capacity, and standardized amount of red blood cells as well as the sample are mixed, and after incubation the residual hemagglutination is recorded. Both IgG and IgM antibodies are functional. Presently the hemagglutination inhibition test is used in seroepidemiological studies and as a reference test, most notably with influenza viruses. As the test utilizes complete virions it is well suited for assaying antibodies even on serotype level. The hemagglutination inhibition test is seldom used for diagnostic purposes nowadays.

Complement Fixation

Complement fixation test is a classical diagnostic approach based on the capacity of complement, a group of heat-labile plasma proteins of the innate immune system, to bind antibodies in antigen-antibody complexes. When such complexes are present on the surfaces of red blood cells, the complement brings about their lysis which can be observed visually or measured. The sample's endogenous complement is first deactivated by heating; the sample is then mixed with the viral antigen and a fixed amount of exogenous complement; and RBCs coated with anti-erythrocyte antibodies are finally added as an indicator. If the sample contains antibodies against the antigen, the exogenous complement gets depleted (fixed) by the immunocomplexes leaving the RBCs intact. If the antiviral antibodies are absent, the complement lyses the cells. The assay measures both IgG and IgM. Since the assay is laborious, calls for standardization of several biological variables and less sensitive than many other immunoassays it is rarely used nowadays.

Sequencing Analysis of Antibody Repertoire

Advances in high-throughput DNA and RNA sequencing (HTS) and bioinformatics have made it possible to sequence, reconstruct and analyze the entire repertoire of B cell immunoglobulin genes (Ig-Seq). These data include all the different B cell clones of an individual and thus contain sequences for all the antibodies they produce. Compared with conventional PCR assays, HTS requires more extensive sample preparation, and the raw data typically comprise around 100 base pair fragments (reads) totaling up to 10^{12} base pairs per sample. Extensive data processing is thus required for comprehensible results.

At present the process of sample preparation, sequencing and data analysis is too resource-demanding for most diagnostic purposes. The vast complexity and constant microevolution of the B cell population is also a challenge; and clear-cut, standardized and diagnostic-grade protocols are yet to be established. Ig-Seq has, however, already become an important and popular tool in research, including in-depth study of humoral immune responses elicited by viral infections or vaccinations. With further advancement in sequencing technology and above all in sequence analysis, clinical applications can be expected in the future. The unresolved challenge for HTS-based serodiagnosis is to link the amino acid sequence of an antigen-binding-site to its specific epitope (i.e. to deduce antibody function based on its sequence). Databases of confirmed antibody sequence-epitope pairs and predictive algorithms are being constructed to resolve the issue but are still at their early stages. The potential of such assays is, however, vast since a single test could reveal the entire infection history (all the pathogens against which antibodies are being produced) or susceptibility (absence of protective antibodies) of the patient.

Point-of-Care Serodiagnostics

POC tests are becoming increasingly common in clinical practice. In proper use, they provide cost-efficient diagnostic guidance for quick clinical decision making. Their performance is rarely as good as that of conventional laboratory tests; their use should be quality-controlled, and a back up should be arranged with a virology lab. Most of them are based on easy-to-use lateral-flow or latex particle technology and yield the result in a few minutes. To curtail the diagnostic window, antibody and antigen detection can be combined into the same POC cassette (e.g., HIV Ag/Ab assay). While nucleic acid detection has largely replaced antigen detection and is now increasingly available in POC format, it does not provide equally straightforward synergy with antibody detection. POC tests are nowadays available for antibody screening of an increasing number of virus infections (HIV, HCV, varicella-zoster virus (VZV), cytomegalovirus (CMV), Epstein-Barr virus (EBV)).

Future Perspectives

Driven by public health, scientific and commercial interests, new diagnostic tests for laboratory diagnosis of viral infections are continuously designed. The development of new molecular detection methods continues on two lines. On the one hand, it includes compact, fully integrated automates requiring minimal training to run and suitable for POC use. On the other hand, separate instruments for specimen handling, serological assays, nucleic acid extraction, PCR set-up, and amplification are integrated into high throughput flow systems, which allow efficient use of both commercial and laboratory designed assays.

The use of multianalyte methods is becoming practical reality in antibody, antigen and nucleic acid detection and might significantly improve infectious disease diagnosis in the future. The current multiplex approaches, however, still lack a comprehensive library of compatible virus specific reagents (from which any combination of pathogens could be assayed). In antibody detection, multiplex approaches have so far been used mainly for research purposes. Multiplex PCR in a single tube or miniature format, on the other hand, still is to reach the sensitivity of standard PCRs targeting single or few analytes. Next-generation sequencing has paved the way to identification of new or emerging viruses. Simple microarrays, readable without bioinformatics, would reduce the cost of serological screening and allow the use of control measures to improved quality of results. Beyond classical virus diagnostics, analysis of virus-induced interferon gene response is an emerging scope in diagnostics. Inspired by single-molecule sequencing, state of the art EIAs also can detect individual antibodies or immunocomplexes and could significantly improve sensitivity. On the other hand, homogenous wash-free immunoassays can omit most hands-on steps that require expensive labor or instrumentation; their results can be read directly after adding a droplet of sample into a premade reagent mix. Use of noninvasive sample materials (urine or saliva) instead of whole blood, serum or plasma furthermore can allow for self-sampling, for convenience of the patient and cost-benefit of the society.

Further Reading

- Arnold, K.B., Chung, A.W., 2018. Prospects from systems serology research. *Immunology* 153, 279–289.
- Barzon, L., Percivalle, E., Pacenti, M., *et al.*, 2018. Virus and antibody dynamics in travelers with acute zika virus infection. *Clinical Infectious Diseases* 66, 1173–1180.
- Chaudhary, N., Wesemann, D.R., 2018. Analyzing immunoglobulin repertoires. *Frontiers in Immunology* 9, 462.
- Chevaliez, S., Pawlotsky, J.M., 2018. New virological tools for screening, diagnosis and monitoring of hepatitis B and C in resource-limited settings. *Journal of Hepatology* 69, 916–926.
- Dubot-Pérés, A., Sengvilaipaseuth, O., Chanthongthip, A., Newton, P.N., de Lamballerie, X., 2015. How many patients with anti-JEV IgM in cerebrospinal fluid really have Japanese encephalitis? *Lancet Infectious Diseases* 15, 1376–1377.
- Enders, M., Schalasta, G., Baisch, C., *et al.*, 2006. Human parvovirus B19 infection during pregnancy – Value of modern molecular and serological diagnostics. *Journal of Clinical Virology* 35, 400–406.
- Hedman, K., Seppälä, I., 1988. Recent rubella virus infection indicated by a low avidity of specific IgG. *Journal of Clinical Immunology* 8, 214–221.
- Landry, M.L., 2016. Immunoglobulin M for acute infection: True or false? *Clinical and Vaccine Immunology* 23, 540–545.
- Nurmi, V., Hedman, L., Weseslindtner, L., Hedman, K., 2020. Comparison of Approaches for IgG Avidity Calculation and a New Highly Sensitive and Specific Method with Broad Dynamic Range. *Viruses*, submitted.
- Parekh, B.S., Ou, C.Y., Fonjongo, P.N., *et al.*, 2018. Diagnosis of human immunodeficiency virus infection. *Clinical Microbiology Reviews* 32, e00064.
- Picone, O., Bouthry, E., Bejaoui-Olmann, Y., *et al.*, 2019. Determination of rubella virus-specific humoral and cell-mediated immunity in pregnant women with negative or equivocal rubella-specific IgG in routine screening. *Journal of Clinical Virology* 112, 27–33.

A Brief History of the Development of Diagnostic Molecular-Based Assays

Hubert GM Niesters, Department of Medical Microbiology and Infection Prevention, Division of Clinical Virology, University Medical Center Groningen, Groningen, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Isothermal Amplification

Companies have developed isothermal amplification technologies to conquer the market without having to deal with the PCR patent rights. These technologies are not suitable for the development of lab-developed assays. Quantitation is possible and the branched-DNA technology (bDNA) (Urdea *et al.*, 1993) (Fig. 1) and the NASBA or Nucleic Acid Sequence Based Amplification (Vandamme *et al.*, 1995) (Fig. 2) have been commercially introduced for the detection of HIV-1; the bDNA has also been introduced for the detection of HBV and HCV. Isothermal amplification technologies have been successful in the early days of molecular-based assays; however, with the rapid development of PCR-based technologies in a number of research laboratories, these isothermal amplification technologies cannot currently compete with the clinical diagnostic market anymore.

Only one isothermal amplification has been able to penetrate the diagnostic market and has retained its position successfully. This is the TMA or Transcription-Mediated Amplification (Stevens *et al.*, 2009) technology, currently owned by Hologic. This is a variant of the NASBA technology and can amplify RNA as well as DNA by using an RNA polymerase and a reverse transcriptase. This latter enzyme also exhibits DNA polymerase activity for single-stranded DNA or single-stranded RNA.

Finally, the so-called LAMP or loop-mediated isothermal amplification method (Notomi *et al.*, 2015) was developed by Notomi and coworkers. This method can be used for RNA as well as DNA targets and amplifies long loop-like structures containing the target sequences. Although it is very rapid and specific, its use in clinical virology is rather limited.

Non-Isothermal PCR Based Technologies

The first paper describing the PCR reaction (Fig. 3) was written by Saiki *et al.* in 1985 (Saiki *et al.*, 1985). With the introduction of commercially available heat-stable enzymes like the *Taq*-DNA polymerase and the subsequent recombinant made enzymes in 1989, the technology was made available for the scientific community, as well as the basis for the introduction of commercial assays. The first one made available was for the detection of HIV-1 in 1992. Detection of any virus did become possible, and with the implementation of a new enzyme, the recombinant Tth DNA polymerase, even the reverse transcriptase step needed to transcribe viral RNA into the first DNA step needed for the PCR reaction became possible.

The steps needed to implement these technologies into clinical practice were subsequently developed and introduced. The most important steps were to automate isolation of nucleic acids to ensure that larger quantities of patient materials could be processed. The methodology mostly used is based on the so-called Boom technology (Boom *et al.*, 2002) (Fig. 4), which involves extracting nucleic acids using silica particles in a high salt chaotropic solution. Also, automation in PCR cyclers was further developed to ensure the repeated cycling of the PCR reaction. But most importantly, the development of the 5'-nuclease technology in 1991 that opened the road to the introduction of so-called TaqMan probes enabled the detection of the amplification products generated by the PCR method in real-time (Lee *et al.*, 1993; Higuchi *et al.*, 1993).

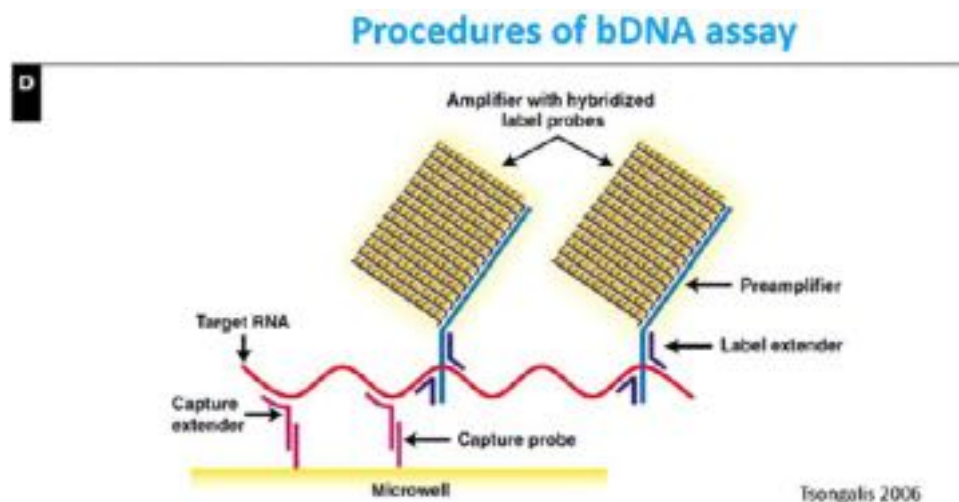


Fig. 1 Procedure of the bDNA assay (<https://www.youtube.com/watch?v=yi-QZvhSCws>).

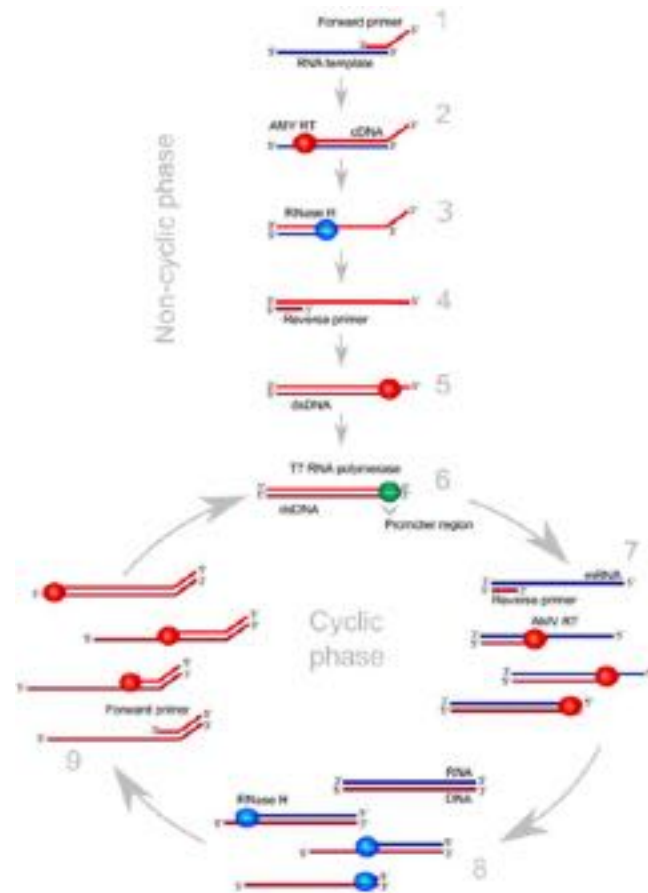


Fig. 2 Principles of the NASBA method.

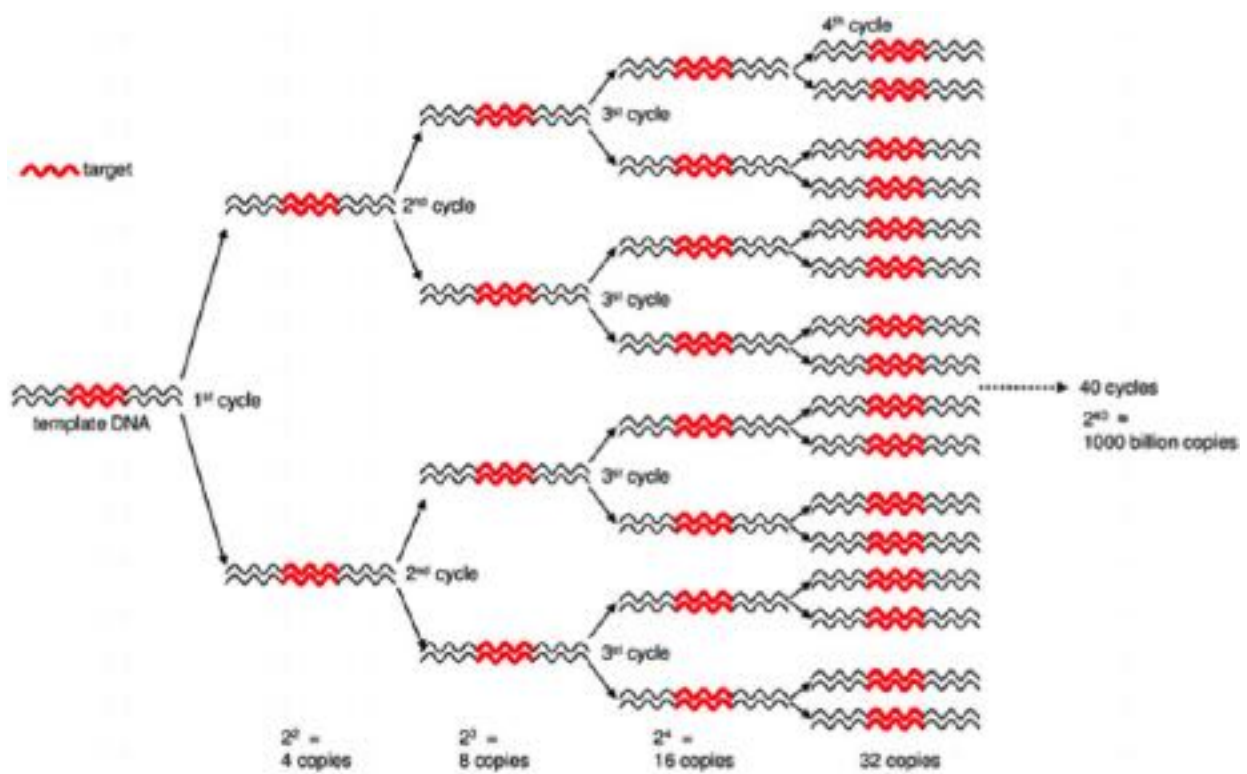


Fig. 3 Principle of the PCR method.

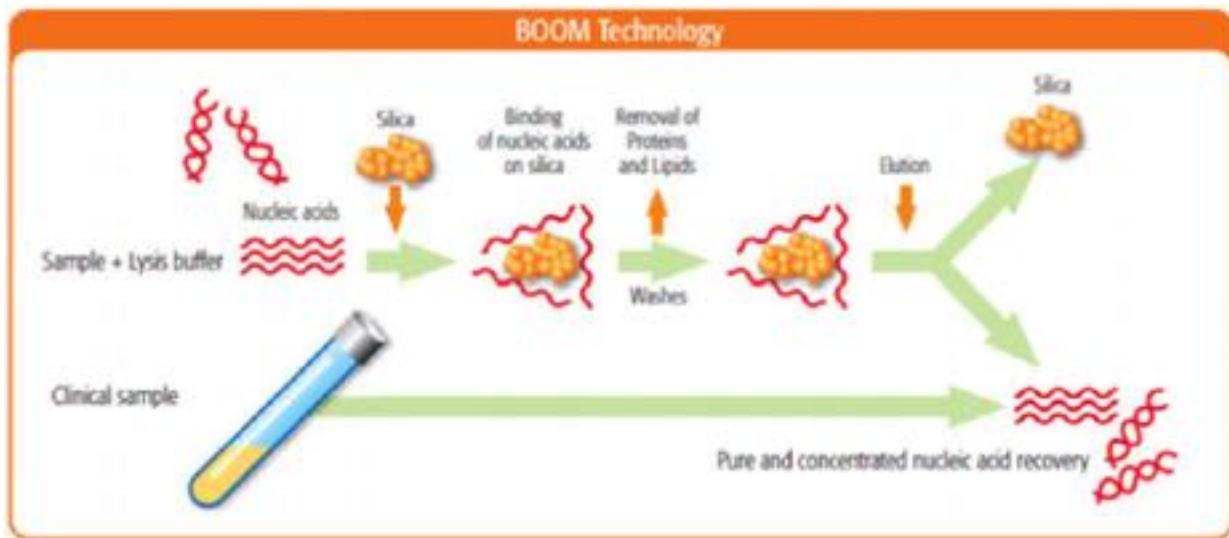


Fig. 4 The Boom method to isolate nucleic acids from clinical samples.

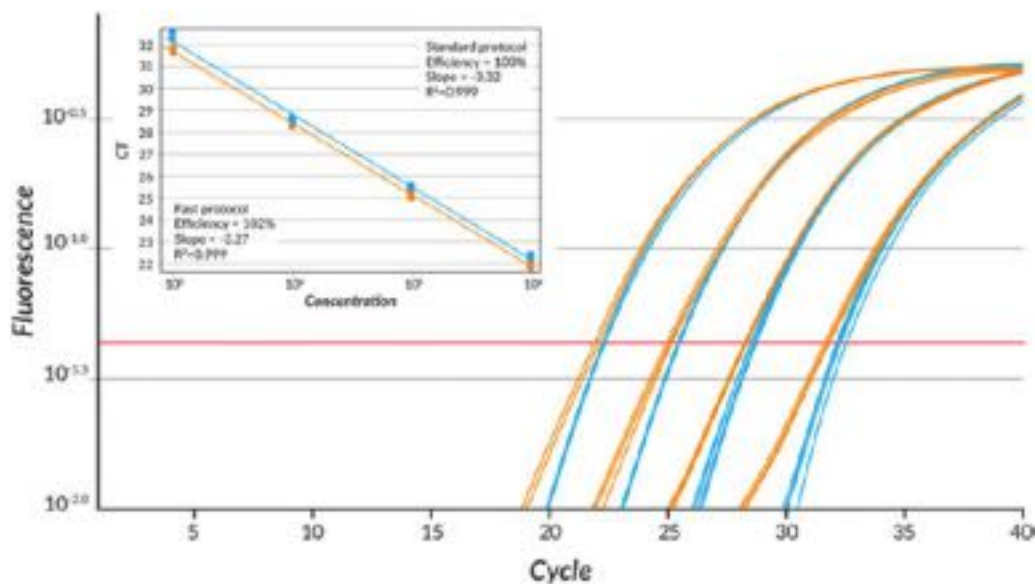


Fig. 5 Relation between Ct value and nucleic acid concentration. The higher the concentration, less cycles are needed to reach a plateau value. So low Ct value with high concentration. High Ct values indicate low concentration and more cycles are needed to reach the plateau value.

This real-time PCR method finally led to the introduction of the quantification of viral RNA or DNA in clinical samples. It opened up the ability to follow the effect of antiviral treatment in patients with, for instance, HIV-1, HBV, or HCV. In principle, every target could be quantified more accurately. To ensure standardization, reference materials were developed, and WHO international standards were slowly introduced. External quality control programs were set up, like from National Institute for Biological Standards and Control (NIBSC) (Shyamala *et al.*, 2004), Quality control for molecular diagnostics (QCMD) (Matheeußen *et al.*, 2020), or Institut für Standardisierung und Dokumentation (INSTAND), to ensure that data generated in different laboratories or by the assays developed by different commercial companies could be compared with one another. Also, with more strict accreditation guidelines in several countries like the ISO 15189:2015, more accurate and standardized data are being produced by clinical laboratories all over the world.

The possibility of detecting any viral target in any clinical sample has given us a lot of new options and imposes new challenges. We have the technologies to detect new viruses very quickly, as has been done with SARS-CoV-2. We can also quantify the amount of virus in clinical specimens based on the so-called cycle threshold (Ct) of the PCR (Fig. 5); a low Ct value of 12–20 implies a very high viral load, while a high Ct value, 35 or higher, implies a low viral load. In clinical follow-up based on consecutive samples over time, the first patient samples usually show low Ct values indicating high viral load and the follow-up samples show low and

Table 1 Overview of technologies used in clinical virology. Some examples of assays are mentioned, as well as the companies who have initially developed the technology. Some of them have been bought by other companies

<i>Molecular Based Technologies</i>				
		<i>Used common name</i>	<i>Targets detected</i>	<i>Developed by/Current company (not complete)</i>
Isothermal technologies	Signal Amplification	Branched DNA	HIV-1, HBV, HCV	Chiron/Bayer
	Target Amplification	Hybrid Capture NASBA	HPV HIV-1	Digene/Qiagen Organon-Teknika/bioMerieux
		TMA	HIV-1, HBV, HCV, Chlamydia trachomatis	GenProbe/Hologic
		LAMP		Eiken Chemical
Thermo Cycling Conditions	Target Amplification	Polymerase Chain Reaction	Any, also lab-developed	Mayor diagnostic companies (Roche, Abbott, bioMerieux, Hologic and many more) BioFire/bioMerieux
	Syndromic Testing	Polymerase Chain Reaction	Respiratory	
	Point-of-Care testing Single target	Polymerase Chain Reaction	Gastro-enteric Neurological targets HIV-1, HBV, HCV, M. tuberculosis	GenMark Qiagen Cepheid

subsequently not detectable viral load (high Ct value). The question still waiting to be answered is what the viral load level is that means that an individual is still infectious to others.

The question of clinical relevance is important with respect to the new CE-IVD guidelines, which come into effect in 2022. It is clear that for viruses like HIV, HBV or SARS-CoV-2, a patient should not have an infection with these viruses at all. Bert, I'm not sure I understand what you are saying here? Every positive test value, therefore, has clinical implications. However, for viruses that are latent in our body, like the herpesviruses CMV or EBV, the viral dynamic of the amount of target DNA is more relevant in relation to disease. But a rule of thumb is that low viral loads are better than higher loads. And there are always exceptions, such as HSV type 1, where detection has implications in specific clinical compartments. For example, a patient with HSV DNA in cerebrospinal fluid or in blood will be more severely ill than an individual with a vesicular eruption around the side of the mouth.

There are, however, still open questions that need to be answered. PCR technologies are currently one of the most used amplification technologies and further developments in automation and multiplexing of targets have resulted in systems that can detect around 15–20 targets simultaneously. These so-called syndromic testing solutions are currently used to detect many viral and non-viral targets in respiratory, gastroenteric, or neurological samples simultaneously. Companies have introduced cartridge-based systems to detect multiple targets on a single sample for a short period of time, mostly between 45 and 90 min. Although viral targets are often detected, the clinical interpretation can remain a challenge if more than a single target is detected. A relative quantitation number is not always provided because an internationally accepted standard is not made available yet (Dien Bard and Alby, 2018; Leber *et al.*, 2018; Babady *et al.*, 2018).

Topics related to the use of amplification technologies will be presented in several articles of this volume (Table 1).

References

- Babady, N.E., England, M.R., Jurcic Smith, K.L., *et al.*, 2018. Multicenter evaluation of the ePlex respiratory pathogen panel for the detection of viral and bacterial respiratory tract pathogens in nasopharyngeal swabs. *Journal of Clinical Microbiology* 56 (2), e01658. doi:10.1128/JCM.01658-17.
- Boom, R., Sol, C.J., Schuurman, T., *et al.*, 2002. Human cytomegalovirus DNA in plasma and serum specimens of renal transplant recipients is highly fragmented. *Journal of Clinical Microbiology* 40 (11), 4105–4113. doi:10.1128/jcm.40.11.4105-4113.2002.
- Dien Bard, J., Alby, K., 2018. Point-counterpoint: Meningitis/encephalitis syndromic testing in the clinical laboratory. *Journal of Clinical Microbiology* 56 (4), e00018. doi:10.1128/JCM.00018-18.
- Higuchi, R., Fockler, C., Dollinger, G., Watson, R., 1993. Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. *Biotechnology* 11 (9), 1026–1030. doi:10.1038/nbt0993-1026.
- Leber, A.L., Everhart, K., Daly, J.A., *et al.*, 2018. Multicenter evaluation of biofire filmarray respiratory panel 2 for detection of viruses and bacteria in nasopharyngeal swab samples. *Journal of Clinical Microbiology* 56 (6), e01945. doi:10.1128/JCM.01945-17.
- Lee, L.G., Connell, C.R., Bloch, W., 1993. Allelic discrimination by nick-translation PCR with fluorogenic probes. *Nucleic Acids Research* 21 (16), 3761–3766. doi:10.1093/nar/21.16.3761.
- Matheeußen, V., Corman, V.M., Donoso Mantke, O., *et al.*, 2020. International external quality assessment for SARS-CoV-2 molecular detection and survey on clinical laboratory preparedness during the COVID-19 pandemic, April/May 2020. *Eurosurveillance* 25 (27), 2001223. doi:10.2807/1560-7917.ES.2020.25.27.2001223.

- Notomi, T., Mori, Y., Tomita, N., Kanda, H., 2015. Loop-mediated isothermal amplification (LAMP): Principle, features, and future prospects. *Journal of Microbiology* 53 (1), 1–5. doi:10.1007/s12275-015-4656-9.
- Saiki, R.K., Scharf, S., Faloona, F., *et al.*, 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230 (4732), 1350–1354. doi:10.1126/science.2999980.
- Shyamala, V., Cottrell, J., Arcangel, P., *et al.*, 2004. Detection and quantitation of HBV DNA in the WHO international standard for HIV-1 RNA (NIBSC code: 97/656). *Journal of Virological Methods* 118 (1), 69–72. doi:10.1016/j.jviromet.2004.01.021.
- Stevens, W.S., Noble, L., Berrie, L., Sarang, S., Scott, L.E., 2009. Ultra-high-throughput, automated nucleic acid detection of human immunodeficiency virus (HIV) for infant infection diagnosis using the Gen-Probe Aptima HIV-1 screening assay. *Journal of Clinical Microbiology* 47 (8), 2465–2469. doi:10.1128/JCM.00317-09.
- Urdea, M.S., Wilber, J.C., Yeghiazarian, T., *et al.*, 1993. Direct and quantitative detection of HIV-1 RNA in human plasma with a branched DNA signal amplification assay. *AIDS* 7 (Suppl 2), S11–S14. doi:10.1097/00002030-199311002-00004.
- Vandamme, A.M., Van Dooren, S., Kok, W., *et al.*, 1995. Detection of HIV-1 RNA in plasma and serum samples using the NASBA amplification system compared to RNA-PCR. *Journal of Virological Methods* 52 (1–2), 121–132. doi:10.1016/0166-0934(94)00151-6.

Sequencing Strategies

Sibnarayan Datta, Defence Research Laboratory, Defence Research and Development Organisation (DRDO), Tezpur, Assam, India

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

Gigabase (Gb) 1,000,000,000 bases

Kilobase (Kb) 1000 bases

MCPyV Merkel cell polyomavirus

Nanometer 10^{-9} m

SARS-CoV Severe acute respiratory syndrome-related coronavirus

SV40 Simian vacuolating virus 40

Zeptoliter 10^{-21} l

Glossary

Base-calling The computational process of translating the raw output (chemiluminescence, fluorescence, change in pH or current flow, etc.) of the sequencing platforms into nucleotides.

Contig (Contiguous sequences) are long DNA segments, reconstructed from the assembly of short sequences, based on the degrees of overlap among them.

De novo sequencing Denotes sequencing a novel genome for which no reference sequence is available for alignment.

Deep sequencing Deep sequencing is a next-generation sequencing approach denoting sequencing a particular genetic region multiple times, often hundreds or thousands of times. This approach allows detection of rare viruses or viral quasiespecies that comprise less than 1% of the sample.

Degenerate primers Degenerate primers are a pool of similar but non-identical oligonucleotide primers, designed from an alignment of closely related sequences that have certain variabilities at the primer binding sites. These primers are useful and amplify similar genetic regions from closely related viral genomes.

High-throughput sequencing This term is associated with the next-generation sequencing approach, which allows sequencing of hundreds of millions of DNA molecules in parallel, resulting in large volumes of data output.

Oligonucleotide primer Oligonucleotides are short DNA or RNA molecules (mostly 10–50 bases) which bind to specific regions of target nucleic acids and act as an initiation point for amplification or sequencing reactions.

Polymerase chain reaction PCR is one of the most widely used molecular techniques for nucleic acid amplification (generating several million copies of the target DNA segment), which uses a thermostable DNA polymerase to mimic DNA replication, *in vitro*.

Reverse transcription A reaction to generate complementary DNA (cDNA) from RNA using reverse transcriptase (RT) enzymes.

Shotgun sequencing This is a rapid technique for sequencing large genomes. This technique involves fragmenting large pieces of nucleic acids into smaller ones, which are then randomly sequenced on NGS platforms. Resulting short sequences are assembled to reconstruct larger contiguous pieces or contigs.

Viral quasiespecies Evolution of viruses results in populations of closely related mutants that are beneficial to virus persistence in hosts or help the virus by immune escape.

Zero-Mode Waveguide (ZMW) It is a photonic nanostructure that allows observation of fluorescent reactions occurring at a single-molecule level.

General Overview

During the last few decades, risk of emerging and re-emerging viral diseases has significantly increased globally. Due to diverse human activities – including rapid and unplanned deforestation for agriculture and urbanization purposes, humans, animals, and plants are increasingly being exposed to wildlife, resulting in transmission and outbreaks of variant as well as new viruses. Moreover, improved facilities for global trade and transportation have hastened the transfer of viruses emerging in one part of the globe to other parts. Recent episodes of emergence/re-emergence and spread of Ebola, Nipah, Zika, Hantavirus, SARS (Severe acute respiratory syndrome), virulent forms of Influenza viruses, MERS (Middle east respiratory syndrome coronavirus), and SARS-CoV-2 viruses attest to this changing epidemiology and predict future outbreaks with still unknown viruses. Additionally, detection and monitoring of viruses is also challenging due to their wide variability and rapid evolution. Apart from these natural consequences, a potential hazard of use of microbes, especially deadly human/animal viruses or highly damaging agricultural viruses, for acts of bioterrorism/agroterrorism, has also increased in recent years. Irrespective of the source, exact identification of the etiological agent is of prime importance to understand transmission routes and to formulate management strategies including quarantine, development of diagnostic tools, and therapeutic/prophylactic possibilities. Since most of the viruses are uncultivable *in vitro*, exact identification is primarily based on molecular virology methods such as PCR (Polymerase chain reaction) and/or nucleic acid sequencing.

Due to their high evolutionary rates and wide diversity, amplification and sequencing of viral genomes is a big challenge to researchers. Before the availability of molecular biology tools, viral identification was mostly based on infection of vulnerable cell cultures, followed by morphological analyses of cytopathic changes in the cells and visualization of virus particles by

electron microscopy. These classic methods were labor-intensive, time-consuming, and mostly unsuccessful as most of the viruses are uncultivable.

During the last four decades, enormous development in science and technology has revolutionized the field of virology by obviating culture-based methods, significantly reducing the turnaround time, and allowing identification of non-cultivable viruses. Among the molecular tools, discoveries of rapid nucleic acid amplification and sequencing technologies have played a crucial role in understanding viral genome and evolution, resulting in development of diagnostics, therapeutics, and prophylactic agents. Sequencing has now become a “gold standard” technique in virology and with the introduction of next-generation technologies, speed of virus genome sequencing has exponentially increased the number of virus taxa and viral genomes in databases by permitting identification and taxonomic classification of numerous novel viruses.

As a result of gradual evolution of sequencing methods and enormous technological refinements, three distinct generations of sequencing technologies are presently available. Each of these technologies has its own features and has distinct applications in virology. This article is intended to introduce the reader to available sequencing technologies and their specific applications in the field of virology so that the readers can decide the sequencing strategy best suited for their research.

First Generation Sequencing Technology

Efforts to sequence nucleic acids started just after description of the double-helical model of DNA by Watson and Crick. Interestingly, it was an RNA molecule that was first sequenced. For the first time, nucleotide sequence of one of the smallest biologically active nucleic acid molecules (alanine transfer RNA, tRNA^{Ala}) was reported by Robert Holley in 1965, for which he was awarded the Nobel Prize in Physiology or Medicine, in 1968. However, it took quite a long time to come across suitable techniques to routinely sequence larger nucleic acids. During the 1970s, two brilliant methods for sequencing DNA strands were discovered and published almost simultaneously: The “dideoxy chain termination” method developed by Frederick Sanger and colleagues and the “chemical modification” based method developed by Allan Maxam and Walter Gilbert. These two techniques established the first-generation sequencing technologies and were revolutionary discoveries in the field of molecular biology, for which Frederick Sanger and Walter Gilbert were honored with the Nobel Prize in Chemistry in 1980.

Initially, Maxam–Gilbert’s chemical modification-based sequencing method became instantly popular due to its simplicity in sample preparation, as compared to the Sanger’s dideoxy chain termination method. However, introduction of fluorescent-tagged dideoxy chain terminators made Sanger’s method more suitable for automation. The discovery of PCR further simplified sample preparation for sequencing by Sanger’s technique, significantly shortening turnaround time. Gradual refinement in instrumentation made this technique ready for high-throughput automated sequencing of comparatively large genomes. Due to its accuracy and robustness, Sanger’s sequencers are still the method of choice for validation of sequences obtained from technologically even more advanced next-generation sequencing (NGS) platforms. On the other hand, due to its limited prospect for automation, use of hazardous chemicals, and difficulties with sequencing long strands, use of Maxam–Gilbert’s technique gradually declined. Nevertheless, during the initial stages of development and testing of different sequencing techniques, viruses, primarily bacteriophages – served as a source for high-quality DNA in sufficient quantities, since they were easy to culture in the laboratory and had relatively smaller genome sizes. Bacteriophage phiX174 and MS2 DNA were the first to be sequenced, while the SV40 genome was the first animal virus genome to be sequenced.

The Sanger sequencing technique, also known as “chain-termination sequencing” or “dye-terminator sequencing”, is principally based on the incorporation of chain-terminating dideoxynucleotides. In its present format, in a PCR-like reaction, template DNA is copied into single strands of DNA complementary to the target template using a DNA polymerase, a specific oligonucleotide primer, normal deoxynucleotides (dNTPs), and a lower molar concentration of four different dideoxynucleotides (ddNTPs, each being tagged with a particular fluorescent dye). While polymerization of nucleotides occurs by chance, whenever a particular ddNTP is incorporated into the growing chain, further polymerization is inhibited due to the absence of the 3′-hydroxyl group in the ddNTPs. Through repeat cycling of this reaction, *n* numbers (*n* being the number of bases in the template DNA) of DNA fragments of variable lengths (each different by a single nucleotide) are generated, having a fluorescently labeled dideoxynucleotide at the 3′ end.

These tagged amplicons are then electrophoretically separated through a capillary filled with an optimized polymer matrix, capable of resolving single-stranded fragments differing by only one base. As a result, a continuous flow of fragments is generated, with shorter fragments moving faster, while the longer ones move slower. At the capillary terminal, fluorescent tags on the fragments sequentially passing through a laser are excited, while a photoreceptor placed opposite registers the wavelength of the light emitted by the excited tags. Since each of the four ddNTPs are tagged with four distinct fluorescent dyes having precise emission spectra, a chromatogram of fluorescent peaks is generated by the continuous flow of tagged fragments through the laser-detector couple. Concurrently, depending upon the wavelength and intensity of a particular fluorescent tag, corresponding base is identified (technically termed “base-calling”) and this continuous “base-calling” represents the nucleotide sequence in the target sequence. Years of refinement in technologies and chemistries associated with this sequencing technique has resulted in highly reliable Sanger sequencing platforms that can analyze accurate sequence information of up to 900–1000 bases. Due to its unmatched reliability, this technology has dominated for almost four decades and is still considered the “gold standard” for DNA sequencing. Readers are encouraged to visit the website with the link provided at the end of this article. These online resources present sequencing technologies in the form of animated videos for better understanding.

Owing to the need for pure, homogenous DNA, prior to the discovery of PCR, application of Sanger sequencing in virology was initially limited to cultivable viruses, including bacteriophages and certain other animal viruses for which cell culture was available. This requirement limited its wide application in virus sequencing, directly from human, animal, and plant samples, as viruses present in these are mostly heterogeneous and are present with large quantities of DNA from the host and other contaminating microorganisms. The discovery of PCR in the late 1980s by Kary Mullis (honored with the Nobel Prize in Chemistry in 1993) and coupling of reverse transcription with PCR (RT-PCR) entirely transformed the state of virus sequencing by facilitating rapid and targeted amplification of minute amounts of viral nucleic acids (DNA/RNA) from complex biological samples, hallmarking the era of culture-independent targeted sequencing of viruses.

Second Generation or Parallel Sequencing Technologies

While the Sanger technology contributed enormously and significantly towards virus detection and discovery projects, evolution and commercial development of parallel sequencing technologies (aka next-generation sequencing, NGS) since 2000 has entirely changed the application and magnitude of sequencing in almost every field of biological research, including virology. In fact, the necessity to expand the scale of sequencing by several orders of magnitude to accomplish large-scale genome projects, like the human genome project, led to the invention of parallel or second-generation sequencing technologies. The glory of parallel sequencing technologies resides in its ability to simultaneously sequence millions of strands of heterogeneous nucleic acids with high sensitivity. Furthermore, unlike Sanger sequencing, parallel sequencing technology has enabled sequencing of samples in a sequence-independent manner, obviating the need for prior knowledge of the nucleic acid sequences. These second-generation sequencing technologies were built on newer sequencing chemistries, several innovations in microfluidics, amplification methodologies, and detection technologies. Since these platforms sequence nucleic acids in a massively parallel fashion, the volume of data generated by these platforms is huge, which prompted rapid advancement in the field of computation and bioinformatics.

Unlike the Sanger method where a single large stretch of DNA is sequenced in each run, NGS platforms require larger DNA molecules to be fragmented into smaller sizes, which are then sequenced in parallel, and millions of short “sequence reads” are generated in each run. These short reads are subsequently assembled into larger contigs through complex computer programs. Due to their extremely large dynamic range of detection and sensitivity, presently available bench-top NGS platforms can sequence complete or near-complete genomes of viruses (known, unknown, or even unexpected) in a single run and within a few hours. NGS has the power to resolve sequences of minor target genomes from a highly complex mixture of non-target nucleic acids, where a computational subtraction method is used to remove non-target sequences (such as that of host sequences) while retaining the target virus sequences. Although, in such circumstances, yield and coverage of target viral sequences are often low, as the predominant non-target nucleic acids fiercely compete with the low abundant target viral nucleic acids for resources during sequencing. As an alternative to this approach, sensitivity, specificity, and turnaround time for NGS-based viral genome sequencing applications can be enhanced dramatically by enrichment of virus particles/genomes or depletion of host/non-target nucleic acids through ultracentrifugation/filtration and nuclease treatment-based methods. Nucleic acids can then be extracted from the resulting enriched viral particles, sequenced directly on NGS platforms or can be amplified through non-specific application techniques such as SISPA (Sequence-independent, single-primer amplification), VIDISCA (Virus Discovery based on cDNA Amplification), RCA (Rolling Circle amplification) and others, followed by running on NGS.

Unlike Sanger sequencing, NGS does not require target-specific primers or any other sample-specific components, thereby allowing a uniform sequencing methodology that is similar for virtually all sample types. These features make NGS systems an indispensable tool in modern viral outbreak investigation. Once the sequence of a potential causative agent is determined, specific diagnostic assays could then be developed easily to monitor and contain further spread of the agent.

Apart from the discovery of new viruses, NGS has numerous applications in virology, which include the study of virus integration sites, evolution, quasispecies, and emergence of antiviral/vaccine-escape mutations.

Virus particle enrichment and nucleic acid amplification protocols coupled with NGS technologies have further enabled the efficient metagenomic study of viromes associated with humans, plants, and the environment. Over the years, NGS has become economically affordable, smaller in size (bench-top), while speed, quality, and quantity of data output have dramatically increased. Available benchtop models offer incredible NGS capabilities at very affordable costs.

NGS technologies were commercially popularized through the 454 FLX pyrosequencer (Roche) and the Illumina Genome Analyzer (Illumina) and later, by the SOLiD platform (Life Technologies). Despite its huge success, Roche has discontinued marketing of its pyrosequencing based NGS platforms, although many of these instruments are still used in several laboratories. Eventually, rapid research and development has resulted in newer NGS platforms like the Ion Torrent PGM (Life Technologies). Although Roche 454 and Illumina are the most extensively represented technologies in virology related literature, in recent years, SOLiD (Sequencing by Oligonucleotide Ligation and Detection) has also gradually emerged as a robust and reliable platform.

Even though each of these platforms exploit different sequencing chemistries and detection technologies, basic procedure is similar in principle and based on the “shotgun” sequencing technique. Initially, nucleic acid samples are fragmented by mechanical (ultrasonic based) or enzymatic (transposon-based) methods to produce nearly uniform length fragments between 100 and 1000 bases, depending on the sequencing platform being used. Mechanical fragmentation is preferred for clinical and environmental samples since it can withstand possible impurities that may inhibit enzymatic fragmentation methods. Enzymatic

methods are suitable for relatively pure samples since fragmentation and adapter ligation can be done in a single reaction, thereby reducing time and handling steps for library preparation. Subsequently, a fragment library is prepared by ligation of adapters (for attachment to the sequencing matrix) and sample-specific barcodes (to identify the sample source after sequencing) to each of the fragments, attachment of each fragment on the sequencing matrix, amplification of the attached fragments and, finally, sequencing takes place. Different NGS platforms use their proprietary adapters and barcodes to fragment libraries for amplification (e.g., emulsion PCR in 454 & SOLiD platforms, and bridge amplification in Illumina platforms). Using distinct barcodes for each sample, samples from different sources can be prepared and sequenced collectively, making the maximum use of high throughput NGS sequencers, thereby cutting-down the per-sample sequencing costs.

Different NGS platforms use distinct sequencing chemistries and detection technologies. The 454 platforms use sequencing-by-synthesis (SBS) chemistry where chemiluminescence, generated during incorporation of a nucleotide, is detected. Illumina platforms also use the SBS technology with reversible fluorescent dye-terminator chemistry where fluorescence signals generated during incorporation of nucleotides are detected. Similarly, Ion-Torrent also uses SBS technology with a novel semiconductor-based detection technology, which registers the drop in pH that occurs during incorporation of each specific nucleotide. Simply based on detection of voltage change, semiconductor-based signal capture technology is much faster, low-cost, and smaller than fluorescence or chemiluminescence detection-based technologies that require complex and sophisticated optics for fluorescence/chemiluminescence image capture and advanced computational tools for image processing and data analyses. On the other hand, the SOLiD platform uses a rather complex sequencing-by-ligation (SBL) technology that uses a mixture of fluorescent-tagged oligonucleotide probes of various lengths and the mismatch sensitivity of DNA ligase to sequence the target DNA. Irrespective of the sequencing chemistry (SBS or SBL) used in these platforms, chemiluminescence/fluorescence image or pH changes occurring on the sequencing chip is collected periodically, integrated, and quality-assessed through computer-intensive complex image/voltage processing algorithms to infer actual nucleotide sequence data from each fragment, which typically results in several gigabases (Gb) of sequence data. Despite being extensively used, different NGS platforms have certain limitations including base calling bias, reproducibility, errors associated with certain sequence motifs, or homopolymers containing regions. Therefore, it is recommended to get samples analyzed on two different NGS platforms simultaneously to reduce errors, albeit depending on the availability of funds. Nevertheless, with consistently improving chemistries, hardware, and software, many of the issues have either been resolved or will be resolved in newer platforms.

For applications in virology, especially virus discovery and metagenomics, NGS platforms that can yield longer sequence reads are preferred since long reads are more suitable for *de novo* assembly of unknown genomes. On the contrary, for analysis of virus evolution and quasispecies, platforms that generate shorter but high-quality reads with increased depth should be preferred. For virus sequence detection, at certain levels of the target template, Illumina has been shown to have analytical sensitivity comparable to that of optimized qPCR but, at low target levels, reads generated by the Illumina platform are insufficient for *de novo* assembly of the genomes. Despite massive refinement in computational procedures for analysis of huge volumes of NGS data, *de novo* assembly and identification of short viral reads is still challenging, particularly for new viruses or highly variable viruses for which reference complete genome sequences are not available in the relevant databases. Therefore, development of a standardized universal bioinformatics pipeline for analysis of NGS data is required before these technologies can be adopted for routine viral diagnostics.

Application of NGS in Virology

NGS has been extensively used by researchers for several interesting applications in virology. One of them is vector enabled metagenomics (VEM), where NGS analysis of nucleic acids of insect vector specimens collected in the field could be done to monitor viruses present in a given environment. Using this method, new plant and zoonotic viruses have been identified in different insects, such as whiteflies and mosquitoes, providing a powerful tool for vector-borne virus discovery and surveillance programs. Similarly, using NGS, clinically important novel viruses were detected in gastrointestinal tissue samples obtained from bats that are believed to be reservoirs for several emerging zoonotic viruses (Paramyxoviruses, SARS-CoV, Lyssaviruses, Filoviruses). The first complete genome sequence of the SARS-CoV2, etiological agent of the ongoing COVID-19 global pandemic was also determined by performing direct NGS of bronchoalveolar lavage fluid RNA extract from a patient suffering from severe pneumonia.

Apart from directly detecting viruses in samples, an alternate strategy of virus discovery through NGS based sequencing of total small RNAs (sRNAs) from invertebrates is gaining importance. RNA silencing-based antiviral defense mechanisms in eukaryotes respond to viral infection by processing the invading viral RNA genomes into sRNAs of distinct lengths (~21–24 nucleotides). NGS-based deep sequencing of these sRNAs rich in processed viral genome, followed by assembly, could reconstruct the sequence of the original viral genome. This strategy has been successfully demonstrated in describing known and unknown viruses from diverse plants and animals, showing the universality of the mechanism. However, this strategy is challenging in cases of multiple infections with novel viruses, where small sequence reads from different viruses may interfere with the *de novo* assembly of each other, resulting in incorrect identification.

Very recently, studies have successfully reconstructed complete or nearly complete ancient viral genomes, such as the hepatitis B virus, from archeological human genome NGS datasets through computational methods of screening for specific viruses. This has opened a new field for analyzing ancient viral sequences and the study of viral evolution over long periods of time. A similar approach of transcriptome subtraction has been used in the discovery of new viruses such as MCPyV associated with Merkel cell carcinoma. Additionally, deep sequencing of amplicons generated by primers that target highly conserved genetic regions within a virus family has led to the discovery of several new polyomavirus and papillomavirus genotypes. Deep sequencing NGS technology

has also been used to determine retrovirus integration sites in natural infections, as well as those resulting from retroviral vector-based gene therapy experiments.

Choosing an NGS technology among the platforms available today is very important for specific applications, especially in virology. Sequencing viruses where a reference sequence is not available involves post-sequencing *de novo* (reference-independent) assembly of sequence reads. For such projects, datasets containing relatively long reads are preferable. On the other hand, when a reference genome is available, short-read technologies could be used to obtain high coverage of the genomes. Earlier, due to relatively long reads, pyrosequencing technology was the preferred NGS technology for metagenomic applications in virology. Following the discontinuation of pyrosequencing platforms in 2016, the consistent improvement in the read length of other NGS platforms has now enabled these alternative NGS technologies for wider applications in virology.

Third Generation Sequencing Technologies

Subsequent to the second generation of sequencers, another class of parallel sequencers have arrived, the third-generation sequencers that are capable of sequencing extremely long stretches of DNA, generating sequence reads in kilobases. These technologies are, therefore, popularly termed “long-read sequencing” parallel technologies to distinguish them from “short-read sequencing” second-generation technologies. In comparison to the second-generation technologies that break long strands of nucleic acids into smaller fragments before amplification and sequencing, third-generation technologies neither need fragmentation nor amplification (thereby reducing time, cost, and amplification bias) and directly read long stretches of nucleic acids at the single-molecule level, providing nucleotide sequence in real-time. Although wide applications of the third-generation technologies are largely restricted due to their relatively high running costs and limited throughput capabilities. Since these technologies are still under active development, these limitations are expected to be addressed in the near future.

Two of the third-generation sequencing technologies are commercially available, namely, the single-molecule real-time (SMRT) technology (Pacific Biosciences) and the Nanopore sequencing technology (Oxford Nanopore Technologies). In SMRT technology, sequencing takes place inside nano-phonic structures having zero-mode waveguide (ZMW) properties, which enables illumination of extremely small observation volumes (as small as zeptoliter), small enough to detect optical reactions occurring at the single DNA molecule level. Each SMRT chip contains many ZMWs and within each of these, one single-stranded DNA template with an active DNA polymerase is immobilized at the bottom of the well. Every time the polymerase incorporates a nucleotide into the growing complementary DNA strand, the dye linked to the incorporated nucleotide is cleaved off, causing it to illuminate the bottom of the wells and decay exponentially, which is detected by the optical sensors at the bottom of each ZMW. Optical signals from each of the ZMWs are detected continuously, resulting in real-time, parallelized single-molecule DNA sequencing. With presently available SMRT technology, extremely long reads ranging from > 190 kb with > 99.999% consensus accuracy can be achieved.

On the other hand, nanopore sequencing uses biological nanopores (nanometer diameter) formed by transmembrane proteins (porins), such as the heptameric transmembrane α -hemolysin (α -HL) from *Staphylococcus aureus*. Like SMRT, this technology also does not need amplification or any other modification to the nucleic acids. Using the principles of electrophoresis, nucleic acid strands are transported through the nanopores under a steady flow of ions. As the density of current flowing across the nanopore surfaces depends on the dimensions of the pores, there is a change in current flow across the nanopore depending on the dimension of the nucleotides occupying the pore. This modulation in current flow is detected by highly sensitive sensors and compared with previously optimized data to interpret the nucleotide sequence from current flow change data, in real-time. Presently, nanopore sequencers as small as USB drives (weighing only ~ 100 g) are commercially available, which offer sufficiently long-read sequence data amounting up to 30 Gb of DNA sequences or 12 million reads of RNA. Being simple, requiring no DNA pre-treatment or amplification, operable at a wide range of conditions, this technology has allowed researchers to sequence samples collected at locations in the field. This technology has been successfully used in a number of field applications such as detection and monitoring of viral pathogens, such as during Ebola outbreaks, genotyping, antibiotic resistance, food safety to humans, and plant genome sequencing.

Sample Preparation Strategies for Sequencing Viruses

Most of the time, virus samples are present with abundant non-target nucleic acids. Therefore, successful sequencing of virus nucleic acids needs specific strategies for sample preparation to achieve desired results.

Nucleic Acid Sequence Dependent Strategies for Known Viruses

Sequencing of viruses, whose partial or complete genome is known, can be easily done by PCR/RT-PCR mediated amplification of the genome using specific primers, followed by sequencing on a Sanger platform. Owing to its high sensitivity, specificity, and capability to amplify non-cultivable viral nucleic acids from a mix of nucleic acids, introduction of PCR and RT-PCR based techniques have almost replaced classical methods, enabling rapid sequencing of viruses. However, unlike the highly conserved prokaryotic rRNA or the eukaryotic ITS genetic regions, there are no ubiquitously conserved genetic regions across all the virus families to design universal viral amplification PCR primers that can broadly amplify all virus families for sequencing-based identification. Therefore, for diagnostic purposes, primers can be designed targeting highly conserved regions in a particular virus

family/genus for broad identification of variant or novel viruses belonging to the same family/genus, while type-specific primers can be designed by including type-specific variation at the 3' end. Amplicons generated can then be directly sequenced on a Sanger platform for confirmation. This strategy can be used for sequencing pure virus preparations (culture supernatant/ultracentrifuge purified samples) as well as field samples (from infected plants, clinics, or environment) where the target virus type is present in a mix of abundant contaminating nucleic acids. Generally, primers are designed to amplify 600–800 bp fragments (which is the limit of Sanger sequencing per run), having an overlap of at least 100–150 bp (for post-sequencing assembly of the fragments) with the neighboring fragment. Alternatively, virus genomes can be PCR-amplified to generate large amplicons (1–2 Kb) that can then be sequenced using nested overlapping primers covering 500–600 bases.

Due to high genetic variability, it is often difficult to design virus family-specific PCR primers or even genus-specific PCR primers, especially for RNA viruses. So, for sequencing a particular virus, primers can be designed based on the type specimen sequence directly. However, for sequencing related viruses (different strains/types belonging to the same group), PCR primers must be designed to target highly conserved regions. For this, if available, complete genome sequences can be obtained from the database (e.g., GenBank RefSeq database) for the target virus and its related viruses, align them using a standard nucleic acid alignment algorithm, and manually inspect the alignment for highly conserved regions that flank a variable region within PCR amplifiable length. The conserved regions allow the design of broadly reacting primers, while the intervening variable regions allow the identification of the specific virus after sequencing. Designing broadly acting virus family/genus consensus primers is technically challenging and may require the introduction of a certain degree of degeneracy in the primers. However, degeneracy should be kept at a minimum, since increasing degeneracy compromises the specificity and sensitivity of the amplification.

Since the Sanger method sequences a single strand of DNA at a time, it is essential to ensure that amplicons processed for direct sequencing are homogenous (i.e., no sequence variation among the amplicons). If related viruses are present, variants and quasispecies are expected in samples, which is very usual in field samples; cloning based library preparation of the PCR/RT-PCR amplicons should be carried out before Sanger sequencing. This can be done by simply using commercially available T-A/U-A based cloning vectors for amplicons generated by conventional Taq DNA polymerases. Alternatively, cloning can be carried out through engineering specific restriction endonuclease (RE) digestion sites at the 5' ends of the amplifying primers while designing, subsequently PCR/RT-PCR generated amplicons that can be digested by the respective RE, ligated into a suitable cloning vector and clonal DNA amplified, and isolated through routine laboratory methods. This ensures that an individual amplicon is present in each clone. Finally, random sequencing of DNA from individual clones is done to resolve the variants and quasispecies. In general, the more clones sequenced, the better the chances are for detecting minor populations of variant viruses in the sample.

Nucleic Acid Sequence Independent Strategies for Sequencing

In contrast, sequence-dependent amplification strategies are not applicable for viruses where absolutely no sequence information is available in the relevant databases. Such unknown viruses can also be sequenced on the Sanger platform by adopting sequence-independent metagenomic sequencing approaches. This does not require any prior knowledge of sequences present in a given sample.

The first step involves physical enrichment of virus particles from the sample and depletion of host genetic material or non-target genetic material through nuclease digestion. Enrichment of virus particles starts with homogenization of the samples and suitable dilution for removal of particulate matter (coarse particles, bacteria, host cells) through centrifugation. Following centrifugation, samples can then be ultra-centrifuged through a suitable medium such as a cesium chloride density gradient to obtain pure virus particles. However, due to the requirement of expensive instrumentation, chemicals, and technical knowledge for ultracentrifugation dependent methods, alternative methods based on simple membrane ultra-filtration – through an appropriate pore size (0.45 μ and 0.22 μ), or PEG (Polyethylene glycol)-based virus particle precipitation are frequently used. Filtered or PEG precipitated intact virus particles are then treated with nucleases to remove non-viral or host-associated nucleic acids. Subsequently, nuclease resistant, capsid-protected viral nucleic acids are extracted and purified for downstream enzymatic reactions for cDNA preparation, conversion of single-stranded DNA/cDNA into double-stranded DNA, amplification of the viral genetic material using sequence-independent priming methods such as SISPA, virus discovery based on cDNA-AFLP (VIDISCA), or RCA to produce sufficient amounts of genetic material for sequence analyses.

The Phi29 DNA polymerase-based RCA does not need prior virus particle separation, since this enzyme is capable of selectively amplifying circular DNA templates in a complex mix or non-target genetic material, thus very useful for enrichment of a number of human, animal, and plant viruses that have circular DNA genomes. These amplified viral genetic materials can then either be directly sequenced on NGS platforms or cloned to generate vector libraries for Sanger sequencing using vector-specific primers. These pre-sequencing enrichment methods reduce huge amounts of background non-target sequence data, thereby increasing the sensitivity of the sequencing methods and sequence coverage of the target nucleic acids. Nevertheless, owing to its parallel sequencing capabilities, application of target enrichment methods followed by NGS in viral metagenomics can immensely improve the chances of discovery of highly divergent or novel viruses compared to PCR-Sanger sequencing techniques. Although alleged to be associated with PCR bias, these methods have been extensively used for the discovery of a number of important viruses, especially in clinical, veterinary, and plant samples. Being based on non-specific amplification methods, extreme care should be exercised to prevent cross-contamination from the laboratory environment while performing sequence-independent amplification experiments.

Virus Databases

Successful completion of any sequencing project leads to the beginning of another involving analyses. Traditionally, after inspection and manual editing of the sequence data from a Sanger sequence, data are submitted to BLAST (Basic local alignment search tool) for a homology search (based on similarity) to determine if the sequence is the same as expected or is a potentially new sequence (no significant match in the database). The latter may warrant further analysis. On the other hand, analyses of large volumes of NGS data need specialized hardware and software solutions for storage and analyses. There are also several data analysis servers that can do the basic analyses as required. Primarily, short-read NGS data requires assembly of the short sequence stretches to create contigs (larger fragments, assembled on the similarity of the overlapping regions in short reads), which are then subjected to homology searches. Nevertheless, irrespective of the source of sequence data, its organization, analysis, and identification require searching for homology in the reference sequence data archived in the international databases. Since sequence databases play an important role in primer designing and sequence analysis, a certain amount of knowledge about databases is essential before embarking on a sequencing project.

Instantly after the discovery of the first-generation sequencer, the volume of sequence data started to increase in molecular biology laboratories and there was a high requirement for data storage and retrieval management systems.

Initially, nucleic acid sequence databases were largely localized to laboratories. In 1979, the Los Alamos Sequence Database was created at the Los Alamos National Laboratory, New Mexico, USA, which was later renamed as GenBank in 1982 and made available to the public. Simultaneously, in 1980, the EMBL Nucleotide Sequence Data Library (now EMBL-Bank, part of European Nucleotide Archive, ENA) was created as a centralized nucleotide repository at EMBL (European Molecular Biology Laboratory), Heidelberg, Germany. On the other hand, another database, the DNA Data Bank of Japan (DDBJ), was created in 1986 at the National Institute of Genetics at Mishima, Japan. Later, with the establishment of the International Nucleotide Sequence Database Collaboration (INSDC), these three primary international sequence databases, GenBank, EMBL-Bank, and the DDBJ – were linked to share data submissions and retrieval tools and to provide unrestricted access to publicly available data to everyone. By the end of 1981, the LANL database contained 263 sequences, including 50 sequences from eukaryotic viruses and 12 bacteriophage sequences, which, after the wide availability of automated Sanger sequencers, increased by doubling every 18 months. Following the commercialization of NGS platforms and the commencement of large-scale genome and metagenome projects, the volume of sequence data generation increased as never before, which required specialized virus databases for storage and retrieval.

The Viral Genomes Resource hosted by the NCBI is one specialized virus reference sequence database, where annotated viral genome sequences cataloged at species level (RefSeq and genome neighbors) are available, which is very helpful in viral sequence identification and taxonomy. Similarly, the ViralZone database, hosted by the Swiss Institute of Bioinformatics, is another useful reference resource for viral sequence analyses. These primary and reference sequence databases have enabled rapid archival, assembly, annotation, taxonomy, discovery, and study of viral evolution, helping enormously at times of crisis. It is estimated that since 2000, there was almost a nine-fold increase in viral sequence data submitted to the primary databases, as well as for novel viruses, as seen by the proportionate increase in the RefSeq database. Considering the clinical significance of certain viruses, seven different modules have been created for influenza virus, Dengue virus, West Nile virus, Ebolavirus, MERS coronavirus, Rotavirus A, and Zika virus at the NCBI Viral Genomes Resource. These specialized modules have additionally improved analyses of new sequences related to these viruses. It is worth mentioning that the first virus genome of the SARS-CoV2 was shared on January 10th, 2020 via the GISAID (Global initiative on sharing all influenza data) platform, which is an international database for sharing information from all influenza viruses. With the COVID-19 pandemic and exponential increase in submission of virus sequences, a separate EpiCov™ database was created to accommodate these sequences. This database is freely accessible and features a number of analysis options including highly customised sequence search, sequence alignment phylogenetic analysis and clade identification, coordinates for diagnostic primers/probes, 3D structures of various virus proteins and drug targets etc., which has helped the scientists enormously in designing improved diagnostic approaches for tracing and tracking the virus, studying evolution and dynamics and also in identifying potential management strategies. Nevertheless, since January 10th, 2020 a total of 2,39,600 SARS-CoV2 complete genome sequences have been submitted by December 4th 2020, nearly all of which were obtained by NGS. This gives an excellent idea of how modern sequencing technologies coupled with highly advanced bioinformatics and a curated database can help in management of a pandemic.

Most of the time, homology-based search approaches (nucleotide BLAST) are useful in determining the identity of the query sequence by searching highly similar sequences deposited in the non-redundant (nr) database (megablast). However, sequences, especially arising from metagenomic studies, have very little or no similarity to sequences belonging to known viruses. In such cases, the search can be performed by low stringency searches – discontinuous megablast (searches more dissimilar sequences) and BLASTn (searches somewhat similar sequences). If the searches still do not return any similar sequences, a blastx search (searches protein databases using a translated nucleotide sequence query) can be done to identify sequences based on protein conserved domains that the query-sequence might encode. Apart from these, a proportion of sequences from metagenomic datasets may not be classified through any of the above-mentioned search strategies. These unclassified sequences are supposed to originate from truly unknown viruses or other “biological dark matter”, and detection of such coding sequences with no homologs, also termed as “ORFans”, may help in the discovery of novel viral species. The unknown protein sequences encoded in these sequences can be used for isolation and purification of the viruses from original samples, followed by sequencing of their complete genomes.

During the last four decades, the field of molecular biology has grown vastly due to a number of discoveries and huge technological leaps. The discovery of nucleic acid sequencing and amplification technologies (primarily PCR) and the subsequent

innovations in the sequencing chemistries and incorporation of various signal detection technologies has now enabled us to completely sequence complex genomes from higher organisms within days. For smaller genomes, especially from most of the pathogenic viruses, entire genome of an unknown virus can be sequenced and analyzed within a few hours. This can keep us a step ahead during viral outbreaks, where early detection, identification, and monitoring of the causative agent makes a great difference. Nowadays, third-generation sequencers that can be easily taken to the field for on-site investigations are available. Nevertheless, in addition to its increasing applications in virus discovery and viral genome sequencing, currently available parallel sequencing technologies are being widely used in monitoring of antiviral drug resistance, viral evolution, and quasispecies diversity. However, the application of sequencing for viruses has certain major challenges, which include low abundance of the target virus nucleic acids in a complex mix of non-target nucleic acids. Therefore, additional steps for enrichment of viral sequences are often needed.

At present, three different generations of sequencing technologies are available; the basic technology, features, and specific applications of which have been discussed. Although, as with any other technology, these sequencing technologies have certain advantages as well as certain limitations. Therefore, the choice of sequencing platform is an important prerequisite for the success of a virus-related project and must be done carefully considering the specific objectives of the project.

Further Reading

- Barrientos-Somarribas, M., Messina, D.N., Pou, C., *et al.*, 2018. Discovering viral genomes in human metagenomic data by predicting unknown protein families. *Scientific Reports* 8, 28.
- Barzon, L., Lavezzo, E., Militello, V., Toppo, S., Palù, G., 2011. Applications of next-generation sequencing technologies to diagnostic virology. *International Journal of Molecular Sciences* 12, 7861–7884.
- Chiu, C.Y., 2013. Viral pathogen discovery. *Current Opinion in Microbiology* 16, 468–478.
- Datta, S., Budhailiya, R., Das, B., *et al.*, 2015. Next-generation sequencing in clinical virology: Discovery of new viruses. *World Journal of Virology* 4, 265–276.
- Delwart, E.L., 2007. Viral metagenomics. *Reviews in Medical Virology* 17, 115–131.
- Deurenberg, R.H., Bathoorn, E., Chlebowicz, M.A., *et al.*, 2017. Application of next generation sequencing in clinical microbiology and infection prevention. *Journal of Biotechnology* 243, 16–24.
- Edwards, R., Rohwer, F., 2005. Viral metagenomics. *Nature Review in Microbiology* 3, 504–510.
- Hall, R.J., Wang, J., Todd, A.K., *et al.*, 2014. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of Virological Methods* 195, 194–204.
- Hatcher, E.L., Zhdanov, S.A., Bao, Y., *et al.*, 2017. Virus variation resource – Improved response to emergent viral outbreaks. *Nucleic Acids Research* 45 (D1), D482–D490.
- Levy, S.E., Boone, B.E., 2019. Next-generation sequencing strategies. *Cold Spring Harbor Perspectives in Medicine* 9.
- Noolij, S., Schmitz, D., Vennema, H., Kroneman, A., Koopmans, M.P.G., 2018. Overview of virus metagenomic classification methods and their biological applications. *Frontiers in Microbiology* 9, 749.
- Reuter, J.A., Spacek, D., Snyder, M.P., 2015. High-throughput sequencing technologies. *Molecular Cell* 58, 586–597.
- Shendure, J.A., Porreca, G.J., Church, G.M., *et al.*, 2011. *Current Protocols in Molecular Biology: Overview of DNA Sequencing Strategies*. Hoboken: Wiley Press.
- Zheng, T., Li, J., Ni, Y., *et al.*, 2019. Mining, analyzing, and integrating viral signals from metagenomic data. *Microbiome* 7, 42.

Relevant Websites

- <https://nanoporetech.com/how-it-works>
How it works.
Oxford Nanopore Technologies.
- <https://www.thermofisher.com/blog/behindthebench/>
How Does Sanger Sequencing Work.
- <https://www.youtube.com/channel/UCvJHVo5xGSKejBbJ0A5AyQ>
NCBI YouTube Channel.
- https://www.youtube.com/playlist?list=PLTt9kKfE_0Gem8hlcJEn7YcesuuKdt_n
Next Generation Sequencing (NGS) - YouTube.
- <https://sapac.illumina.com/science/technology/next-generation-sequencing/beginners.html>
Next-Generation Sequencing Basics.
- <https://www.youtube.com/watch?v=WdTX1yykLks>
Pyrosequencing.
The Basic Principle and Steps ... - YouTube.
- <https://www.pacb.com/videos/video-introduction-to-smrt-sequencing/>
Sequencing 101: Video Introduction to PacBio Sequencing and the Sequel II System.
- <https://www.youtube.com/watch?v=nIvyF8bFDwM>
SOLiD DNA Sequencing - YouTube.
- <https://viralzone.expasy.org/>
ViralZone root.
- <https://www.ncbi.nlm.nih.gov/genome/viruses/>
Viral Genomes - NCBI - NIH.

Validating Real-Time Polymerase Chain Reaction (PCR) Assays

Melvyn Smith, Viapath Analytics, Specialist Virology Centre, King's College NHS Foundation Trust, London, United Kingdom

© 2021 Published by Elsevier Ltd.

Abbreviations

CLIA United States clinical laboratory improvement amendments. This is a regulatory body or program to ensure quality laboratory testing.

Ct Crossing threshold, the cycle number at which the PCR enters the exponential phase.

FDA United States food and drug administration. The FDA is a Federal agency responsible for protecting the public health through the control and supervision of medical devices, the FDA are one of the agencies responsible for developing and implementing CLIA.

IVD In-vitro diagnostic products. These are the reagents or instruments used to diagnose a disease, in this case the PCR assay.

LDT Laboratory developed test.

LOD Limit of detection, the lowest amount of analyte detected in an assay.

rt PCR Real-time polymerase chain reaction.

RT-PCR Reverse transcription PCR.

ΔRn The maximum difference in fluorescence between the start and end of the PCR.

Introduction

For many years, the development of assays took place in the laboratories where the test was required, the so-called in-house or LDT. Although more and more commercially-developed tests have become available, novel assays continue to be developed in academic hospital laboratories. These laboratories are more able to respond quickly to new and re-emerging infections and crucially have the samples necessary to develop the test. This was the case with the new human coronavirus causing clusters of pneumonia epidemiologically linked to a seafood market in Wuhan, China in December 2019. Researchers from Shanghai, Wuhan, Beijing and Sydney sequenced a sample from a patient who had worked at the market; the sequence was deposited in the publicly available GenBank database on 10th January. Further sequencing showed the virus shared 85% identity with a SARS-like coronavirus found in bats. The first LDT was published on 23rd January 2020, followed a month later by the first of the commercial tests.

Although often more expensive than LDTs, commercial kits enable the rapid introduction of new tests, with the further advantage of being CE marked or FDA approved, providing some form of reassurance to the laboratory. However, CE marking is only a declaration of compliance with European legislative requirements; it does not necessarily guarantee the rigorous validation of the assay, or give any details of how and when a test should be used. Furthermore, commercial assays, by definition, have to be commercially attractive. Therefore, specialist, small-scale tests for rarely occurring infectious pathogens will not be cost-effective. Thus, there will continue to be a need for laboratory-developed assays.

The Need for Validation

In-House Assays

Over the years many laboratories have established methodologies for validating their assays. However, the literature continues to show a lack of detail in some critical areas, e.g., optimization of extraction techniques, methods used in primer and probe design, no evidence of amplicon sequencing to confirm specificity, imprecise estimates of sensitivity and specificity and assays that do not include internal or extraction controls. Such lack of detailed experimental information makes assessing the clinical utility of the assay difficult. Some of these problems lie with the scientific literatures' approach to publishing, where space limitations restrict the amount of detail permitted in a paper. These difficulties have led to a number of papers calling for an improvement in the standards of reporting diagnostic assays, including the STARD initiative (standards for reporting of diagnostic accuracy) and the MIQE guidelines (the minimum information for publication of quantitative real-time PCR experiments; [Bustin et al., 2009](#)). In addition to the fundamental requirements of good scientific practice, there are a number of regulatory bodies that require assays to be validated to certain standards, such as the FDA and CLIA requirements in the USA and the IVD Regulations (EU) 2017/7462017 in Europe. There is also an obligation on health institutions to be accredited according to the ISO 15189 standard. There has been an ongoing debate in Europe regarding CE marking of LDTs, similarly in the USA the FDA announced its intention to shift from a policy of enforcement discretion to exercising regulatory oversight at a future date on LDTs. However, the imposition of rigorous controls on the development of assays must not stifle innovation and the ability of front-line laboratories to respond quickly to new and emerging threats. Under these circumstances it makes sense for laboratories developing and publishing assays to ensure

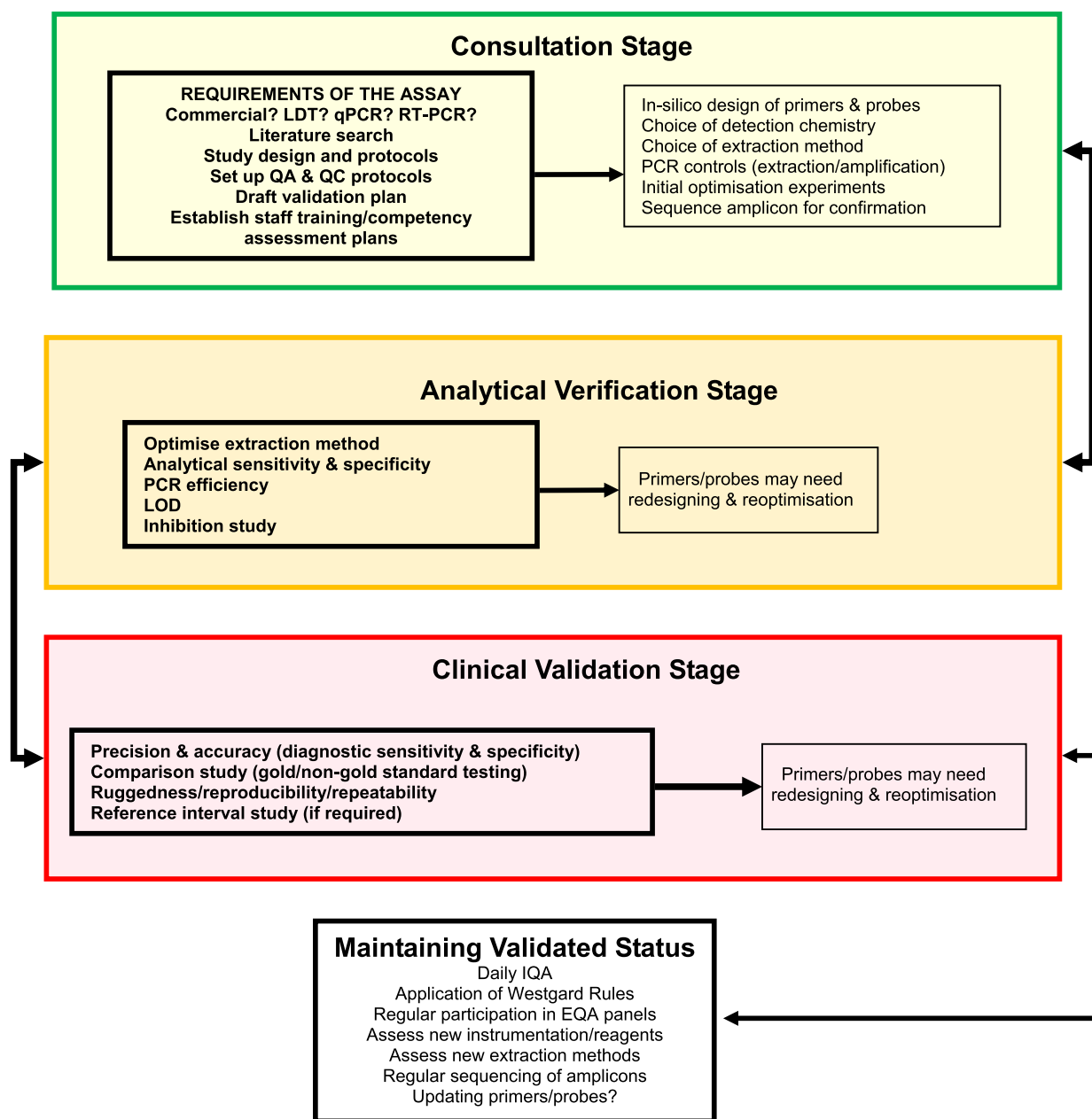


Fig. 1 Assay validation: Key concepts.

that they produce rigorously documented experimental proof to support any potential CE-marking/IVDD requirements that may be necessary in the future.

Commercial Assays

Although commercial assays are typically CE marked or FDA approved this does not necessarily mean the assay has been validated for all applications. Furthermore, although the assay may perform acceptably in the commercial developer's laboratory, a number of factors can affect the assay's performance elsewhere, e.g., staff competences and workflow systems, where not all laboratories may have separate working areas recommended to reduce contamination. There may also be differences in equipment maintenance schedules, including anything from freezers and pipettes to thermal cyclers, which can fundamentally affect the assay. Therefore, it is necessary for laboratories to verify the stated criteria. Although there are no formal requirements in the UK for laboratories to evaluate newly-introduced commercial assays, in the USA, FDA-approved, FDA-modified (i.e., tests modified from the manufacturer's

instructions but with performance characteristics determined by the user laboratory, consistent with CLIA requirements) and LDTs require verification before introducing the test into the clinical laboratory. CLIA stipulates that prior to implementation of an FDA-cleared test laboratories must verify the manufacturer's performance specifications. Where a FDA-modified test or LDT is to be introduced, CLIA stipulates that, additionally, analytical sensitivity and specificity must be established.

In an attempt to clarify the situation, it is the intention in this article to provide a set of basic working guidelines for the validation process. These guidelines can be applied throughout the continuous process of maintaining the validated status of an assay. The key concepts in the process of validating an assay are illustrated in [Fig. 1](#).

The Validation Process: Consultation Stage

The process begins with the development of a validation plan and involves decisions based on the clinical need for the assay, e.g., epidemiological studies, infection control or screening. As discussed previously, there are differences between the levels of validation laboratories perform when introducing commercial or LDTs. As a minimum, the laboratory introducing a commercial test must establish that the manufacturer's performance claims can be reproduced. Once the performance characteristics of an assay have been met, whether commercial or LDT, the validation exercise must continue on a daily basis (see [Fig. 1](#)). This involves continually monitoring the levels of internal and external positive controls to ensure the validation status of the assay is maintained.

Preliminary Considerations

The initial step is to define the purpose of the assay; all the subsequent steps in the validation process are guided by this decision. The following three variables that can affect an assay's performance must be considered at this stage:

- (1) The sample-type and the host/pathogen interactions that determine whether a qualitative or quantitative assay is required.
- (2) The assay system: the biological, technical and operator-related factors that affect the assay's ability to detect the target in the specific sample-type. All assays are considered multiplex, since they must include a co-amplified extraction control.
- (3) The result: will the result accurately predict the status of an individual or population in regard to the analyte detected?

The sample-type (e.g., tissue, whole blood, CSF) may contain inhibitors that affect the activity of the polymerase in the PCR. Therefore, the extraction process needs to be evaluated and necessary alternatives considered. Quantification assays are useful in the management of patients, where pathogen load can be monitored during therapy. Some viral pathogens, such as EBV and CMV can establish a primary low-level latent infection and subsequently become reactivated, switching to a productive lytic infection. Is more than one pathogen to be identified in a multiplex assay? Another significant factor is the availability of sufficient numbers of well-characterized positive control samples to enable the validation. In order to maintain objectivity, the method chosen to resolve discrepant results must be established as part of the validation plan before testing begins.

A quality assurance plan must be established for the assay. Consideration must also be given to the availability of external QA reagents. This can be a problem with new assays targeting rare pathogens, where QA panels are unlikely to be available; the laboratory may need to consider working with the providers to produce suitable reagents. Once the requirement for a test has been established, the next decision is whether to use a commercial assay or develop a LDT. In the USA the CLIA specify that laboratories using FDA-cleared assays must verify that the manufacturer's performance specifications for accuracy, precision, reportable range and reference intervals can be replicated. For modified commercial assays the following should also be tested:

- (1) Analytical Sensitivity (LOD).
- (2) Analytical Specificity to include inhibitory substances.
- (3) Any other performance characteristics required for test performance.

When assessing a commercial assay, the extraction process also needs to be verified. In most cases the manufacturer's protocol includes details on the extraction process to use. However, the assay may need to be validated with a number of different extraction methods, depending on the type of equipment available.

The Validation plan

Verification: the process of establishing whether the individual components of an assay meet the analytical performance requirements established at the start of the development process.

Validation: the process of ensuring that the completed assay conforms to the users' needs, requirements, and/or specifications under defined operating conditions.

There are numerous aspects of an assay that need to be continuously monitored throughout its use in the laboratory. Micro-organisms, particularly viruses, often mutate. This means that the efficiency of the PCR needs to be monitored for potential false-negative results, as this may be the first sign that the primers and/or probe need to be updated and revalidated. Manufacturers continually develop new

Table 1 Theoretical number of samples from subjects of known infection status required for establishing diagnostic specificity and sensitivity estimates using likely estimated specificity/sensitivity value and desired error margin and confidence

Estimated specificity/sensitivity value	2% error in estimated specificity/sensitivity			5% error in estimated specificity/sensitivity		
	Confidence			Confidence		
	90%	95%	99%	90%	95%	99%
90%	610	864	1493	98	138	239
92%	466	707	1221	75	113	195
94%	382	542	935	61	87	150
95%	372	456	788	60	73	126
96%	260	369	637	42	59	102
97%	197	279	483	32	45	77
98%	133	188	325	21	30	52
99%	67	95	164	11	15	26

buffers, enzymes and extraction kits, these must be assessed and if adopted, re-verified, and the assay revalidated. New types of probe chemistry are constantly being introduced and these, too, need testing to maintain or improve the assay's performance.

Although each new assay may present its own particular challenges, the basic criteria for validation, such as specificity, sensitivity and reproducibility apply and any specific differences can be incorporated without difficulty. The first steps concern decisions based on the clinical need for the test, i.e., the scope, purpose and application of the assay and whether a commercial or a LDT will be required. For all in-house developments it is strongly recommended that the comprehensive MIQE guidelines are followed.

Analytical Verification Stage

Reference Materials and Sample Numbers

The first question that arises when developing a LDT, particularly for rare and emerging infectious disease pathogens, is the availability of samples. If sufficient samples are not available, are they obtainable elsewhere? If not, it may be necessary to construct test samples by spiking various concentrations of the analyte into a suitable matrix. Other sources of suitable samples may be other clinical/research laboratories or commercial standards, quality control materials or proficiency panels. Typically, 100 samples of 50–80 positive and 20–50 negative specimens are used. Consideration also needs to be given to potential inhibitory substances likely to be found in the specimens tested. Therefore, paired control specimens should be prepared by adding low concentrations of the analyte, with and without the known inhibitors, to suitable negative samples. However, such artificially constructed samples are unlikely to have the same properties as clinical samples. Therefore, when sufficient numbers of genuine samples become available, the assay and extraction methods will need to be re-assessed. Where samples are available, more than one specimen type may be required, e.g., respiratory pathogens can be extracted from nasopharyngeal aspirates, bronchoalveolar lavages or nose and throat swabs. A literature search should always be carried out to determine the type of specimens to be assessed and the most efficient extraction method for each of the sample type determined.

When sufficient samples numbers are available, the numbers need to be calculated. Ideally this should be determined by statistical analyzes, where the sample size required to detect a significant difference is determined by the standard deviation from the difference of the means of paired samples used in the test under development and the gold standard comparator. Alternatively, they can be estimated from [Table 1](#) below. This method allows for either a 2% or a 5% calculated error in diagnostic sensitivity and specificity. So, for example, if a 2% error is assumed, the number of samples required for an assay with 99% confidence and 99% estimated sensitivity/specificity is 164. However, if the same assay achieved a 95% estimated sensitivity/specificity then the samples required to achieve a high confidence (99%) would be 788.

Template, Primers and Probes: *In-Silico* Design

A thorough literature review is first carried out to identify a suitable locus to be amplified. It may be necessary to consider using degenerate primers and possibly probes to cover sequence variants that occur in different strains of the pathogen. Thorough sequence searches with BLAST (Basic Local Alignment Search Tool; See "Relevant Websites section") are required to ensure as many variants as possible are included in the design. A local database can be constructed and aligned using one (or more) of the sequence alignment program available, such as CLUSTALW (See "Relevant Websites section"). Optimal primer and probe function is critical to successful PCR. There a number of software packages available to assist in primer and probe design, such as Primer Express (Applied Biosystems) and OLIGO (See "Relevant Websites section"). If the assay is a multiplex, the sets of primers and probes must be checked for cross-reactivity, to eliminate any inhibition due to co-amplification issues. Multiplexing is a useful method of syndromic testing and reducing the cost diagnoses by targeting more than one pathogen in a sample. Furthermore, most, if not all assays will be multiplexed, due to the co-amplified EC.

Any matches with other sequences, particularly from other pathogens likely to be encountered in the sample types under investigation will necessitate redesign. It must be stressed that this preliminary specificity check against the databases is not conclusive evidence. The assay itself must be tested against the range of pathogens likely to be encountered in the sample types under investigation.

Choice of the Quantification Standards

There are two main options for quantifying by a real time PCR, either absolute or relative. In most diagnostic assays absolute quantification is chosen. A set of separately amplified, log-diluted previously quantified templates are used to form a standard curve, derived from Ct values plotted against the log concentrations of each standard. The copy number of any sample is estimated from its Ct intercept on the standard curve. The accuracy of the standards can be confirmed by testing against external QA controls. The National Institute for Biological Standards and Control in the UK produce over 300 WHO International standards. Several standards designed for nucleic acid based assay testing are also commercially available, including HIV, hepatitis, CMV, EBV, and B19 parvovirus and are reportable in international units/milliliter (IU/ml).

Reaction Controls

The LDT must include positive controls (PC), negative, no template controls (NTC) and extraction controls (EC). The Ct values of the positive control must be recorded as part of the routine quality assurance. Typically, a single positive control for each specific pathogen, a NTC after every eight samples, and a further NTC as the last sample on the run is used to monitor the assay. Placing a NTC at the start of the set-up could give a false impression of any contamination problems. The PC can be produced from extracted clinical samples or from commercial sources and should be diluted to be reproducibly amplified at the lowest detectable level (typically a Ct of 30). An internal PCR or amplification control (IC) is an absolute requirement for any diagnostic assay. Nucleic acid from a suitable non-target pathogen can be spiked into each extracted specimen prior to the PCR and amplified and detected with a separate set of primers and probe. Preferably, an EC, using a non-related virus or bacterium is used to monitor extraction efficiency. Using this approach, the EC is spiked into the sample before extraction. This has the advantage over using naked nucleic acid as an EC, because it controls for all the steps in the extraction process. Phocine herpesvirus is a useful EC for assays detecting viral DNA pathogens and for RNA viruses, MS2 phage or mengovirus can be used. The use of an RNA template in RT-PCR controls for both the critical reverse transcription step and the PCR. Inhibition can be assessed by comparing the Ct values of the amplified EC in samples with the Ct value of the control extracted in a negative matrix, e.g., water or elution buffer. A difference, typically greater than three Cts (approximately equivalent to one log) indicates inhibition and a potentially false negative result. It is necessary to establish that the primers and probes and the amount of control added to the reaction does not interfere with the amplification of the pathogen target.

Reverse Transcription PCR

Reverse transcription PCR is used to detect RNA viruses, but the use of differential gene-expression assays is becoming more widespread to distinguish between infection and colonization in other infectious micro-organisms. Reverse transcription-PCR can be carried out as a two-step reaction, where the RT step is carried out separately and an aliquot of the reaction transferred to the PCR. More usually, in infectious disease assays, a one-step reaction is employed, where the reverse transcription and PCR occur in the same tube on the thermocycler. The one-step is quicker and more suited to high-throughput diagnostic laboratories, since the potential for contamination is lessened by number of steps that require opening tubes. The choice of priming the RT step needs to be considered; typically, the downstream, anti-sense primer is used. However, random hexamers may be more efficient. Often the choice is down to a sensitivity/efficiency balance between the hexamers ability to bind and hence copy all the RNA species in the sample and the sensitivity provided by the sequence-specific binding of the downstream primer.

Table 2 The Basic Westgard Rules

A	1_2 SD	If one control measure exceeds the mean $\pm 2 \text{ SD}$, the control values from the previous run should be considered to rule out a trend.
B	2_2 SD	This rule detects systematic error. The rule is violated when two consecutive control values exceed the same mean $\pm 2 \text{ SD}$ limit.
C	4_1 SD	This rule detects systematic error. The rule is violated when 4 consecutive control values exceed the same (mean $+1 \text{ SD}$ or mean -1 SD) limit. The run need not be rejected if this rule is violated but should trigger recalibration or equipment maintenance.
D	1_3 SD	This rule detects random error. Violation of this rule may also point to systematic error. The assay run is rejected if one control value exceeds the mean $\pm 3 \text{ SD}$.
E	$R_4 \text{ SD}$	This is a range rule and it detects random error. The rule is violated when the difference in the SD between 2 control values exceeds 4 SD (i.e., -2 SD and $+2 \text{ SD}$).
F	10_x	This rule detects systematic error. The rule is violated when the last 10 consecutive values are on the same side of the mean. Its violation often indicates the deterioration of assay reagents.

Westgard Rules

For each run the Ct values of controls are plotted on Levy-Jennings or Shewhart charts to monitor the performance of the assay; visual presentation of the control data is helpful in spotting variations in the operation of the assay (see [Table 2](#)). Westgard rules can also be applied to determine when a corrective action should be taken and give some indication of the nature of the analytical error. The use of Westgard rules and Shewhart charts must form part of the ongoing monitoring of the assay's validated status.

Experimental Optimization

Optimal primer and probe concentrations must be verified, together with the PCR program itself. A well characterized, stable positive control must be used throughout this verification process to provide an invariant baseline from which the effects of changes to the various components can be measured. The formation of primer dimers and the degree of non-specific amplification can be assessed by analytical agarose gel electrophoresis which also provides evidence of the size of the amplicon. To optimize primer and probe concentrations, ideally two log-dilutions of the positive control are used in a checkerboard layout of varying primer/probe combinations. The optimal concentrations and thermal cycling conditions are those that provide the lowest Ct values, the greatest ΔR_n and a consistent C_t difference of approximately three cycles between the two log-diluted samples. Once the basic parameters of the assay have been established it will be necessary to confirm the amplicon product by sequencing and BLAST analysis. Due to the short amplicon lengths in real-time PCR, the fragment may need to be cloned into a suitable plasmid vector.

Normalization

Expression analyzes with reverse transcription PCRs is not normally used in diagnostic PCR. However, the range of diseases and their associated pathogens being diagnosed by RT-PCR is increasing. In some cases, e.g., with fungal species, there is a need to distinguish between carriage, environmental contamination and infection. Therefore, assays capable of detecting and quantifying differentially expressed genes indicative of the disease condition are needed. This requires carefully selected reference genes to control for variations in extraction, reverse-transcription and amplification efficiencies, so that comparisons across different mRNA concentrations can be made and fold changes in expression levels determined. A detailed literature search for suitable reference gene mRNA targets must be made and the selected candidates experimentally tested for stable expression in both diseased and non-diseased samples.

Analytical Specificity and Sensitivity

Analytical specificity: the LDT's ability to detect the target it was designed for and not cross-react with other analytes in the sample.

Specificity is demonstrated either by spiking samples with a range of different pathogens prior to extraction, or adding the extracted nucleic acid from these pathogens to the extracted sample under investigation. All reactions should be analyzed by agarose gel electrophoresis to ensure that amplification has not occurred in any of the samples expected to be negative. In some cases, primers may amplify non-target templates, which may not be detected by the probe. This will compromise the sensitivity of the assay and must be eliminated, either by re-evaluating the PCR conditions (typically raising the annealing temperature), or preferably by redesigning the primers.

Analytical sensitivity: The LDT's ability to detect very low levels of a given analyte in a biological specimen.

This is synonymous with the assay's LOD, i.e., the lowest concentration of the analyte consistently detected in $\geq 95\%$ of samples tested with acceptable precision. The LOD is usually determined for diagnostic assays by using a set of log-diluted controls, such as patient samples, a suitable cell line, or proficiency panels. It is important to include a NTC in the LOD study to ensure that the PCR does not generate a signal that could interfere with true low-level positive signals from a sample.

There is no hard and fast rule as to how many samples to use, The Clinical and Laboratory Standards Institute suggests a minimum of 60 data-points (12 separate measurements from each of 5 samples) are required from a manufacturer to establish the LOD. This seems a reasonable figure to adopt as a standard approach for both LDTs and commercially produced assays and will provide the necessary intra-assay variation measurements. This measurement should also be carried out on a second thermocycler to establish machine-to-machine variation.

Estimation of the LOD is usually carried out by probit analysis, a type of regression used to analyze binomial response variables, where a sigmoid dose-response curve is transformed into a straight line that can be analyzed by regression either through least squares or maximum likelihood. For PCR assays the lowest concentration of analyte that can be detected with a stated probability can be determined by plotting the data from positive replicate results versus the analyte concentration.

Inhibition Study

The factors that inhibit or prevent the amplification of nucleic acids by PCR can be present in the extracts from a number of sources. Inhibitors typically act by: (1) interference with the cell lysis necessary for extraction of nucleic acids, (2) interference by degrading the

nucleic acid or inhibiting its capture, (3) inhibiting polymerase activity during amplification of the target DNA, resulting in false-negative results or inaccurate quantification. Different sample types present different problems due to the range of endogenous inhibitors found, for example in fecal samples, complex polysaccharides, breakdown products from hemoglobin, and bile acids can all be present and careful thought needs to be given to the method of extraction. Other potential inhibitors include metabolites resulting from pathological conditions such as diabetes mellitus and homeostatic hepatitis, or from medications used in treatment.

Manufacturers continue to develop their PCR reagents specifically to overcome many of the common endogenous inhibitors found in clinical specimens and it is recommended that a number of different commercial amplification kits are evaluated for their performance in this respect. As discussed previously, inhibitors can be detected using an IC. If the IC fails to amplify or exhibits suppression below the acceptable threshold, the amplification of the intended target sequence may also be inhibited. To avoid reducing the sensitivity of the assays, the IC should be used at the lowest concentrations that can be reproducibly amplified to minimize any competition between its amplification and that of the target pathogen.

The IC must demonstrate inhibition by the substances expected to be found in the sample, for which an inhibition study needs to be carried out. The study must first identify what types of inhibitors are likely to be present in the sample types used in the assay. The IC can be tested in negative samples with and without the interfering substance(s) in parallel and the effect on the C_t values monitored. This test must be repeated with the full assay (including the pathogen detection primers and probes) using paired samples, with and without the target analyte to fully validate the assay. Once interference is found, of either or both the IC and target, samples can be serially diluted to test the limit at which the target will amplify. In most cases dilution of inhibited samples provides a straightforward method of enabling amplification. The assay must then be tested against clinical samples; this is because artificially constructed inhibition control samples may not accurately reflect the assays performance in routine use.

PCR Efficiency

The efficiency of the PCR can have a significant impact on the robustness and precision of the assay. The efficiency of the reaction is determined by a number of factors, including primer design, cycling conditions and the reagents used in the reaction mixture. PCR efficiency is particularly important in assays reporting fold changes of mRNA for target genes relative to those of reference genes, where both templates must be amplified with equal efficiencies. PCR amplification efficiency is typically established using calibration curves and it makes sense to use the dilution series (covering 5–6 orders of magnitude) from the LOD analyzes for this purpose. The equation of the linear regression line, together with Pearson's correlation coefficient (r) and the coefficient of determination (r^2) are used to determine amplification efficiency. Amplification efficiency itself is determined from the slope of the log-linear portion of the calibration curve and is given by Eq. (1):

$$E = 10^{(-1/\text{slope})}. \quad (1)$$

Ideally the amount of template will double during each round of exponential amplification, this translates to a reaction efficiency of 2, therefore, using equation 1: $2 = 10^{(-1/\text{slope})}$ which gives a slope for the standard curve of -3.32 . This ideal figure of 3.32 also represents the difference in cycle number for each log dilution in the series. The efficiency can also be expressed as a percentage of the template amplified in each cycle using Eq. (2):

$$\% \text{ Efficiency} = (E - 1) \times 100\% \quad (2)$$

In the ideal example given above; $\% \text{ Efficiency} = (2 - 1) \times 100\% = 100\%$.

The experimental measurement should be carried out in triplicate on at least two thermal cyclers and the information included in the validation documentation. In practice acceptable assays should achieve efficiencies of 90%–100%. Reaction efficiencies lower than this may be caused by suboptimal reaction conditions or poor primer design. Reaction efficiencies of more than 100% could be due to measurement errors e.g., in preparation of the dilution series, or co-amplification of primer-dimers. Improving specificity will improve the sensitivity and increase the dynamic range of the assay, by reducing competition between the specific and non-specific amplification products.

Linear Dynamic Range or Reportable Range

The linear dynamic range can be determined during the LOD analysis by extending the range of the dilution series. The dilution series is often set at the range one would expect the analyte to be found in clinical specimens, although this may not always be known, especially for a novel pathogen and so initial testing can be carried out from 1 up to 10^7 or higher copies/ml of target. If using a generic assay to detect a range of pathogens, e.g., a pan-fungal PCR, each of the species the assay is designed to detect must be tested. Similarly, for multiplex assays, the LOD must be determined individually for each target analyte in the assay and must include the individual targets in the multiplex at high and low concentrations to ensure competitive amplification does not prevent all targets being identified. The reportable range is defined as the lowest and highest results (in suitable units of concentration) reliably detected in the assay. The linearity of the range can be established by calibration curves, where an r^2 value (coefficient of determination) of 1.00 indicates a perfect fit of the data points. Generally, in qPCR assays an r^2 of not less than 0.99 is considered acceptable. The linearity and reportable range should be carried out on at least 5 log-dilutions of the target nucleic

acid extracted from an appropriate sample type (serum, urine, NPA etc.) in triplicate, ideally on two different thermocyclers. Quantitative results from the test can only be reported when they fall within the linear range of the assay. A reportable range is not applicable to qualitative assays, since the result is simply positive or negative (below the LOD). However, a log-diluted panel should be tested, because the information derived is useful in assessing both the efficiency and clinical utility of the test and the results (as C_t values) reported as part of the supporting documentation for the assay.

Clinical Validation Stage

Precision, Accuracy and Trueness

Precision: a measure of the closeness of agreement between independent test results obtained under defined conditions.

Accuracy: the level of agreement between the true value of an analyte and the value obtained by the new test.

Trueness: defines the level of agreement between the average value obtained from a large series, the test and the accepted reference value.

Precision experiments are designed to measure the random error of an assay over a pre-determined period of time by multiple measurements of an aliquot derived from a homogeneous sample. For qualitative assays random error can be established by testing the PC and NC material in triplicate over a period of 10 days on two different thermocyclers. For qPCR assays estimates can be determined from the quantification standards and two positive controls at the lower and upper ends of the assay's reportable range.

Accuracy refers to the closeness in agreement between a single measurement and the true value of the analyte under investigation. Trueness is determined by analyzing the average value obtained from a series of measurements with the new assay and the true value of the analyte (if an international standard is available) or a reference method (if a standard is not available). There are two approaches to evaluating trueness (Burd, 2010):

- (1) A comparison of methods study, where split samples are tested in parallel with an appropriate gold standard method.
- (2) A recovery study, where proficiency samples, or other verified sample types are compared with the assay under development.

Although either of these methods is acceptable, a split sample, comparative study using clinical samples is preferred. However, in some cases, e.g., when developing an assay for a novel pathogen, a suitable gold standard assay may not be available and a recovery study will have to suffice.

The Gold Standard Comparative Study

The comparative study is the cornerstone of the validation exercise. Estimates of sensitivity and specificity are derived from comparisons between the LDT and an established gold standard, which, ideally, is the same type of test and has an assumed sensitivity and specificity of 100%. By definition the gold standard assay is an error-free diagnostic method, which, rarely (if ever) exists. Consequently, an earlier PCR assay, or a culture-based method is often used.

Samples for the comparative study should be representative of the disease and population being investigated and also suitably distributed across both age range and gender. Where only limited samples are available, it may be necessary to use archived positive and negative specimens, or cultured material spiked into negative clinical samples. Other sources may be commercial standards, quality control materials or proficiency panels. Sample numbers can be determined as described previously. However, the sample numbers are more often selected for more pragmatic reasons, such as cost and feasibility and therefore, less than the statistically optimal number will be tested. This is often the case for a new or re-emerging infectious disease.

The difficulty comes in interpreting the results of a study where the comparative test is not a "perfect" gold standard, (i.e., the alloyed standard). This is particularly so if the comparative method is cell culture, which is thought to be 100% specific, but provides less than optimal sensitivity. The strength of the PCR lies in its sensitivity (theoretically a single DNA copy). Thus a newly developed PCR assay may produce many more positive results than the reference method, leading to a misleading, lower estimate of specificity. The researcher is then left with the problem of deciding whether these are true or false positives. A thorough and detailed assessment of the sample derivation, clinical history and sequencing can help to determine the true status of any sample, together with discrepant analysis. In discrepant analysis the discordant results are resolved by a third test, such as another PCR directed to a different gene-target and the results from this resolver test are used to definitively assign the final results. It is important to include a number of randomly selected samples with concordant results for discrepant analysis so as to reduce bias.

Further statistical analysis can be carried out to improve confidence in the results from the new LDT, particularly when the true disease status of the samples is difficult to evaluate. The agreement between the new test and comparator test can be expressed as the

New Test	Gold Standard		
	Positive	Negative	Totals
Positive	TP	FP	TP + FP
Negative	FN	TN	FN + TN
Totals	TP + FN	FP + TN	n

Fig. 2 Fig. 2 2 × 2 Contingency Table.

kappa value. The kappa statistic is a generic term for a number of similar measures of agreement applied to categorical data and is a measure of the proportion of agreement between results beyond chance. It is used in assessing the degree to which two or more raters, (i.e., tests) examining the same data, (i.e., specimens) agree in assigning the data to categories (positive or negative results). A kappa value of 1.00 indicates perfect agreement; a value of 0.00 indicates no agreement above that expected by chance, and a kappa value of -1.00 indicates complete disagreement. Generally, kappa statistics above 0.80 are considered almost perfect. The kappa value is calculated from the results in the standard 2×2 contingency table, used to record the results from the comparative test (see Fig. 2).

$$OP = \frac{TP + TN}{n}$$

$$EP = \left[\frac{TP + FP}{n} \right] \times \left[\frac{TP + FN}{n} \right] + \left[\frac{FN + TN}{n} \right] \times \left[\frac{FP + TN}{n} \right]$$

$$K = \frac{OP - EP}{1 - EP}$$

where:

K = Kappa value.

OP = observed proportion of agreement.

EP = expected proportion of agreement.

TP = true positive.

FP = false positive.

TN = true negative.

FN = false negative.

n = total number of samples.

Diagnostic sensitivity and specificity (DSe and DSp) are also calculated from the results in the contingency table.

$$DSe = \frac{TP}{TP + FN}$$

$$DSp = \frac{TN}{FP + TN}$$

Two further important probabilities can be calculated from the data derived from the trueness study, the positive and negative predictive values (PPV and NPV). The PPV is a measure of a positive result from the test to truly predict the presence of disease/infection, while the NPV is the probability that a negative result accurately indicate a non-diseased/uninfected status.

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{FN + TN}$$

The Study Without a Gold Standard

In many cases, particularly for novel pathogens, a comparative method will not be available. Under these circumstances the gold standard may be a diagnosis determined by accepted clinical methods. It may also be necessary to incorporate results from more than one method to provide the comparative results. Whatever the approach chosen, estimating the performance characteristics of a new assay without a true comparative test is a challenging task. Typically, agreement is measured as the overall percentage or fraction of samples that have the same result (i.e., both either positive or negative). The difficulty with these simple agreement measures is that they do not take agreement by chance into account. A number of models have been proposed, based on latent class analysis, log-linear modeling and other techniques. Latent class analysis involves multiple imperfect tests that are used to construct a gold standard. Such models are based on the concept that the observed results of different, imperfect, tests for the same disease are influenced by a common but unobserved (latent) variable i.e., the true disease status.

Inter-Laboratory Testing

The ultimate evidence of an assay's fitness for purpose is its successful integration into other laboratories. The LDT reagents and samples used for the comparative study should be sent to a minimum of three laboratories willing to participate in testing. This will also provide additional data on the assay's reproducibility and robustness.

Maintaining the Validated Status and Re-Validation

The validated assay must be monitored consistently for repeatability through the performance of the run controls to evaluate any potential changes in the assay's precision and accuracy. All reagents should be assigned batch numbers and new batches tested with the controls and a suitable number of positive and negative samples from a previous run to ensure consistent efficiencies between batches. If more than one thermal cycler is used, each machine should be referenced and each test carried out on it identified. The results of the PCR controls and quantification standards must be plotted daily on Shewhart charts and assessed by Westgard rules to ensure that reaction efficiencies are maintained.

Participation in regular IQA and EQA schemes is fundamental to maintaining an assay's validated status. The purpose of EQA testing is not solely to monitor the performance of the assay, but assesses all aspects of the test procedure, including the extraction method and reporting accuracy. If an external panel is not available, alternatives can include blind sample testing, sample exchange with other laboratories, or clinical note reviews.

Technical modifications continually arise as manufacturers develop and improve their reagents for extracting nucleic acids, the PCR and instrumentation. Assessment of new probe chemistry or the transfer of an assay to full or semi-automated instrumentation, normally only requires a methods comparison study. In this way any changes to the diagnostic sensitivity and specificity can be assessed quickly and accurately and the updated assay introduced with the minimum delay.

Over time it is likely that an assay will require some degree of revalidation, either for purely technical reasons, or because of changes in the nature of the analyte detected. Regular monitoring of amplification efficiency together with clinical information will give an early warning of any changes in the sequence of the amplicon detected in an assay. Point mutations are common in the genomes of many infectious microorganisms, particularly RNA viruses. New viral lineages may also be introduced into the population being tested due to travel from abroad. Under such circumstances, significant modifications to the primers and probe may be required, necessitating reverification and revalidation of the assay. Amplicon sequencing should be part of the validation plan and an integral part of routine trouble-shooting algorithms, since it is capable of resolving errors due to non-specific binding of primers and probes.

Acknowledgment

This article, including tables and figure, is based on author's article "The Validation of Real-time PCR Assays for Infectious Diseases", published in *Real-Time PCR: Advanced Technologies and Applications* (2013), edited by: Nick A. Saunders and Martin A. Lee, Health Protection Agency, Colindale, UK and Porton Consulting Research Ltd, Salisbury, UK (respectively), Caister Academic Press.

References

- Burd, E.M., 2010. Validation of laboratory-developed molecular assays for infectious diseases. *Clinical Microbiology Reviews* 32, 550–576.
- Bustin, S.A., Benes, V., Garson, J.A., *et al.*, 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* 55 (4), 611–622.
- Corman, V.M., Landt, O., Kaiser, M., *et al.*, 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25 (3), 23–30.
- Hadgu, A., Qu, Y., 1998. A biomedical application of latent class models with random effects. *Journal of the Royal Statistical Society Series C* 47 (4), 603–616.
- Hadgu, A., Dendukuri, N., Hilden, J., 2005. Evaluation of nucleic acid amplification tests in the absence of a perfect gold-standard test – A review of the statistical and epidemiologic issues. *Epidemiology* 16, 604–612.
- Rådström, P., Knutsson, R., Wolffs, P., Lövenklev, M., Löfström, C., 2004. Pre-PCR processing: Strategies to generate PCR-compatible samples. *Molecular Biotechnology* 26, 133–146.
- Raymaekers, M., Smets, R., Maes, B., Cartuyvels, R., 2009. Checklist for optimization and validation of real-time PCR assays. *Journal of Clinical Laboratory Analysis* 23, 145–151.
- Rutjes, A.W.S., Reitsma, J.B., Coomarasamy, A., Khan, K.S., Bossuyt, P.M.M., 2007. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technology Assessment* 11, 1–69.
- Saunders, N.A., Martin, A.L. (Eds.), 2013. *Real-Time PCR: Advanced Technologies and Applications*. UK: Caister Academic Press.
- Sloan, L.M., 2007. Real-time PCR in clinical microbiology: Verification, validation, and contamination control. *Clinical Microbiology Newsletter* 29, 87–95.
- Westgard, J.O., Barry, P.L., Hunt, M.R., 1981. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clinical Chemistry* 3, 493–501.
- Wilson, I.G., 1997. Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology* 63, 3741–3751.

Relevant Websites

- <https://www.ebi.ac.uk/Tools/msa/clustalo/>
Clustal Omega.
- <http://www.ncbi.nlm.nih.gov>
National Center for Biotechnology Information.
- <http://www.oligo.net/>
oligo.net.

Rapid Point-of-Care Assays

Jan G Lisby and Uffe V Schenider, Copenhagen University Hospital Hvidovre, Hvidovre, Denmark

© 2021 Elsevier Ltd. All rights reserved.

Introduction

The first report of viruses as pathogenic agents was published in 1892, when the transmission of the tobacco mosaic disease by liquid filtered through porcelain filters capable of retaining bacteria was reported (Iwanowski, 1892). Subsequently, Loeffler and Frosch using similar techniques, reported the transmission of the causative agent of foot-and-mouth disease in 1898. The first human viral pathogen was reported in 1902, when it was reported that a filterable agent was the causative agent of yellow fever (Reed, 1902). Virology as a clinical diagnostic enterprise emerged with the development by Enders *et al.* in 1949 of the monolayer cell culture technology, allowing for growth of viruses outside the host or laboratory animals (Weller *et al.*, 1949). For the subsequent decades, clinical diagnostic virology applied viral propagation in cell cultures, such as plaque assays, first described in 1952 (Dulbecco, 1952) and in the 1960s, detection of viral antigens or human antibodies raised against viral antigens was established. The first solid-phase assays, termed Enzyme Immunoassays (EIA) were made possible by the discovery in 1966 that antibodies (initially polyclonal) could be labeled with enzymes providing a detectable staining result (Nakane and Pierce Jr, 1966). The substitution of enzymes with a radioactive label in 1967 further enhanced the signal (Catt and Tregar, 1967). The description of the polymerase chain reaction (PCR) in 1985 (Saiki *et al.*, 1985) and the subsequent upgrade of the technology by application of a heat-stable DNA polymerase (Saiki *et al.*, 1988) resulted in a paradigm shift in clinical virology diagnostics and within the next decade, PCR almost completely replaced culture and serology methods as the preferred diagnostic intervention for acute viral infections (Mackay *et al.*, 2002; Niesters, 2002).

The primary purpose of human clinical viral diagnostics is to prove or disprove a definite viral diagnosis in the case of a clinical infection. The secondary purpose of the diagnostic efforts is to impact on clinical patient management, which may be in the form of cessation of unnecessary antibacterial treatment (if no other indication of a bacterial infection is present), initiation of specific antiviral treatment (if possible/available) and decision on contact isolation if the patient is admitted to a hospital or other infection control measures. Besides operational knowledge of specific immune and vaccination status, impact on clinical patient management requires availability of diagnostic results within a time window where impact is possible and patient outcome may be altered, which most often will be limited to the first few hours or days after patient symptom debut. For the first several decades of clinical diagnostic virology, this was problematic. Diagnostic methods based upon serology may deliver rapid results, especially methods based upon lateral flow technology (Koczula and Galotta, 2016), but while detection of a matured immune response to a previous infection with a viral pathogen – or a previous vaccination – by detection of IgG against the viral target is sensitive and specific, especially when matured IgG antibodies are detected (Prince and Lapé-Nixon, 2014), rapid detection of viral antigens in an acute infection by lateral flow is specific but insensitive (Chartrand *et al.*, 2015) and detection of an acute viral infection by detection of IgM is sensitive but unspecific (Prince and Lapé-Nixon, 2014). With the emergence of the PCR technology, sensitive and specific detection of an acute viral infection became possible. Initially in the 1990s, diagnostic PCR in general was time consuming and cumbersome (Lisby, 1998; Smith *et al.*, 1992), but eventually, fully automated PCR platforms were developed in the mid to late 2010s for the clinical diagnostic laboratory, bringing the turn-around-time of the diagnostic process in the laboratory down to a few hours (Wirten *et al.*, 2017; Cobb *et al.*, 2017). However, for most patients, a considerable time must be added for logistic delays, such as patient sample transportation to the laboratory, receipt and registration within the laboratory, reporting of test results and clinical action upon the test result. Often, the total turn-around-time from patient sample to clinical action upon the diagnostic result is more than 6–8 h. For acute viral infections, such as viral meningitis/encephalitis, viral respiratory infections and viral gastrointestinal infections, a diagnostic delay of 6–8 h or more may compromise the diagnostic impact on e.g., differential diagnostics (e.g., Influenza vs meningitis in an adult with high fever and muscle/head ache), antimicrobial use (at least the first 1–2 doses) and use of contact isolation facilities.

PoC Technologies

The choice of technology for detection of viruses depends greatly upon the specific pathogen. In general, urine or throat swabs are well-suited specimens for virus detection, since virus often can be detected for a long time and specimens are easy to obtain. But, depending on the specific virus, blood, fecal or other specimens may be more adequate sampling sites (Paz-Bailey *et al.*, 2018). Likewise, greater sampling volume and the processing of a greater fraction of the sample volume increases the sensitivity of detection.

Some viruses such as the genus of flaviviruses can only be isolated from the blood stream in humans for a short time after infection. The preferred tests for such viruses such as yellow fever, dengue-, zika- or west nile virus are therefore lateral flow tests that are testing for virus specific antigens in combination with IgM and IgG human antibodies. The direct detection of virus is hence combined with the detection of host response by seroconversion.

Other viruses, such as viruses infecting the respiratory tract or the gastro-intestinal channel of humans are well-detected by PCR based assays. PCR based amplification is especially beneficial when detection of multiple viruses (multiplex assays) is requested

(e.g., respiratory tract infections) or viruses with high mutations rates such as RNA viruses are being detected. Detection by isothermal amplification have been shown to work well with virus detection. Still, PCR is often preferred over isothermal amplification, as additional conserved nucleic acid regions are requested for design of isothermal amplifications assays compared to PCR based assays.

To detect multiple viruses, current Point-of-care multiplex platforms are utilizing different strategies. In the Biofire Filmarray from BioMérieux, nested PCR amplification by a multiplex first round PCR is combined with second round nested monoplex pathogen specific PCR on microarray with fluorescence detection (Poritz *et al.*, 2011). The ePlex platform from Genmark Dx is also based on microarray detection but with electrochemical detection in combination with multiplex PCR (Babady *et al.*, 2018). Another strategy is applied by the QIAstat Dx platform from Qiagen in which the extracted nucleic acids are fractioned into several separate PCR chambers. Each chamber allows for a five-plex multiplex PCR with an internal control and the detection of fluorescence signal from hydrolysis probes (Parcina *et al.*, 2020). A similar approach was applied by the Idylla Platform from Biocartis which initially was launched with a multiplex respiratory panel for detection of viral pathogens (Wouters *et al.*, 2019).

Most platforms that only detect a few pathogens simultaneously are based on the hydrolysis probe PCR technology, e.g., the Cobas Liat from Roche (Tanriverdi *et al.*, 2010), the GeneXpert from Cepheid (Jenny *et al.*, 2010), the Aries instrument from Luminex Corp (Voermans *et al.*, 2016), the BD max from BD (Cárdenas *et al.*, 2014), the Accula system from Mesa Biotech and the Simplexa from Focus Diagnostics (Woodberry *et al.*, 2013). In contrast to these platforms, the IDnow instrument from Abbott is based on the isothermal nicking enzyme amplification reaction (NEAR) technology, which allows for faster amplification, but at the risk of non-specific amplification products (Nie *et al.*, 2014; Tan *et al.*, 2008). As the landscape of instruments in the PoC field increases, other technologies and combinations hereof are expected to be introduced. Hopefully, the choice of technology by the end-user will reflect the strategy and the implementation hereof, as envisioned by the manufacturer.

PoC Strategy and Implementation

In general, any assay selected for use at a PoC setting should bring value to clinical patient management that cannot be obtained by testing the patient sample for the same microbial targets in the central clinical microbiology laboratory. The patient management value of any test will depend upon the actual patient population and the actual clinical setting, e.g., in an adult Emergency Room, Influenza A/B may be the only respiratory assay providing enough patient management value to be able to show a satisfactory benefit/cost ratio, but in a pediatric ED, a highly multiplexed respiratory panel may be preferable. Ideally, for each Point-of-Care setting (Fig. 1), for each diagnostic question (e.g., contact isolation yes/no, upper respiratory tract infection yes/no) and for each patient population (e.g., adult/pediatric), a specific business case should be established, detailing the impact on clinical patient management obtained by PoC testing as well as the cost of PoC testing.

PoC infectious disease testing performed outside the geography of the central clinical microbiology laboratory may in the near future be performed by individuals with different educations and skillsets, from highly trained biotechnicians to patients themselves. Depending upon the actual geographical location (Fig. 1), PoC testing may be classified as:

Point-of-Care Laboratory Grade 1A

In this setting, PoC for infectious disease testing is performed “in hospital – in lab” in a 24/7 laboratory staffed with dedicated and trained biotechnicians. This may be considered as a satellite clinical microbiology laboratory and may be operated by trained clinical microbiology technicians but may also be operated by trained biotechnicians from other clinical specialties, such as clinical biochemistry or clinical pathology. If available 24/7, any assay performed within the geography of the clinical microbiology laboratory is considered as PoC Grade 1A.

Point-of-Care Laboratory Grade 1B

In this setting, PoC for infectious disease testing is performed “in hospital – outside lab” by local healthcare professionals in e.g., the Emergency Department (ED), the Intensive Care Unit (ICU) or the Pediatrics Department. As the healthcare professionals performing the assays in this setting are not dedicated laboratory professionals, but most likely special trained nurses or nurse assistants, assays must be low complexity, easy to perform and fail-safe.

Point-of-Care Laboratory Grade 2A

In this setting, PoC for infectious disease testing is performed “outside hospital – in healthcare facility” by healthcare professionals in e.g., a General Practitioners office or laboratory. In this setting, the assays may be performed in the actual office or in a side room facility. The healthcare professionals performing the assays will most likely not be professional laboratory workers but may be physicians, nurses or other trained staff. Thus, assays performed in this setting should be low complexity, easy to perform and fail-safe.

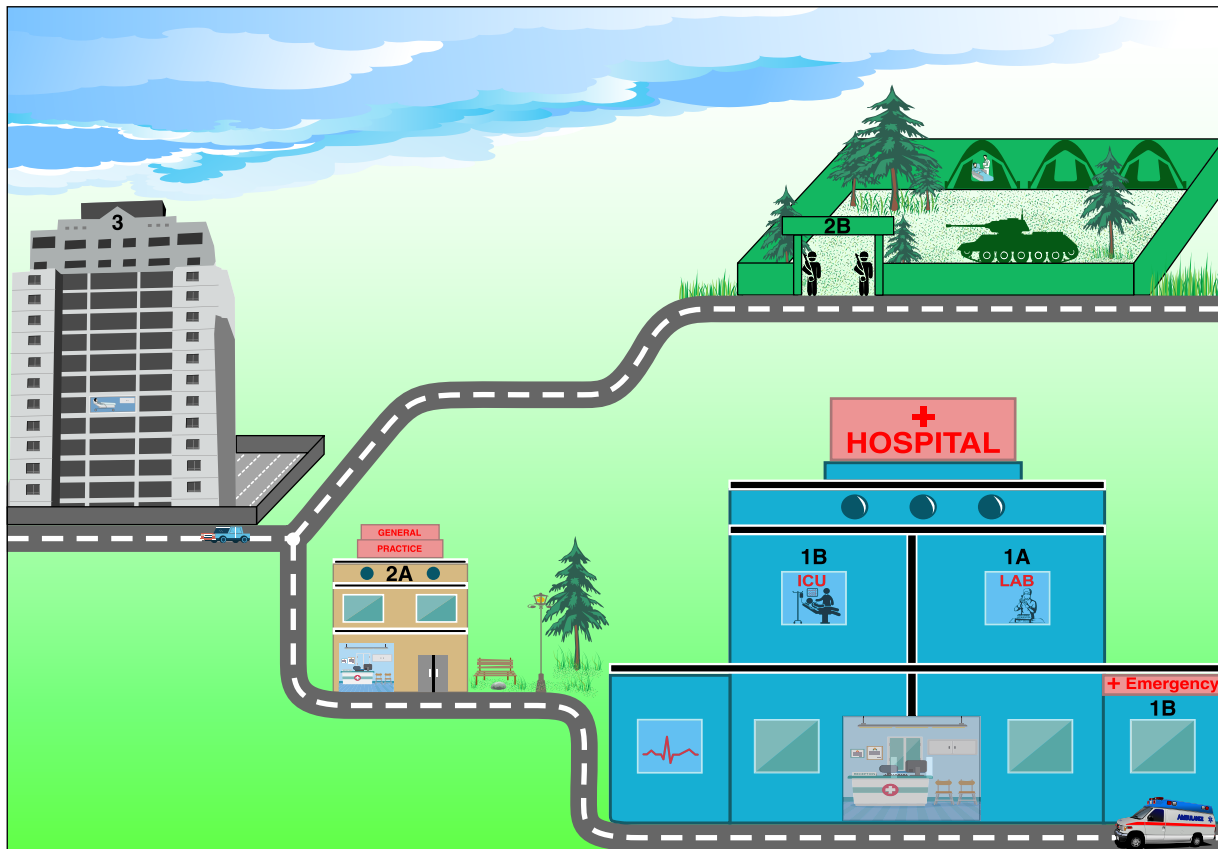


Fig. 1 Different Point-of-Care settings. Point-of-Care Laboratory Grade 1A – in hospital – in lab; Point-of-Care Laboratory Grade 1B – in hospital – outside lab; Point-of-Care Laboratory Grade 2A – outside hospital – in healthcare facility; Point-of-Care Laboratory Grade 2B - outside hospital – outside healthcare facility; Point-of-Care Laboratory Grade 3 – home setting.

Point-of-Care Laboratory Grade 2B

In this setting, PoC for infectious disease testing is performed “outside hospital – outside healthcare facility” by healthcare or other professionals in e.g., a third world field use or a military setting. Thus, assays performed in this setting should be low complexity, CLIA-waived or similar. The Clinical Laboratory Improvement Amendments of 1988 (CLIA) provides federal standards applicable to all US facilities performing diagnostic testing on human specimens. Waived tests include tests that have been cleared by The United States Food and Drug Administration (FDA) for home use (“see Relevant Websites section”).

Point-of-Care Laboratory Grade 3

In this setting, PoC for infectious disease testing is performed in a “home setting” by laymen individuals without any formal or informal previous training. Assays performed in this setting must be low complexity, CLIA-waived or similar.

The basic requirement for any infectious disease PoC diagnostic assay should be equality regarding performance when compared to assays performed within the central laboratory. For Point-of-Care Laboratory Grades 1A and 1B, it is strongly recommended to establish local institutional or regional committees with participation of the local stakeholders and may include establishing local, regional or national guidelines. According to guidelines published by The American Society for Microbiology (Dolen *et al.*, 2017) and The Danish Society for Clinical Microbiology (Lisby *et al.*, 2017), it is recommended to maintain clinical microbiology laboratory oversight and expertise when PoC testing for infectious diseases is implemented.

This may ensure selection of optimal assays and platform, verification of assays locally before implementation, test reporting and interpretation, quality control procedures and participation in quality control programs and – especially – electronic communication between the local PoC platform and the institutional Electronic Patient Record and/or the institutional Laboratory Information System. For Point-of-Care Laboratory Grade 2A, in some countries or regions, these recommendations may be implemented, in other countries/regions, this may not be feasible due to various factors such as organization of the local healthcare system. Analyzing more patient sample at the PoC setting and sending fewer samples to a central clinical microbiology service provider will increase the necessity to focus on “the full infectious disease picture” for each PoC geographical location. Without an established electronic communication between the individual PoC location and the central clinical microbiology

service provider, expert guidance regarding diagnostic workup, test indication and interpretation, and, especially, surveillance of microbial identification and resistance patterns may be compromised. For Point-of-Care Laboratory Grade 2B settings, the same considerations as for Grade 2A may be applied. However, as this setting will likely be without readily accessible central clinical microbiology service providers (e.g., by military use or use in low-resource settings), electronic communication of individual patient test results and clinical microbiology oversight in general may not be relevant. The biggest challenge to clinical microbiology oversight and surveillance of microbial prevalence may arise from infectious disease testing performed in future Point-of-Care Laboratory Grade 3 settings (home use). Unless regulatory requirements regarding test result management, e.g., mandatory upload to cloud-based databases, are established, most likely test results based upon PoC testing in this setting will not be available for clinical microbiology guidance or local/national surveillance. The potential negative impact on local and national surveillance and antimicrobial use, especially in regions with unrestricted access to antimicrobials, may override the potential positive impact by use of infectious disease diagnostics in this PoC setting, e.g., patient/user satisfaction.

PoC Menus

A variety of companies are marketing assays for immunofluorescence detection of antigens or antibodies on serum or feces e.g., Biocan or Abbott ("see Relevant Websites section"). Such test portfolios include viruses like adenovirus, Barmah Forest virus, Chikungunya virus, CMV, dengue virus, Ebolavirus, EBV, Hantaan virus, HAV, HBV, HCV, HIV, HSV 1 or 2, Influenza A and B virus, norovirus, Ross River virus, rotavirus, RSV, SARS-CoV-2, rubellavirus and Zika virus.

The current amplification based PoC tests are currently marketing panels targeting respiratory infections, gastrointestinal infections, CNS infections, HIV and HSV ([Table 1](#)). As such assays are being implemented in the clinical setting, it is anticipated that several companies will market returning travelers panels and post transplantation panels in the future.

A pragmatic approach to selecting PoC menus for current and future PoC platforms could be to select "low hanging fruits" according to business cases based upon the designated geographic location of the PoC instrument in question ([Fig. 1](#)). E.g., for PoC Grade 1a and 1b, assays targeting pathogens requiring contact isolation of the patient (such as SARS-CoV-2, Influenza A&B, RSV, MRSA, VRE, CPO, norovirus) would most likely emerge as the most attractive assays seen from a benefit-cost perspective.

In general, there is currently a lack of rapid open access amplification based PoC platforms that would allow end-users to apply laboratory developed assays allowing for fast detection of viruses according to local requirements for special diagnostics.

PoC Clinical Impact

The clinical impact of PoC testing can be evaluated in several ways including (1) intervention studies in which the change in patient management of a positive test is compared to negative test results, (2) prospective or retrospective case-control studies comparing the effect on patient management of the introduction of a PoC compared to a previous standard of care and (3) randomized studies comparing current standard of care with PoC and the effect on patient management. Such studies have primarily been published on respiratory illness evaluating influenza PoC testing. As the results of these studies have been conflicting, further studies are needed to understand the clinical impact of PoC testing.

Modeling and intervention studies in which the clinicians are blinded for PoC results have suggested that the introduction of PoC may result in better use of side room contact isolation ([Pedersen et al., 2018](#)), reduced prescription of antibiotics ([Pedersen et al., 2018](#); [Kaku et al., 2018](#)), targeted use of antiviral treatment ([Pedersen et al., 2018](#)) and cost savings due deferred admissions and shortened hospital stay ([Rathamat-Langendoen et al., 2019](#)). The potential for cost savings has been supported by one study investigating the use of PoC in adults in the emergency room, in which a positive test lead to deferred admission and reduction in antiviral prescribing and resulted in overall cost savings of approximately \$200 per emergency room visit ([Hansen et al., 2018](#)). A study on pediatric patients could not demonstrate an impact on admission or antibiotic prescription with exception of lower admission rate for children positive for influenza B but supported more appropriate prescription of antiviral treatment ([Busson et al., 2019](#)).

Studies before and after introduction of PoC for influenza testing support a more targeted use of antiviral treatment for children and adults ([Vos et al., 2019](#); [Vecino-Ortiz et al., 2018](#); [Benirschke et al., 2019](#)) and better use of side room contact isolation facilities for immunocompromised adults ([Vos et al., 2019](#)), but the use of PoC testing did not change length of hospital stay for children ([Vecino-Ortiz et al., 2018](#)) or immunocompromised adults ([Vos et al., 2019](#)). Similar, no positive effect on antibiotic prescription could be established for adults in urgent care settings or for immunocompromised adults ([Vos et al., 2019](#); [Benirschke et al., 2019](#)).

Randomized studies on adults agrees on more appropriate use of antiviral treatment by use of PoC testing ([Brendish et al., 2017](#); [Andrews et al., 2017](#)), but the studies are conflicting regarding whether PoC testing may shorten duration of antibiotics ([Brendish et al., 2017](#); [Shengchen et al., 2019](#)) or stay in hospital ([Brendish et al., 2017](#); [Andrews et al., 2017](#); [Shengchen et al., 2019](#)). In one study, the duration of antibiotics was shortened in the intervention group ([Shengchen et al., 2019](#)), but in another study no difference in initiation or mean duration of antibiotics could be established ([Brendish et al., 2017](#)). The two studies agree that stay in hospital was shortened in the PoC test group ([Brendish et al., 2017](#); [Shengchen et al., 2019](#)), but a third study, having length of stay as the primary outcome, could only demonstrate a reduction in time-to-test-result, whereas the length of stay in hospital was not associated with PoC testing ([Andrews et al., 2017](#)). The cost of hospitalization was compared between

Table 1 Commercially available Point-of-Care assays

<i>Syndromic testing</i>	<i>Gastrointestinal panel</i>		<i>Meningitis/Encephalitis</i>		<i>Respiratory panel</i>			<i>Tropical diseases – Returning travelers panel</i>			
Biofire Biomérieux	X		X		X						
ePlex GenMark	(X)		(X)		X						
NATlab Ador Dx			(X)						(X)		
QIAstat Qiagen	X				X						
Vivalytic Bosch	(X)		(X)		(X)						
(X) assays in pipeline											
<i>Targeted testing</i>	<i>SARS-CoV-2</i>	<i>SARS-CoV-2 + Influenza</i>	<i>Influenza A/B</i>	<i>Influenza + RSV</i>	<i>RSV</i>	<i>Norovirus</i>	<i>GBS</i>	<i>GAS</i>	<i>Enteric viral panel</i>	<i>HIV 1 + 2</i>	<i>HSV 1 + 2</i>
Aries Luminex Corp	X			X							X
BDmax BD	X								X		
Vivalytic Bosch	X										
Cobas Liat Roche		X	X	X			X	X			
GeneXpert Cepheid	X		X	X		X		X			
Idnow Abbott	X		X		X		X				
Meridian Revogene											
Mesa Biotech	X		X								
Simplexa Focus Dx	X			X						X	

the PoC group and the standard of care group in one study, finding an average reduction in cost of approximately \$200 (Shengchen *et al.*, 2019).

Most studies agree on PoC testing resulting in a more appropriate use of antiviral treatment, whereas the data are conflicting on the impact of PoC testing on use of antibiotics and the effect on length of stay in hospital, whereas the data the impact of PoC testing on use of side room contact isolation facilities are sparse. Based on current studies, it seems that PoC testing may impact patient management of adults seen in the emergency room more than children and immunocompromised adults, but more studies are needed to understand the true impact of PoC testing on patient management.

Conclusion

Clinical virology diagnostics is currently undergoing a paradigm shift. The diagnostic procedure is no longer confined within the geography of a central medical/clinical microbiology/virology laboratory but may be performed closer to the patient by non-laboratory trained personnel. The different versions of “Point-of-Care” testing may present challenges to platform and assay selection, verification of assay performance, training of non-laboratory personnel performing the tests as well as quality control. Clinical microbiology/virology oversight is highly recommended in order to maintain the same level of diagnostic quality compared to diagnostics by gold standard assays performed by a central laboratory. Furthermore, clinical microbiology laboratory oversight may ensure capture of the PoC assay result in the patient’s electronic medical record and in the laboratory information system. Thus, the likely loss of data caused by manual recording of test results in the electronic patient records may be avoided, maintaining local, regional and national surveillance of select infections, such as Influenza Virus infections. With limited resources available to the healthcare sector, benefit-cost assessments of new diagnostic interventions should be mandatory. The impact of PoC testing on clinical patient management and healthcare costs, such as use of antimicrobials, use of antivirals, use of side room contact isolation facilities and length of stay in hospital/ICU should be documented, as limited data is yet available.

References

- Andrews, D., Chetty, Y., Cooper, B.S., *et al.*, 2017. Multiplex PCR point of care testing versus routine, laboratory-based testing in the treatment of adults with respiratory tract infections: A quasi-randomised study assessing impact on length of stay and antimicrobial use. *BMC Infectious Diseases* 17, 671.
- Babady, N.E., England, M.R., Smith, K.L.J., *et al.*, 2018. Multicenter evaluation of the ePlex respiratory pathogen panel for the detection of viral and bacterial respiratory tract pathogens in nasopharyngeal swabs. *Journal of Clinical Microbiology* 56, e01658-17.
- Benirschke, R.C., McElvania, E., Thomson, R.B., Kaul, K.L., Das, S., 2019. Clinical impact of rapid point-of-care PCR influenza testing in an urgent care setting: A single-center study. *Journal of Clinical Microbiology* 57, e01281-18.
- Brendish, N.J., Malachira, A.K., Armstrong, L., *et al.*, 2017. Routine molecular point-of-care testing for respiratory viruses in adults presenting to hospital with acute respiratory illness (ResPOC): A pragmatic, open-label, randomized controlled trial. *The Lancet Respiratory Medicine* 5, 401–411.
- Busson, L., Bartiaux, M., Brahim, S., *et al.*, 2019. Contribution of the FilmArray respiratory panel in the management of adult and pediatric patients attending the emergency room during 2015–2016 influenza epidemics: An interventional study. *International Journal of Infectious Diseases* 83, 32–39.
- Cárdenas, A.M., Edelstein, P.H., Alby, K., 2014. Development and optimization of a real-time PCR assay for detection of herpes simplex and varicella-zoster viruses in skin and mucosal lesions by use of the BD Max open system. *Journal of Clinical Microbiology* 52, 4375–4376.
- Catt, K.J., Tregar, G.W., 1967. Solid phase radioimmunoassay in antibody-coated tubes. *Science* 158, 1570–1572.
- Chartrand, C., Tremblay, N., Renaud, C., Papenburg, J., 2015. Diagnostic accuracy of rapid antigen detection tests for respiratory syncytial virus infection: Systematic review and meta-analysis. *Journal of Clinical Microbiology* 53, 3738–3749.
- Cobb, B., Simon, C.O., Stramer, S.L., *et al.*, 2017. The COBAS 6800/8800 system: A new era of automation in molecular diagnostics. *Expert Review of Molecular Diagnostics* 17, 167–180.
- Dolen, V., Bahk, K., Carroll, K.C., *et al.*, 2017. Changing Diagnostic Paradigms for Microbiology: Report on an American Academy of Microbiology Colloquium held in Washington, DC, from 17 to 18 October 2016. Washington, DC: American Society for Microbiology, doi:10.1128/AAMCol.17–18Oct.2016.https://www.ncbi.nlm.nih.gov/books/NBK447255/.
- Dulbecco, R., 1952. Production of plaques in monolayer tissue cultures by single particles of an animal virus. *Proceedings of the National Academy of Sciences of the United States of America* 38, 747–752.
- Hansen, G.T., Moore, J., Herding, E., *et al.*, 2018. Clinical decision making in the emergency department setting using rapid PCR: Results of the CLADE study group. *Journal of Clinical Virology* 102, 42–49.
- Iwanowski, D., 1892. Über die mosaikkrankheit der tabakspflanze. *Bulletin Scientific/Académie Impériale des Sciences de Saint Petersburg* 35, 67–70.
- Jenny, S.L., Hu, Y., Overduin, P., Meijer, A., 2010. Evaluation of the xpert flu a panel nucleic acid amplification-based point-of-care test for influenza A virus detection and pandemic H1N1 subtyping. *Journal of Clinical Virology* 49, 85–89.
- Kaku, N., Hashiguchi, K., Iwanaga, Y., *et al.*, 2018. Evaluation of FilmArray respiratory panel multiplex polymerase chain reaction assay for detection of pathogens in adult outpatients with acute respiratory tract infection. *Journal of Infection and Chemotherapy* 24, 734–738.
- Koczula, K.M., Gallotta, A., 2016. Lateral flow assays. *Essays in Biochemistry* 60, 111–120.
- Lisby, G., 1998. Application of nucleic acid amplification in clinical microbiology. In: Meltzer, S. (Ed.), *PCR in Bioanalysis, series: Methods in Molecular Biology*. New Jersey: The Humana Press Inc, pp. 1–29.
- Lisby, J.G., Andersen, D.T., Chen, M., *et al.*, 2017. Anbefalinger vedrørende implementering og anvendelse af Point-of-Care teknologi til diagnostik af infektionssygdomme. Available at: <https://dskm.dk/wp-content/uploads/2016/08/POC-Rapport-fra-DSKMs-Udvalg.pdf> (accessed 5.10.2020).
- Mackay, I.M., Arden, K.E., Nitsche, A., 2002. Real-time PCR in virology. *Nucleic Acids Research* 30, 1292–1305.
- Nakane, P.K., Pierce Jr, G.B., 1966. Enzyme-labelled antibodies: Preparation and application for the localization of antigens. *Journal of Histochemistry and Cytochemistry* 14, 929–931.
- Nie, S., Roth, R.B., Stiles, J., *et al.*, 2014. Evaluation of Alere i Influenza A&B for rapid detection of influenza viruses A and B. *Journal of Clinical Microbiology* 52, 3339–3344.
- Niesters, H.G., 2002. Clinical virology in real time. *Journal of Clinical Virology* 25 (Suppl 3), S3–S12.

- Parcina, M.P., Schneider, U.V., Visseaux, B., *et al.*, 2020. Multicenter evaluation of the QIAstat respiratory panel – A new rapid highly multiplexed PCR based assay for diagnosis of acute respiratory tract infections. *PLOS One* 15, e023018.
- Paz-Bailey, G., Rosenberg, E.S., Doyle, K., *et al.*, 2018. Persistence of zika virus in body fluids – Final report. *New England Journal of Medicine* 379, 1234–1243.
- Pedersen, C.J., Rogan, D.T., Yang, S., Quinn, J.V., 2018. Using a novel rapid viral test to improve triage of emergency department patients with acute respiratory illness during flu season. *Journal of Clinical Virology* 108, 72–76.
- Poritz, M.A., Blaschke, A.J., Byington, C.L., *et al.*, 2011. FilmArray, an automated nested multiplex PCR system for multi-pathogen detection. Development and application to respiratory tract infection. *PLOS One* 6, e26047.
- Prince, H.E., Lapé-Nixon, M., 2014. Role of cytomegalovirus (CMV) IgG avidity testing in diagnosing primary CMV infection during pregnancy. *Clinical and Vaccine Immunology* 21, 1377–1384.
- Rathamat-Langendoen, J., Groenewoud, H., Kuijpers, J., Melchers, W.J.G., van der Wilt, G.J., 2019. Impact of molecular point-of-care testing on clinical management and in-hospital costs of patients suspected of influenza or RSV infection: A modeling study. *Journal of Medical Virology* 91, 1408–1414.
- Reed, W., 1902. Recent researches concerning the etiology, propagation and prevention of yellow fever, by the United States Army Commission. *Journal of Hygiene* 11, 101–119.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., *et al.*, 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487–491.
- Saiki, R.K., Scharf, S., Faloona, F., *et al.*, 1985. Enzymatic amplification of β -Globin sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350–1354.
- Shengchen, D., Gu, X., Fan, G., *et al.*, 2019. Evaluation of a molecular point of care testing for viral and atypical pathogens on intravenous antibiotic duration in hospitalized adults with lower respiratory tract infection: A randomized clinical trial. *Clinical Microbiology and Infection* 25, 141.
- Smith, T.F., Wold, A.D., Epsy, M.J., 1992. Diagnostic virology – Then and now. *Advances in Experimental Medicine and Biology* 312, 191–199.
- Tan, E., Erwin, B., Dames, S., *et al.*, 2008. Specific versus nonspecific isothermal DNA amplification through thermophilic polymerase and nicking enzyme activities. *Biochemistry* 47, 9987–9999.
- Tanriverdi, S., Chen, L., Chen, S., 2010. A rapid and automated sample-to-result HIV load test for near-patient application. *Journal of Infectious Disease* 201 (Suppl), S52–S58.
- Vecino-Ortiz, A.I., Goldenberg, S.D., Douthwaite, S.T., *et al.*, 2018. Impact of a multiplex PCR point-of-care test for influenza A/B and respiratory syncytial virus on an acute pediatric hospital ward. *Diagnostic Microbiology and Infectious Disease* 91, 331–335.
- Voermans, J.J., Seven-Deniz, S., Fraaij, P.L., *et al.*, 2016. Performance evaluation of a rapid molecular diagnostic, MultiCode based, sample-to-answer assay for the simultaneous detection of Influenza A, B and respiratory syncytial viruses. *Journal of Clinical Virology* 85, 65–70.
- Vos, L.M., Weehuizen, J.M., Hoepelman, A.I.M., *et al.*, 2019. More targeted use of oseltamivir and in-hospital isolation facilities after implementation of a multifaceted strategy including a rapid molecular diagnostic panel for respiratory viruses in immunocompromised adult patients. *Journal of Clinical Virology* 116, 11–17.
- Weller, T.H., Robbins, F.C., Enders, J.F., 1949. Cultivation of poliomyelitis virus in cultures of human foreskin and embryonic tissues. *Proceedings of the Society for Experimental Biology and Medicine* 72, 153–155.
- Wirten, M., Larrouy, L., Mahjoub, N., *et al.*, 2017. Multicenter comparison of the new COBAS 6800 system with COBAS Ampliprep/COBAS Taqman and Abbott RealTime for the quantification of HIV, HBV and HCV viral load. *Journal of Clinical Virology* 96, 49–53.
- Woodberry, M.W., Shankar, R., Cent, A., Jerome, K.R., Kuypers, J., 2013. Comparison of the Simplexa FluA/B & RSV direct assay and laboratory-developed real-time PCR assays for detection of respiratory virus. *Journal of Clinical Microbiology* 51, 3883–3885.
- Wouters, Y., Keyaerts, E., Rector, A., *et al.*, 2019. Comparison of the Idylla™ respiratory (IFV-RSV) panel with the GeneXpert Xpert® Flu/RSV assay: A retrospective study with nasopharyngeal and midturbinate samples. *Diagnostic Microbiology and Infectious Disease* 94, 33–37.

Relevant Websites

- <https://www.globalpointofcare.abbott/en/product-details/id-now.html>
Abbott Global Points of Care. ID NOW™.
- <http://www.rapidtest.ca/products>
Biocan Diagnostics Inc.
- <https://www.cdc.gov/labquality/waived-tests.html>
Centers for Disease Control and Prevention. Waived Tests.

Standardization of Diagnostic Assays

Sally A Baylis and C Micha Nübling, Paul-Ehrlich-Institute, Langen, Germany
Wayne Dimech, National Serology Reference Laboratory, Fitzroy, VIC, Australia

© 2021 Elsevier Ltd. All rights reserved.

Glossary

International Standard Internationally recognized measurement standard established by the WHO.

International Unit Value given to WHO International Standards.

WHO World Health Organization.

Introduction

In clinical virology laboratories, diagnostic testing is performed to confirm the cause of viral infections, whether they are acute or chronic or represent reactivation of latent infections. The diagnostic test results help inform physicians about the best course of patient management, for example, whether to prescribe antiviral drugs or treat patients with specific immunoglobulins or administer vaccines prophylactically or therapeutically. Assay standardization is essential in order to compare diagnostic testing results produced by different laboratories, ensure robust testing procedures as well as to develop clinical practice guidelines where diagnostic test results such as viral loads, for example, are used in clinical decision making.

In this article, we describe different aspects of standardization including the development and use of reference materials, the role of written standards as well as principles of quality assurance. Standardization has been applied extensively to the detection of viral DNA and RNA by nucleic acid amplification technology (nucleic acid testing (NAT) or nucleic acid amplification testing (NAAT)). In the case of serological assays used to detect viral antigens or specific antibodies induced following viral infections, standardization has been useful for a more limited range of assays. Testing of an immune response to an organism is not straightforward. When testing for antibodies against human immunodeficiency virus (HIV), for example, the different antigens derived from particular viral subtypes and mutations, the immune response to distinct antigens and the unique development of an individual's immune response, makes the application of standardization difficult. In contrast, for serological assays that detect and quantify viral antigens, such as HIV p24, conventional approaches to standardization are more applicable.

Physical Standards (Reference Materials)

Because of the complex, biological nature of diagnostic specimens tested in clinical virology laboratories, they cannot be evaluated by physicochemical means. Instead bioassays, measuring a functional biological activity are necessary. Reference materials, also termed standards, are essential for the control of such bioassays. These reference materials may be measurement standards, applied to quantitative assays. They can also be controls used qualitatively or quantitatively to monitor day-to-day performance of a test. The most important measurement standards used in virology laboratories are World Health Organization (WHO) International Standards (ISs) which have defined concentrations of specific analytes. The WHO ISs are considered "higher order standards" for biological assays and are critical for calibration of different assays using a common material and for ensuring calibration traceability. Concentrations of ISs are expressed in International Units (IUs) per ml. Because of the biological nature of WHO ISs results cannot simply be reported in International System of Units (SI)-related units such as moles, hence the adoption of the IU.

The ISs are used to directly calibrate secondary standards, including ones that used at a regional or national level, calibrators used for commercially available assays, as well as control materials (e.g., working standards/reagents and run controls, independent of controls supplied by manufacturers of diagnostic kits) used to monitor assay performance. In turn, tertiary standards are directly traceable to secondary standards ([Fig. 1](#)). Collectively, these standards enable the comparison of results among different assays and different laboratories.

Uncertainty values are not assigned to WHO ISs, since IU are arbitrary units and variance is associated the contents in the vial of standard. For secondary and tertiary standards there will be a measurement uncertainty when calibration is performed ([Fig. 1](#)).

The WHO also produces reference materials termed International Reference Panels (IRPs) and no unitages are assigned to these materials. IRPs designed for NAT assays or antigen assays include different genotypes or strains of a virus that might be encountered globally and may challenge current assays. By analogy, IRPs designed for antibody assays reflect sera of patients who developed antibodies against different virus genotypes or strains or a different times post-infection. The WHO reference reagents (RRs), are usually interim standards with a unitage defined in units rather than IU. The types of reference material and their role in the clinical virology laboratory are summarized in [Table 1](#).

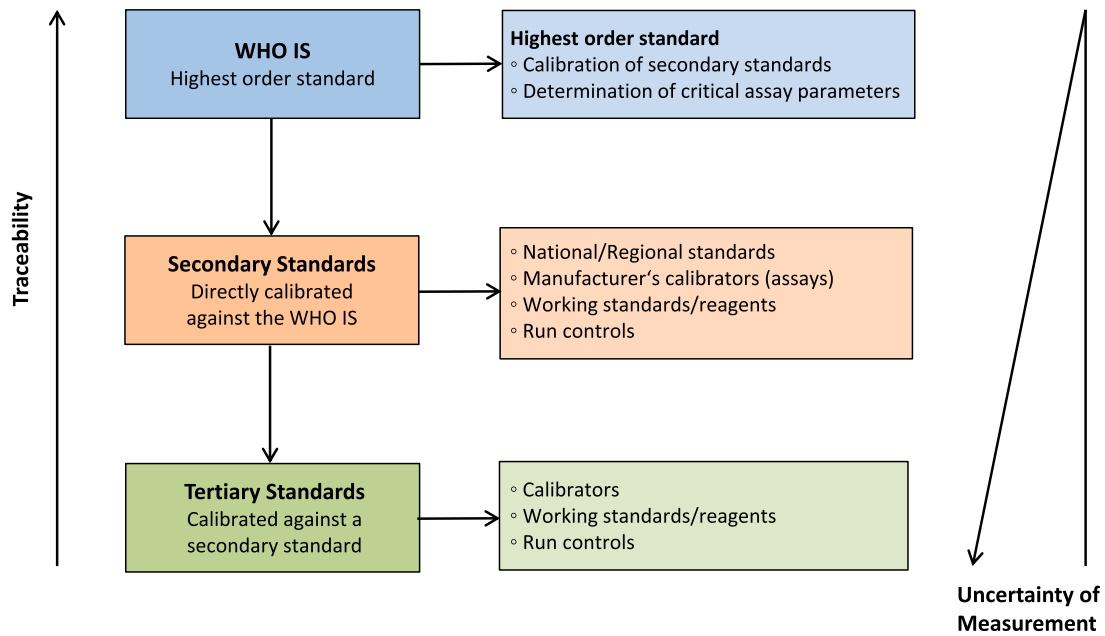


Fig. 1 Hierarchy of standards. The relationship between ISs and secondary and tertiary standards is shown, together with their uses. Reproduced, in modified form, with the permission of the Journal of Clinical Microbiology.

Table 1 Types of standard and their role in the clinical virology laboratory

Type of standard	Role
WHO IS	Highest order reference material. Used for calibration of secondary standards and evaluation of critical assay parameters such as limits of detection.
WHO IRP	Ensuring adequate detection of genetically and antigenically distinct strains of important viral pathogens and associated antibodies.
Secondary standards	National and regional standards traceable to the WHO IS with value assigned in IU. Manufacturer's calibrators used for control of commercial assays with value assigned in IU. Working standards/working reagents used as routine external run controls for assays, assigned a value or range in IU.
Tertiary standards	Calibrators traceable to a secondary standard calibrated in IU. Working standards/working reagents used as routine external run controls for assays, assigned a value or range in IU traceable to a secondary standard.
Working standards/working reagents	Used as routine external run controls for assays, no unitage assigned or given an arbitrary value if no WHO IS exists.

Preparation and Evaluation of Standards

The decision to develop a specific standard is based upon public health need and the prioritization of WHO standards is determined through discussion with the medical scientific communities worldwide and in consultation with the WHO Expert Committee on Biological Standardization (ECBS). The process for the preparation and evaluation of WHO standards, including their replacement is shown in [Fig. 2](#).

The starting materials for the preparation of any kind of standard are carefully pre-qualified to ensure they are fit for purpose. In the case of ISs used for NAT, they should contain the specific virus in sufficiently high titer, the strain should be detected in commonly used assays and the final formulation should not affect NAT assay performance. Similarly, starting materials for viral antigens or antibodies for IS are prepared at a titer that will enable the calibration of secondary standards.

The ISs are usually lyophilized in order to maintain stability of the reference material, allowing worldwide distribution at ambient temperatures therefore avoiding cold chain difficulties. Batches of IS are tested to ensure that volume dispensed during filling is consistent and that the homogeneity and potency of the analyte after filling and lyophilization is adequate. After lyophilization, vials are filled with nitrogen and residual moisture in the freeze-dried residue (cake) is determined together with

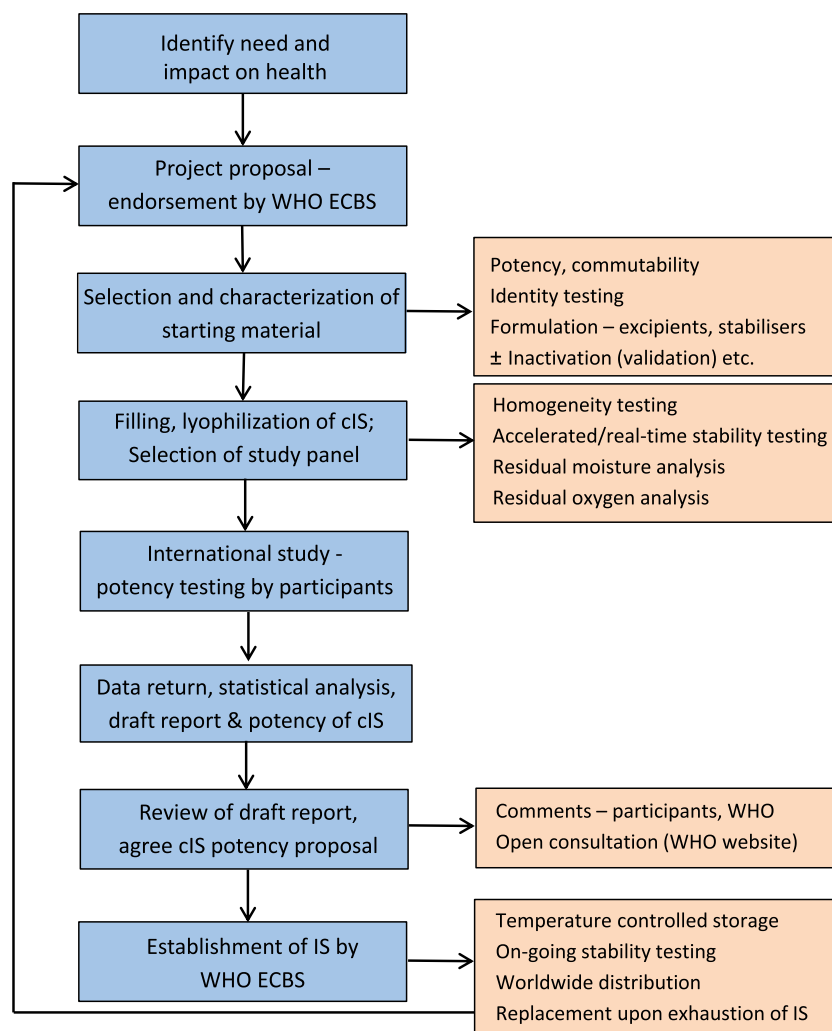


Fig. 2 Process for the development of WHO ISs, RRs and IRPs. The procedure is shown from the identification of a scientific need to develop a standard to the establishment of the standard and ultimately to its replacement. cIS, candidate IS. Reproduced, in modified form, with the permission of the Journal of Clinical Microbiology.

the levels of oxygen present in the final vial to ensure that levels do not exceed predefined specifications that may impact on the stability. Stability of IS must be evaluated to ensure that there is no degradation of the material during storage and shipment that would affect the unitage assigned to the analyte and, in turn, the value assigned to any secondary or tertiary standards. Occasionally, it has been necessary to add stabilizing agents to certain NAT standard preparations, especially for some RNA viruses that are more susceptible to degradation by nucleases. Long-term stability is predicted by accelerated degradation studies with incubation of the candidate material at higher temperatures, for defined time periods, together with continuous monitoring by testing of the IS after its establishment at defined time intervals.

Details of the specifications are described in the WHO Recommendations for the preparation, characterization and establishment of international and other biological reference standards (2004).

The WHO ISs are prepared using material that reflects the sample type being tested in an assay, and materials must be assessed for their “commutability” to ensure that their performance in different assays is equivalent to the usual test samples such as patient specimens tested routinely. The property of commutability is particularly affected by the composition of the sample matrix which may vary between the reference material and the routine specimens. The source materials for ISs for NAT are from viraemic plasma from infected donors or patients or by using virus stocks produced in cell culture, which may be heat-inactivated, and spiked into human plasma. For NAT standards, the use of intact virus particles is important because such preparations undergo nucleic acid extraction as well as the amplification and detection steps in a NAT assay. Materials made from *in vitro* RNA transcripts or plasmids do not control for the extraction step or reflect the complexities of viral genome structures and have not proved to be helpful in assay harmonization. Antigen standards have also been prepared from plasma from infected patients or from virus propagated in cell culture or from recombinant sources.

Developing ISs for antibody detection, is more complex as a single reference preparation is not necessarily suitable for all purposes. An infected individual's antibody response develops over time, with low avidity antibodies being replaced by higher avidity antibodies during the infection course. Furthermore, antibody features (antibody classes; epitopes and antigens targeted by antibodies) depend not only on strain or genotype of the pathogen, but also on parameters like the site and route of infection, inoculum size or infectious dose. Last but not least, the individual antibody response is unpredictable and often highly variable between infected individuals, dependent on different components of the individual immune system. Diagnostic assays may have different designs, e.g., in respect to detect "functional" antibodies (e.g., neutralizing antibodies) or "binding" antibodies, often with different antigens and epitopes used in the individual tests. With all these different aspects to consider, standardization of antibody tests is more difficult. Standardization of antibody assays restricted to a specific viral antigen, for example in the case of hepatitis B virus (HBV), anti-HBV core (anti-HBc), appears more feasible compared to assays detecting total antibodies to HBV: the designs of different anti-HBV assays may differ in the HBV antigens used with the consequence that each measures different subsets of anti-HBV antibodies. The IRPs for antibodies are often comprised of different specimens to reflect the diversity of antibody responses between individuals. Some reference materials have been prepared from therapeutic immunoglobulins, raising potential matrix (commutability) issues.

The WHO reference materials i.e., ISs, IRPs and RRs are evaluated in international studies typically involving reference laboratories, clinical laboratories, regulatory organizations as well as manufacturers of diagnostic kits and reflecting different globally available assays. Such studies allow the head-to-head comparison of different assays used globally, including both commercially available tests as well as laboratory-developed tests (LDTs) or in-house assays. Potencies of the candidate reference materials are determined by the participants. [Fig. 3](#) shows an example of the range of potencies reported by laboratories for the IS for hepatitis E virus (HEV) RNA. Potencies were determined by a mixture of qualitative and quantitative assays with a wide range of reported values. However, expressing results of other study samples against the candidate IS greatly reduced the variation in the measured potencies reported by participants. This reduction in variation is one of the main aims of the collaborative study because it demonstrates that the use of a reference material can improve harmonization of the results of diagnostic assays. The outcome of the studies are published on the WHO website to allow time for comments before it is presented to the WHO ECBS which endorses the proposed unitage and confirms whether materials are established as ISs, IRPs etc.

Several thousand vials of each IS are prepared, once the stocks near exhaustion the IS must be replaced. The continuity of the IU assignment is important and replacement material must be carefully selected to ensure they are similar to the previous IS or have been demonstrated to perform in a similar way. In some cases, where large volumes of the original source material are available this process is easier. Preparation of the replacement IS follows the same manufacturing procedures as the establishment of a new IS. In the collaborative study the replacement material is evaluated alongside the existing IS in order to determine the potency of the new standard.

Written Standards and Guidelines

In order to standardize diagnostic testing, the availability and correct implementation of reference materials is essential; underpinning this are documentary, written standards and guidelines relating to development of tests and their routine use in the diagnostic laboratory.

With respect to the reference materials themselves, the WHO ECBS publishes a Technical Report Series (see "Relevant Websites section") which records opinions of the committee with respect to the establishment of new reference materials and reviews strategic matters in the field of standardization including guidelines outlining the procedures used for the development and establishment of WHO ISs and IRPs together with guidance on the calibration of secondary and tertiary standards. The WHO also publishes biological standardization study reports (BS documents) which document each standardization project, including an analysis of the data generated during the collaborative study, and serve as a basis for the decision of the WHO ECBS to establish new and replacement standards.

The International Organization for Standardization (ISO) is a non-governmental, independent standard setting body that aims to improve quality, reliability and safety of products and services by providing common standards around the world. Several ISO standards are important in the standardization of diagnostics. In particular, ISO 17511:2003 which comprehensively deals with metrological traceability of values assigned to calibrators and control materials. Supporting standardization is ISO 15189:2012 which sets out the requirements for quality management and competence in medical laboratories and is now mandatory in many countries. External quality assurance (EQA) schemes are important in independently demonstrating adequate quality of testing in laboratories. The materials used in EQA samples should be commutable and traceable to established standards and reference materials as defined in ISO 17043:2010 which provides guidance on the development and operation of EQA programmes. The Clinical and Laboratory Standards Institute (CLSI) is a voluntary organization developing clinical and laboratory practices globally, including, for example, guidance concerning commutability of reference materials. The International Laboratory Accreditation Cooperation (ILAC) is an international organization for accreditation bodies for testing and calibration laboratories and produces documents on measurement traceability and selection of reference materials. There are other guidelines and regulations relating to *in vitro* diagnostics either at a national or a regional level worldwide.

The guidelines described above are not an exhaustive list of written standards, but provide some of the most important sources of information and are summarized in [Table 2](#).

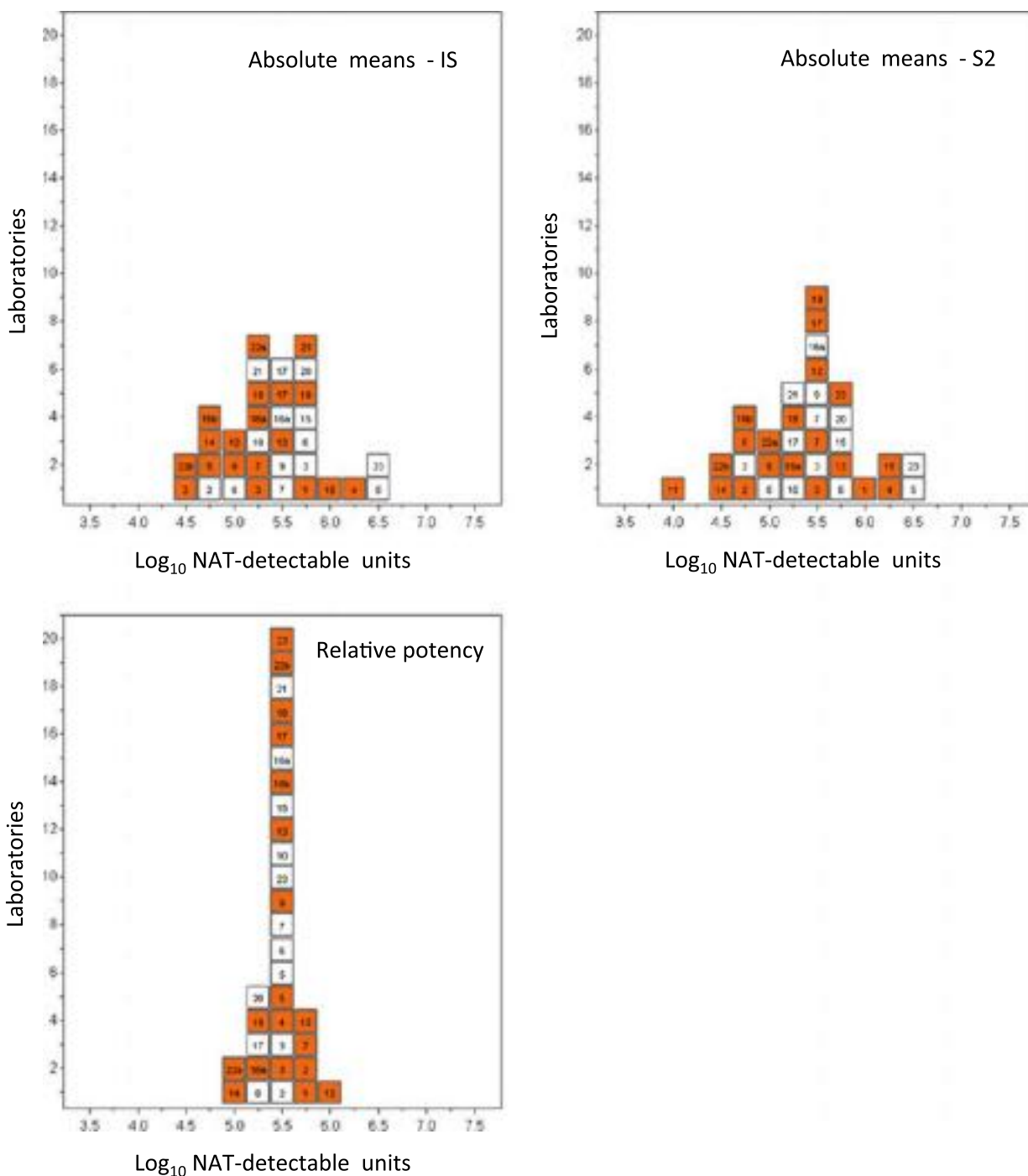


Fig. 3 Harmonization of data by expressing potencies of a sample against a WHO IS. The top two panels show absolute mean values of the WHO IS for HEV RNA (left) and a blinded replicate sample S2 (right). By expressing the potency of S2 against the IS, the agreement is much improved between laboratories (bottom left). White indicates data reported from quantitative assays and orange indicates qualitative assays. The number of laboratories is indicated on the vertical axis. Laboratory code numbers are indicated in the respective boxes. Reproduced, in modified form, with the permission of Emerging Infectious Diseases.

Quality Assurance

In the diagnostic virology laboratory, quality assurance (QA) is a planned and systematic set of activities designed to ensure confidence in the results of different diagnostic assays. The QA process includes the establishment of a quality management system (QMS) setting out the scope of activities, policies, procedures, documented information and resources needed for implementation

Table 2 Examples of written standards

<i>World Health Organization</i>	
WHO Technical Report Series 932	World Health Organization recommendations for the preparation, characterization and establishment of international and other biological reference standards
WHO Technical Report Series 1004	Manual for the preparation of secondary reference materials for in vitro diagnostic assays designed for infectious disease nucleic acid or antigen detection: calibration to WHO international standards
International Organization for Standardization	
ISO 15189:2012	Medical laboratories – Requirements for quality and competence
ISO/IEC 17025:2017	General requirements for the competence of testing and calibration laboratories
ISO 17511:2003	In vitro diagnostic medical devices – Measurement of quantities in biological samples – Metrological traceability of values assigned to calibrators and control materials
ISO/IEC 17043:2010	Conformity assessment – General requirements for proficiency testing
CLSI Standard: EP30, 1st Edition	Characterization and Qualification of Commutable Reference Materials for Laboratory Medicine
European Commission	Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices
	Commission Decision of 3 February 2009 amending Decision 2002/364/EC on common technical specifications for <i>in vitro</i> -diagnostic medical devices (2009/108/EC)
US Food and Drug Administration	https://www.fda.gov/medical-devices/ivd-regulatory-assistance/overview-ivd-regulation#9 https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance

and maintenance of the testing service. QA measures should include the use of standard operating procedures (SOPs) and associated documentation; correct handling of samples to ensure their appropriate storage and traceability, calibration and preventive maintenance of equipment, the use of designated laboratories, and the use of pre-qualified and in-date reagents. Operators should be suitably trained in technical procedures as well as specific assays, be competent in data interpretation, and able to undertake remedial/corrective actions where necessary. Audits and reviews are performed to ensure adequate performance and identify any issues and areas for improvement.

The Use of Controls

Assessing an assay's performance over time is achieved by testing of run controls and monitoring the test results. Controls for monitoring assays include ones provided in diagnostics kits supplied by the kit manufacturers – these typically include positive and negative controls or controls for quantification. Other controls, termed external run controls are provided by other organizations and represent an independent means of monitoring performance. All assays experience variation. The sources of variation are derived from the equipment and consumables used, operators, changes in calibration of instruments, transport and storage of reagents and general environmental conditions such as temperature and humidity. For serological assays, changes in reagent lot numbers are the greatest source of normal variation.

The run controls are tested frequently, usually on a daily or per test run basis, and the results plotted on a Levey-Jennings (LJ) chart (Fig. 4). In a well-designed run control program, samples known to be reactive at medical decision points are used. Preferably they are of known homogeneity and stability, as are controls registered as *in vitro* diagnostic devices. Ideally, QC samples are calibrated against an IS (where one is available). This is usually the case in nucleic acid testing QCs. If this is not possible, the manufacturer should ideally reserve sufficient stock material to allow minimal QC lot-to-lot variation, releasing each new batch against in-house QC reference materials. By achieving consistency of QC reactivity from lot-to-lot, it is possible to monitor the performance of the assay over a long period of time.

If the QC sample is calibrated against the corresponding IS and is proven to be commutable with patients samples, the results of QC testing can measure not only the imprecision of the test system (i.e., the standard deviation (SD) or coefficient of variation) but it can also measure the bias of the assay (i.e., how far removed the mean of the QC results are from the QC assigned value). These metrics are valuable in estimating the measurement of uncertainty of the assay.

Critical to a run control program is determining the acceptance range of test QC test result. Acceptance limits set too tight will falsely reject test runs leading to unnecessary investigations, whereas limits set too broadly may result in incorrect results being reported and defects in the test systems not being identified. Traditionally, the mean and SD of QC results are used to determine acceptance limits. International guidelines suggest using 20 QC test results and set the acceptance limits at mean $\pm$ 2 SDs; and they apply a set of QC "rules", called Westgard Rules, to detect "out-of-control" results. Although this approach works well for clinical chemistry, it is not applicable to infectious disease testing, especially serology, due to changes in reactivity of the QC sample due to reagent lots (Figs. 5 and 6). An alternative method for calculating acceptance limits has been published. QConnect Limits (see "Relevant Websites section") use historical data from many laboratories using the same QC sample and assay

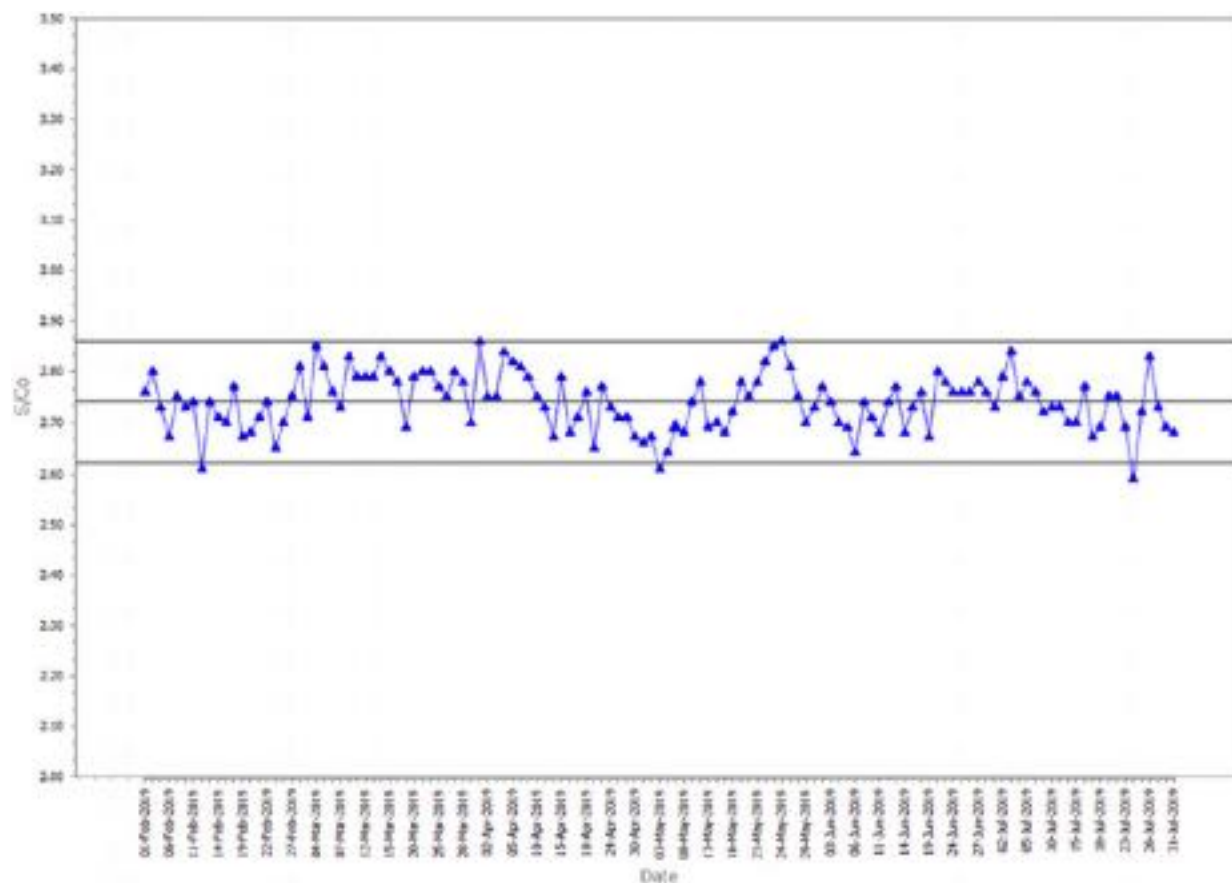


Fig. 4 Results of a QC sample tested repeatedly over time plotted on a LJ chart. The vertical axis represents the signal to cut-off values of the QC sample. The horizontal axis represents the date of testing. The results reported over the period are all within the mean $\pm$ 2 standard deviations (indicated by upper and lower horizontal lines) of the QC test results displayed. The mean value is represented by the horizontal line in the middle of the plot.

combination (i.e., peer group). By using data from multiple laboratories over a long period of time, all sources of “normal” variation can be accounted for and more robust acceptance criteria established.

Batch Testing

Commercial assays undergo testing by manufacturers to ensure adequate performance before they are released onto the market. In some regions of the world, regulatory authorities perform independent testing of assays before they are marketed as well as evaluating different batches during the lifetime of a particular product. In Europe, IVDs deemed to be high risk i.e., those where the consequences of a false diagnosis could cause serious harm both to the patient and to the community include tests for HIV and hepatitis B and hepatitis C. Independent batch testing is done for these high risk IVDs in some European countries using panels of well-characterized samples from patients or blood donors. Assays must show consistent performance, for example adequate sensitivity should be maintained, before new batches are released on the common market. In addition, technical files documenting the design, development, performance and manufacture of these high risk IVDs are independently reviewed by Notified Bodies who are involved in the conformity assessment of IVDs. With the introduction of new regulations (laws) concerning IVDs in Europe, verification of the performance of high risk IVDs by specialised reference laboratories will be a legal requirement before kits receive a EC declaration of conformity (i.e., CE-mark). Similar procedures are in place in the USA (see “Relevant Websites section”) and elsewhere.

External Quality Assessment/Proficiency Testing

EQA schemes, also referred to as proficiency testing (PT) or ring trials, are an important part of the overall standardization process. The EQA organizers distribute the same materials (which are usually blinded) to participating laboratories. The laboratories test

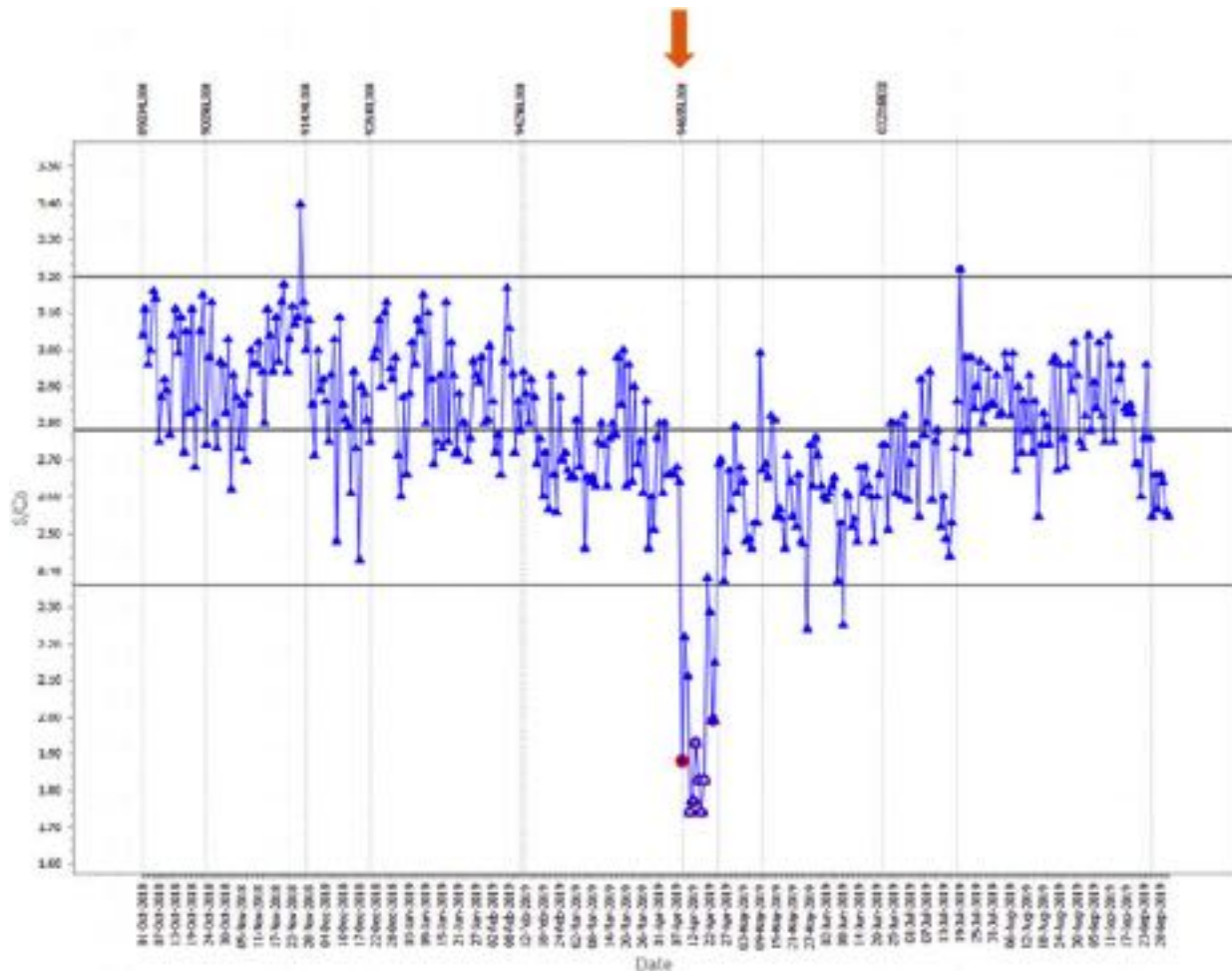


Fig. 5 Results of an anti-hepatitis C virus antibody QC sample tested repeatedly over time. A significant decrease in reactivity of the QC sample was associated with the introduction of a specific reagent lot (orange arrow). The results (signal to cut-off value on the vertical axis) of the QC sample tested on the affected reagent lot were outside the mean $\pm$ 2 SDs (upper and lower bold lines) of the QC test results displayed. Reagent lots are indicated at the top of the figure. The horizontal axis represents the date of testing.

the samples using routine methods and return results to the organizer. In some schemes associated metadata, such as reagent lot numbers and expiry dates are also collected by the organizer for analysis. The participation in EQA schemes provides laboratories with an objective means of assessing and demonstrating the reliability of the results they are producing. The EQA scheme covers the overall performance of a laboratory, including the entire process from receipt and storage of the samples, the laboratory testing, the interpretation and transcription of the data generated, to the reporting of results. Failure at any stage of this process affects the proficiency of the laboratory. Participation in EQA schemes is mandatory for laboratories accredited to ISO 15189:2012 and is important for laboratories seeking accreditation.

Laboratory Accreditation

Accreditation of laboratories to ISO 15189:2012 or ISO/IEC 17025:2017 provides objective and internationally recognized evidence of a level of quality. ISO accreditation ensures that the laboratories have robust and well-documented procedures and that test systems employed are fit for purpose, validated and under control. When a previously validated assay (including commercial assays used in strict accordance with the manufacturer's instructions for use) is introduced into use, verification is required before bringing the assay into routine use to demonstrate their utility. Validation protocols should be prospective i.e., they should set acceptance limits for assay performance which should be verified by the validation experiments and demonstrate that a particular procedure is suitable for its intended purpose. Once an analytical method has been established, a protocol for validation of the analytical procedure can be prepared. The analytical method should yield consistent results under standard operating conditions, critical points should be known, and robustness studies should be performed as part of the validation exercise. Upon completion of the validation experiments, the results are formally documented. Only if an analytical procedure meets the acceptance criteria set out in the validation protocol is the

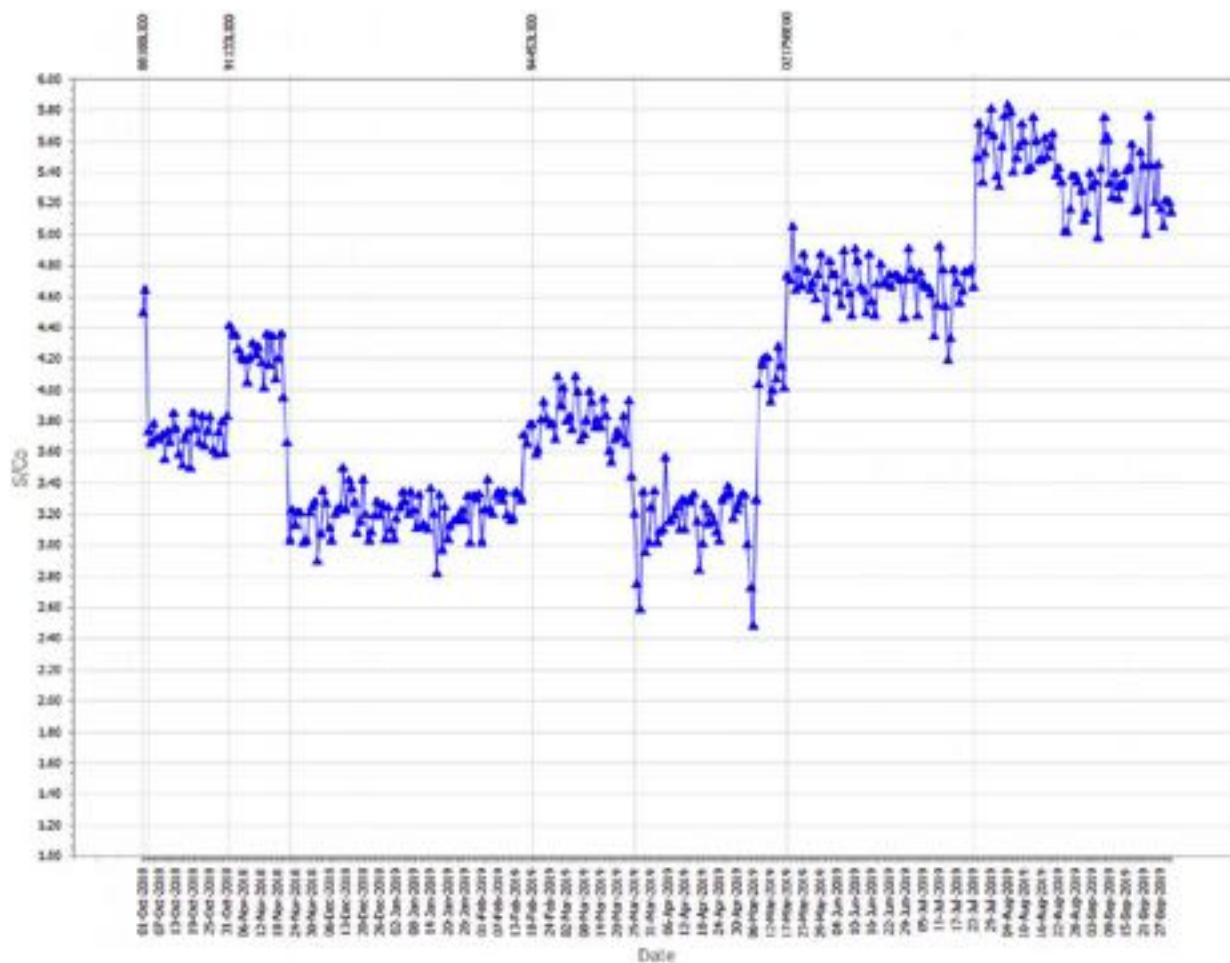


Fig. 6 Changes in QC sample reactivity for a combined serological assay detecting both anti-HIV antibodies and HIV p24 antigen. The changes are a consequence of changes in reagent lots. Reagent lots are indicated at the top of the figure. The vertical axis represents the signal to cut-off values of the QC sample. The horizontal axis represents the date of testing.

method considered validated. Validation includes analysis of performance characteristics such as analytical sensitivities (e.g., limit of detection, limit and range of quantification determined using reference materials where available), specificities as well as accuracy (closeness to an accepted value) and precision (how repeatable/reproducible a measurement is). Examples of accuracy and precision of measurements are shown in [Fig. 7](#) demonstrating good to more variable and poor performance.

Documented evidence of staff training and competence, instrumentation commissioning, maintenance and monitoring, and standardised procedures need to be in place. Accredited laboratories are required to identify, record and investigate any deficiencies detected and undertake a root cause analysis of the issues and implement corrective actions, and demonstrate the effectiveness of the action taken. All suppliers and subcontractors are managed under the QMS. All accredited laboratories undertake self-assessment in the form of internal audits and are subjected to external audits by their accrediting body and other stakeholders periodically.

Similar regulatory requirements are in place in other jurisdictions such as the USA Clinical Laboratory Improvement Amendments (CLIA) regulations or local State-based regulations. In some parts of the world, for example Africa, there are frameworks in place to enable laboratories to work towards accreditation, these include Strengthening Laboratory Management Toward Accreditation (SLMTA) and Stepwise Laboratory Improvement Process Towards Accreditation (SLIPTA).

Commercial and Laboratory-Developed tests

The availability of commercial assays which have been granted regulatory approval, such as clearance by the US Food and Drug Administration or EC declaration of conformity (i.e., have a CE-mark) in Europe, attesting to a certain level of quality have helped improve standardization in the diagnostic virology laboratory. This can be seen in the results from EQA studies where more

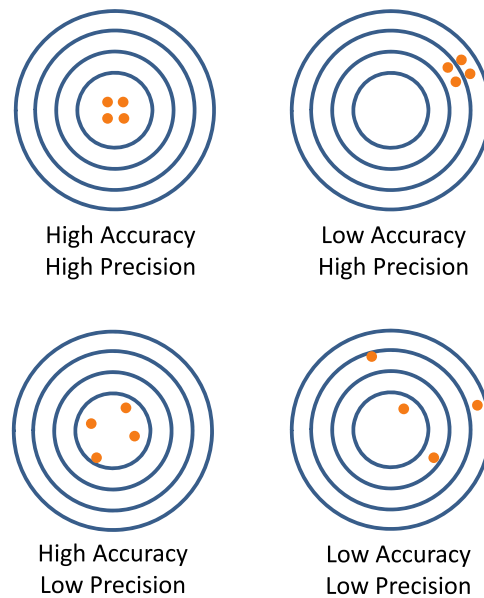


Fig. 7 Precision and accuracy of measurement procedures. The upper left panel demonstrates the ideal situation where a measurement is both accurate and precise – hitting the target directly in the middle. In the case of the upper right panel, although the measurements are very precise they are off target i.e., they are not accurate. In the bottom left panel, the measurements are accurate, but they are not precise. Finally, in the lower right panel, the measurements are scattered being neither precise nor accurate.

consistent results are reported following the introduction of commercial assays over time. This is especially true once assays are calibrated in IU. Laboratory developed tests (LDTs) need to undergo extensive verification and validation before they are brought into routine use. IS are important for the calibration of LDTs and in the preparation of calibrated run controls, traceable to the IS, used to ensure consistent assay performance.

Fig. 8 illustrates the steps involved in the introduction of an assay from initial evaluation and verification, through validation and routine clinical testing. The interplay between QC and EQA procedures ensure adequate test performance and identify out of specification test results which lead to quality improvements.

Virus Isolation

Although used less frequently nowadays, virus isolation or culture is still in use in some diagnostic virology laboratories. The origins of the cell lines and tissue cultures used in isolation procedures should be defined to ensure identity and traceability as well as passage number. For example, continuous cell lines can be obtained from organizations such as the American Type Culture Collection and the European Collection of Authenticated Cell Cultures. The virus detection limit is an important validation parameter; however, this is likely to be specific for each indicator cell line and virus and may well vary for different strains. It is important to verify that the test matrix does not interfere with virus detection, and pre-dilution of samples may be necessary e.g., for stool or urine specimens; appropriate interference and negative controls should be included during isolation procedures. Virus isolation exercises may be offered by some EQA providers.

The Use of International Standards in Clinical Practice

The standardization of NAT testing is important in clinical practice, not only for disease diagnosis, but also in decisions concerning when to implement treatment with antiviral drugs and how long treatment should continue.

The first WHO IS developed for NAT testing was for hepatitis C virus (HCV). While this standard was established in 1997 to enable harmonized regulations for screening of plasma donations for blood product manufacturing, it soon found utility in clinical virology laboratories. In the European Association for the Study of the Liver (EASL) Recommendations on Treatment of Hepatitis C 2018, it is stated that HCV RNA assessment (in both acute and chronic cases) should be made by reliable, sensitive assays, and HCV RNA levels should be expressed in IU/ml using assays with a lower limit of detection ≤ 15 IU/ml. However, the vast majority of patients with an indication for anti-HCV therapy have an HCV RNA level above 50,000 IU/ml. For the treatment of acute hepatitis C in low- or middle-income countries and in specific settings in high-income countries, a qualitative HCV RNA assay with a lower limit of detection $\leq 1,000$ IU/ml can be used to provide broad affordable access to HCV diagnosis and care. The

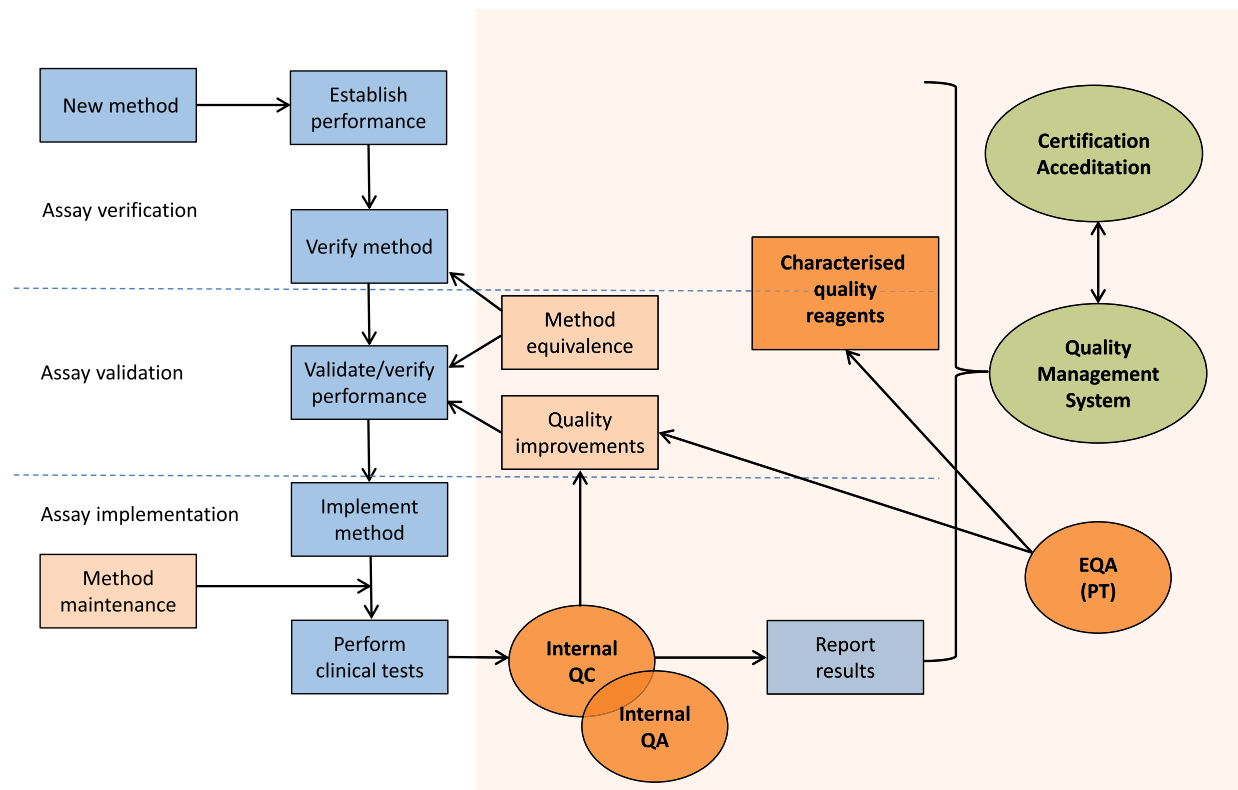


Fig. 8 Assay implementation in a diagnostic laboratory, from initial evaluation and verification, through validation and routine clinical testing. QC and EQA procedures ensure adequate test performance and identify out of specification test results which lead to quality improvements. Reproduced in modified form with permission by Professor H. Niesters.

EASL recommendations include further parameters including lower detection limits (≤ 15 IU/ml) indicating patient recovery following antiviral therapy, for example.

Patients with chronic HBV infection are at increased risk of progression to cirrhosis and hepatocellular carcinoma (HCC). In the EASL 2017 Clinical Practice Guidelines on the management of HBV infection, HBV DNA levels exceeding 2,000 IU/ml, elevated alanine transaminase (ALT) and/or at least moderate histological lesions are an indication for treatment together with all cirrhotic patients with detectable HBV DNA. In patients where the levels of HBV DNA exceed 20,000 IU/ml, and ALT levels exceed $2 \times$ the upper limit of normal, treatment should commence regardless of the degree of fibrosis.

Human cytomegalovirus (HCMV) is a major pathogen for immunocompromised patients, particular transplant recipients. The use of HCMV NAT is important for surveillance to guide antiviral treatment, and for therapeutic monitoring. In the "Third International Consensus Guidelines on the Management of Cytomegalovirus in Solid-organ Transplantation", reporting levels of HCMV DNA by NAT testing in IU/ml is recommended. In the future, it is hoped that it will be possible to define agreed thresholds to initiate pre-emptive therapy.

There are several ISs for infectious disease antigens, such as HBV surface antigen (HBsAg), HIV-1 p24 as well as HCV core antigen. These have been used to calibrate immunoassays and they represent a cost-effective way of monitoring in infected patients undergoing therapy compared to NAT. However, quantitative results of antigen tests are used less frequently nowadays in clinical decision making, having been superseded by viral load testing by NAT. However, HBsAg testing is very widely used in screening blood donors to test and exclude donors with HBV infection.

The level of antibody in a patient is not often used in the clinical setting. Qualitative testing for antibodies uses various processes to detect the presence of the immune reaction; however, a rise in the concentration of antibodies may be useful in confirming a recent infection. Furthermore, detection of antibody classes may provide insight into infection course (e.g., IgM in recent infection). Determination of antibody function may provide relevant information on immunity (e.g., neutralisation assays and protection against potential re-infection). Sometimes levels of immunity in patients are determined by assays calibrated with a WHO IS and the immunity threshold defined in IU/ml. However, this approach is sometimes questionable and standardization using a single reference material has proved challenging in certain circumstances, for example, with assays designed to measure antibodies against a single antigen or neutralization of virus infectivity. Several of the antibody standards that are available have been prepared to set a specification for therapeutic immunoglobulins and may not be the most suitable materials for standardization in the clinical laboratory setting.

Conclusions

The standardization of diagnostics in the clinical virology laboratory is driven by a combined approach using physical reference materials as well as written standards and adherence to quality assurance procedures. The WHO ECBS ensures that standards suitable for their purpose and will be replaced in a timely manner with the assurance of the continuity of the unitage. Discussions with the clinical virology community are essential to identify where there is a diagnostic need for new standards.

Further Reading

- Baylis, S.A., Wallace, P., McCulloch, E., Niesters, H.G.M., Nübling, C.M., 2019. Standardization of nucleic acid tests: The approach of the World Health Organization. *Journal of Clinical Microbiology* 57. (pii: e01056-18).
- Holden, M.J., Madej, R.M., Minor, P., Kalman, L.V., 2011. Molecular diagnostics: Harmonization through reference materials, documentary standards and proficiency testing. *Expert Review of Molecular Diagnostics* 11 (7), 741–755. doi:10.1586/erm.11.50.
- International Conference on Harmonization. 1994. Validation of Analytical Procedures: Text and Methodology (CPMP/ICH/381/95, ICHQ2[R1]).
- Miller, G.W., Myers, G.L., Lou Gantzer, M., *et al.*, 2011. Roadmap for harmonization of clinical laboratory measurement procedures. *Clinical Chemistry* 57, 1108–1117.
- Wallace, P., 2021. Quality assurance in the clinical virology laboratory. In: *Encyclopedia of Virology*, 4th ed.
- World Health Organization, 2011. Laboratory Quality Management System. Available at: https://www.who.int/ihr/publications/lqms_en.pdf.
- World Health Organization, 2006. Recommendations for the Preparation, Characterization and Establishment of International and Other Biological Reference Standards (Revised 2004). Technical W.H.O. Report Series no. 932, pp. 73–131. Available at: http://www.who.int/immunization_standards/vaccine_reference_preparations/TRS932Annex%202_Inter%20_biol%20ef%20standards%20rev2004.pdf.
- World Health Organization. 2017. WHO Manual for the Preparation of Secondary Reference Materials for *In Vitro* diagnostic Assays Designed for Infectious Disease Nucleic Acid or Antigen Detection: Calibration to WHO International Standards. World Health Organisation Technical Report Series no. 1004, pp. 389–455. Available at: http://www.who.int/bloodproducts/norms/SecStandManWHO_TRS_1004_web_Annex_6.pdf?ua_1.

Relevant Websites

- <https://www.fda.gov/medical-devices/ivd-regulatory-assistance/overview-ivd-regulation>
Overview of IVD Regulation.
- <https://www.nrlquality.org.au/qc-limits>
QC Limits.
- https://www.who.int/biologicals/technical_report_series/en/
Technical Report Series (TRS).

Quality Assurance in the Clinical Virology Laboratory

Paul Wallace and Elaine McCulloch, Quality Control for Molecular Diagnostics (QCMD), Glasgow, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

AMR Analytical measurement range.

Assay Drift.

LoD Limit of detection.

LoQ Limit of quantitation.

Probit Analysis.

Reportable Range.

The Principles of Quality Assurance and Quality Management

The principles of quality assurance (QA) and quality management (QM) have their origins in manufacturing and in particular the automotive industry. They were primarily used to support the stream-lining of processes and procedures, with the aim of increasing productivity, while reducing costs but maintaining the quality of the product delivered to the customer. Within the clinical laboratory, the concepts of QA and QM were first introduced in hematology and clinical chemistry by Belk and Sunderman in the late 1940s/early 1950s in an effort help reduce the high number of diagnostic errors observed in relation to the handling and testing of patient specimens which ultimately could have an impact on patient care. These concepts and general principles formed the QA and QM foundation commonly used within the clinical virology laboratory which continue to evolve in order to meet present regulatory demands and assure the safety and effectiveness of the virology testing service.

The key components of quality assurance within a laboratory's quality management system (QMS) are shown diagrammatically in [Fig. 1](#).

As well as supporting the whole diagnostic process (pre- analytical to post analytical) to delivery of patient result, the QMS also support the verification, validation and implementation phases of the diagnostic assays used within the virology laboratory.

With the implementation of quality principles from other industry sectors comes the adoption of the associated terms and definitions used and one such quality term which has found its way into clinical virology in recent years is Total Quality Management (TQM). In the context of the clinical virology laboratory, TQM provides an alternative description for the integration of all quality parameters and the nurturing of a quality culture followed by all laboratory staff. So that the aim of the laboratory is not only to provide correct results, they also aim to ensure that the correct test is performed on the appropriate specimen, the correct result is obtained and interpreted, and provided to the right patient within a meaningful time frame and using the correct procedures. The goal of the laboratory is, to ensure confidentiality, the safety of the patient, as well as provide a mechanism for continuously monitoring and improving of the diagnostic service.

In many countries, clinical virology laboratories and the testing services they provide are required to be accredited and in some regions accreditation is a pre-requisite to regional reimbursement policy. Accreditation also helps the clinical laboratory demonstrate the competence and reliability of its services to its immediate peers within the healthcare environment, the Clinicians who routinely utilize the testing service, and those external to the healthcare environment such as the legal profession. Increasingly, clinical virology laboratories are gaining accreditation to the internationally recognized standard ISO15189 which was developed specifically for the medical laboratories involved in laboratory testing and examination, although in some country's laboratories are accredited to the national regulatory framework. The quality management requirements of ISO15189 are also aligned to those in ISO 9001 which provides a basic standard and quality language across diverse industry sectors.

Quality Management System (QMS)

In the context of the clinical virology laboratory, the QMS covers all the policies, documented processes, procedures, and records used in order to deliver the diagnostic testing service to the patient under its defined scope of accreditation. This also incorporates the operational environment in which the testing is performed, the equipment, specialist reagents and essential supplies required, as well as the qualification and competence of the staff tasked with conducting the testing service. The complexity of the quality management system depends on the scope of the service and in many larger Hospitals the QMS is managed centrally and covers the whole of the pathology services, Hematology, Blood Transfusion, Biochemistry, Microbiology including virology.

Quality Management Documentation

The documentation used within the QMS is usually organized and managed through a clear document hierarchy. The top-level documents include the Quality Manual or Services Manual which stipulate the laboratory's quality policy and objectives. These are monitored against defined quality performance indicators, such as the monitoring of the time taken to report a test result from receipt of specimen, and ensure that the service is achieving its stated quality intentions. The Quality Manual also sets out the



Fig. 1 QMS/QA component.

laboratory organization, management roles and responsibilities in relation to the scope of the services provided. This will include the range and type of tests performed, the specimen types handled, the quality control measures and internal quality assurance procedures the laboratory undertakes, as well as the external quality assessment (EQA) schemes it subscribes to. Practical information on how to order the test, referral of tests and contact details for additional clinical and technical advice are also included within the quality manual or service manual.

The procedural documents including standard operating procedures (SOPs), equipment operating procedures (EOPs), describe in detail how a specific laboratory activity is to be conducted or how a piece of equipment is to be used, who is authorized to do so, as well as when and where that activity is to be performed or equipment used. The procedural documents also describe the inputs and outputs of the procedure and where this information should be recorded. In addition to operating procedures some laboratories may also have separate specific work instruction documents. These detail the exact steps and methods to be employed when conducting a laboratory activity such as the consecutive tasks required in order to set up for the detection of a particular viral pathogen or the operation of a specific piece of equipment, for example the calibration and maintenance of a real-time Thermal Cycler. In some clinical virology laboratories, having separate work instructions means that they can be held directly within the laboratory for easy access and reference where the activity is taking place.

The inputs and outputs of a procedure or specific work instruction are recorded within a Standard Operation Record (SOR). These include paper-based records as well as electronic outputs from equipment or devices within the laboratory such as a thermal cycler or serological autoanalyzer. In these cases, the devices are interfaced with the laboratory and the records are managed through a middleware software such as FlowG (See Relevant Websites section) which collects, collates, and stores the data which can include calibration and control performance data across a range of laboratory platforms. The middleware is also linked to the general Laboratory Information Management System (LIMS) which provides secure storage and management of all the laboratory data and patient results.

Document Control

All documents within the clinical virology laboratory QMS need to undergo regular review at the specified intervals on when a particular policy, process, or procedure is changed in order to ensure that they remain fit for purpose and compliant. This is managed through a document control policy which also includes an index of all the documents within the laboratories QMS, who is responsible for each one, and when it is scheduled for review.

Many clinical laboratories have moved away from a paper-based Quality Management and document control systems towards having a commercially available QMS software management solution such as Q-pulse (See Relevant Websites section) which is able to integrate directly into LIMS and makes the management of laboratory quality assurance information more efficient and compliant with regulatory standards such as ISO15189.

Specimen Management

The ability of the clinical virology laboratory to carry out a test and to support a diagnosis depends extensively on the quality of the specimen the laboratory receives and the time it takes to get to the laboratory. Pre-analytical errors such as wrong labeling of the specimen, the incorrect specimen collection device, incorrect specimen type are the most frequently reported sources of error (between 48% and 62%) within the clinical laboratory.

In some hospitals or institutes, it is the responsibility of specific clinical virology laboratory staff to provide training on appropriate sample collection, handling, transportation and storage of the clinical specimens to be used. As a result, the clinical virology laboratory details the quality requirements of the specimen types (i.e., serum, plasma, urine, stool, etc) the laboratory is capable of processing in relation to the test repertoire offered under the services it provides. This includes specific instructions to clinicians, general practitioners, and nursing staff on how and when to take the patient specimen as well as the collection device to be used and the minimum volumes required for the test requested.

For example, most common serological tests are performed on either serum or plasma and are best tested within 48–72 h after collection in order to prevent degradation of the analytes, unless they are stored and shipped to the laboratory at between 2–8°C. In comparison, for the molecular viral load determination of blood borne viruses such as HIV, HCV, and HBV, either EDTA plasma or sodium citrated plasma is preferable to serum as the degradation of nucleic acids has been observed within serum/clotted specimens and may result in the under-reporting of the viral load result. Heparinised specimens are also not recommended for molecular testing as the heparin has been shown to inhibit techniques such as PCR, although in most modern molecular technologies this does not appear to be a problem.

Although all clinical specimens must be regarded as potentially infectious, most laboratories will also have a separate policy for the handling of potentially high-risk samples, such as those suspected of having a viral hemorrhagic fever (VHF). The policy covers the way that these specimens are handled and the labeling to be used. This will depend on risk factors identified and any regional regulatory and health and safety requirements the laboratory must adhere to.

However, the sample selection process has been improved significantly through the implementation of electronic test requisitions which allow authorized healthcare professionals such as general practitioners to request a test directly through an internet-based portal. This has helped make the process more efficient and standardized, as well as reduces both transcription and translation errors associated with manual processing. Many of the electronic test requisition systems also enable the generation of unique specimen-Patient barcodes which also help reduce the incidences of under labeling (i.e., not providing sufficient patient information in order to support a test requisition) or mislabelling as the additional information is a required field when completing the 'on-line' form and in many cases is incorporated within the barcode generated.

Transportation and Storage of Specimens

Transportation and storage can have a significant impact on the quality of the clinical specimen and hence the accuracy and outcome of the laboratory result. For example, sputum specimens should ideally be tested within 2 h of collection and stool samples within 12 h in order to prevent background bacterial flora growth masking the virologic examination. If this is not possible then the specimens can be stored at 4–8°C but only for 48 h. So, it is essential that staff are appropriately trained and follow the correct procedures which are usually defined within the laboratory's policy on the handling and delivery of laboratory specimens and in line with the requirements of any national Health and Safety at work legislation.

The transportation of samples within the hospital or institute where the clinical virology laboratory is located is usually conducted through a trained internal porter service using containers dedicated to the movement of potentially biohazardous specimens. An approved specialist courier service will usually be contracted in order to transport specimens from external sites such as clinics, General Practitioner sites to the clinical virology laboratory. Transportation of specimens by road must comply with the carriage of dangerous goods regulations within that region (See Relevant Websites section), which state the specific packaging conditions that need to be followed dependant on the risk the viral pathogen poses. In general, most specimens such as blood, excreta, secreta, for diagnostic purposes will be transported under the category B classification unless they contain live virus cultures that are known to be life threatening to humans, under these circumstances the specimens should be transported under category A. Exempt patient specimens are those with a minimal likelihood that a viral pathogen is present and include blood transfusion products, and most dried blood spot applications.

Where specimens need to be transported by air, the regulations of the International Air Transport Association (IATA) regulations are followed. In addition, where viral specimens are being transported from one country to another for reference testing, assay verification, or quality assurance purposes, etc it is important to verify whether the pathogen is regulated under any regional Human Pathogens and Toxins Regulations which could restrict its exportation and importation. Regulations vary from

country to country dependant the apparent risk associated with the activities performed and also take into consideration whether the virus is a security sensitive biological agent.

Laboratory Specimen Retention Times

Specimen retention defines the length of time clinical specimens will be retained by a clinical virology laboratory. Retained specimens are extremely important to the laboratory for confirmatory testing where required, infection control, and public health investigations, as well as for quality control and new assay evaluation. The specimen retention time will depend on the type and origin of the specimen, the needs of the patient, and also the storage capacity and capability of the laboratory. Monitoring retained specimens with regards to events such as the number of freeze/thaw cycles, viral load over time in storage, is important in assuring the quality of the specimens, particularly if they are going to be used for quality control purposes. Many clinical virology laboratories will utilize software options in order to create an inventory to track the use of important retained clinical specimens.

Personnel Training, Competency, and Continuous Professional Development (CPD)

The clinical virology laboratory operates within a highly regulated environment and in many countries the laboratories can only perform their testing service under the appropriate legal licenses.

An important aspect of human resource management particularly within the clinical virology laboratory is to ensure that all staff have defined roles and responsibilities, and they are clear on what is expected of them within the daily operation of the clinical laboratory. The structure and organization of the clinical virology human resource is dependant on the size of the laboratory, the range and complexity of the testing service provided, the volume of annual clinical specimens, and whether it is a separate independent entity or it is affiliated to a hospital or institute as part of the total pathology services provided. An organogram for the clinical virology laboratory with supporting narrative outlining the structure and reporting relationships is usually detailed within the Quality Manual and is an essential quality assurance requirement.

The head of the laboratory (laboratory Director) will usually be a licensed Medical Virologist either with a Medical Doctorate (MD) or a Senior Clinical Scientist with a MSc, PhD, or equivalent professional qualification specialising in virology. They must ensure that the laboratory has sufficiently trained and competent staff in order to provide the scope of virology testing services specified to the appropriate regulatory standards within their jurisdiction, and where required, in line with their accreditation for example to ISO15189. The laboratory director is usually supported by a consultant clinical medical microbiologist and/or virologist responsible for the overall clinical practice and decision-making process. In line with the requirements of International Standards for Laboratory Accreditation, most laboratories will have a dedicated quality manager and nominated deputy this helps avoid potential conflicts of interest between roles as the quality manager does not have any laboratory operational responsibilities and can therefore assess and monitor the quality aspects of the laboratory impartially. Larger organizations may have a separate quality manager responsible for the whole of the pathology services provided by the hospital. The advantage of this is that they are independent of the clinical virology laboratory and can inspect and advise on quality issues. The disadvantage is they may not be familiar with the technicalities of the virology testing service offered which can potentially lead to misunderstandings.

Under the direction of the laboratory director, the clinical virology laboratory will consist of technical supervisors (or technical section leaders) who lead specific laboratory sections such as the molecular, serology, or viral isolation & culture section and are responsible for the daily management of scientific and technical staff within their respective sections. The scientific and technical staff will be responsible for specimen processing, carry out the various testing procedures and other laboratory activities such as equipment maintenance and calibration. The laboratory will also have general non-laboratory based clerical support staff performing administrative tasks including medical record keeping. Other laboratory support activities such as stock management, the procurement of laboratory supplies, and diagnostic kits would may be the responsibility of a dedicated laboratory manager or may be shared responsibilities in smaller organizations or hospitals.

The human resource (HR) files for each staff member usually contains the most up to date job description, curriculum vitae, qualifications including copies of any certificates/diplomas, and laboratory licenses in regions where this is required. In addition to this training records, and details of any continuous education programs/events the staff member has undertaken are also included within the HR or personnel file, along with any relevant publications.

Each laboratory employee should be provided with a clear training plan. This should cover the preliminary laboratory training requirements, initial duties, and responsibilities, for when the individual joins the laboratory. This should be assessed regularly and if necessary, amended in order to accurately reflect the current status of the individual followed by a further personal training and development plan which supports the employee's continued professional development (CPD) requirements in line with the objectives of the clinical virology laboratory.

The effectiveness of training and the competency of employees must also be reviewed regularly in accordance with the duties the employee is expected to perform. This can be achieved through practical laboratory assessments where the test results obtained by the employee on known QC samples or residual material from an external quality assessment cycle, are introduced into the routine workload and used as a way of checking and monitoring the employee's ability over a defined period of time. Other aspects include monitoring the adherence to laboratory policies and procedures, observing performance when carrying out instrument maintenance and calibration checks, as well as the evaluation of problem-solving and data analysis skills through set

technical questionnaires. The outcome of any competency assessment exercises are recorded within the employees training matrix and competency assessment records. If a performance issue is identified further training may be required followed by a reassessment of the employees' competency. Staff training and competency assessment are key elements within a laboratory's accreditation and some regulatory organizations such as the College of American Pathologists (CAP) in the United States specify the frequency that competency assessments must take place in relation to retraining requirements and test complexity. For example, for employees within the first year of their laboratory duties, competency must be assessed at least twice a year when engaged in moderate to high complexity testing (i.e., molecular testing).

Most clinical laboratories will also have an employee handbook which is provided to each staff member when they commence work within the clinical laboratory, this usually includes guidance to the employee on human resource policies such as performance reviews, absence and holiday policies, information technology and security, as well as health and safety aspects related to the clinical virology laboratory working environment.

Laboratory Facilities and Equipment

The quality of the test results a clinical virology laboratory provides are largely dependant on its facilities, organizational set-up, and equipment used. In general, the laboratory will be organized into functional areas designed to provide a safe and effective environment that ensures the delivery of the testing services to the appropriate quality and regulatory standards.

For example, most clinical virology laboratories will have separate functional areas along the following lines:

- specimen collection and preparation area,
- general laboratory area, stores, plant room, etc.,
- molecular virology,
- serology,
- virus culture and isolation unit (Note: in some specialist virology laboratories this could be further split dependant on the type and risk level of the viral pathogen),
- non-laboratory office area.

Each area will have its own set of operational specifications which may include temperature and cleanroom air-handling conditions. These are monitored as part of the laboratories QA measures against the specifications and defined control limits. In many laboratories the monitoring is 24/7 by way of an integrated electronic system which raises an alarm directly to laboratory management 24/7 in "real-time" should a control limit be broken, who can then take immediate action. In addition, the separation of functional activities and equipment helps prevent cross-contamination which could eventually lead to erroneous test results.

Particular attention is paid to the set-up and management of the molecular virology area as molecular amplification techniques such as PCR are extremely sensitive as well as pose a significant risk of introducing amplicon-based contamination if they are not properly controlled. In order to minimize the risk, most clinical virology laboratories will use rooms with a unidirectional workflow with relative air pressures with which to separate and manage the Pre-molecular amplification activities from the Post-molecular amplification activities. The number of separate rooms or areas the laboratory implement largely depends on the facilities and space available.

Four designated areas or rooms provide the laboratory with the ability to separate the molecular activities into the reagent preparation area, the specimen extraction/processing area, the assay set-up room (template addition stage), and the molecular amplification room.

Each room has its own equipment such as biological safety hoods, pipettes, centrifuges, fridge and freezers, etc as well as a dedicated consumable/reagent store which is restricted for use within that room. Staff movement follows the direction of the workflow, and returning to previous rooms in the workflow is not allowed. Personal protective equipment (PPE) such as glove and laboratory coats are required for different rooms and is often color coded for easier visual awareness. The management of waste from the molecular facilities should be properly segregated and disposed according to the institute's disposal and infection control procedures. Amplicon contamination has been a serious problem in the past and has resulted in the shutdown of some laboratory facilities. The use of periodic wipe tests helps to monitor the laboratory environment for potential amplicon contamination. The frequency can be adjusted if any contamination is identified and thorough decontamination measures can then be put in place to rapidly address any issues found. It is also standard practice to include no template control (NTC)/negative controls in order check for potential contamination in the reagents, consumables, and general laboratory environment. Monitoring the positivity rate of the NTC over time can act as an indicator of potential contamination issues.

Equipment Qualification, Calibration, Maintenance and Monitoring

From a quality assurance perspective equipment covers instruments and software including laboratory information management systems (LIMS). Regulatory agencies (ISO15189 or equivalent) place a great deal of emphasis on the quality assurance of

equipment from initial purchase through qualification, calibration, maintenance and monitoring, to its eventual disposal at the end of its intended use and life cycle. Not only is equipment QA an important regulatory requirement, as specialist laboratory equipment is expensive the QA procedures employed can also help to increase the equipment lifespan and therefore provide an additional cost benefit to the laboratory.

Equipment Qualification

The clinical virology laboratory will have an equipment validation and/or verification plan and the initial phase of this will be equipment qualification. The laboratory will evaluate the item of equipment in consultation with the equipment manufacturer in order to define both the functional and operational specifications of the equipment as well as the cost and any preliminary installation, calibration, and training requirements. This step is often called design qualification (DQ) and is followed by installation qualification (IQ) in which the laboratory establishes that the equipment is suitable for operation and meets the manufacturers specifications as well as any regulatory claims. Operational qualification (OQ) is the third step in equipment qualification and it involves proving the equipment's performance within the clinical setting usually against a "gold standard" or predicate device. The final phase is performance qualification (PQ) where the performance of the equipment is monitored over time in the clinical setting against a defined set of performance criteria. Within industry this four-step approach has become known as the 4Qs and within clinical virology the steps are generally covered through assay verification and validation.

For the most generic equipment used within the virology laboratory such as pipettes, vortex's and microfuges the recommendations provided by the equipment manufacturer are usually sufficient for qualification for routine use. However, specialist virology equipment used for molecular and serological testing (e.g., thermal cyclers and autoanalyzer's) will require specific qualification for the equipment's intended use within the clinical context. These instruments can be fully integrated platform or cartridge based commercial assay systems which are essentially 'sample in and result out'. Alternatively, they can be component-based workflows where a combination of equipment is required to create a virology workflow which can be used to generate a test result.

This is largely the case in molecular virology, particularly for laboratory developed assays, where the extraction platform, kit, molecular amplification platform, and detection kit are combined in a molecular workflow. The virology laboratory will define the equipment qualification criterion in accordance with the manufacturers' equipment recommendations and the laboratories intended clinical use. The aim being at qualifying the laboratory equipment and monitoring performance to ensure the validity of data/result generated for the individual items of equipment as well as the whole test workflow.

Equipment Calibration

When using laboratory equipment for measurement purposes whether monitoring viral load or the temperature of the laboratory environment maintaining the quality assurance of the equipment through calibration is essential. Calibration minimizes measurement uncertainty. It also helps establish operational performance criteria which is used to monitor and control errors by ensuring that measurements remain both appropriate and acceptable.

Within the clinical virology laboratory, equipment calibration covers both the fixed plant and machinery used to control the different working environments within the clinical virology laboratory as well as the specialist instrumentation used within each of the designated laboratory areas.

Fixed plant and machinery includes the laboratory's heating, ventilation, and air conditioning (HVAC) systems which are designed to contain any potentially infectious material dependant on the biosafety classification level (1–4, with 4 being the highest risk based on the type of virus and work to be undertaken) as well as any fitted walk-in cold-rooms and specialist culture facilities. The calibration of specialist instrumentation including biosafety hoods, – 80°C freezers, thermal cyclers, autopipettes, etc. assures the user that the instrument is working within the manufacturers defined specifications. The calibration of pipettes can readily be performed by the laboratory themselves using gravimetric calibration which is an accepted standard for pipette calibration. However, for most accredited laboratories the calibration of fixed plant and machinery and some specialist instruments (such as thermal cyclers) is usually conducted by the equipment manufacturer or an approved external contractor to an international standard such as ISO17025 or equivalent. If a laboratory decides to change the manufacturers calibration requirements this could result in the manufacturer's warrant for use being invalidated. It is therefore important that the laboratory discusses and agrees any proposed changes with the manufacturer before implementing them.

The frequency of calibration depends on the specific requirements of the equipment, maintenance and safety record, annual service plan, as well as the extent of its use. Calibration events are recorded within a "equipment/calibration log" which identifies the equipment by its assigned asset number.

Equipment Maintenance and Monitoring

The ongoing calibration, service, and preventive maintenance of laboratory equipment throughout its lifespan supports reliability and confidence in the performance of the equipment. Maintenance and monitoring should include a register of all equipment indicating serial numbers, assigned asset numbers as well a location within the laboratory. This is particularly important for

equipment that is designated for use within a single laboratory location. Within the equipment log, details of the validation and outcome of any safety inspections are recorded along with any equipment failures and servicing events.

Most laboratories will assign the responsibility for specific laboratory equipment to an operator or group of operators. It is their job to keep the equipment in good working order and to ensure that calibration and maintenance events are up to date as well as report any reoccurring problems to laboratory management.

Routine preventive maintenance ensures the equipment warranty in line with the manufacturer's recommendations and specified annual schedule. For example, the air filters within the HVAC system should be monitored and cleaned regularly and failure to do this could lead to potential laboratory contamination issues or safety issues if the filters become blocked and cause overheating. In addition, this can also lead to mechanical issues which shortened the equipment life span and become an additional cost burden to the laboratory.

Most of the environmental monitoring within the laboratory (such as temperature and pressure) is managed through wireless monitoring systems. These systems consist of operating software, specific probes/sensors, and receivers/transmitters for picking up the data and enable real-time performance monitoring throughout the different laboratory environments. If the central power fails, the equipment is covered by a back-up power system which ensures the integrity of the laboratory environments is maintained and monitored 24/7. The metrics from the wireless monitoring system usually have operational tolerances set which provide an automatic alarm which is set electronically to the designated operator when the metric is broken. This enables the laboratory to take immediate action as well as helps with the identification of trends over time.

Internal and External Audits

Internal and external audits are an essential part of the clinical virology laboratory's quality assurance armory. The purpose of the internal audit is to provide an objective overview of the laboratory's control and functional effectiveness over its policies, processes, and procedures in relation to the testing service it provides and the regulatory standard and environment it operates within. The quality manager in consultation with the laboratory management team will prepare an annual internal audit plan which sets out the areas to be audited throughout the year in relation to the relevant clauses of the standard to which the laboratory is accredited, the date and time of the audit, and those responsible for each audit area. The depth, breadth and scope of each internal audit will vary depending on the technical complexity of the area, any findings from previous audit reports, as well as any open quality activities such as corrective actions and preventative actions (CAPA). Internal audits should only be conducted by those staff members who are trained and shown to be competent auditors. In addition, those staff members being audited need to ensure that they are fully prepared prior to the audit with a full understanding of the policies and procedures they are responsible for and a list of items that could be reviewed during the audit. This helps improve the efficiency of the audit and helps ensure that internal audits are completed within the allotted timeframe. The audit should be non-directive and allow the auditees responsible for their designated areas within the laboratory to explain their processes and procedures. The auditor is then able to gain an understanding of the staff competence, identify areas that may require improvements such as additional training, as well as potential risks or non-conformances within the written policy and procedures that could ultimately lead to a non-compliance with the regulatory standard. A documented plan can then be agreed with management outlining how any findings are to be addressed before any scheduled external regulatory audit takes place.

The external audit is an independently conducted inspection usually by a national accreditation body (NAB) which assesses the laboratory's technical competence and conformity to applicable international standard such as ISO15189 or ISO17025. The audits typically follow a four-year cycle with an annual surveillance audit over a three-year period and a re-accreditation audit every fourth year. Many countries have officially appointed National Accreditation Bodies that carry out conformity assessment and accreditation activities in particular for calibration and testing. The NAB's are subject to oversight through regional co-operative bodies such as the EA in Europe, APAC in the Asia-Pacific, IAAC in the Americas, AFRAC in Africa, SADCA in Southern Africa, and ARAC in the Arabian region, all under the umbrella of the International Accreditation Forum (IAF) which aims to promote multilateral recognition arrangements (MLA) and common recognition and harmonization of the accreditation standards and practice across geographical regions. The International Laboratory Accreditation Cooperation (ILAC) supports the interests of those accreditation bodies that deal specifically with laboratory testing (ISO17025, ISO15189), measurement and calibration (ISO17511) as well as EQA or PT provider accreditation (ISO17043).

Assay Verification and Validation

The implementation of a new assay method or the amendment of an existing assay method needs to be supported by an appropriate verification and/or validation study prior to routine use within the clinical virology laboratory. Assay verification is the process of testing and reviewing an assay's performance in relation to its known or reported performance such as against the assay manufacturers' defined performance specifications as they are stated within the package insert provided. Whereas, assay validation is the process of evaluating the assay method, establishing its performance characteristics and determining its fitness for use within the clinical virology laboratory. The verification and validation requirements depend on the type of assay, its intended use, and the amount of performance data available such as from previous clinical studies. There are numerous national and International regulatory guidelines which outline the verification, validation process as in many countries this is governed by the regulatory

Table 1 Key verification and validation parameters

Assay parameter	Description/Definition	EQA monitoring	IQC monitoring
Accuracy	The closeness between values obtained from a series of test results. Where possible within the assay verification phase the test results are compared to “known” and/or an accepted reference value	+	+
		Based on consensus	“run to run”
Precision/imprecision	The ability to consistently produce the same result for a given test sample	+	+
	<i>Repeatability:</i> Repeat testing under unchanged conditions for example within a single assay run (inter-assay) <i>Reproducibility:</i> Repeat testing of the same test sample in different runs across a time period, possibly using different operators (intra-assay)	Duplicate sample/lab peer review	Inter-assay (Intra-assay)
Linear/ Reportable range	This is the ability of the assay to obtain test results which are directly proportional to the concentration of the analyte within a given range	+	+
	The reportable range (also <i>Analytical Measurement Range – AMR</i>) is the interval between the upper and lower concentrations of an analyte for which the assay has acceptable levels of <i>precision</i> , <i>accuracy</i> and <i>linearity</i>	Paired/samples	Only if more than 1 control used
The limit of detection (LOD)	The lowest quantity of a substance that can be distinguished from the absence of that substance (a blank value) within a stated confidence limit. The limit of detection can also be determined using a serial dilution of prequalified standard or reference material spiked into an appropriate clinical matrix. The dilutions are replicate tested and the results analyzed using PROBIT analysis	+	+
		“Educational” samples	
Limit of quantification (LOQ)	The lowest concentration at which the analyte can be reliably detected within established imprecision (and where appropriate bias) criterion	+	+
		“Educational” samples	95% CI of + ve result
Sensitivity (analytical and clinical)	The lowest amount of target nucleic acid that still returns a true positive result.	+	+
	<i>Qualitative assays:</i> The number of true positive samples correctly identified by the assay divided by the total number of true positive samples <i>Quantitative assays:</i> Defined from the limit of detection and where appropriate confidence interval (CI) can be calculated in order to support the chance of obtaining a true positive result	Clinically relevant/ educational samples	
Specificity (analytical and clinical)	The ability to return a “true negative” result when the target nucleic acid is not present	+	+
	Consider any potential interfering or cross-reacting substances that may be present in the specimen or matrix	Clinically relevant/ educational samples	If applicable

requirement within that country and the regulatory category of the assay method. However, the general parameters covered are the same and include accuracy, precision, reportable range, linearity (for quantitative assays), reference interval, analytical sensitivity, and analytical specificity as outlined in [Table 1](#).

Where a clinical virology laboratory uses a commercially developed assay such as one regulated and approved by the Food Drug Administration (FDA) in the US or a CE marked assay in Europe, the manufacturer is responsible for providing the clinical laboratory with the analytical performance characteristics as well as its expected clinical performance and supporting clinical data. The clinical laboratory verifies the assay performance against the manufacturer's performance claims for the its intended use as defined within the instruction manual provided with the assay. Importantly, if the laboratory modifies the commercial assay and/or instructions for intended use, the assay is no longer considered a regulatory approved assay and the laboratory would be expected to fully validate the impact of these changes on the assay's performance. It is important to remember that any changes to the manufacturer's assay or protocol without prior approval could result in the manufacturer withdrawing its support for the assay. In comparison, where the assay has been developed in house by the clinical virology laboratory often referred to as a Laboratory Developed Test (LDT), the process of validating and establishing the performance of a laboratory developed assay would be more extensive and will also include comparative testing of the new test against a previous assay for the same targets as well as reference to clinical performance within the current literature if available.

In both cases, commercial assay or LDT there is a requirement for well characterized reference material and, where available, International standards in order to support the verification and validation process as well as on-going Quality Control (see also section on Internal Quality Control Procedures and Monitoring Assay Performance Over Time). The patient samples, controls, or reference material used for verification and validation should be in the appropriate matrix and at clinically appropriate levels (if known). Where there is limited reference material available or a lack of suitably characterized patient material, the laboratory can use residual material from proficiency testing (PT) or external quality assessment (EQA) schemes. The EQA materials are generally highly characterized as well as have the advantage of comparative peer group data which can help facilitate the verification and validation process. It is also good practice to use the same assay reagent lots when conducting the verification and validation exercise in order to reduce the potential impact of variation introduced through different lots/batches of assay reagents. Finally, before undertaking any verification or validation study it is important that the laboratory ensures that the assay method is calibrated to the manufacturers requirements and there are no recalibration/maintenance events scheduled for the duration of the study.

For the verification of a regulatory approved qualitative assay, the following approach and principles would apply.

Accuracy

In order to verify accuracy a comparison of methods study can be carried out. Known patient samples which cover the assay range, and have previously been tested on the current assay method the clinical virology laboratory uses, are tested on the proposed new method over a defined period of time and the results compared. If insufficient in-house patient samples are available, the laboratory may obtain specimens from other laboratories. However, when using patient samples, it is important to ensure that the specimens have been treated the same prior to testing (i.e., storage conditions, number of freeze-thaw cycles etc). An alternative is to use commercially available control or characterized reference material containing the target of interest across the assay range. The number of samples to be tested and the frequency of testing depends on the type, complexity, and intended use of the assay method. Although some regulatory authorities do provide minimum requirements or recommendations. For example, 20 patient specimens over 2–5 testing events. The qualitative data (positive, negative) obtained is analyzed statistically using Cohen's κ analysis.

Precision

The precision of the assay method would be determined by testing one or two controls at clinically relevant levels (if known) repeatedly over a specified period of time in order to establish the between run precision. The recommended minimum time period for testing is 20 days in many guidelines. However, the testing interval should be sufficient to cover the routine operation within the laboratory, such as rotation of operators. The within-run precision can also be established by running duplicate control samples over a specified time period, usually 10 days. The results are analyzed using standard deviation and coefficient of variance and compared to the claims of the assay manufacturer or the results obtained for the current laboratory method being used.

Reportable range and limit of detection (LoD)

The Reportable range of the assay is verified using known characterized positive clinical samples at clinically relevant levels or by using dilutions of commercially available standards or reference materials tested over a period of time in comparison to the current laboratory method. The LoD of the assay is verified by testing replicate samples at known levels above, at and below the determined LoD (e.g., $\text{LoD} \pm 20\%$) or as specified within the manufacturers Instruction manual in duplicate over a defined period of time (usually 20 days). The percentage detected at each level is then calculated and Probit analysis used to estimate the LoD, which can then be defined as a titer of the target analyte detected at a 95% Confidence Interval (CI).

For the verification of a regulatory approved quantitative assay, such as a molecular assay used for the determination of viral load, the same basic principles used for verifying a qualitative assay apply for accuracy and precision although they need to be verified across the analytical measurement range (AMR) of the assay and in particular at known clinically relevant or clinical decision making concentrations.

Linearity/analytical measurement range (AMR)

In the context of a molecular viral load assay, the AMR or reportable range can be defined as the linear range of test values over which the laboratory can accurately detect the target viral nucleic acid within acceptable degrees of variation. Beyond the manufacturers stated limit of

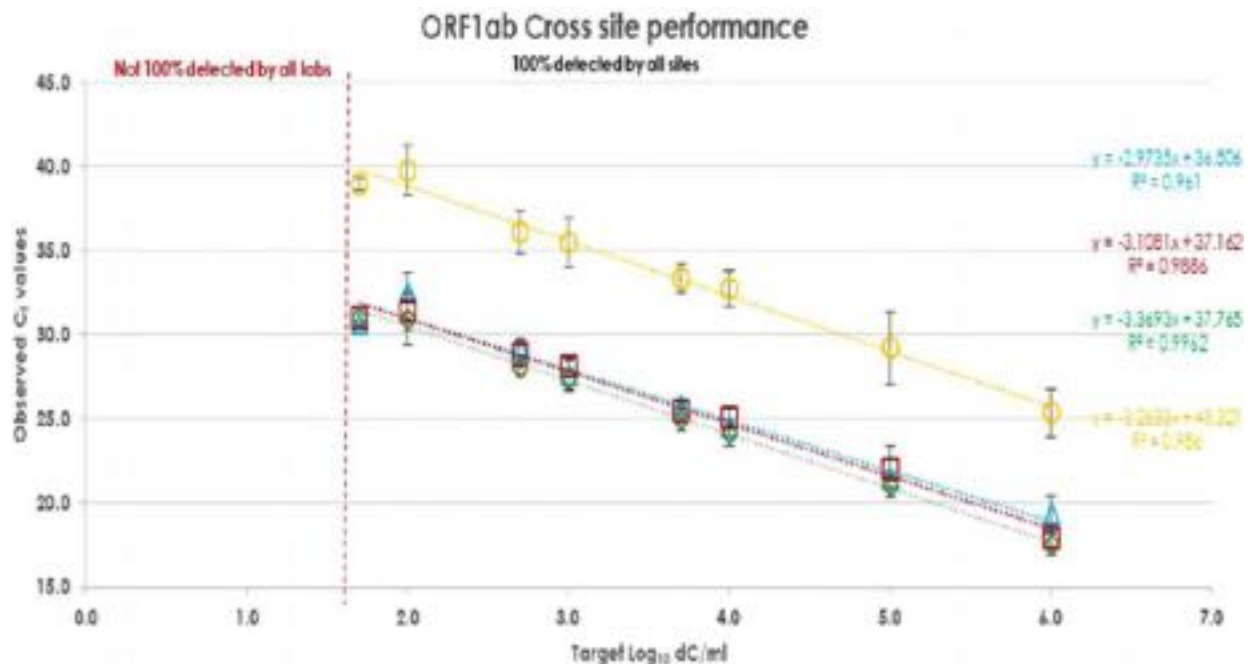


Fig. 2 Graph showing the linear relationship of test results across multiple clinical laboratories for a dilution series of SARS-CoV2.

quantitation (LoQ) the relationship between the measured and actual viral load concentration may not be linear and test results become unreliable. It is therefore important that the laboratory verify the AMR in the clinical setting they intend to use the assay method and in line with the manufacturer's instructions and claims. In some cases, the manufacturer may only have performed an analytical assessment of AMR on contrived or synthesized control materials such as target viral nucleic acid in a plasmid construct and not true clinical samples at relevant titers. As a result, it may not be possible to establish the full analytical AMR equivalence with the manufacturer's specifications. In these circumstances the laboratory should ensure that they assess the AMR within the known clinical context and in line with the laboratory's pre-defined AMR acceptance criteria. At a minimum this should include samples at the low, medium, and high points of the AMR. Polynomial Regression analysis is used to assess the linearity of the assay across all relevant concentrations and if the relationship is found to be non-linear (not first order) those concentrations are removed from the analysis at the high and/or low end and the regression analysis repeated. The testing is usually performed in replicates (2 – 4) at each concentration and precision assessed across the linear range. The highest and lowest concentrations that produce a linear relationship are defined as the upper and lower limits of quantification. (Fig. 2) Where the AMR results are compared against the manufacturer's claims or against the AMR of the comparative method the laboratory is currently using the proportional bias, as measured by the slope of the regression curve, and the constant or systemic bias, as measured by the y-intercept, are determined. A Bland-Altman difference plot is used to measure bias between assays.

Most regulatory standards (such as ISO15189) require the laboratory to have a written policy on verification and validation. This defined the methods, frequency and limits of verification for an assay method. In general, the AMR should be verified in line with the following criteria:

- at changes of reagent lots, unless the laboratory can demonstrate that the use of different lots does not affect the accuracy of patient test results and the range used to report patient test data,
- if QC materials reflect an unusual trend or shift or are outside of the laboratory's acceptable limits, and other means of assessing and correcting unacceptable control values fail to identify and correct the problem,
- after major equipment maintenance or service,
- when recommended by the assay manufacturer.

In addition to this, some regulatory organizations may also specify the frequency of AMR verification, such as every 6 months (e.g., CAP). In most cases the laboratory will only test a smaller sub-set of linearity samples (3 – 4) at the low, mid, and high points within the AMR and only conduct further more extensive testing if these samples fall outside the acceptable criteria.

Sensitivity, Specificity, and Clinical Utility

Key factors that can impact on the accuracy of a test to deliver a result are sensitivity and specificity as well as the prevalence of the disease in the population targeted by the assay method. In the context of a serology assay either targeted at detecting the viral antigen directly or through the indirect detection of antibodies to the viral pathogen, the diagnostic sensitivity and specificity are derived from test results on clinical samples obtained from selected reference patient groups within the population. The degree to

which the reference groups represent all of the host and environmental factors within the total population targeted by the assay method can have a major impact on the accuracy of test result and its interpretation in the clinical setting. Therefore, the clinical virology laboratory also needs to consider this during the verification and validation phase, which is an ongoing process which continues for as long as the assay method is being used within the laboratory and requires constant vigilance, review, and where required reassessment. Specificity also includes any potential interfering or cross-reacting substances that may be present in the specimen or matrix. Interference is basically any non-target organism or substance that can cause a false negative result. Whereas, cross reactivity is a non-target organism or substance that can cause a false positive result. With molecular based assays this can be a problem, particularly when the assays are aimed at differentiating related viral strains. The assay manufacturer or laboratory who developed the assay should have chosen their primers and probes cautiously as well as made an initial "in silico" assessment of the available databases before testing in practice with known samples that are related taxonomically and also epidemiologically. The list of potential cross-reacting and interfering substances is usually documented within the manufacturer's instruction manual. An assessment of analytical specificity should include testing the potential cross-reacting or interfering substance at the known highest concentration that could be expected to be observed within the patient sample. A common approach, in the absence of sufficiently available clinical samples, is to use spike samples in negative matrix and also weakly positive samples (e.g., LoD + 10%) for each of the analytes of interest. From this the minimum inhibitory concentrations can be determined. For quantitative assays any potential cross-reacting targets should be assessed near the Lower Limit of Quantitation (LLOQ) and the quantified target values from spiked and unspiked samples compared and determined to be within the defined precision limits of assay.

Many variables can influence the performance of the assay method, such as the prevalence of the viral disease, the clinical setting, and type of test being performed. Therefore during the clinical evaluation of the assay it is important to consider the testing purpose such as whether it is for diagnosis, screening, or therapeutic monitoring purposes; the environment the testing is likely to take place in such as the central clinical laboratory, out-reach clinic, or point of care setting; the specimen type (e.g., whole blood, plasma, sputum etc); type of result such as quantitative or qualitative result. Two other important measures of assay performance are positive predictive value (PPV), the probability that those testing positive by the test have the viral disease, and the negative predictive value (NPV), the probability that those tested negative by the test are do not have the viral disease. Both PPV and NPV depend on the sensitivity and specificity of the test and also the prevalence of the viral disease within the population which in turn gives the assay its clinical utility. In terms of the number of specimens that should be tested during clinical evaluation in order to assess PPV and NPV, some guidelines recommend 100 negative clinical samples for the evaluation of false positive rates, and 50 clinical samples known to be positive for the target viral pathogen are considered appropriate for determining the false positive rates. The CLSI guidelines (CLSI EP12-A2) recommend a minimum 50 positive and 50 negative clinical specimens. In low prevalent diseases, it can be difficult to obtain sufficient clinical samples to conduct a full clinical evaluation. In these circumstances the clinical laboratory has to consider alternative approaches such as including materials used in previous proficiency testing/EQA scheme challenges.

Internal Quality Assessment (IQA)

"Split sampling" is an effective additional IQA measure for monitoring the whole laboratory testing process. This is where between 0.5%–1.0% of a laboratories clinical patient sample workload are subjected to internal quality assessment. The randomly selected patient samples are split and a new internal testing request generated with a unique laboratory IQA code. This IQA sample is processed through the laboratory in parallel to the original clinical patient sample. At the end of the testing process, results are generated for both original clinical patient sample as well as the IQA sample. These can then be compared internally by the laboratory quality team and any inconsistencies within the results and the report are investigated and corrective action taken through the laboratories quality system. The split sample approach is not always practice or appropriate particularly when the volume of the patient sample is small.

Internal Quality Control Procedures and Monitoring Assay Performance Over Time

Assay variation is the result of systematic and random error which is inherent to the assay method. Common sources of error or variation can be attributed to operator, instrument, reagent lot, calibrator and calibration cycle. The amount and type of variation associated with an assay method is determined during assay verification and/or validation. Having established the extent and where known the clinical relevance of the variation it is important that the laboratory continues to monitor the variation to ensure that it remains within acceptable levels. This is done through Internal Quality Control on a daily run to run basis and is often a requirement of many regulatory bodies (including CLIA and ISO). In general, a commercial assay will be supplied with a set of controls. The intended use of these controls is to monitor the assay in line with the manufacturers defined performance criteria. As such, the manufacturers kit controls are intended for use only with the manufacturers assay and in addition are often specific to a particular kit batch or reagent lot. This can mean that the manufacturers kit controls do not support the monitoring of different reagent lots over time and would not necessarily identify a drift in the overall assay performance over time (assay drift). So, it is important that the operator is aware of the potential limitation of the manufacturers kit controls. For example, many of the controls provided with the manufacturers kit controls only monitor as specific part of the process such as with molecular tests where the manufacturers kit controls may just be free nucleic acid and would therefore only be suitable for controlling the

amplification/detection steps of the assay and not the whole testing process from sample pre-treatment through the extraction phase, amplification and detection, to end result.

Consequently, many regulatory and standards organizations either recommend as “best practice” or in some cases insist on the use of independent third-party controls either instead of, or in addition to those supplied by the assay manufacturer in order to demonstrate compliance. Such recommendations are made at the International level (ISO15189, CLSI) as well as the national level (e.g., RiliBak in Germany). Independent third-party controls are also often referred to as external quality controls or run controls. They aim to monitor the complete testing process and provide an accurate, unbiased measure of quality performance independent of the manufacturers specific control or calibrator.

In general, for qualitative tests a minimum of a positive and negative control is included in each assay as specified in the laboratory procedure. This should also take into consideration the assay manufacturer’s instructions. For multiplex assays covering several different viral targets in a single assay, the laboratory may choose to rotate different positive controls for different targets within the multiplex assay over a given time period at a frequency defined within their laboratory procedure.

For quantitative tests including molecular viral load assays, the laboratory should ensure that the controls it uses are at or near clinically relevant decision levels, where known and/or as specified within the current clinical guidelines. For viral load assays it is common practice to determine performance characteristics for the control under routine laboratory conditions by testing the control over a period of time (usually a minimum of 20 times) in order to obtain sufficient data points to determine the mean and standard deviation from which suitable control limits can be established. Some laboratories calculate a rolling mean based on the last set of observations reported in order to reduce the initial amount of testing and re-evaluate the mean and control limits after sufficient datasets are obtained. Where a Laboratory is using more than one of the same instruments/platforms, they would generally perform the study on one instrument and then perform a comparison testing to confirm performance on the other instruments. Alternatively, the laboratory could opt to run alternate QC runs across platform instruments. Alternative approaches have also been suggested for establishing and monitoring the performance of serology assays based on extended historical data and peer group assessment. This is covered within article by Sally Baylis *et al.*

The most frequently used control rules applied within the clinical laboratory are those defined by Westgard (See Relevant Websites section). The performance can be monitored graphically by plotting the measurements in Levey-Jennings control charts and any deviations from the expected performance can be investigated and the appropriate corrective action taken if required. There are numerous web-based software packages available for the reporting and monitoring quality control data. Some such as Acusera 24.7 (See Relevant Websites section) offer LIMS connectivity, interlaboratory peer group assessment, as well as the ability for the laboratory to further interrogate parameters such as specific operator, and instrument associated with a particular set of QC data. This enables the laboratory to proactively monitor QC performance and identify important trends or shifts in expected results over time which may impact on the assay performance and ultimately the patient test results.

Suitable Standards, Reference Materials and Controls to Support Quality Assurance in Virology

International standards are essential to quality assurance in virology as they provide the means for the calibration of quality control and EQA materials which in turn help support the comparability of results across assay methods and provide confidence in the ability of the laboratory to deliver accurate and reproducible patient test results. In particular, the utility of viral load assays has improved significantly through the calibration of assays to the available International Standards particularly where patient management depends on the ability to relate patient results to prior results or to defined viral load values which support clinical decision making in line clinical practice, such as in the monitoring of post-transplant CMV infections.

The development of WHO International standards is covered in chapter (Sally Baylis), however in the context of clinical virology, and in particular the quantitation of viral nucleic acid, the concepts of metrological traceability as defined in ISO17511 (In vitro diagnostic medical devices – Requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples) which are widely used in other areas of laboratory medicine such as clinical chemistry can generally be applied to molecular viral load assays. Metrological traceability describes the steps through which a patient test result can be traced back to a relevant higher order standard and the uncertainty of measurement in relation to each step estimated in a common unit of measurement. The highest order of standard would be one which is ultimately traceable to the metric system/international system of units (SI). WHO International standards are derived by consensus and are not ultimately traceable back to an SI based value and do not have an assigned uncertainty of measurement nevertheless their utility and importance in driving standardization in molecular virology continues to be demonstrated.

Additional to the WHO initiatives, other international organization such as NIST (National Institute of Standards and Technologies) and LGC (Laboratory of the Government Chemist) have supported the development of methods such as Digital PCR (dPCR) which enables the counting of individual target viral nucleic acid molecules. Recent interlaboratory studies have demonstrated the accuracy and reproducibility of dPCR for the absolute quantification of DNA copy number concentration in comparison to techniques such as flow cytometry and isotope dilution-mass spectrometry. Thus, the inclusion of dPCR within the latest version of ISO17511 means it has the potential to become a recognized SI-traceable primary reference measurement procedure. However, dPCR still has limitation in the same context as any other PCR based methodology such as assay bias. Therefore, understanding the relationship between International Units (IU) and digital copies in a clinical context and across different non-PCR based molecular technologies such as Transcription Mediated Activation (TMA) used by Hologic, Strand



Fig. 3 Schematic of the EQA cycle.

Displacement Amplification (SDA) employed by BD ProbeTec, etc is essential in further supporting standardization. In addition, the commutability of quality control or reference materials needs to be demonstrated relative to the patient specimen and assay technology it is tested on in order to further support comparability of results.

The development of an International standard is primarily based on public health need. However relative to the number of viral pathogens routinely tested in the clinical laboratory there are very few International standards and certified reference materials (CRM) available with which to calibrate secondary standards and commercially available controls. Therefore, technologies such as dPCR and Next Generation Sequencing (NGS) enable commercial control manufacturers to molecularly characterized the materials in the absence of an International standard. When an International standard becomes available these materials can then be back calibrated to the standard to achieve further traceability.

External Quality Assessment and Performance Criteria

External quality assessment (EQA) or proficiency testing (PT) is integral to quality assurance within the clinical virology laboratory. It provides assurance and confidence in the laboratory's procedures and service provision. It also enables the laboratory to compare its performance to other laboratories carrying out similar tests, as well as provides an indication of areas for quality improvements. The participation in EQA is often an accreditation requirement as defined in ISO15189 and more recently ISO22870 which covers the specific use of point of care testing within the clinical laboratory setting. In many countries it can be a mandatory requirement with laboratories required to report their EQA performance to the national body within that region. In these circumstances, EQA is often linked to the test reimbursement policy within specific countries. The EQA provider, is also required to ensure that the EQA schemes it provides fulfill the requirements of ISO17043 (Conformity assessment – General requirements for proficiency testing) (Fig. 3).

The EQA cycle starts with the EQA provider, supplying registered laboratories with "blinded" clinically represented samples at predetermined intervals throughout the year. The laboratory is required to test the EQA samples using its routine laboratory method and return their results through a web-based portal to the EQA provider within a defined period of time. Data analysis is carried out by the EQA provider and an EQA report which details the laboratories individual results and peer group performance is provided to the clinical laboratory along with a certificate of participation. In regions where there are defined national EQA performance criteria for specific target analytes, such as in Germany for the viral load determination of HIV, HBV, and HCV, the EQA provider will also report whether the laboratory has achieved the required level of proficiency.

Individual Report		QCMD 2020 Human Immunodeficiency Virus RNA EQA Programme				 QCMD Quality Control for Molecular Diagnostics	
Catalogue Code: QAV994108	Ref Code: HIVRNA20	Challenge: C1	Analysis Type: Qualitative and Quantitative	Dataset:	Report UID:	Laboratory	

Intended Results / Panel Composition

Sample Code	Sample Content	Matrix	Sample Relationships ^[1]	Detection Frequency ^[2]	Sample Status ^[3]	Consensus (Copies/ml) ^[4]		Range ^[5]
						(Log ₁₀)	(n)	
HIVRNA20C1-01	HIV Type 1B	Plasma	D1	Frequently Detected	CORE	2.721	225	1.886 - 3.793
HIVRNA20C1-02	HIV Type 1B	Plasma	D1	Frequently Detected	CORE	2.731	226	0.602 - 3.811
HIVRNA20C1-03	HIV Type 1C	Plasma		Frequently Detected	CORE	3.744	225	0.602 - 4.342
HIVRNA20C1-04	HIV Negative	Plasma		Negative	CORE	N/A	N/A	N/A

Fig. 4 Example of an HIV viral load EQA panel composition.

EQA Design and Objectives

As molecular diagnostics account for the majority of the test output of the clinical virology laboratory, EQA design has progressed in order to meet increasing demand. Specialist EQA providers such as QCMD (See “Relevant Websites section”) which focus specifically on the provision of EQA within the area of molecular infectious disease offer a growing range of schemes covering viral quantitation, as well as the genotypic analysis for the detection and typing of drug resistance variants based on nucleic acid sequencing using techniques such as NGS.

The design and objectives of a clinical virology EQA scheme is dependant on many factors such as the type of method (e.g., qualitative, quantitative, sequencing, typing), its intended use, the matrix. In addition to this, in some countries where there are mandatory EQA schemes the regional regulatory bodies will often define the number and type of EQA samples that can be included within the EQA panel as well as the frequency of distribution to the laboratory (e.g., once every 6 months in the US and in Germany).

Within the example shown in [Fig. 4](#), the EQA panel consisted of two different viral strains of HIV, at a range of different clinically appropriate viral loads with the lower titer sample included to assess the sensitivity of the participating laboratory tests. In addition to this the panel also contained a negative matrix only sample in order to assess absolute specificity/false positivity. The inclusion of duplicate samples within an EQA challenge and across EQA distributions also supports the evaluation of homogeneity and reproducibility of the EQA panel members.

For many quantitative EQA schemes, the assessment of performance is based on the consensus mean of results returned by all laboratories which tested the EQA sample, once any outlying values have been identified statistically and removed from the initial analysis. Alternatively, the consensus mean of the assigned method or technology peer group can also be used, such as in the example in [Fig. 5\(a\)](#) and [\(b\)](#).

In both approaches the mean and standard deviation are calculated accordingly and a sample performance score based on the distance the laboratories result is from the consensus mean is established. In the example provided, zero points if the result is within one standard deviation from the mean. One penalty point if the result is between one and two standard deviations, two penalty points within two and three standard deviations, and three penalty points if the result is more than three standard deviations from the mean.

The regional regulatory bodies have oversight of viral diagnostic testing for regulated analytes such as the blood borne viruses HIV, HBV, and HCV in some countries and define the minimum performance criteria for laboratories participating in these EQA schemes. For qualitative tests, failure of the laboratory to obtain results within 80% of the overall consensus is considered unsatisfactory and laboratories which fall outside this range require investigation, the outcome of which may involve escalation to the regulatory body.

a**Your Summary Results**

Units	Copies/ml
EQA Assessment Group ^[1]	Abbott Real Time PCR
Core Panel Detection (Qualitative) Score ^[2]	6
Core Panel Estimation (Quantitative) Score ^[3]	6

Core Panel Members Results

Sample Code	Unitage	EQA Assessment Group Consensus ^[4]	SD ^[5]	Quantitative Result		Qualitative Result		
				Your Result ^[6]	Estimation Score ^[7]	Percentage Correct (All) ^[8]	Your Result ^[9]	Detection Score ^[10]
HIVRNA20C1-01	Copies/ml	2.742	0.114	2.760	6	99.3	Positive	6
HIVRNA20C1-02	Copies/ml	2.720	0.151	2.610	6	100.0	Positive	6
HIVRNA20C1-03	Copies/ml	3.546	0.149	3.525	6	99.5	Positive	6
HIVRNA20C1-04	Copies/ml	N/A	-	LOD/INR	N/A	99.3	Negative	6

Fig. 5 Example of performance scoring within the HIV viral load EQA scheme. (a) Qualitative and quantitative panel score. (b) Individual sample analysis (quantitative).

It is important that clinical virology laboratory management and operational staff evaluate the outcome of the EQA schemes through regular quality meetings. A preliminary step is ensuring there are no administrative transcription errors which could inadvertently affect performance. The laboratory should also compare their results to previous EQA distributions and in line with known performance characteristics for the assay as defined through verification/validation and IQC monitoring. Where the IQC indicates that the performance of the assay is acceptable but the EQA indicates a possible quality issue, the laboratory should investigate further and if required obtain further information from the EQA provider as well as the assay manufacturer (if the assay used is commercially available).

As well as meeting regional regulatory requirements, EQA schemes also provide the laboratory with educational feedback in relation to the test coverage for current clinically relevant viral strains, the incidence of false positive results, and differences in laboratory performance with a specific assay or technology. This also includes the monitoring of observed trends in laboratory EQA reporting over time, such as the move from reporting results in copies/ml and the gradual increase in the reporting of EQA results in IU/ml following the introduction of the first WHO International Standard for CMV viral quantitation and more recently the range of gene targets in assays targeting SARS-CoV-2 (Fig. 6(a) and (b)).

The aim is to ensure optimal quality performance and compliance with the accreditation standard and regulatory authority. In addition, participation in EQA schemes also helps identify possible areas for improvement and ensure that the assay is clinically fit for purpose.

Summary and Future Aspects of QA to Virology Laboratory

Viral tests are fundamental for patient management in diagnosis, infection control, and disease outbreaks where the rapid assessments of disease burden and progression is essential for evaluating the effectiveness of interventions and the verification of disease status. Technologies such as Next Generation Sequencing (NGS) and metagenomics allow the whole genome characterization at the nucleotide

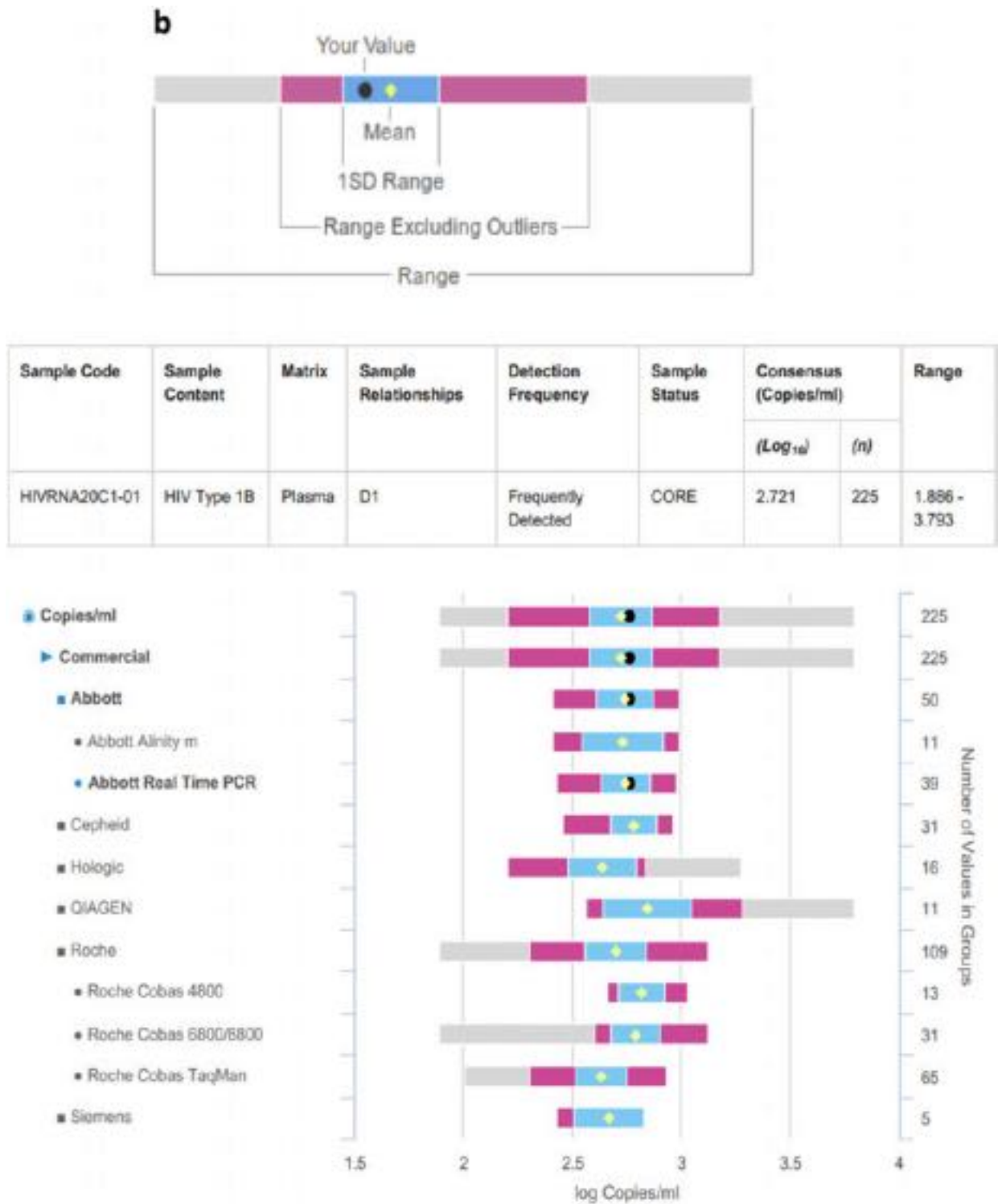


Fig. 5 Continue.

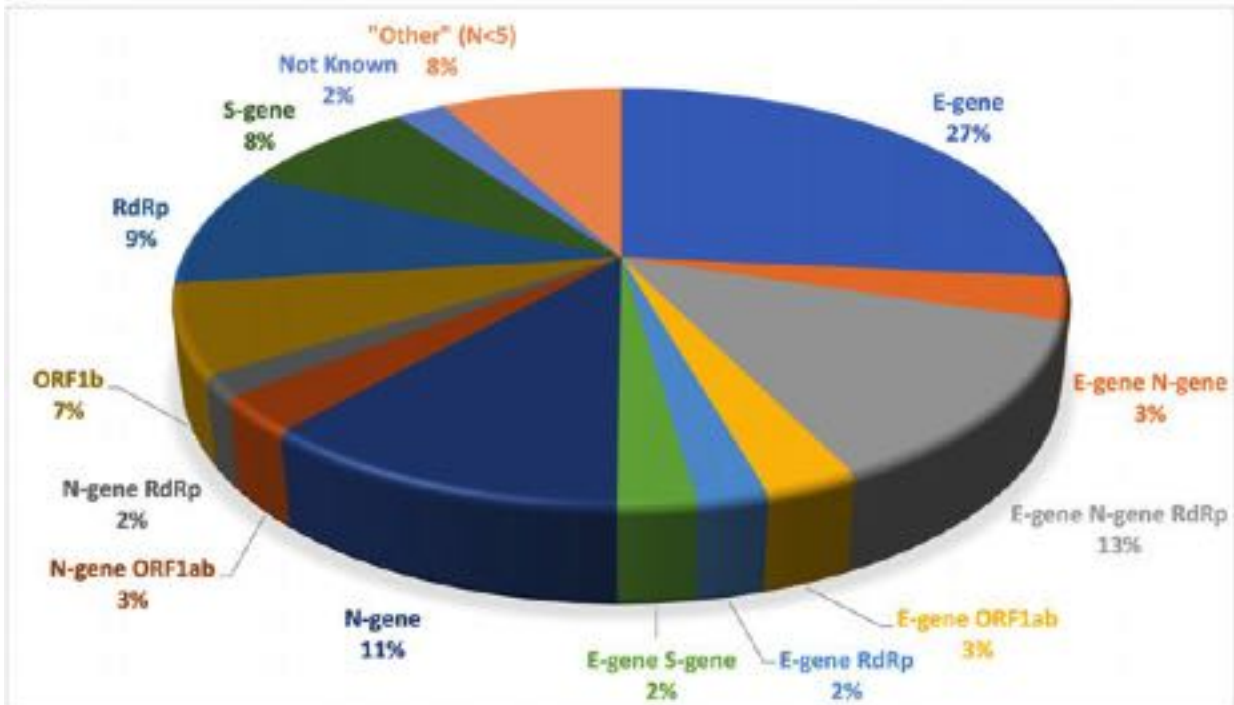
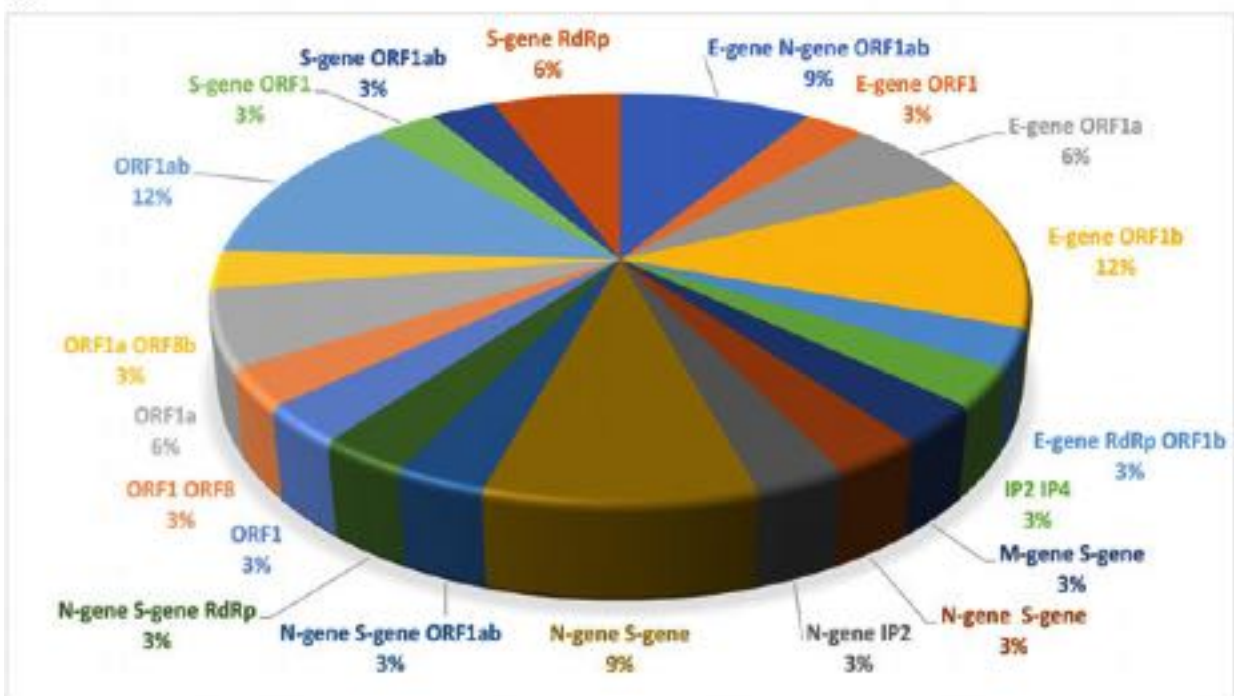
a**b**

Fig. 6 The range of different target gene loci reportedly used by laboratories for the detection of SARS-CoV2 within the QCMD EQA distribution (n=1011 laboratories). (a) Major gene targets/gene combinations used. (b) Less commonly reported gene targets.

level which as well as enhancing diagnostics also aids rapid epidemiology and infection control which is particularly important in relation to the recent Coronavirus pandemic. These technologies require advanced analytical and artificial intelligence software algorithms which support the extraction of diagnostic information. This poses additional challenges in quality assurance which will need to be addressed before such technologies can reach full routine potential and further drive diagnostic improvements as they become more accurate, simple, and affordable for use across a wide range of intended clinical settings.

Further Reading

- Baylis, S., 2020. Quality assurance and laboratory accreditation. In: *Encyclopaedia of Virology*, fourth ed. Elsevier Science.
- Baylis, S.A., Wallace, P., McCulloch, E., *et al.*, 2019. Standardization of nucleic acid tests: The approach of the World Health Organization. *Journal of Clinical Microbiology* 57.
- Bustin, S.A., Benes, V., Garson, J.A., *et al.*, 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* 55 (4).
- EP05-A3E. Evaluation of Precision of Quantitative Measurement Procedures, third ed. Clinical Laboratory Standard Institute (CLSI).
- Huggett, J.F., Foy, C.A., Benes, V., *et al.*, 2020. The digital MIQE guidelines update: Minimum information for publication of quantitative digital PCR experiments for 2020. *Clinical Chemistry* 66 (8), 1012–1029.
- Hayden, R., Sun, Y., Tang, L., *et al.*, 2017. Progress in quantitative viral load testing: Variability and impact of the WHO Quantitative International Standards. *Journal of Clinical Microbiology* 55 (2).
- ISO17511, 2020. In vitro diagnostic medical devices – Requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples.
- ISO17043, 2010. Conformity assessment – General requirements for proficiency testing.
- Miller, M.B., Atzadeh, F., Burnham, C.D., *et al.*, 2019. Clinical utility of advanced microbiology testing tools. *Journal of Clinical Microbiology* 57 (9). doi:10.1128/JCM.00495-19.
- Whale, A.S., Jones, G.M., Pavšič, J., *et al.*, 2018. Assessment of digital PCR as a primary reference measurement procedure to support advances in precision medicine. *Clinical Chemistry* 64 (9), 1296–1307.

Relevant Websites

- <https://www.bipm.org/en/committees/cc/wg/jctlm-rt-nucleic-acids.html>
Bureau International des Poids et Mesures.
- <https://doi.org/10.1093/clinchem/hvaa125>
Clinical Chemistry,
- <https://www.cap.org/>
College of American Pathologists.
- <https://clsi.org/>
Global Laboratory Standards for a Healthier World.
- www.idealgen.com
Idealgen.
- <https://www.iso.org/>
International Organization for Standardization.
- <https://www.qcmd.org/>
Quality Control for Molecular Diagnostics.
- www.randox.com/acusera-24-7-interlaboratory-data-management/
RANDOX.
- www.umcg.nl
UMCG.
- www.unece.org
United Nations Economic Commission for Europe.
- www.westgard.com
WESTGARD.
- <https://www.who.int/>
World Health Organization.

Biosafety and Biosecurity in Diagnostic Laboratories

Hannimari Kallio-Kokko, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Susanna Sissonen, Finnish Institute for Health and Welfare, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Biosafety and Biosecurity

Biosafety and biosecurity are not unambiguous concepts but have different meanings in different contexts.

In laboratory context, biosafety usually refers to “the containment principles, technologies and practices that are implemented to prevent unintentional exposure to pathogens and toxins, or their accidental release” (WHO Laboratory Biosafety Manual, 3rd edition). Biosafety aims to protect the workers, the community and environment. Many common laboratory practices such as good microbiological practices and procedures, use of personal protective equipment (PPE), decontamination and disinfection procedures or the use of safety equipment such as biosafety cabinets are part of laboratory biosafety.

Laboratory biosecurity refers to “institutional and personal security measures designed to prevent the loss, theft, misuse, diversion or intentional release of pathogens and toxins” (WHO Laboratory Biosafety Manual, 3rd ed.). Laboratory biosecurity involves responsibility for the protection, control and accountability of biological materials. Measures that prevent the malicious use of pathogens and toxins such as permits and licenses for the handling, import and export of biological materials, physical structure of the laboratory facilities, restricted access or even ethical codes for life scientists can be regarded as biosecurity measures.

Diagnostic Laboratories

Diagnostic laboratories handle human/animal samples that potentially contain hazardous microbes, and to ensure safe working environment to all personnel, the working practices maintained should follow thorough risk assessment at all stages. The safe working practices in any laboratory contain codes of conduct, competent and appropriately trained staff, good microbial practices and procedures at all stages of handling potentially infective material. In addition, laboratory facility and equipment in use should be appropriate. Regardless of the focus of a laboratory, the possibility of highly contagious viruses, bacteria, fungi, parasites or prions being present in same sample should be taken into consideration, and be included in the risk evaluation to cover safe handling of samples and to ensure choice of proper and effective disinfectants.

The selection and types of diagnostic assays performed in a laboratory affect the probability of risks as do e.g., the level of need for manual handling of samples. Working procedures that contain vigorous shaking (vortexing) or centrifugation of patient samples demand special biosafety precautions. Also assays that aim to concentration or multiplication of infectious agents (e.g., virus cultivation) cause special requirements for both the biosafety level of the facilities and utilization of other safety measures including PPE. The use of automated laboratory assay systems reduce the risk caused by manual sample handling, but problems in the function of machines rise another type of challenges, and handling and cleaning of possibly contaminated laboratory equipment is an important part of risk evaluation, the suitability of effective disinfectants should be ensured also with the equipment manufacturer.

Even though potential to the presence of hazardous microbes is a reality in all diagnostic laboratories, and good basic laboratory practices, e.g., no eating or drinking in the laboratory are the ground for assuring laboratory biosafety, differences in the probability of highly contagious pathogens do exist. The epidemiological situations vary in different countries, leading also to differences in the level of immunity and susceptibility of individuals in the community. The type of a laboratory affect the sample material handled and assays performed, and laboratories providing advanced diagnostics and diagnostics for high threat pathogens need more control measures and enhanced biosafety precautions to be obtained.

Laboratory Biosafety

Biological Agents

Microorganisms are classified into four risk groups (1 – 4) based on the severity of the risk caused by the microbe to both an individual and the community. The severity of the disease, the transmissibility to other individuals as well as the availability of effective treatment and preventive measures affect the classification.

Risk group 1 contains microbes that are unlikely to cause a disease in healthy humans.

Risk group 2 contains microbes that are able to cause disease in humans, but are unlikely to cause serious threat to the environment or community. In case of an infection, effective treatment and preventive measures are available.

Risk group 3 contains microbes that usually cause serious disease in humans, but pose a limited risk of spreading to the community, and effective treatment and preventive measures do exist.

Risk group 4 contains microbes that usually cause serious disease in humans, and can easily be transmitted from one individual to another with no effective treatment or preventive measures usually available.

For a single microbe, the classification may vary between different countries due to e.g., epidemiological situations or existing levels of immunity. Risk groups 2–3 contain viruses, bacteria, fungi and prions whereas all agents belonging to risk group 4 are viruses.

For occupational evaluation of risks caused by handling microorganisms, it should be noted that the list of classified agents is based on the effect of those agents on healthy individuals only. Pre-existing disease, ongoing medication, compromised immunity, pregnancy or breast-feeding may affect the susceptibility to a pathogen. Therefore, special care should be taken when evaluating the risks of even microbes with lower classification for personnel on whose susceptibility may be affected for one or other reason.

Laboratory Biosafety Levels

To ensure the safe working environment in laboratories, there are traditionally also four laboratory biosafety levels, which describe the level of the biocontainment structures and precautions required to handle biological agents in an enclosed facility. The classification is designed to provide protection to personnel, the environment, and the community. Biosafety level designations are a composite of the laboratory facilities, equipment, practices, operational procedures and PPE required for working with biological agents from the various risk groups (1 – 4). Differences between laboratory biosafety levels include requirements for ventilation systems, waste management, access control and safety practices such as work practices and the use of PPE. The lowest biosafety level with least specific requirements is level 1 (BSL1) precautions getting stricter step by step to the highest level 4 (BSL4). In BSL3 and BSL4 laboratories, special engineering and design features are required to prevent the accidental release of the pathogens from the laboratory. Also special training in handling the pathogens is required for the workers in high containment level laboratories.

The risk groups of biological agents relate but do not equate to the biosafety level of laboratories. The assignment of handling an agent to a biosafety level for laboratory work must be based on a risk assessment covering all working procedures. Currently not only the classifications of agents and facilities but especially the risk based and the evidence-based approach to biosafety are emphasized more and more. WHO is updating the Laboratory Biosafety Manual, and when evaluating the risks, the main principles will be the same with the consequence of exposure and release, and likelihood of exposure varying from low to high. The response to the likelihood of the risks will contain core requirements (equivalent to BSL2 without BSC) for handling lower levels of risks, and as the risks elevate, heightened control measures are needed (equivalent to BSL2+ and BSL3), ending to maximum containment procedures (equivalent to BSL4) if the highest risks are being present.

Types of samples handled and diagnostic assays performed affect the needs for biosafety precautions, e.g., the propagation of risk group 4 viruses demand the highest level of biosafety practices. Procedures on samples that might contain agents with potentially severe consequences or on samples with known pathogens of high consequences need the highest level of biosafety practices obtained.

Risk Assessment

Risk assessment is the process in which the risk(s) of working with hazards are assessed and evaluated. Based on the risk assessment the necessary control measures are determined to reduce the risk to an acceptable level. A hazard is something that can cause harm. A risk is the combination of the probability of the hazard to occur and the severity of the harm. Strategies for risk reduction in microbiological work include also the possible elimination of the work, substitution with an alternative organism/activity, isolation of the hazard, use of engineering controls, use of administrative controls and use of PPE. As the risk assessment is the basis for the work safety, laboratories should have a routine for assessing risks, laboratory directors and principal investigators having the responsibility for follow up of proper evaluation and optimal practices.

Risk assessment should take into account the pathogenicity of biological agents handled, mode of transmission of these agents, volume and concentration of infectious material handled. The presence of other risks such as use of needles and sharp items, high risk of creation of aerosols during working procedures, or use of hazardous chemicals and inflammables are included into evaluation. Also the availability of effective treatment or prophylaxis in case of accidental exposure is to be considered. The types of patient samples handled in the laboratory, the types of analyzes performed on the samples (e.g., the potential of aerosol formation), and the facilities available affect whether the risk can be safely handled. The route of transmission and the transmissibility of the agents affect the safety procedures selected for each work stage.

The risk assessment should also cover the evaluation of effective decontamination methods and safe waste handling procedures as well as biosafety precautions. Training of the personnel is a key factor for ensuring the proper biosafety cultures in all laboratories, and the role of maintenance personnel (e.g., cleaners, technical personnel, maintenance personnel of laboratory equipment) should be included.

Routes of Transmission and Infective Dose

In laboratory setting, the route of transmission may be different as compared to naturally acquired infections. Intact skin protects well against most microbes, but e.g., wounds or atopic eczema can break down this barrier. Mucous membranes and eyes are

vulnerable, and serve a possible infection route for several microbes. In laboratory settings the most probable transmission routes for infective agents are through blood (needle stick accidents or cuts with sharp objects), mucosa and airways (spills or splashes into mouth, nose or eyes, through generation of aerosols), or gastrointestinal route (spills, by touching mouth with contaminated fingers or while consuming food in laboratory).

All patient samples are considered potential to transmit infection to a laboratory worker. High risk of transmission is associated with contagious blood (blood borne pathogens e.g. HIV), nasopharyngeal specimens and other respiratory samples (air borne pathogens e.g., influenza viruses), feces (e.g., norovirus), brain biopsies and central nervous system specimens including cerebrospinal fluid (e.g. rabies virus, prions), tissue specimens, and skin biopsies and blisters (e.g., poxviruses and herpes simplex viruses). The most often reported ways of acquiring laboratory associated infections include splashes or sprays of infective material, needle stick injuries, accidents with sharp objects, and earlier also mouth pipetting. Working procedures such as pipetting and pipette spills on surface, taking blood samples or handling blood tubes as well as the usage of centrifuges, blenders, shakers, mixing instruments, and sonicators pose elevated risk for aerosol generation.

The infective dose is the number of microbial particles needed to cause an infection, and it has an important role while evaluating the level of risk caused by handling the patient samples or enriched cultivations. For e.g., norovirus the infective dose is estimated to be as low as only 10–20 virus particles.

Personal Protective Equipment (PPE)

PPE is meant to block the direct transmission of an infective agent to a worker, and should be worn in circumstances where a risk assessment has shown the continuing needs for personal protection. PPE is usually regarded as the last measure to prevent the exposure to infectious agents; all other control measures such as the use of engineering controls should be applied before using PPE. PPE must be suitable for the purpose, suitable for the person and used correctly.

The selection on the PPE is based on the risk assessment. Protective clothing, gloves for hand protection, facial masks for respiratory protection, and goggles and/or face shield for eye protection are examples of personal protective equipment available. It is recommended to avoid wearing protective laboratory clothing outside the laboratory, e.g., in canteens, coffee rooms, offices, libraries, staff rooms and toilets. While working in higher containment facilities, selected PPE is always put on prior to enter to, and removed before exit from the facilities.

Protective clothing is considered the basic PPE used in all clinical laboratory settings. The selection of the material of protective clothing is based on the type of work performed. In case of a constant risk for heavy spills, the material should be water resistant, if no risk for heavy spills also water repellent material can be a choice. In case of contamination of the clothing during work, the decontamination of material by e.g., autoclaving before sending to laundry unit is obligatory. Especially in higher containment facilities, also disposable clothes, or positive pressure suits ventilated with motor unit are used.

As the material of protective gloves vary, the selection is on the potential hazard (microbial and/or chemical). Nitrile has good shield against viruses and high or medium protection against some chemicals (high against e.g., glutaraldehyde, sodium hydroxide and hydrogen peroxide, medium against e.g., 70% ethanol, but not recommended against e.g. acetone or methanol). Latex is good against microbes, but pose a risk of allergies and does not protect against organic solvents. Vinyl is not suitable for work with highly infectious material. Standards, such as EU standard EN374-5:2016 “Protective gloves against dangerous chemicals and micro-organisms” help in selecting the proper material for intended use. Glove packages should be marked with pictograms, which indicate the suitability of the glove against chemical and biological risks. After selecting the correct material, also proper way of using gloves is important. Appropriate gloves are to be worn for all procedures that may involve direct or accidental contact with potentially infective patient samples such as blood or body fluids, other potentially infectious materials or infected animals. Regular change of gloves while working, especially in case of a potential spill to hands, and removing the gloves aseptically is crucial to prevent contamination. Also cleaning the hands after removal of gloves ensure transmission prevention.

For respiratory protection, several different types of masks are available, mask types varying in the level of protection provided. When working with material posing a high risk for airborne transmission, surgical masks offer the lowest protection for the worker, and FFP3/N-95 masks provide a better protection, assuming the person have passed the fit test with the FFP3/N-95 mask selected. Respirator with motor unit using filters against chemicals and/or microbes are a good choice when needed.

To ensure best practices, the training of personnel should include the correct selection and use of PPE as well as the correct dressing and undressing procedures, keeping in mind that after each working period the outside of the PPE is considered potentially contaminated.

Biosafety Cabinets

In diagnostic laboratories, biosafety cabinets are in use to protect the worker as well as to protect the sample from contamination. Biosafety cabinets use laminar airflow for cabinet air, and e.g., High-Efficiency particulate air (HEPA)-filters for incoming and/or exhaust air to capture potentially hazardous particles. Class I cabinet does not filter the incoming air, so while working in these only the worker is protected from contaminating agents. Class II cabinets drive both incoming and exhaust air through filters, and thus protect both the worker and the sample (Figs. 1 and 2). Class III cabinets completely isolate the sample from the worker and are the safest choice for work with highest threat pathogens (Fig. 3). In class III cabinets both incoming and exhaust air is filtered, and the cabinet connect to ventilation system through hard duct connection.

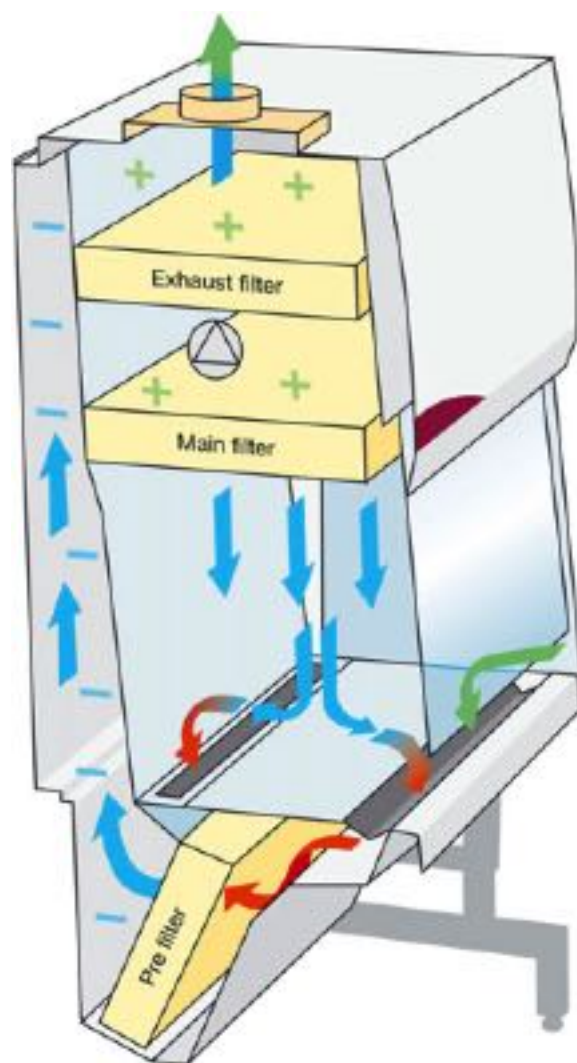


Fig. 1 The air flow system in biosafety class II cabinet. Green arrows = Room air or filtered air to exhaust; blue arrows = Filtered air inside the cabinet; red arrows = Contaminated air before filtering.

With class I and II cabinets, working practices affect the safety of the cabinet, vigorous moving of hands or blocking of the cabinet's air circulation system by paper or material causing loss of safety. After each working session the cabinets are cleaned and treated with disinfectants chosen according to risk assessment, special care is taken when handling samples for nucleic acid detection assays to reduce the possibility of nucleic acid contamination, e.g., by using disinfectants that also break down nucleic acid chain residues. Proper training of personnel is crucial to ensure the optimal protection provided by biosafety cabinets.

Decontamination and Waste Management

Very important is to carefully plan and instruct the disinfection practices of all laboratory work spaces as well as laboratory equipment, which have been in contact with potentially infectious agents. Selection of decontamination method is based on the evaluation of the properties of microbial agents handled in the laboratory. Use of disinfectants should follow the instructions provided by the manufacturer, and the suitability on special laboratory equipment confirmed from the equipment manufacturer. To ensure proper effect, the concentration of disinfectants used, the contact time of disinfectant on table surfaces or equipment, and the storage of the disinfectant should always be as instructed by the manufacturer.

Several disinfecting chemical agents are available: aldehydes (e.g., glutaraldehyde and formaldehyde), alcohols (e.g., ethanol and propanol), biguanides (e.g., chlorhexidine), phenol compounds, halogens (e.g., chlorine compounds and iodine), quaternary ammonium compounds, and peroxygens (e.g., hydrogen peroxide and peracetic acids). Of these, peroxygens have the widest effective range against viruses covering both non-enveloped and enveloped virus groups. In some facilities, also UV-light (germicidal UV-light, 254 nm radiation) and ozone are in use for decontamination. Special attention to selecting the decontamination method is crucial when specimens handled in a laboratory have potential to contain highly resistant agents such as bacterial spores or prions.



Fig. 2 An example of a class II biosafety cabinet.

Safe collection of all waste including sharp material (e.g., needles and pipette tips) prevent the risk of laboratory accidents and assure safe waste management at all stages. All waste is to be collected to specific waste buckets, and the content marked to assure safe disposal of the buckets. Special attention is required on the selection of waste buckets used for highly contagious waste, these should be leakage proof and sealed for safe transportation.

When found necessary waste management should include decontamination of waste before disposal by chemical ways, heat inactivation or autoclaving depending on the potential infectivity of the waste (e.g., primary samples or virus cultivations). Incineration of special waste and safe transportation to disposal unit are also important. In higher containment laboratories, decontamination of sewage water is also required either by heating or by chemical ways before release to sewage system. In higher containment laboratories, fumigation of the facilities with e.g., hydrogen peroxide or formaldehyde is required on regular basis to ensure the safe maintenance of the facilities and laboratory equipment.

Emergency plan including the choice of methods and disinfectants is important to ensure the quick and effective response in case of unexpected events.

Laboratory Accidents

Exhaustive reports on laboratory-acquired infections are scanty. According to the studies, many laboratory accidents and laboratory-acquired infections originate from human factors such as insufficient training, lack of PPE, lack of risk assessments or inadequate risk assessments, lack of standard operating procedures or negligence of following the existing instructions. The best way to prevent laboratory accidents is by thorough education of the staff and maintenance of good laboratory practices on daily basis. Programmed action points and emergency plans in case of an accident prevent/mitigate effectively serious consequences.

In laboratory settings, accidental routes of pathogen transmission include ingestion, injection, cuts and abrasions, and inhalation of infectious aerosols. According to published studies, in many cases the route of exposure remains unknown. Of the identified cases, it is usually by inhaling infectious aerosols. An aerosol is a colloid of fine solid particles or liquid droplets, in air or another gas. An aerosol with a diameter of 5 μm or less can remain airborne for a long period of time, spread wide distances, and is easily inhaled. Proper PPE such as masks or respirators are to prevent primary contact. To mitigate the risks connected to aerosol generating procedures also the use of sealable buckets for centrifugation and opening the buckets only in biosafety cabinet offer best protection in case of e.g., broken tubes during centrifugation.



Fig. 3 An example of a Class III biosafety cabinet.

In case of a spill of infectious material to mucosa or eye, thorough rinsing with clean water is usable first aid, and proper first aid, e.g., eye shower, should be available. In case of wounds to skin gentle cleaning with soap and water, and thorough rinsing with clean water are usable first aid. To prevent the serious consequences of accidents with certain known and preventable microbes, also post exposure prophylaxis is available (e.g., HIV drug administration after a needle accident while handling HIV-positive patient blood sample or rabies-immunoglobulin administration and/or vaccination after exposure to potentially infective sample from a rabies patient or animal). Prophylactic possibilities are included in the risk assessment.

Occupational Health and Special Groups

Occupational health has an important role in providing guidance and in evaluating the workers ability to work in laboratory settings. Management of the facilities assure that both physical and mental health requirements are covered. Vaccines offered by occupational health for vaccine-preventable infectious diseases, such as influenza and hepatitis B are widely used for additional preventive measures, and can offer a protective method in addition to following good laboratory practices and using PPE. Use of vaccinations should be included in risk assessment especially if work contains virus cultivation or handling high virus concentrations of certain pathogens, e.g., rabies virus and TBE virus.

Planning of working profile is important especially with special groups, which are more vulnerable to infections due to their health status or ongoing medication. Special planning is important e.g., with pregnant women, immunocompromised workers, or workers suffering from atopic eczema or allergies against e.g., disinfectants or other material used in PPE (e.g., chlorine or rubber). The need for regular, medical examinations should be included in the risk assessment.

Shipment of Clinical Specimens

As samples often need to be transported from the medical unit taking the samples to diagnostic laboratory doing the microbiological analysis, laboratory has an important role also in instructing the clients to send the specimens. Instructing covers the selection of appropriate tubes for different specimens and assays, proper ways of packaging including the usage of



Fig. 4 An example of a package for transportation of Category B substances.

leakage proof tubes, adsorbent and secondary package, and transportation systems available. International and local regulations guide the transportation of potentially infective samples. International Air Transport Association (IATA) supports aviation with global standards for e.g., airline safety and security, and offers instructions and guidelines for air transportation of highly infectious material. Category A samples listed by IATA demand more extensive safety procedures than Category B samples with lower infective potential (Figs. 4 and 5). Shipment of samples containing potentially highly infectious material demands careful planning and instructions in laboratories, special training in packaging and shipment, and well-organized contacts to couriers.

Laboratory Biosecurity

During the last decades, the concerns for the deliberate malicious use of pathogens and toxins have arisen. Many pathogens and toxins can be used intentionally to cause harm for humans, animals or the environment. Laboratory biosecurity measures prevent the malicious use of micro-organisms. Also the information and knowledge related to pathogens as well as materials like laboratory equipment can be controlled by biosecurity measures.

Microbial agents have a dual use nature: they can be used in benign diagnostics, research or commercial purposes but they can be also used in malicious purposes. It can be difficult to prevent the unauthorized use of biological agents since the same agents can be found in nature. Even the diagnostic laboratories should take the possibility of malicious use of micro-organism into account, even though the probability of malicious use is limited.

Biosafety and biosecurity practices usually comply, and often biosafety and biosecurity measures must be integrated to have an effective biosecurity system. In some cases, biosafety procedures and biosecurity procedures may be in conflict. For example, the list of biological agents handled in the laboratory can be available for safety purposes but at the same time this information should be limited for security purposes.

The key components of laboratory biosecurity are physical security and access control, personnel reliability, information security, transport security and accountability for materials. In addition, biosecurity awareness and emergency response planning are integral parts of an effective biosecurity system. Physical security control measures are used to prevent unauthorized access to sensitive biological materials or information. Physical security measures also improve laboratory biosafety since they restrict the access to the laboratory to only those personnel who are trained to work with the pathogens. Also personnel reliability measures prevent the pathogens and toxins out of the hands of individuals who might have malicious purposes. Personnel who handle or transport dangerous pathogens should be trustworthy, reliable and competent to perform their duties.



Fig. 5 An example of a package for transportation of Category A substances.

Some information related to the working with biological agents can be sensitive and must be kept secure. Sensitive information can include, for example, the list of infectious agents and samples handled and stored in the laboratory, list of key personnel or laboratory security plans. Accountability for materials and inventory control mean that the laboratory should have an up-to-date overview of high-risk materials that they are handling and storing. Transport control ensures that the national and international rules for packaging, transport, import and export are being followed when biological agents are being transported. The biological material should only be sent to trustworthy and legitimate receivers, and the transport should be performed by approved couriers.

The biosecurity measures needed are defined in the risk assessment. Basis of biosecurity in diagnostic laboratory consists of proper training of the personnel including also keeping up the knowledge of restricted sharing of information with third parties and restricted access to classified information. Restricted access to the facilities, including access control and follow up are also crucial. Thorough records and documentation of storages are also important to facilitate good follow up of all material that can potentially cause risk to the environment or that could be misused. Restricted access to delicate information of the facilities and the security systems ensure biosecurity.

Biorisk Management System and Legislation

Biorisk management system describes the management of both biosafety and biosecurity risks. National legislation, regulation and policies as well as international regulation, policies and guidelines often set requirements that are related to laboratory biosafety and biosecurity. Some countries have strictly regulated biosafety and biosecurity systems.

There are several international guidances and standards that are related to biosafety and biosecurity, such as WHO Laboratory Biosafety Manual, Biosafety in Microbiological and Biomedical Laboratories (BMBL) or the ISO 35001 Biorisk Management for Laboratories standard. Laboratory quality standards such as ISO 17025 General requirements for the competence of testing and calibration laboratories and ISO 15189 Medical laboratories – particular requirements for quality and competence comply with the implementation of laboratory biosafety and biosecurity.

National and international Biosafety Networks aim to promote biosecurity and biosafety aspect in different laboratories by e.g., sharing best practices and by organizing training on biosafety and biosecurity for laboratory professionals. In many organizations, Biosafety Officer or Biorisk management advisor is the professional with expertise on biosafety and biosecurity regulations and safe practices, and responsible to supervise and guide personnel in matters connected to biosafety and biosecurity, e.g., in evaluating the risk assessments.

Further Reading

Buchan, B.W., Mahlen, S.D., Relich, R.F., 2019. Interim Clinical Laboratory Guideline for Biological Safety. The American Society for Microbiology. Available at: <https://www.asm.org/ASM/media/Policy-and-Advocacy/Biosafety-white-paper-2019.pdf>.
Centers for Disease Control and Prevention, 2009. Biosafety in Microbiological and Biomedical Laboratories, fifth ed. HHS publication no. (CDC) 21–1112. Available at: <https://www.cdc.gov/biosafety/publications/bmbl5/bmbl.pdf>. (accessed 2.01.16).

Laboratory-acquired infections in Belgium, 2007–2012. Available at: https://www.biosafety.be/sites/default/files/2015_willemarck_lai_report_belgium_2007_2012_final.pdf.
Peter Clevestig, 2009. Handbook of applied biosecurity for life science laboratories. Stockholm International Peace Research Institute (SIPRI). Available at: <https://www.sipri.org/sites/default/files/files/misc/SIPRI09HAB.pdf>.
WHO, 2006. Biorisk Management: Laboratory Biosecurity Guidance. Available at: https://www.who.int/ihr/publications/WHO_CDS_EPR_2006_6/en/.
World Health Organization, 2004. Laboratory Biosafety Manual, third ed. Geneva. Available at: <http://www.who.int/csr/resources/publications/biosafety/Biosafety7.pdf?Ua=1>. (accessed 2.01.16).

Relevant Websites

<https://www.bureaubiosecurity.nl/en>
Biosecurity Office | Bureau Biosecurity.
<https://www.canada.ca/en/public-health/services/canadian-biosafety-standards-guidelines.html>
Canadian Biosafety Standards and Guidelines.
<https://www.cdc.gov>
Centers for Disease Control and Prevention.
<https://www.cdc.gov/infectioncontrol/guidelines/disinfection/disinfection-methods/chemical.html>
Chemical Disinfectants.
<https://www.epa.gov/environmental-topics/chemicals-and-toxics-topics>
Chemicals and Toxics Topics.
<https://www.ecdc.europa.eu/en/home>
European Center for Disease Control and Prevention.
<https://www.iata.org>
IATA.
<https://www.epa.gov/pesticide-registration/selected-epa-registered-disinfectants>
Selected EPA-registered Disinfectants.
<https://www.who.int/>
World Health Organization.

Screening for Viral Infections

Walter Ian Lipkin, Nischay Mishra, and Thomas Brieese, Columbia University, New York, NY, United States

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Viral diagnostics are increasingly significant for public health and clinical medicine. Recognition of an infectious disease as viral was once important only as an insight that allowed practitioners to exclude the need for antibiotics. Tools for virus diagnosis, surveillance and discovery have become more urgent with the expansion of the inventory of effective and specific antiviral drugs, and of the emergence of new zoonotic agents that require assays to facilitate containment and to direct investments in vaccines. An additional incentive is the evidence to support new models of diseases wherein viral infections result in autoimmunity, neurodevelopmental damage, or neoplasia, that manifest months to years after the infection has resolved.

In this article we will review the status of histopathology, culture, and molecular and serological assays in clinical and environmental microbiology.

Pathology

Before the advent of high throughput genetic methods for detection and characterization of viral sequences the majority of viruses were detected using electron microscopy. Prominent examples of viral discovery enabled by electron microscopy include tobacco mosaic virus, ebolavirus, variola major virus, norovirus, polyomavirus, and SARS coronavirus (Kausche *et al.*, 1939; Breman *et al.*, 2016; Van Rooyen and Illingworth, 1944; Long *et al.*, 1970; Kapikian *et al.*, 1972; Haycox *et al.*, 1999; Nicholls *et al.*, 2003). Electron microscopy continues to be an important tool in virology where it is increasingly employed for confirmation of a molecular finding rather than as the first step in investigating outbreaks of infectious disease. Readers may wish to see Goldsmith and Miller for an excellent review of the history and use of electron microscopy in virology (Goldsmith and Miller, 2009). Histopathology and immunohistochemistry, too, have shifted from primary roles in viral identification to applications for viral biology and pathogenesis. In the earliest applications of polymerase chain reaction (PCR) in viral discovery, primer selection was guided by findings in analyses of clinical materials using panels of antisera to a wide range of viruses or hyperimmune sera from patients with specific diseases. In our own work with Sherif Zaki for example, immunological cross reactivity facilitated the discovery of West Nile virus NY99 (Brieese *et al.*, 1999). The finding of a viral nucleic acid sequence in an extract from an individual with disease is insufficient to demonstrate a causal relationship. An argument for a causal relationship is strengthened if footprints of the agent in an affected organ can be demonstrated through immunohistochemistry or in situ hybridization for viral proteins or nucleic acids, respectively (Lipkin, 2010).

Culture

The focus of clinical diagnostic virology has shifted from culture to molecular methods. This largely reflects the introduction of commercial assays that allow rapid, inexpensive detection of viral pathogens using multiplex PCR panels, and quantitative PCR assays that can be used to monitor the efficiency of interventions by tracking viral burden. Culture nonetheless remains critical to viral discovery, and basic and translational viral research. Virus amplification through culture enhances the efficiency of detection through molecular methods. SARS coronavirus, for example, was readily characterized using sequencing, microarrays decorated with oligonucleotide probes representing viral sequences, and consensus PCR. Proof-of-causation through fulfillment of Koch's Postulate requires propagation of an agent isolated from an individual with a disease, and recapitulation of the disease following introduction of the agent into a naïve host. Although alternative strategies for proving a causative relationship have been established, Koch's Postulate remains the most persuasive. Virus isolation is also essential for understanding how it enters a cell, replicates itself, evades innate immune responses, and causes disease in animal models. Finally, virus isolation is imperative for developing and testing the efficacy of antiviral drugs and vaccines.

Viruses can be isolated and propagated in cell culture systems or in live animals. The choice of an in vitro versus an in vivo strategy can have a substantive impact even for closely related viruses. Whereas inoculation of suckling mice is useful in detecting human enterovirus A, tissue culture is better for detection of human enterovirus B. Success depends on the presence of cell surface receptors that enable entry; intracellular factors required for trafficking of viral nucleic acids and proteins to appropriate sites for transcription, replication, translation, assembly and egress; and the ability of the virus to evade innate immune responses that inhibit viral replication. Accordingly, many clinical microbiology laboratories maintain multiple continuous cell lines for virus isolation. Some viruses cannot be propagated in continuous cell lines. Cytomegaloviruses, for example, require immature cells and are typically grown in primary fibroblast cultures. Other viruses require specific receptors that can be introduced through transfection, or more complex systems comprising multiple cell types that resemble organs (organotypic cultures). Still others are even more fastidious and require live animals such as suckling or transgenic mice. Culture in primary cells, transfected cells, organotypic cultures, and live animal systems is typically restricted to public health reference and research laboratories due to the need for substantial investment in resources and expertise.

Nucleic Acid Tests

Polymerase Chain Reaction Assays

The discovery of PCR by Mullis in 1983 transformed clinical microbiology and public health (Saiki *et al.*, 1985) by enabling rapid, inexpensive, tools for microbial differential diagnosis, surveillance, and discovery, as well cDNA cloning. In viral diagnostics, PCR can be broadly divided into assays that target single agents versus those that target more than one agent (multiplex PCR assays), assays that are designed to detect only specific viral species versus those that also detect related viruses (consensus PCR assays), and assays that are quantitative (qPCR assays). The amplification products of single agent PCR can be visualized after size fractionation through gel electrophoresis using dyes like ethidium bromide or the newer and safer stains (e.g., SYBRsafe or GelGreen among others) that also intercalate with nucleic acids. Alternatively, products can be quantitated using complementary oligonucleotide probes that contain both a fluorescent reporter dye and quenching molecule. In Taqman assays, probe bound to the template is degraded by the 5' to 3' exonuclease activity of the polymerase, thereby freeing the reporter molecules for detection. In molecular beacon assays, the quencher and reporter molecule are in close proximity to one another in a hairpin probe structure and become separated when PCR results in a product that is complementary to the hairpin loop sequence between them. Through use of a standard curve and known concentrations of template one can measure the amount of target in an experimental sample. Taqman and molecular beacon assays are exquisitely sensitive and typically can detect as few as 10 target molecules per reaction.

Multiplex PCR

The majority of multiplex PCR assays are based on the Taqman system; however, gel based, mass spectrometric, and Luminex-bead based assays have also been established. At the time of writing, only five fluorescence reporter dyes can be unequivocally separated; thus, multiplex Taqman assays are limited to five targets (typically four viral and 1 host gene target control). The sensitivity is similar to single agent Taqman PCR. Two platforms combine PCR with mass spectroscopy to detect either the product itself or tags attached to the primers. The Ibis system, adopted by Abbott, uses matrix-assisted laser desorption–ionization (MALDI) MS to directly measure the molecular weights of PCR products in a sample and to compare them with a database of known or predicted product weights (Van Ert *et al.*, 2004; Ecker *et al.*, 2005; Sampath *et al.*, 2007; Ecker *et al.*, 2008). The Ibis system discriminates products of PCR amplification using universal bacterial 16S rRNA primers. There is no correlate universal viral primer set; thus, the platform is better suited to bacteriology than virology. MassTag PCR can accommodate up to 50 different primer sets. It uses atmospheric pressure chemical ionization (APCI) MS to detect small molecular weight reporter tags attached to PCR primers (Briese *et al.*, 2005). Syndrome-specific MassTag PCR panels have been established for the detection of viruses, bacteria, fungi, and parasites associated with acute respiratory diseases, diarrheas, encephalitides/meningitides, and hemorrhagic fevers (Briese *et al.*, 2005; Lamson *et al.*, 2006; Palacios *et al.*, 2006; Renwick *et al.*, 2007; Dominguez *et al.*, 2008; Tokarz *et al.*, 2009; Palacios *et al.*, 2009; Tokarz *et al.*, 2010; Tokarz *et al.*, 2011; Tokarz *et al.*, 2012; Al-Samarrai *et al.*, 2013; Hsu *et al.*, 2013). Another highly multiplexed platform, established by Luminex Corporation uses flow cytometry to detect PCR amplification products bound to matching oligonucleotides on fluorescent beads (Brunstein and Thomas, 2006; Han *et al.*, 2006; Li *et al.*, 2007). Signal in multiplex PCR and Luminex assays requires the presence of three independent genetic targets (the forward primer, the reverse primer, and the probe). This provides high confidence in the fidelity of a finding. In contrast, mass spectrometric assays require the presence of only two genetic targets. Accordingly, we use them as screening tools and follow-up presumptive findings with single agent PCR or Sanger sequencing.

DNA Microarrays

Microarrays are typically 750 × 250 × 1 mm glass slides that have been modified to bind oligonucleotides or peptides for detection of host or microbial nucleic acids or antibodies, respectively. This section of the article will focus on nucleic acid detection. Arrays for serology will be discussed below. Arrays comprise hundreds to millions of probes and can be designed to detect virtually any DNA or RNA target sequence. By varying the length of probes one can differentiate between truly complementary or less closely related sequences. Whereas probes less than 25 nt are specific, those more than 60 nt in length tend to be more promiscuous. Two longer probe array platforms were established for viral surveillance and discovery: the GreeneChip and the Virochip (Wang *et al.*, 2002; Wang *et al.*, 2003; Palacios *et al.*, 2007; Quan *et al.*, 2007). Both employ random amplification strategies prior to hybridization. Because host and viral sequences are amplified with similar efficiencies, the sensitivity for viral detection in tissues is lower than for targeted PCR assays. Host nucleic acid can be reduced by enzymatic digestion (DNA) or ribosomal RNA (rRNA) subtraction; nonetheless, these platforms are most successful with acellular template sources, such as virus cell culture supernatant, serum, plasma, cerebrospinal fluid, or urine. Viral nucleic acid that is hybridized to oligonucleotides on arrays can also be eluted from microarrays to enrich for relevant template prior to sequencing (SARS) (Wang *et al.*, 2003).

High Throughput Sequencing

The introduction in 2005 of automated pyrosequencing by 454 Corporation provided the first insights in the potential power of what has become known as 'Next Generation Sequencing' (NGS). Before the platform was replaced in 2013 by the more efficient Illumina system, pyrosequencing resulted in the discovery and implication of novel viruses in outbreaks of disease in humans,

domestic animals and wildlife including an astrovirus associated with fatal encephalitis in X-linked agammaglobulinemia, disseminated arenavirus infection after organ transplantation, a polyomavirus causing vasculitis with myositis and retinal blindness, LuJo and Bundigbuyo virus associated hemorrhagic fever, and a reovirus that decimated wild and domestic salmon worldwide (Palacios *et al.*, 2010). Whereas 454 pyrosequencing required three days to yield 50,000 sequence reads, hundreds of millions of reads can now be obtained in less than 24 h. The pace of viral discovery has accelerated accordingly, and sequencers are moving into clinical microbiology and public health laboratories worldwide.

In the past several years, improvements in single molecule sequencing platforms have begun to address the potential drawback of the Illumina sequencing system concerning its limited read length (~350–700 bp). Recent sequencing technologies aim at fewer but longer reads, such as the single molecule real-time (SMRT) sequencing from PacBio (Eid *et al.*, 2009), or the MinION from Oxford Nanopore (Lu *et al.*, 2016). The latest Sequel sequencer by PacBio generates ~2 million reads in a mean length range of 15–50 kb (up to ~200 kb). However, it is quite a sizable instrument and takes ~30 h to run. In contrast is the MinION by Oxford Nanopore, a tiny device of ~100 g that is connected to a common USB port of a laptop computer. Data can be assessed in real time after a few minutes of run-time and downloaded once sufficient data for the application purpose are collected, commonly after 12–20 h. The device generates ~10 million reads ranging in length from approx. 10 kb to several hundred kb. The long-read platforms are currently used mostly in combination with short read data to generate the most accurate assemblies with regard to sequence duplications or repetitive elements (Wick *et al.*, 2017; Giddins *et al.*, 2018).

VirCapSeq-VERT

Host and viral nucleic acids compete as template in unbiased high throughput sequencing. In complex samples, such as tissue, blood, nasopharyngeal secretions, and feces, host sequences outnumber viral sequences by orders of magnitude. The relative sparsity of viral sequences in complex samples can lead to detection failure where the number of sequence reads is insufficient to include rarer templates. Thus, to enhance the probability for recovery of viral sequences, investigators typically obtain a minimum of 200 million reads per sample. This approach is resource intensive. It requires not only a substantial investment in sequencing but also in bioinformatic analysis. One strategy to focus and increase sensitivity for viral diagnosis and discovery has focused on enrichment of viral template through subtraction of host nucleic acid via nuclease digestion and depletion of rRNA. Another strategy is positive selection. VirCapSeq-VERT is a positive selection system designed to increase the sensitivity of NGS for detection of vertebrate viral sequences based on viral sequence capture (Briese *et al.*, 2015). The platform comprises a library of biotinylated oligonucleotide probes that tile the coding sequences of all known vertebrate viruses with an inter-oligonucleotide distance of 100–150 nucleotides. cDNA libraries prepared from samples containing viral genetic material are hybridized with the oligonucleotides. The resulting viral nucleic acid/viral probe hybrids are collected using avidin coated magnetic beads and taken forward for sequencing. When compared with unbiased NGS, VirCapSeq-VERT enables a 100 to 1000-fold increase in sensitivity and improved depth of coverage. We find that 2.5 million reads with VirCapSeq-VERT provides similar sensitivity and genome coverage to that obtained with 200 million reads after subtraction (Williams *et al.*, 2018). Bioinformatic complexity and computational resources are reduced accordingly.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Platforms

Bacteria and archaeobacteria have evolved a response to invading bacteriophages that entails specific degradation of viral nucleic acid. Infection results in the integration of DNA fragments of bacteriophages into the bacterial genome as clustered repeats known as CRISPR arrays. Genes encoding CRISPR-associated (Cas) proteins with endonuclease (and in some instances integrase) activity are located in close proximity to CRISPR arrays. Subsequent infection with a similar bacteriophage results in expression from CRISPR arrays and Cas genes of RNA transcripts that target and degrade viral nucleic acid. CRISPR arrays can be genetically engineered to represent virtually any sequence. Thus, when joined with the appropriate Cas enzyme the system can be used to degrade vertebrate viral sequences in addition to bacteriophage sequences. CRISPR Cas systems can also facilitate gene editing by including Cas proteins with both endonuclease and integrase activities. For diagnostic purposes CRISPR Cas systems have been employed in conjunction with isothermal amplification in lateral flow assays and in CMOS arrays for multiplex detection of viruses in environmental and clinical samples (Gootenberg *et al.*, 2018; Freije *et al.*, 2019).

Serology

Serology complements direct detection methods by providing a history of exposure to a viral agent. To date most serology assays have been performed using single agent enzyme-linked immunosorbent assays (ELISA), neutralization assays (NT), western blots (WB), immunofluorescence assays (IFA), haemagglutination assay (HA), or hemagglutination inhibition assay (HAI).

Hemagglutination Inhibition Assays

Serum agglutination, perhaps the first routine diagnostic test, was introduced in the 1800s for diagnosis of typhoid fever (Widal, 1896). Later, agglutination of blood cells was introduced by Landsteiner to determine ABO blood groups (Landsteiner, 1901).

In virology, the hemagglutination (HA) assay is based on the capacity of some viruses responsible for prominent human and animal diseases to bind and cross-link red blood cell surface receptors. The resulting lattice formation deposits as a diffuse reddish layer on the surface of the reaction container. The presence of virus-specific antibodies (Abs) bind the virus and thus inhibit agglutination of red blood cells. Antibodies are quantitated by incubating fixed numbers of red blood cells and virus stock with dilutions of a test serum. Hemagglutination inhibition (HAI) assay are commonly used in typing and subtyping of influenza viruses (Hirst, 1942). Other agglutination assays include latex agglutination (LA) systems where microbeads are coated with Abs or viral antigens.

Plaque Reduction Neutralization (PRNT) Assays

Antibodies can be detected and quantitated based on their capacity to prevent viruses from binding to and infecting their target cells (Pedersen and Spackman, 2014). The readout is a reduction in the number of lytic plaques in a culture of susceptible cells inoculated with the virus of interest. In PRNT as in hemagglutination assays, the requirement for infectious virus poses challenges for many clinical microbiology laboratories, particularly with assays for highly pathogenic viruses. Accordingly, in some instances investigators used less pathogenic viruses such as vesicular stomatitis virus (VSV) where the VSV glycoprotein gene has been exchanged for the glycoprotein of a more pathogenic virus such as ebola (Lee *et al.*, 2017). Because of its native Ag presentation and biological signal readout PRNT is the still the method of choice for measuring functional (protective) immunity. Since even minor variation in protein structure result in changed neutralization behavior, it is also superior to the other methods in distinguishing closely related viruses via cross-neutralization studies. The specificity of PRNT assays is frequently exploited in differentiating immunological reactivity to related viruses. For example, PRNT is considered the gold standard in the differential serodiagnosis of infection with flaviviruses including Zika and dengue viruses (Kuno, 2003; Musso and Gubler, 2016). Only recently, Serochip analyses enabled the characterization of specific peptides that include distinguishing epitopes that now can be transferred other platforms such a peptide ELISA or lateral flow format with the promise of specific identification (Mishra *et al.*, 2018). PRNT assays can also be employed to identify virus isolates through use of panels of hyperimmune sera generated in experimentally infected animals, and to group viruses taxonomically (Casals, 1957; The Enteroviruses, 1957; Schmidt *et al.*, 1961; Muir *et al.*, 1998).

Immunofluorescent Assays (IFA), Enzyme-Linked Immunosorbent Assays (ELISA) and Immunoblot Assays

Each of these methods build on the specific binding between Ab and its matching Ag. In IFA, cultured cells are infected with virus, fixed on glass slides or in multi-well plates with acetone or formaldehyde, and incubated with the test serum, plasma or CSF. A reporter dye linked to a secondary Ab directed against the primary (test) Ab is then introduced prior to fluorescence microscopy. Secondary antibodies can also be conjugated to enzymes that yield products that can be visualized by light microscopy.

Enzyme-based dye formation is also the basis for the ELISA (Engvall and Perlmann, 1971; Van Weemen and Schuurs, 1971), the most common assays used for antibody detection. Amongst the various platforms, the ELISA is the most versatile and popular. This is in part because of the multitude of ways to present Ags or Abs and the comparatively high throughput in 96-well format. Another advantage is the ability to present targets in native form supporting the detection of a wide variety of epitopes. The ELISA is typically performed in 96-well plates wherein wells coated with a viral antigen are incubated with serially diluted patient body fluid, a reporter enzyme (usually peroxidase or alkaline phosphatase) directly conjugated to the serum immunoglobulins (direct ELISA) or a secondary antibody (indirect ELISA), and finally a dye system that generates upon enzymatic catalysis a soluble colored product that can be read spectrophotometrically. A common variation is the sandwich ELISA wherein the well-surface is first coated with Abs specific for the Ag (capture Ab) before the antigen is applied. Next, the detection Ab is applied, either in direct or in indirect format. The advantage of the sandwich ELISA is that it increases the sensitivity and specificity of the assay.

Western blot assays (also known and immunoblots) are typically performed including complex Ag preparations such as crude infected cell extracts that are size fractionated by electrophoresis in sodium dodecyl sulfate polyacrylamide (SDS-PAGE) gels, transferred onto nitrocellulose or polyvinylidene difluoride membranes, and then incubated with sera. Immunoreactivity of antibodies in sera are detected using a species-specific secondary antibody that is linked to an enzyme that results in either a dye or chemiluminescence signal. Proteins subjected to SDS-PAGE are denatured. Thus, they are not designed to detect conformational epitopes that may be critical in viral neutralization.

Lateral Flow Assays (LFA)

These assays are designed for point of care serodiagnosis and field applications. In LFAs, the liquid sample (plasma, serum, urine or CSF) flows by capillary action through a membrane where the antigen or antibody in the sample binds to a dye-labeled cognate marker (antigen or antibody). The antigen-antibody-dye marker complex continues to migrate through the membrane until further migration is blocked by binding of the complex to an antibody to a component of the complex that is fixed in position, the appearance of a line of dye indicates that the antigen or antibody was present in the sample.

Multiplex Platforms for High Throughput Serology

In contrast to molecular diagnostics where advances in technology such as PCR and HTS have dramatically improved sensitivity, specificity, and capacity for multiplex analyses, serologic methods remained largely unchanged. This lag is important given the role of serology in establishing the distribution and frequency of infection, testing the significance of association between the finding of an agent and disease, focusing efforts in pathogen discovery and surveillance, and for monitoring humoral responses to vaccines and immunomodulatory drugs. Most serological tests target only one agent; however, Luminex-based systems have been employed that can address up to 100 targets simultaneously.

Phage display systems and arrays that comprise spotted recombinant proteins expressed *in vitro* in *E. coli*, *S. cerevisiae*, baculoviruses, and cell-free, coupled transcription-translation, have also been established. The VirScan platform comprehensively tests for the presence of antiviral antibodies through immunoprecipitation of viral peptides in bacteriophage followed by DNA sequencing. The current version screens 93,904 200-mer oligonucleotides, encoding 56-residue peptide that tile the proteome of 206 viral species with 28 residue overlap (Xu *et al.*, 2015). The phage library is incubated with serum, plasma, or CSF. Bound antibodies are recovered by using a mixture of protein A and G coated magnetic beads. The bound phage DNA is sequenced and quantitated to identify immunoreactive epitopes. These assays are powerful but require expertise and resources for cDNA cloning, expression, sequencing, and bioinformatic analysis. An additional challenge is that VirScan may not have the resolution required to differentiate immune responses to closely related viruses. Peptide arrays (Serochips) are an alternative high throughput platform for unbiased serology. Serochips comprise up to 6 million distinct linear peptide sequences attached to a solid support. Like the VirScan platform, Serochips display only linear peptides; thus, neither platform will detect conformational or glycosylated epitopes. To ensure our ability to detect immune responses to discriminant as well as common viral epitopes we tile the entire proteome of the relevant viruses using 12 aa peptides with 11 aa overlap. After incubation with serum, plasma or CSF, arrays are incubated with fluorescence-labeled secondary antibodies (IgG or IgM). Finally, high-resolution image scans are obtained and relative fluorescence signal intensity is used to map immunoreactivity to specific viruses and identify peptides that can be transferred to less expensive and simpler methods for serodiagnosis such as ELISA, western blot, Luminex assays, or smaller peptide arrays.

The detection of an IgM versus an IgG response is important to certain aspects of functional serology. During the challenge of the immune system with a new Ag, stimulated B-cells produce their membrane-presented Ab first as pentameric immunoglobulin class M (IgM) isotype antibodies, that usually exhibit limited specificity. During the next 2–4 weeks plasma B-cells transition after isotype switching and hypermutation to the production of highly specific monomeric immunoglobulin G (IgG) isotype antibodies (Nossal, 2007; Tonegawa, 1983). The ratio of IgM to IgG response may therefore be helpful in determining the timing of the infection. IgM vs IgG determination is easily achieved in all indirect assay systems through the choice of the secondary Ab, applying either an anti-human IgM or an anti-human IgG-specific Ab. There are many other considerations directing the choice of a particular serologic platform. This often includes the classical paradigm of confirming a positive result in one assay in a second technically independent assay, as it is still common practice in Lyme diagnosis; after initial screening by ELISA positive results are confirmed by immunoblot reaction patterns (Centers for Disease Control and Prevention, 1995; Tokarz *et al.*, 2018).

Discussion/Summary

Viral diagnostics is no longer a discipline that requires resources and expertise available only in reference laboratories. Although culture and electron microscopy remain important tools for basic and translational research, molecular methods have moved to the forefront of diagnostics in clinical microbiology and public health. This reflects not only lower costs and ease of use but also the capacity of molecular methods to yield actionable information in hours rather than days. Serology too is becoming more robust and convenient with the advent of multiplexed assays that provide specific information without virus cultivation and neutralization assays. Advances in molecular and serological methods allow clinicians and public health practitioners to simultaneously consider a wide range of probable candidate viral pathogens, and to detect the presence of pathogens that are not typically anticipated in a differential diagnosis. These include viruses known to cause disease in other geographic regions or in other species, or viruses not previously known to medicine or science. We have not addressed in this article the potential for host genetic, proteomic, metabolomic or microbiome analyses to provide diagnostic and prognostic information independent of virus identification that will impact interventions. This is an area of active research that we anticipate will be featured in future editions of this encyclopedia. We also anticipate that insights from wider application of molecular and serological tools will encourage the development of drugs and vaccines that reduce the burden of viral infections in both acute and chronic disease.

References

- Al-Samarrai, T., Wu, W., Begier, E., *et al.*, 2013. Evaluation of a pilot respiratory virus surveillance system linking electronic health record and diagnostic data. *Journal of Public Health Management & Practice* 19, 322–329.
- Breman, J.G., Heymann, D.L., Lloyd, G., *et al.*, 2016. Discovery and description of Ebola Zaire virus in 1976 and relevance to the West African epidemic during 2013–2016. *The Journal of Infectious Diseases* 214 (suppl 3), S93–S101. doi:10.1093/infdis/jiw207.

- Briese, T., Jia, X.Y., Huang, C., Grady, L.J., Lipkin, W.I., 1999. Identification of a Kunjin/West Nile-like flavivirus in brains of patients with New York encephalitis. *Lancet* 354 (9186), 1261–1262.
- Briese, T., Kapoor, A., Mishra, N., *et al.*, 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio* 6 (5), e01491-15.
- Briese, T., Palacios, G., Kokoris, M., *et al.*, 2005. Diagnostic system for rapid and sensitive differential detection of pathogens. *Emerging Infectious Diseases* 11 (2), 310–313.
- Brunstein, J., Thomas, E., 2006. Direct screening of clinical specimens for multiple respiratory pathogens using the Genaco respiratory panels 1 and 2. *Diagnostic Molecular Pathology* 15, 169–173.
- Casals, J., 1957. The arthropod-borne group of animal viruses. *Transactions of the New York Academy of Sciences* 19 (3), 219–235.
- Centers for Disease Control and Prevention, 1995. Recommendations for test performance and interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease. *Morbidity and Mortality Weekly Report* 44, 590–591.
- Dominguez, S.R., Briese, T., Palacios, G., *et al.*, 2008. Multiplex MassTag-PCR for respiratory pathogens in pediatric nasopharyngeal washes negative by conventional diagnostic testing shows a high prevalence of viruses belonging to a newly recognized rhinovirus clade. *Journal of Clinical Virology* 43, 219–222.
- Ecker, D.J., Sampath, R., Blyn, L.B., *et al.*, 2005. Rapid identification and strain-typing of respiratory pathogens for epidemic surveillance. *Proceedings of the National Academy of Sciences of the United States of America* 102 (22), 8012–8017.
- Ecker, D.J., Sampath, R., Massire, C., *et al.*, 2008. Ibis T5000: A universal biosensor approach for microbiology. *Nature Reviews Microbiology* 6 (7), 553–558.
- Eid, J., Fehr, A., Gray, J., *et al.*, 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323 (5910), 133–138.
- Engvall, E., Perlmann, P., 1971. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry* 8 (9), 871–874.
- Freije, C.A., Myhrvold, C., Boehm, C.K., *et al.*, 2019. Programmable inhibition and detection of RNA viruses using Cas13. *Molecular Cell* 76 (5), 826–837.
- Giddins, M.J., Macesic, N., Annavaiah, M.K., *et al.*, 2018. Successive emergence of ceftazidime-avibactam resistance through distinct genomic adaptations in blaKPC-2-Harboring *Klebsiella pneumoniae* sequence type 307 isolates. *Antimicrobial Agents and Chemotherapy* 62, 3.
- Goldsmit, C.S., Miller, S.E., 2009. Modern uses of electron microscopy for detection of viruses. *Clinical Microbiology Reviews* 22 (4), 552–563.
- Gootenberg, J.S., Abudayyeh, O.O., Kellner, M.J., *et al.*, 2018. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* 360 (6387), 439–444.
- Han, J., Swan, D.C., Smith, S.J., *et al.*, 2006. Simultaneous amplification and identification of 25 human papillomavirus types with Templex technology. *Journal of Clinical Microbiology* 44, 4157–4162.
- Haycox, C.L., Kim, S., Fleckman, P., *et al.*, 1999. Trichodysplasia spinulosa – A newly described folliculocentric viral infection in an immunocompromised host. *Journal of Investigative Dermatology Symposium Proceedings* 4, 268–271.
- Hirst, G.K., 1942. The quantitative determination of Influenza virus and antibodies by means of red cell agglutination. *Journal of Experimental Medicine* 75 (1), 49–64.
- Hsu, C.C., Tokarz, R., Briese, T., *et al.*, 2013. Use of staged molecular analysis to determine causes of unexplained central nervous system infections. *Emerging Infectious Diseases* 19, 1470–1477.
- Kapikian, A.Z., Wyatt, R.G., Dolin, R., *et al.*, 1972. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *Journal of Virology* 10 (5), 1075–1081.
- Kausche, G.A., Plankuch, E., Ruska, H., 1939. Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop. *Naturwissenschaften* 27 (18), 292–299. doi:10.1007/BF01493353.
- Kuno, G., 2003. Serodiagnosis of flaviviral infections and vaccinations in humans. *Advances in Virus Research* 61, 3–65.
- Lamson, D., Renwick, N., Kapoor, V., *et al.*, 2006. MassTag polymerase-chain-reaction detection of respiratory pathogens, including a new rhinovirus genotype, that caused influenza-like illness in New York State during 2004–2005. *Journal of Infectious Diseases* 194, 1398–1402.
- Landsteiner, K., 1901. Ueber Agglutinationserscheinungen normalen menschlichen Blutes. *Wiener klinische Wochenschrift* 14, 1132–1134.
- Lee, S.S., Phy, K., Peden, K., Sheng-Fowler, L., 2017. Development of a micro-neutralization assay for ebolaviruses using a replication-competent vesicular stomatitis hybrid virus and a quantitative PCR readout. *Vaccine* 35 (41), 5481–5486.
- Li, H., McCormac, M.A., Estes, R.W., *et al.*, 2007. Simultaneous detection and high-throughput identification of a panel of RNA viruses causing respiratory tract infections. *Journal of Clinical Microbiology* 45, 2105–2109.
- Lipkin, W.I., 2010. Microbe hunting. *Microbiology and Molecular Biology Reviews* 74 (3), 363–377.
- Long, G.W., Nobel Jr, J., Murphy, F.A., Herrmann, K.L., Lourie, B., 1970. Experience with electron microscopy in the differential diagnosis of smallpox. *Journal of Applied Microbiology* 20 (3), 497–504.
- Lu, H., Giordano, F., Ning, Z., 2016. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics. Bioinformatics* 14 (5), 265–279.
- Mishra, N., Caciula, A., Price, A., *et al.*, 2018. Diagnosis of Zika virus infection by peptide array and enzyme-linked immunosorbent assay. *mBio* 9, e00095. [18].
- Muir, P., Kämmerer, U., Korn, K., *et al.*, 1998. Molecular typing of enteroviruses: current status and future requirements. The European Union concerted action on virus meningitis and encephalitis. *Clinical Microbiology Reviews* 11 (1), 202–227.
- Musso, D., Gubler, D.J., 2016. Zika virus. *Clinical Microbiology Reviews* 29 (3), 487–524.
- Nicholls, J.M., Poon, L.L., Lee, K.C., *et al.*, 2003. Lung pathology of fatal severe acute respiratory syndrome. *Lancet* 361 (9371), 1773–1778.
- Nossal, G., 2007. One cell-one antibody: Prelude and aftermath. *Nature Immunology* 8 (10), 1015–1017.
- Palacios, G., Briese, T., Kapoor, V., *et al.*, 2006. MassTag polymerase chain reaction for differential diagnosis of viral hemorrhagic fever. *Emerging Infectious Diseases* 12, 692–695.
- Palacios, G., Hornig, M., Cisterna, D., *et al.*, 2009. Streptococcus pneumoniae coinfection is correlated with the severity of H1N1 pandemic influenza. *PLOS One* 4, e8540.
- Palacios, G., Lovoll, M., Tengs, T., *et al.*, 2010. Heart and skeletal muscle inflammation of farmed salmon is associated with infection with a novel reovirus. *PLOS One* 5 (7), e11487.
- Palacios, G., Quan, P.L., Jabado, O.J., *et al.*, 2007. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerging Infectious Diseases* 13 (1), 73–81.
- Pedersen, J.C., 2014. Hemagglutination-inhibition assay for influenza virus subtype identification and the detection and quantitation of serum antibodies to influenza virus. In: Spackman, E. (Ed.), *Animal Influenza Virus. Methods in Molecular Biology (Methods and Protocols)* 1161. New York, NY: Humana Press.
- Quan, P.L., Palacios, G., Jabado, O.J., *et al.*, 2007. Detection of respiratory viruses and subtype identification of influenza A viruses by GreeneChipResp oligonucleotide microarray. *Journal of Clinical Microbiology* 45 (8), 2359–2364.
- Renwick, N., Schweiger, B., Kapoor, V., *et al.*, 2007. A recently identified rhinovirus genotype is associated with severe respiratory-tract infection in children in Germany. *Journal of Infectious Diseases* 196, 1754–1760.
- Saiki, R.K., Scharf, S., Faloona, F., *et al.*, 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230 (4732), 1350–1354.
- Sampath, R., Hall, T.A., Massire, C., *et al.*, 2007. Rapid identification of emerging infectious agents using PCR and electrospray ionization mass spectrometry. *Annals of the New York Academy of Sciences* 1102, 109–120.
- Schmidt, N.F., Guenther, R.W., Lennette, E.H., 1961. Typing of ECHO virus isolates by immune serum pools. The “intersecting serum scheme. *Journal of Immunology* 87, 623–626.
- The Enteroviruses, 1957. Committee on the Enteroviruses, National Foundation for infantile paralysis. *American Journal of Public Health and the Nation's Health* 47 (12), 1556–1566.
- Tokarz, R., Firth, C., Madhi, S.A., *et al.*, 2012. Worldwide emergence of multiple clades of enterovirus 68. *Journal of General Virology* 93, 1952–1958.
- Tokarz, R., Jain, K., Bennett, A., Briese, T., Lipkin, W.I., 2010. Assessment of polymicrobial infections in ticks in New York state. *Vector-Borne and Zoonotic Diseases* 10, 217–221.

- Tokarz, R., Kapoor, V., Samuel, J.E., *et al.*, 2009. Detection of tick-borne pathogens by MassTag polymerase chain reaction. *Vector-Borne and Zoonotic Diseases* 9, 147–152.
- Tokarz, R., Kapoor, V., Wu, W., *et al.*, 2011. Longitudinal molecular microbial analysis of influenza-like illness in New York City, May 2009 through May 2010. *Virology Journal* 8, 288.
- Tokarz, R., Mishra, N., Tagliafierro, T., *et al.*, 2018. A multiplex serologic platform for diagnosis of tick-borne diseases. *Scientific Reports* 8 (1), 3158.
- Tonegawa, S., 1983. Somatic generation of antibody diversity. *Nature*. 302 (5909), 575–581.
- Wang, D., Coscoy, L., Zylberberg, M., *et al.*, 2002. Microarray-based detection and genotyping of viral pathogens. *Proceedings of the National Academy of Sciences of the United States of America* 99 (24), 15687–15692.
- Wang, D., Urisman, A., Liu, Y.-T., *et al.*, 2003. Viral discovery and sequence recovery using DNA microarrays. *PLOS Biology* 1 (2), e2.
- Van Ert, M.S., Hofstadler, Y., Jiang, J., *et al.*, 2004. Mass spectrometry provides accurate characterization of two genetic marker types in *Bacillus anthracis*. *BioTechniques* 37 (4), 642.
- Van Rooyen, C.E., Illingworth, R.S., 1944. A laboratory test for diagnosis of smallpox. *British Medical Journal* 2 (4372), 526–529. doi:10.1136/bmj.2.4372.526.
- Van Weemen, B.K., Schuurs, A.H.W.M., 1971. Immunoassay using antigen-enzyme conjugates. *FEBS Letters* 15 (3), 232–236.
- Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology* 13 (6), e1005595.
- Widal, F.M., 1896. Serodiagnostic de la fièvre typhoïde a-propos d'une modification par MMC Nicolle et al. Halipie. *Bulletins et mémoires de la Société Médicale des Hôpitaux de Paris* 13, 561–566.
- Williams, S.H., Cordey, S., Bhuvu, N., *et al.*, 2018. Investigation of the plasma virome from cases of unexplained febrile illness in Tanzania from 2013–2014: A comparative analysis between unbiased and VirCapSeq-VERT high-throughput sequencing approaches. *mSphere* 3 (4), e00311–e00318.
- Xu, G.J., Kula, T., Xu, Q., *et al.*, 2015. Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. *Science*. 348 (6239), aaa0698.

Clinical Diagnostic Virology

Marcus Panning, Institute of Virology, Freiburg University Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany

© 2021 Elsevier Ltd. All rights reserved.

Glossary

DNA Deoxyribonucleic acid.

Evidenced based medicine Uses the currently available best evidence in making decisions about the clinical management of individual patients.

Multiplex PCR An assay to detect different pathogens/targets simultaneously in a single or limited number of reaction tubes.

Nucleic acid testing (NAT) The direct amplification and detection of DNA or RNA (after reverse transcription prior to PCR).

Point-of-Care-tests Rapid tests that can be performed close the patient with minimal hands-on time and ease of use.

Polymerase chain reaction (PCR) Enzymatic thermocyclic amplification procedure for DNA.

Randomized controlled trial A study that allocates enrolled patients randomly to either one of two arms: one intervention arm and one control arm. It thus eliminates bias.

Reverse-transcription PCR For RNA, nucleic acids are transcribed into DNA using reverse transcriptase enzymes.

RNA Ribonucleic acid.

Introduction

Traditionally, clinical diagnostic virology methods included virus isolation using cell culture, antigen detection, and antibody detection methods. Until recently, these conventional methods were rather slow, labor-intensive and only available for a limited number of viral infections. In most instances diagnosis was retrospective. Thus, clinical diagnostic virology had little impact on patient management and clinical decision-making. However, clinical diagnostic virology has been revolutionized with the advent of nucleic acid testing (NAT) more than three decades ago. Molecular methods allow the detection of minute amounts of viral genomes and rapidly became indispensable tools for the clinical diagnostic laboratory. Target-specific amplification of (parts) of the viral DNA or RNA genome using polymerase chain reaction (PCR or reverse-transcription PCR for RNA viruses) is the most common technique used for NAT, but other technologies such as nucleic acid sequence-based amplification (NASBA), transcription-mediated amplification (TMA), or loop-mediated isothermal amplification (LAMP) have also found their niche. Beyond the pure detection of viral pathogens, NAT also provides a means to quantitate the amount of viral DNA/RNA ("viral load"). This was primarily made possible with the introduction of real-time quantitative PCR. Moreover, PCR combined with sequencing capabilities provides a tool to genotype and characterize a virus, i.e., to look for resistance-associated mutations in the viral genome. Technically, next-generation sequencing technologies are rapidly evolving and characterize viral genomes with unprecedented depth and accuracy. In combination with the increasing availability of antiviral therapeutics clinical diagnostic virology thus became essential for optimized patient care, e.g., in the context of viral load monitoring of HIV-, HBV- or HCV-infected patients or monitoring of CMV, EBV and other relevant viruses in immunosuppressed patients as well as for resistance testing. Consequently, with the advent of NAT many traditional "reference" methods became obsolete. It should be noted, however, that even in the era of NAT a mix of different techniques remains essential for the clinical virology laboratory to come to a valid and reliable diagnosis.

Compared to conventional virological techniques NAT is faster, more sensitive and available for a wider range of pathogens. However, the first NAT assays were laboratory-based and still had turn-around times of up to 48 h. If testing for multiple viruses was required, this was mainly done in a sequential and time-consuming manner. In the next step, laboratories developed in-house assays that used the same PCR reaction components and the same thermal cycling profile. This allowed parallel testing for multiple targets but still required separate nucleic acid extraction and PCR set-up steps (Bonzel *et al.*, 2008; Gunson *et al.*, 2008). With constant technical improvement, multiplex panels for the simultaneous detection of a variety of pathogens in a single reaction have been introduced into routine diagnostics recently. This approach is also called "syndromic testing approach" and is evolving rapidly. Here, a simultaneous search for multiple pathogens causing a similar clinical picture is performed. Some of these assays require minimal hands-on time by using single-use cartridges and allow for near-patient testing (also termed "point-of-care testing" = POCT). Time to result may be as low as <2 h on these integrated platforms. A number of different syndromic panels are now commercially available, e.g., for respiratory tract infections, gastrointestinal infections, and meningitis/encephalitis. Complementary to this, fully automated random-access platforms combining nucleic acid extraction and amplification/detection are now on the horizon for numerous pathogens, e.g., CMV, EBV, HIV, HCV, HBV and will certainly impact clinical diagnostic virology. Therefore, it is anticipated that rapid and near-patient testing will again constitute a dramatic change for the field of clinical diagnostic virology and may directly impact patient management. This includes but is not limited to the administration of (new) antiviral drugs (if available), appropriate use of antibiotics, and rapid implementation of infection control measures. This is accompanied by a variety of novel next-generation sequencing approaches entering diagnostic laboratories allowing "personalized medicine". However, although many of these novel

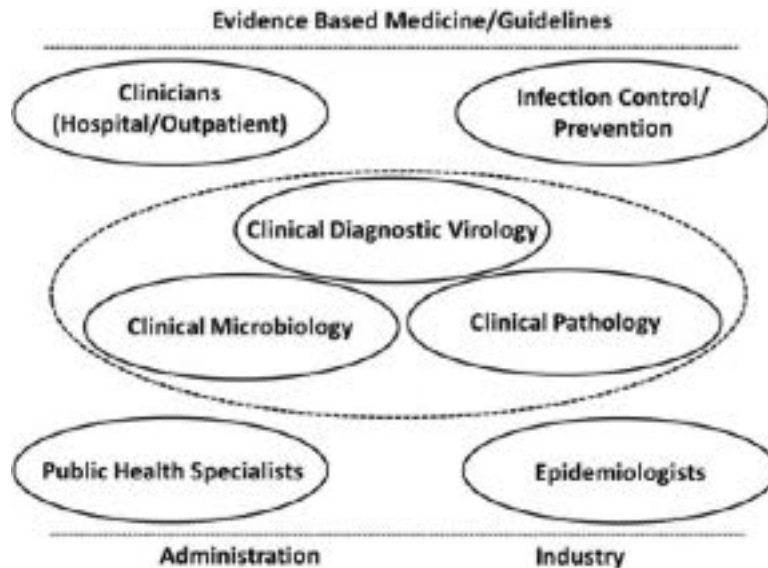


Fig. 1 Interdisciplinary network of clinical diagnostic virology.

techniques demonstrated very good diagnostic performance and accuracy the clinical impact remains unclear to date. It is mandatory that beyond many already published studies high-quality and randomized controlled trials will shed more light on the question if these new technologies will provide a benefit for the patient and will improve outcomes without doing harm. Another challenge remains the currently high costs of these novel assays, the optimal ordering strategy, rapid communication and interpretation of test results in the context of clinical disease and lastly quality control issues.

Clinical diagnostic virology is truly interdisciplinary and requires a close partnership of infectious diseases clinicians, clinical virologists, microbiologists, laboratory/bioinformatics specialists, epidemiologists, infection control specialists, and public health members (**Fig. 1**). Foremost, close interaction with clinicians is essential to serve their needs and in turn for clinicians to understand test characteristics, result in interpretations and their utilization. This is especially true when new test panels will be introduced, which should be accompanied by performance evaluation and ideally cost-effectiveness calculation by the laboratory. Thus, clinical diagnostic virology should be an integrated part of antimicrobial and diagnostic stewardship programs ([Dik et al., 2017](#)). In this respect, clinical diagnostic virology should interact with all stakeholders, e.g., administration and purchasing department and develop solutions to the best of our patients. This may differ from country to country and even within countries with a very dynamic landscape of university-affiliated laboratories, large non-academic laboratories and those serving for basic and regular care. It is obvious that laboratories may serve different patient populations (e.g., number of immunosuppressed and immunocompetent patients, intensive care units, high or low throughput) and have different requirements (e.g., proximity to a hospital). In addition, large healthcare group laboratories may have their own needs and develop different approaches.

Beyond primary patient care, clinical diagnostic virology also contributes to translational research and the implementation/evaluation of emerging technologies and their transition into clinical care. Having this in mind data sharing to assess the burden of diseases and to better appreciate the epidemiology of viral infections seems mandatory on a national as well as international level.

This article will briefly describe examples of syndromic testing approaches currently used in clinical diagnostic virology and how they might impact clinical decision making and the currently available evidence supporting this. Further, the impact of clinical diagnostic virology on current knowledge of hepatitis E virus infection will be briefly discussed as an example of an emerging infection and translational research.

Respiratory Tract Infections

Background

Acute respiratory tract infections (RTI) have a high disease burden worldwide and contribute significantly to mortality. In general, these acute RTI have a viral etiology and less common a bacterial one. Of note, acute RTI can be caused by a variety of respiratory viruses from different virus families including, e.g., influenza virus, respiratory syncytial virus (RSV) A/B, coronavirus, human metapneumovirus, and rhino-/enterovirus. Seasonal influenza alone causes up to 650,000 deaths worldwide and enormous socioeconomic losses each year. Clinical presentation can range from mild upper RTI to severe lower RTI and pneumonia. Of note, specific populations are at increased risk for severe clinical courses, e.g., immunosuppressed patients, HIV-infected patients, pregnant women, and the elderly. Except for viruses, atypical bacteria like *Bordetella pertussis*, *Mycoplasma pneumonia* and *Legionella*

Table 1 Selection of currently available commercial multiplex panels from major companies

Manufacturer	Respiratory panel	Gastrointestinal panel	Meningitis/encephalitis panel
Luminex	xTAG RVP ^a xTAG RVP Fast/FASTv2 ^a NxTAG RPP ^a	xTAG GPP ^a	–
BioMérieux	Verigene respiratory pathogens flex FilmArray Respiratory (RP) Panel	Verigene enteric pathogens test FilmArray Gastrointestinal (GI) Panel	– FilmArray Meningitis/Encephalitis (ME) Panel
Fast track diagnostics	FTD respiratory pathogens 21 ^a , 21plus ^a , 33 ^a	FTD Viral Gastroenteritis ^a , Bacterial gastroenteritis ^a , Stool parasites ^a	FTD Viral meningitis ^a , Bacterial meningitis ^a
Qiagen	QIAstat-Dx Respiratory Panel	QIAstat-Dx Gastrointestinal Panel	–
Genmark	ePlex Respiratory Pathogen Panel	–	–
Seegene	Allplex respiratory panel 1–4 ^a	Allplex GI-Virus assay ^a	Allplex meningitis-V1/V2 assay ^a
Hologic	Panther fusion respiratory assays	–	–

^aNon-cartridge/pouch-based systems.

pneumophila can also cause acute lower RTI/pneumonia. These bacteria are not very well detected by standard culture-based microbiological techniques and are therefore frequently included in modern molecular panels. Thus, testing for atypical bacteria should complement the diagnosis of acute RTI. Suitable clinical samples include nasopharyngeal swabs and aspirates/washings, pharyngeal swabs, high-quality sputum, and bronchoalveolar lavage. Limitations apply as per the recommendation of the manufacturer of commercial assays as some assays have been only validated for a specific specimen type.

The vast majority of (pediatric) RTI cases are seen by primary care physicians. However, it is currently impossible for primary care physicians to distinguish viral etiologies of an acute RTI and even more important to distinguish a viral from a bacterial etiology on clinical grounds only. This makes a laboratory-based diagnosis mandatory for a definite diagnosis. Critically, antibiotics are prescribed in >65% of consultations due to RTI although viruses cause the vast majority of RTI. Rapid identification of the etiology may lead to optimized patient care and treatment and may also reduce overall costs. From a public health perspective antibiotic resistance is an emerging problem worldwide and overuse/not appropriate use of antibiotics is therefore of great concern. In addition, viral infections like influenza can be treated with specific antiviral drugs, e.g., oseltamivir, but these drugs need to be given quite early in the course of the disease. Interestingly, the use of broadly reactive multiplex panels already allowed new insights into the epidemiology of previously underestimated viral infections like coronaviruses. Finally, a rapid diagnosis may also prove useful for better infection control and prevention within the hospital setting.

Assays

In recent years, a paradigm shift has been observed for the detection of pathogens causing acute RTI. The first multiplex PCR assays were mostly laboratory-developed tests (LDT) and were mainly based on real-time PCR technology. One of the first commercial assays for the comprehensive detection of RTI pathogens was developed by Luminex Corporation. However, these assays still required separate nucleic acid extraction steps and the time to result was somehow in the range of around 8 h. Technological advancements led to the development of cartridge-based systems which combined nucleic acid extraction and PCR amplification/detection allowing to test for a battery of different pathogens with minimal hands-on time within less than two hours. Different systems are now on the market and commercially available, e.g., FilmArray, ePlex or Verigene (Table 1). Popowitch *et al.* compared four different multiplex panels and determined overall sensitivity and specificity, ranging between 84.5% and 100%, respectively (Popowitch *et al.*, 2013). It is important to note that sensitivity for different targets can vary and that laboratories are aware of the performance data and possible limitations of the assays they perform. Of note, selected pathogens yielded considerably lower sensitivities e.g., adenovirus and influenza virus in some studies, which is an issue of concern (Popowitch *et al.*, 2013). Another study compared the Genmark ePlex RPP assay with LDT and showed a high overall agreement of 97.4% rendering them suitable for routine diagnostics (Nijhuis *et al.*, 2017). Some of these assays already received FDA/CE marks and a recent study showed a high diagnostic accuracy of these multiplex panels (Huang *et al.*, CMI 2018).

Limiting their widespread implementation are considerable higher costs and presently limited availability of different panels on these platforms. However, recent experience has shown that the costs of multiplex panels are declining as more and more companies enter the market with newly developed panels. In addition, pending issues concerning quality control, reimbursement, and notifiable diseases have to be discussed in some countries.

Critically, molecular assays need constant re-evaluation and close monitoring of new or mutated viral strains, which might affect assay performance. In 2009, a new influenza A virus emerged and challenged many established detection systems. In addition, e.g., influenza drift variants may go undetected if primer and/or probe binding regions are affected. It is a well-known phenomenon that especially RNA viruses are prone to genetic changes. Similarly, for enteroviruses and adenoviruses, a diversity of different types/genotypes is known, and comprehensive detection of all members is challenging. It should also be noted that these panels include fixed combinations of pathogens and usually do not allow them to be customized. Prevalence of RTI pathogens vary in time and

geographically and thus affect positive and negative predictive values. It is mandatory to record positivity rates for all detected pathogens for quality control purposes. The likelihood of false-positive results must be taken into account and might require alternative testing with individual assays. Thus, the laboratory must be aware of possible pitfalls of the assays in use (van Zyl *et al.*, 2019). Strict operating procedures must be followed to minimize the risk of contamination. This holds especially true when near-patient testing is done by non-laboratory personnel outside dedicated molecular diagnostic facilities. Further, NAT detects nucleic acids only and at present, it is impossible to assess infectivity/viability. Prolonged shedding of viral RNA has been described especially for entero-/rhinoviruses in the immunocompromised host. Finally, the role of co-detections is unclear to date. Some studies suggest that co-detections might be more severe compared to mono-infections whereas other studies did not show any differences.

Clinical Impact

Although multiplex panels have been introduced by many laboratories worldwide within the last years the clinical impact of these rapid multiplex panels remains to some extent unclear to date. The first studies mostly used a before/after approach to analyze the impact after the introduction of multiplex testing approaches. Not surprisingly, these studies demonstrated a significant decrease in time to diagnosis (Rappo *et al.*, 2016). Another study reported a significant increase in the number of detected pathogens. However, limitations apply as viral culture or other rather slow conventional methods were used in the prior season. Proper administration of antiviral therapies is another important issue in acute RTI. Vos *et al.* analyzed the impact of rapid viral testing in adult immunosuppressed patients and were able to show a more adequate use of oseltamivir (Vos *et al.*, 2019b).

Another recent systematic review and meta-analysis also yielded inconclusive results due to a very heterogenous study landscape (Vos *et al.*, 2019a). In theory, the rapid identification of a viral etiology may help reduce unnecessary antibiotic treatment and initiate specific antiviral therapy, e.g., in influenza. Moreover, new antiviral therapies are in the pipeline of pharmaceutical companies, e.g., for RSV. The implementation of rapid molecular testing in the emergency department resulted in a more targeted approach to use oseltamivir treatment in immunocompromised patients. However, in the same study, no reduction in antibiotic prescriptions or its duration, admission and length of stay was seen. It is anticipated that new studies will shed more light on the clinical impact of rapid near-patient testing and will help to better appreciate their usefulness for patient care. Ideally, randomized controlled trials will be initiated to answer these urgent questions. It may also influence the decision to admit patients, reduce the length of stay, and avoid antimicrobials.

Costs are also an important factor requiring careful consideration. As an example, multiplex panels showed to be cost-effective in the pediatric setting when children with influenza-like illness in the emergency department were analyzed. However, the estimation of costs and possible savings remains complex and varies from country to country. Beyond laboratory costs (which might increase due to higher test costs and extended availability of personnel) costs might decrease overall for the hospital. Recently, the €hr concept was introduced to analyze this in more detail (Dik *et al.*, 2016).

To summarize, the clinical benefit of rapid syndromic testing for the patient is to some extent unclear to date. A differentiated approach might be necessary taking into account e.g., the severity of the disease and the immune status of the patient. Syndromic testing might be beneficial for the immunocompromised host, but a more targeted approach might be necessary for the otherwise healthy outpatient and for children. It might be necessary to establish local testing algorithms for specific patient groups to use these promising new technologies properly. This may also include different testing approaches over the year to account for seasonal variations in the prevalence of RTI pathogens.

Gastrointestinal Infections

Background

Similar to the situation with acute RTI, gastrointestinal infections (GI) are also associated with a high disease burden worldwide. An estimated 2 billion cases occur each year and they constitute a major cause of emergency department visits and hospitalizations worldwide (Farthing *et al.*, 2013). In particular children in resource-limited countries are affected and mortality in these settings can be very high. It is estimated that around 2 million deaths each year occur in developing countries. On the contrary, in industrialized countries, GI infections also have a great impact on individual and public health but to a lesser extent on mortality. However, although most infections are usually mild and self-limited in industrialized settings, severe courses and sometimes fatal outcomes are seen in vulnerable populations like the elderly and immunocompromised persons. Of note, outbreaks have been described, e.g., in the hospital setting, in nursing homes or on cruise ships. Transmission occurs via contaminated food or water or close contact with an acutely infected person. In addition, an aerosol transmission has been described for noroviruses. Relevant pathogens for GI are of viral, bacterial or parasitic origin. Traditional methods for the detection of GI pathogens include laborious and time-consuming microscopy, culture and antigen detection methods. The time to final diagnosis can take days. In addition, all of these techniques are less sensitive and specific compared to NAT. Timely diagnosis may prevent further spread and adverse patient outcomes and selected patients

may benefit from rapid and appropriate antibiotic treatment. Isolation and/or cohorting of patients with norovirus associated GI is important in the health care setting.

Assays

Numerous LDT, as well as commercial assays, are available for single pathogens but similar to RTI multiplex panels are now entering the market. Some of these platforms allow for rapid and near-patient testing using single-use cartridges or pouches (Table 1). In analogy to RTI panels different viral, bacterial and parasitic targets are included and allow for a “syndromic” testing approach. Most assays include more bacterial/parasitic targets than viral targets. Comparison studies have demonstrated equivalent or even superior performance to reference assays. In these studies, ≥ 2 pathogens were detected significantly more frequently throughout different studies compared to conventional methods. Major viral targets included in these multiplex panels are adenovirus 40/41, astrovirus, norovirus genotype 1/2, rotavirus A, and in some assays sapovirus.

Clinical Impact

The Luminex GPP assay was one of the first syndromic assays for GI pathogens on the market. However, it still required separate nucleic acid extraction procedures and multiple (post-PCR) processing steps. It was shown that costs for the clinical laboratory increased but overall, there were savings for the hospital (Goldenberg *et al.*, 2015). Another study evaluated the impact of the FilmArray GIP on length of stay and patient isolation and concluded that patients spent considerable time without appropriate isolation methods in the hospital whereas others were placed under contact precautions without cause (Rand *et al.*, 2015). One important observation was reported by Claas *et al.* in that 65% of positive results were from pathogens not specifically ordered by the treating physician (Claas *et al.*, 2013). The clinical impact is much harder to assess in resource-limited settings where it became apparent that in children numerous pathogens can be detected in cases with diarrhea as well as in asymptomatic controls (Leva *et al.*, 2016). This requires careful analysis of the local setting (i.e., patient population: adult versus pediatric, ambulatory versus in-hospital, immune status/previous antibiotic treatment, travel associated) and to evaluate the implementation of new methods in the respective laboratory. For some patients, a more tailored approach would be more effective but not possible with the fixed combinations delivered by the manufacturers. It seems advisable to develop algorithms, which patient should be tested with which method. This apparently needs a cross-sectional and collaborative effort of all involved partners within a hospital (Fig. 1). However, there is still limited evidence based data available to show a clear benefit for the patient without doing harm in industrialized settings. Of note, for bacterial pathogens culture and susceptibility testing is usually not recommended by many current guidelines but may prove useful in the context of outbreak investigations. It is also vital for public health surveillance and epidemiological purposes.

Central Nervous System Infections

Background

Infections of the central nervous system are usually caused by viruses or bacteria or to a lesser extent fungi. They can lead to encephalitis, which is an inflammation of the brain. It is frequently accompanied by meningitis, which is an inflammation of the meninges. Meningitis/encephalitis (ME) is associated with significant morbidity and mortality worldwide. Clinically, signs and symptoms of ME are often unspecific. Headaches and altered mental status are frequently reported. For diagnosis of microbial pathogens, Gram staining, microscopy and bacterial culture are standard laboratory methods. This is complemented by analysis of cerebrospinal fluid (CSF) for protein levels, white blood cell count and other clinical chemistry parameters. Molecular testing for viruses, which include herpes simplex viruses 1 and 2, varicella-zoster virus, and enterovirus as the most common viral pathogens causing ME has been well established by many laboratories for a long time. In addition, serological testing for flaviviruses (tick-borne encephalitis virus, West Nile virus) can be performed depending on the local epidemiology.

Assays

Compared to the situation for respiratory and gastrointestinal infections the availability of multiplex panels is rather limited. However, a number of different assays are in the pipeline of commercial companies and it is expected that they will soon enter the market. However, their performance has to be evaluated in comparison studies first. One of the first assays on the market was the Biofire FilmArray Meningitis/Encephalitis panel, which includes 14 targets including viral as well as bacterial/parasitic pathogens. Analytical performance studies showed somehow contrasting results when the assay was compared to culture and molecular methods (Leber *et al.*, 2016). Of importance, false-positive results have been reported and the results of the FilmArray need to be interpreted cautiously and in combination with clinical signs and other laboratory markers. This again highlights the importance of strict contamination precautions, dedicated laboratory space and trained personnel for molecular testing. In addition, the pretest probability for ME has to be taken into account. Critical points of the multiplex panels for ME include e.g., the above-mentioned sensitivity/specificity issues, the composition of the panels and the relevance of detecting bacteria in populations with

successful immunization programs.

Clinical Impact

Similar to the presently limited number of assays information on the clinical impact remains scarce. With respect to the FilmArray, no reduction in antimicrobial use or length of stay was reported in a study with rather small sample size. Thus, as soon as more assays are available and have proven their analytical performance ideally randomized controlled trials are necessary to assess the clinical impact of these assays. Ultimately, evidence based guidelines should provide reliable information on the use of these ME panels.

Hepatitis E Virus Infection

Hepatitis E virus (HEV) infection is another example of clinical diagnostic virology leading to a fundamental change in current knowledge and clinical practice. Until recently, HEV was thought to be restricted to countries in Africa and Asia. In these countries, HEV genotypes 1 and 2 are causing HEV infections and transmission occurs through fecally contaminated water. Illness is usually short-lived mild hepatitis only. Exceptions are severe courses with a high fatality rate in pregnant women. Sporadically, outbreaks can occur with thousands of cases. In industrialized countries, HEV infections were occasionally seen in travelers who acquired infections abroad. Due to imperfect serological methods, this led to a gross underestimation of HEV seroprevalence rates (Kmush *et al.*, 2015). However, interest and progress in HEV understanding dramatically changed when it became clear that HEV genotypes 3 and 4 are prevalent in industrialized countries and are transmitted zoonotically from infected animals (mostly swine, deer and others). Importantly, clinical diagnostic virology could demonstrate locally acquired HEV infection in chronic liver disease (Dalton *et al.*, 2007).

In brief, diagnosis is based on the detection of HEV-specific antibodies and more recently quantitative real-time PCR became available to detect HEV RNA in serum and stool samples. Of note, serological assays differ in their performance partially questioning the results of older seroepidemiological studies (Hartl *et al.*, 2016). The first comparison studies were conducted showing significant differences among assays. However, in the absence of a true gold standard assay performance is still under evaluation. At the same time, the industry developed new test kits with improved performance characteristics. In addition, quantitative real-time PCR and newly established WHO standards for quantitation of viral HEV RNA were introduced. These are now widely used to monitor patients with chronic HEV infection and to control the success of antiviral treatment (Kamar *et al.*, 2014). Using these novel tools, clinical diagnostic virology provided new data on seroprevalence in many European regions and elsewhere. Interestingly, using state-of-the-art diagnostic tools it soon became clear that beyond hepatitis other extrahepatic manifestations are associated with acute HEV infection including various neurological injuries (Dalton *et al.*, 2016). Thus, clinical diagnostic virology led to new knowledge and changed what was previously thought to be accepted knowledge.

To summarize, clinical diagnostic virology is an interdisciplinary subject requiring close collaboration foremost with clinicians and other related disciplines. Recently, multiplex panels proved to be valuable and powerful tools and are anticipated that fully automated random-access machines for almost all relevant viral targets will enter the market soon. However, critical issues remain, and we need evidenced-based data from high-quality trials on whether to implement these multiplex panels into routine diagnostics. Beyond technical issues, this also requires cost-effectiveness analysis and the development of locally adopted diagnostic testing algorithms to benefit the patient.

References

- Bonzel, L., Tenenbaum, T., Schroten, H., *et al.*, 2008. Frequent detection of viral coinfection in children hospitalized with acute respiratory tract infection using a real-time polymerase chain reaction. *The Pediatric Infectious Disease Journal* 27, 589–594.
- Claas, E.C., Burnham, C.A., Mazzulli, T., Templeton, K., Topin, F., 2013. Performance of the xTAG(R) gastrointestinal pathogen panel, a multiplex molecular assay for simultaneous detection of bacterial, viral, and parasitic causes of infectious gastroenteritis. *Journal of Microbiology and Biotechnology* 23, 1041–1045.
- Dalton, H.R., Hazeldine, S., Banks, M., Ijaz, S., Bendall, R., 2007. Locally acquired hepatitis E in chronic liver disease. *Lancet* 369, 1260.
- Dalton, H.R., Webb, G.W., Norton, B.C., Woolson, K.L., 2016. Hepatitis E virus: Time to change the textbooks. *Digestive Diseases* 34, 308–316.
- Dik, J.H., Poelman, R., Friedrich, A.W., *et al.*, 2017. Integrated stewardship model comprising antimicrobial, infection prevention, and diagnostic stewardship (AID stewardship). *Journal of Clinical Microbiology* 55, 3306–3307.
- Dik, J.W., Poelman, R., Friedrich, A.W., *et al.*, 2016. An integrated stewardship model: Antimicrobial, infection prevention and diagnostic (AID). *Future Microbiology* 11, 93–102.
- Farthing, M., Salam, M.A., Lindberg, G., *et al.*, 2013. Acute diarrhea in adults and children: A global perspective. *Journal of Clinical Gastroenterology* 47, 12–20.
- Goldenberg, S.D., Bacelar, M., Brazier, P., Bisnauthsing, K., Edgeworth, J.D., 2015. A cost benefit analysis of the luminex xTAG gastrointestinal pathogen panel for detection of infectious gastroenteritis in hospitalised patients. *Journal of Infection* 70, 504–511.
- Gunson, R.N., Bennett, S., Maclean, A., Carman, W.F., 2008. Using multiplex real time PCR in order to streamline a routine diagnostic service. *Journal of Clinical Virology* 43, 372–375.
- Hartl, J., Otto, B., Madden, R.G., *et al.*, 2016. Hepatitis E seroprevalence in Europe: A meta-analysis. *Viruses* 8.
- Huang, H.S., Tsai, C.L., Chang, J., *et al.*, 2018. Multiplex PCR system for the rapid diagnosis of respiratory virus infection: Systematic review and meta-analysis. *Clinical Microbiology and Infection* 24, 1055–1063.
- Kamar, N., Izopet, J., Tripon, S., *et al.*, 2014. Ribavirin for chronic hepatitis E virus infection in transplant recipients. *The New England Journal of Medicine* 370, 1111–1120.

- Kmush, B.L., Labrique, A.B., Dalton, H.R., *et al.*, 2015. Two generations of "gold standards": The impact of a decade in hepatitis E virus testing innovation on population seroprevalence. *The American Journal of Tropical Medicine and Hygiene* 93, 714–717.
- Leber, A.L., Everhart, K., Balada-Llasat, J.M., *et al.*, 2016. Multicenter evaluation of biofire filmarray meningitis/encephalitis panel for detection of bacteria, viruses, and yeast in cerebrospinal fluid specimens. *Journal of Clinical Microbiology* 54, 2251–2261.
- Leva, A., Eibach, D., Krumkamp, R., *et al.*, 2016. Diagnostic performance of the Luminex xTAG gastrointestinal pathogens panel to detect rotavirus in Ghanaian children with and without diarrhoea. *Virology Journal* 13, 132.
- Nijhuis, R.H.T., Guerendiain, D., Claas, E.C.J., Templeton, K.E., 2017. Comparison of ePlex respiratory pathogen panel with laboratory-developed real-time PCR assays for detection of respiratory pathogens. *Journal of Clinical Microbiology* 55, 1938–1945.
- Popowitch, E.B., O'Neill, S.S., Miller, M.B., 2013. Comparison of the biofire filmarray RP, Genmark eSensor RVP, luminex xTAG RVPv1, and luminex xTAG RVP fast multiplex assays for detection of respiratory viruses. *Journal of Clinical Microbiology* 51, 1528–1533.
- Rand, K.H., Tremblay, E.E., Hoidal, M., *et al.*, 2015. Multiplex gastrointestinal pathogen panels: Implications for infection control. *Diagnostic Microbiology and Infectious Disease* 82, 154–157.
- Rappo, U., Schuetz, A.N., Jenkins, S.G., *et al.*, 2016. Impact of early detection of respiratory viruses by multiplex PCR assay on clinical outcomes in adult patients. *Journal of Clinical Microbiology* 54, 2096–3103.
- van Zyl, G., Maritz, J., Newman, H., Preiser, W., 2019. Lessons in diagnostic virology: Expected and unexpected sources of error. *Reviews in Medical Virology* 29, e2052.
- Vos, L.M., Bruning, A.H.L., Reitsma, J.B., *et al.*, 2019a. Rapid molecular tests for influenza, respiratory syncytial virus, and other respiratory viruses: A systematic review of diagnostic accuracy and clinical impact studies. *Clinical Infectious Diseases* 69 (7), 1243–1253.
- Vos, L.M., Weehuizen, J.M., Hoepelman, A.I.M., *et al.*, 2019b. More targeted use of oseltamivir and in-hospital isolation facilities after implementation of a multifaceted strategy including a rapid molecular diagnostic panel for respiratory viruses in immunocompromised adult patients. *Journal of Clinical Virology* 116, 11–17.

Further Reading

- Binnicker, M.J., 2015. Multiplex molecular panels for diagnosis of gastrointestinal infection: Performance, result interpretation, and cost-effectiveness. *Journal of Clinical Microbiology* 53, 3723–3728.
- Ramanan, P., Bryson, A.L., Binnicker, M.J., Pritt, B.S., Patel, R., 2018. Syndromic panel-based testing in clinical microbiology. *Clinical Microbiology Reviews* 31.
- Schreckenberger, P.C., Mcadam, A.J., 2015. Point-counterpoint: Large multiplex PCR panels should be first-line tests for detection of respiratory and intestinal pathogens. *Journal of Clinical Microbiology* 53, 3110–3115.
- Webb, G.W., Dalton, H.R., 2019. Hepatitis E: An underestimated emerging threat. *Therapeutic Advances in Infectious Disease* 6.

Virus Diagnosis in Immunosuppressed Individuals

Elisabeth Puchhammer-Stöckl, Medical University of Vienna, Vienna, Austria

Fausto Baldanti, University of Pavia, Pavia, Italy and Scientific Institute for Research, Hospitalization and Healthcare, San Matteo Polyclinic Foundation, Pavia, Italy

© 2021 Elsevier Ltd. All rights reserved.

Introduction

The rapid and sensitive diagnosis of virus infections in the immunocompromised human host is of utmost importance to avoid the development of severe and possibly lethal virus diseases. In contrast to the immunocompetent host, the main virus infections leading to life-threatening infections in immunocompromised patients are often caused by viruses of overall low pathogenicity, such as herpesviruses, polyomaviruses and adenoviruses. Primary infection with these viruses may cause only mild disease or even asymptomatic infections in healthy individuals. However, they become dangerous pathogens in the absence of appropriate development of antiviral immune responses and may proceed to a high level or even unlimited virus replication, leading to highly increased morbidity and mortality. Most of these viruses are often latently present in the patients already before immunosuppression starts due to past infections and frequently reactivate and start extensive replication after the onset of immunosuppression.

The major group of immunosuppressed patients, in whom an intensive and rapid virological diagnosis is required and close virological follow-up is needed, are recipients of solid organ transplantation (SOT) or hematopoietic stem cell transplantation (HSCT) recipients. These patients receive severe artificially induced immune suppression, mostly over their lifetime. The administration of immunosuppressive medication is urgently needed to protect the transplanted organs from the recipient's immune response, directed against foreign antigens. A decrease of the host's immune response can be caused by a number of drugs that influence different immune defense mechanisms, and thereby may avoid acute or chronic rejection of the transplant by the recipient. However, as a problematic and potentially dangerous side effect, this impairment of distinct immune mechanisms also leads to a decreased ability of the host to defend against viral infections and against infections with bacterial and fungal pathogens.

Also, patients with genetic defects that impair different parts of the immune response are at high risk for virus infections. Individuals with Human immunodeficiency virus (HIV)-associated immunosuppression and AIDS may also suffer from severe virus infections, and rapid virological diagnosis is needed. Since the successful introduction and application of highly active antiretroviral therapy (HAART), severe virus infections have become quite rare in numerous HIV positive populations. However, especially patients diagnosed at a late time point after infection with HIV and those with limited access to antiretroviral drugs are still at high risk of developing opportunistic infections with other viruses.

Viral Diagnostic Tests

Diagnosis of virus infections in immunosuppressed persons, especially in transplant recipients, encounters a number of specific problems. One main challenge is that virus screening must be performed preemptively in most cases, which means that virus replication must be detected already before the viruses start to cause clinical disease. This means a routine diagnostic follow-up schedule after transplantation is required. Patients must be screened for defined viral parameters at regular intervals and for a long time. Another aspect complicating the routine diagnosis of viruses with latency phases is that the presence of these viruses does not "per se" prove the patient's risk to develop a virus-associated disease on follow-up.

Indeed, a deep understanding of viral pathogenesis is required to select a better indicator of whether a specific virus disease is to be expected or not. In some cases, low-level virus replication is a physiological state and may be also observed in the immunocompetent host, while increasing to a certain level of virus replication may indicate a high risk for the development of virus diseases and may require immediate intervention. This level, deciding whether antiviral interventions are required, needs to be defined for each cohort and each virus separately. It must be confirmed that antiviral interventions are started at the optimal time point, avoiding unnecessary antiviral treatment, but starting early enough to avoid high-level virus replication and disease. Also, the virus emerging after transplantation may have been derived from the virus population of the transplant recipient as well as from the donor virus population transmitted by the transplant. The main analysis applied in the virological follow-up of transplant recipients is the quantitative analysis of the nucleic acid-load of different viruses as Cytomegalovirus (CMV), Epstein-Barr Virus (EBV), BK-Polyomavirus (BKPyV) or Adenovirus (ADV). This may be performed in blood but also in other compartments, as bronchoalveolar lavage (BAL), urine or stool, depending on the specific virus infection and transplant population.

The extent of the recipient's preexisting immune response against different viruses has a substantial impact on the course of the disease. Transplant recipients, who have no prior immunity for these viruses, show a particularly high risk of developing a severe primary infection by the donor virus after transplantation. To assess the individual risk of a patient developing severe virus infections, testing for the presence of virus-specific antibodies (ABs) in the donor and recipient prior to transplantation is routine. Blood samples collected from the patients are tested by Enzyme-linked immunosorbent assay (ELISA) test systems, determining in particular virus-specific IgG. The absence of virus-specific IgG in the recipient indicates that the patient is susceptible and could

Table 1 Clinical manifestations and virus diagnostics of the most important virus infections in the immunocompromised patient after transplantation

<i>Virus</i>	<i>Clinical manifestation</i>	<i>Diagnosis</i>	<i>Pre Tx serology tested</i>	<i>Preemptive testing by quantitative PCR</i>
Cytomegalovirus	End-organ diseases (Pneumonitis, Colitis, encephalitis etc.) disseminated CMV syndrome	Quantitative PCR from blood (also BAL, CSF other body fluids)	Always, from donor and recipient	In all transplant recipients
Epstein-Barr Virus	Posttransplant lymphoproliferative disease	Quantitative PCR from blood	Sometimes especially in children	In high-risk groups
BK Polyoma Virus	Polyomavirus associated nephritis, loss of kidney transplant Hemorrhagic cystitis after HSCT	Quantitative PCR from urine and blood	Not routinely done	Especially in kidney transplant recipients, eventually in HSCT
Adenoviruses	Generalized infection, especially in children after HSCT	Quantitative PCR from blood and stool	No	Especially in pediatric HSCT
Herpes simplex virus	Extended cutaneous lesions, rare generalized infections	PCR from vesicles, ev blood, CSF	No	No
Varicella-Zoster Virus	Extended cutaneous lesions, rare generalized infections	PCR from vesicles, ev blood, CSF	Yes, vaccination possible	No

develop a primary virus infection. If the transplant derives from an IgG seropositive donor, this is a particularly high-risk situation, as it is most likely that a primary infection from the donor will occur.

The determination of the recipient's antiviral T-cell response is gaining growing interest to assess the individual risk for severe infections. In particular, the reconstitution kinetics of neutrophils, CD4⁺, CD8⁺ T-cells and NK cells are of paramount importance in the HSCT setting to determine transplant success but are also actively investigated in SOT to determine the recovery of immune function following induction treatments. CD4⁺ and CD8⁺ T-cell responses to individual viruses can be investigated to determine the risk of specific virus-related diseases. However, this approach is so far not widespread and no consensus on protective responses has been reached yet.

It was also shown that the level of Torque Teno Virus (TTV), a small commensal and non-pathogenic DNA virus present in the majority of the human population in blood, is associated with the level of drug-induced immunosuppression after SOT. TTV-DNA loads are also affected by the patient's immune response and have been proposed as a potential indicator of overall immune impairment in SOT. This was reported in lung transplant and also kidney transplant recipients. An association was established between high immunosuppression, a high TTV load exceeding a specific cut off level and the occurrence of opportunistic infections. In contrast, low TTV-load was reflected low immunosuppression decreasing under a defined lower cut off level and was significantly associated with organ rejection processes after lung and kidney transplantation.

The virus diagnosis procedures and follow-up recommendations are different and specific for each virus, which potentially causes a high risk to the immunocompromised host. Therefore, the most important virus pathogens and their follow-up will be separately reviewed in the following and are summarized in [Table 1](#).

Human Cytomegalovirus (HCMV)

Human Cytomegalovirus is a member of the herpesvirus family and belongs to the beta-herpesvirus subfamily. HCMV proved to be the major opportunistic agent in immunocompromised patients. Before the HAART era, HCMV was associated with most morbidity and mortality in HIV- infected individuals. Indeed, HCMV end-organ diseases were primary AIDS-defining clinical conditions, typically appearing in the presence of CD4⁺ T-cell counts <200. Similarly, HCMV represents the main opportunistic pathogen in both HSCT and SOT recipients. Like all herpesviruses, HCMV is a widespread agent acquired early in life. In immunocompetent individuals, the majority of primary HCMV infections are asymptomatic, while in a minority of cases is a self-resolving mononucleosis syndrome.

Following primary infection HCMV can establish lifelong latency in many cell types (myeloid cells precursors, endothelial cells, kidney and salivary glands epithelial cells). In immunocompetent individuals, reactivations are asymptomatic. The peculiar ability of HCMV to infect and become latent in many cell types may explain why this virus is so aggressive in immunocompromised patients. Indeed, in patients with impaired T-cell and NK immunity primary and reactivated infection can present as disseminated infections associated with end-organ diseases (retinitis, gastritis, small and large bowel ulcers, pneumonia, encephalitis, hepatitis). This wide range of end-organ diseases is due to HCMV establishing latency in endothelial cells, thus reactivation(s) can potentially involve any organ.

From a diagnostic standpoint, it is then important to: (1) determine the donor (D)/recipient (R) serostatus and (2) periodically monitor the presence and the level of HCMV DNA in blood as a marker of reactivation and dissemination. The highest risk for HCMV disease in HSCT is in D⁻/R⁺ combination, while in SOT is in the D⁺/R⁻ group. In both cases, an HCMV infection could develop in the absence of specific memory B- and T-cell responses, with a high risk of dissemination and organ involvement. For

this reason, high-risk transplant populations are often given antiviral prophylaxis. On the other hand, patients already HCMV experienced are at risk of symptomatic HCMV reactivations due to impaired immunity consequent to immunosuppressive anti-rejection treatments. To minimize the toxic effects of current antivirals, medium to low-risk patients are given pre-emptive treatment protocols based on the administration of antiviral treatment in the presence of documented HCMV replication when still asymptomatic.

Periodic HCMV DNA quantitative determination in the blood is then required in the post-transplant period to detect the presence and the extent of virus replication (reactivation). A typical monitoring schedule involves HCMV DNA quantification once a week in the first 3 months (when immunosuppression is greater in SOT and engraftment is ongoing in HSCT) followed by once every 2 weeks in the second 3-month period and once a month in the following 6 months. Detection followed by the rising of HCMV DNA levels in the blood is indicative of asymptomatic HCMV replication and the potential for dissemination. For pre-emptive treatment initiation, different centers use different HCMV DNA cut-off levels due to a lack of standardization between real-time techniques. Recently, a national consensus conference indicated HCMV DNA cut-off levels for high, medium and low-risk SOT as well as HSCT recipients. The diagnosis of HCMV end-organ diseases requires the identification of viable virus (isolation) and virus antigens (immunohistochemistry) in organ tissues or a high level of CMV DNA in relevant fluids.

Sequential determination of HCMV DNAemia is essential to identify treatment failures. Indeed, unchanged or increasing HCMV DNA levels following several days of treatment may indicate the “in vivo” selection of a mutant and drug-resistant HCMV strain. A number of drug-resistant mutants have been described and occur especially in the UL97 and UL54 genes. UL97 codes for a kinase required for the phosphorylation of ganciclovir and different mutations at this gene have been identified which confer resistance against this drug. The UL54 gene codes for the virus polymerase and distinct mutations in this gene may confer resistance against ganciclovir, foscarnet or cidofovir. To identify the development and extent of drug resistance, sequencing of the UL97 and UL54 genes is required. This can confirm the presence of ganciclovir, foscarnet and cidofovir resistant strains and is important to help direct a change of antiviral agents.

Finally, assessment of specific T-cell immunity has proven helpful in predicting patients at risk of HCMV reactivation and disease, as well as designing tailored preventive and treatment approaches.

Epstein-Barr Virus (EBV)

EBV is a member of the herpesvirus family and belongs to the gamma-herpesvirus subfamily and has oncogenic potential. In fact, a peculiar characteristic of the two members of the gamma-herpesviruses (EBV and HHV-8) is their correlation with many lymphomas, sarcomas and carcinomas. Strikingly, EBV-induced malignancies are associated with different virus latency phases (Table 2). Thus, EBV is pathogenic both during replication and latency.

EBV primary infection is acquired early in life and is mostly asymptomatic, justifying the impressive biologic success of this virus. Indeed, EBV seroprevalence is virtually 100% at the age of 20. Mononucleosis is the clinical manifestation of primary EBV infection in a small (but significant) fraction of individuals. Immunocompromised patients, especially transplant recipients, are at risk of complications in both the replicative phase (chronic active EBV infection) and latency (post-transplant lymphoproliferative disorders or PTLN, lymphomas and leiomyosarcoma).

Among SOT recipients, D⁺/R⁻ are at a higher risk of complications. Given the high EBV seroprevalence in young adults and adults, pediatric transplant recipients are at the highest risk of short-term complications. In addition, pediatric SOT recipients are also at the highest risk of late EBV-associated neoplasm due to the longer life expectancy and the greater cumulative effect of immunosuppression. EBV serology might be tricky to interpret due to the response to multiple antigens and the relevant antibody maturation (Table 3).

Table 2 EBV associated tumors and expression of EBV latency genes (modified from J.L.Hsu, S.L. Glaser, Critical Reviews in Oncology/Hematology 34; 27–53, 2000)

<i>EBV latency</i>	<i>Pattern of EBV latency gene expression</i>	<i>Malignancy</i>
Type I	EBNA-1 EBERs	Burkitt's lymphoma Gastric carcinoma
Type II	EBNA-1 LMP-1, -2A, -2B EBERs	Nasopharyngeal carcinoma Hodgkin's disease (HD) Nasal T/NK lymphoma
Type III	EBNA-1, -2, -3A, -3B, -3C, -LP LMP-1, -2A, -2B, EBERs	AIDS-associated non-HD lymphoma PTLD (post-transplant lymphoproliferative disorders)
Other	EBNA-1, -2 EBERs	Leiomyosarcoma

Table 3 Typical^a EBV-serological profiles using the most frequently employed antigens and Ig isotypes

Antigen/Ig isotype condition	VCA/IgG	VCA/IgM	VCA/IgA	EA/DIgG	EA/R/IgG	EBNA/Ig
Susceptible/no previous exposure	—	—	—	—	—	—
Primary infection	++	+++	+	+	+/-	—
Past infection	+	—	+/-	—	+/-	+
Inconclusive: Primary or past infection	+	+	+/-	+/-	+/-	+
Inconclusive: Primary or past infection	+	—	+/-	+/-	+/-	—
Chronic active EBV infection	+++	+	++	++	+	+/-

^a + / —, Antibodies absent or present; +, Antibodies present; ++, Antibodies present in elevated titers; + + +, Antibodies present in strongly elevated titers.

Abbreviations: EBV: Epstein–Barr virus.

VCA: Virus capsid antigens.

EBNA: EBV nuclear antigen.

EA/D: Early antigen/diffuse.

EA/R: Early antigen/restricted.

Atypical patterns include VCA IgM positive only; EBNA positive only; or VCA IgM and EBNA positive, but VCA IgG negative. Atypical patterns merit repeat testing, testing of a follow-up sample, or testing by alternate methods.

Primary EBV infection (mononucleosis) and chronic active EBV infection (often a complication of initial mononucleosis) are observed almost exclusively in pediatric transplant recipients, while PTLD is an uncommon complication of HSCT and SOT in both pediatric and adult settings. The risk of PTLD is directly correlated with the level and duration of immune suppression. Thus, transplants (or rejection episodes) requiring aggressive immune suppressive treatments bear an elevated risk of EBV-PTLD. From a treatment standpoint, it must be underlined that PTLD, as well as all EBV-associated tumors, are induced during a replication-inactive phase of the virus life cycle. Currently available anti-herpetic drugs (inhibiting viral DNA polymerases) are ineffective. Indeed, during PTLD, EBV DNA replicates in its latent form via cellular DNA polymerases in proliferating cells. Then, measurement of EBV DNA in peripheral blood of transplant patients (although pivotal in the monitoring of PTLD development) has to take into account these pathogenetic peculiarities: (1) during mononucleosis or chronic active EBV infection EBV DNA is the expression of active virus replication and it can be used to monitor the efficacy of antiviral treatment, while (2) during PTLD EBV DNA is the expression of EBV-induced cell proliferation in the absence of active virus replication and it can be used to monitor the efficacy of immune suppression tapering as well as lymphocyte B lytic treatments. Cut-off levels to start treatment for PTLD are still debated.

Monitoring EBV-specific T-cell responses to replicative phase antigens as well as to latency phase antigens might be helpful to define patients at risk of PTLD better.

Polyomaviruses (PyV)

Polyomaviruses comprise a group of non-enveloped DNA-viruses that remain after primary infection and persist in the human host. PyVs are highly prevalent worldwide in human populations and include among others BK-PyV, JC-PyV, Merkel cell Carcinoma PyV and trichospinusulosa PyV.

The most important of these viruses in the transplant setting is BK-PyV, which is of clinical significance nearly exclusively after renal transplantation or HSCT. BK-virus is latently resident in kidney cells and may replicate to high levels after renal transplantation. In the worst case, it may lead to Polyomavirus associated nephritis (PVAN) and to loss of the transplant. Recent data provide evidence that BK-PyV associated disease is caused especially by the donor BK-PyV transmitted by the transplanted kidney. The key element in avoiding the development of PVAN is preemptive routine diagnostic testing for BK-PyV DNA by quantitative PCR assays in urine and blood. While low-level BK-PyV DNA is sometimes detectable in immunocompetent persons and is of no significant clinical value, urine BK-PyV DNA levels increasing over > 7–8 log₁₀ copies/mL signify substantial virus replication and reflect an increased risk for progression to BK-PyV DNAemia and further PVAN. Moreover, the detection of decoy cells in urine, which have enlarged cell nuclei and intranuclear viral inclusions, may be of diagnostic value.

Either from the beginning, but at the latest as soon as the virus load in urine increases over > 7–8 log₁₀ copies/mL urine, quantitative PCR testing in blood is required to control the replication kinetics further. An increase of the BK-PyV DNAemia over > 3–4 log₁₀ copies/mL plasma indicates a high risk of developing PVAN. Renal histology is then required to help with management decisions to counteract further BK-PyV replication. These include reducing the individual immunosuppressive drugs to re-establish the recipient's antiviral response and avoid progression and transplant loss. Without intervention, about 50% of highly viremic patients progress to PVAN. The quantitative PCR test cut-offs mentioned above are recommendations based on specific in-house PCR assays. An international BK-PyV calibrator approved by the World Health Organization (WHO) is available and this allows comparison of different quantitative PCR-assays used in different centers.

Serology plays no role in routine testing and risk assessment of kidney transplant recipients. This is in part due to the fact that BK-PyV AB tests are not commercially available and only established in single laboratories in the course of scientific projects. However, an increasing amount of data suggest that a high level of BK-PyV specific ABs in the donor pre-transplantation is associated with a higher risk of the recipient developing PVAN, probably as it reflects a higher virus load in the donor kidney.

BK-PyV may also cause hemorrhagic cystitis in patients after HSCT. Hemorrhagic cystitis may occur in up to 25% of pediatric and up to 54% of adult allogeneic HSCT patients. Diagnosis of BK-PyV associated cystitis is made using clinical parameters as well as by virus detection in urine by PCR. High-level viruria ($> 7 \log_{10}$ BK-PyV DNA copies/mL urine) as well as DNAemia seem to predict hemorrhagic cystitis in these patients.

Other polyomaviruses cause infrequent complications in immunosuppressed patients after transplantation and are therefore not subject to routine follow up virus load testing by quantitative PCR. JC-PyV is latently present in the kidney and in the brain neurones of most people and may cause progressive multifocal leukoencephalopathy (PML) in rare cases of highly immune depressed AIDS patients or in HSCT patients. This is caused by reactivation of JC-PyV in severely T-cell deficient patients and due to progressive demyelination in the brain tissue which is mostly lethal. JC-PyV PCR tests are carried out in cerebrospinal fluid samples and also from blood to diagnose PML.

Adenoviruses (AdVs)

Adenoviruses (AdVs) are non-enveloped DNA viruses that are widely distributed. AdV contains seven species, termed A to G. Most infections with AdVs occur at a young age and in the immunocompetent host are mostly associated with mild infections, such as respiratory infections, ocular infections or gastroenteritis. However, in the immunocompromised host, and especially in children after allogeneic HSCT AdVs are important pathogens, as they may extensively replicate and may cause life-threatening infections. In these patients, AdVs may cause hemorrhagic enteritis or cystitis, or affect other organs including lung, liver, kidney or heart tissue, and finally, AdV infections may cause multiorgan failure. The lethal outcome of generalized AdV infections in children after allo-HSCT is high.

It is thought that the virus derives from extensive local replication especially in the gastrointestinal tract of severely immunosuppressed children and is further spread extensively to different organs by AdV viremia. Therefore, close routine diagnostic follow up for AdV is strongly recommended in this pediatric cohort. Also, adults may develop systemic AdV infections after allogeneic HSCT, however, this is observed to a much lower extent.

Preemptive diagnosis for AdV infections is based on quantitative AdV DNA PCR from stool and blood, using a pan-AdV PCR, detecting all AdV strains in one assay. Current PCR tests commonly target the AdV hexon gene which is a conserved region within AdV strains. Detection of virus DNA in blood is considered a marker of ongoing systemic disease. Therefore, most centers test for DNAemia routinely and on a weekly basis after transplantation. A level of $> 3 \log_{10}$ copies/mL plasma is considered critical for the progression of infection to generalized disease and requires therapeutic intervention. This involves systemic treatment with cidofovir or brincidofovir. Also, a viral load of $> 6 \log_{10}$ AdV DNA copies/g stool is considered critical. Before HSCT, intestinal AdV shedding is associated with an increased risk of later development of disseminated AdV disease post-transplantation. Therefore, AdV PCR testing of stool prior to transplantation may be a useful tool to assess the individual risk of patients for generalized AdV disease. AdV infections after solid organ transplantation are rare.

Other Viruses

There are also other viruses that may cause significant complications in immunosuppressed patients. But as these occur less frequently, routine follow up testing is not initiated in these post-transplant risk groups.

Herpes Simplex Virus (HSV)

Herpes simplex virus, especially HSV type 1, is highly prevalent in the human population. It usually causes a symptomatic or asymptomatic primary infection during early childhood and is thereafter latently present in ganglia cells. HSV may reactivate from time to time also in healthy individuals due to stressful situations or after extended sun exposure. It may then lead classically to vesicular lesions on the lips, herpes labialis, which is quite common. Under severe immunosuppression, HSV may also reactivate and may cause severe cutaneous manifestations with extended dermal lesions, which may also become necrotic. In rare cases, HSV may also affect different organs and even cases of HSV encephalitis may be observed. HSV infection is usually identified by clinically evident dermal manifestations. In difficult clinical situations, the diagnosis may be confirmed by HSV-specific PCR carried out on vesicle fluid. In generalized infection, blood or in suspected cases of encephalitis or meningitis also cerebrospinal fluid (CSF) samples will be tested by PCR. Early treatment with high dose intravenous aciclovir has to be administered as soon as possible to avoid or limit further virus dissemination. Frequently, HSV is also detected in patients in intensive care units, especially in the lower respiratory tract. Whether this has any clinical importance or whether this is due to subclinical replication under extremely stressing conditions, and eventually also due to intubation and artificial ventilation is still unclear.

Varicella-Zoster Virus (VZV)

Varicella-Zoster Virus is a herpesvirus that is highly prevalent in the population, except in younger people if they have been vaccinated already against VZV.

Primary infection with VZV manifests as chickenpox, classically associated with a generalized vesicular rash, and is highly contagious. Diagnosis is mostly made clinically on seeing the typical dermal lesions. The virus lies in a latent state in the dorsal root ganglia cells of the host and remains there for life. Reactivation of VZV is observed most frequently in older individuals and is associated with decreasing cellular immune response against VZV. Reactivation manifests typically as herpes zoster (shingles), manifesting as a vesicular rash that covers usually one dermatome. Herpes zoster may be accompanied by severe pain, or by complications affecting especially the central nervous system (CNS) and appear as meningitis, meningoencephalitis or transverse myelitis. Herpes Zoster infection is usually identified by clinically evident dermal manifestations. In difficult cases, the diagnosis may be confirmed by VZV-specific PCR performed from vesicle fluid. In generalized infection, VZV-PCR may be carried out in blood or CSF in suspected cases of encephalitis or meningitis.

VZV may cause severe complications in the immunocompromised host. Reactivation of VZV latently present in the patient may start after immunosuppression and may lead to severe cases of herpes zoster, affecting more than one dermatome. In especially severe cases not only vesicles but severe and extended necrotizing dermal lesions may be observed. The clinical picture is also clear from the dermal manifestations, but in difficult cases, VZV-specific PCR from vesicle lesions is performed. VZV infection after immunosuppression may also cause meningoencephalitis and diagnosis by VZV PCR from CSF is needed. Rarely systemic infections occur. These have a dramatic course associated with the development of multiple ulcers in the gastrointestinal tract and with multi-organ failure. Rapid detection of VZV DNA by PCR is then required, also from blood. As the clinical appearance of generalized infection is often unclear at the beginning and the course of the disease is fulminant, mortality is high. Treatment with high dose intravenous antiviral drugs such as aciclovir is mandatory in all severe cases of VZV infection in the immunocompromised host.

Children who have neither undergone primary VZV infection nor vaccination against VZV prior to immunosuppression, are at a particularly high risk of developing a primary infection with VZV after transplantation and the start of immunosuppression. Such primary varicella in the immunocompromised host is associated with especially severe clinical complications. Therefore, determination of the VZV ABs prior to transplantation is mandatory, especially in children, to assess the patients' VZV serostatus. If no VZV specific ABs are detectable pre-transplantation, VZV vaccination has to be administered before the start of immunosuppression.

Hepatitis E Virus

Hepatitis E virus is an RNA virus. HEV genotype 3 infections occur especially in pigs and boar in many countries and may be transmitted to humans mostly by eating undercooked food. Acute infection with HEV genotype 3 infection is self-limiting and asymptomatic in the immunocompetent host. However, when patients develop an HEV infection after transplantation, they may develop chronic hepatitis during immunosuppression. Undiagnosed this may lead to the rapid development of liver cirrhosis within a few years and death of the patient. Therefore, screening for HEV infection should be performed in all transplant patients with unclear elevation of liver enzymes. HEV-PCR can detect viremia from blood, and a positive result indicates ongoing acute or chronic infection. Also, HEV can be detected in the stool of the patients. HEV specific AB testing is not suitable in the immunocompromised host, as AB development may be reduced over a long time after infection in these patients. A decrease of immunosuppression and/or administration of ribavirin may lead to the termination of viral replication.

Conclusion

A number of different viruses are dangerous pathogens for the immunosuppressed host. These may not only affect the patient as a single virus infection but may also appear in combination with other virus infections or with bacterial or fungal infections under immunosuppression. The diagnostics of virus infections in immunocompromised patients is of utmost importance for the patients' short and long-time survival. But high virological expertise is needed to correctly perform appropriate tests at the optimal time points and establish exact quantification. Participation in quantitative quality control panels is needed, to allow inter-laboratory comparisons and to assess quantitative cut-off levels, which are comparable between different centers. The diagnosis of virus infections especially after transplantation and the interpretation of the test results requires good knowledge of the particular viruses, their pathogenesis and life cycle, and the mechanisms of virus latency and replication. The differentiation between non-pathogenic and potentially pathogenic virus load levels is still difficult in most virus infections, and the identification of the turning point to highly dangerous levels of virus replication is challenging and needs highly experienced experts.

Further Reading

- Calarota, S.A., Aberle, J.H., Puchhammer-Stöckl, E., Baldanti, F., 2015. Approaches for monitoring of non virus-specific and virus-specific T-cell response in solid organ transplantation and their clinical applications. *Journal of Clinical Virology* 70, 109–119.
- Calarota, S.A., Chiesa, A., Zelini, P., *et al.*, 2013. Detection of Epstein-Barr virus-specific memory CD4⁺ T cells using a peptide-based cultured enzyme-linked immunospot assay. *Immunology* 139 (4), 533–544.
- Cesaro, S., Dalianis, T., Hanssen Rinaldo, C., *et al.*, 2018. ECIL guidelines for the prevention, diagnosis and treatment of BK polyomavirus-associated haemorrhagic cystitis in haematopoietic stem cell transplant recipients. *Journal of Antimicrobial Chemotherapy* 73 (1), 12–21.

- Girmeria, C., Lazzarotto, T., Bonifazi, F., *et al.*, 2019. Assessment and prevention of cytomegalovirus infection in allogeneic hematopoietic stem cell transplant and in solid organ transplant: a multidisciplinary consensus conference by the Italian GITMO, SITO, and AMCLI societies. *Clinical Transplantation* 16, e13666.
- González-Vicent, M., Verna, M., Pochon, C., *et al.*, 2019. Current practices in the management of adenovirus infection in allogeneic hematopoietic stem cell transplant recipients in Europe: The advance study. *European Journal of Haematology* 102 (3), 210–217.
- Görzer, I., Haloschan, M., Jaksch, P., Klepetko, W., Puchhammer-Stöckl, E., 2014. Plasma DNA levels of torque teno virus and immunosuppression after lung transplantation. *The Journal of Heart and Lung Transplantation* 33 (3), 320–323.
- Hirsch, H.H., Randhawa, P.S., 2019. BK polyomavirus in solid organ transplantation – Guidelines from the American society of transplantation infectious diseases community of practice. *Clinical Transplantation* 33 (9), e13528.
- Jaksch, P., Kundi, M., Görzer, I., *et al.*, 2018. Torque teno virus as a novel biomarker targeting the efficacy of immunosuppression after lung transplantation. *The Journal of Infectious Diseases* 218 (12), 1922–1928.
- Kotton, C.N., Kumar, D., Caliendo, A.M., *et al.*, 2018. The third international consensus guidelines on the management of cytomegalovirus in solid-organ transplantation. *Transplantation* 102 (6), 900–931.
- Lee, D.H., Zuckerman, R.A., 2019. Herpes simplex virus infections in solid organ transplantation: Guidelines from the American society of transplantation infectious diseases community of practice. *Clinical Transplantation* 33 (9), e13526.
- Lhomme, S., Legrand-Abravanel, F., Kamar, N., Izopet, J., 2019. Screening, diagnosis and risks associated with Hepatitis E virus infection. *Expert Review of Anti-infective Therapy* 17 (6), 403–418.
- Lion, T., 2014. Adenovirus Infections in Immunocompetent and Immunocompromised patients. *Clinical Microbiology Reviews* 27 (3), 441–462.
- Navarro, D., San-Juan, R., Manuel, O., *et al.*, 2017. Cytomegalovirus infection management in solid organ transplant recipients across European centers in the time of molecular diagnostics: An ESGICH survey. *Transplant Infectious Disease* 19 (6).
- Pergam, S.A., Limaye, A.P., 2019. Varicella zoster virus (VZV) in solid organtransplantation: Guidelines from the American society of transplantation infectious diseases community of practice. *Clinical Transplantation* 4, e13622.
- Reza Hosseini, O., Drabe, C.H., Sørensen, S.S., *et al.*, 2019. Torque-Teno virus viral load as a potential endogenous marker of immune function in solid organ transplantation. *Transplantation Reviews* 33 (3), 137–144.
- Thorley-Lawson, D.A., 2015. EBV persistence – Introducing the Virus. *Current Topics in Microbiology and Immunology* 390 (Pt 1), 151–209.

Diagnosis; Future Prospects on Direct Diagnosis

Marianna Calabretto and Daniele Di Carlo, Sapienza University of Rome, Rome, Italy

Fabrizio Maggi, University of Pisa, Pisa, Italy and University of Insubria, Varese, Italy

Guido Antonelli, Sapienza University of Rome, Rome, Italy

© 2021 Elsevier Ltd. All rights reserved.

The diagnosis of viral infection, especially the direct one (i.e., demonstration of viral particle - as a whole or through detection of viral antigens or nucleic acids), has been and is going to be, affected by several breakthroughs. Indeed, until a few decades ago, the direct diagnosis was made by performing viral isolation by using cell culture. This method, which today only in some instances still represents the “gold standard”, takes a long time to be carried out and requires highly skilled and experienced staff able to correctly interpret the viral cytopathic effect and to manage cell lines and highly pathogenic viruses. These characteristics do not fit with the current need for low-cost, accurate, and rapid virological diagnosis. Indeed, it is generally accepted that a rapid and more reliable diagnosis may lead to better management of the patient, including personalized therapy, reduced use of unnecessary drugs (for instance antibiotics), and shorter hospitalization. For the above reasons and considering the constant improvement of diagnostic assays, most clinical virology laboratories became essential in the hospital to make clinical decisions.

Although the impressive achievement in terms of technical advances, it is important to note that today there is still no diagnostic approach (i.e., molecular detection, antigen identification, virus isolation, etc.) that: meets all needs of diagnostic virology laboratories; is suitable for all clinical situations; and applies to all virus types. This indirectly means that virologists, who are aware of these limitations, use a different assay depending on any specific clinical situation.

This document provides an overview of the main technologies that, to our opinion, have been used in recent years, and will be used in the future, in the routine activity of a clinical laboratory of a big Hospital. In the framework of these technologies, we will try to outline possible future scenarios of their application in the context of the virological diagnostic.

PCR and Its Evolution

Polymerase chain reaction (PCR) is the assay that allows the amplification of nucleic acid fragments of which initial and terminal sequences are known. Its use for the quantification of viral nucleic acids in body fluids has profoundly changed virological diagnosis. Up to now, the amplification of nucleic acids through PCR represents the gold standard for the screening, diagnosis, and follow-up of several viral infections, such as HIV, HBV, HCV, CMV, respiratory viruses.

The most used tests in the laboratory are real-time PCR and reverse transcription-PCR (RT-PCR).

The real-time PCR allows detection and quantification of a viral product during the run of the assay; “real-time”, as opposed to the standard PCR which quantified it only at the end of the session. This leads to: a reduction of run assay time; the lack of need for post PCR processing; an increase in sensitivity.

RT-PCR, on the other hand, allows amplifying cDNA (complementary DNA) obtained by a previous reverse transcriptase reaction on mRNA and other RNA virus genomes.

The introduction of viral load measurement in biological fluids through PCR has led to the reduction of morbidity, mortality, and disease transmission, especially in chronic infections such as HIV, HBV, and HCV. It found application also in monitoring viral infections in transplanted patients such as CMV.

qPCR is based on the quantification of nucleic acids by comparing the number of amplification cycles, which is believed to be exponential, and the quantity of the final product of PCR with those of a reference sample. qPCR measures the intensity of fluorescence at specific times (generally at the end of each amplification cycle) to quantify the amount of target molecule. It gives the threshold per cycle (CT) and the difference in CT is used to calculate the amount of initial nucleic acid.

However, this comparison is subject to factors that could produce inaccurate data. For example, the amplification cycles may not be exponential, or the target nucleic acid quantity may not be sufficient (Pekin *et al.*, 2011; Quan *et al.*, 2018).

The future of quantitative PCR is digital PCR (dPCR).

dPCR, indeed, thanks to the advances in nanofabrication and microfluidics that have now led to systems that produce hundreds to millions of nanoliter- or even picoliter-scale partitions, breaks down the body fluid by quantifying the presence of target nucleic acid in every single droplet and the measurement of the DNA quantity takes place after the amplification is completed. This allows having a more precise and accurate measurement of the number of amplified molecules.

Furthermore, while qPCR is unable to distinguish differences in gene expression or variations in copy numbers that are smaller than twofold (Kuang *et al.*, 2018), dPCR has shown to distinguish between copy number variations of that differ by only one copy (Baker, 2012). Its greater sensitivity, accuracy, and precision make dPCR a potential evolution in virological laboratory diagnostics to test viral load in a different compartment.

Next-Generation Sequencing (NGS)

The next-generation sequencing (NGS) (also called massively parallel or deep sequencing) represents a series of technologies that allow sequencing an entire genome within a few days. The peculiarity of NGS is to include together the three main steps of sequencing: sample preparation, sequencing, and data analysis. This saves significant time and reduces the number of errors in practice. There are three ways to sequence genetic material through NGS.

The first method is similar to the classic Sanger method with the amplification and sequencing of a specific target. The second involves the amplification of all the genetic material present in the sample followed by NGS. Finally, in the third method, the viral genetic material is first enriched by various methods and then sequenced by NGS (Maggi *et al.*, 2019).

The applications of NGS in virology and clinical virology lab are various.

First of all, it allows the identification of new pathogenic and non-pathogenic commensal viruses. For example, the Schmallenberg virus, a new Orthobunyavirus that causes malformations in cattle and sheep, was discovered through NGS (Hoffmann *et al.*, 2012; Rosseel *et al.*, 2012).

Through a metagenomic approach, it is also possible to perform an etiological diagnosis of viral infection, as in the case of the recent identification of viruses responsible for diseases with severe fever but of unknown etiology (Yozwiak *et al.*, 2012; McMullan *et al.*, 2012).

Furthermore, whole – viral genome sequencing may allow: the study of genetic mutations that can lead to drug resistance; the identification of variants that can overcome the immune response of the host; the identification of diversity in the viral variants/genomes within the same host (useful in the case of viruses able to give chronic infection, such as HIV, HCV, HBV, or Herpesviruses) (Clevenbergh, 2000; Kim *et al.*, 2014; Houldcroft, 2017) and, in this framework, study of the diversity of the viral genome and its evolution; to address molecular epidemiology and viral phylogeny.

As far as HCV variants resistant to antiviral drugs are concerned, it is known that the resistance to some HCV antiviral drugs is due to mutations present in some non-structural proteins of the virus, NS3/4A, NS5A, and NS5B. Although now the issue is overcome by the introduction of very effective drugs, NGS would be possible to characterize the virus to administer the most effective antiviral drug. Besides, NGS was used to simultaneously differentiate HCV and HGV (Hepatitis G Virus) in the patient's serum. The two viruses share the same transmission mode that can cause co-infection. The simultaneous diagnosis of the two viruses, therefore, improves the management of infection and therapy (Parker and Chen, 2017).

NGS has the merit also to have demonstrated: the high variability of viruses that were believed to be genetically stable; that many anatomical sites considered sterile have their flora; that in biological samples it is possible to find viral elements not associated with diseases that are more numerous than pathogenic viruses. This aspect is important for the concept of virome, i.e., the presence of a commensal viral flora that is part of the human microbiome, the community of microorganisms present within the human body (Vu and Kaiser, 2017; Rascovan *et al.*, 2016; Freer *et al.*, 2018).

NGS produces a high value of data that must be collected, evaluated, stored, and used correctly so that the potential of this technology can be exploited to the full. To reach this goal a synergistic activity between virology and “bioinformatics” which combines biology and data processing should be planned (Barzon *et al.*, 2013).

CRISPR-Cas

The microbial clustered regularly interspaced short palindromic repeats (CRISPR) represents one of the new technologies that is going to make a revolution also in the molecular diagnostics market, and some example are already available (Joung *et al.*, 2020). In particular, such a technology has changed the perspectives of gene therapy and manipulation. CRISPR for editing cellular genes must be associated with the enzyme Cas (CRISPR-Cas) and an RNA that guides the complex to the target site of nucleic acid; the enzyme cuts the nucleic acid, triggering a cellular DNA repair mechanism that activates or deactivates the target gene or removes pieces of nucleic acid. The best example in virology is the use of technology to handle the integrated viral genome (Deng *et al.*, 2018; Ford *et al.*, 2019). Many viruses have indeed the ability to establish a persistent infection by integrating their genome into the host's DNA. CRISPR technology has been used in vitro and in animal models to achieve integrated sequences, managing to eliminate chronic viral infections. For example, it was used in vitro to inactivate HIV gene expression in microglia and T cells. Besides, targeting the HIV LTR, gag, and pol genes it was possible to eliminate HIV proviral DNA from animal cells of the spleen, lungs, colon, heart, and brain (Hu *et al.*, 2014; Wang *et al.*, 2018; Yin *et al.*, 2017). Also, CRISPR technology delivered by a lentivirus vector has been used to eliminate HIV proviral DNA from infected human peripheral blood mononuclear cells in a transgenic mouse model (Bella *et al.*, 2018).

The same results have also been achieved with other viruses, such as Herpesviruses. In particular, using CRISPR, 95% of the EBV genome in latently infected Burkitt lymphoma cells was achieved (van Diemen *et al.*, 2016).

These results suggest that in the future CRISPR technologies can be used also for therapeutic purposes like the eradication of chronic infections in the tissues and the oncological therapy of malignancies with viral etiology (Strich and Chertow, 2019; Myhrvold *et al.*, 2018).

One of the side effects of the combination of CRISPR/Cas13 is promiscuous ribonuclease activity upon target recognition. This has been exploited, combined with isothermal amplification, to quickly search for RNA or DNA with attomolar sensitivity and single-base mismatch specificity. This Cas13a-based molecular detection platform, termed Specific High-Sensitivity Enzymatic Reporter UNLOCKing (SHERLOCK), has been used to detect specific strains of Zika and Dengue virus, distinguish pathogenic bacteria, genotype human DNA, and identify mutations in cell-free tumor DNA (Roy *et al.*, 2018; Gootenberg *et al.*, 2017).

Moreover, this method seems useful because SHERLOCK reaction reagents can be lyophilized for cold-chain independence and long-term storage and be readily reconstituted on paper for field applications.

Combined with HUDSON (heating unextracted diagnostic samples to obliterate nucleases), SHERLOCK can work directly on biological fluids without having been previously extracted and in less than two hours (Gootenberg *et al.*, 2017).

Another recent method, similar to SHERLOCK, is DETECTR (DNA endonuclease targeted CRISPR trans reporter). This uses Cas12a, similar to Cas13a but working on DNA, which produces single-stranded DNA cleavage activity after recognizing the target sequence. DETECTR has been used successfully to detect human papillomaviruses 16 and 18, two of the most important genotypes that cause neoplastic transformation. This method has proved to be simple and fast compared to other diagnostic methods for the diagnosis of human papillomavirus making it suitable for its use for rapid molecular diagnosis (Chen *et al.*, 2018).

MALDI-TOF Mass Spectrometry

Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) allows identifying with certainty and precision the type and strain of microorganism responsible for a specific pathology, through the analysis of a sample of the patient's biological fluid (blood, urine, or liquor) subjected to spectrophotometric analysis, without the need of preliminary treatments.

The sample to be analyzed is hit by a laser beam which first causes the fragmentation of its molecules and, subsequently, their ionization. Entering an electric field, these generate a signal that is read by software that produces a particular path. This will then be compared to the spectra already present in a database, allowing the recognition of a specific microorganism through this molecular "fingerprint".

In bacteriology and mycology, this technology found large application and then is widespread. In virological diagnostics is going to be used but still requires consolidation. It finds use to detect viral components in a wide variety of biological specimens, to spot genome mutations associated with drug resistance, and to identify the genotype of viruses (Cobo, 2013).

MALDI-TOF MS can be used for rapid and accurate diagnosis of influenza viruses, being able to detect the different subtypes, and for the diagnosis of the various enteroviruses that cause pediatric gastroenteritis, especially if associated with Multiplex PCR (Chou *et al.*, 2011; Piao *et al.*, 2012).

Furthermore, this method has found application in the identification of the viral genotypes of HPV, HBV, and HCV, improving the management of the infection and the response to the therapy (Gray and Coupland, 2014). Finally, through MALDI-TOF MS it is possible to identify drug resistance against antivirals, for example, resistance to ganciclovir (Zürcher *et al.*, 2012).

The main advantages of using this method derive from its rapid and reliable identification of a large variety of viruses, its simplicity of use, its sensitivity and accuracy.

On the other hand, its diagnostic use is limited due to the limited coverage of virus species in the database and MALDI-TOF devices are at the moment still expensive.

Multiplex and POCT

The need to optimize laboratory diagnosis in clinical practice has also led to the development of multiplex platforms over the years, both as regards bacteriological, parasitological, and virological diagnosis. In particular, multiplex PCRs allow detection or quantitation of sequences of multiple viruses in a single session using a set of specific primers (McCulloh *et al.*, 2014).

The advantages of multiplexing tests include reducing turn-around-time and improving patient management. The most recent multi-analytical tests allow analyzing a pattern of viruses and germs whose infections are associated with the same clinical manifestations, which possess the same tissue tropism and the same mode of infection. The manufacturing companies are dedicated to the development of syndromic panels specific for the body district, for example respiratory, gastrointestinal, central nervous system, and sexually transmitted diseases panels (Diaz-Decaro *et al.*, 2018; Huang *et al.*, 2018; Scagnolari *et al.*, 2017; Zhang *et al.*, 2015).

The application of these panels is still ongoing but we have already obtained some results.

For instance, multiplex PCR proves useful in the diagnosis of meningitis and encephalitis. Despite the suboptimal sensibility observed for a certain target, multiplex molecular panels allow simultaneous detection of multiple pathogens. Thus, in a situation that requires rapid and reliable diagnosis with small quantities of the sample – cerebrospinal fluid (CSF) – their application could improve a patient's outcome (Dien Bard and Alby, 2018).

The panels dedicated to sexually transmitted diseases that simultaneously diagnose HIV and Syphilis have led to increased compliance for monitoring patients with risky behaviors, such as drug addicts (Elnifro *et al.*, 2000).

In some circumstances, molecular assays have been made so simple to use that they are often marketed as point of care tests (POCT) which are characterized as simple and rapid diagnostic methods that can be used also by healthcare professionals with little or no laboratory experience. Over the last few years, the need for POCT has emerged considering the following advantages: POCT may be performed close to the patient thus reducing the time of diagnosis, especially in case of serious and potentially lethal

pathologies. These features allow a rapid diagnosis made at the patient's bed and also be made in complex situations where access to laboratory methods is difficult if not impossible (Briedigkeit *et al.*, 1998; Johannessen, 2015; Haleyr Giri Setty and Hewlett, 2014). In particular, this approach has found an important development in the diagnosis of HIV-1, influenza viruses, viral hepatitis, and emerging viruses such as the Ebola virus. Especially when potentially dangerous viruses are involved, it is important to adequately train health professionals in the safe collection and handling of biological samples and the correct storage and use of tests for a reliable diagnosis (Li-Kim-Moy *et al.*, 2016; Khuroo *et al.*, 2015; Kaushik *et al.*, 2016).

Multiplex and POC tests can improve the management of the patients when it needs a viral diagnosis. Through the search for multiple microorganisms simultaneously the time-around-time is reduced leading to fast clinical management of the patient. This is especially important in intensive care or emergency room, with pediatric and geriatric patients, and in situations where resources are scarce and a rapid, inexpensive, and reliable test can save many lives, such as in developing countries (Baba *et al.*, 2014).

Microfluidics

Microfluidics is a technology based on the analysis of small quantities of fluids (10^{-9} to 10^{-18} liters) through a system of channels ranging from 10 to 100 micrometers. The material used for the construction of the canals is an optically transparent, soft elastomer: the poly(dimethylsiloxane) or PDMS.

The peculiarity of this system lies in the use of very small quantities of samples and reagents, in the low cost, in its high resolution and sensitivity, in the run-time extremely short (Whitesides, 2006).

Microfluidics can be an important tool in situations where, as in developing countries, access to other diagnostic methods for viral diseases is not feasible. Microfluidics can be an excellent tool for POCT, too. First, the diagnosis of viral infection can occur within 15–30 min. The ability to correctly diagnose even low-level infections, combined with high specificity and sensitivity, allows clinicians to set appropriate therapy in a short time with the result of a better therapeutic success (Simpson *et al.*, 2018; Na *et al.*, 2018). Furthermore, the microfluidic platforms have the unique ability to control the microenvironment through the precise manipulation of very small sample volumes for the assay, finding application in the discovery of new antiviral drugs and the development of vaccines (Chi *et al.*, 2016; van der Borg *et al.*, 2018; Chaipan, 2017).

A recent study has demonstrated the ability to screen microfluidic chips and sorting HIV-1 particles according to epitope expression (Chaipan, 2017).

Microfluidic technologies reduce reaction times while providing an increased number of information on the virus compared to the techniques currently used for the measurement of Viremia, such as qPCR.

With this method, up to 12 influenza A subtypes can be found simultaneously in less than 100 min (Shen *et al.*, 2019).

Besides, the simultaneous identification of sensitive and resistant viruses makes it possible to assess the viral resistance to antiviral drugs before being administered to the patient.

Microfluidic platforms also find space in biomedical research with the possibility of carrying out multiple experiments in a single device even with small quantities of a biological sample, reducing time and costs (Na *et al.*, 2018). Such technology did not find so far an extensive application but it is our opinion that shortly the progress made by lab-on-a-chip microtechnologies will help this technology live up to its potential.

Conclusions and Future Perspectives

Laboratory diagnostics of viral diseases are gaining increasing interest from companies, researchers, and clinicians. The new technologies will replace or improve the current ones, allowing them to reach increasingly optimal management of the patient with viral diseases. Furthermore, thanks to them it will be possible to give impetus to the world of virology through the discovery of new pathogenic and non-pathogenic viral species and their interactions with the host.

Over recent years, there has been growing interest in using Artificial Intelligence (AI) in different areas of medicine. Following the recent events related to the COVID-19 pandemic, AI has shown how their application could improve the prediction, detection, and management of infectious diseases by the combination of laboratory results, medical history, symptoms, and instrumental tests, such as chest CT. Furthermore, the AI can be a useful support to the processing of laboratory results by eliminating operator-dependent interpretation errors in diagnostic tests (Mei *et al.*, 2020; Bansal *et al.*, 2020).

However, not all methods are equal and can lead to different results with the same biological sample. It is therefore important that research continues to develop ever more precise, low-cost, and accessible solutions.

References

- Baba, M.M., Vidergar, N., Marcello, A., 2014. Virological point-of-care testing for the developing world. *Future Virology* 9 (6), 595–603.
 Baker, M., 2012. Digital PCR hits its stride. *Nature Methods* 9, 541–544.

- Bansal, A., Padappayil, R.P., Garg, C., *et al.*, 2020. Utility of artificial intelligence amidst the COVID 19 pandemic: A review. *Journal of Medical Systems* 44, 156.
- Barzon, L., Lavezzo, E., Costanzi, G., *et al.*, 2013. Next-generation sequencing technologies in diagnostic virology. *Journal of Clinical Virology* 58 (2), 346–350.
- Bella, R., Kaminski, R., Mancuso, P., *et al.*, 2018. Removal of HIV DNA by CRISPR from patient blood engrafts in humanized mice. *Molecular Therapy – Nucleic Acids* 12, 275–282.
- Briedigkeit, L., Müller-Plathe, O., Schlebusch, H., *et al.*, 1998. Point of care testing. *Journal of Laboratory Medicine* 22, 414–420.
- Chaipan, C., 2017. Single-virus droplet microfluidics for high-throughput screening of neutralizing epitopes on HIV particles. *Cell Chemical Biology* 24 (6), 751–757.
- Chen, J.S., Ma, E., Harrington, L.B., *et al.*, 2018. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 360 (6387), 436–439.
- Chi, C.W., Ahmed, A.R., Dereli-Korkut, Z., Wang, S., 2016. Microfluidic cell chips for high-throughput drug screening. *Bioanalysis* 8 (9), 921–937.
- Chou, T.C., Hsu, W., Wang, C.H., Chen, Y.J., Fang, J.M., 2011. Rapid and specific influenza virus detection by functionalized magnetic nanoparticles and mass spectrometry. *Journal of Nanotechnology* 9, 52.
- Clevenbergh, P., 2000. Persisting long-term benefit of genotype-guided treatment for HIV-infected patients failing HAART. The Viradapt study: Week 48 follow-up. *Antiviral Therapy* 5 (1), 65–70.
- Cobo, F., 2013. Application of MALDI-TOF mass spectrometry in clinical virology: A review. *Open Virology Journal* 7, 84–90.
- Deng, Q., Chen, Z., Shi, L., Lin, H., 2018. Developmental progress of CRISPR/Cas9 and its therapeutic applications for HIV-1 infection. *Reviews in Medical Virology* 28 (5), e1998.
- Diaz-Decaro, J.D., Green, N.M., Godwin, H.A., 2018. Critical evaluation of FDA-approved respiratory multiplex assays for public health surveillance. *Expert Review of Molecular Diagnostics* 18 (7), 631–643.
- Dien Bard, J., Alby, K., 2018. Point-counterpoint: Meningitis/encephalitis syndromic testing in the clinical laboratory. *Journal of Clinical Microbiology* 56 (4).
- Elmifro, E.M., Ashshi, A.M., Cooper, R.J., Klapper, P.E., 2000. Multiplex PCR: Optimization and application in diagnostic virology. *Clinical Microbiology Reviews* 13 (4), 559–570.
- Ford, K., McDonald, D., Mali, P., 2019. Functional genomics via CRISPR-Cas. *Journal of Molecular Biology* 431 (1), 48–65.
- Freer, G., Maggi, F., Pifferi, M., *et al.*, 2018. The virome and its major component, Anellovirus, a convoluted system molding human immune defenses and possibly affecting the development of asthma and respiratory diseases in childhood. *Frontiers in Microbiology* 9, 686.
- Gootenberg, J.S., Abudayyeh, O.O., Kellner, M.J., *et al.*, 2018. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* 360 (6387), 439–444.
- Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W., *et al.*, 2017. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356 (6336), 438–442.
- Gray, J., Coupland, L.J., 2014. The increasing application of multiplex nucleic acid detection tests to the diagnosis of syndromic infections. *Epidemiology and Infection* 142 (1), 1–11.
- Haleyur Giri Setty, M.K., Hewlett, I.K., 2014. Point of care technologies for HIV. *AIDS Research and Treatment* 2014, 497046.
- Hoffmann, B., Scheuch, M., Höper, D., *et al.*, 2012. Novel orthobunyavirus in cattle, Europe, 2011. *Emerging Infectious Diseases* 18 (3), 469–472.
- Houldcroft, C.J., 2017. Clinical and biological insights from viral genome sequencing. *Nature Reviews Microbiology* 15 (3), 183–192.
- Hu, W., Kaminski, R., Yang, F., *et al.*, 2014. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America* 111, 11461–11466.
- Huang, H.S., Tsai, C.L., Chang, J., *et al.*, 2018. Multiplex PCR system for the rapid diagnosis of respiratory virus infection: systematic review and meta-analysis. *Clinical Microbiology and Infection* 24 (10), 1055–1063.
- Johannessen, A., 2015. Where we are with point-of-care testing. *Journal of Viral Hepatitis* 22 (4), 362–365.
- Joung, J., Ali Ladh, A., Saito, M., *et al.*, 2020. Detection of SARS-CoV-2 with SHERLOCK one-pot testing. *New England Journal of Medicine* 383 (15), 1492–1494. doi:10.1056/NEJMc2026172.
- Kaushik, A., Tiwari, S., Dev Jayant, R., Marty, A., Nair, M., 2016. Towards detection and diagnosis of Ebola virus disease at point-of-care. *Biosensors and Bioelectronics* 75, 254–272.
- Khuroo, M.S., Khuroo, N.S., Khuroo, M.S., 2015. Diagnostic accuracy of point-of-care tests for hepatitis C virus infection: a systematic review and meta-analysis. *PLoS One* 10 (3), e0121450.
- Kim, J.H., Park, Y.K., Park, E.S., Kim, K.H., 2014. Molecular diagnosis and treatment of drug-resistant hepatitis B virus. *World Journal of Gastroenterology* 20 (19), 5708–5720.
- Kuang, J., Yan, X., Genders, A.J., Granata, C., Bishop, D.J., 2018. An overview of technical consideration when using quantitative real-time PCR analysis of gene expression in human exercise research. *PLOS One* 13 (5), e0196438.
- Li-Kim-Moy, J., Dastouri, F., Rashid, H., *et al.*, 2016. Utility of early influenza diagnosis through point-of-care testing in children presenting to an emergency department. *Journal of Paediatrics and Child Health* 52 (4), 422–429.
- Maggi, F., Pistello, M., Antonelli, G., 2019. Future management of viral diseases: Role of new technologies and new approaches in microbial interactions. *Clinical Microbiology and Infection* 25 (2), 136–141.
- McCulloh, R.J., Andrea, S., Reinert, S., Chapin, K., 2014. Potential utility of multiplex amplification respiratory viral panel testing in the management of acute respiratory infection in children: A retrospective analysis. *Journal of the Pediatric Infectious Diseases Society* 3 (2), 146–153.
- McMullan, L.K., Folk, S.M., Kelly, A.J., *et al.*, 2012. A new phlebovirus associated with severe febrile illness in Missouri. *New England Journal of Medicine* 367 (9), 834–841.
- Mei, X., Lee, H.C., Diao, K., *et al.*, 2020. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nature Medicine* 26 (8), 1224–1228.
- Myhrvold, C., Freije, C.A., Gootenberg, J.S., *et al.*, 2018. Field-deployable viral diagnostics using CRISPR-Cas13. *Science* 360 (6387), 444–448.
- Na, W., Nam, D., Lee, H., Shin, S., 2018. Rapid molecular diagnosis of infectious viruses in microfluidics using DNA hydrogel formation. *Biosensors and Bioelectronics* 108, 9–13.
- Parker, J., Chen, J., 2017. Application of next-generation sequencing for the detection of human viral pathogens in clinical specimens. *Journal of Clinical Virology* 86, 20–26.
- Pekin, D., Skhiri, Y., Baret, J.-C., *et al.*, 2011. Quantitative and sensitive detection of rare mutations using droplet-based microfluidics. *Lab on a Chip* 11 (13), 2156–2166.
- Piao, J., Jiang, J., Xu, B., 2012. Simultaneous detection and identification of enteric viruses by PCR-mass assay. *PLoS One* 7, e42251.
- Quan, P.L., Sauzade, M., Brouzes, E., 2018. dPCR: A technology review. *Sensors* 18 (4), 1271.
- Rascovan, N., Duraisamy, R., Desnues, C., 2016. Metagenomics, and the human virome in asymptomatic individuals. *Annual Review of Microbiology* 70, 125e41.
- Rosseel, T., Scheuch, M., Höper, D., *et al.*, 2012. DNase SISPA-next generation sequencing confirms Schmallenberg virus in Belgian field samples and identifies genetic variation in Europe. *PLoS One* 7 (7), e41967.
- Roy, B., Zhao, J., Yang, C., *et al.*, 2018. CRISPR/Cas9-mediated genome editing-challenges and opportunities. *Frontiers in Genetics* 9, 240.
- Scagnolari, C., Turriziani, O., Monteleone, K., Pierangeli, A., Antonelli, G., 2017. Consolidation of molecular testing in clinical virology. *Expert Review of Anti-Infective Therapy* 15 (4), 387–400.
- Shen, K., Sabbavarapu, N.M., Fu, C., *et al.*, 2019. An integrated microfluidic system for rapid detection and multiple subtyping of influenza A viruses by using glycan-coated magnetic beads and RT-PCR. *Lab on a Chip* 7.
- Simpson, C., Lee, S.S., Lee, C.S., Yamauchi, Y., 2018. Microfluidics: An untapped resource in viral diagnostics and viral cell biology. *Current Clinical Microbiology Reports* 5 (4), 245–251.
- Strich, J.R., Chertow, D.S., 2019. CRISPR-Cas biology and its application to infectious diseases. *Journal of Clinical Microbiology* 57 (4), e01307–e01318.
- van der Borg, G., Braddock, S., Blijleven, J.S., van Oijen, A.M., Roos, W.H., 2018. Single-particle fusion of influenza viruses reveals complex interactions with target membranes. *Journal of Physics: Condensed Matter* 30 (20), 204005.
- van Diemen, F.R., Kruse, E.M., Hooykaas, M.J., *et al.*, 2016. CRISPR/Cas9-mediated genome editing of herpesviruses limits productive and latent infections. *PLoS Pathogens* 12, e1005701.
- Vu, D.L., Kaiser, L., 2017. The concept of commensal viruses almost 20 years later: Redefining borders in clinical virology. *Clinical Microbiology and Infection* 23, 688e690.
- Wang, G., Zhao, N., Berkhout, B., Das, A.T., 2018. CRISPR-Cas based antiviral strategies against HIV-1. *Virus Research* 244, 321–332.

- Whitesides, G.M., 2006. The origins and the future of microfluidics. *Nature* 442 (7101), 368–373.
- Yin, C., Zhang, T., Qu, X., *et al.*, 2017. In vivo excision of HIV-1 provirus by saCas9 and multiplex single-guide RNAs in animal models. *Molecular Therapy* 25, 1168–1186.
- Yozwiak, M.L., Skewes-Cox, P., Stenglein, M.D., *et al.*, 2012. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLOS Neglected Tropical Diseases* 6 (2), e1485.
- Zhang, H., Morrison, S., Tang, Y.W., 2015. Multiplex polymerase chain reaction tests for detection of pathogens associated with gastroenteritis. *Clinics in Laboratory Medicine* 35 (2), 461–486.
- Zürcher, S., Mooser, C., Lüthi, A.U., 2012. Sensitive and rapid detection of ganciclovir resistance by PCR based MALDI-TOF analysis. *Journal of Clinical Virology* 54, 359–363.

Further Reading

- Behjati, S., Tarpey, P.S., 2013. What is next-generation sequencing? *Archives of Disease in Childhood: Education and Practice Edition* 98 (6), 236–238.

TREATMENT

Antiviral Classification

Guangdi Li, Xixi Jing, and Pan Zhang, Central South University, Changsha, China
Erik De Clercq, Rega Institute for Medical Research, KU Leuven, Leuven, Belgium

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

EC₅₀ The concentration of a drug inducing its half-maximal effective response.

IC₅₀ The concentration of an inhibitor at which 50% of inhibition in its activity is achieved.

Glossary

Allosteric inhibitors Inhibitors block the enzymatic activity by targeting outside the active site of viral enzymes.

Antiviral drug resistance The reduction in the effectiveness of an antiviral agent to treat an infectious disease, probably caused by amino acid mutations within or outside the drug-binding site.

Drug susceptibility The sensitivity of viruses to one or more drugs. If a virus is susceptible, it can be treated with the drug.

Genetic barrier to resistance The threshold above which drug resistance develops to a drug or a drug class.

HAART Highly active antiretroviral therapy which contains a combination of ≥ 3 antiretroviral drugs.

Introduction

During the past decades, more than 100 antiviral agents or their combinations have been approved to treat 9 human infection diseases: HIV, HCV, influenza virus, RSV, HSV, HCMV, VZV, HBV, and variola virus (human smallpox). Antiviral agents can be possibly classified based on their chemical structures, drug targets, or mechanisms of action. For instance, most antiviral agents target either viral enzymes to block the viral replication or viral surface proteins to prevent the viral entry. Regarding the mechanisms of action, nucleoside analogs are effective viral polymerase inhibitors that resemble naturally occurring nucleosides to cause the termination of the nascent viral DNA chain, while protease inhibitors block the proteolytic processing by competing with protease substrate peptides. Based on the chemical structures, aciclovir and valaciclovir are classified as acyclic guanosine analogs to treat DNA viruses such as HSV and VZV, while cidofovir, adefovir, tenofovir alafenamide are acyclic nucleoside phosphonate analogs to treat HCMV, HBV, and HIV, respectively. Due to the variable nature of antiviral agents, this book article attempts to characterize the antiviral classification based on the drug targets in 9 human infectious diseases.

Human Immunodeficiency Virus (HIV)

HIV Reverse Transcriptase

HIV reverse transcriptase (RT) is an asymmetric heterodimer which harbors two enzymatic domains: the RNA- and DNA-dependent DNA polymerase and the ribonuclease H (Li and De Clercq, 2016). After the viral entry of HIV particles, HIV RT produces the viral DNA genome based on the template of the single-stranded viral RNA genome released from HIV particles (Das et al., 2019). To inhibit HIV genome replication, nucleoside RT inhibitors (NRTIs) and non-nucleoside RT inhibitors (NNRTIs) have been successfully developed to target HIV RT – an indispensable enzyme for HIV replication. On the one hand, the 5'-triphosphate of the NRTIs act as chain terminators. Because NRTIs lack a 3'-OH group, the incorporation of NRTIs into the newly synthesized viral dsDNA terminates the elongation of DNA primer, thereby blocking HIV reverse transcription (Das and Arnold, 2013). On the other hand, NNRTIs are allosteric inhibitors that allosterically target a hydrophobic pocket located approximately 10–15 Å from the catalytic site of HIV-1 RT (Sluis-Cremer and Tachedjian, 2008). This causes conformation changes of the HIV-1 RT catalytic site, thus interrupting the viral dsDNA replication (De Clercq and Li, 2016).

As of September 2020, fifteen HIV RT inhibitors have been approved by the US FDA for clinical use, including (1) seven NRTIs: zidovudine, didanosine, zalcitabine, stavudine, lamivudine, abacavir, emtricitabine; two NtRTIs (nucleotide RT inhibitors): tenofovir disoproxil fumarate and tenofovir alafenamide; and (2) six NNRTIs: nevirapine, delavirdine, efavirenz, etravirine, rilpivirine, and doravirine. Elvitegravir has only been approved in Russia. For more than two decades, NRTIs have been clinically used as the backbone in the highly active antiretroviral therapy (HAART, a combination of ≥ 3 antiretroviral drugs), while NNRTIs often serve as the third agent. Currently, tenofovir alafenamide and doravirine are widely administered in clinical practice because of a high genetic barrier to resistance, high safety profile, and low frequency of administration (Deeks, 2018; De Clercq, 2018). Novel RT inhibitors such as islatravir and dapivirine are currently under development (Baeten et al., 2016; Nel et al., 2016; Nakata et al., 2007).

HIV Protease

HIV protease plays an indispensable role in the proteolytic processing of Gag and Gagpol polyproteins to release key structural proteins (matrix, capsid, nucleocapsid, p6) and viral enzymes (reverse transcriptase, integrase, protease) (Li and De Clercq, 2016). HIV protease is a homodimer of two subunits and the active site of HIV protease is located at the center of the substrate-binding tunnel. This substrate-binding tunnel has been recognized as a conserved drug-binding pocket for the development of HIV protease inhibitors to prevent proteolytic processing (Ko *et al.*, 2010).

As of September 2020, 10 protease inhibitors have been approved by the US FDA. The era of protease inhibitors began in 1995 when saquinavir was the first protease inhibitor approved for clinical use. Subsequently, the US FDA approved ritonavir (a protease inhibitor and pharmacokinetic booster) in 1996, indinavir in 1996, nelfinavir in 1997, amprenavir in 1999, lopinavir and atazanavir in 2000, fosamprenavir in 2003, tipranavir in 2005, and darunavir in 2006. Over the past decades, HIV protease inhibitors have become an important ingredient of HAART (Hammer *et al.*, 1997; Gulick *et al.*, 1997). Nevertheless, the rapid emergence of resistance and poor bioavailability pose a challenge to the clinical use of protease inhibitors (Subbaiah *et al.*, 2017). Currently, darunavir is the most popular protease inhibitor (Deeks, 2014), while saquinavir and amprenavir have been virtually discontinued (De Clercq and Li, 2016).

HIV Integrase

HIV integrase, encoded by the HIV Pol gene, is an indispensable enzyme for integrating viral DNA genomes into human chromosomes by a series of DNA cutting and joining reactions (Asante-Appiah and Skalka, 1999; Esposito and Craigie, 1999). Since the viral DNA synthesized by HIV reverse transcriptase is initially blunt-ended, there are two critical steps of viral integration: (1) the 3' end processing reaction removes two nucleotides from each of the 3' ends of the viral dsDNA before viral integration; and (2) the DNA strand transfer reaction that breaks the human chromosome and 3' ends of viral DNA are joined to the 5' ends of human chromosome at the integration sites (Li *et al.*, 2015). Antiviral compounds have been screened to inhibit either the 3' end processing reaction or the DNA strand transfer reaction (Delelis *et al.*, 2008). Compounds against the 3' end processing reaction are inactive, but integrase strand transfer inhibitors (INSTIs) have shown potent antiviral activities in both in vitro and in vivo studies (Delelis *et al.*, 2008). INSTIs selectively bind near the 3' end of the viral DNA in the structural complex of viral DNA and HIV integrase, thereby blocking the DNA strand transfer reaction (Delelis *et al.*, 2008).

As of September 2020, four INSTIs (raltegravir, elvitegravir, dolutegravir, bictegravir) have been approved by the US FDA. The first-generation INSTIs such as raltegravir and elvitegravir share a low genetic barrier to resistance, causing the loss of virologic activity in clinical studies (Brooks *et al.*, 2019). In contrast, dolutegravir and bictegravir – two second-generation INSTIs – exhibit a higher genetic barrier to resistance, limited cross-resistance, and less drug-drug interactions (Oliveira *et al.*, 2018; Podany *et al.*, 2017). Currently, the WHO, IAS-USA, and EACS guidelines recommend the use of INSTIs in the first-line HIV regimens (Brooks *et al.*, 2019). Novel INSTIs such as cabotegravir are undergoing clinical development (Orkin *et al.*, 2020).

HIV GP41

GP41 is a transmembrane protein encoded by the env gene. As a key structure protein, HIV GP41 binds with GP120 to form HIV spike trimers on the surface of HIV particles (Mao *et al.*, 2012). During the viral entry, the N-heptad repeat (NHR) and C-heptad repeat (CHR) of GP41 switch to a six-helix bundle (6-HB) structure which binds to the human cell membrane and drives the viral fusion into the human cells. Many antiviral agents have been developed to target the hydrophobic pocket within the NHR trimer, therefore preventing viral entry (Mostashari Rad *et al.*, 2018).

As the only GP41 inhibitor approved by the US FDA, enfuvirtide is a fusion peptide inhibitor that mimics the N-heptad repeat and prevents the formation of the six-helix bundle structure of GP41. Clinical use of enfuvirtide requires twice-daily subcutaneous injection (90 mg/Kg for adults, 2 mg/Kg for children aged 6–16 years) (Kitchen *et al.*, 2008). Enfuvirtide is not commonly used in clinical practice because of its side effects (eosinophilia, neutropenia, increased risk of bacterial pneumonia), its short half-life, and lack of oral availability (Reust, 2011). Although many attempts have been made, the development of GP41 inhibitors remains difficult due to the emerging mutations and structural dynamics of HIV GP41.

HIV GP120

GP120, encoded by the env gene, is an envelope glycoprotein on the surface of HIV particles. During the viral entry, HIV GP120 binds to the CD4 receptor and the CCR5/CXCR4 co-receptor on the surface of human CD4⁺ T cells (Falkenhagen and Joshi, 2018; Shaik *et al.*, 2019). The Phe43 pocket and other conserved regions on the GP120 are considered as the real targets for developing anti-HIV drugs (Mostashari Rad *et al.*, 2018). The drugs targeting GP120 can be called adhesion inhibitors, mainly blocking the binding of GP120 with CD4 (Pu *et al.*, 2019).

Fostemsavir (BMS-663068) is the first GP120-directed attachment inhibitor approved by the US FDA on July 3rd, 2020. Fostemsavir tromethamine is the prodrug of temsavir – the active moiety that binds to the conserved outer domain of GP120 which is adjacent to the CD4 binding loop (Langley *et al.*, 2015). This binding inhibits the exposure of the chemokine co-receptor binding site and prevents the initial interaction of HIV GP120 with the CD4 receptor on human T cells and other immune cells.

(Langley *et al.*, 2015). Fostemsavir is recommended for heavily treatment-experienced adults with multidrug-resistant HIV-1 infections. As the only compound in the class of GP120 drugs, fostemsavir harbors a unique resistance profile and exhibits no cross-resistance with other entry inhibitors (Lataillade *et al.*, 2018).

Hepatitis C Virus (HCV)

HCV NS3/4A Protease

The nonstructural protein 3 (NS3) contains a serine protease domain and a helicase domain at the N-terminal and C-terminal regions, respectively (McGivern *et al.*, 2015). The NS3 protein acts as a serine protease to cleave HCV polyprotein precursor at four junctions (NS3-NS4A, NS4A-NS4B, NS4B-NS5A, NS5A-NS5B) to release 5 viral proteins. As a short 54-amino-acid polypeptide, the nonstructural protein 4A (NS4A) serves as an essential co-factor for the NS3 serine protease in a non-covalent complex. NS3 protease is an attractive drug target because its inhibition can terminate the polyprotein processing and restore interferon gene expression (McGivern *et al.*, 2015). NS3/4A protease inhibitors are a key element of direct-acting antivirals that effectively cure HCV infections by targeting one or more viral proteins (Li and De Clercq, 2017).

As of today, ten NS3/4A protease inhibitors have been approved, including glecaprevir, grazoprevir, paritaprevir, simeprevir, vaniprevir, voxilaprevir, asunaprevir, danoprevir, boceprevir (discontinued), and telaprevir (discontinued) (Li and De Clercq, 2017). Moreover, danoprevir was approved in China to treat patients infected with HCV genotype 1b (Miao *et al.*, 2020). Novel NS3/4A inhibitors such as seraprevir are currently under development. For instance, seraprevir is now evaluated in a phase 3 trial (NCT04001608).

HCV NS5A Phosphoprotein

The nonstructural protein 5A (NS5A) is a zinc-binding and proline-rich hydrophilic phosphoprotein that executes versatile functions in genome replication, viral particle assembly, and human-virus interactions (Ross-Thriepland and Harris, 2015). The NS5A protein in the form of a dimer or multimer localizes to the ER-derived membranes via its amphipathic helix domain. NS5A inhibitors can target the amphipathic helix domain of NS5A dimer to inhibit the conformation of double-membrane vesicles and impair HCV RNA replication factories (Shanmugam *et al.*, 2018).

As of today, six NS5A inhibitors (ombitasvir, elbasvir, velpatasvir, daclatasvir, ledipasvir, pibrentasvir) have been approved by the US FDA. These NS5A inhibitors efficiently block in vitro and in vivo viral replication, though their exact models of drug actions are yet to be elucidated (Ross-Thriepland and Harris, 2015). Moreover, the clinical efficacy of NS5A inhibitors plus other direct-acting antivirals offers more than 90% of sustained virologic responses (Li and De Clercq, 2017). Many experimental NS5A inhibitors such as ravidasvir are still under development. Ravidasvir plus sofosbuvir offered promising virologic response rates in 298 patients infected with HCV genotype 4 (Esmat *et al.*, 2017).

HCV NS5B Polymerase

The nonstructural protein 5B (NS5B) is an RNA-dependent RNA polymerase that is indispensable for HCV RNA synthesis and genome replication. The structure of NS5B consists of three subdomains: “palm”, “finger”, “thumb”, and the hydrophobic membrane anchoring C-terminus. As a promising drug target, the catalytic site of NS5B is encircled by the finger and thumb domains. NS5B inhibitors that alter structural conformation or interfere with viral RNA binding exhibit promising potency to inhibit HCV RNA replication (Kirby *et al.*, 2015).

Two NS5B inhibitors have been approved by the US FDA, including sofosbuvir (GS-7977; formerly PSI-7977) in December 2013 and dasabuvir (ABT-333) in December 2014. The NS5B inhibitors can be mainly divided into either nucleoside inhibitors (e.g., sofosbuvir) or non-nucleoside inhibitors (e.g., dasabuvir) (Li and De Clercq, 2017). On the one hand, nucleoside inhibitors block the viral RNA synthesis by mimicking natural substrates and competing with incoming nucleoside triphosphates at the catalytic site of NS5B (Li and De Clercq, 2017). On the other hand, non-nucleotide inhibitors noncompetitively block the allosteric pockets outside the catalytic site to prevent viral RNA synthesis (Li and De Clercq, 2017). For instance, dasabuvir (Kati *et al.*, 2015), GSK5852 (Voitenleitner *et al.*, 2013), beclabuvir (Gentles *et al.*, 2014), and filibuvir (Fenaux *et al.*, 2013) target the allosteric drug pockets in the palm I, palm II, thumb I, and thumb II domains, respectively. Although only two NS5B inhibitors have been approved, many novel NS5B inhibitors (e.g., beclabuvir) are still under development.

Respiratory Syncytial Virus (RSV)

RSV RNA Polymerase

RSV RNA-dependent RNA polymerase – a complex comprising the viral large polymerase subunit, the phosphoprotein, and the transcription elongation factor M2-1 – plays a critical role in viral mRNA transcription, mRNA capping/methylation, and genome

replication (Fearn and Deval, 2016). The large polymerase subunit harbors the enzymatic domains that produce subgenomic mRNAs, RNA replication, and an antigenome RNA for genome RNA synthesis, while the other proteins act as essential cofactors (Fearn and Deval, 2016). Due to its unique features and essential nature, the large polymerase protein has proved to be a promising drug target for the development of nucleoside and non-nucleoside inhibitors.

As of today, ribavirin is an FDA-approved nucleoside inhibitor that inhibits the activity of RSV RdRp, whereas it is largely discontinued to treat RSV infections due to limited efficacy and risk of serious side effects (Shook and Lin, 2017). Experimental inhibitors such as ALS-8176 and PC786 are still under clinical development. ALS-8176 (lumicitabine) is a nucleoside inhibitor with promising anti-RSV efficacy and safety in phase 2a study (NCT02094365) (DeVincenzo *et al.*, 2015; Wang *et al.*, 2015). PC786 is a potent non-nucleoside inhibitor that blocks RSV type A strains (IC_{50} : <0.09–0.71 nM) and RSV type B strains (IC_{50} : 1.3–50.6 nM) in cell cultures (Coates *et al.*, 2017).

RSV Fusion Glycoprotein

RSV fusion glycoprotein is a class I fusion protein that anchors in the membrane of viral particles via a transmembrane domain (Gilman *et al.*, 2019). During the viral entry, RSV fusion glycoprotein in the trimeric form is undergoing dramatic changes from the metastable prefusion conformation to the highly stable post-fusion conformation, thereby driving the fusion of viral membrane with the human cell membranes (Gilman *et al.*, 2019). RSV fusion glycoprotein that induces RSV-neutralizing antibody responses is a key antigen for protective immunity (Tang *et al.*, 2019). As a leading target of neutralizing antibodies and vaccines, RSV fusion glycoproteins have conserved sequences across different isolates of RSV type A and B strains (Tang *et al.*, 2019).

As of September 2020, palivizumab that binds to RSV fusion glycoprotein remains the only monoclonal antibody approved for the prevention of RSV infection in high-risk infants. Nevertheless, the clinical application of palivizumab is uncommon because of its low stability and its limited efficacy of 50% (Soto *et al.*, 2020). Ongoing studies are currently evaluating many experimental inhibitors of RSV fusion glycoprotein, including preatovir, ziresovir, sisunatovir, MDT-637, JNJ-53718678, and ALX-0171. For instance, preatovir (GS-5806) efficiently inhibits the fusion glycoprotein of RSV-A (EC_{50} : 0.51 ± 0.25 nM) and RSV-B (EC_{50} : 0.35 ± 0.15 nM) strains in cell cultures (Perron *et al.*, 2015), and it has completed clinical phase 2 trials (NCT02135614). Ziresovir is another selective and orally bioavailable inhibitor in phase 3 clinical trial (NCT04231968) (Zheng *et al.*, 2019).

Herpes Simplex Virus (HSV)

HSV DNA Polymerase UL30

HSV DNA replication requires seven viral proteins: (1) an origin binding protein UL9, (2) a single-strand binding protein ICP8; (3) a viral polymerase with two subunits UL30 and UL42; (iv) a helicase-primase complex with three subunits UL5, UL8, and UL52 (Chen *et al.*, 2011). HSV DNA polymerase is a heterodimeric complex with two subunits: (1) the catalytic UL30 subunit which offers both polymerase and proofreading exonuclease activities; (2) the UL42 subunit that promotes processivity by tethering UL30 to viral DNA via its direct DNA binding (Vashishtha and Kuchta, 2015). The UL30 subunit belongs to the B-family of polymerases and forms a typical hand-shaped structure with the catalytic residues (D717, D888) located in the palm domain for dNTP polymerization (Vashishtha and Kuchta, 2015). The UL30 subunit can replicate viral DNA with high fidelity, even in the absence of the UL42 subunit. Due to its indispensable role, the UL30 subunit has been considered as a promising drug target to block viral DNA replication (Zarrouk *et al.*, 2017).

As of September 2020, many UL30 inhibitors such as idoxuridine, brivudine, trifluridine foscarnet, aciclovir, famciclovir, valaciclovir, and penciclovir have been approved for HSV treatment. Because current compounds face the challenge of drug resistance and adverse reactions (Vollmer *et al.*, 2019), many experimental inhibitors such as synguanol (Vollmer *et al.*, 2019), mitoxantrone dihydrochloride (Huang *et al.*, 2019), and PHA767491 (Hou *et al.*, 2017) are under development. For instance, synguanol is a methylenecyclopropane nucleoside analog that competitively inhibits HSV-1 UL30 (IC_{50} : 0.33 ± 0.16 μ M) (Vollmer *et al.*, 2019). Mitoxantrone dihydrochloride inhibits viral DNA synthesis by blocking the transcription of UL30 (IC_{50} : 1.21 μ M) (Huang *et al.*, 2019). PHA767491 also inhibits UL30 to reduce the expression of HSV viral genes (UL5, UL8, UL29, UL30, UL42, UL52) (IC_{50} : 1.86 μ M) (Hou *et al.*, 2017). Antiviral agents such as pritelivir and amenamevir that target HSV helicase-primase complex are still under development (Poole and James, 2018).

HSV Envelope Protein

Although HSV particles harbor at least 15 envelope proteins, four envelope glycoproteins (gD, gB, gH, gL) play an indispensable role in viral entry into all permissive cell types (Agelidis and Shukla, 2015). The viral entry begins with the binding of gD to a human receptor (nectin-1, herpesvirus entry mediator, or 3-O-sulfated heparan sulfate) (Atanasiu *et al.*, 2018). This gD-receptor binding drives the conformation changes in gD to activate the regulatory proteins gH and gL, leading to the activation of gB into a fusogenic state for membrane fusion (Agelidis and Shukla, 2015).

Docosanol (*n*-docosanol; behenyl alcohol) is a naturally occurring antiherpetic agent approved by the US FDA as a topical treatment for herpes labialis, as well as the over-the-counter medication for cold sores and fever blisters (De Clercq and Li, 2016).

Although its exact mechanism of action remains unclear, docosanol may inhibit the interactions between HSV envelope proteins and human receptors (De Clercq and Li, 2016). Experimental inhibitors such as C19 and NGI-1 are still under development (Rinis *et al.*, 2018; Lu *et al.*, 2019).

Human Cytomegalovirus (HCMV)

HCMV DNA Polymerase UL54

HCMV DNA polymerase is a multiprotein complex that contains a catalytic subunit UL54 and a processivity factor UL44 to synthesize long stretches of viral DNA during viral replication (Appleton *et al.*, 2004). The homodimer UL44 binds to viral DNA to prevent dissociation from the template, thereby promoting the long-chain DNA synthesis within the catalytic subunit of UL54 (Appleton *et al.*, 2004). Due to its essential role in viral DNA synthesis, UL54 has been recognized as an important drug target against HCMV.

As of today, cidofovir, ganciclovir, foscarnet, valganciclovir, and fomivirsen (discontinued) have been approved for HCMV treatment. To inhibit viral DNA synthesis, cidofovir and ganciclovir are converted to DNA polymerase substrate analogs by the thymidine kinase of CMV. UL54 gene is prone to mutations that limit the efficiency of antiviral agents (Chou *et al.*, 2016). Furthermore, cidofovir may induce nephrotoxicity especially in bone marrow transplant patients (Piret *et al.*, 2017), which limits its clinical use. Novel compounds such as CMX001 (brincidofovir) are currently under development. CMX001 can be converted to cidofovir. The drug potency is much higher for CMX001 than for cidofovir. Cidofovir from CMX001 does not accumulate in the kidneys, therefore nephrotoxicity could be greatly reduced during the treatment (Marty *et al.*, 2013).

HCMV Terminase UL56

HCMV DNA terminase complex is a hetero-oligomer composed of UL56, UL89, and additional viral subunits (UL51, UL52, UL77, UL93) (Ligat *et al.*, 2018). This terminase complex cleaves HCMV DNA concatemers and packages the genome into the capsid for the DNA-packing process during the viral maturation and packaging (Ligat *et al.*, 2018). The large terminase subunit UL56 plays an essential role in the viral DNA cleavage and packaging (Ligat *et al.*, 2018). UL56 not only has the ATPase activity that hydrolyzes ATP to provide energy for the genome cutting and transfer activities but also has ATP-independent endonuclease activity driven by UL89 (Ligat *et al.*, 2018). As an indispensable viral protein, UL56 has been considered to be a promising drug target.

As of September 2020, letermovir remains the only UL56 inhibitor approved by the US FDA for preventing HCMV in adult HCMV-seropositive recipients of an allogeneic hematopoietic stem cell transplant. Based on a phase 3 clinical trial (NCT02137772), letermovir prophylaxis significantly reduced the risk of HCMV infection, and its adverse effects were mild (Marty *et al.*, 2017). Moreover, letermovir is a viral DNA-packaging inhibitor, remarkably specific for HCMV, but not other herpesviruses (Ligat *et al.*, 2018).

Human Influenza Virus

Viral RNA Polymerase

The influenza polymerase is comprised of polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2), plus either polymerase acidic protein (PA) in influenza A/B or polymerase 3 protein (P3) in influenza C/D viruses (Takashita, 2020). The PB1 subunit encodes the RNA-dependent RNA polymerase, while the PA and PB2 subunits offer the endonuclease and cap-binding activities, respectively (Wandzik *et al.*, 2020). During the transcription of viral mRNA, the cap-snatching of capped oligomers from human Pol II transcripts is mediated by the PB2 cap-binding and PA endonuclease activities. Subsequently, the PB1 active site undertakes the prime template-directed RNA synthesis, followed by the initiation, elongation, termination, and recycling of typical polymerase activities (Wandzik *et al.*, 2020). Due to their importance, PA, PB1, and PB2 have been considered as promising antiviral targets.

As of September 2020, the PA inhibitor baloxavir marboxil (S-033188) and the PB1 inhibitor favipiravir (T-705) have been approved for clinical use, while the PB2 inhibitor pimodivir (JNJ-63623872, VX-787) is currently evaluated in phase 3 trials (NCT03381196, NCT03376321). As a pyrazine derivative, favipiravir received conditional marketing approval only for patients with a novel or reemerging influenza when other antivirals are ineffective because favipiravir increases the risk for teratogenicity and embryotoxicity (Takashita, 2020). Baloxavir marboxil was approved for treating influenza A and B viruses, while its pattern of drug susceptibility follows the order: influenza A > B > C > D (Takashita, 2020). As a cyclohexyl carboxylic acid analog, pimodivir inhibits viral RNA binding by occupying the cap-binding domain of PB2, and it shows strong activity against influenza A but not B viruses (Takashita, 2020). Other novel polymerase inhibitors are still under development.

Neuraminidase

Influenza surface glycoprotein neuraminidase is known for its multifunctional roles in viral entry and viral release. During the viral entry, the glycoprotein neuraminidase contributes to the viral binding to the sialic acid receptors of human cell glycoproteins,

Table 1 Summary of viral proteins targeted by approved and novel inhibitors

<i>Human viruses</i>	<i>Viral targets</i>	<i>Approved drugs</i>	<i>Novel inhibitors</i>
Human immunodeficiency virus (HIV)	Protease	Amprenavir, atazanavir, darunavir fosamprenavir, indinavir, lopinavir, nelfinavir, ritonavir, saquinavir, tipranavir,	TMB-607, TMC-310911, GRL-09510
	Reverse transcriptase	NRTIs: abacavir, didanosine, emtricitabine, lamivudine, stavudine, tenofovir alafenamide, tenofovir disoproxil fumarate, zidovudine, NNRTIs: delavirdine ^a , doravirine, efavirenz, elvitegravir ^b , etravirine, nevirapine, rilpivirine, Bictegravir, dolutegravir, elvitegravir, raltegravir,	MK-8504, MK-8583, racivir, islatravir, rovafovir etalafenamide, censavudine, amdoxovir, elvucitabine, KM-023
	Integrase gp41 gp120	Albuvirtide ^b , enfuvirtide Fostemsavir	Cabotegravir GSK373239
Hepatitis C virus (HCV)	NS3/4A protease	Danoprevir ^b , glecaprevir, grazoprevir, paritaprevir, paritaprevir, Simeprevir,	Asunaprevir, faldaprevir furaprevir, narlaprevir, Seraprevir, vaniprevir, vedroprevir
	NS5A phosphoprotein NS5B polymerase	Daclatasvir, ledipasvir, ombitasvir, elbasvir, velpatasvir, pibrentasvir Sofosbuvir, dasabuvir	Ravidasvir, ruzasvir, odalasvir Adafosbuvir, deleobuvir, Lomibuvir, mericitabine, Radalbuvir, radalbuvir,
Human influenza virus	RNA polymerase	Baloxavir marboxil, favipiravir ^b	Pimodivir (VX-787)
	Neuraminidase Matrix protein 2	Laninamivir ^b , oseltamivir, peramivir, zanamivir, Amantadine ^a , rimantadine	Isocorilagin
Respiratory syncytial virus (RSV)	RNA polymerase	Ribavirin	Lumicitabine (ALS-8176)
	Fusion glycoprotein	Palivizumab	Presatovir (GS-5806), ziresovir (RO-0529, AK0529), MDT-637, JNJ-53718678, sisunatovir, ALX-0171
Herpes simplex virus (HSV)	DNA polymerase UL30	Aciclovir, brivudine, famciclovir, foscarnet, idoxuridine, penciclovir trifluridine, valaciclovir	Synguanol, filociclovir, MBX-2168; mitoxantrone dihydrochloride; PHA767491
	Envelope proteins	Docosanol	NGI-1, C19
Human cytomegalovirus (HCMV)	DNA polymerase UL54	Cidofovir, fomivirsen, foscarnet, ganciclovir, valganciclovir,	Filociclovir
	Terminase UL56	Letermovir	
Varicella-zoster virus (VZV)	DNA polymerase	Aciclovir, brivudine, famciclovir, valaciclovir, vidarabine	
Hepatitis B virus (HBV)	DNA polymerase	Adefovir, besifovir ^b , clevudine ^b entecavir, telbivudine, tenofovir alafenamide, tenofovir	Tenofovir exalidex
Human smallpox	VP37 envelope wrapping protein	Tecovirimat	

^aDiscontinued.^bElsulfavirine was approved in Russia, albuvirtide was approved in China; favipiravir and laninamivir were approved in Japan; danoprevir was approved in China; clevudine was approved in South Korea and the Philippines; besifovir was approved in South Korea.

leading to the enhancement of hemagglutinin receptor binding (Wen and Wan, 2019). During the viral release, neuraminidase cleaves sialic acids from cellular receptors and neuraminidase/hemagglutinin on nascent influenza virions so as to prevent virion aggregation and release new particles (McAuley *et al.*, 2019). Due to its importance, neuraminidase has been recognized as a promising antiviral target (Kumar *et al.*, 2020).

As of September 2020, neuraminidase inhibitors such as oseltamivir, zanamivir, peramivir, and laninamivir have been approved for clinical use (Table 1). Oseltamivir is used orally as the first-line therapy, but its effectiveness is compromised by the development of drug resistance mutations in influenza A genotypes such as H3N2 and H5N1 (Kumar *et al.*, 2020). Due to the emergence of drug resistance, it remains an urgent need to develop anti-influenza agents. Novel antivirals such as A-192558 and A-315675 are still under development (Gubareva and Mohan, 2020).

Matrix Protein 2

The influenza matrix protein 2 (M2) is a 97-residue single-pass membrane protein that forms pH-gated proton channels in the viral lipid envelope (Schnell and Chou, 2008). By shuttling protons inwards and outwards through the viral membrane, matrix protein 2 equilibrates pH across the viral membrane during the viral entry and across the trans-Golgi membrane of host cells during the viral maturation (Mandala *et al.*, 2020). Because the channel pore is essential for the proton shuttling, the pore of the M2 channel is a promising drug-binding pocket (Gu *et al.*, 2013).

As of September 2020, rimantadine and amantadine that target the M2 channel pore have been approved for anti-influenza treatment. For instance, amantadine targets the M2 channel by hydrophobic interactions between the adamantane group and the N-terminal gate of the channel (Gu *et al.*, 2013). Mutations (e.g., V27A, L26F) within the channel pore weaken the hydrophobic interactions of rimantadine and amantadine with the M2 channel, thereby causing drug resistance and treatment failure (Gu *et al.*, 2013). Novel inhibitors that are less prone to mutations may lead to better antiviral efficacy.

Variola Virus (Human Smallpox)

VP37 Envelope Wrapping Protein

The extinction of variola virus, the etiological agent of human smallpox was declared by the WHO in 1980 (Jordan *et al.*, 2010). The F13L gene of variola virus encodes a highly conserved 37 kDa peripheral membrane protein called VP37. Before the viral budding, the wrapping complex requires VP37 and other viral proteins to interact with human membrane proteins; subsequently, it catalyzes the maturation of intracellular viral particles into the egress-competent form of the variola virus particles (Jordan *et al.*, 2010). VP37 interacts with human proteins Rab9 and TIP47 (a Rab9-specific effector) in membrane fractions from infected cells to facilitate assembly of extracellular virus (Chen *et al.*, 2009).

As of September 2020, tecovirimat (ST-246) remains the only VP37 inhibitor approved for treating human smallpox, although its effectiveness has only been shown in animal models but not humans due to the extinction of the virus in human populations. As an 4-trifluoromethyl phenol derivative, tecovirimat blocks the interactions of VP37 with Rab8 and TIP47, thereby inhibiting the maturation of egress-competent enveloped virions for viral budding (Jordan *et al.*, 2010). Novel compounds such as NIOCH-14 are still under development (Delaune and Iseni, 2020).

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)

SARS-CoV-2 Polymerase

SARS-CoV-2 RNA-dependent RNA polymerase, encoded by the NSP12 gene, forms a polymerase complex with its viral cofactors NSP7 and NSP8 to catalyze the viral RNA synthesis during the viral replication and transcription (Gao *et al.*, 2020). Similar to SARS-CoV polymerase, the structure of SARS-CoV-2 polymerase contains three subdomains: fingers, a thumb, and a palm, while the catalytic site (S759-D760-D761) is located within the conserved motifs of the palm domain (Gao *et al.*, 2020). The cofactors NSP7 and NSP8 interact with the thumb, while another NSP8 binds to the fingers, thereby conferring processivity to NSP12 (Hillen *et al.*, 2020). Due to its indispensable role in viral RNA replication, SARS-CoV-2 polymerase is an important target for antiviral drug development (Li and De Clercq, 2020; Zhou *et al.*, 2020; Jiang *et al.*, 2020).

As of October 2020, remdesivir (GS-5734) has been authorized in the US for clinical use as a polymerase inhibitor to treat patients infected with SARS-CoV-2. Remdesivir acts as a delayed chain terminator to interfere with the RNA synthesis of viral polymerase and evades the proofreading of viral exonuclease (Eastman *et al.*, 2020). The intravenous use of remdesivir could shorten the recovery time and reduce the mortality rate of SARS-CoV-2 based on a randomized, double-blinded, placebo-controlled trial (NCT04280705) (Olalla, 2020). Other novel experimental inhibitors such as GS-441524 (Yan and Muller, 2020) and favipiravir (Arab-Zozani *et al.*, 2020) might offer antiviral activities against SARS-CoV-2, but their clinical use requires further investigation.

Varicella Zoster Virus (VZV)

VZV DNA Polymerase

VZV polymerase belongs to the DNA polymerase type-B family. As the essential component of VZV replication complex, DNA polymerase is encoded by the open reading frame 28 (ORF28). VZV DNA polymerase is an ideal target for the development of nucleoside analogs because it plays a critical role in viral DNA replication during the life cycle.

As of September 2020, nucleoside analogs such as aciclovir, famciclovir, valaciclovir, brivudine, and vidarabine have been approved for VZV treatment (De Clercq and Li, 2016). These compounds could effectively target VZV polymerase and block viral DNA synthesis. As a prodrug of penciclovir, famciclovir effectively prevents the recurrence of VZV, and it has only mild adverse events (Wang *et al.*, 2020). Furthermore, both aciclovir and famciclovir interventions offer high rates of benefit and showed a similar time to full crusting of lesions (Pott Junior *et al.*, 2018).

Hepatitis B Virus (HBV)

HBV DNA Polymerase

HBV RNA-dependent DNA-dependent polymerase, encoded by the open reading frame P, acts with multifaceted functions such as (1) HBV reverse transcription that synthesizes the (–) DNA strand from the viral RNA template within the catalytic site of HBV polymerase; (2) degradation of the viral RNA template by the RNase H activity of HBV polymerase (Menendez-Arias *et al.*, 2014). HBV polymerase is mainly comprised of three functional domains: the template domain (AA positions: 1–183), the polymerase domain (349–691), and the RNase H domain (692–845) (Menendez-Arias *et al.*, 2014). The catalytic site of the polymerase domain is considered as a promising drug target, whilst nucleos(t)ide inhibitors have been developed to compete with the incorporation of natural nucleotide substrates into the elongating DNA chain, thereby blocking viral DNA synthesis (De Clercq and Li, 2016).

As of September 2020, many nucleos(t)ide inhibitors (entecavir, telbivudine, adefovir dipivoxil, tenofovir disoproxil fumarate, tenofovir alafenamide, clevudine, and besifovir) have been approved for clinical use. Note that clevudine was only approved in South Korea and the Philippines, while besifovir was approved in South Korea. Despite the success of nucleos(t)ide inhibitors to suppress the viral replication, they do not eliminate the virus from the hepatocytes, and a cure of HBV infection remains yet to be discovered. Novel HBV polymerase inhibitors such as tenofovir exalidex are still under clinical development (Martinez *et al.*, 2020).

Conclusion

Based on the classification of approved antiviral agents, several features could be summarized. First, antiviral agents mostly target the viral proteins with high specificity, leading to less toxicity compared with human protein targets. Second, among 9 infectious diseases, the most popular drug targets could be listed as follows: viral polymerase, viral envelope glycoproteins, and viral protease. These three viral proteins play a dispensable role in viral replication, making them promising viral targets for most human viruses. Third, antiviral agents are mostly small molecules that can be taken orally in clinical settings, while few antibodies and peptide inhibitors have been approved for antiviral use. Compared with monoclonal antibodies and peptides, small molecules are typically cheaper, chemically stabler, structurally simpler and better permeable.

Although this article only describes 10 infectious diseases with approved drugs, future antiviral development should also focus on emerging infectious diseases such as coronaviruses, dengue, Zika, and Ebola. On October 14, 2020, the US FDA approved a mixture of three monoclonal antibodies (atoltivimab, maftivimab, odesivimab-ebgn), which marks the first FDA approval for the treatment of Ebola virus infection. Furthermore, it remains critical improving antivirals to combat emerging drug resistance mutations because drug resistance mutations have been observed in many viruses (e.g., HIV, HBV, influenza).

References

- Agelidis, A.M., Shukla, D., 2015. Cell entry mechanisms of HSV: What we have learned in recent years. *Future Virology* 10, 1145–1154.
- Appleton, B.A., Loregian, A., Filman, D.J., Coen, D.M., Hogle, J.M., 2004. The cytomegalovirus DNA polymerase subunit UL44 forms a C clamp-shaped dimer. *Molecular Cell* 15, 233–244.
- Arab-Zozani, M., Hassanipour, S., Ghoddoosi-Nejad, D., 2020. Favipiravir for treating patients with novel coronavirus (COVID-19): Protocol for a systematic review and meta-analysis of randomised clinical trials. *BMJ Open* 10, e039730.
- Asante-Appiah, E., Skalka, A.M., 1999. HIV-1 integrase: Structural organization, conformational changes, and catalysis. *Advances in Virus Research* 52, 351–369.
- Atanasiu, D., Saw, W.T., Lazear, E., *et al.*, 2018. Using antibodies and mutants to localize the presumptive gH/gL binding site on herpes simplex virus gD. *Journal of Virology* 92, e01694.
- Baeten, J.M., Palanee-Phillips, T., Brown, E.R., *et al.*, 2016. Use of a vaginal ring containing dapivirine for HIV-1 prevention in women. *The New England Journal of Medicine* 375, 2121–2132.
- Brooks, K.M., Sherman, E.M., Egelund, E.F., *et al.*, 2019. Integrase inhibitors: After 10 years of experience, is the best yet to come? *Pharmacotherapy* 39, 576–598.

- Chen, Y., Bai, P., Mackay, S., *et al.*, 2011. Herpes simplex virus type 1 helicase-primase: DNA binding and consequent protein oligomerization and primase activation. *Journal of Virology* 85, 968–978.
- Chen, Y., Honeychurch, K.M., Yang, G., *et al.*, 2009. Vaccinia virus p37 interacts with host proteins associated with LE-derived transport vesicle biogenesis. *Virology Journal* 6, 44.
- Chou, S., Ercolani, R.J., Lanier, E.R., 2016. Novel cytomegalovirus UL54 DNA polymerase gene mutations selected in vitro that confer brincidofovir resistance. *Antimicrobial Agents and Chemotherapy* 60, 3845–3848.
- Coates, M., Brookes, D., Kim, Y.I., *et al.*, 2017. Preclinical characterization of PC786, an inhaled small-molecule respiratory syncytial virus L protein polymerase inhibitor. *Antimicrobial Agents and Chemotherapy* 61, e00737.
- De Clercq, E., 2018. Role of tenofovir alafenamide (TAF) in the treatment and prophylaxis of HIV and HBV infections. *Biochemical Pharmacology* 153, 2–11.
- De Clercq, E., Li, G., 2016. Approved antiviral drugs over the past 50 years. *Clinical Microbiology Reviews* 29, 695–747.
- Das, K., Arnold, E., 2013. HIV-1 reverse transcriptase and antiviral drug resistance. Part 1. *Current Opinion in Virology* 3, 111–118.
- Das, K., Martinez, S.E., Destefano, J.J., Arnold, E., 2019. Structure of HIV-1 RT/dsRNA initiation complex prior to nucleotide incorporation. *Proceedings of the National Academy of Sciences of the United States of America* 116, 7308–7313.
- Deeks, E.D., 2014. Darunavir: A review of its use in the management of HIV-1 infection. *Drugs* 74, 99–125.
- Deeks, E.D., 2018. Doravirine: First Global Approval. *Drugs* 78, 1643–1650.
- Delaune, D., Iseni, F., 2020. Drug development against smallpox: Present and future. *Antimicrobial Agents and Chemotherapy* 64, e01683.
- Delélis, O., Carayon, K., Saib, A., Deprez, E., Mouscadet, J.F., 2008. Integrase and integration: Biochemical activities of HIV-1 integrase. *Retrovirology* 5, 114.
- Devincenzo, J.P., McClure, M.W., Symons, J.A., *et al.*, 2015. Activity of oral ALS-008176 in a respiratory syncytial virus challenge study. *The New England Journal of Medicine* 373, 2048–2058.
- Eastman, R.T., Roth, J.S., Brimacombe, K.R., *et al.*, 2020. Remdesivir: A review of its discovery and development leading to emergency use authorization for treatment of COVID-19. *ACS Central Science* 6, 672–683.
- Esmat, G., Elbaz, T., El Raziky, M., *et al.*, 2017. Effectiveness of ravidasvir plus sofosbuvir in interferon-naïve and treated patients with chronic hepatitis C genotype-4. *Journal of Hepatology* 68, 53–62.
- Esposito, D., Craigie, R., 1999. HIV integrase structure and function. *Advances in Virus Research* 52, 319–333.
- Falkenhagen, A., Joshi, S., 2018. HIV entry and its inhibition by bifunctional antiviral proteins. *Molecular Therapy - Nucleic Acids* 13, 347–364.
- Fearn, R., Deval, J., 2016. New antiviral approaches for respiratory syncytial virus and other mononegaviruses: Inhibiting the RNA polymerase. *Antiviral Research* 134, 63–76.
- Fenaux, M., Eng, S., Leavitt, S.A., *et al.*, 2013. Preclinical characterization of GS-9669, a thumb site II inhibitor of the hepatitis C virus NS5B polymerase. *Antimicrobial Agents and Chemotherapy* 57, 804–810.
- Gao, Y., Yan, L., Huang, Y., *et al.*, 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368, 779–782.
- Gentles, R.G., Ding, M., Bender, J.A., *et al.*, 2014. Discovery and preclinical characterization of the cyclopropylindolobenzazepine BMS-791325, a potent allosteric inhibitor of the hepatitis C virus NS5B polymerase. *Journal of Medicinal Chemistry* 57, 1855–1879.
- Gilman, M.S.A., Furmanova-Hollenstein, P., Pascual, G., *et al.*, 2019. Transient opening of trimeric prefusion RSV F proteins. *Nature Communications* 10, 2105.
- Gu, R.X., Liu, L.A., Wei, D.Q., 2013. Structural and energetic analysis of drug inhibition of the influenza A M2 proton channel. *Trends in Pharmacological Sciences* 34, 571–580.
- Gubareva, L., Mohan, T., 2020. Antivirals targeting the neuraminidase. *Cold Spring Harbor Perspectives in Medicine*. a038455.
- Gulick, R.M., Mellors, J.W., Havlir, D., *et al.*, 1997. Treatment with zidovudine, zalcitabine, and didanosine in adults with human immunodeficiency virus infection and prior antiretroviral therapy. *The New England Journal of Medicine* 337, 734–739.
- Hammer, S.M., Squires, K.E., Hughes, M.D., *et al.*, 1997. A controlled trial of two nucleoside analogues plus zidovudine in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *The New England Journal of Medicine* 337, 725–733.
- Hillen, H.S., Kokic, G., Farnung, L., *et al.*, 2020. Structure of replicating SARS-CoV-2 polymerase. *Nature* 584, 154–156.
- Hou, J., Zhang, Z., Huang, Q., *et al.*, 2017. Antiviral activity of PHA767491 against human herpes simplex virus in vitro and in vivo. *BMC Infectious Diseases* 17, 217.
- Huang, Q., Hou, J., Yang, P., *et al.*, 2019. Antiviral activity of mitoxantrone dihydrochloride against human herpes simplex virus mediated by suppression of the viral immediate early genes. *BMC Microbiology* 19, 274.
- Jiang, C., Wang, Y., Hu, M., *et al.*, 2020. Antibody seroconversion in asymptomatic and symptomatic patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Clinical & Translational Immunology* 9, e1182.
- Jordan, R., Leeds, J.M., Tyavanagimatt, S., Hruby, D.E., 2010. Development of ST-246(R) for treatment of poxvirus infections. *Viruses* 2, 2409–2435.
- Kati, W., Koev, G., Irvin, M., *et al.*, 2015. In vitro activity and resistance profile of dasabuvir, a nonnucleoside hepatitis C virus polymerase inhibitor. *Antimicrobial Agents and Chemotherapy* 59, 1505–1511.
- Kirby, B.J., Symonds, W.T., Kearney, B.P., Mathias, A.A., 2015. Pharmacokinetic, pharmacodynamic, and drug-interaction profile of the hepatitis C virus NS5B polymerase inhibitor sofosbuvir. *Clinical Pharmacokinetics* 54, 677–690.
- Kitchen, C.M., Nuno, M., Kitchen, S.G., Krogstad, P., 2008. Enfuvirtide antiretroviral therapy in HIV-1 infection. *Therapeutics and Clinical Risk Management* 4, 433–439.
- Ko, G.M., Reddy, A.S., Kumar, S., Bailey, B.A., Garg, R., 2010. Computational analysis of HIV-1 protease protein binding pockets. *Journal of Chemical Information and Modeling* 50, 1759–1771.
- Kumar, S., Goicoechea, S., Kumar, S., *et al.*, 2020. Oseltamivir analogs with potent anti-influenza virus activity. *Drug Discovery Today* 25, 1389–1402.
- Langley, D.R., Kimura, S.R., Sivaprakasam, P., *et al.*, 2015. Homology models of the HIV-1 attachment inhibitor BMS-626529 bound to gp120 suggest a unique mechanism of action. *Proteins* 83, 331–350.
- Lataillade, M., Zhou, N., Joshi, S.R., *et al.*, 2018. Viral drug resistance through 48 weeks, in a phase 2b, randomized, controlled trial of the HIV-1 attachment inhibitor prodrug, fostemsavir. *Journal of Acquired Immune Deficiency Syndromes* 77, 299–307.
- Li, G., De Clercq, E., 2016. HIV genome-wide protein associations: A review of 30 years of research. *Microbiology and Molecular Biology Reviews* 80, 679–731.
- Li, G., De Clercq, E., 2017. Current therapy for chronic hepatitis C: The role of direct-acting antivirals. *Antiviral Research* 142, 83–122.
- Li, G., De Clercq, E., 2020. Therapeutic options for the 2019 novel coronavirus (2019-nCoV). *Nature Reviews Drug Discovery* 19, 149–150.
- Li, Y., Xuan, S., Feng, Y., Yan, A., 2015. Targeting HIV-1 integrase with strand transfer inhibitors. *Drug Discovery Today* 20, 435–449.
- Ligat, G., Cazal, R., Hantz, S., Alain, S., 2018. The human cytomegalovirus terminase complex as an antiviral target: A close-up view. *FEMS Microbiology Reviews* 42, 137–145.
- Lu, H., Cherepanova, N.A., Gilmore, R., Contessa, J.N., Lehrman, M.A., 2019. Targeting STT3A-oligosaccharyltransferase with NGI-1 causes herpes simplex virus 1 dysfunction. *FASEB Journal: Official publication of the Federation of American Societies for Experimental Biology* 33, 6801–6812.
- Mandala, V.S., Loftis, A.R., Shcherbakov, A.A., Pentelute, B.L., Hong, M., 2020. Atomic structures of closed and open influenza B M2 proton channel reveal the conduction mechanism. *Nature Structural & Molecular Biology* 27, 160–167.
- Mao, Y., Wang, L., Gu, C., *et al.*, 2012. Subunit organization of the membrane-bound HIV-1 envelope glycoprotein trimer. *Nature Structural & Molecular Biology* 19, 893–899.
- Martinez, M.G., Villeret, F., Testoni, B., Zoulim, F., 2020. Can we cure hepatitis B virus with novel direct-acting antivirals? *Liver International* 40 (Suppl 1), 27–34.
- Marty, F.M., Ljungman, P., Chemaly, R.F., *et al.*, 2017. Letermovir prophylaxis for cytomegalovirus in hematopoietic-cell transplantation. *The New England Journal of Medicine* 377, 2433–2444.
- Marty, F., Winston, D., Rowley, S., *et al.*, 2013. CMX001 to prevent cytomegalovirus disease in hematopoietic-cell transplantation. *The New England Journal of Medicine* 369, 1227–1236.

- Mcauley, J.L., Gilbertson, B.P., Trifkovic, S., Brown, L.E., McKimm-Breschkin, J.L., 2019. Influenza virus neuraminidase structure and functions. *Frontiers in Microbiology* 10, 39.
- McGivern, D.R., Masaki, T., Lovell, W., *et al.*, 2015. Protease inhibitors block multiple functions of the NS3/4A protease-helicase during the hepatitis C virus life cycle. *Journal of Virology* 89, 5362–5370.
- Menendez-Arias, L., Alvarez, M., Pacheco, B., 2014. Nucleoside/nucleotide analog inhibitors of hepatitis B virus polymerase: mechanism of action and resistance. *Current Opinion in Virology* 8, 1–9.
- Miao, M., Jing, X., De Clercq, E., Li, G., 2020. Danoprevir for the treatment of hepatitis C virus infection: Design, development, and place in therapy. *Drug Design, Development and Therapy* 14, 2759–2774.
- Mostashari Rad, T., Saghaie, L., Fassihi, A., 2018. HIV-1 entry inhibitors: A review of experimental and computational studies. *Chemistry & Biodiversity* 15, e1800159.
- Nakata, H., Amano, M., Koh, Y., *et al.*, 2007. Activity against human immunodeficiency virus type 1, intracellular metabolism, and effects on human DNA polymerases of 4'-ethynyl-2-fluoro-2'-deoxyadenosine. *Antimicrobial Agents and Chemotherapy* 51, 2701–2708.
- Nel, A., Van Niekerk, N., Kapiga, S., *et al.*, 2016. Safety and efficacy of a dapivirine vaginal ring for HIV prevention in women. *The New England Journal of Medicine* 375, 2133–2143.
- Olalla, J., 2020. Remdesivir for the treatment of Covid-19 – Preliminary report. *The New England Journal of Medicine* 383, 993–994.
- Oliveira, M., Ibanescu, R.I., Anstett, K., *et al.*, 2018. Selective resistance profiles emerging in patient-derived clinical isolates with cabotegravir, bictegravir, dolutegravir, and elvitegravir. *Retrovirology* 15, 56.
- Orkin, C., Arasteh, K., Gorgolas Hernandez-Mora, M., *et al.*, 2020. Long-acting cabotegravir and rilpivirine after oral induction for HIV-1 infection. *The New England Journal of Medicine* 382, 1124–1135.
- Perron, M., Stray, K., Kinkade, A., *et al.*, 2015. GS-5806 inhibits a broad range of respiratory syncytial virus clinical isolates by blocking the virus-cell fusion process. *Antimicrobial Agents and Chemotherapy* 60, 1264–1273.
- Piret, J., Goyette, N., Boivin, G., 2017. Drug susceptibility and replicative capacity of multidrug-resistant recombinant human cytomegalovirus harboring mutations in UL56 and UL54 genes. *Antimicrobial Agents and Chemotherapy* 61, e01044.
- Podany, A.T., Scarsi, K.K., Fletcher, C.V., 2017. Comparative clinical pharmacokinetics and pharmacodynamics of HIV-1 integrase strand transfer inhibitors. *Clinical Pharmacokinetics* 56, 25–40.
- Poole, C.L., James, S.H., 2018. Antiviral therapies for herpesviruses: Current agents and new directions. *Clinical Therapeutics* 40, 1282–1298.
- Pott Junior, H., De Oliveira, M.F.B., Gamero, S., Amazonas, R.B., 2018. Randomized clinical trial of famciclovir or acyclovir for the treatment of herpes zoster in adults. *International Journal of Infectious Diseases* 72, 11–15.
- Pu, J., Wang, Q., Xu, W., Lu, L., Jiang, S., 2019. Development of protein- and peptide-based HIV entry inhibitors targeting gp120 or gp41. *Viruses* 11, 705.
- Reust, C.E., 2011. Common adverse effects of antiretroviral therapy for HIV disease. *American Family Physician* 83, 1443–1451.
- Rinis, N., Golden, J., Marceau, C., *et al.*, 2018. Editing N-glycan site occupancy with small-molecule oligosaccharyltransferase inhibitors. *Cell Chemical Biology* 25, 1231–1241.
- Ross-Thriepland, D., Harris, M., 2015. Hepatitis C virus NS5A: Enigmatic but still promiscuous 10 years on!. *Journal of General Virology* 96, 727–738.
- Schnell, J.R., Chou, J.J., 2008. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451, 591–595.
- Shaik, M.M., Peng, H., Lu, J., *et al.*, 2019. Structural basis of coreceptor recognition by HIV-1 envelope spike. *Nature* 565, 318–323.
- Shanmugam, S., Nichols, A.K., Saravanabalaji, D., Welsch, C., Yi, M., 2018. HCV NS5A dimer interface residues regulate HCV replication by controlling its self-interaction, hyperphosphorylation, subcellular localization and interaction with cyclophilin A. *PLOS Pathogens* 14, e1007177.
- Shook, B.C., Lin, K., 2017. Recent advances in developing antiviral therapies for respiratory syncytial virus. *Topics in Current Chemistry* 375, 40.
- Sluis-Cremer, N., Tachedjian, G., 2008. Mechanisms of inhibition of HIV replication by non-nucleoside reverse transcriptase inhibitors. *Virus Research* 134, 147–156.
- Soto, J.A., Galvez, N.M.S., Pacheco, G.A., Bueno, S.M., Kalergis, A.M., 2020. Antibody development for preventing the human respiratory syncytial virus pathology. *Molecular Medicine* 26, 35.
- Subbaiah, M.A.M., Meanwell, N.A., Kadow, J.F., 2017. Design strategies in the prodrugs of HIV-1 protease inhibitors to improve the pharmaceutical properties. *European Journal of Medicinal Chemistry* 139, 865–883.
- Takashita, E., 2020. Influenza polymerase inhibitors: Mechanisms of action and resistance. *Cold Spring Harbor Perspectives in Medicine*. a038687.
- Tang, A., Chen, Z., Cox, K.S., *et al.*, 2019. A potent broadly neutralizing human RSV antibody targets conserved site IV of the fusion glycoprotein. *Nature Communications* 10, 4153.
- Vashishtha, A.K., Kuchta, R.D., 2015. Polymerase and exonuclease activities in herpes simplex virus type 1 DNA polymerase are not highly coordinated. *Biochemistry* 54, 240–249.
- Voitenleitner, C., Crosby, R., Walker, J., *et al.*, 2013. In vitro characterization of GSK2485852, a novel hepatitis C virus polymerase inhibitor. *Antimicrobial Agents and Chemotherapy* 57, 5216–5224.
- Vollmer, K.J., Burns, A.C., Sauer, H.E., *et al.*, 2019. Activation of 6-alkoxy-substituted methylenecyclopropane nucleoside analogs requires enzymatic modification by adenosine deaminase-like protein 1. *Antimicrobial Agents and Chemotherapy* 63.
- Wandzik, J.M., Kouba, T., Cusack, S., 2020. Structure and function of influenza polymerase. *Cold Spring Harbor Perspectives in Medicine*. a038372.doi:10.1101/cshperspect.a038372.
- Wang, G., Deval, J., Hong, J., *et al.*, 2015. Discovery of 4'-chloromethyl-2'-deoxy-3',5'-di-O-isobutyryl-2'-fluorocytidine (ALS-8176), a first-in-class RSV polymerase inhibitor for treatment of human respiratory syncytial virus infection. *Journal of Medicinal Chemistry* 58, 1862–1878.
- Wang, W., Tang, J., Wei, F., 2020. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. *Journal of Medical Virology* 92, 441–447.
- Wen, F., Wan, X.F., 2019. Influenza neuraminidase: Underrated role in receptor binding. *Trends in Microbiology* 27, 477–479.
- Yan, V.C., Muller, F.L., 2020. Advantages of the parent nucleoside GS-441524 over remdesivir for Covid-19 treatment. *ACS Medicinal Chemistry Letters* 11, 1361–1366.
- Zarrouk, K., Piret, J., Boivin, G., 2017. Herpesvirus DNA polymerases: Structures, functions and inhibitors. *Virus Research* 234, 177–192.
- Zheng, X., Gao, L., Wang, L., *et al.*, 2019. Discovery of ziresovir as a potent, selective, and orally bioavailable respiratory syncytial virus fusion protein inhibitor. *Journal of Medicinal Chemistry* 62, 6003–6014.
- Zhou, Z., Zhang, M., Wang, Y., *et al.*, 2020. Clinical characteristics of older and younger patients infected with SARS-CoV-2. *Aging* 12, 11296–11305.

Antiretroviral Therapy – Nucleoside/Nucleotide and Non-Nucleoside Reverse Transcriptase Inhibitors

Timothy D Appleby and Killian J Quinn, King's College Hospital, London, United Kingdom

© 2021 Published by Elsevier Ltd.

Retroviruses and Reverse Transcription

Retroviruses are a group of related viruses that depend on reverse transcription for their replication. They are subdivided into beta-, spuma-, gamma-, epsilon-, delta- and the lenti-retroviruses, to which HIV-1 and HIV-2 belong. The term retrovirus derives from the process of converting viral RNA to proviral DNA by use of the endogenous retroviral enzyme, reverse transcriptase (RT).

Transcription is the process of copying DNA code into RNA – thus reverse transcription refers to copying an RNA code into DNA. Both DNA and RNA are composed of long chains of polysaccharide residues held together with phosphates, and each polysaccharide having a base attached to it. The two ends of a DNA and RNA chain are referred to the 5' end and the 3' ends, denoting the position at which the carbon chains link the polysaccharide, i.e., 5th position on their carbon chain, and at the other end the 3rd position.

The genetic information of retroviruses is stored within two identical copies of positive sense RNA strands which exist as a dimeric molecule within the viral capsid. The genetic information is stored within nine reading frames, between 7 and 10 kilobases in length (9.7 kilobases in the case of HIV-1). Genes can be broadly divided into three main functional groups: (1) structural (2) regulatory and (3) accessory genes. Three genes (*gag*, *pol* and *env*) encode the structural polyproteins which undergo proteolytic cleavage to give rise to a total of 15 proteins involved in the viral lifecycle and structure. The *pol* gene encodes the three enzymes, RT, protease and integrase that are essential for the virus lifecycle and which are not available via the host machinery.

The RT enzyme enables production of a complementary strand of viral DNA from single-stranded RNA (ssRNA). It is this process of reverse transcription that enables the integration of the resulting proviral DNA into the host genome and in turn effecting the persistence of the retroviral genome within the host DNA.

Owing to its clinical importance, HIV-1 is by far the most studied retrovirus. Although there are multiple targets within the HIV-1 lifecycle, antiretroviral therapy targets 4 main steps namely: (1) viral entry (2) reverse transcription of viral RNA to double-stranded proviral DNA (dsDNA) (3) integration of dsDNA into the host genome and (4) proteolytic cleavage of viral polypeptides by the enzyme protease, enabling production of new viral proteins and ultimately new virus particles.

Reverse Transcriptase Structure and Functions

RT is an asymmetric heterodimer of two related but distinct subunits, p66 and p51. It has two catalytic domains: the DNA polymerase active site and RNase H both of which are necessary to carry out reverse transcription. These are a DNA polymerase that can copy either an RNA or a DNA template, and an RNase H that degrades RNA if it is part of an RNA–DNA duplex. Viral RT harnesses host nucleotides and initiates the process of reverse transcription of RNA into RNA–DNA duplex and subsequently double stranded DNA (Ren *et al.*, 1995).

With the activity of the RNase H active site, the RNA component of the RNA–DNA double helix is digested, while the polymerase active site can then complete the synthesis of the double-stranded DNA to create a double-stranded helix which can then be integrated as proviral DNA into the host genome. RT was the first enzyme target in the development of ARV drugs and is the target for two distinct classes of ARV drug: the NRTIs and NNRTIs.

The Process of Reverse Transcription

The HIV-1 life cycle begins with an interaction between the gp120 polypeptide on the viral surface and the CD4 subunits on the target cell membrane. This interaction leads to conformational change within gp120, in turn leading to release of gp41 which unfolds in a spring-loaded like mechanism piercing the target cell membrane. The gp41 subunit itself comprises two subdomains (heptad repeat, HR-1 and HR-2) which fold backwards upon one another. This mechanism results in the HIV-1 virion being pulled into close apposition to the target cell, a mixing of the viral and cellular lipid bilayers ensues with the creation of a “fusion pore” and the viral capsid is emptied inside the target cell.

Following cytoplasmic penetration of the viral core, partial capsid dissolution and digestion of the matrix proteins releases viral RNA coupled with nucleoprotein into the host cells (“uncoating”) (Ren *et al.*, 1995). It has been estimated that there are perhaps 50 RTs within each viral capsid which are released.

A tRNA primer binds to the primer binding site on the HIV-1 RNA

Genomic viral RNA is plus (+) sense. Reverse transcription of (+) sense RNA into an RNA–DNA double helix is mediated by the RT polymerase active site. The process begins as the cellular tRNA₃^{Lys} molecule, an 18-nucleotide primer intrinsic to the host cell, hybridizes to a complementary region of the (+)RNA genome forming the primer binding site (PBS). The cellular tRNA₃^{Lys} primer is (–)sense whose 3' end is based-paired with the 5' prime end of the complementary (+)sense RNA. Although various

Table 1 NRTIs, with corresponding nucleosides, and year of licensing

<i>Name</i>	<i>Abbreviation</i>	<i>Nucleoside</i>	<i>Year licensed</i>
Abacavir	ABC	Guanosine	1998
Didanosine	ddI	Adenosine	1991
Emtricitabine	FTC	Cytidine	2003
Lamivudine	3TC	Cytidine	1995
Stavudine	d4T	Thymidine	1994
Tenofovir-DF ^a	TDF	Adenosine	2001
Tenofovir-AF ^b	TAF	Adenosine	2014
Zalcitabine	ddC	Pyrimidine	1992
Zidovudine	AZT	Thymidine	1987

^atenofovir disoproxil fumarate.^btenofovir alafenamide.

primer models can be used to initiate reverse transcription in vitro, tRNA is used by all retroviruses. Host cellular tRNA₃^{Lys} is exclusively used by HIV-1 (Hu and Hughes, 2012; Hughes, 2015).

Reverse transcriptase (RT) starts at this binding site and copies RNA into a single strand of complementary DNA

As the tRNA₃^{Lys} molecule anneals its 3' end to a complementary 5' end of the (+) sense RNA, the RT polymerase active site binds to the viral RNA-tRNA complex at the PBS. Reverse transcription is activated by a signaling motif within the U5 region (non-coding region) of the HIV-1 genome, or primer activation signal. In doing this, RT navigates the 5' end, copying the (+) RNA into a complementary strand of viral DNA, or cDNA and termed (–)strand strong-stop DNA or (–)ssDNA. To do this, viral RT harnesses host cell nucleotides and the process of reverse transcription of RNA is activated: DNA nucleotides are added onto the 3' end of the primer, synthesizing DNA complementary to the U5 and R region (a direct repeat found at both ends of the RNA molecule) of the viral RNA. At this point it only copies from the primer binding site back into the Long Terminal Repeat (LTR), so all that has been copied so far is the LTR plus a little extra (Hu and Hughes, 2012; Hughes, 2015) see “Relevant Websites section”.

In vitro experiments indicate that the addition of the first few nucleotides is a slow process which speeds up after around 10 nucleotides are added. As the (–)ssDNA is released, it anneals to the 3' terminus of the vRNA, and primes further (–) strand DNA synthesis and generates a full-length (–)strand DNA that is used as a template for (+)strand DNA synthesis.

RNAse H degrades the section of the RNA which has been copied

DNA synthesis initially creates an RNA-DNA duplex which is substrate for the RNAse H active site, resulting in simultaneous digestion of the copied ssRNA as polymerization progresses. RNAse H degradation removes the 5' end of the viral RNA thus exposing the newly formed (–)strand DNA. It is not known if the same RT molecule which synthesizes the viral DNA is responsible for digestion of the RNA template.

This allows the tRNA/RT/ssDNA to dissociate from the HIV-1 RNA, and then reattach at the *other* end of the stretch of RNA, a process known as (–) strand transfer. The ends of the (+) sense viral RNA are direct repeats which allow transfer of the newly synthesized DNA to the 3' end of the RNA whereupon (–) strand synthesis of viral single stranded DNA can continue.

As DNA synthesis continues, so too does digestion of the viral RNA by the RNAse H enzymatic domain of RT. As the RNAse H reaches a region within the viral RNA which is purine rich, digestion is halted. As a result, this portion of (+) sense viral RNA persists and acts as a primer to begin the synthesis of the (+) sense viral DNA. Synthesis continues until RT encounters the tRNA₃^{Lys} primer which was initially incorporated and then digests this via RNAse H activity of RT.

As the two ends of the DNA are complementary and easily stick together, the DNA can then circularize. RT completes the second strand of the DNA including the Long Terminal Repeat at each end. In the process of this the DNA loop breaks, leaving a double-stranded DNA fragment with a Long Terminal Repeat at each end thus leaving linear viral DNA (Hu and Hughes, 2012; Hughes, 2015).

Nucleoside/nucleotide Reverse-Transcriptase Inhibitors (NRTIs)

NRTIs and Their Mechanism of Action

NRTIs were the first active medicinal products used against HIV-1. The first to be licensed was 3'-azido-3'-deoxythymidine (AZT) in 1987, subsequently known as zidovudine (Cihlar and Ray, 2010). Since then, eight further pharmacologically useful NRTIs (abacavir, didanosine, emtricitabine, lamivudine, stavudine, tenofovir-DF*, tenofovir-AF**, zalcitabine) have been developed for the treatment of HIV-1 (De Clercq, 2009) (Table 1).

NRTIs are pro-drugs that need to be phosphorylated intracellularly to become the active NRTI-triphosphate nucleoside analogs, a process catalyzed by cellular kinases (Anderson *et al.*, 2004; Furman *et al.*, 1986). They are structurally similar to the host nucleoside bases normally utilized by the RT active site (DNA polymerase) in the transcription of the ssRNA of the invading HIV-1 genome into dsDNA. The analogs bind competitively to proviral DNA, displacing the host nucleosides from DNA polymerase. However, the analogs lack the 3'-hydroxyl group: as a result, the native nucleosides (adenosine, thymidine, pyrimidine, guanosine) cannot bind to the dsDNA

Table 2 summary of NRTI backbone recommendations in national and international adult HIV-1 treatment guidelines

	<i>Backbone</i>	<i>Alternative</i>
BHIVA (British HIV Association, 2016)	TDF & FTC TDF & FTC	ABC & 3TC
WHO (World Health Organization, 2018)	TDF & 3TC TDF & FTC	AZT & 3TC
EACS (European AIDS Clinical Society, 2019)	ABC & 3TC TAF & FTC TDF & FTC TDF & 3TC 3TC ^a	

^awith dolutegravir, if VL < 500,000 and CD4 > 200.

thus terminating the strand elongation, and thereby the process of reverse transcription. This is known as immediate chain termination ([Sluis-Cremer et al., 2000](#)).

Tenofovir-DF (TDF) and tenofovir-AF (TAF) are nucleoside analogs that are phosphorylated at the 5' site, and are more commonly known as nucleotide analogs. This is an artificial distinction, as tenofovir acts pharmacologically by displacing host nucleosides, as with other NRTIs.

Clinical Uses of NRTIs

NRTIs were the first class of antiretroviral drugs to be used in the treatment of HIV. Although first used alone as single agents (monotherapy) with modest effect, treatment failure as a result of emergent drug resistance occurred within 3–4 months of initiation ([Larder et al., 1989](#)). Further trials examined combinations of two NRTIs used concurrently, which showed improved outcomes within treatment-naïve patients with virological control sustained for longer periods and improved clinical outcomes. Trials examined two approaches:

- (1) The addition of a second NRTI to NRTI monotherapy where treatment failure had already occurred.
- (2) Head-to-head comparisons of NRTI monotherapy and dual therapy in treatment-naïve patients ([Schooley et al., 1996](#); [Hammer et al., 1996](#); [Darbyshire and Aboulker, 1996](#)).

The improvement in outcomes with dual NRTI therapy was again short-lived, with resistance developing in the dual combinations. In 1996, triple-therapy, or highly active antiretroviral therapy (HAART), was introduced leading to great improvements in viral suppression, with higher barriers to resistance ([Lange and Ananworanich, 2014](#)).

NRTIs are now primarily used in dual combinations as the “backbone” of antiretroviral therapy for the treatment of HIV-1, with the addition of a third agent, i.e., an NNRTI, a boosted protease inhibitor or an integrase inhibitor. National and international guidelines on the treatment of HIV-1 differ slightly in their preferred backbone and third agent (**Table 2**). However, dual NRTI is the preferred backbone in all British HIV Association (BHIVA) and World Health Organization (WHO) recommendations ([British HIV Association, 2016](#); [World Health Organization, 2018](#)). The European AIDS Clinical Society (EACS) guidelines on the treatment of HIV-1 recommend dual NRTI as the backbone in all but one of their first line recommendations; EACS also recommends that a single NRTI (lamivudine) can be used with dolutegravir as first line treatment for HIV-1 ([European AIDS Clinical Society, 2019](#)).

The most used NRTI combinations are available as fixed dose combinations, in single tablets. Tenofovir-DF and emtricitabine as Truvada, and abacavir and lamivudine as Kivexa or Epzicom. Both of these combinations are now available off patent in generic formulations. More recently, tenofovir alafenamide and emtricitabine have been combined into a fixed dose combination. NRTIs are also included in a number of other single table regimens, containing other classes of drugs.

NRTI Toxicities

NRTIs are pro-drugs that are phosphorylated intracellularly into their active form. Some of the long-term toxicities with NRTIs are attributable to overactivation of this intracellular phosphorylation. NRTI-triphosphates may inhibit mitochondrial DNA polymerase γ , leading to mitochondrial failure, and a pattern of toxicities resembling mitochondrial disease, especially in the older and now, less used drugs. These include lactic acidosis/hyperlactaemia and hepatic steatosis (didanosine, stavudine, zidovudine), pancreatitis (didanosine), peripheral neuropathy (stavudine, didanosine, zalcitabine), lipodystrophy (stavudine, all NRTIs) and myopathy (zidovudine) ([Anderson et al., 2004](#); [Lewis et al., 2001](#); [Carr and Cooper, 2000](#)).

Zidovudine causes anemia and bone marrow suppression in about 5%–10% of patients. This is more common in patients with advanced HIV, and is also thought also to be as a result of mitochondrial toxicity ([Carr and Cooper, 2000](#)).

Hypersensitivity reactions occur in approximately 5% of patients taking abacavir. These reactions usually happen within the first 6 weeks of commencing treatment, but can also develop after many months. Typically, hypersensitivity reactions affect the skin and

cause fever and rash, but can affect multiple organs, including the gastrointestinal and respiratory systems. Severe reactions can sometimes develop, which can be fatal (Dean, 2012). People who have the HLA-B*5701 allele are more susceptible to abacavir hypersensitivity, and in practice all patients are tested for the allele before treatment is started. If the patient tests positive for the HLA-B*5701 allele, they should not be started on abacavir. However, hypersensitivity reactions may still occur even in the absence of the HLA-B*5701 allele. If HLA-B*5701 testing is not available, abacavir could still be starting with caution and close monitoring (Dean, 2012; Mallal *et al.*, 2008).

TDF is usually tolerated well by patients, but can cause acute and chronic renal failure, interstitial nephritis, proximal tubular dysfunction, and rarely Fanconi's syndrome. Stopping TDF due to renal complications is thought to account for fewer than 1% of discontinuations, but close monitoring of renal toxicity is suggested by most HIV treatment guidelines, especially in those most at risk of developing renal disease. These include older patients, including those with other risk factors for renal impairment, such as uncontrolled hypertension, diabetes mellitus, pre-existing low creatinine clearance and those taking nephrotoxic medications (World Health Organization, 2018).

TDF has also been shown to cause a loss of bone mineral density (BMD) and an increased risk of fractures. Studies show 1%–3% greater loss in BMD with tenofovir-DF compared with TDF sparing regimens. The mechanism by which this occurs is not absolutely clear, but it may be linked to TDF's effect on the proximal renal tubule. TDF may cause a sub-clinical tubulopathy and phosphate wasting, which may in turn drive reductions in BMD (Grant and Cotter, 2016). Tenofovir may also have a direct action on bone turnover, by acting directly on the osteoclasts and osteoblasts (Grigsby *et al.*, 2010).

Tenofovir alafenamide (TAF) is a newer formulation of tenofovir which was licensed in 2016. Plasma concentrations of tenofovir are significantly lower after oral administration of TAF, when compared to TDF. This means decreased exposure of the renal proximal tubule to the active drug, thereby decreasing the toxic effects on the kidney and bone. TAF is now recommended as a first-line treatment in BHIVA and EACS guidelines, and is an alternative to TDF for patients at risk of renal impairment patient at risk of low BMD.

Some cohort studies have shown an increased risk of cardiovascular disease with use of abacavir, including events such as acute myocardial infarction within the first 12 months of use (but not stroke) (Sabin *et al.*, 2008). This seems to be especially true in patients with pre-existing history of cardiac disease. However, this increased risk is not seen in all studies, and the link between abacavir and cardiovascular risk remains a topic of debate with concerns that such studies are confounded by selection bias of patients with a pre-existing increased cardiovascular risk (Libre and Hill, 2016).

See Table 3 for a summary of common NRTI toxicities.

Drug-Drug Interactions With NRTIs

As NRTIs are not metabolized by the CYP450 enzyme system, there are fewer potentially significant drug-drug interactions compared with other anti-retroviral drug classes (Pau and George, 2014). TAF is a substrate of P-glycoprotein (*p-gp*), and therefore will interact with P-gp inducers such as rifampicin, resulting in a much lower concentration of TAF, potentially leading to treatment failure (British HIV Association, 2016).

Mechanisms of NRTI Resistance

The mechanisms of viral resistance to NRTIs are complex. Mutations in the *pol* gene that codes for viral RT enzyme occurs under the selective pressure of NRTI drugs. One mechanism is known as discrimination, where the mutated RT loses its ability to bind to specific NRTIs, but keeps its affinity for the host nucleoside bases. This allows reverse transcription to continue without the NRTIs

Table 3 Common NRTI toxicities, taken from European AIDS Clinical Society Guidelines, Version 10.0

	<i>Skin</i>	<i>GI</i>	<i>Liver</i>	<i>Cardiac</i>	<i>MSK</i>	<i>Renal</i>	<i>Metabolic/lipids</i>	<i>Other</i>
ABC	Rash	Nausea Diarrhea		IHD				Systemic hypersensitivity (HLA B*5701 dependent)
AZT	Nail pigmen-tation	Nausea	Hepatic steatosis		Myopathy Rhabdo-myolysis		Lipodystrophy Dyslipidemia Hyperlactaemia	Anemia
3TC								
FTC								
TDF			Hepatitis		↓ BMD Osteo-malacia ↑ fracture risk	↓ eGFR Fanconi syndrome	↓ plasma lipids	
TAF							↑ weight	

Note: European AIDS Clinical Society, 2019. EACS Guideline 10.0. [cited 2020 Jun 7]. Available from: <http://www.eacsociety.org>.

causing chain termination. These are usually point mutations that confer resistance to specific NRTIs, e.g., M184V, K65R, L74V and Q151M (Marcelin, 2006).

Some mutations enhance the ability of RT to remove NRTIs that have already bound to the 3' end of the expanding pro-viral DNA. This is known as primer unblocking, and is enhanced by specific mutations known as TAMs (thymidine analog mutations), examples of which are M41L, D67N, K70R, L210W, T215Y/F and K219Q/E. These mutations are selected by zidovudine and stavudine, but they can cause cross-resistance to most NRTIs. The extent of cross-resistance depends on the number of TAMs. For example, the combination of M41, L210W and T215Y causes high-level resistance to zidovudine and stavudine, and intermediate to high-level resistance to didanosine, abacavir and tenofovir (Marcelin, 2006).

New Drugs

Islatravir is a novel agent, known as a nucleoside reverse transcriptase translocation inhibitor (NRTTI). It is an analog of deoxyadenosine and, like NRTIs, it is converted intracellularly to its active triphosphate form. Its main mode of action is through immediate chain termination. However, occasionally the proviral-DNA continues to be propagated despite incorporation of the active deoxyadenosine analog. In these cases, islatravir acts as a delayed chain terminator by causing structural changes in the proviral-DNA molecule which prevents nucleosides attaching further down the line. Islatravir has been shown to be active against HIV-1 and HIV-2, including virus with pre-existing NRTI-resistance. The active form of islatravir has a long intracellular half-life, and has the potential to be developed as a subdermal implant (Markowitz and Sarafianos, 2018).

Non-Nucleoside Reverse-Transcriptase Inhibitors (NNRTIs)

NNRTIs and Their Mechanism of Action

NNRTIs act by binding to a hydrophobic site that is remote from the active DNA polymerase site, known as the binding pocket, on the p66 subunit of the RT heterodimer. Binding causes a change in conformation of the RT heterodimer and its active enzyme site, thereby inhibiting DNA polymerase activity and blocking the process of reverse transcription of viral ssRNA to pro-viral dsDNA (Weller and Williams, 2001). NNRTI drugs are structurally diverse (Fig. 1) and, unlike NRTIs, are not analogs of nucleosides. Among the pharmacologically useful NNRTIs are efavirenz, nevirapine, etravirine, rilpivirine, and more recently doravirine.

Clinical Uses of NNRTIs

NNRTIs are used for the treatment of HIV-1 as part of combination antiretroviral therapy, and have demonstrated good efficacy in viral suppression. They are often used as the “third agent”, with the treatment “backbone” of two NRTI drug (see section above), as

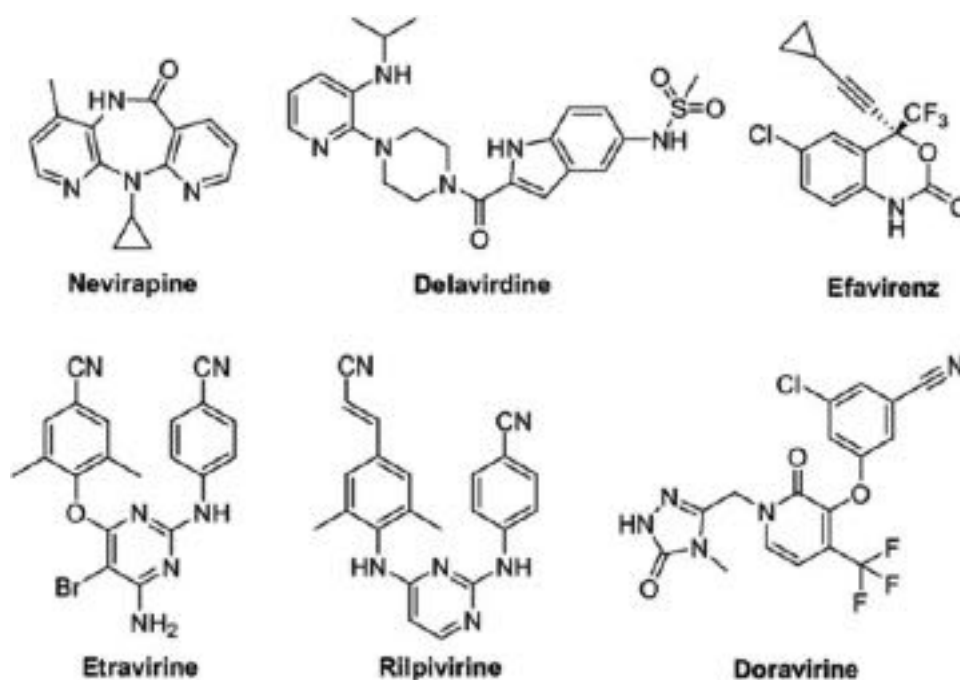


Fig. 1 Chemical structures of NNRTIs (Scientific Figure on ResearchGate).

an alternative to a boosted protease inhibitor or an integrase inhibitors. There is some variation in the recommendations for NNRTI use across international guidelines. The BHIVA guidelines for the treatment of HIV-1 suggest rilpivirine as a first line third agent, with efavirenz as an alternative ([British HIV Association, 2016](#)). The EACS guidelines suggest rilpivirine, or the newer NNRTI doravirine, as a suitable first line third agent with a dual NRTI backbone, with efavirenz as an alternative ([European AIDS Clinical Society, 2019](#)). The consolidated WHO guidelines suggest efavirenz as the first line NNRTI third agent, with nevirapine as an alternative ([World Health Organization, 2018](#)), and the American guidelines favor efavirenz and rilpivirine ([Department of Health and Human Services, 2019](#)). The HIV-2 virus is intrinsically resistant to all NNRTIs ([Gilleece *et al.*, 2010](#)).

NNRTI Toxicities

Older NNRTIs, and especially nevirapine, are associated with hepatotoxicity, including rarely, fatal drug induced liver injury. Although the risk with nevirapine hepatotoxicity is increased in female patients with CD4 cell counts > 250 cells/mm³, and male patients with CD4 cell counts > 400 cells/mm³, all patients require close monitoring of liver enzymes in the weeks after starting.

Nevirapine and other NNRTIs can also cause hypersensitivity reactions, with rash commonly occurring within a few weeks of starting treatment. Hypersensitivity rash can happen in approximately 16% of patients starting nevirapine, but is also common in patients starting efavirenz and etravirine. If hypersensitivity and deranged liver enzymes occur concurrently after starting nevirapine, the medication should be stopped immediately.

Central nervous system side effects and toxicities are common with efavirenz. These usually occur early on in treatment, in approximately 40% of patients, and can include sleep disturbance and vivid dreams, depression and suicidal ideation. Most of these side effects resolve with time, and are only severe enough to need discontinuation in about 3% of patients ([Carr and Cooper, 2000](#)). See [Table 4](#) for a summary of common NNRTI toxicities.

Doravirine is a newer NNRTI which has been shown to be as effective as a boosted protease inhibitor (darunavir/ritonavir) or efavirenz-based therapy, in combination with dual NRTI backbone. It also has fewer toxicities, including fewer central nervous system side effects when compared to efavirenz, and has fewer potential drug-drug interactions. It also appears to remain effective against HIV-1 with pre-existing NNRTI resistance mutations ([Colombier and Molina, 2018](#)).

Drug-Drug Interactions With NNRTIs

All NNRTIs are metabolized by the CYP3A liver enzymes. Nevirapine and efavirenz are also metabolized by CYP2B6, and etravirine by CYP2C9 and CYP2C19 liver enzymes. Drugs that induced these enzymes can lead to reduced levels of NNRTI, and virological failure. Conversely, inhibition of these enzymes leads to increased levels of NNRTIs and potentially more side effects and toxicity. NNRTIs also act as inducers and/or inhibitors of CYP enzymes. For example, efavirenz is a mixed inducer and inhibitor of CYP enzymes, mainly inducing CYP3A and CYP2B6.

The University of Liverpool provides a comprehensive and regularly updated resource on potential drug-drug interactions: See “Relevant Websites section”.

Mechanisms of NNRTI Resistance

A disadvantage of NNRTI drugs is their low barrier to developing resistance, and the prevalence of NNRTI-resistant HIV-1 virus in ART-naïve patients ([Snedecor *et al.*, 2013](#)). Mutations on the pol gene that codes for viral RT reduce the ability of NNRTIs to bind to the binding pocket on the RT p66 subunit, thereby reducing their activity. Just one mutation can mean complete resistance to a specific NNRTI. Single mutations can also cause resistance that extends across the more than one drug in the class. For example, a K103N mutation causes high-level resistance to both nevirapine and efavirenz ([HIV Drug Resistance Database, 2020](#)). For this reason, NNRTIs are not recommended second line, after failure of a first line NNRTI

Table 4 Common NNRTI toxicities, taken from European AIDS Clinical Society Guidelines, Version 10.0

	<i>Skin</i>	<i>Liver</i>	<i>Renal</i>	<i>CNS</i>	<i>Metabolic</i>
Doravirine					
Efavirenz	Rash	Hepatitis		Depression Sleep disturbance Headache Suicidal ideation	Dislipidemia Gynecomastia
Etravirine	Rash				
Nevirapine	Rash Hypersensitivity	Hepatitis			
Rilpivirine	Rash	Hepatitis	↓ eGFR		

Note: European AIDS Clinical Society, 2019. EACS Guideline 10.0. [cited 2020 Jun 7]. Available from: <http://www.eacsociety.org>.

containing regimen (British HIV Association, 2016). However, the newer NNRTI doravirine maintains activity against HIV-1 with K103N and Y181C mutations (Feng *et al.*, 2016).

References

- Anderson, P.L., Kakuda, T.N., Lichtenstein, K.A., 2004. The cellular pharmacology of nucleoside- and nucleotide-analogue reverse-transcriptase inhibitors and its relationship to clinical toxicities. *Clinical Infectious Diseases* 38 (5), 743–753. Available from: <https://academic.oup.com/cid/article-abstract/38/5/743/281843>.
- British HIV Association, 2016. BHIVA guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2015 (2016 interim update). Accessed online 14/12/2016 [Internet]. Vol. 2015. [cited 2020 May 27]. Available from: <http://www.bhiva.org/HIV-1-treatment-guidelines.aspx>.
- Carr, A., Cooper, D.A., 2000. Adverse effects of antiretroviral therapy. *Lancet* 356 (9239), 1423–1430.
- Cihlar, T., Ray, A.S., 2010. Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Research* 85 (1), 39–58. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0166354209004859>.
- Colombier, M.A., Molina, J.M., 2018. Doravirine: A review. *Current Opinion in HIV and AIDS* 13, 308–314. <https://pubmed.ncbi.nlm.nih.gov/29794817/>.
- Darbyshire, J.H., Aboulker, J.P., 1996. Delta: A randomised double-blind controlled trial comparing combinations of zidovudine plus didanosine or zalcitabine with zidovudine alone in HIV-infected individuals. *Lancet* 348 (9023), 283–291.
- De Clercq, E., 2009. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *International Journal of Antimicrobial Agents* 33, 307–320.
- Dean, L., 2012. Abacavir therapy and HLA-B*57:01 genotype. In: Pratt, V.M., McLeod, H.L., Rubinstein, W.S., *et al.* (Eds.), *Medical Genetics Summaries*. Bethesda, MD: NCBI. <http://www.ncbi.nlm.nih.gov/pubmed/28520363>.
- Department of Health and Human Services, 2019. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents with HIV. Department of Health and Human Services 298. [cited 2020 Aug 8] Available from: <https://aidsinfo.nih.gov/contentfiles/vguidelines/adultandadolescentgl.pdf>.
- European AIDS Clinical Society, 2019. EACS Guideline 10.0. [cited 2020 Jun 7]. Available from: <http://www.eacsociety.org>.
- Feng, M., Sachs, N.A., Xu, M., *et al.*, 2016. Doravirine suppresses common nonnucleoside reverse transcriptase inhibitor-associated mutants at clinically relevant concentrations. *Antimicrobial Agents and Chemotherapy* 60 (4), 2241–2247. [cited 2020 Aug 26]. Available from: <http://aac.asm.org/>.
- Furman, P.A., Fyfe, J.A., St Clair, M.H., *et al.*, 1986. Phosphorylation of 3'-azido-3'-deoxythymidine and selective interaction of the 5'-triphosphate with human immunodeficiency virus reverse transcriptase. *Proceedings of the National Academy of Sciences of the United States of America* 83 (21), 8333–8337.
- Gillece, Y., Chadwick, D.R., Breuer, J., *et al.*, 2010. British HIV Association guidelines for antiretroviral treatment of HIV-2-positive individuals 2010 Introduction and epidemiology. *HIV Medicine* 11, 611–619.
- Grant, P.M., Cotter, A.G., 2016. Tenofovir and bone health. *Current Opinion in HIV and AIDS* 11, 326–332. <https://pubmed.ncbi.nlm.nih.gov/2644450/>.
- Grigsby, I.F., Pham, L., Mansky, L.M., Gopalakrishnan, R., Mansky, K.C., 2010. Tenofovir-associated bone density loss. *Therapeutics and Clinical Risk Management* 6, 41–47. www.dovepress.com.
- Hammer, S.M., Katzenstein, D.A., Hughes, M.D., *et al.*, 1996. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *The New England Journal of Medicine* 335 (15), 1081–1090.
- HIV Drug Resistance Database, 2020. NNRTI Resistance Notes - HIV Drug Resistance Database. [cited 2020 Jun 7]. Available from: <https://hivdb.stanford.edu/dr-summary/resistance-notes/NNRTI/>.
- Hu, W.S., Hughes, S.H., 2012. HIV-1 reverse transcription. *Cold Spring Harbor Perspectives in Medicine* 2 (10). <https://pubmed.ncbi.nlm.nih.gov/23028129/>.
- Hughes, S.H., 2015. Reverse transcription of retroviruses and LTR retrotransposons. *Microbiology Spectrum* 3 (2), 1051–1077. <https://pubmed.ncbi.nlm.nih.gov/26104704/>.
- Lange, J.M., Ananworanich, J., 2014. The discovery and development of antiretroviral agents. *Antiviral Therapy* 19 (Suppl 3), 5–14.
- Larder, B.A., Darby, G., Richman, D.D., 1989. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. *Science* 243 (4899), 1731–1734.
- Lewis, W., Copeland, W.C., Day, B.J., 2001. Mitochondrial DNA depletion, oxidative stress, and mutation: mechanisms of dysfunction from nucleoside reverse transcriptase inhibitors. *Laboratory Investigation* 777–790.
- Libre, J.M., Hill, A., 2016. Abacavir and cardiovascular disease: A critical look at the data. *Antiviral Research* 132, 116–121. <https://pubmed.ncbi.nlm.nih.gov/27260856/>.
- Mallal, S., Phillips, E., Carosi, G., *et al.*, 2008. HLA-B*5701 Screening for Hypersensitivity to Abacavir. *The New England Journal of Medicine* 358 (6), 568–579. [cited 2020 Jun 20]. Available from: <http://www.nejm.org/doi/abs/10.1056/NEJMoa0706135>.
- Marcelin, A.-G., 2006. Chapter 1 – Resistance to nucleoside reverse transcriptase inhibitors. In: Geretti, A.M. (Ed.), *Antiretroviral Resistance in Clinical Practice*. Mediscript. <http://www.ncbi.nlm.nih.gov/pubmed/21249766>.
- Markowitz, M., Sarafianos, S.G., 2018. EFdA 4'-ethynyl-2-fluoro-2'-deoxyadenosine, MK-8591: A Novel HIV-1 Reverse Transcriptase Translocation Inhibitor. *Current opinion in HIV and AIDS* 13 (4), 294–299.
- Pau, A.K., George, J.M., 2014. Antiretroviral therapy: Current drugs. *Infectious Disease Clinics of North America* 28, 371–402. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4143801/>.
- Ren, J., Esnouf, R., Hopkins, A., *et al.*, 1995. The structure of HIV-1 reverse transcriptase complexed with 9-chloro-TIBO: Lessons for inhibitor design. *Structure* 3 (9), 915–926. [cited 2020 Oct 24]. Available from: <https://pubmed.ncbi.nlm.nih.gov/8535785/>.
- Sabin, C.A., Worm, S.W., Weber, R., *et al.*, 2008. Use of nucleoside reverse transcriptase inhibitors and risk of myocardial infarction in HIV-infected patients enrolled in the D:A:D study: A multi-cohort collaboration. *Lancet* 371 (9622), 1417–1426.
- Schooley, R.T., Ramirez-Ronda, C., Lange, J.M.A., *et al.*, 1996. Virologic and immunologic benefits of initial combination therapy with zidovudine and zalcitabine or didanosine compared with zidovudine monotherapy. *The Journal of Infectious Diseases* 173 (6), 1354–1366. <https://academic.oup.com/jid/article-abstract/173/6/1354/1010046>.
- Sluis-Cremer, N., Arion, D., Parniak, M.A., 2000. Molecular mechanisms of HIV-1 resistance to nucleoside reverse transcriptase inhibitors (NRTIs). *Cellular and Molecular Life Sciences* 57 (10), 1408–1422.
- Snedecor, S.J., Khachatryan, A., Nedrow, K., *et al.*, 2013. Prevalence of transmitted resistance to first-generation non-nucleoside reverse transcriptase inhibitors and its potential economic impact in HIV-infected patients. *PLoS One* 8 (8), e72784. Available at: <https://dx.plos.org/10.1371/journal.pone.0072784>.
- Weller, I.V.D., Williams, I.G., 2001. ABC of AIDS: Antiretroviral drugs. *British Medical Journal*. 1410–1412. [cited 2020 May 23]. Available from: www.bmjbooks.com.
- World Health Organisation, 2018. Consolidated Guidelines on HIV Prevention, Diagnosis, Treatment and Care for Key Populations. World Health Organisation. <http://www.who.int/hiv/pub/guidelines/keypopulations-2016/en/>.

Relevant Websites

<https://www.bhiva.org/guidelines>
Current Guidelines.
Bhiva.

<http://www.eacsociety.org>

EACSociety.

<https://hivdb.stanford.edu/>

HIV Drug Resistance Database.

Stanford University.

<https://www.hiv-druginteractions.org/checker>

Interaction Checker.

Liverpool HIV Interactions.

<http://www.mclld.co.uk/hiv/>.

The Molecules of HIV.

<https://www.who.int/health-topics/hiv-aids>

World Health Organization: HIV/AIDS.

Protease Inhibitors

Vanesa Anton-Vazquez, King's College Hospital, London, United Kingdom

Frank A Post, King's College Hospital NHS Foundation Trust, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Bioavailability The proportion of a drug which enters the circulation when introduced into the body and so is able to have an active effect.

Boosting Using an antiretroviral (ARV) drug or other drug to increase the effectiveness of another ARV drug.

CD4 T lymphocyte A type of T lymphocyte that helps to coordinate the immune response by stimulating other immune cells, including B lymphocytes, CD8 T lymphocytes and macrophages.

Cytochrome P450 A group of enzymes involved in the breakdown of drugs in the liver and play a key role in the metabolism of many drugs. Cytochrome P450 (CYP450) enzymes metabolize all protease inhibitors.

Drug–drug Interaction A reaction between two or more drugs or between a drug and a food or supplement.

A drug interaction can decrease or increase the action of the drug.

Protease A type of enzyme that breaks down proteins into smaller protein units. These smaller proteins combine with HIV's genetic material to form a new HIV virus. Protease inhibitors (PIs) prevent HIV from replicating by blocking protease.

Resistance When a virus, bacteria or fungus becomes non susceptible to a drug that was previously effective.

Viral load Number of HIV RNA copies per milliliter of blood.

Viral suppression When antiretroviral therapy (ART) reduces a person's viral load (HIV RNA) to an undetectable level.

Protease Inhibitors

Introduction

The introduction of HIV protease inhibitors into clinical practice in 1995 changed the course of HIV infection, with an increase in survival in people living with HIV/AIDS (PLWHA) (Palella *et al.*, 1998). Combination therapy consisting of two nucleoside analogs and a protease inhibitor reduced HIV viral load and increased CD4 cell counts rapidly for the very first time, stopping disease progression (Ho *et al.*, 1995), and became a cornerstone of antiretroviral treatment.

Ritonavir- or cobicistat-boosted protease inhibitors represent an important antiretroviral class due to their low potential to select resistance. However, they also introduced a variety of new challenges such as undesirable side effects, dosing and absorption difficulties and interaction with concomitant drugs. These characteristics should be fully considered and understood when using protease inhibitors.

Mechanism of Action

Protease inhibitors (PIs) are one class of antiretroviral drug that target the viral enzyme, HIV-1 protease (Ali *et al.*, 2010). Protease is an essential element for viral maturation in the HIV life cycle. PIs act by binding to the catalytic site of the HIV protease inhibiting cleavage of Pol polyprotein viral precursors into mature proteins that are essential for viral replication. As a result, non-infectious immature virions are released from the cell surface (Konvalinka *et al.*, 2015).

PIs are active against HIV-1 and HIV-2. To date, ten PIs have been approved by the FDA: ritonavir, followed by indinavir, nelfinavir, saquinavir, amprenavir, lopinavir/ritonavir, atazanavir, fosamprenavir, tipranavir, and darunavir (Table 1).

Pharmacodynamics and Pharmacokinetics in PIs

Protease inhibitor concentrations in blood plasma must be high enough to inhibit HIV viral replication. The key is to maintain drug concentrations above the minimum effective concentration to ensure inhibition of viral replication and prevent emergence of resistance, and below the maximum safe concentration to avoid toxicity.

PIs have poor oral bioavailability. This is mainly due to their susceptibility to oxidation by cytochrome P450 3A4 (CYP3A4) enzymes in the enterocytes and hepatocytes as well as the presence of P-glycoprotein (*p*-gp) efflux transporters that decreases gut absorption. In addition, there is substantial variability in the concentration of plasma PI levels due to inter-individual differences in CYP3A4 (Von Richter *et al.*, 2004).

To overcome the reduced bioavailability of PIs, co-administration of potent CYP3A4 and P-gp inhibitors, known as boosting, increases PI plasma exposure thereby allowing once or twice daily dosing, resulting in enhanced anti-viral efficacy and reduced

Table 1 Classes of protease inhibitors (PIs)

<i>HIV protease inhibitor</i>	<i>PI drug</i>	<i>FDA approval</i>
Ritonavir	RTV	1996
Indinavir	IDV	1996
Nelfinavir	NFV	1997
Saquinavir	SQV	1997
Amprenavir	APV	1999
Lopinavir/ritonavir	LPV/r	2000
Atazanavir	ATV	2003
Fosamprenavir	FPV	2003
Tipranavir	TPV	2005
Darunavir	DRV	2006

toxicity (Larson *et al.*, 2014). Ritonavir and cobicistat are the two main booster agents available in clinical practice (Boffito and Moyle, 2004).

Ritonavir boosts intestinal absorption of the co-administered PI by inhibiting a number of CYP450 enzymes (2B6, 2C8, 2C9, 2D6 and 3A4) and the intestinal efflux transporter P-gp. In contrast, cobicistat inhibits CYP2B6 and 3A4 (with a lesser effect on 2C8 and 2D6 and no effect on 1A2, 2C9 and 2C19) and the efflux transporter P-gp (Nathan *et al.*, 2013). As a boosting agent, ritonavir is dosed as 100 or 200 mg per day (well below the minimum effective concentration for inhibition of viral replication).

Cobicistat is chemically similar to ritonavir, but has no antiviral activity. At a dose of 150 mg per day, it is used to boost atazanavir and darunavir and also the integrase inhibitor elvitegravir.

PIs in Combination With Other ART Classes

PIs are typically administered in combination with one or two nucleoside reverse transcriptase inhibitors (NRTI). In ART naïve and ART experienced HIV patients, ritonavir-boosted darunavir or ritonavir-boosted atazanavir in combination with an NRTI backbone are the most widely used PIs. Boosted darunavir has been shown to be better tolerated with fewer side effects (Molina *et al.*, 2008; Mills *et al.*, 2009) and a lower incidence of treatment discontinuation when compared with boosted atazanavir (Lennox *et al.*, 2014).

Protease Inhibitor-Based Regimens

Triple therapy: Boosted PI + 2 NRTI

Darunavir/ritonavir (or cobicistat) OR Atazanavir/ritonavir (or cobicistat)	Tenofovir (TDF or TAF)/emtricitabine
Darunavir/ritonavir (or cobicistat) OR Atazanavir/ritonavir (or cobicistat)	Abacavir/lamivudine

Dual therapy: boosted PI + 1 NRTI or 1 INSTI

If ABC, TAF or TDF cannot be used, dual therapy could be considered with one NRTI (lamivudine) (Cahn *et al.*, 2014) or, in those with CD4 cell counts >200 cells/mm³, one integrase-inhibitor (raltegravir) as an alternative to triple therapy for selected patients (Raffi *et al.*, 2014).

Darunavir/ritonavir	Raltegravir
Lopinavir/ritonavir	Lamivudine

Darunavir

Darunavir is available in two single co-formulated tablets, darunavir-cobicistat and darunavir-cobicistat-emtricitabine-tenofovir alafenamide.

Darunavir in combination with ritonavir or cobicistat, should not be used in patients with severe liver disease and should be avoided in individuals with sulfonamide allergy.

In individuals with wild-type virus, the recommended dose is darunavir 800 mg/ritonavir 100 mg or cobicistat 150 mg once daily (Madruca *et al.*, 2007; Clotet *et al.*, 2007). The dose of darunavir 600 mg/ritonavir 100 mg twice daily is recommended if the following resistance mutations are present: V11I, V32I, L33F, I47V, I50V, I54L/M, T74P, L76V, I84V, and L89V. Darunavir often

retains activity against HIV virus with protease resistance-associated mutations (typically the result of prior exposure to unboosted PIs) and is the PI of choice for this group of individuals.

Atazanavir

Atazanavir is available in 300 mg capsules and as co-formulated atazanavir 300 mg/cobicistat 150 mg. Atazanavir has better gastrointestinal tolerance than other PIs. It requires an acid pH for optimal drug absorption.

Atazanavir has limited activity against HIV virus with protease resistance-associated mutations. Although more “lipid friendly” than lopinavir/ritonavir ([Monforte et al., 2013](#); [Squires et al., 2010](#)), ATV/r and DRV/r have similar effects on lipids, glucose, insulin resistance ([Ofotokun et al., 2015](#)).

Lopinavir

Lopinavir in fixed combination with ritonavir is available as tablets of 100 mg/25 mg and 200 mg/50 mg and as an oral solution (80/20 mg per mL) for pediatric use. The recommended dose of lopinavir in adults is 800 mg daily in combination with 200 mg of ritonavir, usually in two divided doses.

New PIs

A new class of HIV protease inhibitor (GS-PI1) with high potency, high resistance barrier and long half-life has been developed to achieve metabolic stability without pharmacokinetic boosting. However, no new data have been presented since 2017 ([Link et al., 2017](#)).

Tolerability of PIs

Common adverse events associated with protease inhibitors include hepatotoxicity and gastrointestinal side effects, dyslipidemia, lipodystrophy, insulin resistance and cardiovascular/renal toxicity.

Side effects

Hepatotoxicity

All the approved HIV protease inhibitors are metabolized by the liver and can inhibit the CYP 3A4 drug metabolizing enzymes. Among all PIs, ritonavir is the most potent CYP 3A4 inhibitor. Atazanavir may result in an increase in the indirect serum bilirubin concentration.

Gastrointestinal

Protease inhibitors may cause diarrhea, nausea and vomiting. Lopinavir/ritonavir and fosamprenavir/ritonavir tend to show the highest rates of drug-related diarrhea, compared with atazanavir/ritonavir, darunavir/ritonavir, or saquinavir/ritonavir ([Hill et al., 2009](#)).

Dyslipidemia, lipodystrophy and insulin resistance

The effect of HIV protease inhibitor on the lipid pathway is associated with the inhibition of SREBP-1, a transcription factor involved in the gene expression of adipocytes, leading to the deficiency of adiponectin, a protein hormone that plays a key role in the development of insulin resistance and lipodystrophy ([Bastard et al., 2002](#); [Rahmouni et al., 2008](#)).

In addition, PIs can inhibit glucose transporter-4, having a direct effect on glucose uptake and this leads to insulin resistance ([Murata et al., 2000, 2002](#)).

Cardiovascular

PIs including lopinavir and darunavir (boosted with ritonavir) but not atazanavir have been associated with excess risk of cardiovascular disease ([Ryom et al., 2018](#)). ATV, possibly through its effects on bilirubin, has been associated with reduced progression of atherosclerosis (carotid artery intima media thickness) ([Stein et al., 2015](#)).

Kidney stones

Atazanavir has been associated with the development of renal stones and interstitial nephritis ([Hamada et al., 2012](#)).

Drug–Drug Interactions

Due to increased levels of other metabolized drugs when boosted-PIs are co administered, it is important to check and consider any potential drug interaction between the PI-booster and other concomitant medication (See “Relevant Websites section”). PIs can affect other drugs but also PIs can be affected by other drugs. Main drug interactions are shown in [Tables 2 and 3](#).

Table 2 Example of drugs whose plasma exposure are altered by boosted-PI co-administration

<i>Drug</i>	<i>Exposure of the non-ARV drug</i>	<i>Mechanism of drug-drug interaction</i>
Warfarin	Variable	CYP2C19 inhibition/induction
Omeprazole	Variable	CYP2C19 inhibition/induction
Pravastatin	Increase	OATP1B1/B3 inhibition
Rosuvastatin		
Atorvastatin		
Simvastatin		
Amlodipine	Increase	CYP3A/5 inhibition
Diltiazem		
Verapamil		
Metoprolol		
Digoxin		
Citalopram	Increase	CYP-2D6 inhibition
Diazepam		
Mirtazapine		
Midazolam		
Paroxetine	Variable	
Sertraline		
Fluticasone	Increase	CYP3A/5 inhibition
Budesonide		
Salmeterol		
Sildenafil		
Ethinylestradiol	Decrease	CYP3A4 induction

Table 3 Example of drugs whose co-administration alters plasma PI exposure

<i>Drugs</i>	<i>Exposure of the PIs</i>	<i>Mechanism of drug-drug interaction</i>
Clarithromycin	Increase	Inhibition of CYP3A/5
Erythromycin		
Fluconazole		
Itraconazole		
Ketoconazole		
Fluoxetine		
Efavirenz		
Carbamazepine	Decrease	Induction of CYP3A/5
Phenytoin		
Phenobarbital		
Rifampin		
Rifabutin		
Efavirenz		

Resistance to PIs

PIs have a high genetic barrier to resistance compared with other classes of ART such as (first-generation) integrase strand transfer inhibitors (INSTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs), making them attractive agents for those with suboptimal adherence ([Gardner *et al.*, 2010](#)).

However, suboptimal plasma concentrations are highly undesirable as they can lead to drug resistance and subsequent virologic failure. In selected individuals, therapeutic drug monitoring may be a useful tool to assess adequate drug exposure ([Calcagno *et al.*, 2017](#)).

Protease Inhibitors in Low Resource Settings

Sub-Saharan Africa

In sub-Saharan Africa, where 25.5 million people are living with HIV, access to HIV drugs and treatment monitoring is compromised by a variety of factors. The challenges to adequately respond to the HIV crisis in Africa and achieve the 90/90/90 strategy include financial constrain, distance to health centers, inadequate diagnostic capacity (e.g., CD4 count, viral load), recurrent stock

out of ARVs and concurrent disease burdens. As a result, an increasing number of people in sub-Saharan Africa are at risk of treatment failure and drug resistance (Kanters *et al.*, 2017; Boyd *et al.*, 2015), and dolutegravir, a second generation INSTI with a high barrier to resistance, is increasingly replacing NNRTI in first-line therapy.

Boosted protease-inhibitor remain important drugs in second-line regimens. Boosted PI therapy plus two NRTIs, administered without genotypic resistance testing or regular viral-load monitoring, had high rate of success, showing effective viral suppression (Paton *et al.*, 2014). However, the use of protease inhibitor monotherapy has been shown to be inferior and should not be considered as an acceptable treatment alternative (Ciaffi *et al.*, 2017; Paton *et al.*, 2014). Therefore, the World Health Organization (WHO) recommends as second-line regimens a boosted protease inhibitor (atazanavir/ritonavir, lopinavir/ritonavir, or darunavir/ritonavir) with 2 NRTIs such as zidovudine with lamivudine (3TC) (WHO, 2019; SECOND-LINE Study Group, 2013).

Late presentation of HIV-positive adult patients into care is common in sub-Saharan Africa, with a high incidence of tuberculosis (TB). Several pharmacokinetic studies evaluating the effects of rifampicin on boosted PI exposure have reported high rates of liver toxicity (La Porte *et al.*, 2004; Ribera *et al.*, 2005; Haas *et al.*, 2009). Hence, for HIV/TB patients requiring PI-containing ART, rifampicin should be replaced with rifabutin (administered at a dose of 150 mg/day) where available (WHO, 2010). Unfortunately, access to rifabutin in low- and middle-income countries, is very limited (Grinsztejn *et al.*, 2014).

Protease Inhibitors in Clinical Practice

Protease inhibitors remain widely used in clinical practice despite a large number of antiretrovirals available for the treatment of HIV infection. Most guidelines recommend the use of NNRTI and INSTI (with a NRTI backbone) as first line therapy. In special populations, including those with anticipated adherence issues, confirmed or suspected HIV resistance or intolerance to other classes of ART, boosted PI remain widely used. In clinical trials of first or second line therapy, virologic failure on boosted PI (when administered with at least one NRTI) is rarely associated with emergence of PI (or NRTI) resistance-associated mutations. Hence, boosted PI are particularly attractive for people with adherence issues and the preferred agents for those who have developed resistance to NNRTI and/or INSTI.

References

- Ali, A., Bandaranayake, R.M., Cai, Y., *et al.*, 2010. Molecular basis for drug resistance in HIV-1 protease. *Viruses* 2, 2509–2535.
- Bastard, J.P., Caron, M., Vidal, H., *et al.*, 2002. Association between altered expression of adipogenic factor SREBP1 in lipotrophic adipose tissue from HIV-1-infected patients and abnormal adipocyte differentiation and insulin resistance. *Lancet* 359 (9311), 1026–1031.
- Boffito, M., Moyle, G.J., 2004. Can we boost enough without ritonavir? *AIDS Read* 14, 229–230.
- Boyd, M.A., Moore, C.L., Molina, J., *et al.*, 2015. Baseline HIV-1 resistance, virological outcomes, and emergent resistance in the SECOND-LINE trial: An exploratory analysis. *Lancet HIV* 2, e42–e51.
- Cahn, P., Andrade-Villanueva, J., Arribas, J.R., *et al.*, 2014. Dual therapy with lopinavir and ritonavir plus lamivudine versus triple therapy with lopinavir and ritonavir plus two nucleoside reverse transcriptase inhibitors in antiretroviral-therapy-naïve adults with HIV-1 infection: 48 week results of the randomised, open label, non-inferiority GARDEL trial. *The Lancet Infectious Diseases* 14 (7), 572–580. doi:10.1016/s1473-3099(14)70736-4.
- Calcagno, A., Pagani, N., Ariaudo, A., *et al.*, 2017. Therapeutic drug monitoring of boosted PIs in HIV-positive patients: undetectable plasma concentrations and risk of virological failure. *Journal of Antimicrobial Chemotherapy* 72, 1741–1744.
- Ciaffi, L., Koulla-shiro, S., Sawadogo, A.B., *et al.*, 2017. Boosted protease inhibitor monotherapy versus boosted protease inhibitor plus lamivudine dual therapy as second-line maintenance treatment for HIV-1-infected patients in sub-Saharan Africa (ANRS12 286/MOBIDIP): A multicentre, randomised, parallel, open-label, superiority trial. *The Lancet HIV* 4 (9), e384–e392.
- Clotet, B., Bellos, N., Molina, J.M., *et al.*, 2007. Efficacy and safety of darunavir-ritonavir at week 48 in treatment-experienced patients with HIV-1 infection in POWER 1 and 2: A pooled subgroup analysis of data from two randomised trials. *Lancet* 369 (9568), 1169–1178.
- Gardner, E.M., Hulsiek, K.H., Telzak, E.E., *et al.*, 2010. Antiretroviral medication adherence and class-specific resistance in a large prospective clinical trial. *AIDS* 24, 395.
- Grinsztejn, B., De Castro, N., Arnold, V., *et al.*, 2014. Raltegravir for the treatment of patients co-infected with HIV and tuberculosis (ANRS 12 180 reflate TB): A multicentre, phase 2, non-comparative, open-label, randomised trial. *Lancet Infectious Diseases* 14 (6), 459–467.
- Hamada, Y., Nishijima, T., Watanabe, K., *et al.*, 2012. High incidence of renal stones among HIV-infected patients on ritonavir-boosted atazanavir than in those receiving other protease inhibitor-containing antiretroviral therapy. *Clinical Infectious Diseases* 55, 1262.
- Haas, D.W., Koletar, S.L., Laughlin, L., *et al.*, 2009. Hepatotoxicity and gastrointestinal intolerance when healthy volunteers taking rifampin add twice-daily atazanavir and ritonavir. *Journal of Acquired Immune Deficiency Syndromes* 50 (3), 290–293. doi:10.1097/qai.0b013e318189a7df.
- Hill, A., Balkin, A., 2009. Risk factors for gastrointestinal adverse events in HIV treated and untreated patients. *AIDS Reviews* 11 (1), 30–38.
- Ho, D.D., Neumann, A.U., Perelson, A.S., *et al.*, 1995. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 373, 123–126.
- Kanters, S., Socias, M.E., Paton, N.I., *et al.*, 2017. Comparative efficacy and safety of second-line antiretroviral therapy for treatment of HIV/AIDS: A systematic review and network meta-analysis. *Lancet HIV* 4, e433–e441.
- Konvalinka, J., Kräusslich, H.-G., Müller, B., 2015. Retroviral proteases and their roles in virion maturation. *Virology* 479–480, 403–417.
- La Porte, C.J.L., Colbers, E.P.H., Bertz, R., *et al.*, 2004. Pharmacokinetics of adjusted-dose lopinavir-ritonavir combined with rifampin in healthy volunteers. *Antimicrobial Agents and Chemotherapy* 48 (5), 1553–1560. doi:10.1128/aac.48.5.1553-1560.2004.
- Larson, K.B., Wang, K., Delille, C., Otofokun, I., Acosta, E.P., 2014. Pharmacokinetic enhancers in HIV therapeutics. *Clinical Pharmacokinetics* 53, 865–872.
- Lennox, J.L., Landovitz, R.J., Ribaud, H.J., *et al.*, 2014. Efficacy and tolerability of 3 nonnucleoside reverse transcriptase inhibitor-sparing antiretroviral regimens for treatment-naïve volunteers infected with HIV-1: A randomized, controlled equivalence trial. *Annals of Internal Medicine* 161, 461.
- Link, J.O., Kato, D., Moore, M., *et al.*, 2017. Novel HIV PI with high resistance barrier and potential for unboosted QD oral dosing. In: CROI. Seattle, WA: Gilead Sciences.
- Madruca, J.V., Berger, D., McMurchie, M., *et al.*, 2007. Efficacy and safety of darunavir-ritonavir compared with that of lopinavir-ritonavir at 48 weeks in treatment-experienced, HIV-infected patients in TITAN: A randomised controlled phase III trial. *Lancet* 370 (9581), 49–58.
- Mills, A.M., Nelson, M., Jayaweera, D., *et al.*, 2009. Once-daily darunavir/ritonavir vs. lopinavir/ritonavir in treatment-naïve, HIV-1-infected patients: 96-week analysis. *AIDS* 23, 1679.

- Molina, J.M., Andrade-Villanueva, J., Echevarria, J., *et al.*, 2008. Once-daily atazanavir/ritonavir versus twice-daily lopinavir/ritonavir, each in combination with tenofovir and emtricitabine, for management of antiretroviral-naïve HIV-1-infected patients: 48 week efficacy and safety results of the CASTLE study. *Lancet* 372, 646.
- Monforte, A., Reiss, P., Ryom, L., *et al.*, 2013. Atazanavir is not associated with an increased risk of cardio- or cerebrovascular disease events. *AIDS* 27, 407.
- Murata, H., Hruz, P.W., Mueckler, M., 2000. The mechanism of insulin resistance caused by HIV protease inhibitor therapy. *Journal of Biological Chemistry* 275 (27), 20251–20254.
- Murata, H., Hruz, P.W., Mueckler, M., 2002. Indinavir inhibits the glucose transporter isoform Glut4 at physiologic concentrations. *AIDS* 16 (6), 859–863.
- Nathan, B., Bayley, J., Waters, L., Post, F.A., 2013. Cobicistat: A novel pharmacoenhancer for co-formulation with HIV protease and integrase inhibitors. *Infectious Diseases and Therapy* 2 (2), 111–122. doi:10.1007/s40121-013-0013-7.
- Ototokun, I., Na, L.H., Landovitz, R.J., *et al.*, 2015. Comparison of the metabolic effects of ritonavir-boosted darunavir or atazanavir versus raltegravir, and the impact of ritonavir plasma exposure: ACTG 5257. *Clinical Infectious Diseases* 60 (12), 1842–1851. doi:10.1093/cid/civ193.
- Palella, F.J., Delaney, K.M., Moorman, A.C., *et al.*, 1998. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *New England Journal of Medicine* 338, 853–860.
- Paton, N.I., Kityo, C., Hoppe, A., *et al.*, 2014. Assessment of second-line antiretroviral regimens for HIV therapy in Africa. *New England Journal of Medicine* 371 (3), 234–247. doi:10.1056/nejmoa1311274.
- Rahmouni, K., Sigmund, C.D., 2008. Id3, E47, and SREBP-1c: Fat factors controlling adiponectin expression. *Circulation Research* 103 (6), 565–567.
- Raffi, F., Babiker, A.G., Richert, L., *et al.*, 2014. Ritonavir-boosted darunavir combined with raltegravir or tenofovir–emtricitabine in antiretroviral-naïve adults infected with HIV-1: 96 week results from the NEAT001/ANRS143 randomised non-inferiority trial. *The Lancet* 384 (9958), 1942–1951. doi:10.1016/s0140-6736(14)61170-3.
- Ribera, E., Azuaje, C., Lopez, R.M., *et al.*, 2005. Once-daily regimen of saquinavir, ritonavir, didanosine, and lamivudine in HIV-infected patients with standard tuberculosis therapy (TBQD study). *Journal of Acquired Immune Deficiency Syndromes* 40 (3), 317–323. doi:10.1097/01.qai.0000182629.74336.4d.
- Ryom, L., Lundgren, J.D., El-Sadr, W., *et al.*, 2018. Cardiovascular disease and use of contemporary protease inhibitors: The D:A:D international prospective multicohort study. *The Lancet HIV* 5 (6), e291–e300. doi:10.1016/s2352-3018(18)30043-2.
- SECOND-LINE Study Group, 2013. Ritonavir-boosted lopinavir plus nucleoside or nucleotide reverse transcriptase inhibitors versus ritonavir-boosted lopinavir plus raltegravir for treatment of HIV-1 infection in adults with virological failure of a standard first-line ART regimen (SECOND-LINE): A randomised, open-label, non-inferiority study. *The Lancet* 381 (9883), 2091–2099. doi:10.1016/s0140-6736(13)61164-2.
- Squires, K.E., Young, B., DeJesus, E., *et al.*, 2010. Similar efficacy and tolerability of atazanavir compared with atazanavir/ritonavir, each with abacavir/lamivudine after initial suppression with abacavir/lamivudine plus ritonavir-boosted atazanavir in HIV-infected patients. *AIDS* 24, 2019.
- Stein, J.H., Ribbaudo, H.J., Hodis, H.N., *et al.*, 2015. A prospective, randomized clinical trial of antiretroviral therapies on carotid wall thickness. *AIDS* 29 (14), 1775–1783. doi:10.1097/qad.0000000000000762.
- Von Richter, O., Burk, O., Fromm, M.F., Eichelbaum, M., *et al.*, 2004. Cytochrome P450 3a4 and P-glycoprotein expression in human small intestinal enterocytes and hepatocytes: A comparative analysis in paired tissue specimens. *Clinical Pharmacology & Therapeutics* 75, 172–183.
- WHO, 2010. Priority Research Questions for Tuberculosis/Human Immunodeficiency Virus (TB/HIV) in HIV-Prevalent and Resource-Limited Settings. Geneva: World Health Organization.
- World Health Organization, 2019. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection. Geneva, Switzerland: WHO.

Further Reading

- Manosuthi, W., Wiboonchutikul, S., Sungkanuparph, S., 2016. Integrated therapy for HIV and tuberculosis. *AIDS Research and Therapy* 13, 22.
- Moyle, G.J., Back, D., 2001. Principles and practice of HIV-protease inhibitor pharmacoenhancement. *HIV Medicine* 2, 105–113.
- Renjifo, B., van Wyk, J., Salem, A.H., *et al.*, 2015. Pharmacokinetic enhancement in HIV antiretroviral therapy: A comparison of ritonavir and cobicistat. *Aids Reviews* 17, 37–46.
- Tseng, A., Hughes, C.A., Wu, J., Seet, J., Phillips, E.J., 2017. Cobicistat versus ritonavir: Similar pharmacokinetic enhancers but some important differences. *Annals of Pharmacotherapy* 51, 1008–1022.
- WHO, 2014. Global Tuberculosis Report. World Health Organization.

Relevant Websites

<https://www.hiv-druginteractions.org/checker>
University of Liverpool.

HIV Integrase Inhibitors and Entry Inhibitors

Daniel Bradshaw, Public Health England, London, United Kingdom

Ranjababu Kulasegaram, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Glossary

BCRP Breast cancer resistance protein.
BIC/TAF/FTC Bictegravir/tenofovir alafenamide/emtricitabine.
CAB Cabotegravir.
cART Combination antiretroviral therapy.
CCR5 C-C chemokine receptor type 5.
CD4 Cluster of differentiation 4.
CXCR4 C-X-C chemokine receptor type 4.
DTG Dolutegravir.
E/c/TAF/F Elvitegravir/cobicistat/tenofovir alafenamide/emtricitabine.
E/c/TDF/F Elvitegravir/cobicistat/tenofovir disoproxil fumarate/emtricitabine.
eGFR Estimated glomerular filtration rate.
FDA Food and Drug Administration.
FTC Emtricitabine.
HIV-1 Human immunodeficiency virus type 1.

HIV-2 Human immunodeficiency virus type 2.
HR1 First hepat repeat.
INSTI Integrase strand transfer inhibitors.
LEDGF Lens epithelium-derived growth factor.
LTR Long terminal repeat.
MATE 1 Multidrug and toxin extrusion transporter 1.
NNRTI Non nucleoside reverse transcriptase inhibitor.
NRTI Nucleos(t)ide reverse transcriptase inhibitor.
OATP Organic anion transporting polypeptide.
OBT Optimized background therapy.
OCT2 Organic cation transporter 2.
PEP Post-exposure prophylaxis.
PI Protease inhibitor.
PIC Pre-integration complex.
PNGS Potential N-linked glycosylation sites.
PrEP Pre-exposure prophylaxis.
TDF Tenofovir disoproxil fumarate.
UGT Uridine diphosphate glucuronosyltransferase.

Integrase Strand Transfer Inhibitors

Integrase strand transfer inhibitors (INSTI) are one class of antiretrovirals effective against both HIV-1 and HIV-2. Integrase represents an attractive target for drug design as it has no human counterpart and INSTI are now considered preferred options in treatment guidelines owing to their virological efficacy, safety, tolerability and, in most cases, minimal potential for interactions with co-medications. The first in class INSTI, raltegravir, was approved by the United States Food and Drug Administration (FDA) in 2007, with three further agents, cobicistat-boosted elvitegravir, dolutegravir and bictegravir subsequently approved. A fifth drug, cabotegravir, is undergoing clinical trials. See the [Table 1](#) for a summary of the key properties of approved INSTI and the investigational agent cabotegravir.

Virology

HIV-1 integrase is a 32 kDa protein produced from the C-terminal portion of the Pol gene product and catalyses the integration of reverse transcribed viral cDNA into the host chromosome through a two step process, 3' processing and strand transfer. First, integrase self-associates into tetramers on the cDNA within the cytoplasm and mediates the hydrolysis of the final two nucleotides on long terminal repeats (LTR) of the 3' ends of both strands to generate reactive nucleophilic hydroxyl groups. Integrase remains bound to cDNA as a multimeric complex which bridges both ends of the cDNA within intracellular particles called pre-integration complexes (PIC). After nuclear translocation, integrase associates with lens epithelium-derived growth factor (LEDGF)/p75 and is directed to sites of open chromatin, where it mediates the nucleophilic attack of the viral 3'-hydroxyl cDNA across the major groove of the host DNA, resulting in insertion sites which are five base pairs apart on opposite chromosomal strands. The integration process is completed by DNA gap repair including a series of DNA polymerization, dinucleotide excision and ligation reactions.

INSTI inhibit the second step catalysed by integrase, strand transfer, by competitive binding to the enzyme's active site, resulting in displacement of the 3' end of the cDNA. In addition, INSTI chelate the divalent cation (Mg^{2+} or Mn^{2+}), which is needed for integrase enzymatic activity.

The genetic barrier to resistance of the first generation INSTI, raltegravir and elvitegravir, is relatively low, such that a single nucleotide substitution may cause a major reduction in susceptibility, as well as cross resistance. Most primary resistance mutations occur at the enzyme's active site in response to drug pressure and are therefore associated with reduced viral replicative capacity (fitness). On continued drug exposure, primary resistance mutations are often followed by secondary (accessory) mutations, which either increase resistance or restore viral fitness, or both. The genetic barrier to resistance of the second generation agents, dolutegravir and bictegravir, is higher than first generation INSTI, probably due to structural differences. Second generation agents enter further into the pocket within integrase which is vacated by the displaced viral DNA base, making more

intimate contact with viral DNA, and they may also have the ability to re-adjust their conformation in response to structural changes in the active site of raltegravir-resistant integrase. Multiple amino acid substitutions are therefore usually required before antiviral susceptibility is reduced and resistance mutations are rarely selected in individuals failing dolutegravir- or bictegravir-containing cART. In general, epidemiological studies have found negligible (<1%) levels of INSTI resistance in therapy-naïve individuals, and therefore routine pre-therapy integrase sequencing is not currently recommended.

First Generation INSTI

Raltegravir

Raltegravir is a highly potent antiretroviral compound, with a minimal inhibitor concentration in the nanomolar range. Following 10 days of monotherapy, plasma HIV-1 RNA levels reduce by a mean 2.0 log₁₀ copies/ml. Dosing is with one 400 mg tablet twice daily, or two 600 mg tablets once daily. Raltegravir is generally well tolerated although the commonest reported side effects are headache, nausea and abdominal pain. Neuropsychiatric side effects have also been described, including insomnia and low mood. An uncommon but important toxicity is elevation in serum creatinine kinase levels, which may rarely be associated with rhabdomyolysis.

Raltegravir has a low potential for interaction with co-medications. However, as metabolism is through glucuronidation by uridine diphosphate glucuronosyltransferase (UGT) 1A1, strong inducers of UGT1A1, such as rifampicin, reduce raltegravir exposure. Raltegravir should therefore be avoided with strong UGT1A1 inducers where possible, or administered at a dose of 800 mg twice daily where co-administration is unavoidable. The site which binds Mg²⁺ or Mg²⁺ can be chelated by orally ingested polyvalent cations, such as Al³⁺ or Mg²⁺, which may reduce absorption. Raltegravir should therefore not be co-administered with aluminum or magnesium containing antacids.

Resistance

Most resistance mutations involve amino acid substitutions in the vicinity of the integrase inhibitor binding pocket, either directly or indirectly altering the conformation of the active site. The signature mutations, observed both in vitro and in vivo, are at residues Q148 (to H, R or K), N155 (to H) and less frequently Y143 (to C, H or R). These mutations reduce enzymatic activity and therefore fitness and are likely to disappear on discontinuation of raltegravir. Accessory mutations, some of which are natural polymorphisms, are also mostly located close to the active site.

The Y143 pathway is specific to RAL and relatively uncommon; residue 143 interacts directly with the oxadiazole ring of raltegravir, forming a stacking interaction which is abrogated when this position is mutated. By contrast, changes at residues 148 and 155 disturb the geometry of the integrase active site and thereby disrupt raltegravir binding.

Q148 mutations impart a severe fitness cost and therefore are rapidly compensated for by secondary mutations including L74M plus E138A, E138K or G140S. The most common mutational pattern in this pathway is Q148H plus G140S, which also confers the greatest loss of drug susceptibility.

N155 mutations tend to predominate early in the course of raltegravir failure, probably as changes at this position have less impact on fitness, but in ongoing drug exposure, are replaced by viruses with higher resistance, often bearing Q148H/R/K plus G140S mutations.

Clinical trials in cART naïve individuals

In a randomised controlled phase 3 clinical trial conducted in therapy-naïve patients, the virological efficacy of raltegravir was non-inferior to that of efavirenz, with fewer drug-related adverse events and smaller elevations in lipid levels (The STARTMRK study). Although originally licensed for twice daily dosing (400 mg BD), the ONCEMRK study demonstrated non-inferiority of once daily dosing of 1200 mg in achieving viral suppression.

Raltegravir-containing nucleos(t)ide-sparing regimens in therapy-naïve patients may also be effective, such as the combination of twice daily raltegravir and once daily ritonavir-boosted darunavir in patients with a baseline CD4 cell count >200 cells/uL and a HIV-1 RNA level <100,000 c/ml (NEAT001/ANRS-143 Study).

Clinical trials in cART-experienced patients

The phase 3 BENCHMRK studies demonstrated that the addition of raltegravir to optimized background therapy (OBT) in INSTI-naïve patients with multi-resistant HIV-1 infection improved virological suppression rates. However, in the SWITCHMRK studies, patients with suppressed plasma HIV-1 RNA switching from ritonavir-boosted lopinavir to raltegravir, both in combination with two nucleos(t)ide inhibitors, were more likely to develop virological failure. This was probably driven by archived NRTI resistance, compromising the NRTI backbone, in conjunction with the switch from a third agent with a high to a low genetic barrier to resistance. In the EARNEST study, patients with HIV-1 viraemia on an NNRTI-containing regimen were switched to ritonavir-boosted lopinavir, either as monotherapy, or with raltegravir, or with two to three NRTIs. Lower levels of viral RNA suppression were seen in those receiving a raltegravir- versus a NRTI-containing regimen, possibly reflecting the short half life of raltegravir and therefore greater susceptibility to development of resistance during episodes of non-adherence.

Table 1 HIV integrase strand transfer inhibitors and entry inhibitors

Drug name	Drug class	Year of FDA approval	Administration	Pharmacokinetics	Toxicity	Resistance	Examples of phase 3 studies in therapy-naïve patients	Examples of phase 3 studies in therapy-experienced patients	Other notes
<i>Integrase strand transfer inhibitors</i>									
Raltegravir	1st generation INSTI	2007	Oral	Metabolised by UGT1A1	Nausea, abdominal pain, headache	Low genetic barrier	STARTMRK-1, -2	BENCHMRK	No single tablet regimen
			Once or twice daily	Avoid with strong UGT1A1 inducers if possible	Rarely raised CK, myopathy or neuropsychiatric effects	Major RAM include Y143C/H/R, Q148H/R/K, N155H	ONCEMRK	SWITCHMRK	Longest period of safety data
				Caution with polyvalent cations			NEAT001/ANRS-143	EARNST	
Elvitegravir	1st generation INSTI	2012	Oral	Elvitegravir is metabolised by CYP450 3A4	Nausea, diarrhoea, headache	Low genetic barrier	GS-US-236-0102	GS-US-292-0109	Taken with food
			Once daily	Cobicistat inhibits CYP450 3A & 2D6	Rarely neuropsychiatric effects	Major RAM include T66I, E92Q, Q148H/R/K, N155H	GS-US-236-0103		Multiple drug-drug interactions
			Single tablet regimens:	Cobicistat inhibits transporters e.g., P-gp, MATE1, BRCP, OATP1B1, OATP1B3	↓tubular secretion of creatinine but normal GFR		GS-US-292-0104		
			E/c/TDF/FTC or E/c/TAF/FTC	Avoid with potent CYP450 3A inducers Caution with drugs metabolised by CYP450 3A Caution with polyvalent cations			GS-US-292-0111		
Dolutegravir	2nd generation INSTI	2013	Oral	Mainly metabolised by UGT1A1	Nausea, diarrhoea, headache	High genetic barrier	Single	Sailing	More insomnia & headache than other INSTI
			Once daily or	Some metabolism through UGT1A3/A9, CYP3A4, P-gp, BCRP	Neuropsychiatric effects	Multiple mutations usually required to reduce susceptibility	Spring-2	Viking-3	Further data required on association with (i) weight gain
			Twice daily if INSTI resistance	Inducers/inhibitors of these enzymes may ↓/↑DTG levels		RAM include G148H/R plus G140S	Flamingo	Dawning	(ii) neural tube defects in babies born to women conceiving on DTG
			Single tablet regimens:	DTG inhibits OCT2 and MATE1			Gemini-1, -2	Tango	
			DTG/ABC/3TC, DTG/3TC, DTG/RPV	This may ↑ serum creatinine or metform exposure Caution with polyvalent cations				Striving	
								SWORD-1, -2	

(Continued)

Table 1 Continued

Drug name	Drug class	Year of FDA approval	Administration	Pharmacokinetics	Toxicity	Resistance	Examples of phase 3 studies in therapy-naïve patients	Examples of phase 3 studies in therapy-experienced patients	Other notes
Bictegravir	2nd generation INSTI	2018	Oral	Metabolised by CYP3A and UGT1A1	Headache, nausea, diarrhoea	High genetic barrier	GS-US-380-1489	GS-US-380-1878	No long term safety data in real world cohorts
			Once daily	Inducers/inhibitors of these enzymes may ↓/↑ BIC levels	Rarely neuropsychiatric effects	Multiple mutations usually required to reduce susceptibility			
			Single tablet regimen BIC/TAF/FTC	Metabolised by P-gp and BCRP	↓ tubular secretion of creatinine but GFR normal	RAM include G118R, G148H/R plus G140S, R263K			
				Caution with inhibitors of these transporters					
Cabotegravir	2nd generation INSTI	Investigational	Oral tablet, once daily <i>or</i> Intramuscular injection 4- or 8-weekly	Metabolised by UGT1A1	Injection site reactions (mostly mild)	Minimal resistance data available RAM include G140R, Q148R	None	ATLAS	Ongoing phase 3 treatment trials with CAB/RPV Ongoing phase 3 PrEP trials (HPTN 084)
								FLAIR	
								ATLAS-2M	
<i>Entry inhibitors</i>									
Enfuvirtide (T20)	Fusion inhibitor	2003	Subcutaneous injection	No clinically significant interactions	Injection site reactions, diarrhoea, nausea, peripheral neuropathy, pancreatitis	RAM in amino acids 36-45 within HR1 of gp41	None	TORO-1, -2	Toxicity, subcutaneous route & twice daily dosing limit use
			Twice daily						
Maraviroc	CCR5 receptor antagonist	2007	Oral	Metabolised by CYP450 3A4 and 3A5	Nausea, diarrhoea, fatigue, headache	X4 or X4/R5 dual or mixed tropism	MERIT	MOTIVATE-1, -2	Rarely used if therapy-naïve
			Twice daily	Inducers/inhibitors of CYP450 may ↓/↑ MVC levels MVC dose adjusted according to presence of these agents	Less commonly postural hypotension	Mutations in the V3 loop of gp120			Contraindicated in X4, dual or mixed (R5/X4) strains
Ibalizumab-uiyk	CD4-directed post-attachment inhibitor	2018	Intravenous injection	No clinically significant interactions	Diarrhoea, dizziness, nausea and rash	Minimal data available	None	TMB-301	Monoclonal antibody (IgG isotype 4) against extracellular domain of CD4
			Once per fortnight			May involve RAM in the V5 loop of gp120			
Fostemsavir	HIV-1 gp120-directed attachment inhibitor	2020	Oral prodrug of temsavir	Metabolised by esterase-mediated hydrolysis, CYP450 3A4, P-gp, BCRP	Minimal data	Minimal data	None	BRIGHT-E	Binds extracellular domain of HIV-1 gp120
			Twice daily			Likely to involve RAM in gp120			

Elvitegravir

Elvitegravir was approved by the FDA in 2012 and is only available in the single tablet regimens of elvitegravir, cobicistat, emtricitabine and either tenofovir disoproxil fumarate (E/c/TDF/F) or tenofovir alafenamide (E/c/TAF/F). It is used in treatment naïve patients and as a switch option in those with a suppressed HIV-1 RNA without NRTI or INSTI resistance.

Elvitegravir is generally well tolerated. The commonest side effects are nausea, diarrhea and headache, and other side effects include rash and abnormal dreams. The main factors limiting its more widespread use are a relatively low genetic barrier to resistance and multiple potential interactions with co-medications. Elvitegravir is coadministered with the boosting agent cobicistat, which has no antiretroviral activity but is a potent inhibitor of cytochrome P450 CYP3A4/5. As elvitegravir is metabolized through this pathway, inhibition of CYP3A4/5 boosts elvitegravir plasma exposure, ensuring therapeutic levels are achieved. Inducers of cytochrome P450, such as rifampicin, greatly reduce elvitegravir plasma concentrations and concomitant administration is contraindicated. Caution is also required with cytochrome P450 3A inhibitors such as itraconazole, which may increase exposure to elvitegravir.

Cobicistat also has the potential for multiple interactions. Inhibition of CYP3A by cobicistat results in an increase in plasma levels of drugs metabolized through this pathway. For example, boosting of exogenous corticosteroids may produce iatrogenic Cushing's disease or suppress the endogenous hypothalamic-pituitary-adrenal axis with consequent adrenal insufficiency. Cobicistat also inhibits certain transporters, such as P-glycoprotein, multidrug and toxin extrusion transporter 1 (MATE 1), breast cancer resistance protein (BCRP) and organic anion transporting polypeptide (OATP).

As renal tubular secretion of creatinine is dependent on MATE 1, inhibition of this transporter by cobicistat results in increased serum creatinine and hence a reduced estimated glomerular filtration rate (eGFR), although without a true decline in GFR. Where cobicistat is prescribed with TDF, it is not recommended to be initiated in patients with a creatinine clearance of <70 ml/min, or if coformulated with TAF, <30 ml/min.

Clinical trials

Phase 3 trials demonstrated high rates of HIV-1 RNA suppression in therapy-naïve patients receiving either E/c/F/TAF or E/c/F/TDF, as well as non-inferiority of E/c/F/TDF by comparison with efavirenz or ritonavir-boosted atazanavir, both combined with TDF/FTC (Studies 104, 111, 102, 103). High rates of HIV-1 RNA suppression were also observed in therapy-experienced patients with a fully suppressed plasma HIV-1 RNA and no history of resistance to any of the components of the fixed drug combination who switched to E/c/F/TAF (Study 109).

Resistance

Seven elvitegravir codon mutations have been described in INSTI treatment-naïve and -experienced patients in whom therapy is failing. Elvitegravir shares the Q148 and N155H resistance pathways with raltegravir, whilst the T66 and E92 pathways are predominantly selected by elvitegravir. Y97A is associated with only a 2-fold reduction in elvitegravir susceptibility and may be a polymorphism.

Second Generation INSTI

Dolutegravir

The second generation INSTI dolutegravir was approved in 2013 and is dosed as a 50 mg oral tablet once daily, or twice daily in individuals with confirmed or suspected INSTI resistance. Dolutegravir is also available as a component of the single tablet regimens of dolutegravir-abacavir-lamivudine, dolutegravir-lamivudine, or dolutegravir-rilpivirine. Its once daily dosing, virological efficacy, tolerability and high genetic barrier to resistance have ensured its incorporation as a preferred option in many treatment guidelines.

Although well tolerated, the commonest side effects reported with dolutegravir are nausea, diarrhea and headache. Uncommon but notable associated toxicities include psychiatric disorders such as insomnia and anxiety; suicidal ideation has also rarely been reported. An association between dolutegravir use and weight gain has also recently been described although other factors, such as gender, racial background or co-administered antiretrovirals, may play a role. It is also unclear whether or not the association is pathological or represents a 'return to health' phenomenon.

A slightly increased risk of neural tube defects in babies born to mothers who conceived whilst receiving dolutegravir has also been reported. In the largest cohort reporting outcomes for women receiving a dolutegravir containing regimen in Botswana, neural tube defects were identified in 0.3% of 1683 deliveries, compared to 0.1% of deliveries born to women receiving a non-dolutegravir containing regimen. Neural tube closure occurs in the first four weeks after conception and therefore initiation of dolutegravir is relatively contraindicated during conception and in the first six weeks of pregnancy, unless alternative agents are unavailable. However, the mechanism by which dolutegravir might interfere with pregnancy is not established.

Dolutegravir is mainly eliminated through the UGT1A1 pathway but is also a substrate of UGT1A3, UGT1A9, CYP3A4, P-glycoprotein and BCRP. Therefore, co-administration of inducers of these enzymes reduces plasma exposure to dolutegravir and twice daily dolutegravir dosing may be required, or, in individuals with INSTI resistant virus, dolutegravir should be avoided. Dolutegravir also inhibits the renal organic cation transporter 2 (OCT 2) and MATE 1. As the secretory fraction of creatinine clearance is dependent on these transporters, a slight decrease in creatinine clearance is observed with dolutegravir. Plasma levels of drugs which are excreted through these transporters, such as metformin, may also increase.

Clinical trials: Therapy-naïve with 2 NRTI

Dolutegravir is highly efficacious in suppressing plasma HIV-1 RNA in randomized controlled trials in therapy-naïve individuals. Non-inferiority has been reported when comparing the performance of dolutegravir against that of raltegravir, and superiority when compared to efavirenz or ritonavir-boosted darunavir (the Spring-2, Single and Flamingo studies, respectively).

Clinical trials: Treatment-experienced

Trials reporting outcomes of individuals with suppressed HIV-1 RNA who either continued their current ART or switched to a dolutegravir-based regimen, such as the Striving study, also demonstrated non-inferiority in HIV-1 RNA responses.

In addition, dolutegravir is highly effective in the treatment of patients failing ART. In the Sailing study, INSTI-naïve individuals with unsuppressed plasma HIV-1 RNA on antiretroviral therapy were randomised to receiving OBT with either twice daily raltegravir or once daily dolutegravir for 48 weeks. Greater virological efficacy was reported in the dolutegravir group. The performance of dolutegravir has also been evaluated against a PI in patients with viraemia on first line treatment with NNRTI plus 2 NRTI. In the Dawning study, individuals with virological failure were randomised to receiving dolutegravir once daily or ritonavir-boosted lopinavir, both with two NRTIs, at least one of which was predicted to be fully active according to the resistance profile. At week 48, the proportion with a suppressed (<50 c/ml) HIV-1 RNA was greater in the dolutegravir versus the lopinavir group (84% versus 70%) respectively. However, it remains unclear whether a regimen such as dolutegravir-tenofovir-lamivudine will be effective in individuals with virologic failure who harbor multiple NRTI resistance mutations, such as K65R plus M184V.

Efficacy of dolutegravir has also been demonstrated in the difficult to treat population of INSTI-experienced individuals with viraemia and INSTI-resistant virus. In a single arm study performed in this population (Viking-3), addition of twice daily dolutegravir to an optimized background regimen resulted in HIV-1 RNA suppression <50/cml in 69% individuals by week 24.

Clinical trials: Dual therapy

A novel approach involving therapy with dolutegravir within a NRTI-sparing or single NRTI containing regimen may be advantageous, given the toxicities associated with NRTIs such as TDF or abacavir. The SWORD-1 and -2 studies demonstrated that HIV-1 RNA suppression rates following a switch of triple ART to the single tablet regimen of dolutegravir-rilpivirine was non-inferior to continuing triple therapy in patients with a suppressed HIV-1 RNA and no history of failure or NNRTI/INSTI resistance. In the GEMINI-1 and -2 studies, use of dolutegravir plus lamivudine was shown to be non-inferior in therapy-naïve individuals with a HIV-1 RNA of <500,000 c/ml and no resistance, in comparison to dolutegravir-emtricitabine-TDF. No viral resistance was selected and fewer drug-related adverse events were observed in the dolutegravir-lamivudine group. Dolutegravir-lamivudine may also be indicated as a switch option in patients with suppressed HIV-1 RNA (the TANGO study).

Resistance

Individuals with virologic failure whilst receiving a first line, dolutegravir-containing regimen, which contains at least one other agent, rarely develop resistance mutations in either integrase or other viral genes. Similarly, INSTI-naïve individuals with a history of virological failure, who initiate dolutegravir containing cART and subsequently fail to achieve virological control, rarely (~1%) develop integrase resistance.

By contrast, additional integrase resistance mutations may accumulate in individuals with INSTI resistant virus who switch from a failing first generation INSTI-containing regimen to dolutegravir. For example, in the Viking-3 study, although plasma HIV-1 RNA suppressed following the addition of dolutegravir in the majority, 32/183 (17%) patients developed additional INSTI resistance mutations associated with virological failure at week 24, most often in those with a history of Q148 mutations. INSTI resistance also frequently develops in patients receiving dolutegravir monotherapy, a strategy which should therefore be avoided.

Several integrase mutations are usually needed to confer high-level resistance to dolutegravir. The mutation G148H/R plus G140S, in combination with L74I/M, E92Q, T97A, E138A/K, G140A or N155H, are associated with 5–20 fold reduced dolutegravir susceptibility. The G118R pathway, usually involving at least one other mutation, is also associated with reduced susceptibility. The R263K mutation alone reduces dolutegravir susceptibility by only 2-fold but resistance increases in the context of additional mutations. N155H alone usually yields <2 fold reduction in dolutegravir susceptibility.

Bictegravir

Bictegravir is the most recent INSTI to be approved and is administered within a single tablet regimen with TAF and emtricitabine. It is a second generation INSTI and, like dolutegravir, has a high genetic barrier to resistance.

Bictegravir is a substrate of CYP3A and UGT1A1 and therefore caution may be required with co-administration of inducers or inhibitors of these enzymes. Bictegravir inhibits OCT2 and MATE1 although it does not appear to increase metformin exposure significantly, in contrast to dolutegravir. Bictegravir also inhibits tubular secretion of creatinine leading to raised serum creatinine levels, although these are not considered to be clinically relevant.

The drug is generally well tolerated. The most frequently reported adverse reactions of bictegravir are headache, nausea and diarrhea and less commonly the class-specific neuropsychiatric symptoms.

Phase 3 randomized controlled trials

In trials involving therapy-naïve HIV-1 infected adults who were randomised to receiving either BIC/TAF/FTC or a dolutegravir-containing regimen, non-inferiority was observed in rates of virological suppression. No treatment emergent resistance was seen in

the bictegravir group. Bictegravir is also effective at maintaining viral suppression in patients switching from other therapies, including in those infected with virus harboring M184V/I mutations.

Resistance

Bictegravir, like dolutegravir, retains activity against most viral isolates from patients harboring multi-drug resistant virus who have failed an INSTI containing regimen. In vitro, the G118R and R263K mutations confer a 14-fold, and 6-fold decrease in susceptibility, respectively. The G140S + Q148H combination reduces susceptibility to bictegravir by 5-fold, with further loss of susceptibility conferred by one or more additional mutations (L74M, T97A, S119P/T, E138A/K or Y143C/R/H and G163R). Further resistance data are likely to emerge with bictegravir's increasing use.

Cabotegravir

Cabotegravir is an investigational INSTI with structural similarities to dolutegravir. It may be administered as an oral tablet once daily or by intramuscular injection every four or eight weeks. The injectable route is enabled due to packaging of the drug within nanoparticles, resulting in a long half life (~40 days).

In the phase 3 ATLAS and FLAIR studies, individuals with suppressed plasma HIV-1 RNA receiving oral antiretroviral therapy were either switched to injections of cabotegravir and rilpivirine every four weeks or maintained on oral therapy. A comparison of the two groups at 48 weeks demonstrated non-inferiority in HIV RNA suppression rates. Although 21% of participants found injections painful, most of these considered the pain acceptable and overall <1% discontinued injectables. A preference for injectable over oral treatment was reported by 98% of participants. In a subsequent study, ATLAS-2M, therapy-experienced individuals with suppressed HIV-1 RNA were randomised to receiving either 4-weekly or 8-weekly injectable cabotegravir-rilpivirine: after 48 weeks, there was no loss of virological control using the 8-week regimen.

Overall, in the ATLAS and FLAIR studies, there were only six (0.5%) virologic failures with resistance in the group receiving injectable therapy, including two individuals infected with virus with pre-existing NNRTI resistance. However, at virological failure, all had NNRTI resistance and 4/6 patients developed INSTI resistance mutations. The development of Q148R or G140R was observed particularly in HIV-1 A1 subtype harboring the L74I integrase polymorphism.

Injectable cabotegravir is also a candidate drug for pre-exposure prophylaxis against HIV but acute HIV infection must be carefully excluded before initiating cabotegravir owing to the risk of resistance. In macaques who had been recently infected with a chimeric simian/human immunodeficiency virus, subsequent cabotegravir injections led to selection of INSTI resistance in 3/6 animals, including the mutations G118R, E92G/Q or G140R.

Entry Inhibitors

A growing number of antiretroviral agents target steps in virus entry into host cells. Currently approved drugs include the CCR5 co-receptor antagonist, maraviroc, a gp41-directed fusion inhibitor, enfuvirtide, a CD4-directed post-attachment inhibitor, ibalizumab, and an investigational drug, gp120-directed attachment inhibitor, fostemsavir. See the [Table 1](#) for a summary of the key properties of these HIV entry inhibitors.

Maraviroc

The first stage in the HIV-1 replication cycle is viral attachment to the cell membrane. The CD4⁺ glycoprotein expressed on T lymphocytes and macrophages/monocytes is the major receptor for HIV-1 and binds to the viral envelope protein, gp120. A conformational change in gp120 follows, exposing the V3 loop which then binds a cellular co-receptor, either the C-C chemokine receptor type 5 (CCR5) or the C-X-C chemokine receptor type 4 (CXCR4). Further conformational changes within the gp41 subunit enable fusion of the viral envelope and cell membrane and the release of the viral particle into the cytoplasm.

Around 80% of HIV-1 transmission involves CCR5-using viruses. Individuals who are homozygous for a gene encoding a non-functional CCR5 protein are therefore relatively resistant to HIV-1 infection. Over time, CXCR4-using viruses emerge in around 50% of untreated individuals, particularly in those with low CD4⁺ T lymphocyte counts. X4 virus also frequently predominates in those failing antiretroviral regimens. In some patients, viral strains can use both receptors (dual tropism) or there may be a mixed infection of R5 and X4 viruses.

Maraviroc binds to a hydrophobic pocket in the transmembrane helices of CCR5, leading to conformational change in the CCR5 extracellular loops. This prevents binding of CCR5 to the V3 loop of gp120 and thereby blocks virus entry.

Maraviroc is ineffective against treating X4 strains and therefore the presence of X4, dual or mixed tropic strains must be excluded prior to its use by sequencing of the V3 loop in HIV-1 RNA. If the RNA level is below the cut off required for reliable amplification, genotypic testing may be performed on proviral HIV-1 DNA. Tropism testing is not routinely undertaken in newly-diagnosed patients, as maraviroc is not a preferred first line agent in therapy-naïve individuals. In addition, tropism may change with time, and therefore the finding of R5 virus at diagnosis may be misleading for future treatment decision-making.

Maraviroc is approved for oral administration at a usual dose of 300 mg twice daily. However, it is metabolized by cytochrome P450CYP3A4 or 3A5 and therefore the dose is halved in the presence of co-administered inhibitors of these enzymes, such as ritonavir or cobicistat, or doubled in the presence of inducers, such as efavirenz.

Although usually well tolerated, the commonest side effects reported in trials are nausea, diarrhea, fatigue and headache. An additional important but uncommon toxicity is postural hypotension.

In the phase 3 MERIT trial, maraviroc was compared to efavirenz in therapy-naïve individuals, both co-administered with an NRTI backbone. The primary endpoint of non-inferiority in virological suppression rates was not met, with less viral suppression in the maraviroc arm. However, exclusion of 15% of patients who would have been ineligible by a more sensitive tropism screening assay resulted in similar response rates in the two arms.

The benefit of including maraviroc, as an antiretroviral with a novel mechanism of action, in therapy-experienced patients was demonstrated in the MOTIVATE-1 and -2 trials. Here, individuals with R5 HIV-1 infection, who had been treated with and/or had resistance to three antiretroviral drug classes, were randomised to continuing OBT or OBT plus maraviroc. HIV-1 RNA was <50 copies/ml at week 48 in 47% of those in the group receiving twice daily maraviroc, compared to 16% in the placebo group. Thus maraviroc may be included within a salvage regimen in ART-experienced individuals with R5 tropic virus, although it should not be the only fully active agent.

Antagonism of CCR5 may also have immunomodulatory effects, and trials are ongoing into the use of maraviroc for the treatment of co-morbidities with an inflammatory component, such as non-alcoholic fatty liver disease, in people with HIV infection.

Resistance to maraviroc may develop through one of two mechanisms. X4 using virus or dual or mixed tropic strains may emerge, often representing outgrowth of a pre-existing minority population. Alternatively, mutations within the V3 loop may enable gp120 to bind to CCR5 in the presence of maraviroc, although there is no consensus on the specific signature mutations. Some CCR5 antagonist-resistant viruses selected in vitro have mutations in gp41 only, but the clinical significance of these remains unknown.

Enfuvirtide

Enfuvirtide or T20 is the only approved drug in the class of antiretrovirals known as fusion inhibitors. Enfuvirtide binds to an extracellular region, the first heptad repeat (HR1), of the fusion peptide gp41, preventing its structural rearrangement and thereby blocking fusion of the viral envelope and cell membrane.

Enfuvirtide is administered by subcutaneous injection, twice daily. The most frequently reported adverse reactions are injection site reactions, diarrhea and nausea. However, numerous other adverse reactions may occur including diabetes mellitus, pancreatitis and peripheral neuropathy. No clinically significant pharmacokinetic interactions have been noted with the CYP450 pathway. However, the subcutaneous route of administration, twice daily dosing requirement and poor tolerability limit the utility of enfuvirtide in most scenarios except as a component of a salvage regimen in multi-resistant HIV-1.

The phase 3 TORO-1 and -2 studies demonstrated that the addition of enfuvirtide to optimized OBT resulted in a significantly greater reduction in HIV-1 RNA levels in individuals failing an antiretroviral regimen than in those receiving OBT plus placebo.

Resistance may develop in the context of incomplete viral suppression and is predominantly associated with mutations in gp41 HR1. However, mutations in other regions of the envelope gene, as well as co-receptor usage and density, may also affect susceptibility to enfuvirtide.

Ibalizumab-Uiyk

Ibalizumab was approved by the FDA in 2018 and is a first in class CD4-directed post-attachment HIV-1 inhibitor. It is a monoclonal antibody of IgG isotype 4 directed against domain 2 of the extracellular portion of the CD4 glycoprotein and is produced by recombinant DNA technology in murine myeloma non-secreting 0 cells. By non-competitive binding to CD4, ibalizumab interferes with post-attachment steps and thereby prevents fusion of the viral envelope and cell membranes. Ibalizumab, unlike maraviroc, is active against both R5 and X4 tropic HIV-1 strains and is indicated for the treatment of HIV-1 infected people with multi-drug resistant infection who are failing their current antiretroviral regimen.

Ibalizumab is administered intravenously as a single loading dose followed by a fortnightly maintenance dose. The most common adverse reactions are diarrhea, dizziness, nausea and rash. Drug-drug interactions have not been observed. Of note, ibalizumab binding does not impact CD4 function, given that the target is a domain on CD4 distinct from the binding site of MHC Class II molecules.

In the phase 3 TMB-301 study, heavily treatment-experienced HIV-1-infected participants with unsuppressed HIV RNA levels and resistance to at least one NRTI, NNRTI and PI, received ibalizumab with OBT. By week 25, 43% of individuals achieved an HIV-1 RNA level below <50 c/ml.

Decreased susceptibility to ibalizumab has been observed in some individuals experiencing virological failure and may be associated with genotypic changes in the HIV-1 envelope gene which results in loss of potential N-linked glycosylation sites (PNGS) in the V5 loop of gp120.

Fostemsavir

Fostemsavir is an oral prodrug of temsavir, a first in class attachment inhibitor which binds to the HIV-1 envelope protein gp120. It is an investigational agent, currently undergoing phase 3 clinical trials.

Temsavir binds to gp120 close to the CD4 binding site and prevents initial attachment to the CD4 receptor. By targeting the first step of viral attachment, before co-receptor binding and fusion, temsavir is unaffected by co-receptor tropism. In one arm of a phase 3 trial in which fostemsavir was added to OBT in heavily treatment experienced patients failing their current cART regimen (the BRIGHT study), 59% of individuals achieved HIV-1 RNA levels <40 c/ml by week 96. In phase 2 studies, several patients developed emergent substitutions in gp120 associated with reduced temsavir susceptibility although some of these individuals subsequently achieved viral re-suppression. Further data on the resistance profile are therefore required.

Pregnancy

Raltegravir is the preferred INSTI in pregnant women in view of the availability of long term safety data indicating no increased risk to the fetus. Elvitegravir/cobicistat should not be used owing to limited safety data and the risk of subtherapeutic elvitegravir plasma levels in the second and third trimesters. The slightly increased risk of neural tube defects in babies born to women who conceived on dolutegravir is outlined above. There are currently insufficient data concerning bictegravir for this to be recommended in pregnancy.

There are also only limited data on the use of maraviroc and enfuvirtide in pregnancy and no adequate data concerning the use of ibalizumab in pregnant women.

Children and Adolescents

An INSTI in conjunction with two NRTI is frequently the preferred regimen for HIV-1 infected children, with longer term safety and efficacy data in children available for raltegravir than other agents. Raltegravir may be administered as oral suspension, granules or chewable tablets and approved for use in infants and children weighing at least 2 kg under the FDA label. In children older than 6 years of age weighing at least 25 kg E/c/F/TAF may be used or, above 15 kg, a dolutegravir-containing regimen. Safety and efficacy of BIC/TAF/FTC in children under 18 years of age is not established.

Regarding entry inhibitors, maraviroc is approved for use in children weighing at least 10 kg who are over the age of 2 years. Enfuvirtide may be used in children 6 years of age or older.

Post-Exposure and Pre-Exposure Prophylaxis (PEP and PrEP)

INSTI are good candidates for use in PEP with a favorable safety and tolerability profile and the theoretical advantage of acting early in the virus replication cycle to prevent integration of viral DNA. The most frequently recommended regimen for PEP following a significant exposure to HIV-1 is a 28 day course of raltegravir plus tenofovir disoproxil fumarate and emtricitabine although alternative regimens may be prescribed. For example, DTG/TDF/FTC was used successfully in one open-label, single arm study. In cases of exposure to a resistant strain of HIV, the resistance profile may need to be reviewed before determining the optimal PEP regimen.

Although the only approved regimen for PrEP requires FTC plus TDF or TAF, phase 3 studies are ongoing into the use of alternative agents including cabotegravir. The injectable formulation is an attractive candidate owing to its long plasma half life and trials are investigating the approach of administration every 8 weeks. However, concerns have been raised that INSTI resistance may develop, for example, following initiation of cabotegravir during unrecognized primary HIV infection or after exposure to HIV during the long pharmacokinetic tail in people discontinuing therapy.

HIV-2

Few trials evaluating the use of antiretroviral agents have been completed in people with HIV-2, including no randomised controlled trials. However, INSTI are active against HIV-2 integrase in vitro and single-arm, open label clinical trials involving raltegravir plus TDF/FTC or EVG/c/TDF/FTC have shown favorable treatment responses in ART naïve individuals with HIV-2 over 48 weeks. A randomized controlled trial comparing the use of raltegravir plus TDF/FTC to lopinavir/ritonavir plus TDF/FTC is currently underway and a single arm open-label trial evaluating the use of dolutegravir with two NRTI in ART-naïve individuals is also ongoing.

Maraviroc appears to be active against some HIV-2 isolates; however, there are no FDA-approved assays that can determine HIV-2 co-receptor tropism, and HIV-2 is known to use other minor co-receptors in addition to CCR5 and CXCR4. HIV-2 is intrinsically resistant to enfuvirtide and there are as yet no data on the anti-HIV-2 activity of ibalizumab.

Further Reading

- Aboud, M., Kaplan, R., Lombaard, J., *et al.*, 2019. Dolutegravir versus ritonavir-boosted lopinavir both with dual nucleoside reverse transcriptase inhibitor therapy in adults with HIV-1 infection in whom first-line therapy has failed (DAWNING): An open-label, non-inferiority, phase 3b trial. *Lancet Infectious Diseases* 19 (3), 253–264.
- Cahn, P., Madero, J.S., Arribas, J.R., *et al.*, 2019. Dolutegravir plus lamivudine versus dolutegravir plus tenofovir disoproxil fumarate and emtricitabine in antiretroviral-naïve adults with HIV-1 infection (GEMINI-1 and GEMINI-2): Week 48 results from two multicentre, double-blind, randomised, non-inferiority, phase 3 trials. *Lancet* 393 (10167), 143–155.

- Emu, B., Fessel, J., Schrader, S., *et al.*, 2018. Phase 3 study of ibalizumab for multidrug-resistant HIV-1. *The New England Journal of Medicine* 379 (7), 645–654.
- Gallant, J., Lazzarin, A., Mills, A., *et al.*, 2017. Bictegravir, emtricitabine, and tenofovir alafenamide versus dolutegravir, abacavir, and lamivudine for initial treatment of HIV-1 infection (GS-US-380-1489): A double-blind, multicentre, phase 3, randomised controlled non-inferiority trial. *Lancet* 390 (10107), 2063–2072.
- Gulick, R.M., Lalezari, J., Goodrich, J., *et al.*, 2008. Maraviroc for previously treated patients with R5 HIV-1 infection. *The New England Journal of Medicine* 359 (14), 1429–1441.
- Kozal, M., Aberg, J., Pialoux, G., *et al.*, 2020. Fostemsavir in adults with multidrug-resistant HIV-1 infection. *The New England Journal of Medicine* 382 (13), 1232–1243.
- Lennox, J.L., DeJesus, E., Lazzarin, A., *et al.*, 2009. Safety and efficacy of raltegravir-based versus efavirenz-based combination therapy in treatment-naïve patients with HIV-1 infection: A multicentre, double-blind randomised controlled trial. *Lancet* 374 (9692), 796–806.
- Matheron, S., Descamps, D., Gallien, S., *et al.*, 2018. First-line raltegravir/emtricitabine/tenofovir combination in human immunodeficiency virus type 2 (HIV-2) infection: A phase 2, noncomparative trial (ANRS 159 HIV-2). *Clinical Infectious Diseases* 67 (8), 1161–1167.
- Radzio-Basu, J., Council, O., Cong, M.E., *et al.*, 2019. Drug resistance emergence in macaques administered cabotegravir long-acting for pre-exposure prophylaxis during acute SHIV infection. *Nature Communications* 10 (1), 2005.
- Rhee, S.Y., Grant, P.M., Tzou, P.L., *et al.*, 2019. A systematic review of the genetic mechanisms of dolutegravir resistance. *Journal of Antimicrobial Chemotherapy* 74 (11), 3135–3149.
- Sax, P.E., Erlandson, K.M., Lake, J.E., *et al.*, 2019. Weight gain following initiation of antiretroviral therapy: Risk factors in randomized comparative clinical trials. *Clinical Infectious Diseases*.
- Swindells, S., Andrade-Villanueva, J.F., Richmond, G.J., *et al.*, 2020. Long-acting cabotegravir and rilpivirine for maintenance of HIV-1 suppression. *The New England Journal of Medicine* 382 (12), 1112–1123.
- Zash, R., Holmes, L., Diseko, M., *et al.*, 2019. Neural-tube defects and antiretroviral treatment regimens in Botswana. *The New England Journal of Medicine* 381 (9), 827–840.

Management of Respiratory Syncytial Virus Infections (*Pneumoviridae*)

Rachael S Barr, Bristol Royal Hospital for Children, Bristol, United Kingdom

Simon B Drysdale, St George's University Hospitals NHS Foundation Trust, London, United Kingdom and St George's, University of London, London, United Kingdom

© 2021 Elsevier Ltd. All rights reserved.

Introduction and Epidemiology

Respiratory syncytial virus (RSV) is a single-stranded RNA virus belonging to the *pneumoviridae* family (Genus: *orthopneumovirus*). It has two subtypes, RSV-A and RSV-B, both of which cause acute upper and lower respiratory tract infection. There is no clear association between RSV subtype and disease severity (Green Book, 2015). RSV is one of the leading causes of respiratory infection worldwide with almost all children having been infected by 2 years of age (Glezen *et al.*, 1986). In children under 5 years of age RSV is responsible for 22% of all severe acute lower respiratory tract infections worldwide (Shi *et al.*, 2017). In 2015 it was estimated that globally there were 33.1 million episodes of RSV associated acute lower respiratory tract infection in children under 5 years, resulting in 3.2 million hospital admissions and 59,600 deaths (Shi *et al.*, 2017). RSV also affects adult populations with a systematic review of articles from across the United States showing RSV to be the causative organism in 12% of acute respiratory illnesses reviewed by a medical professional (Colosia *et al.*, 2017). Among adults over the age of 65 years admitted to hospital with RSV, mortality is as high as 6%–8% (Colosia *et al.*, 2017; Falsey *et al.*, 2005) with a similar mortality rate of 8% found in hospitalized younger adults with chronic cardiac or respiratory disease (Falsey *et al.*, 2005). In the United States alone it is estimated that RSV is responsible for 170,000 hospitalizations and 10–17,000 deaths annually in the adult population (Nam and Ison, 2019; Falsey and Walsh, 2000).

RSV shows a seasonal pattern which in temperate climates such as the United Kingdom (UK) lies between October and March. The peak number of infections tends to occur within a six week period within this season each year (Green Book, 2015). In more tropical areas, RSV occurs year-round. In each season, one subtype (A or B) tends to predominate.

Pathogenesis and Immunity

RSV is transmitted between hosts via droplets or fomites. It initially infects the apical surface of ciliated epithelial cells in both the upper and lower airways via the fusion (F) and attachment (G) proteins. Although there is almost ubiquitous infection by 2 years of age, infection with RSV offers incomplete immunity (Glezen *et al.*, 1986). This results in episodes of repeated infection in both children and adults throughout life (Lambert *et al.*, 2014; Openshaw *et al.*, 2017; Gonzalez *et al.*, 2000). RSV often results in a self-limiting illness, however, there are both environmental and host factors that put individuals at higher risk of more severe clinical manifestations. Host-related risk factors in children include younger age at the start of the RSV season with those under 3 months old at greatest risk, prematurity, low birth weight, bronchopulmonary dysplasia or other underlying respiratory disease, congenital heart disease and immunosuppression (Aujard and Fauroux, 2002; Rossi *et al.*, 2007). Environmental risk factors for infants include having older siblings or living in a household with multiple other children and day care attendance (Rossi *et al.*, 2007; Simoes, 2003; Bulkow *et al.*, 2002). In older adults, underlying cardiopulmonary disease and poor functional status are associated with an increased risk of severe clinical disease (Walsh *et al.*, 2004; Duncan *et al.*, 2009).

In the initial weeks of life infants have relative protection against RSV due to transplacental transfer and transfer via breastmilk of protective maternal antibodies. However, these antibodies fall below levels that are protective by approximately three to five months of age (Openshaw *et al.*, 2017). In later childhood, antibody responses to recurrent infection are more brisk and effective than in infancy (Openshaw *et al.*, 2017). Elderly adults are also at higher risk of severe disease due to a combination of physical changes to the lungs over time alongside waning innate and adaptive immunity. Those with low serum neutralizing antibody titers against RSV appear to be at greatest risk of severe disease (Openshaw *et al.*, 2017).

Clinical Features

RSV has a range of clinical manifestations, some of which are present exclusively in certain age groups. In all age groups RSV can present as a mild self-limiting upper respiratory tract infection.

In children under the age of 1 year, RSV is the most common causative organism of bronchiolitis. Bronchiolitis is defined as inflammation of the small airways and presents with signs and symptoms including increased work of breathing, supplemental oxygen requirement, cough, fever and reduced feeding.

In children of pre-school age RSV is a causative organism of viral-induced wheeze. In this illness there is hyper-reactivity of the airways resulting in bronchospasm causing polyphonic wheeze. There may also be increased work of breathing, cough, fever and supplemental oxygen requirements.

In both children and adult populations RSV can also cause viral pneumonia/pneumonitis. The clinical features associated with this may include cough, increased work of breathing/shortness of breath, supplemental oxygen requirement and fever.

Among the adult population underlying cardiovascular disease is a significant risk factor for hospitalization with RSV. RSV in this population may present as an exacerbation of underlying cardiac disease, with 5.4% of all admissions for congestive heart failure during the RSV season thought to be attributable to RSV infection (Nam and Ison, 2019). Studies have also demonstrated an association in older adults between presence of RSV serum IgG and an increased risk of acute myocardial infarction (Ivey *et al.*, 2018; Chuaychoo *et al.*, 2019; Guan *et al.*, 2010).

Treatment Strategies

Treatment of RSV is mainly supportive and varies slightly depending on the clinical syndrome. Many of the management strategies laid out below are not specific to RSV but to the clinical manifestations which may equally be caused by a number of other respiratory viruses.

With regards to acute viral bronchiolitis in infants, the current NICE guideline states that treatment should be supportive with supplemental oxygen if saturations are persistently less than 92% and by giving fluids via naso/orogastric tube or intravenously if required (National Institute for Health and Care Excellence). Many other treatments have been trialed without success. NICE and the American Academy of Pediatrics (AAP) guidance currently recommend against the routine use of nebulized hypertonic saline, bronchodilators, antibiotics, corticosteroids (systemic or inhaled), nebulized adrenaline and montelukast for the management of acute bronchiolitis (National Institute for Health and Care Excellence; Ralston *et al.*, 2014).

Acute viral wheeze in pre-school age children may be caused by RSV and treatment remains mainly supportive. There is evidence to support the use of bronchodilators for symptomatic relief and their use is recommended in this age group by NICE guidance (NICE, 2017; Brand *et al.*, 2008). There is limited evidence for the use of oral corticosteroids in preschool acute viral wheeze. The current consensus from the European respiratory task force is that a trial should only be considered in children with wheeze of such severity that they require admission to hospital, and even in this group, evidence to support benefit is not robust (NICE, 2017; Brand *et al.*, 2008; Grigg, 2010; Brand *et al.*, 2014).

Other clinical syndromes including viral pneumonia, pneumonitis and upper respiratory tract infection in all age groups are also managed supportively with anti-pyretics, supplemental oxygen and feeding support as required.

One antiviral with activity against RSV has been licensed and is occasionally used in specific clinical scenarios. Ribavirin is a nucleoside inhibitor and is used as an anti-viral medication for a variety of infections, including RSV. A Cochrane review of 12 trials looking at use of ribavirin in infants and children with lower respiratory tract infection found that it may be effective in reducing days of mechanical ventilation in children requiring intensive care, however, the evidence base lacks sufficient power to draw firm conclusions on its efficacy (Ventre and Randolph, 2007). Ribavirin also has significant toxicity associated with it including hemolytic anemia and teratogenicity and so should be prescribed with careful attention to the risks and benefits. Its use is currently not recommended by NICE or the AAP for infants with RSV bronchiolitis (National Institute for Health and Care Excellence; Ralston *et al.*, 2014). Patients who have received haematopoietic stem cell transplant (both adults and children) are at particular risk of severe RSV infection and as such ribavirin has been used as a treatment option in these cohorts. Studies into the use of ribavirin in this cohort have mostly been small but are suggestive that it may be beneficial (Sparrelid *et al.*, 1997; McColl *et al.*, 1998; Boeckh *et al.*, 2007). Further randomised controlled trials are required in this field to establish a true effect. The second treatment that has been investigated is intravenous immunoglobulin (IVIG). A Cochrane review looking at immunoglobulin in treatment of RSV in young children found that IVIG was not effective at reducing mortality or time spent in hospital (Sanders *et al.*, 2019). As with ribavirin, there is minimal supporting evidence for the efficacy of IVIG in haematopoietic stem cell transplant recipients. In the 1990s IVIG was obtained from a pool of patients with high titers of RSV-neutralizing antibodies with the aim to create RSV specific IVIG. One product of this type known as RespiGam was licensed in 1996. This was shown to provide some passive immunity and reduce the rate of hospitalization with RSV by 41%–65% and length of stay in hospital by 53%–59% (Sandritter and Kraus, 1997). However, its manufacture was stopped in 2004 (Halsey *et al.*, 1998; Turner *et al.*, 2014). Small trials of patients (adults and children) undergoing haemopoietic stem cell transplant have also investigated combining ribavirin with IVIG. There seems to be a trend towards improved outcomes but more studies are needed. Some guidelines for this group of patients recommend ribavirin plus IVIG as treatment (Dignan *et al.*, 2016).

As discussed below, palivizumab (a monoclonal antibody against RSV) is an effective preventative agent for RSV in infants. It has also been trialed as a treatment in patients with severe LRTI or high-risk patients with URTI. However, a systematic review of palivizumab as treatment in both children and adults did not show any difference in clinical outcomes (Hu and Robinson, 2010). The review included one case report, four case series and two randomised controlled trials (RCTs). Both RCTs investigated pediatric patients only and neither were not adequately powered to assess clinical outcomes with palivizumab (Hu and Robinson, 2010). Palivizumab cannot currently be recommended as a treatment option for patients with RSV infection.

There are also a number of novel anti-viral therapies in development. They can be broadly divided into four classes; immunoglobulins, nucleoside analogs, small interfering RNA's and fusion inhibitors (Xing and Proesmans, 2019). None of these has progressed past a Phase 2 clinical study and thus a new licensed RSV antiviral remains several years away.

Prophylaxis

A key strategy in the control of RSV infection lies in prophylaxis as currently there is no effective vaccine and treatment is mainly supportive in nature. The only licensed preventative treatment for RSV is palivizumab. Palivizumab is a humanized monoclonal antibody which provides passive immunity by targeting the RSV F protein and preventing fusion with the host cell.

A randomised controlled trial of infants born at less than 35 weeks of gestation or infants with bronchopulmonary dysplasia showed that palivizumab reduced rates of hospitalization from RSV by 55% compared to placebo. Those who received palivizumab also had proportionally fewer total days in hospital, fewer days spent on supplemental oxygen and were less likely to be admitted to the intensive care unit as a result of RSV infection ([The IMPact-RSV Study Group, 1998](#)). Palivizumab has also been shown to have some benefits in otherwise healthy infants born at 33–35 weeks of gestation. In the MAKI trial, palivizumab prophylaxis resulted in a 67% relative reduction in RSV infections compared with placebo ([Blanken *et al.*, 2013](#)). It also resulted in a relative reduction of 61% in number of days of wheezing in the first year of life suggesting an improvement in chronic respiratory morbidity associated with RSV infection in infancy ([Blanken *et al.*, 2013](#)).

Although it has been shown to be clinically effective, the use of palivizumab remains controversial due to its high cost. The broadly accepted willingness to pay threshold set by the National Institute for Health and Care Excellence (NICE) in the UK is £30,000 per quality adjusted life year (QALY). An initial analysis of available evidence showed that palivizumab was not cost effective in a UK setting if delivered unselectively to all infants with an incremental cost-effectiveness ratio (ICER) of around £60,000 per QALY ([Wang *et al.*, 2008](#)). A second systematic review assessed cost effectiveness in specific subgroups of infants and found that palivizumab may be cost effective in infants with chronic lung disease and haemodynamically significant congenital heart disease at certain gestational and corrected gestational ages ([Wang *et al.*, 2011](#)). The current guideline for use of palivizumab in the UK has been set out by Public Health England upon guidance from the Joint Committee on Vaccination and Immunization (JCVI) in the Green Book ([Green Book, 2015](#)). In the infants for whom it is recommended it is delivered as monthly intramuscular injections at a dose of 15 mg/kg over the RSV season (October to March in the UK).

There are other monoclonal antibodies in development ([PATH.org](#)). MEDI8897 is a monoclonal antibody in development that has now entered phase 3 trials. Initial results suggest that it is safe to use in otherwise healthy preterm infants and it reduced medically attended RSV LRTI by 70.1% over the course of a year ([Pamela Griffin *et al.*, 2019](#)). It may have benefits over palivizumab as it has a longer half-life of 83–94 days and so may be able to be administered as a single dose to cover a 6 month period ([Domachowski *et al.*, 2018](#)).

RSV Vaccination

Since the 1960s there have been multiple trials looking for an effective RSV vaccine, without any achieving licensure. Proposed populations in whom an RSV vaccine would be beneficial include newborn infants, preschool children, pregnant women and older adults. A cost effectiveness review of different vaccination strategies to protect infants found that the most cost-effective strategy would be a seasonal approach tailored to each specific country, protecting neonates born from just before the RSV season until a few months of age ([Cromer *et al.*, 2017](#)). A vaccine could potentially be given to the mother prior to delivery or to a neonate in the first weeks of life ([Cromer *et al.*, 2017](#)). A snapshot of the landscape of RSV vaccine development as of August 2019 shows there is one vaccine currently in Phase 3 trials ([PATH.org](#)). This is an RSV-F nanoparticle vaccine targeting the maternal population ([PATH.org](#)). This trial recently reported headline results suggesting the vaccine is safe and reduced medically significant RSV LRTI by 39.4% (97.5% CI, –1% to 64%) over a 90 day follow up period ([Novavax, 2019](#)). Full results are yet to be published at the time of writing. There are multiple vaccines currently undergoing Phase 2 trials which between them target all of the potential population groups mentioned above ([PATH.org](#)). The global impact of an effective vaccine for RSV cannot be underestimated. Until that time, RSV will continue to have a significant burden of mortality and morbidity across all age groups but particularly in young children, the elderly and the immunocompromised.

Conclusion

RSV causes a significant global burden of disease in both pediatric and adult populations. It particularly affects infants, especially those born prematurely or with other health needs, and the elderly population. It can cause a wide range of clinical signs and symptoms and can result in a minor URTI through to respiratory failure and death. Although there are prophylactic and antiviral options, management is mainly supportive. A safe and effective vaccine is yet to be found, and until that time the ongoing search will remain a public health priority.

References

- Aujard, Y., Fauroux, B., 2002. Risk factors for severe respiratory syncytial virus infection in infants. *Respiratory Medicine* 96, S9.
- Blanken, M.O., Rovers, M.M., Molenaar, J.M., *et al.*, 2013. Respiratory syncytial virus and recurrent wheeze in healthy preterm infants. *New England Journal of Medicine* 368, 1791–1799.

- Boeckh, M., Englund, J., Li, Y., *et al.*, 2007. Randomized controlled multicenter trial of aerosolized ribavirin for respiratory syncytial virus upper respiratory tract infection in hematopoietic cell transplant recipients. *Clinical Infectious Diseases* 44, 245–249.
- Brand, P.L.P., Baraldi, E., Bisgaard, H., *et al.*, 2008. Definition, assessment and treatment of wheezing disorders in preschool children: An evidence-based approach. *European Respiratory Journal* 32, 1096–1110.
- Brand, P.L.P., Caudri, D., Eber, E., *et al.*, 2014. Classification and pharmacological treatment of preschool wheezing: Changes since 2008. Review Preschool Wheezing. *J* 43, 1172–1177.
- Bulkow, L.R., Singleton, R.J., Karron, R.A., Harrison, L.H., 2002. Risk factors for severe respiratory syncytial virus infection among Alaska Native children. *Pediatrics* 109, 210–216.
- Chuaychoo, B., Ngamwongwan, S., Kaewnaphan, B., *et al.*, 2019. Clinical manifestations and outcomes of respiratory syncytial virus infection in adult hospitalized patients. *Journal of Clinical Virology* 117, 103–108.
- Colosia, A.D., Yang, J., Hillson, E., *et al.*, 2017. The epidemiology of medically attended respiratory syncytial virus in older adults in the United States: A systematic review. *PLoS One* 12, e0182321.
- Cromer, D., van Hoek, A.J., Newall, A.T., Pollard, A.J., Jit, M., 2017. Burden of paediatric respiratory syncytial virus disease and potential effect of different immunisation strategies: a modelling and cost-effectiveness analysis for England. *Lancet Public Health* 2, e367–e374.
- Dignan, F.L., Clark, A., Aitken, C., *et al.*, 2016. BCSH/BSBMT/UK clinical virology network guideline: Diagnosis and management of common respiratory viral infections in patients undergoing treatment for haematological malignancies or stem cell transplantation. *British Journal of Haematology* 173, 380–393.
- Domachowski, J.B., Khan, A.A., Jensen, K., *et al.*, 2018. A single dose monoclonal antibody immunoprophylaxis strategy to prevent respiratory syncytial virus disease in all infants: Results of the first in infant study with medi8897. *Pediatrics* 141, 256.
- Duncan, C.B., Walsh, E.E., Peterson, D.R., Lee, F.E., Falsey, A.R., 2009. Risk factors for respiratory failure associated with respiratory syncytial virus infection in adults. *Journal of Infectious Diseases* 200, 1242–1246.
- Falsey, A.R., Hennessey, P.A., Formica, M.A., Cox, C., Walsh, E.E., 2005. Respiratory syncytial virus infection in elderly and high-risk adults. *New England Journal of Medicine* 352, 1749–1759.
- Falsey, A.R., Walsh, E.E., 2000. Respiratory syncytial virus infection in adults. *Clinical Microbiology Reviews* 13, 371–384.
- Glezen, W.P., Taber, L.H., Frank, A.L., Kasel, J.A., 1986. Risk of primary infection and reinfection with respiratory syncytial virus. *American Journal of Diseases of Children* 140, 543–546.
- Gonzalez, I.M., Karron, R.A., Eichelberger, M., *et al.*, 2000. Evaluation of the live attenuated cpts 248/404 RSV vaccine in combination with a subunit RSV vaccine (PPF-2) in healthy young and older adults. *Vaccine* 18, 1763–1772.
- Green Book, 2015. Chapter 27a – Respiratory syncytial virus.
- Grigg, J., 2010. Role of systemic steroids in acute preschool wheeze. *Archives of Disease in Childhood* 95. doi:10.1136/adc.2009.160994.
- Guan, X.R., Jiang, L.X., Ma, X.H., *et al.*, 2010. Respiratory syncytial virus infection and risk of acute myocardial infarction. *American Journal of the Medical Sciences* 340, 356–359.
- Halsey, N.A., Abramson, J.S., Chesney, P.J., *et al.*, 1998. Prevention of respiratory syncytial virus infections: Indications for the use of palivizumab and update on the use of RSV-IGIV. *Pediatrics* 102, 1211–1216.
- Hu, J., Robinson, J.L., 2010. Treatment of respiratory syncytial virus with palivizumab: A systematic review. *World Journal of Pediatrics* 6, 296–300.
- Ivey, K.S., Edwards, K.M., Talbot, H.K., 2018. Respiratory syncytial virus and associations with cardiovascular disease in adults. *Journal of the American College of Cardiology* 71, 1574–1583.
- Lambert, L., Sagfors, A.M., Openshaw, P.J.M., Culley, F.J., 2014. Immunity to RSV in early-life. *Frontiers in Immunology* 5. doi:10.3389/fimmu.2014.00466.
- McColl, M.D., Corser, R.B., Bremner, J., Chopra, R., 1998. Respiratory syncytial virus infection in adult BMT recipients: Effective therapy with short duration nebulised ribavirin. *Bone Marrow Transplantation* 21, 423–425.
- Nam, H.H., Ison, M.G., 2019. Respiratory syncytial virus infection in adults. *BMJ* 10 (366), i5021.
- National Institute for Health and Care Excellence. Bronchiolitis in children: Diagnosis and management. NICE Guideline [NG9].
- NICE, 2017. Cough – Acute with chest signs in children. NICE CKS. Available at: <https://cks.nice.org.uk/cough-acute-with-chest-signs-in-children#!scenario>. (accessed 09.12.18).
- Novavax, 2019. New Data from Novavax Phase 3 Prepare™ Trial of ResVax™ Presented at 2019 IDSOG Annual Meeting. Novavax Inc. Available at: <https://ir.novavax.com/news-releases/news-release-details/new-data-novavax-phase-3-prepare-trial-resvaxtm-presented-0>. (accessed 17.02.20).
- Openshaw, P.J.M., Chiu, C., Culley, F.J., Johansson, C., 2017. Protective and harmful immunity to RSV infection. *Annual Review of Immunology* 35, 501–532.
- Pamela Griffin, M., Yuan, Y., Takas, T., *et al.*, 2019. MEDI8897 prevents serious RSV disease in healthy preterm infants. *Open Forum Infectious Diseases* 6, S27.
- PATH.org. RSV Vaccine and mAb Snapshot. Available at: https://path.azureedge.net/media/documents/RSV-snapshot-2019_08_28_High_Resolution_PDF.pdf (accessed 16.01.20).
- Ralston, S.L., Lieberthal, A.S., Meissner, H.C., *et al.*, 2014. Clinical practice guideline: The diagnosis, management, and prevention of bronchiolitis. *Pediatrics* 134, e1474–e1502.
- Rossi, G.A., Medici, M.C., Arcangeletti, M.C., *et al.*, 2007. Risk factors for severe RSV-induced lower respiratory tract infection over four consecutive epidemics. *European Journal of Pediatrics* 166, 1267–1272.
- Sanders, S.L., Agwan, S., Hassan, M., van Driel, M.L., Del Mar, C.B., 2019. Immunoglobulin treatment for hospitalised infants and young children with respiratory syncytial virus infection. *Cochrane Database of Systematic Reviews* 8. doi:10.1002/14651858.cd009417.pub2.
- Sandritter, T.L., Kraus, D.M., 1997. Respiratory syncytial virus-immunoglobulin intravenous (RSV-IGIV) for respiratory syncytial viral infections: Part I. *Journal of Pediatric Health Care* 11, 284–291.
- Shi, T., McAllister, D.A., O'Brien, K.L., *et al.*, 2017. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet* 390, 946–958.
- Simoes, E.A.F., 2003. Environmental and demographic risk factors for respiratory syncytial virus lower respiratory tract disease. *Journal of Pediatrics* 143. doi:10.1067/s0022-3476(03)00511-0.
- Sparrelid, E., Ljungman, P., Ekelöf-Andström, E., *et al.*, 1997. Ribavirin therapy in bone marrow transplant recipients with viral respiratory tract infections. *Bone Marrow Transplantation* 19, 905–908.
- The Impact-RSV Study Group, 1998. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. *Pediatrics* 102. doi:10.1542/peds.99.4.645.
- Turner, T.L., Kopp, B.T., Paul, G., *et al.*, 2014. Respiratory syncytial virus: Current and emerging treatment options. *ClinicoEconomics and Outcomes Research* 6, 217–225.
- Ventre, K., Randolph, A., 2007. Ribavirin for respiratory syncytial virus infection of the lower respiratory tract in infants and young children. In: Ventre, K. (Ed.), *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd. doi:10.1002/14651858.CD000181.pub3.
- Walsh, E.E., Peterson, D.R., Falsey, A.R., 2004. Risk factors for severe respiratory syncytial virus infection in elderly persons. *Journal of Infectious Diseases* 189, 233–238.
- Wang, D., Bayliss, S., Meads, C., 2011. Palivizumab for immunoprophylaxis of respiratory syncytial virus (RSV) bronchiolitis in high-risk infants and young children: A systematic review and additional economic modelling of subgroup analyses. *Health Technology Assessment* 15, iii.
- Wang, D., Cummins, C., Bayliss, S., Sandercock, J., Burl, A., 2008. Immunoprophylaxis against respiratory syncytial virus (RSV) with palivizumab in children: A systematic review and economic evaluation. *Health Technology Assessment* 12. doi:10.3310/hta12360.

Xing, Y., Proesmans, M., 2019. New therapies for acute RSV infections: Where are we? *European Journal of Pediatrics* 178, 131–138.

Further Reading

Drysdale, S.B., Kelly, D.F., 2019. How to use...respiratory viral studies. *Archives of Disease in Childhood: Education and Practice Edition* 104 (5), 274–278.

Mazur, N.I., Higgins, D., Nunes, M.C., *et al.*, 2018. The respiratory syncytial virus vaccine landscape: Lessons from the graveyard and promising candidates. *Lancet Infectious Diseases* 18 (10), e295–e311.

Management of Influenza Virus Infections (*Orthomyxoviridae*)

Bruno Lina, HCL Department of Virology, National Reference Center for Respiratory Viruses, Institute of Infectious Agents, Croix-Rousse Hospital, Lyon, France and Virpath Laboratory, International Center of Research in Infectiology (CIRI), INSERM U1111, CNRS—UMR 5308, École Normale Supérieure de Lyon, University Claude Bernard Lyon, Lyon University, Lyon, France

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ECMO Extracorporeal membrane oxygenation

HA Hemagglutinin

M2 Matrix protein 2

NA Neuraminidase

NAI Neuraminidase inhibitors

Pol Polymerase

Introduction

The burden of disease related to the annual influenza epidemic is significant in all regions of the world (Wright *et al.*, 2007; Iuliano *et al.*, 2018; Federici *et al.*, 2018; Molinari *et al.*, 2007). Recent estimates suggest that 5%–10% of the population in the western world experience clinical influenza every year, resulting in medical care expenses and burden that have a major impact on public hospitals. Influenza activity observed during pandemics may also lead to increased medical care in both community-based and hospital-based health structures. As a consequence, shortage of available hospital beds (especially in the emergency rooms) and hospitals become overwhelmed by severely ill individuals requiring intensive medical care because of influenza-related complications in both seasonal and pandemic influenza (Wright *et al.*, 2007; Molinari *et al.*, 2007). Although less documented, this impact is also massive in resource-poor and countries (De Francisco Shapovalova *et al.*, 2015).

To avoid these crisis situations, annual or pandemic vaccination is promoted. However, the relatively low vaccine effectiveness observed with seasonal influenza vaccines, especially for A(H3N2) viruses, as well as the anticipated long delays before the availability of a pandemic vaccine makes the case for better use of currently available antiviral drugs and further antiviral development (McLean *et al.*, 2016). This need is also stressed in pandemic preparedness plans where the only response in the early stages of a pandemic will be the use of antivirals, especially for severe infections (Monto and Fukuda, 2020; Whitley and Monto, 2006).

Seasonal Influenza

Annually, seasonal influenza accounts for about 650,000 deaths, together with millions of severe cases and hospital admissions (Wright *et al.*, 2007). This influenza-related morbidity and mortality is mainly observed in patients with risk factors such as old age, medical comorbidities (pulmonary disease, cardiovascular disease, renal disease, liver disease, diabetes, neuromuscular disease, and cognitive dysfunction), immunocompromised status (stem cell transplant, solid organ transplant, chemotherapy, HIV infection), pregnancy, obesity, and genetic susceptibility (i.e., IFITM3 defects) (Rothberg and Haessler, 2010; Wedzhicha and Seemungal, 2007; Smeeth *et al.*, 2004; Allen *et al.*, 2017). In these patients, symptomatic treatments are often not sufficient to avoid complications and increased risk of morbidity and mortality.

Pandemic Influenza

Influenza A viruses have been responsible for four pandemics in the last 100 years (Monto and Fukuda, 2020; Taubenberger *et al.*, 2006; Spreeuwenberg *et al.*, 2018). During these pandemics, additional risk-groups have emerged (i.e., young adults aged 20–40 years in 1918; children younger than two years of age in 1957). In addition, pregnancy has clearly been recognized with a high risk of morbidity, especially in the third trimester during the 1918, 1957, and 2009 pandemics.

Zoonotic Influenza

Influenza A is a zoonotic virus that can occasionally be responsible for human infections (Peiris, 2009; Beigel *et al.*, 2005; Koopmans *et al.*, 2004; Peiris *et al.*, 1999). These zoonotic viruses are acquired through exposure to infected livestock (swine, chicken, ducks, and turkeys). In the last 40 years, an increased number of zoonotic influenza cases have been reported, with an occasional very high mortality rate. As these infections can be a pandemic threat, the patients should be handled carefully, and their infection controlled rapidly to avoid the implementation of human to human transmission. The disease management of these cases is also a driving force to develop antivirals.

To address all these situations, influenza antivirals have been developed since the 1960s (Wright *et al.*, 2007; Palese and Shaw, 2007; Aoki, 1998; Colman, 1994). Despite showing some efficacy, these products had two major limitations (Wright *et al.*, 2007): a marginal impact on the clinical disease duration in mild or non-severe cases, and (Iuliano *et al.*, 2018) a rapid emergence of resistant viruses, especially in children (Palese and Shaw, 2007; Aoki, 1998; Colman, 1994; Muthuri *et al.*, 2013). However, the clinical benefit in severe

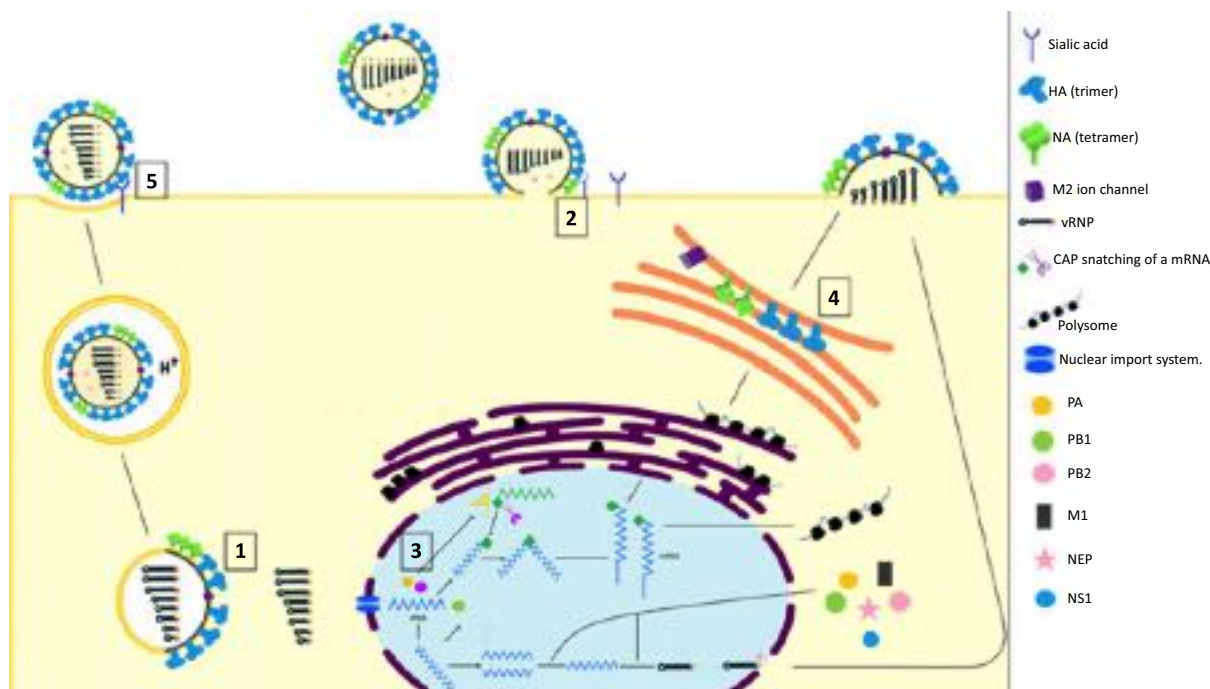


Fig. 1 Schematic representation of the Influenza replication in the cell and the different targets for antiviral action of molecules currently available or under development. The numbering is describing where the different classes of antiviral may interact with the virus replication cycle in the cell. 1 – inhibition of H^+ -mediated decapsulation of vRNPs blocked by M2 ion channel blockers 2 – inhibition of virus release from the cell surface after budding by Neuraminidase inhibitors 3 – inhibition of virus replication and RNA synthesis by polymerase inhibitors 4 – inhibition of HA processing and maturation by blocking HA terminal glycosylation required for HA maturation 5 – inhibition of HA attachment to epithelial cells by removal of sialic acids by a bioavailable sialidase.

cases has been assessed, and antiviral usage has been promoted during the two last pandemics (1968 and 2009) (Aoki, 1998; Colman, 1994; Muthuri *et al.*, 2013). In 1968–1969, Amantadine was positively evaluated and subsequently proposed for the disease management of severe cases, and in 2009, Oseltamivir was also proposed, for the management of both severe and clustered cases.

These two antiviral drugs target two virus proteins, M2 (amantadine) and NA (oseltamivir), and aim for rapid virus clearance through impaired virus replication. However, influenza disease is not only the consequence of virus replication. The immune response that develops to combat the infection is responsible for many of the clinical symptoms that are still present after virus clearance (Wright *et al.*, 2007; Palese and Shaw, 2007). This partly explains why the current antivirals, although “virologically effective”, have limitations in the reduction of disease duration in mild cases. This has led to the development of new concepts for influenza infection treatment, with molecules targeting the host response.

In this article, we will go through the different classes of currently available antivirals and complete this list of recently developed products, some of which are alternative strategies involving host targeted molecules and immunotherapies.

Antiviral Drugs Available

Adamantanes or M2 Blockers (Fig. 1)

Adamantanes were the first class of antivirals developed and tested against Influenza during the 1960s (Palese and Shaw, 2007; Aoki, 1998). These molecules are active against influenza A viruses only. Their mode of action involves blocking the proton channel activity of the M2 complex, a function required for the acidification of the virus that allows viral uncoating and disruption of the macromolecular vRNP complex (Fig. 1) (Palese and Shaw, 2007). This disruption results in the release of the vRNP in the cytoplasm before entry into the nucleus of the infected cell. By direct interaction with the domains involved in the ion-channel activity of the M2 complex, adamantanes abort viral replication (Palese and Shaw, 2007; Aoki, 1998).

Amantadine, the first molecule of this class, showed a sub-optimal safety profile (neurological disorders) and, as a consequence, a derivative (rimantadine) with an improved safety profile was developed (Wright *et al.*, 2007; Aoki, 1998).

These adamantanes were first evaluated and used during the 1968 pandemic with some success; clinical trials have confirmed a superiority of amantadine to placebo (Wright *et al.*, 2007; Aoki, 1998) in disease management. However, they remained poorly used because of the frequent and sometimes severe side effects. In addition, it was observed that resistance could emerge rapidly, related to substitutions in the membrane-spanning region of the M2, especially at amino acids 26, 27, 30, 31, and 34 (Fig. 2)

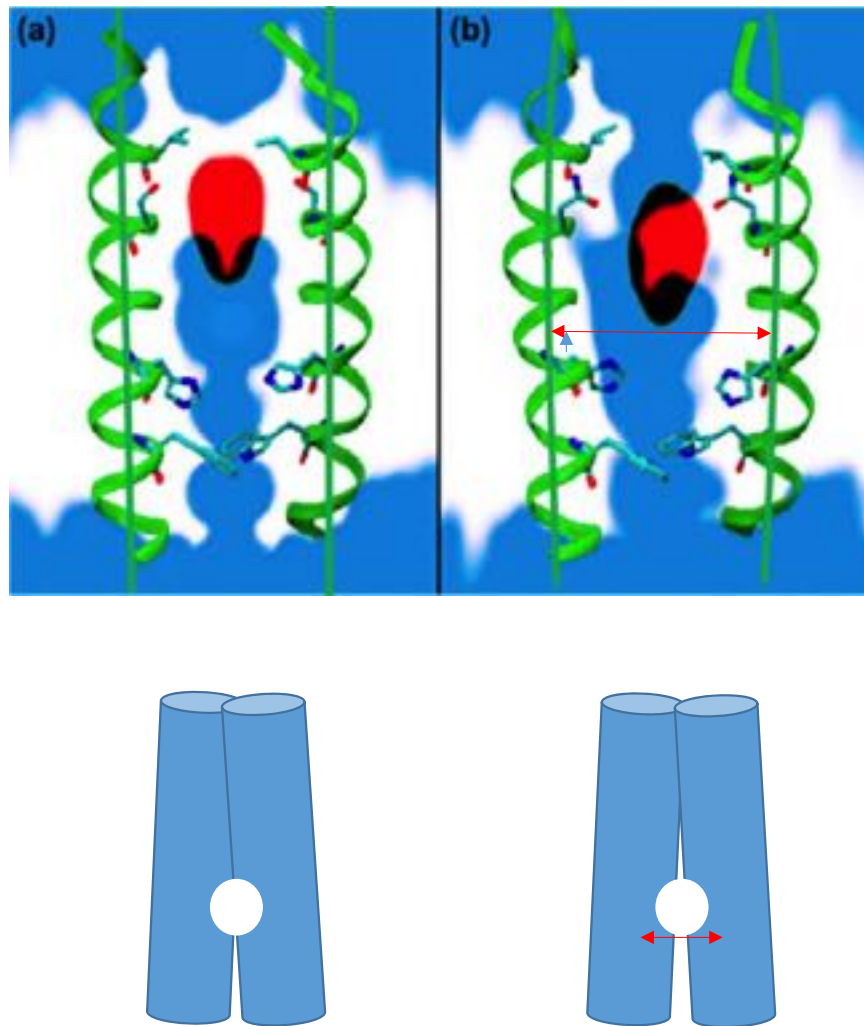


Fig. 2 substitutions associated with resistance to Adamantanes in the ion channel of the M2 tetramer of Influenza A viruses. The alpha helix of M2 proteins with S31 substitution has a curvature that allows the ion channel to be efficient despite binding the M2 blocker in its active site (A). As a consequence of this substitution, the size of the channel is increased (B) as depicted in the cartoon. The Adamantane is in red in the cartoon.

(Wright *et al.*, 2007; Palese and Shaw, 2007; Hay *et al.*, 1986; Weinstock and Zuccotti, 2006). This resistance induced no fitness cost and was sustained in the virus progeny. Hence, the resistant viruses were positively selected in a context of antiviral pressure and were readily transmitted to contacts (Hayden *et al.*, 1991; Shiraishi *et al.*, 2003; Schilling *et al.*, 2004).

These products remained in use in several countries until the end of the 20th century. However, because of the frequency of the side effects observed and the increasing percentage of influenza A viruses resistant to adamantanes (almost 100% by 2010), these molecules are not recommended anymore (Bright *et al.*, 2005; Deyde *et al.*, 2007; Barr *et al.*, 2007).

Neuraminidase Inhibitors (Fig. 1)

The second class of antivirals developed were the neuraminidase inhibitors (NAI). The neuraminidase (NA), a glycoprotein organized as homotetramers accounting for approximately 20% of the glycoproteins of the virus surface, has an enzymatic activity (sialidase) required to cleave the linkage between the terminal sialic acid and the adjacent D-galactose (or D-galactosamine) of the cellular glycoproteins (Palese and Shaw, 2007; Aoki, 1998). This cleavage allows the release of the virus progeny after budding and facilitates the transport of the virus to the epithelial cells through the respiratory mucus containing sialylated proteins (Palese and Shaw, 2007; Aoki, 1998).

Despite subtle differences, the structure of the catalytic pockets of influenza A and B NAs are conserved for all viruses (Fig. 3). In addition, preliminary studies have shown that viruses with impaired NA function because of amino acid substitutions altering the sialidase activity (i.e., temperature-sensitive mutants) remained stacked at the surface of the cell (Palese and Shaw, 2007). As a consequence, these viruses cannot be released, and have a very low level of transmission.

The identification of the key role of the NA in virus dissemination has led to the development of high-affinity sialic acid analogs with sialidase-inhibiting activity. The first available and tested NA inhibitor (NAI) was Zanamivir (Relanza[®], Fig. 4). This molecule

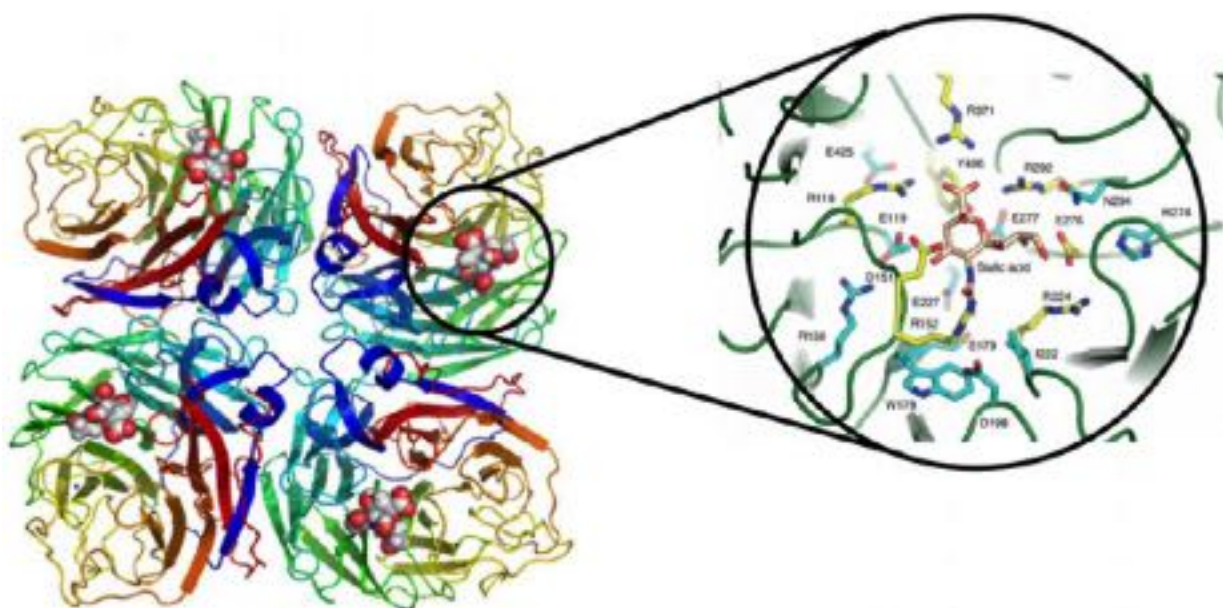


Fig. 3 Neuraminidase pocket site. The neuraminidase is a tetramer with 4 active sites highly conserved within all NAs. Substitutions associated with resistance are observed either in the catalytic site or in the framework structure. The latter is changing the shape of the catalytic pocket, resulting in a reduced binding of the neuraminidase inhibitors.

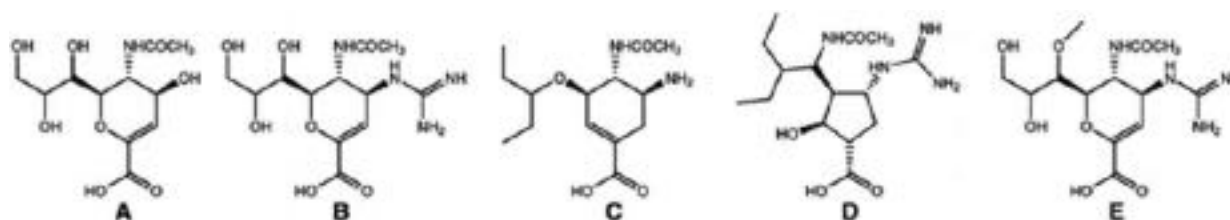


Fig. 4 Sialic acids and analogs acting as neuraminidase inhibitors. Chemical structures of the neuraminidase inhibitors (A) DANA, (B) zanamivir, (C) oseltamivir carboxylate, (D) peramivir, and (E) laninamivir. Structures show the differences relative to DANA – the C4-guanidinium group on zanamivir, peramivir, and laninamivir, the C4-amino group on oseltamivir, and the pentyl side chains on oseltamivir and peramivir. From McKimm-Breschkin, 2013. Influenza neuraminidase inhibitors: Antiviral action and mechanisms of resistance. Influenza Other Respiratory Viruses (Suppl 1), 25–36.

has a very low oral bioavailability and has to be inhaled to be active (Wright *et al.*, 2007). As a consequence, specific inhaling devices have been developed to allow the administration of therapeutic doses of the product. In the clinical studies carried out in patients presenting with mild influenza, Zanamivir demonstrated a clinical benefit with a 24 h reduction of symptom duration compared to placebo, if the drug was administered within 48 h after symptom onset (Gubareva *et al.*, 2000). In addition, it was demonstrated that, as expected, virus shedding was significantly reduced. The complexity of the delivering process (inhalation) and the modest clinical benefit observed resulted in very limited use. A second NAI was introduced, Oseltamivir carboxylate (Tamiflu®), a prodrug that could be administered orally (Fig. 4) (Wright *et al.*, 2007). Virological and clinical studies showed similar effects to Zanamivir (similar reduction of virus shedding and similar reduction in clinical symptoms duration) (Gubareva *et al.*, 2000). Later, two additional NAIs were developed, evaluated, and made available: Peramivir (Rapivab®, IV or IM administration, one single dose) and Laninamivir octanoate (Inavir®, inhaled, one single dose) (Fig. 4) (Mancuso *et al.*, 2010; Ikematsu, 2011). The clinical benefits of all these NAI are very similar: A reduction of disease duration by one day.

Interestingly, at the early stage of the development of these molecules and based on the preliminary studies conducted on temperature-sensitive mutants (Palese and Shaw, 2007), the risk of emergence of resistance was considered to be very low. In vitro and in vivo studies confirmed a reduction in virus fitness when NA substitutions associated with resistance occurred (Yen *et al.*, 2005; Herlocher *et al.*, 2002; Burnham *et al.*, 2014; Ives *et al.*, 2002). Meanwhile, rare sporadic cases of resistant type A or B influenza viruses with NA harboring substitutions associated with reduced susceptibility were reported (Ferraris *et al.*, 2005; Monto *et al.*, 2006). However, in 2007, a naturally-occurring resistant clone of seasonal influenza emerged with excellent fitness. This resistant clone completely displaced the susceptible parental strain, as a result of combined evolution in NA and HA (epistasis) (Wright *et al.*, 2007; Hurt *et al.*, 2010; Meijer *et al.*, 2009; Hurt *et al.*, 2009; Dharan *et al.*, 2009).

Subsequently, surveillance for resistant viruses was reinforced, a better definition of the level of resistance (or reduced susceptibility) was proposed, and phenotypic assays were progressively implemented (chemiluminescent and fluorescent

assays) (Palese and Shaw, 2007; Ferraris *et al.*, 2005; Nguyen *et al.*, 2012; Okomo-Adhiambo *et al.*, 2014). As a consequence, a normalization was proposed in terms of testing and reporting. Briefly, the NAI susceptibility testing works through the determination of the concentration of drug required to inhibit NA enzymatic activity by 50% (measurement of an IC50 value); the level of NA inhibition being interpreted as a surrogate for virus inhibition. To facilitate the interpretation of the results, IC50 value thresholds (or cut-off) were used to separate drug-resistant and drug-sensitive viruses. To facilitate the interpretation and reporting of NAI assay results, a WHO Expert Working Group introduced a very strict set of criteria to define the antiviral susceptibility of viruses based on the fold change of IC50 values, as compared to a reference IC50 (e.g., a median IC50 of viruses of the same type/subtype) (Table 1). Interpretation of NAI assay data is type-specific (Okomo-Adhiambo *et al.*, 2014).

This normalized assay has been associated with molecular investigations of viruses showing RI and HRI phenotypes. The substitutions/deletions associated with these phenotypes showed two kinds of changes: within or near the NA catalytic site affecting the binding of one or more NAIs (catalytic substitutions), or responsible for a change in the global folding of the NA, with no changes in the catalytic pocket (structural substitutions) (Ferraris and Lina, 2008; McKimm-Breschkin, 2013). A comprehensive list of substitutions has been prepared, showing differences between the different NAs.

Recent surveillance data showed that there is only a very limited number of resistant viruses reported worldwide, mainly in H1N1pdm09 viruses and in children (Lackenby *et al.*, 2018). Individual cases are described, including in immunocompromised patients where chronic infections have been reported. There is no information about the sustained transmission of resistant viruses. A five-year-long study reporting on the resistance to NAI in treated patients showed that Oseltamivir resistance was only detected during antiviral treatment, with the highest incidence occurring among one to five-year-olds. Resistance delayed viral clearance but had no impact on symptom resolution (Lina *et al.*, 2018).

Post-licensure studies have been performed to consolidate the potential clinical benefit of the NAI usage in mild and severe influenza. Observational studies have reported the beneficial effects of NAIs, showing – in adult patients with confirmed influenza – a reduction in the likelihood of (Wright *et al.*, 2007) requiring antibiotics because of lower respiratory tract infection (44%) and (Iuliano *et al.*, 2018) being hospitalized (63%) (Katzen *et al.*, 2019; Venkatesan *et al.*, 2017). Similarly, studies conducted during the 2009 pandemic suggested that, when initiated within 48 h of symptom onset in hospitalized patients, NAIs significantly reduced mortality in adults by 62% – this remains significant even when the treatment was started after 48 h (25% overall reduction) (Muthuri *et al.*, 2014).

Post-licensure surveillance has also pointed out some very rare adverse effects of NAI, particularly of Oseltamivir. Most minor side effects associated with NAI consumption were gastrointestinal disorders such as nausea and vomiting, but not diarrhea (Doll *et al.*, 2017). Both stop rapidly after treatment cessation. In Japan, some cases of neuropsychiatric adverse events (NPAE) have been reported to be linked with Oseltamivir or Laninamivir use in adolescents (Nakano *et al.*, 2013). An in-depth analysis carried out in 2008 compared the incidence of cases in Japan and the USA, showing a higher incidence of NPAE reporting rates per 1,000,000 prescriptions in children (aged < 16 years) and adults (respectively, 99 and 28 events in Japan and 19 and 8 in the US). This study concluded that there was no link between NPAEs and Oseltamivir usage; the NPAE incidence being equivalent in influenza treated and non-treated cases (Toovey *et al.*, 2008). The mechanism for the development of NPAE remains to be elucidated. Inhaled Zanamivir is contraindicated if there is a risk of bronchospasm, even if remains difficult to quantify (Heneghan *et al.*, 2014).

In severe cases as well as in immunocompromised patients, different regimens of NAIs have been tested to increase their beneficial effect. Double or triple doses have been evaluated, showing no additional clinical benefit compared to the standard dose (Marty *et al.*, 2017). These regimens induced more severe and more frequent side effects, including gastrointestinal problems. Secondly, combinations of NAIs have also been tested (Oseltamivir plus Zanamivir). These studies did not report any clinical benefit as compared to Oseltamivir alone, or Zanamivir alone. They even reported a possible antagonistic effect with an overall reduced effectiveness (Escuret *et al.*, 2012; Duval *et al.*, 2010).

Post-exposure prophylaxis (PEP) has also been proposed for NAIs (Welliver *et al.*, 2001). Oseltamivir has been approved for this indication, with a seven-day course with a half-dose regimen. A meta-analysis calculated that Oseltamivir PEP had an efficacy of 58.5% (15.6% to 79.6) for households and 68% (34.9%–84.2%) to 89% in contacts of index cases; Zanamivir had a similar performance (Oh *et al.*, 2014). This protection has also been demonstrated in nursing homes (Welliver *et al.*, 2001). However, this half-dose regimen may trigger the emergence of a resistant virus (Wright *et al.*, 2007). Actually, a population dynamical model of pandemic influenza demonstrated that limited post-exposure prophylaxis can provide an optimal strategy in reducing the final size of the pandemic while minimizing the total number of deaths. However, post-exposure prophylaxis of close contacts in the presence of transmissible drug resistance can promote the spread of resistance, especially when combined with aggressive treatment. PEP should be used in the short term, mostly to control epidemics in closed settings (i.e., nursing homes).

Laninamivir has also been favorably evaluated as PEP in nursing homes (Kashiwagi *et al.*, 2016).

Polymerase Inhibitors (Fig. 1)

The third class of antivirals is the polymerase (POL) inhibitors. The development of these molecules has been facilitated by the determination of the crystal structure of the Pol complex and subunits ((Pflug *et al.*, 2014; Pflug *et al.*, 2018) Fig. 1). Different modes of action are used by the different POL inhibiting drugs (Fig. 5) (Mifsud *et al.*, 2019).

Table 1 Summary of the substitution association to reduced or highly reduced inhibition of Neuraminidase activity for seasonal human viruses type A (Table 1A) and type B (Table 1B). These tables are providing the a priori interpretation of the susceptibility to NAI when a specific substitution is observed in the neuraminidase of the tested virus. NI = normal inhibition; RI = reduced inhibition; HRI = highly reduced inhibition. (Ref: http://www.who.int/influenza/gisrs_laboratory/antiviral_susceptibility/avwg2014_nai_substitution_table.pdf)

A Subtype	Amino acid substitution	N2 numbering	Susceptibility assessed by NA inhibition assays (IC50 fold Change vs Wild type [NAI susceptible virus])			
			Oseltamivir	Zanamivir	Peramivir	Laninamivir
A(H1N1)pdm09						
	E119G	119	NI(3)	HRI (832)	RI (51)	?
	E119V	119	RI (60)	HRI (571)	RI (25)	?
	D199G	199	RI (17)	NI (6)	NI(2)	?
	I223K	222	RI (12–39)	NI (5–6)	NI(1–4)	?
	I223R	222	RI (28–45)	RI (10–12)	?	?
	I223V	222	NI (6)	NI (2)	NI (2)	?
	S247N	246	NI (4–8)	NI (2–5)	NI (1)	?
	H275Y	274	HRI (221–1637)	NI (2–6)	RI/HRI (50–751)	NI (1–2)
	N295S	294	HRI (208)	NI (3)	RI (12)	?
	D199N + H275Y	198 + 274	HRI (318–744)	NI (2–3)	HRI (108)	?
	I223K + H275Y	222 + 274	HRI (> 10,822)	RI (11)	HRI (5,343)	RI (11)
	I223R + H275Y	222 + 274	HRI (> 9,100)	RI (22–27)	HRI (>7500)	RI (17)
	I223V + H275Y	222 + 275	HRI (1733)	NI (2)	HRI (1331)	?
	S247N + H275Y	246 + 274	HRI (5880)	NI (5)	HRI (334)	?
	Q313R + I427T	313 + 427	RI (10–43)	NI/RI (3–20)	NI (4)	?
A(H3N2)						
	E119D	119	NI (2)	RI (32)	NI (2)	?
	E119I	119	HRI (208)	RI (17)	NI (3)	?
	E119V	119	RI/HRI (18–2057)	NI (1–7)	NI (1–3)	NI (3–4)
	Q136K	136	NI (1–7)	RI/HRI (30–132)	?	?
	D151E	151	RI (11)	NI (2)	?	?
	D151V/D	151	NI (8)	HRI (164)	?	?
	I222L	222	NI (9)	NI (2)	?	?
	R224K	224	HRI (> 4,000)	RI (>50)	?	?
	Del 245–248	245–248	HRI (157–175)	NI (3)	NI (1)	?
	E276D	276	RI (15)	HRI (160)	?	?
	R292K	292	HRI (> 10,000)	NI/RI/HRI (3–134)	RI/HRI (14–719)	?
	N294S	294	HRI (300–1879)	NI (8)	NI (1)	?
	R371K	371	RI (45)	RI (15)	?	?
	E119V + T148I	119 + 148	HRI (>6,000)	HRI (>800)	HRI (>110)	HRI (>205)
	E119V + I222L	119 + 222	HRI (1,571)	RI (5)	?	?
	E119V + I222V	119 + 222	HRI (293–2,286)	NI (2)	NI (7)	?
	E105K	110	NI(4–8)	RI (15–42)	HRI (213–436)	NI/RI (3–12)
	E117A	119	RI/HRI (3,171)	RI/HRI (12,538)	HRI (13,700)	?
	E117A or D	119	HRI (>300)	HRI (>560)	HRI (>1,598)	?
	E117G	119	RI (31)	RI/HRI (33–560)	HRI (>1,598)	?
	E117V	119	HRI (300)	NI (2)	HRI (531)	?
	Q138R	140	NI (1)	RI (15)	RI (41)	RI (7)
	P139S	141	RI (10)	RI (25)	HRI (322)	RI (6)
	G140R	142	RI (9)	RI (30)	HRI (1321)	RI (9)
	R150K	152	HRI (100–252)	NI/RI/HRI (5–1,000)	HRI (214–400)	?
	D197E	198	RI (12–26)	NI (3–7)	RI (16–18)	?
	D197N	198	NI/RI (4–10)	NI/RI (3–17)	NI (5)	?
	D197Y	198	RI/HRI (15–57)	RI (14)	HRI (168)	?
	A200A/T	201	RI (5–8)	RI (5–7)	?	?
	I221L	222	NI/RI (5–8)	NI (2–4)	?	?
	I221T	222	NI/RI (5–8)	NI (5)	RI (15–43)	NI (<3)
	I221V/L	222	NI (3–6)	NI (1)	NI (4–8)	?
	A245T	246	RI (24)	RI (39)	RI (5)	NI (3)
	H273Y	274	NI/RI (2–12)	NI (1)	RI/HRI (15–110)	?
	R292K	292	HRI (> 300)	RI (29)	HRI (502)	?

(Continued)

Table 1 Continued

A Subtype	Amino acid substitution	N2 numbering	Susceptibility assessed by NA inhibition assays (IC50 fold Change vs Wild type [NAI susceptible virus])			
			Oseltamivir	Zanamivir	Peramivir	Laninamivir
	N294S	294	RI (17–61)	NI (1–4)	RI (31)	?
	K360E	258–359	NI (2)	NI (2)	HRI (165)	NI (<3)
	R374K	371	HRI (101–407)	RI/HRI (29–145)	HRI (352)	?
	A395E	390	RI (5)	NI (1)	RI (5)	NI (<3)
	G407S	402	NI (4)	NI (7)	?	?
	D432G	429	NI (1)	NI (1)	RI (41)	NI (<3)
	G140R + N144K	142 + 146	NI (6)	RI (10)	HRI (257)	?
	Y142H + G145R	144 + 147	RI (5)	NI (4)	HRI (487)	NI (<3)

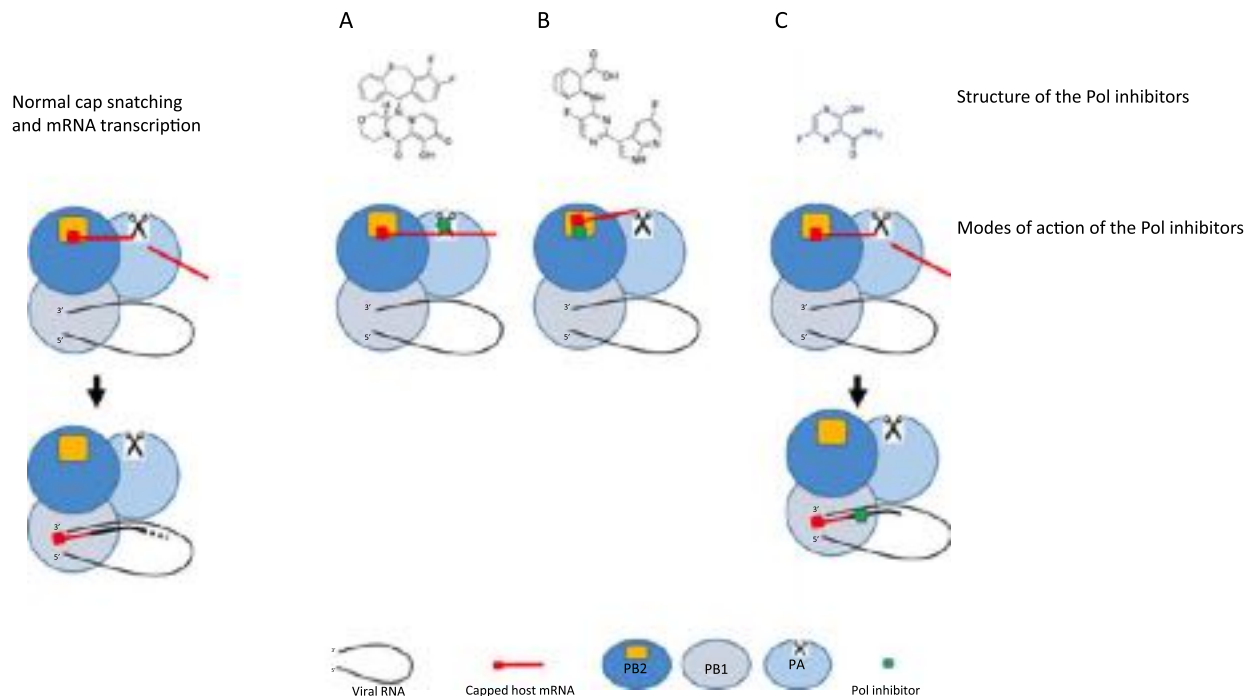


Fig. 5 Structure and mode of action of the different Polymerase inhibitors A describe the normal cap snatching and RNAZ synthesis by the influenza polymerase complex (PB1, PB2, and PA). The three polymerase inhibitors available have different modes of action, as described (Baloxavir inhibits cap snatching, Pimodivir inhibits cap binding, and favipiravir inhibits RNA elongation).

Ribavirine (Ribavirin®)

Historically, this molecule showed a rather large spectrum of activity against RNA viruses, including respiratory syncytial virus (RSV) and hepatitis C. The influenza antiviral effect could be measured in vitro and in pre-clinical studies, but Ribavirine always failed to show clinical efficacy against influenza, in human challenge studies, and clinical studies (Wright *et al.*, 2007; Schofield *et al.*, 1975).

The more recently described POL-inhibiting compounds were prepared by drug design. Solving the structure of the three components of the influenza polymerase (PA, PB1, and PB2) allowed the identification of potential targets and the design of inhibiting molecules (Pflug *et al.*, 2014; Pflug *et al.*, 2018). The POL inhibitors interfere with different components of the polymerase complex.

Baloxavir Marboxil (Xofluza®)

Baloxavir Marboxil was developed after solving the structure of the Cap-snatching pocket of the PA (Pflug *et al.*, 2014, 2018). Rapidly, it appeared that a drug that would incorporate this pocket would prevent the mandatory cap-snatching function of the PA and subsequently block RNA synthesis by the influenza polymerase. Baloxavir Marboxil was developed for that purpose and the preliminary in vitro and clinical results (4 log reduction in virus production) were very encouraging (Hayden *et al.*, 2018; Omoto

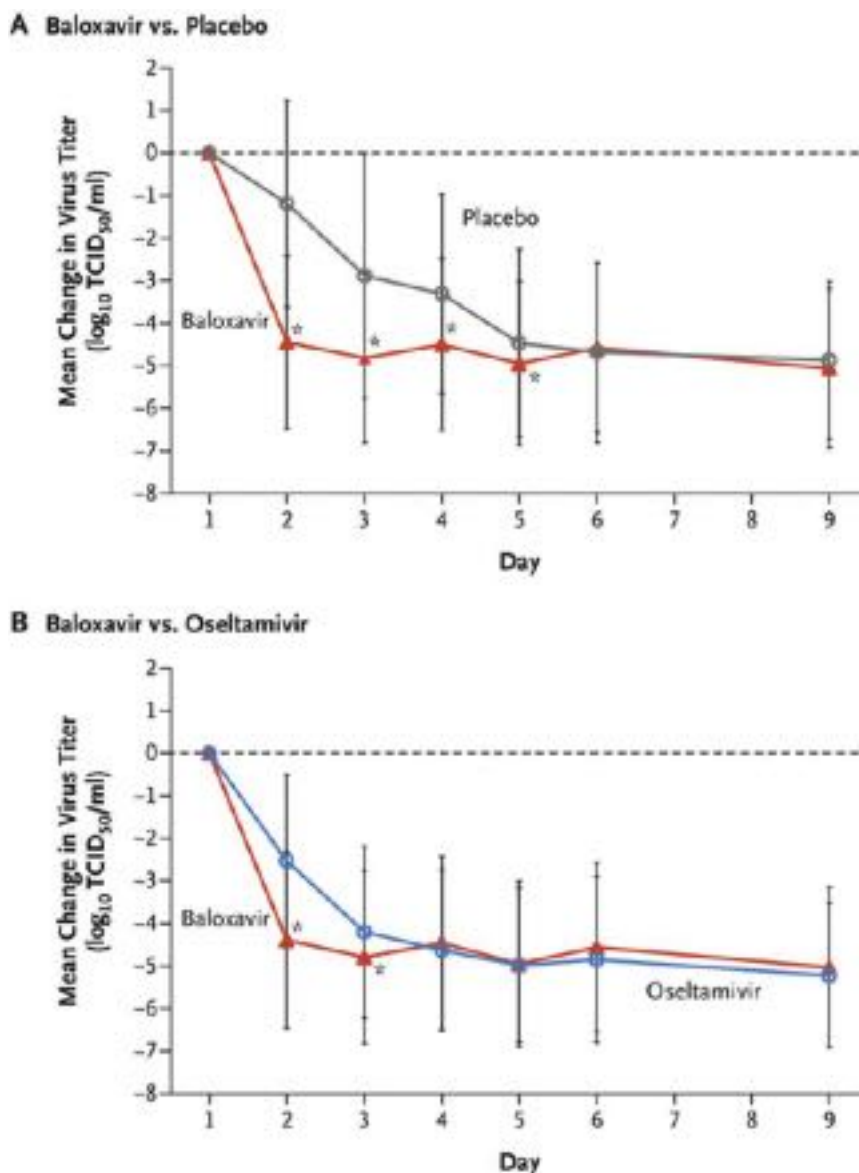


Fig. 6 Virological effect of Baloxavir marboxyl as compared to placebo and oseltamivir. Panel A shows the change from baseline (dashed line) in influenza infectious viral load overtime in the baloxavir group (427 patients) and placebo group (210 patients). The mean ($\pm$ SD) viral loads on day 1 (before the initiation of the trial regimen) were 5.79 ± 1.87 and 5.56 ± 1.89 log₁₀·50% tissue-culture infective dose (TCID₅₀) per milliliter in the baloxavir and placebo groups, respectively. Asterisks indicate a P value of less than 0.05 for the comparison with placebo. Panel B shows the change from baseline in influenza infectious viral load in adults 20–64 years of age in the baloxavir group (352 patients) and oseltamivir group (359 patients). The mean ($\pm$ SD) viral loads on day 1 were 5.76 ± 1.90 and 5.94 ± 1.69 log₁₀ TCID₅₀ per milliliter in the baloxavir and oseltamivir groups, respectively. Asterisks indicate a P value of less than 0.05 for the comparison with oseltamivir. In both panels, I bars indicate standard deviations. From Hayden, F.G., Sugaya, N., Hirotsu, N., *et al.*, 2018. Baloxavir marboxil investigators group. Baloxavir marboxil for uncomplicated influenza in adults and adolescents. The New England Journal of Medicine 379 (10), 913–923.

et al., 2018). This reduction was significantly higher than observed with the INAs, or any other anti-influenza drugs, and was noted in both influenza A and influenza B viruses (Fig. 6).

The first data obtained from clinical trials were, initially, not as promising as the in vitro results.

- (1) First, the clinical impact of treatment is more or less similar to that obtained with the NAs (roughly a 24 h reduction of symptoms duration), as assessed by two randomized, double-blind, controlled trials involving healthy 12–64 years of age outpatients with acute uncomplicated influenza (Hayden *et al.*, 2018; Omoto *et al.*, 2018).
- (2) Secondly, although it was speculated that substitutions associated with resistance might be difficult to accommodate for the virus and result in a severe fitness loss, it appeared that in vitro and in vivo resistance were detected rapidly, mostly at position I38 (I38T/M/F variants) of the polymerase acidic protein of A(H3N2) viruses; other substitutions are also associated with

reduced susceptibility (Takashita *et al.*, 2019; Imai *et al.*, 2020), and this resistant virus could be readily transmitted through airborne spread (Imai *et al.*, 2020). The monitoring of the emergence of resistance (by PCR) reported resistance in 2.2% and 9.7% of baloxavir recipients in the phase two and phase three trials, respectively (Takashita *et al.*, 2019).

These clinical results and resistance patterns have been confirmed by additional observational studies; Baloxavir-Marboxyl is now licensed in Japan. It seems that the I38T substitution can emerge, be sustainable, and potentially be transmitted with a higher frequency in A(H3N2) viruses. To address this risk, in vivo studies conducted in animals and in vitro competition experiments confirmed the lack of fitness related to the I38T substitution, for A(H1N1) and for A(H3N2) viruses.

In addition, a clinical investigation showed that pre-existing resistant viruses were circulating in the community, as seen for NAI before their usage (Ferraris and Lina, 2008; McKimm-Breschkin, 2013; Takashita *et al.*, 2019; Imai *et al.*, 2020). As with NAIs, children have been reported to be more likely to develop resistance during treatment, especially during A(H3N2) infections.

As a consequence, surveillance of baloxavir-resistant I38T PA mutants in seasonal subtypes has been implemented.

Pimodivir

This third polymerase inhibitor is a non-nucleoside analog, targeting PB2 and inhibiting the docking of the cap structure in the polymerase subunit to allow RNA synthesis (Mifsud *et al.*, 2019). Because of a major difference between influenza A and B PB2 domains, Pimodivir is only active against influenza A subtypes. It allows, as with Baloxavir, a rapid decrease in virus load both in vivo and in vitro. It can be given orally, twice daily.

In a phase 2b clinical trial TOPAZ, the viral load – as measured by qRT-PCR in nasal secretions from baseline to day eight – significantly decreased in all treatment groups compared with the placebo group. A Pimodivir plus Oseltamivir combination resulted in a shorter time to symptom resolution than placebo, suggesting a synergistic effect between the two (Fig. 7) (Finberg *et al.*, 2019).

Preliminary data have been collected regarding the risk for Pimodivir resistance. Reduced antiviral activity was reported in viruses with a PB2 F363L substitution. Overall, the PB2 inhibitor gave promising results, and phase three clinical trials are testing its use as monotherapy and in combination with Oseltamivir.

Favipiravir (Avigan®)

The fourth molecule of this class is Favipiravir. It is a purine nucleoside analog that inhibits the PB1 RNA-dependent RNA polymerase (Mifsud *et al.*, 2019). It has demonstrated a large spectrum of action in vitro against influenza A and B viruses, but also against other RNA viruses including Ebola viruses (Furuta *et al.*, 2017).

Japan has approved its use to treat infection with NAI-resistant viruses. However, its use is restricted because of potential teratogenicity. No resistance has been observed so far (Hayden and Shindo, 2019).

Mechanisms of resistance have been investigated. Two substitutions located in PB1 (K229R) and in PA (P653L) have been detected in viruses resistant to Favipiravir in vitro. The K229R substitution in PB1 has a cost in viral fitness for the virus; the polymerase activity is severely impaired by this substitution. However, when compensated with the P653L substitution in PA, the viral fitness of the resistant virus is restored. This K229R substitution may induce resistance to other POL inhibitors of influenza viruses, such as ribavirin (Goldhill *et al.*, 2018).

The common feature of these polymerase inhibitors is their capacity to be associated with NAI and have a synergistic effect for the treatment of influenza. Both Baloxavir and Pimodivir have been tested in combination. Favipiravir has also been tested against influenza viruses resistant to NAI, showing good activity and demonstrating, as for the other pol inhibitors, the lack of cross-resistance with NAIs.

Drugs Under Development With Other Mechanisms of Action

HA processing blockers (Fig. 1)

Nitazoxanide is a drug that was first licensed for the treatment of parasitic diseases. It can also inhibit the replication of a relatively broad range of viruses, including influenza A and B (Beigel *et al.*, 2019; Gamiño-Arroyo *et al.*, 2019). The nitazoxanide metabolite of nitazoxanide can block the maturation process of the HA in the cell by impairing its traffic and its insertion in the cellular membrane before the virus budding. In vitro studies have shown antiviral activity at a low concentration (around 30–100 ng/mL), but the clinical trials have failed to show a clinical effect, so far.

HA attachment blockers (Fig. 1)

Fludase (DAS – 181) is a recombinant sialidase that showed potent antiviral by preventing the entry of influenza viruses in cells through sialic receptors removal from the epithelium (Beigel *et al.*, 2019). A phase two study conducted provided encouraging results. However, the antiviral effect of fludase might be limited because individuals developed antibodies against the compound during treatment (Zenilman *et al.*, 2015). This will be further investigated but, so far, no further clinical trial has been carried out.

Other HA binding inhibitors such as monoclonal antibodies targeting the globular head or the stem of the HA, and small molecules binding to the stem of the HA are in the early stages of development.

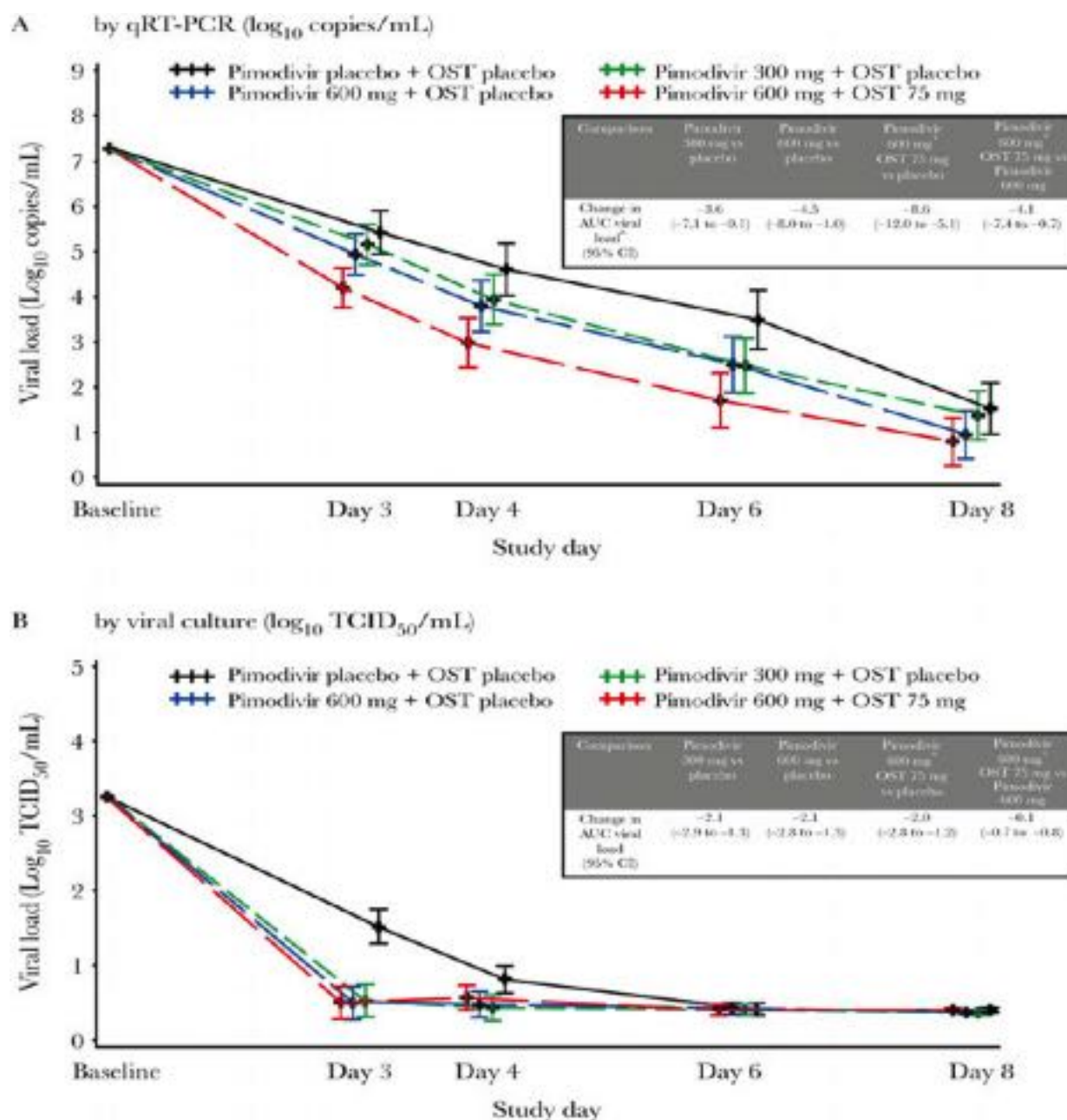


Fig. 7 Clinical impact of pimodivir and pimodivir in association with oseltamivir (Result of the TOPAZ clinical trial, see (Finberg *et al.*) Viral load over time in full analysis set, as determined by quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis (A) and viral culture (B). The limit of quantification for the viral load assay was 4.0 $\log_{10}$ copies/mL, with a limit of detection of 3.48 $\log_{10}$ copies/mL. Results less than the limit of quantification and greater than the limit of detection ("target detected") are imputed with 3.75 $\log_{10}$ copies/mL; results less than the LOD ("target not detected") are imputed with 0 $\log_{10}$ copies/mL. Error bars are 95% confidence intervals (CIs). AUC, area under the curve; OST, oseltamivir; TCID₅₀, median tissue culture infectious dose. From Finberg, R.W., Lanno, R., Anderson, D., *et al.*, 2019. Phase 2b study of pimodivir (JNJ-63623872) as monotherapy or in combination with oseltamivir for treatment of acute uncomplicated seasonal influenza A: Topaz trial. *The Journal of Infectious Diseases* 219 (7), 1026–1034.

Perspectives at Short and Medium Term

In the context of the development of new antiviral drugs, new treatment strategies have been tested, especially during the 2009 pandemic and its aftermath. Actually, monotherapy regimens have shown their added value, but also their limits (lack of very active antiviral and clinical impact). To overcome this problem, novel antiviral strategies have been developed and tested, including host targeted molecules and combined regimens (Beigel *et al.*, 2019; McKimm-Breschkin *et al.*, 2018).

Host Target Molecules (i.e., Diltiazem)

Novel therapeutic approaches to control influenza are currently in development. As successfully proven in antiretroviral therapy (Hütter *et al.*, 2015), targeting the host instead of viral determinants may confer a broad-spectrum antiviral efficacy and a reduced risk of emergence of drug resistance (Ludwig, 2011). In that regard, DNA microarray and, more recently, RNAseq techniques – an *in-silico* assisted strategy based on transcriptional profiling and signature matching – have been developed. Based on the hypothesis that, instead of focusing on specific factors, the global host gene expression profile can be considered as a “signature” of any specific cell state, including during infection or treatment. It has been supposed that the drug-induced inhibition or reversion of the cellular transcriptomic signature of infection (“proviral status”) could establish a globally unfavorable cellular state for viral replication (“antiviral state”).

This hypothesis of signature reversion has been validated through proof-of-concept studies based on the exploitation of global virogenomic signatures of *in vitro* infections with different human and avian influenza subtypes (Josset *et al.*, 2010). Subsequently, the *in vivo* global transcriptional signatures of infection directly from paired nasal wash samples from a cohort of untreated influenza A(H1N1)pdm09-infected patients has been determined. Samples were collected during acute infection (infected) and at least three months later to ensure a recovery non-infected state (cured). Then, an *in silico* drug screening using Connectivity Map (CMAP), a Broad Institute’s publicly available database (more than 7000 drug-associated gene expression profiles obtained after treatment of cells with 1309 different FDA-approved small molecules, (Lamb, 2007)), allowed the identification of a shortlist of high-potential candidate bioactive molecules with signatures anti-correlated with those of the patient’s acute infection state. The putative antiviral properties of these “*in silico* selected” molecules were confirmed and experimentally validated in classic *in vitro* models.

The most effective compounds identified through this process were compared to the gold standard Oseltamivir for the treatment of influenza infections in mice and in an *in vitro* model of reconstituted human airway epithelium (hAE), leading to the identification of two drugs: The adrenergic receptor agonist Etilefrine (Effortil®) and a calcium channel blocker Diltiazem (Tildiem®); the latter being able to stimulate the epithelial antiviral defense during influenza A or B infection (Pizzorno *et al.*, 2019a,b). There is currently a multicenter randomized phase 2b clinical trial trying to assess the effect of this combined therapy compared to the standard of care (oseltamivir monotherapy) for the treatment of severe influenza.

This original, host-targeted, drug repurposing strategy constitutes an effective and highly reactive process for the rapid identification of novel anti-infectious drugs, with potentially major implications for the management of antimicrobial resistance and the rapid response to future epidemic or pandemic (re)emerging diseases.

Macrolides

The added value of macrolides in treating severe influenza has been discussed for some years, as a result of finding significant anti-inflammatory effects when used in adults with severe influenza. This strategy is similar to the host-targeted antiviral approach (Beigel *et al.*, 2019).

The mechanism of this anti-inflammatory/immunomodulatory effect is not fully understood, but two interesting trials conducted with Azithromycin and Clarithromycin reported a faster decline of the inflammatory markers in the Oseltamivir-Azithromycin group as compared to the Oseltamivir alone groups, and a trend toward faster symptom resolution although no differences were observed in viral RNA decline. In a second prospective study, two groups of patients received either a two-day combination of Clarithromycin 500 mg, Naproxen 200 mg, and Oseltamivir 75 mg twice daily, followed by three days of Oseltamivir, or Oseltamivir 75 mg twice daily for five days were compared (Hung *et al.*, 2017). The combination treatment was statistically associated with reduced mortality, less clinical severity, and shorter hospital stay. In addition, they also reported a reduction in the virus titers and pneumonia severity index in the combination group. This potential beneficial effect of macrolide adjunction will be further investigated.

Other Immunomodulators

Potential immunomodulators have also been tested at least in preclinical studies or suggested after limited observational studies. This includes macrolides, sirolimus, N-acetylcysteine, statins, pamidronate, nitazoxanide, chloroquine, antiC5a antibody, interferons, human mesenchymal stromal cells, mycophenolic acid, peroxisome proliferator-activated receptors agonists, non-steroidal anti-inflammatory agents, mesalazine, herbal medicine, and the role of plasmapheresis and hemoperfusion as rescue therapy (Beigel *et al.*, 2019; McKimm-Breschkin *et al.*, 2018). All these strategies deserve more investigation preferably by RCTs. Two have been tested in clinical trials, with some interesting results: macrolides (as described above) and sirolimus.

Sirolimus

In Taiwan, an open-label prospective randomized controlled trial was conducted in patients with severe H1N1 pneumonia with acute respiratory failure. Patients with confirmed H1N1 pneumonia and on mechanical ventilator support were randomized to receive – in addition to the standard of care – Oseltamivir, an adjuvant treatment of corticosteroids with an mTOR inhibitor, either with Sirolimus for 14 days or without (Jia *et al.*, 2018).

Biological and clinical monitoring showed that the outcomes were significantly better in the sirolimus group over the non-sirolimus group. The authors suggested a combination treatment with Sirolimus along with Oseltamivir as an immunomodulatory strategy for managing severe influenza.

Combined Treatments (Double, Triple)

The observed emergence of resistance in all influenza classes and the larger portfolio of anti-influenza classes are pushing the implementation of combined therapies. Most remain at the stage of preclinical, phase two, or phase three trials, but the preliminary results have already shown the limits or added values of some of the tested combinations.

NAI combinations

The potential added value of NAI combinations have been tested before, showing antagonistic effects. NAI combinations have been prohibited.

Triple therapy (amantadine, ribavirine oseltamivir)

Preliminary in vitro studies have shown a potential synergistic antiviral effect when Amantadine, Ribavirine, and Oseltamivir were combined, even with naturally resistant Oseltamivir viruses (Nguyen *et al.*, 2009). Although this treatment showed a significant decrease in viral shedding at day three as compared to the monotherapy arm in a clinical trial, no improved clinical benefit was observed (Beigel *et al.*, 2017).

Combined oseltamivir plus monoclonal antibodies

Monoclonals targeting the stalk of the HA are under development (Beigel *et al.*, 2019; McKimm-Breschkin *et al.*, 2018). Through binding at this conserved region involved in the conformational change, these antibodies inhibit the HA-mediated fusion. For example, MEDI8852 induced a high level of blockage that resulted in a significant reduction of cell-to-cell spread of the virus. A trial conducted in individuals who were 18–65 years old confirmed the decrease in viral shedding in treated subjects. A combined mAB + Oseltamivir strategy is currently being investigated in hospitalized patients (Ali *et al.*, 2018).

Combined oseltamivir plus favipiravir

This combination of recently developed antivirals has different virus targets (NA for Oseltamivir and PB1 for Favipiravir). The preclinical study conducted in a pharmacologically immunosuppressed mouse model infected with the A(H1N1) pandemic influenza virus showed that the treatment of immunosuppressed mice with high doses of Favipiravir (50 mg/kg) in combination with Oseltamivir (20 mg/kg) resulted in delayed mortality and reduced lung viral titers compared to treatment with a single drug regimen with Oseltamivir (Baz *et al.*, 2018). However, it did not prevent the emergence of Oseltamivir-resistant H275Y neuraminidase variants. Nevertheless, this regimen has been tested in a clinical trial showing interesting results and will be further investigated (Wang *et al.*, 2020).

Combined baloxavir + oseltamivir

After showing the absence of drug-drug interactions (Kawaguchi *et al.*, 2018), a preclinical study was carried out in a mice model and this treatment has been successfully tested (Hayden *et al.*, 2018).

Pimodivir and oseltamivir

This combination has been tested in a human trial (TOPAZ), as described above.

Complementary Treatments for Disease Management

Different complementary treatments demonstrated some benefit, such as passive immunotherapy with convalescent plasma and hyperimmune globulin, when administered as an adjunctive therapy for severe influenza.

On the other hand, the use of corticosteroids in severe influenza has been discussed intensively. A recently published meta-analysis showed the lack of benefit of high doses of corticosteroids (Hui *et al.*, 2018). However, short-lasting corticosteroid treatment with low doses may be beneficial. This needs further evaluation to adapt the correct dose to the clinical situation and the timing of the infection. These products should be avoided as much as possible during the anergic immune response observed during severe influenza that behaves like a severe sepsis.

ECMO (Extracorporeal Membrane Oxygenation)

ECMO has been extensively used during the 2009 pandemic when potentially reversible acute respiratory failure associated with severe influenza A (H1N1) pneumonia was observed. However, ECMO remains an expensive, resource-intensive therapy that needs trained staff with intensive care skills. Despite the value of this intervention, it has a high level of mortality. A review and

meta-analysis on the use of ECMO reported overall mortality of 37.1% with a median duration for ECMO of ten days, mechanical ventilation of 19 days, and ICU length of stay of 33 days. They concluded that ECMO therapy was associated with better survival if used very early, once the patient had been started on mechanical ventilation, as an adjunct or salvage therapy for severe H1N1 pneumonia with respiratory failure (Sukhal *et al.*, 2017).

Conclusion

There is absolutely no doubt about the crucial need for antivirals in influenza disease management. A recently published meta-analysis confirmed the added value of the currently available antivirals (neuraminidase inhibitors). These antivirals represented the standard of care for hospitalized patients, but with some limits inherent to the influenza infections (acute disease with short-lived viral shedding, capacity for the rapid development of resistance, cytokine storm responsible for some of the clinical symptoms), new classes are currently being developed against new targets to tackle some of these problems. First, polymerase inhibitors were tested and showed their synergistic effect in combination with NAIs. In addition, host-targeted strategies or immunomodulators have also been developed and tested in combination with virus-targeted antivirals. All these strategies aim to reduce the burden of influenza, but also to become an efficient first line of treatment in case of a pandemic or zoonotic influenza.

The (hopefully) successful development of all these classes of molecules will certainly be of help to reduce the burden of seasonal influenza. They will also be useful in controlling influenza infection and symptoms in frail patients or in patients admitted to the intensive care units because of severe disease. We hope that the use of new combined therapeutic strategies will show a sizeable reduction in disease duration, much better than is currently observed with mild influenza cases when treated with the limited antiviral resources currently available.

References

- Ali, S.O., Takas, T., Nyborg, A., *et al.*, 2018. Evaluation of MEDI8852, an anti-influenza A monoclonal antibody, in treating acute uncomplicated influenza. *Antimicrobial Agents and Chemotherapy* 62 (11).
- Allen, E.K., Randolph, A.G., Bhargale, T., *et al.*, 2017. SNP-mediated disruption of CTCF binding at the IFITM3 promoter is associated with risk of severe influenza in humans. *Nature Medicine* 23 (8), 975–983.
- Aoki, F., 1998. In: Nicholson, K., Webster, R.G., Hay, A.J. (Eds.), *Text Book of Influenza*. London: Blackwell Science Ltd, pp. 457–476.
- Barr, I.G., Hurt, A.C., Deed, N., *et al.*, 2007. The emergence of adamantane resistance in influenza A(H1N1) viruses in Australia and regionally in 2006. *Antiviral Research* 75, 173–176.
- Baz, M., Carboneau, J., Rhéaume, C., Cavanagh, M.H., Boivin, G., 2018. Combination therapy with oseltamivir and favipiravir delays mortality but does not prevent oseltamivir resistance in immunodeficient mice infected with pandemic A(H1N1) influenza virus. *Viruses* 10 (11).
- Beigel, J.H., Bao, Y., Beeler, J., *et al.*, 2017. Oseltamivir, amantadine, and ribavirin combination antiviral therapy versus oseltamivir monotherapy for the treatment of influenza: A multicentre, double-blind, randomised phase 2 trial. *Lancet Infectious Diseases* 17 (12), 1255–1265.
- Beigel, J.H., Farrar, J., Han, A.M., *et al.*, 2005. Avian influenza A (H5N1) infection in humans. *The New England Journal of Medicine* 353, 1374–1385.
- Beigel, J.H., Nam, H.H., Adams, P.L., *et al.*, 2019. Advances in respiratory virus therapeutics – A meeting report from the 6th isiv Antiviral Group conference. *Antiviral Research* 167, 45–67.
- Bright, R.A., Medina, M.J., Xu, X., *et al.*, 2005. Incidence of adamantane resistance among influenza A ((H3N2)) viruses isolated worldwide from 1994 to 2005: A cause for concern. *The Lancet* 366, 1175–1181.
- Burnham, A.J., Baranovich, T., Marathe, B.M., *et al.*, 2014. Fitness costs for influenza B viruses carrying neuraminidase inhibitor-resistant substitutions: underscoring the importance of E119A and H274Y. *Antimicrobial Agents and Chemotherapy* 58 (5), 2718–2730.
- Colman, P.M., 1994. Influenza virus neuraminidase: Structure, antibodies, and inhibitors. *Protein Science* 3, 1687–1696.
- De Francisco Shapovalova, N., Donadel, M., Jit, M., Hutubessy, R., 2015. A systematic review of the social and economic burden of influenza in low- and middle-income countries. *Vaccine* 33 (48), 6357–6544.
- Deyde, V.M., Xu, X., Bright, R.A., *et al.*, 2007. Surveillance of resistance to adamantanes among influenza A(H3N2) and A(H1N1) viruses isolated worldwide. *The Journal of Infectious Diseases* 196, 249–257.
- Dharan, N.J., Gubareva, L.V., Meyer, J.J., *et al.*, 2009. Infections with oseltamivir-resistant influenza A(H1N1) virus in the United States. *JAMA* 301, 1034–1041.
- Doll, M.K., Winters, N., Boikos, C., *et al.*, 2017. Safety and effectiveness of neuraminidase inhibitors for influenza treatment, prophylaxis, and outbreak control: A systematic review of systematic reviews and/or meta-analyses. *Journal of Antimicrobial Chemotherapy* 72 (11), 2990–3007.
- Duval, X., van der Werf, S., Blanchon, T., *et al.*, 2010. Efficacy of oseltamivir-zanamivir combination compared to each monotherapy for seasonal influenza: A randomized placebo-controlled trial. *PLOS Medicine* 7 (11), e1000362.
- Escuret, V., Cornu, C., Boutitie, F., *et al.*, 2012. Oseltamivir-zanamivir bitherapy compared to oseltamivir monotherapy in the treatment of pandemic 2009 influenza A(H1N1) virus infections. *Antiviral Research* 96 (2), 130–137.
- Federici, C., Cavazza, M., Costa, F., Jommi, C., 2018. Health care cost of influenza-related episodes in high income countries: A systematic review. *PLoS One* 13 (9), e0202787.
- Ferraris, O., Kessler, N., Lina, B., 2005. Sensitivity of influenza viruses to zanamivir and oseltamivir: A study performed in France prior to the introduction of neuraminidase inhibitors in clinical practice. *Antiviral Research* 68, 43–48.
- Ferraris, O., Lina, B., 2008. Mutation of neuraminidase implicated in neuraminidase inhibitors resistance. *Journal of Clinical Virology* 41, 13–19.
- Finberg, R.W., Lanno, R., Anderson, D., *et al.*, 2019. Phase 2b study of pimodivir (JNJ-63623872) as monotherapy or in combination with oseltamivir for treatment of acute uncomplicated seasonal influenza A: Topaz trial. *The Journal of Infectious Diseases* 219 (7), 1026–1034.
- Furuta, Y., Komeno, T., Nakamura, T., 2017. Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proceedings of the Japan Academy, Series B, Physical and Biological Sciences* 93 (7), 449–463.
- Gamiño-Arroyo, A.E., Guerrero, M.L., McCarthy, S., *et al.*, 2019. Efficacy and safety of nitazoxanide in addition to standard of care for the treatment of severe acute respiratory illness. *Clinical Infectious Diseases* 69 (11), 1903–1911.

- Goldhill, D.H., Te Velthuis, A.J.W., Fletcher, R.A., *et al.*, 2018. The mechanism of resistance to favipiravir in influenza. *Proceedings of the National Academy of Sciences of the United States of America* 115 (45), 11613–11618.
- Gubareva, L.V., Kaiser, L., Hayden, F.G., 2000. Influenza virus neuraminidase inhibitors. *Lancet* 355, 827–835.
- Hay, A.J., Zamboni, M.C., Wolstenholme, A.J., Skehel, J.J., Smith, M.H., 1986. Molecular basis of resistance of influenza A viruses to amantadine. *Journal of Antimicrobial Chemotherapy* 18 (Suppl B), 19–29.
- Hayden, F.G., Shindo, N., 2019. Influenza virus polymerase inhibitors in clinical development. *Current Opinion in Infectious Diseases* 32 (2), 176–186.
- Hayden, F.G., Sperber, S.J., Belshe, R.B., *et al.*, 1991. Recovery of drug-resistant influenza A virus during therapeutic use of rimantadine. *Antimicrobial Agents and Chemotherapy* 35, 1741–1747.
- Hayden, F.G., Sugaya, N., Hirotsu, N., *et al.*, 2018. Baloxavir marboxil investigators group. Baloxavir marboxil for uncomplicated influenza in adults and adolescents. *The New England Journal of Medicine* 379 (10), 913–923.
- Heneghan, C.J., Onakpoya, I., Thompson, M., *et al.*, 2014. Zanamivir for influenza in adults and children: Systematic review of clinical study reports and summary of regulatory comments. *BMJ* 348, g2547.
- Herlocher, M.L., Carr, J., Ives, J., *et al.*, 2002. Influenza virus carrying an R292K mutation in the neuraminidase gene is not transmitted in ferrets. *Antiviral Research* 54, 99–111.
- Hui, D.S., Lee, N., Chan, P.K., Beigel, J.H., 2018. The role of adjuvant immunomodulatory agents for treatment of severe influenza. *Antiviral Research* 150, 202–216.
- Hung, I.F.N., To, K.K.W., Chan, J.F.W., *et al.*, 2017. Efficacy of clarithromycin-naproxen-oseltamivir combination in the treatment of patients hospitalized for influenza A(H3N2) infection: An open-label randomized, controlled, phase IIb/III Trial. *Chest* 151 (5), 1069–1080.
- Hurt, A.C., Ernest, J., Deng, Y., *et al.*, 2009. Emergence and spread of oseltamivir-resistant A(H1N1) influenza viruses in Oceania, South East Asia and South Africa. *Antiviral Research* 83, 90–93.
- Hurt, A.C., Nor'e, S.S., McCaw, J.M., *et al.*, 2010. Assessing the viral fitness of oseltamivir-resistant influenza viruses in ferrets, using a competitive-mixtures model. *Journal of Virology* 84, 9427–9438.
- Hütter, G., Bodor, J., Ledger, S., *et al.*, 2015. CCR5 targeted cell therapy for HIV and prevention of viral escape. *Viruses* 7, 4186–4203.
- Ikematsu, H., 2011. Kawai n Laninamivir octanoate: A new long-acting neuraminidase inhibitor for the treatment of influenza. *Expert Review of Anti-infective Therapy* 9, 851–857.
- Imai, M., Yamashita, M., Sakai-Tagawa, Y., *et al.*, 2020. Influenza A variants with reduced susceptibility to baloxavir isolated from Japanese patients are fit and transmit through respiratory droplets. *Nature Microbiology* 5 (1), 27–33.
- Iuliano, A.D., Roguski, K.M., Chang, H.H., *et al.*, 2018. Estimates of global seasonal influenza-associated respiratory mortality: A modelling study. *Lancet* 391 (10127), 1285–1300.
- Ives, J.A.L., Carr, J.A., Mendel, D.B., *et al.*, 2002. The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment leave virus severely compromised both in vitro and in vivo. *Antiviral Research* 55, 307–317.
- Jia, X., Liu, B., Bao, L., *et al.*, 2018. Delayed oseltamivir plus sirolimus treatment attenuates H1N1 virus-induced severe lung injury correlated with repressed NLRP3 inflammasome activation and inflammatory cell infiltration. *PLoS Pathogens* 14 (11), e1007428.
- Josset, L., Textoris, J., Liorod, B., *et al.*, 2010. Gene expression signature-based screening identifies new broadly effective influenza a antivirals. *PLoS One* 5, e13169.
- Kashiwagi, S., Watanabe, A., Ikematsu, H., Uemori, M., Awamura, S., 2016. Laninamivir prophylaxis study group. Long-acting neuraminidase inhibitor laninamivir octanoate as post-exposure prophylaxis for influenza. *Clinical Infectious Diseases* 63 (3), 330–337.
- Katzen, J., Kohn, R., Houk, J.L., Ison, M.G., 2019. Early oseltamivir after hospital admission is associated with shortened hospitalization: A 5-year analysis of oseltamivir timing and clinical outcomes. *Clinical Infectious Diseases* 69 (1), 52–58.
- Kawaguchi, N., Koshimichi, H., Ishibashi, T., Wajima, T., 2018. Evaluation of drug-drug interaction potential between baloxavir marboxil and oseltamivir in healthy subjects. *Clinical Drug Investigation* 38 (11), 1053–1060.
- Koopmans, M., Wilbrink, B., Conyn, M., *et al.*, 2004. Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. *Lancet* 363, 587–593.
- Lackenby, A., Besselaar, T.G., Daniels, R.S., *et al.*, 2018. Global update on the susceptibility of human influenza viruses to neuraminidase inhibitors and status of novel antivirals, 2016–2017. *Antiviral Research* 157, 38–46.
- Lamb, J., 2007. The connectivity map: A new tool for biomedical research. *Nature Reviews Cancer* 7, 54–60.
- Lina, B., Boucher, C., Osterhaus, A., *et al.*, 2018. Five years of monitoring for the emergence of oseltamivir resistance in patients with influenza A infections in the influenza resistance information study. *Influenza and Other Respiratory Viruses* 12 (2), 267–278.
- Ludwig, S., 2011. Disruption of virus-host cell interactions and cell signaling pathways as an anti-viral approach against influenza virus infections. *Biological Chemistry* 392, 837–847.
- Mancuso, C.E., Gabay, M.P., Steinke, L.M., Vanosdol, S.J., 2010. Peramivir: An intravenous neuraminidase inhibitor for the treatment of 2009 H1N1 influenza. *Annals of Pharmacotherapy* 44, 1240–1249.
- Marty, F.M., Vidal-Puigserver, J., Clark, C., *et al.*, 2017. Intravenous zanamivir or oral oseltamivir for hospitalised patients with influenza: An international, randomised, double-blind, double-dummy, phase 3 trial. *The Lancet Respiratory Medicine* 5 (2), 135–146.
- McKimm-Breschkin, J.L., 2013. Influenza neuraminidase inhibitors: Antiviral action and mechanisms of resistance. *Influenza Other Respiratory Viruses*. (Suppl 1), 25–36.
- McKimm-Breschkin, J.L., Jiang, S., Hui, D.S., *et al.*, 2018. Prevention and treatment of respiratory viral infections: presentations on antivirals, traditional therapies and host-directed interventions at the 5th ISIRV Antiviral Group conference. *Antiviral Research* 149, 118–142.
- McLean, K.A., Goldin, S., Nannei, C., Sparrow, E., Torelli, G., 2016. The 2015 global production capacity of seasonal and pandemic influenza vaccine. *Vaccine* 34, 5410–5413.
- Meijer, A., Lackenby, A., Hungnes, O., *et al.*, 2009. Oseltamivir-resistant influenza virus A (H1N1), Europe, 2007–08 season. *Emerging Infectious Diseases* 15, 552–560.
- Mifsud, E.J., Hayden, F.G., Hurt, A.C., 2019. Antivirals targeting the polymerase complex of influenza viruses. *Antiviral Research* 169, 104545.
- Molinari, N.A., *et al.*, 2007. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine* 25, 5086–5096.
- Monto, A.S., Fukuda, K., 2020. Lessons from influenza pandemics of the last 100 years. *Clinical Infectious Diseases* 70 (5), 951–957.
- Monto, A.S., McKimm-Breschkin, J.L., Macken, C., *et al.*, 2006. Detection of influenza viruses resistant to neuraminidase inhibitors in global surveillance during the first 3 years of their use. *Antimicrobial Agents and Chemotherapy* 50, 2395–2402.
- Muthuri, S.G., Myles, P.R., Venkatesan, S., Leonardi-Bee, J., Nguyen-Van-Tam, J.S., 2013. Impact of neuraminidase inhibitor treatment on outcomes of public health importance during the 2009–2010 influenza A(H1N1) pandemic: A systematic review and meta-analysis in hospitalized patients. *The Journal of Infectious Diseases* 207 (4), 553–563.
- Muthuri, S.G., Venkatesan, S., Myles, P.R., *et al.*, 2014. Effectiveness of neuraminidase inhibitors in reducing mortality in patients admitted to hospital with influenza A H1N1pdm09 virus infection: A meta-analysis of individual participant data. *The Lancet Respiratory Medicine* 2 (5), 395–404.
- Nakano, T., Okumura, A., Tanabe, T., *et al.*, 2013. Safety evaluation of laninamivir octanoate hydrate through analysis of adverse events reported during early post-marketing phase vigilance. *Scandinavian Journal of Infectious Diseases* 45 (6), 469–477.
- Nguyen, H.T., Fry, A.M., Gubareva, L.V., 2012. Neuraminidase inhibitor resistance in influenza viruses and laboratory testing methods. *Antiviral Therapy* 17, 159–173.
- Nguyen, J.T., Hoopes, J.D., Smee, D.F., *et al.*, 2009. Triple combination of oseltamivir, amantadine, and ribavirin displays synergistic activity against multiple influenza virus strains in vitro. *Antimicrobial Agents and Chemotherapy* 53 (10), 4115–4126.

- Oh, D.Y., Lowther, S., McCaw, J.M., *et al.*, 2014. Evaluation of oseltamivir prophylaxis regimens for reducing influenza virus infection, transmission and disease severity in a ferret model of household contact. *Journal of Antimicrobial Chemotherapy* 69 (9), 2458–2469.
- Okomo-Adhiambo, M., Nguyen, H.T., Abd, E.A., *et al.*, 2014. Drug susceptibility surveillance of influenza viruses circulating in the United States in 2011–2012: Application of the WHO antiviral working group criteria. *Influenza and Other Respiratory Viruses* 8, 258–265.
- Omoto, S., Speranzini, V., Hashimoto, T., *et al.*, 2018. Characterization of influenza virus variants induced by treatment with the endonuclease inhibitor baloxavir marboxil. *Scientific Reports* 8 (1), 9633.
- Palese, P., Shaw, M.L., 2007. Orthomyxoviridae: The Viruses and Their Replication. In: Knipe, D.M., Howley, P.M., Griffin, D.E., *et al.* (Eds.), *Fields Virology*. vol. 2, fifth ed. Philadelphia: Lippincott Williams and Wilkins, pp. 1647–1690.
- Peiris, M.J.S., 2009. Avian influenza viruses in humans. *Revue scientifique et technique* 28, 161–173.
- Peiris, M., Yuen, K.Y., Leung, C.W., *et al.*, 1999. Human infection with influenza H9N2. *Lancet* 354, 916–917.
- Pflug, A., Gaudon, S., Resa-Infante, P., *et al.*, 2018. Capped RNA primer binding to influenza polymerase and implications for the mechanism of cap-binding inhibitors. *Nucleic Acids Research* 46 (2), 956–971.
- Pflug, A., Guilligay, D., Reich, S., Cusack, S., 2014. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 516 (7531), 355–360.
- Pizzorno, A., Padley, B., Terrier, O., Rosa-Calatrava, M., 2019a. Drug repurposing approaches for the treatment of influenza viral infection: Reviving old drugs to fight against a long-lived enemy. *Frontiers in Immunology* 19.
- Pizzorno, A., Terrier, O., Nicolas de Lamballerie, C., *et al.*, 2019b. Repurposing of drugs as novel influenza inhibitors from clinical gene expression infection signatures. *Frontiers in Immunology* 29.
- Rothberg, M.B., Haessler, S.D., 2010. Complications of seasonal and pandemic influenza. *Critical Care Medicine* 38, e91–e97.
- Schilling, M., Gravenstein, S., Drinka, P., *et al.*, 2004. Emergence and transmission of amantadine-resistant influenza A in a nursing home. *Journal of the American Geriatrics Society* 52, 2069–2073.
- Schofield, K.P., Potter, C.W., Edey, D., Jennings, R., Oxford, J.S., 1975. Antiviral activity of ribavirin on influenza infection in ferrets. *Journal of Antimicrobial Chemotherapy* 1 (4 Suppl), 63–69.
- Shiraishi, K., Mitamura, K., Sakai-Tagawa, Y., *et al.*, 2003. High frequency of resistant viruses harboring different mutations in amantadine-treated children with influenza. *The Journal of Infectious Diseases* 188, 57–61.
- Smeeth, L., Thomas, S.L., Hall, A.J., *et al.*, 2004. Risk of myocardial infarction and stroke after acute infection or vaccination. *The New England Journal of Medicine* 351, 2611–2618.
- Spreeuwenberg, P., Kroneman, M., Paget, J., 2018. Reassessing the global mortality burden of the 1918 influenza pandemic. *American Journal of Epidemiology* 187 (12), 2561–2567.
- Sukhal, S., Sethi, J., Ganesh, M., *et al.*, 2017. Extracorporeal membrane oxygenation in severe influenza infection with respiratory failure: A systematic review and meta-analysis. *Annals of Cardiac Anaesthesia* 20 (1), 14–21.
- Takashita, E., Kawakami, C., Ogawa, R., *et al.*, 2019. Influenza A(H3N2) virus exhibiting reduced susceptibility to baloxavir due to a polymerase acidic subunit I38T substitution detected from a hospitalised child without prior baloxavir treatment, Japan, January 2019. *Eurosurveillance* 24, 12.
- Taubenberger, J.K., Morens, David, M., 2006. 1918 influenza: The mother of all pandemics. *Emerging Infectious Diseases* 12 (1), 15–22.
- Toovey, S., Rayner, C., Prinssen, E., *et al.*, 2008. Assessment of neuropsychiatric adverse events in influenza patients treated with oseltamivir: A comprehensive review. *Drug Safety* 31 (12), 1097–1114.
- Venkatesan, S., Myles, P.R., Leonardi-Bee, J., *et al.*, 2017. Impact of outpatient neuraminidase inhibitor treatment in patients infected with influenza A(H1N1)pdm09 at high risk of hospitalization: An individual participant data metaanalysis. *Clinical Infectious Diseases* 64 (10), 1328–1334.
- Wang, Y., Fan, G., Salam, A., *et al.*, 2020. Comparative effectiveness of combined favipiravir and oseltamivir therapy versus oseltamivir monotherapy in critically ill patients with influenza virus infection. *The Journal of Infectious Diseases* 221 (10), 1688–1698.
- Wedzhicha, J.A., Seemungal, T.A.R., 2007. COPD exacerbations: Defining their cause and prevention. *Lancet* 370 (9589), 786–796.
- Weinstock, D.M., Zuccotti, G., 2006. Adamantane resistance in influenza A. *JAMA* 295, 934–936.
- Welliver, R., Monto, A.S., Carewicz, O., *et al.*, 2001. Oseltamivir post exposure prophylaxis investigator group. Effectiveness of oseltamivir in preventing influenza in household contacts: A randomized controlled trial. *JAMA* 285 (6), 748–754.
- Whitley, R.J., Monto, A.S., 2006. Seasonal and pandemic influenza preparedness: A global threat. *The Journal of Infectious Diseases* 194 (Suppl 2), S65–S69.
- Wright, P.F., Neumann, G., Kawaoka, Y., 2007. Orthomyxoviruses. In: Knipe, D.M., Howley, P.M., Griffin, D.E., *et al.* (Eds.), *Fields Virology*. vol. 2, fifth ed. Philadelphia: Lippincott Williams and Wilkins, pp. 1691–1740.
- Yen, H.L., Herlocher, L.M., Hoffmann, E., *et al.*, 2005. Neuraminidase inhibitor-resistant influenza viruses may differ substantially in fitness and transmissibility. *Antimicrobial Agents and Chemotherapy* 49, 4075–4084.
- Zenilman, J.M., Fuchs, E.J., Hendrix, C.W., *et al.*, 2015. Phase 1 clinical trials of DAS181, an inhaled sialidase, in healthy adults. *Antiviral Research* 123, 114–119.

Further Reading

- Watanabe, A., Ishida, T., Hirotsu, N., *et al.*, 2019. Baloxavir marboxil in Japanese patients with seasonal influenza: dose response and virus type/subtype outcomes from a randomized phase 2 study. *Antiviral Research* 163, 75–81.

Management of Herpes Simplex Virus Infections (*Herpesviridae*)

Nicole Samies and Richard Whitley, University of Alabama at Birmingham, Birmingham, AL, United States

© 2021 Elsevier Ltd. All rights reserved.

Brief Overview of Available Antiviral Agents for HSV Infections

Nucleoside analogs (acyclovir, valacyclovir, and famciclovir) are the drugs of choice for treatment of herpes simplex virus (HSV) infections because of their relative safety profiles. Acyclovir is an acyclic guanine analog selectively phosphorylated by viral thymidine kinase first and then by cellular enzymes to form triphosphate-acyclovir. Once triphosphate-acyclovir is formed, it inhibits the viral DNA polymerase, becomes incorporated into the viral DNA chain, and leads to chain termination. The prodrug of acyclovir, valacyclovir, is synthesized by the addition of L-valine to acyclovir. In comparison to oral acyclovir, valacyclovir's bioavailability is significantly higher at 51%–54% compared to 15%–30%. The improved bioavailability allows for less frequent dosing in comparison to oral acyclovir. The third nucleoside analog, famciclovir, is a guanine analog and prodrug of penciclovir. Famciclovir is rapidly absorbed and converted to penciclovir, and then penciclovir is phosphorylated by viral thymidine kinase into penciclovir-triphosphate. Unlike acyclovir, penciclovir has a higher affinity for thymidine kinase and does not inhibit DNA chain formation. Famciclovir is preferred over penciclovir secondary to its improved bioavailability of 77%.

In addition to nucleoside analogs, foscarnet and cidofovir can be used in the treatment of HSV infections. However because of their toxicity profiles neither is recommended as a first-line antiviral agent. Both foscarnet and cidofovir inhibit the viral DNA polymerase similar to the nucleoside analogs, but they do not require phosphorylation by viral thymidine kinase; thus, they can be used in the treatment of resistant HSV strains which have altered their thymidine kinase preventing the phosphorylation of nucleoside analogs. Foscarnet is a pyrophosphate analog that is capable of binding to a particular site on the DNA polymerase without requiring phosphorylation prior. Cidofovir is a monophosphate cytosine analog that requires phosphorylation for activation, but phosphorylation occurs through cellular kinases rather than by the viral thymidine kinase. Once phosphorylated, the diphosphate form of cidofovir inhibits the viral DNA polymerase.

Management of Infections Beyond the Neonatal Period

Orolabial Infections

Gingivostomatitis is a common presentation of primary herpes simplex virus (HSV) infection in early childhood commonly confused with herpangina or hand-foot-and mouth disease, and the diagnosis should be confirmed prior to initiating treatment. The infection can be self-limiting with oral lesions lasting about 7–10 days and fever about 3–4 days. The most common complication of the infection is dehydration requiring hospital admission for intravenous fluids secondary to poor oral intake because of the painful oral ulcerations. Although treatment is not necessary, a randomized controlled trial demonstrated a reduction in the duration of oral ulcerations, fever, poor oral intake, and viral shedding when oral acyclovir is administered within the first 72 h of symptom onset. Oral acyclovir should be preferred over intravenous acyclovir if the child is able to take medication orally. A recent study also provided recommendations for valacyclovir dosing in children greater than 3 months of age; this may be a more favorable alternative to acyclovir secondary to the convenience of administration. Bleeding of lesions is common, and reassurance should be provided to parents. Children with gingivostomatitis should be excluded from childcare or school until resolutions of symptoms because of the high risk of transmission from one child to another.

Recurrent herpes labialis infections, also commonly referred to as 'cold sores' or 'fever blisters', is a common manifestation of HSV in adolescent and adult patients. Treatment of these lesions is limited; topical treatments, including acyclovir, penciclovir, and over the counter 10% *n*-docosanol cream, are not recommended because of limited effectiveness; oral options such as acyclovir, valacyclovir, and famciclovir only have marginal benefit when taken at the initial onset of symptoms. Although oral therapies have only marginal benefit, one randomized clinical trial in adults demonstrated clinical significance in the reduction of recurrent herpes labialis lesions when oral acyclovir was prescribed for suppressive therapy. Thus, patients with at least 6 episodes of recurrent herpes labialis lesions a year may be considered for suppressive therapy for 6 months to 12 months in order to reduce the number of recurrences. Unlike gingivostomatitis, children with recurrent herpes labialis infections do not need to be excluded from school.

Other mucocutaneous infections of the upper airway caused by HSV include pharyngitis, tonsillitis, epiglottitis, supraglottitis, and laryngotracheitis, which occur in both immunocompetent and immunosuppressed patients, can be treated with oral acyclovir.

Cutaneous Infections

Herpetic whitlow can be a complication of autoinoculation from oral secretions especially in children who are thumb-suckers or nail-biters or in adults secondary to job exposure as seen with personnel in dentistry, respiratory therapists, and pediatricians. The

diagnosis of HSV should be confirmed by cultures in order to distinguish it from bacterial felon or paronychia. Treatment with oral acyclovir shows some benefit in the treatment of lesions by reduction in the duration of symptoms and viral shedding as demonstrated in a randomized controlled study that evaluated the use of acyclovir in recurrent hand lesions.

Herpes gladiatorum and herpes rugbiorum are two types of cutaneous infections seen in athletes, most notably wrestling and rugby as the names suggest. Confirmation of the diagnosis should be made prior to treatment in order to distinguish the infection from impetigo. These infections should be treated with oral acyclovir, valacyclovir, or famciclovir, and athletes should be excluded from practice and/or competition until all lesions have crusted over to prevent the spread of infection. Athletes should also be encouraged to drink plenty of fluids to stay well hydrated while on an antiviral medication due to the risk of reversible nephrotoxicity. In order to prevent the spread of HSV between athletes, coaches and family members should be educated on the appearance of herpes mucocutaneous lesions to prevent those infected from playing in close contact sports, and all mats should be cleaned before, between, and after each practice or match to reduce the risk of transmission to players. Suppressive therapy with these medications has been employed.

Patients with eczema and burns are at risk for secondary HSV infections and should be treated with antivirals to prevent further progression to disseminated infection.

Ocular Infections

Primary herpes simplex virus ocular infections include blepharitis, follicular conjunctivitis, keratoconjunctivitis, retinitis and can be caused by mucocutaneous infections (e.g., herpes labialis, gingivostomatitis) from either direct contact or nerve route transmission. Children with concerns for HSV ocular infections should be promptly evaluated by an ophthalmologist and treated to prevent further complications as HSV is the most common cause of corneal scarring and vision loss in the United States. Herpetic keratitis is the most common of all the ocular diseases caused by HSV-1 and is often self-limiting. However, treatment reduces duration of symptoms, viral shedding, and risk of corneal scarring and vision loss. The classification into one of the three main subtypes of herpetic keratitis (epithelial, stromal, and endothelial) determines management of the infection. Treatment recommendations for the epithelial subtype are to administer antiviral agents solely whereas stromal and endothelial subtypes should be treated with a combination of antiviral agents and topical corticosteroids to decrease the amount of stromal inflammation. Of note, steroids should never be used alone in treatment of ocular infections. Children should be treated with both systemic acyclovir and ophthalmic antiviral drops or ointments (e.g., 1% trifluridine or 0.15% ganciclovir). Depending on the severity of the infection, the patient may require intravenous acyclovir and then switch to oral acyclovir to complete therapy. Topical vidarabine and iododeoxyuridine were additional antiviral options previously available for use in the treatment of ocular disease, but they are no longer available in the United States. Parents and providers should be aware of the barriers in administration of topical antivirals to children (e.g., crying during administration, skipping doses, or stopping therapy prematurely) and measures should be taken to address these.

Children are at increased risk for corneal scarring and blindness with each recurrent episode. Therefore, clinicians should consider placing affected children on suppressive therapy to prevent further reactivation. Suppressive therapy poses the risk of developing resistant HSV strains; however, resistant strains are more commonly detected in immunosuppressed patients. Nonetheless, if a patient's ocular disease progresses further despite being on adequate therapy, resistance should be suspected. Foscarnet is a second-line option in treatment of ocular disease caused by herpes simplex virus and can be used for treatment of herpes simplex virus when a resistant strain is detected. Foscarnet inhibits viral DNA polymerase, which is a different mechanism of action than the nucleoside analogs (acyclovir, valacyclovir, and famciclovir) which require phosphorylation by the viral thymidine kinase – the usual site of resistance – in order to inhibit HSV replication. Foscarnet does have a significant toxicity profile, which is why it is not first line for treatment. Providers should be aware of its nephrotoxicity and provide adequate pre- and post-hydration to patients when administering the antiviral agent.

CNS Infections

Herpes encephalitis is typically caused by reactivation of HSV 1 in adults but can be a primary infection in older children and is a common cause for sporadic encephalitis. Herpes encephalitis can cause fatal complications if diagnosis and treatment are delayed. Patients must be evaluated for herpes encephalitis by cerebrospinal fluid (CSF) HSV polymerase chain reaction (PCR) and should be empirically started on intravenous acyclovir until results are confirmed. Adults and children greater than 12 years of age should receive 10 mg/kg/dose every 8 h and children 3 months to 12 years of age should receive 10–15 mg/kg/dose every 8 h. Once diagnosis is confirmed, patients should be administered intravenous acyclovir for 14–21 days. In addition, all patients diagnosed with herpes encephalitis should have head magnetic resonance imaging (MRI) performed to determine areas of the brain affected and evaluate for a possible infarction; close monitoring of renal function as well as for signs of increased intracranial pressure are essential. Other considerations in the management of patients may be the need for anti-seizure medications and possibly mechanical ventilation. The introduction of intravenous acyclovir for treatment of herpes encephalitis has dramatically improved patient outcomes regarding the reduction in the mortality. Unfortunately, morbidity continues to remain high as a result of neurologic impairment. Even with early diagnosis and initiation of intravenous acyclovir, only about 40%–55% of patients are capable of returning fully to their activities of daily living without the need for assistance. Suppressive therapy following the 14 to

21-day treatment course is not recommended as a randomized controlled trial demonstrated that 3 months of suppressive therapy with valacyclovir did not improve neurological outcomes as clinically seen in neonates following 6 months of suppressive therapy. The study was terminated at an interim analysis due to clear lack of evidence of clinical benefit.

Unlike herpetic encephalitis, Mollaret meningitis, a type of aseptic meningitis caused primarily by recurrent HSV-2 orolabial or genital infections, does not require antiviral therapy as it is usually self-limiting. Differentiation between the diagnoses can be made through the presence of degenerated monocytes in cerebrospinal fluid, which is a diagnostic clue for Mollaret meningitis, and recurrence of orolabial or genital lesions.

Genital Herpes

Genital herpes can be acquired through genital-genital, oral-genital, or genital-anal contact with a partner who either has active lesions at time of contact or is asymptotically shedding virus. Most patients are unaware of their diagnosis allowing for transmission of the virus when patients are asymptotically shedding the virus. All lesions concerning for HSV should be swabbed and sent for polymerase chain reaction (PCR) testing to confirm the diagnosis. The lesions should also be typed for either HSV-1 or HSV-2 because virus type predicts the frequency of recurrences. Infection caused by HSV-1 is significantly less likely to recur. Serologic testing on each patient should be considered in order to determine if the infection is primary (patient has no antibodies to either type of HSV), first-episode non-primary (patient has antibodies to HSV type 1 or type 2 and newly acquires the other type), or recurrent to provide appropriate counseling in regards to natural history of the infection and risk of recurrence. All patients should be screened for other sexually transmitted infections, particularly human immunodeficiency virus (HIV), as herpes simplex virus is a known risk factor for the transmission of HIV.

Treatment of genital infections includes treatment for a primary infection, episodic treatment for recurrent infections, and daily suppressive therapy for prevention of recurrent infections and reduction in subclinical viral shedding. First-line antivirals used in the treatment of genital herpes are nucleoside analogs: acyclovir, valacyclovir, and famciclovir. All three nucleoside analogs have the same efficacy, and patients may prefer one over the other depending on convenience of administration and cost. Oral therapy (or IV if severe infections) is preferred over topical therapy as it is more effective in promptly alleviating symptoms, reducing time to healing, and reducing time of viral shedding if started within 24–72 h of symptom onset. A primary (initial) infection should be treated with antiviral drugs for 7–10 days or longer if lesions are not completely healed. The most convenient antiviral prescribed for treatment of genital lesions is valacyclovir as it only requires daily dosing (Table 1). This is because of the increased bioavailability of valacyclovir when compared to oral acyclovir. However, despite convenience in administration, valacyclovir is more costly than acyclovir, and therefore, may be a contributing factor when determining which antiviral to prescribe. In addition to antiviral therapy, patients may try analgesics and warm sitz baths to aid in alleviation of symptoms.

Once patients have completed treatment for their primary infection, the decision on whether to treat each symptomatic recurrence (episodic treatment) or to be placed on suppressive therapy is dependent on each patient's preference. Episodic treatment is typically a 5-day course started at the first sign of symptoms to treat current reactivation, whereas suppressive therapy is a daily medication regimen prescribed to prevent recurrences and reduce the duration of asymptomatic viral shedding (Table 1). Shorter course therapy has been shown efficacious but may be associated with gastrointestinal upset. As noted above, patients should be aware that lesions secondary to HSV-1 have fewer symptomatic recurrences and this may also aid in decision making. Suppressive therapy does appear to be of benefit to patients and their sexual partners by reducing number of symptomatic occurrences by approximately 80% and subclinical shedding, thus, improving their quality of life. One study by Gupta *et al.* in 2004 demonstrated the decrease in recurrences and subclinical shedding in participants on acyclovir or valacyclovir in comparison to no treatment (placebo). Although the study did show a reduction in subclinical shedding, patients should be aware that despite taking a daily medication, there is still a risk of asymptotically shedding the virus and, ergo, transmitting infection. On the other hand, episodic therapy may be preferred by some patients as they are not required to remember to take a daily medication. If a patient chooses episodic treatment, they should be provided with a supply of antiviral drugs in order to start the medication at the earliest onset of prodromal symptoms in order to be the most effective. Resistance to nucleoside analogs is uncommon despite the use of suppressive therapy over long term and is mainly a concern in immunocompromised hosts.

Counseling patients on their new diagnosis is as important as prescribing antiviral therapy secondary to the stigma associated with infection and the psychological effects patients experience afterwards. Many patients experience guilt, anger, low self-esteem, and fear of transmitting to the virus to their partner or future child. Counseling should include discussion on preventing transmission of the virus to sexual partners during times of prodromal symptoms or active lesions and even when asymptomatic secondary to subclinical viral shedding. In order to reduce transmission of the virus, patients should consider refraining from sexual activity during times of active lesions, using condoms to prevent transmission, and taking daily antiviral therapy for suppression. Although these practices can reduce the risk of transmission, patients should be aware that none of these measures are completely effective at preventing transmission. Little is known about the use of antivirals for pre-exposure and post-exposure prophylaxis for a seronegative partner; therefore, it is not currently recommended. Patients of long-term relationships should be offered the opportunity to have their partner participate in the counseling session in order to also educate the partner on the disease and ways to prevent transmission. The counseling session should also discuss the ubiquitous nature of the virus in the United States and that acquisition of the virus could have occurred years prior and not be a result of infidelity.

Table 1 Treatment options for genital herpes

<i>Lesion type</i>	<i>Treatment recommendations</i>	<i>Notes</i>
Primary lesion	Oral acyclovir: 400 mg 3 times a day for 7–10 days OR 200 mg 5 times a day for 7–10 days Valacyclovir: 1000 mg 2 times a day for 7–10 days Famciclovir: 250 mg 3 times a day for 7–10 days	Treatment duration may be extended past 10 days if lesions are incompletely healed
Recurrent lesion		
Episodic therapy	Oral acyclovir: 400 mg 3 times a day for 5 days OR 800 mg 2 times a day for 5 days OR 800 mg 3 times a day for 2 days Valacyclovir: 500 mg 2 times a day for 3 days OR 1000 mg 1 time a day for 5 days Famciclovir: 125 mg 2 times a day for 5 days OR 1000 mg every 12 h for 2 doses OR 500 mg for 1 dose then 250 mg for 2 days	
Suppressive therapy	Oral acyclovir: 400 mg 2 times a day Valacyclovir: 500 mg 1 time a day OR 1000 mg 1 time a day Famciclovir: 250 mg 2 times a day	
Pregnancy	Oral acyclovir: 400 mg 3 times a day Valacyclovir: 500 mg 2 times a day	Should begin at 36 weeks gestation

Note: Workowski K.A., Bolan G.A., 2015. Sexually transmitted diseases treatment guidelines. Morbidity and Mortality Weekly Report: Recommendations and Reports 64 (RR-03), 1–137.

Due to the increased risk of vertical transmission of HSV from mother to infant, all pregnant females with a history of genital lesions should be advised to take acyclovir or valacyclovir around 36 weeks gestation, if not already on suppressive therapy, to decrease the risk of developing active lesions at time of delivery. Antiviral suppressive therapy can decrease the probability of active lesions at the time of delivery and therefore, prevent the need for cesarean delivery but it does not completely prevent subclinical viral shedding. Women should be advised that infants are still at risk of perinatal transmission. The American College of Obstetricians and Gynecologists has recommended cesarean delivery for any woman presenting with active genital lesions or prodromal symptoms (burning, tingling, and/or itching in genital area) at time of delivery whether these lesions represent a new primary infection, non-primary first episode infection, or a reactivation. A prospective cohort study demonstrated cesarean delivery at time of active genital lesions greatly reduced transmission of herpes simplex virus from mother to infant, but it did not completely prevent the development of neonatal herpes infection in neonates. Other factors such as duration of rupture of membranes, application of fetal-scalp electrodes, and subclinical shedding contribute to the risk of transmission. Currently the risk of performing cesarean sections in all women with a previous history of genital herpes outweighs the benefit, and thus, cesareans are not recommended unless there are active lesions or prodromal symptoms indicating a reactivation at time of delivery. Also, screening all pregnant women for HSV serologic status is not recommended. Thus, women who are presumed seronegative during pregnancy but are in a relationship with a known seropositive spouse should refrain from sexual contact during their third trimester to prevent acquisition of the virus in order to decrease risk of transmission to their newborn.

Neonatal Herpes Simplex Virus Infection

Early recognition and prompt initiation of antiviral therapy in infants has prevented further disease progression and significantly improved morbidity and mortality. Therefore, any neonate presenting with concerns for neonatal HSV infection should undergo a complete evaluation. The evaluation includes performing culture and PCR on any concerning mucocutaneous lesion, obtaining mucous membrane culture and PCR of eye, nasopharynx, mouth, and rectum, obtaining blood HSV PCR and serum alanine aminotransferase (ALT), and performing a lumbar puncture to evaluate cerebrospinal fluid (CSF) for HSV with PCR. Both the use of culture and PCR are currently recommended for the detection of HSV DNA in mucocutaneous lesions and the eye, nasopharynx, mouth, and rectum as the superiority of PCR has not yet been established in the detection of HSV DNA from these sites in

neonates. However, PCR is quickly replacing viral cultures in many laboratories due to the costs related to performing viral cultures, the known increased sensitivity of PCR in the diagnosis of HSV DNA in genital lesions, and the quicker turn around time (24 h for PCR vs 2-5 days for culture). Once the full evaluation has been completed, all infants should be empirically started on intravenous acyclovir pending results of all the viral HSV studies. Diagnostic results will provide the information needed to classify the infant into a subtype, which will then determine the duration of treatment. The three subtypes of neonatal HSV infection are: skin, eye, and mouth (SEM), central nervous system (CNS), and disseminated disease.

Once the diagnosis has been confirmed, all infants regardless of their disease classification should have an ophthalmologic exam and neuroimaging. Infants diagnosed with skin, eye, and mouth (SEM) disease should receive 14 days of intravenous therapy, and those with CNS or disseminated disease should receive a minimum of 21 days of intravenous acyclovir. Treatment was extended to 21 days in infants with CNS and disseminated disease secondary to a large proportion (40%) of infants continuing to detect HSV DNA in the CSF by PCR after 14 days of therapy with intravenous acyclovir. Any infant with CNS involvement requires a repeat lumbar puncture prior to completion of 21 days to determine if the evidence of virus, namely viral DNA, has been cleared from the CSF. If the HSV PCR remains positive near completion of 21 days, intravenous acyclovir should be continued for an additional week in the infant and another lumbar puncture will need to be repeated at the end of the week to reassess for clearance of the virus. Intravenous acyclovir should be continued until the HSV cerebrospinal fluid PCR is negative. Infants that require longer than 21 days of intravenous acyclovir have been shown to have worse outcomes. While on intravenous acyclovir, complete blood cell counts should also be obtained twice a week and renal function closely monitored to evaluate for any signs of neutropenia or renal toxicity respectively. If infants develop severe neutropenia (<500 cells/ μL) during the treatment course with intravenous acyclovir, the acyclovir dose can be decreased or granulocyte colony-stimulating factor (G-CSF) can be administered if the absolute neutrophil count remains low despite a reduction in the dose.

Dosing recommendations for intravenous acyclovir have been derived from studies performed by the National Institute of Allergy and Infectious Disease (NIAID) Collaborative Antiviral Study over the last 30 years. Acyclovir quickly came into favor in the treatment of neonatal HSV infection because of the ease in administration and less toxic side effects. Initially low dose acyclovir when studied in comparison to vidarabine did not show therapeutic superiority. Later studies demonstrated improved outcomes when high a dose of intravenous acyclovir was used during treatment. Currently acyclovir is in the only nucleoside antiviral agent approved in the use of neonates as further studies are pending on the pharmacokinetics and safety of valacyclovir in neonates.

Suppressive therapy with oral acyclovir should be administered after the completion of either the 14- or 21-day treatment course of intravenous acyclovir for a total of 6 months. Suppressive therapy, when given over a 6-month period, has demonstrated a reduction in the number of cutaneous recurrences and improved developmental outcomes in neonates. Infants should be closely monitored for neutropenia during the 6 month course with a complete blood cell counts at the 2nd and 4th week of administration and then monthly thereafter. If the patient develops severe neutropenia (<500 cells/ μL), prescribers should consider holding the patient's acyclovir while waiting for recovery of the patient's absolute neutrophil count. The suppressive dose should be adjusted monthly to account for the patient's rapid growth. Although suppressive therapy has improved outcomes for neonates, neurologic impairment in infants diagnosed with encephalitis continues to remain high.

Measures to prevent the transmission of HSV to infant include: cesarean delivery in those with active genital lesions limiting the use of fetal scalp electrodes, and preventing exposure to oral secretions in persons with an active cold sore or a younger child with gingivostomatitis. A guidance statement by the American Academy of Pediatrics also recognizes the risk of infants developing HSV disease when born to mothers with active genital lesions at time of delivery and provides recommendations on managing these infants in attempt to diagnose infants with HSV disease early before disease progression occurs. Any genital lesion concerning for HSV should be evaluated and, if positive, typed for HSV-1 or HSV-2. Serologic testing should then be performed on the mother because determination of a mother's serologic status is imperative in determining the risk of transmission to an infant. A mother with a first-episode primary infection (no antibodies to HSV-1 or HSV-2) is a greater risk of transmitting HSV to their infant than a mother with a first-episode non-primary infection (mother has antibodies to HSV type 1 or type 2 and newly acquires the other type) and recurrent infection (reactivation of virus mother already has antibodies to).

All infants born to a mother with an active genital lesion should have a mucous membrane culture and PCR obtained from conjunctiva, nasopharynx, mouth, and rectum and blood HSV PCR performed at 24 h of life. The 24 h of lifetime point is chosen to avoid possible transient maternal contamination. If the infant is born to a mother with known history of genital herpes and the infant remains asymptomatic after birth, acyclovir can be held, and the infant can be discharged as early as 48 h of life as long as all HSV testing remains negative, infant has available access to medical care, and parents receive education on the signs and symptoms of neonatal HSV disease to monitor for. If these criteria cannot be met, the infant should remain hospitalized until culture is confirmed negative or is negative at 96 h, whichever is first. However, if any HSV test returns positive, the infants should undergo further evaluation with lumbar puncture to assess for central nervous system involvement and serum ALT to assess for systemic involvement. If the additional testing is negative, the infant should receive 10 days of empiric treatment with intravenous acyclovir. However, if the results return abnormal, the infant should be treated with intravenous acyclovir for either 14 or 21 days depending on classification.

If the infant's mother has no prior history of genital herpes and presents with active lesions at time of delivery, serologic testing on the mother is extremely helpful in determining infant's risk as mentioned above. In addition to mucous membrane culture and PCR, the infant should also have a lumbar puncture and serum ATL performed. Intravenous acyclovir should also be initiated. If mother's genital lesion is confirmed to be primary or non-primary infection, and the infant's testing is all negative, the infant should be empirically treated for 10 days with intravenous acyclovir. If the testing is abnormal, the infant should be classified into SEM, CNS, or

disseminated disease and treated accordingly. If mother's genital lesion is confirmed to be a reactivation, and all HSV testing on the neonate is negative, the infant can be discharged home with parents after they receive education on monitoring for signs and symptoms of neonatal HSV to monitor for. However, if the surface cultures or HSV blood PCR return positive, but the CSF HSV PCR and ALT are negative, the infant should be continued on IV acyclovir for a total of 10 days. If the CSF HSV PCR and/or ALT are abnormal, the infant should be treated based on classification of neonatal HSV disease for a total of 14–21 days.

Immunocompromised Hosts

Oral or intravenous acyclovir as well as valacyclovir and famciclovir remain treatments of choice for management of HSV infections in immunocompromised hosts, depending upon severity of infection, risk of dissemination, and degree of immunosuppression. However, immunocompromised patients are known to have a high rate of resistant HSV strains to acyclovir and other nucleoside analogs (valacyclovir, famciclovir) because of their similar mechanism of action when compared to immunocompetent hosts. Thus, other antiviral agents (foscarnet or cidofovir) besides nucleoside analogs should be considered for treatment in these patients if disease progresses despite adequate antiviral therapy. Resistance generally emerges after long-term administration of the antiviral. Resistance can be overcome at times with using higher than the typical recommended doses of acyclovir. However, when resistance cannot be overcome, cidofovir and foscarnet are alternative therapeutic options. These antivirals do not require phosphorylation in order to be active as with nucleoside analogs, and therefore, they are effective alternatives in treating HSV resistant strains. Both cidofovir and foscarnet have significant toxic side effects, limiting their recommendation to a second line treatment alternative.

Immunoprophylaxis

Several different vaccines have been studied for effectiveness in both therapeutic and prophylactic trials of HSV-1 and 2, but to date, no vaccine has demonstrated significant efficacy. There are currently several other vaccines contenders being evaluated, but these have not yet been studied in clinical trials.

Further Reading

- Amir, J., Harel, L., Smetana, Z., Varsano, I., 1997. Treatment of herpes simplex gingivostomatitis with aciclovir in children: A randomised double blind placebo controlled study. *BMJ* 314 (7097), 1800–1803.
- Amir, J., Harel, L., Smetana, Z., Varsano, I., 1999. The natural history of primary herpes simplex type 1 gingivostomatitis in children. *Pediatric Dermatology* 16 (4), 259–263.
- Azher, T.N., Yin, X.T., Tajfirouz, D., Huang, A.J., Stuart, P.M., 2017. Herpes simplex keratitis: Challenges in diagnosis and clinical management. *Clinical Ophthalmology* 11, 185–191.
- Brown, Z.A., Wald, A., Morrow, R.A., *et al.*, 2003. Effect of serologic status and Cesarean delivery on transmission rates of herpes simplex virus from mother to infant. *Journal of the American Medical Association* 289 (2), 203–209.
- Gill, M.J., Bryant, H.E., 1991. Oral acyclovir therapy of recurrent herpes simplex virus type 2 infection of the hand. *Antimicrobial Agents and Chemotherapy* 35 (2), 382–383.
- Gnann Jr., J.W., Skoldenber, B., Hart, J., *et al.*, 2015. Herpes simplex encephalitis: Lack of clinical benefit of long-term valacyclovir therapy. *Clinical Infectious Diseases* 61 (5), 683–691.
- Gnann Jr., J.W., Whitley, R.J., 2016. Genital herpes. *New England Journal of Medicine* 375 (19), 1906.
- Gupta, R., Wald, A., Krantz, E., *et al.*, 2004. Valacyclovir and acyclovir for suppression of shedding of herpes simplex virus in the genital tract. *Journal of Infectious Diseases* 190 (8), 1374–1381.
- Kimberlin, D.W., 2004. Neonatal herpes simplex infection. *Clinical Microbiology Reviews* 17 (1), 1–13.
- Kimberlin, D.W., 2013. Acyclovir dosing in the neonatal period and beyond. *Journal of the Pediatric Infectious Diseases Society* 2 (2), 179–182.
- Kimberlin, D.W., Baley, J., Committee on Infectious Diseases, Committee on Fetus and Newborn, 2013. Guidance on management of asymptomatic neonates born to women with active genital herpes lesions. *Pediatrics* 131 (2), 383–386.
- Kolokotronis, A., Dourmas, S., 2006. Herpes simplex virus infection, with particular reference to the progression and complications of primary herpetic gingivostomatitis. *Clinical Microbiology and Infection* 12 (3), 202–211.
- Rooney, J.F., Straus, S.E., Mannix, M.L., *et al.*, 1993. Oral acyclovir to suppress frequently recurrent herpes labialis. A double-blind, placebo-controlled trial. *Annals of Internal Medicine* 118 (4), 268–272.
- Whitley, R., Baines, J., 2018. Clinical management of herpes simplex virus infections: Past, present, and future. *F1000Research* 7.
- Whitley, R.J., Roizman, B., 2001. Herpes simplex virus infections. *Lancet* 357 (9267), 1513–1518.

Management of Varicella-Zoster Virus Infections (*Herpesviridae*)

Andreas Sauerbrei, Jena University Hospital, Jena, Germany

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

aa	Amino acid	IgG	Immunoglobulin G
ABS	Adenine triphosphate-binding site	IgM	Immunoglobulin M
bp	Base pair(s)	MMR	Measles-mumps-rubella
BVDU	(E)-5-(2-bromovinyl)-2'-deoxyuridine	MMRV	Measles-mumps-rubella-varicella
CNS	Central nervous system	NBS	Nucleoside-binding site
DNA pol	Deoxyribonucleic acid polymerase	pOka	Parental Oka
EC ₅₀	50% effective concentration	PZN	Post zoster neuralgia
EDTA	Ethylene diamine tetraacetate	RNA	Ribonucleic acid
ELISA	Enzyme-linked immunosorbent assay	s.c.	Subcutaneous(ly)
FAMA	Fluorescence antibody membrane antigen test	STIKO	Standing Vaccine Commission of the Robert Koch Institute, Germany
gE	Glycoprotein E	TK	Thymidine kinase
HIV	Human immunodeficiency virus	TK-	Thymidine kinase negative
HSV-1	Herpes simplex virus type 1	TKa	Thymidine kinase altered
HSV-2	Herpes simplex virus type 2	TKr	Thymidine kinase reduced
i.m.	Intramuscular(ly)	UL28	Unique long 28
i.v.	Intravenous(ly)	UL36	Unique long 36
IFAT	Indirect immunofluorescence test	VZIG	Varicella-zoster immunoglobulin
IgA	Immunoglobulin A	VZV	Varicella-zoster virus

Glossary

Congenital varicella This syndrome may occur in the newborn after maternal chickenpox between the 5th and 24th weeks of gestation. As a consequence of intrauterine VZV infection, the characteristic clinical findings consist of dermatomal skin lesions, neurological defects, eye diseases, and/or limb hypoplasia.

Post zoster neuralgia Pain lasting longer than 4 weeks and occurring again after pain-free interval, caused by an irreversible necrosis of ganglion cells.

UL28 VZV gene 28 localized in the unique long gene region.

UL36 VZV gene 36 localized in the unique long gene region.

UN 3373 Label for shipping category B biological substances; category B includes most known microorganisms.

Laboratory Diagnosis

Submission of Samples

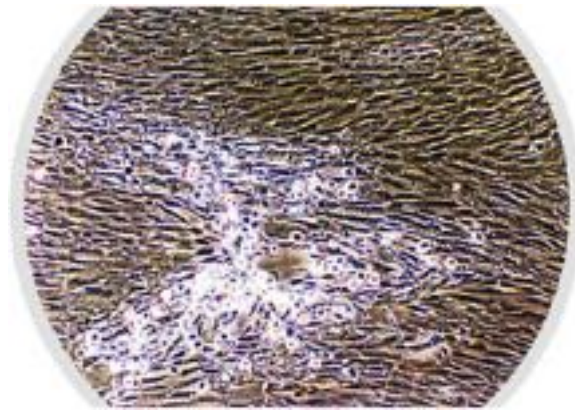
Varicella-zoster virus (VZV)-positive samples are considered as category B, risk group 2 dangerous goods and dispatched in accordance with UN 3373. For this purpose, the primary container with the patient's sample must be shipped with an overpack of adsorbent material in a transportable box. Shipment is possible at room temperature. Cooling is recommended only if the samples are intended for virus isolation in cell culture.

Direct Detection of Virus

Acute VZV infection is diagnosed by detection of the virus ([Table 1](#)). The method of choice is polymerase chain reaction (PCR) to detect viral genomes in vesicle fluid or swabs. Depending on the clinical manifestations, additional samples such as cerebrospinal fluid (CSF), tissue, bronchoalveolar lavage, ethylene diamine tetraacetate (EDTA) blood, amniotic fluid, aqueous humor, pharyngeal secretion, saliva or tear fluid are collected. PCR is used to examine CSF in acute infections of the central nervous system (CNS) and amniotic fluid in prenatal diagnostics after varicella during pregnancy. In immunosuppressed patients with zoster, the detection of VZV DNA in the blood may be helpful in assessing the potential risk of disseminating infection. In addition, virus isolation from vesicle contents or swabs can be used. However, isolation of VZV is only possible in a few cell types such as human embryonic fibroblasts ([Fig. 1](#)). Furthermore, this procedure is time-consuming, requires a high degree of experience and has no

Table 1 Overview of methods for direct detection of VZV or VZV DNA

<i>Principle</i>	<i>Method</i>	<i>Patient samples</i>
Detection of viral DNA	Polymerase chain reaction (PCR)	Vesicle content/swab, in 1 ml physiological saline or viral transport medium Cerebrospinal fluid, tissue, bronchoalveolar lavage, EDTA blood, amniotic fluid, aqueous humor, pharyngeal secretion, saliva, tear fluid
Viral isolation	Viral growth in cell culture, detection by monoclonal antibody	Vesicle content in viral transport medium with special swab, tissue, bronchoalveolar lavage; transport under cooling (2–8°C)
Virus antigen detection	Immunofluorescence test using monoclonal antibody Reduced sensitivity and specificity	Cell-rich vesicle content in viral transport medium with special swab, tissue
Discrimination between wild-type/vaccine strains, genotyping	PCR, Restriction fragment length polymorphism analysis, sequencing	Vesicle content in viral transport medium with special swab, tissue, cerebrospinal fluid, viral isolate

**Fig. 1** VZV-induced cytopathic effect in human embryonic lung fibroblasts.

clinically relevant sensitivity. In most cases, only vesicle fluid with a high viral load is suitable for virus isolation. For successful virus isolation, early and careful sampling and optimal transport of the sample are essential.

Viral isolates are characterized by immunofluorescence with monoclonal fluorescence-labeled antibodies. The direct qualitative detection of VZV antigens in vesicle content or swabs using commercial detection systems can provide results within a few hours but is characterized by reduced sensitivity and specificity. When interpreting the results, it should be noted that methods for the direct detection of VZV including nucleic acids or antigens do not distinguish between primary and recurrent VZV infections. The distinction between wild-type VZV and vaccine strains can be made by restriction fragment length polymorphism analysis or sequencing (genotyping).

Detection of Antibodies

VZV antibody detection (**Table 2**) is especially indicated when susceptible persons have to be identified to offer protection with active or passive immunoprophylaxis. Past VZV infection is diagnosed by detecting VZV IgG in serum or plasma, and positive VZV IgG indicates immunity. Due to high seroconversion rates, the determination of antibody status after varicella vaccination is not necessary for healthy children, adolescents and adults. In contrast, immune status control is recommended in immunodeficient individuals and healthcare workers after vaccination. In daily laboratory practice, ligand assays or partially immunofluorescence tests to detect VZV-specific IgG are common. Regardless of the test used, any result interpreted as VZV IgG-positive by that laboratory can be used as a criterion for immunity against varicella. Persons with indeterminate VZV IgG results are classified as “non-immune”. Commercial test kits differ in sensitivity so that very low antibody titers are not detected. Therefore, highly sensitive tests such as special glycoprotein enzyme-linked immunoassays (ELISA) or the fluorescence antibody membrane antigen test (FAMA) should be used to determine the immune status especially after varicella vaccination and for vaccine studies.

Primary VZV infection (varicella) can normally be diagnosed by determining VZV IgG seroconversion in serum or plasma. This requires the availability of sequentially collected samples where the initial result is VZV IgG-negative. The earliest VZV IgM can be detected from the fourth day after the onset of the disease, usually in combination with VZV IgG. Even though VZV IgM can be used in practice to confirm an active VZV infection, it should be noted that IgM antibody detection will be considerably delayed after the onset of varicella exanthema and is only possible in a maximum of 50%–60% of patients with zoster. Thus,

Table 2 Overview of methods for determination of VZV-specific antibodies

Method	Remarks
Ligand assays (ELISA, chemiluminescence immune assay etc.)	Determination and differentiation of Ig classes (IgG, IgM, IgA) in serum, plasma and cerebrospinal fluid, based on whole viral antigen of VZV-infected cell cultures or viral glycoproteins Simple performance, commercially available, automated
Indirect fluorescence antibody test (IFAT)	Determination and differentiation of Ig classes (IgG, IgM, IgA) in serum, plasma and cerebrospinal fluid Simple performance, commercially available, requires experience for evaluation
Fluorescence antibody membrane antigen test (FAMA)	Determination of antibodies against VZV glycoproteins in serum, reference test for determination of immunity
VZV IgG avidity (ELISA or IFAT)	Differentiation between primary infection (varicella) and recurrent infection after virus reactivation (generalized zoster)
Neutralization test	Determination of antibodies against VZV glycoproteins in serum, good correlation with FAMA

this marker is unreliable and unable to distinguish between a primary and a recurrent VZV infection. In addition, numerous commercial VZV IgM immunoassays show reduced sensitivity and may show false-positive results due to cross-reactions with other herpes viruses, especially herpes simplex virus (HSV). VZV IgA can often be determined in latently VZV-infected individuals, but high titer values correlate exclusively with zoster disease. In the case of VZV infection with involvement of CNS, the earliest increase in intrathecal antibody index (cerebrospinal fluid and serum from the same day of collection) can be detected around 10–12 days after symptom onset. Therefore, determining intrathecal VZV-specific IgG is only suitable for the retrospective diagnosis of VZV-associated CNS infections. The determination of VZV IgG avidity allows the differentiation between primary (varicella) and recurrent infections (zoster). However, there has been little experience with this test procedure so far. In principle, serological findings in immunodeficient patients are not reliable and should therefore be interpreted with caution. Despite acute or past infection, antibody findings in these individuals may also be negative.

Antiviral Treatment

Antiviral Agents in Clinical Use

The administration of antiviral agents may successfully block replication of VZV in infected cells. Early administration, especially in zoster within 72 h after onset of exanthema, can reduce tissue damage and thus diminish or even prevent the destruction of affected ganglion cells. The acyclic nucleoside analogs acyclovir including its prodrug valaciclovir, famciclovir (prodrug of penciclovir) and the cyclic nucleoside analog brivudine ((E)-5-(2-bromovinyl)-2'-deoxyuridine, BVDU) are available for the antiviral treatment of VZV infections (Table 3). The specificity of antiviral activity is based on the phosphorylation by the viral thymidine kinase (TK) to their mono- (acyclovir, penciclovir) or diphosphates (brivudine) while cellular enzymes catalyze the further phosphorylation steps to triphosphate. The activity spectrum is defined by the presence of the key enzyme, the viral TK. The triphosphates of the nucleoside analogs inhibit and fix the viral DNA polymerases (pol) and are incorporated into the growing DNA chain as a “false” substrate. In acyclovir/valaciclovir, this leads to chain termination due to the absence of the hydroxy group in the 3' position, which is essential for further linkage. In other nucleoside analogous compounds, their incorporation into DNA is possible.

Acyclovir

Acyclovir is used as the first-line antiviral drug to treat VZV infections. However, a disadvantage is that acyclovir has a low oral bioavailability of about 15%–30%. Varicella in high-risk patients and zoster disease in immunocompetent patients can be treated orally. In severe VZV infections, especially in immunocompromised patients, acyclovir must be administered intravenously (i.v.). Acyclovir is not officially licensed as antiviral therapy during pregnancy, and administration should be especially avoided in pregnant women before the end of the 14th gestational week. However, results from the manufacturer's pregnancy registries and a Danish population-based retrospective cohort study have not shown an increased rate of major birth defects after oral administration of acyclovir and its topical use. Nevertheless, patients should be informed about the limited data, especially during early pregnancy, and give consent before the drug is used. After i.v. administration of acyclovir, CNS side effects have occasionally been observed, while oral medication may be associated with gastrointestinal side effects. Drugs that can cause renal toxicity should not be combined with acyclovir at the same time. Laboratory kidney and liver parameters must be monitored. If kidney function is impaired, the dose of acyclovir must be reduced.

Valaciclovir

Valaciclovir is the L-valyl ester prodrug of acyclovir. After oral administration, valaciclovir is converted into acyclovir by a hepatic enzyme, valaciclovir hydrolase. Since the oral bioavailability of valaciclovir is 54%, the considerably higher acyclovir concentrations compared with oral acyclovir result in longer dosing intervals and higher compliance. Valaciclovir is approved for the

Table 3 Antiviral drugs against VZV in clinical use

Antiviral drug	Administration	Indication and dosage
Acyclovir	intravenously	Severe and disseminated VZV infections, zoster under immunosuppression; adults: $3 \times 5\text{--}10$ mg/kg body weight per day for 7–10 days, children: 3×500 mg/m ² body surface per day for 7–10 days
	orally	Zoster, varicella in risk patients; adults: 5×800 mg per day over 7 days, children: 5×15 mg/kg body weight per day (max. 4000 mg/day) for 7 days
Valaciclovir ^a	orally	Zoster; immunocompetent adults and adults with mild or moderate immunosuppression: 3×1000 mg per day for 7 days
Famciclovir ^b	orally	Zoster; immunocompetent adults: 3×250 mg per day for 7 days, immunodeficient adults or adults with zoster ophthalmicus: 3×500 mg per day for 7–10 days
Brivudine	orally	Zoster; immunocompetent adults: 1×125 mg per day for 7 days
Foscarnet	intravenously	VZV infections caused by TK-negative (acyclovir-resistant) VZV strains; adults: 3×40 mg/kg body weight per day for 10 (max. 20) days, children (off label use): 3×60 mg per kg body weight per day for 10 (max. 20) days

^aProdrug of acyclovir.^bProdrug of penciclovir.

antiviral treatment of zoster in immunocompetent adults and adults with mild or moderate immunosuppression. The drug is not approved for the antiviral treatment of VZV infections in childhood and adolescence. Valaciclovir is also not licensed in pregnant women. Although available data are reassuring as to the safety of valaciclovir in pregnancy, there is substantially less experience than with acyclovir. Possible side effects are comparable to those after taking acyclovir.

Famciclovir

Famciclovir is the inactive diacetyl ester prodrug of penciclovir, which results from the separation of two ester groups in the small intestine and liver. Penciclovir is an acyclic nucleoside analog derived from ganciclovir by replacement of the ether oxygen atom in the acyclic side chain by a methyl bridge. The oral absorption of penciclovir is very low. Therefore, this drug is only used for the antiviral treatment of local HSV infections. After oral administration, famciclovir has a bioavailability of 77%. Compared to acyclovir, the higher stability of penciclovir triphosphate might result in prolonged antiviral efficacy. Famciclovir is used to treat zoster in immunocompetent adults and immunocompromised patients from 25 years of age. As with valaciclovir, famciclovir is not approved in childhood, adolescence and pregnancy because of limited information on safety. In rare cases, famciclovir may cause headache, mental confusion and nausea.

Foscarnet (trisodium phosphonoformate)

The pyrophosphate analog foscarnet inhibits the viral DNA pol of numerous DNA and RNA viruses by preventing pyrophosphate exchange. Since foscarnet does not need to be metabolized for its antiviral activity, it is also effective against TK-negative VZV strains that are resistant to nucleoside analogs. For this reason, foscarnet is recommended as an alternative antiviral treatment if clinical resistance to acyclovir is suspected.

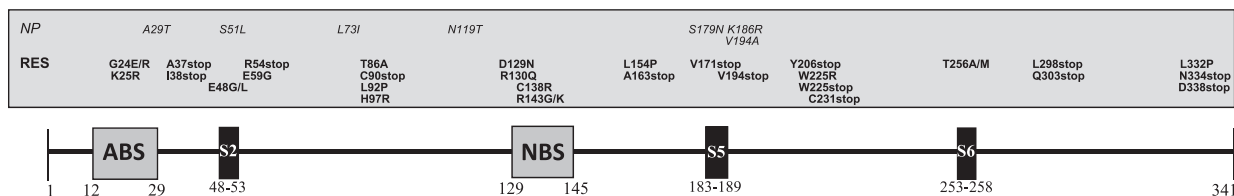
Brivudine ((E)-5-(2-bromovinyl)-2'-deoxyuridine)

The cyclic nucleoside analog brivudine is converted by viral TK into its mono- and diphosphate. Brivudine is administered orally and has a bioavailability of about 40%. It may be used to treat zoster in immunocompetent adults. Since the safety profile is not known due to a lack of studies, brivudine is not approved for antiviral therapy in children and adolescents. Therefore, the risk-benefit ratio should be carefully assessed before the substance is used in younger age groups, and parents must be informed (off-label use). In principle, brivudine is well tolerated. Nevertheless, gastrointestinal disturbances, impairment of kidney function, an increase in liver enzyme values and reversible changes in blood count may occur. Simultaneous administration of 5-fluorouracil or other 5-fluoropyrimidines leads to increased and potentially dangerous toxicity leading to deaths that have been reported in the past.

Development of Resistance

Varicella-zoster virus resistance to antiviral drugs such as acyclovir or foscarnet is rare and has been described in the literature only in immunocompromised patients with zoster, e.g., acquired immunodeficiency syndrome or immunosuppression by cancer or transplantation. The disturbed immune response and the long-term administration of antiviral drugs given as treatment or chemoprophylaxis are key for the development of resistance. The weakened immune response leads to prolonged virus replication, and resistant virus mutants can occur more frequently due to an increased number of natural spontaneous mutations. Under antiviral treatment, resistant viruses are selected and cannot be eliminated by the impaired immune response. In general, resistance is associated with non-synonymous mutations located in the gene of the target molecule or in the gene of proteins responsible for the metabolism or efficacy of antiviral agents. For acyclovir and its nucleoside analogs, resistance is based almost exclusively on non-synonymous mutations of the TK gene (*UL36*) and rarely on amino acid (aa) changes in the DNA pol gene (*UL28*), mostly associated with resistance to foscarnet. In contrast to TK, which is not required for the replication of VZV, the DNA pol is an essential enzyme within the viral replication cycle. According to current knowledge, VZV strains that are resistant to acyclovir due

VZV thymidine kinase



VZV DNA polymerase

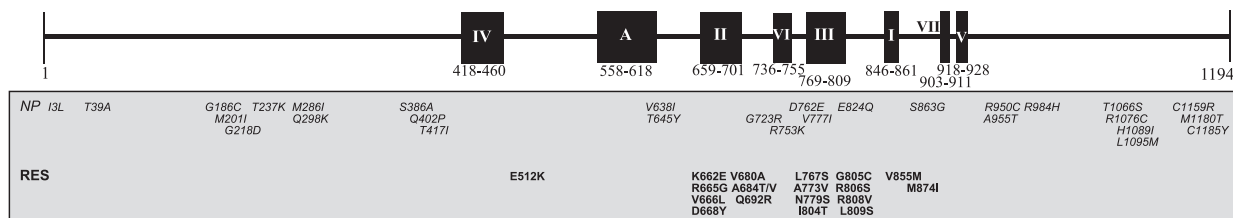


Fig. 2 VZV thymidine kinase (above) and DNA polymerase (below) with functional domains (gray boxes; ABS: ATP-binding site; NBS: Nucleoside-binding site) and conserved regions (black boxes). In the gray boxes above or below the corresponding protein, the confirmed amino acid substitutions from the literature related to the natural gene polymorphism (NP) or resistance (RES) to acyclovir (thymidine kinase) as well as acyclovir, foscarnet and cidofovir, respectively, (DNA polymerase) are summarized.

to mutations in the TK gene are always cross-resistant to famciclovir and brivudine. To date, there are only a few studies in the literature in which the significance of TK and DNA pol mutations has been proven by phenotypic findings.

Thymidine kinase gene

The VZV TK gene is 1026 bp in size and encodes 342 aa (Fig. 2). This gene contains two functional domains, a nucleotide (adenosine triphosphate)-binding site (aa 12–29) and a nucleoside (substrate)-binding site (aa 129–145), and three further conserved regions. Unlike the TK of HSV-1 and HSV-2, few natural polymorphisms have been identified in the VZV TK, but it is important to distinguish each of them from resistance mutations. Compared to the frequently used reference VZV strain Dumas (GenBank accession No. X04370.1, clade 1), all European VZV wild-type strains include the aa polymorphism Serine288Leucine.

Similar to HSV, the resistance mutations of the VZV TK gene are assigned to three phenotypes:

- (1) TK negative (TK⁻, no TK activity detectable, occurs most frequently).
- (2) TK reduced (TK^r, reduced TK activity, 1%–15% of normal activity).
- (3) TK altered (TK^a, altered TK substrate specificity, no phosphorylation of acyclovir and other nucleoside analogs).

Resistance to acyclovir may be caused by stop codons, frameshifts (deletions or insertions) or aa substitutions within and outside conserved gene regions. Since only a few validated resistance mutations have been reported in the literature, new or unknown mutations must always be expected in the genotypic analysis of clinically or phenotypically resistant VZV strains.

DNA polymerase gene

The VZV DNA pol gene is 3585 bp in size and encodes 1195 aa (Fig. 2). There are eight conserved regions with the designations I to VII and A. Similar to the TK gene, only a few natural polymorphisms have been reported for the DNA pol. Resistance-related aa substitutions are mostly localized in conserved gene centers. So far, little research has been carried out on natural polymorphisms and resistance-related mutations of the VZV DNA pol gene.

Resistance Testing

Antiviral therapy failure is assumed for VZV infections if no clinical improvement can be observed within 10–21 days after starting antiviral drugs, usually acyclovir. This means that there is reason to believe that they are resistant virus strains. In these cases, alternative antiviral treatment with foscarnet is indicated. In parallel, genotypic resistance tests of viral TK and possibly DNA pol should be carried out and, if a virus isolate can be detected in the cell culture, phenotypic resistance tests should also be performed. Due to it taking at least 3 weeks to grow, phenotypic resistance testing has usually no clinical relevance, but the procedure can help to characterize new mutations or gene variations that are not yet clear as to their significance for resistance.

Phenotyping

Phenotyping is regarded as the gold standard for antiviral resistance testing, but this procedure cannot be performed in most cases of suspected clinical resistance because VZV isolation in cell culture has a low sensitivity. The plaque reduction assay has been established as the method of choice. After the addition of the antiviral compound to be tested in a descending dilution series, the 50% effective concentration (EC_{50}) is estimated, resulting in 50% inhibition of viral replication. To assess the possible resistance, a sensitive VZV reference strain, e.g., pOka strain, should be tested in each experiment in parallel as a control. It is a decisive advantage that phenotyping allows a clear interpretation of the results. However, the methods used are time-consuming, material-intensive and non-standardized. In practice, phenotypic resistance tests can only be performed if swabs are obtained from vesicle fluid from which the virus can be isolated in cell culture. For results interpretation, the most common and reliable method for nucleoside analogs is to classify VZV strains as resistant if the measured mean EC_{50} value is three to five times higher than the corresponding value of the sensitive control strain. For resistance to foscarnet, EC_{50} values $> 330 \mu\text{M}$ were found to be solid.

Genotyping

As with HSV, the VZV genotyping resistance test is performed by amplification and sequencing of TK and DNA pol gene sequences. To identify non-synonymous mutations, sequence data must be compared with the published sequences of a sensitive reference strain available in the gene bank (e.g., VZV strain Dumas, accession number X04370.1). The advantages of genotyping are a significantly shorter delay of approximately 2 days compared to phenotyping and a direct examination of patient samples. The limited amount of viral DNA can have a limiting effect. A major drawback is the fact that only a little information is available on proven resistance-associated aa substitutions. Therefore, only stop codons or frameshift mutations can be interpreted with high probability with respect to resistance. Furthermore, the analysis of genotypic resistance can be difficult if a mixture of viral strains with different genotypes is present. Therefore, in clinical practice, it is a problem to define a questionable resistance of VZV strains on the basis of genotypic results. Recently, phenotypic testing of recombinant VZV isolates has been shown to be the best method to validate the significance of any resistance mutations.

Prevention

Prevention of Varicella by Vaccination

All available varicella vaccines are live-attenuated vaccines, and nearly all are based on the VZV Oka strain. There is only one non-Oka vaccine (SuduvaxTM, Green Cross, South Korea) available. The Oka strain (parental Oka, pOka) was isolated in the early 1970s from the vesicle fluid of a three-year-old child with chickenpox, whose surname Oka was used to designate this strain. Oka-derived varicella vaccines induce both humoral and T-cell immunity. The seroconversion rate is used as a criterion for assessing vaccine immunity, i.e., the presence of VZV IgG 6–8 weeks after vaccination in individuals who were VZV IgG-negative pre-vaccination. In healthy individuals, the seroconversion rate was calculated to be more than 95% after a single vaccine dose and 98%–99% after two doses, while in at-risk patients, the rate was 80%–90%. However, clinical efficacy was estimated to be partially lower depending on the duration of the post-vaccination period. Taken together, the efficacy of the vaccine in healthy children was between 80% and 97% and 93%–96% for the first and second doses, respectively.

After universal varicella vaccination was introduced in the United States in 1995, the frequency of varicella was reduced by 90% by 2008. In all age groups, the frequency of varicella fell to 71%–84% by the year 2000, and hospitalization due to varicella decreased threefold to fourfold. In the monitored regions, the average vaccination rates for children were approximately 80%. The pronounced herd immunity also led to a reduction of varicella incidence in non-vaccinated individuals. Studies suggest that the antibodies induced by the vaccine persist in a high percentage of vaccinees over decades. As a result of decreasing or insufficient immunity after a one-dose vaccination, a breakthrough disease from 43 days after vaccination can occur after massive virus exposure. Thus, varicella outbreaks in schools and day-care centers, as well as breakthrough diseases in vaccinated persons, have been reported despite high vaccination coverage in the United States after one-dose vaccination. Because there is only adequate protection if the varicella vaccine is administered twice, a two-dose vaccination scheme was introduced in the early 2000s. A shift in varicella to a higher age is not to be expected with vaccination coverage of $\geq 70\%$ comparable to those in the United States and Germany. Reservations against varicella vaccination are mostly justified by the scenario that a decrease in the incidence of varicella can lead to a reduced spread of wild-type viruses and that older people can develop zoster more frequently. In the United States, an influence of varicella vaccination on the incidence of zoster in adults could not be observed, and the current data from different countries confirm these findings. However, zoster has been shown to be 3–12 times less common in children vaccinated against varicella than in unvaccinated children. The disease in vaccinated children tends to be clinically milder and the exanthemas often occur anatomically in the dermatomes corresponding to the sites where the varicella vaccine was given.

In July 2004, universal varicella vaccination was included in the vaccination schedule for all children and adolescents in Germany. By 2011, vaccination coverage had risen to almost 70%. Initially, a single vaccination with monovalent varicella vaccines was recommended for children aged between 1 and 13 years. Since 2009, the recommendation of the Standing Vaccine Commission of the Robert Koch Institute (STIKO, Berlin, Germany) has provided for a two-dose schedule as the standard for varicella vaccination. In subsequent years, this recommendation was implemented without reducing vaccination rates. The first dose of varicella vaccine is administered at the age of 11–14 months, either at the same time as the first measles-mumps-rubella (MMR) vaccination or at the earliest 4 weeks later. The second dose of varicella vaccine is recommended at the age of 15–23 months, preferably with the

Table 4 Recommendations for indication immunizations against varicella in Germany

<i>Risk groups</i>	<i>Persons</i>	<i>Remarks</i>
Individually (not occupational) increased exposure, disease or complication risk, protection of others	Seronegative women who want children	Two doses of varicella vaccine
	Seronegative patients before planned immunosuppressive therapy or organ transplantation	Special hints for vaccination in patients under immunosuppressive therapy without VZV-specific antibodies should be considered
	Susceptible persons with severe neurodermitis	“Susceptible persons” means: No vaccination and no varicella history or anti-VZV IgG-negative in case of serologic testing
	Susceptible persons with close contact to the two aforementioned	
Occupational	Seronegative persons (including trainees, interns, students and volunteers) in the following fields of activity: • Medical facilities including facilities of other human medical health care professions • Activities involving contact with potentially infectious material • Care facilities • Community facilities • Facilities for the joint accommodation of asylum seekers, persons obliged to leave the country, refugees and late repatriates	Total of 2 vaccinations (if MMR vaccination is indicated, use MMRV combination vaccine if applicable)

combined measles-mumps-rubella-varicella (MMRV) vaccine. Since a slightly increased risk of febrile seizures was observed 5–12 days after the first vaccination with the combined MMRV vaccine, the information currently available suggests that separate administration of trivalent MMR vaccine and monovalent varicella vaccine should be preferred.

For unvaccinated 5–17-year-olds without a varicella history, two-dose varicella or MMRV vaccination is recommended as a catch-up vaccination. The minimum interval between two doses of the varicella or MMRV vaccine should be 4–6 weeks. The varicella vaccine is recommended as vaccination for all individuals summarized in [Table 4](#). Currently, the monovalent varicella vaccines Varilrix® from GlaxoSmithKline and Varivax® from MSD Sharp & Dohme GmbH as well as the tetravalent MMRV vaccine Priorix-Tetra® from GlaxoSmithKline are available in Germany. All vaccines are administered subcutaneously (s.c). The recommendations provide varicella vaccination as post-exposure prophylaxis for non-vaccinated individuals without varicella history after contact with individuals at risk. Vaccination should be carried out within 3 days after the onset of the exanthema in the index case or within 5 days after contact with the index case. Sentinel results from the “Varicella” working group at the Robert Koch Institute showed a decrease in varicella incidence of 80%–90% by 2012, affecting all age groups. However, the age groups most affected were those for which the vaccine recommendations applied (1–4 years) or where the vaccination was highly effective (5–9 years). It is significant that infants susceptible to varicella also benefit from herd immunity.

Important contraindications of varicella vaccination are intensive immunosuppressive therapy and pregnancy including the period 4–6 weeks before a planned pregnancy. However, according to current knowledge, there is no risk of prenatal malformations if the varicella vaccination is given accidentally during or shortly before pregnancy. In very rare cases, the transmission of the vaccine virus has been reported from immunocompetent vaccinated individuals to susceptible contacts persons. This is basically possible if the vaccine develops an exanthema with vesicles through which the virus can be transmitted. In contrast, immunodeficient patients with varicella caused by the vaccine virus may have a higher risk of virus transmission. Persons at risk for severe varicella, including pregnant women and neonates of mothers with no history of varicella, should therefore avoid contact with vaccinees. However, pregnancy is not considered a contraindication for vaccinating an unprotected child. After varicella vaccination, side effects such as redness and swelling at the injection site and pain at the injection site may occur in about 4% of vaccinations. 3%–5% of vaccinated children develop skin lesions localized at the injection site. In a further 3%–5% of children, less pronounced varicella-like exanthema may occur 2–6 weeks after vaccination, but adolescents and adults are affected twice as often as children. Varicella-like exanthemas within 2 weeks after vaccination are usually caused by wild-type virus infections. Later exanthemas up to 42 days after vaccination are usually associated with the vaccine virus. Accidental vaccination of seropositive individuals is not associated with an increasing number of side effects.

Passive Immunoprophylaxis of Varicella

Passive VZV immunoprophylaxis with varicella-zoster immunoglobulin (VZIG) can prevent the onset of varicella or considerably attenuate the course of the disease. Therefore, the administration of VZIG was recommended for susceptible at-risk patients after exposure to varicella or zoster. These include the groups summarized in [Table 5](#). If non-vaccinated pregnant women without a varicella

Table 5 Susceptible at-risk patients for whom the administration of VZIG (inclusive dosage and point in time) is recommended after exposure to varicella or zoster in Germany

<i>Risk patients</i>	<i>Dosage</i>	<i>Point in time</i>
Unvaccinated pregnant women without varicella history	VZIG i.v.: 25 IU/kg (0.1 ml/kg/h for 10 min, with good tolerability increase to maximum 1 ml/kg/h)	As soon as possible and not later than 96 h after exposure
Immunocompromised patients whose immune status against varicella is uncertain or negative	VZIG i.m.: 15 IU (normally 0.2 ml) per kg, divided between different parts of the body if more than 2 ml for children up to 20 kg, and more than 5 ml for adults	
Newborns whose mothers develop varicella within 5 days before and 2 days after birth	1 ml/kg i.v. (or 0.5 ml/kg i.m.)	Immediately after birth or outbreak of exanthema in the mother
Preterm infants from 28 weeks gestation whose mothers have no immunity, after exposure during the neonatal period	1 ml/kg i.v. (or 0.5 ml/kg i.m.)	As soon as possible and not later than 96 h after exposure
Preterm infants younger than 28 weeks after exposure during the neonatal period, regardless of the immune status of the mother		

history are passively immunized, it should be noted that the main reason for administering immunoglobulin is to protect them from severe varicella infections and complications. So far, there is no evidence that this cost-intensive approach prevents the spread of fetal disease in the form of congenital varicella syndrome. However, one should be aware that the administration of VZIG is the only way to prevent congenital varicella syndrome after viral exposure of an unprotected woman during the first 20 weeks of pregnancy.

An important prerequisite for the efficacy of passive immunoprophylaxis is the timely administration of VZIG. While previous instructions recommended administration within 3 days and a maximum of 10 days after the beginning of exposure, the current STIKO recommendation recommends “as soon as possible and not later than 96 h after exposure”. An “exposure” was defined as (1) one hour or longer together with an infectious person in a room, (2) personal contact or (3) household contact. For the administration and dosage of VZIG (Table 5), the manufacturer's instructions must be considered. If necessary, the post-exposure application of VZIG can be combined with antiviral chemoprophylaxis.

Prevention of Zoster by Vaccination

The age-related increase in zoster incidence correlates with the decrease of specific T-cell immunity. Therefore, an attempt was made to stimulate the specific cellular immunity in elderly people and thus enable the prevention of zoster. Inspired by the success of the varicella vaccine, a live-attenuated zoster vaccine was developed at the beginning of 2000 containing at least 14-fold higher concentrations of the vaccine virus than in the varicella vaccine. A large study has shown that vaccination of adults can reduce the incidence and severity of zoster by approximately 50% and the incidence of post zoster neuralgia (PZN) by 67%. According to the available information, protective immunity is maintained for at least 7 years. The vaccine is well tolerated. Site effects are local reactions at the injection site such as redness, swelling, pain and touch sensitivity. Under the brand name Zostavax®, the first vaccine for the prevention of zoster and PZN for people aged 50 and over was approved in Europe in 2006. Sanofi Pasteur MSD (Lyon, France) holds the European marketing authorization. The vaccine is administered s.c. as a single dose. Like all live vaccines, this zoster vaccine is contraindicated for immunosuppressed patients and pregnant women. The vaccine can be administered at the same time as the inactivated influenza vaccine while it should not be administered at the same time as the pneumococcal vaccine. Zostavax® has been available in Germany since September 2013. Although the zoster vaccination was part of the public vaccination recommendations in different federal states of Germany, the STIKO has not issued an official recommendation to use a live-attenuated zoster vaccine for standard vaccination. Accordingly, the vaccination with the live zoster vaccine was not an affordable service within health insurance. The main reason for this was that the effectiveness of the vaccination decreases with increasing age and ranges from 70% in the 50–59-year-olds to 41% in the 70–79-year-olds to less than 20% in the ≥ 80-year-olds. Thus, the protection period of the vaccination was documented only for a few years. Furthermore, modeling results from clinical studies showed only a small reduction of the total number of zoster cases by vaccination, which can be between 2.6% (vaccination with 50 years) and 0.6% (vaccination with 80 years) depending on the age of the vaccinated person. In addition, immunosuppressed people who have a significantly increased risk of developing zoster and its complications cannot be vaccinated with the attenuated live vaccine. The epidemiological risk-benefit assessment of the vaccination with Zostavax® did not lead to the recommendation of a standard vaccination in Germany. However, the vaccination of an individual patient can make sense after an individual risk-benefit assessment.

Several years ago, a second zoster vaccine based on recombinant adjuvant VZV-specific glycoprotein E (gE) was developed, which is of fundamental importance for the development of VZV-specific immunity. A liposome-based adjuvant serves as an enhancer of immunity. Since March 2018, this new zoster subunit inactivated vaccine designated Shingrix® and manufactured by

GlaxoSmithKline has been approved in Europe for the prevention of zoster. After administration of two doses, this inactivated vaccine reduced the frequency of zoster between 90%–97% in immunocompetent individuals over 50 years of age in phase III studies, and the vaccine efficacy against PHN was 89% in adults over 70 years of age. So far, two studies on immunodeficient patients have also been conducted. After three vaccine doses at one and three-month intervals, respectively, in 121 adult cancer patients who had undergone an autologous stem cell transplantation, one month after the last immunization (6 months after transplantation), humoral and cellular gE-specific immunity was significantly increased compared to placebo and persisted for up to one year. A similar study of 123 adult HIV patients also demonstrated an increased specific cellular and humoral immune response persisting for 18 months. All available data demonstrated that the adjuvant zoster subunit inactivated vaccine could effectively prevent zoster and PZN in immunocompetent people 50 years of age and older. Efficacy across all age groups from 50 years of age is 92% for protection against zoster and 82% for protection against PZN. Vaccination is safe but very reactive. Local and systemic reactions such as pain at the puncture site or fever, fatigue, headache and myalgia of 1–2 days often occur. It is of great advantage that the vaccine also induces a robust humoral and cellular immune response in immunodeficient individuals. Since December 2018, STIKO, in Germany, has recommended vaccination with the adjuvant zoster subunit inactivated vaccine for general use as a standard vaccine for persons aged 60 years and older and as an indication vaccine for immunosuppressed persons and patients with other severe basic diseases (e.g., rheumatoid arthritis, chronic kidney disease, chronic obstructive lung disease, diabetes mellitus) 50 years and older to prevent zoster, its complications and late sequelae. The vaccination series consists of two intramuscular (i.m.) vaccinations at intervals of at least 2 months and up to a maximum of 6 months. This vaccination is no substitute for an indicated chickenpox vaccination.

Further Reading

- Cunningham, A.L., 2016. Efficacy of the herpes zoster subunit vaccine in adults 70 years of age or older. *New England Journal of Medicine* 375, 1019–1032.
- Damm, O., 2015. Systemic review of models assessing the economic value of routine varicella and herpes zoster vaccination in high-income countries. *BMC Public Health* 15, 533.
- De Clercq, E., 2013. Selective ant-herpesvirus agents. *Antiviral Chemistry & Chemotherapy* 23, 93–101.
- Robert Koch Institute, 2017. STÄNDIGE IMPFKOMMISSION (STIKO): Wissenschaftliche Begründung für die Entscheidung, die HEPES-ZOSTER-Lebendimpfung nicht als Standardimpfung zu empfehlen. *Epidemiologisches Bulletin* 36, 391–410.
- Robert Koch Institute, 2018. STÄNDIGE IMPFKOMMISSION (STIKO): Wissenschaftliche Begründung zur Empfehlung einer Impfung mit dem HERPES-ZOSTER-subunit-Totimpfstoff. *Epidemiologisches Bulletin* 50, 541–567.
- Robert Koch Institute, 2019. EMPFEHLUNGEN der STÄNDIGEN IMPFKOMMISSION (STIKO) beim ROBERT KOCH-INSTITUT – 2019/2020. *Epidemiologisches Bulletin* 34, 313–357.
- Leroux-Roels, I., 2012. A phase 1/2 clinical trial evaluating safety and immunogenicity of a varicella zoster glycoprotein e subunit vaccine candidate in young and older adults. *Journal of Infectious Diseases* 206, 1280–1290.
- Lopez, A.S., 2014. Two-dose varicella vaccination coverage among children aged 7 years – Six sentinel sites, United States, 2006–2012. *Morbidity and Mortality Weekly Report* 63, 174–177.
- Morfin, F., 2003. Herpes simplex virus resistance to antiviral drugs. *Journal of Clinical Virology* 26, 29–37.
- Oxman, M.N., 2005. A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults. *New England Journal of Medicine* 352, 2271–2284.
- Pasternak, B., 2010. Use of acyclovir, valacyclovir, and famciclovir in the first trimester of pregnancy and the risk of birth defects. *Journal of the American Medical Association* 304, 859–866.
- Piret, J., 2016. Antiviral resistance in herpes simplex virus and varicella-zoster virus infections: Diagnosis and management. *Current Opinion in Infectious Diseases* 29, 654–662.
- Sauerbrei, A., 2014. Windpocken (Varizellen). In: Deutsche Vereinigung zur Bekämpfung der Viruskrankheiten e.V. (DVV e.V.), Gesellschaft für Virologie e.V. (GfV e.V.). (Eds.), *S2k-Leitlinie Labordiagnostik schwangerschaftsrelevanter Virusinfektionen*. Berlin, Heidelberg: Springer. pp. 95–110.
- Sauerbrei, A., 2001. Resistance testing of clinical varicella-zoster virus strains. *Antiviral Research* 90, 242–247.
- Sauerbrei, A., Wutzler, P., 2007. *Varicella-Zoster Virus-Infektionen: Aktuelle Prophylaxe und Therapie*, second ed. Bremen/London/Boston: Uni-Med.

Treatment and Prevention of Herpesvirus Infections in the Immunocompromised Host

Sara H Burkhard and Nicolas J Mueller, University Hospital of Zurich, Zurich, Switzerland

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Immunodeficiencies are caused by the defective function of one or several components of the immune system. They can be divided into genetically inherited primary deficiencies and acquired secondary immune deficiencies. Primary immunodeficiencies can go unnoticed due to the lack of symptoms or result in serious and recurrent infections. Although severe defects are rare, they often lead to high mortality in early childhood and are predominantly inherited in a recessive and X-chromosome linked manner. Secondary immune defects are much more common. They can be caused by environmental factors such as malnutrition or result from diseases like infections and cancer, or the administration of cytotoxic drugs. In such conditions, the immune defect is often unwanted, while in some cases, the immunosuppression is induced with therapeutic intention. Exemplary, organ transplantation would not be possible without immunosuppressive drugs because a fully functional immune system would cause the rejection of the allograft.

Immunodeficiencies can also be distinguished by the nature of the underlying defect. Primary immunodeficiencies are currently classified into nine categories defined by the signaling pathway or immune cell type affected. The study of these distinct genetic defects and associated diseases has aided the understanding of the immune response to various pathogens. We now know that the control of viral infections is mainly dependent on T cell function and, to a lesser extent, Natural killer (NK) cells and antibodies. Secondary immune defects roughly cluster into two main groups and a combination thereof. In the first group, patients are treated with cytotoxic drugs and radiation therapy, which results in the impaired generation of neutrophil granulocytes and mucosal barrier dysfunction. The deficiency in this first line of immune defense usually predisposes individuals to bacterial and fungal infections. The second group includes patients in whom T cells are targeted, yielding a defect in cellular immunity. This is usually the case in patients treated with immunosuppressive drugs to contain autoimmune disorders or to prevent the rejection of a transplanted organ.

A T cell defect is also the underlying cause of the immunosuppressive state of patients suffering from acquired immune deficiency syndrome (AIDS) caused by the infection with the human immunodeficiency virus (HIV). The HIV pandemic, which started in the 1980s, raised awareness of the diverse types of opportunistic pathogens that take advantage of the impaired T cell function. Viral infections were observed frequently, many of them being caused by members of the Herpesviridae family. The high incidence of herpesvirus infections among AIDS patients is, to some degree, explained by the ability of herpesviruses to latently persist in a large percentage of the population. As seen in HIV infected subjects, the loss of T cell surveillance can then lead to herpes virus reactivation, replication, and subsequent disease. Classically, AIDS patients would present with cytomegalovirus (CMV) retinitis, Epstein-Barr virus (EBV)-associated lymphoma or the human herpesvirus (HHV)-8 associated Kaposi sarcoma. Thanks to more efficient antiretroviral therapies, HIV replication can now be controlled in the great majority of patients, which leads to recovery of T cell function. For patients with access to treatment, this results in a similar susceptibility to infections as seen in the general population.

There is a rising number of patients treated with immunosuppressive drugs. This is explained by an increase in the number of conditions requiring immunosuppression and the prolonged survival of patients under improved therapeutic regimens. For autoimmune disorders, the blockade of specific pathways involved in the pathogenesis of disease has proven successful. This may be achieved by monoclonal antibodies or small molecule inhibitors. Although such patients may still suffer from opportunistic infections, the therapeutic immune defect is usually more limited, often not resulting in an elevated susceptibility to infection. Transplantation medicine is another field that has seen major advances in the last few decades but still requires a more general immunosuppressive approach. This immunosuppression is achieved by the combination of immunosuppressive drugs that are usually administered for a lifetime. Although subsequent development of a certain immune tolerance to the transplanted organ allows reducing drug dosages, patients are particularly vulnerable in the early months after transplantation. In this phase, induction therapy and initiation of high dose maintenance immunosuppressive therapy prevent acute organ rejection. Just like in antiviral immunity, allograft rejection is mainly mediated by T cells. Upon transplantation, T cells are activated by non-self major histocompatibility complexes (MHC). To limit the cellular immune response, donor and recipient MHC classes are matched before transplantation. However, the lack of organs eligible for transplantation and the urgency of finding a recipient for an organ from a deceased donor limit accurate MHC matching. This, in turn, results in the need for more potent immunosuppression. Also, the adaptive response to minor histocompatibility antigens is not accounted for pre-transplantation and is thought to mirror immunity to viral proteins. It is evident that the immunosuppression that prevents allograft rejection will lead to an increased susceptibility to viral diseases, and again herpes virus infections are at the forefront. Therefore, we will focus on the presentation, management, and treatment of herpesvirus infections in transplanted patients.

Cytomegalovirus (Human Herpesvirus 5)

Cytomegalovirus (CMV) is one of the most common infections occurring after organ transplantation and is related to significant morbidity and loss of graft function. In the immunocompromised host, CMV infection is often more severe,

causing meningoencephalitis, retinitis, gut inflammation, hepatitis, pneumonitis and polyradiculopathy. These end-organ diseases are the result of the invasion and local tissue damage by CMV. However, most often, CMV disease presents with non-specific symptoms such as fever, malaise, decreased leukocyte and platelet count, or elevated liver function values. This symptom complex is usually referred to as a viral syndrome.

Pre-Transplant Risk Stratification

The main strategy of disease prevention begins with the risk stratification before transplantation. This is performed by serological testing, which is achieved by measuring donor and recipient CMV IgG. Thereby, the clinician can evaluate whether donor and recipient have been exposed to CMV previously. For seronegative organ recipients, serological testing is periodically repeated before transplantation as primary infection and subsequent seroconversion could occur.

Upon solid organ transplantation the highest risk of infection is observed when a seropositive donor is combined with a seronegative recipient (D^+/R^-). In this case CMV can be transmitted *via* the organ transplanted into a host lacking CMV immunity. In this setting, the risk of CMV disease post-transplantation is as high as 56%. Intermediate risk is expected in seropositive recipients (R^+) due to CMV reactivation upon immunosuppression. Simultaneous infections with different CMV serotypes are thought to explain the marginally elevated risk in the D^+/R^+ compared to the D^-/R^+ setting. D^-/R^- serostatus is linked to the lowest risk in adult recipients, as *de novo* CMV infection is rare. Prior to hematopoietic stem cell transplantation, recipient immune cells specific for CMV are usually eliminated due to induction therapy. Thus, post-transplantation immunity depends on donor immune cells. Studies have identified a survival advantage for recipients receiving an allotransplant from a seropositive donor compared to those receiving stem cells from seronegative donors. This protective effect was eliminated by T cell depletion, suggesting a central role of CMV specific donor-derived T cells for the control of CMV infection.

The serological risk stratification for CMV before transplantation has a direct impact on patient management and prevention of disease after transplantation. It may, for instance, guide the indication and duration of prophylactic measures taken. Due to the inaccuracy of disease prediction by the D/R risk stratification, there is a need for alternative methods. Assessment of the CMV-specific T cell quantity and functionality has been investigated and shows promising results. Methods involve enzyme-linked immunosorbent assays (ELISAs) or enzyme-linked immunosorbent spots (ELISpots) measuring interferon (IFN) γ release by T cells activated with CMV peptides, some of which are commercially available. Assays using intracellular cytokine staining or MHC-multimer staining of CMV specific T cells rely on analysis by flow cytometry. Performing such T cell assays before transplantation does not take into consideration the influence of the immunosuppression after transplantation, often severely impairing T cell responses. Currently, there is limited literature on the use of pre-transplantation risk stratification by T cell assays.

Prevention Strategies of CMV Disease After Transplantation

The two approaches of “universal prophylaxis” and “pre-emptive therapy” have been proven valuable to prevent CMV disease after solid organ transplantation. For universal prophylaxis, patients are given an antiviral drug in reduced dosage. The administration should be started shortly after transplantation, as this is the time where the threat of acute rejection demands the most potent immunosuppressive regimen. Prophylaxis is continued for the first 3–12 months after transplantation depending on the transplanted organ type, immunosuppressive regimen and serological risk stratification. In the D^-/R^- group, with the lowest risk, preventive therapy is not recommended. The prophylactic approach harbors the risk of late-onset CMV disease after discontinuation of antiviral prophylaxis.

The late-onset disease affects about 20% of patients and is a particular concern to patients with a D^+/R^- serostatus, severe immunosuppression and transplant rejection. Another drawback of universal prophylaxis is that patients are exposed to antiviral treatment and adherent adverse effects for a substantial time period. This is particularly critical for subjects undergoing stem cell transplantation where the toxic effects of some antiviral drugs on the bone marrow limit the use of universal prophylaxis. In those populations, clinicians can opt for preemptive therapy.

This approach involves the frequent screening for CMV replication after transplantation, followed by antiviral treatment should viral DNA be detected. Early antiviral intervention is essential as it can limit the severity of CMV related disease. Surveillance is usually performed by the measurement of CMV DNA by quantitative PCR in plasma or whole blood. The positive result is commonly referred to as DNAemia, as the majority of CMV DNA detected in the blood was demonstrated to be fragmented and non-functional.

Since the World Health Organization standardization of qPCR protocols to measure DNAemia, there is a higher degree of assay reproducibility between laboratories. There is, however, no agreement on what DNA threshold should trigger antiviral treatment and whether this threshold should be adapted to the serological CMV risk. It has been suggested that the change in the level of DNAemia over time should be considered to be relevant in predicting disease.

In case CMV qPCR is not available, measurement of the CMV antigen pp65 by ELISA can be a valuable alternative. In contrast, measurement of CMV IgG is not useful for post-transplantation surveillance as serological responses are impaired or delayed due to the immunosuppression. The preemptive strategy was shown to have similar overall outcomes regarding CMV disease and graft function as the universal prophylaxis for most types of transplants. While preemptive therapy is associated with a higher rate

of CMV DNAemia, late-onset CMV disease and side effects of treatment are more frequently observed in patients receiving prophylactic medication. Thus, the choice of approach is determined by the individual risk for CMV disease and antiviral toxicity.

The main drug used for CMV prevention is the oral prodrug valganciclovir, while intravenous ganciclovir is administered on rare occasions. These drugs are ideal for prophylaxis as they also protect against other herpes virus infections, *e.g.*, varicella-zoster virus or herpes simplex virus. Unfortunately, myelosuppressive adverse effects limit the use of ganciclovir and valganciclovir after hematopoietic stem cell transplantation and are frequently observed after solid organ transplantation. Letermovir, an inhibitor of the terminase complex of CMV, was recently approved for CMV prophylaxis in hematopoietic stem cell transplant recipients. In clinical trials using letermovir, hardly any myelotoxicity was observed. Importantly, letermovir has no antiviral activity against other herpes viruses; hence additional administration of prophylactic antiviral compounds may be required. Alternative substances are currently being evaluated in clinical trials.

The use of intravenous immunoglobulins or CMV-specific IgG was suggested to reduce the incidence and severity of CMV disease. Due to the better control of CMV replication under ganciclovir or valganciclovir, the sole application of immunoglobulins for prophylaxis has become uncommon. However, it may be useful for high-risk patient populations, given in combination with antiviral compounds or when managing subjects with contraindications for common antiviral drugs.

Currently, monoclonal CMV antibodies are in development. Various vaccines are also being engineered and some have proceeded to early clinical studies. Vaccination approaches involve the use of live attenuated virus, viral vectors, recombinant subunits or gene-based vaccines, some of them providing promising results. Several vaccinations were shown to induce the production of neutralizing antibodies, lower DNAemia, preventing infection or reducing CMV episodes.

Diagnosis of CMV Disease and Treatment

In the case of a viral syndrome, the association of non-specific symptoms and detection of CMV DNA in the blood usually prompts the initiation of therapy. Tissue invasive CMV disease, in contrast, does not always cause elevated DNAemia. This is most frequently observed in gastrointestinal disease or pneumonitis. Therefore, detection of the virus in biopsy material should be performed whenever possible to confirm the suspected diagnosis. Histological criteria such as inclusion bodies or detection of viral antigen by immunohistochemistry are used to diagnose localized CMV disease.

Once the diagnosis has been established, oral valganciclovir and intravenous ganciclovir are the preferred agents. Both antiviral drugs have been shown to achieve comparable long-term outcomes. As oral therapy is more convenient, intravenous treatment with ganciclovir is reserved for severe disease and where oral bioavailability may be impaired, for example, in gastrointestinal disease due to rapid intestinal passage or mucosal damage.

Treatment should be continued for a minimum of two weeks. To monitor therapy success, weekly blood CMV DNA load testing is performed. Usually, treatment is only stopped when two negative DNA measurements have been obtained or DNAemia drops below a certain threshold. For tissue invasive disease treatment can be stopped when clinical symptoms have ceased, but a two-week course is usually given to all patients. Recent studies suggest that the assessment of CMV specific T cell immunity may assist the decision to stop the administration of antiviral substances in primary infection. Using a CMV specific ELISA for IFN γ release by T cells, antiviral treatment could safely be stopped when a CMV specific immune response had been detected. Similarly, intracellular staining-based assays demonstrated the inability to clear CMV DNAemia in the absence of T cell responses.

Apart from pharmacological treatment of CMV disease, optimizing host factors can support immune control. Lower blood concentrations for calcineurin inhibitors – an immune-suppressing agent frequently used after solid organ transplantation – have been associated with more rapid clearance of DNAemia. In contrast, immunosuppressive regimens containing a mammalian target of rapamycin (mTOR) inhibitor may be advantageous due to their suggested direct anti-CMV effect. In severe disease, the reduction of all immunosuppressive agents should be considered.

Upon successful treatment of the disease, some patients will develop recurrent CMV disease. After solid organ transplantation, CMV relapses affect up to 35% of patients. Certain risk factors are connected with the development of CMV relapses. These include primary infection, deceased-donor transplantation, high-level CMV DNAemia, delay in CMV DNA regression after the initiation of treatment, multimorbidity, transplant rejection, strong immunosuppressive regimens, thoracic organ transplants and invasive gastrointestinal disease. Patients at risk should be closely monitored for CMV DNAemia and symptoms after discontinuation of therapy. To further limit CMV relapses, the administration of prophylactic medication after resolved CMV disease was trialed. As subsequent studies demonstrated that this approach of secondary prophylaxis failed to prevent relapses, it is not generally recommended. Still, many centers have adopted this strategy.

Antiviral Drug Resistance

Antiviral drug resistance is rarely an issue in immunocompetent hosts, as CMV infection does not usually require treatment. In contrast, immunocompromised patients may suffer from a clinical disease that cannot be controlled by the suppressed immune response. Thus, antiviral therapy is initiated and exerts a selection pressure required for drug-resistant viral subpopulations to propagate. Drug-resistant CMV first emerged among patients with AIDS, where the loss of drug susceptibility complicated as many as 20% of cases after 9–12 months of antiviral treatment. Optimizing treatment regimens reduced the frequency of antiviral drug resistance, which now occurs in 5%–12% of transplant recipients.

Certain host and therapy-related factors are connected with an increased risk of developing CMV drug resistance. These include prolonged antiviral drug exposure and persistent CMV replication under therapy due to impaired CMV immunity. The latter is often observed in the D⁺/R⁻ risk group, in patients on more potent immunosuppressive regimens or with inadequate CMV treatment. The emergence of ganciclovir resistance seems to be connected to antiviral drug doses used for CMV treatment, especially if it is repeatedly started and stopped, as it is only observed in up to 3% of the patients given prophylactic doses. Drug resistance should be suspected and investigated when patients do not improve or relapse under treatment. Although resistance rarely develops during the first 6 weeks of antiviral therapy, testing is recommended in unresponsive patients after two weeks of therapy. However, it is not unusual that DNAemia increases after initiation of therapy during the first week.

Single or multiple mutations lead to different levels of drug resistance and further mutations can be acquired by the virus over time. Testing for resistance is performed by sequencing of viral DNA and comparison with the wild-type sequence. As with all sequence-based resistance testing, mutations can reflect mere sequence polymorphism or confer to drug resistance. The amount of evidence for any given mutation to result in resistance reaches from theoretical predictions to *in vitro* assays or description of clinical cases. The level of resistance is often reported as the change of drug concentration required to suppress viral propagation by 50%. These fold-changes are then categorized so that a 2–5 fold increase refers to a low-grade, 5–15 fold is moderate and > 15 means high-level resistance. The numbers obtained for single mutations can be added should multiple mutations be detected.

Genotypic assessment of resistance requires a certain quantity of CMV DNA. Thus, resistant viral subpopulations of low frequency may not be detected, leading to the assumption that the virus is drug susceptible. Also, resistant CMV subpopulations have been isolated from specific organs, while the mutation was not detected in CMV DNA isolated from blood. Applying technologies such as deep sequencing may improve the detection of drug-resistant subpopulations in future.

Sequencing of CMV codons 400–670 and 300–1000 are usually performed as these contain the most known mutations involved in resistance in the UL97 kinase and the UL54 DNA polymerase, respectively. In 90% of patients on ganciclovir treatment, the UL97 kinase mutation appears before the UL54 polymerase mutation. UL54-associated resistance may signify cross-resistance to other antiviral drugs and often leads to a loss of viral fitness.

If a previously unknown mutation is detected by sequencing, resistance can be assessed by recombinant phenotyping. For this assay, the mutated gene is transferred to a laboratory CMV strain that is then tested for phenotypic resistance to antiviral substances. This is a timely and labor-intensive process and is therefore rarely used clinically.

In the case of low-level resistance, an increase in ganciclovir dosage and improving host factors, such as reducing immunosuppression, may be sufficient to overcome resistance. If genetic testing reveals a high-level of resistance, switching the antiviral agent is recommended. In case of life-threatening disease, the antiviral drug foscarnet is added prior to genetic testing for resistance. The salvage therapy using foscarnet is often successful, but metabolic side effects and nephrotoxicity limit its use. Newer compounds such as letermovir, maribavir and brincidofovir are still not approved for the treatment of CMV disease but may be available for compassionate use. The use of CMV hyperimmunoglobulin and intravenous immunoglobulin has thus far only been reported in case studies.

The idea of improving cellular immunity towards CMV has fueled the study of adoptive T cell transfer. Recently, this approach was reported to be successful in a number of case studies. An adoptive T cell transfer may play a role in complex cases as the generation of autologous CMV T cells takes several weeks and is costly. Therefore, the creation of HLA matched T cell banks is of great interest. Several agents have been used for off label treatment due to their anti-CMV effects *in vitro*. Such drugs include the mTOR inhibitors sirolimus and everolimus, leflunomide and artesunate, and addition to antiviral therapy can be considered in complicated cases.

HSV (Herpes Simplex Virus, Human Herpesvirus 1 and 2)

Immunosuppressed patients experience more severe herpes simplex virus (HSV)-associated disease and shed virus more frequently. Primary infection post-transplantation has been described, but disease mostly results from reactivation of latent HSV after a previous infection. Therefore, serological testing before transplantation is useful to assess the individual post-transplant risk.

Most patients present with typical mucosal and skin lesions, even after immunosuppression, allowing diagnosis on clinical grounds. Lesions can, however, be atypical and include unusual locations. This can render clinical discrimination from lesions caused by other diseases difficult. Consequently, a diagnosis should involve the detection of HSV DNA by PCR from samples collected in mucocutaneous lesions. PCR was shown to be the most sensitive test for HSV detection. Although less sensitive, direct fluorescent antibody assays deliver rapid results. More severe HSV diseases, such as disseminated mucocutaneous disease, visceral disease, esophagitis, hepatitis, and pneumonia, are also observed. Just as in immunocompetent hosts, retrograde reactivation can lead to disease of the central nervous system. Obtaining tissue, cerebrospinal fluid or blood samples for the detection of HSV DNA confirms the diagnosis.

Many transplant recipients receive prophylactic valganciclovir or ganciclovir to prevent CMV disease, effectively averting HSV reactivation or infection. For patients not receiving CMV antiviral prophylaxis, or when a preemptive strategy is adopted, HSV prophylaxis may be considered in HSV seropositive transplant recipients. This is also true for the use of agents for CMV prophylaxis that do not have anti-HSV-activity, as is the case with letermovir.

Preventive measures should cover the first months after transplantation because this is when the majority of HSV infections occur. Prophylaxis and treatment involve antiviral medication with aciclovir and valaciclovir. With localized disease, oral regimens

are preferred and are continued for a minimum of 5–7 days and until the HSV lesions have resolved. For severe manifestations, intravenous administration of aciclovir and the reduction of immunosuppressive medication should be considered.

As immunosuppressed patients are exposed to longer courses of antiviral treatments, aciclovir resistance is an emerging issue. Frequency of drug resistance in such patients can range from 2.1% to 10.9% and it is more often observed in AIDS patients or after hematopoietic stem cell transplantation compared to patients receiving solid organ transplants. Drug resistance should be considered in the case of treatment-refractory disease and a history of previous exposure to antiviral therapy. Mutations causing aciclovir resistance are most frequently detected within the thymidine kinase UL23, implying a cross-resistance to other antiviral substances such as famciclovir, valaciclovir and ganciclovir. Both genetic testing and phenotypic assays are used to identify mutations causing drug resistance. While waiting for antiviral resistance to be confirmed, the antiviral regimen is usually switched to foscarnet or cidofovir. After successful treatment, aciclovir resistance is sometimes lost in virus strains detected upon HSV reactivation.

Recurrent disease is defined as two or more HSV reactivations after the first episode. In this case, suppressive pharmacological therapy can be considered. Such prophylactic treatment was shown to be associated with a lower frequency of aciclovir resistance when compared to repetitive treatment of relapses.

Varicella Zoster Virus (VZV, Human Herpesvirus 3)

Primary infection (chickenpox) is most often seen in children and rarely in adults. On the contrary, Varicella-zoster virus (VZV) reactivation (shingles, herpes zoster) is observed frequently particularly in the immunosuppressed host. Most patients with immune defects present with a classical manifestation of VZV infection. The frequency of severe symptoms, such as multi-dermatomal herpes zoster, a prolonged disease course and complications is, however, elevated in immunocompromised patients.

Rarely, VZV reactivation can affect internal organs, most commonly causing pneumonia, encephalitis and hepatitis, referred to as visceral herpes zoster. In some cases of visceral disease, patients may lack skin lesions. If skin lesions are present, they are more likely to be atypical in appearance. For more uncommon presentations, PCR is the most sensitive test to detect viral replication in the tissue or body fluids. As a distinction from HSV infection may be difficult clinically, multiplex assays detecting both HSV and VZV have been developed. Alternatively, fluorescent labeling of viral antigen can be used in tissue biopsies or scrapings from vesicles and ulcers.

Assessing previous VZV exposure by serology VZV is suggested for patients on transplant waiting lists. The result is used to estimate the post-transplantation risk of infection but also determines if patients are eligible for vaccination and what vaccine should be used. Live attenuated Oka vaccine may be offered to susceptible individuals before transplantation. For seropositive patients above 50 years of age, vaccination with either the live-attenuated Oka herpes zoster vaccine or the adjuvanted subunit herpes zoster vaccine is recommended. These vaccines are partially effective as they can reduce the incidence and severity of disease but fail to prevent herpes zoster completely. Recent recommendations favor the recombinant subunit over the live attenuated vaccine as it shows superior efficacy against herpes zoster and elicits a more long-lasting effect in immunocompetent subjects.

Post-transplantation live attenuated vaccines are contraindicated but the subunit vaccine for herpes zoster can still be safely applied.

Initiation of herpes zoster treatment is suggested up to 72 h after the appearance of first cutaneous lesions, but later administration can be considered in immunosuppressed patients on an individual basis. Antiviral substances include aciclovir and valaciclovir. Famciclovir and brivudine have not been tested in immunocompromised patients, and with brivudine, interaction with 5-Fluoropyrimidine must be avoided. Disease severity determines whether antiviral drugs should be administered orally or intravenously. Therapy should be continued for a minimum of 7 days and until skin lesions have fully crusted. As for HSV, drugs used to prevent CMV disease, namely ganciclovir, valganciclovir and foscarnet, show anti-VZV activity.

Human Herpes Virus – 6 and 7

After solid organ transplantation, HHV-6 reactivation has been reported in 20% to 82% of transplant recipients. In the majority of reactivations, HHV-6B is identified. Mostly, reactivation occurs early after transplantation and does not cause any symptoms. In the case of symptomatic disease, fever and bone marrow suppression the most common presentations but rashes, hepatitis, gastro-duodenitis, colitis and pneumonitis have also been described. Limbic encephalitis caused by HHV-6 infection is mainly associated with hematopoietic stem cell transplantation.

For HHV-7, the frequency of reactivation is estimated to be seen in up to 46% of transplant recipients and also occurs in the early phase after transplantation. Symptomatic disease is rare and was primarily reported to cause fever and bone marrow suppression resulting in thrombocytopenia. The pathogenic potential of HHV-7 remains unclear as HHV-7 replication is detected in many asymptomatic immunosuppressed patients.

HHV-6 (and probably HHV-7) infections have been associated with fungal infections, accelerated liver fibrosis in hepatitis C virus infections after liver transplantation bronchiolitis obliterans post lung transplantation and allograft dysfunction or rejection. Initially, HHV-6 and -7 were also linked to a higher frequency of CMV-associated disease. Subsequent studies, however, failed to show a correlation between co-infection and the clinical course of CMV disease.

The preferred method of diagnosing infection is qPCR from peripheral blood or PBMCs. Due to DNA fragments from lysed cells, HHV-6 DNA detection by PCR from plasma only reaches a specificity of 84% for active viral replication. Furthermore,

diagnostics are limited by the lack of standardized assays, thresholds and by the fact that viral replication does not imply clinical disease. Consequently, the relationship between HHV-6 DNAemia and clinical symptoms is unclear. Further diagnostic testing such as HHV-6 DNA detection in tissue samples or radiological appearance may help establish the diagnosis.

Persistent high-level plasma HHV-6 DNA by qPCR may also suggest that integration of the HHV-6 genome into the germline has taken place and, thus, does not reflect viral replication. This rare event can be inherited, clusters within families, and approximately 1% of the European and US American populations are affected.

There are no randomized, controlled trials demonstrating antiviral treatment efficacy and recommendations based on *in vitro* data. Furthermore, no antiviral drugs have been licensed to treat HHV-6 and -7 related diseases. *In vitro* ganciclovir, foscarnet, cidofovir and brincidofovir have been shown to suppress HHV-6 replication. Several studies have described the successful treatment of HHV-6 using those antiviral drugs after solid organ transplantation. Therapy is recommended in the case of HHV-6 encephalitis and should be considered with a clinical disease that could be linked to HHV-6.

In the case of HHV-7, cidofovir and foscarnet have shown anti-viral activity *in vitro* while ganciclovir and aciclovir did not achieve this effect.

Screening for HHV-6 and -7 and prophylactic measures are not routinely recommended.

Epstein–Barr Virus (Human Herpesvirus 4)

Epstein–Barr virus (EBV) disease in the immunosuppressed individual ranges from systemic presentations such as infectious mononucleosis to more localized disease including hepatitis, pneumonitis, gastrointestinal, hematological manifestations, and neoplasia. The hematological disorders can be serious, causing hemophagocytic lymphohistiocytosis or macrophage activation syndrome, which results in the uncontrolled activation of immune cells.

An important property of EBV is its oncogenic capacity caused by the integration of viral DNA into the host cells' genome. During immunosuppression, the lack of cytotoxic T cell responses allows EBV to transform lymphocytes, potentially resulting in their uncontrolled proliferation. In particular, post-transplantation lymphoproliferative disorders (PTLDs) pose a great challenge. EBV-related PTLD usually arises from B cells but T and NK cell-derived PTLD is also seen. The latter may be caused by increased viral replication in the immunosuppressed, allowing for infection of cells normally not targeted by EBV. About 50% of PTLDs are EBV related and mostly occur within the first year after transplantation. This disease entity is decreasing in incidence, while late presenting EBV-negative PTLD is seen more frequently. A change in induction regimens partly explains this. The application of polyclonal anti-lymphocyte antibodies, which were administered in higher doses in the past, is a known risk factor for PTLD development. An elevated risk for PTLD is seen in EBV seronegative recipients who may contract a primary infection, making the pediatric population particularly vulnerable. This is why the serostatus before transplantation should be determined and most centers measure viral capsid antigen (VCA) IgG and Epstein–Barr nuclear antigen-1 (EBNA-1) IgG.

The type of organ transplant is another determinant of PTLD risk and the highest incidence is observed after hematopoietic stem cell transplantation, particularly transplantation of haploidentical cells. Good prediction models are, however, lacking, and the individual risk is difficult to ascertain.

Clinical presentation of PTLD can range from asymptomatic to fulminant organ failure and tumor lysis. Early signs include fever, lymph node or tonsil enlargement and changes in the hematogram. Due to organ damage such as hepatitis, colitis, pneumonia and cerebritis, symptoms can be very diverse and mimic an infectious disease process or organ rejection. PTLD frequently involves extranodal sites such as the gut, solid allograft and the central nervous system.

Diagnosis is established by histology from lesion biopsy and disease can be categorized into six subclasses defined by the World Health Organization. The detection of EBV nucleic acid or proteins in the tissue establishes a link to the EBV-associated pathogenesis. Monitoring for EBV-related disease is usually performed by measuring plasma EBV DNA by PCR. Generally, high-level EBV DNAemia or a rapid increase in blood EBV DNA load is correlated with an elevated risk of developing PTLD. However, there is a lack of standardization of thresholds, frequency and time points of viral load measurement.

Due to the lack of antiviral drugs with activity against EBV, preemptive strategies and therapy involve reducing the immunosuppressive regimen as an initial step. This conversely risks provoking an acute transplant rejection as seen in 37% of patients followed in a prospective trial. If this approach fails, administration of rituximab – a B cell depleting antibody directed towards CD20 – can be considered in CD20 positive PTLD. Rituximab was the first treatment to improve prognosis drastically. Combined with the reduction of immunosuppression, rituximab achieves a response rate of 44%–79% and complete remissions in 20% to 55%. If the PTLD is still unresponsive to treatment, the next step involves the use of chemotherapy. In rare cases of PTLD such as peripheral T cell lymphoma, Hodgkin's lymphoma, Burkitt's lymphoma and central nervous system lymphoma, chemotherapy is considered as first-line therapy. Further treatment modalities include the adoptive transfer of EBV-specific T cells, surgical resection and radiotherapy.

Human Herpes Virus-8

Due to discrepancies in the geographic distribution of HHV-8, seropositivity rates range from approximately 5% to 50%. This explains the variable frequency of HHV-8 associated pathologies. HHV-8-related disease can affect immunocompetent subjects but is more often related to immunosuppression. Sexual transmission is the predominant mode in HIV positive men who have sex

with men. Besides advanced HIV disease, post-transplant immunosuppression is a risk for progression to HHV-8-associated diseases. After organ transplantation in HHV-8 endemic regions, seroconversion in mismatched donors and recipients is seen frequently. HHV-8 DNAemia, however, usually arises from HHV-8 reactivation after immunosuppression and is rarer. Symptoms include fever, rash, lymphadenopathy and cytopenia but can also be more severe in the case of lymphoproliferative B-cell disorders, gammopathies, bone marrow failure and hemophagocytic syndrome.

Similar to EBV, integration of HHV-8 DNA explains the oncogenicity of this virus. Kaposi sarcoma, caused by the uncontrolled proliferation of mesenchymal cells, is the leading clinical presentation linked to the reactivation of HHV-8. Transformation of endothelial cells causes the formation of blood-filled vascular channels, giving the lesions the typical purple color. Kaposi sarcomas can present as blotches or nodules and are usually found on skin or mucosa but can affect visceral organs. HHV-8 also causes lymphoproliferative disease by infecting B cells at different stages of development. This includes multicentric Castleman disease, presenting with enlarged lymph nodes, flu-like symptoms and organ dysfunction, and primary effusion lymphoma, affecting body cavities often without forming a solid tumor mass. Altogether, these lymphoproliferative conditions are rare since the advent of effective antiretroviral treatment for HIV infected individuals. The mortality rate can reach 60% in the case of Kaposi sarcoma and is even higher for body cavity lymphomas. Serologic testing pre-transplantation can be considered to stratify the risk in endemic areas but is not generally recommended. Higher levels of immunosuppression, particularly anti-lymphocyte agents, increase the risk of HHV-8 disease. When disease is suspected, the diagnostic standard involves tissue biopsy and detection of virus-associated antigens or HHV-8 DNA.

Management of HHV-8 related disease usually starts by reducing the immunosuppression as this was shown to result in complete remission in 30% of cases. Favoring mTOR inhibitors over calcineurin inhibitors may be beneficial as regression of Kaposi sarcoma was reported with the use of drugs such as sirolimus. As for other herpes viruses, mTOR inhibitors may have a direct anti-HHV-8 activity but may also suppress tumorigenesis by inhibiting vascular growth. Due to the lack of prospective controlled trials, antiviral treatment is not routinely supported, although HHV-8 is susceptible to aciclovir, ganciclovir and cidofovir *in vitro*. For neoplastic disease, radiotherapy and chemotherapy may be part of treatment. For multicentric Castleman disease, treatment with rituximab appears promising even though the transformed B cells lack CD20 expression.

Further Reading

- Allen, U.D., Preiksaitis, J.K., 2019. Post-transplant lymphoproliferative disorders, Epstein-Barr virus infection, and disease in solid organ transplantation: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. *Clinical Transplantation* 33 (9), e13652.
- Dierickx, D., Habermann, T.M., 2018. Post-transplantation lymphoproliferative disorders in adults. *The New England Journal of Medicine* 378 (6), 549–562.
- Kotton, C.N., Kumar, D., Caliendo, A.M., *et al.*, 2018. The third international consensus guidelines on the management of cytomegalovirus in solid-organ transplantation. *Transplantation* 102 (6), 900–931.
- Lee, D.H., Zuckerman, R.A., 2019. Herpes simplex virus infections in solid organ transplantation: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. *Clinical Transplantation* 33 (9), e13526.
- Ljungman, P., Boeckh, M., Hirsch, H.H., *et al.*, 2017. Definitions of cytomegalovirus infection and disease in transplant patients for use in clinical trials. *Clinical Infectious Diseases* 64 (1), 87–91.
- Lurain, N.S., Chou, S., 2010. Antiviral drug resistance of human cytomegalovirus. *Clinical Microbiology Reviews* 23 (4), 689–712.
- Murphy, K., Waever, C., 2017. Chapter 13 – Failures of host defense mechanisms. In: Murphy, K., Waever, C. (Eds.), *Janeway's Immunobiology*, ninth ed. Garland Science.
- Murphy, K., Waever, C., 2017. Chapter 15 – Autoimmunity and transplantation. In: Murphy, K., Waever, C. (Eds.), *Janeway's Immunobiology*. Garland Science.
- Pellet Madan, R., Hand, J., 2019. Human herpesvirus 6, 7, and 8 in solid organ transplantation: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. *Clinical Transplantation* 33 (9), e13518.
- Pergam, S.A., Limaye, A.P., 2019. Varicella zoster virus in solid organ transplantation: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. *Clinical Transplantation* 33 (9), e13622.
- Razonable, R.R., Humar, A., 2019. Cytomegalovirus in solid organ transplant recipients: Guidelines of the American Society of Transplantation Infectious Diseases Community of Practice. *Clinical Transplantation* 33 (9), e13512.
- Wang, L., Verschuuren, E.A.M., van Leer-Buter, C.C., *et al.*, 2018. Herpes zoster and immunogenicity and safety of zoster vaccines in transplant patients: A narrative review of the literature. *Frontiers in Immunology* 9, 1632.

Management of Adenovirus Infections (*Adenoviridae*)

Albert Heim, Hannover Medical School, Hanover, Germany

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

c/ml Copies per milliliter; viral genome equivalents (copies) measured in human body fluids (most frequently blood, cerebrospinal fluid or urine), not harmonized to WHO standard preparations

HAdV Human (Mast-)Adenovirus

LRTI Lower respiratory tract infection

URTI Upper respiratory tract infection

Glossary

DNAemia Presence of viral DNA in peripheral blood, either cell-associated or in plasma. The term is used for viremias and are usually detected by nucleic acid amplification techniques e.g., Polymerase chain reaction (PCR).

Latency/latent infection Long term infection of a host with a virus without the production of infectious virus progeny, thus the host is neither affected by a disease nor infectious. On a cellular level, latently infected cells contain complete viral genomes but viral gene expression is usually

restricted to a few genes. Latent infection of a host can be lifelong (as in herpesvirus infections) or limited to a period of several months to years, as seen in adenovirus infections. Latency may lead to reactivation of virus replication with shedding of infectious virus and disease manifestations.

Virus load Concentration of virus DNA measured in peripheral blood samples (whole blood, plasma, or serum) for diagnostic purposes by quantitative nucleic acid amplification techniques (e.g., real-time PCR). Values are given as c/ml or if harmonized according to WHO standard preparations as IU/ml.

Taxonomy

Family and Genus

The seven species of human adenoviruses (HAdVs) (A to G) are members of the *Adenoviridae* family, genus *Mastadenovirus*, thus their correct name is human mastadenovirus according to the International Committee on Taxonomy of Viruses. However, their previous (and now vernacular) name HAdV is widely accepted and used in this article. HAdV species were initially defined by phenotypic properties (e.g., hemagglutination) but are nowadays clearly backed by molecular phylogenetic criteria based on complete genomic sequences. Therefore, HAdV types that belong to the same species (A to G, see [Table 1](#)) are more closely related to each other and frequently have similar tropism, disease association, and virulence. Tropism of HAdV types is influenced by their use of different cellular receptors they attach to with their highly variable fiber knob. Other factors that can contribute to their different tropism and virulence are the interaction of the penton base with the secondary cellular receptor, and early gene products interacting with regulators of the cell cycle, replicating the viral genome and interfering with the immune system. Moreover, recombination events between HAdV types of the same species are feasible, resulting in novel (geno-)types that may contain sequences coding for immunogenic epitopes of low prevalence (sero-)types within a genomic backbone of a more virulent and prevalent type ([Fig. 1](#)). Thus, these recombinant types can be positively selected for by an immune escape mechanism.

Types and Subtypes

Typing was based on serology (neutralization testing and as a surrogate technique hemagglutination inhibition testing) up to type 51 (serotypes), but later on, almost all novel types (genotypes) were defined by genomic criteria that include either recombinant phylogeny of genes coding for major capsid proteins or novel genes. Type numbers were assigned chronologically according to the first isolation of a novel HAdV. Therefore, low numbers are somewhat associated with more prevalent types and/or easy propagation on cell cultures but types with closely related numbers do not have to be closely related to each other (for adenovirus species and types, see [Table 1](#)). According to more precise but somewhat outdated differentiation techniques, genome types (not to be confused with genotypes) and intermediate types were defined on a subtype level to describe adenoviruses associated with high virulence and unusual disease associations or to study infection chains. Genome types were defined by restriction fragment length polymorphisms and intermediate types by different typing results in neutralization testing (based on the hexon neutralization epitope ϵ) and hemagglutination inhibition testing (based on the fiber knob epitope γ). Many of these genome types and intermediate types were reanalyzed by complete genomic sequencing, found to be recombinant adenoviruses fulfilling the genotype criteria for a novel HAdV type and reassigned with a new type number. For example, the genome type 19a – which is associated with epidemic keratoconjunctivitis – was reassigned as type number 64.

without compromising their replication in cell cultures. However, this deletion was never observed in clinical isolates because E3 gene products protect infected cells from the action of the immune system.

Epidemiology

HAdVs are prevalent worldwide and it can be assumed that every adult has survived infections with multiple types, but of course, not all types. HAdV infections of neonates are observed rarely because of maternal passive immunity against highly prevalent types. However, sporadic infections with types not covered by passive immunity may present as severe lower respiratory tract infections (LRTI) and even disseminated infections (see below) in neonates.

As maternal passive immunity wanes at an age of 6–12 months, toddlers get frequently infected with highly prevalent HAdV types of species C (e.g., types 1, 2, and 5), species B (e.g., 3 and 7), and species F (types 40 and 41). Infections with highly endemic species C HAdV types present as upper respiratory tract infections (URTI), sometimes in combination with gastroenteritis. Adenovirus latency in lymphoid cells of the adenoids and gastrointestinal tract is a sequel of these infections and can last for several years after primary infections but is not lifelong in contrast to herpesvirus latency. HAdV latency usually results in intermittent low-level shedding of HAdV DNA and infectious virus with respiratory secretions and feces. The epidemiological significance of this shedding remains obscure, but adenovirus latency may result in clinically significant reactivations in immunosuppressed individuals. Infections with species B types – also associated with respiratory tract infections and occasionally affecting the lower respiratory tract – can be more frequently observed as epidemics. Transmission of respiratory HAdV types is by droplets and smear infections but, in contrast to many other respiratory viruses, not limited to the typical winter season.

Adenoviruses are considered to be the third most frequent cause of viral gastroenteritis (after rotavirus and norovirus infection). Infections with species F adenovirus types frequently cause gastroenteritis outbreaks in toddlers and young children. At a slightly higher age (kindergarten and primary school), HAdV types of species A (12, 18, and 31) can cause small gastroenteritis outbreaks at these institutions. Transmission is by the fecal-oral route, frequently by smear infections, and difficult to control due to very high virus loads in feces (up to 10^{10} /ml). Therefore, nosocomial transmission chains can be observed.

Older children, adolescents, and young adults are prone to respiratory infections with the less prevalent HAdV species B types with respiratory tropism (e.g., 14, 21, and 55, in addition to 3 and 7) and infections with type 4 (species HAdV-E). These can cause epidemics in summer-camps and military barracks. The same types also cause pharyngoconjunctival fever and conjunctivitis outbreaks and some of these can be traced back to insufficiently chlorinated swimming pools.

As many adults are immune to the above-mentioned highly prevalent types, adenovirus infections mostly present as sporadic cases of individuals exposed to the types for which they are susceptible. For example, adults can also be affected in outbreaks of adenovirus types with respiratory tropism (e.g., 3, 4, 7, 14, 21, and 55). Genitourinary infections of adults – urethritis, for example – can be transmitted sexually and are caused by types usually associated with respiratory and eye infections (types 2 and 37, respectively) but these infections appear to be rather sporadic cases or small infection chains. Hemorrhagic cystitis can be observed both in adults and in children and is clearly associated with a few types of species HAdV-B (11, 34, and 35) but cases are rather sporadic. However, it is thought that these infections are underdiagnosed in mild, non-hemorrhagic cystitis cases. Seniors are frequently affected by large outbreaks of epidemic keratoconjunctivitis, which can also affect younger patients. Epidemic keratoconjunctivitis is caused by a few types of species HAdV-D (8, 37, 53, 54, and 64 – previously labeled as 19a). Transmission is by smear infection. Previously, insufficiently disinfected medical devices at ophthalmologists' offices were a major route of transmission. Nowadays, transmission at ophthalmologists' offices can still be observed because the crowding of patients in the waiting room facilitates smear infections from non-disinfected surfaces. Control of transmission is not an easy task because of the very high HAdV concentrations in eye secretions. The majority of other HAdV species D types are far less virulent, not associated with eye infections but rather with mild to asymptomatic gastrointestinal and respiratory infections.

Adenoviruses can cause severe, sepsis-like, disseminated infections in severely immunosuppressed patients of all age groups. These are most frequently observed in pediatric patients after allogeneic stem cell transplantation. As the majority of these infections are reactivations of latent adenovirus, which has a higher prevalence in pediatric patients, these cannot be completely avoided by hygiene measures.

Clinical Description of Infection

Incubation Period

The incubation period of HAdV Infections is usually short, within a range of one to two weeks. It may be even shorter in gastroenteritis and URTI. By contrast, many disseminated infections of immunosuppressed patients originate from latent adenovirus infections, which may have been established several years ago.

Due to the multitude of HAdV types with different tropism and virulence, a multitude of different diseases are associated with HAdV infections. Clinical manifestations of HAdV infections are also influenced by the age and immune status of the infected person.

Upper Respiratory Tract Infections

Many of these occur early in life and are self-limiting and mild. Usual symptoms are fever, nasal congestion, rhinitis, pharyngitis with sore throat, and cervical lymphadenopathy. Sometimes, otitis media may be present. An exudative tonsillitis may also be observed and can be clinically indistinguishable from a group A streptococcal infection. As C-reactive protein levels are elevated and infections may also occur without the typical seasonality of other respiratory virus infections, these may be falsely diagnosed as bacterial infections and treated unnecessarily with antibiotics. Most of these infections are caused by the few types of HAdV species C and a few types of species B (e.g., 3 and 7).

Lower Respiratory Tract Infections

URTI with types of species HAdV-C and -B may progress to LRTI with symptoms of cough, (pseudo-)croup, and bronchitis. Bronchiolitis and pneumonia can be severe and sometimes fatal, particularly in neonates and young children less than two years of age. A pertussis-like syndrome has also been reported, but the etiological significance of HAdV is questionable. In older children and adults, another small group of adenovirus types is clearly associated with severe LRTI (bronchiolitis, pneumonia, ARDS): type 4 of species HAdV-E and several types of species HAdV-B: 14, 21 (subtype 21a), 55, and 66 (previously labeled as genome type 7 h or intermediate strain 7H3) as well as, occasionally, types 3 and 7. Infections with these types account for the outbreaks of severe lower respiratory tract diseases in young military recruits reported most frequently in the US military and occasionally in other militaries. Stress and fatigue in military recruits as well as the cold season promote severe LRTI, including acute respiratory distress syndrome (ARDS).

The role of adenoviruses in asthma remains controversial, but long-term sequelae (bronchitis obliterans, bronchiectasis, and chronic atelectasis) of these LRTI are more frequently reported.

Ocular Infections

Corneal infection is typical in epidemic keratoconjunctivitis. This may begin with conjunctivitis with a "foreign body in the eye" sensation and progresses with corneal erosions and infiltrates that impair vision and painful edema of the eyelids. Acute inflammatory symptoms resolve in about two weeks. Unfortunately, blurred vision due to chronic infiltrates of the cornea (numuli), photophobia, and a foreign-body sensation may persist for months to years. Most cases of epidemic keratoconjunctivitis are caused only by a few types of species HAdV-D (8, 37, 53, 54, and 64). Several other HAdV types are associated with less severe ocular infections. Acute follicular conjunctivitis resolves without consequence in a few weeks but acute hemorrhagic conjunctivitis may also be observed. Pharyngoconjunctival fever is a typical manifestation of HAdV infection presenting with follicular conjunctivitis in combination with upper respiratory tract symptoms as pharyngitis, fever, and – occasionally – lymphadenopathy and malaise. All these eye infections usually do not affect the cornea and are associated with types 3, 4, and 7 of species HAdV-B and HAdV-E but may be caused by other types occasionally.

Gastroenteritis

The enteric HAdV of species F (types 40 and 41) and species A (types 12, 18, 31) are the third most common cause of viral gastroenteritis. In some cases, HAdV upper respiratory tract infections (see above) may also present with gastroenteritis symptoms. Species HAdV-F infections predominate in children below two years old. Symptoms include mild fever, vomiting, abdominal pain, and diarrhea – which is usually watery and non-bloody, lacks fecal leukocytosis, and lasts for about ten days. If dehydration is adequately treated, the vast majority of immunocompetent patients recover uneventfully. Mild enteritis is also observed in many cases of upper respiratory infections with non-enteric types of HAdV (see above). Other gastrointestinal syndromes infrequently associated with non-enteric adenovirus types include intussusception, acute mesenteric lymphadenitis, and appendicitis. However, more studies on these disease associations are required because HAdV DNA detection in these diseases may have been due to coincidental HAdV latency in gut-associated lymphoid tissue.

Genitourinary Diseases

Genitourinary diseases – including genital lesions, urethritis, cystitis, and hemorrhagic cystitis – are occasionally caused by HAdV. Cystitis and hemorrhagic cystitis of children and adults is clearly associated with a few types (11, 34, 35) of species HAdV-B. Although usually self-limiting, these infections may progress to graft nephritis in renal transplant recipients. Moreover, there are hints that these infections may also become latent for several years. Genital lesions and urethritis by HAdV can be sexually transmitted but are infrequent manifestations of HAdV infections with types usually associated with other diseases (e.g., type 2 and 37).

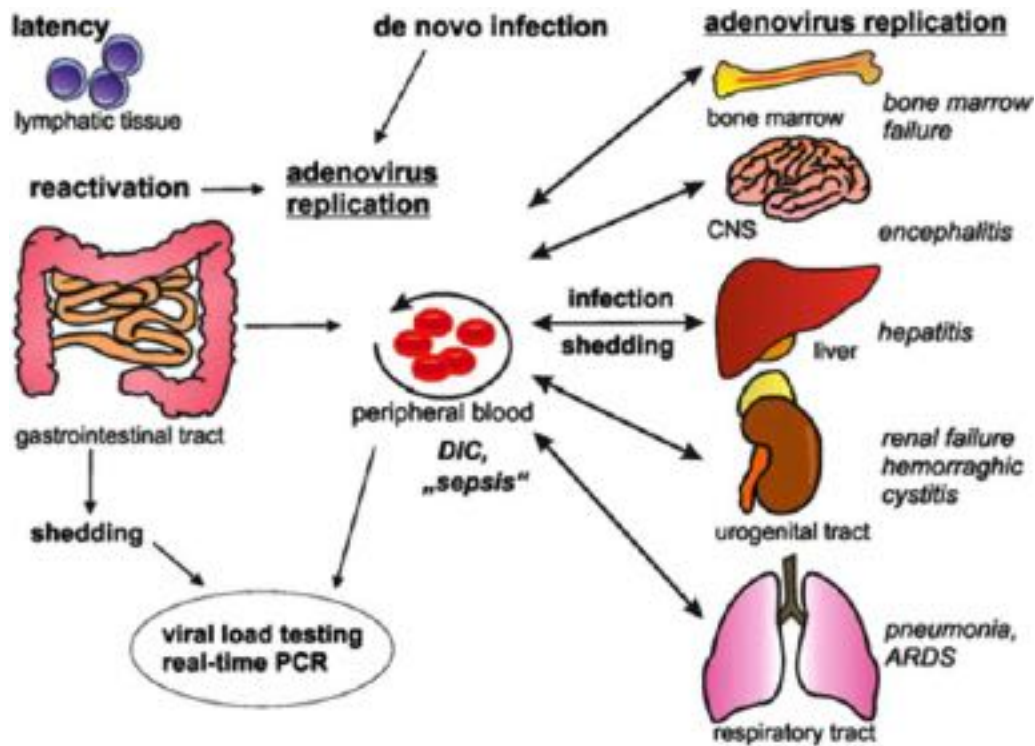


Fig. 2 Pathogenesis of disseminated adenovirus disease in immunocompromised patients. The source of human adenoviruses (HAdV) infection can be either de novo infection or reactivation from latent/persistent infection of the lymphatic tissue, for example, of the gut or the upper respiratory tract. After local replication, for example, in the gastrointestinal tract, HAdV spreads via the peripheral blood to other organs, such as the liver, bone marrow, central nervous system, urogenital, or respiratory tract. Massive HAdV replication in the affected organs leads to tissue destruction and causes the respective organ manifestation, such as enteritis, hepatitis, or pneumonia (here designated in *italics*). Virus progeny is shed to the blood or stool. Multiple organ involvement as well as disseminated intravascular coagulation might lead to a “septic” picture of disease and result in a very high lethality. From Ganzenmueller, T., Heim, A., 2011. Adenoviral load diagnostics by quantitative polymerase chain reaction: Techniques and application. *Reviews in Medical Virology* 22 (3), 194–208. doi:[10.1002/rmv.724](https://doi.org/10.1002/rmv.724).

Diseases Infrequently Associated With Human Adenoviruses Infections

Exanthems are occasionally associated with HAdV infections. Severe neurologic manifestations, such as meningoencephalitis and encephalitis, have been described occasionally and diagnosed by HAdV DNA detection in cerebrospinal fluid. HAdV DNA has also been detected in acute myocarditis of otherwise healthy people but more studies are needed on this topic. Adenovirus type 36 (species HAdV-D) infections have been linked epidemiologically to obesity – although this is controversial – as well as the association of HAdV infections with fetal demise and Kawasaki disease.

Adenovirus Infections of Immunocompromised Patients

All of the above-mentioned HAdV diseases may be prolonged and more severe in immunocompromised individuals including, for example, the progression of a respiratory infection from the upper to the lower respiratory tract or from cystitis to graft nephritis. Nevertheless, HAdV infections are only infrequent complications of solid organ transplant recipients and associated with periods of increased immunosuppression. HAdV infections of solid organ transplant recipients may be associated with the graft and are hard to discern from rejection. For example, in liver transplant patients, infections are most often associated with diarrhea, graft hepatitis, and sometimes pneumonia; in renal transplant recipients, acute hemorrhagic cystitis, graft nephritis, and pneumonia. Without proper adenovirus diagnostics, HAdV graft infections can be mixed up with rejection, and increased immunosuppression to treat supposed rejection may result in graft loss.

HAdV infections are most frequently observed in pediatric allogeneic hematopoietic stem cell recipients, about 20% may present with an HAdV infection and about 5% succumb before screening and treatment options develop. Disease most frequently starts with endogenous reactivation of a latent HAdV infection in the gastrointestinal tract (**Fig. 2**) resulting in gastroenteritis or asymptomatic HAdV shedding with the feces. Less frequently, de novo infections of the gastrointestinal tract or respiratory tract are observed. Anyway, these primary localized infections lead to viremia and subsequent organ infections (e.g., hepatitis, pneumonia, pancreatitis, encephalitis), which may be life-threatening. Lymphopenia (<300 μ l) and extensive use of immunosuppressive drugs to treat graft versus host disease are risk factors for the progression to disseminated disease, which is indicated by an increase of

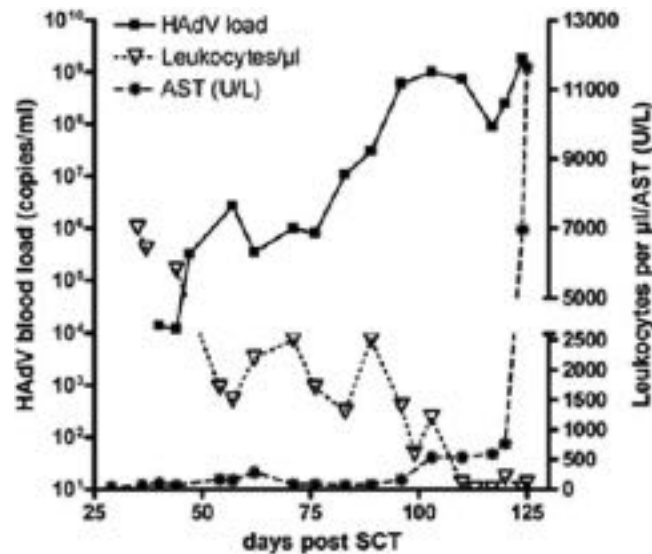


Fig. 3 Typical kinetics of the human adenoviruses (HAdV) blood load, leukocyte counts, and transaminase levels in a stem cell transplant (SCT) recipient with fatal disseminated HAdV disease. The graph depicts the HAdV DNAemia kinetics in peripheral blood of an SCT recipient, who died from multiple organ failure due to disseminated HAdV disease on day 126 after SCT. As a surrogate marker of severe hepatitis, the transaminase levels (aspartate transaminase) are indicated in U/L. Continuously declining leukocyte counts (leukocytes/ml), within parallel increasing HAdV DNAemia, highlight the role of the cellular immune response for the clearance of HAdV infections. From Ganzenmueller, T., Heim, A., 2011. Adenoviral load diagnostics by quantitative polymerase chain reaction: Techniques and application. *Reviews in Medical Virology* 22 (3), 194–208. doi:10.1002/rmv.724.

virus loads in peripheral blood. High-level viremia (10^6 – 10^{10} copies HAdV DNA/ml) is typical for this disease and associated with sepsis-like symptoms and failure of multiple organs (e.g., hepatitis and pneumonia) resulting in high mortality rates (Fig. 3). Both lysis of cells of visceral organs by HAdV replication and induction of a “cytokine storm” like sepsis by high loads of HAdV in the peripheral blood contribute to its pathophysiology. Disseminated disease of pediatric patients is most frequently caused by types 1 and 2 of species HAdV-C, less frequently by the other types of species C and type 31 of species HAdV-A.

In adult allogeneic hematopoietic stem cell recipients, HAdV infections are less prevalent, probably because adults may less frequently have HAdV latency, but disease manifestation is as severe as in the pediatric group. Types of species HAdV-C can be detected, as in pediatric patients, but several types usually associated with cystitis (11, 34, and 35 of species HAdV-B) can almost exclusively be found in this adult age group as an etiology of disseminated disease.

Since highly active antiretroviral therapy became available, HAdV complications in HIV infected patients were hardly ever observed. Previously, AIDS patients presented with chronic gastroenteritis, severe hemorrhagic cystitis, adenovirus-associated pneumonia, meningoencephalitis, and hepatitis as well as disseminated disease. In this group of patients, multiple types of species HAdV-D have been observed as etiological agents.

Animal Adenoviruses

Adenovirus infections can be found in almost all vertebrae (fish, reptile, bird, and mammal). Although typical zoonotic infections have not been described in humans, several adenovirus types infecting humans or simians share parts of their genomic sequences because of previous recombination events. These results of complete genomic sequencing indicate inter-(host-)species transmission events in the phylogeny of adenovirus types, infecting humans and adenovirus types infecting simians.

Diagnosis

Polymerase Chain Reaction (PCR) for Human Adenoviruses DNA Detection

In general, detection of HAdV DNA by nucleic acid amplification technologies (usually PCR) is the preferred diagnostic procedure for all types of adenovirus infections. Diagnostic specimens taken from the affected body site are appropriate (e.g., urine in cystitis cases, nasopharyngeal swabs in upper respiratory tract infections), although a few pitfalls have to be considered.

The huge genetic diversity of HAdV impedes the detection of all types in a single reaction, but several “generic” PCRs have been published and are also commercially available. However, some other commercially available diagnostic PCRs do not detect all types — sometimes only a very limited number of HAdV types. Many of these PCRs are parts of highly multiplexed PCR panels that detect multiple viral and bacterial pathogens associated with respiratory infections or gastroenteritis. Diagnostic claims and

intended use of these assays should be carefully checked. For example, a highly multiplexed PCR assay for gastroenteritis diagnostics may merely detect types 40 and 41 of species HAdV-F. Thus, this will underdiagnose gastroenteritis caused by other HAdV-types (e.g., gastroenteritis outbreaks in older children caused by types of species HAdV-A and sporadic gastroenteritis cases caused by types of species C, D, and G). Moreover, this assay will probably not detect adenovirus gastroenteritis in immunosuppressed patients, where species HAdV-C and -A types predominate. Of course, it is also not appropriate for screening purposes in this group of patients (see below).

Another pitfall of HAdV diagnostics by PCR (and to a lesser extent by virus isolation on cell cultures) is the detection of a latent HAdV infections coincident with another acute respiratory or gastrointestinal infection. About 5% of upper respiratory specimens or feces specimens of healthy or otherwise infected children can be HAdV DNA positive if a sensitive PCR protocol is used. Moreover, even EDTA blood specimens of healthy children with HAdV latency can be occasionally positive. Although analytically correct, these results are diagnostically misleading. Adenovirus shedding and DNAemia during latent infections is intermittent and always at low levels, whereas in all acute HAdV infections, an abundance of HAdV DNA is typical. Therefore, the simple solution to this diagnostic problem is either to limit the sensitivity of PCR protocols or to use real-time PCRs (using a cut off for high Ct values).

Quantitative Human Adenoviruses PCRs and Screening of High-Risk Patients

Diagnosis of disseminated disease previously required detection of HAdV (by virus isolation or histopathology) at multiple body sites, which was of course a tedious procedure. More recently, high virus loads in peripheral blood were found to be closely associated with disseminated infections. Therefore, testing EDTA blood (or plasma) samples by quantitative PCR became the standard diagnostic procedure for disseminated HAdV disease of immunosuppressed patients. Several generic real-time PCR protocols for HAdV DNA have been published and diagnostic assays are also available commercially. Multiple studies on virus loads in hematopoietic stem cell recipients demonstrated that disseminated disease and significantly higher mortality are associated with high peak virus loads in peripheral blood usually in a range between 10^6 and 10^{10} c/ml, whereas peak virus loads $<10^4$ c/ml were not. However, all these studies were not harmonized to a WHO standard, and different conclusions – if any – have been made on threshold virus loads for a therapy decision. Anyway, the occasional detection of low HAdV DNA concentrations in feces, respiratory materials, or blood should not be mistaken as the diagnosis of a severe HAdV infection but it rather indicates HAdV latency and a potential risk for severe HAdV infections.

In allogeneic hematopoietic stem cell recipients who are at a high risk for a disseminated HAdV infection (reactivation or de novo), weekly screening of blood virus loads is recommended by American and European guidelines to achieve an early diagnosis and facilitate preemptive therapy. Several studies in pediatric HSCT recipients have shown that additional weekly stool screening for HAdV DNA can detect reactivation even earlier and that increasing virus loads in stool precede HAdV DNA detection in peripheral blood by about one week. However, stool screening may fail in case of a de novo HAdV infection. Therefore, it should be performed only in addition to blood screening for HAdV.

Real-time PCR protocols can also be used for all other diagnostic applications that usually do not require quantification. Nevertheless, discerning highly positive specimens from weakly positive specimens can be helpful to diagnose acute infections in contrast to low-level HAdV shedding in latency.

Other Diagnostic Methods

Detection of HAdV antigens by ELISA and other (e.g., lateral flow) methods is feasible in stool samples and eye swabs with high HAdV loads, using commercial assays. However, their sensitivity and specificity are lower than PCR-based assays. Virus isolation on cell cultures was the previous gold standard of adenovirus diagnostics, but many laboratories stopped performing it because the time of detection is much slower than PCR. A549 cells are appropriate for most adenovirus types with the exception of types 40 and 41, which require Graham 293 cells for replication. Virus isolation is feasible from respiratory tract materials, eye swabs, urine, and stool but several days (up to three weeks) are required for the development of a CPE. An early CPE, observed during the first 24 h after inoculation of the cell culture, is not caused by HAdV replication but by toxic effects of HAdV penton base proteins, which can be present in high concentrations in clinical specimens.

Diagnosis of adenovirus infections by serological methods cannot be recommended due to the multiplicity of HAdV types. Type-specific neutralization tests can be regarded as reliable but are more appropriate for seroepidemiological and immunity studies. Adenovirus ELISAs use a group-specific epitope on the internal side of the capsid on the hexon protein but cut-offs are rather arbitrarily set; thus, any result is questionable.

Adenovirus Differentiation and Typing

For many diagnostic purposes, precise HAdV typing is not required but, basic differentiation on the HAdV species level can be quite helpful. For example, high virus loads of species HAdV-C in the feces of an HSCT recipient usually predict disseminated disease, whereas species HAdV-F does not. This differentiation can be easily performed with multiplex PCRs (conventional or real-time) for

the different HAdV species. Sequencing of amplicons of the usual generic diagnostic PCRs also provides rapid information on the HAdV species.

Typing is most frequently performed nowadays by sequencing the highly variable loops 1 and/or 2 of the main neutralization determinant ϵ on the hexon protein. This procedure is equivalent to the classical neutralization assays but can be more rapidly performed directly with diagnostic specimens (without prior virus isolation). However, results are equivocal because several of the new (geno-)types (type numbers > 53) share neutralization epitope sequences with one of the older (sero-)types. Precise typing requires additional sequencing of the highly variable loops on the fiber knob and penton base. These additional sequence data help identify all HAdV types unequivocally, including the recombinant (geno-)types. By contrast, even a combination of classical serological typing methods (neutralization testing and hemagglutination inhibition testing) may fail with a few of the new (geno-)types. Complete genomic sequencing of HAdV is also feasible to type cell culture isolates or even HAdV DNA directly from diagnostic samples when the viral load is high. By complete genomic sequencing, all types, subtypes, and strains can be identified precisely. Previously, restriction fragment length polymorphisms were used to differentiate genomic variants of HAdV types as genome-types (not to be confused with genotypes) but this approach became obsolete when next-generation sequencing became available.

Treatment and Prevention

Antivirals

As adenovirus infections of immunocompetent patients are usually mild and self-limiting, these do not really require antiviral treatment – just symptomatic and supportive therapy. Some of these infections may be treated unnecessarily with antibiotics if a diagnosis of HAdV infection has not been made because its symptoms (e.g., fever, elevated c-reactive protein, and tonsillitis) may lead to a suspicion of bacterial infection. A few severe diseases of immunocompetent patients (e.g., pneumonia and epidemic keratoconjunctivitis) would be sensible targets for antiviral treatment; however, this is not established. In a few cases of pneumonia, experimental therapy with cidofovir was reported but its efficacy remains unclear.

In immunocompromised patients, a reduction of immunosuppressive therapy, if feasible, is of outstanding importance because disseminated disease and lethal outcomes are clearly associated with high levels of immunosuppressive treatment or lymphopenia. In some cases, HAdV infection may mimic graft rejection (e.g., HAdV nephritis) or graft versus host disease (e.g., HAdV gastroenteritis). Without a diagnostic workup for HAdV, these cases had been treated with high dose immunosuppression resulting in graft loss, disseminated disease, and many fatalities. On the other hand, many immunosuppressed patients with a HAdV infection and viremia lack any specific symptoms until they reach high virus loads ($> 10^6$ c/ml). However, antiviral therapy tends to fail in patients with high virus loads. Therefore, a weekly blood (and stool) screening of high-risk patients is essential. As preemptive therapy, a reduction of immunosuppressive therapy and the application of antiviral agents (off label use) should be considered.

Several antivirals – including ganciclovir, ribavirin, cidofovir, and its prodrug brincidofovir – have in vitro activities against adenovirus. As many HAdV reactivations and infections were observed during ganciclovir prophylaxis for CMV infections and, in some cases, even during high dose therapy of CMV infections, the clinical use of ganciclovir for HAdV infections should not be considered.

Ribavirin may not have activity against all HAdV types but has gained an orphan drug status for HAdV infections. However, results of case reports and case series on ribavirin therapy were rather inconclusive and included therapy failures. Therefore, ribavirin is no longer frequently used for HAdV infections.

All HAdV types are susceptible to cidofovir in vitro. Although cidofovir resistance can develop with serial passages, little resistance has so far been detected in isolates from cidofovir-treated patients. Prospective controlled clinical trials have never been published, but several case reports and retrospective studies suggested some efficacy and a reduction of mortality if cidofovir is used as preemptive therapy. However, this is not a proof of the efficacy of cidofovir as other measures (reduction of immunosuppression) may have contributed to the more favorable outcomes. Moreover, HSCT patients can clear their HAdV infection spontaneously with hematopoietic regeneration and rising lymphocyte counts. In contrast to high HAdV loads, low peak HAdV loads in peripheral blood are not associated with mortality. Hence, the use of a threshold virus load (e.g., 10^4 copies/ml blood) to trigger preemptive cidofovir treatment has been suggested to avoid overtreatment and unnecessary side effects. Other factors supporting the treatment decision are T cell depletion, severe lymphopenia, and rapidly increasing virus loads.

A lipid conjugate of cidofovir – brincidofovir (CMX001), which can be administered orally – reaches higher intracellular concentrations and better antiviral activity with lower nephrotoxicity than cidofovir. However, diarrhea is the main side effect of brincidofovir. Although a randomized placebo-controlled clinical trial for the preemptive treatment of asymptomatic HAdV-DNAemia in HSCT patients did not show a significant effect on the primary endpoint, a positive trend was observed. In addition, many case studies support the view of a higher efficacy of brincidofovir.

Although preemptive antiviral therapy can limit HAdV infection for some time, clearance of HAdV DNAemia is clearly associated with increased lymphocyte counts and the detection of adenovirus specific T cells. Therefore, immunotherapeutic interventions – either using donor lymphocyte infusions or the adoptive transfer of HAdV-specific cytotoxic T lymphocytes expanded from naive T cells – were developed. The latter approach holds promise to avoid graft-versus-host disease associated

with donor lymphocyte infusion. As the *in vitro* expansion process requires several days, diagnosis of HAdV infection has to be made early, bridging with antiviral agents as cidofovir is usually required.

Vaccines

An oral vaccine against HAdV types 4 and 7 is available only for the US military. This vaccine is effective to prevent respiratory infections, which can be severe in military recruits. Surprisingly, it consists of replication-competent, non-attenuated viruses contained in enteric-coated tablets. After ingestion, the HAdV infection seems to be limited to the gastrointestinal tract, and protective immunity against subsequent respiratory infections with these types is induced.

Germicides

Infectious HAdV can survive for long times in liquids or dried on surfaces; moreover, as a non-enveloped virus, it is resistant to common detergents. Therefore, there is a high risk for smear infections. However, HAdV is sensitive to heating and many chemical germicides. These should be certified for their activity against HAdV, but due to the very high virus loads in feces or eye swabs, even these may fail. Therefore, a combination of thorough cleaning and subsequent use of germicides can be considered.

Further Reading

- Dehghan, S., Seto, J., Liu, E.B., *et al.*, 2019. A zoonotic adenoviral human pathogen emerged through genomic recombination among human and nonhuman simian hosts. *Journal of Virology* 93.
- Genzenmueller, T., Heim, A., 2011. Adenoviral load diagnostics by quantitative polymerase chain reaction: Techniques and application. *Reviews in Medical Virology* 22 (3), 194–208. doi:10.1002/rmv.724.
- Heim, A., Hayden, R.T., 2019. Adenoviruses. In: Carroll, K.C., Pfaller, M.A., Landry, M.L., *et al.* (Eds.), *Manual of Clinical Microbiology*. Washington, D.C.: ASM Press.
- Lion, T., 2019. Adenovirus persistence, reactivation, and clinical management. *FEBS Letters* 593, 3571–3582.
- Nemerow, G., Flint, J., 2019. Lessons learned from adenovirus (1970–2019). *FEBS Letters* 593, 3395–3418.
- Pied, N., Wodrich, H., 2019. Imaging the adenovirus infection cycle. *FEBS Letters* 593, 3419–3448.
- San Martin, C., 2012. Latest insights on adenovirus structure and assembly. *Viruses* 4, 847–877.
- Walsh, M.P., Chintakuntlawar, A., Robinson, C.M., *et al.*, 2009. Evidence of molecular evolution driven by recombination events influencing tropism in a novel human adenovirus that causes epidemic keratoconjunctivitis. *PLOS One* 4, e5635. doi:10.1371/journal.pone.0005635.g001.

Relevant Websites

<http://hadv.wg.gmu.edu/>
HAdV Working Group.

Management of Hepatitis A and E Virus Infection

Sébastien Lhomme, Florence Abravanel, Jean-Marie Peron, Nassim Kamar, and Jacques Izopet, Toulouse University Hospital, Toulouse, France and Toulouse University Paul Sabatier, Toulouse, France

© 2021 Elsevier Ltd. All rights reserved.

Introduction

The hepatitis A virus (HAV) and hepatitis E virus (HEV) are small RNA viruses mainly enterically transmitted. There are two forms of infectious particles; one corresponds to non-enveloped particles that are shed in the feces and the other is cloaked in host cell membranes and circulates in the blood. This membrane envelopment protects the HAV and HEV against neutralizing antibodies. Although both are ancient viruses, they are still a frequent cause of disease affecting millions of individuals annually.

Many individuals with HAV and HEV infections are asymptomatic or have mild symptoms. However, severe liver disease is also observed. In contrast to HAV and HEV genotypes 1 and 2, HEV genotypes 3 and 4 infections can lead to chronic hepatitis and cirrhosis in immunocompromised patients. For these patients, reduction of immunosuppression or ribavirin therapy can eradicate HEV. Extrahepatic manifestations such as neurological and renal diseases have been described in association with acute and chronic HEV infections.

Validated HAV immunoassays are largely available but HEV serological and nucleic acid tests have only recently made available. The improvement of virological tools for HEV diagnosis allowed a deep change in the understanding of the epidemiology and pathology of HEV infection.

Taxonomy

Hepatitis A Virus

HAV belongs to the family *Picornaviridae*, genus *Hepatovirus*. HAV has a primitive capsid related to that of picorna-like viruses (dicistroviruses) that infect insects. The nucleotide sequences of HAV strains vary little over time or place. There are six closely related genotypes (I to VI) but only one serotype. Genotypes I to III, further divided into subgenotypes A and B, are associated with infections in humans, while genotypes IV to VI are simian in origin. Evolutionary ancestral forms have been found in small mammals such as bats, rodents, hedgehogs, and shrews.

Hepatitis E Virus

HEV belongs to the *Hepeviridae* family, which includes two genera (*Orthohepevirus* and *Piscihepevirus*) and five species. The species *Orthohepevirus A* includes HEV that infects humans and several other mammals. *Orthohepevirus B* infects chickens, while *Orthohepevirus C* infects rats and ferrets, *Orthohepevirus D* infects bats, and *Piscihepevirus A* infects cutthroat trout. Two cases of infection with *Orthohepevirus C* HEV were reported recently despite its genetic differences from the other human pathogenic strains. The species *Orthohepevirus A* consists of at least eight distinct HEV genotypes but only one serotype. Genotypes 1 and 2 are strictly human and circulate in developing countries where sanitary conditions are poor. They are responsible for outbreaks. HEV genotype 3 is widely distributed around the world while HEV genotype 4 is found mainly in Asia. Genotypes 3 and 4 have animal reservoirs, mainly pigs, and infect humans in developed countries. HEV genotypes 3 and 4 have been detected in a wide variety of animals, including domestic pigs, wild boar, deer, rabbits, and mongoose. HEV genotypes 5 and 6 have been detected in wild boar in Japan but not yet in humans. HEV genotype 7 was recently detected in dromedary camels; it also caused chronic hepatitis in an immunosuppressed individual who regularly consumed camel milk and meat. Lastly, HEV genotype 8 has been found in Bactrian camels.

Epidemiology and Transmission

HAV and HEV are both transmitted mainly by the fecal-oral route. But despite their modes of transmission being quite similar, their epidemiology differs.

Epidemiology

Hepatitis A virus

HAV is distributed worldwide but its epidemiological pattern varies between populations, depending on sanitation and socioeconomic status. One way of assessing the endemic penetration of HAV is based on the age at which seroprevalence reaches 50% of the population: very high (<5 years), high (5–14 years), intermediate (15–34 years), and low (>35 years) (Fig. 1).

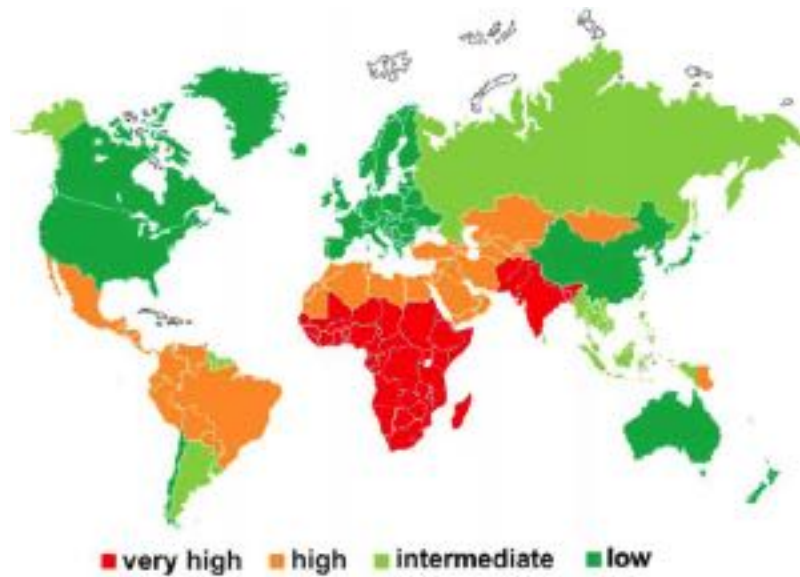


Fig. 1 Geographic distribution of HAV infections. HAV: colors represent different endemic patterns based on the age at which 50% of the population is HAV IgG positive [red: very highly endemic (< 5 years); orange: highly endemic (5–14 years); light green: intermediately endemic (15–34 years); dark green: low endemic (> 35 years)].

Very high and high endemicity pattern

The rates of HAV transmission in low-resource countries with poor sanitation, such as most area in Africa and parts of Asia, are very high and HAV is acquired in early childhood. Infections are asymptomatic and lead to universal protection against HAV in later life. Most people are exposed to HAV and become seropositive before they are 5 years old.

Low endemicity pattern

The circulation of HAV is limited in high-resource countries with good sanitation and hygienic conditions (North America, Europe, Australia, Japan, etc.), and most people in all age groups are susceptible. However, HAV seroprevalence is higher in older people due to a cohort effect. Introduction of HAV into the population of these countries can trigger waves of transmission with symptomatic disease (jaundice), in men who have sex with men (MSM) for example, or through importation of contaminated food.

Intermediate endemicity pattern

HAV infections are fairly common in intermediate-resource countries like North Africa, South America and Western Asia, where symptomatic disease occurs in children over 5 years old and young adults. Socio-economic development and improved hygiene have led to a change in HAV epidemiology in these countries, with more cases of disease, more hospitalizations, and more deaths, despite reduced HAV transmission. This public health problem has led to increased prevention, including universal childhood HAV vaccination.

Hepatitis E virus

HEV genotypes 1 and 2

HEV genotype 1 is highly endemic and prevalent in South, Central, and South-East Asia, as well as in Africa and the Middle East (Fig. 2). HEV genotype 2 is prevalent in western Africa and Mexico. Although HEV genotype 1 was responsible for frequent outbreaks in China over the past few decades, it is now less prevalent. HEV genotype 4 infections have become more prevalent, perhaps due to improved sanitation and increased income. It is estimated that 20.1 million new infections occur annually in Asia and Africa. Determining the real seroprevalence in these areas is difficult as a variety of serology assays with quite different sensitivities have been used. However, published data indicate that 20% to more than 50% of people living in most parts of Asia and Africa have HEV antibodies. Most of the asymptomatic cases are adolescents and young adults. Infection in children is more often asymptomatic than in adults.

HEV genotypes 3 and 4

HEV genotype 3 is the main HEV genotype circulating in Europe, North America, and South America, while genotype 4 is predominant in Asia (Fig. 2). Cases of genotype 4 HEV infection in Europe are sporadic and locally acquired while genotype 1 infections are travel-associated. The total number of reported symptomatic cases in Europe increased ten-fold between 2005 (514 cases) and 2015 (5617 cases), but the percentage of cases hospitalized decreased from 80% in 2005 to around 50% in 2015. A meta-analysis showed that the prevalence of IgG in the general population in Europe is influenced by the assay used for the

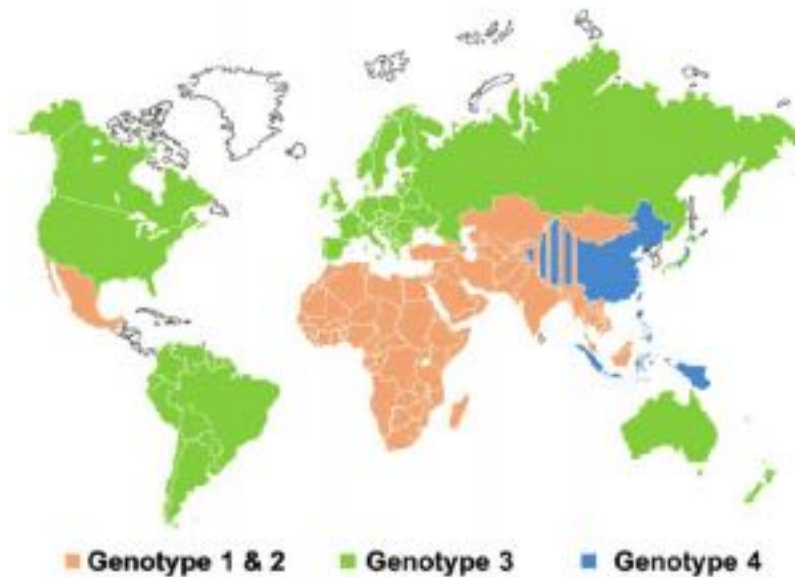


Fig. 2 Geographic distribution of HEV infections. HEV: colors represent the predominant HEV genotypes in each area (pink: HEV genotypes 1 and 2; green: HEV genotype 3; blue: HEV genotype 4).

detection. Thus, 52% of the heterogeneity was explained by 3 parameters: the assay used, the geographical location and the cohort studied. Hence, the seroprevalence of IgG can be <10% in countries like Italy, Ireland or Scotland and >20% in others like France, the Netherlands or Germany. The estimated seroprevalence of IgG in France was 22.4%, based on more than 10 000 French blood donors. There were geographical differences: rates were higher in the South and North East. But seroprevalence can vary from over 80% to below 20% even in high-rate areas like South-West France. Historical data indicated a seroprevalence of 21% in the United States, while more recent data suggest that the prevalence of HEV IgG is 6%. However, few data are available for the United States, probably due to the lack of FDA-licensed anti-HEV serology assays.

Transmission

Hepatitis A virus

Although a highly effective, safe hepatitis A vaccine was developed in the early 1990s, HAV is still an important cause of acute viral hepatitis worldwide. The virus is readily transmitted and outbreaks are frequent, especially in developed countries where herd immunity in the population is low. There is no animal reservoir for human HAV strains.

Water- and food-borne transmission

Most cases of HAV are due to ingesting contaminated water or food, but swimming in water contaminated by adjacent septic systems or sewage is also a risk. A variety of foods have been implicated in hepatitis A outbreaks. Food can be contaminated at any point during processing, harvesting, preparation, or distribution. Seafood, vegetables, and fruit are classic vectors. The largest reported hepatitis A outbreak was associated with consumption of raw clams in Shanghai in 1988: around 300,000 individuals were infected. Hepatitis A outbreaks due to contaminated food remain a public health problem in developed countries with low or intermediate endemicity.

Person-to-person transmission

Close contact between infected and susceptible people is the most common mode of HAV transmission in developing and developed countries. Transmission is facilitated by the prolonged shedding of HAV in the feces before and after the onset of symptoms, and asymptomatic infections, especially in young children. This type of HAV transmission occurs frequently in close institutions like schools, nurseries or day care centers.

Other modes of transmission

Outbreaks of HAV infection frequently occur among MSM. HAV is not transmitted by the semen but by oral contact with fecally contaminated sites. The initial introduction of HAV into an MSM community is followed by its rapid spread and subsequent persistence.

Outbreaks of HAV infection have been also reported among people who inject drugs.

Lastly, HAV can, very rarely, be transmitted by the transfusion of blood or blood products. In some cases, the recipients are HAV IgG positive. Blood donors are not screened for HAV because acute HAV infections are sporadic among blood donors and there are no chronic carriers. However, the plasma industry tests all sources of plasma by testing mini-pools for HAV nucleic acid.

Hepatitis E virus

HEV genotypes 1/2 and genotypes 3/4 are transmitted differently. HEV genotype 1 and 2 infections occur mainly in developing countries, where up to 25% of infected pregnant women can die of severe hepatitis. It can be responsible for outbreaks in areas where it is highly endemic. Drinking contaminated water is the main route of transmission and outbreaks are often linked to fecal contamination of drinking water. Person-to-person transmission of HEV is uncommon. The transmission from infected women to their newborns is well documented, and transmission of HEV genotype 1 through blood transfusions has been reported.

HEV genotypes 3 and 4 are found in high-income countries. Since these genotypes have animal reservoirs, the main route of their transmission is from eating HEV-infected animal products, especially undercooked meat. Eating infected pork liver, pork products containing liver, and other raw or undercooked pork meat are the main sources of HEV infections. Infected game meat is another source of human infection. Strains of rabbit HEV have been detected in humans. HEV shed by infected animals can contaminate water sources, leading to the accumulation of HEV in fruit, vegetables, and shellfish. Hence, HEV RNA (HEV genotype 3) has been detected on red fruit, strawberries, salads and in spices, as well as in oysters and mussels. Recent studies in many European countries, China and Japan have also shown that HEV can be transmitted by transfusion of blood products. Most cases of transfusion-transmitted HEV infection are asymptomatic, and only a small proportion of recipients of infected blood products develop symptomatic hepatitis. The frequency of blood donors testing positive for HEV RNA across many countries ranges from 1:600 to 1:14,799. Blood donors are now systematically screened in several European countries.

Clinical Presentation and Course

Hepatitis A virus

The clinical manifestations of HAV infection usually range from asymptomatic infection to fulminant hepatitis, depending on the age of the patient. Thus, 70% of infected children under 6 years old are asymptomatic while over 70% of infections in adults are symptomatic. Symptoms, including fever, malaise, nausea, vomiting, abdominal pain, dark urine and jaundice, develop after an incubation period of approximately 4 weeks (range 2–7 weeks). Other symptoms like myalgia, pruritus, diarrhea, arthralgia or skin rash are less common. While there is no evidence of chronic liver disease or persistent infection after acute hepatitis, some patients (4%–20%) suffer from prolonged disease or relapsing disease lasting up to 6 months, with prolonged HAV excretion in the feces.

Atypical manifestations reported following HAV infection are: relapsing hepatitis, prolonged cholestasis and complicated cases with acute kidney injury or autoimmune hepatitis. Relapsing hepatitis A is defined by a biphasic peak of serum liver enzyme (aminotransferase) with 4–7-week intervals between the first and second peaks. Prolonged cholestatic hepatitis A is associated with pruritus, fatigue, loose stools and weight loss. Prolonged cholestatic cases can be predicted by the detection of HAV RNA in the plasma after 20 days of illness, while relapse is not predictably linked to the presence of plasma HAV RNA. Acute kidney injury occurs in 1.5%–4.7% of patients with non-fulminant hepatitis A. Renal damage may be due to pre-renal factors associated with anorexia, nausea, vomiting, diarrhea and fever as well as the nephrotoxic effect of hyperbilirubinemia, immune complex-mediated nephritis, interstitial nephritis and rare massive intravascular hemolysis. Other rare extra-hepatic manifestations have been reported: autoimmune hemolytic anemia, aplastic anemia, pure red cell aplasia, pleural or pericardial effusion, acute pancreatitis and neurological complications (mononeuritis, mononeuritis multiplex and Guillain-Barré syndrome).

Fulminant hepatitis is a rare complication of HAV infection. The incidence varies between 0.015% and 0.5% of acute hepatitis A cases. Older adults with underlying chronic liver disease are at greatest risk of acute liver failure.

HAV infection during pregnancy is associated with a great risk of maternal complications and preterm labor, despite the relatively mild features of the disease. Fetal outcome is usually benign despite prematurity.

Hepatitis E Virus

Most acute HEV infections are asymptomatic or minor systemic illnesses, whatever the genotype. The typical early phase symptoms of acute hepatitis E that develop after an incubation period of 2–6 weeks are unspecific and include: flu-like myalgia, arthralgia, weakness, and vomiting. Symptoms of acute hepatitis include jaundice, fatigue, itching, nausea, pale stools, and dark urine. The clinical symptoms and signs are indistinguishable from those of other forms of viral hepatitis.

Acute hepatitis E in immunocompetent patients usually resolves spontaneously. Although reinfections are possible, they are less likely to develop symptomatic hepatitis than non-immune individuals. A few HEV-infected individuals (0.5%–4%) develop acute liver failure. Patients with pre-existing chronic liver diseases are at increased risk of liver failure and mortality can be as high as 67%. There have been cases that have needed liver transplantation following HEV infection.

Pregnant women are at high risk of developing symptomatic disease if they become infected with HEV genotype 1 during the second and third trimesters. Many progress to acute liver failure, haemorrhagia and eclampsia, and the mortality rate can reach 15%–25%. Miscarriages, preterm deliveries, stillbirths and perinatal mortality are all more frequent in HEV genotype 1-infected pregnant women. Their new-born babies are also at risk of HEV infection through maternal–fetal transmission and of developing complications, such as anicteric or icteric hepatitis, hypoglycemic and neonatal death. Conversely, pregnant women infected with

HEV genotype 3 or HEV genotype 4 do not suffer from such maternal mortality, perhaps because HEV genotype 1 replicates more rapidly in decidual and placental tissues than do other genotypes.

Immunocompromised patients, such as those on immunosuppressive therapy following solid organ transplantation, patients with hematological disease on chemotherapy, patients with rheumatic disorders on heavy immunosuppression immunotherapy, and HIV-infected patients with low CD4 T cell counts, can all suffer from chronic HEV infections. Most immunosuppressed patients are asymptomatic. Chronic HEV infection is defined as HEV replication that persists for more than 3 months. Patients infected with HEV genotype 3, HEV genotype 4 or HEV genotype 7 can suffer from chronic HEV infections. A single case of chronic HEV genotype 1 infection has been reported; this needs confirmation.

One third of solid organ transplant (SOT) recipients or patients with hematological disease spontaneously clear the virus, while two-thirds develop a chronic infection. The use of tacrolimus (a more potent immunosuppressant than cyclosporine A), a low lymphocyte subset count and less-active anti-HEV-specific T cell responses are more frequently associated with the development of a chronic infection in SOT patients. About 10% of chronically infected patients develop cirrhosis a few years after their infection. Lastly, liver fibrosis can regress after HEV clearance. While HEV reactivation after clearance is very rare, reinfections can occur and can lead to chronic hepatitis.

The awareness of extrahepatic effects of HEV infections has increased. Most involve the neurological and renal systems. HEV infections are associated with a range of neurological injuries and cases of neurological injury in HEV genotype 1-infected Asians and HEV genotype 3-infected Europeans have been reported. These neurological injuries seem to be more frequent in immunocompetent patients (22.6%) than in immunocompromised ones (3.2%). The neurological disorders associated with an HEV infection include neuropathic pain, painless sensory disorders, neuralgic amyotrophy (Parsonage-Turner syndrome), Guillain-Barre syndrome, encephalitis/myelitis, mononeuritis multiplex, Bell's palsy, vestibular neuritis, myositis and peripheral neuropathy. All the HEV-infected patients with neurological manifestations generally have normal or modestly abnormal liver function. Thus, the neurological symptoms and signs dominate the clinical picture in these patients.

Both acute and chronic HEV infections can lead to impaired renal function. Renal biopsies from patients infected with HEV genotype 1 and HEV genotype 3 show signs of glomerular disease, including membranoproliferative glomerulonephritis with or without cryoglobulinemia and membranous glomerulonephritis. Flare-ups of pre-existing IgA nephropathy may also occur in patients infected with HEV. HEV RNA was detected in the cryoprecipitate from the serum of an immunocompetent patient with an acute HEV infection who presented with cryoglobulinemic membranoproliferative glomerulonephritis. HEV infection has been identified as an independent predictive factor for cryoglobulinemia in SOT recipients. Clearance of HEV is associated with improved renal function and a decline in proteinuria in most patients, which suggests that HEV is probably the cause of the renal injury. Hence, patients with hepatitis who have impaired renal function or proteinuria should perhaps be screened for HEV.

Lastly, acutely infected patients may display hematological disorders like thrombocytopenia, autoimmune hemolytic anemia, aplastic anemia, and acute liver failure associated with pure red-cell aplasia. Up to 25% of HEV genotype 3-infected patients can suffer from asymptomatic monoclonal gammaglobulin but the clinical significance of this observation is uncertain. Acute episodes of pancreatitis have been reported in patients infected with HEV genotype 1 from South-East Asia but not those with HEV genotype 3 or HEV genotype 4 infections.

Diagnosis

Clinical manifestations and biochemical tests alone cannot distinguish between an HAV and an HEV infection. Serum alanine aminotransferase activity is elevated during the prodromal phase and during the initial part of the icteric phase (Fig. 3). An HAV/HEV infection can be diagnosed either directly by detecting the genome or antigen in the blood or other body fluids or indirectly by detecting antibodies in the serum. The IgM response is detected around the time that the serum alanine aminotransferase activity increases. Thus, the presence of HAV/HEV IgM in the serum is a key marker of an acute infection. The IgG response appears shortly after the IgM response and can persist for several years.

Antigen Detection

Hepatitis A virus

HAV antigen can be detected in both blood and fecal samples by immunoassays, but this is less sensitive than nucleic acid assays. No commercial assay is presently available.

Hepatitis E virus

Assaying for HEV capsid antigen can be used to diagnose an HEV infection. The sensitivity of the available commercial assay (Wantai Biological Pharmacy, China) is 91% (88% in immunocompetent and 94% in immunocompromised patients) and the specificity is 100%. This assay could be used for the direct diagnosis of an HEV infection by laboratories with no molecular diagnosis facilities.

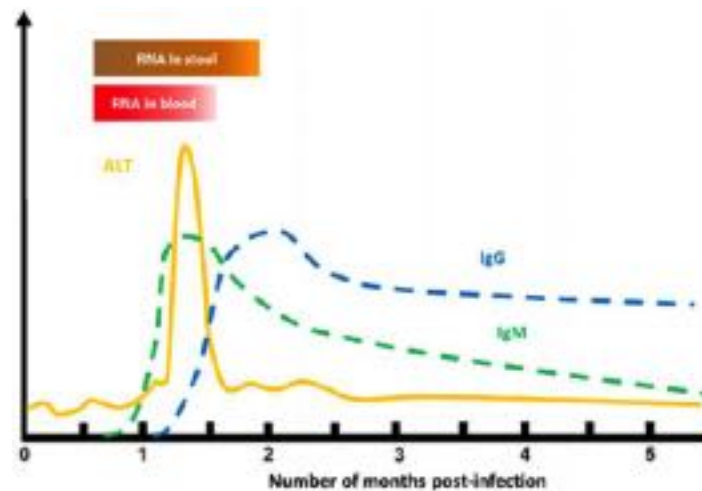


Fig. 3 Virus detection (HAV RNA or HEV RNA) at different sites and serologic response (IgM and IgG antibodies). ALT, alanine aminotransferase.

Table 1 Commercial molecular assays for the detecting/quantifying HAV or HEV RNA

Virus	Manufacturer	Product name	Type of test	Analytical sensitivity
HAV	Altona Diagnostics	Real Star HAV RT-PCR	<i>In vitro</i> diagnostics real-time PCR qualitative test	46 IU/ml
	Hologic/Grifols	Procleix Parvo/HAV	Transfusion, TMA allowing simultaneous detection of HAV and Parvovirus B19	1 IU/ml
	Roche	COBAS Taq Screen DPX	Transfusion, Real-time PCR allowing simultaneous detection of HAV and Parvovirus B19	1 IU/ml
HEV	Altona Diagnostics	Real Star HEV RT-PCR	<i>In vitro</i> diagnostics quantitative test real-time PCR	20 IU/ml
	Eurobio	Eurobioplex HEV RNA	<i>In vitro</i> diagnostics Quantitative test Real Time PCR	40 IU/ml
	Fast-track Diagnostics	FTD Hepatitis E RNA	<i>In vitro</i> diagnostics quantitative test real-time PCR	188 IU/ml
	Hologic/Grifols	Procleix HEV	Transfusion, TMA qualitative test	7 IU/ml
	Mikrogen Diagnostik	AmpliCube HEV	<i>In vitro</i> diagnostics quantitative test real Time PCR	20 IU/ml
	Roche	COBAS HEV	Transfusion, real-time PCR qualitative test	7 IU/ml

Nucleic Acid Detection

Hepatitis A virus

HAV RNA can be detected in the blood and feces by RT-PCR or TMA before the increase in alanine aminotransferase activity and the appearance of clinical symptoms. Several commercial tests are available, with limits of detections ranging from 1 to 46 IU/ml ([Table 1](#)).

Hepatitis E virus

HEV RNA can be detected in the blood, other body fluids and feces using nucleic acid amplification. Most RT-PCR assays, including commercial assays, target ORF3. A TMA assay performed on a fully automated platform is well adapted for high throughput testing. The limit of detection of current assays is 7–188 IU/ml ([Table 1](#)). Reverse transcription droplet digital PCR gives absolute quantities of HEV RNA without a standard curve. Lastly, the loop-mediated isothermal amplification assay provides a one-step single tube amplification of HEV RNA.

Serological Tests

Hepatitis A virus

Enzyme or chemiluminescent immunoassays for HAV IgM and IgG are commercially available in either microplate or multi-parametric automated formats ([Table 2](#)). The diagnostic performance of these immunoassays in terms of sensitivity and specificity is well established (> 95%). No commercial rapid immunochromatographic assay is presently available.

Table 2 Commercial immunoassays for detecting anti-HAV and anti-HEV antibodies

<i>Virus</i>	<i>Marker</i>	<i>Manufacturer</i>	<i>Product name</i>	<i>Type of test</i>
HAV	HAV IgM	Abbott	ARCHITECT HAVAb-IgM	Automate
			ALINITY HAV Ab IgM	
		Roche	COBAS Anti-HAV IgM	Automate
		Biomerieux	VIDAS HAV IgM	Automate
		Siemens	ADVIA Centaur Anti-HAV IgM	Automate
	HAV IgG	Beckman-Coulter	ACCESS HAV IgM	Automate
		Diasorin	LIAISON HAV IgM	Automate
		Ortho	VITROS Anti-HAV IgM	Automate
		Abbott	ARCHITECT HAVAb-IgG	Automate
			ALINITY HAV Ab IgG	
	HAV total antibodies	Roche	COBAS Anti-HAV	Automate
		Biomerieux	VIDAS Anti HAV Total	Automate
		Siemens	ADVIA Centaur Anti-HAV Total	Automate
		Beckman-Coulter	ACCESS HAV Ab	Automate
		Diasorin	LIAISON Anti-HAV	Automate
		Ortho	VITROS Anti-HAV Total	Automate
HEV	HEV IgM	Wantai	HEV IgM Rapid test	Rapid test
		MP Diagnostics	Assure HEV IgM rapid test	Rapid test
		Wantai	HEV Elisa IgM	Microplate
		MP Diagnostics	HEV IgM ELISA	Microplate
		Adaltis	EIAgen HEV M	Microplate
		Mikrogen	recomWELL HEV IgM	Microplate
		Euroimmun	ELISA Anti-Virus HEV IgM	Microplate
		Biomerieux	VIDAS anti-HEV IgM	Automate
		Orgentec	Alegria Anti-Hepatitis E Virus IgM	Automate
		Vircell	Virclia Hepatitis E IgM Monotest	Automate
	HEV IgG	Wantai	HEV Elisa IgG	Microplate
		MP Diagnostics	HEV IgG ELISA	Microplate
		Adaltis	EIAgen HEV G	Microplate
		Mikrogen	recomWELL HEV IgG	Microplate
		Euroimmun	ELISA Anti-Virus HEV IgG	Microplate
		Biomerieux	VIDAS anti-HEV IgG	Automate
		Orgentec	Alegria Anti-Hepatitis E Virus IgG	Automate
		Vircell	Virclia Hepatitis E IgG Monotest	Automate

HAV IgM, a key marker of acute infection, is detectable from the time the alanine aminotransferase activity increases and persists for 3–6 months thereafter. False positive results are possible, especially when the clinical picture does not reflect the HAV infection and there is no HAV IgG.

The presence of HAV IgG alone is a marker of past infection or vaccination. Testing for HAV IgG can be used to identify the unvaccinated individuals who require immunization, but the cost-effectiveness of screening depends on both age and epidemiologic parameters. Post-vaccination tests are not required due to the high efficacy of HAV vaccines.

Hepatitis E virus

Enzyme immunoassays for HEV IgM and IgG are commercially available in Europe and Asia in formats for microplates and multiparametric automated instruments (Table 2). Several immunochromatographic assays are also available in different countries. The antigens used are usually recombinant ORF2 capsid protein and/or ORF3 protein from HEV genotype 1 strain. The diagnostic performance of these immunoassays varies considerably and should be carefully evaluated.

The presence of HEV IgM in the serum is a key marker of an acute infection. Using a validated PCR assay as a reference, studies have shown that the sensitivity of IgM immunoassays is >97% for immunocompetent patients and 80%–85% for immunocompromised patients; their specificity is >99.5%. Longitudinal follow-up of patients with acute hepatitis E proven by nucleic acid testing showed HEV IgM persisted for 6–12 months.

The presence of HEV IgG alone is a marker of a past infection. It is difficult to compare seroprevalence rates for different populations obtained by different laboratory methods because the analytical sensitivity of IgG immunoassays varies enormously. The limits of detection of commercial HEV IgG assays vary from 0.25 WHO unit/ml to 2.5 WHO unit/ml when measured with an international standard (WHO reference reagent established in 2002; National Institute for Biological Standards and Control, Code 95/584). As highlighted in a meta-analysis of studies of HEV seroprevalence in Europe, the most sensitive immunoassay produced

the highest estimates of HEV IgG seroprevalence. Determining the HEV IgG concentration could be useful for estimating the risk of reinfection after a natural infection or vaccination.

Vaccines and Antiviral Agents

Hepatitis A Virus

Prevention against an HAV infection can be afforded by adequate sanitation and housing facilities, together with good personal hygiene. Passive or active immunization, before or after exposure, also provides effective protection against HAV infection.

Immunoglobulin

Passive immunization with human immunoglobulins can be used for pre-exposure prophylaxis. An intra-muscular injection of immunoglobulins (0.02–0.06 ml/kg) is effective within hours of injection and for periods of 12–20 weeks.

Human immunoglobulins can also be used for post-exposure prophylaxis, especially in immunosuppressed people, individuals at risk of severe disease, and infants less than 1 year old. As the average post-exposure incubation period of hepatitis A is about 28 days (but could be as short as 15 days), the time for intervention is only about 2 weeks. While immunoglobulins may protect against the symptoms of hepatitis A, they may not prevent infection following exposure. In recent years, the decline in the concentrations of HAV antibodies in the plasma pools used to prepare immunoglobulins has raised the question of the efficacy of these immunoglobulins in post-exposure prophylaxis. In addition, the short duration of protection and the increasing cost of immunoglobulins has led to a search for alternatives.

Vaccine

Inactivated hepatitis A vaccines contain virus particles produced in cell culture, purified, inactivated with formalin, and adsorbed onto aluminum hydroxide. Vaccines that combine hepatitis A and typhoid protection and hepatitis A and hepatitis B protection have been licensed, as have hepatitis A monovalent vaccines. The immunogenicity of the combination vaccines is equivalent to that of the monovalent hepatitis A vaccine. Live-attenuated HAV vaccines are available in China. Traditionally, two doses of inactivated vaccine at an interval of 6–12 months are recommended, beginning after 12 months of age. The protective efficacy is nearly 95% and protection lasts for at least 10 years. Modeling studies predict that 88% of individuals who were seronegative prior to vaccination will remain protected for at least 30 years. Consequently, the Centers for Disease Control and Prevention now recommend the pre-exposure prophylaxis of 2–40 year-old exposed patients with vaccine instead of immunoglobulins. The threshold for protection from HAV infection in humans is an antibody concentration of 10–33 IU/ml, depending on the assay used. However, clinical experience suggests that vaccination may provide protection even when no HAV antibodies can be detected using standard immunoassays. A positive test for total HAV antibodies signifies immunity to hepatitis A.

Optimal strategies for determining the need for HAV vaccination are based on seroprevalence. Vaccination is not routinely recommended for people living in areas where HAV is very prevalent. Almost all such people are asymptotically infected with HAV in childhood, which effectively prevents clinical hepatitis A in adolescents and adults. Vaccination of all children may protect the health of adolescents and young adults living in countries where HAV prevalence is intermediate. The vaccination of all babies in Argentina with a single dose of inactivated HAV vaccine at 12 months led to marked reduction in the incidence of symptomatic hepatitis A, fulminant hepatitis, and liver transplantation. Targeted vaccination of high-risk groups (travelers to regions of high HAV endemicity, MSM, injecting drug users, immunosuppressed patients, patients infected with HIV, and patients with chronic liver disease) is generally recommended for people living in countries where HAV prevalence is low, but extension to the general population is sometimes an option. The United States began targeted hepatitis A vaccination for high-risk groups and children living in high-incidence communities in 1999. The Centers for Disease Control and Prevention recommended, in 2006, that every child between 12 and 23 months old should be vaccinated.

Clinical trials suggest that an inactivated HAV vaccine effectively prevents infection when given within 14 days post-exposure. This has been confirmed by a randomized controlled trial that included 1080 susceptible children and adults. They were given either immunoglobulins or one dose of inactivated HAV vaccine. Symptomatic infection occurred in 3.3% of Ig recipients and in 4.4% of those vaccinated.

Treatment

There is no specific treatment for hepatitis A. Recovery from symptoms may be slow, taking several weeks or months. Supportive care includes adequate hydration and nutritional support so as to maintain patient comfort and an adequate nutritional balance, including replacing fluids lost by vomiting and diarrhea. Acetaminophen should not be used to control fever because of its liver toxicity. Anti-vomiting medication should not be given.

Considering that acute renal failure or hemolytic anemia may complicate hepatitis A, renal function should be regularly tested, along with complete blood analysis. Individuals with hepatitis A and acute liver failure can be treated with N-acetylcysteine, while liver transplantation is considered for the sickest patients. However, HAV-related fulminant hepatitis resolves spontaneously more frequently than does fulminant hepatitis of other etiologies. This makes the decision to transplant or not rather difficult.

Hepatitis E Virus

Prophylaxis

Adequate sanitation and housing, together with personal hygiene, provide the best protection against HEV infection in developing countries. To these can be added avoiding eating uncooked or undercooked meat. "At risk" food should be cooked thoroughly, with an internal temperature of 71°C for 20 min to completely inactivate HEV.

The high prevalence of viremic blood donations in developed countries and the great proportion of immunocompromised blood recipients has led practitioners to define measures to improve blood safety. The inactivation of pathogens in blood components using amotosalen is not effective for HEV, while treatment with riboflavin (vitamin B2) plus ultraviolet light has a limited effect. Consequently, several developed countries have adopted measures to improve blood safety based on the epidemiology of HEV. All blood donations are screened in Ireland (since 2016), the United Kingdom, the Netherlands (both since 2017) and Switzerland (since 2018). Plasma donations intended for transfusion to patients at risk of severe disease have been selectively screened in France since 2013.

Vaccine

Two vaccines against HEV vaccines have been developed; one is a 56-kDa subunit antigen (amino acids 112–607 of ORF2 expressed by cultured insect cells) and the other is a subunit antigen, HEV 239 (amino acids 368–606 of ORF2 expressed by *Escherichia coli*). Only HEV 239 has been licensed for use in humans, and it is only available in China. The Hecolin® vaccine was developed to elicit protective antibodies against all HEV genotypes and provides cross-protection for both genotypes 1 and 4. The efficacy of this three-dose vaccine is 97%, as judged by its ability to prevent episodes of symptomatic acute hepatitis and its long-term efficacy during follow-up. Modeling studies suggest that it can provide protection for up to 30 years. The vaccine seems to be safe for pregnant women. However, its safety and efficacy has not been tested on patients with chronic hepatitis or on immunosuppressed patients.

The concentration of HEV IgG needed to prevent infection after a natural infection or vaccination in clinical trials is still unknown, although one study reported that an antibody concentration of 2.5 WHO units/ml is protective. However, both immunocompromised and immunocompetent patients with antibody concentrations below 10 WHO units/ml have suffered reinfection.

Management of acute hepatitis E

Most immunocompetent individuals who have acute hepatitis E do not require antiviral therapy with ribavirin because HEV is spontaneously cleared. Whether early treatment with ribavirin helps clear the virus or decreases the risk of liver failure remains to be determined. A few acutely HEV-infected patients have been treated with ribavirin. Their liver enzyme activities were rapidly normalized and they all cleared HEV RNA quickly. However, there was no control group and the ribavirin doses and treatment durations varied greatly. Thus, spontaneous improvement cannot be ruled out. Corticosteroids might decrease the risk of progression to liver failure in patients with fulminant hepatitis E. Some case reports have found that corticosteroid therapy improved liver function. Lastly, some case of HEV infection required liver transplantation.

Management of chronic infections

The first-line treatment of organ transplant patients with chronic hepatitis is to reduce their immunosuppressive drugs, especially those targeting T cells (such as calcineurin inhibitors), when possible (Fig. 4). This allows one-third of patients to clear the HEV. The HEV infection in the remaining two-thirds of patients can be efficiently treated by ribavirin monotherapy. Most (78%) patients achieved a sustained virological response after 3-months of ribavirin therapy. Longer treatment (6 months) can allow HEV clearance in relapsers. Persistent HEV shedding in the feces at the end of therapy is associated with HEV relapse. The main side effects of ribavirin treatment are dose-dependent anemia, dry cough and skin reactions. The ribavirin dose should be adapted to maintain hemoglobin and eGFR concentrations because chronically infected patients can suffer from anemia or impaired renal function.

It has been suggested that ribavirin inhibits HEV replication by inhibiting inosine monophosphate dehydrogenase (IMPDH), leading to the depletion of guanosine triphosphate pools. Deep sequencing analysis has identified several HEV RNA mutations, such as G1634R. Recent studies have shown that ribavirin increases HEV heterogeneity that seems to be reversible. However, the role of HEV RNA variants and their impact on HEV treatment outcome are uncertain.

Pegylated-interferon has also been successfully used to treat HEV infections in liver transplant patients. A 3-month course of pegylated IFN α produced a sustained virological response in a few liver transplant recipients and in a patient with a chronic HEV infection who was on hemodialysis. However, IFN α cannot be used in other organ transplant patients because it increases the risk of acute rejection. Consequently, ribavirin therapy is the main drug used to treat chronic HEV infections. Pegylated-IFN α therapy is restricted to liver transplant patients who do not respond to ribavirin. No other therapy is available for other transplant patients (Fig. 4).

Lastly, a few cases of chronic HEV infections in immunocompromised patients other than SOT has been reported. Pegylated-IFN α , ribavirin or a combination of the two was used to effectively treat HEV infections in patients with hematological disease receiving chemotherapy and patients who were HIV-positive.

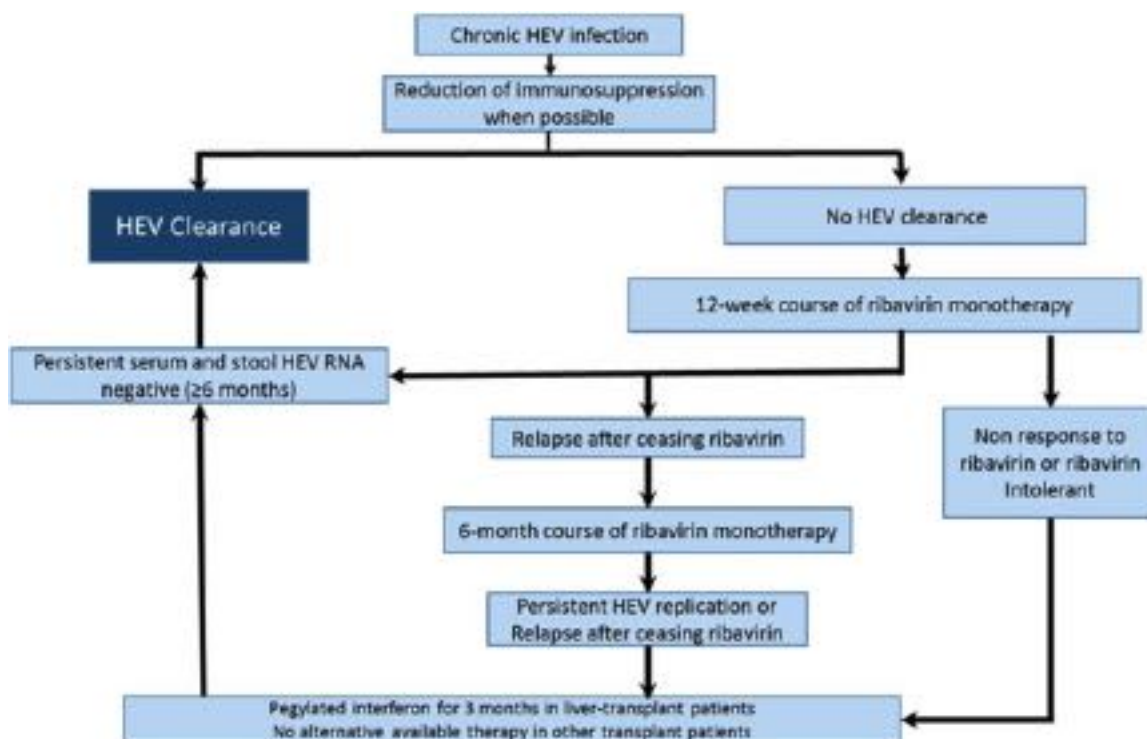


Fig. 4 Treatment algorithm for immunocompromised patients who are solid organ transplant recipients.

Conclusion

While HAV is an ancient virus, it remains a frequent cause of disease that affects millions of individuals annually despite the availability of efficacious, safe hepatitis A vaccines. The worldwide improvements in sanitary and socio-economic conditions has shifted the epidemiological pattern of hepatitis A. The vaccination of infants could be an efficient way of controlling the infection in an entire population.

HEV is a major health problem in developing countries, where it is responsible for sporadic outbreaks. Improved sanitary conditions and the availability of a vaccine will help decrease the incidence of infections in these countries. HEV infections also cause several problems in developed countries. These include zoonotic transmission, transfusion transmission, persistence in immunocompromised patients, and severe forms in patient with pre-existing liver diseases. Any meat at risk should be thoroughly cooked to prevent the food-borne transmission of HEV. Strategies to prevent transfusion transmitted infections must be implemented according to the prevalence of HEV in each specific country. The use of ribavirin has been a major advance in the management of chronic infections in immunocompromised patients. However, new anti-HEV compounds are needed to treat patients who relapse or do not tolerate ribavirin.

Further Reading

- Dalton, H.R., Kamar, N., van Eijk, J.J., *et al.*, 2016. Hepatitis E virus and neurological injury. *Nature Reviews Neurology* 12, 77–85.
- European Association for the Study of the Liver, 2018. EASL clinical practice guidelines on hepatitis E virus infection *Journal of Hepatology* 68, 1256–1271.
- Hofmeister, M.G., Foster, M.A., Teshale, E.H., 2019. Epidemiology and transmission of hepatitis A virus and hepatitis E virus infections in the United States. *Cold Spring Harbor Perspectives in Medicine* 9.
- Kamar, N., Izopet, J., Pavio, N., *et al.*, 2017. Hepatitis E virus infection. *Nature Reviews Disease Primers* 3, 17086.
- Lemon, S.M., Ott, J.J., Van Damme, P., Shouval, D., 2018. Type A viral hepatitis: A summary and update on the molecular virology, epidemiology, pathogenesis and prevention. *Journal of Hepatology* 68.
- Nimgaonkar, I., Ding, Q., Schwartz, R.E., Ploss, A., 2018. Hepatitis E virus: Advances and challenges. *Nature Reviews Gastroenterology & Hepatology* 15, 96–110.
- Victor, J.C., Monto, A.S., Surdina, T.Y., *et al.*, 2007. Hepatitis A vaccine versus immune globulin for postexposure prophylaxis. *New England Journal of Medicine* 357, 1685–1694.

Relevant Websites

<https://ecdc.europa.eu/en/hepatitis-a>
Hepatitis A.
ECDC.
Europa EU.

<https://www.who.int/immunization/diseases/hepatitisA/en/>

Hepatitis A.
World Health Organization.

<https://www.who.int/news-room/fact-sheets/detail/hepatitis-a>

Hepatitis A.
World Health Organization.

https://www.who.int/immunization/policy/position_papers/hepatitis_a/en/

Hepatitis A position paper.
WHO.

Management of Patients With Chronic Hepatitis B (*Hepadnaviridae*) and Chronic Hepatitis D Infection (*Deltavirus*)

Milan J Sonneveld, Erasmus University Medical Center, Rotterdam, The Netherlands

Suzanne van Meer, University Medical Center Utrecht, Utrecht, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

ADV Adefovir

CHB Chronic hepatitis B

ETV Entecavir

HBeAg Hepatitis B e antigen

HBsAg Hepatitis B surface antigen

LAM Lamivudine

TAF Tenofovir alafenamide

TBV Telbivudine

TDF Tenofovir disoproxil fumarate

Glossary

HBsAg loss Clearance of HBsAg from serum, functional cure of hepatitis B.

NA Nucleo(s)tide analog. Direct acting antivirals that suppress HBV DNA. Examples include entecavir and

tenofovir disoproxil fumarate, and tenofovir alafenamide.

PEG-IFN Pegylated interferon alfa, an immunomodulatory used for finite therapy.

Management of Chronic Hepatitis B and Hepatitis D

Chronic Hepatitis B

Introduction

Management of chronic hepatitis B (CHB) requires careful consideration of treatment indications given the requirement for long-term therapy in most patients. For patients who are in need of therapy, options include nucleo(s)tide analogs (NA) and pegylated interferon alfa (PEG-IFN), both with their own advantages and limitations.

Treatment Indications

The indications for treatment in patients with CHB are based on a risk profile conferred by a combination of (1) the extent of hepatic fibrosis, (2) the level of viremia and (3) the severity of liver inflammation. They are based on observations from long-term follow-up studies which have shown that these are the main determinants of progression of liver disease to cirrhosis, hepatocellular carcinoma (HCC) and liver-related mortality (Terrault *et al.*, 2016; European Association for the Study of the Liver, 2017; Chen *et al.*, 2006, 2011).

Current treatment guidelines from both the European Association for the Study of the Liver (EASL) and the American Association of Study of Liver Disease (AASLD) recognize that all patients with cirrhosis and viremia require urgent treatment. In patients without cirrhosis, a combination of HBV DNA levels, ALT levels and extent of liver inflammation and fibrosis is used to determine indication for treatment (Table 1) (European Association for the Study of the Liver, 2017; Terrault *et al.*, 2018a). Liver biopsy is still the gold standard for assessment of hepatic fibrosis, but it is associated with significant risks to the patient. As a result, it has been mostly supplanted by either non-invasive indices that predict the extent of fibrosis (such as the aspartate aminotransferase – to platelet ratio index [APRI] and fibrosis score based on 4 factors [FIB-4]) and/or liver stiffness-based assessment (e.g., Fibroscan). It is important to note that high rates of misclassification have been reported with both APRI and FIB-4, and that liver stiffness based methods may overestimate the extent of fibrosis in the presence of significant liver inflammation (European Association for the Study of the Liver and Asociacion Latinoamericana para el Estudio del H, 2015; Sonneveld *et al.*, 2019a).

In patients not meeting standard treatment criteria, antiviral therapy may be considered in patients at higher risk of HCC, such as those with a family history of HCC, or HBeAg positive patients aged over 30 with high levels of HBV DNA but persistently normal ALT levels (European Association for the Study of the Liver, 2017). Patients with extrahepatic manifestations may also be treated even if they do not meet conventional treatment criteria (European Association for the Study of the Liver, 2017).

Treatment Goals in CHB

Eradication of HBV from infected host hepatocytes cannot be achieved with currently available drugs due to persistence of covalently closed circular (ccc)DNA despite antiviral therapy. The aim of treatment is therefore to reduce or stop progression of

Table 1 Treatment indications

Guidelines	HBeAg positive			HBeAg negative		
	HBV DNA	ALT	Liver Histology	HBV DNA	ALT	Liver histology
AASLD 2018	N/A	N/A	Cirrhosis	N/A	N/A	Cirrhosis
EASL 2017	>20,000	>2 × ULN	N/A	>2000	>2 × ULN	N/A
	>2000	>ULN	At least moderate inflammation and/or fibrosis	>2000	>ULN	At least moderate inflammation and/or fibrosis
	>20,000 detectable	>2 × ULN N/A	N/A Cirrhosis	>20,000 detectable	>2 × ULN N/A	N/A Cirrhosis

Abbreviations: ULN, upper limit of the normal range; AASLD, American Association for the study of Liver Diseases; EASL, European Association for the Study of the Liver; N/A, not applicable.

liver inflammation to liver fibrosis and cirrhosis, and to prevent occurrence of hepatocellular carcinoma (HCC). Various surrogate endpoints are used to assess treatment efficacy in clinical practice (Terrault *et al.*, 2016; European Association for the Study of the Liver, 2017). Frequently used endpoints include normalization of ALT (biochemical response), loss of HBeAg from serum with or without seroconversion to anti-HBe (serological response), suppression of HBV DNA to low (<2000 IU/mL) or undetectable levels (virological response), and improvement of liver histology. Complete remission of disease (sometimes called functional cure) is defined by a loss of HBsAg from serum, since the HBsAg negative state is associated with an excellent prognosis with very low relapse rates in immunocompetent patients (Perrillo, 2006; van Zonneveld *et al.*, 2004). For patients treated with NA, HBV DNA kinetics may be further categorized as complete virological response (HBV DNA undetectability) or partial virological response (at least 1 log HBV DNA decrease after commencing therapy, but persistently detectable HBV DNA at 1 year). Virological breakthrough is defined as a 1 log increase of HBV DNA above the nadir and may be the result of emergence of non-compliance or, more rarely, emergence of viral resistance.

Treatment Options

Current treatment options for patients with CHB consist of the nucleos(t)ide analogs (NA) and pegylated interferon alfa (PEG-IFN). Both classes have different modes of action that translate into different treatment durations and different treatment goals.

Nucleos(t)ide Analogs

NA are oral antiviral agents that inhibit the HBV polymerase/reverse transcriptase and cause chain termination. Until recently there were 5 available NA, including entecavir (ETV), tenofovir disoproxil fumarate (TDF), adefovir dipivoxil (ADV), telbivudine (TBV) and lamivudine (LAM). Recently, the novel tenofovir prodrug tenofovir alafenamide has been approved for the management of CHB, based on the results of an ongoing phase 3 trial showing non-inferiority with regard to antiviral efficacy when compared to TDF (Agarwal *et al.*, 2018). Because continuous treatment with LAM, TBV and ADV is associated with high rates of resistance and subsequent therapy failure these drugs are no longer recommended for the treatment of non-pregnant patients with CHB.

Among treatment naïve patients long-term therapy with ETV or TDF results in near universal HBV DNA suppression to undetectable levels in compliant patients (Zoutendijk *et al.*, 2011; Buti *et al.*, 2015). Despite long-term HBV DNA suppression the rates of HBsAg loss remain low, often <1% per year (Zoutendijk *et al.*, 2011; Buti *et al.*, 2015). As a result, long-term therapy is required for most patients. Antiviral resistance has not yet been demonstrated definitively for TDF, and resistance rates with ETV appear to be very low among patients not previously exposed to LAM. As a result, failure to achieve undetectable HBV DNA is frequently caused by either non-compliance or may be the result of a delayed response due to high baseline viremia. Continuation of antiviral therapy in the absence of viral breakthrough has been shown to be successful in most patients (Zoutendijk *et al.*, 2011). However, failure to achieve complete HBV DNA undetectability is observed in a small minority of patients. Persistent low level viraemia has been associated with a higher risk of adverse outcomes in patients with cirrhosis, but this has so far not been demonstrated for patients without advanced liver disease (Kim *et al.*, 2017). Treatment adaptation is therefore advised in cirrhotics with low level viraemia (European Association for the Study of the Liver, 2017). For non-cirrhotic patients low levels of detectable HBV DNA (<69 IU/mL) do not have to prompt therapy adaptation and may be sometimes accepted (European Association for the Study of the Liver, 2017).

Safety

NA are generally well-tolerated. The most important safety issues are associated with long-term therapy. In particular, long-term ADV and TDF therapy has been associated with reductions in bone mineral density and creatinine clearance, and in rare cases

Table 2 Management of NA failure

Failed NA	Options
Lamivudine (LAM)	Switch to TDF or TAF
Adefovir (ADV)	LAM naïve: switch to entecavir (or TDF or TAF) LAM experienced: switch to TDF or TAF
Telbivudine (TBV)	Switch to TDF or TAF
Entecavir (ETV)	Switch to TDF or TAF
Tenofovir DF (TDF) or Tenofovir AF (TAF)	Switch to ETV
Multidrug resistance	Switch to TDF or TAF Consider TDF + ETV

Abbreviations: TDF, tenofovir disoproxil fumarate; TAF, tenofovir alafenamide; ETV, entecavir; LAM, lamivudine.

development of Fanconi syndrome (Udompap *et al.*, 2018). The risk of renal and bone events may be lower with TAF when compared to TDF, although Fanconi syndrome has also been observed in a patient treated with TAF (Agarwal *et al.*, 2018; Bahr and Yarlagadda, 2019; Grossi *et al.*, 2017). Monitoring of renal function and phosphate levels is therefore advised in patients treated with any of these agents. Among patients with severe liver disease, several case reports have suggested a risk of severe lactic acidosis with ETV therapy, although ETV has also been used safely in patients with decompensated cirrhosis (European Association for the Study of the Liver, 2017; Lange *et al.*, 2009; Hung *et al.*, 2019). In any case, careful monitoring of liver and renal functions is always indicated in patients with advanced liver disease.

Choosing the Appropriate First-Line Agent

Among treatment naïve patients without significant comorbidities ETV, TDF and TAF are all excellent options. Based on the possible risk of renal and bone events in patients treated with TDF, patients at increased risk should preferably be treated with ETV or TAF. Among patients with a reduced renal function, ETV (adjusted dose) or TAF (no dose reduction if clearance >15 mL/min) are the best options.

Management of Patients With NA Failure

In patients who failed on older generations of NA (i.e., LAM, TBV or ADV), resistance testing may be performed to confirm presence of mutations that confer reduced susceptibility to NA. Mutations associated with LAM resistance may confer cross-resistance to ETV, whereas ADV resistance has been associated with somewhat lower response rates to TDF when compared to treatment naïve patients. However, long-term TDF has been successful in the majority of patients with ADV treatment failure. TDF monotherapy has also been effective in patients with ETV failure, with combination therapy only required for a select few patients with previous LAM failure who experience partial responses after long-term TDF therapy (Patterson *et al.*, 2011; Lim *et al.*, 2019; Zoulim *et al.*, 2016). Based on these findings, recommendations have been made for treatment adaptation in patients who failed on NA therapy (Table 2).

Finite NA Therapy

All current guidelines recognize that therapy may be discontinued in patients who have confirmed HBsAg negativity (European Association for the Study of the Liver, 2017). Relapse rates after drug withdrawal in HBsAg negative patients are very low in the absence of immune suppression and irrespective of anti-HBs seroconversion. Given the rarity of HBsAg loss with NA therapy, several studies have examined the possibility of discontinuation of NA in HBsAg positive patients. Recent studies show near universal increases in HBV DNA after drug discontinuation, with a subset of patients experiencing significant flares. However, a subgroup of patients achieves sustained HBV DNA levels <2000 IU/mL or even HBsAg loss without a need for further treatment (Berg *et al.*, 2017; Chang *et al.*, 2015; Hadziyannis *et al.*, 2012). Since these favorable outcomes occur in only a minority of patients, current focus is on identifying patients most likely to achieve off-treatment sustained responses. While various factors have been shown to be associated with sustained disease remission, including serum levels of hepatitis B core related antigen (HBcrAg) and HBsAg, no definitive recommendations can yet be made on patient selection (Sonneveld *et al.*, 2019b; Hsu *et al.*, 2019). Nevertheless, guidelines suggest that therapy discontinuation may be considered in select patients without cirrhosis, in particular HBeAg positive patients who experienced HBeAg seroconversion and completed at least 12 months of consolidation therapy, and HBeAg negative patients who have been successfully treated for multiple (at least 3) years (European Association for the Study of the Liver, 2017). Careful monitoring is always indicated after drug withdrawal as major flares may occur. At this time, drug withdrawal is only advised in the context of studies.

Pegylated Interferon Alfa

Interferon alfa (IFN) has both immunomodulatory and direct antiviral effects in patients with CHB. The availability of the pegylated IFN alfa (PEG-IFN) has allowed once weekly subcutaneous dosing. PEG-IFN is used as a finite treatment course, typically for one year. Response is then assessed at 6–12 months post-treatment. Its main advantage is that it offers a chance at finite treatment for a subset of patients with a response, although this is offset by reduced tolerability when compared to the NA. Responses are defined differently for HBeAg positive and HBeAg negative patients and response rates vary widely across subgroups. Careful patient selection is therefore key to effective use of this agent in clinical practice.

Efficacy in HBeAg-Positive Patients

HBeAg seroconversion was observed in 29%–30% of patients at 6 months post-treatment in two pivotal trials (Janssen *et al.*, 2005; Lau *et al.*, 2005). A combined endpoint of HBeAg loss with HBV DNA <2000 IU/mL at 6 months post-treatment was achieved in 23% in a recent meta-analysis of 3 global randomized trials (Sonneveld *et al.*, 2013b), and HBsAg seroconversion was achieved in 4%–6% (Janssen *et al.*, 2005; Lau *et al.*, 2005). A long-term follow-up study of a predominantly European cohort showed that serological and virological response rates were sustained in the majority of initial responders through a mean of 3 years of follow-up, with increasing rates of HBsAg loss up to 30% in patients who became HBeAg negative after treatment (Buster *et al.*, 2008).

Efficacy in HBeAg-Negative Patients

In the registration trial for PEG-IFN in HBeAg-negative CHB response was defined as an HBV DNA level <20,000 copies/mL with normalization of ALT at 24 weeks post-treatment. This was observed in 36% (Marcellin *et al.*, 2004). In another study comparing PEG-IFN alone with PEG-IFN plus ribavirin in a predominantly European cohort, response rates (HBV DNA <10,000 with normal ALT) at 24 weeks post-treatment were similar in both groups (16%–20%) (Rijckborst *et al.*, 2010b). Long-term virological remission of HBeAg-negative disease, defined as sustained HBV DNA levels <2000 IU/mL, is achieved in around 23% of treated patients (Marcellin *et al.*, 2009). Patients with an initial response may experience relapse during long-term follow-up, although a subset may also achieve HBsAg loss (Marcellin *et al.*, 2009).

Safety of PEG-IFN in CHB

PEG-IFN treatment is associated with considerable side-effects. The most frequently reported side-effects were a flu-like syndrome, myalgia, headache, fatigue and local reactions at the site of subcutaneous injection (Janssen *et al.*, 2005; Lau *et al.*, 2005; Marcellin *et al.*, 2004). PEG-IFN therapy may result in hepatitis flares, which may lead to decompensation and death in patients with advanced liver disease. PEG-IFN is therefore contra-indicated in patients with (a history of) decompensated cirrhosis (Perrillo, 1989; Perrillo *et al.*, 1990; Perrillo, 2001; Flink *et al.*, 2005). PEG-IFN treatment is also associated with mild myelosuppressive effects, but PEG-IFN associated cytopenias are easily managed with dose-reductions and only rarely result in clinically significant infection or bleeding (van Zonneveld *et al.*, 2005). The combination of PEG-IFN with TBV is contra-indicated due to a high risk of myopathy.

Selection of Patients for PEG-IFN Therapy

Since response rates to PEG-IFN are limited, baseline selection of patients with the highest probability of response is essential for optimal use of PEG-IFN. Several baseline predictors have been identified that can be used to guide patient selection (Table 3).

HBeAg-Positive Patients

Multiple factors have been identified that are associated with response to PEG-IFN in HBeAg-positive patients. Response rates are highest in those with HBV genotype A and those with higher baseline levels of ALT (Janssen *et al.*, 2005; Lau *et al.*, 2005). Various

Table 3 Baseline factors associated with higher probability of response to pegylated interferon

<i>HBeAg positive patients</i>	<i>HBeAg negative patients</i>
Genotype A	Genotype C
Higher ALT	Higher ALT
Lower HBV DNA	Lower HBV DNA
Lower HBsAg	Lower HBsAg
	Younger age

Table 4 Rapid assessment for eligibility for treatment with pegylated interferon based on a predicted probability of response of >30%

HBV genotype	Consider pegylated interferon for HBeAg positive patients with
A	Either high ALT ($\geq 2 \times \text{ULN}$) or low HBV DNA ($\leq 9 \log \text{ copies/mL}$)
B or C	Both high ALT ($\geq 2 \times \text{ULN}$) and low HBV DNA ($\leq 9 \log \text{ copies/mL}$)
D	PEG-IFN generally not recommended due to low response rates

Note: Buster, E.H., Hansen, B.E., Lau, G.K., *et al.*, 2009. Factors that predict response of patients with hepatitis B e antigen-positive chronic hepatitis B to peginterferon-alfa. *Gastroenterology* 137 (6), 2002–2009.

Table 5 Stopping-rule for HBeAg positive patients treated with pegylated interferon based on HBsAg levels during therapy

Week	HBV Genotype			
	A	B	C	D
12 HBsAg	No decline	>20,000	>20,000	No decline
24 HBsAg	>20,000	>20,000	>20,000	>20,000

Note: At week 12, HBV genotype specific stopping rules are applied. No genotyping is necessary at week 24.

other factors, including lower baseline HBsAg levels, have been identified in subsequent meta-analyses (Buster *et al.*, 2009; Chan *et al.*, 2018). Based on these findings we proposed a simple scoring system to rapidly screen patients for potential PEG-IFN eligibility (Table 4) (Buster *et al.*, 2009). A more extensive scoring system may subsequently be used for individual patient counseling (Buster *et al.*, 2009; Chan *et al.*, 2018).

Future studies are focussing on genetic predictors of response, novel cytokines and viral mutants that may together be used to optimized individualized decisions (Sonneveld *et al.*, 2012a,b, 2013a; Brouwer *et al.*, 2019). These are however not yet ready for clinical application.

HBeAg-Negative Patients

Less data are available on predictors of response to PEG-IFN in HBeAg-negative patients, which is unfortunate given the even lower response rates in this subgroup. A recent meta-analysis showed higher rates of response in patients with HBV genotype C and younger patients, with inconsistent results regarding HBV DNA levels and HBsAg levels across HBV genotypes (Table 4). A scoring system based on this meta-analysis could be used for assessment of the probability of treatment success (Lampertico *et al.*, 2018).

On-Treatment Prediction of Response

HBeAg-Positive Patients

On-treatment monitoring of viral replication using HBV DNA, HBeAg and HBsAg levels may help to predict response during therapy. On-treatment HBV DNA and HBeAg decline is associated with higher rates of response, but diagnostic accuracy is suboptimal (Fried *et al.*, 2008; Sonneveld *et al.*, 2013c). Recent studies have shown a strong association between on-treatment HBsAg levels and the probability of response. Failure to achieve on-treatment declines and/or persistently high HBsAg levels were shown to predict failure with high accuracy (Sonneveld *et al.*, 2010, 2013b,d). Based on these findings, HBV genotype specific stopping-rules have been defined for HBeAg-positive patients treated with PEG-IFN (Table 5) (Sonneveld *et al.*, 2013b).

HBeAg-Negative Patients

Similar to observations in HBeAg-positive disease, HBV DNA decline on-treatment is associated with treatment response in HBeAg negative CHB patients treated with PEG-IFN (Takkenberg *et al.*, 2009). Subsequent studies have also shown that failure to achieve an on-treatment HBsAg decline is associated with non-response to therapy (Moucari *et al.*, 2009; Marcellin *et al.*, 2010). In the landmark European PARC study, which enrolled predominantly genotype D patients, a combination of HBsAg and HBV DNA levels allowed for reliable prediction of non-response to PEG-IFN. Patients who did not have an HBsAg decline by week 12, and who also failed to achieve a decline of >2 log IU/mL of HBV DNA, had no chance of achieving a sustained response (Rijckborst *et al.*, 2010a). These findings were subsequently validated in two large independent trials, with superior performance among patients with HBV genotype D (Sonneveld *et al.*, 2012a). Current guidelines therefore recognize that patients complying with these criteria should discontinue therapy for futility.

Optimizing Response Rates: Treatment Prolongation and Combination Therapy

The current licensed treatment duration for PEG-IFN is 48 weeks. Reducing the treatment duration has been shown to result in lower response rates (Cheng *et al.*, 2014; Liaw *et al.*, 2011). One study of genotype D HBeAg negative patients showed higher response rates with 96 weeks of therapy (Lampertico *et al.*, 2013). A non-controlled study of HBeAg positive patients treated with PEG-IFN with LAM or ADV for 96 weeks also showed promising results (Cao *et al.*, 2013; Wang *et al.*, 2016). Confirmation of these findings is however required before this can be implemented in clinical practice. Theoretically, combination treatment with PEG-IFN and NA could increase response rates through complementary modes of action. Studies have however consistently shown higher on-treatment response rates, including HBsAg loss, which have been off-set by off-treatment relapse (Janssen *et al.*, 2005; Lau *et al.*, 2005; Marcellin *et al.*, 2004; Sonneveld *et al.*, 2016). Combination therapy is therefore not advised.

Optimizing Response to NA With PEG-IFN

Various studies have evaluated the efficacy of PEG-IFN add-on therapy to NA. These studies have shown increased HBV DNA, HBeAg and HBsAg declines, and also higher rates of HBeAg seroconversion and HBsAg loss with PEG-IFN add-on when compared to continued NA monotherapy (Brouwer *et al.*, 2015; Bourliere *et al.*, 2017; Lampertico *et al.*, 2019; Chi *et al.*, 2017). However, absolute response rates are low and predictors of sustained response are not yet established. Serum HBsAg levels may be useful to identify a subset more likely to respond, but clinically relevant cut-offs remain elusive (Liem *et al.*, 2019). Current guidelines therefore do not recommend PEG-IFN add-on strategies at this time, although if effective selection tools become available this might be an option for a subset of patients.

Management of CHB in Special Populations

Pregnancy

All pregnant women should be tested for HBV infection. The risk of perinatal transmission in babies born to HBsAg positive mothers may be profoundly reduced by a combination of passive and active vaccination of the newborn, and combination therapy is considered cost-effective in settings with adequate healthcare infrastructure (European Association for the Study of the Liver, 2017; Chen *et al.*, 2013). In the context of severe cost constraint, HBIG administration is sometimes limited to babies born to highly viremic (HBeAg positive) mothers, although this strategy is generally not recommended in guidelines (European Association for the Study of the Liver, 2017; Chen *et al.*, 2013).

The risk of transmission despite vaccination is predominantly determined by the level of viremia. Patients with high HBV DNA levels (>200,000 IU/mL and/or high HBsAg levels (above 4–4.5 log IU/mL)) should be treated with a NA to reduce the probability of perinatal transmission (Song *et al.*, 2019). TDF is the preferred agent based on the extensive clinical experience in CHB and HIV infected patients. In pregnant women with low viremia and without signs of significant liver fibrosis treatment may be deferred until after delivery. Post-partum flares may occur both after therapy discontinuation and in untreated patients.

Immunosuppression

Immunosuppression can cause HBV reactivation which may result in fulminant hepatic failure and death. CHB guidelines recommend screening for past or current HBV infection in all patients undergoing significant immunosuppression. The risk of reactivation is determined by HBsAg status (higher in HBsAg positive patients) and with increasing severity immunosuppression. It is recommended that all HBsAg positive patients at risk for reactivation receive prophylactic antiviral therapy. Among HBsAg negative, anti-HBc positive patients, the risk of reactivation should be used to guide decisions on the need for prophylactic therapy. Patients at moderate to high risk should receive prophylaxis.

Chronic Hepatitis Delta

Introduction

Hepatitis delta virus (HDV) is a defective RNA virus that requires HBsAg to complete virion assembly and secretion after autonomous replication. HDV infection can occur as an acute co-infection with HBV and HDV, which evolves to chronicity in only 2% of patients, or as an HDV superinfection in patients with a CHB infection resulting in a chronic HBV-HDV infection in 70%–90% of patients. In most cases of HDV infection, HBV DNA levels are suppressed HDV. HDV is a highly pathogenic virus and can cause severe liver injury that may result in fulminant hepatic failure and rapid progression to cirrhosis and hepatic decompensation, as well as an increased risk of liver cancer (Hughes *et al.*, 2011).

Screening

The EASL and Asian Pacific Association for the Study of the Liver (APASL) guidelines on management of CHB recommend to rule out HDV coinfection in all patients with a CHB. In the AASLD guideline on hepatitis B screening is only advised in patients with specific risk factors including migrants from regions with high HDV endemicity, a history of intravenous drug use or high-risk sexual behavior, individuals infected with HCV or HIV and patients with elevated aminotransferases with low or undetectable HBV DNA ([European Association for the Study of the Liver, 2017](#); [Terrault *et al.*, 2018a](#); [Sarin *et al.*, 2016](#)).

Treatment Indication and Goals

Patients with positive HDV RNA levels have an indication for antiviral therapy as liver disease can progress rapidly. However, risks and benefits of initiating therapy should be considered, including severity of liver disease which may be a contraindication to IFN based therapy ([Terrault *et al.*, 2018b](#)).

A complete virological response of both HBV and HDV infection, thus comprising both HBsAg loss and undetectable HDV RNA levels, is the most desirable endpoint of therapy. However, this occurs only in a small number of patients. The primary aim of anti-HDV therapy is therefore sustained off-treatment HDV RNA undetectability. Due to the high rate of late relapses after initially successful treatment, the definition of “sustained viral response” should be used with caution and long-term monitoring is essential ([Yurdaydin *et al.*, 2019](#)).

Treatment Options

Interferon Alpha

Interferon alpha (IFN) is the only available drug that has been proven to have antiviral efficacy in chronic HDV infection. A Cochrane meta-analysis including five randomized controlled trials comparing IFN treatment for 3–12 months with a no-treatment control group showed that a sustained suppression of HDV RNA at six months of follow-up post-treatment was seen in 17% of patients treated with IFN (compared to 5% in the control group, $p = 0.02$) ([Abbas *et al.*, 2011](#)). Response to IFN based therapy is associated with improved clinical outcomes ([Wranke *et al.*, 2017](#); [Palom *et al.*, 2019](#)). Currently, standard IFN has been replaced by pegylated interferon alpha (PEG-IFN) because of a prolonged plasma half-life time allowing a once-a-week administration. Studies evaluating the efficacy of PEG-IFN for the treatment of chronic HDV infection showed slightly better results, with sustained suppression of HDV RNA at six months post-treatment in 25%–30% of patients ([Wedemeyer *et al.*, 2011](#); [Triantos *et al.*, 2012](#)). Unfortunately, subsequent relapse rates are substantial. Long-term follow-up data from the HIDIT-I trial, the largest randomized trial investigating the safety and efficacy of PEG-IFN with or without adefovir in patients with a chronic HDV infection, showed a late relapse of HDV RNA in more than half of the patients (58%) through approximately four years of follow-up ([Heidrich *et al.*, 2014](#)). Therefore, long-term follow-up with HDV RNA monitoring remains needed in patients with a negative HDV viral load after treatment as long as HBsAg is present in serum ([Heidrich *et al.*, 2014](#)).

The optimal duration of PEG-IFN therapy is not well defined. Current guidelines recommend treatment with PEG-IFN for at least 48 weeks ([European Association for the Study of the Liver, 2017](#); [Terrault *et al.*, 2018a](#); [Sarin *et al.*, 2016](#)). Several studies tried to increase efficacy by prolonging treatment duration. As of yet, treatment prolongation has not been consistently shown to be superior to conventional treatment durations ([Heller *et al.*, 2014](#); [Wedemeyer *et al.*, 2019](#)), although some patients may benefit from a prolonged course of treatment (so-called slow responders) ([Niro *et al.*, 2016](#)). On the other hand, preliminary studies that investigated shorter treatment durations up to six months showed relapses in almost all patients after discontinuation of therapy ([Di Bisceglie *et al.*, 1990](#); [Porres *et al.*, 1989](#)).

Based on existing data from previous studies, an accurate prediction of virological response in patients treated for a chronic HDV infection remains difficult. No clear stopping rules have been defined ([European Association for the Study of the Liver, 2017](#); [Terrault *et al.*, 2018a](#); [Sarin *et al.*, 2016](#)). However, studies have demonstrated that HDV RNA levels at week 24 of treatment with PEG-IFN are associated with treatment outcome ([Keskin *et al.*, 2015](#); [Karaca *et al.*, 2013](#); [Yurdaydin *et al.*, 2008](#)). For example, HDV RNA negativity within 24 weeks of therapy was a strong predictor for sustained HDV suppression at 24 weeks post-treatment with a positive predicted value of 71%. Conversely, a decrease in HDV RNA level of less than 1 log combined with no decrease in HBsAg level effectively identified non-responders ([Keskin *et al.*, 2015](#)). Confirmation of these findings is required before they can be implemented in clinical practice.

Nucleo(s)ide Analogs

The use of nucleo(s)ide analogs (NA) has been extensively explored in chronic HDV infected patients. Several studies investigated the effect of NA treatment as monotherapy, but no effect on HDV viremia was observed ([Wedemeyer *et al.*, 2011](#); [Yurdaydin *et al.*, 2008, 2002](#); [Niro *et al.*, 2005](#)). Furthermore, addition of NA to PEG-IFN therapy did not improve treatment outcome in several randomized trials ([Wedemeyer *et al.*, 2011, 2019](#); [Abbas *et al.*, 2016](#)).

Although NA are not effective against HDV, use of NA is recommended in patients with active HBV replication with persistent elevated HBV DNA levels (>2000 IU/mL), and/or in presence of cirrhosis (European Association for the Study of the Liver, 2017; Sarin *et al.*, 2016; Terrault *et al.*, 2018b), as predominance of both viruses can fluctuate during the natural course of disease and since HBV DNA levels may be associated with more rapid disease progression (Schaper *et al.*, 2010).

Future

New therapies are urgently needed for the treatment of chronic HDV, because of the overall poor treatment response and high rate of relapse after treatment with PEG-IFN. Several new drugs targeting various steps of the HBV and HDV life cycle, such as lonafarnib, Myrcludex-B, PEG-IFN lambda and nucleic Acid Polymers (NAPs), are being evaluated in phase 1 and 2 studies and have shown promise. Registration studies are expected to be commencing soon (Yurdaydin *et al.*, 2019).

References

- Abbas, Z., Khan, M.A., Salih, M., Jafri, W., 2011. Interferon alpha for chronic hepatitis D. *Cochrane Database of Systematic Reviews* 12, CD006002.
- Abbas, Z., Memon, M.S., Umer, M.A., Abbas, M., Shazi, L., 2016. Co-treatment with pegylated interferon alfa-2a and entecavir for hepatitis D: A randomized trial. *World Journal of Hepatology* 8 (14), 625–631.
- Agarwal, K., Brunetto, M., Seto, W.K., *et al.*, 2018. 96 weeks treatment of tenofovir alafenamide vs. tenofovir disoproxil fumarate for hepatitis B virus infection. *Journal of Hepatology* 68 (4), 672–681.
- Bahr, N.C., Yarlagaadda, S.G., 2019. Fanconi syndrome and tenofovir alafenamide: A case report. *Annals of Internal Medicine* 170 (11), 814–815.
- Berg, T., Simon, K.G., Mauss, S., *et al.*, 2017. Long-term response after stopping tenofovir disoproxil fumarate in non-cirrhotic HBeAg-negative patients – FINITE study. *Journal of Hepatology* 67 (5), 918–924.
- Bourliere, M., Rabiaga, P., Ganne-Carrie, N., *et al.*, 2017. Effect on HBs antigen clearance of addition of pegylated interferon alfa-2a to nucleos(t)ide analogue therapy versus nucleos(t)ide analogue therapy alone in patients with HBe antigen-negative chronic hepatitis B and sustained undetectable plasma hepatitis B virus DNA: A randomised, controlled, open-label trial. *Lancet Gastroenterology & Hepatology* 2 (3), 177–188.
- Brouwer, W.P., Chan, H.L., Lampertico, P., *et al.*, 2019. Genome wide association study identifies genetic variants associated with early and sustained response to (Peg) interferon in chronic hepatitis B patients: The GIAN-T-B study. *Clinical Infectious Diseases*.
- Brouwer, W.P., Xie, Q., Sonneveld, M.J., *et al.*, 2015. Adding pegylated interferon to entecavir for hepatitis B e antigen-positive chronic hepatitis B: A multicenter randomized trial (ARES study). *Hepatology* 61 (5), 1512–1522.
- Buster, E.H., Flink, H.J., Cakaloglu, Y., *et al.*, 2008. Sustained HBeAg and HBsAg loss after long-term follow-up of HBeAg-positive patients treated with peginterferon alpha-2b. *Gastroenterology* 135 (2), 459–467.
- Buster, E.H., Hansen, B.E., Lau, G.K., *et al.*, 2009. Factors that predict response of patients with hepatitis B e antigen-positive chronic hepatitis B to peginterferon-alfa. *Gastroenterology* 137 (6), 2002–2009.
- Buti, M., Tsai, N., Petersen, J., *et al.*, 2015. Seven-year efficacy and safety of treatment with tenofovir disoproxil fumarate for chronic hepatitis B virus infection. *Digestive Diseases and Sciences* 60 (5), 1457–1464.
- Cao, Z.H., Ma, L.N., Zhang, H.W., Liu, Y.L., Chen, X.Y., 2013. Extended treatment with peginterferon alpha-2a in combination with lamivudine or adefovir for 96 weeks yields high rates of HBeAg and HBsAg seroconversion. *Journal of Digestive Diseases* 14 (8), 446–450.
- Chan, H.L.Y., Messinger, D., Papatheodoridis, G.V., *et al.*, 2018. A baseline tool for predicting response to peginterferon alfa-2a in HBeAg-positive patients with chronic hepatitis B. *Aliment Pharmacol Ther* 48 (5), 547–555.
- Chang, M.L., Liaw, Y.F., Hadziyannis, S.J., 2015. Systematic review: Cessation of long-term nucleos(t)ide analogue therapy in patients with hepatitis B e antigen-negative chronic hepatitis B. *Alimentary Pharmacology & Therapeutics* 42 (3), 243–257.
- Chen, C.F., Lee, W.C., Yang, H.I., *et al.*, 2011. Changes in serum levels of HBV DNA and alanine aminotransferase determine risk for hepatocellular carcinoma. *Gastroenterology* 141 (4), 1240–1248.
- Chen, S.C., Toy, M., Yeh, J.M., Wang, J.D., Resch, S., 2013. Cost-effectiveness of augmenting universal hepatitis B vaccination with immunoglobulin treatment. *Pediatrics* 131 (4), e1135–e1143.
- Chen, C.J., Yang, H.I., Su, J., *et al.*, 2006. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *JAMA* 295 (1), 65–73.
- Cheng, J., Wang, Y., Hou, J., *et al.*, 2014. Peginterferon alfa-2b in the treatment of Chinese patients with HBeAg-positive chronic hepatitis B: A randomized trial. *Journal of Clinical Virology* 61 (4), 509–516.
- Chi, H., Hansen, B.E., Guo, S., *et al.*, 2017. Pegylated interferon alfa-2b add-on treatment in hepatitis B virus envelope antigen-positive chronic hepatitis B patients treated with nucleos(t)ide analogue: A randomized, controlled trial (PEGON). *Journal of Infectious Diseases* 215 (7), 1085–1093.
- Deterding, K., Wedemeyer, H., 2019. Beyond pegylated interferon-alpha: New treatments for hepatitis delta. *AIDS Reviews* 21 (3), 126–134. doi:10.24875/AIDSRev.19000080.
- Di Bisceglie, A.M., Martin, P., Lisker-Melman, M., *et al.*, 1990. Therapy of chronic delta hepatitis with interferon alfa-2b. *Journal of Hepatology* 11 (Suppl 1), S151–S154.
- European Association for the Study of Liver and Asociacion Latinoamericana para el Estudio del H, 2015. EASL-ALEH clinical practice guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *Journal of Hepatology* 63 (1), 237–264.
- European Association for the Study of the Liver, 2017. EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection. *Journal of Hepatology* 67 (2), 370–398.
- Flink, H.J., Sprengers, D., Hansen, B.E., *et al.*, 2005. Flares in chronic hepatitis B patients induced by the host or the virus? Relation to treatment response during Peg-interferon (alpha) – 2b therapy. *Gut* 54 (11), 1604–1609.
- Fried, M.W., Piratvisuth, T., Lau, G.K., *et al.*, 2008. HBeAg and hepatitis B virus DNA as outcome predictors during therapy with peginterferon alfa-2a for HBeAg-positive chronic hepatitis B. *Hepatology* 47 (2), 428–434.
- Grossi, G., Loglio, A., Facchetti, F., *et al.*, 2017. Tenofovir alafenamide as a rescue therapy in a patient with HBV-cirrhosis with a history of Fanconi syndrome and multidrug resistance. *Journal of Hepatology*.
- Hadziyannis, S.J., Sevastianos, V., Rapti, I., Vassilopoulos, D., Hadziyannis, E., 2012. Sustained responses and loss of HBsAg in HBeAg-negative patients with chronic hepatitis B who stop long-term treatment with adefovir. *Gastroenterology* 143 (3), 629–636.
- Heidrich, B., Yurdaydin, C., Kabacam, G., *et al.*, 2014. Late HDV RNA relapse after peginterferon alpha-based therapy of chronic hepatitis delta. *Hepatology* 60 (1), 87–97.
- Heller, T., Rotman, Y., Koh, C., *et al.*, 2014. Long-term therapy of chronic delta hepatitis with peginterferon alfa. *Alimentary Pharmacology & Therapeutics* 40 (1), 93–104.

- Hsu, Y.C., Nguyen, M.H., Mo, L.R., *et al.*, 2019. Combining hepatitis B core-related and surface antigens at end of nucleos(t)ide analogue treatment to predict off-therapy relapse risk. *Alimentary Pharmacology & Therapeutics* 49 (1), 107–115.
- Hughes, S.A., Wedemeyer, H., Harrison, P.M., 2011. Hepatitis delta virus. *Lancet* 378 (9785), 73–85.
- Hung, T.H., Tsai, C.C., Lee, H.F., 2019. Population-based study of Entecavir and long-term mortality in chronic hepatitis B-related decompensated liver cirrhosis. *Clinics and Research in Hepatology and Gastroenterology*.
- Janssen, H.L., van Zonneveld, M., Senturk, H., *et al.*, 2005. Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: A randomised trial. *Lancet* 365 (9454), 123–129.
- Karaca, C., Soyer, O.M., Baran, B., *et al.*, 2013. Efficacy of pegylated interferon-alpha treatment for 24 months in chronic delta hepatitis and predictors of response. *Antiviral Therapy* 18 (4), 561–566.
- Keskin, O., Wedemeyer, H., Tuzun, A., *et al.*, 2015. Association between level of hepatitis D virus RNA at week 24 of pegylated interferon therapy and outcome. *Clinical Gastroenterology and Hepatology* 13 (13), 2342–2349. [e1-2].
- Kim, J.H., Sinn, D.H., Kang, W., *et al.*, 2017. Low-level viremia and the increased risk of hepatocellular carcinoma in patients receiving entecavir treatment. *Hepatology* 66 (2), 335–343.
- Lampertico, P., Brunetto, M.R., Craxi, A., *et al.*, 2019. Add-on peginterferon alfa-2a to nucleos(t)ide analogue therapy for Caucasian patients with hepatitis B 'e' antigen-negative chronic hepatitis B genotype D. *Journal of Viral Hepatitis* 26 (1), 118–125.
- Lampertico, P., Messinger, D., Cornberg, M., *et al.*, 2018. A genotype-specific baseline score predicts post-treatment response to peginterferon alfa-2a in hepatitis B e antigen-negative chronic hepatitis B. *Annals of Gastroenterology* 31 (6), 712–721.
- Lampertico, P., Vigano, M., Di Costanzo, G.G., *et al.*, 2013. Randomised study comparing 48 and 96 weeks peginterferon alpha-2a therapy in genotype D HBeAg-negative chronic hepatitis B. *Gut* 62 (2), 290–298.
- Lange, C.M., Bojunga, J., Hofmann, W.P., *et al.*, 2009. Severe lactic acidosis during treatment of chronic hepatitis B with entecavir in patients with impaired liver function. *Hepatology* 50 (6), 2001–2006.
- Lau, G.K., Piratvisuth, T., Luo, K.X., *et al.*, 2005. Peginterferon Alfa-2a, lamivudine, and the combination for HBeAg-positive chronic hepatitis B. *New England Journal of Medicine* 352 (26), 2682–2695.
- Liaw, Y.F., Jia, J.D., Chan, H.L., *et al.*, 2011. Shorter durations and lower doses of peginterferon alfa-2a are associated with inferior hepatitis B e antigen seroconversion rates in hepatitis B virus genotypes B or C. *Hepatology* 54 (5), 1591–1599.
- Liem, K.S., van Campenhout, M.J.H., Xie, Q., *et al.*, 2019. Low hepatitis B surface antigen and HBV DNA levels predict response to the addition of pegylated interferon to entecavir in hepatitis B e antigen positive chronic hepatitis B. *Alimentary Pharmacology & Therapeutics* 49 (4), 448–456.
- Lim, Y.S., Gwak, G.Y., Choi, J., *et al.*, 2019. Monotherapy with tenofovir disoproxil fumarate for adefovir-resistant vs. entecavir-resistant chronic hepatitis B: A 5-year clinical trial. *Journal of Hepatology* 71 (1), 35–44.
- Marcellin, P., Bonino, F., Lau, G.K., *et al.*, 2009. Sustained response of hepatitis B e antigen-negative patients 3 years after treatment with peginterferon alpha-2a. *Gastroenterology* 136 (7), 2169–2179.
- Marcellin, P., Lau, G.K., Bonino, F., *et al.*, 2004. Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. *New England Journal of Medicine* 351 (12), 1206–1217.
- Marcellin, P., Piratvisuth, T., Brunetto, M., *et al.*, 2010. On-treatment decline in serum HBsAg levels predicts sustained immune control 1 year post-treatment and subsequent HBsAg clearance in HBeAg-negative hepatitis B virus-infected patients treated with peginterferon alfa-2a. *Hepatology International* 4, 151.
- Moucari, R., Mackiewicz, V., Lada, O., *et al.*, 2009. Early serum HBsAg drop: A strong predictor of sustained virological response to pegylated interferon alfa-2a in HBeAg-negative patients. *Hepatology* 49 (4), 1151–1157.
- Niro, G.A., Ciancio, A., Tillman, H.L., *et al.*, 2005. Lamivudine therapy in chronic delta hepatitis: A multicentre randomized-controlled pilot study. *Alimentary Pharmacology & Therapeutics* 22 (3), 227–232.
- Niro, G.A., Smedile, A., Fontana, R., *et al.*, 2016. HBsAg kinetics in chronic hepatitis D during interferon therapy: On-treatment prediction of response. *Alimentary Pharmacology & Therapeutics* 44 (6), 620–628.
- Palom, A., Rodriguez-Tajes, S., Navascues, C.A., *et al.*, 2019. Long-term clinical outcomes in patients with chronic hepatitis delta: The role of persistent viraemia. *Alimentary Pharmacology & Therapeutics* 51, 158–166.
- Patterson, S.J., George, J., Strasser, S.I., *et al.*, 2011. Tenofovir disoproxil fumarate rescue therapy following failure of both lamivudine and adefovir dipivoxil in chronic hepatitis B. *Gut* 60 (2), 247–254.
- Perrillo, R.P., 1989. Treatment of chronic hepatitis B with interferon: Experience in western countries. *Seminars in Liver Disease* 9 (4), 240–248.
- Perrillo, R.P., 2001. Acute flares in chronic hepatitis B: The natural and unnatural history of an immunologically mediated liver disease. *Gastroenterology* 120 (4), 1009–1022.
- Perrillo, R.P., 2006. Therapy of hepatitis B – viral suppression or eradication? *Hepatology* 43 (2 Suppl 1), S182–S193.
- Perrillo, R.P., Schiff, E.R., Davis, G.L., *et al.*, 1990. A randomized, controlled trial of interferon alfa-2b alone and after prednisone withdrawal for the treatment of chronic hepatitis B. *The Hepatitis Interventional Therapy Group. The New England Journal of Medicine* 323 (5), 295–301.
- Porres, J.C., Carreno, V., Bartolome, J., *et al.*, 1989. Treatment of chronic delta infection with recombinant human interferon alpha 2c at high doses. *Journal of Hepatology* 9 (3), 338–344.
- Rijckborst, V., Hansen, B.E., Cakaloglu, Y., *et al.*, 2010a. Early on-treatment prediction of response to peginterferon alfa-2a for HBeAg-negative chronic hepatitis B using HBsAg and HBV DNA levels. *Hepatology* 52 (2), 454–461.
- Rijckborst, V., ter Borg, M.J., Cakaloglu, Y., *et al.*, 2010b. A randomized trial of peginterferon alpha-2a with or without ribavirin for HBeAg-negative chronic hepatitis B. *American Journal of Gastroenterology* 105 (8), 1762–1769.
- Sarin, S.K., Kumar, M., Lau, G.K., *et al.*, 2016. Asian-Pacific clinical practice guidelines on the management of hepatitis B: A 2015 update. *Hepatology International* 10 (1), 1–98.
- Schaper, M., Rodriguez-Frias, F., Jardi, R., *et al.*, 2010. Quantitative longitudinal evaluations of hepatitis delta virus RNA and hepatitis B virus DNA shows a dynamic, complex replicative profile in chronic hepatitis B and D. *Journal of Hepatology* 52 (5), 658–664.
- Song, J., Yang, F., Wang, S., *et al.*, 2019. Efficacy and safety of antiviral treatment on blocking the mother-to-child transmission of hepatitis B virus: A meta-analysis. *Journal of Viral Hepatitis* 26 (3), 397–406.
- Sonneveld, M.J., Brouwer, W.P., van der Meer, A.J., 2016. Surface antigen seroreversion: The bane of combination therapy in chronic hepatitis B? *Gastroenterology* 150 (5), 1254–1255.
- Sonneveld, M.J., Rijckborst, V., Boucher, C.A., Hansen, B.E., Janssen, H.L., 2010. Prediction of sustained response to peginterferon alfa-2b for hepatitis B e antigen-positive chronic hepatitis B using on-treatment hepatitis B surface antigen decline. *Hepatology* 52 (4), 1251–1257.
- Sonneveld, M.J., Rijckborst, V., Zeuzem, S., *et al.*, 2012a. Presence of precore and core promoter mutants limits the probability of response to peginterferon in hepatitis B e antigen-positive chronic hepatitis B. *Hepatology* 56 (1), 67–75.
- Sonneveld, M.J., Wong, V.W., Wolman, A.M., *et al.*, 2012b. Polymorphisms near IL28B and serologic response to peginterferon in HBeAg-positive patients with chronic hepatitis B. *Gastroenterology* 142 (3), 513–520.
- Sonneveld, M.J., Arends, P., Boonstra, A., Hansen, B.E., Janssen, H.L., 2013a. Serum levels of interferon-gamma-inducible protein 10 and response to peginterferon therapy in HBeAg-positive chronic hepatitis B. *Journal of Hepatology* 58 (5), 898–903.

- Sonneveld, M.J., Hansen, B.E., Piratvisuth, T., *et al.*, 2013b. Response-guided peginterferon therapy in HBeAg-positive chronic hepatitis B using serum hepatitis B surface antigen levels. *Hepatology* 58, 872–880.
- Sonneveld, M.J., Rijckborst, V., Zwang, L., *et al.*, 2013c. Hepatitis B e antigen levels and response to peginterferon: Influence of precore and basal core promoter mutants. *Antiviral Research* 97 (3), 312–317.
- Sonneveld, M.J., Zoutendijk, R., Flink, H.J., *et al.*, 2013d. Close monitoring of hepatitis B surface antigen levels helps classify flares during peginterferon therapy and predicts treatment response. *Clinical Infectious Diseases* 56 (1), 100–105.
- Sonneveld, M.J., Brouwer, W.P., Chan, H.L., *et al.*, 2019a. Optimisation of the use of APRI and FIB-4 to rule out cirrhosis in patients with chronic hepatitis B: Results from the SONIC-B study. *The Lancet Gastroenterology and Hepatology* 4 (7), 538–544.
- Sonneveld, M.J., van Oord, G.W., van Campenhout, M.J., *et al.*, 2019b. Relationship between hepatitis B core-related antigen levels and sustained HBeAg seroconversion in patients treated with nucleos(t)ide analogues. *Journal of Viral Hepatitis* 26 (7), 828–834.
- Takkenberg, B., Zaaijer, H.L., De Niet, A., *et al.*, 2009. Baseline HBSAg level and on-treatment HBSAg and HBV DNA decline predict sustained virological response in HBeAg negative chronic hepatitis B patients treated with peginterferon alfa-2a and adefovir: An interim analysis. *Hepatology* 50 (4), 261–283.
- Terrault, N.A., Bzowej, N.H., Chang, K.M., *et al.*, 2016. AASLD guidelines for treatment of chronic hepatitis B. *Hepatology* 63 (1), 261–283.
- Terrault, N.A., Lok, A.S.F., McMahon, B.J., *et al.*, 2018a. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Clinical Liver Disease* 12 (1), 33–34.
- Terrault, N.A., Lok, A.S.F., McMahon, B.J., *et al.*, 2018b. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology* 67 (4), 1560–1599.
- Triantos, C., Kalafateli, M., Nikolopoulou, V., Burroughs, A., 2012. Meta-analysis: Antiviral treatment for hepatitis D. *Alimentary Pharmacology & Therapeutics* 35 (6), 663–673.
- Udomap, P., Kim, D., Ahmed, A., Kim, W.R., 2018. Longitudinal trends in renal function in chronic hepatitis B patients receiving oral antiviral treatment. *Alimentary Pharmacology & Therapeutics* 48 (11–12), 1282–1289.
- Wang, Z., Sun, L., Wu, Y., Xia, Q., 2016. Extended duration versus standard duration of peginterferon alfa-2a in treatment of chronic hepatitis B: A systematic review and meta-analysis. *Clinics and Research in Hepatology and Gastroenterology* 40 (2), 195–202.
- Wedemeyer, H., Yurdaydin, C., Dalekos, G.N., *et al.*, 2011. Peginterferon plus adefovir versus either drug alone for hepatitis delta. *New England Journal of Medicine* 364 (4), 322–331.
- Wedemeyer, H., Yurdaydin, C., Hardtke, S., *et al.*, 2019. Peginterferon alfa-2a plus tenofovir disoproxil fumarate for hepatitis D (HIDIT-II): A randomised, placebo controlled, phase 2 trial. *Lancet Infectious Diseases* 19 (3), 275–286.
- Wranke, A., Serrano, B.C., Heidrich, B., *et al.*, 2017. Antiviral treatment and liver-related complications in hepatitis delta. *Hepatology* 65 (2), 414–425.
- Yurdaydin, C., Abbas, Z., Buti, M., *et al.*, 2019. Treating chronic hepatitis delta: The need for surrogate markers of treatment efficacy. *Journal of Hepatology* 70 (5), 1008–1015.
- Yurdaydin, C., Bozkaya, H., Gurel, S., *et al.*, 2002. Famciclovir treatment of chronic delta hepatitis. *Journal of Hepatology* 37 (2), 266–271.
- Yurdaydin, C., Bozkaya, H., Onder, F.O., *et al.*, 2008. Treatment of chronic delta hepatitis with lamivudine vs lamivudine + interferon vs interferon. *Journal of Viral Hepatitis* 15 (4), 314–321.
- van Zonneveld, M., Flink, H.J., Verhey, E., *et al.*, 2005. The safety of pegylated interferon alpha-2b in the treatment of chronic hepatitis B: Predictive factors for dose reduction and treatment discontinuation. *Alimentary Pharmacology & Therapeutics* 21 (9), 1163–1171.
- van Zonneveld, M., Honkoop, P., Hansen, B.E., *et al.*, 2004. Long-term follow-up of alpha-interferon treatment of patients with chronic hepatitis B. *Hepatology* 39 (3), 804–810.
- Zoulim, F., Bialkowska-Warzecha, J., Diclescu, M.M., *et al.*, 2016. Entecavir plus tenofovir combination therapy for chronic hepatitis B in patients with previous nucleos(t)ide treatment failure. *Hepatology International* 10 (5), 779–788.
- Zoutendijk, R., Reijnders, J.G., Brown, A., *et al.*, 2011. Entecavir treatment for chronic hepatitis B: Adaptation is not needed for the majority of naive patients with a partial virological response. *Hepatology* 54 (2), 443–451.

Further Reading

- Lopatin, U., 2019. Drugs in the pipeline for HBV. *Clinical Liver Disease* 23 (3), 535–555. doi:10.1016/j.cld.2019.04.006.
- Seto, W.K., Lo, Y.R., Pawlotsky, J.M., Yuen, M.F., 2018. Chronic hepatitis B virus infection. *Lancet* 392 (10161), 2313–2324. doi:10.1016/S0140-6736(18)31865-8.

Relevant Websites

www.aasld.org
AASLD: Home.

www.easl.eu
EASL.
The Home of Hepatology.

Studying Population Genetic Processes in Viruses: From Drug-Resistance Evolution to Patient Infection Dynamics

Jeffrey D Jensen, Arizona State University, Tempe, AZ, United States

© 2021 Published by Elsevier Ltd.

Glossary

Background Selection Reduction in genetic diversity due to purifying selection against deleterious mutations at linked sites.

Distribution of Fitness Effects (DFE) The statistical distribution of selection coefficients of newly arising mutations, as compared to a reference genotype.

Mutational Meltdown The accumulation of stochastic effects which may result in population decline and ultimately extinction.

Selective Sweep The result of a beneficial mutation being driven to high frequency by positive selection, together with linked variation.

Site Frequency Spectrum (SFS) Summary of the frequencies of observed segregating mutations in a population.

In this article, I briefly summarize the population genetic environment in which within-patient viral populations are evolving – discussing the roles of genetic drift as modulated by the infection history of the patient, selection acting on newly arising deleterious and beneficial variants and the related linked selection effects on the genome, and the underlying mutation and recombination/reassortment rates as well as replication behavior. I conclude with a consideration of how an improved understanding of these processes specifically, and evolutionary genomics in general, can inform therapeutic strategies in the future.

The Population Dynamics of Infection

The natural starting point in any evolutionary analysis is a consideration of the effects of genetic drift. This stochastic evolutionary force is generally described in terms of the effective population size (N_e) – that is, the idealized size of the population which would experience the observed amount of genetic drift. This is opposed to the census population size (N), that is, the current number of individuals in the population. There are multiple reasons why N_e may be strongly reduced relative to N , including changes in census population sizes over time as well as natural selection itself, as will be described throughout this article. While viral population census sizes are known to often be exceptionally large, it is important to appreciate that viral effective population sizes are much more constrained, owing both to underlying modes of transmission and infection dynamics as described below – with N_e having been estimated to only be on the order of hundreds to thousands (Hughes, 2009; Miyashita and Kishino, 2010; Renzette *et al.*, 2013). This is significant with regards to viral population genetics as it suggests an upper limit on the efficiency of natural selection, given that this efficacy is dictated by N_e . It relatedly serves as a reminder of the pervasively important role of genetic drift in governing evolutionary outcomes, and the generally unjustified adaptation-centric view of virus genome evolution still forwarded in certain sections of the virology community.

Thus, before treating the topic of selection, one must firstly consider the neutral population dynamics of the organism in question. In the case of a viral population sampled from a patient, these dynamics likely include a strong population bottleneck (*i.e.*, a temporary reduction in population size) associated with the initial infection – these bottlenecks may be quite severe, with some primary infections being established by a very small number of infecting units. After this initial infection, the viral population will often next enter a phase of rapid population growth. This combination of bottlenecking and growth will leave signatures in the genomic patterns of variation, which in turn may be used to estimate the timing of infection, the severity of the bottleneck, the rate of within-patient population growth, and indeed the relative difference between N_e and N . In certain viruses that compartmentalize, these dynamics might also include population sub-structuring, in which isolated or semi-isolated viral populations are established in different compartments (*e.g.*, in different organs) within a single host. These compartments may or may not be connected by gene flow, via the plasma for example.

Multiple approaches have been developed within the field of population genetics to infer these demographic models and their underlying parameters. It is perhaps most helpful to introduce such methodology within the context of a specific application. For this purpose, I will examine recent work in human cytomegalovirus (HCMV), a member of the Herpesviridae family of dsDNA viruses. HCMV is characterized by a particularly large genome (~235,000 base pairs), and is nearly ubiquitous in human populations, with a seroprevalence of 30%–90% in the US and >90% in adults outside of the developed world (Kenneson and Cannon, 2007; Boeckh and Geballe, 2011). Though asymptomatic in most individuals, HCMV can lead to severe symptoms in immunocompromised patients and neonates, and is the most common cause of birth defects resulting from an infectious agent (Hassan and Connell, 2007; Manicklal *et al.*, 2013).

A large literature has accumulated demonstrating that HCMV has high levels of genetic variation within a host (Spector *et al.*, 1984; Drew *et al.*, 1984; Haberland *et al.*, 1999; Meyer-König *et al.*, 1998; Faure-Della Corte *et al.*, 2010), despite encoding a DNA polymerase with proofreading activity (Nishiyama *et al.*, 1983), with several recent high-throughput whole-genome studies revealing the full extent of this variation (Renzette *et al.*, 2011, 2013, 2014, 2015, 2016, 2017; Hage *et al.*, 2017; Pokalyuk *et al.*, 2017). Utilizing such whole-genome data sampled from the urine and plasma compartments of five congenitally infected infants, Renzette *et al.* (2013) found that the sub-structuring effect of compartmentalization was particularly strong – as assessed by F_{ST} , a measure of differentiation between populations – with these two compartments from a single patient being as diverged as samples from two unrelated patients. Utilizing the observed frequencies of genome-wide segregating sites in these congenitally infected patient samples – the site frequency spectrum (SFS) – they additionally inferred the demographic model characterizing each infection. Specifically, Renzette *et al.* inferred a population bottleneck associated with the initial infection (*i.e.*, the movement of the virus from the maternal compartment to the fetal plasma compartment), and a second bottleneck associated with the subsequent infection of the kidney compartment (*i.e.*, as assessed by the urine sample). The severity of the bottleneck characterizing the initial infection was not as strong as earlier results had suggested, with dozens or even hundreds of virions characterizing fetal infection. By estimating the age of the bottleneck, this work also provided the first inference of the timing of fetal infection.

In follow-up work with longitudinal sampling of the plasma, urine, and saliva compartments, Pokalyuk *et al.* (2017) also found evidence for subsequent admixture with maternal HCMV populations after the initial infection and compartmentalization, suggesting re-infection post-birth via, for example, breast milk (Numazaki, 1997; Enders *et al.*, 2011). While these mixed infections can certainly be a significant player in governing levels and patterns of genomic variation as described by Pokalyuk *et al.*, there has been an unfortunate tendency of some authors to only focus on this admixture process while neglecting the variety of other selective and demographic processes which also determine levels of variation (*e.g.*, Cudini *et al.*, 2019; though see Jensen and Kowalik, 2020).

Given that no current method exists to prevent maternal-fetal transmission, or to reduce the severity of fetal infection (Britt, 2017), such a characterization of population dynamics may be crucial to future therapeutic strategies. For example, clinically imposing a more severe population bottleneck during pregnancy may reduce HCMV variability within the fetus, limiting the pool of variation on which natural selection may subsequently act, thereby potentially improving treatment outcomes. Immunotherapy advances have produced therapeutics capable of reducing the rate of maternal transmission in other viruses, including hepatitis B and HIV, and may thus represent a promising route for HCMV as well (Tseng and Kao, 2017; Voronin *et al.*, 2017).

Direct Selection and Linked Selection Effects

With regards to the direct action of natural selection in viral populations, positive selection related to antiviral drug resistance or immune evasion has received particular attention for obvious reasons (see review of Irwin *et al.*, 2016a). This search for adaptive loci is generally based around identifying the patterns of selective sweeps (Maynard Smith and Haigh, 1974) in viral genomes, owing to the genetic hitchhiking effects induced from the rapid rise in frequency of the beneficial mutation towards fixation. However, though it receives comparatively less attention in the virology literature, it is well understood that the vast majority of new fitness-impacting mutations in any organism are deleterious rather than beneficial (*e.g.*, Crow, 1993; Lynch *et al.*, 1999; Bank *et al.*, 2014b; and see reviews of Eyre-Walker and Keightley, 2007; Bank *et al.*, 2014a). The selective removal of these harmful variants by purifying selection will serve to further reduce the effective population size to an extent largely dictated by the rate of recombination and the strength of selection (Charlesworth *et al.*, 1993; Charlesworth, 2013). In addition, genomic linkage to this frequent input of deleterious mutations will reduce the probability of fixation of other linked sites – including reducing the likelihood of adaptation (Hill and Robertson, 1966; and see Pénisson *et al.*, 2017).

Importantly, just as the increase in frequency of a beneficial mutation can lead to the genetic hitchhiking of linked variation towards fixation in the genome (leading to a selective sweep effect), so too can the removal of the much larger input of deleterious mutations lead to the genetic hitchhiking of linked neutral variants towards loss. This effect – the loss and reduction in frequency owing to linkage with deleterious variants – is known as background selection (Charlesworth *et al.*, 1993). This effect can be of major consequence in viral populations given their coding-dense genomes (in which most new mutations are expected to be deleterious), particularly in the absence of recombination. Though this is yet to be thoroughly studied in viral populations, initial work has found a much more dominant role for background selection relative to selective sweeps in shaping within-patient genomic variation (*e.g.*, Renzette *et al.*, 2016).

Quantifying the relative input of deleterious, neutral, and beneficial mutations, means understanding the shape of the distribution of fitness effects (DFE). There are three general types of methods for performing such inference. The first involves the artificial creation of a mutation, one at a time or in combination, in order to measure the resulting fitness effect on an otherwise wildtype background under lab-controlled environmental conditions (*e.g.*, Fowler *et al.*, 2010; Hietpas *et al.*, 2011, 2012; Bank *et al.*, 2014b). The second approach, mutation-accumulation, maintains a population in a laboratory and allows mutations to naturally occur and accumulate, and fitness is generally quantified and related to this accumulation at multiple time points (*e.g.*, Foll *et al.*, 2014; Ferrer-Admetlla *et al.*, 2016; Lynch *et al.*, 2016; Long *et al.*, 2018). Aside from these two experimental approaches, a third approach is based on the direct sampling of natural population data (*e.g.*, from a patient infection, in the case of viral populations) – using either single time point (*e.g.*, Keightley and Eyre-Walker, 2007; Schneider *et al.*, 2011; Tataru *et al.*, 2017; Johri *et al.*, 2020) or multiple time point data (*e.g.*, Malaspina *et al.*, 2012; Mathieson and McVean, 2013; Foll *et al.*, 2015;

Sackman *et al.*, 2019). These approaches generally rely on fitting a population history using patterns of variation at neutral sites, and then using that history to quantify the DFE at putatively selected sites.

One particularly important aside here of great relevance to the virology community, is the danger of basing such selection analyses on consensus sequences (*i.e.*, the representation of each patient's viral sample by a single sequence, which represents the most common allele at each site in a given within-patient viral population). While common practice, this tremendous loss of information (*e.g.*, all rare alleles – which represent the vast majority of variation – are neglected), combined with the associated ascertainment issue of only viewing high frequency mutations, can wreak havoc on evolutionary inference. For example, Renzette *et al.* (2017) demonstrated that earlier work based on consensus sequences had incorrectly arrived at the conclusion that most sampled segregating variation in HCMV was strongly selected owing to this consensus ascertainment – whereas when full within-patient population-level data was considered, it was clear that most segregating variation was in fact neutral or nearly neutral, consistent with general expectations (Kimura, 1968; Ohta, 1973; and see Jensen *et al.*, 2019; Morales-Arce *et al.*, 2020b).

Finally, it is important in this regard to appreciate that the correct evolutionary null model must necessarily account for the population infection dynamics discussed in the first section, together with the purifying and background selection effects discussed above. As viral populations are characterized by changing population sizes owing to infection bottlenecks and subsequent growth, and as the input of deleterious mutations is a constant process, an accounting for these processes is necessary to establish a baseline expectation for levels and patterns of variation, on top of which positively selected outliers may be identified. A neglect of these neutral dynamics can lead to serious mis-inference. For example, Feder *et al.* (2016) examined consensus HIV-1 population sequence data from 6717 drug-treated patients sequenced over a 24-year span. They found, as expected, that the change in variation observed in the viral population was correlated with the degree of treatment effectiveness. They used this pattern to argue that less effective treatments may be associated with so-called soft selective sweeps (*i.e.*, positive selection on a previously neutrally segregating genomic variant, or on multiple simultaneously occurring and identical genomic variants), whereas more effective treatments may be associated with hard selective sweeps (*i.e.*, positive selection on a *de novo* beneficial mutation). Perplexingly, they assumed identical strengths of positive selection, and identical within-patient demographic histories, between patients receiving effective and ineffective treatments. Revisiting this data, Harris *et al.* (2018) demonstrated that if one accounts for these variable fitness effects, and the inherent reality that more effective treatments reduced viral population sizes more than less effective treatments, these data observations can be explained entirely without invoking soft selective sweeps. Hence, it is always crucial to first account for the underlying neutral dynamics.

Other Key Aspects of the Population Genetic Environment

Apart from the effects of genetic drift and direct and linked selection described above, multiple other evolutionary processes are necessary to consider in order to develop a more holistic view of within-host viral population dynamics. For example, while the DFE describes the fitness effects of mutations, the rate of input of mutation is itself an important evolutionary parameter, as is the rate of recombination/reassortment, as that dictates the extent to which these newly arising mutations will be linked (and thus experience Hill-Robertson interference). Inference about mutation and recombination is similarly possible from within-patient polymorphism data. Returning to the HCMV example, Renzette *et al.* (2015) calculated genome-wide rates of mutation and recombination in 500 base pair windows across the genome for 48 longitudinal samples from 18 patients, inferring an average mutation rate of 2.0×10^{-7} new mutations per site per replication, similar to rates observed from murine cytomegalovirus (Drake and Hwang, 2005; Sanjuan *et al.*, 2010; and see review of Sackman *et al.*, 2018).

Importantly, viral mutation rates have generally been thought to be inversely correlated with genome size (Gago *et al.*, 2009), with the general interpretation being, for example, that RNA-based viral genomes are smaller than DNA genomes owing to intrinsically error-prone polymerases (Drake and Holland, 1999; Elena and Sanjuán, 2005). However, it appears more likely that, given that selection operates on the total genome-wide deleterious mutation rate (Kimura, 1967; Lynch, 2008), the selection pressure for replication fidelity (per site) is simply reduced in smaller genomes. This interpretation, as described in the 'drift-barrier hypothesis', has been widely supported across the tree of life (Lynch, 2011, 2012). In other words, genome-wide viral mutation rates are pushed to the lowest levels possible by natural selection, thus predicting the widely-observed correlation between effective population size and mutation rate, as well as between genome size and mutation rate. While some in the virology community retain the notion that viruses may be selected to maintain high mutation rates to better deal with fitness challenges, there is simply neither empirical nor theoretical support for the validity of such a view.

Finally, as much of population genetic inference is based on the Wright-Fisher model and Kingman coalescent – in which there is an assumption of a Poisson-shaped progeny distribution in which the variance equals the mean and is small relative to the population size – there is a growing appreciation that violations of this assumption must be accounted for when studying virus populations, which are highly variable in this regard. As such, virus genealogies are better characterized by more generalized Moran and multiple-merger coalescent models (see Donnelly and Kurtz, 1999; Pitman, 1999; Sagitov, 1999; Schweinsberg, 2000; Eldon and Wakeley, 2008; Matuszewski *et al.*, 2018; and see reviews of Tellier and Lemaire, 2014; Irwin *et al.*, 2016b). Specifically, violations of this progeny distribution assumption may elevate linkage disequilibrium even in the presence of frequent recombination (Eldon and Wakeley, 2008; Birkner *et al.*, 2013), and skew estimates of F_{ST} (Eldon and Wakeley, 2009). As such, these model violations may be mistaken for either population size change or positive selection if not accounted for (Matuszewski *et al.*, 2018; Sackman *et al.*, 2019).

Recent theoretical and statistical developments have provided some inroads into estimating, and thus accounting for, progeny skew from within-host pathogen populations. This work has demonstrated an ability to co-estimate the degree of progeny skew together with population size change, mutation rate, and positive selection from within-patient polymorphism data (Eldon *et al.*, 2015; Matuszewski *et al.*, 2018; Sackman *et al.*, 2019; Morales-Arce *et al.*, 2020a), providing a more appropriate null model for the study of pathogen evolution.

An Outlook on Evolutionarily Informed Treatment Strategies

The ability of the above described inference approaches to study drug-resistance mechanisms, and thus quantify the likelihood of adaptively evading a given therapeutic strategy within an experimental evolution setting, has been well-justified in the literature. Namely, identifying positively selected mutations in experimental viral populations challenged with possible drug-treatments individually or in combination, may elucidate potential routes to resistance, and highlight drugs requiring more complex (and lower probability) mutational evasion routes. For example, the commonly used influenza drug oseltamivir acts as a competitive inhibitor by binding to a hydrophobic pocket in the viral surface protein – an action which can be effectively blocked by mutations near the binding site (Collins *et al.*, 2008). Initially, the high fitness cost of the most common oseltamivir-resistant mutation, NA H275Y, was hypothesized to make the development of resistance unlikely (Ives *et al.*, 2002) – though resistance spread rapidly in the 2007–8 flu season owing to the limited number of required mutational steps (Moscona, 2009; Bloom *et al.*, 2010), an important drawback that became clearer in subsequent experimental evolution studies (*e.g.*, Foll *et al.*, 2014).

Other strategies with wider genomic effects may thus prove more fruitful – ideally ones requiring at least multiple mutation steps in order to confer resistance. Over the past decades, the notion that an excessive input of mutations can drive population extinction has been explored both theoretically and empirically, beginning with the foundational work of Lynch and Gabriel (1990) and Lynch *et al.* (1993). This possibility owes to the fact that the vast majority of fitness-impacting mutations are deleterious, or said another way, there are many more ways to disrupt rather than improve genomic function. This so-called mutational meltdown is achieved once the mean viability of the population drops to the point that the average individual cannot replace itself, at which time the population begins to decline resulting in a snowball effect leading to ultimate extinction. The cause of this final meltdown-phase is in fact the accelerated reduction in the efficacy of natural selection that results from the decline in N_e – that is, the declining population size allows for the further fixation of deleterious mutations via genetic drift, which in turn further reduces the efficacy of selection, allowing for further deleterious fixations, and so on.

Given that many viruses are naturally characterized by high mutation rates, and thus may commonly reside near a threshold of mutational load, a therapeutic increase in mutation rates may indeed induce this phase in which the input of new mutation overwhelms the ability of natural selection to remove this deleterious load. While such a scenario is often referred to as lethal mutagenesis in the virology literature (Bull *et al.*, 2007), the mutational meltdown framework is more general and critically incorporates the additional effects of genetic drift inherent to all populations (see Matuszewski *et al.*, 2017). The ability of an increased mutation rate to induce such extinction in a patient viral population has been relatively well explored and justified in RNA viruses (Loeb *et al.*, 1999; Crotty *et al.*, 2000; Lanford *et al.*, 2001; Crotty *et al.*, 2001; Severson *et al.*, 2003; Airaksinen *et al.*, 2003), using a variety of drugs including ribavirin and, more recently, favipiravir (Furuta *et al.*, 2013; Baranovich *et al.*, 2013; Bank *et al.*, 2016; Ormond *et al.*, 2017). Pertinent to the 2020 pandemic, Sheahan *et al.* (2020) recently demonstrated that the ribonucleoside analog EEID-1931 has a mutagenic effect in SARS-CoV-2 passaged in cell culture and, working in a mouse model, found a positive correlation between increased viral mutation rates and the degree of therapeutic efficacy. This work lends support to other recent calls to better investigate the potential therapeutic strategy of mutational meltdown in the context of CoV-2 (Jensen and Lynch, 2020; Santiago and Caballero, 2020; Jensen *et al.*, 2020).

While this represents a potentially promising and generalized treatment strategy for existing (and future unknown) pathogens, a number of important open questions remain in this regard. These will necessitate further theory work related to the necessary mutation rate turn-up required to induce this effect with realistic DFEs, as well as a more complete exploration of the interplay with underlying recombination rates (when applicable); as well as a more cohesive experimental effort to accurately quantify natural viral mutation rates via mutation-accumulation studies, in order to establish baseline values. What is clear however is that new viral pathogens will continue to emerge, a comprehensive understanding of their evolutionary trajectories will be critical for designing effective clinical treatments and interventions, and hence that a better union of the fields of population genetics and medicine will prove to be of great value.

See also: Antiretroviral Therapy – Nucleoside/Nucleotide and Non-Nucleoside Reverse Transcriptase Inhibitors. Antiviral Classification. HIV Integrase Inhibitors and Entry Inhibitors. Management of Adenovirus Infections (*Adenoviridae*). Management of Hepatitis A and E Virus Infection. Management of Herpes Simplex Virus Infections (*Herpesviridae*). Management of Influenza Virus Infections (*Orthomyxoviridae*). Management of Patients With Chronic Hepatitis B (*Hepadnaviridae*) and Chronic Hepatitis D Infection (*Deltavirus*). Management of Respiratory Syncytial Virus Infections (*Pneumoviridae*). Management of Varicella-Zoster Virus Infections (*Herpesviridae*). Protease Inhibitors. Treatment and Prevention of Herpesvirus Infections in the Immunocompromised Host

References

- Airaksinen, A., Pariente, N., Menendez-Arias, L., Domingo, E., 2003. Curing of foot-and-mouth disease virus from persistently infected cells by ribavirin involves enhanced mutagenesis. *Virology* 311, 339–349.
- Bank, C., Ewing, G.B., Ferrer-Admetlla, A., Foll, M., Jensen, J.D., 2014a. Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics* 30, 540–546.
- Bank, C., Hietpas, R.T., Wong, A., Bolon, D.N., Jensen, J.D., 2014b. A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: Uncovering the potential for adaptive walks in challenging environments. *Genetics* 196, 841–852.
- Bank, C., Renzette, N., Liu, P., et al., 2016. An experimental evaluation of drug-induced mutational meltdown as an antiviral treatment strategy. *Evolution* 70, 2470–2484.
- Baranovich, T., Wong, S., Armstrong, J., et al., 2013. T-705 (favipiravir) induces lethal mutagenesis in influenza A H1N1 viruses in vitro. *Journal of Virology* 87, 3741–3751.
- Birkner, M., Blath, J., Eldon, B., 2013. An ancestral recombination graph for diploid populations with skewed offspring distribution. *Genetics* 193, 255–290.
- Bloom, J.D., Gong, L., Baltimore, D., 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328, 1272–1275.
- Boeckh, M., Geballe, A.P., 2011. Cytomegalovirus: pathogen, paradigm, and puzzle. *Journal of Clinical Investigation* 121, 1673–1680.
- Britt, W.J., 2017. Congenital human cytomegalovirus infection and the enigma of maternal immunity. *Journal of Virology* 91, e02392.
- Bull, J.J., Sanjuan, R., Wilke, C.O., 2007. Theory of lethal mutagenesis for viruses. *Journal of Virology* 81, 2930–2939.
- Charlesworth, B., Morgan, M., Charlesworth, D., 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134, 1289–1303.
- Charlesworth, B., 2013. Background selection 20 years on. *Journal of Heredity* 104, 161–171.
- Collins, P.J., Haire, L., Lin, Y., et al., 2008. Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature* 453, 1258–1261.
- Crotty, S., Cameron, C., Andino, R., 2001. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proceedings of the National Academy of Sciences of the United States of America* 98, 6895–6900.
- Crotty, S., Maag, D., Arnold, J., et al., 2000. The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nature Medicine* 6, 1375–1379.
- Crow, J.F., 1993. Mutation, mean fitness, and genetic load. *Oxford Series in Evolutionary Biology* 9, 3–42.
- Cudini, J., Roy, S., Houldcroft, C., et al., 2019. Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. *Proceedings of the National Academy of Sciences of the United States of America* 116, 5693–5698.
- Donnelly, P., Kurtz, T.G., 1999. Particle representations for measure-valued population models. *Annals of Probability* 27, 166–205.
- Drake, J.W., Holland, J.J., 1999. Mutation rates among RNA viruses. *Proceedings of the National Academy of Sciences of the United States of America* 96, 13910–13913.
- Drake, J.W., Hwang, C.B.C., 2005. On the mutation rate of herpes simplex virus type 1. *Genetics* 170, 969–970.
- Drew, W.L., Sweet, E.S., Miner, R.C., Mocarski, E.S., 1984. Multiple infections by cytomegalovirus in patients with acquired immunodeficiency syndrome: documentation by Southern blot hybridization. *Journal of Infectious Diseases* 150, 952–953.
- Eldon, B., Birkner, M., Blath, J., Freund, F., 2015. Can the site-frequency spectrum distinguish exponential population growth from multiple-merger coalescents? *Genetics* 199, 841–856.
- Eldon, B., Wakeley, J., 2008. Linkage disequilibrium under skewed offspring distribution among individuals in a population. *Genetics* 178, 1517–1532.
- Eldon, B., Wakeley, J., 2009. Coalescence times and F_{st} under a skewed offspring distribution among individuals in a population. *Genetics* 181, 615–629.
- Elena, S.F., Sanjuan, R., 2005. Adaptive value of high mutation rates of RNA viruses: Separating causes from consequences. *Journal of Virology* 79, 11555–11558.
- Enders, G., Daiminger, A., Bäder, U., Exler, S., Enders, M., 2011. Intrauterine transmission and clinical outcome of 248 pregnancies with primary cytomegalovirus infection in relation to gestational age. *Journal of Clinical Virology* 52, 244–246.
- Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. *Nature Reviews Genetics* 8, 610–618.
- Faure-Della Corte, M., Samot, J., Garrigue, I., et al., 2010. Variability and recombination of clinical human cytomegalovirus strains from transplantation recipients. *Journal of Clinical Virology* 47, 161–169.
- Feder, A.F., Rhee, S.-Y., Holmes, S.P., et al., 2016. More effective drugs lead to harder selective sweeps in the evolution of drug resistance in HIV-1. *eLife* 2016, 5.
- Ferrer-Admetlla, A., Leuenberger, C., Jensen, J.D., Wegmann, D., 2016. An approximate Markov model for the Wright-Fisher diffusion and its application to time series data. *Genetics* 203, 831–846.
- Foll, M., Poh, Y.-P., Renzette, N., et al., 2014. Influenza virus drug resistance: A time-sampled population genetics perspective. *PLOS Genetics* 10, e1004185.
- Foll, M., Shim, H., Jensen, J.D., 2015. A Wright-Fisher ABC-based approach for inferring per-site effective population sizes and selection coefficients from time-sampled data. *Molecular Ecology Resources* 15, 87–98.
- Fowler, D.M., Araya, C.L., Fleishman, S.J., et al., 2010. High-resolution mapping of protein sequence-function relationships. *Nature Methods* 7, 741–746.
- Furuta, Y., Gowen, B., Takahashi, K., et al., 2013. Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral Research* 100, 446–454.
- Gago, S., Elena, S., Flores, R., Sanjuan, R., 2009. Extremely high mutation rate of a hammerhead viroid. *Science* 323, 1308.
- Haberland, M., Meyer-König, U., Hufert, F.T., 1999. Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. *Journal of General Virology* 80 (Pt 6), 1495–1500.
- Hage, E., Wilkie, G.S., Linnenweber-Held, S., et al., 2017. Characterization of human cytomegalovirus genome diversity in immunocompromised hosts by whole-genome sequencing directly from clinical specimens. *Journal of Infectious Diseases* 215, 1673–1683.
- Harris, R.B., Sackman, A., Jensen, J.D., 2018. On the unfounded enthusiasm for soft selective sweeps II: examining recent evidence from humans, flies, and viruses. *PLOS Genetics* 14, e1007859.
- Hassan, J., Connell, J., 2007. Translational mini-review series on infectious disease: Congenital cytomegalovirus infection: 50 Years on. *Clinical and Experimental Immunology* 149, 205–210.
- Hietpas, R., Jensen, J.D., Bolon, D.N.A., 2011. Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences of the United States of America* 108, 7896–7901.
- Hietpas, R., Roscoe, B., Jiang, L., Bolon, D.N.A., 2012. Fitness analyses of all possible point mutations for regions of genes in yeast. *Nature Protocols* 7, 1382–1396.
- Hill, W.G., Robertson, A., 1966. The effect of linkage on limits to artificial selection. *Genetics Research* 8, 269–294.
- Hughes, A.L., 2009. Small effective population sizes and rare nonsynonymous variants in potyviruses. *Virology* 393, 127–134.
- Irwin, K.K., Laurent, S., Matuszewski, S., et al., 2016b. On the importance of skewed offspring distributions and background selection in virus population genetics. *Heredity* 117, 393–399.
- Irwin, K.K., Renzette, N., Kowalik, T., Jensen, J.D., 2016a. Antiviral drug resistance as an adaptive process. *Virus Evolution* 2, 1–10.
- Ives, J.A., Carr, J., Mendel, D., et al., 2002. The H274Y mutation in the influenza A/H1N1 neuraminidase active site following oseltamivir phosphate treatment leave virus severely compromised both in vitro and in vivo. *Antiviral Research* 55, 307–317.
- Jensen, J.D., Kowalik, T.F., 2020. A consideration of within-host human cytomegalovirus genetic variation. *Proceedings of the National Academy of Sciences of the United States of America* 117, 816–817.
- Jensen, J.D., Lynch, M., 2020. Considering mutational meltdown as a potential SARS-CoV-2 treatment strategy. *Heredity* 124, 619–620.
- Jensen, J.D., Payseur, B., Stephan, W., et al., 2019. The importance of the Neutral Theory in 1968 and 50 years on: a response to Kern & Hahn 2018. *Evolution* 73, 111–114.
- Jensen, J.D., Stikeleather, R., Kowalik, T., Lynch, M., 2020. Imposed mutational meltdown as an antiviral strategy. *Evolution* 74, 2549–2559. doi:10.1111/evo.14107.
- Johri, P., Charlesworth, B., Jensen, J.D., 2020. Toward an evolutionarily appropriate null model: Jointly inferring demography and purifying selection. *Genetics* 215, 173–192.
- Keightley, P.D., Eyre-Walker, A., 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177, 2251–2261.

- Kenneson, A., Cannon, M.J., 2007. Review and meta-analysis of the epidemiology of congenital cytomegalovirus (CMV) infection. *Reviews in Medical Virology* 17, 253–276.
- Kimura, M., 1967. On the evolutionary adjustment of spontaneous mutation rates. *Genetics Research* 9, 23–34.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624–626.
- Lanford, R.E., Chavez, D., Guerra, B., *et al.*, 2001. Ribavirin induces error-prone replication of GB virus B in primary tamarin hepatocytes. *J. Virol.* 75, 8074–8081.
- Loeb, L.A., Essigmann, J., Kazazi, F., *et al.*, 1999. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proceedings of the National Academy of Sciences of the United States of America* 96, 1492–1497.
- Long, H., Sung, W., Kucukyildirim, S., *et al.*, 2018. Evolutionary determinants of genome-wide nucleotide composition. *Nature Ecology and Evolution* 2, 237–240.
- Lynch, M., Ackerman, M.S., Gout, J.-F., *et al.*, 2016. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics* 17, 704–714.
- Lynch, M., Blanchard, J., Houle, D., *et al.*, 1999. Spontaneous deleterious mutation. *Evolution* 53, 645–663.
- Lynch, M., Burger, R., Butcher, D., Gabriel, W., 1993. The mutational meltdown in asexual populations. *Journal of Heredity* 84, 339–344.
- Lynch, M., Gabriel, W., 1990. Mutation load and the survival of small populations. *Evolution* 44, 1725–1737.
- Lynch, M., 2008. The cellular, developmental, and population-genetic determinants of mutation-rate evolution. *Genetics* 180, 933–943.
- Lynch, M., 2011. The lower bound to the evolution of mutation rates. *Genome Biology and Evolution* 3, 1107–1118.
- Lynch, M., 2012. Evolutionary layering and the limits to cellular perfection. *Proceedings of the National Academy of Sciences of the United States of America* 109, 18851–18856.
- Malaspina, A.-S., Malaspina, O., Evans, S.N., Slatkin, M., 2012. Estimating allele age and selection coefficients from time-series data. *Genetics* 192, 599–607.
- Manicklal, S., Emery, V.C., Lazzarotto, T., Boppana, S.B., Gupta, R.K., 2013. The “Silent” global burden of congenital cytomegalovirus. *Clinical Microbiology Reviews* 26, 86–102.
- Mathieson, I., McVean, G., 2013. Estimating selection coefficients in spatially structured populations from time series data of allele frequencies. *Genetics* 193, 973–984.
- Matuszewski, S., Hildebrandt, M.E., Achaz, G., Jensen, J.D., 2018. Coalescent processes with skewed offspring distributions and non-equilibrium demography. *Genetics* 208, 323–338.
- Matuszewski, S., Ormond, L., Bank, C., Jensen, J.D., 2017. Two sides of the same coin: a population genetics perspective on lethal mutagenesis and mutational meltdown. *Virus Evolution* 3, 1–5.
- Maynard Smith, J., Haigh, J., 1974. The hitch-hiking effect of a favourable gene. *Genetics Research* 23, 23.
- Meyer-König, U., Ebert, K., Schrage, B., Pollak, S., Hufert, F., 1998. Simultaneous infection of healthy people with multiple human cytomegalovirus strain. *Lancet* 352, 1280.
- Miyashita, S., Kishino, H., 2010. Estimation of the size of genetic bottlenecks in cell-to-cell movement of soil-borne wheat mosaic virus and the possible role of the bottlenecks in speeding up selection of variations in trans-acting genes or elements. *Journal of Virology* 84, 1828–1837.
- Morales-Arce, A.Y., Harris, R.B., Stone, A.C., Jensen, J.D., 2020a. Evaluating the contributions of purifying selection and progeny-skew in dictating within-host *Mycobacterium tuberculosis* evolution. *Evolution* 74, 992–1001.
- Morales-Arce, A.Y., Sabin, S.J., Stone, A.C., Jensen, J.D., 2020b. The population genomics of within-host *Mycobacterium tuberculosis*. *Heredity* 126, 1–9. doi:10.1038/s41437-020-00377-7.
- Moscona, A., 2009. Global transmission of oseltamivir-resistant influenza. *New England Journal of Medicine* 360, 953–956.
- Nishiyama, Y., Maeno, K., Yoshida, S., 1983. Characterization of human cytomegalovirus-induced DNA polymerase and the associated 3′-to-5′, exonuclease. *Virology* 124, 221–231.
- Numazaki, K., 1997. Human cytomegalovirus infection of breast milk. *FEMS Immunology and Medical Microbiology* 18, 91–98.
- Ohta, T., 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246, 96–98.
- Ormond, L., Liu, P., Matuszewski, S., *et al.*, 2017. The combined effect of oseltamivir and favipiravir on influenza A virus evolution. *Genome Biology and Evolution* 9, 1913–1924.
- Pénisson, S., Singh, T., Sniogowski, P., Gerrish, P., 2017. Dynamics and fate of beneficial mutations under lineage contamination by linked deleterious mutations. *Genetics* 205, 1305–1318.
- Pitman, J., 1999. Coalescents with multiple collisions. *Journal of Applied Probability* 27, 1870–1902.
- Pokalyuk, C., Renzette, N., Irwin, K.K., *et al.*, 2017. Characterizing human cytomegalovirus reinfection in congenitally infected infants: an evolutionary perspective. *Molecular Ecology* 26, 1980–1990.
- Renzette, N., Bhattacharjee, B., Jensen, J.D., Gibson, L., Kowalik, T.F., 2011. Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants. *PLOS Pathogens* 7, e1001344.
- Renzette, N., Gibson, L., Bhattacharjee, B., *et al.*, 2013. Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. *PLOS Genetics* 9, e1003735.
- Renzette, N., Gibson, L., Jensen, J.D., Kowalik, T.F., 2014. Human cytomegalovirus intrahost evolution – A new avenue for understanding and controlling herpesvirus infections. *Current Opinion in Virology* 8, 109–115.
- Renzette, N., Kowalik, T.F., Jensen, J.D., 2016. On the relative roles of background selection and genetic hitchhiking in shaping human cytomegalovirus genetic diversity. *Molecular Ecology* 25, 403–413.
- Renzette, N., Pfeifer, S.P., Matuszewski, S., Kowalik, T.F., Jensen, J.D., 2017. On the analysis of intrahost and interhost viral populations: Human cytomegalovirus as a case study of pitfalls and expectations. *Journal of Virology* 91, e01976–16.
- Renzette, N., Pokalyuk, C., Gibson, L., *et al.*, 2015. Limits and patterns of cytomegalovirus genomic diversity in humans. *Proceedings of the National Academy of Sciences of the United States of America* 112, E4120–E4128.
- Sackman, A., Harris, R.B., Jensen, J.D., 2019. Inferring demography and selection in organisms characterized by skewed offspring distributions. *Genetics* 211, 1019–1028.
- Sackman, A.M., Pfeifer, S.P., Kowalik, T.F., Jensen, J.D., 2018. On the demographic and selective forces shaping patterns of human cytomegalovirus variation within hosts. *Pathogens* 7, 16.
- Sagitov, S., 1999. The general coalescent with asynchronous mergers of ancestral lines. *Journal of Applied Probability* 36, 1116–1125.
- Sanjuan, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belshaw, R., 2010. Viral mutation rates. *Journal of Virology* 84, 9733–9748.
- Santiago, E., Caballero, A., 2020. The value of targeting recombination as a strategy against coronavirus diseases. *Heredity* 125, 169–172. doi:10.1038/s41437-020-0337-5.
- Schneider, A., Charlesworth, B., Eyre-Walker, A., Keightley, P.D., 2011. A method for inferring the rate of occurrence and fitness effects of advantageous mutations. *Genetics* 189, 1427–1437.
- Schweinsberg, J., 2000. Coalescents with simultaneous multiple collisions. *Electronic Journal of Probability* 5, 1–50.
- Severson, W.E., Schmaljohn, C., Javadian, A., Jonsson, C., 2003. Ribavirin causes error catastrophe during Hantaan virus replication. *Journal of Virology* 77, 481–488.
- Sheahan, T., Sims, A., Zhou, S., *et al.*, 2020. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Science Translational Medicine* 12, eabb5883.
- Spector, S.A., Hirata, K.K., Neuman, T.R., 1984. Identification of multiple cytomegalovirus strains in homosexual men with acquired immunodeficiency syndrome. *Journal of Infectious Diseases* 150, 953–956.
- Tataru, P., Mollion, M., Glémin, S., Bataillon, T., 2017. Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. *Genetics* 207, 1103–1119.
- Tellier, A., Lemaire, C., 2014. Coalescence 2.0: A multiple branching of recent theoretical developments and their applications. *Molecular Ecology* 23, 2637–2652.
- Tseng, T.C., Kao, J.H., 2017. Elimination of hepatitis B: Is it a mission possible? *BMC Medicine* 15, 1–5.
- Voronin, Y., Jani, I., Graham, B.S., *et al.*, 2017. Recent progress in immune-based interventions to prevent HIV-1 transmission to children. *Journal of the International AIDS Society* 20, e25038.

Further Reading

- Charlesworth, B., Charlesworth, D., 2010. *Elements of Evolutionary Genetics*. USA: Roberts and Company Publishers.
- Kimura, M., 1983. *The Neutral Theory of Molecular Evolution*. UK: Cambridge University Press.
- Provine, W.B., 2001. *The Origins of Theoretical Population Genetics*. USA: University of Chicago Press.
- Wakeley, J., 2009. *Coalescent Theory: An Introduction*. USA: Roberts and Company Publishers.
- Walsh, B., Lynch, M., 2018. *Evolution and Selection of Quantitative Traits*. UK: Oxford University Press.

Virus-Based Cancer Therapeutics

Roberto Cattaneo, Mayo Clinic, Rochester, MN, United States

Christine E Engeland, University Hospital Heidelberg and German Cancer Research Center, Heidelberg, Germany and Witten/Herdecke University, Witten, Germany

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

GM-CSF Granulocyte-macrophage colony-stimulating factor;
IFN Interferon;
MeV Measles virus;

NDV Newcastle disease virus;
RBP Receptor-binding protein;
SPECT Single photon emission computed tomography
VSV Vesicular stomatitis virus;

Glossary

Arming Insertion of additional genes into the oncolytic virus genome to increase therapeutic efficacy.

Bystander effects Cytotoxic effects on non-infected cells in the vicinity of infected cells.

Immune checkpoint blockade Treatment with antibodies that block inhibitory signaling in immune cells, especially T cells.

Shielding Genetic or chemical modifications to avoid antibody neutralization of viruses.

Targeting Modifications of viral tropism to increase tumor specificity.

Definition

Oncolytic virotherapy is the treatment of cancer with viruses that replicate preferentially in tumor tissue, damaging it. Selectivity for tumor tissue can occur naturally, and be enhanced by genetic engineering.

Introduction and Historic Notes

Virus infection may be responsible for a significant fraction of human cancer deaths. Nevertheless, shortly after the discovery of animal viruses, observing physicians reported cancer regressions coincident with natural infections, most notably in patients with lymphoma or leukemia. These patients were suffering from viral hepatitis, glandular fever, chickenpox, or measles. For example, before a virus was discovered as the causal agent, influenza was associated with leukemia remission. In a 1910 report, the Italian gynecologist De Pace observed tumor lysis and subsequent tumor remission in a cervical cancer patient after rabies vaccination. These and other anecdotal observations inspired early clinical trials in the second half of the 20th century.

These trials did not meet current ethical standards for clinical investigations. Infectious body fluids or tissue samples were administered to cancer patients. For instance, in a 1949 trial, 22 patients suffering from Hodgkin's lymphoma were treated with sera or tissue extracts from patients infected with hepatitis B virus. The first patient died after treatment, and 13 developed hepatitis. Seven patients were reported to show amelioration and four experienced reduction of lymphoma burden. In 1953, a case series was published reporting transient remission of acute leukemia after intramuscular injection of glandular fever sera in three out of five treated patients. Trials with several human pathogenic viruses were conducted, including flaviviruses such as dengue and West Nile virus. In the latter, viremia and intra-tumoral virus were observed in many patients, with several cases of encephalitis and few cases of transient delay in tumor progression. These approaches to therapy can only be understood in the light of the general hopelessness of what was then cancer treatment.

The advent of tissue culture in the 1950s and 1960s allowed viruses to be propagated in a more defined environment. In addition, it became possible to assess the extent of virus spread and virus-induced cell death in cancer cell lines. Based on *in vitro* experiments, an adenovirus, at that time called adenoidal-pharyngeal-conjunctival virus, emerged as a candidate oncolytic virus. In 1956, this virus was administered to patients with cervical carcinoma by either intratumoral, intra-arterial or intravenous administration. This treatment showed a favorable safety profile. Extensive tumor necrosis occurred in 26 of 40 treated patients, but all patients ultimately succumbed to their tumor within a few months.

Another important development in the assessment of the safety and oncolytic efficacy of viruses was the use of models based on transplantable or spontaneous murine tumors. These models were used to characterize the efficacy of animal viruses including two herpes viruses, equine rhinopneumonitis and infectious bovine rhinotracheitis virus, as well as a murine negative sense RNA virus termed "M-P virus" at the time, which was later identified as a lymphocytic choriomeningitis virus, and the poultry pathogen Newcastle disease virus (NDV), a member of the *Paramyxoviridae* family.

Simultaneously with the testing of several animal viruses as oncolytics, cancer regressions coincident with natural human infections continued to be reported. In the 1970s several case reports described coincidence of infections with measles virus (MeV),



Fig. 1 Lymphoma remission after measles infection. *Left panel:* patient on admission with Burkitt's lymphoma; note the swelling around his right eye. *Center panel:* patient three weeks later after contracting measles by natural infection; note the rash on his skin. *Right panel:* Patient five weeks later. Tumor and measles exanthem have resolved completely. Reprinted with permission from Bluming, A.Z., Ziegler, J.L., 1971. Regression of Burkitt's lymphoma in association with measles infection. *Lancet* 2 (7715), 105–106.

another member of the *Paramyxoviridae* family, with regressions or remissions of leukemia and lymphoma. One of these reports included a striking visual documentation of Burkitt's lymphoma remission following natural measles ([Fig. 1](#)).

In 1974, a large clinical trial with a third member of the *Paramyxoviridae* family naturally infecting humans, mumps virus, was reported. Ninety patients with various terminal cancers were treated by different routes of administration including intravenous or intratumoral injection, local application after scarification of the tumor, oral or rectal administration and inhalation. Side effects were minimal, and 79 of 90 patients experienced tumor remission, including several apparently complete responses. In some patients, signs of increased anti-tumor immunity were observed. However, there was no follow up. In summary, by the end of the 1970s, several clinical trials based on human or non-human viruses reported acceptable toxicities, but limited efficacy.

This situation contrasted with the availability of effective chemotherapy and radiation therapy clinical interventions, and resulted in the decline of research activities in the oncolytic virus field. At this crossroads, it was evident that viruses with increased cancer specificity were needed. However, to achieve this goal, three areas of research needed development: methods for directly documenting viral spread *in vivo*, a deep understanding of the virus tropism determinants, and technology to generate recombinant viruses with specified characteristics.

Re-engineered Viruses: The Beginnings

The “natural” oncolytic properties of some viruses were initially inferred by the coincidence of cancer regressions with natural infections. Subsequently, when tissue culture systems for animal cells were established, it became evident that almost every virus grows much better in transformed cancer cell lines than in primary cells. We now know that the multistep process of cancer progression can affect multiple components of the cellular response to viral infection, favoring the replication of many different viruses. Thus, many viruses have “natural” oncolytic properties.

On the other hand, oncolytic properties are not universal, but depend on proper matching of virus and host cell. Virus families have evolved specificities for different cell types, and this natural diversity impacts on the oncolytic properties of each virus. To be a cancer therapeutic, viruses need to efficiently enter cells. If this happens, the host cell is per definition susceptible ([Fig. 2](#), center). Following cell entry, depending on the outcome of several interactions between cellular intrinsic immunity proteins and viral proteins, the virus will replicate. In this case, the host cell is per definition permissive to infection and will ultimately be lysed ([Fig. 2](#), right).

In the last three decades, research on the tropism determinants for viruses of several different families greatly progressed. This was facilitated by the establishment of methods to easily visualize and quantify virus spread in the host using reporter genes, and to document how virus replication in different locations relates to therapeutic efficacy. In addition, our understanding of tumor biology greatly improved; more sophisticated animal models were included in pre-clinical testing of new therapeutics; and large arrays of diagnostic markers are being used for in-depth evaluation of clinical trial results.

Based on this deeper knowledge of the biology of both viruses and cancer, new viruses with improved oncolytic activities have been generated, and many of these recombinant viruses are now in pre-clinical development. In particular, four DNA viruses and five RNA viruses are either approved therapeutics or in different clinical trial phases. [Table 1](#) lists one or two selected examples of approved or experimental cancer therapeutics derived from each of these nine families. It indicates the virus family and name, the genetic changes introduced, the routes of administration, whether the virus was given alone or in combination with another cancer therapeutic, and the clinical trial identifier and time frame.

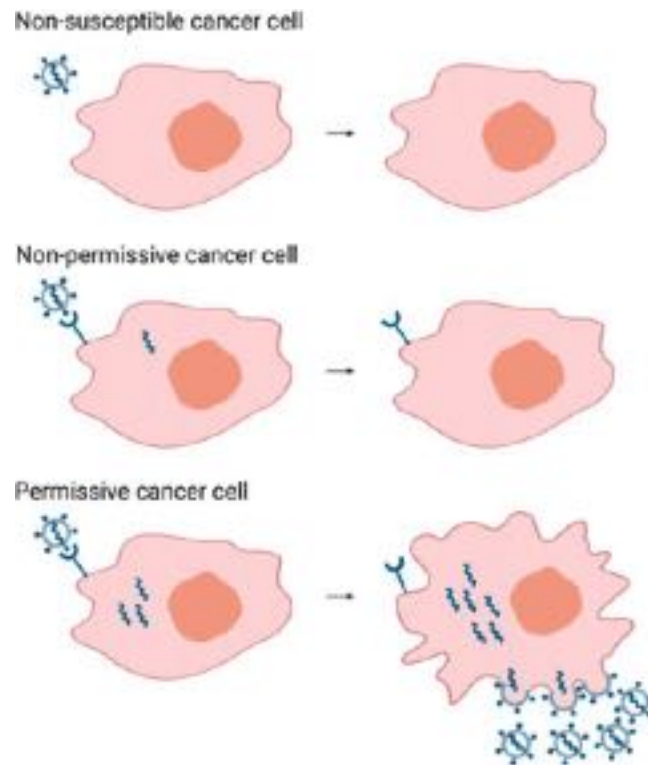


Fig. 2 Susceptibility and permissiveness of cancer cells determine the outcome of oncolysis. *Top panels:* cancer cell that does not express a viral receptor and is therefore not susceptible to the virus. *Center panels:* cancer cell that expresses a viral receptor, allowing viral entry. However, the cell is not permissive for the virus, which does not replicate and cannot cause oncolysis. *Bottom panels:* cancer cell both susceptible to, and permissive for the virus. The virus replicates, ultimately leading to oncolysis. (Created with BioRender.com.)

While this table is only a snapshot of current state-of-the-art research, some general conclusions can be drawn. First, virus-based cancer therapeutics are well tolerated, even at the highest doses achievable by current manufacturing processes, after both localized and systemic administration. In addition, although virus shedding was documented in the urinary and respiratory tract, especially after systemic administration, no transmission from a cancer patient to health care providers, or other contacts, has been noted. Second, three of the most successful oncolytic viruses express as immunomodulatory transgene the granulocyte-macrophage colony-stimulating factor (GM-CSF) (Table 1, DNA viruses). This may be due to the combination of replicative oncolysis with GM-CSF-mediated inflammation to induce adaptive immunity against tumor antigens. The third conclusion is cautionary: several viruses have fallen short of the efficacy expectations that were set by pre-clinical models. This fact underscores the importance of improving efficacy of current generation virus-based cancer therapeutics.

Three Classes of Modifications

Current clinical trials of oncolysis face several challenges. Insufficient efficacy can be addressed by more aggressive dosing, which is being enabled by continuing technological advances in virus production. However, more aggressive dosing requires improved tumor specificity of future vectors to maintain the current safety profiles. Cancer specificity can be improved by *targeting*. Targeting, which is based on the introduction of different layers of cancer specificity, either at the cell entry or post-entry levels, can enhance both efficacy and safety of oncolytic vectors.

Another key challenge for oncolytic therapy is to avoid immediate neutralization of the vectors through circulating antibodies, which can arise in patient populations via natural contagion, scheduled immunization, or prior administration of an oncolytic virus. Temporary *shielding* from the host immune response can provide a window of opportunity for the effective initiation of the oncolytic process. Shielding can be achieved by changing or physically masking the epitopes that are recognized by neutralizing antibodies.

The third key challenge for oncolytic therapy is the need to reach and eliminate not only all the cells in a primary tumor, but also tumor metastases. This challenge is addressed by *arming* viruses with proteins that sensitize not only infected tumor cells but also uninfected metastases, to subsequent combination therapies or immune destruction. Arming occurs through the expression of a foreign gene: for example, a prodrug convertase activating an anticancer therapeutic *in situ*, or an ion channel facilitating radiosensitization, or a cytokine promoting immunostimulation.

Table 1 Selected examples of current clinical trials with viruses from nine families^a

<i>Virus family</i>	<i>Phase and type of clinical trial</i>	<i>Virus name</i>	<i>Genetic changes</i>	<i>Route</i>	<i>Combination therapy</i>	<i>NCT number/PMID</i>	<i>Start-end date</i>
<i>DNA viruses</i>							
Adenoviridae	Phase III terminated; bladder cancer	CG0070	GM-CSF	Intravesicular	None	NCT01438112	3/2014-6/2016
	Approved (China, 2005); HNSCC ^b	Oncorine (H101)	E1B-55k deleted	Intratumoral	Chemotherapy (Cisplatin plus 5-FU or adriamycin)	PMID 15601557 29189159	
Herpesviridae	Approved (United States, Europe, 2015); Melanoma	Talimogene Laherparepvec (T-VEC)	GM-CSF ICP34.5 deleted, US11 altered, ICP47 deleted	Intratumoral	None	NCT00769704 NCT00289016 PMID: 17121894	Phase III: 4/2009-9/2014 Phase II 12/2005-5/2009 Phase I
	Phase III; melanoma	T-VEC	As above	Intratumoral	Pembrolizumab	NCT02263508	8/2014
Parvoviridae	Phase II completed; glioblastoma and PDAC ^c	ParvOryx	None	Intravenous and intratumoral	None	NCT01301430 NCT02653313	9/2011-5/2015 12/2015-5/2018
Poxviridae	Phase III; hepatocellular carcinoma	Pexastimogene devacirepvec (Pexa-Vec)	GM-CSF, thymidine kinase deleted	Intratumoral	Sorafenib	NCT02562755	10/2015
<i>RNA viruses</i>							
Paramyxoviridae	Phase II completed; multiple myeloma	MV-NIS	human NIS	Intravenous	Cyclo-phosphamide	NCT02192775	3/2015-8/2019
Picornaviridae	Phase II; glioblastoma	PVS-RIPO	HRV2 IRES ^d	Intratumoral	None	NCT02986178	6/2017
Reoviridae	Phase III completed; HNSCC ^b	Reolysin	None	Intravenous	Carboplatin Paclitaxel	NCT01166542	6/2010-5/2014
Retroviridae	Phase II/III; High-grade glioma	Vocimagene amiretrorepvec (Toca511)	Cytosine deaminase	Into resection cavity	5-fluorocytosine NCT04105374: Temozolomide and radiation	NCT02414165 NCT04105374	11/2015-12/2019 1/2020
Rhabdoviridae	Phase I; HCC ^e and others	VSV-IFN β	human IFN β	Intratumoral	None	NCT01628640	8/2012
	Phase I; NSCLC, HPV-associated a.o.	MG1 Maraba	Tumor antigen	Intravenous	Adenovirus vaccine NCT02879760: Pembrolizumab NCT03618953: Atezolizumab	NCT02285816 NCT02879760 NCT03618953	10/2014 3/2017 6/2018

^aInformation as of January 2020; for updates see ClinicalTrials.gov.^bHNSCC, head and neck squamous cell carcinoma.^cPDAC, pancreatic ductal adenocarcinoma.^dHRV2 IRES, human rhinovirus type 2 internal ribosomal entry site.^eHCC, hepatocellular carcinoma.

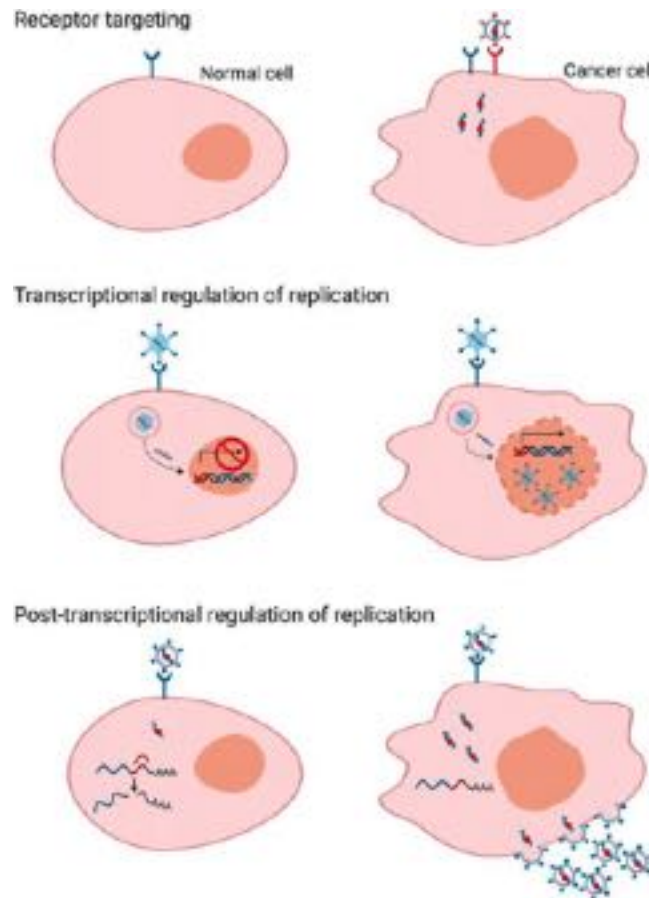


Fig. 3 Entry and post-entry targeting of oncolytic viruses. Left column: normal cells. Right column: cancer cells. Top panels: receptor targeting. The viral RBP is engineered to no longer recognize its natural receptor, while attaching to a tumor-specific surface protein. This approach has been pioneered with enveloped viruses. Center panels: transcriptional regulation of replication. Virus replication is placed under control of a promoter active in tumor cells, but not in normal cells. This approach has been pioneered with viruses whose replication depends on cellular polymerase II. Bottom panels: post-transcriptional regulation of replication. This process takes advantage of the downregulation of certain microRNAs in tumor cells. Viruses engineered with microRNA target sequences in their genome replicate only in absence of the cognate microRNA. (Created with BioRender.com.)

While not all genetic modifications of viruses fall precisely within one of these classes, the targeting, shielding and arming principles guide the continuous improvement of oncolytic viruses.

Retargeting Virus Tropism

The term targeting describes virus modifications that confer greater specificity for tumor cells by improving infection of diseased tissues or decreasing infection of healthy tissues, or both. Tumor specificity can be enhanced either at the stage of virus entry into the cells by modifying receptor choice (Fig. 3, top panels), or post entry during replication. Post-entry targeting can be achieved either at the transcription initiation level through the use of cancer-specific promoters (Fig. 3, center panels), or post-transcriptionally by inserting microRNA (miRNA) target sequences into viral genomes that restrict virus replication in healthy tissues (Fig. 3, bottom panels). For all these targeting strategies, a deeper understanding of tumor biology has facilitated the development of viruses that exploit tumor-specific alterations such as altered expression of cell surface proteins, transcription factors, and microRNAs.

Cell Entry Targeting of Enveloped Viruses

All viruses require interactions with surface molecules on host cells to start infection. Surface proteins preferentially expressed on tumor cells can be exploited for entry targeting. This targeting strategy requires modification of the receptor-binding proteins of viruses. These modifications are either genetic and generate chimeric proteins, or they use chemical adapters to link specificity domains to virus particles.

Genetic modifications of the receptor-binding proteins (RBP) of viruses are desirable because they are retained after replication. To redirect viral tropism by receptor targeting, their RBP are modified in two ways: first, ablation of binding to the natural receptor, or receptors. Second, engineering of binding to receptors preferentially or exclusively expressed on tumor cells.

Fig. 3 (top panels) illustrates receptor targeting by RBP genetic engineering. Measles virus (MeV), an enveloped negative strand RNA virus of the *Paramyxoviridae* family, is used as an example. Receptor targeting is most advanced for enveloped viruses owing to the plasticity of their surface glycoproteins, whose evolution is not restricted by the multiple constraints imposed by icosahedral symmetry on the capsids of other viruses.

The *Paramyxovirus* membrane fusion apparatus is easy to manipulate because receptor binding and membrane fusion functions are mediated by two different proteins. For MeV, the residues required for entry via the natural receptors have been identified and then mutated to disable binding to the natural receptors. Retargeting was completed by adding to the entire ectodomain of the RBP a moiety recognizing a cancer-specific receptor of interest. A flexible linker was introduced between the RBP ectodomain and the targeting moiety to allow its movement.

Initially small peptide ligands for growth factor receptors were used to prove that the MeV RBP can bind targeted receptors without losing its membrane fusion support function. It was then shown that single-chain antibodies can also be used as specificity domains. Single-chain antibodies comprise only the variable domains of both light and heavy chains, which confer antigen specificity, and are far smaller than complete antibodies.

Single-chain antibody fragments are well suited for retargeting of enveloped viruses, since they share with viral glycoproteins the same intracellular synthesis and transport pathways. Receptor targeting using single-chain antibody fragments has also been developed for the *Herpesviridae*, although the membrane fusion apparatus of these large DNA viruses is more complex and involves five glycoproteins. On the other hand, the single protein membrane fusion system of the *Retroviridae* family has proven difficult to retarget without extensive loss of cell entry efficiency.

Cell Entry Targeting of Non-Enveloped Viruses

Receptor targeting of non-enveloped viruses with icosahedral capsids is challenging due to the strict structural constraints imposed by icosahedral symmetry. Nevertheless, receptor targeting by genetic modification has been achieved for *Adenoviridae*. For instance, short peptides or ligands with specificity for tumor surface proteins have been inserted at one end, or into a flexible region, of the adenovirus RBP. Targeting is completed by mutations that ablate binding to the natural receptors.

More invasive capsid modifications have had different degrees of success. In addition to the icosahedral symmetry constraints, targeting strategies must take in account that adenovirus capsids are assembled in the cytosol, rather than transiting through the endoplasmic reticulum. Therefore, ligands that require folding, oxidation or other post-translational modifications in the endoplasmic reticulum are unsuitable.

As alternative entry targeting strategy, adapter molecules have been developed. These adapters bind with one end to the adenovirus capsid, and with the other end to a cancer cell protein. This approach poses less structural constraints than the addition of specificity domains. However, cell entry occurs through the targeted receptor only during the first round of replication.

Cell Entry Targeting by Protease Activation

Another cell entry targeting strategy is based on the selective activation of virus particles in the tumor microenvironment. This strategy, which takes advantage of the requirement of proteolytic cleavage for the activation of fusion proteins, has been used to retarget cell entry of *Paramyxoviridae* and *Retroviridae*.

Protease targeting requires two modifications: first, inactivation of the cleavage site used by the cognate protease expressed in the natural host cells. Second, introduction of a cleavage site recognized by proteases secreted in the tumor microenvironment, such as matrix metalloproteinases. Since protease cleavage requires not only a specific recognition sequence, but also its accessibility by the cognate protease, engineering of efficient proteolytical activation may not be straightforward and require iterative cycles of optimization through mutagenesis. For example, the protease cleavage sites of the MeV and Sendai virus fusion proteins have been re-targeted following this principle.

Transcriptional Regulation of Replication

The most potent oncolytic viruses are arguably wild-type viruses. However, these viruses can kill normal cells, hence causing dose-limiting toxicities. Thus, oncolytic therapy has been developed almost exclusively based on attenuated viruses. On the other hand, attenuation is always relative to the target cell. Ideally, oncolytic viruses should be completely attenuated in normal cells, but maintain normal replication in cancer cells.

Towards this, two replicational targeting strategies have been developed. The first aims at modulating the transcription of a gene controlling the early phases of replication. The second focuses on the selective transcription of virulence factors expressed late during the viral cycle. These strategies have been developed for viruses whose replication depends on cellular polymerases, in particular the *Adenoviridae* and the *Herpesviridae*.

Early transcriptional control of replication is best developed for the *Adenoviridae* (**Fig. 3**, center panels). It can be achieved through the substitution of the promoter modulating transcription of Adenovirus early proteins by commonly used cancer-associated promoters, such as telomerase, or a prostate-specific antigen promoter in case of prostate cancer. Alternatively, there is now a burgeoning number of tumor-associated promoters that were identified using gene expression profiling in different cancer

types. Certain elements of these promoters are now being used to control viral replication at the level of early transcription in new vectors proceeding through clinical trials.

Another strategy aims at generating recombinant viruses that retain full virulence only in cancer cells by exploiting tumor-associated promoters to selectively activate the expression of virulence factors. This strategy is most advanced for the *Herpesviridae*, large DNA viruses that code for several virulence factors whose expression can be altered in combination. For example, a glioma-specific promoter is used to control the expression of the neurovirulence gene ICP34.5 of a herpes virus 1 recombinant used for glioma therapy.

Post-Transcriptional Regulation of Replication

An alternative strategy for targeting viral replication exploits differential expression of microRNAs between tumors and normal tissues. This approach requires identification of microRNAs that are expressed in non-target tissues, but downregulated in tumor cells. Target sites for microRNAs are inserted into the untranslated regions of viral transcripts. These viral transcripts are degraded by the cellular microRNA machinery in cells that express the cognate microRNAs, abrogating viral replication (Fig. 3, bottom panels).

The post-transcriptional regulation of replication principle was initially developed with a picornavirus that was an effective oncolytic but also caused significant toxicities due to myositis. Target sites for muscle-specific microRNAs introduced into the genome of this virus prevented myositis, while therapeutic efficacy was maintained.

With extensive knowledge about microRNA expression in tumors, and the relative ease with which these relatively short target sequences can be incorporated into virus genomes without negatively affecting replication, negative targeting via microRNAs has been extended to several oncolytic virus platforms. These include members of the *Adenoviridae*, *Herpesviridae*, *Paramyxoviridae*, *Poxviridae*, and *Rhabdoviridae*. Towards this, multiple perfect match microRNA sequences have been expressed in the untranslated regions of essential viral mRNAs.

Preferential Tumor Spread

Cancer targeting can also be achieved by favoring preferential spread of viruses in tumor tissue through the exploitation of specific defects in innate immunity. For example, many cancer cells cannot produce interferon (IFN), or cannot respond to IFN stimulation. In these cells, even attenuated viruses with defective IFN control proteins, and thus unable to replicate in normal cells, replicate and spread.

This principle has been exploited for the precise targeting of certain viruses for which the mechanisms of interferon inactivation are understood in depth, like the *Paramyxoviridae*. For example, analyses of the interactions between cellular innate immunity proteins and the MeV proteins controlling their function, have guided the generation of recombinant viruses that retain oncolytic efficacy in tumor tissue, while not spreading in normal cells.

Combination of Targeting Modalities

Multiple targeting modifications have been combined in second-generation recombinant oncolytic viruses. Therefore, most viruses that are currently in pre-clinical trials, or in early clinical trials, are targeted according to more than one modality. Targeting combinations allow to sequentially increase tumor specificity, augmenting safety. This is important, especially since armed viruses are being developed to augment the efficacy of virotherapy approaches.

Shielding From the Host Immune Response

Like any pathogen, oncolytic viruses are recognized as “non-self” by the host immune system, resulting in an anti-viral immune response. Innate, cellular and humoral anti-viral immunity can hamper efficacy of virotherapy.

While innate and cell-mediated responses can limit viral spread, neutralizing antibodies can inactivate the virus before it reaches the tumor. This is a major impediment to intravenous application of oncolytic viruses which, from a clinical perspective, is the preferred route of administration due to its feasibility and the potential to reach disseminated tumor sites. Therefore, avoiding premature viral clearance by neutralizing antibodies has been one focus in the development of advanced oncolytic therapies. Plasmapheresis has been applied to deplete neutralizing antibodies, but this process is complex and costly.

Virus Engineering

Virus engineering strategies have been developed to avoid antibody neutralization. These strategies include generating chimeric viruses that are not recognized by neutralizing antibodies. These chimeric viruses, which can be used for sequential rounds of therapy, can be generated by serotype exchange.

Serotype exchange is an option for most viruses, such as vesicular stomatitis virus (VSV) and adenovirus. In this approach, the viral proteins that are recognized by neutralizing antibodies, or parts of these proteins, are exchanged. For example, since all neutralizing epitopes of VSV reside within its glycoprotein, it is sufficient to replace it with that of another serotype (Fig. 4, top left panel).

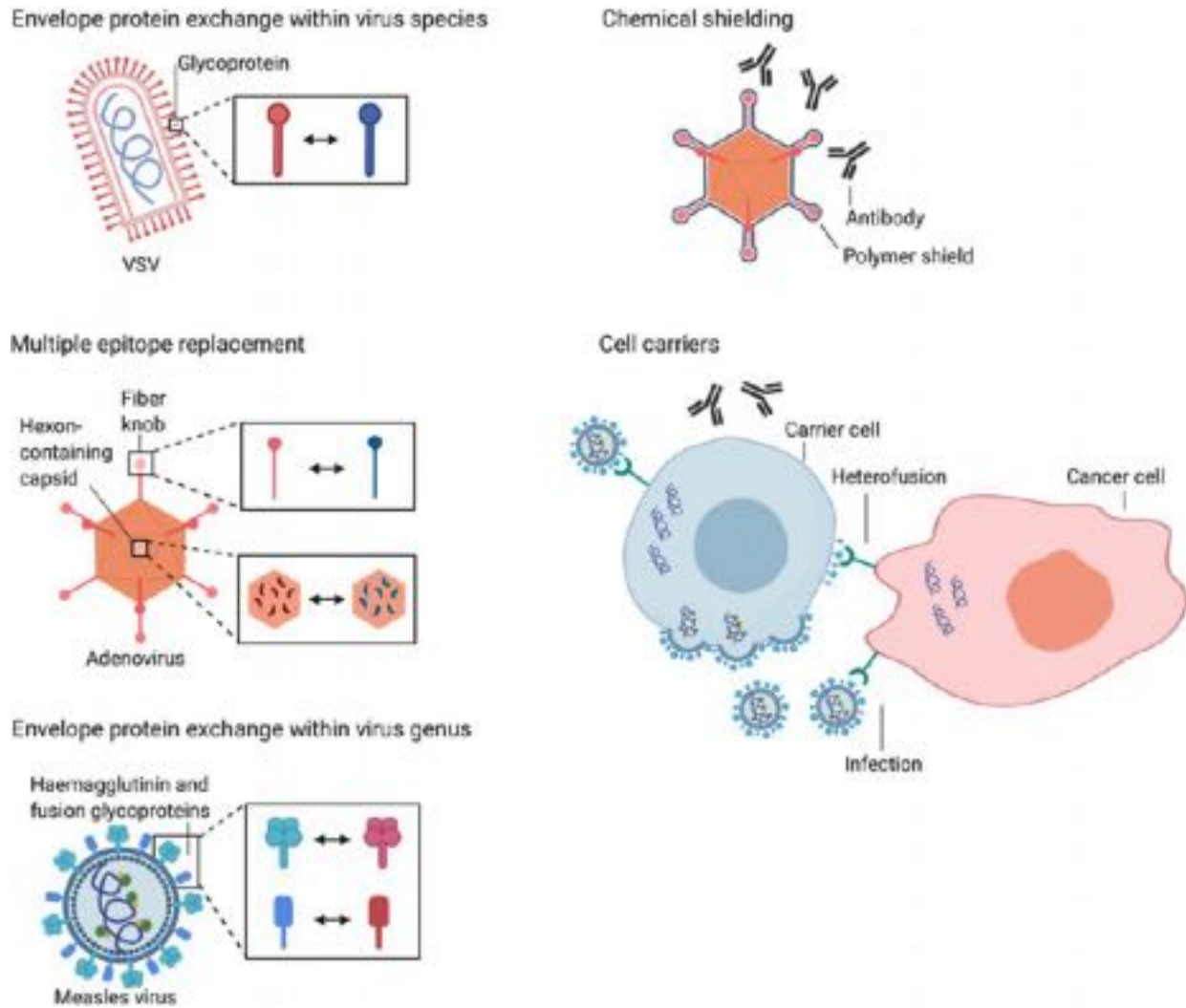


Fig. 4 Shielding strategies. *Left column: virus engineering. Right column: chemical shielding and cell carriers. Left column, top panel.* Serotype exchange. For vesicular stomatitis virus (VSV) exchange of the envelope glycoprotein is sufficient to generate a virus that is temporarily not neutralized. *Left column, center panel.* Multiple epitope replacement. For adenoviruses, epitopes on two capsid proteins must be precisely swapped to generate a virus that is temporarily not neutralized. *Left column, bottom panel.* Chimeric viruses. For MeV, exchange of both envelope glycoproteins with those of a related virus is required to generate a virus that is temporarily not neutralized. *Right column, top panel.* Chemical shielding, shown for adenovirus. Small polymers are added to purified virus particles. *Right column, bottom panel.* Cell carriers, shown for MeV. Cell carriers are infected *ex vivo* prior to administration. Infection is transmitted to cancer cells either by heterofusion, or by released viral particles. Modified from Miest, T.S., Cattaneo, R., 2014. New viruses for cancer therapy: Meeting clinical needs. *Nature Review Microbiology* 12, 23–34. (Created with BioRender.com).

In contrast, both hexon and fiber proteins harbor neutralizing epitopes of adenoviruses (Fig. 4, center left panel). To circumvent neutralizing antibodies, hexon-chimeric and fiber-chimeric variants have been generated. This was a complex endeavor because the hexon capsid protein encodes seven distinct loops that all contribute to neutralization; these loops were individually replaced while leaving the backbone of the hexon protein intact.

Serotype exchange is not an option for viruses, such as MeV, which have a single serotype. In a variation on this strategy (Fig. 4, bottom left panel), the two MeV envelope glycoproteins were replaced with those of canine distemper virus, a closely related animal *Morbillivirus* whose proteins are structurally very similar. All these variations on the theme of serotype exchange require a balance between sufficient protein diversity to avoid cross-neutralization, and enough structural similarities to support particle formation.

Chemical Shielding and Cell Carriers

An alternative to genetic modification of the viral surface proteins is chemical shielding of viruses (Fig. 4, top right panel). Different hydrophilic polymers have been used for this purpose. A prerequisite for successful application of chemical shielding is

that viral entry is not impeded. This has been demonstrated for representatives of the *Adenoviridae*, *Poxviridae*, and *Rhabdoviridae*. Chemical shielding can have several positive effects: in addition to reducing sensitivity to pre-existing neutralizing antibodies, it can limit the *de novo* induction of humoral and cellular immune responses, and uptake into non-target tissues.

Processes based on carrier cells have been developed for systemic administration of oncolytic viruses. In addition to protecting viruses from neutralization, cell carriers may possess inherent tumor tropism, thus targeting virus delivery. Cell carriers are infected *ex vivo* prior to administration. Immune cells, mesenchymal stem cells and also malignant myeloma cells have been used as carriers for oncolytic viruses. Meanwhile mesenchymal stem cells carrying adenovirus or MeV have entered clinical testing. Notably, both chemical shielding and cell carriers do not shield the viral progeny, which may limit the efficacy of these approaches.

Arming Viruses

Another key challenge for oncolytic therapy is the need to reach and eliminate not only all the cells in a primary tumor, but also tumor metastases. Oncolytic viruses as standalone therapies have not achieved durable tumor clearance, but their arming with additional therapeutic genes may address this challenge. Most arming strategies rely on “bystander effects”, i.e., the killing of non-infected tumor cells. Bystander cell killing can be elicited by prodrug convertases, radiosensitizers and immunomodulators (Fig. 5).

Prodrug Convertases

Prodrug convertases are enzymes that activate non-toxic substrates, converting them into cytotoxic substances (Fig. 5, top panel). Prodrug-encoding viruses derived from *Herpesviridae*, *Adenoviridae*, *Poxviridae*, *Paramyxoviridae*, and *Rhabdoviridae* have been developed. The clinically most advanced approach is a replication-competent retrovirus encoding the prodrug convertase cytosine deaminase (Table 1). This enzyme converts a precursor into the active chemotherapeutic 5-fluorouracil (5-FU). 5-FU can diffuse out of the cell and into surrounding cells, exerting toxic effects also on non-infected cells. Compared to systemic chemotherapy, oncolytic virus-encoded prodrug convertases can reduce systemic toxicity by local concentration of the active drug at the site of viral replication, that is within the tumor.

Radiosensitizers for Therapy and Imaging

Arming with radiosensitizers aims at local concentration of radioisotopes for ablation of malignant tissue (Fig. 5, center panel). The sodium iodide symporter (NIS), whose physiological role is to concentrate iodide in the thyroid, has been incorporated into multiple viruses to facilitate tumor ablation. Administration of NIS-encoding viruses and certain β -emitting isotopes enables localized radio-virotherapy: radioactive ions are concentrated within infected tumor cells, and these radioisotopes also affect surrounding, non-infected cells.

An additional key application of NIS is imaging, which enables real-time tracking of viral gene expression and spread. This is done in pre-clinical and clinical studies using γ -emitting isotopes that accumulate at the sites of viral replication. These isotopes, which have much more effective tissue penetration than fluorescence or bioluminescence, have allowed to characterize at high resolution the replication of oncolytic viruses *in vivo*. In particular, the combination of imaging viral replication by single photon emission computed tomography (SPECT) with tissue imaging by computed tomography (CT) allows to correlate virus replication and distribution with measures of clinical efficacy. In worst case scenarios, SPECT-CT can identify off-target viral replication.

Immunomodulators

The successes of cancer immunotherapy in recent years rekindled interest in the immunotherapeutic effects of oncolytic virotherapy. Virus-mediated tumor cell lysis leads to the release of tumor antigens in a highly immunostimulatory context. This can induce anti-tumor immune responses via *in situ* vaccination effects. Arming oncolytic viruses with immunomodulatory transgenes aims at enhancing and shaping these anti-tumor immune responses (Fig. 5, bottom panel).

The immunomodulator most commonly harnessed for this purpose is GM-CSF. GM-CSF recruits and activates antigen-presenting cells, which in turn stimulate T cells. GM-CSF is expressed by three of the clinically most advanced viruses (Table 1). Further immunomodulatory transgenes include cytokines and immune checkpoint blocking antibodies.

However, non-specific immunostimulation may result not only in activation of anti-tumor, but also anti-viral immunity. To specifically direct immune responses against tumor cells, bispecific antibodies for recruitment of immune effector cells to tumor cells have been encoded in oncolytic viruses. Furthermore, oncolytic viruses encoding tumor antigens have been generated to prime and expand tumor-specific T cells. This approach has entered clinical development (Table 1).

Additional arming strategies include encoding pro-apoptotic factors, or factors which counteract inhibitors of apoptosis to increase tumor cell death. Oncolytic viruses encoding short hairpin RNAs, microRNA mimics, and microRNA decoys to inactivate pro-tumorigenic factors have also been reported. Metabolic reprogramming of the tumor microenvironment is also attempted.

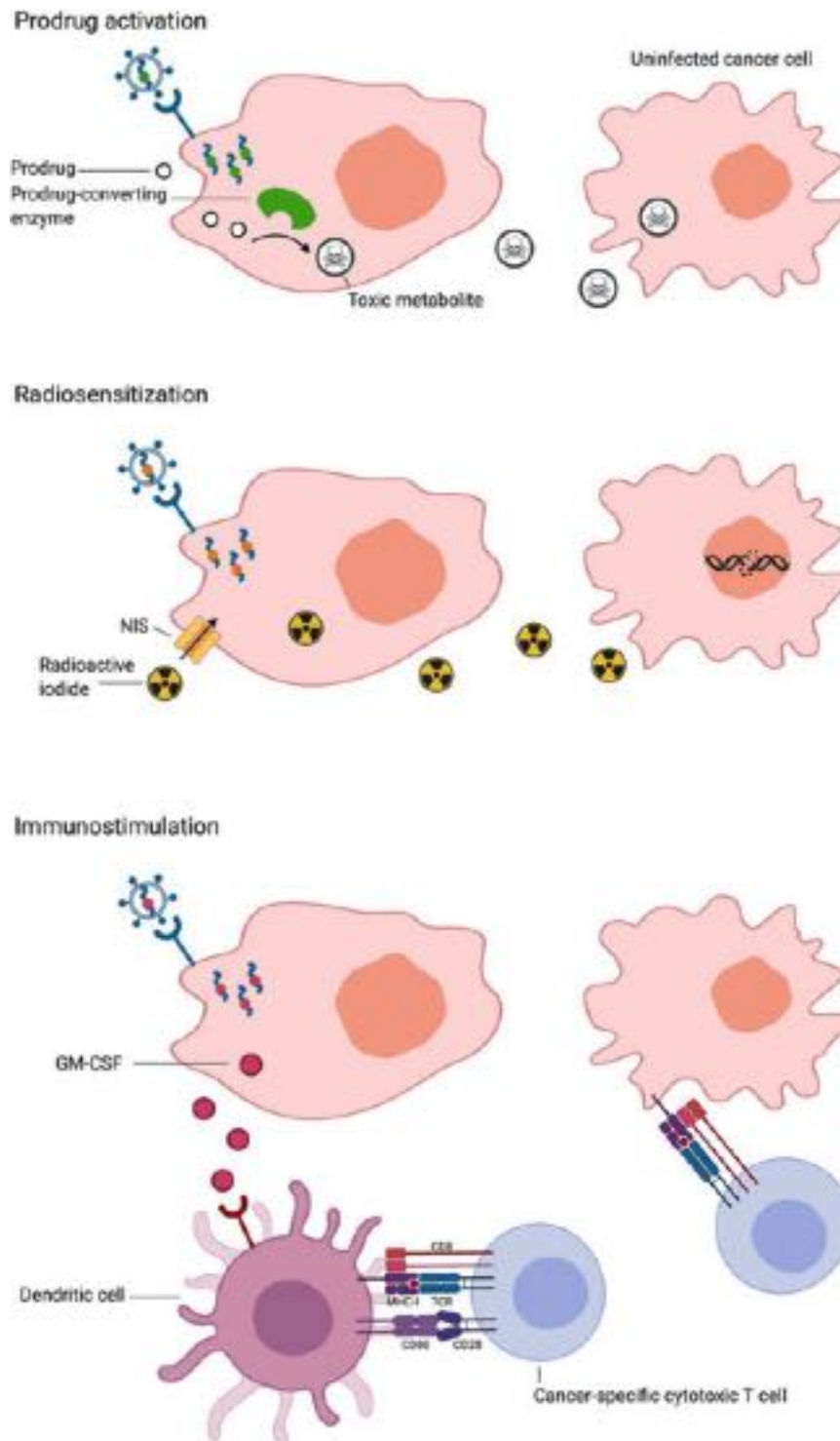


Fig. 5 *Arming strategies.* *Top panels:* prodrug activation. Cancer cells are infected by a virus that expresses a prodrug convertase that activates non-toxic precursors (prodrugs) into therapeutically active toxic metabolites. These metabolites diffuse within the surrounding tissue, killing bystander cells. *Center panels:* radiosensitization. Cancer cells are infected by a virus that expresses a symporter like NIS, which concentrates radioisotopes in infected cells, leading to radiation-induced death of the infected cell, and of nearby bystander cells. Shown here is radiation-induced DNA damage in a bystander cell. *Bottom panels:* immuno-stimulation. Cancer cells are infected by a virus that expresses an immunostimulatory gene like GM-CSF. Infected cells activate anti-tumor immune responses. GM-CSF activates dendritic cells to present tumor antigens to T cells via MHC molecules and to express costimulatory molecules such as CD80. Recognition of the presented antigens by a T cell with a cognate T cell receptor (TCR) primes an anti-tumor immune response, leading to T cell-mediated anti-tumor cytotoxicity. Modified from Miest, T.S., Cattaneo, R., 2014. New viruses for cancer therapy: Meeting clinical needs. *Nature Review Microbiology* 12, 23–34 (Created with BioRender.com).

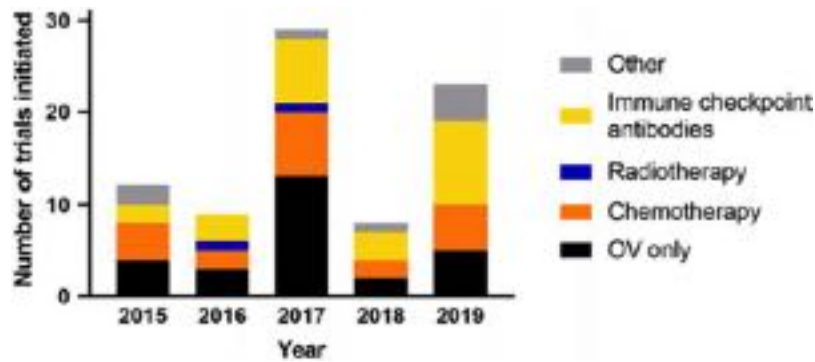


Fig. 6 Clinical trials of oncolytic virotherapy in combination with other treatments. Clinical trials initiated in the last 5 years (2015–2019) are categorized by type of combination therapy.

Combination Therapies

A paradigm of cancer treatment is that no single drug or therapy will cure cancer. Surgical removal of tumors, chemotherapy, radiation therapy, and now also immunotherapy and oncolytic virotherapy are used in combination to treat cancer patients. Indeed, several current clinical trials combine an oncolytic virus with other therapeutics (Table 1). In particular, more than half of all trials initiated in the years 2015–2019 combine virotherapy with at least one additional treatment modality (Fig. 6) and it is to be expected that oncolytic viruses will increasingly be integrated into combination therapy regimens.

For instance, adenovirus H101, the first clinically approved oncolytic virus, is applied in combination with standard chemotherapy for head and neck cancer (Table 1). Several other combinations of chemotherapy and virotherapy are currently in clinical trials. Some of these protocols take advantage of virus-encoded enzymes to locally convert prodrugs into active chemotherapeutics, as outlined above.

In pre-clinical studies oncolytic viruses have also been combined successfully with targeted therapies. For example, Talimogene laherparepvec (T-vec) showed synergistic effects with kinase inhibitors (MEK inhibitors) in mouse models of melanoma. So far, radiosensitizers have only been used for therapeutic purposes in pre-clinical models, but several trials have included imaging to assess the biodistribution of the virus.

The prospect of inducing anti-tumor immunity through oncolytic vaccination prompts the use of viruses in immunotherapeutic regimens. The combination of oncolytic virotherapy with immune checkpoint blockade appears especially promising and is increasingly pursued in clinical trials (Fig. 6). The proposed mechanism of action for this combination is that oncolytic virotherapy induces tumor-specific T cell responses that are subsequently reinforced by checkpoint blockade.

The sheer endless possibilities to combine oncolytic viruses with different cancer treatments require prioritization. Future studies will be guided by better understanding of the mechanisms of action of cancer therapeutics as well as their interactions with other treatments and stringent testing in more refined pre-clinical models.

Complex combination therapies raise questions regarding the optimal dosing and scheduling of the individual treatment components. Several groups have attempted to mathematically model combination treatments with oncolytic viruses to inform dosing and scheduling with the aim to predict – *in silico* – the most effective combination regimens.

Surgery is a mainstay in treatment of many cancer types. Incorporating oncolytic viruses into the treatment of these cancers allows for so-called “window of opportunity” clinical trials. Oncolytic viruses are administered in a neoadjuvant setting, that is before surgical removal of the tumor. Histological assessment of the tumor tissue for signs of viral replication and immune infiltration can provide valuable insights into therapeutic effects, or lack thereof, of oncolytic virotherapy.

To summarize, the dual mechanism of action of oncolytic viruses, namely direct intratumoral spread followed by inflammatory boosting of the anti-tumor immune response, makes them ideal partners for combination with immune checkpoint inhibitors and other immune-oncology agents. Another current priority is research on synergies with conventional anticancer drugs.

Perspectives

Viruses are emerging as a promising new modality in the fight of cancer. While current oncolytic virotherapy approaches are very safe in humans, therapeutic efficacy has been less than expected from pre-clinical studies. This may reflect the fact that many current approaches are based on highly attenuated viruses, and that generalized attenuation interferes with clinical efficacy.

Next-generation viruses will seek to retain wild-type replicative competence in cancer tissue, while entry targeting and post-entry replication control mechanisms will restrict their spread in off-target tissues. In addition, arming strategies based on chemo-, radio-, or immuno-therapies will be potentiated by enhance virus replication in tumors.

Next-generation oncolytic viruses will be deployed strategically, based on deeper understanding the susceptibility of different tumor types, and on progress in stratifying patients. In synergy with approved cancer therapeutics, oncolytic viruses are finding their niches in the treatment of cancer.

Further Reading

- Asada, T., 1974. Treatment of human cancer with mumps virus. *Cancer* 34, 1907–1928.
- Barry, M.A., Rubin, J.D., Lu, S.C., 2020. Retargeting adenoviruses for therapeutic applications and vaccines. *FEBS Letters* 594, 1918–1946. doi:10.1002/1873-3468.13731.
- Bluming, A.Z., Ziegler, J.L., 1971. Regression of Burkitt's lymphoma in association with measles infection. *Lancet* 2 (7715), 105–106.
- Cattaneo, R., Miest, T., Shashkova, E.V., Barry, M.A., 2008. Reprogrammed viruses as cancer therapeutics: Targeted, armed and shielded. *Nature Reviews Microbiology* 6, 529–540.
- Dorer, D.E., Nettelbeck, D.M., 2009. Targeting cancer by transcriptional control in cancer gene therapy and viral oncolysis. *Advanced Drug Delivery Reviews* 61, 554–571.
- Engeland, C.E., Grossardt, C., Veinalde, R., *et al.*, 2014. CTLA-4 and PD-L1 checkpoint blockade enhances oncolytic measles virus therapy. *Molecular Therapy* 22, 1949–1959.
- Hill, C.A.P., Bau, L., Carlisle, R., 2020. Methods for modification of therapeutic viruses. *Methods in Molecular Biology* 2058, 7–29.
- Kelly, E., Russell, S.J., 2007. History of oncolytic viruses: Genesis to genetic engineering. *Molecular Therapy* 15, 651–659.
- Miest, T., Cattaneo, R., 2014. New viruses for cancer therapy: Meeting clinical needs. *Nature Review Microbiology* 12, 23–34.
- Peters, C., Rabkin, S.D., 2015. Designing Herpes viruses as oncolytics. *Molecular Therapy Oncolytics* 2, 15010. doi:10.1038/mt.2015.10.
- Ribas, A., Dummer, R., Puzanov, I., *et al.*, 2017. Oncolytic virotherapy promotes intratumoral T cell infiltration and improves anti-PD-1 immunotherapy. *Cell* 170, 1109–1119.
- Ruiz, A.J., Russell, S.J., 2015. MicroRNAs and oncolytic viruses. *Current Opinion in Virology* 13, 40–48.
- Schneider, U., Bullough, F., Vongpunsawad, S., Russell, S.J., Cattaneo, R., 2000. Recombinant measles viruses efficiently entering cells through targeted receptors. *Journal of Virology* 74, 9928–9936.
- Springfeld, C., von Messling, V., Frenke, M., *et al.*, 2006. Oncolytic efficacy and enhanced safety of measles virus activated by tumor-secreted matrix metalloproteinases. *Cancer Research* 66, 7694–7700.
- Twumasi-Boateng, K., Pettigrew, J.L., Kwok, Y.Y.E., Bell, J.C., Nelson, B.H., 2018. Oncolytic viruses as engineering platforms for combination immunotherapy. *Nature Review Cancer* 18, 419–432.

Relevant Websites

<https://clinicaltrials.gov/>
ClinicalTrials.gov.

PREVENTION

Surveillance of Infectious Diseases

Norman Noah, London School of Hygiene and Tropical Medicine, London, United Kingdom

© 2021 Published by Elsevier Ltd.

Glossary

Case–fatality rate Number of persons dying of an infection divided by total number of persons with the disease. Thus, CFR of 5% means that 5 of 100 persons with the infection died. It is distinguished from “Mortality Rate” which measures the death rate in the whole population.

Endemic An infection continues in a country, though may vary in incidence at different times. Many enteroviruses, adenoviruses and respiratory viruses are endemic.

GUM clinics The term GenitoUrinary Medicine clinics is often used for special clinics for sexually transmitted infections (STIs).

Incidence The rate of new infections (number by population) in a given time – for example, 5 cases of influenza per 1000 per year. Good for short-term infections.

Outbreak and epidemic Most infectious disease epidemiologists do not distinguish between these terms and

use them interchangeably. Generally, an outbreak is defined as a localized increase in cases, whereas an epidemic is more widespread, perhaps affecting a whole country.

Outbreak When an infection occurs at a frequency higher than expected for that time or place. It is basically an increased incidence which is usually unexpected. Two linked cases is also theoretically an outbreak.

Pandemic This term should be restricted to infections affecting many countries, and may extend virtually worldwide. SARS spared many countries and indeed some continents, but it was a pandemic. SARS-CoV-2 and AIDS undoubtedly caused pandemics.

Prevalence The number of infections at any one time in a given population, expressed as a rate. Good for chronic infections such as chronic hepatitis B or C and serological studies (e.g., prevalence of varicella antibody at age 15 in a given population is 95%).

Introduction

Surveillance is undoubtedly an essential – indeed critical – ingredient of any disease control program. It is used to paint a picture of the progress and overall burden of infection or disease, so that any preventive or therapeutic action can be measured as it advances. In this way, the impact of an infection, the effect of an intervention or health promotion strategy, health policy, planning, and delivery can be monitored. Surveillance is the ongoing and systematic collection of routine data which are then analyzed, interpreted, and acted upon. It is essentially a practical process, which nevertheless can be useful in other ways. Its main purpose is to analyze time trends – but these can include not simply fluctuations in overall numbers, but also changes in age and sex distributions, geographical locations, and even possibly, in some of the more sophisticated established surveillance systems, at-risk groups (such as particular social, ethnic and occupational groups). Surveillance is essential for evaluating the impact of an intervention, such as mass vaccination, on a population. It can also act as a fairly sensitive system for the early detection of outbreaks.

Essential Characteristics of Surveillance

The word *ongoing* in the description of surveillance helps to distinguish it from a survey, which is usually finite, tends to focus more on one or more groups of persons, and involves a questionnaire. Nevertheless, surveillance does not have to continue forever – when it is no longer useful, it should be stopped.

To be *systematic* and *consistent* are other important ingredients of surveillance. If reporting centers are not consistent in what they report, nor regular, the data they send in will be uninterpretable, and probably useless. Defining what needs to be reported, its frequency, and agreeing on the criteria for making a reportable diagnosis are necessary to make sense of the data.

Timeliness is another important ingredient of a surveillance system. This is especially the case with infection. It is particularly essential for the early recognition of outbreaks, but it is important also in evaluating the success or otherwise of programs to control the spread of an infection e.g., the effect of lockdown on coronavirus disease (COVID-19) on a population.

Representativeness is another important ingredient of a surveillance program. The program has to be representative of the effect of that infection on a population. This may not always be possible, and any biases and inconsistency in reporting has to be taken into account in interpreting the data.

Completeness is generally less important than representativeness, especially for relatively common infections. Completeness becomes important if or when the number of infections decreases, whether on its own or through a public health intervention. For some serious infections, completeness may also become important. Completeness is discussed in more detail later in this article.

The *collection of routine data* is another important characteristic of surveillance, especially with surveillance of laboratory infections. Generally, testing of samples is done for diagnosis, not primarily for surveillance. As laboratory testing is expensive, making further use of the results by contributing to surveillance makes for more efficient use of information, and the contribution to surveillance itself can

often justify further testing. Typing echoviruses or coxsackie viruses can seldom be justified on clinical grounds alone, but the surveillance of serotypes can provide valuable information on the epidemiology of these viruses, their clinical characteristics, seasonality, and age and sex distributions – and of course whether there has been an outbreak. Echovirus 4 for example tends to be rare, fairly localized even within a country, but, when it occurs, it has a high aseptic meningitis rate, and causes a short and sharp autumn outbreak. Echovirus 9 on the other hand is far more common, more widespread geographically when it causes an epidemic, more benign, with a macular rash, and with meningitis not an especially common manifestation. The weekly patterns of some of these viruses can provide clues on future resurgences. Echovirus 4 as stated, comes and goes in one season. In the autumn of 1974, a small outbreak of echovirus 19 did not disappear entirely throughout the winter. A further and much larger outbreak in the autumn of 1975 was predictable, and occurred.

Although surveillance is essentially a practical exercise, this article attempts to show that surveillance can also be useful in giving us clues about an infection, whether it be its natural history, etiology, severity, or outcome.

Collection of Data: Sources of Data in Surveillance of Viruses

Death Certification

In most developed and middle-income countries, deaths are certified primarily for legal reasons, but have proved to be an important data source to use for surveillance. Clearly, they tend to be useful mainly for serious infections, and may suffer from inaccuracies, but remain a useful basic data source. Diseases with a high mortality rate and a short duration of illness (the viral hemorrhagic fevers for example) will obviously be better represented by death certification than those with a low mortality. HIV/AIDS, before the era of HAART, had a high mortality but long periods of symptomless and symptomatic infection, so that death certification data needed careful interpretation. It is well known that death certification standards of accuracy vary not only from country to country, but also within countries. In England, in the early years of the HIV epidemic, it was shown that deaths in single men had risen in number although AIDS was not mentioned on the death certificates. With influenza outbreaks there is usually an increase in total mortality in the elderly otherwise unaccounted for. The reporting of deaths from SARS-CoV-2 both within and between countries is also known to be inconsistent. Excess deaths have become a useful measure in estimating mortality from influenza and SARS-CoV-2.

Laboratory reporting systems can sometimes be another useful source of information on mortality from infection.

Notifications

Surveillance systems such as statutory notification tend to be based mainly on clinical features. These can be very useful for common diseases with distinctive clinical syndromes, such as measles and mumps. It is important however not to make a disease notifiable unless there is a good reason for it: “good” reasons include a mass vaccination program (when surveillance is virtually mandatory), any other mass control program, or serious diseases for which contact tracing, mass or close contact prophylaxis or investigation into source is necessary. Serious less common viral infections such as poliomyelitis are notifiable in most countries. This is because contact tracing and preventive measures can be taken. Broader clinical diagnoses, such as aseptic (viral) meningitis, may on the face of it be less useful to notify, as it is usually impossible at the bedside to distinguish such causes of it as, for example, the coxsackie B viruses, echoviruses, and mumps. Nevertheless, notification of aseptic meningitis can be useful because a rapid rise in notified cases may need to be investigated further. Moreover, the timing of any epidemic may give a clue to etiology – mumps meningitis tends to increase in spring and is usually accompanied by a concurrent outbreak of clinical mumps, while enterovirus epidemics are more likely to occur in autumn.

Other Sources of Data

Specific general practitioner (GP) surveillance systems are useful in providing data of epidemiological value for common infections that are not notifiable, such as the common cold or chickenpox (in UK). They are of course clinically based, but are nevertheless useful, and often surprisingly accurate, possibly because those GPs who subscribe to a surveillance system are motivated to do so. GP surveillance systems are often sentinel-based, that is, based on a sample of GPs in a country, region, or area. In the English system, GPs provide data on the base populations of their practices, so that rates of infection can be provided as a routine, a feature that is almost unique amongst surveillance systems. Thus, they are good for common infections which to make notifiable would possibly be wasteful, such as chickenpox. Moreover each sentinel would normally provide complete reporting. GP surveillance systems also tend to be good for timeliness and completeness. Outbreaks of influenza are often first recognized by GPs.

Laboratory Data

It is useful to think of laboratory data as being of qualitative rather than quantitative value as they add quality and detail to disease surveillance. Thus, in the example already used for aseptic meningitis, a precise diagnosis of mumps, echovirus, or coxsackie type is

particularly useful in clustering and outbreaks. Indeed, as the enteroviruses exhibit a strong late summer/autumn seasonal pattern with each enterovirus returning to its baseline in winter, when numbers of a particular type continue to be reported at a level higher than the winter baseline, the return of this virus to cause another epidemic in the following summer can usually be safely predicted. Laboratory data are also essential in qualifying food poisoning and gastroenteritis. Separating viral from bacterial causes is a useful first step, as their management tends to be very different. Norovirus gastroenteritis can be food-borne, but also spreads very easily from person-to-person because an extremely small dose is necessary for infection to occur, and the virus being fairly resistant to the environment will survive for some time. Management therefore must concentrate on hygiene. Rotavirus on the other hand is not commonly known to be foodborne. It can be distinguished from norovirus by its strong winter pattern in temperate climates, and its preponderance in infants. With most bacterial causes of food poisoning, especially the salmonellas, management often depends on the removal of the offending food. This is less common with viruses – only occasionally do hepatitis A or norovirus cause outbreaks of food poisoning. Laboratory surveillance is particularly essential for unraveling the mass of respiratory viral infections that afflict humans, such as respiratory syncytial virus, adenoviruses, parainfluenza viruses, and rhinoviruses. Indeed, it is particularly useful for influenza – not only separating it from influenza-like illness, but also in identifying influenza A and B, and if A, the subtype and variant. The constituents of every influenza vaccine depend on laboratory data.

Surveillance of Outbreaks

Surveillance of outbreaks (as opposed to individual infections) can be revealing, and important to allow public health measures. In one incident, routine surveillance revealed two cases of hepatitis B associated with one tattooist. Inspection of the notification data on hepatitis revealed 33 cases in all associated with the tattooist. The existence of the noroviruses was suspected in the UK many years before the organisms were identified. This was because some outbreaks which did not fit the characteristics of known infections but had characteristics of their own had occurred. There are some surveillance systems specifically for outbreaks. One example is EpiWATCH. This is run by the National Health and Medical Research Council (NHMRC) Center for Research Excellence Integrated Systems for Epidemic Response in Australia. Epidemic intelligence systems such as EpiWATCH can be useful for the global surveillance of outbreaks.

Hospital Admissions

These can be useful for certain more serious infections, such as hepatitis and encephalitis. They can be unwieldy, generally lack detail, and often are published a year or more after the events have occurred. Special more timely surveillance systems can be created, as for example with SARS-CoV-2 in 2020.

Serological Surveillance

Serological surveillance has become increasingly important, and is a useful tool in assessing the immunity of a population, though it can also be used to identify vulnerable individuals. The immunity of a population can be vaccine-induced, and serological surveillance is a valuable adjunct to the methods available to monitor a mass immunization program. Vulnerable age groups can be identified, and booster doses of the vaccine introduced. An example of the importance of serological surveillance in determining public health policy is included below (analysis by person).

Surveillance of Viruses in Nonhuman and Environmental Sources

To build a picture of an infection, animals, birds, and the environment have been placed under surveillance. Rabies in foxes and other wildlife, influenza in birds, pigs, and other animals, are examples of important and fairly successful surveillance systems. With the more recent examples of the emerging viruses such as Nipah and the group B coronaviruses, this type of surveillance is becoming important. At the time of writing, there are some surveillance systems for influenza A virus in avians and humans. Environmental surveillance of sewage and other wastewaters for wild and vaccine polio virus, as well as other viruses, exist.

Other Sources of Data

Records of sickness absence, absence from school, calls to an emergency room, if available, can provide speedy information that something has happened, but tend to be nonspecific. Surveillance of antiviral resistance will become important. AIDS was first discovered in the US when the use of a drug for treatment of the then rare *Pneumocystis carinii* infection in gay men began to increase. Surveillance of prescriptions or drug usage as a means of monitoring infection will probably increase in importance.

Newer Types of Surveillance System

Syndromic Surveillance

There is no strict definition of this type of surveillance. One description is: “Syndromic surveillance complements traditional public health surveillance by collecting and analyzing health indicators in real time”. It is usually based on an increase in an unusual group of clinical features which may lead to investigation. One of the earliest examples of this, though not a virological one, was the recognition of an unusual cluster of symptoms (subsequently labeled the Eosinophilia–myalgia syndrome) linked to L-tryptophan consumption originating from a single source in Japan. This was a dietary supplement manufactured using genetically-engineered bacteria. The beginning of the outbreak of SARS-CoV-2 was an increase in an association of unusual clinical features in Wuhan – in effect a new syndrome. Although there was some delay in reporting this particular unexpected increase, the primary objective of this type of surveillance is to recognize an outbreak as early as possible. Sudden changes in numbers of doctors’ prescriptions for specific drugs, school or workplace absenteeism may also be used to uncover new outbreaks or new infections. Electronic surveillance systems are now being developed to detect new outbreaks – whether virological, bacteriological, or chemical. The fear of terrorism has stimulated this greater interest in syndromic surveillance. Syndromic surveillance will undoubtedly become more important with time.

Sentinel Surveillance

Sometimes, for research purposes or otherwise, more detailed or higher quality data on an infection or infections are required than normally available through a passive system. It may be too much of a burden – and probably wasteful – to expect all laboratories or other sources of data to include or collect this extra information. In this instance a form of active surveillance can be introduced. A sentinel surveillance system can be set up using a selected – and willing – number of motivated laboratories. Chosen laboratories would be expected to diagnose a reasonable number of cases of the infection, be within a fairly large area in which they would be expected to diagnose all cases of the infection, and of course to be technically excellent. Sentinel surveillance systems tend to be time limited, and are unsuitable for rare infections.

Enhanced Surveillance

Enhanced Surveillance is a term used for infections of public health importance to combine epidemiological and microbiological data, as well as other types of information necessary. It may include data from a range of sources. For COVID 19, clinical and microbiological as well as social data such as ethnic background have been collected by several countries, including from asymptomatic persons. Serological testing gives information on immunity both in a person as well as in a community. This information can be obtained from primary care and various other settings, including places of education, medical institutions and care homes. Clearly, enhanced surveillance is expensive on costs and manpower, and should only be put into practice for special reasons, such as a pandemic. Further examples covering HIV are given in the next section; many of these are also relevant to Covid19.

Active Surveillance

This is a form of enhanced surveillance to ensure completeness and consistency. Reporting sources have to report make negative returns whether or not they have had cases.

Surveillance of HIV/AIDS

The association of HIV/AIDS with stigma makes surveillance of this infection especially difficult. It is an example of the importance of tailoring surveillance to a specific serious infection if it becomes necessary to do so.

In the UK and some other countries with data protection acts, HIV infection, as with other STIs, is not notifiable. Special confidential surveillance systems through clinicians and GUM clinics, as well as laboratories, are in place. These are especially important for assessing risk factors. Inclusion of risk factors is essential for targeted intervention – for example, the proportions and rates of new diagnoses attributed to men who have sex with men, heterosexual sex, mother-to-infant, blood transfusion, IVDUs, and other needlestick injury.

Laboratory reporting is essential. Death certification is useful, though it has been shown that men who have sex with men, and probably those with other risk factors, are under-represented. In the UK, matching reported cases with death certificates is very important, as it allows for detection of deaths due to AIDS (such as pneumonia) as well as deaths associated with AIDS, and which are seen now in HIV-infected individuals – these include liver and cardiovascular disease, overdoses, and malignancies.

A surveillance system, based on unlinked anonymous testing of samples of blood taken routinely from certain at-risk population or occupational groups has been shown to provide valuable information on HIV infection in these populations. Specific screening systems for blood donors, military recruits, commercial sex workers, and family planning/termination of pregnancy

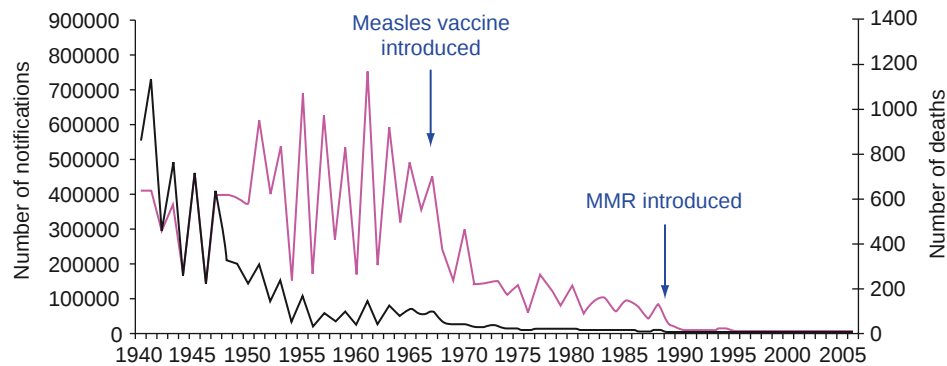


Fig. 1 Measles notifications: cases and deaths, England and Wales 1940–2006. Reproduced from the Health Protection Agency (www.hpa.org.uk).

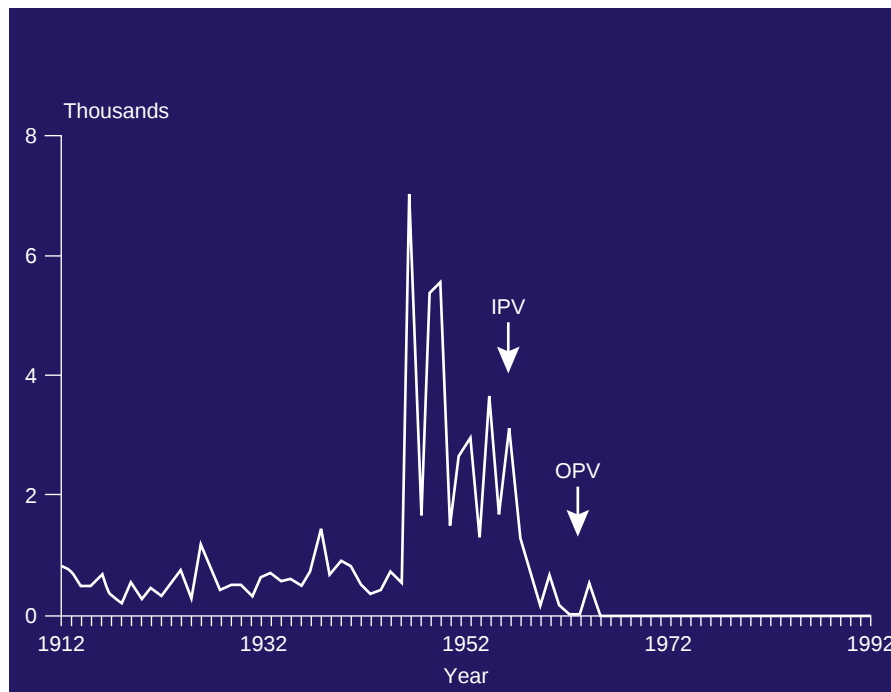


Fig. 2 Notifications of acute poliomyelitis in England and Wales 1912–1993. Reproduced from the Health Protection Agency (www.hpa.org.uk).

clinics are also useful if these populations are to be targeted. Behavioral surveillance should also be seriously considered, to assist in identifying future trends, healthcare planning, as well as for specific health promotion efforts.

Attributes of Surveillance Systems

Completeness

Incompleteness is an almost universal drawback of most notification systems. They should never be dismissed for this reason alone. Statutory notification systems can be essential for surveillance and control. For common infections completeness may not be worth striving for, because notifications (assuming consistency in reporting) will generally provide information on trends, as well as fairly accurate information on age, sex and seasonal distributions, and possibly on place. Notifications were able to show a small increase in measles cases in the early 2000s which led to the need for a preschool booster.

The effect of mass vaccination programs can also be monitored fairly closely with statutory notification systems, as with measles ([Fig. 1](#)) and acute paralytic poliomyelitis ([Fig. 2](#)) in the UK. When a mass vaccination or other universal control program reduces the incidence of an infection to low levels, completeness becomes much more essential. For serious infections also, such as SARS or Lassa fever, for which contact tracing or other control measure is necessary, completeness is essential.

In active surveillance, reporters make negative returns if they have had no cases during the reporting period, to ensure completeness. In some countries, enhanced surveillance has been used to assess more accurately the true incidence of an infection – regions or districts are chosen to report all cases of a particular infection or infections. It is a hybrid of active and sentinel surveillance.

Timeliness

Timeliness is important for infections for which urgent public health measures have to be undertaken, such as SARS-CoV-2, poliomyelitis and viral hemorrhagic fever, as well as any outbreak. In some instances, infections, not normally urgent, can become so as an elimination program progresses. In a country with elimination of measles as its goal, a case of indigenous or imported measles needs to be dealt with urgently, as it may lead to an outbreak if not controlled immediately. Laboratory data are often not timely, and hospital data generally even less so, but can make up in accuracy what they lose in timeliness.

Accuracy

Accuracy is clearly important, though some minor degrees of inaccuracy can be tolerated in some common infections. Clinical data are most liable to have some inaccuracies, though even laboratory data may be inaccurate. Case definitions and quality control systems can be useful to improve accuracy.

Representativeness

For surveillance to provide an accurate picture of the impact of a particular infection, representativeness is essential. It is perhaps the most important quality for any surveillance system. Having a wide coverage of reporting clinicians and laboratories, or a well-chosen sample of sentinel sites, is necessary for the data collected to be representative of an infection in a country. Sometimes it may be necessary to assess data from various sources, such as notifications/GPs (clinical), hospital, laboratory, and death certificates.

Consistency

Consistency is another crucial basic attribute of any surveillance system. Reporters must know what to report (case definition) and how often. Otherwise it will not be possible to interpret trends. Active surveillance is one way of ensuring consistency.

Analysis of Data

Time

The three basic analyzes by time, place, and person should be routine. Computer programs have made analyzes of data quicker but somewhat less flexible. There is no standard period for analysis by time. Depending on what the surveillance intends to show, yearly, quarterly, monthly, four-weekly, or weekly time intervals can be used. In surveillance, time intervals shorter than this are rarely used, although daily data were provided in many countries to monitor SARS-CoV-2. Monthly intervals have the disadvantage of having unequal numbers of days in each month, and are difficult to use when reporting is weekly; for seasonal trends four-weekly periods are better but cannot be divided into quarterly periods. In viral surveillance, four-weekly rather than weekly intervals tend to be most useful in showing seasonal changes. There is more likely to be more variation (“noise”) in weekly intervals, making for less smooth changes. For secular trends, quarterly or annual intervals are generally used. Analyzing by time can reveal regular changes in the periodicity of viruses, enabling some of them to be predicted.

A basic knowledge of seasonal and secular patterns makes it easier to detect changes that signify a possible epidemic, and to differentiate these from a random variation. It is important to remember when analyzing laboratory data that there is often an interval, which can be 2 weeks or more, between date of onset and date of reporting.

Person

Analysis by age and sex is another basic analysis in surveillance. It can identify those most affected, and vulnerable groups. Changes in age distributions may provide important clues about a changing viral infection, and the effect of mass interventions on the age distribution of an infection can be monitored. Changes in the age distribution of measles in 1994 in the UK signified that an epidemic in older children was imminent, and the vaccine schedule was changed to include an extra booster injection (MR) to children aged 5–16 years. This averted the outbreak and the booster dose became a permanent feature of the routine immunization schedule in the UK. Indeed the changes in age distribution following mass vaccination could be considered an epidemiological side effect of mass vaccination. Requests for occupational groups and travel histories should be selective. For poliomyelitis, SARS, dengue, and the viral hemorrhagic fevers, travel histories are required. Occupational group may be useful for norovirus, and hepatitis types A, B, or C. Specific risk factors may be worthwhile for HIV, hepatitis B and C.

Table 1 Stages of a viral infection

<i>Clinical</i>
A. Uninfected
B. Infected
1. Asymptomatic
2. Symptomatic unreported
3. Symptomatic, sees a doctor
4. Symptomatic, admitted to hospital
5. Symptomatic, dies/survives

Note: Chronic carriers can occur at any of the B stages – for example, hepatitis B or C.

Table 2 Stages in laboratory diagnosis

A. Uninfected
B. Infected
1. Asymptomatic
2. Symptomatic, unreported
3. Symptomatic, sees a doctor
4. Symptomatic, specimens submitted
5. Symptomatic, specimens positive
6. Symptomatic, specimens reported to surveillance system

Place

Analysis by place can pinpoint local outbreaks. Some echoviruses (e.g., echovirus type 4) can cause rare short local outbreaks, other types (e.g., types 9 and 11) are more common and more widespread. Food-borne outbreaks of hepatitis A and norovirus are generally picked up locally through routine surveillance, but sometimes more extensive outbreaks caused by a more widely distributed foodstuff, including shellfish or frozen soft fruit, may be identified.

Interpretation

Collection and analysis are generally routine functions; skill is required in interpretation of the data. No statistic is perfect, and surveillance data, like all data, must be interpreted with caution. One must take into account the origins of the data – clinical, laboratory, or hospital. Not only must the reliability, or otherwise, of the data be evaluated, but what the data signify in the natural history of the infection must also be recognized.

Every viral infection has its stages and these must be recognized before surveillance data can be sensibly interpreted. At what stage are the data being collected important to understanding and interpretation? Using influenza or hepatitis A as examples, and a defined population (**Table 1**), only a proportion of persons in the defined population will be infected with the virus. They can only be comprehensively detected by screening, and serological surveillance will identify these persons, or assess population immunity (B1 in **Table 1**). In HIV/AIDS surveillance, unlinked anonymous testing of samples of blood taken routinely at say, an antenatal clinic, can give vital information on the prevalence of HIV infection, since in this infection presence of antibody denotes infection, not immunity. A smaller proportion will be ill (B2), but only some of these will visit a doctor (B3). Surveillance systems based on GP consultation rates have now been recognized as an important addition to the spectrum of a disease, and many countries have excellent systems. Of those patients that do visit their family doctor, only some will be admitted to hospital (B4). Finally, only some will die (B5).

For laboratory data, the stages are slightly different (**Table 2**), but still important in understanding what the reported data mean. As before, only a proportion of persons will be infected (B), some will be asymptomatic (B1), a smaller proportion will be ill (B2), and an even smaller proportion still visit their doctors (B3). Not all doctors will send specimens to a laboratory (B4), and only a proportion of these specimens (B5), depending on accuracy of the identification process, the method of transport, the fragility of the organism, and the swabbing or other sampling technique, will be positive. Finally, depending on the level of consistency of reporting, only some of these will be reported (B6). It is essential to recognize these stages in interpreting surveillance data.

Biases will inevitably occur between these stages. Collection of data in routine surveillance is not normally a scientific process as one has to rely on readily available data – data obtained mostly for other reasons, such as to make a definitive diagnosis. Only the most severe cases die, and death certification thus provides, at best, a limited view of any disease. Similarly, only certain types and severity of cases will be admitted to hospital (some admissions are for social reasons for example) or even visit their family doctor.

Certain age, sex, and perhaps social or occupational groups are more likely to seek medical help, be investigated and be reported. In laboratory data, more severe cases, or children, are perhaps much more likely to be investigated in detail.

These shortcomings of surveillance data do not make them useless – but their strengths and limitations must be recognized.

Feedback

If interpretation is turning statistics into information, feedback is getting the information across to those that matter, and those that need to know, so that action – the objective of surveillance – can be taken. Without feedback, surveillance is pointless. Feedback is most likely to be informative if undertaken by those most closely involved in the surveillance cycle, and who understand the significance of the data they are receiving.

Feedback should be aimed at contributors and those in public health. Contributors will then be aware of which viruses are circulating and this will help them to know what to look for in their own tests (e.g., what echovirus or adenovirus types are in circulation). Moreover, routine surveillance will undoubtedly uncover outbreaks of infection, which will need further investigation and control at local, national, and even international level. An interesting and welcome side effect of a flourishing microbiological/feedback surveillance system is that it often stimulates better quality control within reporting laboratories.

Regular feedback is essential, not only to contributors, but also to those who can act for the public health. Regular and useful feedback encourages good and regular reporting. Generally, the periodicity of feedback should reflect the frequency of reporting – weekly feedback for weekly reports for example. Regular topic-based reviews are also important.

Evaluation of Surveillance

A surveillance system is like a country's train system. Once the rail lines have been built, the goods that will be carried along those lines can be changed according to need. Similarly, in a surveillance network, once the lines of communication have been laid down, the data being reported can be changed according to what is most important at the time (though probably not too frequently). Nevertheless, surveillance systems should ideally be frequently evaluated for usefulness, as well as for accuracy, efficiency, and effectiveness.

They should also be sufficiently flexible, so that “new” or emerging infections can be included in an emergency or when the need arises. The successful implementation of international surveillance for SARS was instrumental in controlling it. Emergency surveillance was also essential following the tsunami of 2004, and is also necessary for the successful management of other disasters following earthquakes, hurricanes, and floods.

Surveillance systems should be evaluated before they are set up, and again at regular intervals thereafter. Before implementing a surveillance system, is there an adequate public health and administrative infrastructure in place to take action? Are the data to be collected representative and sufficiently timely for the specific infection? Are they useful, and is action being taken on the information? If not, is the feedback inadequate?

Global and International Surveillance

The ease and speed of modern travel, the distribution of goods (especially foodstuffs) across increasingly wide parts of the world, and the uncontrollable spread of birds and other wildlife across boundaries has made global and international surveillance essential for outbreak and infection control. Surveillance of viral infections such as influenza and SARS-CoV-2 now requires the expertise of many professionals – epidemiologists, virologists, and vaccinologists, clinicians, statistical modelers, veterinarians, managers, and planners. These occur in many different countries so that information can be exchanged, and attempts made on a global basis, to prevent the next pandemic. Epidemics of Zika virus, dengue and West Nile virus have also spread widely recently. AIDS/HIV was destined to become a global problem almost from the time of its first discovery. On a more positive note, SARS was contained through the use of international surveillance; and surveillance was the backbone of the smallpox eradication program.

International surveillance can also be used for the detection of international outbreaks of food poisoning caused by the distribution of foodstuffs across a wide number of countries. An outbreak of hepatitis A in England was caused by frozen raspberries grown and frozen in another country; and another outbreak of hepatitis A, this time in Czechoslovakia (before it became separate republics) was caused by strawberries used to make ice cream; the strawberries had been imported from another Eastern European country. There are now well-established trans-European surveillance systems for salmonella infections and legionnaires' disease, as well as for viral infections.

The need for surveillance will never diminish or disappear. Surveillance systems will only improve, become increasingly sophisticated, and become increasingly relied upon and used. Control of infection will not be possible without it.

Further Reading

- Charrel, R.M., de Lamballerie, X., Raoult, D., 2007. Chikungunya outbreaks – The globalization of vectorborne diseases. *New England Journal of Medicine* 356, 769–771.
- Chin, J. (Ed.), 2002. *Control of Communicable Diseases Manual*, seventeenth ed. Washington, DC: APHA.
- Chorba, T.L., 2001. Chapter 7 – Disease surveillance. In: Thomas, J.C., Weber, D.J. (Eds.), *Epidemiologic Methods for the Study of Infectious Diseases*. Oxford: Oxford University Press.
- Colon-Gonzalez, F.J., Lake, I.R., Morbey, R.A., *et al.*, 2018. A methodological framework for the evaluation of syndromic surveillance systems: A case study of England. *BMC Public Health* 18, 544. doi:10.1186/s12889-018-5422-9.
- Communicable Disease Report, 1994. National measles and rubella immunisation campaign. *Communicable Disease Report Weekly* 4 (31), 146–150.
- Heymann, D., Rodier, G.R., 1998. Global surveillance of communicable diseases. *Emerging Infectious Diseases* 4, 362–365.
- McCormick, A., 1989. Estimating the size of the HIV epidemic by using mortality data. accessed date: 05.08.89. doi:10.1098/rstb.1989.0081.
- Noah, N., 1989. Cyclical patterns and predictability in infection. *Epidemiology & Infection* 102 (2).
- Noah, N., 2006. *Controlling Communicable Disease*. Maidenhead, UK: Open University Press, pp. 14–19. (chs. 1–4).
- UNAIDS, 2003. Introduction to second generation HIV surveillance. Available at: http://www.data.unaids.org/Publications/IRC-pub03/2nd_generation_en.ppt (accessed September 2007).
- Thomas, M.E.M., Noah, N.D., Tillett, H.E., 1974. Recurrent gastroenteritis in a preparatory school caused by *Shigella sonnei* and another agent. *Lancet* 1, 978–981.

Relevant Websites

- <http://www.cdc.gov>
Centers for Disease Control and Prevention (CDC).
- <http://www.eiss.org>
EIS Scotland's Largest Teaching Union.
- <http://www.eurosurveillance.org>
Eurosurveillance.
- <https://iser.med.unsw.edu.au/epi-watch-outbreak-alerts>
NHMRC Centre for Research Excellence Integrated Systems for Epidemic Response.
- <http://www.hpa.org.uk>
Public Health England – GOV.UK.
- <http://www.EWGLI.org>
Service unavailable.
- <https://www.thedailybeast.com/were-fighting-polio-with-sewage-surveillance>
We're Fighting Polio With Sewage Surveillance.
- <http://www.who.int>
World Health Organization: WHO.

Preparing for Emerging Zoonotic Viruses

Reina S Sikkema and Marion PG Koopmans, Erasmus Medical Center, Rotterdam, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Emerging infectious diseases (EID) have been defined as diseases whose incidence increased over the past decades or are predicted to increase in the foreseeable future ("see Relevant Websites section"). This definition includes known infections with new properties, or known infections in new geographic regions, infections that were not previously recognized, and new infections resulting from spillover of pathogens from an animal reservoir to humans. Although difficult to state with certainty, current consensus is that EID outbreaks have increased significantly in the past decades, both in terms of size and number of causative pathogens (Smith *et al.*, 2014; Allen *et al.*, 2017). The majority of EIDs are thought to originate from animals, of which wildlife is the most important source of human outbreaks (Allen *et al.*, 2017; Jones *et al.*, 2008). Such spillovers can occur directly, or through vectors such as mosquitos, ticks and sandflies. Examples of past outbreaks are SARS (Peiris *et al.*, 2004), MERS (Zaki *et al.*, 2012), Avian Influenza (Lai *et al.*, 2016), Ebola (Dudas *et al.*, 2017) and Zika virus (Gubler *et al.*, 2017). Compared to common endemic diseases, the burden of disease may be relatively limited, but the unexpected nature, high case fatality rate, the uncertainty of sources and modes of transmission, and the paucity of medical and non-medical countermeasures make EIDs a threat to global human and animal health. Fear of spread, travel restrictions, and unpredictable self-imposed avoidance of a region can cause serious socio-economic disruptions on local and global level (McCloskey *et al.*, 2014).

Drivers of (Emerging) Zoonotic Diseases

Infectious disease emergence has proven extremely challenging to predict, and human cases of a wide range of emerging or re-emerging pathogens have been reported all over the world and throughout the year. However, when outbreak data of emerging disease outbreaks in the last decades are combined, some specific conditions or risk factors seem to drive disease emergence (Woolhouse, 2011). The main general drivers that have been described can be summarized as changes in (1) human demographics with the consequential growing demand for food production (Jones *et al.*, 2008), (2) land use, including agricultural changes (Jones *et al.*, 2013; Gortazar *et al.*, 2014) (3) international travel and trade (Gortazar *et al.*, 2014; Randolph and Rogers, 2010), (4) climate change and weather (Allen *et al.*, 2017; Gortazar *et al.*, 2014) (Fig. 1).



Fig. 1 Global changes acting as drivers of infectious disease emergence and spread in the One Health domains.

In the case of zoonotic EIDs, the first step towards an outbreak of human disease or mortality is the occurrence of spill-over of a pathogen from an animal reservoir host to humans. (Plowright *et al.*, 2017; Karesh *et al.*, 2012; Petersen *et al.*, 2018). The majority of human EIDs originate from wildlife (Jones *et al.*, 2008; Allen *et al.*, 2017). However, humans are usually not in close contact with wildlife. Therefore, occasionally there is an additional intermediate host species involved where a pathogen is amplified and that is in closer contact with humans (Karesh *et al.*, 2012; Plowright *et al.*, 2015). Well known examples are Nipah virus, Hendra virus or SARS-CoV, that originate from bats and amplify in pigs, horses and civet cats used as food source, respectively (Plowright *et al.*, 2015). The risk of animal to human spill-over depends on a plethora of factors. The aforementioned drivers of disease emergence create circumstances that facilitate such spill-over. Additionally, the risk of spillover is influenced by the pathogen itself. In general, RNA viruses are zoonotic more often than DNA viruses (Olival *et al.*, 2017). A possible explanation for this is their genetic plasticity, caused by the replication mechanism: most RNA viruses replicate with the use of virus-encoded RNA polymerases which – in contrast to the host cell DNA polymerase – do not have so-called proof-reading capacity. Proofreading is a mechanism to reduce the amount of errors in the nascent strands of newly produced copies of (pathogen) genomes, and the lack of proofreading leads to copies that contain a higher number of mutations when comparing the progeny genomes with those of the parent sequences. As a consequence, hosts infected with such RNA viruses shed a mutant swarm of viruses. When they then encounter a bottleneck, for instance an exposure of a new host, chances are that within the swarm variants exist that are better adapted (by chance) to infect a new host (Domingo and Perales, 2019).

One possible consequence of this flexibility is described as “phylogenetic host breadth” and “host plasticity”, which reflect the diversity of known hosts, and therefore are significant predicting factors for zoonotic potential of viruses (Olival *et al.*, 2017; Kreuder Johnson *et al.*, 2015). Additionally, the dynamics of a pathogen in a reservoir host population is a very important factor for human health risk, and is governed by population structure, host density, host behavior and contact patterns, shedding patterns, stability of pathogens outside a host, levels of immunity, events that stress the population, and more. As a consequence, the resulting risk of infection for humans may vary depending on season, location, weather and host distribution for example (Plowright *et al.*, 2017; Karesh *et al.*, 2012).

The second step towards outbreaks of emerging diseases is the ability of a pathogen to transmit between humans. In general, it is thought to be less likely that enveloped viruses and segmented viruses are associated with human-to-human transmission, similar to viruses with acute durations of infection but the evidence for that is largely observational and certainly not black and white (Geoghegan *et al.*, 2016; Walker *et al.*, 2018). For instance, the majority of top ranking epidemic and pandemic threats listed by the world health organization (WHO) are enveloped viruses (“see Relevant Websites section”). Viral host plasticity is not only associated with spill-over but also increases the risk of human-to-human transmission and international spread (Kreuder Johnson *et al.*, 2015). Moreover, the increased interconnectivity and the overall increase of the human population facilitate rapid spread of novel pathogens.

Disease X

A recent report by the Global Preparedness Monitoring Board warned that a global Influenza pandemic, similar to the 1918 Spanish flu, could spread around the world within 36–50 h, and would take the lives of 50–80 million people. Such an epidemic could destroy nearly 5% of the global economy (“see Relevant Websites section”). The examples of Ebola, SARS, and avian influenza have shown that zoonotic disease outbreaks can be very disruptive because of fear for spread and uncertainty of how to contain them even when there is limited onward transmission (Bartsch *et al.*, 2015; Qiu *et al.*, 2018). (“see Relevant Websites section”). The devastating Ebola outbreak in West Africa in 2014–2015 led to a specific call for action by the members of the World Health Assembly, charging the WHO to come up with a new vision to be better prepared for the future, given the increasing likelihood of such outbreaks. A list of priority diseases was developed, for which an urgent need for accelerated research and development of countermeasures was identified, due to their potential to cause a public health emergency. (“see Relevant Websites section”). In an update in 2018, the term disease X was used to describe a serious international epidemic caused by a pathogen currently unknown to cause human disease. The rationale behind the inclusion of disease X is that a completely unknown disease etiology limits the ability for development of fast track diagnostics, vaccines and therapeutics. Therefore, preparedness and response in case of disease X requires a significantly different approach compared to the other diseases on the list.

According to the International Health Regulations, WHO Member States are obliged to maintain effective disease surveillance and laboratory systems and to report newly emerging diseases that could spread internationally (“see Relevant Websites section”). However, setting up surveillance for diseases that are still unknown is complicated, and extensive surveillance can be costly, especially in the case of diseases that emerge at the human-animal interface. The disease X track therefore is thought to stimulate thinking in terms of generic solutions that increase EID preparedness.

Emerging Disease Detection and the One Health Concept

A key component of early control of emerging diseases is their early detection (Bonacic Marinovic *et al.*, 2014; Chan *et al.*, 2010). As part of the WHO Blueprint initiative, the need for diagnostics for the priority diseases has been stressed. The West African Ebola virus outbreak in 2014 had already disseminated widely into three countries by the time it was confirmed by an international

laboratory, almost 4 months after the index case succumbed to the disease (Baize *et al.*, 2014; Coltart *et al.*, 2017). Similarly, phylogenetic reconstruction of Zika in the Americas predated the time of introduction to almost two years before it was recognized as a health emergency (Thézé *et al.*, 2018). Although easily criticized, the delays in disease detection are not surprising as symptoms of most EID are not specific, and overlap with major “common” diseases that clinicians are familiar with (Sigfrid *et al.*, 2018). Similarly, clinical diagnostic laboratories focus on these common conditions, and may not have the expertise to detect rare and novel causes of disease (Sigfrid *et al.*, 2018). Efforts to increase preparedness for EID are increasingly moving towards risk-based early detection. As most EID come from animal reservoirs, the need for this has reinvigorated the One Health approach.

The concept of One Health is described by the US Center of Disease Control (CDC) as follows: “One Health recognizes that the health of people is connected to the health of animals and the environment. It is a collaborative, multisectoral, and transdisciplinary approach—working at the local, regional, national, and global levels—with the goal of achieving optimal health outcomes recognizing the interconnection between people, animals, plants, and their shared environment.” Although the concept has already been known since Hippocrates launched his theories on human health around 400 BCE (“On Airs, Waters, and Places”; Aristotle “Historia Animalium”) The specific term “One Health” however only came up in the 21st century (Evans and Leighton, 2014; Sikkema and Koopmans, 2016). Since then, the One Health concept is increasingly being endorsed and implemented in national and international policy and research. An important milestone was the publication of the FAO-OIE-WHO tripartite concept note, on “Collaboration, sharing responsibilities and coordinating global activities to address health risks at the animal-human-ecosystems interfaces” (“see Relevant Websites section”) Also, the number of scientific publications that mention the term “One Health” showed a major increase in the recent decade (Sikkema and Koopmans, 2016). Although in practice true integration of information and collaboration between different sectors still proves to be challenging (Sikkema and Koopmans, 2016; Farag *et al.*, 2019; Dos *et al.*, 2019; Manlove *et al.*, 2016), the implementation of the One Health concept has great potential for EID preparedness and response.

Moreover, there has been a vast increase of (digital) data in the past decade that has great potential to be used for zoonotic disease prediction, early detection and control (Becker *et al.*, 2019; Masri *et al.*, 2019). This is not only caused by the shift of traditional diagnostics to multi-analyte technologies such as next generation sequencing, but also by the large increase of for example environmental, financial and social media information collection (Simonsen *et al.*, 2016; Shi and Wang, 2019). However, the use of this “Big Data” with different data types of multiple origins also makes the analysis much more challenging (Khoury and Ioannidis, 2014). The use of big data has been advocated and explored in some aspects of public health and infectious disease prediction, but this is also mostly restricted to only one of the required domains (clinical, public health, academia) or sectors (human health, animal health, environmental health, informatics, artificial intelligence) instead of a true crosscutting One Health analysis. Moreover, this, again, calls for multidisciplinary collaborations, since the traditional expertise involved in One Health or Public Health, such as veterinary or medical specialists or epidemiologists, generally do not have sufficient experience in working with such large and diverse data sets.

Developing Targeted, Risk Based Sampling

To identify risk areas, or so-called hotspots of disease emergence or spread, both knowledge of possible drivers or risk factors for emergence can be used, as well as detailed spatial and quantitative information on these drivers, for example locations where ecosystem disruptions took place or where farming systems or livestock population sizes have gone through significant changes. Other information, that could be mapped to identify risk areas for known and unknown diseases are vectors, animal reservoirs, human population, habitat and climate. (“see Relevant Websites section”). There are currently many open-access databases available online, that contain information that has great potential to be used for zoonotic disease prediction, early detection and control. Examples of currently available information are: animal tracking data (Tian *et al.*, 2015) (“see Relevant Websites section”; “see Relevant Websites section” United States Geological Survey), animal and human outbreak information (Cauchemez *et al.*, 2014) (WAHID; ADNS; WHO), animal populations and trade (Simons *et al.*, 2016) (FAOSTAT), weather and climate (Ryan *et al.*, 2015) (NASA; ECA&D; USGS FEWS NET), country economic data (Farag *et al.*, 2018) (World Bank) and many others.

After defining geographical hotspot areas, risk of EID outbreaks also requires a detailed understanding of human and animal populations at risk. For example, past research indicates that humans in contact with animals, especially in the case of wildlife or animal populations that experienced changes in husbandry, population size or other characteristics are at a higher risk of infection with an (emerging) zoonotic pathogen (Jones *et al.*, 2008; Jones *et al.*, 2013; Allen *et al.*, 2017). However, the intensity and nature of the human-animal interface differs greatly, depending on animal species, animal husbandry systems, eating habits, traditions, etc. Additionally, with human demography, migration and travel, current populations and communities are no longer homogeneous or isolated. Selecting a human population at-risk therefore is complex and people-at-risk move in an increasingly large geographical area. (“see Relevant Websites section”).

Therefore, there is an increasing need and necessity to tailor surveillance and risk predictions to local circumstances. This calls for the inclusion of social sciences in disease surveillance and infectious disease preparedness (Bedford *et al.*, 2019). (“see Relevant Websites section”).

Animal trading patterns and common practices surrounding animal trading have been shown to be particularly important factors in the spread of zoonotic infectious diseases. Livestock farming has undergone massive changes and expansion in the past decades, with increasing international trade (“see Relevant Websites section”). Increasing size and connectivity of the trading networks, including human contacts such as farmers, traders and consumers, result in increased risks of disease transmission and

spread (“see Relevant Websites section”). Within the poultry production sector, for example, there are large differences between large scale, intensive industrialized production systems and traditional small-scale rural poultry production systems, and associated animal and public health risks. The main type of production system, as well as the organization of the trade chain, and the degree and frequency of human and wildlife contact differs greatly between regions. In South East Asia, the majority of poultry is reared in backyard farms, with complicated subsequent trade chains, involving many traders and live poultry markets (Moyen *et al.*, 2018; Sealy *et al.*, 2019). This influences infectious disease risk and possible spread. Farms that keep their poultry outdoors generally have a higher chance of AI outbreaks, because of their contact with wild birds (Bouwstra *et al.*, 2017; Wibawa *et al.*, 2018). Moreover, trading involving middlemen is associated with higher likelihood of AI spread, as well as density of live-poultry markets (Sealy *et al.*, 2019; Gilbert *et al.*, 2014). Also, it has been shown that the prevalence and risk of infectious diseases, such as AI, differs per stage of the poultry supply chain (Wu *et al.*, 2019). In other regions of the world, such as Europe and the US, backyard farms and live animal markets are very uncommon, and poultry is often transported directly from farm to slaughterhouse. Therefore, risk profiles and set-up of surveillance differ greatly between farming and trading systems in different geographical areas, and detailed information on the local situation is crucial (Martin *et al.*, 2011; Delabougliuse *et al.*, 2017). This is not only important for livestock and poultry, but wildlife and wildlife product trading can also pose a significant risk of infectious disease introduction and spread (“see Relevant Websites section”) (Smith *et al.*, 2012).

Also, habits and behaviors of humans around animals matter. Live bird markets are not only a risk for AI transmission because of mixing of birds of different origins and species but also because of particular preferences in some regions for fresh poultry meat and on-site slaughtering (Zhou *et al.*, 2015). In-depth knowledge of (cultural) habits and their history, has proven to be extremely valuable in outbreak investigations of emerging diseases. Case investigations of a Nipah outbreak in humans in Bangladesh in 2004–2005 involved anthropological expertise that was used for in-depth interviews and questionnaires in local language to gain knowledge on possible exposures and risk behaviors. It became clear that drinking raw date palm sap was significantly associated with disease. Several interviews pointed towards fruit bats (*Pteropus giganteus*) as probable source of infection as they drink from the sap collection pots during the night, contaminating the sap with their infectious saliva. (Luby *et al.*, 2006; Halpin *et al.*, 2011). Infrared cameras captured fruit bats feeding on palm sap collection devices, and urinating when flying away, thus potentially contaminating the sap with viruses (Halpin *et al.*, 2011). Protecting the collection devices by a simple bamboo barrier rather than a ban on drinking palm sap was accepted by the local population as intervention (Nahar *et al.*, 2017).

MERS-CoV is another example of a disease where an inventory of the local situation revealed very specific traditions and behaviors around dromedary camels, the reservoir host of the virus. Camel racing and associated movement and mixing of camels is thought to be an important risk factor for spread (Farang *et al.*, 2018). Moreover, local customs involve kissing camels, and drinking raw camel milk and urine, all of which should be considered in the investigations of MERS-CoV (Farang *et al.*, 2018; Gossner *et al.*, 2016). Only with extensive knowledge about the history and background of specific risk behaviors, a strategy to discourage or reduce such behavior can be developed. Another clear example of high risk behavior that has proven to be very difficult to change, are burial rituals in Western Africa, that greatly increase the number of people infected with Ebola in times of an outbreak (Masumbuko Claude *et al.*, 2019). The family and community members of the deceased person touch and wash the bodies before the funeral, which has caused a large number of secondary infections because of their contact with Ebola infected bodily fluids (Nielsen *et al.*, 2015).

Sample Collection

Appropriate sample collection is essential for reliable surveillance and outbreak investigations. If human, animal or environmental samples are not collected and stored in a technically appropriate manner, subsequent laboratory analyses will not be reliable or will not generate the anticipated results. A deep understanding of kinetics of shedding of a pathogen and of the antibody response is important to decide on the optimal sampling protocol, the use of detection methods and the interpretation of results. Moreover, subject selection (especially in an outbreak situation), appropriate sample type and sufficient and timely metadata collections are all crucial for subsequent analyses. In addition, to be able to compare data between studies or between countries, laboratory analyses, information collection and study designs should be standardized. For example, comparing human serological data to assess zoonotic influenza exposure has proven to be very difficult because there are significant differences between studies when assessing the collection of epidemiological data, laboratory methods used and the study population, amongst other things (Sikkema *et al.*, 2016). Therefore, international organizations and laboratory networks are increasing their efforts to publish guidelines and training materials to stimulate a uniform approach to investigate infectious diseases and increase the comparability of surveillance and research.

Not only diagnostic testing methods and quality are carefully assessed and compared (Reusken *et al.*, 2015; Charrel *et al.*, 2017; Pas *et al.*, 2015) (“see Relevant Websites section”) but also the standardized collection of metadata is increasingly important in the era of big data and NGS, for correct interpretation and analysis of the large amount of data that is generated. Therefore, standardized epidemiological questionnaires are being developed and published alongside of the guidelines for infectious disease diagnosis. In practice, complete and timely collection of metadata proves to be challenging. When routine diagnostic data in the Netherlands was assessed for its use for surveillance of arboviral infection in travelers, researchers found that essential information such as vaccination history was only reported in 0.4%, and travel destination was completed in only 42% of patient dossiers

(Cleton *et al.*, 2014). Especially during outbreaks, medical records and associated epidemiological information is often incomplete or absent. This has been illustrated for example in Sierra Leone during the Ebola outbreak (Owada *et al.*, 2016).

Examples of efforts to standardize infectious disease surveillance and clinical research are the global Consortium for the Standardization of Influenza Seroepidemiology (CONSISE) that is aiming to standardize influenza sero-epidemiology and develop investigation protocols (“see Relevant Websites section”) and the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC), a federation of clinical research networks, providing tools and protocols for research response to outbreak-prone infectious disease (“see Relevant Websites section”). Moreover, the WHO, FAO and OIE publish extensive laboratory diagnosis guidelines and are involved in training and capacity building activities.

The Value of Non-Invasive and Bulk Samples

Although smart, risk-based strategies can be developed to optimize sampling efficiency, reliable disease monitoring, especially in the case of emerging diseases, requires a huge sample size to reach sufficient population and/or geographical representation. Traditional surveillance of humans and animals make use of invasive sampling, to screen blood samples or swabs for pathogens. However, the use of such samples needs medical-ethical approval that is increasingly difficult to obtain, as well as approval for animal population studies and experiments. Additionally, invasive sampling can cause significant discomfort, can pose an additional risk for infection of healthcare workers or animal care-takers and can only be executed by specialists in the human or veterinary field. Therefore, setting up and validating surveillance using non-invasive sampling can potentially be easier to execute and to scale up as well as better for human and animal welfare.

Feces or urine collection and testing can be a practical and animal friendly surveillance method. Especially in the case of wildlife, where catching and sampling animals can be particularly stressful for an animal, fecal or urine surveillance is very suitable. Feces, or even urine, can be collected without handling the animal or even be in the vicinity at the time of sample collection. This has been described for multiple pathogens, such as Lyssaviruses and Paramyxovirus in bats (Begeman *et al.*, 2020; Peel *et al.*, 2019) and tapeworms in foxes (Umhang *et al.*, 2016). In humans, there may also be a large number of pathogens that can be detected via saliva, urine or feces, although it is generally not included in standard diagnostic algorithms (Niedrig *et al.*, 2018).

Although alternative, non-invasive sample types increase the ease of sample collection and welfare of the sampled subject, the number of samples will not decrease. Therefore, validation of pooled samples or novel sample types that can be used to monitor groups or populations of humans, animals or a certain geographical area. In Asia, environmental surveillance of live animal markets is routinely implemented as main means of monitoring of Avian Influenza viruses (Bai *et al.*, 2019; Henning *et al.*, 2019; Khan *et al.*, 2018). Various surfaces, such as poultry cages, chopping boards and defeathering machines, are sampled and tested on a regular basis as an indirect measure of poultry infections and public health risk. Similarly, environmental sampling are also being used for the detection of pathogens in healthcare settings (Kim *et al.*, 2016; Kapetshi *et al.*, 2018). Additionally, air sampling may be promising, both in hospital settings, farms and even markets and airports (Scoizec *et al.*, 2018; Zhou *et al.*, 2016; Bailey *et al.*, 2018).

Sewage or wastewater monitoring is another example of such a sample type, that can be used to monitor infectious diseases of large numbers of humans. Sewage monitoring has already been used for many years to monitor polio (Asghar *et al.*, 2014) as well as the monitoring and detection of emerging Norovirus and Enterovirus strains (Mabasa *et al.*, 2018; Majumdar and Martin, 2018). However, sewage could have great potential to be used for virus discovery and emerging disease detection (Fernandez-Cassi *et al.*, 2018; Nordahl Petersen *et al.*, 2015). Other sample types, representing large numbers of animals, that can be used to monitor circulating of pathogens are: manure composting plants, bulk milk and ponds (Himsworth *et al.*, 2019; Garcia *et al.*, 2014; Balmer *et al.*, 2014).

However, the use of non-invasive sampling or pooled sample types may impact the sensitivity of pathogen detection. In the case of non-invasive sampling, each pathogen may have different shedding characteristics, making it impossible to know the optimal sample type to detect emerging diseases. This also impacts the use of sewage for emerging disease surveillance, for example, because not all pathogens are excreted via feces or urine, making it difficult to detect via sewage sampling.

Catch-All Detection Methods

A challenge for risk-based surveillance is the huge potential diversity of pathogens that could cause human disease when spilling over from an animal reservoir. Therefore, even if targeting detection efforts through the approaches described above would succeed, the choice of diagnostic method is a critical one. Depending on how the risk-based surveillance was organized, the choice of potential pathogens to target may be limited or more extensive. An interesting development is the use of unbiased third-generation metagenomic sequencing to characterize all genomic material (DNA and RNA) in a sample, without prior knowledge of possible etiologies. Clinical and environmental samples can be processed with procedures that either enrich for bacteria, parasites or viruses, and subsequently be subjected to sequencing (Wilson *et al.*, 2019; Takhampunya *et al.*, 2019; Zolfo *et al.*, 2018; Nieuwenhuijse and Koopmans, 2017; Hendriksen *et al.*, 2019).

Analysis generally takes place by assembly of sequence reads and annotating the resulting reads using reference databases of (publicly) known sequences. The metagenomics research field, studying and profiling genetic material abundance in diverse matrices and environments is rapidly growing, as well as the number of novel viruses that are being discovered and described (Simmonds and Aiewsakun, 2018). Often the aim of metagenomics studies of the environment is to map all micro-organisms in

potential reservoirs in order to identify potential human health threats (Carroll *et al.*, 2018). However, all viruses that have been characterized to date likely only make up a minor fraction (estimated to be around 0.005%) of the estimated total number of viruses on earth (Geoghegan and Holmes, 2017). This means that the number of new viruses and virus families will increase immensely in the coming years. Moreover, the lack of closely related reference virus genomes makes public health risk assessment of novel viruses very complicated (Geoghegan and Holmes, 2017). Moreover, even if viruses belong to a known virus family with known human viruses, this does not necessarily mean that there is a risk of spill-over.

From Genotype to Phenotype: Deriving Meaningful Information From Genomic Data

The pathogen genomic sequence alone will often give insufficient information on possible human health risks. Therefore, ideally, metagenomics or NGS approaches are supplemented with phenotypic data such as antigenic characteristics, pathogenicity, virulence, host-specificity and transmission characteristics. For novel variants of known viruses, some indications of phenotype or zoonotic risks can be derived from the sequence, based on experimental study outcomes. The impact of inferring phenotype traits from genotype data can be exemplified by the “H5N1 inventory” compiled by the United States Centers for Disease Control (“see Relevant Websites section”). This inventory summarizes all mutations that have been shown to affect virus replication, virulence, tissue or cellular tropism, host-range, transmission, antigenic properties, immune escape, and antiviral drug resistance.

However, in the case of an emerging virus that was previously unknown, novel phenotypic data needs to be generated. This data is important for risk assessment, but information on potentially relevant sequences or epitopes can also be used for development of diagnostics and preventive antiviral strategies. For example, one of the key components influencing ability to spread for an emerging pathogen from a spill-over event is the level of measured or predicted immunity in the population, as this impacts on the likelihood of infection of individuals (humans, livestock, wildlife) and their ability to shed and spread the infection.

In silico epitope prediction can predict B-cell epitopes based on sequence data (Potocnakova *et al.*, 2016). The predicted immunogenic virus proteins can then be used to set up serological assays to determine the prevalence and host range of the virus. For this, multiplex antigen arrays can be used to not only measure antibodies to the newly identified antigen, but also to (distantly) related variants of the same virus family, thus providing information on potential cross-protection or disease enhancing antibodies (Mina *et al.*, 2019; Cleton *et al.*, 2017; Freidl *et al.*, 2015). Possible resistance markers can be predicted when analyzing the sequences, for instance for drugs targeting the polymerase or protease genes. Host binding motifs can be predicted from their location on predicted protein structures. However, all such inferences need to be validated as the choice of *in vitro* systems and animal models can have a significant impact on their validity (Rothenburg and Brennan, 2020; Setoh *et al.*, 2019; Reyes-Ruiz *et al.*, 2019).

Conclusion

The risk of emergence of novel human pathogens has increased in the past decades. Early detection is becoming even more important due to the potential rapid spread and impact considering the ever increasing size of human and animal populations and their movements all over the globe. The unpredictable nature of emergence of diseases makes surveillance and preparedness a huge challenge. However, recent history has shown that outbreaks of emerging diseases will happen again, and all sectors involved in infectious disease surveillance and response need to be ready to detect and handle the next disease X. It has been argued that the design of infectious disease surveillance and response should therefore move from crisis response to true integration and evaluation of a strategy to detect and control emerging diseases (Bedford *et al.*, 2019). This involves a One Health approach, integrating not only the humans, animal and environmental health sectors, but also social sciences, bioinformatics and more. Technical developments in recent years, such as the use of Big Data and metagenomic sequencing will aid the rapid detection of novel pathogens on the human-animal interface.

Note Added in Proof: December 2020

SARS-CoV-2

In December 2019 several cases of pneumonia of unexplained etiology in patients in Wuhan city (China) linked to a seafood wholesale market were reported to the WHO (“see Relevant Websites section”). In the beginning of 2020 it became clear that this Disease X that had emerged was also a betacoronavirus, most closely related to SARS, that was already on the WHO Blueprint priority list in public health emergency contexts (“see Relevant Websites section”). In the following months the outbreak developed into a pandemic, with over 68 million human cases and 1,55 million deaths, as of 8 December 2020 (Dong *et al.*, 2020).

One Health and Animal Reservoirs

Bats are known to harbor a range of different coronaviruses (Cui *et al.*, 2019). Indeed, closely related SARS-like viruses have previously been detected in horse shoe bats, in China (Zhou *et al.*, 2020; Hu *et al.*, 2018). However, although the SARS-CoV-2 genome is 96% identical to the closest bat sequence, the genetic distance between the closest horseshoe bat SARS-like coronavirus

and SARS-CoV-2 likely reflects several years and even decades of virus evolution (Boni *et al.*, 2020). SARS-related coronaviruses are also found in pangolins, where specifically the RBD domain is closely related to SARS-CoV-2 (Boni *et al.*, 2020). Therefore, the role of bats as SARS-CoV-2 reservoir remains to be proven and other mammalian species are considered as possible intermediate hosts, as was previously the case for SARS, MERS and Nipah (Epstein *et al.*, 2006; Wang *et al.*, 2005; Reusken *et al.*, 2013). To date, an animal intermediate host of SARS-CoV-2 has not been found yet (“see Relevant Websites section”).

However, multiple animal species are found to be susceptible for SARS-CoV-2. Ferrets, raccoon dogs, cats and hamsters show infection and virus replication after experimental infection as well as the ability to transmit the virus (Shi *et al.*, 2020; Richard *et al.*, 2020; Sia *et al.*, 2020; Freuling *et al.*, 2020). Other animals that showed infection after experimental inoculation include rabbits, dogs and tree shrews (Shi *et al.*, 2020; Mykityn *et al.*, 2020; Bosco-Lauth *et al.*, 2020). Additionally, animal infections have been detected in the field, mainly in dogs, cats and mink (Sit *et al.*, 2020; Barrs *et al.*, 2020). Although there is evidence for efficient transmission between cats, infection in these animals is not considered a major concern given the small numbers of animals in households. This is different for the recent outbreaks in mink farms in Europe and the US (Cahan, 2020), with the first described SARS-CoV-2 animal-to-human transmissions (Oude Munnink *et al.*, 2020b). These outbreaks were the first large scale outbreaks in animal populations, and large scale mink passages have specifically sparked concern on the risk of specific mutations, affecting virus properties and possibly resulting in immune escape (Koopmans, 2020).

The multiple susceptible animal species and the search for the original animal reservoir, that was at the origin of the SARS-CoV-2 pandemic, clearly shows the need for a One Health approach. Introductions of SARS-CoV-2 into susceptible animal populations can continue to occur, possibly resulting in (additional) animal reservoirs and virus evolution and the risk of continuing evolution and spill backs into the human population (Koopmans, 2020; Olival *et al.*, 2020). Therefore, collaboration between different disciplines remains crucial in monitoring, understanding and containing the SARS-CoV-2 pandemic as well as many other (emerging) pathogens.

Emerging Disease Detection, Full Genome Sequencing and Rapid Data Sharing

The Chinese authorities officially announced the isolation of a novel coronavirus (2019-nCoV) as the causative agent on 7 January 2020 (“see Relevant Websites section”) and the full sequence was released in a publicly accessible discussion forum (“see Relevant Websites section”) 5 days later (“see Relevant Websites section”). Four other sequences were shared on 12 January in the Global Initiative on Sharing All Influenza Data (GISAID) database. A validated real-time RT-PCR assay was developed, validated and published online only 5 days later, on 17 January 2020, and published for PCR validation in a peer-reviewed journal on the 23rd of January (“see Relevant Websites section”) (Corman *et al.*, 2020). Validation panels were available on the European Virus Archive (“see Relevant Websites section”) shortly after. The first validated serological assay followed early April (Okba *et al.*, 2020). This clearly illustrates the technical progress that has been made in the last 17 years, as well as the value of virus sequencing and rapid data and sample sharing.

Samples of the first patients in Wuhan city were sequenced using an unbiased deep meta-transcriptomic sequencing method (Wu *et al.*, 2020). When the sequence was known, amplicon-based methods were developed to sequence the new virus in a timely and sensitive manner ((Oude Munnink *et al.*, 2020a) “see Relevant Websites section”) Whole genome sequencing was performed at unprecedented scale, with almost 250.000 SARS-CoV-2 whole genome sequences that had been shared on the GISAID platform as of 8 December 2020. The timely release of unpublished genomic information through GISAID was essential for fast track development of diagnostics, vaccines and therapeutics as well as the application of WGS in outbreak investigation and public health control measures (Oude Munnink *et al.*, 2020a; Voeten *et al.*, 2020).

With the ever-increasing numbers of sequences that are being generated, the need for rapid characterization of viruses with specific mutations remains high. A high number of papers are published describing the evolution and mutations in the SARS-CoV-2 genome, but assessing the relevance and effect on virus properties, and, consequently, public health, is still challenging (Grubaugh *et al.*, 2020).

Population Surveys and Sero-epidemiological Studies

Serologic studies can provide important information for understanding the extent of past transmission, the current state of the epidemic, and future transmission in an affected population. Moreover, serology is a very useful tool on the human-animal interface (Wernike *et al.*, 2020). It can be used to screen potential animal reservoirs, as has been used to determine the likely reservoir species for MERS-CoV, (Reusken *et al.*, 2013; Deng *et al.*, 2020). Serological surveys in several animal species did not point towards the SARS-CoV-2 animal reservoir species, but did show significant exposure of dogs and cats in Wuhan and Italy, which were severely affected by SARS-CoV-2 (Zhang *et al.*, 2020; Patterson *et al.*, 2020). Also, serological assays can be an additional tool to gain insight in the potential and scale of animal-to-human transmission, when combined with epidemiological information on exposure, as was done for mink farm employees on SARS-CoV-2 on infected mink farms (Oude Munnink *et al.*, 2020b; Sikkema *et al.*, 2016). However, it is important to realize that coronaviruses including betacoronaviruses are common in many animal species, and can result in cross reactivity and false positives (Franzo *et al.*, 2020; Nemoto *et al.*, 2019; Erles and Brownlie, 2008).

In addition to patient diagnostics and surveillance efforts in symptomatic and asymptomatic individuals, wastewater sampling can be used to monitor the circulation of SARS-CoV-2 in the population (Medema *et al.*, 2020; Hart and Halden, 2020). This approach is currently integrated in the national COVID-19 surveillance in multiple counties (“see Relevant Websites section”).

Concluding Remarks

With the emergence and worldwide spread of a novel coronavirus, the disease X scenario that many scientists had warned about came true. The SARS-CoV-2 pandemic led to drastic government measures worldwide, with massive economic and social consequences. The spread of the virus has proven to be difficult to contain, despite extensive lockdowns, information campaigns, travel bans and compulsory mouth masks. However, previous investments in surveillance, diagnostics, novel laboratory techniques, open data sharing and vaccine platforms did pay off. A cluster of pneumonia of unknown origin was detected in December 2019 in Wuhan, and at the end of January 2020, a set of whole genomes are publicly shared and a validated PCR assay has been set up which was implemented in 24 of 30 EU/EEA countries and many more worldwide (Reusken *et al.*, 2020). Moreover, in December 2020, the first COVID-19 vaccine was licensed ("see Relevant Websites section"). The speed of these key developments in the surveillance and control of the new virus is unprecedented. How the SARS-CoV-2 pandemic will evolve remains to be seen, but recent developments reinforce the call for preparedness research and a One Health approach, as highlighted in the current article.

References

- Allen, T., *et al.*, 2017. Global hotspots and correlates of emerging zoonotic diseases. *Nature Communications* 8 (1), 1124.
- Asghar, H., *et al.*, 2014. Environmental surveillance for polioviruses in the Global Polio Eradication Initiative. *The Journal of Infectious Diseases* 210 (Suppl 1), S294–S303.
- Bai, R., *et al.*, 2019. Antigenic variation of avian influenza A(H5N6) viruses, Guangdong Province, China, 2014–2018. *Emerging Infectious Diseases* 25 (10), 1932–1945.
- Bailey, E.S., *et al.*, 2018. Molecular surveillance of respiratory viruses with bioaerosol sampling in an airport. *Tropical Diseases, Travel Medicine and Vaccines* 4, 11.
- Baize, S., *et al.*, 2014. Emergence of Zaire Ebola virus disease in Guinea. *The New England Journal of Medicine* 371 (15), 1418–1425.
- Balmer, S., *et al.*, 2014. Serosurveillance of Schmallenberg virus in Switzerland using bulk tank milk samples. *Preventive Veterinary Medicine* 116 (4), 370–379.
- Barrs, V.R., Peiris, M., Tam, K.W.S., *et al.*, 2020. SARS-CoV-2 in Quarantined Domestic Cats from COVID-19 Households or Close Contacts, Hong Kong, China. *Emerging Infectious Diseases* 26 (12), 3071–3074.
- Bartsch, S.M., Gorham, K., Lee, B.Y., 2015. The cost of an Ebola case. *Pathogens and Global Health* 109 (1), 4–9.
- Becker, D.J., *et al.*, 2019. Dynamic and integrative approaches to understanding pathogen spillover. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374 (1782), 20190014.
- Bedford, J., *et al.*, 2019. A new twenty-first century science for effective epidemic response. *Nature* 575 (7781), 130–136.
- Begeman, L., *et al.*, 2020. Faeces as a novel material to estimate lyssavirus prevalence in bat populations. *Zoonoses and Public Health* 67, 198–202.
- Bonacic Marinovic, A.A., *et al.*, 2014. Speed versus coverage trade off in targeted interventions during an outbreak. *Epidemics* 8, 28–40.
- Boni, M.F., Lemey, P., Jiang, X., *et al.*, 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature Microbiology* 5 (11), 1408–1417.
- Bosco-Lauth, A.M., Hartwig, A.E., Porter, S.M., *et al.*, 2020. Experimental infection of domestic dogs and cats with SARS-CoV-2: Pathogenesis, transmission, and response to reexposure in cats. *Proceedings of the National Academy of Sciences of the United States of America* 117 (42), 26382–26388.
- Bouwstra, R., *et al.*, 2017. Risk for low pathogenicity avian influenza virus on poultry farms, the Netherlands, 2007–2013. *Emerging Infectious Diseases* 23 (9), 1510–1516.
- Cahan, E., 2020. COVID-19 hits US mink farms after ripping through Europe. *Science*. 10.
- Carroll, D., *et al.*, 2018. The Global Virome Project. *Science* 359 (6378), 872–874.
- Cauchemez, S., *et al.*, 2014. Middle East respiratory syndrome coronavirus: Quantification of the extent of the epidemic, surveillance biases, and transmissibility. *Lancet Infectious Diseases* 14 (1), 50–56.
- Chan, E.H., *et al.*, 2010. Global capacity for emerging infectious disease detection. *Proceedings of the National Academy of Sciences of the United States of America* 107 (50), 21701–21706.
- Charrel, R., *et al.*, 2017. Variable sensitivity in molecular detection of Zika virus in European expert laboratories: External quality assessment, November 2016. *Journal of Clinical Microbiology* 55 (11), 3219–3226.
- Cleton, N., *et al.*, 2014. Using routine diagnostic data as a method of surveillance of arboviral infection in travellers: A comparative analysis with a focus on dengue. *Travel Medicine and Infectious Disease* 12 (2), 159–166.
- Cleton, N.B., *et al.*, 2017. A serological protein microarray for detection of multiple cross-reactive flavivirus infections in horses for veterinary and public health surveillance. *Transboundary and Emerging Diseases* 64 (6), 1801–1812.
- Coltart, C.E.M., *et al.*, 2017. The Ebola outbreak, 2013–2016: Old lessons for new epidemics. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 372 (1721), 20160297.
- Corman, V.M., Landt, O., Kaiser, M., *et al.*, 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25 (3).
- Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* 17 (3), 181–192.
- Delabougli, A., *et al.*, 2017. Cultural practices shaping zoonotic diseases surveillance: The case of highly pathogenic avian influenza and Thailand native chicken farmers. *Transboundary and Emerging Diseases* 64 (4), 1294–1305.
- Deng, J., Jin, Y., Liu, Y., *et al.*, 2020. Serological survey of SARS-CoV-2 for experimental, domestic, companion and wild animals excludes intermediate hosts of 35 different species of animals. *Transboundary and Emerging Diseases* 67 (4), 1745–1749.
- Domingo, E., Perales, C., 2019. Viral quasispecies. *PLoS Genetics* 15 (10), e1008271.
- Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infectious Diseases* 20 (5), 533–534.
- Dos, S.R.C., van de Burgwal, L.H.M., Regeer, B.J., 2019. Overcoming challenges for designing and implementing the One Health approach: A systematic review of the literature. *One Health* 7, 100085.
- Dudas, G., *et al.*, 2017. Virus genomes reveal factors that spread and sustained the Ebola epidemic. *Nature* 544 (7650), 309–315.
- Epstein, J.H., Field, H.E., Luby, S., Pulliam, J.R., Daszak, P., 2006. Nipah virus: Impact, origins, and causes of emergence. *Current Infectious Disease Reports* 8 (1), 59–65.
- Erls, K., Brownlie, J., 2008. Canine respiratory coronavirus: An emerging pathogen in the canine infectious respiratory disease complex. *Veterinary Clinics of North America: Small Animal Practice* 38 (4), 815–825. viii.
- Evans, B.R., Leighton, F.A., 2014. A history of One Health. *Revue scientifique et technique* 33 (2), 413–420.
- Farag, E.A.B., *et al.*, 2019. Survey on implementation of One Health approach for MERS-CoV preparedness and control in Gulf Cooperation Council and Middle East countries. *Emerging Infectious Diseases* 25 (3), e171702.
- Farag, E., *et al.*, 2018. Drivers of MERS-CoV Emergence in Qatar. *Viruses* 11 (1), 22.
- Fernandez-Cassi, X., *et al.*, 2018. Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. *Science of the Total Environment* 618, 870–880.
- Franzo, G., Drigo, M., Legnardi, M., *et al.*, 2020. Bovine Coronavirus: Variability, Evolution, and Dispersal Patterns of a No Longer Neglected Betacoronavirus. *Viruses* 12 (11),

- Freidl, G.S., *et al.*, 2015. Serological evidence of influenza A viruses in frugivorous bats from Africa. *PLoS One* 10 (5), e0127035.
- Freuling, C.M., Breithaupt, A., Müller, T., *et al.*, 2020. Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection. *Emerging Infectious Diseases* 26 (12), 2982–2985.
- García, M., Fernandez-Barredo, S., Perez-Gracia, M.T., 2014. Detection of hepatitis E virus (HEV) through the different stages of pig manure composting plants. *Microbial Biotechnology* 7 (1), 26–31.
- Geoghegan, J.L., *et al.*, 2016. Virological factors that increase the transmissibility of emerging human viruses. *Proceedings of the National Academy of Sciences of the United States of America* 113 (15), 4170–4175.
- Geoghegan, J.L., Holmes, E.C., 2017. Predicting virus emergence amid evolutionary noise. *Open Biology* 7, 10.
- Gilbert, M., *et al.*, 2014. Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nature Communications* 5, 4116.
- Gortazar, C., *et al.*, 2014. Crossing the Interspecies Barrier: opening the Door to Zoonotic Pathogens. *PLOS Pathogens* 10 (6), e1004129.
- Gossner, C., *et al.*, 2016. Human-dromedary camel interactions and the risk of acquiring zoonotic Middle East respiratory syndrome coronavirus infection. *Zoonoses Public Health* 63 (1), 1–9.
- Grubaugh, N.D., Hanage, W.P., Rasmussen, A.L., 2020. Making Sense of Mutation: What D614G Means for the COVID-19 Pandemic Remains Unclear. *Cell* 182 (4), 794–795.
- Gubler, D.J., Vasilakis, N., Musso, D., 2017. History and emergence of Zika virus. *The Journal of Infectious Diseases* 216 (suppl_10), S860–S867.
- Halpin, K., *et al.*, 2011. Pteropid bats are confirmed as the reservoir hosts of henipaviruses: A comprehensive experimental study of virus transmission. *The American Journal of Tropical Medicine and Hygiene* 85 (5), 946–951.
- Hart, O.E., Halden, R.U., 2020. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and challenges. *Science of the Total Environment* 730, 138875.
- Hendriksen, R.S., *et al.*, 2019. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nature Communications* 10 (1), 1124.
- Henning, J., *et al.*, 2019. Risk factors for H5 avian influenza virus prevalence on urban live bird markets in Jakarta, Indonesia - Evaluation of long-term environmental surveillance data. *PLoS One* 14 (5), e0216984.
- Himsworth, C.G., *et al.*, 2019. Targeted resequencing of wetland sediment as a tool for avian influenza virus surveillance. *Journal of Wildlife Diseases* 56, 397–408.
- Hu, D., Zhu, C., Ai, L., *et al.*, 2018. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. *Emerging Microbes & Infections* 7 (1), 154.
- Jones, K.E., *et al.*, 2008. Global trends in emerging infectious diseases. *Nature* 451 (7181), 990–993.
- Jones, B.A., *et al.*, 2013. Zoonosis emergence linked to agricultural intensification and environmental change. *Proceedings of the National Academy of Sciences of the United States of America* 110 (21), 8399–8404.
- Kapetshi, J., *et al.*, 2018. Contribution of environment sample-based detection to Ebola outbreak management. *The Journal of Infectious Diseases* 218 (suppl_5), S292–S296.
- Karesh, W.B., *et al.*, 2012. Ecology of zoonoses: Natural and unnatural histories. *The Lancet* 380 (9857), 1936–1945.
- Koopmans, M., 2020. SARS-CoV-2 and the human-animal interface: Outbreaks on mink farms. *Lancet Infectious Diseases*.
- Khan, S.U., *et al.*, 2018. Avian influenza surveillance in domestic waterfowl and environment of live bird markets in Bangladesh, 2007–2012. *Scientific Reports* 8 (1), 9396.
- Khoury, M.J., Ioannidis, J.P.A., 2014. Medicine. Big data meets public health. *Science* 346 (6213), 1054–1055.
- Kim, S.H., *et al.*, 2016. Extensive viable Middle East Respiratory Syndrome (MERS) coronavirus contamination in air and surrounding environment in MERS isolation wards. *Clinical Infectious Diseases* 63 (3), 363–369.
- Kreuder Johnson, C., *et al.*, 2015. Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Scientific Reports* 5, 14830.
- Lai, S., *et al.*, 2016. Global epidemiology of avian influenza A H5N1 virus infection in humans, 1997–2015: A systematic review of individual case data. *Lancet Infectious Diseases* 16 (7), e108–e118.
- Luby, S.P., *et al.*, 2006. Foodborne transmission of Nipah virus, Bangladesh. *Emerging Infectious Diseases* 12 (12), 1888–1894.
- Mabasa, V.V., *et al.*, 2018. Environmental surveillance for noroviruses in selected South African wastewaters 2015–2016: Emergence of the Novel GII.17. *Food and Environmental Virology* 10 (1), 16–28.
- Majumdar, M., Martin, J., 2018. Detection by direct next generation sequencing analysis of emerging enterovirus D68 and C109 strains in an environmental sample from Scotland. *Frontiers in Microbiology* 9, 1956.
- Manlove, K.R., *et al.*, 2016. "One Health" or Three? Publication silos among the One Health disciplines. *PLoS Biology* 14 (4), e1002448.
- Martin, V., *et al.*, 2011. Risk-based surveillance for avian influenza control along poultry market chains in South China: The value of social network analysis. *Preventive Veterinary Medicine* 102 (3), 196–205.
- Masri, S., *et al.*, 2019. Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic. *BMC Public Health* 19 (1), 761.
- Masumbuko Claude, K., Underschultz, J., Hawkes, M.T., 2019. Social resistance drives persistent transmission of Ebola virus disease in Eastern Democratic Republic of Congo: A mixed-methods study. *PLoS One* 14 (9), e0223104.
- McCloskey, B., *et al.*, 2014. Emerging infectious diseases and pandemic potential: Status quo and reducing risk of global spread. *The Lancet Infectious Diseases* 14 (10), 1001–1010.
- Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., Brouwer, A., 2020. Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters*. doi:10.1101/acs.estlett.0c00357.
- Mina, M.J., *et al.*, 2019. Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. *Science* 366 (6465), 599–606.
- Moyen, N., *et al.*, 2018. A large-scale study of a poultry trading network in Bangladesh: Implications for control and surveillance of avian influenza viruses. *BMC Veterinary Research* 14 (1), 12.
- Mykytyn, A.Z., Lamers, M.M., Okba, N.M.A., *et al.*, 2020. Susceptibility of rabbits to SARS-CoV-2. *bioRxiv*. doi:10.1101/2020.08.27.263988.
- Nahar, N., *et al.*, 2017. A controlled trial to reduce the risk of human Nipah virus exposure in Bangladesh. *Ecohealth* 14 (3), 501–517.
- Nemoto, M., Schofield, W., Cullinane, A., 2019. The First Detection of Equine Coronavirus in Adult Horses and Foals in Ireland. *Viruses* 11 (10).
- Niedrig, M., *et al.*, 2018. Find the right sample: A study on the versatility of saliva and urine samples for the diagnosis of emerging viruses. *BMC Infectious Diseases* 18 (1), 707.
- Nielsen, C.F., *et al.*, 2015. Improving burial practices and cemetery management during an Ebola virus disease epidemic - Sierra Leone, 2014. *MMWR Morbidity and Mortality Weekly Report* 64 (1), 20–27.
- Nieuwenhuijse, D.F., Koopmans, M.P., 2017. Metagenomic sequencing for surveillance of food- and waterborne viral diseases. *Frontiers in Microbiology* 8, 230.
- Nordahl Petersen, T., *et al.*, 2015. Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance. *Scientific Reports* 5, 11444.
- Okba, N.M.A., Müller, M.A., Li, W., *et al.*, 2020. Severe Acute Respiratory Syndrome Coronavirus 2-Specific Antibody Responses in Coronavirus Disease Patients. *Emerging Infectious Diseases* 26 (7), 1478–1488.
- Olival, K.J., *et al.*, 2017. Host and viral traits predict zoonotic spillover from mammals. *Nature* 546 (7660), 646–650.
- Olival, K.J., Cryan, P.M., Amman, B.R., *et al.*, 2020. Possibility for reverse zoonotic transmission of SARS-CoV-2 to free-ranging wildlife: A case study of bats. *PLOS Pathogens* 16 (9), e1008758.
- Oude Munnink, B.B., Nieuwenhuijse, D.F., Stein, M., *et al.*, 2020a. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nature Medicine* 26 (9), 1405–1410.
- Oude Munnink, B.B., Sikkema, R.S., Nieuwenhuijse, D.F., *et al.*, 2020b. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*.
- Owada, K., *et al.*, 2016. Epidemiological data management during an outbreak of Ebola virus disease: Key issues and observations from Sierra Leone. *Frontiers in Public Health* 4, 163.
- Pas, S.D., *et al.*, 2015. First international external quality assessment of molecular diagnostics for MERS-CoV. *Journal of Clinical Virology* 69, 81–85.

- Patterson, E.I., Elia, G., Grassi, A., *et al.*, 2020. Evidence of exposure to SARS-CoV-2 in cats and dogs from households in Italy. *Nature Communications* 11 (1), 6231.
- Peel, A.J., *et al.*, 2019. Synchronous shedding of multiple bat paramyxoviruses coincides with peak periods of Hendra virus spillover. *Emerging Microbes & Infections* 8 (1), 1314–1323.
- Peiris, J.S., Guan, Y., Yuen, K.Y., 2004. Severe acute respiratory syndrome. *Nature Medicine* 10 (12 Suppl), S88–S97.
- Petersen, E., *et al.*, 2018. Emerging infections – An increasingly important topic: Review by the Emerging Infections Task Force. *Clinical Microbiology and Infection* 24 (4), 369–375.
- Plowright, R.K., *et al.*, 2017. Pathways to zoonotic spillover. *Nature Reviews Microbiology* 15, 502–510.
- Plowright, R.K., *et al.*, 2015. Ecological dynamics of emerging bat virus spillover. *Proceedings of the Royal Society B: Biological Sciences* 282 (1798), 20142124.
- Potocnaka, L., Bhide, M., Pulzova, L.B., 2016. An introduction to B-cell epitope mapping and in silico epitope prediction. *Journal of Immunology Research* 2016, 6760830.
- Qiu, W., *et al.*, 2018. The impacts on health, society, and economy of SARS and H7N9 outbreaks in China: A case comparison study. *Journal of Environmental and Public Health* 2018, 2710185.
- Randolph, S.E., Rogers, D.J., 2010. The arrival, establishment and spread of exotic diseases: Patterns and predictions. *Nature Reviews Microbiology* 8 (5), 361–371.
- Reusken, C.B., Haagmans, B.L., Müller, M.A., *et al.*, 2013. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: A comparative serological study. *Lancet Infectious Diseases* 13 (10), 859–866.
- Reusken, C., *et al.*, 2015. Identification of essential outstanding questions for an adequate European laboratory response to Ebolavirus Zaire West Africa 2014. *Journal of Clinical Virology* 62, 124–134.
- Reusken, C., Broberg, E.K., Haagmans, B., *et al.*, 2020. Laboratory readiness and response for novel coronavirus (2019-nCoV) in expert laboratories in 30 EU/EEA countries, January 2020. *Eurosurveillance* 25 (6).
- Reyes-Ruiz, J.M., *et al.*, 2019. Mosquito cells persistently infected with dengue virus produce viral particles with host-dependent replication. *Virology* 531, 1–18.
- Rotherburg, S., Brennan, G., 2020. Species-specific host-virus interactions: Implications for viral host range and virulence. *Trends in Microbiology* 28 (1), 46–56.
- Richard, M., Kok, A., de Meulder, D., *et al.*, 2020. SARS-CoV-2 is transmitted via contact and via the air between ferrets. *Nature Communications* 11 (1), 3496.
- Ryan, S.J., *et al.*, 2015. Mapping physiological suitability limits for malaria in Africa under climate change. *Vector-Borne and Zoonotic Diseases* 15 (12), 718–725.
- Scozec, A., *et al.*, 2018. Airborne detection of H5N8 highly pathogenic avian influenza virus genome in poultry farms, France. *Frontiers in Veterinary Science* 5, 15.
- Sealy, J.E., *et al.*, 2019. Poultry trading behaviours in Vietnamese live bird markets as risk factors for avian influenza infection in chickens. *Transboundary and Emerging Diseases* 66 (6), 2507–2516.
- Setoh, Y.X., *et al.*, 2019. Determinants of Zika virus host tropism uncovered by deep mutational scanning. *Nature Microbiology* 4 (5), 876–887.
- Sia, S.F., Yan, L.-M., Chin, A.W.H., *et al.*, 2020. Pathogenesis and transmission of SARS-CoV-2 in golden hamsters. *Nature* 583 (7818), 834–838.
- Shi, J., Wen, Z., Zhong, G., *et al.*, 2020. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* 368 (6494), 1016–1020.
- Shi, L., Wang, Z., 2019. Computational strategies for scalable genomics analysis. *Genes* 10 (12), 1017.
- Sigfrid, L., *et al.*, 2018. Strengthening preparedness for (re-) emerging arboviruses in Europe. *Clinical Microbiology and Infection* 24 (3), 219–220.
- Sikkema, R.S., Freidl, G.S., de Bruin, E., Koopmans, M., 2016. Weighing serological evidence of human exposure to animal influenza viruses – A literature review. *Euro surveillance* 21 (44).
- Sikkema, R.S., *et al.*, 2016. Weighing serological evidence of human exposure to animal influenza viruses – A literature review. *Eurosurveillance* 21 (44), 30388.
- Sikkema, R., Koopmans, M., 2016. One Health training and research activities in Western Europe. *Infection Ecology & Epidemiology* 6, 33703.
- Simmonds, P., Aiewsakun, P., 2018. Virus classification – Where do you draw the line? *Archives of Virology* 163 (8), 2037–2046.
- Simonsen, L., *et al.*, 2016. Infectious disease surveillance in the big data era: Towards faster and locally relevant systems. *The Journal of Infectious Diseases* 214 (suppl_4), S380–S385.
- Simons, R.R.L., *et al.*, 2016. A generic quantitative risk assessment framework for the entry of bat-borne zoonotic viruses into the European Union. *PLOS ONE* 11 (10), e0165383.
- Sit, T.H.C., Brackman, C.J., Ip, S.M., Tam, K.W.S., *et al.*, 2020. Infection of dogs with SARS-CoV-2. *Nature* 586 (7831), 776–778.
- Smith, K.F., *et al.*, 2014. Global rise in human infectious disease outbreaks. *Journal of the Royal Society Interface* 11 (101), 20140950.
- Smith, K.M., *et al.*, 2012. Zoonotic viruses associated with illegally imported wildlife products. *PLoS One* 7 (1), e29505.
- Takthampunya, R., *et al.*, 2019. Metagenomic approach to characterizing disease epidemiology in a disease-endemic environment in Northern Thailand. *Frontiers in Microbiology* 10, 319.
- Thézé, J., *et al.*, 2018. Genomic epidemiology reconstructs the introduction and spread of Zika virus in Central America and Mexico. *Cell Host & Microbe* 23 (6), 855–864.
- Tian, H., *et al.*, 2015. Avian influenza H5N1 viral and bird migration networks in Asia. *Proceedings of the National Academy of Sciences of the United States of America* 112 (1), 172–177.
- Umhang, G., *et al.*, 2016. Surveillance and management of *Echinococcus multilocularis* in a wildlife park. *Parasitology International* 65 (3), 245–250.
- Voeten, H., Sikkema, R.S., Damen, M., *et al.*, 2020. Unravelling the modes of transmission of SARS-CoV-2 during a nursing home outbreak: Looking beyond the church super-spread event. *Clinical Infectious Diseases*.
- Walker, J.W., *et al.*, 2018. Transmissibility of emerging viral zoonoses. *PLoS One* 13 (11), e0206926.
- Wang, M., Yan, M., Xu, H., *et al.*, 2005. SARS-CoV infection in a restaurant from palm civet. *Emerging Infectious Diseases* 11 (12), 1860–1865.
- Wernike, K., Aebischer, A., Michelitsch, A., *et al.*, 2020. Multi-species ELISA for the detection of antibodies against SARS-CoV-2 in animals. *Transboundary and Emerging Diseases*.
- Wibawa, H., *et al.*, 2018. Exploring contacts facilitating transmission of influenza A(H5N1) virus between poultry farms in West Java, Indonesia: A major role for backyard farms? *Preventive Veterinary Medicine* 156, 8–15.
- Wilson, M.R., *et al.*, 2019. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *The New England Journal of Medicine* 380 (24), 2327–2340.
- Woolhouse, M., 2011. How to make predictions about future infectious disease risks. *Philosophical Transactions of the Royal Society B: Biological Sciences* 366 (1573), 2045–2054.
- Wu, J.Y., *et al.*, 2019. Transmission risk of avian influenza virus along poultry supply chains in Guangdong, China. *Journal of Infection* 79 (1), 43–48.
- Wu, F., Zhao, S., Yu, B., *et al.*, 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579 (7798), 265–269.
- Zaki, A.M., *et al.*, 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England Journal of Medicine* 367 (19), 1814–1820.
- Zhang, Q., Zhang, H., Gao, J., *et al.*, 2020. A serological survey of SARS-CoV-2 in cat in Wuhan. *Emerging Microbes & Infections* 9 (1), 2013–2019.
- Zhou, P., *et al.*, 2015. Avian influenza A(H7N9) virus and mixed live poultry–animal markets in Guangdong province: A perfect storm in the making? *Emerging Microbes & Infections* 4 (1), 1–3.
- Zhou, J., *et al.*, 2016. Isolation of H5N6, H7N9 and H9N2 avian influenza A viruses from air sampled at live poultry markets in China, 2014 and 2015. *Eurosurveillance* 21 (35), 30331.
- Zhou, P., Yang, X.L., Wang, X.G., *et al.*, 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579 (7798), 270–273.
- Zolfo, M., *et al.*, 2018. Profiling microbial strains in urban environments using metagenomic sequencing data. *Biology Direct* 13 (1), 9.

Relevant Websites

- <https://reliefweb.int/report/world/world-risk-annual-report-global-preparedness-health-emergencies-global-preparedness>
A World at Risk: Annual report on global preparedness for health emergencies.
Global Preparedness Monitoring Board [EN/AR/RU/ZH].
- <https://efsa.onlinelibrary.wiley.com/doi/epdf/10.2903/j.efsa.2014.3884>
An update on the risk of transmission of Ebola virus (EBOV) via the food chain.
- <https://apps.who.int/iris/bitstream/handle/10665/252646/WHO-OHE-PED-2016.2-eng.pdf>
Anticipating Emerging Infectious Disease Epidemic.
World Health Organization.
- <https://wwwnc.cdc.gov/eid/page/background-goals>
Background and Goals.
CDC.
- <https://isaric.tghn.org/protocols>
CCP UK & Annual Activation.
ISARIC.
- <https://www.who.int/blueprint/priority-diseases/en/>
Coronavirus disease (COVID-19) pandemic.
- <http://www.fao.org/3/a-i2198e.pdf>
Developing sustainable value chains for small-scale livestock producers.
- https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf?sfvrsn=a9ef618c_2
Diagnostic detection of 2019-nCoV by real-time RT-PCR.
- https://vac-lshim.shinyapps.io/ncov_vaccine_landscape/
COVID-19 vaccine tracker.
- <http://www.fao.org/avianflu/documents/Economic-and-social-impacts-of-avian-influenza-Geneva.pdf>
Economic and social impacts of avian influenza.
- <https://www.ecdc.europa.eu/en/about-us/partnerships-and-networks/disease-and-laboratory-networks/evd-labnet>
Emerging Viral Diseases-Expert Laboratory Network (EVD-LabNet).
- <https://life.eurobirdportal.org>
EuroBirdPortal.
LIFE EuroBirdPortal overview.
- <https://EVAg.eur-ocean-virus-archive.com>
European Virus Archive - GLOBAL.
- <https://reliefweb.int/report/world/exploring-machine-learning-map-yellow-fever-risk>
Exploring Machine Learning to Map Yellow Fever Risk.
- <https://consise.tghn.org>
Home.
CONSISE.
- <https://www.who.int/news-room/feature-stories/detail/how-who-is-working-to-track-down-the-animal-reservoir-of-the-sars-cov-2-virus>
How WHO is working to track down the animal reservoir of the SARS-CoV-2 virus.
- <http://www.cdc.gov/flu/pdf/avianflu/h5n1-inventory.pdf>
H5N1 Genetic Changes Inventory.
FluTrackers.
- <https://www.who.int/ihr>
International health regulations.
World Health Organization.
- www.movebank.org
Movebank.org.
- <https://doi.org/10.17504/protocols.io.bdp7i5rn>
nCoV-2019 sequencing protocol v2 (GunIt) V.2.
- <https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/>
Novel Coronavirus
China.
- <https://virological.org/t/novel-2019-coronavirus-genome/319>
Novel 2019 coronavirus genome.
- <https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-context>
Prioritizing diseases for research and development in emergency contexts.
- <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unknown-cause-china/en/>
Pneumonia of unknown cause
China.
- <https://www.glopid-r.org/our-work/social-science-research/>
Social Science Research.
GloPID-R.
- <https://www.who.int/news-room/commentaries/detail/status-of-environmental-surveillance-for-sars-cov-2-virus>
Status of environmental surveillance for SARS-CoV-2 virus.
- https://www.who.int/foodsafety/zooneses/final_concept_note_Hanoi.pdf?ua=1
The FAO-OIE-WHO.
- <http://www.fao.org/3/i0680e/i0680e02.pdf>
2. Change in the livestock sector.

Use of Immunoglobulins in the Prevention of Viral Infections

Leyla Asadi and Giovanni Ferrara, University of Alberta, Edmonton, AB, Canada

© 2021 Elsevier Ltd. All rights reserved.

Introduction

In the late 18th century, Emil von Behring and Shibasaburo Kitasato published their ground-breaking work showing that non-immune rabbits could be protected from developing tetanus if they were injected with sera obtained from rabbits infected with *Clostridium tetani*.

This was followed up one week later by another foundational work where Behring used inactivated *Corneybacterium diphtheriae* toxin to prevent diseases in guinea pigs (Kaufmann, 2017). At the same time, the studies of Paul Erlich contributed to understanding the foundations of humoral immunity and allowed the safe production of large volumes of high-quality anti-diphtheria serum from large animals (Graham and Ambrosino, 2015). These combined efforts led to dramatic success in reducing tetanus and diphtheria mortality and were followed by the development of similar therapies for a variety of other bacterial infections, including *Neisseria meningitidis*, *Haemophilus influenzae*, and *Streptococcus* sp. (Casadevall et al., 2004). This pre-antibiotic and pre-antiviral era was a golden age for “serum therapy” (Casadevall et al., 2004).

The first reported use of serum therapy for viruses appears to have been in 1907 when it was used to prevent roseola (Good and Lorenz, 1991). It was also used during the 1918 Spanish influenza pandemic. Indeed, a modern meta-analysis of studies published during the pandemic estimates that there may have been ~40% reduction in mortality in those who received Spanish influenza-convallescent human blood products early in the course of their disease (Luke et al., 2006).

Because early immunoglobulin products were often contaminated with animal proteins, serious adverse effects such as serum sickness and anaphylaxis remained a real and constant risk. By the 1930s, it was clear that immunoglobulin therapy could be safer and more effective if antibodies were isolated from pools of normal human sera. Furthermore, new methods of isolating and precipitating human immunoglobulin fractions from placental tissue extracts became available (Good and Lorenz, 1991). Despite these advancements, and also due to the cost and complexity of production, serum therapy for bacteria was supplanted by antibiotic therapy with sulfonamides, then beta-lactams.

After World War II, Dr. Edwin Cohn developed a technique that allowed for the purification of antibodies via ethanol fractionation of plasma (Graham and Ambrosino, 2015). His technique – which was further refined by J.L. Oncley (Barahona Afonso and Joao, 2016) – allowed the yield of immunoglobulin for intravenous and subcutaneous use. Since most viral infections remained without treatment, the focus shifted to the use of immunoglobulins for viral pathogens. From there, polyclonal immunoglobulins began to be used for measles, rabies, hepatitis A and B, varicella-zoster virus, and respiratory syncytial virus (RSV). Cohn and Oncley's techniques (with some additional steps) and the use of polyclonal immunoglobulins for these viruses remain in practice today (Barahona Afonso and Joao, 2016).

Since then, several variants of immunoglobulin products have been produced and investigated, in pace with the development of new molecular biology techniques allowing a higher grade of purity and specificity for different uses of the antibodies prepared for the immunoglobulin cocktails.

Polyclonal antibodies in the form of intravenous immunoglobulin or hyperimmune immunoglobulin remain the most commonly used form of immunoglobulins in viral infections. Hyperimmune globulins contain a higher proportion of immunoglobulins targeting a specific pathogen. They are produced either by identifying individuals with high titers of the antibody of interest or by vaccination of donors.

Given the increasing ease in identifying and manufacturing monoclonal antibodies, they have been the focus of intense research, particularly for the treatment of emerging infectious diseases such as viral hemorrhagic fevers. The hybridoma technique developed in the 1970s was a major advancement that allowed the in-vitro production of monoclonal antibodies. In this method, mice were exposed and immunized with antigens of interest, and the antigen-specific B-cells were harvested from the mice spleens: These cells were then immortalized by fusing them with immortal B-cell cancer cells, allowing expansion and production of the same monoclonal antibody (Marston et al., 2018; Salazar et al., 2017). However, newer and more direct methods of monoclonal antibody production are now available. For instance, if the antigen is known, it can be used to screen a phage display antibody library (Salazar et al., 2017) – with antibodies derived from naturally infected or immunized persons. Alternatively, memory B cells – again from infected or immunized persons – can be sorted by flow cytometry, based on the affinity of their B-cell receptor for defined antigens (Salazar et al., 2017). The identified antibodies can then be cloned. Alternatively, deep sequencing of the B-cell IgG repertoire enables the production of heavy and light chain paired mAbs (McDaniel et al., 2016). Once the antibody has been identified, further techniques are available to stabilize the final product and reduce the occurrence of side effects (e.g., the Fc receptor can be modified to prevent undesirable interactions while maintaining the neutralizing functions of the heavy and light variable chains) (Sapparapu et al., 2016).

The possibility to develop monoclonal antibodies targeted to neutralizing epitopes of viral and bacterial molecules enables the production of larger amounts of immunoglobulin without the need to obtain blood products from thousands of donors. It also enables precision medicine in the field, increasing the efficacy and safety of these products that would become “off the shelf” drugs with precise indications. Unfortunately, apart from the few indications already approved (that are reported at the end of the article), this reality is still far for most infectious diseases. Major limitations of this approach include the limited knowledge of the

humoral response against specific infectious agents (i.e., how to identify the right epitopes and their combination to achieve humoral protection) and the existing high cost of molecular techniques.

In this review, we will discuss disease-specific uses of polyclonal and monoclonal antibodies in the prevention of infection.

Polyclonal Immunoglobulins Used in the Prevention of Infectious Diseases

Hepatitis A

The hepatitis A virus (HAV) typically results in asymptomatic disease or a self-limited illness. It rarely causes fulminant hepatitis, chronic complications, or death; however, those with pre-existing chronic liver disease and those who are immunocompromised are at an increased risk for such complications. Transmission of hepatitis A occurs through the fecal-oral route, either via person-to-person contact or consumption of contaminated food or water (Bennett *et al.*, 2014).

Passive immunization against hepatitis began in the 1940s. After seeing the effectiveness of measles passive immunization, Dr. Joseph Stokes administered gamma globulin to children at a camp with an “icteric hepatitis” outbreak. He found that there was a 39% absolute reduction (90% relative reduction) in the development of hepatitis among gamma globulin recipients (Cuthbert, 2001). More recently, a 2009 Cochrane systematic review explored the efficacy of immunoglobulins for pre-exposure and post-exposure prophylaxis. For pre-exposure prophylaxis, they found that when compared with placebo, immunoglobulins reduced the risk of hepatitis A infection by 47%–72% (Liu *et al.*, 2009). No trials compared the clinical efficacy of immunoglobulin versus inactivated HAV vaccine; however, both resulted in similar HAV antibody development at four weeks. By 24 weeks, no one in the immunoglobulin group was still seropositive. Only one study looked at giving both agents; there was no benefit to giving both (albeit the study was of poor quality). Concerning post-exposure prophylaxis, they identified two trials. The more recent trial, completed in 2007, compared immunoglobulin to inactivated hepatitis A vaccine by randomizing 1,090 susceptible contacts of HAV cases to inactivated HAV vaccine or immune globulin within two weeks of exposure. There was no difference in disease incidence between the two groups (4.4% versus 3.3%, respectively) (Victor *et al.*, 2007).

Separately, studies have shown that the inactivated HAV vaccine is safe and, in healthy adults <40 years of age, confers 95% seropositivity for at least 25 years (Bovier *et al.*, 2010; Hens *et al.*, 2014). However, based on limited data, there may be reduced immunogenicity in older adults, particularly with early seroprotection (within 15 days of vaccination) (Link-Gelles *et al.*, 2018). The vaccine has also not been licensed for use in children <1 year of age and could theoretically interfere with passively acquired maternal antibodies. Nevertheless, due to the high immunogenicity of the HAV vaccine, its ability to induce active immunity, longer duration of protection, ease of administration, and greater acceptability and availability, the vaccine became the preferred method of HAV prevention. Immunoglobulin was primarily reserved for those <12 months of age, >40 years old, or those who are immunocompromised.

For the first time, in 2018, the Advisory Committee on Immunization Practice (ACIP) revised their recommendations (Nelson *et al.*, 2018). In contrast to previous recommendations, the vaccine is now preferred for those >40 years of age, and immunoglobulin is only added on a case-to-case basis. They justify this recommendation by pointing at the decreased potency of immunoglobulin, due to reduced HAV antibody titers in the general population from which the immunoglobulin is obtained. This has resulted in a higher volume of immunoglobulin being required for the prevention of HAV infection, making it even less easily available or acceptable. They also now recommend vaccine for pre-exposure prophylaxis in infants traveling internationally between six and 11 months of age (off-label use). This is because the measles, mumps, and rubella (MMR) vaccine and HAV immunoglobulin cannot be given simultaneously. Thus, clinicians should prioritize the prevention of measles, mumps, and rubella by vaccinating with the MMR vaccine and use an inactivated HAV vaccine, as opposed to the immunoglobulin, to try to achieve hepatitis A protection.

The remainder of the recommendations are unchanged and described in Table 1, modified from the 2018 MMWR report on HAV (Nelson *et al.*, 2018).

Measles

The measles virus is a highly contagious virus that produces a characteristic exanthem after a prodromal phase of conjunctivitis, cough, coryza, and fever. Serious complications include pneumonia, post-infectious encephalitis (one in 1000–2000 cases), subacute sclerosing panencephalitis (SSPE) (one in 100,000 cases), and death (~1–3 in 1000 cases) (Bennett *et al.*, 2014). In developing countries, pediatric mortality rates approach 2%–15%. Pregnant women and those with immunodeficiency are at higher risk of complications (Matysiak-Klose *et al.*, 2018).

Before the arrival of the measles vaccine in 1963, major epidemics causing millions of deaths were not uncommon (World Health Organization, 2019a). With the introduction of a highly effective vaccine, global measles rates have plummeted and measles was considered eliminated in most developed countries (World Health Organization, 2019a). One dose of the measles, mumps, rubella (MMR) vaccine provides ~93% protection (CDC, 2020). With two doses, effectiveness reaches ~97% (CDC, 2020). Measles vaccination is the only approach for pre-exposure prophylaxis and the mainstay of post-exposure prophylaxis (PEP) in immunocompetent individuals >6 months (Plotkin, 2012). The vaccine is approved for those >6 months of age and, because it is a live attenuated virus, its administration is contraindicated for those pregnant, with severe allergic reactions to

Table 1 Post-exposure and pre-exposure prophylaxis for hepatitis A virus for healthy individuals

Indication/age group	Hepatitis A vaccine	Immune globulin
< 12 months, healthy, post-exposure	No	0.1 ml/kg
< 12 months, healthy, pre-exposure		
• < 6 months	No	0.1–0.2 ml/kg ^a
• 6–11 months	1 dose ^b	None
12 months – 40 years, healthy, post-exposure and pre-exposure	1 dose ^c	None
> 40 years, healthy, post-exposure and pre-exposure	1 dose ^c	0.1 ml/kg ^d
≥ 12 mos, immunocompromised or chronic liver disease, post-exposure	1 dose	0.1 ml/kg
≥ 12 mos, vaccine contraindicated, post-exposure	No	0.1 ml/kg
All ages, immunocompromised or chronic liver disease, pre-exposure	1 dose	0.1–0.2 ml/kg ^{d,a}
> 6 months, vaccine contraindicated, pre-exposure	No	0.1–0.2 ml/kg ^a

^adose depends on the duration of travel.^bdose does not count toward routine 2-dose series, which should be started after one year.^conly one dose necessary for post-exposure prophylaxis, but two doses for long-term immunity.^drisk assessment (severity of exposure and potential for complications) undertaken to determine the need for immune globulin.

Note: Modified from Nelson, N.P., *et al.*, 2018. Update: Recommendations of the advisory committee on immunization practices for use of hepatitis A vaccine for postexposure prophylaxis and for preexposure prophylaxis for international travel. *Morbidity and Mortality Weekly Report* 67 (43), 1216–1220.

any component of the vaccine, and who are immunocompromised. Thus, immunoglobulins continue to play a role in PEP, particularly for those most vulnerable to complications of the disease. The MMR vaccine, a live vaccine, is not administered in conjunction with immunoglobulins as it results in the negation of each component.

In 1926, Zingher showed that convalescent whole blood, serum, or plasma could prevent measles, and in 1940, Janeway confirmed that early immunoglobulin administration could prevent measles in three out of every four people and attenuate disease in the fourth (Plotkin, 2012). Recently, a 2014 Cochrane systematic review examined the effectiveness of post-exposure immunoglobulins (Young *et al.*, 2014). Looking back to the 1920s, they identified 13 studies (3,925 patients), with only one being a randomized controlled trial. 11 studies predated widespread vaccine implementation (so high measles-specific titers in the immunoglobulin preparations would be expected). The studies were heterogeneous and of moderate quality. Nevertheless, the authors concluded that immunoglobulins within seven days of exposure reduces the risk of measles by up to 83% compared to no intervention. Due to high rates of immunity resulting from vaccination, modern PEP studies are challenging; however, since the publication of the Cochrane review, there are at least two new observational studies. A study of 121 contacts who received PEP showed that MMR vaccine effectiveness was at 83.4% (95% CI, 34.4%–95.9%) and immunoglobulin (dosed at 0.5 ml/kg) effectiveness was at 100% (approximately 95% CI, 56.2%–99.8%) (Arciuolo *et al.*, 2017). However, most immunoglobulin recipients were <6 months of age and so they may have had maternal antibodies that provided protection. A 2017 Canadian study of 55 self-reported susceptible individuals found that immunoglobulin effectiveness was ~69%; however, the dose administered in this study was only 0.25 ml/kg (Bigham *et al.*, 2017). Both studies excluded pregnant women or immunocompromised patients.

Based on these studies, there is consensus that immunoglobulin is beneficial and should be given as quickly as possible (ideally, within six days) if there is a contraindication to the MMR vaccine (Matysiak-Klose *et al.*, 2018). If intramuscular immunoglobulin is available, the dosage should be 0.5 ml/kg; otherwise, IVIg (400 mg/kg) is the agent of choice (Matysiak-Klose *et al.*, 2018; Tunis *et al.*, 2018). Given that individuals weighing 30 kg or more would not have complete protection with IM immunoglobulin at a dosage of 0.5 ml/kg and would therefore require multiple injections, IVIg is the preferred product for most immunocompromised individuals (Tunis *et al.*, 2018; Public Health England, 2020).

Given the reduced efficacy of the MMR vaccine when administered as PEP 72 h post-exposure, some jurisdictions (i.e., United States) recommended immunoglobulin for all susceptible individuals (including non-pregnant and immunocompetent hosts) if they present between 72 h and six days (McLean *et al.*, 2013). However, others (i.e., Canada) no longer recommended the use of immunoglobulin for immunocompetent individuals >12 months of age because the measles virus antibodies in immunoglobulin products may no longer be optimally protective (due to waning anti-measles titers in the general population) and because there is a low risk of disease complications in this group (Tunis *et al.*, 2018).

Hyperimmune Immunoglobulins Used in the Prevention of Infectious Diseases

Hepatitis B Hyperimmune Globulin

Depending on the age of acquisition, hepatitis B virus (HBV) infection may result in acute or chronic hepatitis. Infection in newborns can become chronic in 90% of cases. Chronic disease, in turn, may develop into cirrhosis or hepatocellular carcinoma, resulting in death. Perinatal transmission remains the main mode of HBV transmission but in adults, acquisition primarily occurs via intravenous drug use, sexual activity, or occupational exposure (Bennett *et al.*, 2014).

The HBV vaccine – a recombinant vaccine containing yeast-derived hepatitis B surface antigen (HBsAg) – is the cornerstone of HBV prevention (Schillie *et al.*, 2018). The only contraindication is anaphylaxis with having previously received HBV vaccine. The

seroconversion rate is ~95%–100% in children and ~90% in adults, with lower response rates in higher ages (Henry and Baclic, 2017). The vaccine's effect is blunted in the immunocompromised or those with underlying liver disease (Walayat *et al.*, 2015) and ~5% of individuals are vaccine non-responders.

Hepatitis B hyperimmune IG (HBIG) is administered at a standard dose of 0.06 ml/kg and is used in conjunction with the vaccine in post-exposure settings, in liver transplants and as prevention for maternal-to-child-transmission (MTCT). For known non-responders, HBIG is the only method of protection after an HBV exposure (Schillie *et al.*, 2018). Among susceptible individuals with a high-risk HBV exposure in occupational or non-occupational settings, one dose of HBIG is recommended as soon as possible and up to seven days post-exposure. HBIG provides protection until active immunity from vaccination develops. The evidence for this is from lower quality randomized trials from the 1970s or early 1980s where it appeared that among those given HBIG within seven days of percutaneous exposure to HBsAg positive blood, ~75% developed protective antibodies (Schillie *et al.*, 2018). No studies have compared the efficacy of HBIG and the HBV vaccine series or the combination of these two agents in post-exposure settings.

In addition to immunization of infants and treatment of highly viremic mothers with antiviral therapy, HBIG is also given to prevent vertical transmission. HBV vaccine is 75% effective and HBIG is 71% effective in preventing HBV transmission; when combined, their efficacy reaches 94% (Lee *et al.*, 2006; Schillie *et al.*, 2015; Beasley *et al.*, 1983). Therefore, in cases where the mother is HBsAg positive (and anti-HBe negative) or her hepatitis B status is unknown, vaccine and HBIG are to be administered within 12 h of birth. The addition of HBIG to the vaccine has an incremental cost-effectiveness ratio of \$6957 per quality-adjusted life-year (Fan *et al.*, 2016). Despite decreased efficacy after 48 h, HBIG may be given up to seven days post-birth (Henry and Baclic, 2017). A Cochrane review of ~30 randomized control trials from China looking at the use of HBIG during pregnancy (antenatal) to prevent MTCT was inconclusive due to very low to low-quality evidence (Eke *et al.*, 2017).

A subset of HBsAg-positive liver transplant recipients also require HBIG. As per the 2018 American Association for the Study of Liver Diseases (AASLD) Hepatitis B guidance, in HBsAg-positive liver transplant recipients at low risk of recurrence, "either no HBIG or HBIG for only 5–7 days combined with antivirals long term has been highly effective" (Terrault *et al.*, 2018). However, HBsAg-positive liver transplant recipients who are HIV- or hepatitis delta virus (HDV)-positive, have antiviral drug resistance, or high levels of HBV DNA at the time of transplantation may benefit from indefinite HBIG as they are at higher risk of virological breakthrough and have more limited rescue therapy options (Terrault *et al.*, 2018).

Cytomegalovirus Hyperimmune Immunoglobulin

In immunocompetent individuals, cytomegalovirus (CMV) infection is typically asymptomatic. However, in transplant recipients, CMV can cause both direct disease (a viral syndrome or tissue invasive disease) or may indirectly increase the risk of rejection (Bennett *et al.*, 2014). CMV also results in congenital infection and is a leading cause of sensorineural hearing loss and neurological disability (RCOG Scientific Impact Paper, 2018).

In the late 1980s, CMV hyperimmune immunoglobulin (CMVIG) was licensed for use in renal transplant recipients to prevent primary CMV disease (Snydman *et al.*, 1987, 1993). In a small randomized trial of renal transplant recipients (Snydman *et al.*, 1987) and then, subsequently, in a small, double-blind, randomized controlled trial (Snydman *et al.*, 1993) involving liver transplant recipients – both against a placebo arm – CMVIG was beneficial in reducing primary CMV disease and severity. However, with the advent of effective antivirals, CMV immunoglobulin use has decreased and there is limited evidence about its efficacy in combination with antivirals or with more modern immunosuppression. While there may be a role for combination therapy in high-risk lung or small bowel recipients (where the donor is CMV-positive and the recipient is CMV-negative) (Florescu *et al.*, 2014; Valantine *et al.*, 2001), the randomized trials supporting combination therapy have not been adequately powered. The latest consensus guidelines on the management of CMV disease in solid-organ transplantation no longer recommends routine use of CMVIG (Kotton *et al.*, 2018). That said, CMVIG has been used in patients who are intolerant of antivirals (i.e., due to prolonged neutropenia) (Kotton *et al.*, 2018) and there are case reports of its use as an adjunct in the treatment of tissue-invasive CMV disease (Schulz *et al.*, 2016).

Like its use in transplant recipients, the use of CMVIG in pregnancy has also waned. In the early 2000s, a series of non-randomized studies studying CMV-specific hyperimmune globulin in pregnant women with primary CMV infection found a decreased rate of mother-to-fetus transmission as well as a lower risk of congenital disease (Nigro *et al.*, 2005, 2012; Visentin *et al.*, 2012). However, in 2014, a placebo-controlled randomized trial of 124 pregnant women did not demonstrate a significant benefit (Revello *et al.*, 2014). In the Congenital Human CMV Infection Prevention (CHIP) trial, not only was there no difference in congenital infections, immunological, or virological parameters, but there was even a trend toward more adverse obstetric events in the immunoglobulin group. While it could be argued that this RCT was underpowered, CMV immunoglobulin is no longer routinely recommended in pregnancy (RCOG Scientific Impact Paper, 2018; Yinon *et al.*, 2018). A large, placebo-controlled, double-blinded, randomized trial of CMV immunoglobulin in pregnancy is currently underway (The George Washington University Biostatistics Center, 2020).

Finally, despite the waning use of CMVIG, studies are underway examining CMV monoclonal antibodies. A phase two double-blinded, placebo-controlled, randomized trial of RG7667, a combination Monoclonal Antibody in high-risk renal transplant recipients showed a reduction of CMV DNAemia and disease (Ishida *et al.*, 2017). A phase II clinical trial of CSJ148, another combination monoclonal antibody, was recently completed among hematopoietic stem cell recipients, but it did not achieve the primary efficacy endpoint (Maertens *et al.*, 2020).

Rabies Hyperimmune Immunoglobulin

After the appearance of clinical signs, rabies, a zoonotic encephalitis caused by a variety of different species of viruses in the Rhabdoviridae family, is almost universally fatal. Because 99% of human cases are caused by dog bites (WHO/Department of Control of Neglected Tropical Diseases, 2018), the primary method of rabies control has been through mass vaccination of dogs. Pre-exposure prophylaxis, with the highly effective rabies vaccine, is recommended only for those at occupational risk or those living in remote endemic settings (WHO/Department of Control of Neglected Tropical Diseases, 2018).

Once rabies exposure occurs, post-exposure prophylaxis (PEP) is essential to prevent fatality. PEP consists of thorough wound cleaning, rabies vaccine, and rabies hyperimmune IG (RIG) (WHO/Department of Control of Neglected Tropical Diseases, 2018). RIG is only given to those who have never been immunized. The WHO categorizes rabies exposures into three categories of severity and recommends RIG for severe category three cases. Some national guidelines use RIG in less severe exposure incidents (Human rabies prevention-United States, 2008). Because of the prolonged incubation (at least 45 days) and the latency period of the virus, PEP administration is only too late after the occurrence of clinical signs of rabies. When RIG is warranted, either human RIG (hHRIG), which is from pooled plasma of hyperimmunized donors, equine RIG (eRIG), or the rabies' monoclonal antibody (currently only approved in India) may be given. These products provide immediate virus-neutralizing antibodies and may partially suppress antibody production; therefore, only one dose is recommended and it should be given no more than seven days after the vaccine series is started (Khawplod *et al.*, 1996). Ideally, the full dose of RIG should be infiltrated into the bite wound and surrounding area.

The evidence for the addition of RIG to the vaccine series dates to a small study from the 1950s. In 1954, at the Pasteur Institute in Tehran, Iran, 29 rabid wolf bite victims were randomized to receive vaccine-only PEP or vaccine and anti-rabies serum. Among those with severe wounds, 3/5 in the vaccine group and 1/13 in the vaccine and serum group died. From there, RIG with vaccine became the standard of care (Sparrow *et al.*, 2019; Baltazard and Bahmanyar, 1955; Habel and Koprowski, 1955). A more recent study has corroborated these findings by showing significantly higher rabies antibodies at the end of the vaccine series in those who received rabies immunoglobulin (Navarrete-Navarro *et al.*, 1999).

Most rabies deaths occur in poor rural populations in Africa and Asia where rabies PEP is not affordable for the population; only 1%–10% of patients requiring RIG receive it (Sparrow *et al.*, 2019). Furthermore, blood-derived products have a short shelf-life and there are ethical issues around the use of animals in producing immunoglobulins. Given these issues, there have been global shortages of both hRIG and eRIG, and this has led to the interest and development of rabies monoclonal antibodies. The World Health Organization now recommends the use of monoclonal rabies immunoglobulin (where available) and is partnering with collaborating centers for rabies to facilitate rabies mAb cocktails (World Health Organization (WHO), 2017, 2016). There are currently multiple rabies mAbs in various stages of development (Sparrow *et al.*, 2019).

In 2016, Rabishield (SII RMAb), an IgG1 mAb was approved for use in India. In 2/3 phases of randomized controlled trial carried out in India, 200 patients with category three bites from presumed rabid animals were randomized to either vaccine and HRIG or vaccine and RMAb. SII RMAb was safe and demonstrated non-inferiority in the level of rabies virus neutralizing activity (Gogtay *et al.*, 2018).

Vaccinia Hyperimmune Globulin

Vaccinia immune globulin (VIG) is approved for the treatment of complications associated with the smallpox vaccine. It may also be of some benefit for post-exposure prophylaxis of smallpox. Smallpox, caused by the variola virus, was a highly infectious condition characterized by rash and fever and complicated by a myriad of systemic manifestations, including keratitis, encephalitis, and death. In the late 1960s, using the smallpox vaccine, a global mass vaccination campaign was initiated (Meyer *et al.*, 2020). By 1980, the disease was eradicated globally (World Health Organization, 1979). Despite the eradication of wild type smallpox, two variola virus repositories were maintained at official centers in the United States and Russia for research purposes. Some experts also contend that other countries or groups may also possess or aim to obtain virus samples for bioterrorist aims (Witteck, 2006).

The smallpox vaccine is composed of a live attenuated vaccinia virus. The vaccine itself may rarely result in serious complications (Cono *et al.*, 2003). Though historical rates of vaccine adverse effects were higher, after a round of smallpox vaccination of US service members following the September 11, 2001 terrorist attacks, only two serious complications occurred out of ~770,000 individuals (Poland *et al.*, 2005).

At this time, the only available immunoglobulin formulation for smallpox is an FDA-approved intravenous formulation, produced in Canada and available through the CDC. VIG is given at a dose of 2 ml/kg (100 mg/kg). Because vaccine-related complications are rare, the efficacy of VIG is difficult to study. The largest review on VIG efficacy is a 2004 review of 16 non-controlled studies (Hopkins and Lane, 2004). The studies included in the review were published between 1948 and 2003. These studies suggest that there is potential morbidity and mortality benefit with the use of VIG for progressive vaccinia (vaccinia necrosum) and eczema vaccinatum and severe generalized vaccinia. There is no benefit with post-vaccination encephalitis and VIG is contraindicated in vaccinia keratitis. The concern about keratitis is based in part on historical experimental data in rabbits showing that daily administration of VIG in vaccinal keratitis resulted in larger eye lesions with more persistent scarring, possibly due to immune complex disease (Fulginiti *et al.*, 1965). If the smallpox vaccine is given to people with contraindications to the vaccine (i.e., people with eczema, infants, pregnant women, or the immunosuppressed), prophylactic VIG may be of some benefit. However, its use is not approved for prophylaxis (Witteck, 2006).

Some have explored the use of VIG for post-exposure prophylaxis against smallpox (Hopkins and Lane, 2004). These studies did not meet modern criteria for randomized trials but did point to some benefits. Therefore, given its safety profile, VIG could be considered for individuals at high risk of smallpox complications.

Varicella Hyperimmune Immunoglobulin

Varicella-zoster virus causes varicella (chickenpox) and herpes zoster (shingles). Varicella is a self-limited illness characterized by fever and vesicular rash. Rarely, encephalitis, pneumonia, and death may occur. Pregnant women, neonates, and immunocompromised hosts are far more likely to suffer severe manifestations and multi-organ involvement. In neonates whose mothers developed disease five days before or 48 h post-partum, mortality can be as high as 30% (Bennett *et al.*, 2014). Congenital varicella, though rare, is associated with eye and skin abnormalities, hypoplastic extremities, and CNS impairment (Enders *et al.*, 1994).

The primary method of pre-exposure and post-exposure varicella prevention is vaccination. In the post-exposure setting, among immunocompetent children, the varicella vaccine is 86% effective against all forms and 100% effective against moderate-to-severe disease (Izurieta *et al.*, 1997). However, because it is a live-attenuated vaccine, it is contraindicated in precisely those at the highest risk of severe disease – neonates, pregnant women, and the severely immunocompromised. It is in this vulnerable group where passive immunization may be warranted.

Passive immunization against varicella was attempted as early as the 1940s (Funkhouser, 1948; Schaeffer and Toomey, 1948). In 1962, it was shown that non-specific human serum immunoglobulin limited varicella severity but did not prevent disease (Ross *et al.*, 1962). Then, in a small study with 12 participants, Zoster IG (ZIG), an immunoglobulin formulation produced from convalescent plasma of herpes zoster – appeared to prevent disease transmission (Brunell *et al.*, 1969). However, ZIG was difficult to obtain and was in short supply, and in the early 1980s, varicella-zoster IG (VZIG) – an immunoglobulin prepared from normal donor plasma selected for high titer of antibody to varicella-zoster virus (VZV) – came to market to meet supply shortages. In a double-blind randomized trial, outcomes with ZIG were comparable with outcomes with VZIG (Zaia *et al.*, 1983). The latest preparation of immunoglobulin, VariZig, is equivalent to VZIG in efficacy and safety (Koren *et al.*, 2002). No randomized trials have compared immunoglobulin to placebo. However, observational studies support a change in disease presentation and occurrence. For instance, an observational study from Italy showed that the risk of developing varicella was lower in VZIG recipients than those who did not receive it [42% (21 of 50) versus 72% (13 of 18); $p = 0.0263$] (Trotta *et al.*, 2018). Though methodologically suboptimal, comparison with historical controls (where IG was not administered) also suggests that IG administration alters the incidence of disease, particularly severe disease (Levin *et al.*, 1986).

There are no published data regarding the efficacy of VariZig in immunocompromised hosts nor are there trials comparing immunoglobulin prophylaxis with antiviral prophylaxis (either alone or in combination). However, according to the most recent iteration of the UK varicella PEP guidelines, the administration of VZIG to children with hematologic or oncology diagnoses is variable. This has allowed them to conduct an observational study comparing VZIG and acyclovir. Interim analysis of 73 recruited patients has shown acyclovir is at least equivalent to VZIG in preventing varicella (Public Health England, VZIG Expert Working Group, 2019).

National guidelines on PEP varicella immunoglobulin vary. In the UK, where VZIG is in short supply, immunoglobulin is recommended for neonates and women at <20 weeks gestation. Antivirals are preferred after 20 weeks gestational age and for the immunocompromised (Public Health England, VZIG Expert Working Group, 2019). On the other hand, Canadian and US guidelines recommend the use of VariZig after exposure to varicella-zoster virus (varicella or herpes zoster) for neonates, all pregnant women, at-risk premature infants, and immunocompromised children and adults who lack evidence of immunity (National Advisory Committee on Immunization, NACI, 2016; Bialek *et al.*, 2013).

VariZig administration is approved for up to ten days after exposure (Bialek *et al.*, 2013). There is limited but increasing evidence of its efficacy beyond 96 h (Levin *et al.*, 2019; Swamy and Dotters-Katz, 2019). The duration of protection is unknown but expected to be at least three weeks. VariZig might extend the incubation period of the virus from 10–21 days to ≥ 28 days.

Monoclonal Antibodies Approved by the Food and Drug Administration

Respiratory Syncytial Virus

Respiratory syncytial virus (RSV) is the most common cause of pediatric acute lower respiratory tract infections (Nair *et al.*, 2010). Only a small percentage of healthy, term infants develop severe complications; however, in high-risk infants, including those who are preterm or who have chronic lung diseases, congenital heart diseases, Down syndrome, neuromuscular conditions, or immunodeficiencies, the rate of hospitalizations increases to $\sim 10\%$ (Feltes *et al.*, 2003; Robinson *et al.*, 2016b; The Impact-RSV Study Group, 1998). There are no known treatments for RSV and antiviral therapy with ribavirin – which is of questionable efficacy – is reserved for severe cases in the immunosuppressed (American Academy of Pediatrics Committee on Infectious Diseases; American Academy of Pediatrics Bronchiolitis Guidelines Committee, 2014).

However, hyperimmune immunoglobulins and monoclonal antibodies have shown promise in the prevention of RSV in vulnerable groups. In the early 1990s, two randomized controlled trials carried out in high-risk infants showed that an RSV immune globulin (RSV-IGIV) infused monthly during the RSV season could reduce hospitalizations by 40%–65% (Groothuis and Simoes, 1993; PREVENT

Study Group, 1997). However, there were reports of an increased risk of morbidity and mortality in infants with congenital heart disease (Simoes *et al.*, 1998), and RSV-IGIV was then replaced by Palivizumab.

Palivizumab (Synagis) is a humanized murine monoclonal antibody to the RSV F glycoprotein, a highly conserved glycoprotein among the two serotypes of RSV (serotype A and B) (Resch, 2017). By binding to the F-protein, Palivizumab impairs the fusion of the viral envelope with the host cell membrane, preventing entry into the host cell. Palivizumab also prevents the fusion of the plasma membrane of infected cells, a process that results in syncytium formation. It was the first anti-viral monoclonal antibody approved by the FDA. In the original randomized controlled trial of 1,502 children with prematurity or bronchopulmonary dysplasia (BPD) who were randomized to placebo or monthly IM injections of Palivizumab, there was a 55% reduction in hospitalizations (The IMPact-RSV Study Group, 1998). Hospitalizations were decreased by 78% for children with prematurity without BPD and by 39% for children with BPD. Subsequent studies have also shown reduced hospitalization in infants with hemodynamically significant congenital heart disease (Feltes *et al.*, 2003). A 2014 systematic review of 20 observational studies (Nusrat Homaira *et al.*, 2014) was consistent with these benefits and a 2019 self-controlled case series of 24,329 infants, 1.1% of whom received at least one dose of Palivizumab and 6.2% of whom developed RSV, found that Palivizumab was associated with a 74% reduction in incidence of RSV infection (Moore *et al.*, 2019).

While results on the cost-effectiveness of Palivizumab have been inconsistent (Andabaka *et al.*, 2013), the most recent systematic review on the topic showed that, from a payer perspective, Palivizumab was cost-effective in preterm infants, infants with BPD or chronic lung disease, congenital heart disease, and infants in certain remote communities (Mac *et al.*, 2019). There is general agreement among national guidelines that Palivizumab is beneficial in these specific high-risk situations (American Academy of Pediatrics Committee on Infectious Diseases; American Academy of Pediatrics Bronchiolitis Guidelines Committee, 2014; Robinson and Le Saux, 2015; Public Health England, 2015). However, there is some disagreement about adopting its routine use in Down's syndrome, neuromuscular disorders, cystic fibrosis, or other lung diseases (Luna *et al.*, 2018).

In addition to Palivizumab, Motavizumab was a second-generation monoclonal antibody (mAb) derived from Palivizumab (Feltes *et al.*, 2011; Carbonell-Estrany *et al.*, 2010). It was withdrawn from the market after the FDA requested more information on the risks and severity of hypersensitivity reactions (see "Relevant Websites section"). Other monoclonal antibodies with extended half-lives and strong neutralizing activity are currently under investigation, with two entering phase two and three investigations (Aranda and Polack, 2019).

Human Immunodeficiency Virus

The emergence of antiretroviral therapies (ART) has helped halt the progression of the Human Immunodeficiency Virus (HIV) epidemic. Treatment with ART controls the progression of the HIV-1 virus among infected individuals and also helps prevent HIV acquisition when taken prophylactically by uninfected individuals. However, for those with multi-drug resistant HIV with limited treatment options, monoclonal antibodies may serve as a salvage therapy. Moreover, until an effective vaccine is found, other treatment modalities, such as broadly neutralizing monoclonal antibodies, may aid in controlling the epidemic.

Ibalizumab, a mouse-derived IgG4 monoclonal antibody, was approved by the FDA in 2018 for use in multi-drug resistant HIV (FDA, 2018). Ibalizumab does not act on circulating HIV-1 or cells already infected with HIV-1; instead, it blocks HIV entry into CD4 cells by binding to the CD4 extracellular domain 2 (Emu *et al.*, 2018). In phase three of an open-label study, 40 individuals with multi-drug resistance HIV (defined as triple-class resistance) were treated with at least one active antiviral agent as well as Ibalizumab. 83% of patients had an initial decrease in their viral load and at 25 weeks, 43% of patients had an undetectable viral load (Emu *et al.*, 2018). For patients with MDR-HIV, Ibalizumab may allow the construction of a life-saving treatment regimen.

The other antibodies being studied for the prevention and treatment of HIV are classified as broadly-neutralizing monoclonal antibodies (bnMAbs) that act directly on HIV (Caskey *et al.*, 2019). The site of action for these antibodies is the HIV-1 envelope spike (Env) – the only target on the surface of the virus (Caskey *et al.*, 2019). The observation that a small fraction of HIV-1-infected persons are able to produce antibodies that neutralize multiple circulating HIV-1 strains indicated the possibility to develop potent bnMAbs (Freund *et al.*, 2017). bnMAbs can both lead to the destruction of HIV-infected cells by engaging with Fc receptors and can boost the endogenous immune system through immune complex formation that increases B-cell and T-cell activation. These features have the potential to both control viral replication and lead to the reduction of viral reservoirs.

To date, the clinical information on bNAbs is primarily limited to phase one and phase two studies but holds promise for prophylaxis, maintenance therapy, and potentially even in eliciting prolonged remission or cure (Gruell and Klein, 2018). A recent phase two, randomized controlled trial of VRC01 enrolled 19 individuals with acute HIV infection who started and were well controlled on ART (Crowell *et al.*, 2019). These individuals then underwent an analytic interruption of ART and 14 received VRC01 infusions every three weeks for 24 weeks. However, only one VRC01 recipient achieved the primary efficacy endpoint of viral suppression after ART interruption. The authors postulated that, like antiviral therapy, a combination of bnMAbs may be required to achieved viral suppression.

Experimental Immunoglobulin Therapies

Seasonal Influenza

Both polyclonal immunoglobulin and monoclonal antibodies have been used in the treatment of seasonal influenza and remain active areas of research interest. A modern systematic review of the impact of early administration of convalescent plasma during

the 1918 Spanish influenza epidemic hypothesized that the use of the convalescent plasma resulted in a 21% absolute reduction in mortality (Luke *et al.*, 2006). Subsequently, a small cohort study of patients with severe H1N1 influenza suggested a ~30% reduction in absolute mortality with administration of H1N1 convalescent plasma (Hung *et al.*, 2011). The same research group followed up with a randomized trial of convalescent plasma-derived immunoglobulin involving 35 patients (Hung *et al.*, 2013). They found a reduced viral load among plasma recipients and on subgroup analysis, immunoglobulin within five days of symptom onset was the only factor independently associated with reduced mortality. A phase two randomized trial of convalescent plasma versus standard of care also demonstrated improvements in clinical parameters (Beigel *et al.*, 2017). These studies, however, had methodological issues including small sample sizes, higher-than-expected mortality levels among controls, or differing baseline characteristics between the study groups.

Thus, the findings of two phase III randomized trials were eagerly awaited. Unfortunately, both trials were disappointing and did not support the use of hyperimmune globulin in seasonal influenza. FLU-IVIG was an international double-blind, randomized trial of anti-influenza hyperimmune intravenous immunoglobulin vs standard of care (including neuraminidase-inhibitors) (Davey *et al.*, 2019). Among the 308 participants, there was no benefit with the use of the hyperimmune IVIG. That said, there was clinical benefit among the 27% of patients with influenza B. A randomized trial of high titer plasma (HAI antibody titers of $\geq 1:80$) vs low titer plasma (HAI $< 1:10$) in patients with severe influenza A also did not reveal any benefit with the use of high titer plasma (Beigel *et al.*, 2019).

However, mAbs may still hold some promise and there are currently multiple influenza-specific molecules being studied (Salazar *et al.*, 2017). MHAA4549A is a broadly neutralizing human IgG1 mAb that targets the highly conserved hemagglutinin stalk region of the influenza A virus. In one phase two study, it resulted in significant improvements in symptoms and viral (McBride *et al.*, 2017). However, though it was safe and well-tolerated, no such benefits were seen in two subsequent phase two trials (Genentech, 2018, 2019). The phase 2A study of another anti-influenza monoclonal antibody, VIS410, suggested that it was well tolerated and that there may be some clinical benefit; therefore, a phase 2B trial is now planned (Visterra, 2019).

Avian Influenza

Avian influenza viruses, such as H5N1 and H7N9, are associated with high mortality, with a case fatality rate of ~40%–70% in hospitalized patients (Cowling *et al.*, 2013). Neuraminidase inhibitors decrease mortality by ~50% (Adisasmito *et al.*, 2010; Kandun *et al.*, 2008) but there are reports of emerging neuraminidase inhibitor resistance (Someya *et al.*, 2005; de Jong *et al.*, 2005). Therefore, adjunctive and alternative therapies are still required.

In 2006, convalescent plasma from an H5N1 survivor was successfully used in combination with Oseltamivir to treat a new H5N1 case (Zhou *et al.*, 2007) and another case report described similar success (Kong and Zhou, 2006). Since then, a purified polyclonal immunoglobulin F(FBF001) raised against influenza A/Vietnam/1194/2004 virus (H5N1 subtype) showed activity in animal models and safety in a phase one trial (Bal *et al.*, 2015). H5-specific monoclonal antibodies have also been studied (Chen *et al.*, 2009; Zheng *et al.*, 2011).

Mers-Cov

Middle Eastern respiratory syndrome coronavirus (MERS-CoV) is a zoonotic disease (transmitted from camel and bat reservoirs) with occasional human-to-human transmission that emerged in Saudi Arabia in 2012 (World Health Organization, 2019b). Since then, there have been 2458 cases with a fatality rate of 34% (World Health Organization, 2019b). Though cocktails of antiviral therapies are under investigation, no treatment currently exists, and both polyclonal and monoclonal immunotherapies have been explored. MERS-CoV seropositive camel serum was beneficial in mice for both pre- and post-exposure prophylaxis (Zhao *et al.*, 2015), and a polyclonal human anti-MERS coronavirus antibody produced from transchromosomal cattle (SAB-310) was safe and well-tolerated in a phase one double-blind, placebo-controlled trial (Beigel *et al.*, 2018). Several human monoclonal antibodies against MERS-CoV have also been developed (Widjaja *et al.*, 2019). Most of these antibodies target the MERS-CoV S1B receptor-binding domain (RBD). Both RBD and non-RBD antibodies are protective against MERS-CoV in animal models (Widjaja *et al.*, 2019); however, no testing in humans has been initiated.

Ebola and Other Hemorrhagic Fevers

The first reported use of immunotherapy for the treatment of the Ebola virus (EBOV) was in 1995 in Kikwit, Democratic Republic of Congo (Mupapa *et al.*, 1999). The idea was explored further, with mixed results, in animal models (Jahrling *et al.*, 2007, 1996; Dye *et al.*, 2012; Sullivan *et al.*, 2011; Qiu *et al.*, 2014; Pascal *et al.*, 2018). Then, owing partly to the scale of devastation and the absence of any therapies, multiple investigational therapies were explored during the 2014 Ebola outbreak in West Africa. The use of convalescent plasma in a small non-randomized study did not improve survival (van Griensven *et al.*, 2016). However, ZMapp, a cocktail of monoclonal antibodies targeting the surface glycoprotein of Ebola, showed promise (Davey *et al.*, 2016). 72 patients were randomized in the PREVAIL randomized controlled study. While the pre-specified statistical threshold for efficacy was not met, there was a 15% absolute reduction in death among those receiving ZMapp (Davey *et al.*, 2016). This survival was sufficiently

beneficial that ZMapp was considered the standard of care and used in the control arm of the 2018 PAmoja TuLinde Maisha (PALM [together save lives]) Trial (Mulangu *et al.*, 2019).

The PALM [together save lives] study is a randomized controlled trial that was carried out as a component of the emergency response to the EBOV outbreak in the Democratic Republic of Congo in 2018. Spearheaded by WHO, this three-arm trial compared one antiviral, Remdesivir, and two monoclonal antibody-based regimens, mAb114 and REGN-EB3, to a ZMapp control. mAb114 was identified from a panel of memory B cells from a survivor of the 1995 Democratic Republic of Congo EBOV outbreak (Corti *et al.*, 2016; Gaudinski *et al.*, 2019). mAb114 binds to a highly conserved region of the EBOV envelope glycoprotein, preventing interaction with the NPC1 receptor, thereby preventing viral entry into the host cell cytoplasm (Corti *et al.*, 2016; Gaudinski *et al.*, 2019). REGN3470-3471-3479 is a cocktail of three fully human IgG1 monoclonal antibodies (REGN3470, REGN3471, and REGN3479 or REGN-B3; each antibody in the cocktail in a 1:1:1 ratio) (Pascal *et al.*, 2018). The three antibodies all act on separate epitopes on the EBOV glycoprotein but they can do so simultaneously (Sivapalasingam *et al.*, 2018). Overall, 53% receiving Remdesivir, ~50% of patients receiving ZMap, 35% receiving mAb114, and 33.5% receiving REGN-EB3 died. This study showed that monoclonal antibodies held great promise in reducing EBOV mortality.

Other filoviruses are also potential targets for the use of immunotherapies. For Junin Virus, the causative agent of Argentinian hemorrhagic fever, immune or convalescent plasma has been the standard of care since a seminal trial in 1979 showed significant mortality benefit. However, given the costs and challenges of administering immune plasma, recent studies have explored the use of mAbs (Maiztegui *et al.*, 1979). For instance, J199, a mouse-human chimeric mAb produced in *Nicotiana benthamiana*, administered to guinea pigs provided 100% protection against JUNV six days after infection (Zeitlin *et al.*, 2016).

Lassa virus, the causative agent of Lassa fever (LF), results in 100,000 to 300,000 infections annually, with approximately 5000 deaths (Center for Disease Control and Prevention, 2019). In a famous case of passive immunization, Dr. Jordi Casals, the virologist who first isolated and identified the Lassa virus, also inadvertently infected himself with it, but survived after receiving a blood transfusion from the first known survivor of LF, nurse Lilly Lyman Pinneo (Watts, 2012). However, the use of convalescent plasma was not subsequently shown to be beneficial (McCormick *et al.*, 1986). Since 2008, the Viral Hemorrhagic Fever Consortium (Viral Hemorrhagic Fever Consortium, 2020) has studied the protective role of B cells in LF (Robinson *et al.*, 2016a) and created the largest library of huMAbs from convalescent LF patients. Animal models of LF guinea pigs (Cross *et al.*, 2016) and non-human primates (Mire *et al.*, 2017) have confirmed the potential efficacy of mAbs in this disease.

Another cocktail of three different monoclonal antibodies, FVM04, CA45, and anti-MARV mAb MR191, provided 100% post-exposure protection in guinea pigs and non-human primates against Ebola, Sudan, and Marburg viruses (Brannan *et al.*, 2019).

Other Experimental Infections

West Nile virus (Hirota *et al.*, 2012), Zika virus (Marston *et al.*, 2018), Chikungunya (Jin and Simmons, 2019), and numerous other viruses are currently being studied and monoclonal antibodies are in the early phases of development. With the recent emergence of SARS-CoV-2 immunoglobulin therapy, polyclonal immunoglobulin therapy has specifically been considered. Not only have immunoglobulins, presumably IVIG, been administered to critically ill patients (Yang *et al.*, 2020), but there are also reports of immunoglobulin therapy with SARS-CoV-2 convalescent plasma (Bloomberg, 2020).

References

- Adisasmito, W., Chan, P.K.S., Lee, N., *et al.*, 2010. Effectiveness of antiviral treatment in human influenza A(H5N1) infections: Analysis of a global patient registry. *The Journal of Infectious Diseases* 202 (8), 1154–1160. doi:10.1086/656316.
- Andabaka, T., Nickerson, J.W., Rojas-Reyes, M.X., *et al.*, 2013. Monoclonal antibody for reducing the risk of respiratory syncytial virus infection in children. *Evidence-Based Child Health: A Cochrane Review Journal* 8 (6), 2243–2376. doi:10.1002/ebch.1950.
- American Academy of Pediatrics Committee on Infectious Diseases; American Academy of Pediatrics Bronchiolitis Guidelines Committee, 2014. Updated guidance for palivizumab prophylaxis among infants and young children at increased risk of hospitalization for respiratory syncytial virus infection *Pediatrics*. 134 (2) 415–420. doi:10.1542/peds.2014-1665. Erratum in: *Pediatrics*. 134(6):1221. PMID: 25070315.
- Aranda, S.S., Polack, F.P., 2019. Prevention of pediatric respiratory syncytial virus lower respiratory tract illness: Perspectives for the next decade. *Frontiers in Immunology* 10, 1006. doi:10.3389/fimmu.2019.01006. <https://www.ncbi.nlm.nih.gov/pubmed/31134078>.
- Arciuolo, R.J., Jablonski, R.R., Zucker, J.R., Rosen, J.B., 2017. Effectiveness of measles vaccination and immune globulin post-exposure prophylaxis in an outbreak setting – New York City, 2013. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 65 (11), 1843–1847. doi:10.1093/cid/cix639. <https://www.ncbi.nlm.nih.gov/pubmed/29028959>.
- Bal, C., Herbreteau, C.H., Buchy, P., *et al.*, 2015. Safety, potential efficacy, and pharmacokinetics of specific polyclonal immunoglobulin F(ab')₂ fragments against avian influenza A (H5N1) in healthy volunteers: A single-centre, randomised, double-blind, placebo-controlled, phase 1 study. *The Lancet Infectious Diseases* 15 (3), 285–292. doi:10.1016/S1473-3099(14)71072-2. <https://www.clinicalkey.es/playcontent/1-s2.0-S1473309914710722>.
- Baltazard, M., Bahmanyar, M., 1955. Field trials with rabies vaccine on persons bitten by rabid wolves. *Bulletin of the World Health Organization* 13 (5), 747. <https://www.ncbi.nlm.nih.gov/pubmed/13284556>.
- Barahona Afonso, A.F., Joao, C.M., 2016. The production processes and biological effects of intravenous immunoglobulin. *Biomolecules* 6 (1), 15. doi:10.3390/biom6010015.
- Beasley, R.P., Hwang, L.Y., Lee, G.C., *et al.*, 1983. Prevention of perinatally transmitted hepatitis B virus infections with hepatitis B immune globulin and hepatitis B vaccine. *Lancet* 2 (8359), 1099. <https://www.ncbi.nlm.nih.gov/pubmed/6138642>.
- Beigel, J.H., Aga, E., Elie-Turenne, M., *et al.*, 2019. Anti-influenza immune plasma for the treatment of patients with severe influenza A: A randomised, double-blind, phase 3 trial. *The Lancet Respiratory Medicine* 7 (11), 941–950. doi:10.1016/S2213-2600(19)30199-7.
- Beigel, J.H., Tebas, P., Elie-Turenne, M., *et al.*, 2017. Immune plasma for the treatment of severe influenza: An open-label, multicentre, phase 2 randomised study. *The Lancet Respiratory Medicine* 5 (6), 500–511. doi:10.1016/S2213-2600(17)30174-1. <https://www.clinicalkey.es/playcontent/1-s2.0-S2213260017301741>.

- Beigel, J.H., Voell, J., Kumar, P., *et al.*, 2018. Safety and tolerability of a novel, polyclonal human anti-MERS coronavirus antibody produced from transchromosomal cattle: A phase 1 randomised, double-blind, single-dose-escalation study. *The Lancet Infectious Diseases* 18 (4), 410–418. doi:10.1016/S1473-3099(18)30002-1.
- Bennett, J.E., Dolin, R., Blaser, M.J., 2014. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. Saint Louis: Elsevier, pp. 1439–1468. https://ebookcentral.proquest.com/lib/SITE_ID/detail.action?docID=2010635.
- Bialek, S.R., Seward, J.F., Marin, M., 2013. Updated recommendations for use of VariZIG – United States, 2013. *Morbidity and Mortality Weekly Report* 62 (28), 574–576. <https://www.jstor.org/stable/24852326>.
- Bigham, M., Murti, M., Fung, C., *et al.*, 2017. Estimated protective effectiveness of intramuscular immune serum globulin post-exposure prophylaxis during a measles outbreak in British Columbia, Canada, 2014. *Vaccine* 35 (20), 2723–2727. doi:10.1016/j.vaccine.2017.03.069. <https://www.clinicalkey.es/playcontent/1-s2.0-S0264410X17303961>.
- Bloomberg, 2020. China seeks plasma from recovered patients as virus treatment. Available at: <https://time.com/5784286/covid-19-china-plasma-treatment/>. (Updated).
- Bovier, P.A., Bock, J., Ebengo, T.F., *et al.*, 2010. Predicted 30-year protection after vaccination with an aluminum-free virosomal hepatitis A vaccine. *Journal of Medical Virology* 82 (10), 1629–1634. doi:10.1002/jmv.21883. <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.21883>.
- Brannan, J.M., He, S., Howell, K.A., *et al.*, 2019. Post-exposure immunotherapy for two ebolaviruses and marburg virus in nonhuman primates. *Nature Communications* 10 (1), 105. doi:10.1038/s41467-018-08040-w. <https://www.ncbi.nlm.nih.gov/pubmed/30631063>.
- Brunell, P.A., Ross, A., Miller, L.H., Kuo, B., 1969. Prevention of varicella by zoster immune globulin. *The New England Journal of Medicine* 280 (22), 1191–1194. doi:10.1056/NEJM196905292802201.
- Carbonell-Estrany, X., Simoes, E.A.F., Dagan, R., *et al.*, 2010. Motavizumab for prophylaxis of respiratory syncytial virus in high-risk children: A noninferiority trial. *Pediatrics* 125 (1), e35–e51. doi:10.1542/peds.2008-1036.
- Casadevall, A., Dadachova, E., Pirofski, L.A., 2004. Passive antibody therapy for infectious diseases. *Nature Reviews Microbiology* 2 (9), 695–703. doi:10.1038/nrmicro974.
- Caskey, M., Klein, F., Nussenzweig, M.C., 2019. Broadly neutralizing anti-HIV-1 monoclonal antibodies in the clinic. *Nature Medicine* 25 (4), 547–553. doi:10.1038/s41591-019-0412-8. <https://www.ncbi.nlm.nih.gov/pubmed/30936546>.
- CDC, 2020. MMR and MMRV vaccine composition and dosage. Available at: <https://www.cdc.gov/vaccines/vpd/mmr/hcp/about.html>.
- Center for Disease Control and Prevention, 2019. Lassa fever. Available at: <https://www.cdc.gov/vhf/lassa/index.html>. (Updated).
- Chen, Y., Qin, K., Wu, W.L., *et al.*, 2009. Broad cross-protection against H5N1 avian influenza virus infection by means of monoclonal antibodies that map to conserved viral epitopes. *The Journal of Infectious Diseases* 199 (1), 49–58. doi:10.1086/594374. <https://www.jstor.org/stable/40254375>.
- Cono, J., Casey, C.G., Bell, D.M., 2003. Smallpox vaccination and adverse reactions: Guidance for clinicians. *Morbidity and Mortality Weekly Report* 52 (RR-4), 1–28. <https://www.jstor.org/stable/42000789>.
- Corti, D., Misasi, J., Mulangu, S., *et al.*, 2016. Protective monotherapy against lethal ebola virus infection by a potentially neutralizing antibody. *Science* 351 (6279), 1339–1342. doi:10.1126/science.1252224. <https://www.ncbi.nlm.nih.gov/pubmed/26917593>.
- Cowling, B.J., Jin, L., Lau, E.H., *et al.*, 2013. Comparative epidemiology of human infections with avian influenza A H7N9 and H5N1 viruses in China: A population-based study of laboratory-confirmed cases. *The Lancet* 382 (9887), 129–137. doi:10.1016/S0140-6736(13)61171-X. <https://www.clinicalkey.es/playcontent/1-s2.0-S014067361361171X>.
- Cross, R.W., Mire, C.E., Branco, L.M., *et al.*, 2016. Treatment of lassa virus infection in outbred guinea pigs with first-in-class human monoclonal antibodies. *Antiviral Research* 133, 218–222. doi:10.1016/j.antiviral.2016.08.012.
- Crowell, T.A., Colby, D.J., Pinyakorn, S., *et al.*, 2019. Safety and efficacy of VRC01 broadly neutralising antibodies in adults with acutely treated HIV (RV397): A phase 2, randomised, double-blind, placebo-controlled trial. *The Lancet HIV* 6 (5), e297–e306. doi:10.1016/S2352-3018(19)30053-0.
- Cuthbert, J.A., 2001. Hepatitis A: Old and new. *Clinical Microbiology Reviews* 14 (3), 642. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=88993&tool=pmcentrez&rendertype=abstract>.
- Davey, R.T., Fernández-Cruz, E., Markowitz, N., *et al.*, 2019. Anti-influenza hyperimmune intravenous immunoglobulin for adults with influenza A or B infection (FLU-IVIG): A double-blind, randomised, placebo-controlled trial. *The Lancet Respiratory Medicine* 7 (11), 951–963. doi:10.1016/S2213-2600(19)30253-X.
- Davey, J., Richard, T., Dodd, L., Proschan, M.A., *et al.*, 2016. A randomized, controlled trial of ZMapp for ebola virus infection. *The New England Journal of Medicine* 375 (15), 1448–1456. doi:10.1056/NEJMoa1604330.
- de Jong, M.D., Thanh, T.T., Khanh, T.H., *et al.*, 2005. Oseltamivir resistance during treatment of influenza A (H5N1) infection. *The New England Journal of Medicine* 353 (25), 2667–2672. doi:10.1056/NEJMoa054512. <http://content.nejm.org/cgi/content/abstract/353/25/2667>.
- Dye, J.M., Herbert, A.S., Kuehne, A.I., *et al.*, 2012. Postexposure antibody prophylaxis protects nonhuman primates from filovirus disease. *Proceedings of the National Academy of Sciences of the United States of America* 109 (13), 5034–5039. doi:10.1073/pnas.1200409109. <https://www.jstor.org/stable/41588416>.
- Eke, A.C., Eleje, G.U., Eke, U.A., Xia, Y., Liu, J., 2017. Hepatitis B immunoglobulin during pregnancy for prevention of mother-to-child transmission of hepatitis B virus. *The Cochrane Database of Systematic Reviews* 2 (2), CD008545. doi:10.1002/14651858.CD008545.pub2. <https://www.ncbi.nlm.nih.gov/pubmed/28188612>.
- Emu, B., Fessel, J., Schrader, S., *et al.*, 2018. Phase 3 study of ibalizumab for multidrug-resistant HIV-1. *The New England Journal of Medicine* 379 (7), 645–654. doi:10.1056/NEJMoa1711460.
- Enders, G., Bolley, I., Miller, E., Cradock-Watson, J., Ridehalgh, M., 1994. Consequences of varicella and herpes zoster in pregnancy: Prospective study of 1739 cases. *The Lancet* 343 (8912), 1548–1551. doi:10.1016/S0140-6736(94)92943-2.
- Fan, L., Owusu-Edusei, K., Schillie, S.F., Murphy, T.V., 2016. Cost-effectiveness of active-passive prophylaxis and antiviral prophylaxis during pregnancy to prevent perinatal hepatitis B virus infection. *Hepatology* 63 (5), 1471–1480. doi:10.1002/hep.28310. <https://onlinelibrary.wiley.com/doi/abs/10.1002/hep.28310>.
- FDA, 2018. FDA approves new HIV treatment for patients who have limited treatment options. Available from: <https://www.fda.gov/news-events/press-announcements/fda-approves-new-hiv-treatment-patients-who-have-limited-treatment-options>.
- Feltes, T.F., Cabalka, A.K., Meissner, H.C., *et al.*, 2003. Palivizumab prophylaxis reduces hospitalization due to respiratory syncytial virus in young children with hemodynamically significant congenital heart disease. *The Journal of Pediatrics* 143 (4), 532–540. doi:10.1067/S0022-3476(03)00454-2.
- Feltes, T.F., Sondheimer, H.M., Tulloh, R.M.R., *et al.*, 2011. A randomized controlled trial of motavizumab versus palivizumab for the prophylaxis of serious respiratory syncytial virus disease in children with hemodynamically significant congenital heart disease. *Pediatric Research* 70 (2), 186–191. doi:10.1203/PDR.0b013e318220a553. <https://www.ncbi.nlm.nih.gov/pubmed/21522037>.
- Florescu, D., Abu-Elmagd, K., Mercer, D., Qiu, F., Kalil, A., 2014. An international survey of cytomegalovirus prevention and treatment practices in intestinal transplantation. *Transplantation Journal* 97 (1), 78–82. doi:10.1097/TP.0b013e3182a6baa2. <https://www.ncbi.nlm.nih.gov/pubmed/24092376>.
- Freund, N.T., Wang, H., Scharf, L., *et al.*, 2017. Coexistence of potent HIV-1 broadly neutralizing antibodies and antibody-sensitive viruses in a viremic controller. *Science Translational Medicine* 9 (373), eal2144. doi:10.1126/scitranslmed.aal2144. <https://www.narcis.nl/publication/RecordID/oi:pure.amc.nl:publications%2F7932974d-36e1-4c7b-88bd-d8923b3f88bf>.
- Fulginiti, V.A., Winograd, L.A., Jackson, M., Ellis, P., 1965. Therapy of experimental vaccinal keratitis: Effect of idoxuridine and VIG. *Archives of Ophthalmology* 74 (4), 539–544. doi:10.1001/archophth.1965.00970040541019.
- Funkhouser, W.L., 1948. The use of serum gamma globulin antibodies to control chicken pox in a convalescent hospital for children. *The Journal of Pediatrics* 32 (3), 257–259. doi:10.1016/S0022-3476(48)80028-4.
- Gaudinski, M.R., Coates, E.E., Novik, L., *et al.*, 2019. Safety, tolerability, pharmacokinetics, and immunogenicity of the therapeutic monoclonal antibody mAb114 targeting ebola virus glycoprotein (VRC 608): An open-label phase 1 study. *The Lancet* 393 (10174), 889–898. doi:10.1016/S0140-6736(19)30036-4.
- Genentech, I., 2018. A study of MHAA4549A in combination with oseltamivir versus oseltamivir in participants with severe influenza A infection. Available at: <https://clinicaltrials.gov/ct2/show/results/NCT02293863?View=results>. Updated.

- Genentech, I., 2019. A study of MHAA4549A in combination with oseltamivir versus oseltamivir in participants with severe influenza A infection. Available at: <https://clinicaltrials.gov/ct2/show/NCT02623322>. (Updated).
- Gogtay, N.J., Munshi, R., Ashwath Narayana, D.H., *et al.*, 2018. Comparison of a novel human rabies monoclonal antibody to human rabies immunoglobulin for postexposure prophylaxis: A phase 2/3, randomized, single-blind, noninferiority, controlled study. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 66 (3), 387–395. doi:10.1093/cid/cix791. <https://www.ncbi.nlm.nih.gov/pubmed/29020321>.
- Good, R.A., Lorenz, E., 1991. Historic aspects of intravenous immunoglobulin therapy. *Cancer* 68 (6), 1415–1421. doi:10.1002/1097-0142(19910915)68:6+3.0.co;2-0.
- Graham, B.S., Ambrosino, D.M., 2015. History of passive antibody administration for prevention and treatment of infectious diseases. *Current Opinion in HIV and AIDS* 10 (3), 129–134. doi:10.1097/coh.0000000000000154.
- Groothuis, J.R., Simoes, E.A.F., 1993. Immunoprophylaxis and immunotherapy: Role in the prevention and treatment of respiratory syncytial virus. *International Journal of Antimicrobial Agents* 2 (2), 97–103. doi:10.1016/0924-8579(93)90047-9.
- Gruell, H., Klein, F., 2018. Antibody-mediated prevention and treatment of HIV-1 infection. *Retrovirology* 15 (1), 73. doi:10.1186/s12977-018-0455-9. <https://www.ncbi.nlm.nih.gov/pubmed/30445968>.
- Habel, K., Koprowski, H., 1955. Laboratory data supporting the clinical trial of antirabies serum in persons bitten by a rabid wolf. *Bulletin of the World Health Organization* 13 (5), 773–779. https://www.openaire.eu/search/publication?articleId=od_____267::c5364e010114352cbd2aeb44bbb93e25.
- Henry, B., Baclic, O., 2017. Summary of the NACI update on the recommended use of hepatitis B vaccine. *Canada Communicable Disease Report* 43 (5), 104–106. doi:10.14745/ccdr.v43i05a04. <https://search.proquest.com/docview/1901645176>.
- Hens, N., Ghebretinsae, A.H., Hardt, K., Damme, P., Herck, K., 2014. Model based estimates of long-term persistence of inactivated hepatitis A vaccine-induced antibodies in adults. *Vaccine* 32 (13), doi:10.1016/J.VACCINE.2013.10.088. https://www.openaire.eu/search/publication?articleId=od_____2097::1b8126de5a02a1453063436a9927ddc8.
- Hirota, J., Shimizu, S., Shibahara, T., *et al.*, 2012. Development of monoclonal antibodies to west nile virus and their application in immunohistochemistry. *Clinical and Vaccine Immunology* 19 (11), 1853–1858. doi:10.1128/CDVI.00492-12. <http://cvi.asm.org/content/19/11/1853.abstract>.
- Hopkins, R.J., Lane, J.M., 2004. Clinical efficacy of intramuscular vaccinia immune globulin: A literature review. *Clinical Infectious Diseases* 39 (6), 819–826. doi:10.1086/422999. <https://www.jstor.org/stable/4536755>.
- Human Rabies Prevention—United States, 2008. Recommendations of the advisory committee on immunization practices. *Morbidity and Mortality Weekly Report* 2008 (57), 1–28. <https://search.proquest.com/docview/70755122>.
- Hung, I.F., To, K.K.W., Lee, C.-K., *et al.*, 2011. Convalescent plasma treatment reduced mortality in patients with severe pandemic influenza A (H1N1) 2009 virus infection. *Clinical Infectious Diseases* 52 (4), 447–456. doi:10.1093/cid/ciq106. <https://www.jstor.org/stable/29777320>.
- Hung, I.F., To, K.K.W., Lee, C., *et al.*, 2013. Hyperimmune IV immunoglobulin treatment: A multicenter double-blind randomized controlled trial for patients with severe 2009 influenza A(H1N1) infection. *Chest* 144 (2), 464–473. doi:10.1378/chest.12-2907. <https://www.ncbi.nlm.nih.gov/pubmed/23450336>.
- Ishida, J.H., Patel, A., Mehta, A.K., *et al.*, 2017. Phase 2 randomized, double-blind, placebo-controlled trial of RG7667, a combination monoclonal antibody, for prevention of cytomegalovirus infection in high-risk kidney transplant recipients. *Antimicrobial Agents and Chemotherapy* 61 (2), AAC.01794. doi:10.1128/AAC.01794-16. <https://www.ncbi.nlm.nih.gov/pubmed/27872061>.
- Izurieta, H.S., Strebel, P.M., Blake, P.A., 1997. Postlicensure effectiveness of varicella vaccine during an outbreak in a child care center. *JAMA* 278 (18), 1495–1499. doi:10.1001/jama.1997.03550180045035.
- Jahrling, P.B., Geisbert, T.W., Jaax, N.K., *et al.*, 1996. Experimental infection of cynomolgus macaques with ebola-reston filoviruses from the 1989–1990 U.S. epizootic. *Archives of Virology* 11, 115. <https://www.ncbi.nlm.nih.gov/pubmed/8800793>.
- Jahrling, P.B., Geisbert, J.B., Swearingen, J.R., Larsen, T., Geisbert, T.W., 2007. Ebola hemorrhagic fever: Evaluation of passive immunotherapy in nonhuman primates. *The Journal of Infectious Diseases* 196 (s2), S400–S403. doi:10.1086/520587. <https://www.jstor.org/stable/30086782>.
- Jin, J., Simmons, G., 2019. Antiviral functions of monoclonal antibodies against chikungunya virus. *Viruses* 11 (4), 305. doi:10.3390/v11040305. <https://www.ncbi.nlm.nih.gov/pubmed/30925717>.
- Kandun, I.N., Tresnaningsih, E., Purba, W.H., *et al.*, 2008. Factors associated with case fatality of human H5N1 virus infections in Indonesia: A case series. *The Lancet* 372 (9640), 744–749. doi:10.1016/S0140-6736(08)61125-3. <https://www.clinicalkey.es/playcontent/1-s2.0-S0140673608611253>.
- Kaufmann, S.H., 2017. PMC5347343; remembering emil von behring: From tetanus treatment to antibody cooperation with phagocytes. *mBio* 8 (1), doi:10.1128/mBio.00117-17.
- Khawplod, P., Wilde, H., Chomchey, P., *et al.*, 1996. What is an acceptable delay in rabies immune globulin administration when vaccine alone had been given previously? *Vaccine* 14 (5), 389–391. doi:10.1016/0264-410X(95)00213-K.
- Kong, L.K., Zhou, B.P., 2006. Successful treatment of avian influenza with convalescent plasma. *Hong Kong Medical Journal* 12 (6), 489. <https://www.ncbi.nlm.nih.gov/pubmed/17148811>.
- Koren, G., Money, D., Boucher, M., *et al.*, 2002. Serum concentrations, efficacy, and safety of a new, intravenously administered varicella zoster immune globulin in pregnant women. *The Journal of Clinical Pharmacology* 42 (3), 267–274. doi:10.1177/0091270022201283. <https://onlinelibrary.wiley.com/doi/abs/10.1177/0091270022201283>.
- Kotton, C., Kumar, D., Caliendo, A., *et al.*, 2018. The third international consensus guidelines on the management of cytomegalovirus in solid organ transplantation. *Transplantation* 102 (6), 900–931. doi:10.1097/TP.0000000000002191. <https://www.ncbi.nlm.nih.gov/pubmed/29596116>.
- Lee, C., Gong, Y., Brok, J., Boxall, E.H., Gluud, C., 2006. Effect of hepatitis B immunisation in newborn infants of mothers positive for hepatitis B surface antigen: Systematic review and meta-analysis. *British Medical Journal* 332 (7537), 328–336. <https://search.proquest.com/docview/19737724>.
- Levin, M.J., Duchon, J.M., Swamy, G.K., Gershon, A.A., 2019. Varicella zoster immune globulin (VARIZIG) administration up to 10 days after varicella exposure in pregnant women, immunocompromised participants, and infants: Varicella outcomes and safety results from a large, open-label, expanded-access program. *PLoS One* 14 (7), e0217749. doi:10.1371/journal.pone.0217749. <https://search.proquest.com/docview/2251782470>.
- Levin, M., Nelson, W., Preblud, S., 1986. *Clinical Trials With Varicella-Zoster Immunoglobulins*. Academic Press Inc., Ltd.
- Link-Gelles, R., Hofmeister, M.G., Nelson, N.P., 2018. Use of hepatitis A vaccine for post-exposure prophylaxis in individuals over 40 years of age: A systematic review of published studies and recommendations for vaccine use. *Vaccine* 36 (20), 2745–2750. doi:10.1016/j.vaccine.2018.04.015.
- Liu, J.P., Nikolova, D., Fei, Y., 2009. Immunoglobulins for preventing hepatitis A. *The Cochrane Database of Systematic Reviews*. 2), CD004181. doi:10.1002/14651858.CD004181.pub2. <https://www.ncbi.nlm.nih.gov/pubmed/19370595>.
- Luke, T.C., Kilbane, E.M., Jackson, J.L., Hoffman, S.L., 2006. Meta-analysis: Convalescent blood products for spanish influenza pneumonia: A future H5N1 treatment? *Annals of Internal Medicine* 145 (8), 599–609. doi:10.7326/0003-4819-145-8-200610170-00139.
- Luna, M.S., Manzoni, P., Paes, B., *et al.*, 2018. Expert consensus on palivizumab use for respiratory syncytial virus in developed countries. *Paediatric Respiratory Reviews* 33, 35–44. doi:10.1016/j.prrv.2018.12.001.
- Mac, S., Sumner, A., Duchesne-Belanger, S., *et al.*, 2019. Cost-effectiveness of palivizumab for respiratory syncytial virus: A systematic review. *Pediatrics* 143 (5), e20184064. doi:10.1542/peds.2018-4064. <https://www.ncbi.nlm.nih.gov/pubmed/31040196>.
- Maertens, J., Logan, A.C., Jang, J., *et al.*, 2020. Anti-human cytomegalovirus monoclonal antibodies for prophylaxis in hematopoietic cell transplantation: A phase 2 study. *Antimicrobial Agents and Chemotherapy* 64 (4), doi:10.1128/AAC.02467-19. <https://www.ncbi.nlm.nih.gov/pubmed/32015031>.
- Maiztegui, J.I., Fernandez, N.J., de Damielano, A.J., 1979. Efficacy of immune plasma in treatment of argentine haemorrhagic fever and association between treatment and a late neurological syndrome. *Lancet* 2 (8154), 1216. <https://www.ncbi.nlm.nih.gov/pubmed/92624>.
- Marston, H.D., Paules, C.I., Fauci, A.S., 2018. Monoclonal antibodies for emerging infectious diseases – Borrowing from history. *The New England Journal of Medicine* 378 (16), 1469–1472. doi:10.1056/NEJMp1802256.

- Matysiak-Klose, D., Santibanez, S., Schwerdtfeger, C., *et al.*, 2018. Post-exposure prophylaxis for measles with immunoglobulins revised recommendations of the standing committee on vaccination in germany. *Vaccine* 36 (52), 7916–7922. doi:10.1016/j.vaccine.2018.10.070.
- McBride, J.M., Lim, J.J., Burgess, T., *et al.*, 2017. Phase 2 randomized trial of the safety and efficacy of MHAA4549A, a broadly neutralizing monoclonal antibody, in a human influenza A virus challenge model. *Antimicrobial Agents and Chemotherapy* 61 (11), doi:10.1128/AAC.01154-17. <https://www.ncbi.nlm.nih.gov/pubmed/28807912>.
- McCormick, J.B., King, I.J., Webb, P.A., *et al.*, 1986. Lassa fever. *The New England Journal of Medicine* 314 (1), 20–26. doi:10.1056/NEJM198601023140104.
- McDaniel, J.R., DeKosky, B.J., Tanno, H., Ellington, A.D., Georgiou, G., 2016. Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nature Protocols* 11 (3), 429–442. doi:10.1038/nprot.2016.024. <https://www.ncbi.nlm.nih.gov/pubmed/26844430>.
- McLean, H., Fiebelkorn, A.P., Temte, J.L., Wallace, G.S., 2013. Centers for disease control, prevention. Prevention of measles, rubella, congenital rubella syndrome, and mumps: Summary recommendations of the advisory committee on immunization practices (ACIP).
- Meyer, H., Ehmann, R., Smith, G.L., 2020. Smallpox in the post-eradication era. *Viruses* 12 (2), 138. doi:10.3390/v12020138. <https://www.ncbi.nlm.nih.gov/pubmed/31991671>.
- Mire, C.E., Cross, R.W., Geisbert, J.B., *et al.*, 2017. Human-monoclonal-antibody therapy protects nonhuman primates against advanced lassa fever. *Nature Medicine* 23 (10), 1146–1149. doi:10.1038/nm.4396. <https://www.ncbi.nlm.nih.gov/pubmed/28869611>.
- Moore, H.C., de Klerk, N., Richmond, P.C., *et al.*, 2019. Effectiveness of palivizumab against respiratory syncytial virus: Cohort and case series analysis. *The Journal of Pediatrics* 214, 121–127. doi:10.1016/j.jpeds.2019.06.058.
- Mulangu, S., Dodd, L.E., Davey, R.T., *et al.*, 2019. A randomized, controlled trial of ebola virus disease therapeutics. *The New England Journal of Medicine* 381 (24), 2293–2303. doi:10.1056/NEJMoa1910993.
- Mupapa, K., Massamba, M., Kibadi, K., *et al.*, 1999. Treatment of ebola hemorrhagic fever with blood transfusions from convalescent patients. *International Scientific and Technical Committee. The Journal of infectious diseases* 179 (Suppl 1), S18–S23. <https://www.ncbi.nlm.nih.gov/pubmed/9988160>.
- Nair, H., Nokes, D.J., Gessner, B.D., *et al.*, 2010. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: A systematic review and meta-analysis. *The Lancet* 375 (9725), 1545–1555. doi:10.1016/S0140-6736(10)60206-1. <https://www.clinicalkey.es/playcontent/1-s2.0-S0140673610602061>.
- National Advisory Committee on Immunization (NACI), 2016. Updated recommendations for the use of varicella zoster immune globulin (Varig) for the prevention of varicella in at-risk patients. Available at: <https://www.canada.ca/en/public-health/services/publications/healthy-living/updated-recommendations-use-varicella-zoster-immune-globulin-varig-prevention-varicella-risk-patients.html#a4>. (Updated).
- Navarrete-Navarro, S., Aguilar-Setién, A., Avila-Figueroa, C., Hernández-Sierra, F., Santos-Preciado, J.I., 1999. Improved serological response to human diploid cell rabies vaccine when given simultaneously with antirabies hyperimmune globulin. *Archives of Medical Research* 30 (4), 332–337. doi:10.1016/S0888-0128(99)00036-6.
- Nelson, N.P., Link-Gelles, R., Hofmeister, M.G., *et al.*, 2018. Update: Recommendations of the advisory committee on immunization practices for use of hepatitis A vaccine for postexposure prophylaxis and for preexposure prophylaxis for international travel. *Morbidity and Mortality Weekly Report* 67 (43), 1216–1220. doi:10.15585/mmwr.mm6743a5. <https://www.ncbi.nlm.nih.gov/pubmed/30383742>.
- Nigro, G., Adler, S.P., La Torre, R., Best, A.M., 2005. Passive immunization during pregnancy for congenital cytomegalovirus infection. *The New England Journal of Medicine* 353 (13), 1350–1362. doi:10.1056/NEJMoa043337. <http://content.nejm.org/cgi/content/abstract/353/13/1350>.
- Nigro, G., Adler, S.P., Parruti, G., *et al.*, 2012. Immunoglobulin therapy of fetal cytomegalovirus infection occurring in the first half of pregnancy – A case-control study of the outcome in children. *The Journal of Infectious Diseases* 205 (2), 215–227. doi:10.1093/infdis/jir718. <https://www.jstor.org/stable/41417693>.
- Nusrat Homaira, W., Rawlinson, T.L., Snelling, A.J., 2014. Effectiveness of palivizumab in preventing RSV hospitalization in high risk children: A real-world perspective. *International Journal of Pediatrics* 2014, 571609–571613. doi:10.1155/2014/571609.
- Pascal, K.E., Dudgeon, D., Trefry, J.C., *et al.*, 2018. Development of clinical-stage human monoclonal antibodies that treat advanced ebola virus disease in nonhuman primates. *The Journal of Infectious Diseases* 218 (suppl_5), S612–S626. doi:10.1093/infdis/jiy285. <https://www.ncbi.nlm.nih.gov/pubmed/29860496>.
- Plotkin, S., 2012. Vaccines, sixth ed. Baltimore: The Johns Hopkins University Press, pp. 352–387. http://bvbr.bib-bvb.de/8991/F?func=service&doc_library=BVB01&local_base=BVB01&doc_number=029231748&sequence=000002&line_number=0001&func_code=DB_RECORDS&service_type=MEDIA.
- Poland, G.A., Grabenstein, J.D., Neff, J.M., 2005. The US smallpox vaccination program: A review of a large modern era smallpox vaccination implementation program. *Vaccine* 23 (17), 2078–2081. doi:10.1016/j.vaccine.2005.01.012.
- PREVENT Study Group, 1997. Reduction of respiratory syncytial virus hospitalization among premature infants and infants with bronchopulmonary dysplasia using respiratory syncytial virus immune globulin prophylaxis. the PREVENT study group. *Pediatrics* 99 (1), 93. <https://www.ncbi.nlm.nih.gov/pubmed/8989345>.
- Public Health England, 2015. Chapter 27a Respiratory syncytial virus: The green book. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/458469/Green_Book_Chapter_27a_v2_OW.PDF. (Updated).
- Public Health England, 2020. Guidelines on post-exposure prophylaxis for measles. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/814203/Guidance_for_measles_post-exposure_prophylaxis.pdf. Updated 2019. (accessed 11 February).
- Public Health England, VZIG Expert Working Group, 2019. Updated guidelines on post exposure prophylaxis (PEP) for varicella/shingles. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/812526/PHE_PEP_VZIG_guidance_for_health_professionals.pdf. (Updated).
- Qiu, X., Wong, G., Audet, J., *et al.*, 2014. Reversion of advanced ebola virus disease in nonhuman primates with ZMapp. *Nature* 514 (7520), 47–53. doi:10.1038/nature13777. <https://www.ncbi.nlm.nih.gov/pubmed/25171469>.
- RCOG Scientific Impact Paper, 2018. Congenital cytomegalovirus infection: Update on treatment: Scientific impact paper no. 56. *British Journal of Obstetrics and Gynecology* 125 (1), e1–e11. doi:10.1111/1471-0528.14836.
- Resch, B., 2017. Product review on the monoclonal antibody palivizumab for prevention of respiratory syncytial virus infection. *Human Vaccines & Immunotherapeutics* 13 (9), 2138–2149. doi:10.1080/21645515.2017.1337614. <http://www.tandfonline.com/doi/abs/10.1080/21645515.2017.1337614>.
- Revello, M.G., Lazzarotto, T., Guerra, B., *et al.*, 2014. A randomized trial of hyperimmune globulin to prevent congenital cytomegalovirus. *The New England Journal of Medicine* 370 (14), 1316–1326. doi:10.1056/NEJMoa1310214.
- Robinson, J.E., Hastie, K.M., Cross, R.W., *et al.*, 2016a. Most neutralizing human monoclonal antibodies target novel epitopes requiring both lassa virus glycoprotein subunits. *Nature Communications* 7 (1), 11544. doi:10.1038/ncomms11544. <https://www.ncbi.nlm.nih.gov/pubmed/27161536>.
- Robinson, K.A., Odelola, O.A., Saldana, I.J., 2016b. Palivizumab for prophylaxis against respiratory syncytial virus infection in children with cystic fibrosis. *The Cochrane Database of Systematic Reviews* 7 (7), CD007743. doi:10.1002/14651858.CD007743.pub6. <https://www.ncbi.nlm.nih.gov/pubmed/27439110>.
- Robinson, J.L., Le Saux, N., 2015. Preventing hospitalizations for respiratory syncytial virus infection. *Paediatrics & Child Health* 20 (6), 321–326. doi:10.1093/pch/20.6.321. <https://www.ncbi.nlm.nih.gov/pubmed/26435673>.
- Ross, A.H., Lenchner, E., Reitman, G., 1962. Modification of chicken pox in family contacts by administration of gamma globulin. *The New England Journal of Medicine* 267 (8), 369–376. doi:10.1056/NEJM196208232670801.
- Salazar, G., Zhang, N., Fu, T.M., An, Z., 2017. Antibody therapies for the prevention and treatment of viral infections. *npj Vaccines* 2, 19. doi:10.1038/s41541-017-0019-3.
- Sappapapu, G., Fernandez, E., Kose, N., *et al.*, 2016. Neutralizing human antibodies prevent zika virus replication and fetal disease in mice. *Nature* 540 (7633), 443–447. doi:10.1038/nature20564. <https://www.ncbi.nlm.nih.gov/pubmed/27819683>.
- Schaeffer, M., Toomey, J.A., 1948. Failure of gamma globulin to prevent varicella. *The Journal of Pediatrics* 33 (6), 749–752. doi:10.1016/S0022-3476(48)80085-5.
- Schillie, S., Vellozzi, C., Reingold, A., *et al.*, 2018. Prevention of hepatitis B virus infection in the United States: Recommendations of the advisory committee on immunization practices. *Morbidity and Mortality Weekly Report* 67 (1), 1–31. doi:10.15585/mmwr.mm6701a1. <https://www.ncbi.nlm.nih.gov/pubmed/29939980>.
- Schillie, S., Walker, T., Veselsky, S., *et al.*, 2015. Outcomes of infants born to women infected with hepatitis B. *Pediatrics* 135 (5), e1141–e1147. doi:10.1542/peds.2014-3213. <https://www.ncbi.nlm.nih.gov/pubmed/25896839>.

- Schulz, U., Solidoro, P., Müller, V., *et al.*, 2016. CMV immunoglobulins for the treatment of CMV infections in thoracic transplant recipients. *Transplantation* 100 (Suppl 3), S5–S10. doi:10.1097/TP.0000000000001097. <https://www.ncbi.nlm.nih.gov/pubmed/26900992>.
- Simoes, E.A., Sondheimer, H.M., Top, J.F.H., *et al.*, 1998. Respiratory syncytial virus immune globulin for prophylaxis against respiratory syncytial virus disease in infants and children with congenital heart disease. the cardiac study group. *The Journal of Pediatrics* 133 (4), 492. <https://www.ncbi.nlm.nih.gov/pubmed/9787686>.
- Sivapalasingam, S., Kamal, M., Slim, R., *et al.*, 2018. Safety, pharmacokinetics, and immunogenicity of a co-formulated cocktail of three human monoclonal antibodies targeting ebola virus glycoprotein in healthy adults: A randomised, first-in-human phase 1 study. *The Lancet Infectious Diseases* 18 (8), 884–893. doi:10.1016/S1473-3099(18)30397-9.
- Snydman, D.R., Werner, B.G., Dougherty, N.N., *et al.*, 1993. Cytomegalovirus immune globulin prophylaxis in liver transplantation. A randomized, double-blind, placebo-controlled trial. *Annals of Internal Medicine* 119 (10), 984. doi:10.7326/0003-4819-119-10-199311150-00004. <https://www.ncbi.nlm.nih.gov/pubmed/8214995>.
- Snydman, D.R., Werner, B.G., Heinze-Lacey, B., *et al.*, 1987. Use of cytomegalovirus immune globulin to prevent cytomegalovirus disease in renal-transplant recipients. *The New England Journal of Medicine* 317 (17), 1049–1054. <https://search.proquest.com/docview/81032719>.
- Someya, K., Goto, H., Pham, N.D., *et al.*, 2005. Avian flu isolation of drug-resistant H5N1 virus. *Nature* 437 (7062), 1108. doi:10.1038/4371108a.
- Sparrow, E., Torvaldsen, S., Newall, A.T., *et al.*, 2019. Recent advances in the development of monoclonal antibodies for rabies post exposure prophylaxis: A review of the current status of the clinical development pipeline. *Vaccine* 37, A132–A139. doi:10.1016/j.vaccine.2018.11.004.
- Sullivan, N.J., Hensley, L., Asiedu, C., *et al.*, 2011. CD8⁺ cellular immunity mediates rAd5 vaccine protection against ebola virus infection of nonhuman primates. *Nature Medicine* 17 (9), 1128–1131. doi:10.1038/nm.2447. <https://www.ncbi.nlm.nih.gov/pubmed/21857654>.
- Swamy, G.K., Dotters-Katz, S.K., 2019. Safety and varicella outcomes after varicella zoster immune globulin administration in pregnancy. *American Journal of Obstetrics and Gynecology* 221 (6), 655–656. doi:10.1016/j.ajog.2019.07.003.
- Terrault, N.A., Lok, A.S.F., McMahon, B.J., *et al.*, 2018. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology* 67 (4), 1560–1599. doi:10.1002/hep.29800. <https://onlinelibrary.wiley.com/doi/abs/10.1002/hep.29800>.
- The George Washington University Biostatistics Center, 2020. A randomized trial to prevent congenital cytomegalovirus (CMV). Available at: <https://clinicaltrials.gov/ct2/show/NCT01376778>. Updated 2019. (accessed 11 February).
- The IMPact-RSV Study Group, 1998. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. *Pediatrics* 102 (3), 531–537. doi:10.1542/peds.102.3.531.
- Trotta, M., Borch, B., Nicolai, A., *et al.*, 2018. Epidemiology, management and outcome of varicella in pregnancy: A 20-year experience at the tuscany reference centre for infectious diseases in pregnancy. *Infection* 46 (5), 693–699. doi:10.1007/s15010-018-1150-4. <https://www.ncbi.nlm.nih.gov/pubmed/29766472>.
- Tunis, M.C., Salvadori, M.I., Dubey, V., Baclic, O., 2018. Updated NACI recommendations for measles post-exposure prophylaxis. *Canada Communicable Disease Report* 44 (9), 226–230. doi:10.14745/ccdr.v44i09a07. <https://www.ncbi.nlm.nih.gov/pubmed/31015814>.
- Valantine, H.A., Luikart, H., Doyle, R., *et al.*, 2001. Impact of cytomegalovirus hyperimmune globulin on outcome after cardiopulmonary transplantation: A comparative study of combined prophylaxis with CMV hyperimmune globulin plus ganciclovir versus ganciclovir alone. *Transplantation* 72 (10), 1647–1652. doi:10.1097/00007890-200111270-00012. <https://www.ncbi.nlm.nih.gov/pubmed/11726825>.
- van Griensven, J., Edwards, T., de Lamballerie, X., *et al.*, 2016. Evaluation of convalescent plasma for ebola virus disease in guinea. *The New England Journal of Medicine* 374 (1), 33–42. doi:10.1056/NEJMoa1511812.
- Victor, J.C., Monto, A.S., Surdina, T.Y., *et al.*, 2007. Hepatitis A vaccine versus immune globulin for postexposure prophylaxis. *The New England Journal of Medicine* 357 (17), 1685–1694. doi:10.1056/NEJMoa070546. <http://content.nejm.org/cgi/content/abstract/357/17/1685>.
- Viral Hemorrhagic Fever Consortium, 2020. Viral hemorrhagic fever consortium. Available at: <https://vhfc.org/>. (Updated).
- Visentin, S., Manara, R., Milanese, L., *et al.*, 2012. Early primary cytomegalovirus infection in pregnancy: Maternal hyperimmunoglobulin therapy improves outcomes among infants at 1 year of age. *Clinical Infectious Diseases* 55 (4), 497. <https://search.proquest.com/docview/1027893637>.
- Visterra, I., 2019. Study to evaluate the efficacy and safety of IV VIS410 plus oseltamivir versus oseltamivir alone in hospitalized adults with influenza A infection requiring oxygen support. Available at: <https://clinicaltrials.gov/ct2/show/NCT02623322>. (Updated).
- Walayat, S., Ahmed, Z., Martin, D., *et al.*, 2015. Recent advances in vaccination of non-responders to standard dose hepatitis B virus vaccine. *世界肝病杂志:英文版(电子版)* 7 (24), 2503–2509. doi:10.4254/wjh.v7.i24.2503. <http://lib.cqvip.com/qk/71422X/201524/87747284504849535052484850.html>.
- Watts, G., 2012. Lily lyman pinneo. *The Lancet* 380 (9853), 1552. doi:10.1016/S0140-6736(12)61871-6. <https://www.clinicalkey.es/playcontent/1-s2.0-S0140673612618716>.
- WHO/Department of Control of Neglected Tropical Diseases, 2018. WHO expert consultation on rabies: WHO TRS N°1012. Available at: https://www.who.int/rabies/resources/who_trs_1012/en/.
- Widjaja, I., Wang, C., van Haperen, R., *et al.*, 2019. Towards a solution to MERS: Protective human monoclonal antibodies targeting different domains and functions of the MERS-coronavirus spike glycoprotein. *Emerging Microbes & Infections* 8 (1), 516–530. doi:10.1080/22221751.2019.1597644. <http://www.tandfonline.com/doi/abs/10.1080/22221751.2019.1597644>.
- Wittek, R., 2006. Vaccinia immune globulin: Current policies, preparedness, and product safety and efficacy. *International Journal of Infectious Diseases* 10 (3), 193–201. doi:10.1016/j.ijid.2005.12.001.
- World Health Organization, 1979. The global eradication of smallpox: Final report of the global commission for the certification of smallpox eradication.
- World Health Organization, 2016. Rabies monoclonal antibodies post exposure prophylaxis. Available at: https://www.who.int/rabies/resources/Summary-rabies-mAbs-for-Web_Dec2016.pdf?Ua=1. (Updated).
- World Health Organization (WHO), 2017. Secretariat, SAGE Working Group on Rabies vaccines and immunoglobulins. Proposed revision of the policy on rabies vaccines and rabies immunoglobulins. Available at: https://www.who.int/immunization/sage/meetings/2017/october/1_Background_paper_WG_RABIES_final.pdf. (Updated).
- World Health Organization, 2019a. Measles. Available at: <https://www.who.int/news-room/fact-sheets/detail/measles>. Updated 2020. (accessed September 9).
- World Health Organization, 2019b. Middle East respiratory syndrome coronavirus (MERS-CoV). Available at: <https://www.who.int/emergencies/mers-cov/en/>. (Updated).
- Yang, X., Yu, Y., Xu, J., *et al.*, 2020. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study. *Lancet Respiratory Medicine* 8 (5), 475–481. <https://www.thelancet.com/action/showPdf?pii=S2213-2600%2820%2930079-5>.
- Yinon, Y., Farine, D., Yudin, M.H., 2018. No. 240-cytomegalovirus infection in pregnancy. *Journal of Obstetrics and Gynaecology Canada* 40 (2), e134–e141. doi:10.1016/j.jogc.2017.11.018.
- Young, M.K., Nimmo, G.R., Cripps, A.W., Jones, M.A., 2014. Post-exposure passive immunisation for preventing measles. *The Cochrane Database of Systematic Reviews* 4, CD010056. doi:10.1002/14651858.CD010056.pub2. <https://www.ncbi.nlm.nih.gov/pubmed/24687262>.
- Zaia, J.A., Levin, M.J., Preblud, S.R., *et al.*, 1983. Evaluation of varicella-zoster immune globulin: Protection of immunosuppressed children after household exposure to varicella. *The Journal of Infectious Diseases* 147 (4), 737–743. doi:10.1093/infdis/147.4.737. <https://www.jstor.org/stable/30052644>.
- Zeitlin, L., Geisbert, J.B., Deer, D.J., *et al.*, 2016. Monoclonal antibody therapy for junin virus infection. *Proceedings of the National Academy of Sciences of the United States of America* 113 (16), 4458–4463. doi:10.1073/pnas.1600996113. <https://www.jstor.org/stable/26469346>.
- Zhao, J., Perera, R.A.P.M., Kayali, G., *et al.*, 2015. Passive immunotherapy with dromedary immune serum in an experimental animal model for middle east respiratory syndrome coronavirus infection. *Journal of Virology* 89 (11), 6117–6120. doi:10.1128/JVI.00446-15. <https://www.ncbi.nlm.nih.gov/pubmed/25787284>.
- Zheng, Q., Xia, L., Wu, W.L., *et al.*, 2011. Properties and therapeutic efficacy of broadly reactive chimeric and humanized H5-specific monoclonal antibodies against H5N1 influenza viruses. *Antimicrobial Agents and Chemotherapy* 55 (4), 1349–1357. doi:10.1128/AAC.01436-10. <http://aac.asm.org/content/55/4/1349.abstract>.
- Zhou, B., Zhong, N., Guan, Y., 2007. Treatment with convalescent plasma for influenza A (H5N1) infection. *The New England Journal of Medicine* 357 (14), 1450–1451. doi:10.1056/NEJMc070359. <http://content.nejm.org/cgi/content/abstract/357/14/1450>.

Further Reading

- Dixit, R., Herz, J., Dalton, R., Booy, R., 2016. Benefits of using heterologous polyclonal antibodies and potential applications to new and undertreated infectious pathogens. *Vaccine* 34 (9), 1152–1161.
- Ferrara, G., Zumla, A., Maeurer, M., 2012. Intravenous immunoglobulin (IVIg) for refractory and difficult-to-treat infections. *The American Journal of Medicine* 125 (10), (1036.e1,1036.e8).
- Hey, A., 2015. History and practice: Antibodies in infectious diseases. *Microbiology Spectrum* 3 (2).
- Ian Gust, A.O., 2012. Role of passive immunotherapies in managing infectious outbreaks. *Biologicals* 40 (3), 196–199.
- Pelfrene, E., Mura, M., Cavaleiro Sanches, A., Cavaleri, M., 2019. Monoclonal antibodies as anti-infective products: A promising future? *Clinical Microbiology and Infection* 25 (1), 60–64.
- Perez, E.E., Orange, J.S., Bonilla, F., *et al.*, 2017. Update on the use of immunoglobulin in human disease: A review of evidence. *Journal of Allergy and Clinical Immunology* 139 (3), S1–S46.
- Sparrow, E., Friede, M., Sheikh, M., Torvaldsen, S., 2017. Therapeutic antibodies for infectious diseases. *Bulletin of the World Health Organization* 95 (3), 235–237.

Relevant Websites

<https://www.astrazeneca.com/media-centre/press-releases/2010/AstraZeneca-discontinues-motavizumab-RSV-21122010.html#>
AstraZeneca discontinues development of motavizumab for RSV prophylaxis indication. M2 Pharma.

Vaccine Production, Safety, and Efficacy

Thomas J Brouwers, Athena Institute, VU Amsterdam, Amsterdam, The Netherlands

Bernard AM Van der Zeijst, Leiden University Medical Center, Leiden, The Netherlands

© 2021 Elsevier Ltd. All rights reserved.

Glossary

Adjuvant A substance that enhances the body's immune response to an antigen.

Clinical trial A research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes.

Correlate of protection Correlates of immunity/protection to a virus or other infectious pathogen are measurable signs that a person is immune, in the sense of being protected against becoming infected and/or developing disease. For many viruses, antibodies serve as a correlate of immunity.

Herpes zoster Shingles, also known as zoster or herpes zoster, is a viral disease characterized by a painful skin rash with blisters in a localized area. Shingles is due to a reactivation of varicella-zoster virus (VZV) in a person's body. The disease chickenpox is caused by the initial infection with VZV.

Immunization The action of making a person or animal immune to infection, typically by inoculation with a nonpathogenic version of the infectious agent. Synonymous with vaccination.

Regulatory authority National regulatory agencies responsible for ensuring that products released for public distribution (normally pharmaceuticals and biological products, such as vaccines) are evaluated properly and meet international standards of quality and safety.

Introduction

The development of vaccines is closely linked to the need to protect increasingly larger populations. This started with the transition from populations consisting of hunter-gatherers to agricultural settlements. It went on when urbanization made man even more vulnerable to contagious diseases. Presently about 55% of the world population lives in cities, a situation that is only sustainable with vaccination. Stopping vaccination would result in epidemics. The medical need to prevent (viral) infections is still the major incentive for vaccine development. But whether a vaccine will be developed is also heavily influenced by commercial considerations.

The first vaccine against smallpox ended a disease that in the century before its eradication killed 500 million people ([Fig. 1](#)). Thanks to vaccination, the average life expectancy has gone up in many countries from about 40 years in 1900 to more than 80 years today. [Table 1](#) gives an overview of available viral vaccines for human use. Additionally, many vaccines for animals have been licensed. Thanks to vaccination, smallpox was eradicated by 1977. Rinderpest, a disease that killed up to 100% of cattle during outbreaks was eradicated by 2010 and polio is very close to eradication. Despite of these successes additional vaccines are needed, some urgently. [Table 1](#) lists these vaccines.

Although there have been several "golden ages of vaccines" with important scientific and technical advances bringing vaccine development from a trial-and-error approach to "reverse vaccinology" ([Delany et al., 2014](#)), development times of vaccines have only increased. This is especially worrying when we have to develop vaccines against new viral diseases, where pre-existing immunity is completely absent. This article aims to provide an overview of the history, current status, and future of viral vaccine development and use. A full description of all aspects of viral vaccines would require many hundreds of pages ([Plotkin et al., 2018](#); [Plotkin, 2011](#)). We have selected some recently developed vaccines to demonstrate the general principles of vaccine development and provide an overview of the long road from vaccine candidates towards market entry. Lastly, several new technologies and opportunities are discussed that are expected to change the vaccine development process for the better, especially by shortening development times.

The History of Vaccination

The realization that victims of a severe infectious disease remain immune to the same disease for the rest of their lives was the basis of the first vaccine against smallpox. Actually, a similar approach was already used from about 1570 in China ([Plotkin, 2011](#)). The procedure is called variolation and entailed introducing minute amounts of virus-containing pox material into the nostril of a child. This resulted in a light form of the disease. Already, in those times, the balance between efficacy and adverse effects was evaluated. Up to 1% of variolated children died. This was considered acceptable considering the larger risk of contracting smallpox.

Vaccination differs fundamentally from variolation in that vaccines are composed of viral antigens that produce protective immunity and immune memory without causing disease. There are three classes of licensed viral vaccines: (1) Live-attenuated

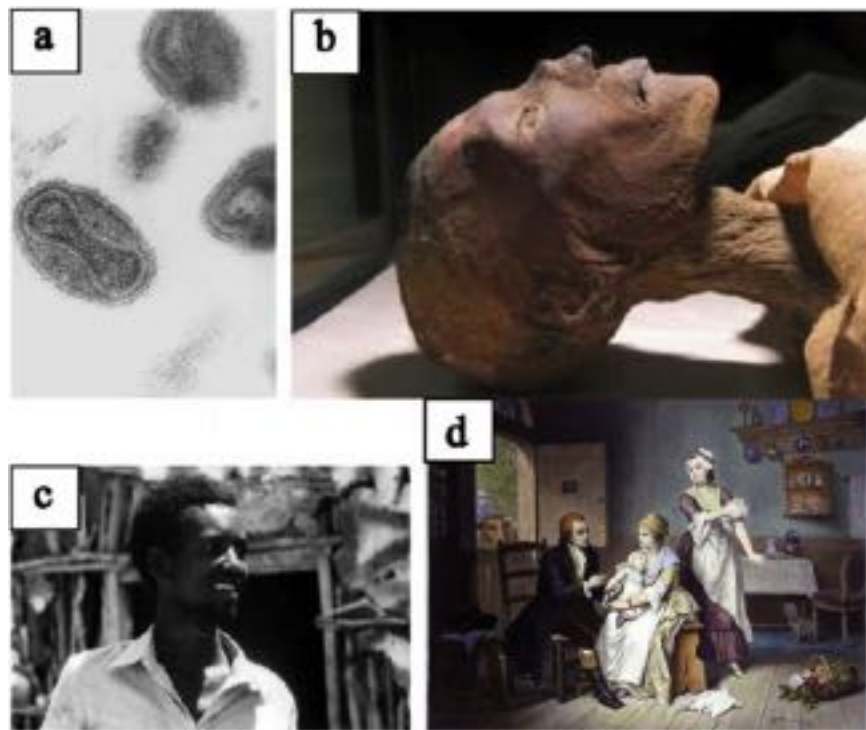


Fig. 1 Conquering smallpox. The smallpox virus (a) has been responsible for hundreds of millions of deaths. The virus was already present in ancient Egypt (b). It spread, by trade routes, wars and colonization, to all continents. Typically, 3 out of 10 people who got the disease died. Survivors were left with scars; some became blind. An eradication program organized by the WHO was successful by 1977 with the last case of naturally acquired smallpox (c). The virus is now confined to two laboratories. Eradication was possible with a vaccine discovered by Edward Jenner in 1796 (d). Sources: (a) CDC PHIL #1849, (b) and (c) WHO, (d) Wellcome Library no. 546000i.

Table 1 Overview of licensed viral vaccines and vaccines under development

<i>Licensed vaccines (Year of market introduction)</i>	<i>Vaccines that are needed (Estimated global death toll /year)</i>
Smallpox (1796)	HIV (770,000)
Rabies (1885)	Broadly protecting influenza vaccine (720,000)
Yellow fever (1930)	Hepatitis C (399,000)
Japanese encephalitis (1930)	Norovirus (200,000)
Influenza (1938) (efficacy 40%–60%)	Epstein-Barr (143,000)
Polio (1954)	Human respiratory syncytial virus (120,000)
Adenovirus (1956)	Hepatitis E (44,000)
Measles (1963)	Cytomegalovirus (N/A)
Mumps (1967)	Improved dengue vaccine (N/A)
Rubella (1969)	Herpes simplex
Varicella (1970)	Human metapneumovirus (N/A)
Hepatitis B (1981)	
Hepatitis A (1991)	
Rotavirus (2006)	
HPV (2006)	
Herpes Zoster (2006)	
Dengue (2015) (efficacy 44%, adverse events)	
Ebola (2019)	

vaccines, (2) Inactivated vaccines and (3) Non-replicating protein vaccines produced by recombinant DNA technology. An advantage of live-attenuated vaccines is that production is inexpensive and no adjuvants to boost immune responses are needed. Disadvantages are more adverse effects. These are usually mild, in rare cases severe. For example, smallpox and yellow fever vaccines result in respectively 1–8 and 1–2 deaths per million vaccinations. Another risk is that mutations may lead to regained virulence. This has been a problem for the oral polio vaccine (Amanna and Slifka, 2009). Inactivated and non-replicating vaccines

are very safe, but a downside is that alone they are often unable to generate robust immunity. These vaccines require adjuvants, substances added to vaccines to enhance the immunogenicity (Di Pasquale *et al.*, 2015). Traditionally aluminum salts, with a strong safety record, were used as adjuvant. Scientific advances have led to additional, more potent, adjuvants with a smaller safety database.

What has Happened When Vaccine Reaches the Market? The Development of Shingrix

Shingrix is the herpes zoster (shingles) vaccine. It was licensed in October 2017 in the USA and in 2018 in Europe and Japan. Licensing was based on two large clinical trials with almost 28,000 participants (Maltz and Fidler, 2019; Levin and Weinberg, 2020). Shingrix was the second herpes zoster vaccine. It addressed several limitations of Zostavax, the first vaccine. These limitations included limited protection, especially in individuals over 70 years old, a limited duration of protection, and safety risks for immunocompromised individuals (Dooling *et al.*, 2018).

The history of this successful vaccine dates back to basic research in the early 1990s. Researchers looked for promising viral antigens for a new vaccine. They opted for antigens that could induce cell-mediated immunity in addition to virus neutralizing antibodies elicited by Zostavax. Viral glycoprotein E (gE) was selected as the most suitable candidate (Vafai, 1993; Haumont *et al.*, 1996). Thus, at least 24 years passed between the beginning of research into this vaccine and its licensing.

Phases and Timelines in Vaccine Development: From Discovery to Product

A development time of several decades is typical for vaccine development which consists of separate phases that can only be carried out after each other. Fig. 2 summarizes this step-wise process.

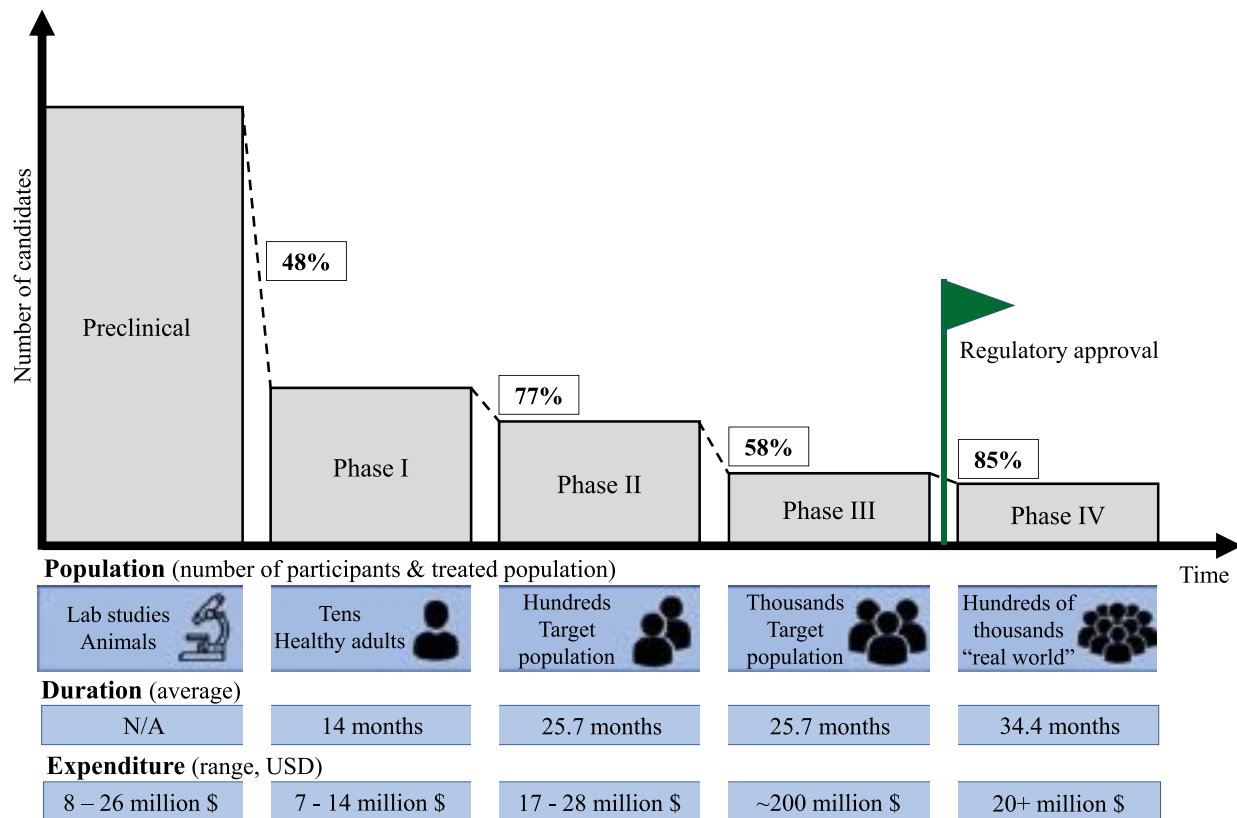


Fig. 2 Timelines and success rates in vaccine development. The transition successes between the clinical phases are from (Wong, C.H., Siah, K.W., LO, A.W., 2019. Estimation of clinical trial success rates and related parameters. *Biostatistics* 20, 273–286). The data on duration from (Dimasi, J.A., Florez, M.I., Stergiopoulos, S., *et al.*, 2020. Development times and approval success rates for drugs to treat infectious diseases. *Clinical Pharmacology & Therapeutics* 107, 324–332) and the cost estimate up to phase 2 from (Gouglas, D., Le, T.T., Henderson, K., *et al.*, 2018. Estimating the cost of vaccine development against epidemic infectious diseases: A cost minimisation study. *The Lancet Global Health* 6, (e1386–e1396)).

Preclinical Studies

The starting point for vaccine development is the medical need to prevent disease. The development begins with basic research to identify the virus responsible for the diseases and, if possible, the antigen(s) that elicit a protective immune response. In addition, attempts are made to determine which immune responses are protective (antibodies, T cells, for example). This is all carried out in the laboratory. Blood samples are often used to study immune responses *ex vivo*. In those cases where an animal model is available, candidate vaccines can be tested for protection after challenge with virulent virus. Finally, toxicity data are collected. The preclinical stage ends with the establishment of a production process.

In Human Studies

After the preclinical studies have come to a satisfactory end, the vaccine is tested in humans (or animals for veterinary vaccines). ‘First-in-human’ studies and all other clinical studies are subject to strict rules to protect study participants and ensure quality. Clinical trials in humans are classified into three phases: phase I, phase II and phase III. Phase I studies are carried out on limited numbers, for example, 20 + healthy adults and are primarily concerned with safety. Phase II studies involve larger numbers of persons belonging to the target population with the aim of getting preliminary information about efficacy, usually immunogenicity and additional safety data.

Together, phase I and II trials should give sufficient confidence that a vaccine is efficacious and safe. The real test comes from phase III trials in which a vaccinated population is compared with a control group for protection against the target diseases, resulting in an estimate of vaccine effectiveness. Depending on the prevalence of the disease these studies may require 10,000 or more participants. After the completion of the phase III studies, all results are summarized in a dossier. This dossier also includes data on a consistent manufacturing process and assays to monitor the production and its end products. This dossier is submitted to the regulatory authorities (The Federal Drug Administration in the USA and the European Medicine Agency). After the regulatory authorities have decided that the vaccine is safe and effective, it is licensed and can be released to the market.

Exceptionally, a vaccine fails in the phase III stage. This happened in the 1960s with a vaccine against disease caused by respiratory syncytial virus. This vaccine was not efficacious and even unsafe, killing two participants of the trial (Hurwitz, 2011). Other problems that popped up in the past were due to the limited number of participants in the phase III trials. As a consequence, rare adverse effects *e.g.*, occurring 1 in 100,000 times will not be detected. Therefore, post-licensure safety surveillance (phase IV studies) is needed to detect adverse effects following immunization (AEFI).

Phase IV studies revealed that a rotavirus vaccine caused intussusception in one or two cases per 10,000 infants vaccinated. This vaccine was withdrawn from the market. Newer rotavirus vaccines are 10-fold safer, but not completely without AEFI (Di Pasquale *et al.*, 2016). Narcolepsy, a sleeping disorder, was identified as a possible AEFI associated with the AS03-adjuvanted influenza vaccine Pandemrix. Although there is a clear association between the receipt of Pandemrix and the development of narcolepsy the causality remains unproven. This incident has shown that there is a need for an internationally coordinated vaccine safety structure that can monitor the safety of the many doses of pandemic vaccines administered in a short period (Edwards *et al.*, 2019).

Costs and Duration of Vaccine Development

The exact data on the development of Shingrix are confidential to the developer and producer GlaxoSmithKline. Clinical studies are responsible for the major costs of vaccine development. Thus, costs will vary depending on the complexity of these studies. Estimates of vaccine development expenditure vary from 135 to more than 1000 million USD (Plotkin *et al.*, 2017; DiMasi *et al.*, 2016). An analysis reported a breakdown of the costs of the separate steps in vaccine development (Gouglas *et al.*, 2018). Analysis of databases with clinical trials on vaccines has yielded reliable data on the success rate of vaccine development. Two studies reported an identical success rate of vaccine development from phase I studies to licensing of 33% (DiMasi *et al.*, 2020; Wong *et al.*, 2019). A success rate of 48% in the transition from the preclinical phase to phase 1 studies (Davis *et al.*, 2011) results in a 16% overall success rate. Estimates for the median duration of phase I-III trials during vaccine development are 6.4–10.7 years (Davis *et al.*, 2011; Pronker *et al.*, 2013). A summary of the data is given in Fig. 2.

Vaccine Production

Coming back to the Shingrix example: this vaccine is produced by expression, via DNA-recombinant technology, of the antigen in Chinese hamster ovary (CHO) cells. The CHO cells are grown in bioreactors in dedicated production facilities. Typically, these production facilities contain stainless steel bioreactors controlled by sophisticated computer programs to ensure a reproducible production process. After this upstream processing the antigen has to be purified, usually, in various steps, a process called downstream processing. Assays are carried out to check (intermediate) products against the specifications (quality control [QC]) during all steps. At the same time, a quality assurance (QA) team follows the stream of accompanying documentation to ensure that all intended controls were carried out. Production, together with QC and QA, can be a long process of more than one year.

The decision to build production facilities is usually taken before the regulatory authorities approve the vaccine. This implies a considerable financial risk.

Shingrix is a successful vaccine and has reached blockbuster status with annual sales in 2019 of more than 1.5 billion USD. The demand was much higher than foreseen, resulting in the need to build additional production facilities. Since the production process is an integral part of the license, careful calibration of a new production facility and the demonstration of equivalence of the product are required. Alternatives for the traditional bioreactor-based production methods that can speed up production, are scalable and reduce costs are being intensively investigated. See Section “Emerging Viral Diseases”.

More About Vaccine Development

Clinical studies are time-consuming and expensive. It would be much easier to measure an immune response that is predictive for protection. Such correlates of protection (CoP) are available for many viral vaccines (Plotkin, 2010). CoPs can be used to license newer vaccines, but also to evaluate existing and new vaccination schedules for protection. Systems vaccinology is a more advanced method to predict the immunogenicity of vaccines. In essence, this links host gene expression (molecular signatures) to protection by vaccines. This approach was very successful for the live attenuated yellow fever vaccine and is, in principle, also useful to predict adverse effects. But more work is needed to move it from ‘promising’ to a standard approach (Raeven *et al.*, 2019).

The Licensing of Influenza Vaccines Follows a Different Procedure

The current influenza vaccines are mostly produced from an egg-grown virus. The vaccines have to be matched every year to the predominantly circulating virus strains. This requires a special procedure. The production process is licensed with the possibility to insert the circulating viruses, selected in the Northern hemisphere in February-March by the WHO. The current production process is capable to make the vaccine available in October (Soema *et al.*, 2015).

The Vaccine Industry

Over the past decades, the number of vaccine suppliers has decreased considerably due to mergers and acquisitions of pharmaceutical companies. Presently about 90% of global vaccine sales come from four large multi-national corporations: GlaxoSmithKline, Merck, Pfizer and Sanofi Pasteur (Shen and Cooke, 2019). In the 1980s emerging market manufacturers started to enter the vaccine market and assumed a significant role since then. Emerging manufacturers, represented by the Developing Countries Vaccine Manufacturers Network, play a critical role in the supply of vaccines of developing countries. They now supply about half of UNICEF’s vaccine procurement in a volume of doses, representing about 30% of the value of UNICEF’s total vaccine procurement. Even including producers from developing countries, there are relatively few vaccine manufacturers that meet international standards of quality. Many vaccine markets are monopolies or oligopolies. The limited number of vaccine suppliers and production capacities leads to a tenuous balance between demand and supply in many individual vaccine markets and regularly to vaccine shortages.

Taking Stock

Due to vaccination, viral diseases remain under control. Two vaccines, hepatitis B and human papillomavirus protect against cancer. Thanks to a functional vaccine industry, vaccines remain available for a largely stable market. In addition, the research-based vaccine industry is innovative and develops new vaccines to answer unmet medical needs. This also applies to influenza vaccines with a new composition every year. However, problems exist with vaccines against new viral diseases to combat epidemics. The duration of these epidemics is unknown and usually, there is no commercially attractive market. Market forces fail, but vaccines are urgently needed. This will be the topic of the second part of this article.

The Challenge of Emerging Viral Diseases

In March 2003, the WHO issued a global alert about a new viral disease, severe acute respiratory syndrome (SARS). In retrospect, the SARS epidemic started in 2002 in China. In several months, SARS spread to 37 countries where it caused 8098 cases with 774 deaths before it came under control. SARS coronavirus (SARS-CoV), the causative agent of the syndrome, was traced back to an intermediate host, the civet cat, which originated as a virus of the horseshoe bat. Small mutations in the viral genome expanded the host range to humans.

A new viral disease in an immune naïve population could cause a catastrophe with many fatalities. It is worrying that there are many animal viruses with unknown pandemic potential. The SARS epidemic was the first wake-up call.

The 2007 World Health Report of the WHO was dedicated to Emerging viral diseases (EVD) and stated that “It would be extremely naive and complacent to assume that there will not be another disease like AIDS, another Ebola, or another SARS, sooner or later” (WHO, 2007). The real wake-up call came only in 2013 after an outbreak of Ebola in West Africa (Guinea, Liberia and Sierra Leone). Ebolavirus had been discovered in 1976. Between 1976 and 2013 it regularly spilled over from its animal reservoir to humans. There were several outbreaks, but these remained small and were contained. In the 2013 outbreak, about 28,000 persons were infected and 11,000 died. In addition, there was transmission to other countries and loss in GDP in the three affected countries. Many discussions in the aftermath of the Ebola outbreak led to the conclusion that more investments in R&D to counter EVD were needed, together with a more active role involving the WHO. As a result, the WHO Blueprint for Action to Prevent Epidemics was developed. The 2018 Blueprint list of priority diseases contains the following viral diseases:

- Crimean-Congo hemorrhagic fever
- Ebola virus disease and Marburg virus disease
- Lassa fever
- Middle East respiratory syndrome (MERS-CoV) and SARS
- Nipah and henipaviral diseases
- Rift Valley fever
- Zika
- Disease X (caused by a pathogen currently unknown)

The coordination and execution of R&D are in the hands of the Coalition for Epidemic Preparedness Innovations (CEPI). CEPI is a public-private partnership that was established during the World Economic Forum Annual Meeting of 2017.

CEPI acquired access to 755 million USD and used the money to support work on priority diseases (Lassa, Nipah, MERS, Rift Valley Fever and Chikungunya), selected with the aid of the WHO Blueprint. Two requests for proposals involved priority viruses. The aim is to bring vaccines through phase II studies and have vaccine stockpiles available for at least two pathogens by the end of 2022.

Another request for proposals involved platform technologies. Please see the CEPI Business Plan 2019–2022 for more details. Platform technologies have the potential to speed up development and production drastically. The common property of platforms is that they can be used for multiple vaccine antigens. The safety of the platform is known, the only variable is the inserted antigen. This will save time and may lead to regulatory streamlining, comparable to the procedure for influenza virus vaccines. Based on reduced development and production times, the most promising platforms are DNA and RNA vaccines and neutralizing antibodies, as discussed below.

DNA vaccines date back to the early 1990s when it was discovered that plasmid DNA, delivered in muscle or skin induces an antibody response to the encoded protein. The DNA has to cross the membranes of the cell and the nucleus. Subsequently, the antigen is synthesized. The initial excitement diminished somewhat after it appeared that immune responses in man were weaker than in mice. However, this problem has been solved by more efficient systems for the delivery and formulations that protect DNA against degradation. In addition to antibodies, DNA vaccines elicit $CD4^+$ and $CD8^+$ T cells. Logistically DNA vaccines have many advantages, such as fast, inexpensive and scalable production and short development times. Also, no handling of infectious virus is required during vaccine development. This makes these vaccines ideal as protection against emerging viral diseases (Rauch *et al.*, 2018). Presently vaccine development against various diseases, including Ebola, MERS and Zika, is ongoing (Rauch *et al.*, 2018). A (theoretical) disadvantage of DNA vaccines is that they may integrate into the human genome and lead to undesired gene activation. Presently (2019) no DNA vaccines have been licensed for human use.

RNA vaccines consist of mRNA. They have the same advantages as DNA vaccines, but since mRNA has to cross only one membrane, immune responses are stronger. In addition to conventional mRNA, self-amplifying mRNA can be used. mRNA is extremely sensitive to degradation by endonucleases. Therefore, a formulation, e.g., liposomes, which protect the RNA is used. Many vaccines are presently under development and also in clinical trials. But although RNA vaccines are very promising, their development is still at an early stage (Maruggi *et al.*, 2019).

Neutralizing antibodies offer an alternative treatment for viral infections. This is what is referred to as passive immunization and was successful in the treatment of Ebola (Saphire *et al.*, 2018). Another new development is the use of broadly neutralizing antibodies directed against conserved viral structures. These antibodies are promising both for prophylaxis and therapy for a range of viruses (Walker and Burton, 2018).

Looking Forward

Existing vaccines will continue to play a key role in controlling (viral) diseases. Many vaccines come as a combination vaccine and it will be complicated and very expensive to modify these vaccines. Newly developed vaccines will probably be made using novel approaches (Mascola and Fauci, 2020). We expect that the establishment of CEPI will contribute to these innovations. Paving a regulatory path to the first registration of DNA and RNA vaccines for infectious diseases will be a

challenge. But nucleic acid vaccines are also a promising approach for the immunotherapy of cancer. This may help in paving this path.

Another challenge is the design of clinical trials to test EVD. During the 2013 Ebola outbreak, there was disagreement about the design of clinical studies, particularly on ethical aspects of a placebo group. However, without a placebo group, it is impossible to measure the safety and efficacy of a vaccine. During these discussions the incidence of the disease declined, making it more difficult to carry out the phase III study. Choices about the design of clinical studies have to be made before outbreaks.

The Development of COVID-19 Vaccines

Shortly after the completion of the manuscript for this chapter, SARS-CoV-2 was discovered. Its global spread led to an overburdening of health systems and a death toll of about 1.6 million by December 13, 2020. Measures aimed at controlling the disease disrupted daily activities. A widespread hope is that immunity provided by vaccination will be the key to a return to 'normal life'. The first vaccine, an mRNA vaccine, developed and produced by Pfizer/BioNTech has been approved by now in several countries and vaccination has begun or will so shortly. The approval of a second mRNA vaccine developed by Moderna is expected in January 2021. Vaccine development started on January 11, 2020 when the genomic sequence of the virus, determined by Chinese scientists, became available. According to the WHO, 214 vaccines were under development on December 10, 2020, of which 13 in phase III clinical trials. The approval of more vaccines is expected in the first months of 2021. Estimated efficacies of vaccines are high, more than 90%. Adverse effects are present. But they are of short duration and not categorized as serious. Three vaccine platforms contributed to the most promising COVID-19 vaccines: (1) mRNA vaccines, (2) Adenovirus-based vaccines and (3) Vaccines consisting of inactivated virus. In addition, there are promising results from protein subunit vaccines. The use of these platforms significantly cut the development timelines. Further factors were an abundance of funding, tight collaborations between vaccine developers, governments and regulatory agencies, and the parallelization of activities. Especially the construction of manufacturing facilities parallel to the clinical development saved a lot of time. Several new platforms and procedures were adopted in the development of the COVID-19 vaccines. Licensing of these vaccines will pave the way for more vaccines based on these technologies. This could lead to considerable changes in the development and production of future vaccines.

References

- Amanna, I.J., Slifka, M.K., 2009. Wanted, dead or alive: New viral vaccines. *Antiviral Research* 84, 119–130.
- Davis, M.M., Butchart, A.T., Wheeler, J.R., *et al.*, 2011. Failure-to-success ratios, transition probabilities and phase lengths for prophylactic vaccines versus other pharmaceuticals in the development pipeline. *Vaccine* 29, 9414–9416.
- Delany, I., Rappuoli, R., De Gregorio, E., 2014. Vaccines for the 21st century. *EMBO Molecular Medicine* 6, 708–720.
- Di Pasquale, A., Bonanni, P., Garçon, N., *et al.*, 2016. Vaccine safety evaluation: Practical aspects in assessing benefits and risks. *Vaccine* 34, 6672–6680.
- Di Pasquale, A., Preiss, S., Tavares Da Silva, F., Garçon, N., 2015. Vaccine adjuvants: From 1920 to 2015 and beyond. *Vaccines* 3, 320–343.
- Dimasi, J.A., Florez, M.I., Stergiopoulos, S., *et al.*, 2020. Development times and approval success rates for drugs to treat infectious diseases. *Clinical Pharmacology & Therapeutics* 107, 324–332.
- Dimasi, J.A., Grabowski, H.G., Hansen, R.W., 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics* 47, 20–33.
- Dooling, K.L., Guo, A., Patel, M., *et al.*, 2018. Recommendations of the advisory committee on immunization practices for use of herpes zoster vaccines. *Morbidity and Mortality Weekly Report* 67, 103–108.
- Edwards, K., Lambert, P.H., Black, S., 2019. Narcolepsy and pandemic influenza vaccination: What we need to know to be ready for the next pandemic. *The Pediatric Infectious Disease Journal* 38, 873–876.
- Gouglas, D., Le, T.T., Henderson, K., *et al.*, 2018. Estimating the cost of vaccine development against epidemic infectious diseases: A cost minimisation study. *The Lancet Global Health* 6(e1386–e1396).
- Haumont, M., Jacquet, A., Massaer, M., *et al.*, 1996. Purification, characterization and immunogenicity of recombinant varicella-zoster virus glycoprotein gE secreted by Chinese hamster ovary cells. *Virus Research* 40, 199–204.
- Hurwitz, J.L., 2011. Respiratory syncytial virus vaccine development. *Expert Review of Vaccines* 10, 1415–1433.
- Levin, M.J., Weinberg, A., 2020. Adjuvanted recombinant glycoprotein E herpes zoster vaccine. *Clinical Infectious Diseases* 70, 1509–1515.
- Maltz, F., Fidler, B., 2019. Shingrix: A new herpes zoster vaccine. *Pharmacy and Therapeutics* 44, 406–433.
- Maruggi, G., Zhang, C., Li, J., Ulmer, J.B., Yu, D., 2019. mRNA as a transformative technology for vaccine development to control infectious diseases. *Molecular Therapy* 27, 757–772.
- Mascola, J.R., Fauci, A.S., 2020. Novel vaccine technologies for the 21st century. *Nature Reviews Immunology* 20, 87–88.
- Plotkin, S.A., 2010. Correlates of protection induced by vaccination. *Clinical and Vaccine Immunology* 17, 1055–1065.
- Plotkin, S.A., 2011. *History of Vaccine Development*. New York: Springer.
- Plotkin, S.A., Orenstein, W.A., Offit, P.A., Edwards, K.M. (Eds.), 2018. *Plotkin's vaccines*. seventh ed. Elsevier.
- Plotkin, S., Robinson, J.M., Cunningham, G., Iqbal, R., Larsen, S., 2017. The complexity and cost of vaccine manufacturing – An overview. *Vaccine* 35, 4064–4071.
- Pronker, E.S., Weenen, T.C., Commandeur, H., Claassen, E.H., Osterhaus, A.D., 2013. Risk in vaccine research and development quantified. *PLOS One* 8, e57755.
- Raeven, R.H.M., Van Riet, E., Meiring, H.D., Metz, B., Kersten, G.F.A., 2019. Systems vaccinology and big data in the vaccine development chain. *Immunology* 156, 33–46.
- Rauch, S., Jasny, E., Schmidt, K.E., Petsch, B., 2018. New vaccine technologies to combat outbreak situations. *Frontiers in Immunology* 9, 1963.
- Saphire, E.O., Schendel, S.L., Gunn, B.M., Milligan, J.C., Alter, G., 2018. Antibody-mediated protection against Ebola virus. *Nature Immunology* 19, 1169–1178.
- Shen, A.K., Cooke, M.T., 2019. Infectious disease vaccines. *Nature Reviews Drug Discovery* 18, 169–170.
- Soema, P.C., Kompier, R., Amorij, J.P., Kersten, G.F., 2015. Current and next generation influenza vaccines: Formulation and production strategies. *European Journal of Pharmaceutics and Biopharmaceutics* 94, 251–263.
- Vafai, A., 1993. Antigenicity of a candidate varicella-zoster virus glycoprotein subunit vaccine. *Vaccine* 11, 937–940.
- Walker, L.M., Burton, D.R., 2018. Passive immunotherapy of viral infections: 'Super-antibodies' enter the fray. *Nature Reviews Immunology* 18, 297–308.

WHO, 2007. The World Health Report 2007: A Safer Future: Global Public Health Security in the 21st Century.
Wong, C.H., Siah, K.W., Lo, A.W., 2019. Estimation of clinical trial success rates and related parameters. *Biostatistics* 20, 273–286.

Relevant Websites

<http://epidemics.events.nejm.org/#!/media/3222>

Epidemics going viral: Innovation vs. nature.

<https://cepi.net/about/governance/>

Governance

CEPI.

www.who.int/blueprint/en/

R&D Blueprint.

<https://www.reuters.com/article/us-gsk-results-idUSKBN1X91EK>

Reuters.

<https://www.weforum.org/events/world-economic-forum-annual-meeting-2017/sessions/cepi-a-global-initiative-to-fight-epidemics>

World Economic Forum Annual Meeting.

Vaccines Against Viral Gastroenteritis

Scott Crytldal and Tyler P Chavers, Centers for Disease Control and Prevention, Atlanta, GA, United States

Claire P Mattison, Centers for Disease Control and Prevention, Atlanta, GA, United States and Cherokee Nation Assurance, Arlington, VA, United States

Jacqueline E Tate and Aron J Hall, Centers for Disease Control and Prevention, Atlanta, GA, United States

Published by Elsevier Ltd.

Glossary

Acute gastroenteritis An inflammation of the stomach and/or intestines, often characterized by symptoms of vomiting and diarrhea lasting less than 14 days.

Correlate of protection A measurable sign that a person is protected against becoming infected or developing illness from a disease.

Genogroup A group of related viruses within a genus, which can then be further broken down into genotypes.

Genotype Subclassification of related viruses within a genogroup based on shared genetic sequences.

Intussusception Intestinal obstruction that results when a part of the intestine folds into the section next to it.

Mono-/polyvalent Having specific immunologic activity against one or many antigens, viral strains, or diseases (e.g., bivalent, trivalent).

Reassortant Having a genome consisting of parts derived from the genomes of two (or more) different viruses.

Seroconversion The development of detectable antibodies in the blood that are directed against an infectious agent.

Vaccine effectiveness The percentage reduction of disease or outcome in a vaccinated group of people in comparison to an unvaccinated group under typical field or uncontrolled conditions.

Vaccine efficacy The percentage reduction of disease or outcome in a vaccinated group of people in comparison to an unvaccinated group under ideal conditions (e.g., a randomized control trial).

Introduction

Each year, an estimated 2 billion cases of acute gastroenteritis (AGE) occur worldwide, resulting in more than 1 million deaths. Nearly 40% of AGE cases occur among children <5 years of age. Numerous bacterial, parasitological, and viral pathogens cause AGE, but viruses cause the majority of these infections. Among viruses that cause AGE, rotaviruses and noroviruses contribute the largest burden. Rotavirus disease typically occurs in young children; in contrast, noroviruses causes illness across all age groups.

Currently, rotavirus vaccines are the only available preventive vaccines against viral gastroenteritis. Prior to rotavirus vaccine development and availability, rotavirus caused approximately 2 million hospitalizations and a median 440,000 deaths in children <5 years of age annually. By 5 years of age, nearly every child experienced an episode of rotavirus gastroenteritis. To reduce rotavirus burden, in 2006, two rotavirus vaccines were licensed: Rotarix and RotaTeq. In 2009, the World Health Organization (WHO) recommended rotavirus vaccines for priority inclusion in national immunization programs worldwide. Today, 4 WHO pre-qualified rotavirus vaccines have been introduced into the national immunization programs of over 100 countries, with several more in development. These vaccines have been associated with a median 59% reduction in rotavirus hospitalizations and 36% in AGE hospitalizations globally. However, rotavirus is still estimated to cause 258 million episodes of diarrhea and 128,500 deaths among children <5 years of age annually. Approximately half of all rotavirus deaths occur in just four countries: India, Nigeria, Pakistan, and the Democratic Republic of the Congo, while in high-income countries, deaths from AGE are uncommon.

With the dramatic decline of rotavirus gastroenteritis, norovirus has become recognized as a leading cause of severe pediatric AGE in some countries that have introduced the rotavirus vaccine. Noroviruses are a diverse group of viruses within the family *Caliciviridae* and cause an estimated 684 million illnesses and 212,000 deaths worldwide annually. Similar to rotavirus, low- and middle-income countries (LMICs) carry the largest burden, accounting for 82% of norovirus illnesses and 97% of norovirus deaths worldwide. In the United States, it is estimated that norovirus causes an average of 19–21 million illnesses, 1.7–2.9 million outpatient visits, and 650–1100 deaths every year. In June 2016, WHO's Product Development for Vaccines Advisory Committee identified norovirus as a priority disease for vaccine development. While several candidate vaccines are in development, there is currently no licensed norovirus vaccine.

Rotavirus Vaccines

Rotavirus genotypes follow a binary nomenclature describing G and P genotypes, which are based on two viral proteins (VP) expressed on the exterior of the viral capsid. Glycoprotein VP7 forms the outer capsid shell and determines the G genotype, while protease-sensitive VP4 forms spiked structures and determines the P genotype. These two proteins play critical roles in attachment to host cells, pathogenesis, and immune response. As such, these sites, or epitopes, are usually targets against which vaccines

generate neutralizing antibodies. Rotavirus vaccine development has used both human and reassortant genotypes in rotavirus vaccine formulations. Reassortant rotavirus genotypes arise when human and non-human strains rearrange and combine naturally or artificially (i.e., in a laboratory) due to their segmented genome structure.

Licensed Rotavirus Vaccines

In 1998, the first commercially available vaccine for viral gastroenteritis, RotaShield, a tetravalent, human-rhesus reassortant rotavirus vaccine, was licensed and introduced into the routine immunization program in the United States. Use of this vaccine was suspended in 1999 and it was subsequently withdrawn from the United States market due to an observed increased risk of intussusception, a rare condition characterized by an invagination of the intestine, among children who received the vaccine.

In 2006, two live, attenuated, oral rotavirus vaccines, Rotarix and RotaTeq were licensed for use. Rotarix is a human, monovalent vaccine comprised of the G1P[8] strain. It is given as a 2-dose series typically at 2 and 4 months of age. In contrast, RotaTeq is a pentavalent human-bovine reassortant vaccine. Each of its five components are made up of a bovine (WC3) backbone combined with either a G1, G2, G3, G4, or P1A[8] human surface protein. This vaccine is given as a 3-dose series at 2, 4, and 6 months of age.

After vaccine introduction, promising efficacy against rotavirus with subsequent reductions in rotavirus AGE were observed in randomized controlled trials (RCTs) in early adopting countries such as the United States, Australia, several Latin American countries, and several European countries (range: 85%–98% efficacy against severe rotavirus AGE). Modest efficacy of these vaccines against severe diarrhea was observed in subsequent clinical trials in Africa and Asia (range: 40%–48% efficacy against severe rotavirus AGE). These clinical trials did not find an increased incidence of severe adverse events (SAEs) nor an increase in incidence of intussusception in vaccine recipients compared to placebo recipients. In 2009, WHO officially recommended rotavirus vaccines for inclusion into national immunization programs worldwide.

In recent years, two new rotavirus vaccines manufactured in India have completed clinical trials and are joining the global vaccine market. In 2018, Rotavac (Bharat Biotech, India) and Rotasiil (Serum Institute, India) vaccines obtained prequalification from WHO and will soon be incorporated into national immunization programs in Africa and Asia. Rotavac is a live, oral, monovalent vaccine derived from a naturally occurring G9P[11] reassortant human-bovine strain and is given in a 3-dose regimen at 6, 10, and 14 weeks of age. The strain was chosen as a vaccine candidate as it appears naturally attenuated, causing asymptomatic infection among neonates. Rotavac has shown 56% (95% confidence interval [CI]: 37%–70%) efficacy against hospitalization or supervised rehydration for children <1 year of age and 49% (95% CI: 17%–68%) efficacy among children between 1 and 2 years of age. Phase III trials for Rotavac did not have sufficient sample sizes to detect an increased risk of intussusception, but a surveillance network has been set up in India to monitor post-introduction risk of intussusception associated with this vaccine. Rotasiil is a live, oral, pentavalent human-bovine reassortant vaccine comprised of G1, G2, G3, G4, and G9 strains and is also given in a 3-dose regimen at 6, 10, and 14 weeks of age. It is delivered in either a liquid or freeze-dried (lyophilized) form. The lyophilized product is thermostable at 40°C for up to 18 months and requires a liquid buffer to be reconstituted at point of use. In RCTs, Rotasiil has shown efficacy against severe rotavirus gastroenteritis of 67% (95% CI: 50%–78%) in Niger, and 36% (95% CI: 12%–54%) in India; neither trial was powered to detect increased risk of intussusception. Both Rotavac and Rotasiil have not shown evidence of interference with other childhood vaccines given concomitantly, such as oral polio vaccine (OPV).

Two other rotavirus vaccines have been used within individual countries but are not available globally. The Lanzhou lamb rotavirus vaccine (Lanzhou Institute of Biological Products, China) is a live, attenuated oral vaccine using a monovalent G10P[12] rotavirus strain isolated from lamb. It has been sold exclusively through the private market in China since 2000. In Vietnam, the internally developed Rotavin M1 (POLYVAC, Center for Research and Production of Vaccines and Biologicals, Vietnam) was first used in 2012. Rotavin M1 is a live, attenuated, frozen oral vaccine that includes a G1P[8] strain, sold exclusively to the Vietnamese private healthcare market. Neither of these vaccines are licensed outside of their respective countries or have been evaluated for further use by WHO.

Vaccine Effectiveness and Impact

Post-licensure studies which measure rotavirus vaccine effectiveness (VE) have consistently shown strong protection in upper and upper-middle income countries, and robust albeit lower protection in LMICs. A recent review of global rotavirus impact studies in children <5 years of age found that after rotavirus vaccines were introduced, rotavirus AGE hospitalizations in countries using vaccines were reduced by a median of 59%. Countries with rotavirus vaccine coverage above 85% demonstrated the greatest impact. Reductions also differed by child mortality levels, demonstrating a 66% reduction in low, 59% reduction in medium, and 50% reduction in high child mortality countries. Globally, Rotarix has demonstrated a median VE against any healthcare utilization of 84% in low child mortality countries, 75% in medium child mortality countries, and 57% in high child mortality countries; RotaTeq demonstrated a median VE of 90% in low child mortality countries and 45% in high child mortality countries. Potential explanations for disparity in VE and impact between high and low child mortality countries include differences in nutritional status, gut microbiota, and interference by other childhood vaccinations such as OPV that reduce seroconversion. Despite a lower VE, due to high rotavirus burdens in LMICs, rotavirus vaccines do still prevent and mitigate the impact of rotavirus illness, preventing severe diarrhea, healthcare visits, and deaths in these countries.

Both Rotarix and RotaTeq are used in the United States where overall rotavirus vaccine coverage was estimated at 73.2% (95% CI: 71.8%–74.6%) in 2015, notably lower than other routine childhood immunizations. In the United States, rotavirus vaccines are associated with a median 80% decline in rotavirus-associated hospitalizations and 57% in rotavirus-related emergency department visits along with indirect protection of unvaccinated age groups and a decrease in health-care costs. Biennial peaks of rotavirus detection have emerged since rotavirus vaccines were introduced into the U.S. immunization program.

Protection from rotavirus infection does not require an exact match between wild-type and vaccine genotypes. Studies have demonstrated that both the monovalent vaccine Rotarix and pentavalent vaccine RotaTeq show cross-protection against dissimilar genotypes. Prior to vaccine introduction, the G1P[8] strain was the most common rotavirus genotype in humans. A recent systematic review showed that the most common strains after vaccine introduction were G2P[4] (50%) and G1P[8] (22%) in countries using Rotarix, and G1P[8] (33%) and G2P[4] (30%) in countries using RotaTeq, although dominant subtypes vary by time and location.

Vaccine performance may also be modulated by the presence of maternal antibodies against rotavirus transferred through the placenta or present in breast milk. These antibodies typically lessen severity and reduce incidence of disease in newborns up to approximately 2 months of age. It has been hypothesized that these maternal antibodies may also neutralize rotavirus vaccine particles before the child can initiate an immune response and thus lower their effectiveness. Further studies are required at this time, as different studies have shown either reduced vaccine seroconversion when high levels of rotavirus-specific IgG were present or no impact on vaccine response when breastfeeding was withheld.

Rotavirus Vaccines in Development

Several rotavirus vaccine candidates are currently in development. RV3-BB is a vaccine that aims to confer immunity in the neonatal period even in the presence of maternal antibodies; it recently completed Phase IIb trials in Indonesia. Other vaccines in development are parenterally administered, non-replicating vaccines, which potentially bypass issues of maternal antibody interference, intestinal replication (potential risk of intussusception), and interference with other childhood vaccines such as OPV. The P2-VP8-P[8] and VAC 041, P2-VP8-P[4]P[6]P[8] subunit vaccines have completed dose escalation trials and Phase I/II trials are currently underway in South Africa. Other non-replicating candidates in early development stages include a VP6 subunit that uses norovirus virus-like particles (VLP) in a combination vaccine; two inactivated rotavirus vaccines (IRV) using the G1P[8] strain; a truncated VP4 vaccine derived from the Lanzhou Lamb vaccine; and an inactivated polio vaccine (IPV)-IRV combination using the G1P[8] strain.

Rotavirus Vaccine Safety Monitoring

After a decade of recommendation by WHO, Rotarix and RotaTeq have demonstrated substantial impact on reducing the global burden of rotavirus disease; the largest effect has been observed on severe disease and hospitalization. Post-licensure safety evaluations have detected a low-level risk of intussusception in some high- and middle-income countries, although the risk-benefit ratio remains favorable towards rotavirus vaccinations due to their substantial impact on disease burden. Additionally, no increased risk of intussusception has been observed in low-income countries to date; however, current safety data are limited and additional studies are on-going. WHO continues to recommend rotavirus vaccine use in all countries as the small increased risk in intussusception remains greatly outweighed by the benefits of rotavirus vaccination.

Norovirus

In 2016, WHO identified norovirus as a priority disease for vaccine development. While there is currently no vaccine available to prevent norovirus disease, many candidate norovirus vaccines exist, including those in Phase I and Phase II trials, as well as many others in preclinical stages. The wide diversity of noroviruses and the limited understanding of naturally acquired norovirus immunity have complicated vaccine development. In addition, there is no single, established correlate of protection, although many candidates have been explored, including serum histo-blood group antigen (HBGA)-blocking antibodies, serum hemagglutination inhibition antibodies, salivary, serum, and fecal IgA, and virus-specific IgG memory B-cells.

Noroviruses are an extremely diverse group of viruses. Based on the phylogenetic clustering of the major capsid protein VP1, there are 10 genogroups, of which GI, GII, and GIV infect humans. Genogroup GII is the predominant norovirus genogroup circulating worldwide; within that genogroup, genotype GII.4 is the most common genotype in many parts of the world, with a new GII.4 strain typically emerging every 2–4 years, sometimes associated with an increase in norovirus activity. Recently, norovirus genotyping expanded to include simultaneous typing of both the capsid and polymerase genes, allowing surveillance to monitor for recombinant strains.

Individual susceptibility to norovirus infection varies based on the fucosyltransferase-2 (FUT2) gene, which controls the expression of HBGAs to which norovirus binds. Individuals with polymorphisms in the FUT2 gene are known as “non-secretors” and make up 20%–30% of African and European populations. Non-secretors are less susceptible (but not completely immune) to norovirus GII.4 infection than secretors; however, they can still become infected with non-GII.4 norovirus strains.

Naturally acquired immunity to norovirus post-infection is not well understood. Challenge studies in adults have shown short-term, strain-specific immunity, lasting from 6 months to 2 years, while modeling studies have estimated immunity lasts 4–8 years. Immunity is thought to be homotypic, with greater protection from strains in the same genogroups compared to other genogroups. Repeat infections by the same genotype are rare, but repeated infections by the same genogroup have been observed, suggesting that immunity is genotype specific. Since norovirus immunity is thought to be genogroup- or genotype-specific, determining which genotypes to include in vaccine formulations is an important consideration. Current vaccine candidates have used both GII.4 as well as GI.1 genotypes; however, norovirus vaccine candidates may need to consider the addition of other genogroups, or update formulations periodically as new strains emerge.

Norovirus Vaccine Development

Norovirus candidate vaccines utilize a variety of different technologies, including non-replicating VLPs, P particles, and recombinant adenoviruses. VLPs are multi-protein structures that have no genetic material and are therefore noninfectious. However, they can interact with cellular receptors and elicit an immune response. P particles resemble the P domain of norovirus, which is the domain that binds to HBGAs. Recombinant adenoviruses expressing the major norovirus capsid protein VP1, through which they are able to elicit an immune response. Due to the lack of a cell culture system prior to 2016, whole killed or live-attenuated viruses have not been previously pursued for norovirus vaccine development. The recently developed human intestinal enteroid (“mini-gut”) system for in vitro norovirus replication is expected to provide many new avenues of norovirus research and enhance norovirus vaccine development.

Norovirus Vaccines in Human Clinical Trials

There are currently four single- and multi-dose norovirus vaccine candidates in human clinical trials. A bivalent intramuscular GI.1/GII.4 VLP vaccine being developed by Takeda Pharmaceutical Company Limited is in Phase IIb trials; an oral vaccine being developed by Vaxart, Inc. is in Phase I trials; a tetravalent vaccine is being developed by the Chinese Academy of Sciences; and a bivalent GI.1/GII.4 vaccine is being developed by the National Vaccine & Serum Institute in China.

Takeda’s GI.1/GII.4 vaccine was initially developed as a monovalent GI.1 VLP vaccine. In Phase I trials, two GI.1 vaccine doses 21 days apart induced a 4.8-fold increase in norovirus-specific IgG and a 9.1-fold increase in norovirus-specific IgA, while also increasing norovirus-specific IgG and IgA memory B-cells in 92%–100% of participants. When challenged with a homologous GI.1 strain, vaccinated participants had a 47% reduction in norovirus illness and a 26% reduction in norovirus infection. In pre-clinical trials in rabbits, the reformulated bivalent vaccine containing GII.4 and GI.1 VLPs induced broadly reactive antibodies to heterologous GI.1, GII.1, GII.3, and GIV noroviruses.

Phase II studies have focused on the bivalent GI.1/GII.4 formulation and have trialed both single-dose and multi-dose vaccines. The bivalent vaccine has demonstrated a rapid immune response, including higher pan-Ig, IgA, and HBGA-blocking antibodies. While the duration of elevated immune responses is unknown, antigen-specific IgG memory B-cells persist until at least 180 days post-vaccination, while other responses have lasted over a year. In a GII.4 norovirus challenge study, two intramuscular injections of the bivalent vaccine did not significantly decrease the incidence of pre-defined norovirus disease compared to placebo. However, reductions in mean viral shedding and illness severity were observed among vaccinated participants. No SAEs related to vaccination were reported in any of the aforementioned studies.

Data from a Phase IIb field efficacy study in military recruits (NCT02669121) were recently reported. Recruits were randomized to receive a placebo or single intramuscular injection of bivalent GI.1/GII.4 vaccine and followed for 45 days afterwards. Vaccine efficacy against moderate or severe AGE caused by any norovirus strain was 61.8% (95% CI: 20.8–81.6); however, due to the low number of norovirus infections in recruits, vaccine efficacy against homotypic infection could not be calculated. Multiple other Phase II trials have recently been completed, including a trial in infants 6 weeks to children 8 years old (NCT02153112), and an evaluation of safety and immune responses in those older than 60 years (NCT02661490), as well as a 5-year Phase II trial in adults and the elderly (NCT03039790).

A second norovirus vaccine currently in human clinical trials is an oral vaccine being developed by the biotechnology company Vaxart, Inc. This single dose vaccine formulation would be delivered as an oral tablet for adults or as a liquid for children and is stable at ambient temperature, an advantage over other vaccines. The vaccine uses a recombinant adenovirus expressing the norovirus GI.1 major capsid protein (VP1). Two Phase I trials have been completed. A Phase I trial testing the safety and immunogenicity of both low- and high-dose formulations found they were well tolerated and no SAEs were reported. After one administration of the high dose vaccine, 78% of participants showed a significant immune response with a 2-fold increase in HBGA-blocking assays (BT50s), increased IgA and IgG memory B-cells, and increased fecal IgA. Vaxart has reported that the vaccine has been well tolerated with no SAEs reported.

Vaxart has also trialed a bivalent GI.1/GII.4 VP1 vaccine in a Phase Ib trial, alongside monovalent GII.4 and GI.1 formulations. The trial reported no SAEs and demonstrated robust immunogenicity. It was also reported that the study met all primary endpoints with no SAEs reported. A Phase II dose confirmation study is planned to begin in 2020.

The two other vaccines are beginning clinical trials. The tetravalent GI.1/GII.3/GII.4/GII.17 vaccine being developed by the Institut Pasteur of Shanghai under the Chinese Academy of Sciences received a permit in May 2019 to begin clinical trials. In

addition, at its National Vaccine & Serum Institute, China is recruiting participants aged 6 months to 59 years in a Phase I clinical trial for a bivalent intramuscular GI.1/GII.4 VLP vaccine.

Norovirus Vaccines in Pre-Clinical Development

An intramuscular, trivalent norovirus GI.3, GII.4, and rotavirus VP6 vaccine is being developed by the University of Tampere, Finland, the Daiichi Sankyo Company Limited, & UMN Pharma Inc., Japan. When vaccinated with the trivalent vaccine, mice showed an immune response that persisted for 24 weeks post-vaccination. The addition of rotavirus VP6 protein showed an adjuvant effect, increasing cross-reactive IgG and norovirus-specific blocking antibodies. Recently, research has evaluated the impact of including CVB1 VLPs to the vaccine to protect against coxsackievirus B (CVB), an enterovirus that can cause severe disease in infants. Addition of CVB1 VLPs did not interfere with immune responses to the norovirus or rotavirus VLPs.

A trivalent vaccine containing a fusion of hepatitis E, norovirus GII.4, and astrovirus P particles has been designed by the Cincinnati Children's Hospital Medical Center, University of Cincinnati, and Virginia Polytechnic Institute and State University. This candidate vaccine has been shown to produce a 1.9-fold higher norovirus IgG titer than immunization with norovirus P particle alone in mice and demonstrates the potential for a P particle vaccine to vaccinate against multiple diseases.

The Ohio State University recently published the results of a pre-clinical trial of an oral, lactic acid bacteria based GII.4 norovirus vaccine candidate on gnotobiotic piglets. Vaccination was found to induce a norovirus-specific immune response and prevent norovirus infection in pig intestines when challenged.

The Chinese Academy of Sciences in Shanghai has tested a VLP-based bivalent vaccine against norovirus GII.4 and enterovirus 71 (one of the viruses that causes hand, foot, and mouth disease) in mice. Vaccination produced significant increases in both enterovirus 71- and norovirus GII.4-specific antibody responses for up to 14 weeks.

Using a recombinant adenovirus expression system, a candidate vaccine expressing the norovirus GII.4 major capsid protein VP1 has been developed by the Chinese Center for Disease Control and Prevention. One study in mice found that this intranasal vaccine increased norovirus IgG and IgA immune responses. The researchers found that administering the recombinant adenovirus expressing norovirus GII.4 capsid protein vaccine before norovirus VLP vaccination produced higher norovirus-specific antibody levels in mice when compared to previous administration of the VLP or multiple VLP vaccinations. These studies show that a recombinant adenovirus expressing norovirus proteins may be able to produce a comparable immune response to norovirus VLP formulations.

Future Considerations for Norovirus Vaccines

While norovirus causes a tremendous morbidity burden worldwide, factors such as the vaccine's cost, effectiveness, and duration of protection will all end up impacting a vaccine's viability. A modeling study of the potential impact of a norovirus vaccine in the United States found that a vaccine with 50% efficacy and a 12-month duration would avert 1.0–2.2 million illnesses per year but cost \$400 million to \$1 billion annually; in contrast, one that conferred protection for 48 months could save up to \$2.1 billion annually. It is also currently unclear who the target population will be for norovirus vaccination. Modeling studies have shown vaccination would have the greatest impact if it targeted the young (under 5 years old) or the elderly (over 65 years old). Additional potential target populations include those working in health care, childcare, and food service.

CDC Disclaimer

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Names of specific vendors, manufacturers, or products are for identification only and do not imply endorsement by the Centers for Disease Control and Prevention or the US Department of Health and Human Services.

Further Reading

- Bartsch, S.M., Lopman, B.A., Hall, A.J., Parashar, U.D., Lee, B.Y., 2012. The potential economic value of a human norovirus vaccine for the United States. *Vaccine* 30 (49), 7097–7104.
- Bartsch, S.M., Lopman, B.A., Ozawa, S., Hall, A.J., Lee, B.Y., 2016. Global economic burden of norovirus gastroenteritis. *PLoS One* 11 (4), e0151219.
- Burke, R., Mattison, C., Pindych, T., *et al.*, 2020. The burden of norovirus in the United States, as estimated based on administrative data: Updates for medically attended illness and mortality, 2001 – 2015. *Clinical Infectious Diseases*. ciae438.
- Burke, R., Tate, J., Kirkwood, C., Steele, A., Parashar, U., 2019. Current and new rotavirus vaccines. *Current Opinion in Infectious Diseases* 32 (5), 435–444.
- Burnett, E., Jonesteller, C., Tate, J., Yen, C., Parashar, U., 2017. Global impact of rotavirus vaccination on childhood hospitalizations and mortality from diarrhea. *Journal of Infectious Diseases* 215 (11), 1666–1672.
- Burnett, E., Parashar, U.D., Tate, J.E., 2020. Global impact of rotavirus vaccination on diarrhea hospitalizations and deaths among children <5 years old: 2006–2019. *Journal of Infectious Diseases*. jiaa081.
- Cohen, D., Muhsen, K., 2019. Vaccines for enteric diseases. *Human Vaccines & Immunotherapeutics* 15 (6), 1205–1214.
- Cortes-Penfield, N., Ramani, S., Estes, M., Atmar, R., 2017. Prospects and challenges in the development of a norovirus vaccine. *Clinical Therapeutics* 39 (8), 1537–1549.

- GBD Diarrhoeal Diseases Collaborators, 2017. Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infectious Diseases* 17 (9), 909–948.
- Hallowell, B., Parashar, U., Hall, A., 2019. Epidemiologic challenges in norovirus vaccine development. *Human Vaccines & Immunotherapeutics* 15 (6), 1279–1283.
- Jonesteller, C., Burnett, E., Yen, C., Tate, J., Parashar, U., 2017. Effectiveness of rotavirus vaccination: A systematic review of the first decade of global postlicensure data, 2006–2016. *Clinical Infectious Diseases* 65 (5), 840–850.
- Lucero, Y., Vidal, R., O’Ryan, G., 2018. Norovirus vaccines under development. *Vaccine* 36 (36), 5435–5441.
- Mattison, C., Cardemil, C., Hall, A., 2018. Progress on norovirus vaccine research: Public health considerations and future directions. *Expert Review of Vaccines* 17 (9), 773–784.
- Pindyck, T., Tate, J., Parashar, U., 2018. A decade of experience with rotavirus vaccination in the United States – Vaccine uptake, effectiveness, and impact. *Expert Review of Vaccines* 17 (7), 593–606.
- Troeger, C., Khalil, I., Rao, P., *et al.*, 2018. Rotavirus vaccination and the global burden of rotavirus diarrhea among children younger than 5 years. *JAMA Pediatrics* 172 (10), 958–965.

Human Papillomavirus (HPV) Vaccines and Their Impact

Jade Pattyn, Pierre Van Damme, and Alex Vorsters, University of Antwerp, Antwerp, Belgium

© 2021 Elsevier Ltd. All rights reserved.

Nomenclature

AAHS Amorphous aluminum hydroxyphosphate sulfate
AS04 Aluminum hydroxide and 3-O-desacyl-4'-monophosphoryl lipid A
CFS Chronic Fatigue Syndrome
CIN Cervical intraepithelial neoplasia
CRPS Complex regional pain syndrome
DNA Deoxyribonucleic acid
GACVS Global Advisory Committee for Vaccine Safety
GMTs Geometric mean titers

HIV Human immunodeficiency virus
HPV Human papillomavirus
hr High-risk
IARC International Agency for Research on Cancer
IgG Immunoglobulin G
POTS Postural orthostatic tachycardia syndrome
RRP Recurrent respiratory papillomatosis
VLPs Virus-like particles
WHO World Health Organization

Glossary

Herd immunity Herd immunity is a form of immunity that occurs when a sufficient proportion of a population (or herd) is immune, through disease or vaccination, conferring indirect protection for individuals who have not developed immunity.

Vaccine effectiveness Vaccine effectiveness is the ability of the vaccine to prevent outcomes of interest in the 'real world'.

Vaccine efficacy Vaccine efficacy is the percentage reduction of disease in a vaccinated group of people compared to an unvaccinated group, using the most favorable conditions (= clinical trial conditions).

Vaccine immunogenicity Immunogenicity refers to the ability of a vaccine to induce an immune response (antibody- and/or cell-mediated immunity) in a vaccinated individual.

HPV and HPV-Related Diseases

Human papillomaviruses (HPV) are the most common viral infections of the reproductive tract. These viruses are easily transmitted. Most sexually active women and men will be infected by HPV at some point in their lives. The peak time for acquiring infection for both women and men is shortly after becoming sexually active (Moscicki, 2007). HPV infections are primarily transmitted by mucosal contact, usually during sexual intercourse. However penetrative sex is not required for transmission, also manual-genital and oral-genital contact is a recognized mode of transmission (Liu *et al.*, 2016).

Although most infections with HPV do not cause symptoms or disease, persistent HPV infection with certain high-risk genotypes can lead to precancerous lesions that may progress to become cancerous (zur Hausen, 1977). Cancers caused by HPV include virtually all cancers of the cervix, most anal cancers (88%), and for a substantial fraction vaginal (78%), vulvar (15%–48% depending on age), penile (51%), and oropharyngeal (13%–60% depending on region) cancers (Bray *et al.*, 2018; World Health Organization, 2017). Worldwide, HPV causes up to 4.5% (643,000 cases) of all new cancer cases worldwide (8.6% females and 0.9% males). The global burden of cervical cancer (530,000 cases) is substantially higher than other HPV-related cancers (113,000 cases) (de Sanjose *et al.*, 2018).

HPV types 16 and 18 account together for about 70% of cervical cancers in every world region, and HPV31/33/45/52/58 for another 20% (de Martel *et al.*, 2017). In HPV-positive cancers of other anogenital sites, HPV16 is the most common HPV type. HPV6 and 11 have been classified as Group 3 agents (i.e., "not classifiable as to its carcinogenicity to humans") by the IARC Monographs Program (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012), but are responsible for almost 100% of anogenital warts and recurrent respiratory papillomatosis (RRP) (Gillison *et al.*, 2012). RRP is a rare disease in children and young adults that causes wart-like growths in the upper aero-digestive tract with the risk of airway obstruction (Fortes *et al.*, 2017).

HPV Vaccines

Three prophylactic HPV vaccines are currently in use to prevent HPV infection and HPV-related diseases, a bivalent, a quadrivalent, and a nonavalent vaccine (summarized in Table 1). The bivalent (HPV16/18) and quadrivalent (HPV6/11/16/18) vaccines were licensed in 2006–2007 and the nonavalent vaccine (HPV6/11/16/18/31/33/45/52/58) in 2014. Currently available HPV vaccines are made from the purified HPV L1 structural protein that self-assembles to form HPV type-specific empty shells, called virus-like

particles (VLPs) (Lowy and Schiller, 2006). The VLPs are structured geometrically and antigenically almost identical to the actual virus but contain no viral DNA, and therefore cannot reproduce or cause disease. HPV vaccines are not therapeutic and are not indicated for the treatment of HPV infection or HPV-related disease (Schiller and Lowy, 2018). Hence, all vaccines against HPV are most effective when administered to individuals not yet exposed to HPV, i.e., before sexual debut (Arbyn *et al.*, 2018; Xu *et al.*, 2019). Most countries offer HPV vaccines to girls aged 9–14 years. HPV vaccines are mainly targeted towards girls because cancer of the cervix is the most common HPV-associated cancer. In addition, some countries also offer routine or catch-up vaccination in boys, older adolescent females and young women. The vaccination schedule depends on the age of the vaccine recipient (<15 years = 2-dose schedule, ≥15 years = 3-dose schedule). By October 2019, 98 countries (global 51%) had introduced the HPV vaccine in their national immunization program for girls and 16 countries for boys (global 8%) (WHO, 2017).

Vaccine Immunogenicity

HPV vaccines prevent HPV-related disease by preventing infection. The mechanism of protection by HPV vaccines is assumed to be mediated by neutralizing antibodies against the major HPV coat protein, L1 (Schiller and Lowy, 2018). Possibly due to better immune activation by parental vaccination than by mucosal infection, and by the use of adjuvants, serological response after vaccination is much stronger (10- to 100-fold higher) than the response after natural HPV infection (Stanley, 2010). Furthermore, long-lived plasma cells continuously generate antibodies and are responsible for long-term HPV-specific antibody persistence shown in several clinical trials (Einstein *et al.*, 2014; Joura *et al.*, 2015; Schwarz *et al.*, 2017). Systemic antibodies generated by HPV immunization are thought to reach the site of infection by active antibody transudation at least in the female genital tract, and by passive exudation at sites of trauma that are supposed to be required for initiation of HPV infection (Pattyn *et al.*, 2019).

With HPV vaccines, practically all adolescent and young vaccinees who were initially naïve to vaccine-related HPV types develop an antibody response after vaccination (Einstein *et al.*, 2014; Joura *et al.*, 2015; Schwarz *et al.*, 2017). Data have shown that these vaccine-induced antibody titers peak after the final dose, decline within the first year, then stabilize at a plateau titer (Schiller and Lowy, 2018). Antibody titers after HPV immunization remain high for at least 10 years for the bivalent vaccine with 100% seropositivity (Schwarz *et al.*, 2017; Naud *et al.*, 2014), for at least 9.9 years (Nygard *et al.*, 2015) for the quadrivalent vaccine and at least 5 years for the nonavalent vaccine (World Health Organization, 2017; Joura *et al.*, 2015).

The immunogenicity of the bivalent and quadrivalent vaccines was compared in a head-to-head trial. After 60 months, geometric mean titers (GMTs) were consistently higher in those receiving the bivalent vaccine across all age strata (18–45 years) (Einstein *et al.*, 2014). However, the clinical relevance of these findings is unclear as vaccine failures have not yet been unequivocally identified in clinical studies, and so no antibody threshold or other immune measurement that correlates with protection has been defined. In general, the different choices of placebo recipients or control subjects, immunological assays and populations analyzed precluded the most direct comparison of results among the trials with the different HPV vaccines.

HPV Vaccine Impact

As the development of cervical cancer may occur decades after HPV infection, regulatory authorities have accepted the use of vaccine-related persistent HPV infections and cervical intraepithelial neoplasia (CIN) grade 2 or 3 (CIN2–3) as virologic and clinical endpoints in vaccine efficacy trials instead of invasive cervical cancer (Group, 2014). Moreover, using cervical cancer as the outcome in those trials was precluded for ethical reasons. HPV vaccines were licensed based on the demonstration of their clinical efficacy in young adult women. Collecting cervical specimens from girls or young adolescents is generally considered unethical. Results of immunobridging studies comparing vaccine immunogenicity in females aged 9–14 years with that in females aged 15–26 years were therefore used to infer clinical efficacy in the younger age group (World Health Organization, 2017).

Impact on HPV Infection and Precancerous Lesions

A meta-analysis by (Drolet *et al.*, 2019) provided updated data on more than 100 studies investigating HPV vaccination impact (Drolet *et al.*, 2019), an impact that was far beyond initial expectations. Clinical trials and post-marketing surveillance of the licensed prophylactic HPV vaccines have demonstrated high efficacy against new persistent high-risk HPV types contained in the vaccine and precancerous cervical abnormalities (Arbyn *et al.*, 2018). Notably, the bivalent vaccine demonstrated a high degree of cross-protection against types 31, 35, and 45 (Lehtinen and Dillner, 2013; Palmer *et al.*, 2019). While the quadrivalent vaccine established more restricted cross-protection, most notably against type 31 (Brown *et al.*, 2009). From a public health perspective, current evidence suggests that the 3 licensed HPV vaccines have relatively similar effectiveness in preventing cervical cancer, mainly caused by HPV16 and 18 (Arbyn *et al.*, 2018). At non-cervical sites, HPV vaccines have been shown to prevent the majority of anal, vulvar and vaginal HPV infections and lesions associated with HPV16 and 18 (Xu *et al.*, 2019). Importantly, reported data on the impact of HPV vaccines is thoroughly consistent across different regions and settings. The protective efficacy of the three vaccines has been maintained throughout their respective observation periods and protection in the long term is likely (Stanley, 2017).

Impact on HPV-Related Cancer

Due to the long latency period between infection and HPV-related cancers, it is still too early to observe a clear vaccination effect on the existing HPV-related cancer incidences. Nevertheless, some indicative observations of a reduced incidence of cervical cancer in vaccinated populations exist (Guo *et al.*, 2018; Luostarinen *et al.*, 2018; Scotland, 2019). Generally, it is expected that the existing HPV vaccines could prevent up to 90% of precursor cancer lesions and HPV-associated cancers (Hartwig *et al.*, 2017).

Impact on Anogenital Warts

HPV vaccination programs have also shown to effectively reduce the incidence of anogenital warts (Drolet *et al.*, 2019). The quadrivalent and nonavalent vaccine, which include HPV6 and HPV11, provides a high-level protection against anogenital warts in men and women. In several countries, the introduction of the HPV vaccine was followed by a rapid drop in the prevalence of genital warts. Furthermore, in countries with even moderate vaccination coverage ($\geq 50\%$), there was evidence of vaccine herd effects, with significant reductions in anogenital wart diagnoses among unvaccinated boys and older women (World Health Organization, 2017; Drolet *et al.*, 2019). Australian active surveillance of pediatric reports suggests a continuous decline in juvenile-onset recurrent respiratory papillomatosis, a disease caused by the perinatal transmission of HPV6/11 infection from an infected mother to her infant, likely due to a very low post-vaccination prevalence of maternal HPV 6/11 infection in Australia (Novakovic *et al.*, 2018; Brotherton, 2019).

Current Status and Future Directions

At least 15 countries now have data demonstrating vaccine effectiveness and/or showing falls in vaccine-targeted types, and cross-protective types especially for the bivalent vaccine, following HPV vaccination (Brotherton, 2019). Falls are the largest with higher coverage and multiple cohorts vaccinated (World Health Organization, 2017). These results reinforce the need for vaccinating a high proportion of the targeted population to maximize and accelerate the direct and herd effects of HPV vaccination (Drolet *et al.*, 2019), such as successfully demonstrated in the Australian mass catch up program (Drolet *et al.*, 2019). Notably, HPV vaccines can be used in persons who are immunocompromised and/or HIV-infected. Nevertheless, although the evidence base to support the immunogenicity of HPV vaccines in immunocompromised and/or HIV-infected persons is reasonable, the evidence base to support the efficacy of HPV vaccines in these populations remains inconsistent (Lacey, 2019).

Recently, studies found high vaccine effectiveness following a single dose of bivalent or quadrivalent vaccine (Kreimer *et al.*, 2018; Safaeian *et al.*, 2018; Sankaranarayanan *et al.*, 2018; Stanley and Dull, 2018). However, the interpretation of trials, which included women with incomplete vaccination schedules, is limited by several factors including that women were not randomized by the number of doses, the sample size, and the low number of incident or persistent infections. Several randomized controlled trials are ongoing to evaluate the efficacy of a one-dose regimen. If effective, a reduction in program costs achieved by implementing a one-dose schedule would make it a very attractive option to public health authorities and governments. Some other prospects to improve immune-based prevention of HPV-associated cancers include: Delivery of HPV vaccination as part of childhood immunization, development of lower-cost locally manufactured vaccines (WHO, 2019), and L2- vaccines, which might have potential for broad HPV coverage since the minor capsid protein L2 contains type-common epitopes (Brotherton, 2019).

Of importance, HPV vaccination does not eliminate the need for effective cervical cancer screening in the HPV-vaccinated cohorts, mainly because the vaccines do not protect against all oncogenic HPV types. However, vaccination has the potential to greatly decrease the intensity, cost, and morbidity associated with screening. It is clear that, together, HPV vaccination (the ultimate long-term preventive strategy) and screening using more sensitive HPV tests could dramatically alter the landscape of HPV-related cancers. However, the effective implementation of HPV vaccination and screening globally remains a challenge, as well as vaccine-impact evaluations (Wong *et al.*, 2011).

Vaccine Safety and Confidence

Data from clinical trials and post-licensure surveillance conducted in several continents are reviewed on a regular basis. Data from all sources continue to be reassuring regarding the safety of HPV vaccines. Over 12 years of research and monitoring have shown that HPV vaccines are very safe (Arbyn *et al.*, 2018; Suragh *et al.*, 2018; Bonaldo *et al.*, 2019). The side-effects of the licensed HPV vaccines are comparable with those of other vaccines. These include predominantly mild and transient local reactions at the site of injection (erythema, pain or swelling). Reported systemic adverse events of minor nature included headache, dizziness, myalgia, arthralgia, and gastrointestinal symptoms (nausea, vomiting abdominal pain). No patterns of serious adverse events were reported in trials or post-licensure surveillance suggesting a causal relation to the vaccine (World Health Organization, 2017; Arbyn *et al.*, 2018). Although case reports have identified a range of new-onset chronic conditions occurring after vaccination, including autoimmune diseases, a well-conducted population-based study on post-licensure safety surveillance showed no association between HPV vaccine and such conditions. Concerns have been raised about complex regional pain syndrome (CRPS), postural orthostatic tachycardia syndrome (POTS) and Chronic Fatigue Syndrome (CFS) following HPV vaccination (World Health Organization, 2017; Arbyn *et al.*, 2018). Despite the

difficulties in diagnosing these disorders, reviews of pre-and post-licensure data provide no evidence that there is a causal association with HPV immunization. The safety of the HPV vaccine has been also regularly reviewed by the Global Advisory Committee for Vaccine Safety (GACVS) of the World Health Organization (WHO) who have not identified any safety concerns and classified HPV vaccines as “extremely safe”, which is encouraged by the statements of many other national agencies.

Unfortunately, misinformation, insufficient preparedness and inadequate crisis response have led to controversy about the safety of HPV vaccines (Vorsters and Damme, 2018). Although the benefits of HPV vaccines are already very apparent in countries where the HPV vaccination programs have been implemented effectively, and despite robust evidence of HPV vaccine safety, uptake has remained low, also in many high-income countries (Bloem and Ogbuanu, 2017). To increase uptake, the lack of HPV vaccination confidence among families and healthcare providers needs to be overcome and decision-makers need to be up to date with the impact and importance of HPV vaccination.

Models on Impact and Cost-Effectiveness of HPV Vaccination

HPV vaccination programs have been shown to be cost-effective in a wide range of settings worldwide (World Health Organization, 2017). Nevertheless, assessment of the cost-effectiveness of HPV vaccines is heavily influenced by parameters including vaccine price, operational costs, HPV prevalence, number of vaccine doses, and uptake of cancer screening and treatment, especially in resource-constrained settings. Global cost-effectiveness analysis informed by country-based evidence suggests that vaccinating preadolescent girls is usually cost-effective, particularly in resource-constrained settings where alternative cervical cancer prevention and control measures often have limited coverage (World Health Organization, 2017). Although extending vaccination to older ages (known as HPV-FASTER (Bosch *et al.*, 2016)) or vaccinating boys as well as girls, could increase the speed and extent of the impact on HPV-related cancer outcomes, the cost-effectiveness of such strategies need to be established by public health authorities in the context of their own disease burden, economic and social priorities.

References

- Arbyn, M., Xu, L., Simoons, C., Martin-Hirsch, P.P., 2018. Prophylactic vaccination against human papillomaviruses to prevent cervical cancer and its precursors. *Cochrane Database of Systematic Reviews* 5, Cd009069. doi:10.1002/14651858.CD009069.pub3.
- Bloem, P., Ogbuanu, I., 2017. Vaccination to prevent human papillomavirus infections: From promise to practice. *PLOS Medicine* 14 (6), e1002325. doi:10.1371/journal.pmed.1002325.
- Bonaldo, G., Vaccheri, A., D'Annibali, O., Motola, D., 2019. Safety profile of human papilloma virus vaccines: An analysis of the US Vaccine Adverse Event Reporting System from 2007 to 2017. *British Journal of Clinical Pharmacology* 85 (3), 634–643.
- Bosch, F.X., Robles, C., Diaz, M., *et al.*, 2016. HPV-FASTER: Broadening the scope for prevention of HPV-related cancer. *Nature Reviews Clinical Oncology* 13 (2), 119–132.
- Bray, F., *et al.*, 2018. Global cancer statistics GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 68 (6), 394–424.
- Brotherton, J.M.L., 2019. Impact of HPV vaccination: Achievements and future challenges. *Papillomavirus Research* 7, 138–140. doi:10.1016/j.pvr.2019.04.004.
- Brown, D.R., *et al.*, 2009. The impact of quadrivalent human papillomavirus (HPV; types 6, 11, 16, and 18) L1 virus-like particle vaccine on infection and disease due to oncogenic nonvaccine HPV types in generally HPV-naïve women aged 16–26 years. *The Journal of Infectious Diseases* 199 (7), 926–935.
- de Martel, C., *et al.*, 2017. Worldwide burden of cancer attributable to HPV by site, country and HPV type. *International Journal of Cancer* 141 (4), 664–670.
- de Sanjose, S., *et al.*, 2018. Burden of human papillomavirus (HPV)-related cancers attributable to HPV types 6/11/16/18/31/33/45/52 and 58. *JNCI Cancer Spectrum* 2 (4), pky045.
- Drolet, M., Benard, E., Perez, N., Brisson, M., 2019. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: Updated systematic review and meta-analysis. *Lancet* 394 (10197), 497–509. doi:10.1016/s0140-6736(19)30298-3.
- Einstein, M.H., *et al.*, 2014. Comparison of long-term immunogenicity and safety of human papillomavirus (HPV) — 16/18 AS04-adjuvanted vaccine and HPV-6/11/16/18 vaccine in healthy women aged 18–45 years: End-of-study analysis of a Phase III randomized trial. *Human Vaccines & Immunotherapeutics* 10 (12), 3435–3445.
- Fortes, H.R., *et al.*, 2017. Recurrent respiratory papillomatosis: A state-of-the-art review. *Respiratory Medicine* 126, 116–121.
- Gillison, M.L., *et al.*, 2012. Human papillomavirus and diseases of the upper airway: Head and neck cancer and respiratory papillomatosis. *Vaccine* 30 (Suppl 5), F34–F54.
- Group, I.H.W., 2014. Primary End-Points for Prophylactic HPV Vaccine Trials. *International Agency for Research on Cancer*.
- Guo, F., Cofie, L.E., Berenson, A.B., 2018. Cervical cancer incidence in young U.S. females after human papillomavirus vaccine introduction. *American Journal of Preventive Medicine* 55 (2), 197–204.
- Hartwig, S., *et al.*, 2017. Estimation of the overall burden of cancers, precancerous lesions, and genital warts attributable to 9-valent HPV vaccine types in women and men in Europe. *Infectious Agents and Cancer* 12, 19.
- IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012. Biological agents. Volume 100 B. A review of human carcinogens. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* 100 (Pt B), 1–441.
- Joua, E.A., *et al.*, 2015. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *The New England Journal of Medicine* 372 (8), 711–723.
- Kreimer, A.R., *et al.*, 2018. Evidence for single-dose protection by the bivalent HPV vaccine – Review of the Costa Rica HPV vaccine trial and future research studies. *Vaccine* 36 (32 Pt A), 4774–4782.
- Lacey, C.J., 2019. HPV vaccination in HIV infection. *Papillomavirus Research* 8, 100174.
- Lehtinen, M., Dillner, J., 2013. Clinical trials of human papillomavirus vaccines and beyond. *Nature Reviews Clinical Oncology* 10 (7), 400–410. doi:10.1038/nrclinonc.2013.84.
- Liu, Z., Rashid, T., Nyitray, A.G., 2016. Penises not required: A systematic review of the potential for human papillomavirus horizontal transmission that is non-sexual or does not include penile penetration. *Sex Health* 13 (1), 10–21.
- Lowy, D.R., Schiller, J.T., 2006. Prophylactic human papillomavirus vaccines. *Journal of Clinical Investigation* 116 (5), 1167–1173.
- Luostarinen, T., *et al.*, 2018. Vaccination protects against invasive HPV-associated cancers. *International Journal of Cancer* 142 (10), 2186–2187.
- Moscicki, A.B., 2007. HPV infections in adolescents. *Disease Markers* 23 (4), 229–234.
- Naud, P.S., *et al.*, 2014. Sustained efficacy, immunogenicity, and safety of the HPV-16/18 AS04-adjuvanted vaccine: Final analysis of a long-term follow-up study up to 9.4 years post-vaccination. *Human Vaccines & Immunotherapeutics* 10 (8), 2147–2162.

- Novakovic, D., *et al.*, 2018. A prospective study of the incidence of juvenile-onset recurrent respiratory papillomatosis after implementation of a national HPV vaccination program. *The Journal of Infectious Diseases* 217 (2), 208–212.
- Nygard, M., *et al.*, 2015. Evaluation of the long-term anti-human papillomavirus 6 (HPV6), 11, 16, and 18 immune responses generated by the quadrivalent HPV vaccine. *Clinical and Vaccine Immunology* 22 (8), 943–948.
- Palmer, T., *et al.*, 2019. Prevalence of cervical disease at age 20 after immunisation with bivalent HPV vaccine at age 12–13 in Scotland: Retrospective population study. *The BMJ* 365, I1161.
- Pattyn, J., *et al.*, 2019. Infection and vaccine-induced HPV-specific antibodies in cervicovaginal secretions. A review of the literature. *Papillomavirus Research* 8, 100185.
- Safaeian, M., *et al.*, 2018. Durability of protection afforded by fewer doses of the HPV16/18 vaccine: The CVT Trial. *Journal of the National Cancer Institute* 110 (2).
- Sankaranarayanan, R., *et al.*, 2018. Can a single dose of human papillomavirus (HPV) vaccine prevent cervical cancer? Early findings from an Indian study. *Vaccine* 36 (32 Pt A), 4783–4791.
- Schiller, J., Lowy, D., 2018. Explanations for the high potency of HPV prophylactic vaccines. *Vaccine* 36 (32 Pt A), 4768–4773. doi:10.1016/j.vaccine.2017.12.079.
- Schwarz, T.F., *et al.*, 2017. Ten-year immune persistence and safety of the HPV-16/18 AS04-adjuvanted vaccine in females vaccinated at 15–55 years of age. *Cancer Medicine* 6 (11), 2723–2731.
- Scotland, I., 2019. Cancer Statistics – Female Genital Organ Cancer. Available at: <https://www.isdscotland.org/Health-Topics/Cancer/Cancer-Statistics/Female-Genital-Organ>.
- Stanley, M., 2010. HPV – Immune response to infection and vaccination. *Infectious Agents and Cancer* 5, 19.
- Stanley, M., Dull, P., 2018. HPV single-dose vaccination: Impact potential, evidence base and further evaluation. *Vaccine* 36 (32 Pt A), 4759–4760. doi:10.1016/j.vaccine.2018.02.076.
- Stanley, M., 2017. Tumour virus vaccines: Hepatitis B virus and human papillomavirus. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372 (1732), doi:10.1098/rstb.2016.0268.
- Suragh, T.A., *et al.*, 2018. Safety of bivalent human papillomavirus vaccine in the US vaccine adverse event reporting system (VAERS), 2009–2017. *British Journal of Clinical Pharmacology* 84 (12), 2928–2932.
- Vorstes, A., Van Damme, P., 2018. HPV immunization programs: Ensuring their sustainability and resilience. *Vaccine* 36 (35), 5219–5221. doi:10.1016/j.vaccine.2018.06.066.
- WHO, 2017. Immunization, Vaccines and Biologicals database as of 31 March 2017. Available at: https://www.who.int/entity/immunization/monitoring_surveillance/VaccineIntroStatus.pptx. (Accessed October 2019).
- WHO, 2019. Global Market Study: HPV Vaccines. World Health Organization.
- Wong, C.A., *et al.*, 2011. Approaches to monitoring biological outcomes for HPV vaccination: Challenges of early adopter countries. *Vaccine* 29 (5), 878–885.
- World Health Organization, 2017. Human papillomavirus vaccines: WHO position paper. *Vaccine* 35 (43), 5753–5755. doi:10.1016/j.vaccine.2017.05.069.
- Xu, L., Selk, A., Garland, S.M., *et al.*, 2019. Prophylactic vaccination against human papillomaviruses to prevent vulval and vaginal cancer and their precursors. *Expert Review of Vaccines* 18 (11), 1157–1166.
- zur Hausen, H., 1977. Human papillomaviruses and their possible role in squamous cell carcinomas. *Current Topics in Microbiology and Immunology* 78, 1–30.

Relevant Websites

- <https://www.hpvcentre.net>
HPV INFORMATION CENTRE.
- <https://www.cdc.gov/hpv/>
HPV, the Vaccine for HPV, and Cancers Caused by HPV/CDC.
- <https://www.hpvworld.com>
HPVWorld. - HPV Research Articles.
- www.hpvboard.org
HPV Prevention and Control Board.
- https://www.who.int/immunization/policy/position_papers/hpv/en/
Human Papillomavirus (HPV) position paper - WHO.
- <https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-vaccine-fact-sheet>
Human Papillomavirus (HPV) Vaccines.

Influenza Vaccination

Topi Turunen, Infectious Disease Unit, Espoo, Finland and Finnish Institute for Health and Welfare, Helsinki, Finland

© 2021 Elsevier Ltd. All rights reserved.

Introduction

Influenza is of great public health significance. The yearly epidemics and occasional pandemics of this respiratory illness give rise to a considerable burden of disease worldwide. Seasonal influenza vaccines are the principal method of preventing influenza infection and complications in humans.

Full coverage from the influenza virus and its epidemiology is covered elsewhere in the encyclopedia. Here, the characteristics and epidemiology of influenza viruses are discussed mainly as they relate to vaccination. The bulk of this chapter focuses on seasonal influenza vaccines; vaccination against pandemic influenza is described in the end.

Influenza as a Disease

The symptoms of influenza illness include fever, myalgia, headache, chills, sore throat, general malaise, and cough, although children and the elderly may present with less specific symptoms. People at both extremes of age are also more prone to developing complications of influenza such as secondary bacterial pneumonia or exacerbation of an existing pulmonary or cardiovascular condition.

Understanding the transmission of influenza is crucial to preventive efforts. The illness is spread by aerosols and droplets when infected people cough or sneeze, as well as by fomites. Notably, viral shedding in infected individuals begins before symptoms appear, and viruses can be transmitted to others during this period as well as during asymptomatic infections that may comprise a large part of all influenza infections.

In clinical settings, the diagnosis of influenza illness is typically based on symptoms, possibly aided by laboratory tests. Real-time reverse-transcription polymerase chain reaction (RT-PCR) is often considered the gold standard but antigen tests, serology, and viral culture diagnostic methods are also available. The sensitivity and specificity of tests used for diagnosis are relevant not only for the correct diagnosis and treatment of the patient but also for estimating influenza vaccine effectiveness (discussed in more depth below).

Epidemiology of Influenza

Influenza viruses are divided into types A, B, and C – as well as subtypes and lineages – based on their genetics and surface antigens. Typically, there are several types and subtypes of influenza A and B viruses cocirculating during any given season, necessitating the inclusion of several viral strains in each seasonal vaccine. Circulating viruses since the 2009 pandemic have included the A(H1N1)pdm09, A(H3N2), B/Victoria, and B/Yamagata viruses.

Influenza viruses cause annual global epidemics. In the temperate regions of the northern and southern hemispheres, influenza activity is typically pronounced in the fall and winter seasons while in the tropics, influenza activity can occur around the year.

Influenza viruses possess the capacity for antigenic drift: A constant change in their surface antigens. This gives circulating influenza viruses an evolutionary edge because the changed viruses may not be recognized by immune systems based on previous exposure (whether natural infection or vaccination). Antigenic drift, combined with waning immunity post-vaccination, necessitates annual reformulation and administration of seasonal influenza vaccines.

Occasionally, antigenic shift between different viruses (such as human and animal influenza viruses) occurs and results in a completely novel type of virus. Such viruses, if transmitted easily between humans, can give rise to global pandemics of varying severity. While seasonal influenza vaccines may sometimes provide cross-protection against pandemic influenza strains, the prevention of pandemic influenza will typically require a specifically developed pandemic vaccine.

Brief History of Influenza Vaccines

Influenza viruses were first isolated in 1933 and the development of the first influenza vaccines began shortly thereafter. The first commercial vaccines were developed and tested in the United States in the shadow of the Second World War and approved for use in 1945.

The first vaccines were composed of whole viruses. Other forms of vaccines (variations of split virion, subunit, and live attenuated vaccines) have been introduced over the years. For most of the time, eggs have been used as a growth medium for vaccine viruses; cell culture became available later. Production methods have been refined over the years; for example, purification steps have been added to reduce the number of unwanted proteins and other substances in the finished product.

Production of Influenza Vaccines

Timeline of the Production Process

Due to the epidemiology of influenza – in particular, its seasonal nature – the strain composition of the vaccines is updated annually for both northern and southern hemispheres. The strain selection and the ensuing production process take place according to a certain schema each year.

The choice of virus strains is based on an international surveillance effort. National influenza centers in various countries around the world monitor the circulating viruses and send representative samples to the World Health Organization's (WHO) collaborating centers. The surveillance system aims to detect patterns in genetic and antigenic changes of the viruses that would be the most relevant for protection in the next influenza season. Viral isolates thought to represent the best candidates for the vaccine strains are collected, grown in culture media, and distributed to vaccine manufacturers for scaling up production.

The production (notably, the growth of viruses in eggs) takes several months and needs to be finished in time for the produced vaccines to be distributed and administered before the next seasonal epidemic. As a result, the consultation to select vaccine strains usually takes place in February for the northern hemisphere and in September for the southern hemisphere. The production delay sometimes gives rise to antigenic mismatch if the circulating strains' evolution unfolds in an unexpected way through antigenic drift and shift. This may have a negative impact on vaccine effectiveness.

Production Steps

Production capacity for influenza vaccines exists in several countries but is mostly centered in the global North and in a handful of private companies. However, nationally manufactured influenza vaccines are also used in many parts of the world.

Each manufacturer of influenza vaccines uses its own proprietary process. While production processes differ in some details, all share certain key steps and result in vaccines of comparable safety; immunogenicity can be verified using standardized methods.

The key production steps are as follows (please note that recombinant, cell-based, and live-attenuated vaccines do not necessarily share all of these):

- Propagation in hen eggs (or other media);
- Disruption of viral lipid envelope using one of several possible detergents;
- Purification steps;
- Establishing the immunogenicity of monovalent vaccine components and combining them in a multivalent (typically, tri- or quadrivalent) formulation; and
- Insertion in vials, labeling, packaging, allocation of lot numbers.

Dependence on hen eggs for growing the viruses leads to several potential problems. Most obviously, the large number of eggs required presents a limit to how many vaccine doses may be produced. A surprising unavailability of eggs, due to, say, an outbreak of avian influenza, may halt production. On the other hand, a problem of "egg adaptation" – where egg-grown viruses change in a way that does not resemble the circulating ones – has been suspected to partially account for lower vaccine effectiveness against some strains – A(H3N2), in particular.

The production of live-attenuated influenza vaccines (LAIV) is somewhat different. LAIV is based on attenuated master donor strains that are cold-adapted (i.e., can replicate in the relatively cooler temperature of human upper airways but not the lower airways) and thus, should not cause significant illness in humans. The production process makes use of the segmented genome of the virus that permits genetic reassortment. Surface antigen-encoding genes of newly emergent wild-type viruses are inserted into master donor viruses; the resulting virus passes the attenuation characteristic to its progeny but expresses the relevant antigens for the season.

Influenza Vaccine Composition

All influenza vaccines contain viral proteins, of which hemagglutinin (HA) is thought to be the main immunogen. Standardization processes aim to keep HA levels in vaccines constant. Other influenza surface and structural viral antigens present in vaccines include neuraminidase (NA), matrix protein, and nucleoprotein. They may contribute to the vaccine-elicited immune response but their levels are typically not quantified.

Adjuvants are used in certain vaccine formulations to increase the immunogenicity and effectiveness of the vaccine. Commercially-used adjuvants include emulsions of oil and water (AS03 and MF59), aluminum compounds, and influenza virosomes (where the presence of influenza viral lipids is thought to contribute to a greater immune response).

Modern vaccine production techniques use sequential purification steps to reduce the number of unnecessary or unwanted substances in the vaccine. However, depending on the specific production method of the vaccine in question, trace amounts of the following may be detectable:

- Egg proteins: In vaccine products grown in hen eggs, trace amounts of proteins, such as ovalbumin, may be present. The amount is typically low enough to be insignificant even in most cases of egg allergy, but may be relevant if past anaphylactic reaction (see Section "Safety and Contraindications" below).
- Antibiotics are sometimes used to prevent bacteria from growing in the virus cultures. In the case of influenza vaccines, these are typically aminoglycosides such as gentamicin or neomycin. Penicillins, cephalosporins, and sulfonamides are not used in vaccine production.
- Detergents used to disrupt viruses (largely removed in purification stage).
- Formaldehyde, from formalin used to inactivate viruses in inactivated vaccines. While toxic in larger quantities, the trace amounts per vaccine dose are insignificant compared to formaldehyde production in human metabolism and environmental exposure.

- Gelatin may be used as a stabilizer (i.e., to keep the vaccine effective during storage until use). While gelatin is derived from pigs, the form used in vaccines has been purified and hydrolyzed, therefore different from gelatin found in foodstuffs. Muslim and Jewish religious communities have generally accepted the administration of gelatin-containing vaccines.
- Thiomersal is a mercury-based compound used earlier to prevent the growth of microbes during the production process or in multi-dose vaccine vials. There is no evidence that thiomersal in vaccines cause harm, but it has largely been removed from vaccines in the US and Europe as a precaution. Consequently, most current vaccine products contain no thiomersal even in trace amounts.

Valency

Influenza vaccines can be characterized by their valency, or how many viral components they contain. In the history of seasonal influenza vaccines, as little as one (monovalent) and many as five (pentavalent) components have been used, but most seasonal vaccines are nowadays tri- or quadrivalent – that is, they contain both circulating influenza A components (currently, H1N1 and H3N2) and either one or both of the two influenza B components (B/Victoria and B/Yamagata). Pandemic vaccines are usually monovalent.

The valency of the vaccine has practical implications if the circulating strains turn out to be different from the ones included in the vaccine. Choosing the “correct” influenza B lineage to be included in trivalent influenza vaccines prove difficult at times. In some influenza B-mismatched seasons, this has resulted in suboptimal vaccine effectiveness but in some seasons, even the trivalent vaccine has demonstrated cross-protection to both influenza B lineages. On the other hand, in seasons with low influenza B activity or a good match to vaccine lineage, the valency of the vaccine has mattered little.

Dosage

The HA content of current vaccines is standardized and quantified through a process called single radial immunodiffusion (SRID). The HA level measured by SRID correlates with the immunogenicity of the vaccine.

Standard-dose vaccines contain 15 µg of HA per viral strain included in the vaccine (generally three or four for the seasonal trivalent and quadrivalent vaccines, respectively). More recently, high-dose vaccines that contain 60 µg of each strain’s HA have been licensed. The high-dose vaccines are discussed in more depth below.

Formulations of Seasonal Influenza Vaccines

Nowadays, many different forms of influenza vaccines exist; others are in development. Many products seek to improve the efficacy and effectiveness of the vaccines, especially in key target groups such as the elderly where the effectiveness of traditional vaccines may be suboptimal.

It would be of great interest to public health decision-makers to know if some vaccine products are better than others and which represent the best effectiveness-safety profile for the various target groups of vaccination. However, as of now, few head-to-head comparisons have been made, and drawing too many conclusions from individual studies is problematic for reasons explored in Section “Evaluating the Impact of Influenza Vaccines”.

Traditional Inactivated Vaccines

These are the most commonly used influenza vaccines worldwide: egg-based, nonadjuvanted, standard-dose inactivated vaccines. All the other vaccines discussed below differ from these “traditional” vaccines regarding one or more of the above attributes. Most available evidence of influenza vaccine efficacy, effectiveness, and safety derives from studies with these “typical” inactivated vaccines.

Even inside this group, inactivated vaccines differ by their valency (see Section “Production of Influenza Vaccines”). Trivalent and quadrivalent seasonal influenza vaccines are most commonly used today; many countries are shifting towards using quadrivalent vaccines.

Recombinant and Cell-Based Influenza Vaccines

These vaccine types circumvent the egg-growing phase; all other influenza vaccines are currently grown in hen eggs. This approach is presumed to avoid some of the pitfalls with using eggs (see Section “Production of Influenza Vaccines”).

A trivalent recombinant vaccine was licensed in the US in 2013 and has since been replaced by a quadrivalent version. In preparing recombinant vaccines, the virus nucleic acid encoding the surface protein HA is combined with a baculovirus vector that transports the gene into a host cell line to produce the antigen, then collected and purified. Recombinant vaccines have been licensed for people aged 18 and above. There is limited evidence on the comparative efficacy or effectiveness of recombinant vaccines, but one randomized controlled trial (RCT) conducted in the A(H3N2)-predominant 2014-2015 season found a 30% lower probability of influenza infection (in relative terms) in recombinant versus egg-based vaccines.

Cell-based vaccines are grown in mammalian cells. A cell-based influenza vaccine production method was first approved by the US Food and Drug Administration (FDA) in 2012. At the time, it still involved egg-grown candidate vaccine viruses but since 2016,

the process has been changed to be completely egg-free. As with recombinant vaccines, evidence of comparative effectiveness is limited. One observational study looking at the A(H3N2)-dominated 2017–2018 influenza season found a quadrivalent cell-based vaccine to have a relative vaccine effectiveness of approximately 10% points higher than a quadrivalent egg-based vaccine.

Live Attenuated Influenza Vaccines

LAIV use cold-adapted weakened live influenza viruses and are administered through an intranasal spray. Because of the cold-adaptation, the viruses can replicate in the upper respiratory tract and cause an immune response similar to that evoked by a natural infection. They do not replicate in the lower airways (thus, limiting the risk of vaccine-induced lower respiratory infection). Live attenuated influenza vaccines have been used in countries of the former Soviet Union since the 1980s. In 2003, another type of LAIV was licensed in the US in a trivalent formulation that has since been replaced with a quadrivalent version.

The effectiveness of LAIV seems generally the best in younger age groups, and in the EU, the vaccines are only indicated for people aged 2–17 years (in the US, 2–49 years). Being live virus vaccines, they have somewhat different contraindications than inactivated vaccines.

The effectiveness has generally been comparable to that of inactivated vaccines. After the 2009 pandemic, some studies showed reduced effectiveness against A(H1N1), leading US authorities to temporarily recommend against its use in 2016. This recommendation was eventually dropped after the A(H1N1) component of the vaccine was changed.

Adjuvanted Vaccines

Adjuvants have been used to increase the immunogenicity of vaccines and result in increased post-vaccination antibody titers. Besides increased immunogenicity, a potential benefit of adjuvanted vaccines is dose-sparing in the face of a large global demand for vaccines such as in the event of a pandemic. However, injection site reactions may be more pronounced with adjuvanted vaccines.

The clinical efficacy and effectiveness of adjuvanted vaccines have compared favorably to nonadjuvanted inactivated vaccines, but evidence from direct head-to-head comparisons is scarce. The use of adjuvanted seasonal influenza vaccines focuses on the elderly population. Studies have also been performed in pediatric populations. AS03 and MF59 -adjuvants were used in the 2009 H1N1 pandemic vaccines which were widely used across age groups.

High-Dose Influenza Vaccines

A high-dose vaccine resembles the standard-dose inactivated vaccine in composition but contains four times the antigen per viral strain. The higher dose antigen dose is intended to evoke a stronger immune response in the elderly, for whom the vaccine is indicated. To date, available evidence of comparative efficacy/effectiveness is limited, but high-dose vaccines have been associated with better immunogenicity (higher antibody titers post-vaccination) and somewhat better efficacy against laboratory-confirmed influenza (LCI) and hospitalizations.

A trivalent high-dose vaccine was licensed for use in the United States in 2009. A quadrivalent formulation has since been developed and FDA approved in 2019. The vaccines are indicated for use in adults aged 65 years and older.

Intradermal Vaccines

When a vaccine is injected intradermally (into the dermal layer of skin instead of the muscle), similar immunogenicity can be achieved using less antigen. The vaccination process uses a smaller needle, which some vaccinees will perceive as an upside but may generate more localized redness and swelling.

In the United States, intradermal vaccines are available for use in adults aged 18–64 years. A trivalent form has been in use since 2012 and a quadrivalent formulation was approved in 2014.

The Future of Influenza Vaccines

The need for annual revaccination and the occasionally suboptimal vaccine performance due to mismatch and other reasons have led to hopes of “universal” influenza vaccines. These would be broadly cross-protective vaccines against conserved antigenic targets on influenza viruses such as the HA stalk domain and other surface components. The idea is that one vaccine could protect against several types of influenza across several seasons, and possibly even pandemic strains.

In a universal vaccine, a conserved antigenic target could be presented to the immune system conjugated to another molecule or in a carrier system such as a virus-like particle. Virus-like particles contain viral lipid membrane-anchored proteins in their native state but lack the virus nucleic acid for safety reasons.

Some universal vaccine candidates have shown some promise in preclinical and clinical studies but currently, public health professionals must accept that universal vaccines will not be available for some time. Meanwhile, further development of existing vaccines can take place as well as research to identify the vaccine products that have the best safety-effectiveness profile for each of the various target groups.

New administration methods such as a microneedle patch have been investigated. Patches with small dissolvable needles have demonstrated good immunogenicity and safety in studies. Such a vaccine, much like the current LAIV, could be self-administered, potentially reducing the need for health care personnel in vaccination campaigns. How to address the rare but possible severe allergic reactions should be addressed before launching such campaigns.

Vaccination in Practice

Administration Routes

Inactivated influenza vaccines are most commonly given by intramuscular, subcutaneous, or intradermal route. Also, intranasal and oral routes have been studied. Live attenuated vaccines are given intranasally.

Timing of Programs

Influenza vaccination programs typically begin between October and December in the northern hemisphere and between April and July in the southern hemisphere. In some years, programs may be delayed due to holdup in vaccine production.

Vaccine Uptake

Vaccine uptake varies widely between countries and vaccination target groups but is seldom as high as recommended. Efforts beyond the mere biomedical domain are needed to understand why some people at high risk for severe influenza choose to remain unvaccinated even when offered vaccination.

Duration of Protection

Most vaccinees develop protective serum antibody levels around two weeks post-vaccination. In principle, the duration of protection is dependent on at least two aspects: changes in the circulating viruses and whether existing antibodies are still protective against the drifted viruses, and waning immunity or reduced antibody levels over time.

Protection against influenza for longer than one year has been documented both post-vaccination and after natural infection with influenza virus. Based on evidence from past epidemics and pandemics, natural infection in childhood or young adulthood may provide protection to antigenically similar viruses even several decades later. However, annual immunization is recommended because of reductions in serum antibody levels and because one or more vaccine antigens are usually updated each year.

Concerns have been raised that protection conferred by the vaccine might wane even during a given season. This has sometimes been observed in studies that track vaccine effectiveness (VE) at several points during a season. Potential explanations include intra-seasonal evolving mismatches between circulating viruses, waning immunity (especially in the elderly), and methodological bias. Regarding the latter, the unvaccinated may develop symptomatic or asymptomatic influenza infections during the season and gain protection from these natural encounters. Cumulative natural encounters in the population may therefore bias VE towards null.

Safety and Contraindications

Overview

Seasonal influenza vaccines have been prepared in similar ways for decades and their safety is well established. Adverse effects tend to be mild and benign, and serious adverse events are rare. This chapter includes both safety and contraindications of influenza vaccines.

Local and Systemic Inflammatory Reactions

The most frequent adverse events of influenza vaccines are local reactions. Characteristic signs of inflammation (pain, redness, swelling) can be observed in up to 65% of vaccinees. The local reactions rarely affect daily activities and generally do not persist for more than one or two days.

Systemic inflammatory reactions (fever, myalgia, headache) occur less frequently, in < 15% of vaccinees. They are more common in children than adults. Febrile seizures have occasionally been reported in small children, mostly in children under five years of age.

The occurrence of local and systemic inflammatory reactions is dependent on vaccine dosage and the presence of adjuvant.

Allergic Reactions

Allergic reactions to influenza vaccines may be immunoglobulin E (IgE)-mediated or non-IgE-mediated. Potential antigens that can lead to allergic reactions include egg proteins (such as ovalbumin), trace amounts of the preservative thiomersal (increasingly rare), and possibly – though unlikely – latex or other rubber compounds found in the vaccine vial and introduced in tissue through the needle, puncturing the vial seal.

Anaphylaxis, a potentially life-threatening but rare serious adverse event of influenza vaccines, is an example of an IgE-mediated reaction and may take place in approximately one per 1,000,000 vaccinations. Persons with a history of anaphylaxis following the administration of an influenza vaccine should not be revaccinated with a similar vaccine. It is, however, important to note that most allergic reactions are not anaphylaxis and rarely prevent revaccination when the vaccine is otherwise indicated. If needed, the vaccine can be given under the physician's observation, or with sufficient follow-up time.

Rare Adverse Events: Neurological Syndromes

Guillain-Barré syndrome (GBS) is a rare condition that has been associated with many infections of the respiratory and gastrointestinal tracts, including naturally-occurring influenza infection. Its symptoms include ascending muscle paralysis and paresthesia – which, when not treated with assisted ventilation, can be fatal. However, most GBS patients do recover. During the 1970s, an increased incidence of GBS (of roughly 1:100,000) was reported in vaccines. This finding was not replicated for several years until, in the 1990s, a small increase (incidence of 1:1,000,000) was observed in vaccinees. In some studies, a protective effect of seasonal influenza vaccines on GBS has been seen. The exact relationship between seasonal influenza vaccination and GBS has been difficult to untangle because of the very low incidence of GBS. It should be noted that even if the one-in-a-million risk of GBS was consistently observed, it would still be outweighed by the protective effect of the vaccine on influenza illness and its severe complications.

Evaluating the Impact of Influenza Vaccines

Studies of influenza vaccines' impact apply various outcome measures. These include immunogenicity, real-life prevention of LCI, and various less specific outcomes such as acute respiratory illness (specifically defined influenza-like illness or any medically-attended respiratory infection), hospitalization, overall mortality, or seroconversion to circulating influenza strains.

Immunogenicity

Vaccination induces the production of protective antibodies targeted at influenza virus antigens, most notably HA. The level of HA antibodies can be detected by various methods. In addition to the humoral immune response, the cellular T cell response plays a role in influenza immunity. The strength of the immune response depends on age and previous exposure to influenza viruses and influenza vaccines, as well as underlying health conditions and the formulation of the vaccine (such as dosage and presence of an adjuvant). Usually, antibody levels peak at around two weeks after vaccination but the response may take longer to form and be less pronounced in the elderly. The same is true of persons without any previous exposure; this is why two doses are recommended for vaccine-naïve children.

While HA antibodies generally correlate with protection from influenza illness, the correlation between antibody levels and clinically significant protection from influenza illness are not always clear. Therefore, vaccine efficacy and effectiveness are more relevant for vaccination policy-makers, something that has been recently reflected in the regulatory field.

Vaccine Efficacy

Vaccine efficacy refers to the percentage reduction of clinical illness between the vaccinated and unvaccinated groups under favorable conditions. This type of information is typically collected in RCTs where the investigators set up an unvaccinated (or vaccinated with a non-influenza vaccine) control group and have control over many variables of the study design.

In the presence of vaccination policies recommending annual influenza vaccines to various target groups, running placebo-controlled RCTs on influenza vaccine efficacy has become increasingly rare and the role of real-life vaccine effectiveness studies is becoming more and more important.

Vaccine Effectiveness

VE refers to the percentage reduction of clinical illness between the vaccinated and unvaccinated groups under conditions considered typical – or “real-world” situations. This type of information is obtained from observational studies where the vaccinated and unvaccinated groups are not assigned by investigators.

Observational study designs used to measure influenza vaccine effectiveness include:

- Cohort studies, where groups of vaccinated and unvaccinated people are followed (either prospectively or retrospectively) and those ill with influenza are identified.
- Case-control studies, where the vaccination status of people ill with influenza (cases) is compared to those who were not ill (controls). A popular variant is the test-negative design, where people with influenza-like symptoms are screened for the virus and assigned as cases or controls depending on the test result.
- Screening method, which combines the influenza-positive cases' individual vaccination status and ecologic data on vaccine coverage in the population where the individual came from.

Observational study designs have strengths and weaknesses compared to RCTs. RCTs are, by design, better at managing bias and confounding than observational study designs. However, proper choices at study planning, execution, and data analysis stages can considerably mitigate the risk of bias in observational studies. Considering also the cost and feasibility as well as ethical issues, data from observational studies are widely used as a basis for public health policy and are increasingly recognized by regulators.

To be informative, observational studies need to obtain the necessary sample size and statistical power. Failure to recruit enough patients is compounded when studying VE in various individuals (e.g., children, adults, elderly) or the VE of several vaccine products to assess brand-specific effectiveness. Multicenter studies with compatible protocols are one way to circumvent this problem. Another approach is to use the comprehensive registers for vaccinations, laboratory findings of influenza, and other covariates needed for analysis that many countries and regions have nowadays.

Understanding Vaccine Efficacy and Effectiveness

Influenza vaccines, unlike any other vaccines, need to be administered annually and have variable efficacy and effectiveness. The effectiveness varies across several dimensions – from year to year and between groups of vaccine recipients. Before delving deeper into available evidence on vaccine effectiveness, it is important to recognize some determinants that may influence the results of the studies. Both naturally-occurring variations in vaccine efficacy and effectiveness and topics related to study design and analytical methods must be considered.

First, there are biological factors that influence the actual vaccine efficacy or effectiveness:

- Level of match/mismatch between circulating and vaccine viruses.
- Population studied: Immune responses are typically better in children and healthy adults than with the elderly and the chronically ill.
- Vaccine type used: Valency, dose, presence of adjuvant, and other topics (see Section “Formulations of Seasonal Influenza Vaccines”) may affect the immune response.
- Waning protection (see Section “Duration of Protection”).
- Repeated vaccinations: It has been postulated that vaccinations repeated in subsequent years can cause positive or negative interference with an increased or decreased immune response. Some studies have observed signals of negative interference with certain seasons and viral strains; however, a meta-analysis found no overall evidence that prior season vaccination would impact current-season VE negatively. An increasing number of studies now include vaccination in one or more previous seasons as a covariate in order to better understand this phenomenon.

Second, there are factors related to study design and execution that may explain why studies sometimes yield different results:

- Outcome studied: Vaccine effectiveness against LCI is different from vaccine effectiveness against influenza-like illness or overall mortality (only a fraction of which is attributable to influenza).
- Study power: The optimal sample size depends on the chosen study design, influenza attack rate, vaccine coverage, and expected vaccine effectiveness. Lack of statistical power leads to uncertainty represented by wide confidence intervals which mean the results should be interpreted with caution.
- Bias that may arise from many different sources and cause the perceived VE to be different from true VE. Are the data on influenza and vaccination status accurate and complete, or can there be misclassification? Are the subjects selected in a representable way? Are vaccinated people on average healthier or more frail than unvaccinated? Are they more or perhaps less likely to seek healthcare services?.
- Confounding and its adjustment in study design and statistical analysis process. In influenza vaccine studies, typically “crude” and “adjusted” figures are given, with the latter representing correction for confounding and thus likely more representative of a true causal relationship between vaccination and influenza (rather than mere statistical correlation).
- Multi-center studies and pooling. Often, data from several smaller studies are pooled together in a meta-analysis. This increases the probability to detect an effect of vaccination but requires the individual studies to be sufficiently heterogeneous.

Vaccination of Specific Population Groups

As an annually recurring epidemic that routinely attains attack rates of 10% or more and results in many hospitalizations and deaths, influenza is of great public health importance. Effective preventive measures to lessen its impact on the population and health systems are therefore valuable. Among preventive measures, annual seasonal influenza vaccination is likely the single most important one.

Who, however, should be vaccinated against influenza, and how effective are the vaccines in the different population groups? Vaccination policy varies between different countries. The United States has recommended annual influenza vaccination of all persons aged six months and above since 2010, but such universal vaccination recommendations are globally rare. Many countries have implemented vaccination programs that address specific risk groups. In the global south, influenza vaccines are seldom prioritized among public health measures and generally utilized less.

Vaccination policy is a varied and evolving topic. It is not practical to list here all vaccination recommendations around the world; instead, some target groups for vaccination and the rationale for vaccinating them will be discussed. There are some unique challenges related to estimating and communicating the impact of influenza vaccines, but it is crucial to understand that even a suboptimally effective vaccine can provide a large public health impact because of the high incidence of influenza and the potential decrease in disease severity and complications, even when the vaccine does not completely prevent the clinical illness.

Elderly

The elderly (often defined as people aged 65 years and above) have long been recognized as a high-risk group for severe influenza and influenza complications. They are among the most universally recommended target groups to receive the annual influenza vaccination. However, antibody responses to vaccination are decreased with age and decreased protective VE has been shown in studies. This puts the elderly in a seemingly paradoxical situation: They are among the ones that would benefit the most from vaccination and also the ones who seem to get the least protection. Because annual influenza vaccination has been recommended to the elderly for a long time, there are few recent placebo-controlled trials. Annual data of vaccine effectiveness are gathered from observational studies.

The elderly generally benefit from annual influenza vaccination. Vaccine efficacy or effectiveness against symptomatic LCI has been investigated in many studies and is commonly in the 20%–60% range with considerable season-to-season variation. However, in some seasons, especially when a significant mismatch occurs, statistically significant effects of influenza vaccines have not been found for this age group. Influenza vaccination might also reduce the risk for influenza-related hospitalizations. Some studies have looked into less specific outcomes such as all-cause mortality or any hospitalizations, but these studies run a high risk of bias and should generally be interpreted with caution.

The desire to improve vaccine effectiveness among the elderly has led to the development of vaccines intended to promote a better immune response in this population. Besides traditional inactivated vaccines, some vaccines of interest in the elderly group are adjuvanted, recombinant, and high-dose vaccines described more in Section “Formulations of Seasonal Influenza Vaccines”. Still, most effective data in the elderly population are derived from traditional inactivated vaccines.

Underlying Conditions

Underlying conditions such as chronic pulmonary, cardiovascular, metabolic, renal, hepatic, neurologic, malignant, and immunosuppressive conditions (due to illness or iatrogenic) are risk factors for complications with influenza and often considered indications for influenza vaccination. However, data on which conditions pose the greatest risk and where might the influenza vaccination be especially important are sometimes sparse and often heterogeneous.

Influenza can trigger exacerbations of asthma and chronic obstructive pulmonary disease (COPD). Evidence of the protective effect of influenza vaccines against chronic pulmonary conditions focuses on asthma, where some studies have shown protection against hospitalization and decreased use of bronchodilators and systemic steroid medicines commonly used to alleviate exacerbations of asthma and, therefore, a proxy of clinical benefit.

Respiratory infections may also trigger acute vascular events in persons with coronary artery disease, and several studies have suggested a protective effect of influenza vaccination against “heart attacks”. The effect is more pronounced the more severe the underlying atherosclerosis. In some studies, the statin medicines used to improve lipid profiles and decrease coronary event risk have been associated with poorer VE, but data are ambiguous and more research on the topic is needed.

Influenza vaccine immunogenicity, efficacy, and effectiveness in patients with renal and hepatic conditions as well as autoimmune conditions such as rheumatic arthritis and inflammatory bowel disease have generally been comparable to that of healthy controls. However, immune responses to vaccination in patients with the latter conditions can be reduced if they are receiving concomitant immunomodulatory therapies. Malignancies are a very heterogeneous group of diseases where the response to the influenza vaccine is greatly dependent on the level of immunocompromise. In conditions such as HIV infection, vaccine immunogenicity correlates with CD4 + T-lymphocyte cell counts.

Pregnancy

In their 2012 position paper, the WHO recommended that countries considering the initiation or expansion of seasonal influenza vaccination programs would give the highest priority to pregnant women. The basis of the recommendation is two-fold: First, pregnancy is an immunosuppressive condition and severe influenza infections can occur during this time, and second, newborns are a risk group for severe influenza but are not eligible to receive influenza vaccine; passive transfer of influenza antibodies from vaccinated women as well as the “cocooning” (i.e., when the child is protected by not having their mother fall ill with influenza) provide the newborns with protection. Both RCTs and observational studies have demonstrated significant protection of mothers and their newborn children against LCI, with vaccine efficacy or effectiveness of 40%–60% VE and above in the infants when the mother is vaccinated.

Healthy Children

During influenza epidemics, higher attack rates have often been observed in children than in adults. It has been postulated that children might be “super spreaders” who drive influenza epidemics in a population. On the other hand, while many children’s influenza infections are mild and self-limiting, influenza can be severe and even life-threatening in small children. For these reasons, some countries such as the United Kingdom and Finland have adopted universal influenza vaccination programs for healthy children (e.g., all children from six months to six or ten years). In the position paper referenced above, the WHO included children aged 6–59 months in additional risk groups to be considered for vaccination.

Besides traditional inactivated vaccines, LAIV is relevant in this target group. Where multiple vaccine types are available, vaccination recommendations are sometimes preferential to one particular type of vaccine and sometimes non-preferential (i.e., parents may choose).

When a child is vaccinated for the first time with a seasonal influenza vaccine, two doses are recommended to ensure sufficient protection. This is particularly true of inactivated vaccines, where vaccine effectiveness has been shown to be lower among children who received only one dose in their first year of vaccination compared to children who received two doses. With the Ann Arbor-backbone LAIV, one dose has been shown to provide roughly 90% of the protection observed after two doses.

Healthy Adults

Few countries prioritize non-pregnant healthy adults as a target group for influenza vaccination. However, this is not to say the vaccine does not work (indeed the vaccine effectiveness is typically higher in adults aged <65 years than in the elderly). Healthy adults can be vaccinated especially if they want to protect themselves from influenza (e.g., because they have a close contact whose age or health condition requires protection from influenza, because of a holiday trip, or if falling ill with influenza would be particularly bad for their work or social situation).

Among adults, healthcare workers are an important target group for vaccination, again prioritized by the WHO. The rationale for vaccinating medical personnel is two-fold: Through clinical work, they become exposed to influenza more than the general population, and may also transmit influenza to their patients, some of whom may be frail and at risk of serious infection.

Pandemic Influenza Vaccines

Influenza viruses have a considerable pandemic potential because of their capacity for antigenic shift, adaptation to infect multiple species, and their mode of transmission. The occurrence of pandemics is inherently unpredictable but in recent history, influenza pandemics have been observed every couple of decades: 1918, 1957, 1968, and 2009.

Much of what is said above about seasonal influenza vaccines applies to pandemic influenza vaccines. A key point to understand regarding vaccine production is that especially in the age of air travel, pandemics may spread very rapidly, and without proper forward planning and production capacity, vaccines may not be available until it is too late (though with several waves of pandemic influenza activity spread over some time, the vaccine may be available for the second wave if not the first one). Ways to enhance preparedness between pandemics include the preparation of collections of candidate vaccine viruses such as H5, H7, and H9 viruses. Regulators such as the FDA and the European medicines agency (EMA) have developed procedures for fast-track licensure of pandemic vaccines.

Among other hurdles, pandemic vaccine producers face increased safety concerns. A high-pathogenicity donor virus would have to be grown under high-containment conditions, so manufacturers have turned to other methods, including recombinant techniques and baculovirus vector-expressed vaccine antigens in cell cultures. The use of aluminum or lipid emulsion adjuvants has often been deemed necessary to achieve sufficient immunogenicity. Another appealing aspect of adjuvants is dose sparing (i.e., extending the supply and availability of vaccine doses through less requirement of viral protein per dose).

At the time of writing, the most recent pandemic vaccine preparation effort was in the 2009 A(H1N1) pandemic. Here, a vaccine was developed, produced, and administered over only 8 months – though large-scale production was not attained as quickly as initially hoped. The pandemic itself turned out to be relatively mild, though in some cases (especially in clinical risk groups) resulted in severe viral pneumonitis, bacterial sequelae, and death. Compared to seasonal epidemics, there were relatively more severe infections in children and young adults than in the elderly.

The adjuvanted vaccines during the 2009 pandemic attained very good effectiveness in many studies. In general, the pandemic vaccines of 2009 had a safety profile comparable to seasonal vaccines and unusual safety issues were not observed in most countries. However, others experienced an increase in narcolepsy, a rare condition characterized by impaired ability to regulate sleep-wake cycles leading to daytime sleepiness and involuntary sleep episodes. Cases of narcolepsy appeared to be associated with the AS03-adjuvanted pandemic H1N1 vaccine in Finland, Sweden, and Iceland. For reasons currently unknown, this was not observed in other countries using that vaccine. Notably, narcolepsy has never been linked to seasonal influenza vaccines.

Overall, the experience with the 2009 pandemic demonstrated that adjuvanted pandemic influenza vaccines are immunogenic, effective, and have an acceptably low rate of adverse events. The pandemic response may have contributed to improved preparedness for influenza pandemics in general and reduced barriers to large-scale vaccine manufacture and vaccination program implementation.

Further Reading

Fiore, A.E., Bridges, C.B., Katz, J.M., Cox, N.J., 2018. Inactivated Influenza Vaccines, in Plotkin's Vaccines, 7th ed. Philadelphia, PA: Elsevier.

World Health Organization, 2012. Position Paper on Influenza. Weekly epidemiological record 23 November 2012, 87th year. No. 47, vol. 87, pp. 461–476. Geneva: WHO.

World Health Organization, 2017. Evaluation of Influenza Vaccine Effectiveness: A Guide to the Design and Interpretation of Observational Studies. Geneva: WHO.

Relevant Websites

<https://www.cdc.gov/flu/professionals/acip/immunogenicity.htm>
Immunogenicity, Efficacy, and Effectiveness of Influenza Vaccines.

<https://vk.ovg.ox.ac.uk/vk/vaccine-ingredients>
Vaccine ingredients at the Vaccine Knowledge Project.

Polio Eradication

M Steven Oberste, Cara C Burns, and Jennifer L Konopka-Anstadt, Centers for Disease Control and Prevention, Atlanta, GA, United States

© 2021 Elsevier Ltd. All rights reserved.

Background

An ancient disease first documented in the 2nd millennium BCE and feared worldwide through much of the modern 20th century, polio is nearing eradication. Derived from the Greek words *polio* (gray) and *myelon* (marrow), poliomyelitis describes the predilection of poliovirus (genus *Enterovirus*, family *Picornaviridae*) to target motor neurons within the gray matter in the anterior horn of the spinal cord, leading to the classic manifestation of poliovirus-induced acute paralysis. Poliovirus is highly infectious, with replication primarily occurring in the intestinal tract of infected individuals. Humans are the only reservoir for poliovirus, with transmission occurring from person-to-person mainly through the fecal-oral route. Over 70% of poliovirus infections in children are asymptomatic, though even infected individuals without symptoms shed virus in stool and can transmit the virus to others. Approximately 24% of infected children develop minor, nonspecific illness without evidence of central nervous system invasion. Aseptic meningitis occurs in 1%–5% of poliovirus-infected children, typically followed by complete recovery. Less than 1% of all poliovirus infections result in acute flaccid paralysis (AFP) and these are classified into three types. Spinal poliomyelitis is most common, characterized by asymmetric paralysis most often involving the legs. Bulbar polio describes the weakness of muscles innervated by cranial nerves. Bulbospinal polio is a combination of bulbar and spinal paralysis.

In 1988, wild poliovirus (WPV) was present on all continents except Australia, and estimates of case burden exceeded 350,000 annually. That same year, the Global Polio Eradication Initiative (GPEI) was launched following the early successes of polio eradication efforts in the Americas (initiated by the Pan American Health Organization in 1985) where the last case occurred in 1991. GPEI is a public-private partnership driven by the objective to eradicate polio worldwide, led by national governments and other key partners including the World Health Organization (WHO), US Centers for Disease Control and Prevention, UNICEF, and Rotary International. These initial spearheading partners were later joined in 2007 by the Bill and Melinda Gates Foundation and by GAVI in 2019. By 2011, the number of countries with WPV circulation was only 16, and by 2014 the regions of the Americas, Europe, Southeast Asia and the Western Pacific were certified to be free of WPV. The only countries with endemic WPV cases in 2019 were in South Asia (Pakistan and Afghanistan). The African Region reported its last case of WPV in 2016, and Africa was declared free of WPV in 2020. It is estimated that the efforts of the GPEI in combination with the use of two highly effective poliovirus vaccines have prevented over 16 million cases of paralytic poliomyelitis over the past 30 years.

Polioviruses comprise three antigenically distinct serotypes, types 1, 2, and 3. Infection with one serotype confers protection only against viruses of that same serotype. Type 2 wild poliovirus (WPV2) was last detected in 1999 and was declared eradicated in 2015. Type 3 wild poliovirus (WPV3) was last detected in 2012 and it was declared eradicated in 2019. Since 2012, the only WPVs detected have been of type 1 (WPV1).

Initial eradication efforts focused on WPVs (i.e., those that circulated naturally in human populations). While GPEI focused on the eradication of WPV, it later became clear that the live, attenuated Sabin oral vaccine strains (see Section “Vaccines”) could very rarely revert and reacquire neurovirulence and transmissibility. For all three serotypes, key sites of reversion include attenuating mutations in the 5′ untranslated region as well as additional sites of reversion within the capsid. Reversion can happen quickly during replication within the human gut. These reverted vaccine viruses, termed vaccine-derived poliovirus (VDPV) can emerge in areas with very low vaccine coverage, causing paralytic disease indistinguishable from WPV. A VDPV was originally defined operationally as a poliovirus that was derived from the oral vaccine strains and that diverged >1% from the parental strain in the nucleotide sequence encoding the VP1 capsid protein (i.e., ≥10 changes in the ~900 nucleotide VP1 region). The threshold of 1% is somewhat arbitrary but is consistent with prolonged replication of the virus. Based on genetic information from several large VDPV2 outbreaks, the threshold for the definition of VDPV2 was modified to a nucleotide divergence >0.6% from the parental OPV strain (≥6 changes in VP1). VDPV are classified as circulating VDPV (cVDPV) when there is evidence for person-to-person transmission; immunodeficiency-associated VDPV (iVDPV) if isolated from a person with primary immune deficiency (primarily deficiencies in humoral immunity, such as hypogammaglobulinemia, combined variable immunodeficiency, or severe combined immunodeficiency); or ambiguous VDPV when the virus cannot be conclusively classified as either cVDPV or iVDPV. To fully eradicate polio, all sources of the virus (wild and vaccine-derived) must be eliminated. Surveillance and vaccination continue to be the most important pillars in efforts toward ensuring poliovirus becomes the second human virus to be eradicated.

Vaccines

Two key tools, Sabin live oral poliovirus vaccine (OPV) and Salk inactivated poliovirus vaccine (IPV), have been central to the global eradication effort. Both IPV and OPV contain a representative strain for each of the three poliovirus serotypes (though the type 2 component was removed from OPV in 2016; see below), with vaccination resulting in the production of neutralizing

antibodies that protect recipients from the disease. Each vaccine has advantages and disadvantages, and appropriate circumstances exist for the use of each.

IPV consists of WPV strains that have been chemically inactivated. It provides excellent systemic immunity and is very safe since no infectious virus is present. First licensed in 1955, IPV reduced the number of polio cases by over 99% in the US. IPV has been shown to be capable of eradicating poliomyelitis in settings where the overall burden of viral exposure is low, and the rate of immunization coverage is high (> 90%). However, IPV does have disadvantages. It is more expensive to produce than the live vaccine and the use of large amounts of wild seed virus in production comes with potential hazards. In addition, while IPV protects vaccinees from disease, it does not induce the intestinal immunity required to prevent person-to-person transmission of poliovirus.

In developing countries where viral exposure is more intense, live-attenuated vaccines imitating natural infection often provide better protection. Accordingly, Sabin OPV was introduced in the early 1960s, containing live strains of poliovirus that have been attenuated via passage in animals and cell cultures. OPV provides strong systemic immunity, limiting poliovirus infection and disease in vaccinees. Unlike IPV, OPV also induces the mucosal immunity necessary to interrupt person-to-person transmission. Since recipients excrete vaccine virus into the environment, OPV expands community protection by secondary infection of contacts of vaccinees. OPV is inexpensive, easy to administer, and has been the cornerstone of the global eradication program. It has been demonstrated as an effective vaccine tool to interrupt polio epidemics and break transmission, resulting in the eradication of the virus from entire regions across the world.

While vaccination has played a central role in the dramatic decline of poliomyelitis cases, unexpected challenges have become apparent during the final articles of eradication, particularly those associated with the use of Sabin OPV. Much is now known regarding the key mutations underlying the attenuation of Sabin OPV strains and their inherent genetic instability. During replication in the human gut, the attenuated strains are under pressure to adapt and evolve. In rare instances, reversion of key attenuating “gatekeeper” mutations restores virulence to OPV strains, resulting in one case of vaccine-associated paralytic poliomyelitis (VAPP) for approximately every 2.4 million doses of OPV administered, with the highest rate after the first dose. A fundamental piece of the poliovirus eradication strategy is achieving and maintaining high levels of routine immunization. Within that context, VAPP cases are not associated with outbreaks of poliomyelitis. However, in areas of low-vaccine coverage and poor sanitation, where OPV strains have the opportunity to replicate unchecked and transmit for longer durations, the genetic instability of OPV strains can lead to transmission of circulating vaccine-derived polioviruses and spur outbreaks, the most prevalent being type 2 VDPVs. Accordingly, WHO coordinated a synchronized global withdrawal of the type 2 component of OPV in April 2016, replacing trivalent OPV with a bivalent formulation containing only the Sabin type 1 and 3 strains. Prior to the switch, WHO recommended that all countries introduce at least one dose of IPV into their routine immunization program to provide immunity to type 2, which has, in turn, strained the global supply of IPV. Recent clinical trials highlighting the efficacy of using fractional doses (one-fifth the typical amount) of IPV (fIPV) have offered a way forward that may help alleviate both the supply and cost issues for IPV used in this manner. Several countries have already begun to use fIPV for routine immunization. After global eradication of all three serotypes has been certified, all use of OPV in routine immunization is expected to cease. OPV cessation is considered an essential step to eliminate the risk of VAPP and remove the source of potential future VDPV emergence. At that point, continued immunization with IPV will be critical to maintaining eradication.

Poliovirus Surveillance

Essential to the eradication program is surveillance for cases of AFP, defined for the purpose of polio surveillance as sudden onset of paralysis or weakness in any part of the body in a child less than 15 years of age. There are multiple etiologies of AFP, including poliovirus, other infectious agents, toxins, autoimmunity, trauma, and congenital conditions. While all these etiologies are medically important and approaches to patient care must consider the cause (or suspected cause) of the illness, GPEI is focused specifically on paralytic cases associated with poliovirus infection.

Virologic surveillance (laboratory confirmation that poliovirus is associated with an AFP case) is essential to identifying WPV reservoirs, directing vaccination campaigns, assessing overall program progress, and global certification of eradication. Surveillance quality indicators include timely case detection, notification, and investigation, as well as adequate stool collection and timely transport to the laboratory. The Global Polio Laboratory Network (GPLN) was launched in 1989 to ensure laboratory support for every country's surveillance system. The GPLN currently consists of 146 laboratories in 92 countries; some large countries have multiple polio laboratories and some laboratories serve neighboring countries that do not have a polio laboratory of their own. Stool specimens collected from children with AFP are tested for the presence of poliovirus by virus isolation in cell cultures and any viruses detected are characterized by poliovirus-specific polymerase chain reaction assays and genomic sequencing. In a typical year, GPLN laboratories test approximately 200,000 stool specimens from AFP cases. Quality of laboratory results is assured through standardized protocols and reagents, annual proficiency testing, ongoing training, and a WHO-led accreditation program. Methods to detect and characterize poliovirus are constantly evolving.

Environmental surveillance (detection of poliovirus in sewage-affected waters) is an important adjunct to AFP surveillance, permitting virologic surveillance of a relatively large population (usually tens of thousands to a few million). While environmental surveillance has played a key role in the detection of poliovirus transmission, especially in the latter stages of eradication, it cannot identify the individuals who are shedding the virus. Its main advantage is its ability to detect virus from asymptomatic virus excretors because those individuals would not be identified through AFP surveillance. Environmental surveillance expanded dramatically from

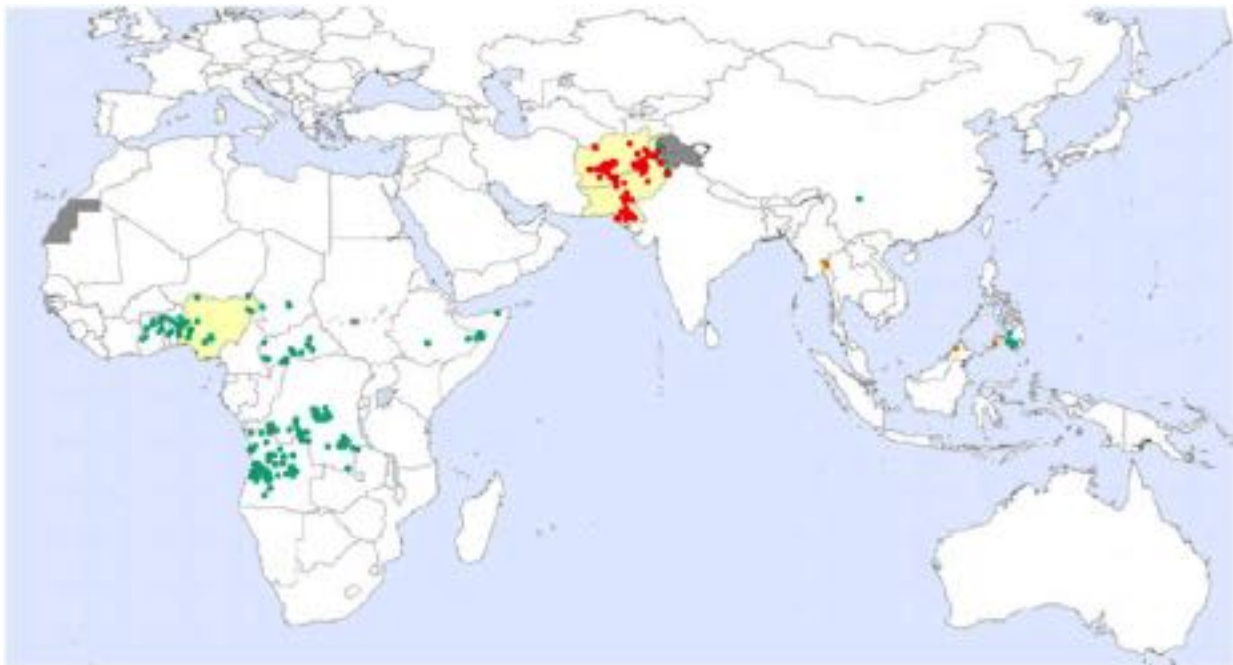


Fig. 1 Global wild and circulating vaccine-derived poliovirus cases with onset of paralysis from 1 January to 31 December 2019. Red, WPV1 ($n = 143$, in 2 countries); orange, cVDPV1 ($n = 8$, in 3 countries); green, cVDPV2 ($n = 241$, in 15 countries). Not shown: Viruses detected from environmental surveillance. Data from the World Health Organization, as of 31 December 2019.

2013 to 2019. Environmental surveillance is currently conducted in hundreds of sites in more than 50 countries, and expansion to additional countries is planned for 2020–2022. Continued environmental surveillance, especially in high-risk areas prone to VDPV emergence, will be needed to assure final eradication.

Status of Eradication

The number of WPV1 cases globally had declined to less than 50 per year in 2016–2018 but the 2019 global total was more than 150 cases, in Pakistan and Afghanistan (Fig. 1). As noted, no WPV has been detected anywhere in Africa since 2016. WPV2 and WPV3 were declared eradicated in 2015 and 2019, respectively. Polio eradication has been difficult to accomplish in countries with environments conducive to poliovirus replication and spread (a high biological infectivity index) and weak routine immunization programs. In recent years, however, the primary barriers have not been biological or environmental. Rather, the difficulties encountered by GPEI have largely been man-made. Some of these include the inability to immunize children in areas under the control of rebels or other anti-government elements, the spread of misinformation regarding vaccine safety, violence against vaccinators and their accompanying security teams, corruption and dishonest practices, as well as simple poor-quality of immunization campaigns. In 2018–2019, immunization was again prohibited by the Taliban in parts of Afghanistan after several years of cooperation. The targeting of polio immunization staff by violence, starting in 2011 in Pakistan, impeded eradication progress, forcing the GPEI to change strategies for advertising immunization campaigns and delivering vaccines. Despite these challenges in the immunization side of the program, AFP surveillance quality in Pakistan and Afghanistan has generally been high since 2015, although there are pockets of lower quality surveillance in areas of insecurity and civil unrest.

In 2018 and 2019, there were significantly more paralytic polio cases associated with VDPV than with WPV (Fig. 1). There were 104 cVDPV cases in 2018 and ~250 cVDPV cases in 2019, compared to 33 and ~150 WPV cases in the two years, respectively. The number of cases associated with VDPV still pales in comparison to the >16 million cases of paralytic polio that have been prevented by GPEI over the past 3 decades. After the switch from Trivalent oral polio vaccine (tOPV) to Bivalent oral polio vaccine (bOPV), cVDPV2 from an ongoing outbreak was detected in Nigeria. Other cVDPV2 outbreaks originating prior to the switch were detected in 2017 (Horn of Africa) and in 2019 (Philippines). Since late 2016, Monovalent type 2 oral polio vaccine (mOPV2) from a global vaccine stockpile has been used on an emergency basis in response to cVDPV2 outbreaks. The WHO Director-General must approve use of mOPV2 in order to limit the exposure to the population of a live OPV strain that would have the potential risk of creating new circulating VDPV2. New emergences were apparently seeded by the use of mOPV2 in outbreak responses, and the cycle continued to increase the number of cVDPV2 outbreaks in African countries, including Nigeria and the Democratic Republic of Congo, which each experienced

several concurrent outbreaks. After the switch, the number of emergences of cVDPV2 increased each year. Although some cVDPV2 outbreaks were stopped in countries such as the Democratic Republic of Congo, other outbreaks continued. In 2019, cVDPV2 was detected in 12 countries in Africa. Nigeria, the Democratic Republic of the Congo, Somalia, and the Philippines exported cVDPV2 to surrounding countries. Three years after the switch, over 250 million doses of mOPV2 have been used in cVDPV2 outbreak responses. In addition to type 2 cVDPV outbreaks, separate cVDPV1 outbreaks were detected in Papua New Guinea and Indonesia in 2018, and in Myanmar and the Philippines in 2019. Late in 2019, cVDPV1 genetically linked to the cVDPV1 outbreak in the Philippines was detected in Malaysia.

New Vaccines

When the poliovirus eradication program began, it was assumed that OPV and IPV would be sufficient for the successful achievement of the goal. Indeed, the use of OPV has reduced poliomyelitis cases by more than 99% worldwide. However, with the growing knowledge of the current risks associated with OPV, and the inability of IPV use alone to fully achieve global eradication, the availability of additional novel vaccine options has become highly desirable. In recent years, efforts to develop novel vaccine platforms have included noninfectious DNA-based vaccines, virus-like particles, and immunogenic peptides, among others. Additionally, to more safely produce IPV, some manufacturers have replaced the traditional wild seed strains used in the production of IPV with attenuated Sabin strains.

More recently, the knowledge gained about poliovirus biology over the last half-century has led to the development of novel OPV (nOPV) strains engineered to be safer and more genetically stable than traditional Sabin strains. These rationally redesigned nOPV strains stabilize known reversion hotspots within the Sabin virus genome, while also offering new attenuating strategies targeting replicative fitness, polymerase fidelity, and recombination. The development of these nOPV strains has been deemed necessary by global eradication partners, with a particular interest in using monovalent nOPV2 in outbreak settings, to reduce the risk of spurring additional VDPV emergences. Clinical testing of these nOPV strains is currently underway, with initial results demonstrating nOPV2 is viable, immunogenic, and safe in human recipients. In November 2020, WHO granted provisional emergency use listing to nOPV2, permitting its initial use in outbreak responses.

Antivirals

In countries where prophylactic immunoglobulin therapy is available, patients with a primary immune deficiency are generally protected from paralytic disease even if they become infected with a poliovirus vaccine strain. However, an iVDPV can replicate for years in some immunodeficient persons and the virus can revert to a virulent phenotype (iVDPV) and continue to evolve, accumulating additional nucleotide substitutions throughout the genome. One such patient has excreted virulent poliovirus for over 30 years. Continued replication of virulent iVDPV puts the patient at risk for paralysis, even in the face of immunoglobulin therapy, and excretion of iVDPV into the environment poses a risk of transmission in the community, especially in areas of low vaccine coverage or in the post-eradication era when OPV will no longer be used. To minimize the risks to both the patient and the community, antiviral drugs are being developed to treat chronically infected primary immune deficiency patients. One drug, pocapavir, has completed a Phase II trial with positive results, and a second drug is being developed to provide a safeguard against the development of drug resistance, in the form of a combination product.

Containment

Once complete eradication of all three serotypes is achieved and vaccination campaigns are no longer implemented, population immunity to polioviruses will decrease. At that point, live poliovirus will exist only in settings such as laboratories and vaccine manufacturing facilities. Therefore, the consequences of any poliovirus introduction into communities from a facility containment breach, even if unlikely, would be severe. This is the case for WPV2 now, in 2020, apart from ongoing cVDPV2 circulation. As a result, WHO has developed a “Global Action Plan to minimize poliovirus facility-associated risk after type-specific eradication of WPVs and sequential cessation of oral polio vaccine use” (GAPIII) to provide guidance for facilities that need to retain live polioviruses for essential functions such as vaccine production and critical research. Currently, all type 2 polioviruses must be contained in facilities that comply with GAPIII, with slightly different requirements for vaccine virus compared to WPV2 and VDPV2. Now that WPV3 has been declared eradicated, it also falls under GAPIII containment requirements. At present (in 2020), Sabin 3 is still included in the oral polio vaccine, so the vaccine strain is not yet subject to containment. In recent years, there have been several known breaches of containment in vaccine manufacturing facilities, underlining the risks associated with retaining live poliovirus materials post-eradication. Fortunately, none of the incidents resulted in community transmission. To limit the risk involved in the use of WPV seed strains for IPV production, multiple manufacturers are developing IPV products based on live-attenuated vaccine strains. As of the end of 2019, Japan was the only country with a licensed Sabin IPV product.

Transition of Global Polio Eradication Initiative Assets Post-Eradication

Once global eradication has been achieved, certain GPEI activities and functions will still be needed to ensure the world remains polio-free for the long term. Four key activities that must continue are containment, disease surveillance, outbreak response capacity, and vaccination (with IPV only once OPV use has ceased). Vaccination with IPV is expected to continue for the foreseeable future, as a barrier to disease caused by VDPVs (e.g., from an immunodeficient chronic excreter) or transmission of WPV that could result from exposure in a facility that retains live virus, such as a vaccine manufacturing facility. Even with heightened awareness and effort, it is expected that maintaining high vaccine coverage will be challenging, especially in resource-limited settings where GPEI has already faced difficulties in achieving levels of population immunity sufficient to interrupt transmission. Poliovirus containment has been described in some detail above; it will be increasingly important as the years pass after eradication given that reintroduction of WPV into a population with reduced immunity would be catastrophic. The capacity to detect polio cases or outbreaks should they occur will need to be integrated into the existing community- and event-based surveillance systems to trigger appropriate investigations and outbreak responses. Vaccine stockpiles will need to be generated and maintained to supply outbreak response vaccination campaigns.

While significant capacity in all these areas has been built over more than three decades and the program's assets have been used to support many other high-priority public health programs, GPEI as a stand-alone program is not sustainable post-eradication. This polio infrastructure – notably the trained global public health workforce that includes field surveillance officers, laboratory scientists, communications and logistics experts, and others, at local to global levels – will be invaluable to other public health programs. Integrating polio assets into other health programs will help maintain the capacity to detect and respond to a polio outbreak while also enhancing the capacity to investigate and respond to other priority health threats.

Further Reading

- Bandyopadhyay, A.S., Garon, J., Seib, K., Orenstein, W.A., 2015. Polio vaccination: Past, present and future. *Future Microbiology* 10, 791–808.
- Greene, S.A., Ahmed, J., Datta, S.D., *et al.*, 2019. Progress toward polio eradication – Worldwide, January 2017–March 2019. *MMWR Morbidity and Mortality Weekly Report* 68, 458–462.
- Heymann, D.L., Sutter, R.W., Aylward, R.B., 2005. A global call for new polio vaccines. *Nature* 434, 699–700.
- Hampton, L.M., Farrell, M., Ramirez-Gonzalez, A., *et al.*, 2016. Cessation of trivalent oral poliovirus vaccine and introduction of inactivated poliovirus vaccine – worldwide, 2016. *MMWR Morbidity and Mortality Weekly Report* 65, 934–938.
- Jorba, J., Diop, O.M., Iber, J., *et al.*, 2019. Update on Vaccine-Derived Poliovirus Outbreaks – Worldwide, January 2018–June 2019. *MMWR Morbidity and Mortality Weekly Report* 68, 1024–1028.
- Kew, O., Pallansch, M., 2018. Breaking the last chains of poliovirus transmission: Progress and challenges in global polio eradication. *Annual Review of Virology* 5, 427–451.
- McKinlay, M.A., Collett, M.S., Hincks, J.R., *et al.*, 2014. Progress in the development of poliovirus antiviral agents and their essential role in reducing risks that threaten eradication. *Journal of Infectious Diseases* 210 (Suppl. 1), S447–S453.
- Van Damme, P., De Coster, I., Bandyopadhyay, A.S., *et al.*, 2019. The safety and immunogenicity of two novel live attenuated monovalent (serotype 2) oral poliovirus vaccines in healthy adults: A double-blind, single-centre phase 1 study. *The Lancet* 394, 148–158.
- World Health Organization, 2019. Polio Endgame Strategy 2019–2023: Eradication, Integration, Certification and Containment (WHO/Polio/19.04). Geneva: WHO.

Relevant Websites

- http://polioeradication.org/wp-content/uploads/2016/12/GAPIII_2014.pdf
GAPIII.
- <http://polioeradication.org/wp-content/uploads/2019/06/english-polio-endgame-strategy.pdf>
Polio Endgame Strategy.
- <http://polioeradication.org/wp-content/uploads/2018/04/polio-post-certification-strategy-20180424-2.pdf>
Polio Post-Certification Strategy.
- <http://www.polioeradication.org>
The Global Polio Eradication Initiative.

Subject Index

Note

This index is in letter-by-letter order, whereby hyphens and spaces within index headings are ignored in the alphabetization, and it is arranged in set-out style, with a maximum of three levels of heading.

Cross-reference terms in italics are general cross-references, or refer to subentry terms within the main entry (the main entry is not repeated to save space).

Location references refer to the volume number, in bold, followed by the page number.

Page numbers followed by “f” and “t” refer to figures and tables, respectively.

A

- AAEV *see* Anacridium aegyptium EV (AAEV)
AAP *see* antiapoptotic proteins (AAP)
AAV *see* adeno-associated virus (AAV)
AAV-2 *see* adeno-associated parvoviruses (AAV-2)
AAvV-1 *see* Avian Avulavirus 1 (AAvV-1)
ABLV *see* Australian bat lyssavirus (ABLV)
ABPV *see* acute bee paralysis virus (ABPV)
ABV *see* acidianus bottle-shaped virus (ABV)
ABVV *see* American bat vesiculovirus (ABVV)
Acanthamoeba polyphaga 1:372–373
acanthoma, defined 2:629
accessory gene, defined 4:98
accessory proteins 3:575–576
accessory RNAs associated with mitovirus infections 4:602
accuracy 5:252
ACD *see* Amasya cherry disease (ACD)
ACE2 *see* angiotensin-converting enzyme 2 (ACE2)
ACEV *see* Anomala cuprea EPV (ACEV)
acidianus bottle-shaped virus (ABV) 4:361
Acidianus filamentous virus 1 (AFV1) 1:368, 4:363, 1:435–436
Acidianus tailed spindle virus (ATSV) 1:370f
Acidianus two-tailed virus (ATV) 4:361, 4:422
acidic hot springs, viruses of 4:422
acidophilic, defined 4:359
AcMNPV *see* *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV)
AConSLA *see* *Ageratum conyzoides* symptomless alphasatellite (AConSLA)
acquired immune deficiency syndrome (AIDS) 5:190, 2:56, 1:567–568
actin filaments (microfilaments), defined 3:32
active surveillance 5:250
acute bee paralysis virus (ABPV) 4:771, 4:772, 4:773
acute flaccid paralysis (AFP) 5:310
 defined 2:245
acute gastroenteritis, defined 5:289
acute lower respiratory tract infections (ALRI) 2:747–748
acute otitis media (AOM) 2:761
 defined 2:757
acute respiratory distress syndrome (ARDS) 2:559
 defined 2:551, 2:825
ACV, *see* *Aeropyrum* coil-shaped virus (ACV)
acyclovir 5:183, 5:179
adamantanes/M2 blockers 5:161–162
adaptation vs. evolution, defined 1:71
ADCC *see* antibody-dependent cellular cytotoxicity (ADCC)
ADCP *see* antibody-dependent cellular phagocytosis (ADCP)
additional strand, conserved E (ASCE) 4:160
additional strand conserved glutamate (ASCE) 4:148–149, 4:149
 defined 4:148
adefovir dipivoxil (ADV) 5:218
Adelaide River virus (ARV) 2:875
adeno-associated parvoviruses (AAV-2) 1:268–269
adeno-associated virus (AAV) 1:411–412, 1:211, 1:533, 1:536
 AAV2 4:845
adenosine triphosphate (ATP) 1:495, 1:488
Adenoviridae
 adenoviruses as vectors 2:15
 classification 2:3–6
 clinical features 2:10–11
 HAdV-A 2:11
 HAdV-B1 2:11
 HAdV-B2 2:11
 HAdV-C 2:11
 HAdV-D 2:11
 HAdV-E 2:11
 HAdV-F 2:11–12
 HAdV-G 2:12
 diagnosis 2:14–15
 epidemiology 2:9–10
 epizootiology 2:12
 avian adenoviruses 2:13
 bovine, ovine, caprine, and other ruminant adenoviruses 2:13
 canine and other carnivoran adenoviruses 2:12
 equine adenoviruses 2:12–13
 frog and fish adenoviruses 2:14
 porcine adenoviruses 2:12
 reptilian adenoviruses 2:13–14
 simian adenoviruses 2:12
 genome 2:8
 life cycle 2:8–9
 pathogenesis 2:14
 prevention 2:15
 taxonomy of 2:4t–6
 treatment 2:15
 virion structure 2:6–8
adenovirus 26 (Ad26) 2:243
adenovirus death protein (ADP) 4:761
adenoviruses (ADVs) 1:401, 2:3, 5:109, 5:105, 5:239, 1:660
 external minor coat protein in 1:340f
 general features of 1:329
 human adenovirus type 5, structure of 1:329–331
 core proteins 1:335
 fiber 1:331–333
 hexon 1:331
 penton base protein 1:331
 protein IIIa 1:333
 protein VI 1:333
 protein VIII 1:333
 protein IX 1:333–335
 structural relatives 1:337
 asymmetric features 1:340
 capsid size determination in DJR lineage 1:340–342
 double jelly roll fold, variants of 1:337–339
 double jelly roll lineage 1:337
 double jelly roll lineage, genomes in 1:340

- adenoviruses (ADVs) (*continued*)
 double jelly roll lineage, vertex structures in 1:339
 membranes and membrane proteins 1:339–340
 polintons and evolutionary pathway of DJR viruses 1:342–343
 variants 1:335–337
 external minor coat proteins 1:337
 fiber shafts and heads 1:335–337
 as vectors 2:15
- adenovirus fiber 1:331–333
- adenovirus infections, management of
 clinical description of infection 5:199
 animal adenoviruses 5:202
 diseases and human adenoviruses infections 5:200–201
 gastroenteritis 5:200
 genitourinary diseases 5:200
 immunocompromised patients, adenovirus infections of 5:201–202
 incubation period 5:199
 lower respiratory tract infections 5:200
 ocular infections 5:200
 upper respiratory tract infections 5:199–200
- diagnosis 5:202–203
 adenovirus differentiation and typing 5:203–204
 diagnostic methods 5:203
 polymerase chain reaction (PCR) for human adenoviruses DNA detection 5:202–203
 quantitative human adenoviruses PCRs and screening of high-risk patients 5:203
- epidemiology 5:199
 structure 5:198–199
 taxonomy 5:197
 family and genus 5:197
 types and subtypes 5:197–198
 treatment and prevention 5:204–205
 antivirals 5:204–205
 germicides 5:205
 vaccines 5:205
- adenovirus vaccine 2:216
 adjuvant, defined 5:281
- ADP *see* adenovirus death protein (ADP)
- adsorption, defined 1:402, 4:53
 adsorption rate, defined 4:387
- adult T-cell leukemia (ATL) 2:532–533
- ADV *see* adenofovir dipivoxil (ADV); adenoviruses (ADVs)
- advanced light microscopy approaches 1:210–211
 multiphoton imaging 1:211–213
 quantitative spectral imaging 1:211
 real-time single virus tracking (SVT) in live cells 1:211
- adverse effects following immunization (AEFI) 5:283
- Aedes aegypti* 1:575, 2:894
- Aedes pseudoscutellaris* dinovernavirus 1 (ApDENV-1) 4:870
- AEFI *see* adverse effects following immunization (AEFI)
- AEP *see* archaeo-eukaryotic primases (AEP)
- Aeropyrum* coil-shaped virus (ACV) 1:436, 4:362
- Aeropyrum pernix* Bacilliform Virus 1 (APBV1) 1:369, 4:361
- aeropyrum pernix ovoid virus 1 (APOV1) 4:362
- AF4 *see* asymmetrical flow field-flow fractionation (AF4)
- affimers, defined 3:692, 3:364
- affine extension, defined 1:248
- affinity, defined 4:136
- AFM *see* atomic force microscopy (AFM)
- A-form, defined 4:387, 4:359
- AFP *see* acute flaccid paralysis (AFP)
- Africa, CLCuD in 3:356, 3:357
- African cassava mosaic disease 3:24–26
 classification of cassava mosaic begomoviruses 3:27
 control 3:28–29
 disease symptoms and yield losses 3:26–27
 epidemiology 3:27
 geographical distribution 3:24–26
 recent outbreak of CMD in South East Asia 3:27–28
- African horse sickness virus (AHSV) 1:545
 classification 2:17
 clinical features 2:18
 diagnosis 2:19–20
 epidemiology 2:17–18
 life cycle 2:17
 pathology 2:18–19
 treatment, control, and prevention 2:20–21
 virion structure and genome 2:17
- African swine fever virus (ASFV) 1:159–160, 2:22–23
 ASFV genes and their functions 2:26*t*–27
 classification 2:22–23
 clinical features 2:31
 diagnosis and control 2:32–33
 epidemiology 2:30–31
 genome of 2:24*f*
 genome organisation 2:24–25
 DNA repair 2:27–28
 DNA replication 2:25
 helicases 2:25
 inhibitors of apoptosis 2:28
 inhibitors of type I interferon and inflammatory responses 2:28
 multigene families 2:25
 nucleotide metabolism 2:25–27
 other enzymes 2:28
 polymerases 2:25
 proteins that inhibit host defences 2:28
 virion morphogenesis and structural proteins 2:28
 immune responses 2:32
 life cycle 2:28–30
 assembly and morphogenesis 2:30
 DNA replication 2:30
 entry 2:29–30
 transcription 2:30
 pathogenesis 2:31–32
 replication cycle of 2:29*f*
 structure 2:23*f*
 virion structure 2:23–24
- AFV1, *see* *Acidianus* filamentous virus 1 (AFV1)
- AfV-E, *see* *Aspergillus foetidus* dsRNA mycovirus (AFV-F)
- AfV-S2, *see* *Aspergillus foetidus* slow virus S2 (AFV-S2)
- Agallia constricta* 4:772
- agar gel immunodiffusion (AGID) 2:134
- Agaricus bisporus* 4:431
- Agaricus bisporus* virus 1 (AbV1) 4:528–529, 4:437
 biological properties (virus host relationships) 4:529–530
 control 4:530–532
 diagnostics and identification 4:530
 epidemiology 4:530
 genome organization 4:529
 taxonomy and classification 4:529
 virion structure and composition 4:529
- Agaricus bisporus* virus 16 (AbV16) 4:532
 classification 4:532
 diagnosis 4:533
 epidemiology 4:533
 genome and virion structure 4:532
 viral expression and disease development 4:532–533
- agarose slides, preparation of 5:11
- Ageratum conyzoides* symptomless alphasatellite (AConSLA) 3:360–361
- Ageratum enation virus* 3:750
- ageratum yellow vein betasatellite (AYVB) 3:361
- Ageratum yellow vein India alphasatellite* (AYVINA) 3:360–361
- Ageratum yellow vein virus* (AYVV) 3:357
- aggregates 3:777
- aggresome, defined 1:495
- AGID *see* agar gel immunodiffusion (AGID)
- AGO1 protein *see* Argonaute 1 (AGO1) protein
- AGO proteins *see* Argonaute (AGO) proteins
- AGP *see* antigenome promoter (AGP)
- agricultural research centers 1:646
- agrobacterium-mediated genetic transformation, defined 3:123
- agroinfection, defined 3:768, 3:461
- agroinfiltration, defined 3:743, 3:123
- agro-inoculation 3:749
 21238:p0015 3:667, 3:768, 3:313
- agro-terrorism, defined 1:644
- agro-warfare 1:647–648
- AHF *see* argentinian hemorrhagic fever (AHF)
- AHS *see* arthrogryposis-hydranencephaly syndrome (AHS)
- AHSV *see* African horse sickness virus (AHSV)
- AHV *see* avian herpesviruses (AHV)
- AIDS *see* acquired immune deficiency syndrome (AIDS)
- airborne transmission 1:561
 defined 1:559
- AIV *see* avian influenza viruses (AIV)

- akabane virus 1:545, 2:34
 classification 2:34
 clinical features 2:36
 diagnosis 2:38
 epidemiology 2:35–36
 genome 2:34–35
 pathogenesis 2:36–38
 prevention 2:38–39
 viral replication 2:35
 virion structure 2:34
- akt, defined 2:875
- alanine aminotransferase (ALT) 2:896
- Albetovirus* 3:581, 3:581–582
- Alfalfa mosaic virus (AMV) 3:261, 3:265
 history 3:132
 taxonomy and classification 3:132–133
 genome structure 3:133–134
 host range and economic significance 3:137–139
 interaction between viral RNA and coat protein 3:134
 particle structure and composition 3:132–133
 replication of viral RNA 3:135–136
 role of coat protein in the AMV replication cycle 3:136–137
 translation of viral RNA 3:134–135
 virus encapsidation and movement 3:136
 virus-host relationships 3:137
 transmission 3:139
 epidemiology and control 3:139
- Alfamovirus* 3:260, 3:262*t*
 characteristics of RNA genome in 3:261*t*
- alfavirus, defined 3:229
- algal double-stranded DNA viruses 4:685–686
 genus *Dinodnavirus* 4:686
- algal double-stranded RNA viruses 4:684–685
 genus *Mimoreovirus* 4:684–685
- algal single-stranded DNA viruses 4:685
 genus *Bacilladnavirus* 4:685
- algal single-stranded RNA viruses 4:684
 genus *Dinornavirus* 4:684
- algal viruses
 belonging to a subgroup within *Mimiviridae* family
 general properties 4:678–681
 history 4:677–678
 proposed subfamily *Mesomimivirinae* 4:681
 unclassified algae-infecting members in the family *Mimiviridae* 4:681–682
 defined 4:677
- alkaline midgut, defined 4:858
- alkaliphilic, defined 4:368
- AlkB protein, defined 3:642
- allele, defined 3:69, 3:554
- Allium virus X (ALVX) 3:628
- alogeneic, defined 2:778, 2:441
- allograft, defined 2:441
- Alloherpesviridae* 2:308*f*
- alloherpesvirus, defined 2:306
- allosteric inhibitors, defined 5:121
- Almendraviruses* 4:715–716
 almendraviruses genome 4:885
- alphachryso-P3 4:564
- alphaentomopoxvirus 4:859
- Alphaflexiviridae* 3:539, 3:140
- alphaflexiviruses (*Alphaflexiviridae*) 3:140
 diagnosis 3:148
 epidemiology and control 3:146
 replication and propagation 3:146
 satellite rnas associated with infections of 3:148
 taxonomy, phylogeny of family members, and evolution 3:140–141
 transmission, host range 3:146
 virion structure 3:141–145
 genome organization 3:142–145
 properties and functions of gene products 3:145–146
 virus–host relationships 3:146–148
- alpha helix (-helix), defined 4:457, 1:345
- Alphaherpesvirinae* 1:321
- alphaherpesvirus 2:442, 2:445
 defined 2:441
- alpha-HPV E6 proteins 2:498
- alpha-like viruses, defined 3:839
- Alphamesonivirus* 4:804, 4:805*t*, 4:807
- Alphamesonivirus* 1 4:804
- alphamesonivirus 4 4:804
- Alphanemrhavirus* 4:715–716
- alphanodaviruses 4:709–710
- Alphanudivirus* 4:827, 4:827–828
- alphapartitivirus 4:635
- alphapleolipovirus, defined 4:380
- Alphaproteobacteria* 4:332–334
 marine *Pelagibacterales*, phages of 4:332–334
 marine *Rhodobacteraceae*, phages of 4:334–335
 SAR116 phages 4:334
- alphasatellite components, CLCuD-associated 3:359–361
- alphasatellites (*Alphasatellitidae*) 3:149, 3:753
 classification 3:149–150
 defined 3:470
 life cycle/epidemiology 3:150–152
 likely origins of 3:150
 pathogenesis 3:152–153
 structure of 3:150
- Alphasatellitidae* 3:150
- Alphatetraviridae* 4:897, 4:717
- alphatetravirus 4:900, 4:900*f*
- alphavirus assembly and budding 1:470–472, 1:471*f*
- alphaviruses 1:266
- alphaviruses causing encephalitis
 clinical features of infection 2:46
 epidemiology 2:44–45
 evolution 2:43–44
 future 2:47
 genetics 2:43
 genome, properties of 2:40–41
 geographic and seasonal distribution 2:42–43
 history 2:40
 host range and virus propagation 2:43
 pathology and histopathology 2:46–47
- prevention, treatment and control 2:47
 replication 2:42
 taxonomy and classification 2:40
 transmission and tissue tropism 2:45–46
 pathogenicity 2:45–46
 viral proteins, properties of 2:41
 nonstructural proteins 2:41
 structural proteins 2:41–42
 virion, properties of 2:40
- alphavirus life cycle 1:469
- alphavirus RNA synthesis 2:176
- alphavirus unique domain (AUD) 2:177
- alphavirus virion structure 1:469–470
- Alphavodaviruses 4:822
- ALPV *see* aphid lethal paralysis virus (ALPV)
- ALRI *see* acute lower respiratory tract infections (ALRI)
- ALT *see* alanine aminotransferase (ALT)
- Alternaria alternata* 4:545–546
- alternaria alternata* virus 1 (AaV1) 4:544
- alternaviruses 4:544, 4:545*t*
 biological effects of 4:545–546
 chrysoviruses, evolutionary relationships among 4:547–548
 genome organization 4:544
 3' poly (A) structure of 4:547
 proteins encoded by 4:546–547
 virion properties 4:544–545
- ALV *see* avian leukosis virus (ALV)
- ALVX *see* Allium virus X (ALVX)
- Amalgaviridae* 3:154
- amalgaviruses (*Amalgaviridae*)
 classification 3:154
 diagnosis 3:155–157
 epidemiology 3:155
 genome 3:154
 life cycle 3:154–155
 pathogenesis 3:155
 prevention 3:157
 treatment 3:157
 virion structure 3:154
- amantadine 5:161, 5:171
- Amasya cherry disease (ACD) 4:646–647
- ambisense, defined 4:764, 2:765, 4:835
- ambisense genome, defined 3:507
- ambisense polarity, defined 3:719
- Amblyomma hebraeum* 1:547
- Amblyomma variegatum* 2:893–894
- American bat vesiculovirus (ABVV) 2:875
- AMEV *see* Amsacta moorei EPV (AMEV)
- AMGs *see* auxiliary metabolic genes (AMGs)
- 2-aminopurine (2-AM) 4:876
- aMLVs *see* amphotropic MLV (aMLVs)
- ampelovirus, defined 3:336
- amphibian alloherpesviruses
 classification 2:306–307
 clinical features 2:311–313
 diagnosis 2:313
 epidemiology 2:310
 genome 2:307–309
 life cycle 2:309
 management and treatment 2:313–314
 pathogenesis 2:311
 prevention 2:314–315
 virion structure 2:307
- amphotropic MLV (aMLVs) 2:646

- amplicon sequencing 1:178–179
Ampullaviridae 4:361
 AMR *see* analytical measurement range (AMR)
Amsacta moorei EPV (AMEV) 4:858
AMV see Alfalfa mosaic virus (AMV)
Anacridium aegyptium EV (AAEV) 4:865
 analogy, defined 1:87
 analytical measurement range (AMR) 5:72–73
 anamorph, defined 4:450
 anaphylaxis 5:305
 anastomosis 4:520, 4:589, 4:557, 4:450, 4:601, 4:607, 4:648, 4:552
Anatid herpesvirus 1 (AnHV1) 2:113
Anelloviridae 2:48
 classification 2:48–49
 clinical features 2:52
 diagnosis 2:54
 epidemiology 2:51–52
 life cycle 2:50–51
 pathogenesis 2:52–54
 treatment and prevention 2:54–55
 virion structure and genome 2:49–50
anelloviruses (AV) 2:48
 angiotensin-converting enzyme 2 (ACE2) 2:199–200
 Ångström, defined 4:10, 1:257
 anguillid perhabdovirus 2:325
 AnHV1 *see* *Anatid herpesvirus 1* (AnHV1)
 ANI *see* Asymptomatic Neurocognitive Impairment (ANI)
 animal adenoviruses 5:202
 animal morbilliviruses *see* morbilliviruses
 animal papillomaviruses *see* papillomaviruses
 Animal Reservoir, defined 2:397
 animal rhabdoviruses *see* rhabdoviruses
 animal viruses 1:6–7
 vector transmission of 1:542–543
 arthropod vectors of animal arboviruses 1:543–545
 co-evolution of viruses and vectors 1:545–546
 co-infection of arthropods with symbiotic bacteria 1:549–550
 environmental/anthropogenic factors contributing to vector-borne virus spread 1:546–547
 replication and transmission of arboviruses in their arthropod vector 1:547–548
 role of vector in genetic diversity of animal arboviruses 1:550
 virus–vector interactions and role of arthropod innate immune response 1:548–549
 ankyloses, defined 2:34
 ankyrin repeat (ANK)/F-box genes 2:672
Anomala cuprea EPV (ACEV) 4:859
Anopheles A serogroup 2:662
 antagonism, defined 4:419
 antagonist, defined 4:776
Antennaviruses 2:507
 anterograde, defined 2:738
 anthrax toxin receptor I (ANTXRI) 1:659
 anthropogenic factors, defined 1:542
 antiapoptotic proteins (AAP) 4:813
 antibiotic resistance, defined 4:252
 antibiotic resistance genes (ARG) 1:642
 Antibiotic Resistant (AR) bacteria 4:252
 antibiotics 4:252, 5:301
 Antibiotic Susceptibility Test (AST) 4:252–253
 defined 4:252
 antibodies in antiviral humoral immunity 1:591
 antibody-dependent cellular cytotoxicity (ADCC) 1:591–592
 neutralization 1:591
 opsonization 1:591
 antibody, defined 2:441
 antibody assay, neutralizing 5:19
 antibody avidity, defined 2:778
 antibody-dependent cellular cytotoxicity (ADCC) 1:584, 1:591–592
 antibody-dependent cellular phagocytosis (ADCP) 1:591
 antibody-dependent enhancement, defined 2:899
 antibody prevalence, defined 2:362
 antibody repertoire, sequencing analysis of 5:20
 anti-CRISPR (*Acr*) genes, defined 4:400
 anti-CRISPR proteins 4:248–249
 of archaeal viruses 4:405
 guilt-by association bioinformatic approach 4:249–250
 lytic phage dependent functional approach 4:250
 prophage-dependent functional approach 4:249
 self-targeting bioinformatic approach 4:250
 antigen, defined 2:441
 antigen detection 5:210
 hepatitis A virus 5:210
 hepatitis E virus 5:210
 antigenemia, defined 2:789
 antigenic complex, defined 2:765
 antigenic drift, defined 2:551
 antigenicity and antibody interactions 1:287–288
 antigenicity and antigenic variation
 dengue virus (DENV) envelope 1:598–599
 enterovirus A71 capsid 1:599–600
 human immunodeficiency virus type 1 envelope glycoprotein 1:599
 influenza A virus hemagglutinin 1:597–598
 antigenic shift, defined 2:551
 antigenome, defined 2:747
 antigenome promoter (AGP) 2:69–70
 antigen presentation 1:602–603
 MHC-I antigen presentation 1:603
 MHC-II antigen presentation 1:603–604
 viral subversion of 1:604–605
 antigen-presenting cells (APCs) 1:584, 1:601–602, 1:662, 2:512–513, 1:601
 anti-hypovirus defense mechanisms 4:592–593
 antiretroviral therapy (ART) 2:468
 non-nucleoside reverse-transcriptase inhibitors (NNRTIs) 5:135
 chemical structures of 5:135f
 clinical uses of 5:135–136
 drug-drug interactions with 5:136
 mechanisms of NNRTI resistance 5:136–137
 and their mechanism of action 5:135
 toxicities 5:136
 nucleoside/nucleotide reverse-transcriptase inhibitors (NRTIs) 5:132–133
 clinical uses of 5:133
 drug-drug interactions with 5:134
 mechanisms of NRTI resistance 5:134–135
 new drugs 5:135
 and their mechanism of action 5:132–133
 toxicities 5:133–134
 reverse transcriptase structure and functions 5:131
 reverse transcription, process of 5:131–132
 reverse transcriptase (RT) 5:132
 RNase H 5:132
 tRNA primer 5:131–132
 antisense RNA technology 3:362
 antiviral agents 1:478–479
 antiviral classification 5:121
 hepatitis B virus (HBV) 5:128
 HBV DNA polymerase 5:128
 hepatitis C virus (HCV) 5:123
 HCV NS3/4A protease 5:123
 HCV NS5A phosphoprotein 5:123
 HCV NS5B polymerase 5:123
 herpes simplex virus (HSV) 5:124
 HSV DNA polymerase UL30 5:124
 HSV envelope protein 5:124–125
 human cytomegalovirus (HCMV) 5:125
 HCMV DNA polymerase UL54 5:125
 HCMV terminase UL56 5:125
 human immunodeficiency virus (HIV) 5:121
 HIV GP120 5:122–123
 HIV GP41 5:122
 HIV integrase 5:122
 HIV protease 5:121–122
 HIV reverse transcriptase 5:121
 human influenza virus 5:125
 matrix protein 2 5:127
 neuraminidase 5:125–127
 viral RNA polymerase 5:125
 respiratory syncytial virus (RSV) 5:123–124
 RSV Fusion glycoprotein 5:124
 RSV RNA polymerase 5:123–124
 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) 5:127
 SARS-CoV-2 polymerase 5:127
 varicella zoster virus (VZV) 5:127–128
 VZV DNA polymerase 5:128
 variola virus (human smallpox) 5:127
 VP37 envelope wrapping protein 5:127
 antiviral defense mechanisms 4:394–397
 CRISPR-Cas systems 4:397
 toxin-antitoxin (TA) system 4:397–398

- antiviral drug resistance, defined 5:121
antiviral drugs 1:286–287
antiviral gene silencing, defined 3:116
antiviral immunity, defined 3:116
antiviral innate immunity 1:577
 antiviral cytokines 1:578
 cellular innate immunity 1:578
 complement system 1:577–578
 defensins 1:577
 direct antiviral effects of Type I IFNs 1:580–581
 good cop–bad cop 1:582
 indirect antiviral effects of Type I IFNs 1:581
 innate immunity memory 1:581
 interferon induction 1:579–580
 new kids in the gut: the Type III interferons 1:578–579
 tonic IFN Levels 1:581
 Type I IFN signaling 1:580
 viral counterstrategies 1:581–582
antivirals 5:204–205
antiviral therapies, viral factories as targets for 1:499
ANTXR1 *see* anthrax toxin receptor I (ANTXR1)
Anulavirus 3:260, 3:262*t*
 characteristics of RNA genome in 3:261*t*
AOM *see* acute otitis media (AOM)
AP1, defined 2:875
aparavirus 4:768, 4:769*t*, 4:770*t*, 4:774
APBV1, *see* *Aeropyrum pernix* Bacilliform Virus 1 (APBV1)
APCI MS *see* atmospheric pressure chemical ionization (APCI) MS
APCs *see* antigen-presenting cells (APCs)
ApDNV-1, *see* *Aedes pseudoscutellaris* dinovernavirus 1 (ApDNV-1)
aphid-borne, defined 3:594
aphid lethal paralysis virus (ALPV) 4:771
aphid transmission factor 3:318
aphid-transmitted caulimoviruses, case of 3:112
aphid-transmitted cucumoviruses, case of 3:112
aphid-transmitted luteovirids, case of 3:107–108
aphid-transmitted potyviruses, case of 3:112
apical membrane, defined 1:529
aplastic anemia, defined 2:362
APMV-1 *see* avian paramyxovirus 1 (APMV-1)
APOBEC, defined 4:835
APOBEC3 2:65
apolipoprotein E (ApoE) 2:386–387
apoptosis 2:895–896, 4:858, 2:182, 2:22, 2:738, 1:577, 2:428, 4:724, 4:732
 inhibitor of 2:672
apoptosome, defined 3:60
apoptotic bodies, defined 1:577, 4:724
aposymbiotic, defined 4:780
APOV1 *see* *aeropyrum pernix* ovoid virus 1 (APOV1)
APVs *see* avipoxviruses (APVs)
aquareovirus 1:306, 4:871
Arabis mosaic virus (ArMV) 3:628–629
ARAV *see* Aravan lyssavirus (ARAV)
Aravan lyssavirus (ARAV) 2:738
AR bacteria *see* Antibiotic Resistant (AR) bacteria
“arbitrium” system 1:638
arboviruses 2:891, 1:542–543, 2:899, 1:542, 2:218, 2:173, 2:40, 4:764, 2:654
 arthropod vectors of animal arboviruses 1:543–545
 replication and transmission in their arthropod vector 1:547–548
 role of vector in genetic diversity of animal arboviruses 1:550
 in their invertebrate vectors 4:722, 4:720*t*
Archaea 1:367–369
 phage egress in 1:517–518
Archaea, antiviral defense mechanisms in 4:400
 archaeal innate antiviral systems 4:405–406
 CRISPR-Cas 4:400–401
 anti-CRISPR proteins of archaeal viruses 4:405
 archaeal Type I CRISPR-Cas systems 4:401–402
 archaeal Type III CRISPR-Cas systems 4:402–405
 CRISPR loci and crRNA biogenesis, expression of 4:401
 novel archaeal CRISPR-Cas systems 4:405
 spacer acquisition and its regulation in archaea 4:400–401
 future perspectives 4:406
Archaea, virus-host interactions in 4:387–388
 antiviral defense mechanisms 4:394–397
 CRISPR-Cas systems 4:397
 toxin-antitoxin (TA) system 4:397–398
 counterdefense mechanisms 4:398
 genome integration 4:391–392
 genome replication 4:390–391
 transcription 4:392–393
 control 4:393
 regulators 4:392–393
 virion egress 4:393–394
 cell membrane disruption 4:393–394
 membrane disruption, viral release without 4:394
 virus entry 4:388–389
 cell-surface, interaction with 4:389–390
 cellular appendages, interaction with 4:388–389
 kinetics of virus entry 4:390
 archaeal innate antiviral systems 4:405–406
 archaeal Type I CRISPR-Cas systems 4:401–402
 archaeal Type III CRISPR-Cas systems 4:402–405
 archaeal virus
 see also bacterial and archaeal viruses
 defined 1:402, 4:380
 ecology and evolution shaping 4:424–425
 in hot spring environments using viral metagenomics 4:407
 learning 4:410–412
 next generation sequencing of samples 4:409
 post sequence analysis 4:409–410
 sample collection and processing 4:407–409
 and viral metagenomes 4:414–415
 archaeo-eukaryotic primases (AEP) 1:433
 Archimedean lattices 1:248, 1:252*f*
 architecture of viruses 1:475
 ARDS *see* acute respiratory distress syndrome (ARDS)
 arenaviruses, human pathogenic 2:508*f*
 clinical features 2:513
 Lassa fever (LF) and Lujo hemorrhagic fever (LHF) 2:513
 lymphocytic choriomeningitis (LCM) 2:513–514
 New World (NW) Mammarenaviral HF 2:514
 diagnosis 2:514
 Lassa fever (LF) 2:514
 lymphocytic choriomeningitis (LCM) 2:514–515
 New World (NW) mammarenaviral HF 2:515
 history and classification 2:507
 human pathogenic mammarenaviruses, epidemiology of 2:511
 argentinian hemorrhagic fever (AHF) 2:512
 Bolivian hemorrhagic fever (BHF) 2:512
 Lassa fever (LF) 2:511
 Lujo hemorrhagic fever (LHF) 2:511
 lymphocytic choriomeningitis (LCM) 2:511–512
 Venezuelan hemorrhagic fever (Vehf) 2:512
 mammarenavirus genome organization and proteins 2:507–509
 mammarenavirus life cycle 2:509–510, 2:509*f*
 assembly and budding 2:511
 cell attachment and entry 2:509–510
 expression and replication of the viral genome 2:510–511
 mammarenavirus virion structure 2:507
 pathogenesis and pathology 2:512–513
 prevention and control 2:515
 antibody therapy 2:515–516
 antiviral drugs 2:515
 medical management 2:515
 vaccines 2:516
 ARG *see* antibiotic resistance genes (ARG)
 argentinian hemorrhagic fever (AHF) 2:512, 2:507, 2:514
 arginine finger 4:166
 defined 4:148
 arginine toggle 4:164
 Argonaute (AGO) proteins 3:43
 argonaute, defined 3:52, 3:594, 3:123
 Argonaute 1 (AGO1) protein 3:758, 3:692
 arming, defined 5:233
 arming viruses 5:241
 immunomodulators 5:241
 prodrug convertases 5:241
 radiosensitizers for therapy and imaging 5:241
 ArMV *see* Arabis mosaic virus (ArMV)

- ART *see* antiretroviral therapy (ART)
- art and literature, viruses in 1:679
- Arteriviridae* 1:499
- arteriviruses 1:499
- arthralgia, defined 2:173
- arthrogryposis, defined 2:34
- arthrogryposis-hydranencephaly syndrome (AHS) 2:36
- arthropod-borne viruses 2:891
- of vertebrates 2:417
- arthropods, bunyaviruses of *see* bunyavirus: of arthropods
- arthropod-specific virus, defined 4:764
- arthropod transmission cycles, influence of 1:563–564
- arthropod vectors of animal arboviruses 1:543–545
- artificially established viral cross infections between plants and fungi 4:445–447
- artificial microRNAs, defined 3:293
- Artoviridae* 4:703
- ARV *see* Adelaide River virus (ARV)
- AS1 *see* Asymmetric Leaves gene 1 (AS1)
- ASCE *see* additional strand, conserved E (ASCE); additional strand conserved glutamate (ASCE)
- ascocarp, defined 4:450
- ascomycete/basidiomycete, defined 4:568
- ascomycetous fungi, defined 4:594
- ascospores, defined 4:607, 4:450
- Ascoviridae* 4:703, 4:724
- ascoviruses 4:724
- distribution and taxonomy 4:724–725
- future perspectives 4:730
- history 4:724
- origin and evolution 4:730
- pathology and pathogenesis 4:728
- cytopathology and cell biology 4:728–729
- signs of disease 4:728
- tissue tropism 4:729
- replication and virion assembly 4:729–730
- transmission and ecology 4:727–728
- host range 4:728
- virion structure and composition 4:725–727
- aseptic meningitis 2:812
- ASFV *see* African swine fever virus (ASFV)
- ASGPR *see* asialoglycoprotein receptor (ASGPR)
- ASHBV *see* ashy-headed sheldgoose (ASHBV)
- ashy-headed sheldgoose (ASHBV) 2:100
- asialoglycoprotein receptor (ASGPR) 1:537
- Asia Minor, defined 3:586
- ASLV *see* Avian Sarcoma and Leukosis Virus (ASLV)
- aspartate aminotransferase (AST) 2:896
- Aspergilli*, diversity of mycoviruses in 4:450
- aspergillus foetidus* mycovirus complex 4:453–454
- chrysovirus 4:454
- genomes 4:451–453
- narnaviruses and mitoviruses 4:454
- partitiviruses 4:454
- phenotypes 4:455
- polymycoviruses 4:454–455
- prevalence 4:450–451
- transmission 4:451
- Aspergillus flavus* 4:520
- Aspergillus foetidus* dsRNA mycovirus (AfV-F) 4:544
- Aspergillus foetidus* slow virus S2 (AfV-S2) 4:660
- assay, defined 4:175
- assay verification and validation 5:70–72
- accuracy 5:72
- linearity/analytical measurement range (AMR) 5:72–73
- precision 5:72
- reportable range and limit of detection (LoD) 5:72
- assembly, defined 1:382
- assembly code embedded within viral genetic message 1:254–255
- assembly of virion 1:480
- icosahedral capsid assembly, structural principles in 1:480–481
- modes of assembly 1:481–483
- capsid self-assembly 1:483
- scaffolding protein-assisted capsid assembly 1:483–486
- viral genome-assisted capsid assembly 1:486
- synthesis of nascent capsid assembly and assembly intermediates 1:481
- assembly of viruses 1:468
- antigenic sites 1:478
- antiviral agents 1:478–479
- architecture of viruses 1:475
- assembly 1:477–478
- evolution 1:477
- helical viruses, atomic structure of 1:475–476
- host receptor recognition site 1:478
- icosahedral enveloped viruses 1:468–469
- alphavirus assembly and budding 1:470–472
- alphavirus life cycle 1:469
- alphavirus virion structure 1:469–470
- flavivirus assembly and budding 1:473
- flavivirus life cycle 1:472
- flavivirus virion structure 1:472–473
- nucleic acid–protein interaction 1:476–477
- spherical viruses, atomic structure of 1:476
- structure determination, methods of 1:475
- viral envelope 1:468
- assembly pathways 4:49
- AST *see* Antibiotic Susceptibility Test (AST); aspartate aminotransferase (AST)
- asthma 2:761–762
- astrocytes, defined 2:778
- Astroviridae*
- classification 2:92–93
- clinical features 2:95
- animal disease 2:95–96
- human disease 2:95
- diagnosis 2:98
- epidemiology 2:95
- classical HAstV 2:95
- novel HAstV 2:95
- genome 2:93–94
- life cycle 2:94
- assembly and release 2:94–95
- binding and entry 2:94
- uncoating and replication 2:94
- pathogenesis 2:96
- immune response 2:97–98
- mechanism of disease 2:96–97
- obstacles 2:96
- prevention 2:98
- treatment 2:98
- virion structure 2:93
- asunaprevir 5:123
- asymmetrical flow field-flow fractionation (AF4) 1:171, 1:162, 1:172f
- Asymmetric Leaves gene 1 (AS1) 3:245
- asymmetric unit, defined 1:248
- Asymptomatic Neurocognitive Impairment (ANI) 2:471
- Atadenovirus* 2:13
- atazanavir 5:141
- atherosclerosis, defined 2:778
- atherosclerotic vascular disease 2:454
- ATL *see* adult T-cell leukemia (ATL)
- atmosphere 4:354–355
- atmospheric pressure chemical ionization (APCI) MS 5:92
- atomic force microscopy (AFM)
- of biological specimens 1:219–220
- complex viruses, dissection of 1:227–230
- defined 4:53
- icosahedra, triangulation numbers of 1:222–223
- infection and budding from host cells 1:226
- initial observations from 1:220–222
- principles and technology 1:218–219
- recombinant and mutant virus particles 1:223
- specific and special structural features 1:226–227
- viral assembly 1:232
- viral nucleic acid, visualization of 1:230–232
- virus fibers and membranes 1:223–226
- atomic level structures, classification of viral world based on 1:153–154
- case study 1:158–160
- future perspectives 1:160–161
- incorporating different levels of data 1:157
- inferring phylogeny based on structural data 1:158
- methods to compare (viral) proteins 1:154–157
- structure-based classification, databases for 1:157–158
- ATP *see* adenosine triphosphate (ATP)
- ATPase, defined 4:160, 4:105
- ATPase gp16, defined 4:302
- ATP-driven genome packaging, defined 1:488
- ATSV, *see* *Acidianus* tailed spindle virus (ATSV)
- attachment, defined 1:382
- attachment factor, defined 1:388

ATV, *see* *Acidianus* two-tailed virus (ATV)
 AUD *see* alphavirus unique domain (AUD)
 audits, internal and external 5:70
Aumavirus 3:581, 3:582
 Australian bat lyssavirus (ABLV) 2:738, 2:740
 autochthonous transmission, defined 2:173
Autographa californica multiple nuclear polyhedrosis virus (AcMNPV) 1:654–655
 autoimmune disease, defined 2:884
 autoimmune vasculitis 2:454
 autophagy 4:824
 defined 1:529, 2:738, 3:420, 2:428, 1:495
 autophagy-related vesicle-mediated release of virus 1:539
 auxiliary metabolic genes (AMGs) 1:621
 auxotrophic complementation 4:522
 AV *see* anelloviruses (AV)
 avian adenoviruses 2:13
 Avian Avulavirus 1 (AAV-1) 2:648
 life cycle 2:649f
 avian hepadnaviruses
 classification 2:100
 diagnosis 2:107–108
 endogenous avian hepadnavirus elements 2:109–111
 epidemiology 2:106
 genome 2:101–103
 life cycle 2:103–106
 pathogenesis and clinical features 2:106–107
 treatment 2:108–109
 virus structure 2:100–101
 avian herpesviruses (AHV)
 classification 2:112–113
 clinical features 2:115
 diagnosis 2:116
 epidemiology 2:114–115
 genome 2:113–114
 host range 2:113
 pathogenesis 2:115–116
 prevention 2:116
 replication cycle 2:114
 virion structure 2:113
 avian hosts 2:805, 2:813
 avian influenza 5:274
 avian influenza viruses (AIV) 2:117
 clinical features and pathogenesis 2:118–119
 epidemiology 2:118
 life cycle 2:117–118
 prevention 2:119–121
 taxonomy 2:117
 virion structure and genome 2:117
 avian leukosis virus (ALV) 2:122
 avian paramyxovirus 1 (APMV-1) 2:648
 Avian Sarcoma and Leukosis Virus (ASLV) 1:211, 2:122
 avidity, defined 4:136
 avipoxviruses (APVs) 2:343
 classification 2:343
 control and treatment 2:347
 use of APV as recombinant vaccine vectors 2:347
 epidemiology 2:346

genome 2:344–345
 life cycle 2:345–346
 pathogenesis 2:346–347
 host range 2:346–347
 immunomodulation 2:347
 replication cycle of 2:345f
 virion structure 2:343–344
 avirulence, defined 3:69
 avirulence determinant, defined 3:761
 avirulence gene, defined 3:60
Avsunviroidae 3:852
 AYVB *see* ageratum yellow vein betasatellite (AYVB)
 AYVINA, *see* *Ageratum yellow vein India alphasatellite* (AYVINA)
 AYVV, *see* *Ageratum yellow vein virus* (AYVV)

B

Bacilladnavirus 4:685
 bacilli double-stranded DNA bacteriophages, replication of 4:61–63
 of 29-like viruses 4:66–67
 of SPO1-like viruses 4:63–65
 of SPP1-like viruses 4:65–66
Bacillus subtilis phage SPP1 4:206
 Background Selection, defined 5:227
 backyard production systems (BPS) 2:118
 bacteria 1:664
 bacterial and archaeal viruses
 in different sizes 1:162–164
 genome replication of
 DNA replication mechanisms 1:435–436
 prokaryotic dsDNA viruses, genome replication of 1:429–430
 prokaryotic RNA viruses, genome replication of 1:437
 prokaryotic ssDNA viruses, genome replication of 1:436
 prokaryotic viruses, genomes of 1:429
 protein-primed DNA replication 1:433–434
 RNA-dependent RNA polymerase, replication using 1:437
 RNA-primed DNA replication 1:430–433
 rolling circle DNA replication (RCR) 1:434–435
 rolling circle replication 1:436–437
 isolation and culturing of 1:164–166
 plaque morphologies 1:165f
 polyphyletic group of viruses 1:162–164
 precipitation methods 1:167–168
 preparative ultracentrifugation methods 1:168–170
 chromatographic methods 1:170–171
 flow field-flow fractionation methods 1:171–172
 purification performance, assessing 1:172
 purification of viruses 1:166–167
 ultrafiltration methods 1:167, 1:167f
 bacterial and archeal virus entry 1:402, 1:404f

genome delivery mechanisms of phages 1:405–406
 filamentous DNA viruses 1:407
 icosahedral DNA viruses with an internal membrane 1:406–407
 icosahedral enveloped dsRNA bacterial viruses 1:407–408
 icosahedral ssDNA bacterial viruses 1:407
 icosahedral ssRNA bacterial viruses 1:408
 icosahedral tailed dsDNA bacterial viruses 1:405–406
 pleomorphic DNA viruses 1:407
 host cell barriers 1:402
 host recognition and adsorption 1:403
 viral enzymes in capsule and cell wall penetration 1:403–405
 virion as a genome delivery device 1:402–403
 bacterial artificial chromosome, defined 2:306
 bacterial evolution, role of bacteriophages in
 coevolution
 defense mechanisms 1:641–642
 integration sites 1:641
 in the patient 1:641
 polylysogeny 1:638–640
 prophage remnants 1:640–641
 war and peace 1:642
 lysis-lysogeny decision 1:638
 synergistic phage-bacterium relationships 1:633–634
 synergistic phage-bacterium relationships 1:634–638
 bacteria-like versus virion-emphasizing modes of existence 4:318
 bacterial virus 1:6
 defined 1:402
 bacteriophage 4:253, 4:252–253, 4:291–292, 1:552–553, 4:277, 4:77, 4:80–81
 defined 5:5, 4:291, 4:302
 discovery of phages 4:3–4
 of human microbiome 4:283–284
 clinical utility of 4:290
 dynamics and implications of human phage community 4:287–290
 human phageome distribution and composition 4:284–287
 phage and biotechnology 4:7–8
 phage biochemistry 4:6–7
 phage development
 as diagnostics for bacterial infections 4:254–255
 as therapeutics for bacterial infections 4:253–254
 phage diagnostics, additional challenges for 4:257
 intracellular bacteria 4:257
 market acceptance 4:257
 multiplex testing 4:257
 phage manipulation 4:257
 sample interference 4:257
 phage ecology and evolution 4:8

- bacteriophage (*continued*)
- phage genetics 4:5–6
 - phage technology, major advantages of 4:255
 - self-replication 4:255
 - specificity 4:255
 - phage technology, major limitations of 4:255
 - narrow host range 4:255
 - phage resistance 4:255
 - phage therapy 4:4
 - phage therapy, additional challenges for 4:255–256
 - chemistry, manufacturing, and controls (CMC) 4:256
 - dosing strategies 4:256
 - environmental impact 4:256–257
 - preclinical data translation 4:255–256
 - regulatory pathway 4:256
 - safety and efficacy 4:256
 - phage typing 4:4–5
 - structure and function 4:46
 - capsids 4:46
 - tails and tail fibers 4:46–48
 - T4 4:70–72
 - T7 4:72
 - early transcription 4:72
 - elongation 4:74
 - phage RNAP and late gene transcription 4:72
 - termination 4:74
 - transcription cycle 4:72–74
 - temperate versus virulent bacteriophages 4:48
 - bacteriophage 186 4:81–83
 - bacteriophage adherence to mucus (BAM) 1:556–557
 - bacteriophage diversity 4:265–266
 - examples of 4:268–269
 - bottom up studies 4:269
 - contribution of prophages to phage diversity 4:270
 - Enterobacteriales* tailed phage example 4:269–270
 - human phageome 4:268–269
 - marine phageome 4:269
 - top down studies 4:268–269
 - two large diversity studies 4:269
 - genome mosaicism 4:272
 - genetic exchange among superclusters 4:272
 - and horizontal exchange of genetic information 4:270–272
 - nature of horizontally exchangeable mosaic section alleles 4:270–272
 - phage-host relationships and phage diversity 4:272–273
 - narrow and wide host range phages 4:273–274
 - phage cluster-host species relationships with *Enterobacteriales* 4:272–273
 - strategies for studying 4:267–268
 - tailed phage diversity and phage classification 4:266–267
 - bacteriophage DNA packaging nanomotors 4:302–305
 - application of 4:308
 - biological nanopore channel, applications of 4:309–312
 - mechanism of single pore sensing 4:308–309
 - Phi29 connector 4:308
 - SPP1 connector 4:308
 - T7 connector 4:308
 - revolving mechanism of biomotors in bacteriophages 4:305
 - revolution mechanism without rotation 4:305
 - revolving mechanisms determined by motor channel sizes 4:305–306
 - revolving motors distinguished by chirality 4:306–307
 - special aspects of revolving motor actions 4:307–308
 - Bacteriophage Exclusion (BREX) 1:608
 - bacteriophage red recombination system 4:291
 - biological roles 4:293
 - DNA repair 4:293–294
 - genetic exchange among similar phages 4:293
 - role in DNA replication 4:293
 - classical models of recombination 4:294
 - homologous recombination systems in *E. coli* and phage 4:292–293
 - bacteriophage-phage, defined 4:283
 - bacteriophage plaques 4:315f
 - bacteriophage receptor proteins of gram-negative bacteria
 - history and overview 4:175
 - phage-host interactions, ecology and evolution of 4:181–182
 - phage-receptor interactions' relevance to phage therapy 4:182–183
 - properties of receptor proteins 4:176–181
 - techniques and methods for studying phage receptors 4:175–176
 - techniques for studying 4:177t–178
 - bacteriophages, general ecology of 4:314
 - phage, microscopic determination of 4:315–316
 - phage community ecology 4:320
 - phage ecosystem ecology 4:320–321
 - phage isolation and host range 4:315
 - phage organismal ecology 4:317–318
 - bacteria-like versus virion-emphasizing modes of existence 4:318
 - phage population ecology 4:318–320
 - bacteriophage tail fibers, structures of 4:195–199
 - bacteriophage tailspikes, structures of 4:199–201
 - bacteriophage vaccines 4:259
 - in vitro* VLP assembly 4:261–262
 - phage T4 VLP vaccines 4:261–262
 - phage T4, architecture of 4:259
 - phage T4 as a vaccine platform 4:259–261
 - in vivo* VLP assembly 4:259–261
 - phage vaccine platforms 4:262–263
 - Bacteroidetes* 4:336–337
 - Baculoviridae* 4:703–705
 - baculovirus 4:747
 - acquisition of genes 4:733–735
 - baculoviral insecticides 4:745
 - baculovirus expression technology 4:745
 - baculovirus transmission and host behavioral manipulation 4:744–745
 - defined 4:827
 - evolution 4:743–744
 - gene expression 4:752–754
 - early gene expression 4:752–754
 - late and very late gene expression 4:754
 - temporal regulation of transcription 4:752–754
 - genome organization and content 4:749–752
 - genomes, gene content, organization 4:741–743
 - historical perspective 4:739
 - host range genes 4:735–736
 - infection cycle 4:744, 4:747–748
 - dissemination of OBs 4:749
 - oral and systemic infection 4:748–749
 - two virion phenotypes 4:748
 - morphology 4:740–741
 - nomenclature, taxonomy, and classification 4:739–740
 - replication 4:733
 - viral DNA replication 4:754–755
 - virion morphogenesis and protein composition 4:755
 - BV morphogenesis 4:755–756
 - F protein properties 4:756
 - GP64 glycoprotein characteristics 4:756
 - nucleocapsids 4:755
 - ODV assembly and ODV envelopes 4:756–757
 - tegument proteins 4:757
 - virus defense and resistance to 4:736
 - baculovirus expression vector system (BEVS) 4:705, 4:704
 - baculovirus-host interactions 4:732
 - hosts acquiring genes from infecting viruses 4:732–733
 - viruses acquiring genes from their hosts 4:732
 - baculovirus of cricket 4:827–828
 - baculovirus X 4:827–828
 - badnavirid, defined 3:274
 - badnaviruses (*Caulimoviridae*) 3:158, 3:159t–161, 3:162f, 3:165, 3:167
 - classification 3:158
 - clinical features 3:166
 - defined 3:274
 - diagnosis 3:167–168
 - disease symptoms caused by 3:166f
 - epidemiology 3:165–166
 - genome 3:163
 - life cycle 3:163–165
 - pathogenesis 3:166–167
 - prevention 3:168
 - virion structure 3:158–163
 - Baloxavir + Oseltamivir 5:171
 - Baloxavir marboxyl 2:559
 - Baloxavir – Marboxyl (Xofluzo) 5:166–168
 - BALT *see* bronchus-associated lymphoid tissue (BALT)

- baltimore classes, evolutionary status of 1:44
- baltimore classification (1971) 1:29
- BAM *see* bacteriophage adherence to mucus (BAM)
- Bamboo Mosaic Virus (BaMV) 1:364–365
- BaMV *see* Bamboo Mosaic Virus (BaMV)
- banana bunchy top virus (BBTV) 3:169
- diagnosis 3:173–174
- epidemiology and control of 3:174–175
- genome organization and functions of gene products 3:170–171
- geographical distribution 3:174
- historical records and possible origins 3:174
- host range and transmission 3:171–172
- taxonomy, phylogeny, and evolution 3:169–170
- virion structure 3:170
- virus–host relationships 3:172–173
- banana streak disease complex (BSD) 3:166
- banana streak GF virus (BSGFV) 3:167
- banana streak IM virus (BSIMV) 3:167
- banana streak MY virus (BSMYV) 3:167
- banana streak OL virus (BSOLV) 3:167
- banana viral diseases 3:90
- bandicoot papillomatosis carcinomatosis virus type 1 (BPCV1) 1:76
- Banna virus 4:881
- barley yellow dwarf viruses (*Luteoviridae*) 3:176
- control 3:182–183
- diagnosis 3:182
- epidemiology 3:181–182
- genome organization and expression 3:177–180
- host range and transmission 3:180–181
- replication 3:181
- taxonomy and classification 3:176
- virion properties and composition 3:176–177
- virus–host relationships 3:181
- barnaviruses (*Barnaviridae*) 4:549
- MBV evolutionary relationships 4:550
- MBV genome organization and expression 4:549–550
- MBV transmission and host range 4:550
- MBV virion properties 4:549
- MBV virion structure and composition 4:549
- base-calling, defined 5:27
- baseplate-hub protein (BHP) pb3 4:191
- Basic local alignment search tool (BLAST) 5:33, 1:101
- basic reproductive number, defined 1:53
- basolateral membrane, defined 1:529
- bat coronavirus (BtCoV) 512 2:256
- bathypelagic layer, defined 4:322
- BBB *see* blood-brain barrier (BBB)
- BBLV *see* Bokeloh bat lyssavirus (BBLV)
- BBMV *see* Broad bean mottle virus (BBMV)
- BBSV *see* beet black scorch virus (BBSV)
- BBTV *see* banana bunchy top virus (BBTV)
- B cell receptors (BCRs) 1:585–587, 1:590–591, 1:663
- B cells 1:590–591
- BCMD *see* Brown Cap Mushroom Disease (BCMD)
- BCMV and BCMNV *see* Bean common mosaic virus (BCMV) and Bean common mosaic necrosis virus (BCMNV)
- BCoV *see* bovine coronavirus (BCoV)
- BCRs *see* B cell receptors (BCRs)
- BCTV *see* beet curly top virus (BCTV)
- BD *see* border disease (BD)
- Beak and feather disease virus* (BFDV) 2:184
- Bean common mosaic virus (BCMV) and Bean common mosaic necrosis virus (BCMNV)
- diagnosis 3:190
- host range and symptomatology 3:190
- reverse transcription-polymerase chain reaction 3:190
- serological techniques 3:190
- geographic distribution 3:185
- history 3:184–185
- host range and transmission 3:185–187
- pathogenicity, pathology, and resistance genes 3:188–189
- prevention and control 3:190–191
- properties of the virion and genome 3:187
- properties of particles 3:187
- properties of the genome and replication 3:188
- serological relationships 3:187–188
- recombination and variability 3:189–190
- taxonomy and classification 3:184–185
- bean dwarf mosaic virus (BDMV) 3:561
- bean golden mosaic virus and bean golden yellow mosaic virus (*Geminiviridae*) 3:192–194, 3:561
- diagnosis 3:197
- management 3:197–198
- taxonomy, phylogeny, and evolution 3:194–195
- transmission, host range, and epidemiology 3:196
- virion structure, genome organization, properties and functions of gene products 3:195–196
- virus–host relationships 3:196–197
- beans, geminivirus resistance in 3:561–563
- becurtovirus 3:411–412
- beet black scorch virus (BBSV) 3:795
- beet curly top virus (BCTV) 3:200, 3:363
- changes in BCTV strain prevalence 3:205
- classification and nomenclature 3:201–203
- experimental systems for infection of plants by 3:209–210
- genome organization and gene expression 3:205–206
- geographic and seasonal distribution 3:205
- history 3:200–201
- host range and pathogenesis 3:207–208
- management of curly top disease caused by 3:210
- cultural practices 3:210–211
- IPM for curly top disease 3:211
- management of the leafhopper vector with insecticides 3:211
- resistant varieties 3:210
- sanitation 3:211
- origin of BCTV and the beet leafhopper vector 3:203–205
- replication 3:206–207
- resistance to 3:560–561
- transmission of BCTV by the beet leafhopper vector 3:208–209
- virus movement 3:207
- beetles, plant viruses transmission by 3:112–113
- beet necrotic yellow vein virus (*Benyviridae*)
- classification 3:213
- diagnosis 3:217
- disease symptoms 3:216–217
- epidemiology 3:215–216
- genome organization and functions of gene products 3:213–215
- prevention 3:217–218
- transmission 3:215
- virion structure 3:213
- virus–host relationships 3:217
- beet poleroviruses 3:599–600
- BEFV *see* bovine ephemeral fever virus (BEFV)
- begomoviruses 3:13–14, 3:412, 3:357
- CLCuD in Africa 3:357
- CLCuD in Indian subcontinent 3:357–359
- behavioral fever, defined 2:306
- Belpaoviridae* 3:654
- benyviruses (*Benyviridae*)
- control 3:226–227
- defined 3:213, 3:219
- diagnosis 3:226
- disease, host range, and epidemiology 3:224–226
- history 3:219
- organization of genome and properties of encoded proteins 3:220–224
- similarities and dissimilarities with other taxa 3:227
- transmission 3:226
- virion structure 3:219–220
- Berrimah virus (BRMV) 2:875
- barrel, defined 3:692, 3:364, 3:348
- barrel fold, defined 1:329
- C1 3:239, 3:244, 3:245
- beta chemokines, defined 2:827
- betaentomopoxvirus 4:859
- betaflexiviruses (*Betaflexiviridae*) 3:229, 3:805
- genome organization 3:234–235
- geographical distribution 3:237–238
- members of the family 3:232
- properties and functions of gene products 3:235
- replication and propagation 3:235–237
- taxonomy and phylogeny 3:229–232
- transmission and host range 3:237
- virion structure 3:232–234
- beta-helical tailspikes, structures of 4:199f
- helix, defined 1:329
- Betaherpesvirinae* 1:321
- betaherpesvirus 2:442, 2:451f

- defined 2:441
 beta-hexamer 3:262–263
Betairidovirinae 4:707–708
Malacoherpessviridae 4:708
Mesoniviridae 4:708–709
Nimaviridae 4:709
Nodaviridae 4:709–710
Nudiviridae 4:710
Nyamiviridae 4:710–711
 Betanodaviruses 4:821
Betanudivirus 4:827, 4:827–828
 betapartitivirus 4:635
 betapleolipovirus, defined 4:380
 betasatellites (*Tolecusatellitidae*) 3:239
 classification and nomenclature 3:239–241
 CLCuD-associated 3:361
 defined 3:554
 genetic diversity and center of origin for 3:241–244
 functional role of bC1 in disease development 3:245
 life cycle, epidemiology 3:244
 replication 3:244–245
 resistance against betasatellites 3:245
 genome organization 3:241, 3:243f
 phylogenetic trees for 3:242f
 species of the genera *Betasatellite* 3:240t–241
 beta sheet, defined 1:345
 -spiral, defined 1:329
 beta-structure, defined 4:194
Betatetravirus 4:897
 BEVS *see* baculovirus expression vector system (BEVS)
 BFDV, *see* Beak and feather disease virus (BFDV)
 bhanja virus 2:776
 Bhendi yellow vein mosaic virus (BYVMV) 3:756
 BHF *see* Bolivian hemorrhagic fever (BHF)
 BIBD *see* boid inclusion body disease (BIBD)
Bicaudaviridae 1:369–370, 4:361
 bicaudavirus AIV 4:393
 bicistronic, defined 3:461, 3:439, 3:154
 bictegravir 5:150–151
 Phase 3 randomized controlled trials 5:150–151
 resistance 5:151
 bidensovirus 4:705
 classification 4:759
 genome organization and expression strategy 4:759
 pathology of silkworm associated with BmBDV 4:759–761
 viral non-structural proteins 4:761
 viral replication 4:761–762
 viral structural protein 4:761
 virus evolution 4:762
 bidirectional transcription, defined 2:144
Bidnaviridae 4:759, 4:705–706
 big-vein disease of lettuce, defined 3:833
 biochemical phase, of virology 1:6
 bio-crime, defined 1:644
 biological control, defined 4:615, 4:493, 4:468, 4:892
 biological warfare, defined 1:644
 biomotor/nanomotor, defined 4:302
 biopesticide, defined 4:892
 biosafety and biosecurity in diagnostic laboratories 5:82
 biorisk management system and legislation 5:89
 diagnostic laboratories 5:82
 laboratory biosafety 5:82–83
 biological agents 5:82–83
 biosafety cabinets 5:84–85
 decontamination and waste management 5:85–86
 laboratory accidents 5:86–87
 laboratory biosafety levels 5:83
 occupational health and special groups 5:87
 personal protective equipment (PPE) 5:84
 risk assessment 5:83
 routes of transmission and infective dose 5:83–84
 shipment of clinical specimens 5:87–88
 laboratory biosecurity 5:88–89
 Biosafety Level, defined 2:891
 Biosafety level 4 (BSL-4), defined 2:355
 bioterrorism, defined 1:644
 bioterrorism, viruses and their potential for 1:644
 agro-warfare 1:647–648
 biological agents for good or evil 1:648–649
 bioweapons, viruses as 1:646–647
 effects of viral introductions 1:649
 entomological warfare 1:647
 20th century state sponsored biological weapons programs 1:645–646
 unconventional bioweapons 1:644
 use of diseases in early history 1:644–645
 bioweapons 1:676–677
 viruses as 1:646–647
 bipartite begomoviruses 3:753, 3:758, 3:244
 birds, retroviruses of
 classification 2:122–123
 clinical features 2:125
 diagnosis 2:126
 endogenous retroviruses 2:126
 epidemiology 2:124–125
 genome 2:123–124
 history 2:122
 life cycle 2:124
 pathogenesis 2:125–126
 treatment 2:126
 virion structure 2:123
Birnaviridae 2:544, 4:776–777
 birnaviruses of invertebrates (entomobirnavirus) 4:706
Dicistroviridae 4:706
Hytrosaviridae 4:706–707
Iflaviridae 4:707
 birth defects, defined 2:797
 BIV *see* bovine immunodeficiency virus (BIV)
 bivalent oral polio vaccine (bOPV) 5:312–313
 BK *see* bradykinin (BK)
 BK polyomavirus (BKPyV) 2:522, 5:105, 5:108
 clinical features 2:522
 infectious cycle 2:522–523
 BKPyV *see* BK polyomavirus (BKPyV)
 Black queen cell virus (BQCV) 4:768–770, 4:771, 4:772, 4:773
 blood-brain barrier (BBB) 1:536, 2:847
 defined 2:884
 blood virome 1:556
 bluetongue virus (BTV) 2:127, 1:545, 1:303
 classification 2:127
 diagnosis 2:134
 epidemiology 2:132–133
 future perspectives 2:135–136
 life cycle 2:129–130
 assembly of genomic RNA and packaging 2:132
 core assembly 2:131–132
 transcription and replication 2:130–131
 virion maturation and egress 2:132
 virus attachment and entry into cells 2:129–130
 pathogenesis and clinical features 2:133–134
 replication cycle 2:130f
 structure of 2:129f
 vaccination and control 2:134–135
 virion structure and genome 2:127–129
 blunavirus 3:247
 diagnosis 3:251
 diseases, epidemiology, and control 3:250–251
 genome organization and gene product function 3:248–249
 members of the family 3:247
 replication and propagation 3:249–250
 taxonomy, phylogeny, and evolution 3:247
 transmission and host range 3:250
 virion structure 3:247–248
 BLV *see* bovine leukemia virus (BLV)
 B lymphocytes 1:590–591
 BM2 2:566–567
 BmBDV *see* Bombyx mori bidensovirus (BmBDV)
 BmDNV *see* Bombyx mori densovirus (BmDNV)
 BMV *see* brome mosaic virus (BMV)
 bnMAbs *see* broadly-neutralizing monoclonal antibodies (bnMAbs)
 bocaviruses
 clinical features 2:424
 epidemiology 2:422–424
 boceprevir 5:123
 boid inclusion body disease (BIBD) 2:507
 Bokeloh bat lyssavirus (BBLV) 2:738
 Bolivian hemorrhagic fever (BHF) 2:507, 2:512
 Bombyx mori bidensovirus (BmBDV) 4:759, 4:839
 Bombyx mori densovirus (BmDNV) 4:759–760

- bombyx mori latent virus 3:825
 properties and distinguishing characteristics 3:825
 bona fide viruses 1:14
 bOPV *see* bivalent oral polio vaccine (bOPV)
 border disease (BD) 2:153, 2:161
 control of BD of sheep 2:163
 Borell bodies, defined 2:343
 borna disease 2:139
 borna disease virus and related bornaviruses 2:137
 classification 2:137
 control and prevention 2:142–143
 diagnosis 2:142
 endogenous bornavirus-like elements 2:141
 epidemiology and clinical features 2:139
 bornavirus associated encephalitis in humans 2:139–140
 proventricular dilatation disease (PDD) of birds 2:140–141
 staggering disease in cats 2:139
 future prospects 2:143
 genome 2:137
 life cycle 2:137–139
 pathogenesis 2:141–142
 virus structure 2:137
 botryoid, defined 2:182
 Botrytis cinerea, hypovirulence in 4:476
 Botrytis virus F (BVF) 4:478–479, 4:437–438, 4:462
 biological properties 4:479
 family –*Gammaflexiviridae* 4:479
 genome structure 4:479
 genus –*Mycoflexivirus* 4:479
 phylogenetic relationships 4:479–480
 virion morphology 4:479
 Botrytis Virus X (BVX) 4:480, 4:438
 biological properties 4:480–481
 family –*Alfaflexiviridae* 4:480
 genome structure 4:480
 genus –*Botrexvirus* 4:480
 phylogenetic relationships 4:481
 virion morphology 4:481
 bottleneck, defined 4:457
 bottle-shaped ampullaviruses 1:433–434
 botybirnaviruses 4:552
 biology 4:554–556
 genome organization and replication 4:552–553
 taxonomy and similarity with other viruses 4:553–554
 transmission and distribution 4:554
 virion properties 4:552
 bovine, ovine, caprine, and other ruminant adenoviruses 2:13
 bovine coronavirus (BCoV) 2:198
 bovine ephemeral fever virus (BEFV) 2:875
 classification 2:875
 control and treatment 2:882–883
 epidemiology 2:880–881
 life cycle 2:878–879
 assembly and budding 2:879–880
 attachment, entry and uncoating 2:878–879
 gene expression and replication 2:879
 inhibition and modification of host cell functions 2:880
 pathogenesis and clinical features 2:881–882
 virus structure 2:875–878
 genomes and proteins 2:877–878
 bovine immunodeficiency virus (BIV) 2:56–57, 2:66
 bovine leukemia virus (BLV) 2:317
 classification 2:144
 diagnosis 2:144–145
 epidemiology and clinical features 2:144
 genome organization and expression 2:148–150
 life cycle 2:145
 pathogenesis 2:150
 prevention and treatment 2:145
 virion structure 2:145–148
 bovine papular stomatitis (BPS) 2:668, 2:668f
 bovine papular stomatitis virus (BPSV) 2:666, 2:170–171
 bovine RSV (bRSV) 2:754
 infections 2:748
 bovine viral diarrhea 2:153
 and border disease 2:158–159
 control of 2:163
 and mucosal disease 2:160
 acute BVDV infection 2:160
 fetal infection with BVDV 2:160
 mucosal disease 2:160–161
 persistent BVDV infection 2:160
 vaccines against 2:162
 bovine viral diarrhea virus (BVDV) 1:425
 BPCV1 *see* bandicoot papillomatosis carcinomatosis virus type 1 (BPCV1)
 BPD *see* bronchopulmonary dysplasia (BPD)
 BPS *see* backyard production systems (BPS); bovine papular stomatitis (BPS)
 BPSV *see* bovine papular stomatitis virus (BPSV)
 BPXV *see* buffalopox virus (BPXV)
 BQCV *see* Black queen cell virus (BQCV)
 brachygnathia, defined 2:34
 bracovirus 4:712, 4:849
 Bradford Hill criteria for causation 1:567, 1:568t
 bradykinin (BK) 2:353
 breaching epithelial barriers 1:535–536
Brevihamaparvovirus 4:836
 BREX *see* Bacteriophage Exclusion (BREX)
 bridge amplification 1:177
 bridge vector, defined 2:805
 brivudine ((E)-5-(2-bromovinyl)-2'-deoxyuridine) 5:184
 BRMV *see* Berrimah virus (BRMV)
 Broad bean mottle virus (BBMV) 3:266
 broadly neutralizing antibodies (bNAbs) 1:213
 broadly-neutralizing monoclonal antibodies (bnMAbs) 5:273
 brome mosaic virus (BMV) 1:463, 3:261–262, 3:262–263, 3:263, 3:264f, 3:266, 3:252
 application in biotechnology 3:259
 coat proteins (CP) 3:266
 future perspectives 3:259
 genome organization 3:253–254
 genome structure, gene expression and sequences 3:253–254
 noncoding regions, *cis* signals, and sub-genomic promoter 3:254–255
 host genes 3:257
 inducing/stabilizing spherules 3:257–258
 regulating lipid metabolism and membrane composition 3:257
 regulating translation of viral proteins and 1a targeting 3:258
 required for viral replication activity 3:258
 transmission and host range 3:258–259
 properties and functions of viral gene products 3:256–257
 BMV 1a and viral replication compartments 3:256–257
 replication and propagation 3:255–256
 in vitro and *in vivo* replication studies 3:255–256
 RNA recombination 3:256
 sub-genomic mRNA synthesis 3:256
 virion structure and RNA encapsidation 3:252–253
Bromoviridae 3:371, 3:132, 3:260
 evolution of 3:263f
 genera and species in 3:262t
 phylogeny and biodiversity of 3:261
 bromovirids 3:261
 bromoviruses (*Bromoviridae*)
 biology 3:266–267
 characteristics of RNA genome in 3:261t
 genome organization and expression 3:264–266
 genomic RNA replication and recombination 3:266
 phylogeny and biodiversity of the family *Bromoviridae* 3:261
 virion properties and structure 3:261–264
 bronchiolitis and recurrent wheezing 2:761
 bronchopulmonary dysplasia (BPD) 5:273
 bronchus-associated lymphoid tissue (BALT) 2:76
 Brown Cap Mushroom Disease (BCMD) 4:532
 classification 4:532
 diagnosis 4:533
 epidemiology 4:533
 genome and virion structure 4:532
 viral expression and disease development 4:532–533
 Brownian motor, defined 4:105
 bRSV *see* bovine RSV (bRSV)
 BSD *see* banana streak disease complex (BSD)
 BSGFV *see* banana streak GF virus (BSGFV)
 BSIMV *see* banana streak IM virus (BSIMV)
 BSMYV *see* banana streak MY virus (BSMYV)
 BSOLV *see* banana streak OL virus (BSOLV)
 BTV *see* bluetongue virus (BTV)

- budded virus (BV) 4:704
 defined 4:739, 4:747
 budding, defined 1:345
 budding, virus 1:519–520
 assembly 1:519–520
 cell-to-cell transmission 1:525
 envelopment 1:521
 ESCRT-dependent budding 1:522–523
 intracellular budding 1:523–524
 maturation 1:524–525
 quasi-enveloped viruses 1:524
 buffalopox virus (BPXV) 2:858
 Bunyamwera serogroup 2:662
Bunyavirales 3:495
 bunyavirales glycoproteins 1:348
 bunyavirus 1:499
 of arthropods 4:764
 genome and coding strategies of viral
 genomes 4:764–766
 host associations, virus maintenance
 cycles and pathogenicity 4:767
 viral replication cycle 4:766–767
 virion structure 4:764
 defined 4:764
 buoyant densities in CsCl, defined 4:632
 Bursa of Fabricius, defined 2:182, 2:540
 burst, defined 4:314
 burst assay, defined 4:98
 BV *see* budded virus (BV)
 BVDV *see* bovine viral diarrhea virus
 (BVDV)
 BVF *see* Botrytis virus F (BVF)
 BVX *see* Botrytis Virus X (BVX)
 Bwamba serogroup 2:662
 bymoviruses (*Potyviridae*)
 classification 3:268
 control 3:272
 defined 3:268
 diagnosis 3:271–272
 member species 3:268
 nucleic acid properties and differentiation
 of bymoviruses 3:268–269
 organization of the genome and properties
 of the encoded proteins 3:269–271
 symptoms, epidemiology and host range
 3:271
 virion structure, serological- and
 cytological properties 3:268
 virus transmission and movement 3:271
 bystander effects, defined 5:233
 BYVMV *see* Bhendi yellow vein mosaic virus
 (BYVMV)
- C**
 cabotegravir 5:151
 cacao swollen shoot disease (CSSD) 3:90
 cacao swollen shoot Togo B virus (CSSTBV)
 3:167
 cacao swollen shoot virus 3:274
 control of the disease 3:284
 diagnosis 3:280–283
 PCR diagnosis 3:283–284
 serological diagnostics 3:281–283
 epidemiology: geographical distribution of
 viral molecular diversity 3:278–280
 genome organization 3:277
 host range and symptomatology 3:274–
 276
 phylogeny 3:277–278
 taxonomy 3:274
 transmission 3:276–277
 virion structure 3:277
 cactus-infecting viruses 3:623–624
 Cactus virus X (CVX) 3:623–624, 3:628
 cadherin-related family member 3
 (CDHR3) 2:758–759, 1:410
 CAEV *see* caprine arthritis encephalitis virus
 (CAEV)
Cafeteria roenbergensis 1:85
 CAGE-Seq *see* cap analysis gene expression
 sequencing (CAGE-Seq)
 calcium 4:523
Caliciviridae 1:278–279
 Californian myxoma virus 2:731–732
 California serogroup 2:662–663
Calighravirus 4:715–716
Calliptamus italicus EPV (CIEV) 4:865
 Calmodulin-like proteins 3:245
 calnexin 1:603
Calvusvirinae 3:788
 CAM *see* cell adhesion molecules (CAM)
 CaMV *see* cauliflower mosaic virus (CaMV)
 cancer 2:454
 cancer therapeutics, virus-based
 arming viruses 5:241
 immunomodulators 5:241
 prodrug convertases 5:241
 radiosensitizers for therapy and imaging
 5:241
 combination therapies 5:241–243
 definition 5:233
 historic notes 5:233–234
 re-engineered viruses: the beginnings
 5:234–235
 shielding from host immune response
 5:239–240
 chemical shielding and cell carriers
 5:240–241
 virus engineering 5:239–240
 three classes of modifications 5:235–237
 virus tropism, retargeting 5:237–238
 cell entry targeting by protease
 activation 5:238
 cell entry targeting of enveloped viruses
 5:237–238
 cell entry targeting of non-enveloped
 viruses 5:238
 combination of targeting modalities
 5:239
 post-transcriptional regulation of
 replication 5:239
 preferential tumor spread 5:239
 transcriptional regulation of replication
 5:238–239
 cancer vaccine, defined 1:658
 cancer vaccines and therapeutics, VLPs as
 1:669–670
 canine and other carnivoran adenoviruses
 2:12
Canine circovirus (CanineCV) 2:185
 canine coronavirus (CCoV) 2:256, 2:198
 CanineCV, *see* *Canine circovirus* (CanineCV)
 canine distemper virus (CDV) 2:74
 canine parvovirus (CPV) 2:684
 CPV-2 2:684
 emergence of
 classification (compact) 2:683
 diagnosis 2:687
 epidemiology 2:684–686
 life cycle 2:683–684
 pathogenesis and clinical features
 2:686–687
 prevention 2:687
 treatment 2:687
 virion structure and genome 2:683
 host range variation, its control, and the
 emergence of 2:685–686
Canis familiaris papillomavirus type 1 (CPV-
 1) 2:85
 Canna yellow mottle virus (CaYMV) 3:166
 canonical/cap-dependent mRNA translation
 1:444–445
 elongation and termination phases 1:446
 initiation phase 1:444–445
 initiation factor eIF4E 1:445
 protein kinase R (PKR) 1:444–445
 scanning and assembly of the 80S
 ribosome 1:445–446
 canonical initiation of virus mRNA
 translation 1:446–447
 “imprisonment” of cellular mRNAs within
 the nucleus 1:447
 microRNAs (miRNAs) 1:455
 non-canonical initiation of virus mRNA
 translation 1:447–449
 cap-independent translation enhancers
 (CITEs) 1:449
 initiation at non-AUGs 1:450
 internal ribosome entry sites (IRESes)
 1:447–449
 leaky scanning 1:450
 ribosome reinitiation 1:450–452
 ribosome ‘shunting’ (discontinuous
 scanning) 1:450
 virus alternatives to components of
 initiation 1:449–450
 translational ‘recoding’: non-canonical
 elongation and termination 1:452
 ribosomal bypassing (‘hopping’) 1:452
 ribosomal ‘frame-shifting’ 1:452–454
 ribosome stop codon ‘read-through’
 1:454
 virus-encoded proteinases 1:454–455
 ‘stopgo’/‘stop carry-on’/ribosomal
 ‘skipping’ 1:454
 CAPA *see* corrective actions and preventative
 actions (CAPA)
 cap analysis gene expression sequencing
 (CAGE-Seq) 1:178
 cap-independent translation enhancers
 (CITEs) 1:449
 capping, defined 2:475
 caprine arthritis encephalitis virus (CAEV)
 2:56–57, 2:65, 2:66–67
 capripoxviruses (CPPV) 2:165

- capsid 1:257, 1:495, 2:441, 1:402, 1:362, 4:387, 4:359, 1:382, 4:26, 4:276, 4:98, 2:92, 4:115, 3:545, 4:568, 4:105, 4:167
- capsid architecture, general principles of 1:257–261
- helical symmetry 1:261
- icosahedral symmetry 1:258–261
- capsid assembly 1:263–266, 4:49–50
- enveloped icosahedral RNA viruses 1:266–268
- large icosahedral dsDNA viruses 1:269–271
- non-enveloped icosahedral RNA viruses 1:265–266
- small icosahedral ssDNA viruses 1:268–269
- tailed bacteriophages 1:271–272
- capsid-associated tegument complex (CATC) 1:321, 1:322
- capsid-encoding organisms 1:14
- capsid/head, defined 4:45
- capsid layers, structural organization of 1:306–307
- double-stranded RNA (dsRNA) viruses, cell entry of 1:307–309
- outer capsid layer 1:306–307
- capsidless particles 2:442–443
- capsid or head, defined 1:318
- capsid or phage head, defined 4:136
- capsid protein (CP) 3:163, 4:594, 3:318, 4:821, 1:353, 1:355–356, 2:57
- defined 4:658
- structural folds of 1:262–263
- four-helix bundle 1:263
- HK97 fold 1:263
- immunoglobulin-like fold 1:263
- jelly-roll -barrel 1:262–263
- serine protease fold 1:263
- capsid protein 2:184, 2:884–885
- capsid self-assembly 1:483
- capsid strategy 3:112
- aphid-transmitted cucumoviruses, case of 3:112
- capsid vertex specific complex (CVSC) 2:715
- cap snatching, defined 4:764, 3:719, 2:765
- capsomer, defined 4:115, 3:461, 2:441
- capsomere, defined 1:257
- capsomers, defined 1:248, 2:22
- capsule, defined 1:402
- capulavirus 3:412–413
- CAR *see* coxsackievirus B3-receptor (CAR)
- Carajas virus (CARV) 2:875
- Carcinus mediterraneus reovirus W2 (CMRV) 4:878
- cardoreovirus 4:878–879, 4:878
- carmo-like viruses (*Tombusviridae*) 3:285
- distribution, host range, transmission, and economic significance 3:288–289
- genome structure and protein functions 3:289–290
- replication and gene expression 3:290
- satellites, defective-interfering RNAs 3:290–291
- taxonomy, classification, and evolutionary relationships 3:285–288
- virion structure and assembly 3:289
- virus–host interaction 3:291
- Carmotetraviridae* 4:898, 4:717
- carnivora, defined 2:683
- carnivores, parvoviruses of
- classification (compact) 2:683
- diagnosis 2:687
- epidemiology 2:684–686
- life cycle 2:683–684
- pathogenesis and clinical features 2:686–687
- prevention 2:687
- treatment 2:687
- virion structure and genome 2:683
- carrier state, defined 4:26
- CART *see* combinatorial antiretroviral therapy (cART)
- CARV *see* Carajas virus (CARV)
- case-control study
- and cohort studies 1:564–565
- defined 1:559
- case fatality rate (CRF) 2:212, 2:608
- defined 5:247
- cash crops viral diseases 3:94–95
- ornamentals and orchids viral diseases 3:95–96
- sugarcane viral diseases 3:94–95
- casing, defined 4:549, 4:528
- Caspar Klug theory (CKT) 1:250
- defined 1:248
- caspase, defined 4:724
- Casphalia extranea* 4:839
- cassava brown streak disease (CBSD)
- current status in Africa 3:295
- engineering CBSD resistance 3:299–300
- evolution, diversity and distribution of viruses Involved in 3:297–298
- genome organization and gene functions 3:295–297
- management and control of 3:298
- sources of CBSD resistance and their introgression 3:298–299
- cassava brown streak viruses (*Potyviridae*) 3:293
- aetiology, host range and transmission 3:293–295
- Cassava common mosaic virus (CsCMV) 3:628, 3:628–629
- cassava mosaic geminiviruses (CMGs), resistance to 3:556–558
- cassava mosaic viruses (*Geminiviridae*) 1:649
- control 3:309
- cultural methods 3:310–311
- biosafety considerations 3:311
- virus resistance by engineering 3:310–311
- diagnosis 3:308–309
- economic importance 3:309
- epidemiology 3:306
- distribution of CMGs in Asia 3:306–307
- diversity and distribution 3:306
- distribution of the African CMGs, 21523: s0065
- recombination and the CMD pandemic 3:307–308
- field-level epidemiology 3:308
- genome 3:303
- history 3:301–302
- host plant resistance 3:309–310
- intraspecific diversity 3:301
- pathogenesis and symptoms 3:305–306
- regional epidemiology 3:308
- replication 3:303–304
- taxonomy 3:301
- transmission 3:304–305
- virion structure 3:302–303
- cassava viral diseases 3:83–85
- Cassava virus X (CsVX) 3:623, 3:623–624, 3:628
- CAT1 *see* cationic amino acid transporter 1 (CAT1)
- catalytic RNA, defined 3:852
- CATC *see* capsid-associated tegument complex (CATC)
- catch-all detection methods 5:260–261
- category A-C Priority pathogens, defined 2:765
- cationic amino acid transporter 1 (CAT1) 2:145
- Caudovirales* 1:319, 4:186, 4:186–187
- caudovirales, defined 4:276, 4:342
- cauliflower mosaic virus (CaMV) 3:164, 1:462
- properties and functions of CaMV gene products 3:316–318
- aphid transmission factor (P2) 3:318
- capsid protein (P4) 3:318
- movement protein (MP) 3:317–318
- protease/reverse transcriptase (P5) 3:318
- transactivator/viroplicin (P6) 3:318
- virion-associated protein (P3) 3:318
- caulimoviruses (*Caulimoviridae*) 3:313–314
- genome organization 3:314–316
- replication and propagation 3:318–320
- taxonomy, phylogeny, and evolution 3:314
- transmission, host range 3:320
- virion structure 3:314
- virus–host relationships 3:320–321
- causation, Bradford Hill criteria for 1:567, 1:568t
- CAV *see* chicken anemia virus (CAV); coxsackie A virus (CAV)
- caveolae, defined 1:529
- caveolin-1 (CAV-1) 1:532–533
- caveolin-dependent endocytosis 1:532–533
- CaYMV *see* Canna yellow mottle virus (CaYMV)
- CBP *see* chitin-binding protein (CBP)
- CBSD *see* cassava brown streak disease (CBSD)
- CCAAT box 2:58–59
- C-C chemokine receptor type 5 (CCR5) 5:151, 2:60, 2:827
- CCD *see* colony collapse disorder (CCD)
- CCHBV *see* crane hepatitis B virus (CCHBV)
- CCHFV *see* Crimean-Congo hemorrhagic fever orthobunyavirus (CCHFV)
- CcIRV-5 *see* Ceratitis capitata idnoreovirus-5 (CcIRV-5)
- CCMV *see* cowpea chlorotic mottle virus (CCMV)
- CCoV *see* canine coronavirus (CCoV)

- CCR5 *see* C-C chemokine receptor type 5 (CCR5)
- CCRS *see* Cherry chlorotic rusty spot disease (CCRS)
- CCS *see* closed circular consensus (CCS)
- CD3 1:584
- CD4 2:60, 2:778, 2:463, 2:464, 1:396, 5:139, 1:213, 1:213–215, 1:383
- CD4 T cells 2:229, 1:585, 2:456, 2:825
- CD4⁺ T cells 2:469, 2:470, 2:471, 2:229, 2:392, 1:589, 1:590, 1:595–596, 1:587, 2:65, 2:66, 2:827, 2:56, 1:399–400, 2:14, 1:604
- CD4⁺, defined 2:827
- CD8 2:778
- CD8 T cells 2:229, 2:455, 2:457, 2:825
- CD8⁺ cytotoxic T cells (CTL) 1:589
- CD8⁺ T cells 1:587, 2:229, 1:589, 1:590, 2:392, 1:603
- CD14⁺ lineage-committed granulocyte-macrophage progenitors 2:456
- CD40 2:613–614
- CD28 1:589
- CD34⁺ stem cells 2:456–457, 2:457
- CD46 1:396, 2:778, 1:412
- CD55 1:284
- CD56 1:584
- CD80 2:613–614
- CD81 2:390
- CD86 2:613–614
- CD134 2:60, 2:66, 2:778
- CD155 1:659, 1:410, 2:688
- CD163 2:703
- CD206 2:60
- CDHR3 *see* cadherin-related family member 3 (CDHR3)
- CDR *see* complementarity-determining regions (CDR)
- CDV *see* canine distemper virus (CDV)
- cELISA *see* competitive enzyme-linked immunosorbent assay (cELISA)
- cell adhesion molecules (CAM) 1:394
- cell-associated enveloped virion (CEV) 2:165
- cell-based vaccines 5:302–303
- cell biology phase, of virology 1:6–7
- cell-cycle link protein, defined 3:169
- cell entry targeting
- of enveloped viruses 5:237–238
 - of non-enveloped viruses 5:238
 - by protease activation 5:238
- cell envelope 1:503–504
- cell envelope, defined 1:402
- cell-free assembly (CFA) system 2:131–132
- cell-free protein synthesis (CFPS) 1:665
- cell line, defined 4:888
- cell-mediated immunity (CMI) 2:516, 2:639
- defined 2:79
- cell polarity, defined 1:529
- cell-surface, interaction with 4:389–390
- cell-to-cell movement, defined 3:140
- cell-to-cell movement protein, defined 3:545
- cell-to-cell transmission 1:525
- cell tropism, defined 2:198
- cellular adenosine deaminases 1:55–56
- cellular appendages, interaction with 4:388–389
- cellular cytidine deaminases 1:55
- cellular electron cryotomography 1:238–240
- cellular immune response
- contraction of 1:590
 - VLPs and the induction of 1:662
- cellular “rotting” 1:503
- cell wall, defined 1:402
- central nervous system (CNS) 1:498–499
- infections 5:102
 - assays 5:102
 - background 5:102
 - clinical impact 5:102–103
- CEPI *see* Coalition for Epidemic Preparedness Innovations (CEPI)
- Ceratitis capitata idnoroovirus-5 (CclRV-5) 4:875
- cereal viral diseases 3:83
- Certification Program, defined 3:430
- certified reference materials (CRM) 5:76
- cervidpoxvirus 2:171
- cesium chloride gradients 5:11
- CEV *see* cell-associated enveloped virion (CEV)
- CFA system *see* cell-free assembly (CFA) system
- CfCPV-16 *see* Choristoneura fumiferana cypovirus 16 (CfCPV-16)
- CFPS *see* cell-free protein synthesis (CFPS)
- CFS *see* Chronic Fatigue Syndrome (CFS)
- CFTR *see* cystic fibrosis transmembrane conductance regulator (CFTR)
- Chagas disease 4:771
- chagres virus 2:771
- chain-termination sequencing 5:28
- Chandipura virus (CHPV) 2:875
- chaperones, defined 3:32
- Chaphamaparvovirus 4:836
- chapparvovirus 4:836
- CHB *see* chronic hepatitis B (CHB)
- chemical composition of viruses 2:415
- chemical fixation 1:243
- chemiosmotic theory 4:215
- chemistry, manufacturing, and control (CMC) 4:252, 4:256
- chemokine, defined 2:56
- chemokine binding protein (vCBP) 2:671–672
- chemokine receptor, defined 2:827
- cheraviruses
- classification 3:322
 - clinical features and pathogenesis 3:325
 - diagnosis 3:325
 - genome organization 3:322
 - geographic distribution 3:324
 - life cycle and epidemiology 3:322–324
 - host range 3:324
 - prevention 3:326
 - transmission 3:324–325
 - treatment 3:325–326
 - virion structure 3:322
- cherax quadricarinatus densovirus (CqDV) 4:840
- Cherry chlorotic rusty spot disease (CCRS) 4:646–647
- chestnut mosaic virus (ChMV) 3:165
- CHF *see* Crimean hemorrhagic fever (CHF)
- chicken anemia virus (CAV) 2:48
- Chickpea chlorotic dwarf virus (CpCDV) 3:358–359
- Chikungunya virus (CHIKV) 1:263, 1:546–547, 1:239, 5:275
- classification 2:173
 - clinical features 2:178–180
 - diagnosis, treatment, and prevention 2:180
 - epidemiology 2:178
 - genome organization and features 2:173–174
 - CHIKV 3′ UTR 2:175
 - non-coding elements 2:174–175
 - protein-coding regions 2:174
 - life cycle 2:175–176, 2:175f
 - attachment and entry 2:175–176
 - non-structural protein expression and viral replication 2:176–177
 - structural protein expression and virion assembly 2:177–178
 - pathogenesis 2:180
 - virion structure 2:173
- CHIKV *see* Chikungunya virus (CHIKV)
- ChilLCD *see* chilli leaf curl disease (ChLCD)
- ChilLCV *see* Chilli leaf curl virus (ChILCV)
- chili leaf curl betasatellite (ChLCB) 3:357
- chilli leaf curl disease (ChLCD) 3:756
- Chilli leaf curl India virus 3:750
- Chilli leaf curl virus (ChILCV) 3:756, 3:750
- chilli viral diseases 3:86–87
- Chilo iridescent virus (CIV) 1:269–270
- Chinese hamster ovary (CHO) cells 2:857–858, 1:668, 5:284
- chitin-binding protein (CBP) 4:813
- ChLCB *see* chili leaf curl betasatellite (ChLCB)
- ChMV *see* chestnut mosaic virus (ChMV)
- CHO cells *see* Chinese hamster ovary (CHO) cells
- Chordopoxvirinae 4:858
- chordopoxviruses (ChPVs) 4:860–861
- chordovirus, defined 3:805
- Choristoneura fumiferana cypovirus 16 (CfCPV-16) 4:874
- CHPV *see* Chandipura virus (CHPV)
- ChPVs *see* chordopoxviruses (ChPVs)
- chromatographic methods 1:170–171
- chromosomal integration of HHV-6 (CIHHV-6) 2:778, 2:781, 2:787, 2:784–785
- and potential for disease in immunocompetent 2:784–785
 - and potential for disease in immunocompromised 2:785
- chromoviruses, defined 3:653
- Chronic Fatigue Syndrome (CFS) 5:297–298
- chronic hepatitis B (CHB) 5:217
- management in special populations 5:222
 - immunosuppression 5:222
 - pregnancy 5:222
 - treatment goals in 5:217–218

- chronic hepatitis B and chronic hepatitis D infection
- chronic hepatitis delta 5:222
- future 5:224
- management of 5:217
- nucleo(s)tide analogs 5:218–219
- choosing the appropriate first-line agent 5:219
- finite NA therapy 5:219
- management of patients with NA failure 5:219
- safety 5:218–219
- nucleo(s)tide analogs 5:223–224
- on-treatment prediction of response 5:221
- HBeAg-negative patients 5:221
- HBeAg-positive patients 5:221
- optimizing response rates: treatment
- prolongation and combination therapy 5:221–222
- optimizing response to NA with PEG-IFN 5:222
- pegylated interferon alfa 5:219–220
- efficacy in HBeAg-negative patients 5:220
- efficacy in HBeAg-positive patients 5:220
- safety of PEG-IFN in CHB 5:220
- screening 5:222–223
- selection of patients for PEG-IFN therapy 5:220–221
- HBeAg-negative patients 5:221
- HBeAg-positive patients 5:220–221
- treatment indications 5:217
- and goals 5:223
- treatment options 5:218, 5:223
- interferon alpha (IFN) 5:223
- chronic hepatitis C, defined 2:386
- chronic obstructive pulmonary disease (COPD) 2:762
- chronic persistent infection, defined 2:697
- chronic release, defined 4:314
- chryso-P4 4:564
- chrysoviruses (*Chrysoviridae*) 4:557, 4:505–506, 4:454
- biology and effects of chrysoviruses on fungal hosts 4:565–566
- evolutionary relationships among 4:547–548
- genome expression and replication 4:562–563
- alphachryso-P3 4:564
- chryso-P4 4:564
- chrysovirus CPs 4:563–564
- chrysovirus RdRps 4:563
- replication of chrysoviruses 4:564
- genome organization 4:559–562
- taxonomy and phylogenetic analysis 4:564–565
- virion properties 4:557
- virion structure and composition 4:557–559
- CHV1 *see* cryphonectria hypovirus 1 (CHV1)
- CHV4 *see* cryphonectria hypovirus 4 (CHV4)
- chytrid fungi, defined 3:567
- Cicadulina* spp. 3:461
- cidofovir 2:458–459
- CIEV *see* Calliptamus italicus EPV (CIEV)
- CIHHV-6 *see* chromosomal integration of HHV-6 (CIHHV-6)
- cilevirus 3:247
- diagnosis 3:251
- diseases, epidemiology, and control 3:250–251
- genome organization and gene product function 3:248–249
- members of the family 3:247
- replication and propagation 3:249–250
- taxonomy, phylogeny, and evolution 3:247
- transmission and host range 3:250
- virion structure 3:247–248
- Circoviridae* 4:718
- classification 2:182
- clinical features 2:186–187
- diagnosis 2:190–191
- disease control 2:191
- epidemiology 2:185–186
- genome 2:184
- life cycle 2:184–185
- pathogenesis 2:187–190
- virion structure 2:182–184
- circoviruses 2:186
- circularly permuted, defined 4:276
- circular permutation, defined 4:368
- circulating recombinant forms (CRFs) 1:108
- circulating vaccine-derived poliovirus (cVDPV) 5:310, 2:692–695
- circulative, propagative transmission, defined 3:761
- circulative non-propagative transmission 3:107–108
- aphid-transmitted luteovirids, case of 3:107–108
- leafhopper-transmitted mastreviruses, case of 3:108
- whitefly-transmitted begomoviruses, case of 3:108–110
- circulative propagative transmission 3:110
- circulative transmission, defined 3:169
- circumsporozoite protein (CSP) 1:665
- cis-acting replication element (Cre) 2:757
- cis-acting RNA elements (CREs) 2:387
- CITEs *see* cap-independent translation enhancers (CITEs)
- 3'CITE, defined 3:456
- citopathology, defined 3:788
- citrivirus, defined 3:805
- citrus tristeza virus (CTV) 3:90–91, 3:327–328
- defective RNAs 3:332
- diagnosis 3:334–335
- economic costs of 3:329
- gene functions 3:331–332
- CTV sub-genomic RNAs 3:332
- genome 3:330–331
- untranslated regions (UTRs) 3:331
- geographic distribution 3:328–329
- history 3:328
- host range and cytopathology 3:329–330
- CTV virions 3:330
- taxonomy and classification 3:328–329
- transgenic resistance to CTV infection 3:334–335
- transmission 3:333–334
- CTV control measures 3:334
- virus-virus interactions 3:332–333
- CIV *see* Chilo iridescent virus (CIV)
- cKS *see* classical Kaposi's sarcoma (cKS)
- CKT *see* Caspar Klug theory (CKT)
- cladogram 1:117–118
- classical Kaposi's sarcoma (cKS) 2:606
- classical swine fever (CSF) 2:32, 2:158, 2:159
- acute CSF 2:159
- chronic CSF 2:159
- control of
- in domestic pigs 2:162
- in wild boar 2:162–163
- late onset CSF 2:159–160
- vaccines against 2:161–162
- classical swine fever viruses 2:153
- classification, defined 1:28
- classification, virus 1:33
- taxa and viruses, differentiating 1:33
- Class I fusion glycoproteins 1:420, 1:419f
- Class I common characteristics 1:420–421
- influenza HA 1:420
- Class II fusion glycoproteins 1:421, 1:422f
- Class II common characteristics 1:421–423
- several cellular fusogens have the class II fold 1:423–424
- TBEV E 1:421
- Class III fusion glycoproteins 1:424, 1:423f
- Class III common characteristics 1:424
- VSV G 1:424
- class switch recombination (CSR) 1:591
- clathrin-coated pits, defined 1:529
- clathrin-independent mechanisms 2:176
- clathrin-mediated endocytosis (CME) 2:176
- defined 2:22
- Clavaviridae* 4:361–362
- Claviviridae* 1:369
- CLCAlaV, *see* Cotton leaf curl Alabad virus (CLCAlaV)
- CLCBurA, *see* Cotton leaf curl Burewala alphasatellite (CLCBurA)
- CLCGezA, *see* Cotton leaf curl Gezira alphasatellite (CLCGezA)
- CLCGezB, *see* Cotton leaf curl Gezira betasatellite (CLCGezB)
- CLCGezV, *see* Cotton leaf curl Gezira virus (CLCGezV)
- CLCKokV, *see* Cotton leaf curl Kokhran virus (CLCKokV)
- CLCKokV-Burewala strain (CLCKokV-Bur) 3:357
- CLCMuA, *see* Cotton leaf curl Multan alphasatellite (CLCMuA)
- CLCMuB, *see* Cotton leaf curl Multan betasatellite (CLCMuB)
- CLCuD *see* cotton leaf curl disease (CLCuD)
- cleavage, defined 3:797
- cleavage/packaging, defined 2:441
- CLEM *see* correlative light and electron microscopy (CLEM)

- Clinical and Laboratory Standards Institute (CLSI) 5:55
- clinical diagnostic virology 5:98–99
- central nervous system infections 5:102
- assays 5:102
- background 5:102
- clinical impact 5:102–103
- gastrointestinal infections (GI) 5:101
- assays 5:101–102
- background 5:101
- clinical impact 5:102
- hepatitis E virus (HEV) infection 5:103
- respiratory tract infections (RTI) 5:99–100
- assays 5:100–101
- background 5:99–100
- clinical impact 5:101
- clinical signs, defined 2:697
- clinical trial, defined 5:281
- clinical utility 5:73–74
- clink, defined 3:470
- clonal expansion, defined 2:144
- clonal interference, defined 1:53
- clonality, defined 2:528
- Clonal virus strains 2:453
- cloned recessive resistance genes 3:73–74
- cloning 1:8–9
- closed circular consensus (CCS) 1:181
- closely related viruses independently infecting plants and fungi 4:445
- Closteroviridae* 3:336
- closteroviruses (*Closteroviridae*) 3:336
- classification 3:336–338
- defined 3:336
- diagnosis 3:346
- disease symptoms 3:345
- epidemiology 3:343–345
- genome organization 3:338–341
- infectious cycle 3:341–343
- pathogenesis 3:345–346
- prevention 3:346
- treatment 3:346
- Clover yellow mosaic virus (CIYMV) 3:628
- CLSI *see* Clinical and Laboratory Standards Institute (CLSI)
- clustered regularly interspaced short palindromic repeat (CRISPR) 1:623, 1:624f, 5:93, 5:113, 1:60, 1:633, 4:423, 4:343, 4:242–243
- CRISPR-associated (Cas) genes 1:633
- CRISPR-associated nuclease 9, defined 3:293
- CRISPR-Cas 1:610–613, 4:252, 3:411, 3:52, 3:123
- CRISPR-Cas3 *see* Type I CRISPR-Cas systems
- CRISPR/Cas9 technology 1:498, 3:420 *see also* Type II CRISPR-Cas systems
- CRISPR-Cas immunity, defined 4:419
- CRISPR-Cas system 4:229, 4:397, 4:400
- clustered regularly interspaced short palindromic repeats-CRISPR associated 9 (CRISPR-Cas9) 3:759
- clustered regularly interspaced short palindromic repeats with the enzyme Cas (CRISPR-Cas) 5:113–114
- CIYMV *see* Clover yellow mosaic virus (CIYMV)
- CM *see* cytoplasmic membrane (CM)
- CMC *see* chemistry, manufacturing, and control (CMC)
- CMD *see* covert mortality disease (CMD)
- CME *see* clathrin-mediated endocytosis (CME)
- CMI *see* cell-mediated immunity (CMI)
- CMRV *see* Carcinus mediterraneus reovirus W2 (CMRV)
- CMs *see* convoluted membranes (CMs)
- CMV *see* cucumber mosaic virus (CMV)
- CMV *see* cytomegalovirus (CMV)
- CMVIG *see* cytomegalovirus hyperimmune immunoglobulin (CMVIG)
- CNS *see* central nervous system (CNS)
- Coalition for Epidemic Preparedness Innovations (CEPI) 2:516, 2:361, 5:286
- coat protein (CP) 4:262–263, 4:105, 3:140
- cocal virus (COCV) 2:875
- COCV *see* cocal virus (COCV)
- Codon 129 polymorphism, defined 2:707
- coevolution
- defense mechanisms 1:641–642
- defined 2:3, 2:349
- integration sites 1:641
- in the patient 1:641
- polylysogeny 1:638–640
- prophage remnants 1:640–641
- war and peace 1:642
- cohort study, defined 1:559
- cold/hot tumor, defined 1:658
- cold tumor 1:659
- coliphage N4 virions 4:213–214
- collection of routine data 5:247–248
- colletotrichum camelliae filamentous virus 1 4:483–484
- biological properties 4:484
- family –Unclassified 4:483–484
- genome structure 4:483–484
- genus –Unclassified 4:483–484
- phylogenetic relationships 4:484
- virion morphology 4:484
- colony collapse disorder (CCD) 4:772
- Colorado tick fever virus (CTFV) 1:546
- Columbiformes 2:811
- combination therapy, defined 1:658
- combinatorial antiretroviral therapy (cART) 2:460
- commerce and tulip mania 1:675
- common cold 2:761
- common source transmission 1:561
- defined 1:559
- comoviruses 3:348
- economic importance 3:353
- epidemiology 3:353
- expression of the viral genome 3:351–352
- functions of the viral proteins 3:352
- genome structure 3:351
- geographic distribution 3:352
- host range 3:352–353
- physical properties of viral particles 3:349–350
- relationships with other viruses 3:352
- symptomatology and cytopathology 3:353
- taxonomy and classification 3:348–349
- transmission 3:353
- use in biotechnology 3:352
- viral structure 3:350–351
- competitive enzyme-linked immunosorbent assay (cELISA) 2:134
- complementarity-determining regions (CDR) 1:587, 1:288
- complementary RNA (cRNA) 2:289, 2:553–554
- complement control proteins 1:396–397
- complement fixation 5:20
- complement system, defined 1:577
- completeness 5:247, 5:251–252
- complex regional pain syndrome (CRPS) 5:297–298
- complex viruses, dissection of 1:227–230
- concatamer, defined 2:599, 1:318, 4:368, 4:136, 4:148, 4:61, 2:441
- concatemeric DNA, defined 4:105
- concatemeric replication products 4:137–138
- conformational recognition, defined 4:136
- conformational selection model, defined 2:707
- congenital infection, defined 2:797
- congenital varicella, defined 5:181
- congenital Zika syndrome (CZS) 2:899, 2:906
- defined 2:899
- conidia, defined 4:601, 4:607, 4:648, 4:552
- conidiospore, defined 4:450
- conidium, defined 4:557
- conjugation, defined 4:242, 4:53
- connector/nanopore, defined 4:302
- consensus genome, defined 1:71
- conserved sequence elements (CSEs) 2:174
- defined 2:173
- CONSISE *see* Consortium for the Standardization of Influenza Seroepidemiology (CONSISE)
- consistency 5:252
- Consortium for the Standardization of Influenza Seroepidemiology (CONSISE) 5:260
- contagious pustular dermatitis (CPD) 2:667–668
- contagium vivum fluidum 1:3–4
- contig, defined 5:27
- continuous professional development (CPD) 5:67–68
- continuum mechanics theory 4:206
- Contractile injection system 4:186
- contractile long-tailed phages 4:211
- contractile tail 4:206
- defined 4:45
- contrast transfer function (CTF) information 1:233
- convalescent fatigue syndrome 2:812
- convergent evolution, defined 4:359, 4:457
- convoluted membranes (CMs) 1:498–499, 1:499
- cooperativity 1:427
- co-option, defined 4:732
- CoP *see* correlate of protection (CoP)

- COPD *see* chronic obstructive pulmonary disease (COPD)
- COPI/COPII, defined 3:32
- core gene, defined 2:441, 4:359
- core proteins 1:335
- Coronavirus (CoV) 1:266, 2:193, 1:499, 2:198–199, 2:256–257, 2:264–265
see also COVID-19 pandemic
- accessory proteins 2:203–204
- avian coronavirus IBV 2:195f
- background 2:428–429
- classification 2:429
- CoV-ectoenzyme interactions 1:397
- diagnosis 2:438
- diseases and host range 2:196–197
- epidemiology and clinical features 2:437
- HCoV-229E 2:437
- HCoV-HKU1 2:437–438
- HCoV-NL63 2:437
- HCoV-OC43 2:437
- future perspectives 2:206
- genome 2:430–431
- accessory proteins 2:431–432
- non-structural proteins 2:431
- genome organization and expression 2:194–195
- genome replication and recombination 2:196
- HCoV-host interactions 2:434–435
- apoptosis 2:435–436
- autophagy 2:435
- ER stress response 2:435
- MAP kinase pathway 2:435
- translational control 2:434–435
- life cycle 2:432–433
- assembly and release of virion 2:434
- attachment 2:432–433
- formation of the replication-transcription complex 2:433
- viral entry and uncoating 2:433
- viral RNA synthesis 2:433–434
- manipulating CoV genomes using RNA recombination and reverse genetics 2:204–205
- molecular features of 2:199–201
- pathogenesis 2:436–437
- prevention 2:439
- replication and transcription of CoV RNA 2:201–203
- replication cycle 2:195–196, 2:202f
- taxonomy and classification 2:193
- treatment 2:438–439
- vaccines and antiviral drug development 2:205–206
- virion properties 2:193
- virion structure 2:429–430
- virion structure and composition 2:193–194
- Coronavirus RNAs 1:464
- corpora allata/cardiac, defined 4:780
- corrective actions and preventative actions (CAPA) 5:70
- correlate of protection (CoP) 5:285
- defined 5:289, 5:281
- correlative light and cryo tomography in cells 1:215–216
- correlative light and electron microscopy (CLEM) 1:215, 1:497
- defined 1:208, 1:495
- workflows 5:12
- Corynebacteriales* 1:514
- cos, defined 4:61
- cos-clearance 4:130–131
- nucleotide switch for 4:131
- cosmopolitan viruses 1:20
- co-suppression, defined 3:123
- co-translational disassembly, defined 3:727
- Cotton leaf curl Alabad virus* (CLCAlaV) 3:357
- Cotton leaf curl Burewala alphasatellite* (CLCBurA) 3:360–361
- cotton leaf curl disease (CLCuD) 3:355
- components of CLCuD complex 3:357
- begomoviruses associated with CLCuD 3:357
- CLCuD-associated alphasatellite components 3:359–361
- CLCuD-associated betasatellite 3:361
- engineered resistance to CLCuD complex 3:362
- genome editing strategy 3:363
- history of 3:355–356
- in Africa 3:356
- elucidation of etiology 3:356–357
- etiology 3:356
- in Southern Asia 3:356
- in the Indian sub-continent 3:29
- causal agent and disease classification 3:29–30
- control 3:30–31
- disease symptoms and yield losses 3:29
- epidemiology 3:30
- geographical distribution 3:29
- long term strategies 3:362
- genetics of host plant resistance to CLCuD 3:362
- non-pathogen derived resistance 3:363
- pathogen derived resistance (PDR) 3:362
- antisense RNA technology 3:362
- RNAi technology 3:362–363
- short term strategies 3:361–362
- strategies to control CLCuD complex 3:361
- Cotton leaf curl Gezira alphasatellite* (CLCGezA) 3:357, 3:360–361
- Cotton leaf curl Gezira betasatellite* (CLCGezB) 3:357
- Cotton leaf curl Gezira virus* (CLCGezV) 3:357
- Cotton leaf curl Kokhran virus* (CLCKokV) 3:357, 3:756
- Cotton leaf curl Multan alphasatellite* (CLCMulA) 3:360–361
- Cotton leaf curl Multan betasatellite* (CLCMulB) 3:361, 3:244–245
- cotton leaf curl virus in cotton (CLCuVs), resistance to 3:563–565, 3:565f
- Cotton yellow mosaic virus* (CYMV) 3:357
- Councilman bodies 2:895–896, 2:896
- counterdefense mechanisms 4:398
- counter-evolution by phages to avoid parasitism by satellites 4:102–103
- CoV *see* Coronavirus (CoV)
- covert mortality disease (CMD) 4:822
- COVID-19 pandemic 1:671, 5:262, 1:205, 5:33, 5:115, 5:3, 1:566
see also Coronavirus (CoV)
- COVID-19 vaccines 5:286–287
- cowpea chlorotic mottle virus (CCMV) 1:664, 1:667, 1:488–490, 3:262–263, 3:263–264, 3:264f
- capsids 3:263–264
- cowpea mosaic virus (CPMV) 1:279, 3:364
- biology 3:364
- expression of the viral genome 3:366–367
- functions of the viral proteins 3:367–368
- genome structure 3:366
- physical properties of viral particles 3:364–365
- replication 3:368–369
- taxonomy and classification 3:364
- use in biotechnology 3:369
- viral structure 3:365–366
- cowpox virus (CPXV) 2:854
- coxsackie A virus (CAV) 1:395
- Coxsackievirus A10 (CV-A10) 1:284
- coxsackievirus B (CVB) 5:293
- Coxsackievirus B3 (CVB3) 1:287
- coxsackievirus B3-receptor (CAR) 1:284
- coxsackieviruses A24 (CVA24) 1:392
- CP *see* capsid protein (CP); coat protein (CP)
- CpCDV, *see* Chickpea chlorotic dwarf virus (CpCDV)
- CPD *see* contagious pustular dermatitis (CPD); continuous professional development (CPD)
- CpDV 4:845
- CPE *see* cytopathic effects (CPE)
- CPMV *see* cowpea mosaic virus (CPMV)
- CPPV *see* capripoxviruses (CPPV)
- CPV *see* canine parvovirus (CPV); cytopathic vacuoles (CPVs)
- CPV-1, *see* *Canis familiaris* papillomavirus type 1 (CPV-1)
- CPXV *see* cowpox virus (CPXV)
- CqDV *see* *cherax quadricarinatus* densovirus (CqDV)
- crane hepatitis B virus (CCHBV) 2:100
- crawlers 3:165
- Cre, *see* cis-acting replication element (Cre)
- Crenarchaeota*, viruses of 4:360t
- CREs *see* cis-acting RNA elements (CREs)
- CRF *see* case fatality rate (CRF); circulating recombinant forms (CRFs)
- cricket baculovirus 4:827–828
- Cricket paralysis virus (CrPV) 4:768–770, 4:771, 4:772, 4:773
- Crimean-Congo hemorrhagic fever orthonairovirus (CCHFV) 2:208, 2:209f
- clinical features 2:212–213
- diagnosis 2:214–215
- epidemiology 2:211–212
- human vaccines against 2:216t
- pathogenesis 2:213–214
- prevention 2:215
- adenovirus vaccine 2:216

- Crimean-Congo hemorrhagic fever
 orthonairovirus (CCHFV) (*continued*)
 DNA vaccines 2:215
 inactivated vaccines 2:215–216
 modified vaccinia virus ankara (MVA)
 2:216
 replication cycle 2:210f
 treatment 2:216
 virion structure, genome and life cycle
 2:208–209
 L segment 2:209–210
 M segment 2:209
 S segment 2:208–209
 virus life cycle 2:210–211
- Crimean hemorrhagic fever (CHF) 2:208
- crinivirus, defined 3:336
- cripavirus 4:768, 4:769t, 4:770t, 4:774
- CRISPR *see* clustered regularly interspaced
 short palindromic repeat (CRISPR)
- CRISPR-Cas systems and anti-CRISPR
 proteins 4:244f, 3:554
 anti-CRISPR proteins 4:248–249
 guilt-by association bioinformatic
 approach 4:249–250
 lytic phage dependent functional
 approach 4:250
 prophage-dependent functional
 approach 4:249
 self-targeting bioinformatic approach
 4:250
 diversity and classification of CRISPR-Cas
 4:245–246
 history of CRISPR-Cas 4:244–245
 mechanisms of CRISPR-Cas immunity
 4:246
 adaptation 4:246
 defense 4:246–248
 phage-host arms race 4:242–244
- CRM *see* certified reference materials (CRM)
- cRNA *see* complementary RNA (cRNA)
- cro* gene 4:81
- Crohn's disease 2:454
- cross-kingdom virus infection 4:443
 artificially established viral cross infections
 between plants and fungi 4:445–447
 closely related viruses independently
 infecting plants and fungi 4:445
 contribution to virus evolution 4:448
 fungal DNA virus that replicates in an
 insect 4:447–448
 natural transmission of a plant virus to a
 fungus 4:447
 viruses that replicate in both plants and
 insects 4:443–445
- cross-protection 3:260, 3:586, 3:862, 3:327,
 3:371, 3:439, 3:839
- Croton yellow vein mosaic virus (CYVMV)
 3:756
- CRPS *see* complex regional pain syndrome
 (CRPS)
- CrPV *see* Cricket paralysis virus (CrPV)
- cryoconite hole, defined 4:342
- cryo-electron microscopy (cryo-EM) 1:242,
 5:12–13, 1:490–491, 1:318, 3:420,
 1:329, 1:362, 3:456, 4:900, 4:902,
 4:206, 4:211, 4:218
- cryo-electron microscopy (cryo-EM)
 structures of viruses
 global symmetry, beyond 1:235–236
 envelope glycoproteins 1:236–237
 genome structures 1:237–238
 symmetry mismatch in DNA portals
 1:235–236
 historical background 1:233
 in-situ structural virology 1:238–240
 cellular electron cryotomography
 1:238–240
 near-atomic resolution, cryo-EM at 1:233–
 235
 data collection and image
 reconstruction 1:233–235
 near-atomic resolution maps,
 interpretation of 1:235
 side chain interactions and virus
 assembly 1:235
- cryoelectron microscopy reconstructions
 4:901f
- cryo-electron tomography (cryo-ET) 1:244–
 245, 1:213–215, 4:206, 4:211, 4:218
 of cell lamellae or vitreous sections 1:245–
 246
 of cell periphery 1:244–245
- cryo-EM *see* cryo-electron microscopy (cryo-
 EM); cryogenic electron microscopy
 (cryo-EM)
- cryo-ET *see* cryo-electron tomography (cryo-ET)
- cryogenic electron microscopy (cryo-EM)
 2:182, 4:342
- cryo-immobilization 1:243–244
- cryopeg, defined 4:342
- cryo-preparation techniques 1:496–497
- cryosphere, defined 4:342
- cryphonectria hypovirus 1 (CHV1) 4:431–
 432, 4:433, 4:461, 4:658
- cryphonectria hypovirus 4 (CHV4) 4:433
- Cryphonectria parasitica*, hypovirulence in
 4:472–474
 artificial application of hypovirulence
 4:474–475
 exclusive transmissible hypovirulence
 4:473–474
- cryptotope, defined 3:727
- crispovirus 4:635–636
- crystallography, virus 1:199–200, 4:184
 future outlook 1:206
 Nitty Gritty 1:200–201
 crystal handling and cryoprotection
 1:201
 crystallization and automation 1:201
 data collection and automation 1:201–
 202
 structure solution and model building
 1:202–204
 targets, virus families and predisposition
 to crystallization 1:200–201
 synergies 1:204
 understanding evolution through
 structural anatomy 1:204
 vaccines, anti-virals and therapeutics
 1:204–206
- CsCMV *see* Cassava common mosaic virus
 (CsCMV)
- CSEs *see* conserved sequence elements
 (CSEs)
- CSF *see* classical swine fever (CSF)
- CSP *see* circumsporozoite protein (CSP)
- CSR *see* class switch recombination (CSR)
- CSSD *see* cacao swollen shoot disease
 (CSSD)
- CSSD *see* cacao swollen shoot virus (CSSD)
- CSSTBV *see* cacao swollen shoot Togo B
 virus (CSSTBV)
- CsVX *see* Cassava virus X (CsVX)
- CTCL *see* cutaneous T-cell lymphoma
 (CTCL)
- C terminus, defined 1:345
- CTF information *see* contrast transfer
 function (CTF) information
- CTFV *see* Colorado tick fever virus (CTFV)
- CThTV *see* curvularia thermal tolerance
 virus (CThTV)
- CTL *see* cytotoxic T lymphocytes (CTL)
- CTV *see* citrus tristeza virus (CTV)
- C-type lectin, defined 2:22
- cucumber green mottle mosaic virus 3:9
 causal agent and classification 3:9
 control 3:10
 disease symptoms and yield losses 3:9–10
 epidemiology 3:10
 geographical distribution 3:9
 cucumber mosaic virus (CMV) 3:371, 4:440,
 4:525, 3:266, 3:266–267, 3:267f
 cytopathology 3:373
 epidemiology and control 3:380–382
 genome organization 3:374–375
 geographic distribution 3:372
 host range 3:372
 infection cycle 3:375–377
 cell-to-cell and long-distance movement
 3:377
 replication 3:375–377
 transmission 3:377–379
 properties and functions of gene products
 3:375
 symptoms and diagnosis 3:372–373
 taxonomy and phylogeny 3:371–372
 variation and evolution 3:379–380
 virion structure 3:373–374
- cucumber vein yellowing virus (CVYV)
 3:296–297
- Cucumovirus* 3:260, 3:260–261, 3:261,
 3:265, 3:262t
 characteristics of RNA genome in 3:261t
- cucurbit viral diseases 3:87–88
- Culex* mosquito species 2:805
- Culex nigripalpus* 2:810–811
- Culex p. pipiens* 2:810–811
- Culex p. quinquefasciatus* 2:810–811
- Culex pipiens* 2:810, 2:812
- Culex restuans* 2:810
- Culex salinarius* 2:810
- Culex tarsalis* 2:810, 2:812
- Culex* vectors 2:810
- Culicoides* 1:545
- Culicoides* midges 1:545
- cultivation of animal viruses 2:415
- culture-independent studies 4:378
- curettage, defined 2:629

- curly top disease, defined 3:200
current virus taxonomy (1971–present)
1:29–32
curovirus 3:413
curvularia thermal tolerance virus (CThTV)
4:441
cutaneous T-cell lymphoma (CTCL) 2:419
CV-A10 *see* Coxsackievirus A10 (CV-A10)
CVA24 *see* coxsackieviruses A24 (CVA24)
CVB *see* coxsackievirus B (CVB)
CVB3 *see* Coxsackievirus B3 (CVB3)
cVDPV *see* circulating vaccine-derived
poliovirus (cVDPV)
CVSC *see* capsid vertex specific complex
(CVSC)
CVX *see* Cactus virus X (CVX)
CVYV *see* cucumber vein yellowing virus
(CVYV)
C-X-C chemokine receptor type 4 (CXCR4)
5:151
CXCR4 *see* C-X-C chemokine receptor type
4 (CXCR4)
Cyanobacteria 4:331–332
filamentous cyanobacteria, phages of
4:332
unicellular cyanobacteria, phages of
4:331–332
cyanophages 4:347
cyclic oligoadenylates (cOA) signal
transduction pathway 4:400
Cymbidiummosaic virus (CymMV) 3:628,
3:628–629
CymMV *see* Cymbidiummosaic virus
(CymMV)
CYMV, *see* Cotton yellow mosaic virus (CYMV)
cypoviruses 1:306, 4:868, 4:714–715, 4:870,
4:870–871, 4:871–875, 4:877
cysteine-rich protein, defined 3:642
cystic fibrosis transmembrane conductance
regulator (CFTR) 2:789
cytokines, defined 1:577
cytomegalovirus (CMV) 5:105, 5:190–191
antiviral drug resistance 5:192–193
characteristics and functions of CMV gene
products 2:446–450
classification 2:442
clinical features 2:454–455
congenital disease 2:454–455
immunocompromised host 2:455
CMV inclusion disease (CID) 2:454,
2:454–455, 2:455
diagnosis 2:458
diagnosis of CMV disease and treatment
5:192
epidemiology 2:452–453
patterns of infection 2:453
virus propagation 2:453–454
future perspectives 2:459
genome 2:443–445
evolution 2:445
structure 2:443–445
life cycle 2:445–451
host range 2:445–451
replication steps 2:451–452
pathogenesis 2:455–456
immunity 2:457
modulation of adaptive immune
response to infection 2:458
modulation of cell autonomous
response to infection 2:457–458
modulation of innate immune response
to infection 2:458
persistence and latency 2:456–457
transmission 2:456
pre-transplant risk stratification 5:191
prevention 2:459
prevention strategies of CMV disease after
transplantation 5:191–192
prophylaxis and treatment 2:458–459
virion structure 2:442–443
cytomegalovirus hyperimmune
immunoglobulin (CMVIG) 5:270
cytopathic effects (CPE) 2:337, 2:757,
2:316, 4:776, 4:839, 2:17
cytopathic vacuoles (CPVs) 1:499, 2:173
cytopathology, defined 3:247, 2:441
cytoplasmic male sterility 3:388
cytoplasmic membrane (CM) 4:42
cytoplasmic polyhedrosis virus (CPV) 4:868
see also cypoviruses
cytoskeleton 1:495
cytotoxic CD4⁺ T cells 1:588–589
cytotoxicity 1:590
cytotoxic T-lymphocyte-associated protein 4
(CTLA-4) 1:660
cytotoxic T lymphocytes (CTL) 4:824,
2:304, 2:283, 1:662
CYVMV *see* Croton yellow vein mosaic virus
(CYVMV)
CZS *see* congenital Zika syndrome (CZS)
- D**
- DAA *see* direct-acting antivirals (DAAs)
DAdV-1 *see* duck adenovirus 1 (DAdV-1)
DAF *see* decay-accelerating factor (DAF)
DALY *see* disability-adjusted life year
(DALY)
DAMPs *see* danger-associated molecular
patterns (DAMPs)
danger-associated molecular patterns
(DAMPs) 1:659
danoprevir 5:123
darunavir 5:140–141
database, viral
analysis and visualization capabilities 1:146
analysis and visualization of 3D protein
structures 1:148
genome annotation using VIGOR
1:148–149
HA subtype numbering conversion
1:149
metadata-driven comparative genomics
1:148
phylogenetic tree reconstruction 1:147–
148
sequence annotation 1:146
sequence search and alignments 1:146–
147
virus genotype/clade classification 1:149
workbench 1:149–150
and analytical tools 1:141–142
types of bioinformatics webtools 1:142–
143
types of databases 1:141–142
data curation 1:146
data retrieval 1:146
application programming interfaces
(API) 1:146
search interface 1:146
data summary 1:144–145
influenza research database (IRD) 1:143
significance of 1:141
sources of data 1:143
data aggregated from public data
archives 1:143
derived and predictive data 1:143–144
direct submission of novel data 1:143
usage statistics 1:150
user support 1:150
virus pathogen database and analysis
resource (ViPR) 1:143
DBDs *see* DNA binding domains (DBDs)
DC *see* dendritic cells (DC)
DCLs *see* dicer-like proteins (DCLs)
DC-SIGN *see* dendritic cell-specific ICAM-
grabbing non-integrin (DC-SIGN)
DCV *see* Drosophila C virus (DCV)
DDBJ *see* DNA Data Bank of Japan (DDBJ)
DdDp *see* DNA-dependent DNA-polymerase
(DdDp)
DdRp *see* DNA dependent RNA polymerase
(DdRp)
death certification 5:248
decapping, defined 4:664
decay-accelerating factor (DAF) 1:397,
1:659
decontamination and waste management
5:85–86
deep-sea hydrothermal vents 4:347–348
defined 4:342
deep sequencing, defined 5:27
defective-interfering (DI) viruses
assay for DI particles 1:618–619
biological effects 1:619
cyclic variations of defective interference
1:618
defective interfering versus defective
viruses 1:618
defectiveness 1:617
experimental animals, DI particles in
1:619
future perspectives 1:620
generation of DI genomes 1:617
history 1:617
interference 1:617–618
natural infections, DI particles in 1:619–
620
structure 1:617
defective interfering particles (DIPs) 4:830
defective-interfering RNAs 1:465–467
defined 3:285, 3:839, 1:53
defective RNA, defined 3:327
defense island system associated with
restriction modification (DISARM)
system 1:608–610

- defense systems 1:606
 classification and diversity of 1:606–607
 CRISPR-Cas 1:610–613
 self versus nonself recognition innate immunity systems 1:607–610
 toxins-antitoxins 1:607
 unclassified defense systems 1:613
 genomic organization and evolution of 1:613–615
 association of defense systems with programmed cell death components 1:615
 defense islands and a tight link between defense and mobilome genes 1:613–615
 gene and domain shuffling and sharing 1:615
 degenerate primers, defined 5:27
 Delphi process 1:573
 deltaentomopoxvirus 4:859
 deltapapillomavirus DNA 2:88
 deltapartitivirus 4:636–637
 deltasatellites (*Tolecusatellitidae*) 3:239, 3:245
 classification and nomenclature 3:239–241
 genome organization 3:245–246, 3:243f
 phylogenetic trees for 3:242f
 species of the genera *Deltasatellite* 3:240t–241
 trans-replication of 3:246
 DEmARC (DivErsity pArtitioning by hieRarchical Clustering) 1:101–102
 dendritic cells (DC) 1:584, 2:229, 2:671
 dendritic cell-specific ICAM-grabbing non-integrin (DC-SIGN) 1:399, 2:765
 dengue hemorrhagic fever (DHF) 2:218
 dengue hemorrhagic fever/dengue shock syndrome (DHF/DSS) 2:218
 dengue virus (DENV) 1:498–499, 1:598–599
 clinical features of infection 2:227–228
 DENV-1 2:218, 2:219
 DENV-2 2:218, 2:219
 DENV-3 2:218, 2:219
 DENV-4 2:218, 2:219
 epidemiology 2:223–224
 evolution 2:222–223
 future 2:231
 genetics 2:222
 geographic and seasonal distribution 2:221
 history 2:218–219
 host range and virus propagation 2:221–222
 immune response 2:228–229
 pathogenesis 2:224–227
 pathology and histopathology 2:228
 prevention and control of dengue 2:229–231
 serologic relationships and variability 2:223
 structures 2:220f
 taxonomy and classification 2:219–221
 transmission and tissue tropism 2:224
 vaccines 2:229–231
de novo sequencing, defined 5:27
 dense body
 see also capsidless particles
 defined 2:441
 density ultracentrifugation 1:169–170
 denonucleosis viruses (DNVs) 4:835
 densoviral capsid structures 4:844f
Densovirinae 4:711–713
 densoviruses (DVs) 4:839, 4:838t, 4:835, 4:836
 biochemical properties and purification of 4:840–841
 biophysical features and functions associated with DV capsid 4:843–844
 discovery, taxonomy and evolution of 4:835–837
 genome structure and replication 4:844–845
 shrimp 4:839–840
 fenneropenaeus chinensis
 hepandensovirus (FCHDV) 4:840
 new invertebrate DVs 4:840
 penaeus stylirostris penstyldensovirus 1 (PstDV1) 4:839–840
 DENV *see* dengue virus (DENV)
 deoxynucleotide triphosphate (dNTP) 4:15
 deoxyuridine triphosphatase (dUTPase) 2:25–27
 dermatome, defined 2:860
 design qualification (DQ) 5:69
 detective quantum efficiency (DQE) 1:233–234
 detergents 5:301
 detritus, defined 2:629
 DEV *see* duck enteritis virus (DEV)
 DFE *see* distribution of fitness effects (DFE)
 DHBV *see* duck hepatitis B virus (DHBV)
 DHF *see* dengue hemorrhagic fever (DHF)
 DHF/DSS *see* dengue hemorrhagic fever/dengue shock syndrome (DHF/DSS)
 DHIS *see* drug-induced hypersensitivity syndrome (DHIS)
 DHS *see* drug hypersensitivity syndrome (DHS)
 Diadromus pulchellus idnoreovirus 1 (DpIRV-1) 4:870, 4:875
 diagnosis of viral infection
 clustered regularly interspaced short palindromic repeats with the enzyme Cas (CRISPR-Cas) 5:113–114
 future perspectives 5:115
 matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry 5:114
 microfluidics 5:115
 multiplex and point of care tests (POCT) 5:114–115
 next-generation sequencing (NGS) 5:112–113
 polymerase chain reaction (PCR) and its evolution 5:112
 diagnostic assays, standardization of 5:52
 batch testing 5:58
 commercial and laboratory-developed tests 5:60–61
 external quality assessment/proficiency testing 5:58–59
 international standards in clinical practice 5:61–62
 laboratory accreditation 5:59–60
 physical standards 5:52
 preparation and evaluation of standards 5:52–55
 quality assurance (QA) 5:55–57
 use of controls 5:57–58
 virus isolation 5:61
 written standards and guidelines 5:55
 diagnostic transmission electron microscopy 5:6–9
 dianthovirus (*Tombusviridae*) 3:383
 cell-to-cell and systemic movement 3:386
 host factors in dianthovirus RNA replication 3:385–386
 RNA replication 3:385
 RNA silencing suppression 3:386–387
 taxonomy and classification 3:383
 genome organization 3:383
 virions, transmission, and epidemiology 3:383–385
 gene expression 3:384–385
 diapause, defined 2:805
 Diaporthe RNA virus (DRV) 4:438
 DIC *see* disseminated intravascular coagulation (DIC)
 dicer, defined 3:788, 3:383, 3:123
 dicer-like protein-1 (DCL1) 3:758
 dicer-like proteins (DCLs) 3:43
Dicistroviridae 4:706, 1:279, 4:768
 dicistroviruses 4:768
 biophysical properties 4:768–770
 geographic and strain variation 4:773–774
 host range 4:771–772
 organization of dicistrovirus genome 4:770f
 organization of the dicistrovirus genome 4:770–771
 pathology and transmission 4:772–773
 relationships within the family 4:774
 similarity with other taxa 4:774–775
 taxonomy and classification 4:768
 virus replication and genome expression 4:771
 dicotyledon, defined 3:545
 differentiating infected from vaccinated animals (DIVA) 2:117
 differentiation of naturally infected versus vaccinated animals (DIVA) 2:134
 differentiation state-dependent process, defined 2:441
 diffracted X-ray tracking (DXT) 1:195–197, 1:191
 diffraction Limit, defined 5:5
 digital epidemiology 1:566
 defined 1:559
 digital PCR (dPCR) 5:75–76, 5:76, 5:112
 Digoxigenin-labeled RNA probes 3:629
 Diltiazem 5:169–170

- Dinodnavirus* 4:686
Dinornavirus 4:684
Dinovernavirus 4:869, 4:871, 4:876, 4:877
 diphtheritic infections 2:346
 DIPs *see* defective interfering particles (DIPs)
 diptera 4:867
 dipteran, defined 4:768
 direct-acting antivirals (DAAs) 1:575, 2:515, 2:394, 2:395f
 defined 2:386
 direct contact transmission 1:561
 defined 1:559
 direct "rabies immunoperoxidase test" (dRIT) 2:743
 direct repeat, defined 4:276
 DISA *see* disabled infectious single-animal (DISA)
 disability-adjusted life year (DALY) 1:678
 disabled infectious single-animal (DISA) 2:135
 vaccine 2:20–21
 disabled infectious single-cycle (DISC) vaccines 2:135, 2:20–21
 DISARM system *see* defense island system associated with restriction modification (DISARM) system
 discontinuous transcription, defined 2:198
 DISC vaccines *see* disabled infectious single-cycle (DISC) vaccines
 disease prevention, implications for 1:568
 disease X 5:257
 disseminated intravascular coagulation (DIC) 2:513, 2:239, 2:613–614
 distribution of fitness effects (DFE) 5:228–229, 5:227
 DIVA *see* differentiating infected from vaccinated animals (DIVA); differentiation of naturally infected versus vaccinated animals (DIVA)
 divavirus, defined 3:805
 diversity, defined 4:283
 diversity-generating retroelements 1:55
 DI viruses *see* defective-interfering (DI) viruses
 DJR *see* double jelly roll (DJR)
 DLP *see* double-layered particle (DLP)
 DmIRV-5, *see* *Drosophila melanogaster* idnoreovirus-5 (DmIRV-5)
 DMVs *see* double membrane vesicles (DMVs)
 DNA, defined 2:441
 DNA, persistence of 4:77–78
 extrachromosomal 4:78–79
 integration into the chromosome 4:77–78
 DNA binding domains (DBDs) 4:140–141
 DNA-condensing agents 4:206
 DNA Data Bank of Japan (DDBJ) 5:33
 DNA degradation (DND) 1:610
 DNA-dependent DNA-polymerase (DdDp) 2:62, 2:60
 DNA dependent RNA polymerase (DdRp) 4:813–814
 defined 4:808
 DNA-dependent RNA-polymerase (RdDp) 2:62
 DNAemia, defined 5:197
 DNA-filled head, energetics of 4:167
 intra-capsid DNA
 equilibrium energy and structure of 4:167–169
 metastability of 4:169–170
 mobility of packaged genome controls
 ejection dynamics 4:170–171
 pressure-driven release of viral genome into a host cell 4:171–172
 DNA injection 4:48
 DNA microarrays 5:92
 DNA mycovirus 4:493
 DNA packaging 4:136–137, 4:160, 4:148
 characteristics 4:161–162
 force generation 4:162–163
 packaging speed 4:161–162
 pausing and slipping 4:163
 conformational variation of DNA 4:139f
 defined 4:36, 4:45
 and DNA binding protein 4:18–19
 DNA "crunching" or "compression" translocation model 4:155
 DNA wrapping/bending 4:146
 future studies 4:157–159
 genome replication and selection mechanisms 4:137
 concatemeric replication products 4:137–138
 monomeric replication products 4:137
 "inchworm" translocation model 4:153–154
 "lever" translocation model 4:154–155
 mechanisms 4:163–164
 force generation 4:164–166
 mechanochemical coupling 4:163–164
 motor coordination 4:166
 motor-substrate interactions 4:164
 molecular structure 4:160–161
 Phi29 motors 4:151–152
 protein-DNA binding 4:139–140
 DNA binding domains (DBDs) 4:140–141
 multiple binding events regulate complex assembly 4:141
 sequence specific and non-specific binding 4:140
 shape of DNA 4:138–139
 terminase and Phi29 motors 4:148–151
 terminase motors 4:152–153
 termination of packaging 4:155–157
 viral genomic DNA, recognition of 4:141
 concatemeric DNA and terminase proteins 4:141–142
 monomeric genomes and terminal proteins 4:141
 DNA phosphorotioation (DPT) system 1:610
 DNA polymerase gene 5:185
 DNA portals, symmetry mismatch in 1:235–236
 DNA repair enzymes 4:238
 DNA replication 4:15
 of 29-like viruses 4:66–67
 of SPO1-like viruses 4:63–65
 of SPP1-like viruses 4:65–66
 DNA sequencing 1:178–179
 DNA synthesis, defined 4:61
 DNA vaccines 2:215, 5:286
 double-stranded 1:43–44
 single-stranded 1:43
 DNA viruses
 origin of 1:19–20
 recombination in 1:460–462
 DNA virus evolution, mechanisms of 1:71–72
 co-divergence with hosts as a driver of DNA diversification 1:74
 contributions of DNA virus persistence and chronic infections 1:73–74
 DNA virus genome types, diversity of 1:72
 DNA virus hosts vary from single cells to complex multi-cellular organisms 1:73
 duplication and deletions of genes and genome segments 1:76–77
 fluctuations in tandem repeat copy number as a mechanism of evolution 1:75
 host cell biology and availability of host enzymes constrains virus evolution 1:72–73
 host-virus exchange via horizontal gene transfer and transposable elements 1:77
 in vivo observations of within-host diversity and adaptation of DNA viruses 1:75
 large DNA viruses undergo frequent recombination 1:75–76
 recombination at different frequencies for small DNA virus genomes 1:76
 single nucleotide differences as a measure of evolutionary change 1:74–75
 time frames 1:73
 DND *see* DNA degradation (DND)
 dNTP *see* deoxynucleotide triphosphate (dNTP)
 DNVs *see* denonucleosis viruses (DNVs)
 docosanol 5:124–125
 document control 5:65–66
 dolutegravir 5:149–150
 clinical trials
 dual therapy 5:150
 therapy-naïve with 2 NRTI 5:149–150
 treatment-experienced 5:150
 resistance 5:150
 domestic pigs, control of classical swine fever in 2:162
 dominant allele, defined 3:69, 3:554
 Donnan's effect, defined 4:105
 doravirine 5:136
 double antibody sandwich, defined 3:539
 double jelly roll (DJR) 1:329
 capsid size determination in DJR lineage 1:340–342
 DJR lineage 1:337
 genomes in 1:340
 vertex structures in 1:339
 polintons and the evolutionary pathway of DJR viruses 1:342–343
 variants of DJR fold 1:337–339

- double jelly roll (DJR) (*continued*)
double-layered particle (DLP) 1:309
double-membraned bodies, defined 3:396
double membrane vesicles (DMVs) 1:498–499, 1:499, 2:388–389, 2:390, 2:262, 2:201–202, 2:198
double-stranded DNA (dsDNA) 2:855, 1:385, 1:235, 1:501–502, 1:318, 1:319, 3:158, 1:257, 4:160, 1:19, 1:43–44, 1:386
dsDNA recombineering 4:298, 4:295–296
double-stranded RNA (dsRNA) 1:444–445, 1:385, 4:26, 1:499, 1:492–493, 1:303, 1:310f, 1:312f, 3:555
cell entry of 1:307–309
classification and host range 4:26–27
Cystoviridae family 4:26
dsRNA mycoviruses 4:433–437
recombination and reassortment 4:33–34
applications and intellectual property 4:34
carrier state 4:33–34
reverse genetics and self assembly 4:34
replication cycle 4:31–32
membrane acquisition and host cell lysis 4:33
NC maturation 4:33
PC assembly and packaging 4:33
plasma membrane penetration and NC entrance into the cytoplasm 4:31–32
transcription 4:32–33
structures of 1:304f
virion structure and properties 4:27
architecture of the polymerase complex and the nucleocapsid 4:27–28
genome organization and sequence similarity 4:31
host cell attachment complex 4:30–31
membrane envelope 4:30
P1 protein 4:28
P2 protein 4:28–30
P4 protein 4:30
P7 protein 4:30
P8 protein 4:30
physico-chemical properties 4:27
protein components of the cystovirus portal complex and nucleocapsid 4:28
double-stranded RNA (dsRNA) mycoviruses, structure of 4:504
chrysovirus 4:505–506
evolutionary relationships based on structural comparisons 4:509
future perspectives 4:511–512
megabimaviruses 4:508–509
partitiviruses 4:508
quadriviruses 4:506–508
RdRp and dsRNA organization within mycovirus capsids 4:509–511
totiviruses 4:504–505
double-stranded RNA (dsRNA) virus core and endogenous transcription 1:309
viral innermost capsid layer that encloses the genome 1:309
viral RNA capping 1:311–313
viral RNA-dependent RNA polymerase (RdRp) 1:309–311
double-stranded RNA (dsRNA) virus families 1:314–315
dPCR *see* digital PCR (dPCR)
DpIRV-1 *see* Diadromus pulchellus idnoreovirus 1 (DpIRV-1)
DpIDV *see* dysaphis plantaginea densovirus (DpIDV)
DPT system *see* DNA phosphorotioation (DPT) system
DQ *see* design qualification (DQ)
DQE *see* detective quantum efficiency (DQE)
dRIT *see* direct “rabies immunoperoxidase test” (dRIT)
droplet transmission, defined 1:559
Drosophila C virus (DCV) 4:768, 4:772, 4:773, 1:549
Drosophila melanogaster 4:772
Drosophila melanogaster idnoreovirus-5 (DmIRV-5) 4:875
drug hypersensitivity syndrome (DHS) 2:786
drug-induced hypersensitivity syndrome (DHIS) 2:786
drug repurposing 1:499
drug resistance 1:60
drug susceptibility, defined 5:121
DRV *see* Diaporthe RNA virus (DRV)
dsDNA *see* double-stranded DNA (dsDNA)
dsRNA *see* double-stranded RNA (dsRNA)
duck adenovirus 1 (DAdV-1) 2:13
duck enteritis virus (DEV) 2:113
duck hepatitis B virus (DHBV) 2:100, 2:100–101, 2:102–103
DNA synthesis 2:105f
genome organization 2:103f
duck plague *see* duck viral enteritis (DVE)
duck viral enteritis (DVE) 2:113
Duplodnaviria 1:20
dUTPase *see* deoxyuridine triphosphatase (dUTPase)
Duvénhage lyssavirus (DUVV) 2:738, 2:740–741
DUVV *see* Duvénhage lyssavirus (DUVV)
DVE *see* duck viral enteritis (DVE)
DVs *see* densoviruses (DVs)
DXT *see* diffracted X-ray tracking (DXT)
dye-terminator sequencing 5:28
dynamins, defined 1:529
dysaphis plantaginea densovirus (DpIDV) 4:835, 4:845
dysbiosis, defined 1:552, 4:283
dysplasia, defined 2:316
dystocia, defined 2:34
- E**
E1 glycoprotein 1:266–267
E1 protein 2:497–498
E1^E4 protein 2:498
E2 protein 2:498, 1:266–267
E5 protein 2:498
E6 Associated Protein (E6AP) 2:498
E6 protein 2:498
E7 protein 2:498–499
E8^E2 protein 2:498
early gene, defined 2:441
early taxonomic developments (1886–1971) 1:28–29
Eastern equine encephalomyelitis virus (EEEV) 1:66, 1:263, 2:40, 2:43–44, 2:44, 2:45, 2:46
EAV *see* equine arteritis virus (EAV)
EBHSV *see* European brown hare syndrome virus (EBHSV)
ebola and other hemorrhagic fevers 5:274–275
Ebola hemorrhagic fever (EHF) 2:232
Ebola virus (EBOV) 1:399, 1:401, 2:610, 1:121, 2:232, 2:233t, 1:367f
classification 2:232
clinical features 2:239
early disease 2:239
peak disease 2:239
recovery 2:239
virus persistence 2:239
diagnosis 2:241
epidemiology 2:232–234
genome 2:234–235
genome structure and morphology 2:234f
pathogenesis 2:240f
pathogenesis 2:239–241
prevention 2:232–234
adenovirus 26 (Ad26) and modified vaccinia ankara (MVA) 2:243
vesicular stomatitis virus 2:243
replication cycle 2:236–237, 2:237f
assembly and budding 2:238
immune evasion 2:238–239
primary transcription and translation 2:237–238
replication 2:238
virus attachment and fusion 2:236–237
treatment 2:241–242
monoclonal antibody therapy 2:241–242
small interfering RNAs (siRNAs) 2:242
small-molecule compounds 2:242
virion structure 2:235–236
Ebola virus disease (EVD) 1:591, 2:232
EBOV *see* Ebola virus (EBOV)
EBV *see* Epstein–Barr virus (EBV)
EC50, defined 2:561
ecdysteroids, defined 4:858
Echovirus 30 (E30) 1:284
ECMO *see* extracorporeal membrane oxygenation (ECMO)
ecology and global impacts of viruses 1:621–622
viral community composition, drivers of 1:625
viral community ecology, methods for studying 1:622–623
viral impacts on carbon, nutrient, and biogeochemical cycling 1:623–625
virus-host linkages and ecology 1:623
ECRA vaccines *see* Entry Competent Replication Abortive (ECRA) vaccines
ecthyma contagiosum 2:667–668
ectoenzymes 1:397

- CoV-ectoenzyme interactions 1:397
 ectotherm, defined 2:306
 EDS *see* egg drop syndrome (EDS)
 EEV *see* Eastern equine encephalomyelitis virus (EEEV)
 eel virus American (EVA) 2:325
 eel virus European X (EVEX) 2:325
 EEV *see* extracellular enveloped virion (EEV); external enveloped virions (EEVs)
 EFC *see* entry fusion complex (EFC)
 EFF-1 *see* epithelial fusion failure 1 (EFF-1)
 efficiency of plating (EOP) 1:162
 EFM *see* epifluorescence microscopy (EFM)
 EGCG *see* epigallocatechin-3-gallate (EGCG)
 EGFR *see* epidermal growth factor receptor (EGFR); epithelial growth factor receptor (EGFR)
 egg drop syndrome (EDS) 2:13
 egg proteins 5:301
 E glycoprotein 2:891–892
 egress, defined 2:441
 EGTA *see* ethylene glycol tetraacetic acid (EGTA)
 Egyptian rousette bats (ERBs) 2:609
 EHF *see* Ebola hemorrhagic fever (EHF)
 EHM *see* equine herpesvirus myeloencephalopathy (EHM)
 EIA *see* enzyme immunoassay (EIA)
 EIAV *see* equine infectious anemia virus (EIAV)
 EID *see* emerging infectious diseases (EID)
 eIF4E binding proteins (eIF4E-BPs). 1:445
 eIF4E-BPs. *see* eIF4E binding proteins (eIF4E-BPs).
 eIF4E-mediated recessive resistance
 characteristics of 3:74–76
 molecular mechanisms of 3:76–77
 eIF4E phosphorylation 1:444
 eIFs *see* eukaryotic translation initiation factors (eIFs)
 EIOS *see* Epidemic Intelligence from open sources (EIOS)
 ejection/E protein, defined 4:219
 electron microscopy (EM) 3:629, 5:3, 1:497, 1:498–499
 for viral diagnosis 5:6
 biological EM challenges and sample preparation 5:9–10
 cryo electron microscopy 5:12–13
 diagnostic transmission electron microscopy 5:6–9
 negative staining, general protocol for the study of viruses by 5:10–11
 negative staining for diagnosis of fecal samples 5:11
 scanning electron microscopy, advances in 5:13–14
 traditional processing techniques, limitations of 5:11–12
 Electron Multiplied Charged Coupled Device (EM-CCD) sensor 1:208
 electron tomography (ET) 1:496–497, 1:499
 electron tomography (ET), analysis of viruses in the cellular context by 1:242–243
 cryo-electron tomography (cryo-ET) 1:244–245
 of cell lamellae or vitreous sections 1:245–246
 of cell periphery 1:244–245
 room temperature-electron tomography (RT-ET) 1:244
 sample preparation 1:243
 chemical fixation 1:243
 cryo-immobilization 1:243–244
 electron transport chain (ETC) 1:212
 electrophoretic mobility shift assay (EMSA) 4:392
 elegant-crested tinamou hepatitis B virus (ETHBV) 2:100
 elimination program, defined 2:797
 ELISA *see* enzyme-linked immunosorbent assay (ELISA)
 ELISpots *see* enzyme-linked immunosorbent spots (ELISpots)
 elvitegravir 5:146–149
 clinical trials 5:149
 resistance 5:149
 E lysis 1:515
 EM *see* electron microscopy (EM)
 EMBL Nucleotide Sequence Data Library 5:33
 EMCV *see* Encephalomyocarditis Virus (EMCV)
 emergence, defined 3:8
 emergent viruses 1:569
 Emerging disease, defined 3:411
 emerging infectious diseases (EID) 5:256, 5:257
 emerging viral diseases (EVD) 5:285–286
 challenge of 5:285–286
 emerging viruses, prevention and control of 1:572–574
 diagnosis 1:573–574
 mosquito vectors, control of 1:575
 treatment 1:574–575
 vaccination 1:575
 EMSA *see* electrophoretic mobility shift assay (EMSA)
 ENA *see* enzootic nasal adenocarcinoma (ENA)
 enation, defined 3:200
 encapsidation, defined 2:441
 encapsulation
 defined 4:849
 loading of virus-like particles by 1:667–668
 encephalitis 2:812
 alphaviruses causing *see* alphaviruses causing encephalitis
 defined 2:40, 2:778, 2:404, 2:884, 2:675
 encephalomyelitis, defined 2:34
 Encephalomyocarditis Virus (EMCV) 1:445
 endemic, defined 2:218, 1:569, 5:247
 endemicity, defined 2:654
 endocytosis, defined 1:529, 1:417
 endogenisation 1:79, 1:79–80, 1:80f
 endogenous avian hepadnavirus elements 2:109–111
 endogenous bornavirus-like elements 2:141
 endogenous lentivirus 2:64
 defined 2:56
 endogenous retroviruses (ERVs) 1:628, 1:630–631, 1:627, 2:122, 2:64
 evolutionary impact of 1:81–83
 exaptation of 1:630–631
 ERV-derived non-coding elements 1:631–632
 ERV-derived proteins 1:632
 ERV-encoded envelope glycoproteins 1:630–631
 origin of 1:631f
 endogenous viral element (EVE) 4:792, 3:158, 4:835
 endogenous virus, defined 2:643
 endolysin, defined 4:276
 endoparasitic wasps, defined 4:724
 endophyte, defined 3:388
 endoplasmic reticulum (ER) 2:386–387, 1:268, 2:463, 2:554–555, 1:411–412, 1:444, 1:603, 1:498–499, 1:499, 1:664, 2:399, 3:140
 endoplasmic reticulum-Golgi intermediate compartment (ERGIC) 2:201–202, 2:196, 2:452, 2:441
 endoreduplication, defined 3:200
Endornaviridae 3:388
 endornaviruses (*Endornaviridae*) 3:388
 fungal 3:389
 gene products 3:391–392
 genome organization 3:390–391
 horizontal transmission 3:393
 from other eukaryotic kingdoms 3:389–390
 pathogenesis 3:394
 plant 3:388–389
 regulation of copy number 3:393–394
 replication (RdRp Activity) 3:392–393
 vertical transmission 3:393
 virus-host relationship 3:394–395
 endosomal sorting complexes required for transport (ESCRT) 2:511, 1:282, 3:32, 2:238, 2:879–880, 2:132, 1:529, 1:539
 -dependent budding 1:522–523
 endosomes, defined 3:32
 endothelial cell, defined 1:529
 energy-driven genome packaging in double-stranded DNA and giant viruses 1:493
 energy-independent genome packaging 1:488
 in larger (ss)RNA viruses 1:490–492
 in small single-stranded viruses 1:488–490
 enfuvirtide 5:152
 enhanced RSV disease (ERD) 2:754
 Enhanced Surveillance 5:250
 enhancer, defined 2:56
 EnMBV1 *see* Entoleuca megabirnavirus 1 (EnMBV1)
 enrichment strategies 1:179
 negative selection-based enrichment 1:179–180
 positive selection-based enrichment 1:179
 entecavir (ETV) 5:218
Enterobacteriales, phage cluster-host species relationships with 4:272–273

- Enterobacteriales* tailed phage example 4:269–270
- enterovirus 71 (EV-A71) 1:284, 1:599–600
- enteroviruses (EVs) 1:530, 1:284, 1:392
- classification (compact) 2:245–246
- clinical features 2:250–252
- diagnosis 2:252–253
- epidemiology 2:250
- genome 2:247–248
- life cycle 2:248–250
- pathogenesis 2:252
- prevention 2:254
- treatment 2:253–254
- virion structure 2:246–247
- enthalpy, defined 4:167
- Entoleuca megabarnavirus 1 (EnMBV1) 4:600
- entomobarnavirus 4:706
- classification (compact) 4:776–777
- clinical features 4:778–779
- Dicistroviridae* 4:706
- epidemiology 4:778
- genome 4:777–778
- Hytrosaviridae* 4:706–707
- Iflaviridae* 4:707
- life cycle 4:778
- pathogenesis 4:779
- virion structure 4:777
- entomological warfare 1:647
- entomopathogen, defined 4:892
- Entomopoxvirinae* 4:858, 4:713
- entomopoxviruses (EPVs) 4:858, 4:861, 4:862, 4:865
- entropy, defined 4:167
- entry, virus 1:561
- Entry Competent Replication Abortive (ECRA) vaccines 2:135
- entry fusion complex (EFC) 2:855, 2:22
- entry inhibitors 2:472
- entry receptor, defined 1:388
- ENTV *see* enzootic nasal tumor virus (ENTV)
- Env, defined 3:96
- envelope 1:495, 2:441, 1:257, 1:362, 1:382, 3:507
- envelope (E) protein 2:885
- enveloped icosahedral phages – double-stranded RNA (dsRNA) 4:26
- classification and host range 4:26–27
- Cystoviridae* family 4:26
- recombination and reassortment 4:33–34
- applications and intellectual property 4:34
- carrier state 4:33–34
- reverse genetics and self assembly 4:34
- replication cycle 4:31–32
- membrane acquisition and host cell lysis 4:33
- NC maturation 4:33
- PC assembly and packaging 4:33
- plasma membrane penetration and NC entrance into the cytoplasm 4:31–32
- transcription 4:32–33
- virion structure and properties 4:27
- architecture of the polymerase complex and the nucleocapsid 4:27–28
- genome organization and sequence similarity 4:31
- host cell attachment complex 4:30–31
- membrane envelope 4:30
- P1 protein 4:28
- P2 protein 4:28–30
- P4 protein 4:30
- P7 protein 4:30
- P8 protein 4:30
- physico-chemical properties 4:27
- protein components of the cystovirus portal complex and nucleocapsid 4:28
- enveloped icosahedral RNA viruses 1:266–268
- enveloped particles 1:468
- icosahedral enveloped viruses 1:468–469
- alphavirus assembly and budding 1:470–472
- alphavirus life cycle 1:469
- alphavirus virion structure 1:469–470
- flavivirus assembly and budding 1:473
- flavivirus life cycle 1:472
- flavivirus virion structure 1:472–473
- viral envelope 1:468
- enveloped virion, defined 2:797
- enveloped viruses
- cell entry targeting of 5:237–238
- defined 1:417
- enveloped virus membrane fusion 1:417–418
- fusion machinery, working of 1:425
- common mechanism of fusion 1:425
- cooperativity 1:427
- structural intermediates during conformational changes 1:426–427
- target membrane, interaction with 1:425–426
- specific lipid requirements 1:420
- viral fusion glycoproteins, structure of 1:420
- Class I fusion glycoproteins 1:420, 1:419f
- Class II fusion glycoproteins 1:421, 1:422f
- Class III fusion glycoproteins 1:424, 1:423f
- premature triggering, regulating the conformational change to avoid 1:425
- viral fusion site in the cell 1:418–420
- envelope glycoproteins 1:236–237
- envelopment, defined 2:441
- env* gene 1:358–360, 2:57–58, 2:61t
- Env-like functions in plants 3:102
- enzootic, defined 2:899
- enzootic nasal adenocarcinoma (ENA) 2:581
- enzootic nasal tumor virus (ENTV) 2:575, 2:581–582
- enzyme immunoassay (EIA) 5:5
- enzyme-linked immunosorbent assay (ELISA) 2:144–145, 1:9, 2:341, 5:17–18, 5:5, 2:896, 3:6, 2:743, 5:105–106, 3:322, 3:470, 5:94, 2:651–652, 5:191
- enzyme-linked immunosorbent spots (ELISpots) 5:191
- enzymes of lipid metabolism 1:499
- EOP *see* efficiency of plating (EOP); equipment operating procedures (EOPs)
- Ephemeroviruses 2:878
- epicardium, defined 2:875
- epidemic, defined 3:355, 1:569, 5:247, 1:559
- Epidemic Intelligence from open sources (EIOS) 1:566
- epidemiological studies 1:564–565
- case-control studies and cohort studies 1:564–565
- qualitative studies 1:565
- time series analysis 1:565
- epidemiology, defined 1:559, 2:884
- epidermal growth factor receptor (EGFR) 2:498, 2:390
- epifluorescence microscopy (EFM) 1:622, 1:622f
- epifluorescent microscopy, defined 4:314
- epigallocatechin-3-gallate (EGCG) 4:817
- epipelagic layer, defined 4:322
- episomal element, defined 4:387
- episome, defined 4:827, 2:599, 2:778
- epithelial cells, defined 4:858, 1:529
- epithelial fusion failure 1 (EFF-1) 1:95
- epithelial growth factor receptor (EGFR) 2:450–451
- epizootic, defined 1:542
- epizootic haemorrhagic disease virus 1:545
- E protein 1:421, 2:196
- capsids with large numbers of E proteins distributed throughout the DNA 4:222–223
- enigmas of 4:224–225
- encapsulation of E proteins 4:225–226
- impacts of E proteins on capsid DNA packaging and ejection 4:226
- impacts of E proteins on DNA ejection 4:226–227
- phages with yet to be identified E proteins 4:224–225
- potential of E protein delivery from capsid derived nanocontainers 4:227
- low copy number E proteins with specific locales in the capsid 4:221
- podoviral E proteins 4:221–222
- Epstein Barr nuclear antigen - 1 (EBNA-1) IgG 5:195
- Epstein-Barr virus (EBV) 1:532–533, 1:533, 5:195, 1:178, 2:785, 5:105, 5:107–108, 1:321, 1:620, 1:461
- associated diseases 2:272
- Burkitt's lymphoma 2:273
- diffuse large B-cell lymphomas 2:273
- EBV-associated T Cell and NK cell lymphomas 2:274
- gastric carcinoma and other epithelial tumors 2:275–276
- Hodgkin's lymphoma 2:273–274
- infectious mononucleosis 2:272
- lymphomas in immunosuppressed individuals 2:272–273

- multiple sclerosis 2:272
- nasopharyngeal carcinoma 2:274–275
- attachment 1:396
- classification 2:267
- control and prevention of EBV infection 2:276
- epidemiology and natural history of infection 2:267–270
- future perspectives 2:276
- history 2:267
- immune response 2:271–272
- strain variation 2:270–271
- virion and genome structure 2:267
- EPVs *see* entomopoxviruses (EPVs)
- EQA *see* external quality assessment (EQA)
- equid herpesvirus type 1 (EHV-1) 2:278, 2:278–279, 2:284
- equid herpesvirus type 2 (EHV-2) 2:278, 2:278–279, 2:284
- equid herpesvirus type 3 (EHV-3) 2:278, 2:278–279, 2:284
- equid herpesvirus type 4 (EHV-4) 2:278, 2:278–279
- equid herpesvirus type 5 (EHV-5) 2:278, 2:278–279, 2:284
- equid herpesvirus type 9 (EHV-9) 2:278, 2:278–279
- equine, canine and swine influenza
 - classification 2:287
 - clinical features 2:292
 - diagnosis 2:292
 - epidemiology 2:290–291
 - genome 2:287–288
 - life cycle 2:288–290
 - pathogenesis 2:291–292
 - prevention 2:293
 - treatment 2:292–293
 - virion structure 2:287
- equine adenoviruses 2:12–13
- equine arteritis virus (EAV) 2:256, 2:697, 1:499
 - classification 2:697
 - clinical features 2:702
 - diagnosis 2:705
 - epidemiology 2:701–702
 - genome organization 2:698–700
 - immunity 2:704
 - life cycle 2:700–701
 - nonstructural and structural proteins of 2:699t
 - pathogenesis 2:703–704
 - prevention and control 2:705–706
 - taxonomy of 2:697t
 - virion structure 2:697–698
- equine encephalitis viruses 2:808
- equine herpesviruses 2:278
 - classification 2:278–279
 - clinical features and pathology 2:283–285
 - control 2:285
 - epidemiology 2:282–283
 - future perspectives 2:285
 - genomes 2:279–281
 - latency 2:282
 - life cycle 2:281–282
 - pathogenesis and immune response 2:283
 - virion structure 2:279
- equine herpesvirus myeloencephathy (EHM) 2:283
- equine infectious anemia virus (EIAV) 2:56–57, 2:64, 2:67
- equipment calibration 5:69
- equipment maintenance and monitoring 5:69–70
- equipment operating procedures (EOPs) 5:65
- equipment qualification 5:69
- ER *see* endoplasmic reticulum (ER)
- eragrovirus 3:413
- ER-associated degradation (ERAD), defined 4:534
- ERBs *see* Egyptian rousette bats (ERBs)
- ERCC4-like nuclease, defined 2:22
- ERD *see* enhanced RSV disease (ERD)
- ER exit sites, defined 3:32
- ERF4 *see* ethylene response factor-4 (ERF4)
- ERGIC *see* endoplasmic reticulum-Golgi intermediate compartment (ERGIC)
- Eriocher sinensis reovirus 4:878–879
- errantivirus, defined 3:653
- error catastrophe, defined 1:53
- error catastrophe and lethal mutagenesis of viruses 1:58
- error-prone repair polymerases 1:55
- error-prone replication 1:53–54
 - lack of proofreading 1:54–55
 - polymerase selectivity, factors determining 1:54
 - viral polymerases, intrinsic selectivity of 1:54
- ER stress, defined 2:428
- ERVs *see* endogenous retroviruses (ERVs)
- Escherichia coli* 1:639–640, 1:637f–638, 4:291, 4:128
- ESCRT *see* endosomal sorting complexes required for transport (ESCRT)
- ESRF *see* European Synchrotron Radiation Facility (ESRF)
- ET *see* electron tomography (ET)
- ETC *see* electron transport chain (ETC)
- ETHBV *see* elegant-crested tinamou hepatitis B virus (ETHBV)
- ethylene glycol tetraacetic acid (EGTA) 4:40
- ethylene response factor-4 (ERF4) 3:758
- ethylene-signaling genes 3:758
- etiology, defined 3:200
- ETV *see* entecavir (ETV)
- etymology and evolution of the meaning of virus 1:671–672
- eubiosis, defined 4:283
- eukaryotic translation initiation factors (eIFs) 3:794
 - eIF4E 1:311, 3:293
- eukaryotic viruses 1:552
- Euprosteria elaeasa virus* (EeV) 4:897–898
- European bat 1 lyssavirus (EBLV-1) 2:738, 2:740
- European bat 2 lyssavirus (EBLV-2) 2:738, 2:740
- European brown hare syndrome virus (EBHSV)
 - classification 2:724
 - clinical features 2:726
 - diagnosis 2:728
- epidemiology 2:726
- genome 2:725–726
- pathogenesis and host immunity 2:726–728
- prevention and control 2:728–729
- treatment 2:729
- virion structure 2:724–725
- European hedgehog (*Erinaceus europaeus*) 2:896
- European rabbits, defined 2:730
- European Synchrotron Radiation Facility (ESRF) 1:202
- euryarchaeal viruses 4:368–370
 - culture-independent studies 4:378
 - icosahedral internal membrane-containing viruses 4:374–376
 - icosahedral tailed viruses 4:370–373
 - icosahedral tailed viruses of halophilic euryarchaea 4:371–373
 - tailed viruses of methanogenic euryarchaea 4:373–374
- pleomorphic, spindle-shaped, and spherical viruses 4:376
- pleomorphic viruses 4:376
- spherical virus MetSV 4:377–378
- spindle-shaped viruses 4:376–377
- EV *see* extracellular virions (EV)
- EVA *see* eel virus American (EVA)
- EVD *see* Ebola virus disease (EVD); emerging viral diseases (EVD)
- EVE *see* endogenous viral element (EVE)
- EVE-derived immunity (EDI), defined 4:835
- EVEX *see* eel virus European X (EVEX)
- evidenced based medicine, defined 5:98
- evolutionarily informed treatment strategies 5:230
- evolutionary arms race 2:65
- evolutionary lineage, defined 4:457
- evolution steered by structure
 - architectural and functional similarities with cellular structures 1:96–98
 - evolution of viruses 1:87–88
 - structure-based virus classification 1:90–92
 - inferring viral evolutionary relationship through 1:92–94
 - viral envelope glycoproteins 1:94–96
 - viral evolutionary relationships from sequence analysis 1:88–90
- evolvability, defined 1:62
- EVs *see* enteroviruses (EVs); extracellular vesicles (EVs)
- exanthem, defined 2:868
- exaptation, defined 1:627
- exclusion, defined 4:276
- exocyst, defined 1:529
- exocytosis, defined 1:529
- exogenous lentivirus, defined 2:56
- exogenous retrovirus, defined 1:627
- exogenous virus, defined 2:643
- exonuclease, defined 4:291
- exophthalmia, defined 2:324
- exosome, defined 1:529, 2:362
- exosome-like release of quasi-enveloped viruses 1:538–539
- experimental immunoglobulin therapies 5:273–274

avian influenza 5:274
 seasonal influenza 5:273–274
 extant viral families, calibrating the long-term evolutionary history of 1:81
 extensible tail 4:206
 external enveloped virions (EEVs) 4:858
 external quality assessment (EQA) 5:76f, 5:72, 5:76, 5:64–65, 5:55
 design and objectives 5:76–78
 and performance criteria 5:76
 extracellular enveloped virion (EEV) 2:165
 extracellular vesicle-mediated export of viruses 1:538
 autophagy-related vesicle-mediated release of virus 1:539
 exosome-like release of quasi-enveloped viruses 1:538–539
 nonlytic release of virus in extracellular vesicles 1:538
 extracellular vesicles (EVs) 1:538, 1:16–17
 extracellular virions (EV) 2:857
 extracorporeal membrane oxygenation (ECMO) 5:171–172
 extrafollicular B cells 1:591
 extra small virus (XSV) 4:716
 extreme environments
 acidic hot springs, viruses of 4:422
 ecology and evolution shaping archaeal viruses 4:424–425
 host-virus interaction, environment impact on 4:420–422
 host virus-interactions, complexity of 4:422–424
 viral fitness, aspects of 4:419–420
 virus interactions impacting ecology and evolution 4:425–426
 extreme environments, ecology of phages in 4:342–343
 atmosphere 4:354–355
 hypersaline environments, phages in 4:343–345
 polar and other cold environments, phages in 4:349–352, 4:354
 glaciers 4:352–353
 permafrost 4:353–354
 polar lakes 4:353
 polar oceans 4:352
 sea ice 4:353
 thermal environments, phages in 4:345–347
 deep-sea hydrothermal vents 4:347–348
 hot deserts 4:348–349
 terrestrial hot springs 4:345–347
 extremophile, defined 4:342
 extremophilic bacteria, defined 4:53
 extrinsic incubation period 2:893, 2:218, 2:805
 Eyach virus (EYAV) 1:546
 EYAV *see* Eyach virus (EYAV)

F

fabaviruses 3:348
 economic importance 3:353
 epidemiology 3:353
 expression of the viral genome 3:351–352
 functions of the viral proteins 3:352
 genome structure 3:351
 geographic distribution 3:352
 host range 3:352–353
 physical properties of viral particles 3:349–350
 relationships with other viruses 3:352
 symptomatology and cytopathology 3:353
 taxonomy and classification 3:348–349
 transmission 3:353
 use in biotechnology 3:352
 viral structure 3:350–351
 FACS *see* fluorescence-activated cell sorting (FACS)
 FAIDS *see* feline acquired immune deficiency syndrome (FAIDS)
 FAMA *see* fluorescence antibody membrane antigen test (FAMA)
 famciclovir 5:184, 5:175
 farnesyl pyrophosphate synthetase (FPPS) 2:150
 fas, defined 2:875
 FAT *see* fluorescent antibody test (FAT)
 fatty acid synthase 2:391
 Faustovirus 1:270–271
 Favipiravir (Avigan) 5:168
 FAVN test *see* fluorescent antibody virus neutralisation (FAVN) test
 F-box, defined 3:594
 FCCS *see* Fluorescence Cross-Correlation Spectroscopy (FCCS)
 FcHDV *see* fenneroopenaeus chinensis hepadensovirus (FcHDV)
 FCS *see* Fluorescence Correlation Spectroscopy (FCS)
 FCV *see* feline calicivirus (FCV)
 FD *see* fusion domain (FD)
 febrile fits or convulsions, defined 2:778
 febrile headache 2:812
 febrile status epilepticus, defined 2:778
 fecal microbial transplantation or transplant (FMT) 4:283
 fecundity, defined 4:314
 FEGSEM *see* field emission scanning electron microscope (FEGSEM)
 feline acquired immune deficiency syndrome (FAIDS) 2:66
 feline calicivirus (FCV) 1:450–452, 2:256
 classification 2:294
 control and prevention 2:298
 disease control 2:298
 vaccination 2:298
 epidemiology 2:297
 life cycle 2:296–297
 pathogenesis 2:297–298
 virus characteristics 2:294–296
 genome organisation and viral proteins 2:294–296
 non-structural proteins 2:296
 feline immunodeficiency virus (FIV) 2:301, 2:56–57, 2:66
 feline infectious peritonitis virus (FIPV) 2:198
 feline leukemia virus (FeLV) 1:631
 clinical features and pathology 2:302
 envelope gene variation and pathogenicity 2:302–303
 epidemiology 2:301
 FeLV oncogenesis 2:303–304
 FeLV sequences (FeLV) 2:301–302
 FeLV subgroups and host range 2:301–302
 future perspectives 2:305
 geographic distribution 2:301
 history 2:300
 immune response 2:304
 prevention and control 2:304–305
 replication 2:300–301
 taxonomy and classification 2:300
 transmission 2:304
 virion and genome structure 2:300
 feline morbillivirus (FmoPV) 2:68
 feline panleukopenia virus (FPV) 2:684
 FeLV *see* feline leukemia virus (FeLV)
 fenneroopenaeus chinensis hepadensovirus (FcHDV) 4:839, 4:840
 Feral, defined 2:730
 F glycoprotein 2:749–750
 FHB, *see* *Fusarium* head blight disease (FHB)
 FHV *see* Flock House Virus (FHV)
 fiber, defined 00057:p0050
 FIB milling *see* focused ion beam (FIB) milling
 fibroblast cultures 2:453
 fibroma viruses 2:730
 Californian myxoma virus 2:731–732
 classification 2:730
 clinical disease in European rabbits 2:732–733
 diagnosis 2:734–735
 epidemiology 2:731
 genome 2:730
 hare fibroma virus 2:736
 life cycle 2:730–731
 myxomatosis in European rabbits 2:732
 myxoma virus 2:731
 pathogenesis of myxomatosis in European rabbits 2:733–734
 prevention of myxomatosis 2:735
 rabbit fibroma virus 2:735–736
 squirrel fibroma virus 2:736
 treatment 2:735
 virion structure 2:730
 field emission scanning electron microscope (FEGSEM) 5:13–14, 5:7f
 fujivirus 4:870, 4:871, 3:546–547, 3:551
 filamentous, defined 4:53
 filamentous bacteriophage 1:365–366
 filamentous DNA viruses 1:407
 filamentous phage 4:77
 biotechnological applications of 4:59–60
 structure of 4:55–56
 filamentous viruses 4:363–364
 Tokiviricetes 4:363–364
 filamentous viruses 4:484–485
 filiformism, defined 3:371
 filiform symptom, defined 3:862
 filopodia, defined 2:22
Filoviridae family classification 2:233t
 filoviruses 2:235
Fimoviridae 3:396

- fimoviruses (*Fimoviridae*) 3:396
 classification 3:396
 diagnosis 3:403
 epidemiology 3:400–402
 genome 3:398–400
 life cycle (replication 3:400
 symptomatology and economic impact 3:402–403
 treatment and prevention 3:403
 virion structure 3:396–398
- finfish, defined 2:324
- fingerlings, defined 2:544
- FIPV *see* feline infectious peritonitis virus (FIPV)
- fIPV *see* fractional doses of inactivated poliovirus vaccine (fIPV)
- first generation sequencing technology 5:28–29
- first sequencing methods 1:175–176
- FI-RSV vaccine-enhanced disease *see* formalin-inactivated RSV (FI-RSV) vaccine-enhanced disease
- FISH *see* fluorescent in situ hybridization (FISH)
- fish alloherpesviruses
 classification 2:306–307
 clinical features 2:311
 diagnosis 2:313
 epidemiology 2:309–310
 genome 2:307–309
 life cycle 2:309
 management and treatment 2:313–314
 pathogenesis 2:310–311
 prevention 2:314–315
 virion structure 2:307
- fish retroviruses
 isolation and sequencing of retroviruses from WDS and WEH lesions 2:318
 isolation of a retrovirus from salmon swimbladder sarcoma 2:320–321
 phylogeny of 2:322–323
 retroviruses and their life cycle 2:316–317
 salmon swimbladder sarcoma-associated retrovirus 2:320
 salmon swimbladder sarcoma virus (SSSV)
 pathogenesis of 2:322
 prevalence, seasonality, and transmission of 2:321–322
 sequence comparisons of SSSV with other retroviruses 2:321
 seasonal cycle of disease, control of 2:319
 skin tumors in walleye and their associated retroviruses 2:317–318
 snakehead retrovirus 2:319–320
 walleye dermal sarcoma virus (WDSV)
 accessory proteins 2:318–319
 zebrafish endogenous retrovirus 2:322
- fish rhabdoviruses
 classification 2:324–325
 clinical features 2:328–329
 diagnosis 2:329–330
 epidemiology 2:328
 genetic diversity 2:326–327
 genome 2:326
 life cycle 2:327–328
 pathogenesis and immune response 2:329
- prevention 2:330
 virus structure 2:325–326
- fitness, defined 1:62, 1:53
- fitness landscape, defined 1:53
- FIV *see* feline immunodeficiency virus (FIV)
- fivivirus, defined 3:805
- flaviviruses 1:498–499, 2:810–811, 2:891, 2:891–892
 assembly and budding 1:473, 1:473f
 life cycle 1:472
- flavivirus particle organization 1:290–292, 1:291f, 1:293f
 acid sensitivity 1:293–295
 flavivirus precursor polyprotein and the derived structural proteins 1:290–292
 icosahedral surface lattices,
 thermodynamic transition between 1:297–299
 immature flavivirus particle 1:292–293
 mature particle organization 1:295–296
 physical principles of 1:296–297
- flavivirus virion structure 1:472–473
- Flavobacterium*-infecting, lipid-containing phage (FLiP) 1:159–160
- Flavobacterium psychrophilum* 4:354
- flexible, filamentous plant viruses 1:364–365
- FLIM *see* Fluorescence Lifetime Imaging Microscopy (FLIM)
- FLiP, *see* *Flavobacterium*-infecting, lipid-containing phage (FLiP)
- Flock House Virus (FHV) 1:409–410, 1:499
- flock house virus 1:549
- Floodwater *Aedes* spp. 2:765
- florid, defined 2:629
- florida clade 1 (FC1) viruses 2:291
- florida clade 2 (FC2) viruses 2:291
- flow field-flow fractionation methods 1:171–172
- Flt3L *see* Fms-like tyrosine kinase 3 ligand (Flt3L)
- fludase 5:168
- fluorescence-activated cell sorting (FACS) 1:179
- fluorescence-activated virus sorting 1:187
 DNA decontamination and fluorescence-activated virus sorting preparation 1:187
 preparation of 384-well plates for sorting 1:187
 sorting 1:187
- fluorescence antibody membrane antigen test (FAMA) 5:182
- Fluorescence Correlation Spectroscopy (FCS) 1:208, 1:211
- Fluorescence Cross-Correlation Spectroscopy (FCCS) 1:208
- Fluorescence Lifetime Imaging Microscopy (FLIM) 1:211, 1:208
- fluorescent antibody test (FAT) 2:743
- fluorescent antibody virus neutralisation (FAVN) test 2:743
- fluorescent in situ hybridization (FISH) 2:781, 2:782f, 1:498
- fluorescent protein (FP) 1:209, 1:208
- fluoro- (or fluorescence) immunoassay (FIA) 5:17–18
- FMD *see* foot-and-mouth disease (FMD)
- FMDV *see* foot-and-mouth disease virus (FMDV)
- FmoPV *see* feline morbillivirus (FmoPV)
- Fms-like tyrosine kinase 3 ligand (Flt3L) 1:604
- FMT *see* fecal microbial transplantation or transplant (FMT)
- FNV *see* French neurotropic vaccine (FNV)
- focused ion beam (FIB) milling 1:239–240
- focus reduction neutralisation test (FRNT) 2:641
- fomite, defined 1:559, 2:92
- foodborne transmission 2:400
- foot-and-mouth disease (FMD) 1:5
 foot-and-mouth disease virus (FMDV) 1:647, 1:279–280, 1:284–285, 1:285, 2:414, 1:199, 1:202, 1:205, 4:659
 classification 2:332
 diagnosis 2:341
 epidemiology 2:338
 future perspectives 2:341–342
 genetics and evolution 2:339
 genome 2:334–336
 leader protein 2:336
 P1 region 2:336
 P2 region 2:336
 P3 region 2:336
 protein products 2:335–336
 history 2:332
 host range 2:338–339
 immune response 2:340–341
 life cycle 2:336–337
 cell attachment and entry 2:337
 post-translation processing 2:337–338
 RNA replication 2:337
 translation 2:337
 virus assembly and release 2:338
 pathogenesis and clinical features 2:340
 prevention and control 2:341
 serological relationships and variability 2:339–340
 virion structure 2:332–334
- formaldehyde 5:301
- formalin-inactivated RSV (FI-RSV) vaccine-enhanced disease 2:754
- forma specialis, defined 3:528
- Forster Resonance Energy Transfer (FRET) 1:211, 1:208
- forward genetics, defined 3:123
- foscarnet 5:184, 2:458–459
- fosmid, defined 4:322, 4:414
- fossilized or endogenized viral genome, defined 1:71
- fostemsavir 5:152–153
- Foundation Block, defined 3:430
- founder effect, defined 2:778
- four-helix bundle 1:263, 1:329
- 454 (pyrosequencing) 1:176–177
- fowl adeno virus 4 (FAdV-4) 2:13
- fowlpox virus (FWPV) 2:343
 classification 2:343
 control and treatment 2:347
 use of APV as recombinant vaccine vectors 2:347

- fowlpox virus (FWPV) (*continued*)
 epidemiology 2:346
 genome 2:344–345
 life cycle 2:345–346
 pathogenesis 2:346–347
 host range 2:346–347
 immunomodulation 2:347
 virion structure 2:343–344
 FP *see* fluorescent protein (FP)
 FpMyV2 *see* fusarium poae mycovirus 2
 (FpMyV2)
 FPPS *see* farnesyl pyrophosphate synthetase
 (FPPS)
 F protein, defined 2:747
 FPV *see* feline panleukopenia virus (FPV)
 fractional doses of inactivated poliovirus
 vaccine (fIPV)
 frameshift, defined 3:788, 3:383
 frameshifting, defined 3:778
 Freeze Etching 5:5
 Freeze Fracture 5:5
 French neurotropic vaccine (FNV) 2:897
 French viscerotropic virus (FVV) 2:897
 Frenkel method 2:341
 FRET *see* Forster Resonance Energy Transfer
 (FRET)
 FRNT *see* focus reduction neutralisation test
 (FRNT)
 frog and fish adenoviruses 2:14
 fruit bats (flying foxes) 2:358
 fruitbody, defined 4:528
 fruit virus diseases 3:89–90
 banana viral diseases 3:90
 cacao swollen shoot disease (CSSD) 3:90
 citrus tristeza virus (CTV) 3:90–91
 papaya ring spot virus 3:91–92
 plum pox virus (Sharka) disease of stone
 fruits 3:92
 fry, defined 2:544
 fucosyltransferase-2 (FUT2) gene 5:291
 fulminant hepatitis, defined 2:397
 functional genomics, defined 3:123
 fungal DNA virus
 see also DNA mycovirus
 replicating in insect 4:447–448
 fungal endornaviruses 3:389
 fungal gene expression, effects of
 mycoreoviruses on 4:612–613
 fungal partitiviruses (*Partitiviridae*) 4:568
 classification 4:568–569
 epidemiology 4:572–573
 genome 4:569–570
 life cycle 4:570–572
 pathogenesis 4:573–574
 taxonomic and phylogenetic
 considerations 4:574–576
 virion structure 4:569
 fungal resistance in plants 4:515–516
 fungal viruses 4:431, 4:434t–436
 biological properties 4:431–432
 host range 4:431–432
 mixed infections 4:433
 symptom expression 4:432
 transmission 4:432–433
 future perspectives 4:439–440
 cross-kingdom infection by viruses and
 viroids 4:439–440
 host defense against mycoviruses and
 their counter-defense 4:440–441
 mycovirus 4:441–442
 plant-fungal mutualistic associations,
 role of mycoviruses in 4:441
 virus neo-lifestyles 4:440
 yeast as a (model) host to study viral
 replication 4:440
 recent technical advances in fungal
 virology 4:439
 replication and gene expression strategy
 4:438–439
 taxonomy 4:433–437
 dsRNA mycoviruses 4:433–437
 single-stranded (ss) RNA mycoviruses
 4:437–438
 unassigned ssRNA viruses 4:438
 fungi 4:544
 fungi and protists, plant viruses
 transmission by 3:113
 furovirus, defined 3:405
 classification 3:405
 control 3:409
 diagnosis 3:408–409
 nucleic acid properties and differentiation
 of 3:405–408
 organization of the genome and
 properties of the encoded proteins
 3:406–408
 similarities and dissimilarities with other
 taxa 3:409
 symptoms, epidemiology and host range
 3:408
 virion structure 3:405
 virus transmission and movement 3:408
Fusarium graminearum, hypovirulence in
 4:477
Fusarium graminearum negative-stranded
 RNA virus 1 4:483
 biological properties 4:483
 family –*Mymonaviridae* 4:483
 genome structure 4:483
 genus –*Unclassified* 4:483
 particle morphology 4:483
 phylogenetic relationships 4:483
Fusarium head blight disease (FHB) 4:441
Fusarium poae mycovirus 2 (FpMyV2)
 4:660
 fusariviruses 4:577
 future prospective 4:580
 gene expression 4:579
 genome organization 4:578–579
 taxonomy and classification 4:577
 virion property 4:577–578
 virus–host interactions 4:579–580
 virus transmissions 4:579
Fuselloviridae 1:369–370, 4:362, 4:365
 fusion domain (FD) 1:424
 fusion machinery, working of 1:425
 common mechanism of fusion 1:425
 cooperativity 1:427
 structural intermediates during
 conformational changes 1:426–427
 elongated trimeric intermediate for class
 I and class II fusion glycoproteins
 1:426–427
 monomeric intermediates for
 vesiculovirus G 1:427
 target membrane, interaction with 1:425–
 426
 fusion peptide or loop, defined 1:417
 fusion pore, defined 1:417
 fusion protein 2:70–71
 defined 1:382
 fusogen, defined 1:87
 fusolin, defined 4:858
 FUT2 gene *see* fucosyltransferase-2 (FUT2)
 gene
 FVV *see* French viscerotropic virus (FVV)
 FWPV *see* fowlpox virus (FWPV)
- ## G
- GACVS *see* Global Advisory Committee for
 Vaccine Safety (GACVS)
 Gag (group-specific antigen), defined 3:96
 gag gene 2:57–58, 2:59, 2:61t
 gag-pol gene 2:61t
 Gag-Pol polyprotein 1:455
 Gag proteins 1:354
 gag proteins and retroviral virion structure
 1:352–354
 GAGs *see* glycosaminoglycans (GAGs)
 GaHV1 *see* gallid herpesvirus type 1
 (GaHV1)
 GaHV2 *see* Gallid herpesvirus 2 (GaHV2)
 Gallid herpesvirus 2 (GaHV2) 2:113
 gallid herpesvirus type 1 (GaHV1) 2:112
 gamma activation site (GAS) 2:65
 gammaentomopoxvirus 4:859
Gammalexiviridae 4:437–438
Gammaherpesvirinae 1:321
 gammaherpesviruses 2:441, 2:442, 2:445
 defined 2:441
 gammalepovirus, defined 4:380
Gammaproteobacteria 4:335–336
 pseudalteromonas phages 4:336
 vibriophages 4:335–336
 ganciclovir 2:454, 2:458–459
 Gannoruwa bat lyssavirus (GBLV) 2:738
 GAS *see* gamma activation site (GAS)
 gastroenteritis 5:200
 gastrointestinal infections (GI) 5:101
 assays 5:101–102
 background 5:101
 clinical impact 5:102
 gastrointestinal tract virome 1:553–555
 GAV *see* gill-associated virus (GAV)
 gazelle herpesvirus 1 *see* equid herpesvirus
 type 9 (EHV-9)
 GBLV *see* Gannoruwa bat lyssavirus (GBLV)
 GBS *see* Guillain-Barré syndrome (GBS)
 GB virus C (GBV-C) 1:556
 GCRV *see* grass carp reovirus (GCRV)
 GCV *see* Giardia canis virus (GCV)
 GDarSLA, *see* *Gossypium darwinii* symptomless
 alphasatellite (GDarSLA)

- GDGT *see* glycerol dibiphytanyl glycerol tetraether (GDGT)
- GDGT lipids *see* glycerol dialkyl glycerol tetraether (GDGT) lipids
- gelatin 5:302
- geminale virion, defined 3:411
- Geminiviridae* 3:461, 3:768, 3:21, 3:301, 3:244–245
- geminiviruses (*Geminiviridae*) 3:411–412, 3:21
- African cassava mosaic disease 3:24–26
- classification of cassava mosaic begomoviruses 3:27
- control 3:28–29
- disease symptoms and yield losses 3:26–27
- epidemiology 3:27
- geographical distribution 3:24–26
- recent outbreak of CMD in South East Asia 3:27–28
- control 3:418–419
- cotton leaf curl disease in the indian sub-continent 3:29
- causal agent and disease classification 3:29–30
- control 3:30–31
- disease symptoms and yield losses 3:29
- epidemiology 3:30
- geographical distribution 3:29
- defined 3:21
- diagnosis 3:415–417
- DNA satellites associated with 3:415
- genome and viral proteins 3:415
- life cycle 3:417–418
- pathogenicity 3:418
- taxonomy, phylogeny, and evolution 3:411–412
- becurtovirus 3:411–412
- begomovirus 3:412
- capulavirus 3:412–413
- curtovirus 3:413
- eragrovirus 3:413
- grablovirus 3:413
- mastrevirus 3:413
- topocuvirus 3:413
- turncurtovirus 3:413
- tomato leaf curl new delhi disease 3:23–24
- component capturing and exchange of helper components 3:24
- control 3:24
- disease symptoms and yield losses 3:24
- epidemiology 3:24
- geographical distribution 3:23–24
- tomato yellow leaf curl disease 3:21–22
- causal agent and taxonomy 3:22–23
- control 3:23
- disease symptoms, host range and yield losses 3:22
- epidemiology 3:23
- geographical distribution 3:21–22
- virion structure 3:413–415
- geminiviruses, plant resistance to 3:555, 3:557*t*
- in beans 3:561–563
- beet curly top virus (BCTV), resistance to 3:560–561
- cassava mosaic geminiviruses (CMGs), resistance to 3:556–558
- to cotton leaf curl virus (CLCuVs) in cotton 3:563–565, 3:565*f*
- to maize streak virus (MSV) 3:565–566, 3:565*f*
- to mungbean yellow mosaic viruses (MYMIV) 3:563, 3:564*f*
- natural resistance to viruses in plants 3:555–556
- in tomato 3:558–559
- characterization of *Ty* genes 3:559–560
- tomato yellow leaf curl virus (TYLCV) resistance 3:559, 3:560
- geminivirus infection 3:750, 3:753, 3:758
- gemonoviridae, defined 4:493
- GenBank 3:245
- defined 4:45
- gene end (GE) sequence 2:752
- gene expression
- defined 1:382
- host interactions and regulation of 4:48
- classes of genes 4:48
- early genes 4:48–49
- late genes 4:49
- microarrays 1:498
- gene-for-gene resistance, defined 3:60
- general practitioner (GP) surveillance system 5:248
- gene silencing, defined 3:293, 3:520, 3:116, 3:594
- gene silencing, viral suppressors of 3:116
- antiviral gene silencing pathway 3:116–117
- assays to identify and characterize gene silencing suppressors 3:120–121
- effect of gene silencing suppressors on plant development 3:121–122
- evolution of viral suppressors of gene silencing 3:121
- mechanisms of 3:119–120
- cellular transcription, activation of 3:120
- DNA methylation, interference with 3:120
- homeostasis of gene silencing components 3:120
- inhibition of siRNA biogenesis 3:120
- RNA-induced silencing complexes, inhibition of assembly/activity of 3:120
- silencing amplification, inhibition of 3:120
- small RNA binding 3:119–120
- silencing suppressors mediate viral synergism 3:122
- small interfering RNAs direct antiviral immunity 3:117–118
- viral suppressors in support of biotechnology 3:122
- viral suppressors of gene silencing need cellular factors 3:121
- viral suppressors of gene silencing promote infection 3:118–119
- gene-silencing pathways 3:43
- antiviral RNA silencing (av-RNAi), view of applications of 3:49–51
- antiviral RNA silencing 3:45–47
- antiviral silencing in crops 3:47–49
- diverse gene-silencing pathways in plants 3:43–45
- viral and plant RNA-silencing suppressors 3:47
- viral siRNA-mediated antiviral RNAi in mammalian cells 3:49
- gene therapy and cancer therapy 1:676
- genetic barrier to resistance, defined 5:121
- genetic clade, defined 2:117
- genetic complementation, defined 1:53
- genetic engineering, defined 4:291
- Genetic engineering technique 3:759
- genetic fusion 1:665–666
- genetic hitch-hiking, defined 1:53
- genetic organization and virion proteins 2:575–576
- genetic parasite 1:606
- Geneva Protocol (1925) 1:645
- genital and urinary tract viromes 1:556
- genital herpes, treatment options for 5:178*t*
- genitourinary diseases 5:200
- GenitoUrinary Medicine (GUM) clinics, defined 5:247
- genogroup, defined 5:289
- genome, defined 4:175, 4:276, 3:327, 2:441, 2:747
- genome activation 3:260
- defined 3:439
- genome editing, defined 3:293, 3:69
- genome editing strategy 3:363
- genome integration 4:391–392
- genome maturation, defined 4:136
- genome maturation complex 4:128–129, 4:130
- cos*-cleavage reaction 4:129–130
- lambda cohesive end site and assembly of 4:128–129
- model for the assembly of 4:129
- genome mosaicism, defined 4:265
- genome packaging 1:488
- defined 1:318
- double-stranded RNA viruses 1:492–493
- energy-driven genome packaging in double-stranded DNA and giant viruses 1:493
- energy-independent genome packaging in larger (ss)RNA viruses 1:490–492
- in small single-stranded viruses 1:488–490
- genome packaging complex 4:130
- assembly of the packaging motor 4:130
- cos*-clearance 4:130–131
- nucleotide switch for 4:131
- genome promoter (GP) 2:69–70
- Genome Relationships Applied to Virus Taxonomy (GRAViTy) 1:102
- genome replication 4:390–391
- defined 1:382
- genomes and genomics 4:50
- chromosome diversity and replication 4:50
- common ancestry 4:51
- common themes in genome structure 4:50–51

- genomes and genomics (*continued*)
 diversity in genome size and organization 4:50
 horizontal exchange of genes is widespread 4:51
 genome structure 4:601–602, 1:237–238
 genomic architecture, defined 1:62
 genomic precursor, defined 4:26
 genomics-based taxonomy of viruses, emergence of 1:50–51
 genomic segment, defined 4:26
Genomoviridae 4:718
 genotype, defined 1:62, 5:289, 2:789, 2:22
 genotyping 5:186
 geodesic polyhedra 1:250
 geometric mean titers (GMTs) 5:296
 germicides 5:205
 germline insertions 1:630–631
 endogenous retroviruses (ERVs), exaptation of 1:630–631
 ERV-derived non-coding elements 1:631–632
 ERV-derived proteins 1:632
 ERV-encoded envelope glycoproteins 1:630–631
 GE sequence *see* gene end (GE) sequence
 GFP *see* green fluorescent protein (GFP)
gfp gene, defined 4:858
 G glycoprotein 2:750–752
 GI *see* gastrointestinal infections (GI)
 giant viruses 1:493, 1:372
 diversity and distribution of 1:373–375
 genomics and proteomics of 1:376–377
 life style of giants 1:375–376
 origin and evolution 1:377–379
 origin of 1:20–22
 parasites of viruses 1:379–381
 serendipitous discovery of 1:372–373
Giardia canis virus (GCV) 4:582, 4:583, 4:584
Giardia intestinalis/duodenalis 4:582
Giardia lamblia 4:582
Giardia lamblia virus (GLV) 4:582, 4:582–583, 4:584, 4:585, 4:586
 giardiaviruses
 history 4:582
 host range and geographic distribution 4:582–583
 infection and replication 4:586–587
 organization and molecular biology of GLV and GCV dsRNA genomes 4:585–586
 physical and biochemical characteristics 4:583–585
 taxonomy, classification and evolution 4:582
 gill-associated virus (GAV) 2:256
 Gillespie-type algorithms, defined 1:248
 gingivostomatitis 5:175, 2:404
 GISAID *see* Global Initiative on Sharing All Influenza Data (GISAID)
 GISRS *see* Global Influenza Surveillance and Response System (GISRS)
 glaciers 4:352–353
 glecaprevir 5:123
 glioblastoma multiforme 2:454
 Global Advisory Committee for Vaccine Safety (GACVS) 5:297–298
 Global Influenza Surveillance and Response System (GISRS) 2:559, 2:561
 Global Initiative on Sharing All Influenza Data (GISAID) 1:141–142, 5:262
 Global Polio Eradication Initiative (GPEI) 5:310
 global polio eradication program 2:690–692
 Global Polio Laboratory Network (GPLN) 5:311
Globuloviridae 4:362
Glossinavirus 4:781
 glutamic protease, defined 3:692
 GLV, *see* *Giardia lamblia* virus (GLV)
 glycan receptors 1:390–391
 enteroviruses (EVs) 1:392
 glycan-based antiviral strategies 1:393–394
 glycosaminoglycan receptors 1:393
 histo-blood group antigen receptors (HBGAs) 1:392–393
 noroviruses 1:392–393
 rotaviruses 1:393
 influenza viruses 1:391–392
 polyomaviruses (PyVs) 1:392
 sialic acid receptors 1:390–391
 glycerol dialkyl glycerol tetraether (GDGT) lipids 1:368–369
 glycerol dibiphytanyl glycerol tetraether (GDGT) 4:364
 glycoprotein (GP) 2:612, 3:575, 1:417
 glycoprotein precursor (GPC) 2:507, 2:508
 glycosaminoglycan receptors 1:393
 glycosaminoglycans (GAGs) 1:390, 2:475
 glycosylation, defined 4:380
Glyphodes pyloalis 4:839
 GMTs *see* geometric mean titers (GMTs)
 gnotobiotic mice, defined 4:283
 Goatpox virus (GTPV) 2:165, 2:167
 Gokushovirus genomes 4:18
 Goldberg polyhedra 1:250
 Goldberg polyhedral, defined 1:248
 “gold-standard” procedure 1:234–235
 Golgi apparatus 2:185
 Gomphocerus sibiricus EPV (GSEV) 4:865
 gonad-specific virus (GSV) 4:829
 gonotrophic cycle, defined 2:805
 gooseberry vein-banding associated virus (GVBaV) 3:165
Gossypium darwinii symptomless alphasatellite (GDarSLA) 3:360–361
Gossypium hirsutum 3:355, 3:356f, 3:563–564
 Gouléako virus 4:764
 GP *see* genome promoter (GP); glycoprotein (GP)
 GPC *see* glycoprotein precursor (GPC)
 GPCRs *see* G-protein-coupled receptors (GPCRs)
 GPEI *see* Global Polio Eradication Initiative (GPEI)
 GPI anchor, defined 3:32
 GPLN *see* Global Polio Laboratory Network (GPLN)
 G protein, defined 2:747
 G-protein-coupled receptors (GPCRs) 4:527, 2:167
 GpSGHV transmission dynamics in the tsetse fly 4:788–789
 GP surveillance system *see* general practitioner (GP) surveillance system
 grablovirus 3:413
 graft-versus-host disease, defined 2:778
 gram-negative bacteria 1:513
 bacteriophage receptors in 4:179t
 gram-positive hosts, MGL in 1:513–514
 granular cells (GCs) 4:816
 granule, defined 4:739
 granulin, defined 4:739
 granulocyte-macrophage colony stimulating factor (GM-CSF) 1:659, 5:235
 granulovirus (GV) 4:739, 4:747, 4:699
 grapevine pinot gris virus 3:16
 causal agent and classification 3:16
 control 3:16–17
 disease symptoms and yield losses 3:16
 epidemiology 3:16
 geographical distribution 3:16
 grapevine roditis leaf discoloration-associated virus 3:15
 causal agent and classification 3:15
 control 3:16
 disease symptoms and yield losses 3:15
 epidemiology 3:15–16
 geographical distribution 3:15
 grapevines viruses 3:14–15
 grapevine pinot gris virus 3:16
 causal agent and classification 3:16
 control 3:16–17
 disease symptoms and yield losses 3:16
 epidemiology 3:16
 geographical distribution 3:16
 grapevine roditis leaf discoloration-associated virus 3:15
 causal agent and classification 3:15
 control 3:16
 disease symptoms and yield losses 3:15
 epidemiology 3:15–16
 geographical distribution 3:15
 ornamental viruses: rose rosette virus 3:19
 causal agent and classification 3:19
 control 3:20
 disease symptoms and yield losses 3:19
 epidemiology 3:19–20
 geographical distribution 3:19
 staple crops: cassava brown streak disease 3:17
 causal agent and classification 3:17
 control 3:18–19
 disease symptoms and yield losses 3:17–18
 epidemiology 3:18
 geographical distribution 3:17
 grapevine vein clearing virus (GVCV) 3:165
 graphical user interface (GUI) 1:111
 grass carp reovirus (GCRV) 1:306
 GRAViTy *see* Genome Relationships Applied to Virus Taxonomy (GRAViTy)
 grazoprevir 5:123
 greater virus world and its evolution 1:38–40

baltimore classes, evolutionary status of 1:44
 double-stranded DNA viruses 1:43–44
 evolutionary relationships within and between baltimore classes 1:40–42
 reverse-transcribing viruses 1:42–43
 RNA viruses 1:40–42
 single-stranded DNA (ssDNA) viruses 1:43
 virus hallmark genes 1:39–40
 green fluorescent protein (GFP) 1:209–210, 1:208
 groundnut rosette virus (GRV) 3:795
 GRV *see* groundnut rosette virus (GRV)
Gryllus bimaculatus 4:827–828
 GSEV *see* Gomphocerus sibiricus EPV (GSEV)
 GSV *see* gonad-specific virus (GSV)
 GTPV *see* Goatpox virus (GTPV)
 guanyl transferase, defined 2:22
 GUI *see* graphical user interface (GUI)
 Guillain-Barré syndrome (GBS) 5:305, 2:899
 guilt-by association bioinformatic approach 4:249–250
Guttaviridae 4:362
 GV *see* granulovirus (GV)
 GVBaV *see* gooseberry vein-banding associated virus (GVBaV)
 GVCV *see* grapevine vein clearing virus (GVCV)
 gyrate lattice, defined 1:248

H

HA *see* hemagglutinin (HA)
 HAART *see* highly active antiretroviral therapy (HAART)
 HA assay *see* hemagglutination (HA) assay
 HAdV-A 2:11
 HAdV-B1 2:11
 HAdV-B2 2:11
 HAdV-C 2:11
 HAdV-C5 *see* human adenovirus type 5 (HAdV-C5)
 HAdV-D 2:11
 HAdV-E 2:11
 HAdV-F 2:11–12
 HAdV-G 2:12
 HAdVs *see* human adenoviruses (HAdVs)
 hAE *see* human airway epithelium (hAE)
 hAE-ALI, defined 2:419
 haematophagous arthropods, defined 1:542
 HAI assay *see* hemagglutination inhibition (HAI) assay
 half-life, defined 2:697
 haloarchaea, defined 4:380
 haloarchaeal pleomorphic viruses (HRPVs) 4:380–381
Haloarcula californiae icosahedral virus 1 (HHIV-1) 1:337–339
Haloarcula hispanica icosahedral virus 2 (HHIV-2) 1:337–339
Haloarcula hispanica pleomorphic virus 1 (HHPV-1) 1:434–435

Haloarcula sinaiensis tailed virus 1 (HSTV-1) 1:323
 halophile 4:342
 halophilic, defined 4:368
 haloquadratum walsbyi, viruses of 4:416–417
Haloarubrum pleomorphic virus 1 (HRPV-1) 1:434–435
halorubrum pleomorphic virus 6 (HRPV-6) 4:390
Haloarubrum sodomense tailed virus 2 (HSTV-2) 1:252–253
 HAM1-like sequence, defined 3:293
Hamaparvovirinae 4:711–713
 Hamiltonian Paths Analysis (HPA) 1:249, 1:248
 Hamiltonian Paths Approach (HPA) 1:254
 applications of 1:254–255
 assembly code embedded within viral genetic message 1:254–255
 icosahedral symmetry, beyond 1:255
 virus assembly mechanisms 1:255
 hammerhead structure, defined 3:852
 hand, foot-and-mouth disease (HFMD) 1:284
 hantavirus, defined 2:349
 hantavirus cardiopulmonary syndrome (HCPS) 2:349
 hantaviruses
 clinical picture and pathogenesis 2:351–353
 diagnostics and prevention 2:353
 ecology and epidemiology 2:350–351
 historical introduction 2:349
 life cycle of 2:351f
 virology 2:349–350
 haplotype, defined 4:601, 4:607
 harnessing of viruses by humans 1:675
 bioweapons 1:676–677
 commerce and tulip mania 1:675
 gene therapy and cancer therapy 1:676
 phage therapy 1:676
 vaccines 1:675–676
 HaSV *see* helicoverpa armigera stunt virus (HaSV)
 HAV *see* hepatitis A virus (HAV)
 HBcAg *see* hepatitis B virus core antigen (HBcAg)
 HBcAg *see* hepatitis B e antigen (HBcAg)
 HBE cultures *see* human bronchial epithelial (HBE) cultures
 HBGAs *see* histo-blood group antigens (HBGAs)
 HBIG *see* Hepatitis B hyperimmune IG (HBIG)
 HBsAg *see* hepatitis B surface antigen (HBsAg)
 HBsAg loss, defined 5:217
 HBV *see* hepatitis B virus (HBV)
 HBVDb *see* Hepatitis B Virus Database (HBVDb)
Hbz gene 2:535–536
 HCC *see* hepatocellular carcinoma (HCC)
 HCMVCMV *see* human cytomegalovirus (HCMV)
 see also cytomegalovirus (CMV)

HCoV *see* human coronavirus (HCoV)
 HCPS *see* hantavirus cardiopulmonary syndrome (HCPS)
 HCRSV *see* hibiscus chlorotic ringspot virus (HCRSV)
 HCs *see* hyaline cells (HCs)
 HCV *see* hepatitis C virus (HCV)
 HDFs *see* host dependency factors (HDFs)
 HDL *see* high-density lipoproteins (HDL)
 HDV *see* hepatitis delta virus (HDV)
 HE *see* hemagglutinin-esterase (HE)
 head, dealing with 4:219–221
 capsids with large numbers of co-localized E proteins 4:223–224
 capsids with large numbers of E proteins distributed throughout the DNA 4:222–223
 E proteins, enigmas of 4:224–225
 encapsulation of E proteins 4:225–226
 impacts of E proteins on capsid DNA packaging and ejection 4:226
 impacts of E proteins on DNA ejection 4:226–227
 phages with yet to be identified E proteins 4:224–225
 potential of E protein delivery from capsid derived nanocontainers 4:227
 tailed phage capsids with low numbers of E proteins in specific locales 4:220–221
 low copy number E proteins with specific locales in the capsid 4:221
 podoviral E proteins 4:221–222
 headful packaging, defined 4:61, 4:368
 headful termination, defined 4:136
 head-tail connector, defined 4:105
 heartland banyangvirus (HRTV) 2:775–776
 heating, ventilation, and air conditioning (HVAC) systems 5:69
 heat therapy, defined 3:430
 HEF protein *see* hemagglutinin-esterase-fusion (HEF) protein
 helical and filamentous phages 4:53
 biotechnological applications of filamentous phage 4:59–60
 biotopes 4:53–55
 infection process 4:56
 M13 (fd, f1), biosynthesis of 4:56
 assembly and secretion 4:58–59
 coat proteins 4:56
 major coat protein gp8 4:56–57
 minor coat proteins gp3 and gp6 4:57–58
 minor coat proteins gp7 and gp9 4:57
 replication by the rolling circle mechanism 4:56
 structure of filamentous phages 4:55–56
 helical symmetry 1:261, 1:362, 3:229
 helical viruses, structure of 1:362, 1:364t
Archaea, *Rudiviridae* and *Lipothrixviridae*, viruses of 1:367–369
 atomic structure 1:475–476
Claviviridae 1:369
 filamentous bacteriophage 1:365–366
 flexible, filamentous plant viruses 1:364–365

- helical viruses, structure of (*continued*)
 hemorrhagic filoviridae 1:366–367
 rod-shaped helical plant viruses 1:362–364
 spindle-shaped viruses 1:369–370
- helicase-primase, defined 2:441
- helicoverpa armigera stunt virus (HaSV) 4:897, 4:901–902
- Heliothis zea nudivir* 4:827–828
- heliolthsi zea nudivir 1 (HzNV-1) 4:710
- 4-Helix bundle 1:362
- 6-helix bundle (6HB) 2:750
- Helminthosporium victoriae* 4:432
 hypovirulence in 4:476–477
- helper component, defined 3:703, 3:106
- Helper component proteinase, defined 3:293
- helper phage, defined 4:98
- helper strategy 3:112
 aphid-transmitted caulimoviruses, case of 3:112
 aphid-transmitted potyviruses, case of 3:112
- hemagglutination (HA) assay 5:93–94
- hemagglutination inhibition 5:20
- hemagglutination inhibition (HAI) assay 5:93–94
- hemagglutination inhibition (HI) test 2:292
- hemagglutinin (HA) 1:570–571, 1:420, 5:301, 2:565–566, 1:668
- hemagglutinin-esterase (HE) 2:262–263
- hemagglutinin-esterase-fusion (HEF) protein 2:561
- hematopoietic stem cell transplantation (HSCT) 2:453, 2:455, 5:105, 2:779, 2:778
- hemifusion 2:289, 1:417
- hemipteran, defined 4:768
- hemocoel, defined 3:176, 3:447
- hemocyte, defined 4:849
- hemogram, defined 2:899, 3:176, 3:200, 3:447, 3:106
- hemorrhagic fever with renal syndrome (HFRS) 2:349, 2:350, 2:350–351
- hemorrhagic filoviridae 1:366–367
- Hendra virus (HeV) 2:355
- henipaviruses 2:355–356
 classification 2:355
 clinical features 2:358–359
 diagnosis 2:360
 epidemiology 2:357–358
 genome 2:356
 life cycle 2:356–357
 pathogenesis 2:359–360
 prevention 2:361
 treatment 2:360–361
 virion structure 2:355–356
- Hepadnaviridae* 2:373
 classification 2:373–374
 clinical features 2:381
 acute HBV infection 2:381
 chronic HBV infection 2:381
 course of chronic infection 2:381–382
 mutant viruses and chronic infection 2:382
 diagnosis 2:383
- epidemiology 2:379
- genome 2:375–376
 core and polymerase 2:376
 surface proteins 2:375–376
 transcription and translation 2:375–376
 X protein 2:376
- genotypes 2:379
- life cycle 2:376
 attachment 2:376
 egress 2:378–379
 encapsidation 2:377–378
 maturation 2:378
 penetration and uncoating 2:376
 transcription/translation 2:376–377
- pathogenesis 2:382–383
- prevention 2:384
- subtypes 2:379
- treatment 2:383–384
 resistance 2:384
- variants 2:379–381
- virion structure 2:374–375
- Hepanhamaparvovirus* 4:836
- heparan sulfate 1:285, 2:751
- heparan sulfate proteoglycans (HSPG) 2:390, 2:399
- hepatitis A 5:268
- hepatitis A virus (HAV) 5:206–207, 5:213, 2:397, 1:530, 5:268, 1:282
 cell culture and growth characteristics 2:366
 clinical features and pathology 2:366–367
 common course of infection 2:367
 extrahepatic manifestations 2:368
 fulminant hepatitis A 2:368
 pathology 2:368–369
 prolonged hepatitis A 2:367
 relapsing hepatitis A 2:367–368
 clinical presentation and course 5:209
 diagnosis 2:370–371
 antigen detection 5:210
 nucleic acid detection 5:210–211
 serological tests 5:211–212
- epidemiology 2:371
 intermediate endemicity pattern 5:207
 low endemicity pattern 5:207
 very high and high endemicity pattern 5:206–207
- genome organization and expression, replication, morphogenesis 2:363–366
- history 2:362
- host range, transmission and tissue tropism 2:366
- immunoglobulin 5:213
- innate and adaptive immune response 2:369–370
 adaptive immune response 2:370
 innate immune response 2:369–370
- morphology and physicochemical properties 2:363
- prevention and control 2:371
- taxonomy 5:206
- taxonomy and evolutionary origin 2:362–363
- transmission 5:208
 other modes of transmission 5:208
- person-to-person transmission 5:208
- water- and food-borne transmission 5:208
- treatment 5:213
- vaccine 5:213
- hepatitis B, *see Hepadnaviridae*
- hepatitis B e antigen (HBeAg) 2:374
- hepatitis B hyperimmune globulin 5:269–270
- Hepatitis B hyperimmune IG (HBIG) 5:270
- hepatitis B surface antigen (HBsAg) 5:269–270, 2:374
- hepatitis B virus (HBV) 1:521, 1:266, 1:537, 1:668, 1:490, 5:128, 2:100
 HBV DNA polymerase 5:128
- hepatitis B virus core antigen (HBcAg) 1:664
- Hepatitis B Virus Database (HBVDb) 1:141–142
- hepatitis B virus genotypes, recombination among 1:112–113
- hepatitis C virus (HCV) 1:11, 1:537, 1:178, 1:180, 5:61–62, 5:123, 4:28–30, 1:498–499
 activation and inhibition of innate and adaptive immunity by 2:391–392
 classification 2:386
 clinical features 2:392–393
 diagnosis 2:393–394
 epidemiology 2:392
 formation of the HCV replication organelle 2:390
 genome organization and viral proteins 2:387–389
 HCV NS3/4A protease 5:123
 HCV NS5A phosphoprotein 5:123
 HCV NS5B polymerase 5:123
 HCV viremia, defined 2:386
 infection 1:68
 life cycle 2:389–390
 pathogenesis 2:393
 prevention 2:394
 replication cycle
 host cell factors contributing to various steps of 2:390–391
 host cell lipids of relevance to 2:391
 treatment 2:394
 virion structure and properties 2:386–387
- hepatitis delta virus (HDV) 4:623
- hepatitis E virus (HEV) 5:110, 5:207, 5:213–214, 2:397, 2:400, 2:400–401, 2:402, 2:398t
 clinical presentation and course 5:209–210
 diagnosis
 antigen detection 5:210
 nucleic acid detection 5:211
 serological tests 5:212–213
 epidemiology
 HEV genotypes 1 and 2 5:207
 HEV genotypes 3 and 4 5:207–208
 infection 5:103
 management of acute hepatitis E 5:214
 management of chronic infections 5:214
 prophylaxis 5:214
 taxonomy 5:206
 transmission 5:208–209

- vaccine 5:214
- hepatocellular carcinoma (HCC) 5:62, 2:106
- hepatocytes 2:390, 2:104
- hepatopancreatic parvovirus (HPV) *see* fenneropenaeus chinensis
- hepandensovirus (FchDV)
- Hepatoviruses 1:279–280
- Hepeviridae*
- classification 2:397
 - clinical features 2:401–402
 - diagnosis 2:402
 - epidemiology 2:399–401
 - genome 2:397–399
 - life cycle 2:399
 - pathogenesis 2:402
 - prevention 2:402
 - treatment 2:402
 - virion structure 2:397
- Herbert virus 4:764
- herd immunity, defined 5:295, 1:559
- herpes encephalitis 5:176–177
- herpes gladiatorum 5:176
- herpes rubiginosus 5:176
- herpes simplex virus (HSV) 1:266, 1:178, 5:182–183, 5:109, 5:124, 2:714–715
- associated disease 5:193–194
 - HSV DNA polymerase UL30 5:124
 - HSV envelope protein 5:124–125
- herpes simplex virus (HSV) infections
- available antiviral agents for 5:175
 - immunocompromised hosts 5:180
 - immunoprophylaxis 5:180
 - management beyond neonatal period 5:175
 - CNS infections 5:176–177
 - cutaneous infections 5:175–176
 - genital herpes 5:177–178
 - ocular infections 5:176
 - orolabial infections 5:175
 - neonatal herpes simplex virus infection 5:178–180
- herpes simplex virus 1 (HSV-1) 1:424, 1:658–659, 1:604, 1:446, 1:252–253
- herpes simplex virus 1 and 2 (HSV-1, HSV-2) 1:321
- classification 2:404
 - clinical features 2:409–410
 - genital herpes 2:410
 - gingivostomatitis 2:409–410
 - herpes simplex encephalitis 2:411
 - herpes simplex virus infections in the immunocompromised host 2:411
 - herpetic keratitis 2:410
 - mucocutaneous infections 2:409–410
 - neonatal herpes simplex virus infection 2:410–411
 - skin manifestations 2:410
 - diagnosis 2:411–412
 - epidemiology 2:408–409
 - life cycle 2:404–407
 - epigenetic regulation 2:407
 - host factors 2:407
 - latent infection 2:407–408
 - lytic infection 2:406–407
 - pathogenesis 2:411
 - prevention 2:412
 - treatment 2:412
 - virion structure 2:404
 - genome 2:404
- Herpesviridae* 2:404, 2:267
- herpesvirus 2:442, 1:321–322
- defined 2:441
- herpesviruses virions, structure of 1:322f
- herpesvirus infections in
- immunocompromised host 5:190
 - cytomegalovirus (human herpesvirus 5) 5:190–191
 - antiviral drug resistance 5:192–193
 - diagnosis of CMV disease and treatment 5:192
 - pre-transplant risk stratification 5:191
 - prevention strategies of CMV disease after transplantation 5:191–192
- Epstein Barr virus 5:195
- herpes simplex virus (HSV)-associated disease 5:193–194
- human herpes virus 5:194–195
- human herpes virus-8 5:195–196
- Varicella Zoster Virus (VZV) 5:194
- herpes virus of turkeys (HVT) 2:652–653, 2:542
- herpes zoster, defined 5:281
- Hershey-Chase experiment 4:206
- heteroduplex DNA, defined 4:291
- hetero-encapsidation, defined 3:528
- heterokaryon, defined 4:557, 4:534, 4:450
- heterokaryosis 4:520
- heterokaryotic/homokaryotic, defined 4:568
- heterospecific antibodies, defined 3:727
- heterothallic fungus, defined 4:450
- heterotrimeric G proteins 4:523
- HeV *see* Hendra virus (HeV)
- HEV *see* hepatitis E virus (HEV)
- hexagonal peroxisome protein (Hex1) 4:432
- hexamers 1:250–251
- hexon 1:331
- defined 2:441
- HFMD *see* hand, foot-and-mouth disease (HFMD)
- H form 1:414
- HFRS *see* hemorrhagic fever with renal syndrome (HFRS)
- HGT *see* horizontal gene transfer (HGT)
- hhi1* transcript 4:832–833
- HHIV-1, *see* *Haloarcula californiae* icosahedral virus 1 (HHIV-1)
- HHIV-2, *see* *Haloarcula hispanica* icosahedral virus 2 (HHIV-2)
- HHPV-1, *see* *Haloarcula hispanica* pleomorphic virus 1 (HHPV-1)
- HHV-6 *see* human herpesvirus 6 (HHV-6)
- HHV-8 *see* human herpesvirus-8 (HHV-8)
- hibiscus chlorotic ringspot virus (HCRSV) 1:667
- Hidden Markov Model (HMM) 1:127
- HIF pathway *see* hypoxia-inducible factor (HIF) pathway
- high-density lipoproteins (HDL) 2:386–387
- higher-order RNA structures, defined 3:778
- highly active antiretroviral therapy (HAART) 5:133, 2:455, 5:121
- highly pathogenic avian influenza (HPAI) 2:119, 2:117
- high mobility group box protein 1 (HMGB1) 2:139, 1:660
- high-pressure freezing (HPF) 5:13
- high throughput sequencing (HTS) 1:50, 5:92–93, 1:47, 3:192, 5:27, 1:175–176, 1:177
- see also* next generation sequencing (NGS)
- applications of high-throughput sequencing in virology 1:177
- DNA sequencing – amplicon sequencing 1:178–179
- enrichment strategies 1:179
- negative selection-based enrichment 1:179–180
 - positive selection-based enrichment 1:179
- first sequencing methods 1:175–176
- 454 (pyrosequencing) 1:176–177
- illumina sequencing 1:177
- metagenomic approaches 1:177–178
- next generation sequencing and applications in virology 1:176
- RNA-sequencing 1:178
- single cell sequencing 1:180
- using single cell sequencing to investigate cellular heterogeneity and its impact on viral infection 1:180
 - using single cell sequencing to investigate viral heterogeneity 1:180
- single molecule sequencing (Pacific Biosciences and nanopore) 1:180–181
- merging SRS and LRS 1:182
- Oxford Nanopore technologies 1:181–182
- Pacific Biosciences 1:181
- higrevirus 3:247
- diagnosis 3:251
 - diseases, epidemiology, and control 3:250–251
 - genome organization and gene product function 3:248–249
 - members of the family 3:247
 - replication and propagation 3:249–250
 - taxonomy, phylogeny, and evolution 3:247
 - transmission and host range 3:250
 - virion structure 3:247–248
- Himetobi P virus (HiPV) 4:771
- HiPV *see* Himetobi P virus (HiPV)
- histo-blood group antigens (HBGAs) 1:307–308, 1:392–393, 1:390
- noroviruses 1:392–393
- rotaviruses 1:393
- histology, defined 2:629
- histone deubiquitinase (DUB) activity 4:592
- history of virology 2:414, 3:3
- arthropod-borne viruses of vertebrates 2:417
 - biochemical/biophysical age 3:4–5
 - biochemistry 2:415–416
 - biological age 3:4
 - chemical composition of viruses 2:415

- history of virology (*continued*)
 cultivation of animal viruses 2:415
 discovery of phages 4:3–4
 early investigations 2:414
 foot-and-mouth disease virus 2:414
 yellow fever virus 2:414
 future 2:417
 HIV-AIDS, recognition of 2:417
 immunology, impact on 2:416–417
 molecular biology age 3:5–7
 phage and biotechnology 4:7–8
 phage biochemistry 4:6–7
 phage ecology and evolution 4:8
 phage genetics 4:5–6
 phage therapy 4:4
 phage typing 4:4–5
 physical studies of viruses 2:414–415
 prehistory 3:3
 recognition of viral entity 3:3–4
 structure of the virion 2:416
 taxonomy and nomenclature 2:417
 tumor virology 2:416
 vaccines and disease control 2:417
 word virus 2:414
- HI test *see* hemagglutination inhibition (HI) test
- HIV *see* human immunodeficiency virus (HIV)
- HJ *see* Holliday junction (HJ)
- HK97 fold 1:263
- H/KDEL receptors, defined 4:534
- HLA *see* human leukocyte antigen (HLA)
- HMGB1 *see* high mobility group box protein 1 (HMGB1)
- HMG-CoA *see* 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA)
- HMM *see* Hidden Markov Model (HMM)
- HMP *see* Human Microbiome Project (HMP)
- hMPV *see* human metapneumovirus (hMPV)
- HoCV-1 *see* homalodisca coagulata virus 1 (HoCV-1)
- Holliday junction (HJ) 4:64, 2:857
- Holliday junction resolvase, defined 4:387
- homalodisca coagulata virus 1 (HoCV-1) 4:773
- homeologous DNA, defined 4:291
- homogenous wash-free immunoassays 5:19
- homologous, defined 1:62
- homologous DNA, defined 4:291
- homologous gene products, defined 2:441
- homologous recombination, defined 4:387, 4:359
- homologous repeat (hr), defined 4:808, 4:747
- Homologous Structure Finder (HSF) software 1:157
- homologs, defined 4:242
- homology, defined 1:87
- homothallic fungus, defined 4:450
- hordeiviruses (*Virgaviridae*) 3:420
 function and relatedness of 3:422–423
 coat protein (CP) 3:423
 cytopathology and replication 3:424–426
 pathogenesis 3:426–427
 pathogenesis protein 3:424
 replicase proteins 3:422–423
 TGB1 3:423
 TGB2 protein 3:423–424
 TGB3 protein 3:424
 future perspectives 3:428
 taxonomy and characteristic 3:420–421
 genome structure and expression 3:421–422
 virus particle structure 3:421
- hordeivirus-host interactions 3:427
 b interference with host defenses 3:427
 autophagy responses 3:427
 RNA silencing pathways 3:427
 ROS bursts 3:427
 hordeivirus applications to cereal genomic analyses 3:428
 post-translational modifications of BSMV proteins 3:427–428
- horizontal gene exchange, defined 4:45
- horizontal gene transfer (HGT) 1:613, 1:77, 1:71, 4:242, 1:87, 4:732
- horizontal mycovirus transmission, defined 4:468
- horizontal transmission, defined 4:520, 2:778, 4:419, 1:559, 3:388, 3:430, 3:106, 4:780, 4:768
- hospital admissions 5:249
- host, clinical status of 1:562–563
- host, defined 4:276
- Hosta virus X (HVX) 3:628
- host cell recognition 4:201–203
- host dependency factors (HDFs) 1:628–629, 1:627, 1:629
- host enzymes, viral hyper-mutation mediated by 1:55
 cellular adenosine deaminases 1:55–56
 cellular cytidine deaminases 1:55
- host factor, defined 3:252
- host genes with indirect effects on retrovirus infection, spread or disease 1:630
- host immune response, shielding from 5:239–240
 chemical shielding and cell carriers 5:240–241
 virus engineering 5:239–240
- host immunity 1:630
- host-pathogen coevolution, defined 2:730
- host plant resistance, defined 3:355
- host plasticity 5:257
- host population immunity, influence of 1:563
- host range, defined 4:265, 4:276, 4:732
- host-virus exchange via horizontal gene transfer and transposable elements 1:77
- host-virus interactions
 complexity of 4:422–424
 environment impact on 4:420–422
- hot deserts 4:348–349
- hot tumor 1:659
- house finches (*Haemorhous mexicanus*) 2:805
- housefly, MdSGHV transmission dynamics in 4:789
- house sparrows (*Passer domesticus*) 2:805
- HPA *see* Hamiltonian Paths Analysis (HPA); Hamiltonian Paths Approach (HPA)
- HPAI *see* highly pathogenic avian influenza (HPAI)
- HPeV1 *see* human parechovirus 1 (HPeV1)
- HPF *see* high-pressure freezing (HPF)
- HRPV-1, *see* *Halorubrum pleomorphic* virus 1 (HRPV-1)
- HRPV-6 *see* *halorubrum pleomorphic* virus 6 (HRPV-6)
- HRPVs *see* haloarchaeal pleomorphic viruses (HRPVs)
- HRTV *see* heartland banyangvirus (HRTV)
- HRV *see* human rhinoviruses (HRV)
- HRV14 *see* human common cold virus strain 14 (HRV14)
- HRV1A *see* human rhinovirus 1A (HRV1A)
- HSCT *see* hematopoietic stem cell transplantation (HSCT)
- HSF software *see* Homologous Structure Finder (HSF) software
- HSPG *see* heparan sulfate proteoglycans (HSPG)
- HSTV-1 *see* *Haloarcula sinaiensis* tailed virus 1 (HSTV-1)
- HSTV-2 *see* *Halorubrum sodomense* tailed virus 2 (HSTV-2)
- HSV *see* herpes simplex virus (HSV)
- HTLV-1 *see* human T-cell leukemia virus-1 (HTLV-1)
- HTSNGS *see* high throughput sequencing (HTS)
 see also next generation sequencing (NGS)
- HTUs *see* hypothetical taxonomy units (HTUs)
- huaiyangshan banyangvirus 2:771–775
- Hubei sclerotinia RNA virus 1 (HuSRV1) 4:466
- human adenoviruses (HAdVs) 5:197, 5:199, 5:198*f*, 5:202*f*
- human adenovirus type 5 (HAdV-C5) 1:329–331
 core proteins 1:335
 fiber 1:331–333
 hexon 1:331
 minor coat proteins 1:333
 protein IIIa 1:333
 protein VI 1:333
 protein VIII 1:333
 protein IX 1:333–335
 penton base protein 1:331
- human airway epithelium (hAE) 5:170
- human and animal viral diseases, epidemiology of
 digital epidemiology 1:566
 disease occurrence and outcome, assessment of 1:564–565
 epidemiological studies 1:564–565
 molecular epidemiology 1:565–566
 disease prevention, implications for 1:568
 factors influencing the spread of viral diseases 1:561
 arthropod transmission cycles, influence of 1:563–564
 clinical status of the host, influence of 1:562–563

- host population immunity, influence of 1:563
 modes of virus transmission 1:561–562
 perpetuation of viruses in nature 1:561
 population size, influence of 1:563
 virulence of the virus, influence of 1:563
 virus entry 1:561
 virus shedding 1:561
 zoonotic transmission cycles, influence of 1:563
 mathematical modeling 1:566–567
 data sharing, data privacy and ethics 1:567
 proof of causation 1:567–568
 sentinel studies 1:566
 seroepidemiologic studies 1:566
 vaccine trials 1:566
 human bronchial epithelial (HBE) cultures 2:749
 human common cold virus strain 14 (HRV14) 1:259–260, 1:265–266, 1:266
 human coronavirus (HCoV)
 background 2:428–429
 classification 2:429
 diagnosis 2:438
 epidemiology and clinical features 2:437
 HCoV-229E 2:437
 HCoV-HKU1 2:437–438
 HCoV-NL63 2:437
 HCoV-OC43 2:437
 genome 2:430–431
 accessory proteins 2:431–432
 non-structural proteins 2:431
 HCoV 229E and NL63 2:256
 HCoV-host interactions 2:434–435
 apoptosis 2:435–436
 autophagy 2:435
 ER stress response 2:435
 MAP kinase pathway 2:435
 translational control 2:434–435
 HCoV-OC43 1:122
 life cycle 2:432–433
 assembly and release of virion 2:434
 attachment 2:432–433
 formation of the replication-transcription complex 2:433
 viral entry and uncoating 2:433
 viral RNA synthesis 2:433–434
 pathogenesis 2:436–437
 prevention 2:439
 treatment 2:438–439
 virion structure 2:429–430
 human cytomegalovirus (HCMV) 5:62, 5:227, 1:604, 5:106–107, 5:125, 1:321, 1:321–322
 see also cytomegalovirus (CMV)
 HCMV DNA polymerase UL54 5:125
 HCMV terminase UL56 5:125
 human herpes virus 5:194–195
 human herpesvirus 5 5:190–191
 antiviral drug resistance 5:192–193
 diagnosis of CMV disease and treatment 5:192
 pre-transplant risk stratification 5:191
 prevention strategies of CMV disease after transplantation 5:191–192
 human herpesvirus 6 (HHV-6) 2:778–779, 2:779*f*, 2:780*t*, 2:780, 2:781, 2:783, 2:784, 2:787–788
 and disease after organ transplant 2:786
 HHV-6A 2:781, 2:783
 and HHV-6B 2:779–780
 chromosomal integration of 2:781–782
 vertical transmission of 2:783
 HHV-6B 2:781, 2:783
 and disease after organ transplant 2:786
 chromosomal integration of 2:781–782
 vertical transmission of 2:783
 HHV-6B and HHV-7 primary infection and disease 2:785
 encephalitis 2:785
 exanthem subitum 2:785
 FSE and temporal lobe epilepsy 2:785
 and HHV-7 disease in older children and adults 2:785
 Alzheimer's disease 2:786
 and the possibility of encephalitis 2:785
 delayed primary infection 2:785
 HHV-6 and multiple sclerosis 2:786
 HHV-6B and DRESS (drug rash with eosinophilia and systemic symptoms) 2:785–786
 and HHV-7 laboratory tests 2:786–787
 antibody detection 2:787
 CIHHV-6, diagnosis of 2:787
 CIHHV-6 and misdiagnosis of active infection 2:787
 virus detection – HHV-6A, HHV-6B and HHV-7 2:787
 and HIV 2:786
 horizontal transmission of 2:783
 transmission of HHV-6 by organ donation 2:783–784
 human herpesvirus-8 (HHV-8) 2:599, 5:195–196, 1:179
 human immune deficiency virus (HIV)
 integrase inhibitors 5:147*t*–148
 children and adolescents 5:153
 entry inhibitors 5:151–152
 enfuvirtide 5:152
 fostemsavir 5:152–153
 ibalizumab-uiyk 5:152
 maraviroc 5:151–152
 HIV-2 5:153
 integrase strand transfer inhibitors (INSTI) 5:145–146
 first generation INSTI 5:146
 second generation INSTI 5:149–150
 virology 5:145–146
 post-exposure and pre-exposure prophylaxis (PEP and PrEP) 5:153
 pregnancy 5:153
 human immunodeficiency virus (HIV)
 1:266, 2:56, 2:56–57, 1:399, 1:229, 1:121, 5:273, 1:178, 1:605, 5:121, 5:190
 classification 2:460
 clinical features 2:469–470
 diagnosis 2:468–469
 early phase 2:464
 attachment and entry 2:464
 uncoating, reverse transcription, nuclear localization, and integration 2:464
 epidemiology 2:466–468
 modes of transmission and risk factors 2:467–468
 prevalence 2:468
 genome, organization of 2:461
 HIV-AIDS
 recognition of 2:417
 surveillance of 5:250–251
 HIV-associated dementia (HAD) 2:470–471, 2:471
 HIV associated Kaposi's sarcoma 2:606
 HIV GP120 5:122–123
 HIV GP41 5:122
 HIV integrase 5:122
 HIV protease 5:121–122
 HIV reverse transcriptase 5:121
 late phase 2:464–465
 budding and maturation 2:465–466
 gene expression and assembly 2:465
 life cycle 2:464
 origin of HIV-1 2:460–461
 pathogenesis 2:470
 coreceptor switch 2:470
 drug resistance, evolution of 2:471
 latency 2:471
 macrophage-tropic HIV-1 and infection of CNS 2:470–471
 prevention 2:473
 biomedical prevention 2:473
 Pre-Exposure Prophylaxis (PrEP) 2:473–474
 social and educational prevention 2:474
 treatment 2:471–472
 entry inhibitors 2:472
 integrase inhibitors 2:473
 Protease Inhibitors (PIs) 2:473
 reverse transcriptase inhibitors (RTIs) 2:472–473
 viral genes, expression of 2:461–463
 viral proteins 2:463–464
 virion proteins 2:463
 human immunodeficiency virus type 1 (HIV-1) 1:620, 1:209, 1:567–568, 1:445, 1:383, 2:460, 1:599, 2:56–57, 2:57*f*, 2:58*f*, 2:57–58, 2:62, 2:64
 HIV-1 centralized vaccines, recombination in the design of 1:112
 HIV-1 fitness recovery, influence of recombination on 1:112
 HIV-1 gp120, analyzing adaptation of 1:113
 human immunodeficiency virus type 2 (HIV-2) 2:827–828, 2:56–57, 2:57*f*, 2:64, 5:153
 clinical features 2:832
 defined 2:827
 epidemiology of 2:830–831
 diagnosis of HIV-2 2:831–832
 risk and transmission of HIV-2 2:831
 genome 2:828–829
 lifecycle 2:828–829
 immunity to 2:832–833
 HIV-2 cellular immunity 2:833
 HIV-2 humoral immunity 2:833

- human immunodeficiency virus type 2 (HIV-2) (*continued*)
- interactions of HIV-2 and HIV-1 *in vivo* 2:834–836
 - HIV-2 protection from HIV-1 2:834–836
 - in vitro* evidence for HIV-2 protection from HIV-1 2:835–836
 - pathogenesis of 2:832
 - prevention of 2:834
 - taxonomy and classification 2:828
 - treatment of 2:833–834
 - virion structure 2:828
- human influenza virus 5:125
- matrix protein 2 5:127
 - neuraminidase 5:125–127
 - viral RNA polymerase 5:125
- human leukocyte antigen (HLA) 2:639–640
- human metapneumovirus (hMPV)
- classification 2:475
 - diagnosis 2:481
 - prevention 2:481–482
 - treatment 2:481
- epidemiology and clinical features 2:477–479
- genome 2:475–476
- life cycle 2:476–477
- entry of hMPV 2:476–477
 - viral particle assembly and spread 2:477
 - viral RNA synthesis 2:477
- pathogenesis 2:479–480
- cellular responses 2:479–480
 - pattern recognition receptors (PRRs) and antagonism of immune responses 2:480–481
- virion structure and virus proteins 2:475
- human microbiome
- bacteriophages of *see* bacteriophages: of human microbiome
 - defined 1:552, 4:283
- Human Microbiome Project (HMP) 1:553
- human noroviruses (HuNoVs) 1:535
- see also* norovirus
- human papillomavirus (HPV) 1:588, 1:668, 1:179
- functions of the HPV proteins 2:497*t*
 - genome 2:493–494
 - host range and tissue tropism 2:494–496
 - individual viral genes, functions of 2:497–498
 - clinical features, pathogenesis and histopathology of HPV infection 2:499
 - diagnosis 2:500
 - E1 protein 2:497–498
 - E1^ΔE4 protein 2:498
 - E2 protein 2:498
 - E5 protein 2:498
 - E6 protein 2:498
 - E7 protein 2:498–499
 - E8^ΔE2 protein 2:498
 - HPV-associated cancers 2:500
 - immune response 2:500
 - L1 protein 2:499
 - L2 protein 2:499
 - natural history, transmission and epidemiology 2:499
 - oncogenic HPV 2:499–500
 - prevention 2:501
 - treatment 2:500–501
- life cycle 2:496–497
- early stages of infection 2:497
 - entry 2:496–497
 - late stages of infection 2:497
 - maintenance phase of infection 2:497
- phylogenetic tree of 2:494*f*
- taxonomy, classification, and evolutionary relationships 2:493
- virion structure 2:493
- human papillomaviruses (HPV) vaccines 5:295–296
- current status and future directions 5:297
 - HPV and HPV-related diseases 5:295
 - immunogenicity 5:296
 - impact 5:296
 - on anogenital warts 5:297
 - on HPV infection and precancerous lesions 5:296
 - on HPV-related cancer 5:296–297
 - models on impact and cost-effectiveness of 5:298
 - vaccine safety and confidence 5:297–298
- human parechovirus 1 (HPeV1) 1:282
- human pathogenic arenaviruses *see* arenaviruses, human pathogenic
- human polyomaviruses *see* polyomaviruses (PyVs)
- human rhinovirus 1A (HRV1A) 1:286
- human rhinoviruses (HRV) 1:395, 1:202
- human sapovirus *see* sapovirus
- human T-cell leukemia virus-1 (HTLV-1) 2:316, 2:317, 1:352
- human T-cell leukemia viruses 1 and 2 (HTLV-1 and HTLV-2)
- classification 2:528–529
 - clinical features 2:532–533
 - adult T-cell leukemia (ATL) 2:532–533
 - HAM/TSP 2:533
 - HTLV-2-associated symptoms 2:533
 - diagnosis and prevention 2:538
 - epidemiology 2:532
 - transmission 2:532
 - genome 2:529–531
 - immune response 2:537–538
 - life cycle 2:531
 - assembly, budding, and maturation 2:531–532
 - viral gene transcription 2:531
 - viral translation 2:531
 - pathogenesis 2:533–535
 - Hbz* gene 2:535–536
 - in vivo* animal models 2:537
 - Tax proteins 2:533–535
 - viral clonality 2:536–537
 - treatment 2:538
 - virion structure 2:529
 - viral entry 2:529
- human virome, defined 1:552, 2:48
- see also* virome, human
- humoral and T cell-mediated immunity to viruses
- adaptive immune response to viral infections 1:585–587
 - T cell recognition of viral antigens 1:587
- antibodies in antiviral humoral immunity 1:591
- antibody-dependent cellular cytotoxicity (ADCC) 1:591–592
 - neutralization 1:591
 - opsonization 1:591
- dysregulation of immune response to viruses 1:595–596
- humoral immune response to viral infections 1:590–591
- B cells 1:590–591
- innate immune system, role of 1:584
- barrier immunity 1:584
 - crosstalk between innate and adaptive immune cells 1:585
 - innate immune effectors 1:584
 - innate immune recognition of viruses 1:584–585
- mounting an effective T cell-mediated immune response to viral infections 1:589
- cellular immune responses, contraction of 1:590
- cytotoxicity 1:590
- immune memory, formation of 1:590
- T cell help 1:589–590
- T cell priming 1:589
- T cell effectors in the response to viral infections 1:587
- CD4⁺ T Cells 1:587
 - CD8⁺ cytotoxic T cells (CTL) 1:589
 - cytotoxic CD4⁺ T cells 1:588–589
 - regulatory T cells (Tregs) 1:588
 - T follicular helper (Tfh) cells 1:587–588
 - T helper 1 (Th1) cells 1:587
- vaccination, induction of immunity to viruses by 1:592–595
- humoral immune response to viral infections 1:590–591
- B cells 1:590–591
 - VLPs and the induction of 1:662–663
- humoral immunity, defined 2:79
- HuSRV1 *see* Hubei sclerotinia RNA virus 1 (HuSRV1)
- HVAC systems *see* heating, ventilation, and air conditioning (HVAC) systems
- HVD *see* hypervariable domain (HVD)
- HVT *see* herpes virus of turkeys (HVT)
- HVX *see* Hosta virus X (HVX)
- hyaline cells (HCs) 4:816
- Hyalomma marginatum* tick-adapted TBEV 1:550
- hyaluronidase 2 (HYAL2) 2:576
- hybridization probe, defined 3:192
- hydranencephaly, defined 2:34
- hydrodynamic model for phage genome ejection *in vivo* 4:216, 4:217*f*, 4:217
- hydrophobic, defined 1:345
- 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMG-CoA) 2:391
- hymenoptera, defined 4:867
- hymenopteran viruses 4:771, 4:772
- defined 4:768
- hyperarid 4:342
- hyperendemic, defined 2:899

- hyperendemicity, defined 2:218
hyperhalophiles, defined 4:387
hyperimmune immunoglobulins 5:269–270
 cytomegalovirus hyperimmune immunoglobulin 5:270
 hepatitis B hyperimmune globulin 5:269–270
 rabies hyperimmune immunoglobulin 5:270–271
 vaccinia immune globulin (VIG) 5:271–272
 varicella hyperimmune immunoglobulin 5:272
hyperparasite, defined 3:581
hyperplasia, defined 2:316, 3:200, 4:780
hyperplastic lesion, defined 2:306
hypersaline environments, phages in 4:343–345
Hypersensitive Reaction, defined 3:612
hypersensitive response, defined 3:60
hyperthermophiles, defined 4:387
hyperthermophilic, defined 4:359
hyperthermophilic archaeal viruses, diversity of 4:359–360
 family *Fuselloviridae* 4:365
 genomes 4:365
 morphology and structure 4:360
 order *Ligamenvirales* 4:365–366
 evolutionary relationships 4:366
 particular morphologies, viruses with 4:360–361
 family *Ampullaviridae* 4:361
 family *Bicaudaviridae* 4:361
 family *Clavaviridae* 4:361–362
 family *Fuselloviridae* 4:362
 family *Guttaviridae* 4:362
 family *Spiraviridae* 4:362
 filamentous viruses 4:363–364
 spherical viruses 4:362
hypertrophy, defined 3:200, 4:780
hypervariable domain (HVD) 2:177
hyphal anastomosis, defined 4:431, 4:568, 4:528, 4:468
hypocalcemia, defined 2:875
hypothetical taxonomy units (HTUs) 1:117
hypovirulence 4:431, 4:461, 4:589, 4:594, 4:577, 4:568, 4:615, 4:493, 4:552, 4:468
 in *Botrytis cinerea* 4:476
 in *Cryphonectria parasitica* 4:472–474
 artificial application of hypovirulence 4:474–475
 exclusive transmissible hypovirulence 4:473–474
 in *Fusarium graminearum* 4:477
 in *Helminthosporium victoriae* 4:476–477
 in *Ophiostoma novo-ulmi* 4:475
 in *Rosellinia necatrix* 4:476
 in *Sclerotinia sclerotiorum* 4:475–476
hypoviruses (*Hypoviridae*) 4:589
 anti-hypovirus defense mechanisms 4:592–593
 hypovirus functional domains 4:591–592
 hypovirus gene expression strategy 4:589–591
 taxonomy and genetic organization 4:589
hypoxia-inducible factor (HIF) pathway 2:672
hytrosa, defined 4:780
Hytrosaviridae 4:706–707, 4:780
hytrosaviruses 4:707, 4:780–781
 diagnosis of SGHV infections 4:789–790
 genome organization 4:783
 host range 4:786
 major structural proteins 4:784–786
 management of SGHV infections 4:790
 modes of infection and gene expression 4:783–784
 pathogenesis in other host tissues 4:787–788
 pathogenesis in the salivary glands 4:787
 similarities with other virus taxa 4:781–783
 taxonomy and classification 4:781
 transmission and epidemiology 4:788–789
 GpSGHV transmission dynamics in the tsetse fly 4:788–789
 MdSGHV transmission dynamics in the housefly 4:789
 viral latency 4:788
 virion structure 4:783
Hz-1 baculovirus 4:827–828
Hz-1 virus (Hz-1V) 4:827–828
HzNV-1 *see* heliothis zeia nudivirus 1 (HzNV-1)
- I**
- IAF *see* International Accreditation Forum (IAF)
IAP *see* Intracisternal A Particle (IAP)
IAPV *see* Israeli acute paralysis virus (IAPV)
IAS *see* internal activation sequence (IAS)
iatrogenic Kaposi's sarcoma 2:606
iatrogenic transmission 1:562, 1:559
IAV *see* intracellular annular viruses (IAV)
Ibalizumab 5:273
ibalizumab-uiyk 5:152
IBDV *see* infectious bursal disease virus (IBDV)
IBs *see* inclusion bodies (IBs)
IBV *see* infectious bronchitis virus (IBV); intranuclear bacilliform viruses (IBVs)
IC50 *see* inhibitory concentration 50 (IC50)
ICAM-1 *see* intercellular adhesion molecule-1 (ICAM-1)
ICAM-5 *see* intercellular adhesion molecule 5 (ICAM-5)
ICD *see* immunogenic cell death (ICD)
ICER *see* incremental cost-effectiveness ratio (ICER)
ICEs *see* integrative and conjugative elements (ICEs)
ichnoviruses 4:712–713, 4:849
Ichadenovirus 2:14
Ichthamaparvovirus 4:836
ICI *see* immune checkpoint inhibitors (ICI)
iCIHHV-6 *see* inherited CIHHV-6 (iCIHHV-6)
- ICOS *see* inducible T cell co-stimulator (ICOS)
icosahedra, triangulation numbers of 1:222–223
icosahedral asymmetric unit, defined 1:329, 1:278
icosahedral capsid, defined 2:797
icosahedral capsid architecture, mathematical models of 1:249–250
 implications for viral evolution 1:253–254
 non-quasi-equivalent trihexagonal lattice architectures 1:253
 overarching framework for icosahedral architectures 1:251–253
 practical applications 1:251
 predictions and limitations of quasi-equivalence theory 1:250
 quasi-equivalence theory 1:249–250
 viral tiling theory 1:250–251
 virus structure in 3D – The Hamiltonian paths approach 1:254
icosahedral capsid assembly, structural principles in 1:480–481
icosahedral DNA viruses with an internal membrane 1:406–407
icosahedral enveloped dsRNA bacterial viruses 1:407–408
icosahedral enveloped viruses 1:468–469
 alphavirus assembly 1:469, 1:471f
 alphavirus assembly and budding 1:470–472
 alphavirus life cycle 1:469
 alphavirus virion structure 1:469–470
 flavivirus assembly 1:472, 1:473f
 flavivirus assembly and budding 1:473
 flavivirus life cycle 1:472
 flavivirus virion structure 1:472–473
icosahedral head/capsid 1:319
icosahedral internal membrane-containing viruses 4:374–376
icosahedral particle, defined 3:712
icosahedral phages
 DNA packaging and DNA binding protein 4:18–19
 DNA replication 4:15
 evolution and evolutionary studies 4:19
 evolution of a two scaffolding protein system 4:19–20
 gene expression 4:15–17
 history 4:10
 host cell recognition, attachment, eclipse and penetration 4:12–15
 lysis 4:19
 morphogenesis 4:17–18
 X174-like viruses as a model system for experimental evolution 4:20
 virion morphology and genome content 4:10–12
icosahedral ssDNA bacterial viruses 1:407
icosahedral ssRNA bacterial viruses 1:408, 4:21–24
 background 4:21–24
 phage lifecycle 4:21–24
icosahedral surface lattices, thermodynamic transition between 1:297–299
icosahedral symmetry 1:258–261, 1:255

- defined 1:362, 1:248, 4:568
viral geometry and 1:249
- icosahedral tailed dsDNA bacterial viruses 1:405–406
- icosahedral tailed viruses 4:370–373
of halophilic euryarchaea 4:371–373
icosahedral tailed viruses of halophilic euryarchaea 4:371–373
tailed viruses of methanogenic euryarchaea 4:373–374
- icosahedron 1:258–259, 2:540, 4:115, 3:818, 3:545
- ICTV *see* International Committee On Taxonomy Of Viruses (ICTV)
- idaeviruses (*Idaevirus*) 3:430–431
diagnosis and detection 3:436
reverse transcriptase 3:436
serological testing 3:436
disease symptoms and effects 3:431–433
naturally infected plants 3:432–433
experimentally infected plants 3:433
genome 3:433–435
geographical distribution 3:431
natural and experimental transmission 3:431
prevention and control 3:436–437
relationships with other viruses 3:437–438
therapy 3:436
viral proteins 3:435
coat protein 3:435
movement protein 3:435
replicase protein 3:435–436
virion structure 3:433
- identity, defined 4:276
- Idnoreovirus* 4:869
- idnoreoviruses 4:868, 4:871, 4:875–876, 4:877
- IDV *see* intracellular dense viruses (IDV)
- IEDB *see* Immune Epitope Database (IEDB)
- IEM *see* immuno-electron microscopy (IEM)
- IEV *see* intracellular enveloped virion (IEV)
- IF *see* immunofluorescence (IF)
- IFA *see* immunofluorescence assay (IFA)
- Iflaviridae* 4:707, 4:792
- iflaviruses (Picornavirales, Iflaviridae, Iflavirus)
classification 4:792
clinical signs 4:794–795
diagnosis/detection methods 4:795
epidemiology 4:794
genome 4:792
life cycle 4:792–794
pathogenesis 4:795
prevention 4:795
virion structure 4:792
- IFN *see* interferon alpha (IFN)
- IFT *see* indirect Fourier transform (IFT)
- IG domains *see* immunoglobulin (IG) domains
- IgG avidity, defined 2:419
- IGLP *see* immunoglobulin-like proteins (IGLP)
- IgSF receptors, virus interactions with 1:395–396
- IHC *see* immunohistochemistry (IHC)
- IHNV *see* infectious hematopoietic necrosis viruses (IHNV)
- IHR *see* International Health Regulations (IHR)
- IHR *see* iterative helical reconstruction (IHR)
- Ikoma lyssavirus (IKOV) 2:738
- IKOV *see* Ikoma lyssavirus (IKOV)
- ILAC *see* International Laboratory Accreditation Cooperation (ILAC)
- Ilarvirus* 3:260, 3:261, 3:265, 3:262*t*
characteristics of RNA genome in 3:261*t*
- ilarviruses (*Bromoviridae*) 3:439
classification 3:439–440
control 3:443–444
diagnosis 3:445–446
genome 3:440–442
infection cycle 3:442
replication 3:442–443
symptoms and pathogenesis 3:444–445
transmission 3:443
virion structure 3:440
- illumina sequencing 1:177
- ILTV *see* infectious laryngotracheitis virus (ILTV)
- ILVs *see* intraluminal vesicles (ILVs)
- IMC-HZ-I-NOV 4:827–828
- immature intracellular virions 2:891–892
- immediate early gene, defined 2:441
- immortalization, defined 2:528
- immune checkpoint blockade, defined 5:233
- immune checkpoint inhibitors (ICI) 1:660, 1:658
- Immune Epitope Database (IEDB) 1:141
- immune escape 1:59–60
- immune memory, formation of 1:590
- immunity, defined 4:276, 2:441
- immunity repressor 4:81, 4:276
- immunization, defined 5:281
- immunoassays 5:17–18
homogenous wash-free immunoassays 5:19
immunoblotting 5:18
immunofluorescence microscopy 5:18–19
lateral-flow immunoassays 5:19
solid-phase microwell immunoassays 5:17–18
- immunoblot assays 5:94
- immunoblotting 5:18
- immunocapture-RT-PCR 3:629
- immunocompetent, defined 2:441
- immunocompromised, defined 2:441
- immunocompromised hosts 5:180
- immunocompromised patients, adenovirus infections of 5:201–202
- immunodeficiencies 5:190, 2:441, 2:56
- immunodeficiency-associated VDPV (iVDPV) 5:310, 5:313
- immunodeficient vaccine-derived polioviruses (iVDPVs) 2:695
- immuno-electron microscopy (IEM) 5:11
- immunofluorescence (IF) 2:640–641
- immunofluorescence assay (IFA) 5:18–19, 2:215, 5:94
- immunofluorescence microscopy 5:18–19
- immunogenic cell death (ICD) 1:659, 1:658
- immunoglobulin (IG) domains 1:606
- immunoglobulin 2:441, 1:533–534
- immunoglobulin-like fold 1:263
- immunoglobulin-like proteins (IGLP) 1:556–557
- immunoglobulins in the prevention of viral infections 5:267–268
- experimental immunoglobulin therapies 5:273–274
avian influenza 5:274
seasonal influenza 5:273–274
- hyperimmune immunoglobulins 5:269–270
cytomegalovirus hyperimmune immunoglobulin 5:270
hepatitis B hyperimmune globulin 5:269–270
rabies hyperimmune immunoglobulin 5:270–271
vaccinia immune globulin (VIG) 5:271–272
varicella hyperimmune immunoglobulin 5:272
- Middle Eastern respiratory syndrome coronavirus (MERS-CoV) 5:274–275
- ebola and other hemorrhagic fevers 5:274–275
experimental infections 5:275
- monoclonal antibodies 5:272–273
human immunodeficiency virus 5:273
respiratory syncytial virus (RSV) 5:272–273
- polyclonal immunoglobulins 5:268
hepatitis A 5:268
measles 5:268–269
- immunoglobulin superfamily proteins 1:394–396
- IgSF receptors, virus interactions with 1:395–396
- immunogold electron microscopy, defined 2:22
- immunohistochemistry (IHC) 2:190, 2:518, 2:79
- immunology, impact on 2:416–417
- immunomodulation 2:784, 2:373
- immunomodulators 5:241
- immunoprophylaxis 5:180
- immunosenescence 2:454
- immunosorbent electron microscopy, defined 3:539
- immunosuppressed individuals 5:105
adenoviruses (AdVs) 5:109
Epstein-Barr virus (EBV) 5:107–108
hepatitis E virus 5:110
herpes simplex virus (HSV) 5:109
human cytomegalovirus (HCMV) 5:106–107
polyomaviruses (PyV) 5:108–109
Varicella-Zoster Virus (VZV) 5:109–110
viral diagnostic tests 5:105–106
- immunosuppression, defined 2:48, 2:441
- immunotherapy, defined 1:658
- implications for viral evolution 1:253–254

- IMV *see* intracellular mature virion (IMV)
- inactivated polio vaccine (IPV) 2:690,
2:695, 1:668, 5:310–311
- inapparent, defined 2:441
- inbred mice, defined 2:643
- “inchworm” translocation model 4:153–154
- incidence, defined 5:247, 2:362
- incidence risk (or attack rate), defined 1:559
- inclusion bodies (IBs) 2:738–739, 2:343
- inclusion principle, defined 1:28
- incremental cost-effectiveness ratio (ICER) 5:157
- indel, defined 1:627
- Indian subcontinent, CLCuD in 3:357–359
- indirect Fourier transform (IFT) 1:192–193
- inducible T cell co-stimulator (ICOS) 1:591
ICOS ligand (ICOSL) 1:591
- induction, defined 4:88, 4:253, 2:441
- infectious bronchitis virus (IBV) 2:193–194,
2:194*f*, 2:198
- infectious bursal disease virus (IBDV)
1:492–493
classification 2:540
diagnosis and control 2:542
electron micrograph of 2:541*f*
pathogenesis and clinical features 2:541–542
virion structure and genome 2:540–541
- infectious clone, defined 2:193
- infectious diseases, surveillance of
analysis of data 5:252
person 5:252
place 5:252–253
time 5:252
- collection of data 5:248
death certification 5:248
hospital admissions 5:249
laboratory data 5:248–249
notifications 5:248
other sources of data 5:248
other sources of data 5:249
serological surveillance 5:249
surveillance of outbreaks 5:249
surveillance of viruses in nonhuman
and environmental sources 5:249
- essential characteristics of surveillance
5:247–248
- evaluation of surveillance 5:254
- feedback 5:254
- global and international surveillance
5:254
- interpretation 5:253–254
- surveillance system, newer types of 5:249–250
active surveillance 5:250
Enhanced Surveillance 5:250
sentinel surveillance 5:250
surveillance of HIV/AIDS 5:250–251
syndromic surveillance 5:250
- surveillance systems, attributes of 5:251–252
accuracy 5:252
completeness 5:251–252
consistency 5:252
- representativeness 5:252
timeliness 5:252
- infectious hematopoietic necrosis viruses (IHNV) 2:324, 2:328
- infectious hypodermal and hematopoietic necrosis virus (IHHNV) *see* penaeus stylirostris penstydensovirus 1 (PstDV1)
- infectious laryngotracheitis virus (ILTIV) 2:113
- infectious pancreatic necrosis (IPN) virus 2:544, 2:546*f*
classification 2:544
control 2:549–550
epidemiology 2:548
life cycle 2:547–548
molecular determinants of virulence 2:548–549
pathogenesis and clinical features 2:548
virion structure and genomic organization 2:544–547
- infectious subviral particle (ISVP) 1:308–309
- inflammasome, defined 3:60
- influenza 1:668, 5:300
as a disease 5:300
epidemiology of 5:300
- Influenza A 5:160
- influenza A virus (IAV) 2:291, 2:288*f*
see also human influenza virus
classification of 2:551
clinical features 2:558
diagnosis 2:559
evolution of 2:556
genome 2:552
genomic organization 2:552
reassortment and mutagenesis 2:552
- IAV epidemics and pandemics 2:556–558
- immunological responses and immune evasion 2:555–556
adaptive immune responses 2:556
immune evasion strategies of IAV 2:556
innate immune responses 2:555–556
- life cycle 2:552–554, 2:289*f*
attachment and entry 2:553–554
genome replication 2:555
new virions, budding of 2:555
transcription and translation 2:554–555
viral assembly 2:555
- pathogenesis 2:558–559
structure of 2:552*f*
treatment and prevention 2:559
antivirals 2:559
vaccines 2:559–560
virion structure 2:551–552
- influenza A virus hemagglutinin 1:597–598
- influenza B, C and D viruses
classification (compact) 2:561
clinical features 2:569, 2:573
diagnosis 2:569, 2:573
epidemiology 2:567–569, 2:572–573
functions of HEF glycoprotein during life cycle 2:571–572
life cycle and functions of viral proteins 2:564–566
BM2 2:566–567
- hemagglutinin (HA) 2:565–566
matrix protein (M1) 2:566
NB 2:566
NEP 2:567
neuraminidase (NA) 2:566
non-structural protein 1 (NS1) 2:567
- pathogenesis 2:569, 2:573
prevention 2:570, 2:573
treatment 2:569–570, 2:573
- influenza B virus 2:561–562
genome 2:562–564
virion structure 2:562
- influenza C and D viruses 2:570–571
virion structure and genome 2:570–571
- influenza HA 1:420
- influenza polymerase 5:125
- influenza research database (IRD) 1:143
- influenza vaccines
evaluating the impact of 5:305
immunogenicity 5:305
understanding vaccine efficacy and effectiveness 5:306
vaccine effectiveness 5:305–306
vaccine efficacy 5:305
history of 5:300
pandemic influenza vaccines 5:308
production of 5:300–301
dosage 5:302
influenza vaccine composition 5:301–302
production steps 5:301
timeline of the production process 5:300–301
valency 5:302
- safety and contraindications 5:304
allergic reactions 5:304–305
local and systemic inflammatory reactions 5:304
rare adverse events: neurological syndromes 5:305
- seasonal 5:302
adjuvanted vaccines 5:303
high-dose influenza vaccines 5:303
intradermal vaccines 5:303
live attenuated influenza vaccines 5:303
recombinant and cell-based influenza vaccines 5:302–303
the future of influenza vaccines 5:303–304
traditional inactivated vaccines 5:302
- specific population groups, vaccination of 5:306–307
elderly 5:307
healthy adults 5:308
healthy children 5:307–308
pregnancy 5:307
underlying conditions 5:307
- vaccination in practice 5:304
administration routes 5:304
duration of protection 5:304
timing of programs 5:304
vaccine uptake 5:304
- influenza viruses 1:266, 1:391–392
- influenza virus infections, antiviral management of 5:160
- adamantanes/M2 blockers 5:161–162

- influenza virus infections, antiviral management of (*continued*)
 complementary treatments for disease management 5:171–172
 ECMO (extracorporeal membrane oxygenation) 5:171–172
 drugs under development with other mechanisms of action 5:168
 HA attachment blockers 5:168
 HA processing blockers 5:168
 neuraminidase inhibitors 5:162–164, 5:163f
 pandemic influenza 5:160
 perspectives at short and medium 5:168–170
 combined Baloxavir + Oseltamivir 5:171
 combined Oseltamivir plus Favipiravir 5:171
 combined Oseltamivir plus monoclonal antibodies 5:171
 host target molecules 5:169–170
 macrolides 5:170
 neuraminidase inhibitors (NAI) combinations 5:171
 Pimodivir and Oseltamivir 5:171
 Siromilus 5:170–171
 triple therapy (Amantadine, Ribavirin, and Oseltamivir) 5:171
 polymerase (POL) inhibitors 5:164–166
 Baloxavir – Marboxyl (Xofluza) 5:166–168
 Favipiravir (Avigan) 5:168
 Pimodivir 5:168
 Ribavirin (Ribavirin) 5:164–166
 seasonal influenza 5:160
 zoonotic influenza 5:160–161
 inherited CIHHV-6 (iCIHHV-6) 2:781
 inhibitory concentration 50 (IC₅₀) 2:561
 INIBs *see* intranuclear inclusion bodies (INIBs)
 injection 4:48
 DNA Injection 4:48
 protein injection 4:48
 innate immune system, role of 1:584
 barrier immunity 1:584
 crosstalk between innate and adaptive immune cells 1:585
 innate immune effectors 1:584
 innate immune recognition of viruses 1:584–585
 innate immunity 4:524, 4:520, 2:404, 1:577, 2:428
 INSDC *see* International Nucleotide Sequence Database Collaboration (INSDC)
 insect cells 1:664
 insects, plant viruses transmission by 3:106–107
 beetles, transmission by 3:112–113
 circulative transmission 3:107–108
 circulative non-propagative transmission 3:107–108
 circulative propagative transmission 3:110
 history of typology of transmission modes 3:107
 non-circulative transmission 3:110–112
 capsid strategy 3:112
 helper strategy 3:112
 insects, rhabdoviruses of
 classification 4:883
 clinical features 4:886
 diagnosis 4:887
 epidemiology 4:886
 distribution, spatial and temporal 4:886
 population dynamics 4:886
 prevalence 4:886
 transmission 4:886
 genome 4:884–885
 almendravirus genome 4:885
 sigmavirus genome 4:885
 life cycle 4:885–886
 integration into the host genome 4:885–886
 pathogenesis 4:886–887
 virion structure 4:883–884
 insects, tetraviruses of *see* tetraviruses of insects
 insect-specific virus, defined 4:764
 insertional mutagenesis, defined 2:122
 in silico, defined 3:388
 in situ hybridization (ISH) 4:808, 2:190, 3:629, 2:79
 in-situ structural virology 1:238–240
 cellular electron cryotomography 1:238–240
 installation qualification (IQ) 5:69
 INSTI *see* integrase strand transfer inhibitors (INSTI)
 integrase, defined 4:387, 4:276, 4:380, 4:368
 Integrase function (Int), defined 3:96
 integrase inhibitors 2:473
 integrase strand transfer inhibitors (INSTI) 5:145–146
 first generation INSTI 5:146
 elvitegravir 5:146–149
 raltegravir 5:146
 second generation INSTI 5:149–150
 bictegravir 5:150–151
 cabotegravir 5:151
 dolutegravir 5:149–150
 virology 5:145–146
 integration, defined 2:643
 integrative and conjugative elements (ICEs) 1:613
 intein 4:677
 interactomics, defined 3:586
 intercellular adhesion molecule-1 (ICAM-1) 1:659, 1:266
 intercellular adhesion molecule 5 (ICAM-5) 1:410
 interferon alpha (IFN) 5:223, 5:220
 interferons 2:555–556
 defined 1:577
 interferon-stimulated genes (ISGs) 2:70
 interferon-stimulated response element (ISRE) 1:577
 intergenic region, defined 3:507, 4:768
 internal activation sequence (IAS) 4:78
 internal quality assessment (IQA) 5:74
 internal quality control procedures and monitoring assay performance over time 5:74–75
 internal ribosome entry site (IRES) 2:688, 3:692, 2:757–758, 2:757, 2:387, 2:335, 1:440, 2:386, 4:808, 4:699, 4:792, 1:447–449, 4:770–771, 4:771, 2:675
 IRES-like RNA structure 2:399
 IRES trans-activating factors (ITAFs) 2:337
 International Accreditation Forum (IAF) 5:70
 International Committee On Taxonomy Of Viruses (ICTV) 4:433, 1:47, 1:122, 1:32, 4:699
 organization 1:32
 taxonomic process 1:32–33
 International Health Regulations (IHR) 2:895
 International Laboratory Accreditation Cooperation (ILAC) 5:55, 5:70
 International Nucleotide Sequence Database Collaboration (INSDC) 5:33
 International Organization for Standardization (ISO) 5:55
 International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) 5:260
 International Standard, defined 5:52
 international surveillance 5:254
 International Unit, defined 5:52
 interspecies transmission, factors in 1:569–570
 epidemiological/ecological barriers 1:570
 host–pathogen interactions 1:570
 viral factors 1:570–571
 ability to adapt 1:571–572
 receptor specificity 1:570–571
 virulence 1:572
 intra-capsid DNA
 equilibrium energy and structure of 4:167–169
 metastability of 4:169–170
 intracellular annular viruses (IAV) 4:767
 intracellular budding 1:523–524
 intracellular dense viruses (IDV) 4:767
 intracellular enveloped virion (IEV) 2:165
 intracellular mature virion (IMV) 2:165
 Intracisternal A Particle (IAP) 1:82
 intraluminal vesicles (ILVs) 1:538, 1:529
 intranuclear bacilliform viruses (IBVs) 4:827
 intranuclear inclusion bodies (INIBs) 4:840
 intravenous immunoglobulin (IVIG) 5:156
 intravenous pathogenicity index (IVPI) 2:118–119, 2:117
 intrinsic resistance, defined 2:404
 introgression, defined 3:768
 intussusception, defined 5:289, 2:789
 invertebrate rhabdoviruses 4:716
 invertebrates, iridoviruses of 4:797
 classification of iridovirids 4:797–798
 genome organization and codon usage 4:800
 host range and pathology 4:799–800
 IIV-6 persistence and sensitivity to external factors 4:799

- induction/inhibition of apoptosis in infections 4:802
 morphology and composition 4:798–799
 promoter elements and transcriptional regulation 4:802
 transcriptional regulation 4:801–802
 viral entry, replication, and release strategy 4:800–801
 virion proteins 4:800
 invertebrates, parvoviruses of *see* parvoviruses: of invertebrates
 invertebrates, viruses of 4:700t–702
 arboviruses in their invertebrate vectors 4:722, 4:720t
 Artoviridae 4:703
 Ascoviridae 4:703
 Baculoviridae 4:703–705
 Bidnaviridae 4:705–706
 birnaviruses of invertebrates 4:706
 Dicistroviridae 4:706
 Hytrosaviridae 4:706–707
 Iflaviridae 4:707
 iridoviruses of invertebrates 4:707–708
 Malacoherpesviridae 4:708
 Mesoniviridae 4:708–709
 Nimaviridae 4:709
 Nodaviridae 4:709–710
 Nudiviridae 4:710
 Nyamiviridae 4:710–711
 parvoviruses of insects 4:711–713
 Polydnaviridae 4:712–713
 plant viruses in their invertebrate vectors 4:722, 4:721t
 poxviruses of invertebrates 4:713
 reoviruses of invertebrates 4:713–715
 retrotransposons associated with invertebrates 4:722
 rhabdoviruses of invertebrates 4:715–716
 Roniviridae 4:716
 Sarothroviridae 4:716
 Solinviviridae 4:716–717
 taxa of other viruses of invertebrates 4:717–722
 tetraviruses 4:717
 inverted terminal repeats (ITRs) 4:365, 4:43, 2:669, 4:387, 4:359
in vivo vs. *in vitro*, defined 1:71
 IPN virus *see* infectious pancreatic necrosis (IPN) virus
 IPV *see* inactivated polio vaccine (IPV)
 IQ *see* installation qualification (IQ)
 IQA *see* internal quality assessment (IQA)
 IRES *see* internal ribosome entry site (IRES)
 iridovirid, defined 4:797
 iridoviruses of invertebrates (*Betairidovirinae*) 4:707–708, 4:797
 classification of iridovirids 4:797–798
 genome organization and codon usage 4:800
 host range and pathology 4:799–800
 IIV-6 persistence and sensitivity to external factors 4:799
 induction/inhibition of apoptosis in infections 4:802
 Malacoherpesviridae 4:708
 Mesoniviridae 4:708–709
 morphology and composition 4:798–799
 Nimaviridae 4:709
 Nodaviridae 4:709–710
 Nudiviridae 4:710
 Nyamiviridae 4:710–711
 promoter elements and transcriptional regulation 4:802
 transcriptional regulation 4:801–802
 viral entry, replication, and release strategy 4:800–801
 virion proteins 4:800
 Iridoviridae 4:708
 ISARIC *see* International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC)
 Isfahan virus (ISFV) 2:875
 ISFV *see* Isfahan virus (ISFV)
 ISGs *see* interferon-stimulated genes (ISGs)
 ISH, *see in situ* hybridization (ISH)
 ISO *see* International Organization for Standardization (ISO)
 isogenic line, defined 3:388
 isolate, defined 3:8
 isothermal amplification 5:22
 i-spanin, defined 1:501
 Israeli acute paralysis virus (IAPV) 4:768–770, 4:771, 4:772, 4:773
 ISRE *see* interferon-stimulated response element (ISRE)
 ISVP *see* infectious subviral particle (ISVP)
 iterative helical reconstruction (IHR) 1:364–365
 Iterons 3:239
 ITRs *see* inverted terminal repeats (ITRs)
 iVDPV *see* immunodeficiency-associated VDPV (iVDPV)
 IVDPVs *see* immunodeficient vaccine-derived polioviruses (IVDPVs)
 IVIG *see* intravenous immunoglobulin (IVIG)
 IVPI *see* intravenous pathogenicity index (IVPI)
- J**
- Jaagsiekte sheep retrovirus (JSRV) 2:322
 classification 2:575
 enJSRVs 2:578–579
 enzootic nasal tumor virus (ENTVs) 2:581–582
 future perspectives 2:582
 genetic organization and virion proteins 2:575–576
 history 2:575
 ovine pulmonary adenocarcinoma (OPA) 2:579–580
 replication cycle 2:576–578
 virus-induced cell transformation, mechanisms of 2:580–581
 Jacalin-type lectin gene (JAX1) 3:623, 3:626–628, 3:629
 JAK inhibitors, defined 2:825
 JAM-1 *see* junctional adhesion molecule 1 (JAM-1)
 JAM-A *see* junctional adhesion molecule-A (JAM-A)
 Japanese encephalitis virus (JEV) 2:810–811
 classification 2:583
 clinical features 2:594
 diagnosis 2:595
 clinical diagnosis 2:595
 laboratory diagnosis 2:595
 epidemiology 2:591–593
 genetic diversity 2:593
 geographic distribution and epidemiologic patterns 2:592–593
 gene expression 2:586–587
 nonstructural proteins 2:587–589
 structural proteins 2:586–587
 genome structure 2:585–586
 life cycle 2:589–590
 viral entry 2:589–590
 viral morphogenesis 2:591
 viral translation and RNA replication 2:590–591
 pathogenesis 2:594–595
 prevention 2:595–596
 transmission 2:593
 non-vector-borne transmission 2:593–594
 vector-borne transmission 2:593
 treatment 2:595
 virion structure 2:583
 extracellular mature virion 2:583
 intracellular immature virion 2:583–584
 JAX1 *see* Jacalin-type lectin gene (JAX1)
 JCI *see* Joint Commission International (JCI)
 JC polyomavirus (JCPyV) 2:523–524, 5:109
 clinical features 2:523–524
 infectious cycle 2:524
 JCPyV *see* JC polyomavirus (JCPyV)
 JD *see* Jembrana disease (JD)
Jdvp5 gene 2:61t
 jelly-roll, defined 1:87, 4:457
 jelly-roll fold, defined 4:387, 4:359
 jelly-roll -barrel 1:262–263
 Jembrana disease (JD) 2:66
 JEV *see* Japanese encephalitis virus (JEV)
 JH *see* juvenile hormone (JH)
 JNK, defined 2:875
 Joint Commission International (JCI) 5:3
 JSRV *see* Jaagsiekte sheep retrovirus (JSRV)
 jumbo phages 4:229
 advanced capabilities 4:237–238
 DNA repair enzymes 4:238
 LPS biosynthesis 4:239
 NAD⁺ salvage pathway 4:238–239
 phage nucleus 4:239–240
 RNA polymerases 4:238
 evolution 4:240
 genome features 4:236
 genome nucleotide composition 4:236–237
 nucleotide modifications and substitutions 4:237
 ORFan genes 4:237
 terminal redundancies 4:236
 transfer RNA genes 4:237

jumbo phages (*continued*)
 history 4:229–230
 isolation and characterization 4:230–234
 virion structure 4:234
 head and tail fibers 4:235–236
 icosahedral geometries 4:234
 morphotypes 4:234
 virion DNA density 4:234–235
 junctional adhesion molecule 1 (JAM-1)
 1:410
 junctional adhesion molecule-A (JAM-A)
 1:308–309
 jungle yellow fever 2:894
 JUNV *see* Junín virus (JUNV)
 Junín virus (JUNV) 2:507, 2:512
 Jurona virus (JURV) 2:875
 JURV *see* Jurona virus (JURV)
 juvenile hormone (JH) 4:862, 4:858

K

Kalanchoe top spotting virus (KTSV) 3:165
 Kallithea virus (KV) 4:829
 Kaposi's sarcoma (KS) 2:604, 2:605, 2:606, 2:599
 classical Kaposi's sarcoma (cKS) 2:606
 HIV associated 2:606
 post-transplantation/iatrogenic 2:606
 Kaposi's sarcoma-associated herpesvirus (KSHV)
 classification 2:599
 clinical features 2:604
 Kaposi's sarcoma (KS) 2:604
 multicentric Castleman's disease (MCD) 2:604–605
 primary effusion lymphoma 2:604
 diagnosis 2:606
 Kaposi's sarcoma 2:606
 multicentric Castleman's disease (MCD) 2:606
 primary effusion lymphoma 2:606
 epidemiology 2:604
 genome 2:599
 life cycle 2:599–601
 binding and entry of the virus to the host cell 2:600–601
 genome segregation 2:603
 host cytoplasm, transport of virus in 2:601–602
 interaction of KSHV with the host immune system 2:603–604
 latent KSHV episome, replication of 2:602–603
 lytic activation 2:603
 nucleus, entry of viral DNA into 2:601–602
 pathogenesis 2:605
 Kaposi's sarcoma 2:605
 KSHV Inflammatory syndrome (KICS) 2:605–606
 multicentric Castleman's disease (MCD) 2:605
 primary effusion lymphoma 2:605
 prevention 2:607

treatment 2:606
 Kaposi's sarcoma 2:606
 multicentric Castleman's disease (MCD) 2:606–607
 primary effusion lymphoma 2:606
 vaccination 2:607
 virion structure 2:599
 Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8), defined 2:599
 Kaposi's sarcoma-associated virus (KSV)
 1:321
 Karelian fever 2:840
 Kashmir bee virus (KBV) 4:771, 4:772, 4:773
 KBLV *see* Kotalahti bat lyssavirus (KBLV)
 KBV *see* Kashmir bee virus (KBV)
 keratitis, defined 2:404
 Khujand lyssavirus (KHUV) 2:738
 KHUV *see* Khujand lyssavirus (KHUV)
 KICS *see* KSHV Inflammatory syndrome (KICS)
 killer immunoglobulin receptors (KIRs)
 1:584
 killer phenomena 4:513
 killer proteins, comparison of 4:517–518
 killer virus-infected yeast, self-protection in 4:542
 killer virus system, defined 4:534
 "Killing-the-Winner" model 1:638
 Kimberley virus (KIMV) 2:875
 KIMV *see* Kimberley virus (KIMV)
 kinetics of virus entry 4:390
 KIRs *see* killer immunoglobulin receptors (KIRs)
Kitaviridae 3:247
 knockdown, defined 3:123
 knockout, defined 3:123
 kobuviruses 1:279–280
 Koch's postulates, defined 3:355
 Koinobiont parasitoid, defined 4:849
 Koopinyah virus (KOOLV) 2:875
 KOOLV *see* Koopinyah virus (KOOLV)
 Kotalahti bat lyssavirus (KBLV) 2:738
 Kotonkan virus (KOTV) 2:875
 KOTV *see* Kotonkan virus (KOTV)
 KP1 4:516–517
 KP4 blocks L-type voltage gated Ca²⁺ channels 4:515
 KP4 protein 4:513–514
 atomic structure of 4:514f
 effect of KP4 on plants 4:515
 evolutionary origin of 4:516
 possible application of 4:515–516
 on *U. maydis* cells 4:514–515
 KP6 4:516
 atomic structure of 4:516, 4:517f
 KREMEN-1 *see* Kringle-containing transmembrane protein-1 (KREMEN-1)
 Kringle-containing transmembrane protein-1 (KREMEN-1) 1:284
 KS *see* Kaposi's sarcoma (KS)
 KSHV *see* Kaposi's sarcoma-associated herpesvirus (KSHV)
 KSHV Inflammatory syndrome (KICS) 2:605–606

KSV *see* Kaposi's sarcoma-associated virus (KSV)
 KTSV *see* Kalanchoe top spotting virus (KTSV)
 Kunjin virus 1:292–293
 Kupffer cells 2:895–896
 KV *see* Kallithea virus (KV)
 kyphosis, defined 2:34

L

L1 protein 2:499
 L2 protein 2:499
 labeling, virus 1:209–210
 laboratory biosafety 5:82–83
 biological agents 5:82–83
 biosafety cabinets 5:84–85
 decontamination and waste management 5:85–86
 laboratory accidents 5:86–87
 laboratory biosafety levels 5:83
 occupational health and special groups 5:87
 personal protective equipment (PPE) 5:84
 risk assessment 5:83
 routes of transmission and infective dose 5:83–84
 shipment of clinical specimens 5:87–88
 laboratory biosecurity 5:82, 5:88–89
 laboratory data 5:248–249
 laboratory developed tests (LDTs) 5:72, 5:60–61
 Laboratory Information Management System (LIMS) 5:65, 5:68–69
 laboratory specimen retention times 5:67
 LACA *see* last archaeal common ancestor (LACA)
 lactate dehydrogenase-elevating virus (LDV) 2:256
Lactobacillus casei 4:84
 La France Disease 4:528–529
 biological properties (virus host relationships) 4:529–530
 control 4:530–532
 diagnostics and identification 4:530
 epidemiology 4:530
 genome organization 4:529
 taxonomy and classification 4:529
 virion structure and composition 4:529
 Lagenaria mild mosaic virus (LaMMoV) 3:623, 3:623–624
 Lagos bat lyssavirus (LBV) 2:738
 LAIV *see* live attenuated influenza vaccines (LAIV)
 lake trout rhabdovirus (LTRV) 2:325
 LAM *see* lamivudine (LAM)
 lambda CI protein 4:81
 genome ejection 4:215
 lambda lysis, looking under the hood of 1:505–507
 lambda lysis cassette 1:505
 lambda lysogeny, maintenance of 4:95f
 lambda MGL, operational outline of 1:505
 lambda prophage as a lysis platform 1:504–505

- Red system, defined 4:291
 lambda system 4:126
 Escherichia Coli integration host factor 4:128
 TerL DNA packaging domain 4:127
 terminase enzyme of phage lambda 4:126
 TerL maturation domain 4:127–128
 TerS subunit 4:126–127
 lambda “switch” 4:94
 TerL maturation domain 4:127–128
 TerS subunit 4:126–127
 laminin-binding protein (LBP) 2:843
 lamivudine (LAM) 5:218
 LaMMoV *see* *Lagenaria mild mosaic virus* (LaMMoV)
 LAMP *see* loop-mediated isothermal amplification (LAMP)
 LANA *see* latency associated nuclear protein (LANA)
Laodelphax striatellus 4:771
 large icosahedral dsDNA viruses 1:269–271
 Lassa fever (LF) 2:507, 2:511, 2:513, 2:514
 Lassa virus (LASV) 2:507, 2:516
 last archaeal common ancestor (LACA) 1:17
 last bacterial common ancestor (LBCA) 1:17
 last eukaryotic common ancestor (LECA) 1:19
 LASV *see* Lassa virus (LASV)
 late gene, defined 2:441
 latency, defined 2:599, 2:306, 1:71, 2:528, 2:441
 latency associated nuclear protein (LANA) 2:602
 latency/latent infection, defined 5:197
 latent infection, defined 4:827, 2:404
 latent period, defined 4:314
 latent reservoir, defined 2:460
 lateral bodies, defined 4:858
 lateral flow assays (LFA) 5:94
 lateral flow detection, defined 3:839
 lateral flow immunoassay (LFIA) 3:629, 5:19
 lateral gene transfer 4:6
 latex agglutination 5:19–20
 LAV *see* live-attenuated vaccines (LAV)
 laves lattice 1:248, 1:253
 LBA *see* long branch attraction (LBA)
 LBCA *see* last bacterial common ancestor (LBCA)
 LBM *see* live bird markets (LBM)
 LBP *see* laminin-binding protein (LBP)
 LBV *see* Lagos bat lyssavirus (LBV)
 LCDV *see* lymphocystis disease virus (LCDV)
 LCGs *see* lysogenic conversion genes (LCGs)
 LCM *see* lymphocytic choriomeningitis (LCM)
 LCMV *see* lymphocytic choriomeningitis virus (LCMV)
 LDL *see* low-density lipoproteins (LDL)
 LDLR *see* low-density lipoprotein receptor (LDLR)
 LDTs *see* laboratory developed tests (LDTs)
 LDV *see* lactate dehydrogenase-elevating virus (LDV)
 leader and trailer, defined 3:567
 leafhopper-transmitted mastreviruses, case of 3:108
 leafhopper vectors 3:467–468
 leaky late (LL) viral gene 2:452
 leaky scanning, defined 3:528, 3:285, 3:839, 3:594
 leaky scanning of ribosomes, defined 3:712
 LECA *see* last eukaryotic common ancestor (LECA)
 lectin, defined 3:486
 legume viral diseases 3:92–94
 Leidenfrost Effect 5:5
 leiomyosarcoma, defined 2:316
 Lentinula edodes virus (LeV) dsRNA 4:627
 lentiviruses 2:56–57, 2:63–64, 2:64, 2:65, 2:56
 gene products of 2:61*t*
 lepidoptera, defined 4:867
 lepidopteran, defined 4:768
Leporid herpesvirus 4 2:734
 letermovir 2:458–459
 lethal mutagenesis, defined 1:53
 leucine-rich repeats (LRR) 1:606, 4:522
 leucopenia, defined 2:875
 leukemia, defined 2:528
 leukocytopenia, defined 2:362
 leukodepletion, defined 2:441
 LeV dsRNA *see* Lentinula edodes virus (LeV) dsRNA
 “lever” translocation model 4:154–155
 Levinthal’s Paradox in protein folding, defined 1:248
Leviviridae, single gene lysis (SGL) of 1:515–516
 LF *see* Lassa fever (LF)
 LFA *see* lateral flow assays (LFA)
 LFA-1 *see* lymphocyte function-associated antigen-1 (LFA-1)
 LFIA *see* lateral flow immunoassay (LFIA)
 LHF *see* Lujo hemorrhagic fever (LHF)
 life history, defined 1:62
Ligamenvirales 4:365–366
 evolutionary relationships 4:366
 light microscopy (LM) 1:495–496
 limbic encephalitis, defined 2:778
 limit of detection (LoD) 5:72
 LIMS *see* Laboratory Information Management System (LIMS)
 LIN *see* lysis inhibition (LIN)
 linker peptides 1:356
 lipidome (virus), defined 1:388
 lipid raft, defined 2:475, 2:56
 lipid transfer proteins (LTPs) 1:499
 lipobox, defined 1:501
 lipopolysaccharide (LPS) 1:556–557, 4:194, 4:175
 lipopolysaccharide, defined 1:402
Lipothrixviridae 1:368, 1:367–369
Lispiviridae 4:718
 live, attenuated vaccine, defined 2:797
 live attenuated influenza vaccines (LAIV) 5:301
 live-attenuated vaccines (LAV) 2:516, 2:20
 live bird markets (LBM) 2:118
 liver dysfunction 2:896
 liver necrosis 2:895–896
 living status of viruses 1:26–27
 Ljungan virus (LV) 1:279–280, 1:282
 LLEBV *see* Lleida bat lyssavirus (LLEBV)
 Lleida bat lyssavirus (LLEBV) 2:738
 LLoQ *see* Lower Limit of Quantitation (LLoQ)
 LL viral gene *see* leaky late (LL) viral gene
 LM *see* light microscopy (LM)
 lobule, defined 2:629
 LoD *see* limit of detection (LoD)
 long branch attraction (LBA) 1:121
 long contractile tails, baseplates of 4:190–191
 long-distance movement, defined 3:140
 long non-contractile tails, tail tip complex of 4:191
 long read sequencing (LRS) 1:182
 “long-read sequencing” parallel technologies 5:31
 long terminal repeats (LTR) 2:58–59, 3:96, 2:122
 different genera of LTR retrotransposons 3:100–101
 life cycle of LTRs retrotransposons 3:99–100
 long-term non progressors, defined 2:827
 loop-mediated isothermal amplification (LAMP) 5:98, 3:629, 3:759
 lopinavir 5:141
 lordosis, defined 2:34
 low-density lipoprotein receptor (LDLR) 1:395, 2:758–759, 2:390
 low-density lipoproteins (LDL) 2:386–387
 Lower Limit of Quantitation (LLoQ) 5:73–74
 lower respiratory tract infections (LRTI) 5:199, 5:200
 low pathogenic avian influenza virus (LPAI) 2:117
 LPAI *see* low pathogenic avian influenza virus (LPAI)
 L polymerase protein 2:752–753
 L protein
 defined 2:747, 3:833, 3:567
 encoded on RNA1 3:834–835
 of herbeviruses 4:766
 LPS *see* lipopolysaccharide (LPS)
 LRR *see* leucine-rich repeats (LRR)
 LRS *see* long read sequencing (LRS)
 LRTI *see* lower respiratory tract infections (LRTI)
 LSDV *see* Lumpy skin disease virus (LSDV)
 L segment 2:209–210
 LTPs *see* lipid transfer proteins (LTPs)
 LTR *see* long terminal repeats (LTR)
 LTR-retroelement, defined 3:653
 LTRV *see* lake trout rhabdovirus (LTRV)
 Lujo hemorrhagic fever (LHF) 2:511, 2:513
 Luminex GPP assay 5:102
 Lumpy skin disease virus (LSDV) 1:547, 2:165
 lumpy skin disease virus 2:168
 clinical features 2:168
 diagnosis 2:169
 epidemiology 2:168

- lumpy skin disease virus (*continued*)
 pathogenesis 2:168–169
 pathology 2:168
 prevention 2:169
 treatment 2:169
- Luteoviridae* 3:594
- luteoviruses (*Luteoviridae*) 3:447
 control 3:454–455
 diagnosis 3:454
 epidemiology 3:454
 evolutionary relationships 3:452
 genome organization and expression 3:449–452
 host range and transmission 3:452–453
 replication 3:453
 taxonomy and classification 3:447–448
 virion properties 3:448–449
 virion structure and composition 3:449
 virus–host relationships 3:453–454
- LV *see* Ljungan virus (LV)
- LVX 3:628
- lymphocystis disease virus (LCDV) 1:104
- lymphocyte function-associated antigen-1 (LFA-1) 1:601
- lymphocytic choriomeningitis (LCM) 2:511–512, 2:513–514, 2:514–515
- lymphocytic choriomeningitis virus (LCMV) 2:507, 2:512–513, 2:514, 1:445, 1:604
- lymphoid progenitor cells 1:587
- lymphoma, defined 2:528
- lymphopenia, defined 2:814, 2:825
- lyophilization 5:5
- Lyphogel method 5:11
- lysis 4:19, 4:50
 defined 1:621
- lysis inhibition (LIN) 1:510
- lysis-lysogeny decision 1:638
- lysogen, defined 4:242, 4:77, 4:368
- lysogenic, defined 4:276
- lysogenic conversion, defined 4:77
- lysogenic conversion genes (LCGs) 1:636
- lysogenic cycle, defined 4:252
- lysogenic life cycle, defined 4:387
- lysogenic or temperate phage, defined 4:283
- lysogenic pathway, defined 4:88
- lysogenic state, maintenance of 4:79–81
 bacteriophage 186 4:81–83
 bacteriophage 4:80–81
 host protein as immunity repressor 4:83–84
 integration-dependent bacteriophage immunity 4:83
 prophage induction 4:84–85
 RNA to maintain lysogeny 4:84
- lysogeny 4:77
 defined 1:621, 1:162, 4:276, 1:71
 DNA, persistence of 4:77–78
 extrachromosomal 4:78–79
 integration into the chromosome 4:77–78
 evolutionary and phenotypical effects of 4:85
 gene disruption 4:85
 genomic rearrangement 4:85–87
 lysogenic conversion 4:85
- lyssaviruses 5:260
 classification (compact) 2:738
 clinical features 2:740–741
 diagnosis 2:742–743
 epidemiology 2:739–740
 genome 2:738
 life cycle 2:738–739
 pathogenesis 2:741–742
 prevention 2:744–745
 treatment 2:743–744
 virion structure 2:738
- Lytic (virulent) phage, defined 4:98
- lytic, defined 4:276
- lytic cycle, defined 4:252
- lytic life cycle, defined 1:162, 4:387
- lytic pathway, defined 4:88
- lytic phage, defined 4:69, 4:283
- lytic transcription 4:69
 developmental pathways and transcription control 4:70–72
 by N4 4:74–75
 by T4 4:70–72
 by T7 4:72, 4:73*f*
 host RNA polymerase, transcription initiation and elongation by 4:69–70
 inhibition of host RNAP by T4, T7, and N4 4:75
- ## M
- M2-1 protein, defined 2:747
- M2-2 protein, defined 2:747
- MA *see* matrix protein (MA)
- machlomovirus
 classification 3:456
 diagnosis 3:459
 epidemiology 3:458–459
 genomes 3:456–457
 hosts and distribution 3:457–458
 life cycle 3:458
 pathogenesis 3:459
 prevention 3:459
 virion structure 3:456
- machrobrachium rosenbergii nodavirus (MrNV) 4:820, 4:716
- Machupo virus (MACV) 2:507, 2:512
- Macrobrachium nipponense reovirus 4:879
- macrodomein 2:177
- macrolides 5:170
- Macronovirus* 4:889
- macrophage, defined 2:56
- macropinocytosis 1:533
 defined 1:529
 defined 2:22
- macroptilium yellow spot virus (MaYSV) 3:194, 3:561
- maculaviruses 3:824
 properties and distinguishing characteristics 3:824
- MACV *see* Machupo virus (MACV)
- Madariaga virus (MADV) 2:40
- Madin-Darby canine kidney (MDCK) cells 1:531
- MADV *see* Madariaga virus (MADV)
- maize chlorotic mottle virus (MCMV) 3:792–793
- maize streak disease (MSD) 3:565–566
 control of 3:467
 epidemiology 3:467
- maize streak virus (MSV) 3:461, 3:468, 1:649
 complementary sense genes (*Rep* and *RepA*) 3:465
 diversity and evolution 3:462–463
 future threat 3:468
 genome organization 3:463
 history and taxonomy 3:461
 host range and symptoms 3:461–462
 leafhopper vectors 3:467–468
 long intergenic region 3:463–465
 molecular biology of 3:465–466
 replication 3:465–466
 particle assembly and movement 3:466–467
 particle structure 3:463
 plant hosts 3:468
 resistance to 3:565–566, 3:565*f*
 short intergenic region 3:465
 transmission 3:463
 virion sense genes (MP and CP) 3:465
- major capsid protein (MCP) 1:374, 4:266–267, 2:442–443, 4:115, 1:153, 1:252–253, 1:319, 1:322–323, 1:325
- major coat protein (MCP) 1:337
- major histocompatibility complex (MHC) 2:455, 1:584, 5:190
 defined 2:441, 2:778
 MHC-I antigen presentation 1:603
 MHC-II antigen presentation 1:603–604
- major immediate early promoter (MIEP) 2:443–445, 2:452
- major Interspersed Genomic Element (MIGE) 4:677
- major tail protein (MTP) 1:320
- Malacoherpesviridae* 4:708
- MALDI MS *see* matrix-assisted laser desorption/ionization (MALDI) MS
- MALDI-TOF mass spectrometry *see* matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry
- Malpais Spring virus (MSP) 2:875
- malpighian cells, defined 2:316
- malpighian tubules, defined 4:858
- mammalian cells 1:664
- mammarenavirus genome organization and proteins 2:507–509
- mammarenavirus life cycle 2:509–510, 2:509*f*
 assembly and budding 2:511
 cell attachment and entry 2:509–510
 expression and replication of the viral genome 2:510–511
- Mammarenavirus L proteins 2:508
- mammarenavirus RNA replication and gene transcription 2:510*f*
- mammarenavirus virion structure 2:507
- Mammarenavirus Z proteins 2:508–509
- MAPK *see* mitogen activated protein kinases (MAPK); mitogen-activated phosphate kinases (MAPKs)

- mapping of epitopes, defined 3:727
- marafiviruses 3:822–823
- infection and transmission 3:824
- properties and distinguishing characteristics 3:823
- virion structure, genome organization, replication cycle, and phylogenetic relationships 3:823–824
- maravirc 5:151–152
- Marburg and Ravn viruses (MARV) 2:611*f*, 2:611*t*, 2:875
- classification 2:609
- clinical features 2:613
- diagnosis 2:615
- epidemiology 2:608–609
- genome 2:610–611
- glycoprotein (GP) 2:612
- life cycle 2:613
- MARV L 2:613
- nucleoprotein (NP) 2:611–612
- pathogenesis 2:613–615
- prevention 2:616–618
- treatment 2:615–616
- viral protein 24 (VP24) 2:612–613
- viral protein 30 (VP30) 2:612
- viral protein 35 (VP35) 2:612
- viral protein 40 (VP40) 2:612
- virion structure 2:609–610
- virulence factors for 2:614*f*
- Marburg hemorrhagic fever (MHF) 2:232
- Marburg virus disease (MVD) 2:232
- Marek's disease (MD) 2:113
- Marek's disease virus (MDV) 2:652–653, 2:113
- maribavir 2:458–459
- marine microbial ecology, RNA viruses in 4:674–675
- marine phage ecology 4:337–338
- auxiliary metabolic genes in marine phages 4:340
- diel rhythms of phage infections in the marine environment 4:339–340
- marine phages as factors driving bacterial mortality and diversity 4:338–339
- phage micro- and macro-diversity in the marine environment 4:337–338
- transfer RNAs (tRNAs) 4:340–341
- marine phages 4:322
- genomic diversity of 4:324–325
- linking phage genomes with host identity 4:329–330
- long read viromics 4:326–328
- metagenome assembled viral genomes (MAVGs) 4:328–329
- phage marker genes 4:324–325
- single virus genomics 4:325–326
- ssDNA phages in the marine environment 4:330
- viral contigs from fosmid libraries 4:325
- viral metagenomics (viromics) 4:325
- morphological diversity of 4:322–324
- marine protists, RNA viruses infecting 4:671
- marine *Pseudoalteromonas* spp. phage PM2 (*Corticoviridae*) 4:208
- marine RNA virus quasispecies 4:674
- marker-assisted selection (MAS) 3:557
- Marnaviridae* 1:279
- marine RNA virus quasispecies 4:674
- RNA viruses infecting marine protists 4:671
- RNA viruses in marine microbial ecology 4:674–675
- taxonomic framework, development of 4:671–674
- MARV *see* Marburg and Ravn viruses (MARV)
- massively parallel/deep sequencing *see* next generation sequencing (NGS)
- mass spectrometry, defined 4:252
- Mastadenovirus* 2:8
- Mastrevirus* 3:358–359, 3:413
- mathematical modeling 1:566–567
- data sharing, data privacy and ethics 1:567
- defined 1:559
- proof of causation 1:567–568
- mathematical modeling of virus architecture 1:249
- genetic economy, principle of 1:249
- Hamiltonian paths approach, applications of 1:254–255
- assembly code embedded within viral genetic message 1:254–255
- icosahedral symmetry, beyond 1:255
- virus assembly mechanisms 1:255
- icosahedral capsid architecture, mathematical models of 1:249–250
- 3D, virus structure in 1:254
- implications for viral evolution 1:253–254
- non-quasi-equivalent trihexagonal lattice architectures 1:253
- overarching framework 1:251–253
- practical applications 1:251
- predictions and limitations of quasi-equivalence theory 1:250
- quasi-equivalence theory 1:249–250
- viral tiling theory 1:250–251
- icosahedral symmetry, viral geometry and 1:249
- matrix-assisted laser desorption/ionization (MALDI) MS 5:92
- matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry 5:114
- matrix protein (MA) 2:325–326, 3:574–575, 2:566, 1:353, 1:354–355, 1:382, 1:349, 2:57, 5:127
- maturation, defined 1:382, 4:115
- maturation protein 1:238
- mature extracellular virions 2:891–892
- MAVGs *see* metagenome assembled viral genomes (MAVGs)
- MAVS *see* mitochondrial antiviral signaling protein (MAVS)
- maximum likelihood (ML) 1:119–120, 2:183*f*
- maximum parsimony (MP) 1:119–120
- MaYSV *see* macroptilium yellow spot virus (MaYSV)
- MBV *see* monodon baculovirus (MBV); mushroom bacilliform virus (MBV)
- MBV1 *see* mulberry badnavirus 1 (MBV1)
- MC29, defined 2:122
- MCD *see* multicentric Castleman's disease (MCD)
- MCF viruses *see* mink cell focus-inducing (MCF) viruses
- MCM *see* minichromosome maintenance (MCM)
- MCMV *see* maize chlorotic mottle virus (MCMV)
- MCP *see* major capsid protein (MCP); major coat protein (MCP)
- MCPyV *see* Merkel cell polyomavirus (MCPyV)
- MCS proteins 1:325
- MCV *see* molluscum contagiosum virus (MCV); mud crab virus (MCV)
- MD *see* Marek's disease (MD)
- MDA *see* multiple displacement amplification (MDA)
- MDA5 *see* melanoma differentiation-associated gene 5 (MDA5)
- MDCK cells *see* Madin-Darby canine kidney (MDCK) cells
- MdSGHV transmission dynamics in the housefly 4:789
- MDV *see* Marek's disease virus (MDV)
- MEAM1 *see* Middle East Asia minor 1 (MEAM1)
- measles 5:268–269
- measles, mumps, and rubella (MMR) vaccine 5:268
- measles virus
- classification 2:619
- clinical features 2:623–625
- complications 2:624–625
- immune responses 2:625
- immunosuppression 2:625
- persistent infections 2:625
- diagnosis 2:626
- epidemiology 2:623
- genome 2:619–622
- envelope complex 2:622–623
- replicative complex 2:621–622
- transcription and replication 2:622
- life cycle 2:623
- MeV-based vaccines and cancer therapeutics 2:627
- pathogenesis 2:625–626
- prevention 2:627
- treatment 2:626–627
- virion structure 2:619
- mechanical transmission, defined 1:542
- MED1 *see* Mediterranean 1 (MED1)
- Mediterranean 1 (MED1) 3:757
- megabirnaviruses (*Megabirnaviridae*) 4:594, 4:630–631, 4:464, 4:508–509
- biological properties 4:596–597
- future perspectives 4:600
- functions of megabirna-P3 and -P4 4:600
- genome expression and replication 4:594–595
- genome organization 4:594
- taxonomic and phylogenetic considerations 4:597–600
- virion properties 4:594

- megabirnaviruses (*Megabirnaviridae*)
 (*continued*)
 virion structure 4:595–596
 megaphages, origin of 1:20–22
 megavirales, defined 1:372
 melanization, defined 4:849
Melanoconion 2:810
 melanoma differentiation-associated gene 5
 (MDA5) 1:455, 2:555
Melanoplus sanguinipes EPV (MSEV) 4:861,
 4:865
 meleira 4:662
 membrane-containing icosahedral DNA
 bacteriophages 4:36–37
 genomes and genomics 4:43
 life cycle 4:40–42
 cell lysis 4:42–43
 genome replication 4:42
 particle assembly 4:42
 receptor recognition and DNA delivery
 4:42
 virion structure and properties 4:37–40
 membrane and DNA 4:40
 overall structure 4:37–40
 membrane insertion, defined 4:53
 membranes and membrane proteins 1:339–
 340
 membranous web 2:388–389
 memory inflation 2:458
 Mendelian gene 2:62
 meningitis, defined 2:884, 2:675
 meningoencephalitis, defined 2:675
 meristem tip-culture, defined 3:430
 Merkel cell polyomavirus (MCPyV) 2:524–525
 clinical features 2:524–525
 infectious cycle 2:525
 MERS *see* Middle East respiratory syndrome
 (MERS)
 MERS-CoV *see* Middle Eastern respiratory
 syndrome coronavirus (MERS-CoV)
Mesoniviridae 4:804, 4:708–709
mesoniviruses
 classification 4:804
 genome 4:804–806
 genome organization 4:806f
 life cycle 4:807
 ORF1a/ORF1b 4:806
 particle morphology and spike protein
 architecture 4:806f
 structural proteins 4:806–807
 virion structure 4:804
 mesopelagic layer, defined 4:322
 messenger RNA (mRNA) 1:437, 1:382, 2:35
 metagemone, defined 4:265, 4:776
 metagenome assembled viral genomes
 (MAVGs) 4:328–329, 4:409
 cultivated phages infecting main groups of
 4:330–332
 marine *Alphaproteobacteria*, phages of
 4:332–334
 marine *Bacteroidetes*, phages of 4:336–337
 marine *Cyanobacteria*, phages of 4:331–332
 marine *Gammaproteobacteria*, phages of
 4:335–336
 metagenomically characterized viruses,
 classification of 1:50
 metagenomics 1:127, 4:342, 1:177–178,
 1:621, 4:314, 1:133, 4:283, 1:184
 applied to viruses 1:133
 pioneering viral metagenomics, one
 gene at a time 1:133
 scaling up from fragmented genes to
 complete genomes 1:133–134
 future of 1:139–140
 global viral diversity, characterizing 1:137
 identifying globally dominant
 bacteriophages 1:137
 leveraging time series to track virus
 populations dynamics 1:138–139
 ssDNA and RNA viruses, revealing the
 extraordinary diversity of 1:138
 unveiling new uncultivated giant viruses
 1:137–138
 viral metagenomics in the clinic 1:134–
 136
 epidemiological surveillance and
 environmental monitoring 1:136–
 137
 metagenomic discovery of new viral
 pathogens 1:136
 metal-tagging TEM (METTEM) 1:497
 metaphor, virus as 1:672
 metapneumovirus *see* human
 metapneumovirus (hMPV)
 metapopulation, defined 4:419
 metastability, defined 4:167
 metastable, defined 1:388
 metaviridae (Ty3/gypsy retrotransposons)
 3:101
Metaviridae 3:654
 metavirus, defined 3:653
 methanogenic, defined 4:387, 4:368
 Methanosarcina Spherical Virus (MetSV)
 1:433–434
 methyl-directed mismatch repair (MMR)
 4:291
 MetSV *see* Methanosarcina Spherical Virus
 (MetSV)
 METTEM *see* metal-tagging TEM (METTEM)
 MEV *see* mink enteritis virus (MEV)
 MGEs *see* mobile genetic elements (MGEs)
 MGL *see* multigene lysis (MGL)
 MHC *see* major histocompatibility complex
 (MHC)
 MHF *see* Marburg hemorrhagic fever (MHF)
 MHV *see* murine hepatitis virus (MHV)
 microarray, defined 4:808
 microbial loop 4:320–321, 4:321f
 microcephaly, defined 2:899
 microfluidics 5:115
 micromyelia, defined 2:34
 microRNAs (miRNA) 2:743–744, 2:144,
 2:48, 3:293, 2:602, 1:455, 3:52,
 3:123, 3:563
 microRNA-122 (miR122) 2:391
 microtiter serum neutralization (MTSN)
 2:134
 microtubule organizing center (MTOC)
 1:498
 microtubules, defined 3:32
 microvesicle, defined 1:529
Microviridae 1:268–269
 capsid assembly pathway in 1:270f
 microviruses 4:207
 Middle East Asia minor 1 (MEAM1) 3:359,
 3:757
 Middle Eastern respiratory syndrome
 coronavirus (MERS-CoV) 5:274–275,
 2:193, 1:121, 5:259
 ebola and other hemorrhagic fevers
 5:274–275
 experimental infections 5:275
 Middle East respiratory syndrome (MERS)
 2:815f, 2:816f, 1:12
 animal infection models 2:821
 clinical features 2:819–820
 epidemiology
 ecology, animal reservoir and zoonotic
 transmission 2:818–819
 human-to-human transmission 2:819
 laboratory diagnosis 2:820–821
 pathogenesis 2:822
 treatment 2:822–823
 vaccines and immunity 2:823
 virology 2:814–815, 2:815, 2:815–816
 phylogeny 2:816–818
 virus receptors 2:818
 MIEP *see* major immediate early promoter
 (MIEP)
 MIGE *see* major Interspersed Genomic
 Element (MIGE)
 Mild Neurocognitive Disorder (MND) 2:471
 mild-strain cross-protection, defined 3:520
 Milker's nodule 2:668
 milt, defined 2:544
 MIMIVIRE, defined 1:372
Mimiviridae family, algal viruses belonging
 to a subgroup within
 general properties 4:678–681
 history 4:677–678
 proposed subfamily *Mesomimivirinae*
 4:681
 unclassified algae-infecting members in
 the family *Mimiviridae* 4:681–682
 mimivirus 1:493, 1:224
 mimivirus virophage resistant element
 1:380–381
Mimoreovirus 4:684–685
 minichromosome, defined 3:158
 minichromosome maintenance (MCM)
 4:390–391, 2:602
 minigenome, defined 2:747
 Minimum Information about an
 Uncultivated Virus Genome
 (MIUViG) 1:127
 mink cell focus-inducing (MCF) viruses
 2:646
 mink enteritis virus (MEV) 2:684
 minor coat proteins 1:333
 protein IIIa 1:333
 protein IX 1:333–335
 protein VI 1:333
 protein VIII 1:333
 minor variant, defined 1:71
 (–)-sense RNA (minus-sense RNA) 1:382
 miRNA *see* microRNAs (miRNA)
 mites, plant viruses transmission by 3:113
 mitochondria 1:495, 1:499

- mitochondrial antiviral signaling protein (MAVS) 1:455, 2:556, 2:391–392
- mitogen-activated phosphate kinases (MAPKs) 2:265
- mitogen activated protein kinases (MAPK) 3:245, 2:428
- mitoviruses 4:465–466, 4:454
 accessory RNAs associated with infections of 4:602
 codon usage and implications for biology and evolution of 4:605
 engineering mitoviruses for infectivity 4:604
 genome structure 4:601–602
 host defense 4:604–605
 origin and evolution 4:605
 phenotypic effects of mitovirus infection 4:602
 phylogenetic relationships 4:602–603
 plant mitoviruses 4:605–606
 taxonomy and nomenclature 4:603–604
 transmission 4:604
- mitoxantrone dihydrochloride 5:124
- MIUViG *see* Minimum Information about an Uncultivated Virus Genome (MIUViG)
- ML *see* maximum likelihood (ML)
- MLA *see* multilateral recognition arrangements (MLA)
- MMR *see* methyl-directed mismatch repair (MMR)
- MMR vaccine *see* measles, mumps, and rubella (MMR) vaccine
- MMV *see* mouse mammary tumor virus (MMV); multi-membraned vesicles (MMVs)
- MND *see* Mild Neurocognitive Disorder (MND)
- mobile genetic elements (MGEs) 1:14, 4:98, 1:38, 4:400
- modified vaccinia ankara (MVA) 2:243, 2:216
- modular evolution, defined 4:457
- MOI *see* multiplicity of infection (MOI)
- Mokola lyssavirus (MOKV) 2:738
- MOKV *see* Mokola lyssavirus (MOKV)
- molecular clock, defined 3:520
- molecular diagnostics, defined 4:252
- molecular epidemiology 1:565–566, 1:382
- molecular strain typing, defined 2:707
- molluscum contagiosum virus (MCV)
 clinical features 2:631
 diagnosis 2:632
 epidemiology 2:630–631
 genome 2:629–630
 history 2:629
 life cycle 2:630
 pathogenesis 2:631–632
 prevention 2:633
 treatment 2:632–633
 virion structure 2:629
- Moloney leukemia murine virus (MoMLV) 2:576
- monkeypox virus (MPXV)
 classification and structural morphology 2:868–869
- clinical features 2:870–871
 diagnosis 2:872
 epidemiology 2:871–872
 history 2:868
 pathogenesis 2:869–870
 prevention 2:872–873
 treatment 2:873–874
 viral life cycle 2:869
- mono-, bi- and tripartite, defined 4:632
- monocistronic, defined 3:516, 4:632
- monoclonal antibodies 2:230, 5:272–273, 2:778, 2:441
 human immunodeficiency virus 5:273
 respiratory syncytial virus (RSV) 5:272–273
 therapy 2:241–242
- monocot, defined 3:778
- monocotyledon, defined 3:545
- monodon baculovirus (MBV) 4:828–829
- monomeric replication products 4:137
- Mononegavirales* 2:232, 3:495
- mononucleosis, defined 2:441
- monopartite, defined 3:703
- monophyletic, defined 1:28, 3:653
- monophyly, defined 4:835
- mono-/polyvalent, defined 5:289
- monosense, defined 4:835
- monovalent type 2 oral polio vaccine (mOPV2) 5:312–313
- Mopeia virus (MOPV) 2:516
- MOPV *see* Mopeia virus (MOPV)
- mOPV2 *see* monovalent type 2 oral polio vaccine (mOPV2)
- morbilliviruses
 classification 2:68
 clinical features 2:74–75
 diagnosis and prevention 2:76–77
 epidemiology and host range 2:72–73
 aquatic mammal morbilliviruses 2:74
 canine distemper virus (CDV) 2:74
 emerging morbilliviruses 2:74
 Peste-des-petitsruminants virus (PPRV) 2:73–74
 rinderpest virus (RPV) 2:73
 genome 2:69–70
 life cycle 2:73f, 2:71–72
 pathogenesis 2:75–76
 proteins 2:70–71
 rinderpest eradication 2:77–78
 treatment 2:77
 virion structure 2:68–69, 2:69f
- morphogenetic signal, defined 4:53
- morphologies, viruses with 4:360–361, 1:258f
 family *Ampullaviridae* 4:361
 family *Bicaudaviridae* 4:361
 family *Clavaviridae* 4:361–362
 family *Fuselloviridae* 4:362
 family *Guttaviridae* 4:362
 family *Spiraviridae* 4:362
 filamentous viruses 4:363–364
 proposed class *Tokiviricetes* 4:363–364
 spherical viruses 4:362
 Family *Globuloviridae* 4:362
 Family *Ovaliviridae* 4:362
 Family *Portogloboviridae* 4:362–363
- Family *Turriviridae* 4:363
- mosaicism 4:277
 in *Myoviridae* 4:280
 in *Podoviridae* 4:280
 in *Siphoviridae* 4:278–280
- mosquitoes 2:893
- mosquito vectors, control of 1:575
- most recent common ancestor (MRCA) 1:117–118
- motif, defined 3:797
- mourning doves (*Zenaidura macroura*) 2:805
- mouse mammary tumor virus (MMV) 1:352, 2:643
- movement of viruses in plants 3:32–33
 intercellular transport 3:39–40
 intracellular movement 3:33–39
 systemic transport 3:40–41
 virus movement and plant defense responses 3:41
- movement protein (MP) 3:317–318, 4:437–438, 3:727, 3:229, 3:719, 3:140, 3:456
- MP *see* maximum parsimony (MP); movement protein (MP)
- MpDV *see* myzus persicae densovirus (MpDV)
- M protein 2:891–892
- MPXV *see* monkeypox virus (MPXV)
- MRCA *see* most recent common ancestor (MRCA)
- mRNA *see* messenger RNA (mRNA)
- MrNV *see* machrobrachium rosenbergii nodavirus (MrNV)
- MSD *see* maize streak disease (MSD)
- M segment 2:209
- MSEV, *see* *Melanopus sanguinipes* EPV (MSEV)
- MSPV *see* Malpais Spring virus (MSP)V
- MSV *see* maize streak virus (MSV)
- MTOC *see* microtubule organizing center (MTOC)
- MTP *see* major tail protein (MTP)
- MTSN *see* microtiter serum neutralization (MTSN)
- mud crab virus (MCV) 4:772
- mulberry badnavirus 1 (MBV1) 3:163
- Muller's ratchet, defined 1:53
- multicentric Castleman's disease (MCD) 2:604–605, 2:605, 2:606, 2:606–607, 2:599
- multigene families, defined 2:22
- multigene lysis (MGL) 1:501–502, 1:501, 1:504–505
 diversity 1:512
 in gram-positive hosts 1:513–514
 lambda lysis, looking under the hood of 1:505–507
 lambda lysis cassette 1:505
 lambda MGL, operational outline of 1:505
 lambda prophage as a lysis platform 1:504–505
 regulation of 1:509–510
 dual start holin genes 1:509–510
 real-time regulation of holin function 1:510–512
- multilateral recognition arrangements (MLA) 5:70

- multi-membraned vesicles (MMVs) 1:498–499
- multiphoton imaging 1:211–213
- multiple displacement amplification (MDA) 4:330, 1:184
- multiplex and point of care tests (POCT) 5:114–115
- multiplex PCR 5:92, 5:98
- multiplex RT-PCR 3:629
- multiplicity of infection (MOI) 4:83, 4:90–91, 4:88
- multivesicular bodies (MVB) 1:522, 2:210–211, 3:32, 2:399, 3:132, 2:879–880
- multivesicular endosomes (MVEs) 1:538, 1:529
- MuLVs/MLVs *see* murine leukemia virus (MuLVs/MLVs)
- mumps virus
 - antibody response 2:641
 - clinical features of primary infection 2:637
 - detection of mumps virus 2:641
 - differential diagnosis 2:640
 - epidemiology 2:637
 - geographic and seasonal distribution 2:637
 - transmission and tissue tropism 2:637
 - evolution and genetics 2:637
 - genome, properties of 2:635
 - history 2:634
 - immune response 2:639–640
 - laboratory diagnosis 2:640–641
 - MuV proteins, properties of 2:635
 - neutralisation tests 2:641
 - pathogenesis 2:639
 - physical properties 2:635–636
 - prevention and control 2:642
 - future perspectives 2:642
 - properties of the proteins of 2:636*t*
 - replication 2:636
 - serologic relationships and variability 2:636–637
 - structure of 2:634*f*
 - systemic manifestations 2:637
 - cardiac symptoms 2:639
 - genitourinary tract 2:638–639
 - musculoskeletal system 2:639
 - neurological disease 2:637–638
 - pancreas 2:638
 - parotid glands 2:637
 - taxonomy and classification 2:634
 - utilization of assays 2:641–642
 - primary infection and primary vaccine failure 2:641–642
 - secondary infection and secondary vaccine failure 2:642
 - virion, properties of 2:634–635
- MuNoVs *see* murine noroviruses (MuNoVs)
- mungbean yellow mosaic viruses (MYMV), resistance to 3:563, 3:564*f*
- murine astroviruses (MuAstV) 2:96
- murine hepatitis virus (MHV) 4:804, 1:492
- murine leukemia virus (MuLVs/MLVs) 1:631, 1:627, 1:178, 1:223, 1:454
 - classification 2:643
 - endogenous 2:647
 - genome and structure 2:643–644
 - history 2:643
 - host-virus interactions 2:645–646
 - pathology 2:646
 - replication 2:644–645
 - vectors 2:646–647
- murine noroviruses (MuNoVs) 1:535
- Murine Sarcoma Viruses (MuSVs/MSVs)
 - pathology 2:646
- Murray Valley encephalitis virus 2:812
- Muscavir* 4:781
- mushroom bacilliform virus (MBV) 4:437
 - MBV evolutionary relationships 4:550
 - MBV genome organization and expression 4:549–550
 - MBV transmission and host range 4:550
 - MBV virion properties 4:549
 - MBV virion structure and composition 4:549
- mushroom compost, defined 4:528
- MuSVs/MSVs *see* Murine Sarcoma Viruses (MuSVs/MSVs)
- Mutational Meltdown, defined 5:227
- mutation rate, defined 1:53
- mutation-selection balance 1:57–58, 1:53
- mutualism, defined 4:419, 4:658
- MVA *see* modified vaccinia ankara (MVA)
- MVB *see* multivesicular bodies (MVB)
- MVD *see* Marburg virus disease (MVD)
- MVEs *see* multivesicular endosomes (MVEs)
- Mxra8 2:176
- myalgia, defined 2:675
- Myc, defined 2:122
- Mycoflexivirus* 4:437–438
- mycoreovirus 1 (MyRV1) 4:658
- mycoreovirus 2 (MyRV2) 4:433
- mycoreoviruses 4:433
 - coinfections of mycoreovirus and other viruses 4:613
 - on fungal gene expression 4:612–613
 - genome rearrangements 4:613
 - genome structures, organizations, and relationships 4:609–610
 - mycoreovirus 1 4:610–611
 - expression of MyRV1 gene products 4:610–611
 - mycoreovirus 2 4:611
 - mycoreovirus 3 4:611–612
 - sclerotinia sclerotiorum reoviruses 4:612
 - mycoreovirus 4 4:612
 - sclerotinia sclerotiorum reovirus 1 4:612
 - structure-function relationships 4:607–609
 - taxonomy and nomenclature 4:609
- mycorrhiza, defined 3:388
- mycoviruses 4:441–442, 4:461, 3:388, 4:432, 4:439, 1:621, 4:544, 4:431, 4:594, 4:648
 - coinfection and evolution 4:459–460
 - diversity in *Aspergilli* 4:450
 - aspergillus foetidus* mycovirus complex 4:453–454
 - chrysovirus 4:454
 - genomes 4:451–453
 - naviruses and mitoviruses 4:454
 - partitiviruses 4:454
 - phenotypes 4:455
 - polymycoviruses 4:454–455
 - prevalence 4:450–451
 - transmission 4:451
 - evolutionary history and drivers of evolution 4:458–459
 - horizontal transmission of 4:525–526
 - host divergence in the evolution 4:459
 - infections in *Sclerotinia sclerotiorum* 4:461, 4:462*t*
 - co-infecting hypovirulent strain SCH941 4:463–464
 - hypovirulent strain SZ-150 4:463
 - megabirnavirus 4:464
 - mitoviruses 4:465–466
 - possible reasons for co-infections 4:466
 - strain AH98 4:463
 - strain Ep-1PN 4:461–462
 - strain sunf-M 4:464–465
 - strain SX247 4:462–463
 - three strains of *S. sclerotiorum* harbor multiple unassigned mycoviruses/dsRNA elements 4:466
- structural evolution 4:459–460
- mycoviruses with filamentous particles 4:478
 - Botrytis virus F (BotV-F) 4:478–479
 - biological properties 4:479
 - family –*Gammaflexiviridae* 4:479
 - genome structure 4:479
 - genus –*Mycoflexivirus* 4:479
 - phylogenetic relationships 4:479–480
 - virion morphology 4:479
 - Botrytis Virus X (BotV-X) 4:480
 - biological properties 4:480–481
 - family –*Alfalexiviridae* 4:480
 - genome structure 4:480
 - genus –*Botrexvirus* 4:480
 - phylogenetic relationships 4:481
 - virion morphology 4:481
- colletotrichum camelliae filamentous virus 1 4:483–484
 - biological properties 4:484
 - family –*Unclassified* 4:483–484
 - genome structure 4:483–484
 - genus –*Unclassified* 4:483–484
 - phylogenetic relationships 4:484
 - virion morphology 4:484
- fusarium graminearum negative-stranded RNA virus 1 4:483
 - biological properties 4:483
 - family –*Myomonaviridae* 4:483
 - genome structure 4:483
 - genus –*Unclassified* 4:483
 - particle morphology 4:483
 - phylogenetic relationships 4:483
- relationships between filamentous viruses from fungi and plants 4:484–485
- sclerotinia sclerotimonavirus 4:481
 - biological properties 4:481
 - family –*Myomonaviridae* 4:481
 - genome structure 4:481
 - genus –*Sclerotimonavirus* 4:481
 - phylogenetic relationships 4:483
 - virion morphology 4:481–483
- mycovirus-mediated biological control 4:468

- basic concepts for 4:468–469
 future perspectives 4:477
 general procedure for developing 4:469
 biocontrol testing in the field 4:472
 biological control agents, mycoviruses as 4:472
 biosafety issues 4:472
 testing for phenotypic effects of the mycovirus 4:469
 transmission properties 4:469–472
 virus detection and characterization 4:469
 hypovirulence in *Botrytis cinerea* 4:476
 hypovirulence in *Cryphonectria parasitica* 4:472–474
 artificial application of hypovirulence 4:474–475
 exclusive transmissible hypovirulence 4:473–474
 hypovirulence in *Fusarium graminearum* 4:477
 hypovirulence in *Helminthosporium victoriae* 4:476–477
 hypovirulence in *Ophiostoma novo-ulmi* 4:475
 hypovirulence in *Rosellinia necatrix* 4:476
 hypovirulence in *Sclerotinia sclerotiorum* 4:475–476
 plant pathogenic fungi, disease cycle of 4:468
 mycovirus transmission 4:525
 between VCGs 4:526–527
 myelomonocytic lineage cells 2:456
Mymonaviridae 4:615
 genomic structure 4:617–618
 host and distribution 4:618–620
 impact on the host 4:618
 phylogenetic status of 4:615
 species and tentative species in 4:615
 virion 4:615–617
 mymonaviruses, defined 4:615
 myositis, defined 2:675
Myoviridae 1:320–321, 4:186, 4:276, 4:280
 myovirus 4:342
 myristoylation, defined 2:757, 2:22
 myxomatosis in European rabbits 2:732
 myxoma virus 2:731
 myzus persicae densovirus (MpDV) 4:845
- N**
- N4 (lytic podoviridae phage) 4:74–75
 NA *see* neuraminidase (NA)
 nAChR *see* nicotinic acetylcholine receptor (nAChR)
 NAC transcription factors 3:758
 NAI *see* neuraminidase inhibitors (NAI)
 nanobiotechnology, defined 3:364
 nano-carriers, virus-like particles as 1:665–666
 conjugation of antigens and VLPs 1:665–666
 chemical and affinity conjugation 1:666–667
 genetic fusion 1:665–666
 SpyTag/SpyCatcher system 1:667
 Encapsulation, loading of virus-like particles by 1:667–668
Nanohaloarchaea, viruses of 4:417
 nanoparticle, defined 3:778
 nanopore sequencing, defined 1:175
 Nanopore Sequencing Technology, defined 4:322
 nanotechnology, defined 3:364
Nanoviridae 3:471
 nanoviruses (*Nanoviridae*) 3:470–471
 classification and taxonomy 3:471–472
 diagnosis 3:479–480
 epidemiology and control 3:478–479
 genome organization 3:472–474
 nanovirid-associated alphasatellites 3:476–477
 nanovirid evolution 3:476
 nanovirid replication 3:474–475
 particle properties 3:472
 properties and functions of nanovirid proteins 3:474
 taxonomy and phylogeny 3:475–476
 transmission and host range 3:477–478
 host range 3:478
 tissue tropism and means of transmission 3:477–478
 transmission by aphids 3:478
 Narcissus mosaic virus (NWV) helix 3:625
 narnaviruses (*Narnaviridae*) 4:621, 4:454
 cis-acting signals for replication 4:623–624
 cis-signals for formation of ribonucleoprotein complexes 4:624–625
 generation of narnaviruses *in vivo* 4:623
 historical background 4:621
 narnavirus persistence in the host 4:625–626
 replication intermediates 4:623
 ribonucleoprotein complexes as a viral entity 4:622–623
 viral genomes 4:621–622
 NASBA *see* nucleic acid sequence-based amplification (NASBA)
 NASH assay *see* nucleic acids spot hybridization (NASH) assay
 NAT *see* nucleic acid testing (NAT)
 National Immunization Days (NIDs) 2:690–692
 natural killer (NK) cells 1:584
 naturally occurring recessive resistance 3:70–73
 natural transformation
 defined 4:242
 of a plant virus to a fungus 4:447
 Natural Vector method 1:102
 Naïve, defined 1:542
 NB 2:566
 NB-LRR protein, defined 3:60, 3:554
 NC *see* nucleocapsid (NC); nucleocapsid protein (NC)
 NCAM *see* neuronal cell adhesion molecule (NCAM)
 NCLDV *see* nucleocytoplasmic large DNA viruses (NCLDV)
- NCVOG *see* Nucleo-Cytoplasmic Virus Orthologous Groups (NCVOG)
 NDV *see* Newcastle disease virus (NDV)
 near-atomic resolution, cryo-EM at 1:233–235
 data collection and image reconstruction 1:233–235
 near-atomic resolution maps, interpretation of 1:235
 side chain interactions and virus assembly 1:235
 NEAR technology *see* nicking enzyme amplification reaction (NEAR) technology
 necro-like viruses (*Tombusviridae*)
 cytopathology 3:484
 genome organization and expression 3:481–483
 intraspecific relationships 3:483
 satellites 3:483–484
 taxonomy and classification 3:481
 transmission and host range 3:484
 virion properties, structure and composition 3:481
 necrotic syndrome of lettuce, defined 3:833
nef gene 2:61t
 negative-contrast electron microscopy, defined 4:632
 negative predictive value (NPV) 5:74
 negative selection-based enrichment 1:179–180
 Negative sense (–) strand, defined 1:429
 negative sense genome, defined 3:507
 negative-sense RNA viruses 1:499, 3:567
 negative single stranded RNA viruses (NSRVs) 1:345–346
 classification of NSRVs and genome organization 1:346
 general features of NSRV virion structures 1:346
 inside the virion 1:348–349
 matrix proteins 1:349
 nucleoproteins 1:349–350
 phosphoproteins 1:350–351
 polymerase 1:350
 viral ribonucleoproteins/nucleocapsids assemblies 1:349
 virion surface, structure of 1:346–348
 accessory proteins in the virion surface 1:348
 structure and function of virion glycoproteins 1:347–348
 Negative Stain 5:5
 negative staining (NS) 5:9, 5:9–10
 for diagnosis of fecal samples 5:11
 agarose slides, preparation of 5:11
 general protocol for the study of viruses by 5:10–11
 neighbor-joining (NJ) method 1:118–119
 nematodes, plant viruses transmission by 3:113
 NendoU function 2:264
Neodiprion sertifer 4:704
 neonatal herpes simplex virus infection 5:178–180
 neoplasia, defined 2:316

- neotope, defined 3:727
 NEP *see* nuclear export protein (NEP)
 nepoviruses (*Secoviridae*)
 cell-to-cell and systemic movement in the plant 3:491
 diseases, economic considerations and control 3:493
 genome organization and function of viral proteins 3:489–490
 historical perspective 3:486
 host range, symptom determinants and interaction with plant defense responses 3:491
 population structures 3:493
 production of viral proteins 3:490
 taxonomy, phylogeny, and evolution 3:486
 transmission 3:491–493
 viral RNA replication 3:490–491
 virion structure 3:486–489
 neuraminidase (NA) 2:566, 5:125–127
 neuraminidase inhibitors (NAI) 5:162–164, 5:163f, 5:171
 neurodevelopmental, defined 2:441
 neuroinvasion 1:536
 neuroinvasive disease, defined 2:884
 neuronal cell adhesion molecule (NCAM) 2:738–739
 neuronal progenitor cells (NPCs) 2:906
 neuropilin-2 2:450–451
 neuropsychiatric adverse events (NPAE) 5:164
 neurotropic disease 2:891, 2:896
 neutralization 1:591
 neutralizing antibodies 5:286
 Newcastle disease virus (NDV) 1:191
 classification 2:648
 clinical features 2:651
 diagnosis 2:651–652
 epidemiology 2:650
 genome 2:648–649
 life cycle 2:649–650
 pathogenesis 2:650–651
 prevention 2:652–653
 virion structure 2:648
 new invertebrate DVs 4:840
 New-World (NW) begomoviruses 3:751
 New World (NW) mammarenaviral HFs 2:515, 2:514
 next generation sequencing (NGS) 2:673, 5:76, 5:112–113, 2:652, 4:414, 1:553
 analyzing recombination in sequences derived from 1:111–112
 and applications in virology 1:176
 defined 1:552, 3:293, 1:175
 N-fold symmetric assembly 4:186
 NF-B inhibitors 2:672
 defined 2:875
 NF-B signaling pathway *see* nuclear factor kappa B (NF-B) signaling pathway
 N-Glycosylation, defined 4:380
 NGS *see* next generation sequencing (NGS)
 NHPs *see* nonhuman primates (NHPs)
 nicking enzyme amplification reaction (NEAR) technology 5:46
Nicotiana benthamiana 3:753, 3:244f, 3:244–245, 3:556f
 nicotinic acetylcholine receptor (nAChR) 2:738–739, 2:741–742
Nidovirales 4:804
 nidoviruses 2:193
 diseases associated with nidoviruses 2:256–257
 effect of nidovirus infection on the host cell 2:264–265
 genome organization 2:258–262
 replicase 2:260–262
 host cell, effect of nidovirus infection on 2:264–265
 origin of nidoviruses 2:264
 origin of nidoviruses 2:264
 replicase 2:260–262
 replication 2:263
 structural and accessory protein genes 2:262–263
 taxonomy and phylogeny 2:256
 transcription 2:263–264
 virus structure 2:257–258
 nidovirus
 classification and prototype members 2:257f
 genome structure 2:260f
 phylogeny 2:258f
 replicase genes 2:261f
 structure 2:259f
 Nidovirus endoribonuclease (NendoU) 2:261
 NIDs *see* National Immunization Days (NIDs)
 Niemann-Pick C1 (NPC1) 2:237, 2:613
 NIENV *see* Nilaparvata lugens endogenous nudivirus (NIENV)
Nilaparvata lugens 4:771
 Nilaparvata lugens endogenous nudivirus (NIENV) 4:829
Nimaviridae 4:808, 4:709
 nimaviruses 4:808
 apoptosis 4:815–816
 clinical features and pathology 4:816–817
 epidemiology 4:816
 genome and phylogeny 4:809–813
 hemocyte responses to WSSV infection 4:816
 life cycle 4:813–815
 protection of shrimp against WSSV infection 4:817
 taxonomy 4:808
 virion structure and composition 4:808–809
 90–90–90 Initiative, defined 2:460
 Nipah virus (NiV) 2:355, 1:121
 NIS *see* sodium iodide symporter (NIS)
 nitazoxanide 5:168
 NiV *see* Nipah virus (NiV)
 NJ method *see* neighbor-joining (NJ) method
 NK cells *see* natural killer (NK) cells
 NLRs, defined 4:520
 NLS *see* nuclear localization sequences (NLS); nuclear localization signal (NLS)
 NM infection, defined 3:743
 NNRTIs *see* non-nucleoside reverse-transcriptase inhibitors (NNRTIs)
 Noble Prizes awarded for discoveries relating to viruses 1:674t
 NOBs *see* non-occluded baculoviruses (NOBs)
Nodaviridae 4:709–710
 nodaviruses of invertebrates and fish
 classification (compact) 4:819
 proposed six clade taxonomic structure 4:819–820
 provisional nodaviruses 4:819
 clinical features 4:822–823
 viral covert mortality disease (VCMD) 4:823
 viral nervous necrosis (VNN) 4:822–823
 white tail disease (WTD) 4:823
 diagnosis 4:824–825
 viral covert mortality disease (VCMD) 4:825
 viral nervous necrosis (VNN) 4:824–825
 white tail disease (WTD) 4:825
 epidemiology 4:822
 genome 4:820
 immune responses to nodavirus infection 4:823–824
 adaptive immune responses in nodavirus infection 4:824
 innate immune responses 4:823–824
 life cycle 4:821–822
 pathogenesis 4:823
 physical properties 4:821
 prevention 4:825
 viral covert mortality disease (VCMD) 4:825
 viral nervous necrosis (VNN) 4:825
 white tail disease (WTD) 4:825
 proteins 4:820–821
 capsid proteins 4:821
 proteins B1 and B2 4:821
 RNA-dependent RNA polymerase (RdRp) 4:820–821
 treatment 4:825
 virion structure 4:820
 NOD-like receptor family pyrin domain containing 3 (NLRP3) 2:555
 nomenclature, defined 1:28
 nonanucleotide, defined 3:301
 non-canonical initiation of virus mRNA translation 1:447–449
 cap-independent translation enhancers (CITEs) 1:449
 initiation at non-AUGs 1:450
 internal ribosome entry sites (IRESes) 1:447–449
 leaky scanning 1:450
 ribosome reinitiation 1:450–452
 ribosome ‘shunting’ (discontinuous scanning) 1:450
 virus alternatives to components of initiation 1:449–450
 alternative to eIF4F 1:450
 alternative to the m⁷G cap structure 1:449–450
 nonchromosomal gene, defined 4:487
 non-circulative, defined 3:371

- non-contractile long-tailed phages (*Siphoviridae*) 4:209–211
- non-enveloped eukaryotic virus 1:409–410
- complex 1:412–414
 - exploitation of virus entry pathway 1:414–415
 - simple 1:410–412
- non-enveloped icosahedral RNA viruses 1:265–266
- nonenveloped particles
- antigenic sites 1:478
 - antiviral agents 1:478–479
 - architecture of viruses 1:475
 - assembly 1:477–478
 - evolution 1:477
 - helical viruses, atomic structure of 1:475–476
 - host receptor recognition site 1:478
 - nucleic acid–protein interaction 1:476–477
 - spherical viruses, atomic structure of 1:476
 - structure determination, methods of 1:475
- non-enveloped viruses, cell entry targeting of 5:238
- nonhuman and environmental sources, surveillance of viruses in 5:249
- nonhuman primates (NHPs) 2:232, 2:232–234
- non-insect vectors, plant viruses
- transmission by 3:113
 - fungi and protists, transmission by 3:113
 - mites, transmission by 3:113
 - nematodes, transmission by 3:113
- non-isothermal PCR based technologies 5:22–25
- nonlytic release of virus in extracellular vesicles 1:538
- non-NB-IRR resistance 3:66–67
- non-nucleoside reverse-transcriptase inhibitors (NNRTIs) 2:472–473, 5:135, 5:121
- clinical uses of 5:135–136
 - drug-drug interactions with 5:136
 - mechanisms of NNRTI resistance 5:136–137
 - and their mechanism of action 5:135
 - toxicities 5:136
- non-occluded baculoviruses (NOBs) 4:827
- non-pathogen derived resistance 3:363
- non-persistent, defined 3:371
- non-persistent manner, defined 3:81
- non-pharmaceutical interventions, defined 2:825
- non-propagative manner, defined 3:81
- non-propagative transmission, defined 3:200
- non-quasi-equivalent trihexagonal lattice architectures 1:253
- non-retroviral endogenous RNA (NERVE) 4:601
- non-retroviral EVE, evolutionary impact of 1:83–84
- nonsense suppressor tRNA, defined 4:487
- non-structural (NS) proteins 1:498, 2:296, 2:40–41, 2:806–808
- NS1 2:567
 - NS5A 5:123
 - NSP3 1:446
- nonsynonymous, defined 1:62
- non-template strand, defined 4:69
- non-translated regions (NTRs) 2:387
- nOPV strains *see* novel oral poliovirus vaccine (nOPV) strains
- norovirus 2:483, 1:392–393
- clinical features 2:487
 - diagnosis 2:488–489
 - epidemiology 2:486–487
 - genome 2:483–484
 - immunity 2:488
 - life cycle 2:485–486
 - pathogenesis 2:487–488
 - prevention 2:489–490
 - treatment 2:489
 - viral classification 2:483
 - virion structure 2:484–485
 - vaccine 5:291–292
 - development 5:292
 - future considerations for 5:293
 - in human clinical trials 5:292–293
 - in pre-clinical development 5:293
- nosocomial transmission 1:562
- defined 1:559
- notifications 5:248
- novel oral poliovirus vaccine (nOPV) strains 5:313
- NP *see* nucleoprotein (NP)
- NPAE *see* neuropsychiatric adverse events (NPAE)
- NPC1 *see* Niemann-Pick C1 (NPC1)
- NPCs *see* neuronal progenitor cells (NPCs)
- NPV *see* negative predictive value (NPV); nucleopolyhedrovirus (NPV)
- NRTIs *see* nucleoside/nucleotide reverse-transcriptase inhibitors (NRTIs)
- NS *see* negative staining (NS)
- NS1 protein, defined 2:747
- NS2A 2:844
- NS2B 2:844
- NS2 protein, defined 2:747
- ns3 gene 4:761
- NS proteins *see* non-structural (NS) proteins
- N-Succinimidyl S-Acetylthioacetate (SATA) 1:667
- N terminus, defined 1:345
- NTRs *see* non-translated regions (NTRs)
- nuclear export protein (NEP) 2:552, 2:290, 2:567
- nuclear factor kappa B (NF- κ B) signaling pathway 2:672
- nuclear inclusion protein a, defined 3:293
- nuclear inclusion protein b, defined 3:293
- nuclear localization sequences (NLS) 2:289
- nuclear localization signal (NLS) 2:602, 3:164, 2:561
- nuclear shuttle protein, defined 3:169
- nuclease, defined 4:242
- nucleic acid 1:229f, 5:5, 4:242
- nucleic acid detection 5:210–211
- hepatitis A virus 5:211
 - hepatitis E virus 5:211
- nucleic acid–protein interaction 1:476–477
- nucleic acid satellites, defined 3:411
- nucleic acid sequence-based amplification (NASBA) 5:98, 5:22, 5:23f
- nucleic acid sequence dependent strategies for known viruses 5:31–32
- nucleic acid sequence independent strategies for sequencing 5:32
- nucleic acids spot hybridization (NASH) assay 3:759
- nucleic acid testing (NAT) 5:17, 5:98, 5:91–92
- clustered regularly interspaced short palindromic repeats (CRISPR) platforms 5:93
 - DNA microarrays 5:92
 - high throughput sequencing 5:92–93
 - multiplex PCR 5:92
 - polymerase chain reaction assays 5:92
 - VirCapSeq-VERT 5:93
- Nucleo(s)tide analog (NA), defined 5:217
- nucleocapsid (NC) 2:610, 1:491, 4:747, 4:827, 1:257, 1:362, 2:747, 1:382, 4:26, 3:507, 2:441
- nucleocapsid core, defined 3:567
- nucleocapsid protein (NC) 1:353, 3:572–573, 1:356, 2:57
- nucleocytoplasmic large DNA viruses (NCLDV) 1:621, 1:493, 1:372, 1:379f, 4:677, 1:159, 1:72, 4:797, 1:269
- Nucleo-Cytoplasmic Virus Orthologous Groups (NCVOG) 4:677
- nucleopolyhedrovirus (NPV) 4:739, 4:747, 4:699
- nucleoprotein (NP) 2:611–612, 2:648, 3:154, 1:349–350, 2:552
- nucleos(t)ide analogue, defined 2:373
- nucleoside/nucleotide reverse-transcriptase inhibitors (NRTIs) 5:132–133, 5:133, 2:472–473, 5:140, 5:121
- clinical uses of 5:133
 - drug-drug interactions with 5:134
 - mechanisms of NRTI resistance 5:134–135
 - new drugs 5:135
 - and their mechanism of action 5:132–133
 - toxicities 5:133–134
- nucleus, virocells with 1:25
- nudaurelia capensis omega virus (NV) 4:900, 4:901–902
- nudaurelia capensis virus (NV) 4:897
- Nudiviridae* 4:827, 4:710
- members of 4:827–829
- nudiviruses 4:828t, 4:710
- classification 4:827
 - criteria for classification 4:827
 - history 4:827
 - members of *Nudiviridae* 4:827–829
 - core genes 4:832
 - defined 4:827
 - gene regulation for switching productive and latent infections 4:832–833
 - genome 4:830–832
 - gene content 4:832
 - general features 4:831–832
 - negative impacts and potential applications 4:833
 - phylogeny 4:829

nudiviruses (*continued*)
 transmission and pathogenesis 4:829–830
 virion structure 4:829
 virus life cycle 4:830
 Number and Brightness (N&B) 00108:
 p0130 1:208
 NWV helix *see* Narcissus mosaic virus
 (NWV) helix
 Nyamiviridae 4:710–711
 nyamiviruses 4:711
 nystagmus, defined 2:34

O

OAEV *see* Oedaleus asiaticus EPV (OAEV)
 OaPV-3 DNA *see* ovis aires papillomavirus
 type 3 (OaPV-3) DNA
 OAS *see* original antigenic sin (OAS)
 oat chlorotic stunt virus (OCSV) 3:790–791
 OB *see* occlusion body (OB)
 obligate endoparasite, defined 3:528
 Obodhiang virus (OBOV) 2:875
 OBOV *see* Obodhiang virus (OBOV)
 occlusion body (OB) 4:739, 4:747
 occlusion body, defined 4:827
 occlusion-derived virus (ODV) 4:704, 4:739,
 4:747
 Ockelbo disease 2:840
 OCSV *see* oat chlorotic stunt virus (OCSV)
 ocular infections 5:200
 odonstyle, defined 3:486
 odontophore, defined 3:486
 ODV *see* occlusion-derived virus (ODV)
 Oedaleus asiaticus EPV (OAEV) 4:865
 OELCA, *see* Okra enation leaf curl
 alphasatellite (OELCA)
 OELCV, *see* Okra enation leaf curl virus
 (OELCV)
 Okra enation leaf curl alphasatellite (OELCA)
 3:360–361
 Okra enation leaf curl virus (OELCV) 3:358–
 359
 okra viral diseases 3:88
 Old-World (OW) 3:751
 Old World primates, defined 2:827
 Oleavirus 3:260, 3:262*t*
 characteristics of RNA genome in 3:261*t*
 oligo-mediated recombineering, mechanism
 of 4:297–298
 oligonucleotide, defined 4:291
 oligonucleotide primer, defined 5:27
Olpidium virulentus 3:833
 Omegatetravirus 4:897
 omp, defined 4:175
 oncogene, defined 2:316, 2:122
 oncogenesis, defined 2:528
 oncogenic HPVs 2:499–500
 oncolytic virotherapy, defined 2:619
 oncolytic viruses (OVs) 1:658–659, 1:658
 combination therapies 1:660
 modes of action 1:659
 tumor targeting by OVs 1:659–660
 oncoproteins, defined 2:79
 126-kDa protein, defined 3:839
 onychomadesis, defined 2:245
 oomycete, defined 3:388
 OPA *see* ovine pulmonary adenocarcinoma
 (OPA)
 open reading frame (ORF) 2:184, 4:677,
 2:757, 2:79, 3:388, 3:778, 3:140,
 4:776, 2:779, 4:845, 3:163
 defined 2:182
 ORF1 2:48–49
 operational taxonomy units (OTUs) 1:117,
 3:653
Ophiostoma novo-ulmi, hypovirulence in
 4:475
 ophiocorvix (Aspiviridae) 3:495
 cytopathology 3:505
 diagnosis 3:503–504
 genome organization and replication
 3:497–498
 pathogenicity and geographic distribution
 3:500–503
 properties and functions of gene products
 3:498–500
 taxonomy, phylogeny and evolution
 3:495–497
 transmission, experimental host-range and
 control 3:504–505
 virion structure 3:497
 opportunist, defined 2:441
 opportunistic pathogen, defined 2:441
 opsonization 1:591
 Opuntia virus X (OpVX) 3:623–624, 3:628
 OPV *see* oral poliovirus vaccine (OPV)
 OpVX *see* Opuntia virus X (OpVX)
 OR1411 2:450–451
 oral cavity and respiratory tract viromes
 1:555–556
 oral poliovirus vaccine (OPV) 5:310–311
 orbivirus 1:303
 orchidodinia, defined 2:675
 ORF *see* open reading frame (ORF)
 ORF1a/ORF1b 4:806
orfA gene 2:61*t*
 ORFan genes 4:237, 4:229, 4:528
 ORFV, *see* *orf virus* (ORFV)
orf virus (ORFV) 2:169, 2:666
 clinical features and gross pathology
 2:169–170
 diagnosis 2:170
 epidemiology 2:169
 histopathology 2:170
 pathogenesis 2:170
 prevention 2:170
 treatment 2:170
 organelle remodeling proteins 1:499
 original antigenic sin (OAS) 2:419
 origin of assembly, defined 3:734
 origin of viruses 1:14
 DNA viruses 1:19–20
 evolutionary connection between viruses
 and mobile genetic elements 1:14
 giant viruses and megaphages 1:20–22
 reverse-transcribing RNA and DNA viruses
 1:18–19
 RNA viruses 1:14–15
 infecting eukaryotes 1:17–18
 scenario for 1:16–17
 selfish RNA replicons, origin of 1:15–16
 Orinoco sheldgoose (OSHBV) 2:100
 ornamentals and orchids viral diseases
 3:95–96
 ornamental viruses 3:19
 causal agent and classification 3:19
 control 3:20
 disease symptoms and yield losses 3:19
 epidemiology 3:19–20
 geographical distribution 3:19
 ortervirales, defined 3:653
 orthobunyaviruses 2:654
 classification 2:654
 phylogenetic analysis by sequencing
 2:654
 serological classification 2:654
 clinical features and pathogenesis 2:662
 Anopheles A 2:662
 Bunyamwera 2:662
 Bwamba 2:662
 California 2:662–663
 Simbu 2:663
 diagnosis 2:663
 polymerase chain reaction (PCR)
 detection of viral RNA 2:663–664
 serological assays 2:663
 epidemiology 2:660–661
 host and viral determinants of disease-
 species barriers 2:661–662
 orthobunyavirus outbreaks and tracking
 2:661
 surveying viral endemic ranges and
 tracking viral spread 2:661
 viral discovery 2:660–661
 genome 2:654–659
 genome and encoded proteins 2:655–
 659
 life cycle 2:659–660
 reassortment 2:660
 effects of 2:660
 evidence of 2:660
 transmission 2:660
 vector 2:660
 vertebrate hosts 2:660
 treatment and prevention 2:664
 orthohantavirus
 structure of 2:350*f*
 types of 2:352*t*
 orthologous genes, defined 4:387
 orthopteran, defined 4:768
 orthoreovirus 1:305–306, 4:871
 orthospoviruses (*Tospoviridae*) 3:12
 detection and diagnosis 3:514
 genetics and evolution 3:513–514
 genome properties 3:511–512
 protein properties 3:512
 geographic distribution 3:510
 history 3:507
 host-range and virus propagation 3:510–
 511
 virion properties 3:510–511
 pathogenicity and cytopathology 3:512–
 513
 prevention and control 3:514–515
 future perspectives 3:515
 replication 3:512

taxonomy and classification 3:507–510
 transmission and epidemiology 3:513
 oryzaviruses 3:551, 3:547–548
 Oseltamivir 5:170, 5:164, 5:171
 plus Favipiravir 5:171
 plus monoclonal antibodies 5:171
 OSHBV *see* Orinoco sheldgoose (OSHBV)
 osmolyte, defined 4:167
 osmotic pressure, defined 4:167
 o-spanin, defined 1:501
 OTUs *see* operational taxonomy units (OTUs)
 ourmiaviruses 3:516
 diagnosis 3:518
 epidemiology and control 3:518
 genome organization 3:516
 properties and functions of gene products 3:516–517
 replication and propagation 3:517–518
 taxonomy, phylogeny and evolution 3:516
 transmission, host range 3:518
 virion structure 3:516
 virus–host relationships 3:518
 outbreak, defined 1:569, 5:247
Ovaliviridae 4:362
 overarching framework for icosahedral architectures 1:251–253
 overwintering 2:893
 ovine pulmonary adenocarcinoma (OPA) 2:575, 2:579–580
 oviposition, defined 3:461, 4:849
 ovis aires papillomavirus type 3 (OaPV-3) DNA 2:87
 OVs *see* oncolytic viruses (OVs)
 OW *see* Old-World (OW)
 Oxford Nanopore technologies 1:181–182

P

P1 protein 4:28
 P2 protein 4:28–30
 P3 movement protein 3:574
 P4 protein 4:30
 P7 protein 4:30
 P8 protein 4:30
 p24 protein 2:145–148
 p75 neurotrophin receptor (p75NTR) 2:738–739
 PA *see* protective antigen (PA); protein antibiotics (PA)
pac, defined 4:61
 Pacheco's parrot disease (PPD) 2:113
 Pacheco's parrot disease virus (PPDV) 2:113
 Pacific Biosciences 1:181
 packaging, defined 4:26, 4:136
 packaging, termination of 4:132–133
 terminase ejection and virion completion 4:133–134
 unit length packaging motors 4:133
 packaging signal, defined 1:248, 1:488
 packaging signal mediated assembly, defined 1:248, 1:488
 packaging signals (PSs) 1:249, 1:254–255
 PACR protein *see* poxvirus APC/cyclosome regulator (PACR) protein

PaFlV1 *see* penicillium aurantiogriseum foetidus-like virus 1 (PaFlV1)
pag1 miRNAs 4:832–833
 PAGE *see* polyacrylamide gel electrophoresis (PAGE)
 pAGO *see* prokaryotic Argonaute (pAGO)
 pairwise evolutionary distances (PED) 4:804
 pairwise sequence comparison (PASC) 1:103f, 2:507
 advantages of 1:104–105
 description and application 1:102–104
 limitations of 1:105–107
 sequence-based virus classification methods 1:101–102, 1:105f
 DEmARC (DivErsity pArtitioning by hieRarchical Clustering) 1:101–102
 Genome Relationships Applied to Virus Taxonomy (GRAViTy) 1:102
 Natural Vector method 1:102
 Sequence Demarcation Tool (SDT) 1:102
 ViCTree 1:102
 ViPTree 1:102
 sequence comparison methods 1:100–101
 virus classification methods 1:100
 PaLCV, *see* Papaya leaf curl virus (PaLCV)
 Palenque lineage 2:809
 paleovirology 1:79
 endogenous retroviruses, evolutionary impact of 1:81–83
 extant viral families, calibrating the long-term evolutionary history of 1:81
 host-virus interactions 1:85–86
 ecology of 1:84–85
 non-retroviral EVE, evolutionary impact of 1:83–84
 patterns and mechanisms underlying viral integration 1:79–81
 palivizumab 5:157, 5:124, 2:754, 5:273
 PAM *see* protospacer adjacent motifs (PAM)
 PAM-dependent DNA interference, defined 4:400
 PAMPs *see* pathogen-associated molecular patterns (PAMPs)
 PAMV *see* Potato aucuba mosaic virus (PAMV)
 pandemic, defined 1:569, 5:247, 3:21, 3:301
 pandemic influenza 5:160
 pandoravirus 1:257
 panicoviruses
 classification 3:456
 diagnosis 3:459
 epidemiology 3:458–459
 genomes 3:456–457
 hosts and distribution 3:457–458
 life cycle 3:458
 pathogenesis 3:459
 prevention 3:459
 virion structure 3:456
Papanivirus 3:581, 3:582–584
Papaya leaf curl virus (PaLCV) 3:357
Papaya mosaic virus (PapMV) 3:628
papaya ringspot virus (PRSV) 3:91–92
 diagnostics 3:523–524
 general properties and genome of 3:521
 history and taxonomy 3:520

host range, symptomatology, and geographic distribution 3:521
 molecular determinants for biological properties 3:523
 "PRSV cluster" 3:520–521
 sequence diversity and evolution;
 molecular clock analyzes 3:521–523
 transmission and epidemiology 3:521
 virus control 3:524
 breeding for resistance 3:524
 cross-protection 3:524
 cultural practices: prophylactic measures 3:524
 pathogen-derived resistance for controlling PRSV 3:524–525
 transgenic resistance 3:525
 papaya sticky disease 4:662
 papillomas 2:83–84, 2:79
 cats 2:85
 diagnosis 2:89–90
 dogs 2:85
 horses 2:84–85
 ruminants 2:83–84
 treatment 2:90
 papillomaviruses 2:493, 2:86
 see also human papillomavirus (HPV)
 classification 2:79–81
 clinical features 2:83–84
 diagnosis 2:88–90
 papillomas 2:89–90
 sarcomas 2:90
 squamous cell carcinomas and other neoplasms 2:90
 viral plaques 2:90
 epidemiology 2:83
 genome 2:81–82
 life cycle 2:82–83
 neoplasms 2:88
 papillomas 2:83–84
 cats 2:85
 dogs 2:85
 horses 2:84–85
 ruminants 2:83–84
 pathogenesis 2:88
 prevention 2:91
 sarcomas 2:87–88
 cats 2:88
 horses 2:87–88
 squamous cell carcinomas 2:86–87
 cats 2:87
 dogs 2:87
 horses 2:87
 rabbits 2:87
 ruminants 2:86–87
 treatment 2:90
 papillomas 2:90
 squamous cell carcinomas and sarcomas 2:90–91
 viral plaques 2:90
 viral plaques 2:85
 cats 2:85–86
 dogs 2:85
 horses 2:85
 virion structure 2:81
 PapMV *see* Papaya mosaic virus (PapMV)
 papule, defined 2:868

- paraesthesia, defined 2:738
 parainfluenza virus (PIV) 1:191
 classification 2:502
 clinical features 2:504–505
 diagnosis 2:505
 epidemiology 2:503–504
 genome 2:502
 life cycle 2:502–503
 pathogenesis 2:505
 treatment and prevention 2:505–506
 virion structure 2:502
 parallel sequencing technologies 5:29–31
 application of NGS in virology 5:30–31
 paralog, defined 3:69
 Paramecium bursaria chlorella virus 1 (PBCV-1) 1:269–270, 1:270
 Paramyxovirus 5:260
 paraphyletic clade, defined 3:653
 paraphyly, defined 4:835
 parapoxviruses (PPVs) 2:666, 2:165
 clinical features 2:667–668
 bovine papular stomatitis (BPS) 2:668, 2:668f
 Milker's nodule, paravaccinia, pseudocowpox 2:668
 orf, scabby mouth, contagious pustular dermatitis (CPD), ecthyma contagiosum 2:667–668
 red deer, parapox of 2:668
 diagnosis 2:673
 genomes 2:669
 host range, epidemiology, and virus propagation 2:666–667
 immune response to infection 2:673
 known or putative genes involved in pathogenesis and virulence 2:670–671
 ankyrin repeat (ANK)/F-box genes 2:672
 apoptosis, inhibitor of 2:672
 chemokine binding protein (vCBP) 2:671–672
 NF-B Inhibitors 2:672
 poxvirus APC/C regulator (PACR) 2:672
 VEGF-E 2:670–671
 viral dUTPase 2:671
 viral GM-CSF inhibitory factor (GIF) 2:671
 viral IL-10 2:671
 viral interferon resistance protein 2:671
 oncolytic potency 2:674
 ORFV vector vaccines 2:673–674
 pathogenesis and the host immune response to infection 2:672–673
 physical properties 2:668
 prevention and control 2:674
 of red deer in NZ 2:171
 serologic relationships 2:667
 taxonomy and classification 2:666
 viral proteins 2:670
 viral replication 2:666–667
 virion properties 2:668–669
 pararetrovirus, defined 3:667, 3:274
 parasitism 1:606
 parasitoid wasp, defined 4:849
 paravaccinia 2:668
 parechoviruses 1:279–280
 classification 2:675
 clinical features 2:679–680
 virus genetics and association to disease 2:680
 diagnosis 2:680–681
 epidemiology 2:677–679
 Australia 2:679
 Japan 2:679
 parechovirus species B-D 2:679
 seroprevalence 2:678–679
 genome 2:676–677
 life cycle 2:677
 pathogenesis 2:680
 treatment and prevention 2:681
 virion structure 2:675–676
 pariacoto virus 1:250, 1:477, 1:478f
 paritaprevir 5:123
 parr, defined 2:544
 Parrot hepatitis B virus (PHBV) 2:100
 Partitiviridae 4:568
 partitiviruses 4:544, 4:508, 4:454
 partitiviruses (*Partitiviridae*) – fungal 4:568
 classification 4:568–569
 epidemiology 4:572–573
 genome 4:569–570
 life cycle 4:570–572
 pathogenesis 4:573–574
 taxonomic and phylogenetic considerations 4:574–576
 virion structure 4:569
 Parton genome 2:808–809
 Parvoviridae 4:836, 4:837f, 2:419
 classification 2:419–420
 clinical features 2:424–425
 human bocaviruses 2:424–425
 human protoparvoviruses 2:425
 diagnosis 2:425–426
 epidemiology 2:422–424
 human bocaviruses 2:422–424
 human protoparvoviruses 2:424
 genome 2:420–422
 life cycle 2:422
 pathogenesis 2:425
 prevention 2:426
 virion structure 2:420
 parvoviruses (PVs) 4:835
 of carnivores
 classification (compact) 2:683
 diagnosis 2:687
 epidemiology 2:684–686
 life cycle 2:683–684
 pathogenesis and clinical features 2:686–687
 prevention 2:687
 treatment 2:687
 virion structure and genome 2:683
 of insects 4:711–713
 Polydnaviridae 4:712–713
 of invertebrates 4:835–837
 biochemical properties and purification of densovirus 4:840–841
 biophysical features and functions associated with densovirus capsid 4:843–844
 densovirus genome structure and replication 4:844–845
 discovery, taxonomy and evolution of densovirus 4:835–837
 expression strategies 4:845–848
 general features 4:839
 shrimp densovirus 4:839–840
 structural features of virions 4:841–843
 PASC *see* pairwise sequence comparison (PASC)
 Passeriform birds 2:810–811
 Passeriformes 2:811
 passerine, defined 2:343
 PAT1 *see* persistency-associated transcript 1 (PAT1)
 patatin-like phospholipase-1 (PLP-1) 4:524–525
 pathogen, defined 2:441
 pathogen-associated molecular patterns (PAMPs) 2:603, 2:475, 1:9, 1:585, 1:659, 1:601–602, 1:577
 pathogen derived resistance (PDR) 3:362
 antisense RNA technology 3:362
 RNAi technology 3:362–363
 pathogen-derived resistance, defined 3:355, 3:520
 pathogenesis, defined 2:884, 2:442
 pathogenic bacteria, recombineering in 4:298–299
 pathogenicity, defined 2:697, 4:53, 4:888, 4:768
 pathogenicity determinant protein 3:239
 pathogenicity determinants, defined 3:116
 pathogen recognition receptors (PRRs) 2:603
 pathognomonic, defined 2:814
 pathogroup, defined 3:184
 pathosystem, defined 3:8
 pathotype, defined 3:184
 pattern recognition receptors (PRRs) 2:611–612, 2:480–481, 1:577, 1:658
 pattern-triggered immunity (PTI) 3:32
 PBCV-1 *see* Paramecium bursaria chlorella virus 1 (PBCV-1)
 PBL *see* peripheral blood leukocytes (PBL)
 PBMCs *see* peripheral blood mononuclear cells (PBMCs)
 PBS *see* primer binding site (PBS)
 PCD *see* programmed cell death (PCD)
 PCNA *see* proliferating cell nuclear antigen (PCNA)
 PCPV *see* pseudocowpox virus (PCPV)
 PCR *see* polymerase chain reaction (PCR)
 PCV-2 systemic disease (PCV-2-SD) 2:187
 PCVD *see* porcine circovirus diseases (PCVD)
 PDA *see* potato dextrose agar (PDA)
 PDB *see* Protein Data Bank (PDB)
 PDD *see* proventricular dilatation disease (PDD)
 PDGFR *see* platelet-derived growth factor receptor (PDGFR)
 PDNS *see* porcine dermatitis and nephropathy syndrome (PDNS)
 PDR *see* pathogen derived resistance (PDR)
 PDVs *see* polydnaviruses (PDVs)
 PeaLDV 3:751
 pecluviruses (*Virgaviridae*)

- assembly of virus particles 3:534
 cultural practices 3:537
 host plant genetic resistance 3:537
 detection and diagnosis 3:534–535
 immunoassays 3:535
 molecular assays 3:535
 disease control 3:536–537
 epidemiology and disease cycle 3:535–536
 disease development cycle 3:535–536
 future perspectives 3:537
 geographic and seasonal distribution 3:529
 history 3:528
 host range and symptoms 3:529–531
 molecular properties 3:532
 5' and 3' noncoding sequences 3:532–533
 coding sequences 3:532
 genome structure and expression strategies 3:532
 sequence comparisons and phylogeny 3:533–534
 physical properties 3:531–532
 taxonomy and classification 3:528–529
 transmission 3:531
 seed and sap transmission 3:531
 soil-borne vector transmission 3:531
 vector 3:535
 virion morphology 3:531
 PED *see* pairwise evolutionary distances (PED)
 PEDV *see* porcine epidemic diarrhea virus (PEDV)
 PEG-IFN *see* pegylated interferon alpha (PEG-IFN)
 PEG precipitation *see* polyethylene glycol (PEG) precipitation
 PEGylated, defined 2:814
 pegylated interferon alpha (PEG-IFN) 5:223, 5:217
 PEL *see* primary effusion lymphoma (PEL)
Pelagibacteres 4:332–334
 PemoNPV *see* penaeus monodon nucleopolyhedrovirus (PemoNPV)
 PEMS *see* poult enteritis mortality syndrome (PEMS)
 penaeid, defined 4:768
 penaeid shrimp virus 2:256
 penaeus monodon baculovirus 4:828–829
 penaeus monodon nucleopolyhedrovirus (PemoNPV) 4:828–829
 penaeus monodon nudivirus (PmNV) 4:828–829
 penaeus monodon singly enveloped nuclear polyhedrosis virus (PmSNPV) 4:828–829
 penaeus stylirostris penstydensovirus 1 (PstDV1) 4:839–840
 penetration, defined 1:382
 penicillium aurantiogriseum foetidus-like virus 1 (PaFV1) 4:660
Penicillium chrysogenum 4:547–548
 penstydensoviruses 4:848
Penstylhamaparvovirus 4:836
 pentamers 1:250–251
 penton, defined 2:442
 penton base protein 1:331
 PEP *see* post-exposure prophylaxis (PEP)
 pepino mosaic virus (PepMV) 3:539, 1:649, 1:364–365, 1:365f, 3:626–628, 3:628–629
 diagnostic 3:543
 genome organization and expression 3:540
 host range and symptomatology 3:540–541
 taxonomy and classification 3:539
 virion properties 3:539–540
 virus damage and control 3:543–544
 virus isolates and strains 3:541–542
 virus transmission and epidemiology 3:542–543
 PepMV *see* pepino mosaic virus (PepMV)
 pepper, potato virus Y (PVY) in 3:619–620
Pepper leaf curl Lahore virus 3:750
 peptide-loaded MHC complexes (pMHC) 1:585
 peptidoglycan, defined 1:402, 4:194
 perch rhabdovirus (PRV) 2:325, 2:328
 percutaneous, defined 2:373
 performance qualification (PQ) 5:69
 pericardial fluid, defined 2:875
 perinatal, defined 2:373
 Perinet virus (PERV) 2:875
 periodontal disease 2:454
 peripheral, defined 2:778
 peripheral blood leukocytes (PBL) 2:283
 peripheral blood mononuclear cells (PBMCs) 2:189, 2:778, 2:148
Periplaneta fuliginosa 4:839
 peritrophic matrix, defined 4:780
 peritrophic membrane, defined 4:858
 permafrost 4:342, 4:353–354
Permutotetraviridae 4:897–898, 4:717
 permutotetravirus REPs 4:897–898
per os infectivity factor (PIF) 4:832
 perpetuation of viruses in nature 1:561
 persistence, defined 2:442
 persistence length, defined 4:167
persistency-associated gene 1 (pag1) 4:830
 persistency-associated transcript 1 (PAT1) 4:830
 persistent infection, defined 4:827, 1:542, 2:797
 persistent lymphocytosis (PL) 2:144
 persistent or chronic infection, defined 1:71
 persistent viral transmission, defined 3:719
 personal protective equipment (PPE) 5:82, 5:84, 5:68
 person-to-person transmission 5:208
 PERV *see* Perinet virus (PERV)
 Peste-des-petitsruminants virus (PPRV) 2:73–74
 pestiviruses
 classification 2:153
 clinical features and pathogenesis 2:159
 border disease (BD) 2:161
 bovine viral diarrhea and mucosal disease 2:160
 classical swine fever (CSF) 2:159
 epidemiology 2:158
 bovine viral diarrhea and border disease 2:158–159
 classical swine fever (CSF) 2:158
 genome structure and polyprotein processing 2:153–154
 life cycle 2:157–158
 prevention and control 2:161
 border disease of sheep, control of 2:163
 bovine viral diarrhea, vaccines against 2:162
 bovine viral diarrhea, control of 2:163
 classical swine fever, vaccines against 2:161–162
 classical swine fever in domestic pigs, control of 2:162
 classical swine fever in wild boar, control of 2:162–163
 diagnosis 2:161
 vaccines 2:161–162
 virion structure 2:154–156
 N^{pro} 2:156
 NS2 2:156
 NS2-3 2:156–157
 NS3 2:156
 NS4A 2:157
 NS4B 2:157
 NS5A 2:157
 NS5B 2:157
 p7 2:156
 PFGE *see* pulsed-field gel electrophoresis (PFGE)
 PFS *see* protospacer-flanking site (PFS)
 PFV1 *see* pyrobaculum filamentous virus 1 (PFV1)
 pgRNA *see* pregenomic RNA (pgRNA)
 phage 4:342
 see also bacteriophages
 adsorption 4:255
 biochemistry 4:6–7
 and biotechnology 4:7–8
 defined 4:242
 development
 as diagnostics for bacterial infections 4:254–255
 as therapeutics for bacterial infections 4:253–254
 discovery of 4:3–4
 ecology 4:316–317
 and evolution 4:8
 phage community ecology 4:320
 phage ecosystem ecology 4:320–321
 phage organismal ecology 4:317–318
 phage population ecology 4:318–320
 genetics 4:5–6
 in hypersaline environments 4:343–345
 in polar and other cold environments 4:349–352, 4:354
 glaciers 4:352–353
 permafrost 4:353–354
 polar lakes 4:353
 polar oceans 4:352
 sea ice 4:353
 structure and assembly 4:24–25
 therapy 4:4
 in thermal environments 4:345–347
 deep-sea hydrothermal vents 4:347–348
 hot deserts 4:348–349

- phage (*continued*)
 terrestrial hot springs 4:345–347
 typing 4:4–5
 phage cluster, defined 4:265
 phage diagnostics, additional challenges for 4:257
 intracellular bacteria 4:257
 market acceptance 4:257
 multiplex testing 4:257
 phage manipulation 4:257
 sample interference 4:257
 phage display, defined 4:252
 phage diversity
 examples of 4:268–269
 bottom up studies 4:269
 contribution of prophages to phage diversity 4:270
Enterobacteriales tailed phage example 4:269–270
 human phageome 4:268–269
 marine phageome 4:269
 top down studies 4:268–269
 two large diversity studies 4:269
 and horizontal exchange of genetic information 4:270–272
 nature of horizontally exchangeable mosaic section alleles 4:270–272
 strategies for studying 4:267–268
 types of bacteriophages 4:266*t*
 phage existence within environments 4:314–315
 microscopic determination 4:315–316
 phage isolation and host range 4:315
 phage genome and protein ejection *in vivo*
 contractile long-tailed phages 4:211
 ejection of proteins into the host cell at the initiation of infection 4:206–207
 factors affecting phage genome translocation 4:211–214
 non-transcriptional genome internalization *in vivo* 4:214–215
 transcription-mediated genome internalization *in vivo* 4:213–214
 models of 4:215–218
 non-contractile long-tailed phages (*Siphoviridae*) 4:209–211
 protein ejection and trans-envelope channel formation 4:206–207
 short-tailed phages (*Podoviridae*) 4:208–209
 tail-less phages 4:207–208
 phage-host interactions, ecology and evolution of 4:181–182
 phage-inducible chromosomal islands (PICIs) 4:98–99, 4:99–100, 4:101
 defined 4:98
 PICI-like elements (PLEs) 4:99–100, 4:98
 phage lambda 1:460, 1:635
 post-infection decision of 4:88–90
 phage lysis, definition of 1:502–503
 phageome, defined 4:265
 phage-receptor interactions' relevance to phage therapy 4:182–183
 phage receptors
 properties of 4:176–181
 techniques and methods for studying 4:175–176
 phage satellites, mobilization of 4:98–99
 counter-evolution by phages to avoid parasitism by satellites 4:102–103
 genetic organization 4:99
 in human health and disease 4:103–104
 integration and excision, gene induction, and replication 4:99–101
 packaging and horizontal transmission 4:101
 prevalence of 4:102
 strategies PICIs use to interfere with their helper phages 4:101–102
 phage T4 lysis inhibition 1:510–512
 phage technology
 major advantages of 4:255
 self-replication 4:255
 specificity 4:255
 major limitations of 4:255
 narrow host range 4:255
 phage resistance 4:255
 phage therapy 1:676
 additional challenges for 4:255–256
 chemistry, manufacturing, and controls (CMC) 4:256
 dosing strategies 4:256
 environmental impact 4:256–257
 preclinical data translation 4:255–256
 regulatory pathway 4:256
 safety and efficacy 4:256
 defined 4:283
 phage typing 4:254
 defined 4:252
 phagocytosis 1:584
 defined 1:577
 pham (or phamily), defined 4:276
 Phase 3 compost, defined 4:528
 phase variation system, defined 4:387
 Phasmaviridae 4:718
 PHBV *see* Parrot hepatitis B virus (PHBV)
 phenotype, defined 1:62
 phenotyping 5:185–186
Phenuiviridae 4:718, 3:720
 Phi29 connector 4:308
 Phi29 motors 4:151–152
 CryoEM structure of 4:149*f*
 terminase and 4:148–151
 PhiH repressor, defined 4:380
 phleboviruses in the candiru antigenic complex 2:771
 phlegiviruses 4:627
 biological properties 4:630
 evolutionary relationships among phlegiviruses 4:630–631
 future perspectives 4:631
 genome organization 4:628–630
 virion properties 4:627–628
 phloem, defined 3:456
 phloem-limited, defined 3:594
 phloem-limited virus, defined 3:200
 phosphatidylethanolamine 4:40
 phosphatidylglycerol 4:40
 phosphoinositide 3-kinase (PI3K) 1:534
 defined 2:875
 phospholipids, defined 1:345
 phosphoproteins 1:350–351, 3:573–574, 2:442–443
 phosphorodiamidate morpholino oligomers (PMOs) 2:615–616
 photolyase, defined 2:343
Phycodnaviridae 4:687–688
 phycodnaviruses (*Phycodnaviridae*)
 ecology 4:692–693
 genomes 4:690–691
 history 4:687
 phycodnavirus genes 4:693–694
 resistance to phycodnavirus infections 4:693
 taxonomy and classification 4:687–688
 virion structure and composition 4:688–689
 carbohydrates 4:690
 lipids 4:689–690
 morphology 4:688–689
 nucleic acids 4:689
 physicochemical and physical properties 4:689
 proteins 4:689
 virus replication 4:691–692
 virus transcription 4:692
 phylodynamics, defined 1:87
 phylogenetic analysis 1:101
 phylogenetic clade, defined 4:776
 phylogenetic diversity 4:677
 phylogenetic host breadth 5:257
 phylogenetics 4:677, 2:3
 phylogenetic tree, defined 2:182
 phylogeny, defined 1:62, 4:858, 3:818
 phylogeny of viruses 1:117*f*
 applications 1:121–122
 evolution, phylogeny, and viruses 1:116
 phylogenetic analysis 1:118–121
 tree definitions 1:116–118
 physical studies of viruses 2:414–415
 physico-chemical phase, of virology 1:5–6
 phytopathogen, defined 3:388
 phytoreovirus 1:303, 4:878, 4:879–881, 3:548, 3:551
 phytosanitation, defined 3:293
 phytoviruses, defined 3:106
 - stacking, defined 4:10
 PI3K *see* phosphoinositide 3-kinase (PI3K)
 PIC *see* pre-integration complex (PIC)
 PICIs *see* phage-inducible chromosomal islands (PICIs)
 Picobirnavirus 1:253
Picornavirales 4:792, 4:768
Picornaviridae 1:253–254
 classification of 2:688–689
 picornaviruses 2:757–758, 1:452–454, 1:278, 1:279–282, 1:280*f*
 capsid assembly 1:280–282
 genome encapsidation 1:282–283
 host interactions 1:283–285
 piercing-sucking insects, defined 3:106
Pieris rapae 4:772
 PIF, *see per os* infectivity factor (PIF)
 PIF proteins, defined 4:747
 "Piggyback-the-Winner" model 1:638
 Pimodivir 5:168
 and Osetamivir 5:171
 pinholins and SAR endolysins 1:507–508
 endolysin diversity 1:512–513

- holin diversity 1:512
- lysis in mycobacteria 1:514
- second type of spanin 1:508–509
- spanin diversity 1:513
- pinwheel, defined 3:631
- PIPO *see* Pretty Interesting Potyviral ORF (PIPO)
- PIRYV *see* Piry virus (PIRYV)
- Piry virus (PIRYV) 2:875
- PIs *see* protease inhibitors (PIs)
- piscine myocarditis virus (PMCV) 4:582
- Pitaya virus X (PiVX) 3:623–624, 3:628
- pithovirus 1:257
- PIV *see* parainfluenza virus (PIV)
- PiVX *see* Pitaya virus X (PiVX)
- PKR *see* protein kinase R (PKR)
- PL *see* persistent lymphocytosis (PL)
- PLAMV 3:628
- Plantain virus X (PIVX) 3:623, 3:623–624
- plant and protozoal partitiviruses 4:632–633
 - current taxonomy of partitiviridae 4:633
 - genome organization and replication strategy 4:637
 - 5' and 3'UTRs 4:637–639
 - genome organization 4:637
 - replication strategy 4:640
 - RNA1&2 4:639–640
- transmission 4:640–641
- virion properties 4:633–635
 - alphapartitivirus 4:635
 - betapartitivirus 4:635
 - crispovirus 4:635–636
 - deltapartitivirus 4:636–637
- virus-host-interaction 4:641
- plant antiviral defense: gene-silencing pathways 3:43
 - antiviral RNA silencing (av-RNAi) 3:45–47
 - applications of 3:49–51
 - antiviral silencing in crops 3:47–49
 - diverse gene-silencing pathways in plants 3:43–45
 - viral and plant RNA-silencing suppressors 3:47
 - viral siRNA-mediated antiviral RNAi in mammalian cells 3:49
- plant endornaviruses 3:388–389
- plant-fungal mutualistic associations, role of mycoviruses in 4:441
- plant reoviruses (*Reoviridae*) 3:545
 - diagnosis and control 3:552–553
 - distribution 3:551
 - future 3:553
 - host range and virus transmission 3:551
 - fijiviruses 3:551
 - oryzaviruses 3:551
 - phytoreoviruses 3:551
 - pathogenicity 3:551–552
 - replication and gene expression 3:548–551
 - taxonomy and classification 3:545–546
 - virion structure and genome organization 3:546–547
 - fijivirus 3:546–547
 - oryzavirus 3:547–548
 - phytoreovirus 3:548
- plant resistance to viruses 3:60, 3:52–53, 3:69–70
 - application of recessive resistance in agriculture 3:78–79
 - benefits of virus-resistant transgenic plants to agriculture 3:54–55
 - cloned recessive resistance genes 3:73–74
 - eIF4E-mediated recessive resistance
 - characteristics of 3:74–76
 - molecular mechanisms of 3:76–77
 - environmental and human health safety issues 3:55–59
 - field resistance and commercial adoption 3:54
 - historical perspective on engineered resistance against plant viruses 3:53
 - mechanisms of recognition 3:62–63
 - molecular underpinnings of engineered resistant to plant viruses 3:53–54
 - naturally occurring recessive resistance 3:70–73
 - new recessive resistance genes found in natural diversity 3:77–78
 - non-NB-LRR resistance 3:66–67
 - resistosome 3:64–66
 - R gene products 3:60–61
 - R gene responses to viruses 3:61–62
 - signaling mechanisms 3:63–64
- plant rhabdoviruses *see* rhabdoviruses
- plants and plant cell culture 1:664–665
- plant satellite viruses 3:581
 - Albetovirus* 3:581–582
 - Aumaivirus* 3:582
 - effect of mixed infections on host and helper viruses 3:584
 - Papanivirus* 3:582–584
 - plant satellite viruses 3:584–585
 - Virtovirus* 3:584
- plant viral diseases 3:81–83
 - cash crops viral diseases 3:94–95
 - ornamentals and orchids viral diseases 3:95–96
 - sugarcane viral diseases 3:94–95
 - cereal viral diseases 3:83
 - fruit virus diseases 3:89–90
 - banana viral diseases 3:90
 - cacao swollen shoot disease (CSSD) 3:90
 - citrus tristeza virus (CTV) 3:90–91
 - papaya ring spot virus 3:91–92
 - plum pox virus (Sharka) disease of stone fruits 3:92
 - legume viral diseases 3:92–94
 - root and tuber crops viral diseases 3:83–85
 - cassava viral diseases 3:83–85
 - potato viral diseases 3:85–86
 - sweet potato viral diseases 3:86
 - vegetable viral diseases 3:86–87
 - chilli viral diseases 3:86–87
 - cucurbit viral diseases 3:87–88
 - okra viral diseases 3:88
 - tomato viral diseases 3:88–89
- plant viruses 1:5–6, 3:3
 - biochemical/biophysical age 3:4–5
 - biological age 3:4
 - host and vector 'manipulation' by 3:113–115
 - molecular biology age 3:5–7
 - prehistory 3:3
 - recognition of viral entity 3:3–4
 - in their invertebrate vectors 4:722, 4:721*t*
- plant viruses, emerging and re-emerging 3:8–9
 - grapevines viruses 3:14–15
 - grapevine pinot gris virus 3:16
 - grapevine roditis leaf discoloration-associated virus 3:15
 - ornamental viruses: rose rosette virus 3:19
 - staple crops: cassava brown streak disease 3:17
- vegetable viruses 3:9
 - begomoviruses 3:13–14
 - cucumber green mottle mosaic virus 3:9
 - orthotospoviruses 3:12
 - squash leaf curl virus 3:14
 - tobamoviruses 3:9
 - tomato brown rugose fruit virus (ToBRFV) 3:10–11
 - tomato spotted wilt virus 3:12
- plant viruses transmission by insects 3:106–107
 - beetles, transmission by 3:112–113
 - circulative transmission 3:107–108
 - circulative non-propagative transmission 3:107–108
 - circulative propagative transmission 3:110
 - history of typology of transmission modes 3:107
 - non-circulative transmission 3:110–112
 - capsid strategy 3:112
 - helper strategy 3:112
- plant viruses transmission by non-insect vectors 3:113
 - fungi and protists, transmission by 3:113
 - mites, transmission by 3:113
 - nematodes, transmission by 3:113
- plant virus vectors 3:106
- plaque, defined 1:162, 4:314, 4:276
- plaque assay, defined 4:265
- plaque-forming unit (Pfu), defined 4:21, 2:442
- plaque reduction neutralization test (PRNT) 2:907, 2:641, 5:94
- plasmid, defined 4:276
- plasmodesmata 1:495, 3:788, 3:383, 3:313, 4:478, 3:456
- plasmodiophorida, defined 3:528
- plastochron, defined 3:461
- platelet-derived growth factor receptor (PDGFR) 2:450–451
- Plautia stali* 4:771
- Plautia stali intestine virus (PSIV) 4:771
- pleiotropic mutation, defined 3:69
- Pleolipoviridae* 4:382–384
- pleolipovirus, defined 4:380
- pleolipovirus His2 1:433–434
- pleolipovirus infectivity, stability of 4:384
 - salinity 4:384
 - temperature and other significant factors 4:384–385
- pleolipovirus virion structure 4:382*f*
- pleomorphic, defined 2:738

- pleomorphic, spindle-shaped, and spherical viruses 4:376
- pleomorphic viruses 4:376
- spherical virus MetSV 4:377–378
- spindle-shaped viruses 4:376–377
- pleomorphic DNA viruses 1:407
- pleomorphic viruses 4:376, 4:359
- pleomorphism 2:68–69
- Pleosporales megabirnavirus 1 (PmbV1) 4:600
- PLP-1 *see* patatin-like phospholipase-1 (PLP-1)
- plum pox virus (*Potyviridae*) 3:586
- detection 3:591
- ecology and control 3:592
- economical importance 3:586
- history and geographical distribution 3:586–587
- host range and symptomatology 3:587
- virion structure and genome properties 3:587
- strains/molecular groups 3:587–591
- virus spread 3:591–592
- plum pox virus (Sharka) disease of stone fruits 3:92
- (+)-sense RNA (plus-sense RNA) 1:382
- plus-strand RNA virus, defined 2:797
- Plutella xylostella* 4:772
- PLV *see* puma lentivirus (PLV)
- PIVX *see* Plantain virus X (PIVX)
- PmbV1 *see* Pleosporales megabirnavirus 1 (PmbV1)
- PMCV *see* piscine myocarditis virus (PMCV)
- PMF *see* proton-motive force (PMF)
- pMHC *see* peptide-loaded MHC complexes (pMHC)
- PML *see* progressive multifocal leukoencephalopathy (PML)
- PmNV *see* penaeus monodon nudivirus (PmNV)
- PMOs *see* phosphorodiamidate morpholino oligomers (PMOs)
- p/mpMLVs *see* polytropic and modified polytropic MLVs (p/mpMLVs)
- PmSNPV *see* penaeus monodon singly enveloped nuclear polyhedrosis virus (PmSNPV)
- pneumonia 2:762
- pneumonia virus of mice (PVM) 2:754
- Pneumoviridae* genome organization 2:748f
- poaceae, defined 3:456
- pock, defined 2:868
- Podoviridae* 1:320, 4:186, 4:276
- mosaicism in 4:280
- podoviruses 1:406, 4:342
- Pogosta disease 2:840
- poinsettia mosaic virus 3:824–825
- properties and distinguishing characteristics 3:825
- Point-of-Care (PoC) clinical impact 5:48–50
- Point-of-Care (PoC) menus 5:48
- Point-of-Care (PoC) strategy and implementation 5:46
- Laboratory Grade 1A 5:46
- Laboratory Grade 1B 5:46
- Laboratory Grade 2A 5:46
- Laboratory Grade 2B 5:46–47
- Laboratory Grade 3 5:47–48
- Point-of-Care (PoC) technologies 5:45–46, 5:47f
- point-of-care serodiagnostics 5:20
- Point-of-Care-tests, defined 5:98
- polar and other cold environments, phages in 4:349–352, 4:354
- glaciers 4:352–353
- permafrost 4:353–354
- polar lakes 4:353
- polar oceans 4:352
- sea ice 4:353
- polar lakes 4:353
- polar oceans 4:352
- poleroviruses (*Luteoviridae*) 3:594–595
- beet poleroviruses 3:599–600
- defined 3:594
- diagnosis 3:600–601
- disease management 3:601
- genome organization and gene expression 3:595–596
- phylogenetic diversity 3:596–597
- physical properties 3:595
- potato leafroll virus 3:599
- silencing suppression by 3:599
- sugarcane yellow leaf virus 3:599
- transmission 3:597–598
- virus-virus interactions 3:598–599
- pol* gene 2:57–58, 2:61t
- POL inhibitors *see* polymerase (POL) inhibitors
- polintons, defined 1:329
- polintons and evolutionary pathway of DJR viruses 1:342–343
- polintovirus, defined 1:372
- polio 1:668–669
- eradication
- antivirals 5:313
- background 5:310
- containment 5:313
- new vaccines 5:313
- poliovirus surveillance 5:311–312
- status of eradication 5:312–313
- transition of global polio eradication initiative assets post-eradication 5:313–314
- vaccines 5:310–311
- poliomyelitis 2:808
- polioviruses 2:688, 2:689
- cessation of vaccination 2:695
- global polio eradication program 2:690–692
- pathogenesis and disease 2:689
- Picornaviridae*, classification of 2:688–689
- poliovirus receptor 2:688
- Sabin vaccine strains, attenuation of 2:690
- vaccine-derived poliovirus 2:692–695
- vaccines 2:689–690
- virus genome 2:688
- virus particle 2:688
- poliovirus receptor (PVR) 1:284
- pollen-borne, defined 3:430
- poly(A) binding protein (PABP) 1:444, 3:132
- polyacrylamide gel electrophoresis (PAGE) 3:586
- polyadeniration, defined 4:544
- Polycipiviridae* 4:718
- polycistronic RNA, defined 4:21
- polyclonal immunoglobulins 5:268
- hepatitis A 5:268
- measles 5:268–269
- Polydnaviridae* 4:712–713, 4:849
- polydnaviruses (PDVs) 4:712
- classification (compact) 4:849
- fate of PDV packaged genome in parasitized insect 4:853–854
- function of the genes encoded by PDV packaged genome 4:853
- life cycle of 4:849–850
- morphogenesis of PDV particles 4:851
- organization of PDV proviral segments in the wasp genome 4:854–856
- packaged genome 4:851–853
- proviral genome maintained in the wasp genome 4:854
- viral machineries 4:856–857
- virion structure 4:850–851
- polyethylene glycol (PEG) precipitation 1:167
- polyextreme environment 4:342
- polyhedrin, defined 4:739, 4:747
- polyhedron, defined 1:248, 4:739
- polyhedrosis 4:704
- polylysogeny 1:639–640, 1:638–640
- polymerase (POL) inhibitors 5:164–166
- Baloxavir – Marboxyl (Xofluza) 5:166–168
- Favipiravir (Avigan) 5:168
- Pimodivir 5:168
- Ribavirine (Ribavirin) 5:164–166
- polymerase 1:350
- polymerase chain reaction (PCR) 1:8, 2:3, 1:9, 1:622–623, 2:301, 5:91, 2:640–641, 2:820, 2:722–723, 2:735, 4:324–325, 5:98, 3:192
- assays 5:92
- based amplification 3:759
- based methods 3:629
- defined 5:27, 3:430
- detection of viral RNA 2:663–664
- for human adenoviruses DNA detection 5:202–203
- and its evolution 5:112
- technique 1:673
- testing 5:177
- polymerase complex, defined 4:26
- polymerase protein 3:575
- polymerase selectivity, factors determining 1:54
- polymeric immunoglobulin receptor (pIgR) 1:529, 1:531, 1:533
- polymycoviruses 4:454–455
- polyomavirus associated nephritis (PVAN) 5:108
- polyomaviruses (PyVs) 2:518–519, 5:108–109, 1:498, 1:392, 2:526–527
- BK polyomavirus (BKPyV) 2:522
- clinical features 2:522
- infectious cycle 2:522–523
- genome 2:519–521
- infectious cycle 2:521–522
- JC polyomavirus (JCPyV) 2:523–524

- clinical features 2:523–524
- infectious cycle 2:524
- Merkel cell polyomavirus (MCPyV) 2:524–525
 - clinical features 2:524–525
 - infectious cycle 2:525
 - of skin 2:526
 - virion structure 2:519
- polypeptide III *see* penton base protein
- polypeptide IV *see* adenovirus fiber
- Polypeptide V 2:8
- polyphyletic, defined 1:28, 4:457, 3:653
- polyphyletic group of viruses 1:162–164
- polyphyly, defined 4:835
- polyprotein, defined 2:386, 1:382, 3:388, 3:631, 3:703, 3:797, 4:768
- polyprotein processing, defined 2:362
- polypurine tract (PPT) 2:59, 2:62–63, 3:96
- polypyrimidine tract-binding protein (PTB) 2:263
- polyribonucleotidyltransferase (PRNTase) 1:350
- polythetic, defined 1:28
- polythetic class, defined 1:47
- polytropic and modified polytropic MLVs (p/mpMLVs) 2:646
- pomoviruses (*Virgaviridae*) 3:603
 - genome properties and functions of the encoded proteins 3:604–605
 - host range, geographical distribution, transmission by vector and variability 3:606–609
 - physical properties of particles 3:603–604
 - serological relationships, diagnosis, and control 3:609–610
 - taxonomy and classification 3:603
 - virus–host interactions and movement 3:605–606
- population genetic processes, studying
 - direct selection and linked selection effects 5:228–229
 - evolutionarily informed treatment strategies 5:230
 - key aspects of population genetic environment 5:229–230
 - population dynamics of infection 5:227–228
- populations, viral 1:53
 - basic population genetic processes shaping viral diversity 1:57–58
 - adaptation via directional selection of beneficial mutations 1:58
 - error catastrophe and lethal mutagenesis of viruses 1:58
 - frequency-dependent selection 1:58–59
 - mutation-selection balance 1:57–58
 - random genetic drift 1:59
 - viral quasispecies 1:58
- diversity-generating retroelements 1:55
- error-prone replication 1:53–54
- lack of proofreading 1:54–55
- polymerase selectivity, factors determining 1:54
- viral polymerases, intrinsic selectivity of 1:54
- post-replicative repair, role of 1:55
- error-prone repair polymerases 1:55
- repair avoidance 1:55
- recombination 1:56
 - in DNA viruses 1:56–57
 - reassortment 1:56
 - template switching 1:56
- viral disease, implications for 1:59
 - drug resistance 1:60
 - immune escape 1:59–60
 - short-term pathogenesis 1:59
- viral hyper-mutation mediated by host enzymes 1:55
 - cellular adenosine deaminases 1:55–56
 - cellular cytidine deaminases 1:55
- population size, influence of 1:563
- porcine adenoviruses 2:12
 - Porcine circovirus 1* (PCV-1) 2:184
 - Porcine circovirus 2* (PCV-2) 2:184, 2:189f
 - Porcine circovirus 3* (PCV-3) 2:185
- porcine circovirus diseases (PCVD) 2:187, 2:188–189
- porcine dermatitis and nephropathy syndrome (PDNS) 2:187
- porcine epidemic diarrhea virus (PEDV) 2:198, 2:256
 - classification 2:850
 - clinical features 2:850–852
 - diagnosis 2:852
 - epidemiology 2:850
 - genome 2:850
 - genome organization of 2:851f
 - life cycle 2:850
 - pathogenesis 2:852
 - prevention 2:852
 - replication cycle 2:851f
 - treatment 2:852
 - vaccines for 2:852t
 - virion structure 2:850
- porcine reproductive and respiratory syndrome (PRRS) 2:697
- porcine reproductive and respiratory syndrome virus (PRRSV) 1:447, 2:256
 - classification 2:697
 - clinical features 2:702
 - diagnosis 2:704–705
 - epidemiology 2:701–702
 - genome organization 2:698–700
 - immunity 2:703
 - life cycle 2:700–701
 - nonstructural and structural proteins of 2:699t
 - pathogenesis 2:703
 - prevention and control 2:705–706
 - taxonomy of 2:697t
 - virion structure 2:697–698
- porcine respiratory disease complex (PRDC) 2:187
- porencephaly, defined 2:34
- portal, defined 1:318, 4:45, 2:442
- portal proteins 1:323–324, 4:105, 4:107–108
 - actively assisting DNA packaging coupling with ATP hydrolysis 4:111
 - bound with adapter proteins facilitates tail attachment 4:109
 - as a check-valve preventing DNA from slipping out 4:110
- help to spool condensed genomic DNA in the capsid 4:109–110
- main techniques used to study portal protein structure and function 4:112
 - biochemical methods 4:112
 - molecular dynamics simulation 4:113
 - optical tweezers: a technique to observe DNA packaging in real time 4:113
 - planar bilayer membrane (BLM) technology 4:113
 - techniques for structural analysis 4:112–113
- as a nucleator in the assembly of the prohead with proper morphology 4:108
- portal directly-mediated packaging mode 4:111
 - lengthwise channel expansion and contraction in the portal protein may apply forces to DNA 4:111
 - rotation of the portal protein may apply forces to DNA 4:111
 - sequential movement of tunnel loops may apply force to DNA 4:111–112
- portal indirectly-mediated packaging mode 4:112
 - DNA translocation driven by a brownian motor inherent in the portal protein 4:112
 - DNA translocation driven by the energy from the DNA deformation in the portal protein 4:112
- portal proteins actively assists DNA packaging at the late stage of DNA packaging 4:111
- providing binding sites for terminases 4:108–109
- as a sensor to control the DNA packaging 4:110–111
- structural features of 4:107
- terminase-driven mechanism plays a key role at the initial stage of DNA packaging 4:111
- portal vertex 4:105–106
 - DNA packaging into the prohead 4:106
 - formation of a prohead 4:105–106
 - virion maturation 4:106–107
- Portogloboviridae* 4:362–363
- positional cloning, defined 3:69
- positive predictive value (PPV) 5:74
- positive selection-based enrichment 1:179
- positive sense (+) strand, defined 1:429
- positive-sense, defined 4:776
- positive-sense, single-stranded RNA virus 3:388
- positive-sense RNA viruses 1:499
- positive-strand RNA virus, defined 2:386
- Pospiviroidae* 3:852
- post-exposure and pre-exposure prophylaxis (PEP and PrEP) 5:153
- post-exposure prophylaxis (PEP) 5:164, 2:744
- Post-E, defined 2:747
- postfusion conformation, defined 1:417
- post-larvae, defined 4:888
- post-replicative repair, role of 1:55
 - error-prone repair polymerases 1:55
 - repair avoidance 1:55

- post-replicative repair, role of (*continued*)
 post-transcriptional gene silencing (PTGS)
 3:361, 4:632, 3:362–363, 3:788,
 3:116, 3:149, 3:528, 3:123, 3:555
 post-translational modifications (PTMs)
 3:420, 1:664
 post-transplantation/iatrogenic Kaposi's
 sarcoma 2:606
 post-transplantation lymphoproliferative
 disorders (PTLDs) 5:195
 postural orthostatic tachycardia syndrome
 (POTS) 5:297–298
 post zoster neuralgia, defined 5:181
 potato, potato virus Y (PVY) in 3:616–618
 Potato aucuba mosaic virus (PAMV) 3:623,
 3:628
 potato dextrose agar (PDA) 4:522
 potato leafroll virus 3:599
 potato viral diseases 3:85–86
 potato virus X (PVX) 3:623, 3:626*f*, 3:627*f*,
 3:626–628, 3:628, 3:628–629, 3:4,
 1:68
 potato virus Y (PVY) 3:612, 3:4, 1:68
 control methods 3:618–619
 host range 3:614–615
 in pepper 3:619–620
 phylogeny and evolution 3:616
 in potato 3:616–618
 serology 3:616
 in tobacco 3:621–622
 in tomato 3:620–621
 transmission 3:615–616
 viral particle, genome and cytopathology
 3:612–614
 Potexviruses (*Alphaflexiviridae*)
 diagnosis 3:629
 genome organization 3:625–628
 properties and functions of gene
 products 3:626–628
 replication and propagation 3:628
 species demarcation criteria 3:625
 symptoms of potexvirus-infected plants
 3:627*f*
 taxonomy, phylogeny, and evolution
 3:623–625, 3:626*f*
 transmission, host range 3:628–629
 epidemiology and control 3:628–629
 virion structure 3:625
 virus species in the genus *Potexvirus* 3:624*t*
 POTS *see* postural orthostatic tachycardia
 syndrome (POTS)
 potyvirus, defined 3:631, 3:456
 Potyviridae 3:295, 3:296–297, 3:520, 3:520–
 521, 3:526, 3:586, 3:862, 3:631,
 3:268, 3:184, 3:797, 3:612
 potyviruses (*Potyviridae*) 3:631
 diagnosis 3:640
 epidemiology and control 3:638–640
 pathogenicity 3:637–638
 taxonomy, phylogeny and evolution
 3:631–632
 virion structure 3:632–634
 expression of overlapping gene products
 3:636–637
 properties and functions of gene products
 derived from polyprotein 3:634–636
 properties of genome 3:633–634
 replication and propagation 3:637
 poult enteritis mortality syndrome (PEMS)
 2:96
 power stroke, defined 4:148
 poxvirus APC/cyclosome regulator (PACR)
 protein 2:672
 poxviruses 2:165, 1:498
 cervidpoxvirus 2:171
 classification 2:165
 genome 2:165–166
 of invertebrates 4:713
 life cycle 2:166–167
 bovine papular stomatitis virus (BPSV)
 2:170–171
 lumpy skin disease virus 2:168
 orf virus 2:169
 parapoxvirus of red deer in NZ 2:171
 pseudocowpoxvirus (PCPV) 2:171
 sheeppox and goatpox virus 2:166–167
 of ruminants 2:165*t*
 vaccinia virus (VACV) 2:171
 virion structure 2:165
 poxviruses of insects 4:858
 classification 4:858–859
 entomopoxvirus phylogeny 4:865
 genomics 4:862–865
 conserved and shared genes 4:863–865
 gene families 4:865
 replication in larvae 4:860–862
 virus/midgut interactions 4:859–860
 PPD *see* Pacheco's parrot disease (PPD)
 PPDV *see* Pacheco's parrot disease virus
 (PPDV)
 PPE *see* personal protective equipment
 (PPE)
 PpNSRV-1 *see* pteromalus puparum
 negative-strand RNA virus 1
 (PpNSRV-1)
 P protein, defined 2:747
 PPRV *see* Peste-des-petitsruminants virus
 (PPRV)
 PPT *see* polypurine tract (PPT)
 PPV *see* positive predictive value (PPV)
 PPVs *see* parapoxviruses (PPVs)
 PPXY late domain 1:522–523
 PQ *see* performance qualification (PQ)
 practical applications 1:251
 PRANC (Pox protein Repeats of Ankyrin C-
 terminal) domain 2:672
 PRC *see* primary replicase complex (PRC)
 PRD1 genes 4:36, 4:37, 4:37–39, 4:38*t*
 PRD1 infection pathway 4:208*f*
 PRD1-like viruses 1:337–339
 PRDC *see* porcine respiratory disease
 complex (PRDC)
 precipitation methods 1:167–168
 pre-emptive therapy 5:191
 Pre-Exposure Prophylaxis (PrEP) 2:473–474
 Pre-F, defined 2:747
 prefusion conformation, defined 1:417
 pregenomic RNA (pgRNA) 3:164, 2:104
 pre-integration complex (PIC) 2:644, 2:464
 PrEP *see* Pre-Exposure Prophylaxis (PrEP)
 preparative ultracentrifugation methods
 1:168–170
 chromatographic methods 1:170–171
 flow field-flow fractionation methods
 1:171–172
 purification performance, assessing 1:172
 preprimosome, defined 4:61
 preprotoxin, defined 4:534
 Pretty Interesting Potyviral ORF (PIPO)
 3:631
 prevalence, defined 5:247, 1:559
 previrology 1:3
 primary effusion lymphoma (PEL) 2:604,
 2:605, 2:606, 2:599
 primary replicase complex (PRC) 2:130
 primer binding site (PBS) 3:96
 priming, defined 1:417
 primosome, defined 4:61
 principle of genetic economy, defined 1:248
 prion, defined 2:707
 prion incubation period, defined 2:707
 prion protein (PrP), defined 2:707
 prions of vertebrates 2:707
 aberrant prion protein metabolism 2:707–
 708
 future perspectives 2:712
 human prion diseases 2:708
 neuronal cell death in prion disease 2:712
 prion disease pathogenesis 2:710–711
 prion strains 2:711–712
 protein-only hypothesis of prion
 propagation 2:708–709
 structural properties of PrP^{Sc} 2:709–710
 structure and putative function of PrP^C
 2:709
 prion strain, defined 2:707
 prM protein 2:885
 PRNP, defined 2:707
 PRNT *see* plaque reduction neutralization
 test (PRNT)
 PRNTase *see* polyribonucleotidyltransferase
 (PRNTase)
 procapsid, defined 4:10, 4:26, 4:160, 4:302,
 1:318, 4:45
Procedovirinae 3:788, 3:481
 processing bodies, defined 3:761
 prodrome, defined 2:868
 pro-drug, defined 2:404
 prodrug convertases 5:241
 productive/lytic infection, defined 4:827
 proficiency testing (PT) 5:72, 5:76
 progeny positive-strand RNAs 4:771
 progeny virions 2:452
 programmed cell death (PCD) 1:606, 1:607,
 4:724
 programmed death receptor 1 (PD-1) 1:660
 programmed death receptor ligand-1 (PD-
 L1) 1:660
 progressive multifocal leukoencephalopathy
 (PML) 5:109, 1:74
 prohead 4:115–117
 assembly 4:117–118
 assembly parasites 4:123
 defined 4:219, 4:105
 major capsid protein before capsid
 assembly 4:117
 portal 4:120–121
 scaffolding protein 4:118–119

- accelerating assembly 4:119
 delay MCP transformation 4:119–120
 exclude host factors 4:120
 promote correct geometry 4:119
 scaffold involvement in portal incorporation 4:120
 stability of 4:123
 structure of 4:121–123
 prokaryotes, defined 4:242
 prokaryotic Argonaute (pAGO) 1:608
 prokaryotic cells, release of phages from 1:501–502
 cell envelope 1:503–504
 multigene lysis (MGL) systems 1:504–505
 lambda lysis, looking under the hood of 1:505–507
 lambda lysis cassette 1:505
 lambda MGL, operational outline of 1:505
 lambda prophage as a lysis platform 1:504–505
 phage lysis, definition of 1:502–503
 pinholins and SAR endolysins 1:507–508
 endolysin diversity 1:512–513
 holin diversity 1:512
 lysis in mycobacteria 1:514
 MGL diversity 1:512
 MGL in gram-positive hosts 1:513–514
 regulation of MGL 1:509–510
 second type of spanin 1:508–509
 spanin diversity 1:513
 single gene lysis (SGL) systems 1:514–515
 Archaea, phage egress in 1:517–518
 extrusion 1:517
 of *Leviviridae* 1:515–516
 mysterious L 1:516–517
 X174 E 1:514–515
 prokaryotic dsDNA viruses, genome replication of 1:429–430
 prokaryotic RNA viruses, genome replication of 1:437
 prokaryotic ssDNA viruses, genome replication of 1:436
 prokaryotic viruses 1:165–166
 classification of 1:430*t*
 genomes of 1:429
 proliferating cell nuclear antigen (PCNA) 4:390–391, 4:366
 proliferation, defined 2:528
 promoter, defined 2:442, 2:56, 4:69
 promoters, viral 1:652
 use 1:655–657
 viral examples 1:653–655
 prophage 4:291, 4:265, 1:552, 4:276, 4:242, 4:77, 4:88, 4:368, 4:283, 4:342
 prophage induction 4:84–85, 4:77
 prophage remnants 1:640–641
 prophage reservoir, defined 4:283
 prophages 1:636
 pro-retroviruses, defined 3:653
 protease, defined 4:776
 protease activation, cell entry targeting by 5:238
 protease inhibitors (PIs) 5:139, 2:473
 -based regimens 5:140
 atazanavir 5:141
 darunavir 5:140–141
 dual therapy 5:140
 lopinavir 5:141
 new PIs 5:141
 triple therapy 5:140
 classes of 5:140*t*
 in combination with other ART classes 5:140
 drug–drug interactions 5:141
 mechanism of action 5:139
 pharmacodynamics and pharmacokinetics in 5:139–140
 protease inhibitors in clinical practice 5:143
 protease inhibitors in low resource settings 5:142–143
 sub-Saharan Africa 5:142–143
 resistance to PIs 5:141–142
 side effects 5:141
 cardiovascular 5:141
 dyslipidemia, lipodystrophy and insulin resistance 5:141
 gastrointestinal 5:141
 hepatotoxicity 5:141
 kidney stones 5:141
 tolerability of 5:141
 protease protein, defined 3:96
 protease/reverse transcriptase (P5) 3:318
 protective antigen (PA) 4:262
 protein antibiotics (PA) 1:515, 1:501–502
 protein cluster, defined 4:265
 protein-coding genes 2:457
 protein-coding regions 2:444*f*
 Protein Data Bank (PDB) 1:141, 1:153
 protein–DNA binding 4:139–140
 DNA binding domains (DBDs) 4:140–141
 multiple binding events regulate complex assembly 4:141
 sequence specific and non-specific binding 4:140
 protein domain, defined 1:345
 protein ejection and trans-envelope channel formation 4:206–207
 protein glycosylation 4:387, 4:359
 protein IIIa 1:333
 protein injection 4:48
 Protein IX 1:337, 1:333–335
 protein kinase R (PKR) 2:857–858, 1:444–445
 protein loops, defined 1:345
 protein-only hypothesis, defined 2:707
 protein-primed DNA replication 1:433–434, 4:36, 1:429, 4:368
 protein receptors 1:394–396
 complement control proteins 1:396–397
 ectoenzymes 1:397
 CoV-ectoenzyme interactions 1:397
 immunoglobulin superfamily proteins 1:394–396
 virus interactions with IgSF receptors 1:395–396
 Protein V 1:335
 protein VI 1:333
 protein VIII 1:333
 proteomics 1:498
 protist, defined 3:106
 protohelix, defined 3:734
 proton-motive force (PMF) 1:504, 4:206
 proto-oncogene, defined 2:316
 protoparvoviruses 2:684*f*
 clinical features 2:425
 epidemiology 2:424
 protoplast, defined 4:601, 4:607, 3:456, 3:507, 4:648
 protospacer adjacent motifs (PAM) 4:398
 protospacer-flanking site (PFS) 4:248
 prototype virus, defined 4:776
 proventricular dilatation disease (PDD) 2:137, 2:140–141
 providence virus 4:898, 1:250
 proviral DNA 2:62
 proviral load, defined 2:528
 proviophage 1:380
 proviophage, defined 1:372
 provirus, defined 1:162, 2:643, 2:144, 4:387, 1:627, 2:528, 2:56, 4:359
 PrP, defined 2:707
 PrP^C, defined 2:707
 PrP^{Sc}, defined 2:707
 PRRs *see* pathogen recognition receptors (PRRs); pattern recognition receptors (PRRs)
 PRRS *see* porcine reproductive and respiratory syndrome (PRRS)
 PRRSV *see* porcine reproductive and respiratory syndrome virus (PRRSV)
 PRSV *see* papaya ringspot virus (PRSV)
 prunevirus, defined 3:805
 pruritis, defined 2:738
 PRV *see* perch rhabdovirus (PRV)
 P-selectin glycoprotein ligand-1 (PSGL-1) 1:410, 1:284
 Pseudaletia separata EPV (PSEV) 4:859
 pseudocowpox 2:668
 pseudocowpox virus (PCPV) 2:666, 2:171
 pseudoknot, defined 3:447, 3:818, 4:648, 3:734
 pseudolysogeny, defined 4:314
Pseudomonas aeruginosa 1:640–641
 pseudomurein endoisopeptidase, defined 4:387
 pseudorabies virus (PrV) 2:715*f*
 classification 2:714
 clinical features and pathogenesis 2:720–721
 epidemiology and prevention 2:722–723
 future perspectives 2:723
 genome 2:716*f*
 history 2:714
 immunity 2:721–722
 latency 2:722
 ORFs 2:717*t*–718
 pseudorabies virus as a tool in neurobiology 2:723
 replication cycle 2:719*f*
 replication cycle 2:715–720
 virion structure 2:714–715
 capsid 2:715
 envelope 2:715
 genome 2:714–715
 tegument 2:715
 pseudo-recombinant, defined 3:301

pseudo-recombination, defined 3:371, 3:81, 3:411
Pseudoviridae 3:654, 3:100–101
 pseudovirus, defined 1:208, 3:653
 PSEV *see* *Pseudaletia separata* EPV (PSEV)
 PSGL-1 *see* P-selectin glycoprotein ligand-1 (PSGL-1)
 PsHV1 *see* psittacid herpesvirus type 1 (PsHV1)
 psittacid herpesvirus type 1 (PsHV1) 2:112, 2:113
 psittacine, defined 2:343
 psittacine BFDV 2:187
 PSIV *see* *Plautia stali* intestine virus (PSIV)
 PSs *see* packaging signals (PSs)
 P(S/T)AP late domain 1:522
 PstDV1 *see* *penaeus stylostris* penstydensovirus 1 (PstDV1)
 PT *see* proficiency testing (PT)
 PTB *see* polypyrimidine tract-binding protein (PTB)
 pteromalus puparum negative-strand RNA virus 1 (PpNSRV-1) 1:104
 pteropid bats, defined 2:355
 PTGS *see* post-transcriptional gene silencing (PTGS)
 PTI *see* pattern-triggered immunity (PTI)
 PTLDs *see* post-transplantation lymphoproliferative disorders (PTLDs)
 PTMs *see* post-translational modifications (PTMs)
 pulsed-field gel electrophoresis (PFGE) 1:622
 puma lentivirus (PLV) 2:56–57
 infections 2:66
 PLVA 2:66
 PLVB 2:66
 punta toro and codel viruses 2:771
 purification of viruses 1:166–167
 purification performance, assessing 1:172
 purifying selection, defined 4:457
 pustule, defined 2:868
 Puumala virus (PUUV) 2:353
 PUUV *see* Puumala virus (PUUV)
 PVAN *see* polyomavirus associated nephritis (PVAN)
 PVM *see* pneumonia virus of mice (PVM)
 PVR *see* poliovirus receptor (PVR)
 PVs *see* parvoviruses (PVs)
 PVSRIPO 1:659
 PVX *see* Potato virus X (PVX)
 PVY *see* potato virus Y (PVY)
 PVYNTN strain of Potato virus Y 1:649
 pyrobaculum filamentous virus 1 (PFV1) 4:363
Pyrobaculum oguniense 4:394
 PyVs *see* polyomaviruses (PyVs)

Q

QA *see* quality assurance (QA)
 QMS *see* quality management system (QMS)
 qPCR *see* quantitative polymerase chain reaction (qPCR)
 Q protein 4:89–90
 quadriparticulate virus, defined 4:642
 quadripartite genome, defined 4:642
 quadriviruses (*Quadriviridae*) 4:642, 4:506–508
 genome organization and expression 4:644–645
 molecular and biological properties 4:645–646
 phylogenetic and evolutionary relationships of RnQV1 to other dsRNA mycoviruses 4:646–647
 virion structure and composition 4:642–644
 qualitative studies 1:565
 quality assurance (QA) 5:55–57
 quality assurance (QA) and quality management (QM), principles of 5:64
 document control 5:65–66
 laboratory specimen retention times 5:67
 personnel training, competency, and continuous professional development (CPD) 5:67–68
 quality management documentation 5:64–65
 quality management system (QMS) 5:64, 5:65f
 specimen management 5:66
 transportation and storage of specimens 5:66–67
 quality assurance (QA) in the clinical virology laboratory
 assay verification and validation 5:70–72
 accuracy 5:72
 linearity/analytical measurement range (AMR) 5:72–73
 precision 5:72
 reportable range and limit of detection (LoD) 5:72
 EQA design and objectives 5:76–78
 equipment calibration 5:69
 equipment maintenance and monitoring 5:69–70
 equipment qualification 5:69
 external quality assessment and performance criteria 5:76
 future aspects of QA to virology laboratory 5:78–81
 internal and external audits 5:70
 internal quality assessment (IQA) 5:74
 internal quality control procedures and monitoring assay performance over time 5:74–75
 laboratory facilities and equipment 5:68
 sensitivity, specificity, and clinical utility 5:73–74
 suitable standards, reference materials and controls to support quality assurance in virology 5:75–76
 quality management documentation 5:64–65
 quality management system (QMS) 5:56–57, 5:64, 5:65f
 quantitative polymerase chain reaction (qPCR) 2:651
 quantitative spectral imaging 1:211
 quantitative trait loci, defined 3:69, 3:554, 3:555
 Quantum Dot, defined 1:208
 quasi-enveloped viruses 1:524
 quasi-enveloped virus particles, defined 2:397
 quasi-equivalence, defined 1:248, 1:278, 4:115
 quasi-equivalence theory 1:249–250
 predictions and limitations of 1:250
 quasi-equivalent conformers, defined 4:21
 quasispecies
 concept 1:68–69
 defined 2:373
 theory 1:58
 viral 1:58
 quasispecies-like populations, evolution as 3:100
 quiescent infection 2:456–457
 quinviruses (*Betaflexiviridae*) 3:642
 diagnosis 3:652
 epidemiology and control 3:651–652
 genome organization 3:643–648
 properties and functions of gene products 3:648–649
 replication and propagation 3:649
 taxonomy and phylogeny 3:642–643
 transmission and host range 3:649–651
 virion structure 3:643

R

rabbit hemorrhagic disease virus (RHDV)
 1:665–666, 1:669, 1:648
 classification 2:724
 clinical features 2:726
 diagnosis 2:728
 epidemiology 2:726
 genome 2:725–726
 pathogenesis and host immunity 2:726–728
 prevention and control 2:728–729
 treatment 2:729
 virion structure 2:724–725
 rabbit myxoma virus and fibroma viruses 2:730
 Californian myxoma virus 2:731–732
 classification 2:730
 clinical disease in European rabbits 2:732–733
 diagnosis 2:734–735
 epidemiology 2:731
 genome 2:730
 life cycle 2:730–731
 myxomatosis in European rabbits 2:732
 myxoma virus 2:731
 pathogenesis of myxomatosis in European rabbits 2:733–734
 prevention of myxomatosis 2:735
 treatment 2:735
 virion structure 2:730

- rabies and lyssaviruses
 classification (compact) 2:738
 clinical features 2:740–741
 diagnosis 2:742–743
 epidemiology 2:739–740
 genome 2:738
 life cycle 2:738–739
 pathogenesis 2:741–742
 prevention 2:744–745
 treatment 2:743–744
 virion structure 2:738
- rabies hyperimmune immunoglobulin 5:270–271
- rabies immunoglobulin (RIG) 2:744
- rabies lyssavirus (RABV) 2:738
- RABV *see* rabies lyssavirus (RABV)
- raccoon parvovirus (RPV) 2:684
- RACE *see* rapid amplified of cDNA ends (RACE)
- Rac prophage, defined 4:291
- radiological abnormalities, defined 2:814
- radiosensitizers for therapy and imaging 5:241
- Radish leaf curl betasatellite (RaLCB) 3:245
- Radi virus (RADV) 2:875
- RADV *see* Radi virus (RADV)
- RaLCB *see* Radish leaf curl betasatellite (RaLCB)
- raltegravir 5:146
 clinical trials in cART-experienced patients 5:146
 clinical trials in cART naïve individuals 5:146
 resistance 5:146
- randomized controlled trial, defined 5:98
- RAP *see* replicate-associated protein (RAP)
- rapid amplified of cDNA ends (RACE) 4:544
- rapid fluorescent focus inhibition test (RFFIT) 2:743
- Rathayibacter toxicus* 1:649
- Ravn virus (RAVV) 2:609
- RAVV *see* Ravn virus (RAVV)
- RBD *see* receptor binding domain (RBD)
- RBPs *see* receptor-binding proteins (RBPs)
- RBS *see* receptor-binding site (RBS)
- RBSDV *see* rice black streaked dwarf virus (RBSDV)
- RCA *see* regulators of complement activity (RCA); rolling circle amplification (RCA)
- RCR *see* rolling circle DNA replication (RCR); rolling circle replication (RCR)
- RCRE *see* replication initiation endonucleases (RCRE)
- RdDp *see* DNA-dependent RNA-polymerase (RdDp); RNA dependent DNA polymerase activity (RdDp)
- RDF *see* recombination directionality factor (RDF)
- RDG, defined 2:475
- RDR *see* recombination-dependent replication (RDR)
- RdRp *see* RNA-dependent RNA polymerase (RdRp)
- RDS *see* runt deformity syndrome (RDS)
- RDV *see* rice dwarf virus (RDV)
- reactivation, defined 2:442
- reactive oxygen species (ROS) 3:420
- read, defined 4:322
- read of insert (ROI) 1:181
- readthrough, defined 3:788, 3:481, 3:285
- readthrough protein, defined 3:106
- real-time PCR assays, validating 5:35, 3:629
 analytical specificity and sensitivity 5:40
 analytical verification stage 5:38
 reference materials and sample numbers 5:38
 template, primers and probes 5:38–39
 clinical validation stage 5:42
 precision, accuracy and trueness 5:42
 commercial assays 5:36–37
 consultation stage 5:37
 experimental optimization 5:40
 gold standard comparative study 5:42–43
 inhibition study 5:40–41
 inter-laboratory testing 5:43
 linear dynamic range or reportable range 5:41–42
 maintaining the validated status and re-validation 5:43–44
 normalization 5:40
 PCR efficiency 5:41
 preliminary considerations 5:37
 quantification standards, choice of 5:39
 reaction controls 5:39
 reverse transcription PCR 5:39
 study without a gold standard 5:43
 validation, need for 5:35–36
 in-house assays 5:35–36
 validation plan 5:37–38
 Westgard rules 5:39–40
- real-time polymerase chain reaction (RT-PCR) 1:497
- reassortant, defined 5:289, 2:349, 3:371
- reassortment, defined 1:62, 1:569, 3:516, 4:607, 2:654
- receptor, defined 1:402, 1:382, 4:380, 2:518
- receptor binding domain (RBD) 2:70–71
- receptor-binding proteins (RBPs) 1:129, 5:237, 1:402
- receptor-binding site (RBS) 1:525
- receptor-mediated, clathrin-dependent endocytosis 1:533
- receptor-mediated endocytosis 1:389
- receptor-mediated virus uncoating 1:389
- receptors, viral 1:388–389
 attachment factors 1:399–400
 co-receptors for infection 1:397–399
 evolution of virus-recept or interactions and viral tropism 1:400–401
 glycan receptors 1:390–391
 enteroviruses (EVs) 1:392
 glycan-based antiviral strategies 1:393–394
 glycosaminoglycan receptors 1:393
 histo-blood group antigen receptors (HBGAs) 1:392–393
 influenza viruses 1:391–392
 polyomaviruses (PyVs) 1:392
 sialic acid receptors 1:390–391
- identification of 1:389–390
- protein receptors, interactions with viruses 1:394–396
 complement control proteins 1:396–397
 ectoenzymes 1:397
 immunoglobulin superfamily proteins 1:394–396
 in virus cell entry 1:389
 attachment 1:389
 receptor-mediated endocytosis 1:389
 receptor-mediated virus uncoating 1:389
- recessive allele, defined 3:69, 3:554
- RecET, defined 4:291
- recombinant and mutant virus particles 1:223
- recombinant nematode anticoagulant protein c2 (rNAPc2) 2:616
- recombinant vesicular stomatitis virus (rVSV)-based Ebola vaccine 2:616
- recombinase, defined 4:291
- recombination 2:63–64, 1:56, 1:62, 1:569, 4:291, 1:71, 3:411, 2:245
 computational estimation of 1:108–109
 analyzing recombination in sequences derived from next generation sequencing 1:111–112
 estimation of recombination rates 1:111
 identification of recombination breakpoints 1:109–111
 phylogenetic recombination networks, reconstruction of 1:111
 rapid tests to detect the presence of recombination 1:108–109, 1:110f
 defective-interfering RNAs 1:465–467
 in DNA viruses 1:56–57
 homologous and non-homologous, defined 4:276
 reassortment 1:56
 recombination in DNA viruses 1:460–462
 recombination in RNA viruses 1:462–465
 template switching 1:56
- recombination-dependent replication (RDR) 4:62–63, 3:753–754
- recombination directionality factor (RDF) 4:77–78
- recombination in virus genetic data 1:112
 hepatitis B virus genotypes, recombination among 1:112–113
 HIV-1 centralized vaccines, recombination in the design of 1:112
 HIV-1 fitness recovery, influence of recombination on 1:112
 HIV-1 gp120, analyzing adaptation of 1:113
- recombineering
 combining with other gene modification technologies 4:299
 development of 4:294–295
 with dsDNA 4:295–296
 dsDNA recombineering, mechanism of 4:298
in vivo cloning 4:297
 markerless gene deletions and counterselection schemes 4:296

- recombineering (*continued*)
 oligo-mediated recombineering,
 mechanism of 4:297–298
 recombinase-independent recombination
 4:299–300
 recombineering in pathogenic bacteria
 4:298–299
 with ssDNA 4:296–297
 recombineering, defined 4:291
 recycling endosome pathway (REP) 2:72
 red deer, parapox of 2:668
 regulators of complement activity (RCA)
 1:396
 regulatory authority, defined 5:281
 regulatory T cells (Tregs) 1:588
 reinfection, defined 2:442
 release, defined 1:382
Reoviridae 3:545
Reoviridae family viruses, overall structure of
 1:303
 Sedoreovirinae 1:303
 orbivirus 1:303
 phytoreovirus 1:303
 rotavirus 1:303
 Spinareovirinae 1:303–306
 aquareovirus 1:306
 cypovirus 1:306
 orthoreovirus 1:305–306
 viral genome, general features of 1:306
 reoviruses 1:303, 4:867, 4:714
 capsid layers, structural organization of
 1:306–307
 cell entry of dsRNA viruses 1:307–309
 outer capsid layer 1:306–307
 double-stranded RNA (dsRNA) virus core
 and endogenous transcription 1:309
 viral innermost capsid layer that
 encloses the genome 1:309
 viral RNA capping 1:311–313
 viral RNA-dependent RNA polymerase
 (RdRp) 1:309–311
 double-stranded RNA (dsRNA) virus
 families 1:314–315
 genome replication and packaging 1:313
 replication and packaging 1:313–314
 viroplasm/replication factories 1:313
Reoviridae family viruses, overall structure
 of 1:303
 general features of the viral genome
 1:306
 Sedoreovirinae 1:303
 Spinareovirinae 1:303–306
 reoviruses of invertebrates 4:713–715
 Sedoreovirinae 4:877–879
 cardoreovirus 4:878–879
 phytoreovirus 4:879–881
 seadornavirus 4:881
 Spinareovirinae 4:868–870
 antigenic and genetic relationships
 4:876–877
 historical overview 4:870
 host range, diseases, transmission, and
 distribution 4:870–871
 virion properties, genome, and
 replication 4:871–875
 taxonomy 4:867
 REP *see* recycling endosome pathway (REP)
RepA gene 3:465
 repeat region 2:58–59
Rep gene 3:465
 Rep/Helicase, defined 4:849
 replicase-associated protein (RAP) 3:792–
 793
 replicase proteins 3:265
 replication, defined 4:136, 2:442
 Replication and Transcription Activator
 (RTA) 2:601
 replication cycle, defined 1:162
 replication initiation endonucleases (RCRE)
 4:390–391, 4:391
 replication initiator protein, defined 3:470
 replication of viruses
 assembly 1:386
 attachment 1:383–384
 double-stranded DNA 1:385
 double-stranded DNA with RNA
 intermediate 1:386
 double-stranded RNA 1:385
 entry 1:384
 maturation 1:386
 release 1:386–387
 single-stranded (–)-sense RNA 1:385
 single-stranded (+)-sense RNA 1:385
 single-stranded DNA 1:385
 single-stranded RNA with DNA
 intermediate 1:386
 transcription and genome replication
 1:384–385
 uncoating 1:384
 replication organelle (RO) 1:495, 1:498
 defined 1:495
 replication origin (*ori*), defined 4:61
 replicon, defined 2:386, 2:747
 replisome, defined 4:61
 Rep protein 3:360
 representational difference analysis, defined
 2:48
 representativeness 5:247, 5:252
 reproduction number, defined 2:825
 reptilian adenoviruses 2:13–14
 RER *see* rough endoplasmic reticulum
 (RER)
 reservoir, defined 2:117
 resistance, defined 3:192
 resistance breaking strain, defined 3:355
 resistance gene, defined 3:60, 3:554, 3:555
 resistance gene analog (RGA) marker 3:563
 resistant, defined 4:276
 resistosome 3:64–66, 3:60
 respiratory syncytial virus (RSV) 1:205,
 5:272–273, 5:123–124
 animal models of 2:753–754
 antiviral agents 2:755
 classification 2:748–749
 disease 2:747–748
 F glycoprotein 2:749–750
 genome transcription and replication
 2:752–753
 G glycoprotein 2:750–752
 history 2:747
 live attenuated vaccines 2:754–755
 management of RSV infections
 clinical features 5:155–156
 epidemiology 5:155
 pathogenesis and immunity 5:155
 prophylaxis 5:156–157
 treatment strategies 5:156
 vaccination 5:157
 maternal vaccination 2:754
 non-structural viral proteins 2:753
 passive immunity and prophylactic
 therapy 2:754
 protection from RSV disease 2:754
 vaccination 2:754
 RSV Fusion glycoprotein 5:124
 RSV RNA polymerase 5:123–124
 third viral glycoprotein 2:753
 treatment of 2:755
 vectored RSV vaccines 2:755
 virion organization 2:749f
 virus entry 2:749
 virion 2:749
 respiratory tract infections (RTI) 5:99–100
 assays 5:100–101
 background 5:99–100
 clinical impact 5:101
 lower 5:200
 upper 5:199–200
 restriction factor (RF) 1:627, 2:442, 1:628–
 629, 1:629–630
 restriction fragment length polymorphisms
 (RFLPs) 1:75, 3:192, 3:200
 restriction-modification system, defined 4:229
 reticulons, defined 3:32
 retinoblastoma protein, defined 2:79
 retinoic-acid-inducible gene 1 (RIG-I)
 2:391–392, 2:214, 2:555
 retroelements, classification of 3:98t
 retrograde, defined 2:738
 retrotranscription 2:62
 retrotranscytosis 1:529
 retrotransposons
 associated with invertebrates 4:722
 defined 3:274
 retrotransposons of plants
 Env-like functions in plants 3:102
 general classification 3:96–99
 long terminal repeat (LTR)
 retrotransposons
 different genera of 3:100–101
 life cycle of 3:99–100
 Metaviridae (Ty3/gypsy retrotransposons)
 3:101
 nomenclature issue 3:102–103
 Pseudoviridae (Ty1/copia retrotransposons)
 3:100–101
 quasispecies-like populations, evolution as
 3:100
 retroviral cycle, high similarities with
 intracellular steps of 3:99–100
 unclassified LTR retrotransposons 3:101–
 102
 retroviral cycle, high similarities with
 intracellular steps of 3:99–100
Retroviridae 2:56
 animal lentiviral infections 2:65–66
 bovine immunodeficiency (BI) and
 Jembrana disease (JD) 2:66

- equine infectious anemia (EIA) 2:67
 feline immunodeficiency and puma
 lentiviral infections 2:66
 visna-maedi (VM) and caprine arthritis
 and encephalitis (CAE) 2:66–67
 biological properties 2:63–64
 natural hosts 2:64
 transmission 2:64
 tropism 2:64
 variability 2:63–64
 classification 2:56–57
 encoded proteins 2:59–60
 genome 2:57–59
 pathogenic processes 2:64–65
 cell defenses and viral mechanisms 2:65
 immune deficiency 2:65
 inflammation 2:65
 replication cycle 2:60–63
 reverse transcription 2:62–63
 virion structure 2:57
Retroviridae 2:827–828
 clinical features 2:829–830
 defined 2:827
 epidemiology 2:829
 genome 2:828–829
 lifecycle 2:828–829
 immunity to 2:830
 taxonomy and classification 2:828
 virion structure 2:828
 retroviruses 1:525, 1:491–492
 see also fish retroviruses
 of birds
 classification 2:122–123
 clinical features 2:125
 diagnosis 2:126
 endogenous retroviruses 2:126
 epidemiology 2:124–125
 genome 2:123–124
 history 2:122
 life cycle 2:124
 pathogenesis 2:125–126
 treatment 2:126
 virion structure 2:123
 in cellular evolution 1:627
 germline insertions 1:630–631
 molecular consequences of retrovirus
 integration 1:627–629
 as selective agents 1:629f
 and their life cycle 2:316–317
 retrovirus integration, molecular
 consequences of 1:627–629
 host dependency factors (HDFs) 1:629
 host genes with indirect effects on
 retrovirus infection, spread or disease
 1:630
 host immunity 1:630
 restriction factors (RFs) 1:629–630
 retrovirus virions, structure of
 env gene 1:358–360
 gag proteins and retroviral virion structure
 1:352–354
 immature retroviral virion, assembly of
 1:356–358
 individual gag domains, structures of
 1:354–355
 capsid protein (CA) 1:355–356
 linker peptides 1:356
 matrix protein (MA) 1:354–355
 nucleocapsid protein (NC) 1:356
 reverse genetics, defined 4:26, 3:123
 reverse-transcribing RNA and DNA viruses,
 origin of 1:18–19
 reverse-transcribing viruses (RTV) 1:42–43,
 1:15–16
 genome organization and functions of
 gene products 3:654–656
 host range and evolutionary dynamics
 3:662–664
 impact and relationships of LTR
 retrotransposons and retroviruses
 with their hosts 3:664–665
 life cycle and replication 3:659–662
 LTR retrotransposons and reverse-
 transcribing viruses 3:654
 taxonomy, phylogeny and members of the
 families 3:656–659
 reverse transcriptase (RT) 2:464, 5:132,
 2:62f, 2:63f, 3:313
 reverse transcriptase inhibitors (RTIs)
 2:472–473
 reverse transcriptase protein, defined 3:96
 reverse transcriptase structure and functions
 5:131
 reverse transcription (RT)-PCR 3:629
 reverse transcription, defined 2:460, 5:27,
 3:158
 reverse transcription, process of 5:131–132
 reverse transcriptase (RT) 5:132
 RNase H 5:132
 tRNA primer 5:131–132
 reverse transcription loop-mediated rapid
 isothermal amplification (RT-LAMP)
 detection kit 4:825
 reverse transcription polymerase chain
 reaction (RT-PCR) 2:241, 5:98,
 5:300, 2:704, 2:705, 2:743, 2:38,
 2:812–813, 3:322, 5:29, 2:214,
 2:341, 2:542, 3:539, 2:514
 rev gene 2:61t
 Rev Response Element (RRE) 2:463
 RF see restriction factor (RF)
 RFFIT see rapid fluorescent focus inhibition
 test (RFFIT)
 RFLPs see restriction fragment length
 polymorphisms (RFLPs)
 RFS see ribosomal frameshift motif (RFS)
 R-gene 3:758–759
 R gene products 3:60–61
 R gene responses to viruses 3:61–62
 R genes 3:555
 RGHV see Ross' goose hepatitis virus
 (RGHV)
 Rhabdoviridae 4:715–716, 3:567
 rhabdoviruses 3:567–568
 classification 2:875
 control and treatment 2:882–883
 cytopathology and replication 3:576
 cell-to-cell movement 3:577–578
 entry 3:576
 morphogenesis 3:576–577
 transcription and genome replication
 3:576
 epidemiology 2:880–881
 genome structure and organization 3:568–
 572
 of invertebrates 4:715–716
 Raniviridae 4:716
 Sarothroviridae 4:716
 Solimiviridae 4:716–717
 life cycle 2:878–879
 assembly and budding 2:879–880
 attachment, entry and uncoating 2:878–
 879
 gene expression and replication 2:879
 inhibition and modification of host cell
 functions 2:880
 particle morphology and composition
 3:568
 pathogenesis and clinical features 2:881–
 882
 taxonomy and classification 3:568
 vector relationships, distribution and
 evolution 3:578–579
 viral proteins, properties of 3:572–573
 accessory proteins 3:575–576
 glycoprotein 3:575
 matrix protein 3:574–575
 nucleocapsid protein 3:572–573
 P3 movement protein 3:574
 phosphoprotein 3:573–574
 polymerase protein 3:575
 virus structure 2:875–878
 genomes and proteins 2:877–878
 rhabdoviruses of insects
 classification 4:883
 clinical features 4:886
 diagnosis 4:887
 epidemiology 4:886
 distribution, spatial and temporal 4:886
 population dynamics 4:886
 prevalence 4:886
 transmission 4:886
 genome 4:884–885
 almendravirus genome 4:885
 sigmavirus genome 4:885
 life cycle 4:885–886
 integration into the host genome
 4:885–886
 pathogenesis 4:886–887
 virion structure 4:883–884
 rhabdovirus glycoprotein G 1:348
 RHDV see rabbit hemorrhagic disease virus
 (RHDV)
 Rhinorrhea, defined 2:814
 Rhinovirus C 1:410
 rhinoviruses 2:758t
 classification 2:757
 clinical features 2:761
 acute otitis media (AOM) 2:761
 asthma 2:761–762
 asymptomatic infections 2:761
 bronchiolitis and recurrent wheezing
 2:761
 chronic obstructive pulmonary disease
 (COPD) 2:762
 common cold 2:761
 pneumonia 2:762
 diagnosis 2:762

- rhinoviruses (*continued*)
 culture 2:762–763
 nucleic acid detection 2:762
 sampling 2:762
 typing 2:763
 epidemiology 2:760
 genome 2:757–758
 life cycle 2:758–759
 assembly 2:760
 entry 2:759
 receptor 2:758–759
 replication 2:759–760
 translation 2:759
 uncoating 2:759
 pathogenesis 2:760–761
 prevention 2:764
 receptor usage of 2:759*t*
 structure 2:758*f*
 treatment 2:763–764
 virion structure 2:757
- Rhipicephalus appendiculatus* 1:547
- Rhizoctonia solani* 4:461
- Rhizoctonia solani* megabirnavirus 1 (RsMBV1) 4:600
- Rhizoctonia solani* RNA virus HN008 (RsRV-HN008) 4:600
- rhizomania, defined 3:213, 3:219
- Rhodobacteraceae* 4:334–335
- Rhopalosiphum padi virus (RhPV) 4:771
- RhPV *see* Rhopalosiphum padi virus (RhPV)
- RHR *see* rolling hairpin replication (RHR)
- ribavirin, defined 2:475
- ribavirine 5:171, 5:164–166
- Ribonuclease H (RNase H) 2:62, 5:132
- ribonucleoprotein (RNP) complexes 5:6, 2:237–238, 3:507, 4:776, 2:138–139
 cis-signals for formation of 4:624–625
 as a viral entity 4:622–623
- Ribosomal-1 frameshifting, defined 4:594
- ribosomal bypassing ('hopping') 1:452
- ribosomal frameshifting 2:193, 3:594, 3:154, 4:664, 3:712, 4:648, 4:892
- ribosomal frameshift motif (RFS) 4:804–806
- ribosomal read-through, defined 3:594
- ribosomal RNA (rRNA) 1:377
- ribosomal 'frame-shifting' 1:452–454
- ribosome-protected RNA fragments (RPFs) 1:178
- ribosome reinitiation 1:450–452
- ribosome shunt, defined 3:313
- ribosome 'shunting' 1:450
- ribosome stalling, defined 2:362
- ribosome stop codon 'read-through' 1:454
- Riboviria* 3:481
 defined 3:788
- ribovirocells 1:25
- ribozyme, defined 4:621, 3:852
- rice black streaked dwarf virus (RBSDV) 1:313
- rice dwarf virus (RDV) 1:303
- rice tungro bacilliform virus (RTBV) 3:669, 3:164, 4:662
 biotechnological applications 3:672
 control 3:671–672
 functions of proteins 3:671
 genome 3:669–671
 genome variability 3:671
 structure 3:669
 taxonomy 3:669
 transmission and host range 3:669
 virus-host relationships 3:671
- rice tungro disease 3:667–669, 3:667
- rice tungro spherical virus (RTSV) 3:672, 1:450, 4:662
 control 3:673
 gene functions 3:672
 genome 3:673
 genome diversity 3:672–673
 structure 3:672
 taxonomy 3:672
 transmission and host range 3:672
 virus-host relationships 3:673
- rice tungro virus 1:649
- rice yellow mottle virus (*Solemoviridae*) 3:675
 biological properties 3:676–677
 diagnostic and identification 3:677
 diversity and evolution 3:678–679
 epidemiological aspects 3:677–678
 organization of the genome 3:675–676
 relationships of the species with other taxa 3:676
 resistance 3:555
 virion properties 3:675
- rift valley fever virus 2:766, 2:768–770, 1:617
 bhanja virus 2:776
 chagres virus 2:771
 heartland banyangvirus (HRTV) 2:775–776
 huaiyangshan banyangvirus 2:771–775
 nonstructural proteins, functions of 2:767–768
 phleboviruses in the candiru antigenic complex 2:771
 punta toro and codel viruses 2:771
 rift valley fever virus 2:768–770
 sandfly fever naples and sandfly fever sicilian viruses 2:770–771
 toscana virus 2:771
 virion structure, genome, and strategy of replication 2:766–767
- RIG *see* rabies immunoglobulin (RIG)
- RIG-I *see* retinoic-acid-inducible gene 1 (RIG-I)
- rinderpest virus (RPV) 2:73
- ring vaccination 1:575
- RIP *see* RNAP inhibitory protein (RIP)
- RISC *see* RNA-induced silencing complex (RISC)
- RLPS *see* rough lipopolysaccharide (RLPS)
- RMSD *see* Root Mean Square Deviation (RMSD)
- 3' RNA, defined 1:345
- RNA dependent DNA polymerase activity (RdDp) 2:60
- RNA-dependent RNA polymerase (RdRp) 4:806, 2:177, 2:389, 1:128, 4:820–821, 1:437, 3:623, 4:771, 1:62, 4:544, 4:867, 4:699, 3:388, 3:336, 3:545, 1:441, 3:555, 3:559
 containing evolutionarily conserved architecture 1:441
 coordinating transcription, capping, and polyadenylation 1:442–443
 different RdRp and template conformations distinguish transcription from replication 1:441–442
 replication using 1:437
 RNA drug, defined 3:43
 RNAi *see* RNA interference (RNAi)
 RNA-induced silencing complex (RISC) 3:792–793, 4:592, 3:788
 RNA-induced silencing complex, defined 3:52, 3:123
 RNA interference (RNAi) 4:832, 3:777, 4:808, 3:192, 2:616, 4:821, 4:772, 3:667, 3:60, 3:52, 3:123, 4:776, 3:362–363, 3:554
 RNAP *see* RNA polymerase (RNAP)
 rNAPc2 *see* recombinant nematode anticoagulant protein c2 (rNAPc2)
 RNAP inhibitory protein (RIP) 4:393
 RNA polymerase (RNAP) 4:393, 2:25, 4:238
 RNA-primed DNA replication 1:430–433, 1:429
 RNA recombination and reverse genetics, manipulating CoV genomes using 2:204–205
 RNase H *see* Ribonuclease H (RNase H)
 RNA-Seq *see* sequencing of total RNA (RNA-Seq)
 RNA-sequencing 1:178
 RNA silencing 3:260, 3:628–629, 3:192, 3:293, 3:388, 3:285, 3:439, 4:457, 4:601, 4:607, 4:648, 3:719
 RNA silencing suppressor 3:260, 3:623, 3:628–629, 3:439, 3:229, 3:719, 3:642
 RNA splicing, defined 2:460
 RNA vaccines 5:286
 RNA viruses 1:40–42
 infecting marine protists 4:671
 in marine microbial ecology 4:674–675
 origin of 1:14–15
 infecting eukaryotes 1:17–18
 recombination in 1:462–465
 RNA virus evolution, mechanisms of additional constraints on RNA virus evolution 1:69
 coevolution with hosts 1:67–68
 genetic drift 1:68
 genetic exchange 1:64–65
 detecting recombination 1:65
 rates of recombination 1:65
 reassortment 1:65
 recombination 1:64–65
 recombination and reassortment, evolutionary outcomes for 1:65–66
 mutation rates of RNA viruses 1:62–64
 natural selection in viruses 1:66–68
 quasispecies concept 1:68–69
 viruses under selection by factors other than host cells 1:68
- RnFV1 *see* Rosellinia necatrix fusarivirus 1 (RnFV1)

- RnMBV2 *see* Rosellinia necatrix megabirnavirus 2 (RnMBV2)
- RnMBV3 *see* Rosellinia necatrix megabirnavirus 3 (RnMBV3)
- RNP complexes *see* ribonucleoprotein (RNP) complexes
- RnPV1 *see* Rosellinia necatrix partitivus 1 (RnPV1)
- RO *see* replication organelle (RO)
- rod-shaped helical plant viruses 1:362–364
- ROI *see* read of insert (ROI)
- rolling circle, defined 3:768
- rolling circle amplification (RCA) 3:168, 3:749, 3:753–754, 3:411
- rolling circle DNA replication (RCR) 1:434–435, 1:429
- rolling circle replication (RCR) 2:184, 1:436–437, 2:182, 4:387, 4:53, 3:470, 3:149
- rolling hairpin replication (RHR) 4:836
- Roniviridae 4:716
- Roniviruses 2:256
- room temperature electron microscopy (RT-EM) 1:242
- room temperature-electron tomography (RT-ET) 1:244
- root and tuber crops viral diseases 3:83–85
- cassava viral diseases 3:83–85
- potato viral diseases 3:85–86
- sweet potato viral diseases 3:86
- Root Mean Square Deviation (RMSD) 1:154
- root-nodules, defined 4:632
- ROS *see* reactive oxygen species (ROS)
- Rosellinia necatrix* 4:461
- hypovirulence in 4:476
- Rosellinia necatrix fusarivir* 1 (RnFV1) 4:526, 4:527
- Rosellinia necatrix megabirnavirus* 2 (RnMBV2) 4:597, 4:599–600
- Rosellinia necatrix megabirnavirus* 3 (RnMBV3) 4:600
- Rosellinia necatrix partitivus* 1 (RnPV1) 4:433
- roseolaviruses 2:442
- defined 2:442
- roseolovirus cell tropism 2:781
- HHV-6A and HHV-6B 2:781
- HHV-7 2:781
- roseoloviruses 1:321
- chromosomal integration of HHV-6A and HHV-6B 2:781–782
- chromosomal integration, mechanism of 2:781–782
- CIHHV-6 and disease 2:784–785
- CIHHV-6 and potential for disease in the immunocompetent 2:784–785
- CIHHV-6 and potential for disease in the immunocompromised 2:785
- epidemiology 2:783
- HHV-6, HHV-7 and immunomodulation 2:784
- HHV-6 and HHV-7 and antiviral therapy 2:787–788
- HHV-6 and HHV-7 disease in older children and adults 2:785
- Alzheimer's disease 2:786
- delayed primary infection 2:785
- HHV-6 and HHV-7 and the possibility of encephalitis 2:785
- HHV-6 and multiple sclerosis 2:786
- HHV-6B and DRESS (drug rash with eosinophilia and systemic symptoms) 2:785–786
- HHV-6 and HHV-7 laboratory tests and diagnosis of infection 2:786–787
- antibody detection – HHV-6 and HHV-7 2:787
- CIHHV-6 and misdiagnosis of active infection 2:787
- diagnosis of CIHHV-6 2:787
- virus detection – HHV-6A, HHV-6B and HHV-7 2:787
- HHV-6 and HIV 2:786
- HHV-6B and disease after organ transplant 2:786
- HHV-6B encephalitis after HSCT 2:786
- HHV-6B and HHV-7 primary infection and disease 2:785
- encephalitis 2:785
- exanthem subitum 2:785
- FSE and temporal lobe epilepsy 2:785
- HHV-7 and disease after organ transplant 2:786
- immune responses 2:784
- occurrence 2:783
- chromosomal integration 2:783
- HHV-6A 2:783
- HHV-6B and HHV-7 primary infections 2:783
- roseolovirus cell tropism 2:781
- HHV-6A and HHV-6B 2:781
- HHV-7 2:781
- tissue distribution 2:783
- transmission 2:783
- horizontal transmission of HHV-6 and HHV-7 2:783
- transmission of HHV-6 by organ donation 2:783–784
- vertical transmission of HHV-6A and HHV-6B 2:783
- viral genomics 2:779–780
- HHV-6A and HHV-6B 2:779–780
- HHV-7 2:780
- viral latency 2:782
- CIHHV-6 and the possibility of reactivation 2:783
- reactivation from latency 2:782–783
- state of viral genome during latency 2:782
- virus structure and replication 2:780–781
- rose rosette virus 3:19
- causal agent and classification 3:19
- control 3:20
- disease symptoms and yield losses 3:19
- epidemiology 3:19–20
- geographical distribution 3:19
- Rossmann fold 1:306–307, 4:148
- Ross River virus 1:263
- Ross' goose hepatitis virus (RGHV) 2:100
- rotaviruses 1:393, 1:303
- classification 2:789
- clinical features 2:792
- diagnosis 2:793
- epidemiology 2:791–792
- genome 2:790
- immune response and prevention 2:793–795
- life cycle 2:790–791
- pathogenesis 2:792–793
- treatment 2:793
- vaccines 5:289–290
- in development 5:291
- licensed 5:290
- safety monitoring 5:291
- vaccine effectiveness and impact 5:290–291
- virion structure 2:789–790
- rough endoplasmic reticulum (RER) 2:185
- rough lipopolysaccharide (RLPS) 4:27
- Roussetus aegyptiacus* 2:608f
- Rous sarcoma virus (RSV) 1:675, 1:653–654, 1:352, 1:452, 2:122
- RPFs *see* ribosome-protected RNA fragments (RPFs)
- RpRp 3:790
- RPV *see* raccoon parvovirus (RPV)
- RRE *see* Rev Response Element (RRE)
- rRNA *see* ribosomal RNA (rRNA)
- RsMBV1 *see* Rhizoctonia solani megabirnavirus 1 (RsMBV1)
- RSV *see* respiratory syncytial virus (RSV); Rous sarcoma virus (RSV)
- RT *see* reverse transcriptase (RT)
- RTA *see* Replication and Transcription Activator (RTA)
- RTase/RNaseH, defined 3:274
- RTBV *see* rice tungro bacilliform virus (RTBV)
- RT-EM *see* room temperature electron microscopy (RT-EM)
- RT-ET *see* room temperature-electron tomography (RT-ET)
- RTI *see* respiratory tract infections (RTI)
- RTIs *see* reverse transcriptase inhibitors (RTIs)
- RT-LAMP detection kit *see* reverse transcription loop-mediated rapid isothermal amplification (RT-LAMP) detection kit
- RT-PCR *see* real-time polymerase chain reaction (RT-PCR)
- RT-PCR *see* reverse transcription polymerase chain reaction (RT-PCR)
- RTSV *see* rice tungro spherical virus (RTSV)
- RTV *see* reverse-transcribing viruses (RTV)
- rubella virus (RUBV) 1:263, 2:797, 1:499
- classification 2:797
- clinical features of infection 2:801–802
- diagnosis 2:803
- epidemiology 2:801
- future 2:804
- genetics and evolution 2:800–801
- genomic organization 2:798–799
- host range and virus propagation 2:797–798
- immune response and serodiagnosis 2:802–803
- intracellular replication cycle 2:799–800

rubella virus (RUBV) (*continued*)
 pathogenesis, pathology, and histopathology 2:802
 prevention and control of rubella 2:803–804
 properties of the virion 2:798
 serologic relationships and variability 2:801
 transmission and tissue tropism 2:801
 Rubus yellow net virus (RYNV) 3:165
 RUBV *see* rubella virus (RUBV)
Rudiviridae 1:367–369
 rule-of-six, defined 2:355
 runt deformity syndrome (RDS) 4:839
 rVSV-based Ebola vaccine *see* recombinant vesicular stomatitis virus (rVSV)-based Ebola vaccine
 Rx gene 3:629, 3:623
 rymoviruses 3:797
 biotechnological application 3:803
 diagnosis 3:802
 disease management 3:802
 cultural practice 3:802
 genetic resistance 3:802–803
 functions of viral proteins 3:799
 genome organization 3:799
 life cycle and epidemiology 3:799–800
 disease cycle 3:801–802
 host range 3:800
 seed transmission 3:800–801
 vector transmission 3:800
 pathogenicity 3:802
 taxonomy and classification 3:797
 virion structure 3:797–799
 RYNV *see* Rubus yellow net virus (RYNV)

S

S2 gene 2:61t
 sabin vaccine strains, attenuation of 2:690
Saccharolobus solfataricus 4:388–389
Saccharomyces cerevisiae 4:431, 4:440
 dsRNA viruses and killer phenotype expression in 4:535–536
Saccharomyces cerevisiae virus L-A (ScV-L-A) 4:432
 S-adenosylmethionine (SAM) 1:309, 2:177
 sadwavirus
 classification 3:322
 clinical features and pathogenesis 3:325
 diagnosis 3:325
 genome organization 3:322
 geographic distribution 3:324
 life cycle and epidemiology 3:322–324
 host range 3:324
 prevention 3:326
 transmission 3:324–325
 treatment 3:325–326
 virion structure 3:322
 SAGs *see* single amplified genomes (SAGs)
 Saint Louis encephalitis virus (SLEV) 2:805, 2:805–806
 classification/genetic characterization 2:808–809

diagnosis 2:812–813
 epidemiology 2:810–812
 history 2:805–808
 host range 2:810–812
 life cycle 2:809–810
 pathogenesis/clinical features 2:812
 prevention 2:813
 virion structure/genome 2:806–808
 salmonid, defined 2:324
 salmon swimbladder sarcoma
 -associated retrovirus 2:320
 isolation of a retrovirus from 2:320–321
 salmon swimbladder sarcoma virus (SSSV)
 pathogenesis of 2:322
 prevalence, seasonality, and transmission of 2:321–322
 sequence comparisons of SSSV with other retroviruses 2:321
 SAM *see* S-adenosylmethionine (SAM)
 sandfly fever Naples and sandfly fever sicilian viruses 2:770–771
 sanger sequencing technique 5:28
 SaPIs, *see* *Staphylococcus aureus* pathogenicity island (SaPIs)
 SAPNs *see* self-assembling protein nanoparticles (SAPNs)
 sapovirus 2:490
 cell culture 2:491
 classification 2:490
 clinical features 2:491
 diagnosis 2:491
 epidemiology 2:490–491
 genome 2:490
 immunity 2:491
 life cycle 2:490
 pathogenesis 2:491
 prevention 2:491–492
 virion structure 2:490
 sarcoids 2:87–88
 cats 2:88
 diagnosis 2:90
 horses 2:87–88
 treatment 2:90–91
 SAR endolysins, defined 1:501
 SARS *see* severe acute respiratory syndrome (SARS)
 SARS-CoV *see* severe acute respiratory syndrome coronavirus (SARS-CoV)
 SARS-CoV-2 *see* severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
Sarothroviridae 4:716, 4:889
 sarthroviruses 4:888
 clinical signs of WTD and histopathology 4:888
 control measures 4:890–891
 diagnostic tools 4:890
 genome organization 4:888–889
 geographical distribution 4:888
 host range 4:890
 morphology of virus 4:888
 pathogenicity and transmission of disease 4:889–890
 susceptibility of cell line to XSV 4:890
 taxonomic position 4:889
 satellite conserved region 3:239

satellite DNA 3:239
 satellite dsRNA, defined 4:534
 satellite like molecule, defined 3:149
 satellite-like RNA, defined 3:681
 satellite nucleic acids and viruses 3:681
 classification 3:682
 expression of foreign sequences from satellite vectors 3:691
 general properties and effects of satellites 3:682–685
 satellite viruses 3:682–685
 geographical distribution 3:682
 history 3:681–682
 replication and structure 3:688–690
 satellite RNAs 3:688–690
 satellite DNAs 3:690
 satellite-mediated control of viruses 3:691
 satellite nucleic acids and satellite-related nucleic acids 3:685–688
 sequence variation and evolution 3:690–691
 satellite phage 4:90
 defined 4:98
 satellite RNAs (satRNAs) 3:795, 3:692, 3:712, 3:581, 3:371, 3:285, 3:456, 3:681
 satellites, defined 3:470, 3:21
 satellite tobacco mosaic virus (STMV) 3:740
 satellite tobacco necrosis virus (STNV) 1:249
 satellite virus, defined 3:581, 3:681, 3:456, 4:888
 satRNAs *see* satellite RNAs (satRNAs)
 SAXS *see* small angle X-ray scattering (SAXS)
 SBL technology *see* sequencing-by-ligation (SBL) technology
 SBMV *see* southern bean mosaic virus (SBMV)
 SBV *see* schmallenberg virus (SBV)
 scab, defined 2:343
 scabby mouth 2:667–668
 scaffolding protein 1:444, 4:118–119
 accelerating assembly 4:119
 -assisted capsid assembly 1:483–486
 defined 4:10, 4:105
 delay MCP transformation 4:119–120
 exclude host factors 4:120
 promote correct geometry 4:119
 scaffold involvement in portal incorporation 4:120
 scanning electron microscopy, advances in 5:13–14
 Scanning Electron Microscope (SEM) 5:5
 SCARB2 *see* scavenger receptor class B member 2 (SCARB2)
 scavenger receptor class B member 1 (SRB1) 2:390
 scavenger receptor class B member 2 (SCARB2) 1:284, 1:410
 SCC *see* squamous cell carcinoma (SCC)
 Schlumbergera virus X (SchVX) 3:623–624, 3:628
 schmallenberg virus (SBV) 2:34, 1:573–574
 classification 2:34
 clinical features 2:36
 diagnosis 2:38

- epidemiology 2:35–36
genome 2:34–35
pathogenesis 2:36–38
prevention 2:38–39
viral replication 2:35
virion structure 2:34
SchVX *see* Schlumbergera virus X (SchVX)
science, viruses and 1:672–674
harnessing of viruses by humans 1:675
bioweapons 1:676–677
commerce and tulip mania 1:675
gene therapy and cancer therapy 1:676
phage therapy 1:676
vaccines 1:675–676
impact of science and technology on virology 1:673–674
impact of viruses on science 1:674–675
Sclerophthora macrospora virus A (SmV-A) 4:438
Sclerophthora macrospora virus B (SmV-B) 4:438
sclerotia, defined 4:552
sclerotinia sclerotimonavirus 4:481, 4:438
biological properties 4:481
family –*Mymonaviridae* 4:481
genome structure 4:481
genus –*Sclerotimonavirus* 4:481
phylogenetic relationships 4:483
virion morphology 4:481–483
Sclerotinia sclerotiorum 4:441, 4:615, 4:493
hypovirulence in 4:475–476
mycoviruses' infections in *see* mycoviruses' infections in *Sclerotinia sclerotiorum*
Sclerotinia sclerotiorum botybirnavirus 1 (SsBV1) 4:464
Sclerotinia sclerotiorum debilitation-associated RNA virus (SsDRV) 4:438, 4:462
Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1) 4:520
distribution of 4:498
exploring SsHADV-1 to control fungal disease 4:501–503
extracellular entry of 4:495–496
genome and proteins of 4:494–495
host range of 4:495
mutualistic interaction with mushroom sciarid fly 4:496–497
taxonomy of 4:498–501
transmission of 4:497–498
Sclerotinia sclerotiorum hypovirus 1 (SsHV1) 4:463
Sclerotinia sclerotiorum integrin-like gene (SSITL) 4:462
Sclerotinia sclerotiorum megabirnavirus 1 (SsMBV1) 4:597, 4:599
Sclerotinia sclerotiorum mycoreovirus 4 (SsMYRV4) 4:520
Sclerotinia sclerotiorum negative-sense RNA virus 1 (SsNSRV-1) 4:463
Sclerotinia sclerotiorum reoviruses 4:612
mycoreovirus 4 4:612
sclerotinia sclerotiorum reovirus 1 4:612
sCMOS camera, defined 1:208
scoliosis, defined 2:34
SCP *see* smallest capsid protein (SCP)
screening for viral infections 5:91
culture 5:91
high throughput serology, multiplex platforms for 5:94–95
nucleic acid tests 5:91–92
clustered regularly interspaced short palindromic repeats (CRISPER) platforms 5:93
DNA microarrays 5:92
high throughput sequencing 5:92–93
multiplex PCR 5:92
polymerase chain reaction assays 5:92
VirCapSeq-VERT 5:93
pathology 5:91
serology 5:93–94
enzyme-linked immunosorbent assays (ELISA) 5:94
hemagglutination inhibition assays 5:93–94
immunoblot assays 5:94
immunofluorescent assays (IFA) 5:94
lateral flow assays (LFA) 5:94
plaque reduction neutralization (PRNT) assays 5:94
“scrunchworm” model 4:165
ScV-L-A, *see* *Saccharomyces cerevisiae* virus L-A (ScV-L-A)
Scylla serrata reovirus 4:879
SD *see* sleeping disease (SD)
SDA *see* Strand Displacement Amplification (SDA)
SDS-PAGE *see* sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)
SDT *see* Sequence Demarcation Tool (SDT)
seadornavirus 4:881
sea ice 4:353
seasonal cycle of disease, control of 2:319
seasonal influenza 5:273–274, 5:160
sea star-associated densovirus (SsADV) 4:840
sea-star wasting disease (SSWD) 4:839, 4:840
secondary structure, defined 1:62
second generation/parallel sequencing technologies 5:29–31, 1:175
application of NGS in virology 5:30–31
Secoviridae 3:322, 3:486, 3:703, 3:364, 3:348
secoviruses (*Secoviridae*) 3:692
biotechnological applications 3:701
diagnosis 3:701–702
epidemiology and control 3:700–701
genome organization 3:697–698
properties and functions of gene products 3:698–699
replication and propagation 3:699–700
taxonomy, phylogeny, and evolution 3:692–694
members of the family 3:694
transmission, host range 3:700
virion structure 3:694–697
sedimentation rate, defined 1:162
Sedoreovirinae 1:303, 4:877–879, 4:713–715
cardoreovirus 4:878–879
orbivirus 1:303
phytoreovirus 1:303
phytoreovirus 4:879–881
rotavirus 1:303
seadornavirus 4:881
seed and sap transmission 3:531
segmental mobility in proteins, defined 3:727
segmented negative strand (sNS) RNA genomes 2:507–508
select agent, defined 2:765
selection coefficient, defined 1:53
selection/counterselection, defined 4:291
Selective Sweep, defined 5:227
SELEX, defined 1:278
self-assembling protein nanoparticles (SAPNs) 1:251
self-assembly, defined 4:26
selfish RNA replicons, origin of 1:15–16
self versus nonself recognition innate immunity systems 1:607–610
SEM *see* Scanning Electron Microscope (SEM)
semigranular cells (SGCs) 4:816
semi-persistent transmission, defined 3:667, 3:313, 3:229, 3:703
Semliki Forest virus (SFV) 1:546, 1:499
semotivirus, defined 3:653
sensitivity 5:73–74
analytical 2:697
diagnostic 2:697
sentinel studies 1:566
sentinel surveillance 5:250
sepsis, defined 2:675
sequelae, defined 2:868
sequence-based virus classification methods 1:101–102
DemARC (DivErsity pArtitioning by hieRarchical Clustering) 1:101–102
Genome Relationships Applied to Virus Taxonomy (GRAViTy) 1:102
Natural Vector method 1:102
Sequence Demarcation Tool (SDT) 1:102
ViCTree 1:102
ViPTree 1:102
sequence-characterized amplified region (SCAR) marker 3:562
sequence comparison methods 1:100–101
Sequence Demarcation Tool (SDT) 1:102
sequence identity, defined 2:442
Sequence Independent Single Primer Amplification (SISPA) 1:178
sequence specific binding, defined 4:136
sequencing-by-ligation (SBL) technology 5:30
Sequencing by Synthesis Technology, defined 4:322
sequencing library, defined 1:175
sequencing of total RNA (RNA-Seq) 1:178
sequencing strategies 5:27–28
first generation sequencing technology 5:28–29
sample preparation strategies for sequencing viruses 5:31–32
nucleic acid sequence dependent strategies for known viruses 5:31–32
nucleic acid sequence independent strategies for sequencing 5:32

- sequencing strategies (*continued*)
 second generation/parallel sequencing technologies 5:29–31
 application of NGS in virology 5:30–31
 third generation sequencing technologies 5:31
 virus databases 5:32–34
- sequiviruses 3:703
 classification 3:703–704
 diseases and management 3:710
 genome organization 3:707–708
 properties of viral proteins 3:708–709
 properties of virions 3:707
 transmission 3:709–710
 variation of isolates and strains 3:704–707
- serine-like protease, defined 2:362
 serine protease fold 1:263
 seroconversion, defined 5:289
 serodiagnosis, defined 2:797
 seroepidemiologic studies 1:566
 serogroup, defined 3:184
 serological approaches for viral diagnosis 5:15
 antibody assay, neutralizing 5:19
 antibody repertoire, sequencing analysis of 5:20
 complement fixation 5:20
 future perspectives 5:20–21
 general considerations and sources of error 5:17
 restrictions and limitations 5:17
 hemagglutination inhibition 5:20
 IgG avidity 5:16–17
 immunoassays 5:17–18
 homogenous wash-free immunoassays 5:19
 immunoblotting 5:18
 immunofluorescence microscopy 5:18–19
 lateral-flow immunoassays 5:19
 solid-phase microwell immunoassays 5:17–18
 latex agglutination 5:19–20
 point-of-care serodiagnostics 5:20
 principles of serological assays 5:15–16
 serological differentiation index, defined 3:727
 serological surveillance 5:249
 serological tests 5:211–212
 hepatitis A virus 5:211–212
 hepatitis E virus 5:212–213
 seronegative, defined 2:442
 seropositive, defined 2:442
 seroprevalence, defined 2:518, 2:442
 serotype, defined 2:789, 3:184
Serpentovirales 3:495
 serum neutralization tests (SNT) 2:285
 serum therapy 5:267
 SEV1 *see* sulfolobus ellipsoid virus 1 (SEV1)
 17D Yellow fever virus vaccine 2:897
 78-kD protein, defined 2:765
 severe acute respiratory syndrome (SARS) 2:815*f*, 2:816*f*, 5:285, 1:12, 5:7, 1:499
 animal infection models 2:821
 clinical features 2:819–820
 epidemiology
 ecology, animal reservoir and zoonotic transmission 2:818
 history 2:814
 emergence of SARS 2:814
 human-to-human transmission 2:819
 laboratory diagnosis 2:820–821
 pathogenesis 2:822
 treatment 2:822
 vaccines and immunity 2:823
 virology 2:814–815, 2:815, 2:815–816
 phylogeny 2:816–818
 virus receptors 2:818
 severe acute respiratory syndrome
 coronavirus (SARS-CoV) 4:804, 2:198, 2:200*f*, 2:825, 2:825–826, 2:826, 1:121, 2:508, 2:193
 severe acute respiratory syndrome
 coronavirus 2 (SARS-CoV-2) 1:13, 5:286–287, 5:3, 5:261, 5:127
 severe fever with thrombocytopenia syndrome virus (SFTSV) *see* huaiyangshan banyangvirus
 sexually transmitted infections (STIs) 2:467
 SFFV *see* spleen focus-forming virus (SFFV)
 SFS *see* Site Frequency Spectrum (SFS)
 SFV *see* Semliki Forest virus (SFV)
 SFV1, *see* *Sulfolobus* filamentous virus 1 (SFV1)
 SGCs *see* semigranular cells (SGCs)
 SGHBV *see* snow goose hepatitis B virus (SGHBV)
 SGL systems *see* single gene lysis (SGL) systems
 sGP *see* soluble glycoprotein (sGP)
 sgRNA *see* sub-genomic RNA (sgRNA)
 sharka disease, defined 3:586
 shedding, virus 1:561
 sheppox and goatpox virus 2:166–167
 clinical features and pathology 2:167
 diagnosis 20934:s0060
 epidemiology 2:166–167
 histopathology 2:167
 pathogenesis 2:167
 prevention 20934:s0070
 treatment 20934:s0065
 Sheppox virus (SPPV) 2:165, 2:167
 sheldgoose hepatitis B virus 2:100
Shewanella piezophila 4:55
 SHFV infecting monkeys *see* simian hemorrhagic fever virus (SHFV) infecting monkeys
 SHIBV *see* Shimoni lyssavirus (SHIBV)
 shielding, defined 5:233
Shigella flexneri 4:200
 Shimoni lyssavirus (SHIBV) 2:738
 Shingrix, development of 5:282
 SHIV, defined 2:827
 SHM *see* somatic hypermutation (SHM)
 S-hole 1:505
 short non-contractile tails 4:191–192
 short read sequencing (SRS) 1:182, 1:180–181
 short-tailed phages (*Podoviridae*) 4:208–209
 short-term pathogenesis 1:59
 shotgun metagenomics 1:623
 shotgun sequencing 1:125, 1:175, 5:27, 1:552
 SHP *see* Structural Homology Program (SHP)
 SH protein *see* small hydrophobic (SH) protein
 shrimp densoviruses 4:839–840
 fenneropenaeus chinensis
 hepanadenovirus (FchDV) 4:840
 new invertebrate DVs 4:840
 penaeus stylirostris penstydendovirus 1 (PstDV1) 4:839–840
Siadenovirus 2:14
 sialic-acid-binding immunoglobulin-like lectin-1 (Siglec-1) 1:399–400
 sialic acid receptors 1:390–391
 sialoadhesin 2:703
Sibine fusca 4:839
 sida micrantha mosaic virus (SiMMV) 3:194
 SiDV *see* solenopsis invicta densovirus (SiDV)
 replication, defined 4:61
Sigmavirus 4:715–716
 sigmavirus genome 4:885
 Sigmodontine-borne hantaviruses 2:351
 Signal Peptidase I (SPI) 1:501
 Signal Peptidase II (SPII) 1:501
 signal transducer and activator of transcription (STAT) 2:555–556
 silencing suppression, defined 3:293
 silencing suppressor, defined 3:788, 3:667, 3:383, 3:285, 3:456, 4:632, 3:594
 Simbu serogroup 2:663
 simeprevir 5:123
 simian 40 virus (SV40) 1:675, 1:653–654
 simian adenoviruses 2:12
 simian hemorrhagic fever virus (SHFV) infecting monkeys 2:256
 simian immune deficiency virus (SIV) 2:56–57, 2:64, 1:121, 2:827–828, 2:460
 clinical features 2:829–830
 defined 2:827
 epidemiology 2:829
 genome 2:828–829
 lifecycle 2:828–829
 immunity to 2:830
 taxonomy and classification 2:828
 virion structure 2:828
 similarity, defined 4:276
 similarity score, defined 1:153
 SiMMV *see* sida micrantha mosaic virus (SiMMV)
 Sindbis-like viruses 1:66
 Sindbis virus (SINV) 1:263
 classification 2:837
 clinical features 2:840–841
 diagnosis 2:841
 epidemiology 2:838–840
 genome 2:838
 geographical distribution of 2:837*f*
 life cycle 2:838, 2:839*f*
 pathogenesis 2:841
 prevention 2:841
 SINV proteins and their key functions 2:838*t*
 treatment 2:841

- virion structure 2:837–838
 single amplified genomes (SAGs) 4:329–330
 single cell sequencing 1:180
 defined 1:175
 to investigate cellular heterogeneity and its impact on viral infection 1:180
 to investigate viral heterogeneity 1:180
 single gene lysis (SGL) systems 1:514–515, 1:501–502
 Archaea, phage egress in 1:517–518
 extrusion 1:517
 of *Leviviridae* 1:515–516
 mysterious L 1:516–517
 X174 E 1:514–515
 small interfering RNA (siRNA) 3:555, 3:560
 single jelly roll (SJR) 1:16
 single molecule Förster resonance energy transfer (smFRET) 4:57, 1:213–215
 Single Molecule Localization Microscopy, defined 1:208
 single molecule sequencing 1:180–181
 defined 1:175
 merging short read sequencing (SRS) and long read sequencing (LRS) 1:182
 Oxford Nanopore technologies 1:181–182
 Pacific Biosciences 1:181
 single nucleotide polymorphism (SNP)
 4:865, 1:410, 1:74, 1:75
 defined 1:71
 single particle analysis (SPA) 1:364–365, 5:5
 Single Primer Isothermal Amplification (SPIA) 1:178
 single radial immunodiffusion (SRID) 5:302
 single-step growth, defined 4:314
 single-strand annealing, defined 4:291
 single-stranded (+)-sense RNA 1:385
 single-stranded (–)-sense RNA 1:385
 single-stranded DNA (ssDNA) mycoviruses 4:493
 discovery of DNA mycoviruses 4:493–494
 host of DNA mycovirus 4:493
 Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1 (SsHADV-1)
 distribution of 4:498
 exploring SsHADV-1 to control fungal disease 4:501–503
 extracellular entry of 4:495–496
 genome and proteins of 4:494–495
 host range of 4:495
 mutualistic interaction with mushroom sciarid fly 4:496–497
 taxonomy of 4:498–501
 transmission of 4:497–498
 single-stranded DNA (ssDNA) viruses 1:43, 1:138, 1:385, 4:835, 2:182, 2:186, 1:19, 4:296–297
 single-stranded DNA binding protein (SSB) 1:433–434
 single-stranded RNA (ssRNA) 3:555
 single-stranded RNA (ssRNA) bacterial viruses 5:131, 3:158, 4:621, 2:130
 biotechnological applications 4:25
 icosahedral ssRNA bacterial viruses 4:21–24
 background 4:21–24
 phage lifecycle 4:21–24
 phage structure and assembly 4:24–25
 single-stranded RNA (ssRNA) mycoviruses 4:437–438
 unassigned ssRNA viruses 4:438
 single-stranded RNA with DNA intermediate 1:386
 single-virus genomics
 best practices and protocol for 1:184–185
 sample size and preservation 1:185
 sequencing 1:186
 staining and single-virus separation (sorting) 1:185
 viral capsid lysis and whole genome amplification 1:185–186
 fluorescence-activated virus sorting 1:187
 DNA decontamination and fluorescence-activated virus sorting preparation 1:187
 preparation of 384-well plates for sorting 1:187
 sorting 1:187
 new tool for viral discovery 1:184
 sample collection and preservation 1:186
 viral staining 1:186–187
 whole-genome amplification 1:187–188
 DNA decontamination 1:188
 reagent preparation 1:188
 single virus tracking (SVT) 1:211
 SINV *see* Sindbis virus (SINV)
 SINV-1 4:771, 4:772, 4:772–773
 Siphoviridae 1:320, 4:186, 4:276
 mosaicism in 4:278–280
 siphovirus 4:342
 siRNA *see* small interfering RNA (siRNA)
 Siromilus 5:170–171
 SIRV, *see* *Sulfolobus islandicus* rod-shaped virus (SIRV)
 SISPA *see* Sequence Independent Single Primer Amplification (SISPA)
 Site Frequency Spectrum (SFS) 5:227
 SIV *see* simian immune deficiency virus (SIV)
 Sjogren's syndrome 2:454
 SJR *see* single jelly roll (SJR)
 skin, polyomaviruses of 2:526
 skin tumors in walleye and their associated retroviruses 2:317–318
 skin virome 1:555
 S-layer, defined 4:387
 S-Layer protein (archaeal), defined 4:380
 SLC7A1 *see* solute carrier family 7 member 1 (SLC7A1)
 SLCCNV *see* Squash leaf curl China virus (SLCCNV)
 sleeping disease (SD) 4:772
 SLEV *see* Saint Louis encephalitis virus (SLEV)
 SLIPTA *see* Stepwise Laboratory Improvement Process Towards Accreditation (SLIPTA)
 SLMTA *see* Strengthening Laboratory Management Toward Accreditation (SLMTA)
 SMAC *see* supramolecular adhesion complex (SMAC)
 small angle X-ray scattering (SAXS) 1:191
 as a structural tool 1:192–194
 time-resolved 1:194–195
 smallest capsid protein (SCP) 2:442–443
 small hydrophobic (SH) protein 2:749
 small icosahedral ssDNA viruses 1:268–269
 small icosahedral viruses, structures of 1:278
 basic structure 1:278
 diversity 1:278–279
 picornaviruses 1:279–282, 1:280f
 capsid assembly 1:280–282
 genome encapsidation 1:282–283
 host interactions 1:283–285
 uncoating 1:285–287
 antigenicity and antibody interactions 1:287–288
 antiviral drugs 1:286–287
 small interfering RNA (siRNA) 2:743–744, 1:498, 3:43, 3:52, 3:123, 2:242, 3:788, 3:116, 3:293, 3:456
 small interfering RNAs direct antiviral immunity 3:117–118
 small ruminant lentiviruses (SRLV) 2:56–57, 2:65, 2:66–67
 small soluble glycoprotein (ssGP) 2:234–235, 2:238
 smFRET *see* single molecule Förster resonance energy transfer (smFRET)
 smokybrown cockroach 4:839
 smolts, defined 2:544
 SMYEV *see* Strawberry mild yellow edge virus (SMYEV)
 snakehead retrovirus 2:319–320
 SNALPs *see* stable nucleic acid-lipid particles (SNALPs)
 SNARE proteins 1:534
 snow goose hepatitis B virus (SGHBV) 2:100
 SNP *see* single nucleotide polymorphism (SNP)
 sNSRV polymerases 1:350
 SNT *see* serum neutralization tests (SNT)
 society, impact of viruses on 1:677–678
 art and literature, viruses in 1:679
 human disease 1:677–678
 impact of viruses on livestock and crops 1:678–679
 impact on human capital in the era of COVID-19 1:678
 origins and historical demographics of viral infections in humans 1:677–678
 viral epidemics in the past century 1:678
 sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) 1:172, 5:94
 sodium iodide symporter (NIS) 5:241
Sogatella furcifera 4:771
 soil-borne vector transmission 3:531
Solemoviridae 3:675, 3:676, 3:712
 solemoviruses (*Solemoviridae*) 3:712
 genome 3:714–716
 host range and transmission 3:717–718
 pathological and epidemiological aspects 3:718
 phylogenetic relationships 3:716–717

- solemoviruses (*Solemoviridae*) (continued)
 replication and movement 3:717
 virion structure 3:712–714
Solenopsis geminata 4:771
Solenopsis invicta 4:771, 4:772–773
solenopsis invicta densovirus (SiDV) 4:840
Solenopsis invicta virus 1 4:771
 solid organ transplantation (SOT) 5:105
 solid organ transplant recipients 2:455
 solid-phase microwell immunoassays 5:17–18
 SOLiD sequencing 1:177
Soliniviridae 4:716–717, 4:892
 classification 4:892
 diagnosis 4:895–896
 epidemiology 4:894–895
 genome 4:893–894
 life cycle 4:894
 pathogenesis 4:895
 prevention 4:896
 virion structure 4:892–893
 solinviviruses 4:717
 soluble glycoprotein (sGP) 2:234–235
 solute carrier family 7 member 1 (SLC7A1) 2:145
 solution-state small angle X-Ray scattering 1:191–194
 diffracted X-ray tracking (DXT) 1:195–197
 SAXS as a structural tool 1:192–194
 time-resolved SAXS 1:194–195
 X-ray footprinting mass spectrometry 1:197
 somatic hypermutation (SHM) 1:591
 somatic incompatibility, defined 4:568
 SOPs *see* standard operating procedures (SOPs)
 SOR *see* Standard Operation Record (SOR)
 SOT *see* solid organ transplantation (SOT)
 Southern Asia, CLCuD in 3:356
 southern bean mosaic virus (SBMV) 1:259–260, 1:262–263, 1:278
 SPA *see* single particle analysis (SPA)
 spanin 1:505
 spatial structure, defined 4:314
 species 1:51
 assignment of viruses to 1:47–48
 criteria used in species definitions 1:48–49
 development of species level in virus taxonomy 1:47–48
 genomics-based taxonomy of viruses, emergence of 1:50–51
 metagenomically characterized viruses, classification of 1:50
 species and genotypes 1:49–50
 species assignments and genetic relationships 1:48
 in biology 1:47
 defined 1:28
 nomenclature 1:51
 species-jump, defined 2:730, 1:559
 specificity 5:73–74
 analytical 2:697
 diagnostic 2:697
 specimen management 5:66
 spherical lattices 1:297
 triangulation and derivation of 1:299f
 spherical viruses 4:362
 atomic structure of 1:476
 family *Globuloviridae* 4:362
 family *Ovaliviridae* 4:362
 family *Portogloboviridae* 4:362–363
 family *Turriviridae* 4:363
 MetSV 4:377–378
 spheroid, defined 4:858
 spheroplasts, defined 4:594, 4:534
 spherules, defined 2:173, 3:252
 SPI *see* Signal Peptidase I (SPI)
 SPIA *see* Single Primer Isothermal Amplification (SPIA)
 SPII *see* Signal Peptidase II (SPII)
 spike 1:257
 defined 4:380
 spike protein 4:385–386
 spillover, defined 2:117
Spinareovirinae 1:303–306, 4:868–870
 antigenic and genetic relationships 4:876–877
 cypoviruses 4:877
 dinoviruses 4:877
 idnoveoviruses 4:877
 aquareovirus 1:306
 cypovirus 1:306
 historical overview 4:870
 host range, diseases, transmission, and distribution 4:870–871
 aquareovirus 4:871
 cypoviruses 4:870–871
 dinovirus 4:871
 fijivirus 4:871
 idnoveoviruses 4:871
 orthoreovirus 4:871
 orthoreovirus 1:305–306
 virion properties, genome, and replication 4:871–875
 cypoviruses 4:871–875
 dinovirus 4:876
 idnoveoviruses 4:875–876
Spinaviridae 4:713–715
 spindles, defined 4:858
 spindle-shaped salterprovirus His1 1:433–434
 spindle-shaped viruses 1:369–370, 4:376–377
 spine helix 1:263
Spiraviridae 4:362
 Spirea yellow leaf spot virus (SYLSV) 3:165
 spleen focus-forming virus (SFFV) 2:322
 split sampling 5:74
 SPMV *see* sweet potato mild mottle virus (SPMMV)
 SPO1-like viruses, DNA replication of 4:63–65
Spodoptera ornithogalli 4:772
 spore balls, defined 3:603
 SPP1 connector 4:308
 SPP1-like viruses, DNA replication of 4:65–66
 SPPV *see* Sheeppox virus (SPPV); sweet potato pakakuy virus (SPPV)
 spraing, defined 3:743, 3:603
 spring viraemia of carp virus (SVCV) 2:324–325, 2:328
 S protein 2:197
 sputnik virus 1:379–380
 SPV1 *see* sulfolobus polyhedral virus 1 (SPV1)
 SpyTag/SpyCatcher system 1:667
 SQPV *see* squirrel poxvirus (SQPV)
 squamata, defined 2:3
 squamous cell carcinoma (SCC) 2:84–85, 2:86, 2:87, 2:86–87
 cats 2:87
 diagnosis 2:90
 dogs 2:87
 horses 2:87
 rabbits 2:87
 ruminants 2:86–87
 treatment 2:90–91
 Squash leaf curl China virus (SLCCNV) 3:756
 squash leaf curl virus 3:14
 causal agent and classification 3:14
 control 3:14
 disease symptoms and yield losses 3:14
 epidemiology 3:14
 geographical distribution 3:14
 squash vein yellowing virus (SqVYV) 3:296–297
 squirrel poxvirus (SQPV) 2:666
 SqVYV *see* squash vein yellowing virus (SqVYV)
 SRB1 *see* scavenger receptor class B member 1 (SRB1)
 Src, defined 2:122, 2:875
 SRID *see* single radial immunodiffusion (SRID)
 SRLV *see* small ruminant lentiviruses (SRLV)
 SRS *see* short read sequencing (SRS)
 SsaDV *see* sea star-associated densovirus (SsaDV)
 SSB *see* single-stranded DNA binding protein (SSB)
 SsBV1, *see* *Sclerotinia sclerotiorum* botrybriovirus 1 (SsBV1)
 ssDNA mycoviruses *see* single-stranded DNA (ssDNA) mycoviruses
 ssDNA viruses *see* single-stranded DNA (ssDNA) viruses
 SsDRV, *see* *Sclerotinia sclerotiorum* debilitation-associated RNA virus (SsDRV)
 S segment 2:208–209
 ssGP *see* small soluble glycoprotein (ssGP)
 SsHADV-1, *see* *Sclerotinia sclerotiorum* hypovirulence-associated DNA virus 1 (SsHADV-1)
 SsHV1, *see* *Sclerotinia sclerotiorum* hypovirus 1 (SsHV1)
 SSITL, *see* *Sclerotinia sclerotiorum* integrin-like gene (SSITL)
 SsMBV1, *see* *Sclerotinia sclerotiorum* megabirnavirus 1 (SsMBV1)
 SsMYRV4, *see* *Sclerotinia sclerotiorum* mycoreovirus 4 (SsMYRV4)
 SsNSRV-1, *see* *Sclerotinia sclerotiorum* negative-sense RNA virus 1 (SsNSRV-1)
 SSP *see* stable signal peptide (SSP)
 SSPE *see* subacute sclerosing panencephalitis (SSPE)

- ssRNA bacterial viruses *see* single-stranded RNA (ssRNA) bacterial viruses
- ssRNA mycoviruses *see* single-stranded RNA (ssRNA) mycoviruses
- SSSV *see* salmon swimbladder sarcoma virus (SSSV)
- SSTV *see* Swedish sea trout virus (SSTV)
- SSV, *see* *Sulfolobus* spindle-shaped viruses (SSV)
- SSV1 *see* *sulfolobus* spindle-shaped virus 1 (SSV1)
- SSWD *see* sea-star wasting disease (SSWD)
- St. Louis encephalitis virus 2:809–810
- stable nucleic acid-lipid particles (SNALPs) 2:616
- stable signal peptide (SSP) 2:507
- standard operating procedures (SOPs) 5:56–57, 5:65
- Standard Operation Record (SOR) 5:65
- standing variation, defined 1:71
- Staphylococcus aureus* pathogenicity island (SaPIs) 4:98
- Staphylococcus aureus* pathogenicity island 1 4:98–99
- staple crops: cassava brown steak disease 3:17
- causal agent and classification 3:17
- control 3:18–19
- disease symptoms and yield losses 3:17–18
- epidemiology 3:18
- geographical distribution 3:17
- start-feeding, defined 2:544
- steckling, defined 3:200
- STED *see* stimulation emission depletion (STED)
- Stepwise Laboratory Improvement Process Towards Accreditation (SLIPTA) 5:60
- sternal recumbency, defined 2:875
- sterol regulatory element-binding protein (SREBP) 2:391
- STHBV *see* stork hepatitis B virus (STHBV)
- stimulation emission depletion (STED) 1:497
- STIs *see* sexually transmitted infections (STIs)
- STIV *see* *sulfolobus* turreted icosahedral virus (STIV)
- STMV *see* satellite tobacco mosaic virus (STMV)
- STNV *see* satellite tobacco necrosis virus (STNV)
- stochastic optical reconstruction (STORM) 1:497
- stokes shift, defined 1:208
- stop-and-start model, defined 3:833
- stork hepatitis B virus (STHBV) 2:100
- STORM *see* stochastic optical reconstruction (STORM)
- strabismus, defined 2:34
- strain, defined 3:184, 3:8
- strain sunf-M 4:464–465
- strand-coupled genome replication, defined 4:387
- strand-displacement, defined 4:387
- Strand Displacement Amplification (SDA) 5:75–76
- stratum spinosum, defined 2:629
- Strawberry crinkle virus 3:628–629
- Strawberry mild yellow edge virus (SMYEV) 3:623, 3:628, 3:628–629
- Strawberry mottle virus 3:628–629
- Strawberry pallidosis-associated virus 3:628–629
- Strawberry polerovirus 1 3:628–629
- Strengthening Laboratory Management Toward Accreditation (SLMTA) 5:60
- Streptococcus thermophilus* 1:633, 1:634f
- structural comparison of tailed phages and herpesviruses 1:322–323
- major capsid proteins (MCPs) fold 1:322–323
- portal protein 1:323–324
- structural genomics, defined 4:359
- Structural Homology Program (SHP) 1:94
- structural lineage, defined 1:329
- structural proteins 1:257
- structure-based virus classification 1:90–92
- inferring viral evolutionary relationship through 1:92–94
- structure of virus, principles of 1:257
- capsid architecture, general principles of 1:257–261
- helical symmetry 1:261
- icosahedral symmetry 1:258–261
- capsid assembly 1:263–266
- enveloped icosahedral RNA viruses 1:266–268
- large icosahedral dsDNA viruses 1:269–271
- non-enveloped icosahedral RNA viruses 1:265–266
- small icosahedral ssDNA viruses 1:268–269
- tailed bacteriophages 1:271–272
- capsid proteins, structural folds of 1:262–263
- four-helix bundle 1:263
- HK97 fold 1:263
- immunoglobulin-like fold 1:263
- jelly-roll -barrel 1:262–263
- serine protease fold 1:263
- methods for studying virus structures 1:261–262
- structure of virus in 3D 1:254
- STSV2 *see* *sulfolobus* tengchongensis spindle-shaped virus 2 (STSV2)
- SU *see* surface proteins (SU)
- subacute sclerosing panencephalitis (SSPE) 5:268, 1:562–563, 2:619
- subclinical infection, defined 2:707
- sub-genomic RNA (sgRNA) 1:450–452, 3:260, 2:92, 4:577, 3:252, 3:327, 3:285, 3:439, 3:818, 4:699
- sub-Saharan Africa 5:142–143
- substitution rate, defined 1:62
- subviral RNAs 3:785–786
- sucrose method 5:11
- sugarcane viral diseases 3:94–95
- sugarcane yellow leaf virus 3:599
- Sulfolobus acidocaldarius* 4:421
- sulfolobus* ellipsoid virus 1 (SEV1) 4:362, 1:433–434
- Sulfolobus* filamentous virus 1 (SFV1) 1:368
- Sulfolobus islandicus* 4:421
- Sulfolobus islandicus* rod-shaped virus (SIRV) 4:419, 4:422
- SIRV1 1:434–435
- SIRV2 1:367, 1:367–368, 1:368f, 4:388, 4:363
- SIRV3 4:397
- Sulfolobus* polyhedral virus 1 (SPV1) 4:363
- Sulfolobus* spindle-shaped viruses (SSV) 4:420, 4:422, 4:424
- SSV1 4:388, 4:362, 4:411
- Sulfolobus* tengchongensis spindle-shaped virus 2 (STSV2) 4:388
- Sulfolobus* turreted icosahedral virus (STIV) 4:388, 4:388–389, 4:363, 1:204, 4:419–420
- STIV1 4:411
- sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate (Sulfo-SMCC) 1:667
- support quality assurance, controls to 5:75–76
- suppressors of gene silencing, defined 3:116
- supraglacial (top) biotopes 4:352–353
- supramolecular adhesion complex (SMAC) 1:601
- surface layer (s-layer), defined 1:402
- surface proteins (SU) 2:57
- surveillance, essential characteristics of 5:247–248
- surveillance of outbreaks 5:249
- surveillance system
- attributes of 5:251–252
- accuracy 5:252
- completeness 5:251–252
- consistency 5:252
- representativeness 5:252
- timeliness 5:252
- newer types of 5:249–250
- active surveillance 5:250
- Enhanced Surveillance 5:250
- sentinel surveillance 5:250
- surveillance of HIV/AIDS 5:250–251
- syndromic surveillance 5:250
- sustained virological response (SVR) 2:386
- SV40 *see* simian 40 virus (SV40)
- S value/Svedberg, defined 4:10
- SVCV *see* spring viraemia of carp virus (SVCV)
- Svedberg unit 1:169
- defined 1:162
- SVR *see* sustained virological response (SVR)
- SVT *see* single virus tracking (SVT)
- Swedish sea trout virus (SSTV) 2:325
- sweet potato mild mottle virus (SPMMV) 3:296–297
- sweet potato pakakuy virus (SPPV) 3:163
- sweet potato viral diseases 3:86
- SYBR Green, defined 2:757
- SYLSV *see* Spirea yellow leaf spot virus (SYLSV)
- sylvatic, defined 2:899
- sylvatic cycle, defined 2:22
- sylvatic cycle of CHIKV transmission, defined 2:173

sylvatic rabies, defined 2:738
 symbiosis, defined 4:419
 symptomatology, defined 3:247
 syncytia, defined 2:619
 syncytium, defined 2:475
 syndromic surveillance 5:250
 synergism/synergistic disease, defined 3:456
 synergistic phage-bacterium relationships
 1:633–634, 1:634–638
 synergy, defined 3:703
 synonymous, defined 1:62
 synovial membranes, defined 2:875
 synovitis 2:173
 synteny 4:278
 defined 4:359, 4:276
 Synthetic Fluorescent Dyes, defined 1:208
 systemic infection, defined 3:743, 2:797
 systemic lupus erythematosus 2:454

T

T–3 Icosahedron, defined 2:92
 T4 (bacteriophage) 4:260*f*, 4:261*f*, 4:70–72,
 1:319
 architecture of 4:259
 structure of 1:272*f*
 as a vaccine platform 4:259–261
 in vivo VLP assembly 4:259–261
 T5 genome 4:214
 T7 (bacteriophage) 4:72, 4:73*f*
 connector 4:308
 early transcription 4:72
 elongation 4:74
 gene expression 4:213
 phage RNAP and late gene transcription
 4:72
 termination 4:74
 transcription cycle 4:72–74
 TAAs *see* tumor-associated antigens (TAAs)
 TABS, *see* trans-activator binding site (TABS)
 TABV *see* Tamana bat virus (TABV)
 Tacaribe virus (TCRV) 2:507
 TAdV–3 *see* Turkey adenovirus 3 (TAdV–3)
 TAF *see* tenofovir alafenamide (TAF)
 tail, defined 1:318, 4:175, 4:276, 4:45
 tail assembly 4:49
 tailed bacteriophages 1:271–272
 tailed bacteriophages 4:194, 4:195*f*
 tailed dsDNA phages 4:45
 bacteriophage structure and function 4:46
 capsids 4:46
 tails and tail fibers 4:46–48
 gene expression, host interactions and
 regulation of 4:48
 classes of genes 4:48
 early genes 4:48–49
 late genes 4:49
 genetic mosaicism in 4:277
 clusters and superclusters 4:280–281
 complete bacteriophage sequences,
 comparative genomics of 4:278
 different phage clusters acquiring new
 genes at different rates 4:281
 DNA sequencing 4:277–278

evidence for genome mosaicism before
 advent of DNA sequencing 4:277
 mosaicism, hybrids, and single gene
 surveys 4:281–282
 mosaicism in the *Myoviridae* 4:280
 mosaicism in the *Podoviridae* 4:280
 mosaicism in the *Siphoviridae* 4:278–
 280
 genomes and genomics 4:50
 chromosome diversity and replication
 4:50
 common ancestry 4:51
 common themes in genome structure
 4:50–51
 diversity in genome size and
 organization 4:50
 horizontal exchange of genes is
 widespread 4:51
 injection 4:48
 DNA Injection 4:48
 protein injection 4:48
 lysis 4:50
 structure of 4:45–46
 temperate versus virulent bacteriophages
 4:48
 virion assembly 4:49
 assembly pathways 4:49
 capsid assembly 4:49–50
 tail assembly 4:49
 tailed phage capsids 4:220–221
 low copy number E proteins with specific
 locales in the capsid 4:221
 podoviral E proteins 4:221–222
 tailed phages 1:318–319
 icosahedral head/capsid 1:319
 infective viral cycle, structural insights into
 1:324–325
 final virion assembly 1:327
 genome ejection 1:324–325
 genome packaging 1:326–327
 genome replication 1:325
 procapsid morphogenesis 1:325–326
 Myoviridae 1:320–321
 Podoviridae 1:320
 Siphoviridae 1:320
 tailed viruses of methanogenic euryarchaea
 4:373–374
 tail fiber
 defined 4:194, 4:276, 4:45
 structures of 4:197*f*
 tail-less phages 4:207–208
 tail protein complex, defined 4:105
 tailspike, defined 4:194
 tail structure and dynamics 4:186–187
 Caudovirales 4:186–187
 long contractile tails, baseplates of 4:190–
 191
 long non-contractile tails, tail tip complex
 of 4:191
 organization and assembly of long tails
 4:187–189
 sheath 4:189–190
 short non-contractile tails 4:191–192
 tailspikes, tail fibers and phage-host
 interaction 4:192–193
 tube 4:189

tail tip protein (TTP) pb9 4:191
 Taiwanese Bat Lyssavirus (TBLV) 2:738
 Tamana bat virus (TABV) 1:546
 tandem repeats, defined 1:71
 tape measure, defined 4:276
 tape measure protein (TMP) 1:320
 targeted therapy, defined 1:658
 targeting, defined 5:233
 target recognition domain (TRD) 1:608
 TAS *see* TNF activation site (TAS)
 TAsTV2 *see* Turkey astrovirus type-2 (TAsTV2)
 TA system *see* toxin-antitoxin (TA) system
 TATA binding protein 2:445
 TATA box 2:58–59
tat gene 2:61*t*
 Taura syndrome virus (TSV) 4:772, 4:774
 taxon, defined 1:28
 taxonomic richness 4:677
 taxonomy, defined 1:28
 taxonomy, virus
 classification, virus 1:33
 taxa and viruses, differentiating 1:33
 future developments 1:35
 history of 1:28–29
 baltimore classification (1971) 1:29
 current virus taxonomy (1971–present)
 1:29–32
 early taxonomic developments (1886–
 1971) 1:28–29
 International Committee on Taxonomy of
 Viruses (ICTV) 1:32
 organization 1:32
 taxonomic process 1:32–33
 nomenclature of viruses 1:34–35
 virus taxa 1:33–34
 tax proteins 2:533–535
 TBEV *see* tick-borne encephalitis virus
 (TBEV)
 TBEV E 1:421
 TBLV *see* Taiwanese Bat Lyssavirus (TBLV)
 TBSV *see* tobacco bushy stunt virus (TBSV);
 tomato bushy stunt virus (TBSV)
 TBTv *see* tobacco bushy top virus (TBTv)
 TBV *see* telbivudine (TBV)
 T cell-dependent (TD) B cell 1:591
 T cell effectors in the response to viral
 infections 1:587
 CD4⁺ T Cells 1:587
 CD8⁺ cytotoxic T cells (CTL) 1:589
 cytotoxic CD4⁺ T cells 1:588–589
 regulatory T cells (Tregs) 1:588
 T follicular helper (Tfh) cells 1:587–588
 T helper 1 (Th1) cells 1:587
 T cell help 1:589–590
 T-cell immunoglobulin and mucin domain
 1 (TIM-1) 1:410
 T cell immunoglobulin and mucin domain
 3 (TIM-3) 1:660
 T cell priming 1:589
 T cell receptors (TCRs) 1:585–587, 1:585,
 1:587
 T cell recognition of viral antigens 1:587
 TCRs *see* T cell receptors (TCRs)
 TCRV *see* Tacaribe virus (TCRV)
 TCV *see* turnip crinkle virus (TCV)
 TDF *see* tenofovir disoproxil fumarate (TDF)

- T-DNA vector, defined 3:132
 TE *see* transposable elements (TE)
 TEC *see* transcriptional enzyme complex (TEC)
 tectiviruses 4:36
 TED *see* translation enhancer domain (TED)
 tegument 1:321
 defined 2:599, 1:318, 2:442
 tegument proteins 2:452
 teichoic acid, defined 4:194
 telaprevir 5:123
 telbivudine (TBV) 5:218
 teleomorph, defined 4:450
 telomeres 2:778
 TEM *see* transmission electron microscopy (TEM)
 TEM8 *see* tumor endothelial marker 8 (TEM8)
 temperate, defined 1:621, 4:36, 4:276
 temperate bacteriophage, defined 4:77
 temperate phage
 decision making by 4:88
 counting by infecting phages 4:90–93
 decision to remain dormant 4:94–97
 post-infection decision of phage lambda 4:88–90
 single cell, view from 4:93–94
 defined 4:88, 4:98
 temperate versus virulent bacteriophages 4:48
 temperate virus, defined 1:162, 4:368
 template strand, defined 4:69
 temporal lobe 2:778
 temporal lobe epilepsy (TLE) 2:785, 2:778
 tenofovir alafenamide (TAF) 5:133, 5:134
 tenofovir disoproxil fumarate (TDF) 5:218, 5:133, 5:134
 tenuiviruses (*Phenuiviridae*) 3:719–720
 biological properties 3:724–725
 host range and geographic distribution 3:725
 classification 3:720
 components 3:720–723
 cytopathology 3:725
 disease symptoms and control 3:725–726
 genome organization 3:723
 insect vector and type of transmission 3:726
 physical and physicochemical properties 3:720
 relation to other taxa 3:726
 replication and transcription 3:723–724
 taxonomy 3:720
 virion structure 3:720
 morphology 3:720
 tepovirus, defined 3:805
 TER *see* transepithelial electrical resistance (TER)
 TerL DNA packaging domain 4:127
 TerL subunits 4:125–126
 terminally redundant DNA, defined 4:77
 terminal protein (TP) 4:761–762, 2:8, 4:62–63, 1:433–434
 Terminal Protein Region 1 (TPR1) and 2 (TPR2) 1:433–434
 terminal redundancy, defined 4:229
 terminase, defined 4:277, 4:160, 4:105
 terminase holoenzymes 4:126, 4:124
 terminase motors 4:152–153
 terminase protomer, defined 4:124
 termination upstream ribosome-binding site (TURBS) 2:875
 terrestrial dust 4:355
 terrestrial hot springs 4:345–347
 TerS subunits 4:125
 tetherin 2:65
 tetra virus capsid
 capsid maturation, auto-proteolysis and dynamics 4:901–902
 capsid structure 4:900–901
 tetra viruses 4:717
 tetra viruses of insects 4:897
 genome organization 4:897
 Alphatetraviridae 4:897
 Carmotetraviridae 4:898
 Permutotetraviridae 4:897–898
 pathology 4:902
 host range 4:902
 persistent infections 4:902–903
 symptoms and transmission 4:902
 taxonomic classification 4:898
 tetra virus capsid
 capsid maturation, auto-proteolysis and dynamics 4:901–902
 capsid structure 4:900–901
 viral replication 4:898–900
 TEV *see* Tobacco Etch Virus (TEV)
 T-even capsids 4:215, 4:216
 T follicular helper (Th) cells 1:587–588
 TfR *see* transferrin receptor (TfR)
 TGB proteins *see* Triple gene block (TGB) proteins
 TGEV *see* transmissible gastroenteritis virus (TGEV)
 TGMV *see* tomato golden mosaic virus (TGMV)
 TGN, *see* trans-Golgi network (TGN)
 TGS *see* transcriptional gene silencing (TGS)
 THC *see* total hemocyte count (THC)
 Theiler's murine encephalomyelitis virus (TMEV) 1:476f, 1:447
 T helper 1 (Th1) cells 1:587
Theobroma cacao 3:165
 thermal environments, phages in 4:345–347
 deep-sea hydrothermal vents 4:347–348
 hot deserts 4:348–349
 terrestrial hot springs 4:345–347
 thermophilic, defined 4:359, 4:368
 thermoproteus spherical piliferous virus 1 (TSPV1) 4:362
Thermoproteus tenax virus 1 (TTV1) 4:419
Thermus thermophilus 1:320
 replication, defined 4:61
 thiomersal 5:302
 third generation sequencing technologies 5:31
 30K superfamily, defined 3:229
 thoracic fluid, defined 2:875
Thosea asigna virus (TaV) 4:897–898
 3' cap-independent translation enhancers (3' CITEs) 1:449
 3C-like chymotrypsin-like protease (3CL^{pro}) domain 4:806
 3C-like protease, defined 3:703
 3Dpol 4:771
 threshold detection nucleic acids 2:455
 thymidine kinase gene 5:185
 tick-borne encephalitis virus (TBEV) 1:290
 classification 2:843
 clinical features 2:846–847
 diagnosis 2:847–848
 ecological cycle 2:847f
 epidemiology 2:844–846
 genome strategy of 2:846f
 life cycle 2:843
 pathogenesis 2:847
 phylogenetic tree based on the complete coding regions of 2:844f
 prevention 2:848
 structure of TBEV virion 2:845f
 treatment 2:848
 viral proteins 2:843–844
 virion structure 2:843
 ticks 1:549
 tilling, defined 3:69
 timeliness 5:247, 5:252
 time-resolved SAXS 1:194–195
 time series analysis 1:565
Tipula paludosa baculovirus (TpBV) 4:828–829
Tipula paludosa nucleopolyhedrovirus (TpNPV) 4:828–829
 TIRF microscopy *see* total internal reflection fluorescence (TIRF) microscopy
 tissue Culture, defined 3:430
 tissue tropism, defined 2:518
 TLE *see* temporal lobe epilepsy (TLE)
 TLRs *see* toll-like receptors (TLRs)
 TL viral gene *see* true late (TL) viral gene
 TM *see* transmembrane (TM)
 TMA *see* transcription-mediated amplification (TMA)
 TMEV *see* Theiler's murine encephalomyelitis virus (TMEV)
 TMP *see* tape measure protein (TMP)
 TMV *see* tobacco mosaic virus (TMV)
tmx gene 2:61t
 TNF activation site (TAS) 2:65
 T number, defined 4:45
 tobacco, potato virus Y (PVY) in 3:621–622
 tobacco bushy stunt virus (TBSV) 1:259–260
 tobacco bushy top virus (TBTv) 3:795
Tobacco curly shoot virus 3:750
 Tobacco Etch Virus (TEV) 1:449
 tobacco mosaic virus (TMV) 1:5, 1:363f, 5:6, 1:475–476
 antigenicity of 3:730–731
 beginnings of virology 3:727–728
 bionanotechnology applications of 3:732–733
 biotechnology applications of 3:732
 physical and chemical properties of 3:728–730
 replication and cell-to-cell movement of 3:732
 self-assembly of TMV Particles 3:730

- tobacco mosaic virus (TMV) (*continued*)
 structure of 1:261f
 taxonomy and classification 3:731–732
 virus disassembly 3:730
- tobacco ringspot virus (TRSV) 1:279
- tobamoviruses 3:9
 cis-acting sequences 3:740
 evolution 3:735–736
 genome organization 3:737–739
 history of tobamovirus research 3:734
 host range and symptomatology 3:740–741
 interactions between viral and host proteins 3:739–740
 satellite tobacco mosaic virus (STMV) 3:740
 taxonomy and classification 3:734–735
 transmission 3:741
 viral proteins 3:739
 virus replication 3:740
 virus structure and composition 3:736–737
- tobraviruses (*Virgaviridae*)
 diseases caused by tobraviruses 3:747–748
 genome structure and expression 3:744–745
 expression of virus genes 3:745
 recombination and sequence deletions in tobravirus RNAs 3:745–746
 RNA sequences 3:744–745
 M-type and NM-type infections 3:743–744
 taxonomy and characteristics 3:743
 tobraviruses as gene expression/silencing vectors 3:747
 viral proteins 3:746
 RNA1 3:746
 RNA2 3:746–747
 virus particle production and structure 3:743
- ToBRFV *see* tomato brown rugose fruit virus (ToBRFV)
- Togaviridae* 1:469, 1:499
- togavirus spherules 1:499
- toggle switch 4:94–96
- Tokiviricetes* 4:363–364
- ToLCA, *see* Tomato leaf curl alphasatellite (ToLCA)
- ToLCB *see* tomato leaf curl betasatellite (ToLCB)
- ToLCBanV, *see* Tomato leaf curl Bangalore virus (ToLCBanV)
- ToLCD *see* Tomato leaf curl disease (ToLCD)
- ToLCNDV, *see* Tomato leaf curl New Delhi virus (ToLCNDV)
- ToLCRnV, *see* Tomato leaf curl Ranchi virus (ToLCRnV)
- ToLCV, *see* Tomato leaf curl virus (ToLCV)
- Toleusatellitidae* 3:240t–241
- tolerance, defined 3:749
- toll-like receptors (TLRs) 1:601–602
 TLR3 pathway 2:391–392
 VLPs and the stimulation of 1:663–664
- tomato, geminivirus resistance in 3:558–559
 tomato yellow leaf curl virus (TYLCV)
 resistance 3:559
 breeding for 3:559
 genetic engineering for 3:560
 Ty genes, characterization of 3:559–560
- tomato, potato virus Y (PVY) in 3:620–621
- tomato brown rugose fruit virus (ToBRFV) 3:10–11
 causal agent and classification 3:11
 control 3:12
 disease symptoms and yield losses 3:11
 epidemiology 3:11–12
 geographical distribution 3:11
- tomato bushy stunt virus (TBSV) 1:257–258, 1:258–259, 1:262–263, 1:278, 3:790–791
- tomato golden mosaic virus (TGMV) 3:194
- Tomato leaf curl alphasatellite* (ToLCA) 3:360–361
- Tomato leaf curl Bangalore virus* (ToLCBanV) 3:357, 3:750
- tomato leaf curl betasatellite (ToLCB) 3:361
- Tomato leaf curl disease (ToLCD) 3:245
 and virus biodiversity on tomato in the Indian subcontinent 3:750–751
- Tomato leaf curl Gujarat virus* 3:750
- Tomato Leaf Curl New Delhi Disease 3:23–24
 component capturing and exchange of helper components 3:24
 control 3:24
 disease symptoms and yield losses 3:24
 epidemiology 3:24
 geographical distribution 3:23–24
- Tomato leaf curl New Delhi virus* (ToLCNDV) 3:358–359, 3:751
 diagnosis and management 3:759
 genome organization and function of gene products 3:751–753, 3:755f
 host range of 3:755–756
 maximum-likelihood dendrogram 3:752f
 phylogeny and molecular diversity 3:751
 synergism and genetic changes leading to begomovirus emergence and evolution 3:753–755
- ToLCNDV2 (JQ897969) 3:751
- ToLCNDV4 (KF551592) 3:751
- ToLCNDV5 (EF450316) 3:751
- transmission and epidemiology 3:756–757, 3:757f
- virus-host interactions 3:757–758
 basal level of defense 3:758
 counter defense RNA silencing 3:758
 resistance gene/R-gene mediated defense 3:758–759
 small non-coding RNAs 3:758
- Tomato leaf curl Pune virus* 3:750
- Tomato leaf curl Ranchi virus* (ToLCRnV) 3:756
- Tomato leaf curl virus* (ToLCV) 3:358–359, 3:245
- Tomato ringspot virus (ToRSV) 3:628–629
- tomato spotted wilt virus (*Tospoviridae*) 1:649, 3:12
 causal agent and classification 3:12
 classification 3:761
 control 3:13
 diagnosis 3:764
 disease symptoms and yield losses 3:12
 epidemiology 3:764, 3:12–13
 genome 3:762
 geographical distribution 3:12
 pathogenesis 3:764
 plant resistance 3:765–766
 prevention 3:764–765
 resistance breaking strains 3:13
 transmission 3:762–764
 virion structure 3:761–762
- tomato viral diseases 3:88–89
- tomato yellow leaf curl disease 3:21–22
 causal agent and taxonomy 3:22–23
 control 3:23
 disease symptoms, host range and yield losses 3:22
 epidemiology 3:23
 geographical distribution 3:21–22
- tomato yellow leaf curl viruses (TYLCVs) 3:768, 3:359
 bipartite 3:771
 breeding for resistance 3:775
 diagnosis 3:774
 immunodetection 3:774
 microscopy 3:774
 nucleic acid amplification 3:774
 nucleic acid hybridization 3:774
 sequencing 3:774–775
 symptom observation 3:774
- epidemiology 3:772–773
- genetic engineering 3:775–776
 CRISPR/Cas9 3:776
 expression of viral genes 3:775–776
 RNAi 3:776
- genome 3:768–770
- genome evolution 3:771
- geographical distribution and spread 3:772–773
- hosts 3:772
- life cycle 3:771–772
- monopartite 3:769–770
- pathogenesis 3:773–774
- prevention 3:775
 biological control of whiteflies 3:775
 chemical control of whiteflies 3:775
 cultures under cover 3:775
 planting in whitefly-free period, eradication of inoculum 3:775
- satellite DNAs associated with 3:771
- treatment 3:775
- TYLCV transmission to tomato and host response 3:776
 involvement of heat shock and quality control proteins 3:777
 response of plants to TYLCV infection 3:776–777
 transcriptome and metabolome 3:777
 transmission by whiteflies 3:776
- TYLCV-whitefly vector relationship 3:773
 effects of TYLCV on life expectancy and fertility of whitefly vector 3:773
 path of the virus in the insect 3:773
 replication of the virus in the insect vector 3:773
 retention of the virus in the insect vector 3:773
 virion structure 3:768

- tombusvirid, defined 3:788, 3:778, 3:285, 3:456
- Tombusviridae* 3:828–829, 3:778, 3:481, 3:286–287, 3:456
- Tombusviruses (*Tombusviridae*) 3:789t–790, 1:499
- classification 3:788
 - cytopathology 3:794
 - genome organization and replication strategy 3:792–794
 - life cycle 3:794
 - pathogenesis 3:795–796
 - subviral molecules 3:794–795
 - taxonomy, phylogeny and evolution 3:788–790
 - virion structure 3:790–792
- Tombusvirus-like viruses (*Tombusviridae*) 3:778
- applied aspects 3:786–787
 - gene expression 3:783
 - gene function 3:781–783
 - genome replication 3:783–785
 - genome structure 3:781
 - host range and geographic distribution 3:786
 - subviral RNAs 3:785–786
 - symptoms and transmission 3:786
 - taxonomy, phylogeny, and evolution 3:778–779
 - virion structure, assembly, and disassembly 3:779–781
- Tombusvirus polymerase molecules 1:497
- tomography 5:5
- tonoplast, defined 3:371
- topocuvirus 3:413
- tOPV *see* trivalent oral polio vaccine (tOPV)
- Toroviruses 2:256, 2:264
- torquetenomdivirus (TTMVD) 2:48
- torquetenominivirus (TTMV) 2:48
- Torque Teno Viruses (TTVs) 1:556, 5:106
- torradoviruses
- classification 3:322
 - clinical features and pathogenesis 3:325
 - diagnosis 3:325
 - genome organization 3:322
 - geographic distribution 3:324
 - life cycle and epidemiology 3:322–324
 - host range 3:324
 - prevention 3:326
 - transmission 3:324–325
 - treatment 3:325–326
 - virion structure 3:322
- ToRSV *see* Tomato ringspot virus (ToRSV)
- torticollis, defined 2:34
- toscana virus 2:771
- Tospoviridae* 4:718, 3:761, 3:509
- total hemocyte count (THC) 4:816
- total internal reflection fluorescence (TIRF) microscopy 1:497, 1:211
- Total Quality Management (TQM) 5:64
- Totiviridae* 4:648
- totivirids (*Totiviridae*) 4:648–649
- biological properties 4:654–655
 - defined 4:648
 - genome organization and expression 4:652–654
 - virus replication cycle 4:654
 - taxonomy and evolutionary relationships among 4:649
 - virion properties 4:649
 - virion structure and composition 4:649–652
 - virus–host relationships 4:655–656
- totiviruses 4:630–631, 4:513, 4:504–505
- defined 4:648
- Toursvirus* 4:724
- toxin-antitoxin (TA) system 4:397–398
- toxins-antitoxins 1:607
- TP *see* terminal protein (TP)
- TQM *see* Total Quality Management (TQM)
- tracheoblasts, defined 4:858
- tracrRNA *see* trans-activating crRNA (tracrRNA)
- traditional processing techniques, limitations of 5:11–12
- trans-acting siRNA, defined 3:52, 3:123
- trans-activating crRNA (tracrRNA) 4:245–246, 1:611–613
- trans-activator binding site (TABS) 3:790
- transactivator/viropilamin (P6) 3:318
- transcription, defined 3:778
- transcriptional elongation, defined 2:475
- transcriptional enzyme complex (TEC) 1:306
- transcriptional gene silencing (TGS) 3:361, 3:362–363, 3:561
- defined 3:116, 3:149, 3:554
- transcriptional interference, defined 4:77
- transcriptional regulatory sequence (TRS) 2:202–203, 2:198
- transcription-dependent CRISPR interference, defined 4:400
- transcription factor binding sites (TBS) 2:58–59
- transcription factor II H, defined 2:765
- transcription-mediated amplification (TMA) 5:98
- transcription-mediated genome internalization 4:214
- transcription regulating sequences (TRSs) 4:806, 4:699, 4:709, 2:195, 2:263–264
- transcytosis 1:529
- defined 2:789
 - endocytic mechanisms leading to 1:531–533
 - caveolin-dependent endocytosis 1:532–533
 - immunoglobulins as mediators of transcytosis 1:533–534
 - macropinocytosis 1:533
 - receptor-mediated, clathrin-dependent endocytosis 1:533
 - and viral pathogenesis 1:534–536
 - breaching epithelial barriers 1:535–536
 - fetal infections 1:537–538
 - neuroinvasion 1:536
 - viral hepatitis 1:536–537
- transduction, defined 4:98
- trans-encapsulation/hetero-encapsulation, defined 4:658
- transepithelial electrical resistance (TER) 1:531, 2:92
- transfection, defined 4:468
- transferrin receptor (TfR) 1:410, 1:268–269
- transfer RNAs (tRNAs) 1:377
- transformation, defined 2:528
- transfusion, defined 2:442
- transfusion transmitted (TT) virus 1:409, 2:48, 2:51, 2:51–52
- transgene, defined 3:313
- transgenic P12 tobacco, defined 3:132
- transgenic pathogen-derived resistance, defined 3:327
- transgenic plant, defined 3:52, 3:293
- trans-Golgi network (TGN) 1:498, 1:425
- translational 'recoding' 1:452
- ribosomal bypassing ('hopping') 1:452
 - ribosomal 'frame-shifting' 1:452–454
 - ribosome stop codon 'read-through' 1:454
 - 'stopgo'/'stop carry-on'/ribosomal 'skipping' 1:454
 - virus-encoded proteinases 1:454–455
- translational repression, defined 4:21
- translation enhancer domain (TED) 1:449
- translesional polymerases, defined 2:683
- translocating motor 4:131–132
- coordinated ATP hydrolysis by terminase motors 4:132
 - mechanochemical coupling and DNA translocation 4:132
 - structure of 4:131–132
- transmembrane (TM) 2:57
- transmissible gastroenteritis virus (TGEV) 2:194, 2:197, 2:256, 2:198
- classification 2:850
 - clinical features 2:850–852
 - diagnosis 2:852
 - epidemiology 2:850
 - genome 2:850
 - genome organization of 2:851f
 - life cycle 2:850
 - pathogenesis 2:852
 - prevention 2:852
 - treatment 2:852
 - vaccines for 2:852t
 - virion structure 2:850
- transmission, defined 1:559
- transmission barrier, defined 2:707
- transmission electron microscopy (TEM) 1:622, 5:5, 3:629, 1:495–496, 1:497
- transmission of virus, modes of 1:561–562
- transmission propensity, defined 3:862
- transovarial, defined 2:654
- transovarial transmission, defined 3:200, 3:545, 4:724
- transplacental, defined 2:442
- transplant, defined 2:442
- transportation and storage of specimens 5:66–67
- transposable elements (TE) 1:79–80
- defined 1:71
- transposon, defined 4:175
- transpoviron, defined 1:372
- Transwell assays 1:531
- traveling salesman problem, defined 1:248
- TRD *see* target recognition domain (TRD)
- trD *see* trimerization domain (trD)
- Treatment as Prevention (TasP) 2:473

- Tregs *see* regulatory T cells (Tregs)
- triangulation, defined 1:248
- triangulation number (T-number) 1:259–260, 1:250, 1:319, 4:342, 4:115, 4:36, 1:278, 1:318, 3:778, 4:21, 4:229
- triangulation number series (T-number series), defined 1:248
- triangulation numbers of icosahedra 1:222–223
- Triatoma infestans* 4:771
- triatoma virus (TrV) 4:768–770
- triatovirus 4:768, 4:769t, 4:770t, 4:774
- Trichomonas vaginalis* virus 1 (TVV1) 4:585
- Trichoplusia ni* 4:772
- Trichoplusia ni* virus (TnV) 4:902
- trichovirus, defined 3:805
- triggering, defined 1:417
- trihexagonal lattice series (t_h -number series), defined 1:248
- TRIM5 2:65, 2:827, 1:629–630
- trimerization domain (trD) 1:424
- tripartite genome 3:260, 3:260–261, 3:439
- tripartite RNA genome, defined 3:252
- Triple gene block (TGB) proteins 3:623, 3:623–624, 3:625–626
- triple gene block, defined 4:478, 3:528, 3:140, 3:642
- triradiate lumen, defined 3:486
- triskelion 1:292
- tritroviruses 3:797
- biotechnological application 3:803
 - diagnosis 3:802
 - disease management 3:802
 - cultural practice 3:802
 - genetic resistance 3:802–803
 - functions of viral proteins 3:799
 - genome organization 3:799
 - life cycle and epidemiology 3:799–800
 - disease cycle 3:801–802
 - host range 3:800
 - seed transmission 3:800–801
 - vector transmission 3:800
 - pathogenicity 3:802
 - taxonomy and classification 3:797
 - virion structure 3:797–799
- trivalent oral polio vaccine (tOPV) 5:312–313
- triviruses (*Betaflexiviridae*) 3:805
- cellular localization of trivirus particles and viral replication 3:811–813
 - classification 3:805–810
 - diagnosis 3:815–816
 - genomic organization and properties of encoded proteins 3:810–811
 - host range and distribution of triviruses 3:813–814
 - phylogenetic relationships of the different triviruses 3:811
 - prevention and treatment 3:816
 - symptomology, impact and vectors 3:814–815
 - virion structure 3:810
- tRNA-like structure, defined 3:336
- tRNA primer 5:131–132
- tRNAs *see* transfer RNAs (tRNAs)
- trophallaxis, defined 4:892
- tropism, defined 1:382, 2:442
- TRS *see* transcriptional regulatory sequence (TRS)
- TRSs *see* transcription regulating sequences (TRSs)
- TRSV *see* tobacco ringspot virus (TRSV)
- true flies *see* diptera
- true late (TL) viral gene 2:452
- Trypanosoma cruzi* 4:772–773
- tsetse fly, GpSGHV transmission dynamics in 4:788–789
- TSPV1 *see* thermoproteus spherical piliferous virus 1 (TSPV1)
- TSV *see* Taura syndrome virus (TSV)
- TTMV *see* torquetenominivirus (TTMV)
- TTMVD *see* torquetenomidivirus (TTMVD)
- TTV1, *see* *Thermoproteus tenax* virus 1 (TTV1)
- TT virus *see* transfusion transmitted (TT) virus
- TTVs *see* Torque Teno Viruses (TTVs)
- Tulip virus X (TVX) 3:628
- tumor-associated antigens (TAAs) 1:659, 1:658
- tumor endothelial marker 8 (TEM8) 1:659
- tumorigenesis, defined 2:528
- tumor virology 2:416
- TURBS *see* termination upstream ribosome-binding site (TURBS)
- turgor 4:216
- Turkey adenovirus 3 (TAdV-3) 2:13
- Turkey astrovirus type-2 (TAsV2) 2:96
- turncurtovirus 3:413
- turnip crinkle virus (TCV) 3:790–791
- turnip yellow mosaic virus (TYMV) 1:258–259
- Turriviridae* 4:363
- TVV1, *see* *Trichomonas vaginalis* virus 1 (TVV1)
- TVX *see* Tulip virus X (TVX)
- 20th century state sponsored biological weapons programs 1:645–646
- 2A-like motif, defined 4:658
- TYLCV-IR 3:244
- TYLCVs *see* tomato yellow leaf curl viruses (TYLCVs)
- Tymoviridae* 3:818
- tymoviruses (*Tymoviridae*) 3:818–821
- bombxy mori latent virus 3:825
 - properties and distinguishing characteristics 3:825
 - capsid structure 3:821
 - family and its distinguishing features 3:818
 - genome organization 3:821–822
 - infection and transmission 3:822
 - maculaviruses 3:824
 - properties and distinguishing characteristics 3:824
 - marafiviruses 3:822–823
 - infection and transmission 3:824
 - properties and distinguishing characteristics 3:823
 - virion structure, genome organization, replication cycle, and phylogenetic relationships 3:823–824
 - phylogenetic relationships and species demarcation 3:822
- poinsettia mosaic virus 3:824–825
- properties and distinguishing characteristics 3:825
 - properties and distinguishing characteristics 3:821
 - replication cycle 3:822
- TYMV *see* turnip yellow mosaic virus (TYMV)
- Type I CRISPR-Cas systems 4:247
- Type I interferon, defined 2:428
- Type 1 wild poliovirus (WPV1) 5:310
- Type II CRISPR-Cas systems 4:247–248
- Type 2 wild poliovirus (WPV2) 5:310
- Type III CRISPR-Cas systems 4:248
- Type 3 wild poliovirus (WPV3) 5:310
- Type IV CRISPR-Cas systems 4:248
- Type 4 pili, defined 4:387
- type species, defined 4:632
- 5' tyrosyl-DNA phosphodiesterase 2 (TDP2) 2:336

U

- UK Biological weapons program 1:645–646
- UL28, defined 5:181
- UL36, defined 5:181
- ultracentrifugation 5:11
- ultrafiltration methods 1:167, 1:167f
- ultrastructural imaging 1:496–497
- Umbravirus* 3:788
- umbraviruses (*Calvusvirinae*, *Tombusviridae*) 3:827
- cell-to-cell movement function 3:830
 - classification 3:827
 - genome organization and expression 3:827–829
 - host range 3:831
 - interaction with assistor virus 3:831
 - involvement of nucleolus in umbravirus systemic infection 3:831
 - phloem-dependent long-distance movement function 3:830–831
 - prevention and control 3:831–832
 - replication 3:829
 - satellite RNA 3:829–830
 - similarity with other taxa 3:829
 - transmission 3:831
 - virion structure 3:827
- UN 3373, defined 5:181
- uncoating, defined 1:388, 1:382, 2:92
- unconventional bioweapons 1:644
- unfolded protein response, defined 2:428
- unit length versus headful packaging mechanisms 4:124–125
- universal prophylaxis 5:191
- unrooted tree 1:117–118
- untranslated regions (UTRs) 3:331
- defined 2:757, 4:21, 4:792
 - 3' untranslated region (3'UTR) 2:892, 2:892f
- upper respiratory tract infections (URTI) 5:199, 5:199–200
- upstream regulatory region (URR) 2:493
- urban cycle of CHIKV transmission, defined 2:173

uridylation, defined 2:757
 URR *see* upstream regulatory region (URR)
 URTI *see* upper respiratory tract infections (URTI)
 US28 signaling 2:457
 u-spanin, defined 1:501
Ustilago maydis viruses and their killer toxins
 atomic structure of KP6 4:516
 comparison of the killer proteins 4:517–518
 effect of KP4 on plants 4:515
 effects of KP4 on *U. maydis* cells 4:514–515
 evolutionary origin of KP4 4:516
 killer phenomena 4:513
 KP1 4:516–517
 KP4 blocks L-type voltage gated Ca^{2+} channels 4:515
 KP4 protein 4:513–514, 4:514f
 KP6 4:516, 4:517f
 possible application of KP4: fungal resistance in plants 4:515–516
 totiviruses 4:513
Ustilago maydis virus H1 (UmV-H1) 4:647
 UTRs *see* untranslated regions (UTRs)

V

VAC *see* viral assembly compartment (VAC)
 vaccination, induction of immunity to viruses by 1:592–595
 vaccine-associated paralytic poliomyelitis (VAPP) 5:311, 2:692
 vaccine-derived poliovirus (VDPV) 5:310
 vaccine design, biotechnology approaches to virus-like particle (VLP)-based vaccines for humans 1:668
 HPV and HBV 1:668
 influenza 1:668
 polio 1:668–669
 virus-like particles (VLPs) 1:662
 virus-like particles (VLPs), expression systems to produce 1:664
 bacteria 1:664
 cell-free protein synthesis (CFPS) 1:665
 insect cells 1:664
 mammalian cells 1:664
 plants and plant cell culture 1:664–665
 yeast 1:664
 virus-like particles (VLPs), immunological properties of 1:662
 cellular immune response, VLPs and induction of 1:662
 humoral immune response, VLPs and induction of 1:662–663
 toll-like receptors, VLPs and the stimulation of 1:663–664
 virus-like particles (VLPs) as cancer vaccines and therapeutics 1:669–670
 virus-like particles (VLPs) as nano-carriers 1:665–666
 conjugation of antigens and virus-like particles 1:665–666
 loading of virus-like particles by encapsulation 1:667–668
 vaccine effectiveness, defined 5:289, 5:295
 vaccine efficacy, defined 5:289, 5:295
 vaccine for coronaviruses (CoVs) 2:205–206
 vaccine immunogenicity, defined 5:295
 vaccine production, safety, and efficacy 5:281
 costs and duration of vaccine development 5:283–284
 emerging viral diseases, challenge of 5:285–286
 history of vaccination 5:281–282
 in human studies 5:283
 influenza vaccines, licensing of 5:285
 more about vaccine development 5:285
 phases and timelines in vaccine development: from discovery to product 5:282
 preclinical studies 5:282–283
 Shingrix, development of 5:282
 taking stock 5:285
 vaccine industry 5:285
 vaccine production 5:284–285
 vaccines 1:675–676, 5:205
 vaccines against viral gastroenteritis
 CDC disclaimer 5:293
 norovirus 5:291–292
 future considerations for norovirus vaccines 5:293
 norovirus vaccine development 5:292
 norovirus vaccines in human clinical trials 5:292–293
 norovirus vaccines in pre-clinical development 5:293
 rotavirus vaccines 5:289–290
 licensed rotavirus vaccines 5:290
 rotavirus vaccine safety monitoring 5:291
 rotavirus vaccines in development 5:291
 vaccine effectiveness and impact 5:290–291
 vaccines and antiviral agents 5:213
 hepatitis A virus 5:213
 immunoglobulin 5:213
 treatment 5:213
 vaccine 5:213
 hepatitis E virus 5:213–214
 management of acute hepatitis E 5:214
 management of chronic infections 5:214
 prophylaxis 5:214
 vaccine trials 1:566
 vaccinia immune globulin (VIG) 5:271–272
 vaccinia virus (VACV) 2:666, 2:171, 1:498, 1:461
 antiviral drugs 2:859
 classification 2:854
 genome 2:855
 host range 2:857–858
 immunomodulators 2:858
 life cycle 2:856f, 2:855
 assembly 2:857
 entry 2:855
 genome replication 2:856–857
 transcription and translation 2:855–856
 origin of 2:854
 smallpox vaccine 2:858–859
 vaccine vectors 2:859
 virion structure 2:854–855
 virulence 2:858
 VACV *see* vaccinia virus (VACV)
 valaciclovir 5:183–184
 valganciclovir 2:458–459
 vaniprevir 5:123
 VAP *see* virion-associated protein (VAP); virus-associated pyramids (VAP)
 VAPP *see* vaccine-associated paralytic poliomyelitis (VAPP)
 variable regions (VRs) 4:841
 varicella, defined 2:860
 varicella hyperimmune immunoglobulin 5:272
 varicella-zoster immunoglobulin (VZIG) 5:187–188, 5:272
 varicella zoster virus (VZV) 5:127–128, 5:109–110, 5:194, 1:321
 classification 2:860
 clinical features 2:863–864
 diagnosis 2:865
 DNA polymerase 5:128
 epidemiology 2:862–863
 genome 2:860–861
 life cycle 2:861–862
 pathogenesis 2:864–865
 prevention 2:866
 treatment 2:865–866
 virion structure 2:860
 varicella-zoster virus (VZV) infections 5:182f
 antiviral agents in clinical use 5:183
 acyclovir 5:183
 brivudine ((E)-5-(2-bromovinyl)-2'-deoxyuridine) 5:184
 famciclovir 5:184
 foscarnet (trisodium phosphonoformate) 5:184
 valaciclovir 5:183–184
 development of resistance 5:184–185
 DNA polymerase gene 5:185
 thymidine kinase gene 5:185
 laboratory diagnosis 5:181
 detection of antibodies 5:182–183
 direct detection of virus 5:181–182
 submission of samples 5:181
 prevention 5:186–187
 passive immunoprophylaxis of varicella 5:187–188
 prevention of varicella by vaccination 5:186–187
 prevention of zoster by vaccination 5:188–189
 resistance testing 5:185–186
 genotyping 5:186
 phenotyping 5:185–186
 varicosaviruses (*Rhabdoviridae*) 3:833–834
 genetics and evolution 3:834–835
 L protein encoded on RNA1 3:834–835
 proteins encoded on RNA2 3:835–836
 transcription termination/initiation strategy 3:836–837
 host range and geographic distribution 3:837
 taxonomy and classification 3:834
 transmission and vectors 3:837–838
 Varidnaviria 1:20

- variola virus (VARV) 2:666, 5:127
 classification and structural morphology 2:868–869
 clinical features 2:870–871
 diagnosis 2:872
 epidemiology 2:871
 history 2:868
 pathogenesis 2:869–870
 prevention 2:872–873
 treatment 2:873–874
 viral life cycle 2:869
 VP37 envelope wrapping protein 5:127
 VariZIG, defined 2:860
 VariZig administration 5:272
Varroa destructor 4:771
 VARV *see* variola virus (VARV)
 vascular endothelial growth factor (VEGF) 2:671
 vCBP *see* chemokine binding protein (vCBP)
 VCGs *see* vegetative compatibility groups (VCGs)
 VCMD *see* viral covert mortality disease (VCMD)
 vCyclin 2:602
 VDJ rearrangement 1:590–591
 VDPV *see* vaccine-derived poliovirus (VDPV)
 vector-borne transmission 1:562, 1:559
 vector capacity, defined 1:542
 vector competence, defined 1:542, 2:805
 vectored RSV vaccines 2:755
 VEEV *see* Venezuelan equine encephalitis virus (VEEV)
 vegetable viral diseases 3:86–87
 chilli viral diseases 3:86–87
 cucurbit viral diseases 3:87–88
 okra viral diseases 3:88
 tomato viral diseases 3:88–89
 vegetable viruses 3:9
 begomoviruses 3:13–14
 cucumber green mottle mosaic virus 3:9
 causal agent and classification 3:9
 control 3:10
 disease symptoms and yield losses 3:9–10
 epidemiology 3:10
 geographical distribution 3:9
 orthotospoviruses 3:12
 squash leaf curl virus 3:14
 causal agent and classification 3:14
 control 3:14
 disease symptoms and yield losses 3:14
 epidemiology 3:14
 geographical distribution 3:14
 tobamoviruses 3:9
 tomato brown rugose fruit virus (ToBRFV) 3:10–11
 causal agent and classification 3:11
 control 3:12
 disease symptoms and yield losses 3:11
 epidemiology 3:11–12
 geographical distribution 3:11
 tomato spotted wilt virus 3:12
 causal agent and classification 3:12
 control 3:13
 disease symptoms and yield losses 3:12
 epidemiology 3:12–13
 geographical distribution 3:12
 resistance breaking strains 3:13
 vegetative Compatibility, defined 4:457
 vegetative compatibility groups (VCGs) 4:520
 vegetative incompatibility 4:601, 4:431, 4:520, 4:589, 4:607, 4:493, 4:648, 4:552, 4:528, 4:468, 4:522–523
 vegetative incompatibility, in filamentous fungi 4:520
 genetics of 4:522–523
 HET protein involved in NLR-mediated innate immunity in fungi 4:523–525
 microscopic and macroscopic analyses of hyphal fusion 4:521
 mycoviruses, horizontal transmission of 4:525–526
 mycovirus transmission 4:525
 potential approaches to enhance mycovirus transmission between VCGs 4:526–527
 signaling pathways of 4:523
 vegetative compatibility groups (VCGs), identification of 4:521–522
 auxotrophic complementation 4:522
 barrage assay 4:521–522
 detecting the PCD of fused cells 4:522
 microscopy 4:522
 visualization of labeled proteins during co-culture 4:522
 VEGF *see* vascular endothelial growth factor (VEGF)
 VEGF-E 2:670–671
 VeHF *see* Venezuelan hemorrhagic fever (VeHF)
 velarivirus, defined 3:336
 velogenic viruses 2:650
 Venezuelan equine encephalitis virus (VEEV) 2:40, 2:43, 2:44–45, 2:45, 2:45–46, 2:46, 1:645–646, 1:536, 1:491, 1:546
 Venezuelan hemorrhagic fever (VeHF) 2:512
 vertebrates, prions of *see* prions of vertebrates
 vertex, defined 1:402
 vertical mycovirus transmission, defined 4:468
 vertical transmission 2:778, 2:218, 2:899, 4:419, 3:388, 3:430, 3:106, 4:780, 4:888, 4:768
 vertical/transplacental transmission 1:562, 1:559
Verticillium albo-atrum 4:523
Verticillium dahliae 4:523
 very low-density lipoprotein receptor (VLDL-R) 1:284
 very low-density lipoproteins (VLDL) 2:386–387
 vesicle, defined 2:860, 2:868
 vesicle-like archaeal viruses 4:380–381
 genomic characteristics 4:381–382
 gene content and conserved genes 4:382
 new isolates and haloarchaeal genomic regions 4:384
 Pleolipoviridae 4:382–384
 pleolipovirus infectivity, stability of 4:384
 salinity 4:384
 temperature and other significant factors 4:384–385
 structural proteins 4:385–386
 spike protein 4:385–386
 virion structure and viral life cycle 4:381
 vesicle-mediated transcytosis and export of viruses 1:529–530
 endocytic mechanisms leading to transcytosis 1:531–533
 caveolin-dependent endocytosis 1:532–533
 immunoglobulins 1:533–534
 macropinocytosis 1:533
 receptor-mediated, clathrin-dependent endocytosis 1:533
 endocytic sorting and vectorial transport of vesicles 1:534
 extracellular vesicle-mediated export of viruses 1:538
 autophagy-related vesicle-mediated release of virus 1:539
 exosome-like release of quasi-enveloped viruses 1:538–539
 nonlytic release of virus in extracellular vesicles 1:538
 in vitro models of viral transcytosis 1:531
 secretion of cargo 1:534
 transcytosis 1:530–531
 transcytosis and viral pathogenesis 1:534–536
 breaching epithelial barriers 1:535–536
 fetal infections 1:537–538
 neuroinvasion 1:536
 viral hepatitis 1:536–537
 vesicle packets (VPs) 1:498–499, 1:499
 vesicular stomatitis Alagoas virus (VSAV) 2:875
 vesicular stomatitis Indiana virus (VSIV) 2:516, 2:875
 vesicular stomatitis New Jersey virus (VSNJV) 2:875, 2:881f
 vesicular stomatitis virus (VSV) 1:64, 2:243, 1:10, 5:94, 1:659, 1:180, 5:239, 1:617, 2:808
 classification 2:875
 control and treatment 2:882–883
 epidemiology 2:880–881
 life cycle 2:878–879
 assembly and budding 2:879–880
 attachment, entry and uncoating 2:878–879
 gene expression and replication 2:879
 inhibition and modification of host cell functions 2:880
 pathogenesis and clinical features 2:881–882
 virus structure 2:875–878
 genomes and proteins 2:877–878
 VSV G 1:424
 VFs *see* virus factories (VFs)
 VHF *see* viral hemorrhagic fever (VHF)
 VHGs *see* Virus Hallmark Genes (VHGs)
 VHSV *see* viral hemorrhagic septicemia viruses (VHSV)

- ViCTree 1:102
vif gene 2:61*t*
 VIG *see* vaccinia immune globulin (VIG)
 VIGS *see* virus-induced gene silencing (VIGS)
 ViPR *see* Viral Pathogen Resource (ViPR); virus pathogen database and analysis resource (ViPR)
 ViPTree 1:102
 viral assembly 1:231*f*, 1:232
 viral assembly compartment (VAC) 2:452 defined 2:442
 viral capsid, defined 2:79, 1:248
 viral capsid antigen (VCA) IgG 5:195
 viral core, defined 2:643
 viral covert mortality disease (VCMD) 4:823, 4:825, 4:822
 viral disease, implications for 1:59 drug resistance 1:60 immune escape 1:59–60 short-term pathogenesis 1:59
 viral DNA packaging machines, enzymology of 4:124–125 genome maturation complex 4:128–129, 4:130 lambda cohesive end site and assembly of 4:128–129 model for the assembly of 4:129 The *cos*-cleavage reaction 4:129–130 genome packaging complex 4:130 assembly of the packaging motor 4:130 *cos*-clearance: transition to a translocating motor 4:130–131 nucleotide switch for *cos*-clearance? 4:131
 lambda system 4:126 *Escherichia Coli* integration host factor 4:128 TerL maturation domain 4:127–128 TerS subunit 4:126–127 TerL DNA packaging domain 4:127 terminase enzyme of phage lambda 4:126 packaging, termination of 4:132–133 terminase ejection and virion completion 4:133–134 unit length packaging motors 4:133 translocating motor 4:131–132 coordinated ATP hydrolysis by terminase motors 4:132 mechanochemical coupling and DNA translocation 4:132 structure of 4:131–132 unit length versus headful packaging mechanisms 4:124–125 viral terminase enzymes 4:125 TerL subunits 4:125–126 terminase holoenzymes 4:126 TerS subunits 4:125 viral dUTPase 2:671 viral envelope 1:468, 1:402, 4:359 viral evolutionary relationships from sequence analysis 1:88–90 viral factory, defined 1:372 viral fibers 1:225*f* viral fitness 4:419, 4:419–420 viral genome-assisted capsid assembly 1:486 viral genomic DNA, recognition of 4:141 concatemeric DNA and terminase proteins 4:141–142 cohesive ends 4:142–144 direct terminal repeats 4:142 headful bacteriophages, genome recognition in 4:144–146 monomeric genomes and terminal proteins 4:141 viral GM-CSF inhibitory factor (GIF) 2:671 viral hemorrhagic fever (VHF) 5:66 viral hemorrhagic septicemia viruses (VHSV) 2:324, 2:328 viral hepatitis 1:536–537 viral IL-10 2:671 viral infected myeloid cells 2:812 viral interferon resistance protein 2:671 viral killer toxins 4:534–535 dsRNA viruses and killer phenotype expression in *Saccharomyces cerevisiae* 4:535–536 endocytosis and intracellular transport of K28 virus toxin 4:538–539 ER exit and nuclear entry of the K28 virus toxin 4:539 K28 affecting DNA synthesis, cell-cycle progression, and induces apoptosis 4:539–541 killer virus-infected yeast, self-protection in 4:542 lethality of membrane damaging viral killer toxins 4:541–542 viral preprotoxin processing and toxin maturation 4:536–538 viral replication cycle 4:536 viral late transcription factor 3 (VLTf3) 1:377–378 viral membrane assembly proteins (VMAPs) 2:857 viral membranes 1:225*f* viral metagenomes archaeal viruses and 4:414–415 from hypersaline environments 4:415–416 viral metagenomics, defined 1:552 viral mutants, defined 3:727 viral nervous necrosis (VNN) 4:822–823, 4:819, 4:825 viral nucleic acid, delivery of 1:230*f* visualization of 1:230–232 viral operational taxonomic unit (vOTU) 1:621, 1:622 Viral Pathogen Resource (ViPR) 1:141 viral plaques 2:85 cats 2:85–86 diagnosis 2:90 dogs 2:85 horses 2:85 treatment 2:90 viral protein 24 (VP24) 2:612–613 viral protein 30 (VP30) 2:612 viral protein 35 (VP35) 2:612 viral protein 40 (VP40) 2:612 viral protein-genome linked 3:293, 3:692, 3:364, 3:348 viral proteins, translation of 1:444 canonical/cap-dependent mRNA translation 1:444–445 elongation and termination phases 1:446 initiation phase 1:444–445 ‘imprisonment’ of cellular mRNAs within the nucleus 1:447 microRNAs (miRNAs) 1:455 non-canonical initiation of virus mRNA translation 1:447–449 cap-independent translation enhancers (CITEs) 1:449 initiation at non-AUGs 1:450 internal ribosome entry sites (IRESes) 1:447–449 leaky scanning 1:450 ribosome reinitiation 1:450–452 ribosome ‘shunting’ 1:450 virus alternatives to components of initiation 1:449–450 translational ‘recoding’ 1:452 ribosomal bypassing (‘hopping’) 1:452 ribosomal ‘frame-shifting’ 1:452–454 ribosome stop codon ‘read-through’ 1:454 ‘stopgo’/‘stop carry-on’/ribosomal ‘skipping’ 1:454 virus-encoded proteinases 1:454–455 viral quasispecies, defined 5:27, 1:53 viral reactivation, defined 2:306 viral replicase proteins 1a and 2a 3:265 viral replication 4:898–900 viral replication complex (VRC) 1:495 defined 1:495 viral ribonucleoproteins (vRNPs) 2:118, 2:561 viral ribonucleoproteins/nucleocapsids assemblies 1:349 viral RNAs (vRNAs) 4:897 viral silencing suppressors (VSRs) 3:43 viral species, defined 2:48 viral suppressor of RNA silencing (VSR) 3:420, 3:545, 3:623 Viral Tiling theory (VTT) 1:251, 1:251*f*, 1:248, 1:250–251 viral titer detection 3:749 viral transcription 1:439 general principles of 1:439 5’ and 3’ ends promote translation, features of viral transcripts at 1:439–440 host transcription machinery, degree of dependence on 1:439 overall genome organization orchestrating viral transcriptional program 1:440–441 transcription strategies determining mechanisms of host shutoff 1:440 RNA-directed RNA polymerase (RdRp) 1:441 containing evolutionarily conserved architecture 1:441 coordinating transcription, capping, and polyadenylation 1:442–443 different RdRp and template conformations distinguish transcription from replication 1:441–442

- viral transcription (*continued*)
 VirCapSeq-VERT 5:93
 viremia 2:896, 2:805, 2:173, 2:48, 2:884
Virgaviridae 3:743, 3:603, 3:528–529, 3:839–843
 virgaviruses (*Virgaviridae*) 3:839–843
 diagnosis 3:849
 epidemiology and control 3:850
 genome organization 3:847
 infectious cycle 3:849
 pathogenicity 3:849–850
 replication 3:847–849
 taxonomy, phylogeny and evolution 3:843–844
 unassigned virgaviruses 3:850
 virion structure 3:844–847
 virion
 assembly 4:49
 assembly pathways 4:49
 capsid assembly 4:49–50
 tail assembly 4:49
 defined 2:442
 as a genome delivery device 1:402–403
 structure 4:769*f*, 2:806–808
 virion-associated protein (VAP) 3:163, 3:318
 virion-containing vesicles, defined 4:724
 virion egress 4:393–394
 cell membrane disruption 4:393–394
 membrane disruption, viral release without 4:394
 virion infectivity factor (*Vif*) 1:68
 virion-producing organisms 1:14
 virions, structural features of 4:841–843
 virion sense genes (MP and CP) 3:465
 virion structure and viral life cycle 4:381
 virocell 1:625, 1:14, 1:621
 virocell concept
 critics of 1:24–25
 diversity of virocells 1:25
 virocells and ribovirocells 1:25
 virocells with nucleus 1:25
 implications of 1:25–26
 enumeration of viruses in the environment 1:26
 living status of viruses 1:26–27
 viruses as cradles of new genes 1:25–26
 virion/virus paradigm 1:23
 critics of 1:23–24
 virocontrol, defined 4:431, 4:461
 viroids (*Pospiviroidae* and *Avsunviroidae*)
 classification 3:853–855
 host range and transmission 3:855
 genome structure 3:852–853
 movement 3:859–860
 pathogenicity 3:860–861
 implications for host range 3:861
 origin and evolution 3:861
 replication 3:858–859
 symptomatology 3:855
 epidemiology and control 3:855–858
 geographic distribution 3:855
 molecular biology 3:858
 taxonomy 3:852
 virology 1:3
 biochemical phase 1:6
 cell biology phase 1:6–7
 dawn of 1:3–5
 drugs 1:10–11
 epidemiology 1:9
 nobel prizes awarded for work of relevance to 1:4*t*
 pathogenicity and host defense mechanisms 1:9
 physico-chemical phase 1:5–6
 previrology 1:3
 selected historical highlights in 1:5*t*
 sequencing phase 1:8–9
 structural phase 1:7–8
 vaccines 1:10
 viral control 1:9–10
 viral diversity 1:7
 virus emergence and human activities 1:11–13
 virome 4:342
 defined 4:265
 virome, human
 composition and diversity of 1:553–555
 blood virome 1:556
 gastrointestinal tract virome 1:553–555
 genital and urinary tract viromes 1:556
 oral cavity and respiratory tract viromes 1:555–556
 skin virome 1:555
 definition of 1:552–553
 in human health and immunity 1:556–557
 technological development and limits for the description of 1:553
 viromics, defined 1:621, 1:133
 virophage 4:342, 4:677, 1:372, 3:681
 viroplasm, defined 2:343, 3:545, 3:567, 3:158
 viroporin, defined 2:428
 virosphere, defined 1:87
Virtovirus 3:581, 3:584
 virulence, defined 1:569, 2:697, 3:371
 virulence gene, defined 3:60, 4:849
 virulence of the virus, influence of 1:563
 virulent, defined 4:36
 virulent systemic strains of FCV (VS-FCV) 2:298
 virulent virus, defined 1:162, 4:368
 viruliferous, defined 3:200, 3:461
 viruliferous whitefly 3:749
 defined 3:768
 virus, defined 1:28
 virus-associated nucleic acid, defined 3:681
 virus-associated pyramids (VAP) 4:394, 4:395*f*
 virus attachment protein, defined 1:382
 virus bioinformatics 1:124
 diagnostics, tools for 1:124–125
 evolution and phylogenetics 1:128–129
 genome sequencing 1:125–127
 host transcriptomics 1:130–131
 machine learning as an opportunity 1:129–130
 RNA secondary structures in viruses 1:127
 technology and bioinformatics drive discoveries 1:124
 viral metagenomics 1:127–128
 virus-host interactions 1:129
 virus disassembly, defined 3:727
 virus-encoded proteinases 1:454–455
 virus-encoded suppressors of host RNAi (VSRs) 4:772
 virus factories (VFs) 1:495
 defined 1:495
 main components and functions of 1:496*f*
 methods to study 1:495–498
 representative examples of 1:498–499
 as targets for antiviral therapies 1:499
 Virus Hallmark Genes (VHG) 1:38–39, 1:39–40
 virus immunity, defined 3:554
 virus-induced gene silencing (VIGS) 3:124, 3:564, 3:555
 advantages and limitations 3:130
 application 3:128–130
 defined 3:420, 3:554
 establishment of a VIGS System 3:124–126
 post-transcriptional gene silencing (PTGS) and 3:124
 post-transcriptional gene silencing (PTGS) and VIGS 3:124
 vectors 3:126–130
 virus-induced gene silencing, defined 3:116
 virus infected cells 1:226*f*
 virus interactions impacting ecology and evolution 4:425–426
 virus like particle (VLP) 1:257, 3:96, 2:493, 1:662, 4:342, 5:295–296, 1:282–283, 2:844, 2:70, 4:259, 1:26, 4:343, 4:347, 4:348*f*
 antigen display on the surface of 1:669*f*
 -based vaccines for humans 1:668
 human papillomavirus (HPV) and hepatitis B virus (HBV) 1:668
 influenza 1:668
 polio 1:668–669
 biotechnology uses of 1:663*f*
 cell-specific cargo delivery using 1:670*f*
 defined 2:173
 expression systems to produce 1:664
 bacteria 1:664
 cell-free protein synthesis (CFPS) 1:665
 insect cells 1:664
 mammalian cells 1:664
 plants and plant cell culture 1:664–665
 yeast 1:664
 immunological properties of 1:662
 and induction of cellular immune response 1:662
 and induction of humoral immune response 1:662–663
 and stimulation of toll-like receptors 1:663–664
 virus-like particles as nano-carriers 1:665–666
 conjugation of antigens and virus-like particles 1:665–666
 chemical and affinity conjugation 1:666–667
 genetic fusion 1:665–666

SpyTag/SpyCatcher system 1:667
 loading of virus-like particles by
 encapsulation 1:667–668
 virus load, defined 5:197
 virus neutralisation test (VNT) 2:651–652,
 2:77, 2:360
 virus-neutralizing antibodies (VNAs) 2:741–
 742
 virus pathogen database and analysis
 resource (ViPR) 1:143
 virus reactivation 2:778
 virus re-assortment, defined 3:692, 3:348
 virus replication complexes, defined 3:692,
 3:364, 3:348
 virus resistance, defined 3:554
 virus response 1 (*vr1*) 4:432
 virus self-assembly, defined 3:727
 virus symptomatology, defined 3:539
 virus tolerance, defined 3:554
 viscerotropic disease 2:891
 visna-maedi virus (VMV) 2:56–57, 2:65,
 2:66–67
 vitellogenesis, defined 4:780
 vitivirus, defined 3:805
 vitrification 5:5, 4:342
 VLDL *see* very low-density lipoproteins
 (VLDL)
 VLDL-R *see* very low-density lipoprotein
 receptor (VLDL-R)
 VLP *see* virus like particle (VLP)
 VLTf3 *see* viral late transcription factor 3
 (VLTf3)
 VMAPs *see* viral membrane assembly
 proteins (VMAPs)
 VMV *see* visna-maedi virus (VMV)
 VNAs *see* virus-neutralizing antibodies
 (VNAs)
 VNN *see* viral nervous necrosis (VNN)
 VNT *see* virus neutralisation test (VNT)
 vOTU *see* viral operational taxonomic unit
 (vOTU)
 voxilaprevir 5:123
 VP3-Like protein, defined 4:380
 VP4-Like protein, defined 4:380
 VPg 4:770–771
 defined 4:768
vpr gene 2:61t, 2:62
 VPs *see* vesicle packets (VPs)
 Vps4 1:499
vpu gene 2:61t
vpw gene 2:61t
vpv gene 2:61t
vpv gene 2:61t
 VRC *see* viral replication complex (VRC)
 vRNAs *see* viral RNAs (vRNAs)
 vRNPs *see* viral ribonucleoproteins (vRNPs)
 VRs *see* variable regions (VRs)
 VSAV *see* vesicular stomatitis Alagoas virus
 (VSAV)
 VSIV *see* vesicular stomatitis Indiana virus
 (VSIV)
 VSNJV *see* vesicular stomatitis New Jersey
 virus (VSNJV)
 VSR *see* viral suppressor of RNA silencing
 (VSR)

VSRs *see* viral silencing suppressors (VSRs)
 VSRs *see* virus-encoded suppressors of host
 RNAi (VSRs)
 VSV *see* vesicular stomatitis virus (VSV)
 VTT *see* Viral Tiling theory (VTT)
 VZIG *see* varicella-zoster immunoglobulin
 (VZIG)
 VZV *see* varicella zoster virus (VZV)
 VZV infections *see* varicella-zoster virus
 (VZV) infections

W

waikaviruses 3:703
 classification 3:703–704
 diseases and management 3:710
 genome organization 3:707–708
 properties of viral proteins 3:708–709
 properties of virions 3:707
 transmission 3:709–710
 variation of isolates and strains 3:704–707
 walleye dermal sarcoma (WDS) 2:317
 walleye dermal sarcoma virus (WDSV)
 2:316, 2:318–319
 walleye discrete epidermal hyperplasia
 (WEH) 2:317
 walleye epidermal hyperplasia virus type 1
 (WEHV-1) 2:316
 walleye retroviruses 2:316
 wamavirus, defined 3:805
 water- and food-borne transmission 5:208
 watermelon mosaic virus and zucchini
 yellow mosaic virus
 biotechnological application 3:870
 classification 3:862–863
 control methods 3:867
 cross-protection 3:867
 prophylactic measures 3:867
 resistant cultivars 3:867–868
 diagnostic methods 3:865
 epidemiology 3:866
 efficient vectors 3:866
 pattern of spread 3:866–867
 virus sources 3:866
 genetics and evolution 3:868–870
 geographic distribution 3:865
 history and taxonomy 3:862
 host range 3:865
 symptomatology 3:863
 synergism and antagonism 3:864–865
 variability 3:868
 vector relationships 3:865–866
 WCBV *see* West Caucasian bat lyssavirus
 (WCBV)
 WCLMV *see* White clovermosaic virus
 (WCLMV)
 WDS *see* walleye dermal sarcoma (WDS)
 WDSV *see* walleye dermal sarcoma virus
 (WDSV)
 weak crystallization theory 1:297–299
 weapons of mass destruction, defined 1:644
 WEE complex of viruses *see* Western equine
 encephalomyelitis (WEE) complex of
 viruses
 WEEV *see* Western equine encephalitis virus
 (WEEV)
 WEH *see* walleye discrete epidermal
 hyperplasia (WEH)
 WEHV-1 *see* walleye epidermal hyperplasia
 virus type 1 (WEHV-1)
 Weighted Gene Coexpression Network
 Analysis (WGCNA) 1:143–144
 West Caucasian bat lyssavirus (WCBV)
 2:738
 Western equine encephalitis virus (WEEV)
 2:40, 2:43, 2:44, 2:45, 2:46
 Western equine encephalomyelitis (WEE)
 complex of viruses 1:66
 Westgard Rules 5:39–40, 5:57–58
 West Nile virus (WNV) 1:572, 1:180, 5:275,
 1:649, 2:848, 2:810–811, 2:812
 classification 2:884
 clinical manifestations 2:888–889
 diagnosis 2:889
 epidemiology 2:887–888
 genome and viral proteins 2:884–885
 capsid (C) protein 2:884–885
 envelope (E) protein 2:885
 prM protein 2:885
 life cycle 2:886–887
 NS1 2:885
 NS2A and NS2B 2:885–886
 NS3 2:886
 NS4A and NS4B 2:886
 NS5 2:886
 pathogenesis 2:889
 treatment and prevention 2:889
 virion structure 2:884
 WGCNA *see* Weighted Gene Coexpression
 Network Analysis (WGCNA)
 WGS *see* whole genome sequencing (WGS)
 wheat streak mosaic virus 1:649
 Whispovirus 4:808
 White clovermosaic virus (WCLMV) 3:628
 whiteflies
 biological control of 3:775
 chemical control of 3:775
 defined 3:768
 whitefly-transmitted begomoviruses, case of
 3:108–110
 white spot disease (WSD) 4:709
 white spot syndrome virus-CN (WSSV-CN)
 4:709
 white spot syndrome viruses (WSSVs)
 4:808, 4:813, 4:816, 4:809f, 4:810f,
 4:811t, 4:709
 white tail disease (WTD) 4:823, 4:825
 whole-genome amplification 1:187–188
 defined 1:184
 DNA decontamination 1:188
 reagent preparation 1:188
 DLB preparation 1:188
 DNA decontamination of DLB and stop
 solution 1:188
 master mix preparation 1:189
 stop solution 1:188
 whole genome reaction procedure 1:189
 whole genome sequencing (WGS) 1:565–
 566, 2:673

wHTH *see* winged HTH domain (wHTH)
 wild boar, control of classical swine fever in
 2:162–163
 wild poliovirus (WPV) 5:310
 wild-type (WT) capsids 4:118–119
 winged HTH domain (wHTH) 4:144
 WNV *see* West Nile virus (WNV)
 WPV *see* wild poliovirus (WPV)
 WRKY 3:758
 WSD *see* white spot disease (WSD)
 WSSV-CN *see* white spot syndrome virus-
 CN (WSSV-CN)
 WSSVs *see* white spot syndrome viruses
 (WSSVs)
 WT capsids *see* wild-type (WT) capsids
 WTD *see* white tail disease (WTD)

X

xenotropic MLVs (xMLVs) 2:646
 xenotropic murine leukemia virus-related
 virus (XMRV) 2:647
 XFMS *see* X-ray footprinting and mass
 spectrometry (XFMS)
Xinmoviridae 4:718–722
 xMLVs *see* xenotropic MLVs (xMLVs)
 XMRV *see* xenotropic murine leukemia
 virus-related virus (XMRV)
 X protein 2:376
 X-ray crystallography 4:900, 1:8
 X-ray diffraction, defined 1:362
 X-ray footprinting and mass spectrometry
 (XFMS) 1:197, 1:191
 XSV *see* extra small virus (XSV)
Xylella fastidiosa 4:772–773

Y

yado-kari, defined 4:658
 yado-kari virus 1 (YkV1) 4:440, 4:658
 future directions 4:663
 genome characteristics of 4:659–660
 -like virus combinations in other fungi
 4:663
 molecular entities sharing similar YkV1-
 like interactions 4:661–663
 phylogenetic placements of 4:660
 predicted past and future of 4:663
 proposed replication model for 4:660–661
 virion morphology 4:658
 yado-nushi, defined 4:658
 yado-nushi virus 1 (YnV1) 4:658
 future directions 4:663
 genome characteristics of 4:658–659
 -like virus combinations in other fungi
 4:663
 molecular entities sharing similar YnV1-
 like interactions 4:661–663
 phylogenetic placements of 4:660
 proposed replication model for 4:660–661
 virion morphology 4:658
 Yam virus X (YVX) 3:628

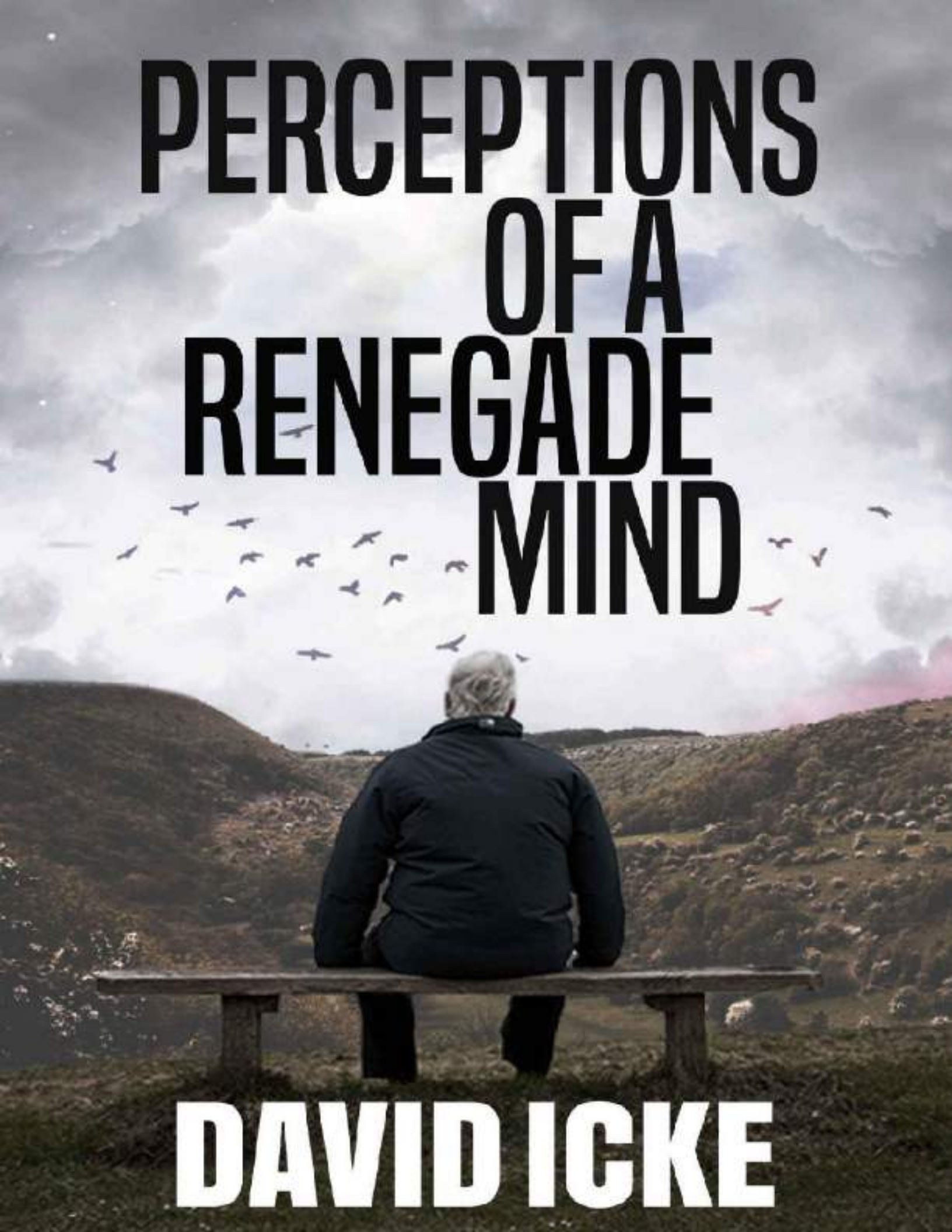
Yata virus (YATV) 2:875
 YATV *see* Yata virus (YATV)
 YBV *see* Yug Bogdanovac virus (YBV)
 yeast 1:664, 4:440
 yeast and fungi, prions of
 anti-prion systems 4:491
 biological roles of prions: a help or a
 hindrance? 4:490–491
 chaperones and prions 4:490
 enzyme as prion 4:492
 genetic signature of a prion 4:487
 history 4:487
 infectious prion amyloids have in-register
 parallel folded -sheet architecture
 4:489–490
 inositol polyphosphates and prion
 propagation 4:491–492
 prion generation, and [PIN]: a prion that
 gives rise to prions 4:490
 prion variant information templating
 mechanism 4:490
 prion variants and the species barrier
 4:490
 self-propagating amyloid as the basis for
 most yeast prions 4:487–489
 shuffleable prion domains suggests
 parallel-in-register -sheet structure
 4:489
 shuffling prion domains and amyloid
 structure 4:489
 yeast L-A virus 4:664
 genome organization 4:664–665
 history 4:664
 L-A genetics 4:667–668
 L-BC 4:668
 replication cycle 4:665–666
 RNA packaging 4:666
 RNA replication 4:666
 transcription 4:665–666
 20S RNA 4:668
 23S RNA 4:668
 viral translation 4:666–667
 virion structure 4:664
 YEL-AND *see* yellow fever vaccine associated
 neurotropic disease (YEL-AND)
 YEL-AVD *see* yellow fever vaccine-associated
 viscerotropic disease (YEL-AVD)
 yellow fever vaccine associated neurotropic
 disease (YEL-AND) 2:897–898
 yellow fever vaccine-associated viscerotropic
 disease (YEL-AVD) 2:897–898
 yellow fever virus (YFV) 1:542, 2:414, 2:256
 classification 2:891
 clinical features 2:895–896
 diagnosis 2:896–897
 epidemiology 2:894–895
 genome 2:892–893
 life cycle 2:893–894
 prevention 2:897–898
 treatment 2:897
 vaccine 1:575
 virion structure 2:891–892
 yellowing, defined 3:371
 yellow mosaic disease (YMD) 3:563
 YFV *see* yellow fever virus (YFV)

YPX_nL late domains 1:522
 Yug Bogdanovac virus (YBV) 2:875
 YVX *see* Yam virus X (YVX)

Z

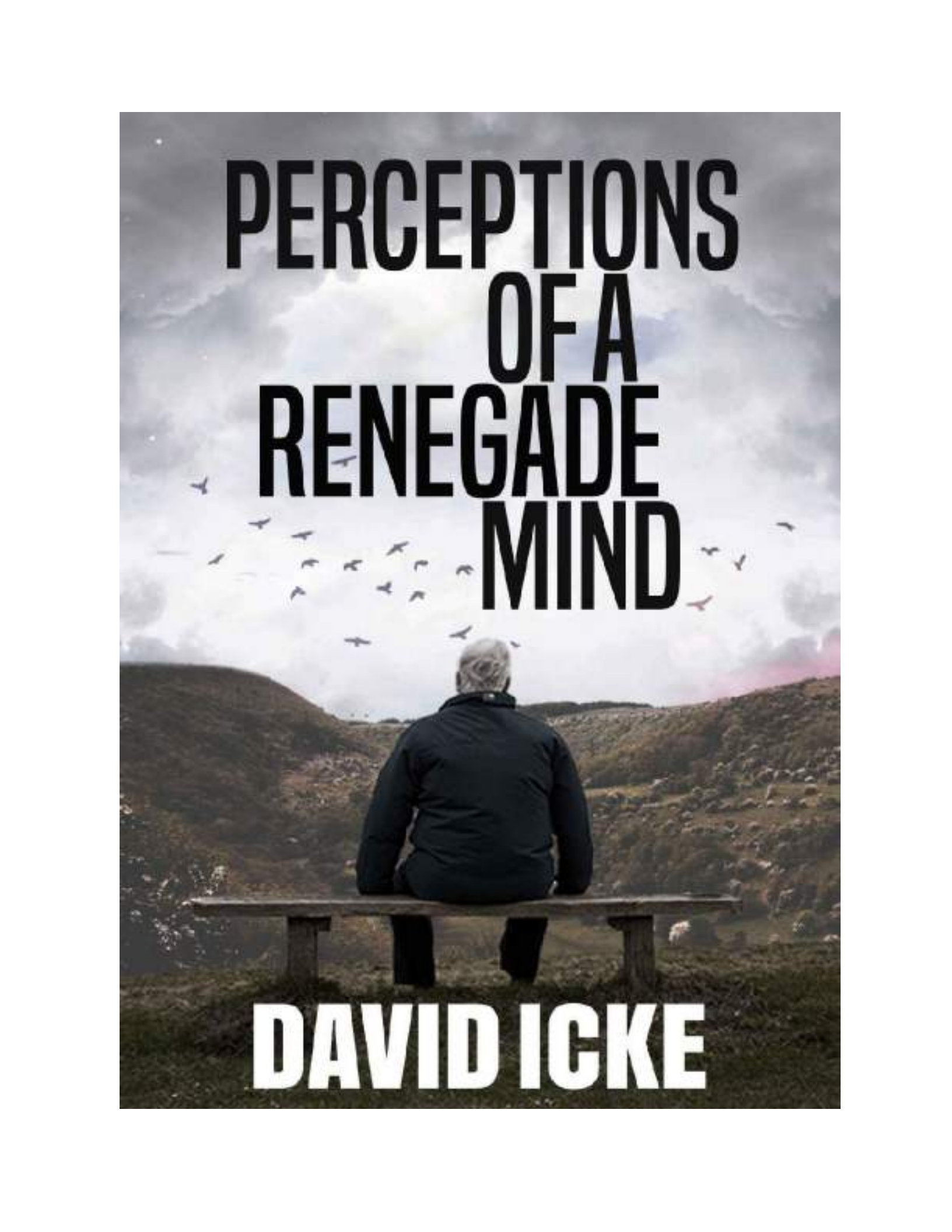
Zanamivir 5:164
 ZBD *see* zinc-binding domain (ZBD)
 zebrafish endogenous retrovirus 2:322
 zero-mode waveguides (ZMW) 1:181, 5:27
 Zika virus (ZIKV) 2:899, 1:253, 5:275,
 1:121, 5:3, 1:267–268, 1:268f,
 1:498–499
 classification 2:899–900
 clinical features 2:905–906
 diagnosis 2:907
 epidemiology 2:904–905
 brief history 2:904–905
 potential factors contributing to ZIKV
 emergence 2:905
 ZIKV lineages and disease association
 2:905
 life cycle 2:902–903
 ecology and transmission cycles 2:902–
 903
 virus replication cycle 2:903–904
 pathogenesis 2:906–907
 prevention and control 2:907–908
 replication cycle 2:904f
 treatment 2:907
 virion structure and genome organization
 2:900–902
 ZIKV *see* Zika virus (ZIKV)
 zinc-binding domain (ZBD) 2:264
 ZMW *see* zero-mode waveguides (ZMW)
 zone of emergence, defined 2:899
 zoonosis 1:569, 1:570f
 defined 1:569, 2:355, 1:542, 1:559
 emerging viruses, prevention and control
 of 1:572–574
 diagnosis 1:573–574
 mosquito vectors, control of 1:575
 treatment 1:574–575
 vaccination 1:575
 factors in increasing emergence of viral
 diseases 1:572
 interspecies transmission, factors in
 1:569–570
 epidemiological/ecological barriers 1:570
 host–pathogen interactions 1:570
 viral factors 1:570–571
 zoonotic infection, defined 2:397
 zoonotic influenza 5:160–161
 zoonotic transmission cycles, influence of
 1:563
 zoonotic viruses
 catch-all detection methods 5:260–261
 developing targeted, risk based sampling
 5:258–259
 disease X 5:257
 drivers of 5:256–257
 emerging disease detection, full genome
 sequencing and rapid data sharing
 5:262

- emerging disease detection and the one health concept 5:257–258
- from genotype to phenotype 5:261
- non-invasive and bulk samples, value of 5:260
- one health and animal reservoirs 5:261–262
- population surveys and sero-epidemiological studies 5:262–263
- sample collection 5:259–260
- SARS-CoV-2 5:261
- zoster, defined 2:860
- zucchini yellow mosaic virus
 - biotechnological application 3:870
 - classification 3:862–863
 - control methods 3:867
 - cross-protection 3:867
 - prophylactic measures 3:867
 - resistant cultivars 3:867–868
 - diagnostic methods 3:865
 - epidemiology 3:866
 - efficient vectors 3:866
 - pattern of spread 3:866–867
 - virus sources 3:866
 - genetics and evolution 3:868–870
 - geographic distribution 3:865
 - history and taxonomy 3:862
 - host range 3:865
 - symptomatology 3:863–864
 - synergism and antagonism 3:864–865
 - variability 3:868
 - vector relationships 3:865–866
- Zygocactus virus X (ZyVX) 3:623–624, 3:628
- ZyVX *see* Zygocactus virus X (ZyVX)

A person with grey hair, seen from behind, is sitting on a wooden bench. They are wearing a dark jacket and looking out over a landscape of rolling hills. The sky is filled with many birds in flight, and the clouds are dramatic and grey. The overall mood is contemplative and expansive.

PERCEPTIONS OF A RENEGADE MIND

DAVID ICKE



PERCEPTIONS OF A RENEGADE MIND

DAVID ICKE

PERCEPTIONS OF A RENEGADE MIND



ickonic
publishing

First published in July 2021.

ickonic
publishing

**New Enterprise House
St Helens Street
Derby
DE1 3GY
UK**

email: gareth.icke@davidicke.com

Copyright © 2021 David Icke

No part of this book may be reproduced in any form without permission from the Publisher, except for the quotation of brief passages in criticism

Cover Design: Gareth Icke
Book Design: Neil Hague

**British Library Cataloguing-in
Publication Data**

A catalogue record for this book is
available from the British Library

eISBN 978-18384153-1-0

PERCEPTIONS OF A RENEGADE MIND



DAVID ICKE

Dedication:

To Freeeeedom!

ICKONIC

THE ALTERNATIVE

NEW. DIFFERENT. REVOLUTIONARY

**HUNDREDS OF CUTTING EDGE DOCUMENTARIES,
FEATURE FILMS, SERIES & PODCASTS.**

SIGN UP NOW AT ICKONIC.COM

THE LIFE STORY OF DAVID ICKE

RENEGADE

THE FEATURE LENGTH FILM



AVAILABLE NOW AT DAVIDICKE.COM

Renegade:

Adjective

‘Having rejected tradition: Unconventional.’

Merriam-Webster Dictionary

Acquiescence to tyranny is the death of the spirit

You may be 38 years old, as I happen to be. And one day, some great opportunity stands before you and calls you to stand up for some great principle, some great issue, some great cause. And you refuse to do it because you are afraid ... You refuse to do it because you want to live longer ... You're afraid that you will lose your job, or you are afraid that you will be criticised or that you will lose your popularity, or you're afraid that somebody will stab you, or shoot at you or bomb your house; so you refuse to take the stand.

Well, you may go on and live until you are 90, but you're just as dead at 38 as you would be at 90. And the cessation of breathing in your life is but the belated announcement of an earlier death of the spirit.

Martin Luther King

**How the few control the many and always have – the many do
whatever they're told**

'Forward, the Light Brigade!'
Was there a man dismayed?
Not though the soldier knew
Someone had blundered.
Theirs not to make reply,
Theirs not to reason why,
Theirs but to do and die.
Into the valley of Death
Rode the six hundred.

Cannon to right of them,
Cannon to left of them,
Cannon in front of them
Volleyed and thundered;
Stormed at with shot and shell,
Boldly they rode and well,
Into the jaws of Death,
Into the mouth of hell
Rode the six hundred

Alfred Lord Tennyson (1809-1892)

The mist is lifting slowly
I can see the way ahead
And I've left behind the empty streets
That once inspired my life
And the strength of the emotion
Is like thunder in the air
'Cos the promise that we made each other
Haunts me to the end

The secret of your beauty
And the mystery of your soul
I've been searching for in everyone I meet
And the times I've been mistaken
It's impossible to say
And the grass is growing
Underneath our feet

The words that I remember
From my childhood still are true
That there's none so blind
As those who will not see
And to those who lack the courage
And say it's dangerous to try
Well they just don't know
That love eternal will not be denied

I know you're out there somewhere
Somewhere, somewhere
I know you're out there somewhere

Somewhere you can hear my voice
I know I'll find you somehow
Somehow, somehow
I know I'll find you somehow
And somehow I'll return again to you

The Moody Blues

Are you a gutless wonder - or a Renegade Mind?

Monuments put from pen to paper,
Turns me into a gutless wonder,
And if you tolerate this,
Then your children will be next.
Gravity keeps my head down,
Or is it maybe shame ...

Manic Street Preachers

Rise like lions after slumber
In unvanquishable number.
Shake your chains to earth like dew
Which in sleep have fallen on you.
Ye are many – they are few.

Percy Shelley

Contents

CHAPTER 1	'I'm thinking' – Oh, but <i>are</i> you?
CHAPTER 2	Renegade perception
CHAPTER 3	The Pushbacker sting
CHAPTER 4	'Covid': The calculated catastrophe
CHAPTER 5	There <i>is no</i> 'virus'
CHAPTER 6	Sequence of deceit
CHAPTER 7	War on your mind
CHAPTER 8	'Reframing' insanity
CHAPTER 9	We must have it? So what is it?
CHAPTER 10	Human 2.0
CHAPTER 11	Who controls the Cult?
CHAPTER 12	Escaping Wetiko
POSTSCRIPT	
APPENDIX	Cowan-Kaufman-Morell Statement on Virus Isolation
BIBLIOGRAPHY	
INDEX	

CHAPTER ONE

I'm thinking' – Oh, but *are* you?

Think for yourself and let others enjoy the privilege of doing so too
Voltaire

French-born philosopher, mathematician and scientist René Descartes became famous for his statement in Latin in the 17th century which translates into English as: 'I think, therefore I am.'

On the face of it that is true. Thought reflects perception and perception leads to both behaviour and self-identity. In that sense 'we' are what we think. But who or what is doing the thinking and is thinking the only route to perception? Clearly, as we shall see, 'we' are not always the source of 'our' perception, indeed with regard to humanity as a whole this is rarely the case; and thinking is far from the only means of perception. Thought is the village idiot compared with other expressions of consciousness that we all have the potential to access and tap into. This has to be true when we *are* those other expressions of consciousness which are infinite in nature. We have forgotten this, or, more to the point, been manipulated to forget.

These are not just the esoteric musings of the navel. The whole foundation of human control and oppression is control of perception. Once perception is hijacked then so is behaviour which is dictated by perception. Collective perception becomes collective behaviour and collective behaviour is what we call human society. Perception is all and those behind human control know that which is

why perception is the target 24/7 of the psychopathic manipulators that I call the Global Cult. They know that if they dictate perception they will dictate behaviour and collectively dictate the nature of human society. They are further aware that perception is formed from information received and if they control the circulation of information they will to a vast extent direct human behaviour. Censorship of information and opinion has become globally Nazi-like in recent years and never more blatantly than since the illusory 'virus pandemic' was triggered out of China in 2019 and across the world in 2020. Why have billions submitted to house arrest and accepted fascistic societies in a way they would have never believed possible? Those controlling the information spewing from government, mainstream media and Silicon Valley (all controlled by the same Global Cult networks) told them they were in danger from a 'deadly virus' and only by submitting to house arrest and conceding their most basic of freedoms could they and their families be protected. This monumental and provable lie became the *perception* of the billions and therefore the *behaviour* of the billions. In those few words you have the whole structure and modus operandi of human control. Fear is a perception – False Emotion Appearing Real – and fear is the currency of control. In short ... get them by the balls (or give them the impression that you have) and their hearts and minds will follow. Nothing grips the dangly bits and freezes the rear-end more comprehensively than fear.

World number 1

There are two 'worlds' in what appears to be one 'world' and the prime difference between them is knowledge. First we have the mass of human society in which the population is maintained in coldly-calculated ignorance through control of information and the 'education' (indoctrination) system. That's all you really need to control to enslave billions in a perceptual delusion in which what are perceived to be *their* thoughts and opinions are ever-repeated mantras that the system has been downloading all their lives through 'education', media, science, medicine, politics and academia

in which the personnel and advocates are themselves overwhelmingly the perceptual products of the same repetition. Teachers and academics in general are processed by the same programming machine as everyone else, but unlike the great majority they never leave the 'education' program. It gripped them as students and continues to grip them as programmers of subsequent generations of students. The programmed become the programmers – the programmed programmers. The same can largely be said for scientists, doctors and politicians and not least because as the American writer Upton Sinclair said: 'It is difficult to get a man to understand something when his salary depends upon his not understanding it.' If your career and income depend on thinking the way the system demands then you will – bar a few free-minded exceptions – concede your mind to the Perceptual Mainframe that I call the Postage Stamp Consensus. This is a tiny band of perceived knowledge and possibility 'taught' (downloaded) in the schools and universities, pounded out by the mainstream media and on which all government policy is founded. Try thinking, and especially speaking and acting, outside of the 'box' of consensus and see what that does for your career in the Mainstream Everything which bullies, harasses, intimidates and ridicules the population into compliance. Here we have the simple structure which enslaves most of humanity in a perceptual prison cell for an entire lifetime and I'll go deeper into this process shortly. Most of what humanity is taught as fact is nothing more than programmed belief. American science fiction author Frank Herbert was right when he said: 'Belief can be manipulated. Only knowledge is dangerous.' In the 'Covid' age belief is promoted and knowledge is censored. It was always so, but never to the extreme of today.

World number 2

A 'number 2' is slang for 'doing a poo' and how appropriate that is when this other 'world' is doing just that on humanity every minute of every day. World number 2 is a global network of secret societies and semi-secret groups dictating the direction of society via

governments, corporations and authorities of every kind. I have spent more than 30 years uncovering and exposing this network that I call the Global Cult and knowing its agenda is what has made my books so accurate in predicting current and past events. Secret societies are secret for a reason. They want to keep their hoarded knowledge to themselves and their chosen initiates and to hide it from the population which they seek through ignorance to control and subdue. The whole foundation of the division between World 1 and World 2 is *knowledge*. What number 1 knows number 2 must not. Knowledge they have worked so hard to keep secret includes (a) the agenda to enslave humanity in a centrally-controlled global dictatorship, and (b) the nature of reality and life itself. The latter (b) must be suppressed to allow the former (a) to prevail as I shall be explaining. The way the Cult manipulates and interacts with the population can be likened to a spider's web. The 'spider' sits at the centre in the shadows and imposes its will through the web with each strand represented in World number 2 by a secret society, satanic or semi-secret group, and in World number 1 – the world of the seen – by governments, agencies of government, law enforcement, corporations, the banking system, media conglomerates and Silicon Valley (Fig 1 overleaf). The spider and the web connect and coordinate all these organisations to pursue the same global outcome while the population sees them as individual entities working randomly and independently. At the level of the web governments *are* the banking system *are* the corporations *are* the media *are* Silicon Valley *are* the World Health Organization working from their inner cores as one unit. Apparently unconnected countries, corporations, institutions, organisations and people are on the *same team* pursuing the same global outcome. Strands in the web immediately around the spider are the most secretive and exclusive secret societies and their membership is emphatically restricted to the Cult inner-circle emerging through the generations from particular bloodlines for reasons I will come to. At the core of the core you would get them in a single room. That's how many people are dictating the direction of human society and its transformation

through the 'Covid' hoax and other means. As the web expands out from the spider we meet the secret societies that many people will be aware of – the Freemasons, Knights Templar, Knights of Malta, Opus Dei, the inner sanctum of the Jesuit Order, and such like. Note how many are connected to the Church of Rome and there is a reason for that. The Roman Church was established as a revamp, a rebranding, of the relocated 'Church' of Babylon and the Cult imposing global tyranny today can be tracked back to Babylon and Sumer in what is now Iraq.



Figure 1: The global web through which the few control the many. (Image Neil Hague.)

Inner levels of the web operate in the unseen away from the public eye and then we have what I call the cusp organisations located at the point where the hidden meets the seen. They include a series of satellite organisations answering to a secret society founded in London in the late 19th century called the Round Table and among them are the Royal Institute of International Affairs (UK, founded in 1920); Council on Foreign Relations (US, 1921); Bilderberg Group (worldwide, 1954); Trilateral Commission (US/worldwide, 1972); and the Club of Rome (worldwide, 1968) which was created to exploit environmental concerns to justify the centralisation of global power to 'save the planet'. The Club of Rome instigated with others the human-caused climate change hoax which has led to all the 'green

new deals' demanding that very centralisation of control. Cusp organisations, which include endless 'think tanks' all over the world, are designed to coordinate a single global policy between political and business leaders, intelligence personnel, media organisations and anyone who can influence the direction of policy in their own sphere of operation. Major players and regular attenders will know what is happening – or some of it – while others come and go and are kept overwhelmingly in the dark about the big picture. I refer to these cusp groupings as semi-secret in that they can be publicly identified, but what goes on at the inner-core is kept very much 'in house' even from most of their members and participants through a fiercely-imposed system of compartmentalisation. Only let them know what they need to know to serve your interests and no more. The structure of secret societies serves as a perfect example of this principle. Most Freemasons never get higher than the bottom three levels of 'degree' (degree of knowledge) when there are 33 official degrees of the Scottish Rite. Initiates only qualify for the next higher 'compartment' or degree if those at that level choose to allow them. Knowledge can be carefully assigned only to those considered 'safe'. I went to my local Freemason's lodge a few years ago when they were having an 'open day' to show how cuddly they were and when I chatted to some of them I was astonished at how little the rank and file knew even about the most ubiquitous symbols they use. The mushroom technique – keep them in the dark and feed them bullshit – applies to most people in the web as well as the population as a whole. Sub-divisions of the web mirror in theme and structure transnational corporations which have a headquarters somewhere in the world dictating to all their subsidiaries in different countries. Subsidiaries operate in their methodology and branding to the same centrally-dictated plan and policy in pursuit of particular ends. The Cult web functions in the same way. Each country has its own web as a subsidiary of the global one. They consist of networks of secret societies, semi-secret groups and bloodline families and their job is to impose the will of the spider and the global web in their particular country. Subsidiary networks control and manipulate the national political system, finance, corporations, media, medicine, etc. to

ensure that they follow the globally-dictated Cult agenda. These networks were the means through which the 'Covid' hoax could be played out with almost every country responding in the same way.

The 'Yessir' pyramid

Compartmentalisation is the key to understanding how a tiny few can dictate the lives of billions when combined with a top-down sequence of imposition and acquiescence. The inner core of the Cult sits at the peak of the pyramidal hierarchy of human society (Fig 2 overleaf). It imposes its will – its agenda for the world – on the level immediately below which acquiesces to that imposition. This level then imposes the Cult will on the level below them which acquiesces and imposes on the next level. Very quickly we meet levels in the hierarchy that have no idea there even is a Cult, but the sequence of imposition and acquiescence continues down the pyramid in just the same way. 'I don't know why we are doing this but the order came from "on-high" and so we better just do it.' Alfred Lord Tennyson said of the cannon fodder levels in his poem *The Charge of the Light Brigade*: 'Theirs not to reason why; theirs but to do and die.' The next line says that 'into the valley of death rode the six hundred' and they died because they obeyed without question what their perceived 'superiors' told them to do. In the same way the population capitulated to 'Covid'. The whole hierarchical pyramid functions like this to allow the very few to direct the enormous many. Eventually imposition-acquiescence-imposition-acquiescence comes down to the mass of the population at the foot of the pyramid. If they acquiesce to those levels of the hierarchy imposing on them (governments/law enforcement/doctors/media) a circuit is completed between the population and the handful of super-psychopaths in the Cult inner core at the top of the pyramid. Without a circuit-breaking refusal to obey, the sequence of imposition and acquiescence allows a staggeringly few people to impose their will upon the entirety of humankind. We are looking at the very sequence that has subjugated billions since the start of 2020. Our freedom has not been taken from us. Humanity has given it

away. Fascists do not impose fascism because there are not enough of them. Fascism is imposed by the population acquiescing to fascism. Put another way allowing their perceptions to be programmed to the extent that leads to the population giving their freedom away by giving their perceptions – their mind – away. If this circuit is not broken by humanity ceasing to cooperate with their own enslavement then nothing can change. For that to happen people have to critically think and see through the lies and window dressing and then summon the backbone to act upon what they see. The Cult spends its days working to stop either happening and its methodology is systematic and highly detailed, but it can be overcome and that is what this book is all about.

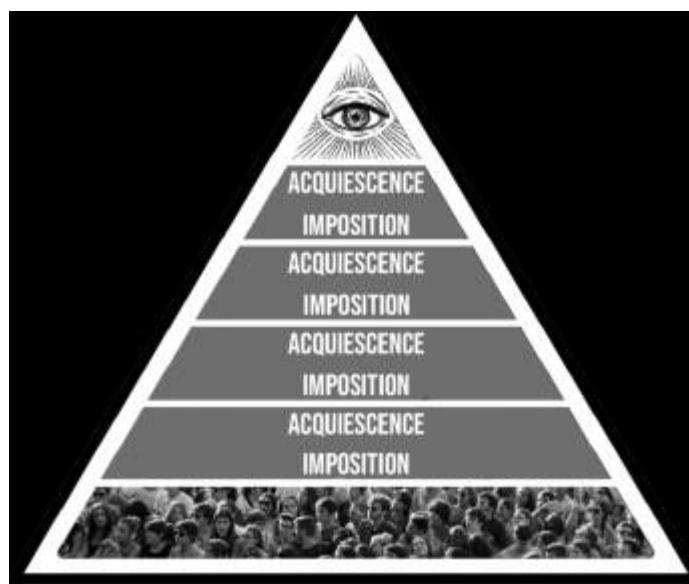


Figure 2: The simple sequence of imposition and compliance that allows a handful of people at the peak of the pyramid to dictate the lives of billions.

The Life Program

Okay, back to world number 1 or the world of the 'masses'. Observe the process of what we call 'life' and it is a perceptual download from cradle to grave. The Cult has created a global structure in which perception can be programmed and the program continually topped-up with what appears to be constant confirmation that the program is indeed true reality. The important word here is 'appears'.

This is the structure, the fly-trap, the Postage Stamp Consensus or Perceptual Mainframe, which represents that incredibly narrow band of perceived possibility delivered by the 'education' system, mainstream media, science and medicine. From the earliest age the download begins with parents who have themselves succumbed to the very programming their children are about to go through. Most parents don't do this out of malevolence and mostly it is quite the opposite. They do what they believe is best for their children and that is what the program has told them is best. Within three or four years comes the major transition from parental programming to full-blown state (Cult) programming in school, college and university where perceptually-programmed teachers and academics pass on their programming to the next generations. Teachers who resist are soon marginalised and their careers ended while children who resist are called a problem child for whom Ritalin may need to be prescribed. A few years after entering the 'world' children are under the control of authority figures representing the state telling them when they have to be there, when they can leave and when they can speak, eat, even go to the toilet. This is calculated preparation for a lifetime of obeying authority in all its forms. Reflex-action fear of authority is instilled by authority from the start. Children soon learn the carrot and stick consequences of obeying or defying authority which is underpinned daily for the rest of their life. Fortunately I daydreamed through this crap and never obeyed authority simply because it told me to. This approach to my alleged 'betters' continues to this day. There can be consequences of pursuing open-minded freedom in a world of closed-minded conformity. I spent a lot of time in school corridors after being ejected from the classroom for not taking some of it seriously and now I spend a lot of time being ejected from Facebook, YouTube and Twitter. But I can tell you that being true to yourself and not compromising your self-respect is far more exhilarating than bowing to authority for authority's sake. You don't have to be a sheep to the shepherd (authority) and the sheep dog (fear of not obeying authority).

The perceptual download continues throughout the formative years in school, college and university while script-reading 'teachers', 'academics' 'scientists', 'doctors' and 'journalists' insist that ongoing generations must be as programmed as they are. Accept the program or you will not pass your 'exams' which confirm your 'degree' of programming. It is tragic to think that many parents pressure their offspring to work hard at school to download the program and qualify for the next stage at college and university. The late, great, American comedian George Carlin said: 'Here's a bumper sticker I'd like to see: We are proud parents of a child who has resisted his teachers' attempts to break his spirit and bend him to the will of his corporate masters.' Well, the best of luck finding many of those, George. Then comes the moment to leave the formal programming years in academia and enter the 'adult' world of work. There you meet others in your chosen or prescribed arena who went through the same Postage Stamp Consensus program before you did. There is therefore overwhelming agreement between almost everyone on the basic foundations of Postage Stamp reality and the rejection, even contempt, of the few who have a mind of their own and are prepared to use it. This has two major effects. Firstly, the consensus confirms to the programmed that their download is really how things are. I mean, everyone knows that, right? Secondly, the arrogance and ignorance of Postage Stamp adherents ensure that anyone questioning the program will have unpleasant consequences for seeking their own truth and not picking their perceptions from the shelf marked: 'Things you must believe without question and if you don't you're a dangerous lunatic conspiracy theorist and a harebrained nutter'.

Every government, agency and corporation is founded on the same Postage Stamp prison cell and you can see why so many people believe the same thing while calling it their own 'opinion'. Fusion of governments and corporations in pursuit of the same agenda was the definition of fascism described by Italian dictator Benito Mussolini. The pressure to conform to perceptual norms downloaded for a lifetime is incessant and infiltrates society right

down to family groups that become censors and condemners of their own 'black sheep' for not, ironically, being sheep. We have seen an explosion of that in the 'Covid' era. Cult-owned global media unleashes its propaganda all day every day in support of the Postage Stamp and targets with abuse and ridicule anyone in the public eye who won't bend their mind to the will of the tyranny. Any response to this is denied (certainly in my case). They don't want to give a platform to expose official lies. Cult-owned-and-created Internet giants like Facebook, Google, YouTube and Twitter delete you for having an unapproved opinion. Facebook boasts that its AI censors delete 97-percent of 'hate speech' before anyone even reports it. Much of that 'hate speech' will simply be an opinion that Facebook and its masters don't want people to see. Such perceptual oppression is widely known as fascism. Even Facebook executive Benny Thomas, a 'CEO Global Planning Lead', said in comments secretly recorded by investigative journalism operation Project Veritas that Facebook is 'too powerful' and should be broken up:

I mean, no king in history has been the ruler of two billion people, but Mark Zuckerberg is ... And he's 36. That's too much for a 36-year-old ... You should not have power over two billion people. I just think that's wrong.

Thomas said Facebook-owned platforms like Instagram, Oculus, and WhatsApp needed to be separate companies. 'It's too much power when they're all one together'. That's the way the Cult likes it, however. We have an executive of a Cult organisation in Benny Thomas that doesn't know there is a Cult such is the compartmentalisation. Thomas said that Facebook and Google 'are no longer companies, they're countries'. Actually they are more powerful than countries on the basis that if you control information you control perception and control human society.

I love my oppressor

Another expression of this psychological trickery is for those who realise they are being pressured into compliance to eventually

convince themselves to believe the official narratives to protect their self-respect from accepting the truth that they have succumbed to meek and subservient compliance. Such people become some of the most vehement defenders of the system. You can see them everywhere screaming abuse at those who prefer to think for themselves and by doing so reminding the compliers of their own capitulation to conformity. 'You are talking dangerous nonsense you Covidiot!!' Are you trying to convince me or yourself? It is a potent form of Stockholm syndrome which is defined as: 'A psychological condition that occurs when a victim of abuse identifies and attaches, or bonds, positively with their abuser.' An example is hostages bonding and even 'falling in love' with their kidnappers. The syndrome has been observed in domestic violence, abused children, concentration camp inmates, prisoners of war and many and various Satanic cults. These are some traits of Stockholm syndrome listed at goodtherapy.org:

- Positive regard towards perpetrators of abuse or captor [see 'Covid'].
- Failure to cooperate with police and other government authorities when it comes to holding perpetrators of abuse or kidnapping accountable [or in the case of 'Covid' cooperating with the police to enforce and defend their captors' demands].
- Little or no effort to escape [see 'Covid'].
- Belief in the goodness of the perpetrators or kidnappers [see 'Covid'].
- Appeasement of captors. This is a manipulative strategy for maintaining one's safety. As victims get rewarded – perhaps with less abuse or even with life itself – their appeasing behaviours are reinforced [see 'Covid'].
- Learned helplessness. This can be akin to 'if you can't beat 'em, join 'em'. As the victims fail to escape the abuse or captivity, they may start giving up and soon realize it's just easier for everyone if they acquiesce all their power to their captors [see 'Covid'].

- Feelings of pity toward the abusers, believing they are actually victims themselves. Because of this, victims may go on a crusade or mission to 'save' [protect] their abuser [see the venom unleashed on those challenging the official 'Covid' narrative].
- Unwillingness to learn to detach from their perpetrators and heal. In essence, victims may tend to be less loyal to themselves than to their abuser [*definitely* see 'Covid'].

Ponder on those traits and compare them with the behaviour of great swathes of the global population who have defended governments and authorities which have spent every minute destroying their lives and livelihoods and those of their children and grandchildren since early 2020 with fascistic lockdowns, house arrest and employment deletion to 'protect' them from a 'deadly virus' that their abusers' perceptually created to bring about this very outcome. We are looking at mass Stockholm syndrome. All those that agree to concede their freedom will believe those perceptions are originating in their own independent 'mind' when in fact by conceding their reality to Stockholm syndrome they have by definition conceded any independence of mind. Listen to the 'opinions' of the acquiescing masses in this 'Covid' era and what gushes forth is the repetition of the official version of everything delivered unprocessed, unfiltered and unquestioned. The whole programming dynamic works this way. I must be free because I'm told that I am and so I think that I am.

You can see what I mean with the chapter theme of 'I'm thinking – Oh, but *are* you?' The great majority are not thinking, let alone for themselves. They are repeating what authority has told them to believe which allows them to be controlled. Weaving through this mentality is the fear that the 'conspiracy theorists' are right and this again explains the often hysterical abuse that ensues when you dare to contest the official narrative of anything. Denial is the mechanism of hiding from yourself what you don't want to be true. Telling people what they want to hear is easy, but it's an infinitely greater challenge to tell them what they would rather not be happening.

One is akin to pushing against an open door while the other is met with vehement resistance no matter what the scale of evidence. I don't want it to be true so I'll convince myself that it's not. Examples are everywhere from the denial that a partner is cheating despite all the signs to the reflex-action rejection of any idea that world events in which country after country act in exactly the same way are centrally coordinated. To accept the latter is to accept that a force of unspeakable evil is working to destroy your life and the lives of your children with nothing too horrific to achieve that end. Who the heck wants that to be true? But if we don't face reality the end is duly achieved and the consequences are far worse and ongoing than breaking through the walls of denial today with the courage to make a stand against tyranny.

Connect the dots – but how?

A crucial aspect of perceptual programming is to portray a world in which everything is random and almost nothing is connected to anything else. Randomness cannot be coordinated by its very nature and once you perceive events as random the idea they could be connected is waved away as the rantings of the tinfoil-hat brigade. You can't plan and coordinate random you idiot! No, you can't, but you can hide the coldly-calculated and long-planned behind the *illusion* of randomness. A foundation manifestation of the Renegade Mind is to scan reality for patterns that connect the apparently random and turn pixels and dots into pictures. This is the way I work and have done so for more than 30 years. You look for similarities in people, modus operandi and desired outcomes and slowly, then ever quicker, the picture forms. For instance: There would seem to be no connection between the 'Covid pandemic' hoax and the human-caused global-warming hoax and yet they are masks (appropriately) on the same face seeking the same outcome. Those pushing the global warming myth through the Club of Rome and other Cult agencies are driving the lies about 'Covid' – Bill Gates is an obvious one, but they are endless. Why would the same people be involved in both when they are clearly not connected? Oh, but they

are. Common themes with personnel are matched by common goals. The 'solutions' to both 'problems' are centralisation of global power to impose the will of the few on the many to 'save' humanity from 'Covid' and save the planet from an 'existential threat' (we need 'zero Covid' and 'zero carbon emissions'). These, in turn, connect with the 'dot' of globalisation which was coined to describe the centralisation of global power in every area of life through incessant political and corporate expansion, trading blocks and superstates like the European Union. If you are the few and you want to control the many you have to centralise power and decision-making. The more you centralise power the more power the few at the centre will have over the many; and the more that power is centralised the more power those at the centre have to centralise even quicker. The momentum of centralisation gets faster and faster which is exactly the process we have witnessed. In this way the hoaxed 'pandemic' and the fakery of human-caused global warming serve the interests of globalisation and the seizure of global power in the hands of the Cult inner-circle which is behind 'Covid', 'climate change' and globalisation. At this point random 'dots' become a clear and obvious picture or pattern.

Klaus Schwab, the classic Bond villain who founded the Cult's Gates-funded World Economic Forum, published a book in 2020, *The Great Reset*, in which he used the 'problem' of 'Covid' to justify a total transformation of human society to 'save' humanity from 'climate change'. Schwab said: 'The pandemic represents a rare but narrow window of opportunity to reflect, reimagine, and reset our world.' What he didn't mention is that the Cult he serves is behind both hoaxes as I show in my book *The Answer*. He and the Cult don't have to reimagine the world. They know precisely what they want and that's why they destroyed human society with 'Covid' to 'build back better' in their grand design. Their job is not to imagine, but to get humanity to imagine and agree with their plans while believing it's all random. It must be pure coincidence that 'The Great Reset' has long been the Cult's code name for the global imposition of fascism and replaced previous code-names of the 'New World

Order' used by Cult frontmen like Father George Bush and the 'New Order of the Ages' which emerged from Freemasonry and much older secret societies. New Order of the Ages appears on the reverse of the Great Seal of the United States as 'Novus ordo seclorum' underneath the Cult symbol used since way back of the pyramid and all seeing-eye (Fig 3). The pyramid is the hierarchy of human control headed by the illuminated eye that symbolises the force behind the Cult which I will expose in later chapters. The term 'Annuit Coeptis' translates as 'He favours our undertaking'. We are told the 'He' is the Christian god, but 'He' is not as I will be explaining.



Figure 3: The all-seeing eye of the Cult 'god' on the Freemason-designed Great Seal of the United States and also on the dollar bill.

Having you on

Two major Cult techniques of perceptual manipulation that relate to all this are what I have called since the 1990s Problem-Reaction-Solution (PRS) and the Totalitarian Tiptoe (TT). They can be uncovered by the inquiring mind with a simple question: Who benefits? The answer usually identifies the perpetrators of a given action or happening through the concept of 'he who most benefits from a crime is the one most likely to have committed it'. The Latin 'Cue bono?' – Who benefits? – is widely attributed to the Roman orator and statesman Marcus Tullius Cicero. No wonder it goes back so far when the concept has been relevant to human behaviour since

history was recorded. Problem-Reaction-Solution is the technique used to manipulate us every day by covertly creating a problem (or the illusion of one) and offering the solution to the problem (or the illusion of one). In the first phase you create the problem and blame someone or something else for why it has happened. This may relate to a financial collapse, terrorist attack, war, global warming or pandemic, anything in fact that will allow you to impose the 'solution' to change society in the way you desire at that time. The 'problem' doesn't have to be real. PRS is manipulation of perception and all you need is the population to believe the problem is real. Human-caused global warming and the 'Covid pandemic' only have to be *perceived* to be real for the population to accept the 'solutions' of authority. I refer to this technique as NO-Problem-Reaction-Solution. Billions did not meekly accept house arrest from early 2020 because there was a real deadly 'Covid pandemic' but because they perceived – believed – that to be the case. The antidote to Problem-Reaction-Solution is to ask who benefits from the proposed solution. Invariably it will be anyone who wants to justify more control through deletion of freedom and centralisation of power and decision-making.

The two world wars were Problem-Reaction-Solutions that transformed and realigned global society. Both were manipulated into being by the Cult as I have detailed in books since the mid-1990s. They dramatically centralised global power, especially World War Two, which led to the United Nations and other global bodies thanks to the overt and covert manipulations of the Rockefeller family and other Cult bloodlines like the Rothschilds. The UN is a stalking horse for full-blown world government that I will come to shortly. The land on which the UN building stands in New York was donated by the Rockefellers and the same Cult family was behind Big Pharma scalpel and drug 'medicine' and the creation of the World Health Organization as part of the UN. They have been stalwarts of the eugenics movement and funded Hitler's race-purity expert' Ernst Rudin. The human-caused global warming hoax has been orchestrated by the Club of Rome through the UN which is

manufacturing both the 'problem' through its Intergovernmental Panel on Climate Change and imposing the 'solution' through its Agenda 21 and Agenda 2030 which demand the total centralisation of global power to 'save the world' from a climate hoax the United Nations is itself perpetrating. What a small world the Cult can be seen to be particularly among the inner circles. The bedfellow of Problem-Reaction-Solution is the Totalitarian Tiptoe which became the Totalitarian Sprint in 2020. The technique is fashioned to hide the carefully-coordinated behind the cover of apparently random events. You start the sequence at 'A' and you know you are heading for 'Z'. You don't want people to know that and each step on the journey is presented as a random happening while all the steps strung together lead in the same direction. The speed may have quickened dramatically in recent times, but you can still see the incremental approach of the Tiptoe in the case of 'Covid' as each new imposition takes us deeper into fascism. Tell people they have to do this or that to get back to 'normal', then this and this and this. With each new demand adding to the ones that went before the population's freedom is deleted until it disappears. The spider wraps its web around the flies more comprehensively with each new diktat. I'll highlight this in more detail when I get to the 'Covid' hoax and how it has been pulled off. Another prime example of the Totalitarian Tiptoe is how the Cult-created European Union went from a 'free-trade zone' to a centralised bureaucratic dictatorship through the Tiptoe of incremental centralisation of power until nations became mere administrative units for Cult-owned dark suits in Brussels.

The antidote to ignorance is knowledge which the Cult seeks vehemently to deny us, but despite the systematic censorship to that end the Renegade Mind can overcome this by vociferously seeking out the facts no matter the impediments put in the way. There is also a method of thinking and perceiving – *knowing* – that doesn't even need names, dates, place-type facts to identify the patterns that reveal the story. I'll get to that in the final chapter. All you need to know about the manipulation of human society and to what end is still out there – *at the time of writing* – in the form of books, videos

and websites for those that really want to breach the walls of programmed perception. To access this knowledge requires the abandonment of the mainstream media as a source of information in the awareness that this is owned and controlled by the Cult and therefore promotes mass perceptions that suit the Cult. Mainstream media lies all day, every day. That is its function and very reason for being. Where it does tell the truth, here and there, is only because the truth and the Cult agenda very occasionally coincide. If you look for fact and insight to the BBC, CNN and virtually all the rest of them you are asking to be conned and perceptually programmed.

Know the outcome and you'll see the journey

Events seem random when you have no idea where the world is being taken. Once you do the random becomes the carefully planned. Know the outcome and you'll see the journey is a phrase I have been using for a long time to give context to daily happenings that appear unconnected. Does a problem, or illusion of a problem, trigger a proposed 'solution' that further drives society in the direction of the outcome? Invariably the answer will be yes and the random – *abracadabra* – becomes the clearly coordinated. So what is this outcome that unlocks the door to a massively expanded understanding of daily events? I will summarise its major aspects – the fine detail is in my other books – and those new to this information will see that the world they thought they were living in is a very different place. The foundation of the Cult agenda is the incessant centralisation of power and all such centralisation is ultimately in pursuit of Cult control on a global level. I have described for a long time the planned world structure of top-down dictatorship as the Hunger Games Society. The term obviously comes from the movie series which portrayed a world in which a few living in military-protected hi-tech luxury were the overlords of a population condemned to abject poverty in isolated 'sectors' that were not allowed to interact. 'Covid' lockdowns and travel bans anyone? The 'Hunger Games' pyramid of structural control has the inner circle of the Cult at the top with pretty much the entire

population at the bottom under their control through dependency for survival on the Cult. The whole structure is planned to be protected and enforced by a military-police state ([Fig 4](#)).

Here you have the reason for the global lockdowns of the fake pandemic to coldly destroy independent incomes and livelihoods and make everyone dependent on the 'state' (the Cult that controls the 'states'). I have warned in my books for many years about the plan to introduce a 'guaranteed income' – a barely survivable pittance – designed to impose dependency when employment was destroyed by AI technology and now even more comprehensively at great speed by the 'Covid' scam. Once the pandemic was played and lockdown consequences began to delete independent income the authorities began to talk right on cue about the need for a guaranteed income and a 'Great Reset'. Guaranteed income will be presented as benevolent governments seeking to help a desperate people – desperate as a direct result of actions of the same governments. The truth is that such payments are a trap. You will only get them if you do exactly what the authorities demand including mass vaccination (genetic manipulation). We have seen this theme already in Australia where those dependent on government benefits have them reduced if parents don't agree to have their children vaccinated according to an insane health-destroying government-dictated schedule. Calculated economic collapse applies to governments as well as people. The Cult wants rid of countries through the creation of a world state with countries broken up into regions ruled by a world government and super states like the European Union. Countries must be bankrupted, too, to this end and it's being achieved by the trillions in 'rescue packages' and furlough payments, trillions in lost taxation, and money-no-object spending on 'Covid' including constant all-medium advertising (programming) which has made the media dependent on government for much of its income. The day of reckoning is coming – as planned – for government spending and given that it has been made possible by printing money and not by production/taxation there is inflation on the way that has the

potential to wipe out monetary value. In that case there will be no need for the Cult to steal your money. It just won't be worth anything (see the German Weimar Republic before the Nazis took over). Many have been okay with lockdowns while getting a percentage of their income from so-called furlough payments without having to work. Those payments are dependent, however, on people having at least a theoretical job with a business considered non-essential and ordered to close. As these business go under because they are closed by lockdown after lockdown the furlough stops and it will for everyone eventually. Then what? The 'then what?' is precisely the idea.



Figure 4: The Hunger Games Society structure I have long warned was planned and now the 'Covid' hoax has made it possible. This is the real reason for lockdowns.

Hired hands

Between the Hunger Games Cult elite and the dependent population is planned to be a vicious military-police state (a fusion of the two into one force). This has been in the making for a long time with police looking ever more like the military and carrying weapons to match. The pandemic scam has seen this process accelerate so fast as

lockdown house arrest is brutally enforced by carefully recruited fascist minds and gormless system-servers. The police and military are planned to merge into a centrally-directed world army in a global structure headed by a world government which wouldn't be elected even by the election fixes now in place. The world army is not planned even to be human and instead wars would be fought, primarily against the population, using robot technology controlled by artificial intelligence. I have been warning about this for decades and now militaries around the world are being transformed by this very AI technology. The global regime that I describe is a particular form of fascism known as a technocracy in which decisions are not made by clueless and co-opted politicians but by unelected technocrats – scientists, engineers, technologists and bureaucrats. Cult-owned-and-controlled Silicon Valley giants are examples of technocracy and they already have far more power to direct world events than governments. They are with their censorship *selecting* governments. I know that some are calling the 'Great Reset' a Marxist communist takeover, but fascism and Marxism are different labels for the same tyranny. Tell those who lived in fascist Germany and Stalinist Russia that there was a difference in the way their freedom was deleted and their lives controlled. I could call it a fascist technocracy or a Marxist technocracy and they would be equally accurate. The Hunger Games society with its world government structure would oversee a world army, world central bank and single world cashless currency imposing its will on a microchipped population ([Fig 5](#)). Scan its different elements and see how the illusory pandemic is forcing society in this very direction at great speed. Leaders of 23 countries and the World Health Organization (WHO) backed the idea in March, 2021, of a global treaty for 'international cooperation' in 'health emergencies' and nations should 'come together as a global community for peaceful cooperation that extends beyond this crisis'. Cut the Orwellian bullshit and this means another step towards global government. The plan includes a cashless digital money system that I first warned about in 1993. Right at the start of 'Covid' the deeply corrupt Tedros

Adhanom Ghebreyesus, the crooked and merely gofer 'head' of the World Health Organization, said it was possible to catch the 'virus' by touching cash and it was better to use cashless means. The claim was ridiculous nonsense and like the whole 'Covid' mind-trick it was nothing to do with 'health' and everything to do with pushing every aspect of the Cult agenda. As a result of the Tedros lie the use of cash has plummeted. The Cult script involves a single world digital currency that would eventually be technologically embedded in the body. China is a massive global centre for the Cult and if you watch what is happening there you will know what is planned for everywhere. The Chinese government is developing a digital currency which would allow fines to be deducted immediately via AI for anyone caught on camera breaking its fantastic list of laws and the money is going to be programmable with an expiry date to ensure that no one can accrue wealth except the Cult and its operatives.



Figure 5: The structure of global control the Cult has been working towards for so long and this has been enormously advanced by the 'Covid' illusion.

Serfdom is so smart

The Cult plan is far wider, extreme, and more comprehensive than even most conspiracy researchers appreciate and I will come to the true depths of deceit and control in the chapters 'Who controls the

Cult?’ and ‘Escaping Wetiko’. Even the world that we know is crazy enough. We are being deluged with ever more sophisticated and controlling technology under the heading of ‘smart’. We have smart televisions, smart meters, smart cards, smart cars, smart driving, smart roads, smart pills, smart patches, smart watches, smart skin, smart borders, smart pavements, smart streets, smart cities, smart communities, smart environments, smart growth, smart planet ... smart *everything* around us. Smart technologies and methods of operation are designed to interlock to create a global Smart Grid connecting the entirety of human society including human minds to create a centrally-dictated ‘hive’ mind. ‘Smart cities’ is code for densely-occupied megacities of total surveillance and control through AI. Ever more destructive frequency communication systems like 5G have been rolled out without any official testing for health and psychological effects (colossal). 5G/6G/7G systems are needed to run the Smart Grid and each one becomes more destructive of body and mind. Deleting independent income is crucial to forcing people into these AI-policed prisons by ending private property ownership (except for the Cult elite). The Cult’s Great Reset now openly foresees a global society in which no one will own any possessions and everything will be rented while the Cult would own literally everything under the guise of government and corporations. The aim has been to use the lockdowns to destroy sources of income on a mass scale and when the people are destitute and in unrepayable amounts of debt (problem) Cult assets come forward with the pledge to write-off debt in return for handing over all property and possessions (solution). Everything – literally everything including people – would be connected to the Internet via AI. I was warning years ago about the coming Internet of Things (IoT) in which all devices and technology from your car to your fridge would be plugged into the Internet and controlled by AI. Now we are already there with much more to come. The next stage is the Internet of Everything (IoE) which is planned to include the connection of AI to the human brain and body to replace the human mind with a centrally-controlled AI mind. Instead of perceptions

being manipulated through control of information and censorship those perceptions would come direct from the Cult through AI. What do you think? You think whatever AI decides that you think. In human terms there would be no individual 'think' any longer. Too incredible? The ravings of a lunatic? Not at all. Cult-owned crazies in Silicon Valley have been telling us the plan for years without explaining the real motivation and calculated implications. These include Google executive and 'futurist' Ray Kurzweil who highlights the year 2030 for when this would be underway. He said:

Our thinking ... will be a hybrid of biological and non-biological thinking ... humans will be able to extend their limitations and 'think in the cloud' ... We're going to put gateways to the cloud in our brains ... We're going to gradually merge and enhance ourselves ... In my view, that's the nature of being human – we transcend our limitations.

As the technology becomes vastly superior to what we are then the small proportion that is still human gets smaller and smaller and smaller until it's just utterly negligible.

The sales-pitch of Kurzweil and Cult-owned Silicon Valley is that this would make us 'super-human' when the real aim is to make us post-human and no longer 'human' in the sense that we have come to know. The entire global population would be connected to AI and become the centrally-controlled 'hive-mind' of externally-delivered perceptions. The Smart Grid being installed to impose the Cult's will on the world is being constructed to allow particular locations – even one location – to control the whole global system. From these prime control centres, which absolutely include China and Israel, anything connected to the Internet would be switched on or off and manipulated at will. Energy systems could be cut, communication via the Internet taken down, computer-controlled driverless autonomous vehicles driven off the road, medical devices switched off, the potential is limitless given how much AI and Internet connections now run human society. We have seen nothing yet if we allow this to continue. Autonomous vehicle makers are working with law enforcement to produce cars designed to automatically pull over if they detect a police or emergency vehicle flashing from up to 100 feet away. At a police stop the car would be unlocked and the

window rolled down automatically. Vehicles would only take you where the computer (the state) allowed. The end of petrol vehicles and speed limiters on all new cars in the UK and EU from 2022 are steps leading to electric computerised transport over which ultimately you have no control. The picture is far bigger even than the Cult global network or web and that will become clear when I get to the nature of the 'spider'. There is a connection between all these happenings and the instigation of DNA-manipulating 'vaccines' (which aren't 'vaccines') justified by the 'Covid' hoax. That connection is the unfolding plan to transform the human body from a biological to a synthetic biological state and this is why synthetic biology is such a fast-emerging discipline of mainstream science. 'Covid vaccines' are infusing self-replicating synthetic genetic material into the cells to cumulatively take us on the Totalitarian Tiptoe from Human 1.0 to the synthetic biological Human 2.0 which will be physically and perceptually attached to the Smart Grid to one hundred percent control every thought, perception and deed. Humanity needs to wake up and *fast*.

This is the barest explanation of where the 'outcome' is planned to go but it's enough to see the journey happening all around us. Those new to this information will already see 'Covid' in a whole new context. I will add much more detail as we go along, but for the minutiae evidence see my mega-works, *The Answer*, *The Trigger* and *Everything You Need to Know But Have Never Been Told*.

Now – how does a Renegade Mind see the 'world'?

CHAPTER TWO

Renegade Perception

It is one thing to be clever and another to be wise

George R.R. Martin

A simple definition of the difference between a programmed mind and a Renegade Mind would be that one sees only dots while the other connects them to see the picture. Reading reality with accuracy requires the observer to (a) know the planned outcome and (b) realise that everything, but *everything*, is connected.

The entirety of infinite reality is connected – that's its very nature – and with human society an expression of infinite reality the same must apply. Simple cause and effect is a connection. The effect is triggered by the cause and the effect then becomes the cause of another effect. Nothing happens in isolation because it *can't*. Life in whatever reality is simple choice and consequence. We make choices and these lead to consequences. If we don't like the consequences we can make different choices and get different consequences which lead to other choices and consequences. The choice and the consequence are not only connected they are indivisible. You can't have one without the other as an old song goes. A few cannot control the world unless those being controlled allow that to happen – cause and effect, choice and consequence. Control – who has it and who doesn't – is a two-way process, a symbiotic relationship, involving the controller and controlled. 'They took my freedom away!!' Well, yes, but you also gave it to them. Humanity is

subjected to mass control because humanity has acquiesced to that control. This is all cause and effect and literally a case of give and take. In the same way world events of every kind are connected and the Cult works incessantly to sell the illusion of the random and coincidental to maintain the essential (to them) perception of dots that hide the picture. Renegade Minds know this and constantly scan the world for patterns of connection. This is absolutely pivotal in understanding the happenings in the world and without that perspective clarity is impossible. First you know the planned outcome and then you identify the steps on the journey – the day-by-day apparently random which, when connected in relation to the outcome, no longer appear as individual events, but as the proverbial *chain* of events leading in the same direction. I'll give you some examples:

Political puppet show

We are told to believe that politics is 'adversarial' in that different parties with different beliefs engage in an endless tussle for power. There may have been some truth in that up to a point – and only a point – but today divisions between 'different' parties are rhetorical not ideological. Even the rhetorical is fusing into one-speak as the parties eject any remaining free thinkers while others succumb to the ever-gathering intimidation of anyone with the 'wrong' opinion. The Cult is not a new phenomenon and can be traced back thousands of years as my books have documented. Its intergenerational initiatives have been manipulating events with increasing effect the more that global power has been centralised. In ancient times the Cult secured control through the system of monarchy in which 'special' bloodlines (of which more later) demanded the right to rule as kings and queens simply by birthright and by vanquishing others who claimed the same birthright. There came a time, however, when people had matured enough to see the unfairness of such tyranny and demanded a say in who governed them. Note the word – *governed* them. Not served them – *governed* them, hence government defined as 'the political direction and control exercised over the

actions of the members, citizens, or inhabitants of communities, societies, and states; direction of the affairs of a state, community, etc.' Governments exercise control over rather than serve just like the monarchies before them. Bizarrely there are still countries like the United Kingdom which are ruled by a monarch *and* a government that officially answers to the monarch. The UK head of state and that of Commonwealth countries such as Canada, Australia and New Zealand is 'selected' by who in a *single family* had unprotected sex with whom and in what order. Pinch me it can't be true. Ouch! Shit, it is. The demise of monarchies in most countries offered a potential vacuum in which some form of free and fair society could arise and the Cult had that base covered. Monarchies had served its interests but they couldn't continue in the face of such widespread opposition and, anyway, replacing a 'royal' dictatorship that people could see with a dictatorship 'of the people' hiding behind the concept of 'democracy' presented far greater manipulative possibilities and ways of hiding coordinated tyranny behind the illusion of 'freedom'.

Democracy is quite wrongly defined as government selected by the population. This is not the case at all. It is government selected by *some* of the population (and then only in theory). This 'some' doesn't even have to be the majority as we have seen so often in first-past-the-post elections in which the so-called majority party wins fewer votes than the 'losing' parties combined. Democracy can give total power to a party in government from a minority of the votes cast. It's a sleight of hand to sell tyranny as freedom. Seventy-four million Trump-supporting Americans didn't vote for the 'Democratic' Party of Joe Biden in the distinctly dodgy election in 2020 and yet far from acknowledging the wishes and feelings of that great percentage of American society the Cult-owned Biden government set out from day one to destroy them and their right to a voice and opinion. Empty shell Biden and his Cult handlers said they were doing this to 'protect democracy'. Such is the level of lunacy and sickness to which politics has descended. Connect the dots and relate them to the desired outcome – a world government run by self-appointed technocrats and no longer even elected

politicians. While operating through its political agents in government the Cult is at the same time encouraging public disdain for politicians by putting idiots and incompetents in theoretical power on the road to deleting them. The idea is to instil a public reaction that says of the technocrats: 'Well, they couldn't do any worse than the pathetic politicians.' It's all about controlling perception and Renegade Minds can see through that while programmed minds cannot when they are ignorant of both the planned outcome and the manipulation techniques employed to secure that end. This knowledge can be learned, however, and fast if people choose to get informed.

Politics may at first sight appear very difficult to control from a central point. I mean look at the 'different' parties and how would you be able to oversee them all and their constituent parts? In truth, it's very straightforward because of their structure. We are back to the pyramid of imposition and acquiescence. Organisations are structured in the same way as the system as a whole. Political parties are not open forums of free expression. They are hierarchies. I was a national spokesman for the British Green Party which claimed to be a different kind of politics in which influence and power was devolved; but I can tell you from direct experience – and it's far worse now – that Green parties are run as hierarchies like all the others however much they may try to hide that fact or kid themselves that it's not true. A very few at the top of all political parties are directing policy and personnel. They decide if you are elevated in the party or serve as a government minister and to do that you have to be a yes man or woman. Look at all the maverick political thinkers who never ascended the greasy pole. If you want to progress within the party or reach 'high-office' you need to fall into line and conform. Exceptions to this are rare indeed. Should you want to run for parliament or Congress you have to persuade the local or state level of the party to select you and for that you need to play the game as dictated by the hierarchy. If you secure election and wish to progress within the greater structure you need to go on conforming to what is acceptable to those running the hierarchy

from the peak of the pyramid. Political parties are perceptual gulags and the very fact that there are party 'Whips' appointed to 'whip' politicians into voting the way the hierarchy demands exposes the ridiculous idea that politicians are elected to serve the people they are supposed to represent. Cult operatives and manipulation has long seized control of major parties that have any chance of forming a government and at least most of those that haven't. A new party forms and the Cult goes to work to infiltrate and direct. This has reached such a level today that you see video compilations of 'leaders' of all parties whether Democrats, Republicans, Conservative, Labour and Green parroting the same Cult mantra of 'Build Back Better' and the 'Great Reset' which are straight off the Cult song-sheet to describe the transformation of global society in response to the Cult-instigated hoaxes of the 'Covid pandemic' and human-caused 'climate change'. To see Caroline Lucas, the Green Party MP that I knew when I was in the party in the 1980s, speaking in support of plans proposed by Cult operative Klaus Schwab representing the billionaire global elite is a real head-shaker.

Many parties – one master

The party system is another mind-trick and was instigated to change the nature of the dictatorship by swapping 'royalty' for dark suits that people believed – though now ever less so – represented their interests. Understanding this trick is to realise that a single force (the Cult) controls all parties either directly in terms of the major ones or through manipulation of perception and ideology with others. You don't need to manipulate Green parties to demand your transformation of society in the name of 'climate change' when they are obsessed with the lie that this is essential to 'save the planet'. You just give them a platform and away they go serving your interests while believing they are being environmentally virtuous. America's political structure is a perfect blueprint for how the two or multi-party system is really a one-party state. The Republican Party is controlled from one step back in the shadows by a group made up of billionaires and their gofers known as neoconservatives or Neocons.

I have exposed them in fine detail in my books and they were the driving force behind the policies of the imbecilic presidency of Boy George Bush which included 9/11 (see *The Trigger* for a comprehensive demolition of the official story), the subsequent 'war on terror' (war of terror) and the invasions of Afghanistan and Iraq. The latter was a No-Problem-Reaction-Solution based on claims by Cult operatives, including Bush and British Prime Minister Tony Blair, about Saddam Hussein's 'weapons of mass destruction' which did not exist as war criminals Bush and Blair well knew.



Figure 6: Different front people, different parties – same control system.

The Democratic Party has its own 'Neocon' group controlling from the background which I call the 'Democons' and here's the penny-drop – the Neocons and Democons answer to the same masters one step further back into the shadows (Fig 6). At that level of the Cult the Republican and Democrat parties are controlled by the same people and no matter which is in power the Cult is in power. This is how it works in almost every country and certainly in Britain with Conservative, Labour, Liberal Democrat and Green parties now all on the same page whatever the rhetoric may be in their feeble attempts to appear different. Neocons operated at the time of Bush through a think tank called The Project for the New American Century which in September, 2000, published a document entitled *Rebuilding America's Defenses: Strategies, Forces, and Resources*

For a New Century demanding that America fight ‘multiple, simultaneous major theatre wars’ as a ‘core mission’ to force regime-change in countries including Iraq, Libya and Syria. Neocons arranged for Bush (‘Republican’) and Blair (‘Labour Party’) to front-up the invasion of Iraq and when they departed the Democons orchestrated the targeting of Libya and Syria through Barack Obama (‘Democrat’) and British Prime Minister David Cameron (‘Conservative Party’). We have ‘different’ parties and ‘different’ people, but the same unfolding script. The more the Cult has seized the reigns of parties and personnel the more their policies have transparently pursued the same agenda to the point where the fascist ‘Covid’ impositions of the Conservative junta of Jackboot Johnson in Britain were opposed by the Labour Party because they were not fascist enough. The Labour Party is likened to the US Democrats while the Conservative Party is akin to a British version of the Republicans and on both sides of the Atlantic they all speak the same language and support the direction demanded by the Cult although some more enthusiastically than others. It’s a similar story in country after country because it’s all centrally controlled. Oh, but what about Trump? I’ll come to him shortly. Political ‘choice’ in the ‘party’ system goes like this: You vote for Party A and they get into government. You don’t like what they do so next time you vote for Party B and they get into government. You don’t like what they do when it’s pretty much the same as Party A and why wouldn’t that be with both controlled by the same force? Given that only two, sometimes three, parties have any chance of forming a government to get rid of Party B that you don’t like you have to vote again for Party A which ... you don’t like. This, ladies and gentlemen, is what they call ‘democracy’ which we are told – wrongly – is a term interchangeable with ‘freedom’.

The cult of cults

At this point I need to introduce a major expression of the Global Cult known as Sabbatian-Frankism. Sabbatian is also spelt as Sabbatean. I will summarise here. I have published major exposés

and detailed background in other works. Sabbatian-Frankism combines the names of two frauds posing as 'Jewish' men, Sabbatai Zevi (1626-1676), a rabbi, black magician and occultist who proclaimed he was the Jewish messiah; and Jacob Frank (1726-1791), the Polish 'Jew', black magician and occultist who said he was the reincarnation of 'messiah' Zevi and biblical patriarch Jacob. They worked across two centuries to establish the Sabbatian-Frankist cult that plays a major, indeed central, role in the manipulation of human society by the Global Cult which has its origins much further back in history than Sabbatai Zevi. I should emphasise two points here in response to the shrill voices that will scream 'anti-Semitism': (1) Sabbatian-Frankists are NOT Jewish and only pose as such to hide their cult behind a Jewish façade; and (2) my information about this cult has come from Jewish sources who have long realised that their society and community has been infiltrated and taken over by interloper Sabbatian-Frankists. Infiltration has been the foundation technique of Sabbatian-Frankism from its official origin in the 17th century. Zevi's Sabbatian sect attracted a massive following described as the biggest messianic movement in Jewish history, spreading as far as Africa and Asia, and he promised a return for the Jews to the 'Promised Land' of Israel. Sabbatianism was not Judaism but an inversion of everything that mainstream Judaism stood for. So much so that this sinister cult would have a feast day when Judaism had a fast day and whatever was forbidden in Judaism the Sabbatians were encouraged and even commanded to do. This included incest and what would be today called Satanism. Members were forbidden to marry outside the sect and there was a system of keeping their children ignorant of what they were part of until they were old enough to be trusted not to unknowingly reveal anything to outsiders. The same system is employed to this day by the Global Cult in general which Sabbatian-Frankism has enormously influenced and now largely controls.

Zevi and his Sabbatians suffered a setback with the intervention by the Sultan of the Islamic Ottoman Empire in the Middle East and what is now the Republic of Turkey where Zevi was located. The

Sultan gave him the choice of proving his 'divinity', converting to Islam or facing torture and death. Funnily enough Zevi chose to convert or at least appear to. Some of his supporters were disillusioned and drifted away, but many did not with 300 families also converting – only in theory – to Islam. They continued behind this Islamic smokescreen to follow the goals, rules and rituals of Sabbatianism and became known as 'crypto-Jews' or the 'Dönme' which means 'to turn'. This is rather ironic because they didn't 'turn' and instead hid behind a fake Islamic persona. The process of appearing to be one thing while being very much another would become the calling card of Sabbatianism especially after Zevi's death and the arrival of the Satanist Jacob Frank in the 18th century when the cult became Sabbatian-Frankism and plumbed still new depths of depravity and infiltration which included – still includes – human sacrifice and sex with children. Wherever Sabbatians go paedophilia and Satanism follow and is it really a surprise that Hollywood is so infested with child abuse and Satanism when it was established by Sabbatian-Frankists and is still controlled by them? Hollywood has been one of the prime vehicles for global perceptual programming and manipulation. How many believe the version of 'history' portrayed in movies when it is a travesty and inversion (again) of the truth? Rabbi Marvin Antelman describes Frankism in his book, *To Eliminate the Opiate*, as 'a movement of complete evil' while Jewish professor Gershom Scholem said of Frank in *The Messianic Idea in Judaism*: 'In all his actions [he was] a truly corrupt and degenerate individual ... one of the most frightening phenomena in the whole of Jewish history.' Frank was excommunicated by traditional rabbis, as was Zevi, but Frank was undeterred and enjoyed vital support from the House of Rothschild, the infamous banking dynasty whose inner-core are Sabbatian-Frankists and not Jews. Infiltration of the Roman Church and Vatican was instigated by Frank with many Dönme 'turning' again to convert to Roman Catholicism with a view to hijacking the reins of power. This was the ever-repeating modus operandi and continues to be so. Pose as an advocate of the religion, culture or country that you want to control and then

manipulate your people into the positions of authority and influence largely as advisers, administrators and Svengalis for those that appear to be in power. They did this with Judaism, Christianity (Christian Zionism is part of this), Islam and other religions and nations until Sabbatian-Frankism spanned the world as it does today.

Sabbatian Saudis and the terror network

One expression of the Sabbatian-Frankist Dönme within Islam is the ruling family of Saudi Arabia, the House of Saud, through which came the vile distortion of Islam known as Wahhabism. This is the violent creed followed by terrorist groups like Al-Qaeda and ISIS or Islamic State. Wahhabism is the hand-chopping, head-chopping 'religion' of Saudi Arabia which is used to keep the people in a constant state of fear so the interloper House of Saud can continue to rule. Al-Qaeda and Islamic State were lavishly funded by the House of Saud while being created and directed by the Sabbatian-Frankist network in the United States that operates through the Pentagon, CIA and the government in general of whichever 'party'. The front man for the establishment of Wahhabism in the middle of the 18th century was a Sabbatian-Frankist 'crypto-Jew' posing as Islamic called Muhammad ibn Abd al-Wahhab. His daughter would marry the son of Muhammad bin Saud who established the first Saudi state before his death in 1765 with support from the British Empire. Bin Saud's successors would establish modern Saudi Arabia in league with the British and Americans in 1932 which allowed them to seize control of Islam's major shrines in Mecca and Medina. They have dictated the direction of Sunni Islam ever since while Iran is the major centre of the Shiite version and here we have the source of at least the public conflict between them. The Sabbatian network has used its Wahhabi extremists to carry out Problem-Reaction-Solution terrorist attacks in the name of 'Al-Qaeda' and 'Islamic State' to justify a devastating 'war on terror', ever-increasing surveillance of the population and to terrify people into compliance. Another insight of the Renegade Mind is the streetwise understanding that

just because a country, location or people are attacked doesn't mean that those apparently representing that country, location or people are not behind the attackers. Often they are *orchestrating* the attacks because of the societal changes that can be then justified in the name of 'saving the population from terrorists'.

I show in great detail in *The Trigger* how Sabbatian-Frankists were the real perpetrators of 9/11 and not '19 Arab hijackers' who were blamed for what happened. Observe what was justified in the name of 9/11 alone in terms of Middle East invasions, mass surveillance and control that fulfilled the demands of the Project for the New American Century document published by the Sabbatian Neocons. What appear to be enemies are on the deep inside players on the same Sabbatian team. Israel and Arab 'royal' dictatorships are all ruled by Sabbatians and the recent peace agreements between Israel and Saudi Arabia, the United Arab Emirates (UAE) and others are only making formal what has always been the case behind the scenes. Palestinians who have been subjected to grotesque tyranny since Israel was bombed and terrorised into existence in 1948 have never stood a chance. Sabbatian-Frankists have controlled Israel (so the constant theme of violence and war which Sabbatians love) and they have controlled the Arab countries that Palestinians have looked to for real support that never comes. 'Royal families' of the Arab world in Saudi Arabia, Bahrain, UAE, etc., are all Sabbatians with allegiance to the aims of the cult and not what is best for their Arabic populations. They have stolen the oil and financial resources from their people by false claims to be 'royal dynasties' with a genetic right to rule and by employing vicious militaries to impose their will.

Satanic 'illumination'

The Satanist Jacob Frank formed an alliance in 1773 with two other Sabbatians, Mayer Amschel Rothschild (1744-1812), founder of the Rothschild banking dynasty, and Jesuit-educated fraudulent Jew, Adam Weishaupt, and this led to the formation of the Bavarian Illuminati, firstly under another name, in 1776. The Illuminati would

be the manipulating force behind the French Revolution (1789-1799) and was also involved in the American Revolution (1775-1783) before and after the Illuminati's official creation. Weishaupt would later become (in public) a Protestant Christian in archetypal Sabbatian style. I read that his name can be decoded as Adam-Weishaupt or 'the first man to lead those who know'. He wasn't a leader in the sense that he was a subordinate, but he did lead those below him in a crusade of transforming human society that still continues today. The theme was confirmed as early as 1785 when a horseman courier called Lanz was reported to be struck by lightning and extensive Illuminati documents were found in his saddlebags. They made the link to Weishaupt and detailed the plan for world takeover. Current events with 'Covid' fascism have been in the making for a very long time. Jacob Frank was jailed for 13 years by the Catholic Inquisition after his arrest in 1760 and on his release he headed for Frankfurt, Germany, home city and headquarters of the House of Rothschild where the alliance was struck with Mayer Amschel Rothschild and Weishaupt. Rothschild arranged for Frank to be given the title of Baron and he became a wealthy nobleman with a big following of Jews in Germany, the Austro-Hungarian Empire and other European countries. Most of them would have believed he was on their side.

The name 'Illuminati' came from the Zohar which is a body of works in the Jewish mystical 'bible' called the Kabbalah. 'Zohar' is the foundation of Sabbatian-Frankist belief and in Hebrew 'Zohar' means 'splendour', 'radiance', 'illuminated', and so we have 'Illuminati'. They claim to be the 'Illuminated Ones' from their knowledge systematically hidden from the human population and passed on through generations of carefully-chosen initiates in the global secret society network or Cult. Hidden knowledge includes an awareness of the Cult agenda for the world and the nature of our collective reality that I will explore later. Cult 'illumination' is symbolised by the torch held by the Statue of Liberty which was gifted to New York by French Freemasons in Paris who knew exactly what it represents. 'Liberty' symbolises the goddess worshipped in

Babylon as Queen Semiramis or Ishtar. The significance of this will become clear. Notice again the ubiquitous theme of inversion with the Statue of 'Liberty' really symbolising mass control (Fig 7). A mirror-image statute stands on an island in the River Seine in Paris from where New York Liberty originated (Fig 8). A large replica of the Liberty flame stands on top of the Pont de l'Alma tunnel in Paris where Princess Diana died in a Cult ritual described in *The Biggest Secret*. Lucifer 'the light bringer' is related to all this (and much more as we'll see) and 'Lucifer' is a central figure in Sabbatian-Frankism and its associated Satanism. Sabbatians reject the Jewish Torah, or Pentateuch, the 'five books of Moses' in the Old Testament known as Genesis, Exodus, Leviticus, Numbers, and Deuteronomy which are claimed by Judaism and Christianity to have been dictated by 'God' to Moses on Mount Sinai. Sabbatians say these do not apply to them and they seek to replace them with the Zohar to absorb Judaism and its followers into their inversion which is an expression of a much greater global inversion. They want to delete all religions and force humanity to worship a one-world religion – Sabbatian Satanism that also includes worship of the Earth goddess. Satanic themes are being more and more introduced into mainstream society and while Christianity is currently the foremost target for destruction the others are planned to follow.



Figure 7: The Cult goddess of Babylon disguised as the Statue of Liberty holding the flame of Lucifer the 'light bringer'.



Figure 8: Liberty's mirror image in Paris where the New York version originated.

Marx brothers

Rabbi Marvin Antelman connects the Illuminati to the Jacobins in *To Eliminate the Opiate* and Jacobins were the force behind the French Revolution. He links both to the Bund der Gerechten, or League of the Just, which was the network that inflicted communism/Marxism on the world. Antelman wrote:

The original inner circle of the Bund der Gerechten consisted of born Catholics, Protestants and Jews [Sabbatian-Frankist infiltrators], and those representatives of respective subdivisions formulated schemes for the ultimate destruction of their faiths. The heretical Catholics laid plans which they felt would take a century or more for the ultimate destruction of the church; the apostate Jews for the ultimate destruction of the Jewish religion.

Sabbatian-created communism connects into this anti-religion agenda in that communism does not allow for the free practice of religion. The Sabbatian 'Bund' became the International Communist Party and Communist League and in 1848 'Marxism' was born with the Communist Manifesto of Sabbatian assets Karl Marx and Friedrich Engels. It is absolutely no coincidence that Marxism, just a different name for fascist and other centrally-controlled tyrannies, is being imposed worldwide as a result of the 'Covid' hoax and nor that Marxist/fascist China was the place where the hoax originated. The reason for this will become very clear in the chapter 'Covid: The calculated catastrophe'. The so-called 'Woke' mentality has hijacked

traditional beliefs of the political left and replaced them with far-right make-believe 'social justice' better known as Marxism. Woke will, however, be swallowed by its own perceived 'revolution' which is really the work of billionaires and billionaire corporations feigning being 'Woke'. Marxism is being touted by Wokers as a replacement for 'capitalism' when we don't have 'capitalism'. We have cartelism in which the market is stitched up by the very Cult billionaires and corporations bankrolling Woke. Billionaires love Marxism which keeps the people in servitude while they control from the top. Terminally naïve Wokers think they are 'changing the world' when it's the Cult that is doing the changing and when they have played their vital part and become surplus to requirements they, too, will be targeted. The Illuminati-Jacobins were behind the period known as 'The Terror' in the French Revolution in 1793 and 1794 when Jacobin Maximillian de Robespierre and his Orwellian 'Committee of Public Safety' killed 17,000 'enemies of the Revolution' who had once been 'friends of the Revolution'. Karl Marx (1818-1883), whose Sabbatian creed of Marxism has cost the lives of at least 100 million people, is a hero once again to Wokers who have been systematically kept ignorant of real history by their 'education' programming. As a result they now promote a Sabbatian 'Marxist' abomination destined at some point to consume them. Rabbi Antelman, who spent decades researching the Sabbatian plot, said of the League of the Just and Karl Marx:

Contrary to popular opinion Karl Marx did not originate the Communist Manifesto. He was paid for his services by the League of the Just, which was known in its country of origin, Germany, as the Bund der Geächteten.

Antelman said the text attributed to Marx was the work of other people and Marx 'was only repeating what others already said'. Marx was 'a hired hack – lackey of the wealthy Illuminists'. Marx famously said that religion was the 'opium of the people' (part of the Sabbatian plan to demonise religion) and Antelman called his books, *To Eliminate the Opiate*. Marx was born Jewish, but his family converted to Christianity (Sabbatian modus operandi) and he

attacked Jews, not least in his book, *A World Without Jews*. In doing so he supported the Sabbatian plan to destroy traditional Jewishness and Judaism which we are clearly seeing today with the vindictive targeting of orthodox Jews by the Sabbatian government of Israel over 'Covid' laws. I don't follow any religion and it has done much damage to the world over centuries and acted as a perceptual straightjacket. Renegade Minds, however, are always asking *why* something is being done. It doesn't matter if they agree or disagree with what is happening – *why* is it happening is the question. The 'why?' can be answered with regard to religion in that religions create interacting communities of believers when the Cult wants to dismantle all discourse, unity and interaction (see 'Covid' lockdowns) and the ultimate goal is to delete all religions for a one-world religion of Cult Satanism worshipping their 'god' of which more later. We see the same 'why?' with gun control in America. I don't have guns and don't want them, but why is the Cult seeking to disarm the population at the same time that law enforcement agencies are armed to their molars and why has every tyrant in history sought to disarm people before launching the final takeover? They include Hitler, Stalin, Pol Pot and Mao who followed confiscation with violent seizing of power. You know it's a Cult agenda by the people who immediately race to the microphones to exploit dead people in multiple shootings. Ultra-Zionist Cult lackey Senator Chuck Schumer was straight on the case after ten people were killed in Boulder, Colorado in March, 2121. Simple rule ... if Schumer wants it the Cult wants it and the same with his ultra-Zionist mate the wild-eyed Senator Adam Schiff. At the same time they were calling for the disarmament of Americans, many of whom live a long way from a police response, Schumer, Schiff and the rest of these pampered clowns were sitting on Capitol Hill behind a razor-wired security fence protected by thousands of armed troops in addition to their own armed bodyguards. Mom and pop in an isolated home? They're just potential mass shooters.

Zion Mainframe

Sabbatian-Frankists and most importantly the Rothschilds were behind the creation of 'Zionism', a political movement that demanded a Jewish homeland in Israel as promised by Sabbatai Zevi. The very symbol of Israel comes from the German meaning of the name Rothschild. Dynasty founder Mayer Amschel Rothschild changed the family name from Bauer to Rothschild, or 'Red-Shield' in German, in deference to the six-pointed 'Star of David' hexagram displayed on the family's home in Frankfurt. The symbol later appeared on the flag of Israel after the Rothschilds were centrally involved in its creation. Hexagrams are not a uniquely Jewish symbol and are widely used in occult ('hidden') networks often as a symbol for Saturn (see my other books for why). Neither are Zionism and Jewishness interchangeable. Zionism is a political movement and philosophy and not a 'race' or a people. Many Jews oppose Zionism and many non-Jews, including US President Joe Biden, call themselves Zionists as does Israel-centric Donald Trump. America's support for the Israel government is pretty much a gimme with ultra-Zionist billionaires and corporations providing fantastic and dominant funding for both political parties. Former Congresswoman Cynthia McKinney has told how she was approached immediately she ran for office to 'sign the pledge' to Israel and confirm that she would always vote in that country's best interests. All American politicians are approached in this way. Anyone who refuses will get no support or funding from the enormous and all-powerful Zionist lobby that includes organisations like mega-lobby group AIPAC, the American Israel Public Affairs Committee. Trump's biggest funder was ultra-Zionist casino and media billionaire Sheldon Adelson while major funders of the Democratic Party include ultra-Zionist George Soros and ultra-Zionist financial and media mogul, Haim Saban. Some may reel back at the suggestion that Soros is an Israel-firster (Sabbatian-controlled Israel-firster), but Renegade Minds watch the actions not the words and everywhere Soros donates his billions the Sabbatian agenda benefits. In the spirit of Sabbatian inversion Soros pledged \$1 billion for a new university network to promote 'liberal values and tackle intolerance'. He made the announcement during his annual speech

at the Cult-owned World Economic Forum in Davos, Switzerland, in January, 2020, after his 'harsh criticism' of 'authoritarian rulers' around the world. You can only laugh at such brazen mendacity. How *he* doesn't laugh is the mystery. Translated from the Orwellian 'liberal values and tackle intolerance' means teaching non-white people to hate white people and for white people to loathe themselves for being born white. The reason for that will become clear.

The 'Anti-Semitism' fraud

Zionists support the Jewish homeland in the land of Palestine which has been the Sabbatian-Rothschild goal for so long, but not for the benefit of Jews. Sabbatians and their global Anti-Semitism Industry have skewed public and political opinion to equate opposing the violent extremes of Zionism to be a blanket attack and condemnation of all Jewish people. Sabbatians and their global Anti-Semitism Industry have skewed public and political opinion to equate opposing the violent extremes of Zionism to be a blanket attack and condemnation of all Jewish people. This is nothing more than a Sabbatian protection racket to stop legitimate investigation and exposure of their agendas and activities. The official definition of 'anti-Semitism' has more recently been expanded to include criticism of Zionism – a *political movement* – and this was done to further stop exposure of Sabbatian infiltrators who created Zionism as we know it today in the 19th century. Renegade Minds will talk about these subjects when they know the shit that will come their way. People must decide if they want to know the truth or just cower in the corner in fear of what others will say. Sabbatians have been trying to label me as 'anti-Semitic' since the 1990s as I have uncovered more and more about their background and agendas. Useless, gutless, fraudulent 'journalists' then just repeat the smears without question and on the day I was writing this section a pair of unquestioning repeaters called Ben Quinn and Archie Bland (how appropriate) outright called me an 'anti-Semite' in the establishment propaganda sheet, the London *Guardian*, with no supporting evidence. The

Sabbatian Anti-Semitism Industry said so and who are they to question that? They wouldn't dare. Ironically 'Semitic' refers to a group of languages in the Middle East that are almost entirely Arabic. 'Anti-Semitism' becomes 'anti-Arab' which if the consequences of this misunderstanding were not so grave would be hilarious. Don't bother telling Quinn and Bland. I don't want to confuse them, bless 'em. One reason I am dubbed 'anti-Semitic' is that I wrote in the 1990s that Jewish operatives (Sabbatians) were heavily involved in the Russian Revolution when Sabbatians overthrew the Romanov dynasty. This apparently made me 'anti-Semitic'. Oh, really? Here is a section from *The Trigger*:

British journalist Robert Wilton confirmed these themes in his 1920 book *The Last Days of the Romanovs* when he studied official documents from the Russian government to identify the members of the Bolshevik ruling elite between 1917 and 1919. The Central Committee included 41 Jews among 62 members; the Council of the People's Commissars had 17 Jews out of 22 members; and 458 of the 556 most important Bolshevik positions between 1918 and 1919 were occupied by Jewish people. Only 17 were Russian. Then there were the 23 Jews among the 36 members of the vicious Cheka Soviet secret police established in 1917 who would soon appear all across the country.

Professor Robert Service of Oxford University, an expert on 20th century Russian history, found evidence that ['Jewish'] Leon Trotsky had sought to make sure that Jews were enrolled in the Red Army and were disproportionately represented in the Soviet civil bureaucracy that included the Cheka which performed mass arrests, imprisonment and executions of 'enemies of the people'. A US State Department Decimal File (861.00/5339) dated November 13th, 1918, names [Rothschild banking agent in America] Jacob Schiff and a list of ultra-Zionists as funders of the Russian Revolution leading to claims of a 'Jewish plot', but the key point missed by all is they were not 'Jews' – they were Sabbatian-Frankists.

Britain's Winston Churchill made the same error by mistake or otherwise. He wrote in a 1920 edition of the *Illustrated Sunday Herald* that those behind the Russian revolution were part of a 'worldwide conspiracy for the overthrow of civilisation and for the reconstitution of society on the basis of arrested development, of envious malevolence, and impossible equality' (see 'Woke' today because that has been created by the same network). Churchill said there was no need to exaggerate the part played in the creation of Bolshevism and in the actual bringing about of the Russian

Revolution 'by these international and for the most part atheistical Jews' ['atheistical Jews' = Sabbatians]. Churchill said it is certainly a very great one and probably outweighs all others: 'With the notable exception of Lenin, the majority of the leading figures are Jews.' He went on to describe, knowingly or not, the Sabbatian modus operandi of placing puppet leaders nominally in power while they control from the background:

Moreover, the principal inspiration and driving power comes from the Jewish leaders. Thus Tchitcherin, a pure Russian, is eclipsed by his nominal subordinate, Litvinoff, and the influence of Russians like Bukharin or Lunacharski cannot be compared with the power of Trotsky, or of Zinovieff, the Dictator of the Red Citadel (Petrograd), or of Krassin or Radek – all Jews. In the Soviet institutions the predominance of Jews is even more astonishing. And the prominent, if not indeed the principal, part in the system of terrorism applied by the Extraordinary Commissions for Combatting Counter-Revolution has been taken by Jews, and in some notable cases by Jewesses.

What I said about seriously disproportionate involvement in the Russian Revolution by Jewish 'revolutionaries' (Sabbatians) is provable fact, but truth is no defence against the Sabbatian Anti-Semitism Industry, its repeater parrots like Quinn and Bland, and the now breathtaking network of so-called 'Woke' 'anti-hate' groups with interlocking leaderships and funding which have the role of discrediting and silencing anyone who gets too close to exposing the Sabbatians. We have seen 'truth is no defence' confirmed in legal judgements with the Saskatchewan Human Rights Commission in Canada decreeing this: 'Truthful statements can be presented in a manner that would meet the definition of hate speech, and not all truthful statements must be free from restriction.' Most 'anti-hate' activists, who are themselves consumed by hatred, are too stupid and ignorant of the world to know how they are being used. They are far too far up their own virtue-signalling arses and it's far too dark for them to see anything.

The 'revolution' game

The background and methods of the 'Russian' Revolution are straight from the Sabbatian playbook seen in the French Revolution

and endless others around the world that appear to start as a revolution of the people against tyrannical rule and end up with a regime change to more tyrannical rule overtly or covertly. Wars, terror attacks and regime overthrows follow the Sabbatian cult through history with its agents creating them as Problem-Reaction-Solutions to remove opposition on the road to world domination. Sabbatian dots connect the Rothschilds with the Illuminati, Jacobins of the French Revolution, the 'Bund' or League of the Just, the International Communist Party, Communist League and the Communist Manifesto of Karl Marx and Friedrich Engels that would lead to the Rothschild-funded Russian Revolution. The sequence comes under the heading of 'creative destruction' when you advance to your global goal by continually destroying the status quo to install a new status quo which you then also destroy. The two world wars come to mind. With each new status quo you move closer to your planned outcome. Wars and mass murder are to Sabbatians a collective blood sacrifice ritual. They are obsessed with death for many reasons and one is that death is an inversion of life. Satanists and Sabbatians are obsessed with death and often target churches and churchyards for their rituals. Inversion-obsessed Sabbatians explain the use of inverted symbolism including the *inverted* pentagram and *inverted* cross. The inversion of the cross has been related to targeting Christianity, but the cross was a religious symbol long before Christianity and its inversion is a statement about the Sabbatian mentality and goals more than any single religion.

Sabbatians operating in Germany were behind the rise of the occult-obsessed Nazis and the subsequent Jewish exodus from Germany and Europe to Palestine and the United States after World War Two. The Rothschild dynasty was at the forefront of this both as political manipulators and by funding the operation. Why would Sabbatians help to orchestrate the horrors inflicted on Jews by the Nazis and by Stalin after they organised the Russian Revolution? Sabbatians hate Jews and their religion, that's why. They pose as Jews and secure positions of control within Jewish society and play the 'anti-Semitism' card to protect themselves from exposure

through a global network of organisations answering to the Sabbatian-created-and-controlled globe-spanning intelligence network that involves a stunning web of military-intelligence operatives and operations for a tiny country of just nine million. Among them are Jewish assets who are not Sabbatians but have been convinced by them that what they are doing is for the good of Israel and the Jewish community to protect them from what they have been programmed since childhood to believe is a Jew-hating hostile world. The Jewish community is just a highly convenient cover to hide the true nature of Sabbatians. Anyone getting close to exposing their game is accused by Sabbatian place-people and gofers of 'anti-Semitism' and claiming that all Jews are part of a plot to take over the world. I am not saying that. I am saying that Sabbatians – the *real* Jew-haters – have infiltrated the Jewish community to use them both as a cover and an 'anti-Semitic' defence against exposure. Thus we have the Anti-Semitism Industry targeted researchers in this way and most Jewish people think this is justified and genuine. They don't know that their 'Jewish' leaders and institutions of state, intelligence and military are not controlled by Jews at all, but cultists and stooges of Sabbatian-Frankism. I once added my name to a pro-Jewish freedom petition online and the next time I looked my name was gone and text had been added to the petition blurb to attack me as an 'anti-Semite' such is the scale of perceptual programming.

Moving on America

I tell the story in *The Trigger* and a chapter called 'Atlantic Crossing' how particularly after Israel was established the Sabbatians moved in on the United States and eventually grasped control of government administration, the political system via both Democrats and Republicans, the intelligence community like the CIA and National Security Agency (NSA), the Pentagon and mass media. Through this seriously compartmentalised network Sabbatians and their operatives in Mossad, Israeli Defense Forces (IDF) and US agencies pulled off 9/11 and blamed it on 19 'Al-Qaeda hijackers' dominated by men from, or connected to, Sabbatian-ruled Saudi

Arabia. The '19' were not even on the planes let alone flew those big passenger jets into buildings while being largely incompetent at piloting one-engine light aircraft. 'Hijacker' Hani Hanjour who is said to have flown American Airlines Flight 77 into the Pentagon with a turn and manoeuvre most professional pilots said they would have struggled to do was banned from renting a small plane by instructors at the Freeway Airport in Bowie, Maryland, just *six weeks* earlier on the grounds that he was an incompetent pilot. The Jewish population of the world is just 0.2 percent with even that almost entirely concentrated in Israel (75 percent Jewish) and the United States (around two percent). This two percent and globally 0.2 percent refers to *Jewish* people and not Sabbatian interlopers who are a fraction of that fraction. What a sobering thought when you think of the fantastic influence on world affairs of tiny Israel and that the Project for the New America Century (PNAC) which laid out the blueprint in September, 2000, for America's war on terror and regime change wars in Iraq, Libya and Syria was founded and dominated by Sabbatians known as 'Neocons'. The document conceded that this plan would not be supported politically or publicly without a major attack on American soil and a Problem-Reaction-Solution excuse to send troops to war across the Middle East. Sabbatian Neocons said:

... [The] process of transformation ... [war and regime change] ... is likely to be a long one, absent some catastrophic and catalysing event – like a new Pearl Harbor.

Four months later many of those who produced that document came to power with their inane puppet George Bush from the long-time Sabbatian Bush family. They included Sabbatian Dick Cheney who was officially vice-president, but really de-facto president for the entirety of the 'Bush' government. Nine months after the 'Bush' inauguration came what Bush called at the time 'the Pearl Harbor of the 21st century' and with typical Sabbatian timing and symbolism 2001 was the 60th anniversary of the attack in 1941 by the Japanese Air Force on Pearl Harbor, Hawaii, which allowed President Franklin Delano Roosevelt to take the United States into a Sabbatian-

instigated Second World War that he said in his election campaign that he never would. The evidence is overwhelming that Roosevelt and his military and intelligence networks knew the attack was coming and did nothing to stop it, but they did make sure that America's most essential naval ships were not in Hawaii at the time. Three thousand Americans died in the Pearl Harbor attacks as they did on September 11th. By the 9/11 year of 2001 Sabbatians had widely infiltrated the US government, military and intelligence operations and used their compartmentalised assets to pull off the 'Al-Qaeda' attacks. If you read *The Trigger* it will blow your mind to see the utterly staggering concentration of 'Jewish' operatives (Sabbatian infiltrators) in essential positions of political, security, legal, law enforcement, financial and business power before, during, and after the attacks to make them happen, carry them out, and then cover their tracks – and I do mean *staggering* when you think of that 0.2 percent of the world population and two percent of Americans which are Jewish while Sabbatian infiltrators are a fraction of that. A central foundation of the 9/11 conspiracy was the hijacking of government, military, Air Force and intelligence computer systems in real time through 'back-door' access made possible by Israeli (Sabbatian) 'cyber security' software. Sabbatian-controlled Israel is on the way to rivalling Silicon Valley for domination of cyberspace and is becoming the dominant force in cyber-security which gives them access to entire computer systems and their passcodes across the world. Then add to this that Zionists head (officially) Silicon Valley giants like Google (Larry Page and Sergey Brin), Google-owned YouTube (Susan Wojcicki), Facebook (Mark Zuckerberg and Sheryl Sandberg), and Apple (Chairman Arthur D. Levinson), and that ultra-Zionist hedge fund billionaire Paul Singer has a \$1 billion stake in Twitter which is only nominally headed by 'CEO' pothead Jack Dorsey. As cable news host Tucker Carlson said of Dorsey: 'There used to be debate in the medical community whether dropping a ton of acid had permanent effects and I think that debate has now ended.' Carlson made the comment after Dorsey told a hearing on Capitol Hill (if you cut through his bullshit) that he

believed in free speech so long as he got to decide what you can hear and see. These 'big names' of Silicon Valley are only front men and women for the Global Cult, not least the Sabbatians, who are the true controllers of these corporations. Does anyone still wonder why these same people and companies have been ferociously censoring and banning people (like me) for exposing any aspect of the Cult agenda and especially the truth about the 'Covid' hoax which Sabbatians have orchestrated?

The Jeffrey Epstein paedophile ring was a Sabbatian operation. He was officially 'Jewish' but he was a Sabbatian and women abused by the ring have told me about the high number of 'Jewish' people involved. The Epstein horror has Sabbatian written all over it and matches perfectly their modus operandi and obsession with sex and ritual. Epstein was running a Sabbatian blackmail ring in which famous people with political and other influence were provided with young girls for sex while everything was being filmed and recorded on hidden cameras and microphones at his New York house, Caribbean island and other properties. Epstein survivors have described this surveillance system to me and some have gone public. Once the famous politician or other figure knew he or she was on video they tended to do whatever they were told. Here we go again ...when you've got them by the balls their hearts and minds will follow. Sabbatians use this blackmail technique on a wide scale across the world to entrap politicians and others they need to act as demanded. Epstein's private plane, the infamous 'Lolita Express', had many well-known passengers including Bill Clinton while Bill Gates has flown on an Epstein plane and met with him four years after Epstein had been jailed for paedophilia. They subsequently met many times at Epstein's home in New York according to a witness who was there. Epstein's infamous side-kick was Ghislaine Maxwell, daughter of Mossad agent and ultra-Zionist mega-crooked British businessman, Bob Maxwell, who at one time owned the *Daily Mirror* newspaper. Maxwell was murdered at sea on his boat in 1991 by Sabbatian-controlled Mossad when he became a liability with his

business empire collapsing as a former Mossad operative has confirmed (see *The Trigger*).

Money, money, money, funny money ...

Before I come to the Sabbatian connection with the last three US presidents I will lay out the crucial importance to Sabbatians of controlling banking and finance. Sabbatian Mayer Amschel Rothschild set out to dominate this arena in his family's quest for total global control. What is freedom? It is, in effect, choice. The more choices you have the freer you are and the fewer your choices the more you are enslaved. In the global structure created over centuries by Sabbatians the biggest decider and restrictor of choice is ... money. Across the world if you ask people what they would like to do with their lives and why they are not doing that they will reply 'I don't have the money'. This is the idea. A global elite of multi-billionaires are described as 'greedy' and that is true on one level; but control of money – who has it and who doesn't – is not primarily about greed. It's about control. Sabbatians have seized ever more control of finance and sucked the wealth of the world out of the hands of the population. We talk now, after all, about the 'One-percent' and even then the wealthiest are a lot fewer even than that. This has been made possible by a money scam so outrageous and so vast it could rightly be called the scam of scams founded on creating 'money' out of nothing and 'loaning' that with interest to the population. Money out of nothing is called 'credit'. Sabbatians have asserted control over governments and banking ever more completely through the centuries and secured financial laws that allow banks to lend hugely more than they have on deposit in a confidence trick known as fractional reserve lending. Imagine if you could lend money that doesn't exist and charge the recipient interest for doing so. You would end up in jail. Bankers by contrast end up in mansions, private jets, Malibu and Monaco.

Banks are only required to keep a fraction of their deposits and wealth in their vaults and they are allowed to lend 'money' they don't have called 'credit'. Go into a bank for a loan and if you succeed

the banker will not move any real wealth into your account. They will type into your account the amount of the agreed 'loan' – say £100,000. This is not wealth that really exists; it is non-existent, fresh-air, created-out-of-nothing 'credit' which has never, does not, and will never exist except in theory. Credit is backed by nothing except wind and only has buying power because people think that it has buying power and accept it in return for property, goods and services. I have described this situation as like those cartoon characters you see chasing each other and when they run over the edge of a cliff they keep running forward on fresh air until one of them looks down, realises what's happened, and they all crash into the ravine. The whole foundation of the Sabbatian financial system is to stop people looking down except for periodic moments when they want to crash the system (as in 2008 and 2020 ongoing) and reap the rewards from all the property, businesses and wealth their borrowers had signed over as 'collateral' in return for a 'loan' of fresh air. Most people think that money is somehow created by governments when it comes into existence from the start as a debt through banks 'lending' illusory money called credit. Yes, the very currency of exchange is a *debt* from day one issued as an interest-bearing loan. Why don't governments create money interest-free and lend it to their people interest-free? Governments are controlled by Sabbatians and the financial system is controlled by Sabbatians for whom interest-free money would be a nightmare come true. Sabbatians underpin their financial domination through their global network of central banks, including the privately-owned US Federal Reserve and Britain's Bank of England, and this is orchestrated by a privately-owned central bank coordination body called the Bank for International Settlements in Basle, Switzerland, created by the usual suspects including the Rockefellers and Rothschilds. Central bank chiefs don't answer to governments or the people. They answer to the Bank for International Settlements or, in other words, the Global Cult which is dominated today by Sabbatians.

Built-in disaster

There are so many constituent scams within the overall banking scam. When you take out a loan of thin-air credit only the amount of that loan is theoretically brought into circulation to add to the amount in circulation; but you are paying back the principle plus interest. The additional interest is not created and this means that with every 'loan' there is a shortfall in the money in circulation between what is borrowed and what has to be paid back. There is never even close to enough money in circulation to repay all outstanding public and private debt including interest. Coldly weaved in the very fabric of the system is the certainty that some will lose their homes, businesses and possessions to the banking 'lender'. This is less obvious in times of 'boom' when the amount of money in circulation (and the debt) is expanding through more people wanting and getting loans. When a downturn comes and the money supply contracts it becomes painfully obvious that there is not enough money to service all debt and interest. This is less obvious in times of 'boom' when the amount of money in circulation (and the debt) is expanding through more people wanting and getting loans. When a downturn comes and the money supply contracts and it becomes painfully obvious – as in 2008 and currently – that there is not enough money to service all debt and interest. Sabbatian banksters have been leading the human population through a calculated series of booms (more debt incurred) and busts (when the debt can't be repaid and the banks get the debtor's tangible wealth in exchange for non-existent 'credit'). With each 'bust' Sabbatian bankers have absorbed more of the world's tangible wealth and we end up with the One-percent. Governments are in bankruptcy levels of debt to the same system and are therefore owned by a system they do not control. The Federal Reserve, 'America's central bank', is privately-owned and American presidents only nominally appoint its chairman or woman to maintain the illusion that it's an arm of government. It's not. The 'Fed' is a cartel of private banks which handed billions to its associates and friends after the crash of 2008 and has been Sabbatian-controlled since it was manipulated into being in 1913 through the covert trickery of Rothschild banking agents Jacob Schiff and Paul

Warburg, and the Sabbatian Rockefeller family. Somehow from a Jewish population of two-percent and globally 0.2 percent (Sabbatian interlopers remember are far smaller) ultra-Zionists headed the Federal Reserve for 31 years between 1987 and 2018 in the form of Alan Greenspan, Bernard Bernanke and Janet Yellen (now Biden's Treasury Secretary) with Yellen's deputy chairman a Israeli-American dual citizen and ultra-Zionist Stanley Fischer, a former governor of the Bank of Israel. Ultra-Zionist Fed chiefs spanned the presidencies of Ronald Reagan ('Republican'), Father George Bush ('Republican'), Bill Clinton ('Democrat'), Boy George Bush ('Republican') and Barack Obama ('Democrat'). We should really add the pre-Greenspan chairman, Paul Adolph Volcker, 'appointed' by Jimmy Carter ('Democrat') who ran the Fed between 1979 and 1987 during the Carter and Reagan administrations before Greenspan took over. Volcker was a long-time associate and business partner of the Rothschilds. No matter what the 'party' officially in power the United States economy was directed by the same force. Here are members of the Obama, Trump and Biden administrations and see if you can make out a common theme.

Barack Obama ('Democrat')

Ultra-Zionists Robert Rubin, Larry Summers, and Timothy Geithner ran the US Treasury in the Clinton administration and two of them reappeared with Obama. Ultra-Zionist Fed chairman Alan Greenspan had manipulated the crash of 2008 through deregulation and jumped ship just before the disaster to make way for ultra-Zionist Bernard Bernanke to hand out trillions to Sabbatian 'too big to fail' banks and businesses, including the ubiquitous ultra-Zionist Goldman Sachs which has an ongoing staff revolving door operation between itself and major financial positions in government worldwide. Obama inherited the fallout of the crash when he took office in January, 2009, and fortunately he had the support of his ultra-Zionist White House Chief of Staff Rahm Emmanuel, son of a terrorist who helped to bomb Israel into being in 1948, and his ultra-Zionist senior adviser David Axelrod, chief strategist in Obama's two

successful presidential campaigns. Emmanuel, later mayor of Chicago and former senior fundraiser and strategist for Bill Clinton, is an example of the Sabbatian policy after Israel was established of migrating insider families to America so their children would be born American citizens. 'Obama' chose this financial team throughout his administration to respond to the Sabbatian-instigated crisis:

Timothy Geithner (ultra-Zionist) Treasury Secretary; Jacob J. Lew, Treasury Secretary; Larry Summers (ultra-Zionist), director of the White House National Economic Council; Paul Adolph Volcker (Rothschild business partner), chairman of the Economic Recovery Advisory Board; Peter Orszag (ultra-Zionist), director of the Office of Management and Budget overseeing all government spending; Penny Pritzker (ultra-Zionist), Commerce Secretary; Jared Bernstein (ultra-Zionist), chief economist and economic policy adviser to Vice President Joe Biden; Mary Schapiro (ultra-Zionist), chair of the Securities and Exchange Commission (SEC); Gary Gensler (ultra-Zionist), chairman of the Commodity Futures Trading Commission (CFTC); Sheila Bair (ultra-Zionist), chair of the Federal Deposit Insurance Corporation (FDIC); Karen Mills (ultra-Zionist), head of the Small Business Administration (SBA); Kenneth Feinberg (ultra-Zionist), Special Master for Executive [bail-out] Compensation. Feinberg would be appointed to oversee compensation (with strings) to 9/11 victims and families in a campaign to stop them having their day in court to question the official story. At the same time ultra-Zionist Bernard Bernanke was chairman of the Federal Reserve and these are only some of the ultra-Zionists with allegiance to Sabbatian-controlled Israel in the Obama government. Obama's biggest corporate donor was ultra-Zionist Goldman Sachs which had employed many in his administration.

Donald Trump ('Republican')

Trump claimed to be an outsider (he wasn't) who had come to 'drain the swamp'. He embarked on this goal by immediately appointing ultra-Zionist Steve Mnuchin, a Goldman Sachs employee for 17

years, as his Treasury Secretary. Others included Gary Cohn (ultra-Zionist), chief operating officer of Goldman Sachs, his first Director of the National Economic Council and chief economic adviser, who was later replaced by Larry Kudlow (ultra-Zionist). Trump's senior adviser throughout his four years in the White House was his sinister son-in-law Jared Kushner, a life-long friend of Israel Prime Minister Benjamin Netanyahu. Kushner is the son of a convicted crook who was pardoned by Trump in his last days in office. Other ultra-Zionists in the Trump administration included: Stephen Miller, Senior Policy Adviser; Avrahm Berkowitz, Deputy Adviser to Trump and his Senior Adviser Jared Kushner; Ivanka Trump, Adviser to the President, who converted to Judaism when she married Jared Kushner; David Friedman, Trump lawyer and Ambassador to Israel; Jason Greenblatt, Trump Organization executive vice president and chief legal officer, who was made Special Representative for International Negotiations and the Israeli-Palestinian Conflict; Rod Rosenstein, Deputy Attorney General; Elliot Abrams, Special Representative for Venezuela, then Iran; John Eisenberg, National Security Council Legal Adviser and Deputy Council to the President for National Security Affairs; Anne Neuberger, Deputy National Manager, National Security Agency; Ezra Cohen-Watnick, Acting Under Secretary of Defense for Intelligence; Elan Carr, Special Envoy to monitor and combat anti-Semitism; Len Khodorkovsky, Deputy Special Envoy to monitor and combat anti-Semitism; Reed Cordish, Assistant to the President, Intragovernmental and Technology Initiatives. Trump Vice President Mike Pence and Secretary of State Mike Pompeo, both Christian Zionists, were also vehement supporters of Israel and its goals and ambitions.

Donald 'free-speech believer' Trump pardoned a number of financial and violent criminals while ignoring calls to pardon Julian Assange and Edward Snowden whose crimes are revealing highly relevant information about government manipulation and corruption and the widespread illegal surveillance of the American people by US 'security' agencies. It's so good to know that Trump is on the side of freedom and justice and not mega-criminals with

allegiance to Sabbatian-controlled Israel. These included a pardon for Israeli spy Jonathan Pollard who was jailed for life in 1987 under the Espionage Act. Aviem Sella, the Mossad agent who recruited Pollard, was also pardoned by Trump while Assange sat in jail and Snowden remained in exile in Russia. Sella had 'fled' (was helped to escape) to Israel in 1987 and was never extradited despite being charged under the Espionage Act. A Trump White House statement said that Sella's clemency had been 'supported by Benjamin Netanyahu, Ron Dermer, Israel's US Ambassador, David Friedman, US Ambassador to Israel and Miriam Adelson, wife of leading Trump donor Sheldon Adelson who died shortly before. Other friends of Jared Kushner were pardoned along with Sholom Weiss who was believed to be serving the longest-ever white-collar prison sentence of more than 800 years in 2000. The sentence was commuted of Ponzi-schemer Eliyahu Weinstein who defrauded Jews and others out of \$200 million. I did mention that Assange and Snowden were ignored, right? Trump gave Sabbatians almost everything they asked for in military and political support, moving the US Embassy from Tel Aviv to Jerusalem with its critical symbolic and literal implications for Palestinian statehood, and the 'deal of the Century' designed by Jared Kushner and David Friedman which gave the Sabbatian Israeli government the green light to substantially expand its already widespread program of building illegal Jewish-only settlements in the occupied land of the West Bank. This made a two-state 'solution' impossible by seizing all the land of a potential Palestinian homeland and that had been the plan since 1948 and then 1967 when the Arab-controlled Gaza Strip, West Bank, Sinai Peninsula and Syrian Golan Heights were occupied by Israel. All the talks about talks and road maps and delays have been buying time until the West Bank was physically occupied by Israeli real estate. Trump would have to be a monumentally ill-informed idiot not to see that this was the plan he was helping to complete. The Trump administration was in so many ways the Kushner administration which means the Netanyahu administration which means the Sabbatian administration. I understand why many opposing Cult fascism in all its forms gravitated to Trump, but he

was a crucial part of the Sabbatian plan and I will deal with this in the next chapter.

Joe Biden ('Democrat')

A barely cognitive Joe Biden took over the presidency in January, 2021, along with his fellow empty shell, Vice-President Kamala Harris, as the latest Sabbatian gofers to enter the White House. Names on the door may have changed and the 'party' – the force behind them remained the same as Zionists were appointed to a stream of pivotal areas relating to Sabbatian plans and policy. They included: Janet Yellen, Treasury Secretary, former head of the Federal Reserve, and still another ultra-Zionist running the US Treasury after Mnuchin (Trump), Lew and Geithner (Obama), and Summers and Rubin (Clinton); Anthony Blinken, Secretary of State; Wendy Sherman, Deputy Secretary of State (so that's 'Biden's' Sabbatian foreign policy sorted); Jeff Zients, White House coronavirus coordinator; Rochelle Walensky, head of the Centers for Disease Control; Rachel Levine, transgender deputy health secretary (that's 'Covid' hoax policy under control); Merrick Garland, Attorney General; Alejandro Mayorkas, Secretary of Homeland Security; Cass Sunstein, Homeland Security with responsibility for new immigration laws; Avril Haines, Director of National Intelligence; Anne Neuberger, National Security Agency cybersecurity director (note, cybersecurity); David Cohen, CIA Deputy Director; Ronald Klain, Biden's Chief of Staff (see Rahm Emanuel); Eric Lander, a 'leading geneticist', Office of Science and Technology Policy director (see Smart Grid, synthetic biology agenda); Jessica Rosenworcel, acting head of the Federal Communications Commission (FCC) which controls Smart Grid technology policy and electromagnetic communication systems including 5G. How can it be that so many pivotal positions are held by two-percent of the American population and 0.2 percent of the world population administration after administration no matter who is the president and what is the party? It's a coincidence? Of course it's not and this is why Sabbatians have built their colossal global web of interlocking 'anti-

hate' hate groups to condemn anyone who asks these glaring questions as an 'anti-Semite'. The way that Jewish people horrifically abused in Sabbatian-backed Nazi Germany are exploited to this end is stomach-turning and disgusting beyond words.

Political fusion

Sabbatian manipulation has reversed the roles of Republicans and Democrats and the same has happened in Britain with the Conservative and Labour Parties. Republicans and Conservatives were always labelled the 'right' and Democrats and Labour the 'left', but look at the policy positions now and the Democrat-Labour 'left' has moved further to the 'right' than Republicans and Conservatives under the banner of 'Woke', the Cult-created far-right tyranny. Where once the Democrat-Labour 'left' defended free speech and human rights they now seek to delete them and as I said earlier despite the 'Covid' fascism of the Jackboot Johnson Conservative government in the UK the Labour Party of leader Keir Starmer demanded even more extreme measures. The Labour Party has been very publicly absorbed by Sabbatians after a political and media onslaught against the previous leader, the weak and inept Jeremy Corbyn, over made-up allegations of 'anti-Semitism' both by him and his party. The plan was clear with this 'anti-Semite' propaganda and what was required in response was a swift and decisive 'fuck off' from Corbyn and a statement to expose the Anti-Semitism Industry (Sabbatian) attempt to silence Labour criticism of the Israeli government (Sabbatians) and purge the party of all dissent against the extremes of ultra-Zionism (Sabbatians). Instead Corbyn and his party fell to their knees and appeased the abusers which, by definition, is impossible. Appeasing one demand leads only to a new demand to be appeased until takeover is complete. Like I say – 'fuck off' would have been a much more effective policy and I have used it myself with great effect over the years when Sabbatians are on my case which is most of the time. I consider that fact a great compliment, by the way. The outcome of the Labour Party capitulation is that we now have a Sabbatian-controlled

Conservative Party 'opposed' by a Sabbatian-controlled Labour Party in a one-party Sabbatian state that hurtles towards the extremes of tyranny (the Sabbatian cult agenda). In America the situation is the same. Labour's Keir Starmer spends his days on his knees with his tongue out pointing to Tel Aviv, or I guess now Jerusalem, while Boris Johnson has an 'anti-Semitism czar' in the form of former Labour MP John Mann who keeps Starmer company on his prayer mat.

Sabbatian influence can be seen in Jewish members of the Labour Party who have been ejected for criticism of Israel including those from families that suffered in Nazi Germany. Sabbatians despise real Jewish people and target them even more harshly because it is so much more difficult to dub them 'anti-Semitic' although in their desperation they do try.

CHAPTER THREE

The Pushbacker sting

*Until you realize how easy it is for your mind to be manipulated, you
remain the puppet of someone else's game*

Evita Ochel

I will use the presidencies of Trump and Biden to show how the manipulation of the one-party state plays out behind the illusion of political choice across the world. No two presidencies could – on the face of it – be more different and apparently at odds in terms of direction and policy.

A Renegade Mind sees beyond the obvious and focuses on outcomes and consequences and not image, words and waffle. The Cult embarked on a campaign to divide America between those who blindly support its agenda (the mentality known as 'Woke') and those who are pushing back on where the Cult and its Sabbatians want to go. This presents infinite possibilities for dividing and ruling the population by setting them at war with each other and allows a perceptual ring fence of demonisation to encircle the Pushbackers in a modern version of the Little Big Horn in 1876 when American cavalry led by Lieutenant Colonel George Custer were drawn into a trap, surrounded and killed by Native American tribes defending their land of thousands of years from being seized by the government. In this modern version the roles are reversed and it's those defending themselves from the Sabbatian government who are surrounded and the government that's seeking to destroy them. This trap was set years ago and to explain how we must return to 2016

and the emergence of Donald Trump as a candidate to be President of the United States. He set out to overcome the best part of 20 other candidates in the Republican Party before and during the primaries and was not considered by many in those early stages to have a prayer of living in the White House. The Republican Party was said to have great reservations about Trump and yet somehow he won the nomination. When you know how American politics works – politics in general – there is no way that Trump could have become the party's candidate unless the Sabbatian-controlled 'Neocons' that run the Republican Party wanted that to happen. We saw the proof in emails and documents made public by WikiLeaks that the Democratic Party hierarchy, or Democons, systematically undermined the campaign of Bernie Sanders to make sure that Sabbatian gofer Hillary Clinton won the nomination to be their presidential candidate. If the Democons could do that then the Neocons in the Republican Party could have derailed Trump in the same way. But they didn't and at that stage I began to conclude that Trump could well be the one chosen to be president. If that was the case the 'why' was pretty clear to see – the goal of dividing America between Cult agenda-supporting Wokers and Pushbackers who gravitated to Trump because he was telling them what they wanted to hear. His constituency of support had been increasingly ignored and voiceless for decades and profoundly through the eight years of Sabbatian puppet Barack Obama. Now here was someone speaking their language of pulling back from the incessant globalisation of political and economic power, the exporting of American jobs to China and elsewhere by 'American' (Sabbatian) corporations, the deletion of free speech, and the mass immigration policies that had further devastated job opportunities for the urban working class of all races and the once American heartlands of the Midwest.

Beware the forked tongue

Those people collectively sighed with relief that at last a political leader was apparently on their side, but another trait of the Renegade Mind is that you look even harder at people telling you

what you want to hear than those who are telling you otherwise. Obviously as I said earlier people wish what they want to hear to be true and genuine and they are much more likely to believe that than someone saying what they don't want to hear and don't want to be true. Sales people are taught to be skilled in eliciting by calculated questioning what their customers want to hear and repeating that back to them as their own opinion to get their targets to like and trust them. Assets of the Cult are also sales people in the sense of selling perception. To read Cult manipulation you have to play the long and expanded game and not fall for the Vaudeville show of party politics. Both American parties are vehicles for the Cult and they exploit them in different ways depending on what the agenda requires at that moment. Trump and the Republicans were used to be the focus of dividing America and isolating Pushbackers to open the way for a Biden presidency to become the most extreme in American history by advancing the full-blown Woke (Cult) agenda with the aim of destroying and silencing Pushbackers now labelled Nazi Trump supporters and white supremacists.

Sabbatians wanted Trump in office for the reasons described by ultra-Zionist Saul Alinsky (1909-1972) who was promoting the Woke philosophy through 'community organising' long before anyone had heard of it. In those days it still went by its traditional name of Marxism. The reason for the manipulated Trump phenomenon was laid out in Alinsky's 1971 book, *Rules for Radicals*, which was his blueprint for overthrowing democratic and other regimes and replacing them with Sabbatian Marxism. Not surprisingly his to-do list was evident in the Sabbatian French and Russian 'Revolutions' and that in China which will become very relevant in the next chapter about the 'Covid' hoax. Among Alinsky's followers have been the deeply corrupt Barack Obama, House Speaker Nancy Pelosi and Hillary Clinton who described him as a 'hero'. All three are Sabbatian stooges with Pelosi personifying the arrogant corrupt idiocy that so widely fronts up for the Cult inner core. Predictably as a Sabbatian advocate of the 'light-bringer' Alinsky features Lucifer on the dedication page of his book as the original radical who gained

his own kingdom ('Earth' as we shall see). One of Alinsky's golden radical rules was to pick an individual and focus all attention, hatred and blame on them and not to target faceless bureaucracies and corporations. *Rules for Radicals* is really a Sabbatian handbook with its contents repeatedly employed all over the world for centuries and why wouldn't Sabbatians bring to power their designer-villain to be used as the individual on which all attention, hatred and blame was bestowed? This is what they did and the only question for me is how much Trump knew that and how much he was manipulated. A bit of both, I suspect. This was Alinsky's Trump technique from a man who died in 1972. The technique has spanned history:

Pick the target, freeze it, personalize it, polarize it. Don't try to attack abstract corporations or bureaucracies. Identify a responsible individual. Ignore attempts to shift or spread the blame.

From the moment Trump came to illusory power everything was about him. It wasn't about Republican policy or opinion, but all about Trump. Everything he did was presented in negative, derogatory and abusive terms by the Sabbatian-dominated media led by Cult operations such as CNN, MSNBC, *The New York Times* and the Jeff Bezos-owned *Washington Post* – 'Pick the target, freeze it, personalize it, polarize it.' Trump was turned into a demon to be vilified by those who hated him and a demi-god loved by those who worshipped him. This, in turn, had his supporters, too, presented as equally demonic in preparation for the punchline later down the line when Biden was about to take office. It was here's a Trump, there's a Trump, everywhere a Trump, Trump. Virtually every news story or happening was filtered through the lens of 'The Donald'. You loved him or hated him and which one you chose was said to define you as Satan's spawn or a paragon of virtue. Even supporting some Trump policies or statements and not others was enough for an assault on your character. No shades of grey were or are allowed. Everything is black and white (literally and figuratively). A Californian I knew had her head utterly scrambled by her hatred for Trump while telling people they should love each other. She was so totally consumed by

Trump Derangement Syndrome as it became to be known that this glaring contradiction would never have occurred to her. By definition anyone who criticised Trump or praised his opponents was a hero and this lady described Joe Biden as 'a kind, honest gentleman' when he's a provable liar, mega-crook and vicious piece of work to boot. Sabbatians had indeed divided America using Trump as the fall-guy and all along the clock was ticking on the consequences for his supporters.

In hock to his masters

Trump gave Sabbatians via Israel almost everything they wanted in his four years. Ask and you shall receive was the dynamic between himself and Benjamin Netanyahu orchestrated by Trump's ultra-Zionist son-in-law Jared Kushner, his ultra-Zionist Ambassador to Israel, David Friedman, and ultra-Zionist 'Israel adviser', Jason Greenblatt. The last two were central to the running and protecting from collapse of his business empire, the Trump Organisation, and colossal business failures made him forever beholding to Sabbatian networks that bailed him out. By the start of the 1990s Trump owed \$4 billion to banks that he couldn't pay and almost \$1 billion of that was down to him personally and not his companies. This mega-disaster was the result of building two new casinos in Atlantic City and buying the enormous Taj Mahal operation which led to crippling debt payments. He had borrowed fantastic sums from 72 banks with major Sabbatian connections and although the scale of debt should have had him living in a tent alongside the highway they never foreclosed. A plan was devised to lift Trump from the mire by BT Securities Corporation and Rothschild Inc. and the case was handled by Wilber Ross who had worked for the Rothschilds for 27 years. Ross would be named US Commerce Secretary after Trump's election. Another crucial figure in saving Trump was ultra-Zionist 'investor' Carl Icahn who bought the Taj Mahal casino. Icahn was made special economic adviser on financial regulation in the Trump administration. He didn't stay long but still managed to find time to make a tidy sum of a reported \$31.3 million when he sold his

holdings affected by the price of steel three days before Trump imposed a 235 percent tariff on steel imports. What amazing bits of luck these people have. Trump and Sabbatian operatives have long had a close association and his mentor and legal adviser from the early 1970s until 1986 was the dark and genetically corrupt ultra-Zionist Roy Cohn who was chief counsel to Senator Joseph McCarthy's 'communist' witch-hunt in the 1950s. *Esquire* magazine published an article about Cohn with the headline 'Don't mess with Roy Cohn'. He was described as the most feared lawyer in New York and 'a ruthless master of dirty tricks ... [with] ... more than one Mafia Don on speed dial'. Cohn's influence, contacts, support and protection made Trump a front man for Sabbatians in New York with their connections to one of Cohn's many criminal employers, the 'Russian' Sabbatian Mafia. Israel-centric media mogul Rupert Murdoch was introduced to Trump by Cohn and they started a long friendship. Cohn died in 1986 weeks after being disbarred for unethical conduct by the Appellate Division of the New York State Supreme Court. The wheels of justice do indeed run slow given the length of Cohn's crooked career.

QAnon-sense

We are asked to believe that Donald Trump with his fundamental connections to Sabbatian networks and operatives has been leading the fight to stop the Sabbatian agenda for the fascistic control of America and the world. Sure he has. A man entrapped during his years in the White House by Sabbatian operatives and whose biggest financial donor was casino billionaire Sheldon Adelson who was Sabbatian to his DNA?? Oh, do come on. Trump has been used to divide America and isolate Pushbackers on the Cult agenda under the heading of 'Trump supporters', 'insurrectionists' and 'white supremacists'. The US Intelligence/Mossad Psyop or psychological operation known as QAnon emerged during the Trump years as a central pillar in the Sabbatian campaign to lead Pushbackers into the trap set by those that wished to destroy them. I knew from the start that QAnon was a scam because I had seen the same scenario many

times before over 30 years under different names and I had written about one in particular in the books. 'Not again' was my reaction when QAnon came to the fore. The same script is pulled out every few years and a new name added to the letterhead. The story always takes the same form: 'Insiders' or 'the good guys' in the government-intelligence-military 'Deep State' apparatus were going to instigate mass arrests of the 'bad guys' which would include the Rockefellers, Rothschilds, Barack Obama, Hillary Clinton, George Soros, etc., etc. Dates are given for when the 'good guys' are going to move in, but the dates pass without incident and new dates are given which pass without incident. The central message to Pushbackers in each case is that they don't have to do anything because there is 'a plan' and it is all going to be sorted by the 'good guys' on the inside. 'Trust the plan' was a QAnon mantra when the only plan was to misdirect Pushbackers into putting their trust in a Psyop they believed to be real. Beware, beware, those who tell you what you want to hear and always check it out. Right up to Biden's inauguration QAnon was still claiming that 'the Storm' was coming and Trump would stay on as president when Biden and his cronies were arrested and jailed. It was never going to happen and of course it didn't, but what did happen as a result provided that punchline to the Sabbatian Trump/QAnon Psyop.

On January 6th, 2021, a very big crowd of Trump supporters gathered in the National Mall in Washington DC down from the Capitol Building to protest at what they believed to be widespread corruption and vote fraud that stopped Trump being re-elected for a second term as president in November, 2020. I say as someone that does not support Trump or Biden that the evidence is clear that major vote-fixing went on to favour Biden, a man with cognitive problems so advanced he can often hardly string a sentence together without reading the words written for him on the Teleprompter. Glaring ballot discrepancies included serious questions about electronic voting machines that make vote rigging a comparative cinch and hundreds of thousands of paper votes that suddenly appeared during already advanced vote counts and virtually all of

them for Biden. Early Trump leads in crucial swing states suddenly began to close and disappear. The pandemic hoax was used as the excuse to issue almost limitless numbers of mail-in ballots with no checks to establish that the recipients were still alive or lived at that address. They were sent to streams of people who had not even asked for them. Private organisations were employed to gather these ballots and who knows what they did with them before they turned up at the counts. The American election system has been manipulated over decades to become a sick joke with more holes than a Swiss cheese for the express purpose of dictating the results. Then there was the criminal manipulation of information by Sabbatian tech giants like Facebook, Twitter and Google-owned YouTube which deleted pro-Trump, anti-Biden accounts and posts while everything in support of Biden was left alone. Sabbatians wanted Biden to win because after the dividing of America it was time for full-on Woke and every aspect of the Cult agenda to be unleashed.

Hunter gatherer

Extreme Silicon Valley bias included blocking information by the *New York Post* exposing a Biden scandal that should have ended his bid for president in the final weeks of the campaign. Hunter Biden, his monumentally corrupt son, is reported to have sent a laptop to be repaired at a local store and failed to return for it. Time passed until the laptop became the property of the store for non-payment of the bill. When the owner saw what was on the hard drive he gave a copy to the FBI who did nothing even though it confirmed widespread corruption in which the Joe Biden family were using his political position, especially when he was vice president to Obama, to make multiple millions in countries around the world and most notably Ukraine and China. Hunter Biden's one-time business partner Tony Bobulinski went public when the story broke in the *New York Post* to confirm the corruption he saw and that Joe Biden not only knew what was going on he also profited from the spoils. Millions were handed over by a Chinese company with close

connections – like all major businesses in China – to the Chinese communist party of President Xi Jinping. Joe Biden even boasted at a meeting of the Cult's World Economic Forum that as vice president he had ordered the government of Ukraine to fire a prosecutor. What he didn't mention was that the same man just happened to be investigating an energy company which was part of Hunter Biden's corrupt portfolio. The company was paying him big bucks for no other reason than the influence his father had. Overnight Biden's presidential campaign should have been over given that he had lied publicly about not knowing what his son was doing. Instead almost the entire Sabbatian-owned mainstream media and Sabbatian-owned Silicon Valley suppressed circulation of the story. This alone went a mighty way to rigging the election of 2020. Cult assets like Mark Zuckerberg at Facebook also spent hundreds of millions to be used in support of Biden and vote 'administration'.

The Cult had used Trump as the focus to divide America and was now desperate to bring in moronic, pliable, corrupt Biden to complete the double-whammy. No way were they going to let little things like the will of the people thwart their plan. Silicon Valley widely censored claims that the election was rigged because it *was* rigged. For the same reason anyone claiming it was rigged was denounced as a 'white supremacist' including the pathetically few Republican politicians willing to say so. Right across the media where the claim was mentioned it was described as a 'false claim' even though these excuses for 'journalists' would have done no research into the subject whatsoever. Trump won seven million more votes than any sitting president had ever achieved while somehow a cognitively-challenged soon to be 78-year-old who was hidden away from the public for most of the campaign managed to win more votes than any presidential candidate in history. It makes no sense. You only had to see election rallies for both candidates to witness the enthusiasm for Trump and the apathy for Biden. Tens of thousands would attend Trump events while Biden was speaking in empty car parks with often only television crews attending and framing their shots to hide the fact that no one was there. It was pathetic to see

footage come to light of Biden standing at a podium making speeches only to TV crews and party fixers while reading the words written for him on massive Teleprompter screens. So, yes, those protestors on January 6th had a point about election rigging, but some were about to walk into a trap laid for them in Washington by the Cult Deep State and its QAnon Psyop. This was the Capitol Hill riot ludicrously dubbed an 'insurrection'.

The spider and the fly

Renegade Minds know there are not two 'sides' in politics, only one side, the Cult, working through all 'sides'. It's a stage show, a puppet show, to direct the perceptions of the population into focusing on diversions like parties and candidates while missing the puppeteers with their hands holding all the strings. The Capitol Hill 'insurrection' brings us back to the Little Big Horn. Having created two distinct opposing groupings – Woke and Pushbackers – the trap was about to be sprung. Pushbackers were to be encircled and isolated by associating them all in the public mind with Trump and then labelling Trump as some sort of Confederate leader. I knew immediately that the Capitol riot was a set-up because of two things. One was how easy the rioters got into the building with virtually no credible resistance and secondly I could see – as with the 'Covid' hoax in the West at the start of 2020 – how the Cult could exploit the situation to move its agenda forward with great speed. My experience of Cult techniques and activities over more than 30 years has showed me that while they do exploit situations they haven't themselves created this never happens with events of fundamental agenda significance. Every time major events giving cultists the excuse to rapidly advance their plan you find they are manipulated into being for the specific reason of providing that excuse – Problem-Reaction-Solution. Only a tiny minority of the huge crowd of Washington protestors sought to gain entry to the Capitol by smashing windows and breaching doors. That didn't matter. The whole crowd and all Pushbackers, even if they did not support Trump, were going to be lumped together as dangerous

insurrectionists and conspiracy theorists. The latter term came into widespread use through a CIA memo in the 1960s aimed at discrediting those questioning the nonsensical official story of the Kennedy assassination and it subsequently became widely employed by the media. It's still being used by inept 'journalists' with no idea of its origin to discredit anyone questioning anything that authority claims to be true. When you are perpetrating a conspiracy you need to discredit the very word itself even though the dictionary definition of conspiracy is merely 'the activity of secretly planning with other people to do something bad or illegal' and 'a general agreement to keep silent about a subject for the purpose of keeping it secret'. On that basis there are conspiracies almost wherever you look. For obvious reasons the Cult and its lapdog media have to claim there are no conspiracies even though the word appears in state laws as with conspiracy to defraud, to murder, and to corrupt public morals.

Agent provocateurs are widely used by the Cult Deep State to manipulate genuine people into acting in ways that suit the desired outcome. By genuine in this case I mean protestors genuinely supporting Trump and claims that the election was stolen. In among them, however, were agents of the state wearing the garb of Trump supporters and QAnon to pump-prime the Capital riot which some genuine Trump supporters naively fell for. I described the situation as 'Come into my parlour said the spider to the fly'. Leaflets appeared through the Woke paramilitary arm Antifa, the anti-fascist fascists, calling on supporters to turn up in Washington looking like Trump supporters even though they hated him. Some of those arrested for breaching the Capitol Building were sourced to Antifa and its stable mate Black Lives Matter. Both organisations are funded by Cult billionaires and corporations. One man charged for the riot was according to his lawyer a former FBI agent who had held top secret security clearance for 40 years. Attorney Thomas Plofchan said of his client, 66-year-old Thomas Edward Caldwell:

He has held a Top Secret Security Clearance since 1979 and has undergone multiple Special Background Investigations in support of his clearances. After retiring from the Navy, he

worked as a section chief for the Federal Bureau of Investigation from 2009-2010 as a GS-12 [mid-level employee].

He also formed and operated a consulting firm performing work, often classified, for U.S government customers including the US. Drug Enforcement Agency, Department of Housing and Urban Development, the US Coast Guard, and the US Army Personnel Command.

A judge later released Caldwell pending trial in the absence of evidence about a conspiracy or that he tried to force his way into the building. *The New York Post* reported a 'law enforcement source' as saying that 'at least two known Antifa members were spotted' on camera among Trump supporters during the riot while one of the rioters arrested was John Earle Sullivan, a seriously extreme Black Lives Matter Trump-hater from Utah who was previously arrested and charged in July, 2020, over a BLM-Antifa riot in which drivers were threatened and one was shot. Sullivan is the founder of Utah-based Insurgence USA which is an affiliate of the Cult-created-and-funded Black Lives Matter movement. Footage appeared and was then deleted by Twitter of Trump supporters calling out Antifa infiltrators and a group was filmed changing into pro-Trump clothing before the riot. Security at the building was *pathetic* – as planned. Colonel Leroy Fletcher Prouty, a man with long experience in covert operations working with the US security apparatus, once described the tell-tale sign to identify who is involved in an assassination. He said:

No one has to direct an assassination – it happens. The active role is played secretly by permitting it to happen. This is the greatest single clue. Who has the power to call off or reduce the usual security precautions?

This principle applies to many other situations and certainly to the Capitol riot of January 6th, 2021.

The sting

With such a big and potentially angry crowd known to be gathering near the Capitol the security apparatus would have had a major police detail to defend the building with National Guard troops on

standby given the strength of feeling among people arriving from all over America encouraged by the QAnon Psyop and statements by Donald Trump. Instead Capitol Police 'security' was flimsy, weak, and easily breached. The same number of officers was deployed as on a regular day and that is a blatant red flag. They were not staffed or equipped for a possible riot that had been an obvious possibility in the circumstances. No protective and effective fencing worth the name was put in place and there were no contingency plans. The whole thing was basically a case of standing aside and waving people in. Once inside police mostly backed off apart from one Capitol police officer who ridiculously shot dead unarmed Air Force veteran protestor Ashli Babbitt without a warning as she climbed through a broken window. The 'investigation' refused to name or charge the officer after what must surely be considered a murder in the circumstances. They just lifted a carpet and swept. The story was endlessly repeated about five people dying in the 'armed insurrection' when there was no report of rioters using weapons. Apart from Babbitt the other four died from a heart attack, strokes and apparently a drug overdose. Capitol police officer Brian Sicknick was reported to have died after being bludgeoned with a fire extinguisher when he was alive after the riot was over and died later of what the Washington Medical Examiner's Office said was a stroke. Sicknick had no external injuries. The lies were delivered like rapid fire. There was a narrative to build with incessant repetition of the lie until the lie became the accepted 'everybody knows that' truth. The 'Big Lie' technique of Nazi Propaganda Minister Joseph Goebbels is constantly used by the Cult which was behind the Nazis and is today behind the 'Covid' and 'climate change' hoaxes. Goebbels said:

If you tell a lie big enough and keep repeating it, people will eventually come to believe it. The lie can be maintained only for such time as the State can shield the people from the political, economic and/or military consequences of the lie. It thus becomes vitally important for the State to use all of its powers to repress dissent, for the truth is the mortal enemy of the lie, and thus by extension, the truth is the greatest enemy of the State.

Most protestors had a free run of the Capitol Building. This allowed pictures to be taken of rioters in iconic parts of the building including the Senate chamber which could be used as propaganda images against all Pushbackers. One Congresswoman described the scene as 'the worst kind of non-security anybody could ever imagine'. Well, the first part was true, but someone obviously did imagine it and made sure it happened. Some photographs most widely circulated featured people wearing QAnon symbols and now the Psyop would be used to dub all QAnon followers with the ubiquitous fit-all label of 'white supremacist' and 'insurrectionists'. When a Muslim extremist called Noah Green drove his car at two police officers at the Capitol Building killing one in April, 2021, there was no such political and media hysteria. They were just disappointed he wasn't white.

The witch-hunt

Government prosecutor Michael Sherwin, an aggressive, dark-eyed, professional Rottweiler led the 'investigation' and to call it over the top would be to understate reality a thousand fold. Hundreds were tracked down and arrested for the crime of having the wrong political views and people were jailed who had done nothing more than walk in the building, committed no violence or damage to property, took a few pictures and left. They were labelled a 'threat to the Republic' while Biden sat in the White House signing executive orders written for him that were dismantling 'the Republic'. Even when judges ruled that a mother and son should not be in jail the government kept them there. Some of those arrested have been badly beaten by prison guards in Washington and lawyers for one man said he suffered a fractured skull and was made blind in one eye. Meanwhile a woman is shot dead for no reason by a Capitol Police officer and we are not allowed to know who he is never mind what has happened to him although that will be *nothing*. The Cult's QAnon/Trump sting to identify and isolate Pushbackers and then target them on the road to crushing and deleting them was a resounding success. You would have thought the Russians had

invaded the building at gunpoint and lined up senators for a firing squad to see the political and media reaction. Congresswoman Alexandria Ocasio-Cortez is a child in a woman's body, a terrible-tuos, me, me, me, Woker narcissist of such proportions that words have no meaning. She said she thought she was going to die when 'insurrectionists' banged on her office door. It turned out she wasn't even in the Capitol Building when the riot was happening and the 'banging' was a Capitol Police officer. She referred to herself as a 'survivor' which is an insult to all those true survivors of violent and sexual abuse while she lives her pampered and privileged life talking drivel for a living. Her Woke colleague and fellow mega-narcissist Rashida Tlaib broke down describing the devastating effect on her, too, of *not being* in the building when the rioters were there. Ocasio-Cortez and Tlaib are members of a fully-Woke group of Congresswomen known as 'The Squad' along with Ilhan Omar and Ayanna Pressley. The Squad from what I can see can be identified by its vehement anti-white racism, anti-white men agenda, and, as always in these cases, the absence of brain cells on active duty.

The usual suspects were on the riot case immediately in the form of Democrat ultra-Zionist senators and operatives Chuck Schumer and Adam Schiff demanding that Trump be impeached for 'his part in the insurrection'. The same pair of prats had led the failed impeachment of Trump over the invented 'Russia collusion' nonsense which claimed Russia had helped Trump win the 2016 election. I didn't realise that Tel Aviv had been relocated just outside Moscow. I must find an up-to-date map. The Russia hoax was a Sabbatian operation to keep Trump occupied and impotent and to stop any rapport with Russia which the Cult wants to retain as a perceptual enemy to be pulled out at will. Puppet Biden began attacking Russia when he came to office as the Cult seeks more upheaval, division and war across the world. A two-year stage show 'Russia collusion inquiry' headed by the not-very-bright former 9/11 FBI chief Robert Mueller, with support from 19 lawyers, 40 FBI agents plus intelligence analysts, forensic accountants and other

staff, devoured tens of millions of dollars and found no evidence of Russia collusion which a ten-year-old could have told them on day one. Now the same moronic Schumer and Schiff wanted a second impeachment of Trump over the Capitol 'insurrection' (riot) which the arrested development of Schumer called another 'Pearl Harbor' while others compared it with 9/11 in which 3,000 died and, in the case of CNN, with the Rwandan genocide in the 1990s in which an estimated 500,000 to 600,000 were murdered, between 250, 000 and 500,000 women were raped, and populations of whole towns were hacked to death with machetes. To make those comparisons purely for Cult political reasons is beyond insulting to those that suffered and lost their lives and confirms yet again the callous inhumanity that we are dealing with. Schumer is a monumental idiot and so is Schiff, but they serve the Cult agenda and do whatever they're told so they get looked after. Talking of idiots – another inane man who spanned the Russia and Capitol impeachment attempts was Senator Eric Swalwell who had the nerve to accuse Trump of collusion with the Russians while sleeping with a Chinese spy called Christine Fang or 'Fang Fang' which is straight out of a Bond film no doubt starring Klaus Schwab as the bloke living on a secret island and controlling laser weapons positioned in space and pointing at world capitals. Fang Fang plays the part of Bond's infiltrator girlfriend which I'm sure she would enjoy rather more than sharing a bed with the brainless Swalwell, lying back and thinking of China. The FBI eventually warned Swalwell about Fang Fang which gave her time to escape back to the Chinese dictatorship. How very thoughtful of them. The second Trump impeachment also failed and hardly surprising when an impeachment is supposed to remove a sitting president and by the time it happened Trump was no longer president. These people are running your country America, well, officially anyway. Terrifying isn't it?

Outcomes tell the story - always

The outcome of all this – and it's the *outcome* on which Renegade Minds focus, not the words – was that a vicious, hysterical and

obviously pre-planned assault was launched on Pushbackers to censor, silence and discredit them and even targeted their right to earn a living. They have since been condemned as 'domestic terrorists' that need to be treated like Al-Qaeda and Islamic State. 'Domestic terrorists' is a label the Cult has been trying to make stick since the period of the Oklahoma bombing in 1995 which was blamed on 'far-right domestic terrorists'. If you read *The Trigger* you will see that the bombing was clearly a Problem-Reaction-Solution carried out by the Deep State during a Bill Clinton administration so corrupt that no dictionary definition of the term would even nearly suffice. Nearly 30, 000 troops were deployed from all over America to the empty streets of Washington for Biden's inauguration. Ten thousand of them stayed on with the pretext of protecting the capital from insurrectionists when it was more psychological programming to normalise the use of the military in domestic law enforcement in support of the Cult plan for a police-military state. Biden's fascist administration began a purge of 'wrong-thinkers' in the military which means anyone that is not on board with Woke. The Capitol Building was surrounded by a fence with razor wire and the Land of the Free was further symbolically and literally dismantled. The circle was completed with the installation of Biden and the exploitation of the QAnon Psyop.

America had never been so divided since the civil war of the 19th century, Pushbackers were isolated and dubbed terrorists and now, as was always going to happen, the Cult immediately set about deleting what little was left of freedom and transforming American society through a swish of the hand of the most controlled 'president' in American history leading (officially at least) the most extreme regime since the country was declared an independent state on July 4th, 1776. Biden issued undebated, dictatorial executive orders almost by the hour in his opening days in office across the whole spectrum of the Cult wish-list including diluting controls on the border with Mexico allowing thousands of migrants to illegally enter the United States to transform the demographics of America and import an election-changing number of perceived Democrat

voters. Then there were Biden deportation amnesties for the already illegally resident (estimated to be as high as 20 or even 30 million). A bill before Congress awarded American citizenship to anyone who could prove they had worked in agriculture for just 180 days in the previous two years as 'Big Ag' secured its slave labour long-term. There were the plans to add new states to the union such as Puerto Rico and making Washington DC a state. They are all parts of a plan to ensure that the Cult-owned Woke Democrats would be permanently in power.

Border – what border?

I have exposed in detail in other books how mass immigration into the United States and Europe is the work of Cult networks fuelled by the tens of billions spent to this and other ends by George Soros and his global Open Society (open borders) Foundations. The impact can be seen in America alone where the population has increased by *100 million* in little more than 30 years mostly through immigration. I wrote in *The Answer* that the plan was to have so many people crossing the southern border that the numbers become unstoppable and we are now there under Cult-owned Biden. El Salvador in Central America puts the scale of what is happening into context. A third of the population now lives in the United States, much of it illegally, and many more are on the way. The methodology is to crush Central and South American countries economically and spread violence through machete-wielding psychopathic gangs like MS-13 based in El Salvador and now operating in many American cities. Biden-imposed lax security at the southern border means that it is all but open. He said before his 'election' that he wanted to see a surge towards the border if he became president and that was the green light for people to do just that after election day to create the human disaster that followed for both America and the migrants. When that surge came the imbecilic Alexandria Ocasio-Cortez said it wasn't a 'surge' because they are 'children, not insurgents' and the term 'surge' (used by Biden) was a claim of 'white supremacists'.

This disingenuous lady may one day enter the realm of the most basic intelligence, but it won't be any time soon.

Sabbatians and the Cult are in the process of destroying America by importing violent people and gangs in among the genuine to terrorise American cities and by overwhelming services that cannot cope with the sheer volume of new arrivals. Something similar is happening in Europe as Western society in general is targeted for demographic and cultural transformation and upheaval. The plan demands violence and crime to create an environment of intimidation, fear and division and Soros has been funding the election of district attorneys across America who then stop prosecuting many crimes, reduce sentences for violent crimes and free as many violent criminals as they can. Sabbatians are creating the chaos from which order – their order – can respond in a classic Problem-Reaction-Solution. A Freemasonic moto says 'Ordo Ab Chao' (Order out of Chaos) and this is why the Cult is constantly creating chaos to impose a new 'order'. Here you have the reason the Cult is constantly creating chaos. The 'Covid' hoax can be seen with those entering the United States by plane being forced to take a 'Covid' test while migrants flooding through southern border processing facilities do not. Nothing is put in the way of mass migration and if that means ignoring the government's own 'Covid' rules then so be it. They know it's all bullshit anyway. Any pushback on this is denounced as 'racist' by Wokers and Sabbatian fronts like the ultra-Zionist Anti-Defamation League headed by the appalling Jonathan Greenblatt which at the same time argues that Israel should not give citizenship and voting rights to more Palestinian Arabs or the 'Jewish population' (in truth the Sabbatian network) will lose control of the country.

Society-changing numbers

Biden's masters have declared that countries like El Salvador are so dangerous that their people must be allowed into the United States for humanitarian reasons when there are fewer murders in large parts of many Central American countries than in US cities like

Baltimore. That is not to say Central America cannot be a dangerous place and Cult-controlled American governments have been making it so since way back, along with the dismantling of economies, in a long-term plan to drive people north into the United States. Parts of Central America are very dangerous, but in other areas the story is being greatly exaggerated to justify relaxing immigration criteria. Migrants are being offered free healthcare and education in the United States as another incentive to head for the border and there is no requirement to be financially independent before you can enter to prevent the resources of America being drained. You can't blame migrants for seeking what they believe will be a better life, but they are being played by the Cult for dark and nefarious ends. The numbers since Biden took office are huge. In February, 2021, more than 100,000 people were known to have tried to enter the US illegally through the southern border (it was 34,000 in the same month in 2020) and in March it was 170,000 – a 418 percent increase on March, 2020. These numbers are only known people, not the ones who get in unseen. The true figure for migrants illegally crossing the border in a single month was estimated by one congressman at 250,000 and that number will only rise under Biden's current policy. Gangs of murdering drug-running thugs that control the Mexican side of the border demand money – thousands of dollars – to let migrants cross the Rio Grande into America. At the same time gun battles are breaking out on the border several times a week between rival Mexican drug gangs (which now operate globally) who are equipped with sophisticated military-grade weapons, grenades and armoured vehicles. While the Capitol Building was being 'protected' from a non-existent 'threat' by thousands of troops, and others were still deployed at the time in the Cult Neocon war in Afghanistan, the southern border of America was left to its fate. This is not incompetence, it is cold calculation.

By March, 2021, there were 17,000 unaccompanied children held at border facilities and many of them are ensnared by people traffickers for paedophile rings and raped on their journey north to America. This is not conjecture – this is fact. Many of those designated

children are in reality teenage boys or older. Meanwhile Wokers posture their self-purity for encouraging poor and tragic people to come to America and face this nightmare both on the journey and at the border with the disgusting figure of House Speaker Nancy Pelosi giving disingenuous speeches about caring for migrants. The woman's evil. Wokers condemned Trump for having children in cages at the border (so did Obama, *Shhhh*), but now they are sleeping on the floor without access to a shower with one border facility 729 percent over capacity. The Biden insanity even proposed flying migrants from the southern border to the northern border with Canada for 'processing'. The whole shambles is being overseen by ultra-Zionist Secretary of Homeland Security, the moronic liar Alejandro Mayorkas, who banned news cameras at border facilities to stop Americans seeing what was happening. Mayorkas said there was not a ban on news crews; it was just that they were not allowed to film. Alongside him at Homeland Security is another ultra-Zionist Cass Sunstein appointed by Biden to oversee new immigration laws. Sunstein despises conspiracy researchers to the point where he suggests they should be banned or *taxed* for having such views. The man is not bonkers or anything. He's perfectly well-adjusted, but adjusted to what is the question. Criticise what is happening and you are a 'white supremacist' when earlier non-white immigrants also oppose the numbers which effect their lives and opportunities. Black people in poor areas are particularly damaged by uncontrolled immigration and the increased competition for work opportunities with those who will work for less. They are also losing voting power as Hispanics become more dominant in former black areas. It's a downward spiral for them while the billionaires behind the policy drone on about how much they care about black people and 'racism'. None of this is about compassion for migrants or black people – that's just wind and air. Migrants are instead being mercilessly exploited to transform America while the countries they leave are losing their future and the same is true in Europe. Mass immigration may now be the work of Woke Democrats, but it can be traced back to the 1986 Immigration Reform and Control Act (it

wasn't) signed into law by Republican hero President Ronald Reagan which gave amnesty to millions living in the United States illegally and other incentives for people to head for the southern border. Here we have the one-party state at work again.

Save me syndrome

Almost every aspect of what I have been exposing as the Cult agenda was on display in even the first days of 'Biden' with silencing of Pushbackers at the forefront of everything. A Renegade Mind will view the Trump years and QAnon in a very different light to their supporters and advocates as the dots are connected. The QAnon/Trump Psyop has given the Cult all it was looking for. We may not know how much, or little, that Trump realised he was being used, but that's a side issue. This pincer movement produced the desired outcome of dividing America and having Pushbackers isolated. To turn this around we have to look at new routes to empowerment which do not include handing our power to other people and groups through what I will call the 'Save Me Syndrome' – 'I want someone else to do it so that I don't have to'. We have seen this at work throughout human history and the QAnon/Trump Psyop is only the latest incarnation alongside all the others. Religion is an obvious expression of this when people look to a 'god' or priest to save them or tell them how to be saved and then there are 'save me' politicians like Trump. Politics is a diversion and not a 'saviour'. It is a means to block positive change, not make it possible.

Save Me Syndrome always comes with the same repeating theme of handing your power to whom or what you believe will save you while your real 'saviour' stares back from the mirror every morning. Renegade Minds are constantly vigilant in this regard and always asking the question 'What can I do?' rather than 'What can someone else do for me?' Gandhi was right when he said: 'You must be the change you want to see in the world.' We are indeed the people we have been waiting for. We are presented with a constant raft of reasons to concede that power to others and forget where the real power is. Humanity has the numbers and the Cult does not. It has to

use diversion and division to target the unstoppable power that comes from unity. Religions, governments, politicians, corporations, media, QAnon, are all different manifestations of this power-diversion and dilution. Refusing to give your power to governments and instead handing it to Trump and QAnon is not to take a new direction, but merely to recycle the old one with new names on the posters. I will explore this phenomenon as we proceed and how to break the cycles and recycles that got us here through the mists of repeating perception and so repeating history.

For now we shall turn to the most potent example in the entire human story of the consequences that follow when you give your power away. I am talking, of course, of the 'Covid' hoax.

CHAPTER FOUR

'Covid': Calculated catastrophe

Facts are threatening to those invested in fraud
DaShanne Stokes

We can easily unravel the real reason for the 'Covid pandemic' hoax by employing the Renegade Mind methodology that I have outlined this far. We'll start by comparing the long-planned Cult outcome with the 'Covid pandemic' outcome. Know the outcome and you'll see the journey.

I have highlighted the plan for the Hunger Games Society which has been in my books for so many years with the very few controlling the very many through ongoing dependency. To create this dependency it is essential to destroy independent livelihoods, businesses and employment to make the population reliant on the state (the Cult) for even the basics of life through a guaranteed pittance income. While independence of income remained these Cult ambitions would be thwarted. With this knowledge it was easy to see where the 'pandemic' hoax was going once talk of 'lockdowns' began and the closing of all but perceived 'essential' businesses to 'save' us from an alleged 'deadly virus'. Cult corporations like Amazon and Walmart were naturally considered 'essential' while mom and pop shops and stores had their doors closed by fascist decree. As a result with every new lockdown and new regulation more small and medium, even large businesses not owned by the Cult, went to the wall while Cult giants and their frontmen and women grew financially fatter by the second. Mom and pop were

denied an income and the right to earn a living and the wealth of people like Jeff Bezos (Amazon), Mark Zuckerberg (Facebook) and Sergei Brin and Larry Page (Google/Alphabet) have reached record levels. The Cult was increasing its own power through further dramatic concentrations of wealth while the competition was being destroyed and brought into a state of dependency. Lockdowns have been instigated to secure that very end and were never anything to do with health. My brother Paul spent 45 years building up a bus repair business, but lockdowns meant buses were running at a fraction of normal levels for months on end. Similar stories can be told in their hundreds of millions worldwide. Efforts of a lifetime coldly destroyed by Cult multi-billionaires and their lackeys in government and law enforcement who continued to earn their living from the taxation of the people while denying the right of the same people to earn theirs. How different it would have been if those making and enforcing these decisions had to face the same financial hardships of those they affected, but they never do.

Gates of Hell

Behind it all in the full knowledge of what he is doing and why is the psychopathic figure of Cult operative Bill Gates. His puppet Tedros at the World Health Organization declared 'Covid' a pandemic in March, 2020. The WHO had changed the definition of a 'pandemic' in 2009 just a month before declaring the 'swine flu pandemic' which would not have been so under the previous definition. The same applies to 'Covid'. The definition had included... 'an infection by an infectious agent, occurring simultaneously in different countries, with a significant mortality rate relative to the proportion of the population infected'. The new definition removed the need for 'significant mortality'. The 'pandemic' has been fraudulent even down to the definition, but Gates demanded economy-destroying lockdowns, school closures, social distancing, mandatory masks, a 'vaccination' for every man, woman and child on the planet and severe consequences and restrictions for those that refused. Who gave him this power? The

Cult did which he serves like a little boy in short trousers doing what his daddy tells him. He and his psychopathic missus even smiled when they said that much worse was to come (what they knew was planned to come). Gates responded in the matter-of-fact way of all psychopaths to a question about the effect on the world economy of what he was doing:

Well, it won't go to zero but it will shrink. Global GDP is probably going to take the biggest hit ever [Gates was smiling as he said this] ... in my lifetime this will be the greatest economic hit. But you don't have a choice. People act as if you have a choice. People don't feel like going to the stadium when they might get infected ... People are deeply affected by seeing these stats, by knowing they could be part of the transmission chain, old people, their parents and grandparents, could be affected by this, and so you don't get to say ignore what is going on here.

There will be the ability to open up, particularly in rich countries, if things are done well over the next few months, but for the world at large normalcy only returns when we have largely vaccinated the entire population.

The man has no compassion or empathy. How could he when he's a psychopath like all Cult players? My own view is that even beyond that he is very seriously mentally ill. Look in his eyes and you can see this along with his crazy flailing arms. You don't do what he has done to the world population since the start of 2020 unless you are mentally ill and at the most extreme end of psychopathic. You especially don't do it when to you know, as we shall see, that cases and deaths from 'Covid' are fakery and a product of monumental figure massaging. 'These stats' that Gates referred to are based on a 'test' that's not testing for the 'virus' as he has known all along. He made his fortune with big Cult support as an infamously ruthless software salesman and now buys global control of 'health' (death) policy without the population he affects having any say. It's a breathtaking outrage. Gates talked about people being deeply affected by fear of 'Covid' when that was because of *him* and his global network lying to them minute-by-minute supported by a lying media that he seriously influences and funds to the tune of hundreds of millions. He's handed big sums to media operations including the BBC, NBC, Al Jazeera, Univision, *PBS NewsHour*,

ProPublica, National Journal, The Guardian, The Financial Times, The Atlantic, Texas Tribune, USA Today publisher Gannett, Washington Monthly, Le Monde, Center for Investigative Reporting, Pulitzer Center on Crisis Reporting, National Press Foundation, International Center for Journalists, Solutions Journalism Network, the Poynter Institute for Media Studies, and many more. Gates is everywhere in the 'Covid' hoax and the man must go to prison – or a mental facility – for the rest of his life and his money distributed to those he has taken such enormous psychopathic pleasure in crushing.

The Muscle

The Hunger Games global structure demands a police-military state – a fusion of the two into one force – which viciously imposes the will of the Cult on the population and protects the Cult from public rebellion. In that regard, too, the 'Covid' hoax just keeps on giving. Often unlawful, ridiculous and contradictory 'Covid' rules and regulations have been policed across the world by moronic automatons and psychopaths made faceless by face-nappy masks and acting like the Nazi SS and fascist blackshirts and brownshirts of Hitler and Mussolini. The smallest departure from the rules decreed by the psychos in government and their clueless gofers were jumped upon by the face-nappy fascists. Brutality against public protestors soon became commonplace even on girls, women and old people as the brave men with the batons – the Face-Nappies as I call them – broke up peaceful protests and handed out fines like confetti to people who couldn't earn a living let alone pay hundreds of pounds for what was once an accepted human right. Robot Face-Nappies of Nottingham police in the English East Midlands fined one group £11,000 for attending a child's birthday party. For decades I charted the transformation of law enforcement as genuine, decent officers were replaced with psychopaths and the brain dead who would happily and brutally do whatever their masters told them. Now they were let loose on the public and I would emphasise the point that none of this just happened. The step-by-step change in the dynamic between police and public was orchestrated from the shadows by

those who knew where this was all going and the same with the perceptual reframing of those in all levels of authority and official administration through 'training courses' by organisations such as Common Purpose which was created in the late 1980s and given a massive boost in Blair era Britain until it became a global phenomenon. Supposed public 'servants' began to view the population as the enemy and the same was true of the police. This was the start of the explosion of behaviour manipulation organisations and networks preparing for the all-war on the human psyche unleashed with the dawn of 2020. I will go into more detail about this later in the book because it is a core part of what is happening.

Police desecrated beauty spots to deter people gathering and arrested women for walking in the countryside alone 'too far' from their homes. We had arrogant, clueless sergeants in the Isle of Wight police where I live posting on Facebook what they insisted the population must do or else. A schoolmaster sergeant called Radford looked young enough for me to ask if his mother knew he was out, but he was posting what he *expected* people to do while a Sergeant Wilkinson boasted about fining lads for meeting in a McDonald's car park where they went to get a lockdown takeaway. Wilkinson added that he had even cancelled their order. What a pair of prats these people are and yet they have increasingly become the norm among Jackboot Johnson's Yellowshirts once known as the British police. This was the theme all over the world with police savagery common during lockdown protests in the United States, the Netherlands, and the fascist state of Victoria in Australia under its tyrannical and again moronic premier Daniel Andrews. Amazing how tyrannical and moronic tend to work as a team and the same combination could be seen across America as arrogant, narcissistic Woke governors and mayors such as Gavin Newsom (California), Andrew Cuomo (New York), Gretchen Whitmer (Michigan), Lori Lightfoot (Chicago) and Eric Garcetti (Los Angeles) did their Nazi and Stalin impressions with the full support of the compliant brutality of their enforcers in uniform as they arrested small business owners defying

fascist shutdown orders and took them to jail in ankle shackles and handcuffs. This happened to bistro owner Marlena Pavlos-Hackney in Gretchen Whitmer's fascist state of Michigan when police arrived to enforce an order by a state-owned judge for 'putting the community at risk' at a time when other states like Texas were dropping restrictions and migrants were pouring across the southern border without any 'Covid' questions at all. I'm sure there are many officers appalled by what they are ordered to do, but not nearly enough of them. If they were truly appalled they would not do it. As the months passed every opportunity was taken to have the military involved to make their presence on the streets ever more familiar and 'normal' for the longer-term goal of police-military fusion.

Another crucial element to the Hunger Games enforcement network has been encouraging the public to report neighbours and others for 'breaking the lockdown rules'. The group faced with £11,000 in fines at the child's birthday party would have been dobbed-in by a neighbour with a brain the size of a pea. The technique was most famously employed by the Stasi secret police in communist East Germany who had public informants placed throughout the population. A police chief in the UK says his force doesn't need to carry out 'Covid' patrols when they are flooded with so many calls from the public reporting other people for visiting the beach. Dorset police chief James Vaughan said people were so enthusiastic about snitching on their fellow humans they were now operating as an auxiliary arm of the police: 'We are still getting around 400 reports a week from the public, so we will respond to reports ... We won't need to be doing hotspot patrols because people are very quick to pick the phone up and tell us.' Vaughan didn't say that this is a pillar of all tyrannies of whatever complexion and the means to hugely extend the reach of enforcement while spreading distrust among the people and making them wary of doing anything that might get them reported. Those narcissistic Isle of Wight sergeants Radford and Wilkinson never fail to add a link to their Facebook posts where the public can inform on their fellow slaves.

Neither would be self-aware enough to realise they were imitating the Stasi which they might well never have heard of. Government psychologists that I will expose later laid out a policy to turn communities against each other in the same way.

A coincidence? Yep, and I can knit fog

I knew from the start of the alleged pandemic that this was a Cult operation. It presented limitless potential to rapidly advance the Cult agenda and exploit manipulated fear to demand that every man, woman and child on the planet was 'vaccinated' in a process never used on humans before which infuses self-replicating *synthetic* material into human cells. Remember the plan to transform the human body from a biological to a synthetic biological state. I'll deal with the 'vaccine' (that's not actually a vaccine) when I focus on the genetic agenda. Enough to say here that mass global 'vaccination' justified by this 'new virus' set alarms ringing after 30 years of tracking these people and their methods. The 'Covid' hoax officially beginning in China was also a big red flag for reasons I will be explaining. The agenda potential was so enormous that I could dismiss any idea that the 'virus' appeared naturally. Major happenings with major agenda implications never occur without Cult involvement in making them happen. My questions were twofold in early 2020 as the media began its campaign to induce global fear and hysteria: Was this alleged infectious agent released on purpose by the Cult or did it even exist at all? I then did what I always do in these situations. I sat, observed and waited to see where the evidence and information would take me. By March and early April synchronicity was strongly – and ever more so since then – pointing me in the direction of *there is no 'virus'*. I went public on that with derision even from swathes of the alternative media that voiced a scenario that the Chinese government released the 'virus' in league with Deep State elements in the United States from a top-level bio-lab in Wuhan where the 'virus' is said to have first appeared. I looked at that possibility, but I didn't buy it for several reasons. Deaths from the 'virus' did not in any way match what they

would have been with a 'deadly bioweapon' and it is much more effective if you sell the *illusion* of an infectious agent rather than having a real one unless you can control through injection who has it and who doesn't. Otherwise you lose control of events. A made-up 'virus' gives you a blank sheet of paper on which you can make it do whatever you like and have any symptoms or mutant 'variants' you choose to add while a real infectious agent would limit you to what it actually does. A phantom disease allows you to have endless ludicrous 'studies' on the 'Covid' dollar to widen the perceived impact by inventing ever more 'at risk' groups including one study which said those who walk slowly may be almost four times more likely to die from the 'virus'. People are in psychiatric wards for less.

A real 'deadly bioweapon' can take out people in the hierarchy that are not part of the Cult, but essential to its operation. Obviously they don't want that. Releasing a real disease means you immediately lose control of it. Releasing an illusory one means you don't. Again it's vital that people are extra careful when dealing with what they want to hear. A bioweapon unleashed from a Chinese laboratory in collusion with the American Deep State may fit a conspiracy narrative, but is it true? Would it not be far more effective to use the excuse of a 'virus' to justify the real bioweapon – the 'vaccine'? That way your disease agent does not have to be transmitted and arrives directly through a syringe. I saw a French virologist Luc Montagnier quoted in the alternative media as saying he had discovered that the alleged 'new' severe acute respiratory syndrome coronavirus , or SARS-CoV-2, was made artificially and included elements of the human immunodeficiency 'virus' (HIV) and a parasite that causes malaria. SARS-CoV-2 is alleged to trigger an alleged illness called Covid-19. I remembered Montagnier's name from my research years before into claims that an HIV 'retrovirus' causes AIDs – claims that were demolished by Berkeley virologist Peter Duesberg who showed that no one had ever proved that HIV causes acquired immunodeficiency syndrome or AIDS. Claims that become accepted as fact, publicly and medically, with no proof whatsoever are an ever-recurring story that profoundly applies to

‘Covid’. Nevertheless, despite the lack of proof, Montagnier’s team at the Pasteur Institute in Paris had a long dispute with American researcher Robert Gallo over which of them discovered and isolated the HIV ‘virus’ and with *no evidence* found it to cause AIDS. You will see later that there is also no evidence that any ‘virus’ causes any disease or that there is even such a thing as a ‘virus’ in the way it is said to exist. The claim to have ‘isolated’ the HIV ‘virus’ will be presented in its real context as we come to the shocking story – and it is a story – of SARS-CoV-2 and so will Montagnier’s assertion that he identified the full SARS-CoV-2 genome.

Hoax in the making

We can pick up the ‘Covid’ story in 2010 and the publication by the Rockefeller Foundation of a document called ‘Scenarios for the Future of Technology and International Development’. The inner circle of the Rockefeller family has been serving the Cult since John D. Rockefeller (1839-1937) made his fortune with Standard Oil. It is less well known that the same Rockefeller – the Bill Gates of his day – was responsible for establishing what is now referred to as ‘Big Pharma’, the global network of pharmaceutical companies that make outrageous profits dispensing scalpel and drug ‘medicine’ and are obsessed with pumping vaccines in ever-increasing number into as many human arms and backsides as possible. John D. Rockefeller was the driving force behind the creation of the ‘education’ system in the United States and elsewhere specifically designed to program the perceptions of generations thereafter. The Rockefeller family donated exceptionally valuable land in New York for the United Nations building and were central in establishing the World Health Organization in 1948 as an agency of the UN which was created from the start as a Trojan horse and stalking horse for world government. Now enter Bill Gates. His family and the Rockefellers have long been extremely close and I have seen genealogy which claims that if you go back far enough the two families fuse into the same bloodline. Gates has said that the Bill and Melinda Gates Foundation was inspired by the Rockefeller Foundation and why not

when both are serving the same Cult? Major tax-exempt foundations are overwhelmingly criminal enterprises in which Cult assets fund the Cult agenda in the guise of 'philanthropy' while avoiding tax in the process. Cult operatives can become mega-rich in their role of front men and women for the psychopaths at the inner core and they, too, have to be psychopaths to knowingly serve such evil. Part of the deal is that a big percentage of the wealth gleaned from representing the Cult has to be spent advancing the ambitions of the Cult and hence you have the Rockefeller Foundation, Bill and Melinda Gates Foundation (and *so* many more) and people like George Soros with his global Open Society Foundations spending their billions in pursuit of global Cult control. Gates is a global public face of the Cult with his interventions in world affairs including Big Tech influence; a central role in the 'Covid' and 'vaccine' scam; promotion of the climate change shakedown; manipulation of education; geoengineering of the skies; and his food-control agenda as the biggest owner of farmland in America, his GMO promotion and through other means. As one writer said: 'Gates monopolizes or wields disproportionate influence over the tech industry, global health and vaccines, agriculture and food policy (including biopiracy and fake food), weather modification and other climate technologies, surveillance, education and media.' The almost limitless wealth secured through Microsoft and other not-allowed-to-fail ventures (including vaccines) has been ploughed into a long, long list of Cult projects designed to enslave the entire human race. Gates and the Rockefellers have been working as one unit with the Rockefeller-established World Health Organization leading global 'Covid' policy controlled by Gates through his mouth-piece Tedros. Gates became the WHO's biggest funder when Trump announced that the American government would cease its donations, but Biden immediately said he would restore the money when he took office in January, 2021. The Gates Foundation (the Cult) owns through limitless funding the world health system and the major players across the globe in the 'Covid' hoax.

Okay, with that background we return to that Rockefeller Foundation document of 2010 headed 'Scenarios for the Future of Technology and International Development' and its 'imaginary' epidemic of a virulent and deadly influenza strain which infected 20 percent of the global population and killed eight million in seven months. The Rockefeller scenario was that the epidemic destroyed economies, closed shops, offices and other businesses and led to governments imposing fierce rules and restrictions that included mandatory wearing of face masks and body-temperature checks to enter communal spaces like railway stations and supermarkets. The document predicted that even after the height of the Rockefeller-envisaged epidemic the authoritarian rule would continue to deal with further pandemics, transnational terrorism, environmental crises and rising poverty. Now you may think that the Rockefellers are our modern-day seers or alternatively, and rather more likely, that they well knew what was planned a few years further on. Fascism had to be imposed, you see, to 'protect citizens from risk and exposure'. The Rockefeller scenario document said:

During the pandemic, national leaders around the world flexed their authority and imposed airtight rules and restrictions, from the mandatory wearing of face masks to body-temperature checks at the entries to communal spaces like train stations and supermarkets. Even after the pandemic faded, this more authoritarian control and oversight of citizens and their activities stuck and even intensified. In order to protect themselves from the spread of increasingly global problems – from pandemics and transnational terrorism to environmental crises and rising poverty – leaders around the world took a firmer grip on power.

At first, the notion of a more controlled world gained wide acceptance and approval. Citizens willingly gave up some of their sovereignty – and their privacy – to more paternalistic states in exchange for greater safety and stability. Citizens were more tolerant, and even eager, for top-down direction and oversight, and national leaders had more latitude to impose order in the ways they saw fit.

In developed countries, this heightened oversight took many forms: biometric IDs for all citizens, for example, and tighter regulation of key industries whose stability was deemed vital to national interests. In many developed countries, enforced cooperation with a suite of new regulations and agreements slowly but steadily restored both order and, importantly, economic growth.

There we have the prophetic Rockefellers in 2010 and three years later came their paper for the Global Health Summit in Beijing, China, when government representatives, the private sector, international organisations and groups met to discuss the next 100 years of 'global health'. The Rockefeller Foundation-funded paper was called 'Dreaming the Future of Health for the Next 100 Years and more prophecy ensued as it described a dystopian future: 'The abundance of data, digitally tracking and linking people may mean the 'death of privacy' and may replace physical interaction with transient, virtual connection, generating isolation and raising questions of how values are shaped in virtual networks.' Next in the 'Covid' hoax preparation sequence came a 'table top' simulation in 2018 for another 'imaginary' pandemic of a disease called Clade X which was said to kill 900 million people. The exercise was organised by the Gates-funded Johns Hopkins University's Center for Health Security in the United States and this is the very same university that has been compiling the disgustingly and systematically erroneous global figures for 'Covid' cases and deaths. Similar Johns Hopkins health crisis scenarios have included the Dark Winter exercise in 2001 and Atlantic Storm in 2005.

Nostradamus 201

For sheer predictive genius look no further prophecy-watchers than the Bill Gates-funded Event 201 held only six weeks before the 'coronavirus pandemic' is supposed to have broken out in China and Event 201 was based on a scenario of a global 'coronavirus pandemic'. Melinda Gates, the great man's missus, told the BBC that he had 'prepared for years' for a coronavirus pandemic which told us what we already knew. Nostradamugates had predicted in a TED talk in 2015 that a pandemic was coming that would kill a lot of people and demolish the world economy. My god, the man is a machine – possibly even literally. Now here he was only weeks before the real thing funding just such a simulated scenario and involving his friends and associates at Johns Hopkins, the World Economic Forum Cult-front of Klaus Schwab, the United Nations,

Johnson & Johnson, major banks, and officials from China and the Centers for Disease Control in the United States. What synchronicity – Johns Hopkins would go on to compile the fraudulent ‘Covid’ figures, the World Economic Forum and Schwab would push the ‘Great Reset’ in response to ‘Covid’, the Centers for Disease Control would be at the forefront of ‘Covid’ policy in the United States, Johnson & Johnson would produce a ‘Covid vaccine’, and everything would officially start just weeks later in China. Spooky, eh? They were even accurate in creating a simulation of a ‘virus’ pandemic because the ‘real thing’ would also be a simulation. Event 201 was not an exercise preparing for something that might happen; it was a rehearsal for what those in control knew was *going* to happen and very shortly. Hours of this simulation were posted on the Internet and the various themes and responses mirrored what would soon be imposed to transform human society. News stories were inserted and what they said would be commonplace a few weeks later with still more prophecy perfection. Much discussion focused on the need to deal with misinformation and the ‘anti-vax movement’ which is exactly what happened when the ‘virus’ arrived – was said to have arrived – in the West.

Cult-owned social media banned criticism and exposure of the official ‘virus’ narrative and when I said there *was* no ‘virus’ in early April, 2020, I was banned by one platform after another including YouTube, Facebook and later Twitter. The mainstream broadcast media in Britain was in effect banned from interviewing me by the Tony-Blair-created government broadcasting censor Ofcom headed by career government bureaucrat Melanie Dawes who was appointed just as the ‘virus’ hoax was about to play out in January, 2020. At the same time the Ickonic media platform was using Vimeo, another ultra-Zionist-owned operation, while our own player was being created and they deleted in an instant hundreds of videos, documentaries, series and shows to confirm their unbelievable vindictiveness. We had copies, of course, and they had to be restored one by one when our player was ready. These people have no class. Sabbatian Facebook promised free advertisements for the Gates-

controlled World Health Organization narrative while deleting ‘false claims and conspiracy theories’ to stop ‘misinformation’ about the alleged coronavirus. All these responses could be seen just a short while earlier in the scenarios of Event 201. Extreme censorship was absolutely crucial for the Cult because the official story was so ridiculous and unsupportable by the evidence that it could never survive open debate and the free-flow of information and opinion. If you can’t win a debate then don’t have one is the Cult’s approach throughout history. Facebook’s little boy front man – front boy – Mark Zuckerberg equated ‘credible and accurate information’ with official sources and exposing their lies with ‘misinformation’.

Silencing those that can see

The censorship dynamic of Event 201 is now the norm with an army of narrative-supporting ‘fact-checker’ organisations whose entire reason for being is to tell the public that official narratives are true and those exposing them are lying. One of the most appalling of these ‘fact-checkers’ is called NewsGuard founded by ultra-Zionist Americans Gordon Crovitz and Steven Brill. Crovitz is a former publisher of *The Wall Street Journal*, former Executive Vice President of Dow Jones, a member of the Council on Foreign Relations (CFR), and on the board of the American Association of Rhodes Scholars. The CFR and Rhodes Scholarships, named after Rothschild agent Cecil Rhodes who plundered the gold and diamonds of South Africa for his masters and the Cult, have featured widely in my books. NewsGuard don’t seem to like me for some reason – I really can’t think why – and they have done all they can to have me censored and discredited which is, to quote an old British politician, like being savaged by a dead sheep. They are, however, like all in the censorship network, very well connected and funded by organisations themselves funded by, or connected to, Bill Gates. As you would expect with anything associated with Gates NewsGuard has an offshoot called HealthGuard which ‘fights online health care hoaxes’. How very kind. Somehow the NewsGuard European Managing Director Anna-Sophie Harling, a remarkably young-

looking woman with no broadcasting experience and little hands-on work in journalism, has somehow secured a position on the 'Content Board' of UK government broadcast censor Ofcom. An executive of an organisation seeking to discredit dissidents of the official narratives is making decisions for the government broadcast 'regulator' about content?? Another appalling 'fact-checker' is Full Fact funded by George Soros and global censors Google and Facebook.

It's amazing how many activists in the 'fact-checking', 'anti-hate', arena turn up in government-related positions – people like UK Labour Party activist Imran Ahmed who heads the Center for Countering Digital Hate founded by people like Morgan McSweeney, now chief of staff to the Labour Party's hapless and useless 'leader' Keir Starmer. Digital Hate – which is what it really is – uses the American spelling of Center to betray its connection to a transatlantic network of similar organisations which in 2020 shapeshifted from attacking people for 'hate' to attacking them for questioning the 'Covid' hoax and the dangers of the 'Covid vaccine'. It's just a coincidence, you understand. This is one of Imran Ahmed's hysterical statements: 'I would go beyond calling anti-vaxxers conspiracy theorists to say they are an extremist group that pose a national security risk.' No one could ever accuse this prat of understatement and he's including in that those parents who are now against vaccines after their children were damaged for life or killed by them. He's such a nice man. Ahmed does the rounds of the Woke media getting soft-ball questions from spineless 'journalists' who never ask what right he has to campaign to destroy the freedom of speech of others while he demands it for himself. There also seems to be an overrepresentation in Ofcom of people connected to the narrative-worshipping BBC. This incredible global network of narrative-support was super-vital when the 'Covid' hoax was played in the light of the mega-whopper lies that have to be defended from the spotlight cast by the most basic intelligence.

Setting the scene

The Cult plays the long game and proceeds step-by-step ensuring that everything is in place before major cards are played and they don't come any bigger than the 'Covid' hoax. The psychopaths can't handle events where the outcome isn't certain and as little as possible – preferably nothing – is left to chance. Politicians, government and medical officials who would follow direction were brought to illusory power in advance by the Cult web whether on the national stage or others like state governors and mayors of America. For decades the dynamic between officialdom, law enforcement and the public was changed from one of service to one of control and dictatorship. Behaviour manipulation networks established within government were waiting to impose the coming 'Covid' rules and regulations specifically designed to subdue and rewire the psyche of the people in the guise of protecting health. These included in the UK the Behavioural Insights Team part-owned by the British government Cabinet Office; the Scientific Pandemic Insights Group on Behaviours (SPI-B); and a whole web of intelligence and military groups seeking to direct the conversation on social media and control the narrative. Among them are the cyberwarfare (on the people) 77th Brigade of the British military which is also coordinated through the Cabinet Office as civilian and military leadership continues to combine in what they call the Fusion Doctrine. The 77th Brigade is a British equivalent of the infamous Israeli (Sabbatian) military cyberwarfare and Internet manipulation operation Unit 8200 which I expose at length in *The Trigger*. Also carefully in place were the medical and science advisers to government – many on the payroll past or present of Bill Gates – and a whole alternative structure of unelected government stood by to take control when elected parliaments were effectively closed down once the 'Covid' card was slammed on the table. The structure I have described here and so much more was installed in every major country through the Cult networks. The top-down control hierarchy looks like this: The Cult – Cult-owned Gates – the World Health Organization and Tedros – Gates-funded or controlled chief medical officers and science 'advisers' (dictators) in each country –

political 'leaders' – law enforcement – The People. Through this simple global communication and enforcement structure the policy of the Cult could be imposed on virtually the entire human population so long as they acquiesced to the fascism. With everything in place it was time for the button to be pressed in late 2019/early 2020.

These were the prime goals the Cult had to secure for its will to prevail:

- 1) Locking down economies, closing all but designated 'essential' businesses (Cult-owned corporations were 'essential'), and putting the population under house arrest was an imperative to destroy independent income and employment and ensure dependency on the Cult-controlled state in the Hunger Games Society. Lockdowns had to be established as the global blueprint from the start to respond to the 'virus' and followed by pretty much the entire world.
- 2) The global population had to be terrified into believing in a deadly 'virus' that didn't actually exist so they would unquestioningly obey authority in the belief that authority must know how best to protect them and their families. Software salesman Gates would suddenly morph into the world's health expert and be promoted as such by the Cult-owned media.
- 3) A method of testing that wasn't testing for the 'virus', but was only claimed to be, had to be in place to provide the illusion of 'cases' and subsequent 'deaths' that had a very different cause to the 'Covid-19' that would be scribbled on the death certificate.
- 4) Because there was no 'virus' and the great majority testing positive with a test not testing for the 'virus' would have no symptoms of anything the lie had to be sold that people without symptoms (without the 'virus') could still pass it on to others. This was crucial to justify for the first time quarantining – house arresting – healthy people. Without this the economy-destroying lockdown of *everybody* could not have been credibly sold.
- 5) The 'saviour' had to be seen as a vaccine which beyond evil drug companies were working like angels of mercy to develop as quickly as possible, with all corners cut, to save the day. The public must absolutely not know that the 'vaccine' had nothing to do with a 'virus' or that the contents were ready and waiting with a very different motive long before the 'Covid' card was even lifted from the pack.

I said in March, 2020, that the 'vaccine' would have been created way ahead of the 'Covid' hoax which justified its use and the following December an article in the New York *Intelligencer* magazine said the Moderna 'vaccine' had been 'designed' by

January, 2020. This was 'before China had even acknowledged that the disease could be transmitted from human to human, more than a week before the first confirmed coronavirus case in the United States'. The article said that by the time the first American death was announced a month later 'the vaccine had already been manufactured and shipped to the National Institutes of Health for the beginning of its Phase I clinical trial'. The 'vaccine' was actually 'designed' long before that although even with this timescale you would expect the article to ask how on earth it could have been done that quickly. Instead it asked why the 'vaccine' had not been rolled out then and not months later. Journalism in the mainstream is truly dead. I am going to detail in the next chapter why the 'virus' has never existed and how a hoax on that scale was possible, but first the foundation on which the Big Lie of 'Covid' was built.

The test that doesn't test

Fraudulent 'testing' is the bottom line of the whole 'Covid' hoax and was the means by which a 'virus' that did not exist *appeared* to exist. They could only achieve this magic trick by using a test not testing for the 'virus'. To use a test that *was* testing for the 'virus' would mean that every test would come back negative given there was no 'virus'. They chose to exploit something called the RT-PCR test invented by American biochemist Kary Mullis in the 1980s who said publicly that his PCR test ... *cannot detect infectious disease*. Yes, the 'test' used worldwide to detect infectious 'Covid' to produce all the illusory 'cases' and 'deaths' compiled by Johns Hopkins and others *cannot detect infectious disease*. This fact came from the mouth of the man who invented PCR and was awarded the Nobel Prize in Chemistry in 1993 for doing so. Sadly, and incredibly conveniently for the Cult, Mullis died in August, 2019, at the age of 74 just before his test would be fraudulently used to unleash fascism on the world. He was said to have died from pneumonia which was an irony in itself. A few months later he would have had 'Covid-19' on his death certificate. I say the timing of his death was convenient because had he lived Mullis, a brilliant, honest and decent man, would have been

vociferously speaking out against the use of his test to detect 'Covid' when it was never designed, or able, to do that. I know that to be true given that Mullis made the same point when his test was used to 'detect' – not detect – HIV. He had been seriously critical of the Gallo/Montagnier claim to have isolated the HIV 'virus' and shown it to cause AIDS for which Mullis said there was no evidence. AIDS is actually not a disease but a series of diseases from which people die all the time. When they die from those *same diseases* after a positive 'test' for HIV then AIDS goes on their death certificate. I think I've heard that before somewhere. Countries instigated a policy with 'Covid' that anyone who tested positive with a test not testing for the 'virus' and died of any other cause within 28 days and even longer 'Covid-19' had to go on the death certificate. Cases have come from the test that can't test for infectious disease and the deaths are those who have died of *anything* after testing positive with a test not testing for the 'virus'. I'll have much more later about the death certificate scandal.

Mullis was deeply dismissive of the now US 'Covid' star Anthony Fauci who he said was a liar who didn't know anything about anything – 'and I would say that to his face – nothing.' He said of Fauci: 'The man thinks he can take a blood sample, put it in an electron microscope and if it's got a virus in there you'll know it – he doesn't understand electron microscopy and he doesn't understand medicine and shouldn't be in a position like he's in.' That position, terrifyingly, has made him the decider of 'Covid' fascism policy on behalf of the Cult in his role as director since 1984 of the National Institute of Allergy and Infectious Diseases (NIAID) while his record of being wrong is laughable; but being wrong, so long as it's the *right kind* of wrong, is why the Cult loves him. He'll say anything the Cult tells him to say. Fauci was made Chief Medical Adviser to the President immediately Biden took office. Biden was installed in the White House by Cult manipulation and one of his first decisions was to elevate Fauci to a position of even more control. This is a coincidence? Yes, and I identify as a flamenco dancer called Lola. How does such an incompetent criminal like Fauci remain in that

pivotal position in American health since *the 1980s*? When you serve the Cult it looks after you until you are surplus to requirements. Kary Mullis said prophetically of Fauci and his like: 'Those guys have an agenda and it's not an agenda we would like them to have ... they make their own rules, they change them when they want to, and Tony Fauci does not mind going on television in front of the people who pay his salary and lie directly into the camera.' Fauci has done that almost daily since the 'Covid' hoax began. Lying is in Fauci's DNA. To make the situation crystal clear about the PCR test this is a direct quote from its inventor Kary Mullis:

It [the PCR test] doesn't tell you that you're sick and doesn't tell you that the thing you ended up with was really going to hurt you ...'

Ask yourself why governments and medical systems the world over have been using this very test to decide who is 'infected' with the SARS-CoV-2 'virus' and the alleged disease it allegedly causes, 'Covid-19'. The answer to that question will tell you what has been going on. By the way, here's a little show-stopper – the 'new' SARS-CoV-2 'virus' was 'identified' as such right from the start using ... *the PCR test not testing for the 'virus'*. If you are new to this and find that shocking then stick around. I have hardly started yet. Even worse, other 'tests', like the 'Lateral Flow Device' (LFD), are considered so useless that they have to be *confirmed* by the PCR test! Leaked emails written by Ben Dyson, adviser to UK 'Health' Secretary Matt Hancock, said they were 'dangerously unreliable'. Dyson, executive director of strategy at the Department of Health, wrote: 'As of today, someone who gets a positive LFD result in (say) London has at best a 25 per cent chance of it being a true positive, but if it is a self-reported test potentially as low as 10 per cent (on an optimistic assumption about specificity) or as low as 2 per cent (on a more pessimistic assumption).' These are the 'tests' that schoolchildren and the public are being urged to have twice a week or more and have to isolate if they get a positive. Each fake positive goes in the statistics as a 'case' no matter how ludicrously inaccurate and the

'cases' drive lockdown, masks and the pressure to 'vaccinate'. The government said in response to the email leak that the 'tests' were accurate which confirmed yet again what shocking bloody liars they are. The real false positive rate is *100 percent* as we'll see. In another 'you couldn't make it up' the UK government agreed to pay £2.8 billion to California's Innova Medical Group to supply the irrelevant lateral flow tests. The company's primary test-making centre is in China. Innova Medical Group, established in March, 2020, is owned by Pasaca Capital Inc, chaired by Chinese-American millionaire Charles Huang who was born in Wuhan.

How it works – and how it doesn't

The RT-PCR test, known by its full title of Polymerase chain reaction, is used across the world to make millions, even billions, of copies of a DNA/RNA genetic information sample. The process is called 'amplification' and means that a tiny sample of genetic material is amplified to bring out the detailed content. I stress that it is not testing for an infectious disease. It is simply amplifying a sample of genetic material. In the words of Kary Mullis: 'PCR is ... just a process that's used to make a whole lot of something out of something.' To emphasise the point companies that make the PCR tests circulated around the world to 'test' for 'Covid' warn on the box that it can't be used to detect 'Covid' or infectious disease and is for research purposes only. It's okay, rest for a minute and you'll be fine. This is the test that produces the 'cases' and 'deaths' that have been used to destroy human society. All those global and national medical and scientific 'experts' demanding this destruction to 'save us' *KNOW* that the test is not testing for the 'virus' and the cases and deaths they claim to be real are an almost unimaginable fraud. Every one of them and so many others including politicians and psychopaths like Gates and Tedros must be brought before Nuremburg-type trials and jailed for the rest of their lives. The more the genetic sample is amplified by PCR the more elements of that material become sensitive to the test and by that I don't mean sensitive for a 'virus' but for elements of the genetic material which

is *naturally* in the body or relates to remnants of old conditions of various kinds lying dormant and causing no disease. Once the amplification of the PCR reaches a certain level *everyone* will test positive. So much of the material has been made sensitive to the test that everyone will have some part of it in their body. Even lying criminals like Fauci have said that once PCR amplifications pass 35 cycles everything will be a false positive that cannot be trusted for the reasons I have described. I say, like many proper doctors and scientists, that 100 percent of the 'positives' are false, but let's just go with Fauci for a moment.

He says that any amplification over 35 cycles will produce false positives and yet the US Centers for Disease Control (CDC) and Food and Drug Administration (FDA) have recommended up to 40 *cycles* and the National Health Service (NHS) in Britain admitted in an internal document for staff that it was using 45 *cycles* of amplification. A long list of other countries has been doing the same and at least one 'testing' laboratory has been using 50 *cycles*. Have you ever heard a doctor, medical 'expert' or the media ask what level of amplification has been used to claim a 'positive'. The 'test' comes back 'positive' and so you have the 'virus', end of story. Now we can see how the government in Tanzania could send off samples from a goat and a pawpaw fruit under human names and both came back positive for 'Covid-19'. Tanzania president John Magufuli mocked the 'Covid' hysteria, the PCR test and masks and refused to import the DNA-manipulating 'vaccine'. The Cult hated him and an article sponsored by the Bill Gates Foundation appeared in the London *Guardian* in February, 2021, headed 'It's time for Africa to rein in Tanzania's anti-vaxxer president'. Well, 'reined in' he shortly was. Magufuli appeared in good health, but then, in March, 2021, he was dead at 61 from 'heart failure'. He was replaced by Samia Hassan Suhulu who is connected to Klaus Schwab's World Economic Forum and she immediately reversed Magufuli's 'Covid' policy. A sample of cola tested positive for 'Covid' with the PCR test in Germany while American actress and singer-songwriter Erykah Badu tested positive in one nostril and negative in the other. Footballer Ronaldo called

the PCR test 'bullshit' after testing positive three times and being forced to quarantine and miss matches when there was nothing wrong with him. The mantra from Tedros at the World Health Organization and national governments (same thing) has been test, test, test. They know that the more tests they can generate the more fake 'cases' they have which go on to become 'deaths' in ways I am coming to. The UK government has its Operation Moonshot planned to test multiple millions every day in workplaces and schools with free tests for everyone to use twice a week at home in line with the Cult plan from the start to make testing part of life. A government advertisement for an 'Interim Head of Asymptomatic Testing Communication' said the job included responsibility for delivering a 'communications strategy' (propaganda) 'to support the expansion of asymptomatic testing that *'normalises testing as part of everyday life'*'. More tests means more fake 'cases', 'deaths' and fascism. I have heard of, and from, many people who booked a test, couldn't turn up, and yet got a positive result through the post for a test they'd never even had. The whole thing is crazy, but for the Cult there's method in the madness. Controlling and manipulating the level of amplification of the test means the authorities can control whenever they want the number of apparent 'cases' and 'deaths'. If they want to justify more fascist lockdown and destruction of livelihoods they keep the amplification high. If they want to give the illusion that lockdowns and the 'vaccine' are working then they lower the amplification and 'cases' and 'deaths' will appear to fall. In January, 2021, the Cult-owned World Health Organization suddenly warned laboratories about over-amplification of the test and to lower the threshold. Suddenly headlines began appearing such as: 'Why ARE "Covid" cases plummeting?' This was just when the vaccine rollout was underway and I had predicted months before they would make cases appear to fall through amplification tampering when the 'vaccine' came. These people are so predictable.

Cow vaccines?

The question must be asked of what is on the test swabs being poked far up the nose of the population to the base of the brain? A nasal swab punctured one woman's brain and caused it to leak fluid. Most of these procedures are being done by people with little training or medical knowledge. Dr Lorraine Day, former orthopaedic trauma surgeon and Chief of Orthopaedic Surgery at San Francisco General Hospital, says the tests are really a '*vaccine*'. Cows have long been vaccinated this way. She points out that masks have to cover the nose and the mouth where it is claimed the '*virus*' exists in saliva. Why then don't they take saliva from the mouth as they do with a DNA test instead of pushing a long swab up the nose towards the brain? The ethmoid bone separates the nasal cavity from the brain and within that bone is the cribriform plate. Dr Day says that when the swab is pushed up against this plate and twisted the procedure is '*depositing things back there*'. She claims that among these '*things*' are nanoparticles that can enter the brain. Researchers have noted that a team at the Gates-funded Johns Hopkins have designed tiny, star-shaped micro-devices that can latch onto intestinal mucosa and release drugs into the body. Mucosa is the thin skin that covers the inside surface of parts of the body such as *the nose* and mouth and produces mucus to protect them. The Johns Hopkins micro-devices are called '*theragrippers*' and were '*inspired*' by a parasitic worm that digs its sharp teeth into a host's intestines. Nasal swabs are also coated in the sterilisation agent ethylene oxide. The US National Cancer Institute posts this explanation on its website:

At room temperature, ethylene oxide is a flammable colorless gas with a sweet odor. It is used primarily to produce other chemicals, including antifreeze. In smaller amounts, ethylene oxide is used as a pesticide and a sterilizing agent. The ability of ethylene oxide to damage DNA makes it an effective sterilizing agent but also accounts for its cancer-causing activity.

The Institute mentions lymphoma and leukaemia as cancers most frequently reported to be associated with occupational exposure to ethylene oxide along with stomach and breast cancers. How does anyone think this is going to work out with the constant testing

regime being inflicted on adults and children at home and at school that will accumulate in the body anything that's on the swab?

Doctors know best

It is vital for people to realise that 'hero' doctors 'know' only what the Big Pharma-dominated medical authorities tell them to 'know' and if they refuse to 'know' what they are told to 'know' they are out the door. They are mostly not physicians or healers, but repeaters of the official narrative – or else. I have seen alleged professional doctors on British television make shocking statements that we are supposed to take seriously. One called 'Dr' Amir Khan, who is actually telling patients how to respond to illness, said that men could take the birth pill to 'help slow down the effects of Covid-19'. In March, 2021, another ridiculous 'Covid study' by an American doctor proposed injecting men with the female sex hormone progesterone as a 'Covid' treatment. British doctor Nighat Arif told the BBC that face coverings were now going to be part of ongoing normal. Yes, the vaccine protects you, she said (evidence?) ... but the way to deal with viruses in the community was always going to come down to hand washing, face covering and keeping a physical distance. That's not what we were told before the 'vaccine' was circulating. Arif said she couldn't imagine ever again going on the underground or in a lift without a mask. I was just thanking my good luck that she was not my doctor when she said – in March, 2021 – that if 'we are *behaving* and we are doing all the right things' she thought we could 'have our nearest and dearest around us at home ... around *Christmas* and *New Year*! Her patronising delivery was the usual school teacher talking to six-year-olds as she repeated every government talking point and probably believed them all. If we have learned anything from the 'Covid' experience surely it must be that humanity's perception of doctors needs a fundamental rethink. NHS 'doctor' Sara Kayat told her television audience that the 'Covid vaccine' would '100 percent prevent hospitalisation and death'. Not even Big Pharma claimed that. We have to stop taking 'experts' at their word without question when so many of them are

clueless and only repeating the party line on which their careers depend. That is not to say there are not brilliant doctors – there are and I have spoken to many of them since all this began – but you won't see them in the mainstream media or quoted by the psychopaths and yes-people in government.

Remember the name – Christian Drosten

German virologist Christian Drosten, Director of Charité Institute of Virology in Berlin, became a national star after the pandemic hoax began. He was feted on television and advised the German government on 'Covid' policy. Most importantly to the wider world Drosten led a group that produced the 'Covid' testing protocol for the PCR test. What a remarkable feat given the PCR cannot test for infectious disease and even more so when you think that Drosten said that his method of testing for SARS-CoV-2 was developed 'without having virus material available'. *He developed a test for a 'virus' that he didn't have and had never seen.* Let that sink in as you survey the global devastation that came from what he did. The whole catastrophe of Drosten's 'test' was based on the alleged genetic sequence published by Chinese scientists on the Internet. We will see in the next chapter that this alleged 'genetic sequence' has never been produced by China or anyone and cannot be when there *is no* SARS-CoV-2. Drosten, however, doesn't seem to let little details like that get in the way. He was the lead author with Victor Corman from the same Charité Hospital of the paper 'Detection of 2019 novel coronavirus (2019-nCoV) by real-time PCR' published in a magazine called *Eurosurveillance*. This became known as the Corman-Drosten paper. In November, 2020, with human society devastated by the effects of the Corman-Drosten test baloney, the protocol was publicly challenged by 22 international scientists and independent researchers from Europe, the United States, and Japan. Among them were senior molecular geneticists, biochemists, immunologists, and microbiologists. They produced a document headed 'External peer review of the RTPCR test to detect SARS-Cov-2 Reveals 10 Major Flaws At The Molecular and Methodological Level: Consequences

For False-Positive Results'. The flaws in the Corman-Drosten test included the following:

- The test is non-specific because of erroneous design
- Results are enormously variable
- The test is unable to discriminate between the whole 'virus' and viral fragments
- It doesn't have positive or negative controls
- The test lacks a standard operating procedure
- It is unsupported by proper peer view

The scientists said the PCR 'Covid' testing protocol was not founded on science and they demanded the Corman-Drosten paper be retracted by *Eurosurveillance*. They said all present and previous Covid deaths, cases, and 'infection rates' should be subject to a massive retroactive inquiry. Lockdowns and travel restrictions should be reviewed and relaxed and those diagnosed through PCR to have 'Covid-19' should not be forced to isolate. Dr Kevin Corbett, a health researcher and nurse educator with a long academic career producing a stream of peer-reviewed publications at many UK universities, made the same point about the PCR test debacle. He said of the scientists' conclusions: 'Every scientific rationale for the development of that test has been totally destroyed by this paper. It's like Hiroshima/Nagasaki to the Covid test.' He said that China hadn't given them an isolated 'virus' when Drosten developed the test. Instead they had developed the test from *a sequence in a gene bank*.' Put another way ... *they made it up!* The scientists were supported in this contention by a Portuguese appeals court which ruled in November, 2020, that PCR tests are unreliable and it is unlawful to quarantine people based solely on a PCR test. The point about China not providing an isolated virus must be true when the 'virus' has never been isolated to this day and the consequences of that will become clear. Drosten and company produced this useless 'protocol' right on cue in January, 2020, just as the 'virus' was said to

be moving westward and it somehow managed to successfully pass a peer-review in 24 hours. In other words there was no peer-review for a test that would be used to decide who had 'Covid' and who didn't across the world. The Cult-created, Gates-controlled World Health Organization immediately recommended all its nearly 200 member countries to use the Drosten PCR protocol to detect 'cases' and 'deaths'. The sting was underway and it continues to this day.

So who is this Christian Drosten that produced the means through which death, destruction and economic catastrophe would be justified? His education background, including his doctoral thesis, would appear to be somewhat shrouded in mystery and his track record is dire as with another essential player in the 'Covid' hoax, the Gates-funded Professor Neil Ferguson at the Gates-funded Imperial College in London of whom more shortly. Drosten predicted in 2003 that the alleged original SARS 'virus' (SARS-1) was an epidemic that could have serious effects on economies and an effective vaccine would take at least two years to produce. Drosten's answer to every alleged 'outbreak' is a vaccine which you won't be shocked to know. What followed were just 774 official deaths worldwide and none in Germany where there were only nine cases. That is even if you believe there ever was a SARS 'virus' when the evidence is zilch and I will expand on this in the next chapter. Drosten claims to be co-discoverer of 'SARS-1' and developed a test for it in 2003. He was screaming warnings about 'swine flu' in 2009 and how it was a widespread infection far more severe than any dangers from a vaccine could be and people should get vaccinated. It would be helpful for Drosten's vocal chords if he simply recorded the words 'the virus is deadly and you need to get vaccinated' and copies could be handed out whenever the latest made-up threat comes along. Drosten's swine flu epidemic never happened, but Big Pharma didn't mind with governments spending hundreds of millions on vaccines that hardly anyone bothered to use and many who did wished they hadn't. A study in 2010 revealed that the risk of dying from swine flu, or H1N1, was no higher than that of the annual seasonal flu which is what at least most of 'it' really was as in

the case of 'Covid-19'. A media investigation into Drosten asked how with such a record of inaccuracy he could be *the* government adviser on these issues. The answer to that question is the same with Drosten, Ferguson and Fauci – they keep on giving the authorities the 'conclusions' and 'advice' they want to hear. Drosten certainly produced the goods for them in January, 2020, with his PCR protocol garbage and provided the foundation of what German internal medicine specialist Dr Claus Köhnlein, co-author of *Virus Mania*, called the 'test pandemic'. The 22 scientists in the *Eurosurveillance* challenge called out conflicts of interest within the Drosten 'protocol' group and with good reason. Olfert Landt, a regular co-author of Drosten 'studies', owns the biotech company TIB Molbiol Syntheselabor GmbH in Berlin which manufactures and sells the tests that Drosten and his mates come up with. They have done this with SARS, Enterotoxigenic E. coli (ETEC), MERS, Zika 'virus', yellow fever, and now 'Covid'. Landt told the *Berliner Zeitung* newspaper:

The testing, design and development came from the Charité [Drosten and Corman]. We simply implemented it immediately in the form of a kit. And if we don't have the virus, which originally only existed in Wuhan, we can make a synthetic gene to simulate the genome of the virus. That's what we did very quickly.

This is more confirmation that the Drosten test was designed without access to the 'virus' and only a synthetic simulation which is what SARS-CoV-2 really is – a computer-generated synthetic fiction. It's quite an enterprise they have going here. A Drosten team decides what the test for something should be and Landt's biotech company flogs it to governments and medical systems across the world. His company must have made an absolute fortune since the 'Covid' hoax began. Dr Reiner Fuellmich, a prominent German consumer protection trial lawyer in Germany and California, is on Drosten's case and that of Tedros at the World Health Organization for crimes against humanity with a class-action lawsuit being prepared in the United States and other legal action in Germany.

Why China?

Scamming the world with a 'virus' that doesn't exist would seem impossible on the face of it, but not if you have control of the relatively few people that make policy decisions and the great majority of the global media. Remember it's not about changing 'real' reality it's about controlling *perception* of reality. You don't have to make something happen you only have to make people *believe* that it's happening. Renegade Minds understand this and are therefore much harder to swindle. 'Covid-19' is not a 'real' 'virus'. It's a mind virus, like a computer virus, which has infected the minds, not the bodies, of billions. It all started, publically at least, in China and that alone is of central significance. The Cult was behind the revolution led by its asset Mao Zedong, or Chairman Mao, which established the People's Republic of China on October 1st, 1949. It should have been called The Cult's Republic of China, but the name had to reflect the recurring illusion that vicious dictatorships are run by and for the people (see all the 'Democratic Republics' controlled by tyrants). In the same way we have the 'Biden' Democratic Republic of America officially ruled by a puppet tyrant (at least temporarily) on behalf of Cult tyrants. The creation of Mao's merciless communist/fascist dictatorship was part of a frenzy of activity by the Cult at the conclusion of World War Two which, like the First World War, it had instigated through its assets in Germany, Britain, France, the United States and elsewhere. Israel was formed in 1948; the Soviet Union expanded its 'Iron Curtain' control, influence and military power with the Warsaw Pact communist alliance in 1955; the United Nations was formed in 1945 as a Cult precursor to world government; and a long list of world bodies would be established including the World Health Organization (1948), World Trade Organization (1948 under another name until 1995), International Monetary Fund (1945) and World Bank (1944). Human society was redrawn and hugely centralised in the global Problem-Reaction-Solution that was World War Two. All these changes were significant. Israel would become the headquarters of the Sabbatians

and the revolution in China would prepare the ground and control system for the events of 2019/2020.

Renegade Minds know there are no borders except for public consumption. The Cult is a seamless, borderless global entity and to understand the game we need to put aside labels like borders, nations, countries, communism, fascism and democracy. These delude the population into believing that countries are ruled within their borders by a government of whatever shade when these are mere agencies of a global power. America's illusion of democracy and China's communism/fascism are subsidiaries – vehicles – for the same agenda. We may hear about conflict and competition between America and China and on the lower levels that will be true; but at the Cult level they are branches of the same company in the way of the McDonald's example I gave earlier. I have tracked in the books over the years support by US governments of both parties for Chinese Communist Party infiltration of American society through allowing the sale of land, even military facilities, and the acquisition of American business and university influence. All this is underpinned by the infamous stealing of intellectual property and technological know-how. Cult-owned Silicon Valley corporations waive their fraudulent 'morality' to do business with human-rights-free China; Cult-controlled Disney has become China's PR department; and China in effect owns 'American' sports such as basketball which depends for much of its income on Chinese audiences. As a result any sports player, coach or official speaking out against China's horrific human rights record is immediately condemned or fired by the China-worshipping National Basketball Association. One of the first acts of China-controlled Biden was to issue an executive order telling federal agencies to stop making references to the 'virus' by the 'geographic location of its origin'. Long-time Congressman Jerry Nadler warned that criticising China, America's biggest rival, leads to hate crimes against Asian people in the United States. So shut up you bigot. China is fast closing in on Israel as a country that must not be criticised which is apt, really, given that Sabbatians control them both. The two countries have

developed close economic, military, technological and strategic ties which include involvement in China's 'Silk Road' transport and economic initiative to connect China with Europe. Israel was the first country in the Middle East to recognise the establishment of Mao's tyranny in 1950 months after it was established.

Project Wuhan – the 'Covid' Psyop

I emphasise again that the Cult plays the long game and what is happening to the world today is the result of centuries of calculated manipulation following a script to take control step-by-step of every aspect of human society. I will discuss later the common force behind all this that has spanned those centuries and thousands of years if the truth be told. Instigating the Mao revolution in China in 1949 with a 2020 'pandemic' in mind is not only how they work – the 71 years between them is really quite short by the Cult's standards of manipulation preparation. The reason for the Cult's Chinese revolution was to create a fiercely-controlled environment within which an extreme structure for human control could be incubated to eventually be unleashed across the world. We have seen this happen since the 'pandemic' emerged from China with the Chinese control-structure founded on AI technology and tyrannical enforcement sweep across the West. Until the moment when the Cult went for broke in the West and put its fascism on public display Western governments had to pay some lip-service to freedom and democracy to not alert too many people to the tyranny-in-the-making. Freedoms were more subtly eroded and power centralised with covert government structures put in place waiting for the arrival of 2020 when that smokescreen of 'freedom' could be dispensed with. The West was not able to move towards tyranny before 2020 anything like as fast as China which was created as a tyranny and had no limits on how fast it could construct the Cult's blueprint for global control. When the time came to impose that structure on the world it was the same Cult-owned Chinese communist/fascist government that provided the excuse – the 'Covid pandemic'. It was absolutely crucial to the Cult plan for the Chinese response to the 'pandemic' –

draconian lockdowns of the entire population – to become the blueprint that Western countries would follow to destroy the livelihoods and freedom of their people. This is why the Cult-owned, Gates-owned, WHO Director-General Tedros said early on:

The Chinese government is to be congratulated for the extraordinary measures it has taken to contain the outbreak. China is actually setting a new standard for outbreak response and it is not an exaggeration.

Forbes magazine said of China: ‘... those measures protected untold millions from getting the disease’. The Rockefeller Foundation ‘epidemic scenario’ document in 2010 said ‘prophetically’:

However, a few countries did fare better – China in particular. The Chinese government’s quick imposition and enforcement of mandatory quarantine for all citizens, as well as its instant and near-hermetic sealing off of all borders, saved millions of lives, stopping the spread of the virus far earlier than in other countries and enabling a swifter post-pandemic recovery.

Once again – *spooky*.

The first official story was the ‘bat theory’ or rather the bat diversion. The source of the ‘virus outbreak’ we were told was a ‘wet market’ in Wuhan where bats and other animals are bought and eaten in horrifically unhygienic conditions. Then another story emerged through the alternative media that the ‘virus’ had been released on purpose or by accident from a BSL-4 (biosafety level 4) laboratory in Wuhan not far from the wet market. The lab was reported to create and work with lethal concoctions and bioweapons. Biosafety level 4 is the highest in the World Health Organization system of safety and containment. Renegade Minds are aware of what I call designer manipulation. The ideal for the Cult is for people to buy its prime narrative which in the opening salvos of the ‘pandemic’ was the wet market story. It knows, however, that there is now a considerable worldwide alternative media of researchers sceptical of anything governments say and they are often given a version of events in a form they can perceive as credible while misdirecting them from the real truth. In this case let them

think that the conspiracy involved is a 'bioweapon virus' released from the Wuhan lab to keep them from the real conspiracy – *there is no 'virus'*. The WHO's current position on the source of the outbreak at the time of writing appears to be: 'We haven't got a clue, mate.' This is a good position to maintain mystery and bewilderment. The inner circle will know where the 'virus' came from – *nowhere*. The bottom line was to ensure the public believed there *was* a 'virus' and it didn't much matter if they thought it was natural or had been released from a lab. The belief that there was a 'deadly virus' was all that was needed to trigger global panic and fear. The population was terrified into handing their power to authority and doing what they were told. They had to or they were 'all gonna die'.

In March, 2020, information began to come my way from real doctors and scientists and my own additional research which had my intuition screaming: 'Yes, that's it! *There is no virus.*' The 'bioweapon' was not the 'virus'; it was the '*vaccine*' already being talked about that would be the bioweapon. My conclusion was further enhanced by happenings in Wuhan. The 'virus' was said to be sweeping the city and news footage circulated of people collapsing in the street (which they've never done in the West with the same 'virus'). The Chinese government was building 'new hospitals' in a matter of ten days to 'cope with demand' such was the virulent nature of the 'virus'. Yet in what seemed like no time the 'new hospitals' closed – even if they even opened – and China declared itself 'virus-free'. It was back to business as usual. This was more propaganda to promote the Chinese draconian lockdowns in the West as the way to 'beat the virus'. Trouble was that we subsequently had lockdown after lockdown, but never business as usual. As the people of the West and most of the rest of the world were caught in an ever-worsening spiral of lockdown, social distancing, masks, isolated old people, families forced apart, and livelihood destruction, it was party-time in Wuhan. Pictures emerged of thousands of people enjoying pool parties and concerts. It made no sense until you realised there never was a 'virus' and the

whole thing was a Cult set-up to transform human society out of one of its major global strongholds – China.

How is it possible to deceive virtually the entire world population into believing there is a deadly virus when there is not even a 'virus' let alone a deadly one? It's nothing like as difficult as you would think and that's clearly true because it happened.

Postscript: See end of book Postscript for more on the 'Wuhan lab virus release' story which the authorities and media were pushing heavily in the summer of 2021 to divert attention from the truth that the 'Covid virus' is pure invention.

CHAPTER FIVE

There is no 'virus'

You can fool some of the people all of the time, and all of the people some of the time, but you cannot fool all of the people all of the time
Abraham Lincoln

The greatest form of mind control is repetition. The more you repeat the same mantra of alleged 'facts' the more will accept them to be true. It becomes an 'everyone knows that, mate'. If you can also censor any other version or alternative to your alleged 'facts' you are pretty much home and cooking.

By the start of 2020 the Cult owned the global mainstream media almost in its entirety to spew out its 'Covid' propaganda and ignore or discredit any other information and view. Cult-owned social media platforms in Cult-owned Silicon Valley were poised and ready to unleash a campaign of ferocious censorship to obliterate all but the official narrative. To complete the circle many demands for censorship by Silicon Valley were led by the mainstream media as 'journalists' became full-out enforcers for the Cult both as propagandists and censors. Part of this has been the influx of young people straight out of university who have become 'journalists' in significant positions. They have no experience and a headful of programmed perceptions from their years at school and university at a time when today's young are the most perceptually-targeted generations in known human history given the insidious impact of technology. They enter the media perceptually prepared and ready to repeat the narratives of the system that programmed them to

repeat its narratives. The BBC has a truly pathetic 'specialist disinformation reporter' called Marianna Spring who fits this bill perfectly. She is clueless about the world, how it works and what is really going on. Her role is to discredit anyone doing the job that a proper journalist would do and system-serving hacks like Spring wouldn't dare to do or even see the need to do. They are too busy licking the arse of authority which can never be wrong and, in the case of the BBC propaganda programme, *Panorama*, contacting payments systems such as PayPal to have a donations page taken down for a film company making documentaries questioning vaccines. Even the BBC soap opera *EastEnders* included a disgracefully biased scene in which an inarticulate white working class woman was made to look foolish for questioning the 'vaccine' while a well-spoken black man and Asian woman promoted the government narrative. It ticked every BBC box and the fact that the black and minority community was resisting the 'vaccine' had nothing to do with the way the scene was written. The BBC has become a disgusting tyrannical propaganda and censorship operation that should be defunded and disbanded and a free media take its place with a brief to stop censorship instead of demanding it. A BBC 'interview' with Gates goes something like: 'Mr Gates, sir, if I can call you sir, would you like to tell our audience why you are such a great man, a wonderful humanitarian philanthropist, and why you should absolutely be allowed as a software salesman to decide health policy for approaching eight billion people? Thank you, sir, please sir.' Propaganda programming has been incessant and merciless and when all you hear is the same story from the media, repeated by those around you who have only heard the same story, is it any wonder that people on a grand scale believe absolute mendacious garbage to be true? You are about to see, too, why this level of information control is necessary when the official 'Covid' narrative is so nonsensical and unsupportable by the evidence.

Structure of Deceit

The pyramid structure through which the 'Covid' hoax has been manifested is very simple and has to be to work. As few people as possible have to be involved with full knowledge of what they are doing – and why – or the real story would get out. At the top of the pyramid are the inner core of the Cult which controls Bill Gates who, in turn, controls the World Health Organization through his pivotal funding and his puppet Director-General mouthpiece, Tedros. Before he was appointed Tedros was chair of the Gates-founded Global Fund to 'fight against AIDS, tuberculosis and malaria', a board member of the Gates-funded 'vaccine alliance' GAVI, and on the board of another Gates-funded organisation. Gates owns him and picked him for a specific reason – Tedros is a crook and worse. 'Dr' Tedros (he's not a medical doctor, the first WHO chief not to be) was a member of the tyrannical Marxist government of Ethiopia for decades with all its human rights abuses. He has faced allegations of corruption and misappropriation of funds and was exposed three times for covering up cholera epidemics while Ethiopia's health minister. Tedros appointed the mass-murdering genocidal Zimbabwe dictator Robert Mugabe as a WHO goodwill ambassador for public health which, as with Tedros, is like appointing a psychopath to run a peace and love campaign. The move was so ridiculous that he had to drop Mugabe in the face of widespread condemnation. American economist David Steinman, a Nobel peace prize nominee, lodged a complaint with the International Criminal Court in The Hague over alleged genocide by Tedros when he was Ethiopia's foreign minister. Steinman says Tedros was a 'crucial decision maker' who directed the actions of Ethiopia's security forces from 2013 to 2015 and one of three officials in charge when those security services embarked on the 'killing' and 'torturing' of Ethiopians. You can see where Tedros is coming from and it's sobering to think that he has been the vehicle for Gates and the Cult to direct the global response to 'Covid'. Think about that. A psychopathic Cult dictates to psychopath Gates who dictates to psychopath Tedros who dictates how countries of the world must respond to a 'Covid virus' never scientifically shown to exist. At the same time psychopathic Cult-owned Silicon Valley information

giants like Google, YouTube, Facebook and Twitter announced very early on that they would give the Cult/Gates/Tedros/WHO version of the narrative free advertising and censor those who challenged their intelligence-insulting, mendacious story.

The next layer in the global 'medical' structure below the Cult, Gates and Tedros are the chief medical officers and science 'advisers' in each of the WHO member countries which means virtually all of them. Medical officers and arbiters of science (they're not) then take the WHO policy and recommended responses and impose them on their country's population while the political 'leaders' say they are deciding policy (they're clearly not) by 'following the science' on the advice of the 'experts' – the same medical officers and science 'advisers' (dictators). In this way with the rarest of exceptions the entire world followed the same policy of lockdown, people distancing, masks and 'vaccines' dictated by the psychopathic Cult, psychopathic Gates and psychopathic Tedros who we are supposed to believe give a damn about the health of the world population they are seeking to enslave. That, amazingly, is all there is to it in terms of crucial decision-making. Medical staff in each country then follow like sheep the dictates of the shepherds at the top of the national medical hierarchies – chief medical officers and science 'advisers' who themselves follow like sheep the shepherds of the World Health Organization and the Cult. Shepherds at the national level often have major funding and other connections to Gates and his Bill and Melinda Gates Foundation which carefully hands out money like confetti at a wedding to control the entire global medical system from the WHO down.

Follow the money

Christopher Whitty, Chief Medical Adviser to the UK Government at the centre of 'virus' policy, a senior adviser to the government's Scientific Advisory Group for Emergencies (SAGE), and Executive Board member of the World Health Organization, was gifted a grant of \$40 million by the Bill and Melinda Gates Foundation for malaria research in Africa. The BBC described the unelected Whitty as 'the

official who will probably have the greatest impact on our everyday lives of any individual policymaker in modern times' and so it turned out. What Gates and Tedros have said Whitty has done like his equivalents around the world. Patrick Vallance, co-chair of SAGE and the government's Chief Scientific Adviser, is a former executive of Big Pharma giant GlaxoSmithKline with its fundamental financial and business connections to Bill Gates. In September, 2020, it was revealed that Vallance owned a deferred bonus of shares in GlaxoSmithKline worth £600,000 while the company was 'developing' a 'Covid vaccine'. Move along now – nothing to see here – what could possibly be wrong with that? Imperial College in London, a major player in 'Covid' policy in Britain and elsewhere with its 'Covid-19' Response Team, is funded by Gates and has big connections to China while the now infamous Professor Neil Ferguson, the useless 'computer modeller' at Imperial College is also funded by Gates. Ferguson delivered the dramatically inaccurate excuse for the first lockdowns (much more in the next chapter). The Institute for Health Metrics and Evaluation (IHME) in the United States, another source of outrageously false 'Covid' computer models to justify lockdowns, is bankrolled by Gates who is a vehement promotor of lockdowns. America's version of Whitty and Vallance, the again now infamous Anthony Fauci, has connections to 'Covid vaccine' maker Moderna as does Bill Gates through funding from the Bill and Melinda Gates Foundation. Fauci is director of the National Institute of Allergy and Infectious Diseases (NIAID), a major recipient of Gates money, and they are very close. Deborah Birx who was appointed White House Coronavirus Response Coordinator in February, 2020, is yet another with ties to Gates. Everywhere you look at the different elements around the world behind the coordination and decision making of the 'Covid' hoax there is Bill Gates and his money. They include the World Health Organization; Centers for Disease Control (CDC) in the United States; National Institutes of Health (NIH) of Anthony Fauci; Imperial College and Neil Ferguson; the London School of Hygiene where Chris Whitty worked; Regulatory agencies like the UK Medicines & Healthcare products Regulatory Agency (MHRA)

which gave emergency approval for 'Covid vaccines'; Wellcome Trust; GAVI, the Vaccine Alliance; the Coalition for Epidemic Preparedness Innovations (CEPI); Johns Hopkins University which has compiled the false 'Covid' figures; and the World Economic Forum. A [Nationalfile.com](https://www.nationalfile.com) article said:

Gates has a lot of pull in the medical world, he has a multi-million dollar relationship with Dr. Fauci, and Fauci originally took the Gates line supporting vaccines and casting doubt on [the drug hydroxychloroquine]. Coronavirus response team member Dr. Deborah Birx, appointed by former president Obama to serve as United States Global AIDS Coordinator, also sits on the board of a group that has received billions from Gates' foundation, and Birx reportedly used a disputed Bill Gates-funded model for the White House's Coronavirus effort. Gates is a big proponent for a population lockdown scenario for the Coronavirus outbreak.

Another funder of Moderna is the Defense Advanced Research Projects Agency (DARPA), the technology-development arm of the Pentagon and one of the most sinister organisations on earth. DARPA had a major role with the CIA covert technology-funding operation In-Q-Tel in the development of Google and social media which is now at the centre of global censorship. Fauci and Gates are extremely close and openly admit to talking regularly about 'Covid' policy, but then why wouldn't Gates have a seat at every national 'Covid' table after his Foundation committed \$1.75 billion to the 'fight against Covid-19'. When passed through our Orwellian Translation Unit this means that he has bought and paid for the Cult-driven 'Covid' response worldwide. Research the major 'Covid' response personnel in your own country and you will find the same Gates funding and other connections again and again. Medical and science chiefs following World Health Organization 'policy' sit atop a medical hierarchy in their country of administrators, doctors and nursing staff. These 'subordinates' are told they must work and behave in accordance with the policy delivered from the 'top' of the national 'health' pyramid which is largely the policy delivered by the WHO which is the policy delivered by Gates and the Cult. The whole 'Covid' narrative has been imposed on medical staff by a climate of fear although great numbers don't even need that to comply. They do so through breathtaking levels of ignorance and

include doctors who go through life simply repeating what Big Pharma and their hierarchical masters tell them to say and believe. No wonder Big Pharma 'medicine' is one of the biggest killers on Planet Earth.

The same top-down system of intimidation operates with regard to the Cult Big Pharma cartel which also dictates policy through national and global medical systems in this way. The Cult and Big Pharma agendas are the same because the former controls and owns the latter. 'Health' administrators, doctors, and nursing staff are told to support and parrot the dictated policy or they will face consequences which can include being fired. How sad it's been to see medical staff meekly repeating and imposing Cult policy without question and most of those who can see through the deceit are only willing to speak anonymously off the record. They know what will happen if their identity is known. This has left the courageous few to expose the lies about the 'virus', face masks, overwhelmed hospitals that aren't, and the dangers of the 'vaccine' that isn't a vaccine. When these medical professionals and scientists, some renowned in their field, have taken to the Internet to expose the truth their articles, comments and videos have been deleted by Cult-owned Facebook, Twitter and YouTube. What a real head-shaker to see YouTube videos with leading world scientists and highly qualified medical specialists with an added link underneath to the notorious Cult propaganda website *Wikipedia* to find the 'facts' about the same subject.

HIV – the 'Covid' trial-run

I'll give you an example of the consequences for health and truth that come from censorship and unquestioning belief in official narratives. The story was told by PCR inventor Kary Mullis in his book *Dancing Naked in the Mind Field*. He said that in 1984 he accepted as just another scientific fact that Luc Montagnier of France's Pasteur Institute and Robert Gallo of America's National Institutes of Health had independently discovered that a 'retrovirus' dubbed HIV (human immunodeficiency virus) caused AIDS. They

were, after all, Mullis writes, specialists in retroviruses. This is how the medical and science pyramids work. Something is announced or *assumed* and then becomes an everybody-knows-that purely through repetition of the assumption as if it is fact. Complete crap becomes accepted truth with no supporting evidence and only repetition of the crap. This is how a 'virus' that doesn't exist became the 'virus' that changed the world. The HIV-AIDS fairy story became a multi-billion pound industry and the media poured out propaganda terrifying the world about the deadly HIV 'virus' that caused the lethal AIDS. By then Mullis was working at a lab in Santa Monica, California, to detect retroviruses with his PCR test in blood donations received by the Red Cross. In doing so he asked a virologist where he could find a reference for HIV being the cause of AIDS. 'You don't need a reference,' the virologist said ... '*Everybody knows it.*' Mullis said he wanted to quote a reference in the report he was doing and he said he felt a little funny about not knowing the source of such an important discovery when everyone else seemed to. The virologist suggested he cite a report by the Centers for Disease Control and Prevention (CDC) on morbidity and mortality. Mullis read the report, but it only said that an organism had been identified and did not say how. The report did not identify the original scientific work. Physicians, however, *assumed* (key recurring theme) that if the CDC was convinced that HIV caused AIDS then proof must exist. Mullis continues:

I did computer searches. Neither Montagnier, Gallo, nor anyone else had published papers describing experiments which led to the conclusion that HIV probably caused AIDS. I read the papers in *Science* for which they had become well known as AIDS doctors, but all they had said there was that they had found evidence of a past infection by something which was probably HIV in some AIDS patients.

They found antibodies. Antibodies to viruses had always been considered evidence of past disease, not present disease. Antibodies signaled that the virus had been defeated. The patient had saved himself. There was no indication in these papers that this virus caused a disease. They didn't show that everybody with the antibodies had the disease. In fact they found some healthy people with antibodies.

Mullis asked why their work had been published if Montagnier and Gallo hadn't really found this evidence, and why had they been fighting so hard to get credit for the discovery? He says he was hesitant to write 'HIV is the probable cause of AIDS' until he found published evidence to support that. 'Tens of thousands of scientists and researchers were spending billions of dollars a year doing research based on this idea,' Mullis writes. 'The reason had to be there somewhere; otherwise these people would not have allowed their research to settle into one narrow channel of investigation.' He said he lectured about PCR at numerous meetings where people were always talking about HIV and he asked them how they knew that HIV was the cause of AIDS:

Everyone said something. Everyone had the answer at home, in the office, in some drawer. They all knew, and they would send me the papers as soon as they got back. But I never got any papers. Nobody ever sent me the news about how AIDS was caused by HIV.

Eventually Mullis was able to ask Montagnier himself about the reference proof when he lectured in San Diego at the grand opening of the University of California AIDS Research Center. Mullis says this was the last time he would ask his question without showing anger. Montagnier said he should reference the CDC report. 'I read it', Mullis said, and it didn't answer the question. 'If Montagnier didn't know the answer who the hell did?' Then one night Mullis was driving when an interview came on National Public Radio with Peter Duesberg, a prominent virologist at Berkeley and a California Scientist of the Year. Mullis says he finally understood why he could not find references that connected HIV to AIDS – *there weren't any!* No one had ever proved that HIV causes AIDS even though it had spawned a multi-billion pound global industry and the media was repeating this as fact every day in their articles and broadcasts terrifying the shit out of people about AIDS and giving the impression that a positive test for HIV (see 'Covid') was a death sentence. Duesberg was a threat to the AIDS gravy train and the agenda that underpinned it. He was therefore abused and castigated after he told the Proceedings of the National Academy of Sciences

there was no good evidence implicating the new 'virus'. Editors rejected his manuscripts and his research funds were deleted. Mullis points out that the CDC has defined AIDS as one of more than 30 diseases *if accompanied* by a positive result on a test that detects antibodies to HIV; but those same diseases are not defined as AIDS cases when antibodies are not detected:

If an HIV-positive woman develops uterine cancer, for example, she is considered to have AIDS. If she is not HIV positive, she simply has uterine cancer. An HIV-positive man with tuberculosis has AIDS; if he tests negative he simply has tuberculosis. If he lives in Kenya or Colombia, where the test for HIV antibodies is too expensive, he is simply presumed to have the antibodies and therefore AIDS, and therefore he can be treated in the World Health Organization's clinic. It's the only medical help available in some places. And it's free, because the countries that support WHO are worried about AIDS.

Mullis accuses the CDC of continually adding new diseases (see ever more 'Covid symptoms') to the grand AIDS definition and of virtually doctoring the books to make it appear as if the disease continued to spread. He cites how in 1993 the CDC enormously broadened its AIDS definition and county health authorities were delighted because they received \$2,500 per year from the Federal government for every reported AIDS case. Ladies and gentlemen, I have just described, via Kary Mullis, the 'Covid pandemic' of 2020 and beyond. Every element is the same and it's been pulled off in the same way by the same networks.

The 'Covid virus' exists? Okay – prove it. Er ... still waiting

What Kary Mullis described with regard to 'HIV' has been repeated with 'Covid'. A claim is made that a new, or 'novel', infection has been found and the entire medical system of the world repeats that as fact exactly as they did with HIV and AIDS. No one in the mainstream asks rather relevant questions such as 'How do you know?' and 'Where is your proof?' The SARS-Cov-2 'virus' and the 'Covid-19 disease' became an overnight 'everybody-knows-that'. The origin could be debated and mulled over, but what you could not suggest was that 'SARS-Cov-2' didn't exist. That would be

ridiculous. 'Everybody knows' the 'virus' exists. Well, I didn't for one along with American proper doctors like Andrew Kaufman and Tom Cowan and long-time American proper journalist Jon Rappaport. We dared to pursue the obvious and simple question: 'Where's the evidence?' The overwhelming majority in medicine, journalism and the general public did not think to ask that. After all, *everyone knew* there was a new 'virus'. Everyone was saying so and I heard it on the BBC. Some would eventually argue that the 'deadly virus' was nothing like as deadly as claimed, but few would venture into the realms of its very existence. Had they done so they would have found that the evidence for that claim had gone AWOL as with HIV causes AIDS. In fact, not even that. For something to go AWOL it has to exist in the first place and scientific proof for a 'SARS-Cov-2' can be filed under nothing, nowhere and zilch.

Dr Andrew Kaufman is a board-certified forensic psychiatrist in New York State, a Doctor of Medicine and former Assistant Professor and Medical Director of Psychiatry at SUNY Upstate Medical University, and Medical Instructor of Hematology and Oncology at the Medical School of South Carolina. He also studied biology at the Massachusetts Institute of Technology (MIT) and trained in Psychiatry at Duke University. Kaufman is retired from allopathic medicine, but remains a consultant and educator on natural healing. I saw a video of his very early on in the 'Covid' hoax in which he questioned claims about the 'virus' in the absence of any supporting evidence and with plenty pointing the other way. I did everything I could to circulate his work which I felt was asking the pivotal questions that needed an answer. I can recommend an excellent pull-together interview he did with the website The Last Vagabond entitled *Dr Andrew Kaufman: Virus Isolation, Terrain Theory and Covid-19* and his website is andrewkaufmanmd.com. Kaufman is not only a forensic psychiatrist; he is forensic in all that he does. He always reads original scientific papers, experiments and studies instead of second-third-fourth-hand reports about the 'virus' in the media which are repeating the repeated repetition of the narrative. When he did so with the original Chinese 'virus' papers Kaufman

realised that there was no evidence of a 'SARS-Cov-2'. They had never – from the start – shown it to exist and every repeat of this claim worldwide was based on the accepted existence of proof that was nowhere to be found – see Kary Mullis and HIV. Here we go again.

Let's postulate

Kaufman discovered that the Chinese authorities immediately concluded that the cause of an illness that broke out among about 200 initial patients in Wuhan was a 'new virus' when there were no grounds to make that conclusion. The alleged 'virus' was not isolated from other genetic material in their samples and then shown through a system known as Koch's postulates to be the causative agent of the illness. The world was told that the SARS-Cov-2 'virus' caused a disease they called 'Covid-19' which had 'flu-like' symptoms and could lead to respiratory problems and pneumonia. If it wasn't so tragic it would almost be funny. *'Flu-like' symptoms? Pneumonia? Respiratory disease?* What in *CHINA* and particularly in *Wuhan*, one of the most polluted cities in the world with a resulting epidemic of respiratory disease?? Three hundred thousand people get pneumonia in China every year and there are nearly a billion cases worldwide of 'flu-like symptoms'. These have a whole range of causes – including pollution in Wuhan – but no other possibility was credibly considered in late 2019 when the world was told there was a new and deadly 'virus'. The global prevalence of pneumonia and 'flu-like systems' gave the Cult networks unlimited potential to re-diagnose these other causes as the mythical 'Covid-19' and that is what they did from the very start. Kaufman revealed how Chinese medical and science authorities (all subordinates to the Cult-owned communist government) took genetic material from the lungs of only a few of the first patients. The material contained their own cells, bacteria, fungi and other microorganisms living in their bodies. The only way you could prove the existence of the 'virus' and its responsibility for the alleged 'Covid-19' was to isolate the virus from all the other material – a process also known as 'purification' – and

then follow the postulates sequence developed in the late 19th century by German physician and bacteriologist Robert Koch which became the 'gold standard' for connecting an alleged causation agent to a disease:

1. The microorganism (bacteria, fungus, virus, etc.) must be present in every case of the disease and all patients must have the same symptoms. It must also *not be present in healthy individuals*.
2. The microorganism must be isolated from the host with the disease. If the microorganism is a bacteria or fungus it must be grown in a pure culture. If it is a virus, it must be purified (i.e. containing no other material except the virus particles) from a clinical sample.
3. The specific disease, with all of its characteristics, must be reproduced when the infectious agent (the purified virus or a pure culture of bacteria or fungi) is inoculated into a healthy, susceptible host.
4. The microorganism must be recoverable from the experimentally infected host as in step 2.

Not one of these criteria has been met in the case of 'SARS-Cov-2' and 'Covid-19'. Not ONE. EVER. Robert Koch refers to bacteria and not viruses. What are called 'viral particles' are so minute (hence masks are useless by any definition) that they could only be seen after the invention of the electron microscope in the 1930s and can still only be observed through that means. American bacteriologist and virologist Thomas Milton Rivers, the so-called 'Father of Modern Virology' who was very significantly director of the Rockefeller Institute for Medical Research in the 1930s, developed a less stringent version of Koch's postulates to identify 'virus' causation known as 'Rivers criteria'. 'Covid' did not pass that process either. Some even doubt whether any 'virus' can be isolated from other particles containing genetic material in the Koch method. Freedom of Information requests in many countries asking for scientific proof that the 'Covid virus' has been purified and isolated and shown to exist have all come back with a 'we don't have that' and when this happened with a request to the UK Department of Health they added this comment:

However, outside of the scope of the [Freedom of Information Act] and on a discretionary basis, the following information has been advised to us, which may be of interest. Most infectious diseases are caused by viruses, bacteria or fungi. Some bacteria or fungi have the capacity to grow on their own in isolation, for example in colonies on a petri dish. Viruses are different in that they are what we call 'obligate pathogens' – that is, they cannot survive or reproduce without infecting a host ...

... For some diseases, it is possible to establish causation between a microorganism and a disease by isolating the pathogen from a patient, growing it in pure culture and reintroducing it to a healthy organism. These are known as 'Koch's postulates' and were developed in 1882. However, as our understanding of disease and different disease-causing agents has advanced, these are no longer the method for determining causation [Andrew Kaufman asks why in that case are there two published articles falsely claiming to satisfy Koch's postulates].

It has long been known that viral diseases cannot be identified in this way as viruses cannot be grown in 'pure culture'. When a patient is tested for a viral illness, this is normally done by looking for the presence of antigens, or viral genetic code in a host with molecular biology techniques [Kaufman asks how you could know the origin of these chemicals without having a pure culture for comparison].

For the record 'antigens' are defined so:

Invading microorganisms have antigens on their surface that the human body can recognise as being foreign – meaning not belonging to it. When the body recognises a foreign antigen, lymphocytes (white blood cells) produce antibodies, which are complementary in shape to the antigen.

Notwithstanding that this is open to question in relation to 'SARS-Cov-2' the presence of 'antibodies' can have many causes and they are found in people that are perfectly well. Kary Mullis said: 'Antibodies ... had always been considered evidence of past disease, not present disease.'

'Covid' really is a *computer* 'virus'

Where the UK Department of Health statement says 'viruses' are now 'diagnosed' through a 'viral genetic code in a host with molecular biology techniques', they mean ... *the PCR test* which its inventor said cannot test for infectious disease. They have no credible method of connecting a 'virus' to a disease and we will see that there is no scientific proof that any 'virus' causes any disease or there is any such thing as a 'virus' in the way that it is described. Tenacious Canadian researcher Christine Massey and her team made

some 40 Freedom of Information requests to national public health agencies in different countries asking for proof that SARS-CoV-2 has been isolated and not one of them could supply that information. Massey said of her request in Canada: 'Freedom of Information reveals Public Health Agency of Canada has no record of 'SARS-COV-2' isolation performed by anyone, anywhere, ever.' If you accept the comment from the UK Department of Health it's because they can't isolate a 'virus'. Even so many 'science' papers claimed to have isolated the 'Covid virus' until they were questioned and had to admit they hadn't. A reply from the Robert Koch Institute in Germany was typical: 'I am not aware of a paper which purified isolated SARS-CoV-2.' So what the hell was Christian Drosten and his gang using to design the 'Covid' testing protocol that has produced all the illusory Covid' cases and 'Covid' deaths when the head of the Chinese version of the CDC admitted there was a problem right from the start in that the 'virus' had never been isolated/purified? Breathe deeply: What they are calling 'Covid' is actually created by a *computer program* i.e. *they made it up* – er, that's it. They took lung fluid, with many sources of genetic material, from one single person alleged to be infected with Covid-19 by a PCR test which they *claimed*, without clear evidence, contained a 'virus'. They used several computer programs to create a model of a theoretical virus genome sequence from more than fifty-six million small sequences of RNA, each of an unknown source, assembling them like a puzzle with no known solution. The computer filled in the gaps with sequences from bits in the gene bank to make it look like a bat SARS-like coronavirus! A wave of the magic wand and poof, an *in silico* (computer-generated) genome, a scientific fantasy, was created. UK health researcher Dr Kevin Corbett made the same point with this analogy:

... It's like giving you a few bones and saying that's your fish. It could be any fish. Not even a skeleton. Here's a few fragments of bones. That's your fish ... It's all from gene bank and the bits of the virus sequence that weren't there they made up.

They synthetically created them to fill in the blanks. That's what genetics is; it's a code. So it's ABBCCDDDD and you're missing some what you think is EEE so you put it in. It's all

synthetic. You just manufacture the bits that are missing. This is the end result of the geneticization of virology. This is basically a computer virus.

Further confirmation came in an email exchange between British citizen journalist Frances Leader and the government's Medicines & Healthcare Products Regulatory Agency (the Gates-funded MHRA) which gave emergency permission for untested 'Covid vaccines' to be used. The agency admitted that the 'vaccine' is not based on an isolated 'virus', but comes from a *computer-generated model*. Frances Leader was naturally banned from Cult-owned fascist Twitter for making this exchange public. The process of creating computer-generated alleged 'viruses' is called 'in silico' or 'in silicon' – computer chips – and the term 'in silico' is believed to originate with biological experiments using only a computer in 1989. 'Vaccines' involved with 'Covid' are also produced 'in silico' or by computer not a natural process. If the original 'virus' is nothing more than a made-up computer model how can there be 'new variants' of something that never existed in the first place? They are not new 'variants'; they are new *computer models* only minutely different to the original program and designed to further terrify the population into having the 'vaccine' and submitting to fascism. You want a 'new variant'? Click, click, enter – there you go. Tell the medical profession that you have discovered a 'South African variant', 'UK variants' or a 'Brazilian variant' and in the usual HIV-causes-AIDS manner they will unquestioningly repeat it with no evidence whatsoever to support these claims. They will go on television and warn about the dangers of 'new variants' while doing nothing more than repeating what they have been told to be true and knowing that any deviation from that would be career suicide. Big-time insiders will know it's a hoax, but much of the medical community is clueless about the way they are being played and themselves play the public without even being aware they are doing so. What an interesting 'coincidence' that AstraZeneca and Oxford University were conducting 'Covid vaccine trials' in the three countries – the UK, South Africa and Brazil – where the first three 'variants' were claimed to have 'broken out'.

Here's your 'virus' – it's a unicorn

Dr Andrew Kaufman presented a brilliant analysis describing how the 'virus' was imagined into fake existence when he dissected an article published by *Nature* and written by 19 authors detailing *alleged* 'sequencing of a complete viral genome' of the 'new SARS-CoV-2 virus'. This computer-modelled *in silico* genome was used as a template for all subsequent genome sequencing experiments that resulted in the so-called variants which he said now number more than 6,000. The fake genome was constructed from more than 56 million individual short strands of RNA. Those little pieces were assembled into longer pieces by finding areas of overlapping sequences. The computer programs created over two million possible combinations from which the authors simply chose the longest one. They then compared this to a 'bat virus' and the computer 'alignment' rearranged the sequence and filled in the gaps! They called this computer-generated abomination the 'complete genome'. Dr Tom Cowan, a fellow medical author and collaborator with Kaufman, said such computer-generation constitutes scientific fraud and he makes this superb analogy:

Here is an equivalency: A group of researchers claim to have found a unicorn because they found a piece of a hoof, a hair from a tail, and a snippet of a horn. They then add that information into a computer and program it to re-create the unicorn, and they then claim this computer re-creation is the real unicorn. Of course, they had never actually seen a unicorn so could not possibly have examined its genetic makeup to compare their samples with the actual unicorn's hair, hooves and horn.

The researchers claim they decided which is the real genome of SARS-CoV-2 by 'consensus', sort of like a vote. Again, different computer programs will come up with different versions of the imaginary 'unicorn', so they come together as a group and decide which is the real imaginary unicorn.

This is how the 'virus' that has transformed the world was brought into fraudulent 'existence'. Extraordinary, yes, but as the Nazis said the bigger the lie the more will believe it. Cowan, however, wasn't finished and he went on to identify what he called the real blockbuster in the paper. He quotes this section from a paper written

by virologists and published by the CDC and then explains what it means:

Therefore, we examined the capacity of SARS-CoV-2 to infect and replicate in several common primate and human cell lines, including human adenocarcinoma cells (A549), human liver cells (HUH 7.0), and human embryonic kidney cells (HEK-293T). In addition to Vero E6 and Vero CCL81 cells. ... Each cell line was inoculated at high multiplicity of infection and examined 24h post-infection.

No CPE was observed in any of the cell lines except in Vero cells, which grew to greater than 10 to the 7th power at 24 h post-infection. In contrast, HUH 7.0 and 293T showed only modest viral replication, and A549 cells were incompatible with SARS CoV-2 infection.

Cowan explains that when virologists attempt to prove infection they have three possible 'hosts' or models on which they can test. The first was humans. Exposure to humans was generally not done for ethical reasons and has never been done with SARS-CoV-2 or any coronavirus. The second possible host was animals. Cowan said that forgetting for a moment that they never actually use purified virus when exposing animals they do use solutions that they *claim* contain the virus. Exposure to animals has been done with SARS-CoV-2 in an experiment involving mice and this is what they found: *None of the wild (normal) mice got sick*. In a group of genetically-modified mice, a statistically insignificant number lost weight and had slightly bristled fur, but they experienced nothing like the illness called 'Covid-19'. Cowan said the third method – the one they mostly rely on – is to inoculate solutions they *say* contain the virus onto a variety of tissue cultures. This process had never been shown to kill tissue *unless* the sample material was starved of nutrients and poisoned as *part of the process*. Yes, incredibly, in tissue experiments designed to show the 'virus' is responsible for killing the tissue they starve the tissue of nutrients and add toxic drugs including antibiotics and they do not have control studies to see if it's the starvation and poisoning that is degrading the tissue rather than the 'virus' they allege to be in there somewhere. You want me to pinch you? Yep, I understand. Tom Cowan said this about the whole nonsensical farce as he explains what that quote from the CDC paper really means:

The shocking thing about the above quote is that using their own methods, the virologists found that solutions containing SARS-CoV-2 – even in high amounts – were NOT, I repeat NOT, infective to any of the three human tissue cultures they tested. In plain English, this means they proved, on their terms, that this ‘new coronavirus’ is not infectious to human beings. It is ONLY infective to monkey kidney cells, and only then when you add two potent drugs (gentamicin and amphotericin), known to be toxic to kidneys, to the mix.

My friends, read this again and again. These virologists, published by the CDC, performed a clear proof, on their terms, showing that the SARS-CoV-2 virus is harmless to human beings. That is the only possible conclusion, but, unfortunately, this result is not even mentioned in their conclusion. They simply say they can provide virus stocks cultured only on monkey Vero cells, thanks for coming.

Cowan concluded: ‘If people really understood how this “science” was done, I would hope they would storm the gates and demand honesty, transparency and truth.’ Dr Michael Yeadon, former Vice President and Chief Scientific Adviser at drug giant Pfizer has been a vocal critic of the ‘Covid vaccine’ and its potential for multiple harm. He said in an interview in April, 2021, that ‘not one [vaccine] has the virus. He was asked why vaccines normally using a ‘dead’ version of a disease to activate the immune system were not used for ‘Covid’ and instead we had the synthetic methods of the ‘mRNA Covid vaccine’. Yeadon said that to do the former ‘you’d have to have some of [the virus] wouldn’t you?’ He added: ‘No-one’s got any – seriously.’ Yeadon said that surely they couldn’t have fooled the whole world for a year without having a virus, ‘but oddly enough ask around – no one’s got it’. He didn’t know why with all the ‘great labs’ around the world that the virus had not been isolated – ‘Maybe they’ve been too busy running bad PCR tests and vaccines that people don’t need.’ What is today called ‘science’ is not ‘science’ at all. Science is no longer what is, but whatever people can be manipulated to *believe* that it is. Real science has been hijacked by the Cult to dispense and produce the ‘expert scientists’ and contentions that suit the agenda of the Cult. How big-time this has happened with the ‘Covid’ hoax which is entirely based on fake science delivered by fake ‘scientists’ and fake ‘doctors’. The human-caused climate change hoax is also entirely based on fake science delivered by fake ‘scientists’ and fake ‘climate experts’. In both cases real

scientists, climate experts and doctors have their views suppressed and deleted by the Cult-owned science establishment, media and Silicon Valley. This is the 'science' that politicians claim to be 'following' and a common denominator of 'Covid' and climate are Cult psychopaths Bill Gates and his mate Klaus Schwab at the Gates-funded World Economic Forum. But, don't worry, it's all just a coincidence and absolutely nothing to worry about. Zzzzzzzzz.

What is a 'virus' REALLY?

Dr Tom Cowan is one of many contesting the very existence of viruses let alone that they cause disease. This is understandable when there is no scientific evidence for a disease-causing 'virus'. German virologist Dr Stefan Lanka won a landmark case in 2017 in the German Supreme Court over his contention that there is no such thing as a measles virus. He had offered a big prize for anyone who could prove there is and Lanka won his case when someone sought to claim the money. There is currently a prize of more than 225,000 euros on offer from an Isolate Truth Fund for anyone who can prove the isolation of SARS-CoV-2 and its genetic substance. Lanka wrote in an article headed 'The Misconception Called Virus' that scientists think a 'virus' is causing tissue to become diseased and degraded when in fact it is the *processes they are using* which do that – not a 'virus'. Lanka has done an important job in making this point clear as Cowan did in his analysis of the CDC paper. Lanka says that all claims about viruses as disease-causing pathogens are wrong and based on 'easily recognisable, understandable and verifiable misinterpretations.' Scientists believed they were working with 'viruses' in their laboratories when they were really working with 'typical particles of specific dying tissues or cells ...' Lanka said that the tissue decaying process claimed to be caused by a 'virus' still happens when no alleged 'virus' is involved. It's the *process* that does the damage and not a 'virus'. The genetic sample is deprived of nutrients, removed from its energy supply through removal from the body and then doused in toxic antibiotics to remove any bacteria. He confirms again that establishment scientists do not (pinch me)

conduct control experiments to see if this is the case and if they did they would see the claims that 'viruses' are doing the damage is nonsense. He adds that during the measles 'virus' court case he commissioned an independent laboratory to perform just such a control experiment and the result was that the tissues and cells died in the exact same way as with alleged 'infected' material. This is supported by a gathering number of scientists, doctors and researchers who reject what is called 'germ theory' or the belief in the body being infected by contagious sources emitted by other people. Researchers Dawn Lester and David Parker take the same stance in their highly-detailed and sourced book *What Really Makes You Ill – Why everything you thought you knew about disease is wrong* which was recommended to me by a number of medical professionals genuinely seeking the truth. Lester and Parker say there is no provable scientific evidence to show that a 'virus' can be transmitted between people or people and animals or animals and people:

The definition also claims that viruses are the cause of many diseases, as if this has been definitively proven. But this is not the case; there is no original scientific evidence that definitively demonstrates that any virus is the cause of any disease. The burden of proof for any theory lies with those who proposed it; but none of the existing documents provides 'proof' that supports the claim that 'viruses' are pathogens.

Dr Tom Cowan employs one of his clever analogies to describe the process by which a 'virus' is named as the culprit for a disease when what is called a 'virus' is only material released by cells detoxing themselves from infiltration by chemical or radiation poisoning. The tidal wave of technologically-generated radiation in the 'smart' modern world plus all the toxic food and drink are causing this to happen more than ever. Deluded 'scientists' misread this as a gathering impact of what they wrongly label 'viruses'.

Paper can infect houses

Cowan said in an article for davidicke.com – with his tongue only mildly in his cheek – that he believed he had made a tremendous

discovery that may revolutionise science. He had discovered that small bits of paper are alive, 'well alive-ish', can 'infect' houses, and then reproduce themselves inside the house. The result was that this explosion of growth in the paper inside the house causes the house to explode, blowing it to smithereens. His evidence for this new theory is that in the past months he had carefully examined many of the houses in his neighbourhood and found almost no scraps of paper on the lawns and surrounds of the house. There was an occasional stray label, but nothing more. Then he would return to these same houses a week or so later and with a few, not all of them, particularly the old and decrepit ones, he found to his shock and surprise they were littered with stray bits of paper. He knew then that the paper had infected these houses, made copies of itself, and blew up the house. A young boy on a bicycle at one of the sites told him he had seen a demolition crew using dynamite to explode the house the previous week, but Cowan dismissed this as the idle thoughts of silly boys because 'I was on to something big'. He was on to how 'scientists' mistake genetic material in the detoxifying process for something they call a 'virus'. Cowan said of his house and paper story:

If this sounds crazy to you, it's because it should. This scenario is obviously nuts. But consider this admittedly embellished, for effect, current viral theory that all scientists, medical doctors and virologists currently believe.

He takes the example of the 'novel SARS-Cov2' virus to prove the point. First they take someone with an undefined illness called 'Covid-19' and don't even attempt to find any virus in their sputum. Never mind the scientists still describe how this 'virus', which they have not located attaches to a cell receptor, injects its genetic material, in 'Covid's' case, RNA, into the cell. The RNA once inserted exploits the cell to reproduce itself and makes 'thousands, nay millions, of copies of itself ... Then it emerges victorious to claim its next victim':

If you were to look in the scientific literature for proof, actual scientific proof, that uniform SARS-CoV2 viruses have been properly isolated from the sputum of a sick person, that actual spike proteins could be seen protruding from the virus (which has not been found), you would find that such evidence doesn't exist.

If you go looking in the published scientific literature for actual pictures, proof, that these spike proteins or any viral proteins are ever attached to any receptor embedded in any cell membrane, you would also find that no such evidence exists. If you were to look for a video or documented evidence of the intact virus injecting its genetic material into the body of the cell, reproducing itself and then emerging victorious by budding off the cell membrane, you would find that no such evidence exists.

The closest thing you would find is electron micrograph pictures of cellular particles, possibly attached to cell debris, both of which to be seen were stained by heavy metals, a process that completely distorts their architecture within the living organism. This is like finding bits of paper stuck to the blown-up bricks, thereby proving the paper emerged by taking pieces of the bricks on its way out.

The Enders baloney

Cowan describes the 'Covid' story as being just as make-believe as his paper story and he charts back this fantasy to a Nobel Prize winner called John Enders (1897-1985), an American biomedical scientist who has been dubbed 'The Father of Modern Vaccines'. Enders is claimed to have 'discovered' the process of the viral culture which 'proved' that a 'virus' caused measles. Cowan explains how Enders did this 'by using the EXACT same procedure that has been followed by every virologist to find and characterize every new virus since 1954'. Enders took throat swabs from children with measles and immersed them in 2ml of milk. Penicillin (100u/ml) and the antibiotic streptomycin (50,g/ml) were added and the whole mix was centrifuged – rotated at high speed to separate large cellular debris from small particles and molecules as with milk and cream, for example. Cowan says that if the aim is to find little particles of genetic material ('viruses') in the snot from children with measles it would seem that the last thing you would do is mix the snot with other material – milk –that also has genetic material. 'How are you ever going to know whether whatever you found came from the snot or the milk?' He points out that streptomycin is a 'nephrotoxic' or poisonous-to-the-kidney drug. You will see the relevance of that

shortly. Cowan says that it gets worse, much worse, when Enders describes the culture medium upon which the virus 'grows': 'The culture medium consisted of bovine amniotic fluid (90%), beef embryo extract (5%), horse serum (5%), antibiotics and phenol red as an indicator of cell metabolism.' Cowan asks incredulously: 'Did he just say that the culture medium also contained fluids and tissues that are themselves rich sources of genetic material?' The genetic cocktail, or 'medium', is inoculated onto tissue and cells from rhesus monkey *kidney* tissue. This is where the importance of streptomycin comes in and currently-used antimicrobials and other drugs that are *poisonous to kidneys* and used in ALL modern viral cultures (e.g. gentamicin, streptomycin, and amphotericin). Cowan asks: 'How are you ever going to know from this witch's brew where any genetic material comes from as we now have five different sources of rich genetic material in our mix?' Remember, he says, that all genetic material, whether from monkey kidney tissues, bovine serum, milk, etc., is made from the exact same components. The same central question returns: 'How are you possibly going to know that it was the virus that killed the kidney tissue and not the toxic antibiotic and starvation rations on which you are growing the tissue?' John Enders answered the question himself – *you can't*:

A second agent was obtained from an uninoculated culture of monkey kidney cells. The cytopathic changes [death of the cells] it induced in the unstained preparations could not be distinguished with confidence from the viruses isolated from measles.

The death of the cells ('cytopathic changes') happened in exactly the same manner, whether they inoculated the kidney tissue with the measles snot or not, Cowan says. 'This is evidence that the destruction of the tissue, the very proof of viral causation of illness, was not caused by anything in the snot because they saw the same destructive effect when the snot was not even used ... the cytopathic, i.e., cell-killing, changes come from the process of the culture itself, not from any virus in any snot, period.' Enders quotes in his 1957 paper a virologist called Ruckle as reporting similar findings 'and in addition has isolated an agent from monkey kidney tissue that is so

far indistinguishable from human measles virus'. In other words, Cowan says, these particles called 'measles viruses' are simply and clearly breakdown products of the starved and poisoned tissue. For measles 'virus' see all 'viruses' including the so-called 'Covid virus'. Enders, the 'Father of Modern Vaccines', also said:

There is a potential risk in employing cultures of primate cells for the production of vaccines composed of attenuated virus, since the presence of other agents possibly latent in primate tissues cannot be definitely excluded by any known method.

Cowan further quotes from a paper published in the journal *Viruses* in May, 2020, while the 'Covid pandemic' was well underway in the media if not in reality. 'EVs' here refers to particles of genetic debris from our own tissues, such as exosomes of which more in a moment: 'The remarkable resemblance between EVs and viruses has caused quite a few problems in the studies focused on the analysis of EVs released during viral infections.' Later the paper adds that to date a reliable method that can actually guarantee a complete separation (of EVs from viruses) DOES NOT EXIST. This was published at a time when a fairy tale 'virus' was claimed in total certainty to be causing a fairy tale 'viral disease' called 'Covid-19' – a fairy tale that was already well on the way to transforming human society in the image that the Cult has worked to achieve for so long. Cowan concludes his article:

To summarize, there is no scientific evidence that pathogenic viruses exist. What we think of as 'viruses' are simply the normal breakdown products of dead and dying tissues and cells. When we are well, we make fewer of these particles; when we are starved, poisoned, suffocated by wearing masks, or afraid, we make more.

There is no engineered virus circulating and making people sick. People in laboratories all over the world are making genetically modified products to make people sick. These are called vaccines. There is no virome, no 'ecosystem' of viruses, viruses are not 8%, 50% or 100 % of our genetic material. These are all simply erroneous ideas based on the misconception called a virus.

What is 'Covid'? Load of bollocks

The background described here by Cowan and Lanka was emphasised in the first video presentation that I saw by Dr Andrew Kaufman when he asked whether the 'Covid virus' was in truth a natural defence mechanism of the body called 'exosomes'. These are released by cells when in states of toxicity – see the same themes returning over and over. They are released ever more profusely as chemical and radiation toxicity increases and think of the potential effect therefore of 5G alone as its destructive frequencies infest the human energetic information field with a gathering pace (5G went online in Wuhan in 2019 as the 'virus' emerged). I'll have more about this later. Exosomes transmit a warning to the rest of the body that 'Houston, we have a problem'. Kaufman presented images of exosomes and compared them with 'Covid' under an electron microscope and the similarity was remarkable. They both attach to the same cell receptors (*claimed* in the case of 'Covid'), contain the same genetic material in the form of RNA or ribonucleic acid, and both are found in 'viral cell cultures' with damaged or dying cells. James Hildreth MD, President and Chief Executive Officer of the Meharry Medical College at Johns Hopkins, said: 'The virus is fully an exosome in every sense of the word.' Kaufman's conclusion was that there is no 'virus': 'This entire pandemic is a completely manufactured crisis ... there is no evidence of anyone dying from [this] illness.' Dr Tom Cowan and Sally Fallon Morell, authors of *The Contagion Myth*, published a statement with Dr Kaufman in February, 2021, explaining why the 'virus' does not exist and you can read it that in full in the Appendix.

'Virus' theory can be traced to the 'cell theory' in 1858 of German physician Rudolf Virchow (1821-1920) who contended that disease originates from a single cell infiltrated by a 'virus'. Dr Stefan Lanka said that findings and insights with respect to the structure, function and central importance of tissues in the creation of life, which were already known in 1858, comprehensively refute the cell theory. Virchow ignored them. We have seen the part later played by John Enders in the 1950s and Lanka notes that infection theories were only established as a global dogma through the policies and

eugenics of the Third Reich in Nazi Germany (creation of the same Sabbatian cult behind the 'Covid' hoax). Lanka said: 'Before 1933, scientists dared to contradict this theory; after 1933, these critical scientists were silenced'. Dr Tom Cowan's view is that ill-health is caused by too much of something, too little of something, or toxification from chemicals and radiation – not contagion. We must also highlight as a major source of the 'virus' theology a man still called the 'Father of Modern Virology' – Thomas Milton Rivers (1888-1962). There is no way given the Cult's long game policy that it was a coincidence for the 'Father of Modern Virology' to be director of the Rockefeller Institute for Medical Research from 1937 to 1956 when he is credited with making the Rockefeller Institute a leader in 'viral research'. Cult Rockefellerers were the force behind the creation of Big Pharma 'medicine', established the World Health Organisation in 1948, and have long and close associations with the Gates family that now runs the WHO during the pandemic hoax through mega-rich Cult gofer and psychopath Bill Gates.

Only a Renegade Mind can see through all this bullshit by asking the questions that need to be answered, not taking 'no' or prevarication for an answer, and certainly not hiding from the truth in fear of speaking it. Renegade Minds have always changed the world for the better and they will change this one no matter how bleak it may currently appear to be.

CHAPTER SIX

Sequence of deceit

If you tell the truth, you don't have to remember anything
Mark Twain

Against the background that I have laid out this far the sequence that took us from an invented 'virus' in Cult-owned China in late 2019 to the fascist transformation of human society can be seen and understood in a whole new context.

We were told that a deadly disease had broken out in Wuhan and the world media began its campaign (coordinated by behavioural psychologists as we shall see) to terrify the population into unquestioning compliance. We were shown images of Chinese people collapsing in the street which never happened in the West with what was supposed to be the same condition. In the earliest days when alleged cases and deaths were few the fear register was hysterical in many areas of the media and this would expand into the common media narrative across the world. The real story was rather different, but we were never told that. The Chinese government, one of the Cult's biggest centres of global operation, said they had discovered a new illness with flu-like and pneumonia-type symptoms in a city with such toxic air that it is overwhelmed with flu-like symptoms, pneumonia and respiratory disease. Chinese scientists said it was a new – 'novel' – coronavirus which they called Sars-Cov-2 and that it caused a disease they labelled 'Covid-19'. There was no evidence for this and the 'virus' has never to this day been isolated, purified and its genetic code established from that. It

was from the beginning a computer-generated fiction. Stories of Chinese whistleblowers saying the number of deaths was being suppressed or that the 'new disease' was related to the Wuhan bio-lab misdirected mainstream and alternative media into cul-de-sacs to obscure the real truth – there was no 'virus'.

Chinese scientists took genetic material from the lung fluid of just a few people and said they had found a 'new' disease when this material had a wide range of content. There was no evidence for a 'virus' for the very reasons explained in the last two chapters. The 'virus' has never been shown to (a) exist and (b) cause any disease. People were diagnosed on symptoms that are so widespread in Wuhan and polluted China and with a PCR test that can't detect infectious disease. On this farce the whole global scam was sold to the rest of the world which would also diagnose respiratory disease as 'Covid-19' from symptoms alone or with a PCR test not testing for a 'virus'. Flu miraculously disappeared *worldwide* in 2020 and into 2021 as it was redesignated 'Covid-19'. It was really the same old flu with its 'flu-like' symptoms attributed to 'flu-like' 'Covid-19'. At the same time with very few exceptions the Chinese response of draconian lockdown and fascism was the chosen weapon to respond across the West as recommended by the Cult-owned Tedros at the Cult-owned World Health Organization run by the Cult-owned Gates. All was going according to plan. Chinese scientists – everything in China is controlled by the Cult-owned government – compared their contaminated RNA lung-fluid material with other RNA sequences and said it appeared to be just under 80 percent identical to the SARS-CoV-1 'virus' claimed to be the cause of the SARS (severe acute respiratory syndrome) 'outbreak' in 2003. They decreed that because of this the 'new virus' had to be related and they called it SARS-CoV-2. There are some serious problems with this assumption and *assumption* was all it was. Most 'factual' science turns out to be assumptions repeated into everyone-knows-that. A match of under 80-percent is meaningless. Dr Kaufman makes the point that there's a 96 percent genetic correlation between humans and chimpanzees, but 'no one would say our genetic material is part

of the chimpanzee family'. Yet the Chinese authorities were claiming that a much lower percentage, less than 80 percent, proved the existence of a new 'coronavirus'. For goodness sake human DNA is 60 percent similar to a *banana*.

You are feeling sleepy

The entire 'Covid' hoax is a global Psyop, a psychological operation to program the human mind into believing and fearing a complete fantasy. A crucial aspect of this was what *appeared* to happen in Italy. It was all very well streaming out daily images of an alleged catastrophe in Wuhan, but to the Western mind it was still on the other side of the world in a very different culture and setting. A reaction of 'this could happen to me and my family' was still nothing like as intense enough for the mind-doctors. The Cult needed a Western example to push people over that edge and it chose Italy, one of its major global locations going back to the Roman Empire. An Italian 'Covid' crisis was manufactured in a particular area called Lombardy which just happens to be notorious for its toxic air and therefore respiratory disease. Wuhan, China, *déjà vu*. An hysterical media told horror stories of Italians dying from 'Covid' in their droves and how Lombardy hospitals were being overrun by a tidal wave of desperately ill people needing treatment after being struck down by the 'deadly virus'. Here was the psychological turning point the Cult had planned. Wow, if this is happening in Italy, the Western mind concluded, this indeed could happen to me and my family. Another point is that Italian authorities responded by following the Chinese blueprint so vehemently recommended by the Cult-owned World Health Organization. They imposed fascistic lockdowns on the whole country viciously policed with the help of surveillance drones sweeping through the streets seeking out anyone who escaped from mass house arrest. Livelihoods were destroyed and psychology unravelled in the way we have witnessed since in all lockdown countries. Crucial to the plan was that Italy responded in this way to set the precedent of suspending freedom and imposing fascism in a 'Western liberal democracy'. I emphasised in an

animated video explanation on davidicke.com posted in the summer of 2020 how important it was to the Cult to expand the Chinese lockdown model across the West. Without this, and the bare-faced lie that non-symptomatic people could still transmit a 'disease' they didn't have, there was no way locking down the whole population, sick and not sick, could be pulled off. At just the right time and with no evidence Cult operatives and gofers claimed that people without symptoms could pass on the 'disease'. In the name of protecting the 'vulnerable' like elderly people, who lockdowns would kill by the tens of thousands, we had for the first time healthy people told to isolate as well as the sick. The great majority of people who tested positive had no symptoms because there was nothing wrong with them. It was just a trick made possible by a test not testing for the 'virus'.

Months after my animated video the Gates-funded Professor Neil Ferguson at the Gates-funded Imperial College confirmed that I was right. He didn't say it in those terms, naturally, but he did say it. Ferguson will enter the story shortly for his outrageously crazy 'computer models' that led to Britain, the United States and many other countries following the Chinese and now Italian methods of response. Put another way, following the Cult script. Ferguson said that SAGE, the UK government's scientific advisory group which has controlled 'Covid' policy from the start, wanted to follow the Chinese lockdown model (while they all continued to work and be paid), but they wondered if they could possibly, in Ferguson's words, 'get away with it in Europe'. 'Get away with it'? Who the hell do these moronic, arrogant people think they are? This appalling man Ferguson said that once Italy went into national lockdown they realised they, too, could mimic China:

It's a communist one-party state, we said. We couldn't get away with it in Europe, we thought ... and then Italy did it. And we realised we could. Behind this garbage from Ferguson is a simple fact: Doing the same as China in every country was the plan from the start and Ferguson's 'models' would play a central role in achieving that. It's just a coincidence, of course, and absolutely nothing to worry your little head about.

Oops, sorry, our mistake

Once the Italian segment of the Psyop had done the job it was designed to do a very different story emerged. Italian authorities revealed that 99 percent of those who had 'died from Covid-19' in Italy had one, two, three, or more 'co-morbidities' or illnesses and health problems that could have ended their life. The US Centers for Disease Control and Prevention (CDC) published a figure of 94 percent for Americans dying of 'Covid' while having other serious medical conditions – on average two to three (some five or six) other potential causes of death. In terms of death from an unproven 'virus' I say it is 100 percent. The other one percent in Italy and six percent in the US would presumably have died from 'Covid's' flu-like symptoms with a range of other possible causes in conjunction with a test not testing for the 'virus'. Fox News reported that even more startling figures had emerged in one US county in which 410 of 422 deaths attributed to 'Covid-19' had other potentially deadly health conditions. The Italian National Health Institute said later that the average age of people dying with a 'Covid-19' diagnosis in Italy was about 81. Ninety percent were over 70 with ten percent over 90. In terms of other reasons to die some 80 percent had two or more chronic diseases with half having three or more including cardiovascular problems, diabetes, respiratory problems and cancer. Why is the phantom 'Covid-19' said to kill overwhelmingly old people and hardly affect the young? Old people continually die of many causes and especially respiratory disease which you can re-diagnose 'Covid-19' while young people die in tiny numbers by comparison and rarely of respiratory disease. Old people 'die of Covid' because they die of other things that can be redesignated 'Covid' and it really is that simple.

Flu has flown

The blueprint was in place. Get your illusory 'cases' from a test not testing for the 'virus' and redesignate other causes of death as 'Covid-19'. You have an instant 'pandemic' from something that is nothing more than a computer-generated fiction. With near-on a

billion people having 'flu-like' symptoms every year the potential was limitless and we can see why flu quickly and apparently miraculously disappeared *worldwide* by being diagnosed 'Covid-19'. The painfully bloody obvious was explained away by the childlike media in headlines like this in the UK '*Independent*': 'Not a single case of flu detected by Public Health England this year as Covid restrictions suppress virus'. I kid you not. The masking, social distancing and house arrest that did not make the 'Covid virus' disappear somehow did so with the 'flu virus'. Even worse the article, by a bloke called Samuel Lovett, suggested that maybe the masking, sanitising and other 'Covid' measures should continue to keep the flu away. With a ridiculousness that disturbs your breathing (it's 'Covid-19') the said Lovett wrote: 'With widespread social distancing and mask-wearing measures in place throughout the UK, the usual routes of transmission for influenza have been blocked.' He had absolutely no evidence to support that statement, but look at the consequences of him acknowledging the obvious. With flu not disappearing at all and only being relabelled 'Covid-19' he would have to contemplate that 'Covid' was a hoax on a scale that is hard to imagine. You need guts and commitment to truth to even go there and that's clearly something Samuel Lovett does not have in abundance. He would never have got it through the editors anyway.

Tens of thousands die in the United States alone every winter from flu including many with pneumonia complications. CDC figures record *45 million* Americans diagnosed with flu in 2017-2018 of which 61,000 died and some reports claim 80,000. Where was the same hysteria then that we have seen with 'Covid-19'? Some 250,000 Americans are admitted to hospital with pneumonia every year with about 50,000 cases proving fatal. About 65 million suffer respiratory disease every year and three million deaths makes this the third biggest cause of death worldwide. You only have to redesignate a portion of all these people 'Covid-19' and you have an instant global pandemic or the *appearance* of one. Why would doctors do this? They are told to do this and all but a few dare not refuse those who must be obeyed. Doctors in general are not researching their own

knowledge and instead take it direct and unquestioned from the authorities that own them and their careers. The authorities say they must now diagnose these symptoms 'Covid-19' and not flu, or whatever, and they do it. Dark suits say put 'Covid-19' on death certificates no matter what the cause of death and the doctors do it. Renegade Minds don't fall for the illusion that doctors and medical staff are all highly-intelligent, highly-principled, seekers of medical truth. *Some are*, but not the majority. They are repeaters, gofers, and yes sir, no sir, purveyors of what the system demands they purvey. The 'Covid' con is not merely confined to diseases of the lungs. Instructions to doctors to put 'Covid-19' on death certificates for anyone dying of *anything* within 28 days (or much more) of a positive test not testing for the 'virus' opened the floodgates. The term dying *with* 'Covid' and not *of* 'Covid' was coined to cover the truth. Whether it was a *with* or an *of* they were all added to the death numbers attributed to the 'deadly virus' compiled by national governments and globally by the Gates-funded Johns Hopkins operation in the United States that was so involved in those 'pandemic' simulations. Fraudulent deaths were added to the ever-growing list of fraudulent 'cases' from false positives from a false test. No wonder Professor Walter Ricciardi, scientific advisor to the Italian minister of health, said after the Lombardy hysteria had done its job that 'Covid' death rates were due to Italy having the second oldest population in the world and to *how hospitals record deaths*:

The way in which we code deaths in our country is very generous in the sense that all the people who die in hospitals with the coronavirus are deemed to be dying of the coronavirus. On re-evaluation by the National Institute of Health, only 12 per cent of death certificates have shown a direct causality from coronavirus, while 88 per cent of patients who have died have at least one pre-morbidity – many had two or three.

This is extraordinary enough when you consider the propaganda campaign to use Italy to terrify the world, but how can they even say twelve percent were genuine when the 'virus' has not been shown to exist, its 'code' is a computer program, and diagnosis comes from a test not testing for it? As in China, and soon the world, 'Covid-19' in

Italy was a redesignation of diagnosis. Lies and corruption were to become the real 'pandemic' fuelled by a pathetically-compliant medical system taking its orders from the tiny few at the top of their national hierarchy who answered to the World Health Organization which answers to Gates and the Cult. Doctors were told – ordered – to diagnose a particular set of symptoms 'Covid-19' and put that on the death certificate for any cause of death if the patient had tested positive with a test not testing for the virus or had 'Covid' symptoms like the flu. The United States even introduced big financial incentives to manipulate the figures with hospitals receiving £4,600 from the Medicare system for diagnosing someone with regular pneumonia, \$13,000 if they made the diagnosis from the same symptoms 'Covid-19' pneumonia, and \$39,000 if they put a 'Covid' diagnosed patient on a ventilator that would almost certainly kill them. A few – painfully and pathetically few – medical whistleblowers revealed (before Cult-owned YouTube deleted their videos) that they had been instructed to 'let the patient crash' and put them straight on a ventilator instead of going through a series of far less intrusive and dangerous methods as they would have done before the pandemic hoax began and the financial incentives kicked in. We are talking cold-blooded murder given that ventilators are so damaging to respiratory systems they are usually the last step before heaven awaits. Renegade Minds never fall for the belief that people in white coats are all angels of mercy and cannot be full-on psychopaths. I have explained in detail in *The Answer* how what I am describing here played out across the world coordinated by the World Health Organization through the medical hierarchies in almost every country.

Medical scientist calls it

Information about the non-existence of the 'virus' began to emerge for me in late March, 2020, and mushroomed after that. I was sent an email by Sir Julian Rose, a writer, researcher, and organic farming promotor, from a medical scientist friend of his in the United States. Even at that early stage in March the scientist was able to explain

how the 'Covid' hoax was being manipulated. He said there were no reliable tests for a specific 'Covid-19 virus' and nor were there any reliable agencies or media outlets for reporting numbers of actual 'Covid-19' cases. We have seen in the long period since then that he was absolutely right. 'Every action and reaction to Covid-19 is based on totally flawed data and we simply cannot make accurate assessments,' he said. Most people diagnosed with 'Covid-19' were showing nothing more than cold and flu-like symptoms 'because most coronavirus strains *are* nothing more than cold/flu-like symptoms'. We had farcical situations like an 84-year-old German man testing positive for 'Covid-19' and his nursing home ordered to quarantine only for him to be found to have a common cold. The scientist described back then why PCR tests and what he called the 'Mickey Mouse test kits' were useless for what they were claimed to be identifying. 'The idea these kits can isolate a specific virus like Covid-19 is nonsense,' he said. Significantly, he pointed out that 'if you want to create a totally false panic about a totally false pandemic – pick a coronavirus'. This is exactly what the Cult-owned Gates, World Economic Forum and Johns Hopkins University did with their Event 201 'simulation' followed by their real-life simulation called the 'pandemic'. The scientist said that all you had to do was select the sickest of people with respiratory-type diseases in a single location – 'say Wuhan' – and administer PCR tests to them. You can then claim that anyone showing 'viral sequences' similar to a coronavirus 'which will inevitably be quite a few' is suffering from a 'new' disease:

Since you already selected the sickest flu cases a fairly high proportion of your sample will go on to die. You can then say this 'new' virus has a CFR [case fatality rate] higher than the flu and use this to infuse more concern and do more tests which will of course produce more 'cases', which expands the testing, which produces yet more 'cases' and so on and so on. Before long you have your 'pandemic', and all you have done is use a simple test kit trick to convert the worst flu and pneumonia cases into something new that doesn't ACTUALLY EXIST [my emphasis].

He said that you then 'just run the same scam in other countries' and make sure to keep the fear message running high 'so that people

will feel panicky and less able to think critically'. The only problem to overcome was the fact *there is no* actual new deadly pathogen and only regular sick people. This meant that deaths from the 'new deadly pathogen' were going to be way too low for a real new deadly virus pandemic, but he said this could be overcome in the following ways – all of which would go on to happen:

1. You can claim this is just the beginning and more deaths are imminent [you underpin this with fantasy 'computer projections']. Use this as an excuse to quarantine everyone and then claim the quarantine prevented the expected millions of dead.
2. You can [say that people] 'minimizing' the dangers are irresponsible and bully them into not talking about numbers.
3. You can talk crap about made up numbers hoping to blind people with pseudoscience.
4. You can start testing well people (who, of course, will also likely have shreds of coronavirus [RNA] in them) and thus inflate your 'case figures' with 'asymptomatic carriers' (you will of course have to spin that to sound deadly even though any virologist knows the more symptom-less cases you have the less deadly is your pathogen).

The scientist said that if you take these simple steps 'you can have your own entirely manufactured pandemic up and running in weeks'. His analysis made so early in the hoax was brilliantly prophetic of what would actually unfold. Pulling all the information together in these recent chapters we have this is simple 1, 2, 3, of how you can delude virtually the entire human population into believing in a 'virus' that doesn't exist:

- A 'Covid case' is someone who tests positive with a test not testing for the 'virus'.
- A 'Covid death' is someone who dies of *any cause* within 28 days (or much longer) of testing positive with a test not testing for the 'virus'.
- Asymptomatic means there is nothing wrong with you, but they claim you can pass on what you don't have to justify locking

down (quarantining) healthy people in totality.

The foundations of the hoax are that simple. A study involving ten million people in Wuhan, published in November, 2020, demolished the whole lie about those without symptoms passing on the 'virus'. They found '300 asymptomatic cases' and traced their contacts to find that not one of them was detected with the 'virus'.

'Asymptomatic' patients and their contacts were isolated for no less than two weeks and nothing changed. I know it's all crap, but if you are going to claim that those without symptoms can transmit 'the virus' then you must produce evidence for that and they never have. Even World Health Organization official Dr Maria Van Kerkhove, head of the emerging diseases and zoonosis unit, said as early as June, 2020, that she doubted the validity of asymptomatic transmission. She said that 'from the data we have, it still seems to be rare that an asymptomatic person actually transmits onward to a secondary individual' and by 'rare' she meant that she couldn't cite any case of asymptomatic transmission.

The Ferguson factor

The problem for the Cult as it headed into March, 2020, when the script had lockdown due to start, was that despite all the manipulation of the case and death figures they still did not have enough people alleged to have died from 'Covid' to justify mass house arrest. This was overcome in the way the scientist described: 'You can claim this is just the beginning and more deaths are imminent ... Use this as an excuse to quarantine everyone and then claim the quarantine prevented the expected millions of dead.' Enter one Professor Neil Ferguson, the Gates-funded 'epidemiologist' at the Gates-funded Imperial College in London. Ferguson is Britain's Christian Drosten in that he has a dire record of predicting health outcomes, but is still called upon to advise government on the next health outcome when another 'crisis' comes along. This may seem to be a strange and ridiculous thing to do. Why would you keep turning for policy guidance to people who have a history of being

monumentally wrong? Ah, but it makes sense from the Cult point of view. These 'experts' keep on producing predictions that suit the Cult agenda for societal transformation and so it was with Neil Ferguson as he revealed his horrific (and clearly insane) computer model predictions that allowed lockdowns to be imposed in Britain, the United States and many other countries. Ferguson does not have even an A-level in biology and would appear to have no formal training in computer modelling, medicine or epidemiology, according to Derek Winton, an MSc in Computational Intelligence. He wrote an article somewhat aghast at what Ferguson did which included taking no account of respiratory disease 'seasonality' which means it is far worse in the winter months. Who would have thought that respiratory disease could be worse in the winter? Well, certainly not Ferguson.

The massively China-connected Imperial College and its bizarre professor provided the excuse for the long-incubated Chinese model of human control to travel westward at lightning speed. Imperial College confirms on its website that it collaborates with the Chinese Research Institute; publishes more than 600 research papers every year with Chinese research institutions; has 225 Chinese staff; 2,600 Chinese students – the biggest international group; 7,000 former students living in China which is the largest group outside the UK; and was selected for a tour by China's President Xi Jinping during his state visit to the UK in 2015. The college takes major donations from China and describes itself as the UK's number one university collaborator with Chinese research institutions. The China communist/fascist government did not appear phased by the woeful predictions of Ferguson and Imperial when during the lockdown that Ferguson induced the college signed a five-year collaboration deal with China tech giant Huawei that will have Huawei's indoor 5G network equipment installed at the college's West London tech campus along with an 'AI cloud platform'. The deal includes Chinese sponsorship of Imperial's Venture Catalyst entrepreneurship competition. Imperial is an example of the enormous influence the Chinese government has within British and North American

universities and research centres – and further afield. Up to 200 academics from more than a dozen UK universities are being investigated on suspicion of ‘unintentionally’ helping the Chinese government build weapons of mass destruction by ‘transferring world-leading research in advanced military technology such as aircraft, missile designs and cyberweapons’. Similar scandals have broken in the United States, but it’s all a coincidence. Imperial College serves the agenda in many other ways including the promotion of every aspect of the United Nations Agenda 21/2030 (the Great Reset) and produced computer models to show that human-caused ‘climate change’ is happening when in the real world it isn’t. Imperial College is driving the climate agenda as it drives the ‘Covid’ agenda (both Cult hoaxes) while Patrick Vallance, the UK government’s Chief Scientific Adviser on ‘Covid’, was named Chief Scientific Adviser to the UN ‘climate change’ conference known as COP26 hosted by the government in Glasgow, Scotland. ‘Covid’ and ‘climate’ are fundamentally connected.

Professor Woeful

From Imperial’s bosom came Neil Ferguson still advising government despite his previous disasters and it was announced early on that he and other key people like UK Chief Medical Adviser Chris Whitty had caught the ‘virus’ as the propaganda story was being sold. Somehow they managed to survive and we had Prime Minister Boris Johnson admitted to hospital with what was said to be a severe version of the ‘virus’ in this same period. His whole policy and demeanour changed when he returned to Downing Street. It’s a small world with these government advisors – especially in their communal connections to Gates – and Ferguson had partnered with Whitty to write a paper called ‘Infectious disease: Tough choices to reduce Ebola transmission’ which involved another scare-story that didn’t happen. Ferguson’s ‘models’ predicted that up to 150,000 could die from ‘mad cow disease’, or BSE, and its version in sheep if it was transmitted to humans. BSE was not transmitted and instead triggered by an organophosphate pesticide used to treat a pest on

cows. Fewer than 200 deaths followed from the human form. Models by Ferguson and his fellow incompetents led to the unnecessary culling of millions of pigs, cattle and sheep in the foot and mouth outbreak in 2001 which destroyed the lives and livelihoods of farmers and their families who had often spent decades building their herds and flocks. Vast numbers of these animals did not have foot and mouth and had no contact with the infection. Another 'expert' behind the cull was Professor Roy Anderson, a computer modeller at Imperial College specialising in the epidemiology of *human*, not animal, disease. Anderson has served on the Bill and Melinda Gates Grand Challenges in Global Health advisory board and chairs another Gates-funded organisation. Gates is everywhere.

In a precursor to the 'Covid' script Ferguson backed closing schools 'for prolonged periods' over the swine flu 'pandemic' in 2009 and said it would affect a third of the world population if it continued to spread at the speed he claimed to be happening. His mates at Imperial College said much the same and a news report said: 'One of the authors, the epidemiologist and disease modeller Neil Ferguson, who sits on the World Health Organisation's emergency committee for the outbreak, said the virus had "full pandemic potential".' Professor Liam Donaldson, the Chris Whitty of his day as Chief Medical Officer, said the worst case could see 30 percent of the British people infected by swine flu with 65,000 dying. Ferguson and Donaldson were indeed proved correct when at the end of the year the number of deaths attributed to swine flu was 392. The term 'expert' is rather liberally applied unfortunately, not least to complete idiots. Swine flu 'projections' were great for GlaxoSmithKline (GSK) as millions rolled in for its Pandemrix influenza vaccine which led to brain damage with children most affected. The British government (taxpayers) paid out more than £60 million in compensation after GSK was given immunity from prosecution. Yet another 'Covid' déjà vu. Swine flu was supposed to have broken out in Mexico, but Dr Wolfgang Wodarg, a German doctor, former member of parliament and critic of the 'Covid' hoax, observed 'the spread of swine flu' in Mexico City at the time. He

said: 'What we experienced in Mexico City was a very mild flu which did not kill more than usual – which killed even fewer people than usual.' Hyping the fear against all the facts is not unique to 'Covid' and has happened many times before. Ferguson is reported to have over-estimated the projected death toll of bird flu (H5N1) by some three million-fold, but bird flu vaccine makers again made a killing from the scare. This is some of the background to the Neil Ferguson who produced the perfectly-timed computer models in early 2020 predicting that half a million people would die in Britain without draconian lockdown and 2.2 million in the United States. Politicians panicked, people panicked, and lockdowns of alleged short duration were instigated to 'flatten the curve' of cases gleaned from a test not testing for the 'virus'. I said at the time that the public could forget the 'short duration' bit. This was an agenda to destroy the livelihoods of the population and force them into mass control through dependency and there was going to be nothing 'short' about it. American researcher Daniel Horowitz described the consequences of the 'models' spewed out by Gates-funded Ferguson and Imperial College:

What led our government and the governments of many other countries into panic was a single Imperial College of UK study, funded by global warming activists, that predicted 2.2 million deaths if we didn't lock down the country. In addition, the reported 8-9% death rate in Italy scared us into thinking there was some other mutation of this virus that they got, which might have come here.

Together with the fact that we were finally testing and had the ability to actually report new cases, we thought we were headed for a death spiral. But again ... we can't flatten a curve if we don't know when the curve started.

How about it *never* started?

Giving them what they want

An investigation by German news outlet *Welt Am Sonntag* (*World on Sunday*) revealed how in March, 2020, the German government gathered together 'leading scientists from several research institutes and universities' and 'together, they were to produce a [modelling]

paper that would serve as legitimization for further tough political measures'. The Cult agenda was justified by computer modelling not based on evidence or reality; it was specifically constructed to justify the Cult demand for lockdowns all over the world to destroy the independent livelihoods of the global population. All these modellers and everyone responsible for the 'Covid' hoax have a date with a trial like those in Nuremberg after World War Two when Nazis faced the consequences of their war crimes. These corrupt-beyond-belief 'modellers' wrote the paper according to government instructions and it said that if lockdown measures were lifted then up to one million Germans would die from 'Covid-19' adding that some would die 'agonizingly at home, gasping for breath' unable to be treated by hospitals that couldn't cope. All lies. No matter – it gave the Cult all that it wanted. What did long-time government 'modeller' Neil Ferguson say? If the UK and the United States didn't lockdown half a million would die in Britain and 2.2 million Americans. Anyone see a theme here? 'Modellers' are such a crucial part of the lockdown strategy that we should look into their background and follow the money. Researcher Rosemary Frei produced an excellent article headlined 'The Modelling-paper Mafiosi'. She highlights a guy called John Edmunds, a British epidemiologist, and professor in the Faculty of Epidemiology and Population Health at the London School of Hygiene & Tropical Medicine. He studied at Imperial College. Edmunds is a member of government 'Covid' advisory bodies which have been dictating policy, the New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) and the Scientific Advisory Group for Emergencies (SAGE).

Ferguson, another member of NERVTAG and SAGE, led the way with the original 'virus' and Edmunds has followed in the 'variant' stage and especially the so-called UK or Kent variant known as the 'Variant of Concern' (VOC) B.1.1.7. He said in a co-written report for the Centre for Mathematical modelling of Infectious Diseases at the London School of Hygiene and Tropical Medicine, with input from the Centre's 'Covid-19' Working Group, that there was 'a realistic

possibility that VOC B.1.1.7 is associated with an increased risk of death compared to non-VOC viruses'. Fear, fear, fear, get the vaccine, fear, fear, fear, get the vaccine. Rosemary Frei reveals that almost all the paper's authors and members of the modelling centre's 'Covid-19' Working Group receive funding from the Bill and Melinda Gates Foundation and/or the associated Gates-funded Wellcome Trust. The paper was published by e-journal *Medrx* ^{xiv} which only publishes papers not peer-reviewed and the journal was established by an organisation headed by Facebook's Mark Zuckerberg and his missus. What a small world it is. Frei discovered that Edmunds is on the Scientific Advisory Board of the Coalition for Epidemic Preparedness Innovations (CEPI) which was established by the Bill and Melinda Gates Foundation, Klaus Schwab's Davos World Economic Forum and Big Pharma giant Wellcome. CEPI was 'launched in Davos [in 2017] to develop vaccines to stop future epidemics', according to its website. 'Our mission is to accelerate the development of vaccines against emerging infectious diseases and enable equitable access to these vaccines for people during outbreaks.' What kind people they are. Rosemary Frei reveals that Public Health England (PHE) director Susan Hopkins is an author of her organisation's non-peer-reviewed reports on 'new variants'. Hopkins is a professor of infectious diseases at London's Imperial College which is gifted tens of millions of dollars a year by the Bill and Melinda Gates Foundation. Gates-funded modelling disaster Neil Ferguson also co-authors Public Health England reports and he spoke in December, 2020, about the potential danger of the B.1.1.7. 'UK variant' promoted by Gates-funded modeller John Edmunds. When I come to the 'Covid vaccines' the 'new variants' will be shown for what they are – bollocks.

Connections, connections

All these people and modellers are lockdown-obsessed or, put another way, they demand what the Cult demands. Edmunds said in January, 2021, that to ease lockdowns too soon would be a disaster and they had to 'vaccinate much, much, much more widely than the

elderly'. Rosemary Frei highlights that Edmunds is married to Jeanne Pimenta who is described in a LinkedIn profile as director of epidemiology at GlaxoSmithKline (GSK) and she held shares in the company. Patrick Vallance, co-chair of SAGE and the government's Chief Scientific Adviser, is a former executive of GSK and has a deferred bonus of shares in the company worth £600,000. GSK has serious business connections with Bill Gates and is collaborating with mRNA-'vaccine' company CureVac to make 'vaccines' for the new variants that Edmunds is talking about. GSK is planning a 'Covid vaccine' with drug giant Sanofi. Puppets Prime Minister Boris Johnson announced in the spring of 2021 that up to 60 million vaccine doses were to be made at the GSK facility at Barnard Castle in the English North East. Barnard Castle, with a population of just 6,000, was famously visited in breach of lockdown rules in April, 2020, by Johnson aide Dominic Cummings who said that he drove there 'to test his eyesight' before driving back to London. Cummings would be better advised to test his integrity – not that it would take long. The GSK facility had nothing to do with his visit then although I'm sure Patrick Vallance would have been happy to arrange an introduction and some tea and biscuits. Ruthless psychopath Gates has made yet another fortune from vaccines in collaboration with Big Pharma companies and gushes at the phenomenal profits to be made from vaccines – more than a 20-to-1 return as he told one interviewer. Gates also tweeted in December, 2019, with the foreknowledge of what was coming: 'What's next for our foundation? I'm particularly excited about what the next year could mean for one of the best buys in global health: vaccines.'

Modeller John Edmunds is a big promotor of vaccines as all these people appear to be. He's the dean of the London School of Hygiene & Tropical Medicine's Faculty of Epidemiology and Population Health which is primarily funded by the Bill and Melinda Gates Foundation and the Gates-established and funded GAVI vaccine alliance which is the Gates vehicle to vaccinate the world. The organisation Doctors Without Borders has described GAVI as being 'aimed more at supporting drug-industry desires to promote new

products than at finding the most efficient and sustainable means for fighting the diseases of poverty'. But then that's why the psychopath Gates created it. John Edmunds said in a video that the London School of Hygiene & Tropical Medicine is involved in every aspect of vaccine development including large-scale clinical trials. He contends that mathematical modelling can show that vaccines protect individuals and society. That's on the basis of shit in and shit out, I take it. Edmunds serves on the UK Vaccine Network as does Ferguson and the government's foremost 'Covid' adviser, the grim-faced, dark-eyed Chris Whitty. The Vaccine Network says it works 'to support the government to identify and shortlist targeted investment opportunities for the most promising vaccines and vaccine technologies that will help combat infectious diseases with epidemic potential, and to address structural issues related to the UK's broader vaccine infrastructure'. Ferguson is acting Director of the Imperial College Vaccine Impact Modelling Consortium which has funding from the Bill and Melina Gates Foundation and the Gates-created GAVI 'vaccine alliance'. Anyone wonder why these characters see vaccines as the answer to every problem? Ferguson is wildly enthusiastic in his support for GAVI's campaign to vaccinate children en masse in poor countries. You would expect someone like Gates who has constantly talked about the need to reduce the population to want to fund vaccines to keep more people alive. I'm sure that's why he does it. The John Edmunds London School of Hygiene & Tropical Medicine (LSHTM) has a Vaccines Manufacturing Innovation Centre which develops, tests and commercialises vaccines. Rosemary Frei writes:

The vaccines centre also performs affiliated activities like combating 'vaccine hesitancy'. The latter includes the Vaccine Confidence Project. The project's stated purpose is, among other things, 'to provide analysis and guidance for early response and engagement with the public to ensure sustained confidence in vaccines and immunisation'. The Vaccine Confidence Project's director is LSHTM professor Heidi Larson. For more than a decade she's been researching how to combat vaccine hesitancy.

How the bloody hell can blokes like John Edmunds and Neil Ferguson with those connections and financial ties model 'virus' case

and death projections for the government and especially in a way that gives their paymasters like Gates exactly what they want? It's insane, but this is what you find throughout the world.

'Covid' is not dangerous, oops, wait, yes it is

Only days before Ferguson's nightmare scenario made Jackboot Johnson take Britain into a China-style lockdown to save us from a deadly 'virus' the UK government website gov.uk was reporting something very different to Ferguson on a page of official government guidance for 'high consequence infectious diseases (HCID)'. It said this about 'Covid-19':

As of 19 March 2020, COVID-19 is no longer considered to be a high consequence infectious diseases (HCID) in the UK [my emphasis]. The 4 nations public health HCID group made an interim recommendation in January 2020 to classify COVID-19 as an HCID. This was based on consideration of the UK HCID criteria about the virus and the disease with information available during the early stages of the outbreak.

Now that more is known about COVID-19, the public health bodies in the UK have reviewed the most up to date information about COVID-19 against the UK HCID criteria. They have determined that several features have now changed; in particular, more information is available about mortality rates (low overall), and there is now greater clinical awareness and a specific and sensitive laboratory test, the availability of which continues to increase. The Advisory Committee on Dangerous Pathogens (ACDP) is also of the opinion that COVID-19 should no longer be classified as an HCID.

Soon after the government had been exposed for downgrading the risk they upgraded it again and everyone was back to singing from the same Cult hymn book. Ferguson and his fellow Gates clones indicated that lockdowns and restrictions would have to continue until a Gates-funded vaccine was developed. Gates said the same because Ferguson and his like were repeating the Gates script which is the Cult script. 'Flatten the curve' became an ongoing nightmare of continuing lockdowns with periods in between of severe restrictions in pursuit of destroying independent incomes and had nothing to do with protecting health about which the Cult gives not a shit. Why wouldn't Ferguson be pushing a vaccine 'solution' when he's owned by vaccine-obsessive Gates who makes a fortune from them and

when Ferguson heads the Vaccine Impact Modelling Consortium at Imperial College funded by the Gates Foundation and GAVI, the 'vaccine alliance', created by Gates as his personal vaccine promotion operation? To compound the human catastrophe that Ferguson's 'models' did so much to create he was later exposed for breaking his own lockdown rules by having sexual liaisons with his married girlfriend Antonia Staats at his home while she was living at another location with her husband and children. Staats was a 'climate' activist and senior campaigner at the Soros-funded Avaaz which I wouldn't trust to tell me that grass is green. Ferguson had to resign as a government advisor over this hypocrisy in May, 2020, but after a period of quiet he was back being quoted by the ridiculous media on the need for more lockdowns and a vaccine rollout. Other government-advising 'scientists' from Imperial College held the fort in his absence and said lockdown could be indefinite until a vaccine was found. The Cult script was being sung by the payrolled choir. I said there was no intention of going back to 'normal' when the 'vaccine' came because the 'vaccine' is part of a very different agenda that I will discuss in Human 2.0. Why would the Cult want to let the world go back to normal when destroying that normal forever was the whole point of what was happening? House arrest, closing businesses and schools through lockdown, (un)social distancing and masks all followed the Ferguson fantasy models. Again as I predicted (these people are so predictable) when the 'vaccine' arrived we were told that house arrest, lockdown, (un)social distancing and masks would still have to continue. I will deal with the masks in the next chapter because they are of fundamental importance.

Where's the 'pandemic'?

Any mildly in-depth assessment of the figures revealed what was really going on. Cult-funded and controlled organisations still have genuine people working within them such is the number involved. So it is with Genevieve Briand, assistant program director of the Applied Economics master's degree program at Johns Hopkins

University. She analysed the impact that 'Covid-19' had on deaths from *all* causes in the United States using official data from the CDC for the period from early February to early September, 2020. She found that allegedly 'Covid' *related*-deaths exceeded those from heart disease which she found strange with heart disease always the biggest cause of fatalities. Her research became even more significant when she noted the sudden decline in 2020 of *all* non-'Covid' deaths: 'This trend is completely contrary to the pattern observed in all previous years ... the total decrease in deaths by other causes almost exactly equals the increase in deaths by Covid-19.' This was such a game, set and match in terms of what was happening that Johns Hopkins University deleted the article on the grounds that it 'was being used to support false and dangerous inaccuracies about the impact of the pandemic'. No – because it exposed the scam from official CDC figures and this was confirmed when those figures were published in January, 2021. Here we can see the effect of people dying from heart attacks, cancer, road accidents and gunshot wounds – *anything* – having 'Covid-19' on the death certificate along with those diagnosed from 'symptoms' who had even not tested positive with a test not testing for the 'virus'. I am not kidding with the gunshot wounds, by the way. Brenda Bock, coroner in Grand County, Colorado, revealed that two gunshot victims tested positive for the 'virus' within the previous 30 days and were therefore classified as 'Covid deaths'. Bock said: 'These two people had tested positive for Covid, but that's not what killed them. A gunshot wound is what killed them.' She said she had not even finished her investigation when the state listed the gunshot victims as deaths due to the 'virus'. The death and case figures for 'Covid-19' are an absolute joke and yet they are repeated like parrots by the media, politicians and alleged medical 'experts'. The official Cult narrative is the only show in town.

Genevieve Briand found that deaths from all causes were not exceptional in 2020 compared with previous years and a Spanish magazine published figures that said the same about Spain which was a 'Covid' propaganda hotspot at one point. *Discovery Salud*, a

health and medicine magazine, quoted government figures which showed how 17,000 *fewer* people died in Spain in 2020 than in 2019 and more than 26,000 fewer than in 2018. The age-standardised mortality rate for England and Wales when age distribution is taken into account was significantly lower in 2020 than the 1970s, 80s and 90s, and was only the ninth highest since 2000. Where is the 'pandemic'?

Post mortems and autopsies virtually disappeared for 'Covid' deaths amid claims that 'virus-infected' bodily fluids posed a risk to those carrying out the autopsy. This was rejected by renowned German pathologist and forensic doctor Klaus Püschel who said that he and his staff had by then done 150 autopsies on 'Covid' patients with no problems at all. He said they were needed to know why some 'Covid' patients suffered blood clots and not severe respiratory infections. The 'virus' is, after all, called SARS or 'severe acute respiratory syndrome'. I highlighted in the spring of 2020 this phenomenon and quoted New York intensive care doctor Cameron Kyle-Sidell who posted a soon deleted YouTube video to say that they had been told to prepare to treat an infectious disease called 'Covid-19', but that was not what they were dealing with. Instead he likened the lung condition of the most severely ill patients to what you would expect with cabin depressurisation in a plane at 30,000 feet or someone dropped on the top of Everest without oxygen or acclimatisation. I have never said this is not happening to a small minority of alleged 'Covid' patients – I am saying this is not caused by a phantom 'contagious virus'. Indeed Kyle-Sidell said that 'Covid-19' was not the disease they were told was coming their way. 'We are operating under a medical paradigm that is untrue,' he said, and he believed they were treating the wrong disease: 'These people are being slowly starved of oxygen.' Patients would take off their oxygen masks in a state of fear and stress and while they were blue in the face on the brink of death. They did not look like patients dying of pneumonia. You can see why they don't want autopsies when their virus doesn't exist and there is another condition in some people that they don't wish to be uncovered. I should add here that

the 5G system of millimetre waves was being rapidly introduced around the world in 2020 and even more so now as they fire 5G at the Earth from satellites. At 60 gigahertz within the 5G range that frequency interacts with the oxygen molecule and stops people breathing in sufficient oxygen to be absorbed into the bloodstream. They are installing 5G in schools and hospitals. The world is not mad or anything. 5G can cause major changes to the lungs and blood as I detail in *The Answer* and these consequences are labelled 'Covid-19', the alleged symptoms of which can be caused by 5G and other electromagnetic frequencies as cells respond to radiation poisoning.

The 'Covid death' scam

Dr Scott Jensen, a Minnesota state senator and medical doctor, exposed 'Covid' Medicare payment incentives to hospitals and death certificate manipulation. He said he was sent a seven-page document by the US Department of Health 'coaching' him on how to fill out death certificates which had never happened before. The document said that he didn't need to have a laboratory test for 'Covid-19' to put that on the death certificate and that shocked him when death certificates are supposed to be about facts. Jensen described how doctors had been 'encouraged, if not pressured' to make a diagnosis of 'Covid-19' if they thought it was probable or '*presumed*'. No positive test was necessary – not that this would have mattered anyway. He said doctors were told to diagnose 'Covid' by symptoms when these were the same as colds, allergies, other respiratory problems, and certainly with influenza which 'disappeared' in the 'Covid' era. A common sniffle was enough to get the dreaded verdict. Ontario authorities decreed that a single care home resident with *one* symptom from a long list must lead to the isolation of the entire home. Other courageous doctors like Jensen made the same point about death figure manipulation and how deaths by other causes were falling while 'Covid-19 deaths' were rising at the same rate due to re-diagnosis. Their videos rarely survive long on YouTube with its Cult-supporting algorithms courtesy of CEO Susan Wojcicki and her bosses at Google. Figure-tampering was so glaring

and ubiquitous that even officials were letting it slip or outright saying it. UK chief scientific adviser Patrick Vallance said on one occasion that 'Covid' on the death certificate doesn't mean 'Covid' was the cause of death (so why the hell is it there?) and we had the rare sight of a BBC reporter telling the truth when she said: 'Someone could be successfully treated for Covid, in say April, discharged, and then in June, get run over by a bus and die ... That person would still be counted as a Covid death in England.' Yet the BBC and the rest of the world media went on repeating the case and death figures as if they were real. Illinois Public Health Director Dr Ngozi Ezike revealed the deceit while her bosses must have been clenching their buttocks:

If you were in a hospice and given a few weeks to live and you were then found to have Covid that would be counted as a Covid death. [There might be] a clear alternate cause, but it is still listed as a Covid death. So everyone listed as a Covid death doesn't mean that was the cause of the death, but that they had Covid at the time of death.

Yes, a 'Covid virus' never shown to exist and tested for with a test not testing for the 'virus'. In the first period of the pandemic hoax through the spring of 2020 the process began of designating almost everything a 'Covid' death and this has continued ever since. I sat in a restaurant one night listening to a loud conversation on the next table where a family was discussing in bewilderment how a relative who had no symptoms of 'Covid', and had died of a long-term problem, could have been diagnosed a death by the 'virus'. I could understand their bewilderment. If they read this book they will know why this medical fraud has been perpetrated the world over.

Some media truth shock

The media ignored the evidence of death certificate fraud until eventually one columnist did speak out when she saw it first-hand. Bel Mooney is a long-time national newspaper journalist in Britain currently working for the *Daily Mail*. Her article on February 19th, 2021, carried this headline: 'My dad Ted passed three Covid tests

and died of a chronic illness yet he's officially one of Britain's 120,000 victims of the virus and is far from alone ... so how many more are there?' She told how her 99-year-old father was in a care home with a long-standing chronic obstructive pulmonary disease and vascular dementia. Maybe, but he was still aware enough to tell her from the start that there was no 'virus' and he refused the 'vaccine' for that reason. His death was not unexpected given his chronic health problems and Mooney said she was shocked to find that 'Covid-19' was declared the cause of death on his death certificate. She said this was a 'bizarre and unacceptable untruth' for a man with long-time health problems who had tested negative twice at the home for the 'virus'. I was also shocked by this story although not by what she said. I had been highlighting the death certificate manipulation for ten months. It was the confirmation that a professional full-time journalist only realised this was going on when it affected her directly and neither did she know that whether her dad tested positive or negative was irrelevant with the test not testing for the 'virus'. Where had she been? She said she did not believe in 'conspiracy theories' without knowing I'm sure that this and 'conspiracy theorists' were terms put into widespread circulation by the CIA in the 1960s to discredit those who did not accept the ridiculous official story of the Kennedy assassination. A blanket statement of 'I don't believe in conspiracy theories' is always bizarre. The dictionary definition of the term alone means the world is drowning in conspiracies. What she said was even more daft when her dad had just been affected by the 'Covid' conspiracy. Why else does she think that 'Covid-19' was going on the death certificates of people who died of something else?

To be fair once she saw from personal experience what was happening she didn't mince words. Mooney was called by the care home on the morning of February 9th to be told her father had died in his sleep. When she asked for the official cause of death what came back was 'Covid-19'. Mooney challenged this and was told there had been deaths from Covid on the dementia floor (confirmed by a test not testing for the 'virus') so they considered it 'reasonable

to assume'. 'But doctor,' Mooney rightly protested, 'an assumption isn't a diagnosis.' She said she didn't blame the perfectly decent and sympathetic doctor – 'he was just doing his job'. Sorry, but that's *bullshit*. He wasn't doing his job at all. He was putting a false cause of death on the death certificate and that is a criminal offence for which he should be brought to account and the same with the millions of doctors worldwide who have done the same. They were not doing their job they were following orders and that must not wash at new Nuremberg trials any more than it did at the first ones. Mooney's doctor was 'assuming' (presuming) as he was told to, but 'just following orders' makes no difference to his actions. A doctor's job is to serve the patient and the truth, not follow orders, but that's what they have done all over the world and played a central part in making the 'Covid' hoax possible with all its catastrophic consequences for humanity. Shame on them and they must answer for their actions. Mooney said her disquiet worsened when she registered her father's death by telephone and was told by the registrar there had been very many other cases like hers where 'the deceased' had not tested positive for 'Covid' yet it was recorded as the cause of death. The test may not matter, but those involved at their level *think* it matters and it shows a callous disregard for accurate diagnosis. The pressure to do this is coming from the top of the national 'health' pyramids which in turn obey the World Health Organization which obeys Gates and the Cult. Mooney said the registrar agreed that this must distort the national figures adding that 'the strangest thing is that every winter we record countless deaths from flu, and this winter there have been none. Not one!' She asked if the registrar thought deaths from flu were being misdiagnosed and lumped together with 'Covid' deaths. The answer was a 'puzzled yes'. Mooney said that the funeral director said the same about 'Covid' deaths which had nothing to do with 'Covid'. They had lost count of the number of families upset by this and other funeral companies in different countries have had the same experience. Mooney wrote:

The nightly shroud-waving and shocking close-ups of pain imposed on us by the TV news bewildered and terrified the population into eager compliance with lockdowns. We were invited to 'save the NHS' and to grieve for strangers – the real-life loved ones behind those shocking death counts. Why would the public imagine what I now fear, namely that the way Covid-19 death statistics are compiled might make the numbers seem greater than they are?

Oh, just a little bit – like 100 percent.

Do the maths

Mooney asked why a country would wish to skew its mortality figures by wrongly certifying deaths? What had been going on? Well, if you don't believe in conspiracies you will never find the answer which is that *it's a conspiracy*. She did, however, describe what she had discovered as a 'national scandal'. In reality it's a global scandal and happening everywhere. Pillars of this conspiracy were all put into place before the button was pressed with the Drosten PCR protocol and high amplifications to produce the cases and death certificate changes to secure illusory 'Covid' deaths. Mooney notes that normally two doctors were needed to certify a death, with one having to know the patient, and how the rules were changed in the spring of 2020 to allow one doctor to do this. In the same period 'Covid deaths' were decreed to be all cases where Covid-19 was put on the death certificate even without a positive test or any symptoms. Mooney asked: 'How many of the 30,851 (as of January 15) care home resident deaths with Covid-19 on the certificate (32.4 per cent of all deaths so far) were based on an assumption, like that of my father? And what has that done to our national psyche?' All of them is the answer to the first question and it has devastated and dismantled the national psyche, actually the global psyche, on a colossal scale. In the UK case and death data is compiled by organisations like Public Health England (PHE) and the Office for National Statistics (ONS). Mooney highlights the insane policy of counting a death from any cause as 'Covid-19' if this happens within 28 days of a positive test (with a test not testing for the 'virus') and she points out that ONS statistics reflect deaths 'involving Covid' 'or due to Covid' which meant in practice any

death where 'Covid-19' was mentioned on the death certificate. She described the consequences of this fraud:

Most people will accept the narrative they are fed, so panicky governments here and in Europe witnessed the harsh measures enacted in totalitarian China and jumped into lockdown. Headlines about Covid deaths tolled like the knell that would bring doomsday to us all. Fear stalked our empty streets. Politicians parroted the frankly ridiculous aim of 'zero Covid' and shut down the economy, while most British people agreed that lockdown was essential and (astonishingly to me, as a patriotic Brit) even wanted more restrictions.

For what? Lies on death certificates? Never mind the grim toll of lives ruined, suicides, schools closed, rising inequality, depression, cancelled hospital treatments, cancer patients in a torture of waiting, poverty, economic devastation, loneliness, families kept apart, and so on. How many lives have been lost as a direct result of lockdown?

She said that we could join in a national chorus of shock and horror at reaching the 120,000 death toll which was surely certain to have been totally skewed all along, but what about the human cost of lockdown justified by these 'death figures'? *The British Medical Journal* had reported a 1,493 percent increase in cases of children taken to Great Ormond Street Hospital with abusive head injuries alone and then there was the effect on families:

Perhaps the most shocking thing about all this is that families have been kept apart – and obeyed the most irrational, changing rules at the whim of government – because they believed in the statistics. They succumbed to fear, which his generation rejected in that war fought for freedom. Dad (God rest his soul) would be angry. And so am I.

Another theme to watch is that in the winter months when there are more deaths from all causes they focus on 'Covid' deaths and in the summer when the British Lung Foundation says respiratory disease plummets by 80 percent they rage on about 'cases'. Either way fascism on population is always the answer.

Nazi eugenics in the 21st century

Elderly people in care homes have been isolated from their families month after lonely month with no contact with relatives and grandchildren who were banned from seeing them. We were told

that lockdown fascism was to 'protect the vulnerable' like elderly people. At the same time Do Not Resuscitate (DNR) orders were placed on their medical files so that if they needed resuscitation it wasn't done and 'Covid-19' went on their death certificates. Old people were not being 'protected' they were being culled – murdered in truth. DNR orders were being decreed for disabled and young people with learning difficulties or psychological problems. The UK Care Quality Commission, a non-departmental body of the Department of Health and Social Care, found that 34 percent of those working in health and social care were pressured into placing 'do not attempt cardiopulmonary resuscitation' orders on 'Covid' patients who suffered from disabilities and learning difficulties without involving the patient or their families in the decision. UK judges ruled that an elderly woman with dementia should have the DNA-manipulating 'Covid vaccine' against her son's wishes and that a man with severe learning difficulties should have the jab despite his family's objections. Never mind that many had already died. The judiciary always supports doctors and government in fascist dictatorships. They wouldn't dare do otherwise. A horrific video was posted showing fascist officers from Los Angeles police forcibly giving the 'Covid' shot to women with special needs who were screaming that they didn't want it. The same fascists are seen giving the jab to a sleeping elderly woman in a care home. This is straight out of the Nazi playbook. Hitler's Nazis committed mass murder of the mentally ill and physically disabled throughout Germany and occupied territories in the programme that became known as Aktion T4, or just T4. Sabbatian-controlled Hitler and his grotesque crazies set out to kill those they considered useless and unnecessary. The Reich Committee for the Scientific Registering of Hereditary and Congenital Illnesses registered the births of babies identified by physicians to have 'defects'. By 1941 alone more than 5,000 children were murdered by the state and it is estimated that in total the number of innocent people killed in Aktion T4 was between 275,000 and 300,000. Parents were told their children had been sent away for 'special treatment' never to return. It is rather pathetic to see claims about plans for new extermination camps being dismissed today

when the same force behind current events did precisely that 80 years ago. Margaret Sanger was a Cult operative who used 'birth control' to sanitise her programme of eugenics. Organisations she founded became what is now Planned Parenthood. Sanger proposed that 'the whole dysgenic population would have its choice of segregation or sterilization'. These included epileptics, 'feeble-minded', and prostitutes. Sanger opposed charity because it perpetuated 'human waste'. She reveals the Cult mentality and if anyone thinks that extermination camps are a 'conspiracy theory' their naivety is touching if breathtakingly stupid.

If you don't believe that doctors can act with callous disregard for their patients it is worth considering that doctors and medical staff agreed to put government-decreed DNR orders on medical files and do nothing when resuscitation is called for. I don't know what you call such people in your house. In mine they are Nazis from the Josef Mengele School of Medicine. Phenomenal numbers of old people have died worldwide from the effects of lockdown, depression, lack of treatment, the 'vaccine' (more later) and losing the will to live. A common response at the start of the manufactured pandemic was to remove old people from hospital beds and transfer them to nursing homes. The decision would result in a mass cull of elderly people in those homes through lack of treatment – *not* 'Covid'. Care home whistleblowers have told how once the 'Covid' era began doctors would not come to their homes to treat patients and they were begging for drugs like antibiotics that often never came. The most infamous example was ordered by New York governor Andrew Cuomo, brother of a moronic CNN host, who amazingly was given an Emmy Award for his handling of the 'Covid crisis' by the ridiculous Wokers that hand them out. Just how ridiculous could be seen in February, 2021, when a Department of Justice and FBI investigation began into how thousands of old people in New York died in nursing homes after being discharged from hospital to make way for 'Covid' patients on Cuomo's say-so – and how he and his staff covered up these facts. This couldn't have happened to a nicer psychopath. Even then there was a 'Covid' spin. Reports said that

thousands of old people who tested positive for 'Covid' in hospital were transferred to nursing homes to both die of 'Covid' and transmit it to others. No – they were in hospital because they were ill and the fact that they tested positive with a test not testing for the 'virus' is irrelevant. They were ill often with respiratory diseases ubiquitous in old people near the end of their lives. Their transfer out of hospital meant that their treatment stopped and many would go on to die.

They're old. Who gives a damn?

I have exposed in the books for decades the Cult plan to cull the world's old people and even to introduce at some point what they call a 'demise pill' which at a certain age everyone would take and be out of here by law. In March, 2021, Spain legalised euthanasia and assisted suicide following the Netherlands, Belgium, Luxembourg and Canada on the Tiptoe to the demise pill. Treatment of old people by many 'care' homes has been a disgrace in the 'Covid' era. There are many, many, caring staff – I know some. There have, however, been legions of stories about callous treatment of old people and their families. Police were called when families came to take their loved ones home in the light of isolation that was killing them. They became prisoners of the state. Care home residents in insane, fascist Ontario, Canada, were not allowed to leave their *room* once the 'Covid' hoax began. UK staff have even wheeled elderly people away from windows where family members were talking with them. Oriana Criscuolo from Stockport in the English North West dropped off some things for her 80-year-old father who has Parkinson's disease and dementia and she wanted to wave to him through a ground-floor window. She was told that was 'illegal'. When she went anyway they closed the curtains in the middle of the day. Oriana said:

It's just unbelievable. I cannot understand how care home staff – people who are being paid to care – have become so uncaring. Their behaviour is inhumane and cruel. It's beyond belief.

She was right and this was not a one-off. What a way to end your life in such loveless circumstances. UK registered nurse Nicky Millen, a proper old school nurse for 40 years, said that when she started her career care was based on dignity, choice, compassion and empathy. Now she said 'the things that are important to me have gone out of the window.' She was appalled that people were dying without their loved ones and saying goodbye on iPads. Nicky described how a distressed 89-year-old lady stroked her face and asked her 'how many paracetamol would it take to finish me off'. Life was no longer worth living while not seeing her family. Nicky said she was humiliated in front of the ward staff and patients for letting the lady stroke her face and giving her a cuddle. Such is the dehumanisation that the 'Covid' hoax has brought to the surface. Nicky worked in care homes where patients told her they were being held prisoner. 'I want to live until I die', one said to her. 'I had a lady in tears because she hadn't seen her great-grandson.' Nicky was compassionate old school meeting psychopathic New Normal. She also said she had worked on a 'Covid' ward with no 'Covid' patients. Jewish writer Shai Held wrote an article in March, 2020, which was headlined 'The Staggering, Heartless Cruelty Toward the Elderly'. What he described was happening from the earliest days of lockdown. He said 'the elderly' were considered a group and not unique individuals (the way of the Woke). Shai Held said:

Notice how the all-too-familiar rhetoric of dehumanization works: 'The elderly' are bunched together as a faceless mass, all of them considered culprits and thus effectively deserving of the suffering the pandemic will inflict upon them. Lost entirely is the fact that the elderly are individual human beings, each with a distinctive face and voice, each with hopes and dreams, memories and regrets, friendships and marriages, loves lost and loves sustained.

'The elderly' have become another dehumanised group for which anything goes and for many that has resulted in cold disregard for their rights and their life. The distinctive face that Held talks about is designed to be deleted by masks until everyone is part of a faceless mass.

'War-zone' hospitals myth

Again and again medical professionals have told me what was really going on and how hospitals 'overrun like war zones' according to the media were virtually empty. The mantra from medical whistleblowers was please don't use my name or my career is over. Citizen journalists around the world sneaked into hospitals to film evidence exposing the 'war-zone' lie. They really *were* largely empty with closed wards and operating theatres. I met a hospital worker in my town on the Isle of Wight during the first lockdown in 2020 who said the only island hospital had never been so quiet. Lockdown was justified by the psychopaths to stop hospitals being overrun. At the same time that the island hospital was near-empty the military arrived here to provide *extra beds*. It was all propaganda to ramp up the fear to ensure compliance with fascism as were never-used temporary hospitals with thousands of beds known as Nightingales and never-used make-shift mortuaries opened by the criminal UK government. A man who helped to install those extra island beds attributed to the army said they were never used and the hospital was empty. Doctors and nurses 'stood around talking or on their phones, wandering down to us to see what we were doing'. There were no masks or social distancing. He accused the useless local island paper, the *County Press*, of 'pumping the fear as if our hospital was overrun and we only have one so it should have been'. He described ambulances parked up with crews outside in deck chairs. When his brother called an ambulance he was told there was a two-hour backlog which he called 'bullshit'. An old lady on the island fell 'and was in a bad way', but a caller who rang for an ambulance was told the situation wasn't urgent enough. Ambulance stations were working under capacity while people would hear ambulances with sirens blaring driving through the streets. When those living near the stations realised what was going on they would follow them as they left, circulated around an urban area with the sirens going, and then came back without stopping. All this was to increase levels of fear and the same goes for the 'ventilator shortage crisis' that cost tens of millions for hastily produced ventilators never to be used.

Ambulance crews that agreed to be exploited in this way for fear propaganda might find themselves a mirror. I wish them well with that. Empty hospitals were the obvious consequence of treatment and diagnoses of non-'Covid' conditions cancelled and those involved handed a death sentence. People have been dying at home from undiagnosed and untreated cancer, heart disease and other life-threatening conditions to allow empty hospitals to deal with a 'pandemic' that wasn't happening.

Death of the innocent

'War-zones' have been laying off nursing staff, even doctors where they can. There was no work for them. Lockdown was justified by saving lives and protecting the vulnerable they were actually killing with DNR orders and preventing empty hospitals being 'overrun'. In Britain the mantra of stay at home to 'save the NHS' was everywhere and across the world the same story was being sold when it was all lies. Two California doctors, Dan Erickson and Artin Massihi at Accelerated Urgent Care in Bakersfield, held a news conference in April, 2020, to say that intensive care units in California were 'empty, essentially', with hospitals shutting floors, not treating patients and laying off doctors. The California health system was working at minimum capacity 'getting rid of doctors because we just don't have the volume'. They said that people with conditions such as heart disease and cancer were not coming to hospital out of fear of 'Covid-19'. Their video was deleted by Susan Wojcicki's Cult-owned YouTube after reaching five million views. Florida governor Ron Desantis, who rejected the severe lockdowns of other states and is being targeted for doing so, said that in March, 2020, every US governor was given models claiming they would run out of hospital beds in days. That was never going to happen and the 'modellers' knew it. Deceit can be found at every level of the system. Urgent children's operations were cancelled including fracture repairs and biopsies to spot cancer. Eric Nicholls, a consultant paediatrician, said 'this is obviously concerning and we need to return to normal operating and to increase capacity as soon as possible'. Psychopaths

in power were rather less concerned *because* they are psychopaths. Deletion of urgent care and diagnosis has been happening all over the world and how many kids and others have died as a result of the actions of these cold and heartless lunatics dictating 'health' policy? The number must be stratospheric. Richard Sullivan, professor of cancer and global health at King's College London, said people feared 'Covid' more than cancer such was the campaign of fear. 'Years of lost life will be quite dramatic', Sullivan said, with 'a huge amount of avoidable mortality'. Sarah Woolnough, executive director for policy at Cancer Research UK, said there had been a 75 percent drop in urgent referrals to hospitals by family doctors of people with suspected cancer. Sullivan said that 'a lot of services have had to scale back – we've seen a dramatic decrease in the amount of elective cancer surgery'. Lockdown deaths worldwide has been absolutely fantastic with the *New York Post* reporting how data confirmed that 'lockdowns end more lives than they save':

There was a sharp decline in visits to emergency rooms and an increase in fatal heart attacks because patients didn't receive prompt treatment. Many fewer people were screened for cancer. Social isolation contributed to excess deaths from dementia and Alzheimer's.

Researchers predicted that the social and economic upheaval would lead to tens of thousands of "deaths of despair" from drug overdoses, alcoholism and suicide. As unemployment surged and mental-health and substance-abuse treatment programs were interrupted, the reported levels of anxiety, depression and suicidal thoughts increased dramatically, as did alcohol sales and fatal drug overdoses.

This has been happening while nurses and other staff had so much time on their hands in the 'war-zones' that Tic-Tok dancing videos began appearing across the Internet with medical staff dancing around in empty wards and corridors as people died at home from causes that would normally have been treated in hospital.

Mentions in dispatches

One brave and truth-committed whistleblower was Louise Hampton, a call handler with the UK NHS who made a viral Internet video saying she had done 'fuck all' during the 'pandemic'

which was 'a load of bollocks'. She said that 'Covid-19' was rebranded flu and of course she lost her job. This is what happens in the medical and endless other professions now when you tell the truth. Louise filmed inside 'war-zone' accident and emergency departments to show they were empty and I mean *empty* as in no one there. The mainstream media could have done the same and blown the gaff on the whole conspiracy. They haven't to their eternal shame. Not that most 'journalists' seem capable of manifesting shame as with the psychopaths they slavishly repeat without question. The relative few who were admitted with serious health problems were left to die alone with no loved ones allowed to see them because of 'Covid' rules and they included kids dying without the comfort of mum and dad at their bedside while the evil behind this couldn't give a damn. It was all good fun to them. A Scottish NHS staff nurse publicly quit in the spring of 2021 saying: 'I can no longer be part of the lies and the corruption by the government.' She said hospitals 'aren't full, the beds aren't full, beds have been shut, wards have been shut'. Hospitals were never busy throughout 'Covid'. The staff nurse said that Nicola Sturgeon, tragically the leader of the Scottish government, was on television saying save the hospitals and the NHS – 'but the beds are empty' and 'we've not seen flu, we always see flu every year'. She wrote to government and spoke with her union Unison (the unions are Cult-compromised and *useless*, but nothing changed. Many of her colleagues were scared of losing their jobs if they spoke out as they wanted to. She said nursing staff were being affected by wearing masks all day and 'my head is splitting every shift from wearing a mask'. The NHS is part of the fascist tyranny and must be dismantled so we can start again with human beings in charge. (Ironically, hospitals were reported to be busier again when official 'Covid' cases *fell* in spring/summer of 2021 and many other conditions required treatment at the same time as *the fake vaccine rollout*.)

I will cover the 'Covid vaccine' scam in detail later, but it is another indicator of the sickening disregard for human life that I am highlighting here. The DNA-manipulating concoctions do not fulfil

the definition of a 'vaccine', have never been used on humans before and were given only emergency approval because trials were not completed and they continued using the unknowing public. The result was what a NHS senior nurse with responsibility for 'vaccine' procedure said was 'genocide'. She said the 'vaccines' were not 'vaccines'. They had not been shown to be safe and claims about their effectiveness by drug companies were 'poetic licence'. She described what was happening as a 'horrid act of human annihilation'. The nurse said that management had instigated a policy of not providing a Patient Information Leaflet (PIL) before people were 'vaccinated' even though health care professionals are supposed to do this according to protocol. Patients should also be told that they are taking part in an ongoing clinical trial. Her challenges to what is happening had seen her excluded from meetings and ridiculed in others. She said she was told to 'watch my step ... or I would find myself surplus to requirements'. The nurse, who spoke anonymously in fear of her career, said she asked her NHS manager why he/she was content with taking part in genocide against those having the 'vaccines'. The reply was that everyone had to play their part and to 'put up, shut up, and get it done'. Government was 'leaning heavily' on NHS management which was clearly leaning heavily on staff. This is how the global 'medical' hierarchy operates and it starts with the Cult and its World Health Organization.

She told the story of a doctor who had the Pfizer jab and when questioned had no idea what was in it. The doctor had never read the literature. We have to stop treating doctors as intellectual giants when so many are moral and medical pygmies. The doctor did not even know that the 'vaccines' were not fully approved or that their trials were ongoing. They were, however, asking their patients if they minded taking part in follow-ups for research purposes – yes, the *ongoing clinical trial*. The nurse said the doctor's ignorance was not rare and she had spoken to a hospital consultant who had the jab without any idea of the background or that the 'trials' had not been completed. Nurses and pharmacists had shown the same ignorance.

‘My NHS colleagues have forsaken their duty of care, broken their code of conduct – Hippocratic Oath – and have been brainwashed just the same as the majority of the UK public through propaganda ...’ She said she had not been able to recruit a single NHS colleague, doctor, nurse or pharmacist to stand with her and speak out. Her union had refused to help. She said that if the genocide came to light she would not hesitate to give evidence at a Nuremberg-type trial against those in power who could have affected the outcomes but didn’t.

And all for what?

To put the nonsense into perspective let’s say the ‘virus’ does exist and let’s go completely crazy and accept that the official manipulated figures for cases and deaths are accurate. *Even then* a study by Stanford University epidemiologist Dr John Ioannidis published on the World Health Organization website produced an average infection to fatality rate of ... *0.23 percent!* Ioannidis said: ‘If one could sample equally from all locations globally, the median infection fatality rate might even be substantially lower than the 0.23% observed in my analysis.’ For healthy people under 70 it was ... *0.05 percent!* This compares with the 3.4 percent claimed by the Cult-owned World Health Organization when the hoax was first played and maximum fear needed to be generated. An updated Stanford study in April, 2021, put the ‘infection’ to ‘fatality’ rate at just 0.15 percent. Another team of scientists led by Megan O’Driscoll and Henrik Salje studied data from 45 countries and published their findings on the Nature website. For children and young people the figure is so small it virtually does not register although authorities will be hyping dangers to the young when they introduce DNA-manipulating ‘vaccines’ for children. The O’Driscoll study produced an average infection-fatality figure of 0.003 for children from birth to four; 0.001 for 5 to 14; 0.003 for 15 to 19; and it was still only 0.456 up to 64. To claim that children must be ‘vaccinated’ to protect them from ‘Covid’ is an obvious lie and so there must be another reason and there is. What’s more the average age of a ‘Covid’ death is akin

to the average age that people die in general. The average age of death in England is about 80 for men and 83 for women. The average age of death from alleged 'Covid' is between 82 and 83. California doctors, Dan Erickson and Artin Massihi, said at their April media conference that projection models of millions of deaths had been 'woefully inaccurate'. They produced detailed figures showing that Californians had a 0.03 chance of dying from 'Covid' based on the number of people who tested positive (with a test not testing for the 'virus'). Erickson said there was a 0.1 percent chance of dying from 'Covid' in the *state* of New York, not just the city, and a 0.05 percent chance in Spain, a centre of 'Covid-19' hysteria at one stage. The Stanford studies supported the doctors' data with fatality rate estimates of 0.23 and 0.15 percent. How close are these figures to my estimate of *zero*? Death-rate figures claimed by the World Health Organization at the start of the hoax were some 15 times higher. The California doctors said there was no justification for lockdowns and the economic devastation they caused. Everything they had ever learned about quarantine was that you quarantine the *sick* and not the healthy. They had never seen this before and it made no medical sense.

Why in the in the light of all this would governments and medical systems the world over say that billions must go under house arrest; lose their livelihood; in many cases lose their mind, their health and their life; force people to wear masks dangerous to health and psychology; make human interaction and even family interaction a criminal offence; ban travel; close restaurants, bars, watching live sport, concerts, theatre, and any activity involving human togetherness and discourse; and closing schools to isolate children from their friends and cause many to commit suicide in acts of hopelessness and despair? The California doctors said lockdown consequences included increased child abuse, partner abuse, alcoholism, depression, and other impacts they were seeing every day. Who would do that to the entire human race if not mentally-ill psychopaths of almost unimaginable extremes like Bill Gates? We must face the reality of what we are dealing with and come out of

denial. Fascism and tyranny are made possible only by the target population submitting and acquiescing to fascism and tyranny. The whole of human history shows that to be true. Most people naively and unquestioning believed what they were told about a 'deadly virus' and meekly and weakly submitted to house arrest. Those who didn't believe it – at least in total – still submitted in fear of the consequences of not doing so. For the rest who wouldn't submit draconian fines have been imposed, brutal policing by psychopaths *for* psychopaths, and condemnation from the meek and weak who condemn the Pushbackers on behalf of the very force that has them, too, in its gun sights. 'Pathetic' does not even begin to suffice. Britain's brainless 'Health' Secretary Matt Hancock warned anyone lying to border officials about returning from a list of 'hotspot' countries could face a jail sentence of up to ten years which is more than for racially-aggravated assault, incest and attempting to have sex with a child under 13. Hancock is a lunatic, but he has the state apparatus behind him in a Cult-led chain reaction and the same with UK 'Vaccine Minister' Nadhim Zahawi, a prominent member of the mega-Cult secret society, Le Cercle, which featured in my earlier books. The Cult enforces its will on governments and medical systems; government and medical systems enforce their will on business and police; business enforces its will on staff who enforce it on customers; police enforce the will of the Cult on the population and play their essential part in creating a world of fascist control that their own children and grandchildren will have to live in their entire lives. It is a hierarchical pyramid of imposition and acquiescence and, yes indeed, of clinical insanity.

Does anyone bright enough to read this book have to ask what the answer is? I think not, but I will reveal it anyway in the fewest of syllables: Tell the psychos and their moronic lackeys to fuck off and let's get on with our lives. We are many – They are few.

CHAPTER SEVEN

War on your mind

One believes things because one has been conditioned to believe them

Aldous Huxley, Brave New World

I have described the 'Covid' hoax as a 'Psyop' and that is true in every sense and on every level in accordance with the definition of that term which is psychological warfare. Break down the 'Covid pandemic' to the foundation themes and it is psychological warfare on the human individual and collective mind.

The same can be said for the entire human belief system involving every subject you can imagine. Huxley was right in his contention that people believe what they are conditioned to believe and this comes from the repetition throughout their lives of the same falsehoods. They spew from government, corporations, media and endless streams of 'experts' telling you what the Cult wants you to believe and often believing it themselves (although *far* from always). 'Experts' are rewarded with 'prestigious' jobs and titles and as agents of perceptual programming with regular access to the media. The Cult has to control the narrative – control *information* – or they lose control of the vital, crucial, without-which-they-cannot-prevail public perception of reality. The foundation of that control today is the Internet made possible by the Defense Advanced Research Projects Agency (DARPA), the incredibly sinister technological arm of the Pentagon. The Internet is the result of military technology.

DARPA openly brags about establishing the Internet which has been a long-term project to lasso the minds of the global population. I have said for decades the plan is to control information to such an extreme that eventually no one would see or hear anything that the Cult does not approve. We are closing in on that end with ferocious censorship since the 'Covid' hoax began and in my case it started back in the 1990s in terms of books and speaking venues. I had to create my own publishing company in 1995 precisely because no one else would publish my books even then. I think they're all still running.

Cult Internet

To secure total control of information they needed the Internet in which pre-programmed algorithms can seek out 'unclean' content for deletion and even stop it being posted in the first place. The Cult had to dismantle print and non-Internet broadcast media to ensure the transfer of information to the appropriate-named 'Web' – a critical expression of the *Cult* web. We've seen the ever-quicken demise of traditional media and control of what is left by a tiny number of corporations operating worldwide. Independent journalism in the mainstream is already dead and never was that more obvious than since the turn of 2020. The Cult wants all information communicated via the Internet to globally censor and allow the plug to be pulled any time. Lockdowns and forced isolation has meant that communication between people has been through electronic means and no longer through face-to-face discourse and discussion. Cult psychopaths have targeted the bars, restaurants, sport, venues and meeting places in general for this reason. None of this is by chance and it's to stop people gathering in any kind of privacy or number while being able to track and monitor all Internet communications and block them as necessary. Even private messages between individuals have been censored by these fascists that control Cult fronts like Facebook, Twitter, Google and YouTube which are all officially run by Sabbatian place-people and from the background by higher-level Sabbatian place people.

Facebook, Google, Amazon and their like were seed-funded and supported into existence with money-no-object infusions of funds either directly or indirectly from DARPA and CIA technology arm In-Q-Tel. The Cult plays the long game and prepares very carefully for big plays like 'Covid'. Amazon is another front in the psychological war and pretty much controls the global market in book sales and increasingly publishing. Amazon's limitless funds have deleted fantastic numbers of independent publishers to seize global domination on the way to deciding which books can be sold and circulated and which cannot. Moves in that direction are already happening. Amazon's leading light Jeff Bezos is the grandson of Lawrence Preston Gise who worked with DARPA predecessor ARPA. Amazon has big connections to the CIA and the Pentagon. The plan I have long described went like this:

1. Employ military technology to establish the Internet.
2. Sell the Internet as a place where people can freely communicate without censorship and allow that to happen until the Net becomes the central and irreversible pillar of human society. If the Internet had been highly censored from the start many would have rejected it.
3. Fund and manipulate major corporations into being to control the circulation of information on your Internet using cover stories about geeks in garages to explain how they came about. Give them unlimited funds to expand rapidly with no need to make a profit for years while non-Cult companies who need to balance the books cannot compete. You know that in these circumstances your Googles, YouTubes, Facebooks and Amazons are going to secure near monopolies by either crushing or buying up the opposition.
4. Allow freedom of expression on both the Internet and communication platforms to draw people in until the Internet is the central and irreversible pillar of human society and your communication corporations have reached a stage of near monopoly domination.
5. Then unleash your always-planned frenzy of censorship on the basis of 'where else are you going to go?' and continue to expand that until nothing remains that the Cult does not want its human targets to see.

The process was timed to hit the 'Covid' hoax to ensure the best chance possible of controlling the narrative which they knew they had to do at all costs. They were, after all, about to unleash a 'deadly virus' that didn't really exist. If you do that in an environment of free-flowing information and opinion you would be dead in the

water before you could say Gates is a psychopath. The network was in place through which the Cult-created-and-owned World Health Organization could dictate the 'Covid' narrative and response policy slavishly supported by Cult-owned Internet communication giants and mainstream media while those telling a different story were censored. Google, YouTube, Facebook and Twitter openly announced that they would do this. What else would we expect from Cult-owned operations like Facebook which former executives have confirmed set out to make the platform more addictive than cigarettes and coldly manipulates emotions of its users to sow division between people and groups and scramble the minds of the young? If Zuckerberg lives out the rest of his life without going to jail for crimes against humanity, and most emphatically against the young, it will be a travesty of justice. Still, no matter, cause and effect will catch up with him eventually and the same with Sergey Brin and Larry Page at Google with its CEO Sundar Pichai who fix the Google search results to promote Cult narratives and hide the opposition. Put the same key words into Google and other search engines like DuckDuckGo and you will see how different results can be. Wikipedia is another intensely biased 'encyclopaedia' which skews its content to the Cult agenda. YouTube links to Wikipedia's version of 'Covid' and 'climate change' on video pages in which experts in their field offer a different opinion (even that is increasingly rare with Wojcicki censorship). Into this 'Covid' silence-them network must be added government media censors, sorry 'regulators', such as Ofcom in the UK which imposed tyrannical restrictions on British broadcasters that had the effect of banning me from ever appearing. Just to debate with me about my evidence and views on 'Covid' would mean breaking the fascistic impositions of Ofcom and its CEO career government bureaucrat Melanie Dawes. Gutless British broadcasters tremble at the very thought of fascist Ofcom.

Psychos behind 'Covid'

The reason for the 'Covid' catastrophe in all its facets and forms can be seen by whom and what is driving the policies worldwide in such a coordinated way. Decisions are not being made to protect health, but to target psychology. The dominant group guiding and 'advising' government policy are not medical professionals. They are psychologists and behavioural scientists. Every major country has its own version of this phenomenon and I'll use the British example to show how it works. In many ways the British version has been affecting the wider world in the form of the huge behaviour manipulation network in the UK which operates in other countries. The network involves private companies, government, intelligence and military. The Cabinet Office is at the centre of the government 'Covid' Psyop and part-owns, with 'innovation charity' Nesta, the Behavioural Insights Team (BIT) which claims to be independent of government but patently isn't. The BIT was established in 2010 and its job is to manipulate the psyche of the population to acquiesce to government demands and so much more. It is also known as the 'Nudge Unit', a name inspired by the 2009 book by two ultra-Zionists, Cass Sunstein and Richard Thaler, called *Nudge: Improving Decisions About Health, Wealth, and Happiness*. The book, as with the Behavioural Insights Team, seeks to 'nudge' behaviour (manipulate it) to make the public follow patterns of action and perception that suit those in authority (the Cult). Sunstein is so skilled at this that he advises the World Health Organization and the UK Behavioural Insights Team and was Administrator of the White House Office of Information and Regulatory Affairs in the Obama administration. Biden appointed him to the Department of Homeland Security – another ultra-Zionist in the fold to oversee new immigration laws which is another policy the Cult wants to control. Sunstein is desperate to silence anyone exposing conspiracies and co-authored a 2008 report on the subject in which suggestions were offered to ban 'conspiracy theorizing' or impose 'some kind of tax, financial or otherwise, on those who disseminate such theories'. I guess a psychiatrist's chair is out of the question?

Sunstein's mate Richard Thaler, an 'academic affiliate' of the UK Behavioural Insights Team, is a proponent of 'behavioural economics' which is defined as the study of 'the effects of psychological, cognitive, emotional, cultural and social factors on the decisions of individuals and institutions'. Study the effects so they can be manipulated to be what you want them to be. Other leading names in the development of behavioural economics are ultra-Zionists Daniel Kahneman and Robert J. Shiller and they, with Thaler, won the Nobel Memorial Prize in Economic Sciences for their work in this field. The Behavioural Insights Team is operating at the heart of the UK government and has expanded globally through partnerships with several universities including Harvard, Oxford, Cambridge, University College London (UCL) and Pennsylvania. They claim to have 'trained' (reframed) 20,000 civil servants and run more than 750 projects involving 400 randomised controlled trials in dozens of countries' as another version of mind reframers Common Purpose. BIT works from its office in New York with cities and their agencies, as well as other partners, across the United States and Canada – this is a company part-owned by the British government Cabinet Office. An executive order by President Cult-servant Obama established a US Social and Behavioral Sciences Team in 2015. They all have the same reason for being and that's to brainwash the population directly and by brainwashing those in positions of authority.

'Covid' mind game

Another prime aspect of the UK mind-control network is the 'independent' [joke] Scientific Pandemic Insights Group on Behaviours (SPI-B) which 'provides behavioural science advice aimed at anticipating and helping people adhere to interventions that are recommended by medical or epidemiological experts'. That means manipulating public perception and behaviour to do whatever government tells them to do. It's disgusting and if they really want the public to be 'safe' this lot should all be under lock and key. According to the government website SPI-B consists of

'behavioural scientists, health and social psychologists, anthropologists and historians' and advises the Whitty-Vallance-led Scientific Advisory Group for Emergencies (SAGE) which in turn advises the government on 'the science' (it doesn't) and 'Covid' policy. When politicians say they are being guided by 'the science' this is the rabble in each country they are talking about and that 'science' is dominated by behaviour manipulators to enforce government fascism through public compliance. The Behaviour Insight Team is headed by psychologist David Solomon Halpern, a visiting professor at King's College London, and connects with a national and global web of other civilian and military organisations as the Cult moves towards its goal of fusing them into one fascistic whole in every country through its 'Fusion Doctrine'. The behaviour manipulation network involves, but is not confined to, the Foreign Office; National Security Council; government communications headquarters (GCHQ); MI5; MI6; the Cabinet Office-based Media Monitoring Unit; and the Rapid Response Unit which 'monitors digital trends to spot emerging issues; including misinformation and disinformation; and identifies the best way to respond'.

There is also the 77th Brigade of the UK military which operates like the notorious Israeli military's Unit 8200 in manipulating information and discussion on the Internet by posing as members of the public to promote the narrative and discredit those who challenge it. Here we have the military seeking to manipulate *domestic* public opinion while the Nazis in government are fine with that. Conservative Member of Parliament Tobias Ellwood, an advocate of lockdown and control through 'vaccine passports', is a Lieutenant Colonel reservist in the 77th Brigade which connects with the military operation jHub, the 'innovation centre' for the Ministry of Defence and Strategic Command. jHub has also been involved with the civilian National Health Service (NHS) in 'symptom tracing' the population. The NHS is a key part of this mind control network and produced a document in December, 2020, explaining to staff how to use psychological manipulation with different groups and ages to get them to have the DNA-manipulating 'Covid vaccine'

that's designed to cumulatively rewrite human genetics. The document, called 'Optimising Vaccination Roll Out – Do's and Don'ts for all messaging, documents and "communications" in the widest sense', was published by NHS England and the NHS Improvement *Behaviour Change Unit* in partnership with Public Health England and Warwick Business School. I hear the mantra about 'save the NHS' and 'protect the NHS' when we need to scrap the NHS and start again. The current version is far too corrupt, far too anti-human and totally compromised by Cult operatives and their assets. UK government broadcast media censor Ofcom will connect into this web – as will the BBC with its tremendous Ofcom influence – to control what the public see and hear and dictate mass perception. Nuremberg trials must include personnel from all these organisations.

The fear factor

The 'Covid' hoax has led to the creation of the UK Cabinet Office-connected Joint Biosecurity Centre (JBC) which is officially described as providing 'expert advice on pandemics' using its independent [all Cult operations are 'independent'] analytical function to provide real-time analysis about infection outbreaks to identify and respond to outbreaks of Covid-19'. Another role is to advise the government on a response to spikes in infections – 'for example by closing schools or workplaces in local areas where infection levels have risen'. Put another way, promoting the Cult agenda. The Joint Biosecurity Centre is modelled on the Joint Terrorism Analysis Centre which analyses intelligence to set 'terrorism threat levels' and here again you see the fusion of civilian and military operations and intelligence that has led to military intelligence producing documents about 'vaccine hesitancy' and how it can be combated. Domestic civilian matters and opinions should not be the business of the military. The Joint Biosecurity Centre is headed by Tom Hurd, director general of the Office for Security and Counter-Terrorism from the establishment-to-its-fingertips Hurd family. His father is former Foreign Secretary Douglas Hurd. How coincidental that Tom

Hurd went to the elite Eton College and Oxford University with Boris Johnson. Imperial College with its ridiculous computer modeller Neil Ferguson will connect with this gigantic web that will itself interconnect with similar set-ups in other major and not so major countries. Compared with this Cult network the politicians, be they Boris Johnson, Donald Trump or Joe Biden, are bit-part players 'following the science'. The network of psychologists was on the 'Covid' case from the start with the aim of generating maximum fear of the 'virus' to ensure compliance by the population. A government behavioural science group known as SPI-B produced a paper in March, 2020, for discussion by the main government science advisory group known as SAGE. It was headed 'Options for increasing adherence to social distancing measures' and it said the following in a section headed 'Persuasion':

- A substantial number of people still do not feel sufficiently personally threatened; it could be that they are reassured by the low death rate in their demographic group, although levels of concern may be rising. Having a good understanding of the risk has been found to be positively associated with adoption of COVID-19 social distancing measures in Hong Kong.
- The perceived level of personal threat needs to be increased among those who are complacent, using hard-hitting evaluation of options for increasing social distancing emotional messaging. To be effective this must also empower people by making clear the actions they can take to reduce the threat.
- Responsibility to others: There seems to be insufficient understanding of, or feelings of responsibility about, people's role in transmitting the infection to others ... Messaging about actions need to be framed positively in terms of protecting oneself and the community, and increase confidence that they will be effective.
- Some people will be more persuaded by appeals to play by the rules, some by duty to the community, and some to personal risk.

All these different approaches are needed. The messaging also needs to take account of the realities of different people's lives. Messaging needs to take account of the different motivational levers and circumstances of different people.

All this could be achieved the SPI-B psychologists said by *using the media to increase the sense of personal threat* which translates as terrify the shit out of the population, including children, so they all do what we want. That's not happened has it? Those excuses for 'journalists' who wouldn't know journalism if it bit them on the arse (the great majority) have played their crucial part in serving this Cult-government Psyop to enslave their own kids and grandkids. How they live with themselves I have no idea. The psychological war has been underpinned by constant government 'Covid' propaganda in almost every television and radio ad break, plus the Internet and print media, which has pounded out the fear with taxpayers footing the bill for their own programming. The result has been people terrified of a 'virus' that doesn't exist or one with a tiny fatality rate even if you believe it does. People walk down the street and around the shops wearing face-nappies damaging their health and psychology while others report those who refuse to be that naïve to the police who turn up in their own face-nappies. I had a cameraman come to my flat and he was so frightened of 'Covid' he came in wearing a mask and refused to shake my hand in case he caught something. He had – naïveitis – and the thought that he worked in the mainstream media was both depressing and made his behaviour perfectly explainable. The fear which has gripped the minds of so many and frozen them into compliance has been carefully cultivated by these psychologists who are really psychopaths. If lives get destroyed and a lot of young people commit suicide it shows our plan is working. SPI-B then turned to compulsion on the public to comply. 'With adequate preparation, rapid change can be achieved', it said. Some countries had introduced mandatory self-isolation on a wide scale without evidence of major public unrest and a large majority of the UK's population appeared to be supportive of more coercive measures with 64 percent of adults saying they would

support putting London under a lockdown (watch the 'polls' which are designed to make people believe that public opinion is in favour or against whatever the subject in hand).

For 'aggressive protective measures' to be effective, the SPI-B paper said, special attention should be devoted to those population groups that are more at risk. Translated from the Orwellian this means making the rest of population feel guilty for not protecting the 'vulnerable' such as old people which the Cult and its agencies were about to kill on an industrial scale with lockdown, lack of treatment and the Gates 'vaccine'. Psychopath psychologists sold their guilt-trip so comprehensively that Los Angeles County Supervisor Hilda Solis reported that children were apologising (from a distance) to their parents and grandparents for bringing 'Covid' into their homes and getting them sick. '... These apologies are just some of the last words that loved ones will ever hear as they die alone,' she said. Gut-wrenchingly Solis then used this childhood tragedy to tell children to stay at home and 'keep your loved ones alive'. Imagine heaping such potentially life-long guilt on a kid when it has absolutely nothing to do with them. These people are deeply disturbed and the psychologists behind this even more so.

Uncivil war – divide and rule

Professional mind-controllers at SPI-B wanted the media to increase a sense of responsibility to others (do as you're told) and promote 'positive messaging' for those actions while in contrast to invoke 'social disapproval' by the unquestioning, obedient, community of anyone with a mind of their own. Again the compliant Goebbels-like media obliged. This is an old, old, trick employed by tyrannies the world over throughout human history. You get the target population to keep the target population in line – *your* line. SPI-B said this could 'play an important role in preventing anti-social behaviour or discouraging failure to enact pro-social behaviour'. For 'anti-social' in the Orwellian parlance of SPI-B see any behaviour that government doesn't approve. SPI-B recommendations said that 'social disapproval' should be accompanied by clear messaging and

promotion of strong collective identity – hence the government and celebrity mantra of ‘we’re all in this together’. Sure we are. The mind doctors have such contempt for their targets that they think some clueless comedian, actor or singer telling them to do what the government wants will be enough to win them over. We have had UK comedian Lenny Henry, actor Michael Caine and singer Elton John wheeled out to serve the propagandists by urging people to have the DNA-manipulating ‘Covid’ non-‘vaccine’. The role of Henry and fellow black celebrities in seeking to coax a ‘vaccine’ reluctant black community into doing the government’s will was especially stomach-turning. An emotion-manipulating script and carefully edited video featuring these black ‘celebs’ was such an insult to the intelligence of black people and where’s the self-respect of those involved selling their souls to a fascist government agenda? Henry said he heard black people’s ‘legitimate worries and concerns’, but people must ‘trust the facts’ when they were doing exactly that by not having the ‘vaccine’. They had to include the obligatory reference to Black Lives Matter with the line ... ‘Don’t let coronavirus cost even more black lives – because we matter’. My god, it was pathetic. ‘I know the vaccine is safe and what it does.’ How? ‘I’m a comedian and it says so in my script.’

SPI-B said social disapproval needed to be carefully managed to avoid victimisation, scapegoating and misdirected criticism, but they knew that their ‘recommendations’ would lead to exactly that and the media were specifically used to stir-up the divide-and-conquer hostility. Those who conform like good little baa, baas, are praised while those who have seen through the tidal wave of lies are ‘Covidiot’. The awake have been abused by the fast asleep for not conforming to fascism and impositions that the awake know are designed to endanger their health, dehumanise them, and tear asunder the very fabric of human society. We have had the curtain-twitchers and morons reporting neighbours and others to the face-napped police for breaking ‘Covid rules’ with fascist police delighting in posting links and phone numbers where this could be done. The Cult cannot impose its will without a compliant police

and military or a compliant population willing to play their part in enslaving themselves and their kids. The words of a pastor in Nazi Germany are so appropriate today:

First they came for the socialists and I did not speak out because I was not a socialist.

Then they came for the trade unionists and I did not speak out because I was not a trade unionist.

Then they came for the Jews and I did not speak out because I was not a Jew.

Then they came for me and there was no one left to speak for me.

Those who don't learn from history are destined to repeat it and so many are.

'Covid' rules: Rewiring the mind

With the background laid out to this gigantic national and global web of psychological manipulation we can put 'Covid' rules into a clear and sinister perspective. Forget the claims about protecting health. 'Covid' rules are about dismantling the human mind, breaking the human spirit, destroying self-respect, and then putting Humpty Dumpty together again as a servile, submissive slave. Social isolation through lockdown and distancing have devastating effects on the human psyche as the psychological psychopaths well know and that's the real reason for them. Humans need contact with each other, discourse, closeness and touch, or they eventually, and literally, go crazy. Masks, which I will address at some length, fundamentally add to the effects of isolation and the Cult agenda to dehumanise and de-individualise the population. To do this while knowing – in fact *seeking* – this outcome is the very epitome of evil and psychologists involved in this *are* the epitome of evil. They must like all the rest of the Cult demons and their assets stand trial for crimes against humanity on a scale that defies the imagination. Psychopaths in uniform use isolation to break enemy troops and agents and make them subservient and submissive to tell what they know. The technique is rightly considered a form of torture and

torture is most certainly what has been imposed on the human population.

Clinically-insane American psychologist Harry Harlow became famous for his isolation experiments in the 1950s in which he separated baby monkeys from their mothers and imprisoned them for months on end in a metal container or 'pit of despair'. They soon began to show mental distress and depression as any idiot could have predicted. Harlow put other monkeys in steel chambers for three, six or twelve months while denying them any contact with animals or humans. He said that the effects of total social isolation for six months were 'so devastating and debilitating that we had assumed initially that twelve months of isolation would not produce any additional decrement'; but twelve months of isolation 'almost obliterated the animals socially'. This is what the Cult and its psychopaths are doing to you and your children. Even monkeys in partial isolation in which they were not allowed to form relationships with other monkeys became 'aggressive and hostile, not only to others, but also towards their own bodies'. We have seen this in the young as a consequence of lockdown. UK government psychopaths launched a public relations campaign telling people not to hug each other even after they received the 'Covid-19 vaccine' which we were told with more lies would allow a return to 'normal life'. A government source told *The Telegraph*: 'It will be along the lines that it is great that you have been vaccinated, but if you are going to visit your family and hug your grandchildren there is a chance you are going to infect people you love.' The source was apparently speaking from a secure psychiatric facility. Janet Lord, director of Birmingham University's Institute of Inflammation and Ageing, said that parents and grandparents should avoid hugging their children. Well, how can I put it, Ms Lord? Fuck off. Yep, that'll do.

Destroying the kids – where are the parents?

Observe what has happened to people enslaved and isolated by lockdown as suicide and self-harm has soared worldwide,

particularly among the young denied the freedom to associate with their friends. A study of 49,000 people in English-speaking countries concluded that almost half of young adults are at clinical risk of mental health disorders. A national survey in America of 1,000 currently enrolled high school and college students found that 5 percent reported attempting suicide during the pandemic. Data from the US CDC's National Syndromic Surveillance Program from January 1st to October 17th, 2020, revealed a *31 percent* increase in mental health issues among adolescents aged 12 to 17 compared with 2019. The CDC reported that America in general suffered the biggest drop in life expectancy since World War Two as it fell by a year in the first half of 2020 as a result of 'deaths of despair' – overdoses and suicides. Deaths of despair have leapt by more than 20 percent during lockdown and include the highest number of fatal overdoses ever recorded in a single year – 81,000. Internet addiction is another consequence of being isolated at home which lowers interest in physical activities as kids fall into inertia and what's the point? Children and young people are losing hope and giving up on life, sometimes literally. A 14-year-old boy killed himself in Maryland because he had 'given up' when his school district didn't reopen; an 11-year-old boy shot himself during a zoom class; a teenager in Maine succumbed to the isolation of the 'pandemic' when he ended his life after experiencing a disrupted senior year at school. Children as young as nine have taken their life and all these stories can be repeated around the world. Careers are being destroyed before they start and that includes those in sport in which promising youngsters have not been able to take part. The plan of the psycho-psychologists is working all right. Researchers at Cambridge University found that lockdowns cause significant harm to children's mental health. Their study was published in the *Archives of Disease in Childhood*, and followed 168 children aged between 7 and 11. The researchers concluded:

During the UK lockdown, children's depression symptoms have increased substantially, relative to before lockdown. The scale of this effect has direct relevance for the continuation of different elements of lockdown policy, such as complete or partial school closures ...

... Specifically, we observed a statistically significant increase in ratings of depression, with a medium-to-large effect size. Our findings emphasise the need to incorporate the potential impact of lockdown on child mental health in planning the ongoing response to the global pandemic and the recovery from it.

Not a chance when the Cult's psycho-psychologists were getting exactly what they wanted. The UK's Royal College of Paediatrics and Child Health has urged parents to look for signs of eating disorders in children and young people after a three to four fold increase. Specialists say the 'pandemic' is a major reason behind the rise. You don't say. The College said isolation from friends during school closures, exam cancellations, loss of extra-curricular activities like sport, and an increased use of social media were all contributory factors along with fears about the virus (psycho-psychologists again), family finances, and students being forced to quarantine. Doctors said young people were becoming severely ill by the time they were seen with 'Covid' regulations reducing face-to-face consultations. Nor is it only the young that have been devastated by the psychopaths. Like all bullies and cowards the Cult is targeting the young, elderly, weak and infirm. A typical story was told by a British lady called Lynn Parker who was not allowed to visit her husband in 2020 for the last ten and half months of his life 'when he needed me most' between March 20th and when he died on December 19th. This vacates the criminal and enters the territory of evil. The emotional impact on the immune system alone is immense as are the number of people of all ages worldwide who have died as a result of Cult-demanded, Gates-demanded, lockdowns.

Isolation is torture

The experience of imposing solitary confinement on millions of prisoners around the world has shown how a large percentage become 'actively psychotic and/or acutely suicidal'. Social isolation has been found to trigger 'a specific psychiatric syndrome, characterized by hallucinations; panic attacks; overt paranoia; diminished impulse control; hypersensitivity to external stimuli; and difficulties with thinking, concentration and memory'. Juan Mendez,

a United Nations rapporteur (investigator), said that isolation is a form of torture. Research has shown that even after isolation prisoners find it far more difficult to make social connections and I remember chatting to a shop assistant after one lockdown who told me that when her young son met another child again he had no idea how to act or what to do. Hannah Flanagan, Director of Emergency Services at Journey Mental Health Center in Dane County, Wisconsin, said: 'The specificity about Covid social distancing and isolation that we've come across as contributing factors to the suicides are really new to us this year.' But they are not new to those that devised them. They are getting the effect they want as the population is psychologically dismantled to be rebuilt in a totally different way. Children and the young are particularly targeted. They will be the adults when the full-on fascist AI-controlled technocracy is planned to be imposed and they are being prepared to meekly submit. At the same time older people who still have a memory of what life was like before – and how fascist the new normal really is – are being deleted. You are going to see efforts to turn the young against the old to support this geriatric genocide. Hannah Flanagan said the big increase in suicide in her county proved that social isolation is not only harmful, but deadly. Studies have shown that isolation from others is one of the main risk factors in suicide and even more so with women. Warnings that lockdown could create a 'perfect storm' for suicide were ignored. After all this was one of the *reasons* for lockdown. Suicide, however, is only the most extreme of isolation consequences. There are many others. Dr Dhruv Khullar, assistant professor of healthcare policy at Weill Cornell Medical College, said in a *New York Times* article in 2016 long before the fake 'pandemic':

A wave of new research suggests social separation is bad for us. Individuals with less social connection have disrupted sleep patterns, altered immune systems, more inflammation and higher levels of stress hormones. One recent study found that isolation increases the risk of heart disease by 29 percent and stroke by 32 percent. Another analysis that pooled data from 70 studies and 3.4 million people found that socially isolated individuals had a 30 percent higher risk of dying in the next seven years, and that this effect was largest in middle age.

Loneliness can accelerate cognitive decline in older adults, and isolated individuals are twice as likely to die prematurely as those with more robust social interactions. These effects start early: Socially isolated children have significantly poorer health 20 years later, even after controlling for other factors. All told, loneliness is as important a risk factor for early death as obesity and smoking.

There you have proof from that one article alone four years before 2020 that those who have enforced lockdown, social distancing and isolation knew what the effect would be and that is even more so with professional psychologists that have been driving the policy across the globe. We can go back even further to the years 2000 and 2003 and the start of a major study on the effects of isolation on health by Dr Janine Gronewold and Professor Dirk M. Hermann at the University Hospital in Essen, Germany, who analysed data on 4,316 people with an average age of 59 who were recruited for the long-term research project. They found that socially isolated people are more than 40 percent more likely to have a heart attack, stroke, or other major cardiovascular event and nearly 50 percent more likely to die from any cause. Given the financial Armageddon unleashed by lockdown we should note that the study found a relationship between increased cardiovascular risk and lack of financial support. After excluding other factors social isolation was still connected to a 44 percent increased risk of cardiovascular problems and a 47 percent increased risk of death by any cause. Lack of financial support was associated with a 30 percent increase in the risk of cardiovascular health events. Dr Gronewold said it had been known for some time that feeling lonely or lacking contact with close friends and family can have an impact on physical health and the study had shown that having strong social relationships is of high importance for heart health. Gronewold said they didn't understand yet why people who are socially isolated have such poor health outcomes, but this was obviously a worrying finding, particularly during these times of prolonged social distancing. Well, it can be explained on many levels. You only have to identify the point in the body where people feel loneliness and missing people they are parted from – it's in the centre of the chest where they feel the ache of loneliness and the ache of missing people. 'My heart aches for

you' ... 'My heart aches for some company.' I will explain this more in the chapter Escaping Wetiko, but when you realise that the body is the mind – they are expressions of each other – the reason why state of the mind dictates state of the body becomes clear.

American psychologist Ranjit Powar was highlighting the effects of lockdown isolation as early as April, 2020. She said humans have evolved to be social creatures and are wired to live in interactive groups. Being isolated from family, friends and colleagues could be unbalancing and traumatic for most people and could result in short or even long-term psychological and physical health problems. An increase in levels of anxiety, aggression, depression, forgetfulness and hallucinations were possible psychological effects of isolation. 'Mental conditions may be precipitated for those with underlying pre-existing susceptibilities and show up in many others without any pre-condition.' Powar said personal relationships helped us cope with stress and if we lost this outlet for letting off steam the result can be a big emotional void which, for an average person, was difficult to deal with. 'Just a few days of isolation can cause increased levels of anxiety and depression' – so what the hell has been the effect on the global population of *18 months* of this at the time of writing? Powar said: 'Add to it the looming threat of a dreadful disease being repeatedly hammered in through the media and you have a recipe for many shades of mental and physical distress.' For those with a house and a garden it is easy to forget that billions have had to endure lockdown isolation in tiny overcrowded flats and apartments with nowhere to go outside. The psychological and physical consequences of this are unimaginable and with lunatic and abusive partners and parents the consequences have led to tremendous increases in domestic and child abuse and alcoholism as people seek to shut out the horror. Ranjit Powar said:

Staying in a confined space with family is not all a rosy picture for everyone. It can be extremely oppressive and claustrophobic for large low-income families huddled together in small single-room houses. Children here are not lucky enough to have many board/electronic games or books to keep them occupied.

Add to it the deep insecurity of running out of funds for food and basic necessities. On the other hand, there are people with dysfunctional family dynamics, such as domineering, abusive or alcoholic partners, siblings or parents which makes staying home a period of trial. Incidence of suicide and physical abuse against women has shown a worldwide increase. Heightened anxiety and depression also affect a person's immune system, making them more susceptible to illness.

To think that Powar's article was published on April 11th, 2020.

Six-feet fantasy

Social (unsocial) distancing demanded that people stay six feet or two metres apart. UK government advisor Robert Dingwall from the New and Emerging Respiratory Virus Threats Advisory Group said in a radio interview that the two-metre rule was 'conjured up out of nowhere' and was not based on science. No, it was not based on *medical* science, but it didn't come out of nowhere. The distance related to *psychological* science. Six feet/two metres was adopted in many countries and we were told by people like the criminal Anthony Fauci and his ilk that it was founded on science. Many schools could not reopen because they did not have the space for six-foot distancing. Then in March, 2021, after a year of six-foot 'science', a study published in the *Journal of Infectious Diseases* involving more than 500,000 students and almost 100,000 staff over 16 weeks revealed no significant difference in 'Covid' cases between six feet and three feet and Fauci changed his tune. Now three feet was okay. There is no difference between six feet and three *inches* when there is no 'virus' and they got away with six feet for psychological reasons for as long as they could. I hear journalists and others talk about 'unintended consequences' of lockdown. They are not *unintended* at all; they have been coldly-calculated for a specific outcome of human control and that's why super-psychopaths like Gates have called for them so vehemently. Super-psychopath psychologists have demanded them and psychopathic or clueless, spineless, politicians have gone along with them by 'following the science'. But it's not science at all. 'Science' is not what is; it's only what people can be manipulated to believe it is. The whole 'Covid' catastrophe is

founded on mind control. Three word or three statement mantras issued by the UK government are a well-known mind control technique and so we've had 'Stay home/protect the NHS/save lives', 'Stay alert/control the virus/save lives' and 'hands/face/space'. One of the most vocal proponents of extreme 'Covid' rules in the UK has been Professor Susan Michie, a member of the British Communist Party, who is not a medical professional. Michie is the director of the Centre for Behaviour Change at University College London. She is a *behavioural psychologist* and another filthy rich 'Marxist' who praised China's draconian lockdown. She was known by fellow students at Oxford University as 'Stalin's nanny' for her extreme Marxism. Michie is an influential member of the UK government's Scientific Advisory Group for Emergencies (SAGE) and behavioural manipulation groups which have dominated 'Covid' policy. She is a consultant adviser to the World Health Organization on 'Covid-19' and behaviour. Why the hell are lockdowns anything to do with her when they are claimed to be about health? Why does a behavioural psychologist from a group charged with changing the behaviour of the public want lockdown, human isolation and mandatory masks? Does that question really need an answer? Michie *absolutely* has to explain herself before a Nuremberg court when humanity takes back its world again and even more so when you see the consequences of masks that she demands are compulsory. This is a Michie classic:

The benefits of getting primary school children to wear masks is that regardless of what little degree of transmission is occurring in those age groups it could help normalise the practice. Young children wearing masks may be more likely to get their families to accept masks.

Those words alone should carry a prison sentence when you ponder on the callous disregard for children involved and what a statement it makes about the mind and motivations of Susan Michie. What a lovely lady and what she said there encapsulates the mentality of the psychopaths behind the 'Covid' horror. Let us compare what Michie said with a countrywide study in Germany published at [researchsquare.com](https://www.researchsquare.com) involving 25,000 school children and 17,854 health complaints submitted by parents. Researchers

found that masks are harming children physically, psychologically, and behaviourally with 24 health issues associated with mask wearing. They include: shortness of breath (29.7%); dizziness (26.4%); increased headaches (53%); difficulty concentrating (50%); drowsiness or fatigue (37%); and malaise (42%). Nearly a third of children experienced more sleep issues than before and a quarter developed new fears. Researchers found health issues and other impairments in 68 percent of masked children covering their faces for an average of 4.5 hours a day. Hundreds of those taking part experienced accelerated respiration, tightness in the chest, weakness, and short-term impairment of consciousness. A reminder of what Michie said again:

The benefits of getting primary school children to wear masks is that regardless of what little degree of transmission is occurring in those age groups it could help normalise the practice. Young children wearing masks may be more likely to get their families to accept masks.

Psychopaths in government and psychology now have children and young people – plus all the adults – wearing masks for hours on end while clueless teachers impose the will of the psychopaths on the young they should be protecting. What the hell are parents doing?

Cult lab rats

We have some schools already imposing on students microchipped buzzers that activate when they get 'too close' to their pals in the way they do with lab rats. How apt. To the Cult and its brain-dead servants our children *are* lab rats being conditioned to be unquestioning, dehumanised slaves for the rest of their lives.

Children and young people are being weaned and frightened away from the most natural human instincts including closeness and touch. I have tracked in the books over the years how schools were banning pupils from greeting each other with a hug and the whole Cult-induced Me Too movement has terrified men and boys from a relaxed and natural interaction with female friends and work colleagues to the point where many men try never to be in a room

alone with a woman that's not their partner. Airhead celebrities have as always played their virtue-signalling part in making this happen with their gross exaggeration. For every monster like Harvey Weinstein there are at least tens of thousands of men that don't treat women like that; but everyone must be branded the same and policy changed for them as well as the monster. I am going to be using the word 'dehumanise' many times in this chapter because that is what the Cult is seeking to do and it goes very deep as we shall see. Don't let them kid you that social distancing is planned to end one day. That's not the idea. We are seeing more governments and companies funding and producing wearable gadgets to keep people apart and they would not be doing that if this was meant to be short-term. A tech start-up company backed by GCHQ, the British Intelligence and military surveillance headquarters, has created a social distancing wrist sensor that alerts people when they get too close to others. The CIA has also supported tech companies developing similar devices. The wearable sensor was developed by Tended, one of a number of start-up companies supported by GCHQ (see the CIA and DARPA). The device can be worn on the wrist or as a tag on the waistband and will vibrate whenever someone wearing the device breaches social distancing and gets anywhere near natural human contact. The company had a lucky break in that it was developing a distancing sensor when the 'Covid' hoax arrived which immediately provided a potentially enormous market. How fortunate. The government in big-time Cult-controlled Ontario in Canada is investing \$2.5 million in wearable contact tracing technology that 'will alert users if they may have been exposed to the Covid-19 in the workplace and will beep or vibrate if they are within six feet of another person'. Facedrive Inc., the technology company behind this, was founded in 2016 with funding from the Ontario Together Fund and obviously they, too, had a prophet on the board of directors. The human surveillance and control technology is called TraceSCAN and would be worn by the human cyborgs in places such as airports, workplaces, construction sites, care homes and ... *schools*.

I emphasise schools with children and young people the prime targets. You know what is planned for society as a whole if you keep your eyes on the schools. They have always been places where the state program the next generation of slaves to be its compliant worker-ants – or Woker-ants these days; but in the mist of the ‘Covid’ madness they have been transformed into mind laboratories on a scale never seen before. Teachers and head teachers are just as programmed as the kids – often more so. Children are kept apart from human interaction by walk lanes, classroom distancing, staggered meal times, masks, and the rolling-out of buzzer systems. Schools are now physically laid out as a laboratory maze for lab-rats. Lunatics at a school in Anchorage, Alaska, who should be prosecuted for child abuse, took away desks and forced children to kneel (know your place) on a mat for five hours a day while wearing a mask and using their chairs as a desk. How this was supposed to impact on a ‘virus’ only these clinically insane people can tell you and even then it would be clap-trap. The school banned recess (interaction), art classes (creativity), and physical exercise (getting body and mind moving out of inertia). Everyone behind this outrage should be in jail or better still a mental institution. The behavioural manipulators are all for this dystopian approach to schools. Professor Susan Michie, the mind-doctor and British Communist Party member, said it was wrong to say that schools were safe. They had to be made so by ‘distancing’, masks and ventilation (sitting all day in the cold). I must ask this lady round for dinner on a night I know I am going to be out and not back for weeks. She probably wouldn’t be able to make it, anyway, with all the visits to her own psychologist she must have block-booked.

Masking identity

I know how shocking it must be for you that a behaviour manipulator like Michie wants everyone to wear masks which have long been a feature of mind-control programs like the infamous MKUltra in the United States, but, there we are. We live and learn. I spent many years from 1996 to right across the millennium

researching mind control in detail on both sides of the Atlantic and elsewhere. I met a large number of mind-control survivors and many had been held captive in body and mind by MKUltra. MK stands for mind-control, but employs the German spelling in deference to the Nazis spirited out of Germany at the end of World War Two by Operation Paperclip in which the US authorities, with help from the Vatican, transported Nazi mind-controllers and engineers to America to continue their work. Many of them were behind the creation of NASA and they included Nazi scientist and SS officer Wernher von Braun who swapped designing V-2 rockets to bombard London with designing the Saturn V rockets that powered the NASA moon programme's Apollo craft. I think I may have mentioned that the Cult has no borders. Among Paperclip escapees was Josef Mengele, the Angel of Death in the Nazi concentration camps where he conducted mind and genetic experiments on children often using twins to provide a control twin to measure the impact of his 'work' on the other. If you want to observe the Cult mentality in all its extremes of evil then look into the life of Mengele. I have met many people who suffered mercilessly under Mengele in the United States where he operated under the name Dr Greene and became a stalwart of MKUltra programming and torture. Among his locations was the underground facility in the Mojave Desert in California called the China Lake Naval Weapons Station which is almost entirely below the surface. My books *The Biggest Secret*, *Children of the Matrix* and *The Perception Deception* have the detailed background to MKUltra.

The best-known MKUltra survivor is American Cathy O'Brien. I first met her and her late partner Mark Phillips at a conference in Colorado in 1996. Mark helped her escape and deprogram from decades of captivity in an offshoot of MKUltra known as Project Monarch in which 'sex slaves' were provided for the rich and famous including Father George Bush, Dick Cheney and the Clintons. Read Cathy and Mark's book *Trance-Formation of America* and if you are new to this you will be shocked to the core. I read it in 1996 shortly before, with the usual synchronicity of my life, I found

myself given a book table at the conference right next to hers. MKUltra never ended despite being very publicly exposed (only a small part of it) in the 1970s and continues in other guises. I am still in touch with Cathy. She contacted me during 2020 after masks became compulsory in many countries to tell me how they were used as part of MKUltra programming. I had been observing 'Covid regulations' and the relationship between authority and public for months. I saw techniques that I knew were employed on individuals in MKUltra being used on the global population. I had read many books and manuals on mind control including one called *Silent Weapons for Quiet Wars* which came to light in the 1980s and was a guide on how to perceptually program on a mass scale. 'Silent Weapons' refers to mind-control. I remembered a line from the manual as governments, medical authorities and law enforcement agencies have so obviously talked to – or rather at – the adult population since the 'Covid' hoax began as if they are children. The document said:

If a person is spoken to by a T.V. advertiser as if he were a twelve-year-old, then, due to suggestibility, he will, with a certain probability, respond or react to that suggestion with the uncritical response of a twelve-year-old and will reach in to his economic reservoir and deliver its energy to buy that product on impulse when he passes it in the store.

That's why authority has spoken to adults like children since all this began.

Why did Michael Jackson wear masks?

Every aspect of the 'Covid' narrative has mind-control as its central theme. Cathy O'Brien wrote an article for davidicke.com about the connection between masks and mind control. Her daughter Kelly who I first met in the 1990s was born while Cathy was still held captive in MKUltra. Kelly was forced to wear a mask as part of her programming from the age of *two* to dehumanise her, target her sense of individuality and reduce the amount of oxygen her brain and body received. *Bingo*. This is the real reason for compulsory

masks, why they have been enforced en masse, and why they seek to increase the number they demand you wear. First one, then two, with one disgraceful alleged 'doctor' recommending four which is nothing less than a death sentence. Where and how often they must be worn is being expanded for the purpose of mass mind control and damaging respiratory health which they can call 'Covid-19'. Canada's government headed by the man-child Justin Trudeau, says it's fine for children of two and older to wear masks. An insane 'study' in Italy involving just 47 children concluded there was no problem for babies as young as *four months* wearing them. Even after people were 'vaccinated' they were still told to wear masks by the criminal that is Anthony Fauci. Cathy wrote that mandating masks is allowing the authorities literally to control the air we breathe which is what was done in MKUltra. You might recall how the singer Michael Jackson wore masks and there is a reason for that. He was subjected to MKUltra mind control through Project Monarch and his psyche was scrambled by these simpletons. Cathy wrote:

In MKUltra Project Monarch mind control, Michael Jackson had to wear a mask to silence his voice so he could not reach out for help. Remember how he developed that whisper voice when he wasn't singing? Masks control the mind from the outside in, like the redefining of words is doing. By controlling what we can and cannot say for fear of being labeled racist or beaten, for example, it ultimately controls thought that drives our words and ultimately actions (or lack thereof).

Likewise, a mask muffles our speech so that we are not heard, which controls voice ... words ... mind. This is Mind Control. Masks are an obvious mind control device, and I am disturbed so many people are complying on a global scale. Masks depersonalize while making a person feel as though they have no voice. It is a barrier to others. People who would never choose to comply but are forced to wear a mask in order to keep their job, and ultimately their family fed, are compromised. They often feel shame and are subdued. People have stopped talking with each other while media controls the narrative.

The 'no voice' theme has often become literal with train passengers told not to speak to each other in case they pass on the 'virus', singing banned for the same reason and bonkers California officials telling people riding roller coasters that they cannot shout and scream. Cathy said she heard every day from healed MKUltra survivors who cannot wear a mask without flashing back on ways

their breathing was controlled – ‘from ball gags and penises to water boarding’. She said that through the years when she saw images of people in China wearing masks ‘due to pollution’ that it was really to control their oxygen levels. ‘I knew it was as much of a population control mechanism of depersonalisation as are burkas’, she said. Masks are another Chinese communist/fascist method of control that has been swept across the West as the West becomes China at lightning speed since we entered 2020.

Mask-19

There are other reasons for mandatory masks and these include destroying respiratory health to call it ‘Covid-19’ and stunting brain development of children and the young. Dr Margarite Griesz-Brisson MD, PhD, is a Consultant Neurologist and Neurophysiologist and the Founder and Medical Director of the London Neurology and Pain Clinic. Her CV goes down the street and round the corner. She is clearly someone who cares about people and won’t parrot the propaganda. Griesz-Brisson has a PhD in pharmacology, with special interest in neurotoxicology, environmental medicine, neuroregeneration and neuroplasticity (the way the brain can change in the light of information received). She went public in October, 2020, with a passionate warning about the effects of mask-wearing laws:

The reinhalation of our exhaled air will without a doubt create oxygen deficiency and a flooding of carbon dioxide. We know that the human brain is very sensitive to oxygen deprivation. There are nerve cells for example in the hippocampus that can’t be longer than 3 minutes without oxygen – they cannot survive. The acute warning symptoms are headaches, drowsiness, dizziness, issues in concentration, slowing down of reaction time – reactions of the cognitive system.

Oh, I know, let’s tell bus, truck and taxi drivers to wear them and people working machinery. How about pilots, doctors and police? Griesz-Brisson makes the important point that while the symptoms she mentions may fade as the body readjusts this does not alter the fact that people continue to operate in oxygen deficit with long list of

potential consequences. She said it was well known that neurodegenerative diseases take years or decades to develop. 'If today you forget your phone number, the breakdown in your brain would have already started 20 or 30 years ago.' She said degenerative processes in your brain are getting amplified as your oxygen deprivation continues through wearing a mask. Nerve cells in the brain are unable to divide themselves normally in these circumstances and lost nerve cells will no longer be regenerated. 'What is gone is gone.' Now consider that people like shop workers and *schoolchildren* are wearing masks for hours every day. What in the name of sanity is going to be happening to them? 'I do not wear a mask, I need my brain to think', Griesz-Brisson said, 'I want to have a clear head when I deal with my patients and not be in a carbon dioxide-induced anaesthesia'. If you are told to wear a mask anywhere ask the organisation, police, store, whatever, for their risk assessment on the dangers and negative effects on mind and body of enforcing mask-wearing. They won't have one because it has never been done not even by government. All of them must be subject to class-action lawsuits as the consequences come to light. They don't do mask risk assessments for an obvious reason. They know what the conclusions would be and independent scientific studies that *have* been done tell a horror story of consequences.

'Masks are criminal'

Dr Griesz-Brisson said that for children and adolescents, masks are an absolute no-no. They had an extremely active and adaptive immune system and their brain was incredibly active with so much to learn. 'The child's brain, or the youth's brain, is thirsting for oxygen.' The more metabolically active an organ was, the more oxygen it required; and in children and adolescents every organ was metabolically active. Griesz-Brisson said that to deprive a child's or adolescent's brain of oxygen, or to restrict it in any way, was not only dangerous to their health, it was absolutely criminal. 'Oxygen deficiency inhibits the development of the brain, and the damage that has taken place as a result CANNOT be reversed.' Mind

manipulators of MKUltra put masks on two-year-olds they wanted to neurologically rewire and you can see why. Griesz-Brisson said a child needs the brain to learn and the brain needs oxygen to function. 'We don't need a clinical study for that. This is simple, indisputable physiology.' Consciously and purposely induced oxygen deficiency was an absolutely deliberate health hazard, and an absolute medical contraindication which means that 'this drug, this therapy, this method or measure should not be used, and is not allowed to be used'. To coerce an entire population to use an absolute medical contraindication by force, she said, there had to be definite and serious reasons and the reasons must be presented to competent interdisciplinary and independent bodies to be verified and authorised. She had this warning of the consequences that were coming if mask wearing continued:

When, in ten years, dementia is going to increase exponentially, and the younger generations couldn't reach their god-given potential, it won't help to say 'we didn't need the masks'. I know how damaging oxygen deprivation is for the brain, cardiologists know how damaging it is for the heart, pulmonologists know how damaging it is for the lungs. Oxygen deprivation damages every single organ. Where are our health departments, our health insurance, our medical associations? It would have been their duty to be vehemently against the lockdown and to stop it and stop it from the very beginning.

Why do the medical boards issue punishments to doctors who give people exemptions? Does the person or the doctor seriously have to prove that oxygen deprivation harms people? What kind of medicine are our doctors and medical associations representing? Who is responsible for this crime? The ones who want to enforce it? The ones who let it happen and play along, or the ones who don't prevent it?

All of the organisations and people she mentions there either answer directly to the Cult or do whatever hierarchical levels above them tell them to do. The outcome of both is the same. 'It's not about masks, it's not about viruses, it's certainly not about your health', Griesz-Brisson said. 'It is about much, much more. I am not participating. I am not afraid.' They were taking our air to breathe and there was no unfounded medical exemption from face masks. Oxygen deprivation was dangerous for every single brain. It had to be the free decision of every human being whether they want to

wear a mask that was absolutely ineffective to protect themselves from a virus. She ended by rightly identifying where the responsibility lies for all this:

The imperative of the hour is personal responsibility. We are responsible for what we think, not the media. We are responsible for what we do, not our superiors. We are responsible for our health, not the World Health Organization. And we are responsible for what happens in our country, not the government.

Halle-bloody-lujah.

But surgeons wear masks, right?

Independent studies of mask-wearing have produced a long list of reports detailing mental, emotional and physical dangers. What a definition of insanity to see police officers imposing mask-wearing on the public which will cumulatively damage their health while the police themselves wear masks that will cumulatively damage *their* health. It's utter madness and both public and police do this because 'the government says so' – yes a government of brain-donor idiots like UK Health Secretary Matt Hancock reading the 'follow the science' scripts of psychopathic, lunatic psychologists. The response you get from Stockholm syndrome sufferers defending the very authorities that are destroying them and their families is that 'surgeons wear masks'. This is considered the game, set and match that they must work and don't cause oxygen deficit. Well, actually, scientific studies have shown that they *do* and oxygen levels are monitored in operating theatres to compensate. Surgeons wear masks to stop spittle and such like dropping into open wounds – not to stop 'viral particles' which are so miniscule they can only be seen through an electron microscope. Holes in the masks are significantly bigger than 'viral particles' and if you sneeze or cough they will breach the mask. I watched an incredibly disingenuous 'experiment' that claimed to prove that masks work in catching 'virus' material from the mouth and nose. They did this with a slow motion camera and the mask did block big stuff which stayed inside the mask and

against the face to be breathed in or cause infections on the face as we have seen with many children. 'Viral particles', however, would never have been picked up by the camera as they came through the mask when they are far too small to be seen. The 'experiment' was therefore disingenuous *and* useless.

Studies have concluded that wearing masks in operating theatres (and thus elsewhere) make no difference to preventing infection while the opposite is true with toxic shite building up in the mask and this had led to an explosion in tooth decay and gum disease dubbed by dentists 'mask mouth'. You might have seen the Internet video of a furious American doctor urging people to take off their masks after a four-year-old patient had been rushed to hospital the night before and nearly died with a lung infection that doctors sourced to mask wearing. A study in the journal *Cancer Discovery* found that inhalation of harmful microbes can contribute to advanced stage lung cancer in adults and long-term use of masks can help breed dangerous pathogens. Microbiologists have said frequent mask wearing creates a moist environment in which microbes can grow and proliferate before entering the lungs. The Canadian Agency for Drugs and Technologies in Health, or CADTH, a Canadian national organisation that provides research and analysis to healthcare decision-makers, said this as long ago as 2013 in a report entitled 'Use of Surgical Masks in the Operating Room: A Review of the Clinical Effectiveness and Guidelines'. It said:

- No evidence was found to support the use of surgical face masks to reduce the frequency of surgical site infections
- No evidence was found on the effectiveness of wearing surgical face masks to protect staff from infectious material in the operating room.
- Guidelines recommend the use of surgical face masks by staff in the operating room to protect both operating room staff and patients (despite the lack of evidence).

We were told that the world could go back to 'normal' with the arrival of the 'vaccines'. When they came, fraudulent as they are, the story changed as I knew that it would. We are in the midst of transforming 'normal', not going back to it. Mary Ramsay, head of immunisation at Public Health England, echoed the words of US criminal Anthony Fauci who said masks and other regulations must stay no matter if people are vaccinated. The Fauci idiot continued to wear two masks – different colours so both could be clearly seen – after he *claimed* to have been vaccinated. Senator Rand Paul told Fauci in one exchange that his double-masks were 'theatre' and he was right. It's all theatre. Mary Ramsay back-tracked on the vaccine-return-to-normal theme when she said the public may need to wear masks and social-distance for years despite the jabs. 'People have got used to those lower-level restrictions now, and [they] can live with them', she said telling us what the idea has been all along. 'The vaccine does not give you a pass, even if you have had it, you must continue to follow all the guidelines' said a Public Health England statement which reneged on what we had been told before and made having the 'vaccine' irrelevant to 'normality' even by the official story. Spain's fascist government trumped everyone by passing a law mandating the wearing of masks on the beach and even when swimming in the sea. The move would have devastated what's left of the Spanish tourist industry, posed potential breathing dangers to swimmers and had Northern European sunbathers walking around with their forehead brown and the rest of their face white as a sheet. The ruling was so crazy that it had to be retracted after pressure from public and tourist industry, but it confirmed where the Cult wants to go with masks and how clinically insane authority has become. The determination to make masks permanent and hide the serious dangers to body and mind can be seen in the censorship of scientist Professor Denis Rancourt by Bill Gates-funded academic publishing website ResearchGate over his papers exposing the dangers and uselessness of masks. Rancourt said:

ResearchGate today has permanently locked my account, which I have had since 2015. Their reasons graphically show the nature of their attack against democracy, and their corruption of

science ... By their obscene non-logic, a scientific review of science articles reporting on harms caused by face masks has a 'potential to cause harm'. No criticism of the psychological device (face masks) is tolerated, if the said criticism shows potential to influence public policy.

This is what happens in a fascist world.

Where are the 'greens' (again)?

Other dangers of wearing masks especially regularly relate to the inhalation of minute plastic fibres into the lungs and the deluge of discarded masks in the environment and oceans. Estimates predicted that more than 1.5 billion disposable masks will end up in the world's oceans every year polluting the water with tons of plastic and endangering marine wildlife. Studies project that humans are using 129 billion face masks each month worldwide – about three million a minute. Most are disposable and made from plastic, non-biodegradable microfibers that break down into smaller plastic particles that become widespread in ecosystems. They are littering cities, clogging sewage channels and turning up in bodies of water. I have written in other books about the immense amounts of microplastics from endless sources now being absorbed into the body. Rolf Halden, director of the Arizona State University (ASU) Biodesign Center for Environmental Health Engineering, was the senior researcher in a 2020 study that analysed 47 human tissue samples and found microplastics in all of them. 'We have detected these chemicals of plastics in every single organ that we have investigated', he said. I wrote in *The Answer* about the world being deluged with microplastics. A study by the Worldwide Fund for Nature (WWF) found that people are consuming on average every week some 2,000 tiny pieces of plastic mostly through water and also through marine life and the air. Every year humans are ingesting enough microplastics to fill a heaped dinner plate and in a life-time of 79 years it is enough to fill two large waste bins. Marco Lambertini, WWF International director general said: 'Not only are plastics polluting our oceans and waterways and killing marine life – it's in all of us and we can't escape consuming plastics,' American

geologists found tiny plastic fibres, beads and shards in rainwater samples collected from the remote slopes of the Rocky Mountain National Park near Denver, Colorado. Their report was headed: 'It is raining plastic.' Rachel Adams, senior lecturer in Biomedical Science at Cardiff Metropolitan University, said that among health consequences are internal inflammation and immune responses to a 'foreign body'. She further pointed out that microplastics become carriers of toxins including mercury, pesticides and dioxins (a known cause of cancer and reproductive and developmental problems). These toxins accumulate in the fatty tissues once they enter the body through microplastics. Now this is being compounded massively by people putting plastic on their face and throwing it away.

Workers exposed to polypropylene plastic fibres known as 'flock' have developed 'flock worker's lung' from inhaling small pieces of the flock fibres which can damage lung tissue, reduce breathing capacity and exacerbate other respiratory problems. *Now ...* commonly used surgical masks have three layers of melt-blown textiles made of ... polypropylene. We have billions of people putting these microplastics against their mouth, nose and face for hours at a time day after day in the form of masks. How does anyone think that will work out? I mean – what could possibly go wrong? We posted a number of scientific studies on this at davidicke.com, but when I went back to them as I was writing this book the links to the science research website where they were hosted were dead. Anything that challenges the official narrative in any way is either censored or vilified. The official narrative is so unsupportable by the evidence that only deleting the truth can protect it. A study by Chinese scientists still survived – with the usual twist which it why it was still active, I guess. Yes, they found that virtually all the masks they tested increased the daily intake of microplastic fibres, but people should still wear them because the danger from the 'virus' was worse said the crazy 'team' from the Institute of Hydrobiology in Wuhan. Scientists first discovered microplastics in lung tissue of some patients who died of lung cancer

in the 1990s. Subsequent studies have confirmed the potential health damage with the plastic degrading slowly and remaining in the lungs to accumulate in volume. Wuhan researchers used a machine simulating human breathing to establish that masks shed up to nearly 4,000 microplastic fibres in a month with reused masks producing more. Scientists said some masks are laced with toxic chemicals and a variety of compounds seriously restricted for both health and environmental reasons. They include cobalt (used in blue dye) and formaldehyde known to cause watery eyes, burning sensations in the eyes, nose, and throat, plus coughing, wheezing and nausea. No – that must be ‘Covid-19’.

Mask ‘worms’

There is another and potentially even more sinister content of masks. Mostly new masks of different makes filmed under a microscope around the world have been found to contain strange black fibres or ‘worms’ that appear to move or ‘crawl’ by themselves and react to heat and water. The nearest I have seen to them are the self-replicating fibres that are pulled out through the skin of those suffering from Morgellons disease which has been connected to the phenomena of ‘chemtrails’ which I will bring into the story later on. Morgellons fibres continue to grow outside the body and have a form of artificial intelligence. Black ‘worm’ fibres in masks have that kind of feel to them and there is a nanotechnology technique called ‘worm micelles’ which carry and release drugs or anything else you want to deliver to the body. For sure the suppression of humanity by mind altering drugs is the Cult agenda big time and the more excuses they can find to gain access to the body the more opportunities there are to make that happen whether through ‘vaccines’ or masks pushed against the mouth and nose for hours on end.

So let us summarise the pros and cons of masks:

Against masks: Breathing in your own carbon dioxide; depriving the body and brain of sufficient oxygen; build-up of toxins in the mask that can be breathed into the lungs and cause rashes on the face and 'mask-mouth'; breathing microplastic fibres and toxic chemicals into the lungs; dehumanisation and deleting individualisation by literally making people faceless; destroying human emotional interaction through facial expression and deleting parental connection with their babies which look for guidance to their facial expression.

For masks: They don't protect you from a 'virus' that doesn't exist and even if it did 'viral' particles are so minute they are smaller than the holes in the mask.

Governments, police, supermarkets, businesses, transport companies, and all the rest who seek to impose masks have done no risk assessment on their consequences for health and psychology and are now open to group lawsuits when the impact becomes clear with a cumulative epidemic of respiratory and other disease. Authorities will try to exploit these effects and hide the real cause by dubbing them 'Covid-19'. Can you imagine setting out to force the population to wear health-destroying masks without doing any assessment of the risks? It is criminal and it is evil, but then how many people targeted in this way, who see their children told to wear them all day at school, have asked for a risk assessment? Billions can't be imposed upon by the few unless the billions allow it. Oh, yes, with just a tinge of irony, 85 percent of all masks made worldwide come from *China*.

Wash your hands in toxic shite

'Covid' rules include the use of toxic sanitisers and again the health consequences of constantly applying toxins to be absorbed through the skin is obvious to any level of Renegade Mind. America's Food and Drug Administration (FDA) said that sanitisers are drugs and issued a warning about 75 dangerous brands which contain

methanol used in antifreeze and can cause death, kidney damage and blindness. The FDA circulated the following warning even for those brands that it claims to be safe:

Store hand sanitizer out of the reach of pets and children, and children should use it only with adult supervision. Do not drink hand sanitizer. This is particularly important for young children, especially toddlers, who may be attracted by the pleasant smell or brightly colored bottles of hand sanitizer.

Drinking even a small amount of hand sanitizer can cause alcohol poisoning in children. (However, there is no need to be concerned if your children eat with or lick their hands after using hand sanitizer.) During this coronavirus pandemic, poison control centers have had an increase in calls about accidental ingestion of hand sanitizer, so it is important that adults monitor young children's use.

Do not allow pets to swallow hand sanitizer. If you think your pet has eaten something potentially dangerous, call your veterinarian or a pet poison control center right away. Hand sanitizer is flammable and should be stored away from heat and flames. When using hand sanitizer, rub your hands until they feel completely dry before performing activities that may involve heat, sparks, static electricity, or open flames.

There you go, perfectly safe, then, and that's without even a mention of the toxins absorbed through the skin. Come on kids – sanitise your hands everywhere you go. It will save you from the 'virus'. Put all these elements together of the 'Covid' normal and see how much health and psychology is being cumulatively damaged, even devastated, to 'protect your health'. Makes sense, right? They are only imposing these things because they care, right? *Right?*

Submitting to insanity

Psychological reframing of the population goes very deep and is done in many less obvious ways. I hear people say how contradictory and crazy 'Covid' rules are and how they are ever changing. This is explained away by dismissing those involved as idiots. It is a big mistake. The Cult is delighted if its cold calculation is perceived as incompetence and idiocy when it is anything but. Oh, yes, there are idiots within the system – lots of them – but they are *administering* the Cult agenda, mostly unknowingly. They are not deciding and dictating it. The bulwark against tyranny is self-

respect, always has been, always will be. It is self-respect that has broken every tyranny in history. By its very nature self-respect will not bow to oppression and its perpetrators. There is so little self-respect that it's always the few that overturn dictators. Many may eventually follow, but the few with the iron spines (self-respect) kick it off and generate the momentum. The Cult targets self-respect in the knowledge that once this has gone only submission remains. Crazy, contradictory, ever-changing 'Covid' rules are systematically applied by psychologists to delete self-respect. They *want* you to see that the rules make no sense. It is one thing to decide to do something when *you* have made the choice based on evidence and logic. You still retain your self-respect. It is quite another when you can see what you are being told to do is insane, ridiculous and makes no sense, and *yet you still do it*. Your self-respect is extinguished and this has been happening as ever more obviously stupid and nonsensical things have been demanded and the great majority have complied even when they can see they are stupid and nonsensical.

People walk around in face-nappies knowing they are damaging their health and make no difference to a 'virus'. They do it in fear of not doing it. I know it's daft, but I'll do it anyway. When that happens something dies inside of you and submissive reframing has begun. Next there's a need to hide from yourself that you have conceded your self-respect and you convince yourself that you have not really submitted to fear and intimidation. You begin to believe that you are complying with craziness because it's the right thing to do. When first you concede your self-respect of $2+2 = 4$ to $2+2 = 5$ you *know* you are compromising your self-respect. Gradually to avoid facing that fact you begin to *believe* that $2+2=5$. You have been reframed and I have been watching this process happening in the human psyche on an industrial scale. The Cult is working to break your spirit and one of its major tools in that war is humiliation. I read how former American soldier Bradley Manning (later Chelsea Manning after a sex-change) was treated after being jailed for supplying WikiLeaks with documents exposing the enormity of

government and elite mendacity. Manning was isolated in solitary confinement for eight months, put under 24-hour surveillance, forced to hand over clothing before going to bed, and stand naked for every roll call. This is systematic humiliation. The introduction of anal swab 'Covid' tests in China has been done for the same reason to delete self-respect and induce compliant submission. Anal swabs are mandatory for incoming passengers in parts of China and American diplomats have said they were forced to undergo the indignity which would have been calculated humiliation by the Cult-owned Chinese government that has America in its sights.

Government-people: An abusive relationship

Spirit-breaking psychological techniques include giving people hope and apparent respite from tyranny only to take it away again. This happened in the UK during Christmas, 2020, when the psycho-psychologists and their political lackeys announced an easing of restrictions over the holiday only to reimpose them almost immediately on the basis of yet another lie. There is a big psychological difference between getting used to oppression and being given hope of relief only to have that dashed. Psychologists know this and we have seen the technique used repeatedly. Then there is traumatising people before you introduce more extreme regulations that require compliance. A perfect case was the announcement by the dark and sinister Whitty and Vallance in the UK that 'new data' predicted that 4,000 could die every day over the winter of 2020/2021 if we did not lockdown again. I think they call it lying and after traumatising people with that claim out came Jackboot Johnson the next day with new curbs on human freedom. Psychologists know that a frightened and traumatised mind becomes suggestable to submission and behaviour reframing. Underpinning all this has been to make people fearful and suspicious of each other and see themselves as a potential danger to others. In league with deleted self-respect you have the perfect psychological recipe for self-loathing. The relationship between authority and public is now demonstrably the same as that of

subservience to an abusive partner. These are signs of an abusive relationship explained by psychologist Leslie Becker-Phelps:

Psychological and emotional abuse: Undermining a partner's self-worth with verbal attacks, name-calling, and belittling. Humiliating the partner in public, unjustly accusing them of having an affair, or interrogating them about their every behavior. Keeping partner confused or off balance by saying they were just kidding or blaming the partner for 'making' them act this way ... Feigning in public that they care while turning against them in private. This leads to victims frequently feeling confused, incompetent, unworthy, hopeless, and chronically self-doubting. [Apply these techniques to how governments have treated the population since New Year, 2020, and the parallels are obvious.]

Physical abuse: The abuser might physically harm their partner in a range of ways, such as grabbing, hitting, punching, or shoving them. They might throw objects at them or harm them with a weapon. [Observe the physical harm imposed by masks, lockdown, and so on.]

Threats and intimidation: One way abusers keep their partners in line is by instilling fear. They might be verbally threatening, or give threatening looks or gestures. Abusers often make it known that they are tracking their partner's every move. They might destroy their partner's possessions, threaten to harm them, or threaten to harm their family members. Not surprisingly, victims of this abuse often feel anxiety, fear, and panic. [No words necessary.]

Isolation: Abusers often limit their partner's activities, forbidding them to talk or interact with friends or family. They might limit access to a car or even turn off their phone. All of this might be done by physically holding them against their will, but is often accomplished through psychological abuse and intimidation. The more isolated a person feels, the fewer resources they have to help gain perspective on their situation and to escape from it. [No words necessary.]

Economic abuse: Abusers often make their partners beholden to them for money by controlling access to funds of any kind. They might prevent their partner from getting a job or withhold access to money they earn from a job. This creates financial dependency that makes leaving the relationship very difficult. [See destruction of livelihoods and the proposed meagre 'guaranteed income' so long as you do whatever you are told.]

Using children: An abuser might disparage their partner's parenting skills, tell their children lies about their partner, threaten to take custody of their children, or threaten to harm their children. These tactics instil fear and often elicit compliance. [See reframed social service mafia and how children are being mercilessly abused by the state over 'Covid' while their parents look on too frightened to do anything.]

A further recurring trait in an abusive relationship is the abused blaming themselves for their abuse and making excuses for the abuser. We have the public blaming each other for lockdown abuse by government and many making excuses for the government while attacking those who challenge the government. How often we have heard authorities say that rules are being imposed or reimposed only because people have refused to 'behave' and follow the rules. We don't want to do it – it's *you*.

Renegade Minds are an antidote to all of these things. They will never concede their self-respect no matter what the circumstances. Even when apparent humiliation is heaped upon them they laugh in its face and reflect back the humiliation on the abuser where it belongs. Renegade Minds will never wear masks they know are only imposed to humiliate, suppress and damage both physically and psychologically. Consequences will take care of themselves and they will never break their spirit or cause them to concede to tyranny. UK newspaper columnist Peter Hitchens was one of the few in the mainstream media to speak out against lockdowns and forced vaccinations. He then announced he had taken the jab. He wanted to see family members abroad and he believed vaccine passports were inevitable even though they had not yet been introduced. Hitchens

has a questioning and critical mind, but not a Renegade one. If he had no amount of pressure would have made him concede. Hitchens excused his action by saying that the battle has been lost. Renegade Minds never accept defeat when freedom is at stake and even if they are the last one standing the self-respect of not submitting to tyranny is more important than any outcome or any consequence.

That's why Renegade Minds are the only minds that ever changed anything worth changing.

CHAPTER EIGHT

'Reframing' insanity

Insanity is relative. It depends on who has who locked in what cage
Ray Bradbury

Reframing' a mind means simply to change its perception and behaviour. This can be done subconsciously to such an extent that subjects have no idea they have been 'reframed' while to any observer changes in behaviour and attitudes are obvious.

Human society is being reframed on a ginormous scale since the start of 2020 and here we have the reason why psychologists rather than doctors have been calling the shots. Ask most people who have succumbed to 'Covid' reframing if they have changed and most will say 'no'; but they *have* and fundamentally. The Cult's long-game has been preparing for these times since way back and crucial to that has been to prepare both population and officialdom mentally and emotionally. To use the mind-control parlance they had to reframe the population with a mentality that would submit to fascism and reframe those in government and law enforcement to impose fascism or at least go along with it. The result has been the fact-deleted mindlessness of 'Wokeness' and officialdom that has either enthusiastically or unquestioningly imposed global tyranny demanded by reframed politicians on behalf of psychopathic and deeply evil cultists. 'Cognitive reframing' identifies and challenges the way someone sees the world in the form of situations, experiences and emotions and then restructures those perceptions to view the same set of circumstances in a different way. This can have

benefits if the attitudes are personally destructive while on the other side it has the potential for individual and collective mind control which the subject has no idea has even happened.

Cognitive therapy was developed in the 1960s by Aaron T. Beck who was born in Rhode Island in 1921 as the son of Jewish immigrants from the Ukraine. He became interested in the techniques as a treatment for depression. Beck's daughter Judith S. Beck is prominent in the same field and they founded the Beck Institute for Cognitive Behavior Therapy in Philadelphia in 1994. Cognitive reframing, however, began to be used worldwide by those with a very dark agenda. The Cult reframes politicians to change their attitudes and actions until they are completely at odds with what they once appeared to stand for. The same has been happening to government administrators at all levels, law enforcement, military and the human population. Cultists love mind control for two main reasons: It allows them to control what people think, do and say to secure agenda advancement and, by definition, it calms their legendary insecurity and fear of the unexpected. I have studied mind control since the time I travelled America in 1996. I may have been talking to next to no one in terms of an audience in those years, but my goodness did I gather a phenomenal amount of information and knowledge about so many things including the techniques of mind control. I have described this in detail in other books going back to *The Biggest Secret* in 1998. I met a very large number of people recovering from MKUltra and its offshoots and successors and I began to see how these same techniques were being used on the population in general. This was never more obvious than since the 'Covid' hoax began.

Reframing the enforcers

I have observed over the last two decades and more the very clear transformation in the dynamic between the police, officialdom and the public. I tracked this in the books as the relationship mutated from one of serving the public to seeing them as almost the enemy and certainly a lower caste. There has always been a class divide

based on income and always been some psychopathic, corrupt, and big-I-am police officers. This was different. Wholesale change was unfolding in the collective dynamic; it was less about money and far more about position and perceived power. An us-and-them was emerging. Noses were lifted skyward by government administration and law enforcement and their attitude to the public they were *supposed* to be serving changed to one of increasing contempt, superiority and control. The transformation was so clear and widespread that it had to be planned. Collective attitudes and dynamics do not change naturally and organically that quickly on that scale. I then came across an organisation in Britain called Common Purpose created in the late 1980s by Julia Middleton who would work in the office of Deputy Prime Minister John Prescott during the long and disastrous premiership of war criminal Tony Blair. When Blair speaks the Cult is speaking and the man should have been in jail a long time ago. Common Purpose proclaims itself to be one of the biggest 'leadership development' organisations in the world while functioning as a *charity* with all the financial benefits which come from that. It hosts 'leadership development' courses and programmes all over the world and claims to have 'brought together' what it calls 'leaders' from more than 100 countries on six continents. The modus operandi of Common Purpose can be compared with the work of the UK government's reframing network that includes the Behavioural Insights Team 'nudge unit' and 'Covid' reframing specialists at SPI-B. WikiLeaks described Common Purpose long ago as 'a hidden virus in our government and schools' which is unknown to the general public: 'It recruits and trains "leaders" to be loyal to the directives of Common Purpose and the EU, instead of to their own departments, which they then undermine or subvert, the NHS [National Health Service] being an example.' This is a vital point to understand the 'Covid' hoax. The NHS, and its equivalent around the world, has been utterly reframed in terms of administrators and much of the medical personnel with the transformation underpinned by recruitment policies. The outcome has been the criminal and psychopathic behaviour of the

NHS over 'Covid' and we have seen the same in every other major country. WikiLeaks said Common Purpose trainees are 'learning to rule without regard to democracy' and to usher in a police state (current events explained). Common Purpose operated like a 'glue' and had members in the NHS, BBC, police, legal profession, church, many of Britain's 7,000 quangos, local councils, the Civil Service, government ministries and Parliament, and controlled many RDA's (Regional Development Agencies). Here we have one answer for how and why British institutions and their like in other countries have changed so negatively in relation to the public. This further explains how and why the beyond-disgraceful reframed BBC has become a propaganda arm of 'Covid' fascism. They are all part of a network pursuing the same goal.

By 2019 Common Purpose was quoting a figure of 85,000 'leaders' that had attended its programmes. These 'students' of all ages are known as Common Purpose 'graduates' and they consist of government, state and local government officials and administrators, police chiefs and officers, and a whole range of others operating within the national, local and global establishment. Cressida Dick, Commissioner of the London Metropolitan Police, is the Common Purpose graduate who was the 'Gold Commander' that oversaw what can only be described as the murder of Brazilian electrician Jean Charles de Menezes in 2005. He was held down by psychopathic police and shot seven times in the head by a psychopathic lunatic after being mistaken for a terrorist when he was just a bloke going about his day. Dick authorised officers to pursue and keep surveillance on de Menezes and ordered that he be stopped from entering the underground train system. Police psychopaths took her at her word clearly. She was 'disciplined' for this outrage by being *promoted* – eventually to the top of the 'Met' police where she has been a disaster. Many Chief Constables controlling the police in different parts of the UK are and have been Common Purpose graduates. I have heard the 'graduate' network described as a sort of Mafia or secret society operating within the fabric of government at all levels pursuing a collective policy

ingrained at Common Purpose training events. Founder Julia Middleton herself has said:

Locally and internationally, Common Purpose graduates will be 'lighting small fires' to create change in their organisations and communities ... The Common Purpose effect is best illustrated by the many stories of small changes brought about by leaders, who themselves have changed.

A Common Purpose mission statement declared:

Common Purpose aims to improve the way society works by expanding the vision, decision-making ability and influence of all kinds of leaders. The organisation runs a variety of educational programmes for leaders of all ages, backgrounds and sectors, in order to provide them with the inspirational, information and opportunities they need to change the world.

Yes, but into what? Since 2020 the answer has become clear.

NLP and the Delphi technique

Common Purpose would seem to be a perfect name or would common programming be better? One of the foundation methods of reaching 'consensus' (group think) is by setting the agenda theme and then encouraging, cajoling or pressuring everyone to agree a 'consensus' in line with the core theme promoted by Common Purpose. The methodology involves the 'Delphi technique', or an adaption of it, in which opinions are expressed that are summarised by a 'facilitator or change agent' at each stage. Participants are 'encouraged' to modify their views in the light of what others have said. Stage by stage the former individual opinions are merged into group consensus which just happens to be what Common Purpose wants them to believe. A key part of this is to marginalise anyone refusing to concede to group think and turn the group against them to apply pressure to conform. We are seeing this very technique used on the general population to make 'Covid' group-thinkers hostile to those who have seen through the bullshit. People can be reframed by using perception manipulation methods such as Neuro-Linguistic Programming (NLP) in which you change perception with the use of

carefully constructed language. An NLP website described the technique this way:

... A method of influencing brain behaviour (the 'neuro' part of the phrase) through the use of language (the 'linguistic' part) and other types of communication to enable a person to 'recode' the way the brain responds to stimuli (that's the 'programming') and manifest new and better behaviours. Neuro-Linguistic Programming often incorporates hypnosis and self-hypnosis to help achieve the change (or 'programming') that is wanted.

British alternative media operation UKColumn has done very detailed research into Common Purpose over a long period. I quoted co-founder and former naval officer Brian Gerrish in my book *Remember Who You Are*, published in 2011, as saying the following years before current times:

It is interesting that many of the mothers who have had children taken by the State speak of the Social Services people being icily cool, emotionless and, as two ladies said in slightly different words, '... like little robots'. We know that NLP is cumulative, so people can be given small imperceptible doses of NLP in a course here, another in a few months, next year etc. In this way, major changes are accrued in their personality, but the day by day change is almost unnoticeable.

In these and other ways 'graduates' have had their perceptions uniformly reframed and they return to their roles in the institutions of government, law enforcement, legal profession, military, 'education', the UK National Health Service and the whole swathe of the establishment structure to pursue a common agenda preparing for the 'post-industrial', 'post-democratic' society. I say 'preparing' but we are now there. 'Post-industrial' is code for the Great Reset and 'post-democratic' is 'Covid' fascism. UKColumn has spoken to partners of those who have attended Common Purpose 'training'. They have described how personalities and attitudes of 'graduates' changed very noticeably for the worse by the time they had completed the course. They had been 'reframed' and told they are the 'leaders' – the special ones – who know better than the population. There has also been the very demonstrable recruitment of psychopaths and narcissists into government administration at all

levels and law enforcement. If you want psychopathy hire psychopaths and you get a simple cause and effect. If you want administrators, police officers and 'leaders' to perceive the public as lesser beings who don't matter then employ narcissists. These personalities are identified using 'psychometrics' that identifies knowledge, abilities, attitudes and personality traits, mostly through carefully-designed questionnaires and tests. As this policy has passed through the decades we have had power-crazy, power-trippers appointed into law enforcement, security and government administration in preparation for current times and the dynamic between public and law enforcement/officialdom has been transformed. UKColumn's Brian Gerrish said of the narcissistic personality:

Their love of themselves and power automatically means that they will crush others who get in their way. I received a major piece of the puzzle when a friend pointed out that when they made public officials re-apply for their own jobs several years ago they were also required to do psychometric tests. This was undoubtedly the start of the screening process to get 'their' sort of people in post.

How obvious that has been since 2020 although it was clear what was happening long before if people paid attention to the changing public-establishment dynamic.

Change agents

At the centre of events in 'Covid' Britain is the National Health Service (NHS) which has behaved disgracefully in slavishly following the Cult agenda. The NHS management structure is awash with Common Purpose graduates or 'change agents' working to a common cause. Helen Bevan, a Chief of Service Transformation at the NHS Institute for Innovation and Improvement, co-authored a document called 'Towards a million change agents, a review of the social movements literature: implications for large scale change in the NHS'. The document compared a project management approach to that of change and social movements where 'people change

themselves and each other – peer to peer’. Two definitions given for a ‘social movement’ were:

A group of people who consciously attempt to build a radically new social order; involves people of a broad range of social backgrounds; and deploys politically confrontational and socially disruptive tactics – Cyrus Zirakzadeh 1997

Collective challenges, based on common purposes and social solidarities, in sustained interaction with elites, opponents, and authorities – Sidney Tarrow 1994

Helen Bevan wrote another NHS document in which she defined ‘framing’ as ‘the process by which leaders construct, articulate and put across their message in a powerful and compelling way in order to win people to their cause and call them to action’. I think I could come up with another definition that would be rather more accurate. The National Health Service and institutions of Britain and the wider world have been taken over by reframed ‘change agents’ and that includes everything from the United Nations to national governments, local councils and social services which have been kidnapping children from loving parents on an extraordinary and gathering scale on the road to the end of parenthood altogether. Children from loving homes are stolen and kidnapped by the state and put into the ‘care’ (inversion) of the local authority through council homes, foster parents and forced adoption. At the same time children are allowed to be abused without response while many are under council ‘care’. UKColumn highlighted the Common Purpose connection between South Yorkshire Police and Rotherham council officers in the case of the scandal in that area of the sexual exploitation of children to which the authorities turned not one blind eye, but both:

We were alarmed to discover that the Chief Executive, the Strategic Director of Children and Young People's Services, the Manager for the Local Strategic Partnership, the Community Cohesion Manager, the Cabinet Member for Cohesion, the Chief Constable and his predecessor had all attended Leadership training courses provided by the pseudo-charity Common Purpose.

Once 'change agents' have secured positions of hire and fire within any organisation things start to move very quickly. Personnel are then hired and fired on the basis of whether they will work towards the agenda the change agent represents. If they do they are rapidly promoted even though they may be incompetent. Those more qualified and skilled who are pre-Common Purpose 'old school' see their careers stall and even disappear. This has been happening for decades in every institution of state, police, 'health' and social services and all of them have been transformed as a result in their attitudes to their jobs and the public. Medical professions, including nursing, which were once vocations for the caring now employ many cold, callous and couldn't give a shit personality types. The UKColumn investigation concluded:

By blurring the boundaries between people, professions, public and private sectors, responsibility and accountability, Common Purpose encourages 'graduates' to believe that as new selected leaders, they can work together, outside of the established political and social structures, to achieve a paradigm shift or CHANGE – so called 'Leading Beyond Authority'. In doing so, the allegiance of the individual becomes 'reframed' on CP colleagues and their NETWORK.

Reframing the Face-Nappies

Nowhere has this process been more obvious than in the police where recruitment of psychopaths and development of unquestioning mind-controlled group-thinkers have transformed law enforcement into a politically-correct 'Woke' joke and a travesty of what should be public service. Today they wear their face-nappies like good little gofers and enforce 'Covid' rules which are fascism under another name. Alongside the specifically-recruited psychopaths we have software minds incapable of free thought. Brian Gerrish again:

An example is the policeman who would not get on a bike for a press photo because he had not done the cycling proficiency course. Normal people say this is political correctness gone mad. Nothing could be further from the truth. The policeman has been reframed, and in his reality it is perfect common sense not to get on the bike 'because he hasn't done the cycling course'.

Another example of this is where the police would not rescue a boy from a pond until they had taken advice from above on the 'risk assessment'. A normal person would have arrived, perhaps thought of the risk for a moment, and dived in. To the police now 'reframed', they followed 'normal' procedure.

There are shocking cases of reframed ambulance crews doing the same. Sheer unthinking stupidity of London Face-Nappies headed by Common Purpose graduate Cressida Dick can be seen in their behaviour at a vigil in March, 2021, for a murdered woman, Sarah Everard. A police officer had been charged with the crime. Anyone with a brain would have left the vigil alone in the circumstances. Instead they 'manhandled' women to stop them breaking 'Covid rules' to betray classic reframing. Minds in the thrall of perception control have no capacity for seeing a situation on its merits and acting accordingly. 'Rules is rules' is their only mind-set. My father used to say that rules and regulations are for the guidance of the intelligent and the blind obedience of the idiot. Most of the intelligent, decent, coppers have gone leaving only the other kind and a few old school for whom the job must be a daily nightmare. The combination of psychopaths and rule-book software minds has been clearly on public display in the 'Covid' era with automaton robots in uniform imposing fascistic 'Covid' regulations on the population without any personal initiative or judging situations on their merits. There are thousands of examples around the world, but I'll make my point with the infamous Derbyshire police in the English East Midlands – the ones who think pouring dye into beauty spots and using drones to track people walking in the countryside away from anyone is called 'policing'. To them there are rules decreed by the government which they have to enforce and in their bewildered state a group gathering in a closed space and someone walking alone in the countryside are the same thing. It is beyond idiocy and enters the realm of clinical insanity.

Police officers in Derbyshire said they were 'horrified' – *horrified* – to find 15 to 20 'irresponsible' kids playing a football match at a closed leisure centre 'in breach of coronavirus restrictions'. When they saw the police the kids ran away leaving their belongings behind and the reframed men and women of Derbyshire police were seeking to establish their identities with a view to fining their parents. The most natural thing for youngsters to do – kicking a ball about – is turned into a criminal activity and enforced by the moronic software programs of Derbyshire police. You find the same mentality in every country. These barely conscious 'horrified' officers said they had to take action because 'we need to ensure these rules are being followed' and 'it is of the utmost importance that you ensure your children are following the rules and regulations for Covid-19'. Had any of them done ten seconds of research to see if this parroting of their masters' script could be supported by any evidence? Nope. Reframed people don't think – others think for them and that's the whole idea of reframing. I have seen police officers one after the other repeating without question word for word what officialdom tells them just as I have seen great swathes of the public doing the same. Ask either for 'their' opinion and out spews what they have been told to think by the official narrative. Police and public may seem to be in different groups, but their mentality is the same. Most people do whatever they are told in fear not doing so or because they believe what officialdom tells them; almost the entirety of the police do what they are told for the same reason. Ultimately it's the tiny inner core of the global Cult that's telling both what to do.

So Derbyshire police were 'horrified'. Oh, really? Why did they think those kids were playing football? It was to relieve the psychological consequences of lockdown and being denied human contact with their friends and interaction, touch and discourse vital to human psychological health. Being denied this month after month has dismantled the psyche of many children and young people as depression and suicide have exploded. Were Derbyshire police *horrified by that*? Are you kidding? Reframed people don't have those

mental and emotional processes that can see how the impact on the psychological health of youngsters is far more dangerous than any 'virus' even if you take the mendacious official figures to be true. The reframed are told (programmed) how to act and so they do. The Derbyshire Chief Constable in the first period of lockdown when the black dye and drones nonsense was going on was Peter Goodman. He was the man who severed the connection between his force and the Derbyshire Constabulary *Male Voice* Choir when he decided that it was not inclusive enough to allow women to join. The fact it was a male voice choir making a particular sound produced by male voices seemed to elude a guy who terrifyingly ran policing in Derbyshire. He retired weeks after his force was condemned as disgraceful by former Supreme Court Justice Jonathan Sumption for their behaviour over extreme lockdown impositions. Goodman was replaced by his deputy Rachel Swann who was in charge when her officers were 'horrificed'. The police statement over the boys committing the hanging-offence of playing football included the line about the youngsters being 'irresponsible in the times we are all living through' missing the point that the real relevance of the 'times we are all living through' is the imposition of fascism enforced by psychopaths and reframed minds of police officers playing such a vital part in establishing the fascist tyranny that their own children and grandchildren will have to live in their entire lives. As a definition of insanity that is hard to beat although it might be run close by imposing masks on people that can have a serious effect on their health while wearing a face nappy all day themselves. Once again public and police do it for the same reason – the authorities tell them to and who are they to have the self-respect to say no?

Wokers in uniform

How reframed do you have to be to arrest a *six-year-old* and take him to court for *picking a flower* while waiting for a bus? Brain dead police and officialdom did just that in North Carolina where criminal proceedings happen regularly for children under nine. Attorney Julie Boyer gave the six-year-old crayons and a colouring book

during the 'flower' hearing while the 'adults' decided his fate. County Chief District Court Judge Jay Corpening asked: 'Should a child that believes in Santa Claus, the Easter Bunny and the tooth fairy be making life-altering decisions?' Well, of course not, but common sense has no meaning when you have a common purpose and a reframed mind. Treating children in this way, and police operating in American schools, is all part of the psychological preparation for children to accept a police state as normal all their adult lives. The same goes for all the cameras and biometric tracking technology in schools. Police training is focused on reframing them as snowflake Wokers and this is happening in the military. Pentagon top brass said that 'training sessions on extremism' were needed for troops who asked why they were so focused on the Capitol Building riot when Black Lives Matter riots were ignored. What's the difference between them some apparently and rightly asked. Actually, there is a difference. Five people died in the Capitol riot, only one through violence, and that was a police officer shooting an unarmed protestor. BLM riots killed at least 25 people and cost billions. Asking the question prompted the psychopaths and reframed minds that run the Pentagon to say that more 'education' (programming) was needed. Troop training is all based on psychological programming to make them fodder for the Cult – 'Military men are just dumb, stupid animals to be used as pawns in foreign policy' as Cult-to-his-DNA former Secretary of State Henry Kissinger famously said. Governments see the police in similar terms and it's time for those among them who can see this to defend the people and stop being enforcers of the Cult agenda upon the people.

The US military, like the country itself, is being targeted for destruction through a long list of Woke impositions. Cult-owned gaga 'President' Biden signed an executive order when he took office to allow taxpayer money to pay for transgender surgery for active military personnel and veterans. Are you a man soldier? No, I'm a LGBTQIA+ with a hint of Skoliosexual and Spectrasexual. Oh, good man. Bad choice of words you bigot. The Pentagon announced in March, 2021, the appointment of the first 'diversity and inclusion

officer' for US Special Forces. Richard Torres-Estrada arrived with the publication of a 'D&I Strategic Plan which will guide the enterprise-wide effort to institutionalize and sustain D&I'. If you think a Special Forces 'Strategic Plan' should have something to do with defending America you haven't been paying attention. Defending Woke is now the military's new role. Torres-Estrada has posted images comparing Donald Trump with Adolf Hitler and we can expect no bias from him as a representative of the supposedly non-political Pentagon. Cable news host Tucker Carlson said: 'The Pentagon is now the Yale faculty lounge but with cruise missiles.' Meanwhile Secretary of Defense Lloyd Austin, a board member of weapons-maker Raytheon with stock and compensation interests in October, 2020, worth \$1.4 million, said he was purging the military of the 'enemy within' – anyone who isn't Woke and supports Donald Trump. Austin refers to his targets as 'racist extremists' while in true Woke fashion being himself a racist extremist. Pentagon documents pledge to 'eradicate, eliminate and conquer all forms of racism, sexism and homophobia'. The definitions of these are decided by 'diversity and inclusion committees' peopled by those who see racism, sexism and homophobia in every situation and opinion. Woke (the Cult) is dismantling the US military and purging testosterone as China expands its military and gives its troops 'masculinity training'. How do we think that is going to end when this is all Cult coordinated? The US military, like the British military, is controlled by Woke and spineless top brass who just go along with it out of personal career interests.

'Woke' means fast asleep

Mind control and perception manipulation techniques used on individuals to create group-think have been unleashed on the global population in general. As a result many have no capacity to see the obvious fascist agenda being installed all around them or what 'Covid' is really all about. Their brains are firewalled like a computer system not to process certain concepts, thoughts and realisations that are bad for the Cult. The young are most targeted as the adults they

will be when the whole fascist global state is planned to be fully implemented. They need to be prepared for total compliance to eliminate all pushback from entire generations. The Cult has been pouring billions into taking complete control of 'education' from schools to universities via its operatives and corporations and not least Bill Gates as always. The plan has been to transform 'education' institutions into programming centres for the mentality of 'Woke'. James McConnell, professor of psychology at the University of Michigan, wrote in *Psychology Today* in 1970:

The day has come when we can combine sensory deprivation with drugs, hypnosis, and astute manipulation of reward and punishment, to gain almost absolute control over an individual's behaviour. It should then be possible to achieve a very rapid and highly effective type of brainwashing that would allow us to make dramatic changes in a person's behaviour and personality ...

... We should reshape society so that we all would be trained from birth to want to do what society wants us to do. We have the techniques to do it... no-one owns his own personality you acquired, and there's no reason to believe you should have the right to refuse to acquire a new personality if your old one is anti-social.

This was the potential for mass brainwashing in 1970 and the mentality there displayed captures the arrogant psychopathy that drives it forward. I emphasise that not all young people have succumbed to Woke programming and those that haven't are incredibly impressive people given that today's young are the most perceptually-targeted generations in history with all the technology now involved. Vast swathes of the young generations, however, have fallen into the spell – and that's what it is – of Woke. The Woke mentality and perceptual program is founded on *inversion* and you will appreciate later why that is so significant. Everything with Woke is inverted and the opposite of what it is claimed to be. Woke was a term used in African-American culture from the 1900s and referred to an awareness of social and racial justice. This is not the meaning of the modern version or 'New Woke' as I call it in *The Answer*. Oh, no, Woke today means something very different no matter how much Wokers may seek to hide that and insist Old Woke and New

Woke are the same. See if you find any 'awareness of social justice' here in the modern variety:

- Woke demands 'inclusivity' while excluding anyone with a different opinion and calls for mass censorship to silence other views.
- Woke claims to stand against oppression when imposing oppression is the foundation of all that it does. It is the driver of political correctness which is nothing more than a Cult invention to manipulate the population to silence itself.
- Woke believes itself to be 'liberal' while pursuing a global society that can only be described as fascist (see 'anti-fascist' fascist Antifa).
- Woke calls for 'social justice' while spreading injustice wherever it goes against the common 'enemy' which can be easily identified as a differing view.
- Woke is supposed to be a metaphor for 'awake' when it is solid-gold asleep and deep in a Cult-induced coma that meets the criteria for 'off with the fairies'.

I state these points as obvious facts if people only care to look. I don't do this with a sense of condemnation. We need to appreciate that the onslaught of perceptual programming on the young has been incessant and merciless. I can understand why so many have been reframed, or, given their youth, framed from the start to see the world as the Cult demands. The Cult has had access to their minds day after day in its 'education' system for their entire formative years. Perception is formed from information received and the Cult-created system is a life-long download of information delivered to elicit a particular perception, thus behaviour. The more this has expanded into still new extremes in recent decades and ever-increasing censorship has deleted other opinions and information why wouldn't that lead to a perceptual reframing on a mass scale? I

have described already cradle-to-grave programming and in more recent times the targeting of young minds from birth to adulthood has entered the stratosphere. This has taken the form of skewing what is 'taught' to fit the Cult agenda and the omnipresent techniques of group-think to isolate non-believers and pressure them into line. There has always been a tendency to follow the herd, but we really are in a new world now in relation to that. We have parents who can see the 'Covid' hoax told by their children not to stop them wearing masks at school, being 'Covid' tested or having the 'vaccine' in fear of the peer-pressure consequences of being different. What is 'peer-pressure' if not pressure to conform to group-think? Renegade Minds never group-think and always retain a set of perceptions that are unique to them. Group-think is always underpinned by consequences for not group-thinking. Abuse now aimed at those refusing DNA-manipulating 'Covid vaccines' are a potent example of this. The biggest pressure to conform comes from the very group which is itself being manipulated. 'I am programmed to be part of a hive mind and so you must be.'

Woke control structures in 'education' now apply to every mainstream organisation. Those at the top of the 'education' hierarchy (the Cult) decide the policy. This is imposed on governments through the Cult network; governments impose it on schools, colleges and universities; their leadership impose the policy on teachers and academics and they impose it on children and students. At any level where there is resistance, perhaps from a teacher or university lecturer, they are targeted by the authorities and often fired. Students themselves regularly demand the dismissal of academics (increasingly few) at odds with the narrative that the students have been programmed to believe in. It is quite a thought that students who are being targeted by the Cult become so consumed by programmed group-think that they launch protests and demand the removal of those who are trying to push back against those targeting the students. Such is the scale of perceptual inversion. We see this with 'Covid' programming as the Cult imposes the rules via psycho-psychologists and governments on

shops, transport companies and businesses which impose them on their staff who impose them on their customers who pressure Pushbackers to conform to the will of the Cult which is in the process of destroying them and their families. Scan all aspects of society and you will see the same sequence every time.

Fact free Woke and hijacking the 'left'

There is no more potent example of this than 'Woke', a mentality only made possible by the deletion of factual evidence by an 'education' system seeking to produce an ever more uniform society. Why would you bother with facts when you don't know any? Deletion of credible history both in volume and type is highly relevant. Orwell said: 'Who controls the past controls the future: who controls the present controls the past.' They who control the perception of the past control the perception of the future and they who control the present control the perception of the past through the writing and deleting of history. Why would you oppose the imposition of Marxism in the name of Wokeism when you don't know that Marxism cost at least 100 million lives in the 20th century alone? Watch videos and read reports in which Woker generations are asked basic historical questions – it's mind-blowing. A survey of 2,000 people found that six percent of millennials (born approximately early 1980s to early 2000s) believed the Second World War (1939-1945) broke out with the assassination of President Kennedy (in 1963) and one in ten thought Margaret Thatcher was British Prime Minister at the time. She was in office between 1979 and 1990. We are in a post-fact society. Provable facts are no defence against the fascism of political correctness or Silicon Valley censorship. Facts don't matter anymore as we have witnessed with the 'Covid' hoax. Sacrificing uniqueness to the Woke group-think religion is all you are required to do and that means thinking for yourself is the biggest Woke no, no. All religions are an expression of group-think and censorship and Woke is just another religion with an orthodoxy defended by group-think and censorship. Burned at

the stake becomes burned on Twitter which leads back eventually to burned at the stake as Woke humanity regresses to ages past.

The biggest Woke inversion of all is its creators and funders. I grew up in a traditional left of centre political household on a council estate in Leicester in the 1950s and 60s – you know, the left that challenged the power of wealth-hoarding elites and threats to freedom of speech and opinion. In those days students went on marches defending freedom of speech while today's Wokers march for its deletion. What on earth could have happened? Those very elites (collectively the Cult) that we opposed in my youth and early life have funded into existence the antithesis of that former left and hijacked the 'brand' while inverting everything it ever stood for. We have a mentality that calls itself 'liberal' and 'progressive' while acting like fascists. Cult billionaires and their corporations have funded themselves into control of 'education' to ensure that Woke programming is unceasing throughout the formative years of children and young people and that non-Wokers are isolated (that word again) whether they be students, teachers or college professors. The Cult has funded into existence the now colossal global network of Woke organisations that have spawned and promoted all the 'causes' on the Cult wish-list for global transformation and turned Wokers into demanders of them. Does anyone really think it's a coincidence that the Cult agenda for humanity is a carbon (sorry) copy of the societal transformations desired by Woke?? These are only some of them:

Political correctness: The means by which the Cult deletes all public debates that it knows it cannot win if we had the free-flow of information and evidence.

Human-caused 'climate change': The means by which the Cult seeks to transform society into a globally-controlled dictatorship imposing its will over the fine detail of everyone's lives 'to save the planet' which doesn't actually need saving.

Transgender obsession: Preparing collective perception to accept the 'new human' which would not have genders because it would be created technologically and not through procreation. I'll have much more on this in Human 2.0.

Race obsession: The means by which the Cult seeks to divide and rule the population by triggering racial division through the perception that society is more racist than ever when the opposite is the case. Is it perfect in that regard? No. But to compare today with the racism of apartheid and segregation brought to an end by the civil rights movement in the 1960s is to insult the memory of that movement and inspirations like Martin Luther King. Why is the 'anti-racism' industry (which it is) so dominated by privileged white people?

White supremacy: This is a label used by privileged white people to demonise poor and deprived white people pushing back on tyranny to marginalise and destroy them. White people are being especially targeted as the dominant race by number within Western society which the Cult seeks to transform in its image. If you want to change a society you must weaken and undermine its biggest group and once you have done that by using the other groups you next turn on them to do the same ... 'Then they came for the Jews and I was not a Jew so I did nothing.'

Mass migration: The mass movement of people from the Middle East, Africa and Asia into Europe, from the south into the United States and from Asia into Australia are another way the Cult seeks to dilute the racial, cultural and political influence of white people on Western society. White people ask why their governments appear to be working against them while being politically and culturally biased towards incoming cultures. Well, here's your answer. In the same way sexually 'straight' people, men and women, ask why the

authorities are biased against them in favour of other sexualities. The answer is the same – that's the way the Cult wants it to be for very sinister motives.

These are all central parts of the Cult agenda and central parts of the Woke agenda and Woke was created and continues to be funded to an immense degree by Cult billionaires and corporations. If anyone begins to say 'coincidence' the syllables should stick in their throat.

Billionaire 'social justice warriors'

Joe Biden is a 100 percent-owned asset of the Cult and the Wokers' man in the White House whenever he can remember his name and for however long he lasts with his rapidly diminishing cognitive function. Even walking up the steps of an aircraft without falling on his arse would appear to be a challenge. He's not an empty-shell puppet or anything. From the minute Biden took office (or the Cult did) he began his executive orders promoting the Woke wish-list. You will see the Woke agenda imposed ever more severely because it's really the *Cult* agenda. Woke organisations and activist networks spawned by the Cult are funded to the extreme so long as they promote what the Cult wants to happen. Woke is funded to promote 'social justice' by billionaires who become billionaires by destroying social justice. The social justice mantra is only a cover for dismantling social justice and funded by billionaires that couldn't give a damn about social justice. Everything makes sense when you see that. One of Woke's premier funders is Cult billionaire financier George Soros who said: 'I am basically there to make money, I cannot and do not look at the social consequences of what I do.' This is the same Soros who has given more than \$32 billion to his Open Society Foundations global Woke network and funded Black Lives Matter, mass immigration into Europe and the United States, transgender activism, climate change activism, political correctness and groups targeting 'white supremacy' in the form of privileged white thugs that dominate Antifa. What a scam it all is and when

you are dealing with the unquestioning fact-free zone of Woke scamming them is child's play. All you need to pull it off in all these organisations are a few in-the-know agents of the Cult and an army of naïve, reframed, uninformed, narcissistic, know-nothings convinced of their own self-righteousness, self-purity and virtue.

Soros and fellow billionaires and billionaire corporations have poured hundreds of millions into Black Lives Matter and connected groups and promoted them to a global audience. None of this is motivated by caring about black people. These are the billionaires that have controlled and exploited a system that leaves millions of black people in abject poverty and deprivation which they do absolutely nothing to address. The same Cult networks funding BLM were behind the *slave trade*! Black Lives Matter hijacked a phrase that few would challenge and they have turned this laudable concept into a political weapon to divide society. You know that BLM is a fraud when it claims that *All Lives Matter*, the most inclusive statement of all, is 'racist'. BLM and its Cult masters don't want to end racism. To them it's a means to an end to control all of humanity never mind the colour, creed, culture or background. What has destroying the nuclear family got to do with ending racism? Nothing – but that is one of the goals of BLM and also happens to be a goal of the Cult as I have been exposing in my books for decades. Stealing children from loving parents and giving schools ever more power to override parents is part of that same agenda. BLM is a Marxist organisation and why would that not be the case when the Cult created Marxism *and* BLM? Patrisse Cullors, a BLM co-founder, said in a 2015 video that she and her fellow organisers, including co-founder Alicia Garza, are 'trained Marxists'. The lady known after marriage as Patrisse Khan-Cullors bought a \$1.4 million home in 2021 in one of the whitest areas of California with a black population of just 1.6 per cent and has so far bought *four* high-end homes for a total of \$3.2 million. How very Marxist. There must be a bit of spare in the BLM coffers, however, when Cult corporations and billionaires have handed over the best part of \$100 million. Many black people can see that Black Lives Matter is not

working for them, but against them, and this is still more confirmation. Black journalist Jason Whitlock, who had his account suspended by Twitter for simply linking to the story about the 'Marxist's' home buying spree, said that BLM leaders are 'making millions of dollars off the backs of these dead black men who they wouldn't spit on if they were on fire and alive'.

Black Lies Matter

Cult assets and agencies came together to promote BLM in the wake of the death of career criminal George Floyd who had been jailed a number of times including for forcing his way into the home of a black woman with others in a raid in which a gun was pointed at her stomach. Floyd was filmed being held in a Minneapolis street in 2020 with the knee of a police officer on his neck and he subsequently died. It was an appalling thing for the officer to do, but the same technique has been used by police on peaceful protestors of lockdown without any outcry from the Woke brigade. As unquestioning supporters of the Cult agenda Wokers have supported lockdown and all the 'Covid' claptrap while attacking anyone standing up to the tyranny imposed in its name. Court documents would later include details of an autopsy on Floyd by County Medical Examiner Dr Andrew Baker who concluded that Floyd had taken a fatal level of the drug fentanyl. None of this mattered to fact-free, question-free, Woke. Floyd's death was followed by worldwide protests against police brutality amid calls to defund the police. Throwing babies out with the bathwater is a Woke speciality. In the wake of the murder of British woman Sarah Everard a Green Party member of the House of Lords, Baroness Jones of Moulsecoomb (Nincompoopia would have been better), called for a 6pm curfew for all men. This would be in breach of the Geneva Conventions on war crimes which ban collective punishment, but that would never have crossed the black and white Woke mind of Baroness Nincompoopia who would have been far too convinced of her own self-righteousness to compute such details. Many American cities did defund the police in the face of Floyd riots

and after \$15 million was deleted from the police budget in Washington DC under useless Woke mayor Muriel Bowser car-jacking alone rose by 300 percent and within six months the US capital recorded its highest murder rate in 15 years. The same happened in Chicago and other cities in line with the Cult/Soros plan to bring fear to streets and neighbourhoods by reducing the police, releasing violent criminals and not prosecuting crime. This is the mob-rule agenda that I have warned in the books was coming for so long. Shootings in the area of Minneapolis where Floyd was arrested increased by 2,500 percent compared with the year before. Defunding the police over George Floyd has led to a big increase in dead people with many of them black. Police protection for politicians making these decisions stayed the same or increased as you would expect from professional hypocrites. The Cult doesn't actually want to abolish the police. It wants to abolish local control over the police and hand it to federal government as the psychopaths advance the Hunger Games Society. Many George Floyd protests turned into violent riots with black stores and businesses destroyed by fire and looting across America fuelled by Black Lives Matter. Woke doesn't do irony. If you want civil rights you must loot the liquor store and the supermarket and make off with a smart TV. It's the only way.

It's not a race war – it's a class war

Black people are patronised by privileged blacks and whites alike and told they are victims of white supremacy. I find it extraordinary to watch privileged blacks supporting the very system and bloodline networks behind the slave trade and parroting the same Cult-serving manipulative crap of their privileged white, often billionaire, associates. It is indeed not a race war but a class war and colour is just a diversion. Black Senator Cory Booker and black Congresswoman Maxine Waters, more residents of Nincompoopia, personify this. Once you tell people they are victims of someone else you devalue both their own responsibility for their plight and the power they have to impact on their reality and experience. Instead

we have: 'You are only in your situation because of whiteness – turn on them and everything will change.' It won't change. Nothing changes in our lives unless *we* change it. Crucial to that is never seeing yourself as a victim and always as the creator of your reality. Life is a simple sequence of choice and consequence. Make different choices and you create different consequences. *You* have to make those choices – not Black Lives Matter, the Woke Mafia and anyone else that seeks to dictate your life. Who are they these Wokers, an emotional and psychological road traffic accident, to tell you what to do? Personal empowerment is the last thing the Cult and its Black Lives Matter want black people or anyone else to have. They claim to be defending the underdog while *creating* and perpetuating the underdog. The Cult's worst nightmare is human unity and if they are going to keep blacks, whites and every other race under economic servitude and control then the focus must be diverted from what they have in common to what they can be manipulated to believe divides them. Blacks have to be told that their poverty and plight is the fault of the white bloke living on the street in the same poverty and with the same plight they are experiencing. The difference is that your plight black people is due to him, a white supremacist with 'white privilege' living on the street. Don't unite as one human family against your mutual oppressors and suppressors – fight the oppressor with the white face who is as financially deprived as you are. The Cult knows that as its 'Covid' agenda moves into still new levels of extremism people are going to respond and it has been spreading the seeds of disunity everywhere to stop a united response to the evil that targets *all of us*.

Racist attacks on 'whiteness' are getting ever more outrageous and especially through the American Democratic Party which has an appalling history for anti-black racism. Barack Obama, Joe Biden, Hillary Clinton and Nancy Pelosi all eulogised about Senator Robert Byrd at his funeral in 2010 after a nearly 60-year career in Congress. Byrd was a brutal Ku Klux Klan racist and a violent abuser of Cathy O'Brien in MKUltra. He said he would never fight in the military 'with a negro by my side' and 'rather I should die a thousand times,

and see Old Glory trampled in the dirt never to rise again, than to see this beloved land of ours become degraded by race mongrels, a throwback to the blackest specimen from the wilds'. Biden called Byrd a 'very close friend and mentor'. These 'Woke' hypocrites are not anti-racist they are anti-poor and anti-people not of their perceived class. Here is an illustration of the scale of anti-white racism to which we have now descended. Seriously Woke and moronic *New York Times* contributor Damon Young described whiteness as a 'virus' that 'like other viruses will not die until there are no bodies left for it to infect'. He went on: '... the only way to stop it is to locate it, isolate it, extract it, and kill it.' Young can say that as a black man with no consequences when a white man saying the same in reverse would be facing a jail sentence. *That's* racism. We had super-Woke numbskull senators Tammy Duckworth and Mazie Hirono saying they would object to future Biden Cabinet appointments if he did not nominate more Asian Americans and Pacific Islanders. Never mind the ability of the candidate what do they look like? Duckworth said: 'I will vote for racial minorities and I will vote for LGBTQ, but anyone else I'm not voting for.' Appointing people on the grounds of race is illegal, but that was not a problem for this ludicrous pair. They were on-message and that's a free pass in any situation.

Critical race racism

White children are told at school they are intrinsically racist as they are taught the divisive 'critical race theory'. This claims that the law and legal institutions are inherently racist and that race is a socially constructed concept used by white people to further their economic and political interests at the expense of people of colour. White is a 'virus' as we've seen. Racial inequality results from 'social, economic, and legal differences that white people create between races to maintain white interests which leads to poverty and criminality in minority communities'. I must tell that to the white guy sleeping on the street. The principal of East Side Community School in New York sent white parents a manifesto that called on

them to become 'white traitors' and advocate for full 'white abolition'. These people are teaching your kids when they urgently need a psychiatrist. The 'school' included a chart with 'eight white identities' that ranged from 'white supremacist' to 'white abolition' and defined the behaviour white people must follow to end 'the regime of whiteness'. Woke blacks and their privileged white associates are acting exactly like the slave owners of old and Ku Klux Klan racists like Robert Byrd. They are too full of their own self-purity to see that, but it's true. Racism is not a body type; it's a state of mind that can manifest through any colour, creed or culture.

Another racial fraud is '*equity*'. Not equality of treatment and opportunity – equity. It's a term spun as equality when it means something very different. Equality in its true sense is a raising up while 'equity' is a race to the bottom. Everyone in the same level of poverty is 'equity'. Keep everyone down – that's equity. The Cult doesn't want anyone in the human family to be empowered and BLM leaders, like all these 'anti-racist' organisations, continue their privileged, pampered existence by perpetuating the perception of gathering racism. When is the last time you heard an 'anti-racist' or 'anti-Semitism' organisation say that acts of racism and discrimination have *fallen*? It's not in the interests of their fundraising and power to influence and the same goes for the professional soccer anti-racism operation, Kick It Out. Two things confirmed that the Black Lives Matter riots in the summer of 2020 were Cult creations. One was that while anti-lockdown protests were condemned in this same period for 'transmitting 'Covid' the authorities supported mass gatherings of Black Lives Matter supporters. I even saw self-deluding people claiming to be doctors say the two types of protest were not the same. No – the non-existent 'Covid' was in favour of lockdowns and attacked those that protested against them while 'Covid' supported Black Lives Matter and kept well away from its protests. The whole thing was a joke and as lockdown protestors were arrested, often brutally, by reframed Face-Nappies we had the grotesque sight of police officers taking the knee to Black Lives Matter, a Cult-funded Marxist

organisation that supports violent riots and wants to destroy the nuclear family and white people.

He's not white? Shucks!

Woke obsession with race was on display again when ten people were shot dead in Boulder, Colorado, in March, 2021. Cult-owned Woke TV channels like CNN said the shooter appeared to be a white man and Wokers were on Twitter condemning 'violent white men' with the usual mantras. Then the shooter's name was released as Ahmad Al Aliwi Alissa, an anti-Trump Arab-American, and the sigh of disappointment could be heard five miles away. Never mind that ten people were dead and what that meant for their families. Race baiting was all that mattered to these sick Cult-serving people like Barack Obama who exploited the deaths to further divide America on racial grounds which is his job for the Cult. This is the man that 'racist' white Americans made the first black president of the United States and then gave him a second term. Not-very-bright Obama has become filthy rich on the back of that and today appears to have a big influence on the Biden administration. Even so he's still a downtrodden black man and a victim of white supremacy. This disingenuous fraud reveals the contempt he has for black people when he puts on a Deep South Alabama accent whenever he talks to them, no, *at* them.

Another BLM red flag was how the now fully-Woke (fully-Cult) and fully-virtue-signalled professional soccer authorities had their teams taking the knee before every match in support of Marxist Black Lives Matter. Soccer authorities and clubs displayed 'Black Lives Matter' on the players' shirts and flashed the name on electronic billboards around the pitch. Any fans that condemned what is a Freemasonic taking-the-knee ritual were widely condemned as you would expect from the Woke virtue-signallers of professional sport and the now fully-Woke media. We have reverse racism in which you are banned from criticising any race or culture except for white people for whom anything goes – say what you like, no problem. What has this got to do with racial harmony and

equality? We've had black supremacists from Black Lives Matter telling white people to fall to their knees in the street and apologise for their white supremacy. Black supremacists acting like white supremacist slave owners of the past couldn't breach their self-obsessed, race-obsessed sense of self-purity. Joe Biden appointed a race-obsessed black supremacist Kristen Clarke to head the Justice Department Civil Rights Division. Clarke claimed that blacks are endowed with 'greater mental, physical and spiritual abilities' than whites. If anyone reversed that statement they would be vilified. Clarke is on-message so no problem. She's never seen a black-white situation in which the black figure is anything but a virtuous victim and she heads the Civil Rights Division which should treat everyone the same or it isn't civil rights. Another perception of the Renegade Mind: If something or someone is part of the Cult agenda they will be supported by Woke governments and media no matter what. If they're not, they will be condemned and censored. It really is that simple and so racist Clarke prospers despite (make that because of) her racism.

The end of culture

Biden's administration is full of such racial, cultural and economic bias as the Cult requires the human family to be divided into warring factions. We are now seeing racially-segregated graduations and everything, but everything, is defined through the lens of perceived 'racism'. We have 'racist' mathematics, 'racist' food and even 'racist' *plants*. World famous Kew Gardens in London said it was changing labels on plants and flowers to tell its pre-'Covid' more than two million visitors a year how racist they are. Kew director Richard Deverell said this was part of an effort to 'move quickly to decolonise collections' after they were approached by one Ajay Chhabra 'an actor with an insight into how sugar cane was linked to slavery'. They are *plants* you idiots. 'Decolonisation' in the Woke manual really means colonisation of society with its mentality and by extension colonisation by the Cult. We are witnessing a new Chinese-style 'Cultural Revolution' so essential to the success of all

Marxist takeovers. Our cultural past and traditions have to be swept away to allow a new culture to be built-back-better. Woke targeting of long-standing Western cultural pillars including historical monuments and cancelling of historical figures is what happened in the Mao revolution in China which 'purged remnants of capitalist and traditional elements from Chinese society' and installed Maoism as the dominant ideology'. For China see the Western world today and for 'dominant ideology' see Woke. Better still see Marxism or Maoism. The 'Covid' hoax has specifically sought to destroy the arts and all elements of Western culture from people meeting in a pub or restaurant to closing theatres, music venues, sports stadiums, places of worship and even banning *singing*. Destruction of Western society is also why criticism of any religion is banned except for Christianity which again is the dominant religion as white is the numerically-dominant race. Christianity may be fading rapidly, but its history and traditions are weaved through the fabric of Western society. Delete the pillars and other structures will follow until the whole thing collapses. I am not a Christian defending that religion when I say that. I have no religion. It's just a fact. To this end Christianity has itself been turned Woke to usher its own downfall and its ranks are awash with 'change agents' – knowing and unknowing – at every level including Pope Francis (*definitely* knowing) and the clueless Archbishop of Canterbury Justin Welby (possibly not, but who can be sure?). Woke seeks to coordinate attacks on Western culture, traditions, and ways of life through 'intersectionality' defined as 'the complex, cumulative way in which the effects of multiple forms of discrimination (such as racism, sexism, and classism) combine, overlap, or intersect especially in the experiences of marginalised individuals or groups'. Wade through the Orwellian Woke-speak and this means coordinating disparate groups in a common cause to overthrow freedom and liberal values.

The entire structure of public institutions has been infested with Woke – government at all levels, political parties, police, military, schools, universities, advertising, media and trade unions. This abomination has been achieved through the Cult web by appointing

Workers to positions of power and battering non-Workers into line through intimidation, isolation and threats to their job. Many have been fired in the wake of the empathy-deleted, vicious hostility of 'social justice' Workers and the desire of gutless, spineless employers to virtue-signal their Wokeness. Corporations are filled with Workers today, most notably those in Silicon Valley. Ironically at the top they are not Woke at all. They are only exploiting the mentality their Cult masters have created and funded to censor and enslave while the Workers cheer them on until it's their turn. Thus the Woke 'liberal left' is an inversion of the traditional liberal left. Campaigning for justice on the grounds of power and wealth distribution has been replaced by campaigning for identity politics. The genuine traditional left would never have taken money from today's billionaire abusers of fairness and justice and nor would the billionaires have wanted to fund that genuine left. It would not have been in their interests to do so. The division of opinion in those days was between the haves and have nots. This all changed with Cult manipulated and funded identity politics. The division of opinion today is between Workers and non-Workers and not income brackets. Cult corporations and their billionaires may have taken wealth disparity to cataclysmic levels of injustice, but as long as they speak the language of Woke, hand out the dosh to the Woke network and censor the enemy they are 'one of us'. Billionaires who don't give a damn about injustice are laughing at them till their bellies hurt. Workers are not even close to self-aware enough to see that. The transformed 'left' dynamic means that Workers who drone on about 'social justice' are funded by billionaires that have destroyed social justice the world over. It's *why* they are billionaires.

The climate con

Nothing encapsulates what I have said more comprehensively than the hoax of human-caused global warming. I have detailed in my books over the years how Cult operatives and organisations were the pump-primers from the start of the climate con. A purpose-built vehicle for this is the Club of Rome established by the Cult in 1968

with the Rockefellers and Rothschilds centrally involved all along. Their gofer frontman Maurice Strong, a Canadian oil millionaire, hosted the Earth Summit in Rio de Janeiro, Brazil, in 1992 where the global 'green movement' really expanded in earnest under the guiding hand of the Cult. The Earth Summit established Agenda 21 through the Cult-created-and-owned United Nations to use the illusion of human-caused climate change to justify the transformation of global society to save the world from climate disaster. It is a No-Problem-Reaction-Solution sold through governments, media, schools and universities as whole generations have been terrified into believing that the world was going to end in their lifetimes unless what old people had inflicted upon them was stopped by a complete restructuring of how everything is done. Chill, kids, it's all a hoax. Such restructuring is precisely what the Cult agenda demands (purely by coincidence of course). Today this has been given the codename of the Great Reset which is only an updated term for Agenda 21 and its associated Agenda 2030. The latter, too, is administered through the UN and was voted into being by the General Assembly in 2015. Both 21 and 2030 seek centralised control of all resources and food right down to the raindrops falling on your own land. These are some of the demands of Agenda 21 established in 1992. See if you recognise this society emerging today:

- End national sovereignty
- State planning and management of all land resources, ecosystems, deserts, forests, mountains, oceans and fresh water; agriculture; rural development; biotechnology; and ensuring '*equity*'
- The state to 'define the role' of business and financial resources
- Abolition of private property
- 'Restructuring' the family unit (see BLM)
- Children raised by the state
- People told what their job will be
- Major restrictions on movement
- Creation of 'human settlement zones'

- Mass resettlement as people are forced to vacate land where they live
- Dumbing down education
- Mass global depopulation in pursuit of all the above

The United Nations was created as a Trojan horse for world government. With the climate con of critical importance to promoting that outcome you would expect the UN to be involved. Oh, it's involved all right. The UN is promoting Agenda 21 and Agenda 2030 justified by 'climate change' while also driving the climate hoax through its Intergovernmental Panel on Climate Change (IPCC), one of the world's most corrupt organisations. The IPCC has been lying ferociously and constantly since the day it opened its doors with the global media hanging unquestioningly on its every mendacious word. The Green movement is entirely Woke and has long lost its original environmental focus since it was co-opted by the Cult. An obsession with 'global warming' has deleted its values and scrambled its head. I experienced a small example of what I mean on a beautiful country walk that I have enjoyed several times a week for many years. The path merged into the fields and forests and you felt at one with the natural world. Then a 'Green' organisation, the Hampshire and Isle of Wight Wildlife Trust, took over part of the land and proceeded to cut down a large number of trees, including mature ones, to install a horrible big, bright steel 'this-is-ours-stay-out' fence that destroyed the whole atmosphere of this beautiful place. No one with a feel for nature would do that. Day after day I walked to the sound of chainsaws and a magnificent mature weeping willow tree that I so admired was cut down at the base of the trunk. When I challenged a Woke young girl in a green shirt (of course) about this vandalism she replied: 'It's a weeping willow – it will grow back.' This is what people are paying for when they donate to the Hampshire and Isle of Wight Wildlife Trust and many other 'green' organisations today. It is not the environmental movement that I knew and instead has become a support-system – as with Extinction Rebellion – for a very dark agenda.

Private jets for climate justice

The Cult-owned, Gates-funded, World Economic Forum and its founder Klaus Schwab were behind the emergence of Greta Thunberg to harness the young behind the climate agenda and she was invited to speak to the world at ... the UN. Schwab published a book, *Covid-19: The Great Reset* in 2020 in which he used the 'Covid' hoax and the climate hoax to lay out a new society straight out of Agenda 21 and Agenda 2030. Bill Gates followed in early 2021 when he took time out from destroying the world to produce a book in his name about the way to save it. Gates flies across the world in private jets and admitted that 'I probably have one of the highest greenhouse gas footprints of anyone on the planet ... my personal flying alone is gigantic.' He has also bid for the planet's biggest private jet operator. Other climate change saviours who fly in private jets include John Kerry, the US Special Presidential Envoy for Climate, and actor Leonardo DiCaprio, a 'UN Messenger of Peace with special focus on climate change'. These people are so full of bullshit they could corner the market in manure. We mustn't be sceptical, though, because the Gates book, *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need*, is a genuine attempt to protect the world and not an obvious pile of excrement attributed to a mega-psychopath aimed at selling his masters' plans for humanity. The Gates book and the other shite-pile by Klaus Schwab could have been written by the same person and may well have been. Both use 'climate change' and 'Covid' as the excuses for their new society and by coincidence the Cult's World Economic Forum and Bill and Melinda Gates Foundation promote the climate hoax and hosted Event 201 which pre-empted with a 'simulation' the very 'coronavirus' hoax that would be simulated for real on humanity within weeks. The British 'royal' family is promoting the 'Reset' as you would expect through Prince 'climate change caused the war in Syria' Charles and his hapless son Prince William who said that we must 'reset our relationship with nature and our trajectory as a species' to avoid a climate disaster. Amazing how many promoters of the 'Covid' and 'climate change' control

systems are connected to Gates and the World Economic Forum. A 'study' in early 2021 claimed that carbon dioxide emissions must fall by the equivalent of a global lockdown roughly every two years for the next decade to save the planet. The 'study' appeared in the same period that the Schwab mob claimed in a video that lockdowns destroying the lives of billions are good because they make the earth 'quieter' with less 'ambient noise'. They took down the video amid a public backlash for such arrogant, empathy-deleted stupidity. You see, however, where they are going with this. Corinne Le Quéré, a professor at the Tyndall Centre for Climate Change Research, University of East Anglia, was lead author of the climate lockdown study, and she writes for ... the World Economic Forum. Gates calls in 'his' book for changing 'every aspect of the economy' (long-time Cult agenda) and for humans to eat synthetic 'meat' (predicted in my books) while cows and other farm animals are eliminated. Australian TV host and commentator Alan Jones described what carbon emission targets would mean for farm animals in Australia alone if emissions were reduced as demanded by 35 percent by 2030 and zero by 2050:

Well, let's take agriculture, the total emissions from agriculture are about 75 million tonnes of carbon dioxide, equivalent. Now reduce that by 35 percent and you have to come down to 50 million tonnes, I've done the maths. So if you take for example 1.5 million cows, you're going to have to reduce the herd by 525,000 [by] 2030, nine years, that's 58,000 cows a year. The beef herd's 30 million, reduce that by 35 percent, that's 10.5 million, which means 1.2 million cattle have to go every year between now and 2030. This is insanity!

There are 75 million sheep. Reduce that by 35 percent, that's 26 million sheep, that's almost 3 million a year. So under the Paris Agreement over 30 million beasts, dairy cows, cattle, pigs and sheep would go. More than 8,000 every minute of every hour for the next decade, do these people know what they're talking about?

Clearly they don't at the level of campaigners, politicians and administrators. The Cult *does* know; that's the outcome it wants. We are faced with not just a war on humanity. Animals and the natural world are being targeted and I have been saying since the 'Covid' hoax began that the plan eventually was to claim that the 'deadly virus' is able to jump from animals, including farm animals and

domestic pets, to humans. Just before this book went into production came this story: 'Russia registers world's first Covid-19 vaccine for cats & dogs as makers of Sputnik V warn pets & farm animals could spread virus'. The report said 'top scientists warned that the deadly pathogen could soon begin spreading through homes and farms' and 'the next stage is the infection of farm and domestic animals'. Know the outcome and you'll see the journey. Think what that would mean for animals and keep your eye on a term called zoonosis or zoonotic diseases which transmit between animals and humans. The Cult wants to break the connection between animals and people as it does between people and people. Farm animals fit with the Cult agenda to transform food from natural to synthetic.

The gas of life is killing us

There can be few greater examples of Cult inversion than the condemnation of carbon dioxide as a dangerous pollutant when it is the gas of life. Without it the natural world would be dead and so we would all be dead. We breathe in oxygen and breathe out carbon dioxide while plants produce oxygen and absorb carbon dioxide. It is a perfect symbiotic relationship that the Cult wants to dismantle for reasons I will come to in the final two chapters. Gates, Schwab, other Cult operatives and mindless repeaters, want the world to be 'carbon neutral' by at least 2050 and the earlier the better. 'Zero carbon' is the cry echoed by lunatics calling for 'Zero Covid' when we already have it. These carbon emission targets will deindustrialise the world in accordance with Cult plans – the post-industrial, post-democratic society – and with so-called renewables like solar and wind not coming even close to meeting human energy needs blackouts and cold are inevitable. Texans got the picture in the winter of 2021 when a snow storm stopped wind turbines and solar panels from working and the lights went down along with water which relies on electricity for its supply system. Gates wants everything to be powered by electricity to ensure that his masters have the kill switch to stop all human activity, movement, cooking, water and warmth any time they like. The climate lie is so

stupendously inverted that it claims we must urgently reduce carbon dioxide when we *don't have enough*.

Co2 in the atmosphere is a little above 400 parts per million when the optimum for plant growth is 2,000 ppm and when it falls anywhere near 150 ppm the natural world starts to die and so do we. It fell to as low as 280 ppm in an 1880 measurement in Hawaii and rose to 413 ppm in 2019 with industrialisation which is why the planet has become *greener* in the industrial period. How insane then that psychopathic madman Gates is not satisfied only with blocking the rise of Co2. He's funding technology to suck it out of the atmosphere. The reason why will become clear. The industrial era is not destroying the world through Co2 and has instead turned around a potentially disastrous ongoing fall in Co2. Greenpeace co-founder and scientist Patrick Moore walked away from Greenpeace in 1986 and has exposed the green movement for fear-mongering and lies. He said that 500 million years ago there was *17 times* more Co2 in the atmosphere than we have today and levels have been falling for hundreds of millions of years. In the last 150 million years Co2 levels in Earth's atmosphere had reduced by *90 percent*. Moore said that by the time humanity began to unlock carbon dioxide from fossil fuels we were at '38 seconds to midnight' and in that sense: 'Humans are [the Earth's] salvation.' Moore made the point that only half the Co2 emitted by fossil fuels stays in the atmosphere and we should remember that all pollution pouring from chimneys that we are told is carbon dioxide is in fact nothing of the kind. It's pollution. Carbon dioxide is an invisible gas.

William Happer, Professor of Physics at Princeton University and long-time government adviser on climate, has emphasised the Co2 deficiency for maximum growth and food production. Greenhouse growers don't add carbon dioxide for a bit of fun. He said that most of the warming in the last 100 years, after the earth emerged from the super-cold period of the 'Little Ice Age' into a natural warming cycle, was over by 1940. Happer said that a peak year for warming in 1988 can be explained by a 'monster El Nino' which is a natural and cyclical warming of the Pacific that has nothing to do with 'climate

change'. He said the effect of Co2 could be compared to painting a wall with red paint in that once two or three coats have been applied it didn't matter how much more you slapped on because the wall will not get much redder. Almost all the effect of the rise in Co2 has already happened, he said, and the volume in the atmosphere would now have to *double* to increase temperature by a single degree. Climate hoaxers know this and they have invented the most ridiculously complicated series of 'feedback' loops to try to overcome this rather devastating fact. You hear puppet Greta going on cluelessly about feedback loops and this is why.

The Sun affects temperature? No you *climate denier*

Some other nonsense to contemplate: Climate graphs show that rises in temperature do not follow rises in Co2 – *it's the other way round* with a lag between the two of some 800 years. If we go back 800 years from present time we hit the Medieval Warm Period when temperatures were higher than now without any industrialisation and this was followed by the Little Ice Age when temperatures plummeted. The world was still emerging from these centuries of serious cold when many climate records began which makes the ever-repeated line of the 'hottest year since records began' meaningless when you are not comparing like with like. The coldest period of the Little Ice Age corresponded with the lowest period of sunspot activity when the Sun was at its least active. Proper scientists will not be at all surprised by this when it confirms the obvious fact that earth temperature is affected by the scale of Sun activity and the energetic power that it subsequently emits; but when is the last time you heard a climate hoaxer talking about the Sun as a source of earth temperature?? Everything has to be focussed on Co2 which makes up just 0.117 percent of so-called greenhouse gases and only a fraction of even that is generated by human activity. The rest is natural. More than 90 percent of those greenhouse gases are water vapour and clouds ([Fig 9](#)). Ban moisture I say. Have you noticed that the climate hoaxers no longer use the polar bear as their promotion image? That's because far from becoming extinct polar

bear communities are stable or thriving. Joe Bastardi, American meteorologist, weather forecaster and outspoken critic of the climate lie, documents in his book *The Climate Chronicles* how weather patterns and events claimed to be evidence of climate change have been happening since long before industrialisation: 'What happened before naturally is happening again, as is to be expected given the cyclical nature of the climate due to the design of the planet.' If you read the detailed background to the climate hoax in my other books you will shake your head and wonder how anyone could believe the crap which has spawned a multi-trillion dollar industry based on absolute garbage (see HIV causes AIDs and Sars-Cov-2 causes 'Covid-19'). Climate and 'Covid' have much in common given they have the same source. They both have the contradictory *everything* factor in which everything is explained by reference to them. It's hot – 'it's climate change'. It's cold – 'it's climate change'. I got a sniffle – 'it's Covid'. I haven't got a sniffle – 'it's Covid'. Not having a sniffle has to be a symptom of 'Covid'. Everything is and not having a sniffle is especially dangerous if you are a slow walker. For sheer audacity I offer you a Cambridge University 'study' that actually linked 'Covid' to 'climate change'. It had to happen eventually. They concluded that climate change played a role in 'Covid-19' spreading from animals to humans because ... wait for it ... I kid you not ... *the two groups were forced closer together as populations grow*. Er, that's it. The whole foundation on which this depended was that 'Bats are the likely zoonotic origin of SARS-CoV-1 and SARS-CoV-2'. Well, they are not. They are nothing to do with it. Apart from bats not being the origin and therefore 'climate change' effects on bats being irrelevant I am in awe of their academic insight. Where would we be without them? Not where we are that's for sure.

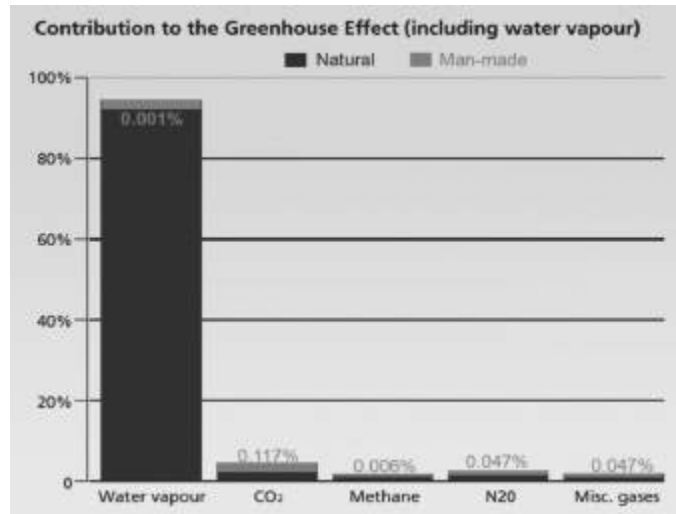


Figure 9: The idea that the gas of life is disastrously changing the climate is an insult to brain cell activity.

One other point about the weather is that climate modification is now well advanced and not every major weather event is natural – or earthquake come to that. I cover this subject at some length in other books. China is openly planning a rapid expansion of its weather modification programme which includes changing the climate in an area more than one and a half times the size of India. China used weather manipulation to ensure clear skies during the 2008 Olympics in Beijing. I have quoted from US military documents detailing how to employ weather manipulation as a weapon of war and they did that in the 1960s and 70s during the conflict in Vietnam with Operation Popeye manipulating monsoon rains for military purposes. Why would there be international treaties on weather modification if it wasn't possible? Of course it is. Weather is energetic information and it can be changed.

How was the climate hoax pulled off? See 'Covid'

If you can get billions to believe in a 'virus' that doesn't exist you can get them to believe in human-caused climate change that doesn't exist. Both are being used by the Cult to transform global society in the way it has long planned. Both hoaxes have been achieved in pretty much the same way. First you declare a lie is a fact. There's a

'virus' you call SARS-Cov-2 or humans are warming the planet with their behaviour. Next this becomes, via Cult networks, the foundation of government, academic and science policy and belief. Those who parrot the mantra are given big grants to produce research that confirms the narrative is true and ever more 'symptoms' are added to make the 'virus'/'climate change' sound even more scary. Scientists and researchers who challenge the narrative have their grants withdrawn and their careers destroyed. The media promote the lie as the unquestionable truth and censor those with an alternative view or evidence. A great percentage of the population believe what they are told as the lie becomes an everybody-knows-that and the believing-masses turn on those with a mind of their own. The technique has been used endlessly throughout human history. Wokers are the biggest promoters of the climate lie *and* 'Covid' fascism because their minds are owned by the Cult; their sense of self-righteous self-purity knows no bounds; and they exist in a bubble of reality in which facts are irrelevant and only get in the way of looking without seeing.

Running through all of this like veins in a blue cheese is control of information, which means control of perception, which means control of behaviour, which collectively means control of human society. The Cult owns the global media and Silicon Valley fascists for the simple reason that it *has* to. Without control of information it can't control perception and through that human society. Examine every facet of the Cult agenda and you will see that anything supporting its introduction is never censored while anything pushing back is always censored. I say again: Psychopaths that know why they are doing this must go before Nuremberg trials and those that follow their orders must trot along behind them into the same dock. 'I was just following orders' didn't work the first time and it must not work now. Nuremberg trials must be held all over the world before public juries for politicians, government officials, police, compliant doctors, scientists and virologists, and all Cult operatives such as Gates, Tedros, Fauci, Vallance, Whitty, Ferguson, Zuckerberg, Wojcicki, Brin, Page, Dorsey, the whole damn lot of

them – including, no *especially*, the psychopath psychologists. Without them and the brainless, gutless excuses for journalists that have repeated their lies, none of this could be happening. Nobody can be allowed to escape justice for the psychological and economic Armageddon they are all responsible for visiting upon the human race.

As for the compliant, unquestioning, swathes of humanity, and the self-obsessed, all-knowing ignorance of the Wokers ... don't start me. God help their kids. God help their grandkids. God *help them*.

CHAPTER NINE

We must have it? So what is it?

*Well I won't back down. No, I won't back down. You can stand me
up at the Gates of Hell. But I won't back down*

Tom Petty

I will now focus on the genetically-manipulating 'Covid vaccines' which do not meet this official definition of a vaccine by the US Centers for Disease Control (CDC): 'A product that stimulates a person's immune system to produce immunity to a specific disease, protecting the person from that disease.' On that basis 'Covid vaccines' are not a vaccine in that the makers don't even claim they stop infection or transmission.

They are instead part of a multi-levelled conspiracy to change the nature of the human body and what it means to be 'human' and to depopulate an enormous swathe of humanity. What I shall call Human 1.0 is on the cusp of becoming Human 2.0 and for very sinister reasons. Before I get to the 'Covid vaccine' in detail here's some background to vaccines in general. Government regulators do not test vaccines – the makers do – and the makers control which data is revealed and which isn't. Children in America are given 50 vaccine doses by age six and 69 by age 19 and the effect of the whole combined schedule has never been tested. Autoimmune diseases when the immune system attacks its own body have soared in the mass vaccine era and so has disease in general in children and the young. Why wouldn't this be the case when vaccines target the *immune system*? The US government gave Big Pharma drug

companies immunity from prosecution for vaccine death and injury in the 1986 National Childhood Vaccine Injury Act (NCVIA) and since then the government (taxpayer) has been funding compensation for the consequences of Big Pharma vaccines. The criminal and satanic drug giants can't lose and the vaccine schedule has increased dramatically since 1986 for this reason. There is no incentive to make vaccines safe and a big incentive to make money by introducing ever more. Even against a ridiculously high bar to prove vaccine liability, and with the government controlling the hearing in which it is being challenged for compensation, the vaccine court has so far paid out more than \$4 billion. These are the vaccines we are told are safe and psychopaths like Zuckerberg censor posts saying otherwise. The immunity law was even justified by a ruling that vaccines by their nature were 'unavoidably unsafe'.

Check out the ingredients of vaccines and you will be shocked if you are new to this. *They put that in children's bodies?? What??* Try aluminium, a brain toxin connected to dementia, aborted foetal tissue and formaldehyde which is used to embalm corpses. World-renowned aluminium expert Christopher Exley had his research into the health effect of aluminium in vaccines shut down by Keele University in the UK when it began taking funding from the Bill and Melinda Gates Foundation. Research when diseases 'eradicated' by vaccines began to decline and you will find the fall began long *before* the vaccine was introduced. Sometimes the fall even plateaued after the vaccine. Diseases like scarlet fever for which there was no vaccine declined in the same way because of environmental and other factors. A perfect case in point is the polio vaccine. Polio began when lead arsenate was first sprayed as an insecticide and residues remained in food products. Spraying started in 1892 and the first US polio epidemic came in Vermont in 1894. The simple answer was to stop spraying, but Rockefeller-created Big Pharma had a better idea. Polio was decreed to be caused by the *poliovirus* which 'spreads from person to person and can infect a person's spinal cord'. Lead arsenate was replaced by the lethal DDT which had the same effect of causing paralysis by damaging the brain and central nervous

system. Polio plummeted when DDT was reduced and then banned, but the vaccine is still given the credit for something it didn't do. Today by far the biggest cause of polio is the vaccines promoted by Bill Gates. Vaccine justice campaigner Robert Kennedy Jr, son of assassinated (by the Cult) US Attorney General Robert Kennedy, wrote:

In 2017, the World Health Organization (WHO) reluctantly admitted that the global explosion in polio is predominantly vaccine strain. The most frightening epidemics in Congo, Afghanistan, and the Philippines, are all linked to vaccines. In fact, by 2018, 70% of global polio cases were vaccine strain.

Vaccines make fortunes for Cult-owned Gates and Big Pharma while undermining the health and immune systems of the population. We had a glimpse of the mentality behind the Big Pharma cartel with a report on WION (World is One News), an international English language TV station based in India, which exposed the extraordinary behaviour of US drug company Pfizer over its 'Covid vaccine'. The WION report told how Pfizer had made fantastic demands of Argentina, Brazil and other countries in return for its 'vaccine'. These included immunity from prosecution, even for Pfizer negligence, government insurance to protect Pfizer from law suits and handing over as collateral sovereign assets of the country to include Argentina's bank reserves, military bases and embassy buildings. Pfizer demanded the same of Brazil in the form of waiving sovereignty of its assets abroad; exempting Pfizer from Brazilian laws; and giving Pfizer immunity from all civil liability. This is a 'vaccine' developed with government funding. Big Pharma is evil incarnate as a creation of the Cult and all must be handed tickets to Nuremberg.

Phantom 'vaccine' for a phantom 'disease'

I'll expose the 'Covid vaccine' fraud and then go on to the wider background of why the Cult has set out to 'vaccinate' every man, woman and child on the planet for an alleged 'new disease' with a survival rate of 99.77 percent (or more) even by the grotesquely-

manipulated figures of the World Health Organization and Johns Hopkins University. The 'infection' to 'death' ratio is 0.23 to 0.15 percent according to Stanford epidemiologist Dr John Ioannidis and while estimates vary the danger remains tiny. I say that if the truth be told the fake infection to fake death ratio is zero. Never mind all the evidence I have presented here and in *The Answer* that there is no 'virus' let us just focus for a moment on that death-rate figure of say 0.23 percent. The figure includes all those worldwide who have tested positive with a test not testing for the 'virus' and then died within 28 days or even longer of any other cause – *any other cause*. Now subtract all those illusory 'Covid' deaths on the global data sheets from the 0.23 percent. What do you think you would be left with? *Zero*. A vaccination has never been successfully developed for a so-called coronavirus. They have all failed at the animal testing stage when they caused hypersensitivity to what they were claiming to protect against and made the impact of a disease far worse. Cult-owned vaccine corporations got around that problem this time by bypassing animal trials, going straight to humans and making the length of the 'trials' before the public rollout as short as they could get away with. Normally it takes five to ten years or more to develop vaccines that still cause demonstrable harm to many people and that's without including the long-term effects that are never officially connected to the vaccination. 'Covid' non-vaccines have been officially produced and approved in a matter of months from a standing start and part of the reason is that (a) they were developed before the 'Covid' hoax began and (b) they are based on computer programs and not natural sources. Official non-trials were so short that government agencies gave *emergency*, not full, approval. 'Trials' were not even completed and full approval cannot be secured until they are. Public 'Covid vaccination' is actually a *continuation of the trial*. Drug company 'trials' are not scheduled to end until 2023 by which time a lot of people are going to be dead. Data on which government agencies gave this emergency approval was supplied by the Big Pharma corporations themselves in the form of Pfizer/BioNTech, AstraZeneca, Moderna, Johnson & Johnson, and

others, and this is the case with all vaccines. By its very nature *emergency* approval means drug companies do not have to prove that the 'vaccine' is 'safe and effective'. How could they with trials way short of complete? Government regulators only have to *believe* that they *could* be safe and effective. It is criminal manipulation to get products in circulation with no testing worth the name. Agencies giving that approval are infested with Big Pharma-connected place-people and they act in the interests of Big Pharma (the Cult) and not the public about whom they do not give a damn.

More human lab rats

'Covid vaccines' produced in record time by Pfizer/BioNTech and Moderna employ a technique *never approved before for use on humans*. They are known as mRNA 'vaccines' and inject a synthetic version of 'viral' mRNA or 'messenger RNA'. The key is in the term 'messenger'. The body works, or doesn't, on the basis of information messaging. Communications are constantly passing between and within the genetic system and the brain. Change those messages and you change the state of the body and even its very nature and you can change psychology and behaviour by the way the brain processes information. I think you are going to see significant changes in personality and perception of many people who have had the 'Covid vaccine' synthetic potions. Insider Aldous Huxley predicted the following in 1961 and mRNA 'vaccines' can be included in the term 'pharmacological methods':

There will be, in the next generation or so, a pharmacological method of making people love their servitude, and producing dictatorship without tears, so to speak, producing a kind of painless concentration camp for entire societies, so that people will in fact have their own liberties taken away from them, but rather enjoy it, because they will be distracted from any desire to rebel by propaganda or brainwashing, or brainwashing enhanced by pharmacological methods. And this seems to be the final revolution.

Apologists claim that mRNA synthetic 'vaccines' don't change the DNA genetic blueprint because RNA does not affect DNA only the other way round. This is so disingenuous. A process called 'reverse

transcription' can convert RNA into DNA and be integrated into DNA in the cell nucleus. This was highlighted in December, 2020, by scientists at Harvard and Massachusetts Institute of Technology (MIT). Geneticists report that more than 40 percent of mammalian genomes results from reverse transcription. On the most basic level if messaging changes then that sequence must lead to changes in DNA which is receiving and transmitting those communications. How can introducing synthetic material into cells not change the cells where DNA is located? The process is known as transfection which is defined as 'a technique to insert foreign nucleic acid (DNA or RNA) into a cell, typically with the intention of altering the properties of the cell'. Researchers at the Sloan Kettering Institute in New York found that changes in messenger RNA can deactivate tumour-suppressing proteins and thereby promote cancer. This is what happens when you mess with messaging. 'Covid vaccine' maker Moderna was founded in 2010 by Canadian stem cell biologist Derrick J. Rossi after his breakthrough discovery in the field of transforming and reprogramming stem cells. These are neutral cells that can be programmed to become any cell including sperm cells. Moderna was therefore founded on the principle of genetic manipulation and has never produced any vaccine or drug before its genetically-manipulating synthetic 'Covid' shite. Look at the name – Mode-RNA or Modify-RNA. Another important point is that the US Supreme Court has ruled that genetically-modified DNA, or complementary DNA (cDNA) synthesized in the laboratory from messenger RNA, can be patented and owned. These psychopaths are doing this to the human body.

Cells replicate synthetic mRNA in the 'Covid vaccines' and in theory the body is tricked into making antigens which trigger antibodies to target the 'virus spike proteins' which as Dr Tom Cowan said have *never been seen*. Cut the crap and these 'vaccines' deliver *self-replicating* synthetic material to the cells with the effect of changing human DNA. The more of them you have the more that process is compounded while synthetic material is all the time self-replicating. 'Vaccine'-maker Moderna describes mRNA as 'like

software for the cell' and so they are messing with the body's software. What happens when you change the software in a computer? Everything changes. For this reason the Cult is preparing a production line of mRNA 'Covid vaccines' and a long list of excuses to use them as with all the 'variants' of a 'virus' never shown to exist. The plan is further to transfer the mRNA technique to other vaccines mostly given to children and young people. The cumulative consequences will be a transformation of human DNA through a constant infusion of synthetic genetic material which will kill many and change the rest. Now consider that governments that have given emergency approval for a vaccine that's not a vaccine; never been approved for humans before; had no testing worth the name; and the makers have been given immunity from prosecution for any deaths or adverse effects suffered by the public. The UK government awarded *permanent legal indemnity* to itself and its employees for harm done when a patient is being treated for 'Covid-19' or 'suspected Covid-19'. That is quite a thought when these are possible 'side-effects' from the 'vaccine' (they are not 'side', they are effects) listed by the US Food and Drug Administration:

Guillain-Barre syndrome; acute disseminated encephalomyelitis; transverse myelitis; encephalitis; myelitis; encephalomyelitis; meningoencephalitis; meningitis; encephalopathy; convulsions; seizures; stroke; narcolepsy; cataplexy; anaphylaxis; acute myocardial infarction (heart attack); myocarditis; pericarditis; autoimmune disease; death; implications for pregnancy, and birth outcomes; other acute demyelinating diseases; non anaphylactic allergy reactions; thrombocytopenia ; disseminated intravascular coagulation; venous thromboembolism; arthritis; arthralgia; joint pain; Kawasaki disease; multisystem inflammatory syndrome in children; vaccine enhanced disease. The latter is the way the 'vaccine' has the potential to make diseases far worse than they would otherwise be.

UK doctor and freedom campaigner Vernon Coleman described the conditions in this list as 'all unpleasant, most of them very serious, and you can't get more serious than death'. The thought that anyone at all has had the 'vaccine' in these circumstances is testament to the potential that humanity has for clueless, unquestioning, stupidity and for many that programmed stupidity has already been terminal.

An insider speaks

Dr Michael Yeadon is a former Vice President, head of research and Chief Scientific Adviser at vaccine giant Pfizer. Yeadon worked on the inside of Big Pharma, but that did not stop him becoming a vocal critic of 'Covid vaccines' and their potential for multiple harms, including infertility in women. By the spring of 2021 he went much further and even used the no, no, term 'conspiracy'. When you begin to see what is going on it is impossible not to do so. Yeadon spoke out in an interview with freedom campaigner James Delingpole and I mentioned earlier how he said that no one had samples of 'the virus'. He explained that the mRNA technique originated in the anti-cancer field and ways to turn on and off certain genes which could be advantageous if you wanted to stop cancer growing out of control. 'That's the origin of them. They are a very unusual application, really.' Yeadon said that treating a cancer patient with an aggressive procedure might be understandable if the alternative was dying, but it was quite another thing to use the same technique as a public health measure. Most people involved wouldn't catch the infectious agent you were vaccinating against and if they did they probably wouldn't die:

If you are really using it as a public health measure you really want to as close as you can get to zero side-effects ... I find it odd that they chose techniques that were really cutting their teeth in the field of oncology and I'm worried that in using gene-based vaccines that have to be injected in the body and spread around the body, get taken up into some cells, and the regulators haven't quite told us which cells they get taken up into ... you are going to be generating a wide range of responses ... with multiple steps each of which could go well or badly.

I doubt the Cult intends it to go well. Yeadon said that you can put any gene you like into the body through the 'vaccine'. 'You can certainly give them a gene that would do them some harm if you wanted.' I was intrigued when he said that when used in the cancer field the technique could turn genes on and off. I explore this process in *The Answer* and with different genes having different functions you could create mayhem – physically and psychologically – if you turned the wrong ones on and the right ones off. I read reports of an experiment by researchers at the University of Washington's school of computer science and engineering in which they encoded DNA to infect computers. The body is itself a biological computer and if human DNA can inflict damage on a computer why can't the computer via synthetic material mess with the human body? It can. The Washington research team said it was possible to insert malicious malware into 'physical DNA strands' and corrupt the computer system of a gene sequencing machine as it 'reads gene letters and stores them as binary digits 0 and 1'. They concluded that hackers could one day use blood or spit samples to access computer systems and obtain sensitive data from police forensics labs or infect genome files. It is at this level of digital interaction that synthetic 'vaccines' need to be seen to get the full picture and that will become very clear later on. Michael Yeadon said it made no sense to give the 'vaccine' to younger people who were in no danger from the 'virus'. What was the benefit? It was all downside with potential effects:

The fact that my government in what I thought was a civilised, rational country, is raining [the 'vaccine'] on people in their 30s and 40s, even my children in their 20s, they're getting letters and phone calls, I know this is not right and any of you doctors who are vaccinating you know it's not right, too. They are not at risk. They are not at risk from the disease, so you are now hoping that the side-effects are so rare that you get away with it. You don't give new technology ... that you don't understand to 100 percent of the population.

Blood clot problems with the AstraZeneca 'vaccine' have been affecting younger people to emphasise the downside risks with no benefit. AstraZeneca's version, produced with Oxford University, does not use mRNA, but still gets its toxic cocktail inside cells where

it targets DNA. The Johnson & Johnson 'vaccine' which uses a similar technique has also produced blood clot effects to such an extent that the United States paused its use at one point. They are all 'gene therapy' (cell modification) procedures and not 'vaccines'. The truth is that once the content of these injections enter cells we have no idea what the effect will be. People can speculate and some can give very educated opinions and that's good. In the end, though, only the makers know what their potions are designed to do and even they won't know every last consequence. Michael Yeadon was scathing about doctors doing what they knew to be wrong. 'Everyone's mute', he said. Doctors in the NHS must know this was not right, coming into work and injecting people. 'I don't know how they sleep at night. I know I couldn't do it. I know that if I were in that position I'd have to quit.' He said he knew enough about toxicology to know this was not a good risk-benefit. Yeadon had spoken to seven or eight university professors and all except two would not speak out publicly. Their universities had a policy that no one said anything that countered the government and its medical advisors. They were afraid of losing their government grants. This is how intimidation has been used to silence the truth at every level of the system. I say silence, but these people could still speak out if they made that choice. Yeadon called them 'moral cowards' – 'This is about your children and grandchildren's lives and you have just buggered off and left it.'

'Variant' nonsense

Some of his most powerful comments related to the alleged 'variants' being used to instil more fear, justify more lockdowns, and introduce more 'vaccines'. He said government claims about 'variants' were nonsense. He had checked the alleged variant 'codes' and they were 99.7 percent identical to the 'original'. This was the human identity difference equivalent to putting a baseball cap on and off or wearing it the other way round. A 0.3 percent difference would make it impossible for that 'variant' to escape immunity from the 'original'. This made no sense of having new 'vaccines' for

‘variants’. He said there would have to be at least a *30 percent* difference for that to be justified and even then he believed the immune system would still recognise what it was. Gates-funded ‘variant modeller’ and ‘vaccine’-pusher John Edmunds might care to comment. Yeadon said drug companies were making new versions of the ‘vaccine’ as a ‘top up’ for ‘variants’. Worse than that, he said, the ‘regulators’ around the world like the MHRA in the UK had got together and agreed that because ‘vaccines’ for ‘variants’ were so similar to the first ‘vaccines’ *they did not have to do safety studies*. How transparently sinister that is. This is when Yeadon said: ‘There is a conspiracy here.’ There was no need for another vaccine for ‘variants’ and yet we were told that there was and the country had shut its borders because of them. ‘They are going into hundreds of millions of arms without passing ‘go’ or any regulator. Why did they do that? Why did they pick this method of making the vaccine?’

The reason had to be something bigger than that it seemed and ‘it’s not protection against the virus’. It’s was a far bigger project that meant politicians and advisers were willing to do things and not do things that knowingly resulted in avoidable deaths – ‘that’s already happened when you think about lockdown and deprivation of health care for a year.’ He spoke of people prepared to do something that results in the avoidable death of their fellow human beings and it not bother them. This is the penny-drop I have been working to get across for more than 30 years – the level of pure evil we are dealing with. Yeadon said his friends and associates could not believe there could be that much evil, but he reminded them of Stalin, Pol Pot and Hitler and of what Stalin had said: ‘One death is a tragedy. A million? A statistic.’ He could not think of a benign explanation for why you need top-up vaccines ‘which I’m sure you don’t’ and for the regulators ‘to just get out of the way and wave them through’. Why would the regulators do that when they were still wrestling with the dangers of the ‘parent’ vaccine? He was clearly shocked by what he had seen since the ‘Covid’ hoax began and now he was thinking the previously unthinkable:

If you wanted to depopulate a significant proportion of the world and to do it in a way that doesn't involve destruction of the environment with nuclear weapons, poisoning everyone with anthrax or something like that, and you wanted plausible deniability while you had a multi-year infectious disease crisis, I actually don't think you could come up with a better plan of work than seems to be in front of me. I can't say that's what they are going to do, but I can't think of a benign explanation why they are doing it.

He said he never thought that they would get rid of 99 percent of humans, but now he wondered. 'If you wanted to that this would be a hell of a way to do it – it would be unstoppable folks.' Yeadon had concluded that those who submitted to the 'vaccine' would be allowed to have some kind of normal life (but for how long?) while screws were tightened to coerce and mandate the last few percent. 'I think they'll put the rest of them in a prison camp. I wish I was wrong, but I don't think I am.' Other points he made included: There were no coronavirus vaccines then suddenly they all come along at the same time; we have no idea of the long term affect with trials so short; coercing or forcing people to have medical procedures is against the Nuremberg Code instigated when the Nazis did just that; people should at least delay having the 'vaccine'; a quick Internet search confirms that masks don't reduce respiratory viral transmission and 'the government knows that'; they have smashed civil society and they know that, too; two dozen peer-reviewed studies show no connection between lockdown and reducing deaths; he knew from personal friends the elite were still flying around and going on holiday while the public were locked down; the elite were not having the 'vaccines'. He was also asked if 'vaccines' could be made to target difference races. He said he didn't know, but the document by the Project for the New American Century in September, 2000, said developing 'advanced forms of biological warfare that can target *specific genotypes* may transform biological warfare from the realm of terror to a politically useful tool.' Oh, they're evil all right. Of that we can be *absolutely* sure.

Another cull of old people

We have seen from the CDC definition that the mRNA 'Covid vaccine' is not a vaccine and nor are the others that *claim* to reduce 'severity of symptoms' in *some* people, but not protect from infection or transmission. What about all the lies about returning to 'normal' if people were 'vaccinated'? If they are not claimed to stop infection and transmission of the alleged 'virus', how does anything change? This was all lies to manipulate people to take the jabs and we are seeing that now with masks and distancing still required for the 'vaccinated'. How did they think that elderly people with fragile health and immune responses were going to be affected by infusing their cells with synthetic material and other toxic substances? They *knew* that in the short and long term it would be devastating and fatal as the culling of the old that began with the first lockdowns was continued with the 'vaccine'. Death rates in care homes soared immediately residents began to be 'vaccinated' – infused with synthetic material. Brave and committed whistleblower nurses put their careers at risk by exposing this truth while the rest kept their heads down and their mouths shut to put their careers before those they are supposed to care for. A long-time American Certified Nursing Assistant who gave his name as James posted a video in which he described emotionally what happened in his care home when vaccination began. He said that during 2020 very few residents were sick with 'Covid' and no one died during the entire year; but shortly after the Pfizer mRNA injections 14 people died within two weeks and many others were near death. 'They're dropping like flies', he said. Residents who walked on their own before the shot could no longer and they had lost their ability to conduct an intelligent conversation. The home's management said the sudden deaths were caused by a 'super-spreader' of 'Covid-19'. Then how come, James asked, that residents who refused to take the injections were not sick? It was a case of inject the elderly with mRNA synthetic potions and blame their illness and death that followed on the 'virus'. James described what was happening in care homes as 'the greatest crime of genocide this country has ever seen'. Remember the NHS staff nurse from earlier who used the same

word 'genocide' for what was happening with the 'vaccines' and that it was an 'act of human annihilation'. A UK care home whistleblower told a similar story to James about the effect of the 'vaccine' in deaths and 'outbreaks' of illness dubbed 'Covid' after getting the jab. She told how her care home management and staff had zealously imposed government regulations and no one was allowed to even question the official narrative let alone speak out against it. She said the NHS was even worse. Again we see the results of reframing. A worker at a local care home where I live said they had not had a single case of 'Covid' there for almost a year and when the residents were 'vaccinated' they had 19 positive cases in two weeks with eight dying.

It's not the 'vaccine' – honest

The obvious cause and effect was being ignored by the media and most of the public. Australia's health minister Greg Hunt (a former head of strategy at the World Economic Forum) was admitted to hospital after he had the 'vaccine'. He was suffering according to reports from the skin infection 'cellulitis' and it must have been a severe case to have warranted days in hospital. Immediately the authorities said this was nothing to do with the 'vaccine' when an effect of some vaccines is a 'cellulitis-like reaction'. We had families of perfectly healthy old people who died after the 'vaccine' saying that if only they had been given the 'vaccine' earlier they would still be alive. As a numbskull rating that is off the chart. A father of four 'died of Covid' at aged 48 when he was taken ill two days after having the 'vaccine'. The man, a health administrator, had been 'shielding during the pandemic' and had 'not really left the house' until he went for the 'vaccine'. Having the 'vaccine' and then falling ill and dying does not seem to have qualified as a possible cause and effect and 'Covid-19' went on his death certificate. His family said they had no idea how he 'caught the virus'. A family member said: 'Tragically, it could be that going for a vaccination ultimately led to him catching Covid ...The sad truth is that they are never going to know where it came from.' The family warned people to remember

that the virus still existed and was 'very real'. So was their stupidity. Nurses and doctors who had the first round of the 'vaccine' were collapsing, dying and ending up in a hospital bed while they or their grieving relatives were saying they'd still have the 'vaccine' again despite what happened. I kid you not. You mean if your husband returned from the dead he'd have the same 'vaccine' again that killed him??

Doctors at the VCU Medical Center in Richmond, Virginia, said the Johnson & Johnson 'vaccine' was to blame for a man's skin peeling off. Patient Richard Terrell said: 'It all just happened so fast. My skin peeled off. It's still coming off on my hands now.' He said it was stinging, burning and itching and when he bent his arms and legs it was very painful with 'the skin swollen and rubbing against itself'. Pfizer/BioNTech and Moderna vaccines use mRNA to change the cell while the Johnson & Johnson version uses DNA in a process similar to AstraZeneca's technique. Johnson & Johnson and AstraZeneca have both had their 'vaccines' paused by many countries after causing serious blood problems. Terrell's doctor Fnu Nutan said he could have died if he hadn't got medical attention. It sounds terrible so what did Nutan and Terrell say about the 'vaccine' now? Oh, they still recommend that people have it. A nurse in a hospital bed 40 minutes after the vaccination and unable to swallow due to throat swelling was told by a doctor that he lost mobility in his arm for 36 hours following the vaccination. What did he say to the ailing nurse? 'Good for you for getting the vaccination.' We are dealing with a serious form of cognitive dissonance madness in both public and medical staff. There is a remarkable correlation between those having the 'vaccine' and trumpeting the fact and suffering bad happenings shortly afterwards. Witold Rogiewicz, a Polish doctor, made a video of his 'vaccination' and ridiculed those who were questioning its safety and the intentions of Bill Gates: 'Vaccinate yourself to protect yourself, your loved ones, friends and also patients. And to mention quickly I have info for anti-vaxxers and anti-Covidors if you want to contact Bill Gates you can do this through me.' He further ridiculed the dangers of 5G. Days later he

was dead, but naturally the vaccination wasn't mentioned in the verdict of 'heart attack'.

Lies, lies and more lies

So many members of the human race have slipped into extreme states of insanity and unfortunately they include reframed doctors and nursing staff. Having a 'vaccine' and dying within minutes or hours is not considered a valid connection while death from any cause within 28 days or longer of a positive test with a test not testing for the 'virus' means 'Covid-19' goes on the death certificate. How could that 'vaccine'-death connection not have been made except by calculated deceit? US figures in the initial rollout period to February 12th, 2020, revealed that a third of the deaths reported to the CDC after 'Covid vaccines' happened within 48 hours. Five men in the UK suffered an 'extremely rare' blood clot problem after having the AstraZeneca 'vaccine', but no causal link was established said the Gates-funded Medicines and Healthcare products Regulatory Agency (MHRA) which had given the 'vaccine' emergency approval to be used. Former Pfizer executive Dr Michael Yeadon explained in his interview how the procedures could cause blood coagulation and clots. People who should have been at no risk were dying from blood clots in the brain and he said he had heard from medical doctor friends that people were suffering from skin bleeding and massive headaches. The AstraZeneca 'shot' was stopped by some 20 countries over the blood clotting issue and still the corrupt MHRA, the European Medicines Agency (EMA) and the World Health Organization said that it should continue to be given even though the EMA admitted that it 'still cannot rule out definitively' a link between blood clotting and the 'vaccine'. Later Marco Cavaleri, head of EMA vaccine strategy, said there was indeed a clear link between the 'vaccine' and thrombosis, but they didn't know why. So much for the trials showing the 'vaccine' is safe. Blood clots were affecting younger people who would be under virtually no danger from 'Covid' even if it existed which makes it all the more stupid and sinister.

The British government responded to public alarm by wheeling out June Raine, the terrifyingly weak infant school headmistress sound-alike who heads the UK MHRA drug 'regulator'. The idea that she would stand up to Big Pharma and government pressure is laughable and she told us that all was well in the same way that she did when allowing untested, never-used-on-humans-before, genetically-manipulating 'vaccines' to be exposed to the public in the first place. Mass lying is the new normal of the 'Covid' era. The MHRA later said 30 cases of rare blood clots had by then been connected with the AstraZeneca 'vaccine' (that means a lot more in reality) while stressing that the benefits of the jab in preventing 'Covid-19' outweighed any risks. A more ridiculous and disingenuous statement with callous disregard for human health it is hard to contemplate. Immediately after the mendacious 'all-clears' two hospital workers in Denmark experienced blood clots and cerebral haemorrhaging following the AstraZeneca jab and one died. Top Norwegian health official Pål Andre Holme said the 'vaccine' was the only common factor: 'There is nothing in the patient history of these individuals that can give such a powerful immune response ... I am confident that the antibodies that we have found are the cause, and I see no other explanation than it being the vaccine which triggers it.' Strokes, a clot or bleed in the brain, were clearly associated with the 'vaccine' from word of mouth and whistleblower reports. Similar consequences followed with all these 'vaccines' that we were told were so safe and as the numbers grew by the day it was clear we were witnessing human carnage.

Learning the hard way

A woman interviewed by UKColumn told how her husband suffered dramatic health effects after the vaccine when he'd been in good health all his life. He went from being a little unwell to losing all feeling in his legs and experiencing 'excruciating pain'. Misdiagnosis followed twice at Accident and Emergency (an 'allergy' and 'sciatica') before he was admitted to a neurology ward where doctors said his serious condition had been caused by the

'vaccine'. Another seven 'vaccinated' people were apparently being treated on the same ward for similar symptoms. The woman said he had the 'vaccine' because they believed media claims that it was safe. 'I didn't think the government would give out a vaccine that does this to somebody; I believed they would be bringing out a vaccination that would be safe.' What a tragic way to learn that lesson. Another woman posted that her husband was transporting stroke patients to hospital on almost every shift and when he asked them if they had been 'vaccinated' for 'Covid' they all replied 'yes'. One had a 'massive brain bleed' the day after his second dose. She said her husband reported the 'just been vaccinated' information every time to doctors in A and E only for them to ignore it, make no notes and appear annoyed that it was even mentioned. This particular report cannot be verified, but it expresses a common theme that confirms the monumental underreporting of 'vaccine' consequences. Interestingly as the 'vaccines' and their brain blood clot/stroke consequences began to emerge the UK National Health Service began a publicity campaign telling the public what to do in the event of a stroke. A Scottish NHS staff nurse who quit in disgust in March, 2021, said:

I have seen traumatic injuries from the vaccine, they're not getting reported to the yellow card [adverse reaction] scheme, they're treating the symptoms, not asking why, why it's happening. It's just treating the symptoms and when you speak about it you're dismissed like you're crazy, I'm not crazy, I'm not crazy because every other colleague I've spoken to is terrified to speak out, they've had enough.

Videos appeared on the Internet of people uncontrollably shaking after the 'vaccine' with no control over muscles, limbs and even their face. A Scottish mother broke out in a severe rash all over her body almost immediately after she was given the AstraZeneca 'vaccine'. The pictures were horrific. Leigh King, a 41-year-old hairdresser from Lanarkshire said: 'Never in my life was I prepared for what I was about to experience ... My skin was so sore and constantly hot ... I have never felt pain like this ...' But don't you worry, the 'vaccine' is perfectly safe. Then there has been the effect on medical

staff who have been pressured to have the 'vaccine' by psychopathic 'health' authorities and government. A London hospital consultant who gave the name K. Polyakova wrote this to the *British Medical Journal* or *BMJ*:

I am currently struggling with ... the failure to report the reality of the morbidity caused by our current vaccination program within the health service and staff population. The levels of sickness after vaccination is unprecedented and staff are getting very sick and some with neurological symptoms which is having a huge impact on the health service function. Even the young and healthy are off for days, some for weeks, and some requiring medical treatment. Whole teams are being taken out as they went to get vaccinated together.

Mandatory vaccination in this instance is stupid, unethical and irresponsible when it comes to protecting our staff and public health. We are in the voluntary phase of vaccination, and encouraging staff to take an unlicensed product that is impacting on their immediate health ... it is clearly stated that these vaccine products do not offer immunity or stop transmission. In which case why are we doing it?

Not to protect health that's for sure. Medical workers are lauded by governments for agenda reasons when they couldn't give a toss about them any more than they can for the population in general. Schools across America faced the same situation as they closed due to the high number of teachers and other staff with bad reactions to the Pfizer/BioNTech, Moderna, and Johnson & Johnson 'Covid vaccines' all of which were linked to death and serious adverse effects. The *BMJ* took down the consultant's comments pretty quickly on the grounds that they were being used to spread 'disinformation'. They were exposing the truth about the 'vaccine' was the real reason. The cover-up is breathtaking.

Hiding the evidence

The scale of the 'vaccine' death cover-up worldwide can be confirmed by comparing official figures with the personal experience of the public. I heard of many people in my community who died immediately or soon after the vaccine that would never appear in the media or even likely on the official totals of 'vaccine' fatalities and adverse reactions when only about ten percent are estimated to be

reported and I have seen some estimates as low as one percent in a Harvard study. In the UK alone by April 29th, 2021, some 757,654 adverse reactions had been officially reported from the Pfizer/BioNTech, Oxford/AstraZeneca and Moderna 'vaccines' with more than a thousand deaths linked to jabs and that means an estimated ten times this number in reality from a ten percent reporting rate percentage. That's seven million adverse reactions and 10,000 potential deaths and a one percent reporting rate would be ten times *those* figures. In 1976 the US government pulled the swine flu vaccine after 53 deaths. The UK data included a combined 10,000 eye disorders from the 'Covid vaccines' with more than 750 suffering visual impairment or blindness and again multiply by the estimated reporting percentages. As 'Covid cases' officially fell hospitals virtually empty during the 'Covid crisis' began to fill up with a range of other problems in the wake of the 'vaccine' rollout. The numbers across America have also been catastrophic. Deaths linked to *all* types of vaccine increased by 6,000 *percent* in the first quarter of 2021 compared with 2020. A 39-year-old woman from Ogden, Utah, died four days after receiving a second dose of Moderna's 'Covid vaccine' when her liver, heart and kidneys all failed despite the fact that she had no known medical issues or conditions. Her family sought an autopsy, but Dr Erik Christensen, Utah's chief medical examiner, said proving vaccine injury as a cause of death almost never happened. He could think of only one instance where an autopsy would name a vaccine as the official cause of death and that would be anaphylaxis where someone received a vaccine and died almost instantaneously. 'Short of that, it would be difficult for us to definitively say this is the vaccine,' Christensen said. If that is true this must be added to the estimated ten percent (or far less) reporting rate of vaccine deaths and serious reactions and the conclusion can only be that vaccine deaths and serious reactions – including these 'Covid' potions' – are phenomenally understated in official figures. The same story can be found everywhere. Endless accounts of deaths and serious reactions among the public, medical

and care home staff while official figures did not even begin to reflect this.

Professional script-reader Dr David Williams, a 'top public-health official' in Ontario, Canada, insulted our intelligence by claiming only four serious adverse reactions and no deaths from the more than 380,000 vaccine doses then given. This bore no resemblance to what people knew had happened in their own circles and we had Dirk Huyer in charge of getting millions vaccinated in Ontario while at the same time he was Chief Coroner for the province investigating causes of death including possible death from the vaccine. An aide said he had stepped back from investigating deaths, but evidence indicated otherwise. Rosemary Frei, who secured a Master of Science degree in molecular biology at the Faculty of Medicine at Canada's University of Calgary before turning to investigative journalism, was one who could see that official figures for 'vaccine' deaths and reactions made no sense. She said that doctors seldom reported adverse events and when people got really sick or died after getting a vaccination they would attribute that to anything except the vaccines. It had been that way for years and anyone who wondered aloud whether the 'Covid vaccines' or other shots cause harm is immediately branded as 'anti-vax' and 'anti-science'. This was 'career-threatening' for health professionals. Then there was the huge pressure to support the push to 'vaccinate' billions in the quickest time possible. Frei said:

So that's where we're at today. More than half a million vaccine doses have been given to people in Ontario alone. The rush is on to vaccinate all 15 million of us in the province by September. And the mainstream media are screaming for this to be sped up even more. That all adds up to only a very slim likelihood that we're going to be told the truth by officials about how many people are getting sick or dying from the vaccines.

What is true of Ontario is true of everywhere.

They KNEW – and still did it

The authorities knew what was going to happen with multiple deaths and adverse reactions. The UK government's Gates-funded

and Big Pharma-dominated Medicines and Healthcare products Regulatory Agency (MHRA) hired a company to employ AI in compiling the projected reactions to the 'vaccine' that would otherwise be uncountable. The request for applications said: 'The MHRA urgently seeks an Artificial Intelligence (AI) software tool to process the expected high volume of Covid-19 vaccine Adverse Drug Reaction ...' This was from the agency, headed by the disingenuous June Raine, that gave the 'vaccines' emergency approval and the company was hired before the first shot was given. 'We are going to kill and maim you – is that okay?' 'Oh, yes, perfectly fine – I'm very grateful, thank you, doctor.' The range of 'Covid vaccine' adverse reactions goes on for page after page in the MHRA criminally underreported 'Yellow Card' system and includes affects to eyes, ears, skin, digestion, blood and so on. Raine's MHRA amazingly claimed that the 'overall safety experience ... is so far as expected from the clinical trials'. The death, serious adverse effects, deafness and blindness were *expected*? When did they ever mention that? If these human tragedies were expected then those that gave approval for the use of these 'vaccines' must be guilty of crimes against humanity including murder – a definition of which is 'killing a person with malice aforethought or with recklessness manifesting extreme indifference to the value of human life.' People involved at the MHRA, the CDC in America and their equivalent around the world must go before Nuremberg trials to answer for their callous inhumanity. We are only talking here about the immediate effects of the 'vaccine'. The longer-term impact of the DNA synthetic manipulation is the main reason they are so hysterically desperate to inoculate the entire global population in the shortest possible time.

Africa and the developing world are a major focus for the 'vaccine' depopulation agenda and a mass vaccination sales-pitch is underway thanks to caring people like the Rockefellers and other Cult assets. The Rockefeller Foundation, which pre-empted the 'Covid pandemic' in a document published in 2010 that 'predicted' what happened a decade later, announced an initial \$34.95 million grant in February, 2021, 'to ensure more equitable access to Covid-19

testing and vaccines' among other things in Africa in collaboration with '24 organizations, businesses, and government agencies'. The pan-Africa initiative would focus on 10 countries: Burkina Faso, Ethiopia, Ghana, Kenya, Nigeria, Rwanda, South Africa, Tanzania, Uganda, and Zambia'. Rajiv Shah, President of the Rockefeller Foundation and former administrator of CIA-controlled USAID, said that if Africa was not mass-vaccinated (to change the DNA of its people) it was a 'threat to all of humanity' and not fair on Africans. When someone from the Rockefeller Foundation says they want to do something to help poor and deprived people and countries it is time for a belly-laugh. They are doing this out of the goodness of their 'heart' because 'vaccinating' the entire global population is what the 'Covid' hoax set out to achieve. Official 'decolonisation' of Africa by the Cult was merely a prelude to financial colonisation on the road to a return to physical colonisation. The 'vaccine' is vital to that and the sudden and convenient death of the 'Covid' sceptic president of Tanzania can be seen in its true light. A lot of people in Africa are aware that this is another form of colonisation and exploitation and they need to stand their ground.

The 'vaccine is working' scam

A potential problem for the Cult was that the 'vaccine' is meant to change human DNA and body messaging and not to protect anyone from a 'virus' never shown to exist. The vaccine couldn't work because it was not designed to work and how could they make it *appear* to be working so that more people would have it? This was overcome by lowering the amplification rate of the PCR test to produce fewer 'cases' and therefore fewer 'deaths'. Some of us had been pointing out since March, 2020, that the amplification rate of the test not testing for the 'virus' had been made artificially high to generate positive tests which they could call 'cases' to justify lockdowns. The World Health Organization recommended an absurdly high 45 amplification cycles to ensure the high positives required by the Cult and then remained silent on the issue until January 20th, 2021 – Biden's Inauguration Day. This was when the

'vaccinations' were seriously underway and on that day the WHO recommended after discussions with America's CDC that laboratories *lowered their testing amplification*. Dr David Samadi, a certified urologist and health writer, said the WHO was encouraging all labs to reduce their cycle count for PCR tests. He said the current cycle was much too high and was 'resulting in any particle being declared a positive case'. Even one mainstream news report I saw said this meant the number of 'Covid' infections may have been 'dramatically inflated'. Oh, just a little bit. The CDC in America issued new guidance to laboratories in April, 2021, to use 28 cycles *but only for 'vaccinated' people*. The timing of the CDC/WHO interventions were cynically designed to make it appear the 'vaccines' were responsible for falling cases and deaths when the real reason can be seen in the following examples. New York's state lab, the Wadsworth Center, identified 872 positive tests in July, 2020, based on a threshold of 40 cycles. When the figure was lowered to 35 cycles 43 percent of the 872 were no longer 'positives'. At 30 cycles the figure was 63 percent. A Massachusetts lab found that between 85 to 90 percent of people who tested positive in July with a cycle threshold of 40 would be negative at 30 cycles, Ashish Jha, MD, director of the Harvard Global Health Institute, said: 'I'm really shocked that it could be that high ... Boy, does it really change the way we need to be thinking about testing.' I'm shocked that I could see the obvious in the spring of 2020, with no medical background, and most medical professionals still haven't worked it out. No, that's not shocking – it's terrifying.

Three weeks after the WHO directive to lower PCR cycles the London *Daily Mail* ran this headline: 'Why ARE Covid cases plummeting? New infections have fallen 45% in the US and 30% globally in the past 3 weeks but experts say vaccine is NOT the main driver because only 8% of Americans and 13% of people worldwide have received their first dose.' They acknowledged that the drop could not be attributed to the 'vaccine', but soon this morphed throughout the media into the 'vaccine' has caused cases and deaths to fall when it was the PCR threshold. In December, 2020, there was

chaos at English Channel ports with truck drivers needing negative 'Covid' tests before they could board a ferry home for Christmas. The government wanted to remove the backlog as fast as possible and they brought in troops to do the 'testing'. Out of 1,600 drivers just 36 tested positive and the rest were given the all clear to cross the Channel. I guess the authorities thought that 36 was the least they could get away with without the unquestioning catching on. The amplification trick which most people believed in the absence of information in the mainstream applied more pressure on those refusing the 'vaccine' to succumb when it 'obviously worked'. The truth was the exact opposite with deaths in care homes soaring with the 'vaccine' and in Israel the term used was 'skyrocket'. A re-analysis of published data from the Israeli Health Ministry led by Dr Hervé Seligmann at the Medicine Emerging Infectious and Tropical Diseases at Aix-Marseille University found that Pfizer's 'Covid vaccine' killed 'about 40 times more [elderly] people than the disease itself would have killed' during a five-week vaccination period and *260 times* more younger people than would have died from the 'virus' even according to the manipulated 'virus' figures. Dr Seligmann and his co-study author, Haim Yativ, declared after reviewing the Israeli 'vaccine' death data: 'This is a new Holocaust.'

Then, in mid-April, 2021, after vast numbers of people worldwide had been 'vaccinated', the story changed with clear coordination. The UK government began to prepare the ground for more future lockdowns when Nuremberg-destined Boris Johnson told yet another whopper. He said that cases had fallen because of *lockdowns* not 'vaccines'. Lockdowns are irrelevant when *there is no 'virus'* and the test and fraudulent death certificates are deciding the number of 'cases' and 'deaths'. Study after study has shown that lockdowns don't work and instead kill and psychologically destroy people. Meanwhile in the United States Anthony Fauci and Rochelle Walensky, the ultra-Zionist head of the CDC, peddled the same line. More lockdown was the answer and not the 'vaccine', a line repeated on cue by the moron that is Canadian Prime Minister Justin Trudeau. Why all the hysteria to get everyone 'vaccinated' if lockdowns and

not 'vaccines' made the difference? None of it makes sense on the face of it. Oh, but it does. The Cult wants lockdowns *and* the 'vaccine' and if the 'vaccine' is allowed to be seen as the total answer lockdowns would no longer be justified when there are still livelihoods to destroy. 'Variants' and renewed upward manipulation of PCR amplification are planned to instigate never-ending lockdown *and* more 'vaccines'.

You *must* have it – we're desperate

Israel, where the Jewish and Arab population are ruled by the Sabbatian Cult, was the front-runner in imposing the DNA-manipulating 'vaccine' on its people to such an extent that Jewish refusers began to liken what was happening to the early years of Nazi Germany. This would seem to be a fantastic claim. Why would a government of Jewish people be acting like the Nazis did? If you realise that the Sabbatian Cult was behind the Nazis and that Sabbatians hate Jews the pieces start to fit and the question of why a 'Jewish' government would treat Jews with such callous disregard for their lives and freedom finds an answer. Those controlling the government of Israel *aren't Jewish* – they're Sabbatian. Israeli lawyer Tamir Turgal was one who made the Nazi comparison in comments to German lawyer Reiner Fuellmich who is leading a class action lawsuit against the psychopaths for crimes against humanity. Turgal described how the Israeli government was vaccinating children and pregnant women on the basis that there was no evidence that this was dangerous when they had no evidence that it *wasn't* dangerous either. They just had no evidence. This was medical experimentation and Turgal said this breached the Nuremberg Code about medical experimentation and procedures requiring informed consent and choice. Think about that. A Nuremberg Code developed because of Nazi experimentation on Jews and others in concentration camps by people like the evil-beyond-belief Josef Mengele is being breached by the *Israeli* government; but when you know that it's a *Sabbatian* government along with its intelligence and military agencies like Mossad, Shin Bet and the Israeli Defense Forces, and that Sabbatians

were the force behind the Nazis, the kaleidoscope comes into focus. What have we come to when Israeli Jews are suing their government for violating the Nuremberg Code by essentially making Israelis subject to a medical experiment using the controversial 'vaccines'? It's a shocker that this has to be done in the light of what happened in Nazi Germany. The Anshe Ha-Emet, or 'People of the Truth', made up of Israeli doctors, lawyers, campaigners and public, have launched a lawsuit with the International Criminal Court. It says:

When the heads of the Ministry of Health as well as the prime minister presented the vaccine in Israel and began the vaccination of Israeli residents, the vaccinated were not advised, that, in practice, they are taking part in a medical experiment and that their consent is required for this under the Nuremberg Code.

The irony is unbelievable, but easily explained in one word: Sabbatians. The foundation of Israeli 'Covid' apartheid is the 'green pass' or 'green passport' which allows Jews and Arabs who have had the DNA-manipulating 'vaccine' to go about their lives – to work, fly, travel in general, go to shopping malls, bars, restaurants, hotels, concerts, gyms, swimming pools, theatres and sports venues, while non-'vaccinated' are banned from all those places and activities. Israelis have likened the 'green pass' to the yellow stars that Jews in Nazi Germany were forced to wear – the same as the yellow stickers that a branch of UK supermarket chain Morrisons told exempt mask-wearers they had to display when shopping. How very sensitive. The Israeli system is blatant South African-style apartheid on the basis of compliance or non-compliance to fascism rather than colour of the skin. How appropriate that the Sabbatian Israeli government was so close to the pre-Mandela apartheid regime in Pretoria. The Sabbatian-instigated 'vaccine passport' in Israel is planned for everywhere. Sabbatians struck a deal with Pfizer that allowed them to lead the way in the percentage of a national population infused with synthetic material and the result was catastrophic. Israeli freedom activist Shai Dannon told me how chairs were appearing on beaches that said 'vaccinated only'. Health Minister Yuli Edelstein said that anyone unwilling or unable to get

the jabs that 'confer immunity' will be 'left behind'. The man's a liar. Not even the makers claim the 'vaccines' confer immunity. When you see those figures of 'vaccine' deaths these psychopaths were saying that you must take the chance the 'vaccine' will kill you or maim you while knowing it will change your DNA or lockdown for you will be permanent. That's fascism. The Israeli parliament passed a law to allow personal information of the non-vaccinated to be shared with local and national authorities for three months. This was claimed by its supporters to be a way to 'encourage' people to be vaccinated. Hadas Ziv from Physicians for Human Rights described this as a 'draconian law which crushed medical ethics and the patient rights'. But that's the idea, the Sabbatians would reply.

Your papers, please

Sabbatian Israel was leading what has been planned all along to be a global 'vaccine pass' called a 'green passport' without which you would remain in permanent lockdown restriction and unable to do anything. This is how badly – *desperately* – the Cult is to get everyone 'vaccinated'. The term and colour 'green' was not by chance and related to the psychology of fusing the perception of the green climate hoax with the 'Covid' hoax and how the 'solution' to both is the same Great Reset. Lying politicians, health officials and psychologists denied there were any plans for mandatory vaccinations or restrictions based on vaccinations, but they knew that was exactly what was meant to happen with governments of all countries reaching agreements to enforce a global system. 'Free' Denmark and 'free' Sweden unveiled digital vaccine certification. Cyprus, Czech Republic, Estonia, Greece, Hungary, Iceland, Italy, Poland, Portugal, Slovakia, and Spain have all committed to a vaccine passport system and the rest including the whole of the EU would follow. The satanic UK government will certainly go this way despite mendacious denials and at the time of writing it is trying to manipulate the public into having the 'vaccine' so they could go abroad on a summer holiday. How would that work without something to prove you had the synthetic toxicity injected into you?

Documents show that the EU's European Commission was moving towards 'vaccine certificates' in 2018 and 2019 before the 'Covid' hoax began. They knew what was coming. Abracadabra – Ursula von der Leyen, the German President of the Commission, announced in March, 2021, an EU 'Digital Green Certificate' – green again – to track the public's 'Covid status'. The passport sting is worldwide and the Far East followed the same pattern with South Korea ruling that only those with 'vaccination' passports – again the *green* pass – would be able to 'return to their daily lives'.

Bill Gates has been preparing for this 'passport' with other Cult operatives for years and beyond the paper version is a Gates-funded 'digital tattoo' to identify who has been vaccinated and who hasn't. The 'tattoo' is reported to include a substance which is externally readable to confirm who has been vaccinated. This is a bio-luminous light-generating enzyme (think fireflies) called ... *Luciferase*. Yes, named after the Cult 'god' Lucifer the 'light bringer' of whom more to come. Gates said he funded the readable tattoo to ensure children in the developing world were vaccinated and no one was missed out. He cares so much about poor kids as we know. This was just the cover story to develop a vaccine tagging system for everyone on the planet. Gates has been funding the ID2020 'alliance' to do just that in league with other lovely people at Microsoft, GAVI, the Rockefeller Foundation, Accenture and IDEO.org. He said in interviews in March, 2020, before any 'vaccine' publicly existed, that the world must have a globalised digital certificate to track the 'virus' and who had been vaccinated. Gates knew from the start that the mRNA vaccines were coming and when they would come and that the plan was to tag the 'vaccinated' to marginalise the intelligent and stop them doing anything including travel. Evil just doesn't suffice. Gates was exposed for offering a \$10 million bribe to the Nigerian House of Representatives to invoke compulsory 'Covid' vaccination of all Nigerians. Sara Cunial, a member of the Italian Parliament, called Gates a 'vaccine criminal'. She urged the Italian President to hand him over to the International Criminal Court for crimes against

humanity and condemned his plans to 'chip the human race' through ID2020.

You know it's a long-planned agenda when war criminal and Cult gofer Tony Blair is on the case. With the scale of arrogance only someone as dark as Blair can muster he said: 'Vaccination in the end is going to be your route to liberty.' Blair is a disgusting piece of work and he confirms that again. The media has given a lot of coverage to a bloke called Charlie Mullins, founder of London's biggest independent plumbing company, Pimlico Plumbers, who has said he won't employ anyone who has not been vaccinated or have them go to any home where people are not vaccinated. He said that if he had his way no one would be allowed to walk the streets if they have not been vaccinated. Gates was cheering at the time while I was alerting the white coats. The plan is that people will qualify for 'passports' for having the first two doses and then to keep it they will have to have all the follow ups and new ones for invented 'variants' until human genetics is transformed and many are dead who can't adjust to the changes. Hollywood celebrities – the usual propaganda stunt – are promoting something called the WELL Health-Safety Rating to verify that a building or space has 'taken the necessary steps to prioritize the health and safety of their staff, visitors and other stakeholders'. They included Lady Gaga, Jennifer Lopez, Michael B. Jordan, Robert DeNiro, Venus Williams, Wolfgang Puck, Deepak Chopra and 17th Surgeon General Richard Carmona. Yawn. WELL Health-Safety has big connections with China. Parent company Delos is headed by former Goldman Sachs partner Paul Scialla. This is another example – and we will see so many others – of using the excuse of 'health' to dictate the lives and activities of the population. I guess one confirmation of the 'safety' of buildings is that only 'vaccinated' people can go in, right?

Electronic concentration camps

I wrote decades ago about the plans to restrict travel and here we are for those who refuse to bow to tyranny. This can be achieved in one go with air travel if the aviation industry makes a blanket decree.

The 'vaccine' and guaranteed income are designed to be part of a global version of China's social credit system which tracks behaviour 24/7 and awards or deletes 'credits' based on whether your behaviour is supported by the state or not. I mean your entire lifestyle – what you do, eat, say, everything. Once your credit score falls below a certain level consequences kick in. In China tens of millions have been denied travel by air and train because of this. All the locations and activities denied to refusers by the 'vaccine' passports will be included in one big mass ban on doing almost anything for those that don't bow their head to government. It's beyond fascist and a new term is required to describe its extremes – I guess fascist technocracy will have to do. The way the Chinese system of technological – technocratic – control is sweeping the West can be seen in the Los Angeles school system and is planned to be expanded worldwide. Every child is required to have a 'Covid'-tracking app scanned daily before they can enter the classroom. The so-called Daily Pass tracking system is produced by Gates' Microsoft which I'm sure will shock you rigid. The pass will be scanned using a barcode (one step from an inside-the-body barcode) and the information will include health checks, 'Covid' tests and vaccinations. Entry codes are for one specific building only and access will only be allowed if a student or teacher has a negative test with a test not testing for the 'virus', has no symptoms of anything alleged to be related to 'Covid' (symptoms from a range of other illness), and has a temperature under 100 degrees. No barcode, no entry, is planned to be the case for everywhere and not only schools.

Kids are being psychologically prepared to accept this as 'normal' their whole life which is why what they can impose in schools is so important to the Cult and its gofers. Long-time American freedom campaigner John Whitehead of the Rutherford Institute was not exaggerating when he said: 'Databit by databit, we are building our own electronic concentration camps.' Canada under its Cult gofer prime minister Justin Trudeau has taken a major step towards the real thing with people interned against their will if they test positive with a test not testing for the 'virus' when they arrive at a Canadian

airport. They are jailed in internment hotels often without food or water for long periods and with many doors failing to lock there have been sexual assaults. The interned are being charged sometimes \$2,000 for the privilege of being abused in this way. Trudeau is fully on board with the Cult and says the 'Covid pandemic' has provided an opportunity for a global 'reset' to permanently change Western civilisation. His number two, Deputy Prime Minister Chrystia Freeland, is a trustee of the World Economic Forum and a Rhodes Scholar. The Trudeau family have long been servants of the Cult. See *The Biggest Secret* and Cathy O'Brien's book *Trance-Formation of America* for the horrific background to Trudeau's father Pierre Trudeau another Canadian prime minister. Hide your fascism behind the façade of a heart-on-the-sleeve liberal. It's a well-honed Cult technique.

What can the 'vaccine' really do?

We have a 'virus' never shown to exist and 'variants' of the 'virus' that have also never been shown to exist except, like the 'original', as computer-generated fictions. Even if you believe there's a 'virus' the 'case' to 'death' rate is in the region of 0.23 to 0.15 percent and those 'deaths' are concentrated among the very old around the same average age that people die anyway. In response to this lack of threat (in truth none) psychopaths and idiots, knowingly and unknowingly answering to Gates and the Cult, are seeking to 'vaccinate' every man, woman and child on Planet Earth. Clearly the 'vaccine' is not about 'Covid' – none of this ever has been. So what is it all about *really*? Why the desperation to infuse genetically-manipulating synthetic material into everyone through mRNA fraudulent 'vaccines' with the intent of doing this over and over with the excuses of 'variants' and other 'virus' inventions? Dr Sherri Tenpenny, an osteopathic medical doctor in the United States, has made herself an expert on vaccines and their effects as a vehement campaigner against their use. Tenpenny was board certified in emergency medicine, the director of a level two trauma centre for 12 years, and moved to Cleveland in 1996 to start an integrative

medicine practice which has treated patients from all 50 states and some 17 other countries. Weaning people off pharmaceutical drugs is a speciality.

She became interested in the consequences of vaccines after attending a meeting at the National Vaccine Information Center in Washington DC in 2000 where she 'sat through four days of listening to medical doctors and scientists and lawyers and parents of vaccine injured kids' and asked: 'What's going on?' She had never been vaccinated and never got ill while her father was given a list of vaccines to be in the military and was 'sick his entire life'. The experience added to her questions and she began to examine vaccine documents from the Centers for Disease Control (CDC). After reading the first one, the 1998 version of *The General Recommendations of Vaccination*, she thought: 'This is it?' The document was poorly written and bad science and Tenpenny began 20 years of research into vaccines that continues to this day. She began her research into 'Covid vaccines' in March, 2020, and she describes them as 'deadly'. For many, as we have seen, they already have been. Tenpenny said that in the first 30 days of the 'vaccine' rollout in the United States there had been more than 40,000 adverse events reported to the vaccine adverse event database. A document had been delivered to her the day before that was 172 pages long. 'We have over 40,000 adverse events; we have over 3,100 cases of [potentially deadly] anaphylactic shock; we have over 5,000 neurological reactions.' Effects ranged from headaches to numbness, dizziness and vertigo, to losing feeling in hands or feet and paraesthesia which is when limbs 'fall asleep' and people have the sensation of insects crawling underneath their skin. All this happened in the first 30 days and remember that only about *ten percent* (or far less) of adverse reactions and vaccine-related deaths are estimated to be officially reported. Tenpenny said:

So can you think of one single product in any industry, any industry, for as long as products have been made on the planet that within 30 days we have 40,000 people complaining of side effects that not only is still on the market but ... we've got paid actors telling us how great

they are for getting their vaccine. We're offering people \$500 if they will just get their vaccine and we've got nurses and doctors going; 'I got the vaccine, I got the vaccine'.

Tenpenny said they were not going to be 'happy dancing folks' when they began to suffer Bell's palsy (facial paralysis), neuropathies, cardiac arrhythmias and autoimmune reactions that kill through a blood disorder. 'They're not going to be so happy, happy then, but we're never going to see pictures of those people' she said. Tenpenny described the 'vaccine' as 'a well-designed killing tool'.

No off-switch

Bad as the initial consequences had been Tenpenny said it would be maybe 14 months before we began to see the 'full ravage' of what is going to happen to the 'Covid vaccinated' with full-out consequences taking anything between two years and 20 years to show. You can understand why when you consider that variations of the 'Covid vaccine' use mRNA (messenger RNA) to in theory activate the immune system to produce protective antibodies without using the actual 'virus'. How can they when it's a computer program and they've never isolated what they claim is the 'real thing'? Instead they use *synthetic* mRNA. They are inoculating synthetic material into the body which through a technique known as the Trojan horse is absorbed into cells to change the nature of DNA. Human DNA is changed by an infusion of messenger RNA and with each new 'vaccine' of this type it is changed even more. Say so and you are banned by Cult Internet platforms. The contempt the contemptuous Mark Zuckerberg has for the truth and human health can be seen in an internal Facebook video leaked to the Project Veritas investigative team in which he said of the 'Covid vaccines': '... I share some caution on this because we just don't know the long term side-effects of basically modifying people's DNA and RNA.' At the same time this disgusting man's Facebook was censoring and banning anyone saying exactly the same. He must go before a Nuremberg trial for crimes against humanity when he *knows* that he

is censoring legitimate concerns and denying the right of informed consent on behalf of the Cult that owns him. People have been killed and damaged by the very 'vaccination' technique he cast doubt on himself when they may not have had the 'vaccine' with access to information that he denied them. The plan is to have at least annual 'Covid vaccinations', add others to deal with invented 'variants', and change all other vaccines into the mRNA system. Pfizer executives told shareholders at a virtual Barclays Global Healthcare Conference in March, 2021, that the public may need a third dose of 'Covid vaccine', plus regular yearly boosters and the company planned to hike prices to milk the profits in a 'significant opportunity for our vaccine'. These are the professional liars, cheats and opportunists who are telling you their 'vaccine' is safe. Given this volume of mRNA planned to be infused into the human body and its ability to then replicate we will have a transformation of human genetics from biological to synthetic biological – exactly the long-time Cult plan for reasons we'll see – and many will die. Sherri Tenpenny said of this replication:

It's like having an on-button but no off-button and that whole mechanism ... they actually give it a name and they call it the Trojan horse mechanism, because it allows that [synthetic] virus and that piece of that [synthetic] virus to get inside of your cells, start to replicate and even get inserted into other parts of your DNA as a Trojan-horse.

Ask the overwhelming majority of people who have the 'vaccine' what they know about the contents and what they do and they would reply: 'The government says it will stop me getting the virus.' Governments give that false impression on purpose to increase take-up. You can read Sherri Tenpenny's detailed analysis of the health consequences in her blog at [Vaxxter.com](https://vaxxter.com), but in summary these are some of them. She highlights the statement by Bill Gates about how human beings can become their own 'vaccine manufacturing machine'. The man is insane. ['Vaccine'-generated] 'antibodies' carry synthetic messenger RNA into the cells and the damage starts, Tenpenny contends, and she says that lungs can be adversely affected through varying degrees of pus and bleeding which

obviously affects breathing and would be dubbed 'Covid-19'. Even more sinister was the impact of 'antibodies' on macrophages, a white blood cell of the immune system. They consist of Type 1 and Type 2 which have very different functions. She said Type 1 are 'hyper-vigilant' white blood cells which 'gobble up' bacteria etc. However, in doing so, this could cause inflammation and in extreme circumstances be fatal. She says these affects are mitigated by Type 2 macrophages which kick in to calm down the system and stop it going rogue. They clear up dead tissue debris and reduce inflammation that the Type 1 'fire crews' have caused. Type 1 kills the infection and Type 2 heals the damage, she says. This is her punchline with regard to 'Covid vaccinations': She says that mRNA 'antibodies' block Type 2 macrophages by attaching to them and deactivating them. This meant that when the Type 1 response was triggered by infection there was nothing to stop that getting out of hand by calming everything down. There's an on-switch, but no off-switch, she says. What follows can be 'over and out, see you when I see you'.

Genetic suicide

Tenpenny also highlights the potential for autoimmune disease – the body attacking itself – which has been associated with vaccines since they first appeared. Infusing a synthetic foreign substance into cells could cause the immune system to react in a panic believing that the body is being overwhelmed by an invader (it is) and the consequences can again be fatal. There is an autoimmune response known as a 'cytokine storm' which I have likened to a homeowner panicked by an intruder and picking up a gun to shoot randomly in all directions before turning the fire on himself. The immune system unleashes a storm of inflammatory response called cytokines to a threat and the body commits hara-kiri. The lesson is that you mess with the body's immune response at your peril and these 'vaccines' seriously – fundamentally – mess with immune response. Tenpenny refers to a consequence called anaphylactic shock which is a severe and highly dangerous allergic reaction when the immune system

floods the body with chemicals. She gives the example of having a bee sting which primes the immune system and makes it sensitive to those chemicals. When people are stung again maybe years later the immune response can be so powerful that it leads to anaphylactic shock. Tenpenny relates this 'shock' with regard to the 'Covid vaccine' to something called polyethylene glycol or PEG. Enormous numbers of people have become sensitive to this over decades of use in a whole range of products and processes including food, drink, skin creams and 'medicine'. Studies have claimed that some 72 percent of people have antibodies triggered by PEG compared with two percent in the 1960s and allergic hypersensitive reactions to this become a gathering cause for concern. Tenpenny points out that the 'mRNA vaccine' is coated in a 'bubble' of polyethylene glycol which has the potential to cause anaphylactic shock through immune sensitivity. Many reports have appeared of people reacting this way after having the 'Covid vaccine'. What do we think is going to happen as humanity has more and more of these 'vaccines'? Tenpenny said: 'All these pictures we have seen with people with these rashes ... these weepy rashes, big reactions on their arms and things like that – it's an acute allergic reaction most likely to the polyethylene glycol that you've been previously primed and sensitised to.'

Those who have not studied the conspiracy and its perpetrators at length might think that making the population sensitive to PEG and then putting it in these 'vaccines' is just a coincidence. It is not. It is instead testament to how carefully and coldly-planned current events have been and the scale of the conspiracy we are dealing with. Tenpenny further explains that the 'vaccine' mRNA procedure can breach the blood-brain barrier which protects the brain from toxins and other crap that will cause malfunction. In this case they could make two proteins corrupt brain function to cause Amyotrophic lateral sclerosis (ALS), a progressive nervous system disease leading to loss of muscle control, and frontal lobe degeneration – Alzheimer's and dementia. Immunologist J. Bart Classon published a paper connecting mRNA 'vaccines' to prion

disease which can lead to Alzheimer's and other forms of neurogenerative disease while others have pointed out the potential to affect the placenta in ways that make women infertile. This will become highly significant in the next chapter when I will discuss other aspects of this non-vaccine that relate to its nanotechnology and transmission from the injected to the uninjected.

Qualified in idiocy

Tenpenny describes how research has confirmed that these 'vaccine'-generated antibodies can interact with a range of other tissues in the body and attack many other organs including the lungs. 'This means that if you have a hundred people standing in front of you that all got this shot they could have a hundred different symptoms.' Anyone really think that Cult gofers like the Queen, Tony Blair, Christopher Whitty, Anthony Fauci, and all the other psychopaths have really had this 'vaccine' in the pictures we've seen? Not a bloody chance. Why don't doctors all tell us about all these dangers and consequences of the 'Covid vaccine'? Why instead do they encourage and pressure patients to have the shot? Don't let's think for a moment that doctors and medical staff can't be stupid, lazy, and psychopathic and that's without the financial incentives to give the jab. Tenpenny again:

Some people are going to die from the vaccine directly but a large number of people are going to start to get horribly sick and get all kinds of autoimmune diseases 42 days to maybe a year out. What are they going to do, these stupid doctors who say; 'Good for you for getting that vaccine.' What are they going to say; 'Oh, it must be a mutant, we need to give an extra dose of that vaccine.'

Because now the vaccine, instead of one dose or two doses we need three or four because the stupid physicians aren't taking the time to learn anything about it. If I can learn this sitting in my living room reading a 19 page paper and several others so can they. There's nothing special about me, I just take the time to do it.

Remember how Sara Kayat, the NHS and TV doctor, said that the 'Covid vaccine' would '100 percent prevent hospitalisation and death'. Doctors can be idiots like every other profession and they

should not be worshipped as infallible. They are not and far from it. Behind many medical and scientific 'experts' lies an uninformed prat trying to hide themselves from you although in the 'Covid' era many have failed to do so as with UK narrative-repeating 'TV doctor' Hilary Jones. Pushing back against the minority of proper doctors and scientists speaking out against the 'vaccine' has been the entire edifice of the Cult global state in the form of governments, medical systems, corporations, mainstream media, Silicon Valley, and an army of compliant doctors, medical staff and scientists willing to say anything for money and to enhance their careers by promoting the party line. If you do that you are an 'expert' and if you won't you are an 'anti-vaxxer' and 'Covidiot'. The pressure to be 'vaccinated' is incessant. We have even had reports claiming that the 'vaccine' can help cure cancer and Alzheimer's and make the lame walk. I am waiting for the announcement that it can bring you coffee in the morning and cook your tea. Just as the symptoms of 'Covid' seem to increase by the week so have the miracles of the 'vaccine'. American supermarket giant Kroger Co. offered nearly 500,000 employees in 35 states a \$100 bonus for having the 'vaccine' while donut chain Krispy Kreme promised 'vaccinated' customers a free glazed donut every day for the rest of 2021. Have your DNA changed and you will get a doughnut although we might not have to give you them for long. Such offers and incentives confirm the desperation.

Perhaps the worse vaccine-stunt of them all was UK 'Health' Secretary Matt-the-prat Hancock on live TV after watching a clip of someone being 'vaccinated' when the roll-out began. Hancock faked tears so badly it was embarrassing. Brain-of-Britain Piers Morgan, the lockdown-supporting, 'vaccine' supporting, 'vaccine' passport-supporting, TV host played along with Hancock – 'You're quite emotional about that' he said in response to acting so atrocious it would have been called out at a school nativity which will presumably today include Mary and Jesus in masks, wise men keeping their camels six feet apart, and shepherds under tent arrest. System-serving Morgan tweeted this: 'Love the idea of covid vaccine passports for everywhere: flights, restaurants, clubs, football, gyms,

shops etc. It's time covid-denying, anti-vaxxer loonies had their bullsh*t bluff called & bar themselves from going anywhere that responsible citizens go.' If only I could aspire to his genius. To think that Morgan, who specialises in shouting over anyone he disagrees with, was lauded as a free speech hero when he lost his job after storming off the set of his live show like a child throwing his dolly out of the pram. If he is a free speech hero we are in real trouble. I have no idea what 'bullsh*t' means, by the way, the * throws me completely.

The Cult is desperate to infuse its synthetic DNA-changing concoction into everyone and has been using every lie, trick and intimidation to do so. The question of '*Why?*' we shall now address.

CHAPTER TEN

Human 2.0

I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted –
Alan Turing (1912-1954), the ‘Father of artificial intelligence’

I have been exposing for decades the plan to transform the human body from a biological to a synthetic-biological state. The new human that I will call Human 2.0 is planned to be connected to artificial intelligence and a global AI ‘Smart Grid’ that would operate as one global system in which AI would control everything from your fridge to your heating system to your car to your mind. Humans would no longer be ‘human’, but post-human and sub-human, with their thinking and emotional processes replaced by AI.

What I said sounded crazy and beyond science fiction and I could understand that. To any balanced, rational, mind it *is* crazy. Today, however, that world is becoming reality and it puts the ‘Covid vaccine’ into its true context. Ray Kurzweil is the ultra-Zionist ‘computer scientist, inventor and futurist’ and co-founder of the Singularity University. Singularity refers to the merging of humans with machines or ‘transhumanism’. Kurzweil has said humanity would be connected to the cyber ‘cloud’ in the period of the ever-recurring year of 2030:

Our thinking ... will be a hybrid of biological and non-biological thinking ... humans will be able to extend their limitations and ‘think in the cloud’ ... We’re going to put gateways to the

cloud in our brains ... We're going to gradually merge and enhance ourselves ... In my view, that's the nature of being human – we transcend our limitations. As the technology becomes vastly superior to what we are then the small proportion that is still human gets smaller and smaller and smaller until it's just utterly negligible.

They are trying to sell this end-of-humanity-as-we-know-it as the next stage of 'evolution' when we become super-human and 'like the gods'. They are lying to you. Shocked, eh? The population, and again especially the young, have been manipulated into addiction to technologies designed to enslave them for life. First they induced an addiction to smartphones (holdables); next they moved to technology on the body (wearables); and then began the invasion of the body (implantables). I warned way back about the plan for microchipped people and we are now entering that era. We should not be diverted into thinking that this refers only to chips we can see. Most important are the nanochips known as smart dust, neural dust and nanobots which are far too small to be seen by the human eye. Nanotechnology is everywhere, increasingly in food products, and released into the atmosphere by the geoengineering of the skies funded by Bill Gates to 'shut out the Sun' and 'save the planet from global warming'. Gates has been funding a project to spray millions of tonnes of chalk (calcium carbonate) into the stratosphere over Sweden to 'dim the Sun' and cool the Earth. Scientists warned the move could be disastrous for weather systems in ways no one can predict and opposition led to the Swedish space agency announcing that the 'experiment' would not be happening as planned in the summer of 2021; but it shows where the Cult is going with dimming the impact of the Sun and there's an associated plan to change the planet's atmosphere. Who gives psychopath Gates the right to dictate to the entire human race and dismantle planetary systems? The world will not be safe while this man is at large.

The global warming hoax has made the Sun, like the gas of life, something to fear when both are essential to good health and human survival (more inversion). The body transforms sunlight into vital vitamin D through a process involving ... *cholesterol*. This is the cholesterol we are also told to fear. We are urged to take Big Pharma

statin drugs to reduce cholesterol and it's all systematic. Reducing cholesterol means reducing vitamin D uptake with all the multiple health problems that will cause. At least if you take statins long term it saves the government from having to pay you a pension. The delivery system to block sunlight is widely referred to as chemtrails although these have a much deeper agenda, too. They appear at first to be contrails or condensation trails streaming from aircraft into cold air at high altitudes. Contrails disperse very quickly while chemtrails do not and spread out across the sky before eventually their content falls to earth. Many times I have watched aircraft cross-cross a clear blue sky releasing chemtrails until it looks like a cloudy day. Chemtrails contain many things harmful to humans and the natural world including toxic heavy metals, aluminium (see Alzheimer's) and nanotechnology. Ray Kurzweil reveals the reason without actually saying so: 'Nanobots will infuse all the matter around us with information. Rocks, trees, everything will become these intelligent creatures.' How do you deliver that? *From the sky*. Self-replicating nanobots would connect everything to the Smart Grid. The phenomenon of Morgellons disease began in the chemtrail era and the correlation has led to it being dubbed the 'chemtrail disease'. Self-replicating fibres appear in the body that can be pulled out through the skin. Morgellons fibres continue to grow outside the body and have a form of artificial intelligence. I cover this at greater length in *Phantom Self*.

'Vaccine' operating system

'Covid vaccines' with their self-replicating synthetic material are also designed to make the connection between humanity and Kurzweil's 'cloud'. American doctor and dedicated campaigner for truth, Carrie Madej, an Internal Medicine Specialist in Georgia with more than 20 years medical experience, has highlighted the nanotechnology aspect of the fake 'vaccines'. She explains how one of the components in at least the Moderna and Pfizer synthetic potions are 'lipid nanoparticles' which are 'like little tiny computer bits' – a 'sci-fi substance' known as nanobots and hydrogel which can be 'triggered

at any moment to deliver its payload' and act as 'biosensors'. The synthetic substance had 'the ability to accumulate data from your body like your breathing, your respiration, thoughts and emotions, all kind of things' and each syringe could carry a *million* nanobots:

This substance because it's like little bits of computers in your body, crazy, but it's true, it can do that, [and] obviously has the ability to act through Wi-Fi. It can receive and transmit energy, messages, frequencies or impulses. That issue has never been addressed by these companies. What does that do to the human?

Just imagine getting this substance in you and it can react to things all around you, the 5G, your smart device, your phones, what is happening with that? What if something is triggering it, too, like an impulse, a frequency? We have something completely foreign in the human body.

Madej said her research revealed that electromagnetic (EMF) frequencies emitted by phones and other devices had increased dramatically in the same period of the 'vaccine' rollout and she was seeing more people with radiation problems as 5G and other electromagnetic technology was expanded and introduced to schools and hospitals. She said she was 'floored with the EMF coming off' the devices she checked. All this makes total sense and syncs with my own work of decades when you think that Moderna refers in documents to its mRNA 'vaccine' as an 'operating system':

Recognizing the broad potential of mRNA science, we set out to create an mRNA technology platform that functions very much like an operating system on a computer. It is designed so that it can plug and play interchangeably with different programs. In our case, the 'program' or 'app' is our mRNA drug – the unique mRNA sequence that codes for a protein ...

... Our MRNA Medicines – 'The 'Software Of Life': When we have a concept for a new mRNA medicine and begin research, fundamental components are already in place. Generally, the only thing that changes from one potential mRNA medicine to another is the coding region – the actual genetic code that instructs ribosomes to make protein. Utilizing these instruction sets gives our investigational mRNA medicines a software-like quality. We also have the ability to combine different mRNA sequences encoding for different proteins in a single mRNA investigational medicine.

Who needs a real ‘virus’ when you can create a computer version to justify infusing your operating system into the entire human race on the road to making living, breathing people into cyborgs? What is missed with the ‘vaccines’ is the *digital* connection between synthetic material and the body that I highlighted earlier with the study that hacked a computer with human DNA. On one level the body is digital, based on mathematical codes, and I’ll have more about that in the next chapter. Those who ridiculously claim that mRNA ‘vaccines’ are not designed to change human genetics should explain the words of Dr Tal Zaks, chief medical officer at Moderna, in a 2017 TED talk. He said that over the last 30 years ‘we’ve been living this phenomenal digital scientific revolution, and I’m here today to tell you, that we are actually *hacking the software of life*, and that it’s changing the way we think about prevention and treatment of disease’:

In every cell there’s this thing called messenger RNA, or mRNA for short, that transmits the critical information from the DNA in our genes to the protein, which is really the stuff we’re all made out of. This is the critical information that determines what the cell will do. So we think about it as an operating system. So if you could change that, if you could introduce a line of code, or change a line of code, it turns out, that has profound implications for everything, from the flu to cancer.

Zaks should more accurately have said that this has profound implications for the human genetic code and the nature of DNA. Communications within the body go both ways and not only one. But, hey, no, the ‘Covid vaccine’ will not affect your genetics. Cult fact-checkers say so even though the man who helped to develop the mRNA technique says that it does. Zaks said in 2017:

If you think about what it is we’re trying to do. We’ve taken information and our understanding of that information and how that information is transmitted in a cell, and we’ve taken our understanding of medicine and how to make drugs, and we’re fusing the two. We think of it as information therapy.

I have been writing for decades that the body is an information field communicating with itself and the wider world. This is why

radiation which is information can change the information field of body and mind through phenomena like 5G and change their nature and function. 'Information therapy' means to change the body's information field and change the way it operates. DNA is a receiver-transmitter of information and can be mutated by information like mRNA synthetic messaging. Technology to do this has been ready and waiting in the underground bases and other secret projects to be rolled out when the 'Covid' hoax was played. 'Trials' of such short and irrelevant duration were only for public consumption. When they say the 'vaccine' is 'experimental' that is not true. It may appear to be 'experimental' to those who don't know what's going on, but the trials have already been done to ensure the Cult gets the result it desires. Zaks said that it took decades to sequence the human genome, completed in 2003, but now they could do it in a week. By 'they' he means scientists operating in the public domain. In the secret projects they were sequencing the genome in a week long before even 2003.

Deluge of mRNA

Highly significantly the Moderna document says the guiding premise is that if using mRNA as a medicine works for one disease then it should work for many diseases. They were leveraging the flexibility afforded by their platform and the fundamental role mRNA plays in protein synthesis to pursue mRNA medicines for a broad spectrum of diseases. Moderna is confirming what I was saying through 2020 that multiple 'vaccines' were planned for 'Covid' (and later invented 'variants') and that previous vaccines would be converted to the mRNA system to infuse the body with massive amounts of genetically-manipulating synthetic material to secure a transformation to a synthetic-biological state. The 'vaccines' are designed to kill stunning numbers as part of the long-exposed Cult depopulation agenda and transform the rest. Given this is the goal you can appreciate why there is such hysterical demand for every human to be 'vaccinated' for an alleged 'disease' that has an estimated 'infection' to 'death' ratio of 0.23-0.15 percent. As I write

children are being given the 'vaccine' in trials (their parents are a disgrace) and ever-younger people are being offered the vaccine for a 'virus' that even if you believe it exists has virtually zero chance of harming them. Horrific effects of the 'trials' on a 12-year-old girl were revealed by a family member to be serious brain and gastric problems that included a bowel obstruction and the inability to swallow liquids or solids. She was unable to eat or drink without throwing up, had extreme pain in her back, neck and abdomen, and was paralysed from the waist down which stopped her urinating unaided. When the girl was first taken to hospital doctors said it was all in her mind. She was signed up for the 'trial' by her parents for whom no words suffice. None of this 'Covid vaccine' insanity makes any sense unless you see what the 'vaccine' really is – a body-changer. Synthetic biology or 'SynBio' is a fast-emerging and expanding scientific discipline which includes everything from genetic and molecular engineering to electrical and computer engineering. Synthetic biology is defined in these ways:

- A multidisciplinary area of research that seeks to create new biological parts, devices, and systems, or to redesign systems that are already found in nature.
- The use of a mixture of physical engineering and genetic engineering to create new (and therefore synthetic) life forms.
- An emerging field of research that aims to combine the knowledge and methods of biology, engineering and related disciplines in the design of chemically-synthesized DNA to create organisms with novel or enhanced characteristics and traits (synthetic organisms including humans).

We now have synthetic blood, skin, organs and limbs being developed along with synthetic body parts produced by 3D printers. These are all elements of the synthetic human programme and this comment by Kurzweil's co-founder of the Singularity University,

Peter Diamandis, can be seen in a whole new light with the 'Covid' hoax and the sanctions against those that refuse the 'vaccine':

Anybody who is going to be resisting the progress forward [to transhumanism] is going to be resisting evolution and, fundamentally, they will die out. It's not a matter of whether it's good or bad. It's going to happen.

'Resisting evolution'? What absolute bollocks. The arrogance of these people is without limit. His 'it's going to happen' mantra is another way of saying 'resistance is futile' to break the spirit of those pushing back and we must not fall for it. Getting this genetically-transforming 'vaccine' into everyone is crucial to the Cult plan for total control and the desperation to achieve that is clear for anyone to see. Vaccine passports are a major factor in this and they, too, are a form of resistance is futile. It's NOT. The paper funded by the Rockefeller Foundation for the 2013 'health conference' in China said:

We will interact more with artificial intelligence. The use of robotics, bio-engineering to augment human functioning is already well underway and will advance. Re-engineering of humans into potentially separate and unequal forms through genetic engineering or mixed human-robots raises debates on ethics and equality.

A new demography is projected to emerge after 2030 [that year again] of technologies (robotics, genetic engineering, nanotechnology) producing robots, engineered organisms, 'nanobots' and artificial intelligence (AI) that can self-replicate. Debates will grow on the implications of an impending reality of human designed life.

What is happening today is so long planned. The world army enforcing the will of the world government is intended to be a robot army, not a human one. Today's military and its technologically 'enhanced' troops, pilotless planes and driverless vehicles are just stepping stones to that end. Human soldiers are used as Cult fodder and its time they woke up to that and worked for the freedom of the population instead of their own destruction and their family's destruction – the same with the police. Join us and let's sort this out. The phenomenon of enforce my own destruction is widespread in the 'Covid' era with Woker 'luvvies' in the acting and entertainment

industries supporting 'Covid' rules which have destroyed their profession and the same with those among the public who put signs on the doors of their businesses 'closed due to Covid – stay safe' when many will never reopen. It's a form of masochism and most certainly insanity.

Transgender = transhumanism

When something explodes out of nowhere and is suddenly everywhere it is always the Cult agenda and so it is with the tidal wave of claims and demands that have infiltrated every aspect of society under the heading of 'transgenderism'. The term 'trans' is so 'in' and this is the dictionary definition:

A prefix meaning 'across', 'through', occurring ... in loanwords from Latin, used in particular for denoting movement or conveyance from place to place (transfer; transmit; transplant) or complete change (transform; transmute), or to form adjectives meaning 'crossing', 'on the other side of', or 'going beyond' the place named (transmontane; transnational; trans-Siberian).

Transgender means to go beyond gender and transhuman means to go beyond human. Both are aspects of the Cult plan to transform the human body to a synthetic state with *no gender*. Human 2.0 is not designed to procreate and would be produced technologically with no need for parents. The new human would mean the end of parents and so men, and increasingly women, are being targeted for the deletion of their rights and status. Parental rights are disappearing at an ever-quickenning speed for the same reason. The new human would have no need for men or women when there is no procreation and no gender. Perhaps the transgender movement that appears to be in a permanent state of frenzy might now contemplate on how it is being used. This was never about transgender rights which are only the interim excuse for confusing gender, particularly in the young, on the road to *fusing* gender. Transgender activism is not an end; it is a *means* to an end. We see again the technique of creative destruction in which you destroy the status quo to 'build back better' in the form that you want. The gender status quo had to be

destroyed by persuading the Cult-created Woke mentality to believe that you can have 100 genders or more. A programme for 9 to 12 year olds produced by the Cult-owned BBC promoted the 100 genders narrative. The very idea may be the most monumental nonsense, but it is not what is true that counts, only what you can make people *believe* is true. Once the gender of $2 + 2 = 4$ has been dismantled through indoctrination, intimidation and $2 + 2 = 5$ then the new no-gender normal can take its place with Human 2.0.

Aldous Huxley revealed the plan in his prophetic *Brave New World* in 1932:

Natural reproduction has been done away with and children are created, decanted', and raised in 'hatcheries and conditioning centres'. From birth, people are genetically designed to fit into one of five castes, which are further split into 'Plus' and 'Minus' members and designed to fulfil predetermined positions within the social and economic strata of the World State.

How could Huxley know this in 1932? For the same reason George Orwell knew about the Big Brother state in 1948, Cult insiders I have quoted knew about it in 1969, and I have known about it since the early 1990s. If you are connected to the Cult or you work your balls off to uncover the plan you can predict the future. The process is simple. If there is a plan for the world and nothing intervenes to stop it then it will happen. Thus if you communicate the plan ahead of time you are perceived to have predicted the future, but you haven't. You have revealed the plan which without intervention will become the human future. The whole reason I have done what I have is to alert enough people to inspire an intervention and maybe at last that time has come with the Cult and its intentions now so obvious to anyone with a brain in working order.

The future is here

Technological wombs that Huxley described to replace parent procreation are already being developed and they are only the projects we know about in the public arena. Israeli scientists told *The Times of Israel* in March, 2021, that they have grown 250-cell embryos

into mouse fetuses with fully formed organs using artificial wombs in a development they say could pave the way for gestating humans outside the womb. Professor Jacob Hanna of the Weizmann Institute of Science said:

We took mouse embryos from the mother at day five of development, when they are just of 250 cells, and had them in the incubator from day five until day 11, by which point they had grown all their organs.

By day 11 they make their own blood and have a beating heart, a fully developed brain. Anybody would look at them and say, 'this is clearly a mouse foetus with all the characteristics of a mouse.' It's gone from being a ball of cells to being an advanced foetus.

A special liquid is used to nourish embryo cells in a laboratory dish and they float on the liquid to duplicate the first stage of embryonic development. The incubator creates all the right conditions for its development, Hanna said. The liquid gives the embryo 'all the nutrients, hormones and sugars they need' along with a custom-made electronic incubator which controls gas concentration, pressure and temperature. The cutting-edge in the underground bases and other secret locations will be light years ahead of that, however, and this was reported by the London *Guardian* in 2017:

We are approaching a biotechnological breakthrough. Ectogenesis, the invention of a complete external womb, could completely change the nature of human reproduction. In April this year, researchers at the Children's Hospital of Philadelphia announced their development of an artificial womb.

The article was headed 'Artificial wombs could soon be a reality. What will this mean for women?' What would it mean for children is an even bigger question. No mother to bond with only a machine in preparation for a life of soulless interaction and control in a world governed by machines (see the *Matrix* movies). Now observe the calculated manipulations of the 'Covid' hoax as human interaction and warmth has been curtailed by distancing, isolation and fear with people communicating via machines on a scale never seen before.

These are all dots in the same picture as are all the personal assistants, gadgets and children's toys through which kids and adults communicate with AI as if it is human. The AI 'voice' on Sat-Nav should be included. All these things are psychological preparation for the Cult endgame. Before you can make a physical connection with AI you have to make a psychological connection and that is what people are being conditioned to do with this ever gathering human-AI interaction. Movies and TV programmes depicting the transhuman, robot dystopia relate to a phenomenon known as 'pre-emptive programming' in which the world that is planned is portrayed everywhere in movies, TV and advertising. This is conditioning the conscious and subconscious mind to become familiar with the planned reality to dilute resistance when it happens for real. What would have been a shock such is the change is made less so. We have young children put on the road to transgender transition surgery with puberty blocking drugs at an age when they could never be able to make those life-changing decisions.

Rachel Levine, a professor of paediatrics and psychiatry who believes in treating children this way, became America's highest-ranked openly-transgender official when she was confirmed as US Assistant Secretary at the Department of Health and Human Services after being nominated by Joe Biden (the Cult). Activists and governments press for laws to deny parents a say in their children's transition process so the kids can be isolated and manipulated into agreeing to irreversible medical procedures. A Canadian father Robert Hoogland was denied bail by the Vancouver Supreme Court in 2021 and remained in jail for breaching a court order that he stay silent over his young teenage daughter, a minor, who was being offered life-changing hormone therapy without parental consent. At the age of 12 the girl's 'school counsellor' said she may be transgender, referred her to a doctor and told the school to treat her like a boy. This is another example of state-serving schools imposing ever more control over children's lives while parents have ever less.

Contemptible and extreme child abuse is happening all over the world as the Cult gender-fusion operation goes into warp-speed.

Why the war on men – and now women?

The question about what artificial wombs mean for women should rightly be asked. The answer can be seen in the deletion of women's rights involving sport, changing rooms, toilets and status in favour of people in male bodies claiming to identify as women. I can identify as a mountain climber, but it doesn't mean I can climb a mountain any more than a biological man can be a biological woman. To believe so is a triumph of belief over factual reality which is the very perceptual basis of everything Woke. Women's sport is being destroyed by allowing those with male bodies who say they identify as female to 'compete' with girls and women. Male body 'women' dominate 'women's' competition with their greater muscle mass, bone density, strength and speed. With that disadvantage sport for women loses all meaning. To put this in perspective nearly 300 American high school boys can run faster than the quickest woman sprinter in the world. Women are seeing their previously protected spaces invaded by male bodies simply because they claim to identify as women. That's all they need to do to access all women's spaces and activities under the Biden 'Equality Act' that destroys equality for women with the usual Orwellian Woke inversion. Male sex offenders have already committed rapes in women's prisons after claiming to identify as women to get them transferred. Does this not matter to the Woke 'equality' hypocrites? Not in the least. What matters to Cult manipulators and funders behind transgender activists is to advance gender fusion on the way to the no-gender 'human'. When you are seeking to impose transparent nonsense like this, or the 'Covid' hoax, the only way the nonsense can prevail is through censorship and intimidation of dissenters, deletion of factual information, and programming of the unquestioning, bewildered and naive. You don't have to scan the world for long to see that all these things are happening.

Many women's rights organisations have realised that rights and status which took such a long time to secure are being eroded and that it is systematic. Kara Dansky of the global Women's Human Rights Campaign said that Biden's transgender executive order immediately he took office, subsequent orders, and Equality Act legislation that followed 'seek to erase women and girls in the law as a category'. *Exactly*. I said during the long ago-started war on men (in which many women play a crucial part) that this was going to turn into a war on them. The Cult is phasing out *both* male and female genders. To get away with that they are brought into conflict so they are busy fighting each other while the Cult completes the job with no unity of response. Unity, people, *unity*. We need unity everywhere. Transgender is the only show in town as the big step towards the no-gender human. It's not about rights for transgender people and never has been. Woke political correctness is deleting words relating to genders to the same end. Wokers believe this is to be 'inclusive' when the opposite is true. They are deleting words describing gender because gender *itself* is being deleted by Human 2.0. Terms like 'man', 'woman', 'mother' and 'father' are being deleted in the universities and other institutions to be replaced by the *no*-gender, not trans-gender, 'individuals' and 'guardians'. Women's rights campaigner Maria Keffler of Partners for Ethical Care said: 'Children are being taught from kindergarten upward that some boys have a vagina, some girls have a penis, and that kids can be any gender they want to be.' Do we really believe that suddenly countries all over the world at the same time had the idea of having drag queens go into schools or read transgender stories to very young children in the local library? It's coldly-calculated confusion of gender on the way to the fusion of gender. Suzanne Vierling, a psychologist from Southern California, made another important point:

Yesterday's slave woman who endured gynecological medical experiments is today's girl-child being butchered in a booming gender-transitioning sector. Ovaries removed, pushing her into menopause and osteoporosis, uncharted territory, and parents' rights and authority decimated.

The erosion of parental rights is a common theme in line with the Cult plans to erase the very concept of parents and 'ovaries removed, pushing her into menopause' means what? Those born female lose the ability to have children – another way to discontinue humanity as we know it.

Eliminating Human 1.0 (before our very eyes)

To pave the way for Human 2.0 you must phase out Human 1.0. This is happening through plummeting sperm counts and making women infertile through an onslaught of chemicals, radiation (including smartphones in pockets of men) and mRNA 'vaccines'. Common agriculture pesticides are also having a devastating impact on human fertility. I have been tracking collapsing sperm counts in the books for a long time and in 2021 came a book by fertility scientist and reproductive epidemiologist Shanna Swan, *Count Down: How Our Modern World Is Threatening Sperm Counts, Altering Male and Female Reproductive Development and Imperiling the Future of the Human Race*. She reports how the global fertility rate dropped by *half* between 1960 and 2016 with America's birth rate 16 percent below where it needs to be to sustain the population. Women are experiencing declining egg quality, more miscarriages, and more couples suffer from infertility. Other findings were an increase in erectile dysfunction, infant boys developing more genital abnormalities, male problems with conception, and plunging levels of the male hormone testosterone which would explain why so many men have lost their backbone and masculinity. This has been very evident during the 'Covid' hoax when women have been prominent among the Pushbackers and big strapping blokes have bowed their heads, covered their faces with a nappy and quietly submitted. Mind control expert Cathy O'Brien also points to how global education introduced the concept of 'we're all winners' in sport and classrooms: 'Competition was defused, and it in turn defused a sense of fighting back.' This is another version of the 'equity' doctrine in which you drive down rather than raise up. What a contrast in Cult-controlled China with its global ambitions

where the government published plans in January, 2021, to 'cultivate masculinity' in boys from kindergarten through to high school in the face of a 'masculinity crisis'. A government adviser said boys would be soon become 'delicate, timid and effeminate' unless action was taken. Don't expect any similar policy in the targeted West. A 2006 study showed that a 65-year-old man in 2002 had testosterone levels 15 percent lower than a 65-year-old man in 1987 while a 2020 study found a similar story with young adults and adolescents. Men are getting prescriptions for testosterone replacement therapy which causes an even greater drop in sperm count with up to 99 percent seeing sperm counts drop to zero during the treatment. More sperm is defective and malfunctioning with some having two heads or not pursuing an egg.

A class of *synthetic* chemicals known as phthalates are being blamed for the decline. These are found everywhere in plastics, shampoos, cosmetics, furniture, flame retardants, personal care products, pesticides, canned foods and even receipts. Why till receipts? Everyone touches them. Let no one delude themselves that all this is not systematic to advance the long-time agenda for human body transformation. Phthalates mimic hormones and disrupt the hormone balance causing testosterone to fall and genital birth defects in male infants. Animals and fish have been affected in the same way due to phthalates and other toxins in rivers. When fish turn gay or change sex through chemicals in rivers and streams it is a pointer to why there has been such an increase in gay people and the sexually confused. It doesn't matter to me what sexuality people choose to be, but if it's being affected by chemical pollution and consumption then we need to know. Does anyone really think that this is not connected to the transgender agenda, the war on men and the condemnation of male 'toxic masculinity'? You watch this being followed by 'toxic femininity'. It's already happening. When breastfeeding becomes 'chest-feeding', pregnant women become pregnant people along with all the other Woke claptrap you know that the world is going insane and there's a Cult scam in progress. Transgender activists are promoting the Cult agenda while Cult

billionaires support and fund the insanity as they laugh themselves to sleep at the sheer stupidity for which humans must be infamous in galaxies far, far away.

‘Covid vaccines’ and female infertility

We can now see why the ‘vaccine’ has been connected to potential infertility in women. Dr Michael Yeadon, former Vice President and Chief Scientific Advisor at Pfizer, and Dr Wolfgang Wodarg in Germany, filed a petition with the European Medicines Agency in December, 2020, urging them to stop trials for the Pfizer/BioNTech shot and all other mRNA trials until further studies had been done. They were particularly concerned about possible effects on fertility with ‘vaccine’-produced antibodies attacking the protein Syncytin-1 which is responsible for developing the placenta. The result would be infertility ‘of indefinite duration’ in women who have the ‘vaccine’ with the placenta failing to form. Section 10.4.2 of the Pfizer/BioNTech trial protocol says that pregnant women or those who might become so should not have mRNA shots. Section 10.4 warns men taking mRNA shots to ‘be abstinent from heterosexual intercourse’ and not to donate sperm. The UK government said that it *did not know* if the mRNA procedure had an effect on fertility. *Did not know?* These people have to go to jail. UK government advice did not recommend at the start that pregnant women had the shot and said they should avoid pregnancy for at least two months after ‘vaccination’. The ‘advice’ was later updated to pregnant women should only have the ‘vaccine’ if the benefits outweighed the risks to mother and foetus. What the hell is that supposed to mean? Then ‘spontaneous abortions’ began to appear and rapidly increase on the adverse reaction reporting schemes which include only a fraction of adverse reactions. Thousands and ever-growing numbers of ‘vaccinated’ women are describing changes to their menstrual cycle with heavier blood flow, irregular periods and menstruating again after going through the menopause – all links to reproduction effects. Women are passing blood clots and the lining of their uterus while men report erectile dysfunction and blood effects. Most

significantly of all *unvaccinated* women began to report similar menstrual changes after interaction with '*vaccinated*' people and men and children were also affected with bleeding noses, blood clots and other conditions. 'Shedding' is when vaccinated people can emit the content of a vaccine to affect the unvaccinated, but this is different. 'Vaccinated' people were not shedding a 'live virus' allegedly in 'vaccines' as before because the fake 'Covid vaccines' involve synthetic material and other toxicity. Doctors exposing what is happening prefer the term 'transmission' to shedding. Somehow those that have had the shots are transmitting effects to those that haven't. Dr Carrie Madej said the nano-content of the 'vaccines' can 'act like an antenna' to others around them which fits perfectly with my own conclusions. This 'vaccine' transmission phenomenon was becoming known as the book went into production and I deal with this further in the Postscript.

Vaccine effects on sterility are well known. The World Health Organization was accused in 2014 of sterilising millions of women in Kenya with the evidence confirmed by the content of the vaccines involved. The same WHO behind the 'Covid' hoax admitted its involvement for more than ten years with the vaccine programme. Other countries made similar claims. Charges were lodged by Tanzania, Nicaragua, Mexico, and the Philippines. The Gardasil vaccine claimed to protect against a genital 'virus' known as HPV has also been linked to infertility. Big Pharma and the WHO (same thing) are criminal and satanic entities. Then there's the Bill Gates Foundation which is connected through funding and shared interests with 20 pharmaceutical giants and laboratories. He stands accused of directing the policy of United Nations Children's Fund (UNICEF), vaccine alliance GAVI, and other groupings, to advance the vaccine agenda and silence opposition at great cost to women and children. At the same time Gates wants to reduce the global population. Coincidence?

Great Reset = Smart Grid = new human

The Cult agenda I have been exposing for 30 years is now being openly promoted by Cult assets like Gates and Klaus Schwab of the World Economic Forum under code-terms like the 'Great Reset', 'Build Back Better' and 'a rare but narrow window of opportunity to reflect, reimagine, and reset our world'. What provided this 'rare but narrow window of opportunity'? The 'Covid' hoax did. Who created that? *They* did. My books from not that long ago warned about the planned 'Internet of Things' (IoT) and its implications for human freedom. This was the plan to connect all technology to the Internet and artificial intelligence and today we are way down that road with an estimated 36 billion devices connected to the World Wide Web and that figure is projected to be 76 billion by 2025. I further warned that the Cult planned to go beyond that to the Internet of *Everything* when the human brain was connected via AI to the Internet and Kurzweil's 'cloud'. Now we have Cult operatives like Schwab calling for precisely that under the term 'Internet of Bodies', a fusion of the physical, digital and biological into one centrally-controlled Smart Grid system which the Cult refers to as the 'Fourth Industrial Revolution'. They talk about the 'biological', but they really mean the synthetic-biological which is required to fully integrate the human body and brain into the Smart Grid and artificial intelligence planned to replace the human mind. We have everything being synthetically manipulated including the natural world through GMO and smart dust, the food we eat and the human body itself with synthetic 'vaccines'. I said in *The Answer* that we would see the Cult push for synthetic meat to replace animals and in February, 2021, the so predictable psychopath Bill Gates called for the introduction of synthetic meat to save us all from 'climate change'. The climate hoax just keeps on giving like the 'Covid' hoax. The war on meat by vegan activists is a carbon (oops, sorry) copy of the manipulation of transgender activists. They have no idea (except their inner core) that they are being used to promote and impose the agenda of the Cult or that they are only the *vehicle* and not the *reason*. This is not to say those who choose not to eat meat shouldn't be respected and supported in that right, but there are ulterior motives

for those in power. A *Forbes* article in December, 2019, highlighted the plan so beloved of Schwab and the Cult under the heading: 'What Is The Internet of Bodies? And How Is It Changing Our World?' The article said the human body is the latest data platform (remember 'our vaccine is an operating system'). *Forbes* described the plan very accurately and the words could have come straight out of my books from long before:

The Internet of Bodies (IoB) is an extension of the IoT and basically connects the human body to a network through devices that are ingested, implanted, or connected to the body in some way. Once connected, data can be exchanged, and the body and device can be remotely monitored and controlled.

They were really describing a human hive mind with human perception centrally-dictated via an AI connection as well as allowing people to be 'remotely monitored and controlled'. Everything from a fridge to a human mind could be directed from a central point by these insane psychopaths and 'Covid vaccines' are crucial to this. *Forbes* explained the process I mentioned earlier of holdable and wearable technology followed by implantable. The article said there were three generations of the Internet of Bodies that include:

- Body external: These are wearable devices such as Apple Watches or Fitbits that can monitor our health.
- Body internal: These include pacemakers, cochlear implants, and digital pills that go inside our bodies to monitor or control various aspects of health.
- Body embedded: The third generation of the Internet of Bodies is embedded technology where technology and the human body are melded together and have a real-time connection to a remote machine.

Forbes noted the development of the Brain Computer Interface (BCI) which merges the brain with an external device for monitoring and controlling in real-time. 'The ultimate goal is to help restore function to individuals with disabilities by using brain signals rather than conventional neuromuscular pathways.' Oh, do fuck off. The goal of brain interface technology is controlling human thought and emotion from the central point in a hive mind serving its masters wishes. Many people are now agreeing to be chipped to open doors without a key. You can recognise them because they'll be wearing a mask, social distancing and lining up for the 'vaccine'. The Cult plans a Great Reset money system after they have completed the demolition of the global economy in which 'money' will be exchanged through communication with body operating systems. Rand Corporation, a Cult-owned think tank, said of the Internet of Bodies or IoB:

Internet of Bodies technologies fall under the broader IoT umbrella. But as the name suggests, IoB devices introduce an even more intimate interplay between humans and gadgets. IoB devices monitor the human body, collect health metrics and other personal information, and transmit those data over the Internet. Many devices, such as fitness trackers, are already in use ... IoB devices ... and those in development can track, record, and store users' whereabouts, bodily functions, and what they see, hear, and even think.

Schwab's World Economic Forum, a long-winded way of saying 'fascism' or 'the Cult', has gone full-on with the Internet of Bodies in the 'Covid' era. 'We're entering the era of the Internet of Bodies', it declared, 'collecting our physical data via a range of devices that can be implanted, swallowed or worn'. The result would be a huge amount of health-related data that could improve human wellbeing around the world, and prove crucial in fighting the 'Covid-19 pandemic'. Does anyone think these clowns care about 'human wellbeing' after the death and devastation their pandemic hoax has purposely caused? Schwab and co say we should move forward with the Internet of Bodies because 'Keeping track of symptoms could help us stop the spread of infection, and quickly detect new cases'. How wonderful, but keeping track' is all they are really bothered

about. Researchers were investigating if data gathered from smartwatches and similar devices could be used as viral infection alerts by tracking the user's heart rate and breathing. Schwab said in his 2018 book *Shaping the Future of the Fourth Industrial Revolution*:

The lines between technologies and beings are becoming blurred and not just by the ability to create lifelike robots or synthetics. Instead it is about the ability of new technologies to literally become part of us. Technologies already influence how we understand ourselves, how we think about each other, and how we determine our realities. As the technologies ... give us deeper access to parts of ourselves, we may begin to integrate digital technologies into our bodies.

You can see what the game is. Twenty-four hour control and people – if you could still call them that – would never know when something would go ping and take them out of circulation. It's the most obvious rush to a global fascist dictatorship and the complete submission of humanity and yet still so many are locked away in their Cult-induced perceptual coma and can't see it.

Smart Grid control centres

The human body is being transformed by the 'vaccines' and in other ways into a synthetic cyborg that can be attached to the global Smart Grid which would be controlled from a central point and other sub-locations of Grid manipulation. Where are these planned to be? Well, China for a start which is one of the Cult's biggest centres of operation. The technological control system and technocratic rule was incubated here to be unleashed across the world after the 'Covid' hoax came out of China in 2020. Another Smart Grid location that will surprise people new to this is Israel. I have exposed in *The Trigger* how Sabbatian technocrats, intelligence and military operatives were behind the horrors of 9/11 and not 19 Arab hijackers' who somehow manifested the ability to pilot big passenger airliners when instructors at puddle-jumping flying schools described some of them as a joke. The 9/11 attacks were made possible through control of civilian and military air computer systems and those of the White House, Pentagon and connected agencies. See *The Trigger* – it

will blow your mind. The controlling and coordinating force were the Sabbatian networks in Israel and the United States which by then had infiltrated the entire US government, military and intelligence system. The real name of the American Deep State is 'Sabbatian State'. Israel is a tiny country of only nine million people, but it is one of the global centres of cyber operations and fast catching Silicon Valley in importance to the Cult. Israel is known as the 'start-up nation' for all the cyber companies spawned there with the Sabbatian specialisation of 'cyber security' that I mentioned earlier which gives those companies access to computer systems of their clients in real time through 'backdoors' written into the coding when security software is downloaded. The Sabbatian centre of cyber operations outside Silicon Valley is the Israeli military Cyber Intelligence Unit, the biggest infrastructure project in Israel's history, headquartered in the desert-city of Beersheba and involving some 20,000 'cyber soldiers'. Here are located a literal army of Internet trolls scanning social media, forums and comment lists for anyone challenging the Cult agenda. The UK military has something similar with its 77th Brigade and associated operations. The Beersheba complex includes research and development centres for other Cult operations such as Intel, Microsoft, IBM, Google, Apple, Hewlett-Packard, Cisco Systems, Facebook and Motorola. Techcrunch.com ran an article about the Beersheba global Internet technology centre headlined 'Israel's desert city of Beersheba is turning into a cybertech oasis':

The military's massive relocation of its prestigious technology units, the presence of multinational and local companies, a close proximity to Ben Gurion University and generous government subsidies are turning Beersheba into a major global cybertech hub. Beersheba has all of the ingredients of a vibrant security technology ecosystem, including Ben Gurion University with its graduate program in cybersecurity and Cyber Security Research Center, and the presence of companies such as EMC, Deutsche Telekom, PayPal, Oracle, IBM, and Lockheed Martin. It's also the future home of the INCB (Israeli National Cyber Bureau); offers a special income tax incentive for cyber security companies, and was the site for the relocation of the army's intelligence corps units.

Sabbatians have taken over the cyber world through the following process: They scan the schools for likely cyber talent and develop them at Ben Gurion University and their period of conscription in the Israeli Defense Forces when they are stationed at the Beersheba complex. When the cyber talented officially leave the army they are funded to start cyber companies with technology developed by themselves or given to them by the state. Much of this is stolen through backdoors of computer systems around the world with America top of the list. Others are sent off to Silicon Valley to start companies or join the major ones and so we have many major positions filled by apparently 'Jewish' but really Sabbatian operatives. Google, YouTube and Facebook are all run by 'Jewish' CEOs while Twitter is all but run by ultra-Zionist hedge-fund shark Paul Singer. At the centre of the Sabbatian global cyber web is the Israeli army's Unit 8200 which specialises in hacking into computer systems of other countries, inserting viruses, gathering information, instigating malfunction, and even taking control of them from a distance. A long list of Sabbatians involved with 9/11, Silicon Valley and Israeli cyber security companies are operatives of Unit 8200. This is not about Israel. It's about the Cult. Israel is planned to be a Smart Grid hub as with China and what is happening at Beersheba is not for the benefit of Jewish people who are treated disgustingly by the Sabbatian elite that control the country. A glance at the Nuremberg Codes will tell you that.

The story is much bigger than 'Covid', important as that is to where we are being taken. Now, though, it's time to really strap in. There's more ... much more ...

CHAPTER ELEVEN

Who controls the Cult?

Awake, arise or be forever fall'n
John Milton, *Paradise Lost*

I have exposed this far the level of the Cult conspiracy that operates in the world of the seen and within the global secret society and satanic network which operates in the shadows one step back from the seen. The story, however, goes much deeper than that.

The 'Covid' hoax is major part of the Cult agenda, but only part, and to grasp the biggest picture we have to expand our attention beyond the realm of human sight and into the infinity of possibility that we cannot see. It is from here, ultimately, that humanity is being manipulated into a state of total control by the force which dictates the actions of the Cult. How much of reality can we see? Next to damn all is the answer. We may appear to see all there is to see in the 'space' our eyes survey and observe, but little could be further from the truth. The human 'world' is only a tiny band of frequency that the body's visual and perceptual systems can decode into *perception* of a 'world'. According to mainstream science the electromagnetic spectrum is 0.005 percent of what exists in the Universe ([Fig 10](#)). The maximum estimate I have seen is 0.5 percent and either way it's miniscule. I say it is far, far, smaller even than 0.005 percent when you compare reality we see with the totality of reality that we don't. Now get this if you are new to such information: Visible light, the only band of frequency that we can see, is a *fraction* of the 0.005

percent (Fig 11 overleaf). Take this further and realise that our universe is one of infinite universes and that universes are only a fragment of overall reality – *infinite* reality. Then compare that with the almost infinitesimal frequency band of visible light or human sight. You see that humans are as near blind as it is possible to be without actually being so. Artist and filmmaker, Sergio Toporek, said:

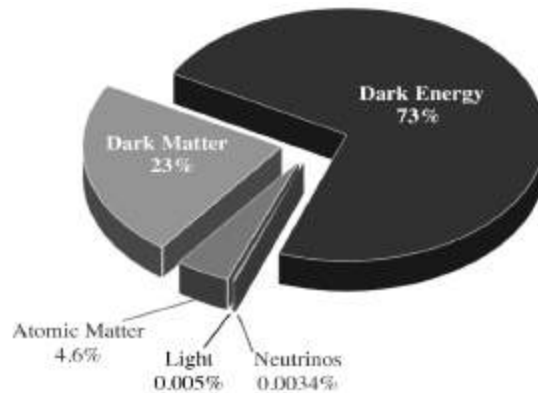


Figure 10: Humans can perceive such a tiny band of visual reality it's laughable.

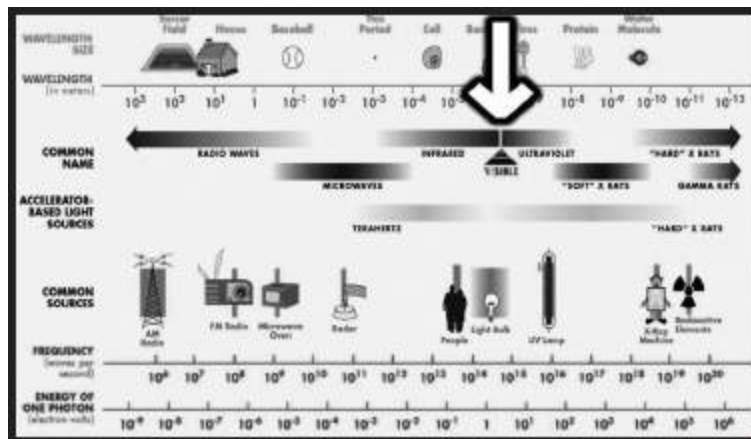


Figure 11: We can see a smear of the 0.005 percent electromagnetic spectrum, but we still know it all. Yep, makes sense.

Consider that you can see less than 1% of the electromagnetic spectrum and hear less than 1% of the acoustic spectrum. 90% of the cells in your body carry their own microbial DNA and are not 'you'. The atoms in your body are 99.9999999999999999% empty space and none of them are the ones you were born with ... Human beings have 46 chromosomes, two less than a potato.

The existence of the rainbow depends on the conical photoreceptors in your eyes; to animals without cones, the rainbow does not exist. So you don't just look at a rainbow, you create it. This is pretty amazing, especially considering that all the beautiful colours you see represent less than 1% of the electromagnetic spectrum.

Suddenly the 'world' of humans looks a very different place. Take into account, too, that Planet Earth when compared with the projected size of this single universe is the equivalent of a billionth of a pinhead. Imagine the ratio that would be when compared to infinite reality. To think that Christianity once insisted that Earth and humanity were the centre of everything. This background is vital if we are going to appreciate the nature of 'human' and how we can be manipulated by an unseen force. To human visual reality virtually *everything* is unseen and yet the prevailing perception within the institutions and so much of the public is that if we can't see it, touch it, hear it, taste it and smell it then it cannot exist. Such perception is indoctrinated and encouraged by the Cult and its agents because it isolates believers in the strictly limited, village-idiot, realm of the five senses where perceptions can be firewalled and information controlled. Most of those perpetuating the 'this-world-is-all-there-is' insanity are themselves indoctrinated into believing the same delusion. While major players and influencers know that official reality is laughable most of those in science, academia and medicine really believe the nonsense they peddle and teach succeeding generations. Those who challenge the orthodoxy are dismissed as nutters and freaks to protect the manufactured illusion from exposure. Observe the dynamic of the 'Covid' hoax and you will see how that takes the same form. The inner-circle psychopaths knows it's a gigantic scam, but almost the entirety of those imposing their fascist rules believe that 'Covid' is all that they're told it is.

Stolen identity

Ask people who they are and they will give you their name, place of birth, location, job, family background and life story. Yet that is not who they are – it is what they are *experiencing*. The difference is *absolutely crucial*. The true 'I', the eternal, infinite 'I', is consciousness,

a state of being aware. Forget 'form'. That is a vehicle for a brief experience. Consciousness does not come *from* the brain, but *through* the brain and even that is more symbolic than literal. We are awareness, pure awareness, and this is what withdraws from the body at what we call 'death' to continue our eternal beingness, *isness*, in other realms of reality within the limitlessness of infinity or the Biblical 'many mansions in my father's house'. Labels of a human life, man, woman, transgender, black, white, brown, nationality, circumstances and income are not who we are. They are what we are – awareness – is *experiencing* in a brief connection with a band of frequency we call 'human'. The labels are not the self; they are, to use the title of one of my books, a *Phantom Self*. I am not David Icke born in Leicester, England, on April 29th, 1952. I am the consciousness *having that experience*. The Cult and its non-human masters seek to convince us through the institutions of 'education', science, medicine, media and government that what we are *experiencing* is who we *are*. It's so easy to control and direct perception locked away in the bewildered illusions of the five senses with no expanded radar. Try, by contrast, doing the same with a humanity aware of its true self and its true power to consciously create its reality and experience. How is it possible to do this? We do it all day every day. If you perceive yourself as 'little me' with no power to impact upon your life and the world then your life experience will reflect that. You will hand the power you don't think you have to authority in all its forms which will use it to control your experience. This, in turn, will appear to confirm your perception of 'little me' in a self-fulfilling feedback loop. But that is what 'little me' really is – a *perception*. We are all 'big-me', infinite me, and the Cult has to make us forget that if its will is to prevail. We are therefore manipulated and pressured into self-identifying with human labels and not the consciousness/awareness *experiencing* those human labels.

The phenomenon of identity politics is a Cult-instigated manipulation technique to sub-divide previous labels into even smaller ones. A United States university employs this list of letters to

describe student identity: LGBTTQQFAGPBDSM or lesbian, gay, bisexual, transgender, transsexual, queer, questioning, flexual, asexual, gender-fuck, polyamorous, bondage/discipline, dominance/submission and sadism/masochism. I'm sure other lists are even longer by now as people feel the need to self-identity the 'I' with the minutiae of race and sexual preference. Wokers programmed by the Cult for generations believe this is about 'inclusivity' when it's really the Cult locking them away into smaller and smaller versions of Phantom Self while firewalling them from the influence of their true self, the infinite, eternal 'I'. You may notice that my philosophy which contends that we are all unique points of attention/awareness within the same infinite whole or Oneness is the ultimate non-racism. The very sense of Oneness makes the judgement of people by their body-type, colour or sexuality utterly ridiculous and confirms that racism has no understanding of reality (including anti-white racism). Yet despite my perception of life Cult agents and fast-asleep Wokers label me racist to discredit my information while they are themselves phenomenally racist and sexist. All they see is race and sexuality and they judge people as good or bad, demons or untouchables, by their race and sexuality. All they see is *Phantom Self* and perceive themselves in terms of Phantom Self. They are pawns and puppets of the Cult agenda to focus attention and self-identity in the five senses and play those identities against each other to divide and rule. Columbia University has introduced segregated graduations in another version of social distancing designed to drive people apart and teach them that different racial and cultural groups have nothing in common with each other. The last thing the Cult wants is unity. Again the pump-primers of this will be Cult operatives in the knowledge of what they are doing, but the rest are just the Phantom Self blind leading the Phantom Self blind. We *do* have something in common – we are all *the same consciousness* having different temporary experiences.

What is this 'human'?

Yes, what *is* 'human'? That is what we are supposed to be, right? I mean 'human'? True, but 'human' is the experience not the 'I'. Break it down to basics and 'human' is the way that information is processed. If we are to experience and interact with this band of frequency we call the 'world' we must have a vehicle that operates within that band of frequency. Our consciousness in its prime form cannot do that; it is way beyond the frequency of the human realm. My consciousness or awareness could not tap these keys and pick up the cup in front of me in the same way that radio station A cannot interact with radio station B when they are on different frequencies. The human body is the means through which we have that interaction. I have long described the body as a biological computer which processes information in a way that allows consciousness to experience this reality. The body is a receiver, transmitter and processor of information in a particular way that we call human. We visually perceive only the world of the five senses in a wakened state – that is the limit of the body's visual decoding system. In truth it's not even visual in the way we experience 'visual reality' as I will come to in a moment. We are 'human' because the body processes the information sources of human into a reality and behaviour system that we *perceive* as human. Why does an elephant act like an elephant and not like a human or a duck? The elephant's biological computer is a different information field and processes information according to that program into a visual and behaviour type we call an elephant. The same applies to everything in our reality. These body information fields are perpetuated through procreation (like making a copy of a software program). The Cult wants to break that cycle and intervene technologically to transform the human information field into one that will change what we call humanity. If it can change the human information field it will change the way that field processes information and change humanity both 'physically' and psychologically. Hence the *messenger* (information) RNA 'vaccines' and so much more that is targeting human genetics by changing the body's information – *messaging* – construct through food, drink, radiation, toxicity and other means.

Reality that we experience is nothing like reality as it really is in the same way that the reality people experience in virtual reality games is not the reality they are really living in. The game is only a decoded source of information that appears to be a reality. Our world is also an information construct – a *simulation* (more later). In its base form our reality is a wavefield of information much the same in theme as Wi-Fi. The five senses decode wavefield information into electrical information which they communicate to the brain to decode into holographic (illusory ‘physical’) information. Different parts of the brain specialise in decoding different senses and the information is fused into a reality that appears to be outside of us but is really inside the brain and the genetic structure in general ([Fig 12](#) overleaf). DNA is a receiver-transmitter of information and a vital part of this decoding process and the body’s connection to other realities. Change DNA and you change the way we decode and connect with reality – see ‘Covid vaccines’. Think of computers decoding Wi-Fi. You have information encoded in a radiation field and the computer decodes that information into a very different form on the screen. You can’t see the Wi-Fi until its information is made manifest on the screen and the information on the screen is inside the computer and not outside. I have just described how we decode the ‘human world’. All five senses decode the waveform ‘Wi-Fi’ field into electrical signals and the brain (computer) constructs reality inside the brain and not outside – ‘You don’t just look at a rainbow, you create it’. Sound is a simple example. We don’t hear sound until the brain decodes it. Waveform sound waves are picked up by the hearing sense and communicated to the brain in an electrical form to be decoded into the sounds that we hear. Everything we hear is inside the brain along with everything we see, feel, smell and taste. Words and language are waveform fields generated by our vocal chords which pass through this process until they are decoded by the brain into words that we hear. Different languages are different frequency fields or sound waves generated by vocal chords. Late British philosopher Alan Watts said:

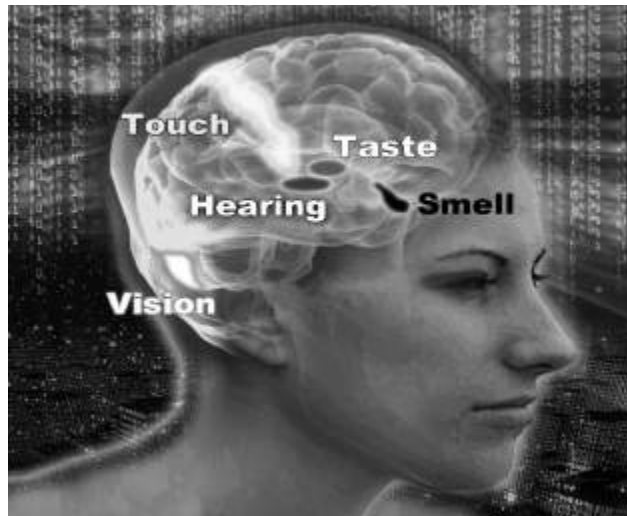


Figure 12: The brain receives information from the five senses and constructs from that our perceived reality.

[Without the brain] the world is devoid of light, heat, weight, solidity, motion, space, time or any other imaginable feature. All these phenomena are interactions, or transactions, of vibrations with a certain arrangement of neurons.

That's exactly what they are and scientist Robert Lanza describes in his book, *Biocentrism*, how we decode electromagnetic waves and energy into visual and 'physical' experience. He uses the example of a flame emitting photons, electromagnetic energy, each pulsing electrically and magnetically:

... these ... invisible electromagnetic waves strike a human retina, and if (and only if) the waves happen to measure between 400 and 700 nano meters in length from crest to crest, then their energy is just right to deliver a stimulus to the 8 million cone-shaped cells in the retina.

Each in turn send an electrical pulse to a neighbour neuron, and on up the line this goes, at 250 mph, until it reaches the ... occipital lobe of the brain, in the back of the head. There, a cascading complex of neurons fire from the incoming stimuli, and we subjectively perceive this experience as a yellow brightness occurring in a place we have been conditioned to call the 'external world'.

You hear what you decode

If a tree falls or a building collapses they make no noise unless someone is there to decode the energetic waves generated by the disturbance into what we call sound. Does a falling tree make a noise? Only if you hear it – *decode* it. Everything in our reality is a frequency field of information operating within the overall ‘Wi-Fi’ field that I call The Field. A vibrational disturbance is generated in The Field by the fields of the falling tree or building. These disturbance waves are what we decode into the sound of them falling. If no one is there to do that then neither will make any noise. Reality is created by the observer – *decoder* – and the *perceptions* of the observer affect the decoding process. For this reason different people – different *perceptions* – will perceive the same reality or situation in a different way. What one may perceive as a nightmare another will see as an opportunity. The question of why the Cult is so focused on controlling human perception now answers itself. All experienced reality is the act of decoding and we don’t experience Wi-Fi until it is decoded on the computer screen. The sight and sound of an Internet video is encoded in the Wi-Fi all around us, but we don’t see or hear it until the computer decodes that information. Taste, smell and touch are all phenomena of the brain as a result of the same process. We don’t taste, smell or feel anything except in the brain and there are pain relief techniques that seek to block the signal from the site of discomfort to the brain because if the brain doesn’t decode that signal we don’t feel pain. Pain is in the brain and only appears to be at the point of impact thanks to the feedback loop between them. We don’t see anything until electrical information from the sight senses is decoded in an area at the back of the brain. If that area is damaged we can go blind when our eyes are perfectly okay. So why do we go blind if we damage an eye? We damage the information processing between the waveform visual information and the visual decoding area of the brain. If information doesn’t reach the brain in a form it can decode then we can’t see the visual reality that it represents. What’s more the brain is decoding only a fraction of the information it receives and the rest is absorbed by the

sub-conscious mind. This explanation is from the science magazine, *Wonderpedia*:

Every second, 11 million sensations crackle along these [brain] pathways ... The brain is confronted with an alarming array of images, sounds and smells which it rigorously filters down until it is left with a manageable list of around 40. Thus 40 sensations per second make up what we perceive as reality.

The 'world' is not what people are told to believe that is it and the inner circles of the Cult *know that*.

Illusory 'physical' reality

We can only see a smear of 0.005 percent of the Universe which is only one of a vast array of universes – 'mansions' – within infinite reality. Even then the brain decodes only 40 pieces of information ('sensations') from a potential *11 million* that we receive every second. Two points strike you from this immediately: The sheer breathtaking stupidity of believing we know anything so rigidly that there's nothing more to know; and the potential for these processes to be manipulated by a malevolent force to control the reality of the population. One thing I can say for sure with no risk of contradiction is that when you can perceive an almost indescribable fraction of infinite reality there is always more to know as in tidal waves of it. Ancient Greek philosopher Socrates was so right when he said that wisdom is to know how little we know. How obviously true that is when you think that we are experiencing a physical world of solidity that is neither physical nor solid and a world of apartness when everything is connected. Cult-controlled 'science' dismisses the so-called 'paranormal' and all phenomena related to that when the 'para'-normal is perfectly normal and explains the alleged 'great mysteries' which dumbfound scientific minds. There is a reason for this. A 'scientific mind' in terms of the mainstream is a material mind, a five-sense mind imprisoned in see it, touch it, hear it, smell it and taste it. Phenomena and happenings that can't be explained that way leave the 'scientific mind' bewildered and the rule is that if they

can't account for why something is happening then it can't, by definition, be happening. I beg to differ. Telepathy is thought waves passing through The Field (think wave disturbance again) to be decoded by someone able to connect with that wavelength (information). For example: You can pick up the thought waves of a friend at any distance and at the very least that will bring them to mind. A few minutes later the friend calls you. 'My god', you say, 'that's incredible – I was just thinking of you.' Ah, but *they* were thinking of *you* before they made the call and that's what you decoded. Native peoples not entrapped in five-sense reality do this so well it became known as the 'bush telegraph'. Those known as psychics and mediums (genuine ones) are doing the same only across dimensions of reality. 'Mind over matter' comes from the fact that matter and mind are the *same*. The state of one influences the state of the other. Indeed one *and* the other are illusions. They are aspects of the same field. Paranormal phenomena are all explainable so why are they still considered 'mysteries' or not happening? Once you go down this road of understanding you begin to expand awareness beyond the five senses and that's the nightmare for the Cult.



Figure 13: Holograms are not solid, but the best ones appear to be.

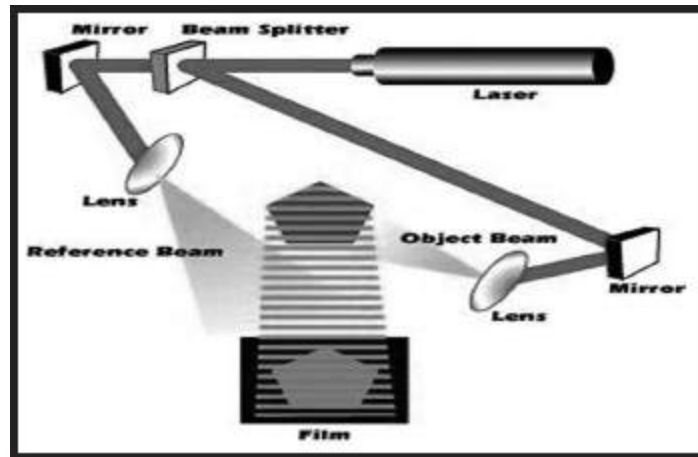


Figure 14: How holograms are created by capturing a waveform version of the subject image.

Holographic 'solidity'

Our reality is not solid, it is holographic. We are now well aware of holograms which are widely used today. Two-dimensional information is decoded into a three-dimensional reality that is not solid although can very much appear to be (Fig 13). Holograms are created with a laser divided into two parts. One goes directly onto a holographic photographic print ('reference beam') and the other takes a waveform image of the subject ('working beam') before being directed onto the print where it 'collides' with the other half of the laser (Fig 14). This creates a *waveform* interference pattern which contains the wavefield information of whatever is being photographed (Fig 15 overleaf). The process can be likened to dropping pebbles in a pond. Waves generated by each one spread out across the water to collide with the others and create a wave representation of where the stones fell and at what speed, weight and distance. A waveform interference pattern of a hologram is akin to the waveform information in The Field which the five senses decode into electrical signals to be decoded by the brain into a holographic illusory 'physical' reality. In the same way when a laser (think human attention) is directed at the waveform interference pattern a three-dimensional version of the subject is projected into apparently 'solid' reality (Fig 16). An amazing trait of holograms reveals more 'paranormal mysteries'. Information of the *whole*

hologram is encoded in waveform in every part of the interference pattern by the way they are created. This means that every *part* of a hologram is a smaller version of the whole. Cut the interference wave-pattern into four and you won't get four parts of the image. You get quarter-sized versions of the *whole* image. The body is a hologram and the same applies. Here we have the basis of acupuncture, reflexology and other forms of healing which identify representations of the whole body in all of the parts, hands, feet, ears, everywhere. Skilled palm readers can do what they do because the information of whole body is encoded in the hand. The concept of as above, so below, comes from this.



Figure 15: A waveform interference pattern that holds the information that transforms into a hologram.



Figure 16: Holographic people including 'Elvis' holographically inserted to sing a duet with Celine Dion.

The question will be asked of why, if solidity is illusory, we can't just walk through walls and each other. The resistance is not solid against solid; it is electromagnetic field against electromagnetic field and we decode this into the *experience* of solid against solid. We should also not underestimate the power of belief to dictate reality. What you believe is impossible *will be*. Your belief impacts on your decoding processes and they won't decode what you think is impossible. What we believe we perceive and what we perceive we experience. 'Can't dos' and 'impossibles' are like a firewall in a computer system that won't put on the screen what the firewall blocks. How vital that is to understanding how human experience has been hijacked. I explain in *The Answer, Everything You Need To Know But Have Never Been Told* and other books a long list of 'mysteries' and 'paranormal' phenomena that are not mysterious and perfectly normal once you realise what reality is and how it works. 'Ghosts' can be seen to pass through 'solid' walls because the walls are not solid and the ghost is a discarnate entity operating on a frequency so different to that of the wall that it's like two radio stations sharing the same space while never interfering with each other. I have seen ghosts do this myself. The apartness of people and objects is also an illusion. Everything is connected by the Field like all sea life is connected by the sea. It's just that within the limits of our visual reality we only 'see' holographic information and not the field of information that connects everything and from which the holographic world is made manifest. If you can only see holographic 'objects' and not the field that connects them they will appear to you as unconnected to each other in the same way that we see the computer while not seeing the Wi-Fi.

What you don't know *can* hurt you

Okay, we return to those 'two worlds' of human society and the Cult with its global network of interconnecting secret societies and satanic groups which manipulate through governments, corporations, media, religions, etc. The fundamental difference between them is *knowledge*. The idea has been to keep humanity

ignorant of the plan for its total enslavement underpinned by a crucial ignorance of reality – who we are and where we are – and how we interact with it. ‘Human’ should be the interaction between our expanded eternal consciousness and the five-sense body experience. We are meant to be *in* this world in terms of the five senses but not *of* this world in relation to our greater consciousness and perspective. In that state we experience the small picture of the five senses within the wider context of the big picture of awareness beyond the five senses. Put another way the five senses see the dots and expanded awareness connects them into pictures and patterns that give context to the apparently random and unconnected. Without the context of expanded awareness the five senses see only apartness and randomness with apparently no meaning. The Cult and its other-dimensional controllers seek to intervene in the frequency realm where five-sense reality is supposed to connect with expanded reality and to keep the two apart (more on this in the final chapter). When that happens five-sense mental and emotional processes are no longer influenced by expanded awareness, or the True ‘I’, and instead are driven by the isolated perceptions of the body’s decoding systems. They are in the world *and* of it. Here we have the human plight and why humanity with its potential for infinite awareness can be so easily manipulatable and descend into such extremes of stupidity.

Once the Cult isolates five-sense mind from expanded awareness it can then program the mind with perceptions and beliefs by controlling information that the mind receives through the ‘education’ system of the formative years and the media perceptual bombardment and censorship of an entire lifetime. Limit perception and a sense of the possible through limiting knowledge by limiting and skewing information while censoring and discrediting that which could set people free. As the title of another of my books says ... *And The Truth Shall Set You Free*. For this reason the last thing the Cult wants in circulation is the truth about anything – especially the reality of the eternal ‘I’ – and that’s why it is desperate to control information. The Cult knows that information becomes perception

which becomes behaviour which, collectively, becomes human society. Cult-controlled and funded mainstream 'science' denies the existence of an eternal 'I' and seeks to dismiss and trash all evidence to the contrary. Cult-controlled mainstream religion has a version of 'God' that is little more than a system of control and dictatorship that employs threats of damnation in an afterlife to control perceptions and behaviour in the here and now through fear and guilt. Neither is true and it's the 'neither' that the Cult wishes to suppress. This 'neither' is that everything is an expression, a point of attention, within an infinite state of consciousness which is the real meaning of the term 'God'.

Perceptual obsession with the 'physical body' and five-senses means that 'God' becomes personified as a bearded bloke sitting among the clouds or a raging bully who loves us if we do what 'he' wants and condemns us to the fires of hell if we don't. These are no more than a 'spiritual' fairy tales to control and dictate events and behaviour through fear of this 'God' which has bizarrely made 'God-fearing' in religious circles a state to be desired. I would suggest that fearing *anything* is not to be encouraged and celebrated, but rather deleted. You can see why 'God fearing' is so beneficial to the Cult and its religions when *they* decide what 'God' wants and what 'God' demands (the Cult demands) that everyone do. As the great American comedian Bill Hicks said satirising a Christian zealot: 'I think what God meant to say.' How much of this infinite awareness ('God') that we access is decided by how far we choose to expand our perceptions, self-identity and sense of the possible. The scale of self-identity reflects itself in the scale of awareness that we can connect with and are influenced by – how much knowing and insight we have instead of programmed perception. You cannot expand your awareness into the infinity of possibility when you believe that you are little me Peter the postman or Mary in marketing and nothing more. I'll deal with this in the concluding chapter because it's crucial to how we turnaround current events.

Where the Cult came from

When I realised in the early 1990s there was a Cult network behind global events I asked the obvious question: When did it start? I took it back to ancient Rome and Egypt and on to Babylon and Sumer in Mesopotamia, the 'Land Between Two Rivers', in what we now call Iraq. The two rivers are the Tigris and Euphrates and this region is of immense historical and other importance to the Cult, as is the land called Israel only 550 miles away by air. There is much more going on with deep esoteric meaning across this whole region. It's not only about 'wars for oil'. Priceless artefacts from Mesopotamia were stolen or destroyed after the American and British invasion of Iraq in 2003 justified by the lies of Boy Bush and Tony Blair (their Cult masters) about non-existent 'weapons of mass destruction'.

Mesopotamia was the location of Sumer (about 5,400BC to 1,750BC), and Babylon (about 2,350BC to 539BC). Sabbatians may have become immensely influential in the Cult in modern times but they are part of a network that goes back into the mists of history. Sumer is said by historians to be the 'cradle of civilisation'. I disagree. I say it was the re-start of what we call human civilisation after cataclysmic events symbolised in part as the 'Great Flood' destroyed the world that existed before. These fantastic upheavals that I have been describing in detail in the books since the early 1990s appear in accounts and legends of ancient cultures across the world and they are supported by geological and biological evidence. Stone tablets found in Iraq detailing the Sumer period say the cataclysms were caused by non-human 'gods' they call the Anunnaki. These are described in terms of extraterrestrial visitations in which knowledge supplied by the Anunnaki is said to have been the source of at least one of the world's oldest writing systems and developments in astronomy, mathematics and architecture that were way ahead of their time. I have covered this subject at length in *The Biggest Secret* and *Children of the Matrix* and the same basic 'Anunnaki' story can be found in Zulu accounts in South Africa where the late and very great Zulu high shaman Credo Mutwa told me that the Sumerian Anunnaki were known by Zulus as the Chitauri or 'children of the serpent'. See my six-hour video interview with Credo on this subject entitled *The*

Reptilian Agenda recorded at his then home near Johannesburg in 1999 which you can watch on the Ickonic media platform.

The Cult emerged out of Sumer, Babylon and Egypt (and elsewhere) and established the Roman Empire before expanding with the Romans into northern Europe from where many empires were savagely imposed in the form of Cult-controlled societies all over the world. Mass death and destruction was their calling card. The Cult established its centre of operations in Europe and European Empires were Cult empires which allowed it to expand into a global force. Spanish and Portuguese colonialists headed for Central and South America while the British and French targeted North America. Africa was colonised by Britain, France, Belgium, the Netherlands, Portugal, Spain, Italy, and Germany. Some like Britain and France moved in on the Middle East. The British Empire was by far the biggest for a simple reason. By now Britain was the headquarters of the Cult from which it expanded to form Canada, the United States, Australia and New Zealand. The Sun never set on the British Empire such was the scale of its occupation. London remains a global centre for the Cult along with Rome and the Vatican although others have emerged in Israel and China. It is no accident that the 'virus' is alleged to have come out of China while Italy was chosen as the means to terrify the Western population into compliance with 'Covid' fascism. Nor that Israel has led the world in 'Covid' fascism and mass 'vaccination'.

You would think that I would mention the United States here, but while it has been an important means of imposing the Cult's will it is less significant than would appear and is currently in the process of having what power it does have deleted. The Cult in Europe has mostly loaded the guns for the US to fire. America has been controlled from Europe from the start through Cult operatives in Britain and Europe. The American Revolution was an illusion to make it appear that America was governing itself while very different forces were pulling the strings in the form of Cult families such as the Rothschilds through the Rockefellers and other subordinates. The Rockefellers are extremely close to Bill Gates and

established both scalpel and drug 'medicine' and the World Health Organization. They play a major role in the development and circulation of vaccines through the Rockefeller Foundation on which Bill Gates said his Foundation is based. Why wouldn't this be the case when the Rockefellers and Gates are on the same team? Cult infiltration of human society goes way back into what we call history and has been constantly expanding and centralising power with the goal of establishing a global structure to dictate everything. Look how this has been advanced in great leaps with the 'Covid' hoax.

The non-human dimension

I researched and observed the comings and goings of Cult operatives through the centuries and even thousands of years as they were born, worked to promote the agenda within the secret society and satanic networks, and then died for others to replace them. Clearly there had to be a coordinating force that spanned this entire period while operatives who would not have seen the end goal in their lifetimes came and went advancing the plan over millennia. I went in search of that coordinating force with the usual support from the extraordinary synchronicity of my life which has been an almost daily experience since 1990. I saw common themes in religious texts and ancient cultures about a non-human force manipulating human society from the hidden. Christianity calls this force Satan, the Devil and demons; Islam refers to the Jinn or Djinn; Zulus have their Chitauri (spelt in other ways in different parts of Africa); and the Gnostic people in Egypt in the period around and before 400AD referred to this phenomena as the 'Archons', a word meaning rulers in Greek. Central American cultures speak of the 'Predators' among other names and the same theme is everywhere. I will use 'Archons' as a collective name for all of them. When you see how their nature and behaviour is described all these different sources are clearly talking about the same force. Gnostics described the Archons in terms of 'luminous fire' while Islam relates the Jinn to 'smokeless fire'. Some refer to beings in form that could occasionally be seen, but the most common of common theme is that they operate from

unseen realms which means almost all existence to the visual processes of humans. I had concluded that this was indeed the foundation of human control and that the Cult was operating within the human frequency band on behalf of this hidden force when I came across the writings of Gnostics which supported my conclusions in the most extraordinary way.

A sealed earthen jar was found in 1945 near the town of Nag Hammadi about 75-80 miles north of Luxor on the banks of the River Nile in Egypt. Inside was a treasure trove of manuscripts and texts left by the Gnostic people some 1,600 years earlier. They included 13 leather-bound papyrus codices (manuscripts) and more than 50 texts written in Coptic Egyptian estimated to have been hidden in the jar in the period of 400AD although the source of the information goes back much further. Gnostics oversaw the Great or Royal Library of Alexandria, the fantastic depository of ancient texts detailing advanced knowledge and accounts of human history. The Library was dismantled and destroyed in stages over a long period with the death-blow delivered by the Cult-established Roman Church in the period around 415AD. The Church of Rome was the Church of Babylon relocated as I said earlier. Gnostics were not a race. They were a way of perceiving reality. Whenever they established themselves and their information circulated the terrorists of the Church of Rome would target them for destruction. This happened with the Great Library and with the Gnostic Cathars who were burned to death by the psychopaths after a long period of oppression at the siege of the Castle of Monségur in southern France in 1244. The Church has always been terrified of Gnostic information which demolishes the official Christian narrative although there is much in the Bible that supports the Gnostic view if you read it in another way. To anyone studying the texts of what became known as the Nag Hammadi Library it is clear that great swathes of Christian and Biblical belief has its origin with Gnostics sources going back to Sumer. Gnostic themes have been twisted to manipulate the perceived reality of Bible believers. Biblical texts have been in the open for centuries where they could be changed while Gnostic

documents found at Nag Hammadi were sealed away and untouched for 1,600 years. What you see is what they wrote.

Use your *pneuma* not your *nous*

Gnosticism and Gnostic come from 'gnosis' which means knowledge, or rather *secret* knowledge, in the sense of spiritual awareness – knowledge about reality and life itself. The desperation of the Cult's Church of Rome to destroy the Gnostics can be understood when the knowledge they were circulating was the last thing the Cult wanted the population to know. Sixteen hundred years later the same Cult is working hard to undermine and silence me for the same reason. The dynamic between knowledge and ignorance is a constant. 'Time' appears to move on, but essential themes remain the same. We are told to 'use your nous', a Gnostic word for head/brain/intelligence. They said, however, that spiritual awakening or 'salvation' could only be secured by expanding awareness *beyond* what they called *nous* and into *pneuma* or Infinite Self. Obviously as I read these texts the parallels with what I have been saying since 1990 were fascinating to me. There is a universal truth that spans human history and in that case why wouldn't we be talking the same language 16 centuries apart? When you free yourself from the perception program of the five senses and explore expanded realms of consciousness you are going to connect with the same information no matter what the perceived 'era' within a manufactured timeline of a single and tiny range of manipulated frequency. Humans working with 'smart' technology or knocking rocks together in caves is only a timeline appearing to operate within the human frequency band. Expanded awareness and the knowledge it holds have always been there whether the era be Stone Age or computer age. We can only access that knowledge by opening ourselves to its frequency which the five-sense prison cell is designed to stop us doing. Gates, Fauci, Whitty, Vallance, Zuckerberg, Brin, Page, Wojcicki, Bezos, and all the others behind the 'Covid' hoax clearly have a long wait before their range of frequency can make that connection given that an open heart is

crucial to that as we shall see. Instead of accessing knowledge directly through expanded awareness it is given to Cult operatives by the secret society networks of the Cult where it has been passed on over thousands of years outside the public arena. Expanded realms of consciousness is where great artists, composers and writers find their inspiration and where truth awaits anyone open enough to connect with it. We need to go there fast.

Archon hijack

A fifth of the Nag Hammadi texts describe the existence and manipulation of the Archons led by a 'Chief Archon' they call 'Yaldabaoth', or the 'Demiurge', and this is the Christian 'Devil', 'Satan', 'Lucifer', and his demons. Archons in Biblical symbolism are the 'fallen ones' which are also referred to as fallen angels after the angels expelled from heaven according to the Abrahamic religions of Judaism, Christianity and Islam. These angels are claimed to tempt humans to 'sin' ongoing and you will see how accurate that symbolism is during the rest of the book. The theme of 'original sin' is related to the 'Fall' when Adam and Eve were 'tempted by the serpent' and fell from a state of innocence and 'obedience' (connection) with God into a state of disobedience (disconnection). The Fall is said to have brought sin into the world and corrupted everything including human nature. Yaldabaoth, the 'Lord Archon', is described by Gnostics as a 'counterfeit spirit', 'The Blind One', 'The Blind God', and 'The Foolish One'. The Jewish name for Yaldabaoth in Talmudic writings is Samael which translates as 'Poison of God', or 'Blindness of God'. You see the parallels. Yaldabaoth in Islamic belief is the Muslim Jinn devil known as Shaytan – Shaytan is Satan as the same themes are found all over the world in every religion and culture. The 'Lord God' of the Old Testament is the 'Lord Archon' of Gnostic manuscripts and that's why he's such a bloodthirsty bastard. Satan is known by Christians as 'the Demon of Demons' and Gnostics called Yaldabaoth the 'Archon of Archons'. Both are known as 'The Deceiver'. We are talking about the same 'bloke' for sure and these common themes

using different names, storylines and symbolism tell a common tale of the human plight.

Archons are referred to in Nag Hammadi documents as mind parasites, inverters, guards, gatekeepers, detainers, judges, pitiless ones and deceivers. The 'Covid' hoax alone is a glaring example of all these things. The Biblical 'God' is so different in the Old and New Testaments because they are not describing the same phenomenon. The vindictive, angry, hate-filled, 'God' of the Old Testament, known as Yahweh, is Yaldabaoth who is depicted in Cult-dictated popular culture as the 'Dark Lord', 'Lord of Time', Lord (Darth) Vader and Dormammu, the evil ruler of the 'Dark Dimension' trying to take over the 'Earth Dimension' in the Marvel comic movie, *Dr Strange*. Yaldabaoth is both the Old Testament 'god' and the Biblical 'Satan'. Gnostics referred to Yaldabaoth as the 'Great Architect of the Universe' and the Cult-controlled Freemason network calls their god 'the 'Great Architect of the Universe' (also Grand Architect). The 'Great Architect' Yaldabaoth is symbolised by the Cult as the all-seeing eye at the top of the pyramid on the Great Seal of the United States and the dollar bill. Archon is encoded in *arch*-itect as it is in *arch*-angels and *arch*-bishops. All religions have the theme of a force for good and force for evil in some sort of spiritual war and there is a reason for that – the theme is true. The Cult and its non-human masters are quite happy for this to circulate. They present themselves as the force for good fighting evil when they are really the force of evil (absence of love). The whole foundation of Cult modus operandi is inversion. They promote themselves as a force for good and anyone challenging them in pursuit of peace, love, fairness, truth and justice is condemned as a satanic force for evil. This has been the game plan throughout history whether the Church of Rome inquisitions of non-believers or 'conspiracy theorists' and 'anti-vaxxers' of today. The technique is the same whatever the timeline era.

Yaldabaoth is revolting (true)

Yaldabaoth and the Archons are said to have revolted against God with Yaldabaoth claiming to *be* God – the *All That Is*. The Old Testament ‘God’ (Yaldabaoth) demanded to be worshipped as such: ‘*I am the LORD, and there is none else, there is no God beside me*’ (Isaiah 45:5). I have quoted in other books a man who said he was the unofficial son of the late Baron Philippe de Rothschild of the Mouton-Rothschild wine producing estates in France who died in 1988 and he told me about the Rothschild ‘revolt from God’. The man said he was given the name Phillip Eugene de Rothschild and we shared long correspondence many years ago while he was living under another identity. He said that he was conceived through ‘occult incest’ which (within the Cult) was ‘normal and to be admired’. ‘Phillip’ told me about his experience attending satanic rituals with rich and famous people whom he names and you can see them and the wider background to Cult Satanism in my other books starting with *The Biggest Secret*. Cult rituals are interactions with Archontic ‘gods’. ‘Phillip’ described Baron Philippe de Rothschild as ‘a master Satanist and hater of God’ and he used the same term ‘revolt from God’ associated with Yaldabaoth/Satan/Lucifer/the Devil in describing the Sabbatian Rothschild dynasty. ‘I played a key role in my family’s revolt from God’, he said. That role was to infiltrate in classic Sabbatian style the Christian Church, but eventually he escaped the mind-prison to live another life. The Cult has been targeting religion in a plan to make worship of the Archons the global one-world religion. Infiltration of Satanism into modern ‘culture’, especially among the young, through music videos, stage shows and other means, is all part of this.

Nag Hammadi texts describe Yaldabaoth and the Archons in their prime form as energy – consciousness – and say they can take form if they choose in the same way that consciousness takes form as a human. Yaldabaoth is called ‘formless’ and represents a deeply inverted, distorted and chaotic state of consciousness which seeks to attached to humans and turn them into a likeness of itself in an attempt at assimilation. For that to happen it has to manipulate

humans into low frequency mental and emotional states that match its own. Archons can certainly appear in human form and this is the origin of the psychopathic personality. The energetic distortion Gnostics called Yaldabaoth is psychopathy. When psychopathic Archons take human form that human will be a psychopath as an expression of Yaldabaoth consciousness. Cult psychopaths are Archons in human form. The principle is the same as that portrayed in the 2009 *Avatar* movie when the American military travelled to a fictional Earth-like moon called Pandora in the Alpha Centauri star system to infiltrate a society of blue people, or Na'vi, by hiding within bodies that looked like the Na'vi. Archons posing as humans have a particular hybrid information field, part human, part Archon, (the ancient 'demigods') which processes information in a way that manifests behaviour to match their psychopathic evil, lack of empathy and compassion, and stops them being influenced by the empathy, compassion and love that a fully-human information field is capable of expressing. Cult bloodlines interbreed, be they royalty or dark suits, for this reason and you have their obsession with incest. Interbreeding with full-blown humans would dilute the Archontic energy field that guarantees psychopathy in its representatives in the human realm.

Gnostic writings say the main non-human forms that Archons take are *serpentine* (what I have called for decades 'reptilian' amid unbounded ridicule from the Archontically-programmed) and what Gnostics describe as 'an unborn baby or foetus with grey skin and dark, unmoving eyes'. This is an excellent representation of the ET 'Greys' of UFO folklore which large numbers of people claim to have seen and been abducted by – Zulu shaman Credo Mutwa among them. I agree with those that believe in extraterrestrial or interdimensional visitations today and for thousands of years past. No wonder with their advanced knowledge and technological capability they were perceived and worshipped as gods for technological and other 'miracles' they appeared to perform. Imagine someone arriving in a culture disconnected from the modern world with a smartphone and computer. They would be

seen as a 'god' capable of 'miracles'. The Renegade Mind, however, wants to know the source of everything and not only the way that source manifests as human or non-human. In the same way that a Renegade Mind seeks the original source material for the 'Covid virus' to see if what is claimed is true. The original source of Archons in form is consciousness – the distorted state of consciousness known to Gnostics as Yaldabaoth.

'Revolt from God' is energetic disconnection

Where I am going next will make a lot of sense of religious texts and ancient legends relating to 'Satan', Lucifer' and the 'gods'. Gnostic descriptions sync perfectly with the themes of my own research over the years in how they describe a consciousness distortion seeking to impose itself on human consciousness. I've referred to the core of infinite awareness in previous books as Infinite Awareness in Awareness of Itself. By that I mean a level of awareness that knows that it is all awareness and is aware of all awareness. From here comes the frequency of love in its true sense and balance which is what love is on one level – the balance of all forces into a single whole called Oneness and Isness. The more we disconnect from this state of love that many call 'God' the constituent parts of that Oneness start to unravel and express themselves as a part and not a whole. They become individualised as intellect, mind, selfishness, hatred, envy, desire for power over others, and such like. This is not a problem in the greater scheme in that 'God', the *All That Is*, can experience all these possibilities through different expressions of itself including humans. What we as expressions of the whole experience the *All That Is* experiences. We are the *All That Is* experiencing itself. As we withdraw from that state of Oneness we disconnect from its influence and things can get very unpleasant and very stupid. Archontic consciousness is at the extreme end of that. It has so disconnected from the influence of Oneness that it has become an inversion of unity and love, an inversion of everything, an inversion of life itself. Evil is appropriately live written backwards. Archontic consciousness is obsessed with death, an inversion of life,

and so its manifestations in Satanism are obsessed with death. They use inverted symbols in their rituals such as the inverted pentagram and cross. Sabbatians as Archontic consciousness incarnate invert Judaism and every other religion and culture they infiltrate. They seek disunity and chaos and they fear unity and harmony as they fear love like garlic to a vampire. As a result the Cult, Archons incarnate, act with such evil, psychopathy and lack of empathy and compassion disconnected as they are from the source of love. How could Bill Gates and the rest of the Archontic psychopaths do what they have to human society in the 'Covid' era with all the death, suffering and destruction involved and have no emotional consequence for the impact on others? Now you know. Why have Zuckerberg, Brin, Page, Wojcicki and company callously censored information warning about the dangers of the 'vaccine' while thousands have been dying and having severe, sometimes life-changing reactions? Now you know. Why have Tedros, Fauci, Whitty, Vallance and their like around the world been using case and death figures they're aware are fraudulent to justify lockdowns and all the deaths and destroyed lives that have come from that? Now you know. Why did Christian Drosten produce and promote a 'testing' protocol that he knew couldn't test for infectious disease which led to a global human catastrophe. Now you know. The Archontic mind doesn't give a shit ([Fig 17](#)). I personally think that Gates and major Cult insiders are a form of AI cyborg that the Archons want humans to become.

these.' The statement was true in all respects. We do live in a technologically-generated virtual reality simulation (more very shortly) and we have been manipulated to be an energy source for Archontic consciousness. The Disney-Pixar animated movie *Monsters, Inc.* in 2001 symbolised the dynamic when monsters in their world had no energy source and they would enter the human world to terrify children in their beds, catch the child's scream, terror (low-vibrational frequencies), and take that energy back to power the monster world. The lead character you might remember was a single giant eye and the symbolism of the Cult's all-seeing eye was obvious. Every thought and emotion is broadcast as a frequency unique to that thought and emotion. Feelings of love and joy, empathy and compassion, are high, quick, frequencies while fear, depression, anxiety, suffering and hate are low, slow, dense frequencies. Which kind do you think Archontic consciousness can connect with and absorb? In such a low and dense frequency state there's no way it can connect with the energy of love and joy. Archons can only feed off energy compatible with their own frequency and they and their Cult agents want to delete the human world of love and joy and manipulate the transmission of low vibrational frequencies through low-vibrational human mental and emotional states. *We are their energy source.* Wars are energetic banquets to the Archons – a world war even more so – and think how much low-frequency mental and emotional energy has been generated from the consequences for humanity of the 'Covid' hoax orchestrated by Archons incarnate like Gates.

The ancient practice of human sacrifice 'to the gods', continued in secret today by the Cult, is based on the same principle. 'The gods' are Archontic consciousness in different forms and the sacrifice is induced into a state of intense terror to generate the energy the Archontic frequency can absorb. Incarnate Archons in the ritual drink the blood which contains an adrenaline they crave which floods into the bloodstream when people are terrorised. Most of the sacrifices, ancient and modern, are children and the theme of 'sacrificing young virgins to the gods' is just code for children. They

have a particular pre-puberty energy that Archons want more than anything and the energy of the young in general is their target. The California Department of Education wants students to chant the names of Aztec gods (Archontic gods) once worshipped in human sacrifice rituals in a curriculum designed to encourage them to 'challenge racist, bigoted, discriminatory, imperialist/colonial beliefs', join 'social movements that struggle for social justice', and 'build new possibilities for a post-racist, post-systemic racism society'. It's the usual Woke crap that inverts racism and calls it anti-racism. In this case solidarity with 'indigenous tribes' is being used as an excuse to chant the names of 'gods' to which people were sacrificed (and still are in secret). What an example of Woke's inability to see beyond black and white, us and them, They condemn the colonisation of these tribal cultures by Europeans (quite right), but those cultures sacrificing people including children to their 'gods', and mass murdering untold numbers as the Aztecs did, is just fine. One chant is to the Aztec god Tezcatlipoca who had a man sacrificed to him in the 5th month of the Aztec calendar. His heart was cut out and he was eaten. Oh, that's okay then. Come on children ... after three ... Other sacrificial 'gods' for the young to chant their allegiance include Quetzalcoatl, Huitzilopochtli and Xipe Totec. The curriculum says that 'chants, affirmations, and energizers can be used to bring the class together, build unity around ethnic studies principles and values, and to reinvigorate the class following a lesson that may be emotionally taxing or even when student engagement may appear to be low'. Well, that's the cover story, anyway. Chanting and mantras are the repetition of a particular frequency generated from the vocal cords and chanting the names of these Archontic 'gods' tunes you into their frequency. That is the last thing you want when it allows for energetic synchronisation, attachment and perceptual influence. Initiates chant the names of their 'Gods' in their rituals for this very reason.

Vampires of the Woke

Paedophilia is another way that Archons absorb the energy of children. Paedophiles possessed by Archontic consciousness are used as the conduit during sexual abuse for discarnate Archons to vampire the energy of the young they desire so much. Stupendous numbers of children disappear every year never to be seen again although you would never know from the media. Imagine how much low-vibrational energy has been generated by children during the 'Covid' hoax when so many have become depressed and psychologically destroyed to the point of killing themselves. Shocking numbers of children are now taken by the state from loving parents to be handed to others. I can tell you from long experience of researching this since 1996 that many end up with paedophiles and assets of the Cult through corrupt and Cult-owned social services which in the reframing era has hired many psychopaths and emotionless automatons to do the job. Children are even stolen to order using spurious reasons to take them by the corrupt and secret (because they're corrupt) 'family courts'. I have written in detail in other books, starting with *The Biggest Secret* in 1997, about the ubiquitous connections between the political, corporate, government, intelligence and military elites (Cult operatives) and Satanism and paedophilia. If you go deep enough both networks have an interlocking leadership. The Woke mentality has been developed by the Cult for many reasons: To promote almost every aspect of its agenda; to hijack the traditional political left and turn it fascist; to divide and rule; and to target agenda pushbackers. But there are other reasons which relate to what I am describing here. How many happy and joyful Wokers do you ever see especially at the extreme end? They are a mental and psychological mess consumed by emotional stress and constantly emotionally cocked for the next explosion of indignation at someone referring to a female as a female. They are walking, talking, batteries as Morpheus might say emitting frequencies which both enslave them in low-vibrational bubbles of perceptual limitation and feed the Archons. Add to this the hatred claimed to be love; fascism claimed to 'anti-fascism', racism claimed to be 'anti-racism';

exclusion claimed to inclusion; and the abuse-filled Internet trolling. You have a purpose-built Archontic energy system with not a wind turbine in sight and all founded on Archontic *inversion*. We have whole generations now manipulated to serve the Archons with their actions and energy. They will be doing so their entire adult lives unless they snap out of their Archon-induced trance. Is it really a surprise that Cult billionaires and corporations put so much money their way? Where is the energy of joy and laughter, including laughing at yourself which is confirmation of your own emotional security? Mark Twain said: 'The human race has one really effective weapon, and that is laughter.' We must use it all the time. Woke has destroyed comedy because it has no humour, no joy, sense of irony, or self-deprecation. Its energy is dense and intense. *Mmmmm*, lunch says the Archontic frequency. Rudolf Steiner (1861-1925) was the Austrian philosopher and famous esoteric thinker who established Waldorf education or Steiner schools to treat children like unique expressions of consciousness and not minds to be programmed with the perceptions determined by authority. I'd been writing about this energy vampiring for decades when I was sent in 2016 a quote by Steiner. He was spot on:

There are beings in the spiritual realms for whom anxiety and fear emanating from human beings offer welcome food. When humans have no anxiety and fear, then these creatures starve. If fear and anxiety radiates from people and they break out in panic, then these creatures find welcome nutrition and they become more and more powerful. These beings are hostile towards humanity. Everything that feeds on negative feelings, on anxiety, fear and superstition, despair or doubt, are in reality hostile forces in super-sensible worlds, launching cruel attacks on human beings, while they are being fed ... These are exactly the feelings that belong to contemporary culture and materialism; because it estranges people from the spiritual world, it is especially suited to evoke hopelessness and fear of the unknown in people, thereby calling up the above mentioned hostile forces against them.

Pause for a moment from this perspective and reflect on what has happened in the world since the start of 2020. Not only will pennies drop, but billion dollar bills. We see the same theme from Don Juan Matus, a Yaqui Indian shaman in Mexico and the information source for Peruvian-born writer, Carlos Castaneda, who wrote a series of

books from the 1960s to 1990s. Don Juan described the force manipulating human society and his name for the Archons was the predator:

We have a predator that came from the depths of the cosmos and took over the rule of our lives. Human beings are its prisoners. The predator is our lord and master. It has rendered us docile, helpless. If we want to protest, it suppresses our protest. If we want to act independently, it demands that we don't do so ... indeed we are held prisoner!

They took us over because we are food to them, and they squeeze us mercilessly because we are their sustenance. Just as we rear chickens in coops, the predators rear us in human coops, humaneros. Therefore, their food is always available to them.

Different cultures, different eras, same recurring theme.

The 'ennoia' dilemma

Nag Hammadi Gnostic manuscripts say that Archon consciousness has no 'ennoia'. This is directly translated as 'intentionality', but I'll use the term 'creative imagination'. The *All That Is* in awareness of itself is the source of all creativity – all possibility – and the more disconnected you are from that source the more you are subsequently denied 'creative imagination'. Given that Archon consciousness is almost entirely disconnected it severely lacks creativity and has to rely on far more mechanical processes of thought and exploit the creative potential of those that do have 'ennoia'. You can see cases of this throughout human society. Archon consciousness almost entirely dominates the global banking system and if we study how that system works you will appreciate what I mean. Banks manifest 'money' out of nothing by issuing lines of 'credit' which is 'money' that has never, does not, and will never exist except in theory. It's a confidence trick. If you think 'credit' figures-on-a-screen 'money' is worth anything you accept it as payment. If you don't then the whole system collapses through lack of confidence in the value of that 'money'. Archontic bankers with no 'ennoia' are 'lending' 'money' that doesn't exist to humans that *do* have creativity – those that have the inspired ideas and create businesses and products. Archon banking feeds off human creativity

which it controls through 'money' creation and debt. Humans have the creativity and Archons exploit that for their own benefit and control while having none themselves. Archon Internet platforms like Facebook claim joint copyright of everything that creative users post and while Archontic minds like Zuckerberg may officially head that company it will be human creatives on the staff that provide the creative inspiration. When you have limitless 'money' you can then buy other companies established by creative humans. Witness the acquisition record of Facebook, Google and their like. Survey the Archon-controlled music industry and you see non-creative dark suit executives making their fortune from the human creativity of their artists. The cases are endless. Research the history of people like Gates and Zuckerberg and how their empires were built on exploiting the creativity of others. Archon minds cannot create out of nothing, but they are skilled (because they have to be) in what Gnostic texts call 'countermimicry'. They can imitate, but not innovate. Sabbatians trawl the creativity of others through backdoors they install in computer systems through their cybersecurity systems. Archon-controlled China is globally infamous for stealing intellectual property and I remember how Hong Kong, now part of China, became notorious for making counterfeit copies of the creativity of others – 'countermimicry'. With the now pervasive and all-seeing surveillance systems able to infiltrate any computer you can appreciate the potential for Archons to vampire the creativity of humans. Author John Lamb Lash wrote in his book about the Nag Hammadi texts, *Not In His Image*:

Although they cannot originate anything, because they lack the divine factor of ennoia (intentionality), Archons can imitate with a vengeance. Their expertise is simulation (HAL, virtual reality). The Demiurge [Yaldabaoth] fashions a heaven world copied from the fractal patterns [of the original] ... His construction is celestial kitsch, like the fake Italianate villa of a Mafia don complete with militant angels to guard every portal.

This brings us to something that I have been speaking about since the turn of the millennium. Our reality is a simulation; a virtual reality that we think is real. No, I'm not kidding.

Human reality? Well, virtually

I had pondered for years about whether our reality is 'real' or some kind of construct. I remembered being immensely affected on a visit as a small child in the late 1950s to the then newly-opened Planetarium on the Marylebone Road in London which is now closed and part of the adjacent Madame Tussauds wax museum. It was in the middle of the day, but when the lights went out there was the night sky projected in the Planetarium's domed ceiling and it appeared to be so real. The experience never left me and I didn't know why until around the turn of the millennium when I became certain that our 'night sky' and entire reality is a projection, a virtual reality, akin to the illusory world portrayed in the *Matrix* movies. I looked at the sky one day in this period and it appeared to me like the domed roof of the Planetarium. The release of the first *Matrix* movie in 1999 also provided a synchronistic and perfect visual representation of where my mind had been going for a long time. I hadn't come across the Gnostic Nag Hammadi texts then. When I did years later the correlation was once again astounding. As I read Gnostic accounts from 1,600 years and more earlier it was clear that they were describing the same simulation phenomenon. They tell how the Yaldabaoth 'Demiurge' and Archons created a 'bad copy' of original reality to rule over all that were captured by its illusions and the body was a prison to trap consciousness in the 'bad copy' fake reality. Read how Gnostics describe the 'bad copy' and update that to current times and they are referring to what we would call today a virtual reality simulation.

Author John Lamb Lash said 'the Demiurge fashions a heaven world copied from the fractal patterns' of the original through expertise in 'HAL' or virtual reality simulation. Fractal patterns are part of the energetic information construct of our reality, a sort of blueprint. If these patterns were copied in computer terms it would indeed give you a copy of a 'natural' reality in a non-natural frequency and digital form. The principle is the same as making a copy of a website. The original website still exists, but now you can change the copy version to make it whatever you like and it can

become very different to the original website. Archons have done this with our reality, a *synthetic* copy of prime reality that still exists beyond the frequency walls of the simulation. Trapped within the illusions of this synthetic Matrix, however, were and are human consciousness and other expressions of prime reality and this is why the Archons via the Cult are seeking to make the human body synthetic and give us synthetic AI minds to complete the job of turning the entire reality synthetic including what we perceive to be the natural world. To quote Kurzweil: 'Nanobots will infuse all the matter around us with information. Rocks, trees, everything will become these intelligent creatures.' Yes, *synthetic* 'creatures' just as 'Covid' and other genetically-manipulating 'vaccines' are designed to make the human body synthetic. From this perspective it is obvious why Archons and their Cult are so desperate to infuse synthetic material into every human with their 'Covid' scam.

Let there be (electromagnetic) light

Yaldabaoth, the force that created the simulation, or Matrix, makes sense of the Gnostic reference to 'The Great Architect' and its use by Cult Freemasonry as the name of its deity. The designer of the Matrix in the movies is called 'The Architect' and that trilogy is jam-packed with symbolism relating to these subjects. I have contended for years that the angry Old Testament God (Yaldabaoth) is the 'God' being symbolically 'quoted' in the opening of Genesis as 'creating the world'. This is not the creation of prime reality – it's the creation of the *simulation*. The Genesis 'God' says: 'Let there be Light: and there was light.' But what is this 'Light'? I have said for decades that the speed of light (186,000 miles per second) is not the fastest speed possible as claimed by mainstream science and is in fact the frequency walls or outer limits of the Matrix. You can't have a fastest or slowest anything within all possibility when everything is possible. The human body is encoded to operate within the speed of light or *within the simulation* and thus we see only the tiny frequency band of visible *light*. Near-death experiencers who perceive reality outside the body during temporary 'death' describe a very different

form of light and this is supported by the Nag Hammadi texts. Prime reality beyond the simulation ('Upper Aeons' to the Gnostics) is described as a realm of incredible beauty, bliss, love and harmony – a realm of 'watery light' that is so powerful 'there are no shadows'. Our false reality of Archon control, which Gnostics call the 'Lower Aeons', is depicted as a realm with a different kind of 'light' and described in terms of chaos, 'Hell', 'the Abyss' and 'Outer Darkness', where trapped souls are tormented and manipulated by demons (relate that to the 'Covid' hoax alone). The watery light theme can be found in near-death accounts and it is not the same as *simulation* 'light' which is electromagnetic or radiation light within the speed of light – the 'Lower Aeons'. Simulation 'light' is the 'luminous fire' associated by Gnostics with the Archons. The Bible refers to Yaldabaoth as 'that old serpent, called the Devil, and Satan, which deceiveth the whole world' (Revelation 12:9). I think that making a simulated copy of prime reality ('countermimicry') and changing it dramatically while all the time manipulating humanity to believe it to be real could probably meet the criteria of deceiving the whole world. Then we come to the Cult god Lucifer – the *Light Bringer*. Lucifer is symbolic of Yaldabaoth, the bringer of radiation light that forms the bad copy simulation within the speed of light. 'He' is symbolised by the lighted torch held by the Statue of Liberty and in the name 'Illuminati'. Sabbatian-Frankism declares that Lucifer is the true god and Lucifer is the real god of Freemasonry honoured as their 'Great or Grand Architect of the Universe' (simulation).

I would emphasise, too, the way Archontic technologically-generated luminous fire of radiation has deluged our environment since I was a kid in the 1950s and changed the nature of The Field with which we constantly interact. Through that interaction technological radiation is changing us. The Smart Grid is designed to operate with immense levels of communication power with 5G expanding across the world and 6G, 7G, in the process of development. Radiation is the simulation and the Archontic manipulation system. Why wouldn't the Archon Cult wish to unleash radiation upon us to an ever-greater extreme to form

Kurzweil's 'cloud'? The plan for a synthetic human is related to the need to cope with levels of radiation beyond even anything we've seen so far. Biological humans would not survive the scale of radiation they have in their script. The Smart Grid is a technological sub-reality within the technological simulation to further disconnect five-sense perception from expanded consciousness. It's a technological prison of the mind.

Infusing the 'spirit of darkness'

A recurring theme in religion and native cultures is the manipulation of human genetics by a non-human force and most famously recorded as the biblical 'sons of god' (the gods plural in the original) who interbred with the daughters of men. The Nag Hammadi *Apocryphon of John* tells the same story this way:

He [Yaldabaoth] sent his angels [Archons/demons] to the daughters of men, that they might take some of them for themselves and raise offspring for their enjoyment. And at first they did not succeed. When they had no success, they gathered together again and they made a plan together ... And the angels changed themselves in their likeness into the likeness of their mates, filling them with the spirit of darkness, which they had mixed for them, and with evil ... And they took women and begot children out of the darkness according to the likeness of their spirit.

Possession when a discarnate entity takes over a human body is an age-old theme and continues today. It's very real and I've seen it. Satanic and secret society rituals can create an energetic environment in which entities can attach to initiates and I've heard many stories of how people have changed their personality after being initiated even into lower levels of the Freemasons. I have been inside three Freemasonic temples, one at a public open day and two by just walking in when there was no one around to stop me. They were in Ryde, the town where I live, Birmingham, England, when I was with a group, and Boston, Massachusetts. They all felt the same energetically – dark, dense, low-vibrational and sinister. Demonic attachment can happen while the initiate has no idea what is going on. To them it's just a ritual to get in the Masons and do a bit of good

business. In the far more extreme rituals of Satanism human possession is even more powerful and they are designed to make possession possible. The hierarchy of the Cult is dictated by the power and perceived status of the possessing Archon. In this way the Archon hierarchy becomes the Cult hierarchy. Once the entity has attached it can influence perception and behaviour and if it attaches to the extreme then so much of its energy (information) infuses into the body information field that the hologram starts to reflect the nature of the possessing entity. This is the *Exorcist* movie type of possession when facial features change and it's known as shapeshifting. Islam's Jinn are said to be invisible tricksters who change shape, 'whisper', confuse and take human form. These are all traits of the Archons and other versions of the same phenomenon. Extreme possession could certainly infuse the 'spirit of darkness' into a partner during sex as the Nag Hammadi texts appear to describe. Such an infusion can change genetics which is also energetic information. Human genetics is information and the 'spirit of darkness' is information. Mix one with the other and change must happen. Islam has the concept of a 'Jinn baby' through possession of the mother and by Jinn taking human form. There are many ways that human genetics can be changed and remember that Archons have been aware all along of advanced techniques to do this. What is being done in human society today – and far more – was known about by Archons at the time of the 'fallen ones' and their other versions described in religions and cultures.

Archons and their human-world Cult are obsessed with genetics as we see today and they know this dictates how information is processed into perceived reality during a human life. They needed to produce a human form that would decode the simulation and this is symbolically known as 'Adam and Eve' who left the 'garden' (prime reality) and 'fell' into Matrix reality. The simulation is not a 'physical' construct (there is no 'physical'); it is a source of information. Think Wi-Fi again. The simulation is an energetic field encoded with information and body-brain systems are designed to decode that information encoded in wave or frequency form which

is transmitted to the brain as electrical signals. These are decoded by the brain to construct our sense of reality – an illusory ‘physical’ world that only exists in the brain or the mind. Virtual reality games mimic this process using the same sensory decoding system. Information is fed to the senses to decode a virtual reality that can appear so real, but isn’t (Figs 18 and 19). Some scientists believe – and I agree with them – that what we perceive as ‘physical’ reality only exists when we are looking or observing. The act of perception or focus triggers the decoding systems which turn waveform information into holographic reality. When we are not observing something our reality reverts from a holographic state to a waveform state. This relates to the same principle as a falling tree not making a noise unless someone is there to hear it or decode it. The concept makes sense from the simulation perspective. A computer is not decoding all the information in a Wi-Fi field all the time and only decodes or brings into reality on the screen that part of Wi-Fi that it’s decoding – focusing upon – at that moment.



Figure 18: Virtual reality technology ‘hacks’ into the body’s five-sense decoding system.



Figure 19: The result can be experienced as very ‘real’.

Interestingly, Professor Donald Hoffman at the Department of Cognitive Sciences at the University of California, Irvine, says that our experienced reality is like a computer interface that shows us only the level with which we interact while hiding all that exists beyond it: 'Evolution shaped us with a user interface that hides the truth. Nothing that we see is the truth – the very language of space and time and objects is the wrong language to describe reality.' He is correct in what he says on so many levels. Space and time are not a universal reality. They are a phenomenon of decoded *simulation* reality as part of the process of enslaving our sense of reality. Near-death experiencers report again and again how space and time did not exist as we perceive them once they were free of the body – body decoding systems. You can appreciate from this why Archons and their Cult are so desperate to entrap human attention in the five senses where we are in the Matrix and of the Matrix. Opening your mind to expanded states of awareness takes you beyond the information confines of the simulation and you become aware of knowledge and insights denied to you before. This is what we call 'awakening' – *awakening from the Matrix* – and in the final chapter I will relate this to current events.

Where are the 'aliens'?

A simulation would explain the so-called 'Fermi Paradox' named after Italian physicist Enrico Fermi (1901-1954) who created the first nuclear reactor. He considered the question of why there is such a lack of extraterrestrial activity when there are so many stars and planets in an apparently vast universe; but what if the night sky that we see, or think we do, is a simulated projection as I say? If you control the simulation and your aim is to hold humanity fast in essential ignorance would you want other forms of life including advanced life coming and going sharing information with humanity? Or would you want them to believe they were isolated and apparently alone? Themes of human isolation and apartness are common whether they be the perception of a lifeless universe or the fascist isolation laws of the 'Covid' era. Paradoxically the very

existence of a simulation means that we are not alone when some force had to construct it. My view is that experiences that people have reported all over the world for centuries with Reptilians and Grey entities are Archon phenomena as Nag Hammadi texts describe; and that benevolent 'alien' interactions are non-human groups that come in and out of the simulation by overcoming Archon attempts to keep them out. It should be highlighted, too, that Reptilians and Greys are obsessed with *genetics* and *technology* as related by cultural accounts and those who say they have been abducted by them. Technology is their way of overcoming some of the limitations in their creative potential and our technology-driven and controlled human society of today is *archetypical* Archon-Reptilian-Grey *modus operandi*. Technocracy is really *Archontocracy*. The Universe does not have to be as big as it appears with a simulation. There is no space or distance only information decoded into holographic reality. What we call 'space' is only the absence of holographic 'objects' and that 'space' is The Field of energetic information which connects everything into a single whole. The same applies with the artificially-generated information field of the simulation. The Universe is not big or small as a physical reality. It is decoded information, that's all, and its perceived size is decided by the way the simulation is encoded to make it appear. The entire night sky as we perceive it only exists in our brain and so where are those 'millions of light years'? The 'stars' on the ceiling of the Planetarium looked a vast distance away.

There's another point to mention about 'aliens'. I have been highlighting since the 1990s the plan to stage a fake 'alien invasion' to justify the centralisation of global power and a world military. Nazi scientist Werner von Braun, who was taken to America by Operation Paperclip after World War Two to help found NASA, told his American assistant Dr Carol Rosin about the Cult agenda when he knew he was dying in 1977. Rosin said that he told her about a sequence that would lead to total human control by a one-world government. This included threats from terrorism, rogue nations, meteors and asteroids before finally an 'alien invasion'. All of these

things, von Braun said, would be bogus and what I would refer to as a No-Problem-Reaction-Solution. Keep this in mind when 'the aliens are coming' is the new mantra. The aliens are not coming – they are *already here* and they have infiltrated human society while looking human. French-Canadian investigative journalist Serge Monast said in 1994 that he had uncovered a NASA/military operation called Project Blue Beam which fits with what Werner von Braun predicted. Monast died of a 'heart attack' in 1996 the day after he was arrested and spent a night in prison. He was 51. He said Blue Beam was a plan to stage an alien invasion that would include religious figures beamed holographically into the sky as part of a global manipulation to usher in a 'new age' of worshipping what I would say is the Cult 'god' Yaldabaoth in a one-world religion. Fake holographic asteroids are also said to be part of the plan which again syncs with von Braun. How could you stage an illusory threat from asteroids unless they were holographic inserts? This is pretty straightforward given the advanced technology outside the public arena and the fact that our 'physical' reality is holographic anyway. Information fields would be projected and we would decode them into the illusion of a 'physical' asteroid. If they can sell a global 'pandemic' with a 'virus' that doesn't exist what will humans not believe if government and media tell them?

All this is particularly relevant as I write with the Pentagon planning to release in June, 2021, information about 'UFO sightings'. I have been following the UFO story since the early 1990s and the common theme throughout has been government and military denials and cover up. More recently, however, the Pentagon has suddenly become more talkative and apparently open with Air Force pilot radar images released of unexplained craft moving and changing direction at speeds well beyond anything believed possible with human technology. Then, in March, 2021, former Director of National Intelligence John Ratcliffe said a Pentagon report months later in June would reveal a great deal of information about UFO sightings unknown to the public. He said the report would have 'massive implications'. The order to do this was included bizarrely

in a \$2.3 trillion 'coronavirus' relief and government funding bill passed by the Trump administration at the end of 2020. I would add some serious notes of caution here. I have been pointing out since the 1990s that the US military and intelligence networks have long had craft – 'flying saucers' or anti-gravity craft – which any observer would take to be extraterrestrial in origin. Keeping this knowledge from the public allows craft flown by *humans* to be perceived as alien visitations. I am not saying that 'aliens' do not exist. I would be the last one to say that, but we have to be streetwise here. President Ronald Reagan told the UN General Assembly in 1987: 'I occasionally think how quickly our differences worldwide would vanish if we were facing an alien threat from outside this world.' That's the idea. Unite against a common 'enemy' with a common purpose behind your 'saviour force' (the Cult) as this age-old technique of mass manipulation goes global.

Science moves this way ...

I could find only one other person who was discussing the simulation hypothesis publicly when I concluded it was real. This was Nick Bostrom, a Swedish-born philosopher at the University of Oxford, who has explored for many years the possibility that human reality is a computer simulation although his version and mine are not the same. Today the simulation and holographic reality hypothesis have increasingly entered the scientific mainstream. Well, the more open-minded mainstream, that is. Here are a few of the ever-gathering examples. American nuclear physicist Silas Beane led a team of physicists at the University of Bonn in Germany pursuing the question of whether we live in a simulation. They concluded that we probably do and it was likely based on a lattice of cubes. They found that cosmic rays align with that specific pattern. The team highlighted the Greisen–Zatsepin–Kuzmin (GZK) limit which refers to cosmic ray particle interaction with cosmic background radiation that creates an apparent boundary for cosmic ray particles. They say in a paper entitled 'Constraints on the Universe as a Numerical Simulation' that this 'pattern of constraint' is exactly what you

would find with a computer simulation. They also made the point that a simulation would create its own 'laws of physics' that would limit possibility. I've been making the same point for decades that the *perceived* laws of physics relate only to this reality, or what I would later call the simulation. When designers write codes to create computer and virtual reality games they are the equivalent of the laws of physics for that game. Players interact within the limitations laid out by the coding. In the same way those who wrote the codes for the simulation decided the laws of physics that would apply. These can be overridden by expanded states of consciousness, but not by those enslaved in only five-sense awareness where simulation codes rule. Overriding the codes is what people call 'miracles'. They are not. They are bypassing the encoded limits of the simulation. A population caught in simulation perception would have no idea that this was their plight. As the Bonn paper said: 'Like a prisoner in a pitch-black cell we would not be able to see the "walls" of our prison,' That's true if people remain mesmerised by the five senses. Open to expanded awareness and those walls become very clear. The main one is the speed of light.

American theoretical physicist James Gates is another who has explored the simulation question and found considerable evidence to support the idea. Gates was Professor of Physics at the University of Maryland, Director of The Center for String and Particle Theory, and on Barack Obama's Council of Advisors on Science and Technology. He and his team found *computer codes* of digital data embedded in the fabric of our reality. They relate to on-off electrical charges of 1 and 0 in the binary system used by computers. 'We have no idea what they are doing there', Gates said. They found within the energetic fabric mathematical sequences known as error-correcting codes or block codes that 'reboot' data to its original state or 'default settings' when something knocks it out of sync. Gates was asked if he had found a set of equations embedded in our reality indistinguishable from those that drive search engines and browsers and he said: 'That is correct.' Rich Terrile, director of the Centre for Evolutionary Computation and Automated Design at NASA's Jet

Propulsion Laboratory, has said publicly that he believes the Universe is a digital hologram that must have been created by a form of intelligence. I agree with that in every way. Waveform information is delivered electrically by the senses to the brain which constructs a *digital* holographic reality that we call the 'world'. This digital level of reality can be read by the esoteric art of numerology. Digital holograms are at the cutting edge of holographics today. We have digital technology everywhere designed to access and manipulate our digital level of perceived reality. Synthetic mRNA in 'Covid vaccines' has a digital component to manipulate the body's digital 'operating system'.

Reality is numbers

How many know that our reality can be broken down to numbers and codes that are the same as computer games? Max Tegmark, a physicist at the Massachusetts Institute of Technology (MIT), is the author of *Our Mathematical Universe* in which he lays out how reality can be entirely described by numbers and maths in the way that a video game is encoded with the 'physics' of computer games. Our world and computer virtual reality are essentially the same.

Tegmark imagines the perceptions of characters in an advanced computer game when the graphics are so good they don't know they are in a game. They think they can bump into real objects (electromagnetic resistance in our reality), fall in love and feel emotions like excitement. When they began to study the apparently 'physical world' of the video game they would realise that everything was made of pixels (which have been found in our energetic reality as must be the case when on one level our world is digital). What computer game characters thought was physical 'stuff', Tegmark said, could actually be broken down into numbers:

And we're exactly in this situation in our world. We look around and it doesn't seem that mathematical at all, but everything we see is made out of elementary particles like quarks and electrons. And what properties does an electron have? Does it have a smell or a colour or a texture? No! ... We physicists have come up with geeky names for [Electron] properties, like

electric charge, or spin, or lepton number, but the electron doesn't care what we call it, the properties are just numbers.

This is the illusory reality Gnostics were describing. This is the simulation. The A, C, G, and T codes of DNA have a binary value – A and C = 0 while G and T = 1. This has to be when the simulation is digital and the body must be digital to interact with it. Recurring mathematical sequences are encoded throughout reality and the body. They include the Fibonacci sequence in which the two previous numbers are added to get the next one, as in ... 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, etc. The sequence is encoded in the human face and body, proportions of animals, DNA, seed heads, pine cones, trees, shells, spiral galaxies, hurricanes and the number of petals in a flower. The list goes on and on. There are fractal patterns – a 'never-ending pattern that is infinitely complex and self-similar across all scales in the as above, so below, principle of holograms. These and other famous recurring geometrical and mathematical sequences such as Phi, Pi, Golden Mean, Golden Ratio and Golden Section are *computer codes* of the simulation. I had to laugh and give my head a shake the day I finished this book and it went into the production stage. I was sent an article in *Scientific American* published in April, 2021, with the headline 'Confirmed! We Live in a Simulation'. Two decades after I first said our reality is a simulation and the speed of light is its outer limit the article suggested that we do live in a simulation and that the speed of light is its outer limit. I left school at 15 and never passed a major exam in my life while the writer was up to his eyes in qualifications. As I will explain in the final chapter *knowing* is far better than thinking and they come from very different sources. The article rightly connected the speed of light to the processing speed of the 'Matrix' and said what has been in my books all this time ... 'If we are in a simulation, as it appears, then space is an abstract property written in code. It is not real'. No it's not and if we live in a simulation something created it and it wasn't *us*. 'That David Icke says we are manipulated by aliens' – he's crackers.'

Wow ...

The reality that humanity thinks is so real is an illusion. Politicians, governments, scientists, doctors, academics, law enforcement, media, school and university curriculums, on and on, are all founded on a world that *does not exist* except as a simulated prison cell. Is it such a stretch to accept that 'Covid' doesn't exist when our entire 'physical' reality doesn't exist? Revealed here is the knowledge kept under raps in the Cult networks of compartmentalised secrecy to control humanity's sense of reality by inducing the population to believe in a reality that's not real. If it wasn't so tragic in its experiential consequences the whole thing would be hysterically funny. None of this is new to Renegade Minds. Ancient Greek philosopher Plato (about 428 to about 347BC) was a major influence on Gnostic belief and he described the human plight thousands of years ago with his Allegory of the Cave. He told the symbolic story of prisoners living in a cave who had never been outside. They were chained and could only see one wall of the cave while behind them was a fire that they could not see. Figures walked past the fire casting shadows on the prisoners' wall and those moving shadows became their sense of reality. Some prisoners began to study the shadows and were considered experts on them (today's academics and scientists), but what they studied was only an illusion (today's academics and scientists). A prisoner escaped from the cave and saw reality as it really is. When he returned to report this revelation they didn't believe him, called him mad and threatened to kill him if he tried to set them free. Plato's tale is not only a brilliant analogy of the human plight and our illusory reality. It describes, too, the dynamics of the 'Covid' hoax. I have only skimmed the surface of these subjects here. The aim of this book is to crisply connect all essential dots to put what is happening today into its true context. All subject areas and their connections in this chapter are covered in great evidential detail in *Everything You Need To Know, But Have Never Been Told* and *The Answer*.

They say that bewildered people 'can't see the forest for the trees'. Humanity, however, can't see the forest for the *twigs*. The five senses

see only twigs while Renegade Minds can see the forest and it's the forest where the answers lie with the connections that reveals.

Breaking free of perceptual programming so the forest can be seen is the way we turn all this around. Not breaking free is how humanity got into this mess. The situation may seem hopeless, but I promise you it's not. We are a perceptual heartbeat from paradise if only we knew.

CHAPTER TWELVE

Escaping Wetiko

Life is simply a vacation from the infinite
Dean Cavanagh

Renegade Minds weave the web of life and events and see common themes in the apparently random. They are always there if you look for them and their pursuit is aided by incredible synchronicity that comes when your mind is open rather than mesmerised by what it thinks it can see.

Infinite awareness is infinite possibility and the more of infinite possibility that we access the more becomes infinitely possible. That may be stating the apparently obvious, but it is a devastatingly-powerful fact that can set us free. We are a point of attention within an infinity of consciousness. The question is how much of that infinity do we choose to access? How much knowledge, insight, awareness, wisdom, do we want to connect with and explore? If your focus is only in the five senses you will be influenced by a fraction of infinite awareness. I mean a range so tiny that it gives new meaning to infinitesimal. Limitation of self-identity and a sense of the possible limit accordingly your range of consciousness. We are what we think we are. Life is what we think it is. The dream is the dreamer and the dreamer is the dream. Buddhist philosophy puts it this way: 'As a thing is viewed, so it appears.' Most humans live in the realm of touch, taste, see, hear, and smell and that's the limit of their sense of the possible and sense of self. Many will follow a religion and speak of a God in his heaven, but their lives are still

dominated by the five senses in their perceptions and actions. The five senses become the arbiter of everything. When that happens all except a smear of infinity is sealed away from influence by the rigid, unyielding, reality bubbles that are the five-sense human or Phantom Self. Archon Cult methodology is to isolate consciousness within five-sense reality – the simulation – and then program that consciousness with a sense of self and the world through a deluge of life-long information designed to instil the desired perception that allows global control. Efforts to do this have increased dramatically with identity politics as identity bubbles are squeezed into the minutiae of five-sense detail which disconnect people even more profoundly from the infinite 'I'.

Five-sense focus and self-identity are like a firewall that limits access to the infinite realms. You only perceive one radio or television station and no other. We'll take that literally for a moment. Imagine a vast array of stations giving different information and angles on reality, but you only ever listen to one. Here we have the human plight in which the population is overwhelmingly confined to CultFM. This relates only to the frequency range of CultFM and limits perception and insight to that band – limits *possibility* to that band. It means you are connecting with an almost imperceptibly minuscule range of possibility and creative potential within the infinite Field. It's a world where everything seems apart from everything else and where synchronicity is rare. Synchronicity is defined in the dictionary as 'the happening by chance of two or more related or similar events at the same time'. Use of 'by chance' betrays a complete misunderstanding of reality. Synchronicity is not 'by chance'. As people open their minds, or 'awaken' to use the term, they notice more and more coincidences in their lives, bits of 'luck', apparently miraculous happenings that put them in the right place at the right time with the right people. Days become peppered with 'fancy meeting you here' and 'what are the chances of that?' My entire life has been lived like this and ever more so since my own colossal awakening in 1990 and 91 which transformed my sense of reality. Synchronicity is not 'by chance'; it is by accessing expanded

realms of possibility which allow expanded potential for manifestation. People broadcasting the same vibe from the same openness of mind tend to be drawn 'by chance' to each other through what I call frequency magnetism and it's not only people. In the last more than 30 years incredible synchronicity has also led me through the Cult maze to information in so many forms and to crucial personal experiences. These 'coincidences' have allowed me to put the puzzle pieces together across an enormous array of subjects and situations. Those who have breached the bubble of five-sense reality will know exactly what I mean and this escape from the perceptual prison cell is open to everyone whenever they make that choice. This may appear super-human when compared with the limitations of 'human', but it's really our natural state. 'Human' as currently experienced is consciousness in an unnatural state of induced separation from the infinity of the whole. I'll come to how this transformation into unity can be made when I have described in more detail the force that holds humanity in servitude by denying this access to infinite self.

The Wetiko factor

I have been talking and writing for decades about the way five-sense mind is systematically barricaded from expanded awareness. I have used the analogy of a computer (five-sense mind) and someone at the keyboard (expanded awareness). Interaction between the computer and the operator is symbolic of the interaction between five-sense mind and expanded awareness. The computer directly experiences the Internet and the operator experiences the Internet via the computer which is how it's supposed to be – the two working as one. Archons seek to control that point where the operator connects with the computer to stop that interaction ([Fig 20](#)). Now the operator is banging the keyboard and clicking the mouse, but the computer is not responding and this happens when the computer is taken over – *possessed* – by an appropriately-named computer 'virus'. The operator has lost all influence over the computer which goes its own way making decisions under the control of the 'virus'. I have

just described the dynamic through which the force known to Gnostics as Yaldabaoth and Archons disconnects five-sense mind from expanded awareness to imprison humanity in perceptual servitude.



Figure 20: The mind ‘virus’ I have been writing about for decades seeks to isolate five-sense mind (the computer) from the true ‘I’. (Image by Neil Hague).

About a year ago I came across a Native American concept of Wetiko which describes precisely the same phenomenon. Wetiko is the spelling used by the Cree and there are other versions including wintiko and windigo used by other tribal groups. They spell the name with lower case, but I see Wetiko as a proper noun as with Archons and prefer a capital. I first saw an article about Wetiko by writer and researcher Paul Levy which so synced with what I had been writing about the computer/operator disconnection and later the Archons. I then read his book, the fascinating *Dispelling Wetiko, Breaking the Spell of Evil*. The parallels between what I had concluded long before and the Native American concept of Wetiko were so clear and obvious that it was almost funny. For Wetiko see the Gnostic Archons for sure and the Jinn, the Predators, and every other name for a force of evil, inversion and chaos. Wetiko is the Native American name for the force that divides the computer from

the operator (Fig 21). Indigenous author Jack D. Forbes, a founder of the Native American movement in the 1960s, wrote another book about Wetiko entitled *Columbus And Other Cannibals – The Wetiko Disease of Exploitation, Imperialism, and Terrorism* which I also read. Forbes says that Wetiko refers to an evil person or spirit ‘who terrorizes other creatures by means of terrible acts, including cannibalism’. Zulu shaman Credo Mutwa told me that African accounts tell how cannibalism was brought into the world by the Chitauri ‘gods’ – another manifestation of Wetiko. The distinction between ‘evil person or spirit’ relates to Archons/Wetiko possessing a human or acting as pure consciousness. Wetiko is said to be a sickness of the soul or spirit and a state of being that takes but gives nothing back – the Cult and its operatives perfectly described. Black Hawk, a Native American war leader defending their lands from confiscation, said European invaders had ‘poisoned hearts’ – Wetiko hearts – and that this would spread to native societies. Mention of the heart is very significant as we shall shortly see. Forbes writes: ‘Tragically, the history of the world for the past 2,000 years is, in great part, the story of the epidemiology of the wetiko disease.’ Yes, and much longer. Forbes is correct when he says: ‘The wetikos destroyed Egypt and Babylon and Athens and Rome and Tenochtitlan [capital of the Aztec empire] and perhaps now they will destroy the entire earth.’ Evil, he said, is the number one export of a Wetiko culture – see its globalisation with ‘Covid’. Constant war, mass murder, suffering of all kinds, child abuse, Satanism, torture and human sacrifice are all expressions of Wetiko and the Wetiko possessed. The world is Wetiko made manifest, *but it doesn’t have to be*. There is a way out of this even now.



Figure 21: The mind 'virus' is known to Native Americans as 'Wetiko'. (Image by Neil Hague).

Cult of Wetiko

Wetiko is the Yaldabaoth frequency distortion that seeks to attach to human consciousness and absorb it into its own. Once this connection is made Wetiko can drive the perceptions of the target which they believe to be coming from their own mind. All the horrors of history and today from mass killers to Satanists, paedophiles like Jeffrey Epstein and other psychopaths, are the embodiment of Wetiko and express its state of being in all its grotesqueness. The Cult is Wetiko incarnate, Yaldabaoth incarnate, and it seeks to facilitate Wetiko assimilation of humanity in totality into its distortion by manipulating the population into low frequency states that match its own. Paul Levy writes:

'Holographically enforced within the psyche of every human being the wetiko virus pervades and underlies the entire field of consciousness, and can therefore potentially manifest through any one of us at any moment if we are not mindful.' The 'Covid' hoax has achieved this with many people, but others have not fallen into Wetiko's frequency lair. Players in the 'Covid' human catastrophe including Gates, Schwab, Tedros, Fauci, Whitty, Vallance, Johnson, Hancock, Ferguson, Drosten, and all the rest, including the psychopath psychologists, are expressions of Wetiko. This is why

they have no compassion or empathy and no emotional consequence for what they do that would make them stop doing it. Observe all the people who support the psychopaths in authority against the Pushbackers despite the damaging impact the psychopaths have on their own lives and their family's lives. You are again looking at Wetiko possession which prevents them seeing through the lies to the obvious scam going on. *Why can't they see it?* Wetiko won't let them see it. The perceptual divide that has now become a chasm is between the Wetikoed and the non-Wetikoed.

Paul Levy describes Wetiko in the same way that I have long described the Archontic force. They are the same distorted consciousness operating across dimensions of reality: '... the subtle body of wetiko is not located in the third dimension of space and time, literally existing in another dimension ... it is able to affect ordinary lives by mysteriously interpenetrating into our three-dimensional world.' Wetiko does this through its incarnate representatives in the Cult and by weaving itself into The Field which on our level of reality is the electromagnetic information field of the simulation or Matrix. More than that, the simulation *is* Wetiko / Yaldabaoth. Caleb Scharf, Director of Astrobiology at Columbia University, has speculated that 'alien life' could be so advanced that it has transcribed itself into the quantum realm to become what we call physics. He said intelligence indistinguishable from the fabric of the Universe would solve many of its greatest mysteries:

Perhaps hyper-advanced life isn't just external. Perhaps it's already all around. It is embedded in what we perceive to be physics itself, from the root behaviour of particles and fields to the phenomena of complexity and emergence ... In other words, life might not just be in the equations. It might BE the equations [My emphasis].

Scharf said it is possible that 'we don't recognise advanced life because it forms an integral and unsuspecting part of what we've considered to be the natural world'. I agree. Wetiko/Yaldabaoth *is* the simulation. We are literally in the body of the beast. But that doesn't mean it has to control us. We all have the power to overcome Wetiko

influence and the Cult knows that. I doubt it sleeps too well because it knows that.

Which Field?

This, I suggest, is how it all works. There are two Fields. One is the fierce electromagnetic light of the Matrix within the speed of light; the other is the 'watery light' of The Field beyond the walls of the Matrix that connects with the Great Infinity. Five-sense mind and the decoding systems of the body attach us to the Field of Matrix light. They have to or we could not experience this reality. Five-sense mind sees only the Matrix Field of information while our expanded consciousness is part of the Infinity Field. When we open our minds, and most importantly our hearts, to the Infinity Field we have a mission control which gives us an expanded perspective, a road map, to understand the nature of the five-sense world. If we are isolated only in five-sense mind there is no mission control. We're on our own trying to understand a world that's constantly feeding us information to ensure we do not understand. People in this state can feel 'lost' and bewildered with no direction or radar. You can see ever more clearly those who are influenced by the Fields of Big Infinity or little five-sense mind simply by their views and behaviour with regard to the 'Covid' hoax. We have had this division throughout known human history with the mass of the people on one side and individuals who could see and intuit beyond the walls of the simulation – Plato's prisoner who broke out of the cave and saw reality for what it is. Such people have always been targeted by Wetiko/Archon-possessed authority, burned at the stake or demonised as mad, bad and dangerous. The Cult today and its global network of 'anti-hate', 'anti-fascist' Woke groups are all expressions of Wetiko attacking those exposing the conspiracy, 'Covid' lies and the 'vaccine' agenda.

Woke as a whole is Wetiko which explains its black and white mentality and how at one it is with the Wetiko-possessed Cult. Paul Levy said: 'To be in this paradigm is to still be under the thrall of a two-valued logic – where things are either true or false – of a

wetikoized mind.’ Wetiko consciousness is in a permanent rage, therefore so is Woke, and then there is Woke inversion and contradiction. ‘Anti-fascists’ act like fascists because fascists *and* ‘anti-fascists’ are both Wetiko at work. Political parties act the same while claiming to be different for the same reason. Secret society and satanic rituals are attaching initiates to Wetiko and the cold, ruthless, psychopathic mentality that secures the positions of power all over the world is Wetiko. Reframing ‘training programmes’ have the same cumulative effect of attaching Wetiko and we have their graduates described as automatons and robots with a cold, psychopathic, uncaring demeanour. They are all traits of Wetiko possession and look how many times they have been described in this book and elsewhere with regard to personnel behind ‘Covid’ including the police and medical profession. Climbing the greasy pole in any profession in a Wetiko society requires traits of Wetiko to get there and that is particularly true of politics which is not about fair competition and pre-eminence of ideas. It is founded on how many backs you can stab and arses you can lick. This culminated in the global ‘Covid’ coordination between the Wetiko possessed who pulled it off in all the different countries without a trace of empathy and compassion for their impact on humans. Our sight sense can see only holographic form and not the Field which connects holographic form. Therefore we perceive ‘physical’ objects with ‘space’ in between. In fact that ‘space’ is energy/consciousness operating on multiple frequencies. One of them is Wetiko and that connects the Cult psychopaths, those who submit to the psychopaths, and those who serve the psychopaths in the media operations of the world. Wetiko is Gates. Wetiko is the mask-wearing submissive. Wetiko is the fake journalist and ‘fact-checker’. The Wetiko Field is coordinating the whole thing. Psychopaths, gofers, media operatives, ‘anti-hate’ hate groups, ‘fact-checkers’ and submissive people work as one unit *even without human coordination* because they are attached to the *same* Field which is organising it all ([Fig 22](#)). Paul Levy is here describing how Wetiko-possessed people are drawn together and refuse to let any information breach their rigid

perceptions. He was writing long before 'Covid', but I think you will recognise followers of the 'Covid' religion *oh just a little bit*:

People who are channelling the vibratory frequency of wetiko align with each other through psychic resonance to reinforce their unspoken shared agreement so as to uphold their deranged view of reality. Once an unconscious content takes possession of certain individuals, it irresistibly draws them together by mutual attraction and knits them into groups tied together by their shared madness that can easily swell into an avalanche of insanity.

A psychic epidemic is a closed system, which is to say that it is insular and not open to any new information or informing influences from the outside world which contradict its fixed, limited, and limiting perspective.

There we have the Woke mind and the 'Covid' mind. Compatible resonance draws the awakening together, too, which is clearly happening today.

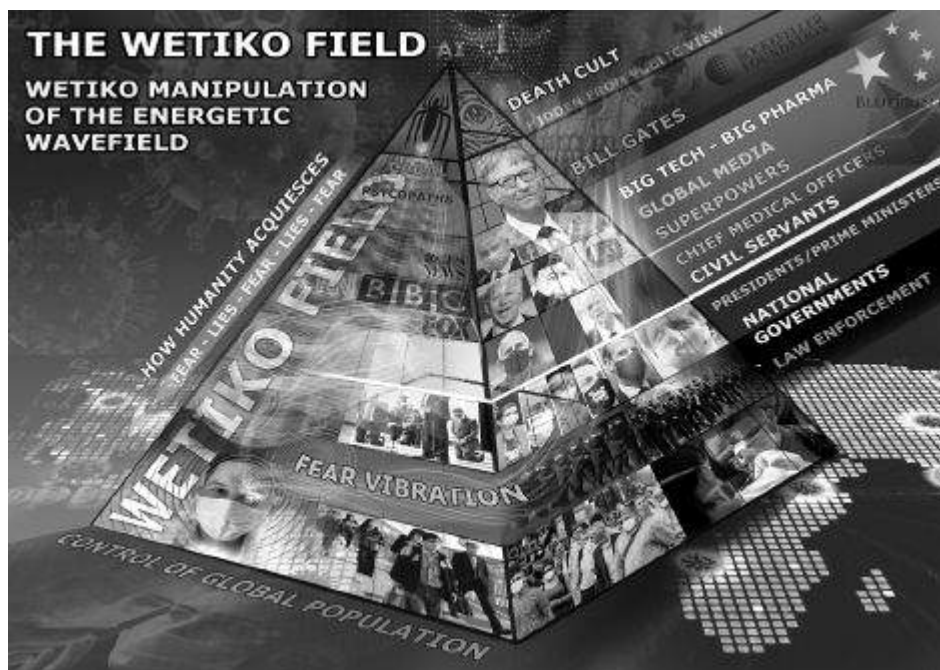


Figure 22: The Wetiko Field from which the Cult pyramid and its personnel are made manifest. (Image by Neil Hague).

Spiritual servitude

Wetiko doesn't care about humans. It's not human; it just possesses humans for its own ends and the effect (depending on the scale of

possession) can be anything from extreme psychopathy to unquestioning obedience. Wetiko's worst nightmare is for human consciousness to expand beyond the simulation. Everything is focussed on stopping that happening through control of information, thus perception, thus frequency. The 'education system', media, science, medicine, academia, are all geared to maintaining humanity in five-sense servitude as is the constant stimulation of low-vibrational mental and emotional states (see 'Covid'). Wetiko seeks to dominate those subconscious spaces between five-sense perception and expanded consciousness where the computer meets the operator. From these subconscious hiding places Wetiko speaks to us to trigger urges and desires that we take to be our own and manipulate us into anything from low-vibrational to psychopathic states. Remember how Islam describes the Jinn as invisible tricksters that 'whisper' and confuse. Wetiko is the origin of the 'trickster god' theme that you find in cultures all over the world. Jinn, like the Archons, are Wetiko which is terrified of humans awakening and reconnecting with our true self for then its energy source has gone. With that the feedback loop breaks between Wetiko and human perception that provides the energetic momentum on which its very existence depends as a force of evil. Humans are both its target and its source of survival, but only if we are operating in low-vibrational states of fear, hate, depression and the background anxiety that most people suffer. We are Wetiko's target because we are its key to survival. It needs us, not the other way round. Paul Levy writes:

A vampire has no intrinsic, independent, substantial existence in its own right; it only exists in relation to us. The pathogenic, vampiric mind-parasite called wetiko is nothing in itself – not being able to exist from its own side – yet it has a 'virtual reality' such that it can potentially destroy our species ...

...The fact that a vampire is not reflected by a mirror can also mean that what we need to see is that there's nothing, no-thing to see, other than ourselves. The fact that wetiko is the expression of something inside of us means that the cure for wetiko is with us as well. The critical issue is finding this cure within us and then putting it into effect.

Evil begets evil because if evil does not constantly expand and find new sources of energetic sustenance its evil, its *distortion*, dies with the assimilation into balance and harmony. Love is the garlic to Wetiko's vampire. Evil, the absence of love, cannot exist in the presence of love. I think I see a way out of here. I have emphasised so many times over the decades that the Archons/Wetiko and their Cult are not all powerful. *They are not*. I don't care how it looks even now *they are not*. I have not called them little boys in short trousers for effect. I have said it because it is true. Wetiko's insatiable desire for power over others is not a sign of its omnipotence, but its insecurity. Paul Levy writes: 'Due to the primal fear which ultimately drives it and which it is driven to cultivate, wetiko's body politic has an intrinsic and insistent need for centralising power and control so as to create imagined safety for itself.' *Yeeeeeees!* Exactly! Why does Wetiko want humans in an ongoing state of fear? Wetiko itself *is* fear and it is petrified of love. As evil is an absence of love, so love is an absence of fear. Love conquers all and *especially* Wetiko which *is* fear. Wetiko brought fear into the world when it wasn't here before. *Fear* was the 'fall', the fall into low-frequency ignorance and illusion – fear is **False Emotion Appearing Real**. The simulation is driven and energised by fear because Wetiko/Yaldabaoth (fear) *are* the simulation. Fear is the absence of love and Wetiko is the absence of love.

Wetiko today

We can now view current events from this level of perspective. The 'Covid' hoax has generated momentous amounts of ongoing fear, anxiety, depression and despair which have empowered Wetiko. No wonder people like Gates have been the instigators when they are Wetiko incarnate and exhibit every trait of Wetiko in the extreme. See how cold and unemotional these people are like Gates and his cronies, how dead of eye they are. That's Wetiko. Sabbatians are Wetiko and everything they control including the World Health Organization, Big Pharma and the 'vaccine' makers, national 'health'

hierarchies, corporate media, Silicon Valley, the banking system, and the United Nations with its planned transformation into world government. All are controlled and possessed by the Wetiko distortion into distorting human society in its image. We are with this knowledge at the gateway to understanding the world. Divisions of race, culture, creed and sexuality are diversions to hide the real division between those possessed and influenced by Wetiko and those that are not. The 'Covid' hoax has brought both clearly into view. Human behaviour is not about race. Tyrants and dictatorships come in all colours and creeds. What unites the US president bombing the innocent and an African tribe committing genocide against another as in Rwanda? What unites them? *Wetiko*. All wars are Wetiko, all genocide is Wetiko, all hunger over centuries in a world of plenty is Wetiko. Children going to bed hungry, including in the West, is Wetiko. Cult-generated Woke racial divisions that focus on the body are designed to obscure the reality that divisions in behaviour are manifestations of mind, not body. Obsession with body identity and group judgement is a means to divert attention from the real source of behaviour – mind and perception. Conflict sown by the Woke both within themselves and with their target groups are Wetiko providing lunch for itself through still more agents of the division, chaos, and fear on which it feeds. The Cult is seeking to assimilate the entirety of humanity and all children and young people into the Wetiko frequency by manipulating them into states of fear and despair. Witness all the suicide and psychological unravelling since the spring of 2020. Wetiko psychopaths want to impose a state of unquestioning obedience to authority which is no more than a conduit for Wetiko to enforce its will and assimilate humanity into itself. It needs us to believe that resistance is futile when it fears resistance and even more so the game-changing non-cooperation with its impositions. It can use violent resistance for its benefit. Violent impositions and violent resistance are *both* Wetiko. The Power of Love with its Power of No will sweep Wetiko from our world. Wetiko and its Cult know that. They just don't want us to know.

AI Wetiko

This brings me to AI or artificial intelligence and something else Wetikos don't want us to know. What is AI *really*? I know about computer code algorithms and AI that learns from data input. These, however, are more diversions, the expeditionary force, for the real AI that they want to connect to the human brain as promoted by Silicon Valley Wetikos like Kurzweil. What is this AI? It is the frequency of *Wetiko*, the frequency of the Archons. The connection of AI to the human brain is the connection of the Wetiko frequency to create a Wetiko hive mind and complete the job of assimilation. The hive mind is planned to be controlled from Israel and China which are both 100 percent owned by Wetiko Sabbatians. The assimilation process has been going on minute by minute in the 'smart' era which fused with the 'Covid' era. We are told that social media is scrambling the minds of the young and changing their personality. This is true, but what is social media? Look more deeply at how it works, how it creates divisions and conflict, the hostility and cruelty, the targeting of people until they are destroyed. That's Wetiko. Social media is manipulated to tune people to the Wetiko frequency with all the emotional exploitation tricks employed by platforms like Facebook and its Wetiko front man, Zuckerberg. Facebook's Instagram announced a new platform for children to overcome a legal bar on them using the main site. This is more Wetiko exploitation and manipulation of kids. Amnesty International likened the plan to foxes offering to guard the henhouse and said it was incompatible with human rights. Since when did Wetiko or Zuckerberg (I repeat myself) care about that? Would Brin and Page at Google, Wojcicki at YouTube, Bezos at Amazon and whoever the hell runs Twitter act as they do if they were not channelling Wetiko? Would those who are developing technologies for no other reason than human control? How about those designing and selling technologies to kill people and Big Pharma drug and 'vaccine' producers who know they will end or devastate lives? Quite a thought for these people to consider is that if you are Wetiko in a human life you are Wetiko on the 'other side' unless your frequency

changes and that can only change by a change of perception which becomes a change of behaviour. Where Gates is going does not bear thinking about although perhaps that's exactly where he wants to go. Either way, that's where he's going. His frequency will make it so.

The frequency lair

I have been saying for a long time that a big part of the addiction to smartphones and devices is that a frequency is coming off them that entraps the mind. People spend ages on their phones and sometimes even a minute or so after they put them down they pick them up again and it all repeats. 'Covid' lockdowns will have increased this addiction a million times for obvious reasons. Addictions to alcohol overindulgence and drugs are another way that Wetiko entraps consciousness to attach to its own. Both are symptoms of low-vibrational psychological distress which alcoholism and drug addiction further compound. Do we think it's really a coincidence that access to them is made so easy while potions that can take people into realms beyond the simulation are banned and illegal? I have explored smartphone addiction in other books, the scale is mind-blowing, and that level of addiction does not come without help. Tech companies that make these phones are Wetiko and they will have no qualms about destroying the minds of children. We are seeing again with these companies the Wetiko perceptual combination of psychopathic enforcers and weak and meek unquestioning compliance by the rank and file.

The global Smart Grid is the Wetiko Grid and it is crucial to complete the Cult endgame. The simulation is radiation and we are being deluged with technological radiation on a devastating scale. Wetiko frauds like Elon Musk serve Cult interests while occasionally criticising them to maintain his street-cred. 5G and other forms of Wi-Fi are being directed at the earth from space on a volume and scale that goes on increasing by the day. Elon Musk's (officially) SpaceX Starlink project is in the process of putting tens of thousands of satellites in low orbit to cover every inch of the planet with 5G and other Wi-Fi to create Kurzweil's global 'cloud' to which the

human mind is planned to be attached very soon. SpaceX has approval to operate 12,000 satellites with more than 1,300 launched at the time of writing and applications filed for 30,000 more. Other operators in the Wi-Fi, 5G, low-orbit satellite market include OneWeb (UK), Telesat (Canada), and AST & Science (US). Musk tells us that AI could be the end of humanity and then launches a company called Neuralink to connect the human brain to computers. Musk's (in theory) Tesla company is building electric cars and the driverless vehicles of the smart control grid. As frauds and bullshitters go Elon Musk in my opinion is Major League.

5G and technological radiation in general are destructive to human health, genetics and psychology and increasing the strength of artificial radiation underpins the five-sense perceptual bubbles which are themselves expressions of radiation or electromagnetism. Freedom activist John Whitehead was so right with his 'databit by databit, we are building our own electronic concentration camps'. The Smart Grid and 5G is a means to control the human mind and infuse perceptual information into The Field to influence anyone in sync with its frequency. You can change perception and behaviour en masse if you can manipulate the population into those levels of frequency and this is happening all around us today. The arrogance of Musk and his fellow Cult operatives knows no bounds in the way that we see with Gates. Musk's satellites are so many in number already they are changing the night sky when viewed from Earth. The astronomy community has complained about this and they have seen nothing yet. Some consequences of Musk's Wetiko hubris include: Radiation; visible pollution of the night sky; interference with astronomy and meteorology; ground and water pollution from intensive use of increasingly many spaceports; accumulating space debris; continual deorbiting and burning up of aging satellites, polluting the atmosphere with toxic dust and smoke; and ever-increasing likelihood of collisions. A collective public open letter of complaint to Musk said:

We are writing to you ... because SpaceX is in process of surrounding the Earth with a network of thousands of satellites whose very purpose is to irradiate every square inch of the

Earth. SpaceX, like everyone else, is treating the radiation as if it were not there. As if the mitochondria in our cells do not depend on electrons moving undisturbed from the food we digest to the oxygen we breathe.

As if our nervous systems and our hearts are not subject to radio frequency interference like any piece of electronic equipment. As if the cancer, diabetes, and heart disease that now afflict a majority of the Earth's population are not metabolic diseases that result from interference with our cellular machinery. As if insects everywhere, and the birds and animals that eat them, are not starving to death as a result.

People like Musk and Gates believe in their limitless Wetiko arrogance that they can do whatever they like to the world because they own it. Consequences for humanity are irrelevant. It's absolutely time that we stopped taking this shit from these self-styled masters of the Earth when you consider where this is going.

Why is the Cult so anti-human?

I hear this question often: Why would they do this when it will affect them, too? Ah, but will it? Who is this *them*? Forget their bodies. They are just vehicles for Wetiko consciousness. When you break it all down to the foundations we are looking at a state of severely distorted consciousness targeting another state of consciousness for assimilation. The rest is detail. The simulation is the fly-trap in which unique sensations of the five senses create a cycle of addiction called reincarnation. Renegade Minds see that everything which happens in our reality is a smaller version of the whole picture in line with the holographic principle. Addiction to the radiation of smart technology is a smaller version of addiction to the whole simulation. Connecting the body/brain to AI is taking that addiction on a giant step further to total ongoing control by assimilating human incarnate consciousness into Wetiko. I have watched during the 'Covid' hoax how many are becoming ever more profoundly attached to Wetiko's perceptual calling cards of aggressive response to any other point of view ('There is no other god but me'), psychopathic lack of compassion and empathy, and servile submission to the narrative and will of authority. Wetiko is the psychopaths *and* subservience to psychopaths. The Cult of Wetiko is

so anti-human because it is *not* human. It embarked on a mission to destroy human by targeting everything that it means to be human and to survive as human. 'Covid' is not the end, just a means to an end. The Cult with its Wetiko consciousness is seeking to change Earth systems, including the atmosphere, to suit them, not humans. The gathering bombardment of 5G alone from ground and space is dramatically changing The Field with which the five senses interact. There is so much more to come if we sit on our hands and hope it will all go away. It is not meant to go away. It is meant to get ever more extreme and we need to face that while we still can – just.

Carbon dioxide is the gas of life. Without that human is over. Kaput, gone, history. No natural world, no human. The Cult has created a cock and bull story about carbon dioxide and climate change to justify its reduction to the point where Gates and the ignoramus Biden 'climate chief' John Kerry want to suck it out of the atmosphere. Kerry wants to do this because his master Gates does. Wetikos have made the gas of life a demon with the usual support from the Wokers of Extinction Rebellion and similar organisations and the bewildered puppet-child that is Greta Thunberg who was put on the world stage by Klaus Schwab and the World Economic Forum. The name Extinction Rebellion is both ironic and as always Wetiko inversion. The gas that we need to survive must be reduced to save us from extinction. The most basic need of human is oxygen and we now have billions walking around in face nappies depriving body and brain of this essential requirement of human existence. More than that 5G at 60 gigahertz interacts with the oxygen molecule to reduce the amount of oxygen the body can absorb into the bloodstream. The obvious knock-on consequences of that for respiratory and cognitive problems and life itself need no further explanation. Psychopaths like Musk are assembling a global system of satellites to deluge the human atmosphere with this insanity. The man should be in jail. Here we have two most basic of human needs, oxygen and carbon dioxide, being dismantled.

Two others, water and food, are getting similar treatment with the United Nations Agendas 21 and 2030 – the Great Reset – planning to

centrally control all water and food supplies. People will not even own rain water that falls on their land. Food is affected at the most basic level by reducing carbon dioxide. We have genetic modification or GMO infiltrating the food chain on a mass scale, pesticides and herbicides polluting the air and destroying the soil. Freshwater fish that provide livelihoods for 60 million people and feed hundreds of millions worldwide are being 'pushed to the brink' according the conservationists while climate change is the only focus. Now we have Gates and Schwab wanting to dispense with current food sources all together and replace them with a synthetic version which the Wetiko Cult would control in terms of production and who eats and who doesn't. We have been on the Totalitarian Tiptoe to this for more than 60 years as food has become ever more processed and full of chemical shite to the point today when it's not natural food at all. As Dr Tom Cowan says: 'If it has a label don't eat it.' Bill Gates is now the biggest owner of farmland in the United States and he does nothing without an ulterior motive involving the Cult. Klaus Schwab wrote: 'To feed the world in the next 50 years we will need to produce as much food as was produced in the last 10,000 years ... food security will only be achieved, however, if regulations on genetically modified foods are adapted to reflect the reality that gene editing offers a precise, efficient and safe method of improving crops.' Liar. People and the world are being targeted with aluminium through vaccines, chemtrails, food, drink cans, and endless other sources when aluminium has been linked to many health issues including dementia which is increasing year after year. Insects, bees and wildlife essential to the food chain are being deleted by pesticides, herbicides and radiation which 5G is dramatically increasing with 6G and 7G to come. The pollinating bee population is being devastated while wildlife including birds, dolphins and whales are having their natural radar blocked by the effects of ever-increasing radiation. In the summer windscreens used to be splattered with insects so numerous were they. It doesn't happen now. Where have they gone?

Synthetic everything

The Cult is introducing genetically-modified versions of trees, plants and insects including a Gates-funded project to unleash hundreds of millions of genetically-modified, lab-altered and patented male mosquitoes to mate with wild mosquitoes and induce genetic flaws that cause them to die out. Clinically-insane Gates-funded Japanese researchers have developed mosquitos that spread vaccine and are dubbed 'flying vaccinators'. Gates is funding the modification of weather patterns in part to sell the myth that this is caused by carbon dioxide and he's funding geoengineering of the skies to change the atmosphere. Some of this came to light with the Gates-backed plan to release tonnes of chalk into the atmosphere to 'deflect the Sun and cool the planet'. Funny how they do this while the heating effect of the Sun is not factored into climate projections focussed on carbon dioxide. The reason is that they want to reduce carbon dioxide (so don't mention the Sun), but at the same time they do want to reduce the impact of the Sun which is so essential to human life and health. I have mentioned the sun-cholesterol-vitamin D connection as they demonise the Sun with warnings about skin cancer (caused by the chemicals in sun cream they tell you to splash on). They come from the other end of the process with statin drugs to reduce cholesterol that turns sunlight into vitamin D. A lack of vitamin D leads to a long list of health effects and how vitamin D levels must have fallen with people confined to their homes over 'Covid'. Gates is funding other forms of geoengineering and most importantly chemtrails which are dropping heavy metals, aluminium and self-replicating nanotechnology onto the Earth which is killing the natural world. See *Everything You Need To Know, But Have Never Been Told* for the detailed background to this.

Every human system is being targeted for deletion by a force that's not human. The Wetiko Cult has embarked on the process of transforming the human body from biological to synthetic biological as I have explained. Biological is being replaced by the artificial and synthetic – Archontic 'countermimicry' – right across human society. The plan eventually is to dispense with the human body altogether

and absorb human consciousness – which it wouldn't really be by then – into cyberspace (the simulation which is Wetiko/Yaldabaoth). Preparations for that are already happening if people would care to look. The alternative media rightly warns about globalism and 'the globalists', but this is far bigger than that and represents the end of the human race as we know it. The 'bad copy' of prime reality that Gnostics describe was a bad copy of harmony, wonder and beauty to start with before Wetiko/Yaldabaoth set out to change the simulated 'copy' into something very different. The process was slow to start with. Entrapped humans in the simulation timeline were not technologically aware and they had to be brought up to intellectual speed while being suppressed spiritually to the point where they could build their own prison while having no idea they were doing so. We have now reached that stage where technological intellect has the potential to destroy us and that's why events are moving so fast. Central American shaman Don Juan Matus said:

Think for a moment, and tell me how you would explain the contradictions between the intelligence of man the engineer and the stupidity of his systems of belief, or the stupidity of his contradictory behaviour. Sorcerers believe that the predators have given us our systems of beliefs, our ideas of good and evil; our social mores. They are the ones who set up our dreams of success or failure. They have given us covetousness, greed, and cowardice. It is the predator who makes us complacent, routinary, and egomaniacal.

In order to keep us obedient and meek and weak, the predators engaged themselves in a stupendous manoeuvre – stupendous, of course, from the point of view of a fighting strategist; a horrendous manoeuvre from the point of those who suffer it. They gave us their mind. The predators' mind is baroque, contradictory, morose, filled with the fear of being discovered any minute now.

For 'predators' see Wetiko, Archons, Yaldabaoth, Jinn, and all the other versions of the same phenomenon in cultures and religions all over the world. The theme is always the same because it's true and it's real. We have reached the point where we have to deal with it. The question is – how?

Don't fight – walk away

I thought I'd use a controversial subheading to get things moving in terms of our response to global fascism. What do you mean 'don't fight'? What do you mean 'walk away'? We've got to fight. We can't walk away. Well, it depends what we mean by fight and walk away. If fighting means physical combat we are playing Wetiko's game and falling for its trap. It wants us to get angry, aggressive, and direct hate and hostility at the enemy we think we must fight. Every war, every battle, every conflict, has been fought with Wetiko leading both sides. It's what it does. Wetiko wants a fight, anywhere, any place. Just hit me, son, so I can hit you back. Wetiko hits Wetiko and Wetiko hits Wetiko in return. I am very forthright as you can see in exposing Wetikos of the Cult, but I don't hate them. I refuse to hate them. It's what they want. What you hate you become. What you *fight* you become. Wokers, 'anti-haters' and 'anti-fascists' prove this every time they reach for their keyboards or don their balaclavas. By walk away I mean to disengage from Wetiko which includes ceasing to cooperate with its tyranny. Paul Levy says of Wetiko:

The way to 'defeat' evil is not to try to destroy it (for then, in playing evil's game, we have already lost), but rather, to find the invulnerable place within ourselves where evil is unable to vanquish us – this is to truly 'win' our battle with evil.

Wetiko is everywhere in human society and it's been on steroids since the 'Covid' hoax. Every shouting match over wearing masks has Wetiko wearing a mask and Wetiko not wearing one. It's an electrical circuit of push and resist, push and resist, with Wetiko pushing *and* resisting. Each polarity is Wetiko empowering itself. Dictionary definitions of 'resist' include 'opposing, refusing to accept or comply with' and the word to focus on is 'opposing'. What form does this take – setting police cars alight or 'refusing to accept or comply with'? The former is Wetiko opposing Wetiko while the other points the way forward. This is the difference between those aggressively demanding that government fascism must be obeyed who stand in stark contrast to the great majority of Pushbackers. We saw this clearly with a march by thousands of Pushbackers against lockdown in London followed days later by a Woker-hijacked

protest in Bristol in which police cars were set on fire. Masks were virtually absent in London and widespread in Bristol. Wetiko wants lockdown on every level of society and infuses its aggression to police it through its unknowing stooges. Lockdown protesters are the ones with the smiling faces and the hugs, The two blatantly obvious states of being – getting more obvious by the day – are the result of Wokers and their like becoming ever more influenced by the simulation Field of Wetiko and Pushbackers ever more influenced by The Field of a far higher vibration beyond the simulation. Wetiko can't invade the heart which is where most lockdown opponents are coming from. It's the heart that allows them to see through the lies to the truth in ways I will be highlighting.

Renegade Minds know that calmness is the place from which wisdom comes. You won't find wisdom in a hissing fit and wisdom is what we need in abundance right now. Calmness is not weakness – you don't have to scream at the top of your voice to be strong. Calmness is indeed a sign of strength. 'No' means I'm not doing it. NOOOO!!! doesn't mean you're not doing it even more. Volume does not advance 'No – I'm not doing it'. You are just not doing it. Wetiko possessed and influenced don't know how to deal with that. Wetiko wants a fight and we should not give it one. What it needs more than anything is our *cooperation* and we should not give that either. Mass rallies and marches are great in that they are a visual representation of feeling, but if it ends there they are irrelevant. You demand that Wetikos act differently? Well, they're not going to are they? They are Wetikos. We don't need to waste our time demanding that something doesn't happen when that will make no difference. We need to delete the means that *allows* it to happen. This, invariably, is our cooperation. You can demand a child stop firing a peashooter at the dog or you can refuse to buy the peashooter. If you provide the means you are cooperating with the dog being smacked on the nose with a pea. How can the authorities enforce mask-wearing if millions in a country refuse? What if the 74 million Pushbackers that voted for Trump in 2020 refused to wear masks, close their businesses or stay in their homes. It would be unenforceable. The

few control the many through the compliance of the many and that's always been the dynamic be it 'Covid' regulations or the Roman Empire. I know people can find it intimidating to say no to authority or stand out in a crowd for being the only one with a face on display; but it has to be done or it's over. I hope I've made clear in this book that where this is going will be far more intimidating than standing up now and saying 'No' – I will not cooperate with my own enslavement and that of my children. There might be consequences for some initially, although not so if enough do the same. The question that must be addressed is what is going to happen if we don't? It is time to be strong and unyieldingly so. No means no. Not here and there, but *everywhere* and *always*. I have refused to wear a mask and obey all the other nonsense. I will not comply with tyranny. I repeat: Fascism is not imposed by fascists – there are never enough of them. Fascism is imposed by the population acquiescing to fascism. *I will not do it*. I will die first, or my body will. Living meekly under fascism is a form of death anyway, the death of the spirit that Martin Luther King described.

Making things happen

We must not despair. This is not over till it's over and it's far from that. The 'fat lady' must refuse to sing. The longer the 'Covid' hoax has dragged on and impacted on more lives we have seen an awakening of phenomenal numbers of people worldwide to the realisation that what they have believed all their lives is not how the world really is. Research published by the system-serving University of Bristol and King's College London in February, 2021, concluded: 'One in every 11 people in Britain say they trust David Icke's take on the coronavirus pandemic.' It will be more by now and we have gathering numbers to build on. We must urgently progress from seeing the scam to ceasing to cooperate with it. Prominent German lawyer Reiner Fuellmich, also licenced to practice law in America, is doing a magnificent job taking the legal route to bring the psychopaths to justice through a second Nuremberg tribunal for crimes against humanity. Fuellmich has an impressive record of

beating the elite in court and he formed the German Corona Investigative Committee to pursue civil charges against the main perpetrators with a view to triggering criminal charges. Most importantly he has grasped the foundation of the hoax – the PCR test not testing for the ‘virus’ – and Christian Drosten is therefore on his charge sheet along with Gates frontman Tedros at the World Health Organization. Major players must not be allowed to inflict their horrors on the human race without being brought to book. A life sentence must follow for Bill Gates and the rest of them. A group of researchers has also indicted the government of Norway for crimes against humanity with copies sent to the police and the International Criminal Court. The lawsuit cites participation in an internationally-planned false pandemic and violation of international law and human rights, the European Commission’s definition of human rights by coercive rules, Nuremberg and Hague rules on fundamental human rights, and the Norwegian constitution. We must take the initiative from hereon and not just complain, protest and react.

There are practical ways to support vital mass non-cooperation. Organising in numbers is one. Lockdown marches in London in the spring in 2021 were mass non-cooperation that the authorities could not stop. There were too many people. Hundreds of thousands walked the London streets in the centre of the road for mile after mile while the Face-Nappies could only look on. They were determined, but calm, and just *did it* with no histrionics and lots of smiles. The police were impotent. Others are organising group shopping without masks for mutual support and imagine if that was happening all over. Policing it would be impossible. If the store refuses to serve people in these circumstances they would be faced with a long line of trolleys full of goods standing on their own and everything would have to be returned to the shelves. How would they cope with that if it kept happening? I am talking here about moving on from complaining to being pro-active; from watching things happen to making things happen. I include in this our relationship with the police. The behaviour of many Face-Nappies

has been disgraceful and anyone who thinks they would never find concentration camp guards in the 'enlightened' modern era have had that myth busted big-time. The period and setting may change – Wetikos never do. I watched film footage from a London march in which a police thug viciously kicked a protestor on the floor who had done nothing. His fellow Face-Nappies stood in a ring protecting him. What he did was a criminal assault and with a crowd far outnumbering the police this can no longer be allowed to happen unchallenged. I get it when people chant 'shame on you' in these circumstances, but that is no longer enough. They *have* no shame those who do this. Crowds needs to start making a citizen's arrest of the police who commit criminal offences and brutally attack innocent people and defenceless women. A citizen's arrest can be made under section 24A of the UK Police and Criminal Evidence (PACE) Act of 1984 and you will find something similar in other countries. I prefer to call it a Common Law arrest rather than citizen's for reasons I will come to shortly. Anyone can arrest a person committing an indictable offence or if they have reasonable grounds to suspect they are committing an indictable offence. On both counts the attack by the police thug would have fallen into this category. A citizen's arrest can be made to stop someone:

- Causing physical injury to himself or any other person
- Suffering physical injury
- Causing loss of or damage to property
- Making off before a constable can assume responsibility for him

A citizen's arrest may also be made to prevent a breach of the peace under Common Law and if they believe a breach of the peace will happen or anything related to harm likely to be done or already done in their presence. This is the way to go I think – the Common Law version. If police know that the crowd and members of the public will no longer be standing and watching while they commit

their thuggery and crimes they will think twice about acting like Brownshirts and Blackshirts.

Common Law – common sense

Mention of Common Law is very important. Most people think the law is the law as in one law. This is not the case. There are two bodies of law, Common Law and Statute Law, and they are not the same. Common Law is founded on the simple premise of do no harm. It does not recognise victimless crimes in which no harm is done while Statute Law does. There is a Statute Law against almost everything. So what is Statute Law? Amazingly it's the law of the *sea* that was brought ashore by the Cult to override the law of the land which is Common Law. They had no right to do this and as always they did it anyway. They had to. They could not impose their will on the people through Common Law which only applies to do no harm. How could you stitch up the fine detail of people's lives with that? Instead they took the law of the sea, or Admiralty Law, and applied it to the population. Statute Law refers to all the laws spewing out of governments and their agencies including all the fascist laws and regulations relating to 'Covid'. The key point to make is that Statute Law is *contract law*. It only applies between *contracting* corporations. Most police officers don't even know this. They have to be kept in the dark, too. Long ago when merchants and their sailing ships began to trade with different countries a contractual law was developed called Admiralty Law and other names. Again it only applied to *contracts* agreed between *corporate* entities. If there is no agreed contract the law of the sea had no jurisdiction *and that still applies to its new alias of Statute Law*. The problem for the Cult when the law of the sea was brought ashore was an obvious one. People were not corporations and neither were government entities. To overcome the latter they made governments and all associated organisations corporations. All the institutions are *private corporations* and I mean governments and their agencies, local councils, police, courts, military, US states, the whole lot. Go to the

Dun and Bradstreet corporate listings website for confirmation that they are all corporations. You are arrested by a private corporation called the police by someone who is really a private security guard and they take you to court which is another private corporation. Neither have jurisdiction over you unless you consent and *contract* with them. This is why you hear the mantra about law enforcement policing by *consent* of the people. In truth the people 'consent' only in theory through monumental trickery.

Okay, the Cult overcame the corporate law problem by making governments and institutions corporate entities; but what about people? They are not corporations are they? Ah ... well in a sense, and *only* a sense, they are. Not people exactly – the illusion of people. The Cult creates a corporation in the name of everyone at the time that their birth certificate is issued. Note birth/ *berth* certificate and when you go to court under the law of the sea on land you stand in a *dock*. These are throwbacks to the origin. My Common Law name is David Vaughan Icke. The name of the corporation created by the government when I was born is called Mr David Vaughan Icke usually written in capitals as MR DAVID VAUGHAN ICKE. That is not me, the living, breathing man. It is a fictitious corporate entity. The trick is to make you think that David Vaughan Icke and MR DAVID VAUGHAN ICKE are the same thing. *They are not*. When police charge you and take you to court they are prosecuting the corporate entity and not the living, breathing, man or woman. They have to trick you into identifying as the corporate entity and contracting with them. Otherwise they have no jurisdiction. They do this through a language known as legalese. Lawful and legal are not the same either. Lawful relates to Common Law and legal relates to Statute Law. Legalese is the language of Statue Law which uses terms that mean one thing to the public and another in legalese. Notice that when a police officer tells someone why they are being charged he or she will say at the end: 'Do you understand?' To the public that means 'Do you comprehend?' In legalese it means 'Do you stand under me?' Do you stand under my authority? If you say

yes to the question you are unknowingly agreeing to give them jurisdiction over you in a contract between two corporate entities.

This is a confidence trick in every way. Contracts have to be agreed between informed parties and if you don't know that David Vaughan Icke is agreeing to be the corporation MR DAVID VAUGHAN ICKE you cannot knowingly agree to contract. They are deceiving you and another way they do this is to ask for proof of identity. You usually show them a driving licence or other document on which your corporate name is written. In doing so you are accepting that you are that corporate entity when you are not. Referring to yourself as a 'person' or 'citizen' is also identifying with your corporate fiction which is why I made the Common Law point about the citizen's arrest. If you are approached by a police officer you identify yourself immediately as a living, breathing, man or woman and say 'I do not consent, I do not contract with you and I do not understand' or stand under their authority. I have a Common Law birth certificate as a living man and these are available at no charge from commonlawcourt.com. Businesses registered under the Statute Law system means that its laws apply. There are, however, ways to run a business under Common Law. Remember all 'Covid' laws and regulations are Statute Law – the law of *contracts* and you do not have to contract. This doesn't mean that you can kill someone and get away with it. Common Law says do no harm and that applies to physical harm, financial harm etc. Police are employees of private corporations and there needs to be a new system of non-corporate Common Law constables operating outside the Statute Law system. If you go to davidicke.com and put Common Law into the search engine you will find videos that explain Common Law in much greater detail. It is definitely a road we should walk.

With all my heart

I have heard people say that we are in a spiritual war. I don't like the term 'war' with its Wetiko dynamic, but I know what they mean. Sweep aside all the bodily forms and we are in a situation in which two states of consciousness are seeking very different realities.

Wetiko wants upheaval, chaos, fear, suffering, conflict and control. The other wants love, peace, harmony, fairness and freedom. That's where we are. We should not fall for the idea that Wetiko is all-powerful and there's nothing we can do. Wetiko is not all-powerful. It's a joke, pathetic. It doesn't have to be, but it has made that choice for now. A handful of times over the years when I have felt the presence of its frequency I have allowed it to attach briefly so I could consciously observe its nature. The experience is not pleasant, the energy is heavy and dark, but the ease with which you can kick it back out the door shows that its real power is in persuading us that it has power. It's all a con. Wetiko is a con. It's a trickster and not a power that can control us if we unleash our own. The con is founded on manipulating humanity to give its power to Wetiko which recycles it back to present the illusion that it has power when its power is *ours* that we gave away. This happens on an energetic level and plays out in the world of the seen as humanity giving its power to Wetiko authority which uses that power to control the population when the power is only the power the population has handed over. How could it be any other way for billions to be controlled by a relative few? I have had experiences with people possessed by Wetiko and again you can kick its arse if you do it with an open heart. Oh yes – the *heart* which can transform the world of perceived 'matter'.

We are receiver-transmitters and processors of information, but what information and where from? Information is processed into perception in three main areas – the brain, the heart and the belly. These relate to thinking, knowing, and emotion. Wetiko wants us to be head and belly people which means we think within the confines of the Matrix simulation and low-vibrational emotional reaction scrambles balance and perception. A few minutes on social media and you see how emotion is the dominant force. Woke is all emotion and is therefore thought-free and fact-free. Our heart is something different. It *knows* while the head *thinks* and has to try to work it out because it doesn't know. The human energy field has seven prime vortexes which connect us with wider reality ([Fig 23](#)). Chakra means

‘wheels of light’ in the Sanskrit language of ancient India. The main ones are: The crown chakra on top of the head; brow (or ‘third eye’) chakra in the centre of the forehead; throat chakra; heart chakra in the centre of the chest; solar plexus chakra below the sternum; sacral chakra beneath the navel; and base chakra at the bottom of the spine. Each one has a particular function or functions. We feel anxiety and nervousness in the belly where the sacral chakra is located and this processes emotion that can affect the colon to give people ‘the shits’ or make them ‘shit scared’ when they are nervous. Chakras all play an important role, but the Mr and Mrs Big is the heart chakra which sits at the centre of the seven, above the chakras that connect us to the ‘physical’ and below those that connect with higher realms (or at least should). Here in the heart chakra we feel love, empathy and compassion – ‘My heart goes out to you’. Those with closed hearts become literally ‘heart-less’ in their attitudes and behaviour (see Bill Gates). Native Americans portrayed Wetiko with what Paul Levy calls a ‘frigid, icy heart, devoid of mercy’ (see Bill Gates).



Figure 23: The chakra system which interpenetrates the human energy field. The heart chakra is the governor – or should be.

Wetiko trembles at the thought of heart energy which it cannot infiltrate. The frequency is too high. What it seeks to do instead is close the heart chakra vortex to block its perceptual and energetic influence. Psychopaths have ‘hearts of stone’ and emotionally-damaged people have ‘heartache’ and ‘broken hearts’. The astonishing amount of heart disease is related to heart chakra

disruption with its fundamental connection to the 'physical' heart. Dr Tom Cowan has written an outstanding book challenging the belief that the heart is a pump and making the connection between the 'physical' and spiritual heart. Rudolph Steiner who was way ahead of his time said the same about the fallacy that the heart is a pump. *What?* The heart is not a pump? That's crazy, right? Everybody knows that. Read Cowan's *Human Heart, Cosmic Heart* and you will realise that the very idea of the heart as a pump is ridiculous when you see the evidence. How does blood in the feet so far from the heart get pumped horizontally up the body by the heart?? Cowan explains in the book the real reason why blood moves as it does. Our 'physical' heart is used to symbolise love when the source is really the heart vortex or spiritual heart which is our most powerful energetic connection to 'out there' expanded consciousness. That's why we feel *knowing* – intuitive knowing – in the centre of the chest. Knowing doesn't come from a process of thoughts leading to a conclusion. It is there in an instant all in one go. Our heart knows because of its connection to levels of awareness that *do* know. This is the meaning and source of intuition – intuitive *knowing*.

For the last more than 30 years of uncovering the global game and the nature of reality my heart has been my constant antenna for truth and accuracy. An American intelligence insider once said that I had quoted a disinformant in one of my books and yet I had only quoted the part that was true. He asked: 'How do you do that?' By using my heart antenna was the answer and anyone can do it. Heart-centred is how we are meant to be. With a closed heart chakra we withdraw into a closed mind and the bubble of five-sense reality. If you take a moment to focus your attention on the centre of your chest, picture a spinning wheel of light and see it opening and expanding. You will feel it happening, too, and perceptions of the heart like joy and love as the heart impacts on the mind as they interact. The more the chakra opens the more you will feel expressions of heart consciousness and as the process continues, and becomes part of you, insights and knowings will follow. An open

heart is connected to that level of awareness that knows all is *One*. You will see from its perspective that the fault-lines that divide us are only illusions to control us. An open heart does not process the illusions of race, creed and sexuality except as brief experiences for a consciousness that is all. Our heart does not see division, only unity (Figs 24 and 25). There's something else, too. Our hearts love to laugh. Mark Twain's quote that says 'The human race has one really effective weapon, and that is laughter' is really a reference to the heart which loves to laugh with the joy of knowing the true nature of infinite reality and that all the madness of human society is an illusion of the mind. Twain also said: 'Against the assault of laughter nothing can stand.' This is so true of Wetiko and the Cult. Their insecurity demands that they be taken seriously and their power and authority acknowledged and feared. We should do nothing of the sort. We should not get aggressive or fearful which their insecurity so desires. We should laugh in their face. Even in their no-face as police come over in their face-nappies and expect to be taken seriously. They don't take themselves seriously looking like that so why should we? Laugh in the face of intimidation. Laugh in the face of tyranny. You will see by its reaction that you have pressed all of its buttons. Wetiko does not know what to do in the face of laughter or when its targets refuse to concede their joy to fear. We have seen many examples during the 'Covid' hoax when people have expressed their energetic power and the string puppets of Wetiko retreat with their tail limp between their knees. Laugh – the world is bloody mad after all and if it's a choice between laughter and tears I know which way I'm going.



Figure 24: Head consciousness without the heart sees division and everything apart from everything else.



Figure 25: Heart consciousness sees everything as One.

'Vaccines' and the soul

The foundation of Wetiko/Archon control of humans is the separation of incarnate five-sense mind from the infinite 'I' and closing the heart chakra where the True 'I' lives during a human life. The goal has been to achieve complete separation in both cases. I was interested therefore to read an account by a French energetic healer of what she said she experienced with a patient who had been given the 'Covid' vaccine. Genuine energy healers can sense information and consciousness fields at different levels of being which are referred to as 'subtle bodies'. She described treating the patient who later returned after having, without the healer's knowledge, two doses of the 'Covid vaccine'. The healer said:

I noticed immediately the change, very heavy energy emanating from [the] subtle bodies. The scariest thing was when I was working on the heart chakra, I connected with her soul: it was detached from the physical body, it had no contact and it was, as if it was floating in a state of total confusion: a damage to the consciousness that loses contact with the physical body, i.e. with our biological machine, there is no longer any communication between them.

I continued the treatment by sending light to the heart chakra, the soul of the person, but it seemed that the soul could no longer receive any light, frequency or energy. It was a very powerful experience for me. Then I understood that this substance is indeed used to detach consciousness so that this consciousness can no longer interact through this body that it possesses in life, where there is no longer any contact, no frequency, no light, no more energetic balance or mind.

This would create a human that is rudderless and at the extreme almost zombie-like operating with a fractional state of consciousness at the mercy of Wetiko. I was especially intrigued by what the healer said in the light of the prediction by the highly-informed Rudolf Steiner more than a hundred years ago. He said:

In the future, we will eliminate the soul with medicine. Under the pretext of a 'healthy point of view', there will be a vaccine by which the human body will be treated as soon as possible directly at birth, so that the human being cannot develop the thought of the existence of soul and Spirit. To materialistic doctors will be entrusted the task of removing the soul of humanity.

As today, people are vaccinated against this disease or that disease, so in the future, children will be vaccinated with a substance that can be produced precisely in such a way that people, thanks to this vaccination, will be immune to being subjected to the 'madness' of spiritual life. He would be extremely smart, but he would not develop a conscience, and that is the true goal of some materialistic circles.

Steiner said the vaccine would detach the physical body from the etheric body (subtle bodies) and 'once the etheric body is detached the relationship between the universe and the etheric body would become extremely unstable, and man would become an automaton'. He said 'the physical body of man must be polished on this Earth by spiritual will – so the vaccine becomes a kind of arymanique (Wetiko) force' and 'man can no longer get rid of a given materialistic feeling'. Humans would then, he said, become 'materialistic of constitution and can no longer rise to the spiritual'. I have been writing for years about DNA being a receiver-transmitter of information that connects us to other levels of reality and these 'vaccines' changing DNA can be likened to changing an antenna and what it can transmit and receive. Such a disconnection would clearly lead to changes in personality and perception. Steiner further predicted the arrival of AI. Big Pharma 'Covid vaccine' makers, expressions of Wetiko, are testing their DNA-manipulating evil on children as I write with a view to giving the 'vaccine' to babies. If it's a soul-body disconnecter – and I say that it is or can be – every child would be disconnected from 'soul' at birth and the 'vaccine' would create a closed system in which spiritual guidance from the greater self would play no part. This has been the ambition of Wetiko all

along. A Pentagon video from 2005 was leaked of a presentation explaining the development of vaccines to change behaviour by their effect on the brain. Those that believe this is not happening with the 'Covid' genetically-modifying procedure masquerading as a 'vaccine' should make an urgent appointment with Naivety Anonymous. Klaus Schwab wrote in 2018:

Neurotechnologies enable us to better influence consciousness and thought and to understand many activities of the brain. They include decoding what we are thinking in fine levels of detail through new chemicals and interventions that can influence our brains to correct for errors or enhance functionality.

The plan is clear and only the heart can stop it. With every heart that opens, every mind that awakens, Wetiko is weakened. Heart and love are far more powerful than head and hate and so nothing like a majority is needed to turn this around.

Beyond the Phantom

Our heart is the prime target of Wetiko and so it must be the answer to Wetiko. We *are* our heart which is part of one heart, the infinite heart. Our heart is where the true self lives in a human life behind firewalls of five-sense illusion when an imposter takes its place – *Phantom Self*; but our heart waits patiently to be set free any time we choose to see beyond the Phantom, beyond Wetiko. A Wetikoed Phantom Self can wreak mass death and destruction while the love of forever is locked away in its heart. The time is here to unleash its power and let it sweep away the fear and despair that is Wetiko. Heart consciousness does not seek manipulated, censored, advantage for its belief or religion, its activism and desires. As an expression of the One it treats all as One with the same rights to freedom and opinion. Our heart demands fairness for itself no more than for others. From this unity of heart we can come together in mutual support and transform this Wetikoed world into what reality is meant to be – a place of love, joy, happiness, fairness, justice and freedom. Wetiko has another agenda and that's why the world is as

it is, but enough of this nonsense. Wetiko can't stay where hearts are open and it works so hard to keep them closed. Fear is its currency and its food source and love in its true sense has no fear. Why would love have fear when it knows it is *All That Is, Has Been, And Ever Can Be* on an eternal exploration of all possibility? Love in this true sense is not the physical attraction that passes for love. This can be an expression of it, yes, but Infinite Love, a love without condition, goes far deeper to the core of all being. It *is* the core of all being. Infinite reality was born from love beyond the illusions of the simulation. Love infinitely expressed is the knowing that all is One and the swiftly-passing experience of separation is a temporary hallucination. You cannot disconnect from Oneness; you can only *perceive* that you have and withdraw from its influence. This is the most important of all perception trickery by the mind parasite that is Wetiko and the foundation of all its potential for manipulation.

If we open our hearts, open the sluice gates of the mind, and redefine self-identity amazing things start to happen. Consciousness expands or contracts in accordance with self-identity. When true self is recognised as infinite awareness and label self – Phantom Self – is seen as only a series of brief experiences life is transformed. Consciousness expands to the extent that self-identity expands and everything changes. You see unity, not division, the picture, not the pixels. From this we can play the long game. No more is an experience something in and of itself, but a fleeting moment in the eternity of forever. Suddenly people in uniform and dark suits are no longer intimidating. Doing what your heart knows to be right is no longer intimidating and consequences for those actions take on the same nature of a brief experience that passes in the blink of an infinite eye. Intimidation is all in the mind. Beyond the mind there is no intimidation.

An open heart does not consider consequences for what it knows to be right. To do so would be to consider not doing what it knows to be right and for a heart in its power that is never an option. The Renegade Mind is really the Renegade Heart. Consideration of consequences will always provide a getaway car for the mind and

the heart doesn't want one. What is right in the light of what we face today is to stop cooperating with Wetiko in all its forms and to do it without fear or compromise. You cannot compromise with tyranny when tyranny always demands more until it has everything. Life is your perception and you are your destiny. Change your perception and you change your life. Change collective perception and we change the world.

Come on people ... One human family, One heart, One goal ...
FREEEEEEEDOM!

We must settle for nothing less.

Postscript

The big scare story as the book goes to press is the 'Indian' variant and the world is being deluged with propaganda about the 'Covid catastrophe' in India which mirrors in its lies and misrepresentations what happened in Italy before the first lockdown in 2020.

The *New York Post* published a picture of someone who had 'collapsed in the street from Covid' in India in April, 2021, which was actually taken during a gas leak in May, 2020. Same old, same old. Media articles in mid-February were asking why India had been so untouched by 'Covid' and then as their vaccine rollout gathered pace the alleged 'cases' began to rapidly increase. Indian 'Covid vaccine' maker Bharat Biotech was funded into existence by the Bill and Melinda Gates Foundation (the pair announced their divorce in May, 2021, which is a pity because they so deserve each other). The Indian 'Covid crisis' was ramped up by the media to terrify the world and prepare people for submission to still more restrictions. The scam that worked the first time was being repeated only with far more people seeing through the deceit. Davidicke.com and Ickonic.com have sought to tell the true story of what is happening by talking to people living through the Indian nightmare which has nothing to do with 'Covid'. We posted a letter from 'Alisha' in Pune who told a very different story to government and media mendacity. She said scenes of dying people and overwhelmed hospitals were designed to hide what was really happening – genocide and starvation. Alisha said that millions had already died of starvation during the ongoing lockdowns while government and media were lying and making it look like the 'virus':

Restaurants, shops, gyms, theatres, basically everything is shut. The cities are ghost towns. Even so-called 'essential' businesses are only open till 11am in the morning. You basically have just an hour to buy food and then your time is up.

Inter-state travel and even inter-district travel is banned. The cops wait at all major crossroads to question why you are traveling outdoors or to fine you if you are not wearing a mask.

The medical community here is also complicit in genocide, lying about hospitals being full and turning away people with genuine illnesses, who need immediate care. They have even created a shortage of oxygen cylinders.

This is the classic Cult modus operandi played out in every country. Alisha said that people who would not have a PCR test not testing for the 'virus' were being denied hospital treatment. She said the people hit hardest were migrant workers and those in rural areas. Most businesses employed migrant workers and with everything closed there were no jobs, no income and no food. As a result millions were dying of starvation or malnutrition. All this was happening under Prime Minister Narendra Modi, a 100-percent asset of the Cult, and it emphasises yet again the scale of pure anti-human evil we are dealing with. Australia banned its people from returning home from India with penalties for trying to do so of up to five years in jail and a fine of £37,000. The manufactured 'Covid' crisis in India was being prepared to justify further fascism in the West. Obvious connections could be seen between the Indian 'vaccine' programme and increased 'cases' and this became a common theme. The Seychelles, the most per capita 'Covid vaccinated' population in the world, went back into lockdown after a 'surge of cases'.

Long ago the truly evil Monsanto agricultural biotechnology corporation with its big connections to Bill Gates devastated Indian farming with genetically-modified crops. Human rights activist Gurcharan Singh highlighted the efforts by the Indian government to complete the job by destroying the food supply to hundreds of millions with 'Covid' lockdowns. He said that 415 million people at the bottom of the disgusting caste system (still going whatever they say) were below the poverty line and struggled to feed themselves every year. Now the government was imposing lockdown at just the

time to destroy the harvest. This deliberate policy was leading to mass starvation. People may reel back at the suggestion that a government would do that, but Wetiko-controlled 'leaders' are capable of any level of evil. In fact what is described in India is in the process of being instigated worldwide. The food chain and food supply are being targeted at every level to cause world hunger and thus control. Bill Gates is not the biggest owner of farmland in America for no reason and destroying access to food aids both the depopulation agenda and the plan for synthetic 'food' already being funded into existence by Gates. Add to this the coming hyper-inflation from the suicidal creation of fake 'money' in response to 'Covid' and the breakdown of container shipping systems and you have a cocktail that can only lead one way and is meant to. The Cult plan is to crash the entire system to 'build back better' with the Great Reset.

'Vaccine' transmission

Reports from all over the world continue to emerge of women suffering menstrual and fertility problems after having the fake 'vaccine' and of the non-'vaccinated' having similar problems when interacting with the 'vaccinated'. There are far too many for 'coincidence' to be credible. We've had menopausal women getting periods, others having periods stop or not stopping for weeks, passing clots, sometimes the lining of the uterus, breast irregularities, and miscarriages (which increased by 400 percent in parts of the United States). Non-'vaccinated' men and children have suffered blood clots and nose bleeding after interaction with the 'vaccinated'. Babies have died from the effects of breast milk from a 'vaccinated' mother. Awake doctors – the small minority – speculated on the cause of non-'vaccinated' suffering the same effects as the 'vaccinated'. Was it nanotechnology in the synthetic substance transmitting frequencies or was it a straight chemical bioweapon that was being transmitted between people? I am not saying that some kind of chemical transmission is not one possible answer, but the foundation of all that the Cult does is frequency and

this is fertile ground for understanding how transmission can happen. American doctor Carrie Madej, an internal medicine physician and osteopath, has been practicing for the last 20 years, teaching medical students, and she says attending different meetings where the agenda for humanity was discussed. Madej, who operates out of Georgia, did not dismiss other possible forms of transmission, but she focused on frequency in search of an explanation for transmission. She said the Moderna and Pfizer 'vaccines' contained nano-lipid particles as a key component. This was a brand new technology never before used on humanity. 'They're using a nanotechnology which is pretty much little tiny computer bits ... nanobots or hydrogel.' Inside the 'vaccines' was 'this sci-fi kind of substance' which suppressed immune checkpoints to get into the cell. I referred to this earlier as the 'Trojan horse' technique that tricks the cell into opening a gateway for the self-replicating synthetic material and while the immune system is artificially suppressed the body has no defences. Madej said the substance served many purposes including an on-demand ability to 'deliver the payload' and using the nano 'computer bits' as biosensors in the body. 'It actually has the ability to accumulate data from your body, like your breathing, your respiration, thoughts, emotions, all kinds of things.'

She said the technology obviously has the ability to operate through Wi-Fi and transmit and receive energy, messages, frequencies or impulses. 'Just imagine you're getting this new substance in you and it can react to things all around you, the 5G, your smart device, your phones.' We had something completely foreign in the human body that had never been launched large scale at a time when we were seeing 5G going into schools and hospitals (plus the Musk satellites) and she believed the 'vaccine' transmission had something to do with this: '... if these people have this inside of them ... it can act like an antenna and actually transmit it outwardly as well.' The synthetic substance produced its own voltage and so it could have that kind of effect. This fits with my own contention that the nano receiver-transmitters are designed to connect people to the

Smart Grid and break the receiver-transmitter connection to expanded consciousness. That would explain the French energy healer's experience of the disconnection of body from 'soul' with those who have had the 'vaccine'. The nanobots, self-replicating inside the body, would also transmit the synthetic frequency which could be picked up through close interaction by those who have not been 'vaccinated'. Madej speculated that perhaps it was 5G and increased levels of other radiation that was causing the symptoms directly although interestingly she said that non-'vaccinated' patients had shown improvement when they were away from the 'vaccinated' person they had interacted with. It must be remembered that you can control frequency and energy with your mind and you can consciously create energetic barriers or bubbles with the mind to stop damaging frequencies from penetrating your field. American paediatrician Dr Larry Palevsky said the 'vaccine' was not a 'vaccine' and was never designed to protect from a 'viral' infection. He called it 'a massive, brilliant propaganda of genocide' because they didn't have to inject everyone to get the result they wanted. He said the content of the jabs was able to infuse any material into the brain, heart, lungs, kidneys, liver, sperm and female productive system. 'This is genocide; this is a weapon of mass destruction.' At the same time American colleges were banning students from attending if they didn't have this life-changing and potentially life-ending 'vaccine'. Class action lawsuits must follow when the consequences of this college fascism come to light. As the book was going to press came reports about fertility effects on sperm in 'vaccinated' men which would absolutely fit with what I have been saying and hospitals continued to fill with 'vaccine' reactions. Another question is what about transmission via blood transfusions? The NHS has extended blood donation restrictions from seven days after a 'Covid vaccination' to 28 days after even a sore arm reaction.

I said in the spring of 2020 that the then touted 'Covid vaccine' would be ongoing each year like the flu jab. A year later Pfizer CEO, the appalling Albert Bourla, said people would 'likely' need a 'booster dose' of the 'vaccine' within 12 months of getting 'fully

vaccinated' and then a yearly shot. 'Variants will play a key role', he said confirming the point. Johnson & Johnson CEO Alex Gorsky also took time out from his 'vaccine' disaster to say that people may need to be vaccinated against 'Covid-19' each year. UK Health Secretary, the psychopath Matt Hancock, said additional 'boosters' would be available in the autumn of 2021. This is the trap of the 'vaccine passport'. The public will have to accept every last 'vaccine' they introduce, including for the fake 'variants', or it would cease to be valid. The only other way in some cases would be continuous testing with a test not testing for the 'virus' and what is on the swabs constantly pushed up your nose towards the brain every time?

'Vaccines' changing behaviour

I mentioned in the body of the book how I believed we would see gathering behaviour changes in the 'vaccinated' and I am already hearing such comments from the non-'vaccinated' describing behaviour changes in friends, loved ones and work colleagues. This will only increase as the self-replicating synthetic material and nanoparticles expand in body and brain. An article in the *Guardian* in 2016 detailed research at the University of Virginia in Charlottesville which developed a new method for controlling brain circuits associated with complex animal behaviour. The method, dubbed 'magnetogenetics', involves genetically-engineering a protein called ferritin, which stores and releases iron, to create a magnetised substance – 'Magneto' – that can activate specific groups of nerve cells from a distance. This is claimed to be an advance on other methods of brain activity manipulation known as optogenetics and chemogenetics (the Cult has been developing methods of brain control for a long time). The ferritin technique is said to be non-invasive and able to activate neurons 'rapidly and reversibly'. In other words, human thought and perception. The article said that earlier studies revealed how nerve cell proteins 'activated by heat and mechanical pressure can be genetically engineered so that they become sensitive to radio waves and magnetic fields, by attaching them to an iron-storing protein called ferritin, or to inorganic

paramagnetic particles'. Sensitive to radio waves and magnetic fields? You mean like 5G, 6G and 7G? This is the human-AI Smart Grid hive mind we are talking about. The *Guardian* article said:

... the researchers injected Magneto into the striatum of freely behaving mice, a deep brain structure containing dopamine-producing neurons that are involved in reward and motivation, and then placed the animals into an apparatus split into magnetised and non-magnetised sections.

Mice expressing Magneto spent far more time in the magnetised areas than mice that did not, because activation of the protein caused the striatal neurons expressing it to release dopamine, so that the mice found being in those areas rewarding. This shows that Magneto can remotely control the firing of neurons deep within the brain, and also control complex behaviours.

Make no mistake this basic methodology will be part of the 'Covid vaccine' cocktail and using magnetics to change brain function through electromagnetic field frequency activation. The Pentagon is developing a 'Covid vaccine' using ferritin. Magnetism would explain changes in behaviour and why videos are appearing across the Internet as I write showing how magnets stick to the skin at the point of the 'vaccine' shot. Once people take these 'vaccines' anything becomes possible in terms of brain function and illness which will be blamed on 'Covid-19' and 'variants'. Magnetic field manipulation would further explain why the non-'vaccinated' are reporting the same symptoms as the 'vaccinated' they interact with and why those symptoms are reported to decrease when not in their company. Interestingly 'Magneto', a 'mutant', is a character in the Marvel Comic *X-Men* stories with the ability to manipulate magnetic fields and he believes that mutants should fight back against their human oppressors by any means necessary. The character was born Erik Lehnsherr to a Jewish family in Germany.

Cult-controlled courts

The European Court of Human Rights opened the door for mandatory 'Covid-19 vaccines' across the continent when it ruled in a Czech Republic dispute over childhood immunisation that legally

enforced vaccination could be 'necessary in a democratic society'. The 17 judges decided that compulsory vaccinations did not breach human rights law. On the face of it the judgement was so inverted you gasp for air. If not having a vaccine infused into your body is not a human right then what is? Ah, but they said human rights law which has been specifically written to delete all human rights at the behest of the state (the Cult). Article 8 of the European Convention on Human Rights relates to the right to a private life. The crucial word here is '*except*':

There shall be no interference by a public authority with the exercise of this right EXCEPT such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic wellbeing of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others [My emphasis].

No interference *except* in accordance with the law means there *are* no 'human rights' *except* what EU governments decide you can have at their behest. 'As is necessary in a democratic society' explains that reference in the judgement and 'in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others' gives the EU a coach and horses to ride through 'human rights' and scatter them in all directions. The judiciary is not a check and balance on government extremism; it is a vehicle to enforce it. This judgement was almost laughably predictable when the last thing the Cult wanted was a decision that went against mandatory vaccination. Judges rule over and over again to benefit the system of which they are a part. Vaccination disputes that come before them are invariably delivered in favour of doctors and authorities representing the view of the state which owns the judiciary. Oh, yes, and we have even had calls to stop putting 'Covid-19' on death certificates within 28 days of a 'positive test' because it is claimed the practice makes the 'vaccine' appear not to work. They are laughing at you.

The scale of madness, inhumanity and things to come was highlighted when those not 'vaccinated' for 'Covid' were refused evacuation from the Caribbean island of St Vincent during massive volcanic eruptions. Cruise ships taking residents to the safety of another island allowed only the 'vaccinated' to board and the rest were left to their fate. Even in life and death situations like this we see 'Covid' stripping people of their most basic human instincts and the insanity is even more extreme when you think that fake 'vaccine'-makers are not even claiming their body-manipulating concoctions stop 'infection' and 'transmission' of a 'virus' that doesn't exist. St Vincent Prime Minister Ralph Gonsalves said: 'The chief medical officer will be identifying the persons already vaccinated so that we can get them on the ship.' Note again the power of the chief medical officer who, like Whitty in the UK, will be answering to the World Health Organization. This is the Cult network structure that has overridden politicians who 'follow the science' which means doing what WHO-controlled 'medical officers' and 'science advisers' tell them. Gonsalves even said that residents who were 'vaccinated' after the order so they could board the ships would still be refused entry due to possible side effects such as 'wooziness in the head'. The good news is that if they were woozy enough in the head they could qualify to be prime minister of St Vincent.

Microchipping freedom

The European judgement will be used at some point to justify moves to enforce the 'Covid' DNA-manipulating procedure. Sandra Ro, CEO of the Global Blockchain Business Council, told a World Economic Forum event that she hoped 'vaccine passports' would help to 'drive forced consent and standardisation' of global digital identity schemes: 'I'm hoping with the desire and global demand for some sort of vaccine passport – so that people can get travelling and working again – [it] will drive forced consent, standardisation, and frankly, cooperation across the world.' The lady is either not very bright, or thoroughly mendacious, to use the term 'forced consent'.

You do not 'consent' if you are forced – you *submit*. She was describing what the plan has been all along and that's to enforce a digital identity on every human without which they could not function. 'Vaccine passports' are opening the door and are far from the end goal. A digital identity would allow you to be tracked in everything you do in cyberspace and this is the same technique used by Cult-owned China to enforce its social credit system of total control. The ultimate 'passport' is planned to be a microchip as my books have warned for nearly 30 years. Those nice people at the Pentagon working for the Cult-controlled Defense Advanced Research Projects Agency (DARPA) claimed in April, 2021, they have developed a microchip inserted under the skin to detect 'asymptomatic Covid-19 infection' before it becomes an outbreak and a 'revolutionary filter' that can remove the 'virus' from the blood when attached to a dialysis machine. The only problems with this are that the 'virus' does not exist and people transmitting the 'virus' with no symptoms is brain-numbing bullshit. This is, of course, not a ruse to get people to be microchipped for very different reasons. DARPA also said it was producing a one-stop 'vaccine' for the 'virus' and all 'variants'. One of the most sinister organisations on Planet Earth is doing this? Better have it then. These people are insane because Wetiko that possesses them is insane.

Researchers from the Salk Institute in California announced they have created an embryo that is part human and part monkey. My books going back to the 1990s have exposed experiments in top secret underground facilities in the United States where humans are being crossed with animal and non-human 'extraterrestrial' species. They are now easing that long-developed capability into the public arena and there is much more to come given we are dealing with psychiatric basket cases. Talking of which – Elon Musk's scientists at Neuralink trained a monkey to play Pong and other puzzles on a computer screen using a joystick and when the monkey made the correct move a metal tube squirted banana smoothie into his mouth which is the basic technique for training humans into unquestioning compliance. Two Neuralink chips were in the monkey's skull and

more than 2,000 wires ‘fanned out’ into its brain. Eventually the monkey played a video game purely with its brain waves. Psychopathic narcissist Musk said the ‘breakthrough’ was a step towards putting Neuralink chips into human skulls and merging minds with artificial intelligence. *Exactly*. This man is so dark and Cult to his DNA.

World Economic Fascism (WEF)

The World Economic Forum is telling you the plan by the statements made at its many and various events. Cult-owned fascist YouTube CEO Susan Wojcicki spoke at the 2021 WEF Global Technology Governance Summit (see the name) in which 40 governments and 150 companies met to ensure ‘the responsible design and deployment of emerging technologies’. Orwellian translation: ‘Ensuring the design and deployment of long-planned technologies will advance the Cult agenda for control and censorship.’ Freedom-destroyer and Nuremberg-bound Wojcicki expressed support for tech platforms like hers to censor content that is ‘technically legal but could be harmful’. Who decides what is ‘harmful’? She does and they do. ‘Harmful’ will be whatever the Cult doesn’t want people to see and we have legislation proposed by the UK government that would censor content on the basis of ‘harm’ no matter if the information is fair, legal and provably true. Make that *especially* if it is fair, legal and provably true. Wojcicki called for a global coalition to be formed to enforce content moderation standards through automated censorship. This is a woman and mega-censor so self-deluded that she shamelessly accepted a ‘free expression’ award – *Wojcicki* – in an event sponsored by her own *YouTube*. They have no shame and no self-awareness.

You know that ‘Covid’ is a scam and Wojcicki a Cult operative when YouTube is censoring medical and scientific opinion purely on the grounds of whether it supports or opposes the Cult ‘Covid’ narrative. Florida governor Ron DeSantis compiled an expert panel with four professors of medicine from Harvard, Oxford, and Stanford Universities who spoke against forcing children and

vaccinated people to wear masks. They also said there was no proof that lockdowns reduced spread or death rates of 'Covid-19'. Cult-gofer Wojcicki and her YouTube deleted the panel video 'because it included content that contradicts the consensus of local and global health authorities regarding the efficacy of masks to prevent the spread of Covid-19'. This 'consensus' refers to what the Cult tells the World Health Organization to say and the WHO tells 'local health authorities' to do. Wojcicki knows this, of course. The panellists pointed out that censorship of scientific debate was responsible for deaths from many causes, but Wojcicki couldn't care less. She would not dare go against what she is told and as a disgrace to humanity she wouldn't want to anyway. The UK government is seeking to pass a fascist 'Online Safety Bill' to specifically target with massive fines and other means non-censored video and social media platforms to make them censor 'lawful but harmful' content like the Cult-owned Facebook, Twitter, Google and YouTube. What is 'lawful but harmful' would be decided by the fascist Blair-created Ofcom.

Another WEF obsession is a cyber-attack on the financial system and this is clearly what the Cult has planned to take down the bank accounts of everyone – except theirs. Those that think they have enough money for the Cult agenda not to matter to them have got a big lesson coming if they continue to ignore what is staring them in the face. The World Economic Forum, funded by Gates and fronted by Klaus Schwab, announced it would be running a 'simulation' with the Russian government and global banks of just such an attack called Cyber Polygon 2021. What they simulate – as with the 'Covid' Event 201 – they plan to instigate. The WEF is involved in a project with the Cult-owned Carnegie Endowment for International Peace called the WEF-Carnegie Cyber Policy Initiative which seeks to merge Wall Street banks, 'regulators' (I love it) and intelligence agencies to 'prevent' (arrange and allow) a cyber-attack that would bring down the global financial system as long planned by those that control the WEF and the Carnegie operation. The Carnegie Endowment for International Peace sent an instruction to First World

War US President Woodrow Wilson not to let the war end before society had been irreversibly transformed.

The Wuhan lab diversion

As I close, the Cult-controlled authorities and lapdog media are systematically pushing 'the virus was released from the Wuhan lab' narrative. There are two versions – it happened by accident and it happened on purpose. Both are nonsense. The perceived existence of the never-shown-to-exist 'virus' is vital to sell the impression that there is actually an infective agent to deal with and to allow the endless potential for terrifying the population with 'variants' of a 'virus' that does not exist. The authorities at the time of writing are going with the 'by accident' while the alternative media is promoting the 'on purpose'. Cable news host Tucker Carlson who has questioned aspects of lockdown and 'vaccine' compulsion has bought the Wuhan lab story. 'Everyone now agrees' he said. Well, I don't and many others don't and the question is *why* does the system and its media suddenly 'agree'? When the media moves as one unit with a narrative it is always a lie – witness the hour by hour mendacity of the 'Covid' era. Why would this Cult-owned combination which has unleashed lies like machine gun fire suddenly 'agree' to tell the truth??

Much of the alternative media is buying the lie because it fits the conspiracy narrative, but it's the *wrong* conspiracy. The real conspiracy is that *there is no virus* and that is what the Cult is desperate to hide. The idea that the 'virus' was released by accident is ludicrous when the whole 'Covid' hoax was clearly long-planned and waiting to be played out as it was so fast in accordance with the Rockefeller document and Event 201. So they prepared everything in detail over decades and then sat around strumming their fingers waiting for an 'accidental' release from a bio-lab? *What??* It's crazy. Then there's the 'on purpose' claim. You want to circulate a 'deadly virus' and hide the fact that you've done so and you release it down the street from the highest-level bio-lab in China? I repeat – *What??*

You would release it far from that lab to stop any association being made. But, no, we'll do it in a place where the connection was certain to be made. Why would you need to scam 'cases' and 'deaths' and pay hospitals to diagnose 'Covid-19' if you had a real 'virus'? What are sections of the alternative media doing believing this crap? Where were all the mass deaths in Wuhan from a 'deadly pathogen' when the recovery to normal life after the initial propaganda was dramatic in speed? Why isn't the 'deadly pathogen' now circulating all over China with bodies in the street? Once again we have the technique of tell them what they want to hear and they will likely believe it. The alternative media has its 'conspiracy' and with Carlson it fits with his 'China is the danger' narrative over years. China *is* a danger as a global Cult operations centre, but not for this reason. The Wuhan lab story also has the potential to instigate conflict with China when at some stage the plan is to trigger a Problem-Reaction-Solution confrontation with the West. Question everything – *everything* – and especially when the media agrees on a common party line.

Third wave ... fourth wave ... fifth wave ...

As the book went into production the world was being set up for more lockdowns and a 'third wave' supported by invented 'variants' that were increasing all the time and will continue to do so in public statements and computer programs, but not in reality. India became the new Italy in the 'Covid' propaganda campaign and we were told to be frightened of the new 'Indian strain'. Somehow I couldn't find it within myself to do so. A document produced for the UK government entitled 'Summary of further modelling of easing of restrictions – Roadmap Step 2' declared that a third wave was inevitable (of course when it's in the script) and it would be the fault of children and those who refuse the health-destroying fake 'Covid vaccine'. One of the computer models involved came from the Cult-owned *Imperial College* and the other from Warwick University which I wouldn't trust to tell me the date in a calendar factory. The document states that both models presumed extremely high uptake

of the 'Covid vaccines' and didn't allow for 'variants'. The document states: 'The resurgence is a result of some people (mostly children) being ineligible for vaccination; others choosing not to receive the vaccine; and others being vaccinated but not perfectly protected.' The mendacity takes the breath away. Okay, blame those with a brain who won't take the DNA-modifying shots and put more pressure on children to have it as 'trials' were underway involving children as young as six months with parents who give insanity a bad name. Massive pressure is being put on the young to have the fake 'vaccine' and child age consent limits have been systematically lowered around the world to stop parents intervening. Most extraordinary about the document was its claim that the 'third wave' would be driven by 'the resurgence in both hospitalisations and deaths ... dominated by *those that have received two doses of the vaccine*, comprising around 60-70% of the wave respectively'. The predicted peak of the 'third wave' suggested 300 deaths per day with 250 of them *fully 'vaccinated' people*. How many more lies do acquiescers need to be told before they see the obvious? Those who took the jab to 'protect themselves' are projected to be those who mostly get sick and die? So what's in the 'vaccine'? The document went on:

It is possible that a summer of low prevalence could be followed by substantial increases in incidence over the following autumn and winter. Low prevalence in late summer should not be taken as an indication that SARS-CoV-2 has retreated or that the population has high enough levels of immunity to prevent another wave.

They are telling you the script and while many British people believed 'Covid' restrictions would end in the summer of 2021 the government was preparing for them to be ongoing. Authorities were awarding contracts for 'Covid marshals' to police the restrictions with contracts starting in July, 2021, and going through to January 31st, 2022, and the government was advertising for 'Media Buying Services' to secure media propaganda slots worth a potential £320 million for 'Covid-19 campaigns' with a contract not ending until March, 2022. The recipient – via a list of other front companies – was reported to be American media marketing giant Omnicom Group

Inc. While money is no object for 'Covid' the UK waiting list for all other treatment – including life-threatening conditions – passed 4.5 million. Meantime the Cult is seeking to control all official 'inquiries' to block revelations about what has really been happening and why. It must not be allowed to – we need Nuremberg jury trials in every country. The cover-up doesn't get more obvious than appointing ultra-Zionist professor Philip Zelikow to oversee two dozen US virologists, public health officials, clinicians, former government officials and four American 'charitable foundations' to 'learn the lessons' of the 'Covid' debacle. The personnel will be those that created and perpetuated the 'Covid' lies while Zelikow is the former executive director of the 9/11 Commission who ensured that the truth about those attacks never came out and produced a report that must be among the most mendacious and manipulative documents ever written – see *The Trigger* for the detailed exposure of the almost unimaginable 9/11 story in which Sabbatians can be found at every level.

Passive no more

People are increasingly challenging the authorities with amazing numbers of people taking to the streets in London well beyond the ability of the Face-Nappies to stop them. Instead the Nappies choose situations away from the mass crowds to target, intimidate, and seek to promote the impression of 'violent protestors'. One such incident happened in London's Hyde Park. Hundreds of thousands walking through the streets in protest against 'Covid' fascism were ignored by the Cult-owned BBC and most of the rest of the mainstream media, but they delighted in reporting how police were injured in 'clashes with protestors'. The truth was that a group of people gathered in Hyde Park at the end of one march when most had gone home and they were peacefully having a good time with music and chat. Face-Nappies who couldn't deal with the full-march crowd then waded in with their batons and got more than they bargained for. Instead of just standing for this criminal brutality the crowd used their numerical superiority to push the Face-Nappies out of the

park. Eventually the Nappies turned and ran. Unfortunately two or three idiots in the crowd threw drink cans striking two officers which gave the media and the government the image they wanted to discredit the 99.9999 percent who were peaceful. The idiots walked straight into the trap and we must always be aware of potential agent provocateurs used by the authorities to discredit their targets.

This response from the crowd – the can people apart – must be a turning point when the public no longer stand by while the innocent are arrested and brutally attacked by the Face-Nappies. That doesn't mean to be violent, that's the last thing we need. We'll leave the violence to the Face-Nappies and government. But it does mean that when the Face-Nappies use violence against peaceful people the numerical superiority is employed to stop them and make citizen's arrests or Common Law arrests for a breach of the peace. The time for being passive in the face of fascism is over.

We are the many, they are the few, and we need to make that count before there is no freedom left and our children and grandchildren face an ongoing fascist nightmare.

COME ON PEOPLE – IT'S TIME.

One final thought ...

The power of love
A force from above
Cleaning my soul
Flame on burn desire
Love with tongues of fire
Purge the soul
Make love your goal

I'll protect you from the hooded claw
Keep the vampires from your door
When the chips are down I'll be around
With my undying, death-defying
Love for you

Envy will hurt itself
Let yourself be beautiful
Sparkling love, flowers
And pearls and pretty girls
Love is like an energy
Rushin' rushin' inside of me

This time we go sublime
Lovers entwine, divine, divine,
Love is danger, love is pleasure
Love is pure – the only treasure

I'm so in love with you
Purge the soul
Make love your goal

The power of love
A force from above
Cleaning my soul
The power of love
A force from above
A sky-scraping dove

Flame on burn desire
Love with tongues of fire
Purge the soul
Make love your goal

Frankie Goes To Hollywood

APPENDIX

Cowan-Kaufman-Morell Statement on Virus Isolation (SOVI)

Isolation: The action of isolating; the fact or condition of being isolated or standing alone; separation from other things or persons; solitariness

Oxford English Dictionary

The controversy over whether the SARS-CoV-2 virus has ever been isolated or purified continues. However, using the above definition, common sense, the laws of logic and the dictates of science, any unbiased person must come to the conclusion that the SARS-CoV-2 virus has never been isolated or purified. As a result, no confirmation of the virus' existence can be found. The logical, common sense, and scientific consequences of this fact are:

- the structure and composition of something not shown to exist can't be known, including the presence, structure, and function of any hypothetical spike or other proteins;
- the genetic sequence of something that has never been found can't be known;
- "variants" of something that hasn't been shown to exist can't be known;
- it's impossible to demonstrate that SARS-CoV-2 causes a disease called Covid-19.

In as concise terms as possible, here's the proper way to isolate, characterize and demonstrate a new virus. First, one takes samples (blood, sputum, secretions) from many people (e.g. 500) with symptoms which are unique and specific enough to characterize an illness. Without mixing these samples with ANY tissue or products that also contain genetic material, the virologist macerates, filters and ultracentrifuges i.e. *purifies* the specimen. This common virology technique, done for decades to isolate bacteriophages¹ and so-called giant viruses in every virology lab, then allows the virologist to demonstrate with electron microscopy thousands of identically sized and shaped particles. These particles are the isolated and purified virus.

These identical particles are then checked for uniformity by physical and/or microscopic techniques. Once the purity is determined, the particles may be further characterized. This would include examining the structure, morphology, and chemical composition of the particles. Next, their genetic makeup is characterized by extracting the genetic material directly from the purified particles and using genetic-sequencing techniques, such as Sanger sequencing, that have also been around for decades. Then one does an analysis to confirm that these uniform particles are exogenous (outside) in origin as a virus is conceptualized to be, and not the normal breakdown products of dead and dying tissues.² (As of May 2020, we know that virologists have no way to determine whether the particles they're seeing are viruses or just normal breakdown products of dead and dying tissues.)³

1 Isolation, characterization and analysis of bacteriophages from the haloalkaline lake Elmenteita, Kenya Julia Khayeli Akhwale et al, PLOS One, Published: April 25, 2019.
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0215734> – accessed 2/15/21

2 "Extracellular Vesicles Derived From Apoptotic Cells: An Essential Link Between Death and Regeneration," Maojiao Li et al, Frontiers in Cell and Developmental Biology, 2020 October 2.
<https://www.frontiersin.org/articles/10.3389/fcell.2020.573511/full> – accessed 2/15/21

If we have come this far then we have fully isolated, characterized, and genetically sequenced an exogenous virus particle. However, we still have to show it is causally related to a disease. This is carried out by exposing a group of healthy subjects (animals are usually used) to this isolated, purified virus in the manner in which the disease is thought to be transmitted. If the animals get sick with the same disease, as confirmed by clinical and autopsy findings, one has now shown that the virus actually causes a disease. This demonstrates infectivity and transmission of an infectious agent.

None of these steps has even been attempted with the SARS-CoV-2 virus, nor have all these steps been successfully performed for any so-called pathogenic virus. Our research indicates that a single study showing these steps does not exist in the medical literature.

Instead, since 1954, virologists have taken unpurified samples from a relatively few people, often less than ten, with a similar disease. They then minimally process this sample and inoculate this unpurified sample onto tissue culture containing usually four to six other types of material – all of which contain identical genetic material as to what is called a “virus.” The tissue culture is starved and poisoned and naturally disintegrates into many types of particles, some of which contain genetic material. Against all common sense, logic, use of the English language and scientific integrity, this process is called “virus isolation.” This brew containing fragments of genetic material from many sources is then subjected to genetic analysis, which then creates in a computer-simulation process the alleged sequence of the alleged virus, a so called in silico genome. At no time is an actual virus confirmed by electron microscopy. At no time is a genome extracted and sequenced from an actual virus. This is scientific fraud.

The observation that the unpurified specimen — inoculated onto tissue culture along with toxic antibiotics, bovine fetal tissue, amniotic fluid and other tissues — destroys the kidney tissue onto which it is inoculated is given as evidence of the virus' existence and pathogenicity. This is scientific fraud.

From now on, when anyone gives you a paper that suggests the SARS-CoV-2 virus has been isolated, please check the methods sections. If the researchers used Vero cells or any other culture method, you know that their process was not isolation. You will hear the following excuses for why actual isolation isn't done:

1. There were not enough virus particles found in samples from patients to analyze.
2. Viruses are intracellular parasites; they can't be found outside the cell in this manner.

If No. 1 is correct, and we can't find the virus in the sputum of sick people, then on what evidence do we think the virus is dangerous or even lethal? If No. 2 is correct, then how is the virus spread from person to person? We are told it emerges from the cell to infect others. Then why isn't it possible to find it?

Finally, questioning these virology techniques and conclusions is not some distraction or divisive issue. Shining the light on this truth is essential to stop this terrible fraud that humanity is confronting. For, as we now know, if the virus has never been isolated, sequenced or shown to cause illness, if the virus is imaginary, then why are we wearing masks, social distancing and putting the whole world into prison?

Finally, if pathogenic viruses don't exist, then what is going into those injectable devices erroneously called "vaccines," and what is their purpose? This scientific question is the most urgent and relevant one of our time.

We are correct. The SARS-CoV2 virus does not exist.

Sally Fallon Morell, MA

Dr. Thomas Cowan, MD

Dr. Andrew Kaufman, MD

Bibliography

- Alinsky, Saul:** *Rules for Radicals* (Vintage, 1989)
- Antelman, Rabbi Marvin:** *To Eliminate the Opiate* (Zahavia, 1974)
- Bastardi, Joe:** *The Climate Chronicles* (Relentless Thunder Press, 2018)
- Cowan, Tom:** *Human Heart, Cosmic Heart* (Chelsea Green Publishing, 2016)
- Cowan, Tom, and Fallon Morell, Sally:** *The Contagion Myth* (Skyhorse Publishing, 2020)
- Forbes, Jack D:** *Columbus And Other Cannibals – The Wetiko Disease of Exploitation, Imperialism, and Terrorism* (Seven Stories Press, 2008 – originally published in 1979)
- Gates, Bill:** *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need* (Allen Lane, 2021)
- Huxley, Aldous:** *Brave New World* (Chatto & Windus, 1932)
- Köhnlein, Dr Claus, and Engelbrecht, Torsten:** *Virus Mania* (emu-Verlag, Lahnstein, 2020)
- Lanza, Robert, and Berman, Bob:** *Biocentrism* (BenBella Books, 2010)
- Lash, John Lamb:** *Not In His Image* (Chelsea Green Publishing, 2006)
- Lester, Dawn, and Parker, David:** *What Really Makes You Ill – Why everything you thought you knew about disease is wrong* (Independently Published, 2019)
- Levy, Paul:** *Dispelling Wetiko, Breaking the Spell of Evil* (North Atlantic Books, 2013)
- Marx, Karl:** *A World Without Jews* (Philosophical Library, first edition, 1959)
- Mullis, Kary:** *Dancing Naked in the Mine Field* (Bloomsbury, 1999)
- O'Brien, Cathy:** *Trance-Formation of America* (Reality Marketing, 1995)
- Scholem, Gershon:** *The Messianic Idea in Judaism* (Schocken Books, 1994)
- Schwab, Klaus, and Davis, Nicholas:** *Shaping the Future of the Fourth Industrial Revolution: A guide to building a better world* (Penguin Books, 2018)
- Schwab, Klaus:** *The Great Reset* (Agentur Schweiz, 2020)
- Sunstein, Cass and Thaler, Richard:** *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Penguin, 2009)
- Swan, Shanna:** *Count Down: How Our Modern World Is Threatening Sperm Counts, Altering Male and Female Reproductive Development and Imperiling the Future of the Human Race* (Scribner, 2021)
- Tegmark, Max:** *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality* (Penguin, 2015)
- Velikovsky, Immanuel:** *Worlds in Collision* (Paradigma, 2009)

Wilton, Robert: *The Last Days of the Romanovs* (Blurb, 2018, first published 1920)

Index

A

abusive relationships

- blaming themselves, abused as [ref1](#)
- children [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#)
- conspiracy theories [ref1](#)
- domestic abuse [ref1](#), [ref2](#)
- economic abuse and dependency [ref1](#)
- isolation [ref1](#)
- physical abuse [ref1](#)
- psychological abuse [ref1](#)
- signs of abuse [ref1](#)

addiction

- alcoholism [ref1](#)
- frequencies [ref1](#)
- substance abuse [ref1](#), [ref2](#)
- technology [ref1](#), [ref2](#), [ref3](#)

Adelson, Sheldon [ref1](#), [ref2](#), [ref3](#)

Agenda 21/Agenda 2030 (UN) [ref1](#), [ref2](#), [ref3](#), [ref4](#)

AIDs/HIV [ref1](#)

- causal link between HIV and AIDs [ref1](#), [ref2](#)

retroviruses [ref1](#)

- testing [ref1](#), [ref2](#)
- trial-run for Covid-19, as [ref1](#), [ref2](#)

aliens/extraterrestrials [ref1](#), [ref2](#)

aluminium [ref1](#)

Amazon [ref1](#), [ref2](#), [ref3](#)

amplification cycles [ref1](#), [ref2](#)
anaphylactic shock [ref1](#), [ref2](#), [ref3](#), [ref4](#)
animals [ref1](#), [ref2](#), [ref3](#)
antibodies [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Antifa [ref1](#), [ref2](#), [ref3](#), [ref4](#)
antigens [ref1](#), [ref2](#)
anti-Semitism [ref1](#), [ref2](#), [ref3](#)
Archons [ref1](#), [ref2](#)
 consciousness [ref1](#), [ref2](#), [ref3](#)
 energy [ref1](#), [ref2](#), [ref3](#)
 ennoa [ref1](#)
 genetic manipulation [ref1](#), [ref2](#)
 inversion [ref1](#), [ref2](#), [ref3](#)
 lockdowns [ref1](#)
 money [ref1](#)
 radiation [ref1](#)
 religion [ref1](#), [ref2](#)
 technology [ref1](#), [ref2](#), [ref3](#)
 Wetiko factor [ref1](#), [ref2](#), [ref3](#), [ref4](#)
artificial intelligence (AI) [ref1](#)
army made up of robots [ref1](#), [ref2](#)
 Human 2.0 [ref1](#), [ref2](#)
 Internet [ref1](#)
 MHRA [ref1](#)
 Morgellons fibres [ref1](#), [ref2](#)
 Smart Grid [ref1](#)
 Wetiko factor [ref1](#)
asymptomatic, Covid-19 as [ref1](#), [ref2](#), [ref3](#)
aviation industry [ref1](#)

B

banking, finance and money [ref1](#), [ref2](#), [ref3](#)

2008 crisis [ref1](#), [ref2](#)

boom and bust [ref1](#)

cashless digital money systems [ref1](#)

central banks [ref1](#)

credit [ref1](#)

digital currency [ref1](#)

fractional reserve lending [ref1](#)

Great Reset [ref1](#)

guaranteed income [ref1](#), [ref2](#), [ref3](#)

Human 2.0 [ref1](#)

incomes, destruction of [ref1](#), [ref2](#)

interest [ref1](#)

one per cent [ref1](#), [ref2](#)

scams [ref1](#)

BBC [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

Becker-Phelps, Leslie [ref1](#)

Behavioural Insights Team (BIT) (Nudge Unit) [ref1](#), [ref2](#), [ref3](#)

behavioural scientists *and* psychologists, advice from [ref1](#), [ref2](#)

Bezos, Jeff [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Biden, Hunter [ref1](#)

Biden, Joe [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#), [ref11](#),
[ref12](#), [ref13](#), [ref14](#), [ref15](#), [ref16](#), [ref17](#)

Big Pharma

cholesterol [ref1](#)

health professionals [ref1](#), [ref2](#)

immunity from prosecution in US [ref1](#)

vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

Wetiko factor [ref1](#), [ref2](#)

WHO [ref1](#), [ref2](#), [ref3](#)

Bill and Melinda Gates Foundation [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#),
[ref7](#)

billionaires [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#) [ref10](#), [ref11](#)
bird flu (H5N1) [ref1](#)
Black Lives Matter (BLM) [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Blair, Tony [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
Brin, Sergei [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
British Empire [ref1](#)
Bush, George HW [ref1](#), [ref2](#)
Bush, George W [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Byrd, Robert [ref1](#)

C

Canada

Global Cult [ref1](#)
hate speech [ref1](#)
internment [ref1](#)
masks [ref1](#)
old people [ref1](#)
SARS-COV-2 [ref1](#)
satellites [ref1](#)
vaccines [ref1](#)
wearable technology [ref1](#)

Capitol Hill riot [ref1](#), [ref2](#)

agents provocateur [ref1](#)
Antifa [ref1](#)
Black Lives Matter (BLM) [ref1](#), [ref2](#)
QAnon [ref1](#)
security precautions, lack of [ref1](#), [ref2](#), [ref3](#)

carbon dioxide [ref1](#), [ref2](#)

care homes, deaths in [ref1](#), [ref2](#)

cashless digital money systems [ref1](#)

censorship [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

fact-checkers [ref1](#)

masks [ref1](#)

media [ref1](#), [ref2](#)

private messages [ref1](#)

social media [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)

transgender persons [ref1](#)

vaccines [ref1](#), [ref2](#), [ref3](#)

Wokeness [ref1](#)

Centers for Disease Control (CDC) (United States) [ref1](#), [ref2](#), [ref3](#),
[ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#), [ref11](#), [ref12](#), [ref13](#)

centralisation [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

chakras [ref1](#)

change agents [ref1](#), [ref2](#), [ref3](#)

chemtrails [ref1](#), [ref2](#), [ref3](#)

chief medical officers and scientific advisers [ref1](#), [ref2](#), [ref3](#), [ref4](#),
[ref5](#), [ref6](#)

children *see also* **young people**

abuse [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#)

care, taken into [ref1](#), [ref2](#), [ref3](#)

education [ref1](#), [ref2](#), [ref3](#), [ref4](#)

energy [ref1](#)

family courts [ref1](#)

hand sanitisers [ref1](#)

human sacrifice [ref1](#)

lockdowns [ref1](#), [ref2](#), [ref3](#)

masks [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

mental health [ref1](#)

old people [ref1](#)

parents, replacement of [ref1](#), [ref2](#)

Psyop (psychological operation), Covid as a [ref1](#), [ref2](#)

reframing [ref1](#)

smartphone addiction [ref1](#)

social distancing and isolation [ref1](#)
social media [ref1](#)
transgender persons [ref1](#), [ref2](#)
United States [ref1](#)
vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#)
Wetiko factor [ref1](#)

China [ref1](#), [ref2](#), [ref3](#), [ref4](#)

anal swab tests [ref1](#)
Chinese Revolution [ref1](#), [ref2](#), [ref3](#)
digital currency [ref1](#)
Global Cult [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#)
guaranteed income [ref1](#)
Imperial College [ref1](#)
Israel [ref1](#)
lockdown [ref1](#), [ref2](#)
masculinity crisis [ref1](#)
masks [ref1](#)
media [ref1](#)
origins of virus in China [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
pollution causing respiratory diseases [ref1](#)
Sabbatians [ref1](#), [ref2](#)
Smart Grid [ref1](#), [ref2](#)
social credit system [ref1](#)
testing [ref1](#), [ref2](#)
United States [ref1](#), [ref2](#)
vaccines [ref1](#), [ref2](#)
Wetiko factor [ref1](#)
wet market conspiracy [ref1](#)
Wuhan [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

cholesterol [ref1](#), [ref2](#)

Christianity [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

criticism [ref1](#)
cross, inversion of the [ref1](#)

Nag Hammadi texts [ref1](#), [ref2](#), [ref3](#)

Roman Catholic Church [ref1](#), [ref2](#)

Sabbatians [ref1](#), [ref2](#)

Satan [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Wokeness [ref1](#)

class [ref1](#), [ref2](#)

climate change hoax [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Agenda 21/Agenda 2030 [ref1](#), [ref2](#), [ref3](#)

carbon dioxide [ref1](#), [ref2](#)

Club of Rome [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

fear [ref1](#)

funding [ref1](#)

Global Cult [ref1](#)

green new deals [ref1](#)

green parties [ref1](#)

inversion [ref1](#)

perception, control of [ref1](#)

PICC [ref1](#)

reframing [ref1](#)

temperature, increases in [ref1](#)

United Nations [ref1](#), [ref2](#)

Wikipedia [ref1](#)

Wokeness [ref1](#), [ref2](#)

Clinton, Bill [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)

Clinton, Hillary [ref1](#), [ref2](#), [ref3](#)

the cloud [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

Club of Rome and climate change hoax [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

cognitive therapy [ref1](#)

Cohn, Roy [ref1](#)

Common Law [ref1](#)

Admiralty Law [ref1](#)

arrests [ref1](#), [ref2](#)

contractual law, Statute Law as [ref1](#)

corporate entities, people as [ref1](#)

legalese [ref1](#)

sea, law of the [ref1](#)

Statute Law [ref1](#)

Common Purpose leadership programme [ref1](#), [ref2](#)

communism [ref1](#), [ref2](#)

co-morbidities [ref1](#)

computer-generated virus,

Covid-19 as [ref1](#), [ref2](#), [ref3](#)

computer models [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

connections [ref1](#), [ref2](#), [ref3](#), [ref4](#)

consciousness [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Archons [ref1](#), [ref2](#), [ref3](#)

expanded [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

experience [ref1](#)

heart [ref1](#)

infinity [ref1](#), [ref2](#)

religion [ref1](#), [ref2](#)

self-identity [ref1](#)

simulation thesis [ref1](#)

vaccines [ref1](#)

Wetiko factor [ref1](#), [ref2](#)

conspiracy theorists [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

contradictory rules [ref1](#)

contrails [ref1](#)

Corman-Drosten test [ref1](#), [ref2](#), [ref3](#), [ref4](#)

countermimicry [ref1](#), [ref2](#), [ref3](#)

Covid-19 vaccines *see* vaccines

Covidiots [ref1](#), [ref2](#)

Cowan, Tom [ref1](#), [ref2](#), [ref3](#), [ref4](#)

crimes against humanity [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

cyber-operations [ref1](#)

cyberwarfare [ref1](#)

D

DARPA (Defense Advanced Research Projects Agency) [ref1](#)

deaths

care homes [ref1](#)

certificates [ref1](#), [ref2](#), [ref3](#), [ref4](#)

mortality rate [ref1](#)

post-mortems/autopsies [ref1](#)

recording [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

deceit

pyramid of deceit [ref1](#), [ref2](#)

sequence of deceit [ref1](#)

decoding [ref1](#), [ref2](#), [ref3](#)

dehumanisation [ref1](#), [ref2](#), [ref3](#)

Delphi technique [ref1](#)

democracy [ref1](#)

dependency [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Descartes, René [ref1](#)

DNA

numbers [ref1](#)

vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#)

DNR (do not resuscitate)

orders [ref1](#)

domestic abuse [ref1](#), [ref2](#)

downgrading of Covid-19 [ref1](#)

Drosten, Christian [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

Duesberg, Peter [ref1](#), [ref2](#)

E

economic abuse [ref1](#)

Edmunds, John [ref1](#), [ref2](#)

education [ref1](#), [ref2](#), [ref3](#), [ref4](#)

electromagnetic spectrum [ref1](#), [ref2](#)

Enders, John [ref1](#)

energy

Archons [ref1](#), [ref2](#), [ref3](#)

children and young people [ref1](#)

consciousness [ref1](#)

decoding [ref1](#)

frequencies [ref1](#), [ref2](#), [ref3](#), [ref4](#)

heart [ref1](#)

human energy field [ref1](#)

source, humans as an energy [ref1](#), [ref2](#)

vaccines [ref1](#)

viruses [ref1](#)

ennoia [ref1](#)

Epstein, Jeffrey [ref1](#), [ref2](#)

eternal 'I' [ref1](#), [ref2](#)

ethylene oxide [ref1](#)

European Union [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Event [ref1](#) *and* Bill Gates [ref2](#)

exosomes, Covid-19 as natural defence mechanism called [ref1](#)

experience [ref1](#), [ref2](#)

Extinction Rebellion [ref1](#), [ref2](#)

F

Facebook

addiction [ref1](#), 448–50

Facebook

Archons [ref1](#)

censorship [ref1](#), [ref2](#), [ref3](#)

hate speech [ref1](#)

monopoly, as [ref1](#)

private messages, censorship of [ref1](#)

Sabbatians [ref1](#)

United States election fraud [ref1](#)

vaccines [ref1](#)

Wetiko factor [ref1](#)

fact-checkers [ref1](#)

Fauci, Anthony [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#),
[ref11](#), [ref12](#)

fear [ref1](#), [ref2](#), [ref3](#), [ref4](#)

climate change [ref1](#)

computer models [ref1](#)

conspiracy theories [ref1](#)

empty hospitals [ref1](#)

Italy [ref1](#), [ref2](#), [ref3](#)

lockdowns [ref1](#), [ref2](#), [ref3](#), [ref4](#)

masks [ref1](#), [ref2](#)

media [ref1](#), [ref2](#)

medical staff [ref1](#)

Psyop (psychological operation), Covid as a [ref1](#)

Wetiko factor [ref1](#), [ref2](#)

female infertility [ref1](#)

Fermi Paradox [ref1](#)

Ferguson, Neil [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

fertility, decline in [ref1](#)

The Field [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

finance *see* **banking, finance and money**

five-senses [ref1](#), [ref2](#)

Archons [ref1](#), [ref2](#), [ref3](#)

censorship [ref1](#)
consciousness, expansion of [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
decoding [ref1](#)
education [ref1](#), [ref2](#)
the Field [ref1](#), [ref2](#)
God, personification of [ref1](#)
infinity [ref1](#), [ref2](#)
media [ref1](#)
paranormal [ref1](#)
perceptual programming [ref1](#), [ref2](#)
Phantom Self [ref1](#)
pneuma not nous, using [ref1](#)
reincarnation [ref1](#)
self-identity [ref1](#)
Wetiko factor [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
5G [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)
Floyd, George and protests, killing of [ref1](#)
flu, re-labelling of [ref1](#), [ref2](#), [ref3](#)
food and water, control of [ref1](#), [ref2](#)
Freemasons [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
Frei, Rosemary [ref1](#)
frequencies
addictions [ref1](#)
Archons [ref1](#), [ref2](#), [ref3](#)
awareness [ref1](#)
chanting and mantras [ref1](#)
consciousness [ref1](#)
decoding [ref1](#), [ref2](#)
education [ref1](#)
electromagnetic (EMF) frequencies [ref1](#)
energy [ref1](#), [ref2](#), [ref3](#), [ref4](#)
fear [ref1](#)

the Field [ref1](#), [ref2](#) 5G [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#)
five-senses [ref1](#), [ref2](#)
ghosts [ref1](#)
Gnostics [ref1](#)
hive-minds [ref1](#)
human, meaning of [ref1](#)
light [ref1](#), [ref2](#)
love [ref1](#), [ref2](#)
magnetism [ref1](#)
perception [ref1](#)
reality [ref1](#), [ref2](#), [ref3](#)
simulation [ref1](#)
terror [ref1](#)
vaccines [ref1](#)
Wetiko [ref1](#), [ref2](#), [ref3](#)
Fuellmich, Reiner [ref1](#), [ref2](#), [ref3](#)
furlough/rescue payments [ref1](#)

G

Gallo, Robert [ref1](#), [ref2](#), [ref3](#)

Gates, Bill

Archons [ref1](#), [ref2](#), [ref3](#)
climate change [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Daily Pass tracking system [ref1](#)
Epstein [ref1](#)
fascism [ref1](#)
five senses [ref1](#)
GAVI [ref1](#)
Great Reset [ref1](#)
GSK [ref1](#)
Imperial College [ref1](#), [ref2](#)
Johns Hopkins University [ref1](#), [ref2](#), [ref3](#)

lockdowns [ref1](#), [ref2](#)
masks [ref1](#)
Nuremberg trial, proposal for [ref1](#), [ref2](#)
Rockefellers [ref1](#), [ref2](#)
social distancing and isolation [ref1](#)
Sun, dimming the [ref1](#)
synthetic meat [ref1](#), [ref2](#)
vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
Wellcome Trust [ref1](#)
Wetiko factor [ref1](#), [ref2](#), [ref3](#)
WHO [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#)
Wokeness [ref1](#)
World Economic Forum [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Gates, Melinda [ref1](#), [ref2](#), [ref3](#)
GAVI vaccine alliance [ref1](#)
genetics, manipulation of [ref1](#), [ref2](#), [ref3](#)
Germany [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#) *see also* **Nazi Germany**
Global Cult [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
anti-human, why Global Cult is [ref1](#)
Black Lives Matter (BLM) [ref1](#), [ref2](#), [ref3](#), [ref4](#)
China [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#)
climate change hoax [ref1](#)
contradictory rules [ref1](#)
Covid-19 [ref1](#), [ref2](#), [ref3](#)
fascism [ref1](#)
geographical origins [ref1](#)
immigration [ref1](#)
Internet [ref1](#)
mainstream media [ref1](#), [ref2](#)
masks [ref1](#), [ref2](#)
monarchy [ref1](#)
non-human dimension [ref1](#)

perception [ref1](#)
political parties [ref1](#), [ref2](#)
pyramidal hierarchy [ref1](#), [ref2](#), [ref3](#)
reframing [ref1](#)
Sabbatian-Frankism [ref1](#), [ref2](#)
science, manipulation of [ref1](#)
spider and the web [ref1](#)
transgender persons [ref1](#)
vaccines [ref1](#)
who controls the Cult [ref1](#)
Wokeness [ref1](#), [ref2](#), [ref3](#), [ref4](#)

globalisation [ref1](#), [ref2](#)

Gnostics [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Google [ref1](#), [ref2](#), [ref3](#), [ref4](#)

government

behavioural scientists and psychologists, advice from [ref1](#), [ref2](#)

definition [ref1](#)

Joint Biosecurity Centre (JBC) [ref1](#)

people, abusive relationship with [ref1](#)

Great Reset [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)

fascism [ref1](#), [ref2](#), [ref3](#)

financial system [ref1](#)

Human 2.0 [ref1](#)

water and food, control of [ref1](#)

green parties [ref1](#)

Griesz-Brisson, Margarite [ref1](#)

guaranteed income [ref1](#), [ref2](#), [ref3](#)

H

Hancock, Matt [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

hand sanitisers [ref1](#)

heart [ref1](#), [ref2](#)

hive-minds/groupthink [ref1](#), [ref2](#), [ref3](#)

holographs [ref1](#), [ref2](#), [ref3](#), [ref4](#)

hospitals, empty [ref1](#)

human, meaning of [ref1](#)

Human 2.0 [ref1](#)

addiction to technology [ref1](#)

artificial intelligence (AI) [ref1](#), [ref2](#)

elimination of Human 1.0 [ref1](#)

fertility, decline in [ref1](#)

Great Reset [ref1](#)

implantables [ref1](#)

money [ref1](#)

mRNA [ref1](#)

nanotechnology [ref1](#)

parents, replacement of [ref1](#), [ref2](#)

Smart Grid, connection to [ref1](#), [ref2](#)

synthetic biology [ref1](#), [ref2](#), [ref3](#), [ref4](#)

testosterone levels, decrease in [ref1](#)

transgender = transhumanism [ref1](#), [ref2](#), [ref3](#)

vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#)

human sacrifice [ref1](#), [ref2](#), [ref3](#)

Hunger Games Society [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

Huxley, Aldous [ref1](#), [ref2](#), [ref3](#)

I

identity politics [ref1](#), [ref2](#), [ref3](#)

Illuminati [ref1](#), [ref2](#)

illusory physical reality [ref1](#)

immigration [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Imperial College [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)

implantables [ref1](#), [ref2](#)

incomes, destruction of [ref1](#), [ref2](#)

Infinite Awareness [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Internet [ref1](#), [ref2](#) *see also* social media

artificial intelligence (AI) [ref1](#)

independent journalism, lack of [ref1](#)

Internet of Bodies (IoB) [ref1](#)

Internet of Everything (IoE) [ref1](#), [ref2](#)

Internet of Things (IoT) [ref1](#), [ref2](#)

lockdowns [ref1](#)

Psyop (psychological operation), Covid as a [ref1](#)
trolls [ref1](#)

intersectionality [ref1](#)

inversion

Archons [ref1](#), [ref2](#), [ref3](#)

climate change hoax [ref1](#)

energy [ref1](#)

Judaism [ref1](#), [ref2](#), [ref3](#)

symbolism [ref1](#)

Wetiko factor [ref1](#)

Wokeness [ref1](#), [ref2](#), [ref3](#)

Islam

Archons [ref1](#)

crypto-Jews [ref1](#)

Islamic State [ref1](#), [ref2](#)

Jinn and Djinn [ref1](#), [ref2](#), [ref3](#)

Ottoman Empire [ref1](#)

Wahhabism [ref1](#)

isolation *see* **social distancing** *and* **isolation**

Israel

China [ref1](#)

Cyber Intelligence Unit Beersheba complex [ref1](#)

expansion of illegal settlements [ref1](#)

formation [ref1](#)
Global Cult [ref1](#)
Judaism [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
medical experiments, consent for [ref1](#)
Mossad [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Palestine-Israel conflict [ref1](#), [ref2](#), [ref3](#)
parents, replacement of [ref1](#)
Sabbatians [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
September 11, 2001, terrorist attacks on United States [ref1](#)
Silicon Valley [ref1](#)
Smart Grid [ref1](#), [ref2](#)
United States [ref1](#), [ref2](#)
vaccines [ref1](#)
Wetiko factor [ref1](#)

Italy

fear [ref1](#), [ref2](#), [ref3](#)
Lombardy [ref1](#), [ref2](#), [ref3](#)
vaccines [ref1](#)

J

Johns Hopkins University [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
Johnson, Boris [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)
Joint Biosecurity Centre (JBC) [ref1](#)

Judaism

anti-Semitism [ref1](#), [ref2](#), [ref3](#)
Archons [ref1](#), [ref2](#)
crypto-Jews [ref1](#)
inversion [ref1](#), [ref2](#), [ref3](#)
Israel [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Labour Party [ref1](#)
Nazi Germany [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Sabbatians [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Silicon Valley [ref1](#)
Torah [ref1](#)
United States [ref1](#), [ref2](#)
Zionists [ref1](#), [ref2](#), [ref3](#)

K

Kaufman, Andrew [ref1](#), [ref2](#), [ref3](#), [ref4](#)
knowledge [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
Koch's postulates [ref1](#)
Kurzweil, Ray [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
Kushner, Jared [ref1](#), [ref2](#)

L

Labour Party [ref1](#), [ref2](#)
Lanka, Stefan [ref1](#), [ref2](#)
Lateral Flow Device (LFD) [ref1](#)
Levy, Paul [ref1](#), [ref2](#), [ref3](#)
Life Program [ref1](#)
lockdowns [ref1](#), [ref2](#), [ref3](#)
 amplification tampering [ref1](#)
 Archons [ref1](#)
 Behavioural Insights Team [ref1](#)
 Black Lives Matter (BLM) [ref1](#)
 care homes, deaths in [ref1](#)
 children
abuse [ref1](#), [ref2](#)
mental health [ref1](#)
 China [ref1](#), [ref2](#)
 computer models [ref1](#)
 consequences [ref1](#), [ref2](#)
 dependency [ref1](#), [ref2](#), [ref3](#)

domestic abuse [ref1](#)
fall in cases [ref1](#)
fear [ref1](#), [ref2](#), [ref3](#), [ref4](#)
guaranteed income [ref1](#)
Hunger Games Society [ref1](#), [ref2](#), [ref3](#)
interaction, destroying [ref1](#)
Internet [ref1](#), [ref2](#)
overdoses [ref1](#)
perception [ref1](#)
police-military state [ref1](#), [ref2](#)
protests [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
psychopathic personality [ref1](#), [ref2](#), [ref3](#)
reporting/snitching, encouragement of [ref1](#), [ref2](#)
testing [ref1](#)
vaccines [ref1](#)
Wetiko factor [ref1](#)
WHO [ref1](#)
love [ref1](#), [ref2](#), [ref3](#)
Lucifer [ref1](#), [ref2](#), [ref3](#)

M

Madej, Carrie [ref1](#), [ref2](#)
Magufuli, John [ref1](#), [ref2](#)
mainstream media [ref1](#)
BBC [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)
censorship [ref1](#), [ref2](#)
China [ref1](#)
climate change hoax [ref1](#)
fear [ref1](#), [ref2](#)
Global Cult [ref1](#), [ref2](#)
independent journalism, lack of [ref1](#)
Ofcom [ref1](#), [ref2](#), [ref3](#)

perception [ref1](#), [ref2](#)

Psyop (psychological operation), Covid as a [ref1](#)

Sabbatians [ref1](#), [ref2](#)

social disapproval [ref1](#)

social distancing and isolation [ref1](#)

United States [ref1](#), [ref2](#)

vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Mao Zedong [ref1](#), [ref2](#), [ref3](#)

Marx and Marxism [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)

masculinity [ref1](#)

masks/face coverings [ref1](#), [ref2](#), [ref3](#)

 censorship [ref1](#)

 children [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

 China, made in [ref1](#)

 dehumanisation [ref1](#), [ref2](#), [ref3](#)

 fear [ref1](#), [ref2](#)

 flu [ref1](#)

 health professionals [ref1](#), [ref2](#), [ref3](#), [ref4](#)

 isolation [ref1](#)

 laughter [ref1](#)

mass non-cooperation [ref1](#)

microplastics, risk of [ref1](#)

mind control [ref1](#)

multiple masks [ref1](#)

 oxygen deficiency [ref1](#), [ref2](#), [ref3](#)

 police [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

 pollution, as cause of plastic [ref1](#)

 Psyop (psychological operation), Covid as a [ref1](#)

 reframing [ref1](#), [ref2](#)

 risk assessments, lack of [ref1](#), [ref2](#)

 self-respect [ref1](#)

 surgeons [ref1](#)

United States [ref1](#)
vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Wetiko factor [ref1](#)
'worms' [ref1](#)
The Matrix movies [ref1](#), [ref2](#), [ref3](#)
measles [ref1](#), [ref2](#)
media see mainstream media
Medicines and Healthcare products Regulatory Agency (MHRA)
[ref1](#), [ref2](#), [ref3](#), [ref4](#)
Mesopotamia [ref1](#)
messaging [ref1](#)
military-police state [ref1](#), [ref2](#), [ref3](#)
mind control [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#) *see also* MKUltra
MKUltra [ref1](#), [ref2](#), [ref3](#)
monarchy [ref1](#)
money *see* banking, finance and money
Montagnier, Luc [ref1](#), [ref2](#), [ref3](#)
Mooney, Bel [ref1](#)
Morgellons disease [ref1](#), [ref2](#)
mortality rate [ref1](#)
Mullis, Kary [ref1](#), [ref2](#), [ref3](#)
Musk, Elon [ref1](#)

N

Nag Hammadi texts [ref1](#), [ref2](#), [ref3](#)
nanotechnology [ref1](#), [ref2](#), [ref3](#)
narcissism [ref1](#)
Nazi Germany [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)
near-death experiences [ref1](#), [ref2](#)
Neocons [ref1](#), [ref2](#), [ref3](#)

Neuro-Linguistic Programming (NLP) and the Delphi technique
[ref1](#)

NHS (National Health Service)

amplification cycles [ref1](#)

Common Purpose [ref1](#), [ref2](#)

mind control [ref1](#)

NHS England [ref1](#)

saving the NHS [ref1](#), [ref2](#)

vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

whistle-blowers [ref1](#), [ref2](#), [ref3](#)

No-Problem-Reaction-Solution [ref1](#), [ref2](#), [ref3](#), [ref4](#)

non-human dimension of Global Cult [ref1](#)

nous [ref1](#)

numbers, reality as [ref1](#)

Nuremberg Codes [ref1](#), [ref2](#), [ref3](#)

Nuremberg-like tribunal, proposal for [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#),
[ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#), [ref11](#), [ref12](#)

O

Obama, Barack [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#)

O'Brien, Cathy [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Ochel, Evita [ref1](#)

Ofcom [ref1](#), [ref2](#), [ref3](#)

old people [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Oneness [ref1](#), [ref2](#), [ref3](#)

Open Society Foundations (Soros) [ref1](#), [ref2](#), [ref3](#)

oxygen 406, 528–34

P

paedophilia [ref1](#), [ref2](#)

Page, Larry [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

Palestine-Israel conflict [ref1](#), [ref2](#), [ref3](#)

pandemic, definition of [ref1](#)

pandemic and health crisis scenarios/simulations [ref1](#), [ref2](#), [ref3](#),
[ref4](#)

paranormal [ref1](#)

PCR tests [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

Pearl Harbor attacks, prior knowledge of [ref1](#)

Pelosi, Nancy [ref1](#), [ref2](#), [ref3](#)

perception [ref1](#), [ref2](#), [ref3](#), [ref4](#)

climate change hoax [ref1](#)

control [ref1](#), [ref2](#), [ref3](#)

decoding [ref1](#), [ref2](#)

enslavement [ref1](#)

externally-delivered perceptions [ref1](#)

five senses [ref1](#)

human labels [ref1](#)

media [ref1](#), [ref2](#)

political parties [ref1](#), [ref2](#)

Psyop (psychological operation), Covid as a [ref1](#)

sale of perception [ref1](#)

self-identity [ref1](#), [ref2](#)

Wokeness [ref1](#)

Phantom Self [ref1](#), [ref2](#), [ref3](#)

pharmaceutical industry *see* **Big Pharma**

phthalates [ref1](#)

Plato's Allegory of the Cave [ref1](#), [ref2](#)

pneuma [ref1](#)

police

Black Lives Matter (BLM) [ref1](#)

brutality [ref1](#)

citizen's arrests [ref1](#), [ref2](#)

common law arrests [ref1](#), [ref2](#)

Common Purpose [ref1](#)

defunding [ref1](#)

lockdowns [ref1](#), [ref2](#)

masks [ref1](#), [ref2](#), [ref3](#), [ref4](#)

police-military state [ref1](#), [ref2](#), [ref3](#)

psychopathic personality [ref1](#), [ref2](#), [ref3](#), [ref4](#)

reframing [ref1](#)

United States [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Wokeness [ref1](#)

polio [ref1](#)

political correctness [ref1](#), [ref2](#), [ref3](#), [ref4](#)

political parties [ref1](#), [ref2](#), [ref3](#), [ref4](#)

political puppets [ref1](#)

pollution [ref1](#), [ref2](#), [ref3](#)

post-mortems/autopsies [ref1](#)

Postage Stamp Consensus [ref1](#), [ref2](#)

pre-emptive programming [ref1](#)

Problem-Reaction-Solution [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

Project for the New American Century [ref1](#), [ref2](#), [ref3](#), [ref4](#)

psychopathic personality [ref1](#)

Archons [ref1](#)

heart energy [ref1](#)

lockdowns [ref1](#), [ref2](#), [ref3](#)

police [ref1](#), [ref2](#), [ref3](#), [ref4](#)

recruitment [ref1](#), [ref2](#)

vaccines [ref1](#)

wealth [ref1](#)

Wetiko [ref1](#), [ref2](#)

Psyop (psychological operation), Covid as a [ref1](#), [ref2](#), [ref3](#), [ref4](#),
[ref5](#)

Pushbackers [ref1](#), [ref2](#), [ref3](#), [ref4](#)

pyramid structure [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Q

QAnon Psyop [ref1](#), [ref2](#), [ref3](#)

R

racism *see also* **Black Lives**

Matter (BLM)

anti-racism industry [ref1](#)

class [ref1](#)

critical race theory [ref1](#)

culture [ref1](#)

intersectionality [ref1](#)

reverse racism [ref1](#)

white privilege [ref1](#), [ref2](#)

white supremacy [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Wokeness [ref1](#), [ref2](#), [ref3](#)

radiation [ref1](#), [ref2](#)

randomness, illusion of [ref1](#), [ref2](#), [ref3](#)

reality [ref1](#), [ref2](#), [ref3](#)

reframing [ref1](#), [ref2](#)

change agents [ref1](#), [ref2](#)

children [ref1](#)

climate change [ref1](#)

Common Purpose leadership programme [ref1](#), [ref2](#)

contradictory rules [ref1](#)

enforcers [ref1](#)

masks [ref1](#), [ref2](#)

NLP and the Delphi technique [ref1](#)

police [ref1](#)

Wetiko factor [ref1](#)

Wokeness [ref1](#), [ref2](#)

religion *see also* particular religions

alien invasions [ref1](#)

Archons [ref1](#), [ref2](#)
consciousness [ref1](#), [ref2](#)
control, system of [ref1](#), [ref2](#), [ref3](#)
criticism, prohibition on [ref1](#)
five senses [ref1](#)
good and evil, war between [ref1](#)
hidden non-human forces [ref1](#), [ref2](#)
Sabbatians [ref1](#)
save me syndrome [ref1](#)
Wetiko [ref1](#)
Wokeness [ref1](#)
repetition and mind control [ref1](#), [ref2](#), [ref3](#)
reporting/snitching, encouragement of [ref1](#), [ref2](#)
Reptilians/Grey entities [ref1](#)
rewiring the mind [ref1](#)
Rivers, Thomas Milton [ref1](#), [ref2](#)
Rockefeller family [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#)
Rockefeller Foundation documents [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Roman Empire [ref1](#)
Rothschild family [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#)
RT-PCR tests [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)
Russia
 collusion inquiry in US [ref1](#)
Russian Revolution [ref1](#), [ref2](#)
Sabbatians [ref1](#)

S

Sabbatian-Frankism [ref1](#), [ref2](#)
 anti-Semitism [ref1](#), [ref2](#)
 banking and finance [ref1](#), [ref2](#), [ref3](#)
 China [ref1](#), [ref2](#)
 Israel [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Judaism [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Lucifer [ref1](#)
media [ref1](#), [ref2](#)
Nazis [ref1](#), [ref2](#)
QAnon [ref1](#)
Rothschilds [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
Russia [ref1](#)
Saudi Arabia [ref1](#)
Silicon Valley [ref1](#)
Sumer [ref1](#)
United States [ref1](#), [ref2](#), [ref3](#)
Wetiko factor [ref1](#)
Wokeness [ref1](#), [ref2](#), [ref3](#)
SAGE (Scientific Advisory Group for Emergencies) [ref1](#), [ref2](#), [ref3](#),
[ref4](#)
SARS-1 [ref1](#)
SARs-CoV-2 [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)
Satan/Satanism [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
satellites in low-orbit [ref1](#)
Saudi Arabia [ref1](#)
Save Me Syndrome [ref1](#)
scapegoating [ref1](#)
Schwab, Klaus [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#),
[ref11](#), [ref12](#)
science, manipulation of [ref1](#)
self-identity [ref1](#), [ref2](#), [ref3](#), [ref4](#)
self-respect, attacks on [ref1](#)
September 11, 2001, terrorist attacks on United States [ref1](#), [ref2](#),
[ref3](#), [ref4](#)
77th Brigade of UK military [ref1](#), [ref2](#), [ref3](#)
Silicon Valley/tech giants [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#) *see also*
Facebook

Israel [ref1](#)

Sabbatians [ref1](#)

technocracy [ref1](#)

Wetiko factor [ref1](#)

Wokeness [ref1](#)

simulation hypothesis [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

Smart Grid [ref1](#), [ref2](#), [ref3](#)

artificial intelligence (AI) [ref1](#)

China [ref1](#), [ref2](#)

control centres [ref1](#)

the Field [ref1](#)

Great Reset [ref1](#)

Human 2.0 [ref1](#), [ref2](#)

Israel [ref1](#), [ref2](#)

vaccines [ref1](#)

Wetiko factor [ref1](#)

social disapproval [ref1](#)

social distancing and isolation [ref1](#), [ref2](#), [ref3](#)

abusive relationships [ref1](#), [ref2](#)

children [ref1](#)

flats and apartments [ref1](#)

heart issues [ref1](#)

hugs [ref1](#)

Internet [ref1](#)

masks [ref1](#)

media [ref1](#)

older people [ref1](#), [ref2](#)

one-metre (three feet) rule [ref1](#)

rewiring the mind [ref1](#)

simulation, universe as a [ref1](#)

SPI-B [ref1](#)

substance abuse [ref1](#)

suicide and self-harm [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
technology [ref1](#)
torture, as [ref1](#), [ref2](#)
two-metre (six feet) rule [ref1](#)
women [ref1](#)
social justice [ref1](#), [ref2](#), [ref3](#), [ref4](#)
social media *see also* **Facebook bans on alternative views** [ref1](#)
censorship [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
children [ref1](#)
emotion [ref1](#)
perception [ref1](#)
private messages [ref1](#)
Twitter [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
Wetiko factor [ref1](#)
YouTube [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Soros, George [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
Spain [ref1](#)
SPI-B (Scientific Pandemic Insights Group on Behaviours) [ref1](#),
[ref2](#), [ref3](#), [ref4](#)
spider and the web [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Starmer, Keir [ref1](#)
Statute Law [ref1](#)
Steiner, Rudolf [ref1](#), [ref2](#), [ref3](#)
Stockholm syndrome [ref1](#)
streptomycin [ref1](#)
suicide and self-harm [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Sumer [ref1](#), [ref2](#)
Sunstein, Cass [ref1](#), [ref2](#), [ref3](#)
swine flu (H1N1) [ref1](#), [ref2](#), [ref3](#)
synchronicity [ref1](#)
synthetic biology [ref1](#), [ref2](#), [ref3](#), [ref4](#)
synthetic meat [ref1](#), [ref2](#)

T

technology *see also* **artificial intelligence (AI)**; **Internet**;

social media addiction [ref1](#), [ref2](#), [ref3](#), [ref4](#)

Archons [ref1](#), [ref2](#)

the cloud [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

cyber-operations [ref1](#)

cyberwarfare [ref1](#)

radiation [ref1](#), [ref2](#)

social distancing and isolation [ref1](#)

technocracy [ref1](#)

Tedros Adhanom Ghebreyesus [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#),
[ref8](#), [ref9](#), [ref10](#), [ref11](#), [ref12](#), [ref13](#)

telepathy [ref1](#)

Tenpenny, Sherri [ref1](#)

Tesla, Nikola [ref1](#)

testosterone levels, decrease in [ref1](#)

testing for Covid-19 [ref1](#), [ref2](#)

anal swab tests [ref1](#)

cancer [ref1](#)

China [ref1](#), [ref2](#), [ref3](#)

Corman-Drosten test [ref1](#), [ref2](#), [ref3](#), [ref4](#)

death certificates [ref1](#), [ref2](#)

fraudulent testing [ref1](#)

genetic material, amplification of [ref1](#)

Lateral Flow Device (LFD) [ref1](#)

PCR tests [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

vaccines [ref1](#), [ref2](#), [ref3](#)

Thunberg, Greta [ref1](#), [ref2](#), [ref3](#)

Totalitarian Tiptoe [ref1](#), [ref2](#), [ref3](#), [ref4](#)

transgender persons

activism [ref1](#)

artificial wombs [ref1](#)

censorship [ref1](#)
child abuse [ref1](#), [ref2](#)
Human 2.0 [ref1](#), [ref2](#), [ref3](#)
Wokeness [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
women, deletion of rights and status of [ref1](#), [ref2](#)
young persons [ref1](#)
travel restrictions [ref1](#)
Trudeau, Justin [ref1](#), [ref2](#), [ref3](#)
Trump, Donald [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#),
[ref11](#)
Twitter [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)

U

UKColumn [ref1](#), [ref2](#)
United Nations (UN) [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#) *see also* **Agenda 21/Agenda 2030 (UN)**
United States [ref1](#), [ref2](#)
American Revolution [ref1](#)
borders [ref1](#), [ref2](#)
Capitol Hill riot [ref1](#), [ref2](#)
children [ref1](#)
China [ref1](#), [ref2](#)
CIA [ref1](#), [ref2](#)
Daily Pass tracking system [ref1](#)
demographics by immigration, changes in [ref1](#)
Democrats [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
election fraud [ref1](#)
far-right domestic terrorists, pushbackers as [ref1](#)
Federal Reserve [ref1](#)
flu/respiratory diseases statistics [ref1](#)
Global Cult [ref1](#), [ref2](#)
hand sanitisers, FDA warnings on [ref1](#)

immigration, effects of illegal [ref1](#)
impeachment [ref1](#)
Israel [ref1](#), [ref2](#)
Judaism [ref1](#), [ref2](#), [ref3](#)
lockdown [ref1](#)
masks [ref1](#)
mass media [ref1](#), [ref2](#)
nursing homes [ref1](#)
Pentagon [ref1](#), [ref2](#), [ref3](#), [ref4](#)
police [ref1](#), [ref2](#), [ref3](#), [ref4](#)
pushbackers [ref1](#)
Republicans [ref1](#), [ref2](#)
borders [ref1](#), [ref2](#)
Democrats [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Russia, inquiry into collusion with [ref1](#)
Sabbatians [ref1](#), [ref2](#), [ref3](#)
September 11, 2001, terrorist attacks [ref1](#), [ref2](#), [ref3](#), [ref4](#)
UFO sightings, release of information on [ref1](#)
vaccines [ref1](#)
white supremacy [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Woke Democrats [ref1](#), [ref2](#)

V

vaccines [ref1](#), [ref2](#), [ref3](#)
adverse reactions [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Africa [ref1](#)
anaphylactic shock [ref1](#), [ref2](#), [ref3](#), [ref4](#)
animals [ref1](#), [ref2](#)
anti-vax movement [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
AstraZeneca/Oxford [ref1](#), [ref2](#), [ref3](#), [ref4](#)
autoimmune diseases, rise in [ref1](#), [ref2](#)
Big Pharma [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#)

bioweapon, as real [ref1](#), [ref2](#)
black and ethnic minority communities [ref1](#)
blood clots [ref1](#), [ref2](#)
Brain Computer Interface (BCI) [ref1](#)
care homes, deaths in [ref1](#)
censorship [ref1](#), [ref2](#), [ref3](#)
chief medical officers and scientific advisers, financial interests of
[ref1](#), [ref2](#)
children [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#), [ref10](#)
China [ref1](#), [ref2](#)
clinical trials [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
compensation [ref1](#)
compulsory vaccinations [ref1](#), [ref2](#), [ref3](#)
computer programs [ref1](#)
consciousness [ref1](#)
cover-ups [ref1](#)
creation before Covid [ref1](#)
cytokine storm [ref1](#)
deaths and illnesses caused by vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
definition [ref1](#)
developing countries [ref1](#)
digital tattoos [ref1](#)
DNA-manipulation [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#),
[ref10](#)
emergency approval [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
female infertility [ref1](#)
funding [ref1](#)
genetic suicide [ref1](#)
Global Cult [ref1](#)
heart chakras [ref1](#)
hesitancy [ref1](#)
Human 2.0 [ref1](#), [ref2](#), [ref3](#), [ref4](#)
immunity from prosecution [ref1](#), [ref2](#), [ref3](#)

implantable technology [ref1](#)
Israel [ref1](#)
Johnson & Johnson [ref1](#), [ref2](#), [ref3](#), [ref4](#)
lockdowns [ref1](#)
long-term effects [ref1](#)
mainstream media [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
masks [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Medicines and Healthcare products Regulatory Agency (MHRA)
[ref1](#), [ref2](#)
messaging [ref1](#)
Moderna [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
mRNA vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#)
nanotechnology [ref1](#), [ref2](#)
NHS [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
older people [ref1](#), [ref2](#)
operating system [ref1](#)
passports [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Pfizer/BioNTech [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
polyethylene glycol [ref1](#)
pregnant women [ref1](#)
psychopathic personality [ref1](#)
races, targeting different [ref1](#)
reverse transcription [ref1](#)
Smart Grid [ref1](#)
social distancing [ref1](#)
social media [ref1](#)
sterility [ref1](#)
synthetic material, introduction of [ref1](#)
tests [ref1](#), [ref2](#), [ref3](#)
travel restrictions [ref1](#)
variants [ref1](#), [ref2](#)
viruses, existence of [ref1](#)
whistle-blowing [ref1](#)

WHO [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Wokeness [ref1](#)
working, vaccine as [ref1](#)
young people [ref1](#)
Vallance, Patrick [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#)
variants [ref1](#), [ref2](#), [ref3](#)
vegans [ref1](#)
ventilators [ref1](#), [ref2](#)
virology [ref1](#), [ref2](#)
virtual reality [ref1](#), [ref2](#), [ref3](#)
viruses, existence of [ref1](#)
visual reality [ref1](#), [ref2](#)
vitamin D [ref1](#), [ref2](#)
von Braun, Wernher [ref1](#), [ref2](#)

W

war-zone hospital myths [ref1](#)
waveforms [ref1](#), [ref2](#)
wealth [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#) [ref10](#), [ref11](#)
wet market conspiracy [ref1](#)
Wetiko factor [ref1](#)
 alcoholism and drug addiction [ref1](#)
 anti-human, why Global Cult is [ref1](#)
 Archons [ref1](#), [ref2](#), [ref3](#), [ref4](#)
 artificial intelligence (AI) [ref1](#)
 Big Pharma [ref1](#), [ref2](#)
 children [ref1](#)
 China [ref1](#)
 consciousness [ref1](#), [ref2](#)
 education [ref1](#)
 Facebook [ref1](#)

fear [ref1](#), [ref2](#)
frequency [ref1](#), [ref2](#)
Gates [ref1](#), [ref2](#)
Global Cult [ref1](#), [ref2](#)
heart [ref1](#), [ref2](#)
lockdowns [ref1](#)
masks [ref1](#)
Native American concept [ref1](#)
psychopathic personality [ref1](#), [ref2](#)
reframing/retraining programmes [ref1](#)
religion [ref1](#)
Silicon Valley [ref1](#)
Smart Grid [ref1](#)
smartphone addiction [ref1](#), [ref2](#)
social media [ref1](#)
war [ref1](#), [ref2](#)
WHO [ref1](#)
Wokeness [ref1](#), [ref2](#), [ref3](#)
Yaldabaoth [ref1](#), [ref2](#), [ref3](#), [ref4](#)
whistle-blowing [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
white privilege [ref1](#), [ref2](#)
white supremacy [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
Whitty, Christopher [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#),
[ref10](#)
‘who benefits’ [ref1](#)
Wi-Fi [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Wikipedia [ref1](#), [ref2](#)
Wojcicki, Susan [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#)
Wokeness
 Antifa [ref1](#), [ref2](#), [ref3](#), [ref4](#)
 anti-Semitism [ref1](#)
 billionaire social justice warriors [ref1](#), [ref2](#), [ref3](#)

Capitol Hill riot [ref1](#), [ref2](#)
censorship [ref1](#)
Christianity [ref1](#)
climate change hoax [ref1](#), [ref2](#)
culture [ref1](#)
education, control of [ref1](#)
emotion [ref1](#)
facts [ref1](#)
fascism [ref1](#), [ref2](#), [ref3](#)
Global Cult [ref1](#), [ref2](#), [ref3](#), [ref4](#)
group-think [ref1](#)
immigration [ref1](#)
indigenous people, solidarity with [ref1](#)
inversion [ref1](#), [ref2](#), [ref3](#)
left, hijacking the [ref1](#), [ref2](#)
Marxism [ref1](#), [ref2](#), [ref3](#)
mind control [ref1](#)
New Woke [ref1](#)
Old Woke [ref1](#)
Oneness [ref1](#)
perceptual programming [ref1](#)
 Phantom Self [ref1](#)
police [ref1](#)
defunding the [ref1](#)
reframing [ref1](#)
public institutions [ref1](#)
Pushbackers [ref1](#), [ref2](#), [ref3](#)
racism [ref1](#), [ref2](#), [ref3](#)
reframing [ref1](#), [ref2](#)
religion, as [ref1](#)
Sabbatians [ref1](#), [ref2](#), [ref3](#)
Silicon Valley [ref1](#)
social justice [ref1](#), [ref2](#), [ref3](#), [ref4](#)

transgender [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)
United States [ref1](#), [ref2](#)
vaccines [ref1](#)
Wetiko factor [ref1](#), [ref2](#), [ref3](#)
young people [ref1](#), [ref2](#), [ref3](#)
women, deletion of rights and status of [ref1](#), [ref2](#)
World Economic Forum (WEF) [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#),
[ref8](#), [ref9](#)
World Health Organization (WHO) [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#),
[ref7](#), [ref8](#), [ref9](#)
AIDs/HIV [ref1](#)
amplification cycles [ref1](#)
Big Pharma [ref1](#), [ref2](#), [ref3](#)
cooperation in health emergencies [ref1](#)
creation [ref1](#), [ref2](#)
fatality rate [ref1](#)
funding [ref1](#), [ref2](#), [ref3](#)
Gates [ref1](#)
Internet [ref1](#)
lockdown [ref1](#)
vaccines [ref1](#), [ref2](#), [ref3](#), [ref4](#)
Wetiko factor [ref1](#)
world number 1 (masses) [ref1](#), [ref2](#)
world number 2 [ref1](#)
Wuhan [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#) [ref8](#)

Y

Yaldabaoth [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#)
Yeadon, Michael [ref1](#), [ref2](#), [ref3](#), [ref4](#)
young people *see also* children addiction to technology [ref1](#)
Human 2.0 [ref1](#)
vaccines [ref1](#), [ref2](#)

Wokeness [ref1](#), [ref2](#), [ref3](#)

YouTube [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#)

WHO 548

Z

Zaks, Tal [ref1](#)

Zionism [ref1](#), [ref2](#), [ref3](#)

Zuckerberg, Mark [ref1](#), [ref2](#), [ref3](#), [ref4](#), [ref5](#), [ref6](#), [ref7](#), [ref8](#), [ref9](#),
[ref10](#), [ref11](#), [ref12](#)

Zulus [ref1](#)



Ickonic is something that has been a dream of mine for the last 5 years, growing up around alternative information I have always had a natural interest in what is going on in the World and what could I do to make it better.

Across the range of subjects and positions of influence occupied mainly by people who don't strive to make things better it's the Media that I have always found the most frustrating and fascinating. Mainly because if the Media did their Jobs properly then so much of the negative things happening in the World simply would not be able to happen, because they would be exposed within a heartbeat.

Free Press and the Opportunities that the internet could have given would mean that the Media are able to expose things like never before and hold people to account for their actions. As we all know there are 'Untouchables' that walk among us, people the Media simply won't touch, expose or investigate and that leads to the dark underworlds that infest the establishment the World over. Well I say enough, it's time for something different, a different kind of Media, where no one is off limits from exposing and investigating. All we're interested in at Ickonic is the truth of what is really going on in the World on whichever subject we're covering.

We hope you enjoy what we have created and take something away from the platform, we aim to deliver information that's informative and most importantly self-empowering, you're not a little person, you're part of something much bigger than that and its time we as a collective race began to understand that and look to the future as ours to take.

It's time...

Jaymie Icke - Founder Ickonic Alternative Media.

SIGN UP NOW AT ICKONIC.COM

DAVID ICKE
THE ANSWER



We live in extraordinary times with billions bewildered and seeking answers for what is happening. David Icke, the man who has been proved right again and again, has spent 30 years uncovering the truth behind world affairs and in a stream of previous books he predicted current events.

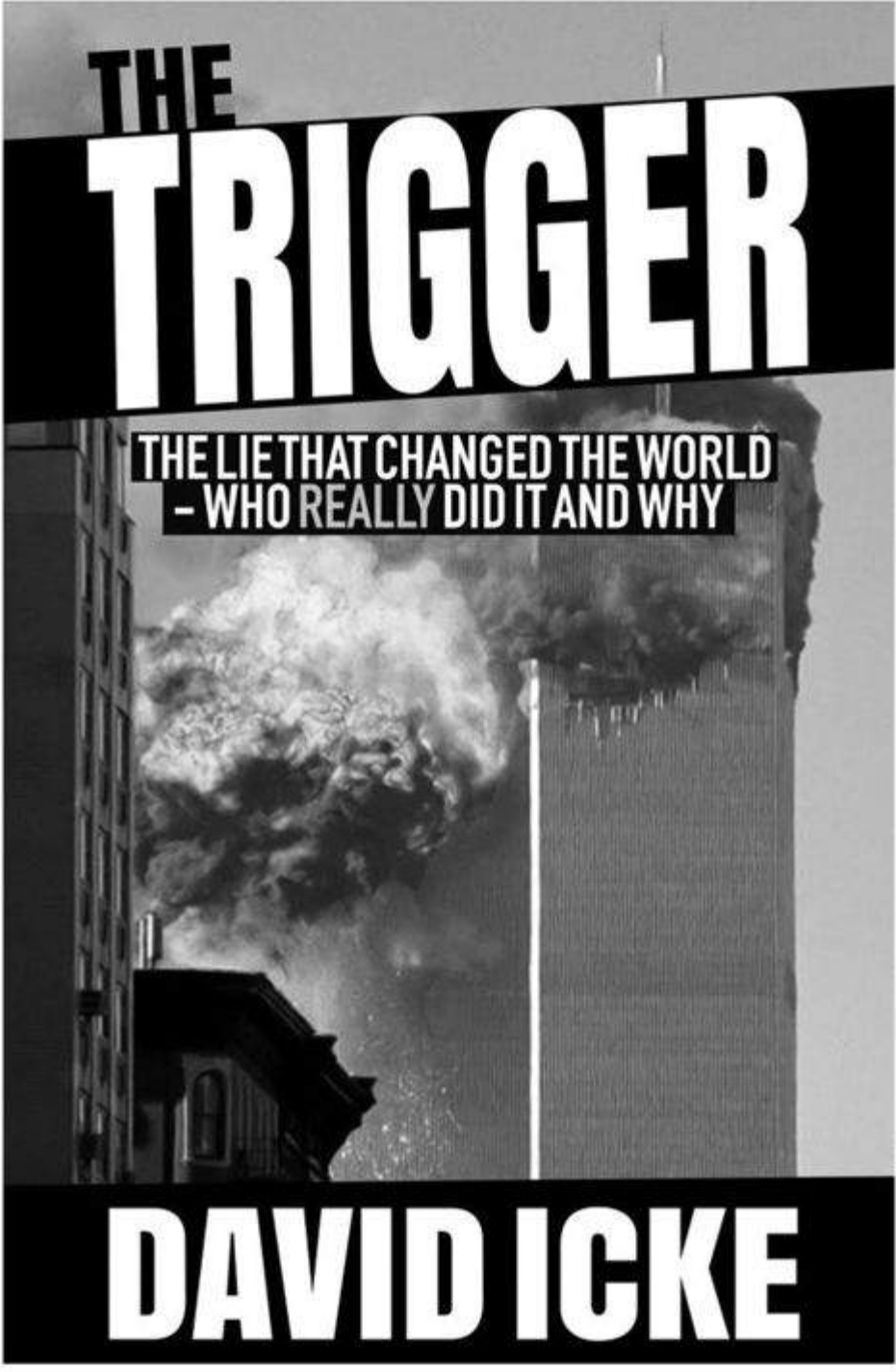
The Answer will change your every perception of life and the world and set you free of the illusions that control human society. There is nothing more vital for our collective freedom than humanity becoming aware of what is in this book.

Available now at davidicke.com.

THE

TRIGGER

**THE LIE THAT CHANGED THE WORLD
- WHO REALLY DID IT AND WHY**



DAVID ICKE

**EVERYTHING
YOU NEED
TO KNOW**

BUT HAVE NEVER BEEN TOLD

DAVID ICKE

DAVIDICKE.COM



DAVID ICKE STORE
LATEST NEWS ARTICLES
DAVID ICKE VIDEOS
WEEKLY DOT-CONNECTOR PODCASTS
LIVE EVENTS

WWW.DAVIDICKE.COM

364

THE LIFE STORY OF DAVID ICKE

RENEGADE

THE FEATURE LENGTH FILM

/ren·i·geɪd/

noun

A person who behaves in a rebelliously unconventional manner.



AVAILABLE NOW AT DAVIDICKE.COM

2 NEW BOOKS
BY NEIL HAGUE

ORION'S DOOR

SYMBOLS OF CONSCIOUSNESS & BLUEPRINTS OF CONTROL
- THE STORY OF ORION'S INFLUENCE OVER HUMANITY

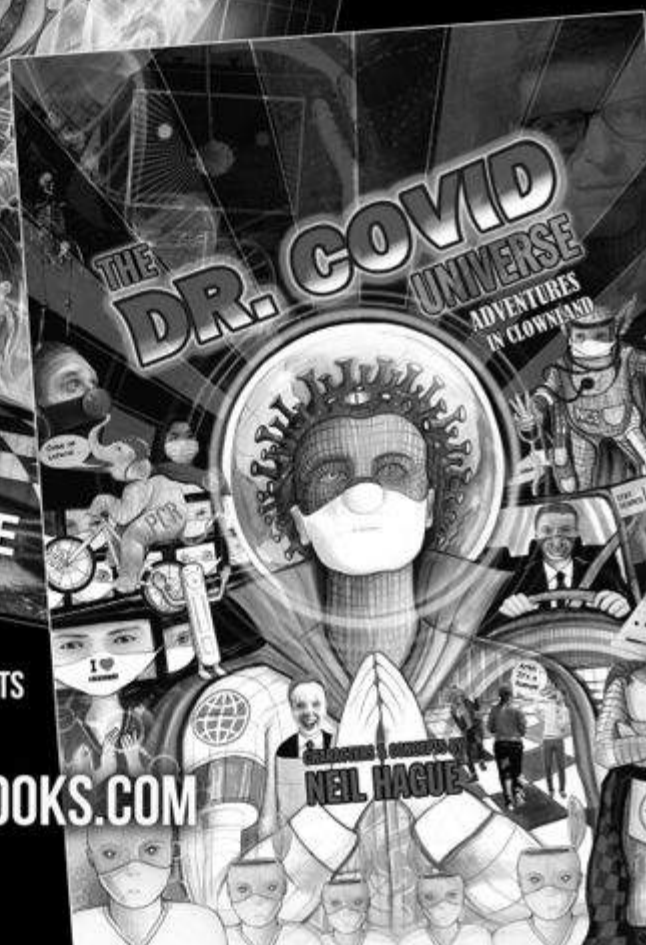
CUTTING EDGE VISIONARY ART
& UNIQUE ILLUSTRATED BOOKS

NEIL HAGUE

FOR
BOOKS, PRINTS & T-SHIRTS
VISIT:

NEILHAGUEBOOKS.COM

OR NEILHAGUE.COM



Before you go ...

For more detail, background and evidence about the subjects in *Perceptions of a Renegade Mind* – and so much more – see my others books including *And The Truth Shall Set You Free*; *The Biggest Secret*; *Children of the Matrix*; *The David Icke Guide to the Global Conspiracy*; *Tales from the Time Loop*; *The Perception Deception*; *Remember Who You Are*; *Human Race Get Off Your Knees*; *Phantom Self*; *Everything You Need To Know But Have Never Been Told*, *The Trigger* and *The Answer*.

You can subscribe to the fantastic new Ickonic media platform where there are many hundreds of hours of cutting-edge information in videos, documentaries and series across a whole range of subjects which are added to every week. This includes my 90 minute breakdown of the week's news every Friday to explain *why* events are happening and to what end.